

Data Science Project 25 – Seat Reservation

Ia Diakonidze, Raela Salja, Dunja Milojevic

Challenge

Main Challenges & Difficulties during the project include the following:

- Business Understanding**
Difficulties associated with comprehending business principles and organization
- Data Understanding**
Challenges in grasping data source structure and relationships due to incomplete documentation
- Data Cleaning**
Data Conversion, handling missing values, choosing relevant attributes
- Data Merging**
Difficulties merging the different data sources, such as Accounting, ARES & Capacity due to data quality

Dataset

Three Main Data Sources

- Accounting Data**
• contains the revenue data for the seat reservations booked across ÖBB and other platforms
37 Attributes
12 Data files
12M Records
- ARES Data**
• contains the reservation data for the seat reservations from the inventory system of ÖBB
63 Attributes
12 Data files
11,6M Records
- Capacities Data**
• contains the capacities of the trains as well as the operating stations
9 Attributes
1 Data file
260K Records

Lookup Files

- ICE Train numbers for Passau border trains
- Sales System Channel and Offices
- Timetables & Itinerary
- Country Code and Border Stations
- Class, Couch and Cancellation

Data Understanding & Preparation

Summary of bookings per travel route

- Analyse data and drop irrelevant columns
- Data Cleaning and Conversion
- Handling Missing Data and data aggregation

Monthly average variance between booking date and departure date


Seat distribution across different sale systems

Distribution of cancellations between confirmed tickets, and cancellations (domestic & abroad)


Workflow

- Analyze data
- Data Cleaning using Python
- Handling Missing Data
- Data Aggregation
- Data Merging (lookup tables)
- Identifying Passau borders
- Extract information about ÖBB trains
- Data Merging (Accounting, ARES, Capacity)
- Visualizations Dashboard (PowerBI)

