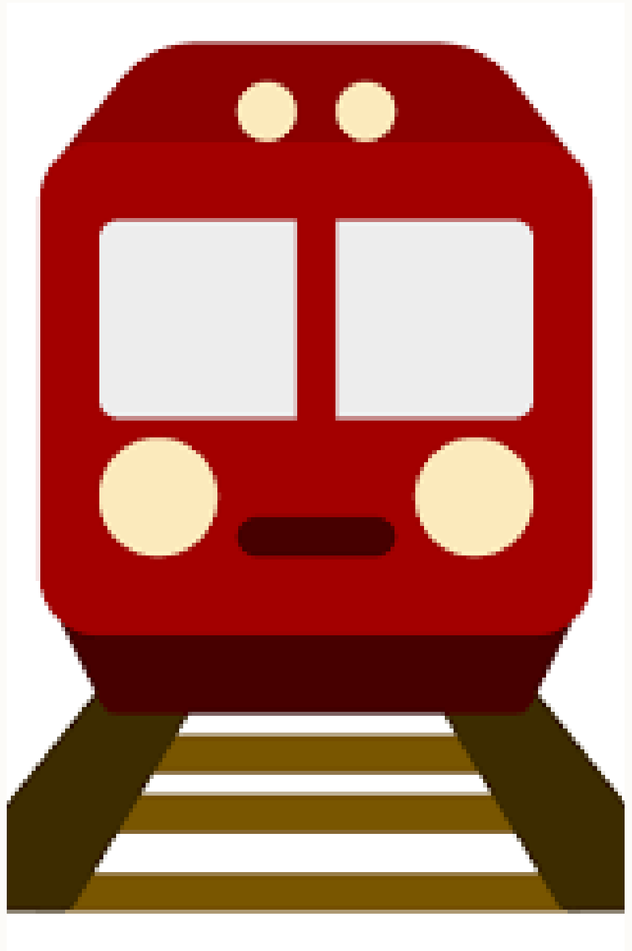
UK Train Rides Analysis



Introduction :

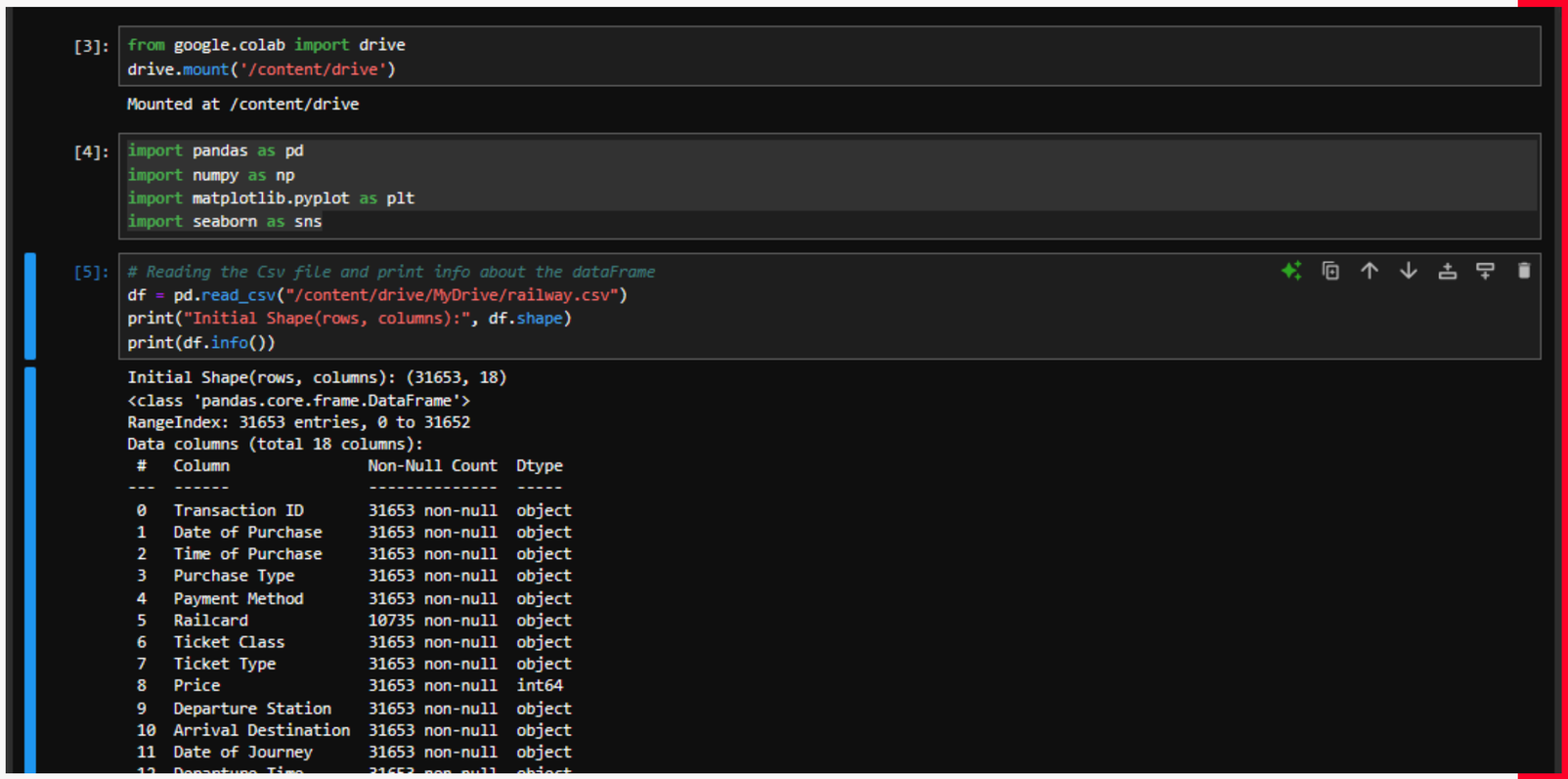
- This project focuses on analyzing a comprehensive dataset of UK train rides, which includes over 31,000 records and 18 attributes related to ticket sales, journey status, passenger details, and travel conditions. The goal is to extract meaningful insights, identify patterns and bottlenecks, and visualize trends that can support operational decision-making and enhance service quality.

