UK Train Rides

Project by:
"Insight Innovators"
YAT422B_GIZ2_DAT1_G3

1. Sarah Farag
2. Sahar Hamdi
3. Nada Saad
4. Mona Hassan
5. Sara Mohamed
6. Dina Yasser

UK Train Rides Dataset

Introduction:

This project aims to analyze railway journey data from January 1st, 2024, to April 30th, 2024. The dataset includes crucial information regarding railway card usage, journey status, delay reasons, and other relevant details. By comprehensively examining this data, we aim to uncover significant insights and patterns to improve railway operations and enhance passenger experience.

Business Outcomes:

- Improve operational efficiency by identifying delay patterns.

- Enhance passenger satisfaction through better resource allocation and scheduling.

- Increase revenue by understanding booking trends and passenger behavior.

Problem Statement:

The dataset reveals significant issues, such as high numbers of delays and missing railway card data. These issues impact operational efficiency and passenger experience.

Pain Points:

- Frequent delays causing passenger dissatisfaction.

- Incomplete data affecting the accuracy of analysis.

- Lack of insights into the reasons for delays and their impact on revenue.

Objectives:

1. ***Data Cleaning and Preparation:***

   - The dataset contains over 20,000 null values in the "Railway Card" column. These null values will be replaced with "No-Railcard" to ensure completeness.

   - All null values in "Reason for Delay" will be replaced with "Arrived"

   - All null values in "Actual Arrival Time" will be replaced with " ".

   - All necessary columns with object data types will be converted to appropriate date/time formats for accurate analysis.

2. *Standardization:*

   o The values in the "Reason for Delay" column will be unified into specific categories:

      1. "Weather" will be "Weather Conditions,"

      2. "Signal failure" will be "Signal Failure,"

      3. and "Staffing" will be "Staff Shortage."

3. *Research and Analysis:*

   o Conduct research on various weather conditions that affected railway journeys during the specified period.

   o Analyze ticket types to study the behavior of passengers, particularly focusing on the demand for refunds and compensations in cases of cancellations or delays.

   o Investigating The Difference between the Online Purchasing and Station Purchasing to Measure the Efficiency of Both methods which can guide us which is the best way for purchasing and how to enhance it.

By achieving these objectives, we aim to gain a comprehensive understanding of the factors influencing railway journeys and delays. The insights derived from this analysis will be instrumental in identifying areas for improvement and implementing strategies to enhance the overall efficiency and passenger satisfaction in railway services.

Roles and Responsibilities with Timelines

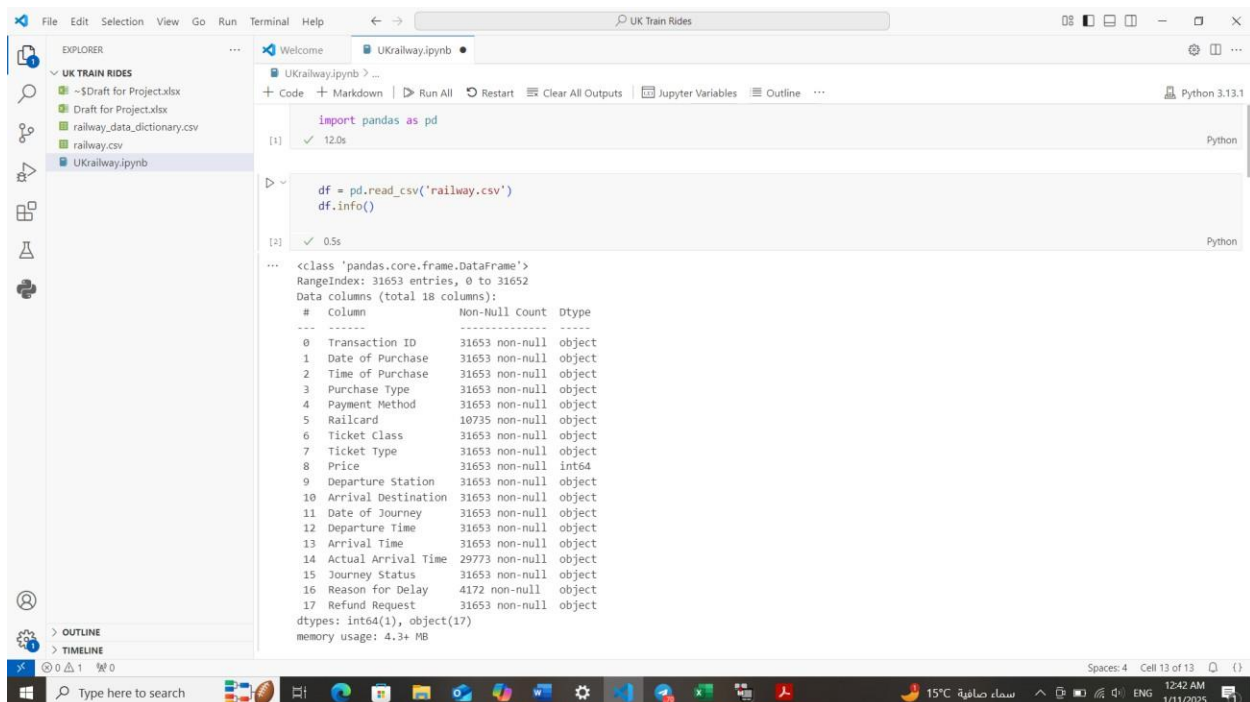| Phases	Milestones	Tools	Timelines	Deliverables	Assigned to
Data Exploration and Preprocessing	Examining the data to understand its structure, main characteristics, and discover patterns or relationships.	MS Excel Python (pandas, Matplotlib)	2 weeks	Project Documentation	The Whole Group
Data Cleaning and Preparation	Perform data cleaning, transformation, and analysis.	MS Excel Python (pandas, Matplotlib)	2 weeks	1. Cleaned dataset ready for analysis 2. Data preprocessing notebook.	Sarah Farag Sahar Hamdi Nada Saad
Data Analysis and Insights Generation	Provide insights into railway operations and help interpret analysis results.	MS Excel Python (pandas, Matplotlib)	2 weeks	Set of analysis questions that can be answered via the dataset	Mona Hassan Sara Mohamed Dina Yasser
Visualization Dashboard and Final Presentation	Build a Tableau visualization dashboard that visualize the answers to all answered questions.	Tableau	2 weeks	1. Visualization dashboard 2. Final report and presentation.	The Whole Group

1st Phase: Data Exploration and Preprocessing

Data Exploration with Python: Unveiling Insights from UK Train Rides

The efficient operation and punctuality of train services play a crucial role in ensuring the seamless transportation of passengers across the UK. By using the powerful capabilities of Python, we can delve into the vast datasets related to UK train rides, uncovering valuable insights that inform decisions, improve services, and enhance customer satisfaction. Python, with its libraries such as Pandas, NumPy, and Matplotlib, serves as a versatile and efficient tool for data exploration and analysis. These libraries enable us to clean, manipulate, and visualize data effectively, transforming raw datasets into meaningful information.

