



UK Train Rides Analysis Applied on

## Digital Egypt Pioneers Initiative

By:

**Omar Ali Abdelhamed**

**Ahmed Ali Mostafa**

**Basma Ahmed Saber**

**Fatma Mohamed Abu Elfadl**

**Areej Elsayed Moro**

**Shahd Saad Bedair**

A Final Project Report Submitted in Partial Fulfillment of the Requirements for the DEPI Graduation.

**(Microsoft Power BI Engineer Track)**

Executive Summary

This project focuses on analyzing train ride data across the United Kingdom to uncover key insights about train travel patterns, delays, ticket pricing, and passenger trends. By leveraging data analysis techniques, we aim to identify trends, optimize scheduling, and improve the overall efficiency of train services.

**Objectives:**

Analyze train ride frequencies and identify peak travel times.

Examine delay patterns and potential causes.

Assess ticket pricing trends based on distance, time, and demand.

Identify the busiest routes and stations.

Provide recommendations for improving train service efficiency.

**Data Sources:**

The dataset includes information on train journeys, schedules, delays, fares, and station details. It covers multiple time periods to help identify seasonal trends and variations.

**Methodology:**

Data Cleaning & Processing: Handling missing values, ensuring data consistency.

Exploratory Data Analysis (EDA): Identifying key trends and relationships.

Visualization: Using charts, maps, and heatmaps to present insights.

Statistical Analysis: Evaluating correlations between factors like delay and peak hours.

Acknowledgement

We would like to extend our deepest gratitude to Dr. Yasser Abdelrahman for their exceptional guidance, continuous support, and unwavering dedication throughout this initiative. Their insightful advice, valuable feedback, and encouragement have played a crucial role in shaping our understanding and enhancing the quality of this project. Their expertise and mentorship have been a source of inspiration, and we truly appreciate the time and effort they have invested in our learning journey.

We are also immensely grateful to Eyouth and everyone involved in its educational programs for providing us with the necessary resources, knowledge, and opportunities that allowed us to successfully complete this project. The support we received from the instructors, coordinators, and staff has been instrumental in overcoming challenges and refining our skills. Their commitment to education and student success is truly commendable, and we deeply appreciate their efforts.

Furthermore, we would like to extend our appreciation to our peers and colleagues who contributed directly or indirectly to the completion of this project. Their collaboration, discussions, and shared insights have enriched our experience and made this journey both productive and enjoyable.

Finally, we acknowledge the effort of everyone who played a role in making this project possible, whether through guidance, encouragement, or assistance. This achievement would not have been possible without the collective effort and dedication of all those who supported us along the way.

# About Data

## Data Sources

The dataset used in this analysis was sourced from Kaggle, a popular platform for open datasets and machine learning competitions.

**Dataset Name:** UK Train Rides Dataset

**Source:** Kaggle - UK Train Rides

**Description:** This dataset contains information about train schedules, delays, station names, and passenger volume across the UK railway system.

**Key Attributes:**

Transaction ID: Unique identifier for each transaction

DepartureTime & ArrivalTime: Scheduled and actual timings

Delay (minutes): Time difference between scheduled and actual arrival

DepartureDestination & ArrivalDestination: Names of the stations

Before analysis, the dataset was cleaned and preprocessed to handle missing values, duplicates, and inconsistencies.

## Data Cleaning & Preprocessing

Several data issues were identified and addressed:

**Missing Values:** Removed records with incomplete timestamps.

**Duplicates**: Eliminated duplicate train records.

**Data Formatting:** Converted timestamps to a uniform format.

**Outliers:** Removed extreme delay values exceeding 300 minutes.

# Data Analysis & Insights

1. **Operational Performance Dashboard**

Key Findings from the Dashboard:

**1. Overall Performance Metrics:**

Average Delay: 3.25 minutes.

On-Time Percentage: 86.82%.

Delayed Percentage: 7.24%.

Cancelled Percentage: 5.94%.

**2. Main Causes of Revenue Loss Due to Delays:**

Weather issues (90.31%) are the biggest contributor to revenue loss.

Technical issues (79.72%) and weather conditions (75.01%) also significantly impact revenue.

