Module Code: CT048-3-M-BIS

Module Name: Business Intelligence Systems

Report

Analyzing Train Delays in UK National Rail

Student Name: Muhammad Yousouf Ali Budullah

TP Number: TP086704

Intake Code: APDMF2501DSBA(BI)(PR)

Programme: MSc Data Science & Business Analytics

Module Lecturer: Assoc. Prof. Dr. Selvakumar Samuel

Date of Submission: June 4, 2025

Table of Contents

List of Figures

1 Introduction

The UK National Rail is one of the most intricate and frequently used transportation networks in the United Kingdom. It connects major cities, hub, and suburban areas through an extensive and vital rail system. Given the large number of daily commuters and intercity travellers, the UK National Rail's operations rely heavily on schedule accuracy and efficient service delivery. This report focuses on the national rail data set representing ticket sales and journeys throughout January to April 2024. It comprises of insightful transactional details such as ticket type, departure and arrival stations, time stamps, ticket price and journey status.

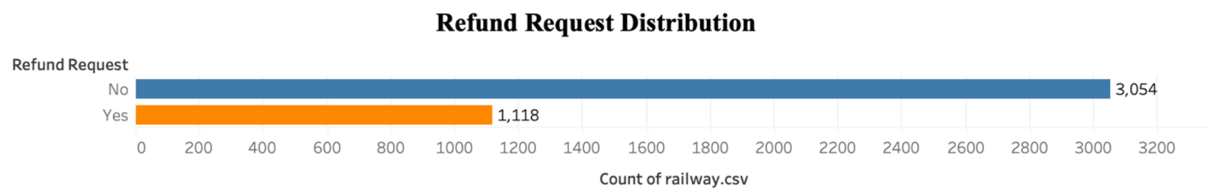


Figure 1: Refund Request Distribution

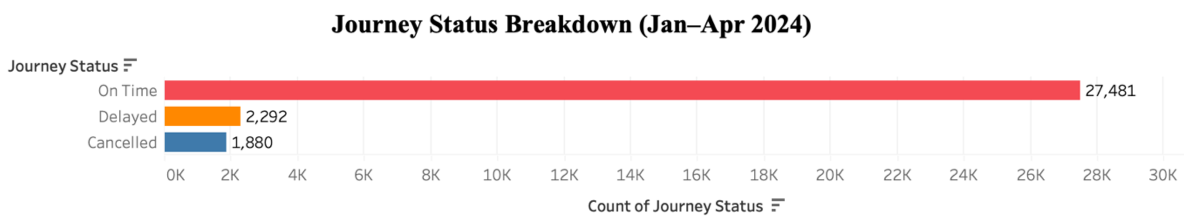


Figure 2: Journey Status Breakdown

During the period which this data was collected it was observed that majority of journeys were completed on time, however a notable amount experienced either delays or cancellations. These disruptions not only impact customer satisfaction but also in some case result in refund request, rendering them key operational and financial concern.

The purpose of this report is to analyse these patterns in depth and discover the reasons behind delays and refund requests. Furthermore, it attempts to distinguish areas where scheduling can be optimised and recommend data driven strategies to prevent common causes of delays. The analysis is assisted by descriptive and diagnostic data methods, contextual information from external sources, and knowledge from academic and industry best practices to guide business intelligence solutions.

Problem Statement: Many train rides during the period January to April 2024 experienced delays due to various reasons.

Objectives:

1. To conduct analysis of train delay trends, causes and associated ticket refunds.
2. To explore and recommend optimized scheduling.
3. To recommend strategies to prevent common causes of train delays.

2 Business and Data Understanding

2.1 Business Understanding

The UK National Rail plays an important role in the United Kingdoms transportation infrastructure facilitating over 1.6 billion journeys in the 2024 financial year (Department for Transport, 2024). This heavy usage highlights the national rails importance to daily commuters, tourists and the broader economy.

Every day, the rail networks facilitates approximately 3 million journeys, contributing a substantial amount to local economies. Passengers spends nearly 100 billion British Pounds every year in local communities throughout their journeys (Network Rail, 2024).

Therefore, ensuring fast and reliable rail services is imperative for retaining public trust and maximising operational efficiency. Delays and cancellations not only cause frustration for passengers, but they also cause financial implications, such as refund requests and substantial revenue loss. To understand the patterns and causes of these delays is important to increasing service quality and operational efficiency.

2.2 Data Understanding

The dataset used to conduct the analysis consists of records of UK train journeys and ticket purchase from January to April 2024. Each records depicts a single ticket transaction and includes the following information.

- **Ticket Purchase Details:** Date and time of purchase, method of payment, purchase channel, and price.
- **Ticket Information:** Ticket type (e.g., Advance, Anytime), class, and railcard usage.
- **Journey Details:** Departure and arrival stations, scheduled and actual times, and journey status (on time, delayed, cancelled).
- **Customer Response:** Whether a refund was requested, and where applicable, a stated reason for delay or cancellation.

The dataset use was mostly clean and well structured, it had no missing values or major formatting issues. The only inconsistency observed was related to the Reason for Delay field, where some values appeared in slightly different formats e.g., “Signal Failure” vs “signal failure”, and “Weather conditions” vs “Weather”). This was resolved during analysis in Tableau by grouping similar delay reasons under categories for consistency.

In Addition, feature creation and transformation was applied to facilitate more in depth analysis.

