Delivery Time Analysis Report

Natalia Klinik

Introduction

The purpose of this analysis was to explore how well current delivery time predictions match the actual times, and to identify which factors might influence delivery duration. The analysis is based on real delivery data, focusing only on delivery moments — that is, route segments labeled as "STOP".

Data Loading and Preparation

To avoid exposing sensitive information, database connection details were stored in a secure .env file. Data was extracted from the *droptime database* using Python and saved into separate CSV files.

The relevant tables were loaded into dataframes and inspected for structure and quality. The main analysis was based on a merged dataset combining information from the *route_segments* and orders tables using the shared *order_id*.

I focused only on rows where the segment type was "STOP", as these represent the moment when a driver actually delivers the order.

I also:

- Converted delivery timestamps to datetime format.

- Calculated actual delivery duration in minutes.

- Removed rows with missing key values to ensure accuracy.

- Converted the predicted delivery time (originally in seconds) to minutes for better readability.

Visualizations and Findings

1. Actual Delivery Time

A histogram shows that the vast majority of deliveries are completed within 20 minutes. Although there are a few outliers (up to 250 minutes), they represent rare cases and are not necessarily errors.

Histogram of Actual Delivery Time


Histogram of Actual Delivery Time

Key takeaway: Most deliveries are quick and within expected timeframes.

2. Prediction Error

I plotted the difference between predicted and actual delivery times. Most deliveries are within ±10 minutes of the prediction, with many even arriving earlier than expected.

Outliers exist but are uncommon. Limiting the graph to the -20 to +20 minute range helped visualize the bulk of the data more clearly.

