UK Train Rides Project

Supervised by

Dr. Anas Ismail

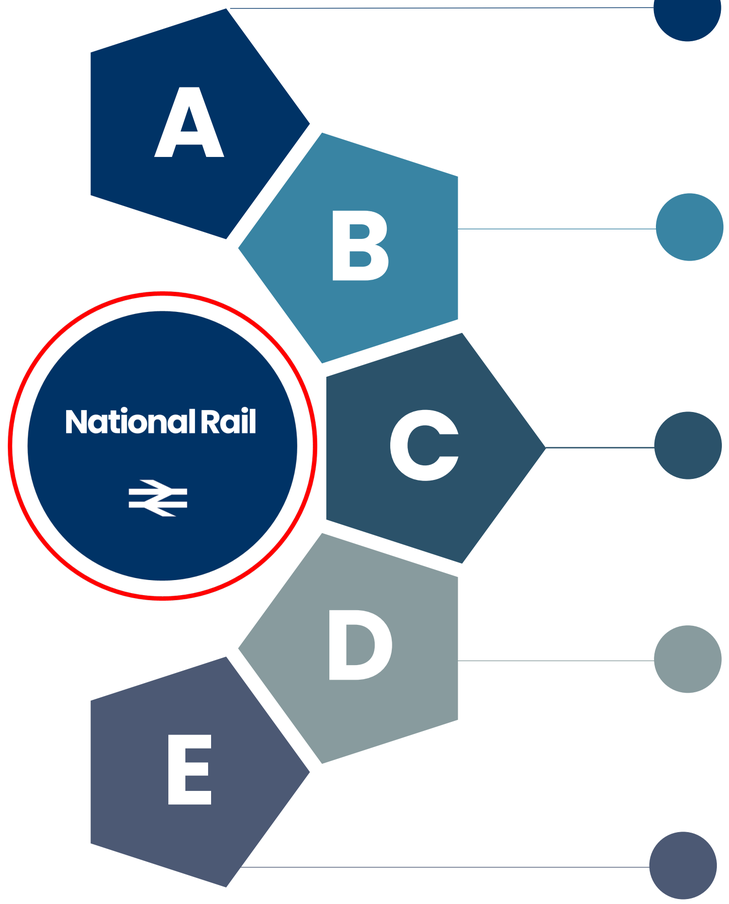
Team Members

- Abdel Moneim Mohamed
- Ashraf Ali
- Kareem Mostafa
- Mohamed Ahmed
- Mohamed Elsourogi
- Mostafa Said El-Ashmawy

Content

- Data Set Description
- Problem Statement
- Analysis Strategy
- Work Stream 1 : Customer Behavior
- Work Stream 2: Train Rides Operations
- Work Stream 3: Revenue Analysis
- Recommendations

Data Set Description



The dataset encompasses comprehensive information on train rides across the UK, capturing various aspects of each transaction with **18 columns** and **31,654 rows** enables in-depth exploration of travel patterns and service performance within the UK rail system

Key columns include the **Transaction ID**, which uniquely identifies each purchase, and timestamps for both the **Date and Time of Purchase**.

The **Purchase Type** and **Payment Method** detail how tickets were acquired, while the **Railcard** and **Ticket Class** provide insight into customer demographics and travel preferences.

Additional columns such as **Price**, **Departure Station**, and **Arrival Destination** reflect the journey specifics, complemented by the **Date of Journey**, **Departure Time**, and **Arrival Time**

The dataset also tracks the **Actual Arrival Time**, and **Journey Status**, allowing for analysis of punctuality, alongside the **Reason for Delay** and any **Refund Requests** submitted

Problem Statement

How to maximize UK railway revenue by:

- ❑ Strengthening the customer loyalty,
- ❑ Attract new customers, and
- ❑ Reducing the number of delayed & cancelled trips.

Analysis Strategy

Customer Behavior

Kareem Mostafa
Mohamed Elsourogi

Understanding customer behavior is essential for enhancing the overall travel experience.

By analyzing transaction data, including purchase types and payment methods, we can identify trends in customer preferences and tailor services accordingly.

Insights gained from this analysis can inform targeted marketing strategies and improve customer engagement.

Train Rides Operations

Ashraf Ali
Mostafa Said El-Ashmawy

Efficient train rides operations are critical for maintaining punctuality and customer satisfaction.

By examining data on departure and arrival times, delays, and journey statuses, we can assess operational performance.

This information helps in identifying bottlenecks and implementing improvements to ensure that services run smoothly and on time

Revenue Analysis

Abdel Moneim Mohamed
Mohamed Ahmed

Revenue analysis plays a pivotal role in the financial health of train services.

By scrutinizing ticket prices, sales data, and refund requests, organizations can gain insights into revenue streams and profitability.

This analysis aids in strategic decision-making, enabling operators to optimize pricing strategies and enhance overall revenue performance.

