



AUDiPY

Analyzer HANDBUCH

Start

Explorer

Data Tools

Speichern

Zurück

Zurücksetzen

Lesenzeichen

Audit Steps

Ask my Data

Visualisieren

Analysieren

AUDiPY

... Superstore

	ROW ID	ORDER ID	ORDER DATE	SHIP DATE	SHIP MODE	CUSTOMER ID	CUSTOMER NAME	SEGMENT	COUNTRY	CITY	STATE	POSTAL CODE	REGION	PROC
1	1	CA-2016-152156	08.11.2016	11.11.2016	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BD-100
2	2	CA-2016-152156	08.11.2016	11.11.2016	Second Class	CG-12520	Claire Gule	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-100
3	3	CA-2016-138688	12.06.2016	16.06.2016	Second Class	DW-13045	Darin Van Hull	Corporate	United States	Los Angeles	California	90036	West	OFF-IA-100
4	4	US-2015-108966	11.10.2015	18.10.2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-100
5	5	US-2015-108966	11.10.2015	18.10.2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-SI-100
6	6	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-100
7	7	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-100
8	8	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	ILL-PH-100
9	9	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-100
10	10	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-100
11	11	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-100
12	12	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-100
13	13	CA-2017-144412	15.04.2017	20.04.2017	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	OFF-PA-100
14	14	CA-2016-161389	05.12.2016	10.12.2016	Standard Class	RA-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98103	West	OFF-AB-100
15	15	US-2015-118983	22.11.2015	26.11.2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-100
16	16	US-2015-118983	22.11.2015	26.11.2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-100
17	17	CA-2014-105891	11.11.2014	18.11.2014	Standard Class	PK-19075	Pete Kitz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-100
18	18	CA-2014-161764	13.05.2014	15.05.2014	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jordan	Utah	84064	West	OFF-SI-100
19	19	CA-2014-143336	27.08.2014	01.09.2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-AR-100
20	20	CA-2014-143336	27.08.2014	01.09.2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	TEC-PH-100
21	21	CA-2014-143336	27.08.2014	01.09.2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-AB-100
22	22	CA-2016-137330	09.12.2016	13.12.2016	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AR-100
23	23	CA-2016-137330	09.12.2016	13.12.2016	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-100

▼

ROW ID

Exakte Übereinstimmung

Cc

✓

✗

Σ

🔄

📄

Disclaimer

AUDIPLY Analyzer ist ein leistungsfähiges Software-Tool zur Analyse und Visualisierung von strukturierten Datensätzen. Es ermöglicht Anwendern, große Datenmengen effizient zu durchsuchen, zu filtern und zu analysieren, um fundierte Entscheidungen auf Basis der gewonnenen Erkenntnisse zu treffen.

Dieses Handbuch wurde mit größter Sorgfalt erstellt. Dennoch können Fehler und Irrtümer nicht vollständig ausgeschlossen werden. Die Semitax GmbH übernimmt keine Gewähr für die Richtigkeit, Vollständigkeit und Aktualität der Inhalte.

Die in diesem Handbuch beschriebenen Funktionen und Eigenschaften der Software können ohne vorherige Ankündigung geändert werden. Alle in diesem Handbuch enthaltenen Informationen und Empfehlungen dienen ausschließlich zu Informationszwecken.

Für den Inhalt von externen Links übernehmen wir keine Verantwortung; für den Inhalt der verlinkten Seiten sind ausschließlich deren Betreiber verantwortlich.

Dieses Handbuch sowie die darin enthaltenen Texte, Bilder und Grafiken sind urheberrechtlich geschützt. Jede Vervielfältigung oder Verbreitung, auch auszugsweise, bedarf der schriftlichen Zustimmung der Semitax GmbH.

AUDIPLY

Semitax GmbH

© 2025 Semitax GmbH
Alle Rechte vorbehalten.

Kontakt: info@semitax.de
Website: www.semitax.de



Inhalt

1	Einstieg AUDIPY	12
1.1	Dateiformate im AUDIPY	12
1.2	Dateiexplorer.....	14
1.2.1	Obere Explorer-Symbole	14
1.2.1.1	Schließen des Explorers	15
1.2.1.2	Projekt wechseln	15
1.2.1.3	Navigation durchs Fallverzeichnis	16
1.2.1.4	Übergeordnete Ordner.....	16
1.2.1.5	Ordner anlegen	17
1.2.2	Rechtsklick Menü im Explorer.....	18
1.2.2.1	Öffnen (mit Spaltenauswahl)	19
1.2.2.2	Stream.....	21
1.2.2.2.1	Extraktion	23
1.2.2.2.2	Gruppieren	29
1.2.2.2.3	Verbinden.....	29
1.2.2.2.4	Anhängen	30
1.2.2.2.5	Pivot	30
1.2.2.3	Dateioperationen und Verwaltung.....	30
1.2.2.4	Kategorisierung und Statuskennzeichnung	32
1.2.3	Untere Explorer-Symbole	33
1.2.3.1	Spaltenauswahl.....	33
1.2.3.2	Stream.....	33
2	Hauptnavigation	34
2.2	Komprimierung:.....	35
3	Start	36
3.1	Öffnen.....	36
3.1.1	Datei öffnen.....	36
3.1.2	Ordner öffnen.....	37
3.1.3	Letzte Projekte.....	37
3.2	Importieren.....	38
3.2.1	Datei.....	38
3.2.2	Datenbank.....	41

3.2.3	Umwandlung von mehreren Dateien.....	42
3.2.4	Data-Wizard „Start“ – CSV / TXT.....	44
3.2.5	Data-Wizard „Optionen“.....	46
3.2.6	Data-Wizard „Spalten verbinden“.....	47
3.2.7	Data-Wizard „Erweitert“ - Streaming.....	48
3.2.8	Drucklisten (PDF / Text) – Report Wizard.....	49
3.2.8.1	Automatische Erkennung.....	49
3.2.8.1.1	Praktische Anwendungsszenarien.....	50
3.2.8.1.2	Blimport mit der automatischen Erkennung.....	51
3.2.8.1.3	Erweiterte Strategien und Optimierung.....	55
3.2.8.2	Manuelle Definition.....	57
3.2.8.2.1	Praktische Anwendungsszenarien.....	57
3.2.8.2.2	Bedienelemente und Konfiguration.....	59
3.2.8.2.3	Import mithilfe der manuellen Definition.....	62
3.2.8.2.4	Erweiterte Strategien und Optimierung.....	65
3.2.9	Excel Import.....	67
3.2.10	XML Import.....	70
3.2.11	GoBD Import.....	72
3.3	Power BI verbinden.....	75
3.4	Projekt wechseln.....	82
3.5	Speichern.....	82
3.6	Speichern unter.....	82
3.7	Exportieren.....	82
3.8	Berichte.....	84
3.8.1	Bericht.....	84
3.8.2	Drucken.....	85
3.8.3	Verlauf.....	87
3.9	Einstellungen.....	89
3.9.1	Allgemein.....	89
3.9.2	Ansicht.....	90
3.9.3	Funktionen.....	92
3.9.4	Pfade.....	93
3.9.5	AI.....	94

3.10	Hilfe	95
3.11	About.....	95
4	Explorer	96
5	Speichern.....	96
6	Zurück.....	96
7	Zurücksetzen	97
8	Lesezeichen.....	98
8.1	Lesezeichen hinzufügen.....	99
8.2	Lesezeichen Tools.....	100
9	Audit Steps.....	101
9.1	My Audit Manager	101
9.2	Audit Trail erstellen.....	103
9.3	Audit Script erstellen.....	107
10	Ask my Data (AMY)	108
10.1	Abo-Modelle	108
10.2	Funktionen	110
10.3	AMY-Tool-Box	111
10.3.1	Magic Prompt Generator.....	111
10.3.1.1	Magic Prompt Spaltenbeschriftungen.....	113
10.3.1.2	Magic Prompt Spaltenauswahl.....	114
10.3.2	Spaltenbeschreibung hinzufügen.....	114
10.3.3	Bestimmte Spalten auswählen.....	114
10.3.4	Anfrage löschen.....	114
10.4	AMY- Tabelle im Analyzer	115
10.5	AMY – Lesezeichen	115
10.6	AMY- Dokumentation.....	116
10.7	AMY-Interaktives Speichern	116
11	Tab-Anzeige	117
11.1	Verschieben von Tabellenblättern.....	117
11.2	Tab-Menü.....	117
12	Datentabelle	119
12.1	Formatierungstypen Spalte	119
12.1.1	Datum (+ Zeit) (Zeitstempel)	119
12.1.2	Zeit.....	121

12.1.3	Nummer.....	122
12.1.4	Zeichenfolge	124
12.1.5	Boolean	124
12.1.6	Link	125
12.2	Sortierung	125
12.2.1	Aufsteigend sortieren	125
12.2.2	Absteigend sortieren.....	125
12.2.3	Sortierung zurücksetzen	125
12.3	Spaltenaktionen	126
12.3.1	Spalte duplizieren.....	126
12.3.2	Spalte umbenennen	126
12.3.3	Spalte löschen	127
12.3.4	Spalte ausblenden.....	127
12.3.5	Alle Spalten einblenden	127
12.3.6	Spalten neu anordnen	127
12.3.7	Beschreibung	128
12.4	Datentabelle: Interaktion	130
12.4.1	Inhalt kopieren	130
12.4.2	Filtern.....	130
12.4.2.1	Nach ausgewähltem Inhalt filtern	130
12.4.2.2	Mehrere Werte filtern	133
12.4.2.2.1	Fuzzy-Methode	134
12.4.2.3	Filter löschen.....	135
12.4.3	Google.....	135
12.4.4	Spaltenstatistiken	136
12.4.4.1	Spaltenstatistik für Textspalte.....	137
12.4.4.2	Spaltenstatistik für numerische Spalte	139
12.4.4.3	Spaltenstatistik für Datumsspalten	141
12.4.4.4	Spaltenstatistik für Zeitspalte	143
13	Aktionsleiste	145
13.1	Filterung.....	145
13.1.1	Mehrere Werte filtern	146
13.1.2	Inhalte zum Filtern auswählen.....	147
13.1.3	Anzeige & Filter verwalten.....	149

13.1.4	Alle Filter löschen	151
13.2	Info: Spaltenstatistik	151
13.3	Summen	152
13.4	KI-Spalten Generator	153
13.5	Table-Tools	153
13.6	Erscheinungsbild AUDIPY	153
13.7	Zoom	154
14	Data Tools	155
14.1	Auswählen	155
14.2	Verbinden	157
14.3	Anhängen	161
14.4	Extrahieren	162
14.5	Vergleichen	162
14.5.1	Praktische Anwendungsszenarien	163
14.5.2	Erstellung & Auswertung	164
14.5.3	Erweiterte Strategien	171
14.6	Table-Tools	173
14.6.1	Transformieren	173
14.6.1.1	Zusammenführen	174
14.6.1.2	Teilen	175
14.6.1.3	Extrahieren	176
14.6.1.4	Leere auffüllen	177
14.6.1.5	Wert ersetzen	178
14.6.2	Bereinigen	181
14.6.2.1	Duplikate	181
14.6.2.2	Leere Zeilen	183
14.6.3	Neue Spalte	184
14.6.3.1	KI-Spalte	185
14.6.3.2	Benutzerdefinierte Spalte	187
14.6.3.3	Editierbare Spalte	190
14.6.3.4	Indexspalte hinzufügen	192
14.6.3.5	Prüfspalte hinzufügen	193
14.6.4	Berechnen	194
14.6.4.1	Spalte berechnen	194

14.6.4.1.1	Wert berechnen	195
14.6.4.1.2	Statistische Werte:	197
14.6.4.1.3	Datum / Zeit	199
14.6.4.1.4	Zeilen verschieben	200
14.6.4.1.5	Formel eingeben:	202
14.6.4.2	Wenn...Dann...	204
14.6.4.3	Kumulative Summe.....	210
14.6.5	Sortieren	212
14.6.6	Pivot	213
14.6.7	Gruppieren.....	217
14.7	Stichprobe.....	221
14.7.1	Zufallsstichprobe	221
14.7.1.1	Praktische Anwendungsszenarien.....	222
14.7.1.2	Erstellung einer Zufallsstichprobe.....	224
14.7.1.3	Erweiterte Strategien.....	229
14.7.2	Monetary Unit Sampling (MUS).....	231
14.7.2.1	Anwendungsbeispiele & Vorteile.....	231
14.7.2.2	Methodische Beschreibung MUS	232
14.7.2.3	Erstellung.....	236
14.7.2.3.1	Planung	236
14.7.2.3.2	Extraktion	239
14.7.2.4	Bewertung.....	242
14.7.2.5	Erweiterte Strategien und Best Practices	244
14.8	Named Entity Recognition	246
14.8.1	Praktische Anwendungsszenarien.....	247
14.8.2	NER Dialog - Möglichkeiten.....	249
14.8.3	Erstellung einer Classic NER	250
14.8.4	Erstellung einer LLM-Konfiguration (Experimentell).....	255
14.8.5	Analyse	257
14.8.6	Erweiterte Strategien und Best Practices	258
15	Visualisieren.....	260
15.1	Arbeitsabläufe und Best Practices	261
15.2	Eindeutige Werte (alle)	262

15.3	Eindeutige Werte (Spalte)	263
15.3.1	Anwendungsszenarien in der Berufspraxis	264
15.4	Fehlende Werte	265
15.5	Chart Builder	266
15.5.1	Zeitreihen-Dekomposition: Erweiterte Analysemethoden	270
16	Analysieren.....	272
16.1	Überblick mit Beschreibung	273
16.2	Ausreißer.....	274
16.2.1	Methoden im Überblick	274
16.2.1.1	Z-Score (Standardisierter Score).....	275
16.2.1.2	Rolling Z-Score (Gleitender Z-Score)	276
16.2.1.3	Modified Z-Score (Robuster Z-Score).....	276
16.2.1.4	Rolling Modified Z-Score (Gleitender Robuster Z-Score)	276
16.2.1.5	Isolation Forest.....	277
16.2.1.6	HIOD (Hierarchical Incremental Outlier Detection)	277
16.2.1.6.1	Methode	278
16.2.1.6.2	Mathematische Beschreibung	278
16.2.1.6.3	Vergleich mit Isolation Forest	280
16.2.1.6.4	Anwendungen im Steuerrecht	281
16.2.2	Wofür nutzt man eine Ausreißer-Analyse?	281
16.2.3	Bedienelemente und Parameterkonfiguratio.....	282
16.2.4	Erstellung einer Ausreißer-Analyse	284
16.2.5	Visualisierung und Ergebnisinterpretation.....	289
16.2.6	Erweiterte Analysestrategien und Integration	290
16.2.7	Best Practices und methodische Empfehlungen.....	291
16.3	Boxplot / Violin.....	293
16.3.1	Wofür nutzt man Boxplot/Violin?.....	294
16.3.2	Qualitätssicherung und sicherer Umgang.....	296
16.3.3	Besonderheiten Box-Plot, Violin-Plot und Punkt-Plot	296
16.3.4	Erstellung vom Boxplot / Violin Diagramm	298
16.3.5	Erweiterte Analyseworkflows und Interpretationsstrategien.....	302
16.4	Verteilungsanalyse	304
16.4.1	Wofür nutzt man eine Verteilungsanalyse?	304
16.4.2	Werteauswahl und numerische Variablen.....	306

16.4.3	Erstellung einer Verteilungsanalyse.....	309
16.4.4	Ergebnisinterpretation und statistische Beurteilung	313
16.4.5	Erweiterte Analysestrategien und statistische Vertiefung	314
16.4.6	Best Practices und methodische Empfehlungen.....	315
16.5	Ziffernanalyse.....	317
16.5.1	Wofür nutzt man eine Ziffernanalyse?.....	317
16.5.2	Bedienelemente und Analysekonfiguration	318
16.5.3	Erstellung einer Ziffernanalyse	321
16.5.4	Automatische Optimierungen innerhalb der Analyse	326
16.5.5	Erweiterte Analysemöglichkeit	326
16.5.6	Interpretation der Ergebnisse und Risikobeurteilung.....	328
16.5.7	Qualitätssicherung und praktische Erfahrung.....	329
16.6	Lückenanalyse.....	330
16.6.1	Wofür nutzt man die Lückenanalyse?.....	331
16.6.2	Bedienelemente und Konfigurationsoptionen	333
16.6.3	Erstellung einer Lückenanalyse	335
16.6.4	Ergebnisinterpretation und Risikobeurteilung	337
16.6.5	Erweiterte Analysestrategien.....	339
16.7	Netzwerkdiagramm	340
16.7.1	Wofür nutzt man ein Netzwerkdiagramm?	340
16.7.2	Bedienelemente und Konfigurationsmöglichkeiten	343
16.7.3	Erstellung eines Netzwerkdiagrammes	345
16.7.4	Visuelle Interpretation und Farbkodierung.....	349
16.7.5	Erweiterte Analysestrategien.....	350
16.7.6	Best Practices und methodische Empfehlungen.....	351
16.8	Geldfluss	352
16.8.1	Wofür nutzt man die Geldflussmethode?	352
16.8.2	Bedienelemente und Datenstruktur-Konfiguration.....	354
16.8.3	Qualitätssicherung und Datenintegrität	356
16.8.4	Erstellung einer Geldfluss-Analyse.....	356
16.8.5	Farbkodierung und visuelle Interpretation	362
16.8.6	Erweiterte Analysestrategien und Best Practices	363
16.9	Parallele Koordinaten.....	364
16.9.1	Warum benötigt man parallele Koordinaten?	365

16.9.2	Erstellung der parallelen Koordinaten	367
16.9.3	Bedienelemente und Datenstruktur-Konfiguration	372
16.9.4	Farbkodierung und visuelle Interpretation	374
16.9.5	Erweiterte Analysestrategien und methodische Vertiefung	376
17	Entwicklermodus.....	378
17.1	Parquet Editor.....	379
17.1.1	Metadaten auslesen:	379
17.1.2	Metadaten aus Datei löschen:	381
17.1.3	Metadaten im Ordner löschen:	381
17.1.4	Metadaten ersetzen:.....	381
17.2	Sonstiges	382
17.2.1	Sicherheitstool: Verschlüsselung	382
17.2.2	Datenwerkzeuge: Anonymisierung.....	382

Vorwort AUDIPY:

Nach der Installation kann der AUDIPY Analyzer entweder über den **AUDIPY-Central** oder über das **Desktop Symbol** gestartet werden. Nach dem Start erscheint ein Ladefenster, das den Initialisierungsprozess anzeigt.

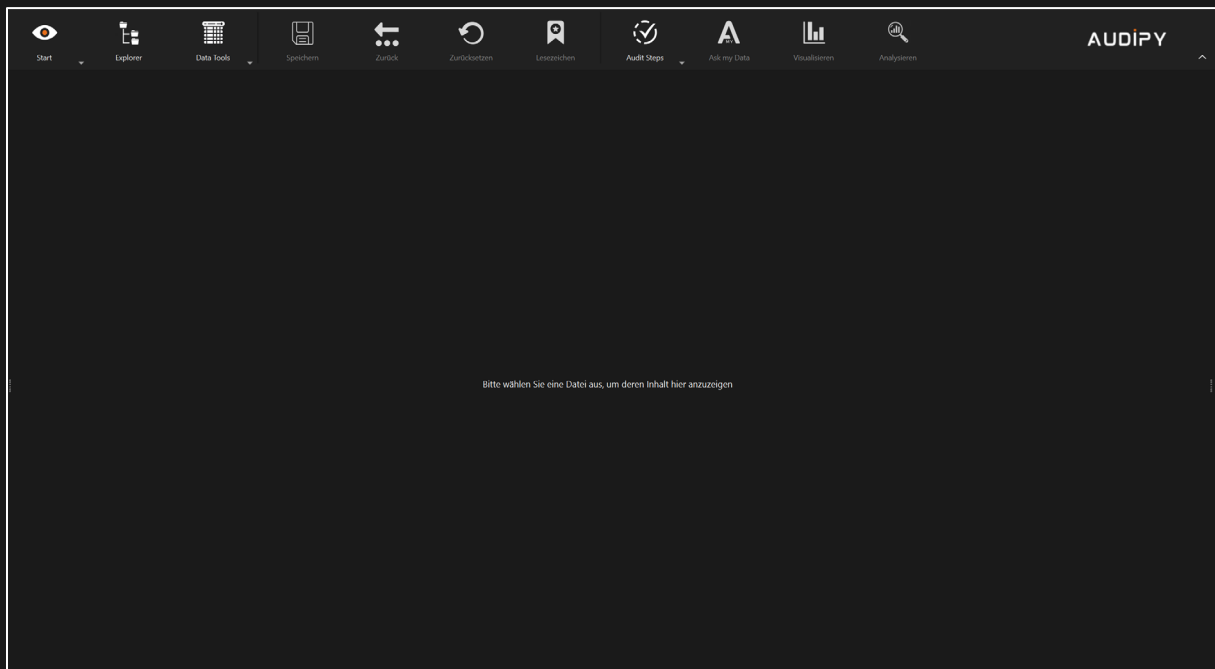


Sobald das Programm vollständig geladen ist, öffnet sich die folgende Hauptoberfläche. Hinweis: Bei Start über den Central öffnet sich der AUDIPY Explorer mit den über den Central importierten Daten. Bei Start über das Desktop Icon bleibt das Programmfenster erst leer (s. **1** Einstieg AUDIPY).

ROW ID	ORDER ID	ORDER DATE	SHIP DATE	SHIP MODE	CUSTOMER ID	CUSTOMER NAME	SEGMENT	COUNTRY	CITY	STATE	POSTAL CODE	REGION	PROD
1	CA-2016-152156	08.11.2016	11.11.2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-1000
2	CA-2016-152156	08.11.2016	11.11.2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-1000
3	CA-2016-138688	12.06.2016	16.06.2016	Second Class	DV-13045	Darin Van Huff	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-1000
4	US-2015-108966	11.10.2015	18.10.2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-1000
5	US-2015-108966	11.10.2015	18.10.2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-1000
6	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-1000
7	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-1000
8	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1000
9	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-1000
10	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-PA-1000
11	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-1000
12	CA-2014-115812	09.06.2014	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1000
13	CA-2017-114412	15.04.2017	20.04.2017	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	OFF-PA-1000
14	CA-2016-161389	05.12.2016	10.12.2016	Standard Class	BM-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-1000
15	US-2015-118983	22.11.2015	26.11.2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-1000
16	US-2015-118983	22.11.2015	26.11.2015	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-1000
17	CA-2014-105893	11.11.2014	18.11.2014	Standard Class	PK-19075	Pete Kutz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-1000
18	CA-2014-161164	13.05.2014	15.05.2014	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jordan	Utah	94084	West	OFF-ST-1000
19	CA-2014-143336	27.08.2014	01.09.2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-AR-1000
20	CA-2014-143336	27.08.2014	01.09.2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	TEC-PH-1000
21	CA-2014-143336	27.08.2014	01.09.2014	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West	OFF-BI-1000
22	CA-2016-137330	09.12.2016	13.12.2016	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AR-1000
23	CA-2016-137330	09.12.2016	13.12.2016	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-1000

1 Einstieg AUDIPY

Wird der AUDIPY direkt gestartet, ohne dass zuvor Daten über den Central-Client importiert wurden, erscheint zunächst ein leeres Programmfenster. In diesem Fall sind noch keine Daten geladen, und es stehen keine Analysefunktionen zur Verfügung. Um mit der Arbeit zu beginnen, muss der Nutzer manuell eine Datei öffnen oder den Import über den Central-Client durchführen. Die Vorgehensweise wird in den folgenden Kapiteln ausführlich dargestellt.



1.1 Dateiformate im AUDIPY

Für die Analyseprozesse nutzt der Analyzer spezifische Dateiformate, die für die effiziente Verarbeitung strukturierter Daten ausgelegt sind. Besonders hervorzuheben ist dabei das **Parquet-Format**.

Parquet ist ein spaltenbasiertes, Open-Source-Dateiformat, das speziell für analytische Workloads entwickelt wurde. Im Gegensatz zu zeilenbasierten Formaten (wie CSV) ermöglicht die spaltenweise Speicherung eine deutlich effizientere Datenkompression und beschleunigte Abfragen – insbesondere bei großen Datenmengen. Aufgrund seiner offenen Spezifikation und breiten Unterstützung eignet sich Parquet nicht nur für den Einsatz im Analyzer, sondern ist auch problemlos in anderen Anwendungen wie Power BI, Apache Spark oder Pandas nutzbar.

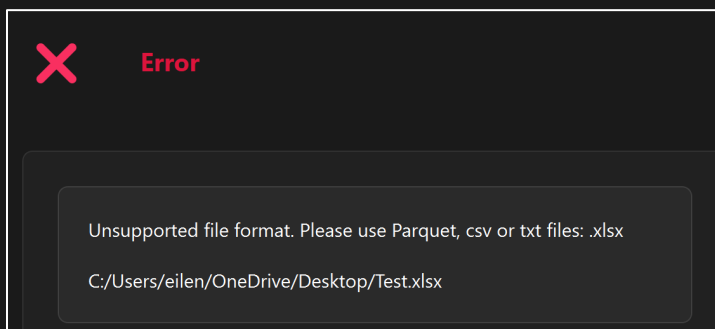
Diese Eigenschaften machen Parquet zur bevorzugten Wahl für den Austausch und die Weiterverarbeitung von Analyseergebnissen im Rahmen moderner Datenanalyseprozesse.

Hinweis:

Neben Parquet-Dateien unterstützt der Analyzer auch das Einlesen von **CSV-Dateien**. Diese können flexibel verwendet werden, etwa bei Datenexporten aus Drittsystemen oder bei manueller Datenbearbeitung. Es ist jedoch zu beachten, dass beim Öffnen von CSV-Dateien eine temporäre Parquet-Datei generiert wird, die bei Bedarf manuell gespeichert werden muss.

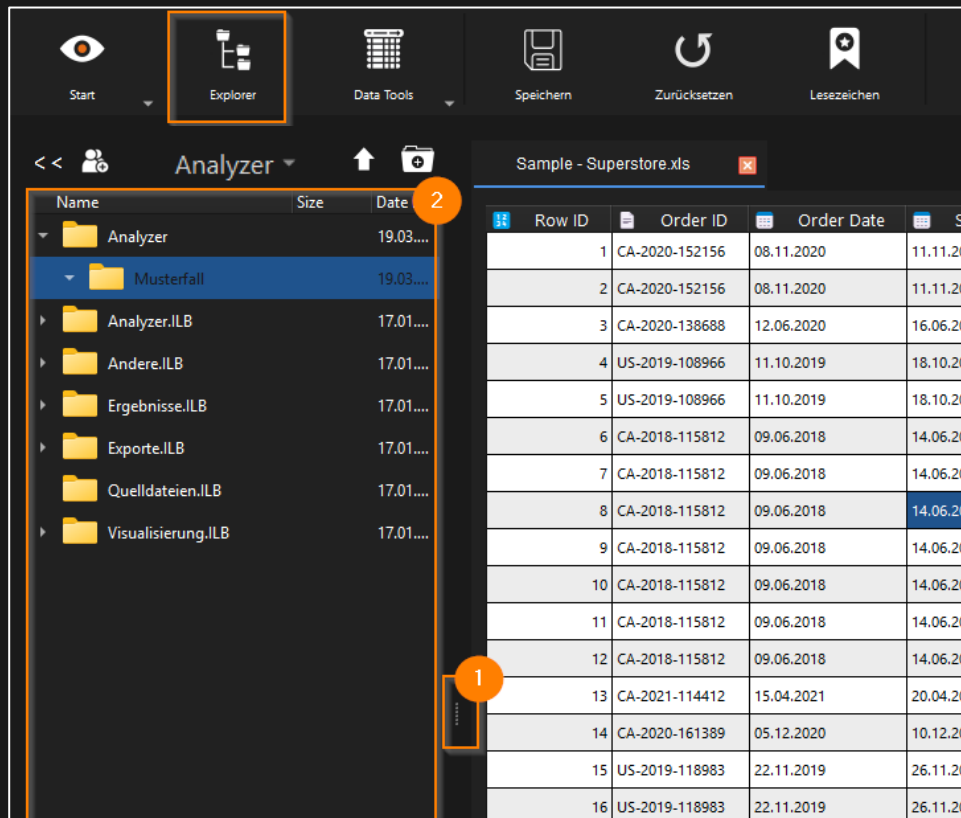
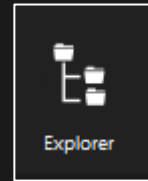
Hinweis Nr. 2:

Beim Öffnen eines anderen Datenformats (z.B. Excel `xlsx`) erscheint ein Hinweis, dass diese nicht geladen werden kann. Hier sollte man entweder im Central oder über die Import Funktion (s. Kapitel 3.2) im Analyzer die Datei einmal in ein Parquet Format umwandeln und erneut öffnen.



1.2 Dateiexplorer

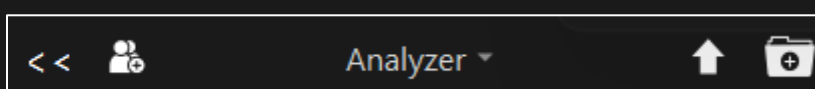
Auf der linken Seite befinden sich kleine Punkte **①**, über die sich die Explorer-Leiste größentechnisch verschieben lässt. Verwenden Sie den **Explorer**-Headerpunkt, um die Explorerleiste zu öffnen. Es öffnet sich eine seitliche Leiste **②**, die Ihnen schnellen Zugriff auf Ihre Dateien und Ordner ermöglicht.



Der **Explorer** bietet eine zentrale Übersicht über verfügbare Dateien und Ordner (mit Änderungsdatum und Uhrzeit), sodass Sie schnell auf die benötigten Ressourcen zugreifen können. Durch einfaches Navigieren lassen sich einzelne Dateien in den Analyzer laden, ohne die Anwendung verlassen zu müssen.

1.2.1 Obere Explorer-Symbole

Zusätzlich bietet der Explorer im **oberen Bereich** folgende Icons:



1.2.1.1 Schließen des Explorers



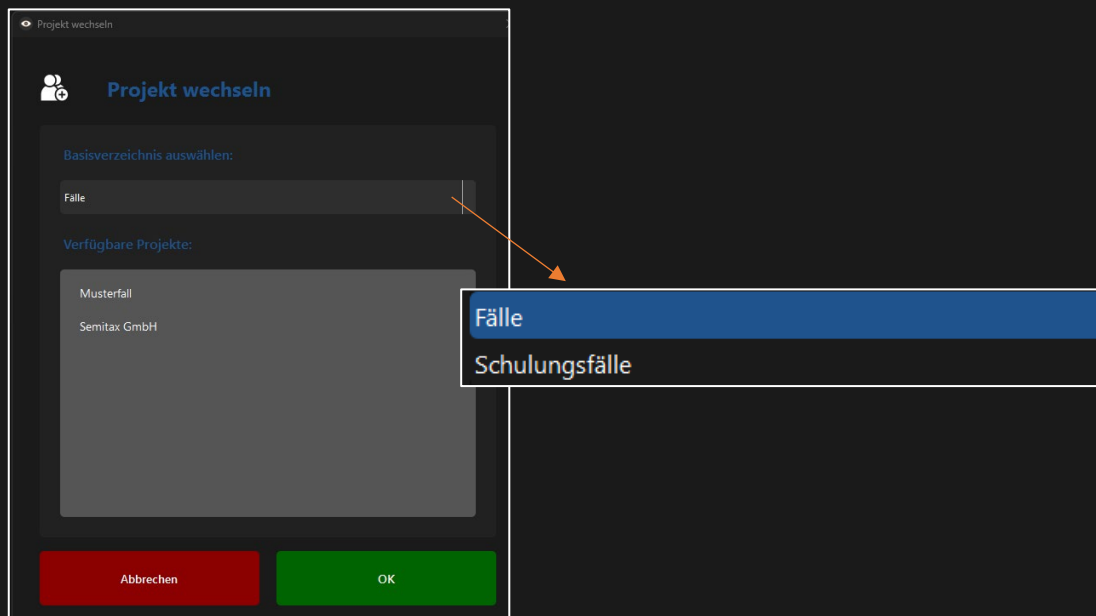
Mit dem **doppelten Pfeilsymbol** lässt sich der Explorer wieder einklappen / schließen, um eine größere Darstellung der Datentabelle zu ermöglichen.

1.2.1.2 Projekt wechseln



Die Funktion „**Projekt wechseln**“ ermöglicht es dem Benutzer, zwischen verschiedenen vorhandenen Projekten (die verschiedenen Tabs im Central unter „Fallverwaltung“) zu wechseln, die in einem vordefinierten Verzeichnis gespeichert sind. Standardisiert eingestellt sind Fälle und Schulungsfälle.

Nach dem Aufruf dieser Funktion öffnet sich ein Dialogfenster, mit einem Auswahlbereich, der es ermöglicht in den angelegten Projektordnern hin und her zu wechseln (im Beispiel Ordner Fälle und den Ordner Schulungsfälle). Die abgelegten Unterprojekte wie hier **Musterfall** und **Semitax GmbH** des Ordners „Fälle“ werden dann im Fenster darunter angezeigt.

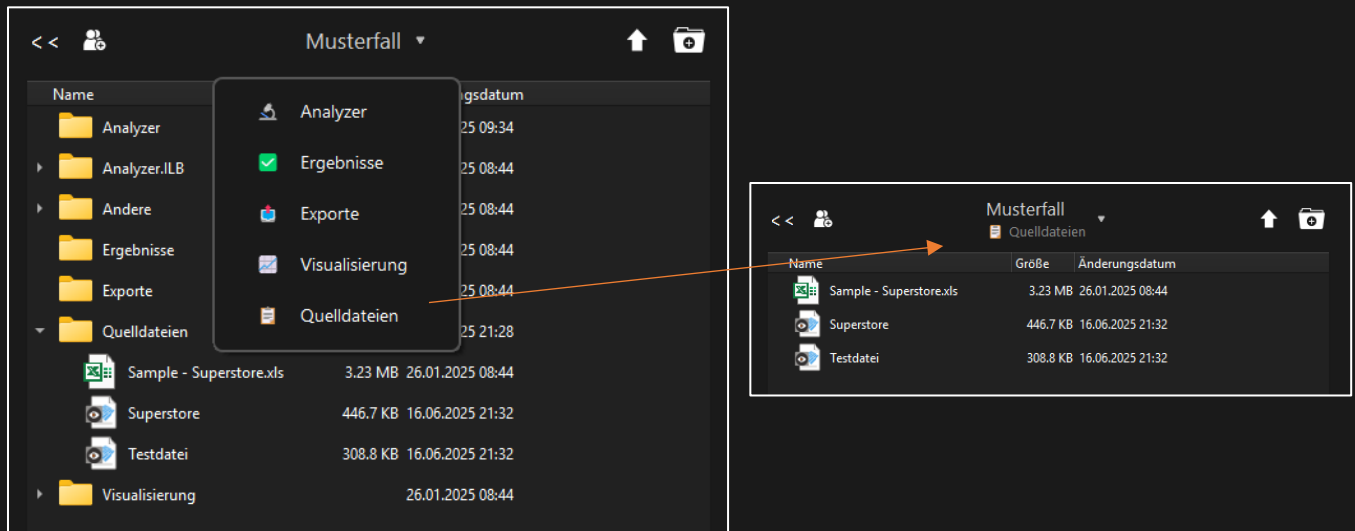


Zur **Auswahl eines Projekts** muss einfach auf den entsprechenden Eintrag in der Liste geklickt werden. Sobald ein Projekt markiert ist, kann die Auswahl durch einen Klick auf die grüne Schaltfläche „**OK**“ bestätigt werden. Die Anwendung lädt daraufhin alle relevanten Daten des gewählten Projekts und stellt diese zur weiteren Bearbeitung bereit.

Falls der Benutzer den Vorgang abbrechen möchte, kann er dies jederzeit über die rote Schaltfläche „**Abbrechen**“ tun. In diesem Fall wird das Fenster geschlossen, ohne Änderungen am aktuell geladenen Projekt vorzunehmen.

1.2.1.3 Navigation durchs Fallverzeichnis

Durch einen einfachen Linksklick auf den Namen des Falls öffnet sich ein neues Fenster, in dem direkt in die Unterordner des jeweiligen Falls navigiert werden kann. So gelangt man **beispielsweise** direkt zum Ordner „**Quelldateien**“ und sieht anschließend die Dateien im Explorer – ganz ohne Umweg über den Central oder des Microsoft Dateixplorer. Ein Wechsel kann so schnell und unkompliziert erfolgen.



1.2.1.4 Übergeordnete Ordner



Durch Klicken auf dieses Symbol wird im Datei-Explorer der übergeordnete Ordner des aktuellen Verzeichnisses geöffnet.

Wiederholtes Klicken ermöglicht das schrittweise Wechseln zu immer höheren Verzeichnisebenen, bis zur höchsten verfügbaren Ebene (z.B. dem Laufwerks- oder Computerübersicht).

Beispiel:

- Aktueller Ordner: C:\Benutzer\Max\Dokumente\Projekte
- Nach Klick: C:\Benutzer\Max\Dokumente
- Weiterer Klick: C:\Benutzer\Max
- usw., bis zur obersten Ebene.
-

Einsatzbereich:

Erleichtert das schnelle Navigieren innerhalb der Ordnerstruktur, ohne dass man manuell zurückwechseln oder neue Explorer-Fenster öffnen muss.

1.2.1.5 Ordner anlegen



Über das **Ordnersymbol mit dem Pluszeichen** kann ein neuer Ordner erstellt werden. Dies ermöglicht es dem Nutzer, eine individuelle Ordnerstruktur anzulegen – etwa zur thematischen Trennung von Analysen oder zur besseren Organisation eigener Dateien innerhalb der Anwendung.

Nach Klick auf das Ordnersymbol öffnet sich ein Fenster, wo der Ordnername hinterlegt werden muss:

Neuer Ordner

Ordnername:

Text hier eingeben...

Abbrechen OK

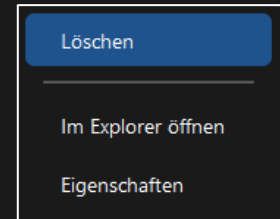
Anschließend wird der Ordner mit dem Erstellungsdatum und der genauen Uhrzeit im Explorer angezeigt:

Name	Größe	Änderungsdatum
eSupport		13.08.2024 18:43
Musterordner		19.04.2025 10:08
PerfLogs		01.04.2024 09:26
Program Files		03.04.2025 14:14
Program Files (x86)		04.03.2025 21:26
Users		04.03.2025 21:28
Windows		18.04.2025 13:51

Nun können die Daten in den Ordner gelegt / verschoben werden und über den Pfeil daneben aufgeklappt werden.

1.2.2 Rechtsklick Menü im Explorer

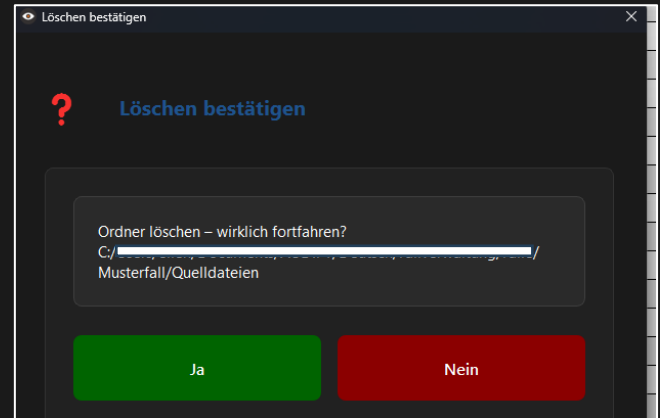
Mit **Rechtsklick auf einen Ordner** erscheint folgendes Menü.



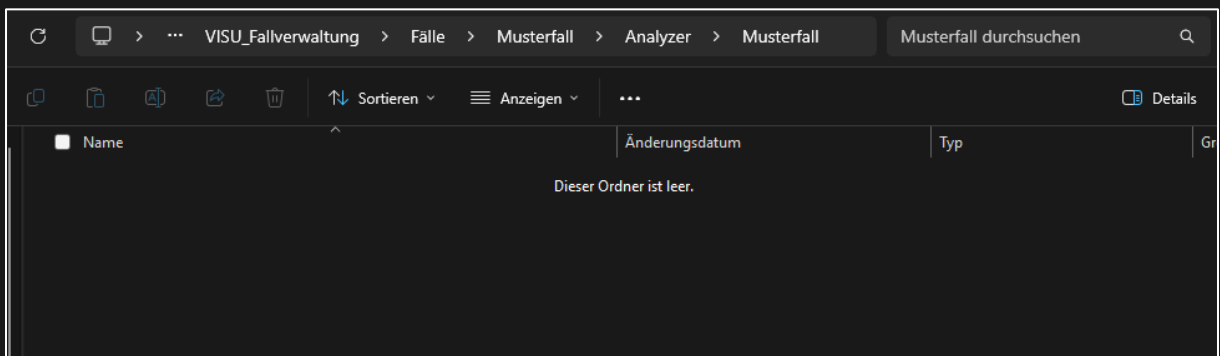
Die folgenden Optionen stehen im Kontextmenü zur Verfügung:

- **Löschen:**
Entfernt die Datei aus dem aktuellen Speicherort (Achtung: je nach Anwendung kann dies dauerhaft oder in den Papierkorb erfolgen).

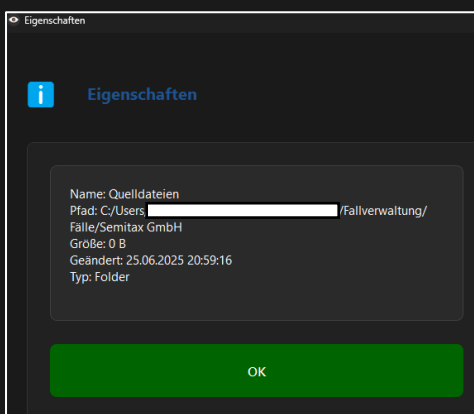
Das folgende Bestätigungsfenster muss für eine endgültige Löschung mit „Ja“ bestätigt werden:



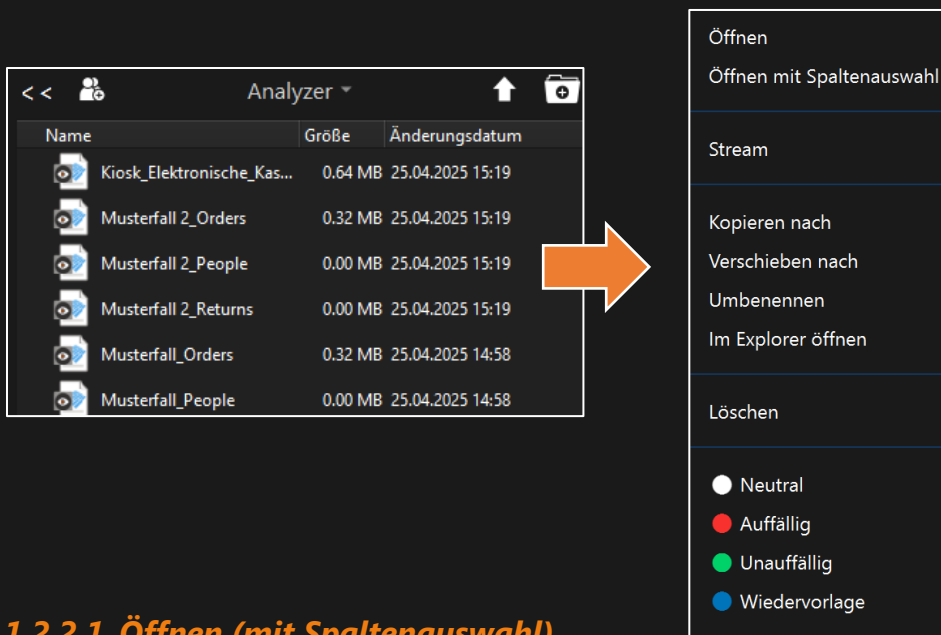
- **Im Explorer öffnen:**
Öffnet den Speicherort der ausgewählten Datei direkt im Windows Explorer.



- **Eigenschaften:**
Öffnet das folgende Fenster und erklärt die Eigenschaften des Ordners.



Wenn Parquet-Dateien angezeigt werden, wie im folgenden Beispiel, können Sie mit **einem Rechtsklick** auf einen Fall eine weitere Menüleiste aufrufen:

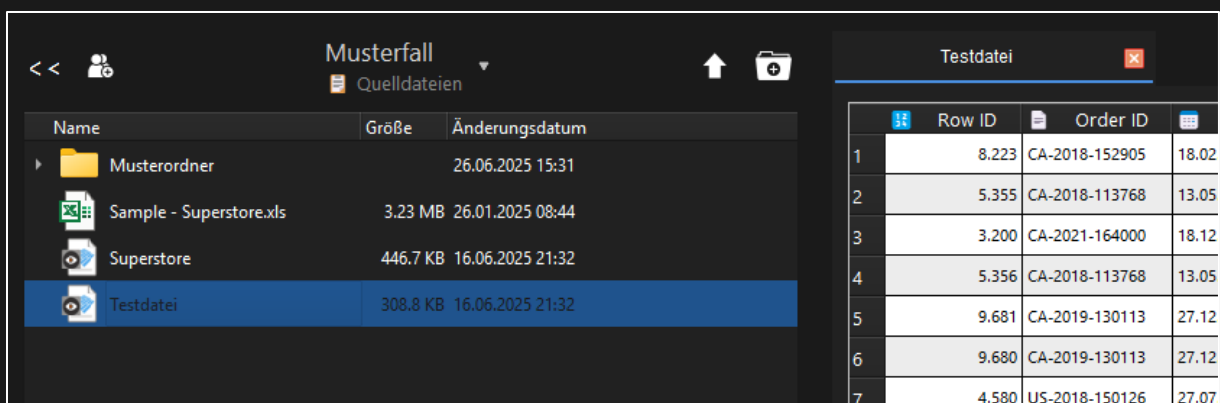


1.2.2.1 Öffnen (mit Spaltenauswahl)

Diese Option dient dazu, den gewählten Fall oder die Datei zu öffnen oder direkt für eine Analyse vorzubereiten:

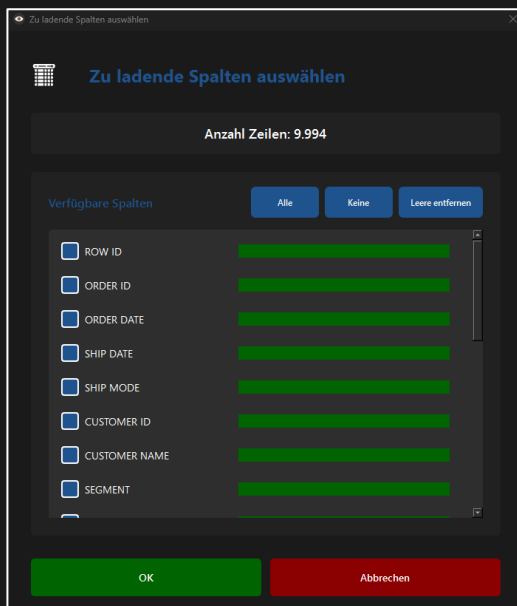
- **Öffnen**

Öffnet die Datei im Standardmodus der Anwendung zur Einsicht oder Bearbeitung. Der Fall wird automatisch geladen und erscheint rechts neben der Explorerleiste:



- **Öffnen mit Spaltenauswahl**

Ermöglicht das gezielte Öffnen der Datei mit Auswahl **bestimmter Spalten** – ideal für große Datenmengen oder gezielte Prüfziele.



Über die Schnellbuttons „**Alle**“ und „**Keine**“ lassen sich Spalten in der Ansicht schnell und komfortabel ein- oder ausblenden.

Diese Funktionen erleichtern die gezielte Auswahl relevanter Spalten, insbesondere bei umfangreichen Datensätzen.

Zusätzlich steht die Option „**Leere entfernen**“ zur Verfügung. Aktiviert man diese Funktion, werden automatisch alle Spalten aus der Ansicht entfernt bzw. ausgehakt, die keine Werte enthalten. Dies trägt zur besseren Übersichtlichkeit

bei und hilft, sich auf die inhaltlich relevanten Daten zu konzentrieren.

Am unteren Rand des Explorers befindet sich zusätzlich ein Button „**Spaltenauswahl**“, als Schnellzugriff dieser Funktion (s. **1.2.3**)

1.2.2.2 Stream

Die **Stream-Ansicht** erlaubt es, sehr große Datenmengen effizient zu verarbeiten, indem sie nur in kleinen Abschnitten geladen werden. Dies reduziert den Speicherverbrauch drastisch und ermöglicht die Bearbeitung von Dateien mit mehreren Millionen Zeilen. Sie kann entweder per **Rechtsklick auf ein Stream-fähiges Objekt** oder über den **Schnellzugriff am unteren Rand des Explorers** gestartet werden.

Im folgenden Screenshot stehen Ihnen hierbei folgende Menüpunkte zur Verfügung.

The screenshot shows the 'Stream' tool interface. On the left, there's a panel with five numbered buttons: 1. Extraktion (Extraction), 2. Gruppieren (Group), 3. Verbinden (Join), 4. Anhängen (Append), and 5. Pivot. Above these buttons, text explains that these tools load data in small chunks to process large files efficiently. On the right, a data table is displayed with columns: event_time, event_type, product_id, category_id, and category_code. The table contains 22 rows of data.

	event_time	event_type	product_id	category_id	category_code
1	01.10.2019 02:00:00	view	44600062	2103807459595387724	
2	01.10.2019 02:00:00	view	3900821	2053013552326770905	appliances.environment.water_heater
3	01.10.2019 02:00:01	view	17200506	2053013559792632471	furniture.living_room.sofa
4	01.10.2019 02:00:01	view	1307067	2053013558920217191	computers.notebook
5	01.10.2019 02:00:04	view	1004237	2053013555631882655	electronics.smartphone
6	01.10.2019 02:00:05	view	1480613	2053013561092866779	computers.desktop
7	01.10.2019 02:00:08	view	17300353	2053013553853497655	
8	01.10.2019 02:00:08	view	31500053	2053013558031024687	
9	01.10.2019 02:00:10	view	28719074	2053013565480109009	apparel.shoes.keds
10	01.10.2019 02:00:11	view	1004545	2053013555631882655	electronics.smartphone
11	01.10.2019 02:00:11	view	2900536	2053013554776244595	appliances.kitchen.microwave
12	01.10.2019 02:00:11	view	1005011	2053013555631882655	electronics.smartphone
13	01.10.2019 02:00:13	view	3900746	2053013552326770905	appliances.environment.water_heater
14	01.10.2019 02:00:15	view	44600062	2103807459595387724	
15	01.10.2019 02:00:16	view	13500240	2053013557099889147	furniture.bedroom.bed
16	01.10.2019 02:00:17	view	23100006	2053013561638126333	
17	01.10.2019 02:00:18	view	1801995	2053013554415534427	electronics.video.tv
18	01.10.2019 02:00:18	view	10900029	2053013555069845885	appliances.kitchen.mixer
19	01.10.2019 02:00:19	view	1306631	2053013558920217191	computers.notebook
20	01.10.2019 02:00:19	view	1005135	2053013555631882655	electronics.smartphone
21	01.10.2019 02:00:20	view	1003306	2053013555631882655	electronics.smartphone
22	01.10.2019 02:00:20	view	4803399	2053013554658804075	electronics.audio.headphone

Folgende Funktionen sind möglich:

Extraktion ① - 1.2.2.2.1

Funktion: Selektives Herausfiltern von Spalten oder Zeilen anhand von Bedingungen (z.B. nur bestimmte Kategorien, Zeiträume oder Wertebereiche).

Warum wichtig in Stream: Beim Arbeiten mit großen Datenmengen ist es ineffizient, immer alle Daten zu laden. Mit der Extraktion können gezielt nur die benötigten Teilmengen verarbeitet werden, wodurch Ressourcen geschont und die Übersichtlichkeit erhöht wird.

Gruppieren ② - 1.2.2.2.2

Funktion: Gruppiert Datensätze nach einem oder mehreren Kriterien und erlaubt Aggregationen wie Summe, Durchschnitt, Minimum, Maximum oder Zählungen.

Warum wichtig in Stream: Bei großen Datenmengen ist es oft hilfreich, einen Überblick zu gewinnen, ohne jede einzelne Zeile anzuschauen. Gruppierung reduziert die Datenmenge auf wesentliche Kennzahlen und erleichtert die Analyse von Mustern oder Trends.

Verbinden ③ - 1.2.2.2.3

Funktion: Kombiniert mehrere Datensätze über gemeinsame Schlüssel (Join-Operation).

Warum wichtig in Stream: Daten sind oft auf mehrere Dateien oder Tabellen verteilt. Das Verbinden ermöglicht es, diese Informationen zu konsolidieren – auch bei großen Quellen – ohne sie vorher vollständig zu laden.

Anhängen ④ - 1.2.2.2.4

Funktion: Fügt mehrere Tabellen mit gleicher Struktur untereinander zusammen.

Warum wichtig in Stream: Viele Datenquellen liefern ihre Informationen in getrennten Dateien (z. B. monatliche Exporte). Durch Anhängen lassen sich diese Dateien effizient zu einem zusammenhängenden Datensatz verarbeiten, ohne den Speicher zu überlasten.

Pivot ⑤ - 1.2.2.2.5

Funktion: Wandelt Zeilen in Spalten um (oder umgekehrt), z.B. zur Bildung von Kreuztabellen.

Warum wichtig in Stream: Pivot-Tabellen bieten eine kompakte Darstellung großer Datenmengen und ermöglichen schnelle Auswertungen. In Stream können solche Transformationen auch bei Millionen Zeilen vorgenommen werden, ohne dass die Anwendung überlastet wird.

1.2.2.2.1 Extraktion

Startet einen automatisierten Prozess zur Datenextraktion oder strukturellen Aufbereitung. Dies kann z.B. für Analysezwecke oder zur Weiterverarbeitung notwendig sein. Wichtig ist hierbei, der Fall ist noch gar nicht geöffnet, sondern er wird mit den konkreten Angaben extrahiert und anschließend geöffnet:

Nach Auswahl „Extraktion“ erscheint folgendes Fenster:

The screenshot shows a software window titled 'Extraktion'. At the top, a status bar indicates 'Anzahl Zeilen: 9.980' (Callout 1). Below this, a large grey area says 'Keine Filter ausgewählt' (Callout 2). To the right of this area are two buttons: 'Neuer Filter' (Callout 2) and 'Spaltenauswahl' (Callout 3). Below the buttons, a prompt reads 'Bitte wählen Sie eine Spalte zum Filtern aus'. Underneath is a table with columns: Ship Mode, Customer ID, Customer Name, Segment, Country/Region, City, State, Postal Code, Region, and a partially visible 'FUR'. The table contains 14 rows of data (Callout 4). At the bottom, there are two checkboxes: 'Im Projekt speichern' (checked) and 'Nach Schlüsselfeld extrahieren' (unchecked). The 'Dateiname:' field contains 'Musterfall 2_Orders_filtered'. To the right of the second checkbox is a 'Rechnungsnummer' field. A large green button labeled 'Extraktion' is at the bottom center.

	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City	State	Postal Code	Region	FUR
1	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR
2	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR
3	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West	OFF
4	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR
5	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF
6	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR
7	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF
8	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC
9	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF
10	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF
11	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR
12	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC
13	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28027	South	OFF
14	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98103	West	OFF

Oberfläche im Überblick

- **Anzahl Zeilen 1:**

Oben im Fenster wird die Gesamtanzahl der geladenen Datensätze angezeigt (hier: **9.980 Zeilen**). Diese Zahl ändert sich dynamisch, wenn Filter angewendet werden.

- **Filterbereich 2:**

Im oberen mittleren Bereich wird angezeigt, ob und welche Filter aktiv sind. Über den Button „**Neuer Filter**“ auf der rechten Seite können benutzerdefinierte Filterkriterien erstellt werden (z.B. „Segment = Consumer“ oder „Ship Mode =

Second Class“). Hierfür erscheint ein neues Fenster, welches unter **1.2.1.1.1** genauer erläutert wird.

- **Spaltenauswahl 3:**

Mit dem Button „**Spaltenauswahl**“ kann gezielt festgelegt werden, welche Spalten in das Ergebnis einbezogen werden sollen. Dies verbessert Übersichtlichkeit und reduziert die Datenmenge auf das Wesentliche. Die Funktion und das sich öffnende Programmfenster wurden bereits unter **1.2.2.1** genauer dargelegt. Für eine bessere Übersichtlichkeit wird hierauf verwiesen.

- **Datentabelle 4:**

Die zentrale Tabelle zeigt die geladenen Datensätze mit allen verfügbaren Spalten. Durch Klicken auf die Spaltennamen wird die Spalte ausgewählt mit ein neuer Filter gesetzt werden soll. Sie erscheint anschließend blau hinterlegt:

er ID	Customer Name	Segment
	Claire Gute	Consumer
	Claire Gute	Consumer
	Darrin Van Huff	Corporate
	Sean O'Donnell	Consumer

Einstellungen am unteren Rand:

☐ Im Projekt speichern	Dateiname): Musterfall 2_Orders_filtered	☐ Nach Schlüsselfeld extrahieren	Rechnungsnummer

- **Im Projekt speichern:**

Aktiviert die Option, die extrahierte Datei direkt im aktuellen Projekt abzulegen. Ansonsten wird die Datei nur temporär abgelegt.

- **Dateiname:**

Hier kann ein Name für die extrahierte Datei vergeben werden. Der Name sollte aussagekräftig gewählt werden, um den Inhalt der Extraktion leicht wiederzuerkennen (z. B. „Orders_Filtered“).

- **Nach Schlüsselfeld extrahieren:**

Optional kann ein **Schlüsselfeld** angegeben werden (z. B. „Rechnungsnummer“), um sicherzustellen, dass nur eindeutige Datensätze berücksichtigt werden – nützlich zur Vermeidung von Dubletten. Funktioniert erst nach Auswahl, dass dies im Projekt gespeichert werden soll (vorher ausgegraut); da eine Schlüsselfeld Extraktion im temporären Speicher nicht sinnvoll ist.

Aktion

- **Extraktion (grüner Button):**

Startet den Extraktionsprozess mit den gewählten Filter- und Spalteneinstellungen. Das Ergebnis wird als neue Datei gespeichert und kann anschließend weiterverwendet oder erneut geöffnet werden.

Extraktion: Neuer Filter

Wählt man eine Spalte in der Tabelle aus und klickt anschließend auf „**Neuer Filter**“, öffnet sich ein Dialogfeld, in dem definiert werden kann, **welche Bedingung** auf die gewählte Spalte angewendet werden soll. Dies ermöglicht eine präzise Steuerung der Datenextraktion. **Das Format** der Spalte (Numerisch, Datum, Zeichen etc.) entscheidet über das Aussehen und der Funktionen innerhalb des Dialogs.

1.2.2.2.1.1 Filter: Zeichen

Beim Zeichenfeld sieht das Dialogfenster wie folgt aus:

The screenshot shows a dialog box titled 'Inhalt zum Filtern auswählen'. It contains several sections and controls:

- Section 1:** A header area with a checkmark icon, the text 'Eindeutige Werte in CATEGORY', and a circled number 1.
- Section 2:** A dropdown menu labeled 'Übereinstimmungstyp:' with the selected value 'Exakte Übereinstimmung' and a circled number 2.
- Section 3:** A dropdown menu labeled 'Sortieren nach:' with the selected value 'Alphabetisch aufsteigend' and a circled number 3.
- Section 4:** A search input field labeled 'Search:' with the placeholder text 'Tippen um Werte zu filtern...' and a circled number 4.
- Section 5:** A list of available values under the heading 'Verfügbare Werte'. It includes three items: 'Furniture (2121)', 'Office Supplies (6026)', and 'Technology (1847)', each with an unchecked checkbox. Above the list are three buttons: 'Alle auswählen', 'Alle abwählen', and 'Wert hinzufügen'. A circled number 5 is placed over the list area.
- Buttons:** At the bottom, there are two large buttons: a green 'OK' button and a red 'Abbrechen' button.

Elemente des Dialogs im Überblick:

- **Spaltenname ①:**

Zeigt an, auf welche Spalte sich der Filter bezieht (hier z.B. *Category*). Die Auswahl wurde zuvor in der Tabelle getroffen.

- **Übereinstimmungstyp ②:**

Bestimmt, wie genau der ausgewählte Wert mit den tatsächlichen Daten abgeglichen wird. Optionen sind hier:

Exakte Übereinstimmung
Stimmt nicht überein
Enthält
Enthält nicht

- **Sortieren nach ③:**

Legt fest, wie die Einträge in der Liste darunter sortiert werden – z.B. alphabetisch aufsteigend oder nach Häufigkeit (Anzahl).

Alphabetisch aufsteigend
Alphabetisch absteigend
Anzahl aufsteigend
Anzahl absteigend

- **Suchfeld ④:**

Ermöglicht das schnelle Auffinden bestimmter Werte durch Eingabe eines Suchbegriffs. Die Liste aktualisiert sich dynamisch, um passende Einträge hervorzuheben.

- **Liste eindeutiger Werte ⑤:**

Zeigt alle in der Spalte vorkommenden **eindeutigen Werte** (hier Produktkategorien), jeweils mit der Anzahl der Vorkommen in Klammern. Beispiel: Technology (1847) bedeutet, dass dieser Name 1847-mal im Datensatz enthalten ist.

Schnellfunktionen:

Alle auswählen	Alle abwählen	Wert hinzufügen

- **Alle auswählen / Alle abwählen:**

Schnellauswahl für die gesamte Liste – ideal zur Umkehrung oder Komplettauswahl.

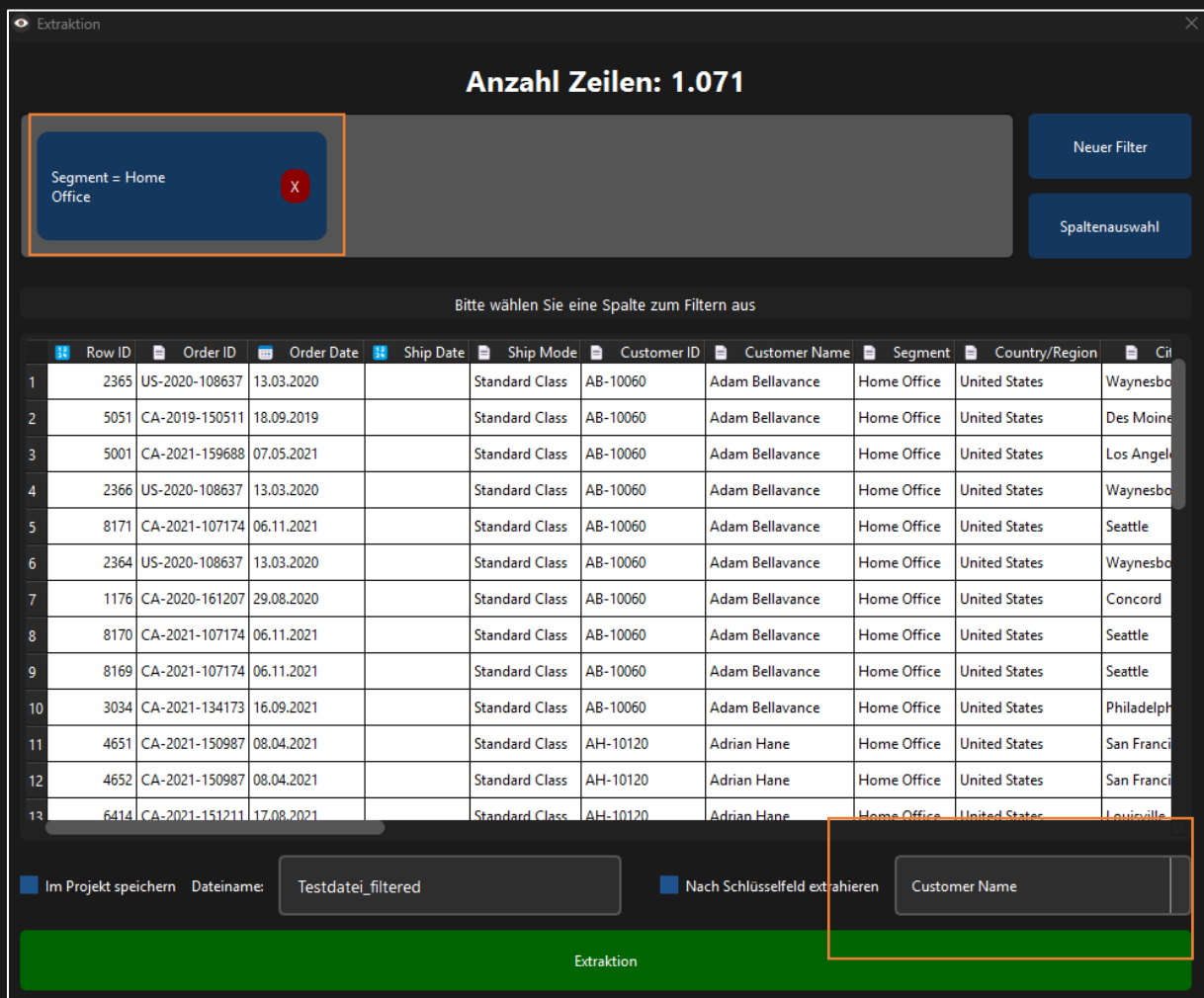
- **Wert hinzufügen:**

Ermöglicht das manuelle Hinzufügen eines Werts, der eventuell nicht in der Liste steht (z.B. bei später zu erwartenden Daten oder spezifischen Anforderungen).

Extraktion Beispiel

Im Beispielsfall soll eine vorherige Extraktion nach Schlüsselfeld „Customer Name“ und nach dem „Segment“ Homeoffice stattfinden.

Hierfür wählt man Extraktion und gibt folgende Parameter ein:



Anzahl Zeilen: 1.071

Segment = Home Office

Neuer Filter

Spaltenauswahl

Bitte wählen Sie eine Spalte zum Filtern aus

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City
1	2365	US-2020-108637	13.03.2020		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Waynesbo
2	5051	CA-2019-150511	18.09.2019		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Des Moines
3	5001	CA-2021-159688	07.05.2021		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Los Angeles
4	2366	US-2020-108637	13.03.2020		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Waynesbo
5	8171	CA-2021-107174	06.11.2021		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Seattle
6	2364	US-2020-108637	13.03.2020		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Waynesbo
7	1176	CA-2020-161207	29.08.2020		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Concord
8	8170	CA-2021-107174	06.11.2021		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Seattle
9	8169	CA-2021-107174	06.11.2021		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Seattle
10	3034	CA-2021-134173	16.09.2021		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Philadelph
11	4651	CA-2021-150987	08.04.2021		Standard Class	AH-10120	Adrian Hane	Home Office	United States	San Francisco
12	4652	CA-2021-150987	08.04.2021		Standard Class	AH-10120	Adrian Hane	Home Office	United States	San Francisco
13	6414	CA-2021-151211	17.08.2021		Standard Class	AH-10120	Adrian Hane	Home Office	United States	Louisville

Im Projekt speichern Dateiname: Testdatei_filtered

Nach Schlüsselfeld extrahieren Customer Name

Extraktion

Man wählt die Spalte Segment aus und setzt den Filter für „**Segment = Home Office**“. Zusätzlich klickt man auf „im Projekt speichern“ und wählt beim **Schlüsselfeld Customer Name** als Spalte aus.

Nach Abschluss des Vorgangs wird die erzeugte Datei / die erzeugten Dateien im Explorer sichtbar und automatisch geladen, sobald sie erstellt wurde(n).



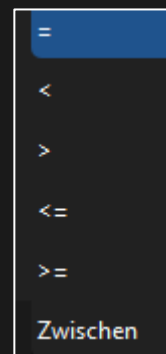
4-2020-129714	01.09.2020	03.09.2020	First Class	AB-10060	Adam Bellavance	Home Office	United States
4-2020-129714	01.09.2020	03.09.2020	First Class	AB-10060	Adam Bellavance	Home Office	United States
4-2020-129714	01.09.2020	03.09.2020	First Class	AB-10060	Adam Bellavance	Home Office	United States
4-2020-129714	01.09.2020	03.09.2020	First Class	AB-10060	Adam Bellavance	Home Office	United States
4-2020-161207	29.08.2020	03.09.2020	Standard Class	AB-10060	Adam Bellavance	Home Office	United States
5-2020-108637	13.03.2020	18.03.2020	Standard Class	AB-10060	Adam Bellavance	Home Office	United States
5-2020-108637	13.03.2020	18.03.2020	Standard Class	AB-10060	Adam Bellavance	Home Office	United States
5-2020-108637	13.03.2020	18.03.2020	Standard Class	AB-10060	Adam Bellavance	Home Office	United States

Ergebnis: Es werden mehrere Einzeldateien erzeugt, und zwar nach dem Namen (Schlüsselfeld) und dem Segment Homeoffice.

1.2.2.2.1.2 Filter Numerisch

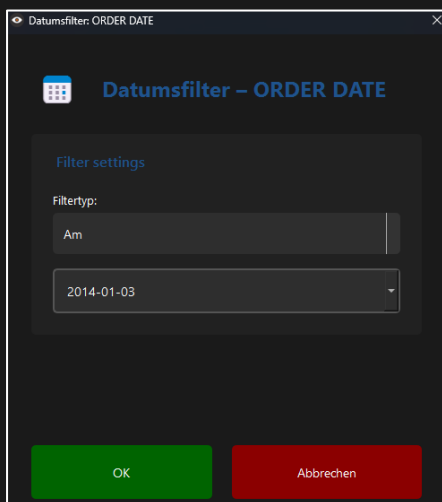
Bei einem numerischen Feld sieht der „Neue Filter“ Dialog wie folgt aus:

Unter Vergleich können die folgenden Parameter ausgewählt werden:
Anschließend folgt dieselbe Vorgehensweise wie unter „Zeichen“ beschrieben.

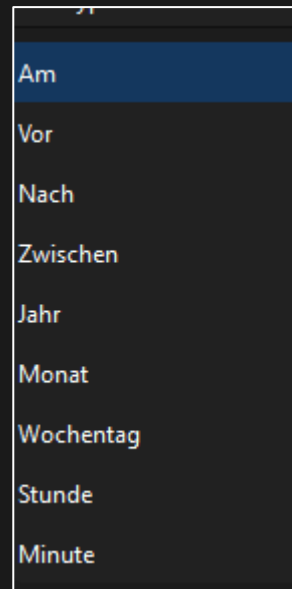


1.2.2.2.1.3 Filter Datum:

Beim Datumsfeld besitzt man unter Filter die nachfolgenden Möglichkeiten:



Es gibt verschiedene Filtertypen, nachdem man eine Datumsspalte filtern kann:



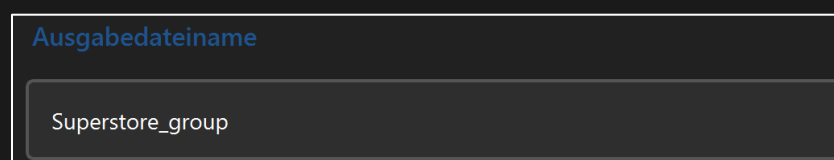
Nach Auswahl eines Filters erfolgt die Extraktion der Werte auf die gleiche Weise wie im Abschnitt „**Zeichen**“ beschrieben (1.2.2.2.1.1).

1.2.2.2.2 Gruppieren

Mit dieser Funktion lassen sich große Datenmengen anhand eines oder mehrerer Kriterien zusammenfassen und z. B. summieren oder zählen.

Die Oberfläche entspricht der regulären **Gruppierungsfunktion**, siehe Kapitel **14.1.6.7**.

In Stream kann der Benutzer im unteren Fensterbereich der neu erzeugten Datei einen individuellen Namen geben. Die neue Datei erscheint nach dem Gruppieren direkt im Explorer und wird erst im Anschluss automatisch geöffnet.



1.2.2.2.3 Verbinden

Diese Funktion verbindet zwei Dateien anhand eines gemeinsamen Schlüssels – vergleichbar mit einem Join. So lassen sich Informationen aus mehreren Quellen effizient zusammenführen.

Die Funktionsweise und Oberfläche entsprechen der im Kapitel **14.1.2** beschriebenen Standardfunktion. Sobald die Verbindungsoperation abgeschlossen ist, taucht die resultierende Datei im Explorer auf und wird dann automatisch eingelesen.

1.2.2.2.4 Anhängen

Hier können mehrere Dateien mit gleicher Struktur untereinander angehängt werden. Die Funktion eignet sich besonders für die Zusammenführung mehrerer Teildateien. Die Bedienoberfläche ist identisch zur regulären **Anfügefunktion**, die im Kapitel **14.1.3** erklärt wird. Die zusammengeführte Datei wird zunächst im Explorer angezeigt und im Anschluss automatisch in die Anwendung geladen.

1.2.2.2.5 Pivot

Diese Funktion transformiert Daten – z. B. indem sie Zeilen in Spalten umwandelt. So lassen sich kompakte Übersichten und Kreuztabellen erstellen. Die Funktionalität entspricht der regulären **Pivot-Funktion**, wie im Kapitel **14.1.6.6** dargestellt.

Ausgabedateiname
Muster_pivot

Im Stream-Kontext kann der Benutzer im unteren Bereich einen neuen Dateinamen vergeben. Nach der Pivotierung wird die neu erzeugte Datei im Explorer angezeigt und danach automatisch geöffnet.

1.2.2.3 Dateioperationen und Verwaltung

Mit **Rechtsklick auf einen Fall** lassen sich organisatorische Aufgaben rund um die Datei direkt durchführen – ohne das Hauptfenster zu verlassen:

Kopieren nach
Verschieben nach
Umbenennen
Im Explorer öffnen

- **Kopieren nach**

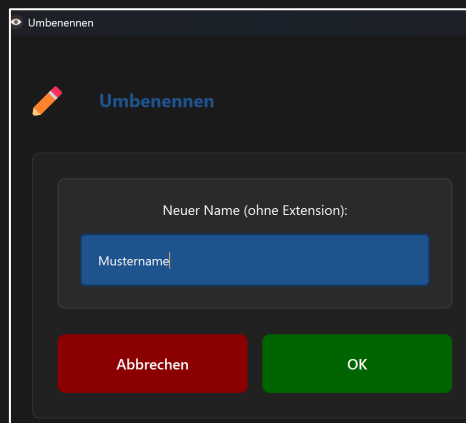
Erstellt eine Kopie der Datei und erlaubt die Auswahl eines Zielordners zur Ablage.

- **Verschieben nach**

Verschiebt die Datei vollständig an einen neuen Speicherort – nützlich für Projektorganisation oder Archivierung.

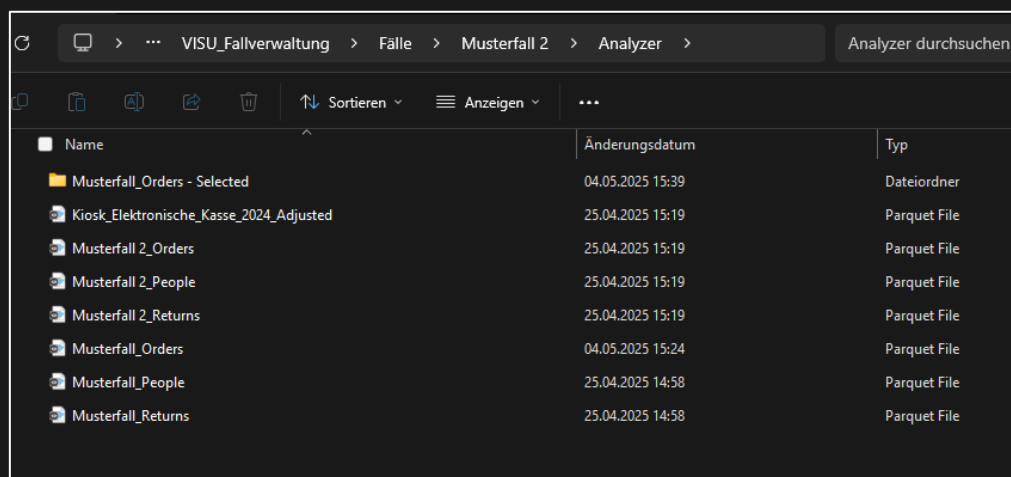
- **Umbenennen**

Öffnet ein Eingabefeld zur Änderung des Dateinamens – hilfreich zur besseren Beschreibung oder Kategorisierung.



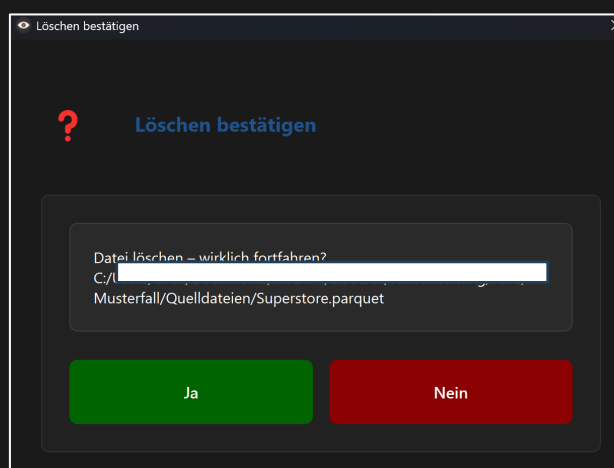
- **Im Explorer öffnen**

Öffnet das Verzeichnis der Datei im Windows Explorer, um dort direkt weiterarbeiten zu können.



- **Löschen**

Entfernt die Datei aus der Anwendung. Bevor die Datei aus der Anwendung entfernt wird, erscheint ein Bestätigungsfenster. Die Löschung erfolgt nur, wenn der Vorgang mit **„Ja“** bestätigt wird, um versehentliche Löschungen zu vermeiden



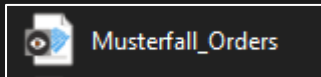
1.2.2.4 Kategorisierung und Statuskennzeichnung

Zur besseren Übersicht und Strukturierung lassen sich Dateien farblich kennzeichnen. Einfach mit **Rechtsklick auf einen Fall** klicken und Sie erhalten folgende Markierungen, welche insbesondere in Analyse- oder Prüfprozessen hilfreich sind:



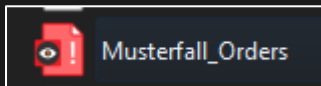
- **Neutral**

Standardstatus ohne besondere Bewertung. Hier bleibt es bei der standardisierten Ansicht:



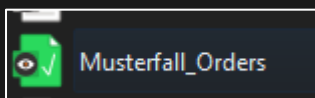
- **Auffällig**

Kennzeichnet den Fall als kritisch oder prüfungswürdig. Diese Markierung dient z. B. der Hervorhebung bei Abweichungen oder Unsicherheiten.



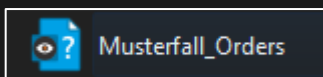
- **Unauffällig**

Weist darauf hin, dass keine besonderen Auffälligkeiten festgestellt wurden – z. B. nach abgeschlossener Prüfung.



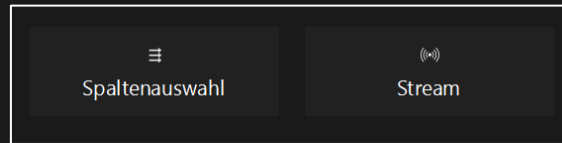
- **Wiedervorlage**

Markiert die Datei zur späteren nochmaligen Prüfung oder Bearbeitung. Praktisch bei offenen Fragen oder zur Überwachung.



1.2.3 Untere Explorer-Symbole

Im unteren Bereich des Explorers befinden sich 2 zusätzliche Funktionen:



1.2.3.1 Spaltenauswahl

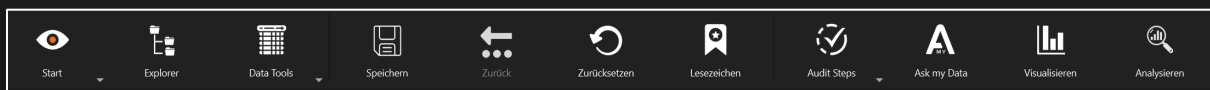
Am unteren Rand des Explorers befindet sich ein zusätzlicher Button „**Spaltenauswahl**“, der als **Schnellzugriff** auf die bereits unter **Kapitel 1.2.2.1** ausführlich erläuterte Funktion dient. Die **Handhabung und der Funktionsumfang** entsprechen vollständig der dort beschriebenen Variante – es handelt sich lediglich um eine komfortable Abkürzung, um die Funktion direktüber den Button nutzen zu können.

1.2.3.2 Stream

Am unteren Rand des Explorers befindet sich der Button „**Stream**“, der als Schnellzugriff auf die in Kapitel **1.2.2.2** ausführlich beschriebene Stream-Funktion dient. Der Button ermöglicht es, den Datenstream direkt aus dem Explorer heraus zu starten – ohne den Umweg über Rechtsklick. Dies spart Zeit und erhöht die Benutzerfreundlichkeit.

Die Stream-Funktion erlaubt es, Daten in Echtzeit zu verarbeiten und hierbei ressourcenschonend vorzugehen, da die gesamten Daten nicht erst zwischengespeichert werden müssen. Dies ist besonders bei großen Datenmengen oder in zeitkritischen Szenarien von Vorteil. Lange Ladezeiten sind dank der Stream-Funktion selbst bei Datenmengen im Milliardenbereich nicht mehr vorhanden.

2 Hauptnavigation



Die **Hauptmenüleiste** von AUDIPY Analyzer bietet verschiedene Werkzeuge für die Datenanalyse und Verwaltung. Die einzelnen Icons repräsentieren spezifische Funktionen, die die Arbeit mit der Software erleichtern. Jedes dieser Symbole ermöglicht den schnellen Zugriff auf zentrale Werkzeuge, sei es zum Importieren, Speichern, Analysieren oder Exportieren von Daten.

In den folgenden Kapiteln werden die einzelnen **Icons und ihre Funktionen** detailliert erläutert, sodass sich Benutzer schnell mit den Möglichkeiten der Software vertraut machen können.

Die Reihenfolge der Kapitel folgt **nicht dem chronologischen Ablauf**, sondern wurde bewusst **nach funktionaler Logik** gegliedert, um eine **schnellere Orientierung und gezieltere Informationssuche** im Handbuch zu ermöglichen.

Kapitalübersicht:

3 Start

4 Explorer

14 Data Tools

5 Speichern

6 Zurück

7 Zurücksetzen

8 Lesezeichen

9 Audit Steps

10 Ask my Data (AMY)

15 Visualisieren

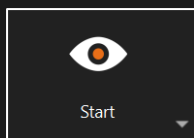
16 Analysieren

2.2 Komprimierung:

Über den kleinen Pfeil an der Seite des Headers kann die Leiste eingeklappt werden und es erscheint zur besseren Veranschaulichung der Datenmenge nur eine komprimierte Ansicht der Icons:



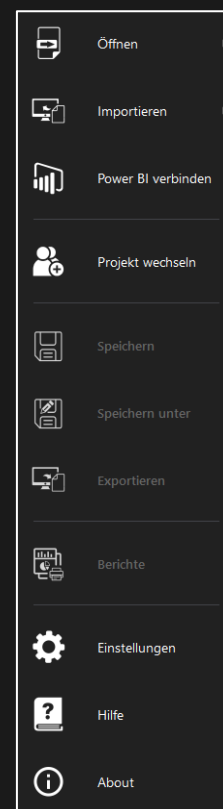
3 Start



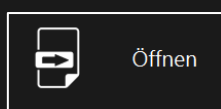
Der **Start** Button steuert den Zugriff auf Programmeinstellungen sowie einigen Zusatzfunktionen.

Bei Klick auf Start öffnet sich folgendes Programmfenster:

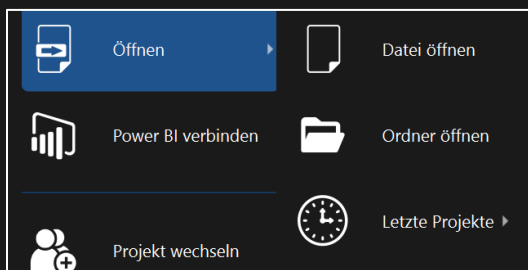
Dieses Menü bietet zentrale Funktionen rund um das Arbeiten mit Dateien und Projekten. Hier können Sie unter anderem neue Dateien oder Ordner öffnen, die Daten exportieren, Verbindungen zu externen Diensten wie Power BI herstellen, Projekte wechseln, Ihre Arbeit speichern oder in ein druckbares PDF-Format (Bericht) umwandeln. Darüber hinaus finden Sie hier den Verlauf Ihrer Arbeitsschritte, allgemeine Einstellungen sowie Informationen zur Anwendung.



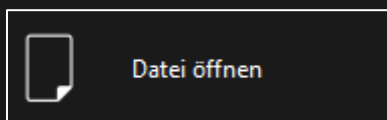
3.1 Öffnen



Bewegt man die Maus auf „Öffnen“ öffnet sich ein seitliches Fenster mit 3 Optionen: **Datei öffnen (3.1.1)** oder einen gesamten **Ordner öffnen (3.1.2)**. Darüber hinaus lassen sich hier auch die **letzten Projekte (3.1.3)** heraus wieder öffnen.

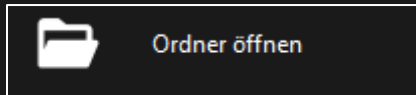


3.1.1 Datei öffnen



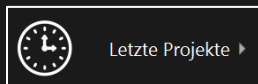
Beim Klick auf **Datei öffnen** startet der Explorer, sodass Sie direkt die gewünschte Datei (entweder Parquet oder csv Dateien) auswählen können. Diese Vorgehensweise ist insbesondere dann wichtig, wenn Sie einen Fall nicht über den Central-Client und dessen Fallmaske öffnen, sondern über das Analyzer-Icon. So haben Sie schnellen Zugriff auf die benötigten Dateien und können diese direkt in der Anwendung bearbeiten. Beim Öffnen von CSV-Dateien ist zu beachten, dass eine temporäre Parquet-Datei erstellt wird, die der Nutzer manuell speichern muss. Der Import über den Central-Client oder über die Import Funktion im Analyzer sollte daher bevorzugt werden, da die Ablage in diesem Fall automatisch im richtigen Ordner erfolgt.

3.1.2 Ordner öffnen



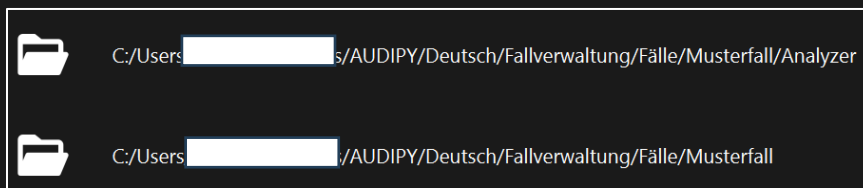
Mit **Ordner öffnen** können Sie direkt den gewünschten Verzeichnisordner in Ihrem Dateisystem aufrufen. Dies erleichtert den Zugriff auf zugehörige Dateien oder weitere Unterordner, ohne die Anwendung verlassen zu müssen bzw. jede einzelne Datei hochzuladen. Dies ist vor allem sinnvoll, wenn mehrere Dateien analysiert werden sollen.

3.1.3 Letzte Projekte

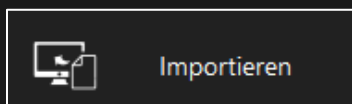


Die Funktion „**Letzte Projekte**“ zeigt dem Benutzer eine Liste der zuletzt geöffneten oder bearbeiteten Projekte an. Dadurch können Projekte schnell erneut geöffnet werden, ohne manuell nach dem Speicherort suchen zu müssen. Außerdem erhält man so direkt die Möglichkeit in den einzelnen Ordner innerhalb der Explorerleiste zu gelangen.

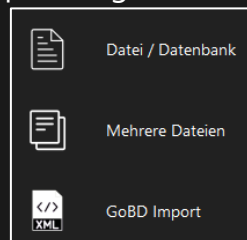
Beispiel:



3.2 Importieren



Beim Start der **Import-Funktion** im AUDIPY Analyzer können Sie zwischen verschiedenen Importmöglichkeiten wählen:



Datei / Datenbank

Import einer einzelnen Datei oder aus einer Datenbank. Hierbei können Sie entweder eine lokale Datei auswählen oder sich mit einer Datenbank verbinden und Daten von dort einlesen.

Mehrere Dateien

Import mehrerer Dateien gleichzeitig. Diese Option eignet sich, wenn Sie mehrere zusammengehörige Dateien in einem Schritt laden möchten.

GoBD Import

Import von Daten im GoBD-konformen XML-Format (Index.xml). Diese Funktion unterstützt speziell den Import von Daten gemäß den Anforderungen der „Grundsätze zur ordnungsmäßigen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff“ (GoBD).

Diese Auswahl ermöglicht es Ihnen, die passende Importmethode für Ihren Anwendungsfall zu wählen — flexibel und direkt im Analyzer, ohne den zusätzlichen Schritt über den AUDIPY Central.

3.2.1 Datei

Bei einer Umwandlung einer **einzelnen Datei** öffnet sich folgendes Dialogfenster:

The screenshot shows the 'Daten Konverter & Importer' dialog box. It has four main sections highlighted with orange boxes and numbered 1 to 4:

- 1. Datenquelle:** A section with the title 'Datenquelle' and the instruction 'Bitte wählen Sie den gewünschten Import aus:'. It contains two radio buttons: 'Einzelne Datei' (selected) and 'Datenbank'.
- 2. Dateityp:** A section with the title 'Dateityp' and a dropdown menu showing '-- Dateityp auswählen --'.
- 3. Ausgangsdatei:** A section with the title 'Ausgangsdatei' and the instruction 'Bitte wählen Sie zuerst einen Dateityp aus'. It contains a text field with the placeholder 'Zuerst Dateityp auswählen...' and a button with three dots '...'.
- 4. Speicherort:** A section with the title 'Speicherort' and a text field showing the path 's:\AUDIPY\Deutsch\Fallverwaltung\Fälle\Musterfall\Analyzer'. It also has a button with three dots '...'.

At the bottom of the dialog, there are two buttons: 'Weiter' (disabled) and 'Abbrechen' (active).

Funktionsweise:

- **Auswahl der Ausgangsdatei ❶:**

Einzelne Datei:

- Wählen Sie diese Option, wenn Sie eine lokal gespeicherte Datei (z. B. eine Excel- oder CSV-Datei) konvertieren möchten.
- Hinweis: Bei CSV / TXT Dateien wird anschließend noch ein Importassistent über den Data-Wizard erfolgen. Hierzu wird auf den gleichen Prozess unter **3.2.4** verwiesen. Auch für Drucklisten (PDF / Text), Excel oder XML Dateien gibt es einen eigenen Importassistenten.

Datenbank (s. **3.2.2** Datenbank):

- Wählen Sie diese Option, wenn Sie die Daten direkt aus einer Datenbank importieren möchten.

- **Festlegen des Dateityps ❷**

- Hier definieren Sie den Typ der Datei, die Sie importieren möchten:
- Es gibt zahlreiche Dateiformate die der AUDIPY unterstützt, um eine Umwandlung zu gewährleisten:
- **Bitte beachten:** Der Dateityp muss immer vor der Auswahl der Datei festgelegt werden. Andernfalls lässt sich keine Datei öffnen.

-- Dateityp auswählen --
Excel Dateien (*.xlsx, *.xls, *.xlsb)
CSV/Text Dateien (*.csv, *.txt, *.del, *.asc, *.tsv)
Drucklisten (PDF / Text Dateien)
Microsoft Access (*.mdb, *.accdb)
SQLite Datenbank (*.db, *.sqlite, *.sqlite3)
JSON Dateien (*.json)
Parquet Dateien (*.parquet)
Arrow/Feather Dateien (*.ipc, *.feather)
Avro Dateien (*.avro)
IDEA Dateien (*.imd)
dBASE Dateien (*.dbf)
XML Dateien (*.xml, *.xaf)
HDF5 Dateien (*.h5, *.hdf5)
Stata Dateien (*.dta)
SPSS Dateien (*.sav)
SAS Dateien (*.sas7bdat)
ORC Dateien (*.orc)

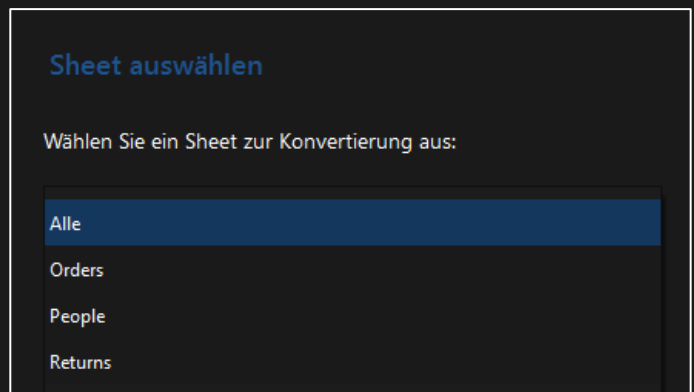
- **Auswahl der Ausgangsdatei ❸:**

- Nach Auswahl des Dateityps wird das Feld erst verfügbar und farblich angezeigt. Auch können dann nur die Dateitypen gefunden werden, die man vorher ausgewählt hat.
- Wählen Sie die Datei aus, die Sie konvertieren möchten. Dies kann über das Eingabefeld erfolgen, indem Sie die Datei direkt angeben, oder durch Klicken auf das Ordner-Symbol, um die Datei aus Ihrem Verzeichnis auszuwählen. Wichtig: Die Datei darf nicht geöffnet sein.

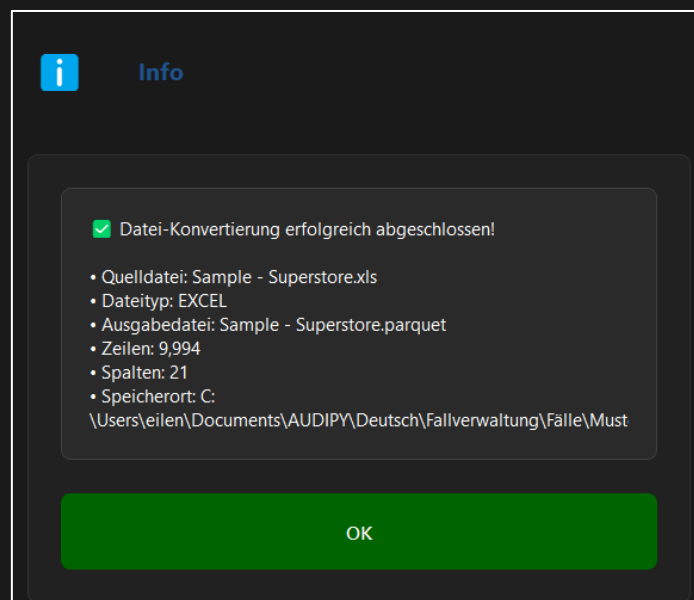
- **Festlegen des Speicherorts ❹:**

- Geben Sie den Ordner an, in dem die konvertierte Datei gespeichert werden soll. Sie können einen Pfad manuell eingeben oder über das Ordner-Symbol auswählen.

- **Voreingestellt** ist bereits der in der Fallverwaltung aktivierte Fall von Ihnen (im Screenshot ist es der Musterfall)
- Starten Sie die Konvertierung anschließend mit „weiter“. Wenn Sie eine Datei mit mehreren Tabellenblättern haben, werden Sie noch mit folgendem Fenster gefragt, ob sie alle Tabellenblätter oder nur ein bestimmtes umwandeln wollen:

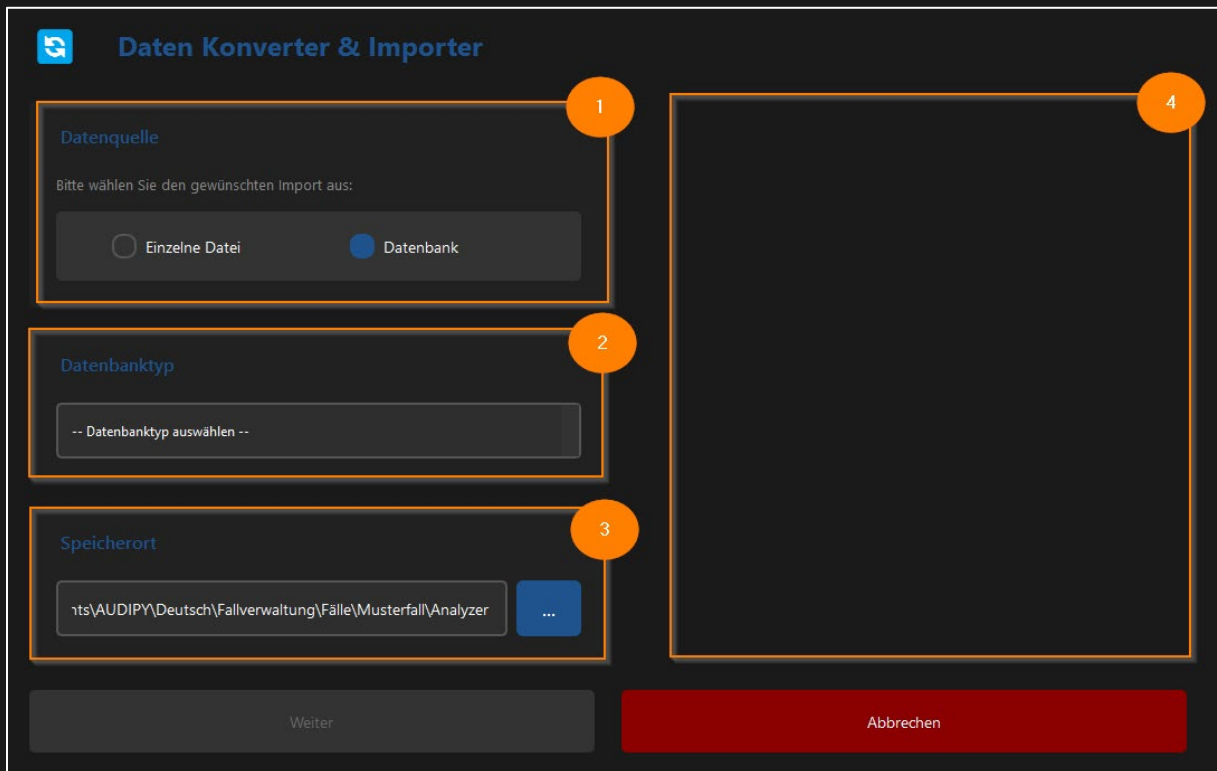


Nach Auswahl erhalten Sie folgendes Bestätigungsfenster:



3.2.2 Datenbank

Bei einer Umwandlung von einer Datei aus einer Datenbank öffnet sich folgendes Dialogfenster:



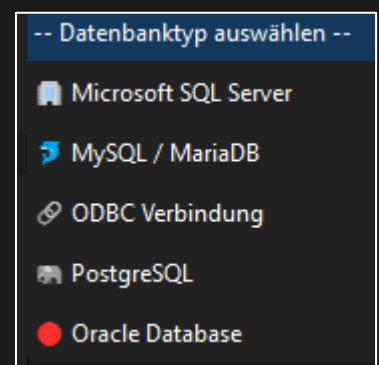
- **Auswahl der Ausgangsdatei ①:**

Datenbank:

- Wählen Sie diese Option, wenn Sie die Daten direkt aus einer Datenbank importieren möchten.

- **Festlegen des Datenbanktyp ②**

- Hier definieren Sie den Typ der Datei, die Sie importieren möchten:
- Es gibt die folgenden Datenbanken die der AUDIPY aktuell unterstützt, um eine Umwandlung zu gewährleisten:
- Je nach Auswahl ändert sich der rechte Teil des Programmfensters ④. Hier sind dann gesonderte Eingaben zur Verbindung der Datenbank notwendig.
- Der Datenbanktyp muss immer ausgewählt werden, ansonsten kann keine Verbindung hergestellt werden.



- **Festlegen des Speicherorts 3:**

- Geben Sie den Ordner an, in dem die konvertierte Datei gespeichert werden soll. Sie können einen Pfad manuell eingeben oder über das Ordner-Symbol auswählen.
- Voreingestellt ist bereits der in der Fallverwaltung aktivierte Fall von Ihnen (im Screenshot ist es der Musterfall)

- **Verbindungsdetails 4**

- Je nach Auswahl des Datenbanktypens (s. 2) werden hier die notwendigen Verbindungsdetails abgefragt. Beispielsweise hier für den Microsoft SQL Server_:

3.2.3 Umwandlung von mehreren Dateien

Bei einer Umwandlung von mehreren Dateien in Parquet nutzen Sie bitte die Schaltfläche „**mehrere Dateien in Parquet**“. Anschließend öffnet sich folgendes Dialogfenster:

Funktionsweise

- **Dateityp-Wahl ①:**

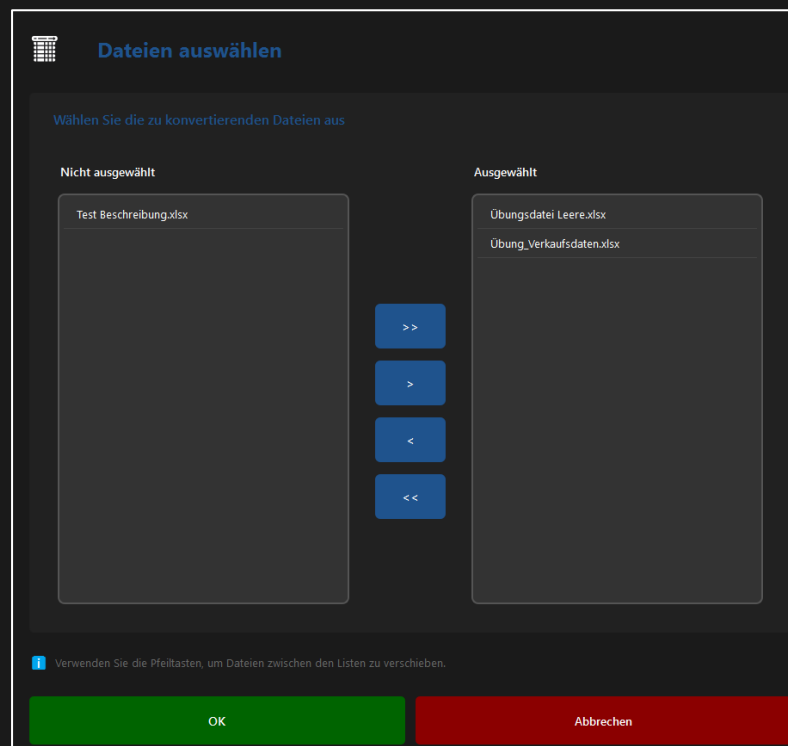
Wählen Sie den Typ der zu konvertierenden Dateien aus. Zur Auswahl stehen **CSV/TXT, Excel oder Drucklisten (PDF/Text)**. Bei den Drucklisten wird zur besseren Übersichtlichkeit an dieser Stelle auf das Kapitel **3.2.8** verwiesen, da zwecks Import noch weitere Maßnahmen über den Report Wizard erfolgen müssen. Auf die Diese Wahl bestimmt, welche Dateien aus dem Ordner verarbeitet werden. **Wichtig: Es werden keine Dateien angezeigt!** Einfach den Ordner mit den mehreren Dateien auswählen.

- **Ausgangsordner ②:**

Geben Sie den Pfad zu dem Ordner an, der die Dateien enthält, die Sie umwandeln möchten. Der Ordner kann direkt eingegeben oder über das Ordner-Symbol ausgewählt werden.

Bei mehreren Dateien im Ordner erscheint folgender Dialog:

Anschließend können Sie aus dem Ordner die entsprechenden Dateien auswählen. Über die Pfeiltasten können bestimmte Tabelle vom Import ausgeschlossen werden. Im folgenden Beispiel werden 2 Dateien importiert und 1 Datei nicht:



- **Beispielsdatei ③ (nur bei Dateityp CSV/TXT & PDF/TXT):**

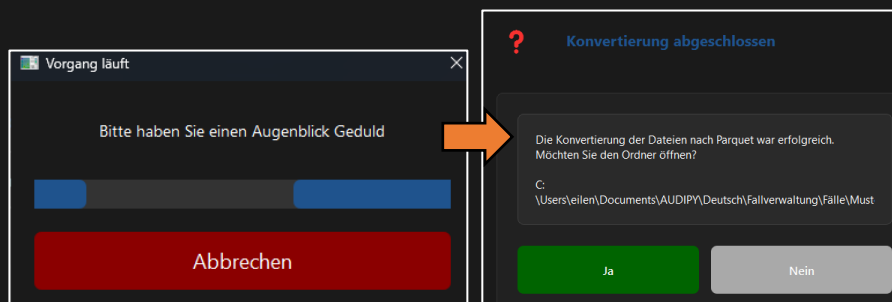
Das Programm zeigt eine enthaltene Beispielsdatei. Sollte keine gefunden werden, wird „Keine Dateien gefunden“ ausgegeben. Diese Datei dient als Referenz, z. B. zur Spaltenstruktur oder für das Zielformat.

- **Optionen ④:**

Gleichartige Dateien zusammenführen: Aktivieren Sie diese Option, wenn mehrere Dateien desselben Typs in eine Datei zusammengeführt werden sollen.

Unterordner durchsuchen: Aktivieren Sie diese Option, wenn auch Dateien in Unterverzeichnissen des gewählten Ausgangsordners berücksichtigt werden sollen. Dies erleichtert **eine schnellere Umwandlung**, weil sie so einfach nur noch den Hauptordner auswählen müssen.

Anschließend erhalten Sie ein kleines Ladefenster und eine Bestätigung:



3.2.4 Data-Wizard „Start“ – CSV / TXT

Hinweis: Nur bei Umwandlung einer CSV Datei in Parquet öffnet sich der Data-Wizard.

Im Data-Wizard erhält man ein **Vorschaufenster** in dem man z.B. unter dem Reiter „**Start ①**“ eine Auswahl zur Kopfzeile ② treffen kann. Es besteht ebenfalls die Möglichkeit, bestimmte Zeilen beim Import zu ignorieren, damit diese nicht in den Datensatz geladen werden. Dies ist besonders nützlich, wenn vor den eigentlichen Daten zusätzliche Beschreibungen oder Kommentare enthalten sind, die für die Analyse nicht benötigt werden.

Mithilfe der Felder „**Kopfzeilen ignorieren**“ und „**Zeilen nach Kopfzeile ignorieren**“ ③ können Sie gezielt festlegen, wie viele Zeilen oberhalb oder unterhalb der Kopfzeile übersprungen werden sollen. So können Sie sicherstellen, dass nur die relevanten Daten in den AUDIPY-Analyzer geladen werden

CSV-zu-Parquet-Konverter

Erweitert

Bitte legen Sie die Trennzeichen für den Datensatz fest und wählen Sie im folgenden Dialog die zu analysierenden Spalten aus: Test.CSV

Start

Optionen

Spalten verbinden

Erweitert

Kopfzeile

☒ Daten enthalten Kopfzeile

Kopfzeilen ignorieren:

1

Zeilen nach Kopfzeile ignorieren:

0

Quick Tips:

- Vorschau der Daten zuerst prüfen
- Rechts-/Linksklick auf Kopfzeilen, um den Typ festzulegen (Datum/Zahl/Text)
- Aktivieren Sie den Streaming-Modus für sehr große Dateien

Tipp: Rechts- oder Linksklick auf einen Spaltenkopf, um das Spaltenformat (Datum/Numerisch/Text) festzulegen.

Zeilen: 14

	CATEGORY	CITY	COUNTRY	CUSTOMER.ID	CUSTOMER.NAM	Category_1	City_1	Country/Region	Customer.ID_1	Customer.NAM_1
1	Furniture	Henderson	United States	CG-12520	Claire Gute					
2	Furniture	Henderson	United States	CG-12520	Claire Gute					
3	Office Supplies	Los Angeles	United States	DV-13045	Darrin Van Huff					
4	Furniture	Fort Lauderdale	United States	SO-20335	Sean O'Donnell					
5	Office Supplies	Fort Lauderdale	United States	SO-20335	Sean O'Donnell					

OK

Abbrechen

Datum / Zeitstempel

Numerisch

Zeichen

Mit Klick (linke oder rechts Maustaste) auf eine Spaltennamen lässt sich der Formatierungstyp verändern. Jeder Formatierungstyp hat sein eigenes Auswahlfenster. Hierfür wird auf Kapitel 12 Datentabelle verwiesen.

semi

 tax

 solutions

Seite 45 | 385

3.2.5 Data-Wizard „Optionen“

Der „**Optionen**“-Reiter ❶ des Data Wizards bietet Einstellungen, um die Datenimportparameter präzise anzupassen. Im Bereich der Trennzeichen kann festgelegt werden, welches Zeichen verwendet wird, um die Datenfelder voneinander zu trennen. Dabei stehen Tabulatoren, Semikolons, Kommata, Leerzeichen oder benutzerdefinierte Trennzeichen zur Auswahl. Diese Einstellung ist entscheidend, um sicherzustellen, dass die Daten korrekt interpretiert werden

CSV-zu-Parquet-Konverter

Bitte legen Sie die Trennzeichen für den Datensatz fest und wählen Sie im folgenden Dialog die zu analysierenden Spalten aus: Test.CSV

Start **Optionen ❶** Spalten verbinden Erweitert

Trennzeichen

☐ Tab ☒ Semikolon ☐ Komma ☐ Leerzeichen ☐ Andere: ❷

Optionen

Einschlusszeichen: " Kodierung: utf8-lossy **Kodierung erkennen**

⚡ Tipp: Rechts- oder Linksklick auf einen Spaltenkopf, um das Spaltenformat (Datum/Numerisch/Text) festzulegen. Zeilen: 14

	CATEGORY	CITY	COUNTRY	CUSTOMER ID	CUSTOMER NAME	Category_1	City_1	Country/Region	Customer ID_1	Customer Name_1
1	Furniture	Henderson	United States	CG-12520	Claire Gute					
2	Furniture	Henderson	United States	CG-12520	Claire Gute					
3	Office Supplies	Los Angeles	United States	DV-13045	Darrin Van Huff					
4	Furniture	Fort Lauderdale	United States	SO-20335	Sean O'Donnell					
5	Office Supplies	Fort Lauderdale	United States	SO-20335	Sean O'Donnell					

OK **Abbrechen**

Zusätzlich können weitere Optionen für die Verarbeitung der Datei ❷ konfiguriert werden. Hierzu gehört die Festlegung des Einschlusszeichens, das beispielsweise Textwerte in Anführungszeichen markiert. Ebenso kann die Zeichenkodierung der Datei angepasst werden, wobei UTF-8 als Standard weit verbreitet ist. Die Schaltfläche „Kodierung erkennen“ ermöglicht es, die Kodierung der Datei automatisch zu bestimmen, falls diese unbekannt ist.

3.2.6 Data-Wizard „Spalten verbinden“

Der Reiter „**Spalten verbinden**“ ① bietet eine Funktion, um zwei ausgewählte Spalten zu einer neuen Spalte zusammenzuführen ②. Dies kann beispielsweise dann nützlich sein, wenn Daten aus getrennten Spalten (wie Vor- und Nachnamen) in einer zusammenhängenden Spalte benötigt werden.

CSV-zu-Parquet-Konverter

Erweitert

Bitte legen Sie die Trennzeichen für den Datensatz fest und wählen Sie im folgenden Dialog die zu analysierenden Spalten aus: Test.CSV

Start Optionen **Spalten verbinden** ① Erweitert

Spalten verbinden

Spalte 1: CATEGORY Spalte 2: CATEGORY Trennzeicher - ☐ neue Spalte ② verbinden

Tipp: Rechts- oder Linksklick auf einen Spaltenkopf, um das Spaltenformat (Datum/Numerisch/Text) festzulegen. Zeilen: 14

	CATEGORY	CITY	COUNTRY	CUSTOMER ID	CUSTOMER NAME	Category_1	Country/R
1	Furniture-Bush Somerset Collection Bookcase	Henderson	United States	CG-12520	Claire Gute		
2	Furniture-Hon Deluxe Fabric Upholstered Stacking ...	Henderson	United States	CG-12520	Claire Gute		
3	Office Supplies-Self-Adhesive Address Labels for ...	Los Angeles	United States	DV-13045	Darrin Van Huff		
4	Furniture-Bretford CR4500 Series Slim Rectangular Table	Fort Lauderdale	United States	SO-20335	Sean O'Donnell		
5	Office Supplies-Eldon Fold 'N' Roll Cart System	Fort Lauderdale	United States	SO-20335	Sean O'Donnell		

OK Abbrechen

Im oberen Bereich des Fensters können die zu verbindenden Spalten ausgewählt werden. **Spalte 1** und **Spalte 2** stehen für die beiden Datenspalten, die miteinander kombiniert werden sollen. Zudem kann ein Trennzeichen definiert werden, das zwischen den Werten der beiden Spalten eingefügt wird, beispielsweise ein Bindestrich, Leerzeichen oder ein anderes gewünschtes Zeichen.

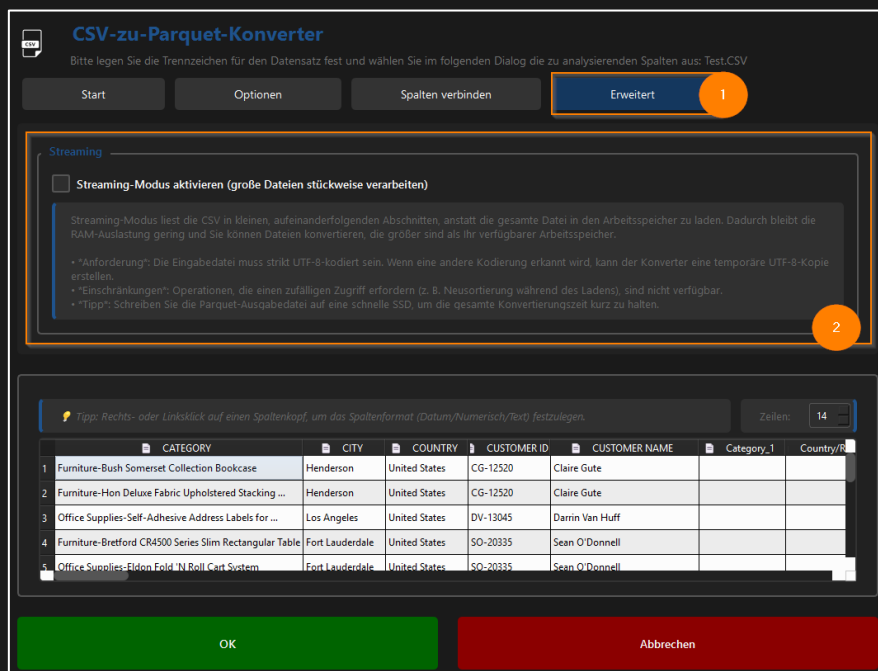
Durch Aktivierung der Option „**neue Spalte**“ wird die kombinierte Spalte als zusätzliche Spalte eingefügt, ohne die bestehenden Daten zu überschreiben. Mit einem Klick auf die Schaltfläche „verbinden“ wird die Konfiguration abgeschlossen und die Spaltenverbindung durchgeführt.

Diese Funktion ermöglicht eine flexible Datenaufbereitung, indem verschiedene Informationen sinnvoll zusammengeführt werden können, bevor die Analyse beginnt.

3.2.7 Data-Wizard „Erweitert“ - Streaming

Im Reiter „Erweitert“ ① steht die Option „Streaming-Modus aktivieren (große Dateien stückweise verarbeiten)“ zur Verfügung.

Der **Streaming-Modus** ② ermöglicht es, CSV-Dateien in kleinen, aufeinanderfolgenden Abschnitten (Chunks) einzulesen, anstatt die gesamte Datei auf einmal in den Arbeitsspeicher zu laden.