Histogram of the prediction error

Histogram of the prediction error
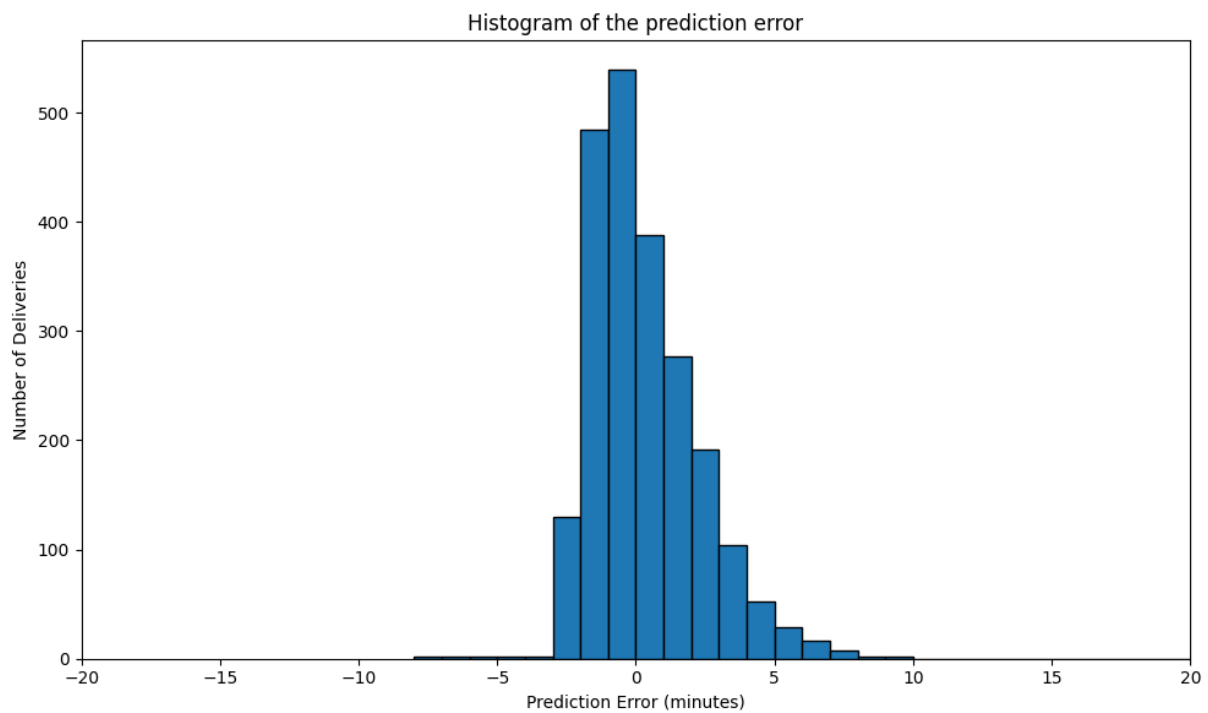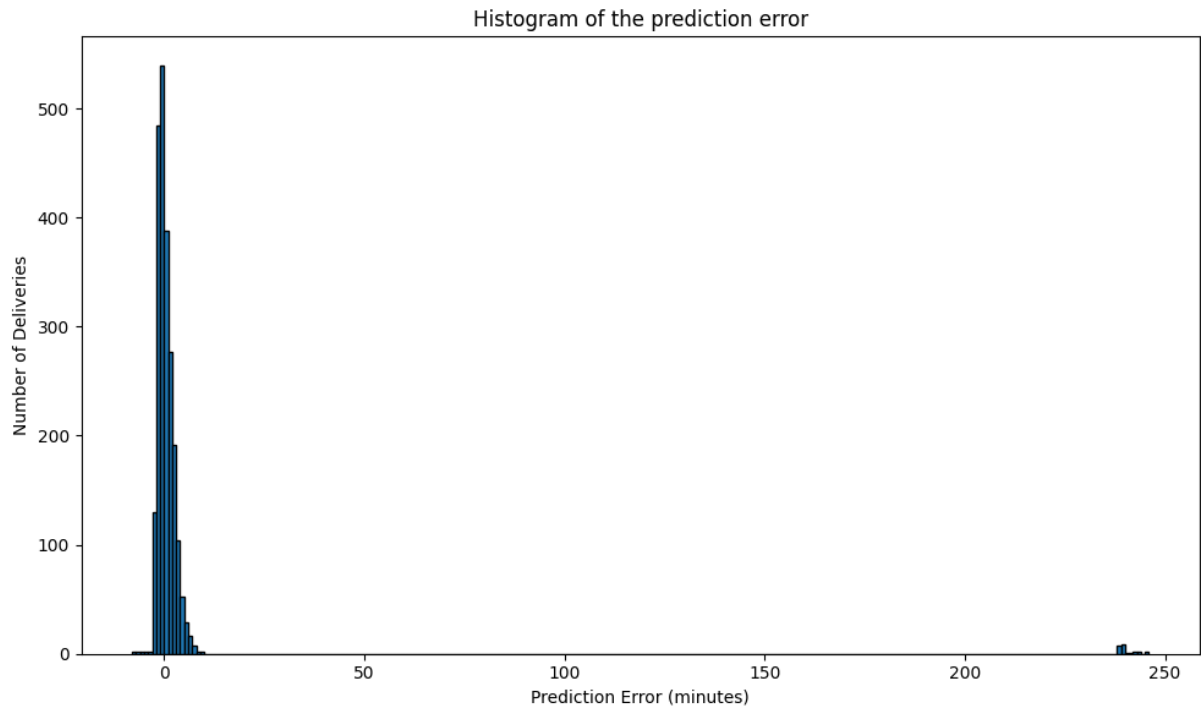
Key takeaway: The current prediction system is mostly accurate, but sometimes underestimates delivery duration.

