Report

Marta Figurska

May 2025

1 Introduction

This report presents an analysis of the delivery data, focusing on various aspects such as time intervals or delivery sectors. It includes visual representations such as histograms and provides an explanation of the methodology used.

1.1 Methodology and assumptions

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 understand the context and structure of the data.

- I assumed that the time periods between segments don't influence delivery time. 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.
- Another aspect of the analysis was to identify deliveries that start in the evening and end 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 Results

After data cleansing, to begin the analysis, I created a general histogram showing the overall distribution of delivery durations. To gain more detailed insights, then I generated separate histograms for each sector, allowing a clearer comparison of delivery patterns. The delivery times were rounded up to minutes.

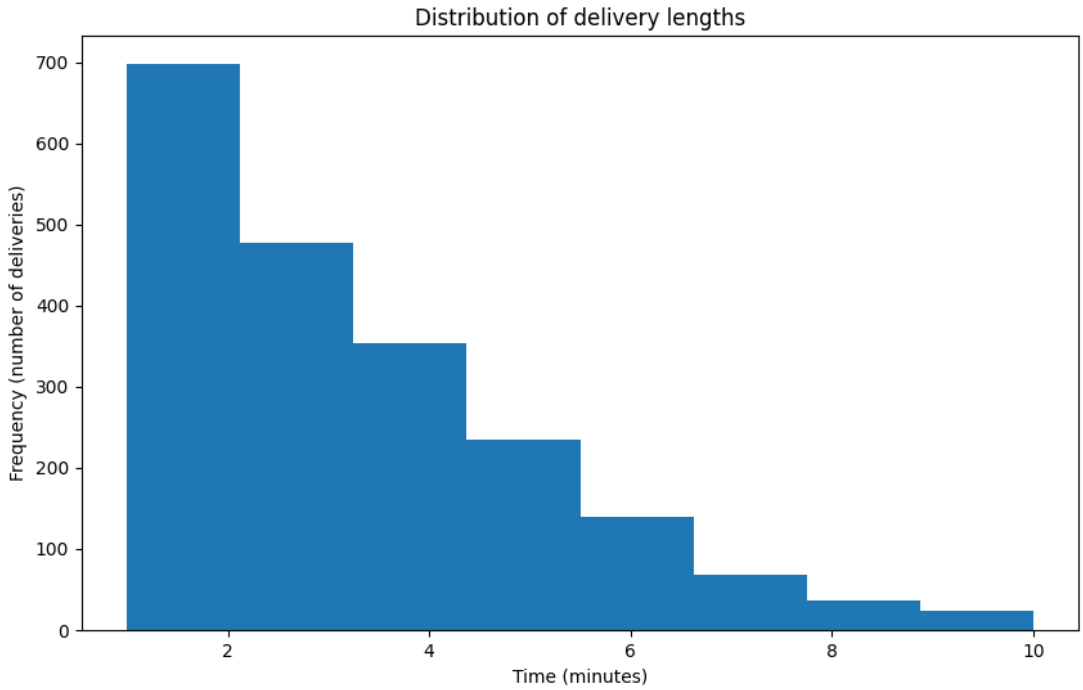
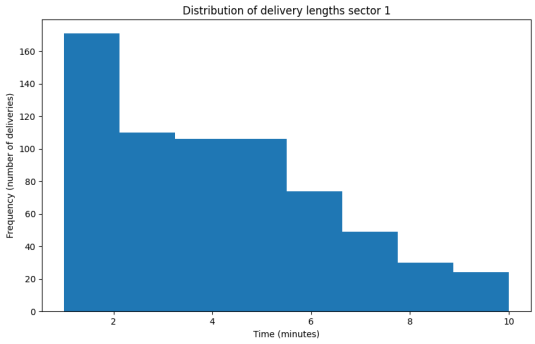


