🎟 Analyzing UK Train Rides

📌 Overview

This document provides a comprehensive overview of the UK Train Rides from January 2024 to April 2024. This dataset shows valuable insights into overall performance, Ridership Analysis, journey performance until the Revenue, Ticket Class analysis, and finally, the Forecast and Related recommendation.

📌 Objectives

The goal of this project is to perform an in-depth analysis of the UK railway services dataset using Power BI, with the following key objectives:

- Present a comprehensive overview of railway performance through key KPIs (Key Performance Indicators) such as ridership volume, revenue trends, and service reliability.
- Conduct a full ridership analysis, identifying:
  - Peak times of the day and week,
  - Seasonal variations,
  - The total number of riders versus the number of cancelled journeys across the dataset's timeframe.
- Perform detailed station-level analysis, evaluating:
  - Traffic to and from each station,
  - Route popularity,
  - Performance variations across different station pairs.
- Analyse revenue patterns alongside related KPIs, highlighting factors that influence financial performance.
- Examine ridership behaviour based on ticket class categories (e.g., First Class, Standard Class, Discounted Tickets), assessing demand patterns and revenue contributions.
- Develop forecasts and strategic recommendations based on historical trends to support operational planning and revenue optimisation.

Data set Dictionary

Field	Description
Transaction ID	Unique identifier for an individual train ticket purchase
Date of Purchase	Date the ticket was purchased
Time of Purchase	Time the ticket was purchased
Purchase Type	Whether the ticket was purchased online or directly at a train station
Payment Method	Payment method used to purchase the ticket (Contactless, Credit Card, or Debit Card)
Railcard	Whether the passenger is a National Railcard holder (Adult, Senior, or Disabled) or not (None). Railcard holders get 1/3 off their ticket purchases.
Ticket Class	Seat class for the ticket (Standard or First)
Ticket Type	When you bought or can use the ticket. Advance tickets are 1/2 off and must be purchased at least a day prior to departure. Off-Peak tickets are 1/4 off and must be used outside of peak hours (weekdays between 6-8am and 4-6pm). Anytime tickets are full price and can be bought and used at any time during the day.
Price	Final cost of the ticket
Departure Station	Station to board the train
Arrival Destination	Station to exit the train
Date of Journey	Date the train departed
Departure Time	Time the train departed
Arrival Time	Time the train was scheduled to arrive at its destination (can be on the day after departure)
Actual Arrival Time	Time the train arrived at its destination (can be on the day after departure)
Journey Status	Whether the train was on time, delayed, or cancelled
Reason for Delay	Reason for the delay or cancellation
Refund Request	Whether the passenger requested a refund after a delay or cancellation

📚 Analysis Dictionary

Calendar: A table that contains all dates starting from the actual start date till the actual end date and this table also includes the holidays
the table includes the following date breakdown that will ease the analysis (Year, Month Number and month Name, Day of week number and day of week name, day of year, week of year, and finally is a weekend or is holiday)
that breakdown will support the analysis and show how exactly a certain period in each station behaves from the ridership and revenue point of view.
In my opinion, it is the grid of the tableau we are drawing.

Holidays: a table that contains the official occasions that are considered a vacation to exclude them from the analysis to get more accurate insights, and also we are dealing with 4 different areas around the UK that have different official vacations.

Railway: it is the Main table and considered the core of our data and contains all data related to the purchasing time, date, price, payment method, class, from, to and type)

Stations: it is a table that includes the location for the station with the region, latitude and longitude to accurately calculate the distance between the station.
This parameter will be used in the analysis as a factor that may cause the delay if the distance is longer and vice versa.

It is also used to analyse the cost of the ticket to determine which is most profitable, the long routes or the short routes.

Data Analysis:

📊 Dashboard Overview: Introduction

This dashboard provides a high-level summary of the key metrics analyzed in the UK Train Rides dataset between January 2024 and April 2024.
It serves as a quick introduction to the main performance indicators before deeper analysis.
The dashboard summarizes:

- **Ridership Overview**:
  - Total number of rides
  - Active vs. Cancelled rides
  - Cancellation rate percentage
- **Revenue Overview**:
  - Total revenue generated
  - Refund amounts
  - Ticket types revenue breakdown (Advance Tickets)
  - Average ticket cost
- **Journey Performance:**
  - Number of routes
  - On-time vs. delayed journeys
  - Delay rate
  - Peak Hour trips
  - Average journey distance
- **Ticket Demand:**
  - Demand distribution between First Class and Standard Class tickets

    - Purchase type (Advance, Anytime, Off-Peak)
    - Railcard user numbers
- **Prediction:**
    - Predicted total ridership (based on historical trends)

Questions to Answer :

Ridership Analysis :

1- What are the major KPIs for the analysis?
2- What are the peak travel hours during the day at each station, and what is the corresponding ridership volume?
3- How does active and cancelled ridership vary across different weekdays?
4- How is the ridership distributed over the project timeframe (January–April 2024)?
5- To what extent do holidays impact ridership cancellations?
6- Which routes show the highest and lowest active-to-cancelled ridership ratios?
7- Which departure stations have the highest ridership volumes?
8- What are the major causes of journey delays, and how frequently does each cause occur?

Journey Performance Analysis :

1- What are the main KPIs affecting the performance?
2- What is the efficiency of the station/route expressed in how many active/cancelled trips occurred?
3- What is the majour factors of delaying and how offten?
4- What is the most crowded station ?
5- What is the highest reserved route ?
6- Which is lowest station has delays and which one is the highest ?
7- The peak hour at which the highest number of ridership when? And how many ridership?

Revenue Analysis :

1- What are the main KPIs affecting the revenue?
2- How much each payment method share at the revenue?
3- Which purchase type gives higher revenue
4- What is the size of refund against the revenue at each station?
5- What is the effect of the holidays on the revenue?
6- What is the month to month revenue?
7- What is the revenue of each station and which is the best?

Ticket Class Demand Analysis:

1- What are the KPIs that support the ticket class analysis?
2- What is the Average ticket cost by station?
3- What is the best-selling ticket type?
4- What is the distripution of the rail card holders between the ridership?
5- What is the percentage share of the ticket class?
6- Is there a relation between the ticket purchase type and refund request? And which is higher?
7- What is the distripution of the active tickets along the time fram of the analysis?

The work done on the data from the cleaning process and building new meausrers that support our analysis till the forcast is as below:

