

Zwischenstand, PSP & Gantt

Projektstrukturplan



21.06.2024

Gantt



Dateien einlesen

Korrupte CSV's

Dateigröße

Correct Rows Wrong Rows

ABCD 709595 9

MNOP 1108949 6

ZYXW 429755 2

>Zeilen manuell reparieren

CSV Größe Parquet Größe

ABCD 982 MB 21 MB

MNOP 1820 MB 30 MB

ZYXW 448 MB 12 MB

- ➤ Git Large File System < 2 GB!
- ➤ Parquet

Actions & Passendes Datenformat

- 1. Auswahl Actions
- 24 Actions
- Auswahlkriterien
 - Relevanz bzgl. der Problemstellung
 - InteressanteZusatzinformationen
 - Ähnlich Verteilung über alle Airlines

2. Extraktion Actions

- Herausforderungen:
 - Unterschiedliche Formate
 - Actionintern
 - Actionübergreifend
 - Buchstabendreher (AD > ZY)
- Semi-standardisierte Extraktionsfunktion

3. Datenschnittstelle

- Informationen zusammenführen
- Spezifische Tabellen:
 - Metadaten
- o TOW / ATOW
- Einheitlich JSON Datenformat

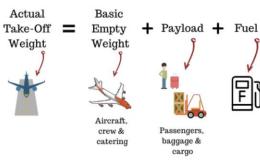
Analyse der Metadaten & Gewichtswerte

Metadaten

- Übersicht aller extrahierten Actions inkl. deren Keys
- Übersetzungstabelle (Abkürzung, Ausgeschrieben, Bedeutung)
- Identifikation der Flugmetadaten pro Action

Gewichtswerte

- ETOW vs ATOW
- Nutzung Übersicht
- Loadsheet&LoadingInstruction (ATOW)
- Berechnung



Ausblick



Wie gehen wir Probleme an?

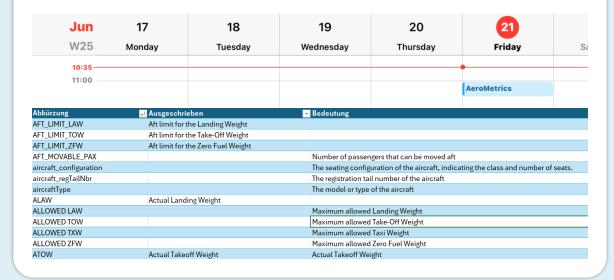
Zeitdruck & Unerwartetes

- Parallellisierung
- Einbindung von allen



Informationsaustausch

- Regelmäßige Termine
- Ausführliche Dokumentation





Q&A