In this exploration, we aim to analyze various aspects of UK train rides, including purchase patterns, ticket pricing, punctuality, and delays. By examining the data, we can identify trends, uncover anomalies, and derive actionable insights that can help optimize train services and improve the overall passenger experience.

Our data exploration journey begins with data cleaning and preprocessing, where we handle missing values, standardize formats, and ensure data consistency. Next, we dive into descriptive statistics and visualizations, providing a comprehensive overview of the dataset. Through this process, we aim to answer key questions about purchase methods, ticket types, station activities, and delay reasons. By harnessing the power of Python, we embark on a systematic and insightful exploration of UK train rides, paving the way for data-driven decision-making and enhanced operational efficiency.

```python
df.isnull().sum()
```

```
Transaction ID          0
Date of Purchase        0
Time of Purchase        0
Purchase Type           0
Payment Method          0
Railcard            20918
Ticket Class            0
Ticket Type             0
Price                   0
Departure Station       0
Arrival Destination     0
Date of Journey         0
Departure Time          0
Arrival Time            0
Actual Arrival Time  1880
Journey Status          0
Reason for Delay    27481
Refund Request          0
dtype: int64
```

```python
df['Journey Status'].value_counts()
```

```
Journey Status
On Time      27481
Delayed       2292
Cancelled     1880
Name: count, dtype: int64
```



```
Reason for Delay    0
Refund Request      0
dtype: int64
```

```python
df['Journey Status'].value_counts()
```

```
Journey Status
On Time      27481
Delayed       2292
Cancelled     1880
Name: count, dtype: int64
```

```python
df['Railcard'].value_counts()
```

```
Railcard
Adult      4846
Disabled   3089
Senior     2800
Name: count, dtype: int64
```

```python
df['Ticket Type'].value_counts()
```

```
Ticket Type
Advance    17561
Off-Peak    8752
Anytime     5340
Name: count, dtype: int64
```

```python
df['Departure Station'].value_counts()
```

```
Departure Station
Manchester Piccadilly    5650
London Euston            4954
Liverpool Lime Street    4561
London Paddington        4500
London Kings Cross       4229
London St Pancras        3891
Birmingham New Street    2136
York                      927
Reading                   594
Oxford                    144
Edinburgh Waverley         51
Bristol Temple Meads       16
Name: count, dtype: int64
```

```python
df['Arrival Destination'].value_counts()
```

```
Arrival Destination
Birmingham New Street    7742
Liverpool Lime Street    5022
York                     4019
Manchester Piccadilly    3968
Reading                  3920
London Euston            1567
London St Pancras         749
Oxford                    623
London Paddington         351
Leicester                 337
Sheffield                 272
```

```python
df['Reason for Delay'].value_counts()
```

```
Reason for Delay
No Delay           27481
Weather              995
Technical Issue      707
Signal Failure       523
Signal failure       447
Staffing             410
Staff Shortage       399
Weather Conditions   377
Traffic              314
Name: count, dtype: int64
```

2nd Phase: Data Cleaning and Preparation

Assigned to (Sarah Farag – Sahar Hamdi – Nada Saad)

After conducting Data exploration using python to see the data in a collective perspective, we've found out that certain fields are containing the most missing values. The decision was to use Python as the first tool to clean the data because Python's efficient data structures and optimized libraries enable us to process this large dataset quickly and effectively. This makes data cleaning more manageable, even when dealing with massive amounts of data.

Also, the date/time format needed to be adjusted for the dataset's usability, enabling more accurate and meaningful data analysis and visualization. We decided on using Power Query as it provides an intuitive, user-friendly interface that allows us to manage and format date/time data without requiring advanced coding skills. Power Query seamlessly integrates with Excel, allowing us to leverage Excel's powerful data analysis and visualization tools. Using Power Query ensures that date/time formatting is consistent across the dataset. This is crucial for accurate data analysis, as inconsistent formats can lead to errors and misinterpretations.

Data Cleaning and Handle Missing Data: Identify and handle missing values, duplicates, and erroneous or outlier data points.

Replacing "Null" values in the "Railcard" column with "No-Railcard" provides a couple of key advantages:

1. **Clarity**: When data analysts and other stakeholders review the dataset, they can clearly see that the value is unknown rather than being unsure if it's a mistake or a genuine missing value.
2. **Consistency**: It ensures that all entries in the "Railcard" column have a value, making data handling and processing more straightforward. This also helps maintain data integrity and reduces the risk of errors during analysis.
3. **Querying**: When querying the dataset, it becomes easier to filter and analyze data. Instead of checking for "Null" values, which can be interpreted differently by various systems or software, users can simply filter for "Unknown".
4. **Documentation**: It provides a clear indication that the data point was either not captured or not applicable, which can be helpful for documenting the dataset and understanding its limitations.

The decision to set the "Actual Arrival Time" column entries to a unique time of "23:59:59" was made to address the issue of missing values in a clear and consistent manner. This approach offers several key benefits:

1. **Distinct Placeholder**: The time "23:59:59" was chosen as it is highly unlikely to be a valid arrival time, thereby clearly indicating that the actual arrival time is not available or was not recorded.

2. **Data Integrity**: By using a consistent placeholder, we ensure that the dataset remains structurally sound and free of null values, which can lead to complications in data processing and analysis.

3. **Ease of Identification**: The unique time "23:59:59" allows for easy identification and filtering of missing or placeholder entries during data analysis. This makes it straightforward to isolate these entries for further investigation or handling.