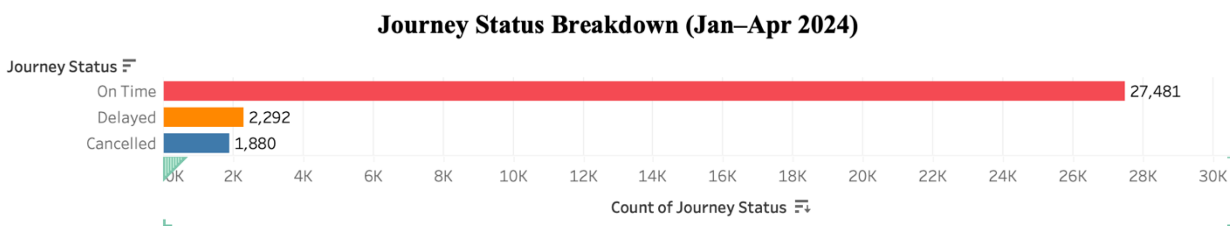
- A delay duration column was calculated by comparing scheduled and actual arrival times.
- A delay severity category was created to classify journeys into Minor, Moderate, Severe, or Cancelled.
- A route identifier was formed by combining departure and arrival station names.

3 Descriptive Analysis

"What patterns exist in the data?"

3.1 Journey Status Overview

The bar chart displays that most train journeys between January and April 2024 were completed on time, while a smaller number experienced delays or were cancelled.



This breakdown confirms that while delays are not predominant, they still occur frequently enough to justify an in-depth analysis especially as they could potential impact customer satisfaction and refund activity.

3.2 Delay Severity Distribution

To better understand the impact of service disruptions, each delayed or cancelled journey was given one of four severity levels based on the calculated delay duration.

- Minor (up to 5 minutes)
- Moderate (6–15 minutes)
- Severe (more than 15 minutes)
- Cancelled (no delay time recorded but considered a disruption)

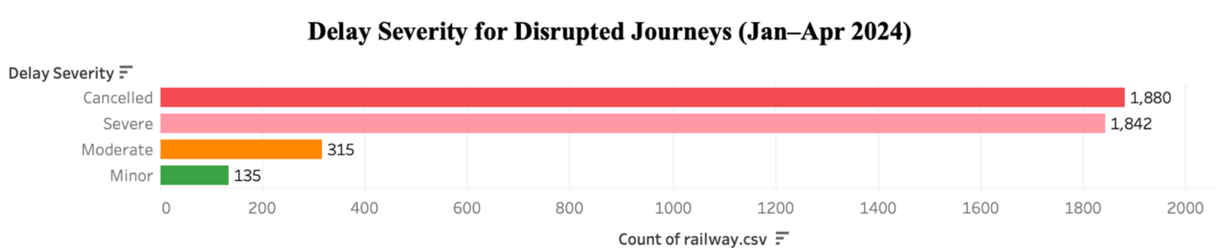


Figure 3: Delay Severity for Disrupted Journeys

The majority of disrupted journeys were either Severely delayed or Cancelled. Minor and Moderate delays were not as common, indicating that when disruptions do occur, they typically of significance.

3.3 Ticket and Purchase Trends

The dataset shows that Advance tickets were most frequently purchased ticket type during the period January to April 2024. These are typically tickets booked ahead of time at lower prices and could be more sensitive to service reliability.

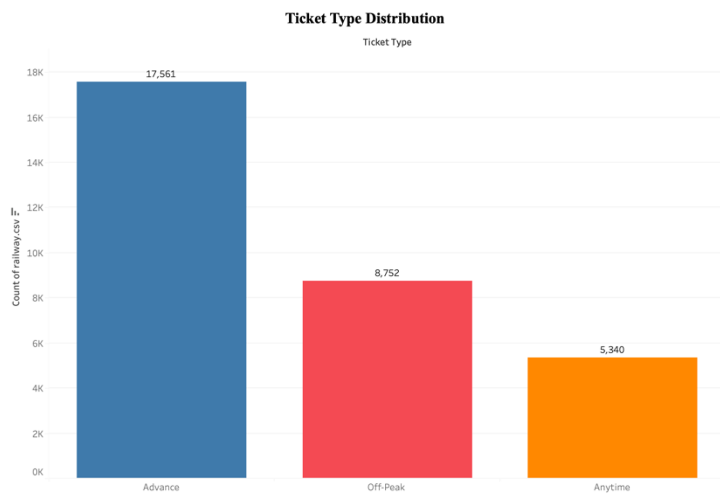


Figure 4: Ticket Type Distribution

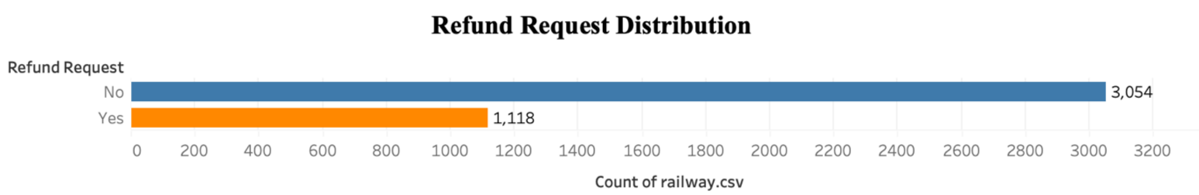
In regards of purchasing behavior, a large number of tickets were bought online, reflecting a shift toward digital services. Station-based purchases were still made albeit less frequently.

