

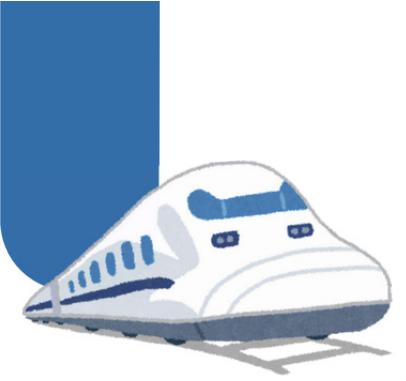


UK Train Rides: What the Data Says

وزارة الاتصالات
 وتكنولوجيا المعلومات



Digital Egypt Pioneers



Our Team

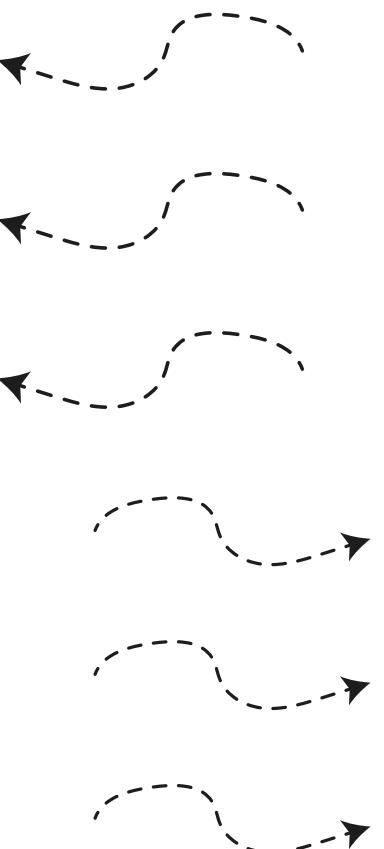
وزارة الاتصالات
هيئة تنمية مجتمع المعلومات



Mahmoud Ahmed Mahmoud

Mennatallah Emam Ebrahim

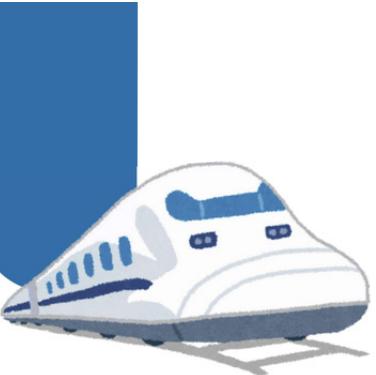
Abdelrahman Mohamed Sayed



Menna Abd El Hameed Abd El Hasib

Iman Abduldayem Abdulhamid

Arwa Hassan Abdelrazek Mohammed



Project objectives

Week 1: Data Modeling and Preprocessing

Perform data cleaning to address missing values, errors, and outliers.

Prepare an analysis-ready dataset.

Week 2: Data Analysis Questions Phase

Explore the dataset using SQL and Python (Pandas, Matplotlib).

Determine relationships and patterns within the data.

Compile a validated list of answerable questions.

Week 3: Visualization & Final Presentation

Build a comprehensive visualization dashboard using Tableau.

Highlight analytical insights

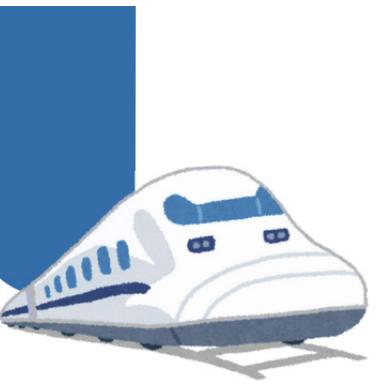
Demonstrate the value of data-driven decisions through visual storytelling.

Week 4: Forecasting Phase

Define key forecasting questions relevant to business needs.

Predict future ride demand, revenue, and ticket class usage.

Visualize forecast results with professional plots and charts.



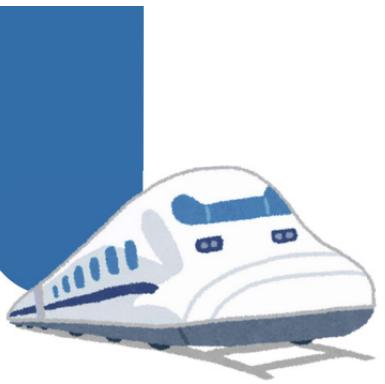
About Data Set

وزارة الاتصالات
 وتكنولوجيا المعلومات



This dataset provides comprehensive UK train journey data from various sources, enabling analysis of performance, routes, and passenger behavior for service enhancement and future forecasting.