3. Average Delivery Time by Sector

I compared delivery durations across the three delivery sectors. Although Sector 3 appears slightly slower, the difference is minor — less than one minute on average.
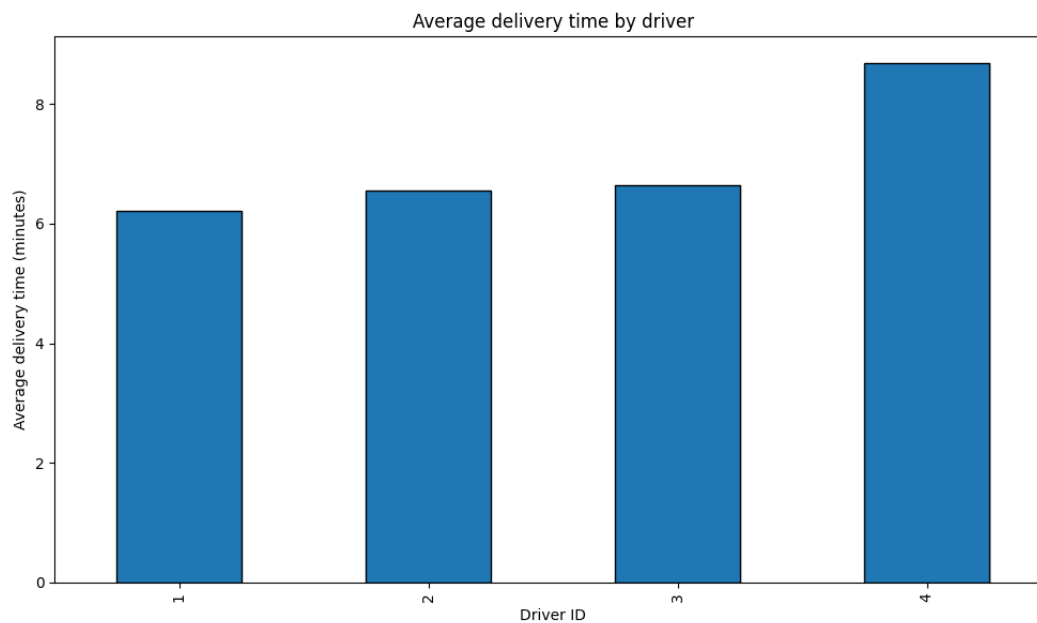

Average Delivery Time by Sector
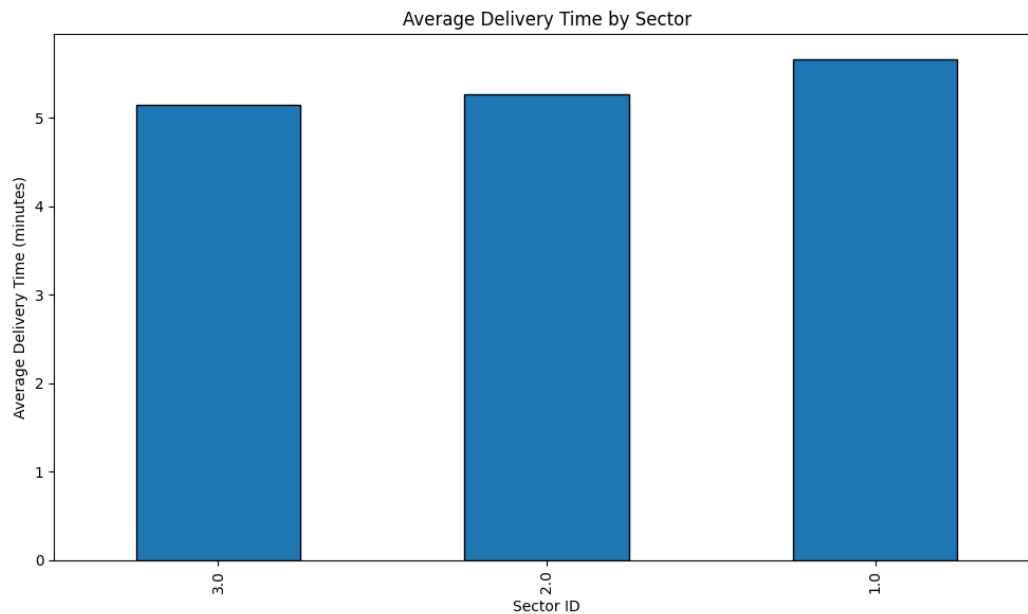
Key takeaway: Sector has minimal influence on delivery duration.

