UK Train Analysis Report Documentation

Christeen Adly
Roqya Farag
Omnia Moustafa
Wesam Sadek

4/28/2025

1. Overview and Objectives

This Power BI report provides an in-depth analysis of train operations in the UK, sourced from CSV data. The key objectives are:

- Understand operational efficiency through trip completion and delays
- Analyze financial performance via revenues and refunds
- Forecast upcoming trends in trips and revenue
- Offer actionable insights for optimization and decision-making

2. Data Source and Description

Source: Railway Comma-Separated Values (CSV) file

Contents:
This dataset encompasses information on approximately 31,653 transactions placed between December 2023 and April 2024. It provides both temporal and categorical variables for multidimensional analysis, including:

Key Columns

Column Name	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 (Contactles, 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

Additional Custom Columns Created:

- **Delay In Minutes:** Difference between scheduled and actual time.
- **Route:** The travel path taken.
- **Journey Duration:** Planned time for the journey.
- **Actual Journey Duration:** Time actually taken.
- **Hour order:** Departure hour in 24-hour format for proper sorting in charts.
- **Departure Hour:** Hour extracted from departure time.

3. Data Analysis Questions

- What is the overall distribution of train trips (on-time, delayed, cancelled)?
- What is the net revenue after refunds?
- Which stations experience the highest number of delayed departures or arrivals?
- At what times of day do most trips depart? What are the peak departure hours?
- What percentage of passengers use each type of railcard (Adult, Senior, Disabled, None)?
- During which hours of the day are the highest number of trip delays and cancellations happening? What are the peak times for delays and cancellations?
- Which routes have the highest delays and what are the major contributing factors?
- What are the most common reasons for train delays, and how many trips were affected by each reason? Also, how much average delay time is caused by each reason?
- How does the average actual journey duration compare to the scheduled (average) journey duration for each route? Which routes experience the most significant delays?
- Which specific routes have the highest number of on-time trips vs delayed vs cancelled?
- What is the total net revenue, and how is it distributed by payment method (Contactless, Credit Card, Debit Card)?
- How did the monthly revenues change from January to April, and what were the month-over-month percentage increases or decreases?
- What is the overall average ticket price, and how does it vary across ticket types (Advance, Anytime, Off-Peak)?
- How did the average ticket price change monthly from January to April, and what were the percentage changes month-over-month?
- What is the total amount refunded, and how is it distributed among different ticket types?
- How did the total amount of refunds trend monthly from January to April, and what were the month-over-month percentage changes?
- How does the average ticket price vary depending on Railcard type (None, Adult, Disabled, Senior) and Ticket Class (First Class vs Standard)?
- How does the average ticket price vary by Purchase Type (Station vs Online) and Railcard type (None, Adult, Disabled, Senior)?
- What is the distribution of refunds based on journey status?
- What are the most common reasons for delay that lead to a refund, and how many refunds are associated with each reason (Technical Issue, Staffing, Signal Failure, Weather, Traffic)?
- What factors most significantly influence the likelihood of a journey being delayed?
- How has the total number of trips changed over time (from Jan 2024 onward), and what is the forecasted trend for future months?

- How has revenue changed over time (from Jan 2024 onward), and what is the expected forecasted trend for future months?

4. Data Cleaning Process

Data preprocessing was a crucial step to ensure accurate analysis. The following steps were undertaken in **Power Query**:

Cleaning Steps: In "Reasons for Delay" column:

- ✓ Renaming column to become "Trip Disruption Reason"
- ✓ Merging "Weather" and "Weather Conditions" using "Replace value".
- ✓ Merging "Signal Failure" with "Signal failure" using "Replace value".
- ✓ Merging "Staffing" with "Staff Shortage" using "Replace value".
- ✓ Replacing all "Blanks" with "On Time".

Added Columns in Power Query:

1. **Delay In Minutes**:
   This column calculates the difference between the scheduled and actual journey times in minutes to help identify how late each trip was.

