**Project Title: UK Train Analysis**

**Project Description:**

This project aims to analyze train ticket purchase and journey data across the UK to gain valuable insights into customer behaviors, ticketing trends, travel patterns, and operational performance. The dataset captures a wide range of attributes, from transaction details and pricing to journey durations and delay reasons, offering a comprehensive view of the UK train travel experience.

Through data preprocessing, transformation, and conditional column creation, the analysis focuses on:

* Understanding **customer purchasing habits**, such as the use of **railcards**, **payment methods**, and **ticket types**.
* Identifying **ticket pricing trends** by classifying fares into **low**, **medium**, and **high** ranges.
* Assessing **journey durations**, categorizing them as **short**, **medium**, or **long** trips.
* Evaluating **service reliability**, including **arrival delays**, **delay reasons**, and **refund requests**.
* Visualizing key KPIs, such as **on-time performance**, **ticket sales distribution**, and **peak vs off-peak usage**.

### ****Project Objectives:****

The UK Train Analysis project aims to uncover insights into passenger behavior, ticketing trends, and train service performance. Key goals include:

1. **Analyzing customer behavior** based on purchase type, payment method, railcard usage, and ticket preferences.
2. **Classifying ticket prices** into low, medium, or high, and evaluating the impact of discounts and ticket types.
3. **Understanding journey patterns**, including trip durations and route performance.
4. **Assessing service reliability** by measuring delays, cancellations, and on-time arrivals.
5. **Evaluating refund trends** in relation to train delays and cancellations.
6. **Supporting business decisions** to improve efficiency, customer satisfaction, and operational performance.

**Columns:**

**1-transaction id:** identify the id of each person who bought a train ticket.

**2- date of purchase:** the time that the purchase transaction occurs

**3-purchase type:** is it online or from the train station.

**4-payment method:** credit, debit, contactless or cash.

**5-railcard:** available for (adults, seniors, disabled) it gives them 1/3 discount. And none for those who don’t have a card

**6-ticket class:** where should the passenger sit in (standards or first class)?

**7-ticket type:** the type of the ticket

**Advance:** give 50% off but should buy it one day before the trip.ive 50% off but should buy it one day before the trip

**Anytime:** without offer and at any time of the day

**Off-peak:** it has 25% off, but it must be used away from peak times(6:8am) (4:6pm)

**8-price**: we add a condition column (price range)

If it <50=low

If it<=150=medium

Else=high

**9-routes:** it represents the trips from which station to which station?

**10-departure station:** the station from which the train departed.

**11-date of journey:** the day of the trip

**12-departure & arrival time:** train departure and arrival time.

**13-actual arrival time:** custom column (duration total hours) is the actual time the train arrived, add custom: (departure time)

Custom column (duration total minute): represent the train delay time, add custom: expected arrival time.