Staffing (64.11%), signal failure (45.37%), traffic (43.56%), and staff shortage (33.16%) are other major factors.

**3. Arrival Delay by Station:**

The stations with the highest delay percentages are:

London Euston (67.90%)

London Kings Cross (60.71%)

Leeds (25.10%)

Other stations have delays below 15%.

**4. Delays by Time Category:**

Morning (710 delays) and Early Morning (608 delays) have the highest number of delays.

Evening (516 delays) follows next.

Afternoon (320 delays) and Night (138 delays) have significantly fewer delays.

**5. Delays by Reason:**

Weather (758 delays) is the most common cause.

Technical issues (472 delays) and signal failure (451 delays) are other major reasons.

Staff shortage (183 delays), staffing (172 delays), weather conditions (169 delays), and traffic (87 delays) contribute to fewer delays.

6. Delay Trends Over Months (Delay LM Chart):

Delays are increasing month by month:

February: 559 delays

March: 561 delays

April: 642 delays

**Main Insights:**

Weather-related issues are the most significant cause of both delays and revenue loss.

Morning and early morning periods experience the highest number of delays.

Delays are increasing month by month, which may indicate worsening operational issues.

London Euston and Kings Cross suffer the most from delays, which may require targeted interventions.

1. **Overview Analysis Dashboard**

Key Findings from the Dashboard:

**1. General Performance Metrics:**

Total Riders: 32K

Net Revenue: 742K

Total Revenue: 696K

Departure Stations: 12

Arrival Stations: 32

**2. Route Distribution by Riders:**

The top three busiest routes are:

Manchester Piccadilly → Liverpool Lime Street (15%)

London Euston → Birmingham New Street (13%)

London Kings Cross → York (12%)

Other significant routes include:

London Paddington → Reading (12%)

London St Pancras → Birmingham New Street (11%)

Liverpool Lime Street → Manchester Piccadilly (9%)

Less popular routes (each 2%) include:

London Euston → Manchester Piccadilly

Birmingham New Street → London St Pancras

London Paddington → Oxford

**3. Passenger Railcard Usage:**

66.09% of passengers do not use a railcard.

15.31% use an adult railcard.

8.85% use a Senior railcard.

Only 0.10% use a Disabled railcard.

**4. Passenger Ticket Type Distribution:**

Advance Tickets: 5K (16.55%)

Off-Peak Tickets: 9K (27.65%)

Anytime Tickets: 18K (55.48%) → This is the most preferred ticket type.

**5. Ticket Class Transactions:**

Standard Class: 29K transactions (90.3%) dominate ticket purchases.

First Class: 3K transactions (9.66%) indicate lower usage of premium services.

**6. Peak Travel Times:**

Travel demand fluctuates throughout the day, with the busiest times occurring around 6:00 PM - 7:00 PM.

Peak times:

6:00 PM: 2.5K passengers (highest)

3:00 PM - 5:00 PM: Increased activity with spikes at 1.3K and 1.7K.

Early morning (12 AM - 6 AM): Low passenger volume, averaging around 0.4K - 1.4K.

**7. Top Departure Stations:**

Manchester Piccadilly: 5.7K departures (highest)

London Paddington: 4.5K

London Kings Cross: 4.2K

London Euston: 5.0K

Liverpool Lime Street: 4.6K

**Main Insights:**

Busiest Routes: Manchester Piccadilly → Liverpool Lime Street and London routes have the highest ridership.

Revenue & Ticket Sales: Most passengers opt for standard class and anytime tickets.

Peak Travel Times: Highest traffic is in the evening (6:00 PM), likely due to work commuters.

Railcard Usage: Majority (66%) do not use a railcard, which could indicate potential for marketing incentives to encourage adoption.

1. **Customer Satisfaction & Refunds Dashboard**

Key Findings from the dashboard:

**1. Overall Refund Requests:**

96.47% of refund requests are denied, while only 3.53% are approved.

There are 31,000 denied refund requests and 1,118 approved refund requests.

**2. Refund Requests by Journey Status:**

30.43% of refund requests are approved for \*Cancelled\* journeys.

23.82% of refund requests are approved for \*Delayed\* journeys.

