****

***UK Train Rides***

*Technical Report*

*Date of submission : April 2025*

*Team Name: Data Wizards*

*Our team:*

*.* Majid Ashraf Abd El Majed

. Mostafa Hossam Esmail Abo Bakr

. Mohamed Khaled Mohamed

. Basmala Khaled kamal

. Nada Ahmed Nour Eldeen

1. **Introduction**

In the dynamic landscape of the United Kingdom's rail network, understanding passenger behavior and operational efficiency is paramount This data aims to support the analysis of train performance over time, understand the causes of delays, and identify potential passenger behavior and preferences based on purchase methods. It captures a wealth of information, including payment methods (online versus station purchases), ticket pricing based on departure and arrival stations, journey durations, delay statuses with corresponding reasons, and the presence or absence of refund requests

1. Bottom of Form
2. Top of Form

Bottom of Form

Objective

The primary objective of this analysis is to evaluate the performance of train services across the United Kingdom. This includes analyzing patterns in delays, identifying causes of cancellations, and understanding passenger behavior related to ticket purchases. The insights derived aim to enhance service quality, improve operational efficiency, and inform strategic decision-making.

### Tool Used

The analysis was conducted using Microsoft Power BI, a robust business analytics platform. Power BI enabled efficient data cleaning, transformation, normalization, modeling, and visualization, providing a scalable and structured approach to derive actionable insights from the dataset.

## 2. Overview

The dataset comprises 31,653 rows and 18 columns, capturing detailed information about train ticket purchases in the United Kingdom. Key attributes include:

* **Transaction Details**: Payment method (online or at the station), ticket price, departure and arrival stations.
* **Journey Information**: Departure time, actual arrival time, journey status (e.g., on time, delayed, cancelled), and delay reasons.
* **Passenger Behavior**: Refund requests and purchase methods.

This dataset supports the analysis of train service performance over time, identification of delay causes, and exploration of passenger preferences based on ticket purchase patterns.

## 2. Business Question

Below are some key questions that the dashboard will answer, which will help in analyzing performance and making data-driven strategic decisions.

**1**. **Overview Dashboard Business Questions:**

How many train trips were completed during the analysis period?

What percentage of trips were delayed or cancelled?

Is the overall performance improving or declining over time?

**2. Delayed & Cancelled Dashboard Business Questions:**

Which months had the highest rates of delayed or cancelled trips?

Which routes experienced frequent service disruptions?

Are there any seasonal patterns or recurring trends in delays?

**3. Route Analysis Dashboard Business Questions:**

Which routes have the highest and lowest number of trips?

Which routes show the best on-time performance?

Are there routes that require resource reallocation or service improvements?

**4. Revenue Dashboard Business Questions:**

Which ticket types generate the most revenue?

How is revenue distributed across months or routes?

Do delays and cancellations negatively impact revenue or refund rates?

**5. Travel Behavior Dashboard Business Questions:**

What are the most common travel times and days?

Which ticket types are preferred by passengers?

How does customer behavior change in response to frequent delays?

Top of Form

Bottom of Form

Top of Form

Bottom of Form

Top of Form

Bottom of Form

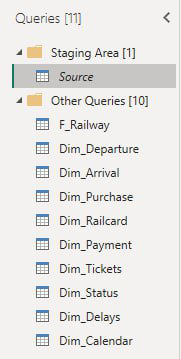
Top of Form

Bottom of Form

## 3. Data Cleaning

The data cleaning process was systematic and thorough to ensure data integrity and reliability for analysis.

### 3.1 Creating a Staging Area

To preserve the original dataset, a staging area labeled "Original Dataset" was created. This ensured that all modifications could be tracked, and the unaltered data remained accessible for reference or reversion if needed.

### 3.2 Creating a Date Table

A dedicated date table was created to support time-based analysis:

* A Blank Query was used in Power Query.
* M code was written to dynamically determine the start and end dates based on:
  + Date of Purchase: Earliest date extracted.
  + Date of Journey: Latest date extracted.

This table ensures consistency in time intelligence calculations across the model.

### 

### 3.3 Modifying Column Headers and Assigning Data Types

The F\_Railway table was cleaned as follows:

* **Header Standardization**: Applied "Use First Row as Headers" to ensure consistent column names.
* **Data Type Assignment**:
  + Date of Purchase and Date of Journey: Changed to Date.
  + Price: Changed to Whole Number.
  + Departure Time and Actual Arrival Time: Changed to Time.

### 3.4 Cleaning and Standardizing Delay Reason Column

The Delay Reason column was standardized to ensure consistency:

* Unified Inconsistent Phrasing:
  + "Signal failure" and "Signal Failure" → Signal Failure.
* Merged Similar Reasons:
  + "Weather" and "Weather Conditions" → Weather Conditions.
  + "Staff" and "Staff Shortage" → Staff Shortage.
* Handled Missing Values:
  + Null values were replaced with "No Delay Reported".

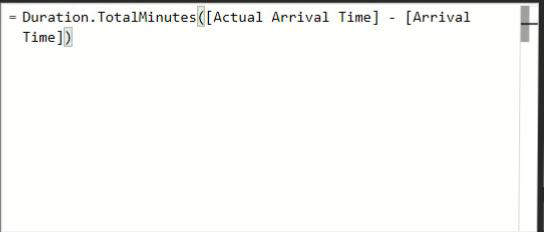
### 3.5 Handling Empty Values in Actual Arrival Time

* **Null Values**: Identified as representing cancelled journeys.
* **Retention**: Null values were retained to preserve the Time data type.
* **Documentation**: Noted that Null = Cancelled Journeys for clarity in subsequent analysis.

### 3.6 Creating a Delay Duration Column

A custom column was created to calculate delay duration:

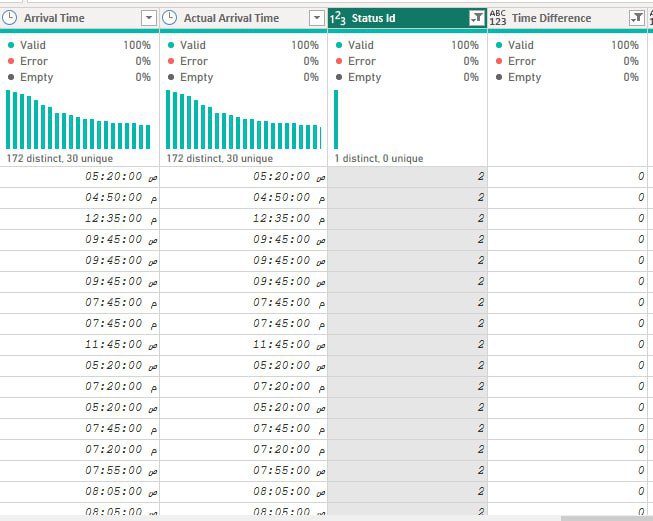
* **Method**: Used Duration.TotalMinutes() in Power Query to compute the time difference between expected and actual arrival times.

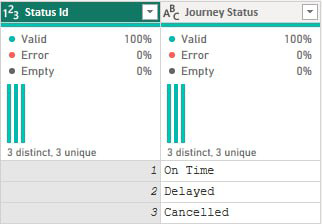


* **Naming**: Renamed the column to Delay Minutes.
* **Outlier Review**:
  + Identified 18 journeys marked as "Delayed" but with 0 delay minutes.
  + Updated their status to "On Time" to reflect accurate performance.