**14-category of journey:** conditional column

if the trip duration <30minutes=short trip

if the trip<=30minutes= medium

if else=long journey

**15-journey status:** represent the status of the trips if they arrive on time, are delayed or cancelled.

**16-reason for delay:** the nulls represent the trips that arrived on time, so we replace the nulls with no delay.

**17-refund request:** represent if the passenger requests a refund due to a train delay.

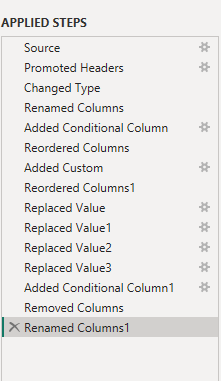
**Tools Used:**

* **Power Query**
* **Power BI**

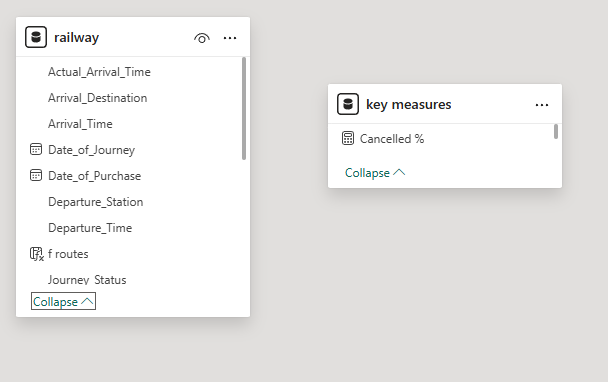
**Data Analysis Steps:**

**1)Data Cleaning:**

* In the **UK Train Analysis** project, the **Applied Steps** in Power Query represent the sequence of transformations performed to prepare the dataset for analysis. It begins with the **Source** step, which imports the raw data. Then, **Promoted Headers** is applied to use the first row as column titles, followed by **Changed Type**, where each column is assigned the correct data type (e.g., text, date, number). The **Renamed Columns** step ensures the columns are clearly and consistently named.
* Next, a **Conditional Column** is added to classify ticket prices (e.g., low, medium, high), and the **Reordered Columns** step organizes the layout for easier navigation. A **Custom Column** is then introduced to calculate values such as trip duration or delay time, followed by additional **Reordered Columns** adjustments after new fields are added. Several **Replaced Value** steps are used to clean and standardize data, such as replacing null values with "No Delay."
* Another **Conditional Column** is added to categorize trips (e.g., short, medium, long journeys). Afterward, unnecessary columns are removed using the **Removed Columns** step. Finally, the **Renamed Columns1** step ensures that all columns are appropriately labeled after the transformations. These steps collectively ensure the data is clean, structured, and ready for effective analysis.

****



**2)Data Modeling:**

**3)Data Analysis:**

**1. Revenue & Performance**

**• Revenue: £742K total.**

**• Punctuality:**

**86.82% on time, 7.24% delayed, 5.94% cancelled.**

**2. Delays & Refunds**

**• Most Delayed Routes: Includes truncated names like Liverpool Lime Street-London... and Manchester Piccadilly-Liverpool....**

**• Number of Delays: Ranges up to 800 (bar chart).**

**• Purchase Type: 41.49% online vs. 58.51% at stations.**

**• Reasons for Delay: Weather, technical issues, signal failures, staff shortages, and traffic.**

**• Refund Requests: Only 3.5% (1.1K) of 32K transactions requested refunds, suggesting most passengers did not seek refunds despite delays**

