# Data Quality Checks im Kontext Customer Experience – realisiert mit Google Cloud

## Ausgangslage und Motivation

Das Team von Customer Experience & Data stellt im Konzern Daten in Bezug auf die Kundenzufriedenheit zur Verfügung. Dabei steht der sogenannte **Net Promoter Score (NPS)** im Fokus mit der zentralen Frage:

„Wie wahrscheinlich ist es, dass Sie die Telekom im Bekannten- und Freundeskreis weiterempfehlen? “ (andere Fragen sind ebenfalls möglich, weswegen es verschiedene NPS-Typen gibt)

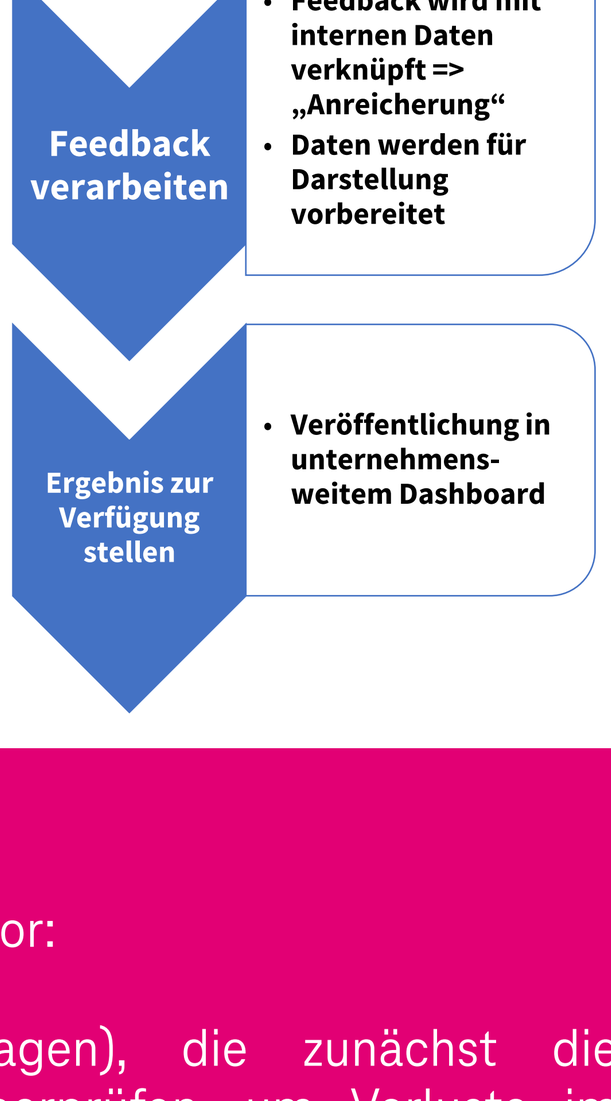
Der NPS berechnet sich dabei aus dem prozentualen Anteil der Promotoren abzüglich des prozentualen Anteiles der Detraktoren (siehe Abbildung):

**Zwei Hauptfaktoren sprachen für eine Konzeptionierung von Ende-zu-Ende Data Quality Checks:**

1. Bis zur Darstellung im Dashboard werden viele Daten verarbeitet, die voneinander abhängig sind, was den Prozess fehleranfällig macht
2. Die Datenbasis wird auf die Google Cloud Platform migriert, was die Nutzung der integrierten Tools von Google Cloud möglich macht

Ziel des Praxisprojekts war daher:
**Konzeptionierung und Einführung von Ende-zu-Ende Data Quality Checks auf Basis der Möglichkeiten von Google Cloud**

## Ablauf des Prozesses




## Umsetzung

1. Überblick über Status-Quo zu bestehenden Checks verschaffen und weitere Bedarfe ermitteln

2. Recherche zu Möglichkeiten und passende Tools, um zukünftigen Aufbau zu evaluieren

3. Erstellung eines Konzepts, welcher den gesamten Prozess abdeckt
   a. Wichtiger Aspekt: Einführung von inhaltlichen Tests, um zukünftig zusätzlich zur Vollständigkeit auch die Sinnhaftigkeit der Daten überprüfen zu können

4. Implementierung des Konzepts und ansprechende Visualisierung der Ergebnisse

## Ergebnis und Ausblick

Das fertiggestellte Konzept sieht vor:

- SQL-Queries (Datenbankabfragen), die zunächst die Vollständigkeit der Daten überprüfen, um Verluste im Prozess auszuschließen
  – Die Nutzung von BigQuery von Google Cloud ermöglicht dabei sehr schnelle Abfrage von Daten

- Mit Hilfe vom Python-Framework PySpark (ermöglicht umfangreiche Datenanalysen) sollen Sollwerte für die Attribute ermittelt werden, um dann mittels sogenannten Assertions in Dataform (Regeln zur Wahrung der Datenqualität in einem von Google Cloud inbegriffenem Tool) Abweichungen feststellen zu können

Das Dashboard zur Überprüfung der Vollständigkeit der Daten im Gesamtprozess ist fertiggestellt (nach folgendem Muster):

Next Step: Ermittlung von Sollwerten für inhaltliche Prüfung

Konzerneinheit: Telekom Deutschland GmbH | Abteilung: CX & Data | Student: Achraf El Hasnaoui

h_da hochschule darmstadt informatik
member of EUT+ EUROPEAN UNIVERSITY OF TECHNOLOGY

**Praxisprojekt 2**
**Informatik Dual**
**2024**

**Gruppe A**