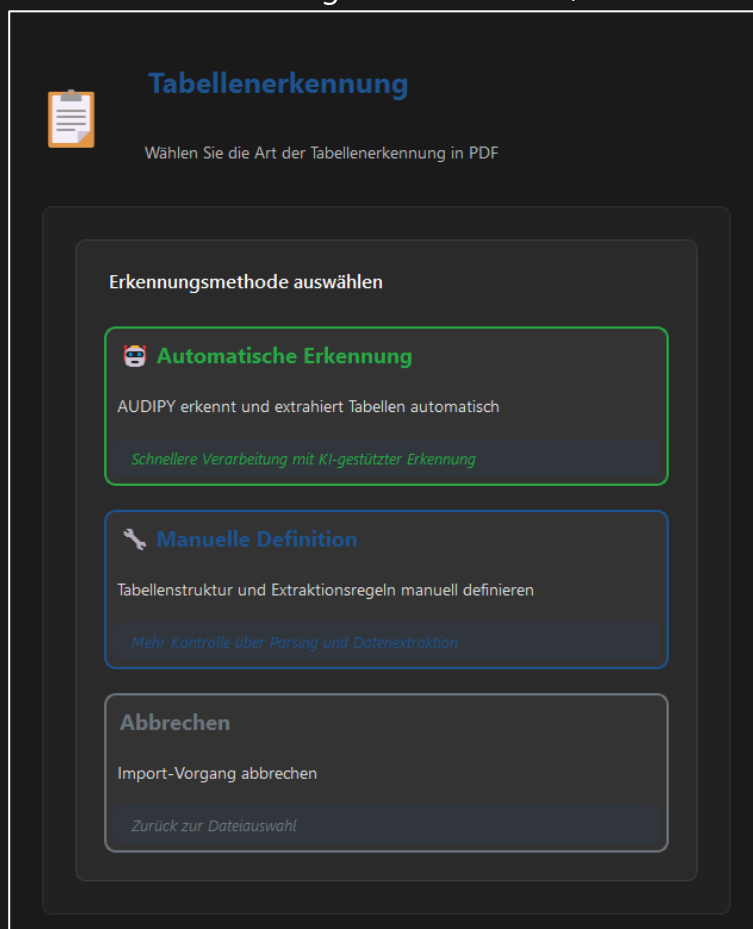
Dies reduziert die RAM-Auslastung erheblich und erlaubt die Verarbeitung sehr großer Dateien, die ansonsten den verfügbaren Speicher überschreiten würden.

Hinweise und Einschränkungen:

- **Anforderung:**
Die Eingabedatei muss strikt **UTF-8-kodiert** sein. Falls eine andere Kodierung erkannt wird, kann der Konverter automatisch eine temporäre UTF-8-Kopie erstellen.
- **Einschränkungen:**
Operationen, die einen zufälligen Zugriff auf die gesamte Datei erfordern (z.B. Neusortierungen während des Ladens), sind im Streaming-Modus **nicht verfügbar**.
- **Tipp:**
Um die Konvertierungsgeschwindigkeit zu optimieren, empfiehlt es sich, die Ausgabedatei (z.B. Parquet-Datei) auf eine schnelle SSD zu schreiben.

3.2.8 Drucklisten (PDF / Text) – Report Wizard

Um Tabellen aus PDF / Texten zu generieren wird nach der generellen Auswahl ein weiteres Dialogfenster geöffnet, mit welchen Sie weitere Möglichkeiten haben, um die die PDF-Datei einzulesen.



3.2.8.1 Automatische Erkennung

Nach Auswahl öffnet sich der Report Wizard. Der **Report Wizard** stellt das Herzstück der automatisierten PDF-Tabellenextraktion dar und wurde speziell für die Anforderungen von

Wirtschaftsprüfern, Steuerberatern und Controllern entwickelt. Diese fortschrittliche Funktion ermöglicht es, aus komplexen PDF-Dokumenten systematisch Tabellendaten zu extrahieren, zu standardisieren und für die weitere Analyse aufzubereiten.

Die Bedeutung dieser Funktionalität wird besonders in der modernen Prüfungs- und Beratungspraxis deutlich: Mandanten übermitteln Daten häufig in heterogenen PDF-Formaten – von Bankauszügen über Umsatzlisten bis hin zu Betriebswirtschaftlichen Auswertungen verschiedener Softwaresysteme. Der Report Wizard automatisiert den traditionell zeitaufwändigen Prozess der manuellen Datenübertragung und reduziert gleichzeitig das Risiko von Übertragungsfehlern erheblich.

Das System arbeitet mit zwei komplementären Extraktionsmethoden, die unterschiedliche Stärken bei verschiedenen PDF-Strukturen aufweisen. Diese **intelligente Dual-Engine-Architektur** gewährleistet eine optimale Erkennungsleistung über das gesamte Spektrum der in der Praxis anzutreffenden Dokumententypen hinweg.

3.2.8.1.1 Praktische Anwendungsszenarien

Die automatische Erkennung ist in der Praxis vor allem aufgrund der einfachen Bedienung und der zeitlichen Komponente sehr nützlich. Sie kann bei den folgenden Bereichen wie folgt eingesetzt werden:

Wirtschaftsprüfung

- **Jahresabschlussprüfung:** Bei der Prüfung von Nebenbüchern und Hilfsbuchhaltungen ermöglicht der Report Wizard die systematische Extraktion von Kontensalden aus verschiedenen Softwaresystemen. Typischerweise liegen diese als PDF-Exporte vor – beispielsweise Summenlisten aus DATEV, SAP oder branchenspezifischen ERP-Systemen. Der Smart Merge harmonisiert automatisch unterschiedliche Kontenbezeichnungen und erstellt eine konsolidierte Datenbasis für analytische Prüfungshandlungen.
- **Umsatzsteuer-Voranmeldungen:** Mandanten übermitteln häufig USt-VA-Daten als PDF-Ausdrucke verschiedener Steuerberatungssoftware. Der Report Wizard extrahiert systematisch die Kennziffern und erstellt eine vergleichbare Datenmatrix für Plausibilitätsprüfungen und Trendanalysen über mehrere Voranmeldungszeiträume hinweg.
- **Bankbestätigungen:** Die Verarbeitung multipler Bankbestätigungen verschiedener Institute wird durch die automatische Extraktion von Kontostammdaten, Salden und Kreditlinien erheblich beschleunigt. Der Lattice-Modus erweist sich hier als besonders effektiv, da Bankformulare typischerweise klare Tabellenstrukturen aufweisen.

Steuerberatung

- **BWA-Analysen:** Betriebswirtschaftliche Auswertungen verschiedener Mandanten können systematisch in eine einheitliche Vergleichsmatrix überführt werden. Dies ermöglicht Benchmarking-Analysen und die Identifikation von Optimierungspotenzialen. Der Smart Merge erkennt automatisch korrespondierende Kontenklassen und harmonisiert unterschiedliche Kontenpläne.
- **Lohnsteuer-Anmeldungen:** Die Extraktion von Lohnsteuer-Anmeldedaten aus PDF-Formularen verschiedener Lohnabrechnungssysteme ermöglicht eine effiziente Erstellung von Jahresstatistiken und die Vorbereitung von Lohnsteuer-Nachschauen. Besonders bei Mandanten mit mehreren Betriebsstätten oder Abrechnungskreisen zeigt sich der Mehrwert der automatisierten Datenkonsolidierung.

- **Vorsteuer-Sammelerfassungen:** Eingangsrechnungen, die als PDF-Listen vorliegen, können systematisch für die Umsatzsteuer-Voranmeldung aufbereitet werden. Der Report Wizard extrahiert Rechnungsdaten, Nettobeträge und Steuerbeträge und erstellt eine prüfbare Datenbasis für die Steueranmeldung.

Controlling

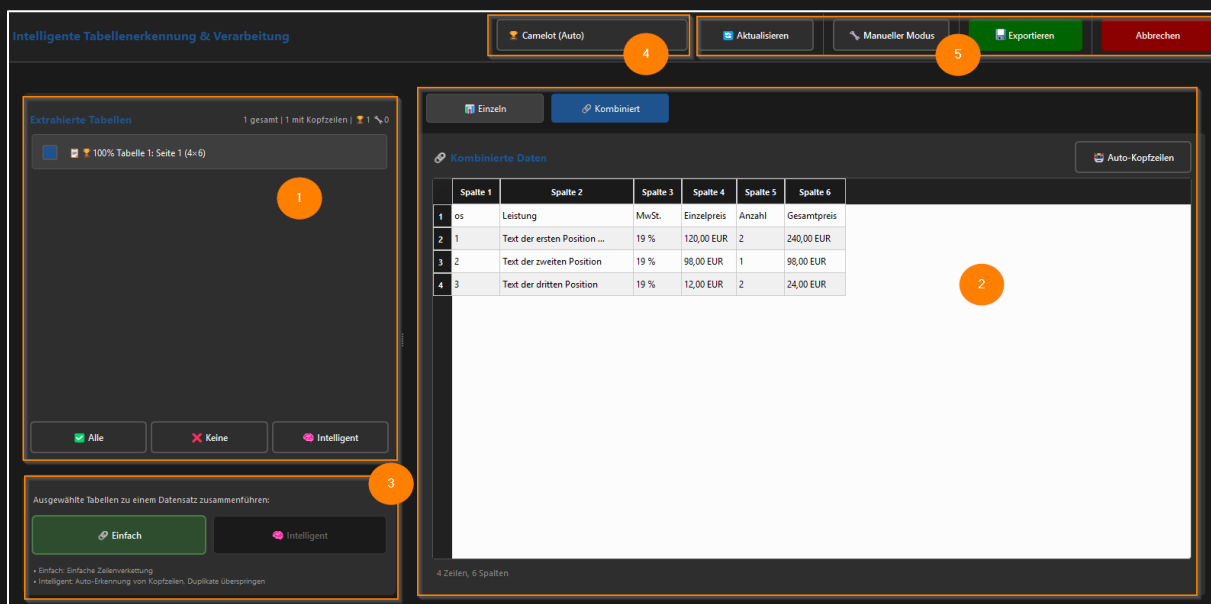
- **Kostenstellenauswertungen:** Kostenstellenberichte aus verschiedenen ERP-Systemen werden in eine einheitliche Controlling-Datenbasis überführt. Dies ermöglicht systemübergreifende Kostenanalysen und die Erstellung konsolidierter Managementberichte. Der Hybrid-Modus bewährt sich bei komplexen Kostenstellenstrukturen mit mehreren Hierarchieebenen.
- **Budget-Ist-Vergleiche:** Planungsberichte und Ist-Auswertungen verschiedener Abteilungen oder Gesellschaften können systematisch in eine Abweichungsanalyse überführt werden. Der Smart Merge harmonisiert unterschiedliche Planungsstrukturen und erstellt eine konsistente Vergleichsbasis.
- **Liquiditätsplanung:** Cashflow-Prognosen verschiedener Geschäftsbereiche werden in eine konsolidierte Liquiditätsplanung integriert. Die automatische Header-Erkennung identifiziert Zeiträume und Kategorien, während der Smart Merge eine einheitliche Periodenstruktur erstellt.

3.2.8.1.2 BImport mit der automatischen Erkennung

Für den reibungslosen Import mit der automatischen Erkennung sollten folgende Schritte vorgenommen werden

1. **Einzeldatei-Modus:** Wählen Sie eine PDF-Datei über den Hauptdialog aus. Der Report Wizard startet automatisch die Tabellenerkennung (3.2.1).
2. **Ordner-Modus:** Navigieren Sie zu einem Verzeichnis mit mehreren PDF-Dateien. Nutzen Sie bei Bedarf die Schaltfläche "📁 Dateien auswählen" für eine selektive Verarbeitung (3.2.3).
3. **Extractor-Konfiguration:** Wählen Sie den geeigneten Erkennungsmodus basierend auf Ihren Dokumentencharakteristika. Bei Unsicherheit empfiehlt sich der Auto-Modus.

Anschließend erhalten Sie folgende Ansicht:



Vorhandende Tabellen ①

- **Ergebnisbewertung:** Prüfen Sie die Erkennungsstatistiken in der Statusleiste. Achten Sie auf die Anzahl erkannter Tabellen und die Genauigkeitswerte bei Camelot-Extraktionen.
- **Qualitätsprüfung:** Nutzen Sie die Einzeltabellen-Vorschau zur stichprobenartigen Kontrolle der Erkennungsqualität. Besondere Aufmerksamkeit verdienen Tabellen mit niedrigen Genauigkeitswerten.

Statussymbole interpretieren:

- : Header wurden erkannt und verarbeitet, Daten sind bereinigt
- : Header wurden automatisch erkannt, aber noch nicht verarbeitet
- : Keine klaren Header identifiziert
- : Camelot-Extraktion mit optionaler Genauigkeitsangabe
- : PDFPlumber-Extraktion

Auswahloptionen:

- **Alle:** Markiert sämtliche erkannte Tabellen für die Weiterverarbeitung
- **Keine:** Hebt alle Markierungen auf
- **Intelligent:** Intelligente Auswahl basierend auf Strukturähnlichkeit oder Dateizugehörigkeit

Die Intelligent-Auswahl analysiert im Ordner-Modus sowohl strukturelle Gemeinsamkeiten (gleiche Spaltenanzahl und Datentypen) als auch dateispezifische Gruppierungen. Dies ermöglicht eine effiziente Verarbeitung homogener Datensätze.

Detailansicht und Headeroptimierung 2

- **Einzel- oder Kombinierte Ansicht:** Entscheiden Sie selbst.
- **Automatische Header-Erkennung:** Aktivieren Sie "🤖 Auto Headers" für einzelne Tabellen oder nutzen Sie die globale Header-Erkennung im kombinierten Modus.
- **Manuelle Header-Anpassung:** Verwenden Sie das Rechtsklick-Menü auf Spaltenüberschriften für präzise Umbenennungen entsprechend Ihrer fachlichen Terminologie.
- **Header-Bereinigung:** Bei erkannten Headern werden diese automatisch aus den Datenzeilen entfernt und als Spaltenüberschriften gesetzt.

Tabellenauswahl und -kombination 3

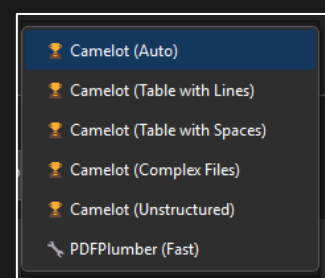
Der **Kombinationsbereich** bietet zwei strategische Ansätze:

-  **Einfach:** Einfache Verkettung aller ausgewählten Tabellen ohne strukturelle Anpassungen. Geeignet für Tabellen mit identischem Aufbau oder wenn manuelle Nachbearbeitung geplant ist.
-  **Intelligent:** Intelligente Kombination mit automatischer Header-Harmonisierung, Duplikatserkennung und struktureller Anpassung. Analysiert Datentypen, erkennt gemeinsame Spaltenschemata und entfernt automatisch redundante Header-Zeilen.

Wichtig: Erst nach Auswahl wird die Exportfunktion oben rechts im Dialogfenster grün hinterlegt und kann entsprechend verwendet werden.

Extractor-Auswahl und Modi 4

Das **Extractor-Auswahlfeld** in der Haupttoolbar ermöglicht die strategische Auswahl der Erkennungsmethode entsprechend der Dokumentencharakteristika und Qualitätsanforderungen. Diese Auswahl beeinflusst fundamental sowohl die Erkennungsgenauigkeit als auch die Verarbeitungsgeschwindigkeit.



Verfügbare Camelot-Modi umfassen:

-  **Camelot (Auto):** Intelligenter Automatik-Modus, der sequenziell verschiedene Erkennungsverfahren testet und das beste Ergebnis auswählt. Ideal für heterogene Dokumentenbestände und als Standard-Einstellung für die meisten Anwendungsfälle.
-  **Camelot (Table with Lines):** Lattice-Methode für Tabellen mit klar definierten Rahmenlinien und Gitternetzstrukturen. Besonders effektiv bei

traditionellen Rechnungsformaten, ERP-System-Ausdrucken und strukturierten Finanzberichten.

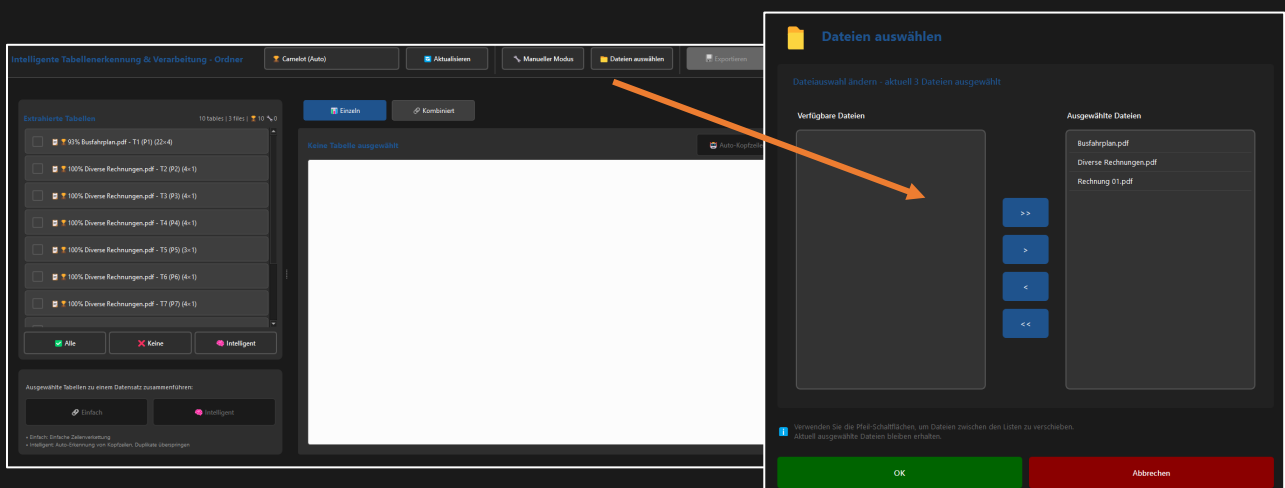
- 🏆 **Camelot (Table with Spaces)**: Stream-Methode für Tabellen, die primär durch Leerräume und Textausrichtung strukturiert sind. Optimal für ASCII-basierte Berichte, Kontoauszüge und ältere Systemausdrucke.
- 🏆 **Camelot (Complex Files)**: Hybrid-Methode für komplexe Dokumentlayouts mit gemischten Strukturelementen. Entwickelt für anspruchsvolle Layouts wie mehrspaltige Berichte oder Dokumente mit eingebetteten Grafiken.
- 🏆 **Camelot (Unstructured)**: Network-Methode mit Machine Learning-Algorithmen für schwer strukturierbare Dokumente. Einsatz bei handschriftlich ergänzten Formularen oder stark variierenden Layoutstrukturen.
- 🛠️ **PDFPlumber (Fast)**: Schnelle und ressourcenschonende Alternative mit hoher Zuverlässigkeit für Standardformate. Empfehlenswert für große Dokumentenbestände und zeitkritische Verarbeitungen.

Die Auswahl beeinflusst fundamental nicht nur die Erkennungsqualität, sondern auch die verfügbaren Genauigkeitsindikatoren und Verarbeitungsgeschwindigkeit. Bei Camelot-Modi wird zusätzlich ein Genauigkeitswert in Prozent angezeigt, der die Verlässlichkeit der Extraktion quantifiziert.

Steuerungselemente der Haupttoolbar 5

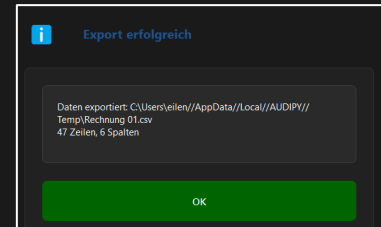
Die **Haupttoolbar** bietet zentrale Steuerungsfunktionen für den Workflow:

- 🔄 **Aktualisieren**: Startet die Tabellenerkennung mit den aktuellen Einstellungen neu. Besonders relevant nach Änderungen der Extractor-Konfiguration oder bei inkonsistenten Ergebnissen.
- 🛠️ **Manueller Modus**: Wechselt zur manuellen Strukturdefinition für Spezialfälle, in denen die automatische Erkennung an ihre Grenzen stößt. (s. auch Kapitel 3.2.8.2)
- 📁 **Dateien auswählen** (nur im Ordner-Modus): Öffnet einen erweiterten Dateiauswahl-Dialog für die selektive Verarbeitung von PDF-Dateien aus einem Verzeichnis.



- **Export** (grün): Exportiert die kombinierten Tabellendaten wahlweise als CSV oder Excel-Datei. Wird erst nach erfolgreicher **Tabellenkombination** 3 aktiviert.
- **Abbrechen** (rot): Schließt den Report Wizard ohne Datenübernahme.

Nach erfolgreichem Export erscheint ein Bestätigungsfenster, und der Data-Wizard öffnet sich, um die weitere Formatierung und die Umwandlung in eine Parquet-Datei durchzuführen (vgl. 3.2.4).



3.2.8.1.3 Erweiterte Strategien und Optimierung

Möchte man den automatischen Import von PDF / TXT Dateien optimieren, sollte mal folgende Details nicht außer Acht lassen:

Optimierung der Erkennungsqualität

- Die **Extractor-Wahl** sollte dokumentenspezifisch erfolgen. Für strukturierte Berichte aus professionellen Buchhaltungssystemen empfiehlt sich primär der Lattice-Modus, während für legacy-Systeme oder ASCII-basierte Exporte der Stream-Modus oft superior ist. Bei unklaren Dokumentenstrukturen bietet der Auto-Modus eine ausgezeichnete Ausgangsbasis.
- **Qualitätsindikatoren** sind systematisch zu berücksichtigen: Camelot-Genauigkeitswerte unter 70% deuten auf strukturelle Herausforderungen hin und erfordern eine kritische Ergebnisvalidierung. In solchen Fällen kann ein Wechsel zu PDFPlumber oder eine manuelle Nachbearbeitung angebracht sein.
- Die **Header-Strategie** beeinflusst maßgeblich die Datenqualität nachgelagerter Analysen. Bei heterogenen Datenquellen sollte zunächst die automatische Header-Erkennung erfolgen, gefolgt von einer fachlichen Normalisierung der Spaltenbezeichnungen entsprechend der unternehmensinternen Terminologie oder branchenüblichen Standards.

Workflow-Integration und Datenqualität

- **Batch-Verarbeitung** großer Dokumentenbestände erfordert eine systematische Vorgehensweise: Gruppieren Sie Dokumente nach Strukturähnlichkeit und verarbeiten Sie homogene Bestände gemeinsam. Dies maximiert die Effektivität der Smart-Auswahl und reduziert den manuellen Nachbearbeitungsaufwand.

- **Validierungsstrategien** sollten stichprobenbasiert erfolgen: Kontrollieren Sie mindestens 10% der erkannten Tabellen detailliert und achten Sie dabei besonders auf numerische Plausibilität und Vollständigkeit. Bei kritischen Datensätzen (beispielsweise Bilanzpositionen) empfiehlt sich eine 100%-Kontrolle.
- Die **Ergebnisdokumentation** sollte sowohl die verwendeten Erkennungsparameter als auch etwaige manuelle Adjustierungen umfassen. Dies gewährleistet die Nachvollziehbarkeit und Reproduzierbarkeit der Datenaufbereitung im Rahmen der beruflichen Qualitätssicherung.
- **Versionskontrolle** der Ausgangsdaten ist besonders bei mandantenbasierten Workflows kritisch: Dokumentieren Sie die Quell-PDFs mit Dateinamen, Erstellungsdatum und Hash-Werten, um bei Nachfragen oder Korrekturen eine eindeutige Zuordnung zu gewährleisten.

Performance und Skalierung

- Bei **umfangreichen Dokumentenbeständen** empfiehlt sich eine gestaffelte Verarbeitung: Beginnen Sie mit einer repräsentativen Stichprobe zur Optimierung der Parameter, bevor Sie die Gesamtverarbeitung starten. Dies vermeidet zeitaufwändige Wiederholungen bei suboptimalen Einstellungen.
- **Ressourcenmanagement** ist besonders bei Camelot-Verarbeitungen relevant: Der Hybrid- und Network-Modus sind rechenintensiv und sollten nur bei entsprechenden Systemressourcen und spezifischen Anforderungen eingesetzt werden. Für Routineverarbeitungen ist PDFPlumber oft die effizientere Alternative.
- Die **Fehlerbehandlung** sollte proaktiv erfolgen: Dokumentieren Sie systematisch auftretende Erkennungsprobleme und entwickeln Sie dokumentenspezifische Lösungsstrategien. Häufige Probleme umfassen rotierte Seiten, mehrsprachige Inhalte oder beschädigte PDF-Strukturen.
- Diese systematische Herangehensweise gewährleistet eine professionelle Integration des Report Wizards in bestehende Prüfungs-, Beratungs- und Controlling-Workflows und maximiert dabei sowohl die Effizienz als auch die Datenqualität der automatisierten Tabellenextraktion.

3.2.8.2 Manuelle Definition

Der **manuelle Strukturdefinitions-Modus** des Report Wizards repräsentiert die präziseste und flexibelste Methode zur Verarbeitung strukturierter Textdaten und PDF-Dokumente. Diese erweiterte Funktionalität wurde speziell für komplexe Dokumentenstrukturen entwickelt, die über die Möglichkeiten der automatischen Tabellenerkennung hinausgehen oder höchste Genauigkeitsanforderungen stellen.

Die fundamentale Stärke dieser Herangehensweise liegt in der **vollständigen Kontrolle** über den Extraktionsprozess: Anwender definieren explizit, welche Zeilen welche Bedeutung haben, wie diese erkannt werden und welche Datenfelder aus ihnen extrahiert werden sollen. Dies ermöglicht die Verarbeitung selbst hochkomplexer Berichtsstrukturen, wie sie beispielsweise in Legacy-Systemen, Mainframe-Ausdrucken oder stark variierenden Berichtsformaten auftreten.

Das System basiert auf einem hierarchischen Strukturmodell, das die natürliche Gliederung von Geschäftsberichten widerspiegelt: Header-Bereiche enthalten Metainformationen, Spaltenüberschriften definieren die Datenstruktur, Detailzeilen beinhalten die eigentlichen Geschäftsdaten und Footer-Bereiche können Summen oder Abschlussinformationen enthalten. Diese logische Trennung ermöglicht nicht nur eine präzise Datenextraktion, sondern auch die Rekonstruktion der ursprünglichen Datenbeziehungen.

3.2.8.2.1 Praktische Anwendungsszenarien

Der manuelle Import bietet in der Praxis einige individuelle Besonderheiten:

Wirtschaftsprüfung

- **Hauptbuch-Analysen:** Bei der Prüfung von Hauptbuchauszügen aus Legacy-Systemen ermöglicht die manuelle Strukturdefinition die präzise Extraktion von Kontonummern, Buchungstexten, Soll- und Habenwerten auch bei komplexen Formatierungen. Header-Definitionen erfassen Datumsangaben und Mandanteninformationen, während Datenzeilen systematisch alle Buchungssätze extrahieren. Footer-Definitionen sichern die Erfassung von Saldovortrag und Endsummen für Abstimmungsrechnungen.
- **Nebenchbuch-Auswertungen:** Debitorenbuchhaltungen oder Kreditorenbuchhaltungen verschiedener Softwaresysteme können durch spezifische Pattern-Definitionen einheitlich verarbeitet werden. Die Feldauswahl ermöglicht die granulare Extraktion von Kundennummern, Rechnungsnummern, Fälligkeitsdaten und offenen Posten. Smart-Templates reduzieren den Aufwand bei wiederkehrenden Mandantensystemen.
- **Steuerliche Nebenerklärungen:** Komplexe Anlagenbuchhaltungen oder Abschreibungsübersichten können systematisch für steuerliche Zwecke aufbereitet werden. Die hierarchische Strukturierung erfasst Anlagenkategorien

im Header, Abschreibungsmethoden in Spaltenüberschriften und detaillierte Bewegungsdaten in den Datenzeilen.

Steuerberatung

- **Kassenbuch-Digitalisierung:** Handschriftlich geführte oder legacy-basierte Kassenbücher können durch präzise Feldauswahl digitalisiert werden. Pattern-Definitionen erkennen Datumsmuster, während Feldauswahl Einnahmen, Ausgaben und Verwendungszwecke systematisch erfasst. Die Datenqualität übertrifft häufig manuelle Übertragungen erheblich.
- **Lohnbuchhaltungs-Konsolidierung:** Verschiedene Lohnabrechnungssysteme können durch System-spezifische Templates einheitlich verarbeitet werden. Header-Bereiche erfassen Abrechnungsperioden, Datenzeilen extrahieren Mitarbeiterdaten, Brutto-/Nettolöhne und Steuerabzüge. Footer-Auswertungen summieren Gesamtbelastungen für die Steueranmeldung.
- **USt-Voranmeldungs-Analysen:** Komplexe Umsatzsteuer-Voranmeldungen verschiedener Mandate können strukturiert für Vergleichsanalysen aufbereitet werden. Die Extraktion erfolgt nach einheitlichen Schemata unabhängig vom verwendeten Steuerberatungsprogramm.

Controlling

- **Management-Reporting:** Heterogene ERP-System-Berichte werden durch definierte Strukturmuster in einheitliche Controlling-Datensätze überführt. Header-Informationen erfassen Gesellschaft und Periode, Datenzeilen extrahieren Kostenstellen-, Kostenträger- und Leistungsartendaten. Die konsolidierte Datenbasis ermöglicht gesellschaftsübergreifende Analysen und Benchmarks.
- **Budget-Controlling:** Plan-Ist-Vergleiche verschiedener Organisationseinheiten können systematisch harmonisiert werden. Template-basierte Verarbeitung gewährleistet konsistente Datenstrukturen auch bei unterschiedlichen Planungstools und Reportingformaten.
- **Liquiditätsplanung:** Cashflow-Berichte verschiedener Geschäftsbereiche oder Tochtergesellschaften werden in einheitliche Liquiditätsmodelle integriert. Die granulare Feldauswahl ermöglicht die Erfassung unterschiedlicher Periodenstrukturen und Währungsangaben.
- **Kostenstellenrechnung:** Komplexe Kostenstellenberichte mit mehrstufigen Hierarchien können durch hierarchische Pattern-Definition systematisch aufgeschlüsselt werden. Dies ermöglicht detaillierte Kostenanalysen und Abweichungsanalysen auf verschiedenen Verdichtungsebenen.

3.2.8.2.2 Bedienelemente und Konfiguration

Die **Hauptwerkzeugleiste** bietet zentrale Steuerungsfunktionen für den gesamten Definitionsprozess:

- **Datei öffnen:** Lädt einzelne Dateien oder Ordner zur Verarbeitung. Unterstützt Textdateien (TXT, PRN, ASC, DAT) und PDF-Dokumente bei installiertem PyMuPDF.
- **Vorlage laden:** Importiert zuvor gespeicherte Strukturdefinitionen für die Wiederverwendung bei ähnlichen Dokumententypen. Templates enthalten alle Definitionsregeln und Feldkonfigurationen.
- **Vorlage speichern:** Exportiert die aktuellen Strukturdefinitionen als JSON-Template für spätere Verwendung oder Übertragung auf andere Arbeitsplätze.
- **Dateien auswählen** (nur im Ordner-Modus): Öffnet einen erweiterten Dateiauswahl-Dialog zur selektiven Verarbeitung spezifischer Dateien aus einem Verzeichnis.
- **Export:** Führt die Datenextraktion durch und erstellt CSV- oder Excel-Dateien. Im Einzeldatei-Modus wird eine Datei verarbeitet, im Ordner-Modus werden alle ausgewählten Dateien in eine konsolidierte Ausgabedatei kombiniert.

Strukturdefinitions-Panel

Das **linke Panel** präsentiert eine kartenbasierte Übersicht aller definierten Strukturen mit aussagekräftigen Statusindikatoren:

Strukturtyp-Symbole:

-  **Header:** Kopfzeilendefinitionen für Dokumentmetadaten und Kontextinformationen
-  **Spaltenüberschrift:** Definition von Spaltenbezeichnungen und Datenfeldbeschreibungen
-  **Datenzeile:** Kernstück der Extraktion - Definition der eigentlichen Geschäftsdaten
-  **Footer:** Fußzeilendefinitionen für Summen, Totals oder Abschlussinformationen

Interaktionsmöglichkeiten:

- **Auswahl:** Klick auf eine Karte lädt die Definition zur Bearbeitung
- **Löschung:** Rotes X-Symbol entfernt die Definition nach Bestätigung
- **Statusanzeige:** Kompakte Darstellung von Pattern und Feldanzahl

Die hierarchische Reihenfolge der Definitionen ist durch die Geschäftslogik vorgegeben: Header → Spaltenüberschriften → Datenzeilen → Footer. Diese Sequenz spiegelt die typische Struktur von Finanz- und Geschäftsberichten wider.

Workflow-Steuerung und Zeilenerstellung

Der **zentrale Workflow-Bereich** leitet systematisch durch den Definitionsprozess:

Zeilentyp-Buttons:

-  **Header:** Startet die Definition einer Kopfzeile. Hierarchisch vor allen anderen Definitionen erforderlich.
-  **Spaltenüberschrift:** Definiert Spaltenbezeichnungen. Muss vor Datenzeilen erstellt werden.
-  **Datenzeile:** Erstellt die zentrale Datenextraktionsdefinition. Zwingend erforderlich für Export.
-  **Footer:** Definiert Fußzeilen. Kann nur nach Datenzeilen-Definition erstellt werden.

Bearbeitungssteuerung:

- **Abschließen (grün):** Speichert die aktuelle Definition und kehrt zur Übersicht zurück
- **Verwerfen (rot):** Bricht die Bearbeitung ab und verwirft alle Änderungen
-  **Prüfen:** Validiert die aktuelle Definition auf Vollständigkeit und Konsistenz

Pattern-Editor und Selektoren

Der **Pattern-Editor** ermöglicht die präzise Definition von Erkennungsmustern für verschiedene Zeilentypen:

Selector-Buttons:

- **§ Text:** Fügt Platzhalter für Buchstaben (A-Z, a-z, äöü) ein
- **# Nummer:** Definiert Positionen für Ziffern (0-9)
- **Leerzeichen:** Markiert Whitespace-Bereiche zur strukturellen Orientierung
- *** Beliebig:** Universeller Platzhalter für jedes Zeichen

Pattern-Eingabefeld: Monospace-formatiertes Eingabefeld für direkte Pattern-Eingabe oder kombinierte Verwendung mit Selector-Buttons. Unterstützt sowohl symbolische Platzhalter als auch Literaltext.

Die Pattern-Syntax ermöglicht hybride Ansätze: "Summe: ########.##" würde beispielsweise Zeilen erkennen, die mit "Summe: " beginnen, gefolgt von einem numerischen Wert in deutschem Zahlenformat.

Live-Vorschau und Feldauswahl

Das **StructureLineDisplay** bietet eine interaktive Vorschau der Feldauswahl mit präziser visueller Kontrolle:

Visualisierungsfeatures:

- **Farbkodierung:** Unterschiedliche Hintergrundfarben signalisieren Match-Status, Fehler oder normale Anzeige
- **Feldmarkierungen:** Definierte Bereiche werden visuell hervorgehoben mit Positionsangaben
- **Resize-Handles:** Feldgrenzen können durch Ziehen an den Rändern angepasst werden

Interaktionsmöglichkeiten (nur bei Datenzeilen):

- **Mausauswahl:** Klicken und Ziehen definiert neue Datenfelder
- **Größenänderung:** Präzise Anpassung der Feldgrenzen durch Drag-and-Drop
- **Doppelklick:** Entfernt bestehende Felddefinitionen nach Bestätigung

Feld-Management

Das **rechte Panel** verwaltet die definierten Datenfelder mit erweiterten Bearbeitungsmöglichkeiten:

Feldkarten-Features:

- **Namensbearbeitung:** Direkte Umbenennung der Feldbezeichnungen für fachliche Klarheit
- **Positionsanzeige:** Von/Bis-Angaben für präzise Spaltenpositionierung
- **Beispielwerte:** Live-Vorschau der extrahierten Daten aus der aktuellen Zeile
- **Löschfunktion:** Entfernung einzelner Felder mit Bestätigungsdialog

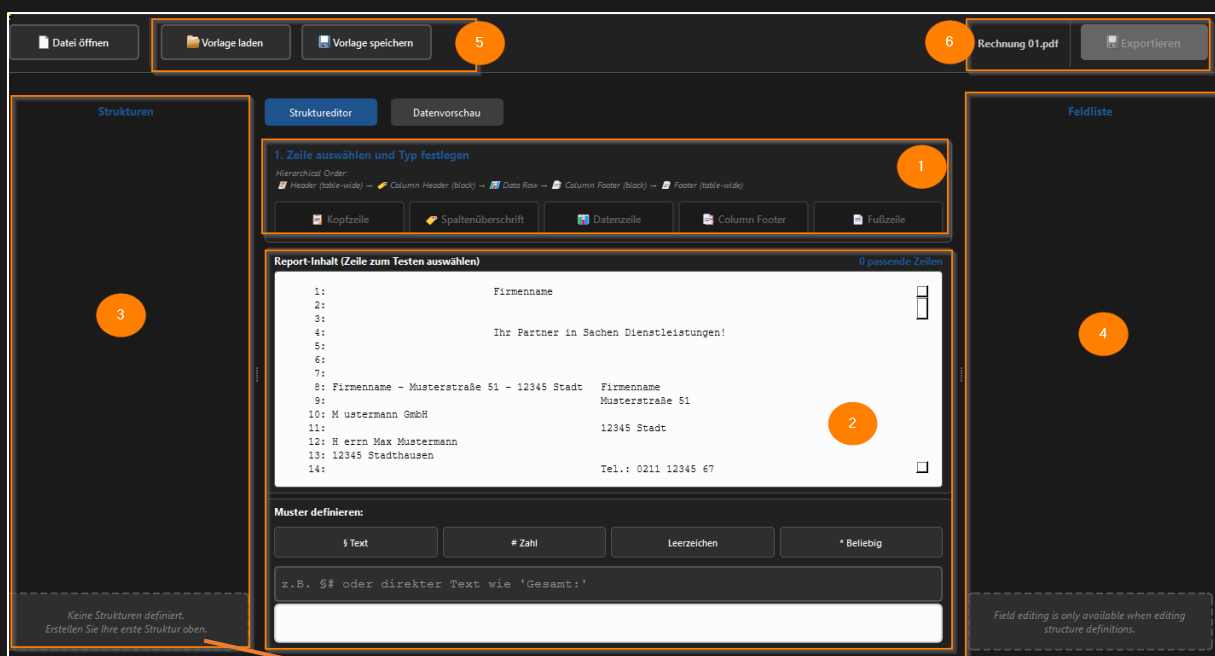
Bearbeitungsmodi:

- **Vollbearbeitung:** Während der Datenzeilen-Definition sind alle Funktionen verfügbar
- **Nur-Anzeige:** Bei anderen Zeilentypen werden Felder schreibgeschützt angezeigt
- **Selektive Löschung:** Feldentfernung ist unabhängig vom Bearbeitungsmodus möglich

3.2.8.2.3 Import mithilfe der manuellen Definition

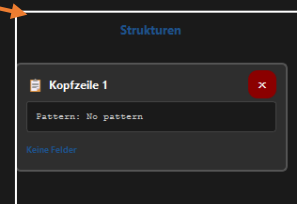
Beim Import über die manuelle Definition gehen Sie wie folgt vor:

1. **Einzeldatei-Modus:** Verwenden Sie "Datei öffnen" zur Auswahl einer spezifischen Datei. Das System lädt den Inhalt automatisch und zeigt eine nummerierte Zeilenübersicht.
2. **Ordner-Modus:** Navigieren Sie zu einem Verzeichnis mit mehreren strukturell ähnlichen Dateien. Wählen Sie eine repräsentative Beispieldatei für die Template-Erstellung aus.
3. **Dateiauswahl verfeinern:** Im Ordner-Modus nutzen Sie "Dateien auswählen" zur Definition der zu verarbeitenden Dateien. Dies ermöglicht selektive Verarbeitung und schließt irrelevante Dateien aus.
4. **Dokumentenanalyse:** Studieren Sie die Dokumentenstruktur in der Zeilenübersicht. Identifizieren Sie wiederkehrende Muster, Kopfzeilen, Datenabschnitte und Fußbereiche.



Hierarchische Strukturdefinition ① & ③

- **Header-Definition (optional):** Beginnen Sie mit Kopfzeilendefinitionen, falls das Dokument Metainformationen enthält. Selektieren Sie eine charakteristische Zeile und verwenden Sie "Header". Die linke Seite wird nach Speicherung verändert. ③
- **Spaltenüberschriften definieren (optional):** Markieren Sie Zeilen mit Spaltenbezeichnungen und erstellen Sie entsprechende Definitionen über "Spaltenüberschrift".



- **Datenzeilen erstellen (erforderlich):** Identifizieren Sie die Kerngeschäftsdaten und definieren Sie diese über "  Datenzeile". Dies ist der zentrale Schritt für die spätere Datenextraktion.
- **Footer-Abschluss (optional):** Definieren Sie Fußzeilen für Summen oder Abschlussinformationen über "  Footer".

Alle Anpassungen müssen gespeichert werden, anschließend erscheint die Speicherung auf der linken Dialogseite.

2. Definition 'Detail 1' bearbeiten

Fertigstellen **Nicht speichern** **Check**

Report-Inhalt (Zeile zum Testen auswählen) 0 passende Zeilen

25: Pos	Leistung	MwSt.	Einzelpreis	Anzahl	Gesamtpreis	
26: 1	Text der ersten Position	19 %	120,00 EUR	2	240,00 EUR	☐
27:	Mehrere Zeilen sind möglich...					☐
28: 2	Text der zweiten Position	19 %	98,00 EUR	1	98,00 EUR	☒
29: 3	Text der dritten Position	19 %	12,00 EUR	2	24,00 EUR	☐
30:						
31:						
32:						
33:						☐

Muster definieren:

z.B. \$# oder direkter Text wie 'Gesamt:'

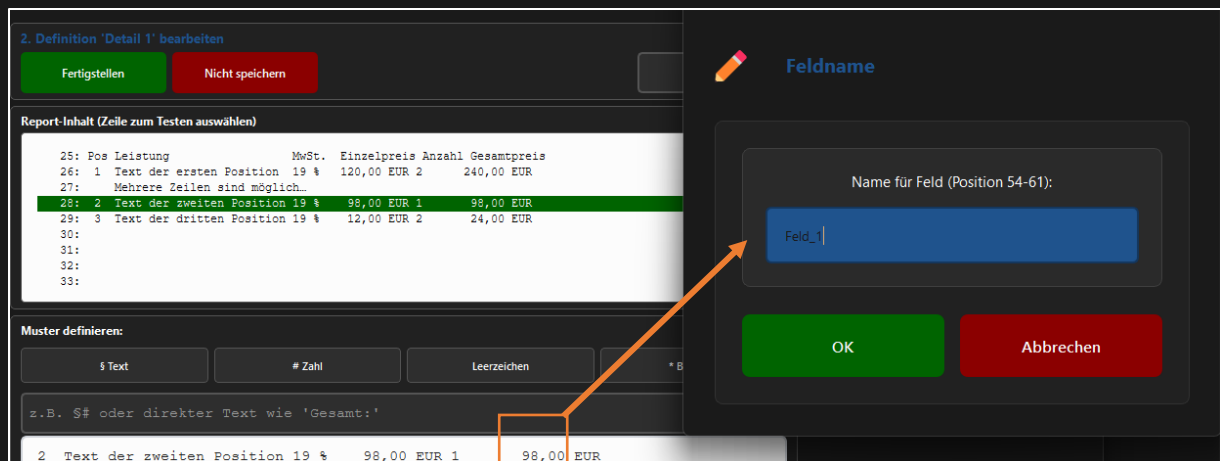
2 Text der zweiten Position 19 % 98,00 EUR 1 98,00 EUR

Pattern-Entwicklung und Validierung ②

1. **Pattern-Analyse:** Untersuchen Sie die ausgewählte Zeile auf charakteristische Merkmale. Achten Sie auf feste Textbestandteile, numerische Bereiche oder strukturelle Elemente.
2. **Selector-Verwendung:** Nutzen Sie die Selector-Buttons für systematische Pattern-Erstellung oder geben Sie Literaltext direkt ein. Kombinieren Sie beide Ansätze für optimale Präzision.
3. **Live-Validierung:** Beobachten Sie die farbliche Rückmeldung im Pattern-Display. Grün signalisiert erfolgreiche Matches, Rot deutet auf Pattern-Probleme hin.
4. **Match-Verifikation:** Prüfen Sie die Match-Statistik in der Statusleiste. Zu wenige oder zu viele Matches deuten auf Pattern-Optimierungsbedarf hin.

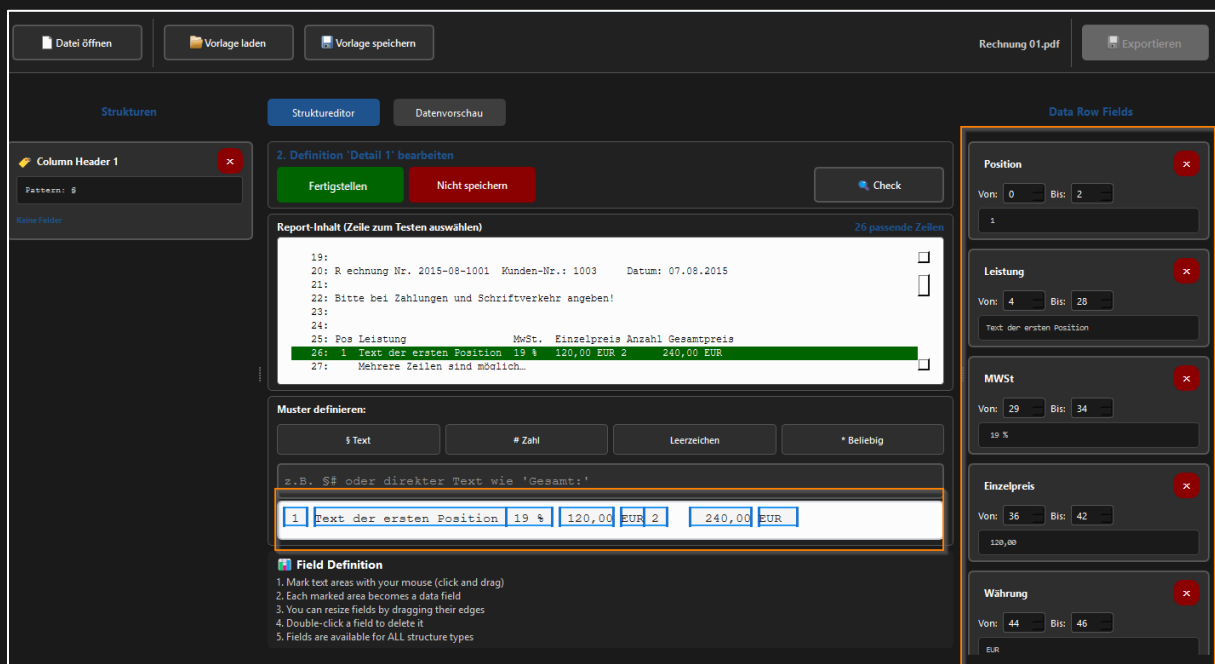
Feldkonfiguration und -optimierung ② & ④

1. **Feldauswahl** (nur bei Datenzeilen): Markieren Sie Datenbereiche durch Mausauswahl im Line-Display. Jeder markierte Bereich wird zu einem extrahierbaren Datenfeld.
2. **Feldverfeinerung**: Justieren Sie Feldgrenzen durch Ziehen an den Rändern. Achten Sie auf präzise Abgrenzung zur Vermeidung von Datenüberschneidungen.



3. **Semantische Benennung**: Vergeben Sie aussagekräftige Feldnamen entsprechend der fachlichen Bedeutung. Nutzen Sie konsistente Namenskonventionen für bessere Nachvollziehbarkeit.

Datenvalidierung: Kontrollieren Sie die Beispielwerte in den Feldkarten auf Plausibilität und Vollständigkeit.



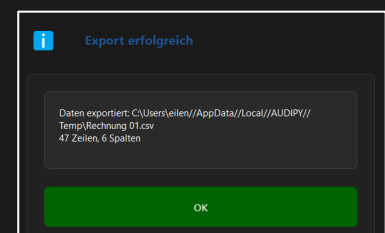
Definition-Abschluss und Testing 5

1. **Vollständigkeitsprüfung:** Verwenden Sie "🔍 Prüfen" zur Validierung der aktuellen Definition. Beheben Sie gemeldete Probleme vor dem Abschluss.
2. **Definition speichern:** Schließen Sie die Bearbeitung über "Abschließen" ab. Die Definition wird in die Strukturübersicht übernommen.
3. **Iterative Verfeinerung:** Bearbeiten Sie bestehende Definitionen durch Klick auf die entsprechenden Karten. Optimieren Sie Pattern und Felder basierend auf Tests mit verschiedenen Dokumentenzeilen.
4. **Template-Erstellung:** Speichern Sie bewährte Definitionssätze als Templates für wiederkehrende Dokumententypen.

Datenextraktion und Export 6

1. **Vorschau-Kontrolle:** Wechseln Sie zum "Datenvorschau"-Tab zur Verifikation der Extraktionsergebnisse. Prüfen Sie Datenqualität und -vollständigkeit.
2. **Export-Konfiguration:** Wählen Sie zwischen Einzeldatei- oder Ordner-Export entsprechend Ihrem Verarbeitungsmodus.
3. **Ausgabeformat:** Definieren Sie CSV- oder Excel-Format basierend auf nachgelagerten Verarbeitungsanforderungen.
4. **Qualitätssicherung:** Validieren Sie das Export-Ergebnis auf Vollständigkeit, korrekte Zeichensatzdarstellung und fachliche Plausibilität.

Nach erfolgreichem Export erscheint ein Bestätigungsfenster, und der Data-Wizard öffnet sich, um die weitere Formatierung und die Umwandlung in eine Parquet-Datei durchzuführen (vgl. 3.2.4).



3.2.8.2.4 Erweiterte Strategien und Optimierung

Möchte man den Import von PDF / TXT Dateien optimieren sollte mal folgende Details nicht außer Acht lassen:

Pattern-Optimierung und Erkennungsqualität

- Die **Pattern-Entwicklung** sollte iterativ erfolgen: Beginnen Sie mit einfachen, charakteristischen Merkmalen und verfeinern Sie schrittweise. Nutzen Sie Literaltext für eindeutige Kennzeichnungen wie "Summe:" oder "Total:" und ergänzen Sie Platzhalter für variable Bereiche. Die Kombination beider Ansätze maximiert sowohl Präzision als auch Flexibilität.

- **Robustes Pattern-Design** berücksichtigt typische Variationen: Führende Leerzeichen, unterschiedliche Zahlenformate oder alternative Schreibweisen. Testen Sie Ihre Pattern mit verschiedenen Dokumentenvarianten derselben Quelle zur Sicherstellung der Stabilität. Dokumentieren Sie erkannte Variationsmuster für die Template-Dokumentation.
- **Fehlerdiagnose** bei unerwarteten Match-Ergebnissen: Zu viele Matches deuten auf zu generische Pattern hin, zu wenige auf zu restriktive Definitionen. Nutzen Sie die Live-Vorschau zur schrittweisen Pattern-Verfeinerung und validieren Sie regelmäßig mit repräsentativen Dokumentenabschnitten.

Feldauswahl und Datenqualität

- **Präzise Feldabgrenzung** vermeidet Datenüberschneidungen und -verluste: Orientieren Sie sich an natürlichen Spaltengrenzen und berücksichtigen Sie Padding-Charakteristika der Quelldokumente. Testen Sie Feldgrenzen mit Extremwerten (längste/kürzeste Inhalte) zur Sicherstellung der Robustheit.
- **Semantische Feldbenennungen** erhöhen die Nachvollziehbarkeit: Verwenden Sie fachliche Terminologie statt technischer Bezeichnungen. Konsistente Namenskonventionen (z.B. "Betrag_Brutto", "Betrag_Netto") erleichtern die spätere Datenanalyse und Automatisierung.
- **Datentyp-Berücksichtigung** bei der Feldkonfiguration: Numerische Felder sollten Vorzeichen, Dezimaltrennzeichen und Währungssymbole angemessen behandeln. Datumsfelder erfordern einheitliche Formatierungen für die nachgelagerte Verarbeitung.

Template-Management und Wiederverwendung

- **Systematische Template-Erstellung** reduziert langfristig den Aufwand: Entwickeln Sie Templates für wiederkehrende Dokumententypen und -quellen. Dokumentieren Sie Template-Anwendungsbereiche und bekannte Limitationen für andere Anwender.
- **Versionierung von Templates** gewährleistet Nachvollziehbarkeit: Speichern Sie bewährte Template-Versionen mit aussagekräftigen Namen und Datumsangaben. Dies ermöglicht Rollbacks bei unerwarteten Formatänderungen der Quelldokumente.
- **Template-Anpassungsstrategien** für evolvierende Dokumentenformate: Entwickeln Sie robuste Basis-Templates mit nachgelagerten Anpassungsebenen. Dies minimiert den Wartungsaufwand bei geringfügigen Formatänderungen.

Ordner-Verarbeitung und Skalierung

- **Beispieldatei-Auswahl** beeinflusst maßgeblich die Template-Qualität: Wählen Sie repräsentative Dateien mit typischen Formatierungen und vollständigen Datensätzen. Vermeiden Sie Ausreißer oder unvollständige Dokumente für die initiale Template-Entwicklung.

3.2.9 Excel Import

Der **Excel Wizard** unterstützt den Import und die strukturierte Vorbereitung von Excel-Dateien zur weiteren Verarbeitung. Er erkennt automatisch Tabellenblätter, analysiert Inhalte auf potenzielle Probleme (z.B. fehlerhafte Header, verbundene Zellen) und bietet eine Vorschau sowie manuelle Korrekturmöglichkeiten.

The screenshot displays the Excel Wizard application interface. At the top, a status bar shows '1 Datei • 1 Sheet • 21.576 Zeilen' and a 'Smart Merge' button. Below this, the interface is divided into two main sections. The left section, labeled 'Excel Sheets', shows a list of files with a warning icon and the text 'Übung_Verkaufsdaten.xlsx - Verkaufsdaten' and '21.576 Zeilen • 8 Spalten • Header erkannt • 1 Probleme'. The right section, labeled 'Datenvorschau', shows a preview of the data with a table. The table has columns: Monat, Laden, Stadt, Land, Manager, Kategorien, and Verkauf. The data is organized by date (2023-01-01) and includes details about the store (Laden), city (Stadt), country (Land), manager (Manager), category (Kategorien), and sales amount (Verkauf). A warning icon at the bottom right indicates 'Verdacht auf verbundene Zellen'.

	Monat	Laden	Stadt	Land	Manager	Kategorien	Verkauf	
2	2023-01-01 00:00:00	Computech	Köln	ABD	DE	Anna Schmidt	Mobiletelefone	4.25
3	2023-01-01 00:00:00	Computech	Düsseldorf	ABD	DE	Ben Koch	PCs	1525.002
4	2023-01-01 00:00:00	Computech	München	ABD	DE	Emilia Schäfer	Laptops	3366.0935
5	2023-01-01 00:00:00	MicroWorld	Stuttgart	ABD	DE	Lukas Wagner	Zubehör	253.572
6	2023-01-01 00:00:00	Computech	Frankfurt	ABD	DE	Hannah Schwarz	Drucker	155.7625
7	2023-01-01 00:00:00	Computech	Leipzig	ABD	DE	Paul Fischer	PCs	1439.7215
8	2023-01-01 00:00:00	MicroWorld	München	ABD	DE	Leon Klein	Kameras	25.5
9	2023-01-01 00:00:00	MicroWorld	Dortmund	ABD	DE	Paul Fischer	Mobiletelefone	10.2
10	2023-01-01 00:00:00	Computech	Stuttgart	ABD	DE	Leon Klein	Drucker	922.318
11	2023-01-01 00:00:00	MicroWorld	Köln	ABD	DE	Laura Weber	Kameras	8.5
12	2023-01-01 00:00:00	MicroWorld	Essen	ABD	DE	Anna Schmidt	Kameras	147.5005
13	2023-01-01 00:00:00	MicroWorld	Hamburg	ABD	DE	Ben Koch	Kameras	2.55
14	2023-01-01 00:00:00	MicroWorld	München	ABD	DE	Mia Becker	Kameras	170.833

Übersicht & Smart Merge 1

Dieser Bereich zeigt auf einen Blick:

- Anzahl der geladenen Dateien und Sheets
- Gesamtanzahl der enthaltenen Datenzeilen

Beispiel:

„1 Datei • 1 Sheet • 21.576 Zeilen“

So erhält der Benutzer sofort eine Übersicht über das Datenvolumen.

- **Smart Merge:** Intelligentes Zusammenführen mehrerer Tabellenblätter

Excel Sheets 2

Hier werden alle erkannten Tabellenblätter (Sheets) der gewählten Excel-Datei aufgelistet, inklusive:

- Dateiname und Blattname
- Anzahl der Zeilen und Spalten
- Hinweise zur Header-Erkennung
- Warnmeldungen (z. B. „Verdacht auf verbundene Zellen“)
- Buttons zur Auswahlsteuerung:
 - **Alle:** alle Sheets markieren
 - **Keine:** keine Auswahl treffen
 - **Saubere:** nur fehlerfreie Tabellen auswählen

Der Benutzer kann pro Tabellenblatt einzeln festlegen, ob es importiert werden soll.

Datenvorschau 3

In diesem Bereich wird eine Vorschau des aktuell ausgewählten Tabellenblatts angezeigt. Ein zentrales Element ist die Einstellung der Header-Zeile, die steuert, ab welcher Zeile die Spaltenüberschriften übernommen werden sollen:

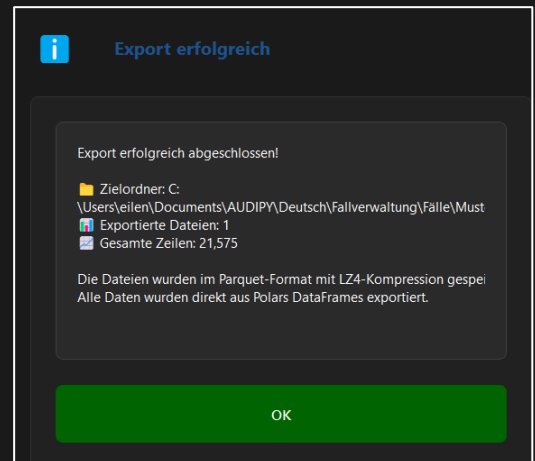
- Header-Zeile = 0
 - Die erste Zeile der Excel-Tabelle wird als Spaltenüberschrift interpretiert.
 - Dies ist die Standardoption, wenn die Excel-Datei korrekt formatiert ist.
- Header-Zeile = 1
 - Die zweite Zeile enthält die Spaltenüberschriften.
 - Dies ist nützlich, wenn die erste Zeile z. B. Überschriften, Titelzeilen oder allgemeine Informationen enthält, aber keine echten Spaltennamen.
- Header-Zeile = -1
 - Es wird kein Header verwendet. Alle Zeilen, inklusive der ersten, gelten als reguläre Daten.
 - Die Spalten erhalten dann automatisch generierte Namen wie Spalte 1, Spalte 2 usw.
- Auto
 - Das System versucht automatisch zu erkennen, welche Zeile am wahrscheinlichsten die Spaltennamen enthält (basierend auf Heuristiken und Datenstruktur).
 - Anzeige der tatsächlichen Dateninhalte – inklusive der automatisch erkannten Datentypen.
- Visuelle Prüfung auf strukturelle Auffälligkeiten wie:

- Falsch interpretierte Datumsfelder
- Doppelte Header-Zeilen
- Verdacht auf verbundene Zellen (wird unten rechts als Warnung angezeigt)

⚠ Verdacht auf verbundene Zellen

Nach Auswahl der gewünschten Einstellungen (z. B. Header-Zeile, Tabellenblatt, Datenvorschau) kann die Datei über die Schaltfläche „**1** **einzel** **exportieren**“ oben rechts als **Parquet-Datei** übernommen werden.

Nach erfolgreichem Import wird die erzeugte **Parquet-Datei im angegebenen Zielordner gespeichert**. Anschließend erscheint ein **Bestätigungsdialog**, der den erfolgreichen Abschluss des Exports bestätigt und Informationen zur erzeugten Datei enthält.



3.2.10 XML Import

Die Funktion **XML Import** (XML-Wizard) dient dazu, eine XML-Datei zu laden, deren Struktur zu analysieren und die Inhalte in einer **tabellarischen, übersichtlichen Form** darzustellen. Nach dem Import werden die XML-Daten automatisch in ihre Bestandteile zerlegt und visuell aufbereitet.

The screenshot shows the XML-Wizard interface. At the top, it says 'XML-Wizard' and '85 Elemente verarbeitet'. There are buttons for 'Exportieren' and 'Schließen'. Below this, a section labeled 'Test.xml' shows '85 Elemente • 4.4 KB • Bereit für Export'. The main area is divided into 'Filter & Optionen' on the left and a table view on the right. The table has columns: @id, Attributes, Children, Element, Level, Path, and Text. It lists various XML elements like catalog, book, author, title, genre, price, publish_date, and description, along with their attributes and content.

	@id	Attributes	Children	Element	Level	Path	Text
1			12	catalog	0	catalog	
2	bk101	id=bk101	6	book	1	catalog/book	
3			0	author	2	catalog/book/author	Gambardella, Matthew
4			0	title	2	catalog/book/title	XML Developer's Guide
5			0	genre	2	catalog/book/genre	Computer
6			0	price	2	catalog/book/price	44.95
7			0	publish_date	2	catalog/book/publish_date	2000-10-01
8			0	description	2	catalog/book/description	An in-depth look at creating applications ...
9	bk102	id=bk102	6	book	1	catalog/book	
10			0	author	2	catalog/book/author	Ralls, Kim
11			0	title	2	catalog/book/title	Midnight Rain
12			0	genre	2	catalog/book/genre	Fantasy
13			0	price	2	catalog/book/price	5.95
14			0	publish_date	2	catalog/book/publish_date	2000-12-16
15			0	description	2	catalog/book/description	A former architect battles corporate zombies, ...
16	bk103	id=bk103	6	book	1	catalog/book	
17			0	author	2	catalog/book/author	Corets, Eva
18			0	title	2	catalog/book/title	Maeyve Ascendant

Importierte Datei ①

Hier wird die aktuell geladene XML-Datei mit zusätzlichen Parametern angezeigt.

Filter & Optionen ②

Dieser Bereich bietet umfassende Möglichkeiten zur Anpassung der Ansicht und zum gezielten Filtern von XML-Inhalten.

- **Suchen:** Ermöglicht die Suche nach bestimmten XML-Elementnamen oder Textinhalten.
- **Ebenen-Filter:** Filtert die Anzeige nach bestimmten Hierarchieebenen (z. B. nur Elemente der Ebene 1 oder 2). Es sind bis zu 5 Ebenen möglich.
- **Anzeigeoptionen:**
 - **Leere Elemente anzeigen:** Blendet auch XML-Knoten ohne Inhalt ein.

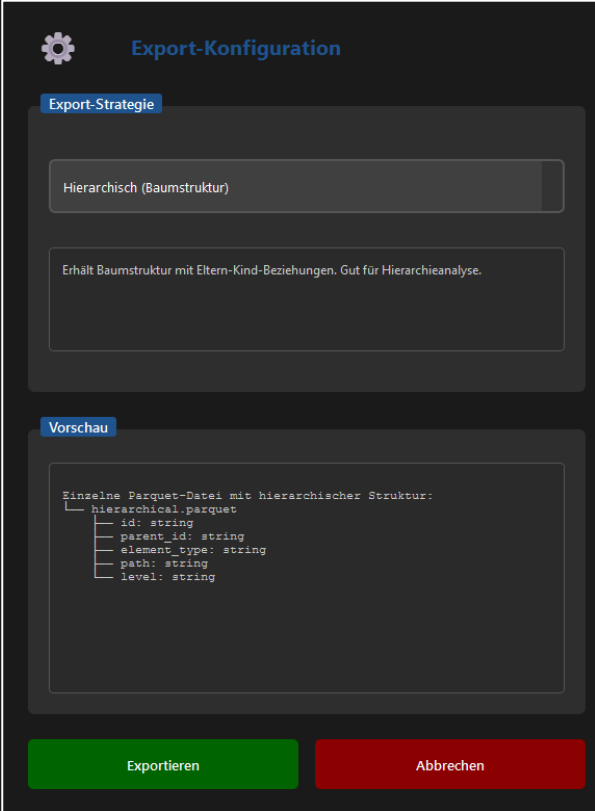
- **Attribute als Spalten:** Zeigt XML-Attribute als separate Spalten in der Tabelle (standardmäßig aktiviert).

Datenansicht 3

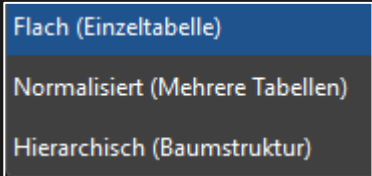
Hier wird der strukturierte Inhalt der XML-Datei tabellarisch dargestellt. Die Ansicht ist je nach gewähltem Modus (z. B. „Flache Ansicht“) anpassbar.

- **Flache Ansicht** – alle Elemente in Tabellenstruktur
- **Hierarchie** – Darstellung in Codestruktur
- **Elemente** – Gruppiert nach Tag-Namen
- **Attribute** – Fokus auf Attributwerte

Nach Bestätigung des Exportbuttons gelangt man in ein weiteres Dialogfenster:



Bei der Verarbeitung oder dem Export von XML-Daten kann zwischen drei verschiedenen Darstellungsformen gewählt werden. Jede dieser Optionen bestimmt, wie die XML-Inhalte strukturiert angezeigt und ggf. exportiert werden:



- **Flach (Einzeltabelle)**

In dieser Ansicht werden sämtliche XML-Elemente in einer einzigen, durchgehenden Tabelle dargestellt – unabhängig davon, wie tief sie im XML-Baum verschachtelt sind.

- Die Daten erscheinen zeilenweise, jedes Element steht für sich.

- Ideal für einfache Analysen, z. B. in Excel oder CSV-Dateien.
- Die ursprüngliche hierarchische Struktur geht dabei verloren, was die Nachvollziehbarkeit der Datenbeziehungen erschweren kann.
- Besonders nützlich, wenn nur Inhalte und Werte wichtig sind, nicht aber ihre Beziehung zueinander.

- **Normalisiert (Mehrere Tabellen)**

Hier werden die XML-Daten logisch aufgeteilt und in mehrere Tabellen zerlegt – typischerweise eine Tabelle für Entitäten, Attribute und Beziehungen.

- Diese Struktur orientiert sich an der relationalen Datenbanklogik.
- Redundante Informationen werden reduziert, verwandte Elemente lassen sich über Schlüsselbeziehungen verbinden.
- Eignet sich besonders gut für komplexe Auswertungen, Datenimporte in relationale Datenbanken oder strukturierte Reports.
- Die Datenstruktur bleibt erhalten, ist aber auf mehrere Ebenen verteilt.

- **Hierarchisch (Baumstruktur)**

Diese Ansicht bildet die XML-Daten so ab, wie sie ursprünglich in der Datei vorliegen: als verschachtelte Baumstruktur mit Eltern-Kind-Beziehungen.

- Jedes Element ist einem Pfad und einer Ebene zugeordnet, was die XML-Logik sehr gut sichtbar macht.
- Optimal geeignet für eine inhaltliche und strukturelle Analyse der Datei.
- Weniger geeignet für einfache tabellarische Exporte, da verschachtelte Strukturen nicht ohne Weiteres in flache Tabellen umgewandelt werden können.
- Besonders hilfreich, wenn die Beziehungen zwischen Elementen eine wichtige Rolle spielen (z. B. bei Konfigurationsdateien oder komplexen Datenmodellen).

Anschließend kann der Import gestartet werden. Die Datei wird im hinterlegten Speicherort als Paquet-Datei abgelegt und kann dann ganz normal geöffnet werden.

3.2.11 GoBD Import

Mit dieser Funktion können Sie Daten aus **GoBD-konformen Dateien** einfach und schnell importieren. Das Tool bietet zwei wesentliche Optionen für den Import sowie die Möglichkeit, mehrere Dateien zu kombinieren.

The screenshot shows the 'GoBD Import' window. It is divided into several sections:

- Importquelle wählen** (Import source selection): Contains a text field labeled 'Ordner' (Folder) and an 'Auswählen ...' (Select ...) button. This area is marked with a red circle '1'.
- Index-Dateien** (Index files): A large empty rectangular area for displaying index files. It has 'Alle auswählen' (Select all) and 'Alle abwählen' (Deselect all) buttons at the top. This area is marked with a red circle '2'.
- Tabellen-Details** (Table details): A large empty rectangular area for displaying details of selected tables. It has 'Alle auswählen' (Select all) and 'Alle abwählen' (Deselect all) buttons at the top. This area is marked with a red circle '3'.
- Optionen** (Options): Contains three checkboxes: 'Dateien kombinieren' (Combine files), 'Nur 1. Ebene kombinieren' (Combine only 1st level), and 'Daten Stream' (Data stream). This area is marked with a red circle '4'.
- Verarbeitung** (Processing): Contains a text field with the value '-- Keine Aufbereitung --' (No processing). This area is marked with a red circle '5'.

At the bottom right, there are two buttons: 'Daten importieren' (Import data) in green and 'Schließen' (Close) in red.

Funktionsweise

1. Auswahl der Importquelle ❶:

- **Ordner:** Sie können einen gesamten Ordner als Quelle auswählen, der die relevanten GoBD-Daten enthält.
- **Index.xml:** Alternativ können Sie eine spezifische index.xml-Datei auswählen, um die Daten zu importieren.
- **SAP DART Dateien:** Der Import von **SAP DART-Dateien** ermöglicht das Einlesen archivierter SAP-Datenstrukturen zur weiteren Analyse und Verarbeitung.

2. Index Dateien ❷:

- Nach der Auswahl der Quelle werden die zu importierenden Index.xml angezeigt.

3. Tabellen-Details ❸:

Ib diesem Bereich können die Details oder zusätzliche Informationen zu den ausgewählten Daten eingesehen werden.

4. Optionen ❹:

- **Dateien kombinieren:** Die Funktion „**Dateien kombinieren**“ dient dazu, mehrere gleichartige Dateien aus unterschiedlichen **Index.xml**-Dateien in einer einzigen Datei zusammenzufassen.

Dies ist besonders dann nützlich, wenn Sie mehrere GOBD-Exporte (z. B. für verschiedene Zeiträume) erhalten haben, diese aber für die weitere Verarbeitung in einem einzigen Satz benötigen.

- **Nur 1. Ebene kombinieren:** Mit der Option **Nur 1. Ebene kombinieren** werden beim Erstellen einer Index.xml nur die Dateien auf der obersten Ebene jedes Ordners berücksichtigt. Die Ordner selbst bleiben getrennt.

1. **Beispiel:**

Ordner „Daten A“ mit 2021.xml und 2022.xml
Ordner „Daten_B“ mit 2021.xml und 2022.xml

Ergebnis:

In „Daten A“ werden die Index.xml aus 2021.xml und 2022.xml kombiniert importiert

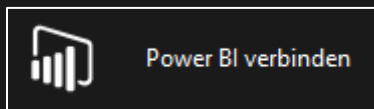
In „Daten B“ werden die Index.xml aus 2021.xml und 2022.xml kombiniert importiert.

Nur Dateien auf der ersten Ebene der Ordner werden kombiniert.
So müssen keine einzelnen Dateien eingelesen werden.

- **Daten Stream:** Mit der Option **können auch größere Datenmengen** z.B. von SAP importiert werden. Die Funktionsweise ist wie beim Data-Wizard mit CSV (s. Kapitel **3.2.4**)

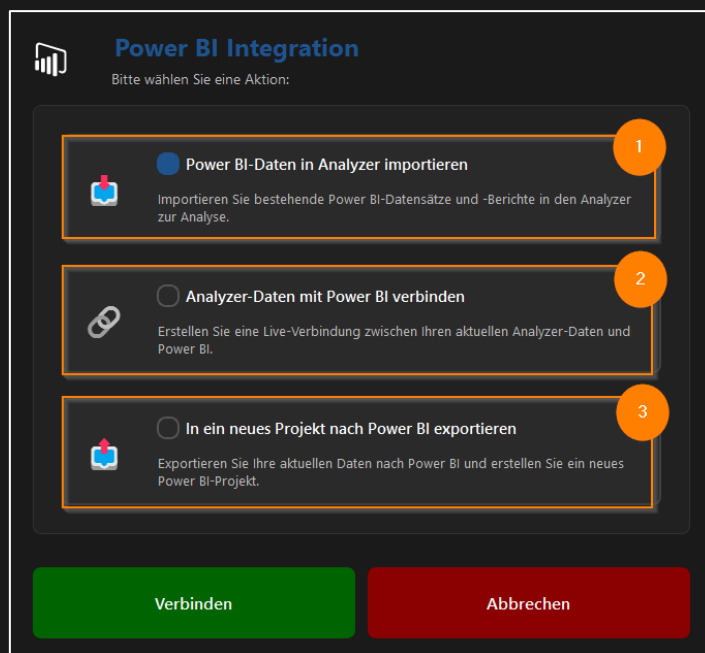
5. Verarbeitung 5: Hier können eigene Audit-Steps ausgewählt werden.

3.3 Power BI verbinden



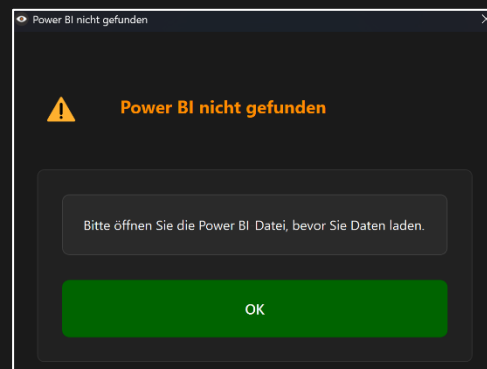
Power BI verbinden

Mit der Funktion **Power BI verbinden** können Sie Daten zwischen dem Analyzer und Microsoft Power BI austauschen und so Ihre Analysen nahtlos miteinander verknüpfen. Beim Aufrufen dieser Funktion stehen Ihnen drei Hauptaktionen zur Auswahl:



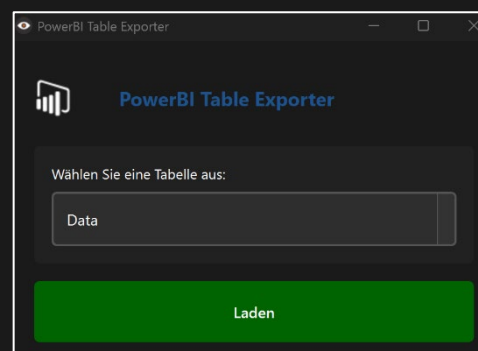
1. Power BI-Daten in Analyzer importieren ①

- Wählen Sie diese Option, um bereits in Power BI vorhandene Daten in den Analyzer zu laden. So können Sie die Daten direkt in der Anwendung weiterbearbeiten oder ergänzen, ohne erneut Daten exportieren zu müssen. **Hierfür muss die gewünschte Power BI Datei mit Daten geöffnet sein, ansonsten erhalten Sie folgenden Fehlerhinweis:**



Sie können auch auswählen welche Datentabelle exportiert werden soll, wenn Sie mehr als eine haben:

Anschließend lädt der Analyzer die Daten von Power BI ein und Sie erhalten Ihre Daten in einer klassischen Tabellenstruktur.

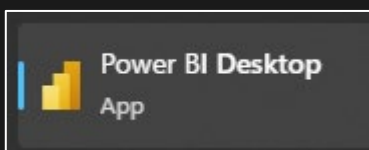


2. Analyzer-Daten mit Power BI verbinden ②

- Diese Option stellt eine direkte Verbindung zwischen Ihren im Analyzer vorliegenden Daten und Power BI her (z.B. wenn man bereits eine Power BI Datei hat und neue Daten hinzufügen möchte). Auch hier ist wichtig, dass eine Power BI Datei geöffnet sein muss, damit die Daten aus dem Analyzer rüber transportiert werden können.

Folgende Schritte helfen Ihnen zur Verwendung der AUDIPY Daten Verbindung mit Power Query:

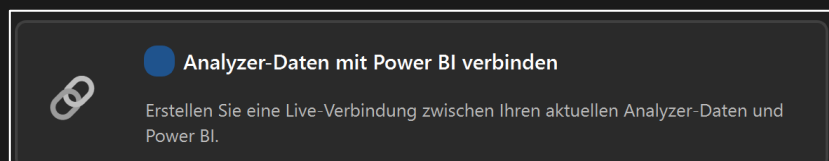
1. Schritt: Power BI öffnen



Für die Verbindung sollten Sie erst einmal **Power BI Desktop** (kostenlose Version) auf Ihrem Rechner installiert haben und dann öffnen.

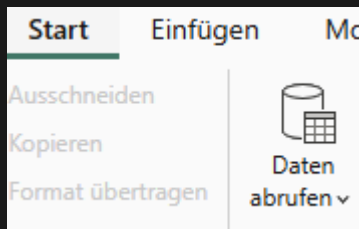
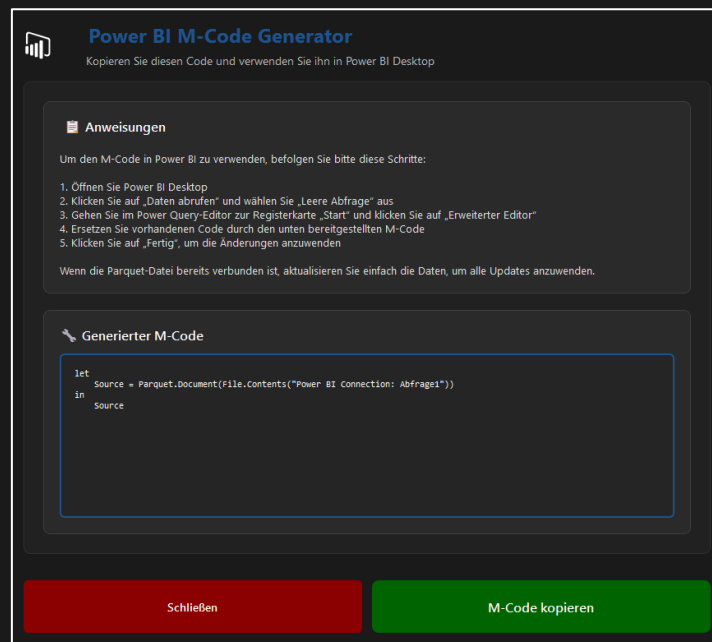
2. Schritt: Verbindung Power BI

Über Start – Power BI Verbinden im nachfolgenden Fenster den zweiten Punkt „**Analyzer Daten mit Power BI verbinden**“ anklicken (blau hinterlegtes Kästchen)



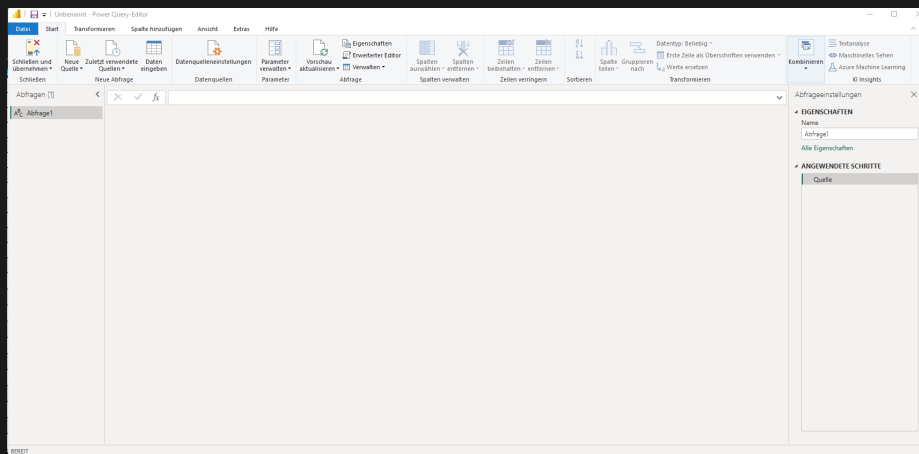
3. Schritt: Anweisungen M-Code befolgen

Es öffnet sich nun ein weitere Programmfenster mit **5 weiteren Schritten**, mit dem Ziel den **generierten M-Code** in Power BI zu integrieren.

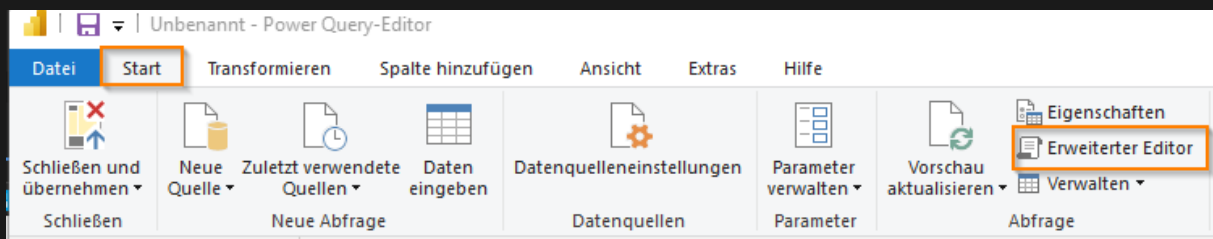


Nach Öffnen von Power BI klicken Sie in Power BI unter dem Reiter Start auf **Daten abrufen** und wählen Sie **leere Abfrage**.

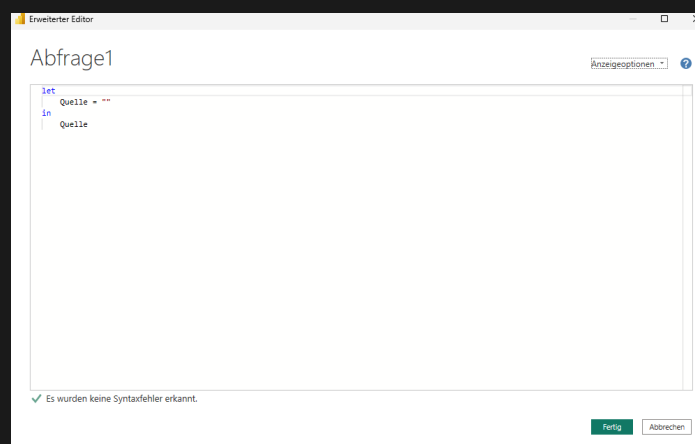
Hier nach öffnet sich der leere **Power Query**:



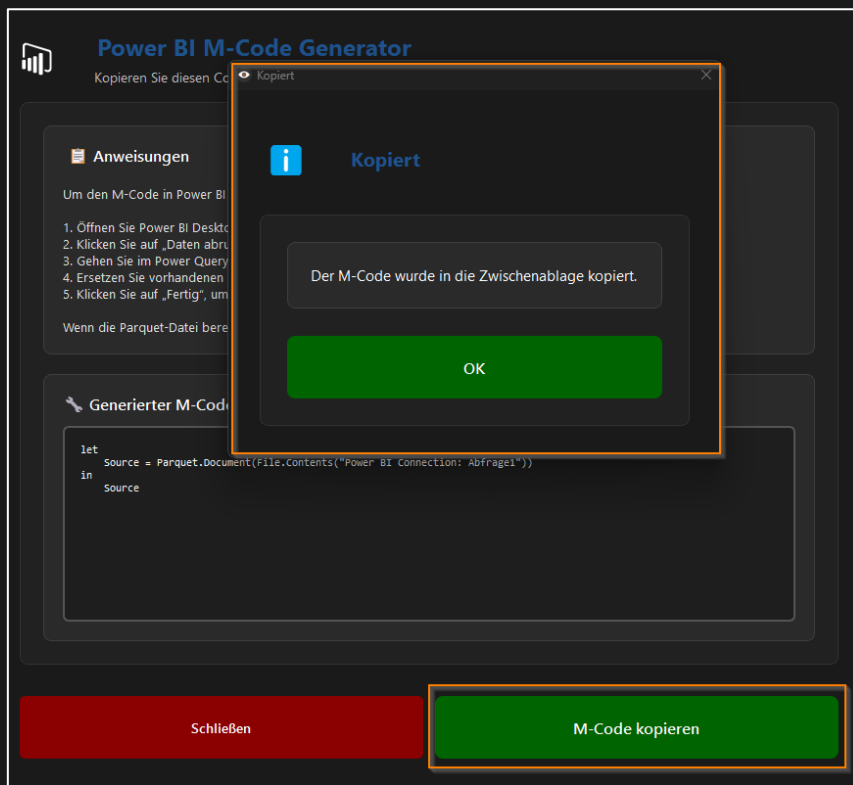
Wählen Sie nun unter Start **Erweiterter Editor**:



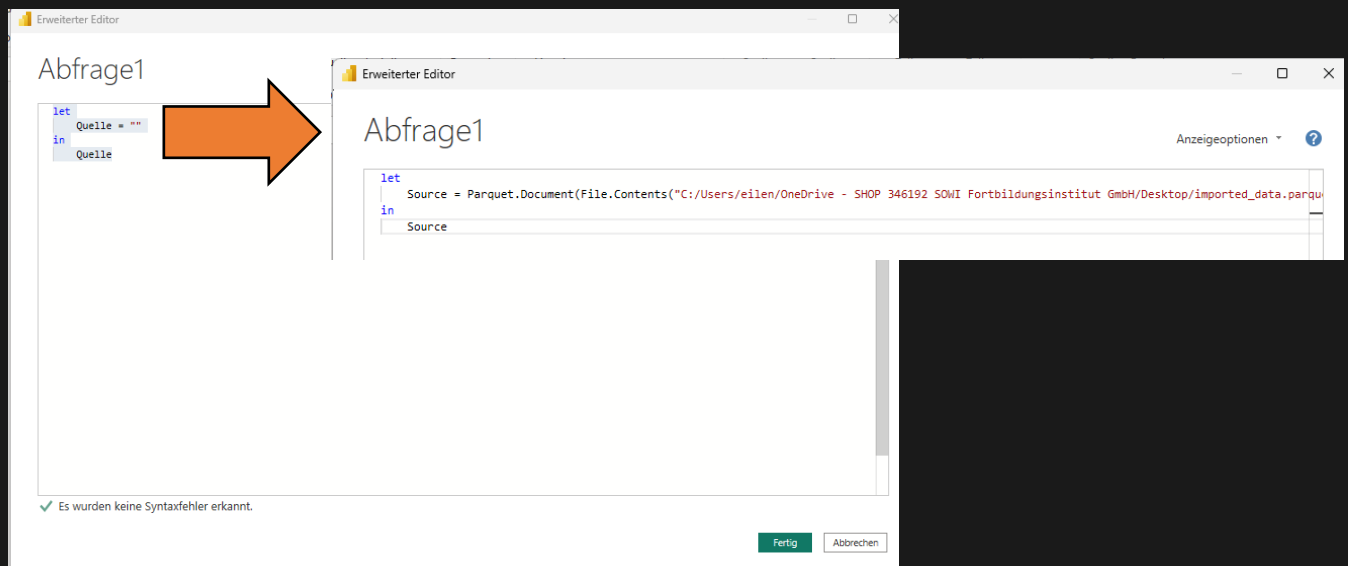
Es öffnet sich nun der Editor:



Klicken Sie nun im AUDIPY auf **M-Code kopieren** (Hinweis: Sofern Sie das Fenster geschlossen haben, öffnen Sie einfach wieder über Start – Power BI Verbinden - Analyzer-Daten mit Power BI verbinden die M-Code Details). Nach Betätigung des grünen „M-Code kopieren“ Buttons erhalten Sie ein Bestätigungsfenster, dass der der Code in die Zwischenablage kopiert wurde.

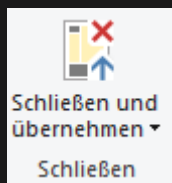


Nun gehen Sie in Power BI in den geöffneten Editor und ersetzen den angezeigten Code komplett durch den in die Zwischenablage kopierten Code und bestätigen dies mit „**Fertig**“.



Anschließend ist die vorherige leere Power Query Datei befüllt mit Daten:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment
1	CA-2020-152156	08.11.2020 00:00:00	null	Second Class	CG-12520	Claire Gute	Consumer
2	CA-2020-152156	08.11.2020 00:00:00	null	Second Class	CG-12520	Claire Gute	Consumer
3	CA-2020-138688	12.06.2020 00:00:00	null	Second Class	DV-13045	Darrin Van Huff	Corporate
4	US-2019-108966	11.10.2019 00:00:00	null	Standard Class	SO-20335	Sean O'Donnell	Consumer
5	US-2019-108966	11.10.2019 00:00:00	null	Standard Class	SO-20335	Sean O'Donnell	Consumer
6	CA-2018-115812	09.06.2018 00:00:00	null	Standard Class	BH-11710	Brosina Hoffman	Consumer
7	CA-2018-115812	09.06.2018 00:00:00	null	Standard Class	BH-11710	Brosina Hoffman	Consumer
8	CA-2018-115812	09.06.2018 00:00:00	null	Standard Class	BH-11710	Brosina Hoffman	Consumer
9	CA-2018-115812	09.06.2018 00:00:00	null	Standard Class	BH-11710	Brosina Hoffman	Consumer
10	CA-2018-115812	09.06.2018 00:00:00	null	Standard Class	BH-11710	Brosina Hoffman	Consumer
11	CA-2018-115812	09.06.2018 00:00:00	null	Standard Class	BH-11710	Brosina Hoffman	Consumer
12	CA-2018-115812	09.06.2018 00:00:00	null	Standard Class	BH-11710	Brosina Hoffman	Consumer
13	CA-2021-144412	15.04.2021 00:00:00	null	Standard Class	AA-10480	Andrew Allen	Consumer
14	CA-2020-161389	05.12.2020 00:00:00	null	Standard Class	IM-15070	Irene Maddox	Consumer
15	US-2019-118983	22.11.2019 00:00:00	null	Standard Class	HP-14815	Harold Pawlan	Home Office
16	US-2019-118983	22.11.2019 00:00:00	null	Standard Class	HP-14815	Harold Pawlan	Home Office
17	CA-2018-105893	11.11.2018 00:00:00	null	Standard Class	PK-19075	Pete Kriz	Consumer
18	CA-2018-167164	13.05.2018 00:00:00	null	Second Class	AG-10270	Alejandro Grove	Consumer
19	CA-2018-143336	27.08.2018 00:00:00	null	Second Class	ZD-21925	Zuschuss Donatelli	Consumer
20	CA-2018-143336	27.08.2018 00:00:00	null	Second Class	ZD-21925	Zuschuss Donatelli	Consumer
21	CA-2018-143336	27.08.2018 00:00:00	null	Second Class	ZD-21925	Zuschuss Donatelli	Consumer
22	CA-2020-137330	09.12.2020 00:00:00	null	Standard Class	KB-16585	Ken Black	Corporate
23	CA-2020-137330	09.12.2020 00:00:00	null	Standard Class	KB-16585	Ken Black	Corporate
24	US-2021-156999	16.07.2021 00:00:00	null	Second Class	SP-20065	Sandra Phangan	Consumer
25	CA-2019-106320	25.09.2019 00:00:00	null	Standard Class	EB-13670	Emily Burns	Consumer
26	CA-2020-121755	16.01.2020 00:00:00	null	Second Class	EH-13945	Eric Hoffmann	Consumer
27	CA-2020-121755	16.01.2020 00:00:00	null	Second Class	EH-13945	Eric Hoffmann	Consumer
28	US-2019-150630	17.09.2019 00:00:00	null	Standard Class	TB-21520	Tracy Blumstein	Consumer
29							



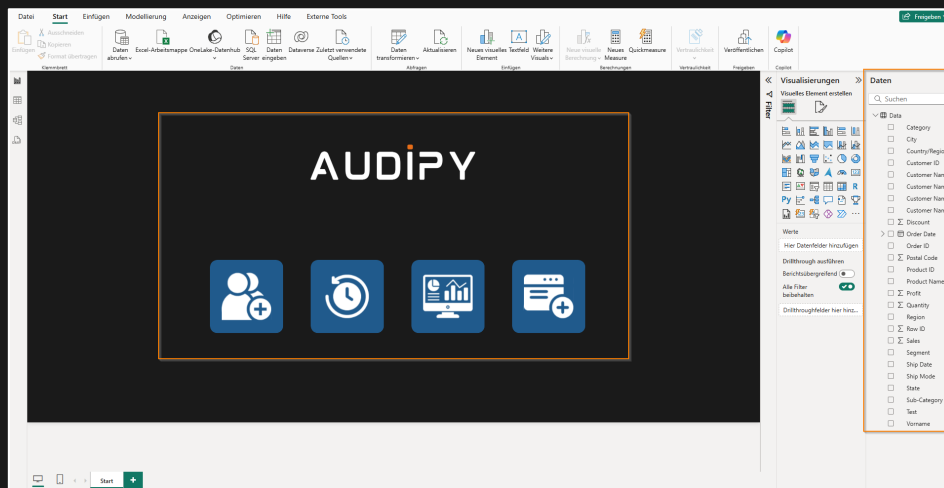
Über **Schließen und übernehmen** (oben links) können die Daten nun nach Power BI transferiert werden.

Ihre AUDIPY Daten wurden nun erfolgreich mit Power BI verbunden und können individuell visualisiert werden.

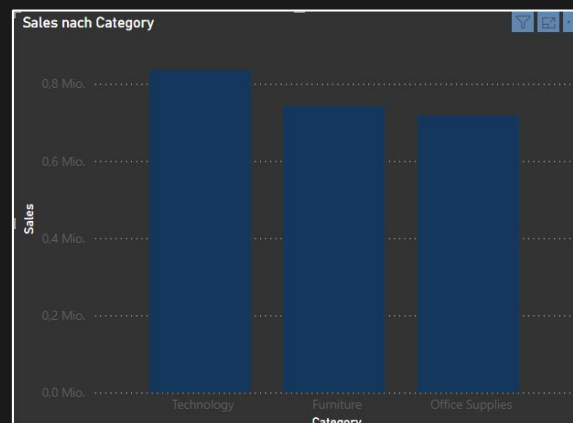
Hinweis: Im Gegensatz zum Schritt „in ein **neues Projekt** nach Power BI exportieren“ wird bei dieser Variante kein Musterformat vom AUDIPY hinterlegt. Lediglich die Daten wurden miteinander verbunden. So können bestehende Power BI Dashboards mit den Daten vom AUDIPY ergänzt werden.

3. In ein neues Projekt nach Power BI exportieren ③

- Möchten Sie Daten oder Ergebnisse aus dem Analyzer für weitere Auswertungen in Power BI bereitstellen, können Sie sie über diese Option in ein **neues Power-BI-Projekt** übertragen. Hierfür muss vorher keine Power BI Datei geöffnet sein. Dies passiert nach einer Speicherung der Datei automatisch. Nachdem Power BI vollständig geladen hat, erhalten Sie ein Titelbild mit „AUDIPY“ (Hinweis: Möchten Sie später nochmals weitere Daten hinzufügen, kann dies über die zweite Variante „**Analyzer-Daten mit Power BI verbinden**“ erfolgen).

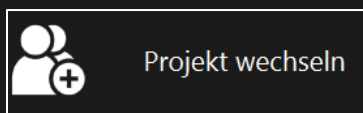


Die **exportierten Daten** vom AUDIPY sind nun unter Daten hinterlegt und können nun individuell verwendet werden. Hinweis: Beim Hinzufügen eines Visual (z.B. Balkendiagramm) wird die Formatierung bereits im entsprechenden Design dargestellt und sieht wie folgt aus:



So können sehr schnell professionelle Dashboards erstellt werden und der Zwischenschritt über den Power Query (wie im Schritt **Analyzer-Daten mit Power BI verbinden**) entfällt.

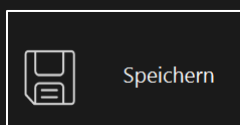
3.4 Projekt wechseln



Die Funktion „**Projekt wechseln**“ ermöglicht es dem Benutzer, zwischen verschiedenen vorhandenen Projekten (auch „Fälle“ genannt) zu wechseln, die in einem vordefinierten Verzeichnis gespeichert sind.

Im **Kapitel 1.2.1.2** wird das einfache Vorgehen ausführlich dargestellt.

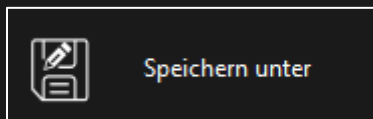
3.5 Speichern



Die **Speichern**-Funktion sichert Ihre aktuellen Einstellungen, Filter und Auswahlen, sodass Sie später problemlos darauf zurückgreifen können. Beim Klicken auf **Speichern** werden alle vorgenommenen Änderungen in einer Datei oder einem Projektstand festgehalten.

Dies stellt sicher, dass Ihre Arbeit nicht verloren geht und jederzeit in dem zuletzt gesicherten Zustand wiederhergestellt werden kann. Eine weitere Speicheroption befindet sich im eigentlichen **Header als Schnellzugriff (2.5)**. Beim erstmaligen Speichern öffnet sich der Explorer, damit der Speicherort bestimmt werden kann.

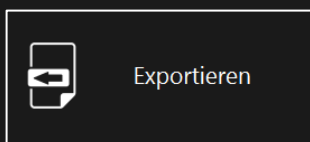
3.6 Speichern unter



Speichern unter ermöglicht es Ihnen, Ihre aktuelle Arbeit in einer neuen Datei abzulegen, ohne dabei die ursprüngliche Datei zu überschreiben. Diese Funktion ist

besonders nützlich, wenn Sie verschiedene Versionen Ihrer Daten oder Projekte anlegen möchten. Nach Auswahl eines Verzeichnisses und Vergabe eines Dateinamens wird eine separate Kopie angelegt, die unabhängig von der Ausgangsdatei weiterbearbeitet oder verteilt werden kann.

3.7 Exportieren



Unter **Exportieren** haben Sie die Möglichkeit, Ihre aufbereiteten Daten und Analysen in unterschiedlichen Formaten abzulegen oder weiterzugeben. Dadurch können Sie den Datenfluss in andere Anwendungen sicherstellen,

Ergebnisse langfristig archivieren oder mit Kollegen teilen.

Hinweis: Beim Export ermittelt **AUDIPY** stets die aktuell angezeigten Daten. Es gilt das Prinzip „*What you see is what you get*“ – das heißt, alle **aktiven Filter** werden auch beim Export berücksichtigt.

Abgrenzung zum Speichern unter:

Beim Speichern werden die gesamten Projekteinstellungen gesichert – inklusive Historie und Lesezeichen. Hierbei wird jedoch **nicht** der aktuelle Filterstatus übernommen. Beim späteren Laden eines gespeicherten Projekts erscheinen die Daten wieder im Ursprungszustand ohne die beim Speichern aktiven Filter. Beim Export ist es genau andersrum – Keine Historie / Lesezeichen, aber dafür der angezeigte Filter.

Im Einzelnen stehen Ihnen beim Export folgende Optionen zur Verfügung:



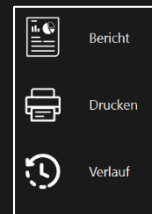
- **Parquet:** Ein spaltenbasiertes Speicherformat, das sich besonders für Big-Data- Analysen eignet, da es hohe Kompressionsraten und schnelle Lesezugriffe ermöglicht (weitere Informationen unter Kapitel 1.1).
- **CSV:** Ein weit verbreitetes Textformat zur Speicherung und zum Austausch strukturierter Daten. In einer CSV-Datei werden die Informationen zeilenweise organisiert, wobei jede Zeile einen Datensatz darstellt. Innerhalb einer Zeile sind die einzelnen Werte durch ein Trennzeichen – üblicherweise ein Komma – voneinander getrennt. Alternativ können je nach Ländereinstellung oder Anwendung auch andere Zeichen wie Semikolon oder Tabulator verwendet werden. CSV-Dateien sind leicht lesbar und können von einer Vielzahl von Programmen, darunter Tabellenkalkulationen (z.B. Microsoft Excel) und Datenbankanwendungen, problemlos verarbeitet werden. Aufgrund ihrer einfachen Struktur eignen sich CSV-Dateien besonders gut für den schnellen Export, Import oder die Weiterverarbeitung von Daten in verschiedenen Systemen.
- **XLSX:** Das Standardformat für Microsoft Excel. Hier können Sie Ihre Daten in einer Tabellenkalkulationsumgebung weiterbearbeiten, Diagramme erstellen oder zusätzliche Formeln anwenden.
- **PBIX:** Dieses Format wird von Microsoft Power BI genutzt. Damit lassen sich die exportierten Daten direkt in Power BI importieren. **Beim Export von PBIX-Dateien** stehen Ihnen dieselben Funktionen zur Verfügung, wie sie auch bei der direkten Verbindung von AUDIPY-Dateien mit Power BI genutzt werden (Verweis auf das Kapitel 3.2). Das bedeutet, dass die aufbereiteten und gefilterten Daten so bereitgestellt werden, dass sie in Power BI unmittelbar weiterverarbeitet und visualisiert werden können.

Durch die Auswahl des passenden Formats können Sie Ihre Ergebnisse optimal für den jeweiligen Anwendungsfall bereitstellen und so die Zusammenarbeit mit unterschiedlichen Teams oder Tools nahtlos gestalten.

3.8 Berichte



Unter „**Berichte**“ stehen Ihnen zentrale Funktionen zur Auswertung und Nachverfolgung zur Verfügung. Hier können Sie Berichte einsehen, ausdrucken und die vorgenommenen Prüfungsabläufe im **Verlauf** einsehen.



3.8.1 Bericht

Wenn Sie auf **Bericht** klicken, öffnet sich zuerst ein Fenster, in dem Sie den **Titel** ❶ für Ihren Bericht festlegen können. Mithilfe der Zusammenfassungsseite (standardisiert eingestellt) kann bereits ein Schnellüberblick gewonnen werden. Sie können hier auch entscheiden ob die Summe (hier muss dann die gewünschte zu summierende Spalte ausgewählt werden) oder ob die Anzahl berücksichtigt werden soll.

Ebenfalls können Sie hier die gewünschten **Spalten** ❷ auswählen, die später im fertigen Bericht erscheinen sollen. Über die Schnellschaltflächen ❸ „Alle“ bzw. „Keine“ können die Daten alle ausgewählt bzw. alle abgewählt werden, um eine neue Auswahl treffen zu können.

Nach dem Bestätigen des grünen Buttons „**PDF erstellen**“ müssen Sie in Ihrem Dateexplorer den gewünschten Speicherort festlegen und optional einen Namen vergeben. Anschließend öffnet sich der Bericht:

PDF-Bericht erstellen

Berichtseinstellungen

Berichtstitel
Datenanalysebericht ❶

☐ Zusammenfassungsseite hinzufügen

Aggregationsmetrik: Anzahl

Spalten auswählen ❸

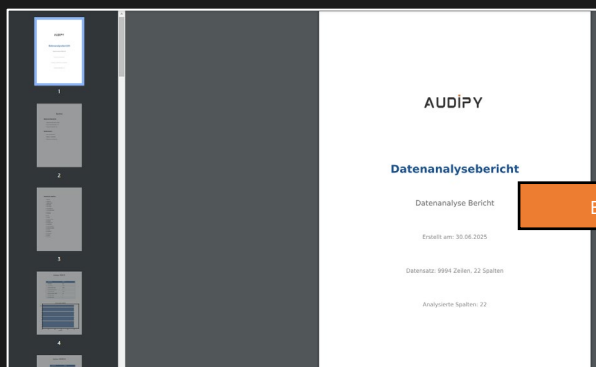
Alle Keine

❷

☐ ROW ID
☐ ORDER ID
☐ ORDER DATE
☐ SHIP DATE
☐ SHIP MODE
☐ CUSTOMER ID

Für jede ausgewählte Spalte wird eine Seite mit Statistiken und Diagramm erstellt.

PDF erstellen **Abbrechen**



Beispiel



Dort haben Sie die Möglichkeit, die Ansicht noch einmal zu überprüfen und den Bericht direkt auszudrucken. Dadurch können Sie schnell und einfach eine ansprechende Zusammenfassung Ihrer Daten erstellen und bei Bedarf weitergeben.

3.8.2 Drucken

Die Funktion „**Drucken**“ dient dazu, einen benutzerdefinierten Bericht auf Basis ausgewählter Spalten eines Datensatzes zu erzeugen und diesen für den Druck vorzubereiten. Sie erlaubt es, den Berichtstitel festzulegen sowie gezielt Spalten auszuwählen, die im Ausdruck erscheinen sollen.

Im Gegensatz zur vorherigen Funktion „Bericht“ steht hier lediglich der IST-Zustand der Spalten im Vordergrund, weshalb die Spalten so wie sie angezeigt werden auch gedruckt werden. Bei „Bericht“ konnte noch eine Aggregationsmetrik (Anzahl und Summe) ausgewählt werden.

Drucken

Berichtseinstellungen

Berichtstitel

Testdatei A

Spalten auswählen **Alle** **Keine**

☐ ROW ID

☐ ORDER ID

☐ ORDER DATE

☐ SHIP DATE

☐ SHIP MODE

☐ CUSTOMER ID

Bericht erzeugen **Abbrechen**

Wählen Sie die Spalten aus, die Sie in Ihren Bericht aufnehmen möchten.

Berichtseinstellungen 1

- **Berichtstitel**

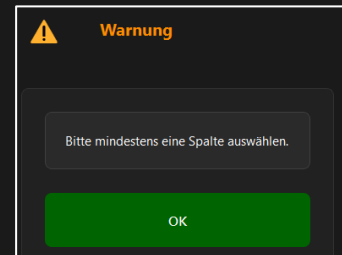
Hier kann ein individueller Titel für den Bericht eingetragen werden. Dieser erscheint später als Überschrift im gedruckten Bericht. (Beispiel im Bild: „Testdatei A“)

Spalten auswählen 2

- In diesem Bereich werden alle verfügbaren Spalten des Datensatzes aufgelistet.

- Sie können per Checkbox auswählen, welche Spalten in den Bericht aufgenommen werden sollen.
- **Zusätzliche Buttons**
- **Alle:** Wählt alle Spalten aus.
- **Keine:** Hebt alle Auswahlen auf.

Es muss mindestens eine Spalte ausgewählt werden, dies zeigt aber ansonsten auch ein Warnhinweis, dass eine Spalte auszuwählen ist.



◆ *Hinweis: Eine blaue Infozeile unterhalb der Liste weist darauf hin, dass nur ausgewählte Spalten in den Bericht übernommen werden.*

Mit Bericht erzeugen öffnet sich das folgende Fenster, welches einen sofortigen Druck oder eine Speicherung ① ermöglicht. Auch kann eine entsprechende Seite über die Seite aufgerufen werden ②.

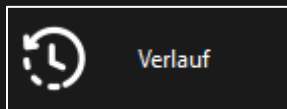
Testdatei A

ROW ID	SHIP MODE	CUSTOMER ID	PROFIT
1	Second Class	CG-12520	41.9136
2	Second Class	CG-12520	219.562
3	Second Class	DV-13045	6.8714
4	Standard Class	SO-20335	-383.031
5	Standard Class	SO-20335	2.5164
6	Standard Class	BH-11710	14.1694
7	Standard Class	BH-11710	1.9656
8	Standard Class	BH-11710	90.7152
9	Standard Class	BH-11710	5.7825
10	Standard Class	BH-11710	34.47
11	Standard Class	BH-11710	85.3092
12	Standard Class	BH-11710	68.3568
13	Standard Class	AA-10480	5.4432
14	Standard Class	IM-15070	132.5922
15	Standard Class	HP-14815	-123.858
16	Standard Class	HP-14815	-3.816
17	Standard Class	PK-19075	13.3176
18	Second Class	AG-10270	9.99
19	Second Class	ZD-21925	2.4824
20	Second Class	ZD-21925	16.011
21	Second Class	ZD-21925	7.384
22	Standard Class	KB-16585	5.0596
23	Standard Class	KB-16585	15.6884
24	Second Class	SF-20065	-1.0196
25	Standard Class	EB-13870	240.2649

ROW ID	SHIP MODE	CUSTOMER ID	PROFIT
26	Second Class	EH-13945	4.2224
27	Second Class	EH-13945	11.7741
28	Standard Class	TB-21520	-1665.0522

Search

3.8.3 Verlauf

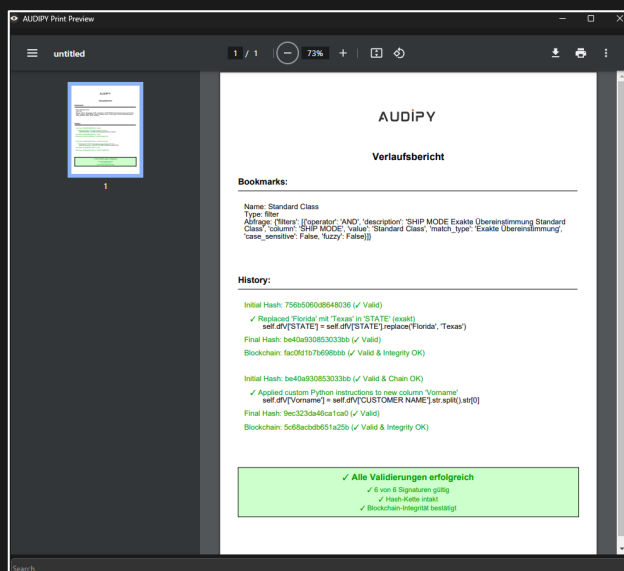


Der **Verlauf-Button** öffnet ein Fenster, in dem alle ausgeführten Aktionen und Datenänderungen chronologisch dokumentiert sind und sich jede Änderung lückenlos nachvollziehen lässt.

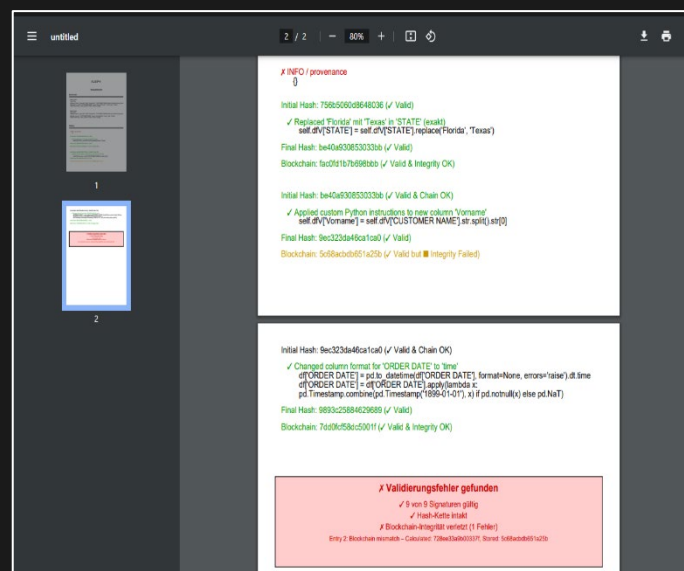
Aufgezeichnet werden neben den Prüfungsschritten auch die Lesezeichen **1**. **Wichtig hierbei ist, dass hier nur Verlaufseinträge und Lesezeichen angezeigt werden, welche vorher gespeichert wurden.**

	Action	Code
1	Replaced 'Florida' mit 'Texas' in 'STATE' (exakt)	<code>self.dfV["STATE"] = self.dfV["STATE"].replace('Florida', 'Texas')</code>
2	Applied custom Python instructions to new column 'Vorname'	<code>self.dfV["Vorname"] = self.dfV["CUSTOMER NAME"].str.split().str[0]</code>
3	Changed column format for 'ORDER DATE' to 'time'	<code>df["ORDER DATE"] = pd.to_datetime(df["ORDER DATE"], format=None, errors='raise')...</code>

Durch den Verlauf behalten Sie jederzeit den Überblick über Ihre Arbeitsschritte **2** und können bei Bedarf nachvollziehen, welche Anpassungen vorgenommen wurden. Zusätzlich haben Sie die Möglichkeit, einen Bericht **3** aus dieser Übersicht zu erstellen oder einen Sicherheitscheck durch den Hash-Check (s. weitere Ausführungen auf der nächsten Seite), um die einzelnen Aktionen in dokumentierter Form zu sichern oder weiterzugeben. Dies wird beispielhaft im folgenden Fenster dargestellt.



Beispiel: für eine gültigen Verlaufsbericht



Beispiel 2: Validierungsfehler gefunden

Der erzeugte Bericht enthält neben den inhaltlichen Auswertungen auch prüfbare Informationen zur Datenintegrität. Hierzu werden die verarbeiteten Daten mit **Hashwerten** versehen. Ein Hashwert ist ein eindeutiger, digitaler Fingerabdruck der Daten, der sich bereits bei kleinsten Änderungen vollständig verändert.

Im Bericht werden zwei Hashwerte ausgewiesen:

- **Initialer Hashwert:** Dieser wird zu Beginn der Analyse basierend auf den Originaldaten erzeugt.
- **Finaler Hashwert:** Dieser wird beim Speichern des Berichts auf Basis der verarbeiteten und gespeicherten Daten erneut gebildet.

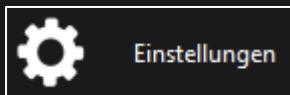
Durch einen späteren Vergleich beider Hashwerte kann überprüft werden, ob die Daten während des Analyseprozesses **unverändert** geblieben sind. Sollten dieselben Daten mit denselben Analyseschritten erneut verarbeitet werden, müssen identische Hashwerte entstehen. Jede noch so kleine **Manipulation der Daten** würde zu einem abweichenden Hashwert führen.

Zusätzlich wird jeder dieser Schritte in einer **Blockchain** verankert, wodurch die gesamte Historie unveränderlich und lückenlos dokumentiert wird. Am Ende des Berichts fasst ein grüner Validierungsblock alle Integritätsprüfungen zusammen:

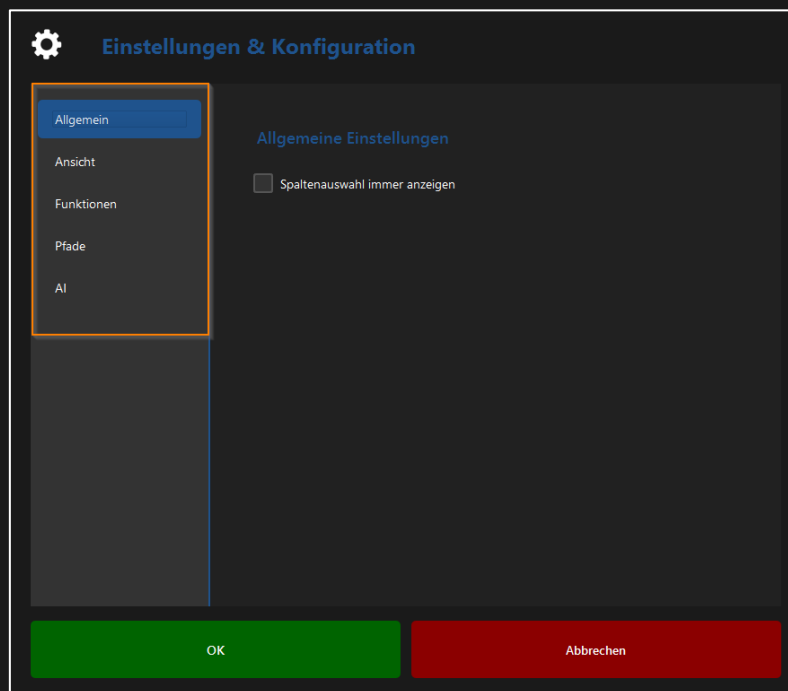
1. ☒ Alle Signaturen sind gültig
2. ☒ Die Hash-Kette ist intakt
3. ☒ Die Blockchain-Integrität wurde verifiziert

Damit stellt der Analyzer einen nachvollziehbaren Verlauf sicher und dokumentiert die Unveränderbarkeit der verwendeten Datenbasis – ein entscheidender Nachweis gerade in prüfungsrelevanten oder rechtssicheren Umgebungen.