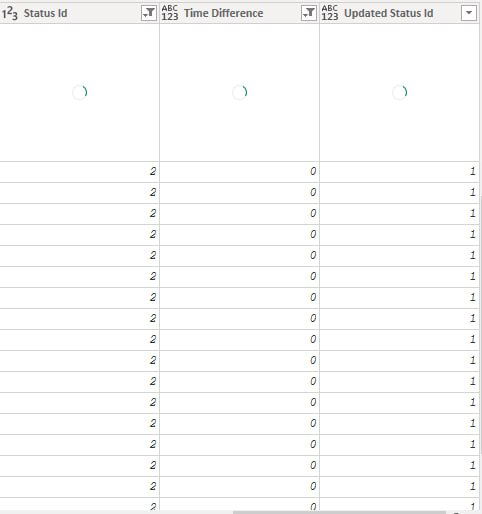
### 3.7 Correcting Journey Status Based on Actual Arrival Time

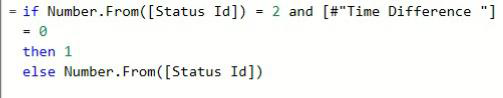
* **Issue Identification**: Journeys marked as Delayed with a 0-minute delay were flagged as inconsistent.





* **Resolution**:
  + Created a conditional column, Updated Status Id, to reclassify these journeys as On Time.
  + Replaced the original Status Id with the corrected column.

Renamed the new column to Status Id for consistency.



### 

### ****3.8 Updating Delay ID Based on Delay Minutes****

To ensure consistency with the adjustments made to the Journey Status ID, the **Delay ID** column was updated using the **Delay Minutes** column according to the following logic:

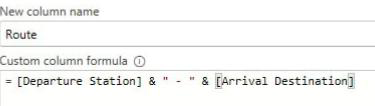
* If **Delay Minutes = 0**, then **Delay ID** is set to 1 (indicating "No Delay Reported").
* Otherwise, retain the original **Delay ID** value.

**Purpose:**  
This step ensures logical alignment between the journey status and delay reason, especially in cases where trips were originally marked as "Delayed" despite having zero actual delay.

### 

### 3.9 Creating a Route Column

Before normalization, a custom column named Route was created in Power Query to enhance analysis:

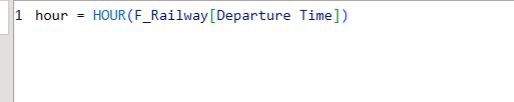
* **Purpose**: To concatenate the Departure Station and Arrival Station for a unified route identifier (e.g., "London - Manchester").
* **Method**:
  + Used M code in Power Query to create the column.
  + Formula: Text.Combine({[Departure Station], " - ", [Arrival Station]})
* **Benefit**: Simplifies route-based analysis and visualization, providing a clear representation of journey paths.

### Creating a Hour Column

Before normalization, a custom column named Route was created in Power Query to enhance analysis:

**Purpose:** The purpose of creating the Hour column is to extract the hour component from the Departure Time field. This enables the analysis of travel patterns across different hours of the day.

**Method:** The column was created using a DAX formula:



This formula returns the hour (from 0 to 23) for each train ride based on its departure time. The column was added as a calculated column within the dataset.

**Benefit:** The Hour column allows for time-based analysis and visualization, such as identifying peak travel hours or low-traffic periods. It enables the creation of insightful hourly charts that support decision-making and scheduling improvements in train operations.

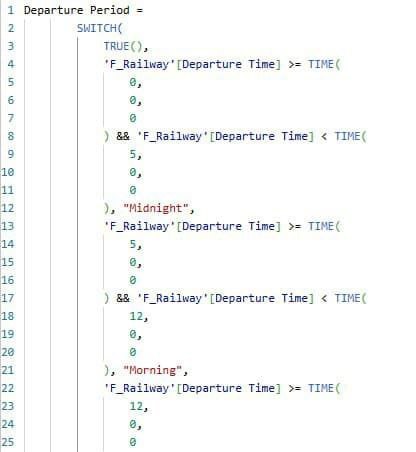
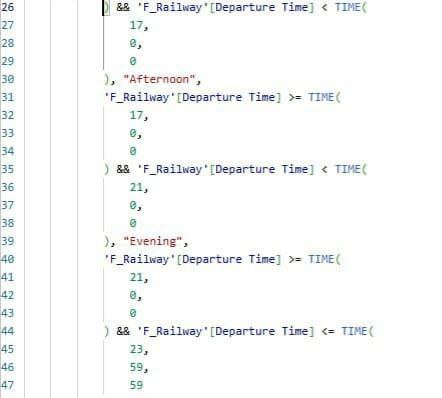
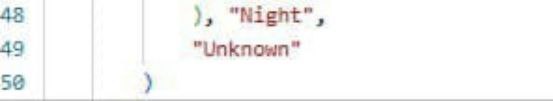
**3.11 Creating Time Period Classification Columns**

To enhance the analysis of passenger behavior based on time-of-day patterns, two custom columns were created using DAX within the **Table View**

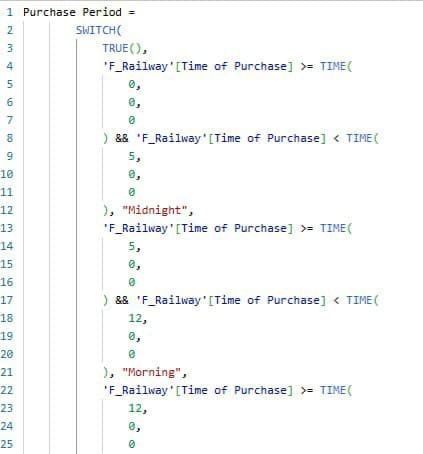
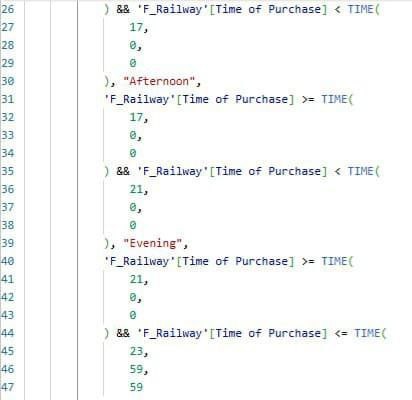
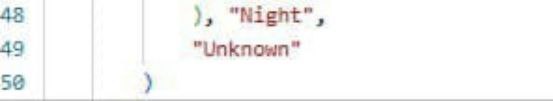
**in Power BI:**

* + Purchase Period
  + Departure Period

**Method:** Both columns were added by selecting “New Column” from the Table View. The logic behind these columns involves categorizing timestamps into descriptive periods of the day such as Morning, Afternoon, Evening, Midnight and Night. This classification was achieved using conditional DAX expressions based on the hour component of the relevant time field (either Date of Purchase or Departure Time).

