

# Data analysis and visualisation

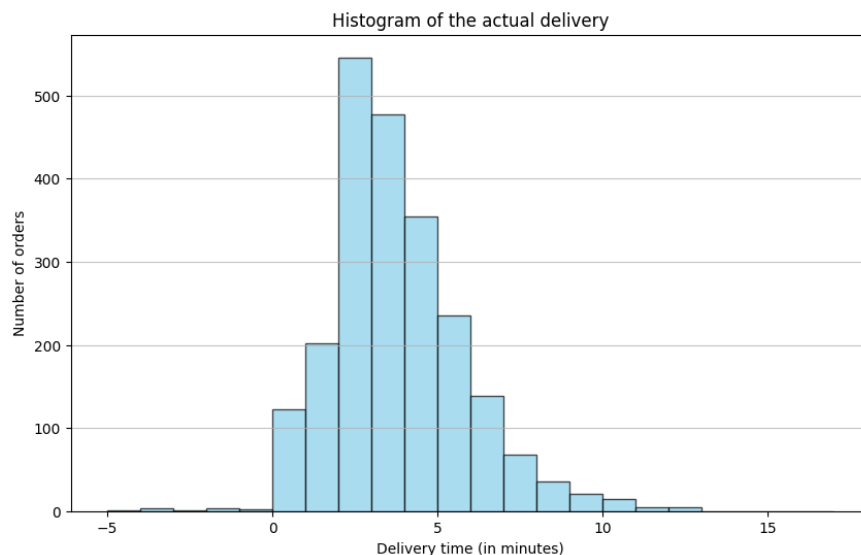
## 1. Introduction:

In this report, a data analysis and visualization of the order delivery process based on the provided dataset is conducted. The aim is to explore trends, identify interesting connections between variables. It will help in forming correct hypotheses about the determinants of delivery time. By thoroughly examining the data, we can gain insights optimize the delivery process. Before the actual analysis we can make an assumption that ..

## 2. Methodology and Data Filtered:

### ***Real delivery time***

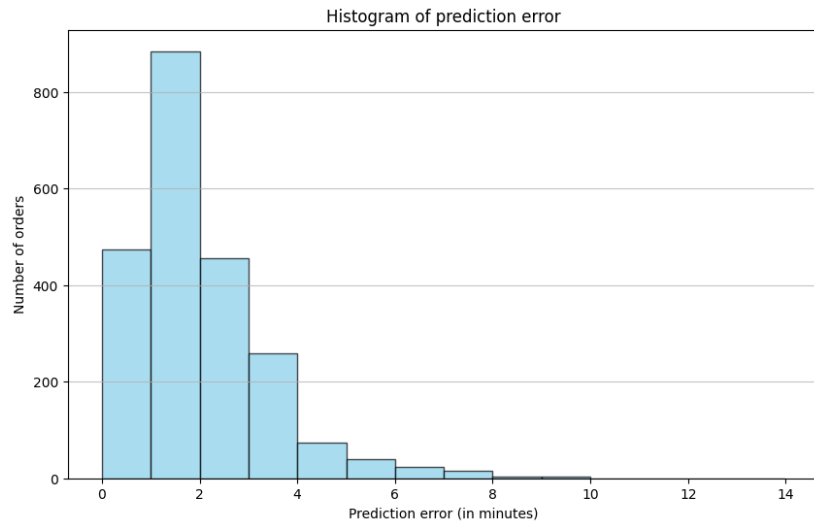
Before conducting the preliminary data analysis, the "route\_segments" table was joined with the "orders" table to calculate the actual delivery time and compare it with the predicted delivery time. The data were filtered because for the problem under consideration, only the time of parcel delivery described by the "STOP" segment type is relevant. Out of 2751 observations marked as "STOP," 494 lacked an assigned order number. This may be due to the driver stopping for a purpose other than delivering an order, so these observations will not be considered in the analysis. Additionally, 38 observations were removed for which the delivery time exceeded 1500 seconds, as they must have resulted from exceptional circumstances, and including them in the analysis could negatively affect its accuracy.



Initially, a histogram showing the delivery time in minutes was generated. For most products, the delivery time does not exceed 10 minutes. Delivery mostly takes 3 minutes. The presented histograms do not contain outliers, where the delivery time was close to 250 minutes.

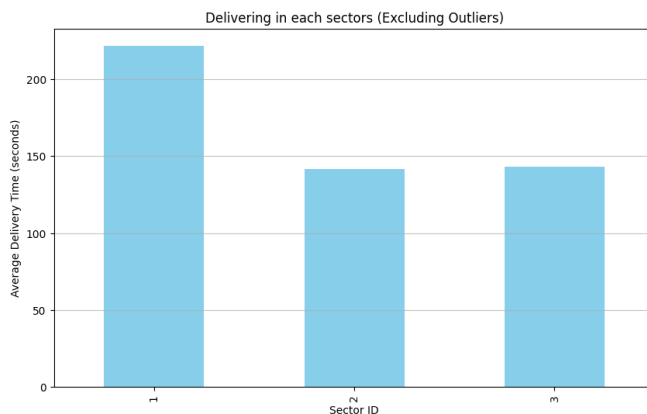
### Prediction error

Histogram analysis for the prediction error also indicates several observations for which the prediction error generated using the mean is significantly higher. The prediction errors range from 0 to 8 minutes, with a concentration of observations with an error of 1-2 minutes.



