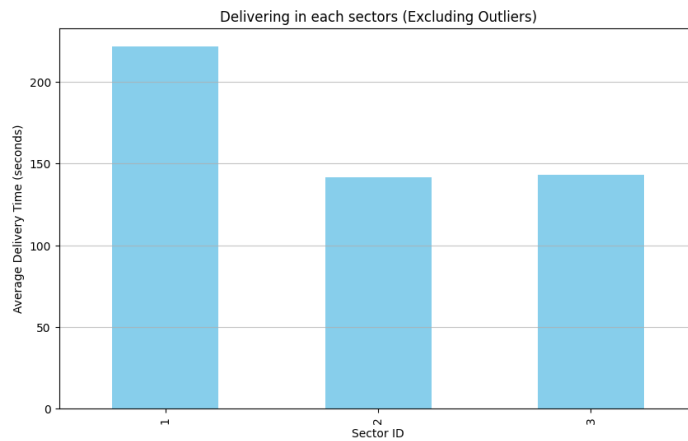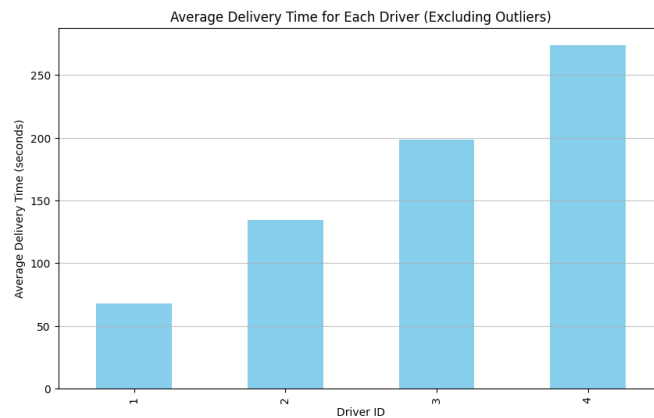# Building and verifying the hypothesis

1. In the first part of the report, a preliminary data analysis was presented. It allowed for the extraction of certain assumptions:
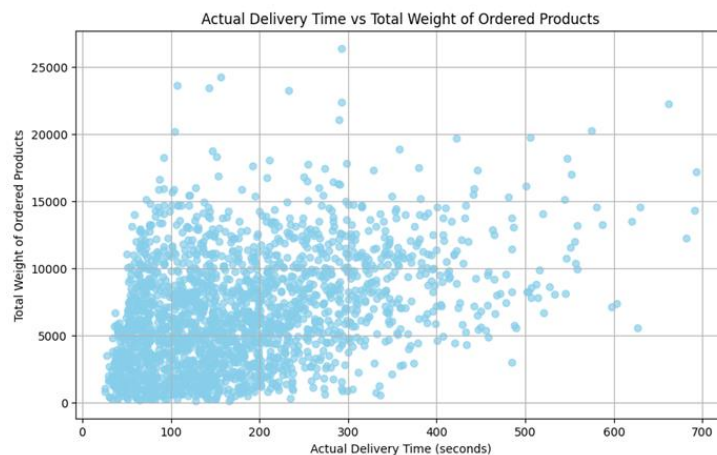
a. The delivery time depends on the sector.

b. The delivery time depends on the driver.

c. The delivery time depends on the total weight of the order

Various factors influencing the delivery time of orders indicate that using the mean for prediction may not be the most effective approach, as this method does not consider the specific characteristics of each order, such as which sector the order pertains to or which driver is delivering it. Based on graphical presentations, the hypothesis that the sector influences the delivery time was verified. Additionally, it was demonstrated that the delivery time of orders varies depending on the driver. Through preliminary data analysis, it was also confirmed that the weight of the order and the number of ordered products affect the delivery time.

These hypotheses can also be confirmed using appropriate statistical tests such as Analysis of Variance (ANOVA) to determine if the mean delivery time differs across different categories. By conducting a T-test, we can verify the statistical significance of weight and calculate how delivery time changes if weight changes by 1.

2. For an accurate estimation of delivery time, it would be reasonable to construct a linear regression model. In this model, sector ID and driver ID variables would be converted into binary variables. It seems that the weight variable can be directly incorporated into the model without transformation, as it appears to have a linear impact on delivery time. Model validation could be performed by separating training and validation observations and calculating prediction error.

Proposal for Model:

The expected delivery time will be equal to the arithmetic mean of 3 values:

Adjusted Delivery Time Based on Sector and **Adjusted Delivery Time Based on Driver, Adjusted Delivery Time Based on Weight.**

Adjusted Delivery Time Based on Sector (a):

- If the delivery is in Sector 1, add 50 seconds to the mean of delivery time.
- If the delivery is in Sector 2 or Sector 3, subtract 30 seconds from the mean delivery time.

Adjusted Delivery Time Based on Driver (b):

- If the driver is Driver 1, subtract 100 seconds from the adjusted delivery time.
- If the driver is Driver 2, subtract 35 seconds from the adjusted delivery time.
- If the driver is Driver 3, add 100 seconds to the adjusted delivery time.
- If the driver is Driver 4, add 35 seconds to the adjusted delivery time.

Linear adjusting of Delivery Time Based on Weight of the order (c):
- 0.02 * Weight of delivery

$$\text{\textit{Predicted delivery time: }} \frac{a+b+c}{3}$$

3. Factors Contributing to Longer Delivery Times:

Challenges in Locating Addresses: Deliveries may take longer due to difficulties in finding specific flats, especially those located on higher floors or in buildings with long corridors.

Security and Access Restrictions: Some delivery locations may have security measures or access restrictions that can impede the delivery process, leading to delays.

4. More specific information about location for example:

- **Building Infrastructure Details:** Gathering specific information about building such as the presence of elevators, staircases
- **Floor Level of Delivery Addresses:** Recording the floor levels of delivery addresses provides insights into the logistical complexities associated with higher or lower floors, including elevator usage and delivery time variations.
- **Distance from Nearest Parking Facilities:** Collecting data on the proximity of delivery locations to parking areas can help estimate the time needed for delivery.

1. Risks of Estimating Delivery Times:

   Under-Estimating:

   - Some groceries, such as frozen or refrigerated items, may deteriorate if delivery times are underestimated, and proper temperature control measures are not implemented during transit.
   - Customers may experience dissatisfaction with the service if they have to wait longer than expected for delivery. This can lead to a negative perception of the company's service quality.

   Over-Estimating:

   - Overestimating delivery times may result in underutilization of delivery resources. Company may allocate more time and resources for deliveries than actually required. This can lead to increase in operational costs and reduce profitability.
   - Underestimating delivery times may lead to customers being absent when delivery attempts are made, further contributing to dissatisfaction and potential delivery complications.