2. **Route**:
   This column indicates the travel path or line taken by the vehicle, added to group and analyze delays by route.

3. **Journey Duration**:
   This shows the planned journey time, included for comparison with actual durations to assess performance.

4. **Actual Journey Duration**:
   This column captures the real time taken for the journey, used to calculate delays and evaluate operational efficiency.
5. **Hour order**:
   This column converts the departure hour into a 24-hour numeric format to ensure correct chronological sorting when visualizing data in charts.

6. **Departure Hour**:
   This extracts the hour from the departure time, allowing for hourly trend analysis in delays or traffic patterns.

These columns enriched the dataset and enabled more flexible, meaningful visualizations and DAX calculations throughout the report.

5. Data Modeling

Data modeling is minimal; only a direct relationship between the date and railway tables is needed.

6. Analysis Methodology:

6.1. DAX Measures (All in `All Measures` Table)

Total Trips	This measure computes the total number of trips recorded in the railway table.
Completed Trips	This measure calculates the number of completed trips by subtracting the total cancelled trips from the total trips.
Delayed Trips	This measure counts the total number of delayed trips by filtering the "railway" dataset for rows where the "Journey Status" is marked as "Delayed."
Cancelled Trips	This measure counts the total number of cancelled trips by filtering the "railway" dataset for rows where the "Journey Status" is marked as "Cancelled."
Average Delay in Minutes	This measure computes the average delay time, in minutes, from the "Delay in minutes" column of the "railway" dataset.
Average Delayed Trips Per Hour	This measure calculates the average number of delayed trips per hour, dividing the total number of delayed trips in a selected hour by the distinct count of days with delays during that hour.
Average Journey Duration	This measure calculates the average journey duration from the "Journey Duration" column of the "railway" dataset.
Average On-Time Trips Per Hour	This measure calculates the average number of on-time trips per hour, dividing the total number of on-time trips in a selected hour by the distinct count of days with on-time trips during that hour.
Average On-Time Trips Per Route	This measure calculates the average number of on-time trips per route, dividing the total number of on-time trips on a selected route by the distinct count of days with on-time trips on that route.
Average Trips Per Hour	This measure calculates the average number of trips per hour, dividing the total number of trips in a selected hour by the distinct count of days with trips during that hour.
Average Delayed Trips Per Route	This measure calculates the average number of delayed trips per route, dividing the total number of delayed trips on a selected route by the distinct count of days with delays on that route.
Average Actual journey duration	This measure calculates the average value of the actual journey duration from the "railway" dataset.
Average Cancelled Trips Per Hour	This measure calculates the average number of cancelled trips per hour, dividing the total number of cancelled trips in a selected hour by the distinct count of days with cancellations during that hour.
Average Cancelled Trips Per Route	This measure calculates the average number of cancelled trips per route, dividing the total number of cancelled trips on a selected route by the distinct count of days with cancellations on that route

