

UK National Rail Analysis

Report by: Brianna Browning & Riley Sample

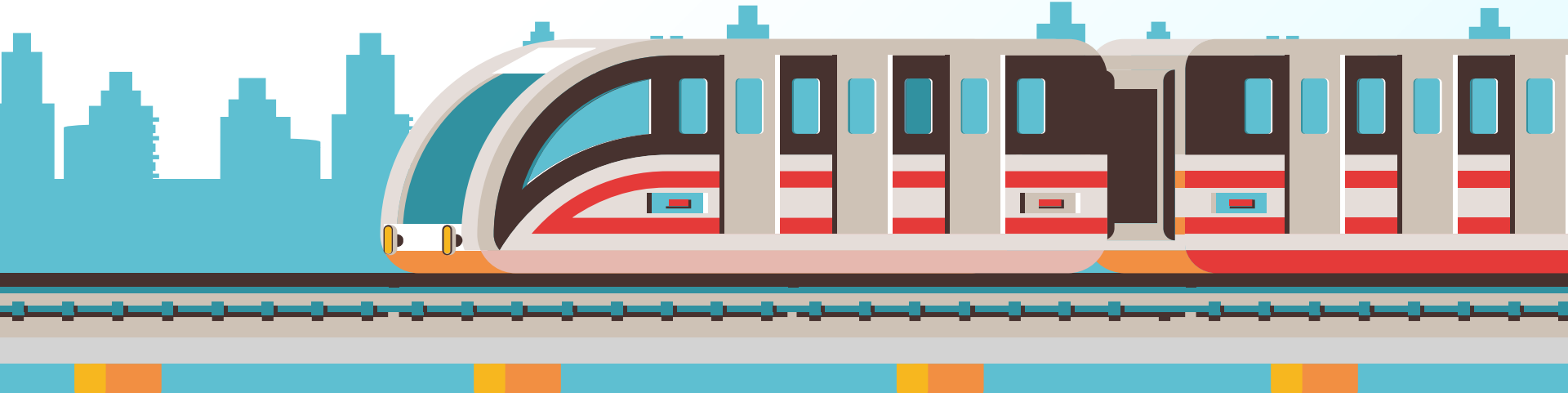


Table of contents

01

**Business
Problem**

02

Data

03

Analysis

04

**Recommendation
s**

STOP

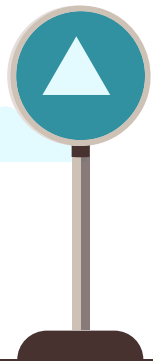


Introduction

The following data was pulled from the UK National Rail System, representing this past week's (7/10/24 - 7/17/24) worth of station, route, & train information.

The analysis is intended for scheduling prediction purposes, directed for the use of station, train, and regional operators.

* Please note that no data was collected on Saturday

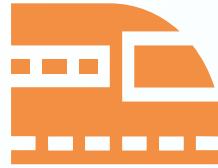


Business Problem

Are the UK train, station, and regional operators equipped to predict delays and cancellations?