3.9 Einstellungen

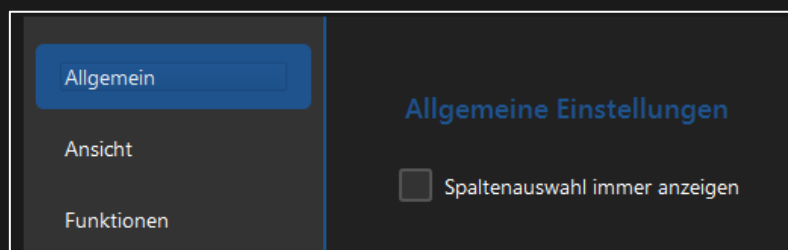


Über den Menüpunkt **Einstellungen** können Sie verschiedene Aspekte der Anwendung an Ihre Bedürfnisse anpassen. Hier lassen sich unter anderem allgemeine Voreinstellungen, Layoutoptionen, Funktionsumfänge sowie Verzeichnis- und KI-Einstellungen festlegen. Dadurch erhalten Sie die Möglichkeit, den Analyzer individuell zu konfigurieren und effizienter zu nutzen.



3.9.1 Allgemein

In diesem Bereich finden Sie grundlegende Optionen, die das Verhalten der Anwendung beeinflussen. Hier können Sie auswählen ob die Spaltenauswahl (Funktionserläuterung siehe unter Screenshot) immer angezeigt wird oder nicht.



Erläuterung der Funktion Spaltenauswahl:

Beim Laden von Parquet-Dateien bietet der Analyzer die Möglichkeit, gezielt nur bestimmte **Spalten auszuwählen**. Dies dient der **Performance-Optimierung** und Übersichtlichkeit, da oft nur ein Teil der Daten für eine konkrete Analyse benötigt wird.

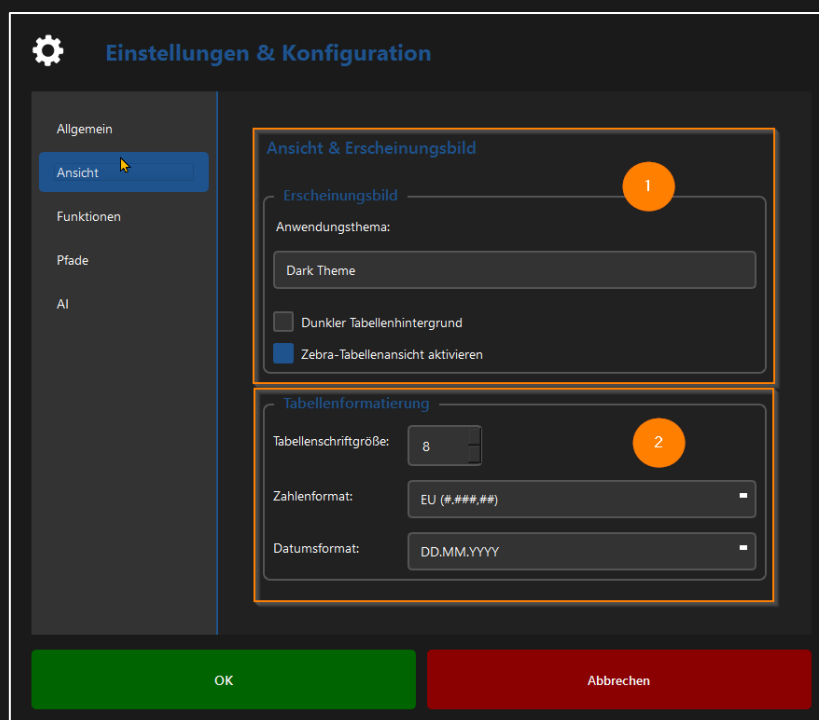
Durch die spaltenbasierte Struktur von Parquet-Dateien ist ein solches selektives Laden besonders effizient.

Hinweis: Beim Speichern des Falls werden **nur die geladenen Spalten** gesichert. Spalten, die beim Laden nicht ausgewählt wurden, gehen beim normalen Speichern verloren.

Wenn der vollständige Datenbestand erhalten bleiben soll, empfiehlt es sich, stattdessen die Funktion „**Speichern unter...**“ zu verwenden und dabei bewusst einen neuen Dateistand zu erzeugen, um den Originalinhalt nicht zu überschreiben.

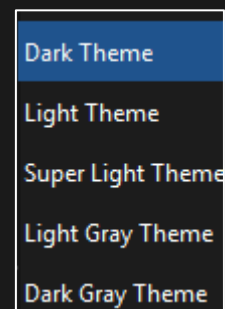
Diese Funktionalität unterstützt einen ressourcenschonenden Analyseprozess und bietet gleichzeitig Flexibilität im Umgang mit umfangreichen Datenbeständen.

3.9.2 Ansicht



Unter **Ansicht** haben Sie die Möglichkeit, das Erscheinungsbild Ihrer Tabellen individuell anzupassen. So können Sie das gesamte Programm in unterschiedlichen Farbdesigns verwenden, je nach individuellen Vorlieben. Im Folgenden wird aus Vereinfachungsgründen die dunkle Ansicht vorgestellt. **Wichtig: Das Programm muss für die Umsetzung neu gestartet werden.**

Ebenfalls können Sie in diesem Bereich ① einen dunklen Tabellenhintergrund (lediglich für die Daten auswählen) oder eine



Zebra-Tabellenansicht (jede zweite Spalte wird gräulich von der vorherigen abgesetzt) aktivieren, um die Lesbarkeit der Zeilen zu erhöhen.

Row ID	Order ID	Order Date
1	CA-2020-152156	08.11.2020
2	CA-2020-152156	08.11.2020
3	CA-2020-138688	12.06.2020
4	US-2019-108966	11.10.2019
5	US-2019-108966	11.10.2019
6	CA-2018-115812	09.06.2018

Ohne Zebra-Tabellenansicht

Row ID	Order ID	Order Date
1	CA-2020-152156	08.11.2020
2	CA-2020-152156	08.11.2020
3	CA-2020-138688	12.06.2020
4	US-2019-108966	11.10.2019
5	US-2019-108966	11.10.2019
6	CA-2018-115812	09.06.2018

Mit Zebra-Tabellenansicht

Row ID	Order ID	Order Date
1	CA-2020-152156	08.11.2020
2	CA-2020-152156	08.11.2020
3	CA-2020-138688	12.06.2020
4	US-2019-108966	11.10.2019
5	US-2019-108966	11.10.2019
6	CA-2018-115812	09.06.2018

Helle Ansicht mit Zebra-Tabellenansicht

Row ID	Order ID	Order Date
1	CA-2020-152156	08.11.2020
2	CA-2020-152156	08.11.2020
3	CA-2020-138688	12.06.2020
4	US-2019-108966	11.10.2019
5	US-2019-108966	11.10.2019
6	CA-2018-115812	09.06.2018

Dunkler Modus mit Zebra-Tabellenansicht

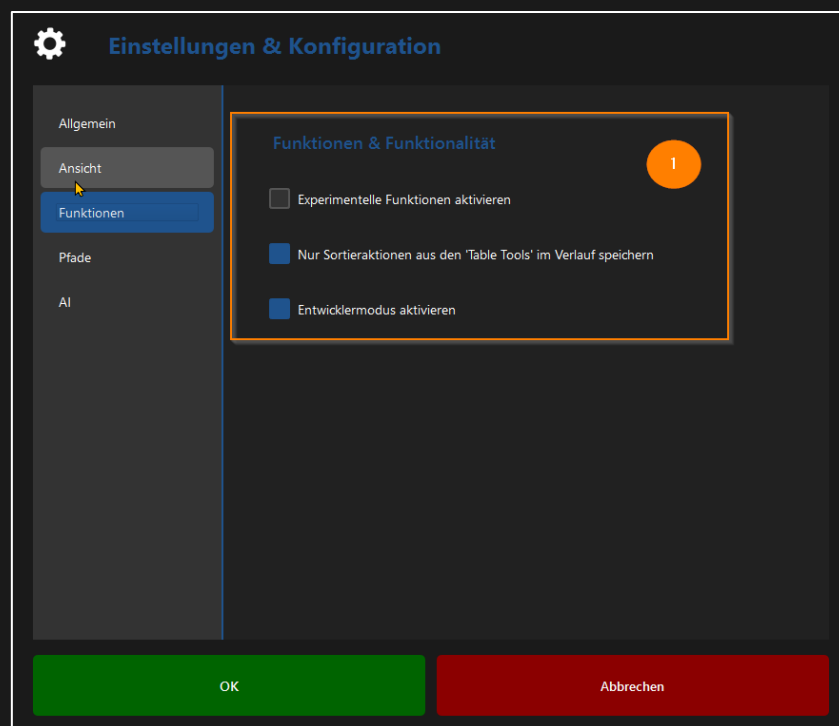
Darüber hinaus lassen sich **Schriftgröße, Zahlenformat und Datumsformat** 2 anpassen. So können Sie das Layout ganz nach Ihren persönlichen Vorlieben oder unternehmensspezifischen Standards gestalten.

3.9.3 Funktionen

Unter **Funktionen** können Sie zusätzliche oder erweiterte Features der Anwendung verwalten ①. Beispielsweise lassen sich **experimentelle Funktionen** aktivieren, die sich noch in der Testphase befinden, aber neue Möglichkeiten bieten können.

Ein Hinweis zur Option „Sortieraktionen aus den Table-Tools im Verlauf aufzeichnen“: **Diese Option dient dazu festzulegen, ob Sortierungen in der Tabellenansicht dokumentiert werden sollen oder nicht.**

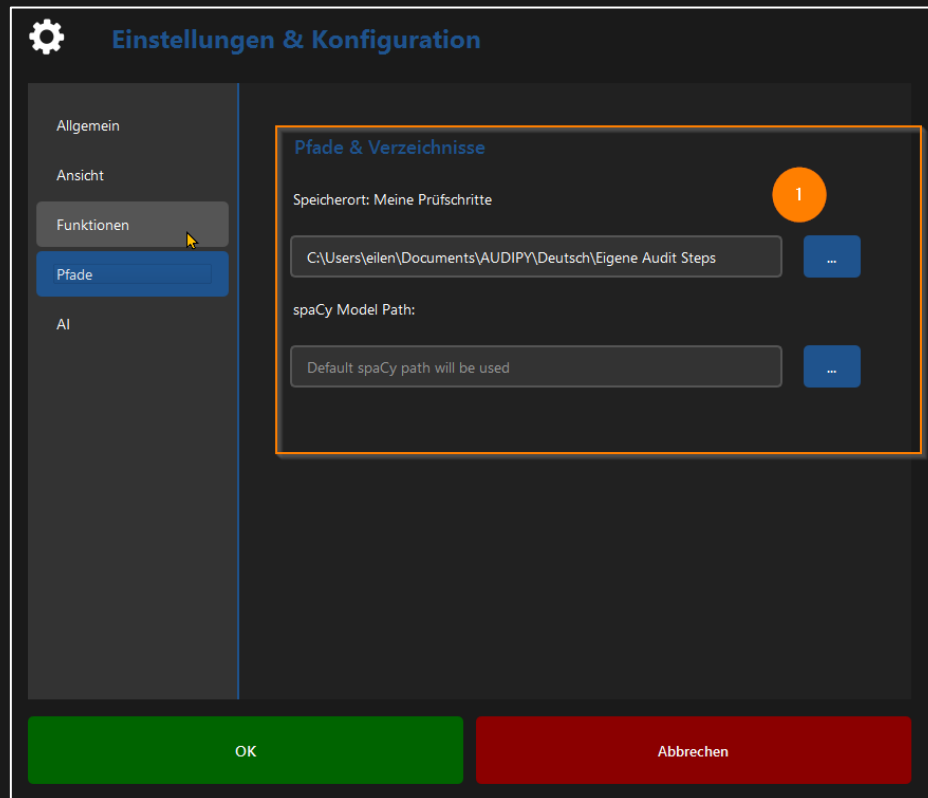
Wenn Sie z.B. eine Sortierung für einen Audit-Schritt benötigen, müssen Sie sich darüber bewusst sein, dass im Standard nur Sortierungen gespeichert werden, die über das **Table Tool** ausgeführt wurden.



Für fortgeschrittene Benutzer gibt es außerdem einen **Entwicklermodus**, der erweiterte Einblicke in die Abläufe der Anwendung ermöglicht. Nach Aktivierung muss das Programm neu gestartet werden. Für weitere Informationen wird auf das Kapitel 6 verwiesen.

3.9.4 Pfade

Im Bereich **Pfade** können Sie festlegen, wo die Anwendung bestimmte Daten und Konfigurationen speichert oder ausliest. Dazu zählt zum Beispiel ein Speicherort **1** für eigene Arbeitsschritte oder Audit Steps (s. Kapitel 9). Darüber hinaus besteht die Möglichkeit, den Pfad für das verwendete **spaCy-Modell im Rahmen der Named Entity Recognition (NER) Analyse** (s. Kapitel 14.1.8) anzupassen.



So können Sie bei Bedarf eine benutzerdefinierte Modellversion einbinden oder einen alternativen Speicherort nutzen, um Ihre Analysetools optimal zu organisieren.

3.9.5 AI

Die Einstellungen richten sich nach dem gewählten Modell. Bezieht man die KI über Semitax, ist **AUDIPY AI** automatisch voreingestellt **1** und ein API-Key entfällt, da dieser im Semitax-Lizenzschlüssel enthalten ist. Nutzt man hingegen ein anderes KI-Modell, kann dieses frei gewählt werden – in diesem Fall muss der zugehörige API-Key hinterlegt werden.

Einstellungen & Konfiguration

Allgemein
Ansicht
Funktionen
Pfade
AI

KI-Konfiguration

AI Model Selection

Wählen Sie das KI-Modell:

AUDIPY AI

AI Services

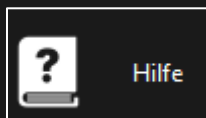
Chat GPT	https://chatgpt.com	Löschen
Google Gemini	https://gemini.google.com/app	Löschen
Claude AI	https://claude.ai/new	Löschen

Add new service

OK Abbrechen

Durch das Hinzufügen neuer Services können Sie den Funktionsumfang flexibel erweitern und die KI-Unterstützung optimal an Ihre Anforderungen anpassen. Diese sind z.B. auch entscheidend für die Funktion der benutzerdefinierten Spalte (s.14.1.6.3.2)

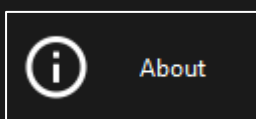
3.10 Hilfe



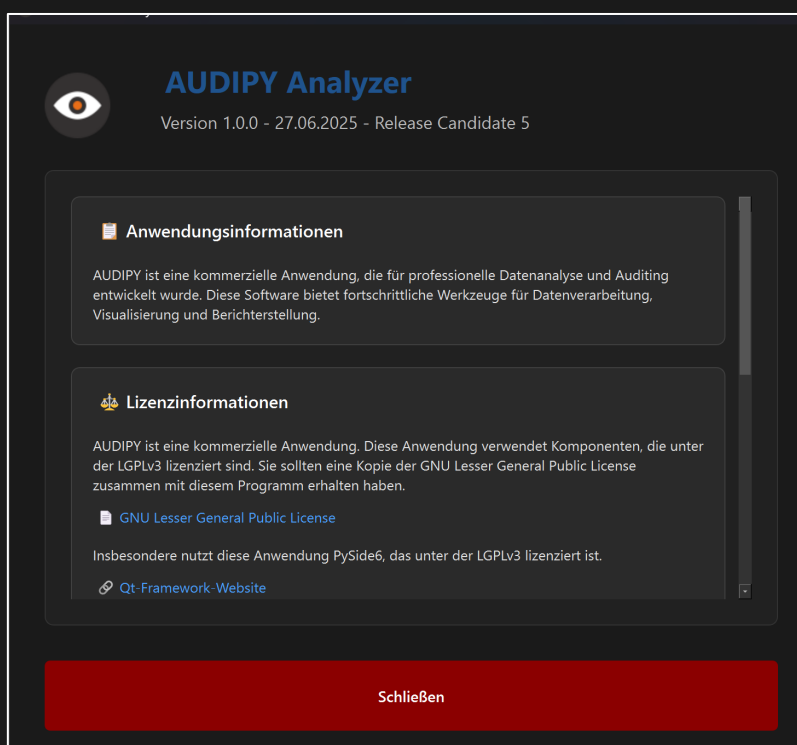
Über **Hilfe** öffnet sich die jeweils aktuelle Fassung des **Handbuchs zum Analyzer**. Das Handbuch enthält alle wichtigen Informationen zur Bedienung, zu Funktionen sowie zu häufigen Anwendungsfällen und unterstützt Sie bei der effektiven Nutzung des Programms.

Bei weiteren Fragen bitte den Kundenservice unter support@semitax.de kontaktieren.

3.11 About



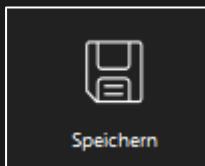
Unter **About** finden Sie Informationen zur aktuellen Version der Anwendung, ihren Lizenzbestimmungen und den verwendeten Komponenten. Darüber hinaus werden Kontaktdaten sowie der Hinweis auf die Website des Herstellers angezeigt. So erhalten Sie einen schnellen Überblick über rechtliche Hinweise, Entwicklerinformationen und Supportmöglichkeiten.



4 Explorer

Die detaillierte Darstellung des Datei-Explorers zur Dateiauswahl erfolgte bereits in Kapitel 1.2 unter Berücksichtigung ergonomischer und nutzerfreundlicher Aspekte. An dieser Stelle wird daher ausschließlich auf die dortigen Ausführungen verwiesen.

5 Speichern



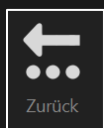
Die **Speichern**-Funktion sichert Ihre aktuellen Einstellungen, Filter und Auswahlen, sodass Sie später problemlos darauf zurückgreifen können. Beim Klicken auf **Speichern** werden alle vorgenommenen Änderungen in einer Datei oder einem Projektstand festgehalten.

Dies stellt sicher, dass Ihre Arbeit nicht verloren geht und jederzeit in dem zuletzt gesicherten Zustand wiederhergestellt werden kann. Beim erstmaligen Speichern von **CSV Dateien** oder beim **Power BI Import** öffnet sich der Explorer, damit der Speicherort bestimmt werden kann.

Hinweis:

Das Speichern sorgt darüber hinaus auch erst für die **Darstellung im Verlauf (3.8.3)**.

6 Zurück



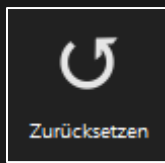
Die Zurück-Funktion ermöglicht es Ihnen, **bis zu 3** Ihrer zuletzt ausgeführten Schritte sofort zurückzunehmen. Unter dem Pfeil-Icon befinden sich drei weiße Punkte, die im orangen leuchtenden Zustand anzeigen, wie viele Rückschritte aktuell möglich sind. Leuchten alle 3 Punkte (●●●), können Sie dreimal hintereinander auf das Pfeil-Symbol klicken und auf die vorherigen Ansichten zurückkommen. Nach jedem Klick wird ein Punkt wieder weiß (●●● → ●●● → ●●● → ●●●), bis keine Zurück-Ebene mehr vorhanden ist. Erst wenn Sie neue Aktionen ausführen, füllen sich die Punkte wieder auf und machen die Funktion erneut verfügbar.

Bitte beachten:

Reine Filterungen werden nicht über die Zurück-Funktion zurückgenommen, da sie nicht als Bearbeitungsschritte im Historienpuffer abgelegt werden. Verwenden Sie stattdessen die rote **Filter löschen**-Schaltfläche , um angewendete Filter rückgängig zu machen und alle Datensätze wieder anzuzeigen.

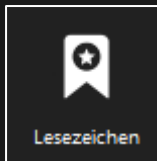
Die Funktionen „**Speichern**“ und „**Zurücksetzen**“ löschen die Historie. Hier sollte also bewusst gespeichert werden.

7 Zurücksetzen



Die **Zurücksetzen**-Funktion verwirft alle aktuell vorgenommenen Änderungen, Filter und Auswahlen und stellt den ursprünglichen Zustand wieder her. Dadurch können Sie schnell zu einer neutralen Ausgangslage zurückkehren, falls falsche Anpassungen vorgenommen wurden oder Sie eine neue Analyse starten möchte

8 Lesezeichen



Der Menüpunkt **Lesezeichen** öffnet eine seitliche Leiste, die Sie – ähnlich wie beim Explorer – sowohl über den Headerpunkt als auch über kleine Punkte **1** oder die Pfeile **2** am Bildschirmrand ein- oder ausklappen können.

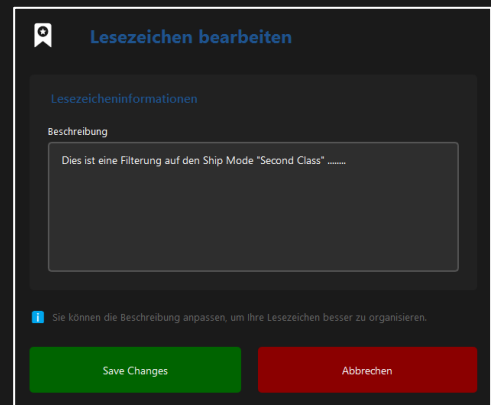
Customer ID	Customer Name	Segment	Country/Region	City	State	Postal Code	Region	
PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53.711,0	Central	OFF-ST
PK-19075	Pete Kriz	Consumer	United States	Detroit	Michigan	48.227,0	Central	FUR-FL
PK-19075	Pete Kriz	Consumer	United States	Detroit	Michigan	48.227,0	Central	OFF-ST
PK-19075	Pete Kriz	Consumer	United States	Detroit	Michigan	48.227,0	Central	OFF-ST
PK-19075	Pete Kriz	Consumer	United States	Detroit	Michigan	48.227,0	Central	OFF-Pk
PK-19075	Pete Kriz	Consumer	United States	Detroit	Michigan	48.227,0	Central	FUR-Cr
PK-19075	Pete Kriz	Consumer	United States	Detroit	Michigan	48.227,0	Central	OFF-LA
PK-19075	Pete Kriz	Consumer	United States	Columbus	Ohio	43.229,0	East	FUR-Ch
PK-19075	Pete Kriz	Consumer	United States	San Francisco	California	94.110,0	West	OFF-En
PK-19075	Pete Kriz	Consumer	United States	Columbia	South Carolina	29.203,0	South	OFF-Bi
PK-19075	Pete Kriz	Consumer	United States	Milwaukee	Wisconsin	53.209,0	Central	OFF-LA
PK-19075	Pete Kriz	Consumer	United States	Newark	Delaware	19.711,0	East	OFF-AF
PK-19075	Pete Kriz	Consumer	United States	San Francisco	California	94.122,0	West	OFF-Bi
PK-19075	Pete Kriz	Consumer	United States	San Francisco	California	94.122,0	West	FUR-TA

Ein Lesezeichen speichert Ihre aktuellen Einstellungen, etwa Filter, Darstellungen oder Berechnungen, sodass Sie diese mit nur einem Klick jederzeit wiederherstellen können **3**. Dabei helfen verschiedene Symbole (z. B. ein Trichtersymbol für gefilterte Ansichten oder ein Amy-Icon für Ergebnisse der Assistentin), die Lesezeichen ihrem Ursprung zuzuordnen.

Folgende Symbole sind hierbei möglich:

- Filterung
- Ask my Data (AMY)
- Visualisierung
- Analyse
- Gruppierung
- Pivot

Ein kleines Info-Symbol neben dem Lesezeichenname zeigt an, dass bei diesem auch eine Beschreibung vorhanden ist. Mit Klick auf das Infosymbol , wird die Beschreibung geöffnet und kann jederzeit wieder angepasst werden. Die Beschreibung ist lediglich **optional**.



Über das Lesezeichen Icon auf der rechten Seite  können Lesezeichen hinzugefügt werden. Zusätzlich lassen sich die Lesezeichen farblich markieren und jederzeit aktualisieren, um Änderungen an der Darstellung zu übernehmen.

Bitte beachten: Ein Lesezeichen kann nur geöffnet werden, wenn auch ein Datenbestand geöffnet ist.

8.1 Lesezeichen hinzufügen

Klickt man auf das **Lesezeichen-Symbol** im Seitenschenkel öffnet sich das folgende Fenster:



Im Dialog **Lesezeichen hinzufügen**, stehen Ihnen zwei Eingabefelder zur Verfügung:

1. **Lesezeichen-Name ①**: Geben Sie hier eine eindeutige Bezeichnung ein, um das Lesezeichen später leichter wiederzufinden. Der Name ist wichtig, ansonsten erhalten Sie einen Fehlerhinweis.
2. **Lesezeichenbeschreibung (optional) ②**: Tragen Sie **bei Bedarf** eine kurze Erläuterung ein, um den Zweck oder Kontext des Lesezeichens zu dokumentieren.

Nach dem Ausfüllen können Sie das neue Lesezeichen mit **Hinzufügen** speichern oder den Vorgang mit **Abbrechen** beenden.

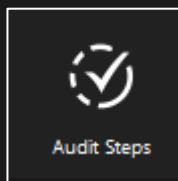
8.2 Lesezeichen Tools

Ein **Rechtsklick** auf ein Lesezeichen öffnet ein Kontextmenü, mit dem Sie das Lesezeichen nachträglich verwalten und anpassen können. Im Einzelnen stehen Ihnen folgende Aktionen zur Verfügung:

- **Löschen**: Entfernt das ausgewählte Lesezeichen vollständig.
- **Umbenennen**: Ändert den Namen des Lesezeichens, um es später leichter wiederzufinden.
- **Beschreibung**: Bearbeiten oder Hinzufügen einer Kurzbeschreibung, um den Zweck oder Inhalt des Lesezeichens zu dokumentieren.
- **Aktualisieren**: Speichert den aktuellen Zustand (z. B. neue Filter, geänderte Darstellungen) im Lesezeichen.
- **Farbe ändern**: Ordnet dem Lesezeichen eine individuelle Farbe zu, damit Sie in der Übersicht schnell zwischen verschiedenen Lesezeichen unterscheiden können. Sie können aus verschiedenen Farben wählen



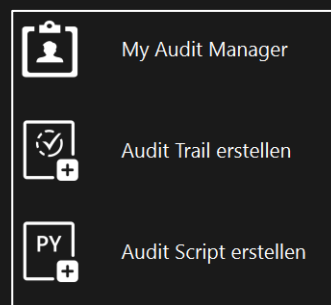
9 Audit Steps



Die **Audit Steps** beinhalten alle Prüfschritte, die Sie auf Ihre Daten anwenden möchten, zum Beispiel zur Qualitätssicherung oder Einhaltung von Vorschriften. Um diese Schritte wiederholt verwenden zu können, müssen Ihre neuen Daten jedoch den gleichen Aufbau aufweisen (z. B. dieselben Spalten wie die zuvor geprüften Datensätze).

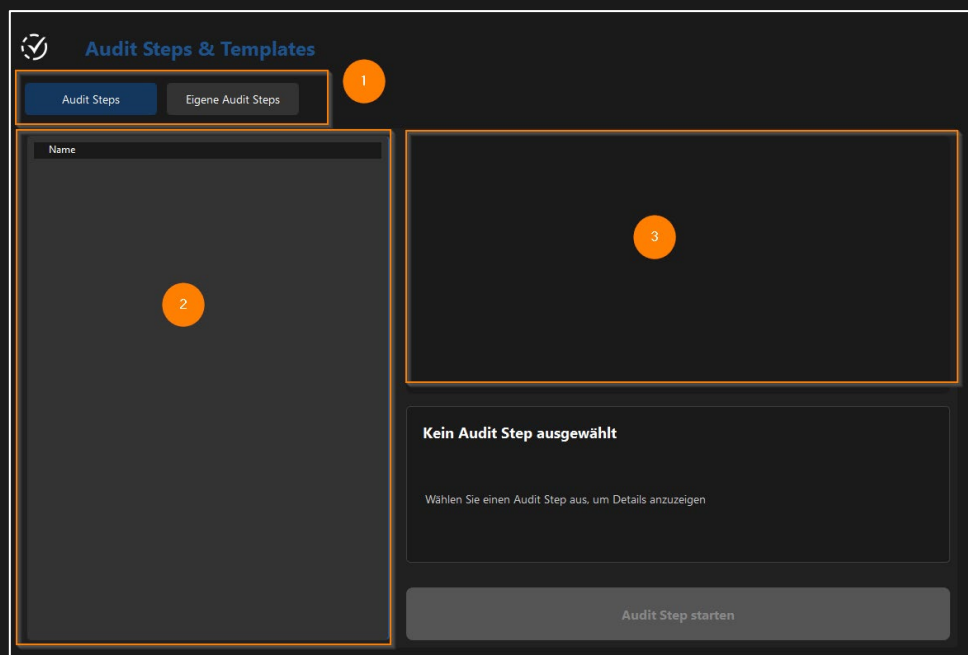
Nur so lassen sich die definierten Schritte reibungslos übertragen und automatisiert abarbeiten, ohne dass manuelle Anpassungen erforderlich werden.

Im Hauptmenü können Sie mit **My Audit Manager** alle Auditschritte zentral verwalten. Über **Audit Trail erstellen** legen Sie neue Prüfpfade an und dokumentieren diese nachvollziehbar. Mit **Audit Script erstellen** definieren Sie eigene Prüfskripte in Python zur Automatisierung von Prüfabläufen.



9.1 My Audit Manager

Über My Audit Manager können Sie verschiedene Prüfschritte organisieren und ausführen. Dabei sind die Bereiche wie folgt aufgeteilt:



1. Audit Steps / Eigene Audit Steps ①

- **Audit Steps:** Enthält zum Beispiel firmeninterne, standardisierte Prüfungsschritte, die für mehrere Projekte oder Abteilungen relevant sind.

- **Eigene Audit Steps:** Hier können Sie persönliche oder projektspezifische Prüfschritte ablegen, die nur für Ihren individuellen Anwendungsfall gelten. Darüber hinaus können Sie hier auch Ihre Audit Steps importieren. Den Speicherort können Sie unter Einstellungen – Pfade festlegen (s. **3.9.1.4** Pfade)

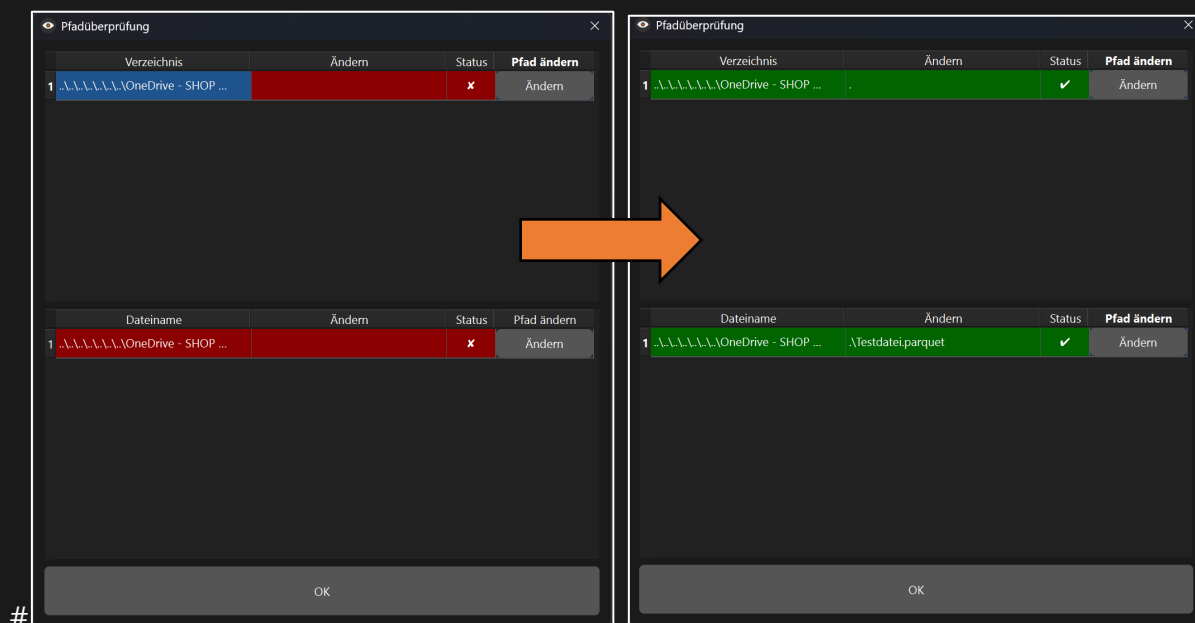
2. Liste der vorhandenen Audit Steps ②

- In dieser Spalte werden alle verfügbaren Prüfungsschritte angezeigt. Sie können einen Schritt auswählen, um dessen Details einzusehen oder ihn direkt zu starten.

3. Detailbereich ③

- Hier wird eine ausführliche Beschreibung des ausgewählten Audit Steps dargestellt. Auf diese Weise sehen Sie, welche Prüfungen im Einzelnen durchgeführt werden und können gezielt nachvollziehen, welche Datenfelder oder Kriterien geprüft werden. Mit Rechtsklick auf einen importierten Audit Step kann dieser umbenannt oder gelöscht werden.

Nach dem Start eines Prüfungsschritts wird folgendes Fenster geöffnet, bei welchem der entsprechende Pfad geändert werden muss.

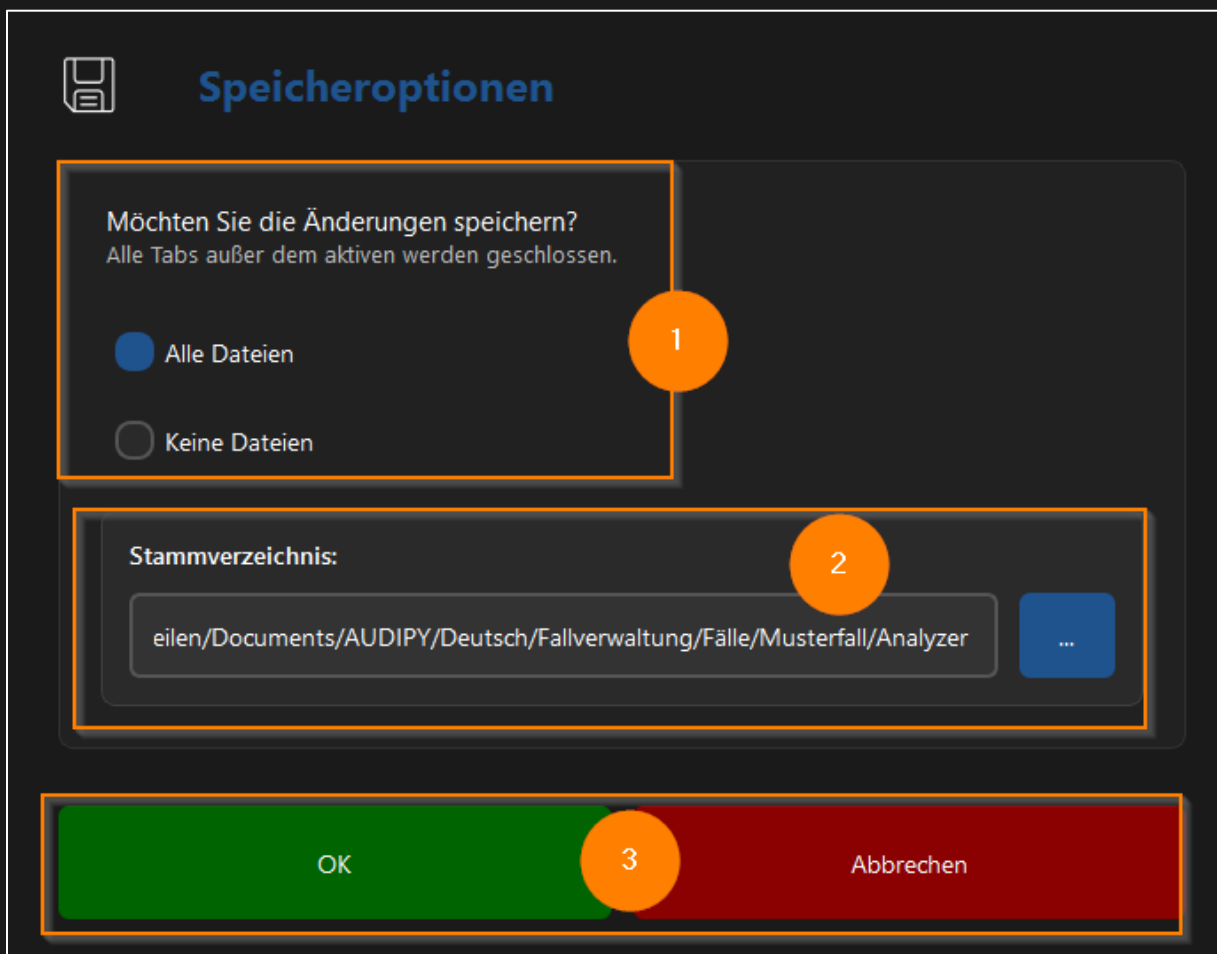


Nach Pfadanpassung sollte alles grün leuchten. Anschließend kann der Dialog mit OK geschlossen werden.

9.2 Audit Trail erstellen

Der **Audit Trail** ermöglicht es, einzelne Prüfungsschritte direkt aufzuzeichnen, nachdem sie durchgeführt wurden. Diese Aufzeichnungen lassen sich im Anschluss als eigenständige **Audit Steps** abspeichern und bei Bedarf wiederverwenden. Auf diese Weise können Prüfprozesse effizient dokumentiert, standardisiert und für zukünftige Audits reproduzierbar gestaltet werden.

Nach Klick auf „Audit Trail erstellen“ erscheint folgendes Fenster:



1. Speicheroptionen ①

- Sie werden gefragt, ob Sie alle offenen Dateien (Tabs) speichern möchten, bevor sie – bis auf den aktiven – geschlossen werden, da im folgenden Fenster nur gespeicherte Aktionen und Lesezeichen angezeigt werden, ähnlich wie bei der Verlaufsansicht.
- Wählen Sie **Alle Dateien**, um sämtliche aktuellen Änderungen zu sichern, oder **Keine Dateien**, wenn Sie keine weiteren Anpassungen speichern wollen.

2. Stammverzeichnis ②

- Unter *dem Stammverzeichnis* geben Sie an, in welchem Verzeichnis die zu exportierenden Audit Steps abgelegt werden sollen. Hinweis: Es wird maximal eine weitere Ordnerstufe durchsucht – bei einem Pfad wie *FiBu 2020/DATEV* würde also nur *FiBu 2020* nach Parquet-Dateien durchsucht.
- Hier können Sie beispielsweise einen lokalen Ordner, ein Netzwerkverzeichnis oder einen Cloud-Speicherpfad angeben, je nachdem, wo Sie Ihre Audit Steps aufbewahren möchten.

3. Bestätigen und Export durchführen

- Klicken Sie auf **OK**, um den Export zu starten.
- Über **Abbrechen** brechen Sie den Vorgang ab, ohne Änderungen vorzunehmen.

Mit Betätigung des OK Buttons öffnet sich das folgende Dialogfenster:

1. Dateiauswahl ①

In diesem Bereich wählen Sie die Datei aus, deren enthaltene Informationen (z. B. Aktionen oder Lesezeichen) Sie exportieren möchten. Mehrere Dateien können angezeigt werden, aber es kann jeweils nur eine gleichzeitig **aktiv** sein. Über den Pfad ② kann der jeweilige Fall ausgewählt werden.

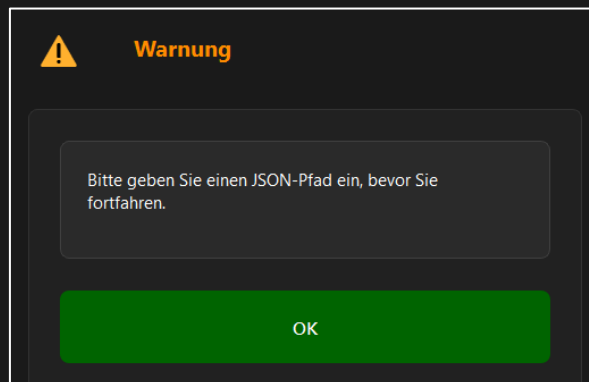
2. JSON Pfad auswählen ②

Über die Schaltfläche „**Pfad auswählen**“ legen Sie fest, wohin die exportierten Daten im JSON-Format gespeichert werden sollen.

`C:/Users/eilen/Documents/AUDIPY/Deutsch - FV/Eigene Audit Steps/Test.audit`

Wichtig: Sie müssen einen Namen vergeben! Ohne Namen können Sie keine Datei speichern (hierfür kommt aber kein Warnhinweis). Im Beispielsfall wurde die Datei **Test** unter eigene Audit Steps gespeichert.

Legen Sie hier keinen Pfad fest und klicken Sie auf „**Ok**“ erhalten Sie folgenden Warnhinweis:



3. Dateipfad und erkannte Inhalte ③

Hier sehen Sie den Pfad der aktuell ausgewählten Datei. Das Programm erkennt automatisch, welche Inhalte in der Datei vorhanden sind, und zeigt diese in den entsprechenden Tabs wie *Lesezeichen*, *Aktionen*, *zusätzliche Metadaten* und *Beschreibung* an.

4. Auswahl innerhalb der Rubriken ④

In der jeweiligen Rubrik (z.B. Aktionen) können Sie **genau auswählen**, welche Elemente Sie exportieren möchten – etwa bestimmte Bearbeitungsschritte oder Metadaten. Nicht ausgewählte Elemente werden nicht mit exportiert.

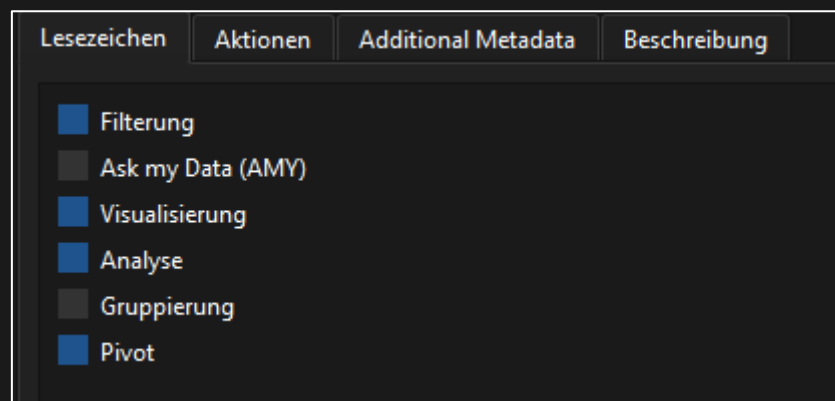


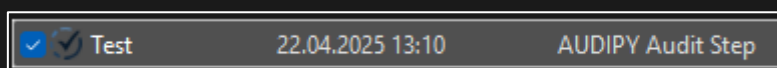
Abb: Hier würden die Lesezeichen Ask my Data und die Gruppierung nicht mit exportiert werden.

5. Schnellzugriffe für Auswahl 5

Für Rubriken wie *Lesezeichen* und *Aktionen* stehen Schnellzugriff-Schaltflächen zur Verfügung, mit denen Sie alle Einträge auf einmal auswählen oder abwählen können. Das erleichtert die Handhabung bei vielen Einträgen.



Nach Abschluss dieses Vorgangs finden Sie die Datei im gewünschten Speicherort (Sie legen unter 2 den gewünschten Pfad fest). Die Datei wird wie folgt abgespeichert:



Diese Datei können Sie später wieder importieren oder mit anderen Personen teilen, um dieselben Prüfschritte in einer anderen Umgebung oder einem neuen Projekt nutzen zu können.

9.3 Audit Script erstellen

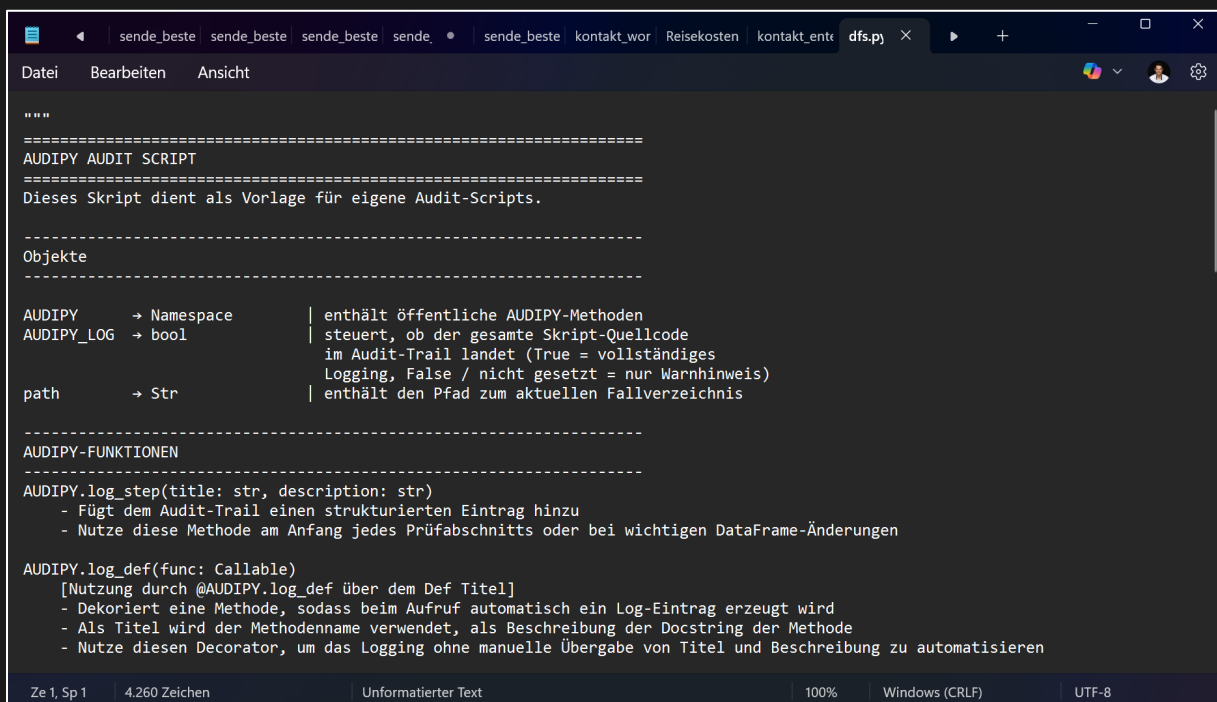
Die AUDIT-Skripte dienen als Grundlage für individuelle Prüf- und Dokumentationsschritte.

Es öffnet sich bei diesem Vorgang ein **Skriptfenster**, in dem du die Möglichkeit hast, die Standardlogik anzupassen oder zu erweitern. Hier kannst du spezielle Prüfungen, Berechnungen oder zusätzliche Protokollierungen direkt im Python-Quellcode definieren.

Um diese Anpassungen vorzunehmen, ist **Spezialwissen in Python** notwendig. Der mitgelieferte Code dient als Vorlage und bietet bereits:

- einen vorbereiteten Namespace (AUDIPY) mit den zentralen Funktionen,
- Variablen zur Steuerung des Loggings und
- Methoden, um eigene Schritte im Audit-Trail zu dokumentieren.

Die Skripte sind so konzipiert, dass auch komplexe Prüf- und Auswertungslogiken umgesetzt werden können, während gleichzeitig eine lückenlose Dokumentation gewährleistet bleibt.



```
"""
=====
AUDIPY AUDIT SCRIPT
=====
Dieses Skript dient als Vorlage für eigene Audit-Skripts.

-----
Objekte
-----

AUDIPY      → Namespace      | enthält öffentliche AUDIPY-Methoden
AUDIPY_LOG  → bool           | steuert, ob der gesamte Skript-Quellcode
                             | im Audit-Trail landet (True = vollständiges
                             | Logging, False / nicht gesetzt = nur Warnhinweis)
path        → Str            | enthält den Pfad zum aktuellen Fallverzeichnis

-----
AUDIPY-FUNKTIONEN
-----
AUDIPY.log_step(title: str, description: str)
- Fügt dem Audit-Trail einen strukturierten Eintrag hinzu
- Nutze diese Methode am Anfang jedes Prüfabschnitts oder bei wichtigen DataFrame-Änderungen

AUDIPY.log_def(func: Callable)
[Nutzung durch @AUDIPY.log_def über dem Def Titel]
- Dekoriert eine Methode, sodass beim Aufruf automatisch ein Log-Eintrag erzeugt wird
- Als Titel wird der Methodenname verwendet, als Beschreibung der Docstring der Methode
- Nutze diesen Decorator, um das Logging ohne manuelle Übergabe von Titel und Beschreibung zu automatisieren
=====
"""
```


10 Ask my Data (AMY)

Hinweis: Diese KI Funktion ist abhängig vom gewählten Abomodell. Falls sie in Ihrer Version nicht verfügbar ist, prüfen Sie bitte Ihr aktuelles Abonnement oder kontaktieren Sie den Support unter support@semitax.de.

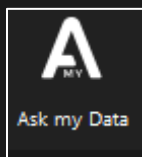
10.1 Abo-Modelle

Folgende Abo Modelle sind hinsichtlich der KI (Programm auch ohne KI Funktionen nutzbar) möglich:

 AI STARTER	 AI PRO
• Amy Column Generator – Automatische Generierung und Optimierung von Datenspalten für strukturierte Analysen. • DSGVO-konformes, auf Daten optimiertes Modell – Entwickelt für den sicheren Einsatz zur Datenanalyse im europäischen Raum. • Offline-Modelle (z. B. GPT4All) – Flexibel nutzbare KI, unabhängig von einer dauerhaften Internetverbindung. • Integration gängiger KI-Dienste – Möglichkeit zur Nutzung verschiedener leistungsfähiger Modelle, je nach individuellem Bedarf (z.B. Open AI, Mistral oder DeepSeek)	• Alle Funktionen von AI Starter – Mit zusätzlichem Fokus auf interaktive und erweiterte Analysefunktionen. • Amy "Ask my Data" Chat – Stellen Sie Fragen zu Ihren Daten in natürlicher Sprache und erhalten Sie nicht nur Antworten, sondern auch automatisch generierte Tabellen und interaktive Diagramme für eine visuell unterstützte Analyse. • Amy Monetary Unit Sampling – Unterstützt gezielt bei der Prüfung finanzieller Daten und hilft bei der Generierung der entsprechenden Auswertungen, um eine präzise Analyse zu ermöglichen. • Power BI Integration – Chatten Sie direkt mit Ihren Power BI-Daten und lassen Sie sich DAX- und M-Code generieren, um Berichte, Analysen und Dashboards gezielt zu optimieren. • Sprachoptimiertes Modell – Speziell für die Verarbeitung von

	Text und Sprache konzipiert, um komplexe Berichte zu generieren. • Kontextabhängige Modellauswahl – Automatische Anpassung an den jeweiligen Anwendungsfall für optimale Ergebnisse.
Ideal für alle, die erste Schritte mit KI machen und dabei auf eine datenschutzkonforme, zukunftsichere Lösung setzen möchten.	AI Pro richtet sich an Unternehmen und Fachkräfte, die über klassische Datenanalysen hinausgehen und KI aktiv in ihre Entscheidungsprozesse einbinden möchten. Besonders die Kombination aus interaktivem Daten-Chat, automatischer Auswertung und visueller Unterstützung macht es zu einer leistungsstarken Lösung für anspruchsvolle Aufgaben.

10.2 Funktionen

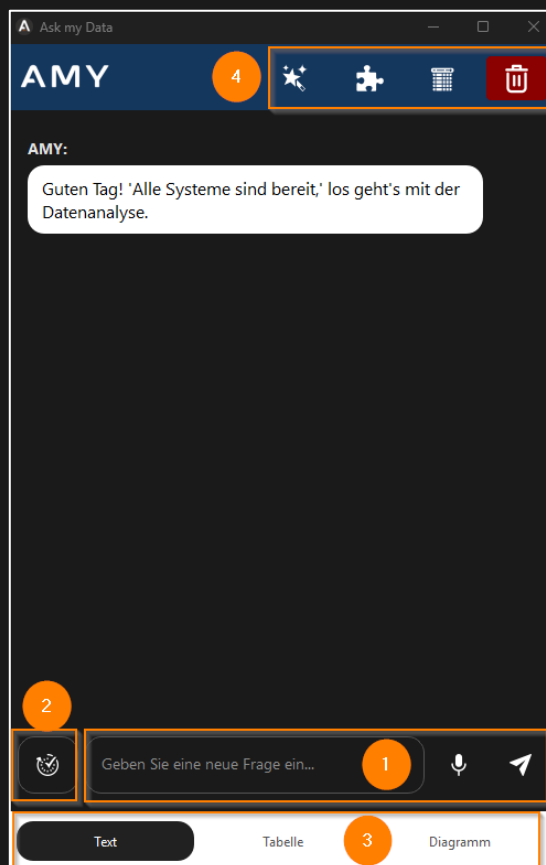


Unter **Ask my Data** steht Ihnen die KI-gestützte Assistentin (je nach gewählten Abomodel) AMY zur Verfügung, die Sie bei der Analyse und Aufbereitung Ihrer Daten unterstützt. In einem interaktiven Chatfenster können Sie auf natürliche Weise **Fragen stellen oder Befehle eingeben**

①, um beispielsweise bestimmte Daten zu filtern, Zusammenhänge zu erkennen oder Visualisierungen zu erstellen. Möchte man dass die KI auf die vorherigen Anfragen im Chat Zugriff hat und nicht immer als eine komplett neue Anfrage wertet, muss die **Chat Historie** ② aktiviert werden.

Darüber hinaus kann Amy auf Wunsch verschiedene Darstellungsformen für die Ergebnisse anbieten – etwa als **Textausgabe, Tabelle oder Diagramm** ③. So erhalten Sie schnell und unkompliziert Einblicke in Ihre Daten, können Ausreißer oder Muster erkennen und fundierte Entscheidungen treffen.

Im Headerbereich gibt es eine **eigene Tool-Box** ④ in der neben der Erstellung von Prompts, auch Spalten ausgewählt, eine Spaltenbeschreibung hinzugefügt oder Anfragen gelöscht werden können.



Mit ihrer nahtlosen Integration in den Analyzer erleichtert AMY den gesamten Arbeitsprozess und macht komplexe Analysen zugänglicher.

Beispiel:

Wenn AMY ein Liniendiagramm erstellen soll, sollte dies auch im Prompt beschrieben werden. Fehlt diese Angabe und erstellt AMY bspw. ein Balkendiagramm kann man auch die Anfrage nochmal mit dem Liniendiagramm verknüpfen und die Antwort wird geändert.

Hinweis: **AMY** interpretiert Ihre Eingaben automatisch und führt die passenden Aktionen durch, ohne dass Sie komplizierte Skripte schreiben oder externe Tools verwenden müssen. Ein intelligenter Algorithmus sorgt dafür, dass auch online basierte KI-Modelle wie von Open AI und DeepSeek uneingeschränkt genutzt werden können, da höchstens Strukturinformationen übersendet werden, nie aber die eigentlichen Daten.

10.3 AMY-Tool-Box

Im oberen Bereich der AMY-Ansicht finden Sie mehrere Symbole, die den Umgang mit Daten und Abfragen erleichtern.



Im Einzelnen stehen Ihnen folgende Funktionen zur Verfügung:

Magic Prompt Generator ①: Erzeugt automatisch passende Befehle oder Fragen für die KI, um Ihnen die Formulierung Ihrer Anfragen zu erleichtern.

Spaltenbeschreibung hinzufügen ②: Ermöglicht das Hinzufügen oder Bearbeiten von Beschreibungen für bestimmte Spalten, damit Sie schneller nachvollziehen können, welche Daten darin enthalten sind.

Bestimmte Spalte auswählen ③: Bietet die Möglichkeit, gezielt einzelne Spalten für die Analyse auszuwählen

Anfrage löschen ④: Bei Bedarf den Datensatz entfernen.

10.3.1 Magic Prompt Generator

Der **Prompt Generator** erlaubt es Ihnen, mithilfe weniger Eingaben eine kontextspezifische Analyse durchzuführen. Dafür stehen folgende Felder zur Verfügung:



- **System ①** (Pflichtfeld): Geben Sie hier das System oder die Umgebung an, in der die Daten verwendet oder verarbeitet werden (z. B. DATEV). Auf Basis dieser Information kann die KI das Vorgehen oder die Analyse speziell auf Ihr System zuschneiden.

- **Branche ①** (Pflichtfeld): Wählen Sie die Branche oder das Fachgebiet aus, auf das sich die Analyse beziehen soll (z. B. Einzelhandel). Dies hilft der KI, relevante Branchendaten oder -prozesse zu berücksichtigen.

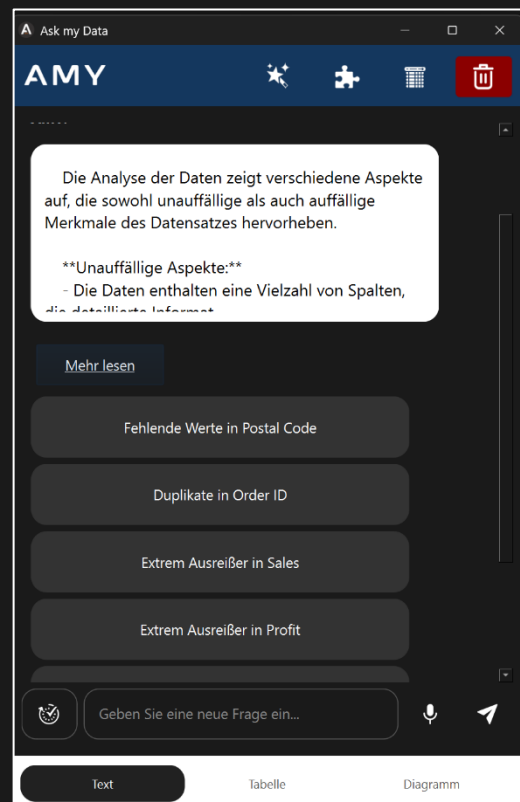
- **Besondere Überlegungen (optional) ②:** Fügen Sie zusätzliche Hinweise hinzu, etwa besondere Anforderungen, Einschränkungen oder Zielsetzungen, die die KI bei der Analyse berücksichtigen soll.

Aus diesen Angaben generiert die KI automatisch passende Schritte und Lösungen, die auf Ihren spezifischen Kontext zugeschnitten sind.

Beispielsergebnis:

Der Magic Prompt Generator hat beim Musterfall automatisch verschiedene Aspekte extrahiert und anschaulich zusammengefasst. Dabei wurde zum einen auf unauffällige Merkmale eingegangen, wie zum Beispiel vollständig ausgefüllte Felder, zum anderen aber auch auf **Auffälligkeiten** hingewiesen. Hierzu gehören unter anderem fehlende Werte im „Postal Code“, doppelte Einträge in „Order ID“ sowie extreme Ausreißer in „Sales“ und „Profit“. Durch diese automatisierte Zusammenstellung erhält man schnell einen Überblick über **potenzielle Probleme im Datensatz** und kann gezielt weiterführende Analysen oder Bereinigungen durchführen.

Die Ergebnisse lassen sich auch wieder speichern oder weiterleiten,



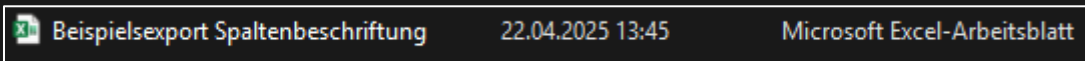
Für noch bessere Analysen sollten Sie im oberen Bereich des Magic Prompt-Fensters zusätzlich gezielt Spaltenbeschriftungen festlegen oder einzelne Spalten auswählen **3**, um bessere Ergebnisse zu erzielen (s. **10.3.1.1** ff). Eine fehlende Spaltenbeschriftung führt schnell zu fehlerhaften Ergebnissen.

10.3.1.1 Magic Prompt Spaltenbeschriftungen

Über **Spaltenbeschriftung** können Sie zu jeder ausgewählten Spalte zusätzliche Informationen oder Beschreibungen hinterlegen (eine Beschreibung pro Spalte nur möglich). Damit schaffen Sie mehr Transparenz für Ihre Datenanalyse, da alle Beteiligten erkennen können, welche Bedeutung oder welchen Zweck eine bestimmte Spalte erfüllt. Im gezeigten Dialog wählen Sie zunächst die gewünschte Spalte aus **1** und verfassen dann eine oder mehrere Beschreibungen **2**. Diese können Sie bei Bedarf über **Importieren** (z.B. eine Verfahrensdokumentation mit Beschreibung der einzelnen Spalten) oder **Exportieren** **3** austauschen oder sichern, um sie beispielsweise in anderen Projekten wiederzuverwenden.

Exportieren

Beim Export dieser Spaltenbeschriftung muss man den Speicherort des Exports vornehmen. Anschließend wird eine **Excel Datei** gespeichert:

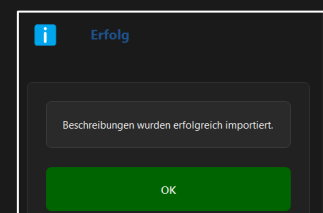


Es werden 2 Spalten in der Datei aufgeführt einmal Column (Spalte) und die Description (Beschreibung).

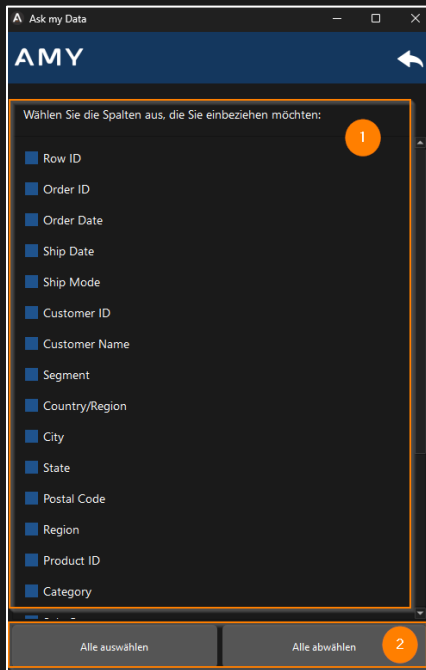
	A	B	C	D	E	F	G	H
1	Column	Description						
2	City	Beschreibt den Ort des Verkaufs						
3	Customer	Hier steht der Vorname und Nachname des Kunden drin						
4								

Importieren

Möchten Sie in einem Fall die Beschriftung importieren klicken Sie auf „**Importieren**“, wählen die Excel Datei aus ihrem Speicherort aus und erhalten anschließend das folgende Bestätigungsfenster:



10.3.1.2 Magic Prompt Spaltenauswahl



Über die **Spaltenauswahl** bestimmen Sie gezielt, welche Spalten ❶ in Ihre Analyse einbezogen werden sollen. Dadurch können Sie sich auf die für Ihre Fragestellung relevanten Daten konzentrieren und unnötige Informationen ausblenden. Mithilfe der Schaltflächen **Alle auswählen** und **Alle abwählen** ❷ lässt sich die gesamte Spaltenliste in einem Schritt aktivieren oder deaktivieren. Dies erleichtert die Verwaltung großer Datenbestände und sorgt für eine übersichtliche, fokussierte Analyse.

Durch diese zusätzlichen Einstellungen wird die KI-gestützte Analyse noch präziser, da sie exakt auf die von Ihnen definierten Datenfelder zugreift. Dies unterstützt Sie dabei, fundierte und zielgerichtete Ergebnisse zu erzielen.

10.3.2 Spaltenbeschreibung hinzufügen

Die **separate Spaltenbeschreibung** für die allgemeinen AMY-Fragen im klassischen Chatfenster, werden genauso behandelt wie unter 10.3.1.1 im **Magic Prompt Generator** beschrieben. Einziger Unterschied ist der Aufruf, der im Chatdialogfenster oben über das **Puzzlezeichen** geöffnet wird.



10.3.3 Bestimmte Spalten auswählen



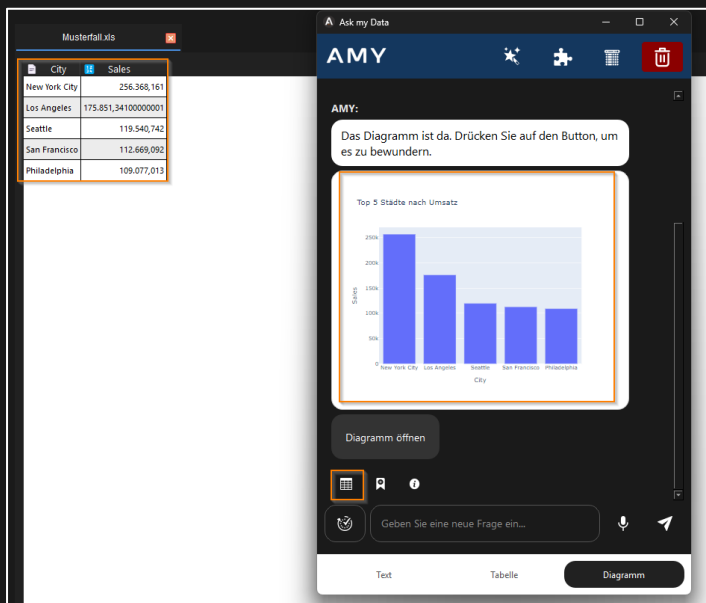
Diese Funktionalität unterscheidet sich inhaltlich nicht von der bereits unter 10.3.1.2 im **Magic Prompt Generator** beschriebenen Spaltenauswahl. Der einzige Unterschied liegt im **Aufruf**, der hier über das Dialogfenster unter dem **Tabellen-Symbol** erfolgt. Die einzelnen Schritte zum Auswählen und Verwenden bestimmter Spalten bleiben jedoch gleich und können analog zum zuvor beschriebenen Verfahren angewendet werden.

10.3.4 Anfrage löschen



Die Funktion **Anfrage löschen** leert das aktuelle Chatfenster, sodass keine zuvor gestellten Fragen oder Antworten mehr angezeigt werden. Auf diese Weise starten Sie mit einer leeren Oberfläche und können neue Anfragen an AMY stellen, ohne von vorherigen Konversationen abgelenkt zu werden.

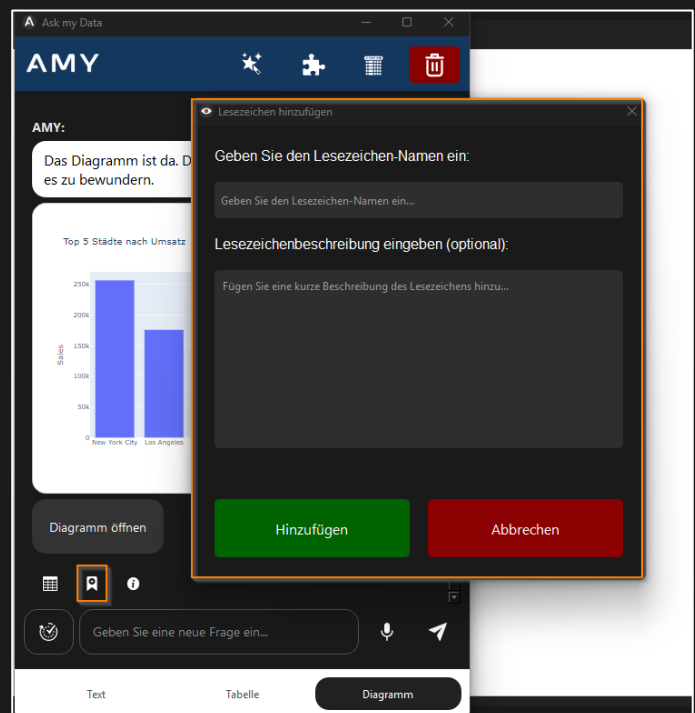
10.4 AMY- Tabelle im Analyzer



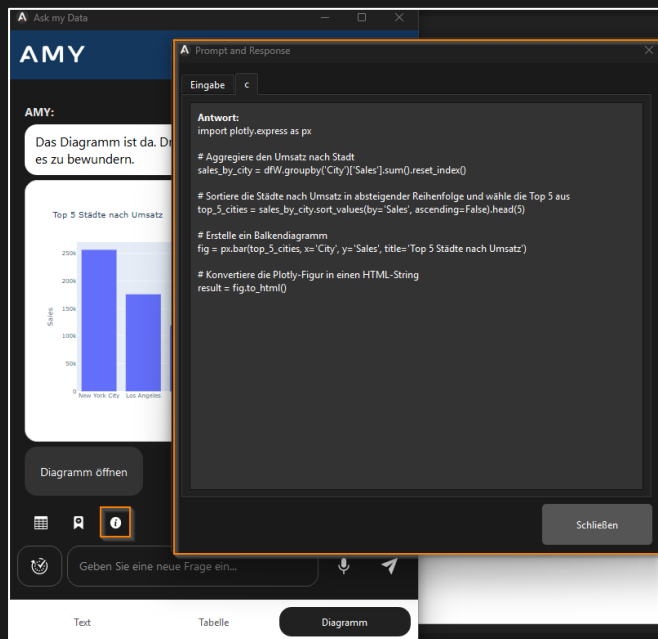
Über die **Tabellenfunktion** einer Abfrage in AMY haben Sie die Möglichkeit, die zugrunde liegenden Daten direkt im Analyzer anzuzeigen. Dabei werden automatisch die gleichen Filter- und Analyseparameter angewendet, die auch für das erzeugte Diagramm genutzt wurden. So können Sie auf einen Blick sowohl die grafische Darstellung als auch die entsprechenden Zahlenwerte einsehen und bei Bedarf unmittelbar weiterverarbeiten.

10.5 AMY – Lesezeichen

Mit einem Klick auf das **Lesezeichen-Icon** im AMY-Fenster können Sie die aktuelle Abfrage ganz einfach als Lesezeichen speichern. Daraufhin öffnet sich ein Dialog, in dem Sie einen Namen und eine optionale Beschreibung für das Lesezeichen angeben können. Auf diese Weise lassen sich Ihre KI-Abfragen dauerhaft **sichern und jederzeit** erneut aufrufen, ohne die ursprüngliche Formulierung oder die Ergebnisse neu eingeben zu müssen.



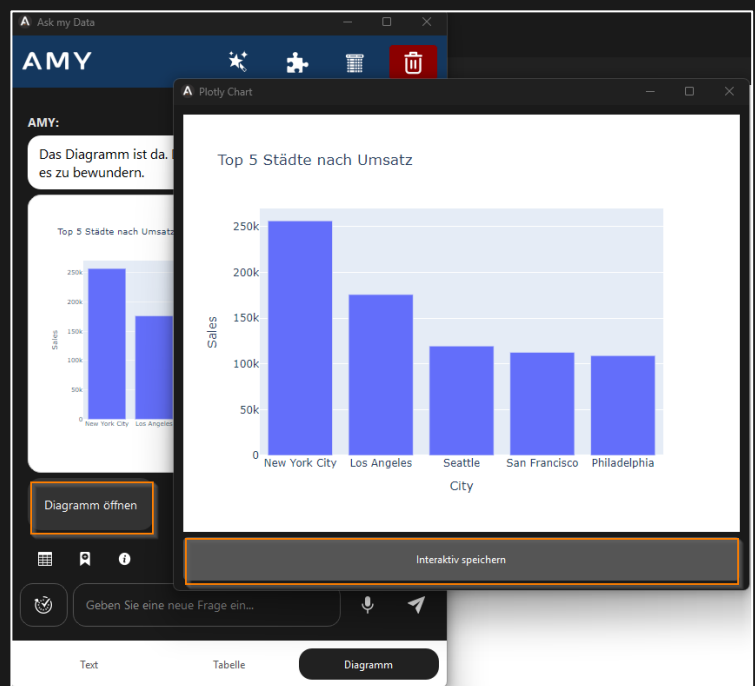
10.6 AMY- Dokumentation



Über das **Info-Symbol** erhalten Sie einen Einblick in die intern ausgeführten Schritte oder das generierte Skript, das für die jeweilige Aktion verwendet wurde. Wie im Beispiel ersichtlich, zeigt das Fenster beispielsweise den Python-Code, mit dem Daten aggregiert und visualisiert wurden. Auf diese Weise können Sie nachvollziehen, welche Befehle oder Logik der KI-Abfrage zugrunde liegen, und bei Bedarf selbst Anpassungen oder Erweiterungen vornehmen.

10.7 AMY-Interaktives Speichern

Über die Funktion **Interaktiv speichern** können Sie das angezeigte Diagramm als eigenständige HTML-Datei exportieren. Dadurch bleibt die Interaktivität erhalten, sodass Anwender – auch ohne Zugriff auf den Analyzer – mit dem Diagramm interagieren und beispielsweise Datenreihen filtern oder Werte hervorheben können. Dieses Format eignet sich hervorragend, um Ergebnisse mit Kollegen oder Kunden zu teilen, ohne dass zusätzliche Software benötigt wird.



11 Tab-Anzeige

Ist ein Fall geöffnet, wird der **Dateiname im Tabellenblatt** angezeigt. Bei mehreren gleichzeitig geöffneten Fällen erscheinen dort mehrere Dateien nebeneinander:

Musterfall.xls			
Row ID	Order ID	Order Date	
1	CA-2020-152156	08.11.2020	11.1
2	CA-2020-152156	08.11.2020	11.1

Musterfall.xls	Musterfall 2.xlsx

1

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
1	1	CA-2020-152156	08.11.2020	11.11.2020	Second Class	CG-12520

Durch einfaches Klicken auf einen der Dateinamen kann zwischen den zugehörigen Tabellen bequem gewechselt werden **①**. So ist ein schneller Zugriff auf unterschiedliche Fallinhalte ohne erneutes Öffnen der Dateien möglich.

Ebenso kann über die linke Seite (drei Punkte an der Seite)  der Dateixplorer (s. Kapitel 1.2) geöffnet werden, in welchem die jeweiligen Dateien enthalten sind (**Wichtig: sofern es sich um Parquet Dateien handelt**). Hier kann man zwischen den Fällen ebenfalls per Klick auf die linke Maustaste wechseln.

11.1 Verschieben von Tabellenblättern

Mit gedrückter linker Maustaste können die Tabellenblätter verschoben werden.

Musterfall 2.xls



Musterfall.xls



 Row ID	 Order ID	 Order Date	 Ship Date	 Ship Mode
1	CA-2020-152156	08.11.2020	11.11.2020	Second Class

11.2 Tab-Menü

Mit Rechtsklick auf einen Tab erhält man ein Menüfenster mit folgenden Funktionen:

Alle Tabs schließen und speichern
 Alle Tabs bis auf den offenen schließen und speichern
 Alle Tabs ohne Speichern schließen
 Alle Tabs bis auf den offenen ohne Speichern schließen

• Alle Tabs schließen und speichern

Schließt sämtliche geöffneten Tabs und speichert zuvor alle vorgenommenen Änderungen.

• Alle Tabs bis auf den offenen schließen und speichern

Speichert und schließt alle Tabs – mit Ausnahme des aktuell aktiven Tabs, der geöffnet bleibt.

- **Alle Tabs ohne Speichern schließen**

Schließt alle Tabs **ohne** vorheriges Speichern. Nicht gespeicherte Änderungen gehen dabei verloren.

- **Alle Tabs bis auf den offenen ohne Speichern schließen**

Schließt alle Tabs außer dem aktuell geöffneten, **ebenfalls ohne** vorheriges Speichern.

12 Datentabelle

Nach dem erfolgreichen Import der Daten, der Übersicht über die zentrale Steuerungsleiste (Header) sowie der Tabbereichen, steht nun die interaktive Datentabelle im Fokus. Sie bildet das Herzstück der Anwendung, da hier sämtliche importierten Datensätze in strukturierter Form angezeigt, gefiltert und vorbereitet werden können. Die Tabelle dient nicht nur der Übersicht, sondern ist auch Ausgangspunkt für viele Analysen und Visualisierungen.

In diesem Abschnitt wird erklärt, wie Sie sich in der Tabelle zurechtfinden, welche Interaktionsmöglichkeiten bestehen und wie Sie gezielt mit einzelnen Spalten oder Zeilen arbeiten können – etwa durch Sortierung, Filterung oder gezielte Auswahl von Werten.

12.1 Formatierungstypen Spalte

Mit **Rechtsklick auf eine Spalte** können verschiedene Formatierungen vorgenommen werden:

- Datum (+ Zeit)
- Zeit
- Nummer
- Zeichenfolge
- Boolean
- Link

12.1.1 Datum (+ Zeit) (Zeitstempel)

- Funktion:** Definiert eine Spalte als Datums- bzw. Zeitspalte (mit Datum und Uhrzeit). Im AUDIPY kann der aus Datum und Uhrzeit bestehen, muss es aber nicht. Es kann auch ein Datum ohne Uhrzeit verwendet werden, daher ist Zeit auch in () gestellt. Uhrzeiten mit 00:00:00 werden ausgeblendet.
- Verwendung:** Diese Einstellung stellt sicher, dass Werte in der Spalte als kombinierte Datums- und Zeitangaben interpretiert werden. Dadurch kann das System entsprechende Filter, Sortierungen oder Formatierungen speziell für Datum und Uhrzeit anwenden.

Mit Klick auf Datum (+Zeit) öffnet sich ein weiteres Dialogfenster, in welchem verschiedene Einstellungen ausgewählt werden können:

Original	Formatiert
2020-11-08 00:00:00	11.08.2020
2020-11-08 00:00:00	11.08.2020
2020-06-12 00:00:00	06.12.2020
2019-10-11 00:00:00	10.11.2019
2019-10-11 00:00:00	10.11.2019
2018-06-09 00:00:00	06.09.2018
2018-06-09 00:00:00	06.09.2018
2018-06-09 00:00:00	06.09.2018
2018-06-09 00:00:00	06.09.2018
2018-06-09 00:00:00	06.09.2018

Automatische Erkennung ①:

Wenn diese Option aktiviert ist, erkennt das System das Datumsformat automatisch anhand der geladenen Daten. Dies ist hilfreich, wenn die Daten in einem klaren,

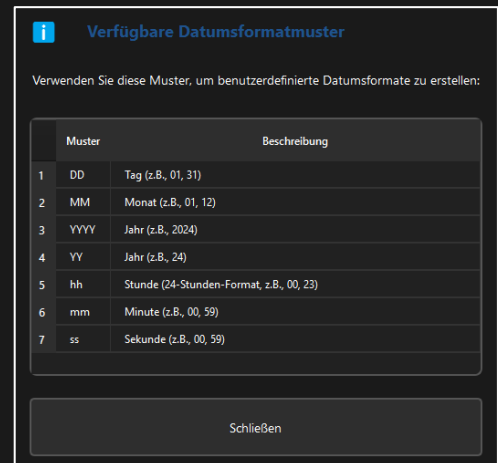
einheitlichen Format vorliegen (z. B. YYYY-MM-DD).
Empfohlen bei gängigen Datumsformaten.

Benutzerdefiniertes Format ②:

Falls die automatische Erkennung fehlschlägt oder ein spezielles Format gewünscht ist, kann hier ein eigenes Formatmuster eingegeben werden. Beispiele sind:

- DD.MM.YYYY → 08.11.2020
- MM/DD/YYYY → 11/08/2020
- YYYY/MM/DD → 2020/11/08

 Der Button „Info“ liefert unterstützende Hinweise zu den möglichen Formatbestandteilen.



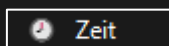
Vorschaufenster ③:

Im unteren Bereich wird in einer Tabelle die Auswirkung der Formatwahl gezeigt:

- **Original:** Das ursprüngliche Format aus den Quelldaten.
- **Formatiert:** Die umgewandelte Darstellung basierend auf der aktuellen Einstellung.

Diese Vorschau ist besonders nützlich, um Fehler frühzeitig zu erkennen – etwa ein vertauschtes Tag-Monat-Jahr-Muster.

12.1.2 Zeit



- **Funktion:** Definiert eine Spalte als reine Zeitspalte (ohne Datum). Diese Funktion ist nur in bestimmten Ausnahmefällen sinnvoll – etwa wenn ausschließlich Uhrzeiten analysiert werden sollen, unabhängig vom Datum.

Hinweis: In den meisten Anwendungsfällen ist es empfehlenswert, stattdessen einen vollständigen **Zeitstempel** (siehe Abschnitt 4.1.1) zu verwenden, da dieser sowohl Datum als auch Uhrzeit enthält und damit deutlich mehr Kontext für Analysen bietet.

- **Verwendung:** Wird genutzt, wenn nur Uhrzeitwerte gespeichert werden sollen. Das System kann so Zeitangaben passend formatieren oder sortieren (z. B. 08:00 < 09:00).

Mit Klick auf Zeit öffnet sich ein weiteres Dialogfenster:

Zeitformatoptionen

Uhrzeitformateinstellungen

☒ Automatisch

☐ Benutzerdefiniert Formatmuster:

Vorschau (Vorher & Nachher)

Original	Formatiert
2020-11-08 00:00:00	
2020-11-08 00:00:00	
2020-06-12 00:00:00	
2019-10-11 00:00:00	
2019-10-11 00:00:00	
2018-06-09 00:00:00	
2018-06-09 00:00:00	
2018-06-09 00:00:00	
2018-06-09 00:00:00	
2018-06-09 00:00:00	
2018-06-09 00:00:00	

Automatische Erkennung ①:

Wählt automatisch das passende Zeitformat basierend auf den vorhandenen Daten. Typische Formate wie HH:MM:SS werden dabei automatisch erkannt. *Empfohlen, wenn Uhrzeitangaben konsistent formatiert sind.*

Benutzerdefiniertes Format ②:

Hier kann ein individuelles Zeitformat definiert werden – zum Beispiel:

- HH:mm → 13:45
- HH:mm:ss → 13:45:00
- hh:mm:ss tt → 01:45:00 PM

Info

Das Eingabefeld wird durch den **Info**-Button unterstützt, der Hinweise zu den verfügbaren Formatoptionen gibt.

Verfügbare Uhrzeitformatmuster

Verwenden Sie diese Muster, um benutzerdefinierte Uhrzeitformate zu erstellen:

	Muster	Beschreibung
1	HH	Stunde (24-Stunden-Format, z.B., 00, 23)
2	hh	Stunde (12-Stunden-Format, z. B. 01, 12)
3	MM	Minute (z.B., 00, 59)
4	SS	Sekunde (z.B., 00, 59)
5	AM/PM	Meridiem-Indikator (z. B. AM, PM)
6	hh:MM	Stunde und Minute (z. B. 13:45)
7	hh:MM:SS	Stunde, Minute und Sekunde (z. B. 13:45:30)

Schließen

Vorschaufenster ③:

Im unteren Bereich wird in einer Tabelle die Auswirkung der Formatwahl gezeigt:

- **Original:** Das ursprüngliche Format aus den Quelldaten.
- **Formatiert:** Die umgewandelte Darstellung basierend auf der aktuellen Einstellung.

Diese Vorschau ist besonders nützlich, um Fehler frühzeitig zu erkennen.

12.1.3 Nummer

Nummer

- **Funktion:** Legt eine Spalte als numerisch fest. Ohne Formatierung werden Zahlen immer mit allen Stellen nach dem Komma angezeigt.
- **Verwendung:** Nützlich, wenn in der Spalte nur Zahlenwerte (z. B. ganze Zahlen, Dezimalwerte) hinterlegt werden. Ermöglicht eine numerische Sortierung und ggf. Berechnungen.

Ganzzahl ①:

Diese Option erlaubt die Auswahl von **Ganzzahlen** als Format. Das bedeutet, dass alle Zahlen ohne Nachkommastellen angezeigt werden.

Optionen für numerische Formate

Einstellungen für Zahlenformate

☒ Ganzzahl

☐ Dezimalzahl Dezimalstellen: 2

Regionale Formatierung EU

Vorschau (Vorher & Nachher)

Original	Formatiert
261.96	261
731.9399999999999	731
14.62	14
957.5775	957
22.368000000000002	22
48.86	48
7.28	7
907.152	907
18.504	18
114.9	114

Format anwenden Abbrechen

Dezimalzahl ②:

Diese Option erlaubt die Verwendung von **Dezimalzahlen**. Rechts daneben befindet sich ein Dropdown, mit dem die **Anzahl der Dezimalstellen** ausgewählt werden kann (z. B. 2 Dezimalstellen → „1,23“). Voreingestellt sind 2, sobald man hier z.B. 3 wählt wird dies automatisch unten im Vorschaufenster ④ angezeigt.

Format ③:

Hier kann die Auswahl des **Zahlenformats nach Europa (EU)oder USA (US)** getroffen werden.

- **Dezimaltrennzeichen:** Die amerikanische Schreibweise ist mit **Punkt**, die europäische mit **Komma**.
- **Tausendertrennzeichen:** Europäisch mit **Punkt** – amerikanisch mit **Komma**

Beispiel	Dezimaltrennzeichen / Tausendertrennzeichen
Europa	10.000,00
Amerika	10,000.00

Vorschaufenster ④:

Im unteren Bereich wird in einer Tabelle die Auswirkung der Formatwahl gezeigt:

- Original: Das ursprüngliche Format aus den Quelldaten.
- Formatiert: Die umgewandelte Darstellung basierend auf der aktuellen Einstellung.

Diese Vorschau ist besonders nützlich, um Fehler frühzeitig zu erkennen.

12.1.4 Zeichenfolge

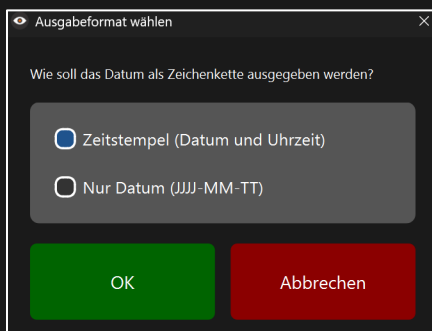
Zeichenfolge

- **Funktion:** Markiert die Spalte als Textspalte.
- **Verwendung:** Hier können beliebige Zeichenketten gespeichert werden. Sortierung und Filterung erfolgen dann als Text, z. B. alphabetisch.

Wenn eine Spalte in ein Zeichen umgewandelt wird und in den Spalten nicht alle Zeilen ausgefüllt sind, werden folgende Abkürzungen eingefügt:

- **NA (Not Available)**
Kennzeichnet beliebige fehlende oder nicht verfügbare Werte (Text, Zahl, Datum).
- **NaN (Not a Number)**
Speziell für undefinierte oder fehlende **numerische** Werte (z. B. infolge ungültiger Rechenoperationen).
- **NaT (Not a Time)**
Markiert fehlende oder ungültige **Datums-/Zeitangaben** in Timestamp-Spalten.

Hinweis: Möchte man aus einer **Zeitstempel-Spalte** eine Zeichenfolge machen, erscheint ein weiteres Zwischenfenster:



Dies hat den Zweck sicherzustellen, dass Datums- und Zeitangaben konsistent und im gewünschten Format als Zeichenkette ausgegeben werden. Gewählt werden kann zwischen Datum und Uhrzeit als Text oder nur Datum.

12.1.5 Boolean

Boolean

- **Funktion:** Bestimmt, dass in dieser Spalte nur Wahr-/Falsch-Werte (Ja/Nein, True/False) gespeichert werden.
- **Verwendung:** Zeigt immer Checkboxes an oder filtert nur nach „True“ oder „False“.

12.1.6 *Link*



- **Funktion:** Stellt eine Spalte für Links bzw. URLs bereit (setzt eine ordnungsgemäße Schreibweise voraus)
- **Verwendung:** Speichert klickbare Links, die beim Anklicken den Standardbrowser öffnen oder eine Aktion ausführen.

12.2 Sortierung

Ebenfalls mit **Rechtsklick** kann auch eine einfache Sortierung vorgenommen werden:

Aufsteigend sortieren
Absteigend sortieren
Sortierung zurücksetzen

Hinweis:

Für die Sortierung der Einträge steht ein eigenes Tool in den Einstellungen des AUDIPYs zur Verfügung (S. Kapitel **2.2.9.3**) Hier können Sie festlegen, ob eine manuelle Sortierung im Verlauf gespeichert werden soll. Standardmäßig werden manuelle Sortierungen *nicht* im Verlauf gespeichert. Sortierungen, die über das Tool vorgenommen werden, *werden* hingegen automatisch gespeichert.

12.2.1 *Aufsteigend sortieren*

- **Funktion:** Sortiert die Werte in der ausgewählten Spalte in aufsteigender Reihenfolge (z. B. A–Z oder 1–9).
- **Verwendung:** Nützlich, um schnell eine Übersicht von klein nach groß oder von A bis Z zu erhalten.

12.2.2 *Absteigend sortieren*

- **Funktion:** Sortiert die Werte in der ausgewählten Spalte in absteigender Reihenfolge (z. B. Z–A oder 9–1).
- **Verwendung:** Verschafft einen schnellen Überblick von groß nach klein oder von Z bis A.

12.2.3 *Sortierung zurücksetzen*

- **Funktion:** Hebt alle aktiven Sortierungen auf und stellt die ursprüngliche Reihenfolge wieder her.
- **Verwendung:** Praktisch, wenn man die Liste wieder in den Zustand bringen möchte, bevor Sortierungen angewendet wurden.

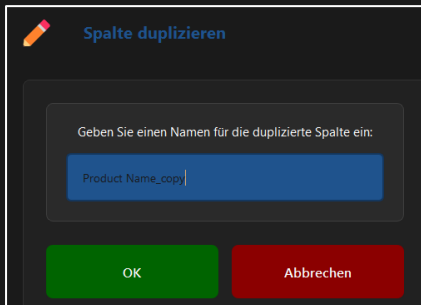
12.3 Spaltenaktionen

Über **Rechtsklick auf den Spaltennamen** lassen sich auch einfache Spaltenaktionen durchführen

Spalte duplizieren
Spalte umbenennen
Spalte löschen
Spalte ausblenden
Alle Spalten einblenden
Spalten neu anordnen
Beschreibung

12.3.1 Spalte duplizieren

- **Funktion:** Erstellt eine Kopie der aktuell ausgewählten Spalte.
- **Verwendung:** Praktisch, um vorhandene Daten zu sichern oder alternative Bearbeitungen an einer Kopie vorzunehmen.



Spalte duplizieren

Geben Sie einen Namen für die duplizierte Spalte ein:

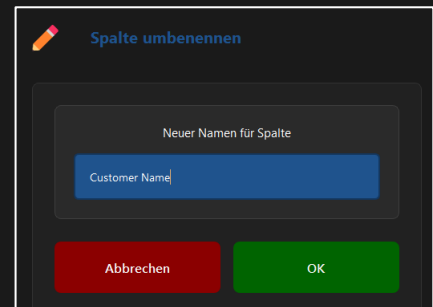
Product Name_copy

OK Abbrechen

Hierfür muss im die gewünschte zu duplizierende Spalte ausgewählt und mit einem neuen Namen versehen werden. Anschließend mit „OK“ bestätigen und die Spalte wird am Ende der Tabelle eingefügt.

12.3.2 Spalte umbenennen

- **Funktion:** Ermöglicht das Ändern des Spaltentitels.
- **Verwendung:** Nützlich, um eine aussagekräftigere Bezeichnung für die Daten in dieser Spalte zu vergeben.



Spalte umbenennen

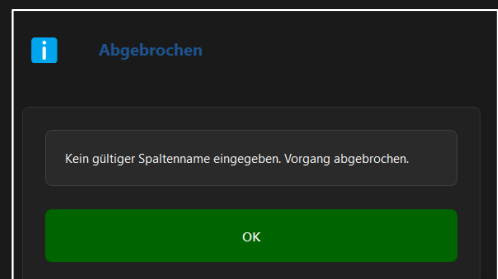
Neuer Namen für Spalte

Customer Name

Abbrechen OK

Hinweis: Mit einem Doppelklick auf den Spaltennamen lässt sich der Name der Spalte ebenfalls ändern.

Falls kein gültiger Spaltenname eingegeben wird, erscheint folgendes Hinweisfenster:



Abgebrochen

Kein gültiger Spaltenname eingegeben. Vorgang abgebrochen.

OK

12.3.3 Spalte löschen

- **Funktion:** Entfernt die ausgewählte Spalte (inklusive aller zugehörigen Daten).
- **Verwendung:** Achtung! Nach dem Löschen können die Spaltenwerte über die Funktion „Zurück“ (Kapitel 6) wiederhergestellt werden.

12.3.4 Spalte ausblenden

- **Funktion:** Blendet die ausgewählte Spalte aus, ohne sie zu löschen.
- **Verwendung:** Nützlich, um temporär bestimmte Spalten zu verbergen, die gerade nicht relevant sind, ohne die Daten zu verlieren.

12.3.5 Alle Spalten einblenden

- **Funktion:** Zeigt alle zuvor ausgeblendeten Spalten wieder an.
- **Verwendung:** Stellt schnell den vollen Tabellenumfang wieder her, wenn mehrere Spalten ausgeblendet wurden.

12.3.6 Spalten neu anordnen

- **Funktion:** Die Spaltenreihenfolge kann individuell angepasst werden (z.B. erste Spalte nach ganz hinten, zweite Spalte als erste usw.).
- **Verwendung:** Nach Betätigung der Schaltfläche öffnet sich ein neues Fenster mit allen Spaltennamen. Dies kann durch die Pfeile auf der rechten Seite beliebig verändert werden.



Die Spalten lassen sich entweder über Drag & Drop an die gewünschte Stelle platzieren ① oder man bedient sich den Schaltflächen auf der Seite. Dabei sind die Nummer 2 und 4 vereinfachte Buttons, welche die Spalte entweder an direkt an den Anfang ② oder direkt ans Ende setzen ④. Die Pfeile unter ③ hingegen verschieben die gewünschte Zeile nach oben oder unten.

12.3.7 Beschreibung

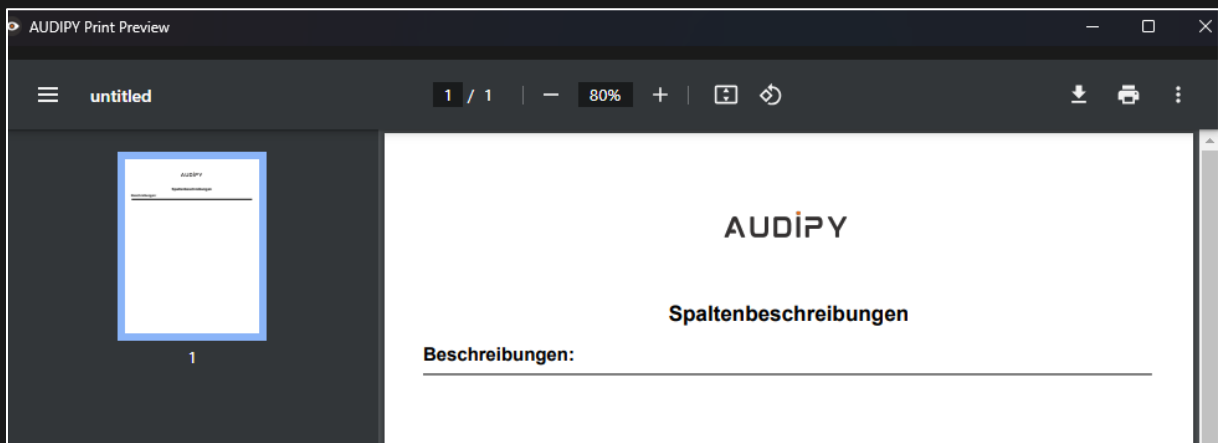
Mit diesem Dialog können Sie die Beschreibungen von Spalten in Ihrer Anwendung ansehen, bearbeiten, importieren, exportieren und verwalten. Auf diese Weise stellen Sie sicher, dass die Spalten Ihrer Datenbank oder Tabelle für alle Nutzer nachvollziehbar dokumentiert sind und Missverständnisse über deren Inhalt vermieden werden. Darüber hinaus hilft es auch den AI Funktionen (hierfür ist ein Abo notwendig), die Daten besser zu verarbeiten.



Spaltenauswahl ①

Im oberen Bereich des Dialogs wählen Sie die gewünschte Spalte aus.

- Das Dropdown-Feld (oder die Liste) zeigt die verfügbaren Spalten an, z.B. „Order Date“.
- Darunter befinden sich drei Schaltflächen:
 - **Importieren:** Lädt Beschreibungen aus einer Datei.
 - **Exportieren:** Speichert die aktuellen Beschreibungen in einer Datei.
 - **Alle anzeigen:** Zeigt die Beschreibungen aller Spalten an. Hierzu öffnet sich ein Vorschaufenster, in welchem dann alle Spaltenbeschriftungen angezeigt werden.



Beschreibung bearbeiten 2

Im mittleren Bereich sehen Sie die aktuelle Beschreibung der ausgewählten Spalte.

- In dem großen Textfeld können Sie die Beschreibung anpassen oder eine neue Beschreibung eingeben.

12.4 Datentabelle: Interaktion

Innerhalb der Tabelle können in den einzelnen Zeilen verschiedene Funktionen ausgewählt werden, um die Daten zu filtern, im Internet über Google zu suchen, zu übersetzen oder zu bereinigen.

12.4.1 Inhalt kopieren

Mit **Rechtsklick auf eine Zeile** kann ein Inhalt kopiert werden..

- **Funktion:** Kopiert den Inhalt der aktuell ausgewählten Zelle in die Zwischenablage.
- **Verwendung:** Ermöglicht das schnelle Übertragen von Zellinhalten in andere Anwendungen (z. B. Texteditor, Excel)

Inhalt kopieren

Nach ausgewähltem Inhalt filtern

Mehrere Werte filtern

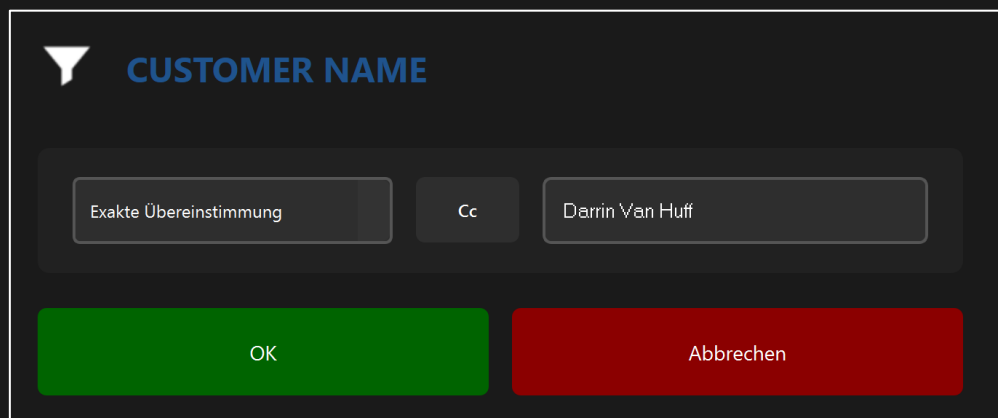
Filter löschen

12.4.2 Filtern

Mit **Rechtsklick auf eine Zeile** können verschiedene Filterungen durchgeführt werden.

12.4.2.1 Nach ausgewähltem Inhalt filtern

- **Funktion:** Setzt einen Filter in der Tabelle, der nur Zeilen anzeigt, die exakt zum ausgewählten Inhalt passen.
- **Verwendung:** Ideal, um schnell alle Zeilen zu finden, die denselben Wert wie die ausgewählte Zelle enthalten.



CUSTOMER NAME

Exakte Übereinstimmung Cc Darrin Van Huff

OK Abbrechen

Der Filterdialog ermöglicht eine gezielte Suche und Einschränkung der Datensätze in der Tabelle. Die Bedienung erfolgt über vier zentrale Steuerelemente:

- **1. Spaltenname**

Der angezeigte Titel (z.B. *Customer Name*) entspricht der Spalte in der Tabelle, auf die der Filter angewendet wird. Diese Beschriftung wird dynamisch gesetzt, abhängig davon, welche Spalte aktuell gefiltert werden soll.

- **2. Vergleichsoperator / Filtertyp**

Je nach Datentyp der Spalte (z.B. Text, Zahl, Datum) erscheinen hier unterschiedliche Vergleichsoperatoren.

Spaltenformat	Verfügbare Filterfunktionen
Text	- Exakte Übereinstimmung - Teilweise Übereinstimmung - Stimmt nicht überein - Enthält nicht - Ungefähre Übereinstimmung
Numerisch	- Exakte Übereinstimmung - Stimmt nicht überein - Größer als - Kleiner als - Größer oder gleich - Kleiner oder gleich
Datum	- Exakte Übereinstimmung - Stimmt nicht überein - Vor - Nach - Am oder vor - Am oder nach
Zeit	- Exakte Übereinstimmung - Stimmt nicht überein - Vor - Nach - Am oder vor - Am oder nach
Boolean	- Exakte Übereinstimmung - Stimmt nicht überein

- **3. Zusatzoptionen / Modifikatoren**

In diesem Feld werden zusätzliche Filteroptionen angezeigt, abhängig vom Spaltentyp.

Datentyp	Modifikator	Funktion / Bedeutung
Text	Cc	Groß-/Kleinschreibung umschalten
Numerisch	Abs	Umschalten auf absolute Werte
Datum	YY	Suche auf Jahresangabe beschränken
Zeit	hh	Suche nach Stunden aktivieren
Boolean	Cc	Groß-/Kleinschreibung (wie bei Textfeldern)

- **4. Filterwert**

Hier wird der **eigentliche Vergleichswert** eingetragen, nach dem gefiltert werden soll – z.B. ein Name, ein Datum oder eine Zahl. Der eingegebene Wert wird mit dem ausgewählten Filtertyp (Punkt 2) verglichen.

Nach der Filterung erhalten Sie ein neues Fenster:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	
6	CA-2018-115812	09.06.2018		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los #
7	CA-2018-115812	09.06.2018		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los #
8	CA-2018-115812	09.06.2018		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los #
9	CA-2018-115812	09.06.2018		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los #
10	CA-2018-115812	09.06.2018		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los #
11	CA-2018-115812	09.06.2018		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los #
12	CA-2018-115812	09.06.2018		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los #
2.844	CA-2021-112865	18.05.2021		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Phila
3.091	CA-2020-112123	03.03.2020		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	New
3.092	CA-2020-112123	03.03.2020		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	New
3.093	CA-2020-112123	03.03.2020		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	New
4.910	CA-2021-127306	14.01.2021		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	John
4.911	CA-2021-127306	14.01.2021		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	John
4.912	CA-2021-127306	14.01.2021		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	John
4.913	CA-2021-127306	14.01.2021		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	John
4.914	CA-2021-127306	14.01.2021		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	John
5.759	CA-2019-149300	22.11.2019		First Class	BH-11710	Brosina Hoffman	Consumer	United States	Hiale
7.612	US-2019-130491	08.02.2019		First Class	BH-11710	Brosina Hoffman	Consumer	United States	Gard
7.613	US-2019-130491	08.02.2019		First Class	BH-11710	Brosina Hoffman	Consumer	United States	Gard
7.614	US-2019-130491	08.02.2019		First Class	BH-11710	Brosina Hoffman	Consumer	United States	Gard
7.855	CA-2021-135587	07.12.2021		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Hatt
7.856	CA-2021-135587	07.12.2021		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Hatt
9.748	US-2018-140914	11.11.2018		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Chic
9.749	US-2018-140914	11.11.2018		Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Chic

Customer Name

Exakte Übereinstimmung

Cc

Brosina Hoffman

✓

✗

Σ

In der **unteren Aktionsleiste** (s. Kapitel 13) wird nun die Filterung angezeigt und man kann entscheiden was jetzt weiter passieren soll. Soll das Fenster wie geschlossen werden kann man über das **rote X** die Filterungen wieder löschen. Man kann jedoch auch noch eine weitere Filterung auf die angezeigten Daten vornehmen.

Möchte man – wie im gezeigten Beispiel – zusätzlich zum Filter auf den Customer Name „Brosina Hoffmann“ auch den Ship Mode „Standard Class“ berücksichtigen, aktiviert man einfach das entsprechende Häkchen.

Es erscheint dann folgendes Fenster:

So können beliebig viele Filterungen hinzugefügt und gelöscht werden. Standardmäßig ist eine doppelte Filterung über „und“ eingestellt. Mit Klick auf „und“ ① ermöglicht es auch eine Filterung nach „oder“

12.4.2.2 Mehrere Werte filtern

- **Funktion:** Erlaubt das Filtern nach mehreren, frei wählbaren Werten in derselben Spalte.
- **Verwendung:** Praktisch, um eine Liste nach mehreren Kriterien zu durchsuchen (z. B. mehrere Namen oder IDs).

Hierzu öffnet sich ein weiteres Programmfenster

1 Übereinstimmungstyp: Legen Sie fest, ob die Auswahl exakt oder gar nicht übereinstimmen soll. Auch ein ungefähre Übereinstimmung (sog. Fuzzy Methode ist möglich) welche unter **12.4.2.2.1** genauer erläutert wird, da sich hier noch ein weiteres Dialogfenster öffnet.

2 Sortierung: Bestimmen Sie, in welcher Reihenfolge die verfügbaren Werte angezeigt werden, beispielsweise alphabetisch aufsteigend oder absteigend (oder Anzahl absteigend / aufsteigend)

Unter **3** erhalten Sie das Vorschauenfenster welche zusätzlichen Optionen erlaubt:

Suchfeld: Geben Sie einen Suchbegriff ein, um die angezeigten Werte einzugrenzen.

Werteliste: Aktivieren oder deaktivieren Sie die gewünschten Einträge über Kontrollkästchen.

Alle auswählen / Alle abwählen: Setzen oder entfernen Sie die Markierung für sämtliche angezeigten Einträge in einem Schritt.

Wert hinzufügen: Ermöglicht es, manuell einen neuen Wert in die Filterliste aufzunehmen.

Eindeutige Werte in Postal Code

Übereinstimmungstyp: Exakte Übereinstimmung

Sortieren nach: Alphabetisch aufsteigend

Suchen: Tippen um Werte zu filtern...

Verfügbare Werte

Alle auswählen Alle abwählen Wert hinzufügen

☐ 10009.0 (229)

☐ 10011.0 (193)

☐ 10024.0 (230)

☐ 10035.0 (263)

☐ 1040.0 (1)

☐ 10550.0 (8)

OK Abbrechen

12.4.2.2.1 Fuzzy-Methode

Die Fuzzy-Suche ermöglicht es, Datensätze auch dann zu finden, wenn die Suchbegriffe nicht exakt mit den vorhandenen Werten übereinstimmen. Sie ist besonders hilfreich bei Tippfehlern, Namensvarianten oder unterschiedlichen Schreibweisen.

Funktionsweise:

Im gezeigten Dialogfenster kann der Benutzer die Fuzzy-Suche auf eindeutige Werte einer Spalte (z.B. „Customer Name“) anwenden. Dabei stehen folgende Optionen zur Verfügung:

- **Fuzzy-Suchbegriffe:**

Im blauen Eingabefeld können mehrere Begriffe eingegeben werden – **jeweils einer pro Zeile**. Jeder Begriff wird durch Drücken der Enter-Taste hinzugefügt.

Die Suche wird dann mit Hilfe eines Fuzzy-Matching-Algorithmus durchgeführt (z.B. Levenshtein-Distanz), um auch ähnlich geschriebene Namen zu erkennen.

Beispiel:

Wenn ein Benutzer Sean O Donnell eingibt, könnten auch Varianten wie Sean O'Donnell, S. O Donnell oder Sean Odonnel gefunden werden.

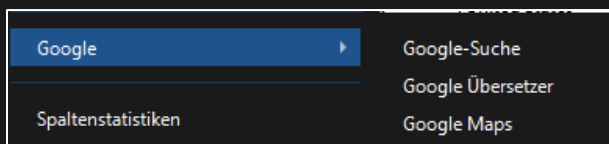
12.4.2.3 Filter löschen

- **Funktion:** Hebt alle aktiven Filter auf und zeigt wieder alle Zeilen in der Tabelle an.
- **Verwendung:** Stellt den ursprünglichen Tabellenzustand her, bevor Filter angewendet wurden.

Hinweis: Dies kann auch über die Aktionsleiste erfolgen, sofern man gerade erst gefiltert hat (s. auch **12.4.2.1** und Kapitel **13.1.4**)

12.4.3 Google

Über Rechtsklick und anschließend über Google erhält man zusätzliche Hilfe:

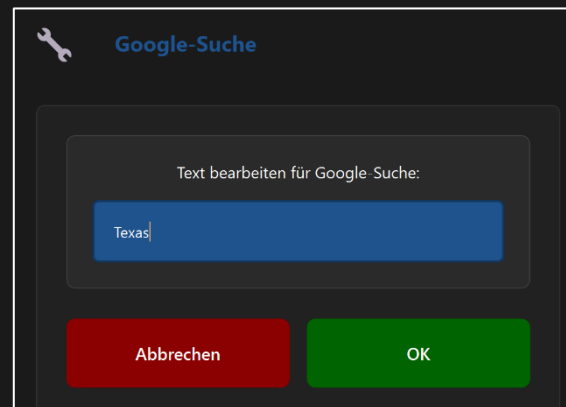


- **Google-Suche:** Öffnet Ihren Webbrowser und führt eine Google-Suche mit dem ausgewählten Inhalt durch.

- **Google Übersetzer:** Übernimmt den ausgewählten Text und öffnet den Google Übersetzer, um den Inhalt automatisch zu übersetzen
- **Google Maps:** Verwendet den ausgewählten Inhalt (z.B. eine Adresse) und öffnet Google Maps, um den entsprechenden Ort anzuzeigen.

Hinweis:

Alle **3 Google-Optionen** haben aufgrund des Datenschutzes ein Vorschaufenster, in welchem der ausgewählte Text aus der Zeile noch geändert werden kann.



12.4.4 Spaltenstatistiken

Mit Rechtsklick auf **Spaltenstatistik** wird ein neues Fenster geöffnet, in dem Ihnen verschiedene Kennzahlen und Auswertungen zu den Daten in der ausgewählten Spalte angezeigt werden. Sinn und Zweck dieser Funktion ist es, schnell einen Überblick über wichtige statistische Informationen zu erhalten.

Wichtig: Je nach Spaltentyp (Text, Numerisch, Datum (+Zeit), s. Kapitel **12.1**) verändern sich auch die angezeigten Statistiken. In den nachfolgenden Kapiteln erhalten Sie so die unterschiedlichen Auswertungsmethoden.

12.4.4.1 Spaltenstatistik für Textspalte

Bei einer Zeichen- bzw. Textspalte erscheint wird folgendes Fenster angezeigt:

Spaltenauswahl

Kategoriespalte: Customer Name

Aggregation: Anzahl

Statistische Zusammenfassung

Statistik	Wert
Anzahl (nicht-NaN)	9.994
Einzigartig	793
Nullwerte	0
Durchschnittliche Länge	12,96
Maximale Länge	22
Minimale Länge	7

Datenvisualisierung - Top 10

Name	Wert
William Brown	37
Matt Abelman	34
John Lee	34
Paul Prost	34
Jonathan Doherty	32
Chloris Kastensmidt	32
Seth Vernon	32
Edward Hooks	32
Emily Phan	31
Zuschuss Carroll	31

Schließen

Spaltenauswahl ①:

In diesem Bereich können Sie die gewünschte Spalte auswählen, für die Sie Statistiken anzeigen lassen möchten. Dadurch müssen Sie nicht erst in der Haupttabelle auf eine Spalte klicken, sondern können direkt im Statistikfenster zwischen verschiedenen Spalten wechseln. Darüber hinaus ist es möglich den Aggregationszustand (Anzahl von Werten oder Summen) zu ändern.

Bei Summen erscheint ein weiteres Auswahlfenster, bei welchem für die Summe die **numerische Spalte** ausgewählt werden kann:

Aggregation: Summe

Numerisches Feld: Sales

Statistische Zusammenfassung ②:

Hier werden verschiedene Kennzahlen zur aktuell ausgewählten Spalte angezeigt:

- **Anzahl (nicht-NaN):** Gibt die Zahl aller vorhandenen Werte an (ohne fehlende Werte).
- **Einzigartig:** Zeigt an, wie viele unterschiedliche Werte es innerhalb der Spalte gibt. Dies ist besonders nützlich, um z.B. Kategorisierungen oder Gruppierungen zu erkennen.

- **Nullwerte:** Gibt an, wie viele Einträge fehlen oder leer sind. Dies hilft dabei, Datenqualitätsprobleme schnell zu identifizieren.
- **Durchschnittliche Länge:** Die mittlere Anzahl der Zeichen pro Eintrag (nur relevant bei Textfeldern).
- **Maximale Länge:** Die höchste gemessene Zeichenanzahl eines Eintrags.
- **Minimale Länge:** Die kleinste gemessene Zeichenanzahl eines Eintrags. Diese Statistiken unterstützen Sie dabei, Auffälligkeiten, Inkonsistenzen oder Besonderheiten in Ihren Daten zu erkennen.

Datenvisualisierung Top 10 3:

In diesem Diagramm wird die Häufigkeit der Werte in der ausgewählten Spalte grafisch dargestellt. Es visualisiert die Verteilung der Einträge und ermöglicht so einen schnellen Überblick darüber, welche Werte besonders häufig vorkommen und wie gleichmäßig die Verteilung insgesamt ist.

Hinweis:

Für Spalten vom Typ **Boolean** (wahr/falsch) sowie für **Links** (z.B. URLs) ist eine detaillierte Statistikauswertung nur bedingt sinnvoll.

Die Darstellung entspricht in diesen Fällen der Ansicht für Textspalten, obwohl der Informationsgehalt aufgrund der speziellen Inhalte eingeschränkt ist.

12.4.4.2 Spaltenstatistik für numerische Spalte

Bei einer numerischen Spalte sieht der Dialog etwas anders aus, wobei der Aufbau gleich ist.



Spaltenauswahl 1:

In diesem Feld können Sie die gewünschte Spalte auswählen, für die Sie Statistiken anzeigen lassen möchten. Dadurch müssen Sie nicht erst in der Haupttabelle auf eine Spalte klicken, sondern können direkt im Statistikfenster zwischen verschiedenen Spalten wechseln. Darüber hinaus ist es möglich den Aggregationszustand (Anzahl von Werten oder Summen) zu ändern.

Bei Summen erscheint ein weiteres Auswahlfenster, bei welchem für die Summe die **numerische Spalte** ausgewählt werden kann:

Aggregation: Summe

Numerisches Feld: Sales

Statistische Zusammenfassung 2:

Hier werden verschiedene Kennzahlen zur aktuell ausgewählten Spalte angezeigt:

- **Summe:** Die Gesamtsumme aller Werte der Spalte.
- **Minimum:** Der kleinste vorhandene Wert.

- **Maximum:** Der größte vorhandene Wert.
- **Mittelwert:** Der Durchschnitt aller Werte.
- **Median:** Der mittlere Wert, bei dem die Hälfte der Werte darunter und die Hälfte darüber liegt.
- **Standardabweichung:** Ein Maß für die Streuung der Werte um den Mittelwert.
- **Anzahl (nicht-NaN):** Die Anzahl aller gültigen (nicht fehlenden) Einträge.
- **Nullwerte:** Die Zahl der fehlenden Werte. Diese Statistiken unterstützen Sie dabei, ein besseres Verständnis über die Verteilung und Struktur Ihrer numerischen Daten zu erlangen, etwa zur Erkennung von Ausreißern oder besonderen Auffälligkeiten.

Datenvisualisierung – Top 10³:

Im unteren Bereich wird die Verteilung der Werte innerhalb der Spalte grafisch dargestellt.

Die horizontale Achse zeigt die Wertebereiche (z.B. von -2000 bis +2000), die vertikale Achse gibt die Häufigkeit der Werte in diesen Bereichen an. Zusätzlich werden normale Werte, sowie untere und obere Grenzen farblich hervorgehoben:

- **Normaler Bereich (Blau):** Werte, die innerhalb eines typischen Bereichs liegen.
- **Untere Grenze (Braun):** Werte, die unterhalb der zu erwartenden Streuung liegen.
- **Obere Grenze (Rot):** Werte, die oberhalb der zu erwartenden Streuung liegen.

