AMSE/SAKI 2023 Project

By Nicolas Bandel

Structure

- Question
- Data sources
- Data processing
- Results
- Problem

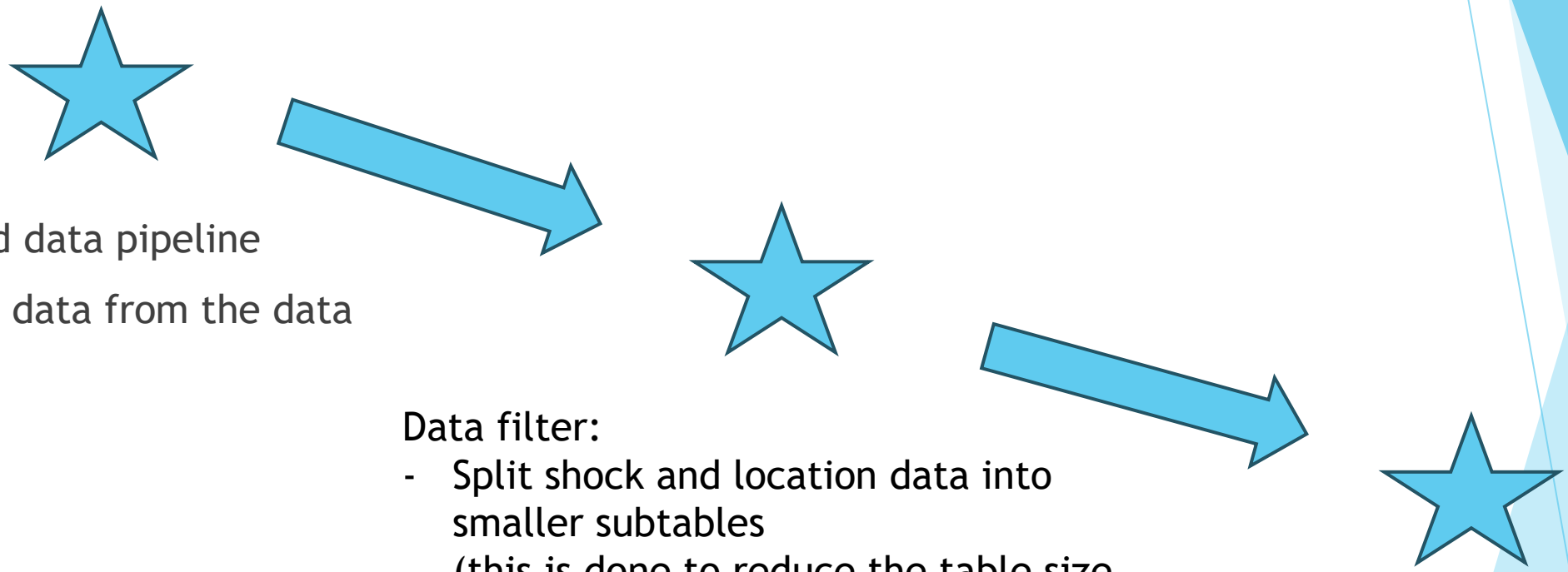
Question

- Determine the most problematic DB Cargo stations regarding measured shock events in Germany

Data sources

- To answer the question the two data sources have been evaluated:
  - Data source 1: Impact data of freight wagons
  - Data source 2: List of freight transport locations