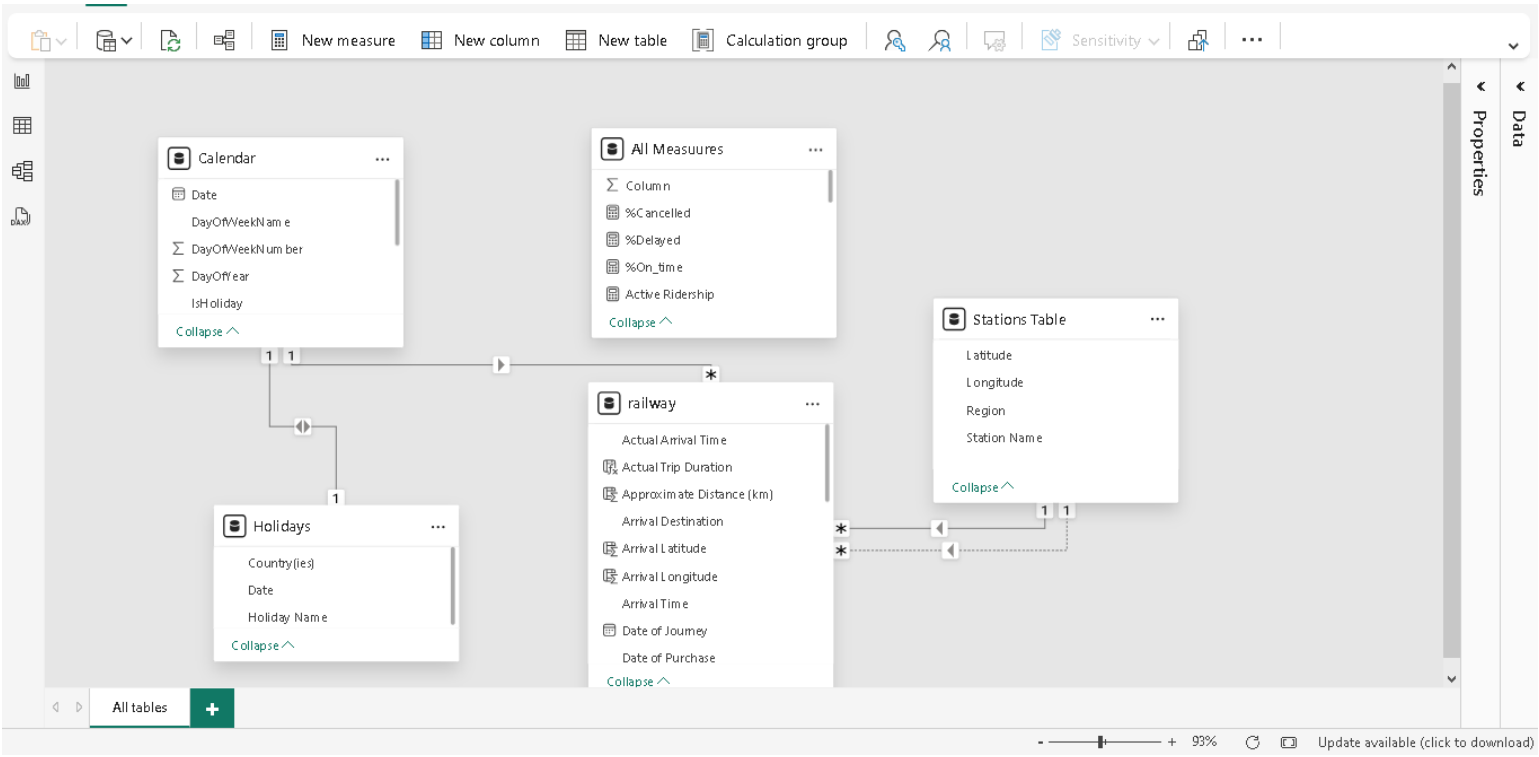
1-Data Cleaning Process

Data cleaning was conducted using **Power Query** to ensure consistency and remove any irrelevant information. Below are the detailed steps performed:

1- Creating the Dates Table to make a continuous timeline from the starting date till the end date using the DAX language functions, starting from naming a variables to resue it without elongate the code and CALENDER function the breakdown the date to year, month, week, and day then use the function FILTER to exclude the holidays from the analysis so as to make the analysis presice and accurate
2- Creating the Holiday Tables to filter all holidays as we are dealing here with different areas in the country that have difrent official occations as a holiday to use it as expressed above.
3- Creating multible measure to calculate the target data that enhance the analysis process and give better insights like the followings:
   - **% cancellation:** is the percentage of cancelled trips of the total trips
   - **% Delayed:** is the percentage of delayed trips to the total trips
   - **% Ontime:** is the percentage of on-time trips to the total trips
   - **Active Ridership:** the calculation of the active ridership by counting the rows of the Journey status, except the cancelled trips
   - **Average Duration Difference:** is the average difference in duration using the AVERAGE function
   - **Average Ticket Cost**: the average of the ticket price from the price column using the function AVERAGE
   - **Cancelled Ridership:** The calculation of the cancelled ridership by counting the rows of the Journey status only includes the cancelled trips
   - **Count_Advance_Ticktype**: is the measure that calculates the ticket of advance type by using the functions CALUCULATE and COUNTROWS
   - **Count_Anytime_Ticktype:** is the measure that calculates the ticket of Anytime type by using the functions CALUCULATE and COUNTROWS
   - **Count_First Class:** is the measure that calculates the ticket of First class by using the functions CALUCULATE and COUNTROWS
   - **Count_off-Peak_Ticktype:** is the measure that calculates the ticket of off-peak type by using the functions CALUCULATE and COUNTROWS
   - **Delay Reason Count: A** Simple measure that counts the delay reasons using the function COUNT
   - **Delayed:** A Simple measure that counts the delayed rides using the function CALCULATE & COUNTROWS
   - **OnTime:** A Simple measure that counts the on-time rides using the function CALCULATE & COUNTROWS
   - **Peak hour:** It is a Measure that calculates the rush hour every day of every station by a functions of FILTER that exclude the cancelled rides, then calculates the rides at every hour, and chooses the time at which the maximum rides occur using the MAX function

- **PeakHourFormatted:** it is a simple function that reformulates the time into HH: MM AM/PM.
- **PeakHourTrips:** calculates the number of trips that occurred during the peak hour, meaning the hour of the day with the highest number of trips, while excluding any trips that were "CANCELLED".
- **Railcard Users:** it is a simple measure that calculates the railcard users using functions CALCULATE and COUNTROWS.
- **RefundRequestCount:** It is a simple measure that calculates the number of refund requests using functions CALCULATE and COUNTROWS.
- **Revenue:** it is a simple measure that calculates the total revenue by using the sum function
- **Revenue_Advance_Ticket_Type:** it is a simple measure that calculates the revenue from the ticket type “Advance” by using the function CALCULATE and COUNTROWS.
- **Route Counts:** it is a simple measure that count the different rout by using function DISTINICTCOUNT.
- **Total Refund Amounts :** a measure that calculates the total refund amounts by using SUM fucntion for the refunded tickets.
- **Total Trips :** it is a simple fuction that calculates the number of total trips excluding the cancelled ones.
- **Total trip % :** it is a fuction that calculates the percentage of the total active trips to the total trips.
- **UniqueJourneyCount:** It calculates the distinct (unique) count of Route values on the current Date of Journey, ignoring any trips that were "Cancelled".

2- 🎯 Data Modeling :



we have five tables:

Table Name	Purpose
Calendar	A calendar table providing dates, days, etc.
Holidays	Special dates and holidays info.
All Measures	A measure table that only holds DAX measures (no raw data).
Railway	The main fact table, containing journey details.
Stations Table	Details about stations (name, location, region).

📈 Relationships Between the Tables

1. **Calendar ➔ Railway**

- Field used: Date from Calendar to Date of Journey in the Railway
- Cardinality: *One-to-Many (1:)*
    - One Date in the Calendar can link to many journeys in Railway.
- Filter Direction: Single Direction (Calendar filters Railway)
- Purpose: This is a very common setup to enable time-based filtering:
  ➔ When you select a date or period, it will filter railway journeys that happened on those dates.

2. **Calendar ➔ Holidays**

- Field used: Date from Calendar to Date in Holidays
- Cardinality: *One-to-Many (1:)*
- Filter Direction: Single Direction (Calendar filters Holidays)
- Purpose: This lets you check if a journey date coincides with a holiday.
  ➔ Important for analysis like holiday ridership spikes, special schedules, etc.

3. Railway ➔ Stations Table (Departure and Arrival)

we have two separate relationships here!

3.1 Railway ➔ Stations Table (Departure)

- Field used: Arrival Destination (or something similar) ➔ Station Name
- Cardinality: Many-to-One (:1)
    - Many railway records can link to one station record.

- Filter Direction: Single Direction (Railway filters Station Table)

Likely Purpose: To connect each journey's arrival station to a station's data (region, location, etc.).

3.2 Railway ➔ Stations Table (Departure)

- Another field (probably Departure Station) ➔ Station Name
- Cardinality: Many-to-One (:1)
- Filter Direction: Single Direction (Railway filters Station Table)

Likely Purpose: To connect each journey's departure station to station attributes.

3-📊 Visulaisation:

- **<u>1- Overall performance dashboard:</u>**

Ridership Card (First Card, Leftmost)

➔ Displays total ridership, active ridership, cancelled trips, and the percentage of cancellations.

Key Influencers (Below Ridership)

➔ Analyzes main drivers affecting revenue, refund requests, and journey status.

Revenue Card (Second Card)

➔ Shows total revenue, refund volumes, advance ticket sales, Railcard users, and average ticket cost.

Journey Performance Card (Third Card)

➔ Summarizes total routes, on-time performance, delayed trips percentage, peak hour trips, and average distance traveled.

Ticket Demand Card (Fourth Card)

➔ Breaks down ticket sales by class (First Class vs Standard) and types (Advance, Anytime, Off-Peak, Railcard users).

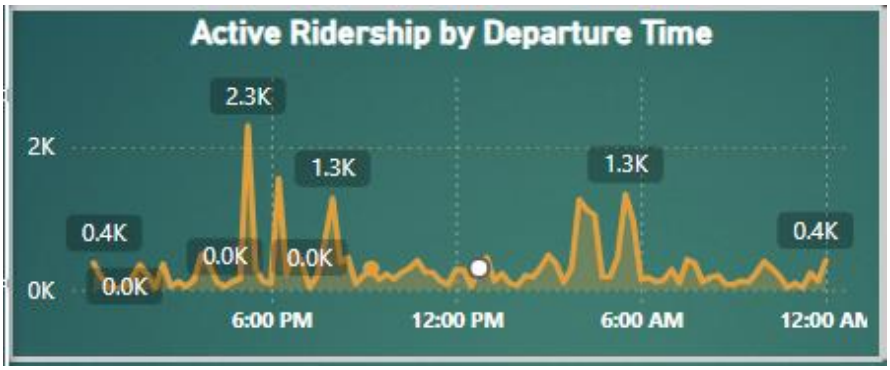
Prediction Section (Fifth Card, Rightmost)

➔ Forecasts active ridership by time and journey date, revenue by purchase date, and advance ticket sales trends.

