

[Ahmed Ali Sadek Ateya] Rail Riddle Report

Major Predictions and Approach

What trends or patterns influence rail delays?

Rail delays appear to be influenced by various factors, including the number of wagons, loading times, and signal problems.

What is the reasoning behind your answer?

The analysis suggests a correlation between the number of wagons and delays, with larger train sizes leading to increased loading and handling times. Signal problems have also been identified as a contributing factor to delays.

How robust are these patterns, and how might they be improved?

In case of total handling delay and cargo movement delay Two patterns are supported and confirmed the influence of wagon numbers. To improve robustness, further investigation into specific signal problems and operational constraints is recommended.

Some types of delay may be more predictable than others. Which types of delay are most and least predictable?

Delays related to signal problems seem less predictable, while delays caused by loading times and train size show stronger predictability.

Key Features or Metrics

What are the key features or signatures in the data that help us predict rail delays?

The number of wagons, loading times, cargo movement delays, and signal problems appear to be key features influencing rail delays.

Did you use any additional external data?

No external data was used for this analysis.

Did this data improve your ability to predict rail delays?

Yes, the data has provided insights into the relationship between various factors and rail delays.

Methods and Approach

What is the best analytic approach to take?

Exploratory data analysis, scatter plots, and bar plots were employed to identify patterns and correlations.

If you think machine learning is viable, what type of algorithm would you use?

If using machine learning, regression algorithms could be suitable for predicting delays based on multiple features.

What did you try?

Exploratory data analysis, scatter plots, and bar plots were created to visualize relationships between variables and delays.

Did anything not work?

The analysis revealed meaningful insights; however, additional domain knowledge might be needed to interpret certain patterns.

Data Preparation

Were there any variables or variable sets that required significant cleaning, or contained significant missing values. How would you recommend dealing with these?

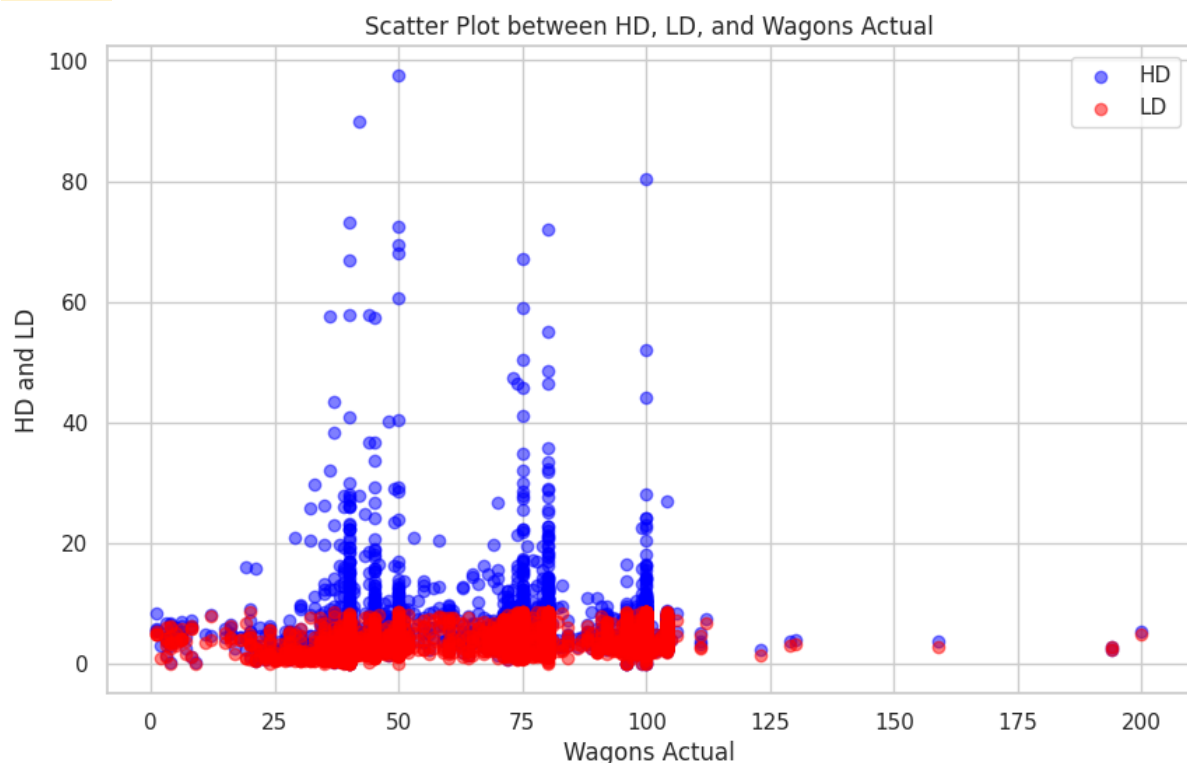
Some variables had missing values, particularly Destination.QuantSum, Cancelled.ReasonCode, NWB trains.NWBPreliminaryDate, NWB trains.NWBConfirmedDate, and Wagons.Actual. Imputation or data exclusion could be considered.