**Purpose:** The purpose of these columns is to: Support temporal analysis by grouping trips and purchases based on common travel patterns. Enable visual segmentation of customer behavior by time of day. Identify peak booking and departure periods to guide staffing, marketing campaigns, and service planning. Departure period :

Purchase period :



### 3.12 Creating Geographical Region Classifications

### To facilitate regional-level analysis of train traffic and passenger flow, two additional columns were introduced to categorize stations based on their geographical locations: Departure Region Arrival Region

### Method: From the Table View in Power BI:

### 1. The fields Departure Station and Arrival Destination were selected.

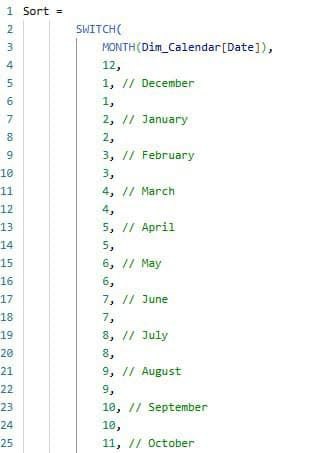
### 2. Using the Data Groups feature, station values were grouped into custom-defined regions based on geography (e.g., North, South, Midlands, etc.).

### 3. As a result, two new columns were generated: Departure Region and Arrival Region.

### Purpose: This classification supports: Regional traffic flow analysis to identify key origin and destination regions. Strategic planning through understanding regional demand variations. Visual storytelling by allowing for aggregation and comparison across geographical zones.

**3.13 Creating a Custom Sort Column in the Date Table:**

**Purpose:** Ensure chronological sorting of dates in visuals for a cleaner and more intuitive dashboard presentation. A new column was created in the Dim\_Calendar table to provide a structured sort order for date fields used in visuals. This ensured that months or dates appeared in the correct chronological order, especially when using fields like "Month Name". The column was hidden from report view since it was only used for sorting and not for direct analysis.Top of Form

Bottom of Form

Top of Form

Bottom of Form

Top of Form

Bottom of Form

Top of Form

Bottom of Form

## 4. Data Normalization

To optimize model performance and adhere to a Star Schema structure, the dataset was normalized into fact and dimension tables.

### 4.1 Creating Dimension Tables

The F\_Railway table was duplicated to extract unique values for the following dimensions:

* **Process**:
  + Removed unnecessary columns.
  + Removed duplicates.
  + Added an Index Column starting from 1, renamed to [Item] Id.
  + Merged the [Item] Id back into F\_Railway and removed the original column.
* **Dimension Tables**:
  + Dim\_Departure: List of departure stations.
  + Dim\_Arrival: List of arrival stations.
  + Dim\_Purchase: Ticket purchase methods (e.g., online, at station).
  + Dim\_Railcard: Discount card types.
  + Dim\_Payment: Payment methods.
  + Dim\_Status: Journey statuses (e.g., On Time, Delayed, Cancelled).
  + Dim\_Delays: Reasons for delays.

### 4.2 Ticket Dimension Design

The Ticket Type was extracted into a dedicated dimension table, Dim\_Tickets:

* **Decision**:
  + Initially considered combining Ticket Class and Ticket Type into one dimension.
  + Final Approach: Retained Ticket Class in the F\_Railway fact table to simplify the model and reduce join operations. Only Ticket Type was moved to Dim\_Tickets.
* **Process**:
  + Removed duplicates from Ticket Type.
  + Added an Index Column as Ticket Id.
  + Merged Ticket Id into F\_Railway and removed the original Ticket Type column.
* **Benefit**: Balances dimensional clarity with model simplicity, improving performance.

### 4.3 Ticket Dimension Note

* **Decision**: Ticket Class was intentionally retained in the F\_Railway fact table.
* **Rationale**: This avoids unnecessary complexity in the data model, reduces join operations, and maintains flexibility for direct analysis of Ticket Class.
* **Benefit**: Enhances model performance while preserving analytical capabilities.

## 5. Data Modeling

The data model was designed to support efficient reporting and analysis using a Star Schema structure.

### 5.1 Setting Up Relationships

* **Configuration**:
  + Linked each [Item] Id in F\_Railway to its corresponding Dim\_ table.
  + All relationships were One-to-Many (One in Dim\_, Many in F\_Railway).
  + Used Single Direction relationships to optimize performance.
* **Outcome**: A robust Star Schema that simplifies queries and enhances analytical capabilities.

### 5.2 Linking the Date Table

The Dim\_Calendar table was integrated as follows:

* **Relationships**:
  + Active: Linked to Date of Purchase.
  + Inactive: Linked to Date of Journey, accessible via the USERELATIONSHIP function in DAX.
* **Purpose**: Enables flexible time-based analysis for both purchase and journey dates.

### 5.3 Organizing the Data Model

To enhance clarity and scalability, tables were grouped in the Model View:

* **Delays**: Dim\_Reason For Delay
* **Stations**: Dim\_Departure, Dim\_Arrival
* **Purchase & Payment**: Dim\_PurchaseType, Dim\_PaymentMethod
* **Tickets**: Dim\_Status, Dim\_Railcard, Dim\_Tickets
* **Date**: Dim\_Calendar
* **Data Dictionary**: T-Dictionary,C-Dictionary,M\_Dectionary And R\_Dictionary

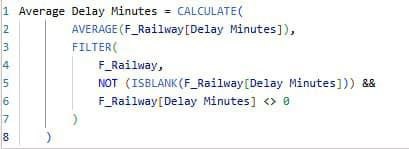
## 6. Creating Measures Using DAX

A set of DAX measures was created to support analytical reporting and visualization in Power BI. These measures cover key performance indicators such as delay durations, cancellation rates, and revenue trends.

### 6.1 List of Measures

* Arrival Route: 

Classifies performance by final destination.

* Average Delay Minute: 

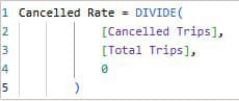
Calculates the average delay time per journey.

* Average Ticket Price:



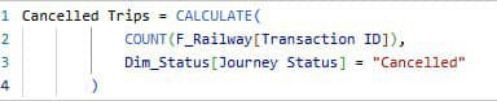
Identifies pricing trends.

* Cancelled %:



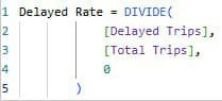
Percentage of cancelled journeys.

* Cancelled Trips:



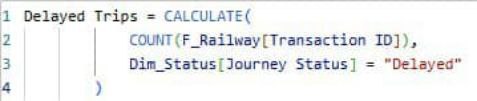
Count of cancelled journeys.

* Delayed %:



Percentage of delayed journeys.

* Delayed Trips:



Count of delayed journeys.

* Max Delay Minutes:



Maximum recorded delay duration.

* Max Ticket Price:



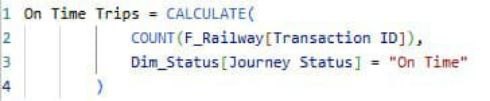
Highest ticket price.

* Min Ticket Price:



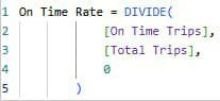
Lowest ticket price.

* On Time Trips:



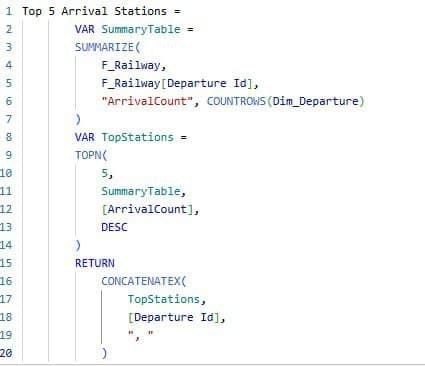
Count of on-time journeys.

