

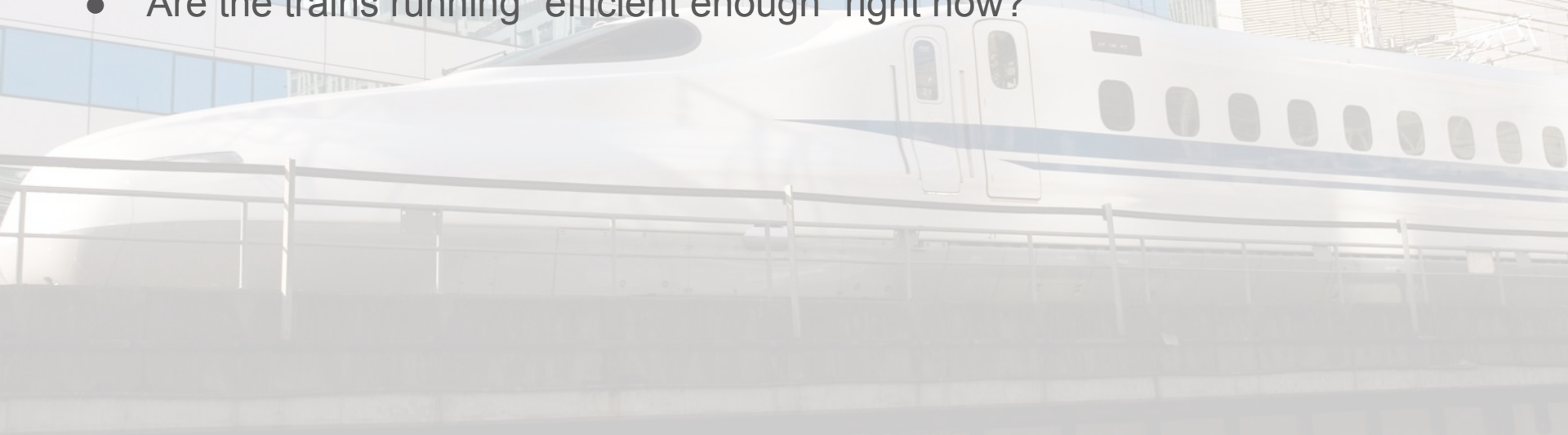
Welcome to the Joevengers

Rail Revisor Project



Business questions

- We have rail lines that have consistent delays.
- We have stations that are underutilized and should be closed/retired.
- How can we collect better data to give the most appropriate analysis?
- Are the trains running “efficient enough” right now?

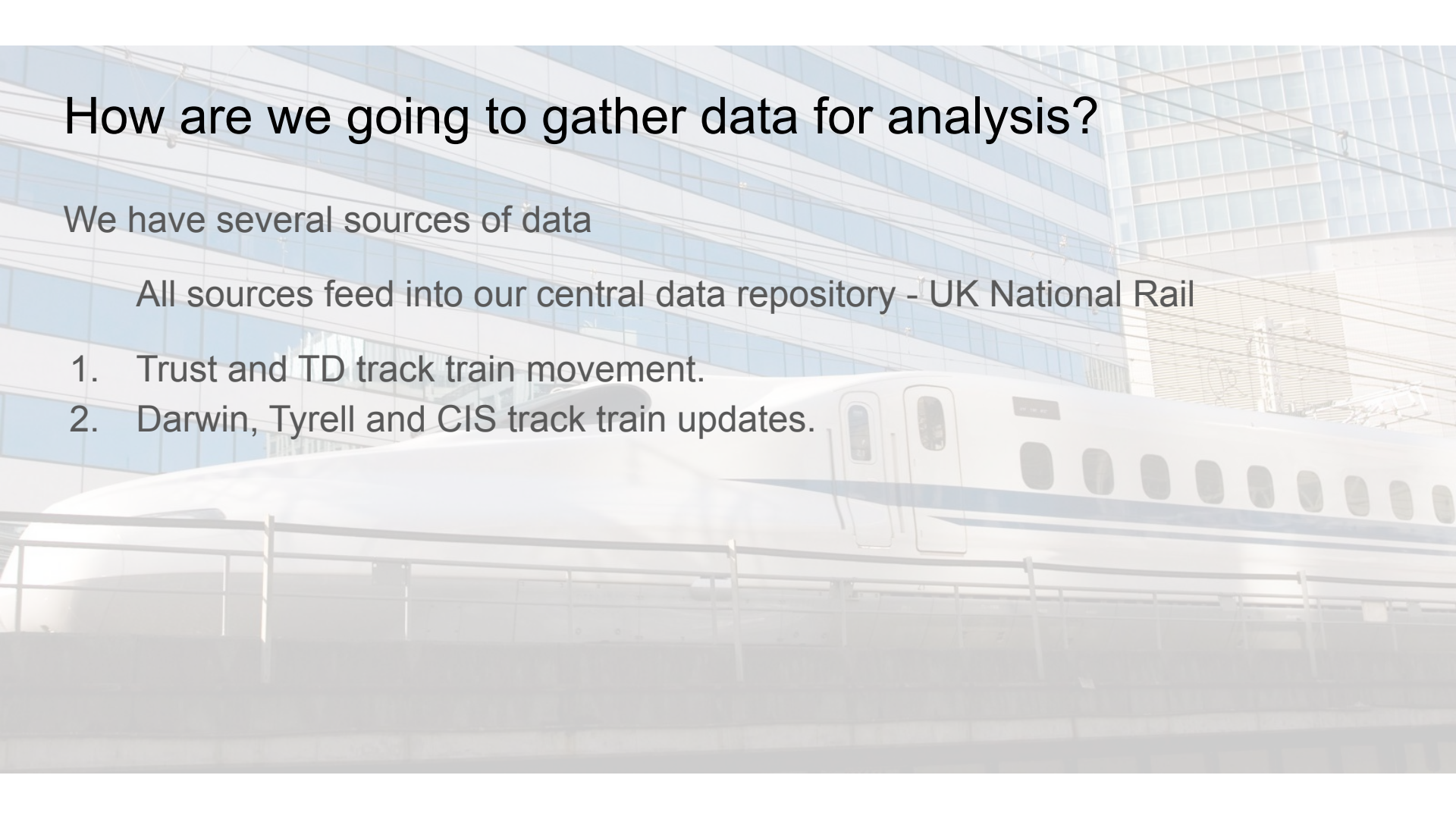


How are we going to gather data for analysis?