Objectives

- Explore and understand the key factors affecting train journeys in the UK.
- Identify common reasons for journey delays and cancellations.
- Analyze sales patterns based on ticket types, passenger classes, and payment methods.
- Create intuitive, interactive dashboards to support business and transport stakeholders.

Data Cleaning & Preprocessing

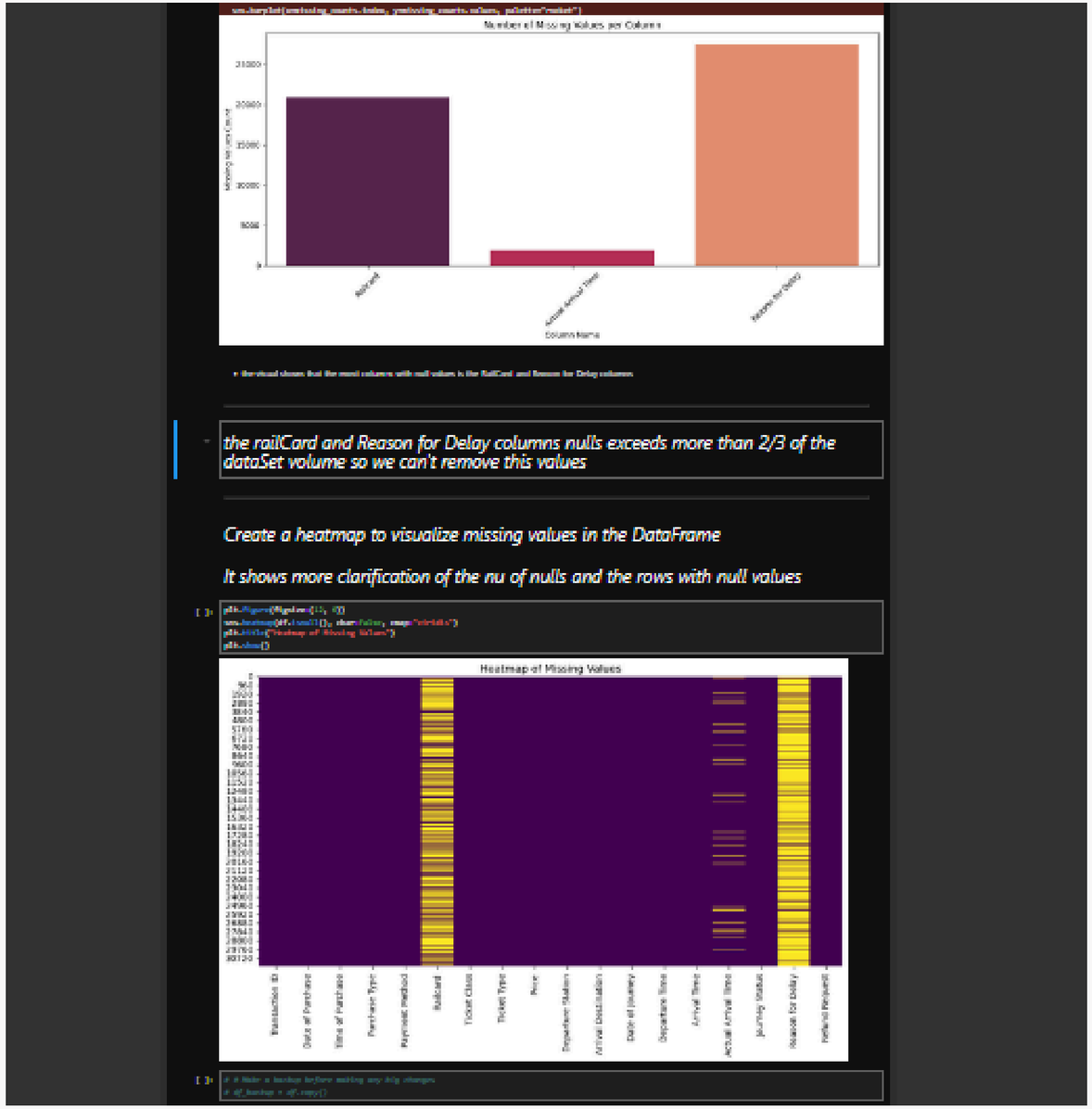
- Tools used: Pandas, NumPy , Matplotlib
- Handled missing values
- Converted date/time columns
- Corrected outliers



```
[3]: from google.colab import drive
     drive.mount('/content/drive')

     Mounted at /content/drive

[4]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns

[5]: # Reading the Csv file and print info about the dataFrame
     df = pd.read_csv("/content/drive/MyDrive/railway.csv")
     print("Initial Shape(rows, columns):", df.shape)
     print(df.info())

     Initial Shape(rows, columns): (31653, 18)
     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 31653 entries, 0 to 31652
     Data columns (total 18 columns):
      #   Column              Non-Null Count  Dtype
     ---  ------              --------------  -----
      0   Transaction ID      31653 non-null  object
      1   Date of Purchase    31653 non-null  object
      2   Time of Purchase    31653 non-null  object
      3   Purchase Type       31653 non-null  object
      4   Payment Method      31653 non-null  object
      5   Railcard            10735 non-null  object
      6   Ticket Class        31653 non-null  object
      7   Ticket Type         31653 non-null  object
      8   Price               31653 non-null  int64
      9   Departure Station   31653 non-null  object
      10  Arrival Destination 31653 non-null  object
      11  Date of Journey     31653 non-null  object
```