Limited Delays &
Cancellations



Meets Passenger
Demand



Operational
Efficiency



Data

108,374

– Total Trains Running

7,841

– Delays

28,805

– Routes

2,356

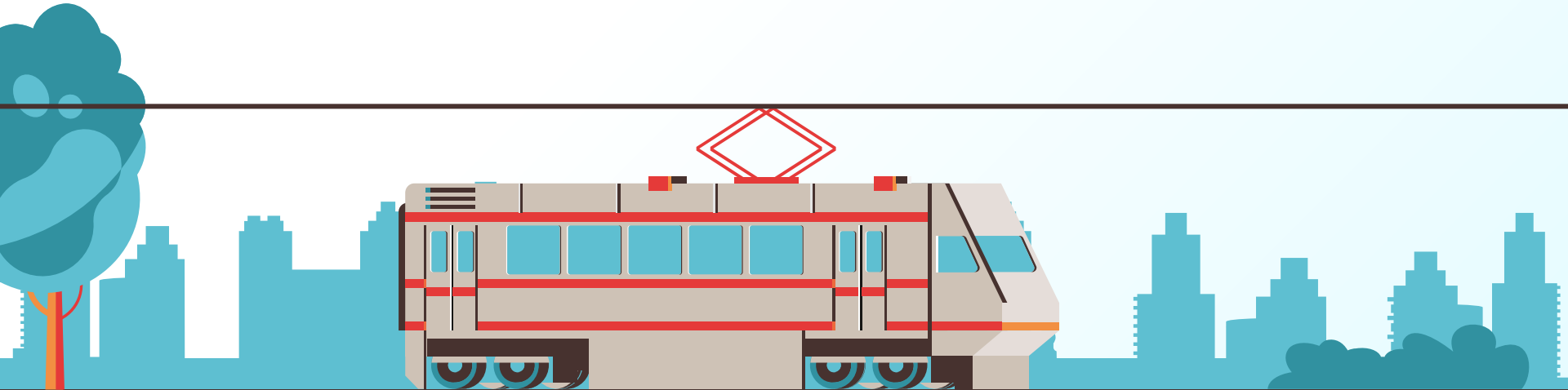
– Stations

5

– Regions Serviced



When are Trains Most Frequently Delayed?



When are Trains Most Frequently Delayed?



Departures

Most frequent during the evening (3pm-7pm).
Most trains run Tuesday & Wednesday



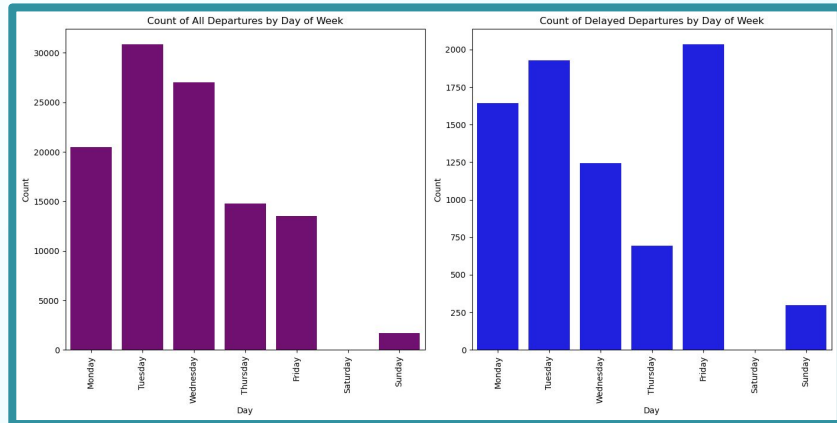
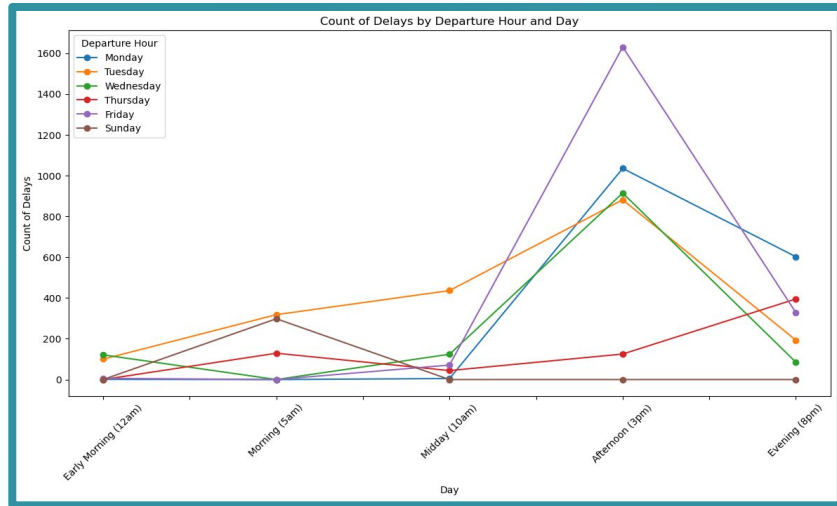
Hourly Delays