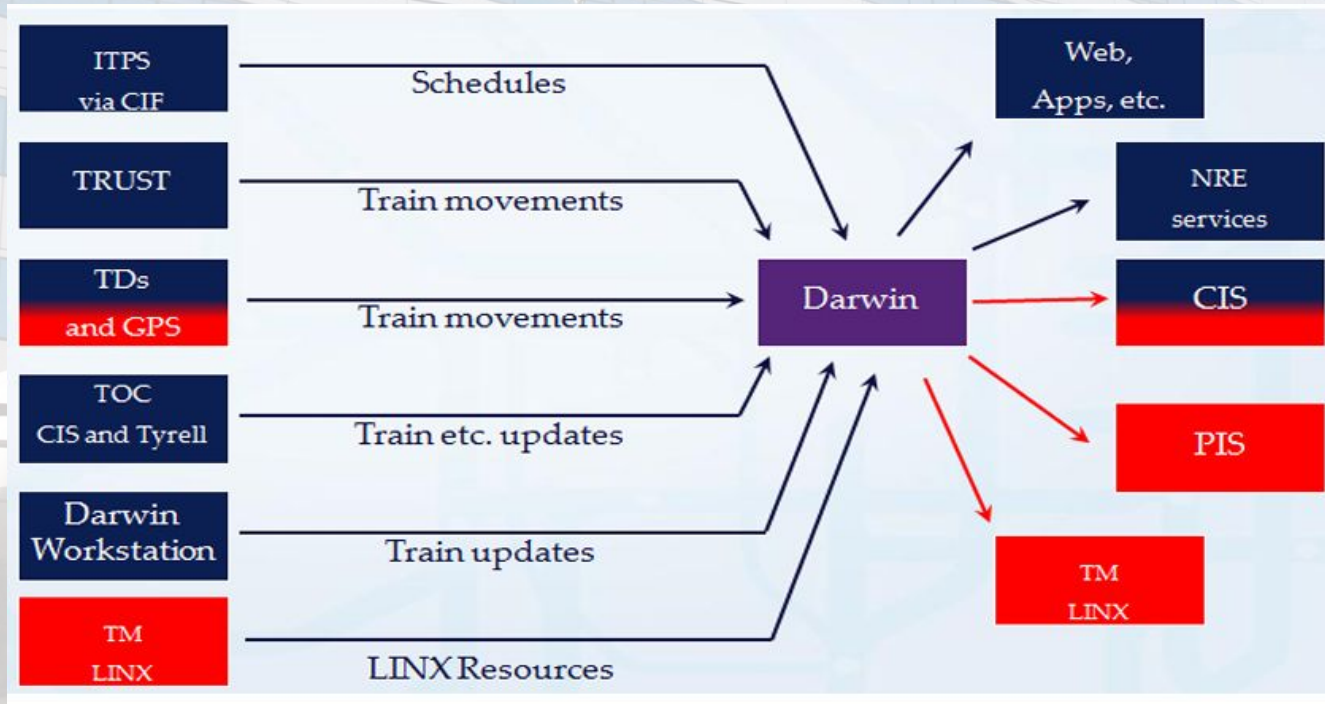
We have several sources of data

All sources feed into our central data repository - UK National Rail

1. Trust and TD track train movement.
2. Darwin, Tyrell and CIS track train updates.



Data flow diagram - Part 1 - There are several systems that feed into darwin.