**3. Refund Requests by Reason for Delay:**

Staffing issues have the highest approval rate for refund requests at 44.39%.

Traffic issues follow with a 38.54% approval rate.

Technical Issues have the lowest approval rate at 5.88%.

**4. Refund Requests by Price:**

Refund requests are more likely to be approved for higher-priced tickets, with the highest approval rate at 90.

**5. Refund Requests by Ticket Type:**

Anytime tickets have the highest approval rate for refund requests at 4.22%.

Advance tickets have a lower approval rate at 2.58%.

**6. Refund Requests by Payment Method:**

Contactless payment methods have the highest approval rate for refund requests at 29.95%.

Debit Card and Credit Card payments have lower approval rates at 2.11% and 1.95%, respectively.

**7. Refund Requests by Railroad:**

Senior railcard holders have the highest approval rate for refund requests at 8.71%.

Adult and None disabled categories have lower approval rates at 2.47% and 1.94%, respectively.

**Main Insights:**

High Denial Rate: The overwhelming majority of refund requests are denied, indicating a strict refund policy or potential issues in the refund approval process.

Journey Status Impact: Refund requests for Cancelled and Delayed journeys have higher approval rates, suggesting that these statuses are more likely to result in refunds.

Reason for Delay: Refund requests due to Staffing and Traffic issues are more likely to be approved, while Technical Issues have a significantly lower approval rate.

Ticket Price Influence: Higher-priced tickets have a higher likelihood of refund approval, possibly due to the greater financial impact on customers.

Ticket Type Variability: Anytime tickets have a higher refund approval rate compared to Advance tickets, which may reflect differences in terms and conditions.

Payment Method Preference: Contactless payments have a notably higher refund approval rate, possibly due to ease of processing or customer trust in this method.

Railcard Holder Benefits: Senior railcard holders experience higher refund approval rates, indicating potential preferential treatment or specific policies for this group.

1. **Customer Behavior Dashboard**

Key Findings from the dashboard:

**1. Average Duration:**

The average travel duration is 74 minutes.

**2. Ticket Type Distribution:**

55% of tickets are Advanced tickets.

27.65% of tickets are Off-Peak tickets.

16.87% of tickets are Anytime tickets.

**3. Top Transactions by Travel Duration:**

The longest route is Liverpool Lime Street to London Exston at 163 minutes with 1,097 transactions.

The shortest route is Liverpool Lime Street to Manchester Piccadilly at 30 minutes with 3,002 transactions.

**4. Total Tickets per Time Category:**

The highest number of tickets is during the Nighttime category with 7.7K tickets.

The lowest number of tickets is during the Afternoon time category with 5.1K tickets.

**5. Passenger Ticket Class:**

93.34% of passengers travel in Standard class (29K).

9.66% of passengers travel in First Class (3K).

**6. Passenger Payment Method:**

60.46% of payments are made using Debit Cards (19K).

24.23% of payments are made using Contactless methods (11K).

5.32% of payments are made using Credit Cards (2K).

**7. Railcard Usage:**

66.09% of passengers do not use a railcard (21K).

9.76% of passengers use an adult railcard (3K).

8.85% of passengers use a Senior railcard (3K).

**Main Insights:**

Ticket Type Preference: The majority of passengers prefer Advanced tickets, likely due to cost savings, followed by Off-Peak and Anytime tickets.

Travel Duration: The most frequent and longest journeys are between major cities like Liverpool, London, and Manchester, indicating high demand for these routes.

Time Category: The Night time category has the highest ticket sales, suggesting that many passengers travel during late hours, possibly for work or leisure.

Ticket Class: The vast majority of passengers opt for the Standard class, with a small percentage choosing First Class, reflecting a preference for more economical travel options.

Payment Methods: Debit Cards are the most popular payment method, followed by Contactless payments, indicating a shift towards digital and card-based transactions.

# Recommendations

**1. Reduce Delays & Improve Punctuality**

- Address Weather-Related Delays (Main Cause - 90.31%)

Invest in better weather forecasting systems to prepare for disruptions in advance.

Implement real-time monitoring of weather conditions to adjust schedules dynamically.

Improve track drainage and heating systems to prevent weather-related disruptions (e.g., ice, floods).