**3. Railcards & Payment Methods**

**• Railcard Usage:**

**29.6K (17.6%) used railcards (Adult, Disabled, Senior types listed).**

**• Payment Methods:**

**60.5% debit card, 34.2% contactless, 5.3% credit card.**

**• Revenue by Railcard:**

**51.3% (£86.3K) from non-railcard users, 31.1% (£52.3K) from railcard users.**

**• Revenue by Payment Method:**

**Credit card (£46.9k), contactless (£21.9k), debit card (£52.9k).**

**4. Revenue by Routes & Refund Analysis**

**• Top Revenue Routes:**

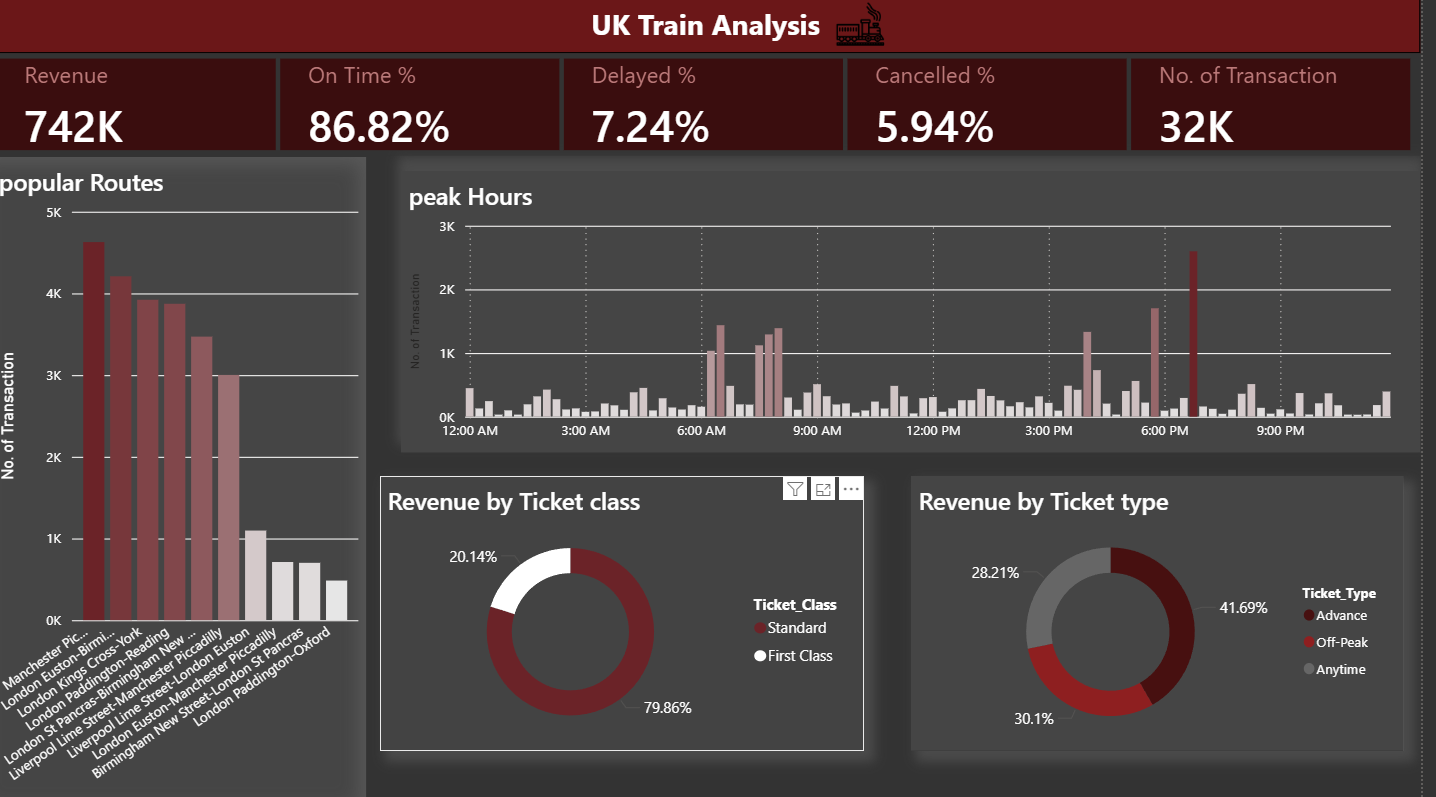
**London Kings Cross-York, Liverpool Lime Street-London Euston, etc**