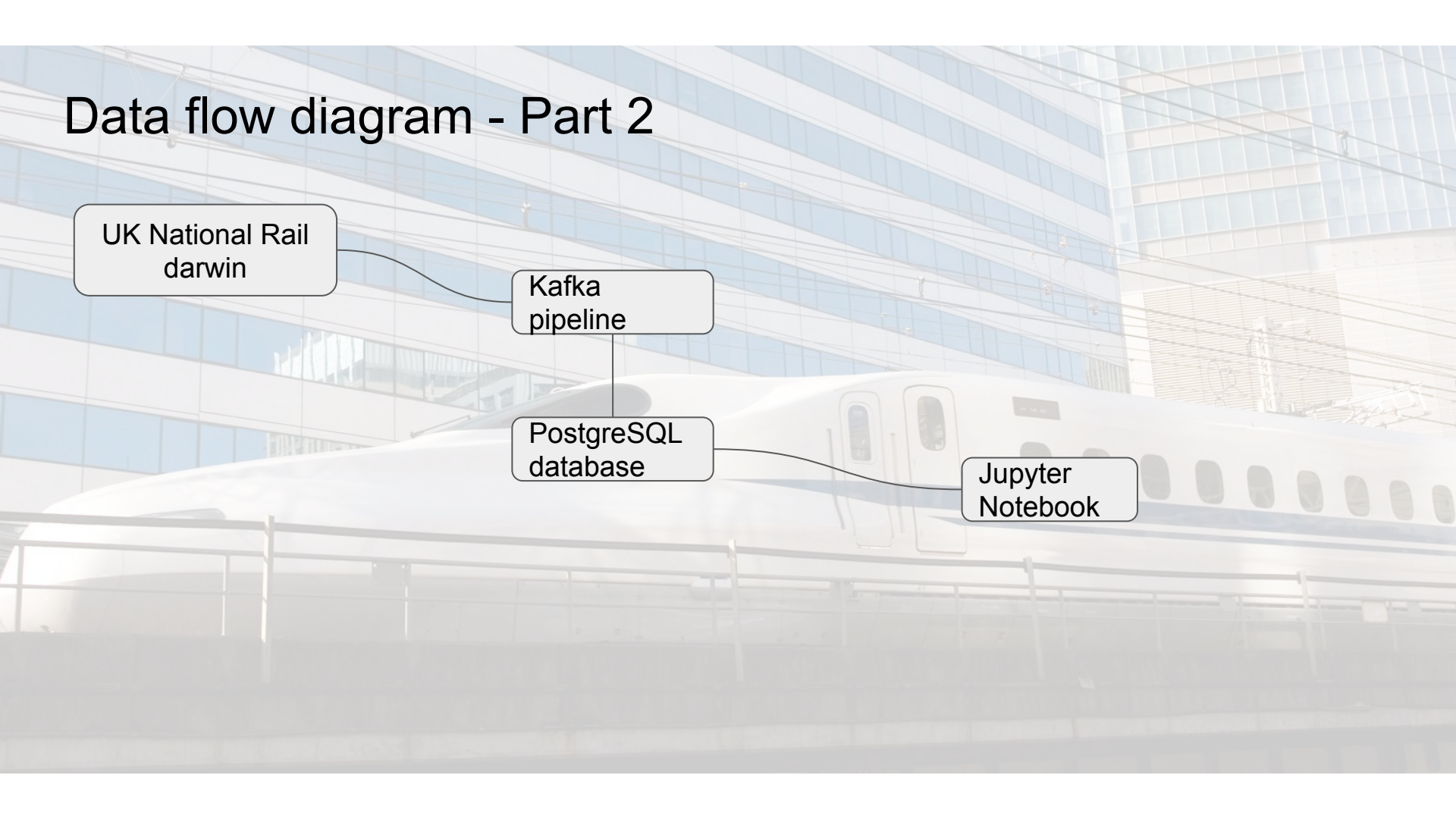
Data flow diagram - Part 2

UK National Rail
darwin

Kafka
pipeline

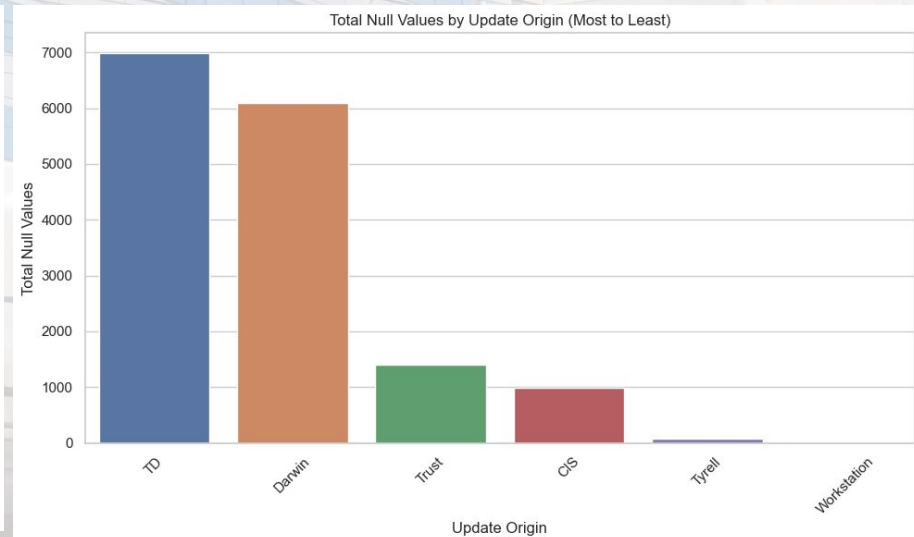
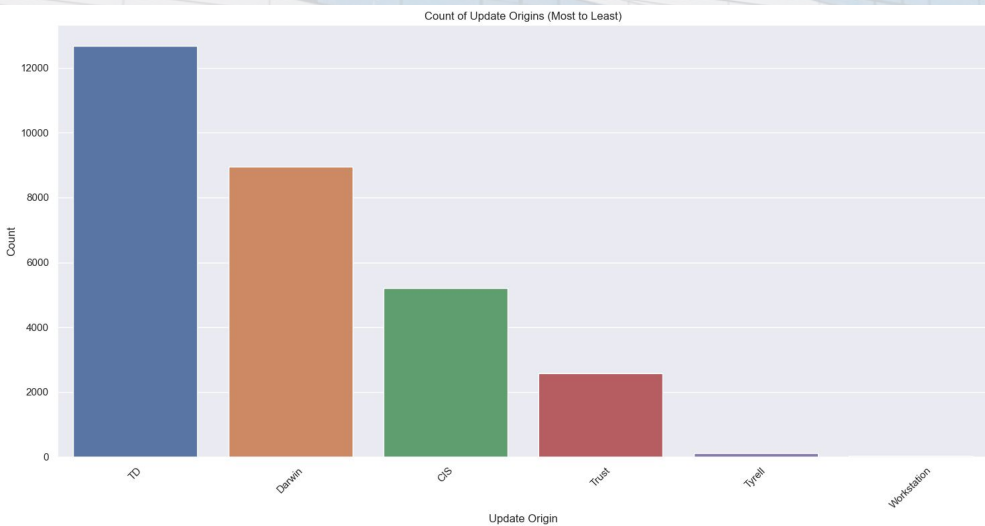
PostgreSQL
database

Jupyter
Notebook



The Sources and It's Failures

We have too much missing time data
from each Update Origin.

