## Part 3. Building and verifying the hypothesis

Methodology

To validate the hypothesis, I generated delivery time estimates based on historical data. I utilized current delivery data as my primary source. Additionally, I established a static mapping correlating the 'sector_id' with the average delivery time per sector.

Part 3.1. Planned delivery per sector

planned_delivery_per_sector_in_minutes — actual_delivery_duration_in_minutes
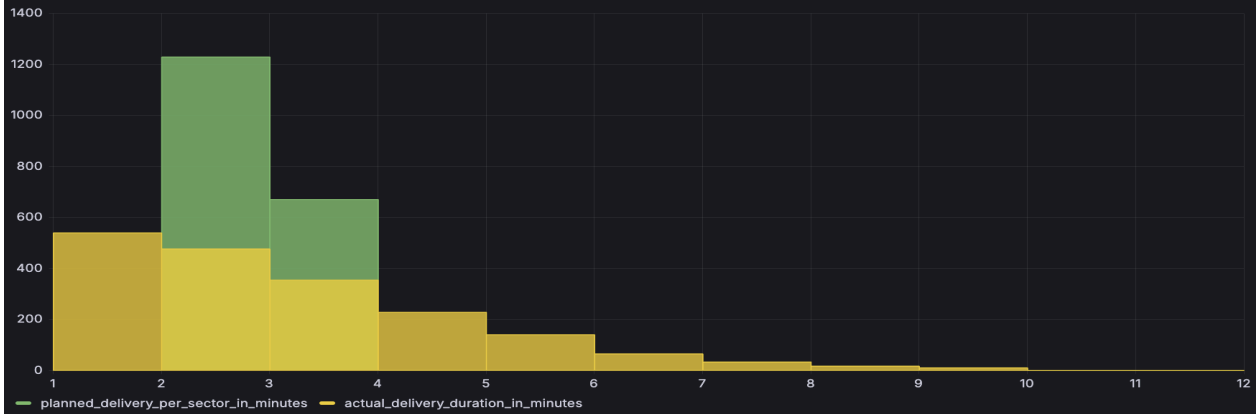
Part 3.1. Sector planned duration comparison

- It looks like taking everate delivery per sector does not improve so much the planned delivery time

```
SELECT
CASE o.sector_id
        WHEN 3 THEN 2.38
        WHEN 2 THEN 2.31
        WHEN 1 THEN 3.54
        ELSE NULL
    END AS planned_delivery_per_sector_in_minutes,
  TIMESTAMPDIFF(
    MINUTE,
    rs.segment_start_time,
    rs.segment_end_time
  ) AS actual_delivery_duration_in_minutes,
  o.planned_delivery_duration / 60 AS planned_delivery_duration_in_minutes
FROM
  route_segments AS rs
  JOIN orders AS o
WHERE
  TIMESTAMPDIFF(
    MINUTE,
    rs.segment_start_time,
    rs.segment_end_time
  ) > 0
AND
    TIMESTAMPDIFF(
    MINUTE,
    rs.segment_start_time,
    rs.segment_end_time
  ) < 200
AND o.order_id = rs.order_id
```

Part 3.2. Predicted delivery time per driver

planned_delivery_per_driver_in_minutes — actual_delivery_duration_in_minutes
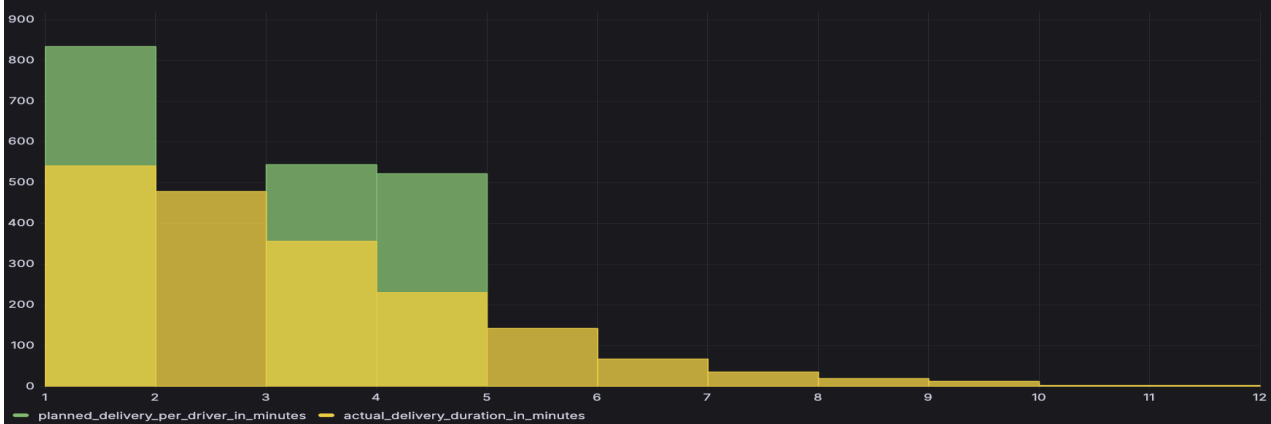
Part 3.2. Delivery duration analyse per driver