All delays are most frequent in the evenings, regardless of the day

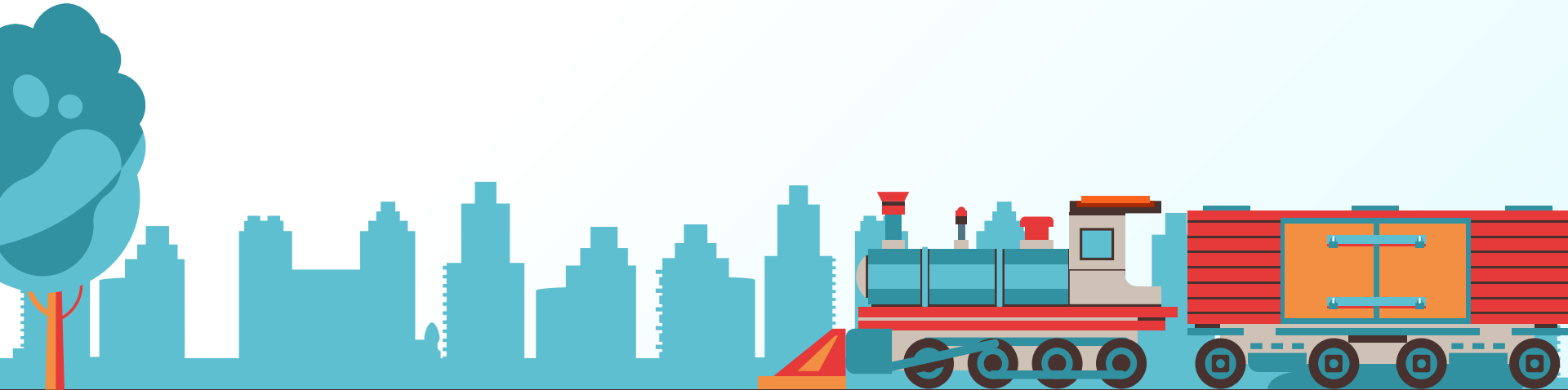


Daily Delays

Most delays occur on Friday & Tuesday



Do Train Cancellations Correlate with Departure Hour?



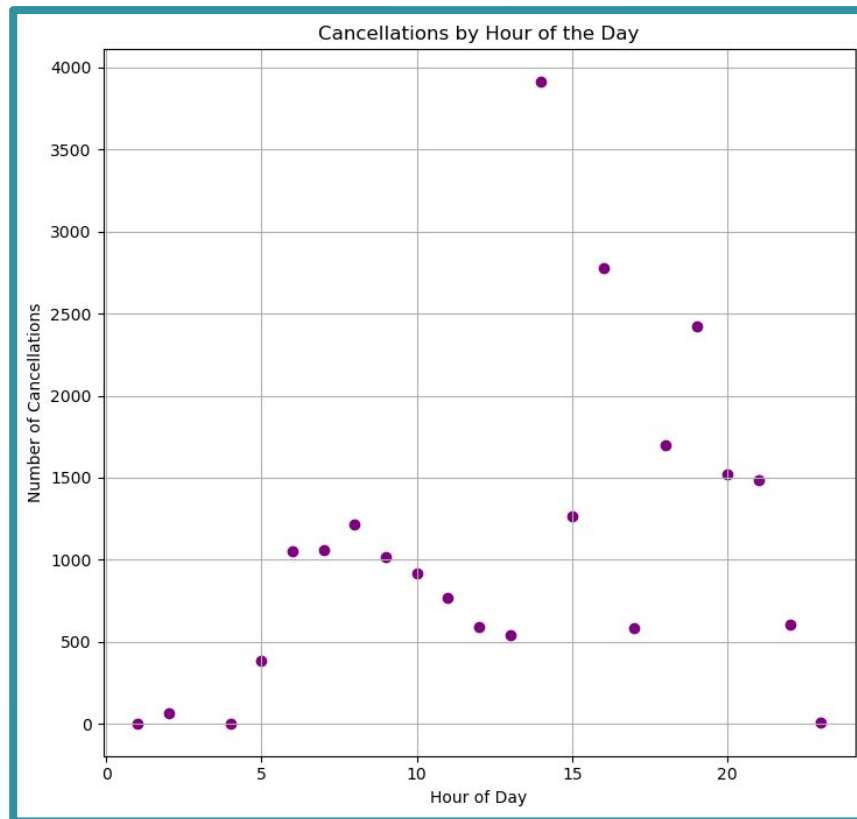
Do Train Cancellations Correlate with Departure Hour?