Exploratory Data Analysis (EDA)

- Heatmap to visualize missing values .
- Histograms for distribution of tickets .
- Scatter plot for outliers detection .

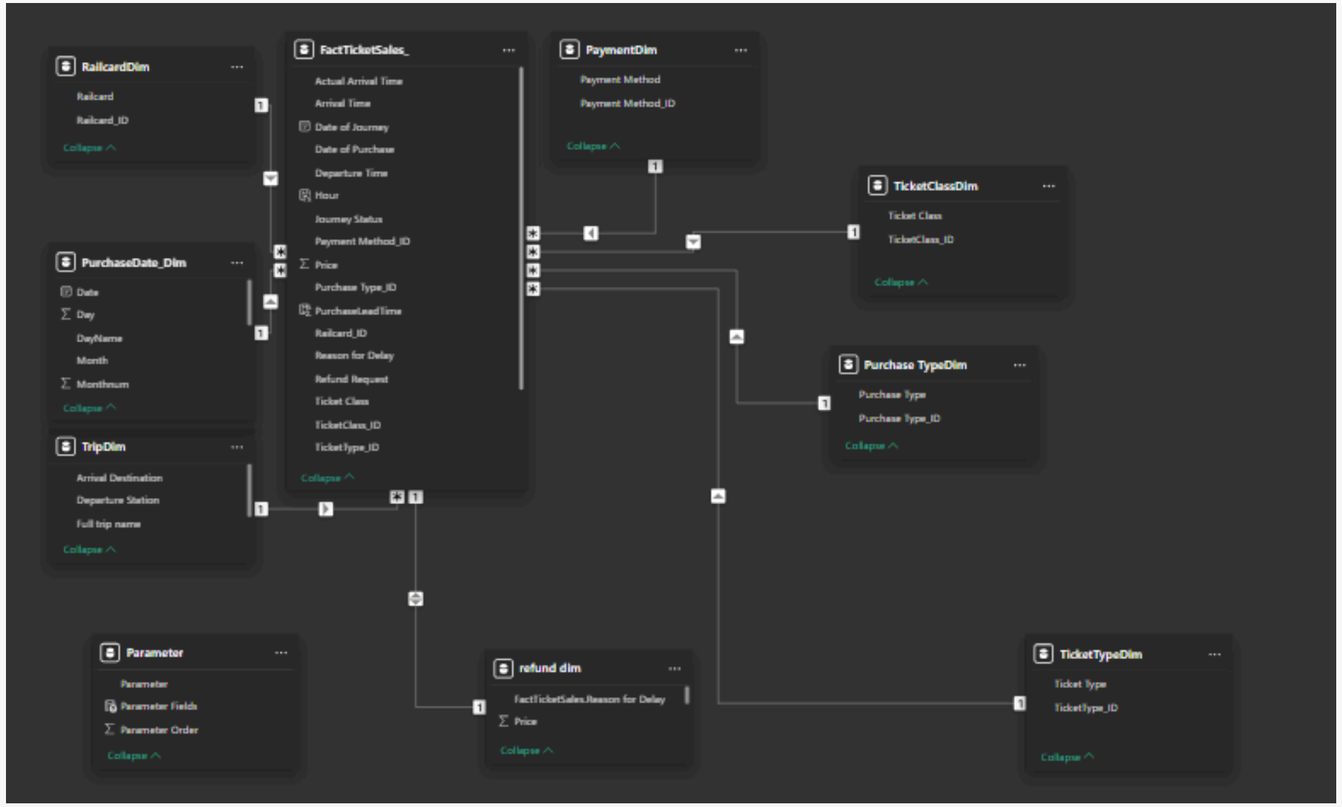


Data Modeling

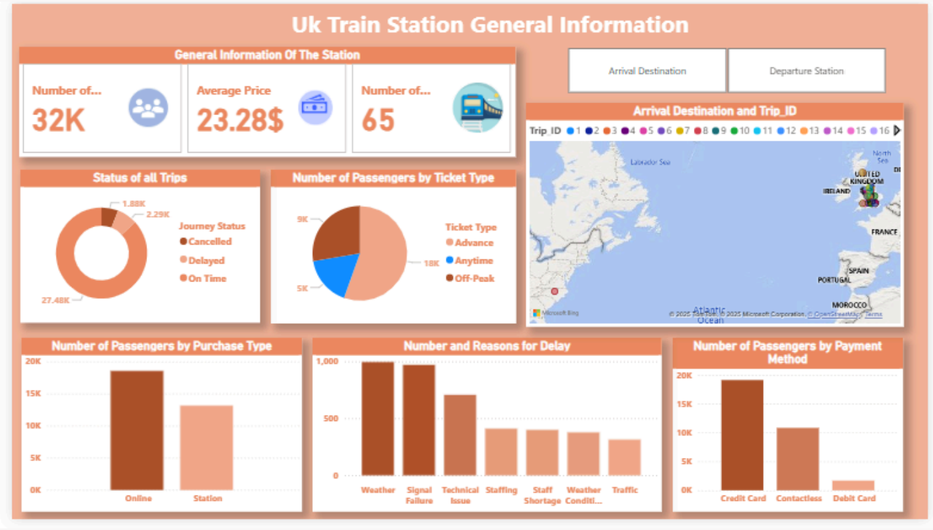
- Used Star Schema

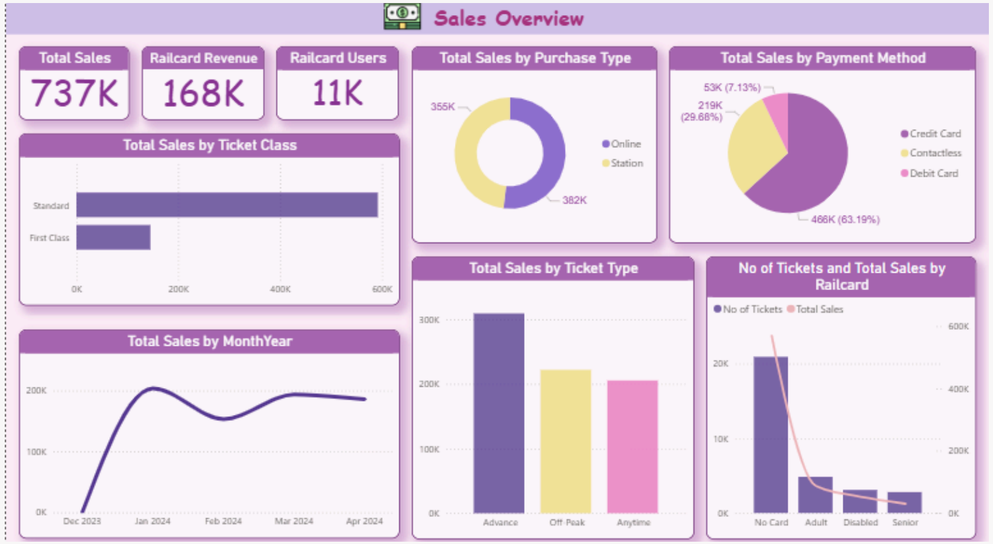
- Fact Table : Ticket Sales / Journey Facts.

