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 implications.
- Data Understanding: challenges in grasping data source structures and relationships due to incomplete documentation.
- Data Cleaning:
  - Data Conversion: handling missing values, cleaning unrelated attributes.
  - Data Merging: difficulties merging the different data sources, such as Accounting, ARES, Capacity.

Data Understanding & Preparation

- Analyze data and its characteristics
- Data Cleaning and Conversion
- Handling Missing Data and Data Aggregation
- Identifying Passau borders

Workflow

- Analyze data
- Data Cleaning using Python
- Handling Missing Data
- Data Aggregation
- Data Merging (lookup tables)
- 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.

Average of Total Capacity and Maximum of Booked Seats

Percentage of maximum booked seats compared to capacities

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Total Capacity</th>
<th>Second Class Capacity</th>
<th>Business Class Capacity</th>
<th>First Class Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2022</td>
<td>October</td>
<td>390</td>
<td>326</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>2022</td>
<td>December</td>
<td>403</td>
<td>328</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>2023</td>
<td>January</td>
<td>383</td>
<td>328</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>2023</td>
<td>February</td>
<td>383</td>
<td>328</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>2023</td>
<td>March</td>
<td>383</td>
<td>328</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>2023</td>
<td>April</td>
<td>383</td>
<td>328</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>2023</td>
<td>May</td>
<td>383</td>
<td>328</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>2023</td>
<td>June</td>
<td>383</td>
<td>328</td>
<td>0</td>
<td>55</td>
</tr>
</tbody>
</table>