4. **Standardization**: This method ensures that all missing actual arrival times are uniformly represented, facilitating more accurate and reliable data analysis and reporting.

The decision to replace the "Reason for Delay" column entries with "Arrived" was made to address the issue of missing values and ensure data consistency. This approach offers several key benefits:

1. **Clarity**: By replacing missing or null entries with "No Delay," we provide a clear indication that there was no recorded reason for a delay, making it easier for analysts to understand the dataset.

2. **Data Integrity**: Ensuring that every entry in the "Reason for Delay" column has a value maintains the structural integrity of the dataset. This helps prevent issues that can arise from null values during data processing and analysis.

3. **Consistency**: Using a consistent placeholder like "No Delay" helps standardize the data, making it easier to analyze and interpret. This uniformity aids in generating accurate and reliable reports.

4. **Ease of Analysis**: The "No Delay" entry allows for straightforward filtering and querying of the dataset, enabling analysts to quickly identify and separate entries with no recorded delays from those with specific reasons for delays.

By adopting this approach, we maintain the integrity and usability of the dataset, enabling more efficient and effective data analysis.

Transform Data: Ensure that our data is in the right format (e.g., date formats, numerical values, categorical encoding).

This decision was driven by several key factors:

1. **Accuracy**: Converting these columns to the "Date/Time" format ensures that all date and time values are accurately recognized and processed. This helps to eliminate errors that can occur when dates and times are stored as text or other formats.

2. **Consistency**: Standardizing the data types across these columns provides consistency in how dates and times are represented. This uniformity is essential for accurate comparisons, calculations, and aggregations.

3. **Enhanced Analysis**: With the data in the correct "Date/Time" format, it becomes easier to perform time-based analyses, such as calculating durations, identifying trends, and generating time series visualizations.

4. **Improved Visualization**: Many data visualization tools and software rely on proper "Date/Time" formats to create accurate and meaningful visual representations. By ensuring that the data is in the correct format, we enable more effective and insightful visualizations.

5. **Data Integrity**: Adjusting the data types helps maintain the integrity of the dataset by ensuring that dates and times are stored in a standardized and reliable manner. This reduces the risk of data corruption and enhances the overall quality of the dataset.

Through these adjustments, we enhance the dataset's usability, enabling more accurate and meaningful data analysis and visualization.

In order to effectively manage and analyze the dataset, we employed a combination of Python and Power Query with Excel to address several key data quality issues and standardization needs.
Using Python:
- To handle missing values, we substituted "Null" entries in specific columns with predefined values:
  - The "Railcard" column entries were replaced with "No-Railcard".
  - The "Actual Arrival Time" column entries were set to a unique time " ".
  - The "Reason for Delay" column entries were replaced with "Arrived".
- To standardize the values in the "Reason for Delay" column, we unified various entries to ensure consistency:
  - Entries like "Weather" and "Weather Conditions" were merged under "Weather Conditions".
  - Entries like "Signal Failure" and "Signal failure" were consolidated under "Signal Failure".
  - Entries like "Staffing" and "Staff Shortage" were unified under "Staff Shortage".

Using Power Query and Excel:
- To ensure proper data analysis and visualization, we adjusted the data types of the following columns to "Date/Time" format:
  - "Date of Purchase"
  - "Time of Purchase"
  - "Date of Journey"
  - "Departure Time"
  - "Arrival Time"
  - "Actual Arrival Time"

Through these steps, we transformed the dataset into a more structured and standardized form, facilitating more accurate and meaningful insights.

3rd Phase: Data Analysis and Insights Generation

Assigned to (Mona Hassan – Sara Mohamed – Dina Yasser)

At this phase we aim to fully understand the dataset to better extract the valuable information that will aid in generating the best insights.

Here are the railway stations affected by Storm Henk:

1. **Winchester and Micheldever**: Landslide blocked all railway lines between these stations on January 4th.

2. **Yeovil Junction and Exeter**: Landslide near Crewkerne in south Somerset blocked all railway lines between these stations on January 5th.

3. **Maidstone East**: Landslide blocked several lines by this station in Kent.

4. **Robertsbridge**: Landslide hit the line at this station in East Sussex.

5. **Arlesey**: Landslide on the railway line between Stevenage and Peterborough.

6. **London Paddington and south Wales**: Flooding caused train services to be diverted between these stations.

7. **Coventry and Birmingham International**: Flooding hit the railway between these stations.

8. **London Paddington and Maidenhead**: Electricity failure caused route blockages between these stations.

9. **Watford Junction and London Euston**: Damage to overhead wires caused earlier closures between these stations.

10. **Richmond and Willesden Junction**: Urgent repairs on the track at Gunnersbury affected services between these stations.

11. **Swindon and Bristol Parkway**: Flooding blocked several parts of the network between these stations.

These disruptions were caused by landslides, flooding, and power failures due to the severe weather conditions brought by Storm Henk.

Source: [Storm Henk triggers landslide chaos across UK rail network | Ground Engineering](#)

Source: [Storm Henk batters UK leading to power outages, travel disruption and flooding - BBC News](#)

The key points from Payment Options:

1. **Pay As You Go with Contactless or Oyster**:
   - No need to buy tickets in advance.
   - Touch in and out with a contactless card/device or Oyster card to be charged the correct fare.
   - Contactless cards/devices are accepted on National Rail, London Underground, DLR, London Buses, and London Trams within London and at some stations across the South East.
   - Daily and weekly (Monday to Sunday) capping ensures you won't be charged more than a certain amount.
2. **Travelcards**:
   - Purchased ahead of your journey for unlimited travel within specific zones.
   - Available for 1 day, 7 days, 1 month, or any period between 1 month and 1 year.
   - Travelcards can be used on National Rail services, London Underground, DLR, London Buses, and London Trams in specified zones.

Contactless Payment Specifics:

- Charged the cheapest adult fare for your journey on the day and time of travel.
- Railcard and other discounts are not available with contactless payment.
- Capping ensures cost-efficiency for frequent travel.

Oyster Card Specifics:

- Allows storage of up to £90 of pay as you go credit.
- Usable on a wide range of transport services within London Zones 1 to 9 and some services beyond.
- Capping limits daily and weekly travel costs.
- Railcard discounts can be applied for up to 1/3 off travel.
- Child rate fares available; under 5s travel free with a fare-paying adult.
- Can store up to 3 Travelcard or London Bus & Tram Pass Season tickets.