MoM_Trip	This measure calculates the month-over-month percentage change in the number of total trips. It formats the result as a percentage and adds an upward (▲) or downward (▼) arrow based on the trend direction.
MoM_Trip2	This measure calculates the raw month-over-month change in the number of total trips without any formatting or additional symbols.
Trip_Color	This measure assigns a color based on the value of [MoM_Trip2]: - "Green" if the month-over-month trip count increased. - "Red" if the month-over-month trip count decreased.
On Time Trips	This measure counts the number of trips where the 'Journey Status' is marked as "on time" in the railway table.
Percentage of Cancelled Journeys	This measure computes the ratio of cancelled trips to total trips in the dataset.
Percentage of Completed Journeys	This measure computes the ratio of completed trips to total trips in the dataset.
Percentage of Delayed Journeys	This measure computes the ratio of delayed trips to total trips in the dataset.
Percentage of On Time Journeys	This measure computes the ratio of on-time trips to total trips in the dataset.
Trips and average delay label	This measure creates a text label combining the total number of trips and the average delay time, formatted for easy reading.
MoM_NetRevenue	This DAX measure calculates the month-over-month percentage change in net revenue, displaying it with one decimal and an up (▲) or down (▼) arrow based on the trend.
MoM_NetRevenue 2	This DAX measure calculates the month-over-month net revenue growth rate as a decimal value, without any formatting or arrows.
Net-Revenue	This DAX measure calculates the net revenue by summing revenue only from trips that had no refund request.
Revenues	This DAX measure calculates the total revenue by summing the price of all train tickets.
NetRevenue_Color	This DAX measure assigns a "Green" label if month-over-month net revenue growth is positive, otherwise it returns "Red".
Percentage of Railcards	This DAX measure calculates the percentage of a specific railcard type relative to the total number of railcards.
Railcards	This DAX measure counts the number of railcards used, excluding entries where the railcard value is "None".
Refund	This measure counts the number of refund requests where the 'Refund Request' field equals "Yes" in the railway table.
Refunds	This measure sums the 'Price' values for all entries in the railway table where the 'Refund Request' is marked as "Yes".
MoM_Refund	This measure calculates the month-over-month percentage change in total refunds. It compares the current month's

	refunds with the previous month, formats the result as a percentage, and adds an upward (▲) or downward (▼) arrow based on the trend.
MoM_Refund2	This measure computes the raw month-over-month change in refunds without formatting. It returns the percentage change value but without symbols or text formatting.
Percentage of Refund	This measure calculates the ratio of the number of refund requests to the total number of refund entries in the dataset.
Refund_Color	This measure assigns a color based on the value of [MoM_Revenue2]: - "Green" if the value is positive, - "Red" if the value is negative.
Avg. Price	This DAX measure calculates the average ticket price across all train journeys.
MoM_Price	This measure calculates the month-over-month price change as a percentage, formatted with an upward or downward arrow based on whether the price has increased or decreased compared to the previous month.
MoM_Price2	This measure calculates the month-over-month price change as a decimal percentage without any formatting or symbols, showing the raw value of the change compared to the previous month.
Price_Color	This measure evaluates the value of [MoM_Price2] (Month-over-Month price change) and returns "Green" if the value is greater than 0, indicating a price increase, and "Red" if it's 0 or negative, indicating no change or a decrease.

These allow consistent metrics across multiple pages and support advanced visuals like KPIs and conditional formatting.

6.2. Dashboard Creation Process

- Created clean slicers for filtering by trip date, routes, Railcard type, journey status, trip disruption reason, ticket class, ticket type, purchase type and payment method.
- Implemented custom themes for consistent design.
- Used bookmarks and buttons for interactivity and focus areas.
- Tooltips with dynamic context on hover.