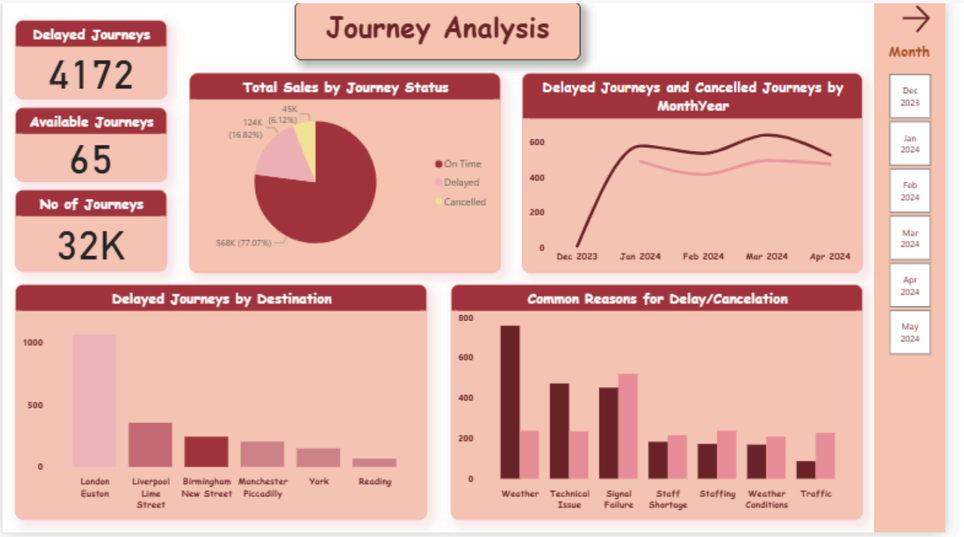
- Dimension Tables : Payment, Railcard, Date, Ticket Type, etc.

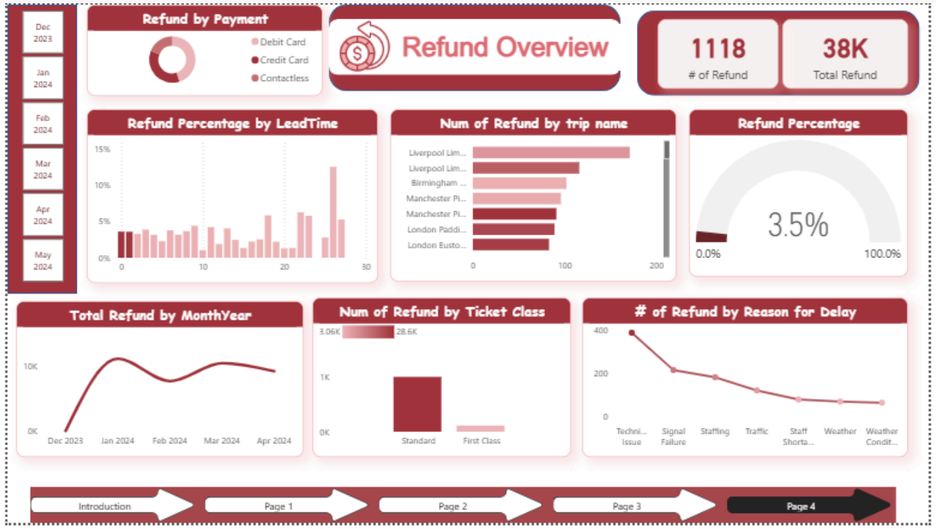


Power BI or Report Pages (4 Pages)