Diese grafische Auswertung hilft Ihnen, ungewöhnliche Ausreißer oder stark verzerrte Verteilungen sofort zu erkennen und unterstützt Sie bei einer schnellen Einschätzung der Datenqualität.

12.4.4.3 Spaltenstatistik für Datumsspalten

Bei einer Datumsspalte werden wiederum andere Statistiken angezeigt:



Spaltenauswahl ①:

In diesem Feld können Sie die gewünschte Spalte auswählen, für die Sie Statistiken anzeigen lassen möchten. Dadurch müssen Sie nicht erst in der Haupttabelle auf eine Spalte klicken, sondern können direkt im Statistikfenster zwischen verschiedenen Spalten wechseln. Darüber hinaus ist es möglich die Aggregation (Anzahl von Werten oder Summen) sowie den Diagrammtyp für Datumsfelder abzuändern.

- Wochentag
- Jahr
- Monat-Jahr
- Tag

Bei Summen erscheint ein weiteres Auswahlfenster, bei welchem für die Summe die **numerische Spalte** ausgewählt werden kann:

Aggregation: Summe

Numerisches Feld: Sales

Beim Wochentag sieht es dann so aus wie auf dem Screenshot (hier Aggregation „Anzahl“), es ist aber auch z.B. das Jahr möglich. Dies gilt aber nur bei Spalten mit Datumsformat. Bei allen anderen Spalten existiert diese Einstellung nicht.

Statistische Zusammenfassung ❷:

Für die aktuell ausgewählte Datums-Spalte werden folgende Werte berechnet:

- **Anzahl (nicht-NaN):** Gibt an, wie viele gültige Datumseinträge vorhanden sind.
- **Nullwerte:** Zeigt, wie viele Werte in der Spalte fehlen.
- **Frühestes Datum:** Das erste (älteste) vorhandene Datum in den Daten.
- **Spätestes Datum:** Das letzte (neueste) vorhandene Datum.
- **Spanne (Tage):** Die Anzahl der Tage zwischen dem frühesten und spätesten Datum.

Diese Kennzahlen helfen, die zeitliche Abdeckung der Daten schnell zu überblicken und mögliche Lücken oder Unstimmigkeiten zu erkennen.

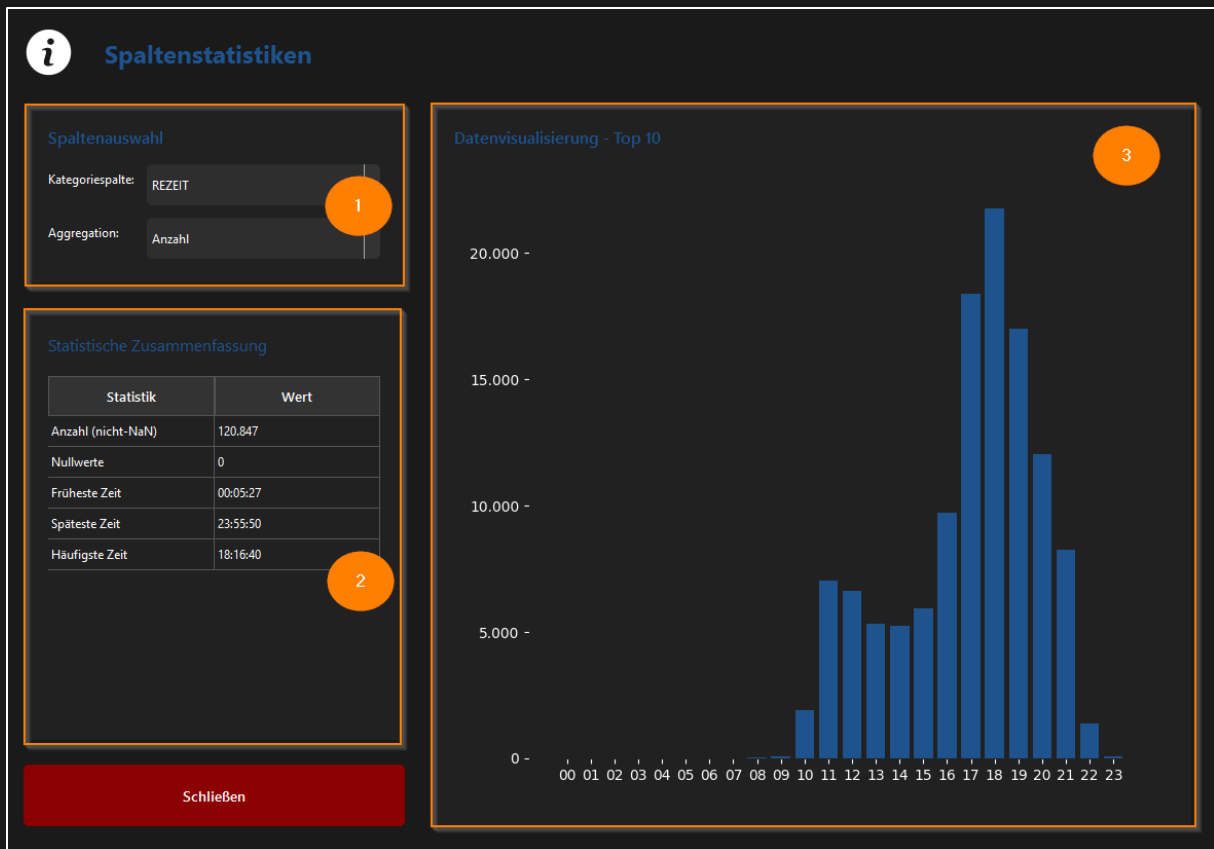
Datenvisualisierung – Top 10 ❸:

Das Diagramm zeigt die Verteilung der Einträge über die Wochentage (Standardeinstellung, siehe Punkt ❶) hinweg. So sehen Sie auf einen Blick, an welchen Tagen besonders viele oder wenige Buchungen, Bestellungen oder Ereignisse erfasst wurden.

Die Wochentage sind dabei auf der vertikalen Achse dargestellt, die Anzahl der Einträge auf der horizontalen Achse.

12.4.4.4 Spaltenstatistik für Zeitspalte

Bei einer Zeitspalte werden wiederum andere Statistiken angezeigt:



Spaltenauswahl ①:

In diesem Feld können Sie die gewünschte Spalte auswählen, für die Sie Statistiken anzeigen lassen möchten. Dadurch müssen Sie nicht erst in der Haupttabelle auf eine Spalte klicken, sondern können direkt im Statistikfenster zwischen verschiedenen Spalten wechseln. Bei der Aggregation kann zwischen **Anzahl** und **Summe** unterschieden werden.

Bei Summen erscheint ein weiteres Auswahlfenster, bei welchem für die Summe die **numerische Spalte** ausgewählt werden kann:

Aggregation: Summe

Numerisches Feld: Sales

Statistische Zusammenfassung ②:

Für die aktuell ausgewählte Zeit-Spalte werden folgende Werte berechnet:

- **Anzahl (nicht-NaN):** Gibt die Anzahl aller vorhandenen Zeitwerte an.
- **Nullwerte:** Zeigt, wie viele Einträge fehlen oder ungültig sind.
- **Früheste Zeit:** Der früheste (kleinste) Zeitwert, der in der Spalte enthalten ist.

- **Späteste Zeit:** Der späteste (größte) Zeitwert, der in der Spalte enthalten ist.
- **Häufigste Zeit:** Die Uhrzeit, die am häufigsten in den Daten vorkommt. Diese Statistiken bieten einen schnellen Überblick über die zeitliche Verteilung Ihrer Daten und helfen, Ausreißer oder Muster zu erkennen.

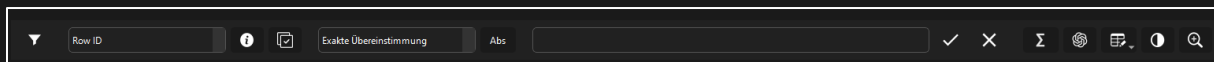
Datenvisualisierung – Top 10 ③:

Im unteren Bereich wird die Verteilung der Werte über den Tagesverlauf grafisch dargestellt.

Auf der horizontalen Achse sehen Sie die Stunden (00 bis 23 Uhr), während die vertikale Achse die Anzahl der Einträge angibt.

So können Sie auf einen Blick erkennen, zu welchen Tageszeiten die meisten Einträge vorkommen. Typische Häufungen, wie beispielsweise zur Abendzeit, werden hier sehr deutlich sichtbar.

13 Aktionsleiste



Die Aktionsleiste oder auch Actionbar genannt gehört zum schnellen Überblick über die Daten und bietet neben Summen, speziellen Filterungsoptionen auch Icons für einen Schnellaufgriff.

Die Leiste passt sich nach dem Prinzip „*What you see is what you get*“ an die jeweils ausgewählte Spalte und deren Datentyp an. Wählen Sie zum Beispiel eine **numerische Spalte** aus, so stehen Ihnen Filteroptionen wie „größer“, „kleiner“ oder „gleich“ zur Verfügung. Handelt es sich dagegen um eine **Textspalte**, ändern sich die verfügbaren Aktionen entsprechend, etwa „enthält nicht“, „stimmt nicht überein“ oder „ungefähre Übereinstimmung“. Auf diese Weise erhalten Sie immer genau die Funktionen, die für den jeweiligen Spaltentyp sinnvoll sind, ohne dass Sie lange nach den passenden Optionen suchen müssen.

13.1 Filterung



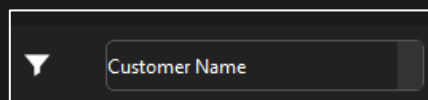
Wie bereits unter **12.4.2** dargestellt, können Sie auch innerhalb des Datensatzes eine Filterung vornehmen – beispielsweise per Rechtsklick auf eine Zeile oder Spalte. Zusätzlich steht Ihnen in der **Aktionsleiste / Actionbar** eine weitere Möglichkeit zur Filterung zur Verfügung. Dabei gibt es zwei Hauptwege:

1. Automatische Anpassung an den Feldtyp

- Positionieren Sie sich direkt in der gewünschten Spalte.
- Die Aktionsleiste passt sich automatisch an den Datentyp dieser Spalte an.
- Anschließend erhalten Sie eine Auswahl an Filteroptionen, die genau zu diesem Feldtyp passen (z. B. „größer“, „kleiner“, „enthält“ etc.).

2. Manuelle Spaltenauswahl

- Alternativ können Sie die Spalte explizit über das Feld neben dem Filtersymbol auswählen (im Beispiel „Customer Name“).
- Dies ermöglicht Ihnen, gezielt die Spalte zu definieren, in der die Filterung vorgenommen werden soll.



Beide Methoden führen zum gewünschten Filterungsergebnis und bieten Ihnen flexible Ansätze, um Ihre Daten effizient zu durchsuchen und zu analysieren.

13.1.1 Mehrere Werte filtern



Auch diese Funktion wurde bereits unter Filtern **12.4.2** beschrieben. Der Datensatz kann so beliebig mit mehreren Werten gefiltert werden. Nach Klick auf das Kästchen Symbol mit Haken öffnet sich ein neues Programmfenster und die gewünschten Filterungen können ausgewählt werden.

Eindeutige Werte in Postal Code

Übereinstimmungstyp: **1** Exakte Übereinstimmung

Sortieren nach: **2** Alphabetisch aufsteigend

Suchen: Tippen um Werte zu filtern...

Verfügbare Werte **3**

Alle auswählen Alle abwählen Wert hinzufügen

☐ 10009.0 (229)

☐ 10011.0 (193)

☐ 10024.0 (230)

☐ 10035.0 (263)

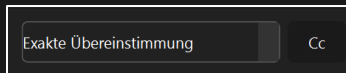
☐ 1040.0 (1)

☐ 10550.0 (8)

OK Abbrechen

Ausführliche Erklärung **unter 12.4.2.2** mehrere Werte filtern

13.1.2 Inhalte zum Filtern auswählen



Da der Feldtyp **automatisch** erkannt wird (nach Auswahl einer Spalte) ändert sich auch die Möglichkeit den Datensatz in der Spalte zu filtern.

Folgende Filterungen sind möglich:

1. Texttyp / Zeichenfolge



- **Exakte Übereinstimmung:** Zeigt nur Datensätze an, deren Text exakt mit dem eingegebenen Suchbegriff übereinstimmt.
- **Teilweise Übereinstimmung:** Zeigt Datensätze an, in denen zumindest ein Teil des Feldinhalts mit dem Suchbegriff übereinstimmt. Beispielsweise würde der Begriff „Haus“ auch „Baumhaus“ erfassen.
- **Stimmt nicht überein:** Schließt alle Datensätze aus, die exakt mit dem Suchbegriff übereinstimmen.
- **Enthält nicht:** Schließt alle Datensätze aus, die den Suchbegriff (auch nur teilweise) enthalten.
- **Ungefähre Übereinstimmung / sog. Fuzzy Methode:** Erfasst ähnlich geschriebene Begriffe, selbst wenn sie nicht exakt mit dem Suchbegriff übereinstimmen (z.B. Tippfehler oder Abweichungen in der Schreibweise). Hierzu wird auf Kapitel **12.4.2.2.1** verwiesen.

Zusätzliche Funktion:



Mit Klick auf das Symbol können bei der Filterung auch **Groß-/Kleinschreibung** berücksichtigt werden.

2. Datumsfelder



- **Exakte Übereinstimmung:** Zeigt nur Einträge an, die genau an einem bestimmten Datum liegen.
- **Stimmt nicht überein:** Schließt alle Einträge aus, die an einem bestimmten Datum liegen.
- **Vor:** Filtert alle Datensätze, deren Datum älter (früher) ist als das angegebene Vergleichsdatum.
- **Nach:** Zeigt alle Datensätze an, deren Datum jünger (später) ist als das Vergleichsdatum.

- **Am oder vor:** Erfasst Datensätze, die bis einschließlich des angegebenen Datums liegen.
- **Am oder nach:** Schließt alle Datensätze ein, die ab dem angegebenen Datum und später liegen.

Zusätzliche Funktion:



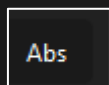
Mit Klick auf das Symbol wird bei der Filterung ausschließlich das **Jahr** berücksichtigt

3. Numerische Felder



- **Exakte Übereinstimmung:** Zeigt nur Datensätze, bei denen der Zahlenwert genau dem gesuchten Wert entspricht.
- **Stimmt nicht überein:** Schließt alle Datensätze aus, die exakt dem angegebenen Zahlenwert entsprechen.
- **Größer als:** Liefert alle Datensätze, deren Zahlenwert über dem Vergleichswert liegt.
- **Kleiner als:** Zeigt Datensätze, deren Wert unter dem Vergleichswert liegt.
- **Größer oder gleich:** Nimmt alle Datensätze auf, deren Zahlenwert mindestens dem Vergleichswert entspricht.
- **Kleiner oder gleich:** Nimmt alle Datensätze auf, deren Zahlenwert höchstens dem Vergleichswert entspricht.

Zusätzliche Funktion:



Mit Klick auf das Symbol können auch **absolute Werte** berücksichtigt werden. So werden die Werte unabhängig vom Vorzeichen verwendet.

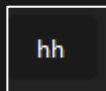
4. Zeitfelder



- **Exakte Übereinstimmung:** Zeigt nur Einträge an, die genau zu einer bestimmten Uhrzeit liegen.
- **Stimmt nicht überein:** Schließt alle Einträge aus, die zu einer bestimmten Uhrzeit liegen.
- **Vor:** Filtert alle Datensätze, deren Uhrzeit früher ist als die angegebene Vergleichszeit.
- **Nach:** Zeigt alle Datensätze an, deren Uhrzeit später ist als die Vergleichszeit.

- **Am oder vor:** Erfasst Datensätze, die bis einschließlich der angegebenen Uhrzeit liegen.
- **Am oder nach:** Schließt alle Datensätze ein, die ab der angegebenen Uhrzeit und später liegen.

Zusätzliche Funktion:



Mit Klick auf das Symbol werden nur die **Stunden** aus dem Zeitfeld berücksichtigt.

13.1.3 Anzeige & Filter verwalten



Das **Filtersymbol** zeigt an, ob aktuell Filter auf den Datensatz angewendet sind. Leuchtet es **rot**, bedeutet dies, dass mindestens ein Filter aktiv ist. Bleibt es neutral, sind keine Filter gesetzt.



Über **Klick auf das Symbol** gelangen Sie in die **Filterverwaltung**. Diese Funktion bietet Ihnen einen schnellen Überblick über alle aktuell gesetzten Filter und ermöglicht deren einfache Anpassung.

Die Filterverwaltung gliedert sich in folgende Bereiche:

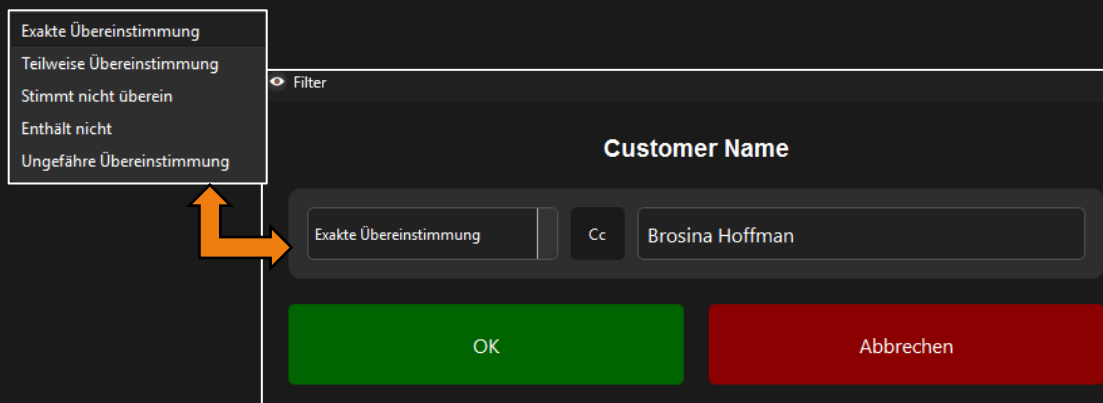
1. Aktive Filter anzeigen ①

- Nach dem Klick auf das Filtersymbol wird eine übersichtliche Liste aller derzeit aktiven Filter angezeigt.

2. Filter bearbeiten ②

- Mit dem **Stift-Symbol** können Sie einzelne Filter bearbeiten.
- In der dargestellten Filterverwaltung (Beispiel Musterfall) wurde nach dem „Customer Name“ Brosina Hoffman gefiltert. Klickt man auf das Stift Symbol öffnet sich das folgende Fenster, in dem sich die Filterkriterien anpassen oder weiter präzisieren lassen. Auch hier wird der Feldtyp automatisch erkannt.



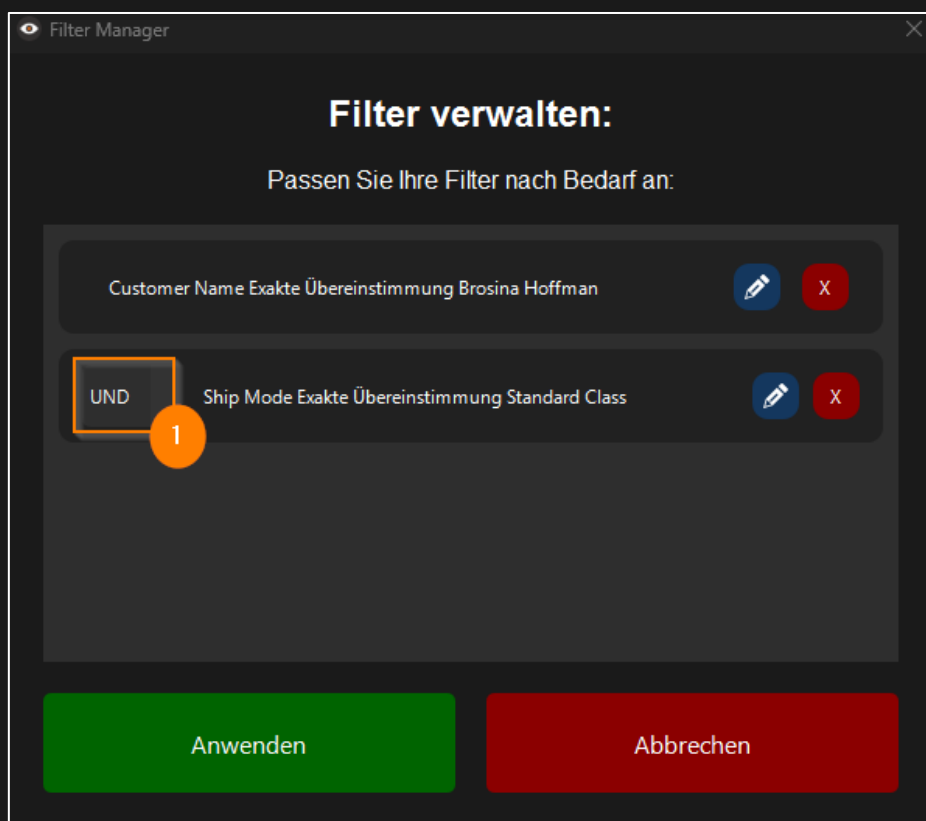


3. Filter löschen ③

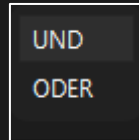
- Über das **X-Symbol** können Sie einen oder mehrere Filter aus der Liste entfernen (**Hinweis: Man kann die Filter auch insgesamt aufheben über das X in der Aktionsleiste / Actionbar**).
- Dadurch wird der Datensatz wieder unbeschränkt angezeigt oder es können neue Filter gesetzt werden.

Hinweis: Mehrere Filter

Wenn Sie mehrere Filter setzen, erscheint ab dem zweiten Filter ein zusätzliches Auswahlfeld (UND/ODER). Damit können Sie festlegen, ob die Filterkriterien gleichzeitig (UND) oder alternativ (ODER) angewendet werden sollen.



Standardmäßig ist eine doppelte Filterung über „und“ eingestellt. Mit Klick auf „und“ ① ermöglicht es auch eine Filterung nach „oder“



Diese strukturierte Ansicht der Filterverwaltung ermöglicht es Ihnen, bestehende Filter schnell zu überprüfen, anzupassen und zu entfernen – alles, ohne dass Sie jeden Filter von Grund auf neu erstellen müssen.

13.1.4 Alle Filter löschen



Das **X** in der Schnellzugriffleiste dient dazu, alle aktiven Filter auf einmal zu löschen. Sobald ein Filter aktiv ist, leuchtet das Symbol rot, was sofort signalisiert, dass der Datensatz derzeit gefiltert ist. Mit einem Klick auf das **X** werden sämtliche Filter entfernt und der ursprüngliche, ungefilterte Datensatz wieder angezeigt.



13.2 Info: Spaltenstatistik



Über das **Info-Symbol** in der Aktionsleiste (Actionbar) können Sie ein separates Fenster öffnen, in dem die **Spaltenstatistiken** kompakt und übersichtlich dargestellt werden. Wählen Sie die gewünschte Spalte aus, um sich einen Überblick über die zugehörigen statistischen Werte zu verschaffen. Die angezeigten Kennzahlen variieren je nach Typ der Spalte und liefern gezielt passende Informationen, beispielsweise zu Anzahl, Verteilung oder Besonderheiten der Einträge.

Eine detaillierte Erläuterung der einzelnen Auswertungen und grafischen Darstellungen finden Sie in **Abschnitt 12.4.4** dieses Handbuchs.

13.3 Summen



Die Funktion **Summe berechnen** ermöglicht es, für alle numerischen Spalten automatisch die Summe der enthaltenen Werte zu ermitteln. Dabei wird die Berechnung dynamisch an die aktuelle Filterung des Datensatzes angepasst. Das heißt, wenn Sie einen Filter anwenden, aktualisiert sich die Summe automatisch und bezieht nur die in der gefilterten Ansicht enthaltenen Werte mit ein. So behalten Sie jederzeit einen präzisen Überblick über aggregierte Kennzahlen, die exakt Ihrer aktuellen Datenauswahl entsprechen.

● Row ID: 49.945.015,00	● Postal Code: 551.513.164,00	● Sales: 2.297.200,86	● Quantity: 37.873,00	● Discount: 1.561,09	● Profit: 286.397,02	Summe

Über ein zusätzliches Feld in der **Summenleiste** lassen sich weitere statistische Kennzahlen auswählen, darunter **Minimum**, **Maximum**, **Mittelwert**, **Median**, **Anzahl**, **Anzahl (eindeutig)**, **Standardabweichung** und **Varianz**. So können Sie flexibel entscheiden, welche Größe für Ihre aktuelle Analyse relevant ist und sich automatisch auf den gefilterten Datenbestand berechnen lassen.



13.4 KI-Spalten Generator



Mit der **KI-Spalte** haben Sie die Möglichkeit, mithilfe von AMY, der künstlichen Intelligenz Assistentin im Analyzer komplexe Transformationen oder Extraktionen an Ihren Spalten vorzunehmen, ohne selbst Code schreiben zu müssen. Dabei beschreiben Sie einfach in natürlicher Sprache, wie die Spalte verändert werden soll, und die KI setzt diese Anweisung automatisch um. Bei dem Feld handelt es sich um ein Schnellzugriff. Die Funktion kann auch über Data-Tools im Header (oder unter dem Schnellzugriff Table-Tools (3.6)) und dort unter den Table-Tools (Berechnen) geöffnet werden.

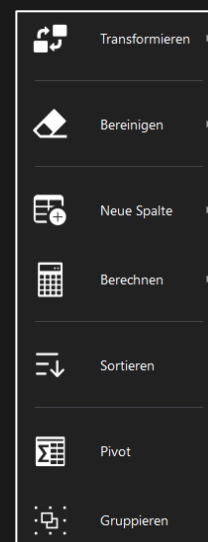
Die genaue Vorgehensweise kann dem **Kapitel 14.1.6.3.110.3** entnommen werden.

13.5 Table-Tools



Die **Table-Tools in der Schnellzugriffleiste** bieten einen schnellen Zugriff auf häufig benötigte Funktionen wie **Gruppieren**, **Pivot**, **Duplikate**, **Sortieren**, **Berechnen** und **Transformieren**. Sie stellen damit eine komfortable Abkürzung dar, da dieselben Werkzeuge alternativ auch über die **Data-Tools** im Header aufgerufen werden können.

Die einzelnen Funktionen werden in den **Kapiteln 14.1.6 ff** detailliert beschrieben.



13.6 Erscheinungsbild AUDIPY



Über das **Kontrast-Icon** lässt sich die Anwendung in einen **dunklen Modus** schalten, sodass das Erscheinungsbild insgesamt abgedunkelt wird. Dies kann insbesondere bei längerer Arbeit am Bildschirm oder in schwach beleuchteten Umgebungen angenehmer sein. Standardmäßig ist jedoch die **helle Version** aktiviert, sodass Sie bei Bedarf manuell zum dunklen Erscheinungsbild wechseln können.

AUDIPY Analyzer

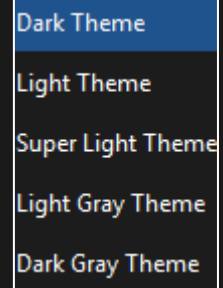
Start Explorer Data Tools Speichern Zurücksetzen Leertisch Audit Steps Ask my Data Visualisieren Analysieren Exportieren

Musterfall.xls

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City	State	Postal Code	Region	Product ID	Category	Sub-Category
7	CA-2018-115812	09.06.2018	14.06.2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90.032.0	West	OFF-AB-10002033	Office Supplies	Art
8	CA-2018-115812	09.06.2018	14.06.2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90.032.0	West	TEC-PH-10002275	Technology	Phones
9	CA-2018-115812	09.06.2018	14.06.2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90.032.0	West	OFF-BI-10003910	Office Supplies	Binders
10	CA-2018-115812	09.06.2018	14.06.2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90.032.0	West	OFF-AP-10002092	Office Supplies	Appliances
11	CA-2018-115812	09.06.2018	14.06.2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90.032.0	West	FUR-TA-10001539	Furniture	Tables
12	CA-2018-115812	09.06.2018	14.06.2018	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California	90.032.0	West	TEC-PH-10002033	Technology	Phones
13	CA-2021-114412	15.04.2021	20.04.2021	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina	28.027.0	South	OFF-PA-10002365	Office Supplies	Paper
14	CA-2020-161389	05.12.2020	10.12.2020	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington	98.103.0	West	OFF-BI-10003656	Office Supplies	Binders
15	US-2019-118883	22.11.2019	26.11.2019	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76.106.0	Central	OFF-AP-10002311	Office Supplies	Appliances
16	US-2019-118883	22.11.2019	26.11.2019	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas	76.106.0	Central	OFF-BI-10000736	Office Supplies	Binders
17	CA-2018-105893	11.11.2018	18.11.2018	Standard Class	PK-19075	Pete Krog	Consumer	United States	Madison	Wisconsin	53.711.0	Central	OFF-ST-10004186	Office Supplies	Storage
18	CA-2018-167164	13.05.2018	15.05.2018	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jordan	Utah	84.084.0	West	OFF-ST-10000107	Office Supplies	Storage
19	CA-2018-143336	27.08.2018	01.09.2018	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94.109.0	West	OFF-PA-10003056	Office Supplies	Art
20	CA-2018-143336	27.08.2018	01.09.2018	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94.109.0	West	TEC-PH-10001949	Technology	Phones
21	CA-2018-143336	27.08.2018	01.09.2018	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94.109.0	West	OFF-BI-10002215	Office Supplies	Binders
22	CA-2020-137330	09.12.2020	13.12.2020	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68.025.0	Central	OFF-AP-10000246	Office Supplies	Art
23	CA-2020-137330	09.12.2020	13.12.2020	Standard Class	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68.025.0	Central	OFF-AP-10001492	Office Supplies	Appliances
24	US-2021-156809	16.07.2021	18.07.2021	Second Class	SF-20065	Sandra Flanagan	Consumer	United States	Philadelphia	Pennsylvania	19.140.0	East	FUR-CH-10002274	Furniture	Chairs
25	CA-2019-106320	25.09.2019	30.09.2019	Standard Class	EB-13870	Emily Burns	Consumer	United States	Orem	Utah	84.057.0	West	FUR-TA-10000577	Furniture	Tables
26	CA-2020-121755	16.01.2020	20.01.2020	Second Class	EH-13945	Eric Hoffmann	Consumer	United States	Los Angeles	California	90.049.0	West	OFF-BI-10001634	Office Supplies	Binders
27	CA-2020-121755	16.01.2020	20.01.2020	Second Class	EH-13945	Eric Hoffmann	Consumer	United States	Los Angeles	California	90.049.0	West	TEC-AC-10003027	Technology	Accessories
28	US-2019-150630	17.09.2019	21.09.2019	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19.140.0	East	FUR-BO-10004634	Furniture	Bookcases
29	US-2019-150630	17.09.2019	21.09.2019	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19.140.0	East	OFF-BI-10000474	Office Supplies	Binders
30	US-2019-150630	17.09.2019	21.09.2019	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19.140.0	East	FUR-FU-10004846	Furniture	Furnishings
31	US-2019-150630	17.09.2019	21.09.2019	Standard Class	TB-21520	Tracy Blumstein	Consumer	United States	Philadelphia	Pennsylvania	19.140.0	East	OFF-EN-10001509	Office Supplies	Envelopes

Exakte Übereinstimmung Abs

Hinweis: Möchte man lieber mit einem insgesamt hellen Design des AUDIPYs arbeiten, so kann man unter den Einstellungen (unter Ansicht **3.9.1.2**) verschiedene Designs auswählen.



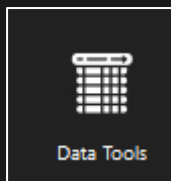
13.7 Zoom



Mit der **Zoom-Funktion** können Sie die Darstellung Ihrer Datentabelle stufenlos vergrößern oder verkleinern. Beim Aktivieren dieser Funktion öffnet sich eine **Skala** mit **+** / **–**-Schaltflächen. Über diese regeln Sie die gewünschte Zoomstufe, sodass Zeilen und Spalten entweder größer (bessere Lesbarkeit) oder kompakter (mehr Inhalte auf einmal) angezeigt werden. Dies erleichtert das Arbeiten mit umfangreichen Datensätzen, da Sie jederzeit die optimale Übersicht einstellen können.



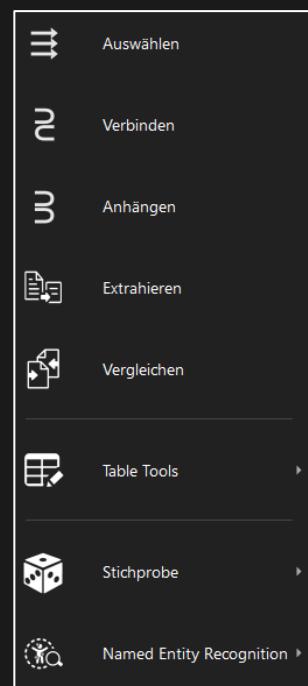
14 Data Tools



Der Menüpunkt **Data Tools** bündelt zentrale Werkzeuge für die Bearbeitung, Anreicherung und Analyse Ihrer Daten.

Hier können Sie unter anderem Datensätze auswählen, zusammenführen, anhängen oder vergleichen und Stichproben für Testzwecke ziehen. Darüber hinaus stehen erweiterte Funktionen wie die **Named Entity Recognition** bereit, um wertvolle Informationen aus unstrukturierten Daten zu extrahieren.

So behalten Sie stets die volle Kontrolle über Ihre Datensätze und können sie bedarfsgerecht aufbereiten. Die einzelnen Menüpunkte werden im folgenden detailliert beschrieben:

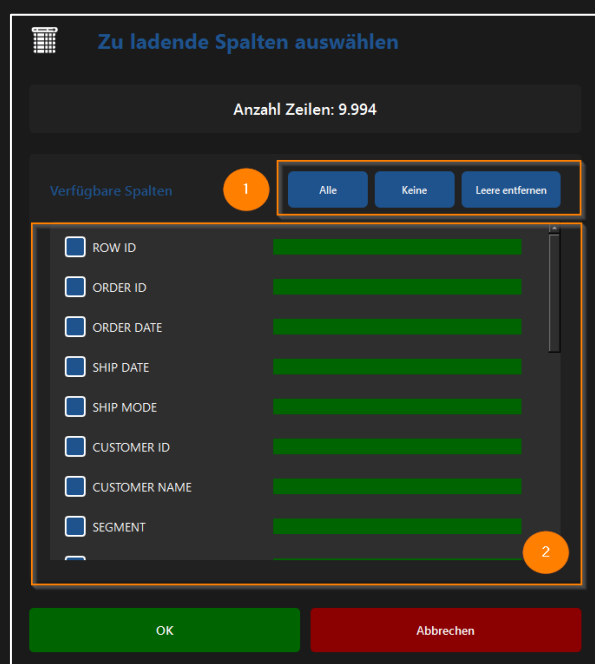


14.1 Auswählen

Unter dem Menüpunkt „**Auswählen**“ haben Sie die Möglichkeit, gezielt zu entscheiden, welche Spalten in den Datenbereich geladen und angezeigt werden sollen. Auf diese Weise können Sie sich auf die für Ihre Analyse relevanten Informationen konzentrieren und gleichzeitig die **Performance steigern**, da unnötige Spalten nicht verarbeitet werden müssen.

Diese Funktion entfaltet insbesondere bei **Parquet-Dateien** ihre volle Stärke: Durch die spaltenbasierte Speicherung kann der Ladevorgang deutlich beschleunigt und der **Arbeitsspeicherverbrauch reduziert** werden, da nur die ausgewählten Spalten tatsächlich eingelesen werden.

Wichtig: Beim anschließenden Speichern des Falls werden **ausschließlich die geladenen Spalten** gespeichert. Spalten, die beim Laden nicht ausgewählt wurden, gehen dabei verloren. Wenn der vollständige Datenbestand erhalten bleiben soll, empfiehlt es sich, die Funktion „**Speichern unter...**“ zu nutzen und den Fall unter einem neuen Namen zu sichern.



1. Übersicht der Spalten ①

- Im angezeigten Dialog werden sämtliche verfügbaren Spalten aufgelistet, zusammen mit einer Anzeige der Zeilenanzahl.
- Grün zeigt an, dass in der Spalte gültige Werte vorhanden sind. Der Balken repräsentiert dabei die relative Anzahl der nicht-leeren Einträge.
- Rot weist auf leere Einträge hin – also Zellen mit NULL oder fehlenden Werten. Je mehr Rot angezeigt wird, desto höher ist der Anteil fehlender Daten in dieser Spalte.
- Jede Spalte kann über ein Kontrollkästchen individuell an- oder abgewählt werden.

2. Schnellaktionen 2

- **Alle:** Aktiviert sämtliche Spalten.
- **Keine:** Deaktiviert alle Spalten.
- **Leere entfernen:** Blendet Spalten aus, die keine oder nur leere Werte enthalten.

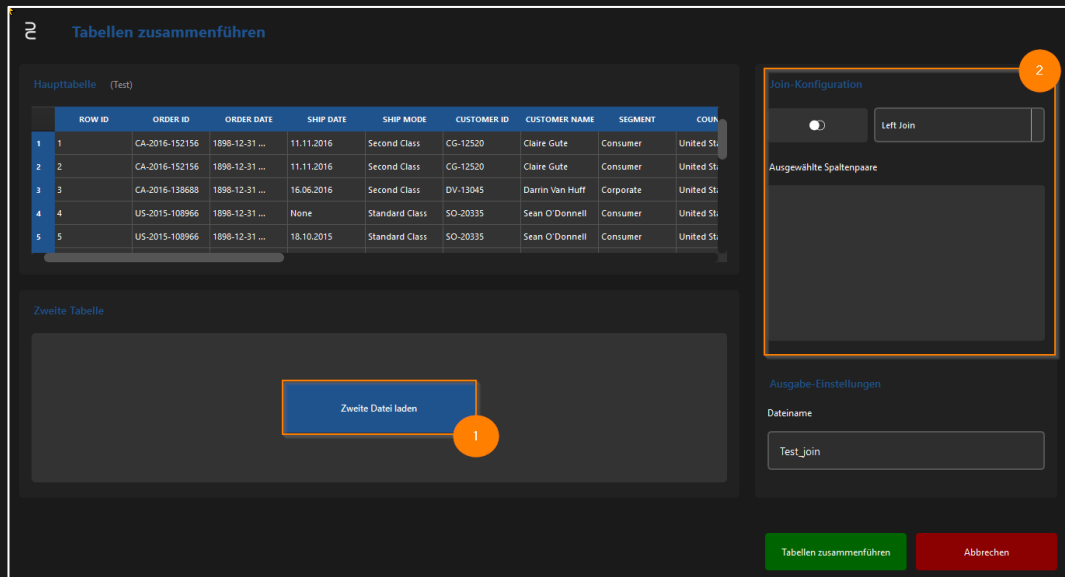
3. Bestätigung

- Klicken Sie auf **OK**, um Ihre Auswahl zu übernehmen. Nur die ausgewählten Spalten werden dann im Datenbereich angezeigt und verarbeitet.
- Mit **Abbrechen** verwerfen Sie Ihre Änderungen und behalten den bisherigen Zustand bei.

Durch das gezielte Auswählen relevanter Spalten reduzieren Sie den Speicherbedarf und verbessern die Arbeitsgeschwindigkeit bei großen Datensätzen. Sollten Sie später feststellen, dass Sie zusätzliche Spalten benötigen, können Sie diesen Vorgang jederzeit wiederholen und die benötigten Spalten erneut aktivieren.

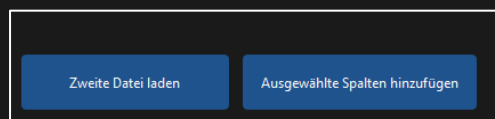
14.2 Verbinden

Unter **Verbinden** können Sie zwei separate Datentabellen zusammenführen, um daraus eine kombinierte Tabelle zu erstellen. Diese Funktion ist besonders nützlich, wenn Sie Informationen aus verschiedenen Quellen oder Dateien in einer einzigen Ansicht benötigen. Im Folgenden finden Sie eine detaillierte Beschreibung der einzelnen Schritte und Optionen:



1. Zweite Datei laden ①

- Um eine zusätzliche Tabelle zu verbinden, laden Sie zunächst die entsprechende Datei (eine Parquet-Datei) über den entsprechenden Button. Die Anwendung zeigt daraufhin beide Tabellen untereinander an.
- Anschließend erhalten Sie am unteren linken Programmfenster weitere Schaltflächen:



2. Verknüpfungstyp auswählen ②



Bei der Zusammenführung von zwei Tabellen können unterschiedliche Join-Typen festgelegt werden, die bestimmen, wie Datensätze miteinander verknüpft werden (die blaue Farbe zeigt das Ergebnis an):



Inner Join

Nur Datensätze, die in beiden Tabellen einen passenden Schlüsselwert besitzen, werden in das Ergebnis übernommen. Alle Einträge ohne Entsprechung in der jeweils anderen Tabelle entfallen.

Left Join (Linker Join)

Alle Datensätze aus der ersten Tabelle bleiben erhalten. Passende Einträge aus der zweiten Tabelle werden hinzugefügt. Fehlen passende Schlüsselwerte, werden die Felder aus der zweiten Tabelle entsprechend leer (NULL) oder nicht belegt.

Right Join (Rechter Join)

Analog zum Left Join, nur in umgekehrter Richtung: Alle Datensätze aus der zweiten Tabelle bleiben erhalten, und passende Einträge aus der ersten Tabelle werden ergänzt. Fehlt eine Übereinstimmung, bleibt der entsprechende Bereich leer.

Full (Outer) Join

Es werden sämtliche Datensätze aus beiden Tabellen zusammengeführt, unabhängig davon, ob passende Schlüsselwerte vorhanden sind. Sofern keine Übereinstimmung besteht, bleiben die Felder der jeweils anderen Tabelle leer.

Left Anti Join

Es werden ausschließlich die Datensätze aus der linken Tabelle übernommen, **für die es keine passenden Schlüsselwerte in der rechten Tabelle gibt**. Enthält ein Eintrag aus der linken Tabelle **keine Übereinstimmung**, wird er übernommen – andernfalls wird er ausgeschlossen.

Es werden ausschließlich die Datensätze aus der rechten Tabelle übernommen, **für die es keine passenden Schlüsselwerte in der linken Tabelle gibt**. Enthält ein Eintrag aus der rechten Tabelle **keine Übereinstimmung**, wird er übernommen – andernfalls wird er ausgeschlossen.

2

Tabellen zusammenführen

Haupttabelle (Testdatei A)

9.201 DUPLIKATE

	Row ID	Order ID	Order Date	Ship Date	Customer ID	Customer Name	Segment	Country	City
1	1	CA-2016-152156	1899-01-01 ...	11.11.2016	CG-12520	Claire Gute	Consumer	United States	Henderson
2	2	CA-2016-152156	1899-01-01 ...	11.11.2016	CG-12520	Claire Gute	Consumer	United States	Henderson
3	3	CA-2016-138688	1899-01-01 ...	16.06.2016	DV-13045	Darrin Van Huff	Corporate	United States	Los Angel
4	4	US-2015-108966	1899-01-01 ...	18.10.2015	SO-20335	Sean O'Donnell	Consumer	United States	Fort Laud
5	5	US-2015-108966	1899-01-01 ...	18.10.2015	SO-20335	Sean O'Donnell	Consumer	United States	Fort Laud

Zweite Tabelle (Testdatei B)

5.194 DUPLIKATE

	Row ID	Order ID	Order Date	Ship Date	Customer ID	Customer Name	Segment	Country/Region	City
1	8223	CA-2018-152905	2018-02-18 ...	nan	AB-10015	Aaron Bergman	Consumer	United States	Arlington
2	5355	CA-2018-113768	2018-05-13 ...	nan	AH-10030	Aaron Hawkins	Corporate	United States	Los Angel
3	3200	CA-2021-164000	2021-12-18 ...	nan	AH-10030	Aaron Hawkins	Corporate	United States	Philadelpi
4	5356	CA-2018-113768	2018-05-13 ...	nan	AH-10030	Aaron Hawkins	Corporate	United States	Los Angel
5	9681	CA-2019-130113	2019-12-27 ...	nan	AH-10030	Aaron Hawkins	Corporate	United States	San Franc

Join-Konfiguration

Left Join

Ausgewählte Spaltenpaare

SHIP MODE -> Ship Mode

Ausgabe-Einstellungen

Dateiname

Testdatei_A_Join

Zweite Datei laden

Ausgewählte Spalten hinzufügen

Tabellen zusammenführen

Abbrechen

- ## 1. Verbindungsspalten festlegen (mindestens eine ist erforderlich) 1
- Um die Tabellen korrekt zu verbinden, müssen sie über mindestens eine gemeinsame Spalte verfügen, anhand derer sich die Datensätze zuordnen lassen (z. B. eine Kunden-ID).
 - Stellen Sie sicher, dass beide Tabellen in der entsprechenden Spalte dieselben Werte (z. B. gleiche IDs) verwenden und im passenden Format vorliegen.
 - Die aktuell ausgewählte Spalte wird durch die blaue Farbe visuell hervorgehoben. Die roten Infobereiche zeigen an, wie viele Duplikate in der jeweiligen Spalte vorhanden sind. Dadurch lässt sich bereits ein erster Eindruck von der Spaltenqualität der Verbindung gewinnen."
 - Über die Schaltfläche „**Ausgewählte Spalten hinzufügen**“ kann eine Verbindung beider Tabellen geschaffen werden.
 - Sobald ein Kriterium hinzugefügt wurde, verschwinden diese in der Vorschau beider Tabellen. Die Verbindung wird jedoch im Vorschaufenster „**Ausgewählte Spaltenpaare**“ 2 angezeigt.

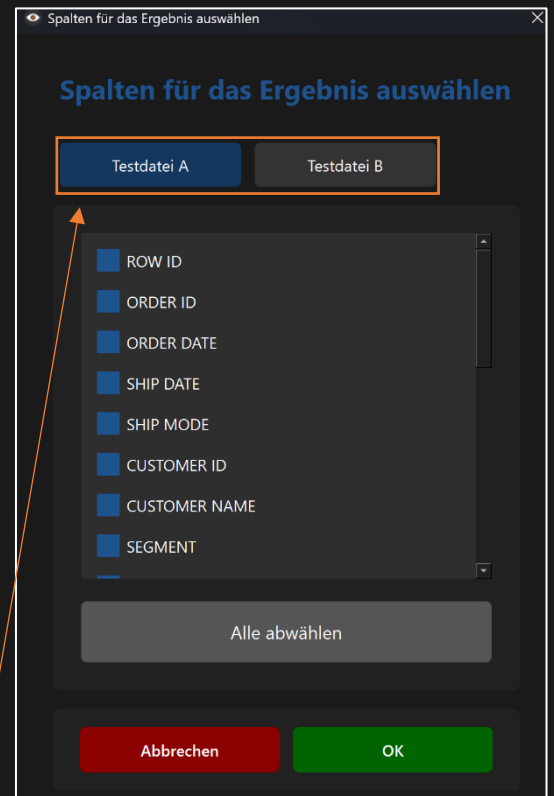
Überprüfen Sie hier, ob die Zuordnung der Spalten Ihren Vorstellungen entspricht und ob alle relevanten Felder korrekt gemappt sind. Auf diese Weise stellen Sie sicher, dass die Zusammenführung der Datensätze wie gewünscht erfolgt und keine wichtigen Informationen fehlen.

2. Ausgabe Einstellungen 3

- Hier kann ein individueller Dateiname eingetragen werden. Andernfalls wird der Dateiname ergänzt durch „_join“ (im Screenshot Test_join)

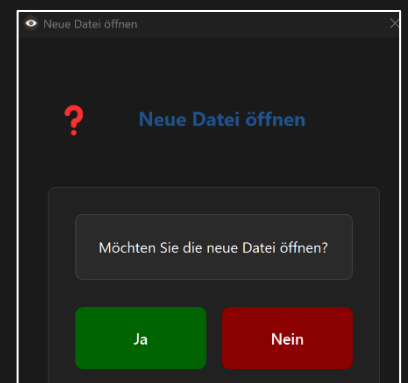
3. Tabellen zusammenführen

- Klicken Sie auf **Tabellen zusammenführen**, um den Zusammenführungsprozess fortzusetzen:
- Im anschließenden Fenster haben Sie die Möglichkeit, bestimmte Spalten für die neue verbundene Datei abzuwählen oder gezielt Wunschspalten auszuwählen.
- Über den Schnellzugriff **„Alle abwählen“** können Sie zunächst alle Spalten deaktivieren, um anschließend einzelne gezielt auszuwählen.
- Im Reiter können Sie die Felder der jeweiligen Einzeltabelle sehen, um so getrennt für jede Tabelle eine Auswahl zu treffen.
- Falls Sie den Vorgang abbrechen möchten, klicken Sie auf **Abbrechen**, ohne Änderungen zu übernehmen.



Nach Auswahl wird die neue Datei erstellt. Anschließend werden Sie gefragt, ob Sie die neue Datei sofort öffnen möchten?

Mit Klick auf Ja wird die Datei direkt geöffnet und befindet sich auf der linken Seite auch im Explorer.



14.3 Anhängen

Mit der Funktion **Anhängen** lassen sich mehrere Parquet-Dateien zu einem einzigen Datenbestand zusammenführen. Im gezeigten Dialogfenster können Sie die gewünschten Dateien auswählen und anschließend per Klick auf **Anhängen** in einer gemeinsamen Tabelle vereinen.

Hier die wichtigsten Punkte im Überblick:

1. Parquet-Dateien hinzufügen ①

- Mit diesem Button fügen Sie weitere Parquet-Dateien hinzu.
- Achten Sie darauf, dass die Dateien möglichst denselben Aufbau (z.B. Spaltenstruktur) aufweisen, damit sie sich problemlos zusammenführen lassen.

2. Ausgabe Einstellungen ②

- Hier können Sie einen individuellen Namen für die neue Datei erstellen (im Beispiel: *Test A und B gesamt*)

3. Anhängen

- Durch Klick auf **Anhängen** werden alle ausgewählten Parquet-Dateien zeilenweise zu einer einzigen Tabelle zusammengefügt und tauchen auch im Explorer als Datei auf.
- Bei abweichenden Spaltenstrukturen versucht die Anwendung, die Daten sinnvoll zu vereinen (z.B. durch das Einfügen leerer Werte in fehlenden Spalten).

4. Abbrechen

- Über **Abbrechen** können Sie den Vorgang jederzeit beenden, ohne dass die Daten zusammengeführt werden.

Dank dieser Funktion können Sie große oder auf mehrere Dateien verteilte Datensätze bequem konsolidieren und im Anschluss als eine einheitliche Tabelle weiterverarbeiten.

Hinweis: Unterschiedliche Spalten

Beinhaltet die angehängte Datei Spalten, welche nicht in der ersten Datei enthalten sind, so werden diese Spalten einfach leer ohne Zuordnung neben den bestehenden Spalten hinzugefügt:

ORDER ID	ORDER DATE	SHIP DATE

Anhängen

Ausgewählte Dateien

C:/Users/eilen/Documents/AUDIPY/De.../Quelldateien/Testdatei B.parquet

1

Parquet-Dateien hinzufügen

Ausgabe-Einstellungen

Dateiname

Test A und B gesamt

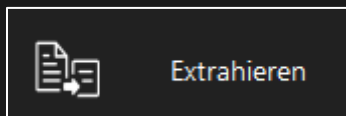
2

Anhängen

Abbrechen

Test A und B gesamt	01.07.2

14.4 Extrahieren



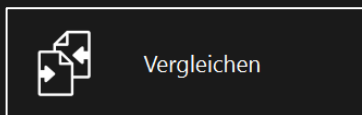
Extrahieren

Die Funktion „**Extrahieren**“ dient dazu, gezielt Daten aus dem aktuellen DataFrame zu entnehmen – basierend auf gesetzten Filtern, Gruppierungen oder einem definierten Schlüsselfeld. Dadurch können relevante Teildatensätze schnell isoliert, gespeichert und für weitere Analysen oder Verarbeitungsschritte verwendet werden.

Ein zentraler Vorteil: Die Extraktion erfolgt **direkt auf dem sichtbaren (gefilterten) Zustand** des DataFrames. So wird exakt das extrahiert, was der Nutzer sieht und benötigt

Diese Funktion wurde ausführlich unter **1.2.2.2.1** erläutert, daher wird an dieser Stelle auf dieses Kapitel verwiesen.

14.5 Vergleichen



Vergleichen

Die **Vergleichs-Funktion** stellt eine zentrale Analysemethode zur systematischen Gegenüberstellung von Datensätzen dar, die insbesondere für Wirtschaftsprüfer, Steuerberater und Controller von zentraler Bedeutung ist. Diese Funktionalität ermöglicht es, Unterschiede zwischen zwei Parquet-Dateien zu identifizieren, zu visualisieren und zu dokumentieren.

Der Vergleich arbeitet intelligent und erkennt nicht nur unterschiedliche Werte, sondern kann auch mit verschobenen, gelöschten oder hinzugefügten Datensätzen umgehen. Dies ist besonders relevant für die Analyse von Zeitreihen, Bestandsveränderungen oder bei der Überprüfung von Datenübernahmen, wo sich nicht nur Werte, sondern auch die Reihenfolge von Datensätzen ändern kann.

Hauptmerkmale:

- **Zwei Dateien gleichzeitig öffnen:**
Wählen Sie zwei beliebige Dateien aus, um sie parallel anzeigen zu lassen. Dies eignet sich ideal zum Auffinden von Unterschieden, Überarbeitungen oder Versionsvergleichen.
- **Flexible Anordnung:**
Die Darstellung der beiden Dateien kann je nach Bedarf erfolgen:
 - **Horizontal:** Eine Datei oben, die andere unten.
 - **Vertikal:** Eine Datei links, die andere rechts.
- **Synchronisiertes Scrollen (optional):**
Beim Vergleich kann das Scrollen synchronisiert erfolgen, um verwandte Abschnitte gleichzeitig zu betrachten.

14.5.1 *Praktische Anwendungsszenarien*

Die Vergleichsfunktion ist für unterschiedlichste Bereiche bei der Datenanalyse von enormer Bedeutung:

Wirtschaftsprüfung

Bilanzvergleiche zwischen Geschäftsjahren: Wirtschaftsprüfer können mit der Compare-Funktion systematisch Bilanzpositionen verschiedener Stichtage analysieren. Besonders wertvoll ist die Funktion bei Kontenplänen, die sich durch Umgliederungen oder neue Konten unterscheiden. Die Spalten-Zuordnung ermöglicht es, auch bei verschiedenen ERP-Systemen oder unterschiedlichen Kontonummerierungen valide Vergleiche durchzuführen.

Plausibilitätsprüfungen: Bei analytischen Prüfungshandlungen können Planwerte mit Istwerten oder Vorjahreswerte mit aktuellen Zahlen systematisch verglichen werden. Die Excel-Exportfunktion liefert eine vollständige Dokumentation für die Prüfungsakte. Besonders bei Konzernen ermöglicht die Funktion den Vergleich von Daten aus verschiedenen Tochterunternehmen, auch wenn diese unterschiedliche Buchungsstrukturen verwenden.

Datenqualitätskontrolle: Vor substantiven Prüfungshandlungen kann die Compare-Funktion zur Überprüfung von Datenmigrations- oder Systemwechselprozessen eingesetzt werden. Unterschiede zwischen Alt- und Neusystem werden transparent dargestellt und können hinsichtlich ihrer Prüfungsrelevanz bewertet werden.

Steuerberatung

Mandantenvergleiche und Benchmarking: Steuerberater können anonymisierte Mandantendaten für Branchenvergleiche nutzen. Die flexible Spalten-Zuordnung ermöglicht es, auch bei unterschiedlichen Kontenplänen oder Buchungslogiken aussagekräftige Vergleiche durchzuführen. Dies unterstützt Beratungsgespräche, in denen Optimierungspotentiale aufgezeigt werden sollen.

Jahresvergleiche für Steuererklärungen: Bei der Erstellung von Steuererklärungen können Vorjahreswerte systematisch mit aktuellen Zahlen verglichen werden. Besonders bei Unternehmen mit umfangreichem Anlagevermögen hilft die Funktion dabei, Veränderungen bei Abschreibungen oder Zu- und Abgängen nachzuvollziehen. Die Excel-Ausgabe kann direkt als Grundlage für Anlagen zur Steuererklärung verwendet werden.

Compliance-Überwachung: Für die Überwachung steuerlicher Vorgaben können Sollwerte mit Istdaten verglichen werden. Dies betrifft beispielsweise die Kontrolle von Kennzahlen-Schwellenwerten oder die Einhaltung steuerlicher Mindestanforderungen bei verschiedenen Rechtsformen.

Controlling

Budget-Ist-Analysen: Controller nutzen die Vergleichsfunktion für systematische Abweichungsanalysen zwischen Planwerten und tatsächlichen Ergebnissen. Die visuelle Hervorhebung hilft bei der schnellen Identifikation kritischer Bereiche, während die Excel-Exportfunktion detaillierte Berichte für die Geschäftsführung ermöglicht.

Kennzahlen-Überwachung: Bei der Analyse von Produktionskennzahlen, Kostenstellenvergleichen oder Projektcontrolling können Zielwerte mit erreichten Werten systematisch gegenübergestellt werden. Die Funktion "Nur Unterschiede anzeigen" reduziert die Datenmenge auf kritische Abweichungen und erhöht die Effizienz der Analyse.

Konsolidierungsunterstützung: Im Konzerncontrolling unterstützt die Funktion bei der Harmonisierung von Daten aus Tochtergesellschaften. Unterschiedliche Kontierungsrichtlinien oder Buchungszyklen können durch die flexible Spalten-Zuordnung ausgeglichen werden, was die Qualität der Konsolidierung erheblich verbessert.

14.5.2 Erstellung & Auswertung

Nach Betätigung der Funktion „Vergleichen“ öffnet sich das folgende Fenster, mit der Aufforderung eine **zweite Datei** auszuwählen:

Dateien vergleichen

Dateiauswahl

Erste Datei

Testdatei A.parquet

Zweite Datei

Wählen Sie eine Parquet-Datei zum Vergleichen aus...

Vergleichen Abbrechen

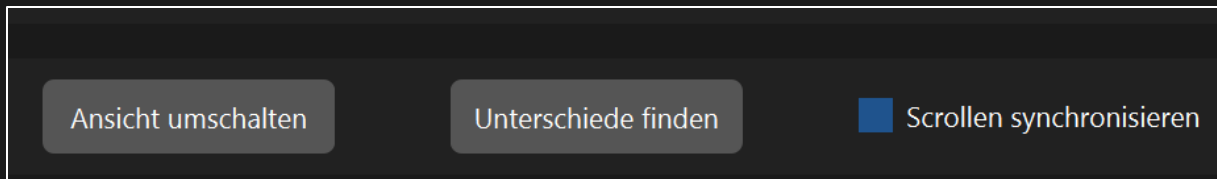
Anschließend werden beide Dateien dargestellt (Vertikal ist Standard, kann aber geändert werden):

Testdatei A.parquet						
ROW ID	ORDER ID	ORDER DATE	SHIP DATE	SHIP MODE	CUSTOMER ID	CUSTOMER NAME
1	CA-2016-152156	00.00.00	11.11.2016	Second Class	CG-12520	Claire Gute
2	CA-2016-152156	00.00.00	11.11.2016	Second Class	CG-12520	Claire Gute
3	CA-2016-138688	00.00.00	16.06.2016	Second Class	DV-13045	Darrin Van Hu
4	US-2015-108966	00.00.00	18.10.2015	Standard Class	SO-20335	Sean O'Donne
5	US-2015-108966	00.00.00	18.10.2015	Standard Class	SO-20335	Sean O'Donne
6	CA-2014-115812	00.00.00	14.06.2014	Standard Class	BH-11710	Brosina Hoff
7	CA-2014-115812	00.00.00	14.06.2014	Standard Class	BH-11710	Brosina Hoff
8	CA-2014-115812	00.00.00	14.06.2014	Standard Class	BH-11710	Brosina Hoff
9	CA-2014-115812	00.00.00	14.06.2014	Standard Class	BH-11710	Brosina Hoff
10	CA-2014-115812	00.00.00	14.06.2014	Standard Class	BH-11710	Brosina Hoff
11	CA-2014-115812	00.00.00	14.06.2014	Standard Class	BH-11710	Brosina Hoff
12	CA-2014-115812	00.00.00	14.06.2014	Standard Class	BH-11710	Brosina Hoff
13	CA-2017-114412	00.00.00	20.04.2017	Standard Class	AA-10480	Andrew Allen
14	CA-2016-161389	00.00.00	10.12.2016	Standard Class	IM-15070	Irene Meddow
15	US-2015-118983	00.00.00	26.11.2015	Standard Class	HP-14815	Harold Pawla
16	US-2015-118983	00.00.00	26.11.2015	Standard Class	HP-14815	Harold Pawla
17	CA-2014-105893	00.00.00	16.11.2014	Standard Class	PK-19075	Pete Kri
18	CA-2014-167164	00.00.00	15.05.2014	Second Class	AG-10270	Alegandro Gro
19	CA-2014-143336	00.00.00	01.09.2014	Second Class	ZD-21925	Zuschuss Dor
20	CA-2014-143336	00.00.00	01.09.2014	Second Class	ZD-21925	Zuschuss Dor
21	CA-2014-143336	00.00.00	01.09.2014	Second Class	ZD-21925	Zuschuss Dor
22	CA-2016-137330	00.00.00	13.12.2016	Standard Class	KB-16585	Ken Black
23	CA-2016-137330	00.00.00	13.12.2016	Standard Class	KB-16585	Ken Black
24	US-2017-156909	00.00.00	18.07.2017	Second Class	SF-20065	Sandra Flana

Testdatei B.parquet						
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
1	8.223	CA-2018-152905	18.02.2018	Standard Class	AB-10015	Aaron Bergman
2	5.355	CA-2018-113766	13.05.2018	Standard Class	AH-10030	Aaron Hawkins
3	3.200	CA-2021-164000	18.12.2021	Standard Class	AH-10030	Aaron Hawkins
4	5.356	CA-2018-113768	13.05.2018	Standard Class	AH-10030	Aaron Hawkins
5	9.681	CA-2019-130113	27.12.2019	Standard Class	AH-10030	Aaron Hawkins
6	9.680	CA-2019-130113	27.12.2019	Standard Class	AH-10030	Aaron Hawkins
7	4.590	US-2018-150126	27.07.2018	Standard Class	AS-10045	Aaron Smayling
8	1.834	CA-2021-162891	01.08.2021	Standard Class	AS-10045	Aaron Smayling
9	1.835	CA-2021-162891	01.08.2021	Standard Class	AS-10045	Aaron Smayling
10	3.040	CA-2021-101749	03.10.2021	Standard Class	AS-10045	Aaron Smayling
11	2.365	US-2020-108637	13.03.2020	Standard Class	AB-10060	Adam Bellavance
12	5.001	CA-2019-150511	18.09.2019	Standard Class	AB-10060	Adam Bellavance
13	5.001	CA-2021-159888	07.05.2021	Standard Class	AB-10060	Adam Bellavance
14	2.366	US-2020-108637	13.03.2020	Standard Class	AB-10060	Adam Bellavance
15	8.171	CA-2021-101774	06.11.2021	Standard Class	AB-10060	Adam Bellavance
16	2.364	US-2020-108637	13.03.2020	Standard Class	AB-10060	Adam Bellavance
17	1.176	CA-2020-161207	29.08.2020	Standard Class	AB-10060	Adam Bellavance
18	8.170	CA-2021-101774	06.11.2021	Standard Class	AB-10060	Adam Bellavance
19	8.169	CA-2021-101774	06.11.2021	Standard Class	AB-10060	Adam Bellavance
20	3.034	CA-2021-134173	16.09.2021	Standard Class	AB-10060	Adam Bellavance
21	4.983	CA-2018-160066	16.11.2018	Standard Class	AH-10075	Adam Hart
22	1.570	CA-2021-165029	26.11.2021	Standard Class	AH-10075	Adam Hart
23	925	CA-2020-149797	15.09.2020	Standard Class	AH-10075	Adam Hart
24	8.415	CA-2020-147109	17.12.2020	Standard Class	AH-10075	Adam Hart

Schließen Ansicht umschalten Unterschiede finden Scrollen synchronisieren

Unter den beiden Tabellen befinden sich neben dem roten Button „**Schließen**“, welcher die Vergleichsfunktion beendet und zur ursprünglichen Ansicht zurückkehrt, zusätzliche Optionen:



Ansicht umschalten:

Diese Funktion ermöglicht es, die Darstellung der verglichenen Dateien jederzeit zwischen **vertikaler** und **horizontaler** Ansicht umzuschalten.

- **Zweck:**

Je nach Bildschirmgröße oder persönlicher Vorliebe kann es sinnvoller sein, die beiden Dateien **nebeneinander (vertikal)** oder **untereinander (horizontal)** anzuzeigen. Diese Umschaltfunktion bietet maximale Flexibilität beim Vergleich.

- **Funktionsweise:**

- **Vertikale Ansicht:**

Eine Datei wird links, die andere rechts angezeigt – ideal für breite Bildschirme (Standardeinstellung)

- **Horizontale Ansicht:**

Eine Datei wird oben, die andere unten angezeigt – hilfreich bei langen Textdateien oder bei beschränkter Breite.

Testdatei A.parquet														
	ROW ID	ORDER ID	ORDER DATE	SHIP DATE	SHIP MODE	CUSTOMER ID	CUSTOMER NAME	SEGMENT	COUNTRY	CITY	STATE	POSTAL CODE	REGION	PRODUCT ID
1	1	CA-2016-152156	00:00:00	11.11.2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-10001798
2	2	CA-2016-152156	00:00:00	11.11.2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10000454
3	3	CA-2016-138688	00:00:00	16.06.2016	Second Class	DV-13045	Darin Van Huff	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10000240
4	4	US-2015-108966	00:00:00	18.10.2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Texas	33311	South	FUR-TA-10000577
5	5	US-2015-108966	00:00:00	18.10.2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Texas	33311	South	OFF-ST-10000760
6	6	CA-2014-115812	00:00:00	14.06.2014	Standard Class	BH-11770	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-10001487
7	7	CA-2014-115812	00:00:00	14.06.2014	Standard Class	BH-11770	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-10002833
8	8	CA-2014-115812	00:00:00	14.06.2014	Standard Class	BH-11770	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-100002275
9	9	CA-2014-115812	00:00:00	14.06.2014	Standard Class	BH-11770	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-100003910
10	10	CA-2014-115812	00:00:00	14.06.2014	Standard Class	BH-11770	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-10002892
11	11	CA-2014-115812	00:00:00	14.06.2014	Standard Class	BH-11770	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-10001539

Testdatei B.parquet															
	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City	State	Postal Code	Region	Product ID	Ca
1	8.223	CA-2018-152905	18.02.2018		Standard Class	AB-10015	Aaron Bergman	Consumer	United States	Arlington	Texas	76.017,0	Central	OFF-ST-10000321	Office Sup
2	5.355	CA-2018-113768	13.05.2018		Standard Class	AH-10030	Aaron Hawkins	Corporate	United States	Los Angeles	California	90.004,0	West	FUR-CH-10000439	Furniture
3	3.206	CA-2021-164000	16.12.2021		Standard Class	AH-10030	Aaron Hawkins	Corporate	United States	Philadelphia	Pennsylvania	19.134,0	East	OFF-AR-10003183	Office Sup
4	5.356	CA-2018-113768	13.05.2018		Standard Class	AH-10030	Aaron Hawkins	Corporate	United States	Los Angeles	California	90.004,0	West	OFF-PA-10003063	Office Sup
5	9.681	CA-2019-130113	27.12.2019		Standard Class	AH-10030	Aaron Hawkins	Corporate	United States	San Francisco	California	94.122,0	West	TEC-PH-10003505	Technology
6	9.680	CA-2019-130113	27.12.2019		Standard Class	AH-10030	Aaron Hawkins	Corporate	United States	San Francisco	California	94.122,0	West	OFF-ST-10000046	Office Sup
7	4.580	US-2018-150126	27.07.2018		Standard Class	AS-10045	Aaron Smayling	Corporate	United States	New York City	New York	10.035,0	East	OFF-PA-10002709	Office Sup
8	1.834	CA-2021-162691	01.08.2021		Standard Class	AS-10045	Aaron Smayling	Corporate	United States	Austin	Texas	78.745,0	Central	TEC-MA-10000488	Technology
9	1.835	CA-2021-162691	01.08.2021		Standard Class	AS-10045	Aaron Smayling	Corporate	United States	Austin	Texas	78.745,0	Central	OFF-PA-10003729	Office Sup
10	3.040	CA-2021-101749	03.10.2021		Standard Class	AS-10045	Aaron Smayling	Corporate	United States	Pasadena	California	91.104,0	West	FUR-TA-10001520	Furniture
11	2.365	US-2020-108637	13.03.2020		Standard Class	AB-10060	Adam Bellavance	Home Office	United States	Waynesboro	Virginia	22.880,0	South	OFF-AR-10000450	Office Sup

Hinweis: Beim Vergleich von zwei Dateien – unabhängig davon, ob sie **horizontal** oder **vertikal** angezeigt werden – kann das Verhältnis der Anzeigeflächen flexibel mit

gedrückter linker Maustaste in der Mitte individuell angepasst werden. Standardisiert sind 50:50 eingestellt.

Scrollen synchronisieren (Funktion unten rechts):

Mit der Funktion „**Scrollen synchronisieren**“ werden beide angezeigten Dateien beim Vergleich gleichzeitig gescrollt, sodass korrespondierende Inhalte stets parallel sichtbar bleiben.

➔ Zweck:

Diese Funktion erleichtert den visuellen Vergleich von Inhalten, insbesondere bei großen Tabellen oder langen Textdateien, in denen sich Unterschiede oft nur in bestimmten Zeilen zeigen. Sie sorgt dafür, dass Sie beim Scrollen in einer Datei automatisch auch an der gleichen Position in der anderen Datei bleiben.

Unterschiede finden:

Diese Funktion unterstützt beim Vergleich zweier Datendateien durch eine visuelle Zuordnung und den Abgleich von Spalten.

Spalten zuordnen

Dateiinformationen

Spalten in der ersten Datei: 22
Spalten in der zweiten Datei: 26

Schnellauswahl

Alle auswählen
Auswahl abgleichen
Alle abwählen

Weitere Optionen

☐ Spalten beistellen, die vom Vergleich ausgeschlossen sind
☐ Spalten als Text vergleichen

Spaltenzuordnung

☐ ROW ID
☐ ORDER ID
☐ ORDER DATE
☐ SHIP DATE
☐ SHIP MODE
☐ CUSTOMER ID
☐ CUSTOMER NAME

OK Abbrechen

Dateiinformationen 1:

Im gezeigten Dialog werden die Spaltennamen aus beiden Dateien einander gegenübergestellt:

- Spalten aus der ersten Datei (z. B. 22 Spalten).
- Spalten aus der zweiten Datei (z. B. 26 Spalten).

Das Ziel ist, **gemeinsame Spalten** zu identifizieren und korrekt zuzuordnen.

Schnellauswahl ②:

- **Alle auswählen:**
Aktiviert alle verfügbaren Spalten für den Vergleich.
- **Auswahl abgleichen:**
Wählt automatisch Spalten mit identischen Namen aus.
- **Alle abwählen:**
Deaktiviert alle Spalten für eine komplett manuelle Auswahl.

Spaltenzuordnung ③:

- Die Zuordnung erfolgt über eine Auswahlliste aller verfügbaren Spalten der zweiten Datei. Spalten ohne Zuordnung werden automatisch mit einem leeren Feld markiert.
- **Farbliche Markierungen** (blaue Kästchen) zeigen an, ob eine Spalte aktiv in den Vergleich miteinbezogen wird.

Weitere Optionen ④:

Spalten beistellen, welche vom Vergleich ausgeschlossen sind: Diese Option ermöglicht es, Spalten aus beiden Dateien, die nicht direkt verglichen werden, trotzdem in der Ergebnisansicht anzuzeigen. Dies ist wertvoll für:

- Referenznummern und Identifikationsspalten
- Kontextinformationen wie Buchungsdaten oder Kostenstellen
- Zusätzliche Informationen, die für die Bewertung der Unterschiede wichtig sind

Spalten als Text vergleichen: Aktiviert einen Vergleichsmodus, der alle Datentypen als Text behandelt. Diese Option hilft bei:

- Unterschiedlichen Zahlenformaten zwischen den Dateien
- Abweichenden Datums- und Zeitdarstellungen
- Rundungsproblemen bei Dezimalzahlen

Nach dem Klick auf den grünen „OK“-Button startet ein mehrstufiger Analyseprozess, der durch eine Fortschrittsanzeige begleitet wird:

Schritt 1: Spalten angleichen (0-20%)

- Harmonisierung der Spaltenstrukturen zwischen beiden Datensätzen
- Anwendung der konfigurierten Zuordnungen
- Überprüfung der Datentypen

Schritt 2: Zeilen zuordnen (20-60%)

- Intelligente Zuordnung gleicher Datensätze zwischen beiden Dateien
- Erkennung verschobener, gelöschter und hinzugefügter Zeilen
- Optimale Gegenüberstellung für den Vergleich

Schritt 3: Datenformate anpassen (60-80%)

- Einheitliche Behandlung von Leer- und Null-Werten
- Konsistente Darstellung aller Datentypen
- Vorbereitung für den eigentlichen Vergleich

Schritt 4: Unterschiede ermitteln (80-100%)

- Zeilenweiser Vergleich der konfigurierten Spalten
- Markierung aller gefundenen Unterschiede
- Erstellung der Ergebnisübersicht

Nach erfolgreichem Abschluss des Vergleichs öffnet sich die **Vergleichsansicht**, die beide Datensätze übersichtlich nebeneinander darstellt:

Hauptkomponenten:

- **Linke Tabelle:** Darstellung der ersten Datei mit Hervorhebung der Unterschiede
- **Rechte Tabelle:** Darstellung der zweiten Datei mit entsprechenden Markierungen
- **Differenz-Übersicht:** Seitliche Miniaturansicht aller Änderungszeilen
- **Gekoppelte Navigation:** Beide Tabellen scrollen synchron

Farbkodierung und Markierungen

Das System verwendet eine intuitive Farbkodierung zur Kennzeichnung von Unterschieden:

Markierungen in den Tabellen:

- **Rote Hinterlegung:** Zellen mit unterschiedlichen Werten zwischen beiden Dateien
- **Grauer Hintergrund:** Spalten, die nicht für den Vergleich konfiguriert wurden
- **Normaler Hintergrund:** Identische Werte in verglichenen Spalten

Differenz-Übersicht (rechter Rand):

- **Grüner Hintergrund:** Bereiche ohne Unterschiede
- **Rote Striche:** Zeilen mit identifizierten Unterschieden
- **Anklickbar:** Direktes Springen zu Problemstellen

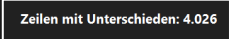
Abhängig von den gesetzten Optionen werden dabei nur bestimmte Spalten berücksichtigt:

Beispiel: Hier wurden alle Positionen abgewählt und nur die ROW ID beider Dateien miteinander verglichen. Das Ergebnis wird wie folgt angezeigt:

Musterfall_Orders.parquet		Test.parquet	
Row ID		Row ID	
24			
25		25	
26			
27			
28		28	
29		29	
30		30	
31		31	

So können leicht Unterschiede (farblich markiert) erkannt werden.

Im unteren Bereich der **Vergleichstabellen** finden Sie zusätzliche Funktionen zur Navigation und Weiterverarbeitung der Ergebnisse:

-  **Roter Schließbutton**
Mit einem Klick auf diesen Button schließen Sie die generelle Vergleichsfunktion und kehren zur ursprünglichen Datenansicht (1. Datei) zurück.
-  **Anzahl der Unterschiede**
Zeigt die Gesamtanzahl der erkannten Unterschiede zwischen den beiden Tabellen an – je nach gewählten Vergleichsoptionen.

Zusätzlich erhalten Sie jedoch weitere Funktionen innerhalb der Unterscheidungsansicht:

Ansicht umschalten	Unterschiede löschen	Nur Unterschiede anzeigen	In Excel öffnen

- **Ansicht umschalten**
Wechselt zwischen verschiedenen Darstellungsformen der Vergleichsergebnisse – z. B. horizontal oder vertikal, um Unterschiede besser zu erkennen.
- **Unterschiede löschen**

Mit einem Klick auf diesen Button kehren Sie zurück zur Startansicht der Vergleichsfunktion. Dabei werden die aktuell angezeigten Vergleichsergebnisse verworfen.

- **Nur Unterschiede anzeigen**

Blendet alle identischen Zeilen aus und zeigt **ausschließlich abweichende Einträge**. So können Sie sich gezielt auf relevante Änderungen konzentrieren.

Musterfall_Orders.parquet	
Row ID	
1	
2	
3	
18	
19	

Hat man diese Ansicht gewählt kann man mit dem Button

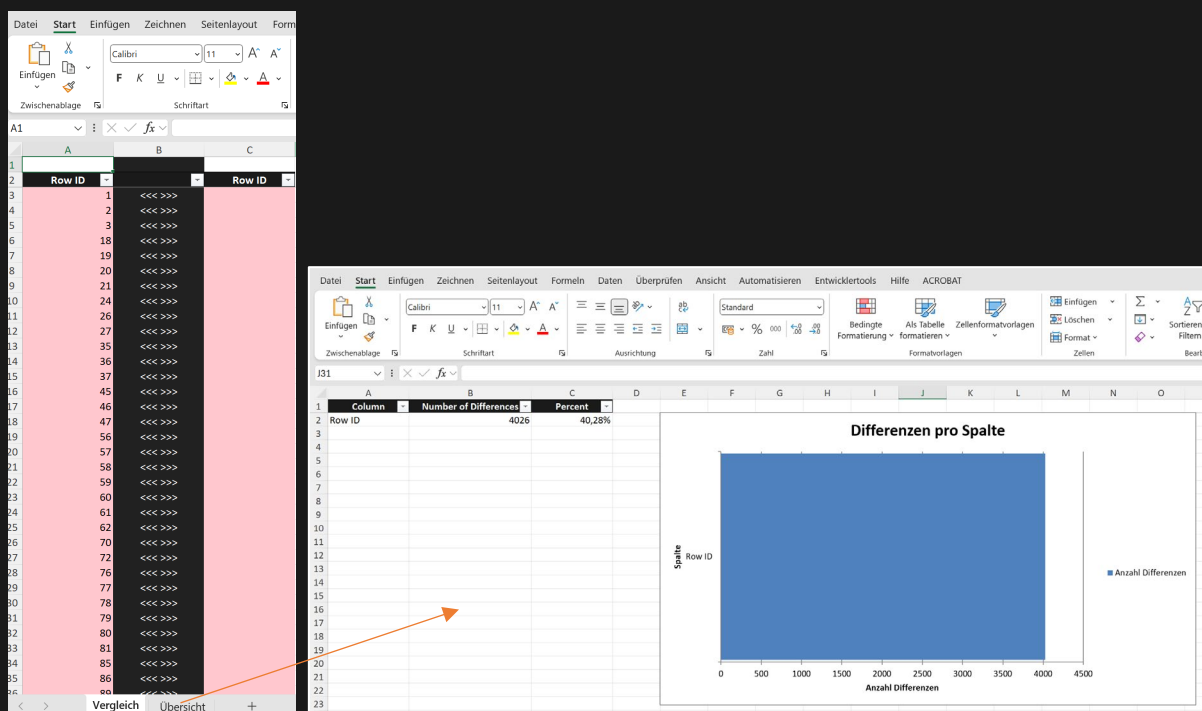
Alle Zeilen anzeigen

wieder zur ursprünglichen Ansicht zurückkehren.

- **In Excel öffnen**

Exportiert das Vergleichsergebnis in eine Excel-Datei und öffnet diese direkt in Microsoft Excel. Ideal zur Weitergabe oder zur Weiterbearbeitung.

Das Ergebnis wird in Excel wie folgt automatisch aufbereitet:



Die Export-Funktion erstellt eine umfassende Excel-Datei mit mehreren Arbeitsblättern für eine vollständige Dokumentation:

Vergleichs-Arbeitsblatt:

- **Nebeneinander-Darstellung:** Beide Datensätze mit sichtbarer Trennlinie
- **Automatische Farbmarkierung:** Alle Unterschiede werden rot hervorgehoben
- **Fixierte Überschriften:** Spaltenköpfe bleiben beim Scrollen sichtbar
- **Filterfunktion:** Jede Spalte kann individuell gefiltert werden

Übersichts-Arbeitsblatt:

- **Spaltenstatistiken:** Anzahl der Unterschiede pro Spalte
- **Prozentangaben:** Anteil der Änderungen bezogen auf alle Datensätze
- **Diagramm:** Balkendiagramm zur visuellen Darstellung der Unterschieds-Verteilung

Automatische Formatierung:

- **Datentyp-gerechte Darstellung:** Zahlen, Datumsangaben und Text werden passend formatiert
- **Optimierte Spaltenbreiten:** Automatische Anpassung an den Inhalt
- **Professionelles Layout:** Einheitliche Gestaltung für Berichte und Präsentationen

14.5.3 *Erweiterte Strategien*

Vorbereitung und Planung

Optimale Datenvorbereitung: Eine erfolgreiche Anwendung der Compare-Funktion beginnt bereits bei der Vorbereitung der Daten. Beide Datensätze sollten über konsistente Schlüsselfelder verfügen, die eine eindeutige Zuordnung ermöglichen. Bei zeitbezogenen Vergleichen empfiehlt sich die vorherige Bereinigung von Dubletten und die Vereinheitlichung von Datums- und Zeitformaten.

Durchdachte Spalten-Auswahl: Die Auswahl der zu vergleichenden Spalten sollte zweckgebunden erfolgen. Für Wertvergleiche sind betragsbezogene Spalten zentral, während beschreibende Informationen (wie Buchungstexte oder Referenzen) als zusätzliche Information einbezogen werden können. Die Option "Spalten ohne Vergleich einbeziehen" ermöglicht eine optimale Balance zwischen Fokussierung auf Wesentliches und vollständiger Information.

Umgang mit großen Datenmengen: Bei sehr umfangreichen Datensätzen sollte der Vergleich strukturiert angegangen werden. Eine Vorab-Segmentierung nach Zeiträumen, Geschäftsbereichen oder anderen relevanten Kriterien reduziert die Verarbeitungszeit und verbessert die Übersichtlichkeit der Ergebnisse.

Ergebnisinterpretation und Weiterverarbeitung

Systematische Bewertung der Unterschiede: Nach dem initialen Vergleich sollten die identifizierten Unterschiede systematisch bewertet werden. Hierbei unterscheidet man

zwischen echten Geschäftsveränderungen, technischen Formatunterschieden und möglichen Datenfehlern. Die Excel-Exportfunktion mit spaltenweisen Statistiken unterstützt diese Bewertung erheblich.

Aufbauende Analysen: Die Ergebnisse der Compare-Funktion bilden häufig den Startpunkt für weiterführende Untersuchungen. Identifizierte Abweichungen können in separaten Arbeitsschritten durch Detailanalysen, Trenduntersuchungen oder Ursachenforschung vertieft werden. Die strukturierte Excel-Ausgabe ermöglicht eine nahtlose Integration in bestehende Analyseprozesse.

Dokumentationsstandards: Für prüfungsrelevante oder compliance-kritische Anwendungen sollten einheitliche Dokumentationsprozesse etabliert werden. Dies umfasst die Archivierung der ursprünglichen Eingabedateien, der Zuordnungs-Konfiguration und der erzeugten Vergleichsergebnisse. Die Excel-Ausgabe sollte mit entsprechenden Metadaten (Vergleichsdatum, verantwortliche Person, Analysezweck) versehen werden.

Integration in den beruflichen Arbeitsablauf

Qualitätssicherung: Die Compare-Funktion sollte als fester Bestandteil von Qualitätssicherungsprozessen verstanden werden. Regelmäßige Vergleiche zwischen Quellsystemen und aufbereiteten Daten helfen dabei, Datenfehler frühzeitig zu erkennen. Besonders bei mandatskritischen Analysen empfiehlt sich eine Kontrolle der Zuordnungs-Konfiguration durch eine zweite Person.

Effizienzsteigerung: Durch die Entwicklung von Standard-Zuordnungsmustern für wiederkehrende Vergleichsszenarien kann die Arbeitseffizienz erheblich gesteigert werden. Häufig verwendete Spalten-Zuordnungen sollten dokumentiert und bei ähnlichen Aufgaben wiederverwendet werden. Die konsequente Anwendung bewährter Muster reduziert Fehlerquellen und beschleunigt den Analyseprozess.

Kommunikation und Berichtswesen: Die visuell aufbereiteten Vergleichsergebnisse eignen sich hervorragend für die Kommunikation mit Mandanten oder der Geschäftsführung. Die intuitive Farbkodierung ermöglicht auch Laien ein schnelles Verständnis der wichtigsten Ergebnisse. Die Excel-Ausgabe kann als Grundlage für Präsentationen oder als Standard-Berichtsformat dienen.

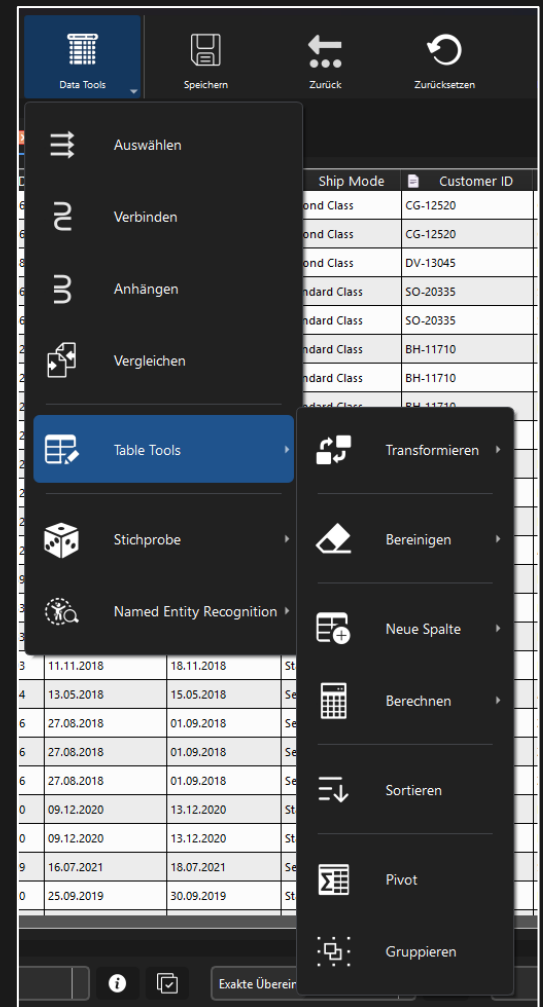
14.6 Table-Tools

Unter **Table Tools** steht Ihnen eine zentrale Sammlung von Funktionen zur Verfügung, um Ihre Daten direkt in der Tabelle zu bearbeiten oder zu analysieren. Mit wenigen Klicks können Sie beispielsweise Spalten transformieren, Berechnungen durchführen, Duplikate aufspüren und entfernen, leere Zeilen entfernen, neue Spalten hinzufügen oder Pivot-Tabellen anlegen. Dadurch lassen sich viele Arbeitsschritte nahtlos in einem einzigen Tool abbilden, ohne dass Sie separate Anwendungen einsetzen müssen.

Die einzelnen Schritte werden in den folgenden Unterkapiteln näher erläutert und mit Screenshots gezeigt.

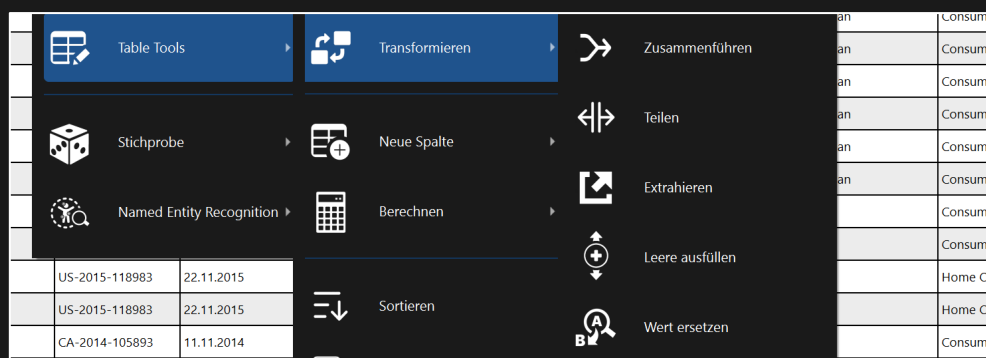


Hinweis: Die Table Tools lassen sich auch über die **Aktionsleiste** (s. unter Kapitel 4.5) per Schnellzugriff öffnen.



14.6.1 Transformieren

Transformieren bietet Ihnen die Möglichkeit, Spalteninhalte gezielt zu bearbeiten und neu zu strukturieren. Dabei können Sie mehrere Spalten zusammenführen, eine bestehende Spalte in mehrere Spalten aufteilen oder bestimmte Teile aus einem Spaltenwert extrahieren. Darüber hinaus können sie leere Zeile mit Werten auffüllen oder bestimmte Werte ersetzen. Diese Funktionen helfen Ihnen dabei, Ihre Daten optimal auf den weiteren Analyseprozess vorzubereiten.



14.6.1.1 Zusammenführen

Mit der Funktion **Zusammenführen** können Sie Inhalte aus zwei Spalten in einer neuen Spalte kombinieren. Dies ist beispielsweise nützlich, wenn Sie eine vollständige Adresse aus mehreren Feldern erzeugen oder andere zusammengehörige Informationen in einer Spalte bündeln möchten. Im gezeigten Dialog stehen Ihnen folgende Eingabemöglichkeiten zur Verfügung:

1. Erste Spalte auswählen / Zweite Spalte auswählen ①

- Wählen Sie hier die beiden Spalten aus, deren Werte Sie zusammenführen möchten. Beispielsweise könnten das eine Ortsangabe (z. B. „City“) und ein Land (z. B. „Country/Region“) sein.

2. Separator / Neuer Spaltenname ②

- **Separator:** Geben Sie das Zeichen an, das zwischen den zusammengeführten Werten stehen soll. Häufig wird ein Schrägstrich (/), ein Komma oder ein Leerzeichen verwendet. Hier können Sie individuell entscheiden.
- **Neuer Spaltenname:** Legen Sie fest, wie die neu erstellte Spalte heißen soll. Hier empfiehlt es sich, einen aussagekräftigen Namen zu vergeben, damit der Schritt klar erkennbar ist (Hinweis: notwendig, ansonsten Fehlerhinweis).

3. Vorschau ③

- Unterhalb der Eingabefelder sehen Sie eine Vorschau auf das Ergebnis der Zusammenführung. Damit können Sie sofort prüfen, ob der gewählte Separator oder Spaltenname Ihren Vorstellungen entspricht.

Abschließend bestätigen Sie Ihre Eingaben mit **OK**, woraufhin ein Fenster erscheint, welches bestätigt, dass eine neue Spalte erstellt wurde, in der die Inhalte aus den beiden Quellspalten entsprechend zusammengeführt sind (**wird ganz hinten angefügt**). Wenn Sie den Vorgang abbrechen möchten, klicken Sie auf **Abbrechen**, ohne dass Änderungen übernommen werden.

Spalten verbinden

Spaltenauswahl

Erste Spalte auswählen:
CITY

Zweite Spalte auswählen:
COUNTRY

Konfiguration

Trennzeichen: - Neuer Spaltenname: Geben Sie den Spaltennamen ein...

Vorschau:

Spalte 1	Spalte 2	Merged
Henderson	United States	Henderson-United States
Henderson	United States	Henderson-United States
Los Angeles	United States	Los Angeles-United States
Fort Lauderdale	United States	Fort Lauderdale-United States
Fort Lauderdale	United States	Fort Lauderdale-United States

OK Abbrechen

Die neue Spalte wurde hinzugefügt.

OK

14.6.1.2 Teilen

Mit der Funktion **Spalte teilen** können Sie den Inhalt einer bestehenden Spalte anhand eines bestimmten Trennzeichens in mehrere Spalten aufsplitten. Das ist besonders hilfreich, wenn beispielsweise kombinierte Informationen getrennt vorliegen sollen. Im gezeigten Dialogfenster stehen Ihnen folgende Eingabemöglichkeiten zur Verfügung:

Spalte teilen

Spaltenauswahl

Spalte auswählen:

CUSTOMER NAME

Trennzeichen-Einstellungen

Trennzeichen:

Geben Sie das Trennzeichen ein (z.B., Komma, Leerzeichen)

Vorschau

Original
Claire Gute
Claire Gute
Darrin Van Huff
Sean O'Donnell
Sean O'Donnell

Teilen Abbrechen

1. Spalte auswählen ①

○ Legen Sie fest, welche Spalte geteilt werden soll (z. B. „Customer Name“).

2. Trennzeichen ②

○ Geben Sie das Zeichen oder die Zeichenfolge an, nach der aufgespalten werden soll (z. B. ein Leerzeichen, Komma oder Semikolon). Im Beispiel ist ein Leerzeichen eingetippt.

3. Vorschau ③

○ Unterhalb der Eingabefelder sehen Sie direkt eine Vorschau darauf, wie die Inhalte aufgeteilt werden. So können Sie überprüfen, ob das gewählte Trennzeichen die Daten korrekt separiert. Im Beispiel werden Vor- und Nachname aufgeteilt.

Wenn Sie mit den Einstellungen zufrieden sind, klicken Sie auf **Teilen**, um die neue Spaltenaufteilung zu übernehmen. Die neuen Spalten werden am **Ende der Tabelle eingefügt** und können mit Rechtsklick auf den Spaltennamen auch unbenannt werden. Möchten Sie keine Änderungen vornehmen, brechen Sie den Vorgang über **Abbrechen** ab.

14.6.1.3 Extrahieren

Mit der Funktion **Extrahieren** können Sie bestimmte Inhalte aus einer vorhandenen Spalte filtern und in einer neuen Spalte ablegen. Dies ist beispielsweise nützlich, wenn Sie aus einem Textfeld nur die numerischen Werte oder andere Teilinformationen benötigen. Im angezeigten Dialog stehen Ihnen folgende Eingabefelder zur Verfügung:

Daten extrahieren

Spalten- und Methodenauswahl

Spalte auswählen:

PRODUCT ID

Option auswählen:

Zahlen

Ausgabekonfiguration

Neuer Spaltenname:

Product_ID Zahlen

Vorschau:

Original	Extracted
FUR-BO-10001798	10001798
FUR-CH-10000454	10000454
OFF-LA-10000240	10000240
FUR-TA-10000577	10000577
OFF-ST-10000760	10000760

Extrahieren

Abbrechen

1. Spalte auswählen ①

Wählen Sie die Spalte aus, aus der Sie die Daten extrahieren möchten (z.B. „Product ID“).

2. Option auswählen ②

Entscheiden Sie, welcher Teil des Spalteninhalts extrahiert werden soll. Je nach Spaltenformat ändern sich die Optionen. Hierbei sind folgende Konstellationen möglich:

Spaltenformat:	Zeichenfolge	Nummern	Datum (+Zeit)	Zeit
Optionen:	- Zahlen - Text - Erstes Wort - Letztes Wort	- Zahlen - Text	- Jahr - Tag - Stunde - Minute - Sekunde - Wochentag	- Jahr - Tag - Stunde - Minute - Sekunde - Wochentag

3. Neuer Spaltenname ③

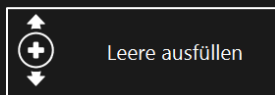
Geben Sie an, wie die neu erstellte Spalte heißen soll. So können Sie die extrahierten Daten eindeutig identifizieren.

4. Vorschau ④

Im Vorschauenfenster sehen Sie, wie die extrahierten Daten aussehen werden. Hier können Sie überprüfen, ob die gewünschte Extraktion (z. B. nur Zahlen) korrekt funktioniert.

Mit einem Klick auf **Extrahieren** wird eine neue Spalte **am Ende der Tabelle** angelegt, die nur die ausgewählten Inhalte enthält. Möchten Sie den Vorgang abbrechen oder keine Änderungen vornehmen, wählen Sie **Abbrechen**.

14.6.1.4 Leere auffüllen



Leere ausfüllen

Die Funktion **Leere Werte füllen** hilft Ihnen, fehlende Einträge in einer Spalte Ihrer Tabelle schnell und zuverlässig zu ergänzen. Statt manuell Zellen zu

kopieren, können Sie hier zwischen zwei einfachen Imputationsmethoden wählen: Forward-Fill (Werte nach unten propagieren) und Backward-Fill (Werte nach oben propagieren).

Bei Klick auf die Funktion öffnet sich das folgende Fenster:

1. Spalte wählen ①

Wählen Sie aus dem Dropdown die Zielspalte aus, in der leere Zellen aufgefüllt werden sollen (im Beispiel die Spalte „Datum“).

2. Füllrichtung ②

- **Nach unten (oben → unten)**

Jeder leere Wert wird mit dem zuletzt darüberstehenden (vorherigen) Wert ersetzt („Forward-Fill“).

Im Beispiel werden die leeren Zeilen mit dem Wert des jeweiligen Datums von oben gefüllt (leere Zeile unter 01.01.2025 mit 01.01.2025)

- **Nach oben (unten → oben)**

Jeder leere Wert wird mit dem nächsten darunter stehenden (nachfolgenden) Wert ersetzt („Backward-Fill“).

Original	Gefüllt
2025-01-01	2025-01-01
NaT	2025-01-01
NaT	2025-01-01
2025-01-02	2025-01-02
NaT	2025-01-02

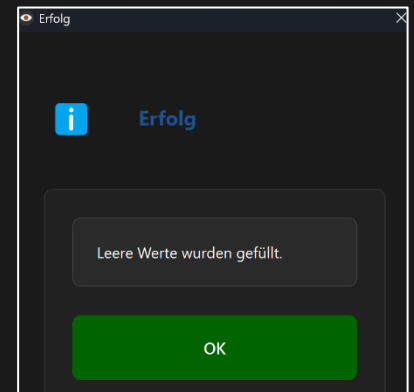
3. Vorschau der Änderungen ③

Im grauen Vorschaufenster erscheint pro Zeile „Alt → Neu“:

- **Alt:** Ursprünglicher Wert (bei leeren Zellen z.B. bei Feldern mit Datum „NaT“ (Zeitstempel - Not a time), bei numerischen Zellen („NaN“ (Not a Number)),
- **Neu:** Wert, mit dem die Lücke gefüllt wird

Diesen Hinweis bestätigen Sie ebenfalls mit **OK**.

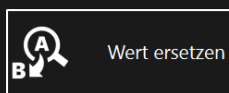
Anschließend wird der Dialog geschlossen und Sie sehen sofort Ihre Tabelle in aktualisierter Form, in der alle zuvor leeren Zellen mit den gewählten Nachbarwerten aufgefüllt wurden.



Beispiel (von oben nach unten gefüllt):

Datum	Wert		Datum	Wert
01.01.2025	20,00		01.01.2025	20,00
	10,00		01.01.2025	10,00
	0,50		01.01.2025	0,50
02.01.2025	60,00		02.01.2025	60,00
	110,00		02.01.2025	110,00
	109,10		02.01.2025	109,10

14.6.1.5 Wert ersetzen



Mit der Funktion „**Werte ersetzen**“ können Sie in Ihren Datentabellen zielgerichtet einzelne Inhalte austauschen – sei es, um Tippfehler zu korrigieren oder Platzhalter durch reale Werte zu ersetzen. Statt mühsam jede Zelle einzeln anzupassen, bietet dieses Werkzeug eine effiziente, transparente und rückverfolgbare Lösung, die Ihnen Arbeitsschritte dokumentiert und vor unerwünschten Änderungen schützt.

Im Programmfenster für die Funktion sind folgende Eintragungen vorzunehmen:

The screenshot shows a dialog box titled "Werte ersetzen" (Replace Values) with a magnifying glass icon. It contains four main sections, each highlighted with an orange circle and a number:

- 1. Spaltenauswahl (Column Selection):** A dropdown menu showing "SUB-CATEGORY".
- 2. Wertekonfiguration (Value Configuration):** Two input fields. The first is labeled "Zu ersetzender Wert:" (Value to be replaced) and contains "Tables". The second is labeled "Ersetzen durch:" (Replace with) and contains "Computer". Both fields have a "Leer" (Empty) checkbox to their right.
- 3. Übereinstimmungstyp (Match Type):** Two radio buttons. The first is "Exakte Übereinstimmung (ganze Zelle)" (Exact match (whole cell)) and is selected. The second is "Enthält (Teilstring)" (Contains (substring)).
- 4. Vorschau (Preview):** A table showing the replacement results.

At the bottom are two buttons: "OK" (green) and "Abbrechen" (red).

Original	Replaced
Bookcases	Bookcases
Chairs	Chairs
Labels	Labels
Tables	Computer
Storage	Storage

1. **Spalte auswählen** ①

Im Dropdown-Menü oben rechts wählen Sie die Zielspalte, z. B. „**Sub-Category**“, in der die Ersetzungen stattfinden sollen.

2. **Alten & Neuen Wert eingeben** ②

Tragen Sie im Feld „**Zu ersetzender Wert**“ exakt den Inhalt ein, der ersetzt werden soll.

- Für **numerische oder Textwerte** geben Sie den entsprechenden Text (z.B. Tables) bzw. die Zahl (z. B. 10,00) ein.
- Für Zeitstempel (Datum oder Zeit) gelten die amerikanischen Formate **YYYY-MM-DD, HH:MM:SS**. Das Ergebnis wird jedoch im deutschen Format angezeigt
- Sofern leere Zeilen ersetzt werden sollen, kann man das Kästchen „Leer“ anklicken, dann wird nach leeren Zeile gesucht.

Unter „**Ersetzen durch**“ tippen Sie den Wert ein, der künftig anstelle des alten Werts stehen soll. Auch hier können Sie „Leer“ anklicken, sofern der Wert durch eine leere Zeile ersetzt werden soll.

3. Übereinstimmungstyp wählen ③

- **Exakte Übereinstimmung (ganze Zelle):** Ersetzt nur Zellen, die in Gänze mit dem alten Wert übereinstimmen (z.B. 10,00 wenn die Zeile auch 10,00 enthält)
- **Enthält (Teilstring):** Ersetzt alle Zellen, in deren Inhalt der alte Wert an beliebiger Stelle vorkommt (z.B. Wert 10, wenn in Zeile 10,00 steht).

4. Vorschau ④

Im Vorschaufenster sehen Sie für jede betroffene Zeile das Schema **Original – Neuer Wert**.

Wichtig das Vorschaufenster ist auf die ersten 5 Zeilen des Datensatzes beschränkt. Ist Ihr zu ersetzender Wert erst später im Datenbestand wird dieser nicht angezeigt, die Funktion wird ihn aber ersetzen.

Mit „**OK**“ erhalten Sie ein Bestätigungsfenster, welches ebenfalls mit „OK“ bestätigt werden muss. Anschließend wird der Dialog geschlossen und Sie sehen sofort Ihre Tabelle mit den angepassten Werten.

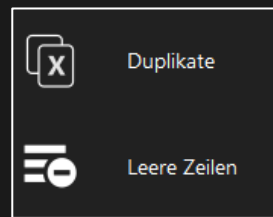
Beispiel: Alter Wert „Tables“ ersetzt durch „Computer“

Vorschau

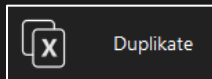
Original	Replaced
Bookcases	Bookcases
Chairs	Chairs
Labels	Labels
Tables	Computer
Storage	Storage

14.6.2 Bereinigen

Die Rubrik „**Bereinigen**“ hilft Ihnen, Ihre Daten effizient aufzuräumen und von unnötigen Einträgen zu befreien. Hier finden Sie Werkzeuge, um typische Probleme wie doppelte Datensätze oder überflüssige Leerzeilen schnell und einfach zu beheben.



14.6.2.1 Duplikate



Duplikate

Die Funktion **Duplikate** hilft Ihnen dabei, doppelte Einträge in Ihren Daten zu identifizieren und gegebenenfalls zu entfernen. Im gezeigten Dialog können Sie verschiedene Einstellungen vornehmen:

1. Aktion ①

- **Duplikate anzeigen:** Listet alle Datensätze auf, die in den ausgewählten Spalten mehrfach vorkommen. So können Sie prüfen, ob und wie viele Duplikate vorhanden sind.
- **Duplikate entfernen:** Löscht alle doppelten Einträge aus den ausgewählten Spalten, sodass jeder Wert nur noch einmal erhalten bleibt.

2. Spalten ②

- Wählen Sie eine oder mehrere Spalten aus, in denen Sie nach Duplikaten suchen möchten. Fügen Sie diese über **Spalte hinzufügen** der Liste hinzu.

- Die ausgewählten Spalten dienen als Grundlage für die Duplikatserkennung (z. B. „Customer Name“).

3. **Ganze Zeile berücksichtigen** (optional) **3**

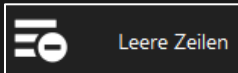
- Ist diese Option aktiviert, wird nicht nur in den ausgewählten Spalten nach Duplikaten gesucht, sondern die gesamte Zeile in die Prüfung einbezogen. Das bedeutet, dass zwei Zeilen nur als Duplikat gelten, wenn alle Werte in allen Spalten übereinstimmen.

4. **Option anwenden** **4**

- **Filter:** Zeigt die gefundenen Duplikate als gefilterte Ansicht an oder blendet sie aus, je nachdem, ob Sie „Duplikate anzeigen“ oder „Duplikate entfernen“ gewählt haben. Die Daten selbst werden dabei nicht dauerhaft verändert. Das anschließende Ergebnis kann als Lesezeichen gespeichert werden (**8** Lesezeichen).
- **Daten überschreiben:** Führt die Aktion direkt auf den Daten durch, indem Duplikate endgültig entfernt werden (bei „Duplikate entfernen“) oder alle Duplikate in eine neue Ansicht übernommen werden (bei „Duplikate anzeigen“).

Abschließend klicken Sie auf **Ausführen**, um die gewählte Aktion anzuwenden, oder auf **Abbrechen**, falls Sie keine Änderungen vornehmen möchten. Diese Funktion unterstützt Sie dabei, Ihre Datensätze sauber zu halten und doppelte Einträge effizient zu handhaben.

14.6.2.2 Leere Zeilen



Leere Zeilen

Mit der Funktion „**Leere Zeilen**“ können Sie gezielt leere oder inhaltlose Zeilen in Ihren Daten identifizieren und entfernen. Dies hilft, die Daten zu bereinigen und unnötige Einträge zu eliminieren.

Mit Klick auf die Funktion erhalten Sie folgendes Dialogfenster

Aktionsauswahl ①

Hier wählen Sie aus, ob Sie:

- **Leere Zeilen entfernen** möchten (die Zeilen werden gelöscht)
- oder **Leere Zeilen anzeigen** möchten (die Zeilen werden gefiltert und nach Abschluss der Funktion angezeigt, ohne sie zu löschen).

Spaltenauswahl ②

Bestimmen Sie, auf welche Spalten die Prüfung erfolgen soll. Über den Button „**Spalte hinzufügen**“ können Sie weitere Spalten angeben.

Ganze Zeile berücksichtigen: eine Zeile wird nur dann als leer gewertet, wenn **alle ausgewählten Spalten leer** sind.

Optionen anwenden ③

Definieren Sie, ob das Ergebnis als Filter angewendet werden soll (Daten bleiben erhalten, nur Ansicht wird gefiltert) oder ob die Daten tatsächlich überschrieben werden. Auch die Überschreibung kann nicht mehr rückgängig gemacht werden.

Leerererkennung ④

Hier legen Sie fest, was als „leer“ gilt:

- **Leere Zeichenfolgen einschließen:** auch Zellen mit "" werden als leer erkannt.

- **Nur Leerzeichen einschließen:** nur Zellen mit Leerzeichen (also keine anderen Inhalte) gelten als leer.

14.6.3 Neue Spalte

In diesem Bereich stehen Ihnen verschiedene Optionen zur Verfügung, um neue Spalten in Ihre Tabelle einzufügen. Diese Funktionen ermöglichen es, Ihre Daten gezielt zu erweitern, zu analysieren oder zu bewerten – je nach Anwendungszweck manuell, regelbasiert oder mithilfe von KI.

Folgende Spaltentypen stehen zur Auswahl:

- **KI-Spalte:** Nutzt Künstliche Intelligenz, um neue Spalten zu erzeugen und/oder bisherige Spalten zu bearbeiten.
- **Benutzerdefinierte Spalte:** Ermöglicht eigene Formeln und Regeln zur automatisierten Berechnung.
- **Editierbare Spalte:** Fügt eine Spalte hinzu, die frei von Hand befüllt und bearbeitet werden kann.
- **Indexspalte:** Erstellt automatisch eine fortlaufende Nummerierung.
- **Prüfspalte:** Dient der gezielten Prüfung einzelner Kriterien oder Auffälligkeiten.



Im weiteren Verlauf dieses Abschnitts finden Sie detaillierte Erläuterungen zu den einzelnen Spaltentypen sowie praxisnahe Anwendungsbeispiele.

14.6.3.1 KI-Spalte

Diese Funktion steht Ihnen nur zur Verfügung, wenn Sie entweder AUDIPY AI Starter oder AUDIPY AI Pro abonniert haben.

Mit der **KI-Spalte** haben Sie die Möglichkeit, mithilfe von AMY, der künstlichen Intelligenz Assistentin im Analyzer komplexe Transformationen oder Extraktionen an Ihren Spalten vorzunehmen, ohne selbst Code schreiben zu müssen. Dabei beschreiben Sie einfach in natürlicher Sprache, wie die Spalte verändert werden soll, und die KI setzt diese Anweisung automatisch um. Im gezeigten Dialog können Sie folgende Schritte durchführen:

1. Spalte auswählen ①

- Wählen Sie zunächst die Spalte aus, die Sie mithilfe der KI transformieren möchten (z. B. „Product ID“).

2. Transformationsbeschreibung eingeben ②

- Geben Sie im Textfeld an, welche Änderungen Sie vornehmen wollen. Formulieren Sie Ihre Anweisung in normaler Sprache, etwa: „Nur das Wort zwischen den Bindestrichen soll als neue Spalte angezeigt werden.“
- Mit **Hinzufügen** wird die Beschreibung in die Liste der gewünschten Transformationen übernommen. Sie können mehrere Transformationen nacheinander definieren, wenn Sie unterschiedliche Änderungen an derselben Spalte vornehmen möchten.

3. Übersicht der Transformationen 3

- Im unteren Bereich wird jede hinzugefügte Anweisung angezeigt (z.B. „Product ID: Nur das Wort zwischen den Bindestrichen soll als neue Spalte angezeigt werden.“).
- Wenn Sie eine Transformation entfernen oder bearbeiten möchten, können Sie diese einfach löschen und bei Bedarf neu hinzufügen.

Nachdem Sie auf „Mit AMY generieren“ geklickt haben, öffnet sich ein Fenster, in dem Sie Ihre neu erzeugte Python-Anweisung überprüfen und anpassen können. Folgende Schritte stehen Ihnen dabei zur Verfügung:

1. Neuer Spaltenname (zwingend) 1

- Geben Sie im Feld *Neuer Spaltenname* einen eindeutigen Namen für die neu zu erstellende Spalte ein (z.B. „Neu_Product ID“).

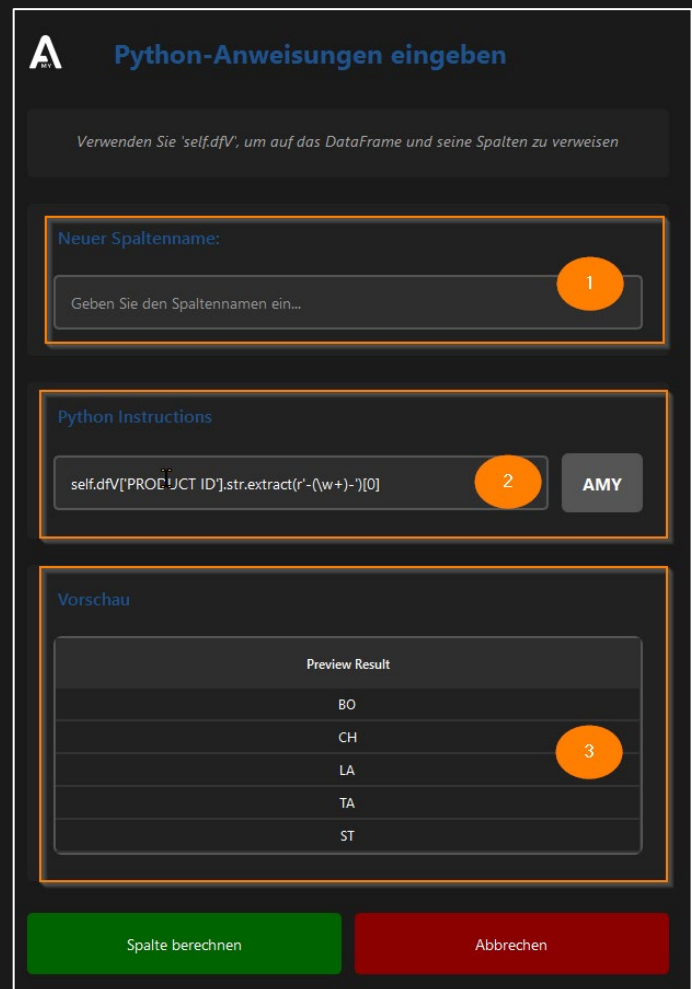
2. Anweisungen (Python-Code) 2

- Das Textfeld *Anweisungen* zeigt den von AMY erzeugten Python-Code, der die gewünschte Transformation an Ihrer Spalte durchführt.
- Sie können den Code bei Bedarf anpassen, um weitere Änderungen vorzunehmen oder die bestehende Logik zu verfeinern.
- Über „AMY“ Sie wieder in den Spaltengenerator von AMY.

3. Vorschau 3

- Im Vorschau-Bereich werden einige Zeilen der neuen Spalte angezeigt, sodass Sie sofort sehen, ob das Ergebnis Ihren Erwartungen entspricht.

Durch die **KI-Spalte**-Funktion können Sie Ihre Daten intuitiv und schnell anpassen, selbst wenn es sich um komplexe oder mehrstufige Transformationen handelt. Dies spart Zeit und verringert die Fehleranfälligkeit, da die KI-Anweisungen direkt in den entsprechenden Code oder Transformationsschritt übersetzt werden.



Python-Anweisungen eingeben

Verwenden Sie 'self.dfV', um auf das DataFrame und seine Spalten zu verweisen

Neuer Spaltenname:

Geben Sie den Spaltennamen ein...

Python Instructions

self.dfV["PRODUCT ID"].str.extract(r'-(w+)-')[0]

AMY

Vorschau

Preview Result
BO
CH
LA
TA
ST

Spalte berechnen **Abbrechen**

14.6.3.2 Benutzerdefinierte Spalte

Mit der **Benutzerdefinierten Spalte** können Sie eigenen Python-Code eingeben, um komplexere Berechnungen oder Datenmanipulationen durchzuführen, die über die vordefinierten Funktionen hinausgehen. Im angezeigten Dialog stehen Ihnen folgende Elemente zur Verfügung:

1. Neuer Spaltenname ①

- Vergeben Sie hier den Namen, unter dem die neu erstellte Spalte in Ihrer Tabelle erscheinen soll (z. B. „Umsatz“).
- Wählen Sie eine aussagekräftige Bezeichnung, damit Sie die Spalte später leicht identifizieren können.

2. Anweisungen (Python-Code) ②

- In diesem Feld fügen Sie den Code ein, der Ihre gewünschte Berechnung oder Transformation umsetzt.
- Typischerweise greifen Sie dabei mit `self.dfV['Spaltenname']` auf vorhandene Spalten zu und definieren die Operation, z.B. `self.dfV['Menge'] * self.dfV['Einzelverkaufspreis']`.

