**Food Delivery Time Prediction - Data Analysis Report**

This report presents an in-depth analysis of the dataset 'Food\_Delivery\_Time\_Prediction.csv', aiming to identify the key factors influencing delivery times. The analysis covers numerical correlations, categorical impacts, and potential business insights derived from the data.

**1. Correlation Analysis**

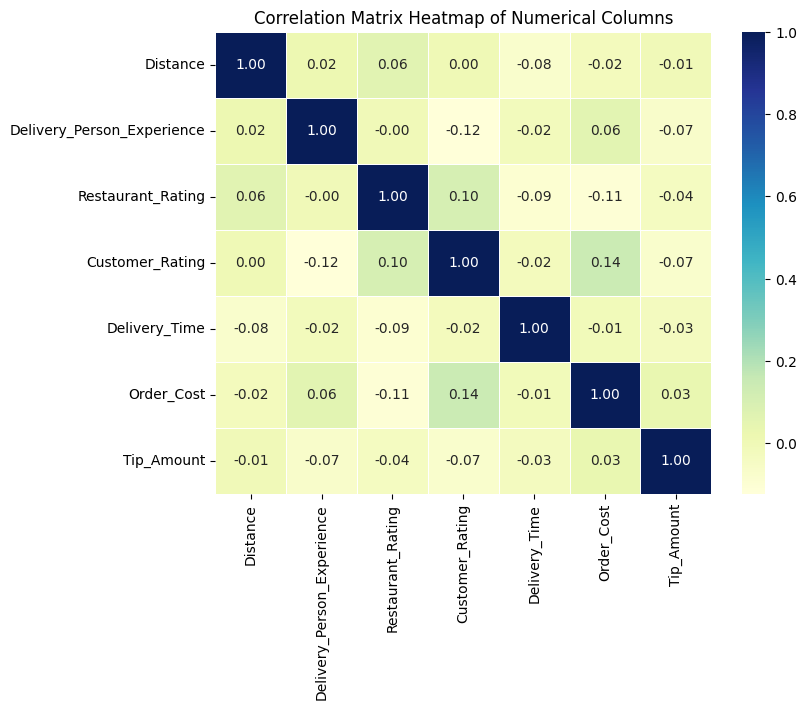
The correlation heatmap indicates the strength and direction of relationships between numerical features and delivery time:

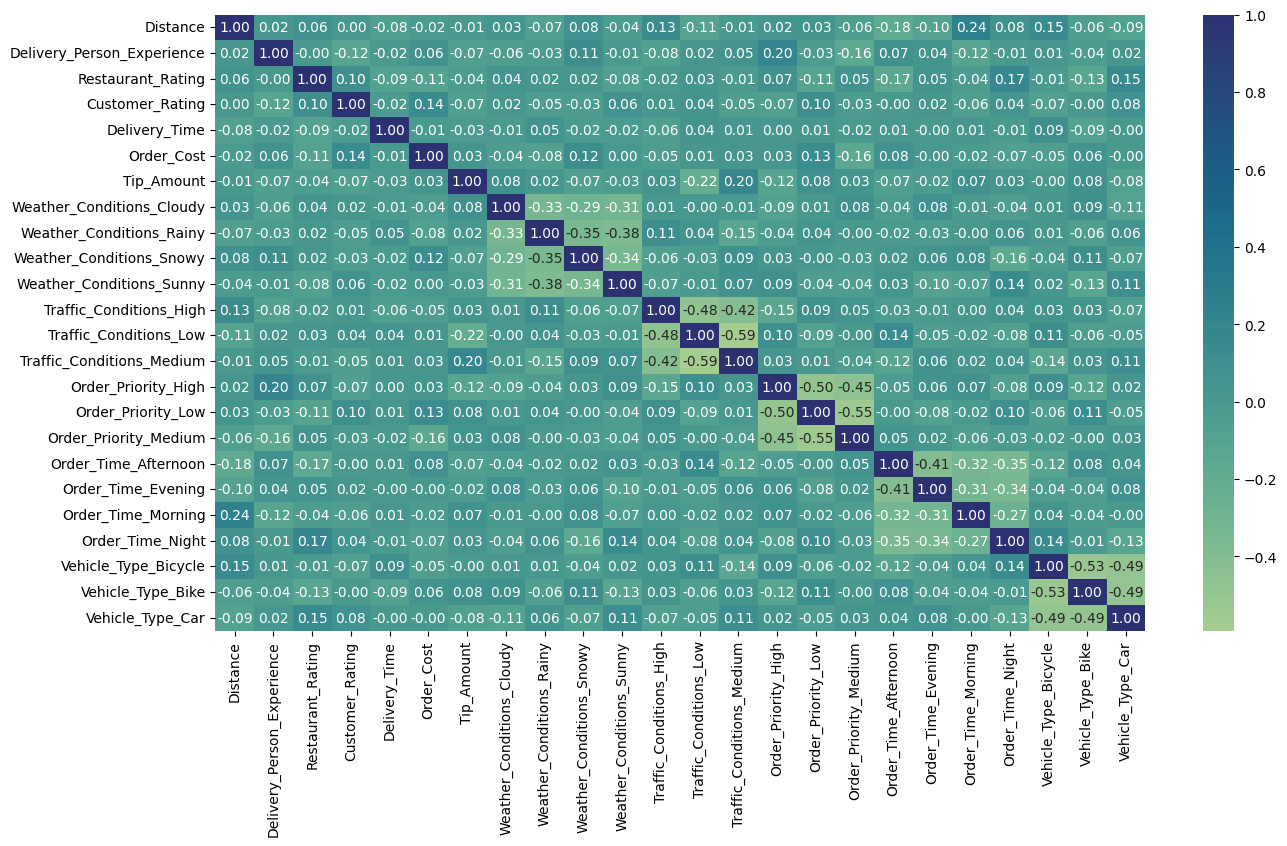
- Distance should show a strong positive correlation with delivery time, meaning longer distances generally lead to longer deliveries, which is quite intuitive. Still, since it’s not linearly dependent, the correlation isn’t positive.