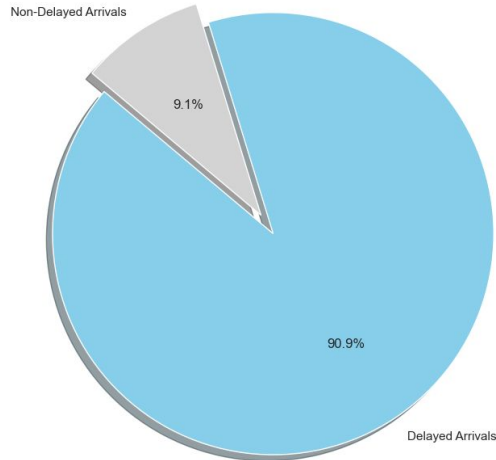


A Slice of Delays Derived From Null Values

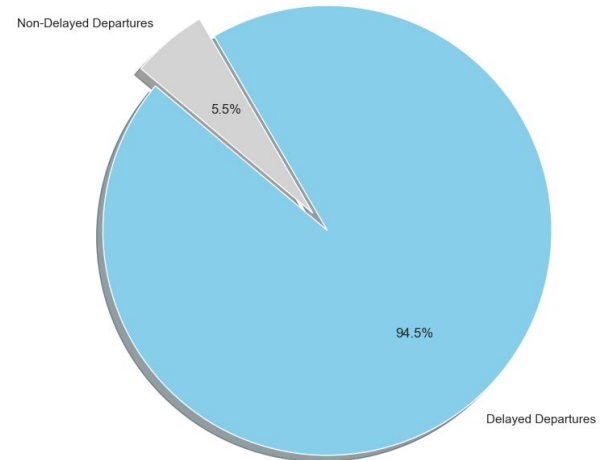
Most of our reported delays are the result of the origin updates pushing null time data.

This creates a challenge in seeing what the actual average delay is.

Percentage of Delayed Arrivals Coming From Null Planned Arrival Values



Percentage of Delayed Departures Coming From Null Planned Departure Values



Lean, Mean, Machine

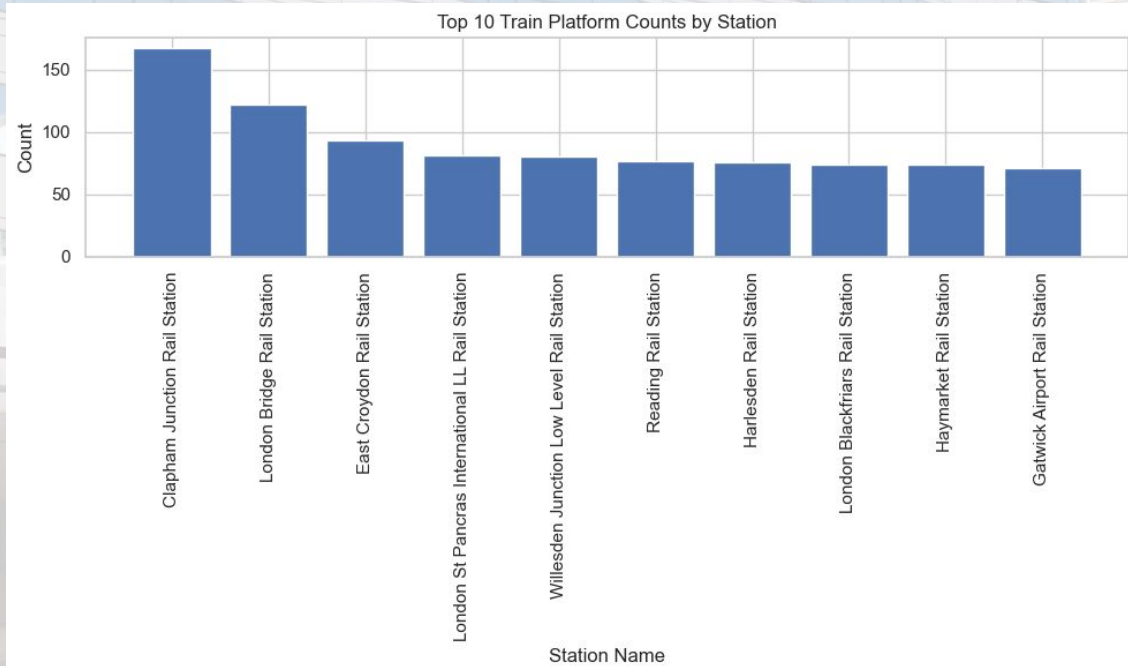
Even with null caused delays, an overwhelming majority of trains are arriving on time.