* On Time %:



Percentage of on-time journeys.

* Top 5 Arrival:



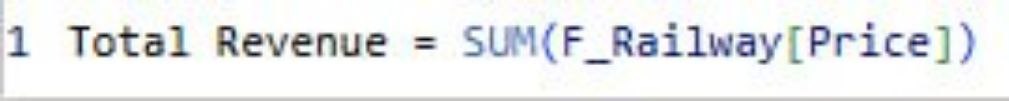
Identifies the five most frequent destinations.

* Total Delay Minutes:



Sum of all delay durations.

* Total Revenue:



Total revenue from ticket sales.

* Total Transactions:

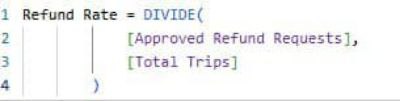


Total number of tickets sold.

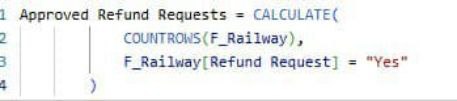
* Routes: 

Total number of unique train routes

* Refund Rate:



Percentage of refunded trips

* Approved Refunds:

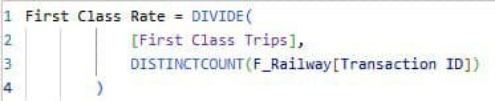
Refund requests that were accepted

* Departure Route:



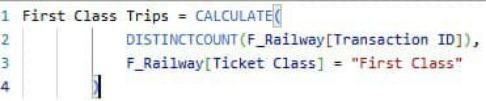
Analysis of train departures categorized by route

* First Class Rate:



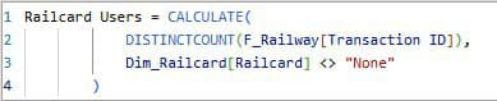
Percentage of tickets booked in first class

* First Class Trips :



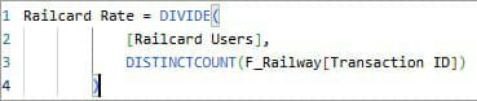
Total number of trips taken in first class

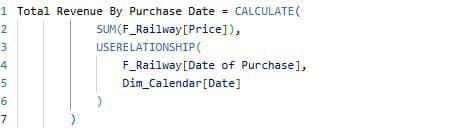
* Railcard Users :



Total passenger who used a railcard

* Railcard Rate :



* ****Total Revenue by Purchase Date :

**Activating Inactive Relationships with USERELATIONSHIP Purpose:** Enable dynamic analysis using the "Date of Purchase" despite the active relationship being with "Date of Journey". In the data model, the Date table was connected to both "Date of Journey" and "Date of Purchase". Since only one active relationship is allowed at a time, the relationship with "Date of Purchase" was inactive by default. To enable accurate time-based analysis by "Date of Purchase", the USERELATIONSHIP function was used within DAX measures to temporarily activate the inactive relationship. This method allowed for flexible analysis without the need to duplicate the Date table**.**

**6.2 Organizing DAX Measures into Folders**

To enhance clarity and navigation within the **Data Model**, DAX measures were systematically grouped into folders in the **Model View**. This organization creates a more structured model, particularly when working with a large number of measures.

**Folder Structure:**

* **Delay Insights**
  + Average Delay Minute
  + Max Delay Minute
  + Total Delay Minutes
* **Fare Insights**
  + Railcard Users
* **Financials**
  + Average Ticket Price
  + Max Ticket Price
  + Min Ticket Price
  + Total Revenue
* **Performance Rate**
  + Cancelled Rate
  + Delayed Rate
  + First Class Rate
  + On Time Rate
  + Railcard Rate
  + Refund Rate
* **Refund insight**
  + Approved Refund Requests
* **Route and Stations Info**
  + Arrival Route
  + Departure Route
  + Top 5 Arrival Stations
* **Trips Summary**
  + Cancelled Trips
  + Delayed Trips
  + First Class Trips
  + On Time Trips
  + Total Trips
  + Unique Routes

**Benefit:**  
This folder structure improves **model readability**, **accelerates report development**, and allows for quicker access to **relevant KPIs** during dashboard creation.

### ****6.3 Creating a Data Dictionary Using DAX Studio****

To document and understand the metadata of the data model, a **Data Dictionary** was generated using **DAX Studio** by executing specialized functions under the INFO.VIEW namespace.

**DAX Functions Used:**

* INFO.VIEW.COLUMNS()
  + Retrieves all columns in the data model, including their names, data types, and associated tables.
* INFO.VIEW.MEASURES()
  + Lists all DAX measures along with their expressions and home tables.
* INFO.VIEW.TABLES()
  + Displays all tables within the model, including their storage mode and row counts.
* INFO.VIEW.RELATIONS()
  + Extracts all relationships between tables, detailing source and target columns, relationship type, and direction.

**Purpose:**  
This provides a **structured, automated** way to generate a Data Dictionary that supports **documentation**, **model validation**, and **handover processes**. It also ensures **transparency and traceability** within the Power BI solution.

### ****6.4 Description Enhancement for Tables, Columns, and Measures****

As part of the Data Dictionary creation process, detailed descriptions were **manually added** to the following components:

* **All Tables:**  
  Each table in the model was annotated with a meaningful description explaining its role and contents.
* **All Columns:**  
  Each column was described with its **purpose**, **data type**, and **relevance to analysis**.
* **All Measures:**  
  Every DAX measure was documented with an explanation of its **calculation logic** and **business interpretation**.

**Purpose:**  
Providing thorough descriptions improves **model transparency**, facilitates **collaboration** with team members, and enhances **report maintainability**, particularly during future updates or handovers.

## 7. Data Visualization

This section presents the key visuals in the dashboard and the insights they provide. The dashboard consists of four main pages, each designed to address specific business questions and provide actionable insights

* **Railway Overview Page**

**Dashboard Summary:**

This dashboard provides a general overview of the train trips during the analysis period. It highlights key metrics such as total trips, delays, cancellations, average delay minutes, and refund approvals. The visualizations help assess the overall performance and identify patterns over time.

**Key Insights from Visuals:**

**KPIs**: (Total Trips, Delayed Trips, Delayed %, Average Delay Minutes, Cancelled Trips, Cancelled %, Approved Refund): Provides an executive summary of train operations with a 7.18% delay rate and 5.94% cancellation rate, indicating room for service efficiency improvements. The average delay duration of 42.55 minutes may affect customer satisfaction.

**Column & Line Chart:** Operational Delay Trends by Month Highlights the peak in delayed trips in March (634 trips), despite a constant average delay time (~41 mins), suggesting operational inefficiencies during this period.