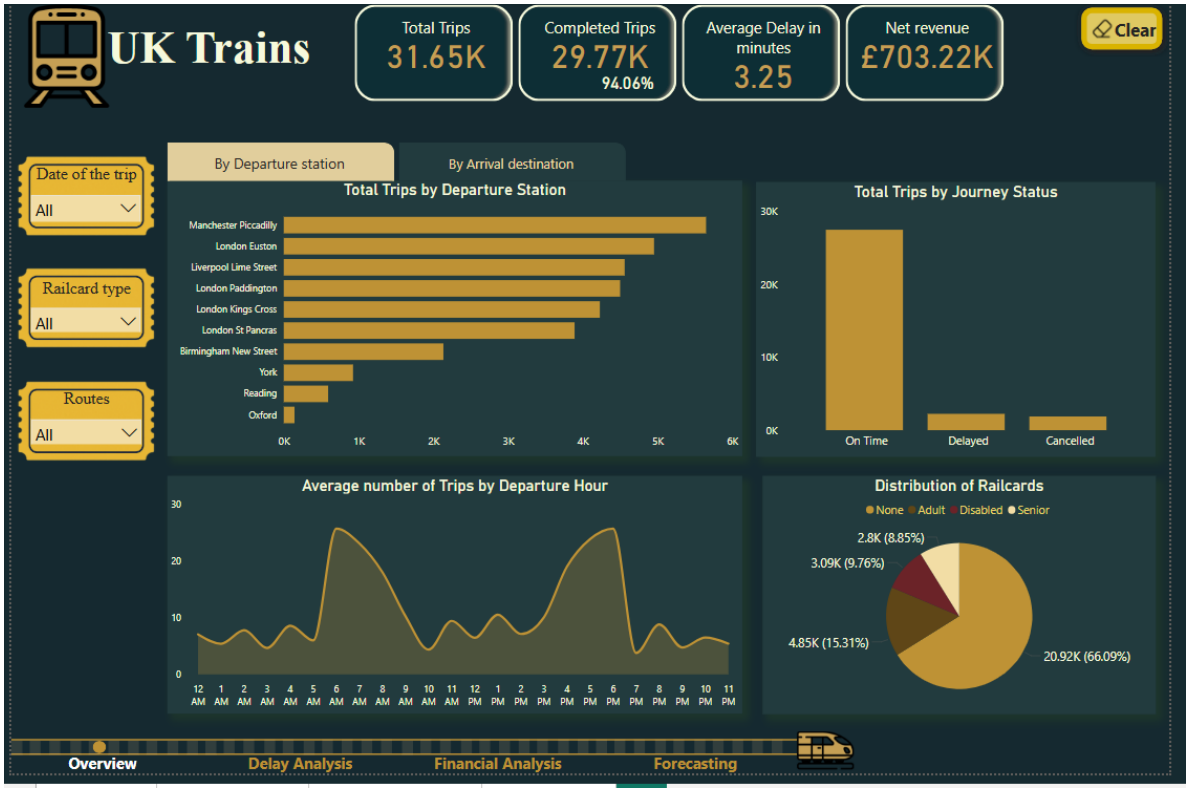
The data was displayed with the four dashboards that were made up:

- **Overview Dashboard**
  Cards: Total Trips, Completed Trips (%), Average Delay in Minutes and Net Revenue.
- **Delay Analysis**
  Cards: Total Trips, On-Time Trips, Delayed Trips and Cancelled Trips.
- **Financial Analysis**
  Cards: Net Revenues, Avg. Price and Refund.
- **Forecasting** and key influencers.

6.3. Page-by-Page Visualization

Page 1: Overview

- Bar chart – Total Trips by Departure Station and Arrival Destination.
- Column chart – Total Trips by Journey Status.
- Stacked Area chart – Average Number of Trips by Departure Hour.
- Pie chart – Distribution of Railcards.



Key Findings:

- **Trips Performance:** Out of 31.65K total trips, **94.06%** were successfully completed, showing strong operational performance. The average delay across all trips is **3.25 minutes**, indicating minimal service disruptions. Net revenue achieved is approximately **£703.22K**.
- **Top Departure Stations:** Manchester Piccadilly, London Euston, and Liverpool Lime Street are the leading stations by trip volume.
- **Top Arrival Destinations:** Birmingham New Street, Liverpool Lime Street and York are the busiest arrival destinations, indicating their central role in the network's operations.
- **Journey Status:** Majority of trips were completed on time with minimal instances of delays and cancellations.
- **Trip Patterns:** Peak trip departures occur around **6 AM** and **6 PM**, reflecting typical commuter travel patterns.