3. AI – Prompt Generator ③

- Über den Button **AI** gelangen Sie in den **Prompt Generator**, wo Sie die gewünschte Funktion in natürlicher Sprache beschreiben können.
- Die KI erzeugt daraus automatisch den passenden Python-Code, den Sie anschließend in das Feld **Anweisungen** ② einfügen können.
- Dies erleichtert die Erstellung komplexerer Berechnungen, ohne dass Sie selbst programmieren müssen. Beispiel:

Python-Anweisungen eingeben

Verwenden Sie 'self.dfV', um auf das DataFrame und seine Spalten zu verweisen

Neuer Spaltenname:

Umsatz

Python Instructions

```
self.dfV['Menge'].astype(float) * self.dfV['Verkauf'].astype(float)
```

AI

Vorschau

Preview Result
523.92
2195.82
29.24
4787.9000000000001
44.74

Spalte berechnen Abbrechen

Prompt-Generator

Wählen Sie die Spalte und geben Sie die Transformationsbeschreibung ein:

Spalte & Anweisung

Anweisung ohne Spalte | Geben Sie die Transformationsbeschreibung ein... |  | **Hinzufügen**

Anweisungen

Die Spalte Menge soll mit der Spalte Verkauf multipliziert werden 

Generieren **Abbrechen**

Bei Auswahl „Generieren“ erscheint ein neues Fenster, mit dem man nun verschiedene Auswahlmöglichkeiten zur Verfügung hat:

Generated Prompt

You are an expert in Python and Pandas. You have a Pandas 2.0 DataFrame called 'self.dFV'. Generate only the expression that comes after the = in an assignment statement. Do not include variable names or the assignment itself-only the right-hand side of the expression. Please help me to generate this column(s) with the following instructions:

Die Spalte Menge soll mit der Spalte Verkauf multipliziert werden

Gib nur den Python-Code für die Transformation an und konzentriere dich dabei auf den Ausdruck rechts vom Gleichheitszeichen.

1

Aktionen

In die Zwischenablage kopieren

Chat GPT | **Online-KI öffnen**

Mit AMY generieren **2**

Schließen

Das Dialogfenster „**Prompt-Vorschau**“ zeigt eine Vorschau des automatisch generierten Prompt-Texts, der an eine KI-Instanz (z. B. ChatGPT) gesendet wird. Es dient zur Überprüfung und ggf. Anpassung des Prompts vor der Ausführung.

Bestandteile des Dialogfensters **1**:

- **Überschrift:**
„Prompt-Vorschau“ – Kennzeichnet das Fenster eindeutig.

- **Textbereich „Erstellter Prompt“:**

Hier wird der generierte Prompt angezeigt, einschließlich aller relevanten Anweisungen in englischer und deutscher Sprache. Der Text erklärt der KI, dass ein Ausdruck zur Berechnung einer neuen Spalte in einem Pandas-DataFrame erstellt werden soll. Besonderer Fokus liegt auf der reinen Ausgabe des rechten Ausdrucksteils einer Zuweisung.

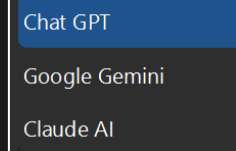
- **Funktionsschaltflächen ②:**

- **„In die Zwischenablage kopieren“**

Kopiert den angezeigten Prompt in die Zwischenablage des Betriebssystems, damit Sie anschließend nur noch die Online KI aufrufen und den Befehl eingegeben müssen.

- **„Modellauswahl“ / „Online-KI öffnen“**

Startet die Online-KI-Schnittstelle zur weiteren Bearbeitung bzw. Kommunikation mit dem KI-Modell. Hierbei können Sie aus den Modellen auswählen, welche Sie unter den Einstellungen hinterlegt haben (s.3.9.1.5). Grundsätzlich sind Chat GPT, Google Gemini und Claude AI hinterlegt.



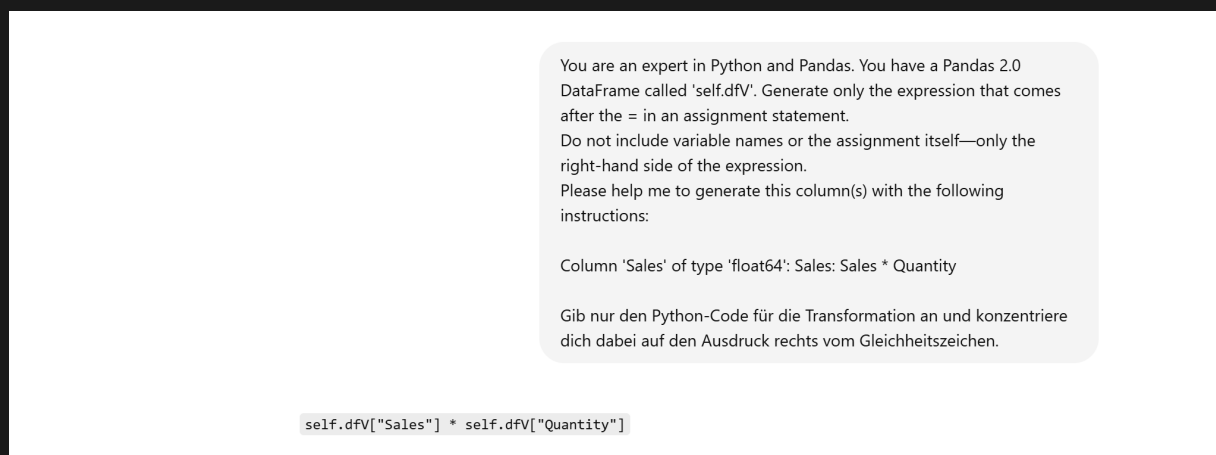
- **„Mit AMY generieren“**

Mit AMY geht der gesamte Schritt schneller, da man nicht vorher in die anderen Online Dienste muss. Bitte beachten diese Funktion ist nur bei den Abo-Modellen AI Starter und AI Pro inklusive.

- **„Schließen“**

Schließt das Dialogfenster ohne weitere Aktion.

Beispiel: Öffnen Sie z.B. Chat GPT und haben den Teil in die Zwischenablage kopiert und im Chatbereich eingefügt erhalten Sie folgendes Ergebnis:



Die Formel kann nun kopiert und unter Anweisung im Generator (s. Bild auf Seite 187) eingefügt werden. Fertig ist Ihre benutzerdefinierte Formel. Schneller geht es aber

über die Funktion: Mit Open AI. Hier wird automatisch die Formel hinterlegt (bitte beachten Sie dass dies evtl. abhängig ist, welche Abo-Modell Sie gewählt haben ist).



Python Instructions

`self.dfV['Menge'].astype(float) * self.dfV['Verkauf'].astype(float)`

AI

4. Vorschau ④ (s. Bild unter 14.1.6.3.2)

- Hier sehen Sie eine kleine Vorschau der neu berechneten Spalte. So können Sie vorab prüfen, ob Ihr Code wie beabsichtigt funktioniert und die gewünschten Werte liefert.

Abschließend klicken Sie auf **Spalte berechnen**, um Ihre Berechnung auszuführen und die neue Spalte in die Tabelle zu übernehmen. Möchten Sie keine Änderungen vornehmen, wählen Sie **Abbrechen**, um den Vorgang zu beenden, ohne eine neue Spalte anzulegen.

14.6.3.3 Editierbare Spalte

Mit dieser Funktion erstellen Sie eine **neue Spalte**, die auf dem Inhalt einer bereits bestehenden Spalte basiert – jedoch **vollständig editierbar** ist. So können Sie bestehende Daten übernehmen und anschließend manuell anpassen, ergänzen oder korrigieren, ohne die Originalspalte zu verändern.

Optionen im Überblick:

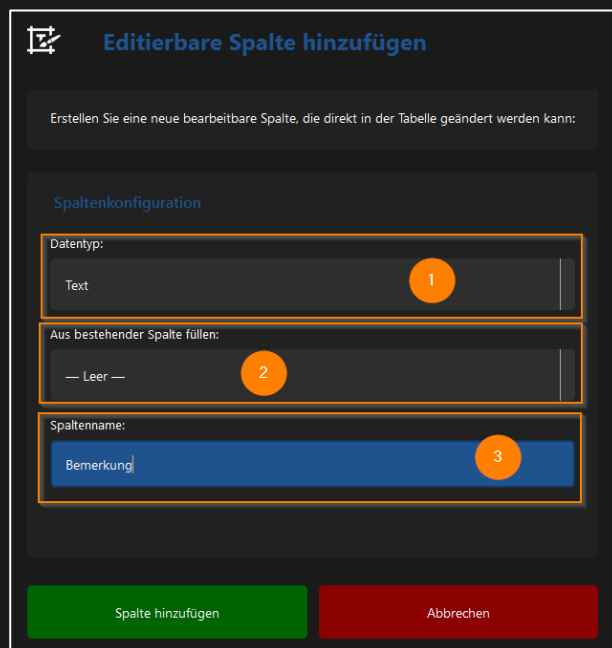
1. Datentyp ①:

Legen Sie fest, ob die neue Spalte als

- **Text** (z. B. für Kommentare oder Bezeichnungen) oder
- **Numerisch** (z. B. für Bewertungen oder Werteanpassungen) verwendet werden soll.

2. Aus bestehender Spalte füllen ②:

Wählen Sie hier eine vorhandene Spalte aus, deren Inhalte in die neue editierbare Spalte übernommen werden sollen. Die Werte werden beim Erstellen kopiert – Änderungen wirken sich ausschließlich auf die neue Spalte aus.



Editierbare Spalte hinzufügen

Erstellen Sie eine neue bearbeitbare Spalte, die direkt in der Tabelle geändert werden kann:

Spaltenkonfiguration

Datentyp:
Text ①

Aus bestehender Spalte füllen:
— Leer — ②

Spaltenname:
Bemerkung ③

Spalte hinzufügen Abbrechen

Wählen Sie „Leer“ aus (wie im Beispiel), wenn Sie freie Eintragungen vornehmen möchten.

3. **Spaltenname 3:**

Geben Sie den Namen der neuen Spalte ein, unter dem sie in der Tabelle angezeigt wird. Im Beispiel soll die Spalte „Bemerkung“ heißen.

Nach dem Klick auf **„Spalte hinzufügen“** wird die editierbare Spalte mit den übernommenen Werten eingefügt und kann direkt angepasst werden – z. B. für individuelle Bewertungen, Korrekturen oder Anmerkungen.

 PROFIT	 Bemerkung
41,9136	
219,582	
6,8714	
-383,031	

14.6.3.4 Indexspalte hinzufügen

Mit der Funktion **Indexspalte hinzufügen** können Sie eine neue Spalte erstellen, die als eindeutiger Index für Ihre Datensätze dient. Diese Spalte kann beispielsweise verwendet werden, um jeden Datensatz eindeutig zu kennzeichnen oder eine Reihenfolge festzulegen. Im Dialog stehen Ihnen folgende Optionen zur Verfügung:

1. Startwert des Index ①

- Geben Sie an, mit welcher Zahl die Indexspalte beginnen soll (z. B. 1).
- Jeder folgende Datensatz erhält dann fortlaufend den nächsthöheren Wert.

2. Verfügbare Spalten (optional) ②

- Sie können die Daten vor dem Hinzufügen der Indexspalte nach bestimmten Spalten sortieren. Fügen Sie hierzu eine oder mehrere Spalten über **Spalte hinzufügen** hinzu und wählen Sie die Sortierreihenfolge (aufsteigend oder absteigend).
- Dadurch werden die Datensätze vor der Indexierung neu angeordnet, sodass der Index mit einer bestimmten Sortierung verknüpft ist.

Index hinzufügen

Eine neue Indexspalte zum DataFrame hinzufügen

Indexkonfiguration

Startwert des Index:

1

Optional: DataFrame vor dem Hinzufügen des Index sortieren

PRODUCT NAME Spalte hinzufügen

PRODUCT NAME Aufsteigend

Index hinzufügen Abbrechen

Auf diese Weise erhalten Sie eine zusätzliche, fortlaufende Spalte, die Sie für Analysen, Filterungen oder eine eindeutige Zuordnung Ihrer Datensätze nutzen können.

14.6.3.5 Prüfspalte hinzufügen

Die **Prüfspalte** ermöglicht es, eine zusätzliche Spalte in Ihre Tabelle einzufügen, mit der Sie jeden Eintrag individuell kennzeichnen können. Dafür stehen drei Zustände zur Verfügung, die Sie durch einfaches Klicken auf die jeweilige Zelle wechseln:

- **X**
- **✓**
- **Leer**

Um die Prüfspalte anzulegen, vergeben Sie zunächst einen **Namen** für die neue Spalte. Anschließend wird diese in der Tabelle hinzugefügt, sodass Sie bei Bedarf jeden einzelnen Datensatz manuell bewerten oder kommentieren können. Dies ist besonders hilfreich für Qualitätskontrollen, manuelle Freigaben oder andere Überprüfungsprozesse, bei denen Sie Einträge schnell und unkompliziert markieren möchten.

Die Prüfspalte wird am Ende der Tabelle eingefügt und ist blau hinterlegt. **Die blaue Farbe bedeutet, dass die Spalte durch den Nutzer editierbar ist.** Mit Klick in eine Zeile wird direkt ein Haken **✓** angezeigt. Mit nochmaligen Klicks erscheint das **X** usw.

Prüfspalte hinzufügen

Prüfspalte hinzufügen

Wie es funktioniert

Geben Sie den Namen der neuen Spalte ein. In dieser Spalte können Sie den Zustand jeder Zelle durch Klicken darauf umschalten. Die Zustände sind:

- Checked
- Unchecked
- Leer

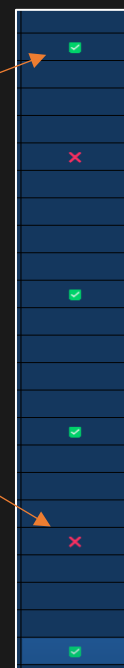
Spaltenkonfiguration

Spaltenname:

Spalte hinzufügen

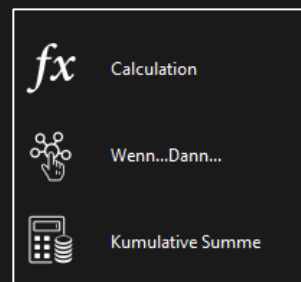
Abbrechen

Sales	Quantity	Discount	Profit	Prüfung
261,96	2	0,00	41,91	
731,94	3	0,00	219,58	
14,62	2	0,00	6,87	
957,58	5	0,45	-383,03	
22,37	2	0,20	2,52	
48,86	7	0,00	14,17	
7,28	4	0,00	1,97	
907,15	6	0,20	90,72	
18,50	3	0,20	5,78	
114,90	5	0,00	34,47	
1.706,18	9	0,20	85,31	
911,42	4	0,20	68,36	
15,55	3	0,20	5,44	
407,98	3	0,20	132,59	



14.6.4 Berechnen

Unter **Berechnen** finden Sie verschiedene Werkzeuge, um neue Spalten in Ihrer Tabelle zu erzeugen oder bestehende Werte mithilfe von Formeln und Regeln zu verarbeiten. Hierzu gehören beispielsweise die **Calculation**-Funktion, bedingte Spalten mithilfe von **Wenn...Dann...**, das Bilden einer **Kumulativen Summe**. Diese Möglichkeiten erlauben es Ihnen, komplexe Berechnungen direkt in der Anwendung durchzuführen, ohne dass Sie auf externe Tools oder manuelle Nachbearbeitung angewiesen sind. Diese werden in den folgenden Kapiteln genauer beschrieben:



14.6.4.1 Spalte berechnen

Die **Spalte berechnen**-Funktion ermöglicht es Ihnen, eine neue Spalte basierend auf bestehenden Daten zu erstellen oder vorhandene Werte zu berechnen. Dabei stehen Ihnen mehrere Vorgehensweisen und Methoden zur Auswahl (s. **Rechteck**), um die Berechnungen flexibel an Ihre Anforderungen anzupassen. Im Folgenden finden Sie eine ausführliche Beschreibung der wichtigsten Schritte in den jeweiligen Kategorien.

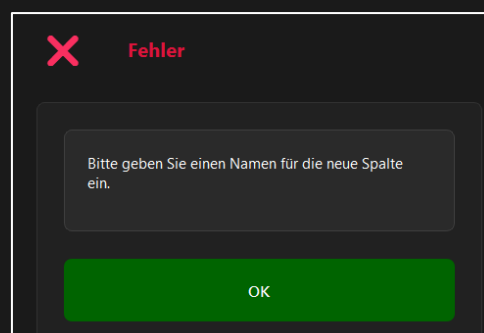
14.6.4.1.1 Wert berechnen

Nutzen Sie **vordefinierte Rechenoperationen**, um zum Beispiel Zahlen zu addieren, zu subtrahieren, zu multiplizieren oder zu dividieren.

The screenshot shows the 'fx Spalte berechnen' (fx Calculate Column) dialog box. It has a title bar with 'fx' and 'Spalte berechnen'. Below the title bar is a text input field labeled 'Neuer Spaltenname:' with the placeholder text 'Geben Sie den Spaltennamen ein...'. This field is highlighted with an orange box and a callout '1'. Below this is a row of five buttons: 'Wert berechnen' (highlighted), 'Statistische Werte', 'Datum/Zeit', 'Zeilen verschieben', and 'Formel eingeben'. Below the buttons is a section with two dropdown menus. The first is labeled 'Methode:' and has 'Absolutwert' selected; it is highlighted with an orange box and a callout '2'. The second is labeled 'Spalte auswählen:' and has 'Row ID' selected; it is highlighted with an orange box and a callout '3'. Below these is a preview area labeled 'Vorschau Ergebnis:' showing 'abs(Row ID) = 2698', highlighted with an orange box and a callout '4'. At the bottom are two buttons: 'Berechnen' (green) and 'Abbrechen' (red).

Neuer Spaltenname 1

- Vergeben Sie im Feld *Neuer Spaltenname* einen eindeutigen und aussagekräftigen Namen für Ihre neue Spalte (z. B. „Umsatz“).
- So können Sie später schnell erkennen, wofür die neu erzeugte Spalte verwendet wird.
- Wird kein Spaltenname ausgewählt, kann auch keine neue Spalte hinzugefügt werden. Sie erhalten dann einen entsprechenden Hinweis:



2. Methode wählen ②

- In der Dropdown-Liste stehen Ihnen verschiedene Rechenmethoden zur Verfügung.
- Zusätzlich legen Sie fest, welche Datenquelle für die Berechnung herangezogen werden soll. Dabei können Sie im Feld **Option** entweder eine **Spalte** auswählen oder einen festen **Wert** eingeben.

Absolutwert
Addition
Subtrahieren
Multiplizieren
Dividieren
Prozent von
Runden
Abrunden
Aufrunden

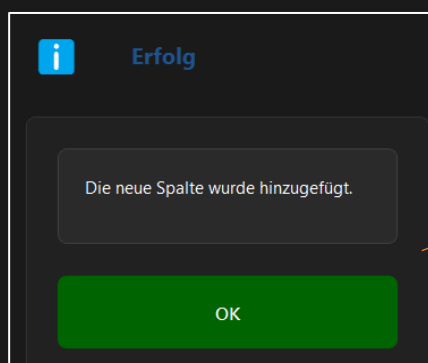
- **Spalte:** Verwenden Sie eine andere Spalte als zweiten Faktor oder Summand (z. B. Einzelverkaufspreis * Menge).
- **Wert:** Geben Sie einen konstanten Wert an, der auf alle Zeilen angewendet wird (z. B. ein pauschaler Aufschlag von 10).

3. Vorschaufenster ③

- Unterhalb der Auswahl sehen Sie eine Vorschau auf das Ergebnis der Berechnung. Hier können Sie kontrollieren, ob die gewählte Methode und die eingegebenen Werte das gewünschte Resultat liefern. Im Beispiel wurde die Spalte „Sales“ mit der Spalte „Quantity“ multipliziert.

Vorschau Ergebnis: Sales * Quantity = 523.92

Mit einem Klick auf **Berechnen** wird eine neue Spalte **am Ende der Tabelle** angelegt, welche die Ergebnisse beinhaltet. Vorher erscheint noch ein Bestätigungsfenster, dass die neue Spalte angelegt wurde.



12 3%	Sales	12 3%	Quantity	12 3%	Umsatz
	261,96		2		523,92
	731,94		3		2.195,82
	14,62		2		29,24
	957,58		5		4.787,89
	22,37		2		44,74
	48,86		7		342,02
	7,28		4		29,12

Möchten Sie den Vorgang abbrechen, wählen Sie **Abbrechen**.

Durch diese flexible Vorgehensweise lassen sich sowohl einfache als auch komplexe Rechenoperationen durchführen, ohne dass Sie externe Tools benötigen. So behalten Sie stets die Kontrolle über Ihre Daten und können schnell neue Kennzahlen oder Berechnungen direkt in der Tabelle abbilden.

14.6.4.1.2 Statistische Werte:

Die Funktion „**statistische Werte**“ beschreibt typischerweise eine Methode, die grundlegende statistische Kennzahlen oder Maße aus einem gegebenen Datensatz berechnet. Solche Werte dienen dazu, einen Datensatz zusammenzufassen, zu charakterisieren und interpretierbar zu machen.

1. Neuer Spaltenname ❶

- Vergeben Sie im Feld *Neuer Spaltenname* einen eindeutigen und aussagekräftigen Namen für Ihre neue Spalte.
- Ohne Spaltennamen kann keine Spalte hinzugefügt werden.

2. Methode wählen ❷

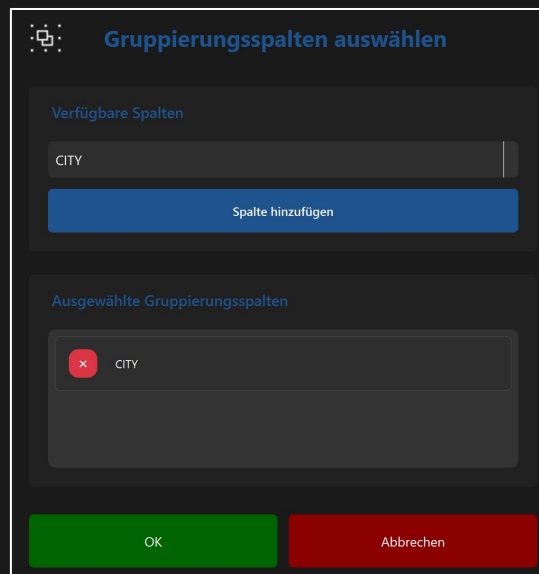
- In der Dropdown-Liste stehen Ihnen verschiedene Methoden zur Verfügung

Summe
Minimum
Maximum
Mittelwert
Median
Standardabweichung
Anzahl
Einzigartige zählen

3. Spalte auswählen ❸

- Zusätzlich legen Sie fest, welche Datenquelle für die Berechnung herangezogen werden soll. Dabei können Sie eine **Spalte** auswählen und optional die Funktion Gruppieren nach benutzen. Hiermit werden die Schritte „Gruppierung“ und „Verbinden“ in einem einzigen Vorgang kombiniert, indem die ausgewählten Elemente gleichzeitig zu einer Gruppe zusammengefasst und unmittelbar miteinander verbunden

werden. Es öffnet sich dann ein neues Fenster, in dem Sie beliebig die vorhandenen Spalten gruppieren können.



Was bedeutet Gruppieren?

Gruppieren (Group By) bedeutet, dass Datensätze basierend auf gleichen Werten in einer oder mehreren ausgewählten Spalten zusammengefasst werden. Anschließend werden auf diesen Gruppen meist aggregierte Werte berechnet.

Beispiel: Angenommen Sie haben folgende Daten:

Region	Umsatz
Nord	100
Süd	200
Nord	150
Süd	100

Wenn Sie neben der Summe von der Spalte „Umsatz“ nach der Spalte „Region“ gruppieren, erhalten Sie etwa Folgendes (je nach Aggregationsfunktion):

Region	Umsatzsumme
Nord	250
Süd	300

Sie fassen also alle Datensätze einer Region zusammen und berechnen daraus Werte wie Summen, Durchschnittswerte, Min/Max etc.

4. Vorschaufenster ④

Unterhalb der Auswahl sehen Sie eine Vorschau auf das Ergebnis der Berechnung. Im Beispiel wurde der Mittelwert von der Spalte „Sales“ berechnet:

Vorschau Ergebnis: $\text{mean}(\text{Sales}) = 229.85800083049833$

14.6.4.1.3 Datum / Zeit

Möchte Sie mit Datum / Zeitspalten rechnen (z.B. die Differenz von 2 Datumsspalten) können Sie die Rubrik Datum/Zeit auswählen. Anschließend stehen Ihnen folgende Möglichkeiten zur Verfügung:

fx Spalte berechnen

Neuer Spaltenname: ①

Wert berechnen Statistische Werte **Datum/Zeit** Zeilen verschieben Formel eingeben

Methode: Differenz ②

Spalte 1: Order Date Spalte 2: Order Date ③ Einheit: Tage

Vorschau Ergebnis: Order Date – Order Date = 0.00 Tage ④

Berechnen Abbrechen

1. Neuer Spaltenname ①

- Vergeben Sie im Feld *Neuer Spaltenname* einen eindeutigen und aussagekräftigen Namen für Ihre neue Spalte.
- Ohne Spaltennamen kann keine Spalte hinzugefügt werden.

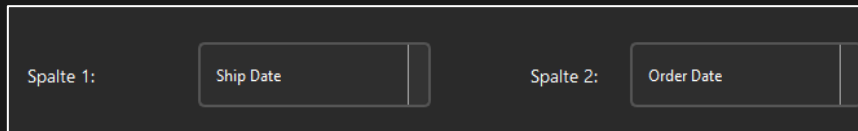
2. Methode wählen ②

- In der Dropdown-Liste stehen Ihnen verschiedene Rechenmethoden zur Verfügung

- Differenz
- Dauer hinzufügen
- Dauer abziehen
- Teil extrahieren
- Zeitstempel (Datum und Uhrzeit)

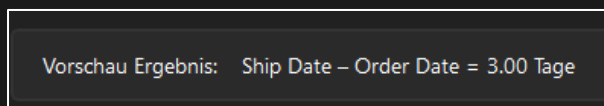
3. Spalte auswählen 3

- Zusätzlich legen Sie fest, welche Datenquelle für die Berechnung herangezogen werden soll. Dabei können Sie die entsprechenden **Spalten** auswählen und optional die Einheit fürs Ergebnis (Tage, Stunden, Minuten, Sekunden) einstellen.



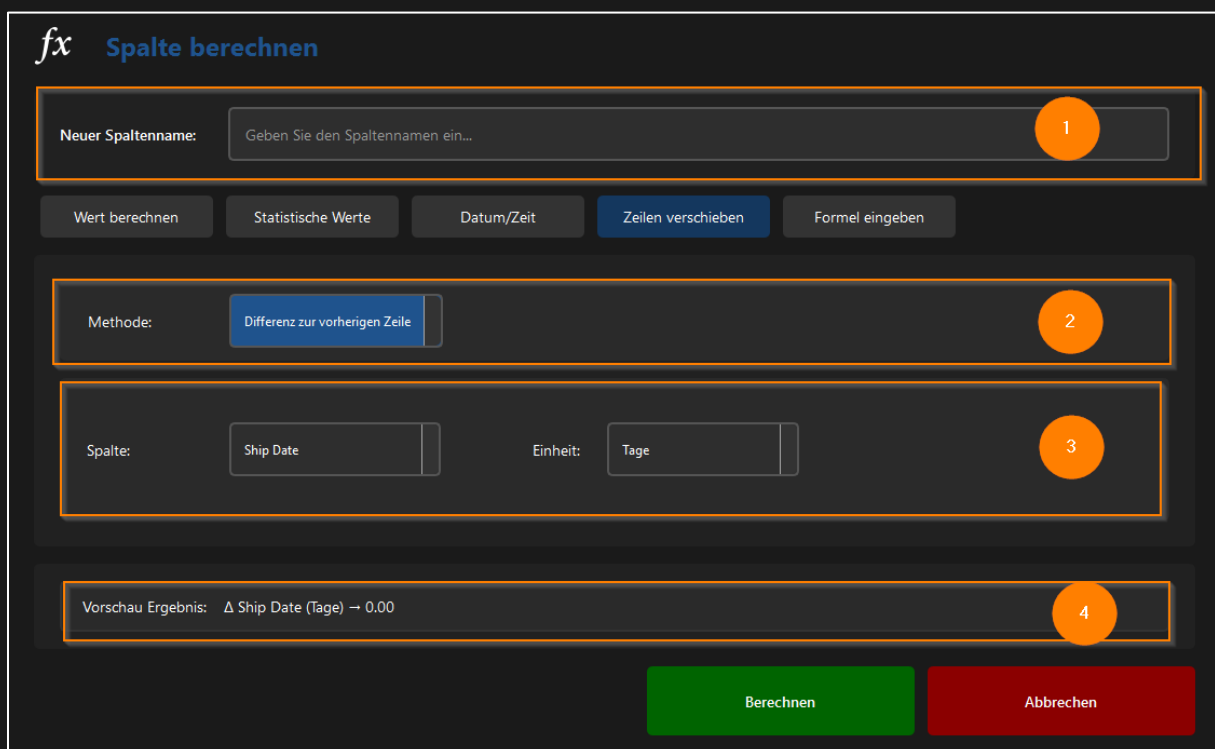
4. Vorschaufenster 4

Unterhalb der Auswahl sehen Sie eine Vorschau auf das Ergebnis der Berechnung. Im Beispiel wurde der Mittelwert von der Spalte „Sales“ berechnet:



14.6.4.1.4 Zeilen verschieben

Die Funktion Zeilen verschieben lässt Zeilen in der neuen Spalte im Gegensatz zur Ursprungsspalte nach unten oder oben verschieben. Gleichzeitig kann hier aber auch eine Differenz zur vorherigen / nächsten Zeile berechnet werden.

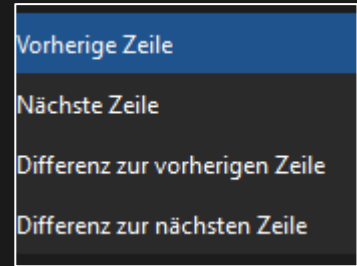


1. Neuer Spaltenname 1

- Vergeben Sie im Feld *Neuer Spaltenname* einen eindeutigen und aussagekräftigen Namen für Ihre neue Spalte.
- Ohne Spaltennamen kann keine Spalte hinzugefügt werden.

2. Methode wählen ②

- In der Dropdown-Liste stehen Ihnen verschiedene Methoden zur Verfügung



3. Spalte auswählen ③

- Zusätzlich legen Sie fest, welche Datenquelle für die Berechnung herangezogen werden soll. Dabei können Sie die entsprechenden **Spalte** auswählen und bei den Methoden „Differenz“ optional die Einheit fürs Ergebnis (Tage, Stunden, Minuten, Sekunden) einstellen.

4. Vorschaufenster ④

Unterhalb der Auswahl sehen Sie eine Vorschau auf das Ergebnis der Berechnung. Im Beispiel wurde der Mittelwert von der Spalte „Sales“ berechnet:

Vorschau Ergebnis: Order Date.shift(-1) → 2020-11-08 00:00:00

14.6.4.1.5 Formel eingeben:

Hier entfällt die Auswahl einer Methode. Definieren Sie eine **individuelle Formel**, wenn die vordefinierten Methoden nicht ausreichen oder Sie eine komplexere Berechnung benötigen.

fx Spalte berechnen

Neuer Spaltenname: 1

Wert berechnen Statistische Werte Datum/Zeit Zeilen verschieben **Formel eingeben**

Verwenden Sie '@', um auf Spalten zu verweisen: @A + @B * 2 2

Formel: f(x)

Vorschau Ergebnis: Berechnung nicht möglich, Formel prüfen 3

Berechnen Abbrechen

1. Neuer Spaltenname 1

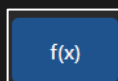
- Vergeben Sie im Feld *Neuer Spaltenname* einen eindeutigen und aussagekräftigen Namen für Ihre neue Spalte (z. B. „Umsatz“).
- Ohne Spaltenname kann keine neue Spalte hinzugefügt werden.

2. Formel eingeben 2

In der Zeile Formel können Sie mithilfe des Zeichen '@', auf bestimmte Spalten im Datenbestand zugreifen (z. B. @SpalteA + @SpalteB * 2). Die werden dann in einem kleinen Vorschaufenster auch angezeigt.

Im Screenshot wurden die Spalten „Sales“ und „Quantity“ multipliziert.

Formel:



Unter f(x) stehen die vom AUDIPY unterstützten Operatoren und Funktionen als Übersicht zur Verfügung:

Unterstützte Operatoren und Funktionen			
Name	Operator	Beispiel	Beschreibung
Addition	+	@A + @B	Addiert zwei Werte.
Subtraktion	-	@A - @B	Subtrahiert den zweiten Wert vom ersten.
Multiplikation	*	@A * @B	Multipliziert zwei Werte.
Division	/	@A / @B	Teilt den ersten Wert durch den zweiten.
Modulus	%	@A % @B	Gibt den Rest der Division zurück.
Exponentiation	**	@A ** @B	Erhöht den ersten Wert auf die Potenz des zweiten Wertes.
Gleich	==	@A == @B	Überprüft, ob zwei Werte gleich sind.
Ungleich	!=	@A != @B	Überprüft, ob zwei Werte ungleich sind.
Größer als	>	@A > @B	Überprüft, ob der erste Wert größer als der zweite ist.
Kleiner als	<	@A < @B	Überprüft, ob der erste Wert kleiner als der zweite ist.
Größer als oder gleich	>=	@A >= @B	Überprüft, ob der erste Wert größer oder gleich dem zweiten ist.
Kleiner als oder gleich	<=	@A <= @B	Überprüft, ob der erste Wert kleiner oder gleich dem zweiten ist.
Und	and	(@A > 0) and (@B > 0)	Gibt True zurück, wenn beide Operanden True sind.
Oder	or	(@A > 0) or (@B > 0)	Gibt True zurück, wenn mindestens einer der Operanden True ist.
Nicht	not	not (@A > 0)	Gibt True zurück, wenn der Operand False ist.

3. Vorschaufenster 3

Unterhalb der Auswahl sehen Sie eine Vorschau auf das Ergebnis der Berechnung. Im Beispiel wurde der Mittelwert von der Spalte „Sales“ berechnet:

Vorschau Ergebnis: $261.96 * 2 = 523.92$

14.6.4.2 Wenn...Dann...

Mit der Funktion **Wenn...Dann...** können Sie eine neue, bedingte Spalte erstellen, deren Werte sich abhängig von festgelegten Bedingungen automatisch anpassen. So lassen sich beispielsweise bestimmte Textausgaben, Zahlenwerte oder andere Spaltenwerte zuweisen, wenn eine Bedingung erfüllt ist. Im angezeigten Dialog stehen Ihnen folgende Elemente zur Verfügung:

1. Neuer Spalte ①

- Vergeben Sie hier den Namen für Ihre neue, bedingte Spalte.
- Sie können hier auch das Format (s. Bild) der Spalte wählen

Text
Nummer
Datum (+ Zeit)
Zeit
Boolean

2. Bedingung hinzufügen (optional für mehrere Bedingungen) ②

Klicken Sie auf **Bedingung hinzufügen**, wenn Sie mehr als eine Bedingung definieren möchten. Jede Bedingung besteht aus einem **Wenn**-Teil und einem **Dann**-Teil. Sie können beliebig viele Bedingungen hinzufügen, um komplexe Fälle abzudecken.

3. Bedingungstabelle ③

- Legen Sie hier fest, unter welchen Voraussetzungen (z. B. **Sub-Category** = „Phones“) ein bestimmter Wert zugewiesen wird (z. B. **Einzelverkaufspreis**).
- Hierbei können sie in jeder Bedingung unterschiedliche Optionen auswählen. Diese sind wieder am Spaltenformat orientiert und verändern sich je nach Auswahl des entsprechenden Spaltenformats:

Zeichenfolge	Nummern	Datum (+Zeit)	Zeit
• Exakte Übereinstimmung • Teilweise Übereinstimmung • Stimmt nicht überein • Enthält nicht	• Exakte Übereinstimmung • Stimmt nicht überein • Größer als • Kleiner als • Größer oder gleich • Kleiner oder gleich	• Exakte Übereinstimmung • Stimmt nicht überein • Vor • Nach • Am oder vor • Am oder nach	• Exakte Übereinstimmung • Stimmt nicht überein • Vor • Nach • Am oder vor • Am oder nach

Anschließend muss ausgewählt werden, welche Voraussetzung für die ausgewählte Spalte greifen soll. Hier können Sie zwischen Wert, Spalte und Formel auswählen. Je nach Auswahl verändert sich dann auch wieder die Darstellung im Rahmen der Bedingungszeile:

Wert

Spalte

Formel

Wert:

Wert

Geben Sie das Ergebnis...

Geben Sie einen Wert ein, der als Bedingung für ‚Wenn‘ dient (z. B. ein bestimmter Name in einer Namensspalte).

Spalte:

Spalte

Verkauf

Wählen Sie die Spalte aus dem Dropdown-Menü aus, die als Grundlage für die ‚Wenn‘-Bedingung dienen soll (z. B. eine Namensspalte für die Suche nach bestimmten

Einträgen)

Formel:

Formel

Hinzufügen

Wählen Sie ‚Formel‘ aus, wenn Sie Ihre Bedingung auf eine benutzerdefinierte Formel stützen möchten. Klicken Sie anschließend auf ‚Hinzufügen‘, um eine eigene Formel zu erstellen. Es öffnet sich ein neues Fenster zur Eingabe.“

In der Zeile Formel können Sie mithilfe des Zeichen ‚@‘, auf bestimmte Spalten zugreifen (z. B. @SpalteA + @SpalteB * 2).

Im Screenshot wurde die Spalte Profit von der Spalte Verkauf subtrahiert. Das Ergebnis wird in der Vorschauzeile angezeigt.

Mit Klick auf Berechnen wird die Formel in die Bedingung eingefügt, allerdings nicht angezeigt. Klickt man wieder auf Hinzufügen, ist die Formel aber weiterhin enthalten und wird für die Voraussetzung verwendet.

Unter dem Fragezeichen-Symbol können die vorhandenen Berechnungen bzw. Funktionen vom AUDIPY eingesehen werden:

fx

Formel eingeben

Verwenden Sie '@', um auf Spalten zu verweisen: @A + @B * 2

Formeleingabe

Formel:

@Verkauf-@PROFIT

?

Vorschau

Result:

Vorschau Ergebnis: 261.96-41.9136 = 220.04639999999998

Berechnen

Abbrechen

Unterstützte Operatoren und Funktionen			
Name	Operator	Beispiel	Beschreibung
Addition	+	@A + @B	Addiert zwei Werte.
Subtraktion	-	@A - @B	Subtrahiert den zweiten Wert vom ersten.
Multiplikation	*	@A * @B	Multipliziert zwei Werte.
Division	/	@A / @B	Teilt den ersten Wert durch den zweiten.
Modulus	%	@A % @B	Gibt den Rest der Division zurück.
Exponentiation	**	@A ** @B	Erhöht den ersten Wert auf die Potenz des zweiten Wertes.
Gleich	==	@A == @B	Überprüft, ob zwei Werte gleich sind.
Ungleich	!=	@A != @B	Überprüft, ob zwei Werte ungleich sind.
Größer als	>	@A > @B	Überprüft, ob der erste Wert größer als der zweite ist.
Kleiner als	<	@A < @B	Überprüft, ob der erste Wert kleiner als der zweite ist.
Größer als oder gleich	>=	@A >= @B	Überprüft, ob der erste Wert größer oder gleich dem zweiten ist.
Kleiner als oder gleich	<=	@A <= @B	Überprüft, ob der erste Wert kleiner oder gleich dem zweiten ist.
Und	and	(@A > 0) and (@B > 0)	Gibt True zurück, wenn beide Operanden True sind.
Oder	or	(@A > 0) or (@B > 0)	Gibt True zurück, wenn mindestens einer der Operanden True ist.
Nicht	not	not (@A > 0)	Gibt True zurück, wenn der Operand False ist.

Für die Folge der ‚Dann‘-Bedingung kann ebenfalls festgelegt werden, was passieren soll – zur Auswahl stehen ein fester Wert, eine Spalte oder eine Formel. Die Funktionalität entspricht dabei der Vorgehensweise bei der ‚Wenn‘-Bedingung

Hinweis:

Die Reihenfolge der Bedingungen ist relevant: Trifft die erste Bedingung zu, wird der zugehörige Wert angewendet und die restlichen Bedingungen nicht mehr geprüft.

4. Sonst-Fall ④

- Geben Sie an, was in der neuen Spalte stehen soll, wenn keine der definierten Bedingungen erfüllt ist. Dies kann ein konstanter Wert oder ein Verweis auf eine andere Spalte sein.

Wert
Spalte
Formel

5. Alle Bedingungen löschen ⑤

- Möchten Sie noch einmal von vorn beginnen, können Sie mit einem Klick auf den Schnellbutton „Alle Bedingungen löschen“ alle Eintragungen zurücksetzen und neu starten.

Auf diese Weise können Sie Ihre Daten automatisiert klassifizieren, gruppieren oder bestimmte Werte zuweisen, ohne manuelle Nachbearbeitung vornehmen zu müssen.

Beispiel:

In diesem Beispiel wird die Bedingung definiert, dass **wenn** in der Spalte **Ship Mode** der Wert **Standard Class** steht, **dann** soll in der neuen Spalte der Wert **Economy Class** erscheinen. Ansonsten soll es bei den anderen Klassen bleiben.

Bedingte Spalte hinzufügen
Definieren Sie Bedingungen für die neue Spalte:

Bedingungen Hinzufügen

☒ Wenn SHIP MODE Exakte Übereinstimmung Cc Wert Standard Class Dann Wert Economy Class

Sonst-Fall
Wird ausgeführt, wenn keine Bedingung erfüllt ist.
Sonst: Spalte SHIP MODE

Alle Bedingungen löschen Spalte hinzufügen Abbrechen

Neue Spalte
Name: Neue Klasse
Format: Text

Neue Klasse

Economy Class
Economy Class
Economy Class
Economy Class
Economy Class
Second Class
Second Class

Ergebnis: Es wird eine neue Spalte mit dem Namen „Neue Klasse“ hinzugefügt.

Kumulative Summe

Die Funktion **Kumulative Summe** ermöglicht es Ihnen, einen fortlaufenden Gesamtwert einer ausgewählten Spalte zu berechnen. So können Sie zum Beispiel verfolgen, wie sich der Profit im Laufe einer Liste oder innerhalb bestimmter Gruppen aufsummiert. Der gezeigte Dialog gliedert sich in folgende Bereiche:

1. Spalte für kumulative Summe (Pflichtfeld) ①

- Wählen Sie hier die Spalte aus, deren Werte aufaddiert werden sollen (z. B. „Profit“).
- Auf Basis dieser Spalte wird eine neue Spalte erstellt, in der pro Zeile der jeweilige Summenstand bis zu dieser Zeile angezeigt wird.

2. Sortierung (optional) ②

- Standardmäßig werden die Zeilen in der Reihenfolge der vorhandenen Tabelle verarbeitet. Möchten Sie die kumulative Summe in einer bestimmten Reihenfolge berechnen, können Sie hier eine oder mehrere Spalten angeben, nach denen sortiert wird.
- Klicken Sie auf **Spalte hinzufügen**, um weitere Spalten für die Sortierung festzulegen. Sie können zudem auswählen, ob die Sortierung **beibehalten**, **aufsteigend** oder **absteigend** erfolgen soll.
- Die Reihenfolge der Einträge ist wichtig für die Berechnung, da die kumulative Summe jeweils den vorherigen Wert weiterführt.

3. Gruppierung (optional) ③

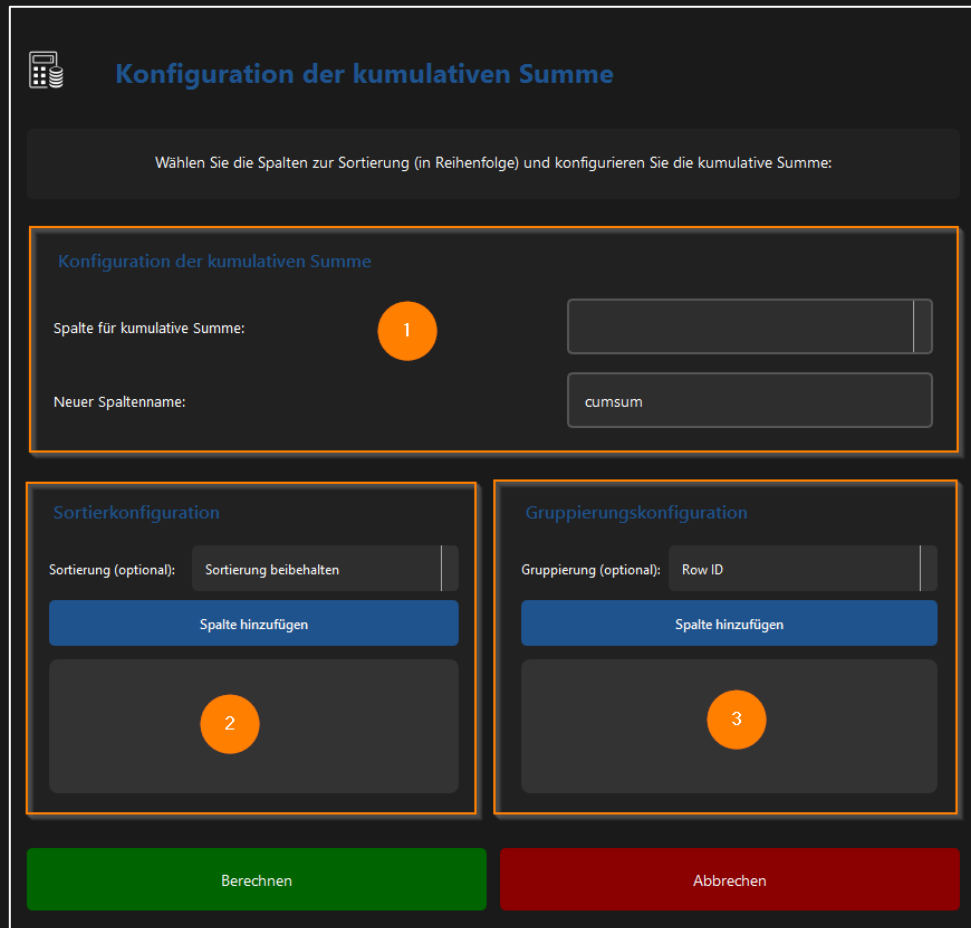
- Geben Sie hier eine oder mehrere Spalten an, nach denen die kumulative Summe getrennt berechnet werden soll.
- Beispiel: Wenn Sie nach „Product Name“ gruppieren, wird für jedes Produkt eine eigene kumulative Summe geführt. Sobald das Produkt wechselt, startet die Summenbildung wieder von null.
- Auf diese Weise erhalten Sie eine fortlaufende Addition pro Gruppe, was besonders hilfreich sein kann, um Teilbereiche (z. B. nach Kategorien oder Kunden) gesondert zu betrachten.

Mit einem Klick auf **Berechnen** wird eine neue Spalte **am Ende der Tabelle** angelegt (es erscheint erst noch ein Fenster, welches die neue Spalte bestätigt), welche die Ergebnisse beinhaltet. Möchten Sie den Vorgang abbrechen, wählen Sie **Abbrechen**

Mit der kumulativen Summe behalten Sie den Überblick über fortlaufende Werte, sei es der aufaddierte Umsatz, Profit oder eine andere Kennzahl. Durch die optionalen Sortierungs- und Gruppierungsmöglichkeiten können Sie die Berechnung flexibel an Ihre Analysebedürfnisse anpassen.

14.6.4.3 Kumulative Summe

Mit dieser Funktion kann eine **kumulative Summe (laufende Summe)** über eine ausgewählte Spalte berechnet werden. Die Berechnung erfolgt optional sortiert und/oder gruppiert, sodass flexible Auswertungen innerhalb von Teilmengen oder über Zeitverläufe hinweg möglich sind.



1. Konfiguration der kumulativen Summe ①

Hier wird festgelegt:

- **Spalte für die kumulative Summe:**
Die numerische Spalte, deren Werte aufsummiert werden sollen.
- **Neuer Spaltenname:**
Name der Spalte, in der das Ergebnis der kumulativen Berechnung gespeichert wird (Standard: cumsum).

2. Sortierkonfiguration (optional) ②

Die Reihenfolge, in der die Werte aufsummiert werden, kann hier definiert werden:

- Bei keiner Angabe wird die ursprüngliche Sortierung beibehalten.
- Über „**Spalte hinzufügen**“ kann eine oder mehrere Spalten zur Sortierung ausgewählt werden (z. B. nach Datum, Kundennummer etc.).
- Die Reihenfolge der Sortierung beeinflusst direkt das Ergebnis der kumulativen Summe.

3. Gruppierungskonfiguration (optional) ③

Zur Berechnung **getrennter kumulativer Summen innerhalb von Gruppen**:

- Gruppierungen können z. B. nach Kunde, Region oder Kategorie erfolgen.
- Jede Gruppe erhält ihre eigene kumulative Berechnung, unabhängig von anderen Gruppen.
- Über „**Spalte hinzufügen**“ kann die Gruppierung beliebig erweitert werden.

Beispiel:

Eingangsdaten:			
Kunde	Monat	Umsatz	
A	Jan	100	
A	Feb	200	
A	Mär	150	
Ausgabe mit kumulativer Summe (<code>cumsum</code>):			
Kunde	Monat	Umsatz	cumsum
A	Jan	100	100
A	Feb	200	300
A	Mär	150	450

14.6.5 Sortieren

Mit der Funktion **Spalten sortieren** können Sie die Reihenfolge Ihrer Daten nach einer oder mehreren Spalten bestimmen. Diese Sortierung wird grundsätzlich (je nach Einstellung s. 2.2.9.3) im Verlauf (s. Kapitel 3.8.1.3) berücksichtigt. Sortiert man hingegen in der Tabelle (mit Rechtsklick auf den Spaltennamen) wird diese Filterung nicht im Verlauf angezeigt.

Im gezeigten Dialog stehen Ihnen folgende Optionen zur Verfügung:

1. Verfügbare Spalten ①

- Wählen Sie aus der Liste die Spalte(n) aus, die Sie sortieren möchten.
- Klicken Sie anschließend auf **Spalte hinzufügen**, um die ausgewählte Spalte in die Sortierliste zu übernehmen.

2. Sortierliste ②

- Hier sehen Sie alle Spalten, die aktuell in die Sortierung einbezogen werden. Jede Spalte kann einzeln aufsteigend oder absteigend sortiert werden.
- Sie können mehrere Spalten hinzufügen, um eine mehrstufige Sortierung vorzunehmen (z. B. zuerst nach „Category“ aufsteigend, dann nach „Sub-Category“ absteigend).

Sortieren bzw. Abbrechen

- **Sortieren:** Übernimmt die ausgewählten Sortierkriterien und aktualisiert die Reihenfolge Ihrer Daten.
- **Abbrechen:** Schließt den Dialog, ohne Änderungen vorzunehmen.

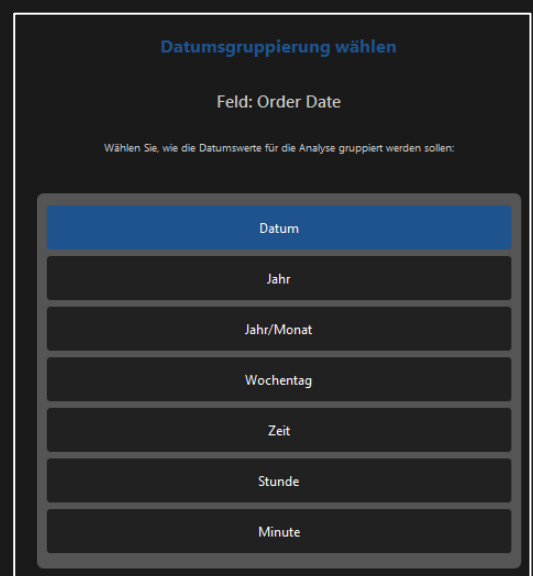
14.6.6 Pivot

Mit der **Pivot-Tabelle** können Sie Ihre Daten flexibel zusammenfassen und aus verschiedenen Blickwinkeln betrachten. Dabei ziehen Sie die gewünschten Felder in **Zeilenfelder**, **Spaltenfelder** und **Datenfelder**, um unterschiedliche Kombinationen und Aggregationen Ihrer Datensätze zu erhalten.



1. Verfügbare Felder ①

- Auf der rechten Seite finden Sie eine Liste aller Spalten, die in Ihrer Tabelle vorhanden sind. Numerische Felder erhalten hierbei ein .
- Ziehen Sie die gewünschten Spalten per Drag-and-Drop in den entsprechenden Bereich (**Zeilen**, **Spalten** oder **Daten**), um das Pivot-Layout zu bestimmen.
- **Hinweis:** In diesem Fenster legen Sie fest, wie die **Datumswerte** für die Analyse gruppiert werden sollen. Sie können auswählen, ob die Daten beispielsweise nach Datum, Jahr oder nach Stunden zusammengefasst werden. Klicken Sie auf die gewünschte Gruppierung, um sie anzuwenden.



2. Zeilenfelder ②

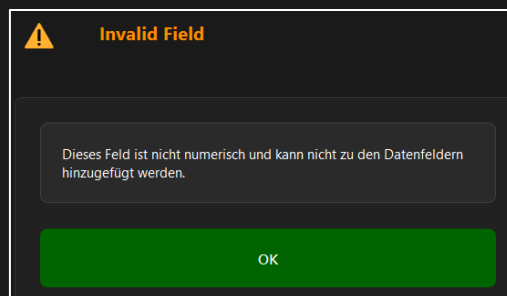
- Spalten, die Sie hier ablegen, bestimmen die **Zeilenkategorien** der Pivot-Tabelle. Beispielsweise könnten Sie hier „Region“ ablegen, sodass jede Zeile eine Region repräsentiert.

3. Spaltenfelder ③

- Spalten, die in diesem Bereich liegen, bestimmen die **Spaltenkategorien** der Pivot-Tabelle. Zum Beispiel können Sie hier „Produktkategorie“ platzieren, sodass jede Spalte eine Kategorie abbildet.

4. Datenfelder ④

- Hier legen Sie die Werte ab, die Sie zusammenfassen möchten (z.B. Umsatz, Anzahl). Standardmäßig wird eine Summe gebildet, aber je nach Bedarf können Sie andere Aggregationen (Durchschnitt, Min, Max, Anzahl usw.) auswählen. In dieses Feld können nur numerische Werte einbezogen werden, ansonsten erhalten Sie folgenden Warnhinweis:



Abschließend brauchen Sie nur noch mit „OK“ die Eingaben bestätigen und erhalten eine Pivot-Tabelle. Bei größeren Datenmengen kann es zu kurzen Wartezeiten kommen.

Die **Detailansicht** wird geöffnet, wenn Sie in einer Pivot Tabelle per Doppelklick auf eine bestimmte Zeile klicken. Dort sehen Sie die einzelnen Datensätze, die in der Pivot Tabelle zusammengefasst wurden.

In diesem Detailfenster haben Sie mehrere Möglichkeiten, die Daten weiterzuverarbeiten ⑤-⑧:

City	(Sales, 'Central')	(Sales, 'East')	(Sales, 'South')	(Sales, 'West')
Aberdeen	25,50	0,00	0,00	0,00
Abilene	1,39	0,00	0,00	0,00
Akron	0,00	2.729,99	0,00	0,00
Albuquerque	0,00	0,00	0,00	2.220,17
Alexandria	0,00	0,00	5.519,57	0,00
Allen	290,21	0,00	0,00	0,00
Allentown	0,00	853,26	0,00	0,00
Altoona	0,00	20,45	0,00	0,00
Amarillo	3.773,06	0,00	0,00	0,00
Anaheim	0,00	0,00	0,00	7.986,88
Andover	0,00	435,85	0,00	0,00
Ann Arbor	889,27	0,00	0,00	0,00
Antioch	0,00	0,00	0,00	19,44
Apopka	0,00	0,00	904,55	0,00
Apple Valley	137,94	0,00	0,00	1.915,08
Appleton	1.671,31	0,00	0,00	0,00
Arlington	5.847,75	0,00	14.366,76	0,00
Arlington Heights	14,11	0,00	0,00	0,00
Arvada	0,00	0,00	0,00	503,40
Asheville	0,00	0,00	1.475,37	0,00
Athens	0,00	0,00	1.720,81	0,00
Atlanta	0,00	0,00	17.197,84	0,00
Atlantic City	0,00	23,36	0,00	0,00
Auburn	0,00	1.384,16	1.766,83	4,18
Aurora	7.572,95	0,00	0,00	4.083,50

5 Schließen
 6 Doppelklick auf eine Zeile, um die Detailsansicht zu öffnen.
 7 Speichern
 Lesezeichen
 In Excel öffnen
 8

- Button Schließen **5**: Mit Betätigung des roten „Schließen“ Buttons gehen die angezeigten Daten verloren, da diese nur temporär gespeichert sind. Für eine Speicherung der Daten sollte ein Lesezeichen zur dauerhaften Verwendung angelegt werden (2.7 Lesezeichen).
- Mit einem **Doppelklick auf eine Zeile** **6** öffnet sich die **Detailsansicht** der zugehörigen Werte. In dieser Ansicht werden alle zu dem jeweiligen Eintrag gespeicherten Informationen übersichtlich dargestellt.

i

Details zur Kategorie:

CITY = Ann Arbor, REGION = Central

Datenansicht

5 Zeilen

	ROW ID	ORDER ID	ORDER DATE	SHIP DATE	SHIP MODE	CUSTOMER ID	CUSTOMER NAME	SEGMENT	COUNTRY	CITY	ST
1	5635	CA-2015-134943	00:00:00	06.12.2015	First Class	SU-20665	Stephanie Ulpright	Home Office	United States	Ann Arbor	Michiga
2	5716	CA-2015-124107	00:00:00	12.10.2015	Second Class	BM-11650	Brian Moss	Corporate	United States	Ann Arbor	Michiga
3	5717	CA-2015-124107	00:00:00	12.10.2015	Second Class	BM-11650	Brian Moss	Corporate	United States	Ann Arbor	Michiga
4	5718	CA-2015-124107	00:00:00	12.10.2015	Second Class	BM-11650	Brian Moss	Corporate	United States	Ann Arbor	Michiga
5	5719	CA-2015-124107	00:00:00	12.10.2015	Second Class	BM-11650	Brian Moss	Corporate	United States	Ann Arbor	Michiga

In Excel öffnen

Filter anwenden

Lesezeichen hinzufügen

Schließen

Die Detailsansicht bietet folgende Möglichkeiten:

- **Export nach Excel:** Die angezeigten Werte können direkt in eine Excel-Datei exportiert werden.
- **Filtern in der Haupttabelle:** Über einen einfachen Klick kann ein Filter auf die Haupttabelle angewendet werden, um nur jene Einträge anzuzeigen, die der Detailansicht entsprechen (Achtung: Auch hier geht die vorherige Pivot-Tabelle verloren, da sie nur temporär gespeichert ist). Es empfiehlt sich ein Lesezeichen zu speichern (8 Lesezeichen).
- **Als Lesezeichen speichern:** Relevante Einträge lassen sich als Lesezeichen ablegen, um sie später schnell wiederzufinden oder in einem anderen Kontext erneut aufzurufen.
- **Speichern, Lesezeichen, in Excel öffnen 7:** Hier lässt sich die gesamte Pivot Ansicht als separate Datei im Explorer speichern. Darüber hinaus stehen auch hier die Erstellung eines Lesezeichen (8 Lesezeichen) sowie die Excelfunktion zur Verfügung.
- **Summe, Dunkelmodus, Zoom 8:** Über die **Summenfunktion** Σ können die Spalten schnell zusammengefasst werden. Voreingestellt ist die Funktion als Summe. Klickt man auf das Symbol öffnet sich eine Zeile mit Summen. Am Rand gibt es die Möglichkeit in ein Dropdown-Menü zu gelangen: Hier lassen sich neben Summe auch folgende Funktionen anzeigen:

Summe

Summe
Minimum
Maximum
Mittelwert
Median
Anzahl
Anzahl (eindeutig)
Standardabweichung
Varianz

So kann die Darstellung und die Berechnung der numerischen Felder beliebig veranschaulicht werden.

Der integrierte **Dunkelmodus**  sorgt für eine augenschonende Darstellung, insbesondere bei längeren Analysephasen oder Arbeiten in dunklen Umgebungen. Zusätzlich ermöglicht die **Zoomfunktion** , die Tabellenansicht flexibel zu vergrößern oder zu verkleinern, um Inhalte besser lesbar zu machen

14.6.7 Gruppieren

Mit der Funktion **Gruppieren** können Sie Ihre Daten nach bestimmten Spalten zusammenfassen und gleichzeitig Kennzahlen (wie **Summe**, **Durchschnitt**, **Anzahl**, etc.) für die ausgewählten Spalten berechnen. Dadurch erhalten Sie eine kompakte Übersicht, in der gleiche Werte (z. B. gleiche Kategorie oder Unterkategorie) nur einmal dargestellt werden, während die zugehörigen Zahlenwerte entsprechend aggregiert sind.

Daten gruppieren und aggregieren

Gruppierungsspalten

SUB-CATEGORY

Spalte hinzufügen

× CATEGORY

× SUB-CATEGORY

1

Aggregatspalten

Verkauf

Spalte hinzufügen

× Verkauf Summe

2

Gruppieren

Abbrechen

1. Gruppieren nach Spalten ①

- Wählen Sie hier die Spalten aus, nach denen gruppiert werden soll (z. B. „Category“, „Sub-Category“). Die Gruppierung erfolgt **hierarchisch** – das bedeutet, dass die erste gewählte Spalte die **oberste Gruppierungsebene** bildet, alle weiteren Spalten ordnen sich dieser unter.
- Je nach Reihenfolge entstehen dadurch unterschiedliche Gruppierungsstrukturen: Wird beispielsweise zuerst nach *Category* und dann nach *Sub-Category* gruppiert, erhält man eine andere Darstellung als bei umgekehrter Reihenfolge (*Sub Category* → *Category*). Die Reihenfolge steuert also, **wie die Daten verschachtelt und aggregiert werden** – was insbesondere bei Auswertungen mit Summen, Zählungen oder Mittelwerten entscheidend ist.
- Fügen Sie bei Bedarf weitere Spalten über **Spalte hinzufügen** hinzu, um eine mehrstufige Gruppierung zu erreichen.

2. Zusammenfassende Spalten 2

- Geben Sie an, welche Spalten aggregiert werden sollen (z. B. „Sales“).
- Wählen Sie anschließend die gewünschte Aggregationsmethode (z. B. **Summe**, **Durchschnitt**, **Min**, **Max**, **Anzahl**).
- Über **Spalte hinzufügen** können Sie weitere Spalten oder Aggregationen hinzufügen, um mehrere Kennzahlen gleichzeitig zu berechnen.

Summe
Minimum
Maximum
Mittelwert
Median
Anzahl
Anzahl (eindeutig)
Standardabweichung
Varianz

Abschließend klicken Sie auf **Gruppieren**, um die gewählte Aktion anzuwenden, oder auf **Abbrechen**, falls Sie keine Änderungen vornehmen möchten.

Danach öffnen sich die gruppierten Werte:

Category	Sub-Category	Sales (Summe)
Furniture	Bookcases	114.879,97
Furniture	Chairs	327.848,53
Furniture	Furnishings	91.705,12
Furniture	Tables	206.965,52
Office Supplies	Appliances	107.486,92
Office Supplies	Art	27.110,54
Office Supplies	Binders	203.312,29
Office Supplies	Envelopes	16.476,38
Office Supplies	Fasteners	3.024,25
Office Supplies	Labels	12.486,30
Office Supplies	Paper	78.288,23
Office Supplies	Storage	223.334,10
Office Supplies	Supplies	46.634,45
Technology	Accessories	166.484,76
Technology	Copiers	149.528,01
Technology	Machines	189.238,65
Technology	Phones	329.970,91

Sales (Summe): 2.294.774,93

Buttons: 3 Schließen, 4 Doppelklick auf eine Zeile, um die Detailansicht zu öffnen, 5 Speichern, Lesezeichen, In Excel öffnen, 6 Summe

- Button Schließen 3: Mit Betätigung des roten „Schließen“ Buttons gehen die angezeigten Daten verloren, da diese nur temporär gespeichert sind. Für eine Speicherung der Daten sollte ein Lesezeichen zur dauerhaften Verwendung angelegt werden (2.7 Lesezeichen).
- Mit einem **Doppelklick auf eine Zeile** 4 öffnet sich die **Detailansicht** der zugehörigen Werte. In dieser Ansicht werden alle zu dem jeweiligen Eintrag gespeicherten Informationen übersichtlich dargestellt.

Details zur Kategorie:
CATEGORY = Furniture, SUB-CATEGORY = Tables

Datenansicht 319 Zeilen

	ROW ID	ORDER ID	ORDER DATE	SHIP DATE	SHIP MODE	CUSTOMER ID	CUSTOMER NAME	SEGMENT	COUNTRY	CITY
1	4	US-2015-108966	31.12.1898		Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale
2	11	CA-2014-115812	31.12.1898	14.06.2014	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
3	25	CA-2015-106320	31.12.1898	30.09.2015	Standard Class	EB-13870	Emily Burns	Consumer	United States	Orem
4	118	CA-2015-110457	31.12.1898	06.03.2015	Standard Class	DK-13090	Dave Kipp	Consumer	United States	Seattle
5	126	US-2014-134614	31.12.1898	25.09.2014	Standard Class	PF-19165	Philip Fox	Consumer	United States	Bloomington
6	202	CA-2014-133690	31.12.1898	05.08.2014	First Class	BS-11755	Bruce Stewart	Consumer	United States	Denver
7	227	CA-2015-163055	31.12.1898	16.08.2015	Standard Class	DS-13180	David Smith	Corporate	United States	Detroit
8	232	US-2017-100930	31.12.1898	12.04.2017	Standard Class	CS-12400	Christopher Schild	Home Office	United States	Tampa
9	233	US-2017-100930	31.12.1898	12.04.2017	Standard Class	CS-12400	Christopher Schild	Home Office	United States	Tampa
10	242	CA-2016-157749	31.12.1898	09.06.2016	Second Class	KL-16645	Ken Lonsdale	Consumer	United States	Chicago
11	283	CA-2015-130890	31.12.1898	06.11.2015	Standard Class	JO-15280	Jas O'Carroll	Consumer	United States	Los Angeles
12	304	US-2017-152380	31.12.1898	23.11.2017	Standard Class	JH-15910	Jonathan Howell	Consumer	United States	Chicago

[In Excel öffnen](#)
[Filter anwenden](#)
[Lesezeichen hinzufügen](#)
[Schließen](#)

Die Detailansicht bietet folgende Möglichkeiten:

- **Export nach Excel:** Die angezeigten Werte können direkt in eine Excel-Datei exportiert werden.
- **Filtern in der Haupttabelle:** Über einen einfachen Klick kann ein Filter auf die Haupttabelle angewendet werden, um nur jene Einträge anzuzeigen, die der Detailansicht entsprechen (Achtung: Auch hier geht die vorherige Gruppierungs-Tabelle verloren, da sie nur temporär gespeichert ist). Es empfiehlt sich ein Lesezeichen zu speichern (2.7 Lesezeichen).
- **Als Lesezeichen speichern:** Relevante Einträge lassen sich als Lesezeichen ablegen, um sie später schnell wiederzufinden oder in einem anderen Kontext erneut aufzurufen.
- **Speichern, Lesezeichen, in Excel öffnen 5:** Hier lässt sich die gesamte Gruppierungsansicht als separate Datei im Explorer speichern. Darüber hinaus stehen auch hier die Erstellung eines Lesezeichen (2.7 Lesezeichen) sowie die Excelfunktion zur Verfügung.
- **Summe, Dunkelmodus, Zoom 6:** Über die **Summenfunktion** Σ können die Spalten schnell zusammengefasst werden. Voreingestellt ist die Funktion als Summe. Klickt man auf das Symbol öffnet sich eine Zeile mit Summen. Am Rand gibt es die Möglichkeit in ein Dropdown-Menü zu gelangen: Hier lassen sich neben Summe auch folgende Funktionen anzeigen:

Summe
Minimum
Maximum
Mittelwert
Median
Anzahl
Anzahl (eindeutig)
Standardabweichung
Varianz

So kann die Darstellung und die Berechnung der numerischen Felder beliebig veranschaulicht werden.

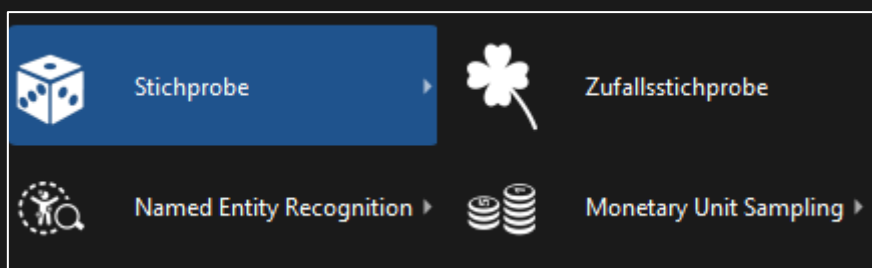
Der integrierte **Dunkelmodus**  sorgt für eine augenschonende Darstellung, insbesondere bei längeren Analysephasen oder Arbeiten in dunklen Umgebungen. Zusätzlich ermöglicht die **Zoomfunktion** , die Tabellenansicht flexibel zu vergrößern oder zu verkleinern, um Inhalte besser lesbar zu machen

	 CATEGORY	 SUB-CATEGORY	 Verkauf (Summe)
1	Furniture	Bookcases	114.879,98
2	Furniture	Chairs	328.449,1
3	Furniture	Furnishings	91.705,12
4	Furniture	Tables	206.965,53
5	Office Supplies	Appliances	107.532,14
6	Office Supplies	Art	27.118,8
7	Office Supplies	Binders	203.412,68
8	Office Supplies	Envelopes	16.476,38
9	Office Supplies	Fasteners	3.024,25
10	Office Supplies	Labels	12.486,3
11	Office Supplies	Paper	78.479,24
12	Office Supplies	Storage	223.843,59
13	Office Supplies	Supplies	46.673,52
14	Technology	Accessories	167.380,31
15	Technology	Copiers	149.528,01
16	Technology	Machines	189.238,6
17	Technology	Phones	330.007,1

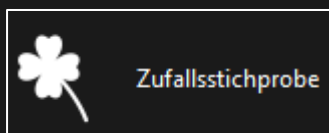
Beispiel: Dunkelmodus Datenbestand

14.7 Stichprobe

In der Prüfungspraxis ist es häufig nicht möglich oder wirtschaftlich sinnvoll, die Gesamtheit einer Grundgesamtheit vollständig zu untersuchen. Stattdessen werden Stichprobenverfahren eingesetzt, um aus einer Teilmenge der Daten verlässliche Aussagen über die Gesamtpopulation treffen zu können. Zwei gängige Verfahren sind das **Monetary Unit Sampling (MUS)** und die **Zufallsstichprobe**. Diese beiden Verfahren können mit dem AUDIPY durchgeführt werden.



14.7.1 Zufallsstichprobe



Die **Zufallsstichproben-Ziehung** stellt eine zentrale Methode zur statistisch fundierten Datenauswahl dar, die insbesondere für Wirtschaftsprüfer, Steuerberater und Controller bei der Durchführung von Stichprobenprüfungen unverzichtbar ist. Diese Funktionalität ermöglicht es, aus großen Datensätzen repräsentative Teilmengen zu ziehen, die für Prüfungshandlungen, Analysen oder Qualitätskontrollen verwendet werden können.

Das System implementiert eine mathematisch korrekte Zufallsauswahl, die den Anforderungen professioneller Prüfungsstandards entspricht. Durch die Verwendung eines definierbaren Seed-Wertes wird sichergestellt, dass Stichprobenziehungen reproduzierbar und damit prüfungskonform dokumentierbar sind. Dies ist besonders wichtig für Compliance-Anforderungen und die Nachvollziehbarkeit von Prüfungshandlungen.

Durch die einfache Anwendung und nachvollziehbare Methodik stellt die Zufallsstichprobe ein bewährtes Instrument zur Unterstützung systematischer Prüfungen dar.

14.7.1.1 *Praktische Anwendungsszenarien*

Der Zufallstichprobenprüfung kommt in der Praxis an vielen verschiedenen Stellen große Bedeutung zu.

Wirtschaftsprüfung

Substantive Prüfungshandlungen: Wirtschaftsprüfer nutzen die Stichprobenziehung für die Auswahl von Buchungen, Belegen oder Geschäftsvorfällen für detaillierte Einzelfallprüfungen. Die mathematisch korrekte Zufallsauswahl gewährleistet, dass die Stichprobe repräsentativ für die Grundgesamtheit ist und valide Rückschlüsse auf die Gesamtpopulation erlaubt. Besonders bei der Prüfung von Forderungen, Verbindlichkeiten oder Umsatzerlösen ermöglicht das Random Sampling eine effiziente und standardkonforme Prüfungsplanung.

Compliance Testing: Bei der Überprüfung von internen Kontrollen können aus großen Datensätzen systematisch Stichproben gezogen werden, um die Wirksamkeit von Geschäftsprozessen zu beurteilen. Die automatische Metadaten-Dokumentation gewährleistet die vollständige Nachvollziehbarkeit der Auswahl und unterstützt die Objektivität bei Qualitätskontrollen oder Peer Reviews.

Analytische Prüfungshandlungen: Für Trendanalysen oder Plausibilitätsprüfungen können aus sehr großen Datensätzen repräsentative Stichproben gezogen werden, die eine effiziente Analyse bei gleichzeitiger statistischer Aussagekraft ermöglichen. Die Bereichseinschränkung unterstützt dabei die Fokussierung auf spezifische Zeiträume oder Geschäftsbereiche.

Steuerberatung

Betriebsprüfungs-Vorbereitung: Steuerberater können im Vorfeld von Betriebsprüfungen repräsentative Stichproben der mandantenseitigen Buchführung ziehen und diese auf steuerliche Risiken oder Besonderheiten analysieren. Die in den Metadaten dokumentierten Parameter ermöglichen es, die Stichprobenauswahl gegenüber Prüfungsbehörden transparent darzustellen und die methodische Korrektheit zu belegen.

Compliance-Überwachung: Bei der laufenden steuerlichen Beratung können aus umfangreichen Buchungsdaten Stichproben für die Überprüfung der ordnungsgemäßen Buchführung oder der Einhaltung steuerlicher Vorschriften gezogen werden. Dies ist besonders relevant bei Mandanten mit hohem Buchungsvolumen oder komplexen Geschäftsstrukturen.

Datenqualitätsprüfung: Vor der Erstellung von Steuererklärungen oder anderen steuerlichen Auswertungen können Stichproben zur Überprüfung der Datenqualität und -konsistenz herangezogen werden. Systematische Fehler oder Unstimmigkeiten in der Buchführung können so frühzeitig erkannt und korrigiert werden.

Controlling

Kostenstellenanalysen: Controller können aus umfangreichen Kostendaten repräsentative Stichproben für detaillierte Kostenanalysen ziehen. Dies ist besonders wertvoll bei der Überprüfung von Kostenzuordnungen, der Analyse von Kostenstellenabrechnungen oder der Vorbereitung von Budgetgesprächen.

Kennzahlen-Validierung: Bei der Erstellung von Management-Berichten können Stichproben zur Validierung der zugrunde liegenden Daten herangezogen werden. Dies gewährleistet die Qualität der berichteten Kennzahlen und stärkt das Vertrauen in die Controlling-Informationen.

Projektcontrolling: Für die Analyse von Projektkosten oder -erlösen können aus großen Transaktionsdatenbeständen systematische Stichproben gezogen werden. Die Bereichseinschränkung ermöglicht dabei die Fokussierung auf spezifische Projektphasen oder Leistungszeiträume.

14.7.1.2 Erstellung einer Zufallsstichprobe

Mit Klick auf die Zufallsstichprobe öffnet sich ein weiteres Fenster:

Zufällige Datensatzstichprobe

Stichprobeneinstellungen

Auszuwählende Datensätze: 10

Zufallswert: 9386

Bereich: 1 → 9994

☐ Duplikate zulassen

Ausgabe

>IPY\Deutsch\Fallverwaltung\Fälle\Semitax GmbH\Analyzer\sample.parquet

Durchsuchen... Felder auswählen...

OK Abbrechen

Der **Zufallsstichproben-Dialog** gliedert sich in übersichtliche Funktionsbereiche, die eine logische Arbeitsreihenfolge von der grundlegenden Konfiguration bis zur Ausgabedefinition unterstützen. Die Benutzeroberfläche ist so gestaltet, dass auch bei komplexen Stichprobenziehungen alle relevanten Parameter auf einen Blick erfassbar sind.

Stichprobeneinstellungen 1:

- **Auszuwählende Datensätze:**

Numerisches Eingabefeld zur Bestimmung der Stichprobengröße. Das System berücksichtigt automatisch die Obergrenze des verfügbaren Datenbestands und verhindert damit ungültige Eingaben. Die Stichprobengröße sollte statistisch fundiert gewählt werden, wobei Faktoren wie Prüfungsrisiko, Materialität und gewünschte Aussagesicherheit zu berücksichtigen sind.

- **Zufallswert:**

Numerischer Wert zur Initialisierung des Zufallsgenerators. Dieser Parameter ist entscheidend für die Reproduzierbarkeit der Stichprobenziehung. Durch Verwendung desselben Seed-Wertes können identische Stichproben zu späteren Zeitpunkten exakt reproduziert werden. Dies ist besonders wichtig für:

- Dokumentation von Prüfungshandlungen
- Nachvollziehbarkeit bei Qualitätskontrollen
- Konsistenz bei Folgeprüfungen
- Compliance mit Prüfungsstandards

- **Bereich:**

Definition des Datenbereichs, aus dem die Stichprobe gezogen werden soll. Diese Funktion ermöglicht eine zeitliche oder sachliche Einschränkung der Grundgesamtheit:

- **Startindex:** Erste Zeile des zu berücksichtigenden Bereichs
- **Endindex:** Letzte Zeile des zu berücksichtigenden Bereichs
- **Pfeil-Symbol:** Visualisiert die Bereichsdefinition von Start bis Ende

Die Bereichseinschränkung ist besonders wertvoll bei chronologisch sortierten Daten, wo nur bestimmte Zeiträume für die Analyse relevant sind.

- **Duplikate zulassen (Checkbox):**

Kontrollkästchen zur Aktivierung der Ziehung mit Zurücklegen. Bei aktivierter Option können einzelne Datensätze mehrfach in der Stichprobe enthalten sein. Diese Einstellung ist relevant für:

- Mathematisch korrekte Zufallsauswahl bei kleinen Grundgesamtheiten
- Spezielle statistische Analyseverfahren
- Simulationsrechnungen

Bei den meisten praktischen Anwendungen bleibt diese Option deaktiviert, da Prüfungshandlungen typischerweise eine Ziehung ohne Zurücklegen erfordern.

- **Ausgabe ②:**

- **Pfad:**
Eingabefeld zur Spezifikation des Speicherorts der Stichprobendatei. Das System schlägt automatisch einen Standard-Dateipfad vor, der sich am Verzeichnis der ursprünglichen Datei orientiert. Dies erleichtert die Zuordnung und Archivierung zusammengehöriger Dateien.

- **Durchsuchen**

Über die Schaltfläche "Durchsuchen..." kann der Zielordner und Dateiname über einen Standard-Dateidialog ausgewählt werden. Das System fügt automatisch die korrekte Dateierweiterung (.parquet) hinzu, falls diese nicht explizit angegeben wurde.

- **Feld auswählen:**

Über die Schaltfläche "**Felder auswählen...**" wird ein separater Dialog geöffnet, der die präzise Definition der in der Stichprobe zu berücksichtigenden Spalten ermöglicht. Diese Funktionalität ist entscheidend für die Optimierung der Stichprobendateien und die Fokussierung auf prüfungsrelevante Informationen.



- **Verfügbare Felder:** Vollständige Liste aller Spalten des ursprünglichen Datensatzes mit individuellen Auswahlmöglichkeiten. Jede Spalte kann über ein Kontrollkästchen aktiviert oder deaktiviert werden. Die Auswahl beeinflusst direkt die Größe und Struktur der resultierenden Stichprobendatei.

- **Schnellauswahl-Funktionen:**

- **Alle:** Aktiviert sämtliche verfügbaren Spalten
- **Keine:** Deaktiviert alle Spalten für eine komplett manuelle Auswahl

Strategische Spaltenauswahl

Die Auswahl der zu berücksichtigenden Spalten sollte zweckgebunden erfolgen und verschiedene Kategorien berücksichtigen:

Prüfungsrelevante Kerndaten:

- Beträge und Mengen

- Buchungsdaten und Perioden
- Kontokennungen und Kostenstellen
- Geschäftspartner-Informationen

Identifikations- und Referenzfelder:

- Eindeutige Schlüssel (IDs, Belegnummern)
- Systemreferenzen
- Verknüpfungen zu anderen Datenquellen

Kontextinformationen:

- Buchungstexte und Beschreibungen
- Benutzer- und Zeitstempel
- Zusätzliche Klassifizierungen

Die durchdachte Spaltenauswahl reduziert die Dateigröße und verbessert die Übersichtlichkeit bei der späteren Analyse, ohne prüfungsrelevante Informationen zu verlieren.

Nach **Bestätigung mit OK**, wird eine neue Datei erzeugt, welche wie folgt aussehen kann:

ROW ID	ORDER ID	ORDER DATE	SHIP DATE	SHIP MODE	CUSTOMER ID	CUSTOMER NAME	SEGMENT	COUNTRY	CITY	STATE	POSTAL CODE	REGION
343	CA-2014-122336	13.04.2014	17.04.2014	Second Class	JD-15895	Jonathan Doherty	Corporate	United States	Philadelphia	Pennsylvania	19140	East
2819	CA-2015-104626	01.09.2015	08.09.2015	Standard Class	DR-12940	Daniel Raglin	Home Office	United States	Franklin	Massachusetts	2038	East
3717	CA-2017-144568	29.05.2017	02.06.2017	Standard Class	JO-15550	Jesus Ocampo	Home Office	United States	Omaha	Nebraska	68104	Central
6882	CA-2015-137113	01.12.2015	05.12.2015	Second Class	TW-21025	Tamara Willingham	Home Office	United States	Seattle	Washington	98105	West
922	CA-2015-111164	11.04.2015	15.04.2015	Standard Class	SE-20110	Sanjit Engle	Consumer	United States	New York City	New York	10009	East
4535	CA-2015-163104	19.03.2015	26.03.2015	Standard Class	MC-18130	Mike Caudle	Corporate	United States	Los Angeles	California	90004	West

Das System hat die Stichprobenziehung in mehreren Schritten durchgeführt:

Schritt 1: Parametervalidierung

- Überprüfung der Stichprobengröße gegen verfügbare Datenmenge
- Validierung der Bereichsdefinition
- Konsistenzprüfung der Spaltenauswahl

Schritt 2: Bereichsextraktion

- Filterung des Datensatzes entsprechend der definierten Start- und Endindizes

- Vorbereitung der Grundgesamtheit für die Zufallsauswahl

Schritt 3: Zufallsziehung

- Initialisierung des Zufallsgenerators mit dem definierten Seed-Wert
- Durchführung der mathematisch korrekten Stichprobenziehung
- Berücksichtigung der Duplikat-Einstellung

Schritt 4: Spaltenfilterung

- Anwendung der konfigurierten Spaltenauswahl
- Optimierung der resultierenden Datenstruktur

Schritt 5: Dateispeicherung

- Export der Stichprobe im Parquet-Format
- Automatische Dokumentation aller verwendeten Parameter in den Datei-Metadaten
- Validierung der erfolgreichen Speicherung

Automatische Metadaten-Dokumentation

Das System dokumentiert automatisch alle relevanten Parameter der Stichprobenziehung in den Metadaten der resultierenden Parquet-Datei. Diese Dokumentation umfasst:

Stichprobenparameter:

- Anzahl der gezogenen Datensätze
- Verwendeter Seed-Wert
- Definierter Auswahlbereich (Start- und Endindex)
- Duplikat-Einstellung
- Zeitstempel der Durchführung

Quelldaten-Informationen:

- Pfad und Name der ursprünglichen Datei
- Gesamtzahl der verfügbaren Datensätze in der Grundgesamtheit
- Ausgewählte Spalten und deren ursprüngliche Positionen

Diese Metadaten sind über die **Verlauf-Funktion** der Anwendung einsehbar und stellen eine vollständige Dokumentation der Stichprobenziehung bereit, ohne dass zusätzliche manuelle Aufzeichnungen erforderlich sind.

Fehlerbehandlung und Validierung

Das System implementiert umfassende Validierungsroutinen:

Eingabevalidierung:

- Überprüfung numerischer Bereiche
- Konsistenzprüfung der Index-Definitionen
- Validierung der Dateipfad-Angaben

Laufzeit-Kontrollen:

- Überwachung der Speicherplatz-Verfügbarkeit
- Behandlung von Dateisystem-Fehlern
- Schutz vor Überschreibung kritischer Dateien

14.7.1.3 *Erweiterte Strategien*

Stichprobenumfang-Bestimmung: Die Festlegung der Stichprobengröße sollte auf statistischen Überlegungen basieren und Faktoren wie gewünschte Aussagesicherheit, tolerable Fehlerrate und Variabilität der Grundgesamtheit berücksichtigen. Bei Prüfungshandlungen sind zusätzlich das Prüfungsrisiko und die Materialitätsschwelle relevant. Für erste Orientierungen können bewährte Faustregeln herangezogen werden, bei kritischen Anwendungen sollte jedoch eine detaillierte statistische Planung erfolgen.

Seed-Management: Die systematische Verwaltung von Seed-Werten ist entscheidend für die Reproduzierbarkeit und Dokumentation. Empfehlenswert ist die Entwicklung einer projektspezifischen Seed-Systematik, die beispielsweise das Datum der Ziehung, die Projektkennung oder andere relevante Parameter berücksichtigt. Der Standard-Seed-Wert von 9386 kann als Ausgangspunkt dienen, sollte aber für produktive Anwendungen projektspezifisch angepasst werden.

Schichtung und Stratifikation: Für inhomogene Grundgesamtheiten kann es sinnvoll sein, vor der Stichprobenziehung eine Schichtung vorzunehmen. Dies bedeutet, dass der Datensatz zunächst in homogene Teilgruppen (Schichten) unterteilt wird, aus denen dann separate Stichproben gezogen werden. Die Bereichseinschränkung der Random Sampling Funktion unterstützt diesen Ansatz durch die Möglichkeit der gezielten Auswahl aus Datensegmenten.

Qualitätssicherung und Nachvollziehbarkeit

Validierung der Stichprobenqualität: Nach der Ziehung sollte die resultierende Stichprobe auf ihre Repräsentativität hin überprüft werden. Dies kann durch Vergleich wichtiger Kennzahlen (Mittelwerte, Verteilungen, Extremwerte) zwischen Stichprobe

und Grundgesamtheit erfolgen. Signifikante Abweichungen können auf systematische Probleme hinweisen und eine Wiederholung der Ziehung erforderlich machen.

Nutzung der Verlauf-Funktion: Die automatische Metadaten-Dokumentation ermöglicht eine vollständige Nachvollziehbarkeit aller Stichprobenziehungen über die **Verlauf-Funktion** der Anwendung. Diese Funktion zeigt alle verwendeten Parameter, Zeitstempel und Quelldaten-Informationen an, wodurch eine manuelle Dokumentation überflüssig wird. Bei Teamarbeiten oder längerfristigen Projekten können so alle Beteiligten jederzeit die Parameter zurückliegender Stichprobenziehungen einsehen.

Reproduzierbarkeit und Compliance: Die in den Metadaten gespeicherten Parameter gewährleisten, dass jede Stichprobenziehung bei Bedarf exakt reproduziert werden kann. Dies ist besonders wichtig für Compliance-Anforderungen, Qualitätskontrollen oder bei Nachfragen von Aufsichtsbehörden. Die automatische Dokumentation eliminiert das Risiko manueller Dokumentationsfehler und erhöht die Glaubwürdigkeit der Prüfungsarbeit.

Integration in den Prüfungsworkflow

Stichprobenplanung: Die Random Sampling Funktion sollte als integraler Bestandteil der Prüfungsplanung verstanden werden. Die Bestimmung der Stichprobenparameter sollte bereits in der Planungsphase erfolgen und auf der Risikobeurteilung, der Materialitätsfestlegung und den geplanten Prüfungshandlungen basieren. Bei iterativen Prüfungsansätzen können mehrere Stichprobenziehungen mit unterschiedlichen Parametern erforderlich sein.

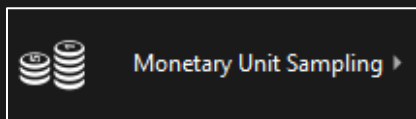
Effizienzoptimierung: Durch die Entwicklung standardisierter Stichprobenverfahren für wiederkehrende Prüfungsaufgaben kann die Effizienz erheblich gesteigert werden. Dies betrifft sowohl die Parameterwahl als auch die Spaltenauswahl. Bewährte Konfigurationen können als Vorlagen dokumentiert und bei ähnlichen Aufgaben wiederverwendet werden. Die automatische Metadaten-Dokumentation unterstützt dabei die Entwicklung von Best-Practice-Ansätzen.

Teamkoordination: Bei Teamprüfungen ist eine koordinierte Stichprobenziehung wichtig, um Doppelarbeiten zu vermeiden und eine gleichmäßige Abdeckung der Grundgesamtheit sicherzustellen. Die Bereichseinschränkung kann dabei zur Aufteilung des Datensatzes auf verschiedene Prüfer genutzt werden. Die automatische Dokumentation in den Metadaten erleichtert die Koordination und gewährleistet Transparenz über alle gezogenen Stichproben.

Kommunikation und Berichtswesen: Die über die Verlauf-Funktion zugänglichen Metadaten können direkt in Prüfungsberichte oder Arbeitspapiere übernommen werden und stellen eine vollständige Dokumentation der angewandten Stichprobenverfahren dar. Die mathematische Korrektheit der Zufallsauswahl in Kombination mit der lückenlosen Dokumentation unterstützt die Glaubwürdigkeit der

Prüfungsergebnisse und erleichtert die Kommunikation mit Mandanten oder Aufsichtsbehörden.

14.7.2 Monetary Unit Sampling (MUS)



Monetary Unit Sampling (MUS) ist eine **statistische Stichprobenmethode**, die vor allem im Prüfungs- und Rechnungswesen eingesetzt wird. Der Kern dieser Methode besteht darin, nicht einzelne Positionen (z. B. Belege oder Buchungen) zu betrachten, sondern stattdessen die Geldeinheiten (z. B. Euro, Dollar) in den Fokus zu rücken. Dadurch steigt die Wahrscheinlichkeit, dass höherwertige Posten ausgewählt werden, da sie mehr Geldeinheiten enthalten und somit öfter im Stichprobenumfang erscheinen können.

Im AUDIPY gibt es nicht nur die Möglichkeit die Analyse durchführen zu können, sondern man kann sich diese auch durch die interne KI bewerten lassen, Der Aufruf erfolgt über die Data-Tools im Header.

Für vertieftes Wissen rund um MUS wird auf Kapitel **14.1.7.2.2** verwiesen.

Monetary Unit Sampling ▶		Erstellung
Second Class	AG-10270	Bewertung
Second Class	ZD-21925	

14.7.2.1 Anwendungsbeispiele & Vorteile

Das MUS-Verfahren erlangt in der Praxis an vielen verschiedenen Stellen große Bedeutung.

Wirtschaftsprüfung

Debitorenprüfung: Ein Wirtschaftsprüfer plant die Prüfung eines Debitorenbestands von 2,5 Millionen Euro mit 15.000 Einzelpositionen. Bei einer Wesentlichkeit von 25.000 Euro und einem Konfidenzniveau von 95% berechnet das System einen Stichprobenumfang von etwa 120 Positionen. Die Anwendung des MUS-Verfahrens stellt sicher, dass größere Debitorenpositionen automatisch eine höhere Auswahlwahrscheinlichkeit erhalten, was der natürlichen Risikoverteilung entspricht.

Lagerbestandsprüfung: Bei der Prüfung eines Lagerbestands mit stark heterogener Wertverteilung (wenige hochwertige Artikel, viele geringwertige Positionen) ermöglicht MUS eine risikoadäquate Auswahl. Hochwertige Lagerpositionen werden entweder als Vollerhebung oder separate Prüfungseinheit behandelt, während die systematische Stichprobe die Grundmasse abdeckt.

Steuerberatung

Umsatzsteuer-Compliance: Steuerberater nutzen MUS für die risikoorientierte Prüfung von Umsatzsteuererklärungen. Bei einer monatlichen Umsatzsumme von 800.000 Euro ermöglicht eine MUS-Stichprobe die effiziente Identifikation von Buchungen mit erhöhtem Korrekturrisiko. Negative Werte (Stornobuchungen, Gutschriften) können separat analysiert werden.

Vorsteuer-Dokumentation: Die systematische Auswahl von Vorsteuer-Belegen basierend auf dem Nettobetrag gewährleistet eine repräsentative Überprüfung der Berechtigung zum Vorsteuerabzug. Großbelege werden automatisch erfasst, während kleinere Beträge proportional zu ihrem Wert berücksichtigt werden.

Controlling und interne Revision

Spesenabrechnung: Controller verwenden MUS zur Stichprobenprüfung von Reisekosten- und Spesenabrechnungen. Bei einem jährlichen Spesenvolumen von 150.000 Euro und einer Toleranzschwelle von 2.000 Euro ermöglicht das Verfahren eine effiziente Identifikation von Abrechnungen mit erhöhtem Prüfungsbedarf.

Kreditoren-Monitoring: Die interne Revision setzt MUS für die risikoorientierte Überprüfung von Kreditorenbuchungen ein. Hohe Einzelbeträge werden vollständig geprüft, während für die Grundmasse eine repräsentative Stichprobe gezogen wird.

14.7.2.2 Methodische Beschreibung MUS

1. Grundidee – Wertproportionale Auswahl

Die gesamte Buchhaltung wird als Zahlenstrahl von 0 bis $N \cdot Y = BV \sum_{i=1}^i$ gedacht. Jede einzelne Geldeinheit belegt dort einen Punkt. Sticht man zufällig in diesen Strahl, so ist die Wahrscheinlichkeit, gerade diese Einheit auszuwählen, für jeden Euro gleich. Ein Posten mit hohem Buchwert besitzt entsprechend viele solcher Punkte – er wird daher proportional häufiger gezogen als ein kleiner Posten. Dieses Prinzip der probability-proportional-to-size (PPS) bildet das Fundament des MUS.

2. Planung des Stichprobenumfangs

2.1. Sicherung gegen Irrtum erster Art (α -Risiko)

Für ein vorgegebenes Vertrauensniveau CL gilt das Irrtumsrisiko $\alpha = 1 - CL$. Damit lässt sich schon unter der extremen Annahme „es wird kein Fehler gefunden“ ein Umfang bestimmen, der die Restunsicherheit auf begrenzt:

$$n = \left\lceil -\ln(\alpha) \frac{Y}{W} \right\rceil, \quad n \geq 1$$

W steht dabei für die Tolerable Misstatement, also den zulässigen Gesamtfehler.

2.2. Gleichzeitige Begrenzung von α und β

Soll zusätzlich das Risiko einer unzulässigen Akzeptanz (β) kontrolliert werden, kommen Poisson-Konfidenzgrenzen zum Einsatz. Für jede potentielle Fehlerzahl Parameter

$$\frac{Y}{W} \lambda_L \leq n \leq \frac{Y}{W} \lambda_U$$

gleichzeitig erfüllbar ist, wird gewählt und auf ganze Posten aufgerundet.

3. Auswahlverfahren – Cell Method

- Intervallbildung: $I = Y/n$.
- Zellenbildung: Der Zahlenstrahl wird in exakt n Zellen der Breite I zerlegt.
- Stichprobenziehung: In jeder Zelle wird unabhängig ein Punkt $h_i = iI + U_i$ mit $U_i \sim U(0, I]$ ermittelt. Das erste Konto, dessen kumulierter Buchwert den Punkt überschreitet, gehört zur Stichprobe. Damit bleibt jede Einschlusswahrscheinlichkeit exakt wertproportional.

4. Überschicht (Top Stratum)

Positionen mit $|BV_i| \geq C$ - standardmäßig $C = I$ - können vollständig geprüft werden. Ihr realer Fehler $\delta_i = BV_i - AV_i$ geht unverändert in die Fehlersumme ein. Alternativ kann man sie in die reguläre Stichprobe aufnehmen, ohne dass sich die nachfolgenden Formeln ändern.

5. Fehlerermittlung und Projektion

Für jede geprüfte Position wird die relative Abweichung (Tainting) berechnet:

$$t_i = \frac{BV_i - AV_i}{BV_i}$$

Normale Posten: Projizierter Fehler: $\hat{\delta}_i = t_i \cdot I$

Top-Stratum: Realer Fehler δ_i ohne Projektionsfaktor.

6. Allowance for Sampling Risk

Für jede α -Risikostufe existiert eine 21-gliedrige Confidence-Factor-Tabelle

$$CF(0) \dots CF(20)$$

Basic Precision: $BP = CF(0) \cdot I$

Dynamische Anhebung: Übersteigt ein einzelner projizierter Fehler $|\hat{\delta}_i|$ diese Basis, wird auf genau diesen Maximalwert erhöht.

Zuschläge bei gefundenen Fehlern: Sortiert man die projizierten Fehler einer Richtung absteigend, erhält Rang k den Zuschlag

$$\Delta_k = [CF(k) - CF(k - 1)] \cdot |\hat{\delta}_{(k)}|.$$

7. Upper Misstatement Limit (UML)

Die obere Fehlergrenze für Überbewertungen lautet:

$$UML_{over} = BP + \sum_k \Delta_k + \sum_{Top} \delta_i$$

Eine analoge Formel gilt für Unterbewertungen; Netto-Grenzen entstehen durch Saldierung.

8. Prüfentscheidung

Erfüllt eine der folgenden Bedingungen

- $UML_{over} > W$
- $|UML_{under}| > W$

ist das Risiko zu hoch; zusätzliche Prüfungsverfahren oder Korrekturen sind erforderlich. Liegen alle UML-Größen unterhalb von W vertretbar, ist die Grundgesamtheit mit dem gewählten Vertrauensniveau

9. Tabellen und Konstanten

Tabelle	Hinterlegte Parameter
Confidence Factors	21 Werte je α -Stufe ($\alpha = 5\%, 10\%, \dots 50\%$)
Poisson-Schranken	λ_L, λ_U für $\alpha, \beta = 1\%, 5\%, 10\%$; Fehlerzahl $k \leq 6$

Zusätzliche Konstanten: Mindestumfang $n \geq 1$, Höchstreihenfolge für Confidence Factors $k_{max} = 20$, Standard-Grenze $C = 1$ für die Überschicht.

10. Zusammenfassung

Die Methode kombiniert eine wertproportionale Ziehung mit wohldefinierten Poisson- und Konfidenzgrenzen.

Aus den Parametern

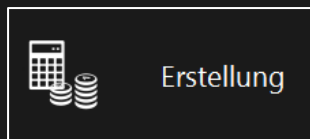
- Gesamtwert Y
- Toleranz W
- α -Risiko (und optional β -Risiko)

lassen sich deterministisch

1. Stichprobenumfang und Intervall
2. Projektionen jeder festgestellten Abweichung,
3. Basic Precision und daraus resultierende Upper Misstatement Limits

ableiten.

14.7.2.3 Erstellung



Erstellung

Im Erstellungsfenster vom MUS gibt es zwei Reiter: **Planung und Extraktion**: Beide haben ihre ganz unterschiedlichen Einstellungen:

14.7.2.3.1 Planung

Im folgenden Bild sieht man, wie MUS konfiguriert wird. Die wichtigsten Elemente im Bereich Planung sind **orange markiert** und wie folgt zu verstehen:

Monetary Unit Sampling

Planung Extraktion

1 Grundgesamtheit

Art der Werte: Positive Werte

Feld: Verkauf

Wert der Grundgesamtheit: 2.988,91

3 Erweitertes Testverfahren (optional)

☐ Fehlerrisiko (β) zusätzlich berücksichtigen

β -Risiko (%): 10,0

2 Einstellungen

Konfidenzniveau (%): 95,00

Materialitätseingabe: Prozent

3,0

4 Vorschau

Stichprobenumfang: 100

Auswahlintervall: 29,89

Hinweis: Berechnet mit Alpha=5,00%, (Wesentlichkeit=89,67); n=100, Intervall=29,89.

Zurücksetzen OK Abbrechen

1. Grundgesamtheit ①

Die Definition der Grundgesamtheit bildet das Fundament jeder MUS-Anwendung. Das System ermöglicht eine differenzierte Herangehensweise an verschiedene Datenstrukturen und Prüfungsziele:

Wertetypen-Auswahl:

- Positive Werte: Standardansatz für die meisten Prüfungshandlungen, insbesondere bei Asset-Prüfungen wie Debitoren, Vorräten oder Anlagevermögen

- Negative Werte: Spezielle Anwendung für Liability-Prüfungen oder Stornobuchungen, beispielsweise bei der Prüfung von Kreditoren oder negativen Bestandsbewegungen
- Absolutwerte: Universeller Ansatz, der sowohl positive als auch negative Beträge in die Stichprobe einbezieht, ideal für Gewinn-und-Verlust-Positionen oder gemischte Bestände

Die Feldauswahl bestimmt, welche numerische Spalte als Basis für die Stichprobenziehung dient. Bei Debitorenprüfungen würde typischerweise der offene Betrag gewählt, bei Lagerprüfungen der Bestandswert, und bei Umsatzprüfungen der Rechnungsbetrag.

Grundgesamtheitswert: Das System berechnet automatisch die Summe aller Werte entsprechend der gewählten Wertetyp-Einstellung. Dieser Wert ist entscheidend für alle nachfolgenden Berechnungen und wird als schreibgeschütztes Informationsfeld angezeigt, um Eingabefehler zu vermeiden.

2. Einstellungen 2

Konfidenzniveau (Vertrauensniveau): Das Konfidenzniveau definiert die statistische Sicherheit der Stichprobe und entspricht dem Komplement des Alpha-Risikos (Risiko falscher Annahme). Typische Werte in der Prüfungspraxis:

- **95%:** Hohe Sicherheit für kritische Prüfungsgebiete oder finale Substantive Tests
- **90%:** Standardniveau für die meisten Prüfungshandlungen
- **85%:** Reduzierte Sicherheit bei vorgelagerten Tests oder unterstützenden Prüfungshandlungen

Wesentlichkeitseingabe: Die Bestimmung der Wesentlichkeit erfolgt wahlweise als absoluter Betrag oder als Prozentsatz der Grundgesamtheit:

Absolute Betragsangabe: Direkte Eingabe des Wesentlichkeitsbetrags in Euro, empfohlen wenn dieser bereits aus der Prüfungsplanung feststeht

Prozentuale Angabe: Berechnung der Wesentlichkeit als Anteil der Grundgesamtheit, nützlich für standardisierte Ansätze oder wenn Wesentlichkeit relativ zur Datenmenge bestimmt werden soll

Die Wesentlichkeit fungiert als Toleranzschwelle für die statistische Bewertung und bestimmt maßgeblich den erforderlichen Stichprobenumfang. Eine niedrigere Wesentlichkeit führt zu größeren Stichproben, bietet aber entsprechend höhere Aufdeckungswahrscheinlichkeit für Fehler.

3. Erweitertes Testverfahren 3

Für anspruchsvolle Prüfungsszenarien bietet das System die Möglichkeit, auch das Beta-Risiko (Risiko der falschen Ablehnung) in die Stichprobenplanung einzubeziehen. Diese Option ist besonders relevant bei:

- **Erstprüfungen** ohne historische Fehlerquoten
- **Risikobereichen** mit unbekannter Fehlerverteilung
- **Regulatorischen Anforderungen** für erhöhte statistische Absicherung

Die Aktivierung der Beta-Berücksichtigung verändert die Berechnungslogik von der standardmäßigen Poisson-Verteilung zu einem erweiterten Alpha-Beta-Modell, was typischerweise zu konservativeren (größeren) Stichprobenumfängen führt.

4. Vorschau 4

Das System berechnet kontinuierlich eine Vorschau der Stichprobenparameter:

Stichprobenumfang (n): Anzahl der zu ziehenden Positionen, berechnet basierend auf den eingestellten Parametern unter Verwendung der statistischen Formeln für MUS

Auswahlintervall: Rechnerischer Abstand zwischen Stichprobenelementen, ermittelt als Quotient aus Grundgesamtheitswert und Stichprobenumfang

Diese Vorschau ermöglicht eine iterative Optimierung der Parameter vor der eigentlichen Stichprobenziehung und hilft bei der Einschätzung des zu erwartenden Prüfungsaufwands.

WICHTIG: Nachdem Sie alle Werte eingetragen haben, müssen Sie zum Reiter Extraktion wechseln, ansonsten ist keine Analyse möglich.

14.7.2.3.2 Extraktion

Im zweiten Reiter „**Extraktion**“ bestimmen Sie die genaue **Zusammenstellung** Ihrer Stichprobe. Die wichtigsten Elemente sind **orange markiert** und wie folgt zu verstehen:

Monetary Unit Sampling

Planung Extraktion

Optionen für Oberschichtwerte

- ☒ Oberschichtwerte aus der Stichprobe entfernen
- ☐ Oberschichtwerte in die Stichprobe einbeziehen
- ☐ Oberschichtwerte in Datei HighValueItems.parquet
- ☐ Betrag der Oberschichtwerte ändern 29,89

Summe

Es sind 0 Elemente mit dem Wert 0 aufgetreten.
Elemente mit dem Wert 0 werden in der Regel nicht gezogen.

	Summe	Datensätze
Positive Werte	2.988,91	10
Negative Werte	0,00	0
Absolutwerte	2.988,91	10

Parameter

Numerisches Feld: Verkauf

Startwert: 26,93

Auswahlintervall: 29,89

Ausgabe

Dateiname: MUS Sample.parquet

Zurücksetzen OK Abbrechen

1. Optionen für Oberschichtwerte ①

Hochwertobjekte (High-Value Items) sind Positionen, deren Einzelwerte das berechnete Auswahlintervall erreichen oder überschreiten. Diese Positionen würden bei der systematischen Ziehung automatisch mehrfach ausgewählt und erfordern daher eine gesonderte Behandlung:

Behandlungsoptionen:

- **Entfernung aus der Stichprobe:** Hochwertobjekte werden von der systematischen Auswahl ausgeschlossen und verbleiben ungeprüft
- **Einbeziehung in die Stichprobe:** Hochwertobjekte werden als Vollerhebungselemente in die finale Stichprobe integriert
- **Separate Datei:** Hochwertobjekte werden in eine eigenständige Parquet-Datei exportiert für separate Prüfungsplanung

Die **Schwellenwertdefinition** für Hochwertobjekte entspricht standardmäßig dem berechneten Auswahlintervall, kann aber bei Bedarf manuell angepasst werden. Eine Reduzierung des Schwellenwerts führt zu mehr Hochwertobjekten, eine Erhöhung zu weniger.

2. Parameter ②

Numerisches Feld: Auswahl des Datenfelds, das als Basis für die Wertberechnung dient - sollte identisch mit der Planungsphase sein

Startwert: Zufällig generierter Ausgangspunkt für die systematische Ziehung, bestimmt welche Position als erste ausgewählt wird. Der Wert muss zwischen 1 und dem Auswahlintervall liegen

Auswahlintervall: Übernommen aus der Planungsphase, kann aber bei Bedarf manuell angepasst werden, wobei Änderungen die statistische Validität beeinflussen

3. Summe ③

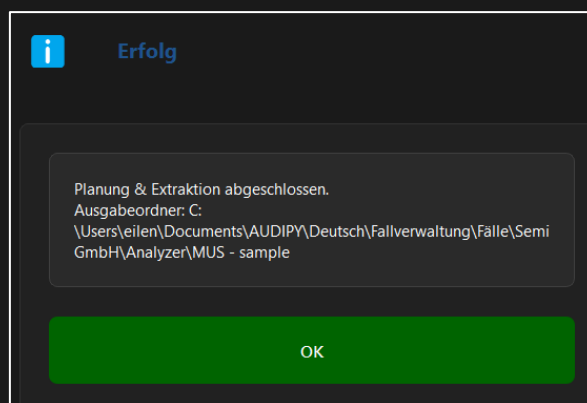
- Zeigt, wie viele Elemente tatsächlich gezogen wurden.
- **Positive Werte, Negative Werte:** Summen und Anzahl der Datensätze, die in die Stichprobe einfließen.
- **Stichprobengröße:** Die Anzahl der Datensätze, die das Tool auf Basis des Intervalls und der Oberschichtoptionen extrahiert.

4. Ausgabe ④

- Gibt an unter welchem Namen die Stichprobe gespeichert wird. So können Sie diese später für die Prüfung oder weitere Analysen öffnen.

Konfigurationsdatei (INI): Vollständige Dokumentation aller verwendeten Parameter für Nachvollziehbarkeit und eventuelle Wiederholung der Stichprobenziehung

Anschließend können Sie mit **Ok die Stichprobe starten** (oder mit „Abbrechen“ den Vorgang wieder schließen) und erhalten folgendes Bestätigungsfenster (inkl. Speicherort). Die Datei wird automatisch geöffnet.



Wenn Sie die Datei direkt im jeweiligen Programm öffnen, werden die **zusätzlichen MUS-Spalten** (z. B. **CumSum**, **Selected**, **AuditedValue**) automatisch am **rechten Ende** der Tabelle angefügt. Um diese Felder zu sehen, müssen Sie daher in der geöffneten Datei **horizontal bis zum rechten Rand** scrollen.

Selected: Boolean-Indikator für die Auswahl

AuditedValue: Kopie des ursprünglichen Buchwertes, vorgesehen für die Eingabe von Prüfwerten

CumSum: Kumulative Summe für nachvollziehbare Dokumentation des Auswahlprozesses

Abhängig von der **Bildschirmauflösung** und der **Tabellenbreite** kann es erforderlich sein, mehrere Spaltenbreiten weiter zu scrollen, bis die MUS-Spalten sichtbar werden. Sobald Sie diese Spalten finden, erhalten Sie einen detaillierten Überblick über die im Rahmen des Monetary Unit Sampling ermittelten Stichproben und deren Kennwerte.

 CumSum	 Selected	 AuditedValue
4....		4.007,8400000000006
7.937,912		3.930,072
11.750,8820000000...		3.812,9700000000003
15.536,1740000000...		3.785,2920000000004
19.147,0220000000...		3.610,848
22.651,9220000000...		3.504,9
26.058,5860000000...		3.406,664
29.452,2660000000...		3.393,68
32.718,6420000000...		3.266,376
35.802,07200000001		3.083,4300000000003

14.7.2.4 Bewertung

Nach Abschluss der Feldarbeit ermöglicht das System die automatisierte Erstellung eines standardisierten Prüfungsberichts. Die PDF-Generierung erfolgt auf Basis der ursprünglichen Planungsparameter und der eingetragenen Prüfwerte:

Berichtsinhalte:

- Vollständige Dokumentation der Planungsparameter
- Statistische Auswertung von Über- und Unterbewertungen
- Berechnung der oberen Fehlergrenze (Upper Error Limit)
- Vergleich mit der definierten Wesentlichkeitsschwelle
- Prüfungsurteil und Empfehlungen

Konfigurationsmöglichkeiten:

- Ersteller-Information für die Dokumentation
- Ausgabeordner für die PDF-Datei
- Vorschau der verwendeten Parameter aus der INI-Konfiguration

Header: Datatools > Stichprobe > Monetary Unit Sampling > Bewertung

Stichprobe		Zufallsstichprobe		20	1.415,43	Adam	Adam
Named Entity Recognition		Monetary Unit Sampling		00	83,28	Cathy	Cathy
							Alex
							Rick
							Gary
							Mari
							Chris
							Darri
	3.023,93	9					
	2.799,96	5					
	2.799,96	4		0.00	1.371,98	Darrin	Darri

Anschließend öffnet sich ein kleines Pogrammfenster mit dem Hinweis, dass Sie die erstellte Datei einmal speichern müssen, bevor sie fortfahren.

Ungespeicherte Änderungen

Haben Sie ungespeicherte Änderungen? Bitte speichern Sie diese, bevor Sie fortfahren.

Ja Nein Abbrechen

Wichtig: Bitte überklicken Sie dieses Fenster nicht nur mit „Ja“, sondern speichern Sie einmal, da ansonsten eine Bewertung nicht durchgeführt werden kann.

Mit Klick auf ja erhalten Sie ein neues Fenster:

MUS - PDF Erstellung

Berichtseinstellungen

Erstellt von:
Max Mustermann

PDF-Ausgabeordner:
"\\Deutsch\Fallverwaltung\Fälle\\Semitax GmbH\\Analyzer\\MUS - sample" ...

Configuration Preview

Beschreibung	Wert
[Planning]	
Konfidenzniveau	95,00
Wesentlichkeitsgrenze als Betrag ...	False
Wesentlichkeitsgrenze	3,0
Positive Werte berücksichtigen	True

PDF erstellen

KI-Analyse erstellen

Berichtseinstellungen ①

- **Erstellt von:**
Eingabefeld für den Namen des Autors / Prüfers.
- **PDF-Ausgabeordner:**
Eingabefeld mit Button ..., um den Speicherort für die erzeugte PDF-Datei auszuwählen. Der Pfad wird im Textfeld angezeigt.

Konfigurationsvorschau ②

In diesem Fenster wird eine tabellarische Vorschau der aktuellen Konfiguration angezeigt. Die Tabelle besteht aus zwei Spalten:

- **Beschreibung:** Enthält die Bezeichnung der einzelnen Parameter.
- **Wert:** Zeigt den aktuell eingestellten Wert für den jeweiligen Parameter.

PDF erstellen – Statistische Bewertungsmethodik ③

Das System implementiert die Tainting-Methode mit separater Behandlung von Über- und Unterbewertungen:

Grundprinzip: Jeder identifizierte Fehler wird als prozentuale Abweichung (Tainting) des Buchwertes erfasst und auf das entsprechende Auswahlintervall

hochgerechnet. Die Aggregation erfolgt unter Berücksichtigung der Confidence Factors entsprechend der AICPA-Tabellen.

Separate Behandlung: Überbewertungen (Buchwert > Prüfwert) und Unterbewertungen (Buchwert < Prüfwert) werden getrennt analysiert, da sie unterschiedliche Risikoimplikationen haben.

Adjustierte Basic Precision: Das System kann die Basis-Präzision automatisch erhöhen, wenn einzelne Projektionen die ursprünglich berechnete Basic Precision überschreiten, was zu konservativeren Hochrechnungen führt.

Erweiterte KI-Analyse 4

Für Anwender mit entsprechenden Berechtigungen (AUDIPY AI PRO) bietet das System eine **automatisierte KI-gestützte Analyse** der Prüfungsergebnisse:

Funktionsumfang:

- Automatisierte Interpretation der statistischen Kennzahlen
- Kontextuelle Einordnung der Ergebnisse
- Risikobewertung und Handlungsempfehlungen
- Integration in den PDF-Bericht

Anwendungsvoraussetzungen:

- Erweiterte Nutzerberechtigungen
- Vollständige Datensätze mit eingetragenen Prüfwerten
- Konfigurierte KI-Schnittstelle

14.7.2.5 *Erweiterte Strategien und Best Practices*

Optimierung der Stichprobeneffizienz

Die Wirksamkeit von MUS-Stichproben kann durch strategische Parameterwahl erheblich gesteigert werden. **Konfidenzniveau-Staffelung** ermöglicht es, für verschiedene Risikobereiche unterschiedliche Sicherheitsniveaus anzuwenden: Während kritische Bilanzpositionen mit 95%iger Sicherheit geprüft werden, können unterstützende Tests mit 90% oder 85% durchgeführt werden.

