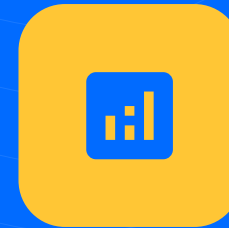


Besseres Data Warehouse dank Künstlicher Intelligenz

Dext.ai

Künstliche Intelligenz für Data Warehouses & Co.



Agenda

- Über uns
- Automatisierung in der Datenqualitätsprüfung
- Proaktive Identifikation verspäteter Datenanlieferungen
- Erkennung lang laufender SQL-Abfragen (AIOps – System Health-Check)

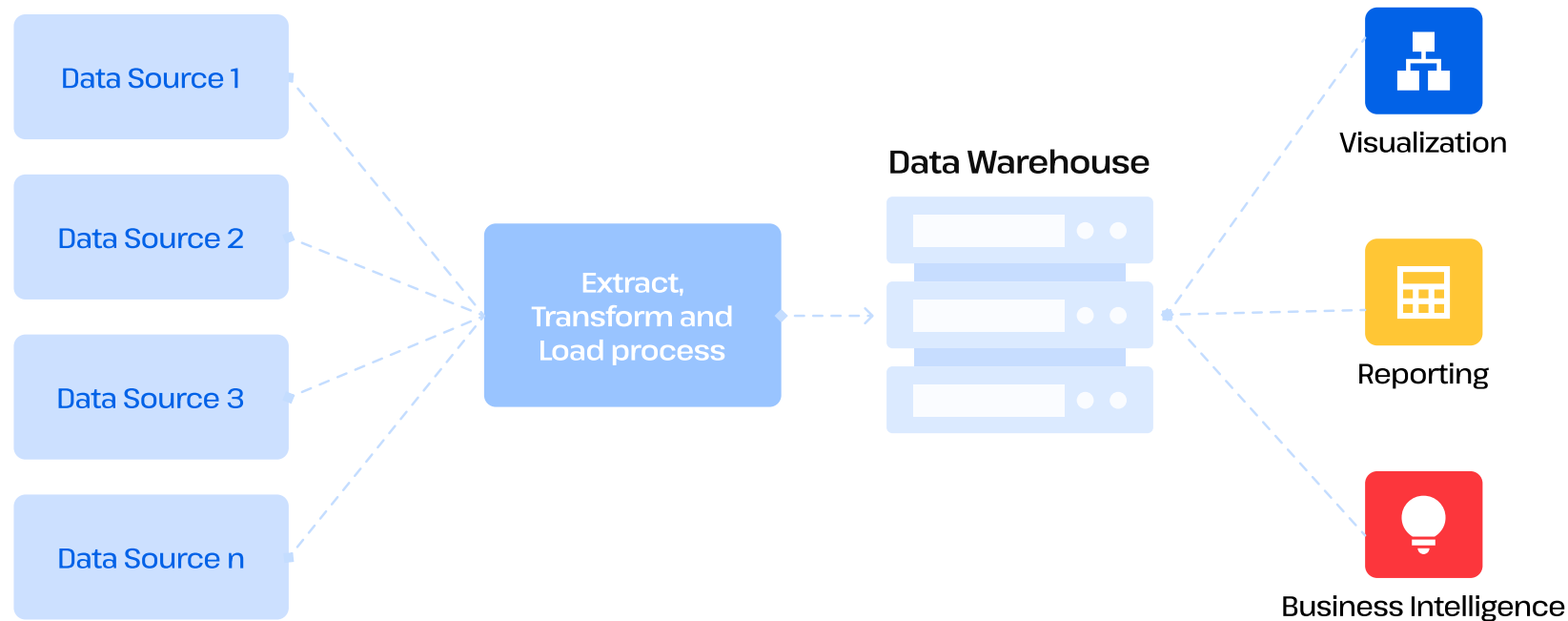
Über uns

- DEXT.AI GmbH wurde 2020 in Wien gegründet
- Ziel: Entwicklung von KI-basierter Software für Datenplattformen
- Grants von Austria Wirtschaftsservice und Wirtschaftsagentur Wien
- Erstes Produkt seit 2021 am Markt
- Erster Kunde: IT-Services der österreichischen Sozialversicherung

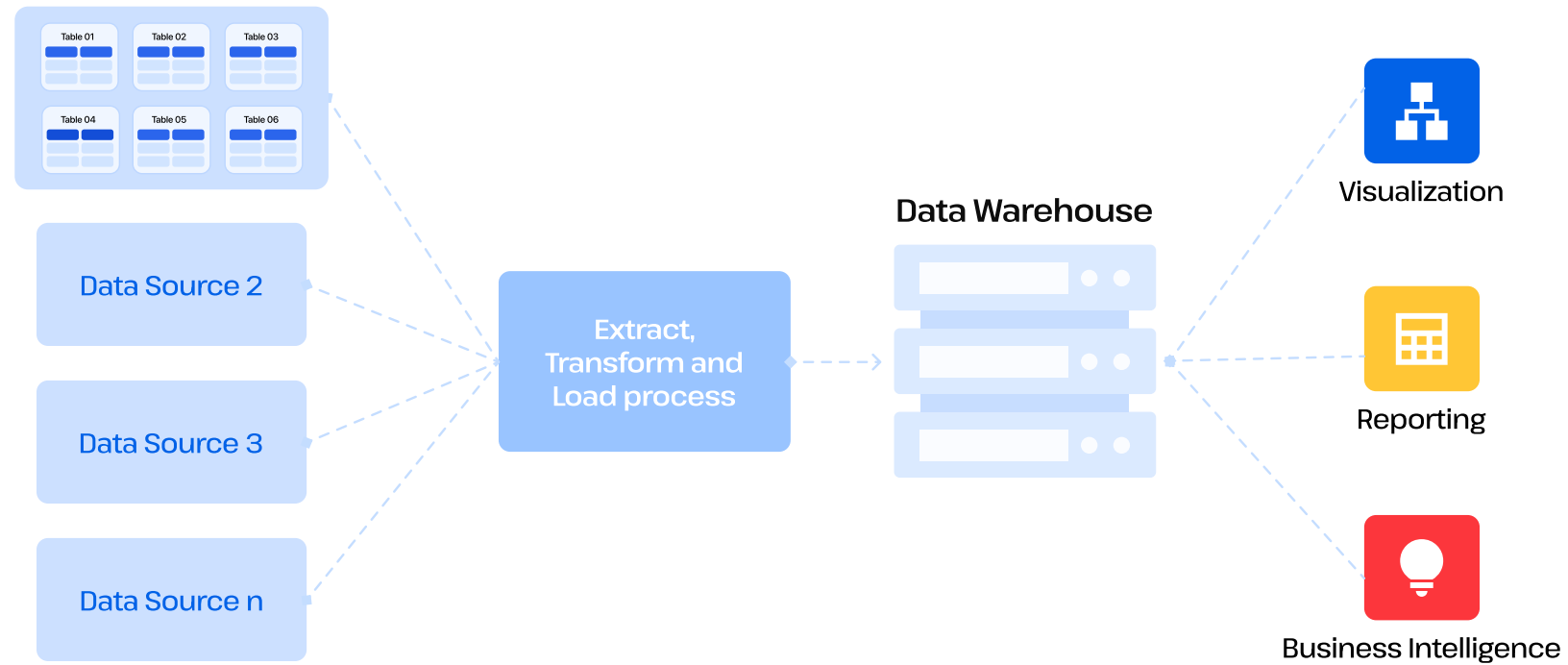
Use Case 1:

Automatisierung in der Datenqualitätsprüfung

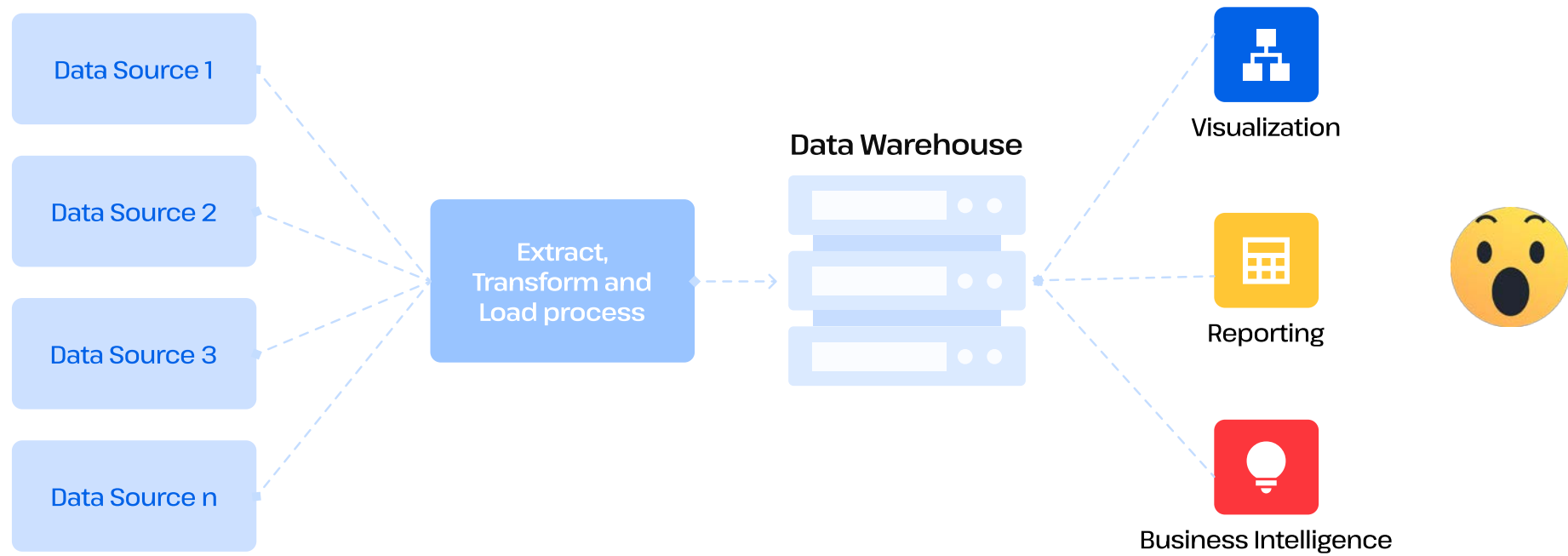
Täglich verarbeitet jedes mittelgroße Data Warehouse Daten aus dutzenden / hunderten Quellsystemen



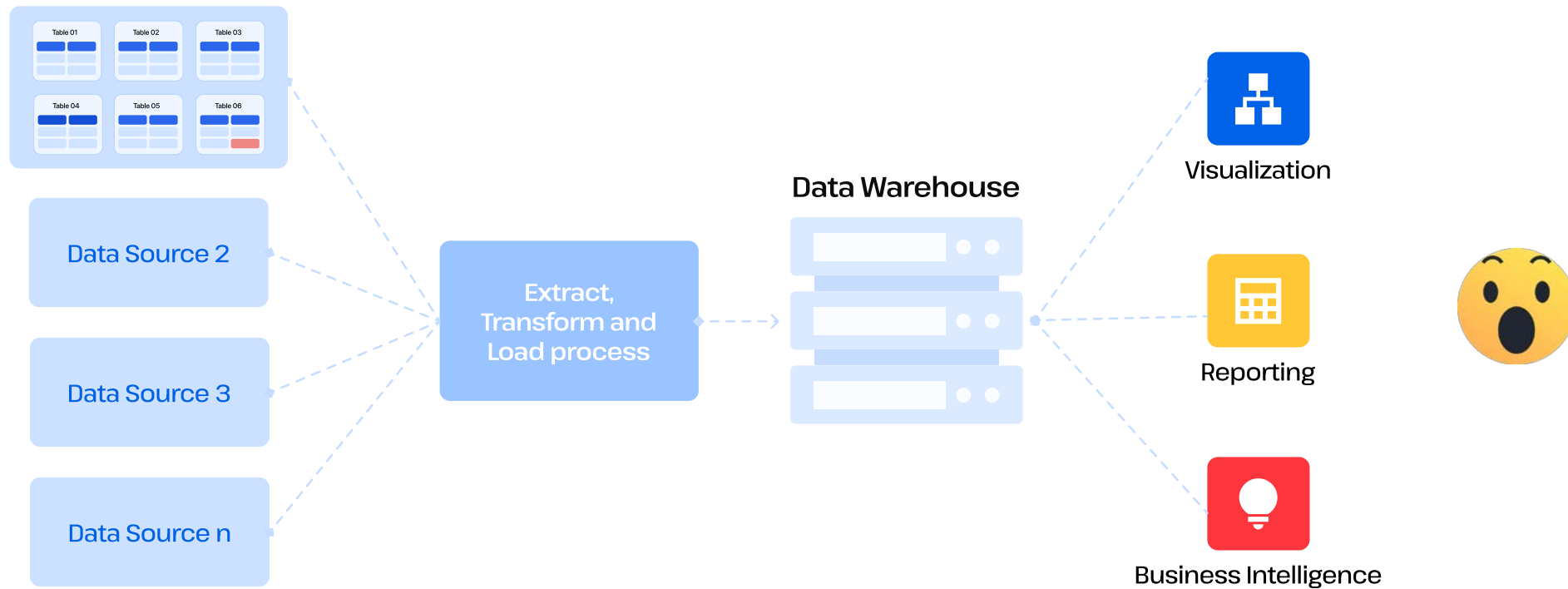
Jedes Quellsystem enthält Daten in vielen verschiedenen Strukturen



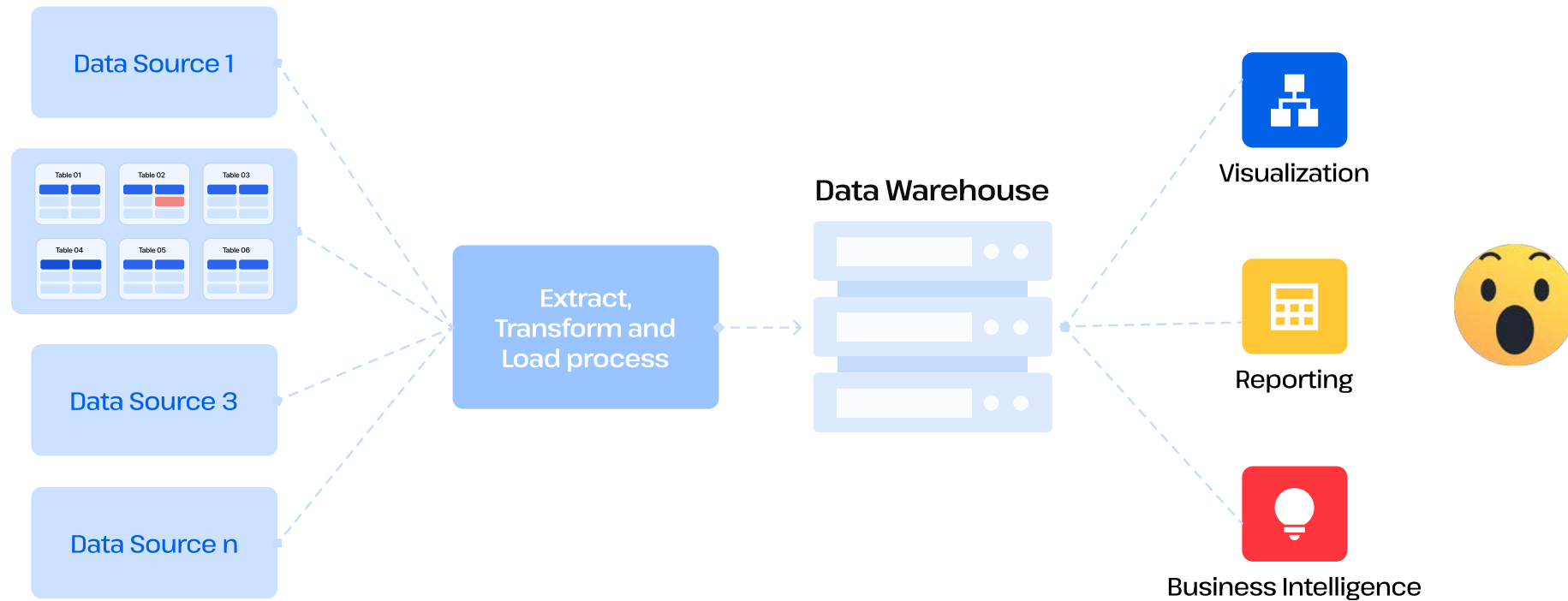
Es kommt vor, dass die Berichte falsch sind