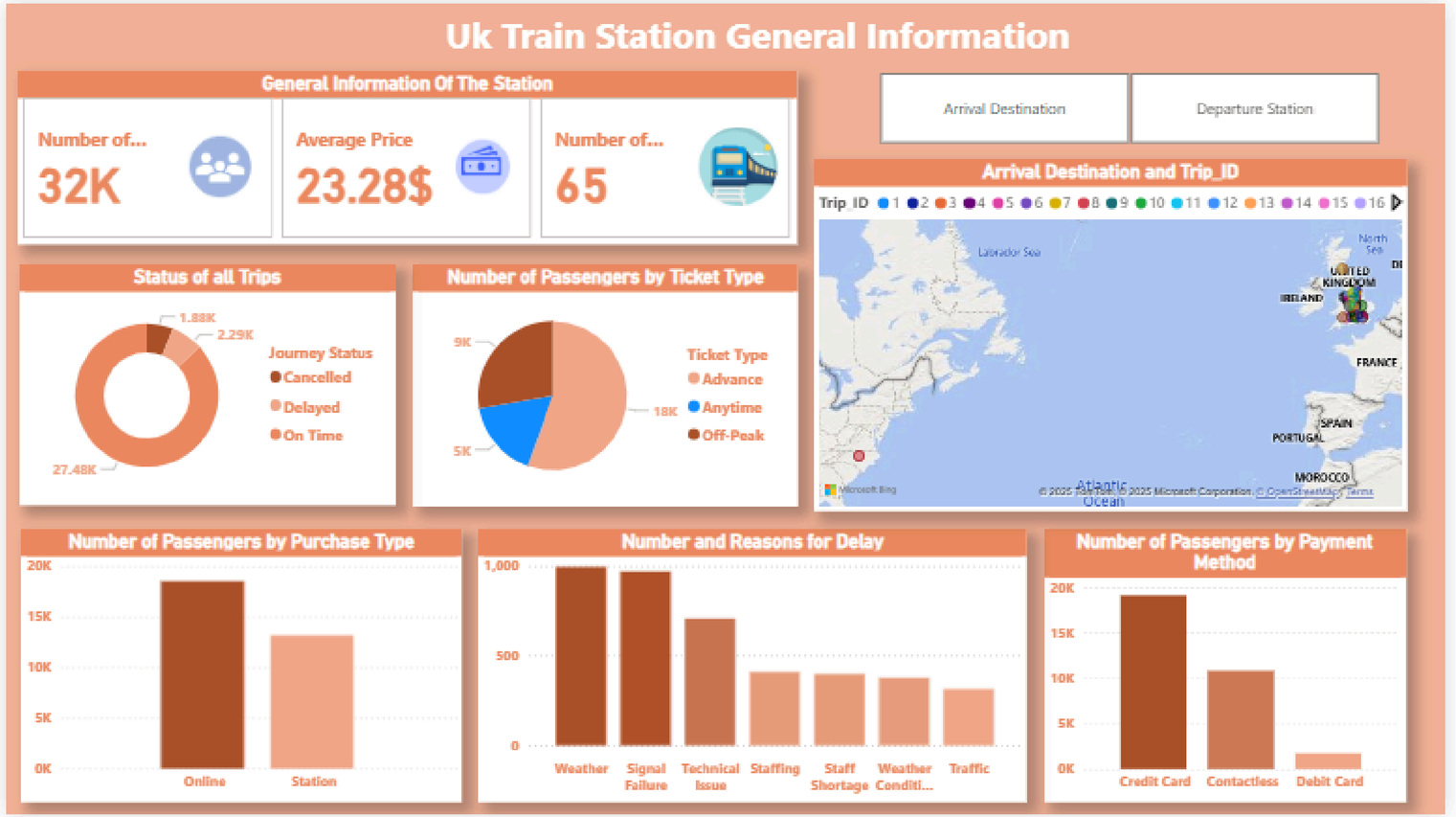






📊 Page 1 : Train Station General Information

- Number of passengers , Avg Price, # of Stations.
- Most used payment types.
- Ticket types distribution.
- Map for destinations.