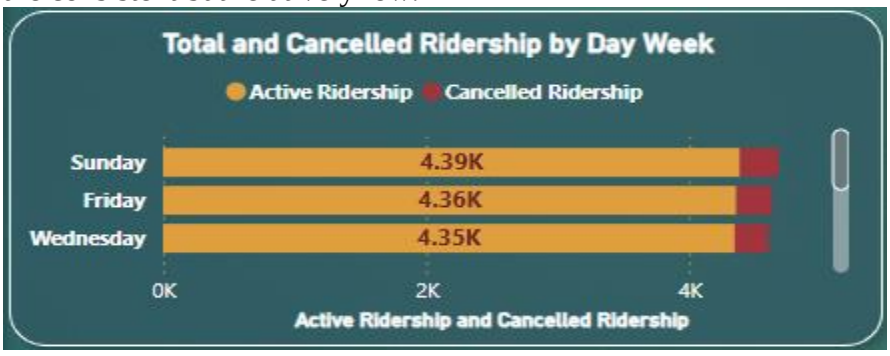
- ## 2- Ridership analysis:

KIPs :

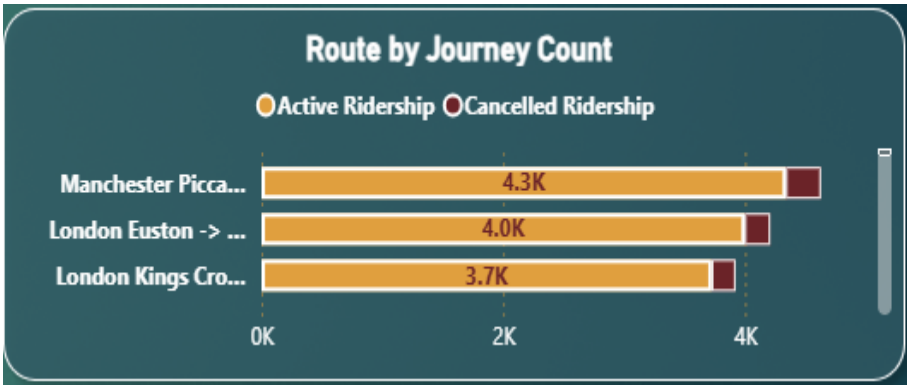
- **Active Ridership by Departure Time**
  - **Summary:** Shows the number of active riders over different times of the day.
  - **Insight:** Peaks around 8 AM, 12 PM, and 6 PM suggest rush-hour travel periods. Early mornings and late evenings have lower ridership.



- **Ridership and Cancelled Ridership by Holiday (Top Right - Bar Chart)**
  - **Summary:** Compares normal ridership against cancellations during major holidays.
  - **Insight:** Minor cancellations occurred during holidays like St. Patrick's Day, Good Friday, and New Year's. Ridership was steady with a small percentage of cancellations.

- **Total and Cancelled Ridership by Day of Week (Middle Left - Horizontal Bar Chart)**
  - **Summary:** Highlights how many trips were active or canceled across different days.
  - **Insight:** Sunday, Friday, and Wednesday show almost equal ridership (around 4.3K each). Cancellations are consistent but relatively low.
  - 

- **Route by Journey Count (Middle Right - Horizontal Bar Chart)**
  - **Summary:** Displays active versus canceled ridership by specific stations.
  - **Insight:**
    - Manchester Piccadilly, London Euston, and London Kings Cross have the highest journey counts.
    - Cancellations are present but smaller compared to active rides.



- **Ridership by Time (Bottom Left - Line Chart)**
  - **Summary:** Trends in daily ridership over time (January to April 2024).
  - **Insight:**
    - Consistent ridership with visible periodic dips (likely weekends or specific disruptions).
    - Drops might correspond with maintenance, holidays, or failures (like the mentioned signal failure).

- **Active Ridership by Departure Station (Bottom Right - Map Visualization)**
  - **Summary:** Geographic representation of ridership volume from various departure stations.
  - **Insight:**
    - Major activity centers around London and the surrounding cities.
    - Larger circles indicate more significant ridership counts.

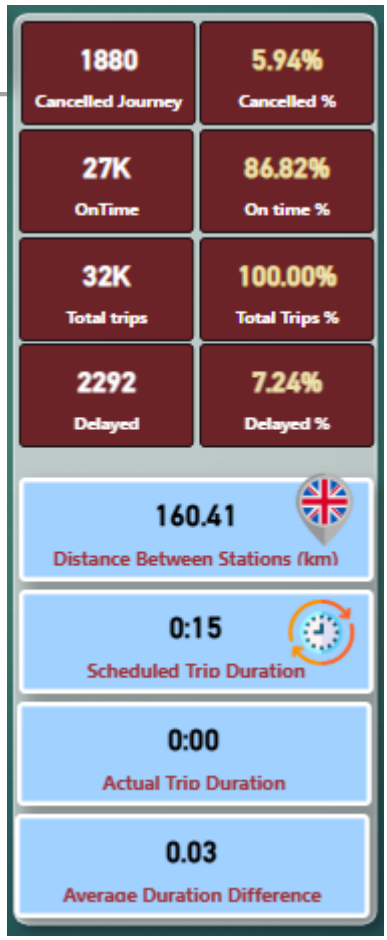
- **3-Journey performance analysis:**

KPIs :

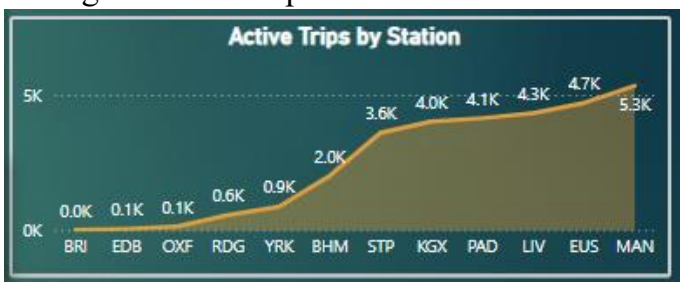
- **Ridership**: Total passengers recorded during the period.
- **Active Ridership**: Passengers who completed their trips successfully.
- **Peak Hour Trips**: Number of trips during peak traffic hours.
- **Peak Hour**: The highest ridership time of day.
- **Cancelled Ridership**: Total passengers affected by journey cancellations.

Visuals and Metrics

- **Cancelled Journey**: Outlines the percentage and total number of cancelled journeys.
- **On-Time**: Displays percentage of trips completed without delay.
- **Total Trips**: Indicates the complete number of trips included in the analysis.
- **Delayed**: Shows the percentage of journeys that experienced delays.
- **Distance Between Stations**: Average distance covered between stations.
- **Scheduled Trip Duration**: Average planned travel time between stations.
- **Actual Trip Duration**: Shows there's no major deviation from the schedule (potential placeholder or data issue).
- **Average Duration Difference**: Minimal time difference between scheduled and actual trips.
- **Reasons for Delay**: Weather, technical issues, and signal failures are the top causes of delays.



- **Active Trips by Station**: Manchester Piccadilly (MAN) and Liverpool Lime Street (LIV) stations have the highest active trips.



-

- **Total Trips by Route** (Bottom Right - Horizontal Bar Chart): The Manchester to Liverpool route has the most trips among the top four listed routes.

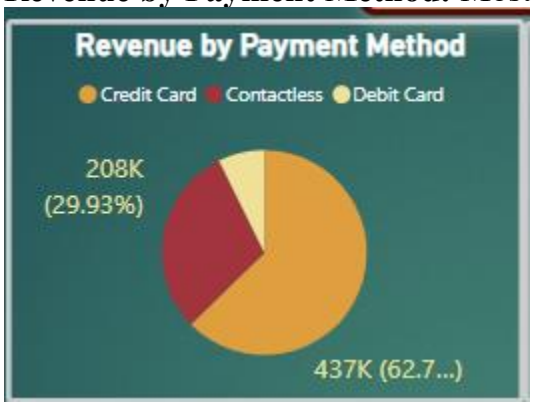
- ## 4-Revenue Overview:

🔵 KPIs

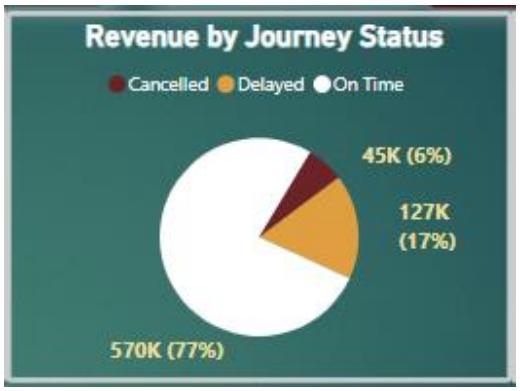
- **Average Ticket Cost (£23):** The mean price per ticket sold.
- **Net Revenue (£703.2K):** Total revenue after deducting refunds.
- **Revenue (£741.9K):** Gross revenue collected.
- **Refund Amount (£38.7K):** Total money refunded to passengers.
- **Refund Requests (1,118):** Total number of refund requests made.

📊 **Visuals and Metrics**

- **Revenue by Payment Method:** Most revenue comes from debit card payments .