Figure 5: Ticket Purchase Method Distribution

Understanding these trends is useful in determining which types of passengers are most affected by disruptions, as refund expectations and communication preferences may vary depending on their ticket type and purchase method.

3.4 Refund Request Summary

This chart is centered around journeys that were delayed or cancelled, as refund requests are only relevant in the event of service disruption. Amid these disrupted journeys, the number of passengers who demanded refunds remained relatively low, albeit substantial.



While many passengers did not file a refund request, the considerable number requests illustrate the importance of service reliability on customer behavior.

3.5 Route and Pricing Overview

There were Several high traffic routes that occurred regularly in the dataset. Manchester Piccadilly to Liverpool Lime Street was among the most popular. These routes could represent notable commuting routes or popular city-to-city transits.

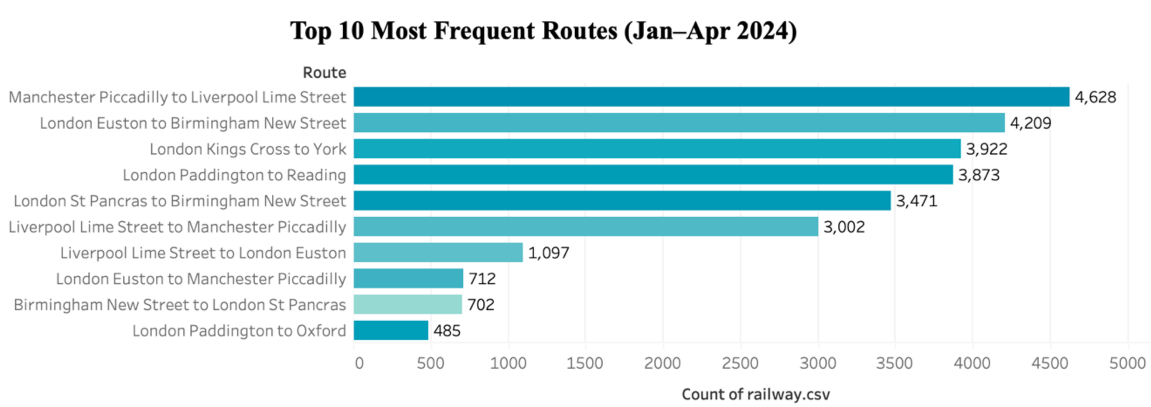


Figure 6: Top 10 Most Frequent Routes

Regarding ticket pricing, Advance tickets are seen to have rather lower average fares, while Anytime and Off-Peak tickets generally saw higher prices. Understanding where these higher priced tickets are used is helpful in determining the monetary effect of delays or cancellations.

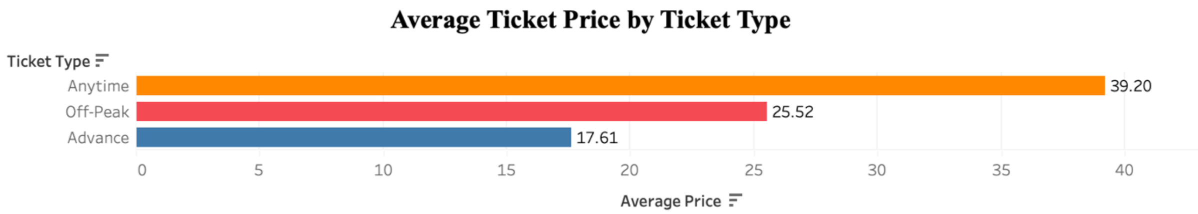


Figure 7: Average Ticket Price (£) by Ticket Type

These routes and pricing trends justify further investigation into which group of customers are most impacted by disruptions and how service can be improved.

4 Diagnostic Analysis

"Why are these patterns occurring?"

Diagnostic analysis improves on previous findings through investigating why delays and refund requests happened. It places an emphasis on identifying the primary causes of disruptions, where and when they occurred the most, and how they impacted customer behaviour.
Specfically, it investigates

- Common reasons for delays and where they occurred most frequently
- Patterns by route and time of day
- What factors were most likely to trigger refund requests

The intent is to determine what is the cause of these disruptions and areas that may require improvements to operations.

4.1 Delay Causes

To simplify the analysis, similar and related delay reasons were grouped into categories. This lowers the similarity between equivalent reasons and gives a better idea of the causes service disruptions.

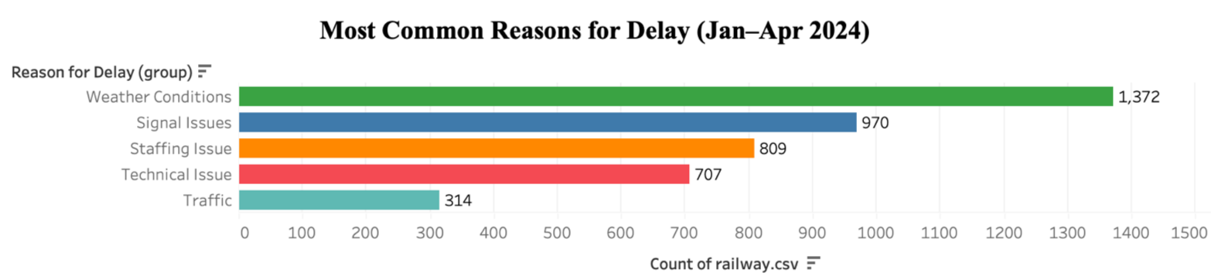


Figure 8: Most Common Reasons for Delays

The most common reasons for disruptions was Weather Conditions, making up 1,372 delays and cancellations. This was followed by Signal Issues (970) and Staffing Issues (809), which demonstrate operational challenges the UK National Rail might be facing. Technical Issues, contributed to 707 delays. While Traffic related delays, appeared less frequently but still notable.
The results indicate that, while external factors such as weather contribute a substantial amount to disruptions, a significant amount are still caused by internal difficulties which can be addressed through better planning and infrastructure adjustments. These topics can be addressed for future research and long-term service improvement plans.