22.05 %

Proportion of trains cancelled

2.65 %

Proportion of delayed trains cancelled

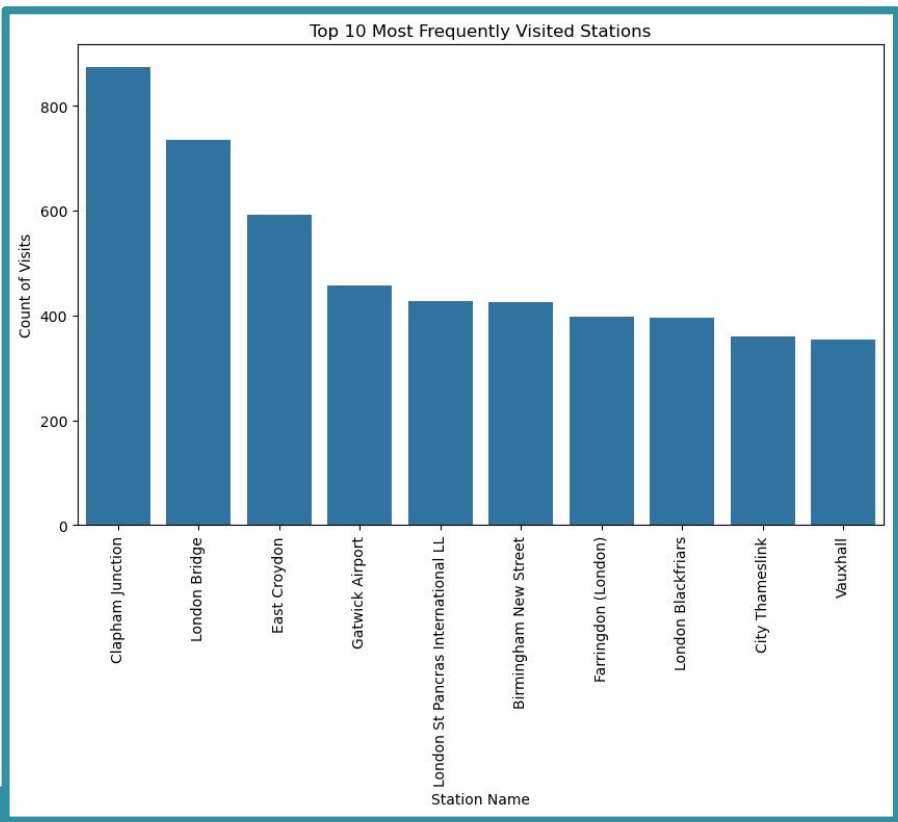


Correlation between departure hour and cancellations = -0.016

Do The Most Popular Stations Also Have The Most Delays?



Do The Most Popular Stations Also Have The Most Delays?



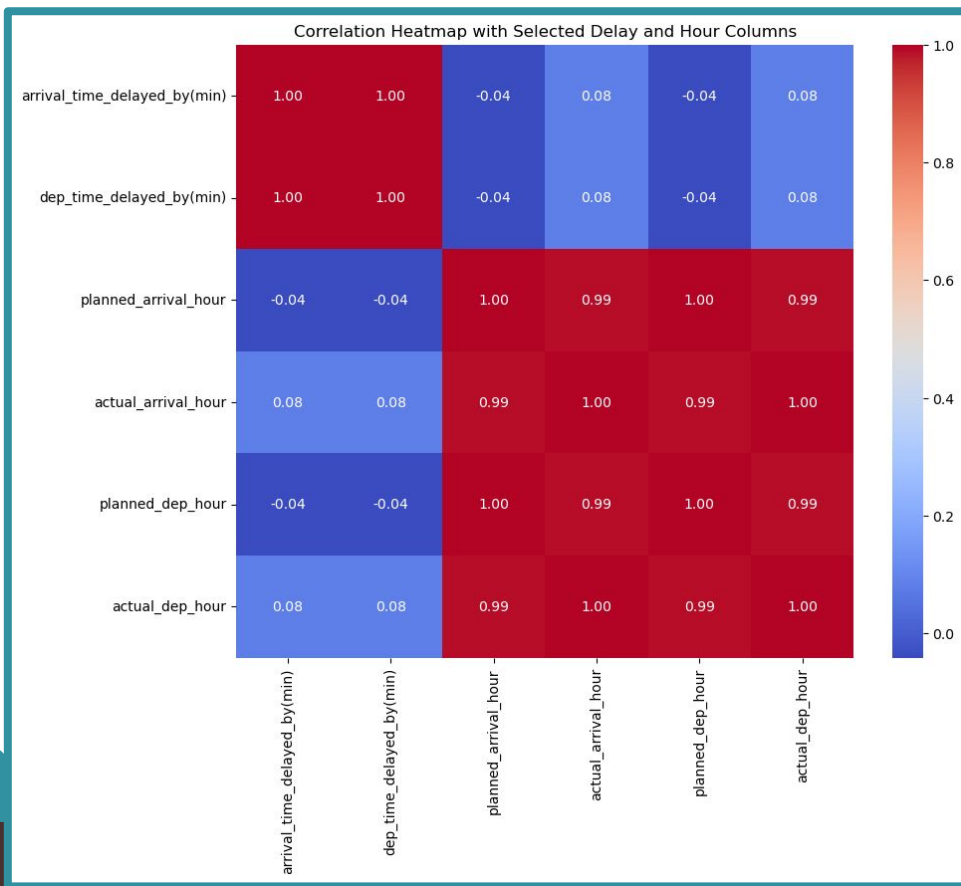
	StationName	is_delayed_departure	dep_time_delayed_by(min)
0	Harlesden	107	7249.0
1	Willesden Junction Low Level	107	7249.0
2	Cardiff Queen Street	101	3826.0
3	Sunderland	84	5484.0
4	Cardiff Central	80	4029.0
5	Trefforest	78	3527.0
6	Pontypridd	78	3611.0
7	Trefforest Estate	77	2079.0
8	Radyr	77	2086.0
9	Taffs Well	77	2078.0