Data processing



Automated data pipeline

- Read the data from the data sources

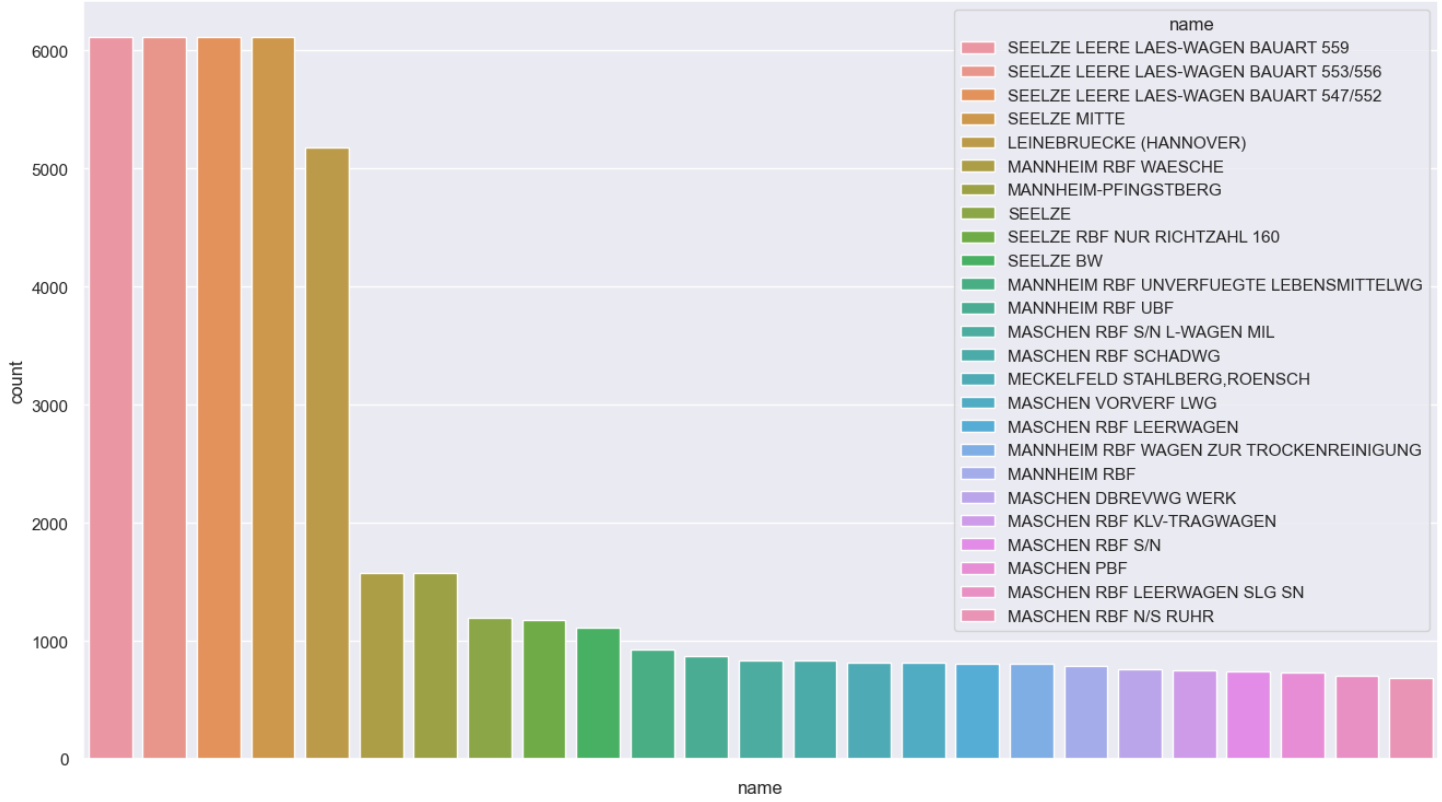
Data filter:

- Split shock and location data into smaller subtables
  (this is done to reduce the table size in the data processor)
- Group shock events by their geo location
- Divide shock data by the speed of the train