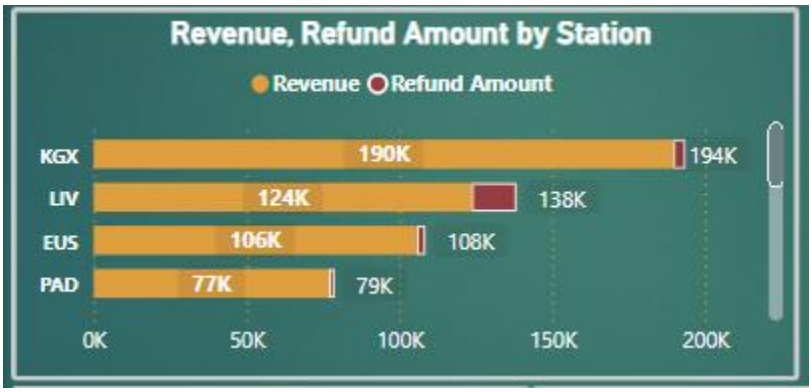


- **Revenue by Journey Status:.** Most revenue is generated from on-time journeys (570K), with smaller portions from delayed and cancelled rides.

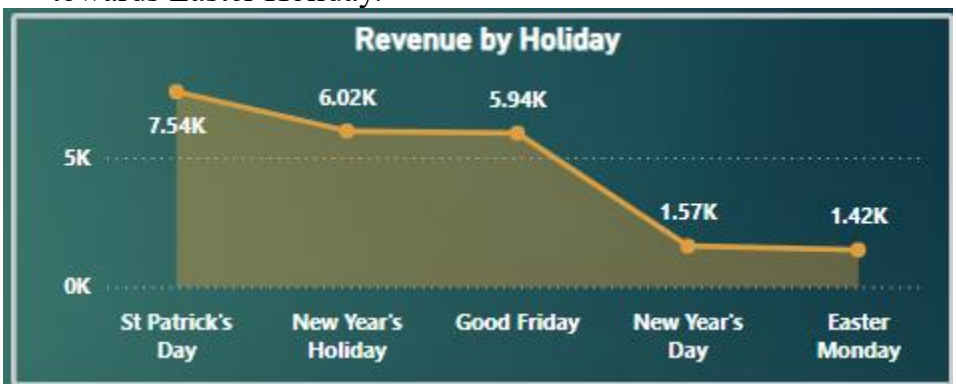


- **Purchase Type:** Revenue split shows slightly more online purchases (£360,234) compared to station purchases (£336,231).

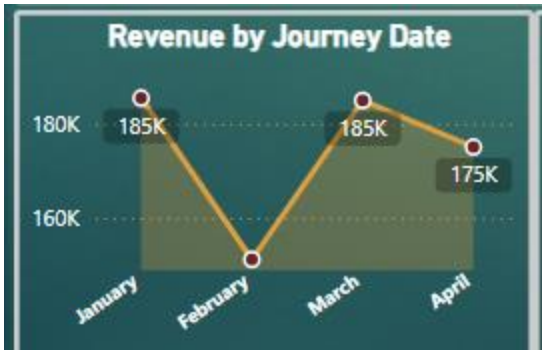
- **Revenue, Refund Amount by Station:** King's Cross (KGX) generates the highest revenue and refund amounts compared to other stations.



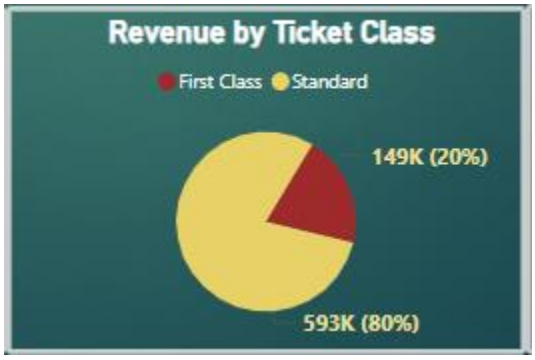
- **Revenue by Holiday:** Highest revenue spikes around St. Patrick’s Day (£7.54K) and gradually drops towards Easter Holiday.



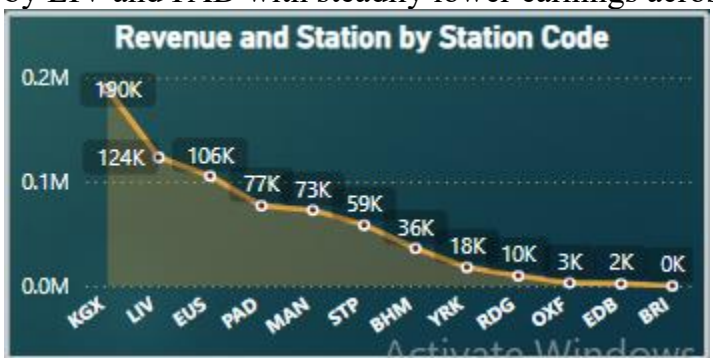
- **Revenue by Journey Date:** Revenue fluctuates monthly, peaking in February and falling sharply in April.

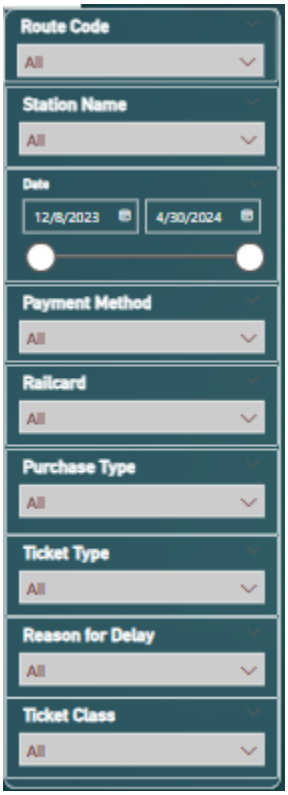


Revenue by Ticket Class: Standard class tickets dominate the revenue (80%), while first-class tickets account for 20%.



- **Revenue and Station by Station Code:** KGX leads in station revenue, followed by LIV and PAD with steadily lower earnings across others.





Filters Panel

- **Route Code, Station Name, Date Range, Payment Method, Railcard, Purchase Type, Ticket Type, Reason for Delay, and Ticket Class**: Dynamic filters allow users to focus the revenue analysis across various factors and travel conditions.

- ## 5- Ticket Demand:

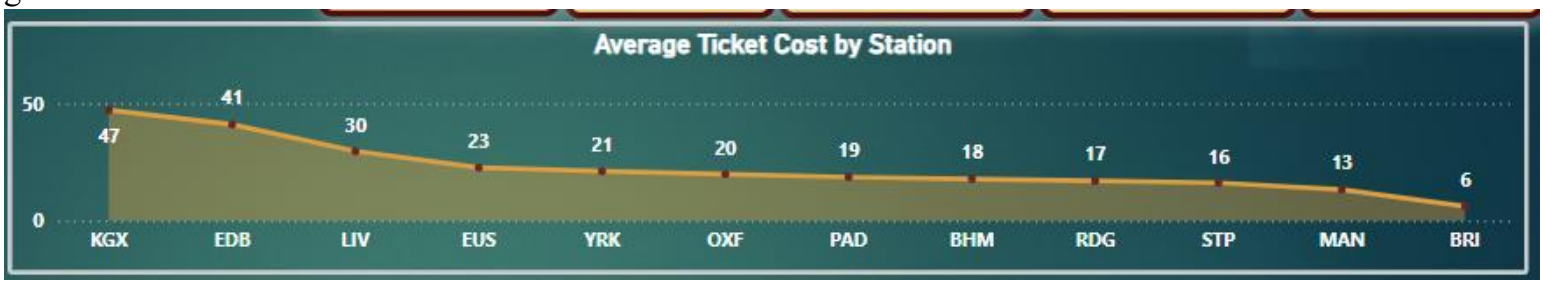
KPIs :

- **Average Ticket Cost (£23):** The mean price paid per ticket across all sales.
- **Railcard Users (10.7K):** Total number of travelers who used a railcard.
- **Advance Ticket Type (17.6K):** Number of advance tickets purchased.

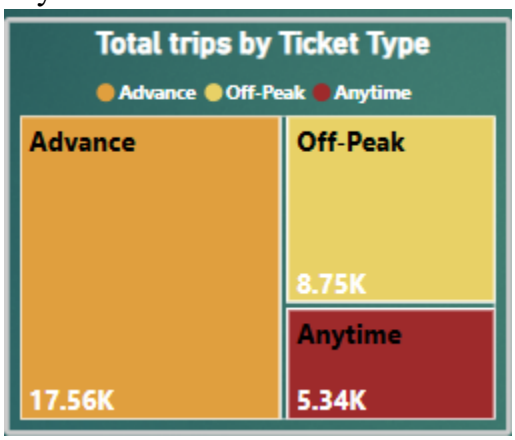
- **Anytime Ticket Type (5.3K):** Number of anytime tickets sold.
- **Off-Peak Ticket Type (8.8K):** Number of off-peak tickets purchased.