The Data Source Before Cleaning

وزارة الاتصالات
بتكنولوجيا المعلومات



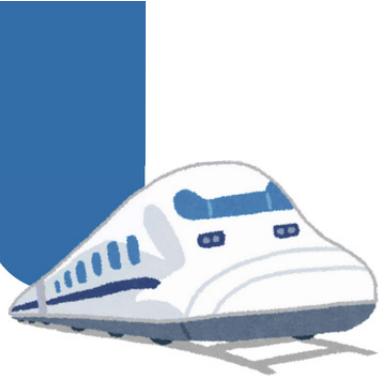
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
action ID	Date of Purchase	Time of Purchase	Purchase Type	Payment Method	Railcard	Ticket Class	Ticket Type	Price	Departure Station	Arrival Destination	Date of Journey	Departure Time	Arrival Time	Actual Arrival Time	Journey Status	Reason for Delay	Refund R
ba8-b3dc-4677-b176	12/8/2023 0:00	0.528599537	Online	Contactless	Adult	Standard	Advance	43	London Paddington	Liverpool Lime Street	1/1/2024 0:00	0.458333333	0.5625	0.5625	On Time	No	
1b0-f214-4197-be53	12/16/2023 0:00	0.47431713	Station	Credit Card	Adult	Standard	Advance	23	London Kings Cross	York	1/1/2024 0:00	0.40625	0.482638889	0.486111111	Delayed	Signal Failure	No
a96-f713-40d9-9629	12/19/2023 0:00	0.827395833	Online	Credit Card	None	Standard	Advance	3	Liverpool Lime Street	Manchester Piccadilly	1/2/2024 0:00	0.760416667	0.78125	0.78125	On Time	No	
11-4f7-4c87-8ab5	12/20/2023 0:00	0.95875	Station	Credit Card	None	Standard	Advance	13	London Paddington	Reading	1/1/2024 0:00	0.895833333	0.9375	0.9375	On Time	No	
b45-0762-485e-a7a3	12/27/2023 0:00	0.765925926	Online	Contactless	None	Standard	Advance	76	Liverpool Lime Street	London Euston	1/1/2024 0:00	0.697916667	0.791666667	0.791666667	On Time	No	
j88-3d95-44ef-99fa	12/30/2023 0:00	0.330625	Online	Credit Card	None	Standard	Advance	35	London Kings Cross	York	1/1/2024 0:00	0.260416667	0.336805556	0.336805556	On Time	No	
79d-85a4-4b41-a607	12/31/2023 0:00	0.001400463	Station	Credit Card	Adult	Standard	Advance	2	London Euston	Oxford	1/1/2024 0:00	0.9375	0.986111111	0.986111111	On Time	No	
lab-f808-46fa-bf2b	12/31/2023 0:00	0.066180556	Station	Contactless	Disabled	Standard	Advance	2	Liverpool Lime Street	Manchester Piccadilly	1/1/2024 0:00	0	0.020833333	0.020833333	On Time	No	
bf-4291-41ec-a37d	12/31/2023 0:00	0.071631944	Station	Credit Card	None	Standard	Advance	37	London Euston	York	1/1/2024 0:00	0	0.076388889	0.088194444	Delayed	Signal Failure	No
ba7-1923-459d-b046	12/31/2023 0:00	0.129074074	Online	Credit Card	None	Standard	Advance	13	London Paddington	Reading	1/1/2024 0:00	0.0625	0.104166667	0.104166667	On Time	No	
a7d-e76c-49f2-b49f	12/31/2023 0:00	0.143483796	Online	Contactless	None	Standard	Advance	8	York	Durham	1/1/2024 0:00	0.072916667	0.107638889	0.107638889	On Time	No	
7ac-d590-4356-9eaa	12/31/2023 0:00	0.161238426	Online	Contactless	Adult	Standard	Advance	8	London Paddington	Reading	1/1/2024 0:00	0.09375	0.135416667	0.135416667	On Time	No	
d75-566a-41aa-3468	12/31/2023 0:00	0.246782407	Online	Contactless	None	Standard	Advance	3	Manchester Piccadilly	Liverpool Lime Street	1/1/2024 0:00	0.177083333	0.197916667	0.197916667	On Time	No	
545-eb6f-49b2-9b5a	12/31/2023 0:00	0.280960648	Online	Contactless	None	Standard	Advance	3	Manchester Piccadilly	Liverpool Lime Street	1/1/2024 0:00	0.208333333	0.229166667	0.229166667	On Time	No	
2a6-88a8-40fb-bacc	12/31/2023 0:00	0.337384259	Online	Credit Card	Disabled	Standard	Advance	15	Birmingham New Street	London St Pancras	1/1/2024 0:00	0.270833333	0.326388889	0.326388889	On Time	No	
3b4-995e-49bf-93a3	12/31/2023 0:00	0.34505787	Online	Credit Card	None	Standard	Advance	13	London Paddington	Reading	1/1/2024 0:00	0.322916667	0.364583333	0.364583333	On Time	No	
611-342a-4b17-90dc	12/31/2023 0:00	0.349479167	Online	Credit Card	None	Standard	Advance	13	London Paddington	Reading	1/1/2024 0:00	0.322916667	0.364583333	0.364583333	On Time	No	
6ac-a976-4909-a26e	12/31/2023 0:00	0.381481481	Online	Credit Card	None	Standard	Advance	8	London St Pancras	Birmingham New Street	1/1/2024 0:00	0.3125	0.368055556	0.368055556	On Time	No	
452-c491-468d-b39c	12/31/2023 0:00	0.383576389	Online	Credit Card	None	Standard	Advance	35	London Kings Cross	York	1/1/2024 0:00	0.3125	0.388888889	0.388888889	On Time	No	
992-6c6c-4569-914e	12/31/2023 0:00	0.446087963	Online	Credit Card	None	Standard	Advance	35	London Kings Cross	York	1/1/2024 0:00	0.375	0.451388889	0.451388889	On Time	No	
fc-aea0-424f-b30e	12/31/2023 0:00	0.498090278	Station	Debit Card	Adult	Standard	Advance	7	Birmingham New Street	Manchester Piccadilly	1/1/2024 0:00	0.46875	0.524305556	0.545833333	Delayed	Technical Issue	Yes
358-044d-4000-b70e	12/31/2023 0:00	0.50818287	Station	Credit Card	Disabled	Standard	Advance	2	Liverpool Lime Street	Manchester Piccadilly	1/1/2024 0:00	0.4375	0.458333333	0.458333333	On Time	No	
940-7228-4lef-9eae	12/31/2023 0:00	0.565034722	Online	Credit Card	Senior	Standard	Advance	23	London Kings Cross	York	1/1/2024 0:00	0.395833333	0.472222222	0.472222222	On Time	No	
0c-666b-4925-86b6	12/31/2023 0:00	0.599409722	Station	Contactless	None	Standard	Advance	7	London Euston	Birmingham New Street	1/1/2024 0:00	0.53125	0.586805556	0.586805556	On Time	No	
34-7056-441c-9191	12/31/2023 0:00	0.620648148	Online	Credit Card	None	Standard	Advance	7	London Euston	Birmingham New Street	1/1/2024 0:00	0.552083333	0.607638889	0.607638889	On Time	No	
93c-b820-42dc-ad4f	12/31/2023 0:00	0.63880787	Online	Contactless	None	Standard	Advance	86	Manchester Piccadilly	London Paddington	1/1/2024 0:00	0.572916667	0.666666667	0.666666667	On Time	No	
231-5241-46f4-8328	12/31/2023 0:00	0.662337963	Online	Credit Card	Senior	First Class	Advance	34	Oxford	Bristol Temple Meads	1/1/2024 0:00	0.59375	0.645833333	0.6625	Delayed	Signal Failure	Yes
284-5f19-4lef-8350	12/31/2023 0:00	0.697280093	Station	Credit Card	None	Standard	Advance	13	London Paddington	Reading	1/1/2024 0:00	0.625	0.666666667	0.666666667	On Time	No	
a4c-7ba3-45e9-a9cd	12/31/2023 0:00	0.751365741	Online	Contactless	None	Standard	Advance	7	London Euston	Birmingham New Street	1/1/2024 0:00	0.6875	0.743055556	0.743055556	On Time	No	
8b1-ee1d-4a6f-8c29	12/31/2023 0:00	0.777083333	Online	Credit Card	None	First Class	Advance	57	London Kings Cross	York	1/1/2024 0:00	0.708333333	0.784722222	0.784722222	On Time	No	
61-b648-41c2-8151	12/31/2023 0:00	0.814756944	Station	Debit Card	None	Standard	Advance	8	London St Pancras	Birmingham New Street	1/1/2024 0:00	0.739583333	0.795138889	0.795138889	On Time	No	
9b5-3ab7-4d40-a949	12/31/2023 0:00	0.823530093	Online	Credit Card	None	Standard	Advance	13	London Paddington	Reading	1/1/2024 0:00	0.760416667	0.802083333	0.802083333	On Time	No	
e71-2e6f-450e-90ed	12/31/2023 0:00	0.847511574	Station	Contactless	None	Standard	Advance	35	London Kings Cross	York	1/1/2024 0:00	0.739583333	0.815972222	0.815972222	On Time	No	
87-d2ad-47be-8cdb	12/31/2023 0:00	0.848587963	Online	Credit Card	None	Standard	Advance	5	Birmingham New Street	Tamworth	1/1/2024 0:00	0.78125	0.795138889	0.795138889	On Time	No	
b1b-bf7f-43d9-b63c	1/1/2024 0:00	0.048842593	Station	Contactless	None	Standard	Off-Peak	10	London Euston	Birmingham New Street	1/1/2024 0:00	0.979166667	0.034722222	0.034722222	On Time	No	
3cf-382b-4794-aa8c	1/1/2024 0:00	0.053009259	Online	Contactless	None	Standard	Off-Peak	4	Manchester Piccadilly	Liverpool Lime Street	1/1/2024 0:00	0.114583333	0.135416667	0.135416667	On Time	No	

Windows Go to Settings to activate Windows

railway

+





Analysis Questions



1

Why does train ticket revenue change each month, and are there certain months or seasons when sales are higher or lower?

2

What causes low ticket sales in some months, and how can we increase revenue during those times?

3

Which train stations or routes generate the highest and lowest revenue, and what strategies can be used to improve the performance of low-revenue stations or routes?

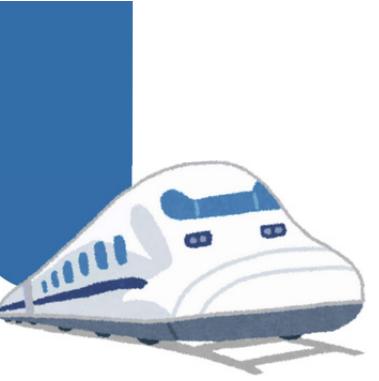
4

What factors—such as location, ticket type, or payment method—contribute to revenue differences between train stations?

5

What are the most common causes of journey delays, cancellations, and refund requests—and how can we reduce them?





Data Cleaning



Import the data



```
[ ] #import the railway.csv
from google.colab import files
uploaded = files.upload()
```

Import libraries, and read the data



```
#import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv('railway.csv')
df.head()
```

	Transaction ID	Date of Purchase	Time of Purchase	Purchase Type	Payment Method	Railcard	Ticket Class	Ticket Type	Price	Departure Station	Arrival Destination	Date of Journey	Departure Time	Arrival Time	Actual Arrival Time	Journey Status	Reason for Delay	Refund Request
0	da8af9ba8- b3dc4677- b176	2023-12- 08	12:41:11	Online	Contactless	Adult	Standard	Advance	43	London Paddington	Liverpool Lime Street	2024-01-01	11:00:00	13:30:00	13:30:00	On Time	Nan	
1	b0cdd1b0- f144-4197- be53	2023-12- 16	11:23:01	Station	Credit Card	Adult	Standard	Advance	23	London Kings Cross	York	2024-01-01	09:45:00	11:35:00	11:40:00	Delayed	Signal Failure	
2	f3ba9a6- f713-40d9- 9629	2023-12- 19	19:51:27	Online	Credit Card	Nan	Standard	Advance	3	Liverpool Lime Street	Manchester Piccadilly	2024-01-02	18:15:00	18:45:00	18:45:00	On Time	Nan	
3	b247111- 4fe7-4cd7- 8ab4	2023-12- 20	23:09:36	Station	Credit Card	Nan	Standard	Advance	13	London Paddington	Reading	2024-01-01	21:30:00	22:30:00	22:30:00	On Time	Nan	
4	2be00b45- 0762-485e- 72	2023-12- 27	18:22:56	Online	Contactless	Nan	Standard	Advance	76	Liverpool Lime Street	London Euston	2024-01-01	16:45:00	19:00:00	19:00:00	On Time	Nan	