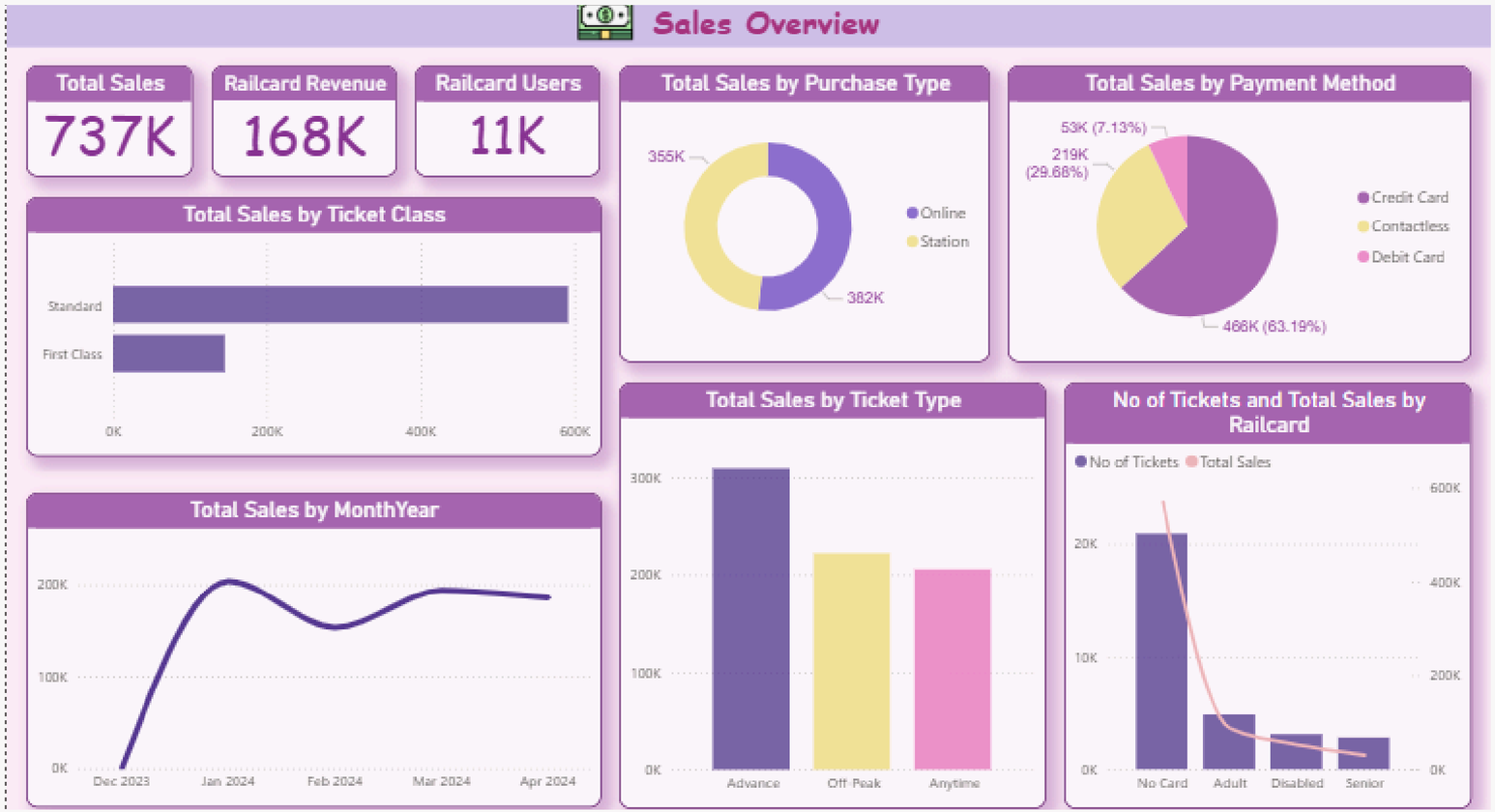


Page 2 : Journey Analysis

- Delayed Journeys (4172).
- Trends overtime (delays by month).
- Common delay reasons