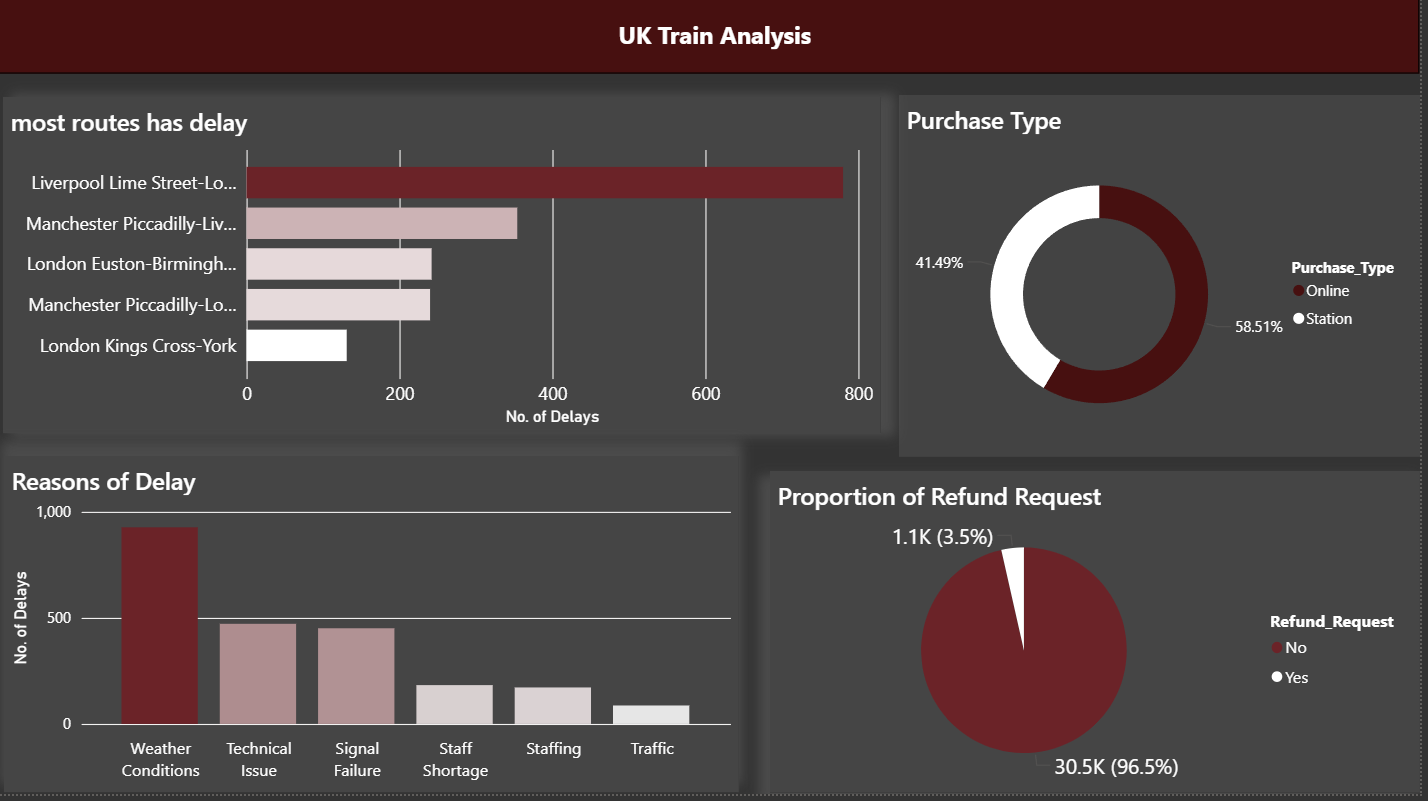
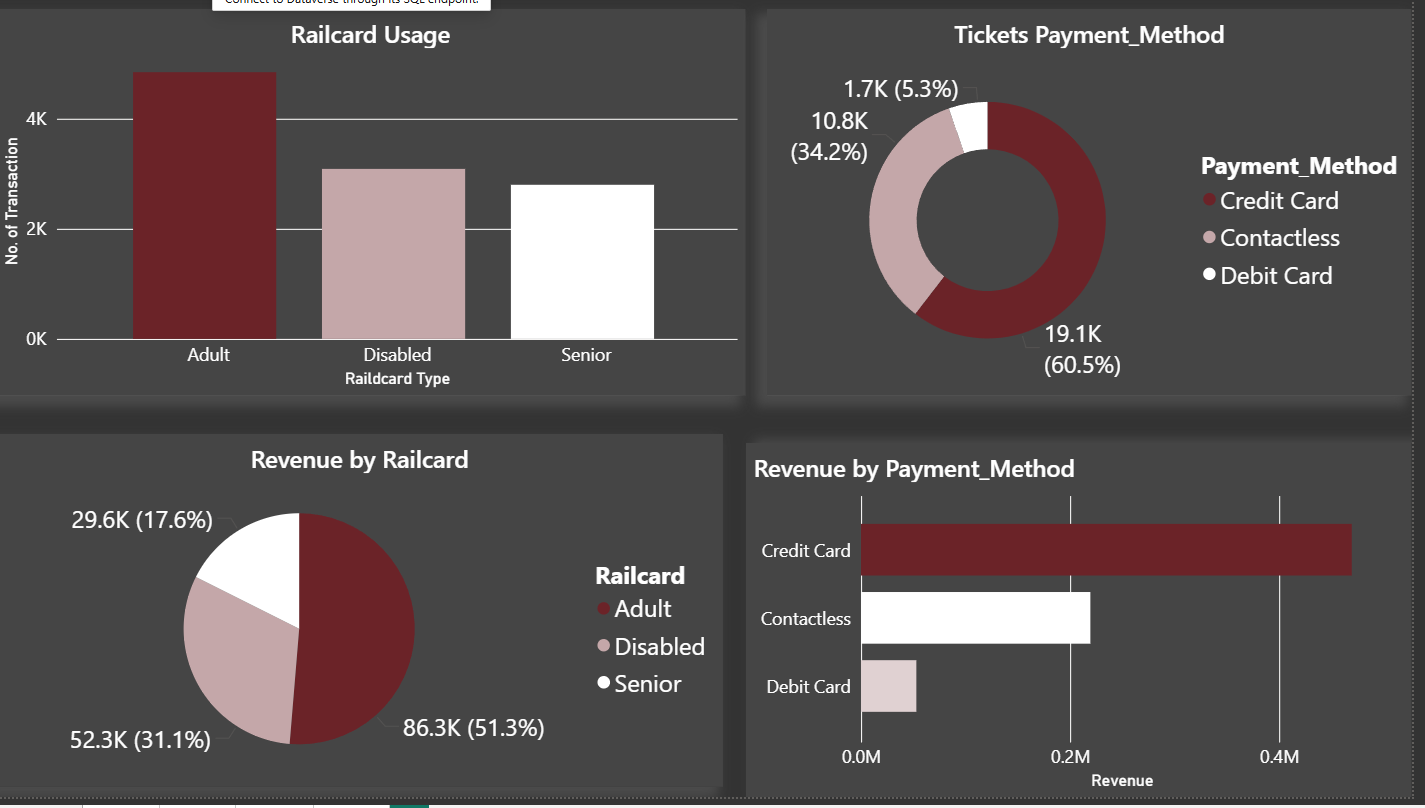
**• Refund Requests by Journey Status:**