4.2 Refund Triggers

To better understand what causes passengers to request a refund, this section compares refund behaviour across both delay severity and ticket types.

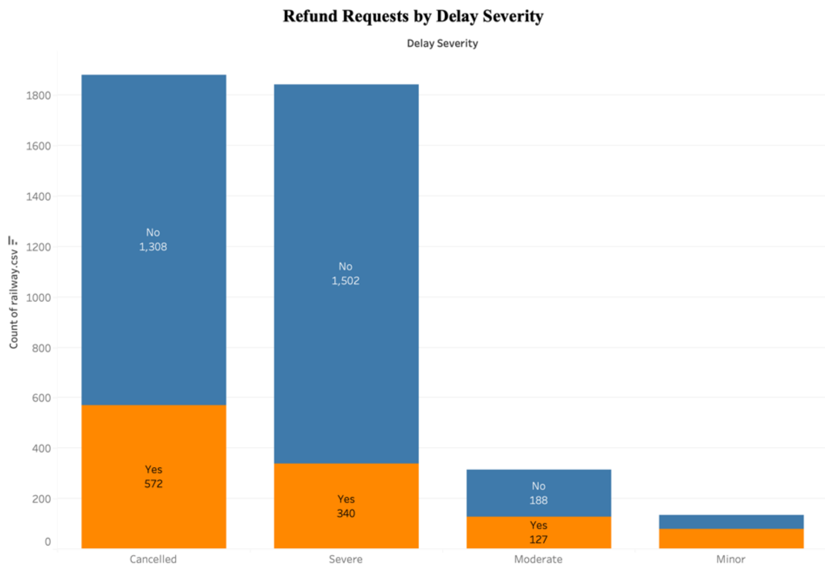


Figure 9: Refund Requests by Delay Severity Distribution

The data illustrated a clear tendency where refund requests increases considerably with the severity of the delay. A large number of refunds were for cancelled or severely delayed journeys, with only a handful being for minor delays. This indicates that more disruptive service interactions are more inclined to result in reimbursement.

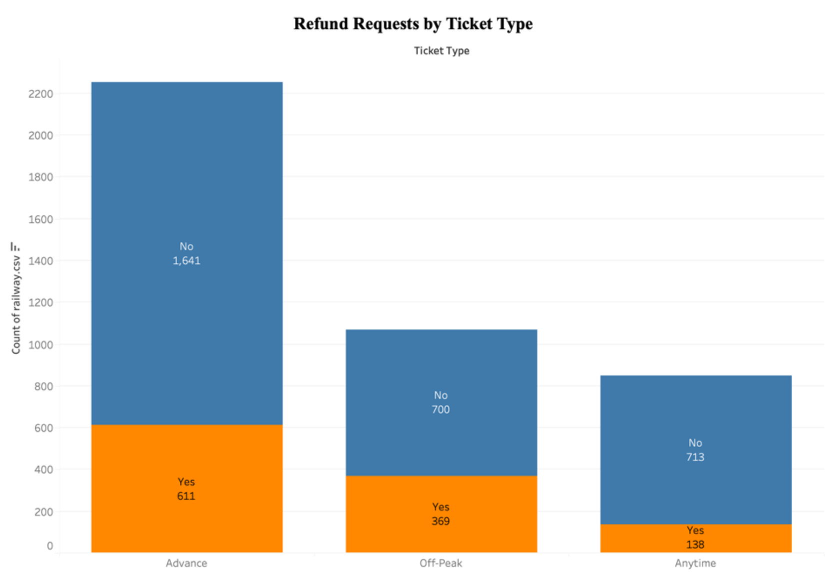


Figure 10: Refund Request by Ticket Type Distribution

The second chart reveals that Advance ticket customers were particularly affected by service disruptions, requesting more refunds than Off-Peak or Anytime ticket holders. This may have been connected to the harsher travel restrictions related to Advance tickets, such as specified departure times, which make those passengers less flexible and more inclined to request refunds when disruptions occurred.

Altogether these findings suggest that refund behaviour is mostly driven by a combination of delay severity and ticket restrictions, offering helpful guidance for customer service policies and compensation planning.

4.3 Route Based Risk

Some train routes experienced longer delays than others, suggesting repeating operational or infrastructure challenges in specific parts of the network.