Work Stream 1: Customer Behavior

Transaction ID	Purchase Type	Payment Method	Railcard	Ticket Class	Ticket Type	Departure Station	Arrival Destination	Date of Journey

Data Understanding

- Understand different variable types, info, and basic Analysis using **Excel**

Data Cleaning

- Apply the routine Data Cleaning using **Power BI**
- Checking data types, missing values, duplicates

Tools

- Apply the Data Analysis using **Power BI**
- Data transformation using **Power Query**
- KPIs, Measures using **DAX**

Work Stream 1: Customer Behavior

Main Objectives

Identify customer behavior and patterns

How 5 items interact with each other

Quantity

Ticket Types **Purchase method**

Time **Location**

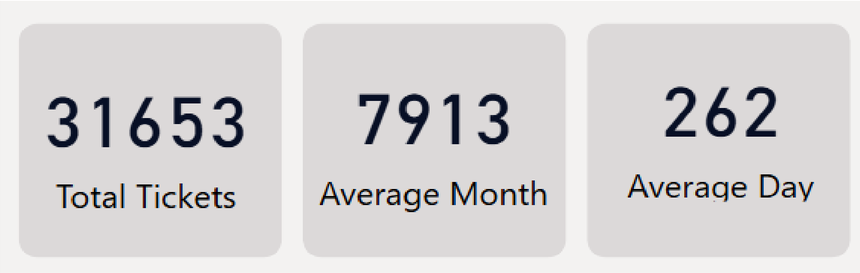
Main Questions

- What is the quantity of tickets?
- What type of tickets prefer?
- How to purchase the tickets?
- When prefer to purchase tickets?
- Where prefer to use stations?

Customer Behavior Insights

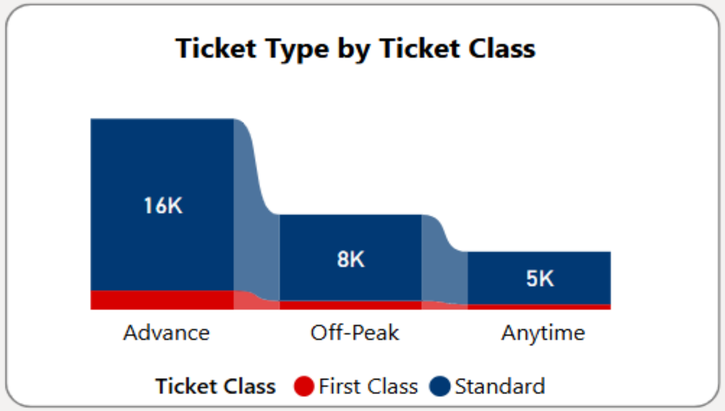
Quantity

Total – Monthly -Daily

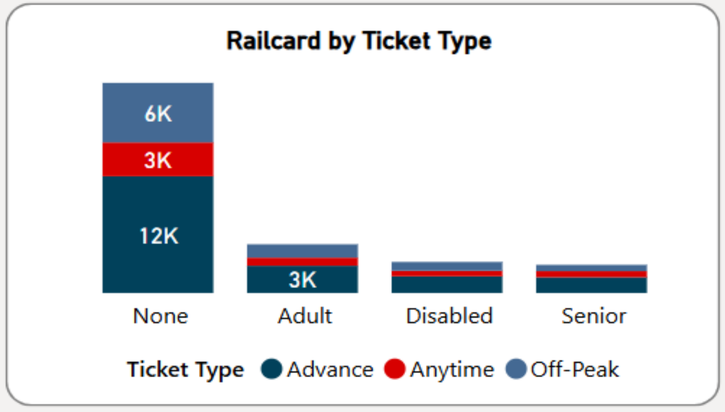


Type

Ticket type and relation
with ticket class



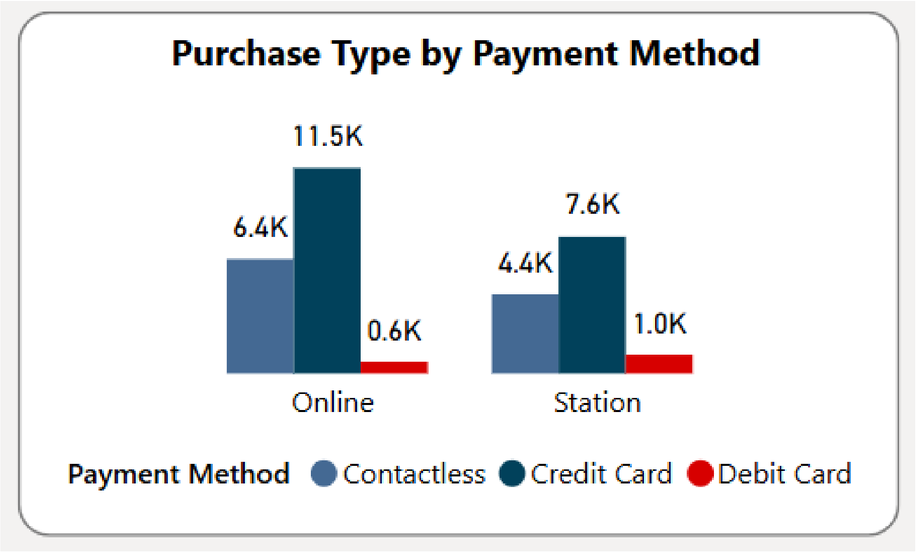
Relation between
Railcard Holder and ticket type