Wesentlichkeits-Splitting ist eine fortgeschrittene Technik, bei der verschiedene Teilpopulationen mit differenzierten Wesentlichkeitsschwellen behandelt werden. Hochfrequente, niederwertige Transaktionen können mit einer relativen Wesentlichkeit von 5% der Teilsumme geprüft werden, während seltene, hochwertige Geschäftsvorfälle eine absolute Wesentlichkeitsschwelle erhalten.

Integration in Prüfungsworkflows

Mehrphasige Prüfungsansätze kombinieren MUS mit anderen Prüfungstechniken: In der ersten Phase wird mittels MUS eine Grundstichprobe gezogen und geprüft. Basierend auf den Ergebnissen können in einer zweiten Phase gezielte Ergänzungsprüfungen für spezifische Risikobereiche durchgeführt werden.

Dokumentations-Standards gewährleisten die Nachvollziehbarkeit und Verteidigbarkeit der Stichprobenentscheidungen. Die automatisch generierten INI-Dateien und PDF-Berichte sollten durch zusätzliche Arbeitsanweisungen ergänzt werden, die die fachliche Begründung für Parameterwahl und Interpretationen dokumentieren.

Qualitätssicherung und Validierung

Parameter-Plausibilisierung vor der Stichprobenziehung umfasst die Überprüfung von Konfidenzniveau und Wesentlichkeit auf Konsistenz mit der Prüfungsplanung sowie die Validierung der Grundgesamtheit auf Vollständigkeit und Korrektheit der Datenaufbereitung.

Ergebnis-Validierung nach der Feldarbeit beinhaltet die kritische Würdigung der statistischen Hochrechnung im Kontext der praktischen Prüfungserfahrungen sowie die Überprüfung der Schlussfolgerungen auf Angemessenheit und Verhältnismäßigkeit.

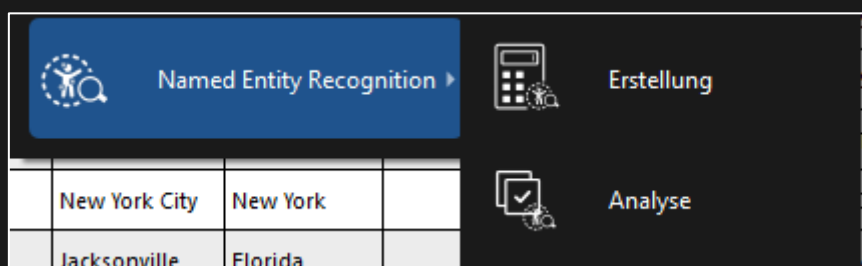
Durch diese systematische Herangehensweise wird MUS zu einem mächtigen Werkzeug für risikoorientierte, effiziente Prüfungshandlungen, das sowohl den statistischen Anforderungen als auch den praktischen Bedürfnissen der beruflichen Praxis gerecht wird.

14.8 Named Entity Recognition

Named Entity Recognition (NER) stellt eine der wichtigsten Funktionen zur automatischen Textanalyse dar und ermöglicht es, aus unstrukturierten Textdaten systematisch relevante Entitäten wie Personen, Organisationen und Orte zu identifizieren und zu kategorisieren. Für Wirtschaftsprüfer, Steuerberater und Controller eröffnet diese Technologie völlig neue Möglichkeiten zur effizienten Bearbeitung großer Dokumentenmengen und zur systematischen Erfassung geschäftsrelevanter Informationen.

Die Funktionalität arbeitet mit zwei verschiedenen Ansätzen: dem klassischen spaCy-NER, das auf bewährten Machine-Learning-Modellen basiert und eine hohe Verarbeitungsgeschwindigkeit bei gleichzeitig geringem Ressourcenverbrauch bietet, sowie dem experimentellen LLM-NER, das auf Large Language Models aufbaut und besonders bei komplexen oder domänenspezifischen Texten überlegene Ergebnisse erzielen kann. Diese technologische Vielfalt ermöglicht es, je nach Anforderung und verfügbaren Ressourcen die optimale Lösung zu wählen.

Der Aufruf der NER erfolgt über die Data-Tools im Header.



Hinweis: Für die NET wird ein spaCy Model benötigt und der Pfad muss bestimmt werden (s. Kapitel [2.2.9.4](#))

Ein **spaCy-Model** ist eine **trainierte Sammlung von Daten und Algorithmen**, die Text automatisch verarbeiten kann, zum Beispiel:

- Wörter erkennen und kategorisieren (Tokenisierung)
- Wortarten bestimmen (z. B. Nomen, Verb, Adjektiv)
- Zusammenhänge im Satz verstehen (Syntaxanalyse)
- Entitäten erkennen (z. B. Namen, Orte, Organisationen)
- Texte vektorisiert darstellen (z. B. für Machine Learning)

Ein Modell enthält alles, was nötig ist, damit **spaCy** schnell und effizient Sprache analysieren kann.

14.8.1 *Praktische Anwendungsszenarien*

Wirtschaftsprüfung

In der Wirtschaftsprüfung ermöglicht NER eine systematische Identifikation von Geschäftspartnern, verbundenen Unternehmen und Schlüsselpersonen aus umfangreichen Dokumentensammlungen. Bei der Prüfung von Konzernabschlüssen können automatisch alle erwähnten Tochtergesellschaften und deren geografische Verteilung extrahiert werden, was die Risikobeurteilung und Wesentlichkeitsanalyse erheblich beschleunigt.

Anwendungsbeispiel Konzernprüfung: Bei der Prüfung eines internationalen Konzerns wird die gesamte Korrespondenz der Geschäftsführung mittels NER analysiert. Das System identifiziert automatisch alle erwähnten Tochtergesellschaften, Joint Ventures und Geschäftspartner. Die anschließende Netzwerkanalyse visualisiert komplexe Beteiligungsstrukturen und deckt potenzielle Interessenkonflikte oder undokumentierte Geschäftsbeziehungen auf.

Compliance-Überwachung: NER kann zur automatischen Überwachung von Compliance-relevanten Erwähnungen eingesetzt werden. Durch die Analyse von E-Mail-Korrespondenz und Besprechungsprotokollen können systematisch Hinweise auf sanktionierte Personen oder Unternehmen identifiziert werden.

Steuerberatung

Steuerberater nutzen NER-Funktionalität zur effizienten Aufbereitung von Betriebsprüfungsunterlagen und zur systematischen Erfassung steuerrelevanter Geschäftsbeziehungen. Die automatische Extraktion von Lieferanten, Kunden und geografischen Bezügen aus Geschäftsunterlagen ermöglicht eine fundierte Vorbereitung auf Außenprüfungen.

Betriebsprüfungsvorbereitung: Vor einer anstehenden Betriebsprüfung werden alle relevanten Geschäftsunterlagen der vergangenen Jahre mittels NER analysiert. Das System extrahiert systematisch alle Geschäftspartner und deren geografische Zuordnung, wodurch grenzüberschreitende Geschäftsbeziehungen und potenzielle Verrechnungspreisthemen frühzeitig identifiziert werden können.

Umsatzsteuer-Analyse: Bei komplexen Umsatzsteuerfällen mit internationalen Bezügen kann NER zur automatischen Identifikation von EU-Ländern und Drittländern aus Geschäftsdokumenten verwendet werden, was die korrekte Zuordnung von Umsatzsteuertatbeständen erheblich vereinfacht.

Controlling

Controller verwenden NER zur automatischen Kategorisierung und Strukturierung von Planungskomentaren, Projektdokumentationen und Budgetbegründungen. Die systematische Extraktion von Abteilungen, Kostenstellen und Verantwortlichen aus unstrukturierten Textdaten ermöglicht eine verbesserte Berichterstattung und Kostenstellenanalyse.

Projektcontrolling: Bei der Analyse umfangreicher Projektdokumentationen extrahiert NER automatisch alle beteiligten Abteilungen, externe Dienstleister und geografischen Standorte. Diese Informationen fließen in die Projektkostenverfolgung ein und ermöglichen eine detaillierte Ressourcenallokationsanalyse.

Budgetanalyse: Aus Kommentaren und Begründungen in Budgetplanungen werden systematisch Kostentreiber, verantwortliche Personen und betroffene Geschäftsbereiche extrahiert. Die Netzwerkanalyse visualisiert Abhängigkeiten zwischen verschiedenen Budgetpositionen und Organisationseinheiten.

14.8.2 NER Dialog - Möglichkeiten

Der NER-Dialog präsentiert sich als modernes, zweiseitiges Interface mit einer klaren Trennung zwischen Modellkonfiguration und Eingabemodus. Die Benutzeroberfläche nutzt Registerkarten zur Unterscheidung zwischen klassischem NER und dem experimentellen LLM-Ansatz, wodurch eine intuitive Navigation zwischen den verschiedenen Verfahren ermöglicht wird.

The screenshot shows a window titled "NER-Spaltenauswahl" with a close button in the top right corner. Inside the window, there is a header area with a logo and the title "NER-Spaltenauswahl". Below the header, there is a tab labeled "Klassische NER". The main content area is divided into two columns. The left column has a section "Modell herunterladen" with dropdowns for "Sprache:" (set to "Deutsch") and "Größe:" (set to "Klein"). Below these is a note: "Das Herunterladen eines Klein-Modells kann einige Minuten dauern." and a button "Modell herunterladen". There is also a section "Modell auswählen" with an empty dropdown. The right column has a section "Eingabemodus auswählen" with three radio buttons: "Datei verwenden", "Ordner verwenden", and "Spalte(n) verwenden" (which is selected). Below this is a section "Spalte auswählen:" with a dropdown showing "Source" and a button "Spalte hinzufügen". There is also a section "Wert hinzufügen:" with a dropdown showing "Originalspalte verwenden". At the bottom of the window, there are two large buttons: "NER extrahieren" (green) and "Abbrechen" (red).

Registerkarte "Classic NER"

Die Registerkarte für klassisches NER umfasst alle Funktionen zur Verwendung traditioneller spaCy-Modelle und ist standardmäßig aktiviert. Sie bietet eine bewährte Balance zwischen Geschwindigkeit und Genauigkeit, die für die meisten Anwendungsfälle in der Wirtschaftsprüfung und Steuerberatung ausreichend ist.

Registerkarte "LLM NER" (Experimentell)

Diese erweiterte Funktionalität steht nur zur Verfügung, wenn experimentelle Features aktiviert sind. LLM-basierte Entitätenerkennung kann bei komplexen Fachtexten oder

branchenspezifischen Dokumenten deutlich bessere Ergebnisse erzielen, erfordert jedoch erheblich mehr Rechenressourcen und bei Cloud-basierten Modellen eine entsprechende API-Konfiguration.

14.8.3 Erstellung einer Classic NER

1. Download Bereich ①

- Der Download-Bereich ermöglicht das direkte Herunterladen und Installieren von spaCy-Sprachmodellen aus dem offiziellen Repository. Diese Modelle bilden die Grundlage für die Entitätenerkennung und müssen einmalig für jede gewünschte Sprache und Modellgröße installiert werden.
- **Sprachauswahl:** Die umfangreiche Sprachunterstützung deckt alle relevanten Geschäftssprachen ab:
 - **Deutsch:** Für deutschsprachige Geschäftskorrespondenz und Verträge
 - **Englisch:** Für internationale Dokumentation und Compliance-Texte
 - **Französisch, Spanisch, Italienisch:** Für europäische Geschäftsbeziehungen

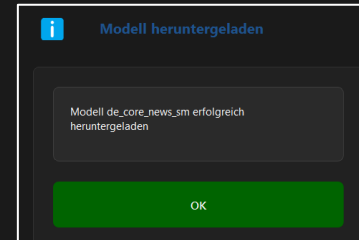
- **Weitere Sprachen:** Chinesisch, Japanisch, Niederländisch und weitere für spezielle Anwendungsfälle
- **Modellgrößen** mit unterschiedlichen Eigenschaften:
 - **Small (Klein):** Kompakte Modelle mit schneller Verarbeitung, ideal für große Dokumentenmengen bei Routineprüfungen
 - **Medium (Mittel):** Ausgewogenes Verhältnis zwischen Geschwindigkeit und Genauigkeit für Standard-Anwendungen
 - **Large (Groß):** Höchste Genauigkeit für kritische Analysen und komplexe Texte
 - **Transformer:** Modernste Technologie mit überlegener Genauigkeit, jedoch deutlich höherem Ressourcenbedarf
- Die Auswahl der Modellgröße sollte strategisch erfolgen: Für Massendatenverarbeitung bei Betriebsprüfungen eignen sich kleine Modelle, während für die Analyse kritischer Verträge oder Compliance-Dokumente große Modelle oder Transformer-Modelle empfohlen werden.
- **Download-Prozess:** Der automatisierte Download erfolgt direkt aus dem spaCy-Repository mit integrierter Fortschrittsanzeige und Konsolen-Output für technische Details. Nach erfolgreichem Download steht das Modell sofort für die Verwendung zur Verfügung.

2. Modellauswahl ②

- Der Bereich zur Modellauswahl zeigt alle lokal verfügbaren Modelle in einer übersichtlichen Dropdown-Liste. Die Darstellung erfolgt im Format "Sprache - Größe" (z.B. "Deutsch - Groß"), was eine schnelle Identifikation und Auswahl ermöglicht. Nicht verfügbare Modelle können direkt über den Download-Bereich nachinstalliert werden.
- Empfehlung: Englisches Transformer-Modell für deutsche Texte: Das englische Transformer-Modell (en_core_web_trf) stellt derzeit die leistungsfähigste Option für die Named Entity Recognition dar und bietet eine bemerkenswerte Eigenschaft: Trotz seiner primären Ausrichtung auf englische Texte erzielt es auch bei deutschsprachigen Dokumenten exzellente Ergebnisse. Diese Fähigkeit resultiert aus der modernen Transformer-Architektur, die auf umfangreichen multilingualen Trainingsdaten basiert und sprachübergreifende Muster in der Entitätserkennung erfasst.
 - **Technische Überlegenheit:** Das Transformer-Modell nutzt Attention-Mechanismen, die kontextuelle Beziehungen zwischen Wörtern deutlich besser erfassen als traditionelle Modelle. Dies führt zu einer signifikant höheren Erkennungsgenauigkeit, insbesondere bei komplexen Entitätsstrukturen wie mehrteiligen Firmennamen, zusammengesetzten Personennamen oder geografischen Bezeichnungen mit Zusätzen.

- Aktueller Stand deutscher Transformer-Modelle: Zum gegenwärtigen Zeitpunkt existiert kein offizielles deutsches Transformer-Modell im spaCy-Repository. Diese Lücke in der Modelllandschaft macht das englische Transformer-Modell zur ersten Wahl für Anwendungen, die höchste Genauigkeit erfordern. Die sprachübergreifende Leistungsfähigkeit kompensiert dabei vollständig das Fehlen eines nativen deutschen Transformer-Modells.

Nach erfolgreichem Download erhält man ein Hinweis-Dialog:



- Anwendungsempfehlungen nach Priorität
 - **Kritische Compliance-Analysen:** Bei Dokumenten mit hoher rechtlicher oder regulatorischer Relevanz sollte ausschließlich das Transformer-Modell verwendet werden. Die zusätzliche Verarbeitungszeit ist hier eindeutig gerechtfertigt durch die überlegene Erkennungsqualität, die falsch-negative oder falsch-positive Ergebnisse minimiert.
 - **Vertragsanalysen und Due Diligence:** Für die Analyse von Kaufverträgen, Fusionsdokumenten oder Compliance-kritischen Vereinbarungen bietet das Transformer-Modell die nötige Präzision zur Identifikation aller relevanten Geschäftspartner und Rechtssubjekte.
 - **Zeitkritische vs. qualitätskritische Prozesse:** Die Entscheidung zwischen Geschwindigkeit und Genauigkeit sollte bewusst getroffen werden. Für explorative Analysen oder Massenverarbeitung können kleinere Modelle ausreichend sein, während für finale Berichte und rechtlich relevante Dokumentationen das Transformer-Modell die erste Wahl darstellt.
- Ressourcenplanung und Performance-Optimierung
 - **Hardwareanforderungen:** Das Transformer-Modell erfordert erheblich mehr Rechenressourcen als konventionelle Modelle. Eine dedizierte GPU kann die Verarbeitungszeit um den Faktor 5-10 reduzieren und sollte bei regelmäßiger Nutzung in Betracht gezogen werden.
 - **Verarbeitungsstrategie:** Für umfangreiche Dokumentensammlungen empfiehlt sich eine hybride Strategie: Erstanalyse mit einem schnellen Modell zur Identifikation relevanter Dokumente, gefolgt von einer Detailanalyse kritischer Inhalte mit dem Transformer-Modell.
- Die Investition in die längere Verarbeitungszeit zahlt sich durch die deutlich höhere Datenqualität und Zuverlässigkeit der Ergebnisse aus, was gerade in professionellen Kontexten der Wirtschaftsprüfung und Steuerberatung von entscheidender Bedeutung ist.

3. Eingabemodi und Datenquellen ③ (verändert sich je nach Auswahl)

- Die Software bietet drei verschiedene Eingabemodi, die unterschiedliche Arbeitsszenarien optimal unterstützen und eine flexible Integration in bestehende Arbeitsabläufe ermöglichen.

Eingabemodus auswählen

☒ Datei verwenden ☐ Ordner verwenden ☐ Spalte(n) verwenden

Textdatei auswählen:

Ausgabeverzeichnis:

- **Einzeldatei-Verarbeitung:** Die Einzeldatei-Verarbeitung eignet sich optimal für die gezielte Analyse spezifischer Dokumente wie Verträge, Gutachten oder Prüfungsberichte. Das System unterstützt sowohl Textdateien (.txt) als auch PDF-Dokumente, wobei bei PDF-Dateien eine automatische Textextraktion erfolgt.
 - **Konfigurationselemente:**
 - **Dateiauswahl:** Browse-Dialog zur Auswahl der zu analysierenden Datei
 - **Ausgabeverzeichnis:** Automatische Vorschlag basierend auf dem Quellverzeichnis mit der Namenskonvention "NER - [Dateiname]"

Für Wirtschaftsprüfer ist dieser Modus besonders relevant bei der Analyse von Jahresabschlüssen, Verträgen oder Compliance-Dokumenten, wo eine präzise Identifikation von Personen, Unternehmen und Orten für die Risikobeurteilung entscheidend ist.

Eingabemodus auswählen

☐ Datei verwenden ☒ Ordner verwenden ☐ Spalte(n) verwenden

Ordner auswählen:

Ausgabeverzeichnis:

- **Ordner-Verarbeitung:** Die Stapelverarbeitung ermöglicht die simultane Analyse aller Textdateien und PDFs innerhalb eines Verzeichnisses. Dies ist besonders effizient bei der Bearbeitung umfangreicher Dokumentensammlungen wie Buchführungsunterlagen, E-Mail-Archiven oder Vertragssammlungen.
- **Verarbeitungslogik:** Das System durchsucht das gewählte Verzeichnis nach allen .txt- und .pdf-Dateien und verarbeitet diese sequenziell. Ein Fortschrittsbalken informiert über den aktuellen Bearbeitungsstand, und die Ergebnisse aller Dateien werden in einer konsolidierten Ausgabedatei zusammengefasst.

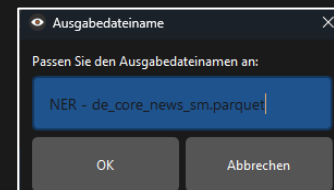
Steuerberater können diese Funktion nutzen, um beispielsweise bei Betriebsprüfungen schnell alle relevanten Geschäftspartner und Transaktionsbeteiligten aus umfangreichen Dokumentensammlungen zu extrahieren.

- **Spalten-Verarbeitung:** Der spaltenbasierte Modus ermöglicht die direkte Verarbeitung von Textdaten aus bereits geladenen Datenrahmen. Dies ist besonders wertvoll, wenn Textinformationen bereits in strukturierter Form vorliegen, etwa aus Datenbank-Exporten oder Excel-Dateien.
 - **Spaltenauswahl:** Über ein intuitives Interface können beliebig viele Textspalten für die NER-Analyse ausgewählt werden. Jede ausgewählte Spalte wird separat verarbeitet, und die Ergebnisse werden entsprechend kategorisiert.

- **Quellenreferenzierung:** Ein wichtiges Feature ermöglicht die Auswahl einer alternativen Spalte für die Quellenangabe. Anstatt der Spaltenbezeichnung kann so beispielsweise eine Dokumenten-ID oder ein Kundennamen als Quelle verwendet werden, was die spätere Nachverfolgung und Zuordnung erheblich erleichtert.

Controller können diese Funktionalität nutzen, um aus Kommentarfeldern in Planungstabellen oder Budgetdokumenten systematisch Abteilungen, Kostenstellen oder Verantwortliche zu extrahieren.

Nach Auswahl des Eingabemodus und des Herunterladens des Modells, kann die NER extrahiert werden (Grüner Button). Anschließend muss man einen Namen vergeben werden.



14.8.4 Erstellung einer LLM-Konfiguration (Experimentell)

Die LLM-Funktionalität unterstützt verschiedene Anbieter und Modellarchitekturen, die jeweils spezifische Vor- und Nachteile aufweisen:

OpenAI GPT-4: Höchste Genauigkeit und Sprachverständnis, jedoch mit Kostenimplikationen und Datenschutzaspekten bei Cloud-Verarbeitung. Erfordert einen gültigen API-Schlüssel.

Alternative Provider: Lokale Modelle von Databricks (Dolly), Meta (Llama2), Mistral AI und anderen, die eine datenschutzkonforme Verarbeitung ermöglichen, jedoch erhebliche lokale Rechenressourcen erfordern.

Sprachkonfiguration

Die Sprachauswahl für LLM-Modelle erfolgt über standardisierte Sprachcodes und beeinflusst die Optimierung des Modells für sprachspezifische Eigenarten und Terminologien. Dies ist besonders relevant bei der Verarbeitung von Fachtexten mit branchenspezifischen Begriffen.

Datenschutz- und Kostenhinweise

Das System integriert kontextspezifische Warnhinweise, die bei der Auswahl cloud-basierter Modelle auf Datenschutzimplikationen und potenzielle API-Kosten hinweisen. Diese Information ist essentiell für Berufsträger, die strengen Verschwiegenheitspflichten unterliegen.

Verarbeitungsprozess und Fortschrittsverfolgung

Automatisierte Verarbeitung

Nach der Konfiguration und Quellenauswahl startet der automatisierte Verarbeitungsprozess. Ein integrierter Fortschrittsbalken informiert über den aktuellen Stand der Bearbeitung, wobei bei umfangreichen Dokumentensammlungen eine präzise Einschätzung der verbleibenden Zeit möglich ist.

Fehlerbehandlung: Das System implementiert robuste Fehlerbehandlungsroutinen, die bei problematischen Dateien (beschädigte PDFs, unlesbare Textcodierung) eine detaillierte Fehlermeldung ausgeben, ohne den gesamten Verarbeitungsprozess zu unterbrechen.

Ergebnisausgabe

Die Verarbeitungsergebnisse werden standardmäßig im hocheffizienten Parquet-Format gespeichert, das eine optimale Balance zwischen Kompression und Zugriffsgeschwindigkeit bietet. Die Ausgabedatei enthält drei zentrale Spalten:

- **Entity:** Die identifizierte Entität (z.B. "Max Mustermann", "Deutsche Bank AG", "Frankfurt am Main")
- **Type:** Die Kategorisierung der Entität (PERSON, ORGANISATION, LOCATION)
- **Source:** Die Quellreferenz (Dateiname oder Spaltenbezeichnung)

Dateinamenskonventionen

Das System schlägt intelligente Dateinamen vor, die den verwendeten Modellnamen und den Verarbeitungsmodus widerspiegeln (z.B. "NER - de_core_news_lg.parquet"). Diese Konvention ermöglicht eine eindeutige Nachverfolgung und Versionierung der Analyseergebnisse.

14.8.5 Analyse

The screenshot shows the 'NER Analyse' interface. It has a dark theme with blue, green, and red buttons. The interface is divided into three main sections: 'Eingabe' (Input), 'Ausgabe' (Output), and 'Netzwerkconfiguration' (Network Configuration). The 'Eingabe' section has two text input fields for 'Einzelne Datei:' and 'Aggregierte Datei:', both containing the path ':/NER - Apple Wiki Artikel/NER Results/NER - Wiki_final.parquet' and a blue button with three dots. The 'Ausgabe' section has one text input field for 'Ausgabedatei:' containing the path '. - Apple Wiki Artikel/NER Results\NER - Wiki_final_final.parquet' and a blue button with three dots. The 'Netzwerkconfiguration' section has a 'Quelle:' field with a button 'Ursprüngliche Spalte verwenden', a 'Ziel:' field, and a checkbox 'Kategorie einbeziehen (Quellspalte)'. At the bottom, there are two large buttons: a green 'OK' button and a red 'Abbrechen' button.

Nach der erfolgreichen NER-Verarbeitung bietet das System eine automatische Aggregationsfunktion, die identische Entitäten zusammenfasst und deren Häufigkeit zählt. Diese Funktionalität ist besonders wertvoll bei der Analyse großer Dokumentensammlungen, wo die gleichen Personen oder Organisationen mehrfach auftreten.

Der Aggregationsprozess erstellt eine separate Datei mit dem Suffix "_user", die zusätzliche Spalten für die manuelle Nachbearbeitung enthält:

- **Entity_user:** Editierbare Version der Entitätsbezeichnung
- **Type_user:** Anpassbare Kategorisierung
- **include:** Boolean-Flag für die Inclusion/Exclusion in nachgelagerte Analysen
- **Count:** Häufigkeit des Auftretens

Manuelle Verfeinerung

Die aggregierten Ergebnisse können anschließend in der Hauptanwendung geladen und manuell verfeinert werden. Dies ermöglicht die Korrektur von Erkennungsfehlern, die Zusammenführung ähnlicher Entitäten (z.B. "Deutsche Bank" und "Deutsche Bank AG") und die kategorische Neuordnung bei Fehlklassifikationen.

14.8.6 *Erweiterte Strategien und Best Practices*

Modellauswahl-Strategien

Die optimale Modellauswahl hängt von verschiedenen Faktoren ab, die strategisch berücksichtigt werden sollten. Für explorative Analysen großer Datenmengen sind kleine bis mittlere Modelle ausreichend und ermöglichen eine schnelle Ersteinschätzung der Datenlage. Für finale Analysen oder kritische Compliance-Prüfungen sollten große Modelle oder Transformer-Varianten verwendet werden, um maximale Genauigkeit zu gewährleisten.

Mehrstufige Analysestrategie: Ein bewährter Ansatz kombiniert verschiedene Modellgrößen in einem mehrstufigen Verfahren. Zunächst wird mit einem kleinen Modell eine Grobanalyse durchgeführt, um relevante Dokumentensubsets zu identifizieren. Anschließend erfolgt eine Detailanalyse der kritischen Dokumente mit einem großen Modell.

Qualitätssicherung und Validierung

Die automatische Entitätenerkennung erreicht je nach Textqualität und Modelleignung unterschiedliche Genauigkeitsgrade. Eine systematische Qualitätskontrolle sollte bei kritischen Anwendungen implementiert werden:

Stichprobenvalidierung: Bei umfangreichen Analysen sollten repräsentative Stichproben manuell validiert werden, um die Erkennungsqualität zu bewerten und systematische Fehler zu identifizieren.

Plausibilitätsprüfungen: Die aggregierten Ergebnisse sollten auf Plausibilität geprüft werden. Ungewöhnlich häufige oder seltene Entitäten können Hinweise auf Erkennungsfehler oder besonders relevante Geschäftsbeziehungen sein.

Integration in Arbeitsabläufe

Die NER-Funktionalität sollte strategisch in bestehende Arbeitsabläufe integriert werden. Bei regelmäßigen Analysen können Vorlagen mit konfigurierten Modellen und Ausgabepfaden erstellt werden, um eine konsistente Bearbeitung zu gewährleisten.

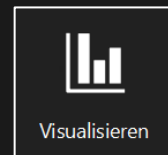
Automatisierung: Für wiederkehrende Analysen kann die NER-Verarbeitung in automatisierte Workflows integriert werden, etwa zur monatlichen Analyse neuer Geschäftskorrespondenz oder zur kontinuierlichen Überwachung von Compliance-relevanten Dokumenten.

Ergebnisarchivierung: Die strukturierte Archivierung von NER-Ergebnissen ermöglicht longitudinale Analysen und die Nachverfolgung der Entwicklung von Geschäftsbeziehungen über längere Zeiträume.

Die Named Entity Recognition stellt damit ein mächtiges Werkzeug für die systematische Erschließung unstrukturierter Textdaten dar und ermöglicht es Wirtschaftsprüfern, Steuerberatern und Controllern, aus großen Dokumentenmengen effizient geschäftsrelevante Informationen zu extrahieren und für weiterführende Analysen zu strukturieren.

15 Visualisieren

Der Bereich **Visualisierung** eröffnet Ihnen zahlreiche Möglichkeiten, Ihre Daten über die klassische Tabellenansicht hinaus anschaulich darzustellen. So können Sie beispielsweise statistische Kennzahlen grafisch aufbereiten und dabei auch leichter Lücken oder fehlende Werte erkennen. Falls Sie eine individuelle Darstellung bevorzugen, bietet der **Chart Builder** die Option, eigene Diagramme zu erstellen, um Ihre Daten gezielt zu analysieren und zu interpretieren.



Hinweis: Beim Aufruf dieser Funktionen ermittelt **AUDIPLY** stets die aktuellen statischen Daten. Es gilt das Prinzip „*What you see is what you get*“ – das heißt, alle **aktiven Filter** werden auch in der Visualisierung berücksichtigt.

Verfügbare Analysemodi

- **Eindeutige Werte (alle Spalten)** - Dieser Modus bietet einen systematischen Überblick über die Datenverteilung sämtlicher Spalten im Datensatz. Die horizontale Balkendiagramm-Darstellung visualisiert die Anzahl eindeutiger Werte pro Spalte, wodurch sich schnell Anomalien, potenzielle Schlüsselfelder und Datenqualitätsprobleme identifizieren lassen.
- **Eindeutige Werte (Spalten-spezifisch)** - Hier erfolgt eine detaillierte Analyse einzelner Spalten mit der Möglichkeit zur seitenweisen Navigation durch umfangreiche Wertelisten. Diese Funktion erweist sich als besonders wertvoll bei der Untersuchung von Kontenplänen, Kostenstellen-Strukturen oder Lieferantenverzeichnissen.
- **Fehlende Werte** - Die Fehlwert-Analyse stellt sowohl absolute Zahlen als auch prozentuale Verteilungen dar und ermöglicht eine sofortige Bewertung der Datenvollständigkeit. Dies ist entscheidend für die Beurteilung der Verlässlichkeit nachgelagerter Analysen.
- **Diagramm-Generator** - Das Herzstück für komplexe Visualisierungen mit umfangreichen Konfigurationsmöglichkeiten für Achsen, Gruppierungen, Zeitreihen-Analysen und spezialisierte Darstellungsformen.

15.1 Arbeitsabläufe und Best Practices

Ein bewährter Ansatz beginnt mit der **Übersichtsanalyse aller Spalten zur Identifikation interessanter Felder**. Anschließend erfolgt die detaillierte Untersuchung auffälliger Spalten mit der Einzelanalyse-Funktion. Die Fehlwert-Analyse sollte parallel durchgeführt werden, um die Verlässlichkeit der Erkenntnisse zu bewerten.

Qualitätssicherung durch Visualisierung

Grafische Darstellungen decken oft Datenqualitätsprobleme auf, die in tabellarischen Ansichten verborgen bleiben. Ungewöhnliche Verteilungen, extreme Ausreißer oder systematische Lücken werden durch Visualisierung sofort erkennbar.

Integration in Prüfungshandlungen

Die verschiedenen Analysemodi unterstützen spezifische Prüfungshandlungen: Vollständigkeitsprüfungen durch Eindeutige-Werte-Analysen, Plausibilitätsprüfungen durch Verteilungsanalysen und Trend-Analysen durch zeitreihenbasierte Visualisierungen.

Das **Modul Visualisierung** stellt somit ein umfassendes Instrumentarium für die moderne datengestützte Prüfungs- und Beratungstätigkeit dar, das sowohl grundlegende Datenqualitätsprüfungen als auch komplexe analytische Verfahren in einer integrierten Benutzeroberfläche vereint.

15.2 Eindeutige Werte (alle)

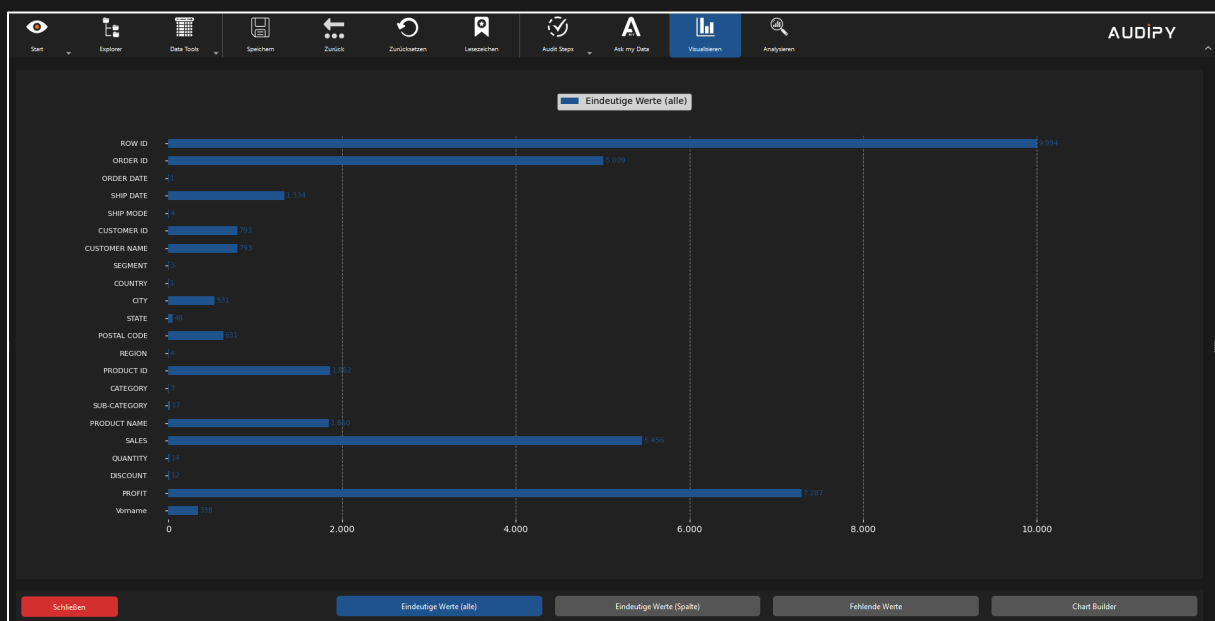
Die **Übersichtsdarstellung aller Spalten** liefert eine fundamentale Einschätzung der Datenstruktur und -qualität. Das horizontale Balkendiagramm ordnet die Spalten nach der Anzahl ihrer eindeutigen Werte, wobei jeder Balken mit präzisen Zahlenwerten annotiert wird.

Interpretationsrichtlinien für die Praxis

Spalten mit sehr hohen eindeutigen Werten (nahe der Gesamtanzahl der Datensätze) deuten typischerweise auf Primärschlüssel, Belegnummern oder Transaktions-IDs hin. In der Wirtschaftsprüfung können solche Felder für Vollständigkeitsprüfungen oder die Identifikation von Duplikaten relevant sein.

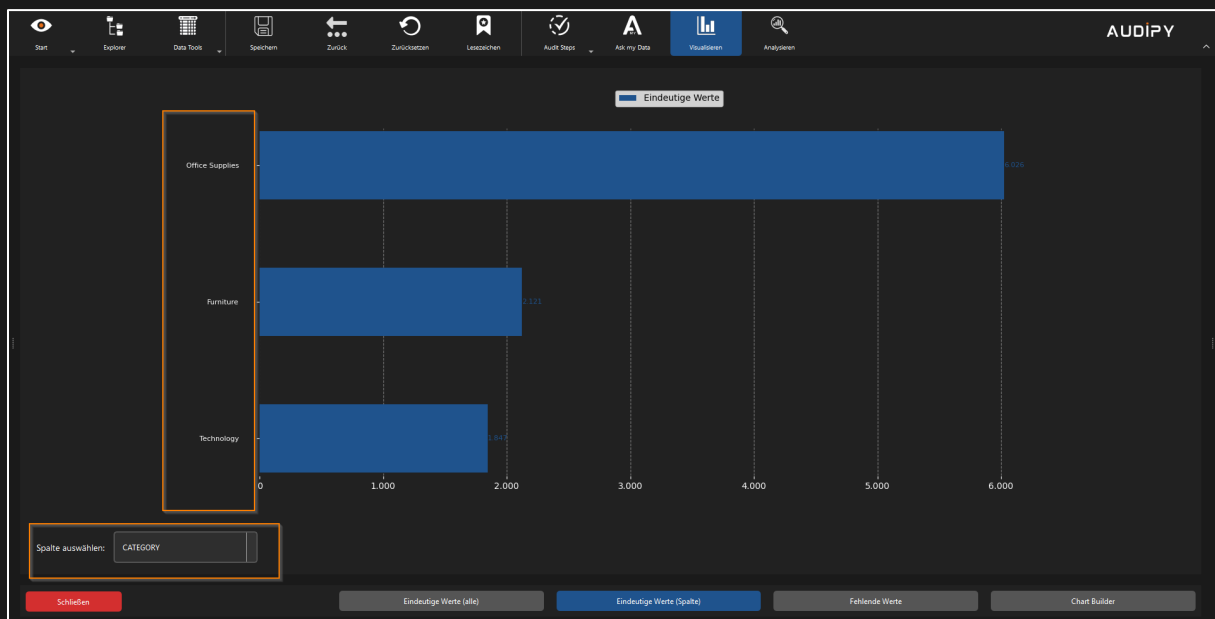
Mittlere Wertebereiche charakterisieren oft kategoriale Variablen wie Kontonummern, Kostenstellen oder Produktgruppen. Diese Felder eignen sich besonders für Gruppierungsanalysen und bilden häufig die Grundlage für risikoorientierte Prüfungsansätze.

Spalten mit sehr wenigen eindeutigen Werten (unter 10) repräsentieren meist Statusfelder, Ja/Nein-Indikatoren oder grobe Klassifikationen. In Controlling-Anwendungen können dies beispielsweise Budgetfreigabe-Status oder Projektphasen-Kennzeichnungen sein.



15.3 Eindeutige Werte (Spalte)

Mit der Darstellung „**Eindeutige Werte (Spalte)**“ können Sie gezielt eine einzelne Spalte auswählen und sich anzeigen lassen, wie viele unterschiedliche Einträge dort vorkommen. Die Einzelspalten-Analyse ermöglicht eine tiefgreifende Untersuchung der Wertverteilungen mit integrierter Paginierung für umfangreiche Datensätze. Die Darstellung erfolgt als absteigende Rangliste der häufigsten Werte mit entsprechenden Frequenzen.



Spalten-Auswahlfeld - Ein Dropdown-Menü am unteren Rand ermöglicht die direkte Auswahl der zu analysierenden Spalte. Das System merkt sich die zuletzt gewählte Spalte für konsistente Arbeitsabläufe.

Paginierung - Bei mehr als 20 eindeutigen Werten aktiviert sich automatisch eine Seitennavigation mit Vor/Zurück-Buttons und Seitenzählung. Dies gewährleistet übersichtliche Darstellungen auch bei umfangreichen Wertelisten.

← Zurück 1 / 500 Weiter →

Label-Trunkierung - Überlange Texteinträge werden automatisch auf 30 Zeichen gekürzt und mit "..." markiert, um die Diagramm-Lesbarkeit zu erhalten.

15.3.1 Anwendungsszenarien in der Berufspraxis

In folgenden Fällen können die Spaltenanalysen eine schnelle und hilfreiche Ergänzung im täglichen Umgang mit Daten sein:

Wirtschaftsprüfung:

Die Analyse von Lieferanten-Spalten kann ungewöhnliche Zahlungsmuster oder potenzielle Scheintransaktionen aufdecken. Hohe Transaktionsvolumen bei seltenen Lieferanten rechtfertigen oft eine vertiefte Prüfung.

Steuerberatung:

Bei der Analyse von Buchungsschlüsseln oder Steuercodes ermöglicht die Frequenzverteilung die Identifikation ungewöhnlicher oder veralteter Codes, die steuerliche Risiken bergen könnten.

Controlling:

Die Untersuchung von Kostenstellen-Buchungen deckt oft unausgewogene Kostenverteilungen oder fehlerhafte Zuordnungen auf, die die Aussagekraft der Kostenstellenrechnung beeinträchtigen.

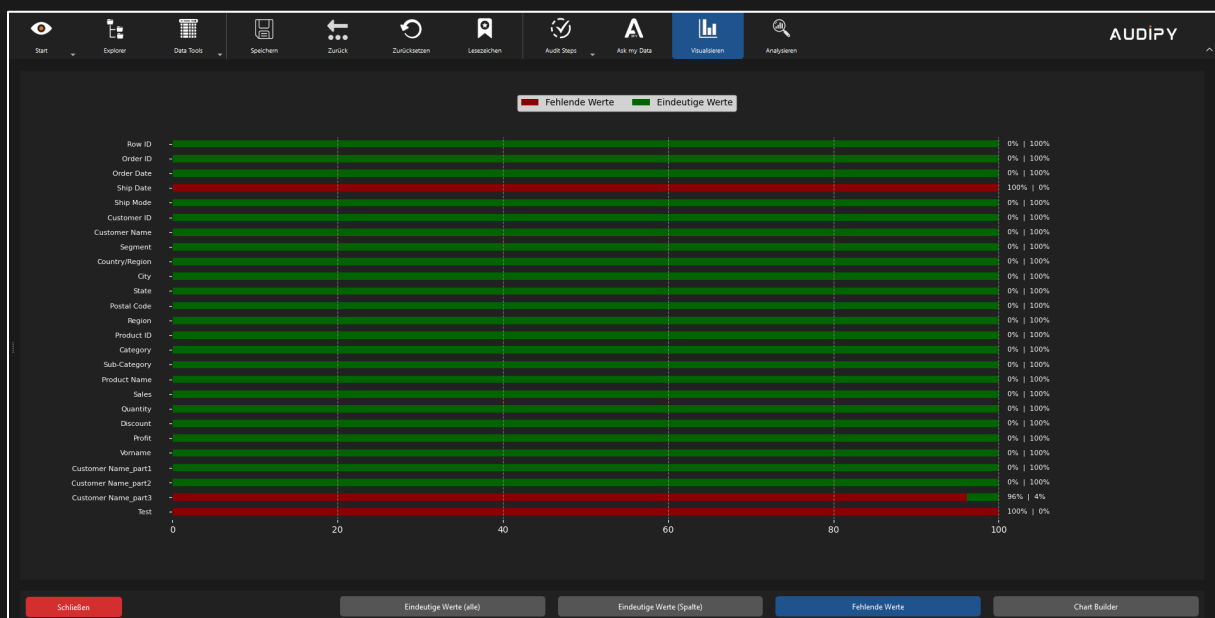
15.4 Fehlende Werte

Die **Fehlwert-Visualisierung** kombiniert absolute und relative Darstellungen in einem 100%-Stapeldiagramm, das für jede Spalte sowohl die Anzahl fehlender Werte als auch deren prozentualen Anteil visualisiert. Diese Darstellung ermöglicht eine sofortige Bewertung der Datenvollständigkeit und identifiziert kritische Qualitätsprobleme.

Darstellungslogik und Interpretation

Das Diagramm teilt jede Spalte in zwei Segmente: einen **roten Bereich** für fehlende Werte und einen **grünen Bereich** für vorhandene eindeutige Werte. Die prozentuale Annotation rechts neben den Balken liefert präzise Zahlen für beide Kategorien.

Spalten mit hohen Fehlwert-Anteilen (über 20%) erfordern besondere Aufmerksamkeit, da sie die Aussagekraft analytischer Verfahren erheblich beeinträchtigen können. In der Wirtschaftsprüfung können solche Lücken auf Systemschwächen, unvollständige Datenmigrationen oder mangelnde Dokumentationsdisziplin hinweisen.



Pflichtfelder:

Fehlwerte in Grunddatenfeldern wie Belegnummern, Buchungsdaten oder Beträgen deuten auf systemische Probleme hin und erfordern sofortige Klärung.

Ergänzungsfelder:

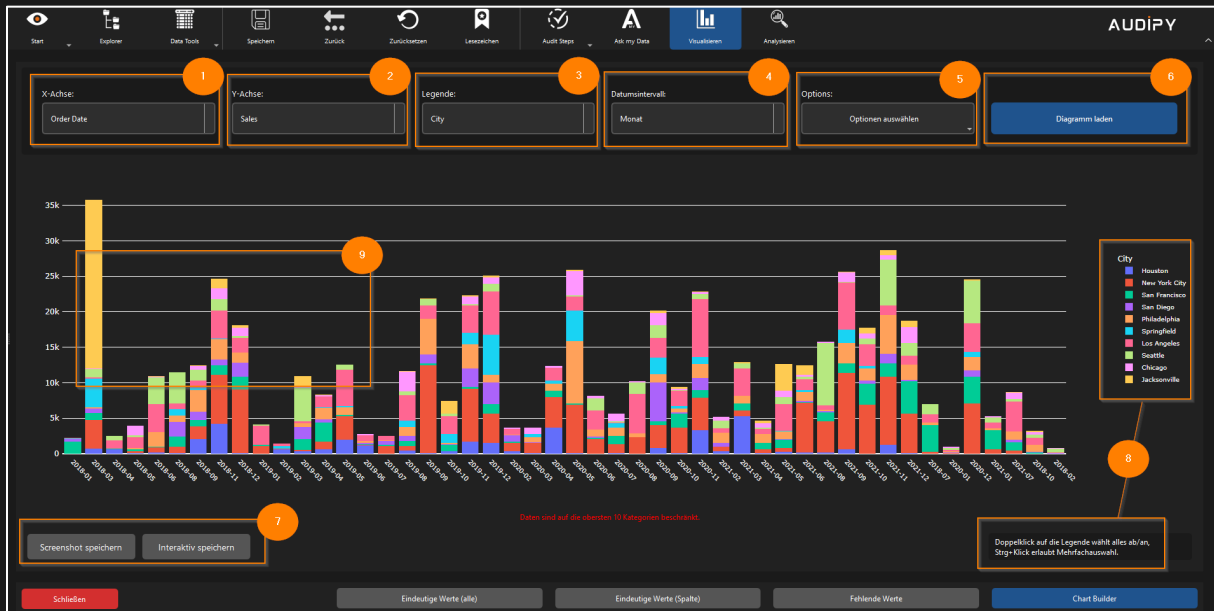
Optionale Felder wie Notizen oder Zusatzinformationen können naturgemäß höhere Fehlwert-Anteile aufweisen, ohne die Datenqualität grundsätzlich zu gefährden.

Berechnete Felder:

Fehlwerte in abgeleiteten Spalten können auf Berechnungsfehler oder unvollständige Algorithmen hinweisen.

15.5 Chart Builder

Der **Chart Builder** stellt das umfangreichste Modul in der Rubrik „**Visualisierung**“ dar und ermöglicht die Erstellung komplexer, interaktiver Visualisierungen mit vielfältigen Konfigurationsmöglichkeiten. Die Benutzeroberfläche gliedert sich in mehrere Konfigurationsbereiche, die eine systematische Diagramm-Erstellung unterstützen.



Im Folgenden sind die wichtigsten Bereiche und Funktionen erläutert:

- **X-Achsen-Auswahl ①**
 - Bestimmt die primäre Gruppierungsvariable für das Diagramm. Die Auswahl beeinflusst alle nachgelagerten Optionen und aktiviert automatisch datumsbasierte Funktionen bei Zeitfeldern.
- **Y-Achse-Auswahl ②**
 - Beschränkt sich auf **numerische Spalten** und definiert die zu aggregierenden Werte. Das System prüft automatisch die Datentypen und zeigt nur geeignete Kandidaten an.
- **Legende (optional) ③**
 - Ermöglicht die zusätzliche Gruppierung nach einer kategorialen Variablen. Diese Option ist besonders wertvoll für die Darstellung mehrdimensionaler Zusammenhänge.
 - Bei Analyse-Dimensionen mit mehr als 30 eindeutigen Kategorien aktiviert das System automatisch ein intelligentes Management-System. Dieses verhindert Performance-Probleme und unübersichtliche Darstellungen durch eine Top-N-Limitierung.

- Das System erkennt automatisch große Kategorien-Sets und bietet dem Anwender die Auswahl zwischen vollständiger Darstellung (mit Warnung vor möglichen Performance-Problemen) und einer Beschränkung auf die relevantesten Kategorien.
- Die Standard-Limitierung auf 30 Kategorien basiert auf visuellen Darstellungsstandards und kognitiver Verarbeitungskapazität. Anwender können diese Grenze nach Bedarf anpassen, wobei das System kontinuierlich über die aktive Limitierung informiert.
- Die Auswahl kann individuell eingestellt werden:

○ **Datumsintervall** 4

- Wird automatisch aktiviert, wenn die X-Achse ein Datumsfeld enthält. Verfügbare Optionen umfassen Monats-, Jahres-, Wochentags-, Stunden- und rollende Fenster-Analysen.

Monatsaggregation - Fasst Datenpunkte zu Monatswerten zusammen und eignet sich besonders für betriebswirtschaftliche Zeitreihen-Analysen. Die Darstellung erfolgt im Format YYYY-MM für internationale Kompatibilität.

Jahresaggregation - Reduziert die Granularität auf Jahreswerte und ermöglicht langfristige Trendanalysen. Besonders relevant für strategische Planungsunterstützung.

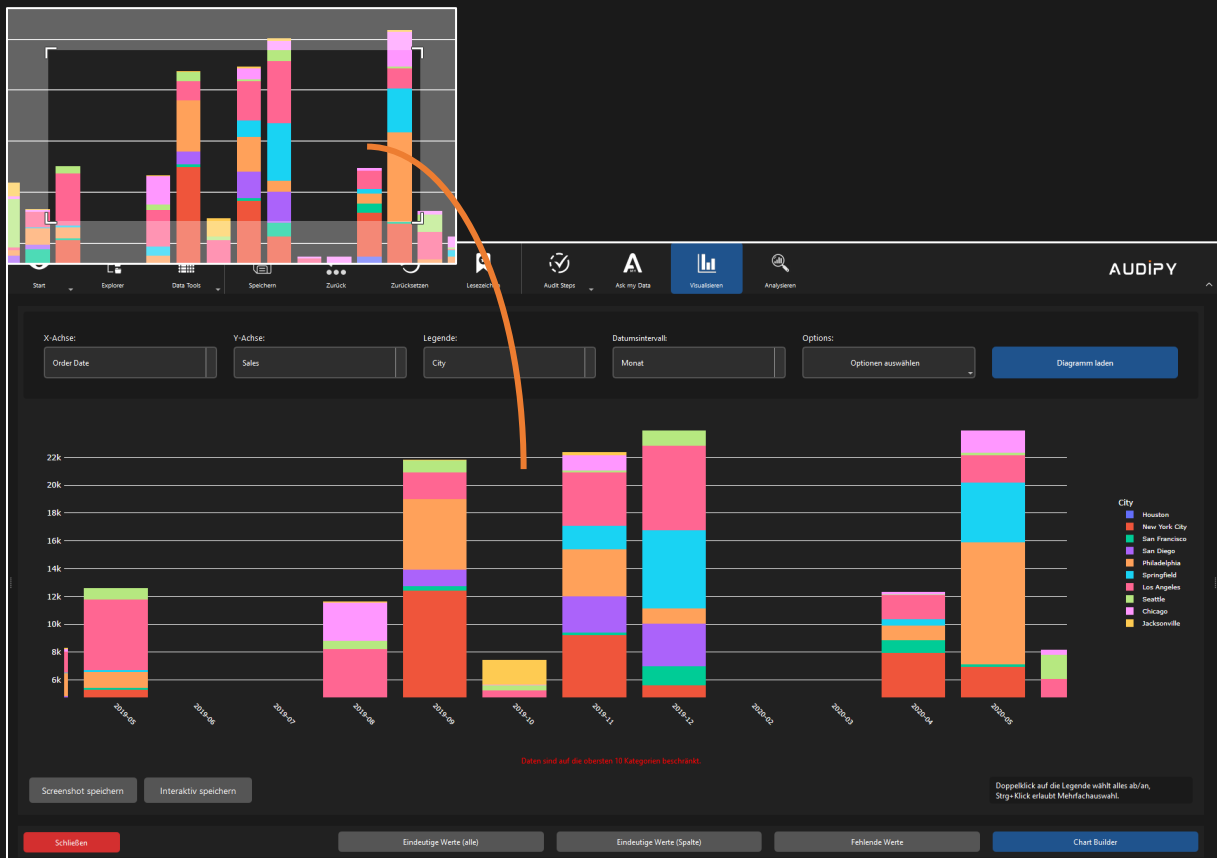
Wochentagsanalyse - Aggregiert Daten nach Wochentagen und deckt zyklische Muster in Geschäftsprozessen auf. Dies kann beispielsweise ungewöhnliche Buchungsmuster an bestimmten Wochentagen identifizieren.

Stundenanalyse - Ermöglicht die Analyse von Tagesverläufen und kann bei der Identifikation von Systemlast-Spitzen oder zeitkritischen Geschäftsprozessen hilfreich sein.

Gleitend - Berechnet gleitende Durchschnitte über benutzerdefinierte Zeiträume (standardmäßig 60 Tage) und glättet kurzfristige Schwankungen für die Trendidentifikation.

- **Optionen auswählen (optional) 5**
 - In Mehrfachauswahl-Dropdown für spezialisierte Darstellungsformen wie **Liniendiagramme** oder **Zeitreihen-Dekomposition**.
 - Die Einstellung **Zeitreihe zerlegen** teilt Ihre zeitbasierten Daten in einzelne Komponenten (z.B. Trend, Saisonalität und Restwerte) auf. Dadurch erhalten Sie einen detaillierten Einblick in die Struktur der Zeitreihe und können Muster oder Auffälligkeiten leichter erkennen.
- **Diagramm laden 6**
 - Über diese Funktion lässt sich das individuelle Diagramm laden bzw. auch wieder abändern, wenn man an den Einstellungen etwas geändert hat.
- **Screenshot speichern / Interaktiv speichern 7**
 - **Screenshot speichern** ermöglicht die sofortige Erstellung hochauflösender PNG-Bilder für Berichtszwecke. Die Funktion respektiert das aktuelle Erscheinungsbild und eignet sich für die Integration in Prüfungsberichte oder Präsentationen.
 - **Interaktiv speichern** erstellt eigenständige HTML-Dateien mit vollständiger Interaktivität. Diese können in Webbrowser geöffnet, mit Kollegen geteilt oder in webbasierte Berichtssysteme eingebettet werden.
- **Filterung: Legende anklicken 8**
 - Durch einen **Klick auf einzelne Legenden-Einträge** können Sie die zugehörigen Datenreihen im Diagramm ein- oder ausblenden. Das ermöglicht eine schnelle, fokussierte Analyse bestimmter Kategorien.
 - Mit **Doppelklick** auf die Legende wählt man die Legende entweder insgesamt ab oder wieder an
 - Mit **Strg + Klick** ist auch eine Mehrfachauswahl möglich (z.B. wenn man im Beispieldiagramm 2 Städte miteinander vergleichen möchte)
- **Interaktive Filterung durch Rechteckauswahl 9**
 - Ziehen Sie mit der Maus ein **Rechteck über einen Bereich im Diagramm**, um nur die darin enthaltenen Daten anzuzeigen und den Rest automatisch auszublenden. Diese Funktion erlaubt es Ihnen, spezifische Ausschnitte zu untersuchen, ohne die gesamte Ansicht zu verlieren.

- Möchte man die ursprüngliche Darstellung wieder haben, bedarf es einen Doppelklick in das Diagramm. So wird die Filterung wieder aufgehoben



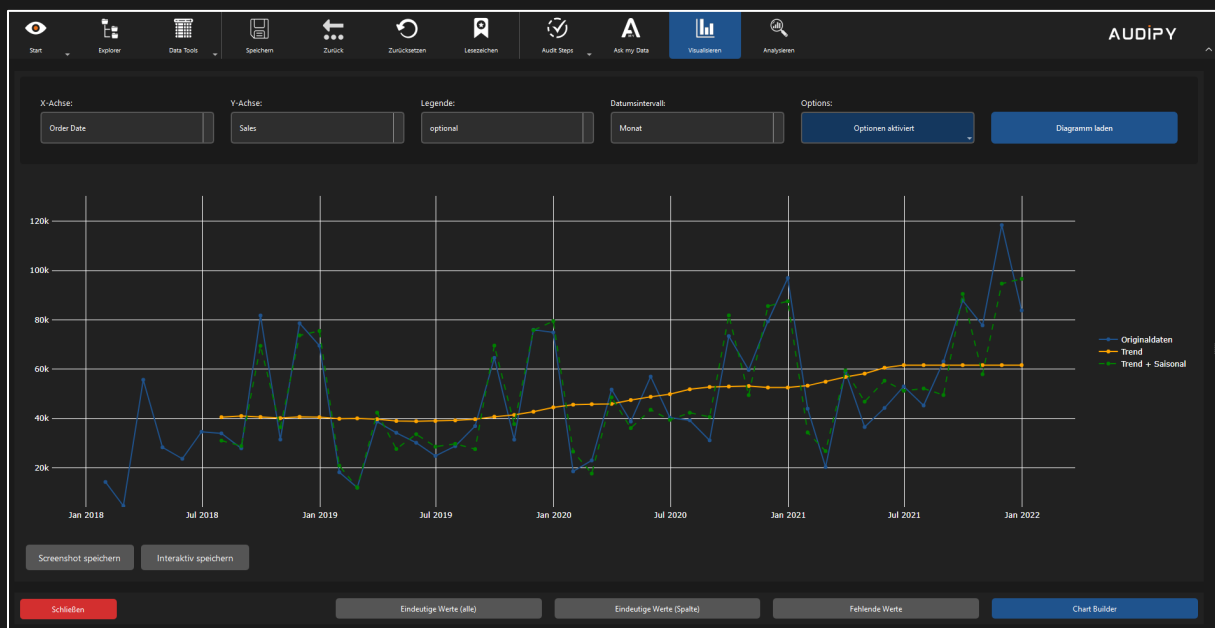
Dank des **Chart Builders** behalten Sie stets die Kontrolle über Ihre Datenvisualisierung. Sie können schnell zwischen verschiedenen Diagrammtypen wechseln, Details ein- und ausblenden und bei Bedarf interaktive Filter anwenden, um Ihre Daten noch präziser zu analysieren.

15.5.1 Zeitreihen-Dekomposition: Analysemethoden

Erweiterte

Unter den erweiterten Optionen im Diagramm kann auch die **Zeitreihen-Dekomposition (Zerlegung)** ausgewählt werden.

Die Zeitreihen-Dekomposition stellt eine hochspezialisierte Analysemethode dar, die komplexe Zeitreihen in ihre Grundkomponenten zerlegt: Trend, saisonale Schwankungen und Residuen. Diese Funktion nutzt statistische Verfahren zur Identifikation langfristiger Entwicklungen und wiederkehrender Muster.



Methodische Grundlagen

Das System implementiert eine additive Dekomposition, die besonders für betriebswirtschaftliche Zeitreihen geeignet ist. Die Methode separiert:

- **Originaldaten**

Die unveränderte Zeitreihe mit allen Schwankungen und Anomalien.

- **Trend-Komponente**

Der langfristige Verlauf nach Eliminierung kurzfristiger Schwankungen. Diese Komponente zeigt die grundlegende Entwicklungsrichtung auf.

- **Trend + Saisonalität**

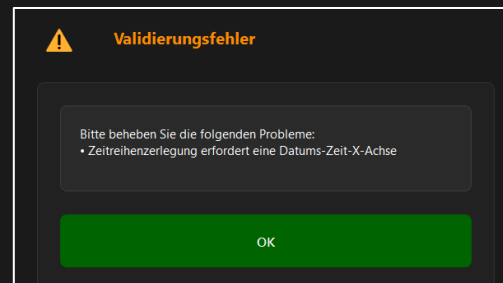
Die Kombination aus langfristigem Trend und wiederkehrenden saisonalen Mustern. Diese Linie sollte idealerweise den Großteil der ursprünglichen Variation erklären.



Praktische Anwendungen der Dekomposition

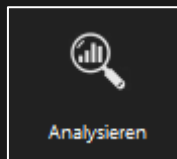
- **Budgetplanung:**
Die Trend-Komponente liefert Basis-Werte für die mittelfristige Planung, während saisonale Muster bei der Liquiditätsplanung berücksichtigt werden können.
- **Anomalie-Erkennung:**
Starke Abweichungen zwischen Original und Trend+Saisonal deuten auf außergewöhnliche Ereignisse hin, die weitere Untersuchung rechtfertigen.
- **Prognose-Unterstützung:**
Die identifizierten Komponenten können als Grundlage für statistische Vorhersagemodelle dienen.

Wichtig: Die Zeitreihen-Zerlegung benötigt eine **Datumsspalte auf der X-Achse** ansonsten erhält man folgenden Hinweis:



Auch hier kann das Ergebnis wie gewohnt als **Screenshot** oder **individuell gespeichert** werden.

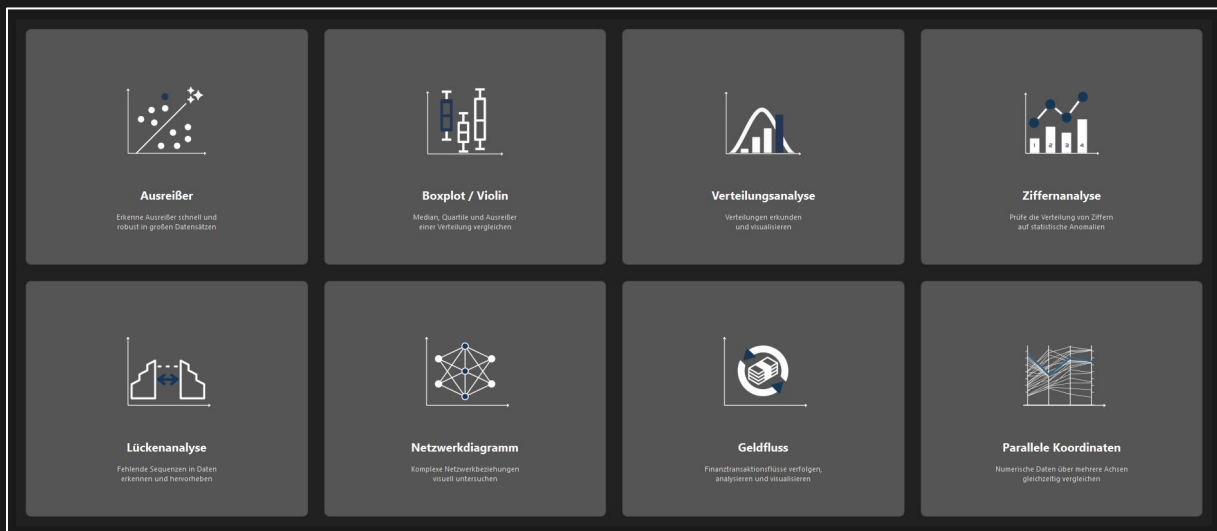
16 Analysieren



Im Bereich **Analysieren** stehen Ihnen verschiedene Werkzeuge und Methoden zur Verfügung, um Ihre Daten gezielt zu untersuchen und weiterführende Erkenntnisse zu gewinnen.

Wie im gesamten Umgang mit AUDIPY gilt auch hier das Grundprinzip „What you see is what you get“ – es wird also nur das analysiert, was zuvor bereits als Datei sichtbar war. Dabei greifen die aktuell gesetzten Filter ebenfalls: Es wird ausschließlich der gefilterte Bereich analysiert.

Folgende **Analysemethoden** bietet der AUDIPY:

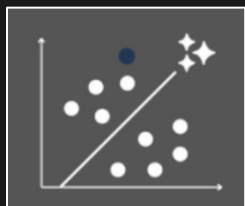


16.1 Überblick mit Beschreibung

Die einzelnen Analysen lassen sich wie folgt beschreiben:

Art	Kurzbeschreibung
Ausreißer	Identifiziert ungewöhnlich hohe oder niedrige Werte, die deutlich vom restlichen Datenbestand abweichen.
Boxplot / Violin	Visualisiert die Verteilung und Streuung der Daten; die Violin-Variante zeigt zusätzlich die Dichteverteilung.
Verteilungsanalyse	Untersucht, wie sich Datenwerte innerhalb eines Merkmals verteilen, z.B. normalverteilt, schief oder mehrgipflig.
Ziffernanalyse	Prüft auf ziffernbezogene Auffälligkeiten, z.B. mit der Benfordschen Regel zur Aufdeckung potenzieller Manipulationen.
Lückenanalyse	Deckt zeitliche oder inhaltliche Lücken in den Daten auf, z. B. fehlende Buchungen oder unregelmäßige Zeitabstände.
Netzwerkdiagramm	Stellt Verbindungen und Beziehungen zwischen verschiedenen Datenpunkten oder Entitäten grafisch dar (z.B. Geschäftspartnerbeziehungen).
Geldfluss	Visualisiert den positiven Fluss von Geldmitteln zwischen Konten, Organisationseinheiten oder Transaktionen zur Erkennung von Mustern und Auffälligkeiten.
Parallele Koordinaten	Visualisiert mehrdimensionale Datenbeziehungen zwischen Spalten, um Muster, Cluster und Abweichungen in komplexen Datensätzen schnell zu erkennen – ideal für Auswertungen mit vielen Einflussfaktoren.

16.2 Ausreißer



Die **Anomalieerkennung** (Ausreißer-Analyse) ist ein hochentwickeltes analytisches Verfahren zur systematischen Identifikation ungewöhnlicher Muster, Ausreißer und verdächtiger Transaktionen in Geschäftsdaten. Diese Methoden nutzen fortgeschrittene statistische und maschinelle Lernalgorithmen, um Datenpunkte zu identifizieren, die signifikant von der normalen Geschäftstätigkeit abweichen. In einer Zeit zunehmender Datenvolumen und komplexer Geschäftsstrukturen ermöglicht die automatisierte Anomalieerkennung eine präzise, objektive und skalierbare Analyse, die manuell nicht mehr effizient durchführbar wäre.

In der Wirtschaftsprüfung revolutioniert die Anomalieerkennung die risikobasierte Prüfungsplanung durch die systematische Identifikation ungewöhnlicher Transaktionen, die weitere Prüfungshandlungen erfordern könnten. Statt auf Stichprobenverfahren oder subjektive Risikoeinschätzungen angewiesen zu sein, können Prüfer mathematisch fundierte Anomalien identifizieren, die auf Fehler, Unregelmäßigkeiten oder sogar betrügerische Aktivitäten hindeuten. Die verschiedenen verfügbaren Algorithmen ermöglichen dabei eine nuancierte Herangehensweise je nach Datencharakteristik und Prüfungszielen.

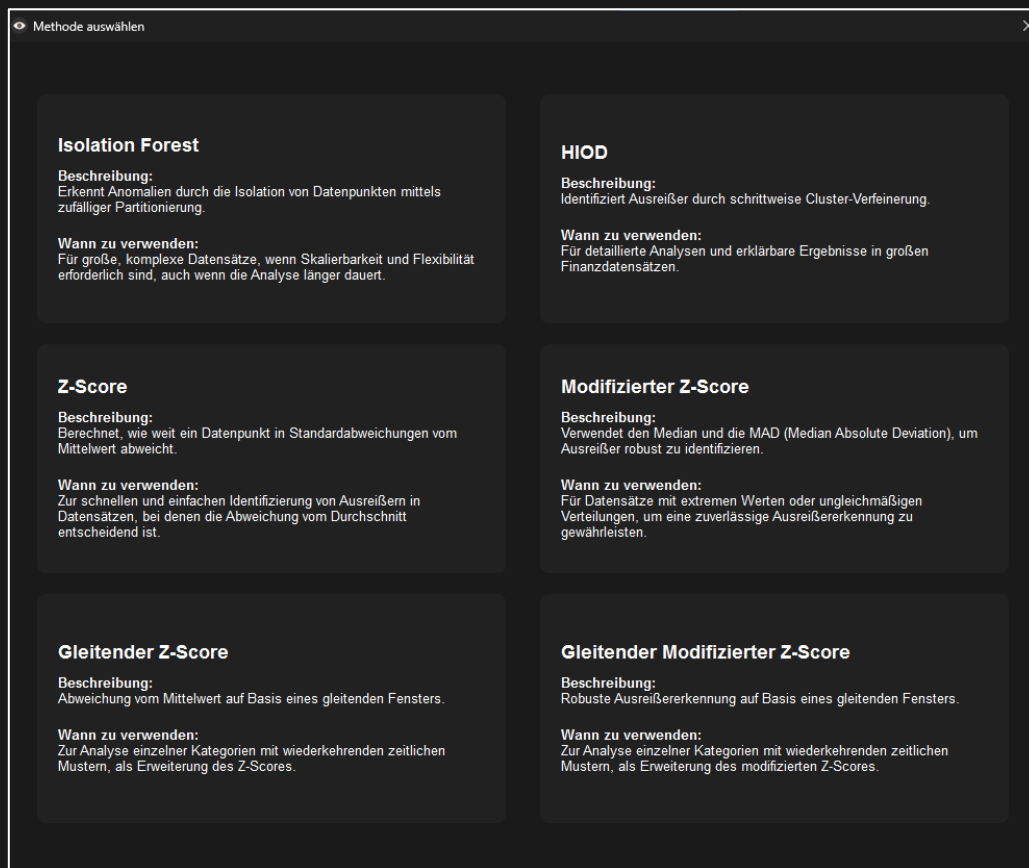
Controller nutzen Anomalieerkennung zur kontinuierlichen Überwachung der Geschäftsprozesse und zur frühzeitigen Identifikation operativer Probleme. Ungewöhnliche Kostenmuster, Budgetabweichungen oder Performance-Anomalien können automatisch erkannt und zur weiteren Untersuchung eskaliert werden. Dies ermöglicht eine proaktive Steuerung und schnelle Reaktion auf sich entwickelnde Probleme, bevor sie sich zu größeren Geschäftsrisiken entwickeln.

Steuerberater verwenden Anomalieerkennung zur Qualitätssicherung der Mandantendaten und zur Vorbereitung auf Betriebsprüfungen. Ungewöhnliche Ausgabenmuster oder verdächtige Transaktionsstrukturen können frühzeitig identifiziert und geklärt werden, wodurch das Risiko von Beanstandungen durch die Finanzverwaltung minimiert wird. Die objektive, mathematische Grundlage der Analyse unterstützt dabei die Argumentation gegenüber Prüfern und erhöht die Glaubwürdigkeit der Dokumentation.

16.2.1 Methoden im Überblick

Über das **Info-Symbol** oben links kann nochmal eine kleine Unterscheidung zwischen den unterschiedlichen Methoden aufgerufen werden. Mit Klick auf eine der Methoden wird diese in die Auswahl aufgenommen.





Die Anwendung bietet sechs verschiedene Anomalieerkennungsmethoden, die jeweils unterschiedliche Stärken und Anwendungsgebiete aufweisen. Die Wahl der optimalen Methode hängt von den Charakteristika Ihrer Daten, der Art der erwarteten Anomalien und den verfügbaren Rechenressourcen ab.

16.2.1.1 Z-Score (Standardisierter Score)

Der Z-Score ist die klassische statistische Methode zur Ausreißererkennung und basiert auf der Berechnung, wie viele Standardabweichungen ein Datenpunkt vom Mittelwert entfernt liegt. Diese Methode ist besonders effektiv bei normalverteilten Daten und bietet eine intuitive Interpretation der Ergebnisse.

Funktionsweise: Für jeden Datenpunkt wird berechnet: $Z = (x - \mu) / \sigma$, wobei μ der Mittelwert und σ die Standardabweichung der Datenreihe ist. Werte mit $|Z| >$ Schwellwert (typisch 3.0) werden als Anomalien klassifiziert.

Anwendungsgebiete:

- Schnelle Identifikation von Ausreißern in Umsätzen, Kosten oder anderen finanziellen Kennzahlen
- Qualitätskontrolle bei gut verstandenen, stabilen Geschäftsprozessen
- Erste Screening-Analysen bei neuen Datensätzen

Stärken: Schnell, einfach interpretierbar, gut geeignet für normalverteilte Daten
Grenzen: Sensibel gegenüber Extremwerten, weniger effektiv bei schiefen Verteilungen

16.2.1.2 *Rolling Z-Score (Gleitender Z-Score)*

Diese erweiterte Version des Z-Scores berechnet Mittelwert und Standardabweichung innerhalb eines gleitenden Zeitfensters, wodurch saisonale Muster und zeitliche Trends berücksichtigt werden können.

Funktionsweise: Für jeden Zeitpunkt wird der Z-Score basierend auf den Daten eines definierten Zeitfensters um diesen Punkt berechnet. Dies ermöglicht die Anpassung an sich ändernde Geschäftsbedingungen und saisonale Schwankungen.

Anwendungsgebiete:

- Analyse von Zeitreihendaten mit saisonalen Mustern
- Überwachung von Geschäftsprozessen mit sich ändernden Baselines
- Identifikation von Anomalien in Kontext ihrer zeitlichen Umgebung

Stärken: Anpassung an zeitliche Trends, Berücksichtigung saisonaler Muster
Grenzen: Höherer Rechenaufwand, benötigt ausreichend Datenpunkte im Zeitfenster

16.2.1.3 *Modified Z-Score (Robuster Z-Score)*

Der Modified Z-Score verwendet Median und Median Absolute Deviation (MAD) anstelle von Mittelwert und Standardabweichung, wodurch er robuster gegenüber Extremwerten wird.

Funktionsweise: $M = 0.6745 \times (x - \text{Median}) / \text{MAD}$, wobei MAD der Median der absoluten Abweichungen vom Median ist. Der Faktor 0.6745 sorgt für Konsistenz mit dem Standard-Z-Score bei normalverteilten Daten.

Anwendungsgebiete:

- Datensätze mit bekannten Extremwerten oder Ausreißern
- Schiefe Verteilungen oder unregelmäßige Datenstrukturen
- Robuste Anomalieerkennung in "verschmutzten" Datensätzen

Stärken: Robust gegenüber Extremwerten, funktioniert bei nicht-normalverteilten Daten
Grenzen: Schwellwerte anders als beim Standard-Z-Score (typisch 3.5-4.0)

16.2.1.4 *Rolling Modified Z-Score (Gleitender Robuster Z-Score)*

Diese Methode kombiniert die Robustheit des Modified Z-Scores mit der zeitlichen Anpassungsfähigkeit des Rolling-Ansatzes.

Funktionsweise: Der Modified Z-Score wird innerhalb eines gleitenden Zeitfensters berechnet, wodurch sowohl zeitliche Trends als auch Robustheit gegenüber Extremwerten erreicht werden.

Anwendungsgebiete:

- Zeitreihendaten mit sowohl saisonalen Mustern als auch Extremwerten
- Kontinuierliche Überwachung unvorhersagbarer Geschäftsprozesse
- Robuste Anomalieerkennung in sich ändernden Umgebungen

Stärken: Kombination aus zeitlicher Anpassung und Robustheit **Grenzen:** Komplexere Parametereinstellung, höherer Rechenaufwand

16.2.1.5 *Isolation Forest*

Der Isolation Forest ist ein moderner Machine-Learning-Algorithmus, der Anomalien durch zufällige Partitionierung des Datenraums identifiziert. Die Grundidee ist, dass Anomalien einfacher zu isolieren sind als normale Datenpunkte.

Funktionsweise: Der Algorithmus erstellt zufällige binäre Entscheidungsbäume und misst, wie viele Schritte nötig sind, um einen Datenpunkt zu isolieren. Anomalien benötigen weniger Schritte und werden entsprechend identifiziert.

Anwendungsgebiete:

- Komplexe, hochdimensionale Datensätze
- Unbekannte Anomaliemuster ohne Vorannahmen über Datenverteilung
- Große Datensätze mit hoher Performance-Anforderung

Stärken: Skalierbar, keine Verteilungsannahmen, multidimensionale Anomalieerkennung **Grenzen:** Längere Berechnungszeit, weniger interpretierbare Ergebnisse

16.2.1.6 *HIOD (Hierarchical Incremental Outlier Detection)*

In der Finanzbuchhaltung spielen präzise Analysen und zuverlässige Methoden zur Erkennung von Anomalien eine zentrale Rolle. Ausreißer in finanziellen Transaktionen können auf Fehler, Betrug oder ungewöhnliche Aktivitäten hinweisen. Die hier vorgestellte Methode, die Hierarchisch-Inkrementelle Outlier Detection (HIOD), bietet eine systematische und skalierbare Technik zur Identifikation solcher Ausreißer. Diese Methode kombiniert k-Means Clustering mit einer schrittweisen Verfeinerung der Klassifikation, um Ausreißer verschiedener Ordnungen zu identifizieren.

16.2.1.6.1 Methode

Die HIOD-Methode folgt einem strukturierten Ansatz zur schrittweisen Erhöhung der Anzahl der Cluster und nutzt dabei das k-Means Clustering, um Ausreißer zu identifizieren. Die Methode besteht aus mehreren Schritten, die aufeinander aufbauen und sukzessive verfeinerte Klassifizierungen erzeugen.

1. Vorbereitung und Mindestanforderungen: Die Methode erfordert eine Mindestanzahl von Datenpunkten, die für eine sinnvolle Clusteranalyse notwendig sind. Hier wurde eine Mindestanzahl von 100 Datenpunkten pro Kategorie (z.B. Konten) festgelegt.
2. Festlegung der Idealklassifizierung: Die Anzahl der idealen Klassen wird durch die Quadratwurzel der Anzahl der Datenpunkte bestimmt. Diese heuristische Methode sorgt für eine ausgeglichene Anzahl von Clustern und Datenpunkten pro Cluster.
3. Erste Klassifikation (50% der Idealklassen): Zu Beginn wird die Anzahl der Cluster auf die Hälfte der idealen Klassen festgelegt. Das k-Means Clustering wird auf die Daten angewendet, und Datenpunkte, die alleine in einem Cluster verbleiben, werden als Ausreißer erster Ordnung klassifiziert.
4. Inkrementelle Erhöhung der Clusterzahl: In den folgenden Schritten wird die Anzahl der Cluster schrittweise um eins erhöht. Nach jeder Erhöhung wird das k-Means Clustering erneut angewendet, und Datenpunkte, die alleine in einem Cluster verbleiben, werden als Ausreißer der jeweiligen Ordnung klassifiziert.
5. Letzte Klassifikation (200% der Idealklassen): Im letzten Schritt wird die Anzahl der Cluster auf das Doppelte der idealen Klassen erhöht. Auch hier werden Einzelpunkte als Ausreißer höchster Ordnung klassifiziert.
6. Erstellung der finalen Tabelle der Ausreißer: Alle identifizierten Ausreißer werden in einer Tabelle zusammengefasst, die nach der Ordnung der Ausreißer sortiert ist.

16.2.1.6.2 Mathematische Beschreibung

Sei n die Anzahl der Datenpunkte in der Analyse. Die Anzahl der idealen Klassen k_{ideal} wird durch die Quadratwurzel von n bestimmt:

$$k_{\text{ideal}} = \lceil \sqrt{n} \rceil$$

Die erste Klassifikation erfolgt mit der Hälfte der idealen Klassen:

$$k_{\text{half}} = \left\lfloor \frac{k_{\text{ideal}}}{2} \right\rfloor$$

Für jede inkrementelle Erhöhung i gilt:

$$k_i = k_{\text{half}} + i \quad \text{für} \quad i \in \{1, 2, \dots, k_{\text{ideal}} - k_{\text{half}}\}$$

Die letzte Klassifikation erfolgt mit dem Doppelten der idealen Klassen:

$$k_{\text{double}} = 2 \cdot k_{\text{ideal}}$$

Datenpunkte, die in einem Cluster der Größe eins verbleiben, werden als Ausreißer der jeweiligen Ordnung klassifiziert. Die finale Tabelle der Ausreißer enthält alle identifizierten Ausreißer, sortiert nach ihrer Ordnung.

Abbildung 1: Zusammengefasste Darstellung aller Ausreißer

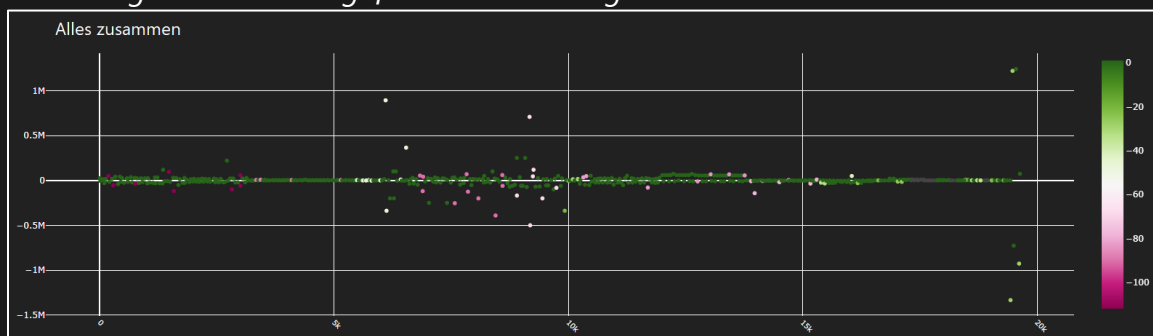


Abbildung 2: Darstellung der Ausreißer für ein Konto (Kfz-Betriebskosten)

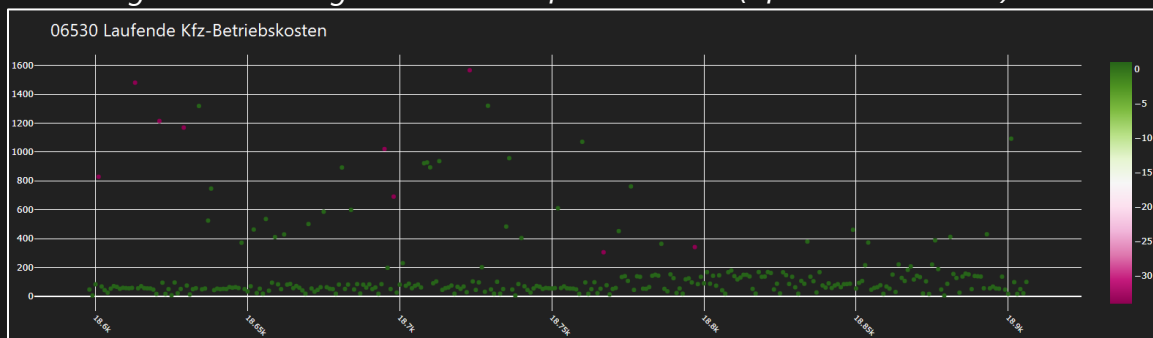
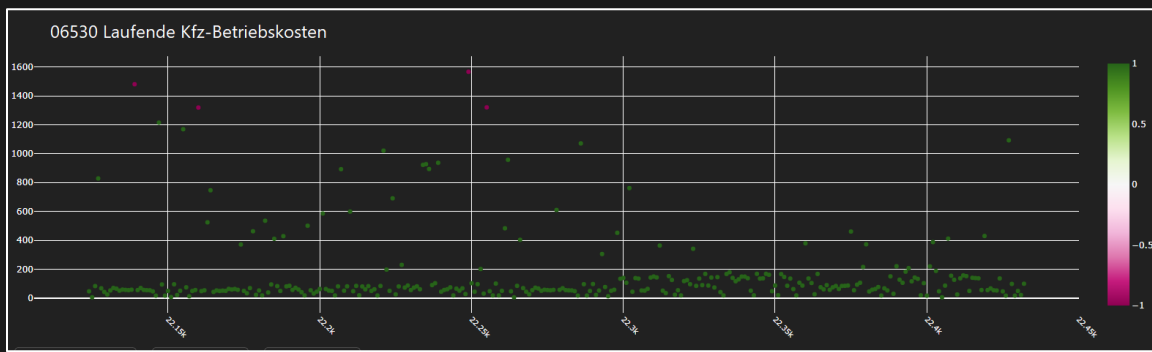


Abbildung 3: Gleiches Konto (Kfz-Betriebskosten) wie in Abbildung 2 mit Isolation Forrest



16.2.1.6.3 Vergleich mit Isolation Forest

Der Isolation Forest ist eine beliebte Methode zur Anomalieerkennung, die auf der Isolation von Datenpunkten basiert. Beide Methoden haben ihre Vor- und Nachteile:

Erklärbarkeit:

- HIOD: Die hierarchisch-inkrementelle Methode ist intuitiv verständlich, da sie auf dem Prinzip der Clusterbildung basiert. Jede Erhöhung der Clusterzahl und die daraus resultierenden Ausreißer können leicht nachvollzogen werden.
- Isolation Forest: Der Isolation Forest basiert auf der Isolation von Datenpunkten durch zufällige Aufteilung der Daten. Die Methode ist weniger intuitiv und schwieriger zu erklären.

Granularität der Analyse:

- HIOD: Die Methode ermöglicht eine feinkörnige Analyse durch die Identifikation von Ausreißern verschiedener Ordnungen. Dies erlaubt eine differenzierte Bewertung der Ausreißer.
- Isolation Forest: Die Methode liefert eine einzige Anomaliewahrscheinlichkeit für jeden Datenpunkt, was weniger Detailtiefe bietet.

Anpassungsfähigkeit:

- HIOD: Durch die schrittweise Erhöhung der Clusterzahl kann die Methode flexibel an die spezifischen Anforderungen und die Struktur der Daten angepasst werden.
- Isolation Forest: Die Methode hat weniger Anpassungsparameter und bietet daher weniger Flexibilität bei der Feinabstimmung.

Rechenaufwand:

- HIOD: Die Methode kann je nach Anzahl der Cluster und Datenpunkte rechenintensiv sein, insbesondere bei großen Datensätzen.

- Isolation Forest: Isolation Forest ist in der Regel schneller und effizienter in Bezug auf den Rechenaufwand, da er weniger komplexe Berechnungen erfordert.

16.2.1.6.4 Anwendungen im Steuerrecht

Die HIOD-Methode kann in der Finanzbuchhaltung angewendet werden, um Anomalien in großen Datensätzen effizient zu identifizieren. Dies ist besonders relevant für Steuerprüfungen und die interne Kontrolle, wo die frühzeitige Erkennung von Unregelmäßigkeiten entscheidend ist. Die hierarchische und inkrementelle Natur der Methode ermöglicht es, Ausreißer verschiedener Ordnungen zu erkennen und eine differenzierte Analyse der Daten durchzuführen.

Die hier vorgestellte Hierarchisch-Inkrementelle Outlier Detection bietet eine robuste und skalierbare Methode zur Erkennung von Ausreißern in großen Finanzdatensätzen. Durch die Kombination von k-Means Clustering mit schrittweiser Verfeinerung der Klassifikation können Ausreißer präzise identifiziert und in ihrer Ordnung klassifiziert werden. Diese Methode ist besonders geeignet für Anwendungen im Steuerrecht, wo die Identifikation von Anomalien in Finanzdaten von großer Bedeutung ist. Im Vergleich zum Isolation Forest bietet die HIOD-Methode eine höhere Erklärbarkeit, größere Granularität der Analyse und Flexibilität, was sie zu einer wertvollen Ergänzung in der Finanzanalyse macht.

16.2.2 Wofür nutzt man eine Ausreißer-Analyse?

Die Ausreißer Analyse hat im beruflichen Alltag eine enorme Bedeutung in verschiedenen Bereichen:

Wirtschaftsprüfung: Automatisierte Risikoidentifikation

In der Jahresabschlussprüfung nutzen Sie die Anomalieerkennung zur systematischen Identifikation prüfungsrelevanter Transaktionen. Sie konfigurieren die Analyse mit "Geschäftsbereich" als Kategorie, "Transaktionsbetrag" als Wert und wählen den Modified Z-Score aufgrund der erwarteten Extremwerte in Finanzdaten.

Die Analyse identifiziert automatisch ungewöhnliche Transaktionen, die in traditionellen Stichprobenverfahren möglicherweise übersehen würden. Ein Geschäftsbereich mit ungewöhnlich vielen Anomalien könnte auf Kontrollschwächen oder sogar Manipulationen hindeuten. Die zeitliche Visualisierung hilft dabei, Anomalien mit spezifischen Geschäftsereignissen oder Systemänderungen zu korrelieren.

Die interaktive Exploration ermöglicht es, von der Makro-Ebene (Gesamtübersicht aller Anomalien) zur Mikro-Ebene (spezifische Transaktionsdetails) zu navigieren. Jede identifizierte Anomalie kann direkt in die weitere Prüfungsplanung integriert werden, wodurch ein evidenzbasierter, risikobasierter Prüfungsansatz realisiert wird.

Controlling: Kontinuierliche Prozessüberwachung

Als Controller implementieren Sie die Anomalieerkennung als Teil Ihres kontinuierlichen Monitoring-Systems für operative Kennzahlen. Sie verwenden Rolling Z-Score für saisonale Geschäftsdaten wie Verkaufszahlen oder Rolling Modified Z-Score für volatilere Kennzahlen wie Produktionskosten.

Das System identifiziert automatisch ungewöhnliche Abweichungen von erwarteten Mustern, lange bevor diese in traditionellen Reporting-Zyklen sichtbar würden. Eine plötzliche Anomalie in den Materialkosten könnte auf Lieferantenprobleme, Qualitätsissues oder Preisvolatilität hindeuten. Die frühe Identifikation ermöglicht proaktive Gegenmaßnahmen anstatt reaktiver Schadensbegrenzung.

Die kategorienbasierte Analyse hilft bei der Priorisierung: Kostenstellen mit vielen Anomalien erhalten erhöhte Aufmerksamkeit, während stabile Bereiche weniger intensive Überwachung benötigen. Das Zeitfenster kann je nach Geschäftstaktung angepasst werden - kürzere Fenster für schnelllebige Bereiche, längere Fenster für strategische Kennzahlen.

Steuerberatung: Datenqualität und Compliance-Überwachung

In der Steuerberatung nutzen Sie Anomalieerkennung zur präventiven Qualitätssicherung der Mandantendaten. Sie analysieren "Ausgabentyp" als Kategorie mit "Ausgabenbetrag" als Wert, wobei der Isolation Forest aufgrund seiner Robustheit gegenüber verschiedenen Ausgabenmustern gewählt wird.

Die Analyse identifiziert ungewöhnliche Ausgabenmuster, die steuerlich problematisch sein könnten oder Dokumentationslücken aufweisen. Regelmäßige Anomalien in bestimmten Ausgabenkategorien können auf systematische Erfassungsfehler oder unzureichende Belegdokumentation hindeuten.

Die präventive Identifikation und Klärung von Anomalien vor Betriebsprüfungen reduziert das Beanstandungsrisiko erheblich. Die mathematische Objektivität der Analyse unterstützt die Argumentation gegenüber Prüfern und demonstriert die Sorgfalt in der Datenqualitätssicherung

16.2.3 Bedienelemente und Parameterkonfiguration

Das **Methoden-Dropdown** bietet die Auswahl zwischen den sechs verfügbaren Algorithmen. Der **Info-Button** neben der Methodenauswahl öffnet einen detaillierten Methodenvergleichsdialog, der Ihnen bei der optimalen Methodenwahl hilft.

Der Methodenvergleichsdialog präsentiert für jede Methode:

- **Detaillierte Beschreibung** der Funktionsweise und mathematischen Grundlagen
- **Anwendungsempfehlungen** basierend auf Datencharakteristika
- **Stärken und Schwächen** für informierte Entscheidungen

Diese Expertenhilfe ist besonders wertvoll für Anwender ohne tiefere statistische Kenntnisse, da sie praxisorientierte Empfehlungen für die Methodenwahl bietet.

Datendimensions-Definition

Das **Kategorie-Feld** ermöglicht die Segmentierung der Anomalieerkennung nach verschiedenen Geschäftsdimensionen:

- **"Alle zusammen"** - Gesamtanalyse über alle Daten hinweg für übergreifende Anomalien
- **Spezifische Kategorien** - Separate Analyse für verschiedene Geschäftsbereiche, Kostenstellen oder Zeiträume

Die kategorienbasierte Analyse ist entscheidend für die praktische Anwendbarkeit, da sich Geschäftsbereiche oft in ihren normalen Schwankungsmustern unterscheiden. Eine Anomalie in einem volatilen Geschäftsbereich könnte in einem stabilen Bereich normal sein.

Das **Wert-Feld** definiert die numerische Variable, die auf Anomalien untersucht wird. Alle numerischen Spalten Ihres Datensatzes stehen zur Verfügung, typischerweise Transaktionsbeträge, Umsätze, Kosten oder andere Geschäftskennzahlen.

Das **Datumsspalte-Feld** ist essentiell für zeitbasierte Analysen und die Darstellung der Ergebnisse. Das System identifiziert automatisch verfügbare Datumsspalten oder zeigt eine entsprechende Warnung an, falls keine vorhanden sind.

Erweiterte Parametersteuerung

Zeitfenster-Konfiguration (Window Size)

Der **Zeitfenster-Slider** steuert die Größe des gleitenden Fensters für Rolling-Methoden und die Kontextberechnung für die Visualisierung. Der Wert wird in Wochen angegeben (2-52 Wochen) und intern in Tage umgerechnet.

Kleine Zeitfenster (2-8 Wochen):

- Höhere Sensitivität für kurzfristige Anomalien
- Schnellere Anpassung an Änderungen
- Geeignet für operative Kontrollen

Große Zeitfenster (26-52 Wochen):

- Stabilere Baseline-Berechnung
- Reduzierte Anzahl false-positive Anomalien
- Geeignet für strategische Langzeitanalysen

Methoden-spezifische Parameter

Der **zweite Slider** passt sich dynamisch an die gewählte Methode an:

Z-Score-Methoden (Z-Score, Rolling Z-Score):

- Schwellwert von 2.0 bis 7.0 (Standard: 3.0)
- Niedrigere Werte: Mehr Anomalien, höhere Sensitivität
- Höhere Werte: Weniger Anomalien, konservativere Erkennung

Modified Z-Score-Methoden:

- Schwellwert von 2.0 bis 7.0 (Standard: 4.0)
- Typischerweise höhere Schwellwerte als beim Standard-Z-Score

Isolation Forest:

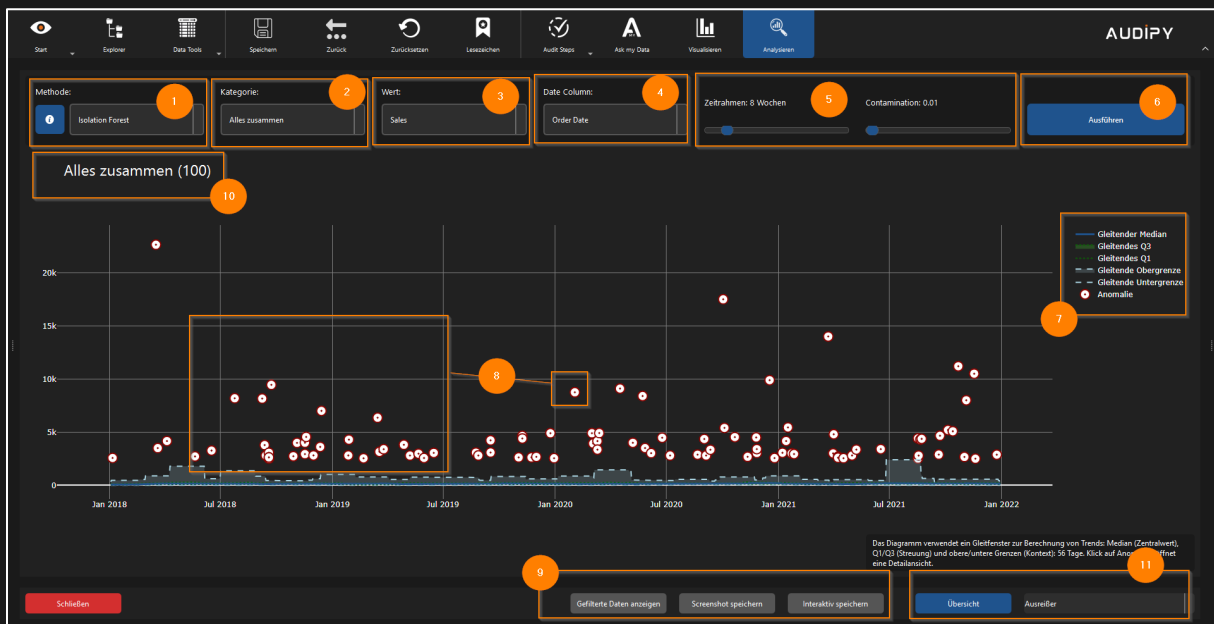
- Contamination-Parameter von 0.01 bis 0.05 (1% bis 5%)
- Repräsentiert den erwarteten Anteil an Anomalien in den Daten
- Niedrigere Werte: Weniger, aber präzisere Anomalien

HIOD:

- Keine Parametereinstellung erforderlich
- Automatische Optimierung der Erkennungsparameter

16.2.4 Erstellung einer Ausreißer-Analyse

Im folgenden Bild sieht man, wie eine Isolation-Forest-Analyse konfiguriert und dargestellt wird. Die wichtigsten Elemente sind **orange markiert** und wie folgt zu verstehen:



1. Methode 1

- Beginnen Sie mit der Klärung Ihrer Analyseziele und nutzen Sie den Info-Button für methodische Unterstützung. Für erste Analysen oder normalverteilte Daten ist der Z-Score ideal. Bei unbekannten Datenstrukturen oder komplexen Mustern wählen Sie Isolation Forest. Für robuste Analysen mit Extremwerten verwenden Sie Modified Z-Score-Varianten.

2. Kategorie 2

- Hier können Sie die **Gruppierung der Daten** festlegen, z.B. nach Produktkategorien, Abteilungen oder Standorten. Beginnen Sie oft mit "Alle zusammen" für einen Gesamtüberblick und segmentieren Sie anschließend nach spezifischen Bereichen für detaillierte Analysen.
- Dies ist eine sehr entscheidende Einstellung, da sie es ermöglicht, Ausreißer gezielt innerhalb bestimmter Gruppen zu untersuchen. **Wichtig, da ohne diese spezifische Gruppierung die Ausreißer oft viel zu grob und wenig aussagekräftig sind. Denken Sie zum Beispiel an eine Kostenanalyse: Ein Ausreißer bei Reisekosten unterscheidet sich grundlegend von einem Ausreißer bei Büromaterial.**

3. Wert 3

- Wählen Sie das **numerische Feld** aus, auf dessen Basis der Algorithmus die Ausreißer bestimmen soll (z. B. **Sales, Umsatz**).

4. Datumsspalte 4

- Wählen Sie die **Datumsspalte** aus, nach der sich das Diagramm ausrichten soll (z.B. *Buchungsdatum*, *Lieferdatum* o.Ä.). Diese Angabe steuert die zeitliche Einordnung der Daten in der Visualisierung.

5. Zeitraum & Filter 5

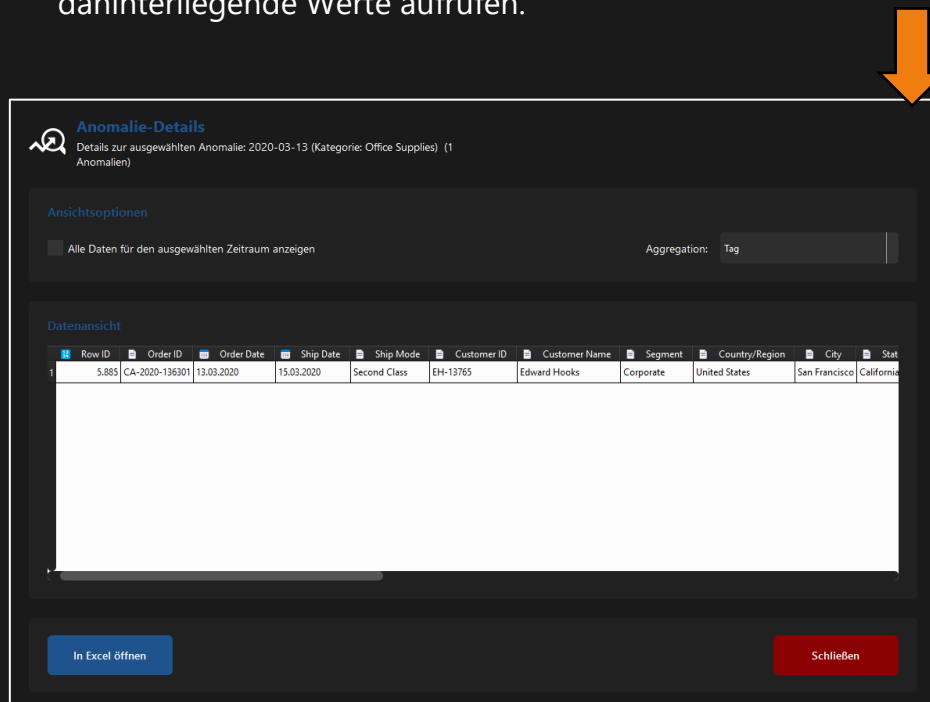
- Konfigurieren Sie das Zeitfenster basierend auf Ihren Datencharakteristika und Analysezielen. Für operative Analysen wählen Sie kleinere Fenster (4-12 Wochen), für strategische Betrachtungen größere Fenster (26-52 Wochen). Justieren Sie die methodenspezifischen Parameter entsprechend Ihrer Risikotoleranz und gewünschten Sensitivität.

6. Ausführen 6

- Starten Sie die Analyse mit "**Ausführen**" und bewerten Sie die Ergebnisse in der interaktiven Visualisierung. Das System präsentiert die Anomalien im Kontext rollierender statistischer Bänder, die die normale Geschäftsvariation repräsentieren.

7. Diagramm & Seitenschenkel 7

- Die **Punkte** im Diagramm repräsentieren einzelne Datenpunkte. Wenn man mit der Maus drüber geht, bekommt man ein kleines Quickview. Darüber hinaus kann man mit Klick auf einen **Datenpunkt** die dahinterliegende Werte aufrufen.



Anomalie-Details
Details zur ausgewählten Anomalie: 2020-03-13 (Kategorie: Office Supplies) (1 Anomalien)

Ansichtsoptionen

☐ Alle Daten für den ausgewählten Zeitraum anzeigen

Aggregation: Tag

Datenansicht

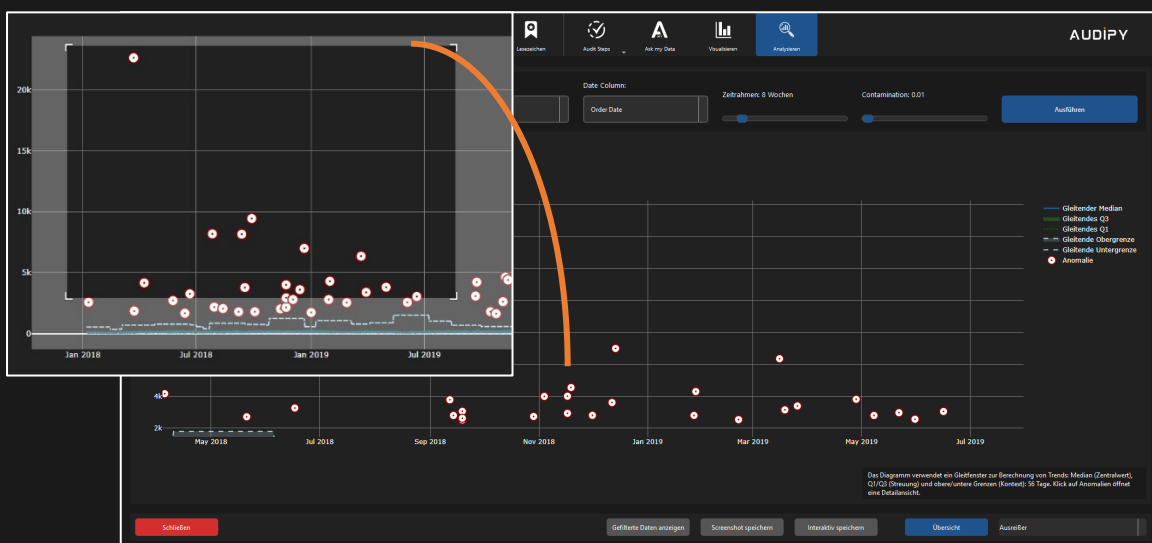
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City	State
1	5.885	CA-2020-136301	13.03.2020	15.03.2020	Second Class	EH-13765	Edward Hooks	Corporate	United States	San Francisco California

[In Excel öffnen](#) [Schließen](#)

- Darüber hinaus kann die Ansicht zur weiteren Bearbeitung auch nach Excel transportiert werden.
- Die **Linie** im Diagramm kann eine **Trendlinie** oder ein gleitender Mittelwert sein, um die Grundtendenz im Zeitverlauf zu verdeutlichen.
- Durch einen **Klick auf einzelne Seitenschenkel-Einträge** können Sie die zugehörigen dazugehörigen Filter / Linien ein-bzw. ausblenden.
-

8. Interaktive Filterung durch Rechteckauswahl 8

- Ziehen Sie mit der Maus ein **Rechteck über einen Bereich im Diagramm**, um nur die darin enthaltenen Daten anzuzeigen und den Rest automatisch auszublenden. Diese Funktion erlaubt es Ihnen, spezifische Ausschnitte zu untersuchen, ohne die gesamte Ansicht zu verlieren.
- Möchte man die ursprüngliche Darstellung wieder haben, bedarf es einen Doppelklick in das Diagramm. So wird die Filterung wieder aufgehoben



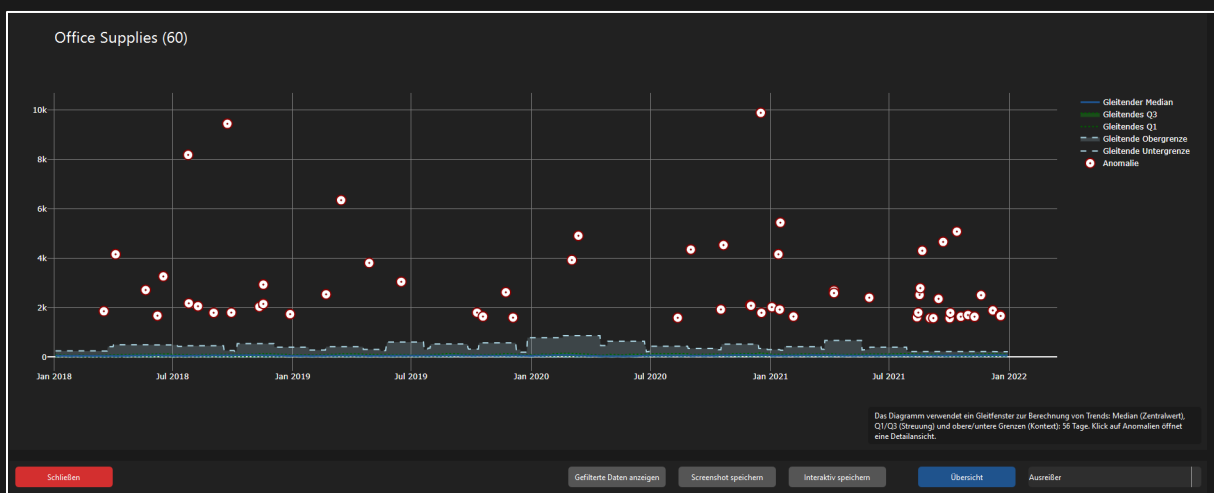
9. Gefilterte Daten / Screenshot speichern / Interaktiv speichern 9

- Über „**Gefilterte Daten anzeigen**“ können Sie gezielt nur die als Ausreißer erkannten Datenpunkte ein- oder ausblenden. Damit lassen sich **verdächtige Werte** schnell isolieren und genauer untersuchen.
- Mit **Screenshot speichern** erstellen Sie ein statisches Bild des Diagramms. Dieses können Sie entweder unter den Falldaten speichern oder können dieses auch an andere Personen zusenden. Hier empfiehlt sich allerdings das interaktive Speichern.
- **Interaktiv speichern** legt das Diagramm als eigenständige HTML-Datei ab, die weiterhin interaktive Funktionen (z.B. Zoom oder Filter) unterstützt. So können Sie es frei verschicken, ohne dass der Empfänger selbst den AUDIPY auf dem Rechner hat. Die HTML-Datei wird ganz normal über den Browser geöffnet.

10. Kategorie Auswahl 10

- Wählt man eine Kategorie so werden standardmäßig die gesamten Datenbestände aus den Kategorien im Diagramm angezeigt, Möchte man nur einzelne / bestimmte Kategorien in der Ansicht haben, so lässt sich dies unter „Alle zusammen“ auswählen.
- Mit einem Klick öffnet sich ein Drill-Down Menü wo man die einzelnen Kategorien angezeigt bekommt (Mit Anzahl der auffälligen Datensätzen in Klammern)
- Im Beispieldiagramm wurde eine Analyse auf die Kategorien gemacht, daher ergibt sich folgendes Bild
- Man kann dann die einzelnen Kategorien auswählen (im Beispiel gibt es z.B. 60 Auffälligkeiten bei Office Supplies):

Alles zusammen (100)
Furniture (22)
Office Supplies (60)
Technology (18)



11. Übersicht / Auswahl (11)

- Mithilfe des blauen Buttons kann zur Analyse Übersicht mit allen möglichen Analysen gewechselt werden. Es kann aber auch das Dropdown Menü benutzt werden um zu einer anderen Analyse zu wechseln.

16.2.5 Visualisierung und Ergebnisinterpretation

Die Anomalieerkennung präsentiert ihre Ergebnisse in einer sophisticateden, interaktiven Zeitreihenvisualisierung, die Anomalien im Kontext der normalen Geschäftsvariation darstellt.

- **Anomalie-Punkte:** Identifizierte Ausreißer werden als hervorgehobene weiße Punkte mit rotem Rand dargestellt. Die Größe und Intensität korrelieren mit der Stärke der Anomalie.
- **Rollierender Median:** Die zentrale blaue Linie zeigt den gleitenden Median über das definierte Zeitfenster und repräsentiert die zentrale Tendenz der normalen Geschäftstätigkeit.
- **Interquartil-Band** (grün): Der Bereich zwischen dem ersten und dritten Quartil (Q1-Q3) zeigt die normale Geschäftsvariation an. Etwa 50% aller normalen Datenpunkte sollten in diesem Bereich liegen.
- **Gesamt-Variationsband** (hellblau): Der äußere Bereich zwischen Minimum und Maximum zeigt die vollständige Bandbreite der historischen Geschäftstätigkeit. Anomalien außerhalb dieses Bereichs sind besonders auffällig.

Interaktive Exploration

Die Visualisierung bietet umfangreiche Interaktionsmöglichkeiten:

- **Klickbare Anomalie-Details:** Klicken auf Anomalie-Punkte öffnet detaillierte Informationen über den spezifischen Datenpunkt, einschließlich Kontext und möglicher Ursachen.
- **Dynamischer Kontext:** Die rollierenden Bänder passen sich automatisch an das gewählte Zeitfenster an und zeigen die historische Variabilität zum Zeitpunkt jeder Anomalie.
- **Zoom und Navigation:** Interaktive Navigation ermöglicht die detaillierte Betrachtung spezifischer Zeiträume und die Analyse von Anomalie-Clustern.

Kategorienbasierte Segmentierung

Das Kategorieauswahlfeld oberhalb der Visualisierung zeigt für jede Kategorie die Anzahl identifizierter Anomalien in Klammern. Dies ermöglicht eine schnelle Priorisierung:

- **Hohe Anomaliezahlen:** Kategorien mit vielen Anomalien verdienen prioritäre Aufmerksamkeit und könnten auf systematische Probleme hindeuten.
- **Null Anomalien:** Kategorien ohne Anomalien zeigen stabilen, vorhersagbaren Betrieb an.
- **Relative Bewertung:** Der Vergleich der Anomaliezahlen zwischen Kategorien hilft bei der Ressourcenallokation für weitere Untersuchungen.

16.2.6 *Erweiterte Analysestrategien und Integration*

Nutzen Sie verschiedene Anomalieerkennungsmethoden für dieselben Daten, um die Robustheit Ihrer Erkenntnisse zu validieren. Konsistente Anomalien über mehrere Methoden hinweg haben höhere Priorität als method-spezifische Ausreißer.

Eine bewährte Strategie ist die Kombination aus schnellem Z-Score-Screening für erste Erkenntnisse, gefolgt von robusteren Modified Z-Score-Analysen für die Detailvalidierung und abschließender Isolation Forest-Analyse für komplexe, multidimensionale Anomalien.

Zeitfenster-Optimierung

Experimentieren Sie systematisch mit **verschiedenen Zeitfenstergrößen**, um die optimale Balance zwischen Sensitivität und Stabilität zu finden. Kleinere Fenster identifizieren mehr kurzfristige Anomalien, während größere Fenster stabilere, langfristige Abweichungen erkennen.

Integration in Workflow-Systeme

Die **Export-Funktionen** ermöglichen die nahtlose Integration der Anomalie-Erkenntnisse in bestehende Workflow- und Dokumentationssysteme. Identifizierte Anomalien können direkt in Prüfungsprogramme, Compliance-Checklisten oder Eskalationsprozesse übertragen werden.

Schwellwert-Kalibrierung

Kalibrieren Sie die **Erkennungsschwellwerte** basierend auf Ihren historischen Erfahrungen und Risikotoleranzen. Dokumentieren Sie die Auswirkungen

verschiedener Schwellwerte auf die Anzahl und Qualität identifizierter Anomalien, um konsistente, reproduzierbare Analysestandards zu entwickeln.

16.2.7 *Best Practices und methodische Empfehlungen*

Investieren Sie in die **Datenqualität vor der Anomalieerkennung**. Unvollständige, inkonsistente oder fehlerhafte Daten können zu false-positive Anomalien führen oder echte Probleme verschleiern. Validieren Sie Datumsspalten, numerische Formate und kategoriale Konsistenz.

Kontextuelle Interpretation

Interpretieren Sie Anomalien immer im Geschäftskontext. Statistische Auffälligkeiten können legitime Geschäftsereignisse, saisonale Effekte oder einmalige Transaktionen repräsentieren. Die **Anomalieerkennung identifiziert statistische Ausreißer** - die geschäftliche Bewertung erfordert Fachwissen.

Dokumentation und Nachvollziehbarkeit

Dokumentieren Sie Ihre **Analyseparameter, -methoden und -erkenntnisse** systematisch. Dies ist besonders wichtig für Prüfungsverfahren oder regulatorische Nachfragen, wo die Nachvollziehbarkeit der Analysen essentiell ist.

Kontinuierliche Validierung

Überprüfen Sie regelmäßig die Effektivität Ihrer **Anomalieerkennungsparameter durch Feedback-Loops**: Wie viele identifizierte Anomalien waren tatsächlich relevant? Welche echten Probleme wurden übersehen? Nutzen Sie diese Erkenntnisse für kontinuierliche Optimierung.

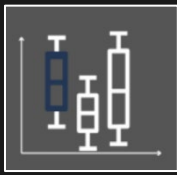
Performance und Skalierbarkeit

Bei **sehr großen Datensätzen** beachten Sie die Performance-Unterschiede zwischen den Methoden. **Z-Score-Varianten** sind am schnellsten, während Machine Learning-Methoden wie Isolation Forest mehr Rechenzeit benötigen, aber auch komplexere Anomalien identifizieren können.