Travelcards:

- Provide unlimited travel within specified zones.
- Available for different durations: 1 day, 7 days, 1 month, etc.
- Paper One Day Travelcards; longer duration Travelcards added to Oyster or Smartcard.
- Boundary Zone tickets available for travel beyond permitted zones.
- Return tickets from home counties, South, South East, or further afield may include a One Day Travelcard.

The ticket details, discounts, and refund policies for different ticket types:

Advance Tickets

- **Details**:
    - Great value for long journeys, must be bought in advance.
    - Valid only on the date and train specified.
    - Single journeys only, can combine tickets for return trips.
    - Sold in limited numbers, subject to availability.
    - Available up to 12 weeks ahead of travel, sometimes until the day of travel.
- **Discounts**:
    - 50% off for children aged 5 to 15.
    - 50% off for 16-17 Saver Railcard holders on adult Standard class Advance fares.
    - 1/3 off Standard class Advance fares with any National Railcard.
    - 1/3 off First Class Advance fares with 16-25, 26-30, Senior, Two Together, HM Forces, Veterans, or Disabled Persons Railcards.
- **Refund Policy**:
    - Non-refundable unless the train is delayed or cancelled and you choose not to travel.
    - Can amend the ticket up to the time of travel, with a £10 fee and any difference in fare.

Off-Peak and Super Off-Peak Tickets

- **Details**:
    - Available for travel at less busy times on weekdays, all day on weekends.
    - Cheaper but may require travel at specified times, days, or routes.
    - Valid for 1 month from the date shown on the ticket.
    - Can buy at any time before travel.
- **Discounts**:
    - 50% off for children aged 5 to 15.
    - 50% off for 16-17 Saver Railcard holders on adult Off-Peak and Super Off-Peak fares.
    - 1/3 off Standard Class Off-Peak and Super Off-Peak fares with any National Railcard.
- **Refund Policy**:
    - Refundable with no admin fee if the service is delayed or cancelled and you choose not to travel.
    - Can get a refund and rebook with no admin fee if traveling on another day not covered by the original ticket.
    - Refundable with a maximum £5 admin fee if canceling for other reasons.

<u>Anytime Tickets</u>

- **Details**:
    - Fully flexible with no time restrictions.
    - Can travel on any train on the route shown.
    - Anytime Singles valid for 2 days, Anytime Returns for 5 days (outward) and 1 month (return).
    - Can buy at any time before travel.
- **Discounts**:
    - 50% off for children aged 5 to 15.
    - 50% off for 16-17 Saver Railcard holders on adult Anytime fares.
    - 1/3 off Standard Class Anytime fares with any National Railcard.
- **Refund Policy**:
    - Refundable with no admin fee if the service is delayed or cancelled and you choose not to travel.
    - Can get a refund and rebook with no admin fee if traveling on a different day.

Find out more about our Railcards.

See the Advance tickets terms and conditions for full information.

Find out more about pay as you go with contactless(external link, opens in a new tab) and check out the pay as you go with contactless map.

Purchase & Ticketing Insights

These questions have been designed to extract valuable insights from the dataset and enhance our understanding of the purchase and ticketing patterns. Here's why each question is useful:

1. **What are the most common purchase types (Online vs. Station)?**

2. **What is the demographic profile of passengers purchasing online versus at stations?**

3. **What Percentage of tickets are purchased with Rail-cards?**

4. **What are the most sold ticket type and class are purchased and why is the most used?**

5. **What are the peak times for ticket purchases?**

6. **What is the most ticket price sold? What are the primary reasons of train delays, and Which is the highest percentage?**

7. **How do reasons affect total delayed hours?**

8. **Which departure stations experience the highest frequency of delays, and what are the associated reasons?**

9. **How do delays and cancellations vary from January to April 2024?**

10. **Which type of delay contributes most significantly to total service problems?**

11. **What are the primary reasons of train delays, and Which is the highest percentage?**

12. **How do reasons affect total delayed hours?**

13. **Which departure stations experience the highest frequency of delays, and what are the associated reasons?**

14. **How do delays and cancellations vary from January to April 2024?**

15. **Which type of delay contributes most significantly to total service problems?**

Assigned to Sara Farag and Nada Saad

First Dashboard Questions, Insights, Recommendations?

1) **What are the most common purchase types (Online vs. Station)?**

2) **What is the demographic profile of passengers purchasing online versus at stations?**

3) **What Percentage of tickets are purchased with Railcards?**

• **Purchase Preferences:** With 58.51% of tickets bought online, passengers favor digital convenience. However, nearly half still purchase at stations, suggesting a need for hybrid service approaches.

• **Purchase with Railcard preference:** 66% of tickets purchased aren't with Railcards, following by a 15.3% Adult Railcard usage.

• **Edinburgh Waverley Station,** which is located in the heart of Edinburgh, Scotland, near Princes Street and Edinburgh Castle, has a 100% Station Purchase with 100% No-Railcard Usage. People might exclusively buy tickets at the station and avoid using Railcards because Edinburgh is a popular tourist destination and tourists aren't unfamiliar with railcards or UK train systems. A significant portion of passengers might be business travelers who prioritize flexibility so they purchase an Advance Ticket and don't use railcards due to eligibility or lack of need for discounts.

• **Bristol Temple Meads station has** 100% Online Purchase with 100% No-Railcard Usage.

• **All London Stations (Paddington, St. Pancras, King's Cross, Huston**) is showing a 61% Online Purchasing for several reasons:

These major London stations, being hubs of travel and commerce, might show a preference for online ticket purchasing for several reasons:

1. Tech-Savvy Demographic: Many passengers passing through these stations, particularly commuters and business travelers, are likely comfortable using digital platforms for efficiency.

2. Convenience: London travelers often plan trips ahead and buy tickets online to save time, avoiding queues and delays at physical ticket counters.

3. Mobile Integration: The widespread use of mobile apps for ticket purchases makes online buying particularly appealing in urban environments like London.

4. Promotion of Digital Services: These stations might actively encourage online purchasing to streamline operations and reduce congestion at ticket counters.