Customer Behavior Insights

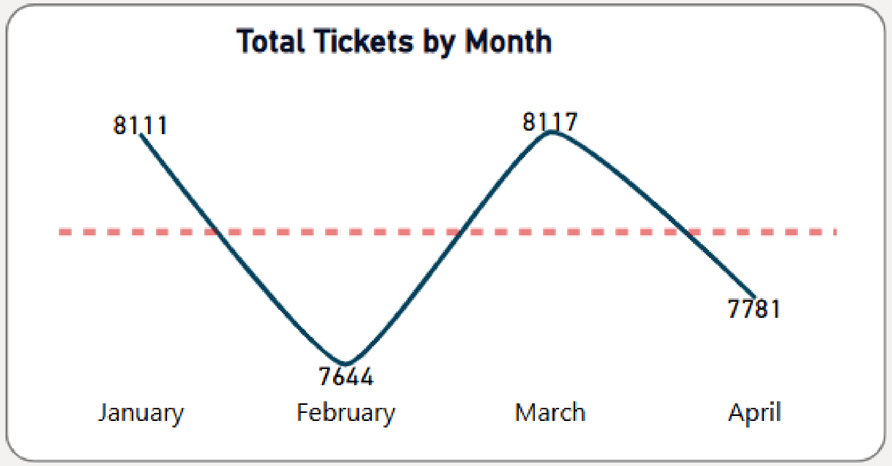
Purchase Method

Relation of Purchase Type by payment Method



Time Analysis

Monthly Trend

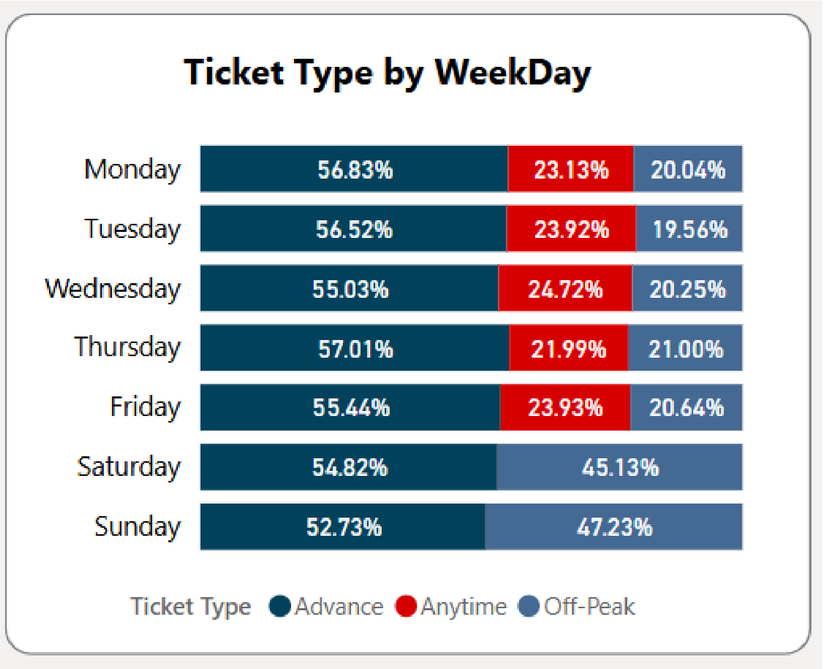


Customer Behavior Insights

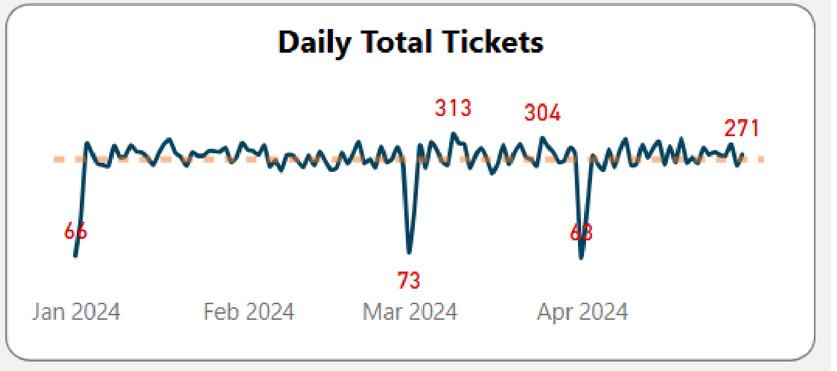
Time Analysis

Week Day – Ticket Class

Interaction between weekday and Ticket class type



Daily Trend

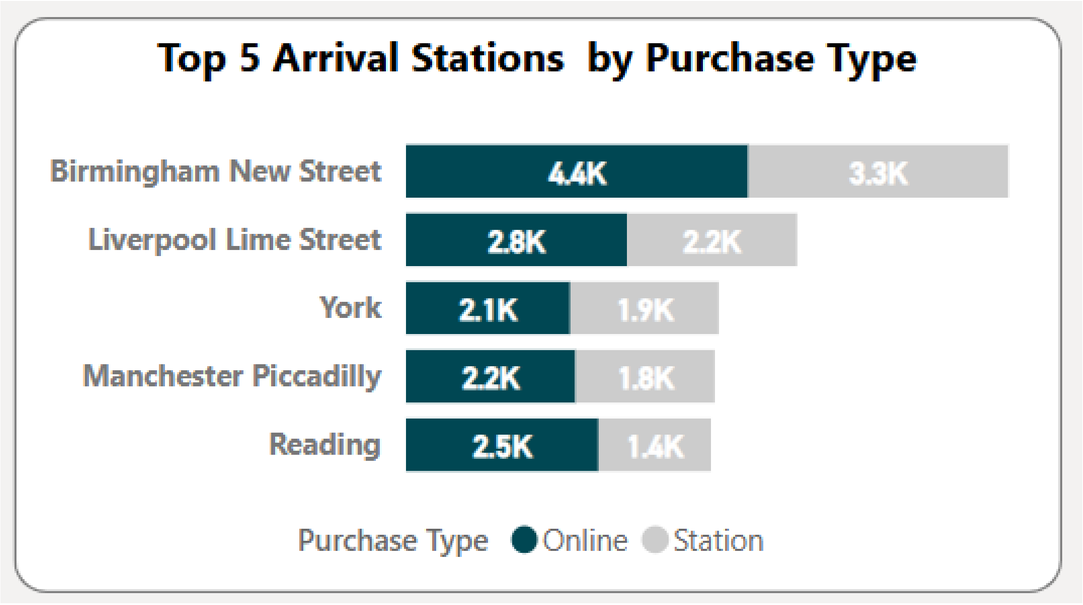


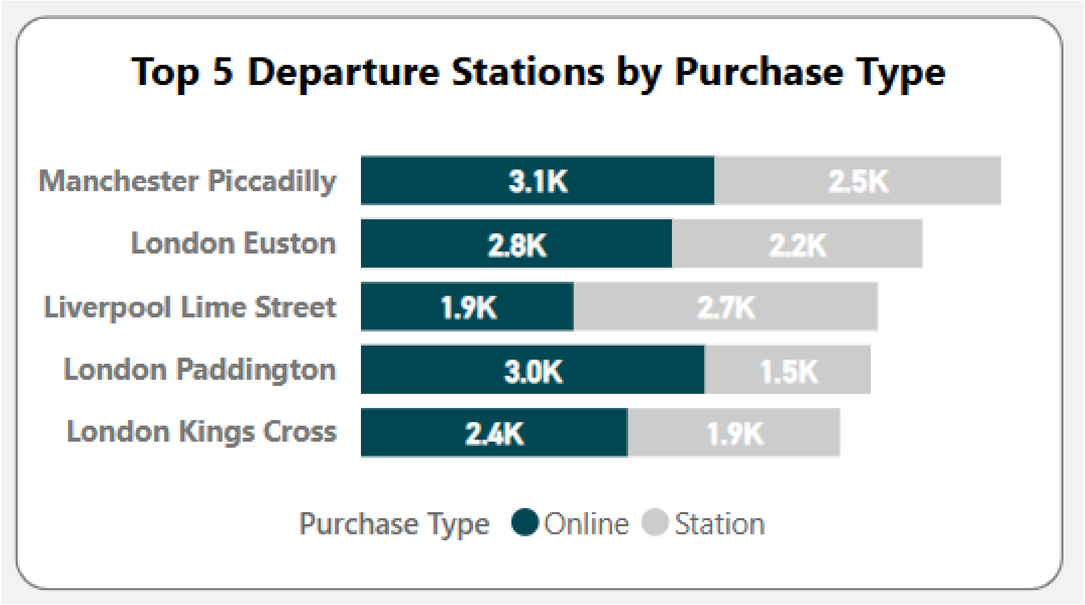
Customer Behavior Insights

Location

Top 5 Departure / Arrival Stations

Connect with purchase type





Work Stream 2: Train Rides Operations