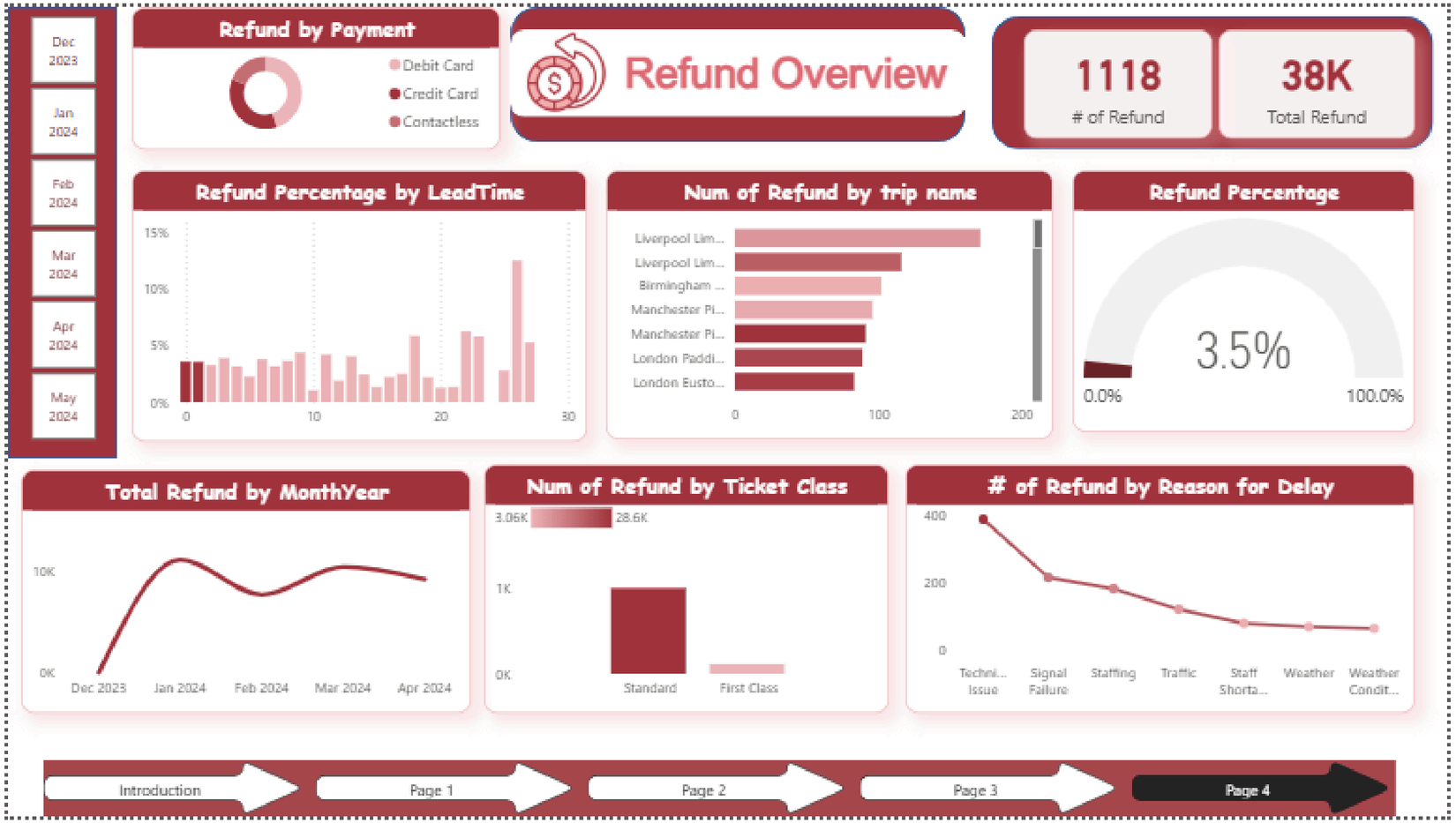
Page 3 : Sales Analysis

- Total Sales.
- Purchase & Payment Behavior.
- Ticket Preferences.
- Sales Trend.