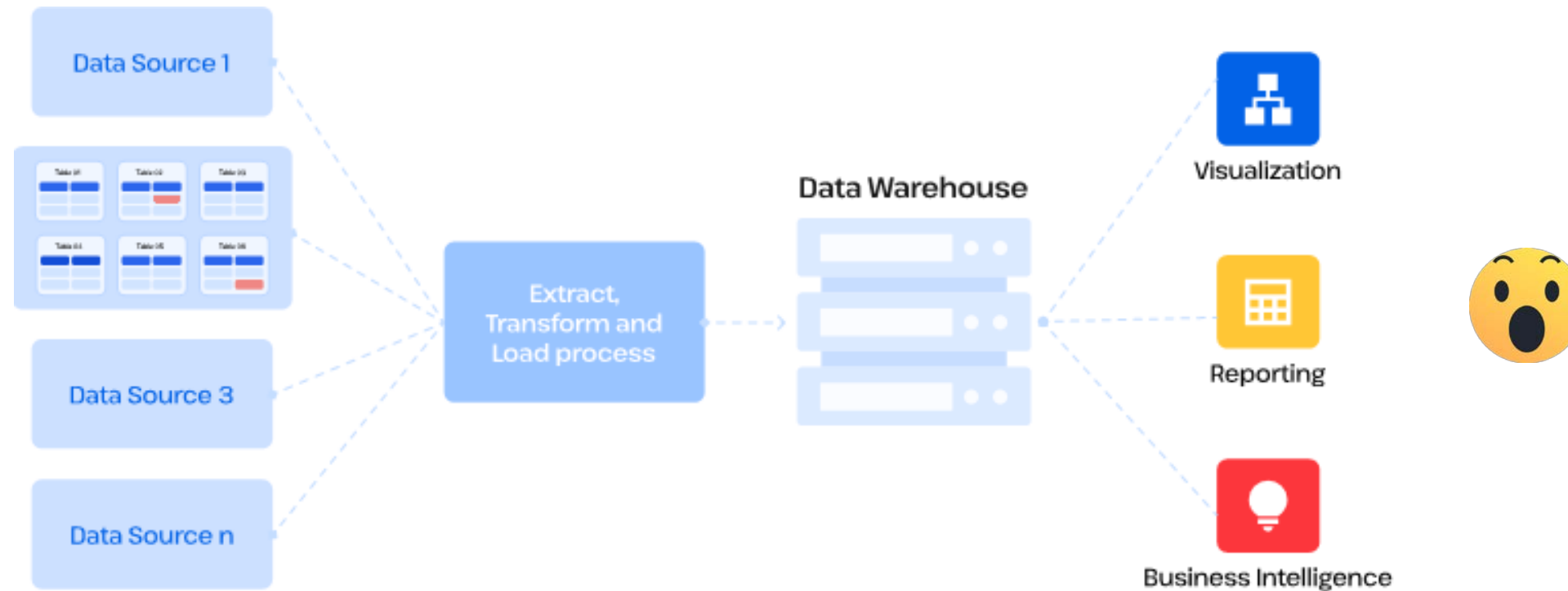
... die u.a. aufgrund von Fehlinformationen im Quellsystem erstellt wurden



Fehler treten in vielen Quellsystemen und verschiedenen Datenstrukturen auf



Ein falscher Bericht kann mehrere Fehlerursachen haben



Was sind typische Fehler in Quellsystemen?



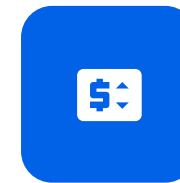
Data Ghosting

Wo sind meine Geschäftskunden?



Leere Spalten

Warum haben meine Angestellten kein Geburtsdatum mehr?



Abgeschnittene Spalten

Ist der Umsatz wirklich gesunken?



Invertierte Werte

Ist es Sommer oder Winter in den Daten?



Inkorrekte Werte

Warum tauchen deutsche Bundesländer auf und nicht österreichische?



Vertauschung von Spalten

Vorname und Nachname aufgrund von Fehlern im Quellsystem vertauscht?

Klassischer Ansatz / Status Quo

Definieren Sie Datenqualitätsregeln

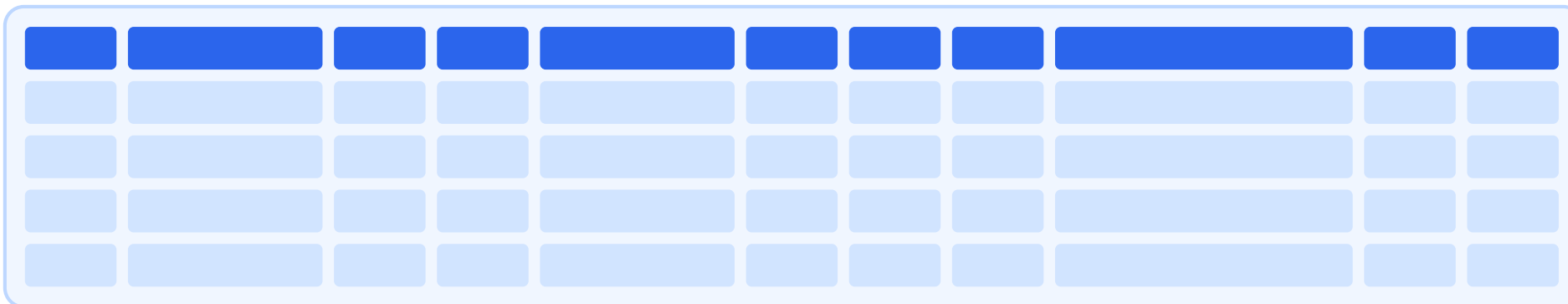
Welche Regeln braucht man?

- Anzahl der Datensätze
- Anzahl der fehlenden Werte
- List gültiger Werte
- Summe, Average, Min und Max Wert
- ...



Wie viele Datenqualitätsregeln werden für eine Tabelle mit 100 Spalten benötigt?

HUNDERTE, TAUSENDE



Die Daten ändern sich mit der Zeit, die Regeln müssen angepasst werden

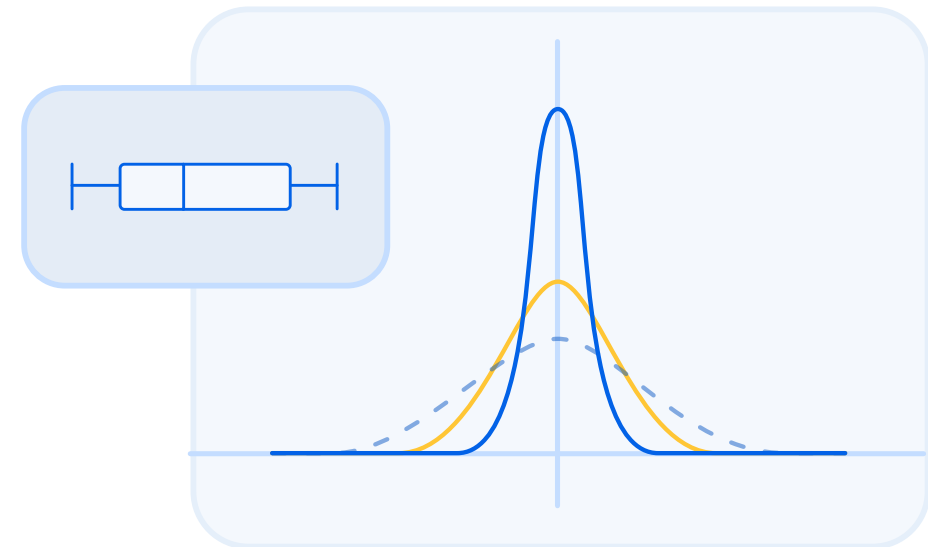
Könnte Künstliche Intelligenz helfen?

