Report - research

Marta Figurska

May 2025

1 Introduction

This part of the report focuses on a deeper analysis of delivery data, aiming to uncover patterns and insights. The primary goal is to evaluate actual delivery performance. The analysis is based on cleaned and filtered data. The process of filtering and methodology is described in "$marta\_figurska\_analysis$" report, I will paste it in 1.1 for clarity.

1.1 Methodology and assumptions

This subsection is copied from "$marta\_figurska\_analysis$".

I began by carefully reviewing the available data, examining the relationships between tables and looking into how the data was inserted into the database. This step was crucial to understanding the context and structure of the data.

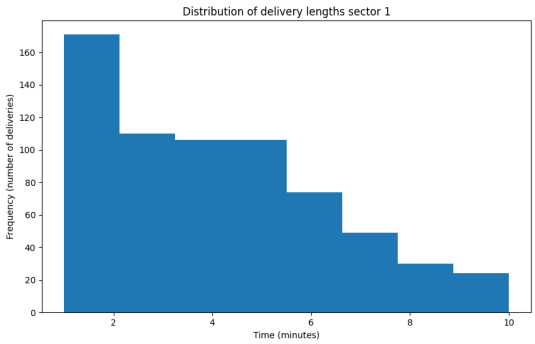
- I assumed that the time between segments is not important. I also ignored overlapping between segments' timelines.
- When the start and end times for a delivery were identical, these records were removed - as I assumed it is invalid data. Also instances where the start time occurred after the stop time were also removed.
- The main focus was on "STOP" data as it was connected to sectors, which were important during the analysis.
- To improve the accuracy of the analysis, 2% of the highest and lowest values were removed. This helped in decreasing the impact of extreme outliers which could have been registered while abnormally huge traffic or after GPS failure. They should not have an influence on the results.
- An important aspect of the analysis was to identify deliveries that starts in the evening and ends after midnight, as these deliveries were problematic. I made sure to handle them.
- I used Seaborn and Matplotlib in Python to generate histograms. I also used Microsoft Power BI for creating charts. These visualizations provide a clear picture of the data distribution and help in understanding insights.

2 How to improve the algorithm

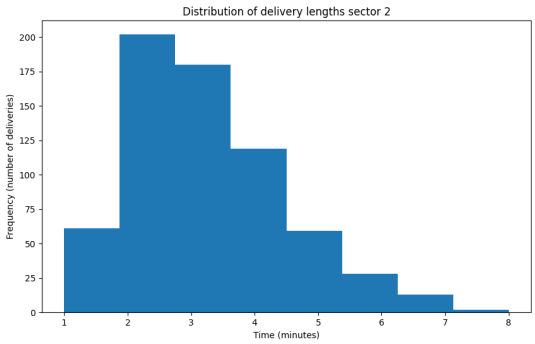
2.1 Considering sectors

The average duration of delivery in each sector.

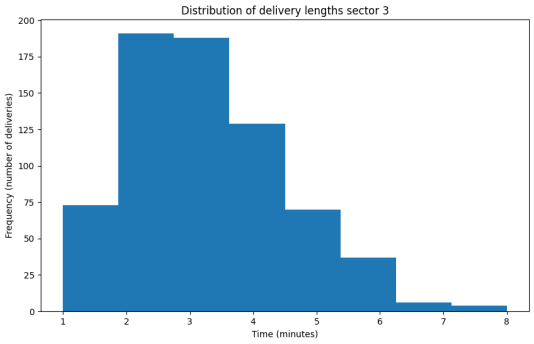
- 4.4 min for the first sector
- 3.1 min for the second sector
- 3.2 min for the third sector



(a) Sector 1



(b) Sector 2



(c) Sector 3

Figure 1: Histograms of delivery durations across sectors

As we can see, the first sector takes significantly more time. Also the distribution of time taken in the first sector is different - the histogram is more equally distributed. That's why predicting the delivery time for all sectors using the whole data is not effective. We need to predict the delivery time for the sectors separately, because they need different treatment.

Dividing into smaller sectors could be considered.

2.2 Different algorithm

First, the algorithm should be connected to available external application that supports planning the route. Our custom algorithm should additionally consider extra factors for each sector separately. Also rethinking the boundaries of sectors could be helpful. Distributing the drivers into each sector will locate them closer to customers.

2.3 Other potentially influential factor

Right now the algorithm to plan the duration of delivery is very naive because it does not take into consideration other factors than duration of previous deliveries. What else should be taken into account:

- **Driver's experience** – More experienced drivers may navigate routes faster and handle deliveries more efficiently.
- **Traffic conditions** – Traffic jams, accidents, or local events can significantly delay delivery times.
- **Customer not present** – If the recipient is not home, the driver may need to wait, call, or return later.
- **Vehicle issues or breakdowns** – Mechanical problems can halt or slow down the delivery process.
- **Address or street name errors** – Incorrect or unclear addresses lead to time lost in locating the destination.
- **Incomplete or missing orders** – If products are not prepared or missing from the store, the driver must wait.
- **Type of vehicle** – Different vehicles (bicycle, car, scooter) have different speed and access limitations.
- **Fuel requirements** – Stopping for fuel or running low may interrupt the delivery schedule.
- **Lack of available parking** – Drivers may need to park far away and deliver on foot, increasing time.
- **Inefficient route planning** – Poorly optimized routes result in unnecessary delays or detours.
- **High number of deliveries** – A larger number of orders in one route extends the overall delivery time.
- **Large or heavy orders** – Bulky items take longer to load, unload, and carry, affecting delivery duration.
- **Time of the day** – Peak hours typically experience higher traffic.

2.4 Data to collect

- **Type of vehicle** – Knowing whether the delivery was made by car, bike, scooter, etc., could help predicting similar courses.
- **Traffic in orders** – Capturing hourly or sectional delivery load can reveal peak periods, enabling better resources and route planning.
- **Loyal customer tracking** – Storing information about frequent customers could help in future deliveries or customer satisfaction analysis.

2.5 Risk of over- or under-estimating the delivery times

Overestimating delivery times can lead to inefficient route planning, wasted resources, and reduced number of deliveries per shift. It may also create negative opinion among customers.

Underestimating delivery times could result in delays, increased stress for drivers and customers. It can also affect logistics coordination and lead to time conflicts across deliveries.

However I believe that longer waiting for delivery is more frustrating.

3 Summary

Several real-world factors that influence delivery duration were identified, such as traffic conditions, vehicle type, delivery time of day etc. These insights highlight the complexity of the problem. The current algorithm needs to be changed.

Future work could include integrating additional data sources to develop more precise predictive models. Such improvements would support better planning and reduce delays.