4. Additional Analyses

4.1. Delivery Time by Driver

Some drivers consistently take longer to complete deliveries. For example, Driver 4 stands out with significantly longer times.
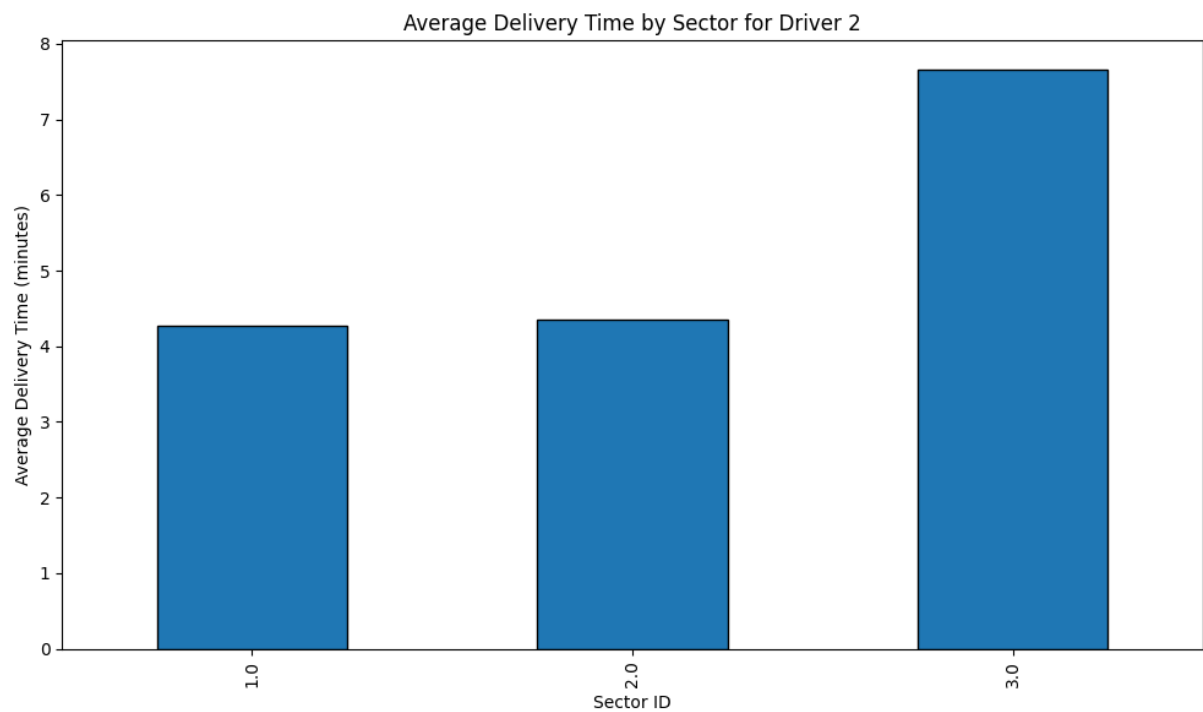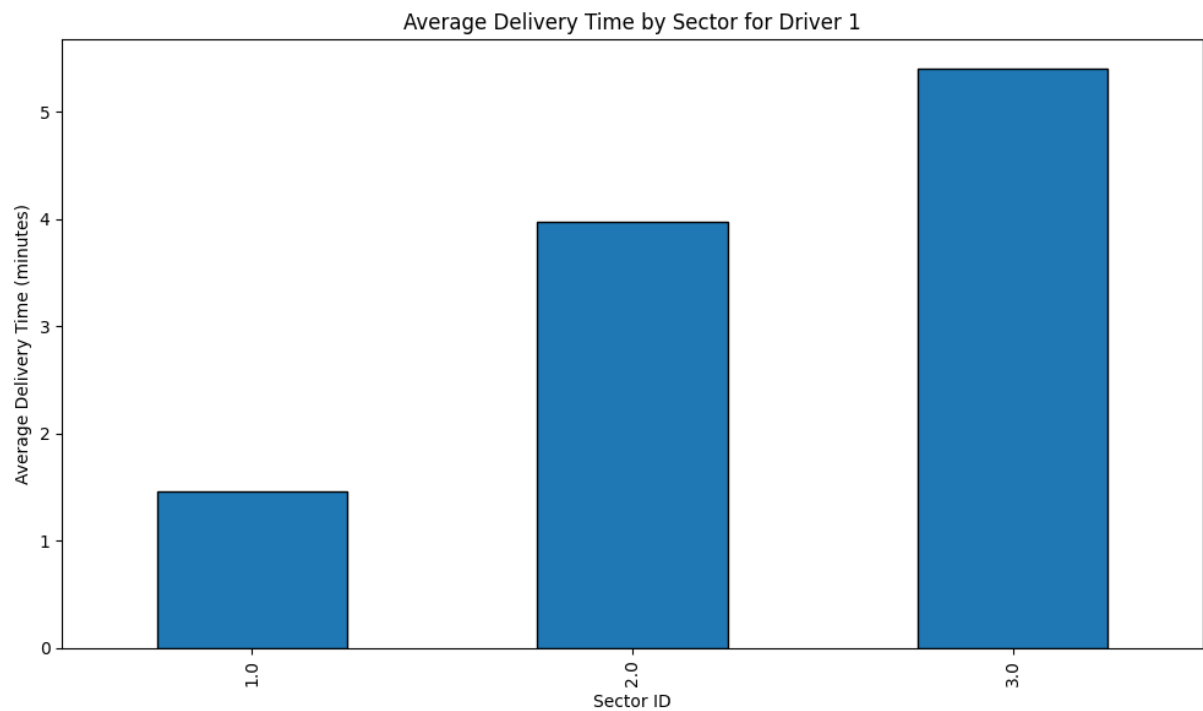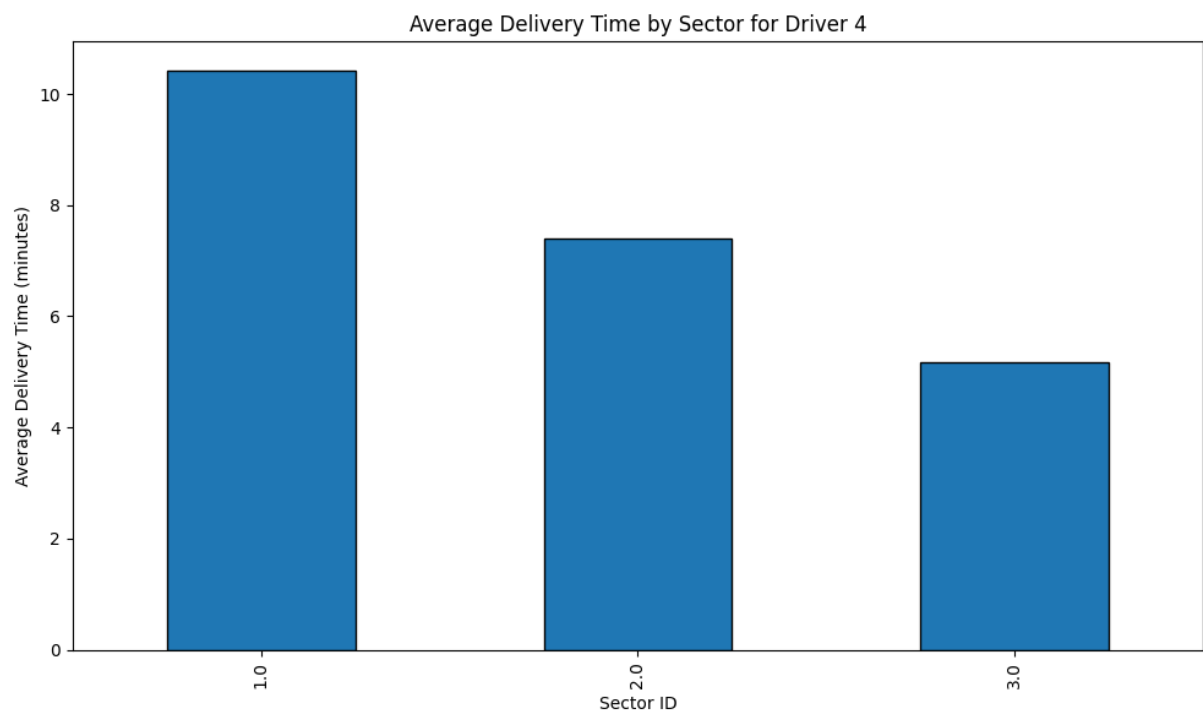