Transaction ID	Departure Station	Arrival Destination	Date of Journey	Departure Time	Arrival Time	Actual Arrival Time	Delay Time	Journey Status	Reason for Delay

Applied routine **Data Cleaning**

Route Delays: What are the patterns in delay times across different routes?

Temporal Delay Variation: How do delays vary by time of day or day of the week?

Delay Reasons: What are the most common reasons for delays, and how frequently do they occur?

Operational Hotspots: Are there specific times or dates with significantly higher operational issues?

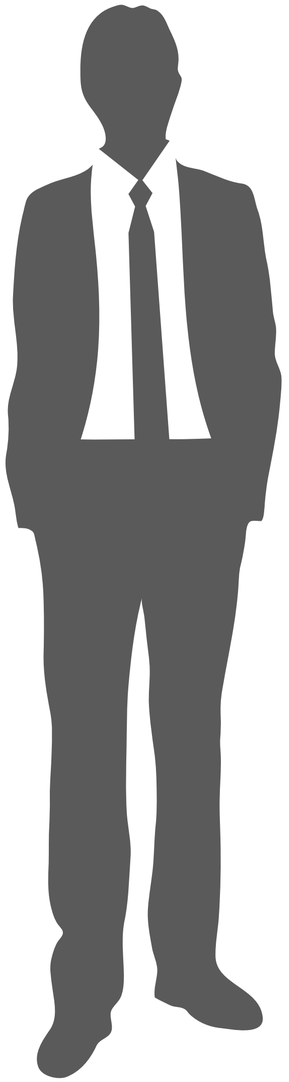
Improvement Opportunities: What operational improvements can be made to reduce delays?