📊 Visuals and Metrics

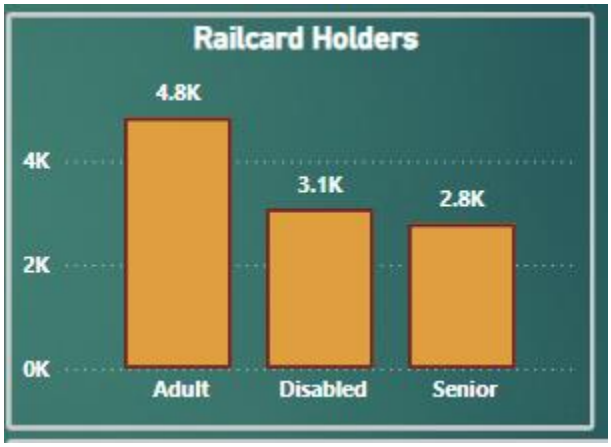
- **Average Ticket Cost by Station:** King's Cross (KGX) shows the highest average ticket cost, with a general decline across other stations.



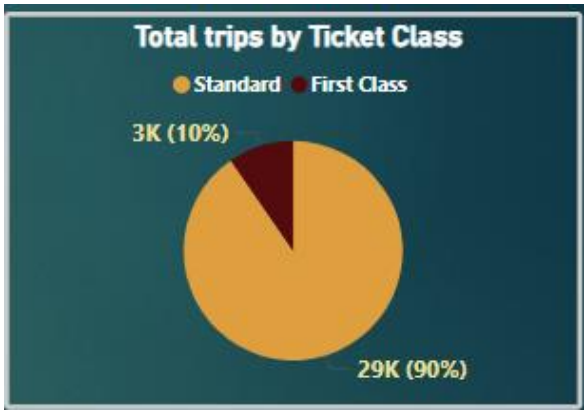
- **Total Trips by Ticket Type:** Advance tickets dominate sales (17.56K), followed by off-peak and anytime tickets.



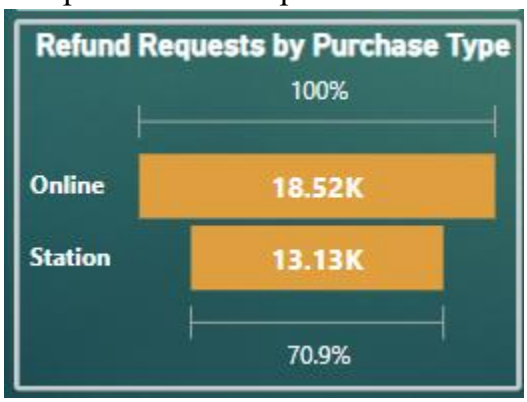
- **Railcard Holders:** Adults represent the largest group of railcard users, followed by disabled and senior travellers.



- **Total Trips by Ticket Class:** Standard class accounts for the vast majority of trips (90%), with only 10% in first class.



- **Refund Requests by Purchase Type:** Most refund requests came from station purchases (70.9%) compared to online purchases.



- **Active Tickets by Date of Purchase:** Peak ticket purchase activity occurred in early February, followed by smaller peaks in March and April.

- **<u>6-Prediction (Forecasting Analysis)</u>**

Introduction:

The forecasts shown are generated based on historical ridership, revenue, and ticket sales data between **December 2023 and April 2024**.
The models use **time-series forecasting techniques** likely based on approaches such mathematical models embedded within Power BI's forecasting engine.

The forecast calculation considers:

- **Historical trends** (repeating patterns in past data)
- **Seasonality** (daily, weekly, and monthly cyclical effects)
- **Growth rates** (positive or negative)
- **Variance and noise** (to compute prediction intervals with upper and lower bounds)

Key parameters influencing the forecast:

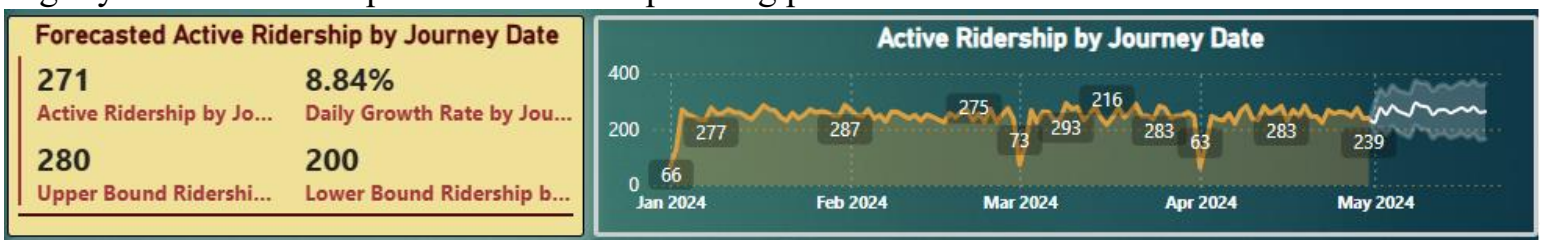
- **Time unit granularity** (hourly, daily, or transaction-based aggregation)
- **Observed seasonality** (e.g., weekday vs weekend travel, holiday impacts)
- **Outliers and anomalies** (signal failures, unusual surges)
- **Confidence intervals** (a range showing uncertainty in predictions)

Forecasted Active Ridership by Departure Time:

The model predicts that 395 passengers will depart per hour on average in future periods, with a strong hourly growth rate of 117.7%.
The forecast exhibits a wide prediction interval (ranging from 0 to 895), indicating high uncertainty, likely due to the high fluctuation in hourly data.

Active Ridership by Departure Time

Historical hourly ridership shows distinct, sharp peaks around typical commuting hours (6 AM–9 AM and 4 PM–7 PM), with the forecast projecting a flattening of peaks, implying more consistent but slightly reduced ridership across hours in upcoming periods.



Forecasted Revenue by Date of Purchase

Historical revenue highlights clear revenue spikes around key travel seasons (February and April), and the forecast shows a flattening trend, implying that unless special events reoccur, revenue may fall back to baseline levels.

Future revenue is forecasted at £3,729 per day, but with a negative growth rate of -38.37%, suggesting declining revenue potential if current patterns persist.
The forecast range between £1,000 and £7,000 reflects significant uncertainty, possibly linked to variable purchasing behaviors around holidays or promotions.

Forecasted Active Tickets by Date of Purchase

The historical ticketing data shows major purchasing peaks tied to seasonality, particularly early in February and around Easter.
The forecast trend is downward-sloping, reinforcing the need for demand stimulation strategies if business objectives are to be met.

Ticket purchases are forecasted to average 126 per day, with a strong decline rate of -53.85%.
This substantial drop suggests reduced demand or ticketing activity, potentially due to seasonal

factors, lower marketing efforts, or passenger behavior changes post-peak seasons.
The forecast band (76 to 176) indicates moderate uncertainty.

- **<u>7-Key Influencer Introduction</u>**

1. What Influences Revenue to Increase

- **Arrival Destination is London Euston (78.23):**
  - When the arrival destination is London Euston, the average revenue increases by 78.23 units (likely British pounds, based on the report context).
- **Departure Time is 7:45:00 AM - 8:00:00 AM (63.47):**
  - When the departure time is between 7:45 AM and 8:00 AM, the average revenue increases by 63.47 units.

- **Arrival Destination is London Paddington (38.96):**
    - When the arrival destination is London Paddington, the average revenue increases by 38.96 units.
- **Journey Status is On Time (Decrease by 36.95):**
    - When the journey status is "On Time," the average revenue decreases by 36.95 units.