### Delivery Time by Sector

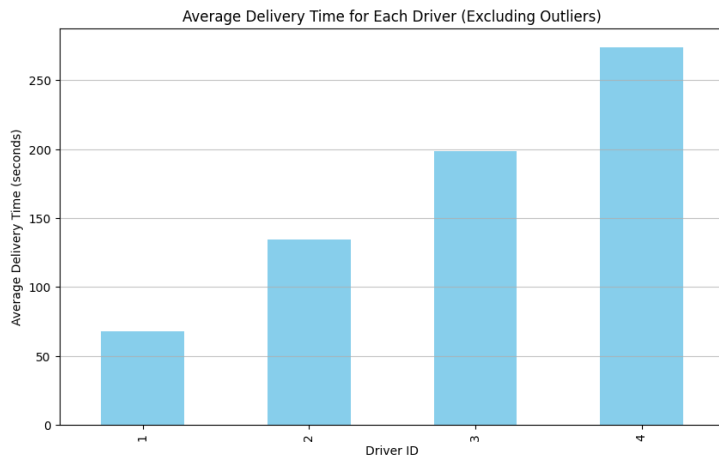
There are noticeable differences in delivery times across different sectors. The most significant difference is mainly observed between Sector 1 and the other two sectors. Thus, the hypothesis that delivery time is longer in one of the sectors is confirmed.



Sector	Average delivery time in seconds
1	222
2	142
3	143

### Delivery Time by Different Drivers

However, the differences in delivery time among individual drivers prove to be more significant. Driver 1 delivers products over twice as fast as Driver 4.



Driver	Average delivery time in seconds
1	68
2	135
3	199
4	274

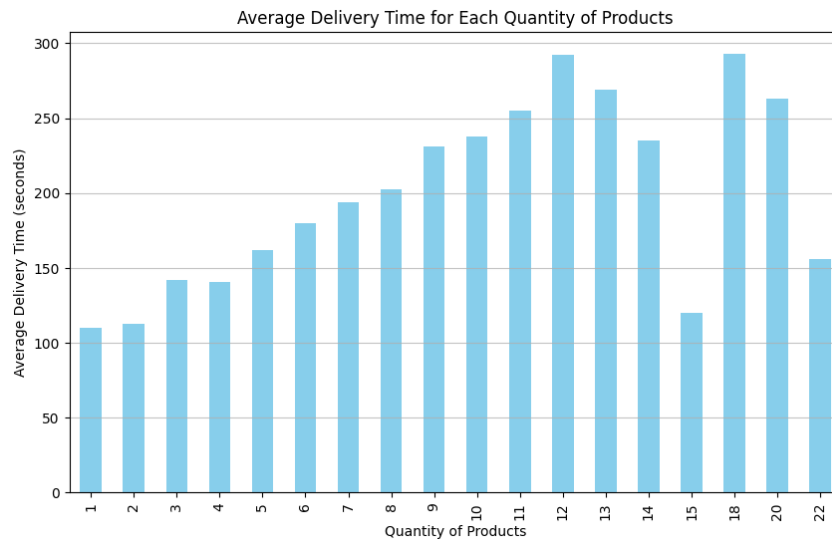
Additionally, we observe a strong negative correlation between the driver ID and delivery time. This may indicate that consecutive driver IDs were assigned as employees were hired, suggesting that drivers with higher IDs are less experienced.

### 3. Other Factors may Influencing Delivery Time

Furthermore, to conduct further analysis, a new table was created to combine information about the actual delivery time with the quantity of ordered products and the total weight of the order. Additionally, the dataset was limited to exclude outliers, defined as observations with a delivery time lower than 0 and exceeding 1500 seconds. Based on the scatter plot, there appears to be a positive relationship between total weight of the order and actual time of delivery. The calculated correlation between them is close to 0,4.



Additionally, there are differences between the quantity of delivered products and the time required for their delivery. We observe a continuous increase in delivery time for orders ranging from 1 to 12 items. Above 12 items, the trend is not noticeable, which may be due to the low number of orders containing such a large number of products. Importantly, there is a positive correlation between the total number of order items and the total weight of the order.



#### 4. Summarizing the Assumptions

- Based on the analyzed data, the following conclusions emerge:
- Orders from Sector 1 take significantly more time than those from the other two sectors.
- Drivers with higher ID numbers require more time to deliver orders.
- The number of products in an order and the total weight are positively correlated with the time required for delivery. This means that the heavier the order, the more time is needed for its delivery.