Capacity vs. Delay : What is the relationship between train capacity and delay occurrence?

Frequent Journeys: What are the top 5 frequent journeys?

Problematic Journeys: What are the top journeys with frequent delays and/or cancellation?

Train Rides Operations Insights



Strong Overall Reliability

The vast majority of journeys (>86%) arrived on time, indicating a generally reliable train operation

Time-Based Delay Concentration

Weekdays do not significantly influence the occurrence of delays or cancellations. However, disruptions are notably more frequent during the morning peak hours compared to the afternoon peak

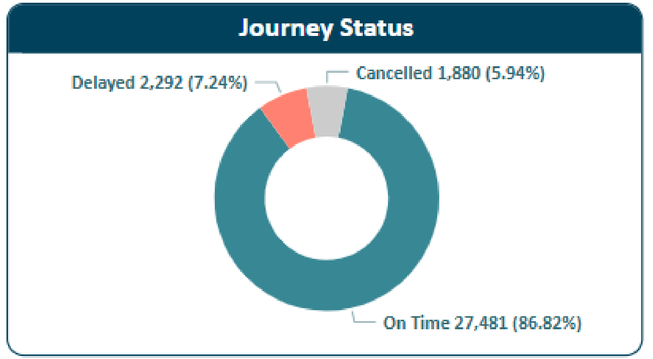
Stations with Elevated Delay Rates

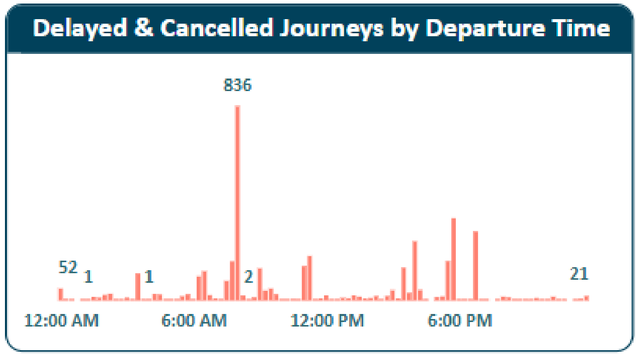
Liverpool Lime Street is the station most affected by delays, with Manchester Piccadilly being the second most impacted

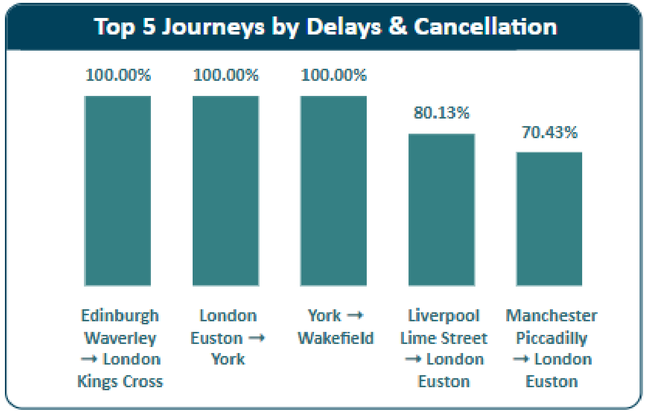
Journeys Prone to Disruption

The ranking of problematic journeys can shift depending on the chosen metric (absolute count, ratio, or revenue impact).

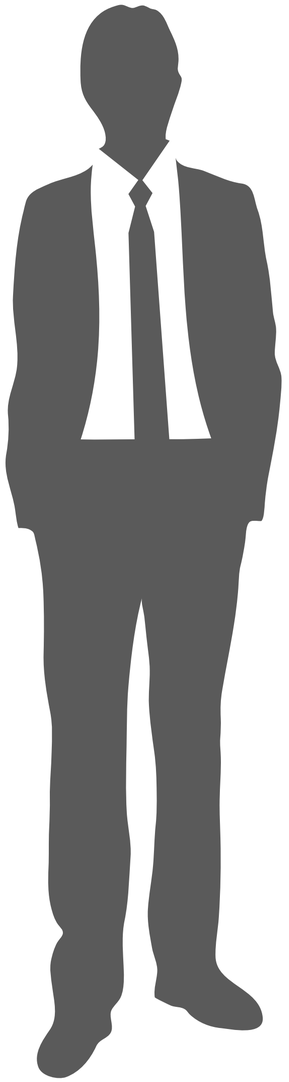
- Liverpool Lime Street to London Euston and Manchester Piccadilly to Liverpool Lime Street experience the highest absolute number of delays and cancellations.
- When considering the ratio of delayed/cancelled journeys to total journeys, Edinburgh Waverley to London Kings Cross, London Euston to York and York to Wakefield show the highest proportion of disruptions.