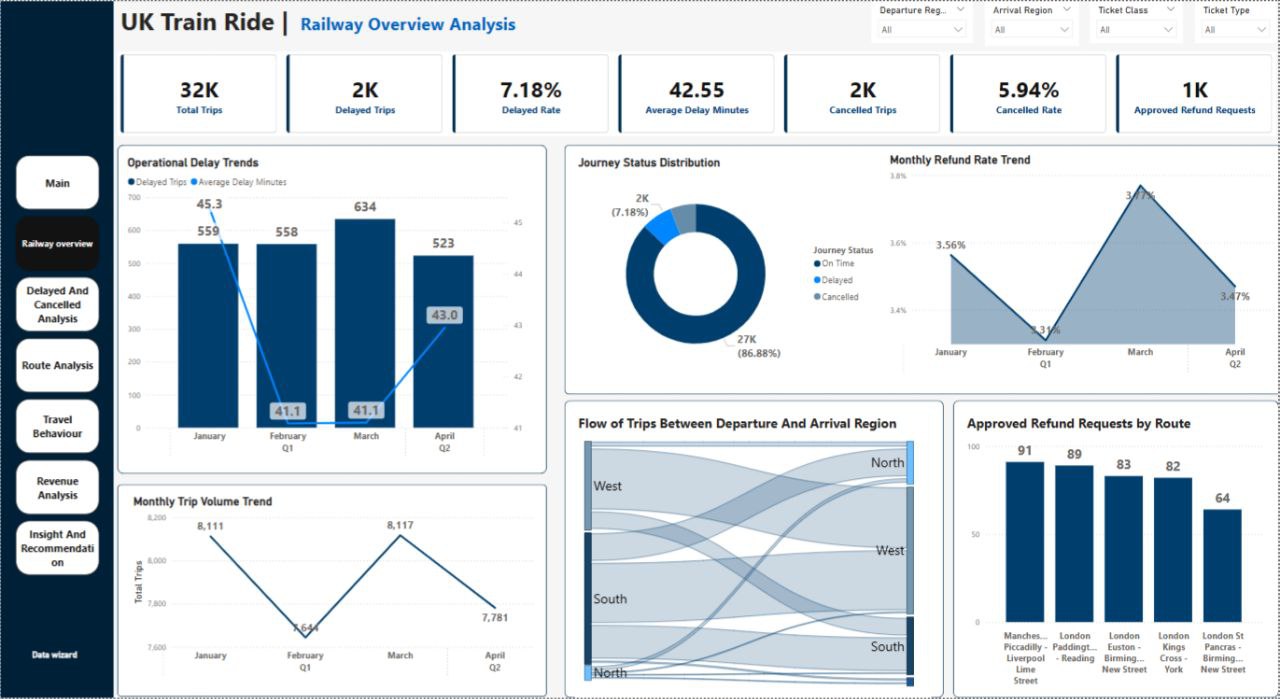
**Donut Chart:** Journey Status Distribution Indicates that the majority of trips (86.88%) are on time, while delays and cancellations collectively make up over 13%, highlighting moderate reliability concerns.

**Area Chart:** Monthly Refund Rate Trend Shows a spike in refund requests during March (3.7%), aligning with the high volume of delayed trips, which implies a direct impact of service reliability on refunds.

**Line Chart:** Monthly Trip Volume Trend Reveals stable demand, peaking in March with 8,117 trips, indicating that despite service issues, overall ridership demand remains consistent.

**Sankey Diagram:** Flow of Trips Between Departure and Arrival Regions Displays traffic concentration towards the North region from all other regions, useful for strategic planning and resource allocation.

**Column Chart**: Approved Refund Requests by Route Identifies routes with the highest refund requests, especially Manchester–Liverpool and London Paddington–Reading, pointing to route-specific reliability issues.

****

***Business Question 1****: How many train trips were completed during the analysis period?*

***Answer****: A total of 32,000 train trips were completed during the analysis period.*

***Visual Used:*** *Card Visual – Total Trips*

***Business Question 2****: What percentage of trips were delayed or cancelled?*

***Answer****: Delayed Trips: 2,000, which is 7.18% of total trips. Cancelled Trips: 2,000, which is 5.94% of total trips.*

***Visual Used****: Card Visuals + Donut Chart (Journey Status)*

***Business Question 3****: Is the overall performance improving or declining over time?*

***Answer:*** *The number of total trips fluctuated slightly, with a peak in March (8,117 trips) and a drop in February (7,642 trips). Delayed trips were highest in March (634 delays), while the average delay minutes remained relatively stable (~41–45 minutes). Refund rates peaked in March, indicating potential customer dissatisfaction during that period. Visuals Used: Line Chart – Total Trips by Month Combo Chart – Delayed Trips & Average Delay Minutes by Month Area Chart – Refund Rate by Month*

***Visuals Used:*** *Line Chart – Total Trips by Month , Combo Chart – Delayed Trips & Average Delay Minutes by Month Area Chart – Refund Rate by MonthTop of Form*

*Bottom of Form*

*Top of Form*

* **Delayed & Cancelled Page***Bottom of FormTop of Form*

**Key Insights from Visuals:**

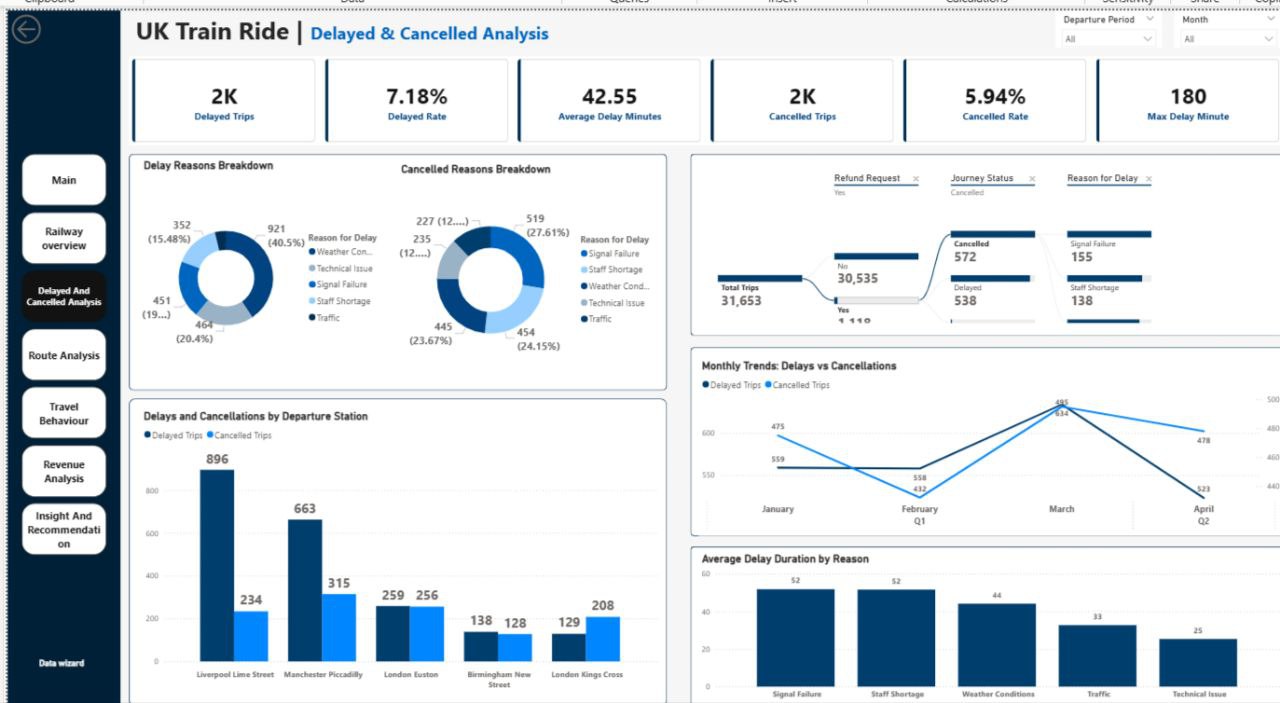
**KPIs**: (Delayed Trips, Delayed %, Average Delay Minutes, Cancelled Trips, Cancelled %, Max Delay Minute): Summarizes the scale of service interruptions, delay frequency, and trip cancellations. Highlights the intensity of delays with maximum delay duration.