- Improve Maintenance & Technical Issue Handling (79.72%)

Implement predictive maintenance using AI to detect and fix technical issues before they cause delays.

Increase on-site technical staff in high-delay stations (London Euston, Manchester Piccadilly).

Upgrade train control systems to reduce failures.

- Optimize Staffing to Prevent Shortages (33.16%)

Introduce flexible shift planning based on demand trends.

Use AI-based scheduling tools to ensure enough staff are available during peak times.

**2. Improve Passenger Experience & Revenue**

- Encourage Railcard Usage (Currently Only 33.91% Use It)

Launch promotional discounts for railcard users to increase adoption.

Improve awareness by targeted marketing campaigns for frequent travelers.

-Optimized Ticket Pricing & Promotions

Introduce dynamic pricing based on peak/off-peak hours to balance demand.

Offer bundle discounts for frequent travelers (e.g., weekly or monthly passes).

-Improve First-Class Experience (Only 9.66% Use First-Class Tickets)

Enhance services (Wi-Fi, snacks, priority boarding) to attract more premium passengers.

Introduce corporate partnerships to encourage business travelers to choose First Class.

**3. Manage Peak Times More Effectively**

- Reduce Overcrowding in the Evening Peak (6:00 PM - 7:00 PM is the busiest)

Increase train frequency during peak hours to reduce congestion.

Introduce discounts for off-peak travel to shift some demand to less crowded times.

Use real-time crowd monitoring and notify passengers about alternative routes.

- Early Morning & Late Night Optimization

Since early morning (12 AM - 6 AM) has low demand, consider reducing service frequency to save costs.

Introduce "Night Rider" discounts to encourage late-night travelers.

**4. Improve Performance at High-Delay Stations**

-Focus on London Euston & Manchester Piccadilly

Since these stations have the most delays, we have implemented faster turnaround procedures for trains.

Improve the platform and signaling efficiency to reduce waiting times.

-Station-Specific Delay Reduction Plans

Analyze why London Kings Cross and Leeds experience frequent delays and implement station-specific solutions (e.g., track maintenance, staff availability).

**5. Customer Experience Improvements:**

Journey Status Updates: Provide real-time updates on journey statuses (delays, cancellations) to keep passengers informed and reduce frustration.

Feedback Mechanism: Implement a robust feedback mechanism to gather customer insights and continuously improve service quality.

**6. Operational Efficiency:**

Staff Training: Invest in training staff to handle refund requests and customer inquiries more effectively, especially for high-impact issues like staffing and traffic delays.

Infrastructure Upgrades: Consider infrastructure upgrades to mitigate common issues such as signal failures and technical problems.

**7. Data-Driven Decisions:**

Regular Analysis: Continuously analyze customer behavior and refund request data to identify trends and areas for improvement.

Personalized Offers: Use data insights to offer personalized travel options and discounts based on customer preferences and travel patterns.

# Conclusion

This report provides a comprehensive analysis of the UK train operational performance and passenger trends, highlighting key insights into delays, revenue, ticketing, and travel patterns.

The analysis shows that while the railway system serves 32K passengers across 12 departure stations and 32 arrival stations, operational inefficiencies, weather conditions, and technical issues contribute to delays and revenue losses. The average delay of 3.25 minutes, with only 86.82% on-time performance, indicates room for improvement in scheduling, maintenance, and traffic management.

Passenger behavior reveals a strong preference for Standard Class (90.3%) and Anytime tickets (55.48%), while railcard adoption remains low (only 33.91%), presenting an opportunity for marketing and pricing strategies to encourage usage. The evening peak (6:00 PM - 7:00 PM) sees the highest congestion, requiring better demand management strategies.

To enhance efficiency and customer experience, strategic recommendations include:

Reducing delays through predictive maintenance, staff optimization, and traffic control improvements.

Optimizing revenue by promoting railcards, adjusting ticket pricing, and enhancing First-Class services.

Managing peak-hour congestion by increasing train frequency and offering off-peak incentives.

Targeting high-delay stations (London Euston, Manchester Piccadilly) with customized solutions.

By implementing these recommendations, the railway system can improve reliability, reduce financial losses, enhance customer satisfaction, and maximize operational efficiency.