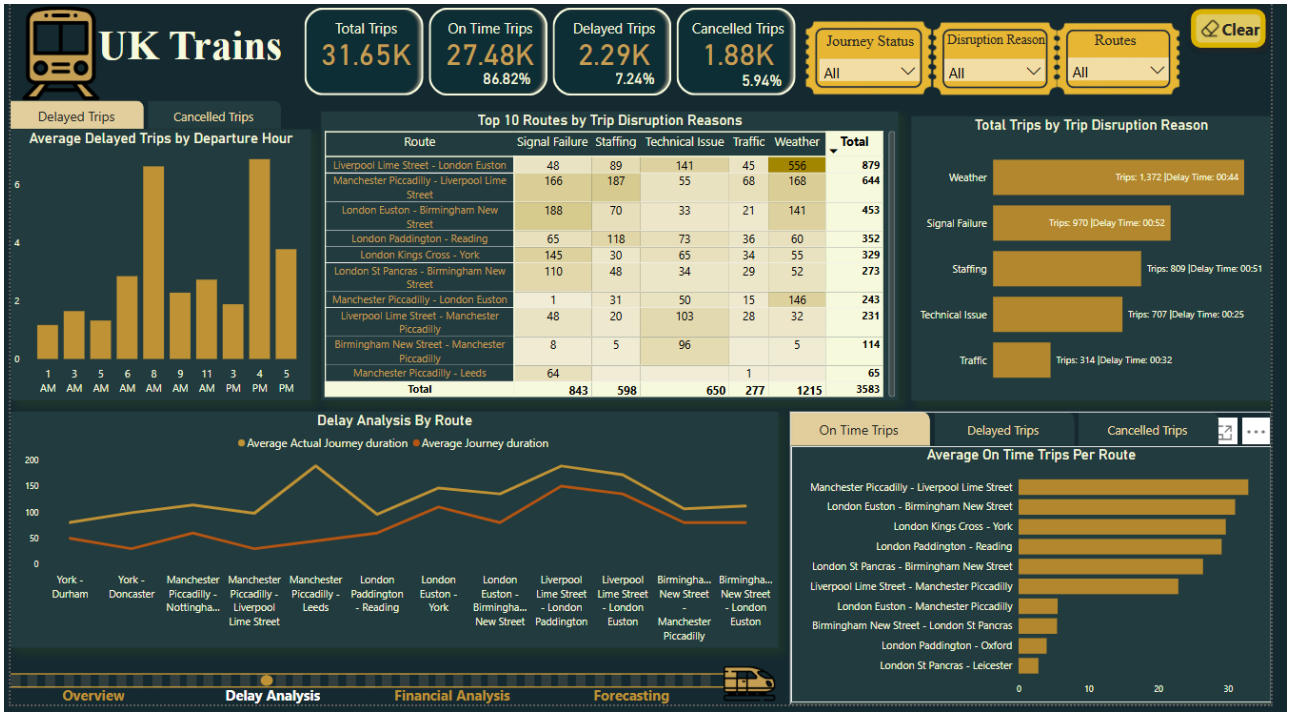
- **Railcard Usage:** 66.09% of travelers did not use a railcard. Among railcard users, Senior and Disabled railcards accounted for a combined share of over **17%**, while Adult railcards are over **15%** becoming the most used railcards type.

Analysis Summary:

- The railway service shows **high efficiency** with **strong trip completion rates** and **minimal average delays**.
- **Manchester Piccadilly** is the most active **departure station**, while **Birmingham New Street** is the key **arrival destination**.
- **Peak travel times** align with **standard commuting hours**, stressing the need for **reliable service** during these periods.
- **The usage of railcards**, especially among **senior and disabled** passengers, highlights a valuable opportunity to expand targeted promotions. By offering more tailored discounts and campaigns, the railway service could encourage greater adoption of railcards, increase customer loyalty, and boost overall ridership.

Page 2: Delay Analysis

- Column chart – Average Delayed and Cancelled Trips by Departure Hour.
- Table – Top 10 Routes by Trip Disruption.
- Bar chart – Total Trips by Trip Disruption Reason.
- Line chart – Delay Analysis by Route (Average Actual Journey Duration vs Scheduled).
- Bar chart – Average Trips per Route for on time, delayed and cancelled trips.