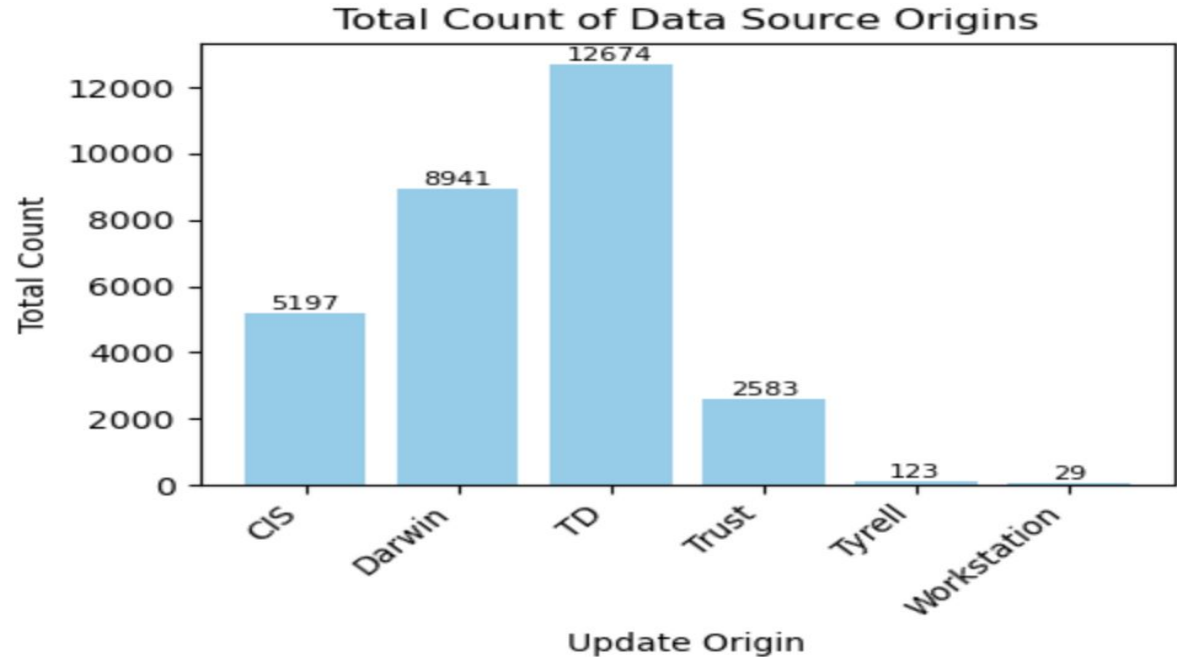
Busy Stations

These are the top 10 business rail stations.

It should be no surprise that most of these stations are in or near London.

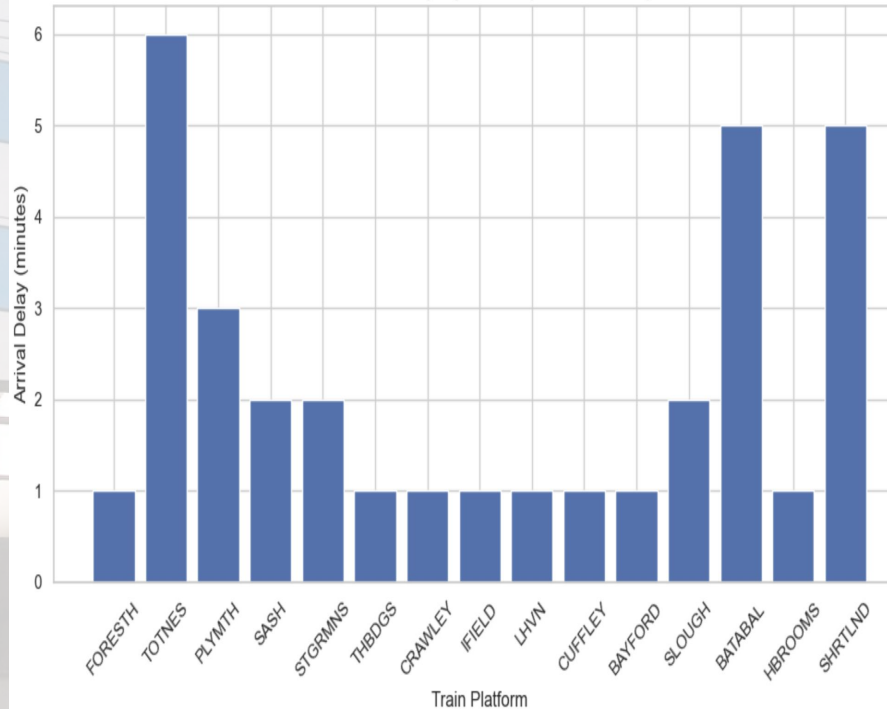


Business Case :Reasons for Delays and Recommendations for Better Customer Services -Source for Data Analysis ?

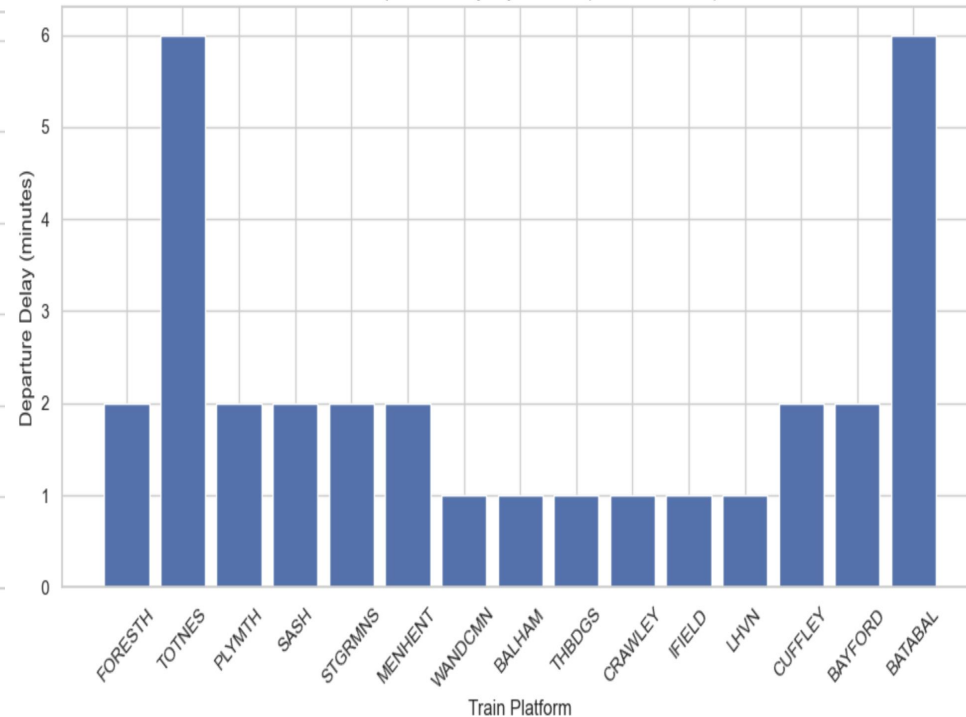


Business Case :Reasons for Delays and Recommendations for Better Customer Services -Arrival and Departure Delay Pattern