Train Rides Operations Insights



Dominant Cause of Disruption

Weather conditions are the most frequent factor leading to both delays and cancellations.

High-Impact Delay Factors

Signal issues and staff shortages, while potentially less frequent, result in the longest average delay time (approximately 52 minutes per incident).

Morning Peak Vulnerability at Liverpool Lime Street

A significant number of cancellations and delays at 8 am specifically at Liverpool Lime Street are attributed to weather conditions.

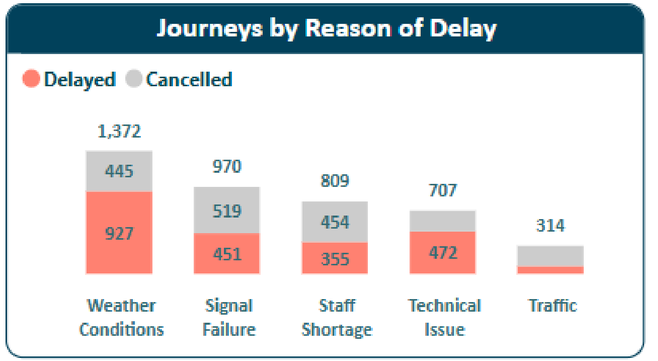
Top Travel Corridors

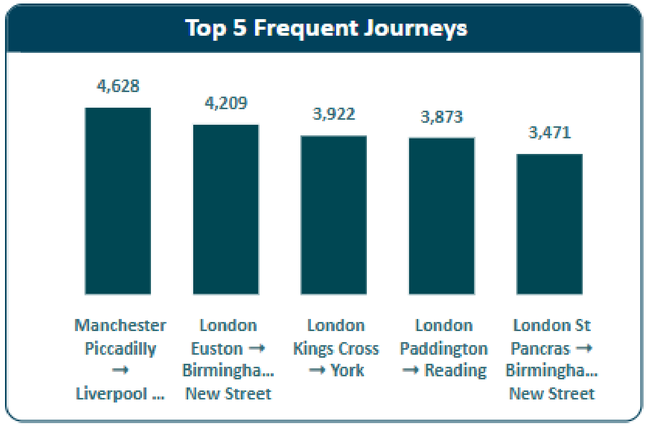
The five most frequently traveled routes are:

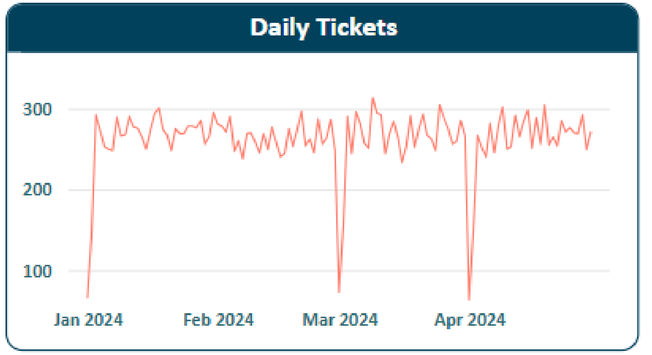
i. Manchester Piccadilly to Liverpool Lime Street
ii. London Euston to Birmingham New Street
iii. London Kings Cross to York
iv. London Paddington to Reading
v. London St Pancras to Birmingham New Street

Potential Data Inconsistencies

Analysis reveals unusual dips in daily ticket numbers during January, March, and April, suggesting potential data errors that warrant further investigation and validation.







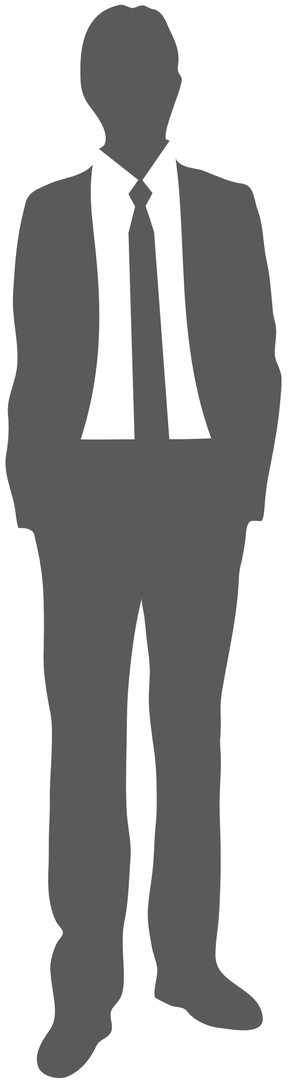
Work Stream 3: Revenue Analysis

Transaction ID	Railcard	Ticket Class	Ticket Type	Price	Departure Station	Arrival Destination	Journey	Date of Journey	Delay Time	Journey Status	Reason for Delay	Refund Request

Applied routine **Data Cleaning**

- What is the overall trend of total sales revenue?
- How does revenue distribution vary across different ticket classes and ticket types?
- What is the revenue contribution of Railcard holders?
- Which routes generate the highest and lowest revenue? And what are the characteristics of these journeys?
- How does revenue fluctuate across different months?
- What is the total financial loss due to refunds?
- What is the percentage of revenue lost due to refunds?
- What are the primary reasons for refunds and their associated financial impact?
- Is there a correlation between delays and refund requests, particularly for specific ticket types?
- Can we identify any patterns or trends in refund requests over time?
- How does delay time correlate with refund requests and customer satisfaction?