Figure 1: Histogram of actual delivery lengths

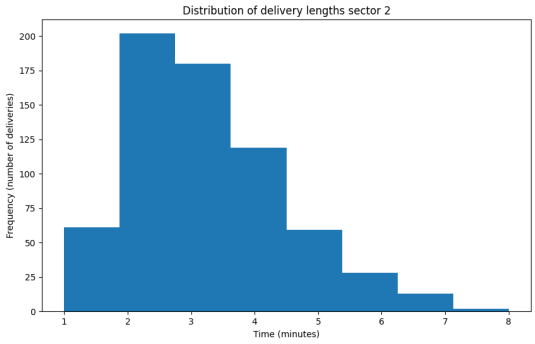
Most of the deliveries are in the first bin. The average delivery time is equal to around 3 minutes.

Below are presented the histograms for each sector. The average delivery times:

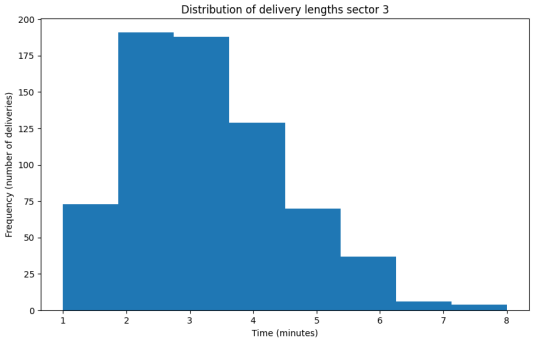
- 4.4 min for the first sector
- 3.1 min for the second sector
- 3.2 min for the third sector



(a) Sector 1



(b) Sector 2



(c) Sector 3

Figure 2: Histograms of delivery durations across sectors

The insights confirm the driver's thesis as one of the sectors takes the most time - the sector number one require significantly more time than other two sectors. It's average delivery time and maximum length of the delivery are the highest.

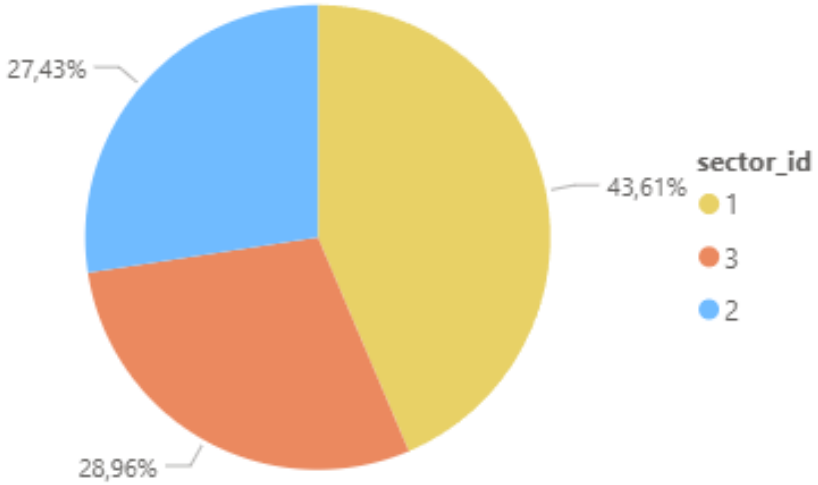


Figure 3: Chart showing proportions between sums of delivery times in sectors

There are presented two charts: first shows proportions between summed delivery times in each sector (above) and the number of deliveries in each sector (below). Number of deliveries

in each sector are comparable (first sector has slightly higher number) but the time consumed by deliveries in the first sector is significantly higher.

Figure 4: Number of deliveries per sector

2.1 Differences between planned and actual delivery duration

The charts presents the differences planned and actual delivery lengths. The positive value means that the driver arrived faster than planned, the negative value means the driver was late.