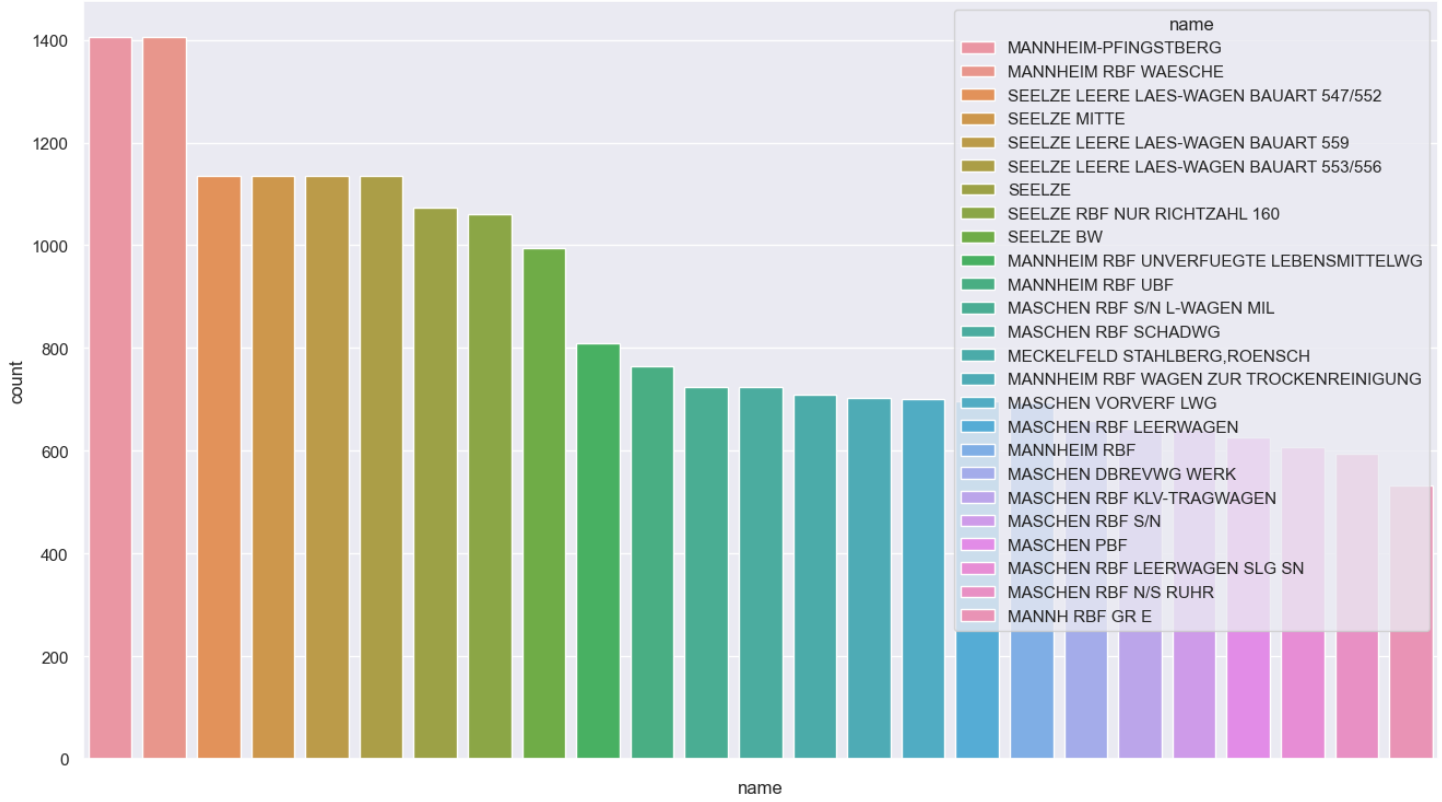
Data processor:

- Calculate distance between shock event and each DB cargo location
- Count the number of shock events close to each location
- Create final result table