Die **Anomalieerkennung** stellt somit ein mächtiges, wissenschaftlich fundiertes Werkzeug dar, das durch seine Kombination aus statistischer Rigorosität, algorithmischer Vielfalt und praktischer Anwendbarkeit optimal geeignet ist, kritische

Ausreißer und verdächtige Muster in komplexen Geschäftsdaten zu identifizieren und damit objektive Grundlagen für weitere Untersuchungen und fundierte Geschäftsentscheidungen zu schaffen.

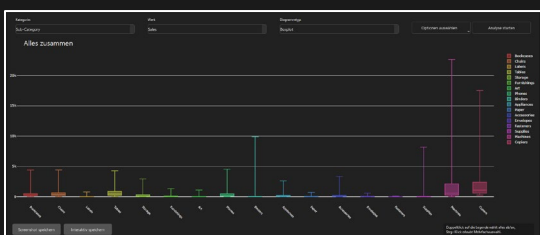
16.3 Boxplot / Violin



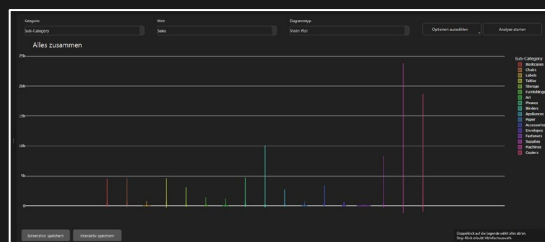
Die **Box-Plot, Violin-Plot und Punkt-Plot Analysefunktion** ist ein essentielles Werkzeug für die moderne Datenanalyse in den Bereichen Wirtschaftsprüfung, Steuerberatung und Controlling. Diese Visualisierungsmethoden ermöglichen es Ihnen, komplexe Finanzdaten schnell zu durchdringen und wichtige Erkenntnisse für Ihre berufliche

Praxis zu gewinnen. Während traditionelle Tabellen und einfache Diagramme oft nur oberflächliche Einblicke bieten, zeigen diese statistischen Visualisierungen die tatsächliche Verteilung Ihrer Daten und decken Anomalien, Trends und Muster auf, die in der Summenbetrachtung verborgen bleiben würden.

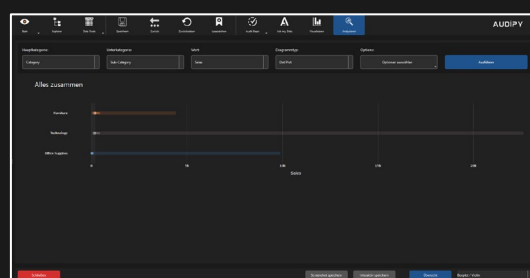
In der Wirtschaftsprüfung können Sie mit diesen Tools beispielsweise Umsatzverteilungen verschiedener Mandanten analysieren, um ungewöhnliche Schwankungen zu identifizieren, die weitere Prüfungshandlungen erfordern könnten. Steuerberater nutzen diese Analysen zur Bewertung von Betriebsausgabenstrukturen ihrer Mandanten im Branchenvergleich oder zur Identifikation steueroptimierender Maßnahmen. Im Controlling ermöglichen diese Visualisierungen die detaillierte Analyse von Kostenstellen, die Bewertung von Budgetabweichungen und die Identifikation von Optimierungspotenzialen in verschiedenen Unternehmensbereichen.



Boxplot



Violin



Punkt Plot (Dot-Plot)

16.3.1 Wofür nutzt man Boxplot/Violin?

In der Praxis können Boxplots/ Violin Diagramme wie folgt sehr nützlich sein:

Wirtschaftsprüfung: Risikoidentifikation und Prüfungsplanung

Ein typisches Szenario in der **Wirtschaftsprüfung** ist die Analyse der monatlichen Umsätze verschiedener Tochtergesellschaften eines Konzerns. Sie beginnen mit einer Box-Plot-Analyse, bei der Sie als Hauptkategorie "Tochtergesellschaft" und als Wert "Monatsumsatz" wählen. Die Option "Ausreißer separat anzeigen" hilft Ihnen, ungewöhnliche Umsatzspitzen oder -einbrüche zu identifizieren.

Das Box-Plot zeigt Ihnen sofort, welche Gesellschaften konsistente Umsätze erzielen (schmale Boxen mit wenigen Ausreißern) und welche große Schwankungen aufweisen (breite Boxen oder viele Ausreißer). Anschließend können Sie über die Kategorieauswahl in einzelne Gesellschaften "hineinzoomen" und mit der Unterkategorie "Monat" die zeitliche Entwicklung analysieren.

Controlling: Kostenstellenanalyse und Budgetcontrolling

Im **Controlling** einer Produktionsgesellschaft steht die Analyse der Personalkosten verschiedener Abteilungen im Fokus. Sie wählen "Abteilung" als Hauptkategorie, "Kostenstelle" als Unterkategorie und "Personalkosten pro Monat" als Wert. Ein Violin-Plot offenbart nicht nur die durchschnittlichen Personalkosten, sondern auch deren Verteilungsmuster.

Eine bimodale Verteilung in der Marketingabteilung könnte beispielsweise auf zwei verschiedene Gehaltsgruppen hinweisen, während eine rechtsschiefe Verteilung in der Produktion auf wenige hochbezahlte Spezialisten neben vielen Standardarbeitsplätzen deutet. Diese Erkenntnisse sind entscheidend für die Budgetplanung und die Bewertung der Kostenstruktur.

Steuerberatung: Mandantenvergleich und Optimierungsberatung

Als **Steuerberater** analysieren Sie die Betriebsausgabenstrukturen verschiedener Mandanten einer Branche, um Benchmarks zu erstellen und Optimierungspotenziale zu identifizieren. Sie wählen "Mandant" als Hauptkategorie, "Ausgabenkategorie" als Unterkategorie und "Jahresausgaben" als Wert.

Ein Punkt-Plot bietet Ihnen eine klare Übersicht über die Ausgabenverteilung und sortiert automatisch nach den medianen Ausgabenhöhen. Sie können schnell erkennen, welche Ausgabenkategorien bei einzelnen Mandanten überdurchschnittlich hoch sind. Die Balkenbreite zeigt Ihnen zusätzlich, wie einheitlich oder unterschiedlich die Ausgaben in jeder Kategorie sind, was für die Beurteilung von Optimierungspotenzialen entscheidend ist.

16.3.2 *Qualitätssicherung und sicherer Umgang*

Bevor Sie mit der Analyse beginnen, sollten Sie die Qualität Ihrer Daten überprüfen. Achten Sie darauf, dass **numerische Spalten** tatsächlich numerische Daten enthalten und kategorische Spalten konsistente Bezeichnungen verwenden. Leere Zellen werden automatisch aus der Analyse ausgeschlossen, aber inkonsistente Kategorienamen (z.B. "IT-Abteilung" vs. "IT Abteilung") werden als separate Kategorien behandelt.

Interpretation von Ergebnissen

Bei der **Interpretation Ihrer Ergebnisse** sollten Sie immer den Kontext Ihrer Daten berücksichtigen. Ein Ausreißer in den Personalkosten könnte auf einen Fehler hinweisen, aber auch auf eine berechtigte Sonderzahlung. Die statistische Auffälligkeit allein ist nicht ausreichend - sie muss immer im fachlichen Kontext bewertet werden.

Dokumentation und Nachvollziehbarkeit

Für die Dokumentation Ihrer Analysen nutzen Sie die integrierten **Export-Funktionen**. Sie können sowohl statische Screenshots als auch interaktive Versionen Ihrer Diagramme exportieren, um diese in Berichte, Präsentationen oder Arbeitspapiere zu integrieren. Die interaktiven Versionen ermöglichen es Kollegen oder Mandanten, Ihre Analysen nachzuvollziehen und eigene Erkundungen durchzuführen.

Diese umfassenden Analysemöglichkeiten bieten Ihnen somit nicht nur ein Werkzeug zur Datenvisualisierung, sondern eine **methodische Grundlage für die systematische Erkenntnisgewinnung**, die Ihre professionelle Urteilsbildung in Wirtschaftsprüfung, Steuerberatung und Controlling wesentlich unterstützt und verbessert.

16.3.3 *Besonderheiten Box-Plot, Violin-Plot und Punkt-Plot*

Die Anwendung bietet drei verschiedene Visualisierungsformen, die jeweils unterschiedliche Stärken für verschiedene Analyseszenarien aufweisen:

Box-Plot für strukturierte Prüfungshandlungen

Der Box-Plot ist das bewährte Standardwerkzeug für die schnelle statistische Analyse in der Wirtschaftsprüfung. Diese Darstellungsform reduziert komplexe Datensätze auf ihre wesentlichen statistischen Kennzahlen und macht Ausreißer sofort sichtbar. Wenn Sie beispielsweise die monatlichen Umsätze verschiedener Filialen eines Einzelhandelsunternehmens prüfen, zeigt Ihnen der Box-Plot auf einen Blick, welche Standorte konsistent performen und welche ungewöhnliche Schwankungen aufweisen.

Die Box-Plot-Darstellung zeigt Ihnen folgende Elemente:

- **Zentrale Linie:** Median (50. Perzentil) - weniger anfällig für Extremwerte als der Durchschnitt
- **Box-Grenzen:** 1. Quartil (25%) und 3. Quartil (75%) - zeigen den Bereich der mittleren 50% der Daten
- **Whisker:** Minimum und Maximum (oder begrenzter Bereich bei "ohne Ausreißer")
- **Einzelne Punkte:** Ausreißer, die besondere Aufmerksamkeit verdienen

Violin-Plot für tiefgreifende Controlling-Analysen

Der Violin-Plot erweitert die Box-Plot-Funktionalität um die Darstellung der tatsächlichen Datenverteilung und ist daher das bevorzugte Werkzeug für tiefgreifende Controlling-Analysen. Während der Box-Plot nur die Quartile zeigt, visualisiert der Violin-Plot die gesamte Wahrscheinlichkeitsdichte der Daten. Dies ist besonders relevant, wenn Sie als Controller verstehen möchten, wie sich Kosten oder Erträge innerhalb bestimmter Bereiche verteilen.

Violin-Plots bieten zusätzliche Erkenntnisse durch:

- **Breite der Form:** Zeigt die Häufigkeitsdichte der Werte an verschiedenen Punkten
- **Symmetrie:** Gibt Hinweise auf die Schiefe der Verteilung
- **Mehrere Spitzen:** Können auf multimodale Verteilungen hinweisen
- **Kombinierte Box-Plot-Elemente:** Alle statistischen Kennzahlen bleiben sichtbar

Punkt-Plot für umfassende Vergleichsanalysen

Der Punkt-Plot ist das ideale Werkzeug für Steuerberater, die große Mengen von Kategorien übersichtlich vergleichen müssen. Diese Darstellungsform reduziert die visuelle Komplexität und konzentriert sich auf die wesentlichen statistischen Maßzahlen, während sie gleichzeitig eine intuitive Rangfolge der analysierten Kategorien erstellt.

Die Punkt-Plot-Darstellung charakterisiert sich durch:

- **Zentraler Punkt:** Median der jeweiligen Kategorie
- **Innerer Balken:** Interquartilsbereich (25%-75% der Daten)
- **Äußerer Balken:** Gesamtbereich ohne Extremausreißer
- **Automatische Sortierung:** Kategorien werden nach Median-Werten geordnet
- **Farbkodierung:** Visuelle Hierarchie durch abgestufte Farben

Optionen und Ausreißer-Behandlung

Die Behandlung von Ausreißern ist in der Finanzanalyse von besonderer Bedeutung, da diese oft die wichtigsten Erkenntnisse liefern. Das System bietet Ihnen verschiedene Optionen, die sich je nach gewähltem Diagrammtyp unterscheiden:

Für Box-Plot und Violin-Plot stehen zur Verfügung:

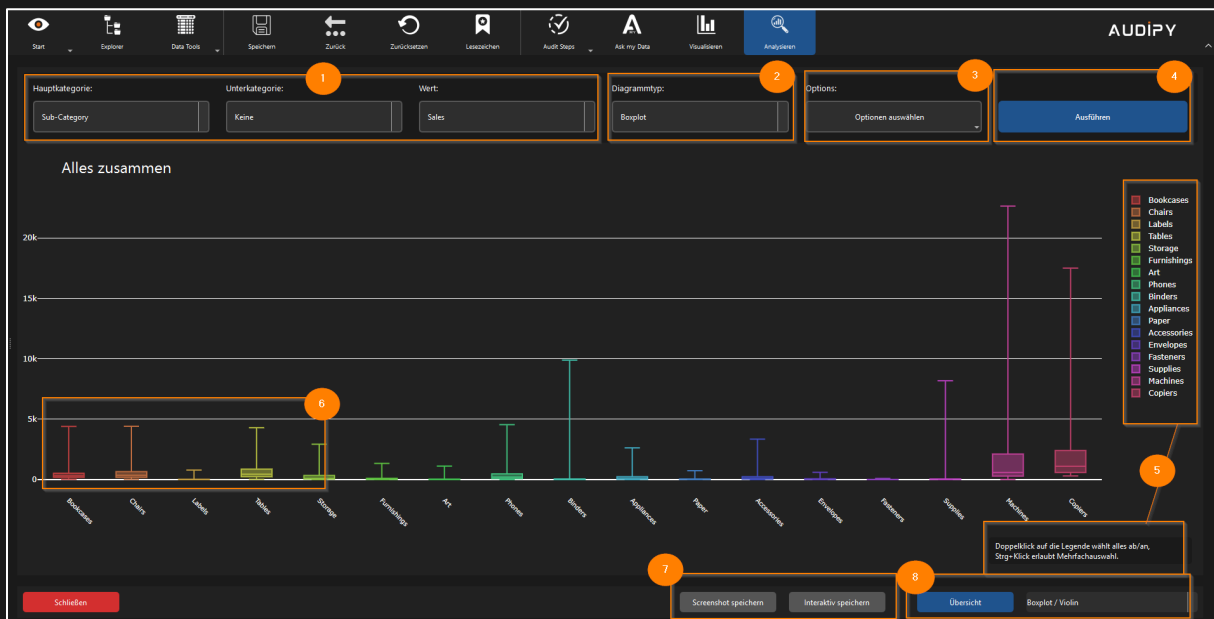
- **"Ausreißer separat anzeigen"**: Diese Option ist das Standardwerkzeug für die Risikoidentifikation in der Wirtschaftsprüfung. Ausreißer werden als einzelne Punkte außerhalb der normalen Verteilung dargestellt und sind oft Hinweise auf Fehler, Unregelmäßigkeiten oder besondere Geschäftsereignisse, die weitere Prüfungshandlungen erfordern.
- **"Ohne Ausreißer"**: Diese Alternative ist wertvoll, wenn Sie die normale Geschäftstätigkeit analysieren möchten, ohne dass außergewöhnliche Ereignisse die Betrachtung verzerren. Die Darstellung begrenzt sich auf den Bereich zwischen dem 1. und 3. Quartil plus dem 1,5-fachen Interquartilsabstand.

Für Punkt-Plot ist verfügbar:

- **"Ohne Ausreißer"**: Dies ist die einzige Option für Punkt-Plots, was der Philosophie dieser Darstellungsform entspricht, eine bereinigte Übersicht über die zentralen Tendenzen zu bieten.

16.3.4 Erstellung vom Boxplot / Violin Diagramm

Bevor Sie mit der eigentlichen Analyse beginnen, sollten Sie Ihre Fragestellung klar definieren. Überlegen Sie sich, welche Erkenntnisse Sie aus der Analyse gewinnen möchten und wie diese in Ihre berufliche Tätigkeit einfließen sollen. In der Wirtschaftsprüfung könnte dies die Identifikation von Prüfungsrisiken sein, im Controlling die Optimierung der Kostenstruktur und in der Steuerberatung die Entwicklung von Optimierungsstrategien.



1. Datenauswahl & strategische Planung ①

- Beginnen Sie mit der Auswahl der **Hauptkategorie**. Wählen Sie die Dimension aus, die für Ihre Fragestellung am relevantesten ist. Für eine erste Orientierung ist es oft sinnvoll, mit "Alle zusammen" zu beginnen, um einen Gesamtüberblick zu erhalten. Anschließend wählen Sie die **numerische Variable** (Wert) aus, die Sie analysieren möchten.
- Falls Sie eine hierarchische Analyse planen, konfigurieren Sie die **Unterategorie**. Lassen Sie dieses Feld zunächst auf "Keine" stehen, wenn Sie sich erst einen Überblick verschaffen möchten.

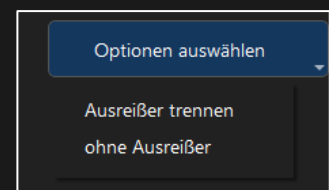
2. Diagrammtyp ②

Wählen Sie den Diagrammtyp basierend auf Ihrem Analyseziel:

- **Box-Plot**: Für schnelle Übersichten und Ausreißer-Identifikation
- **Violin-Plot**: Für detaillierte Verteilungsanalysen
- **Punkt-Plot**: Für Vergleiche zwischen vielen Kategorien

3. Optionen auswählen (Ausreißer) - optional ③

- Entscheiden Sie sich für die Behandlung von Ausreißern basierend auf Ihrer Fragestellung. Klicken Sie anschließend auf "**Ausführen**", um die Analyse zu starten.



4. Analyse starten ④

Sobald alle Einstellungen vorgenommen sind, kann man über den „**Analyse-starten**“ Button das Diagramm **laden** oder **aktualisieren**. Das System nutzt die oben getroffenen Auswahlfelder (Kategorie, Wert, Diagrammtyp und Optionen), um das entsprechende Boxplot bzw. Violin-Diagramm zu erzeugen.

- Das System generiert automatisch das gewählte Diagramm und zeigt oberhalb eine **interaktive Kategorieauswahl**, falls Sie Kategorien definiert haben. Nutzen Sie diese Auswahlmöglichkeit, um zwischen verschiedenen Hauptkategorien zu wechseln und die entsprechenden Unterkategorien zu analysieren (Voreingestellt ist alle zusammen / alle Textspalten).

Hinweis: Das System erkennt automatisch, wenn Ihre Datenanalyse aufgrund der Größe oder Komplexität optimiert werden muss. Bei mehr als 50 Kategorien in der Hauptkategorie erscheint eine Warnung mit folgenden Optionen:

- **Automatische Begrenzung akzeptieren:** Das System begrenzt die Analyse auf die häufigsten oder volumensstärksten Kategorien
- **Eigene Anzahl definieren:** Sie können eine spezifische Anzahl von Kategorien festlegen
- **Vollständige Analyse:** Sie können die Warnung ignorieren und alle Kategorien analysieren (nicht empfohlen bei sehr großen Datensätzen)

Diese Funktion ist besonders wertvoll in der Steuerberatung, wenn Sie beispielsweise hunderte von Ausgabenkategorien haben, aber nur die bedeutendsten für Ihre Analyse relevant sind.

The image shows two screenshots of a software interface. The left screenshot, titled 'Zu viele Kategorien' (Too many categories) with a yellow warning icon, displays a message: 'Es gibt 531 unterschiedliche Kategorien, was die Leistung beeinträchtigen kann. Es wird empfohlen, die Anzahl der Kategorien zu reduzieren.' Below this, it asks 'Möchten Sie die Analyse auf die wichtigsten Kategorien beschränken?' (Do you want to limit the analysis to the most important categories?). At the bottom are two buttons: 'Ja' (Yes) in green and 'Nein' (No) in red. The right screenshot, titled 'Anzahl der Kategorien begrenzen' (Limit number of categories) with a blue information icon, asks 'Geben Sie die Anzahl der anzuzeigenden Top-Kategorien ein:' (Enter the number of top categories to display:). Below this is a text input field containing the number '10'. At the bottom are two buttons: 'OK' in green and 'Abbrechen' (Cancel) in red.

Die **Unterkategorien-Prüfung** erfolgt intelligent erst dann, wenn Sie eine spezifische Hauptkategorie auswählen. Dies verhindert unnötige Warnungen bei der Erstanalyse und ermöglicht Ihnen eine differenzierte Betrachtung. Das System wendet dieselben Optimierungsregeln an wie bei den Hauptkategorien, jedoch kontextspezifisch für die gewählte Hauptkategorie.

Für sehr große Datensätze wendet die Anwendung automatisch intelligente Sampling-Verfahren an:

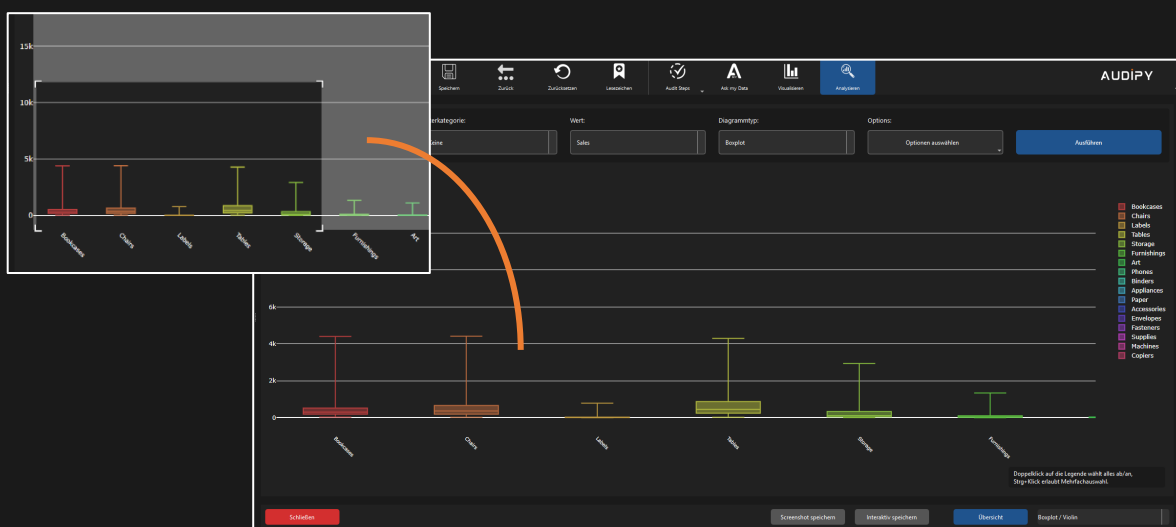
- **Outlier-Sampling:** Bei der Option "Ausreißer separat anzeigen" werden repräsentative Ausreißer ausgewählt
- **Statistische Optimierung:** Für Box-Plots und Punkt-Plots werden Statistiken aus allen Daten berechnet, aber nur optimierte Datenpunkte übertragen
- **Dichtefunktion-Sampling:** Bei Violin-Plots wird die Dichteverteilung effizient berechnet, ohne alle Datenpunkte einzeln zu verarbeiten

5. Legende interaktiv nutzen 5

- **Klick auf eine Legenden-Kategorie:** Blendet die entsprechende Datenreihe im Diagramm ein oder aus.
- Mit **Doppelklick** auf die Legende wählt man die Legende entweder insgesamt ab oder wieder an
- Mit **Strg + Klick** ist auch eine Mehrfachauswahl möglich (z.B. wenn man im Beispieldiagramm Produkte miteinander vergleichen möchte)

6. Interaktive Filterung durch Rechteckauswahl 6

- Ziehen Sie mit der Maus ein **Rechteck über einen Bereich im Diagramm**, um nur die darin enthaltenen Daten anzuzeigen und den Rest automatisch auszublenden. Diese Funktion erlaubt es Ihnen, spezifische Ausschnitte zu untersuchen, ohne die gesamte Ansicht zu verlieren.
- Möchte man die ursprüngliche Darstellung wieder haben, bedarf es einen Doppelklick in das Diagramm. So wird die Filterung wieder aufgehoben.

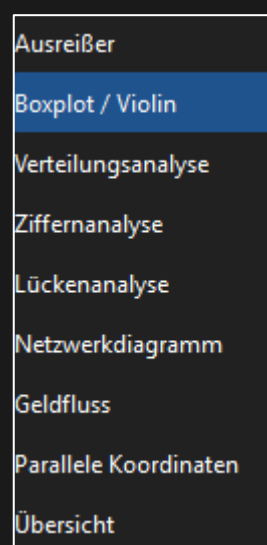


7. Screenshot speichern / Interaktiv speichern 7

- Mit **Screenshot speichern** erstellen Sie ein statisches Bild des Diagramms. Dieses können Sie entweder unter den Falldaten speichern oder können dieses auch an andere Personen zusenden. Hier empfiehlt sich allerdings das interaktive Speichern.
- **Interaktiv speichern** legt das Diagramm als eigenständige HTML-Datei ab, die weiterhin interaktive Funktionen (z.B. Zoom oder Filter) unterstützt. So können Sie es frei verschicken, ohne dass der Empfänger selbst den AUDIPY auf dem Rechner hat. Die HTML-Datei wird ganz normal über den Browser geöffnet.

8. Übersicht / Auswahl 8

- Mithilfe des blauen Buttons kann zur Analyse Übersicht mit allen möglichen Analysen gewechselt werden. Es kann aber auch das Dropdown Menü benutzt werden um zu einer anderen Analyse zu wechseln.



16.3.5 Erweiterte Analyseworkflows und Interpretationsstrategien

Die interaktive Natur der Kategorieauswahl ermöglicht explorative Datenanalysen, die über statische Berichte hinausgehen. Ein bewährter Workflow beginnt mit der Gesamtbetrachtung (Hauptkategorie "Alle zusammen"), um grundlegende Datenverteilungen und Anomalien zu identifizieren. Anschließend wechseln Sie zu spezifischen Kategorien und nutzen Unterkategorien für detaillierte Analysen.

Dieser Workflow ist besonders wertvoll in der Wirtschaftsprüfung, wo Sie oft von einer breiten Risikobeurteilung zu spezifischen Prüfungshandlungen übergehen müssen. Sie können beispielsweise zunächst alle Umsätze betrachten, auffällige Perioden identifizieren und dann systematisch in diese Perioden "hineinzoomen", um die Ursachen zu verstehen.

Kombinierte Diagrammtyp-Analyse

Die Kombination verschiedener Diagrammtypen für dieselben Daten kann zusätzliche Erkenntnisse liefern. Ein bewährtes Vorgehen ist:

1. **Punkt-Plot für den Überblick:** Verschafft Ihnen eine sortierte Gesamtsicht auf alle Kategorien
2. **Box-Plot für statistische Details:** Zeigt Ihnen Quartile, Median und Ausreißer für ausgewählte Kategorien
3. **Violin-Plot für Verteilungsanalysen:** Offenbart die tatsächliche Form der Datenverteilung

Jeder Diagrammtyp beleuchtet andere Aspekte Ihrer Daten, und zusammen ergeben sie ein vollständiges Bild. Sie können zwischen den Diagrammtypen wechseln, ohne Ihre Kategorieauswahl und Konfiguration zu verlieren.

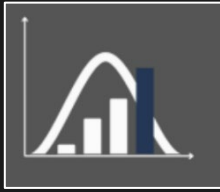
Ausreißer-Strategien für verschiedene Analyseziele

Die Möglichkeit, sowohl mit als auch ohne Ausreißer zu analysieren, ermöglicht es Ihnen, zwischen operativen und strategischen Perspektiven zu wechseln:

- **Operative Sicht ohne Ausreißer:** Zeigt Ihnen das normale Geschäftsverhalten und ist ideal für Budgetplanungen, Benchmarks und Standardprozesse
- **Strategische Sicht mit Ausreißern:** Hebt potenzielle Risiken und Chancen hervor und ist wertvoll für Risikoanalysen, Prüfungsplanungen und Optimierungsstrategien

Diese Dualität ermöglicht es Wirtschaftsprüfern, zwischen Routine-Prüfungshandlungen und Ausnahmeprüfungen zu differenzieren, Controllern zwischen operativer Planung und strategischen Szenarien zu unterscheiden, und Steuerberatern zwischen regulärer Beratung und Optimierungsstrategien zu wechseln.

16.4 Verteilungsanalyse



Die Verteilungsanalyse ist ein fundamentales statistisches Werkzeug zur Untersuchung der Struktur und Charakteristika numerischer Daten. Diese Analysemethode visualisiert, wie sich Werte in einem Datensatz verteilen, und ermöglicht es Fachkräften in Wirtschaftsprüfung, Steuerberatung und Controlling, tiefgreifende Erkenntnisse über die Natur ihrer Geschäftsdaten zu gewinnen. Durch die graphische Darstellung von Häufigkeitsverteilungen können Sie Muster, Anomalien und statistische Eigenschaften identifizieren, die in tabellarischen Übersichten oder einfachen Kennzahlen verborgen bleiben würden.

In der Wirtschaftsprüfung dient die Verteilungsanalyse zur Beurteilung der Plausibilität von Geschäftsdaten und zur Identifikation ungewöhnlicher Verteilungsmuster, die auf Fehler oder Manipulationen hindeuten könnten. Die Form einer Verteilung kann Aufschluss über die zugrundeliegenden Geschäftsprozesse geben: Normalverteilte Umsätze deuten auf stabile, natürliche Geschäftsverläufe hin, während stark schiefe oder multimodale Verteilungen auf strukturelle Besonderheiten oder potenzielle Probleme hinweisen können.

Controller nutzen Verteilungsanalysen zur Bewertung der Variabilität von Kosten, zur Identifikation von Ausreißern in Budgetabweichungen und zur statistischen Modellierung für Prognosen und Planungen. Die Analyse der Verteilungsform hilft bei der Auswahl angemessener statistischer Methoden und bei der Bewertung der Zuverlässigkeit von Durchschnittswerten und anderen Kennzahlen. Steuerberater verwenden diese Methoden zur Analyse von Ausgabenmustern, zur Plausibilitätsprüfung von Betriebsausgaben und zur Vorbereitung fundierter Argumentationen in Betriebsprüfungsverfahren.

Die mathematische Robustheit der Verteilungsanalyse macht sie zu einem objektiven Beurteilungsinstrument, das sowohl explorative Datenanalysen als auch konfirmatorische statistische Tests unterstützt. Moderne Implementierungen ermöglichen dabei nicht nur die klassische Histogramm-Darstellung, sondern auch erweiterte Visualisierungen wie Dichtefunktionen, Normalverteilungsvergleiche und kategoriespezifische Überlagerungen.

16.4.1 Wofür nutzt man eine Verteilungsanalyse?

Die Verteilungsanalyse kommt in der Praxis in verschiedenen Bereichen vor:

Wirtschaftsprüfung: Umsatzplausibilität und Risikobeurteilung

In der Jahresabschlussprüfung verwenden Sie die Verteilungsanalyse zur Beurteilung der Plausibilität von Umsatzdaten. Sie konfigurieren die Analyse mit "Produktgruppe"

als Kategorie und "Monatsumsatz" als Wert. Die normale Verteilungsdarstellung zeigt Ihnen zunächst die Gesamtverteilung aller Umsätze.

Eine stark rechtsschiefe Verteilung mit wenigen extrem hohen Werten könnte auf Sondergeschäfte hindeuten, die weitere Prüfungshandlungen erfordern. Durch Aktivierung der "Alle Kategorien als Dichtelinien"-Option können Sie die Umsatzverteilungen verschiedener Produktgruppen simultan vergleichen. Anomale Verteilungsformen in einzelnen Produktgruppen können Hinweise auf Unregelmäßigkeiten oder besondere Geschäftsereignisse liefern.

Die "Normalverteilung anzeigen"-Option hilft Ihnen bei der Beurteilung, ob die Daten den Annahmen für weitere statistische Tests entsprechen. Signifikante Abweichungen von der Normalverteilung können die Anwendung bestimmter Analyseverfahren einschränken und alternative Prüfungsansätze erforderlich machen.

Controlling: Kostenanalyse und Variabilitätsbewertung

Im Controlling einer Produktionsgesellschaft analysieren Sie die Verteilung der Materialkosten verschiedener Fertigungslinien. Sie wählen "Fertigungslinie" als Kategorie und "Materialkosten pro Einheit" als Wert. Die Verteilungsanalyse mit aktivierter Statistik-Box zeigt Ihnen nicht nur die Form der Verteilung, sondern auch quantitative Kennzahlen wie Standardabweichung und Variationskoeffizient.

Unterschiedliche Verteilungsbreiten zwischen Fertigungslinien können auf verschiedene Effizienzgrade oder Kontrolldichte hindeuten. Eine Fertigungslinie mit sehr breiter, flacher Kostenverteilung könnte Prozessprobleme oder unzureichende Standardisierung aufweisen. Die logarithmische Transformation kann dabei helfen, relative Kostenunterschiede besser sichtbar zu machen und Ausreißer zu identifizieren.

Durch die Binning-Optimierung mit der "Freedman-Diaconis"-Methode können Sie auch bei Vorhandensein von Kostenausreißern eine aussagekräftige Darstellung der normalen Kostenvariabilität erhalten. Dies unterstützt Sie bei der Budgetplanung und der Bewertung der Kostenkontrolle.

Steuerberatung: Ausgabenanalyse und Betriebsprüfungsvorbereitung

Als Steuerberater nutzen Sie die Verteilungsanalyse zur Analyse der Betriebsausgaben Ihres Mandanten in Vorbereitung auf eine Betriebsprüfung. Sie wählen "Ausgabenart" als Kategorie und "Ausgabenbetrag" als Wert. Die Analyse beginnt mit der normalen Verteilungsdarstellung aller Ausgaben.

Eine ungewöhnliche Verteilungsform, wie beispielsweise eine verdächtig gleichmäßige Verteilung bei Bewirtungskosten, könnte auf Schätzungen oder Pauschalierungen hindeuten, die steuerlich problematisch sein könnten. Die "Box-Plot erstellen"-Option

ergänzt die Histogramm-Darstellung um eine kompakte Übersicht über Median, Quartile und Ausreißer, die Sie bei der Diskussion mit der Finanzverwaltung unterstützt.

Die kategorienweise Analyse mit Dichtelinien-Overlay ermöglicht Ihnen den direkten Vergleich verschiedener Ausgabenarten. Auffällige Unterschiede in den Verteilungsmustern können Diskussionspunkte mit dem Mandanten aufzeigen und helfen bei der Vorbereitung fundierter Argumentationen für die Betriebsprüfung.

16.4.2 *Werteauswahl und numerische Variablen*

Das **Wert-Feld** bestimmt die numerische Variable, deren Verteilung analysiert wird. Die Auswahl der richtigen Variable ist entscheidend für aussagekräftige Ergebnisse, da verschiedene Geschäftskennzahlen unterschiedliche statistische Eigenschaften und Interpretationsansätze erfordern.

Das System bietet alle numerischen Spalten Ihres Datensatzes zur Auswahl an. Typische Anwendungen umfassen:

- **Finanzielle Kennzahlen:** Umsätze, Kosten, Gewinne, Deckungsbeiträge
- **Transaktionsdaten:** Rechnungsbeträge, Zahlungsbeträge, Bestellvolumen
- **Operative Kennzahlen:** Produktionsmengen, Bearbeitungszeiten, Qualitätsmessungen
- **Verhältniskennzahlen:** Margen, Renditen, Wachstumsraten

Die Wahl der zu analysierenden Variable sollte immer im Kontext Ihrer spezifischen Fragestellung erfolgen. Absolute Beträge wie Umsätze eignen sich für die Analyse von Geschäftsvolumen und Größeneffekten, während relative Kennzahlen wie Margen oder Wachstumsraten für Effizienz- und Performanceanalysen geeigneter sind.

Verteilungstypen und Transformationen

Das **Verteilungstyp-Feld** bietet Ihnen drei verschiedene Darstellungsformen, die jeweils unterschiedliche Aspekte Ihrer Daten beleuchten:

Normal (Linear)

Die Standarddarstellung ohne mathematische Transformation zeigt die ursprünglichen Datenwerte in ihrer natürlichen Skalierung. Dies ist die am häufigsten verwendete Option und eignet sich für die meisten Geschäftsdaten, die bereits näherungsweise normalverteilt sind oder bei denen die ursprüngliche Skalierung interpretativ wichtig ist.

Logarithmisch (ln)

Die natürliche Logarithmus-Transformation ist besonders wertvoll für rechtsschief verteilte Daten, wie sie bei Umsätzen, Kosten oder anderen finanziellen Kennzahlen häufig auftreten. Diese Transformation:

- Komprimiert große Werte stärker als kleine Werte
- Kann schiefe Verteilungen normalisieren
- Ist ideal für Daten mit exponentiellen Wachstumsmustern
- Ermöglicht die Analyse relativer Veränderungen

Logarithmisch (log10)

Die Zehner-Logarithmus-Transformation bietet ähnliche Vorteile wie der natürliche Logarithmus, ist aber intuitiver interpretierbar, da sie Größenordnungen (Zehnerpotenzen) direkt sichtbar macht. Diese Option ist besonders nützlich für Daten, die über mehrere Größenordnungen streuen, wie Kundengrößen, Projektvolumen oder Unternehmenskennzahlen.

Binning-Methoden und statistische Optimierung

Das **Methode-Feld** ermöglicht Ihnen die Auswahl verschiedener statistischer Verfahren zur optimalen Einteilung der Daten in Histogramm-Balken (Bins). Diese scheinbar technische Entscheidung hat erheblichen Einfluss auf die Interpretierbarkeit Ihrer Ergebnisse:

Auto

Die automatische Methode kombiniert verschiedene statistische Ansätze und wählt die jeweils optimale Bin-Anzahl basierend auf der Datengröße und -verteilung. Dies ist die empfohlene Standardoption für die meisten Anwendungen.

Sturges

Die Sturges-Regel ($k = 1 + 3.322 * \log_{10}(n)$) ist konservativ und eignet sich besonders für normalverteilte Daten. Sie tendiert zu weniger Bins und glättet damit die Darstellung, was bei verrauschten Daten vorteilhaft sein kann.

Freedman-Diaconis

Diese Methode berücksichtigt die Interquartilsspanne und ist robust gegenüber Ausreißern. Sie eignet sich besonders für schiefe Verteilungen und Daten mit Extremwerten, wie sie in Finanzdaten häufig auftreten.

Scott

Die Scott-Regel basiert auf der Standardabweichung und eignet sich gut für kontinuierliche, normalverteilte Daten. Sie ist sensitiver für die Datenstreuung als die Sturges-Regel.

Rice

Die Rice-Regel ($k = 2 * n^{(1/3)}$) ist weniger konservativ als Sturges und zeigt mehr Details in der Verteilungsstruktur. Sie eignet sich für größere Datensätze und detaillierte Verteilungsanalysen.

Erweiterte Optionen und Visualisierungsverbesserungen

Das **Optionen-Feld** bietet vier mächtige Erweiterungen, die Ihre Verteilungsanalyse erheblich bereichern können:

Box-Plot erstellen

Diese Option fügt einen Box-Plot oberhalb oder seitlich des Histogramms hinzu und kombiniert damit zwei komplementäre statistische Visualisierungen. Der Box-Plot zeigt Median, Quartile und Ausreißer und ergänzt die Detailinformationen des Histogramms um kompakte statistische Kennzahlen. Diese Kombination ist besonders wertvoll für Präsentationen und schnelle statistische Beurteilungen.

Normalverteilung anzeigen

Die Überlagerung einer theoretischen Normalverteilungskurve ermöglicht den direkten visuellen Vergleich zwischen Ihren tatsächlichen Daten und einer idealen Normalverteilung. Dies ist essentiell für die Beurteilung der Anwendbarkeit parametrischer statistischer Tests und für die Identifikation von Abweichungen von der Normalverteilungsannahme.

Statistik-Box anzeigen

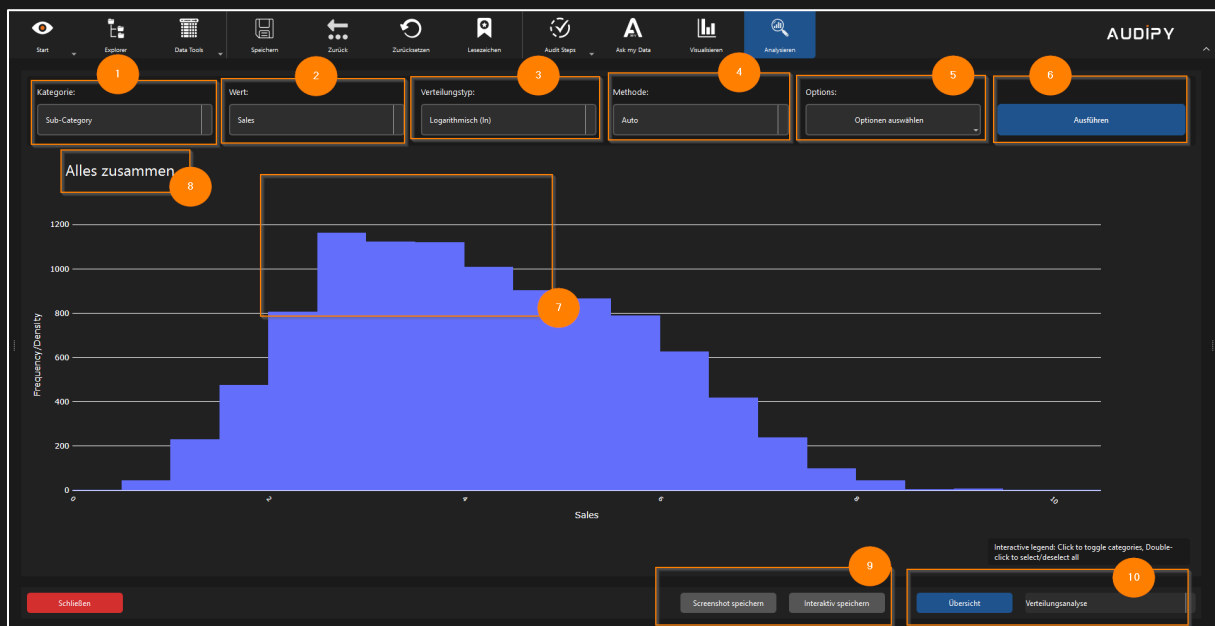
Diese Option fügt eine kompakte Statistik-Übersicht direkt in das Diagramm ein, die wichtige Kennzahlen wie Mittelwert, Standardabweichung, Median, Schiefe und Kurtosis anzeigt. Dies ermöglicht eine quantitative Beurteilung der Verteilungseigenschaften ohne separate Berechnungen.

Alle Kategorien als Dichtelinien anzeigen

Diese erweiterte Visualisierungsoption erstellt ein Overlay-Diagramm mit Dichtefunktionen für alle Kategorien gleichzeitig. Anstatt zwischen einzelnen Kategorien zu wechseln, können Sie alle Verteilungen simultan vergleichen. Dies ist besonders wertvoll für Benchmarking-Analysen und die Identifikation von Ausreißer-Kategorien.

16.4.3 Erstellung einer Verteilungsanalyse

Bevor Sie mit der Verteilungsanalyse beginnen, sollten Sie die Qualität und Eignung Ihrer Daten überprüfen. Stellen Sie sicher, dass die numerische Variable tatsächlich kontinuierliche oder diskrete Werte enthält, die für eine Verteilungsanalyse geeignet sind. Extreme Ausreißer können die Visualisierung dominieren und sollten gegebenenfalls vorab identifiziert und bewertet werden.



1. Kategorie ①

- Wählen Sie hier die **Hauptkategorie**, nach der die Daten untersucht werden sollen (z. B. **Produktgruppe, Abteilung, Region**) oder beginnen Sie mit „Alle zusammen“.

2. Wert ②

- Ihre wichtigste numerische Variable als Wert, "Normal" als Verteilungstyp und "Auto" als Methode. Diese Grundkonfiguration gibt Ihnen einen ersten Überblick über die Datenverteilung und hilft bei der Entscheidung über weitere Analyseschritte.

3. Verteilungstyp ③

- Basierend auf der ersten Visualisierung entscheiden Sie über mögliche Transformationen:
- Bei rechtsschiefen Verteilungen (lange Ausläufer nach rechts) testen Sie die logarithmischen Optionen
- Bei bereits symmetrischen Verteilungen bleiben Sie bei der normalen Darstellung

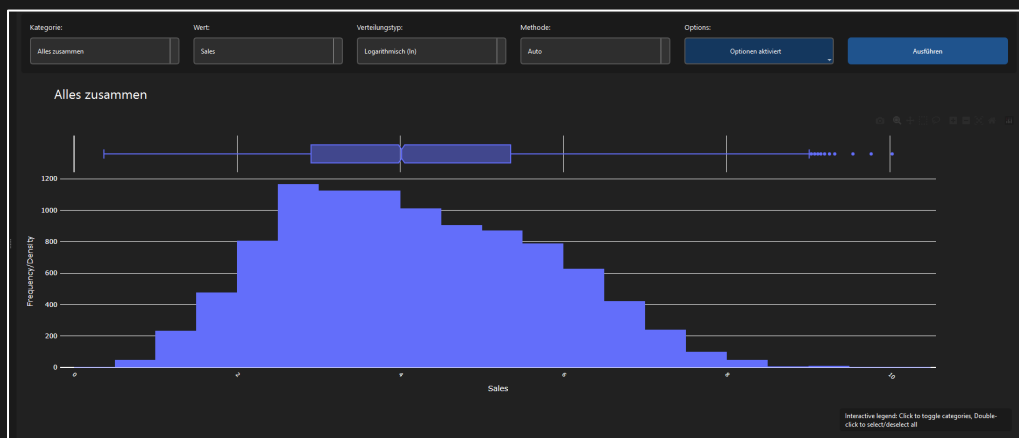
- Bei Daten über mehrere Größenordnungen wählen Sie "Logarithmisch (log10)" für bessere Interpretierbarkeit

4. Methode 4

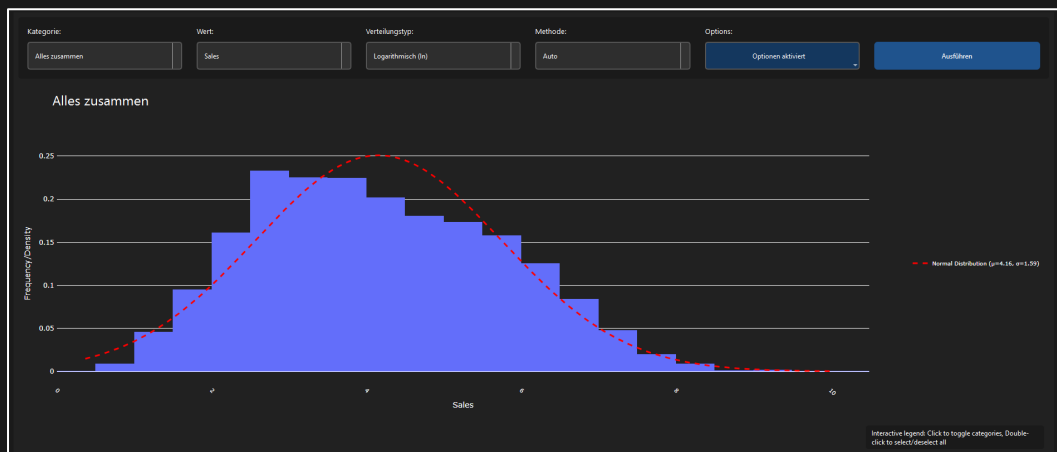
- Experimentieren Sie mit verschiedenen Binning-Methoden, um die optimale Darstellung zu finden:
 - **"Freedman-Diaconis"** für Daten mit Ausreißern
 - **"Sturges"** für geglättete Darstellungen
 - **"Rice"** für detaillierte Strukturanalysen

5. Optionen auswählen (optional) 5

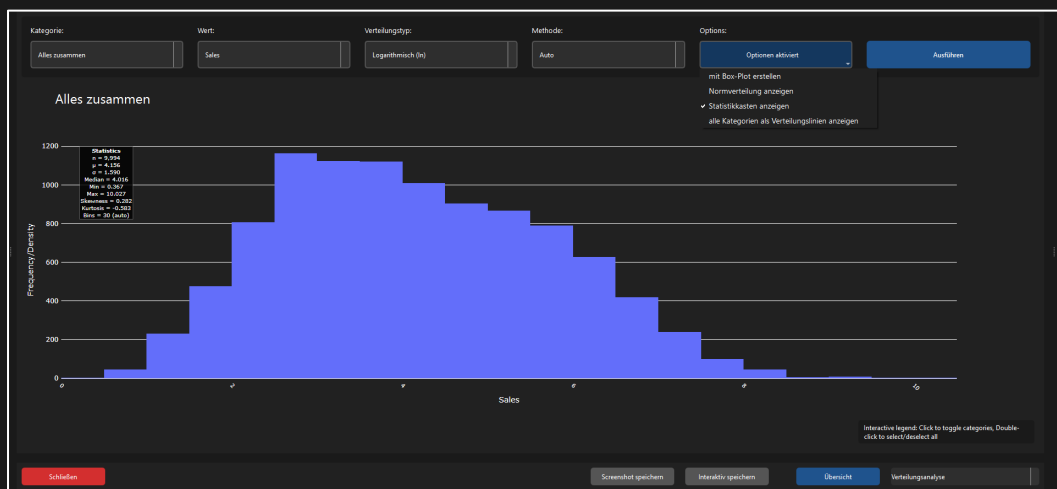
- Aktivieren Sie schrittweise die erweiterten Optionen:
 - Mit **Box-Plot**: für eine bessere Verteilungsübersicht
 - **"Statistik-Box anzeigen"** für quantitative Beurteilungen
 - **"Normalverteilung anzeigen"** für statistische Testvalidierung
 - **"Alle Kategorien als Dichtelinien"** für Vergleichsanalysen zwischen Gruppen



Mit Box-Plot



Mit Normfallverteilung



Mit Statistikkasten



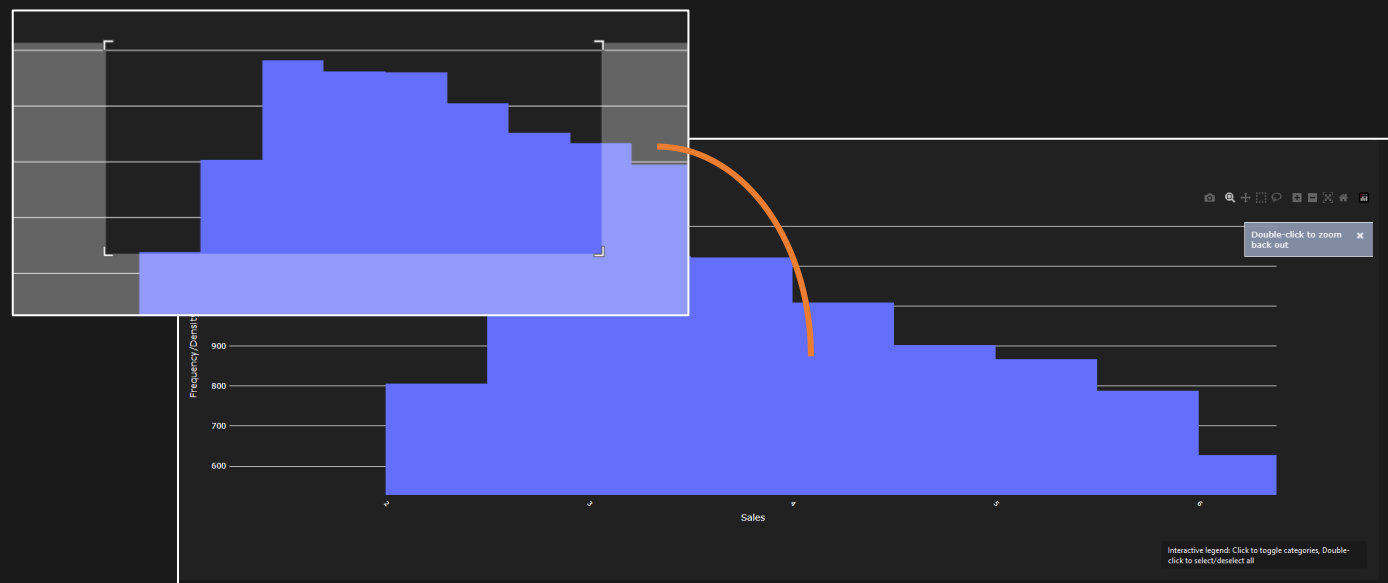
Kategorien als Verteilungslinie anzeigen

6. Verteilungsanalyse erstellen ⑥

- Sobald alle Einstellungen vorgenommen sind, kann man über den „**Verteilungsanalyse ausführen**“ **Button** das Diagramm **laden** oder **aktualisieren**. Das System nutzt die getroffenen Auswahlfelder um die Verteilung graphisch darzustellen.

7. Interaktive Filterung durch Rechteckauswahl ⑦

- Ziehen Sie mit der Maus ein **Rechteck über einen Bereich im Diagramm**, um nur die darin enthaltenen Daten anzuzeigen und den Rest automatisch auszublenden. Diese Funktion erlaubt es Ihnen, spezifische Ausschnitte zu untersuchen, ohne die gesamte Ansicht zu verlieren.
- Möchte man die ursprüngliche Darstellung wieder haben, bedarf es einen
- Doppelklick in das Diagramm. So wird die Filterung wieder aufgehoben.



8. Kategorie-Auswahl ⑧

- Über dieses Dropdown kann man **alle Kategorien** gemeinsam anzeigen oder gezielt **einzelne Kategorien** auswählen.
- Werden z.B. Produkte analysiert kann man mit Mit Klick auf „Alles zusammen“ folgende Auswahlmöglichkeiten erhalten:

Alles zusammen
Bookcases
Machines
Binders
Tables
Appliances
Chairs
Supplies
Storage
Copiers

9. Screenshot speichern / Interaktiv speichern 9

- Mit **Screenshot speichern** erstellen Sie ein statisches Bild des Diagramms. Dieses können Sie entweder unter den Falldaten speichern oder können dieses auch an andere Personen zusenden. Hier empfiehlt sich allerdings das interaktive Speichern.
- **Interaktiv speichern** legt das Diagramm als eigenständige HTML-Datei ab, die weiterhin interaktive Funktionen (z.B. Zoom oder Filter) unterstützt. So können Sie es frei verschicken, ohne dass der Empfänger selbst den AUDIPY auf dem Rechner hat. Die HTML-Datei wird ganz normal über den Browser geöffnet.

10. Übersicht / Auswahl 10

- Mithilfe des blauen Buttons kann zur Analyse Übersicht mit allen möglichen Analysen gewechselt werden. Es kann aber auch das Dropdown Menü benutzt werden um zu einer anderen Analyse zu wechseln.

16.4.4 Ergebnisinterpretation und statistische Beurteilung

Die Form einer Verteilung liefert wichtige Erkenntnisse über die zugrundeliegenden Geschäftsprozesse und Datenqualität:

Normalverteilung (Glockenform): Eine symmetrische, glockenförmige Verteilung deutet auf stabile, natürliche Prozesse hin. Dies ist das Idealbild für viele statistische Analysen und deutet auf gut kontrollierte Geschäftsprozesse hin. In der Wirtschaftsprüfung ist dies ein positives Zeichen für die Plausibilität der Daten.

Rechtschiefe Verteilung: Lange Ausläufer nach rechts (wenige sehr große Werte) sind typisch für Umsätze, Kosten oder Kundengrößen. Dies ist normal für viele Geschäftskennzahlen, kann aber bei extremer Ausprägung auf Konzentrationsrisiken hindeuten.

Linkschiefe Verteilung: Ausläufer nach links sind seltener und können auf Deckeleffekte oder Mindestgrenzen hindeuten. In der Praxis könnte dies beispielsweise bei Rabattsätzen oder Effizienzkennzahlen auftreten.

Multimodale Verteilungen: Mehrere Gipfel deuten auf verschiedene Populationen oder Prozesse innerhalb der Daten hin. Dies kann auf strukturelle Unterschiede hinweisen, die separate Analysen erfordern.

Gleichverteilung (flach): Eine flache Verteilung ohne klare Spitzen kann auf zufällige Daten oder unnatürliche Manipulationen hindeuten und erfordert weitere Untersuchungen.

Statistische Kennzahlen und ihre Interpretation

Die integrierte Statistik-Box liefert quantitative Maße für die Verteilungsbeurteilung:

Schiefe (Skewness):

- Werte nahe 0: symmetrische Verteilung
- Positive Werte: rechtsschiefe Verteilung
- Negative Werte: linksschiefe Verteilung
- Absolute Werte > 1 : stark asymmetrische Verteilung

Kurtosis:

- Werte nahe 0: normale Verteilungsspitze
- Positive Werte: spitzere Verteilung als Normalverteilung
- Negative Werte: flachere Verteilung als Normalverteilung

Kategorienvergleiche und Benchmarking

Die Overlay-Funktion mit Dichtelinien ermöglicht sophisticated Vergleichsanalysen zwischen verschiedenen Kategorien. Dabei sollten Sie auf folgende Aspekte achten:

Positionsunterschiede: Unterschiedliche Mittelwerte oder Mediane zwischen Kategorien deuten auf systematische Unterschiede hin, die betriebswirtschaftlich interpretiert werden sollten.

Streuungsunterschiede: Verschiedene Verteilungsbreiten können auf unterschiedliche Risikoprofile oder Kontrollqualität hindeuten.

Formunterschiede: Unterschiedliche Verteilungsformen zwischen Kategorien können auf verschiedene zugrundeliegende Prozesse oder Datenqualitätsprobleme hinweisen.

16.4.5 Erweiterte Analysestrategien und statistische Vertiefung

Nutzen Sie die **Kombinationsmöglichkeiten** der verschiedenen Optionen für umfassende statistische Beurteilungen. Beginnen Sie mit einer Gesamtverteilungsanalyse einschließlich Normalverteilungsvergleich und Statistik-Box. Diese gibt Ihnen ein quantitatives Fundament für weitere Analysen.

Anschließend wechseln Sie zur kategorienspezifischen Analyse mit Dichtelinien-Overlay, um Unterschiede zwischen Gruppen zu identifizieren. Diese mehrstufige Herangehensweise ermöglicht es Ihnen, sowohl die Gesamtstruktur als auch spezifische Teilmuster in Ihren Daten zu verstehen.

Transformationsanalyse für verschiedene Datentypen

Experimentieren Sie **systematisch mit den verschiedenen Verteilungstypen**, um die optimale Darstellung für Ihre spezifischen Daten zu finden. Finanzielle Zeitreihendaten profitieren häufig von logarithmischen Transformationen, während Verhältniskennzahlen oft in ihrer natürlichen Form am besten analysiert werden.

Die Wahl der **richtigen Transformation** kann entscheidend sein für die Identifikation von Mustern und Anomalien. Eine rechtschiefe Umsatzverteilung kann nach logarithmischer Transformation eine Normalverteilung zeigen, was die Anwendung parametrischer statistischer Tests ermöglicht und die Interpretierbarkeit erhöht.

Longitudinale Verteilungsanalyse

Verwenden Sie **Zeiträume als Kategorien** für die Analyse von Verteilungsveränderungen über die Zeit. Dies ermöglicht es Ihnen, Trends in der Datenqualität, saisonale Effekte oder strukturelle Veränderungen zu identifizieren. Eine sich verändernde Verteilungsform über die Zeit kann auf evolvierende Geschäftsprozesse oder veränderte externe Bedingungen hindeuten.

Diese **zeitliche Perspektive** ist besonders wertvoll für Controller, die die Stabilität ihrer Geschäftsprozesse überwachen möchten, und für Wirtschaftsprüfer, die langfristige Trends in der Datenqualität bewerten müssen.

16.4.6 Best Practices und methodische Empfehlungen

Beachten Sie, dass die **Aussagekraft von Verteilungsanalysen stark von der Stichprobengröße** abhängt. Für zuverlässige Verteilungsbeurteilungen sollten mindestens 100-200 Beobachtungen vorliegen, idealerweise deutlich mehr. Bei kleineren Stichproben können zufällige Schwankungen die Verteilungsform dominieren und zu Fehlinterpretationen führen.

Ausreißer-Management und Robustheit

Identifizieren Sie **systematisch Ausreißer** und bewerten Sie deren Einfluss auf die Verteilungsanalyse. Extreme Werte können die Skalierung dominieren und wichtige Verteilungsdetails überdecken. Die logarithmischen Transformationen können dabei helfen, den Einfluss von Ausreißern zu reduzieren, ohne diese vollständig zu eliminieren.

Kontextuelle Interpretation und fachliche Validierung

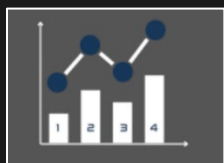
Interpretieren Sie **statistische Auffälligkeiten** immer im Kontext Ihrer fachlichen Expertise. Eine ungewöhnliche Verteilungsform kann statistische Gründe haben, aber auch auf reale Geschäftsbesonderheiten hindeuten. Die Kombination statistischer Analyse mit betriebswirtschaftlichem Know-how ist entscheidend für aussagekräftige Ergebnisse.

Dokumentation und Reproduzierbarkeit

Dokumentieren Sie Ihre **Analyseparameter** systematisch, einschließlich der gewählten Binning-Methoden, Transformationen und Optionen. Dies ist besonders wichtig für die Nachvollziehbarkeit in Prüfungsverfahren oder bei regulatorischen Nachfragen. Die Export-Funktionen sowie Lesezeichen der Software unterstützen Sie bei der Integration der Ergebnisse in Berichte und Präsentationen.

Die **Verteilungsanalyse** stellt somit ein vielseitiges, statistisch fundiertes Werkzeug dar, das durch seine Flexibilität und Tiefe optimal an die komplexen Anforderungen moderner Datenanalyse in Wirtschaftsprüfung, Controlling und Steuerberatung angepasst ist und gleichzeitig sowohl explorative als auch konfirmatorische Analyseansätze unterstützt.

16.5 Ziffernanalyse



Die **Benford-Analyse**, auch bekannt als **Ziffernanalyse** oder **Benfordsches Gesetz**, ist eines der mächtigsten statistischen Werkzeuge in der modernen Wirtschaftsprüfung und Betrugserkennung. Dieses mathematische Phänomen besagt, dass in vielen natürlich vorkommenden Datensätzen die Ziffer "1" als erste Ziffer etwa 30,1% aller Zahlen auftritt, während "9" nur in etwa 4,6% der Fälle die erste Ziffer bildet. Diese scheinbar paradoxe Verteilung folgt einem logarithmischen Muster und ist für Wirtschaftsprüfer von unschätzbarem Wert, da manipulierte oder konstruierte Daten typischerweise nicht diesem natürlichen Muster folgen.

In der Wirtschaftsprüfung hat sich die Benford-Analyse als hocheffektives Instrument zur Aufdeckung von Unregelmäßigkeiten, Betrug und Datenmanipulationen etabliert. Buchhalter, die Zahlen erfinden oder manipulieren, tendieren dazu, intuitiv gleichmäßigere Ziffernverteilungen zu erstellen, was zu signifikanten Abweichungen vom Benfordschen Gesetz führt. Steuerberater nutzen diese Technik zur Überprüfung der Plausibilität von Ausgabenbelegen und Einkommensangaben ihrer Mandanten. Controller verwenden die Benford-Analyse zur Validierung von Kostendaten und zur Identifikation anomaler Budgetbuchungen, die weitere Untersuchungen erfordern könnten.

Die wissenschaftliche Grundlage der Benford-Analyse macht sie besonders wertvoll für die forensische Buchführung. Anders als bei subjektiven Bewertungen oder Stichprobenprüfungen liefert die Benford-Analyse objektive, mathematisch fundierte Hinweise auf potenzielle Unregelmäßigkeiten. Dies macht sie zu einem unverzichtbaren Werkzeug in der risikoorientierten Prüfung, wo die systematische Identifikation von Prüfungsrisiken im Mittelpunkt steht.

16.5.1 Wofür nutzt man eine Ziffernanalyse?

Die Ziffernanalyse ist ein vielseitiges Instrument und wird in der Praxis in unterschiedlichen Bereichen eingesetzt:

Wirtschaftsprüfung: Jahresabschlussprüfung und Risikobeurteilung

In der Jahresabschlussprüfung setzen Sie die Benford-Analyse typischerweise zur Überprüfung von Umsatzerlösen, Ausgabenbelegen und Forderungen ein. Ein praktisches Szenario wäre die Analyse der Ausgangsrechnungen eines Mandanten: Sie wählen "Alle zusammen" als Kategorie, "Rechnungsbetrag" als Wert und beginnen mit der "Erste Ziffer"-Analyse.

Zeigt die Analyse erhebliche Abweichungen vom Benfordschen Gesetz, können Sie anschließend die Kategorie auf "Vertriebsbereich" oder "Monat" umstellen, um zu

identifizieren, ob die Auffälligkeiten systematisch in bestimmten Bereichen oder Zeiträumen auftreten. Eine konzentrierte Abweichung in einem bestimmten Vertriebsbereich könnte auf lokale Unregelmäßigkeiten hindeuten, während zeitlich begrenzte Anomalien möglicherweise auf periodenspezifische Buchungsmanipulationen schließen lassen.

Steuerberatung: Betriebsprüfungsvorbereitung und Plausibilitätskontrolle

Als Steuerberater nutzen Sie die Benford-Analyse zur Vorbereitung auf Betriebsprüfungen und zur internen Qualitätskontrolle. Sie können beispielsweise die Ausgabenbelege Ihres Mandanten analysieren, indem Sie "Ausgabenkategorie" als Kategorie und "Belegbetrag" als Wert wählen. Die "Erste Ziffer"-Analyse hilft Ihnen, problematische Ausgabenbereiche zu identifizieren, bevor diese von der Finanzverwaltung entdeckt werden.

Die "Endziffer-Untersuchung" ist besonders wertvoll zur Aufdeckung von Rundungsmanipulationen bei Belegen. Zeigen sich hier unnatürliche Häufungen bestimmter Endziffern, könnte dies auf systematische Belegmanipulationen oder psychologische Rundungstendenzen hinweisen, die steuerlich problematisch sein könnten.

Controlling: Budgetüberwachung und Prozessvalidierung

Im Controlling verwenden Sie die Benford-Analyse zur Validierung von Kostendaten und zur Überwachung von Budgetbuchungen. Ein typisches Anwendungsszenario ist die Analyse von Reisekosten verschiedener Abteilungen: Sie wählen "Abteilung" als Kategorie, "Reisekostenbetrag" als Wert und führen eine "Erste Ziffer"-Analyse durch.

Abweichungen in bestimmten Abteilungen können auf unzureichende Kontrollen, systematische Abrechnungsfehler oder sogar betrügerische Aktivitäten hinweisen. Die anschließende "Zweite Ziffer"-Analyse derselben Daten kann subtilere Unregelmäßigkeiten aufdecken, die bei der ersten Analyse unentdeckt blieben. Dies ermöglicht Ihnen eine risikobasierte Überwachung Ihrer internen Kontrollsysteme.

16.5.2 Bedienelemente und Analysekonfiguration

Das **Kategorie-Feld** ermöglicht Ihnen die strategische Gruppierung Ihrer Benford-Analyse nach verschiedenen Geschäftsdimensionen. Diese Funktionalität ist besonders wertvoll, wenn Sie vermuten, dass Unregelmäßigkeiten nur in bestimmten Bereichen auftreten oder wenn Sie die Datenqualität verschiedener Abteilungen oder Prozesse vergleichen möchten.

Verfügbare Optionen umfassen:

- **"Alle zusammen"** - Führt die Benford-Analyse über den gesamten Datensatz durch, ideal für eine erste Gesamtbeurteilung
- **Alle Textspalten** Ihres Datensatzes - Ermöglicht die Gruppierung nach Kostenstellen, Abteilungen, Lieferanten, Produktkategorien oder anderen kategorischen Variablen

In der praktischen Anwendung beginnen Wirtschaftsprüfer oft mit der Gesamtanalyse ("Alle zusammen"), um einen ersten Eindruck von der Datenqualität zu gewinnen. Zeigen sich hier Auffälligkeiten, kann anschließend durch die Kategorieauswahl systematisch eingegrenzt werden, in welchen Bereichen die Abweichungen auftreten. Ein Controller könnte beispielsweise zunächst alle Reisekosten analysieren und bei Auffälligkeiten dann nach Abteilungen aufschlüsseln, um problematische Bereiche zu identifizieren.

Werteauswahl und Datenqualifikation

Das **Wert-Feld** bestimmt die numerische Variable, die der Benford-Analyse unterzogen wird. Die Auswahl der richtigen Variable ist entscheidend für aussagekräftige Ergebnisse, da das Benfordsche Gesetz nur bei bestimmten Arten von Daten zuverlässig funktioniert.

Das System bietet alle numerischen Spalten Ihres Datensatzes zur Auswahl an. Für optimale Ergebnisse sollten Sie Daten wählen, die:

- Über mehrere Größenordnungen streuen (von zweistelligen bis zu mehrstelligen Beträgen)
- Natürlichen Geschäftsprozessen entstammen (Umsätze, Ausgaben, Rechnungsbeträge)
- Nicht künstlich begrenzt oder gerundet sind (wie Gehälter oder Produktpreise)

Typische Anwendungen in der Berufspraxis umfassen Rechnungsbeträge, Spesenabrechnungen, Umsatzzahlen, Ausgabenbelege und ähnliche transaktionsbasierte Daten. Die Software filtert automatisch Werte unter 10 aus der Analyse heraus, da sehr kleine Zahlen oft nicht dem Benfordschen Gesetz folgen und die Analyseergebnisse verzerren können.

Analysetypen und ihre spezifischen Anwendungen

Die Anwendung bietet vier verschiedene Analysetypen, die jeweils unterschiedliche Aspekte potenzieller Unregelmäßigkeiten beleuchten:

Erste Ziffer (First Digit)

Die Analyse der ersten Ziffer ist der Klassiker der Benford-Analyse und folgt der bekannten logarithmischen Verteilung, bei der die Ziffer "1" mit 30,1%

Wahrscheinlichkeit auftritt, "2" mit 17,6%, und so weiter bis "9" mit nur 4,6%. Diese Analyse ist besonders effektiv bei der Erkennung von erfundenen oder manipulierten Daten, da Menschen intuitiv dazu neigen, erste Ziffern gleichmäßiger zu verteilen.

In der Wirtschaftsprüfung verwenden Sie die Erste-Ziffer-Analyse typischerweise für:

- Rechnungsbeträge und Ausgabenbelege
- Umsatzzahlen verschiedener Perioden
- Lagerbestände und Bewertungen
- Forderungen und Verbindlichkeiten

Zweite Ziffer (Second Digit)

Die Analyse der zweiten Ziffer folgt einer anderen, aber ebenso mathematisch definierten Verteilung. Hier sollten alle Ziffern von 0-9 nahezu gleichmäßig verteilt sein (jeweils etwa 10%), mit leichten Variationen. Diese Analyse ist besonders sensitiv für subtile Manipulationen, die bei der ersten Ziffer möglicherweise unentdeckt bleiben.

Anwendungsgebiete der Zweite-Ziffer-Analyse:

- Verfeinerte Betrugserkennung bei verdächtigen Datensätzen
- Überprüfung von Daten, die bei der ersten Ziffer unauffällig erscheinen
- Validierung automatisch generierter Daten und Berechnungen

Erste zwei Ziffern (First Two Digits)

Diese umfassendste Analyse betrachtet die ersten beiden Ziffern zusammen (10, 11, 12, ... bis 99) und folgt einer komplexeren logarithmischen Verteilung. Sie ist besonders mächtig bei der Aufdeckung sophistizierter Manipulationen, da es für einen Manipulator nahezu unmöglich ist, die komplexe Verteilung von 90 verschiedenen Ziffernkombinationen intuitiv nachzuahmen.

Die Erste-zwei-Ziffern-Analyse eignet sich besonders für:

- Forensische Untersuchungen mit Betrugsverdacht
- Detaillierte Validierung kritischer Finanzdaten
- Überprüfung von Berechnungen und automatisierten Systemen

Endziffer-Untersuchung (Unit Digit Investigation)

Die Endziffer-Analyse untersucht die letzte Ziffer von Zahlen und erwartet eine gleichmäßige Verteilung aller Ziffern von 1-9 (jeweils etwa 11,1%). Diese Analyse ist besonders wertvoll zur Erkennung von Rundungsmanipulationen und psychologischen Rundungstendenzen.

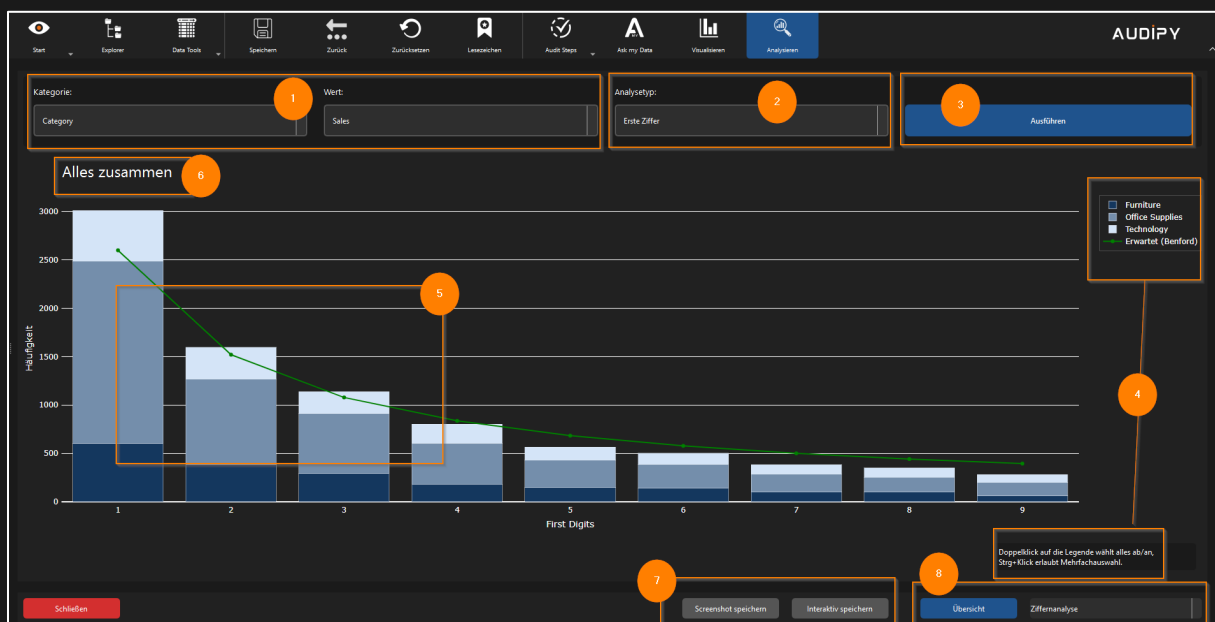
Diese Analyse wird automatisch auf Werte ab 100 angewendet und ist ideal für:

- Erkennung systematischer Rundungsmanipulationen
- Überprüfung von Preisgestaltungsstrategien
- Identifikation von psychologisch motivierten Zahlenanpassungen

16.5.3 Erstellung einer Ziffernanalyse

Bevor Sie eine Benford-Analyse durchführen, sollten Sie sicherstellen, dass Ihre Daten für diese Art der Untersuchung geeignet sind. Prüfen Sie, ob die numerischen Werte über mehrere Größenordnungen streuen und aus natürlichen Geschäftsprozessen stammen. Besonders wichtig ist, dass die Daten nicht künstlich begrenzt oder systematisch gerundet wurden, da dies die Aussagekraft der Analyse erheblich beeinträchtigen kann.

Bsp. Erste Ziffer



1. Kategorie & Wert auswählen ①

- Beginnen Sie Ihre Untersuchung mit einer Gesamtbetrachtung, indem Sie eine Kategorie wählen. Wählen Sie die relevante numerische Variable aus, die Sie analysieren möchten. Für den Einstieg empfiehlt sich die "Erste Ziffer"-Analyse, da diese die bekannteste und am besten interpretierbare Benford-Verteilung liefert.

2. Analysetyp ②

- Im Feld „Auswertung“ legen Sie fest, **welche Ziffernposition** analysiert werden soll. Hier kann ausgewählt werden welche Ziffern einbezogen werden sollen:

Erste Ziffer
Zweite Ziffer
Erste beiden Ziffern
Einerstellenuntersuchung

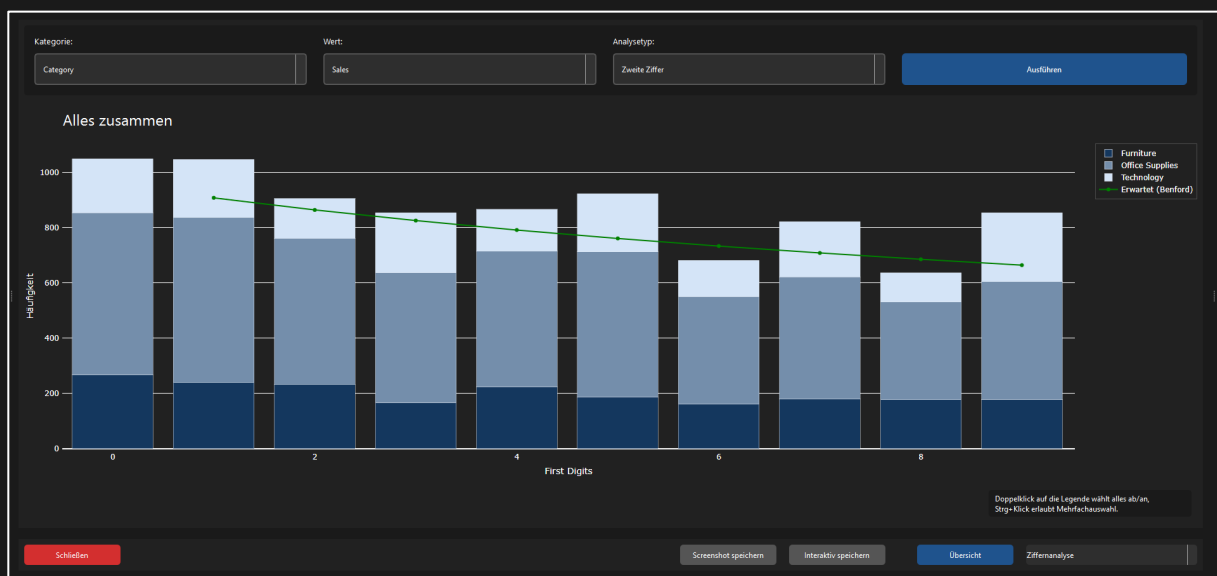
- **Erste Ziffer (s. Screenshot zu Beginn des Kapitels)**

Es wird ausschließlich die **erste (linke) Ziffer eines Betrags** analysiert. Diese klassische **Benford-Analyse** ist ein zentrales Verfahren zur Aufdeckung von Unregelmäßigkeiten oder möglichen Manipulationen in größeren Zahlenmengen.

- **Zweite Ziffer:**

Analysiert die **zweite Ziffer eines Betrags**. Diese Prüfung wird eingesetzt, wenn feinere Unregelmäßigkeiten identifiziert werden sollen, da auch hier typische Verteilungen erwartet werden.

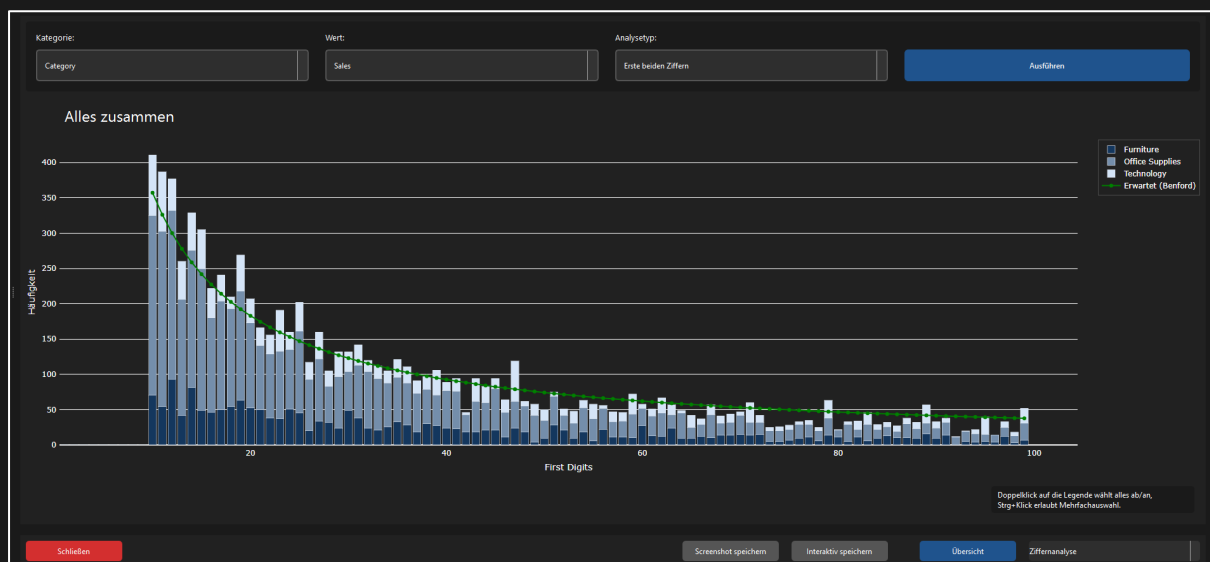
Auffällige Abweichungen können Hinweise auf Datenfehler oder bewusste Eingriffe sein.



- **Erste beiden Ziffern:**

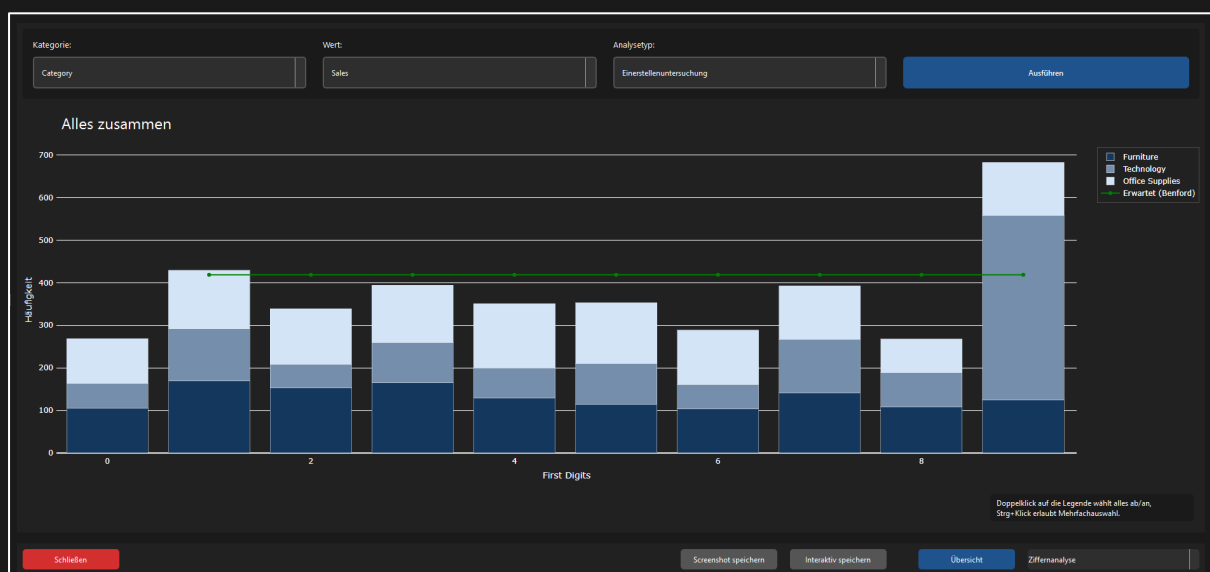
Bewertet die **Kombination der ersten beiden Ziffern** eines Betrags. Diese Variante bietet eine noch genauere Analyse, da eine feingliedrigere Sollverteilung besteht.

Schon kleine Abweichungen können hier auf systematische Manipulationen hinweisen.



▪ Einerstellenuntersuchung

Betrachtet die **Einerstelle** (die letzte Ziffer **vor dem Dezimalkomma**). Diese Analyseform wird gezielt eingesetzt, um **typische Rundungsmuster** (z. B. auffällige Häufungen auf 9, 19, 29...) oder **Verhaltensauffälligkeiten** bei der Betragseingabe zu erkennen.



Vertiefte Informationen, weit über das eben dargestellte, können Sie im folgenden SRP Handbuch (öffentlich zugänglich) von Andreas Wähnert finden und erlernen:

https://www.schleswig-holstein.de/mm/downloads/FM/handbuch_srp_aufgabe_13.pdf

- Die Wahl hängt davon ab, ob Sie beispielsweise eine **Benford-Analyse** oder eine Prüfung auf Manipulationen in den **letzten Ziffern** durchführen möchten.

3. Ziffernanalyse starten ③

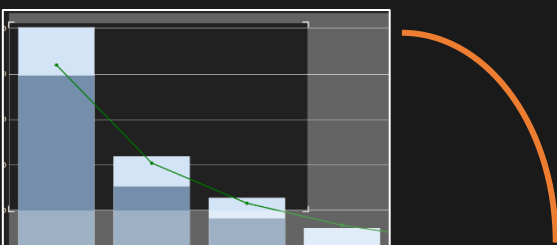
- Klicken Sie auf "**Ausführen**", um die Benford-Analyse zu starten. Das System generiert automatisch ein Balkendiagramm mit der tatsächlichen Verteilung Ihrer Daten (blaue Balken) und der erwarteten Benford-Verteilung (grüne Linie).
- Das System berechnet nun die **Häufigkeiten** der ausgewählten Ziffern und stellt sie in einem Diagramm dar. Signifikante Abweichungen zwischen diesen beiden Darstellungen deuten auf potenzielle Unregelmäßigkeiten hin.

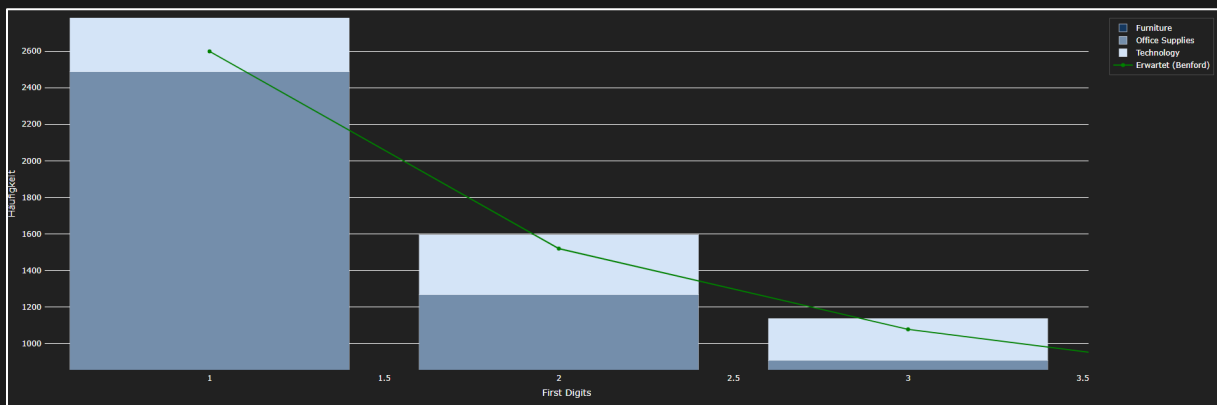
4. Legende interaktiv nutzen ④

- **Klick auf eine Legenden-Kategorie:** Blendet die entsprechende Datenreihe im Diagramm ein oder aus.
- Mit **Doppelklick** auf die Legende wählt man die Legende entweder insgesamt ab oder wieder an
- Mit **Strg + Klick** ist auch eine Mehrfachauswahl möglich (z.B. wenn man im Beispieldiagramm 2 Kategorien miteinander vergleichen möchte)

5. Interaktive Filterung durch Rechteckauswahl ⑤

- Ziehen Sie mit der Maus ein **Rechteck über einen Bereich im Diagramm**, um nur die darin enthaltenen Daten anzuzeigen und den Rest automatisch auszublenden. Diese Funktion erlaubt es Ihnen, spezifische Ausschnitte zu untersuchen, ohne die gesamte Ansicht zu verlieren.
- Möchte man die ursprüngliche Darstellung wieder haben, bedarf es einen Doppelklick in das Diagramm. So wird die Filterung wieder aufgehoben.





6. Kategorie Auswahl 6

- Über dieses Dropdown kann man **alle Kategorien** gemeinsam anzeigen oder gezielt **einzelne Kategorien** auswählen.
- Im Musterfall wurden die Kategorien analysiert. Mit Klick auf „Alles zusammen“, erhält man folgende Auswahlmöglichkeiten:

Alles zusammen
Furniture
Office Supplies
Technology

7. Screenshot speichern / Interaktiv speichern 7

- Mit **Screenshot speichern** erstellen Sie ein statisches Bild des Diagramms. Dieses können Sie entweder unter den Falldaten speichern oder können dieses auch an andere Personen zusenden. Hier empfiehlt sich allerdings das interaktive Speichern.
- Interaktiv speichern** legt das Diagramm als eigenständige HTML-Datei ab, die weiterhin interaktive Funktionen (z.B. Zoom oder Filter) unterstützt. So können Sie es frei verschicken, ohne dass der Empfänger selbst den AUDIPY auf dem Rechner hat. Die HTML-Datei wird ganz normal über den Browser geöffnet.

8. Übersicht / Auswahl 8

- Mithilfe des blauen Buttons kann zur Analyse Übersicht mit allen möglichen Analysen gewechselt werden. Es kann aber auch das Dropdown Menü benutzt werden um zu einer anderen Analyse zu wechseln.

16.5.4 Automatische Optimierungen innerhalb der Analyse

Das System erkennt automatisch, wenn Ihre Kategorieauswahl zu einer übermäßigen Anzahl von Gruppen führt. Bei mehr als 50 Kategorien erscheint eine Warnung mit den folgenden Handlungsoptionen:

- **Automatische Beschränkung:** Das System begrenzt die Analyse auf die volumenstärksten oder häufigsten Kategorien
- **Manuelle Festlegung:** Sie können eine spezifische Anzahl der wichtigsten Kategorien definieren
- **Vollanalyse:** Sie können alle Kategorien analysieren (bei Performanceeinbußen)

Diese Optimierung ist besonders relevant, wenn Sie beispielsweise Lieferantendaten mit hunderten verschiedenem Lieferanten analysieren, aber nur die größten für Ihre Risikobeurteilung relevant sind.

Automatische Datenfilterung

Die Software wendet automatisch verschiedene Filter an, um die Qualität der Benford-Analyse zu optimieren:

- **Mindestwertefilter:** Werte unter 10 werden automatisch ausgeschlossen, da sie typischerweise nicht dem Benfordschen Gesetz folgen
- **Spezialfilter für Endziffer-Analyse:** Bei der Endziffer-Untersuchung werden zusätzlich Werte unter 100 ausgeschlossen

Datenbereinigung: Nicht-numerische Werte und Nullwerte werden automatisch entfernt

Performance-Optimierung

Bei großen Datensätzen mit vielen Kategorien implementiert das System automatische Performance-Optimierungen, einschließlich progressiver Datenverarbeitung mit Fortschrittsanzeige. Dies ermöglicht auch bei umfangreichen Prüfungsunterlagen eine flüssige Analyse ohne Systemblockaden.

16.5.5 Erweiterte Analysemöglichkeit

Ein bewährter Ansatz für umfassende Untersuchungen ist die sequenzielle Anwendung verschiedener Analysetypen.

Mehrstufige Benford-Analyse

Beginnen Sie mit der "Erste Ziffer"-Analyse für einen ersten Überblick. Bei Auffälligkeiten wechseln Sie zur "Zweite Ziffer"-Analyse für eine detailliertere Betrachtung. Die "Erste zwei Ziffern"-Analyse sollten Sie für forensische Untersuchungen reservieren, da sie die umfassendste, aber auch komplexeste Betrachtung bietet.

Die "Endziffer-Untersuchung" ergänzt diese Analysen um eine völlig andere Perspektive und kann Rundungsmanipulationen aufdecken, die bei den anderen Analysetypen unentdeckt bleiben. Diese Kombination verschiedener Benford-Tests erhöht die Wahrscheinlichkeit, auch sophisticated Manipulationen zu entdecken.

Zeitreihenanalyse mit Benford-Methoden

Nutzen Sie die Kategoriefunktion für zeitbasierte Analysen, indem Sie Zeiträume (Monate, Quartale, Jahre) als Kategorien definieren. Dies ermöglicht es Ihnen zu identifizieren, ob Unregelmäßigkeiten kontinuierlich auftreten oder auf bestimmte Perioden beschränkt sind. Temporäre Abweichungen könnten auf einmalige Fehler oder periodenspezifische Manipulationen hindeuten, während kontinuierliche Anomalien auf systematische Probleme schließen lassen.

Vergleichende Benford-Analyse

Die Kategoriefunktion ermöglicht auch vergleichende Analysen zwischen verschiedenen Geschäftsbereichen, Standorten oder Prozessen. Sie können beispielsweise die Datenqualität verschiedener Filialen vergleichen oder die Ausgabenmuster verschiedener Abteilungen benchmarken. Dies ist besonders wertvoll für die Identifikation von Best Practices und die gezielte Verbesserung interner Kontrollen.

16.5.6 Interpretation der Ergebnisse und Risikobeurteilung

Das Benford-Diagramm zeigt Ihnen die tatsächliche Verteilung Ihrer Daten als blaue Balken und die theoretisch erwartete Benford-Verteilung als grüne Linie. Die Interpretation erfolgt durch den Vergleich dieser beiden Darstellungen:

- **Normale Übereinstimmung:** Wenn die blauen Balken eng der grünen Linie folgen, deutet dies auf natürliche, unmanipulierte Daten hin. Kleine Abweichungen sind normal und können durch Zufallsschwankungen oder branchenspezifische Eigenarten erklärt werden.
- **Signifikante Abweichungen:** Große Unterschiede zwischen der tatsächlichen und erwarteten Verteilung erfordern weitere Untersuchungen. Besonders auffällig sind:
 - Übermäßig hohe Häufigkeiten bei bestimmten Ziffern (mögliche Manipulation)
 - Zu gleichmäßige Verteilungen (typisch für erfundene Daten)
 - Ungewöhnliche Spitzen oder Täler in der Verteilung

Kategorienbasierte Risikoanalyse

Wenn Sie mit **Kategorien** arbeiten, ermöglicht das interaktive Auswahlfeld oberhalb des Diagramms eine systematische Risikobeurteilung verschiedener Geschäftsbereiche. Sie können zwischen "Alle zusammen" und spezifischen Kategorien wechseln, um zu identifizieren, in welchen Bereichen Abweichungen auftreten.

Ein praktisches Vorgehen ist die **sequenzielle Analyse**: Beginnen Sie mit der Gesamtbetrachtung, identifizieren Sie Auffälligkeiten, und untersuchen Sie dann systematisch einzelne Kategorien, um die Quelle der Anomalien einzugrenzen. Dies ist besonders wertvoll bei der Prüfungsplanung, da Sie Ihre begrenzten Prüfungsressourcen auf die risikoreichsten Bereiche konzentrieren können.

Statistische Signifikanz und fachliche Beurteilung

Wichtig bei der **Interpretation ist, dass statistische Auffälligkeiten** nicht automatisch Betrug bedeuten. Die Benford-Analyse ist ein Screening-Instrument, das weitere Untersuchungen anzeigt, aber nicht allein als Beweis für Unregelmäßigkeiten dient. Berücksichtigen Sie immer den fachlichen Kontext Ihrer Daten. Natürliche Geschäftsprozesse können durchaus zu Abweichungen führen, beispielsweise wenn Preise typischerweise in bestimmten Bereichen liegen oder wenn regulatorische Grenzen die Datenverteilung beeinflussen.

16.5.7 Qualitätssicherung und praktische Erfahrung

Für aussagekräftige **Benford-Analysen** sollten Ihre **Datensätze mindestens 100-200 Werte** umfassen, idealerweise deutlich mehr. Die Daten sollten über mehrere Größenordnungen streuen und aus natürlichen Geschäftsprozessen stammen. Vermeiden Sie die Analyse von künstlich begrenzten Daten wie Gehältern, standardisierten Preisen oder regulatorisch festgelegten Beträgen.

Dokumentation und Nachweisführung

Dokumentieren Sie Ihre **Benford-Analysen systematisch, einschließlich der verwendeten Parameter, identifizierten Abweichungen und durchgeführten Folgeprüfungen**. Nutzen Sie die Export-Funktionen der Software, um Screenshots oder interaktive Versionen Ihrer Analysen für Arbeitspapiere oder Berichte zu speichern. Dies ist besonders wichtig für die Nachvollziehbarkeit in Prüfungsverfahren oder bei regulatorischen Nachfragen.

Integration in Prüfungs- und Kontrollprozesse

Die Benford-Analyse sollte als systematisches Element in Ihre Prüfungs- oder Kontrollprozesse integriert werden, **nicht als gelegentliches Ad-hoc-Werkzeug**. Definieren Sie klare Kriterien für die Interpretation von Abweichungen und dokumentierte Folgeprozesse für identifizierte Anomalien. Dies stellt sicher, dass die Benford-Analyse konsistent angewendet wird und ihre volle Wirksamkeit als Screening-Instrument entfalten kann.

Grenzen und Einschränkungen beachten

Bedenken Sie immer die Grenzen der Benford-Analyse. Sie ist ein **Screening-Instrument**, das weitere Untersuchungen anzeigt, aber nicht allein als Beweis für Unregelmäßigkeiten dient. Berücksichtigen Sie branchenspezifische Eigenarten, natürliche Geschäftsbegrenzungen und mögliche technische Ursachen für Abweichungen. Die statistische Auffälligkeit muss immer im fachlichen Kontext bewertet und durch zusätzliche Prüfungshandlungen substantiiert werden.

Die Benford-Analyse stellt somit ein mächtiges, wissenschaftlich fundiertes Werkzeug dar, das Ihre Fähigkeit zur Identifikation von Unregelmäßigkeiten und Risiken erheblich verbessert und gleichzeitig objektive, nachvollziehbare Grundlagen für weitere Prüfungshandlungen liefert

16.6 Lückenanalyse



Die **Lückenanalyse** ist ein fundamentales Prüfungswerkzeug zur systematischen Überprüfung der Vollständigkeit und Integrität sequenzieller Nummernkreise in Geschäftsdaten. Diese Analysemethode identifiziert fehlende Nummern in kontinuierlichen Zahlenfolgen und ist von entscheidender Bedeutung für die Beurteilung der Vollständigkeit von Aufzeichnungen und die Aufdeckung möglicher Manipulationen oder Systemfehler.

In der Wirtschaftsprüfung stellt die Lückenanalyse ein unverzichtbares Instrument zur Überprüfung der Vollständigkeitsbehauptung des Managements dar. Rechnungsnummern, Belegnummern, Buchungsnummern und andere sequenzielle Identifikatoren müssen lückenlos sein, um die ordnungsgemäße Erfassung aller Geschäftsvorfälle zu gewährleisten. Fehlende Nummern können auf gelöschte Transaktionen, systemische Probleme oder sogar betrügerische Aktivitäten hindeuten, bei denen Belege bewusst entfernt wurden, um Unregelmäßigkeiten zu verbergen.

Controller nutzen die Lückenanalyse zur Überwachung der Datenintegrität in ihren Systemen und zur Sicherstellung, dass alle Geschäftsvorfälle ordnungsgemäß erfasst wurden. Dies ist besonders wichtig bei der Monats- und Jahresabschlussbearbeitung, wo die Vollständigkeit der Daten für die Genauigkeit der Finanzberichterstattung entscheidend ist. Steuerberater verwenden diese Analyse zur Vorbereitung auf Betriebsprüfungen und zur Sicherstellung, dass alle steuerrelevanten Belege ordnungsgemäß dokumentiert und archiviert sind.

Die mathematische Einfachheit der Lückenanalyse macht sie zu einem besonders robusten Prüfungsinstrument. Anders als komplexere statistische Methoden ist die Identifikation fehlender Nummern in einer Sequenz eindeutig und nicht interpretationsbedürftig. Dies macht die Lückenanalyse zu einem objektiven, nachvollziehbaren Nachweis für Vollständigkeitsprobleme, der auch in rechtlichen Auseinandersetzungen Bestand haben kann.

16.6.1 Wofür nutzt man die Lückenanalyse?

Die Lückenanalyse kann wie folgt in der Praxis nützlich sein:

Wirtschaftsprüfung: Vollständigkeitsprüfung von Ausgangsrechnungen

Ein klassisches Anwendungsszenario in der Wirtschaftsprüfung ist die Überprüfung der Vollständigkeit von Ausgangsrechnungen. Sie konfigurieren die Analyse mit "Rechnungsnummer" als Nummernspalte, wählen "Jahr" als Reset-Feld (falls die Rechnungsnummerierung jährlich zurückgesetzt wird) und "Rechnungsdatum" für die X-Achse zur zeitlichen Darstellung.

Das Ergebnis zeigt Ihnen nicht nur fehlende Rechnungsnummern, sondern auch deren zeitliche Verteilung. Lücken am Jahresende könnten auf Systemprobleme während der Abschlussbuchungen hindeuten, während vereinzelte Lücken über das Jahr verteilt möglicherweise auf normale Stornierungen zurückzuführen sind. Diese Analyse unterstützt die Beurteilung der Vollständigkeitsbehauptung und hilft bei der Planung gezielter Prüfungshandlungen.

Controlling: Überwachung der Buchungsintegrität

Im Controlling verwenden Sie die Lückenanalyse zur Überwachung der Buchungsintegrität in Ihrem ERP-System. Sie wählen "Buchungsnummer" als Nummernspalte, "Kostenstelle" als Nummernkreis (um verschiedene Bereiche getrennt zu analysieren) und "Monat" als Reset-Feld, falls Ihre Buchungsnummerierung monatlich zurückgesetzt wird.

Die Analyse deckt Systemfehler, Datenübertragungsprobleme oder Prozesslücken auf, die die Genauigkeit Ihrer Finanzberichterstattung beeinträchtigen könnten. Regelmäßige Lückenanalysen als Teil Ihrer monatlichen Abschlussprozesse stellen sicher, dass alle Geschäftsvorfälle ordnungsgemäß erfasst wurden und erhöhen die Zuverlässigkeit Ihrer Controlling-Informationen.

Steuerberatung: Betriebsprüfungsvorbereitung

Als Steuerberater nutzen Sie die Lückenanalyse zur Vorbereitung Ihrer Mandanten auf Betriebsprüfungen. Sie analysieren "Belegnummer" als Nummernspalte, verwenden "Belegart" als Nummernkreis (um Eingangs- und Ausgangsrechnungen getrennt zu prüfen) und wählen "Jahr" als Reset-Feld entsprechend der Archivierungspraxis Ihres Mandanten.

Diese präventive Analyse identifiziert potenzielle Vollständigkeitsprobleme, bevor sie von der Finanzverwaltung entdeckt werden. Sie können systematisch jede identifizierte Lücke mit Ihrem Mandanten klären und dokumentieren, wodurch Sie für Nachfragen der Prüfer vorbereitet sind und das Verfahren beschleunigen können.

16.6.2 Bedienelemente und Konfigurationsoptionen

Das **Nummer-Feld** bildet das Herzstück der Lückenanalyse und bestimmt, welche Spalte auf fehlende Werte in der sequenziellen Reihenfolge überprüft wird. Das System bietet Ihnen alle Spalten Ihres Datensatzes zur Auswahl an und behandelt verschiedene Datentypen intelligent:

- **Numerische Spalten:** Werden direkt als Zahlensequenz analysiert
- **Text-Spalten:** Das System extrahiert automatisch numerische Bestandteile aus gemischten Formaten (z.B. "RE-001234" → 1234)
- **Gemischte Formate:** Alphanumerische Bezeichnungen werden auf ihre numerischen Komponenten reduziert

Diese flexible Datentyp-Behandlung ermöglicht es Ihnen, auch komplexe Nummernsysteme zu analysieren, wie sie in der Praxis häufig vorkommen. Rechnungsnummern folgen oft Mustern wie "2024-0001", "INV-001234" oder "RE240001", und das System kann die numerischen Sequenzen in all diesen Formaten identifizieren und auf Lücken überprüfen.

In der praktischen Anwendung wählen Wirtschaftsprüfer typischerweise Spalten wie "Rechnungsnummer", "Belegnummer", "Journalnummer" oder "Kundennummer" für die Analyse. Controller fokussieren oft auf "Kostenstellen-IDs", "Projekt-Nummern" oder "Buchungsreferenzen", während Steuerberater häufig "Beleg-IDs", "Ausgaben-Nummern" oder "Eingangsrechnungs-Nummern" untersuchen.

X-Achse und zeitliche Dimension

Das **X-Achse-Feld** ermöglicht Ihnen die Wahl zwischen einer numerischen und einer zeitlichen Darstellung der identifizierten Lücken:

"Nummer"-Option:

- Zeigt Lücken entlang der numerischen Sequenz an
- Ideal für die Identifikation von Mustern in der Nummernvergabe
- Ermöglicht die schnelle Lokalisierung größerer zusammenhängender Lückenbereiche

Datumsspalten-Option:

- Visualisiert Lücken über die Zeit hinweg
- Besonders wertvoll für die Identifikation zeitlicher Muster oder Anomalien
- Ermöglicht die Korrelation von Lücken mit Geschäftsereignissen oder Systemausfällen

Die zeitliche Darstellung ist besonders aufschlussreich in der Wirtschaftsprüfung, da sie Ihnen ermöglicht, Lücken mit spezifischen Geschäftsereignissen, Systemwartungen oder Personalwechseln zu korrelieren. Ein Controller kann beispielsweise feststellen, ob Lücken in Buchungsnummern mit Urlaubszeiten, Systemupgrades oder organisatorischen Veränderungen zusammenfallen.

Nummernkreis-Gruppierung

Das **Nummernkreis-Feld** bietet Ihnen die Möglichkeit, Ihre Lückenanalyse nach verschiedenen Kategorien oder Bereichen zu segmentieren:

"Alle Nummern"-Option:

- Behandelt alle Nummern als eine kontinuierliche Sequenz
- Geeignet für einfache, einheitliche Nummernsysteme
- Ideal für erste Gesamtübersichten

Kategorische Gruppierung:

- Ermöglicht die separate Analyse verschiedener Nummernkreise
- Jede Kategorie wird als eigenständige Sequenz behandelt
- Unverzichtbar für komplexe Organisationsstrukturen

In der Praxis verwenden Unternehmen oft verschiedene Nummernkreise für unterschiedliche Zwecke: Ausgangsrechnungen könnten mit "AR-" beginnen, Eingangsrechnungen mit "AP-", und interne Belege mit "JE-". Die Nummernkreis-Funktion ermöglicht es Ihnen, jeden dieser Bereiche separat auf Vollständigkeit zu prüfen, da eine durchgehende Nummerierung nur innerhalb des jeweiligen Kreises erwartet wird.

Reset-Feld für periodische Nummerierung

Das **Reset-Feld** berücksichtigt die betriebliche Realität, dass viele Unternehmen ihre Nummerierung periodisch zurücksetzen:

"Reset-Feld auswählen"-Option:

- Ignoriert periodische Zurücksetzungen
- Behandelt alle Nummern als kontinuierliche Sequenz

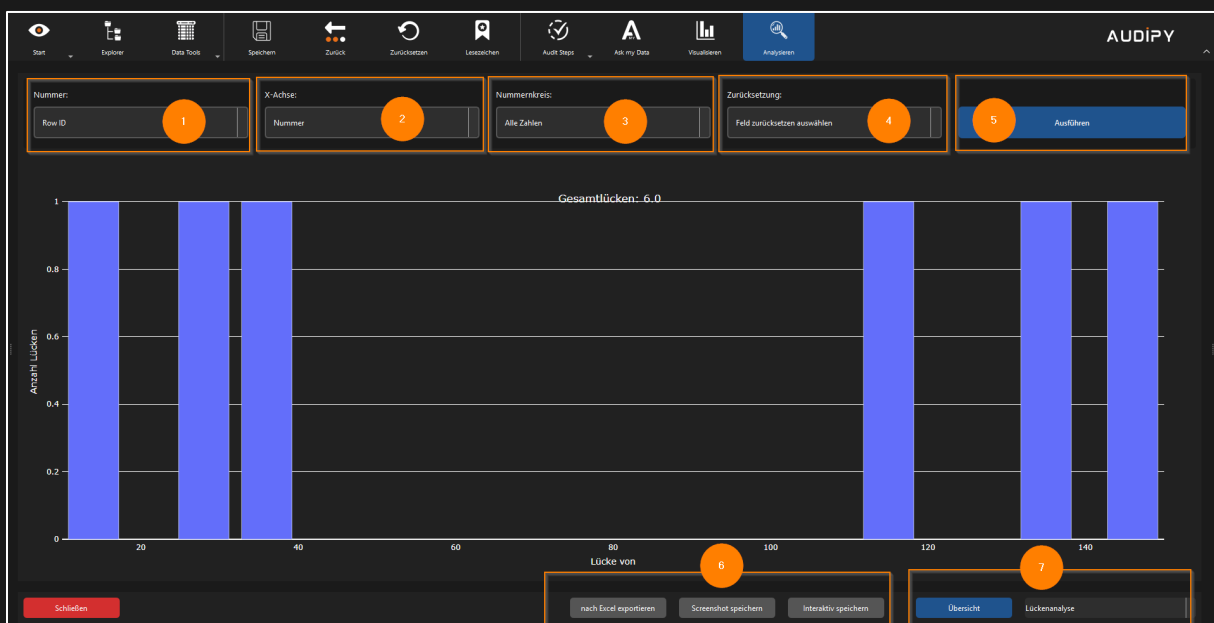
Spezifische Reset-Felder:

- Ermöglicht die Definition von Perioden, in denen Nummerierung neu beginnt
- Typische Reset-Felder sind Jahr, Monat, Quartal oder Geschäftsjahr
- Jede Reset-Periode wird als eigenständige Sequenz analysiert

Diese Funktionalität ist essentiell für die realitätsgerechte Analyse, da viele Unternehmen ihre Rechnungsnummern jährlich oder monatlich zurücksetzen. Ein Steuerberater kann beispielsweise die Rechnungsnummern eines Mandanten analysieren, der jedes Jahr bei "0001" neu beginnt, indem er "Jahr" als Reset-Feld wählt. Dadurch wird jedes Jahr separat auf Vollständigkeit geprüft, ohne dass der Jahreswechsel fälschlicherweise als massive Lücke interpretiert wird.

16.6.3 Erstellung einer Lückenanalyse

Bevor Sie mit der Lückenanalyse beginnen, sollten Sie die Struktur Ihrer Daten verstehen und die relevanten Spalten identifizieren. Prüfen Sie, ob Ihre Nummernspalte tatsächlich sequenzielle Werte enthalten sollte und ob es organisatorische Gründe für erwartete Lücken gibt (z.B. reservierte Nummernbereiche, stornierte Belege mit bewusst freigelassenen Nummern).



1. Nummer ①

- Wählen Sie zunächst die **Nummernspalte** aus, die Sie auf Vollständigkeit überprüfen möchten. Dies sollte eine Spalte mit sequenziellen Identifikatoren sein, wie Rechnungsnummern, Belegnummern oder Buchungsreferenzen
- Diese Nummer dient als **Anker**, um zu überprüfen, ob Nummern **fehlen**, **doppelt** oder **unvollständig** sind.

2. X-Achse ②

- Hier legen Sie fest, wie die **X-Achse** in Ihrem Diagramm dargestellt wird. Sie können beispielsweise **Datumsfelder** verwenden (z.B. „Belegdatum“) oder einfach nur die **Nummern** (z.B. 1, 2, 3 ...).
- Bei einer Datumsachse sehen Sie Zeitlücken (Tage ohne Eintrag). Bei einer numerischen Achse können Sie fehlende Nummern erkennen.

3. Nummernkreis ③

Entscheiden Sie, ob Sie:

- **"Alle Nummern"** wählen für einfache, einheitliche Systeme ohne Kategorisierung
- **Spezifische Nummernkreise** wählen, wenn verschiedene Dokumenttypen oder Bereiche unterschiedliche Sequenzen verwenden

Die Wahl hängt von der Komplexität Ihres Nummernsystems ab. In Zweifelsfällen beginnen Sie mit "Alle Nummern" und verfeinern die Analyse bei Bedarf durch Kategorisierung.

4. Zurücksetzung (optional) ④

Berücksichtigen Sie die Nummerierungspraxis Ihres Unternehmens:

- **Keine Reset-Feld-Auswahl:** Wenn Nummern kontinuierlich über lange Zeiträume vergeben werden
- **Jahr/Monat/Quartal:** Wenn die Nummerierung periodisch zurückgesetzt wird

5. Lückenanalyse durchführen ⑤

- Klicken Sie auf **Lückenanalyse durchführen**, um den Analyseprozess zu starten.
- Das System identifiziert automatisch alle fehlenden Nummern in den definierten Sequenzen und präsentiert diese in einer übersichtlichen Visualisierung.

6. Screenshot speichern / Interaktiv speichern ⑥

- Mit **Screenshot speichern** erstellen Sie ein statisches Bild des Diagramms. Dieses können Sie entweder unter den Falldaten speichern oder können dieses auch an andere Personen zusenden. Hier empfiehlt sich allerdings das interaktive Speichern.

- **Interaktiv speichern** legt das Diagramm als eigenständige HTML-Datei ab, die weiterhin interaktive Funktionen (z.B. Zoom oder Filter) unterstützt. So können Sie es frei verschicken, ohne dass der Empfänger selbst den AUDIPY auf dem Rechner hat. Die HTML-Datei wird ganz normal über den Browser geöffnet.
- **Nach Excel exportieren:** Lassen Sie sich die Daten direkt in Excel zwecks Weiterverarbeitung exportieren. Hierfür einmal den Speicherort und den Namen der Datei auswählen und die Exceldatei öffnet sich.

Nummernkre	Reset Field	Lücke von	Lücke bis	Anzahl der Lücken
		14	16	1
		28	30	1
		36	38	1
		115	117	1
		135	137	1
		146	148	1

Die Tabelle zeigt die **identifizierten Lücken** innerhalb eines **Nummernkreises**. Beispielsweise besteht zwischen **Nr. 14 und 16** eine Lücke von **einer fehlenden Nummer (15)**. Unterschiedliche Nummernkreise würden zu Beginn aufgeführt werden. Leere Felder im „Reset Field“ zeigen, dass keine Tages- oder Intervall-Zurücksetzung der Nummerierung berücksichtigt wurde.

7. Übersicht / Auswahl 7

- Mithilfe des blauen Buttons kann zur Analyse Übersicht mit allen möglichen Analysen gewechselt werden. Es kann aber auch das Dropdown Menü benutzt werden um zu einer anderen Analyse zu wechseln.

16.6.4 Ergebnisinterpretation und Risikobeurteilung

Die Lückenanalyse präsentiert ihre Ergebnisse in Form eines Balkendiagramms, das die Anzahl der Lücken über die gewählte X-Achse (numerisch oder zeitlich) darstellt. Der Titel des Diagramms zeigt die Gesamtanzahl aller identifizierten Lücken an, was einen sofortigen Überblick über das Ausmaß potenzieller Vollständigkeitsprobleme bietet.

- **Keine Lücken gefunden:** Das System zeigt eine entsprechende Meldung an und bestätigt die Vollständigkeit der Nummerierung. Dies ist ein

positives Prüfungsergebnis, das die Integrität des untersuchten Nummernkreises belegt.

- **Vereinzelte Lücken:** Einzelne oder wenige verstreute Lücken können auf normale Geschäftsprozesse hindeuten, wie stornierte Belege, Testbuchungen oder reservierte Nummern. Diese erfordern typischerweise eine Einzelfalluntersuchung zur Klärung der Ursachen.
- **Systematische Lückenmuster:** Regelmäßige oder gehäufte Lücken deuten auf systematische Probleme hin, die weitere Untersuchungen erfordern. Dies könnte auf Systemausfälle, Datenübertragungsprobleme oder sogar bewusste Manipulationen hinweisen.