JA



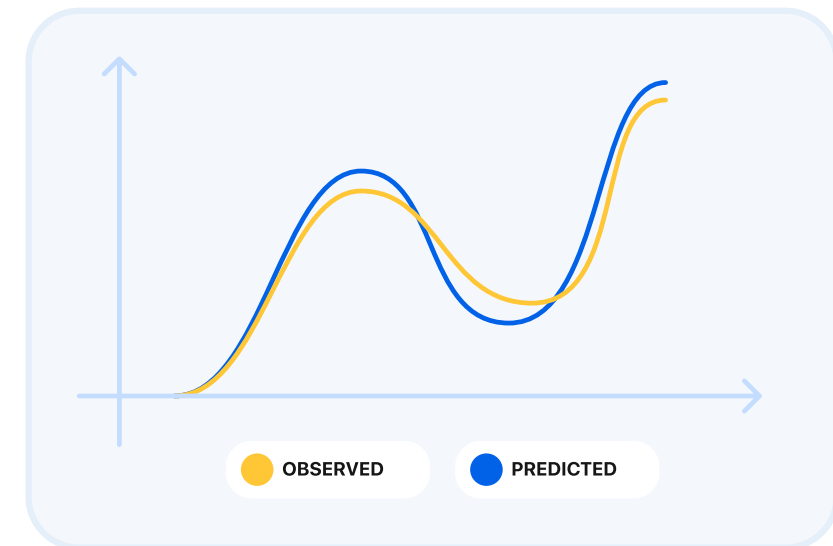
Autometrics

Berechnet Profile der Daten im Laufe der Zeit und erfasst die wichtigsten Kennzahlen für die Analyse



Forecasting Model

Unter Verwendung der Metriken und mit Hilfe von Unsupervised Learning können die zukünftigen Werte vorhergesagt werden



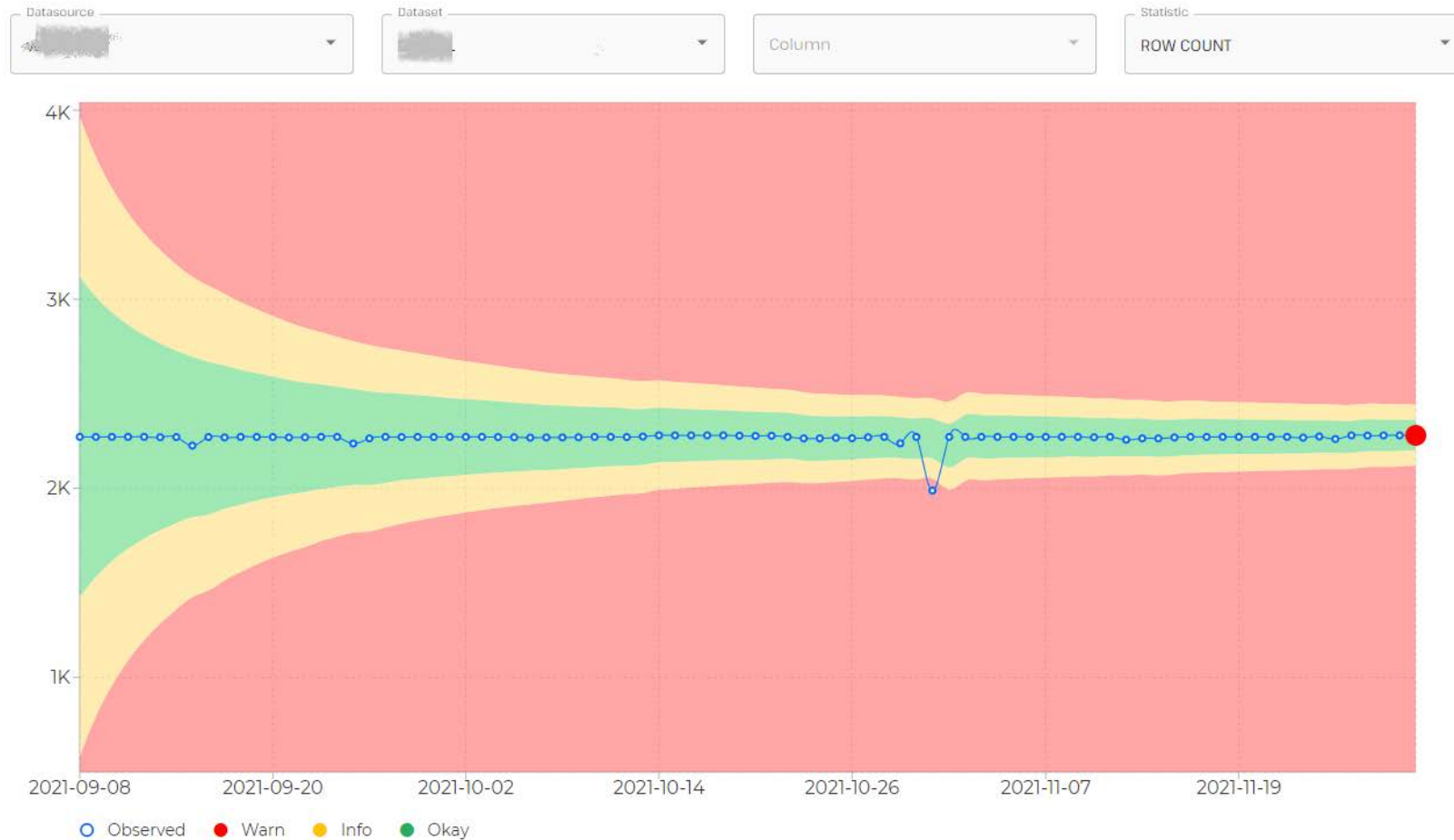
Autothresholds

Passt automatisch die Schwellwerte basierend auf vorhergehenden Prediction-Fehler an

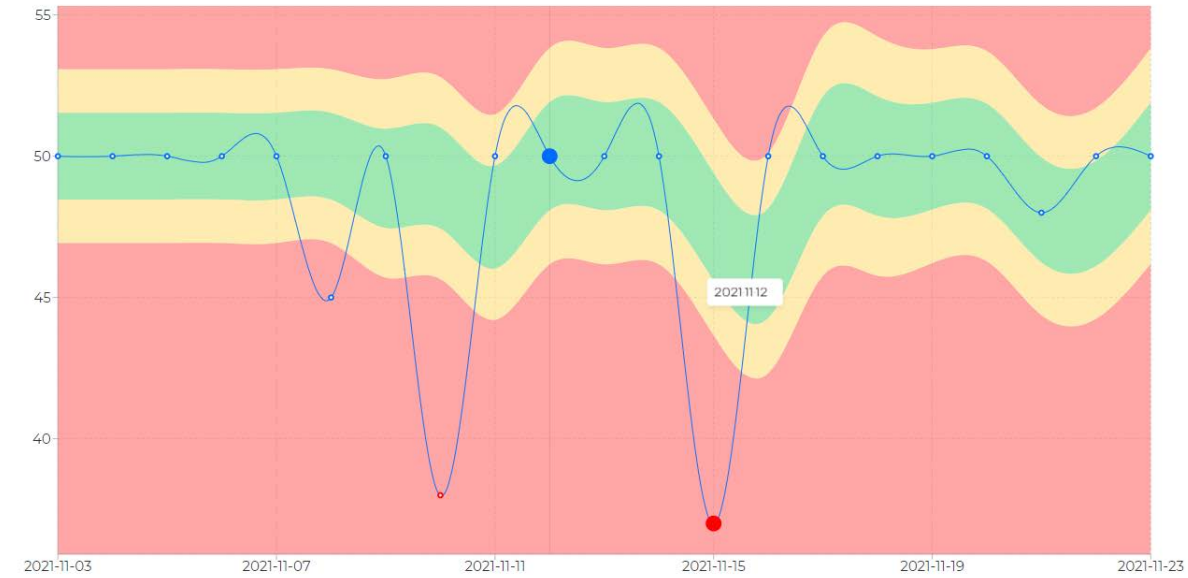
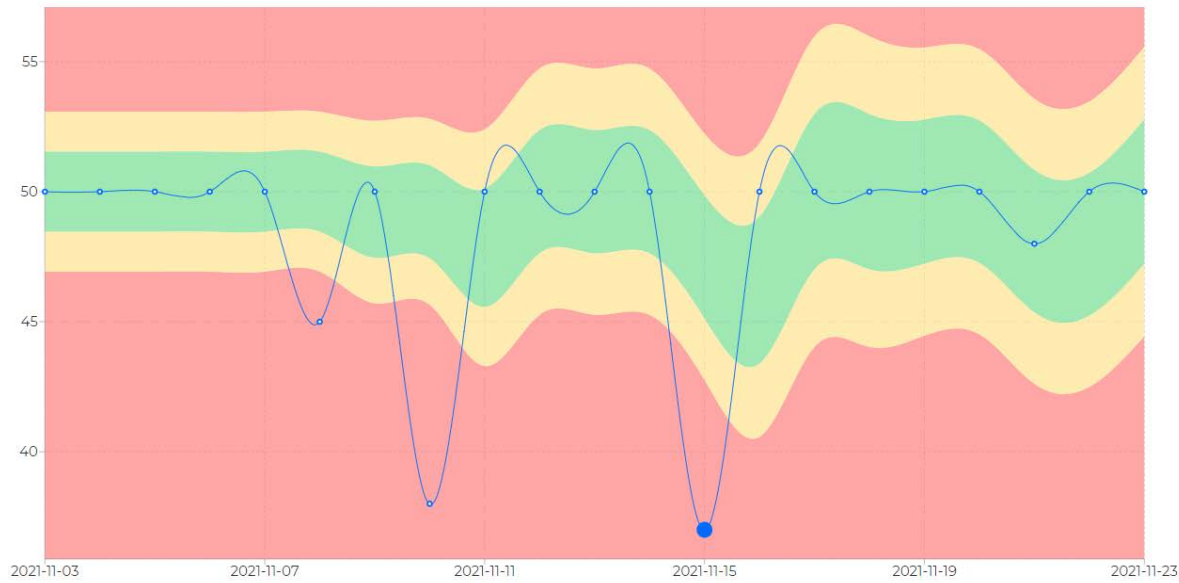