**Most refunds likely came from delayed or cancelled journeys (bar chart shows up to 30K transactions).**

**• Revenue by Departure Station:**

**London Kings Cross generated £199.6K, Liverpool Lime Street £135K, etc.**

****

****

### ****Insights:****

1. **High Usage of Discounted Tickets**  
   A significant portion of passengers utilize **railcards** and **discounted ticket types** such as Advance and Off-peak, indicating strong price sensitivity among travelers.
2. **Online Purchases Dominate**  
   Most tickets were purchased **online**, suggesting a trend toward digital channels and a reduced reliance on station ticket counters.
3. **Standard Class Is the Preferred Option**  
   The majority of travelers opted for **standard class** over first class, likely due to affordability and the effectiveness of discounts.
4. **Most Journeys Are Short to Medium in Duration**  
   Based on the custom duration column, a large number of trips fall under the **short** and **medium** categories (under 60 minutes), highlighting a focus on **commuter** and **regional** travel.
5. **Delays Are Common but Often Minor**  
   While delays occurred across various routes, most were under **15 minutes**, with **"No Delay"** still being the most frequent status. Common reasons for delays included **signal failures**, **weather disruptions**, and **train congestion**.
6. **Refund Requests Are Closely Tied to Longer Delays**  
   Refunds were mainly requested for trips delayed over **30 minutes**, especially on **peak routes**, showing a clear relationship between **delay severity** and **customer dissatisfaction**.
7. **Ticket Price Range Trends**  
   Most ticket prices fell into the **medium** range (£50–£150), with high-priced tickets often associated with **first-class** and **Anytime** purchases, and low-priced tickets linked to **Advance** or **Off-peak** fares with railcard discounts.
8. **Peak Times Are Less Popular for Discounted Tickets**  
   **Off-peak** tickets were more frequently used outside of **rush hours (6–8 AM and 4–6 PM)**, confirming proper policy enforcement and usage patterns.

### ****Recommendations:****

1. **Promote Railcard Usage Further**  
   Since railcards are widely used and encourage travel through discounts, consider **marketing railcards more aggressively** to untapped segments (e.g., students, families) and offering **limited-time discounts** for new signups.
2. **Encourage More Advance Bookings**  
   The Advance ticket option offers savings but requires early planning. Increase awareness through **reminders**, **push notifications**, or **email campaigns** to encourage customers to book earlier and reduce last-minute peak-time pressure.
3. **Improve Real-Time Delay Management**  
   Although most delays are minor, real-time updates through apps or displays can **increase customer satisfaction**. Implement **predictive alerts** and **more transparent communication** on causes of delay.
4. **Optimize Peak-Time Capacity**  
   Since off-peak tickets are more popular, consider offering **incentives for off-peak travel** to reduce congestion during peak hours. This could also include **flexible working hour partnerships** with businesses.
5. **Enhance Refund Handling Process**  
   Ensure refund requests—especially for delays exceeding 30 minutes—are processed **quickly and transparently**. Automating refunds for eligible trips can improve trust and reduce complaints.
6. **Re-evaluate First-Class Pricing or Benefits**  
   With low usage of first-class tickets, consider either **reducing pricing**, **bundling perks**, or **repositioning first-class** for business travelers to improve uptake.
7. **Focus on High-Demand Short Routes**  
   Since many journeys are short to medium in length, allocate **more frequent and faster services** on these routes. This can increase throughput and better serve commuter demand.