Revenue Analysis Insights



Dominance of Standard Class

Standard Class tickets contribute the majority of the revenue (79.86%) despite having a significantly lower average ticket price (£20.72) compared to First Class (£48.86). This suggests high volume in Standard Class sales

Advance Tickets Drive Volume

Advance tickets represent the largest share of revenue (41.69%), albeit with the lowest average ticket price (£17.61). This indicates a strong customer preference for discounted advance purchases

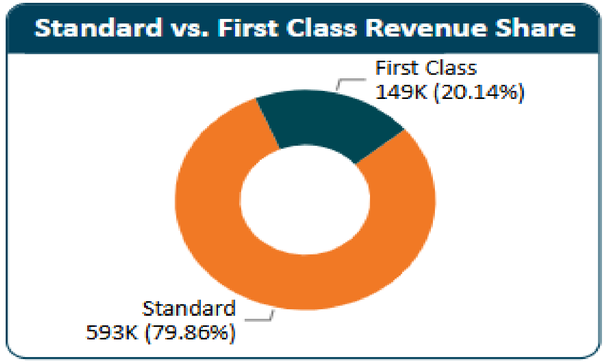
Railcard Impact

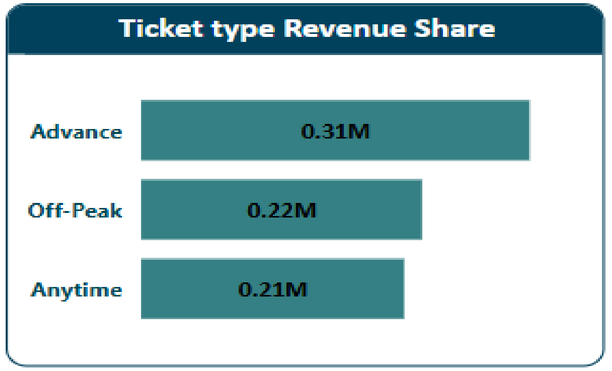
Railcard holders account for a notable portion of the revenue (22.67%), with an average ticket price (£15.67) lower than the overall average, highlighting the importance of this customer segment

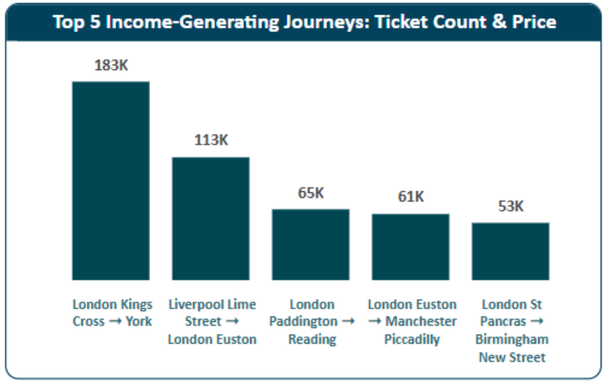
Top Performing Routes

The highest revenue-generating routes are:

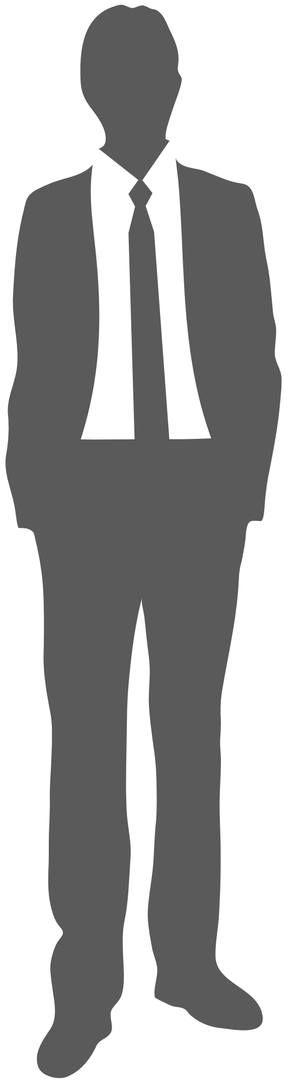
i. London Kings Cross → York
ii. Liverpool Lime Street → London Euston
iii. London Paddington → Reading
iv. London Euston → Manchester Piccadilly







Revenue Analysis Insights



Seasonal Revenue Trends

January exhibits the highest revenue, aligning with the highest number of journeys. Conversely, February shows the lowest revenue and the fewest journeys. This suggests a strong seasonal impact on travel demand

Significant Refund Losses

The company experienced a loss of 5.22% of its total revenue due to refunds. This represents a substantial financial impact that warrants further investigation