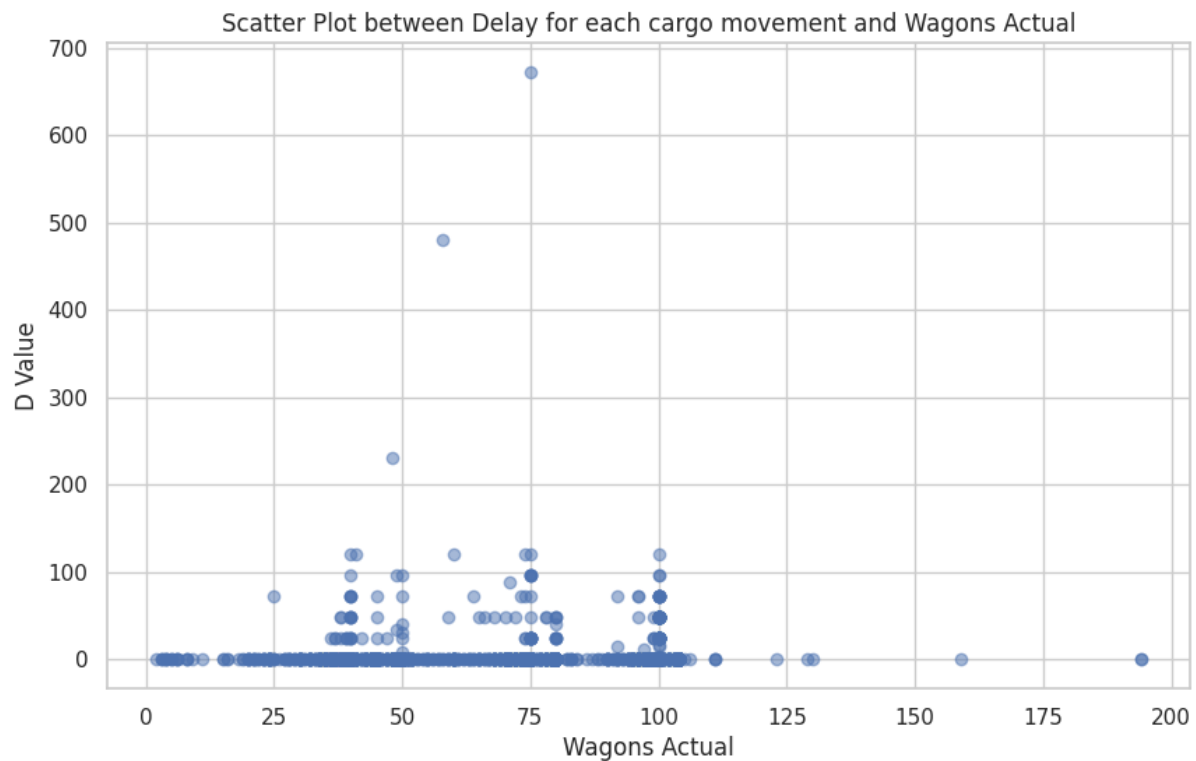
What other sources of data would be most helpful in predicting rail delays?

External factors such as weather conditions, maintenance schedules, and track availability could provide additional predictive power.

Data Understanding

Include 1 or 2 visuals that demonstrate the relationships between variables and the target variable.





Did you create any other features from the dataset?

Yes, I created

LD for loading delay = FL - SL

HD for total handling delay = T2-T1

D = Delay for each cargo movement

Other Findings

Have you found any other valuable insights from the data?

I think wagon type would be useful but it was in old system data

Recommendations

What do you recommend as the next steps?

1. Further investigation into specific signal problems and operational constraints.
2. Consider incorporating external data sources to enhance predictive modeling.
3. Develop machine learning models to predict delays based on identified features.

What information or data is missing that would be needed to develop a machine learning solution?

Additional data related to external factors (weather, maintenance schedules) and operational details (loading equipment efficiency) would enhance the accuracy of machine learning models.