**Donut Charts:** Delay Reasons Breakdown & Cancelled Reasons Breakdown Breaks down primary causes for delays and cancellations. Signal failure and weather conditions are the leading issues.

**Sankey Diagram:** Journey Status & Reason for Delay Visualizes the distribution of delay and cancellation reasons, with notable causes like signal failure and staff shortage.

**Column Chart:** Delays and Cancellations by Departure Station Reveals which stations report the highest number of disrupted services — Liverpool Lime Street and Manchester Piccadilly lead.

**Line Chart:** Monthly Trends – Delays vs Cancellations Compares delays and cancellations month-by-month, showing a peak in March and a relative improvement in April.

**Column Chart:** Average Delay Duration by Reason Highlights the average minutes lost by each delay cause, with signal failure and staff shortage causing the longest delays.

**Business Questions & Answers:**

**1. Which months had the highest rates of delayed or cancelled trips?** March recorded the highest number of delayed trips (634) and cancelled trips (495), indicating a peak in operational disruptions during that month**.**

**2. Which routes experienced frequent service disruptions?** Liverpool Lime Street and Manchester Piccadilly had the highest numbers of delayed trips (896 and 663 respectively), along with significant cancellation figures, highlighting recurring issues on these routes.

**3. Are there any seasonal patterns or recurring trends in delays?** Yes, there is a noticeable increase in delays and cancellations during March (end of Q1), followed by a decline in April. This suggests potential seasonal impacts, such as weather conditions or operational pressure early in the year.

* **Route Analysis Page**

**Key Insights from Visuals:** **KPIs** :(Unique Routes, Total Trips, Arrival Route, Departure Route, Delayed Trips, Average Delay Minutes, Cancelled Trips): Provides an overview of network size and operational performance across routes.

**Donut Chart: Top 3 Revenue-Generating Routes** Highlights routes with the highest revenue, guiding strategic investment and prioritization.

**Column Chart:** **Top 7 Routes by Cancelled Rate** Identifies routes with the highest cancellation rates, signaling operational reliability issues.

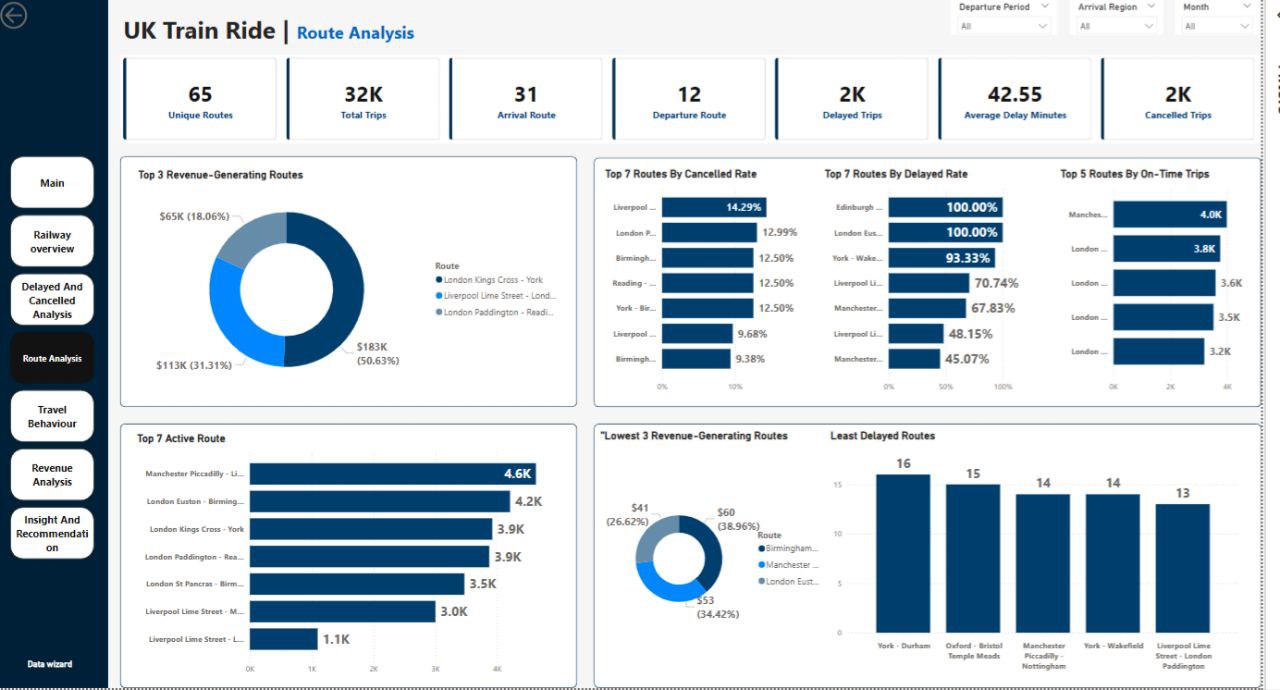
**Column Chart: Top 7 Routes by Delayed** Rate Displays delay frequency by route, crucial for pinpointing underperforming segments.

**Column Chart: Top 5 Routes by On-Time Trips** Showcases most punctual routes, serving as benchmarks for performance standards.

**Column Chart: Top 7 Active Routes** Presents the most frequently usedroutes, essential for capacity planning and resource allocation.

**Donut Chart:** **Lowest 3 Revenue-Generating Routes** Reveals underperforming routes in terms of revenue, indicating areas for review or cost optimization.

**Column Chart:** **Least Delayed Routes** Highlights best-performing routes in terms of delay reduction, useful for identifying best practices.



**Business Questions & Answers:**

**1. Which routes have the highest and lowest number of trips?** The Manchester Piccadilly – Liverpool Lime Street route is the most active, with 4.6K trips. In contrast, the Liverpool Lime Street – London Paddington route appears among the least active with only 1.1K trips**.**

**2. Which routes show the best on-time performance?** Top-performing routes for punctuality include Manchester – London, London – Birmingham, and London – York, with up to 4.0K on-time trips, suggesting strong reliability on these corridors.

**3. Are there routes that require resource reallocation or service improvements?** Routes like Edinburgh – Birmingham and London Euston – Birmingham New Street have 100% delay rates, indicating urgent attention is needed. Additionally, Liverpool Lime Street – London Paddington and London Paddington – Reading show high cancellation and delay rates, which could benefit from resource optimization or operational review.

* **Travel Behaviour Page**

**Key Insights from Visuals:**

**KPIs:** (Total Trips, First Class Trips, First Class Rate, Railcard Users, Railcard Rate): Provides an overview of passenger profile, highlighting premium service uptake and railcard utilization.

**Donut Chart:** Distribution of Trips by Purchase Type Shows purchasing behavior, indicating that the majority of tickets are bought at stations, guiding channel optimization.