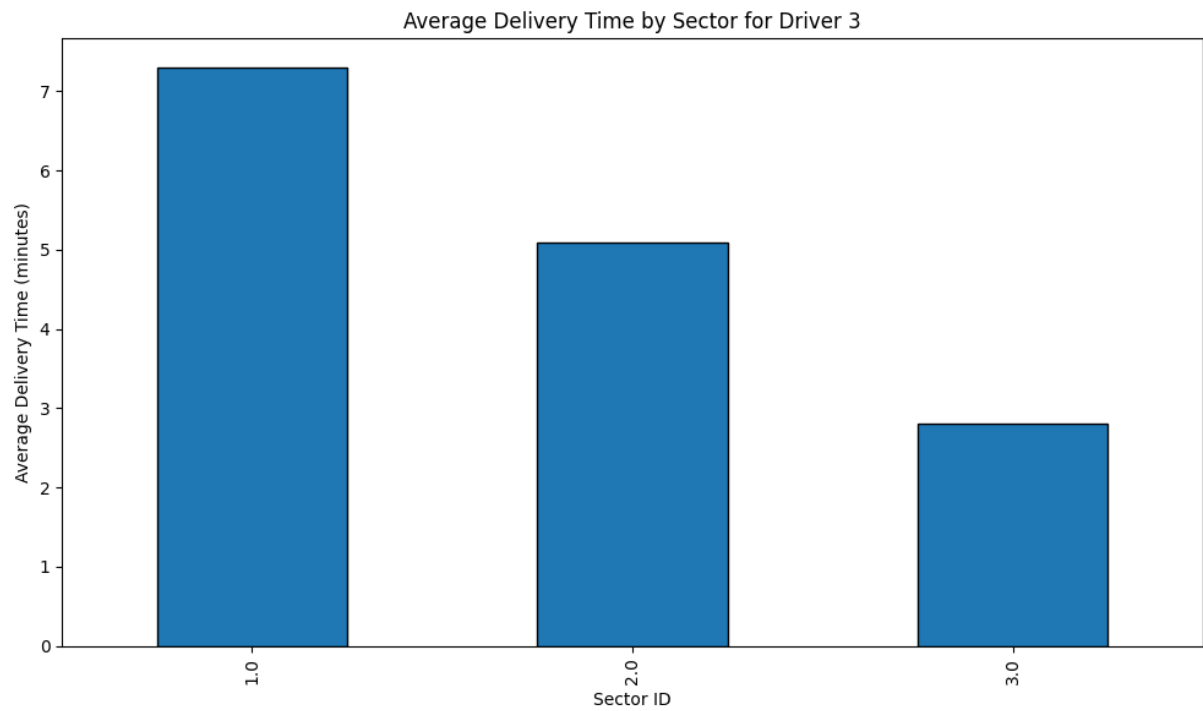Average delivery time by driver

Key takeaway: Individual driver performance can affect delivery time.

4.2. Driver + Sector Combination

By analyzing driver performance within each sector, we noticed patterns:

- Drivers 1 and 2 are slower in Sector 3.

- Drivers 3 and 4 are slower in Sector 1.

Average Delivery Time by Sector for Driver 1

Average Delivery Time by Sector for Driver 2

Average Delivery Time by Sector for Driver 3

Average Delivery Time by Sector for Driver 4

Key takeaway: Performance differences may depend on both driver and delivery area.

4.3. Delivery Time vs. Order Weight

Orders were categorized into:

- Small (<2kg)

- Medium (2–5kg)

- Large (>5kg)

Relationship Between Order Weight and Delivery Time

Average Delivery Time by Order Weight Category

As expected, large orders took longer. However, medium orders were delivered faster than small ones, which suggests other factors also play a role.

Key takeaway: Heavier orders generally take longer, but not always — other variables matter too.