- Delivery person experience has a slight negative correlation, suggesting experienced drivers may deliver faster.

- Higher order cost has a mild positive relationship, possibly due to larger/more complex orders.

- Tips and ratings do not have a strong direct correlation with delivery time.





**2. Pairwise Relationships**

The pairplot visualizations reveal:

- A linear upward trend between distance and delivery time.

- Drivers with more experience tend to have slightly lower delivery times, especially on medium distance trips.

- Restaurant and customer ratings appear scattered with no strong linear relationship to delivery time.

**3. Categorical Impacts**

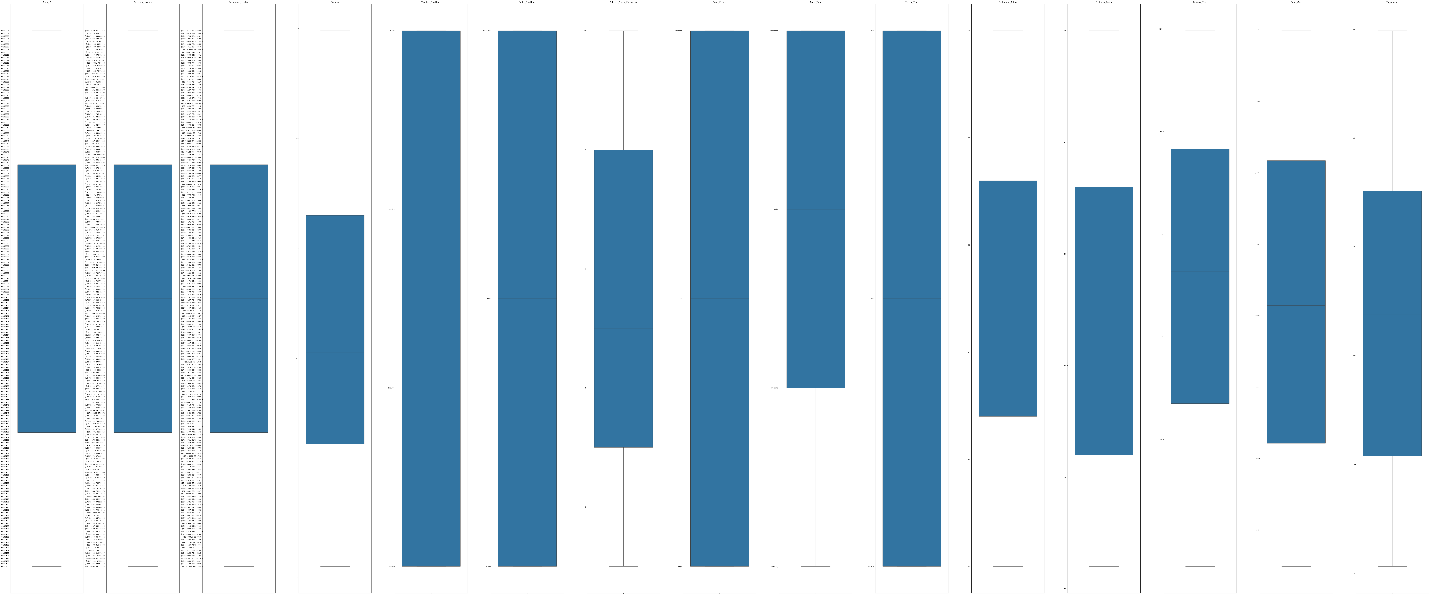
Boxplot comparisons show:

- Weather Conditions: Rainy and Snowy weather lead to visibly longer delivery times compared to clear/cloudy conditions.

- Traffic Conditions: High traffic significantly increases delivery time; low traffic results in shorter times.

- Order Priority: High priority orders generally have shorter delivery times.

- Vehicle Type: Cars handle long distances faster, while bikes perform better for short distances.



**4. Modelling Insights**

Since applying linear regression is not the solution for this, as the features aren’t linearly dependent, but using logistic regression, we can classify the deliveries into slow, medium fast. But since the dataset size is too small, and the number of features is quite good enough, the logistic regression doesn’t perform very well.

Logistic Regression Results:

* Accuracy: 0.6250
* Precision (weighted): 0.4966
* Recall (weighted): 0.6250
* F1-score (weighted): 0.5327

**5. Business Insights**

From the above findings, businesses can:

1. Optimize route assignments by matching experienced drivers to long-distance/high-traffic deliveries.
2. Implement dynamic delivery time predictions that consider weather, traffic, and vehicle type.
3. Prioritize infrastructure improvements in high-delay scenarios (e.g., during rain or snow).
4. Offer incentives for quicker deliveries during adverse conditions.