• **Online purchasing aligns with the fast-paced lifestyle of many Londoners,** and the stations likely benefit from the efficiency it brings to passenger flow.

• **Birmingham New Street and York Station** is showing nearly 70 % and 81% Online Purchasing with 20.97% and 40% respectively Disabled Railcard Usage. Birmingham New Street and York stations offer assistance to older and disabled passengers who need help accessing our facilities and train services. The Passenger Assistance team have access to wheelchairs and Passenger Assistance vehicles to help them assist passengers in and around the station.

1. There is step-free access to all platforms, making it easier for wheelchair users and those with mobility challenges.

2. Ticket machines are designed to be accessible, with height adjustments available.

3. The station includes accessible toilets and lifts to help passengers move between levels.

4. Designated parking spaces for Blue Badge holders are available.

5. According to a study by the disability charity Leonard Cheshire, more than 40% of UK train stations are not accessible to physically disabled people.

• **Liverpool Lime St**. is serving nearly 20 million passengers every year, and it has 58% Purchasing at Station, with 18% of Adult-Railcard Usage.

1- Many passengers might travel on shorter, regional routes, where railcards may offer limited financial benefits compared to longer-distance journeys.

2- Only 18% of passengers using Adult Railcards could indicate a need for better awareness campaigns, as some might not fully understand the potential savings railcards provide.

3- Many commuters might buy daily or flexible tickets at the station, which might not always be eligible for railcard discounts.

4- Advance tickets are sold in limited numbers and are subject to availability. They are typically cheaper but must be purchased before the day of travel or sometimes up to 10 minutes before departure. These tickets are tied to a specific train and time, so you need to know your travel details when purchasing. If you have a railcard, you can get discounts on Advance tickets, even when buying them at the station.

• **Oxford Station** is showing 58% of Online Purchasing with 55% Adult Railcard Usage. Which could be related to Oxford's academic and professional environment attracts individuals who are familiar with online systems and value efficiency. Also, Many passengers, including students and commuters, likely plan their journeys in advance, opting for online purchasing to secure tickets conveniently. Adult Railcards, such as the 16-25 Railcard or the Two Together Railcard, may appeal to Oxford's demographic—students, young professionals, or traveling couples.

- Once you have a Rail-card, you can get 1/3 off the price of your rail travel, making it a great way to secure cheap fares. Some Rail-cards are also available as a three-year option for £80, giving you even greater value for money (therefore costing around £26 per year).

- A very low-rate "2 stars" Mobile App is in need of updates and a more user-friendly interface.

- **Negative Reviews:**

  Andrew Forsythe – 14 March 2025

  "The app is just horrible and frustrating to use. The old well designed easy to use user interface was sacrificed and replaced with one that is complex and cumbersome to use. Also now useless for live service updates. The in app ticket purchase process is excessively cumbersome as well. Quicker and easier just to purchase tickets at the station. Also if the app developers say customers asked for a new app design. I certainly don't ask for a new app. Did anyone?"

  Janine Louie - 3 October 2024

  "This latest update is awful. I've lost all my preset journeys and now when looking up live trains, instead of being able to see 10 journeys in one page, I can only see about 2 and need to scroll a lot to see what used to be a succinct page of results. The UX designer needs to learn what the users actually want to want rather than just making the design more modern looking."

Business Problems and Solutions:

1. Low Railcard Usage (e.g., Edinburgh Waverley, Liverpool Lime St.):

• Problem: Passengers might not be aware of railcard benefits or find the upfront cost a barrier. Business travelers and tourists may avoid railcards due to lack of eligibility or perceived relevance.

• **Enhancement:**

- Launch targeted awareness campaigns emphasizing cost-saving benefits and eligibility criteria for railcards.

- Offer introductory discounts or trial periods for railcards to encourage adoption.

- Ensure railcard promotions are clearly visible at stations and online platforms.

2- Exclusively Station Purchases (e.g., Edinburgh Waverley, Liverpool Lime St.):

• Problem: High reliance on station-based purchasing could lead to congestion during peak times and may reflect a gap in digital convenience.

- **Enhancement:**

  o Invest in user-friendly ticketing apps and websites.

  o Include multi-language support and simplified interfaces to cater to tourists.

  o Add kiosks and help desks to educate passengers about online purchasing benefits.

3. Accessibility Challenges (e.g., Birmingham New Street, York Station):

- Problem: Despite some stations having accessibility features, overall UK rail accessibility remains a challenge for disabled passengers.

- **Enhancement:**

  o Prioritize upgrades for physical station infrastructure across more stations (e.g., lifts, ramps, signage).

  o Extend staff training to assist passengers with diverse needs.

  o Collaborate with charities to identify key areas for improvement and raise funds for accessible upgrades.

4. High Online Purchasing but No Railcard Usage (e.g., Bristol Temple Meads):

- Problem: Despite passengers favoring online purchasing, railcard adoption is negligible.

- **Enhancement:**

  o Embed railcard suggestions directly into online ticketing platforms, with cost-saving comparisons highlighted.

  o Promote railcards on social media channels targeting regions with low adoption rates.

5. Underutilized Discounts for Short-Distance Routes (e.g., Liverpool Lime St.):

- **Problem: Adult railcards may not appear beneficial for shorter or flexible journeys.**

- **Enhancement:**

  o Develop railcard options tailored for short-distance travel or commuter packages.

  o Offer dynamic ticketing discounts tied to railcards for flexible travel scenarios.

General Recommendations:

1. Hybrid Services: Leverage both digital and physical channels for a seamless passenger experience.

2. Customized Offers: Roll out station-specific campaigns to address unique passenger demographics and preferences.

Continue First Dashboard:

4) What are the most sold ticket type and class purchased? Why is it the most preferable?

5) What are the peak times for ticket purchases?

6) What is the most ticket price sold?

- **Ticket Class vs. Ticket Type**

  o Standard Advance tickets (49.93%) are the most sold.

  o Standard Off-Peak (25.24%) and First Class Advance (5.56%) follow.

  o Insight: Passengers prioritize affordability and early booking.

  o First Class still holds a significant chunk, especially with Advance fares.

