UK Train Rides Technical Report

Prepared by:
1. Youssef Ahmed
2. Yomna Saad Abdelazeem

Tools Used:
Python (Pandas, NumPy, Seaborn, Matplotlib), Power BI.

Abstract

This report presents an in-depth analysis of a UK railway dataset with the aim of identifying delay causes, understanding travel behavior, and forecasting demand. A structured data science pipeline was implemented, including data cleaning, exploratory analysis, modeling, and visualization. Forecasts of ride volume and revenue by ticket class were also provided.

1. Introduction

The performance of train services in the UK is a key factor influencing commuter satisfaction and operational efficiency. This project analyzes historical ride data to uncover patterns in delay incidents and forecast future demand, which can support strategic decision-making.

2. Methodology

The tools used include Python for data processing and modeling, and Power BI for interactive dashboard visualization.

3. Data Preprocessing

3.1 Dataset Overview:

The dataset includes information on ticket purchase time, journey dates, departure/arrival times, delays, and ticket types.

3.2 Cleaning Steps:

- Removed or imputed missing values in columns like Railcard and Reason_For_Delay.
- Standardized and renamed columns.
- Converted string-based times and dates into datetime objects.
- Derived key features: journey_duration, actual_journey_duration, and delay_time.

4. Exploratory Data Analysis (EDA)

4.1 Delay Analysis:

- Most common delay reasons were identified (e.g., signal failures, weather).
- Delays were more frequent during peak hours and specific weekdays.

4.2 Customer Behavior:

- Different ticket types exhibited varying delay exposures.
- Time-of-day analysis revealed trends in ticket purchases.

4.3 Visualizations:

- Bar charts, heatmaps and line charts illustrated delay patterns and demand fluctuations.

5. Results and Discussion

- The models predicted a steady increase in demand, particularly on weekdays.
- Revenue forecasts showed higher income from first-class tickets, especially during morning hours.
- Identified actionable insights to reduce delays and optimize scheduling.

6. Dashboard and Data Visualization

To enhance the interpretation of the dataset, an interactive dashboard was developed using Power BI. It provides a comprehensive overview of ticketing patterns, journey statuses, and service performance. The dashboard is divided into two pages: Ticket Analysis and Journey Analysis.

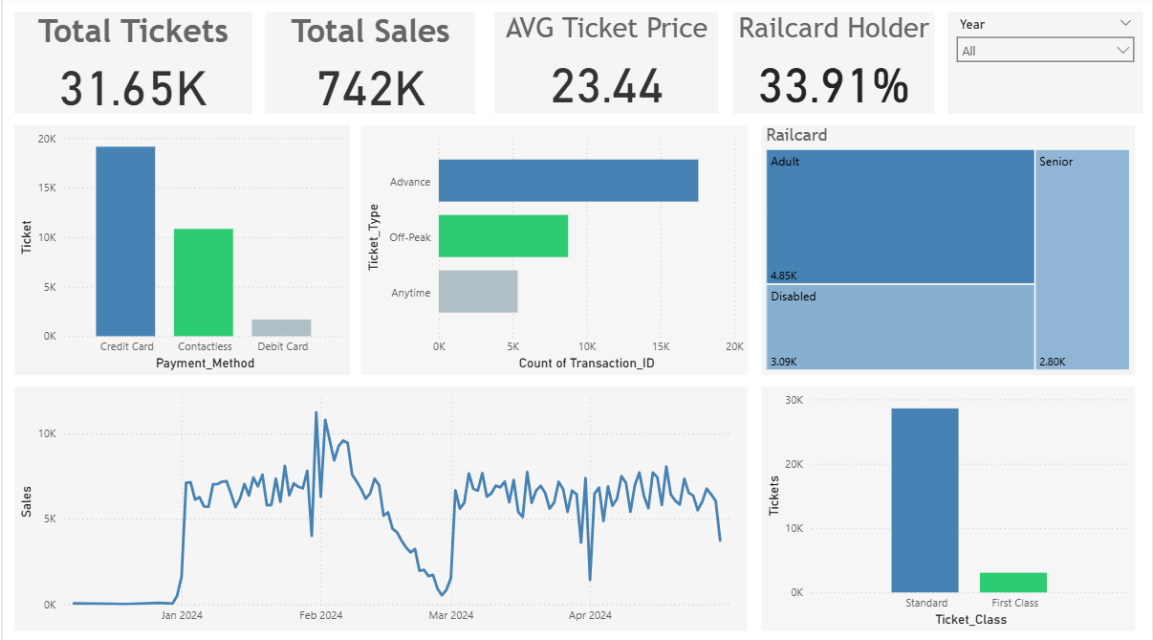
6.1 Ticket Analysis Dashboard

Key Performance Indicators:

- Total Tickets Sold: 31.65K
- Total Sales: 742K
- Average Ticket Price: 23.44 £
- Raicard Holders: 33.91%

Main Visuals:

- Heatmap for Railcards
- Bar charts for payment methods and ticket types
- Sales trend line chart displaying ticket sales over time.
- Ticket class distribution chart (Standard vs. First Class).



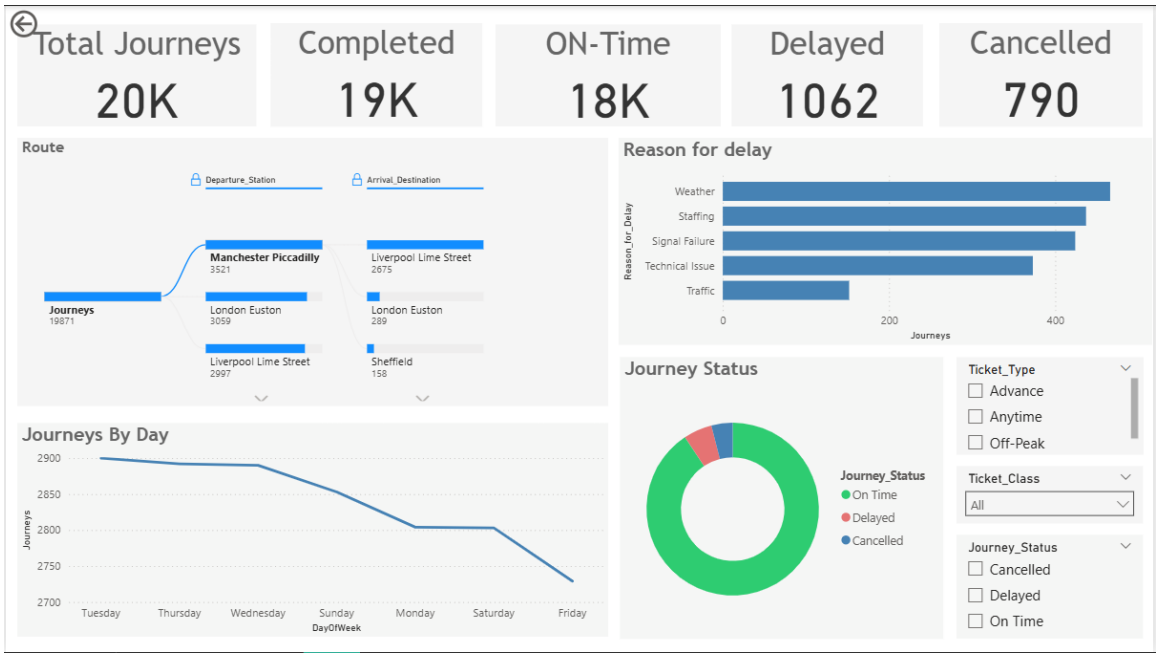
6.2 Journey Analysis Dashboard

Key Metrics:

- Total Journeys: 20K
- Completed: 19K
- On-Time: 18K
- Delayed: 1,062
- Cancelled: 790

Main Visuals:

- Line chart for journeys by weekday
- Delay reason analysis
- Route frequency Decomposition chart
- Journey status donut chart



These dashboards provided meaningful insights for performance tracking, customer behavior understanding, and operations planning.

7. Conclusion

This project successfully processed and analyzed the UK railway dataset, providing insights into delay trends and forecasting future ride volume. These findings can support more efficient train scheduling, targeted marketing, and operational planning.