



DigiHaul Data Analysis and Optimization

Presented by Pushpa NB
Data Scientist
26th May 2024



INTRODUCTION

- ❖ Overview of the data and its analysis methods
- ❖ Presentation of the findings
- ❖ Selection of features and models for predicting new bookings
- ❖ Conclusion with recommendations



Problem Statement

Road Haulage Overview:

- ❖ Road haulage is essential for the UK economy.
- ❖ 90% of goods transported by land in Great Britain are moved by road.
- ❖ DigiHaul specializes in digital transport solutions, integrating data from Carriers and Shippers for seamless logistics.



Data Overview

Datasets Provided

- ❖ GPS_data : Unique shipment, Latitude, Longitude, Timestamp
- ❖ Shipment_bookings : Unique shipment ID, Shipper ID, Carrier ID, Vehicle type, Collection postcode, Delivery postcode, Collection and Delivery Coordinates and Schedules
- ❖ New_bookings: Same format as Shipment_bookings for a different time period



Methodology

- Data Preprocessing:
 - Cleaning and preparing the data for analysis.
 - Merging fault data with rainfall and calendar data.
- Feature Engineering:
 - Selecting the appropriate feature
- Model Selection:
 - Choosing appropriate models (e.g., regression) to predict given new bookings delay
 - Justifying model choices based on data characteristics.



Task 1 - On-time Delivery Analysis

Objective:

Calculate the percentage of shipments that met the on-time delivery threshold (within 30 minutes of the scheduled window) from October 1st to December 31st, 2023.

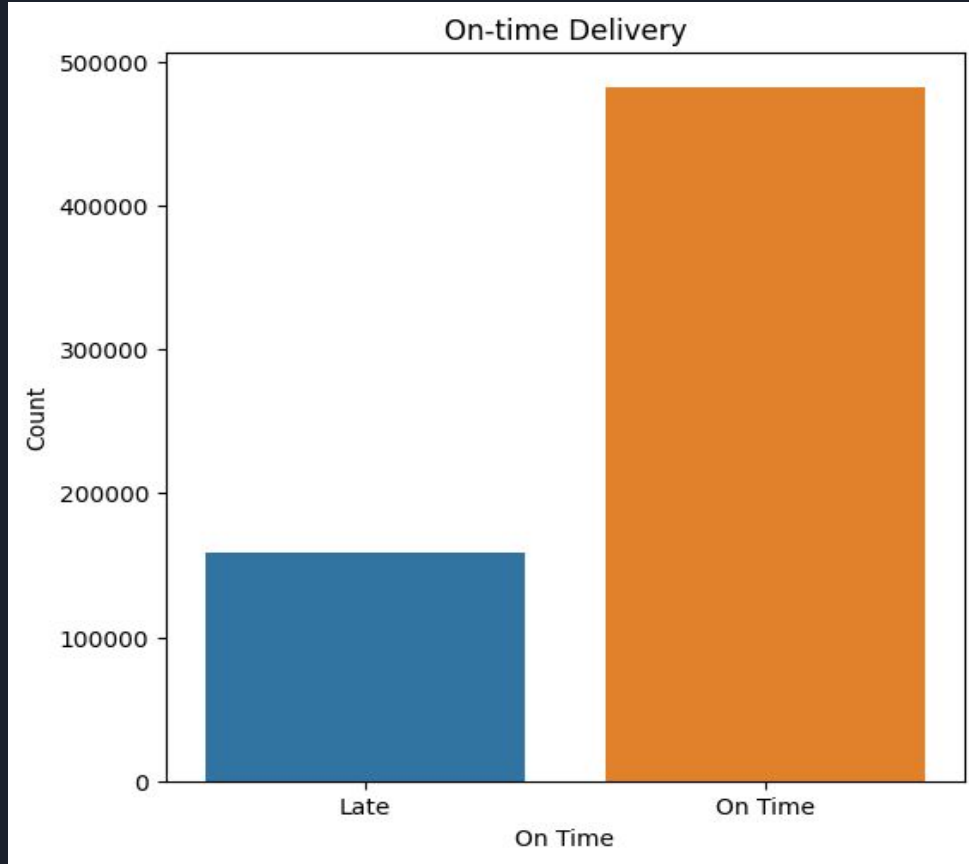
Assumptions:

A delivery is considered on-time if it occurs within the scheduled window plus 30 minutes.

GPS_data timestamps are compared against the delivery schedule.

Results:

Percentage of on-time deliveries: 75.13%



- ❖ After the analysis, we found that the count of delayed deliveries is lower than the count of on-time deliveries, which is a positive indication

Task 2 - Notification of Potential Delays

Objective:

Identify shippers to notify about potential delays during October 1st to December 31st, 2023.

Methodology:

Compare actual delivery times against the scheduled window.

Identify shipments with potential delays.

Determine the shippers for each delayed shipment and the times for notification.

Results:

Example of notified message to the shipper is shown in the image

Please be informed that the following shipments are potentially delayed:

SHIPMENT_NUMBER	LAST_DELIVERY_SCHEDULE_LATEST	\
548065	SEZHUK-231214-273813	2023-12-15 16:00:00+00:00
548066	SEZHUK-231214-273813	2023-12-15 16:00:00+00:00
548067	SEZHUK-231214-273813	2023-12-15 16:00:00+00:00
548068	SEZHUK-231214-273813	2023-12-15 16:00:00+00:00
548069	SEZHUK-231214-273813	2023-12-15 16:00:00+00:00
548070	SEZHUK-231214-273813	2023-12-15 16:00:00+00:00

RECORD_TIMESTAMP

548065	2023-12-15	16:35:00+00:00
548066	2023-12-15	17:19:00+00:00
548067	2023-12-15	17:23:00+00:00
548068	2023-12-15	17:33:00+00:00
548069	2023-12-15	17:34:00+00:00
548070	2023-12-15	17:37:00+00:00

Regards,
DigiHaul Team



Task 3 - Delay Prediction Model

Objective:

Predict the likelihood of delay for new shipments using the "New_bookings.csv" dataset.

Approach:

Train a predictive model using historical shipment data.

Features: Vehicle size, build-up, collection and delivery schedules, historical delay patterns.

Results:

Prediction Accuracy: got around 79% of accuracy from few of the regression models(ex: Linear, Decision Tree, Random Forest etc)



Summary of Findings

Key Insights:

On-time delivery performance and improvement areas.

Shippers needing notification for potential delays as already shown in the image which has to be followed to improve the delivery

Predictive model outcomes for new shipments.

Recommendations:

Enhance real-time tracking to improve on-time delivery rates.

Few details are missing in the data provided like missing shipping id.

Implement automated notifications for potential delays.

Utilize the predictive model for proactive delay management.

Thank You