Was soll man bei der Umsetzung beachten?

Die Schwellwerte sollten sich automatisch anpassen

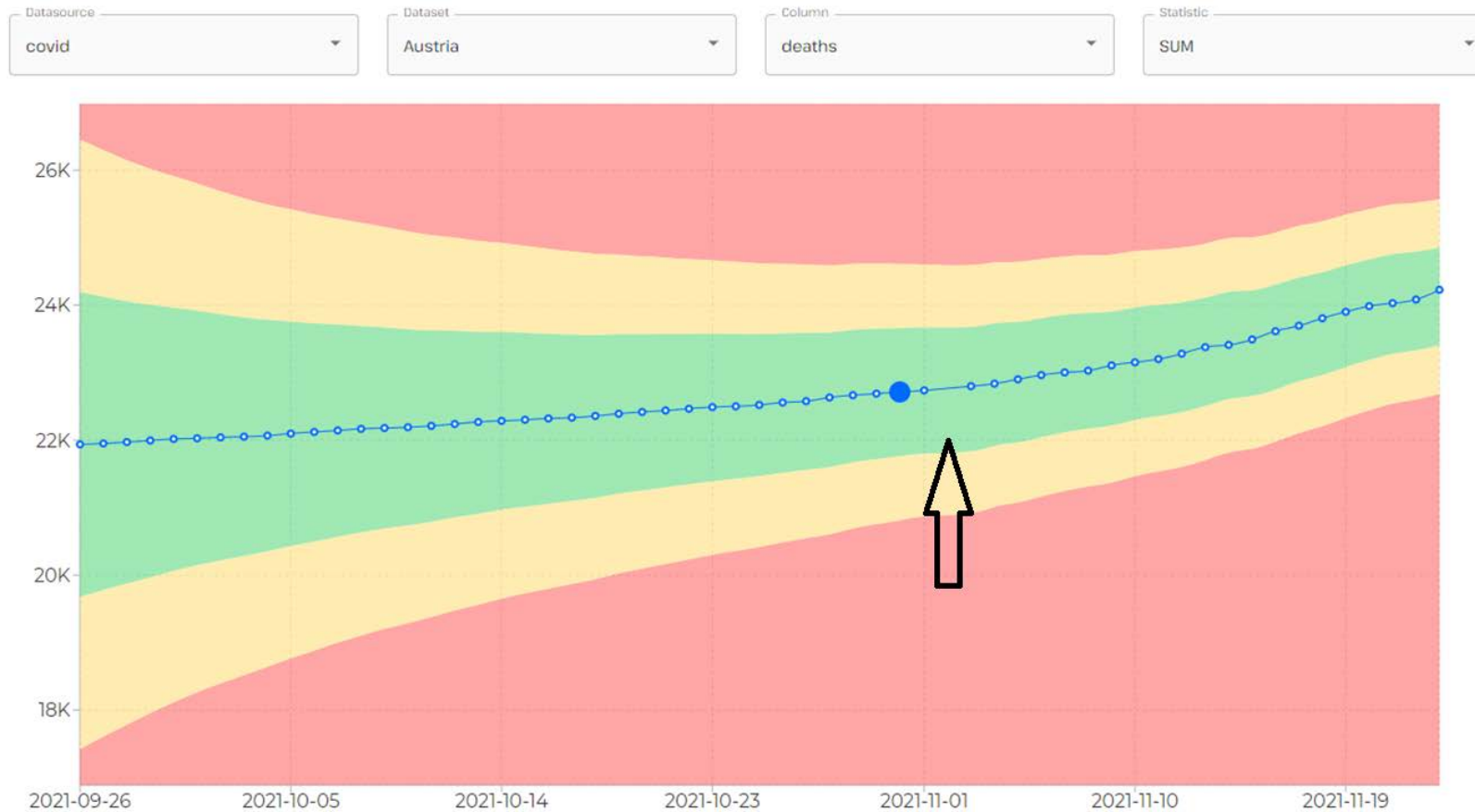


Funktionalität „Confirm anomaly“



Automatisierung in der Datenqualitätprüfung

Daten werden an einem bestimmten Tag nicht angeliefert

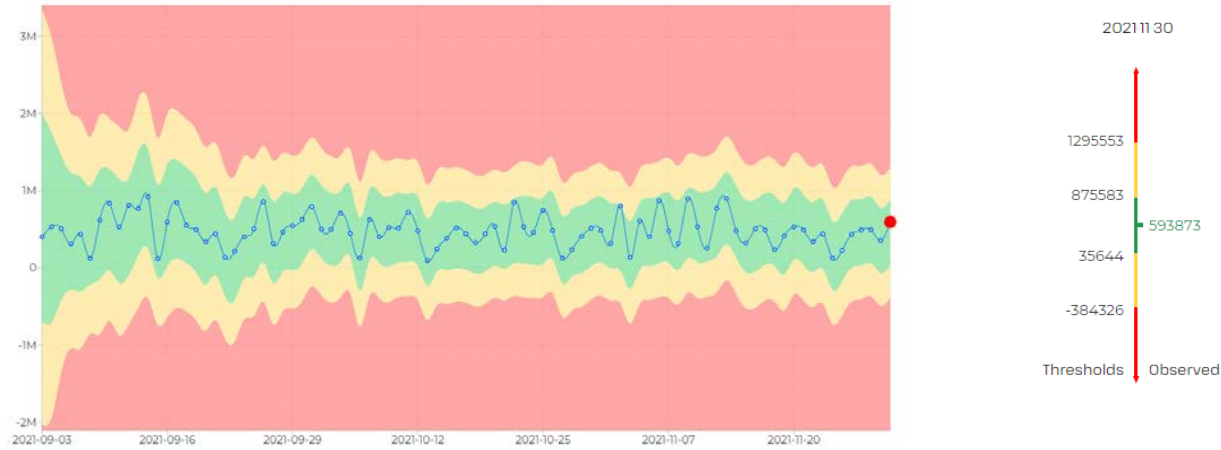


Sensitivity und Memory vom Traffic Light System

Alert Configuration

Sensitivity:

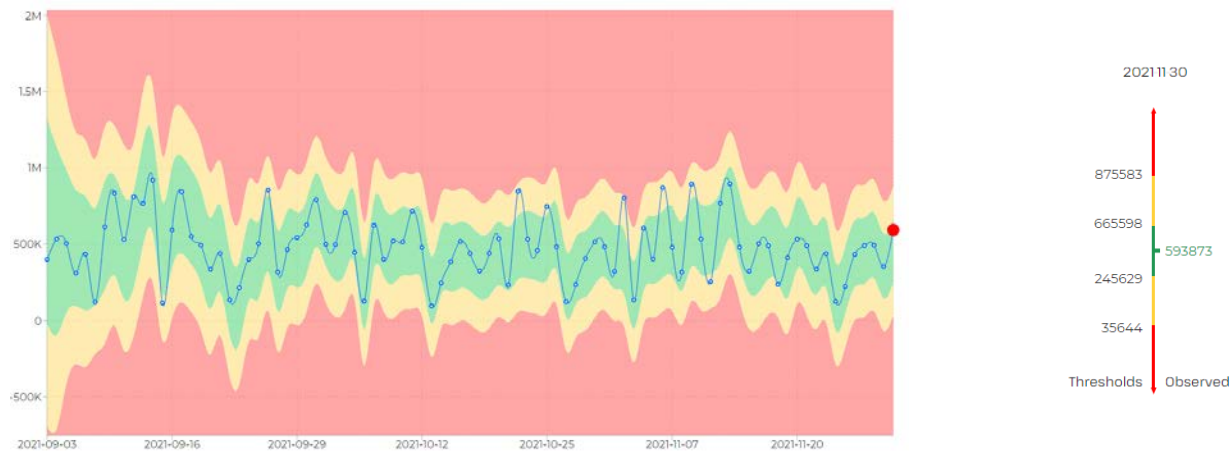
Memory:



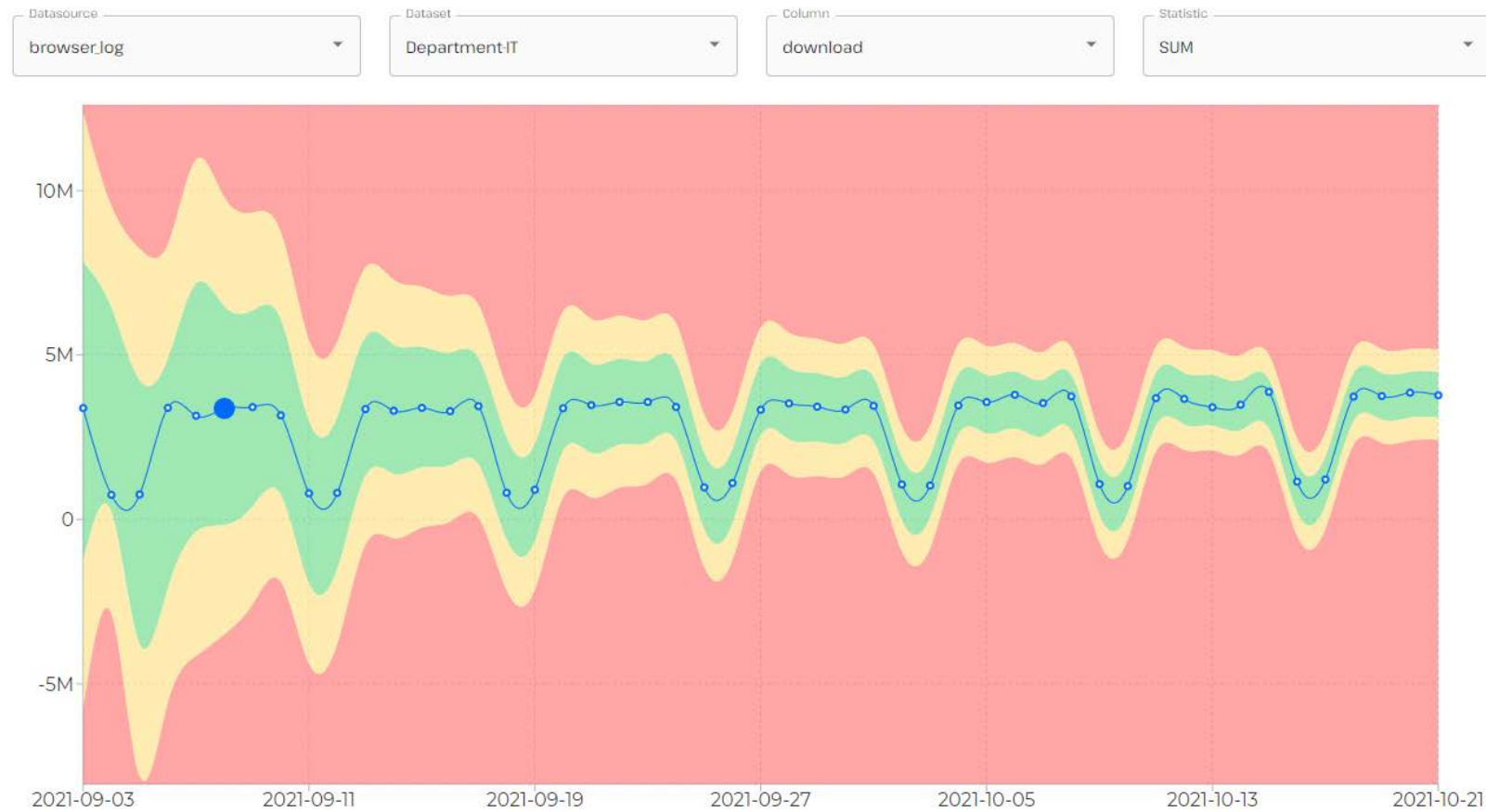
Alert Configuration

Sensitivity:

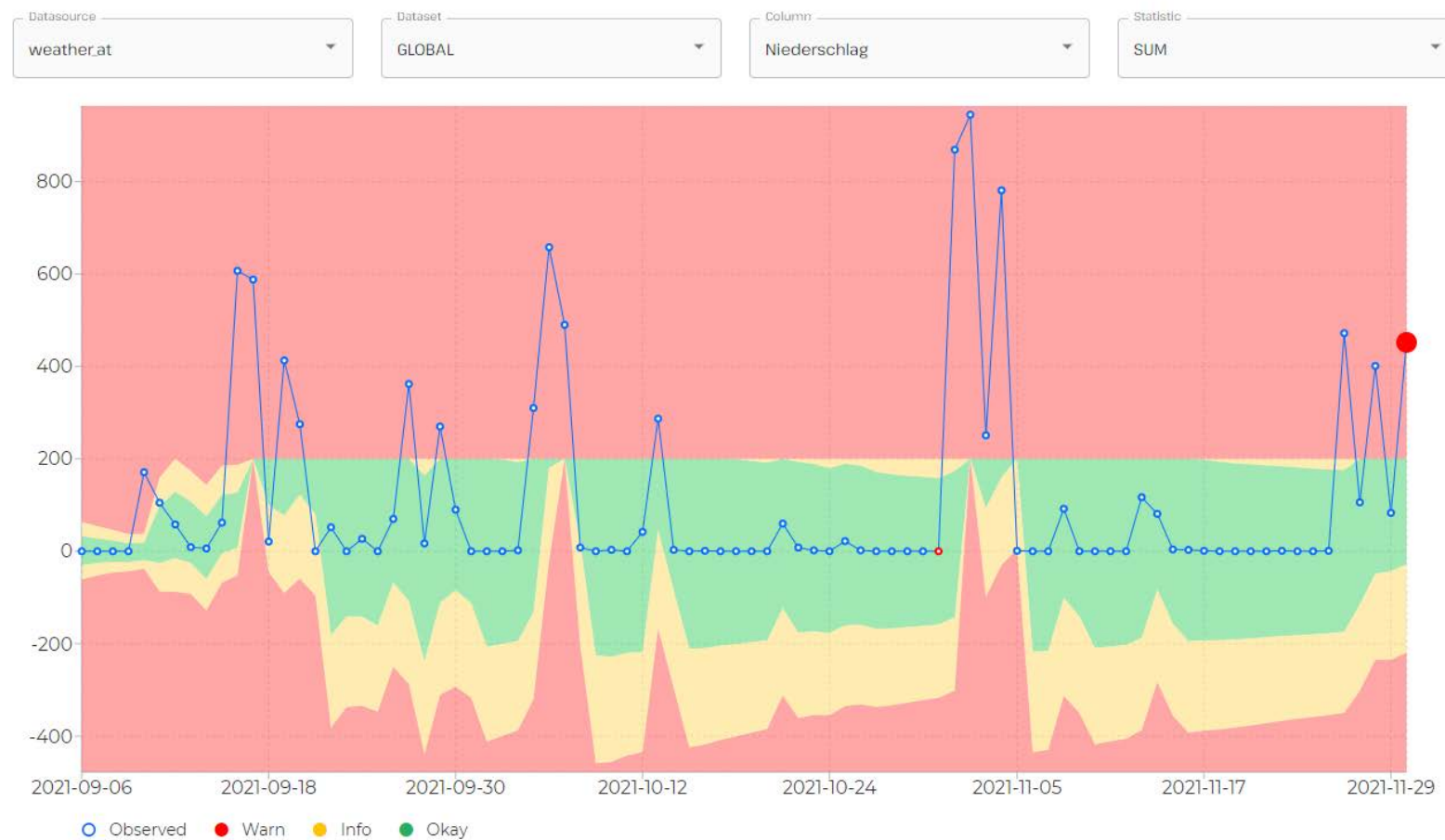
Memory:



Saisonale Effekte



Manuelle Regeln auf aggregierten Metriken



Rules Configuration

Lower Limit

Upper Limit

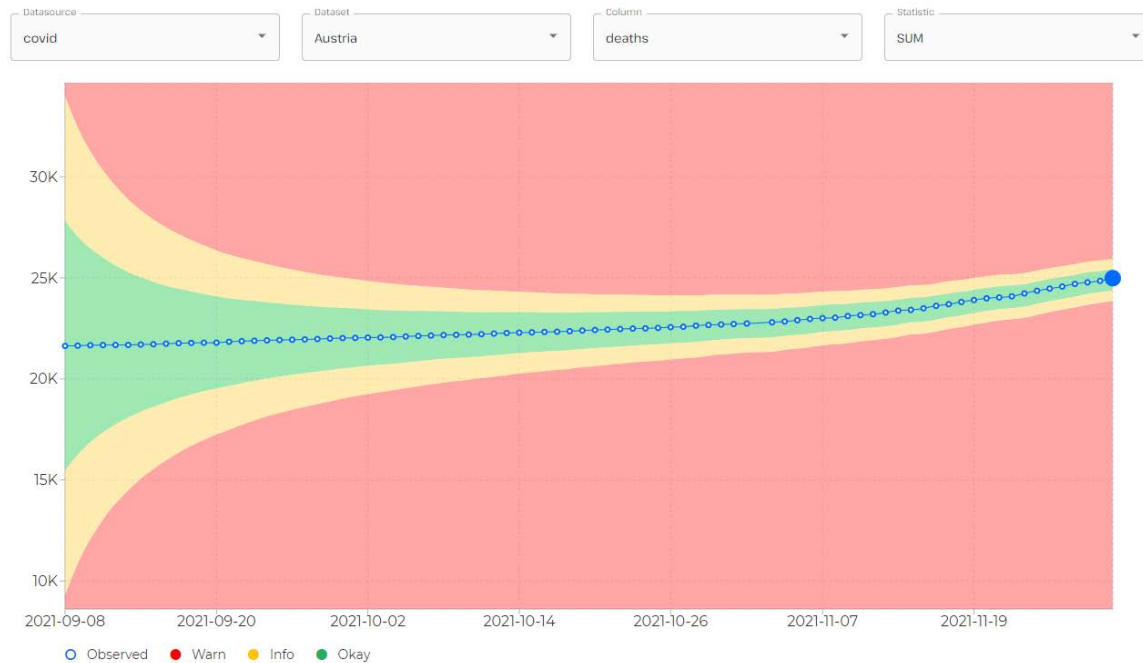
200

Automatisierung in der Datenqualitätprüfung



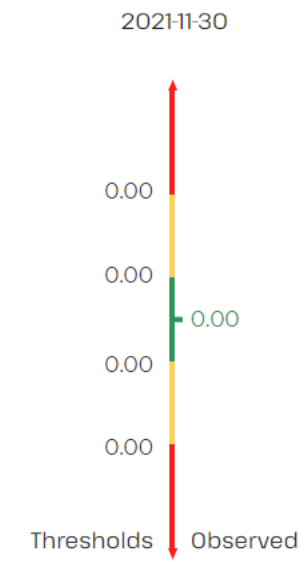
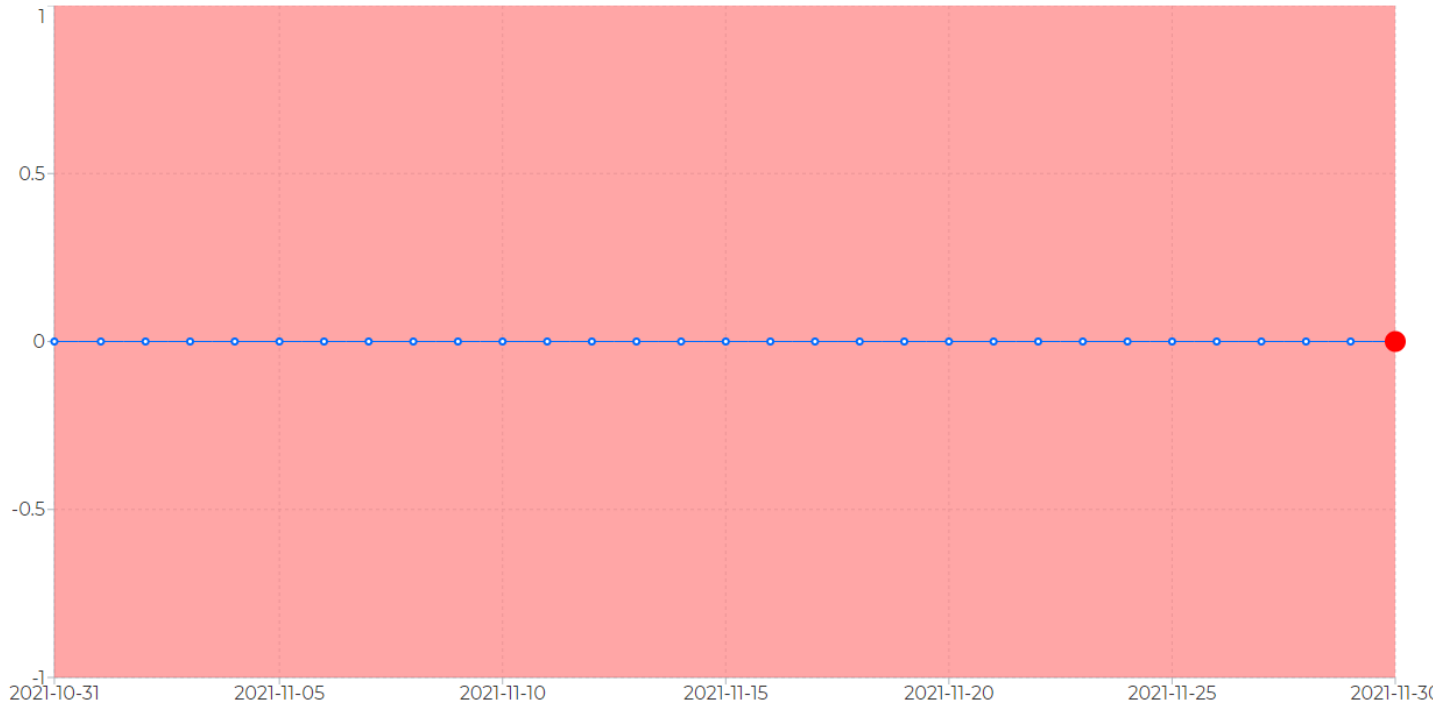
Virtuelle Tabellen:

- Statische Unterteilung
- Dynamische Unterteilung



Befüllungsgrad

Datasource: browser_log | Dataset: Department-Finance | Column: upload | Statistic: NULL COUNT