4.4. Delivery Time by Day of the Week

We grouped deliveries by weekday and found slightly longer times on Tuesdays. However, no consistent pattern was found across the week.
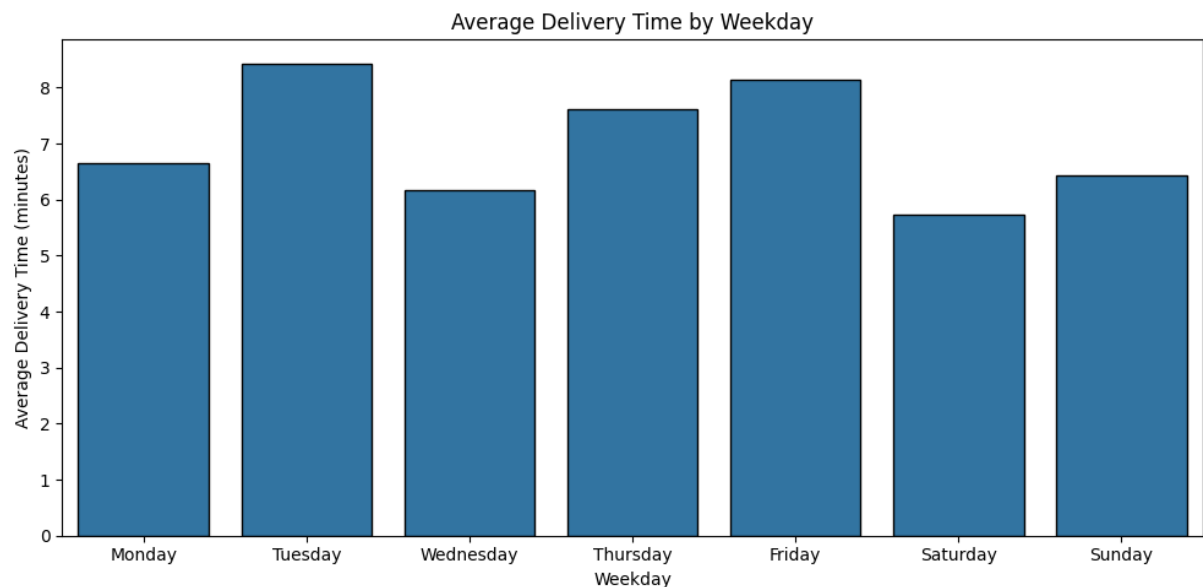
Key takeaway: Weekday does not significantly affect delivery time.

4.5. Delivery Time vs. Product Quantity

We found no strong correlation between how many items were in the order and how long the delivery took.
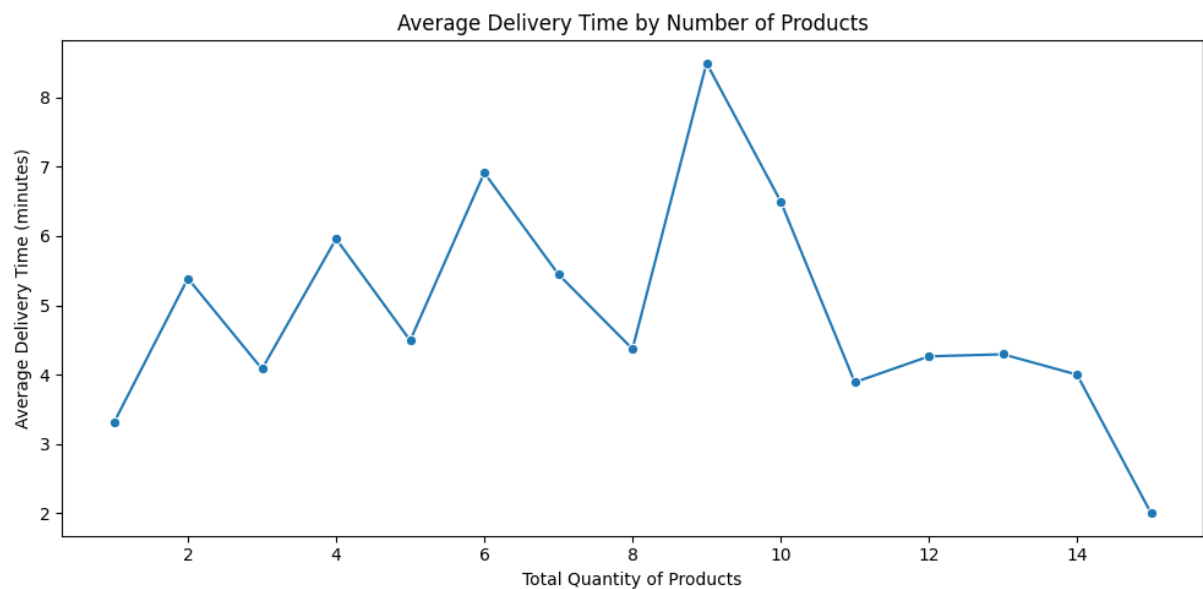
Key takeaway: Quantity of products does not strongly impact delivery duration.