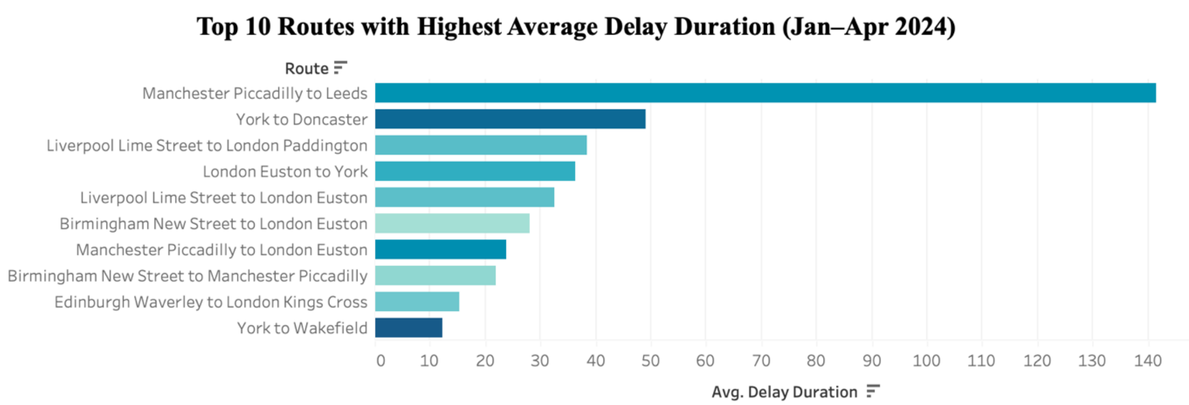


Figure 11: Top 10 Routes with Highest Average Delay Duration

The route from Manchester Piccadilly to Leeds had the longest average delay duration (minutes) , followed by York to Doncaster and Liverpool Lime Street to London Paddington .These delays could be attributed to high commuter numbers, infrastructure challenges, or traffic during peak travel hours.

Routes connecting big cities, especially those between London Euston and Manchester, were frequently in the top ten. This reveals that long-distance intercity routes may be more vulnerable to prolonged delays, either due to more scheduling, longer route times, or rail infrastructure.

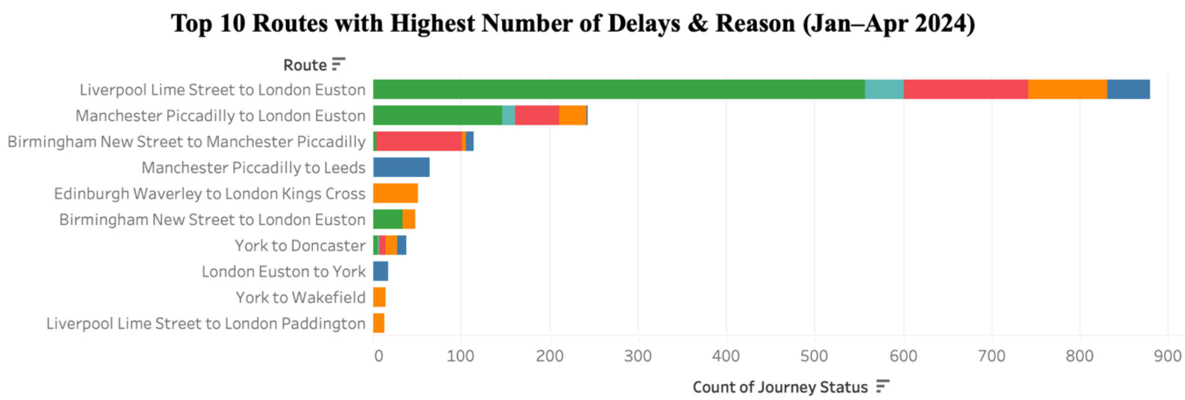


Figure 12: Top 10 Routes with Highest Number of Delays & Reason

In addition to delay duration, it's important to understand which routes experience the highest number of delays. The chart below shows the top 10 most frequently delayed routes, with color-coded bars representing delay reasons,

Green = Weather, Red = Technical, Orange = Staffing, Blue = Signal, Teal = Traffic

Routes such as Liverpool Lime Street to London Euston and Manchester Piccadilly to London Euston experienced the highest volume of delays, mostly due to weather (green) and technical issues (red). While some of these routes didn't top the list for longest delays, their frequency of disruption indicates recurring operational hotspots.

Identifying vulnerable routes is helpful for prioritising operational adjustments, repair, and schedule optimization. Addressing delays on these routes could help reduce total disruptions and improve the customer experience.

4.4 Delay Timing Patterns

In addition to route-based delays, disruptions were also more common during morning peak hours at 8 AM and again during the afternoon peak 5 PM. These periods are typical of daily commuter traffic.

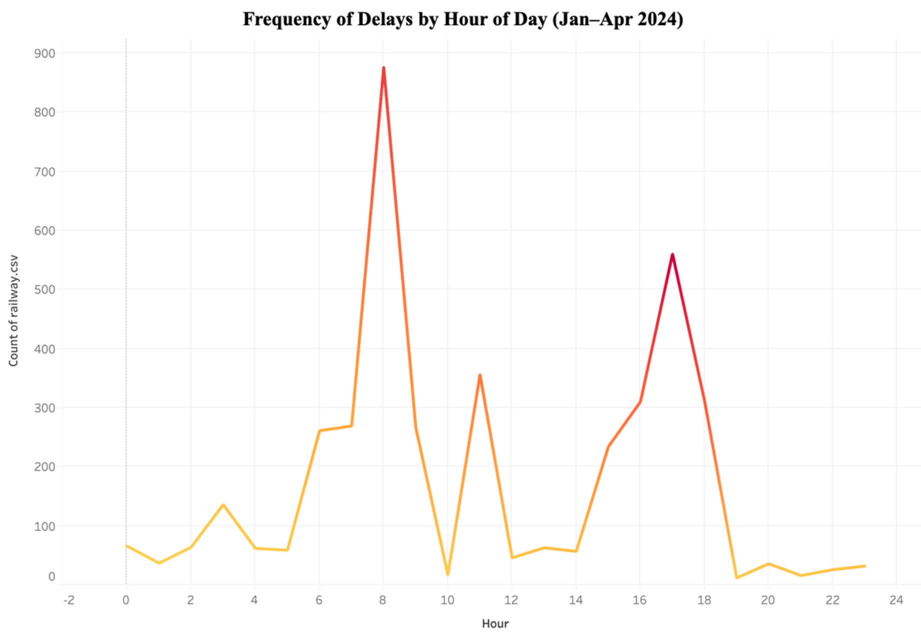


Figure 13: Frequency of Delays by Hour of Day (24H)

This trend emphasizes the significance of managing high-demand service windows more proactively through better scheduling and predictive planning.