check the data



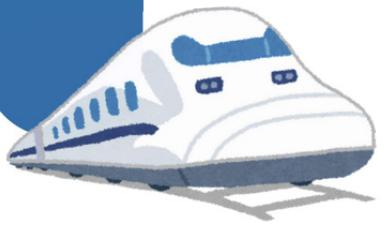
```
df.sample(10) #check the data
```

	Transaction ID	Date of Purchase	Time of Purchase	Purchase Type	Payment Method	Railcard	Ticket Class	Ticket Type	Price	Departure Station	Arrival Destination	Date of Journey	Departure Time	Arrival Time	Actual Arrival Time	Journey Status	Reason for Delay	Refund Request
401	29151685-57b- 4720-5a12	2024-01-03	09:34:38	Online	Credit Card	Disabled	Standard	Advance	48	London Euston	Manchester Piccadilly	2024-01-04	08:00:00	09:50:00	09:50:00	On Time	Nan	
11427	27d5dc96-6783- 4229-a808	2024-02-08	04:28:29	Online	Credit Card	Nan	Standard	Advance	3	Liverpool Lime Street	Manchester Piccadilly	2024-02-22	02:45:00	03:15:00	03:15:00	On Time	Nan	
11006	b94794a1-ac19- 437a-4055	2024-02-07	03:15:20	Online	Contactless	Nan	Standard	Advance	8	York	Durham	2024-02-08	01:45:00	02:35:00	02:35:00	On Time	Nan	
4490	4cc0ff79-9737- 4041-46d0	2024-01-18	03:54:49	Station	Credit Card	Nan	Standard	Advance	3	Liverpool Lime Street	Manchester Piccadilly	2024-01-19	02:15:00	02:45:00	02:45:00	On Time	Nan	
29058	56241315-c8d- 4043-4036	2024-04-20	20:29:11	Station	Debit Card	Nan	Standard	Advance	35	London Kings Cross	York	2024-04-21	18:45:00	20:35:00	20:35:00	On Time	Nan	
28232	7a6196ac-8251- 4006-6055	2024-04-17	21:37:00	Station	Credit Card	Nan	Standard	Advance	13	London Paddington	Reading	2024-04-18	20:00:00	21:00:00	21:00:00	On Time	Nan	
3070	3919ab12-9460- 4915-4414	2024-01-16	09:07:52	Online	Credit Card	Nan	First Class	Advance	57	London Kings Cross	York	2024-01-17	07:30:00	09:20:00	09:20:00	On Time	Nan	
20216	93e2021-8fb6- 4583-8007	2024-03-18	02:59:43	Station	Credit Card	Nan	Standard	Off-Peak	12	London St Pancras	Birmingham New Street	2024-03-18	04:15:00	05:35:00	05:35:00	On Time	Nan	
6218	56e30a1a-e0d- 4008-430	2024-01-24	06:05:17	Station	Contactless	Nan	Standard	Advance	8	London St Pancras	Birmingham New Street	2024-01-25	04:30:00	05:50:00	05:50:00	On Time	Nan	
26444	014440d7-7816- 4226-4648	2024-04-11	09:34:51	Station	Credit Card	Nan	Standard	Advance	76	Liverpool Lime Street	London Euston	2024-04-12	08:00:00	10:15:00	11:15:00	Delayed	Weather	

Activate Windows
Go to Settings to activate Windows.

Data Cleaning

بيانات الاتصالات
بنكنا له جياب المعلومات



Check data type, nulls, and duplicate data



```
print(df.isnull().sum()) # missing values
```

Column	Non-Null Count	Dtype
Transaction ID	0	object
Date of Purchase	0	object
Time of Purchase	0	object
Purchase Type	0	object
Payment Method	0	object
Railcard	20918	object
Ticket Class	0	object
Ticket Type	0	object
Price	0	int64
Departure Station	0	object
Arrival Destination	0	object
Date of Journey	0	object
Departure Time	0	object
Arrival Time	0	object
Actual Arrival Time	1880	object
Journey Status	0	object
Reason for Delay	27481	object
Refund Request	0	object
dtype:	int64(1), object(17)	

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31653 entries, 0 to 31652
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Transaction ID  31653 non-null   object 
 1   Date of Purchase 31653 non-null   object 
 2   Time of Purchase 31653 non-null   object 
 3   Purchase Type    31653 non-null   object 
 4   Payment Method   31653 non-null   object 
 5   Railcard          10735 non-null   object 
 6   Ticket Class     31653 non-null   object 
 7   Ticket Type      31653 non-null   object 
 8   Price             31653 non-null   int64  
 9   Departure Station 31653 non-null   object 
 10  Arrival Destination 31653 non-null   object 
 11  Date of Journey  31653 non-null   object 
 12  Departure Time   31653 non-null   object 
 13  Arrival Time    31653 non-null   object 
 14  Actual Arrival Time 29773 non-null   object 
 15  Journey Status   31653 non-null   object 
 16  Reason for Delay 4172 non-null    object 
 17  Refund Request   31653 non-null   object 
dtypes: int64(1), object(17)
```

Describe the data, and detect the outliers

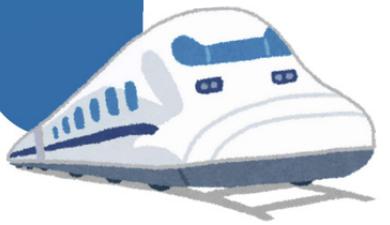


```
print(df.describe())
```

	Price
count	31653.000000
mean	23.439200
std	29.997628
min	1.000000
25%	5.000000
50%	11.000000
75%	35.000000
max	267.000000

```
def detect_outliers_iqr(data, column):
    Q1 = data[column].quantile(0.25)
    Q3 = data[column].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR
    outliers = data[(data[column] < lower) | (data[column] > upper)]
    return outliers
```

Data Cleaning



Convert time columns to time format



```
[ ] # Convert time columns to time format
time_columns = ["Time of Purchase", "Departure Time", "Arrival Time", "Actual Arrival Time"]
for col in time_columns:
    df[col] = pd.to_datetime(df[col], errors="coerce").dt.time
```

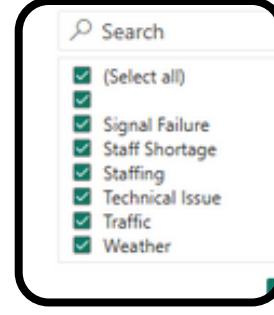
Time of Purchase
0.528599537
0.47431713
0.827395833
0.95875
0.765925926
0.330625

Time of Purchase
8:22:05 PM
8:18:10 PM
8:20:20 PM
8:21:59 PM
8:19:31 PM
8:20:58 PM
8:24:24 PM
8:19:49 PM

Standardize the columns by merging similar entries



```
# Standardize categorical columns (strip spaces and convert to title case where applicable)
categorical_columns = ["Purchase Type", "Payment Method", "Railcard", "Ticket Class", "Ticket Type",
                      "Departure Station", "Arrival Destination", "Journey Status", "Refund Request"]
df[categorical_columns] = df[categorical_columns].apply(lambda x: x.str.strip().str.title())
```



Data Cleaning

جامعة الاتصالات
بتكنولوجيا المعلومات

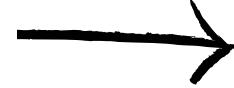


Detect outliers using IQR, and remove them

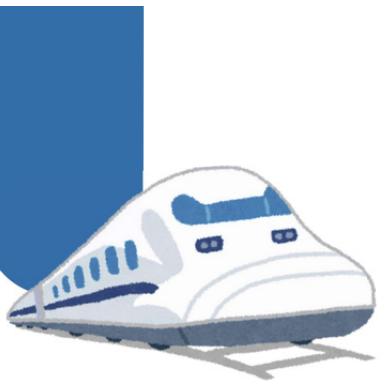


```
for col in columns_to_check:  
    outliers = detect_outliers_iqr(df, col)  
    print(f"\nOutliers in '{col}': {len(outliers)} rows")  
  
df = df[~df.index.isin(outliers.index)]
```

Handle null values (drop rows with any NaNs)