Train Arrival Delays by Platform (First 15 Records)

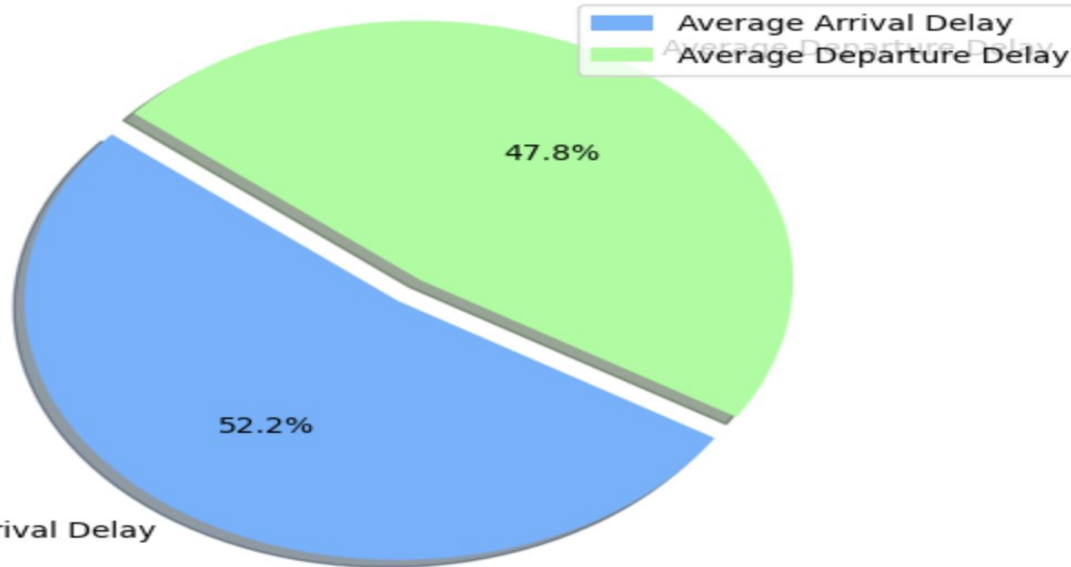


Train Departure Delays by Platform (First 15 Records)



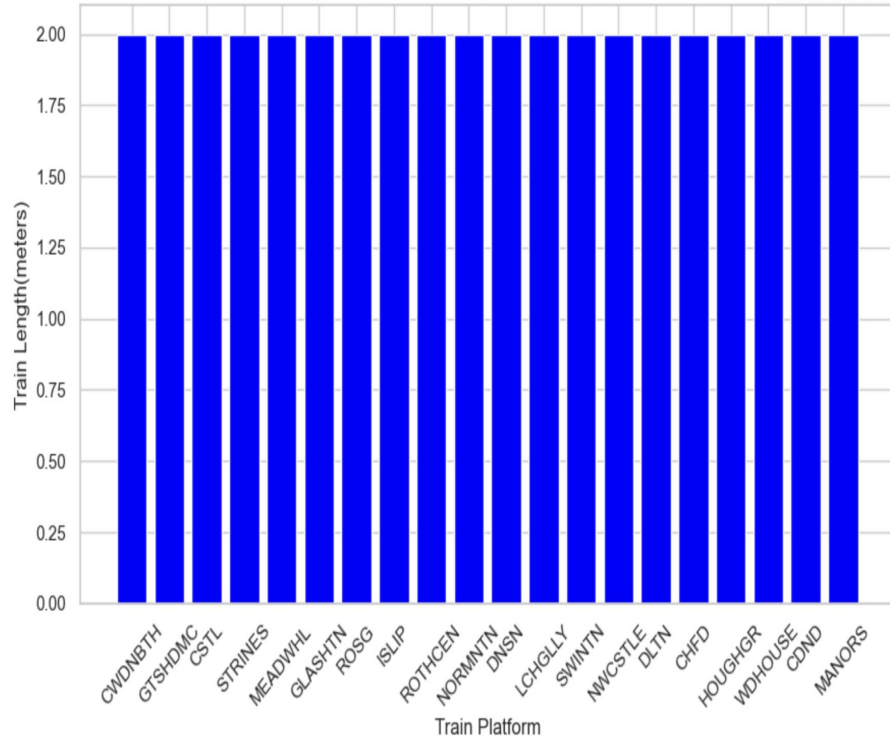
Business Case :Reasons for Delays and Recommendations for Better Customer Services - Average Arrival and Departure Delay Pattern