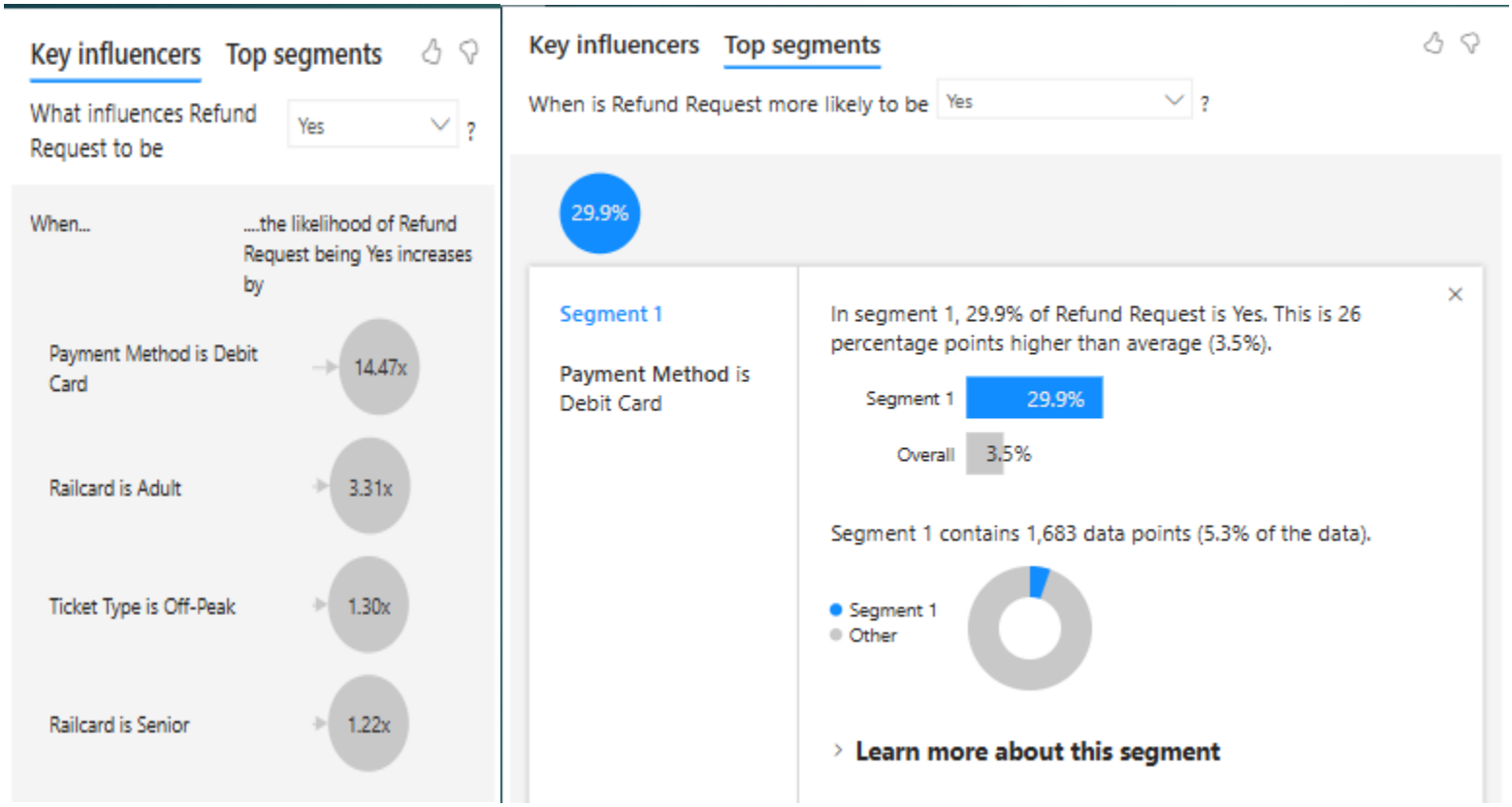
Interpretation:

- **Arrival Destination (London Euston and London Paddington):**
    - Journeys terminating at London Euston and London Paddington generate significantly higher revenues. This is logical since these are two of London's largest and most critical stations, attracting both business and leisure travelers. Higher revenue can be attributed to:
        - Higher ticket prices for these destinations.
        - Greater passenger volumes headed to these key stations.
    - **Connection with other visuals:** In the "Revenue by Date of Purchase" chart, we observe revenue fluctuations (e.g., £11K in February vs. £4K in May). The revenue drop could correlate with fewer journeys to London Euston or Paddington during those periods.
- **Departure Time (7:45 AM - 8:00 AM):**
    - This time window falls within peak commuting hours, explaining why revenue increases:
        - Higher ticket prices during peak hours.
        - Increased number of passengers traveling at this time.
    - **Connection with other visuals:** In the "Active Ridership by Departure Time" chart, there’s a significant rise in ridership during early morning hours (e.g., 1,337 passengers at 6:00 AM), supporting the observation that peak times drive higher revenues.

Journey Status (On Time - Revenue Decrease):

- Journeys classified as "On Time" show a revenue decline. Though counterintuitive, this may be due to:
    - Delayed journeys possibly causing indirect revenue increases (e.g., alternative travel purchases, fees).
    - "On Time" services might offer more competitive (lower) fares to attract passengers, whereas delayed services could correspond to peak routes like London Euston.
- **Connection with other visuals:** This requires further analysis, but may relate to lower passenger counts on punctual journeys, as seen in "Active Ridership by Journey Date."

2. **What Influences Refund Requests to be Approved**



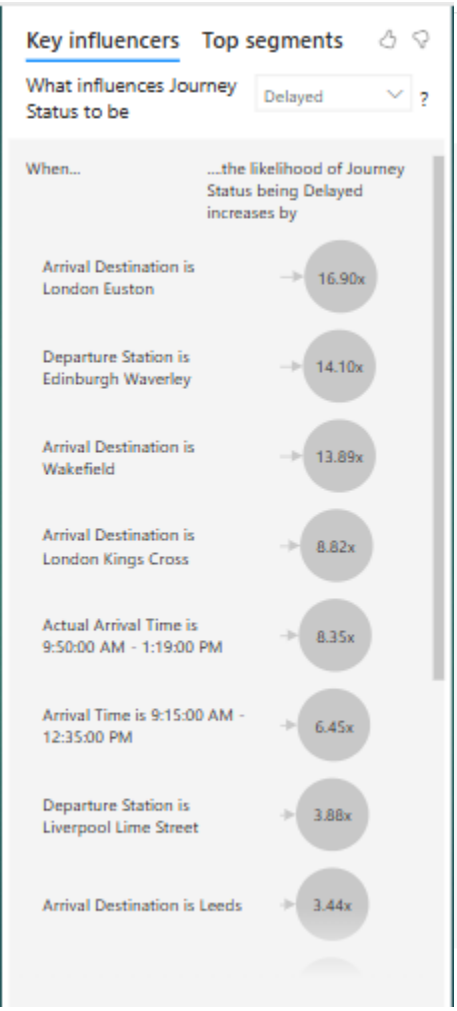
- **Payment Method is Debit Card (14.479x):**
  - Refund requests are 14.479 times more likely when the payment method is Debit Card.
- **Railcard is Adult (3.31x):**
  - Refund requests are 3.31 times more likely when the Railcard type is Adult.
- **Ticket Type is Off-Peak (1.30x):**
  - Refund requests are 1.30 times more likely when the ticket type is Off-Peak.
- **Payment Method is Credit Card (Decrease by 1.04x):**
  - Refund requests are 1.04 times less likely when payment was made by Credit Card.

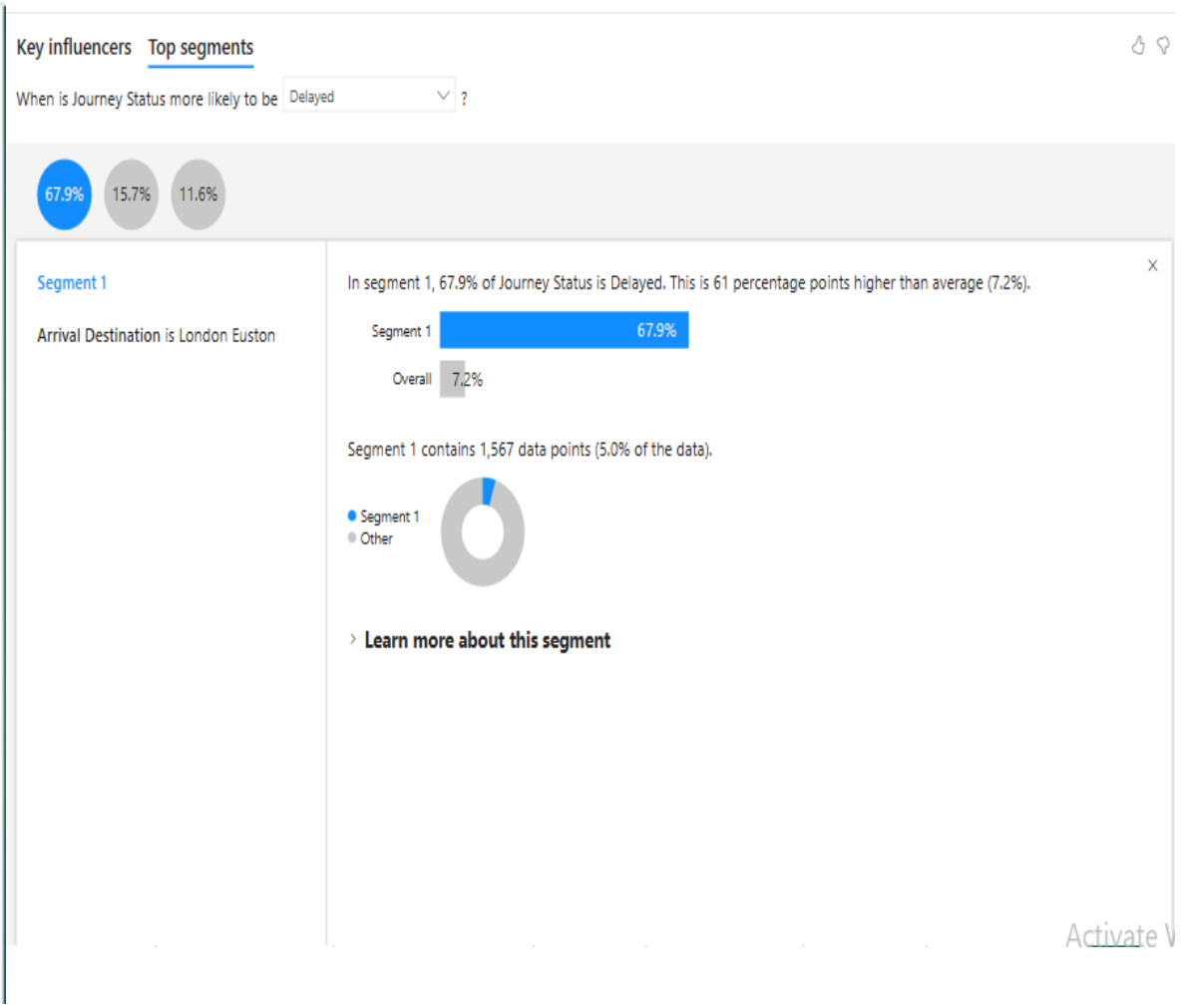
Interpretation:

- **Payment Method - Debit Card:**
  - Passengers using Debit Cards are significantly more prone to request refunds. Potential reasons include:
    - Greater cost sensitivity among debit card users, prompting faster refund claims when expectations (e.g., journey quality) are unmet.
    - Debit card refunds may be processed more simply than credit card transactions.

- **Railcard Type - Adult:**
  - Adult Railcard holders are more likely to seek refunds, possibly because:
    - Adults are more aware of refund rights compared to students or seniors.
    - Business travelers may demand refunds when services are disrupted.
- **Ticket Type - Off-Peak:**
  - Off-Peak tickets slightly increase refund probability. This may be due to:
    - Lower-cost Off-Peak fares leading to less hesitation in refund requests.
    - Lower service reliability during off-peak hours.
- **Payment Method - Credit Card:**
  - Paying by Credit Card slightly reduces refund likelihood, perhaps because:
    - Credit card users are less cost-sensitive.
    - Refund processes for credit cards may be more complex, discouraging claims.

3. **What Influences Journey Status to Be Delayed**





- **Arrival Destination is London Euston (16.90x):**
  - Journeys arriving at London Euston are 16.90 times more likely to be delayed.
- **Departure Station is Edinburgh Waverley (14.10x):**
  - Journeys departing from Edinburgh Waverley are 14.10 times more likely to be delayed.
- **Arrival Destination is Wakefield (13.89x):**
  - Journeys arriving at Wakefield are 13.89 times more likely to be delayed.
- **Arrival Time (Decrease by 0.24, On-Time Status Increases 1.116x):**
  - For each 0.24 decrease in actual arrival time, the likelihood of the journey being "On Time" increases by 1.116 times.

Interpretation:

- **Arrival Destination - London Euston:**
  - Journeys terminating at London Euston are highly prone to delays. Reasons include:
    - High congestion levels at this major hub.
    - Longer-distance journeys (e.g., from Edinburgh) being more exposed to delays.
- **Departure Station - Edinburgh Waverley:**
  - Departures from Edinburgh Waverley also have a high delay probability, possibly due to:
    - Operational congestion or maintenance issues at a major Scottish hub.
    - Longer trip lengths increasing delay risks.
- **Arrival Destination - Wakefield:**
  - Journeys ending at Wakefield face notable delays, potentially due to:
    - Limited infrastructure (e.g., fewer platforms, fewer available trains).
    - Scheduling or operational constraints.
- **Arrival Time Decrease:**
  - A decrease in actual arrival time (earlier or exactly on time) positively influences punctuality.

Connection with other visuals:

- In "Active Ridership by Journey Date," passenger counts decline in April, possibly linked to dissatisfaction stemming from delays (especially to London Euston).

4. Linking Insights with Other Report Visuals

- **Revenue by Date of Purchase:**

- Revenue fluctuations, particularly the March decline (£1K), might result from fewer trips to high-revenue destinations (London Euston, Paddington) or increased delays/cancellations.

- **Active Tickets by Date of Purchase:**
  - The ticket count drop to 112 in April may be associated with a rise in cancellations and refund requests, likely driven by service disruptions.
- **Active Ridership by Journey Date:**
  - The dip in passenger numbers (to 63) in April coincides with increased delays, reducing customer satisfaction and overall demand.

Results and Recommendations

The document and image focus on the train route from **London Kings Cross to York**, analyzing factors influencing **revenue**, **refund requests**, and **journey delays**, along with related KPIs and visuals such as revenue trends, active tickets, and ridership patterns. I'll summarize the results and provide actionable recommendations based on the full scope of the data.

1. Results Summary

What Influences Revenue to Increase

- **Arrival Destination is London Euston (78.23)**: Journeys arriving at London Euston increase average revenue by 78.23 units (likely GBP). This is likely due to higher ticket prices and a larger passenger volume, as London Euston is a major hub for commuters and tourists.
- **Departure Time is 7:45 AM - 8:00 AM (63.47):** Peak departure times between 7:45 AM and 8:00 AM increase revenue by 63.47 units, driven by higher fares during peak hours and increased passenger numbers.
- **Arrival Destination is London Paddington (38.96):** Journeys to London Paddington increase revenue by 38.96 units, reflecting its status as another key London station with high passenger demand.
- **Journey Status is On Time (Decrease by 36.95):** On-time journeys decrease revenue by 36.95 units, possibly because delayed journeys lead to higher revenue through additional fees or demand for alternative journeys, while on-time journeys may have lower, competitive pricing.

- **Insights from Related Visuals:**
  - The "**Revenue by Date of Purchase**" chart shows fluctuations (e.g., 11K in February, 4K in May, and 1K in March). Revenue drops may be linked to fewer journeys to high-revenue destinations like London Euston or Paddington, or increased delays/cancellations.
  - The "**Active Ridership by Departure Time**" chart indicates higher passenger numbers in the morning (e.g., 1,337 at 6:00 AM), supporting the revenue increase during peak times.
  - The "**Active Ridership by Journey Date**" chart shows a drop in passengers in April (63), potentially due to delays or cancellations affecting on-time journeys, which in turn reduces revenue.

What Influences Refund Requests to be "Yes"

- **Payment Method is Debit Card (14.479x)**: Debit card payments increase the likelihood of a refund request by 14.479 times, likely due to cost-sensitivity among debit card users and the ease of requesting refunds with direct payment methods.
- **Railcard is Adult (3.31x)**: Adult railcard holders are 3.31 times more likely to request refunds, possibly because adults are more aware of their rights or travel for work, necessitating refunds for delayed/canceled journeys.
- **Ticket Type is Off-Peak (1.30x)**: Off-Peak tickets increase refund likelihood by 1.30 times, as they are cheaper and may be linked to less reliable journeys (more prone to delays/cancellations).
- **Payment Method is Credit Card (Decrease by 1.04x)**: Credit card payments decrease refund likelihood by 1.04 times, as credit card users may be less cost-sensitive or face more complex refund processes.

- **Insights from Related Visuals:**
  - The "**Active Tickets by Date of Purchase**" chart shows a significant drop in ticket numbers in April (from 294 to 112), likely due to cancellations (e.g., Actual Arrival Time = 0 in prior visuals), which correlates with increased refund requests, especially among debit card users.
  - The drop in ridership in April (63 passengers, as seen in "**Active Ridership by Journey Date**") may also reflect dissatisfaction from delays/cancellations, prompting more refund requests.

What Influences Journey Status to be "Delayed"

- **Arrival Destination is London Euston (16.90x)**: Journeys to London Euston are 16.90 times more likely to be delayed, likely due to the station's high congestion and operational challenges.
- **Departure Station is Edinburgh Waverley (14.10x)**: Journeys departing from Edinburgh Waverley are 14.10 times more likely to be delayed, possibly due to operational issues at this major station or the long distances involved (e.g., to London Euston).
- **Arrival Destination is Wakefield (13.89x)**:Journeys to Wakefield are 13.89 times more likely to be delayed, potentially due to limited resources (e.g., fewer platforms) or scheduling issues at this secondary station.
- **Arrival Time (Decrease by 0.24, Journey Status On Time 1.116x)**: A decrease in Actual Arrival Time by 0.24 increases the likelihood of a journey being on time by 1.116 times, indicating that better adherence to schedules reduces delays.

- **Insights from Related Visuals:**
  - The "**Active Ridership by Journey Date**" chart shows a drop in passengers in April (63), likely due to delays in journeys to London Euston or from Edinburgh Waverley, which negatively impacts passenger satisfaction.
  - The "**Revenue and Total Refund Amount by Route**" chart (top right in the image) shows £17 in total refunds for the London Kings Cross to York route, suggesting that delays (e.g., to London Euston) may contribute to refund requests, aligning with the refund insights above.