Key Findings:

- **Disruption Causes**: Weather-related issues are the leading cause of trip disruptions, followed by signal failures and technical issues, indicating the need for operational resilience against external and internal factors.
- **Peak Delay Hours**: Delays are more frequent during early morning (8 AM) and late afternoon (4 PM) hours, suggesting peak commuter times are particularly vulnerable.
- **Peak Cancellation Hours:** Cancellations are most frequent during early morning (7 AM) and early evening (6 PM), indicating operational challenges during peak travel periods.
- **Route-Specific Impact**: Certain routes, such as Liverpool Lime Street to London Euston and London Euston to Birmingham New Street, experience higher disruption frequencies, highlighting critical areas for focused improvements.
- **Journey Duration Variability**: Analysis shows that some routes consistently face longer actual journey durations compared to their average, signalling possible systemic issues requiring route-specific interventions.
- **High On-Time Performance Routes**: Despite overall disruptions, certain routes like Manchester Piccadilly to Liverpool Lime Street maintain a strong on-time performance, offering models for best practices.

Analysis Summary:

- **Weather conditions** are the main cause of **train delays**, highlighting the need for **infrastructure upgrades** and **better contingency planning**.
- **Delays during commuter peaks** stress the importance of **optimizing scheduling** and **staffing** during critical hours.
- **Frequently disrupted routes** should be **prioritized** for **targeted operational improvements**.
- The **variation between actual and average journey times** across routes shows the need for **ongoing monitoring** and **tailored performance initiatives**.
- **Consistently strong-performing routes** offer **benchmarks** for **system-wide improvements**.

Page 3: Financial Analysis

- Bar chart – Net Revenues by Payment Method.
- Bar chart – Average Price by Ticket Type.
- Bar chart – Refund by Ticket Type.
- Column chart – Revenues Trend by Month.
- Column chart – Price Trend by Month.
- Column chart – Refund Trend by Month.
- Bar chart – Average Price by Railcard and Ticket Class.
- Column chart – Average Price by Purchase Type and Railcard.

Key Findings:

- **Highest Net Revenue Source:** Credit cards contribute the largest share of net revenues (65.48%), followed by contactless payments.
- **Average Ticket Price:** The overall average ticket price is £23.44, with Anytime tickets being the most expensive (£39.20) and Advance tickets the cheapest (£17.61).
- **Refund Distribution:** Refunds total £38.70K, mainly from Advance and Off-Peak tickets.
- **Revenue Trend:** March saw the highest revenue growth (+21.8%), while February had the sharpest drop (-19.7%).
- **Price Trend:** Ticket prices decreased in February (-15.3%) but rebounded in March (+15.3%) and slightly increased in April (+4%).
- **Refunds Trend:** Refunds increased sharply in March (+34.5%) but dropped again in April (-11.9%).
- **Railcard Discounts Impact:** Tickets without railcards are the most expensive. Senior and Disabled railcards significantly reduce average ticket prices.
- **Purchase Method Impact:** Tickets bought online generally have lower average prices compared to station purchases.

Analysis Summary:

- **Credit card payments** generate the majority of revenues (**65%**), followed by **contactless payments**.
- The **average ticket price** is **£23.44**, with **Anytime tickets** being the most expensive and **Advance tickets** the cheapest.
- **Refunds** amount to **£38.70K**, mainly from **Advance** and **Off-Peak** tickets.
- **Revenue** and **price trends** dropped in **February** but **recovered** in **March**.
- **Refunds** spiked in **March** and slightly **declined** in **April**.
- **Prices** are higher for passengers **without railcards** and for those **buying at stations**.
- **Key actions**: promote **online purchases**, encourage **railcard use**, and improve **refund policies**.

Page 4: Forecasting

- Pie chart – Refund by Journey Status (Cancelled vs Delayed).
- Bar chart – Refund by Trip Disruption Reason (Technical, Signal Failure, etc.).
- Bubble chart – Key Influences on Journey Status.
- Line chart – Total Trips by Month (with Forecast).
- Line chart – Refunds by Month (with Forecast).
- Line chart – Revenues by Month (with Forecast).