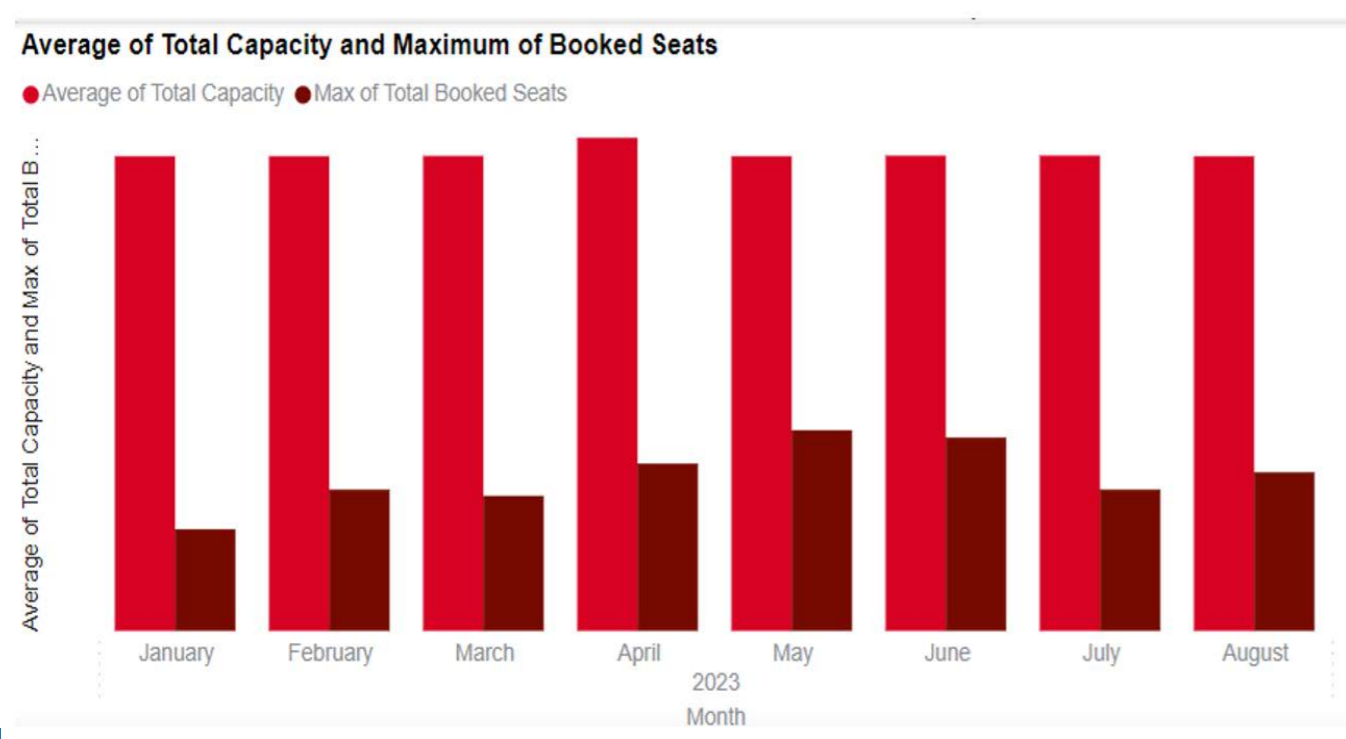
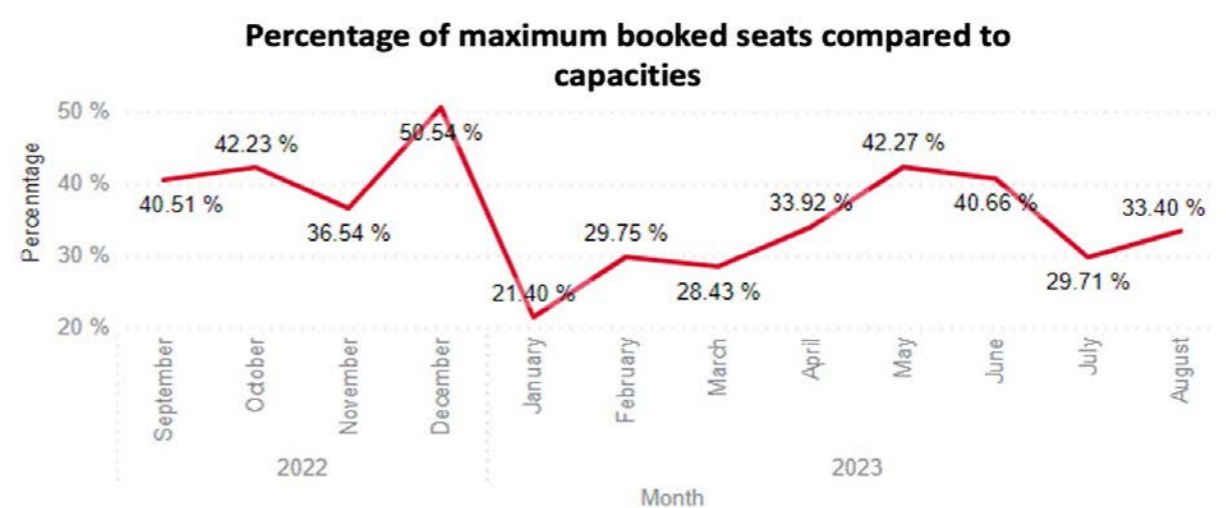
Results and Conclusions



We achieved our primary objective by enhancing our comprehension of the provided datasets and gaining deeper insights into the distribution of data across various attributes.



In relation to the secondary objective, challenges arose due to data quality issues; nevertheless, we effectively communicated noteworthy insights to the project owner regarding the process of data merging



TrainNumber

- 22
- 228
- 229
- 23
- 26
- 27
- 28

TRAIN	Year	Month	Total Capacity	Second Class capacity	Business Class capacity	First Class capacity
228	2022	October	390	335	0	55
228	2022	December	403	328	0	75
228	2023	January	383	328	0	55
228	2023	February	383	328	0	55
228	2023	March	383	328	0	55
228	2023	April	383	328	0	55
228	2023	May	383	328	0	55
228	2023	June	383	328	0	55

TrainNumber

- 22
- 228
- 229
- 23
- 26
- 27
- 28

TRAIN	Year	Month	Total Capacity	Second Class capacity	Business Class capacity	First Class capacity
228	2023	March	383	328	0	55
228	2023	March	390	335	0	55