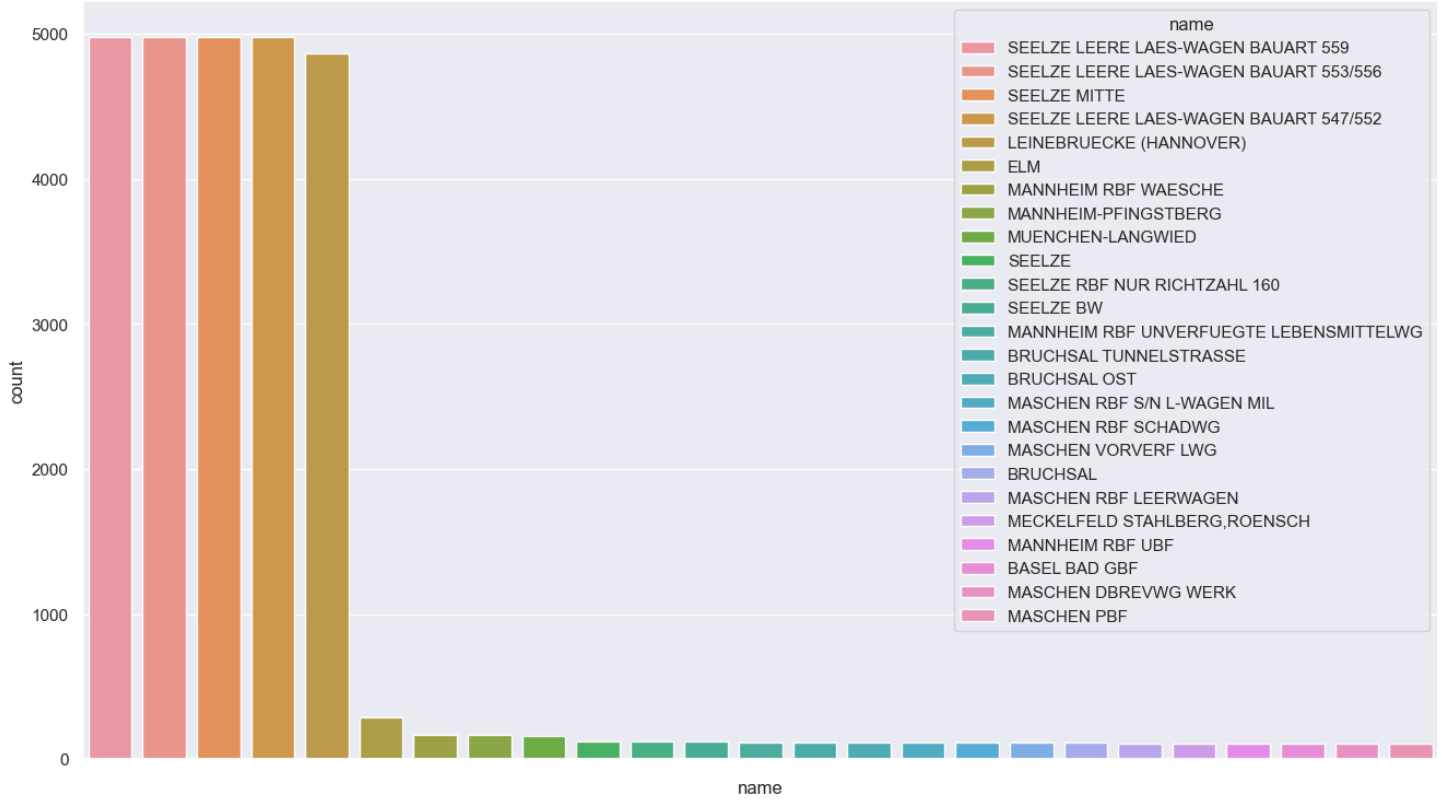
Results
What are the most problematic stations? (All)



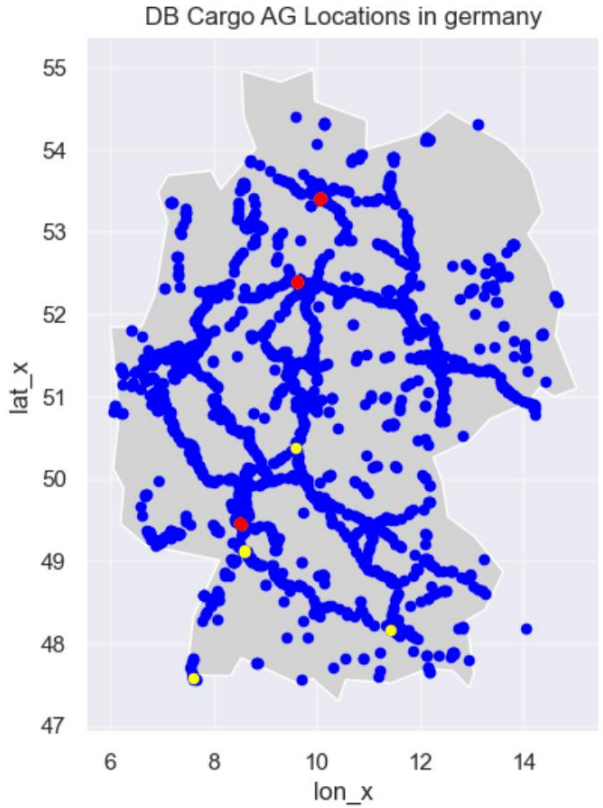
Results
What are the most problematic stations? (Standing)



Results
What are the most problematic stations? (Moving)



Where are the most problematic locations in Germany?



What problems occured during the project?

- Data-sets became to big
  - Solution: splitting the data into tables based on their geo location
- The locations in the DB-Data sets have similar positions with different names
  - No proper solution found