```
print("\nNull values before dropping:")  
print(df.isnull().sum())  
  
df.dropna(inplace=True)  
  
print("\nNull values after dropping:")  
print(df.isnull().sum())
```



The Data Source After Cleaning

بيانات الاتصالات
بتكنولوجيا المعلومات

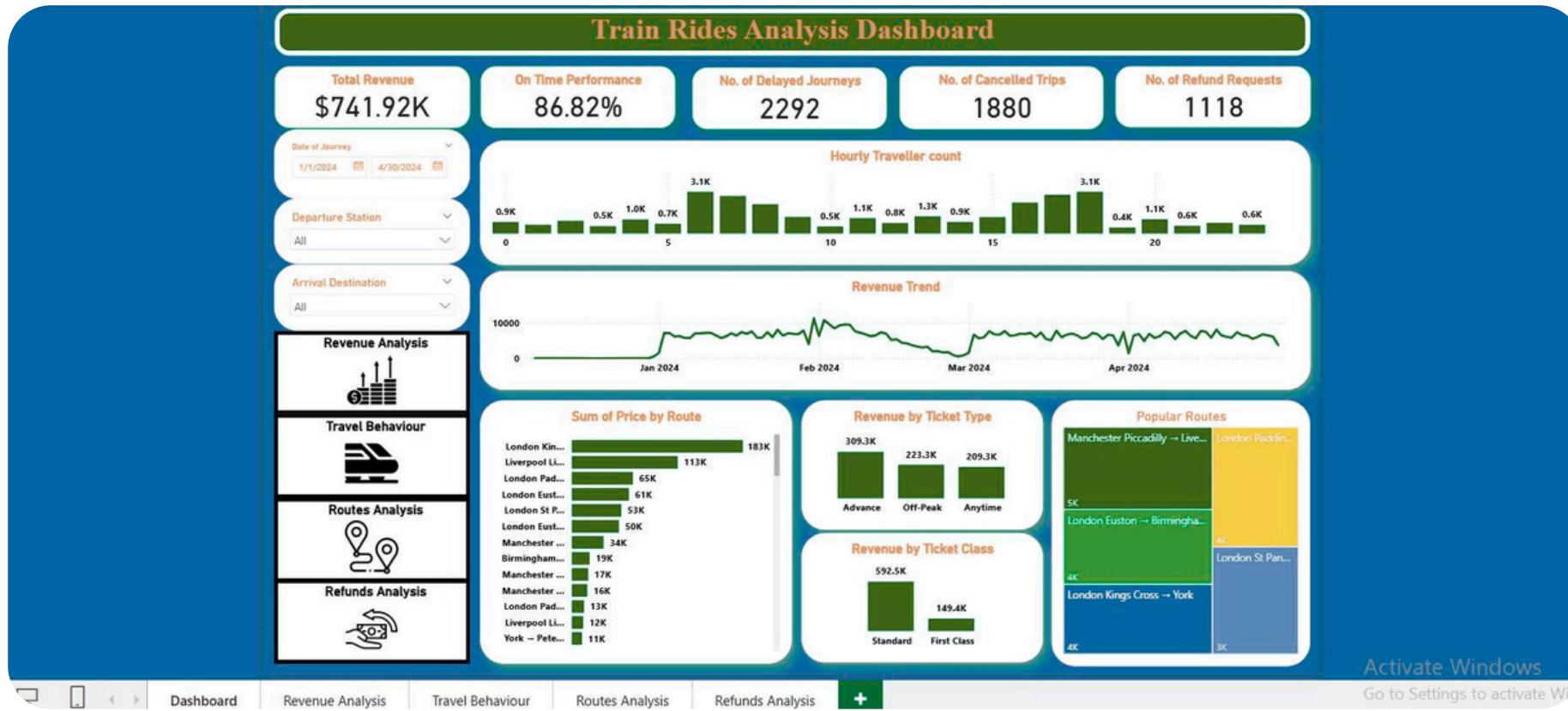
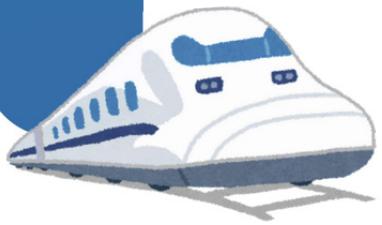