5 Information Analysis

“What external events or industry factors may help explain what we saw in the data?”

To better understand the data's delay and refund insights gathered from diagnostic analysis, this section explores the events and trends that influenced UK train services between January and April 2024. External factors, such as bad weather, industry actions, and labor shortages, to give a broader context that data analysis alone cannot explain.

5.1 Weather Disruptions

Two large storms, Storm Henk in January and Storm Kathleen in April, had severe effects on train infrastructure throughout the UK. Storm Henk and Storm Kathleen produced torrential rain and strong gusts, that led to severe flooding and transportation disruption, notably in the north-west of England, including cities such as Manchester and Liverpool. These incidents explain the data's long average delays on Manchester-connected routes and large number of disruptions north-west.

5.2 Industrial Action and Staffing Shortages

National rail strikes took place in January and February 2024, interrupting operations in several locations and significantly decreased service availability. Furthermore, once services were restored, the system experienced repercussions well into March, resulting in more delays. This is consistent with the large number of "Staffing Issues" and "Signal Problems" observed in the grouped delay analysis. Due to this, the UK government stated that the minimum train driver age will be reduced from 20 to 18, with the intent of resolving the industry's well-known personnel shortage. This policy further confirms the data's trend of delays due to staffing issues.

5.3 Public Sentiment and Passenger Experience

Public frustration is apparent on social media platforms such as TikTok and Reddit, where people complained about cancellations, bus replacements, and longer travel times. In one post someone mentioned having a 30-minute train commute replaced by a ninety-minute bus ride. These cases illustrate a growing gap between expected service and actual experience, which is particularly troubling given the large frequency of refund claims associated with Advance tickets in the data. In general, the sentiment analysis using Brand24 revealed considerable discontent services.

6 Knowledge Analysis

"What existing research or frameworks can help address these issues?"

This section explores literatures and best practices that correspond directly to the patterns discovered in the data and information analysis. It is organised around the disruptions mentioned throughout the report, such as delays caused by traffic, weather, and staffing issues. It presents established solutions that can inspire UK National Rail's strategy for the future.

6.1 Smart Operations & Scheduling

The diagnostic analysis uncovered long average delay times on important intercity routes such as Manchester Piccadilly to Leeds. These delays were often caused by traffic and scheduling constraints because of weather conditions. Rail technology companies such as Cyient have found solutions to identical challenges by implementing AI-powered scheduling systems that dynamically change train schedules based on real-time traffic and resource availability (Keerthi Thimmegowda, 2024). This strategy reduces traffic and train clashes, especially during peak hours, which the time-based analysis identified as an important pain point.

Furthermore, literature supports the adoption of real-time rerouting algorithms to handle unexpected disruptions. These systems allows trains to be rerouted when problems occur, such as track obstacles or delays in neighboring routes. This is an ideal solution for the operational delays seen in vulnerable routes in the dataset (Lövétei et al., 2025).

6.2 Infrastructure Reliability & Early Disruption Detection

Technical issues, were one of the most common reasons of data delays. Companies such as as VLink Inc. have successfully built predictive maintenance systems that utilise IoT sensors and machine learning to identify indicators of component failure before they impact rail service (Gopalakrishnan, 2025). This could assist prevent many of the technical delays identified in the delay severity breakdown.

Similarly, the information analysis helped us associate delays on north-west routes to Storm Henk and Storm Kathleen, which caused significant weather disruption. Tomorrow.io provides advanced weather intelligence solutions to rail operators, allowing them to prepare for and mitigate such events. Implementing this type of solution could mitigate the impact of severe storms and other weather conditions, which were identified as the most prevalent cause of delay (Favela, 2024)

6.3 Customer-Focused Service Recovery

The refund trigger analysis found that advance ticket buyers were the most inclined to seek compensation, particularly after significant delays or cancellations. Rail operators such as Network Rail have implemented automated reimbursement systems in wherein refunds are automatically granted when a service surpasses a specific delay limit (Network Rail, 2025). This method could solve refund processing delays and boost consumer trust in the event of a service disruption.

Finally, sentiment analysis based on information analysis (for instance, social media feedback via Brand24) highlighted substantial frustration among people during peak hours. According to literature, incorporating social media sentiment analysis into BI dashboards could assist rail operators evaluate commuter satisfaction in real time and respond proactively to sudden problems (Collins et al., 2013)

7 Critical Analysis & Insight Creation

7.1 Context and Scope

This project sought to uncover patterns and trends of delay across the UK National Rail system between January and April 2024 and to identify solutions that would improve service and passenger satisfaction. The analysis had three objectives: To conduct analysis of train delay trends, causes and associated ticket refunds, To explore and recommend optimized scheduling, To recommend strategies to prevent common causes of train delays.