4.6. Correlation Matrix

We created a correlation matrix to explore how delivery duration relates to:

- Driver ID

- Order weight

- Product quantity

- Sector

- Day of the week

Correlation Matrix: Delivery Duration vs Features
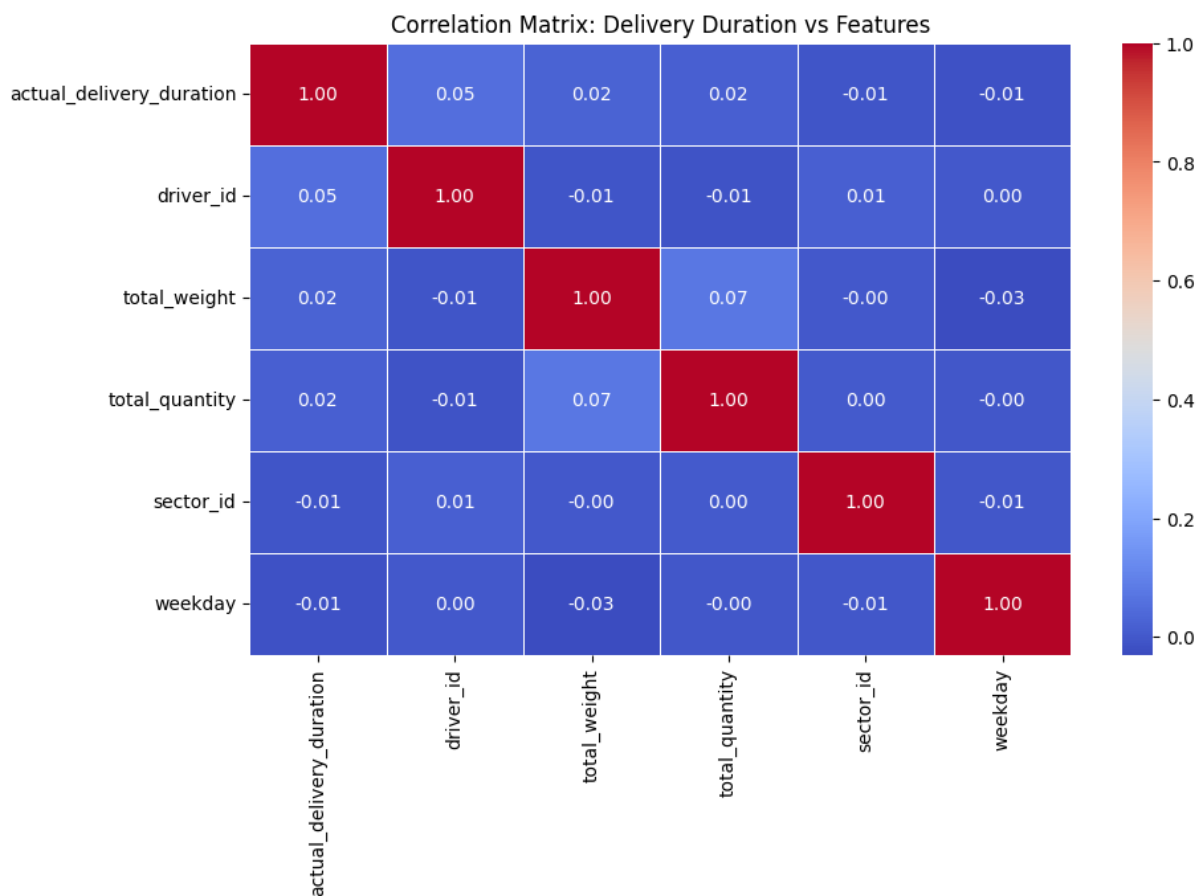
No individual factor showed a strong correlation. This suggests that delivery time is influenced by a combination of factors rather than one dominant variable.

Conclusion

The current delivery time prediction system performs fairly well, with most predictions close to actual results. However, there's room for improvement — especially in edge cases involving specific drivers or heavier orders.

Key insights:

- Most deliveries are completed within 20 minutes.

- Prediction errors are usually small, but outliers exist.

- Driver behavior impacts delivery time more than location or weekday.

- Order weight plays a role, but not consistently.

- No single factor determines delivery time — a more complex prediction model could improve accuracy.

Note: For more detailed and technical analysis check the *part2.ipynb* file. It can be found in the *solutions/raw/* directory.