```
re(10) # check data after cleaning
```

transaction ID	Date of Purchase	Time of Purchase	Purchase Type	Payment Method	Railcard	Ticket Class	Ticket Type	Price	Departure Station	Arrival Destination	Date of Journey	Departure Time	Arrival Time	Actual Arrival Time	Journey Status	Reason for Delay
590e9209-0524-455e-9154	2024-02-03	04:19:04	Online	Credit Card	Adult	Standard	Advance	3	Liverpool Lime Street	Manchester Piccadilly	2024-02-04	02:45:00	03:15:00	03:15:00	On Time	No Delay
e20a3aa5-cba0-4621-84cf	2024-03-02	20:32:24	Station	Contactless	Adult	Standard	Advance	35	London Kings Cross	York	2024-03-03	17:45:00	19:35:00	19:37:00	Delayed	Signal failure
cc975841-caa5-4f05-8ea1	2024-02-14	18:54:36	Online	Contactless	Disabled	Standard	Advance	8	London Paddington	Reading	2024-02-15	17:15:00	18:15:00	18:15:00	On Time	No Delay
f270ade3-a2bc-42e3-800b	2024-03-19	08:04:06	Online	Credit Card	Adult	Standard	Advance	3	Manchester Piccadilly	Liverpool Lime Street	2024-03-20	06:30:00	07:00:00	07:00:00	On Time	No Delay
16172ecc-7ed3-43d9-8dbe	2024-01-27	20:06:28	Station	Credit Card	Adult	Standard	Off-Peak	10	London Euston	Birmingham New Street	2024-01-27	21:30:00	22:50:00	22:50:00	On Time	No Delay
e4a941ae-931a-4c37-8a2c	2024-03-14	09:07:52	Online	Credit Card	Adult	Standard	Off-Peak	10	London Euston	Birmingham New Street	2024-03-14	10:30:00	11:50:00	11:50:00	On Time	No Delay
341e09e2-5f03-42e1-b456	2024-04-13	14:49:38	Online	Contactless	Adult	Standard	Off-Peak	10	London Euston	Birmingham New Street	2024-04-13	16:15:00	17:35:00	17:35:00	On Time	No Delay

Data Analysis

وزارة الاتصالات
وتكنولوجيا المعلومات

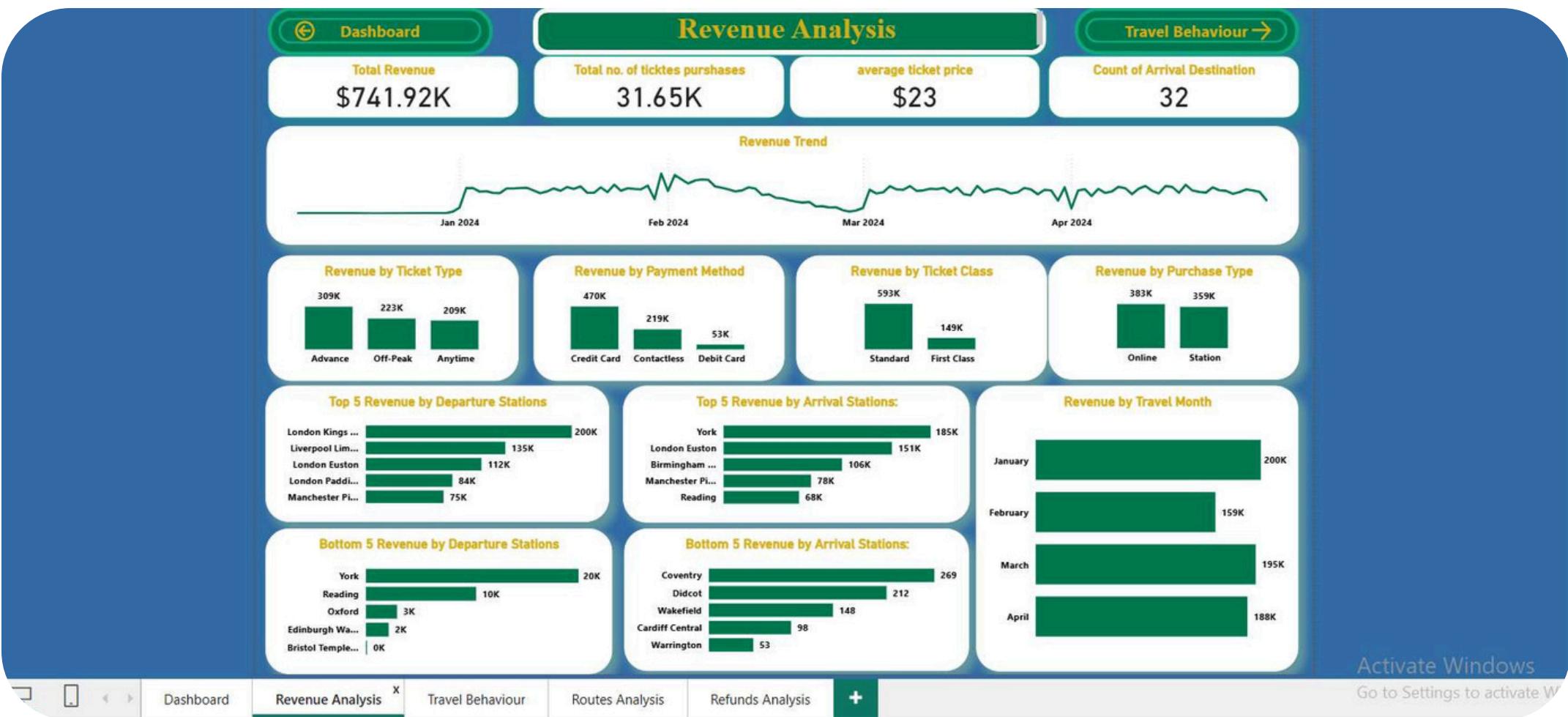
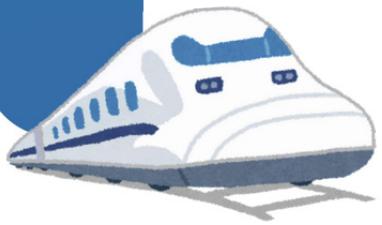


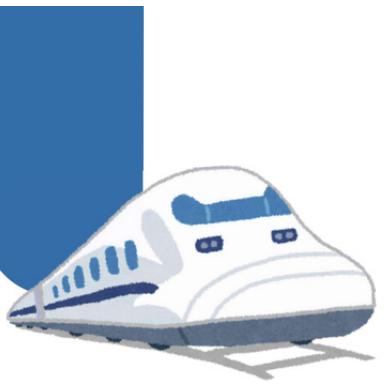
The entire analysis was subdivided into four main sections: Revenue Analysis, Travel Behaviours, Routes Analysis, and Refunds Analysis. Each section included various visualizations that provided insights into different aspects of the railway operations. Below is an elaboration on the detailed analysis, including the interpretation of results and useful insights derived from each visualization.



Revenue Analysis

بيانات الاتصالات
بتكنولوجيا المعلومات





Revenue Analysis

Total Revenue	Total no. of ticktes purshases	average ticket price	Count of Arrival Destination
\$741.92K	31.65K	\$23	32

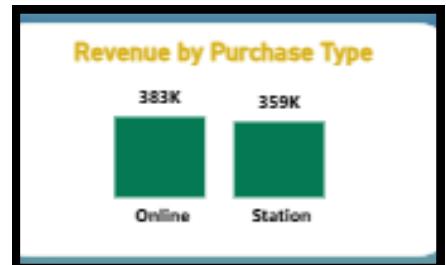
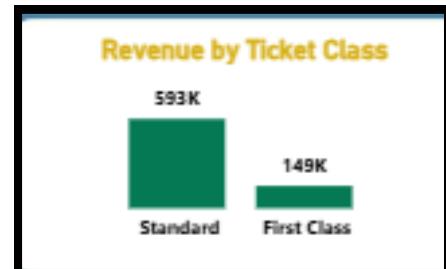
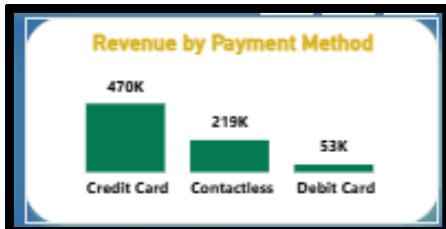
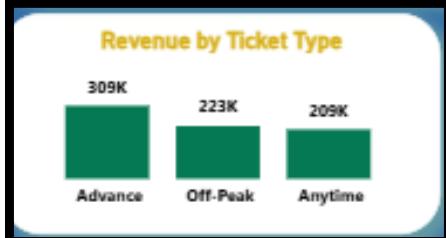


Using new measures and represented as cards, it was found that between January and April 2024, the railway generated a total revenue of **£741,921** from **31,653** ticket purchases. The average ticket price during this period was **£23**, reflecting customer spending per transaction. The network size included **32** train stations, indicating a broad service area that potentially attracted more customers.



The highest revenue was recorded in **January 2024**, while **the lowest** was in **February 2024**. Fluctuations in monthly revenue could be due to seasonal variations, special events, or economic conditions. January's high revenue might be due to post-holiday travel, while February's dip could indicate a low travel season or economic downturn.

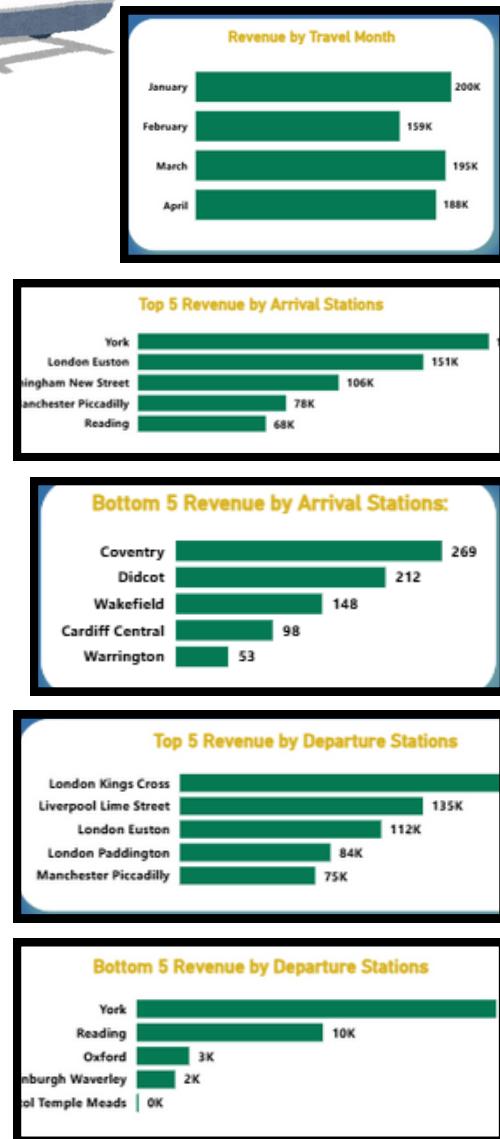
Revenue Analysis



- Advance tickets generated the **most revenue**, followed by Off-Peak and Anytime tickets. **Advance tickets being the highest revenue contributor** suggests that customers prefer planning their trips in advance, likely due to lower prices or availability of discounts. Off-Peak tickets also contribute significantly, indicating a preference for travel during less busy times.
- Credit cards** were the **most popular payment method**, followed by contactless payments and debit cards. **The popularity of credit cards and contactless payments** suggests a customer preference for quick and convenient payment options. This information can be useful for tailoring payment processing services.