- It looks like predicting the delivery time base on driver give as a better estimation.

```sql
SELECT
  CASE
    rs.driver_id
    WHEN 4 THEN 4.35
    WHEN 3 THEN 3.03
    WHEN 2 THEN 1.96
    WHEN 1 THEN 1.11
    ELSE NULL
  END AS planned_delivery_per_driver_in_minutes,
  TIMESTAMPDIFF(
    MINUTE,
    rs.segment_start_time,
    rs.segment_end_time
  ) AS actual_delivery_duration_in_minutes
FROM
  route_segments AS rs
  JOIN orders AS o
WHERE
  TIMESTAMPDIFF(
    MINUTE,
    rs.segment_start_time,
    rs.segment_end_time
  ) > 0
  AND TIMESTAMPDIFF(
    MINUTE,
    rs.segment_start_time,
    rs.segment_end_time
  ) < 200
  AND o.order_id = rs.order_id
```

Part 3.3. Why could some deliveries take more time?

1. **Limited Access:** Some buildings may have restricted access to certain floors or areas, requiring delivery personnel to navigate through security checkpoints or obtain special permissions before completing the delivery.
2. **Parking Constraints:** Limited parking availability near the building can make it challenging for delivery vehicles to unload packages efficiently, resulting in delays as drivers search for suitable parking spots or wait for loading zones to become available.
3. **Weather Conditions:** Adverse weather conditions, such as extreme heat or cold, can exacerbate the physical strain on delivery personnel when navigating stairs, potentially leading to slower delivery times.

Part 3.4. What additional data would be worth collecting for future analysis of this domain?

1. **Building Characteristics:** Information about the building's structure, such as the number of floors, presence of elevators, access points, and any accessibility features, can help delivery companies better understand the logistical challenges associated with each location.
2. **Weather Data:** Gathering historical weather patterns can provide insights into delivery trends, and extreme weather events, which can influence delivery time. This data can include temperature, rain and snow.
3. **More accurate Geographical Data:** Analyzing geographical data such as building locations, traffic patterns, and proximity to delivery hubs can inform route planning and resource allocation strategies.

Part 3.5. What is the risk of over- or under-estimating the delivery times?

Overestimating Delivery Times:

1. **Customer Dissatisfaction:** Customers may become frustrated or dissatisfied if they perceive delivery times as excessively long. Overestimating delivery times can lead to lost sales opportunities and damage to customer relationships.
2. **Reduced Efficiency:** Overestimating delivery times may lead to inefficient resource allocation and underutilization of transportation and logistics assets. This can negatively impact operational efficiency and increase costs.

Underestimating Delivery Times:

1. **Customer Disappointment and Complaints:** Customers may experience disappointment or frustration if deliveries arrive later than promised. This can lead to complaints, negative reviews, and damage to the company's reputation.
2. **Missed Deadlines:** Underestimating delivery times increases the risk of missing deadlines, especially for time-sensitive orders or projects. This can result in contractual penalties, loss of future business opportunities, and damage to business relationships.
3. **Increased Transportation Costs:** Rushing deliveries to meet underestimated timelines may require expedited shipping methods, which can be significantly more expensive than standard shipping options. This can erode profit margins and increase overall transportation costs.