- **Ticket Price Variation by Type:** Yes, ticket prices vary significantly by type:

  o Advance Tickets: These are usually the cheapest option under £10 but must be booked in advance. They are non-flexible and tied to specific trains.

  o Off-Peak Tickets: These are moderately priced and allow travel during less busy times, offering flexibility compared to Advance tickets. Off-Peak Tickets are favored by leisure travelers who can avoid rush hours. In general, Off-Peak hours begin at 09:30 from Monday to Friday in cities and large towns, and at 09:00 everywhere else. If your train is scheduled to depart after this time, you can travel with an Off-Peak ticket. Weekends and bank holidays are Off-Peak all day. If you travel outside of the permitted times or routes you will have to pay the difference between the price of your ticket and the Anytime ticket. You may also be liable for a Penalty Fare. The price of Off-Peak and Super Off-Peak tickets will vary. Where those tickets are available, they will cost less than an Anytime ticket.

  o Anytime Tickets: These are the most expensive as they provide maximum flexibility, allowing travel at any time of the day without restrictions. Anytime Tickets are often purchased by business travelers needing flexibility

  o This pricing structure is designed to balance demand and encourage passengers to travel during less busy periods.

- Peak Times for Ticket Purchases: Peak times for ticket purchases typically align **with:**

  - Tickets Sold per Hour

  - Peak ticket sale hours:

  - 6:00–7:00 AM (top)

  - 5:00–6:00 PM

  - 6:00–7:00 PM

  - Insight: Strong alignment with commuter rush hours.

- Morning Commutes: (5,466) trips Between 6:00 AM and 8:00 AM, as passengers travel to work or school.

- Evening Commutes: (5,511) trips Between 4:00 PM and 6:00 PM, when people return home.

- **Most Ticket Price Sold: Ticket prices varied from £1 to £267, with £3 tickets being the most sold for many reasons:**

  - Represent a low-cost option to travelers for short or off-peak trips

  - Suitable for short distance or commuter journeys

  - Might to be priority for quick trips that doesn't require additional features

  - Tickets could align with online or advanced booking systems where lower prices are offered for early reservations.

The most commonly sold ticket price depends on the type of journey and passenger demographics.

Here's an analysis of the business needs, problems, and recommendations based on your data insights:

- **Business Needs**

  1. Dynamic Ticket Pricing Strategies:

  - A pricing structure tailored to passenger behavior can balance demand and ensure efficient utilization of train capacity.

  2. Targeted Marketing Campaigns:

  - Promote specific ticket types (e.g., Advance or Off-Peak) and railcards to the right audience segments based on their preferences and travel patterns.

3. Operational Adjustments During Peak Times:

- o Enhance station and service readiness during peak morning and evening commute hours, as well as holiday periods, to handle the increased passenger flow.

4. Encourage Economical Travel Options:

- o Focus on promoting affordable ticket options like £3 tickets for short journeys to attract budget-conscious travelers.

Business Problems and Solutions:

1. Ineffective Utilization of Ticket Types

- • Problem: Higher-priced Anytime Tickets may not be fully utilized by cost-conscious travelers, while Off-Peak tickets might have lower visibility.

- • Solution:

  - o Develop personalized offers via online platforms to encourage uptake of Off-Peak and Advance tickets.

  - o Highlight cost savings during ticket purchase processes (e.g., "Save £X by traveling Off-Peak").

2. Peak-Time Congestion

- Problem: High ticket sales during morning and evening commutes could lead to overcrowded services and operational strain.

- **Solution:**

  - o Enhance scheduling by adding extra services during peak periods.

  - o Offer off-peak incentives (e.g., discounts for travel just outside peak hours) to spread demand more evenly.

3. Pricing Perception for Frequent Travelers

- • Problem: Frequent short-distance travelers might feel ticket prices don't justify their journeys.

- • Solution:

  - o Introduce multi-trip passes or discounted subscription plans tailored to commuter needs.

  - o Bundle services (e.g., rail tickets combined with parking discounts) to enhance the value proposition.

Recommendations

1. **Passenger-Centric Pricing and Marketing:**

   o Use predictive analytics to segment passenger demographics and optimize pricing strategies for each group.

   o Offer periodic promotions for specific ticket types to balance train occupancy.

2. **Enhanced Technology Integration:**

   o Upgrade online ticketing systems to better recommend economical travel options, railcards, and journey flexibility.

3. **Strategic Capacity Management:**

   o Monitor ticket sales to proactively adjust train frequency and capacity during high-demand periods.

4. **Flexible Solutions for Commuters:**

   o Introduce new commuter-focused products like digital season passes or credits for unused trips.

Recommendations:

- **Digital Sales Strategy**

   o Invest in digital platform UX/UI to further boost online purchases.

   o Launch app-exclusive deals or early bird discounts to reinforce this channel.

- **Ticket Class Optimization**

   o Promote Advance Standard tickets with flexible features (e.g., partial refunds or easier changes).

   o Offer First Class upgrades at off-peak times for a small premium to encourage upselling.

- **Railcard Adoption**

   o Run educational campaigns showing savings with Railcards.

   o Offer bundle promotions (e.g., buy ticket + Railcard at discounted rate).

- **Peak Hour Planning**

   o Increase support staffing and automated kiosks between 6–7 AM and 5–7 PM.

## **Suggestions and Recommendations for the first dashboard:**

1- Enhance the UK Railway Mobile App by making it user-friendly, and smooth the process of purchasing and booking trips through the official website.

2- Launch targeted awareness campaigns emphasizing cost-saving benefits and eligibility criteria for railcards via social media platforms and mobile ads, also make sure they are clearly visible at stations and online platforms.

3- Offer introductory discounts or trial periods for railcards to encourage subscription.

4- Add kiosks and help desks to educate passengers about online purchasing benefits, or aid in helping passengers in self-purchasing their tickets in the station.

5- Extend staff training to assist passengers with diverse needs especially disabled people.

6- Collaborate with charities to help raise funds for accessible upgrades in the major cities such as London and Liverpool.

7- Develop railcard options tailored for short-distance travel or commuter packages.

8- Launch app-exclusive deals or early bird discounts to reinforce this channel.

9- Promote Advance Standard tickets with flexible features (e.g., partial refunds or easier changes).

10- Offer First Class upgrades at off-peak times for a small premium to encourage upselling.

11- Offer dynamic ticketing discounts tied to Railcards for flexible travel scenarios.