Zeitliche Musteranalyse

Wenn Sie eine Datumsspalte für die X-Achse gewählt haben, ermöglicht die zeitliche Darstellung zusätzliche Erkenntnisse:

- **Periodische Häufungen:** Lücken, die zu bestimmten Zeiten auftreten (z.B. Monatsende, Quartalsende), können auf Systemwartungen, Personalwechsel oder Arbeitsüberlastung hindeuten.
- **Kontinuierliche Zeiträume ohne Lücken:** Bestätigen ordnungsgemäße Systemfunktion und Prozessdisziplin in diesen Perioden.
- **Plötzliche Lückenhäufungen:** Können auf Systemausfälle, Personalengpässe oder andere betriebliche Störungen hinweisen, die eine nähere Untersuchung verdienen.

Export und Detailanalyse

Das System bietet umfassende Export-Funktionen für die weitere Analyse der identifizierten Lücken:

- **Excel-Export:** Alle identifizierten Lücken werden in einer strukturierten Tabelle exportiert, die folgende Informationen enthält:
 - Nummernkreis-Zuordnung (falls konfiguriert)
 - Reset-Feld-Werte (falls konfiguriert)
 - Lücken-Startpunkt ("Lücke von")
 - Lücken-Endpunkt ("Lücke bis")
 - Anzahl fehlender Nummern
 - Zeitstempel (falls Datumsspalte gewählt)

Diese detaillierte Aufstellung ermöglicht es Ihnen, systematisch jede identifizierte Lücke zu untersuchen und zu dokumentieren. Die Excel-Datei enthält automatisch Autofilter und formatierte Spalten für eine benutzerfreundliche Weiterbearbeitung.

16.6.5 *Erweiterte Analysestrategien*

Nutzen Sie die **verschiedenen Konfigurationsoptionen** für eine mehrstufige Analyse derselben Daten. Beginnen Sie mit einer Gesamtübersicht ("Alle Nummern", kein Reset-Feld), um das Gesamtbild zu erfassen. Verfeinern Sie anschließend durch Nummernkreis-Gruppierung und Reset-Feld-Konfiguration, um spezifische Problembereiche zu identifizieren.

Diese **mehrstufige Herangehensweise** ist besonders wertvoll bei komplexen Organisationen mit verschiedenen Geschäftsbereichen und Nummernsystemen. Sie können systematisch von der Makro- zur Mikroebene vordringen und dabei die Ursachen identifizierter Probleme eingrenzen.

Zeitreihenanalyse von Lückenmustern

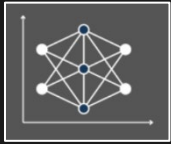
Verwenden Sie die **zeitliche X-Achse** für Trendanalysen der Datenqualität. Regelmäßige Lückenanalysen über mehrere Perioden hinweg können Trends aufzeigen: Verbessert oder verschlechtert sich die Datenintegrität über die Zeit? Gibt es saisonale Muster oder Korrelationen mit Personalwechseln oder Systemänderungen?

Diese longitudinale Betrachtung ist besonders wertvoll für Controller, die die Effektivität ihrer internen Kontrollsysteme überwachen möchten, und für Wirtschaftsprüfer, die das Risikoprofil eines Mandanten über mehrere Jahre bewerten.

Korrelationsanalyse mit Geschäftsereignissen

Nutzen Sie die **exportierten Lückendaten für Korrelationsanalysen** mit bekannten Geschäftsereignissen. Fallen Lücken mit Systemwartungen, Personalwechseln, Reorganisationen oder anderen bedeutsamen Ereignissen zusammen? Diese Analyse kann helfen, die Ursachen von Lücken zu verstehen und präventive Maßnahmen zu entwickeln.

16.7 Netzwerkdiagramm



Die **Netzwerkdiagramm-Analyse** ist ein hochentwickeltes Visualisierungswerkzeug zur Darstellung und Analyse komplexer Beziehungsstrukturen in Geschäftsdaten. Diese Analysemethode transformiert tabellarische Daten in interaktive, dreidimensionale Netzwerkdarstellungen, die verborgene Verbindungen, Abhängigkeiten und Muster sichtbar machen, welche in traditionellen Auswertungen unentdeckt bleiben würden. Durch die graphische Darstellung von Knoten (Entitäten) und Kanten (Beziehungen) ermöglicht sie eine intuitive Erfassung komplexer Datenstrukturen und Geschäftszusammenhänge.

In der Wirtschaftsprüfung revolutioniert die Netzwerkanalyse die Aufdeckung von Unregelmäßigkeiten und Betrugsmustern durch die Visualisierung von Transaktionsketten, Lieferantennetzwerken und Zahlungsströmen. Verdächtige Muster wie Scheintransaktionen, Rundumbuchungen oder ungewöhnliche Geschäftsbeziehungen werden durch die Netzwerkstruktur sofort erkennbar. Die Analyse kann komplexe Corporate-Governance-Strukturen, verbundene Unternehmen und potenzielle Interessenkonflikte aufdecken, die in herkömmlichen Prüfungsansätzen möglicherweise übersehen werden.

Controller nutzen Netzwerkdiagramme zur Analyse von Kostenflüssen, Lieferkettenstrukturen und internen Verrechnungsbeziehungen. Diese Visualisierungen helfen bei der Identifikation von Prozessengpässen, ineffizienten Strukturen und Optimierungspotenzialen in komplexen Organisationen. Die Darstellung von Kunden-Lieferanten-Beziehungen kann strategische Abhängigkeiten aufzeigen und bei der Risikobewertung unterstützen.

Steuerberater verwenden Netzwerkanalysen zur Untersuchung von Beteiligungsstrukturen, Cashflow-Ketten und steueroptimalen Organisationsformen. Die Visualisierung hilft bei der Bewertung von Steuergestaltungen, der Identifikation von nahestehenden Personen und der Analyse komplexer internationaler Unternehmensstrukturen. Besonders wertvoll ist die Methode bei der Vorbereitung auf Betriebsprüfungen, wo komplexe Geschäftsbeziehungen verständlich dargestellt werden müssen.

16.7.1 Wofür nutzt man ein Netzwerkdiagramm?

Das Netzwerkdiagramm wird in der Praxis wie folgt eingesetzt:

Wirtschaftsprüfung: Betrugs- und Unregelmäßigkeiten-Aufdeckung

In einer forensischen Untersuchung analysieren Sie Zahlungsströme zwischen verschiedenen Bankkonten. Sie konfigurieren das Netzwerk mit "Zahlendes Konto" als

Quelle und "Empfangendes Konto" als Ziel, verwenden "Transaktionsbetrag" als Kantengewicht und "Transaktionshäufigkeit" (durch "Anzahl"-Option) zur Darstellung der Beziehungsintensität.

Das resultierende Netzwerk macht verdächtige Muster sofort sichtbar: Rundumbuchungen erscheinen als geschlossene Kreise, Strohmänner als zentrale blaue Knoten mit ungewöhnlich vielen Verbindungen, und Geldwäsche-Strukturen als komplexe Verzweigungen mit auffälligen Volumenkonzentrationen. Die interaktive Exploration ermöglicht es Ihnen, von auffälligen Mustern ausgehend die beteiligten Konten zu identifizieren und weitere Prüfungshandlungen zu planen.

Die Ähnlichkeitsspalte könnte "Transaktionszeitpunkt" enthalten, wodurch zeitlich nahe Transaktionen visuell gruppiert werden und koordinierte Aktivitäten erkennbar werden. Export-Funktionen ermöglichen die Übertragung komplexer Netzwerke an Ermittlungsbehörden oder spezialisierte Forensik-Teams.

Controlling: Lieferketten- und Kostenfluss-Analyse

Als Controller eines Produktionsunternehmens analysieren Sie die Komplexität Ihrer Lieferkette. Sie definieren "Lieferant" als Quelle, "Empfänger" (interne Abteilungen oder externe Kunden) als Ziel, verwenden "Liefervolumen" als Kantengewicht und "Lieferzuverlässigkeit" als Ähnlichkeitsmaß.

Das Netzwerkdiagramm offenbart kritische Abhängigkeiten: Ein einzelner großer blauer Knoten könnte einen systemkritischen Lieferanten repräsentieren, dessen Ausfall die gesamte Produktion gefährden würde. Lange Ketten zwischen roten und grünen Knoten zeigen komplexe Abhängigkeiten auf, die Risiken bergen oder Optimierungspotenziale bieten.

Die physikbasierte Anordnung gruppiert ähnliche Lieferanten automatisch, wodurch alternative Beschaffungsquellen oder Diversifikationsmöglichkeiten sichtbar werden. Die Kantengewichtung durch Volumen macht Hauptlieferströme sofort erkennbar und unterstützt bei der Priorisierung von Supplier-Relationship-Management-Aktivitäten.

Steuerberatung: Beteiligungsstrukturen und Steueroptimierung

In der internationalen Steuerberatung visualisieren Sie komplexe Beteiligungsstrukturen eines multinationalen Konzerns. Sie verwenden "Muttergesellschaft" als Quelle, "Beteiligungsgesellschaft" als Ziel, "Beteiligungsquote" als Kantengewicht und "Steuerbelastung" des jeweiligen Standorts als Ähnlichkeitsmaß.

Das resultierende Netzwerk macht Optimierungspotenziale sichtbar: Gesellschaften mit hoher Steuerbelastung (dargestellt durch größere Distanzen aufgrund niedriger Ähnlichkeitswerte) könnten durch Umstrukturierungen optimiert werden. Komplexe

Verschachtelungen werden als dichte Knotencluster sichtbar, die möglicherweise vereinfacht werden können.

Die Farbkodierung hilft bei der Identifikation verschiedener Gesellschaftstypen: Operative Gesellschaften (oft grün als finale Empfänger), Holdingstrukturen (blau als Vermittler) und reine Finanzierungsgesellschaften (rot als reine Quellen). Export-Funktionen ermöglichen die Integration in spezialisierte Steuerplanungssoftware oder die Dokumentation für Beratungsgespräche.

16.7.2 Bedienelemente und Konfigurationsmöglichkeiten

Das **Quellspalte-Feld** bestimmt den Ausgangspunkt der Beziehungen in Ihrem Netzwerk und definiert die Entitäten, von denen Verbindungen ausgehen. Diese Spalte sollte kategorische Daten enthalten, die eine Seite Ihrer Geschäftsbeziehungen repräsentieren.

Das **Zielspalte-Feld** definiert die Endpunkte der Beziehungen und komplettiert die grundlegende Netzwerkstruktur. Die Kombination aus Quell- und Zielspalte erstellt die fundamentalen Verbindungen (Kanten) Ihres Netzwerks.

Verfügbare Optionen umfassen alle Textspalten Ihres Datensatzes. Typische Anwendungsszenarien:

- **Transaktionsanalyse:** "Zahlender" als Quelle, "Empfänger" als Ziel
- **Lieferkettenanalyse:** "Lieferant" als Quelle, "Kunde" als Ziel
- **Organisationsstrukturen:** "Abteilung" als Quelle, "Kostenstelle" als Ziel
- **Beteiligungsstrukturen:** "Muttergesellschaft" als Quelle, "Tochtergesellschaft" als Ziel

Die Wahl der Quell- und Zielspalten bestimmt fundamental die Interpretation Ihres Netzwerks. In der Wirtschaftsprüfung könnte die Analyse "Lieferant → Unternehmen → Kunde" komplexe Transaktionsketten aufdecken, während Controller mit "Kostenstelle → Projekt → Budgetverantwortlicher" interne Verantwortungsstrukturen visualisieren können.

Ähnlichkeits- und Gewichtungsmechanismen

Das **Ähnlichkeitsspalte-Feld** ermöglicht die Integration numerischer Beziehungsmaße, die die Stärke oder Ähnlichkeit von Verbindungen quantifizieren. Diese erweiterte Funktionalität transformiert einfache binäre Verbindungen in gewichtete, nuancierte Beziehungsdarstellungen.

Verfügbare Optionen:

- **"Keine"** - Erstellt ungewichtete Verbindungen für einfache Beziehungsanalysen
- **Alle numerischen Spalten** - Ermöglicht die Integration quantitativer Beziehungsmaße

Die Ähnlichkeitswerte beeinflussen die visuelle Darstellung des Netzwerks, indem sie die Länge der Verbindungslinien steuern: Höhere Ähnlichkeitswerte führen zu kürzeren Verbindungen und bringen verwandte Knoten näher zusammen. Dies ist besonders wertvoll für die Analyse von Korrelationen, Übereinstimmungsgraden oder Kooperationsintensitäten.

Das **Kantengewicht-Feld** bietet zusätzliche Gewichtungsmöglichkeiten für die Netzwerkdarstellung und beeinflusst die Dicke der Verbindungslinien. Diese Funktion ist essentiell für die Darstellung von Volumen, Häufigkeiten oder Intensitäten in Geschäftsbeziehungen.

Verfügbare Optionen:

- **"Keine"** - Alle Verbindungen werden mit einheitlicher Stärke dargestellt
- **"Anzahl"** - Automatische Zählung der Verbindungshäufigkeiten zwischen Entitäten
- **Alle numerischen Spalten** - Verwendung spezifischer Werte als Kantengewichte

Die "Anzahl"-Option ist besonders mächtig, da sie automatisch berechnet, wie oft bestimmte Verbindungen in Ihren Daten auftreten. In der Praxis bedeutet dies: Wenn "Lieferant A" zehnmal an "Unternehmen B" liefert, wird diese Verbindung als dickere Linie dargestellt als eine einmalige Lieferbeziehung. Dies macht Hauptgeschäftsbeziehungen sofort visuell erkennbar.

Export-Optionen und Interoperabilität

Das **Optionen-Feld** bietet spezialisierte Export-Funktionen für die Integration mit professionellen Netzwerkanalyse-Tools:

Gephi (GEXF)

Gephi ist die weltweit führende Open-Source-Software für Netzwerkanalyse und -visualisierung. Der GEXF-Export überträgt Ihr Netzwerk mit allen Attributen, Farben und Gewichtungen in Gephi, wo Sie erweiterte Analysen durchführen können:

- Community-Detection-Algorithmen zur Gruppenidentifikation
- Zentralitätsmaße zur Identifikation wichtiger Knoten
- Layoutalgorithmen für optimierte Darstellungen
- Umfangreiche Filteroptionen und statistische Analysen

GraphML

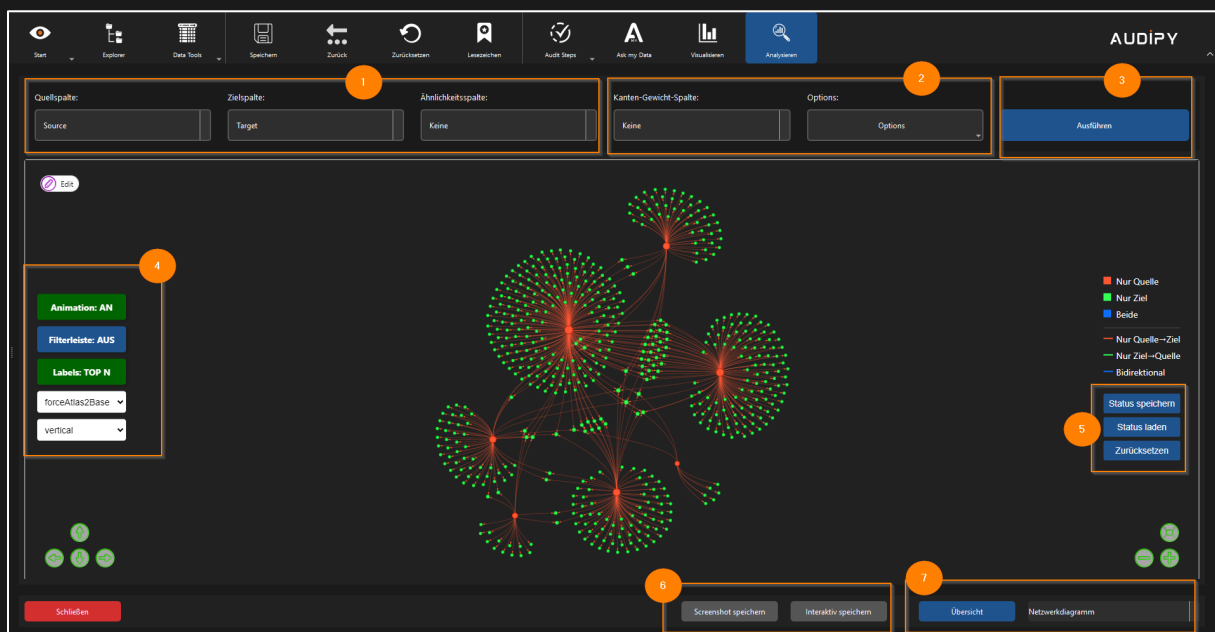
GraphML ist ein standardisiertes XML-Format für Netzwerkdaten, das von vielen wissenschaftlichen und kommerziellen Analyseprogrammen unterstützt wird. Dieser Export ermöglicht die Integration in spezialisierte Analysesoftware und die Archivierung von Netzwerkstrukturen für spätere Auswertungen.

Beide Export-Optionen bewahren alle in der Anwendung definierten Attribute, Gewichtungen und Farbkodierungen, sodass Ihre Analysen nahtlos in spezialisierte Tools übertragen werden können.

16.7.3 Erstellung eines Netzwerkdigrammes

Bevor Sie mit der Netzwerkerstellung beginnen, analysieren Sie Ihre Datenstruktur sorgfältig. Identifizieren Sie die Entitäten (die zu Knoten werden) und die Beziehungen zwischen ihnen (die zu Kanten werden). Ihre Daten sollten mindestens zwei kategorische Spalten enthalten, die Beziehungen zwischen verschiedenen Entitäten definieren.

Stellen Sie sicher, dass Ihre Daten bereinigt sind: Fehlende Werte in den Quell- oder Zielspalten werden automatisch ausgeschlossen, da sie keine gültigen Netzwerkverbindungen erstellen können. Konsistente Bezeichnungen sind essentiell, da "Kunde A" und "Kunde_A" als verschiedene Knoten behandelt werden.



1. Grundkonfiguration des Netzwerk ①

- Beginnen Sie mit der Definition der grundlegenden Netzwerkstruktur durch Auswahl der **Quellspalte** und **Zielspalte**. Diese Entscheidung ist fundamental, da sie die Richtung und Interpretation aller Beziehungen bestimmt.
- Für eine erste Exploration wählen Sie "**Keine**" für Ähnlichkeit und Kantengewicht, um die grundlegende Struktur ohne zusätzliche Komplexität zu verstehen.

2. Erweiterte Parameter-Konfiguration ②

- Nach der erfolgreichen Erstellung des Grundnetzwerks erweitern Sie die Analyse durch:
 - **Ähnlichkeitsspalte:** Wenn Sie numerische Beziehungsmaße haben (Korrelationen, Bewertungen, Intensitäten)

- **Kantengewicht:** Für die Darstellung von Volumen, Häufigkeiten oder Wichtigkeiten
- **"Anzahl"** als Kantengewicht für die automatische Häufigkeitsanalyse von Beziehungen

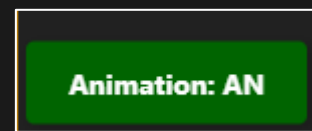
3. Netzwerkerstellung und Optimierung 3

- Klicken Sie auf **"Ausführen"**, um das interaktive Netzwerkdiagramm zu generieren. Das System erstellt automatisch ein physikbasiertes Layout, das ähnliche Knoten zusammenführt und das Netzwerk für optimale Lesbarkeit organisiert.
- Die Erstellung kann je nach Datengröße einige Momente dauern. Das System verarbeitet dabei automatisch Knoteneigenschaften, Kantenstärken und Farbkodierungen.

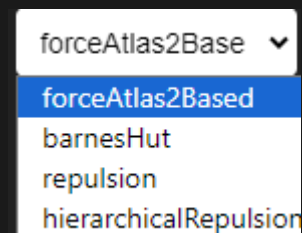
4. Interaktive Exploration und Analyse 4

- Das generierte Netzwerk bietet umfangreiche Interaktionsmöglichkeiten zur detaillierten Analyse. Nutzen Sie die verfügbaren Steuerelemente für verschiedene Analyseperspektiven und optimieren Sie die Darstellung für Ihre spezifischen Erkenntnisziele.

- Das **Animationssteuerung** ermöglicht das Ein- und Ausschalten der physikbasierten Animation. In der aktiven Position simuliert das System physikalische Kräfte zwischen den Knoten, wodurch sich das Netzwerk dynamisch reorganisiert und optimale Positionen für verwandte Knoten findet. Die Deaktivierung fixiert alle Positionen und ermöglicht präzise manuelle Anpassungen ohne störende Bewegungen.



- Das **Physik-Dropdown** bietet verschiedene Layoutalgorithmen für unterschiedliche Netzwerktypen:

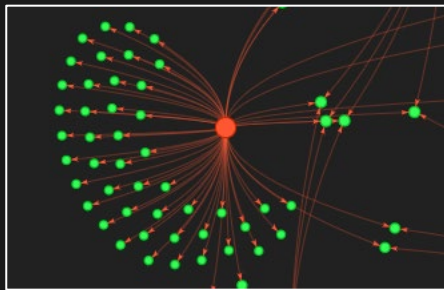


- **forceAtlas2Based:** Optimiert für große Netzwerke mit klarer Community-Struktur
- **barnesHut:** Effizient für mittlere bis große Netzwerke, gute Balance zwischen Performance und Qualität
- **repulsion:** Einfacher Abstoßungsalgorithmus, ideal für kleinere Netzwerke
- **hierarchicalRepulsion:** Speziell für hierarchische Strukturen wie Organigramme

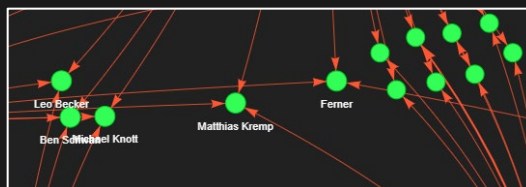
- Das **Kantenglättungs-Dropdown** steuert die visuelle Darstellung der Verbindungslinien:



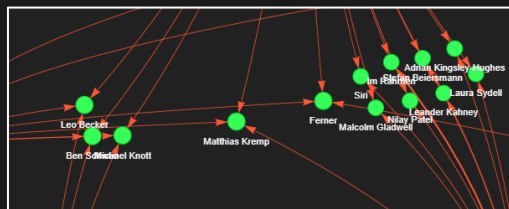
- **vertical/horizontal**: Geradlinige Verbindungen für klare, technische Darstellungen
 - **dynamic**: Adaptive Kurven, die sich automatisch an die Netzwerkdichte anpassen
 - **cubicBezier**: Elegante Kurven für ästhetische Darstellungen
 - **curvedCW/curvedCCW**: Gerichtete Kurven für die Darstellung von Richtungsflüssen
- Die Label-Steuerung (Beschriftung) bietet **drei Modi** für die Knotenbeschriftung:
- **"Labels: AUS"**: Minimale Darstellung ohne Beschriftungen für Übersichtlichkeit bei dichten Netzwerken



- **"Labels: TOP N"**: Zeigt nur die wichtigsten Knoten (basierend auf Verbindungsgrad) beschriftet an



- **"Labels: ALLE"**: Vollständige Beschriftung aller Knoten für detaillierte Analysen



- Das **Filter-Menü** ermöglicht die dynamische Anpassung der Netzwerkdarstellung:
 - Knotenfilterung nach Eigenschaften oder Verbindungsgrad
 - Kantenfilterung nach Gewicht oder Typ
 - Temporäre Ausblendung bestimmter Netzwerkteile für fokussierte Analysen

5. Interaktive Exploration und Analyse 5

- Die **Zustandsspeicherung** ermöglicht die Archivierung spezifischer Netzwerkkonfigurationen:
 - **"Status speichern"**: Erstellt eine JSON-Datei mit allen Knotenpositionen, Einstellungen und Anpassungen
 - **"Status laden"**: Wiederherstellung gespeicherter Konfigurationen für konsistente Darstellungen
 - **"Zurücksetzen"**: Rückkehr zur ursprünglichen automatischen Anordnung
- Diese Funktionen sind essentiell für die Dokumentation von Analyseergebnissen und die Reproduzierbarkeit von Erkenntnissen in Prüfungsverfahren oder Präsentationen

6. Screenshot speichern / Interaktiv speichern 6

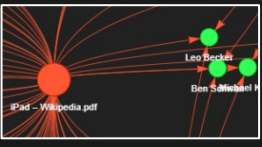
- Mit **Screenshot speichern** erstellen Sie ein statisches Bild des Diagramms. Dieses können Sie entweder unter den Falldaten speichern oder können dieses auch an andere Personen zusenden. Hier empfiehlt sich allerdings das interaktive Speichern.
- **Interaktiv speichern** legt das Diagramm als eigenständige HTML-Datei ab, die weiterhin interaktive Funktionen (z.B. Zoom oder Filter) unterstützt. So können Sie es frei verschicken, ohne dass der Empfänger selbst den AUDIPY auf dem Rechner hat. Die HTML-Datei wird ganz normal über den Browser geöffnet.

7. Übersicht / Auswahl 7

- Mithilfe des blauen Buttons kann zur Analyse Übersicht mit allen möglichen Analysen gewechselt werden. Es kann aber auch das Dropdown Menü benutzt werden, um zu einer anderen Analyse zu wechseln.

16.7.4 Visuelle Interpretation und Farbkodierung

Das Netzwerkdiagramm verwendet ein intelligentes Farbkodierungssystem zur sofortigen Identifikation verschiedener Knotentypen:

- **Rote Knoten (Source-only):** Entitäten, die nur als Ausgangspunkte von Beziehungen auftreten. In Geschäftsnetzwerken könnten dies reine Lieferanten, externe Geldgeber oder initiiierende Instanzen sein. Diese Knoten sind oft besonders interessant für Risikoanalysen, da sie Abhängigkeiten oder externe Einflüsse repräsentieren. 
- **Grüne Knoten (Target-only):** Entitäten, die ausschließlich als Endpunkte von Beziehungen fungieren. Dies könnten Endkunden, finale Kostenträger oder absorbierende Einheiten sein. Ihre Analyse hilft bei der Identifikation von Senken und Endpunkten in Geschäftsprozessen. 
- **Blaue Knoten (Bidirektional):** Entitäten, die sowohl Ausgangspunkte als auch Endpunkte von Beziehungen sind. Diese Knoten sind oft zentrale Akteure im Netzwerk und verdienen besondere Aufmerksamkeit, da sie als Vermittler, Hubs oder kritische Verbindungsknoten fungieren.

Kanten-Darstellung und Bedeutung

Die Verbindungslinien (Kanten) folgen einem entsprechenden Farbschema und zeigen durch Pfeile die Richtung der Beziehungen an:

- **Rote Pfeile:** Verbindungen, die ausschließlich von Source-only zu Target-only Knoten führen. Diese repräsentieren direkte, unidirektionale Geschäftsbeziehungen ohne Rückflüsse.
- **Grüne Pfeile:** Umgekehrte Verbindungen von Target-only zu Source-only Knoten, die ungewöhnliche oder besondere Geschäftsbeziehungen anzeigen können.
- **Blaue Doppelpfeile:** Bidirektionale Verbindungen, die wechselseitige Geschäftsbeziehungen anzeigen. Diese sind oft Indikatoren für enge Partnerschaften, interne Verrechnungen oder komplexe Abhängigkeiten.

Knotengröße und Netzwerk-Zentralität

Die Größe der Knoten korreliert mit ihrer Bedeutung im Netzwerk, gemessen am Verbindungsgrad (Anzahl der ein- und ausgehenden Verbindungen). Größere Knoten sind zentraler im Netzwerk und verdienen erhöhte Aufmerksamkeit in der Analyse, da sie:

- Potenzielle Schwachstellen oder Single Points of Failure darstellen
- Einflussreiche Akteure oder wichtige Geschäftspartner repräsentieren
- Optimierungspotenziale oder Risikoschwerpunkte anzeigen

16.7.5 Erweiterte Analysestrategien

Nutzen Sie **verschiedene Abstraktionsebenen für umfassende Analysen**. Beginnen Sie mit einer groben Übersicht (z.B. "Geschäftsbereich → Kostenstelle") und verfeinern Sie anschließend die Analyse auf spezifischeren Ebenen (z.B. "Kostenstelle → Projekt → Mitarbeiter"). Diese mehrstufige Herangehensweise ermöglicht es, sowohl strategische Muster als auch operative Details zu erfassen.

Temporale Netzwerkanalyse

Für **zeitbasierte Analysen** erstellen Sie separate Netzwerke für verschiedene Zeiträume und vergleichen die Strukturveränderungen. Dies ist besonders wertvoll für die Analyse von Geschäftsentwicklungen, saisonalen Mustern oder der Auswirkungen von Organisationsänderungen.

Kombinierte Gewichtungsanalysen

Experimentieren Sie mit **verschiedenen Gewichtungs- und Ähnlichkeitsmaßen** für dieselben Grunddaten. Ein Netzwerk könnte nach Volumen gewichtet werden, während ein zweites nach Häufigkeit oder Profitabilität gewichtet wird. Diese verschiedenen Perspektiven können unterschiedliche strategische Erkenntnisse liefern.

Integration mit statistischen Analysen

Nutzen Sie die **Export-Funktionen** zur Integration mit spezialisierten Netzwerkanalyse-Tools. Gephi und andere Software bieten erweiterte statistische Analysen wie Community-Detection, Zentralitätsmaße und Netzwerk-Metriken, die über die visuelle Analyse hinausgehen.

16.7.6 *Best Practices und methodische Empfehlungen*

Investieren Sie ausreichend Zeit in die Datenbereinigung vor der Netzwerkerstellung. Inkonsistente Bezeichnungen, Tippfehler oder unterschiedliche Schreibweisen führen zu fragmentierten Netzwerken mit künstlich vielen Knoten. Standardisieren Sie Entitätsnamen und verwenden Sie eindeutige Identifikatoren wo möglich.

Skalierbarkeit und Performance

Bei **sehr großen Netzwerken (>1000 Knoten)** sollten Sie die Analyse schrittweise angehen. Beginnen Sie mit Filtern oder Aggregationen, um die Komplexität zu reduzieren, und fokussieren Sie auf die relevantesten Teilnetzwerke für Ihre spezifische Fragestellung.

Dokumentation und Reproduzierbarkeit

Dokumentieren Sie Ihre **Analyseparameter** systematisch und nutzen Sie die Zustandsspeicherung für wichtige Erkenntnisse. Dies ist besonders wichtig für Prüfungsverfahren oder regulatorische Nachfragen, wo die Nachvollziehbarkeit von Analysen essentiell ist.

Integration in Arbeitsabläufe

Etablieren Sie die Netzwerkanalyse als regelmäßigen Bestandteil Ihrer analytischen Prozesse. Regelmäßige Netzwerkanalysen können Trends, Veränderungen und neue Muster frühzeitig identifizieren, bevor sie in traditionellen Auswertungen sichtbar werden.

Die **Netzwerkdiagramm-Analyse** stellt somit ein mächtiges, vielseitiges Werkzeug dar, das durch seine Kombination aus mathematischer Präzision, visueller Intuition und interaktiver Exploration optimal geeignet ist, komplexe Geschäftsbeziehungen zu verstehen und strategische Erkenntnisse zu gewinnen, die in traditionellen Analysemethoden verborgen bleiben würden.

16.8 Geldfluss



Die **Geldfluss-Analyse** mittels Sankey-Diagramme ist eine hochspezialisierte Visualisierungsmethode zur Darstellung von Finanzströmen, Cashflows und Werttransfers zwischen verschiedenen Entitäten. Diese Analysemethode transformiert komplexe Finanztransaktionen in intuitive, proportionale Flussdarstellungen, bei denen die Breite der Verbindungslinien das Volumen der Geldströme widerspiegelt. Dadurch werden Hauptfinanzierungsquellen, kritische Zahlungsströme und Liquiditätskonzentrationen auf den ersten Blick erkennbar.

16.8.1 Wofür nutzt man die Geldflussmethode?

Die Geldflussmethode kann in unterschiedlichen Bereichen zahlreiche Transaktionen analysieren.

Wirtschaftsprüfung: Konzern-Cashflow und Verrechnungspreisanalyse

In der **Konzernprüfung** analysieren Sie interne Verrechnungen und Cashflows zwischen Tochtergesellschaften. Sie konfigurieren die Analyse mit "Zahlende Gesellschaft" als Quelle, "Empfangende Gesellschaft" als Ziel und "Verrechnungsbetrag" als Wert. Das resultierende Sankey-Diagramm macht sofort sichtbar, welche Gesellschaften als primäre Cashflow-Quellen fungieren und welche hauptsächlich Empfänger sind.

Ungewöhnlich **breite Flüsse** zwischen bestimmten Gesellschaften können auf überhöhte Verrechnungspreise oder verdächtige Transaktionen hindeuten. Die Kategorieauswahl ermöglicht es Ihnen, sich auf einzelne Gesellschaften zu fokussieren und deren spezifische Transaktionsmuster zu analysieren. Gesellschaften, die sowohl große Zu- als auch Abflüsse aufweisen, verdienen besondere Aufmerksamkeit als potenzielle Durchleitungsgesellschaften.

Die Farbkodierung hilft bei der Verfolgung von Geldströmen durch komplexe Strukturen: Sie können visuell verfolgen, wie Mittel von der Muttergesellschaft durch verschiedene Holdings zu operativen Einheiten fließen und dabei potenzielle Optimierungen oder Compliance-Probleme identifizieren.

Controlling: Working Capital und Lieferkettenfinanzierung

Im **Working Capital Management** analysieren Sie die Geldflüsse zwischen Lieferanten, Ihrem Unternehmen und Kunden. Die Konfiguration "Lieferant" → "Unternehmen" → "Kunde" mit "Zahlungsvolumen" als Wert zeigt die Liquiditätszyklen und Cashflow-Timing.

Das Sankey-Diagramm macht Liquiditätsengpässe sichtbar: Wenn die eingehenden Flüsse von Kunden schmaler sind als die ausgehenden Flüsse zu Lieferanten, deutet dies auf Working Capital-Bedarf hin. Die zeitliche Komponente kann durch separate Analysen für verschiedene Perioden dargestellt werden.

Die **Kategorieauswahl** ermöglicht die Fokussierung auf kritische Lieferanten oder Hauptkunden. Ein Lieferant mit ungewöhnlich breiten Zuflüssen könnte auf Abhängigkeiten hindeuten, die risikomanagement-technisch bedenklich sind. Umgekehrt können Kunden mit sehr breiten Abflüssen als strategische Partnerschaften identifiziert werden, die besondere Aufmerksamkeit verdienen.

Steuerberatung: Internationale Cashpool-Optimierung

Als **internationaler Steuerberater** optimieren Sie die Cashpool-Struktur eines multinationalen Konzerns. Sie analysieren "Land der Quelle" → "Zentraler Cashpool" → "Land des Ziels" mit "Cashflow-Volumen" als Wert. Das Sankey-Diagramm zeigt sofort, welche Länder netto Einzahler in den Pool sind und welche netto Empfänger.

Die **Visualisierung** hilft bei der Identifikation steueroptimaler Strukturen: Länder mit hohen Steuerlasten sollten idealerweise als Quellen erscheinen, während Länder mit günstigen Steuersätzen als Ziele fungieren sollten. Unausgewogene Flüsse können auf Optimierungspotenziale hindeuten.

Die Kategorieauswahl ermöglicht die **länderspezifische Analyse**: Sie können sich auf ein einzelnes Land fokussieren und dessen gesamte ein- und ausgehende Cashflows analysieren. Dies ist besonders wertvoll für die Argumentation gegenüber lokalen Steuerbehörden, da Sie die Geschäftszwecke und wirtschaftlichen Substanz der Transaktionen visuell darstellen können.

16.8.2 Bedienelemente und Datenstruktur-Konfiguration

Das **Quelle-Feld** bestimmt die Ausgangspunkte der Geldströme und sollte Entitäten repräsentieren, von denen finanzielle Mittel ausgehen. Diese könnten Geschäftsbereiche, Kostenstellen, Bankkonten, Lieferanten oder externe Finanzierungsquellen sein.

Das **Ziel-Feld** definiert die Empfänger der Geldströme und komplettiert die fundamentale Flussrichtung. Die Kombination aus Quelle und Ziel erstellt die Basis-Architektur Ihres Finanzfluss-Netzwerks.

Verfügbare Optionen umfassen alle Textspalten Ihres Datensatzes. Strategische Überlegungen zur Auswahl:

- **Hierarchische Strukturen:** "Konzern → Tochtergesellschaft → Geschäftsbereich" für Konzernanalysen
- **Operative Flüsse:** "Lieferant → Unternehmen → Kunde" für Wertketten-Analysen
- **Interne Allokationen:** "Zentralfunktion → Kostenstelle → Projekt" für interne Verrechnungen
- **Finanzierungsstrukturen:** "Investor → Holding → Operative Gesellschaft" für Kapitalfluss-Analysen

Die **Qualität Ihrer Geldfluss-Analyse** hängt maßgeblich von der logischen Konsistenz dieser Quelle-Ziel-Beziehungen ab. Inkonsistente oder zufällige Zuordnungen führen zu uninformativen Visualisierungen, während durchdachte Definitionen aussagekräftige Einsichten in Ihre Finanzarchitektur liefern.

Wertdefinition und Volumenquantifizierung

Das **Wert-Feld** bestimmt die numerische Größe, die durch die Breite der Sankey-Flüsse dargestellt wird. Diese Auswahl ist entscheidend für die Aussagekraft Ihrer Analyse, da sie definiert, was als "Volumen" des Geldflusses interpretiert wird.

Das System bietet alle numerischen Spalten Ihres Datensatzes zur Auswahl an. Typische Anwendungsszenarien umfassen:

- **Absolute Beträge:** Transaktionsvolumen, Umsätze, Kosten, Investitionen für die Darstellung tatsächlicher Geldmengen
- **Periodische Flüsse:** Monatliche Zahlungen, Quartalsvolumen, Jahresbudgets für zeitraumbezogene Analysen
- **Kumulative Werte:** Gesamtvolumen, YTD-Beträge, Lebenszeit-Transaktionen für langfristige Betrachtungen
- **Relative Kennzahlen:** Margen, Renditen, Anteile für proportionale Analysen

Die Wahl des Wertfeldes sollte immer im Kontext Ihrer Analyseziele erfolgen. Für Liquiditätsanalysen sind tatsächliche Zahlungsbeträge relevant, während für strategische Bewertungen möglicherweise Margen oder Renditen aussagekräftiger sind. Das System aggregiert automatisch mehrfache Transaktionen zwischen denselben Quelle-Ziel-Paaren, wodurch die Gesamtstärke der Beziehung dargestellt wird.

Kategoriebasierte Fokussierung

Das interaktive **Kategorieauswahlfeld** oberhalb des Diagramms ermöglicht die dynamische Fokussierung auf spezifische Teilbereiche Ihres Finanzflusses. Diese Funktionalität ist besonders wertvoll bei komplexen Organisationen mit vielen parallelen Geldströmen.

- **"Alle zusammen"** zeigt das vollständige Sankey-Diagramm mit allen identifizierten Geldströmen. Diese Gesamtsicht ist ideal für strategische Übersichten und die Identifikation der größten Finanzflüsse in Ihrer Organisation.
- **Spezifische Quellkategorien** filtern das Diagramm auf alle Flüsse, die von einer bestimmten Quelle ausgehen. Dies ermöglicht es, sich auf einzelne Geschäftsbereiche, Kostenstellen oder Finanzierungsquellen zu konzentrieren und deren spezifische Allokationsmuster zu analysieren.

Diese Fokussierungsfunktion ist essentiell für die systematische Analyse komplexer Finanzstrukturen. Sie können mit der Gesamtsicht beginnen, auffällige Muster identifizieren und dann gezielt in spezifische Bereiche "hineinzoomen", um die Ursachen und Details zu verstehen.

16.8.3 Qualitätssicherung und Datenintegrität

Stellen Sie sicher, dass Ihre Sankey-Diagramme die **tatsächlichen Geldflüsse vollständig** repräsentieren. Vergleichen Sie die Summen in der Visualisierung mit Ihren Quelldaten, um sicherzustellen, dass keine wesentlichen Transaktionen durch Datenqualitätsprobleme verloren gegangen sind.

Plausibilitätskontrollen

Nutzen Sie die **Sankey-Visualisierung** für Plausibilitätskontrollen: Ungewöhnliche Flussmuster, unlogische Richtungen oder extreme Volumenkonzentrationen können auf Datenfehler oder echte Geschäftsprobleme hindeuten, die weitere Untersuchungen erfordern.

Dokumentation und Reproduzierbarkeit

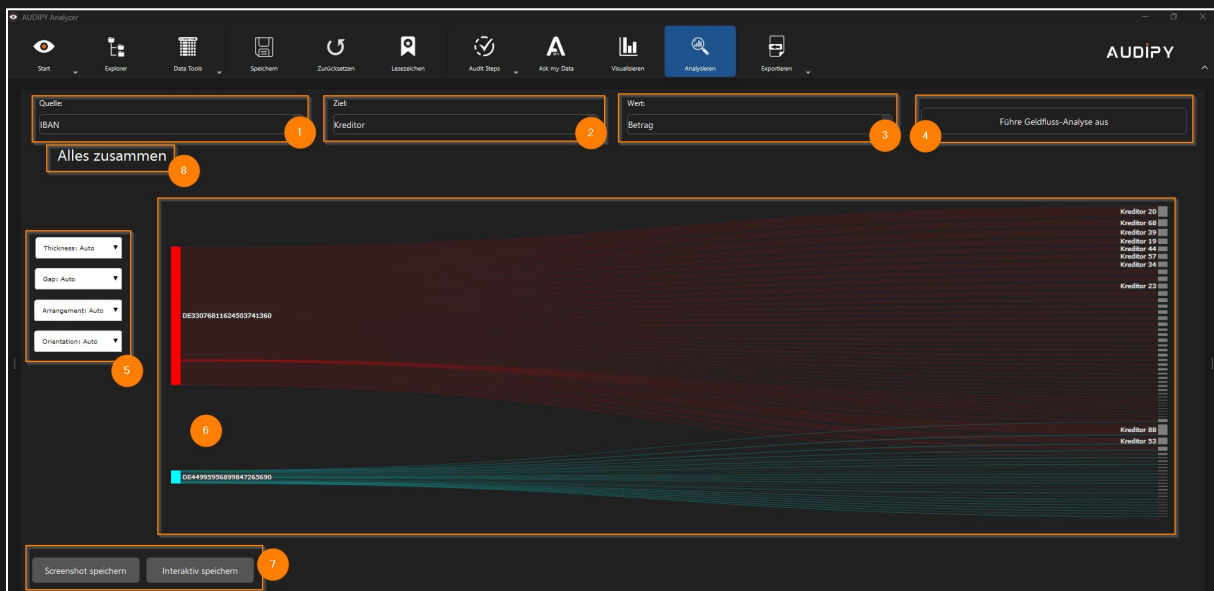
Dokumentieren Sie Ihre Analyseparameter und Export-Konfigurationen systematisch. Die Export-Funktionen ermöglichen es Ihnen, interaktive Versionen Ihrer Analysen für Präsentationen oder weitere Untersuchungen zu speichern, während Screenshots für statische Dokumentationen geeignet sind.

Die **Geldfluss-Analyse mit Sankey-Diagrammen** stellt somit ein mächtiges Werkzeug dar, das komplexe Finanzstrukturen in intuitive, proportionale Visualisierungen transformiert und dabei sowohl strategische Übersichten als auch detaillierte operative Einblicke ermöglicht, die für fundierte Geschäftsentscheidungen und Compliance-Anforderungen unerlässlich sind.

16.8.4 Erstellung einer Geldfluss-Analyse

Bevor Sie mit der **Geldfluss-Analyse** beginnen, validieren Sie die Struktur und Qualität Ihrer Finanzdaten. Stellen Sie sicher, dass Ihre Quelle-Ziel-Beziehungen logisch konsistent sind und echte Geldflüsse repräsentieren. Negative Werte können Rückzahlungen oder Korrekturen darstellen, sollten aber in ihrer Bedeutung klar sein.

Prüfen Sie die **Vollständigkeit Ihrer Daten**: Fehlende Werte in Quelle oder Ziel werden automatisch ausgeschlossen, können aber wichtige Finanzflüsse repräsentieren, die separate Behandlung erfordern. Konsistente Entitätsnamen sind essentiell, da "Abteilung Marketing" und "Marketing Abt." als separate Knoten behandelt werden.



8. Quelle 1

- Hier wird die **Datenquelle** oder **Eingangsgröße** definiert. Beispielsweise kann dies ein **Bankkonto**, eine **Abteilung** oder ein **Produktsegment** sein, von dem aus der Geldfluss ausgeht.

9. Ziel 2

- Dieser Punkt gibt an, **wer** das Geld erhält bzw. wohin das Geld fließt. Das können **Kunden**, **Regionen**, **Abteilungen** oder andere **Empfänger** sein.
- Dadurch wird deutlich, wie sich die Geldströme **aufteilen** und welche Ziele (Empfänger) am **stärksten** beteiligt sind.

10. Wert 3

- Hier wird das **numerische Feld** angegeben, das den Geldbetrag (z. B. **Währungsbeträge**) repräsentiert.
- Absolute Beträge für Volumenanalysen, relative Kennzahlen für Effizienzbetrachtungen oder spezifische Kostenarten für detaillierte Kostenverfolgung.

11. Geldfluss-Analyse starten 4

- Sobald alle Einstellungen vorgenommen sind, kann man über den „**Führe Geldfluss-Analyse aus**“ Button das Diagramm **laden** oder **aktualisieren**. Das System nutzt die getroffenen Auswahlfelder um den Geldfluss graphisch darzustellen.
- Diese Gesamtsicht zeigt Ihnen sofort die dominierenden Finanzflüsse und die allgemeine Struktur Ihrer Finanzarchitektur.
- Achten Sie auf die Proportionalität der Flüsse: Ungewöhnlich breite Ströme können auf Konzentrationrisiken oder Effizienzprobleme

hindeuten, während viele dünne Ströme auf Fragmentierung oder komplexe Strukturen hinweisen können.

12. Zusätzliche Diagramm-Einstellungen 5

- **Thickness (Dicke):** Beeinflusst die Breite der **Knoten (Entitäten)** in Ihrem Sankey-Diagramm und kann die Lesbarkeit erheblich verbessern:
 - i. Über die Einstellung der Balkenstärke lässt sich das Layout anpassen, um große und kleine Flüsse übersichtlich und korrekt darzustellen. Folgende Einstellungen sind möglich:

Auto (Standard):

Automatische Optimierung basierend auf der Datenkomplexität und Displaygröße
Thick (Dick): Breitere Knoten für bessere Lesbarkeit bei komplexen Diagrammen mit vielen Entitäten
Thin (Dünn): Schmale Knoten für die Maximierung des Platzes für Flussdarstellungen

Die optimale Einstellung hängt von der Anzahl Ihrer Entitäten und der Komplexität der Geldflüsse ab. Bei vielen kleinen Transaktionen sind dünne Knoten oft vorteilhaft, während bei wenigen großen Flüssen dickere Knoten die Übersichtlichkeit erhöhen.

Thick (Dicke Balken):



Erzwingt eine gleichmäßig dicke Darstellung der Flüsse, unabhängig vom tatsächlichen Betrag.

→ Gut geeignet, wenn visuelle Klarheit wichtiger ist als die absolute Betragsdifferenz.

Thin (Dünne Balken):



Alle Balken werden relativ dünn dargestellt.

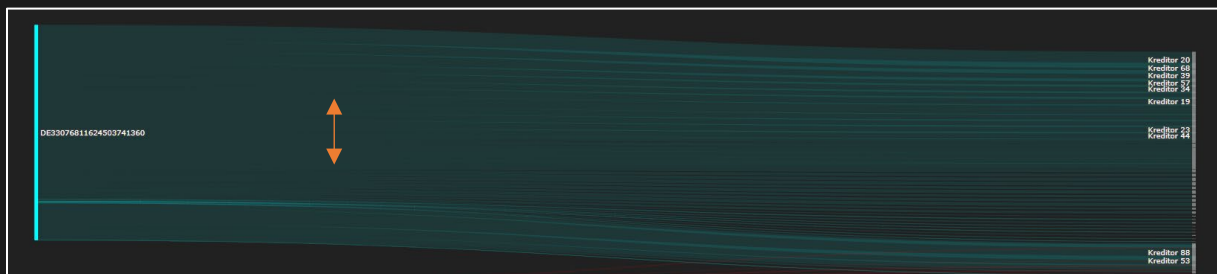
→ Praktisch, wenn viele kleine Flüsse dargestellt werden und Platz gespart werden muss.

- **Gap (Abstand):** Die Abstand-Steuerung reguliert den Platz zwischen den Knoten und beeinflusst die Gesamtstruktur der Darstellung:

- i. **Auto:** Automatische Optimierung für ausgewogene Darstellung
Small gap (Kleiner Abstand): Kompakte Darstellung für maximale Informationsdichte
Large gap (Großer Abstand): Weiträumige Anordnung für verbesserte Lesbarkeit einzelner Flüsse
- ii. Kleinere Abstände sind vorteilhaft bei einfachen Strukturen mit wenigen Entitäten, während größere Abstände bei komplexen Netzwerken mit vielen überlappenden Flüssen die Klarheit verbessern.

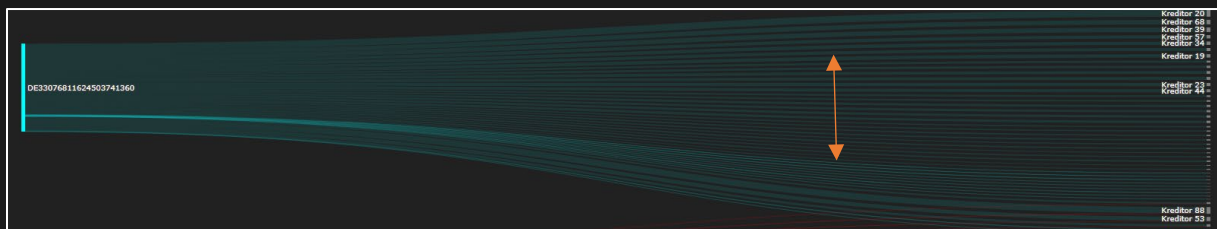
Small Gap:

Geringe Abstände – geeignet für kompakte Darstellungen mit vielen kleinen Flüssen.



Large Gap:

Größere Abstände – sorgt für mehr Übersicht bei wenigen, aber bedeutenden Flüssen.



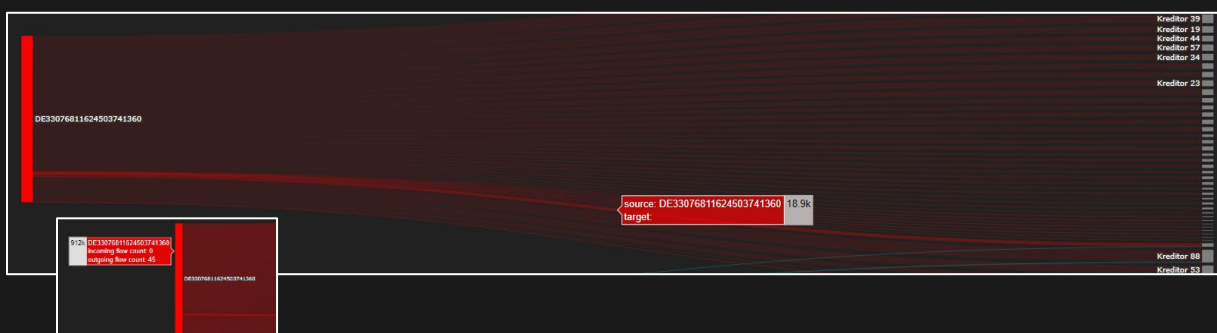
Option	Beschreibung
Snap <i>Anheften</i>	Automatische Optimierung mit Einrasterung für saubere Linienführung. Sorgt für aufgeräumte, gleichmäßige Verbindungen.
Perpendicular <i>Rechtwinklig</i>	Flüsse verlaufen rechtwinklig (90°-Winkel). Rechtwinklige Anordnung für technische, schematische Darstellungen

Freeform <i>Freiform</i>	Freie Positionierung für organische, natürliche Flussdarstellungen
Fixed <i>Fixiert</i>	Feste Positionen basierend auf Datenreihenfolge für reproduzierbare Layouts

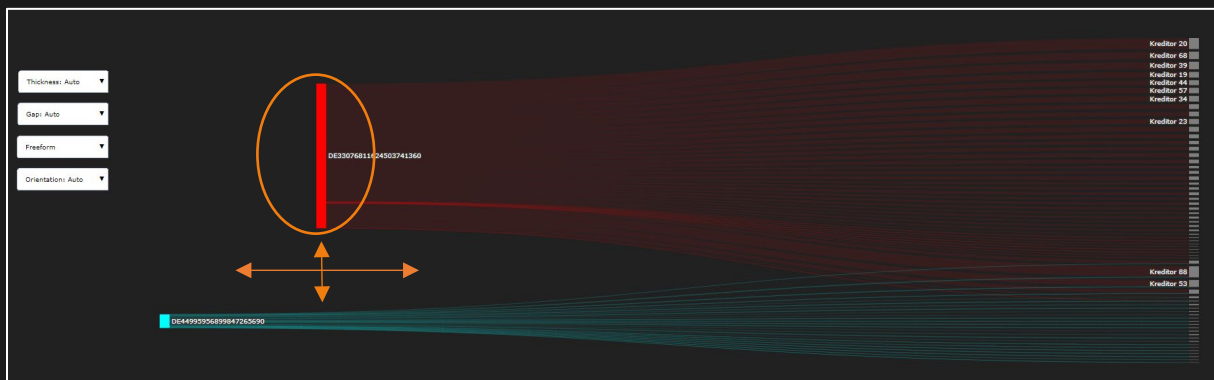
- **Orientation (Ausrichtung):** ermöglicht die Anpassung der Flussrichtung an verschiedene Darstellungsanforderungen:
 - i. **Horizontal:** Flüsse von links nach rechts, ideal für zeitliche Abfolgen oder Prozessschritte
 - ii. **Vertikal:** Flüsse von oben nach unten, geeignet für hierarchische Strukturen oder Wasserfalldarstellungen
- Die Orientierungswahl sollte die natürliche Leserichtung Ihrer Daten unterstützen und die intuitive Interpretation fördern.

13. Sankey-Diagramm ⑥

- Das zentrale Element der **Geldfluss-Analyse** ist das **Sankey-Diagramm**. Die **dicken Linien** (Flows) zwischen den **Knoten** verdeutlichen, **wie viel Geld** von einer Quelle zur nächsten Ebene fließt.
- Typischerweise beginnt der Fluss links (z.B. „Root“), fließt durch **Zwischenknoten** (z.B. „Central“) und mündet rechts in den **Endknoten** (z.B. „Leaf“).
- Die Dicke einer Linie spiegelt die **Menge** des Geldflusses wider – je **dicker** die Linie, desto **größer** der Geldbetrag.
- **Mouseover:** Geht man mit der Maus über eine Linie erhält man entsprechende Informationen in einem kleinen Fenster.



- Die Linien können je nach Einstellung **frei bewegt werden**, um ein individuelles Bild zu erstellen (hierfür auf den **Start der Linien klicken** und mit gedrückter Maus verschieben).



14. Screenshot speichern / Interaktiv speichern 7

- Mit **Screenshot speichern** erstellen Sie ein statisches Bild des Diagramms. Dieses können Sie entweder unter den Falldaten speichern oder können dieses auch an andere Personen zusenden. Hier empfiehlt sich allerdings das interaktive Speichern.
- **Interaktiv speichern** legt das Diagramm als eigenständige HTML-Datei ab, die weiterhin interaktive Funktionen (z. B. Zoom oder Filter) unterstützt. So können Sie es frei verschicken, ohne dass der Empfänger selbst den AUDIPY auf dem Rechner hat. Die HTML-Datei wird ganz normal über den Browser geöffnet.

15. Bezeichnung 8

- Über dieses Dropdown kann man **alle Kategorien** gemeinsam anzeigen oder gezielt **einzelne Kategorien** auswählen.
- Wechseln Sie systematisch zwischen verschiedenen Quellkategorien, um Vergleiche zwischen Bereichen zu ermöglichen und Best Practices oder Problemzonen zu identifizieren.
- Wenn man zum Beispiel nur den Geldfluss eines bestimmten Kontos sehen möchte, wählt man dieses gezielt aus. Möchte man jedoch die **gesamten Geldströme** in einer Ansicht sehen, bleibt man bei „Alles zusammen“.
- Im Musterfall wurde der Geldfluss von verschiedenen IBAN Nr. dargestellt, daher kann man dann die einzelnen Konten auswählen. Die Farben aus der zusammengefassten Darstellung werden auch für die einzelnen Kategorien verwendet, um eine einheitliche Struktur zu haben.

Alles zusammen
DE33076811624503741360
DE44995956899847265690

16. Übersicht / Auswahl 9

- Mithilfe des blauen Buttons kann zur Analyse Übersicht mit allen möglichen Analysen gewechselt werden. Es kann aber auch das Dropdown Menü benutzt werden um zu einer anderen Analyse zu wechseln.

16.8.5 Farbkodierung und visuelle Interpretation

Das Sankey-Diagramm verwendet ein sophistiziertes Farbkodierungssystem zur Verbesserung der analytischen Erkenntnisse:

- **Rainbow-Farbpalette für Quellen:** Jede eindeutige Quelle erhält eine spezifische Farbe aus einer Rainbow-Palette, die auch bei der fokussierten Betrachtung beibehalten wird. Dies ermöglicht es Ihnen, Geldströme verschiedener Quellen auch in komplexen Diagrammen zu verfolgen.
- **Konsistente Farbzuordnung:** Die Farbzuordnungen bleiben während der gesamten Analysesitzung konsistent, sodass Sie zwischen verschiedenen Fokussierungen wechseln können, ohne die Orientierung zu verlieren.
- **Transparenz und Intensität:** Flüsse werden mit reduzierter Transparenz dargestellt, um Überlappungen sichtbar zu machen, ohne die Lesbarkeit zu beeinträchtigen. Die Intensität der Farben korreliert mit der Bedeutung der jeweiligen Quelle.

Fokussierungs-Effekte

Bei der Auswahl spezifischer Quellkategorien ändert sich das Farbschema strategisch:

- **Hervorhebung der fokussierten Quelle:** Die ausgewählte Quelle behält ihre ursprüngliche Rainbow-Farbe bei und wird damit visuell hervorgehoben.
- **Dämpfung nicht-fokussierter Elemente:** Alle anderen Quellen und deren Flüsse werden in gedämpftem Grau dargestellt, wodurch die Aufmerksamkeit auf die analysierten Ströme gelenkt wird.
- **Kontextuelle Sichtbarkeit:** Trotz der Dämpfung bleiben alle Elemente sichtbar, sodass Sie den Kontext und die relativen Proportionen verstehen können.

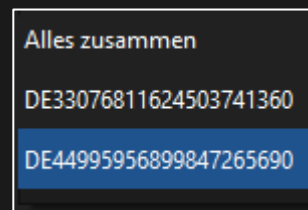
16.8.6 Erweiterte Analysestrategien und Best Practices

Erstellen Sie separate Sankey-Diagramme für verschiedene Zeiträume (Monate, Quartale, Jahre) und vergleichen Sie die Strukturveränderungen. Dies ermöglicht die Identifikation von Trends, saisonalen Mustern oder strukturellen Veränderungen in Ihren Geldflüssen.

Achten Sie auf Flüsse, die über die Zeit breiter oder schmaler werden, auf neue Verbindungen, die entstehen, oder auf etablierte Verbindungen, die verschwinden. Diese Veränderungen können auf strategische Entwicklungen, operative Verbesserungen oder potenzielle Probleme hindeuten.

Hierarchische Dekomposition

Nutzen Sie die **Kategorieauswahl** für eine systematische Dekomposition komplexer Geldflüsse. Beginnen Sie mit der Gesamtsicht, identifizieren Sie die größten Flüsse, und analysieren Sie diese dann kategorienweise im Detail. Diese Top-Down-Herangehensweise stellt sicher, dass Sie sich auf die wesentlichen Finanzströme konzentrieren.



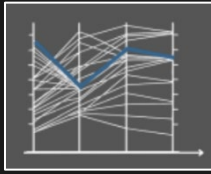
Threshold-Analyse für Fokussierung

Bei sehr **komplexen Geldfluss-Strukturen** können Sie Ihre Daten vorab filtern, um nur Flüsse oberhalb bestimmter Schwellenwerte zu berücksichtigen. Dies reduziert die visuelle Komplexität und konzentriert die Analyse auf die volumenmäßig wichtigsten Transaktionen.

Benchmarking und Vergleichsanalysen

Verwenden Sie **konsistente Farbschemata** und Layout-Einstellungen für Vergleiche zwischen verschiedenen Organisationseinheiten, Zeiträumen oder Szenarien. Die konsistente Darstellung ermöglicht es Stakeholdern, schnell Unterschiede und Gemeinsamkeiten zu erkennen.

16.9 Parallele Koordinaten



Die Analyse **paralleler Koordinaten** ist eine hochentwickelte Visualisierungsmethode zur Darstellung und Exploration multidimensionaler Datenstrukturen, die traditionelle zweidimensionale Diagrammgrenzen überwindet. Diese Technik ordnet mehrere numerische Variablen als parallele vertikale Achsen an und verbindet die Datenpunkte durch Linien, wodurch komplexe Muster, Korrelationen und Anomalien in hochdimensionalen Datenräumen sichtbar werden. Jede Linie repräsentiert dabei einen vollständigen Datensatz (z.B. ein Unternehmen, eine Kostenstelle, einen Kunden) und zeigt dessen Werte über alle analysierten Dimensionen hinweg.

In der Wirtschaftsprüfung revolutioniert die Parallel-Koordinaten-Analyse die Identifikation von Unregelmäßigkeiten und Risikomustern durch die simultane Betrachtung multipler Finanzkennzahlen. Statt einzelne Kennzahlen isoliert zu prüfen, können Sie Kombinationen von Umsatz, Gewinn, Cashflow, Verschuldungsgrad und anderen Metriken gleichzeitig analysieren. Unternehmen oder Transaktionen mit ungewöhnlichen Kennzahlenkombinationen werden als Linien sichtbar, die vom normalen Muster abweichen, was auf Manipulationen, Fehler oder besondere Geschäftsereignisse hindeuten kann.

Controller nutzen parallele Koordinaten zur ganzheitlichen Performance-Analyse und Benchmarking. Die Methode ermöglicht es, komplexe Geschäftseinheiten anhand multipler KPIs zu bewerten und dabei Trade-offs, Synergien und Optimierungspotenziale zu identifizieren. Eine Kostenstelle könnte beispielsweise niedrige Kosten aufweisen, aber gleichzeitig schlechte Qualitätskennzahlen oder lange Durchlaufzeiten – ein Muster, das nur durch die multidimensionale Betrachtung erkennbar wird.

Steuerberater verwenden die Analyse zur Bewertung komplexer Unternehmensstrukturen und zur Identifikation optimaler Steuergestaltungen. Die simultane Betrachtung von Steuerlasten, Rentabilitäten, Risikoexpositionen und regulatorischen Anforderungen verschiedener Jurisdiktionen ermöglicht eine ganzheitliche Bewertung internationaler Strukturen und die Identifikation von Win-Win-Optimierungen.

16.9.1 Warum benötigt man parallele Koordinaten?

Wie können parallele Koordinaten im beruflichen Alltag helfen:

Wirtschaftsprüfung: Multidimensionale Risikobewertung

In der **Konzernprüfung** analysieren Sie die Risikomuster verschiedener Tochtergesellschaften durch die simultane Betrachtung multipler Finanzkennzahlen. Sie konfigurieren die Analyse mit "Tochtergesellschaft" als Hauptkategorie und wählen im Aggregations-Dialog Kennzahlen wie Umsatz (Summe), Verschuldungsgrad (Mittelwert), Gewinnmarge (Mittelwert), Cashflow-Volatilität (Standardabweichung) und Working Capital Ratio (Mittelwert).

Das **resultierende Parallel-Koordinaten-Diagramm** macht Risikomuster sofort sichtbar: Gesellschaften mit ungewöhnlichen Kennzahlenkombinationen erscheinen als Linien, die vom normalen Cluster abweichen. Eine Gesellschaft mit hohem Umsatz, aber negativem Cashflow und hoher Volatilität würde eine charakteristische Linie zeigen, die weitere Prüfungshandlungen rechtfertigt.

Die **Range-Filter-Funktionalität** ermöglicht präzise Risikoanalysen: Sie können beispielsweise alle Gesellschaften identifizieren, die eine Verschuldung über 70%, aber eine Marge unter 5% aufweisen. Diese multidimensionalen Filter decken Risikokombinationen auf, die in eindimensionalen Analysen übersehen werden könnten.

Controlling: Performance-Benchmarking und Optimierungsidentifikation

Als Controller eines diversifizierten Unternehmens benchmarken Sie die Performance verschiedener Geschäftsbereiche anhand multipler KPIs. Die Konfiguration umfasst "Geschäftsbereich" als Hauptkategorie und Kennzahlen wie Umsatzwachstum (Mittelwert), EBITDA-Marge (Mittelwert), Mitarbeiterproduktivität (Mittelwert), Kundenzufriedenheit (Median) und Innovationsindex (Summe).

Das **Parallel-Koordinaten-Diagramm** deckt Trade-offs und Synergien zwischen verschiedenen Performance-Dimensionen auf. Ein Geschäftsbereich könnte hohe Margen aufweisen, aber niedrige Mitarbeiterproduktivität und Kundenzufriedenheit – ein Muster, das auf kurzfristige Gewinnoptimierung zu Lasten nachhaltiger Performance hindeutet.

Die Clustering-Funktion identifiziert automatisch Geschäftsbereiche mit ähnlichen Performance-Profilen, wodurch Best-Practice-Transfer und gemeinsame Optimierungsansätze möglich werden. Bereiche mit ausgewogenen Profilen (hohe Performance in allen Dimensionen) können als Benchmarks für andere Bereiche dienen.

Steuerberatung: Internationale Strukturoptimierung

In der **internationalen Steuerberatung** optimieren Sie komplexe Holdingstrukturen durch die multidimensionale Analyse verschiedener Jurisdiktionen. Sie analysieren "Land" als Hauptkategorie mit Kennzahlen wie effektive Steuerbelastung (Mittelwert), regulatorische Komplexität (Index), politische Stabilität (Rating), Doppelbesteuerungsabkommen (Anzahl) und Compliance-Kosten (Summe).

Das **Parallel-Koordinaten-Diagramm** visualisiert die multidimensionalen Trade-offs zwischen Steueroptimierung und anderen Faktoren. Ein Land könnte niedrige Steuern bieten, aber hohe regulatorische Komplexität und Compliance-Kosten aufweisen. Die Visualisierung hilft bei der Identifikation optimaler Jurisdiktionen, die ein ausgewogenes Verhältnis zwischen allen relevanten Faktoren bieten.

Die **Range-Filter** ermöglichen präzise Suchkriterien: Sie können beispielsweise Länder identifizieren, die eine **Steuerbelastung unter 15%, aber gleichzeitig eine politische Stabilität über 8/10 und weniger als 50 Compliance-Anforderungen** aufweisen. Diese multidimensionale Filterung unterstützt evidenzbasierte Beratung und Strukturoptimierung.

16.9.2 Erstellung der parallelen Koordinaten

Bevor Sie mit der Konfiguration beginnen, definieren Sie klar Ihre Analyseziele und die zu vergleichenden Entitäten. Überlegen Sie sich, welche Kombinationen von Kennzahlen Ihnen strategische Erkenntnisse liefern könnten und welche Trade-offs oder Synergien Sie identifizieren möchten.

Parallel-Koordinaten-Analysen sind besonders wertvoll für:

- Multi-Kriterien-Bewertungen und Ranking-Analysen
- Identifikation von Ausreißern in mehrdimensionalen Räumen
- Cluster-Analysen und Segmentierung
- Korrelations- und Abhängigkeitsanalysen zwischen Kennzahlen



Hauptkategorie & Zusatzkategorie ①

- Wählen Sie die **Hauptkategorie** basierend auf der Entitätsebene, die Sie analysieren möchten. Für strategische Analysen wählen Sie übergeordnete Kategorien wie Geschäftsbereiche oder Tochtergesellschaften. Für operative Analysen fokussieren Sie auf Kostenstellen, Projekte oder Prozesse.

- Konfigurieren Sie die **Zusatzkategorie** nur dann, wenn Sie eine hierarchische Analyse benötigen oder verschiedene Aggregationsebenen vergleichen möchten. Eine zu granulare Kategorisierung kann zu überkomplexen Visualisierungen führen.

Kennzahlen-Portfolio im Aggregations-Dialog ②

Summe
Mittelwert
Minimum
Maximum
Median
Anzahl
Anzahl eindeutig
Standardabweichung
Varianz
Modus
1st Quartile (INC)
1st Quartile (EXC)
3rd Quartile (INC)
3rd Quartile (EXC)

- Öffnen Sie den **Aggregations-Dialog** und wählen Sie systematisch die numerischen Dimensionen Ihrer Analyse aus. Eine **ausgewogene Kennzahlen-Kombination** sollte **verschiedene Geschäftsaspekte** abdecken:
 - **Performance-Kennzahlen:** Umsatz, Gewinn, Produktivität für Ergebnisbewertung
 - **Effizienz-Kennzahlen:** Margen, Umschlagshäufigkeiten, Durchlaufzeiten für Prozessqualität
 - **Risiko-Kennzahlen:** Schwankungen, Verschuldungsgrade, Konzentrationen für Risikoeinschätzung
 - **Struktur-Kennzahlen:** Größen, Anteile, Verhältnisse für Strukturanalysen
- Vermeiden Sie zu viele ähnliche Kennzahlen, da diese das Diagramm unübersichtlich machen ohne zusätzliche Erkenntnisse zu liefern. 4-8 gut gewählte Achsen sind oft optimal für aussagekräftige Analysen.

Wähle numerische Spalten und ihre Aggregation:

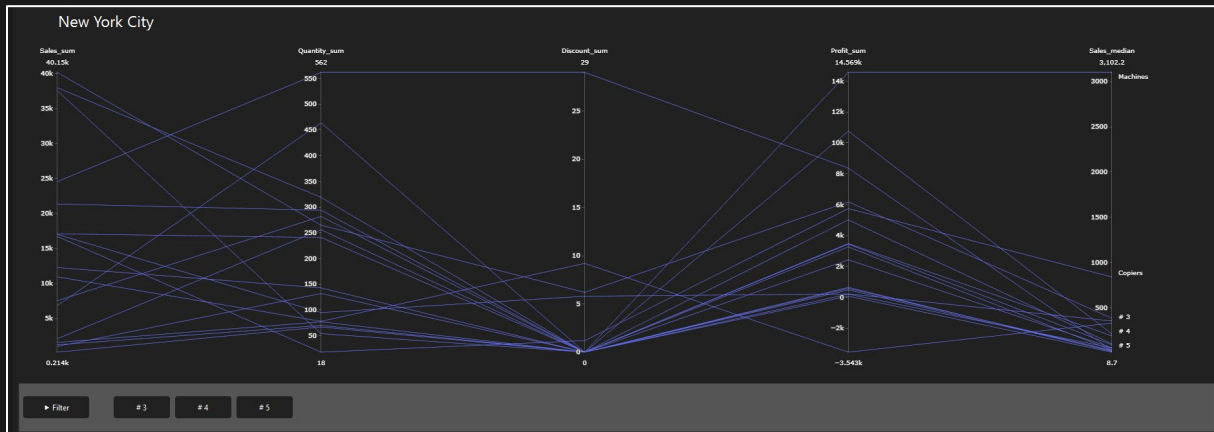
Verfügbare Spalten: Row ID

☒	Sales	Summe
☒	Quantity	Summe
☒	Discount	Summe
☒	Profit	Summe
☒	Sales	Median

Erste Visualisierung und Muster-Identifikation ③

- Starten Sie die Analyse mit der **"Alle zusammen"**-Einstellung ④, um einen Gesamtüberblick über alle Entitäten zu erhalten. Das resultierende Diagramm zeigt jede Entität als farbkodierte Linie, die alle Achsen verbindet. Nutzen Sie die Kategorieauswahl für systematische Detailanalysen spezifischer Entitäten oder Gruppen. Die interaktiven Filterfunktionen ermöglichen es Ihnen, sich auf relevante Teilmengen zu konzentrieren und dabei den Kontext der Gesamtanalyse zu behalten.

Beispiel: Filterung nach einer bestimmten Stadt (bspw. New York City)



Achten Sie auf grundlegende Muster:

- **Parallele Linien:** Deuten auf ähnliche Profile und konsistente Geschäftsmodelle hin
- **Kreuzende Linien:** Zeigen Trade-offs oder verschiedene Strategieansätze
- **Ausreißer-Linien:** Kennzeichnen außergewöhnliche Entitäten, die weitere Analyse verdienen
- **Cluster-Bildung:** Gruppierungen ähnlicher Entitäten mit vergleichbaren Profilen

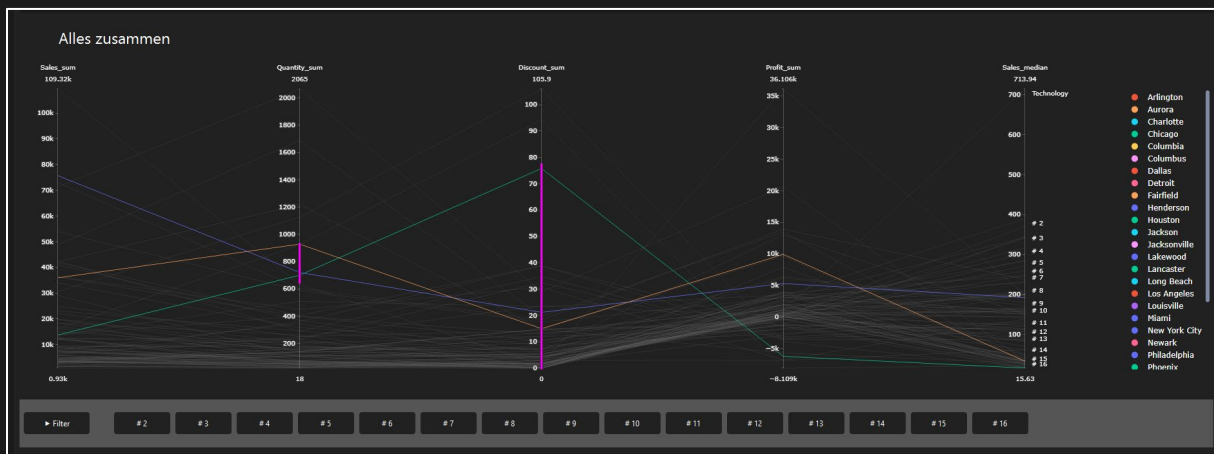
Fokussierte Detailanalyse mit Filterung 5 & 6

- Das Parallel-Koordinaten-System bietet umfangreiche Interaktionsmöglichkeiten zur Optimierung der Analyse:
 - **Achsen-Neuordnung:** Durch Drag-and-Drop können Sie die Reihenfolge der Achsen ändern, um verschiedene Korrelationsmuster sichtbar zu machen. Benachbarte Achsen zeigen Korrelationen deutlicher als entfernte Achsen.

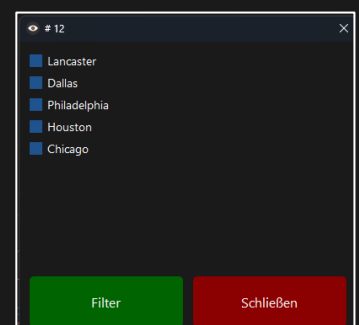


- **Achsen-Skalierung:** Jede Achse kann individuell skaliert werden, um Unterschiede zwischen Kennzahlen verschiedener Größenordnungen auszugleichen und alle Dimensionen gleichwertig darzustellen.

- **Dynamische Filterung:** Durch das Ziehen von Bereichen auf einzelnen Achsen können Sie interaktiv Teilmengen der Daten auswählen und fokussieren. Diese Range-Filter ermöglichen es, spezifische Kombinationen von Kennzahlenwerten zu isolieren.



- Das System identifiziert automatisch Entitäten mit ähnlichen Kennzahlenprofilen und bietet erweiterte Gruppierungsfunktionen:
 - **Automatische Cluster-Erkennung:** Entitäten mit ähnlichen Linienverläufen werden automatisch gruppiert und können als Sets analysiert werden. Dies ist besonders wertvoll für die Identifikation von Geschäftsmodell-Typen oder Risikoprofilen.
 - **Interaktive Gruppen-Buttons:** Das System generiert automatisch Buttons für identifizierte Cluster, die eine schnelle Fokussierung auf spezifische Entitätsgruppen ermöglichen. Dies ist besonders nützlich bei komplexen Analysen mit vielen Entitäten.
 - **Hierarchische Filterung:** Die Kombination aus Achsen-Filtern und Kategorie-Gruppierung ermöglicht mehrstufige Filterprozesse, bei denen Sie zunächst nach Kennzahlen-Kriterien filtern und anschließend nach kategorialen Merkmalen segmentieren können.



- Die **Range-Filter-Funktionalität** ermöglicht präzise multidimensionale Selektionen:
 - **Multi-Achsen-Filter:** Sie können gleichzeitig auf mehreren Achsen Bereiche definieren und dadurch sehr spezifische Kombinationen von Kennzahlenwerten identifizieren. Beispielsweise könnten Sie nach Unternehmen suchen, die hohe Umsätze (>1M), niedrige Kosten (<500K) und mittlere Risikokennzahlen (0.2-0.4) aufweisen.
 - **Logische Verknüpfung:** Mehrere Range-Filter werden automatisch mit UND-Logik verknüpft, wodurch präzise multidimensionale Selektionen möglich werden.
 - **Dynamische Rückkopplung:** Gefilterte Entitäten werden sofort in der Visualisierung hervorgehoben, während nicht-gefilterte Entitäten gedämpft dargestellt werden. Dies ermöglicht die sofortige visuelle Bewertung der Filtereffekte.

Gefilterte Daten / Screenshot speichern / Interaktiv speichern 7

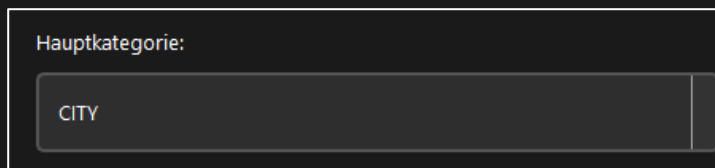
- Über „**Gefilterte Daten anzeigen**“ können Sie gezielt nur die als Ausreißer erkannten Datenpunkte ein- oder ausblenden. Damit lassen sich **verdächtige Werte** schnell isolieren und genauer untersuchen.
- Mit **Screenshot speichern** erstellen Sie ein statisches Bild des Diagramms. Dieses können Sie entweder unter den Falldaten speichern oder können dieses auch an andere Personen zusenden. Hier empfiehlt sich allerdings das interaktive Speichern.
- **Interaktiv speichern** legt das Diagramm als eigenständige HTML-Datei ab, die weiterhin interaktive Funktionen (z.B. Zoom oder Filter) unterstützt. So können Sie es frei verschicken, ohne dass der Empfänger selbst den AUDIPY auf dem Rechner hat. Die HTML-Datei wird ganz normal über den Browser geöffnet.

Übersicht / Auswahl 8

- Mithilfe des **blauen Buttons** kann zur Analyse Übersicht mit allen möglichen Analysen gewechselt werden. Es kann aber auch das Dropdown Menü benutzt werden, um zu einer anderen Analyse zu wechseln.

16.9.3 Bedienelemente und Datenstruktur-Konfiguration

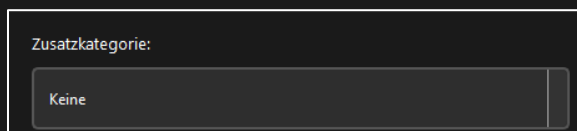
Das **Hauptkategorie-Feld** bestimmt die primäre Gruppierungsebene Ihrer multidimensionalen Analyse und definiert, welche Entitäten als individuelle Linien im Koordinatensystem dargestellt werden. Diese Auswahl ist fundamental, da sie bestimmt, was als "Datenpunkt" in Ihrer Analyse behandelt wird.

The image shows a user interface element for selecting a main category. It consists of a rectangular box with a light gray background. Inside the box, at the top left, is the label 'Hauptkategorie:'. Below this label is a horizontal scrollable list. The first item in the list is 'CITY', which is highlighted with a darker background. To the right of the list is a small, light gray rectangular button.

Verfügbare Optionen umfassen alle Textspalten Ihres Datensatzes. Strategische Überlegungen zur Auswahl:

Unternehmensebene: "Mandant", "Tochtergesellschaft", "Geschäftsbereich" für Konzern- und Portfolio-Analysen **Operative Ebene:** "Kostenstelle", "Projekt", "Produktlinie" für interne Performance-Vergleiche **Transaktionsebene:** "Lieferant", "Kunde", "Vertrag" für Beziehungs- und Risiko-Analysen **Zeitliche Ebene:** "Monat", "Quartal", "Jahr" für Trend- und Entwicklungsanalysen

Das **Zusatzkategorie-Feld** ermöglicht eine hierarchische Strukturierung Ihrer Analyse durch die Definition einer zweiten Gruppierungsebene. Diese erweiterte Kategorisierung ist besonders wertvoll bei komplexen Organisationsstrukturen oder wenn Sie verschiedene Aggregationsebenen vergleichen möchten.

The image shows a user interface element for selecting an additional category. It consists of a rectangular box with a light gray background. Inside the box, at the top left, is the label 'Zusatzkategorie:'. Below this label is a horizontal scrollable list. The first item in the list is 'Keine', which is highlighted with a darker background. To the right of the list is a small, light gray rectangular button.

Verfügbare Optionen:

- **"Keine"** - Verzichtet auf hierarchische Strukturierung für einfache Analysen
- **Alle Textspalten** des Datensatzes - Ermöglicht die Kombination zweier kategorialer Dimensionen

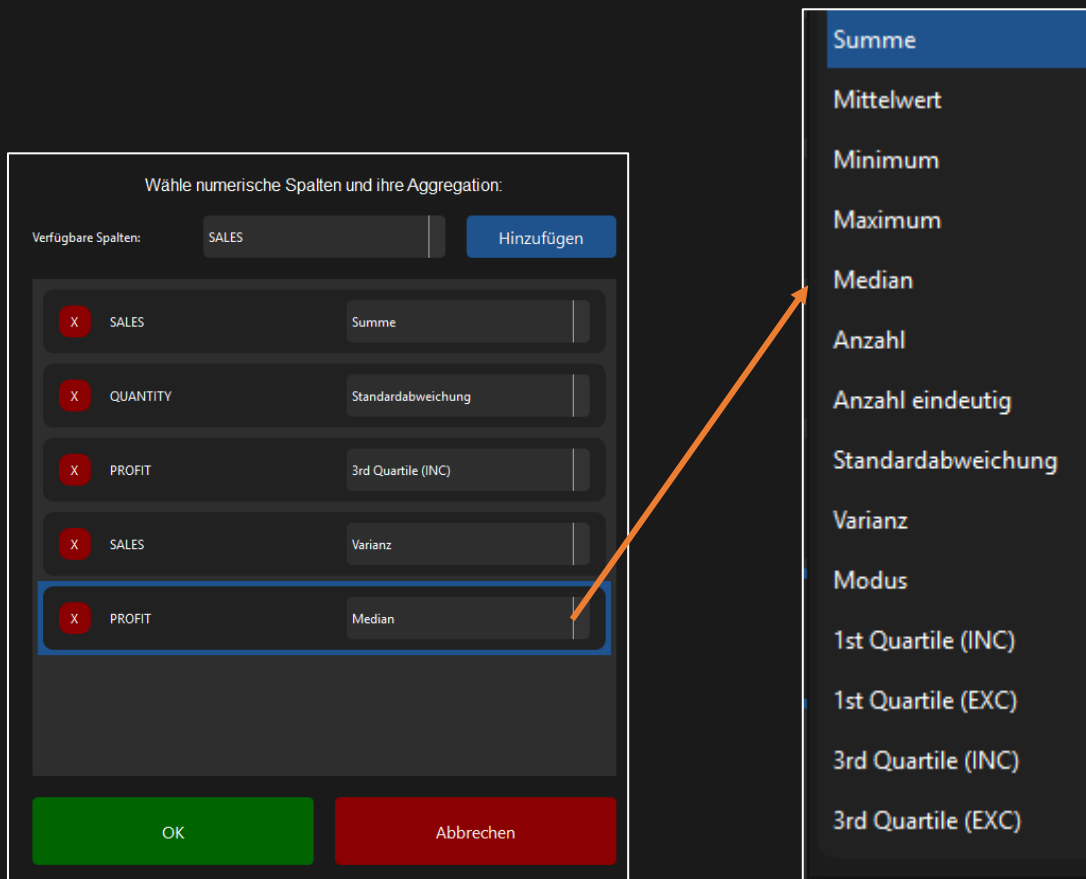
Die **Kombination von Haupt- und Zusatzkategorie erstellt eine Zwei-Ebenen-Hierarchie**, bei der jede Kombination der beiden Kategorien als separate Linie dargestellt wird. Dies ermöglicht beispielsweise die simultane Analyse von **"Abteilung × Kostenstelle"** oder **"Region × Produktgruppe"**, wodurch sowohl übergeordnete als auch detaillierte Muster erkennbar werden.

Aggregations-Dialog und Kennzahlen-Definition

Der **Aggregations-Button** öffnet einen spezialisierten Dialog zur Definition der numerischen Achsen Ihres Parallel-Koordinaten-Systems. Diese Konfiguration ist das

Herzstück der Analyse, da sie bestimmt, welche Kennzahlen verglichen werden und wie sie aus Ihren Rohdaten berechnet werden.

Der Aggregations-Dialog bietet für jede numerische Spalte verschiedene statistische Funktionen:



Zentrale Tendenz-Maße

- **Summe:** Gesamtvolumen oder kumulative Werte für Volumensvergleiche
- **Mittelwert:** Durchschnittliche Performance oder typische Werte für Effizienzanalysen
- **Median:** Robuste zentrale Tendenz, weniger anfällig für Ausreißer

Streuungs- und Risiko-Maße

- **Standardabweichung:** Variabilität und Risikobewertung für Stabilitätsanalysen
- **Varianz:** Quadrierte Abweichungen für mathematische Analysen
- **Min/Max:** Extreme Werte für Bandbreiten-Betrachtungen

Erweiterte statistische Maße

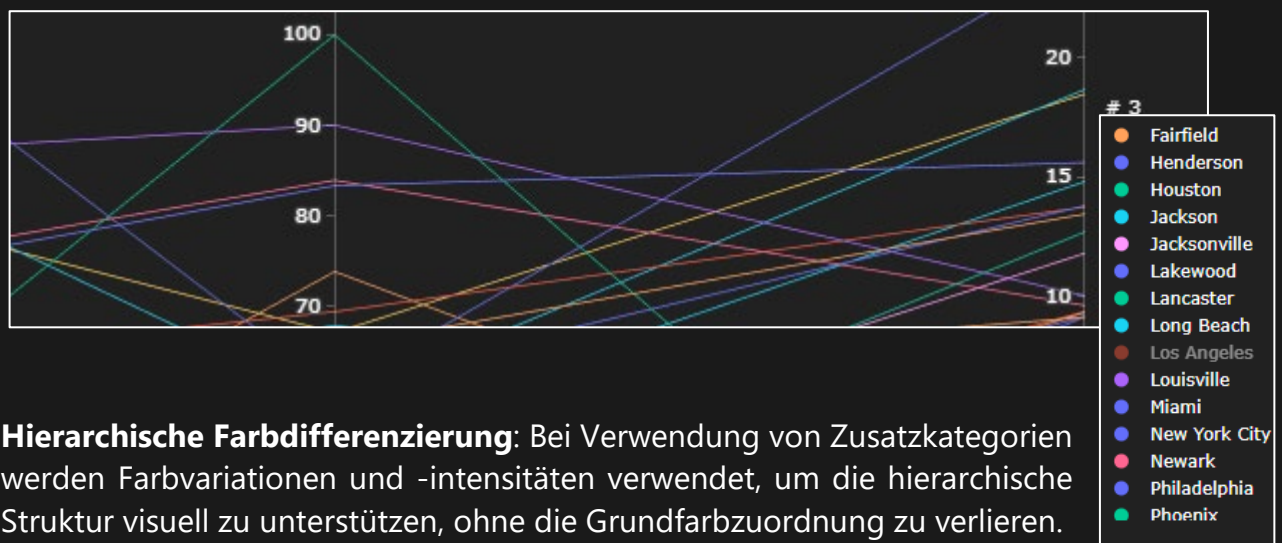
- **Quartile (Q1/Q3):** Verschiedene Berechnungsmethoden (Excel-kompatibel) für Verteilungsanalysen
- **Modus:** Häufigster Wert für kategorische Kennzahlen
- **Anzahl:** Häufigkeiten und Volumen-Indikatoren

Die Flexibilität des Aggregations-Dialogs ermöglicht es Ihnen, für verschiedene Spalten unterschiedliche Funktionen zu wählen. Sie könnten beispielsweise für Umsätze die Summe verwenden, für Margen den Mittelwert, für Risikokennzahlen die Standardabweichung und für Qualitätsmessungen den Median.

16.9.4 Farbkodierung und visuelle Interpretation

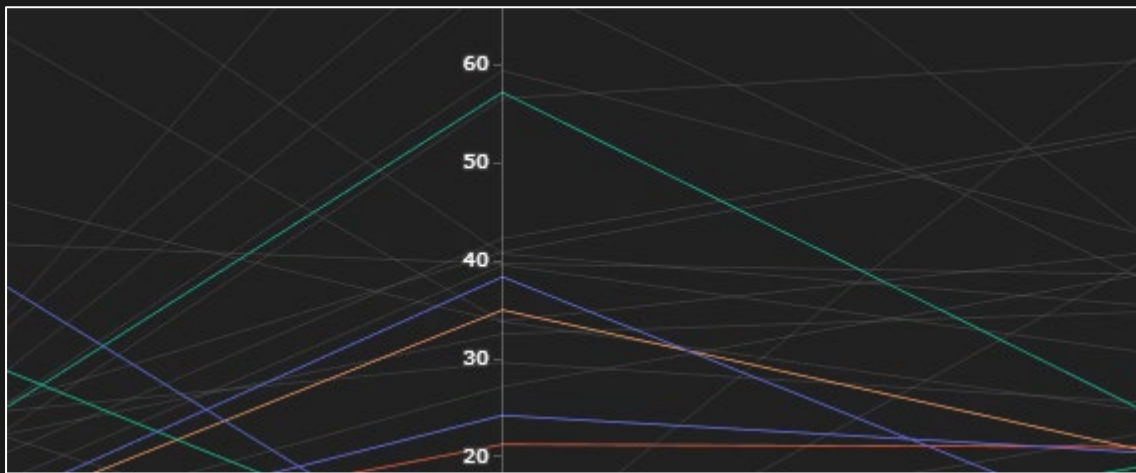
Das Parallel-Koordinaten-System verwendet ein sophistiziertes Farbkodierungssystem zur Verbesserung der analytischen Erkenntnisse:

Kategoriebasierte Farbpalette: Jede Hauptkategorie erhält eine eindeutige Farbe aus einer optimierten Palette, die auch bei vielen Kategorien unterscheidbar bleibt. Die Farbzuoordnung bleibt während der gesamten Analysesitzung konsistent.



Hierarchische Farbdifferenzierung: Bei Verwendung von Zusatzkategorien werden Farbvariationen und -intensitäten verwendet, um die hierarchische Struktur visuell zu unterstützen, ohne die Grundfarbzuoordnung zu verlieren.

Fokussierungs-Effekte: Bei der Auswahl spezifischer Kategorien werden relevante Linien hervorgehoben, während andere gedämpft dargestellt werden. Dies reduziert visuelle Komplexität ohne Informationsverlust.



Linieninterpretation und Muster-Erkennung

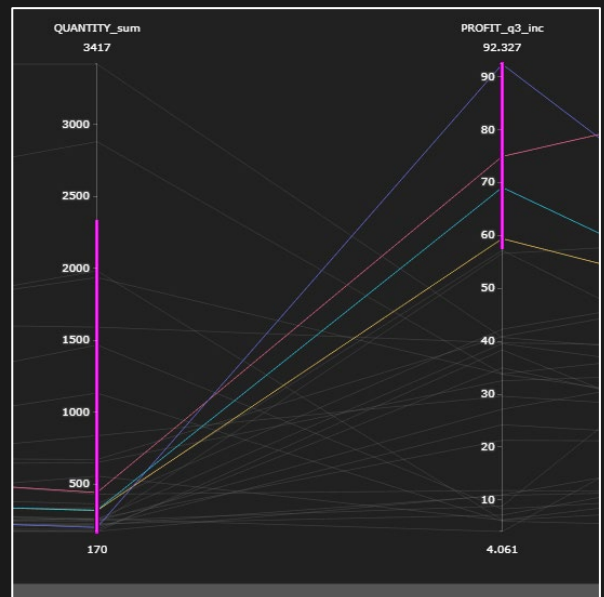
Die Interpretation der Linienverläufe folgt etablierten Prinzipien der multidimensionalen Datenanalyse:

Steile Anstiege/Abfälle: Große Änderungen zwischen benachbarten Achsen deuten auf starke Variationen oder Trade-offs zwischen den entsprechenden Kennzahlen hin.

Parallele Segmente: Ähnliche Steigungen zwischen denselben Achsen bei verschiedenen Entitäten deuten auf systematische Zusammenhänge oder gemeinsame Einflussfaktoren hin.

Kreuzungsmuster: Linien, die sich zwischen Achsen kreuzen, zeigen verschiedene Strategieansätze oder Optimierungsrichtungen auf und sind oft besonders interessant für Benchmarking-Analysen.

Cluster-Formationen: Bündel von Linien, die ähnliche Verläufe zeigen, identifizieren Entitäten mit vergleichbaren Profilen und können auf gemeinsame Erfolgsrezepte oder Problemstellungen hinweisen.



16.9.5 Erweiterte Analysestrategien und methodische Vertiefung

Nutzen Sie die Achsen-Neuordnung für Sequential Pattern Mining, bei dem Sie die Achsen in logischer oder zeitlicher Reihenfolge anordnen und dabei Entwicklungsmuster oder Ursache-Wirkungs-Ketten identifizieren. Dies ist besonders wertvoll für Prozessanalysen und Performance-Entwicklungen.

Sensitivity Analysis durch Achsen-Variation

Experimentieren Sie systematisch mit verschiedenen Achsen-Kombinationen und Aggregationsfunktionen, um die Robustheit Ihrer Erkenntnisse zu validieren. Konsistente Muster über verschiedene Kennzahlen-Kombinationen hinweg sind vertrauenswürdiger als isolierte Beobachtungen.

Comparative Benchmarking

Erstellen Sie mehrere Parallel-Koordinaten-Analysen für verschiedene Zeiträume, Märkte oder Organisationseinheiten und vergleichen Sie die resultierenden Muster. Dies ermöglicht die Identifikation zeitlicher Trends, regionaler Unterschiede oder struktureller Veränderungen.

Correlation Discovery durch Achsen-Proximität

Platzieren Sie vermutlich korrelierte Kennzahlen als benachbarte Achsen, um Korrelationsmuster durch die Parallelität oder Divergenz der Linien visuell zu bewerten. Diese visuelle Korrelationsanalyse kann quantitative Analysen ergänzen und neue Hypothesen generieren.

Integration in analytische Workflows

Filtering und Datenexploration

Die gefilterten Ergebnisse aus der Parallel-Koordinaten-Analyse können direkt in andere Analysewerkzeuge übertragen werden. Der "**Gefilterte Daten anzeigen**"-Button ermöglicht die nahtlose Integration in tabellarische Analysen oder andere Visualisierungen.

Hypothesis Generation und Testing

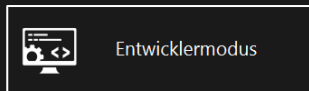
Nutzen Sie die Parallel-Koordinaten-Analyse zur Generierung von Hypothesen über multidimensionale Zusammenhänge und validieren Sie diese anschließend mit spezialisierten statistischen Tests oder Detailanalysen.

Progressive Disclosure

Beginnen Sie mit einer groben Übersicht und verfeinern Sie systematisch durch Filterung und Fokussierung. Diese **Progressive-Disclosure-Strategie** ermöglicht es, auch in sehr komplexen Datenstrukturen relevante Muster zu identifizieren, ohne von der Komplexität überwältigt zu werden.

Die **Parallel-Koordinaten-Analyse** stellt somit ein mächtiges Werkzeug für die multidimensionale Datenexploration dar, das durch seine Kombination aus visueller Intuition, statistischer Rigorosität und interaktiver Flexibilität optimal geeignet ist, komplexe Geschäftsmuster zu identifizieren und fundierte Entscheidungen auf Basis multikriterieller Bewertungen zu treffen.

17 Entwicklermodus

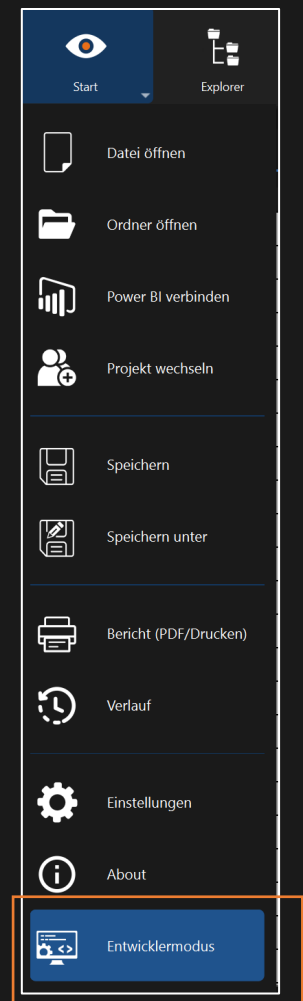


Der Entwicklermodus stellt spezielle Funktionen und Werkzeuge bereit, die über den normalen Funktionsumfang der Anwendung hinausgehen, aus diesem Grund wurde ein separates Kapitel für diesen Modus erstellt.

Es richtet sich in erster Linie an fortgeschrittene Nutzer und Entwickler, die tiefere Einblicke in interne Abläufe benötigen oder erweiterte Einstellungen vornehmen möchten. Um den normalen Anwender nicht mit dieser Modus zu konfrontieren, **muss dieser erst einmal aktiviert** werden (s. Kapitel **3.9.1.3**). **Wichtig:** Das Programm muss nach Aktivierung neu gestartet werden!

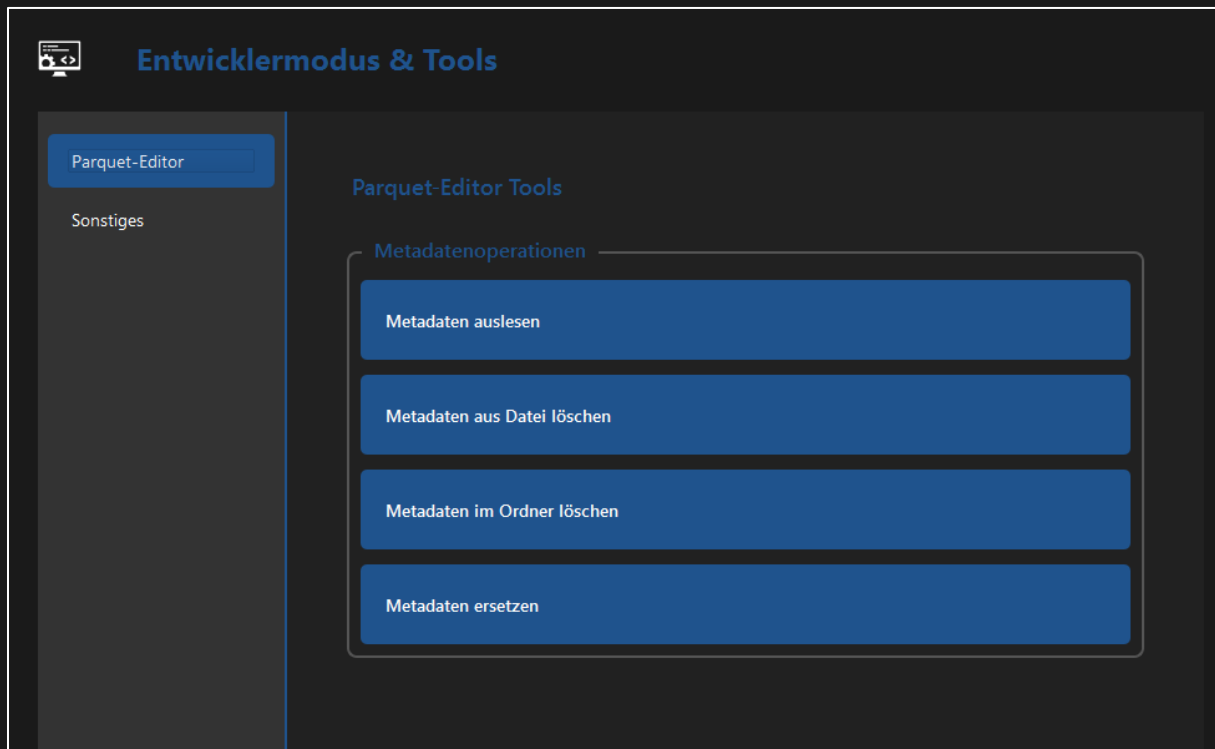
Nach Neustart kann der Entwicklermodus über „**Start**“ ganz unten aufgerufen werden:

Nach Aufruf erhalten Sie ein neues Dialogfenster.



17.1 Parquet Editor

Im Bereich „**Parquet-Editor**“ stehen Funktionen zur Verfügung, um Metadaten in Parquet-Dateien gezielt zu verwalten. Diese Werkzeuge ermöglichen sowohl die Auswertung als auch das gezielte Löschen oder Ersetzen von Metadaten innerhalb einzelner Dateien oder ganzer Ordnerstrukturen.



17.1.1 Metadaten auslesen:

Parquet-Dateien enthalten nicht nur strukturierte Daten, sondern auch sogenannte Metadaten, die zusätzliche Informationen über den Aufbau, Inhalt und Ursprung der Datei liefern. Diese Metadaten sind insbesondere bei der Datenprüfung, Analyseautomatisierung oder beim Debugging hilfreich, da sie beispielsweise das Schema, Datentypen, Komprimierungsarten oder Ursprungssysteme dokumentieren.

Nach Auswahl der entsprechenden Parquet Datei werden die Metadaten ausgelesen und hat man hat anschließend folgende Ansicht:

i Benutzerdefinierte Metadaten

Metadaten erfolgreich ausgelesen:


```

{
  "user_defined_metadata": {
    "Original_Columns": {
      "Row ID": "int64",
      "Order ID": "object",
      "Order Date": "datetime64[ns]",
      "Ship Date": "datetime64[ns]",
      "Ship Mode": "object",
      "Customer ID": "object",
      "Customer Name": "object",
      "Segment": "object",
      "Country/Region": "object",
      "City": "object",
      "State": "object",
      "Postal Code": "float64",
      "Region": "object",
      "Product ID": "object",
      "Category": "object",
      "Sub-Category": "object",
      "Product Name": "object",
      "Sales": "float64",
      "Quantity": "int64",
      "Discount": "float64",
      "Profit": "float64"
    }
  }
}
        
```

Speichern
 Schließen

i Erfolg

Datei erfolgreich gespeichert.

OK

Speicherort aus öffnen.

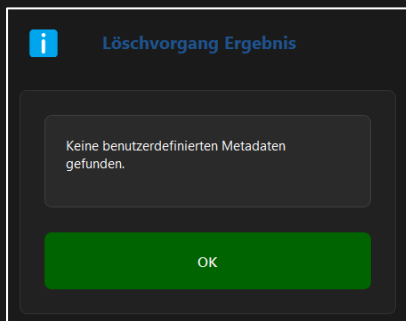
Nach erfolgreicher Speicherung einer **JSON Datei**, kann man diese per Doppelklick vom

Datei	Bearbeiten	Ansicht
<pre> { "user_defined_metadata": { "Original_Columns": { "Row ID": "int64", "Order ID": "object", "Order Date": "datetime64[ns]", "Ship Date": "datetime64[ns]", "Ship Mode": "object", "Customer ID": "object", "Customer Name": "object", "Segment": "object", "Country/Region": "object", "City": "object", "State": "object", "Postal Code": "float64", "Region": "object", "Product ID": "object", "Category": "object", "Sub-Category": "object", "Product Name": "object", "Sales": "float64", "Quantity": "int64", "Discount": "float64", "Profit": "float64" } }, "column_user": { "Vorname": { "info": "user_column" } } } </pre>		

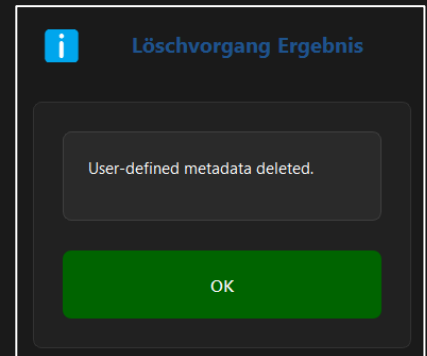
17.1.2 Metadaten aus Datei löschen:

Entfernt die Metadaten gezielt aus einer einzelnen Parquet-Datei, ohne die eigentlichen Dateninhalte zu verändern.

Nach Auswahl der Parquet Datei wird vom Programm die gesamten Metadaten aus der Datei gelöscht und man erhält das folgende Programmfenster:



Sollten keine Metadaten in der ausgewählten Datei vorhanden sein, bekommen Sie folgendes Fenster angezeigt:



17.1.3 Metadaten im Ordner löschen:

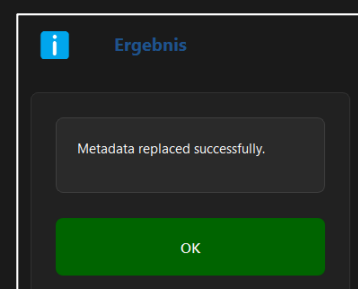
Löscht die Metadaten aus **allen Parquet-Dateien** in einem angegebenen Ordner. Ideal zur Bereinigung größerer Datenmengen in einem Schritt. Die Vorgehensweise ist wie im Punkt „**Metadaten aus Datei löschen**“ beschrieben.

17.1.4 Metadaten ersetzen:

Überschreibt bestehende Metadaten in einer Datei mit neuen, vordefinierten Metadaten. Diese Funktion erlaubt eine gezielte Aktualisierung oder Standardisierung von Metainformationen.

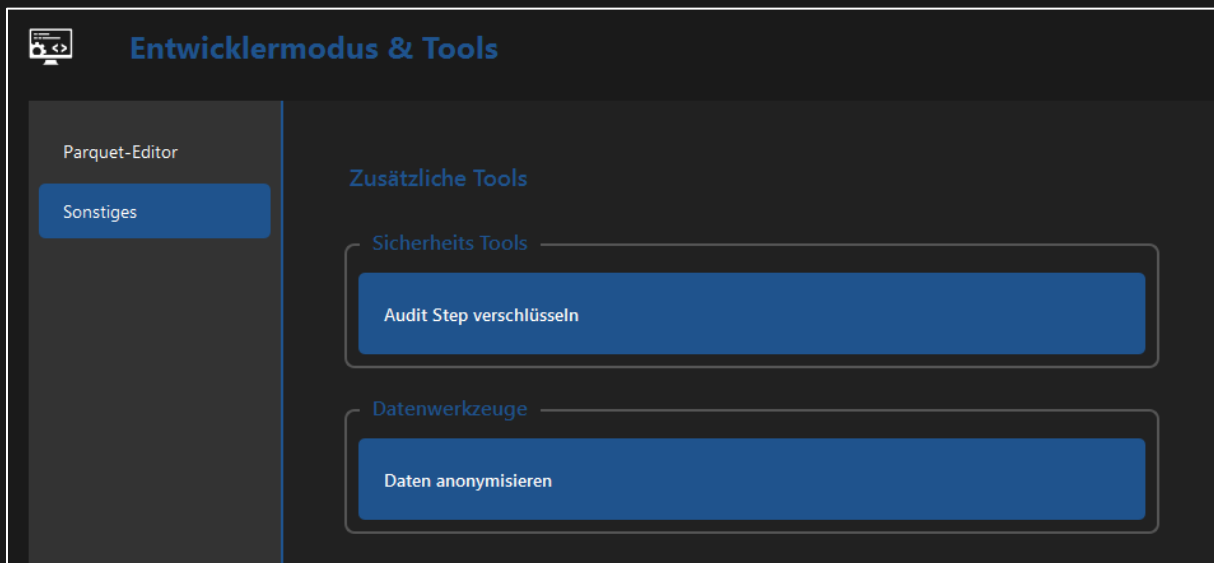
Bei der Funktion „**Metadaten ersetzen**“ wählen Sie zunächst die JSON-Datei aus, welche die neuen Metadaten enthält.

Anschließend wählen Sie die Parquet-Datei, in der die zuvor ausgewählte JSON-Datei als neue Metadaten eingefügt werden soll. Der eigentliche Austausch der Metadaten erfolgt anschließend automatisch durch das Programm, ohne dass weitere Eingaben erforderlich sind. Mit folgendem Fenster wird der „Austausch“ bestätigt und der Vorgang ist erledigt:



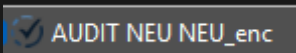
17.2 Sonstiges

Unter der Rubrik „Sonstiges“ kann man die **AUDIT Steps (9)** verschlüsseln.



17.2.1 Sicherheitstool: Verschlüsselung

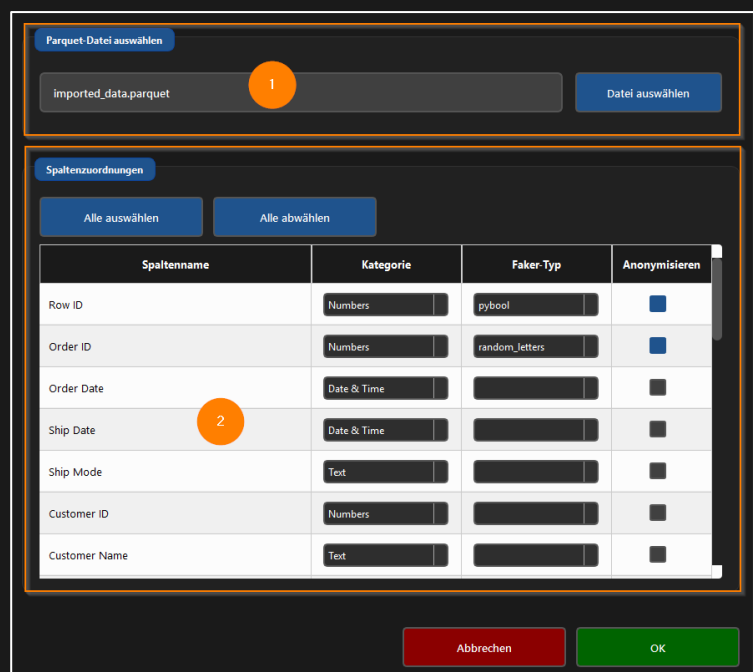
Um die **Verschlüsselung** durchzuführen, klicken Sie einfach auf „**Audit Step verschlüsseln**“. Wählen Sie anschließend die gewünschte Audit-Datei aus und legen Sie den Speicherort für die neu erstellte, verschlüsselte Datei mit der Endung **.enc** fest. Danach ist die Verschlüsselung abgeschlossen.



17.2.2 Datenwerkzeuge: Anonymisierung

Die **Anonymisierungsfunktion** dient dazu, sensible oder personenbezogene Daten in einer Parquet-Datei unkenntlich zu machen, ohne dabei die Struktur oder Auswertbarkeit der Daten zu verlieren. Sie ist insbesondere für datenschutzkritische Szenarien relevant – etwa bei der Weitergabe von Daten zu Test-, Analyse- oder Schulungszwecken.

Nach Auswahl einer **Parquet-Datei** **1** kann der Nutzer gezielt



festlegen, welche Spalten anonymisiert werden sollen. Dies erfolgt über folgende **Einstellmöglichkeiten 2**:

- **Spaltenauswahl:**

In der Tabelle werden alle Spalten der Datei aufgelistet. Über die Checkbox in der Spalte „**Anonymisieren**“ wird gesteuert, ob eine Spalte anonymisiert werden soll.

- **Kategorie & Faker-Typ:**

Für jede Spalte kann eine **Datenkategorie** (z.B. Datum, Person, Organisation) sowie ein entsprechender **Faker-Typ** ausgewählt werden. Der Faker-Typ bestimmt, **wie** die Werte ersetzt werden – z. B. durch zufällige Namen, zufällige Zahlen, Datumswerte oder strukturgetreue Platzhalter.

Kategorie: Faker-Typ: unterschiedlich je nach Kategorie (z.B. Nummern)

Personal	boolean
Address	random_digit
Internet	random_digit_not_null
Company	random_element
Job	random_int
Phone	random_letter
Date & Time	random_letters
Text	random_lowercase_letter
Color	random_number
Credit Card	random_uppercase_letter
Currency	pybool
Bank	pydecimal
Barcode	pyfloat
ISBN	pyint
File	pystr
Automotive	pystr_format
SSN	
UUID	
Numbers	
Profile	
User Agent	
Geo	
Emoji	
Misc	

Im Idealfall sollten für die Anonymisierung reine Textspalten ausgewählt werden.

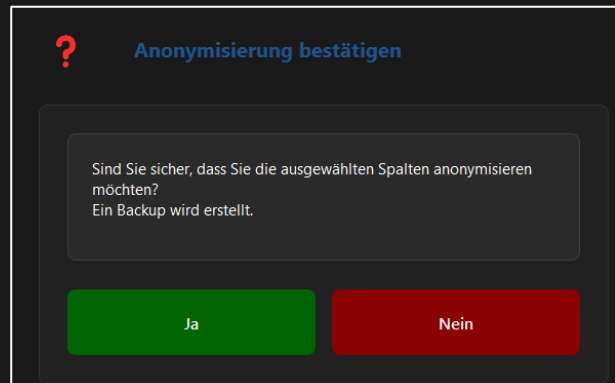
- **Beispiel:**

- Eine Transaktionsnummer kann durch eine Zufallsfolge ersetzt werden.
- Ein Datum bleibt im Format erhalten, aber mit verändertem Wert.
- Eine Kostenstelle oder ein Kundenschlüssel wird durch synthetische IDs ersetzt.

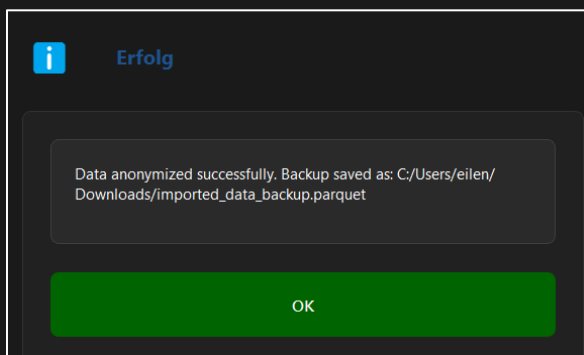
Nach Auswahl der gewünschten Spalten kann der Anonymisierungsprozess mit einem Klick auf „**OK**“ gestartet werden.

Vor der tatsächlichen Durchführung erscheint ein **Sicherheitsdialog**, der zur finalen Bestätigung auffordert und darauf hinweist, dass eine Anonymisierung vorgenommen wird.

Zusätzlich wird automatisch ein **Backup der Originaldatei** erstellt, um die Daten im Bedarfsfall wiederherstellen zu können.



Die Anonymisierung wird mit dem folgenden Fenster bestätigt:



Anschließend kann die anonymisierte Datei aus dem Speicherpfad geöffnet werden.

Sie haben weitere Fragen oder Anregungen? Dann kontaktieren Sie uns unter info@semitax.de. Gerne helfen wir Ihnen weiter.

AUDIPLY

REALTIME AUDIT. WITH AI.

