

KONSISTENTE DATEN - STAMMDATENPFLEGE - STEIGERN SIE IHRE DATENQUALITÄT!



WAS BIETEN WIR?

✓ Beratung bei der Konsolidierung des bestehenden Datenbestands

✓ Identifikation und Aufzeigen von Inkonsistenzen in den Daten

- Fehlende Einträge
Postleitzahl, aber kein Ort
- Mehrfach vorhandene Einträge
Kunde XY-AG ist doppelt angelegt
- Fehlerhafte Formatierungen in Namen und Adressen
*0134/324 234 34 vs. 0134 - 32423434 vs +49 134 32423434,
Max Mustermann vs. Mustermann, Max*
- Probleme bei Umlauten und Sonderzeichen
Müller und MÃ¼ller || Mueller und Müller
- Unterschiedliche Schreibweisen
*MUELLER und Müller (klein vs. groß)
Straße und Strasse (ß vs. ss) und Str. (Abkürzung)
GRADC und °C*

✓ Automatisiertes Korrigieren, Vereinheitlichen, Konsolidieren und Vervollständigen von Daten nach abgesprochenen Regeln

✓ Datenbasierte Unterstützung bei der Identifikation von Anomalien in den Daten

- Ausreißer erkennen
Ungewöhnlich hoher Wert in einer Spalte: 100.123 statt 100,123 (z.B. durch Buchungs- oder Eingabefehler bei der Eingabe)
- Zusammenhänge in den Daten identifizieren und diese auszunutzen, um Unregelmäßigkeiten zu erkennen
Wenn in Spalte A ein Eintrag steht, so gibt es für gewöhnlich auch in Spalte D einen Eintrag bis auf wenige Ausnahmen. Fehlender Eintrag?

Eine hohe Datenqualität ist die Grundvoraussetzung, dass Sie Mehrwert aus Ihren Daten generieren und eine optimale Grundlage für Analytics und AI Anwendungen schaffen. Eine gute Datenqualität ist die Basis für maschinelles Lernen und der Schlüsselfaktor, um aus den Daten sinnvolle Muster zu erlernen und verwertbare Ergebnisse zu erzeugen.

Deshalb unterstützen wir sehr gerne dabei, Ihre Daten hierfür unter die Lupe zu nehmen und fit zu machen, um aus ihnen zukünftig wertvolle Erkenntnisse zu ziehen und ihr Potenzial bestmöglich zu nutzen.

Die Werte in Spalte B sind gewöhnlich doppelt so hoch wie diejenigen in Spalte C. Es gibt allerdings ein paar Ausnahmen, bei denen der Wert in Spalte C halb so groß ist wie in Spalte B.

- Auffälligkeiten in der zeitlichen Struktur
Keine Buchung für Kunde A im Dezember 2018

✓ Unterstützung beim Design eines für Sie passenden Datenmodells im Hinblick auf analytische Use Cases

✓ Kategorisierung von Daten mittels Natural Language Processing

Beispiel: Kategorisieren eines Freitextfeldes 'Kundenfeedback' in

1: positives Feedback

0: neutrales Feedback

-1: negatives Feedback

→ *"Ich fand den Vortrag spitze"*

TECHNISCHE UMSETZUNG

- ★ Gesteigerte Datenqualität und Datenkonsistenz
- ★ Höhere Transparenz in den Daten
- ★ Möglichkeit für performante Abfragen
- ★ Grundlage für Analytics & AI Anwendungen
- ★ Bestmögliche Ausschöpfung des Potenzials der vorhandenen Daten

KONSISTENTE DATEN - STAMMDATENPFLEGE - STEIGERN SIE IHRE DATENQUALITÄT!

WAS BIETEN WIR?

Ursprünglicher Datensatz

Name	Adresse 1	Adresse 2	Adresse 3	Adresse 4	Telefon
Mayer	Berliner Platz	5	45678	Musterhausen	0170-1234-567
Müller	Bahnhofstrasse	1	12345	Würzburg	0049 931 1234
Mayer	Bahnhofplatz	3	78945	Muster	0170/1234567
Müller	Bahnhofstraße	1	12567	Würzburg	(0)931 1234
Maier	Bahnhofplatz	78945	3	Muster	+49 170 1234567
Mayer	Bahnhofplatz	3	78945	Muster	0170-1234-567
Maier	Berliner Platz	5	45678	Musterhausen	+49 170 1234567
Maier	Berliner Platz	5	45678	Musterhausen	0170 1234567
Mueller	Bahnhofstr	1	12345	Wuerzburg	0931 1234
Mayer	Bahnhofplatz	3	78945		0170 1234567
Mayer	Berlin Platz	5	45678	Musterhausen	0170-1234-567

Konsolidierter Datensatz

ID	Name	Adresse 1	Adresse 2	Adresse 3	Adresse 4	Telefon
1	Müller	Bahnhofstrasse	1	12345	Würzburg	0049 931 1234
2	Maier	Bahnhofplatz	3	78945	Muster	0049 170 1234567
3	Mayer	Berliner Platz	5	45678	Musterhausen	0049 170 1234567