Distribution of Average Delays

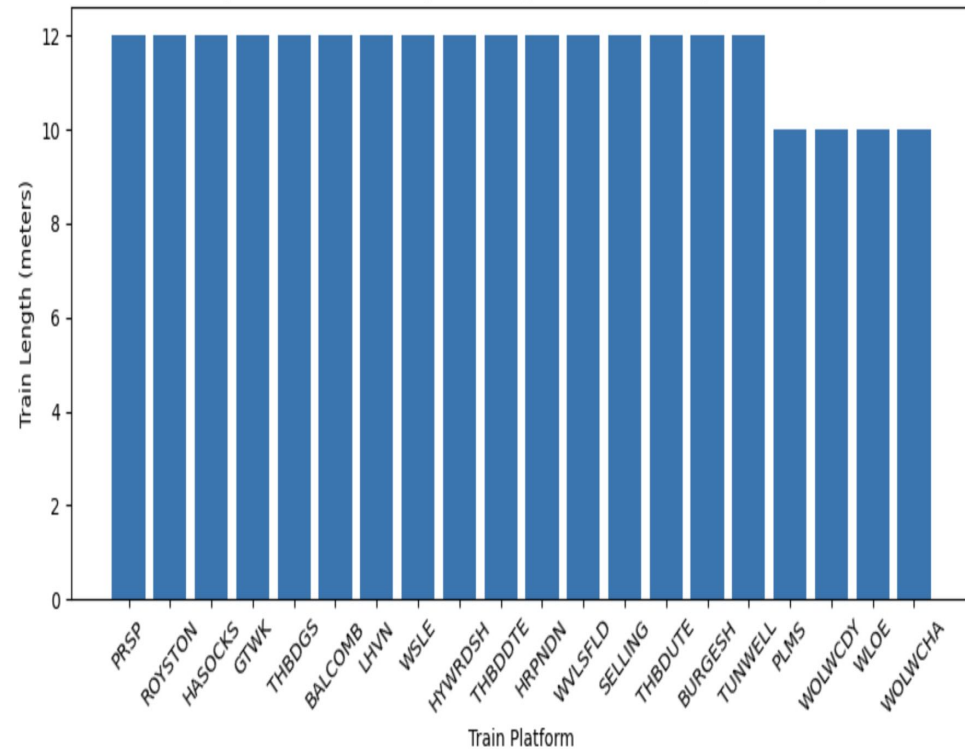


Business Case :Reasons for Delays and Recommendations for Better Customer Services -Train Length Pattern

Top 20 Train Platforms by Train Length with No Arrival&Departure Delays

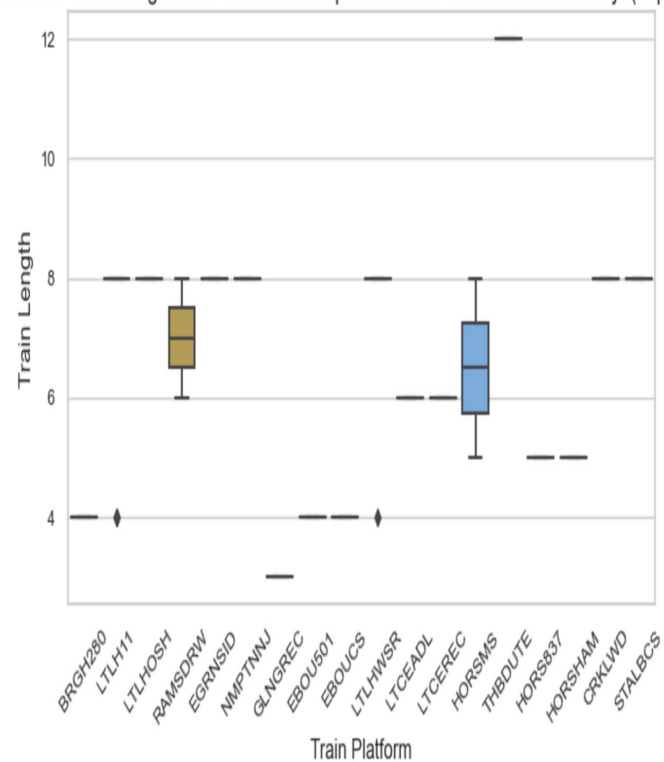
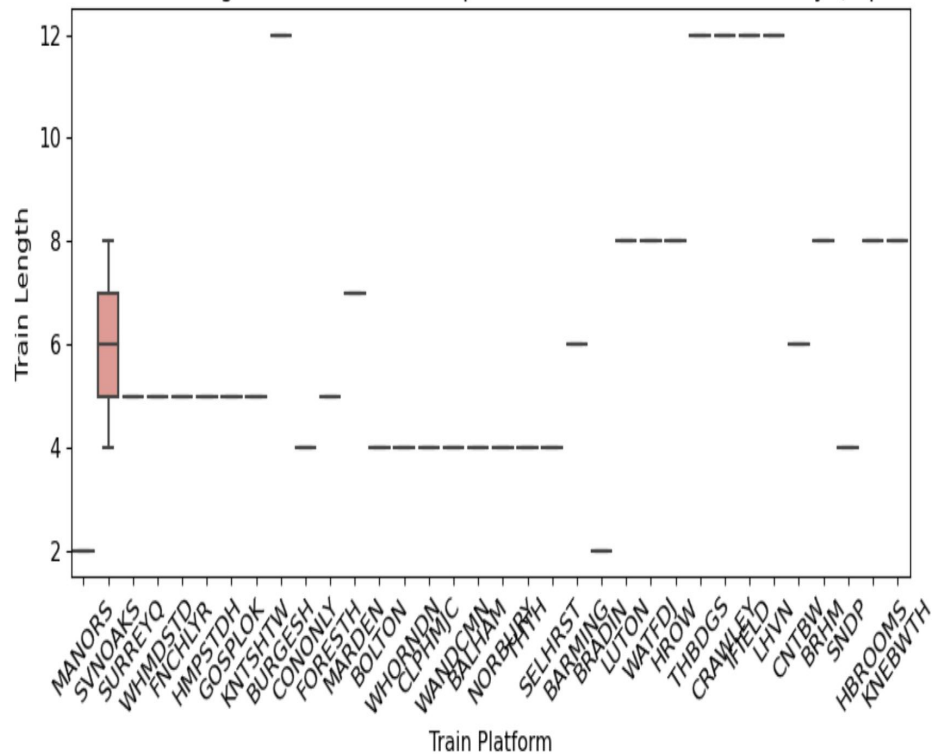