Additional Insights from Visuals and KPIs

- **Revenue Trends**:Revenue fluctuates significantly by date of purchase (11K in February, 1K in March, 4K in May). This volatility suggests seasonal or operational factors (e.g., delays, cancellations, or fewer journeys to high-revenue destinations like London Euston).

- **Active Tickets**:The drop in active tickets in April (112) aligns with increased cancellations and delays, particularly for journeys involving London Euston and Edinburgh Waverley.
- **Ridership Patterns**: Ridership drops in April (63 passengers) correlate with delays and cancellations, indicating a direct impact on passenger satisfaction and demand.
- **Refund Amount**:The £17 refund amount for the London Kings Cross to York route suggests that delays and cancellations (e.g., to London Euston) are driving refund requests, particularly among debit card users.

2. Recommendations

Based on the full analysis of the document, visuals, and KPIs, here are actionable recommendations to improve revenue, reduce refund requests, and minimize journey delays for the London Kings Cross to York route.

To Increase Revenue

1. Increase Journeys to High-Revenue Destinations:

- Prioritize scheduling more journeys to London Euston and London Paddington, as these destinations generate the highest revenue (78.23 and 38.96 units, respectively).
- Ensure these journeys are reliable by addressing operational challenges (e.g., congestion at London Euston) to maintain passenger trust and demand.
- Example Action: Add more direct trains to London Euston during high-demand periods and improve platform availability to reduce bottlenecks.

2. Capitalize on Peak Times:

- Schedule additional trains between 7:45 AM and 8:00 AM, as this peak period increases revenue by 63.47 units due to higher fares and passenger volume.
- Optimize pricing strategies during peak hours to maximize revenue without deterring passengers.
- Example Action: Introduce dynamic pricing for peak-hour tickets to balance demand and revenue, ensuring sufficient train capacity during these times.

3. Investigate On-Time Journey Revenue Drop:

- Conduct a deeper analysis into why on-time journeys reduce revenue by 36.95 units. This could involve reviewing pricing strategies (e.g., are on-time journeys priced too low to compete?) or assessing if delayed journeys generate additional revenue (e.g., through fees or alternative bookings).
- Adjust pricing or service offerings for on-time journeys to make them more profitable without compromising reliability.
- Example Action: Test a slight fare increase for on-time journeys to offset the revenue drop, while offering incentives (e.g., loyalty points) to maintain passenger satisfaction.

4. Address Seasonal Revenue Volatility:

- Analyze the reasons behind revenue drops in March (1K) and May (4K) compared to February (11K). This may involve reviewing journey schedules, marketing efforts, or external factors (e.g., weather, holidays).
- Implement targeted promotions during low-revenue months to boost ridership to high-revenue destinations like London Euston and Paddington.
- Example Action: Launch a spring travel campaign offering discounted fares for early bookings to London Euston, encouraging travel during typically low-revenue periods.

<u>To Reduce Refund Requests</u>

1. Minimize Delays and Cancellations:

- Focus on reducing delays and cancellations, especially for journeys paid with debit cards, as debit card users are 14.479 times more likely to request refunds.
- Address operational issues at high-delay stations like London Euston and Edinburgh Waverley (see delay recommendations below) to improve reliability.
- Example Action: Introduce a real-time monitoring system to identify and resolve potential delays before they impact passengers, reducing the need for refunds.

2. Target Adult Railcard Holders:

- Offer incentives for adult railcard holders (3.31 times more likely to request refunds) to reduce refund requests. This could include faster compensation processes, discounts on future journeys, or priority rebooking for delayed/canceled journeys.
- Example Action: Create an "Adult Railcard Loyalty Program" offering a 10% discount on the next journey if a refund is requested due to delays, encouraging retention over refunds.

3. Improve Off-Peak Journey Reliability:

- Enhance the reliability of Off-Peak journeys, as their ticket type increases refund likelihood by 1.30 times. This may involve better scheduling to avoid delays or ensuring sufficient train capacity during non-peak hours.
- Example Action: Allocate additional trains for Off-Peak routes to Wakefield (a high-delay destination) to reduce delays and improve passenger satisfaction, thereby lowering refund requests.

4. Encourage Credit Card Usage:

- Promote credit card payments, as they reduce refund likelihood by 1.04 times. Offer small incentives (e.g., a 2% discount) for credit card payments to shift passenger behavior.
- Example Action: Partner with credit card companies to offer cashback rewards for train ticket purchases, making credit card payments more attractive and reducing refund requests.

5. Address High Refund Amounts on Key Routes:

- The £17 refund amount for the London Kings Cross to York route indicates a need to address underlying issues like delays and cancellations. Focus on improving service reliability on this route to reduce refund requests.
- Example Action: Conduct a root cause analysis of refunds on this route and implement targeted improvements (e.g., better scheduling, more staff support) to enhance passenger experience.

To Reduce Journey Delays

1. **Optimize Operations at London Euston:**

- Journeys to London Euston are 16.90 times more likely to be delayed. Address congestion and operational challenges at this busy station by improving scheduling, increasing platform availability, and enhancing coordination with other train services.
- Example Action: Invest in infrastructure upgrades at London Euston (e.g., additional platforms, better signaling systems) to reduce delays and improve journey reliability.

2. **Improve Departure Operations at Edinburgh Waverley:**

- Journeys departing from Edinburgh Waverley are 14.10 times more likely to be delayed. Tackle operational issues at this station, such as congestion, maintenance schedules, or staffing shortages, to ensure timely departures.
- Example Action: Implement a dedicated departure management team at Edinburgh Waverley to streamline operations and reduce delays, especially for long-distance journeys to London Euston.

3. **Enhance Service to Wakefield:**

- Journeys to Wakefield are 13.89 times more likely to be delayed. Improve resources at this secondary station (e.g., more platforms, better scheduling) to reduce delays and ensure smoother operations.
- Example Action: Review and adjust the scheduling of trains stopping at Wakefield to minimize conflicts with other routes, ensuring more reliable service.

4. **Monitor and Improve Actual Arrival Time:**

- A decrease in Actual Arrival Time by 0.24 increases the likelihood of a journey being on time by 1.116 times. Set up daily monitoring of Actual Arrival Time to identify and address deviations early, ensuring journeys stay on schedule.
- Example Action: Use predictive analytics to forecast potential delays based on historical data and adjust schedules proactively to maintain on-time performance.

5. **Address Long-Distance Journey Challenges:**

- Long-distance journeys (e.g., from Edinburgh Waverley to London Euston) are prone to delays due to distance and external factors (e.g., weather). Implement contingency plans, such as backup trains or alternative routes, to mitigate delays on these routes.

- Example Action: Develop a weather contingency plan for long-distance routes, including pre-scheduled backup trains to minimize disruptions during adverse conditions.

3. Final Opinion :

- **Revenue**: Revenue is heavily influenced by journeys to London Euston (78.23 units) and London Paddington (38.96 units), as well as peak departure times (7:45-8:00 AM, 63.47 units). However, on-time journeys reduce revenue by 36.95 units, which warrants further investigation into pricing and operational strategies. Seasonal volatility (e.g., 1K in March vs. 11K in February) suggests a need for targeted interventions during low-revenue periods.
- **Refund Requests**: Refund requests are significantly driven by debit card payments (14.479x likelihood), adult railcards (3.31x), and Off-Peak tickets (1.30x), while credit card payments slightly reduce requests (1.04x decrease). The £17 refund amount on the London Kings Cross to York route highlights the impact of delays and cancellations on passenger satisfaction.
- **Journey Delays**: Delays are most likely for journeys to London Euston (16.90x), from Edinburgh Waverley (14.10x), and to Wakefield (13.89x). Improving Actual Arrival Time (e.g., reducing it by 0.24) increases on-time likelihood by 1.116x, emphasizing the importance of schedule adherence.
- **Relation to Visuals and KPIs**: The drop in ridership (63 passengers in April), ticket numbers (112 in April), and revenue (1K in March) is directly tied to delays and cancellations, particularly involving London Euston and Edinburgh Waverley. These issues also contribute to higher refund requests, as seen in the £17 refund amount for the route.

4. Strategic Roadmap for Implementation

To operationalize the recommendations, here's a prioritized roadmap:

1. **Short-Term (0-3 Months) :**
   - Schedule additional peak-time trains (7:45-8:00 AM) to boost revenue.
   - Implement real-time monitoring of Actual Arrival Time to reduce delays.
   - Offer incentives for adult railcard holders to reduce refund requests.
2. **Medium-Term (3-6 Months):**
   - Address operational issues at London Euston and Edinburgh Waverley through better scheduling and resource allocation.
   - Launch a spring travel campaign to boost revenue during low-revenue months (e.g., March, May).
   - Promote credit card payments with small incentives to reduce refund requests.
3. **Long-Term (6-12 Months):**
   - Invest in infrastructure upgrades at London Euston and Wakefield to reduce delays.
   - Develop a weather contingency plan for long-distance routes to minimize disruptions.
   - Conduct a detailed pricing analysis for on-time journeys to address the revenue drop.