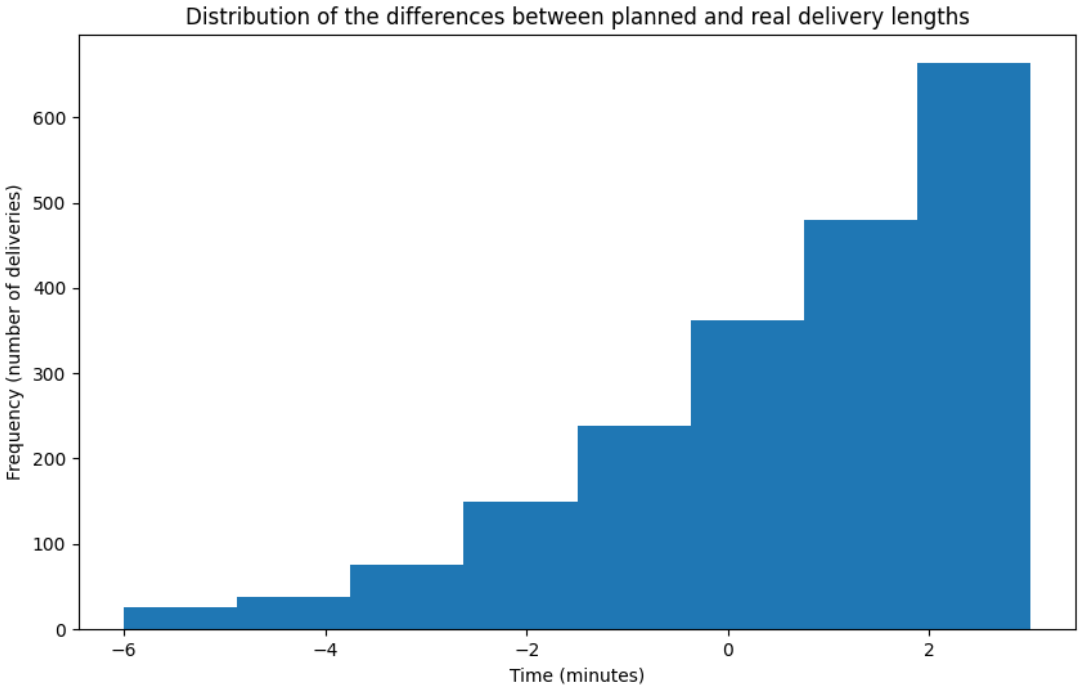
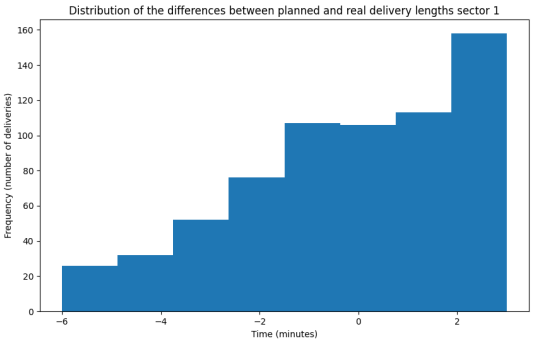
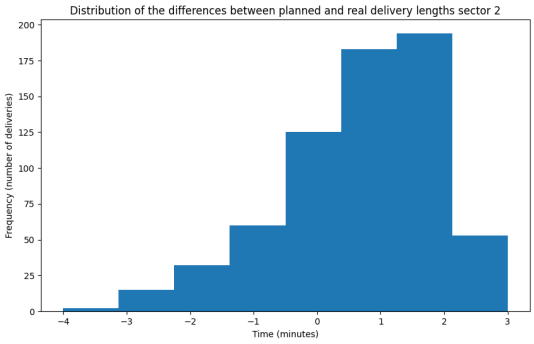


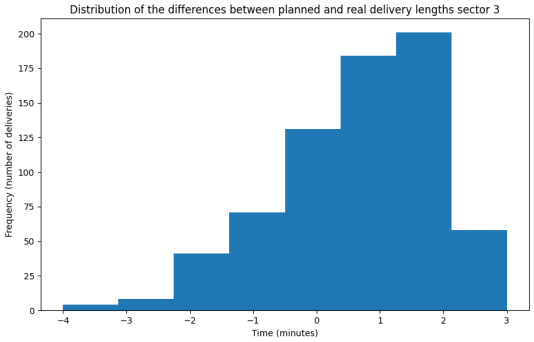
Figure 5: Histogram of differences between planned and actual delivery lengths