Top 20 Train Platforms by Train Length with Arrival&Departure Delays



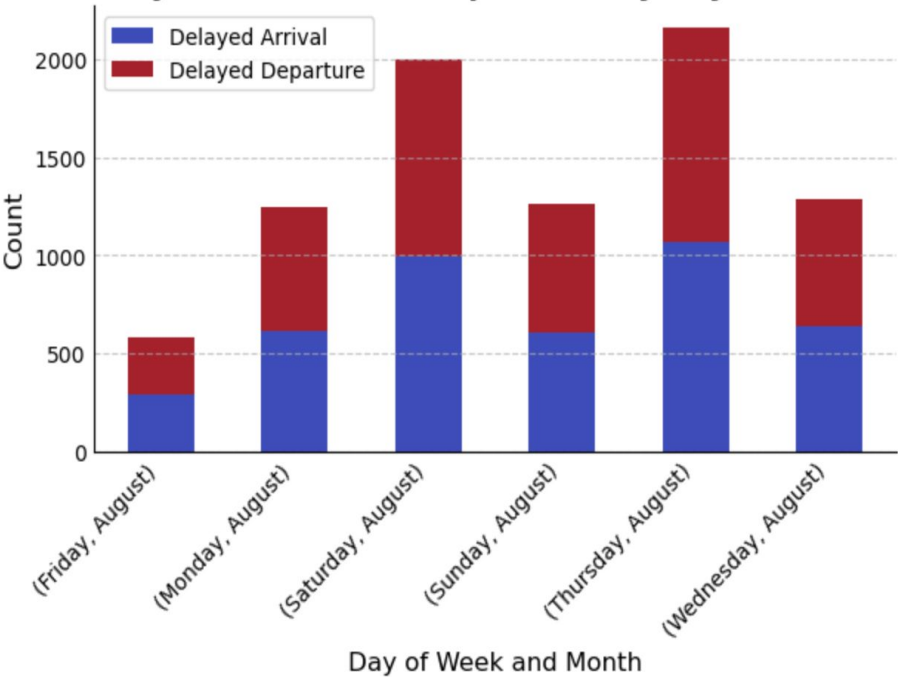
Business Case :Reasons for Delays and Recommendations for Better Customer Services -Train Length Correlations Pattern

Distribution of Train Lengths for Arrival and Departure Train Platforms - No Delays(Top 35 Records) Distribution of Train Lengths for Arrival and Departure Train Platforms with Delays(Top 35 Records)



Business Case :Reasons for Delays and Recommendations for Better Customer Services -Analyzing Maximum Delays by Day of the Week and Month

Delayed Arrivals and Departures by Day and Month



	Day	Month	Platform Name	Service Day	Arrival Delay	Departure Delay
1	Wednesday	August	WELLING	2023-08-16	False	False
2	Wednesday	August	MANORS	2023-08-16	False	False
3	Wednesday	August	SVNOAKS	2023-08-16	False	False
4	Wednesday	August	CLPHMJM	2023-08-16	False	False
5	Wednesday	August	SURREYQ	2023-08-16	False	False
6	Wednesday	August	WHMDSTD	2023-08-16	False	False
7	Wednesday	August	FNCHLYR	2023-08-16	False	False
8	Wednesday	August	HMPSTDH	2023-08-16	False	False
9	Wednesday	August	GOSPLOK	2023-08-16	False	False
10	Wednesday	August	KNTSHTW	2023-08-16	False	False
11	Wednesday	August	BURGESH	2023-08-16	False	False
12	Thursday	August	TMPLORR	2023-08-17	False	False
13	Wednesday	August	CONONLY	2023-08-16	False	False
14	Wednesday	August	FORESTH	2023-08-16	False	False
15	Thursday	August	MNKRISB	2023-08-17	False	False

Business Case :Reasons for Delays and Recommendations for Better Customer Services

Insights from Stacked Bar Chart Analysis:

The stacked bar chart highlights a notable trend of increased delays on Saturdays and Thursdays during the month of August. This observation suggests the presence of underlying factors contributing to higher delays on these specific days and in that particular month.

Factors Possibly Contributing to Delays:

Several potential factors could influence this pattern:

- **Weather Conditions:** Adverse weather conditions, common during certain months, might impact train operations and cause delays.
- **Peak Travel Season:** August might coincide with peak travel season, leading to higher passenger volume and potential operational challenges.
- **Maintenance and Repair:** Scheduled maintenance or unexpected repair work could disrupt regular operations.
- **Operational Challenges:** Operational inefficiencies, staffing issues, or infrastructure limitations might be more pronounced on Saturdays and Thursdays.
- **External Factors:** External events, such as local events or holidays, might affect demand and contribute to delays

Business Case :Reasons for Delays and Recommendations for Better Customer Services -

Recommendations for Improved Customer Service: Based on the analysis, recommendations could include:

- Weather Monitoring: Implement real-time weather tracking and contingency plans to mitigate weather-related disruptions.
- Seasonal Preparedness: Enhance resources and staff during peak travel seasons to manage increased passenger traffic effectively.
- Maintenance Planning: Optimize maintenance schedules to minimize disruptions during peak travel periods.
- Operational Efficiency: Identify and address operational inefficiencies specific to Wednesdays and Thursdays.
- Event Management: Coordinate with local event organizers to better manage potential demand fluctuations.
- Passenger Data: We need ridership and pricing data to make financial decisions and accurately examine passenger related delays

Train Length and Delays: Additionally, investigating the relationship between train length and delays could provide valuable insights. Longer trains might require more time for boarding and departing, impacting schedules. Ensuring efficient passenger flow and boarding procedures could help reduce delays associated with train length.

Conclusion: The EDA reveals a distinct pattern of delays on some days in August. By considering factors like weather, peak travel season, maintenance, and operational challenges, we can formulate strategies to mitigate delays and enhance customer experiences. Further analysis of train length's impact on delays offers opportunities for process optimization and better service delivery.



WHERE

WHAT

WHEN

WHY

HOW

WHO

Notes & links

[Train picture](#)

[Image transparency](#)

[Move Image to back](#)