- Standard class tickets** generated more revenue than First Class tickets. This indicates a higher demand for Standard class travel, possibly due to its affordability. The relatively **lower revenue from First Class** might suggest that premium pricing needs to be re-evaluated to attract more customers.
- Both online and station purchases are almost equal, with **online purchases slightly higher**. This suggests that while a significant portion of customers still prefer buying tickets at stations, there is a growing trend towards online purchases, indicating the importance of maintaining and enhancing digital sales channels.

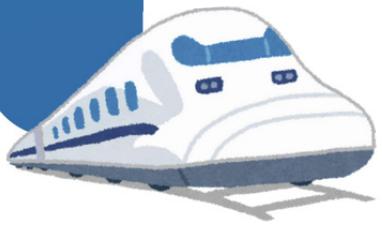


Revenue Analysis

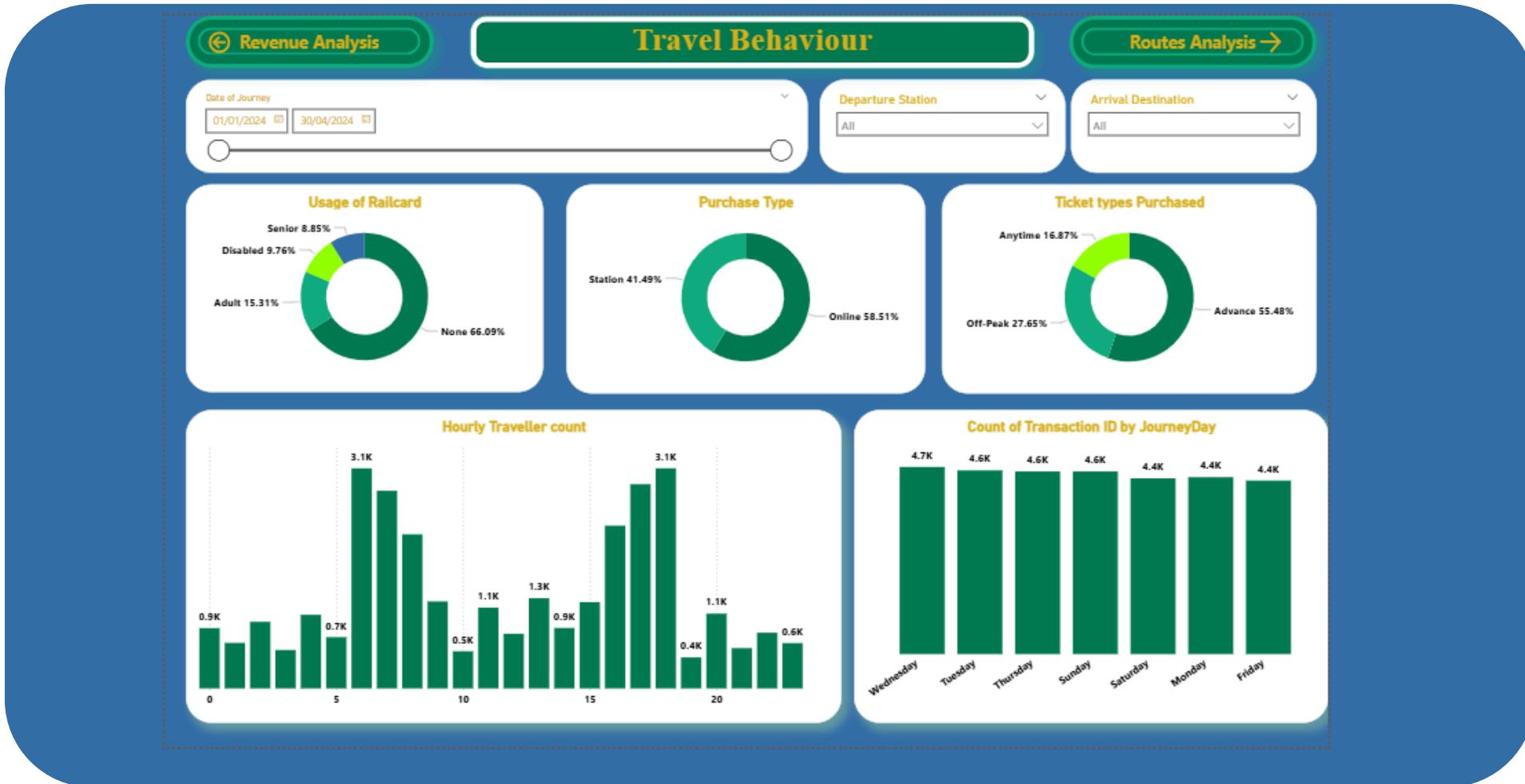


- Data reveals fluctuations throughout the period from January to April 2024. **January** generated the **highest revenue** at £200k, followed by **March** with £195k, **April** with £188k, and **February** with the **lowest revenue** at £159k. These variations highlight the need for targeted marketing and pricing strategies to maximize revenue across different times of the year.
- York, London Euston, and Birmingham New Street** are major destinations, reflecting their importance as travel hubs or key business/tourist locations.
- Low revenue** at stations like **Coventry**, and **Didcot** suggests these destinations are less popular. Analysing factors such as local attractions, business hubs, or the quality of services might help in formulating strategies to boost their appeal.
- The **high revenue** at **London Kings Cross**, **Liverpool Lime Street**, and **London Euston** indicates these are major starting points for many passengers, likely due to their strategic locations in major cities and connectivity to important routes.
- Stations like **Bristol Temple Meads**, **Edinburgh Waverley**, and **Oxford** generate the **least revenue**, suggesting these locations either have fewer travellers starting their journeys there or the ticket prices are lower. It may be beneficial to investigate why these stations have low revenue and explore strategies to increase their usage, such as improving connectivity, services, or marketing.

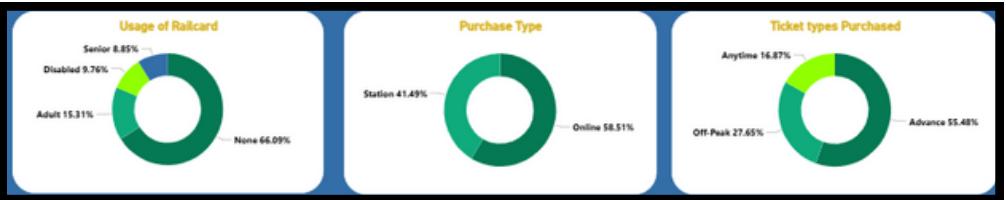
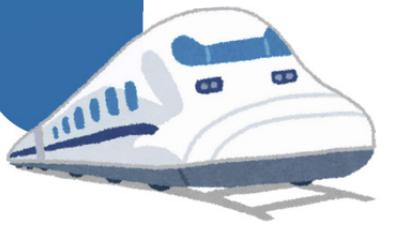
Travel Behaviour



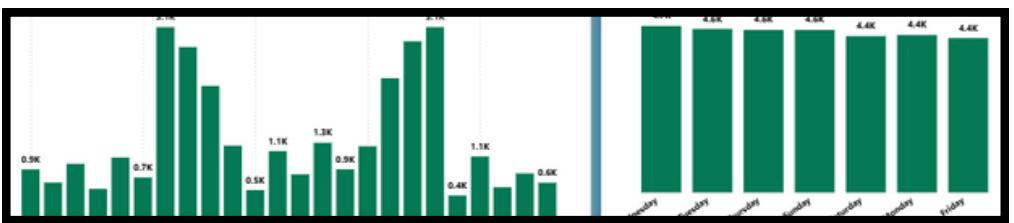
بيانات الاتصالات
بتكنولوجيا المعلومات



Travel Behaviour

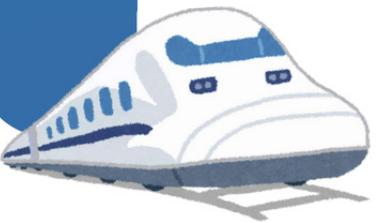


- The detailed breakdown shows that **adult railcards** are **the most used**, followed by disabled and senior railcards.
- 58.51% of tickets were **purchased online**, while 41.49% were bought at stations. The preference for online purchases highlights **the importance** of a **robust and user-friendly** online ticketing system.
- **Advance tickets (55.48%)** are **the most purchased**, followed by Off-Peak (27.65%) and Anytime (16.87%). The high percentage of Advance ticket purchases indicates that customers are price-sensitive and prefer to plan ahead to get the best deals.
- **Peak travel times** are **6 AM and 6 PM**. These peak times correspond to typical commuting hours, suggesting that a significant portion of travellers are daily commuters. This information can be used to optimize train schedules and manage crowding during peak hours.
- Ticket purchases and travel are fairly consistent throughout the week, with **Wednesday** being **slightly higher**. Consistent travel patterns throughout the week suggest a steady demand for services. The slight peak on Wednesdays might be due to mid-week business travel or other recurring events.



Routes Analysis

بيانات الاتصالات
بتكنولوجيا المعلومات



Travel Behaviour

Routes Analysis

Refunds Analysis →