12- Increase support staffing and automated kiosks between 6–7 AM and 5–7 PM.

13- Promote off-peak travel incentives to spread demand (discounts, loyalty points).

14- Offer bundle promotions (e.g., buy ticket + Railcard at discounted rate).

Second Dashboard

Assigned to Sarah Farag and Dina Yasser

Second Dashboard Questions, Insights, Recommendations?

1. What is the total revenue generated by UK train ticket sales in January – April 2024, including and excluding refunds?

2. Which departure station generates the highest revenue, and how does it compare to other stations?

3. What are the revenue contributions of different ticket types, such as Standard Advance or First Class Anytime?

4. How does revenue vary across different hours of the day, and which time slot generates the most revenue?

5. How do payment methods (credit card, contactless, debit card) influence total revenue, and which is the most profitable?

6. What impact do refund requests have on net revenue, particularly at high-refund stations like Liverpool Lime Street and Manchester Piccadilly?
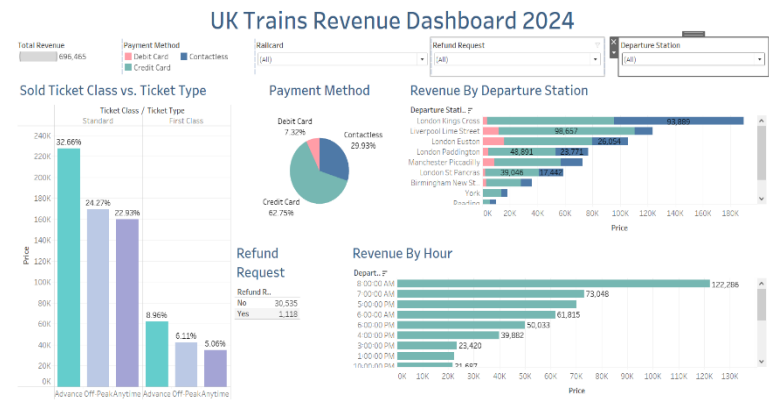
Initial Insights:

1. Total Revenues: £741.92K, with £696.465K after subtracting refunded tickets.

2. Trips from **London King's Cross** Station ranked first in revenue generation, contributing over £199K.

3. Advance tickets are favored by the majority for several compelling reasons, despite their non-refundable nature:

   - **Cost Savings**: Advance tickets are usually the most economical option, offering significant discounts compared to Off-Peak or Anytime tickets.

Budget-conscious travelers are willing to trade flexibility for affordability.

- **Planned Journeys**: Passengers who book their trips ahead often have fixed schedules and are confident they won't need to make changes. This makes the non-refundable condition less of a concern.

- **Guaranteed Seats**: Advance tickets typically come with a seat reservation, which is attractive for passengers who prioritize comfort and assurance, especially on popular routes during peak times.

- **Behavioral Habits**: Regular travelers, like commuters or students, often prefer advance tickets as they are accustomed to planning trips well in advance to maximize savings.

- **Popularity of Online Platforms**: Advance tickets are widely promoted on digital booking systems, encouraging their adoption among tech-savvy users.

- While Advance tickets are nonrefundable unless the ride is cancelled or delayed, It remains the popular option for passengers.

4. 31.653K Tickets sold: Credit card purchases generated 62.7% (£437K) of total revenues, while standard tickets accounted for more than 79% (£590K). Advanced Standard Tickets generated the highest share, at 32% (£242K).

5. Purchasing train tickets in the UK with a debit card generated 7.3% and This method might be less preferable for a few reasons:
   - Ticket Collection Issues: Some systems require the physical debit card used for payment to collect tickets from machines. If the card is lost or unavailable, this can cause inconvenience.
   - Refund Options: Debit cards may have fewer protections compared to credit cards when it comes to disputes or refunds for canceled or delayed journeys.
   - Security Concerns: Debit cards are directly linked to your bank account, so some travelers prefer using credit cards or contactless methods for added security.
   - Preference for Contactless Payments: Many travelers opt for contactless payment methods which are faster and more convenient.

6. 90.25% of debit card users, who didn't benefit from Railcard usage, requests a refund (516 request) of total refund requests (1118 requests).

There are several reasons why passengers purchasing tickets with debit cards might be more likely to request refunds:

- **Direct Account Impact**: Debit card transactions withdraw funds directly from the user's bank account. If there's an issue with the service or ticket, passengers may feel a stronger urgency to reclaim their money compared to credit card users, whose transactions don't immediately affect their available balance.
- **Refund Processing Speed**: Refunds to debit cards are typically processed faster than credit cards, which might encourage passengers to request refunds more readily.
- **Budget-Conscious Travelers**: Debit card users may be more budget-conscious and sensitive to service quality, prompting them to seek refunds if expectations aren't met.
- **Transaction Errors**: Mistakes like duplicate charges or incorrect ticket details might occur more frequently with debit card transactions, leading to refund requests.

7. Disabled Railcard passengers never requested a refund.
8. Departure Stations: Passengers departing from stations like **Liverpool Lime St. and Manchester Piccadilly** have a significantly higher likelihood of requesting a refund (500 Requests), compared to passengers from other stations with 95.64% purchasing with a debit card.
9. 77% of passengers paying by debit card are likely to request a refund if their journey is delayed or cancelled, while over 70% of those using credit or contactless cards do not.
10. Standard off-peak tickets are the most requested for a refund (32.81%), while Standard Advance tickets is the second highest request for a refund (31.82%).

Business Needs:

1. **Optimize Payment Methods:**

   - Enhance systems to support popular payment methods like credit cards (62.75%) and contactless payments (29.93%).
   - Promote lesser-used payment options like debit cards (7.32%) for wider accessibility.

2. **Improve Ticket Distribution:**

   - Focus on increasing sales for ticket types with lower percentages, such as First Class Off-Peak (6.11%) and First Class Anytime (5.06%), to diversify revenue sources.

3. **Maximize Revenue at High-Performing Stations:**

   - Stations like Liverpool Lime Street (£98,657) and London Kings Cross (£93,889) generate significant revenue. Ensuring top-tier services at these stations can further boost profitability.

4. **Address Refund Issues:**

   - With 1,118 refund requests, analyzing reasons for dissatisfaction and resolving key issues could enhance customer retention.