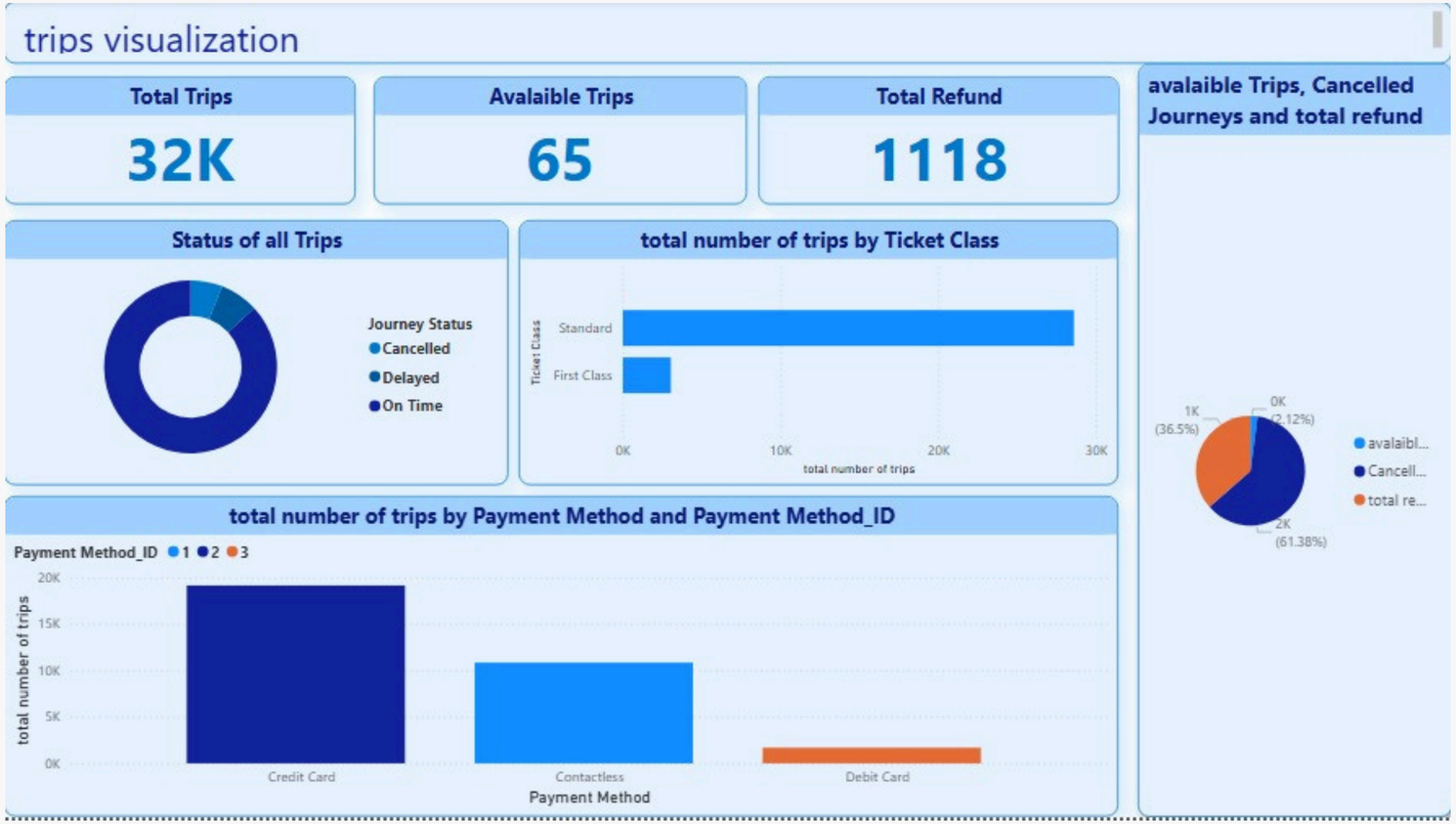
Page 4 : Refund Analysis

- Refund Statistics.
- Refund Timing.
- Payment Method for Refunds.
- Delay Reason Impact.



Additional Page : Trips Analysis

- Trips Statistics .
- Status of all the Trips .
- Nu of trips by tickets Class.



📊 Key Insights & Recommendations

🚦 Delay Analysis

- Weather is the most frequent cause of delays, especially during winter months.
- London Euston and Birmingham New Street record the highest number of delayed journeys.
- Most delays occur during evening peak hours (5 PM – 7 PM).

Recommendations

- Implement automated alerts for weather-affected regions to adjust timetables proactively.
- Invest in infrastructure resilience in weather-sensitive zones.

💳 Payment & Ticketing Trends

- Credit card is the most used payment method, followed by mobile payments.
- Single tickets are purchased more frequently than return or group tickets.
- Ticket purchases spike during Monday mornings and Friday afternoons.

Recommendations:

- Promote contactless and mobile payments with incentives (e.g., discounts).