Date of Journey
1/1/2024 4/30/2024

No. of Delayed Journeys: 2292

No. of Cancelled Trips: 1880

Top departure stations:

Station	Count
Manchester Piccadilly	5.7K
London Euston	5.0K
Liverpool Lime Street	4.6K
London Paddington	4.5K
London Kings Cross	4.2K

Top Arrival stations:

Station	Count
Birmingham New Street	7.7K
Liverpool Lime Street	5.0K
York	4.0K
Manchester Piccadilly	4.0K
Reading	3.9K

Least Arrival stations:

Station	Count
Coventry	65
Didcot	48
Cardiff Central	16
Wakefield	15
Warrington	15

Routes with most delay:

Route	Count of Reason for Delay
Liverpool Lime Street → London Euston	879
Manchester Piccadilly → Liverpool Lime Street	644
London Euston → Birmingham New Street	453
London Paddington → Reading	352
London Kings Cross → York	329

Routes with most cancellation:

Route	Cancelled Journeys
Manchester Piccadilly → Liverpool Lime Street	290
London Paddington → Reading	286
London St Pancras → Birmingham New Street	273
London Euston → Birmingham New Street	211
London Kings Cross → York	198

Popular Routes:

Route	Count
Manchester Piccadilly → Liverpool Lime Street	5K
London Euston → Birmingham New Street	4K
London King's Cross → York	4K

Cancelled Journeys by Reason for Delay:

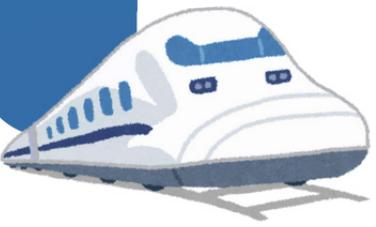
Reason	Count
Signal Failure	519
Weather	445
Staffing	238
Technical Issue	235

Reasons for delays:

Reason	Count
Weather	1372
Signal Failure	970
Technical Issue	707
Staffing	410
Staff Shortage	399
Traffic	314

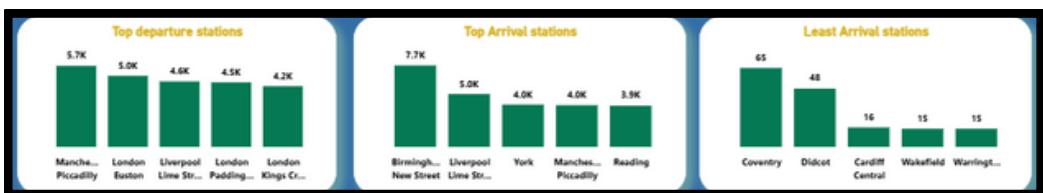
Activate Windows
Go to Settings to activate

Routes Analysis

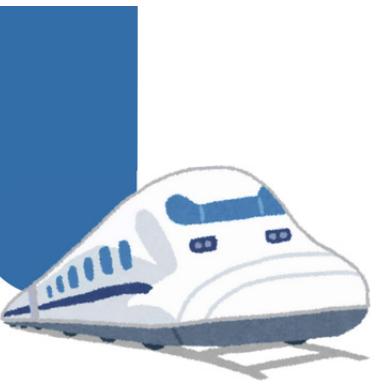


No. of Delayed Journeys	No. of Cancelled Trips
2292	1880

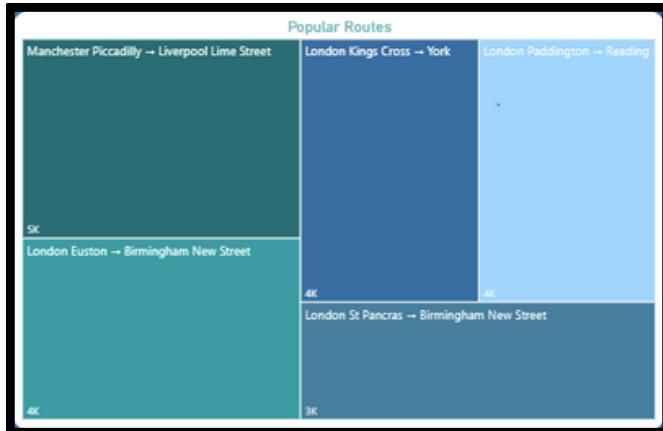
- Total delay 2,292 delayed journeys and 1,880 cancellations. The data on delays and cancellations highlights operational challenges. Frequent delays and cancellations can significantly impact customer satisfaction and operational efficiency. Identifying the reasons and specific routes with frequent delays and cancellations helps in targeting improvements and mitigating issues.



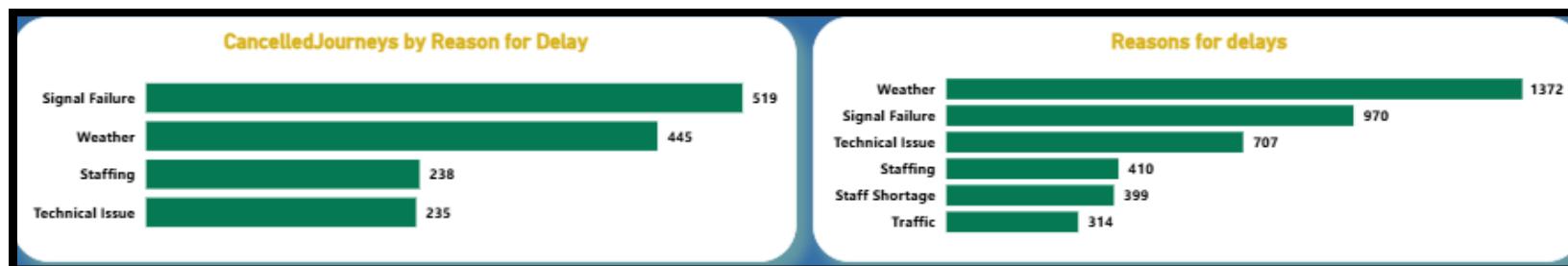
- Manchester Piccadilly, London Euston, Liverpool Lime Street, London Paddington and London Kings Cross are the major departure points, reflecting high passenger traffic and significant revenue potential. Focusing on enhancing services, facilities, and connectivity at these stations can further boost customer satisfaction and operational efficiency.
- Birmingham New Street, Liverpool Lime Street, York, Manchester Piccadilly and Reading are the top arrival stations. The high arrival rates at these stations indicate popular destinations. Improving services and facilities at these stations can enhance the travel experience, attracting more passengers and boosting revenue.
- Coventry, Didcot, Cardiff Central, Wakefield and Warrington have low arrival rates, indicating less popularity. Investigating the causes, such as connectivity issues, lack of attractions, or poor facilities, can help in formulating strategies to increase their usage.



Routes Analysis



- These routes are highly frequented, indicating strong demand. Enhancing services, increasing train frequency, and offering promotional fares on these routes can capitalize on their popularity and boost revenue.

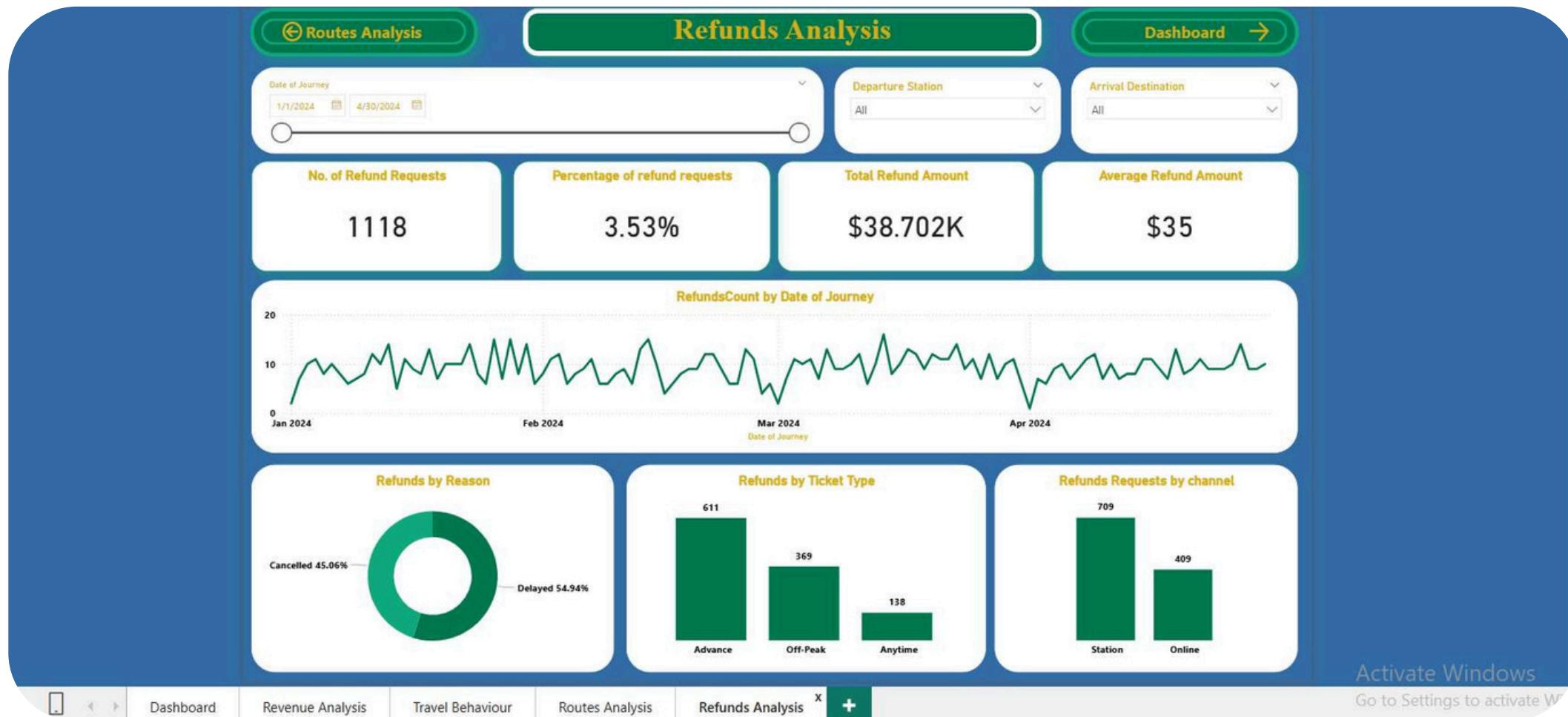


- Signal failure is the primary cause of cancellations, highlighting the need for robust signal systems and contingency plans. Addressing staff shortages and improving infrastructure can further reduce cancellations.

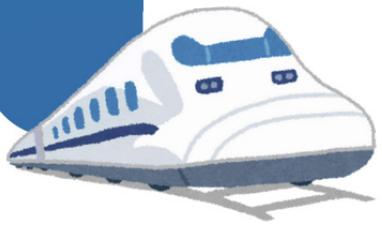
- Weather is the leading cause of delays, followed by Signal Failure.
- Implementing weather-resilient infrastructure, regular maintenance, and upgrading signal systems can mitigate these delays.

Refunds analysis

وزارة الاتصالات
وتكنولوجيا المعلومات