5. **Use Peak Hour Demand:**

   o   Revenue spikes at 8:00 AM (£122,286) and 7:00 AM (£73,048) highlight peak travel times. Offering tailored promotions during these hours can enhance profitability.

Important Insights:

1. **Ticket Preferences:**

   o   Standard Advance tickets lead sales at 32.66%, indicating passengers favor cost-effective pre-booking options.

   o   Standard Anytime tickets (22.93%) show consistent demand for flexibility.

2. **Revenue Patterns by Departure Station:**

   o   London-centric stations like Paddington (£48,891), St Pancras (£39,046), and Euston (£26,054) generate substantial revenue, reflecting strong urban travel demand.

   o   Stations with lower revenue, such as York (£9,046) and Reading (£8,657), could benefit from targeted marketing efforts.

3. **Revenue by Hour:**

   o   Morning hours dominate revenue generation (8:00 AM being the highest). This reinforces the importance of efficient service scheduling and staffing during early hours.

4. **Refund Impact:**

   o   Refunds requested are relatively low, but addressing even minor service issues could enhance customer loyalty and reduce loss.

Recommendations:

1. **Promotion Strategies:**

   o   Highlight Advance ticket benefits for cost-conscious travelers while emphasizing perks for First Class options to attract premium passengers.

   o   Offer a discounted ticket to First-Class to the most loyal customer who purchased the last 10 trips via the Online Purchasing method.

2. **Station-Specific Campaigns:**

   o   Use insights from high-revenue stations like Liverpool Lime Street to implement best practices across other stations.

3. **Service Improvements:**

- o Offer promotional campaigns, bundles, packaged discounts to commuters and students who are traveling through the same route at 8:00 Am from Monday to Friday.

- o Provide affordable offers on Breakfast through the vending machines at the Railways stations in the early mornings.

- o Enhance accessibility and passenger services during peak hours to maintain efficiency and reduce bottlenecks.

4. **Feedback Collection:**

- o Proactively gather feedback to address refund-related concerns and improve overall satisfaction.

Suggestions and Recommendations for the second dashboard:

1. Highlight Advance ticket benefits for cost-conscious travelers while emphasizing perks for First Class options to attract premium passengers.
2. Offer a discounted ticket to First-Class to the most loyal customer who purchased the last 50 trips via the Online Purchasing method.
3. Use insights from high-revenue stations like Liverpool Lime Street to implement best practices across other stations.
4. Offer promotional campaigns, bundles, packaged discounts to commuters and students who are traveling through the same route daily at 8:00 Am from Monday to Friday.
5. Provide affordable offers on Breakfast through the vending machines at the Railways stations in the early mornings.
6. Enhance accessibility and passenger services during peak hours to maintain efficiency and reduce congestion.
7. Proactively gather feedback to address refund-related concerns and improve overall satisfaction.

Third Dashboard

Assigned to Sahar Hamdi and Sara Mohmmed

Third Dashboard Questions, Insights, and Recommendations

1. What are the primary reasons of train delays, and Which is the highest percentage?

2. How do reasons affect total delayed hours?

3. Which departure stations experience the highest frequency of delays, and what are the associated reasons?

4. How do delays and cancellations vary from January to April 2024?

5. Which type of delay contributes most significantly to total service problems?
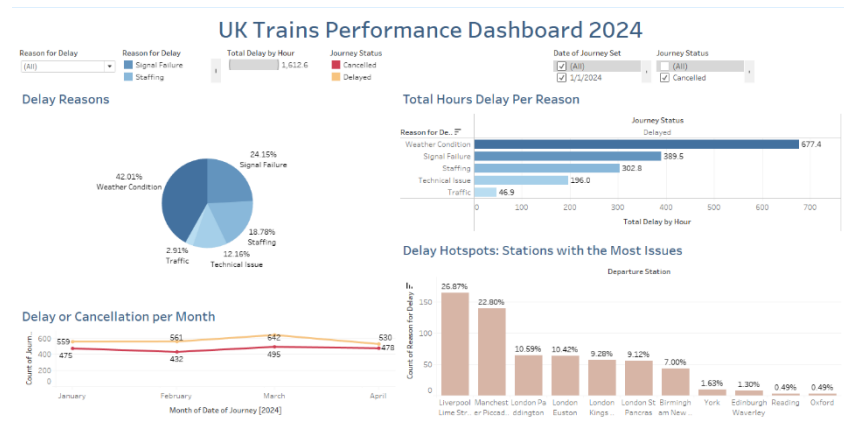
Insights

1. Delay Reasons

- Insight: The primary reasons for train delays are clearly highlighted.

  - Weather Conditions account for 42.01% of all delays, making it the leading cause.

  - Signal Failure is the second most common reason, contributing to 24.15% of delays.

  - Other significant causes include Starting Delays (18.79%) , Technical Issues (13.16%) , and Traffic (2.51%).

2. Total Hours Delay Per Reason

- Insight: The total delay hours caused by each reason provide a quantitative perspective.

    - Weather Conditions resulted in 677.4 hours of delays.

    - Signal Failure caused 359.8 hours of delays.

    - Starting Delays contributed 302.8 hours , while Technical Issues added 136.9 hours .

    - Traffic caused 46.9 hours of delays.

3. Delay or Cancellation Per Month

- Insight: The monthly trend shows fluctuations in delays and cancellations.

    - January had the highest number of delays/cancellations (605).

    - February saw a decrease (561), followed by another increase in March (642) and April (630).

    - There is no clear seasonal pattern, but January appears to be the most problematic month.

4. Delay Hotspots: Stations with the Most Issues

- Insight: Certain stations experience disproportionately high delays.

    - Liverpool Lime Street is the top hotspot, accounting for 36.07% of all delays.

    - Manchester Piccadilly follows closely with 22.80%.

    - Other major hotspots include London Paddington (10.59%), London Euston (10.42%), and London King's Cross (9.28%).

    - Smaller stations like York and Wolverhampton also appear on the list.

Suggestions

1. Insulate power stations with protective materials to withstand storms.

2. Invest in hybrid train engines to reduce reliance on traditional power systems.

3. Address staff needs by improving work conditions and compensation.

4. Offer additional bonuses or support for handling storm-related disruptions and high work pressure.