- No overlap in the top 10 most frequently visited stations and stations with the highest delays.
- This is a positive outcome, as it indicates the most popular train stations maintain higher operational efficiency to mitigate delays.

Can We Predict If & How Long a Train Will be Delayed Based on Its Station?



Can We Predict If & How Long a Train Will be Delayed Based on Its Station?



Model prediction results:

92.9% accuracy

- True Negatives: 30138 (Predicted 0, actual 0)
- False Positives: 88 (Predicted 1, actual 0)
- False Negatives: 2229 (Predicted 0, actual 1)
- True Positives: 58 (Predicted 1, actual 1)

Performance on Delayed:

- Precision: 40% of Delay predictions were correct
- Recall: 3% of Not Delayed were correctly identified

Performance on Not Delayed:

- Precision: 93% of Not Delayed instances were correct
- Recall: 100% of trains Not Delayed were correctly identified

Improving Train Status

Identify Delay Causes Predictions Identify Station Patterns

Use feature engineering/ feature selection to identify important variables

Gather more data over time to identify variability through common station patterns

01

02

03

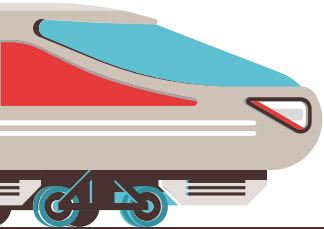
04

Identify
Cancellation
Causes

Implement
Changes

Find a reliable way to gather and incorporate cancellation data and factors

Implement changes and improve model performance/ practicality



Thank you!



Meet the Team!



Riley Sample

rsc622.rcs@gmail.com
[linkedin.com/in/rileysample](https://www.linkedin.com/in/rileysample)



Brianna Browning

briannabrowning223@gmail.com
[linkedin.com/in/brianna-browning](https://www.linkedin.com/in/brianna-browning)



Recommendations

The following recommendation are intended for scheduling predicting purposes

Identify delay causes	While Tuesday has the highest number of departures, it also has a significant number of delays. Friday has a low number of departures, but many delays. Focus on the efficiency if more trains were to be added on Fridays.
Identify cancellation causes	Work to reduce the rate of cancellations around peak hours by identifying strong correlating external factors such as station foot traffic, weather conditions, and maintenance schedules.
Identify why less frequently visited stations have more delays	Such delay factors may include: economy, population, poverty, crime, and lifestyle/amenities of the area
Adjust maintenance schedules	Use the morning dead hours for maintenance fixes
Analyze external impacting factors	Analyze outside causes for delay effects, such as weather, station foot traffic, geographical locations, area population/congestion, and local events



As more data is collected, the analysis will begin to show patterns in delays and cancellation, leading to more effective predictions and resolution options for inefficiencies.