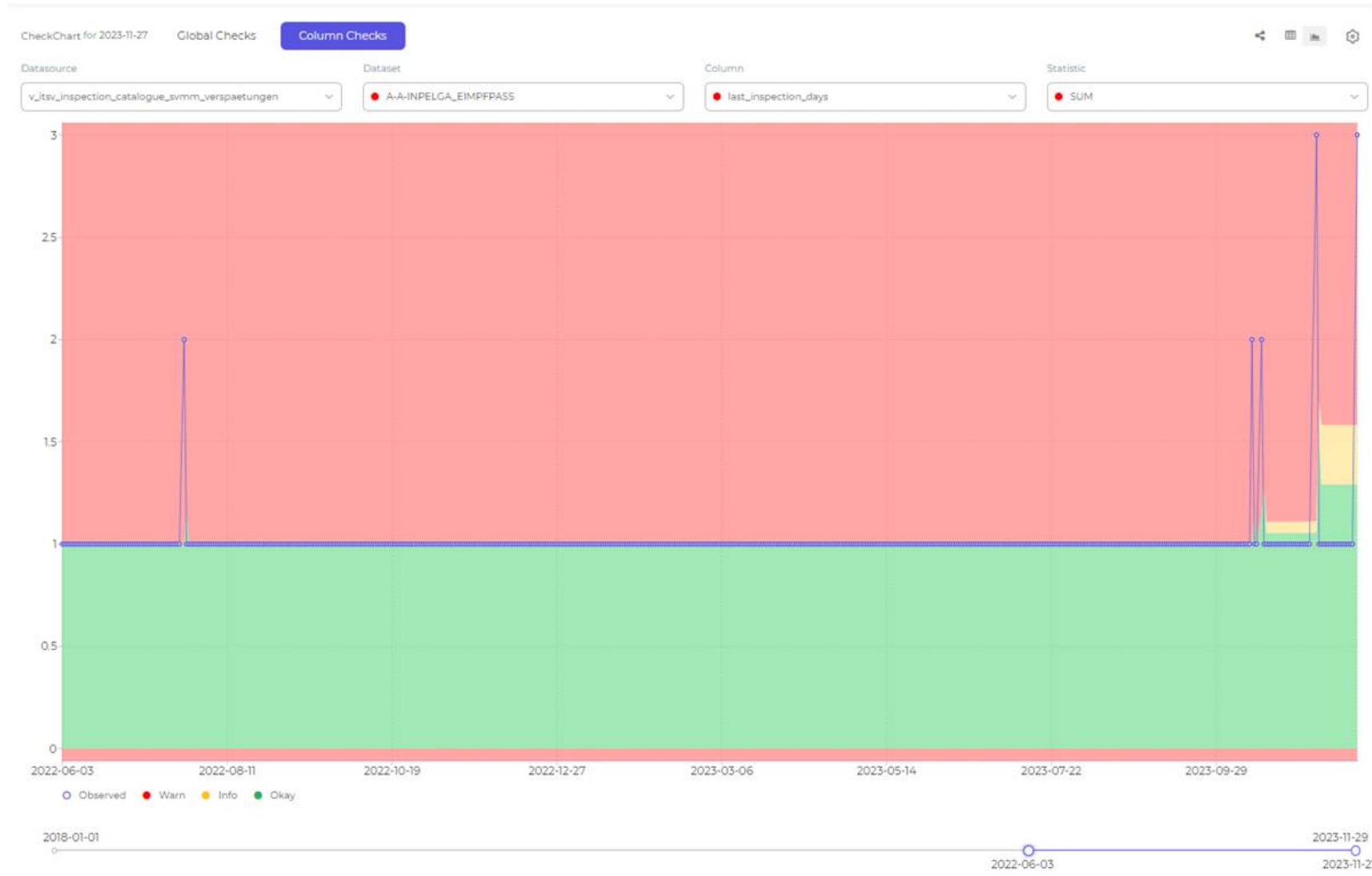


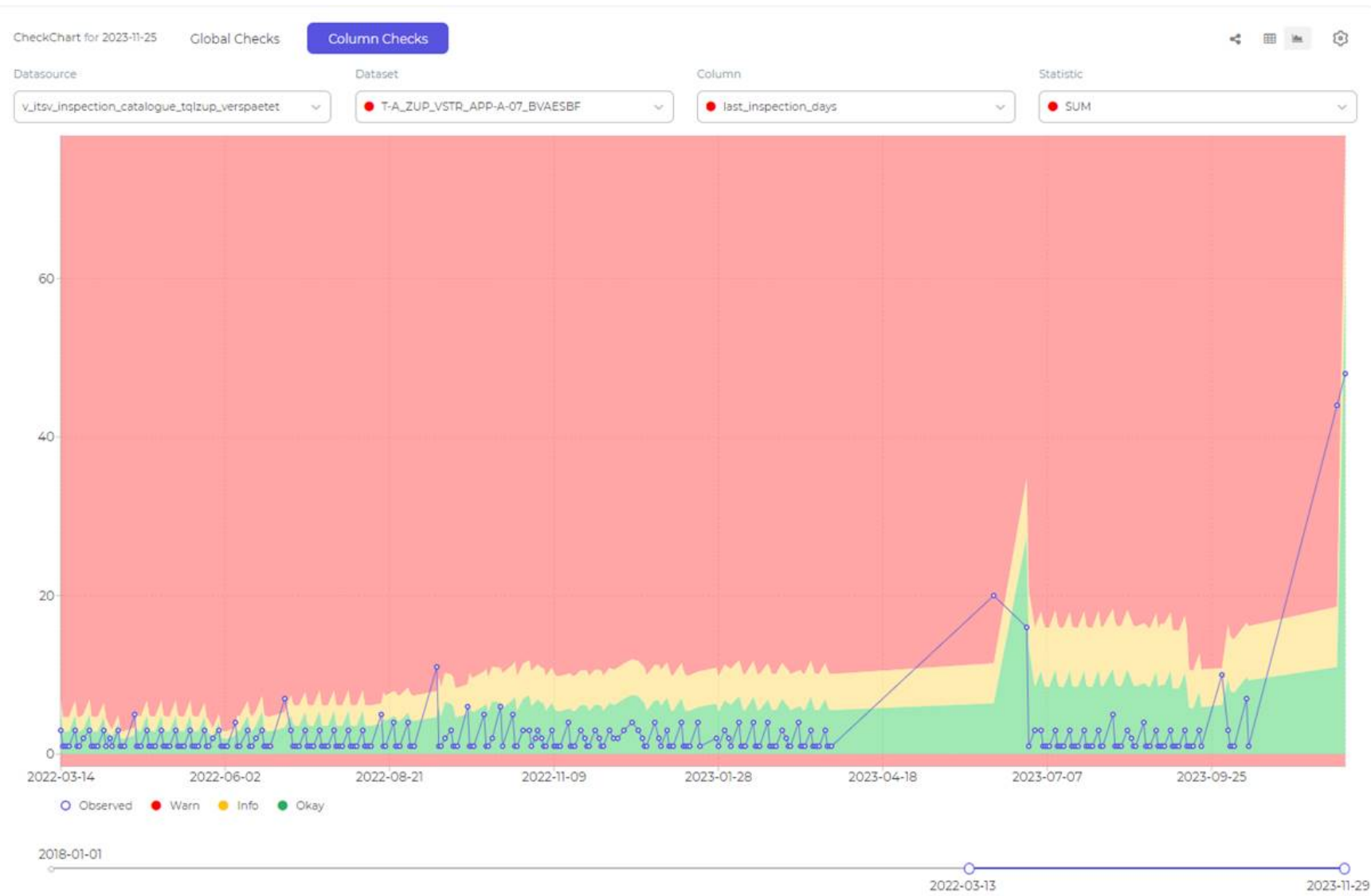
Use Case 2:

Proaktive Identifikation verspäteter Datenanlieferungen

Was kann gemessen werden?

- Die Zeit zwischen den Anlieferungen





Use Case 3:

Erkennung lang laufender
SQL-Abfragen
(AI Ops – System Health-Check)

Was kann gemessen werden?

- Anzahl der SQL-Abfragen, die länger als 15 min dauern
- CPU Consumption
- IO Consumption
- Anzahl der gelockten SQL-Abfragen



Use Case 1:

Automatisierung in der Datenqualitätsprüfung

Use Case 2:

Proaktive Identifikation verspäteter Datenanlieferungen

Use Case 3:

Erkennung lang laufender SQL-Abfragen (AI Ops)