**Donut Chart:** Trips Breakdown by Ticket Type Highlights ticket type preferences, with a dominance of “Anytime” tickets, reflecting flexible travel demand.

**Donut Chart:** Distribution of Trips by Payment Method Reveals payment preferences, where credit card is the most used, suggesting transaction convenience priorities.

**Column Chart:** Trips by Railcard Eligibility Shows segmentation of travelers by eligibility, with most trips not using railcards—opportunity for awareness campaigns.

**Table:** Departure Period vs. Class and Railcard Use Breaks down traveler preferences across time periods, showing higher railcard use in afternoons and more first-class trips in the morning and evening.

**Column Chart:** Trips by Purchase Period Identifies peaks in ticket purchase times, with evening being the most active period, aiding marketing and staffing strategies.

**Column Chart:** Trips by Departure Hour Provides insights into travel timing trends, with clear peak hours in the morning and late afternoon, supporting schedule planning.

**Business Questions Answered:**

**1. What are the most common travel times and days?** The most common departure hours are during morning (6–9 AM) and late afternoon (4–6 PM), reflecting peak commuting and return times. The afternoon period sees the highest volume of railcard usage**.**

**2.Which ticket types are preferred by passengers?** “Anytime” tickets are the most commonly used, indicating a preference for flexibility. This is followed by Advance and Off-Peak tickets.

**3. How does customer behavior change in response to frequent delays?** Although the dashboard doesn’t directly tie delays to behavior, railcard and class usage during peak vs. off-peak periods suggest that regular commuters may adjust travel times. Refund rate data also implies a reaction to service disruptions through refund claims.

* **Revenue Analysis Page**

**Key Insights from Visuals:**

**KPIs:** (Total Revenue, Refund Rate, Avg Ticket Price, Max Ticket Price, Min Ticket Price): Provides a summary of overall revenue performance, pricing strategy, and refund activity, reflecting service profitability.

**Column Chart:** Total Revenue by Departure Station Highlights top-performing departure stations in revenue generation, with London Kings Cross as the leading station.

**Donut Chart:** Total Revenue by Payment Method Shows distribution of revenue by payment method, emphasizing dominance of credit card usage.

**Donut Chart:** Total Revenue by Ticket Type Breaks down revenue by ticket type, showing Advance tickets as the largest revenue contributor.

**Donut Chart:** Total Revenue by Ticket Class Displays revenue contribution by class, highlighting Standard class as the major revenue driver.

**Column Chart:** Total Revenue by Arrival Destination Identifies destinations generating the most revenue, supporting route profitability analysis.

**Line Chart:** Monthly Revenue Trend Monitors changes in revenue across months, capturing seasonal effects and performance fluctuations.

**Treemap:** Total Revenue by Railcard Shows revenue distribution based on railcard eligibility, highlighting reliance on non-railcard users.



**Business Questions Answered:**

**1.Which ticket types generate the most revenue?** Advance tickets generate the most revenue, contributing approximately 41.6% ($309K) of the total. This indicates strong adoption of early booking incentives.

**2.How is revenue distributed across months or routes?** Revenue peaks in January ($200K) and March ($195K), while February sees a dip. Route-wise, London Kings Cross and York lead in revenue generation, highlighting high-demand or premium routes.

**3. Do delays and cancellations negatively impact revenue or refund rates?** Yes, there is a 3.53% refund rate, and a notable share of refunds are concentrated on specific routes, suggesting that delays and cancellations do have a financial impact—both via direct refunds and potential customer dissatisfaction.

*Top of Form*

*Bottom of Form*

## 8. Insight & Recommendation

**1. Overall Trip Performance (On-Time, Delayed, Cancelled)**

**Insight 1**: On-time trips account for 86.88% (27,499 out of 31,653), while delays and cancellations stand at 7.18% and 5.94%, respectively. Some routes, such as Edinburgh Waverley - London Kings Cross and London Euston - York, show a 100%delay rate. **Recommendation:** Prioritize performance improvements on high-delay routes by identifying root causes (e.g., signal failures, weather conditions). Implement preventative measures such assignal system upgrades and increasedstaffing.

**Insight 2:** Top causes of delays are adverse weather (40.5%) and cancellations are largely due to signal failures (27.61%) and staff shortages (24.15%). Signal failures and staff shortages also show the highest average delay duration (over 51 minutes). **Recommendation:** Focus investments on updating signal infrastructure and addressing staffing gaps through better shift planning and recruitment.

**2. Monthly Performance Trends**

**Insight 3**: March had the highest trip volume (8,117) and delays (634), while February had the lowest (7,644 trips and 558delays). January had the highest average delay duration (45.3 minutes).

**Recommendation:** Analyze seasonal demand and external factors affecting March. Implement contingency plans such as additional trains or preventive maintenance during high-demand months.

**Insight 4:** Refund rates peaked in March (3.77%) and were lowest in February (3.31%), correlating with increased delays and cancellations.

**Recommendation:** Enhance customer service in months with high refunds through proactivecommunication and automated compensation.

**3. Route-Level Performance**

**Insight 5:** Manchester Piccadilly - Liverpool Lime Street had the highest trip count (4,628) and on-time trips(3,987), but also the highest refunds (91). In contrast, London Euston - York showed poor punctuality despite lower traffic.

**Recommendation:** Maintain strong operational standards on high-traffic routes. Conduct root-cause analysis for underperforming routes to establish corrective actions.

**Insight 6:** Low-revenue routes (e.g., London Euston - Oxford: $41) tend to have high delay or cancellation rates (e.g., London Paddington - Oxford: 12.99% cancellations), potentially reducing demand.

**Recommendation:** Reassess the viability of low-revenue, low-performance routes. Consider resource reallocation to high-yield corridors.

**4. Station-Level Insights**

**Insight 7:** Liverpool Lime Street recorded the highest departure delays (896) and cancellations (234), whereas London Kings Cross led in revenue ($199,650), showing no direct correlation between performanceand revenue.

**Recommendation:** Improve infrastructure and operational efficiency at high-delay stations such as Liverpool Lime Street.

**Insight 8:** York generated the highest arrival revenue ($185,403) but showed significant delays on specific routes. Recommendation: Enhance punctuality to high-revenue destinations like York to boost customer satisfaction and repeatbusiness.

**5. Revenue and Ticket Type Analysis**

**Insight 9:** Advance tickets generated the highest revenue ($309,274; 55.48% of trips), while Anytime tickets were the lowest ($209,309; 16.87%). The average ticket price is $23.44, indicating a preference for budget options.

**Recommendation:** Promote Advance tickets through targeted marketing and dynamic pricing to encourage early bookings and optimize revenue.

**Insight 10:** Credit card payments dominate ($469,511), compared to debit cards ($52,966), likely due to associated benefits like points or installment plans.

**Recommendation:** Offer small incentives for debit card usage to reduce credit card processing fees.

**6. Railcard Usage**

**Insight 11:** 33.91% of passengers used Railcards (10,735 users), but non-Railcard users generated higher revenue ($573,697), indicating substantial discounts for Railcard holders.

**Recommendation:** Reevaluate discount levels for Railcards to balance affordability with profitability. Consider targeted offers for frequent user groups (e.g., Adults: 4,846 users).

**Insight 12:** Railcard usage peaked in the afternoon (49.68%) and was lowest at midnight (24.36%), reflecting usage patterns of students and commuters.