(a) Sector 1



(b) Sector 2



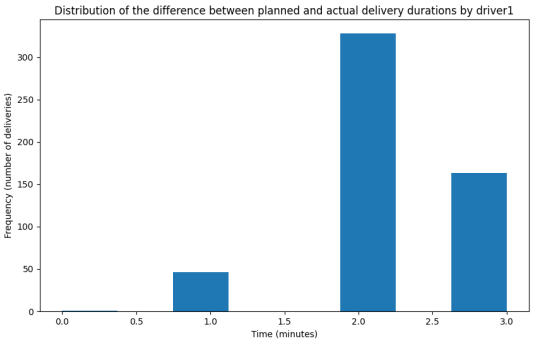
(c) Sector 3

Figure 6: Histograms of differences between planned and actual delivery durations across sectors*

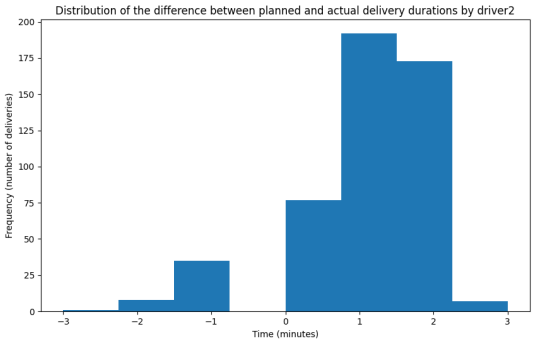
Most of the deliveries are slightly too fast. It is better than being late. The charts show that additional conditions need to be taken into consideration - not only the type of sector.

2.2 Drivers' delivery times

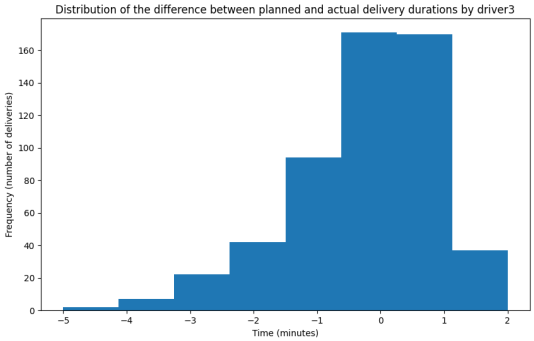
After playing and modeling the data I captured difference between drivers' performances. The positive value on time axis means that the driver was faster than planned.



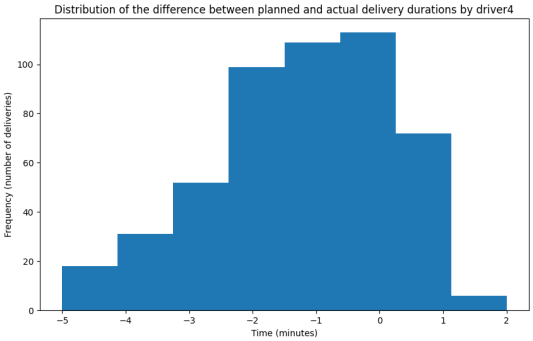
(a) driver 1



(b) driver 2



(c) driver 3



(d) driver 4

Figure 7: Histograms of delivery durations across drivers*

Driver number 4 has the worst performance as I believe that being too late is worse than delivering order too soon (from the customers' perspective). The driver number 3 is the most accurate.

3 Summary

The analysis proved that current algorithm needs to be changed as the errors between planned and actual delivery durations are high. The data shows that there is a lot of aspects to consider to prepare accurate algorithm predicting delivery time. For example sector number or the drivers' experience. As a potential extension, future analysis could explore how delivery times are influenced by the time of day, allowing for even more precise planning and optimization.

* Title and axis names should be bigger or removed.