Refunds analysis

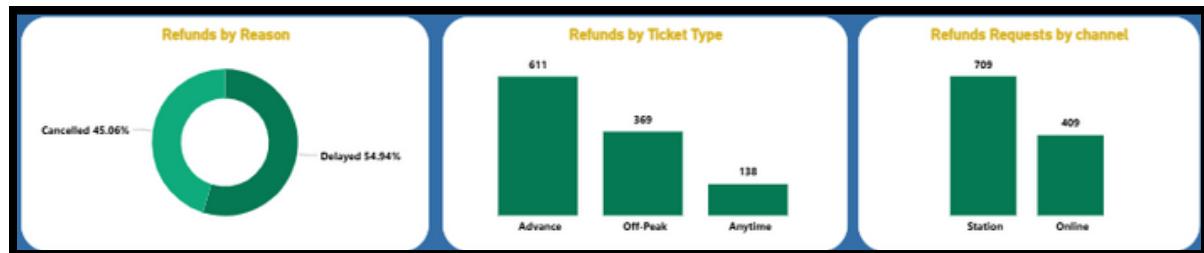


No. of Refund Requests	Percentage of refund requests	Total Refund Amount	Average Refund Amount
1118	3.53%	\$38.702K	\$35

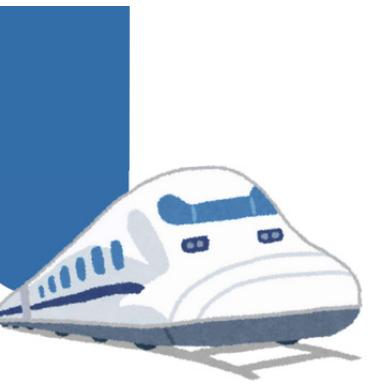
- There were **1,118 refund requests**, which is **3.53% of total travellers**, with a **total refund amount of £38,702** and an **average refund amount of £35**. The relatively low percentage of refund requests suggests **overall customer satisfaction**. However, the total refund amount indicates areas where service improvements are necessary to reduce refund claims.



- The refund trend is relatively uniform, with no significant peaks or troughs. **A uniform refund trend indicates consistent service performance**. Monitoring this trend helps in identifying any emerging issues that might lead to an increase in refunds.



- Delays (54.94%)** and **cancellations (45.06%)** are the main reasons for refunds. Focusing on reducing delays and cancellations can directly impact the number of refund requests, leading to higher customer satisfaction and reduced revenue loss from refunds.
- Advance tickets** have the **highest number** of refund requests, followed by **Off-Peak** and **Anytime tickets**. The high refund rate for Advance tickets might be due to the longer time between purchase and travel, increasing the likelihood of changes or cancellations. Offering flexible ticket options or insurance could mitigate this.

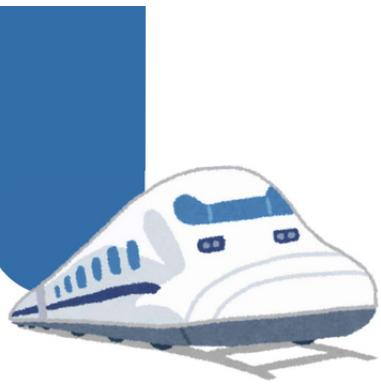


Key Findings



- Total Revenue reached £741.92K, mostly from Advance Tickets (309K).
- Peak travel hours are at 6 AM and 6 PM, indicating commuter-based usage.
- Online ticket purchases dominate with 58.51%, showing a strong shift to digital platforms.
- Standard class tickets generated significantly more revenue than First Class (593K vs 149K).
- Manchester Piccadilly and London Euston are among the top departure stations.
- Most delayed route: Liverpool Lime Street → London Euston with 879 delays.
- Top delay causes: Weather (1372), Signal Failure (970), and Technical Issues (707).
- Most cancellations: Manchester Piccadilly → Liverpool Lime Street.
- Refund requests make up 3.53% of total trips, mainly due to delays and cancellations.



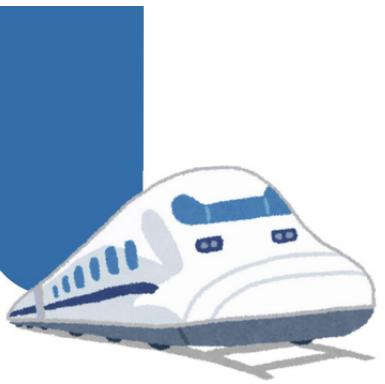


Recommendations



1. Promote Advance Ticket Sales further as they yield the highest revenue.
2. Optimize Peak Hours by adding more trains or increasing capacity at 6 AM and 6 PM.
3. Invest in Digital Experience — enhance the online booking system and mobile app.
4. Reduce Delays & Cancellations by:
 - Upgrading signaling systems.
 - Improving weather-resilient infrastructure.
 - Ensuring proper staffing and maintenance.
5. Target High Revenue Stations (e.g., London Kings Cross) for premium services.
6. Streamline Refund Process, especially online, to reduce congestion at stations.
7. Introduce Flexible Advance Tickets to lower the high refund rate on these tickets.





PROGRAMMING LANGUAGES + FRAMEWORKS



1. Programming Languages

- Python: For data cleaning, handling missing values, outliers, and transformations.

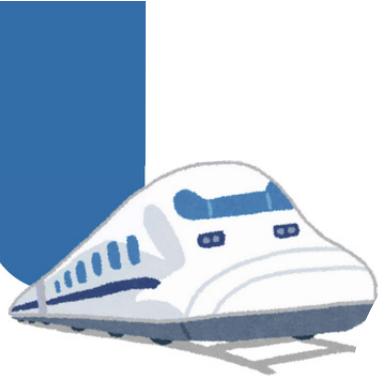
2. Tools & Frameworks

- Power BI: For building interactive dashboards and creating relationships between data tables.
- Excel: For initial data structuring and review before import into Power BI.

3. Supporting Technologies

- Power BI Data Model: Used to establish relationships and enable cross-table analysis.
- Power Query (in Power BI): used for further data shaping and filtering.

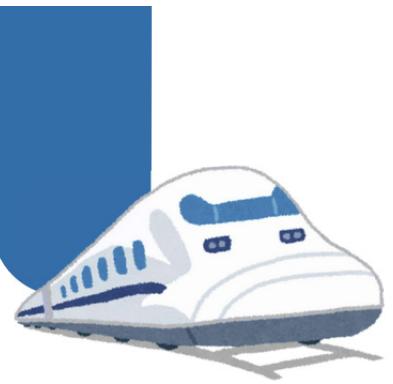




Here in this link, you will be able to view all the files we worked on and the report we created. To download the data, click on the link here.



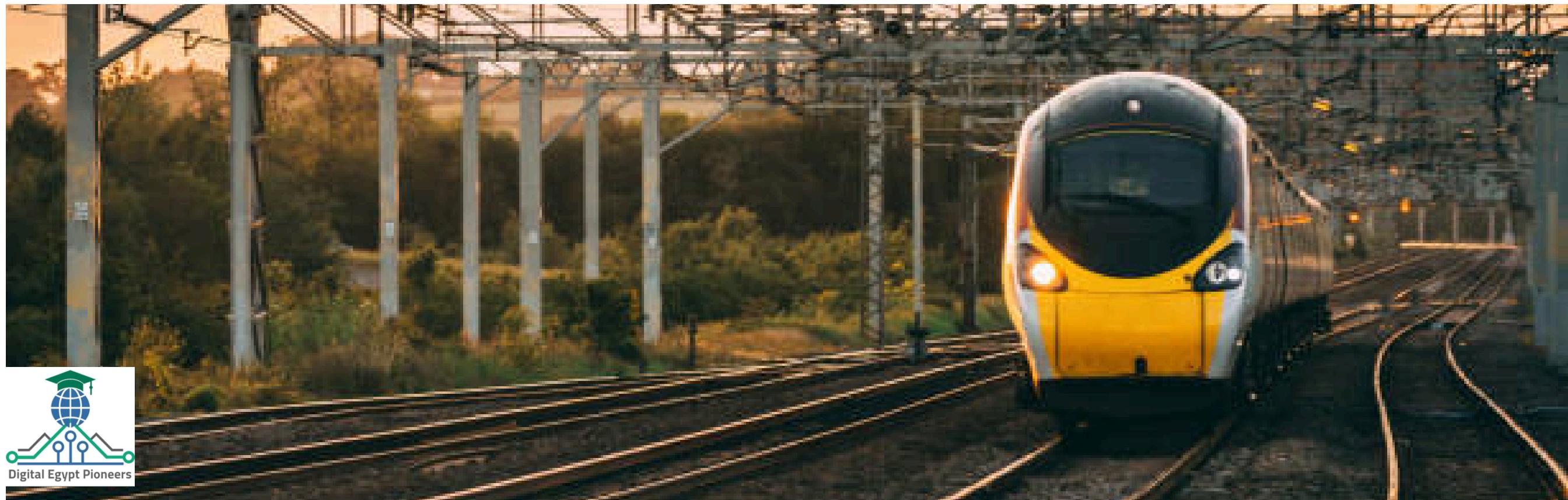
GitHub



وزارة الاتصالات
وتكنولوجيا المعلومات



Thank You



Digital Egypt Pioneers