Train delays have a substantial impact, ranging from operational costs to reputational damage and lost revenue. Because UK National Rail serves nearly a billion passenger travels each year, even minor inefficiencies can have far-reaching repercussions. Addressing this issue from a Business Intelligence (BI) perspective ensures that decisions are evidence based and outcome oriented.

The dataset used was structured and contained journey status, delay causes, ticket categories, and refund requests. While it facilitated broad analysis, it only provided a four-month sample and does not contain real-time system behavior or all external factors.

7.2 Insights Across All Analyses

The analysis led to several insights that offer opportunities when approached through a Business Intelligence lens.

- **Insight 1: High Delay Frequency in Peak Hours and Intercity Routes**
    - Delays were most frequent during morning and afternoon commuter peaks, particularly on intercity routes such as Manchester to Leeds. This reflects congestion bottlenecks that may be mitigated through dynamic scheduling. This is significant as these delays affect large number of daily commuters, compounding disruption and network strain.
- **Insight 2: Severe Delays and Refunds Correlate Strongly**
    - Refund requests were more frequent for journeys rated as "Severe" or "Cancelled." In particular, advance ticket holders had the highest refund rate. This is substantial, as The observed pattern implies that the degree of the delay has a direct impact on revenue and customer trust.

- **Insight 3: Weather and Staffing Are Top External Disruptors**
    - Storms (e.g., Henk and Kathleen) and national rail strikes coincided with increased delays. Staffing and signal difficulties were also among the most commonly reported.This matters as, although external, these factors are foreseeable and should be handled proactively through available resources and scheduling.
- **Insight 4: Public Sentiment Aligns with Delay Trends**
    - Sentiment analysis from platforms such as Reddit and TikTok displayed considerable discontent during disruptions, particularly delays that led to lengthy bus replacements. Real-time mood trends could act as early warning signs of a service breakdown.

7.3 Role of AI and Automation

Artificial intelligence (AI) and automation provide transformative possibilities for eliminating and minimising rail service disruptions. Several reoccurring issues have been identified during this project, which led to use cases in which intelligent systems can proactively manage disruptions, optimise resources, and improve customer satisfaction.

- **AI-Based Dynamic Scheduling**
    - Technical and signal failures are leading causes of disruption. Predictive maintenance systems use IoT sensors and AI models to detect early signs of faults or component wear. This allows maintenance to be conducted proactively, preventing unexpected breakdowns and improving overall service reliability.
- **Predictive Maintenance Using IoT and AI**
    - Technical and signal failures are leading causes of disruption. Predictive maintenance systems use IoT sensors and AI models to detect early signs of faults or component wear. This allows maintenance to be conducted proactively, preventing unexpected breakdowns and improving overall service reliability.
- **Automated Refund Processing Systems**
    - Manually reviewing refund eligibility is time consuming and inconsistent. AI can help by automatically triggering refunds when specific conditions are met (for instance, a delay of more than 15 minutes for Advance ticket holders). Network

Rail and other operators are already adopting this model to guarantee transparent quick and fair business.

- **Sentiment Analysis Integration**
  - Analysing social media platforms such as Reddit and TikTok in real time could reveal passenger dissatisfaction during delays or cancellations. Tools like Brand24, can detect emotional tone and complaint trends, providing early warning of service issues. When integrated to BI dashboards, this would facilitate a faster, more informed response from operators.

7.4 Limitations and Considerations

While the analysis yielded actionable insights, several limitations must be acknowledged to ensure accurate interpretation and responsible application of the findings.

7.4.1 Data Scope

The dataset covers a four-month period (January to April 2024), limiting insights into seasonal patterns or long-term trends. Insights might fail to recognize issue that occur outside of this period.

7.4.2 External Factors

Real-world events like as storms and strikes were manually discovered through additional research, although they were not formally linked to dataset records. This reduces the precision of cause-effect analysis.

7.4.3 Data Quality Assumptions

The research assumes that the reasons for delays and refund requests are regularly and accurately reported. In real-world operational contexts, manual entry or regional variations can lead to inconsistencies.

7.4.4 Insight Generalisability

Observed trends, such as refund behavior or delay clustering by time, may not apply to all routes, areas, or times. This needs to be taken into consideration before implementing recommendations on a large scale.

7.4.5 Ethical Considerations

While AI and automation increase efficiency, their implementation must address transparency and data privacy, particularly for refund decisions and sentiment monitoring based on public platforms.