**Recommendation:** Introduce special afternoon incentives to further boost usage during high-demand periods.

### 7. Delays and Cancellations by Station

**Insight 13:** Liverpool Lime Street and Manchester Piccadilly are the most affected stations in terms of delays and cancellations.

**Recommendation:** Deploy dedicated maintenance teams and reinforce staffing at high-impact stations to enhance reliability.

### 8. Refund Trends

**Insight 14:** Of 1,118 approved refunds, 572 were due to cancellations and 538 to delays. Only 8 were for on-time trips. Manchester Piccadilly - Liverpool Lime Street had the highest refund count (91).

**Recommendation:** Improve customer communications and automate refund eligibility checks to reduce complaints and boost trust.

### 9. Revenue by Class and Ticket Type

**Insight 15:** Standard Class accounts for 79.86% of revenue ($592,522), while First Class contributes 20.14% ($149,399), despite a consistent 9.66% occupancy rate.

**Recommendation:** Promote First Class through added-value services like complimentary meals and boarding priority to boost appeal.

### 10. Performance by Departure Time

**Insight 16:** Morning has the highest number of trips (11,709) and Railcard users (3,714). Night periods have the lowest volume (1,999). First Class usage remains stable across time slots (~9%).

**Recommendation:** Increase morning train capacity and offer time-based promotions for Railcard users to distribute demand more evenly.

### Summary of Key Recommendations

1. Enhance performance on high-delay routes (e.g., London Euston - York).
2. Strengthen operations at underperforming stations (e.g., Liverpool Lime Street).
3. Boost revenue through Advance and First Class ticket promotions.
4. Improve customer experience via proactive communication and compensation.
5. Optimize service capacity and distribution based on time-of-day demand patterns.

**1. Overall Trip Performance (On-Time, Delayed, Cancelled)**

**Insight 1**: On-time trips account for 86.88% (27,499 out of 31,653), while delays and cancellations stand at 7.18% and 5.94%, respectively. Some routes, such as Edinburgh Waverley - London Kings Cross and London Euston - York, show a 100%delay rate. **Recommendation:** Prioritize performance improvements on high-delay routes by identifying root causes (e.g., signal failures, weather conditions). Implement preventative measures such assignal system upgrades and increasedstaffing.

**Insight 2:** Top causes of delays are adverse weather (40.5%) and cancellations are largely due to signal failures (27.61%) and staff shortages (24.15%). Signal failures and staff shortages also show the highest average delay duration (over 51 minutes). **Recommendation:** Focus investments on updating signal infrastructure and addressing staffing gaps through better shift planning and recruitment.

**2. Monthly Performance Trends**

**Insight 3**: March had the highest trip volume (8,117) and delays (634), while February had the lowest (7,644 trips and 558delays). January had the highest average delay duration (45.3 minutes).

**Recommendation:** Analyze seasonal demand and external factors affecting March. Implement contingency plans such as additional trains or preventive maintenance during high-demand months.

**Insight 4:** Refund rates peaked in March (3.77%) and were lowest in February (3.31%), correlating with increased delays and cancellations.

**Recommendation:** Enhance customer service in months with high refunds through proactivecommunication and automated compensation.

**3. Route-Level Performance**

**Insight 5:** Manchester Piccadilly - Liverpool Lime Street had the highest trip count (4,628) and on-time trips(3,987), but also the highest refunds (91). In contrast, London Euston - York showed poor punctuality despite lower traffic.

**Recommendation:** Maintain strong operational standards on high-traffic routes. Conduct root-cause analysis for underperforming routes to establish corrective actions.

**Insight 6:** Low-revenue routes (e.g., London Euston - Oxford: $41) tend to have high delay or cancellation rates (e.g., London Paddington - Oxford: 12.99% cancellations), potentially reducing demand.

**Recommendation:** Reassess the viability of low-revenue, low-performance routes. Consider resource reallocation to high-yield corridors.

**4. Station-Level Insights**

**Insight 7:** Liverpool Lime Street recorded the highest departure delays (896) and cancellations (234), whereas London Kings Cross led in revenue ($199,650), showing no direct correlation between performanceand revenue.

**Recommendation:** Improve infrastructure and operational efficiency at high-delay stations such as Liverpool Lime Street.

**Insight 8:** York generated the highest arrival revenue ($185,403) but showed significant delays on specific routes. Recommendation: Enhance punctuality to high-revenue destinations like York to boost customer satisfaction and repeatbusiness.

**5. Revenue and Ticket Type Analysis**

**Insight 9:** Advance tickets generated the highest revenue ($309,274; 55.48% of trips), while Anytime tickets were the lowest ($209,309; 16.87%). The average ticket price is $23.44, indicating a preference for budget options.

**Recommendation:** Promote Advance tickets through targeted marketing and dynamic pricing to encourage early bookings and optimize revenue.

**Insight 10:** Credit card payments dominate ($469,511), compared to debit cards ($52,966), likely due to associated benefits like points or installment plans.

**Recommendation:** Offer small incentives for debit card usage to reduce credit card processing fees.

**6. Railcard Usage**

**Insight 11:** 33.91% of passengers used Railcards (10,735 users), but non-Railcard users generated higher revenue ($573,697), indicating substantial discounts for Railcard holders.

**Recommendation:** Reevaluate discount levels for Railcards to balance affordability with profitability. Consider targeted offers for frequent user groups (e.g., Adults: 4,846 users).

**Insight 12:** Railcard usage peaked in the afternoon (49.68%) and was lowest at midnight (24.36%), reflecting usage patterns of students and commuters.

**Recommendation:** Introduce special afternoon incentives to further boost usage during high-demand periods.

### 7. Delays and Cancellations by Station

**Insight 13:** Liverpool Lime Street and Manchester Piccadilly are the most affected stations in terms of delays and cancellations.

**Recommendation:** Deploy dedicated maintenance teams and reinforce staffing at high-impact stations to enhance reliability.

### 8. Refund Trends

**Insight 14:** Of 1,118 approved refunds, 572 were due to cancellations and 538 to delays. Only 8 were for on-time trips. Manchester Piccadilly - Liverpool Lime Street had the highest refund count (91).

**Recommendation:** Improve customer communications and automate refund eligibility checks to reduce complaints and boost trust.

### 9. Revenue by Class and Ticket Type

**Insight 15:** Standard Class accounts for 79.86% of revenue ($592,522), while First Class contributes 20.14% ($149,399), despite a consistent 9.66% occupancy rate.

**Recommendation:** Promote First Class through added-value services like complimentary meals and boarding priority to boost appeal.

### 10. Performance by Departure Time

**Insight 16:** Morning has the highest number of trips (11,709) and Railcard users (3,714). Night periods have the lowest volume (1,999). First Class usage remains stable across time slots (~9%).

**Recommendation:** Increase morning train capacity and offer time-based promotions for Railcard users to distribute demand more evenly.

### Summary of Key Recommendations

1. Enhance performance on high-delay routes (e.g., London Euston - York).
2. Strengthen operations at underperforming stations (e.g., Liverpool Lime Street).
3. Boost revenue through Advance and First Class ticket promotions.
4. Improve customer experience via proactive communication and compensation.
5. Optimize service capacity and distribution based on time-of-day demand patterns.