Cancellation vs. Delay Refunds

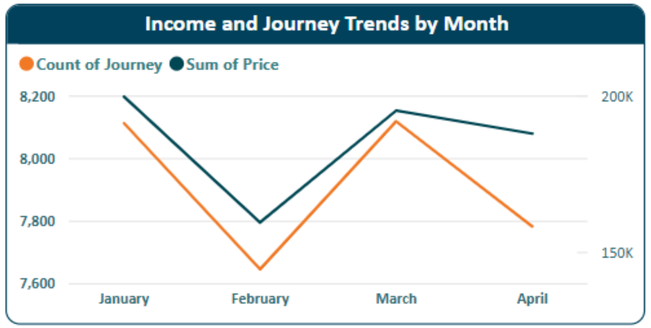
Refunds due to cancellations (51.16% of all refunds) slightly outnumber those caused by delays

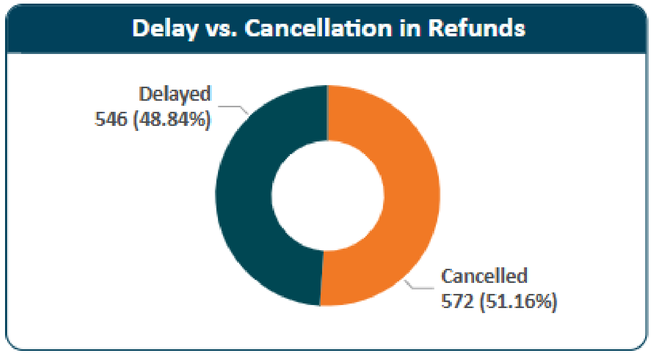
Primary Refund Drivers

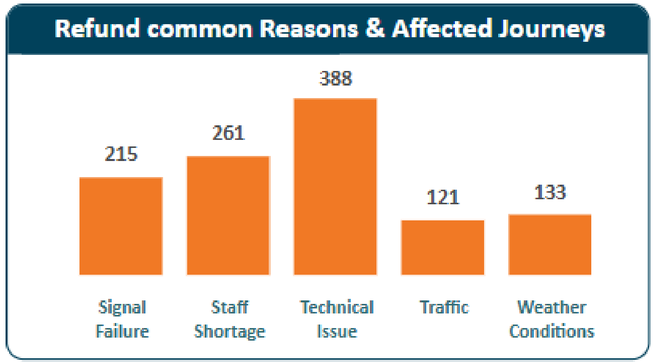
The main reasons cited for refunds are technical Issue and staff Shortage

Advance Tickets and Refunds/Delays

Advance tickets appear to be more frequently associated with refund requests and are also more likely to be affected by delays. This could be due to the longer booking window and potential for disruptions.







Recommendations - Enhancing Revenue and Customer Engagement

Boost Advance Ticket Sales

Capitalize on Advance ticket popularity through intensified online promotions and potentially dynamic pricing strategies

Strategic Ticket Pricing Review

Evaluate the price gap between Standard and First Class to optimize revenue and potentially attract more First Class passengers

Loyalty Program for Railcard Users

Implement a loyalty program or exclusive benefits to incentivize Railcard usage and increase customer spend

Route Performance Optimization (Revenue Focus)

Review underperforming routes for pricing and marketing adjustments to improve revenue generation

External Factor Analysis (Pricing)

Utilize data on holidays and station classifications to implement a more dynamic and responsive pricing strategy

Investigate Sales Decline

Conduct a study to understand the reasons behind the observed decline in ticket sales at the beginning of each month and implement targeted strategies to mitigate it

Recommendations - Improving Operational Efficiency

Invest in Proactive Maintenance

Implement a robust maintenance schedule for infrastructure and trains to minimize technical issues causing delays and cancellations.

Targeted Mitigation Strategies

Develop specific strategies to address the most frequent and high-impact delay reasons

- **Weathe**r: Implement robust communication protocols and alternative transportation plans for weather-related disruptions, especially at vulnerable locations like Liverpool Lime Street.
- **Signal Issues**: Prioritize maintenance and upgrades of signaling infrastructure to minimize failures
- **Staff Shortage**: Review staffing levels and implement strategies to ensure adequate personnel availability, particularly during peak hours

Route-Specific Performance Monitoring (Operations)

Continuously monitor top problematic routes to identify recurring operational issues and implement targeted improvements

Proactive Weather Management

Engage with meteorologists to develop predictive models for weather conditions, particularly for Liverpool Lime Street station during the 8 am peak. This will enable proactive measures to mitigate potential delays and cancellations

Recommendations - Enhancing Data Insights and Analysis Capabilities

Detailed Profitability Analysis

Request granular data on revenue, expenses, and operating costs per journey for deeper profitability analysis

Comprehensive Weather Data Analysis

Obtain a full year of data to accurately assess the impact of weather patterns on delays and cancellations across all seasons. The current data represents a limited timeframe.

Enhance Data Granularity (Trip Itinerary)

Incorporate trip itinerary data into future data collection efforts. This more detailed information can facilitate a deeper analysis and the identification of more granular performance indicators

Further Analysis with Capacity Data

Once train capacity data is available, conduct an analysis to understand its potential correlation with delay occurrences

Thank You