8 Final Deliverables

8.1 Strategic Recommendations

Figure 14: Strategic Recommendations Infographic

8.2 BI Scorecard: Key Performance Indicators (KPIs)

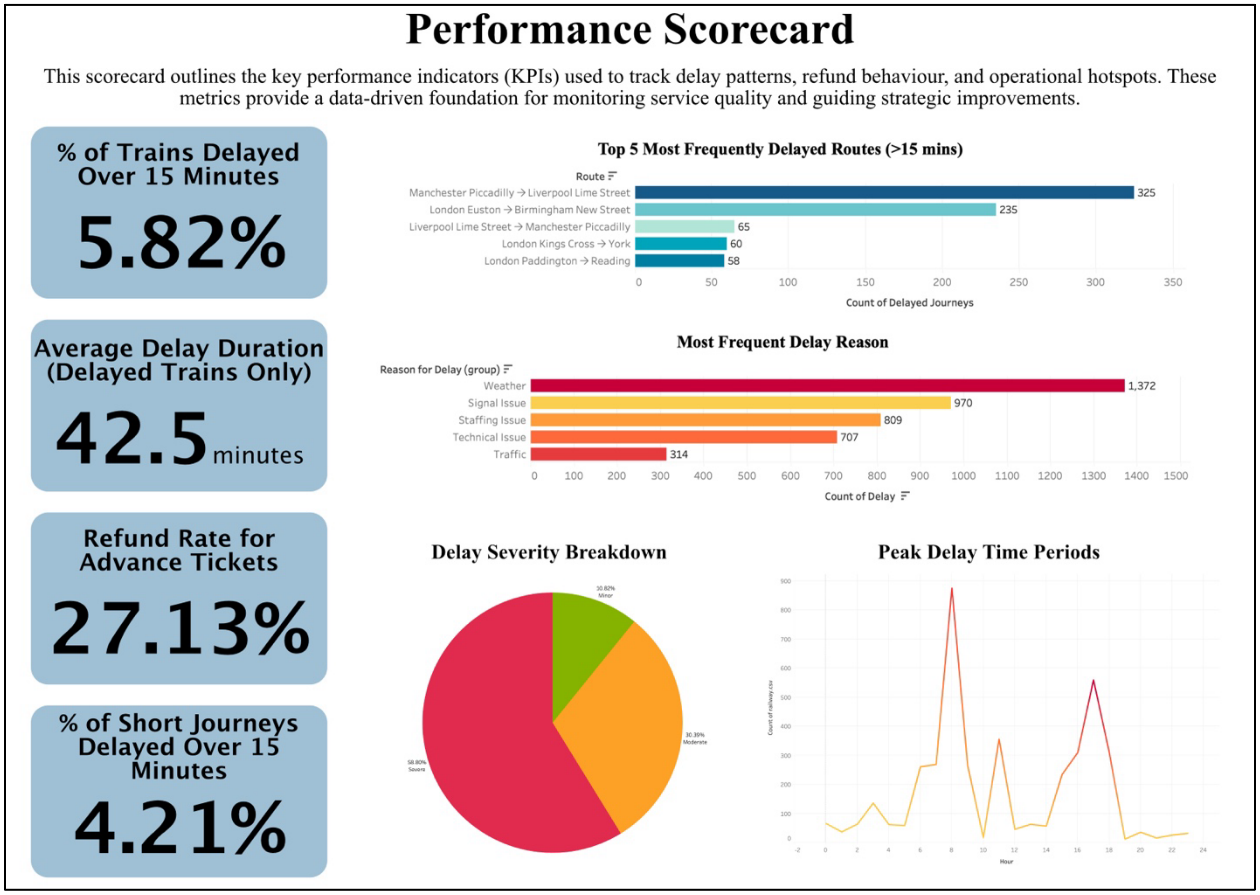


Figure 15: Performance Scorecard

Bibliograhy

#1 AI Social Listening Tool. Brand24. (n.d.). https://brand24.com/

Collins, C., Hasan, S., & Ukkusuri, S. v. (2013). A Novel Transit Rider Satisfaction Metric A Novel Transit Rider Satisfaction Metric: Rider Sentiments Measured from Online Social Media Data. https://doi.org/10.5038/2375-0901.16.2.2

Favela, R. (2024, August 9). *How the rail industry can solve for weather-related disruptions*. Tomorrow.io. https://www.tomorrow.io/blog/how-the-rail-industry-can-solve-for-weatherrelated-disruptions/

Thimmegowda, K. (2024, July 26). *Revolutionizing Railways with AI: Predictive analysis insights*. Delivering Intelligent Engineering and Technology Solutions. https://www.cyient.com/blog/revolutionizing-railways-harnessing-the-power-of-ai-for-predictive-analysis

Lövétei, I. F., Lindenmaier, L., & Aradi, S. (2025). Efficient real-time rail traffic optimization: Decomposition of rerouting, reordering, and rescheduling problems. Journal of Rail Transport Planning and Management, 33. https://doi.org/10.1016/j.jrtpm.2024.100496

Lowering the minimum age requirement for train drivers – letter to the rail sector from lord Hendy. GOV.UK. (2025, May 7). https://www.gov.uk/government/consultations/lowering-the-minimum-age-requirement-for-train-drivers-from-20-to-18/outcome/lowering-the-minimum-age-requirement-for-train-drivers-letter-to-the-rail-sector-from-lord-hendy

Meierhans, J. (2025, May 8). *Train driver minimum age to be lowered to 18*. BBC News. https://www.bbc.com/news/articles/cx2w2prj470o

O'Hare, P. (2024, April 6). *Storm Kathleen brings 90mph winds and Scottish travel disruption*. BBC News. https://www.bbc.com/news/articles/c2808nl1gd4o

Gopalakrishnan, S. (2025, January 22). *Ai in railways: Predictive analytics for maintenance*. VLink. https://vlinkinfo.com/blog/ai-in-railways/

Payments for disruption on the railway. Network Rail. (2025, February 3). https://www.networkrail.co.uk/industry-and-commercial/information-for-operators/payments-for-disruption-on-the-railway/

Rail strikes in January and February 2024. UCL News. (2024, January 22). https://www.ucl.ac.uk/news/2024/jan/rail-strikes-january-and-february-2024

Suzanne Leigh, M. J. (2024, January 3). *Storm Henk Batters UK leading to power outages, travel disruption and flooding*. BBC News. https://www.bbc.com/news/uk-67861206