Key Findings:

- **Journey Status:** 51.16% of trips resulted in cancellations, slightly higher than delays at 48.84%.
- **Top Disruption Factors for Delays:** Weather disruptions increase delay likelihood by **14.81x**. Purchases made at the station increase delay likelihood by **13.84x**. Technical issues contribute **11.03x** higher chances of delay.
- **Primary Disruption Reason for Refunds:** Technical Issues lead the causes, followed by Staffing and Signal Failures.
- **Forecasting Trends: Total Trips** are expected to steadily grow from May 2024 onwards. **Refunds** are forecasted to slightly increase with visible seasonal variations. **Revenue** shows a positive trend with a notable upward projection mid-2024.

Analysis Summary:

- **Delayed journeys** have slightly exceeded **cancellations**, mainly due to **weather conditions** and **technical challenges**.
- **Refund requests** are mostly triggered by **technical disruptions**.
- **Forecasted trends** show a **consistent increase** in **trips**, **refunds**, and **revenues**.
- This suggests **growing service demand** but also emphasizes the need for **operational improvements** to better handle **disruptions**.

7. Challenges and How to Overcome Them

1. **Weather-Related Delays**:
   - **Challenge**: Weather disruptions are a recurring cause of delays.
   - **Solution**: Invest in weather-resistant infrastructure and real-time monitoring systems to better predict and manage delays due to adverse weather conditions.
2. **Operational Challenges During Peak Times**:
   - **Challenge**: Delays and cancellations peak during early morning and late afternoon due to high passenger volume.
   - **Solution**: Implement better scheduling and increase staffing during peak hours to ensure smoother operations and minimize delays.
3. **Route-Specific Disruptions**:
   - **Challenge**: Certain high-traffic routes are more prone to disruptions.
   - **Solution**: Focus on improving operational efficiency and reliability for these routes, possibly by analyzing historical data for root causes and implementing preventive measures.
4. **Railcard Underutilization**:
   - **Challenge**: 66.09% of passengers do not use railcards, limiting opportunities for discount promotions.
   - **Solution**: Launch targeted marketing campaigns to raise awareness about railcard benefits, and offer exclusive promotions or discounts to encourage usage, particularly for senior and disabled customers.
5. **Refund Management**:
   - **Challenge**: High volume of refunds during certain months, especially in cases of cancellations.

- **Solution**: Refine refund policies and streamline processes to ensure faster response times while maintaining customer satisfaction.

8. Conclusions and Recommendations

Conclusions:

1. **Overall Performance**: The railway service shows strong operational performance with a 94.06% trip completion rate and minimal average delays (3.25 minutes), resulting in a net revenue of £703.22K.
2. **Peak Travel Patterns**: Peak travel times occur during typical commuting hours (6 AM and 6 PM), stressing the need for robust service during these periods.
3. **Weather Impact**: Weather-related disruptions are the primary cause of delays, with significant impact during peak times, necessitating improvements in weather resilience.
4. **Railcard Usage**: A large proportion of passengers (66.09%) do not use railcards. Expanding railcard usage, especially among senior and disabled passengers, could drive customer loyalty and revenue growth.
5. **Revenue Trends**: Credit card payments generate the highest revenue, with fluctuations observed in ticket prices and refund trends over the months.

Recommendations:

1. **Service Resilience**: Enhance infrastructure to address weather-related disruptions and improve contingency planning to minimize delays, particularly during peak hours.
2. **Railcard Promotion**: Increase marketing efforts around railcard usage, particularly for senior and disabled passengers, to boost adoption and increase revenue.
3. **Peak-Time Optimization**: Focus on optimizing scheduling and staffing during the early morning and late afternoon hours to reduce delays and cancellations.
4. **Targeted Operational Improvements**: Prioritize routes with higher disruption frequencies for specific operational improvements, including route-specific monitoring and management.
5. **Encourage Online Purchases**: Encourage passengers to purchase tickets online as these tend to have lower prices, improving both customer satisfaction and revenue.
6. **Refund Policy Review**: Streamline the refund process, particularly for Advance and Off-Peak tickets, to minimize customer dissatisfaction and financial losses.