**Week 3**

**Food Delivery Prediction using Regression Modelling**

**Team Leader**

1. Bhavya Sri

**Team Members**

J. Sinduja

K. Nanda Kishore Reddi

Ejumalla Saikiran

Lakshanya S

Naveen Rampa

Submission Date :

22-12-2023

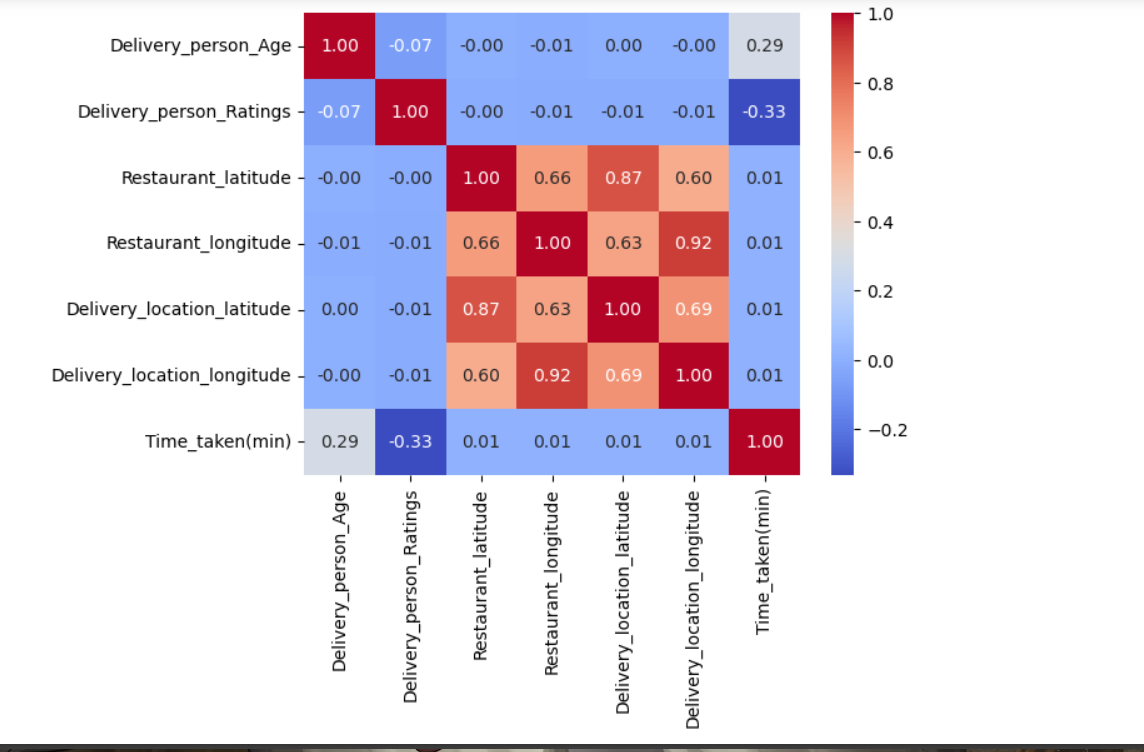
**1. Initial Exploration**

* Loaded the dataset and conducted an initial exploration using describe() and info().
* We ensured that there were no missing values in the dataset.

**2. Exploratory Data Analysis**

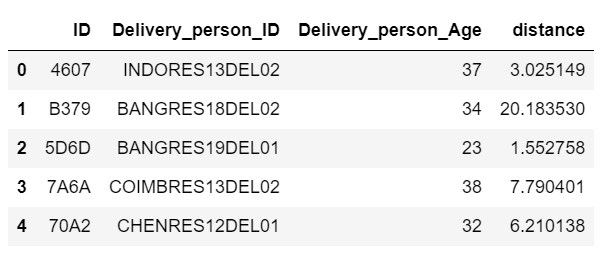
2.1.Correlation matrix

The dataset's exploratory data analysis reveals strong positive correlations, like 0.66 between restaurant latitude and longitude, 0.87 between delivery location latitude and restaurant latitude, and a significant 0.92 between restaurant and delivery location longitudes, indicate the importance of geography in food delivery predictions, suggesting potential implications for service efficiency in the industry. Further investigation into these relationships could offer valuable insights for improving overall delivery processes.



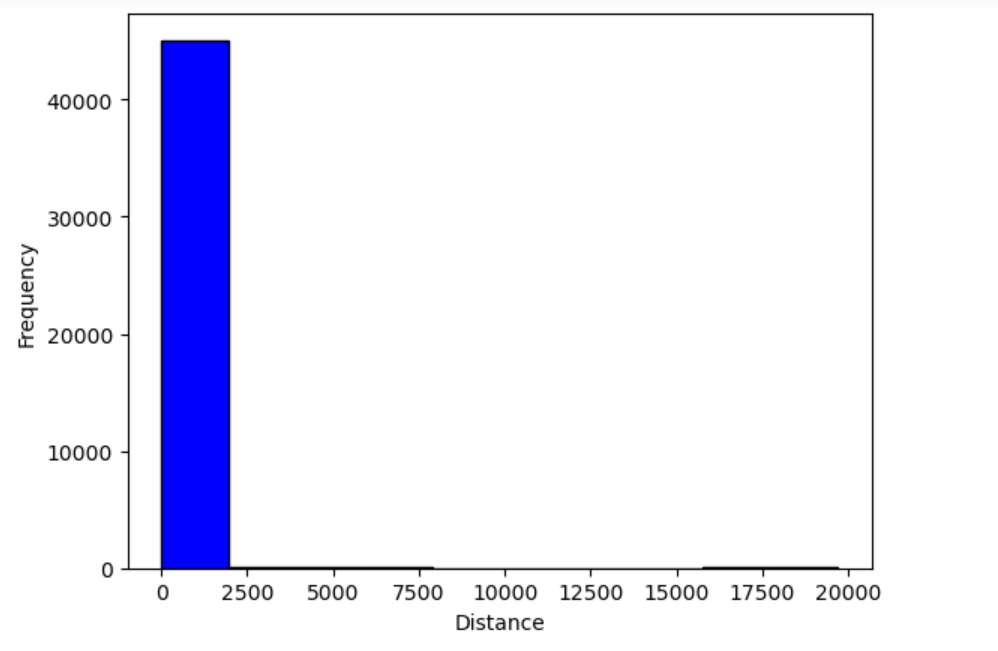
2.2 Calculating Distances

This 'distance' column could be useful in a food delivery prediction analysis, as it quantifies the spatial separation between the restaurant and the delivery location. It may help in understanding delivery times, optimizing delivery routes, or factoring distance into predictive models, hence taking advantage of the close relations found in correlation matrix we calculated the distance between the restaurant and the delivery location using the Haversine formula.



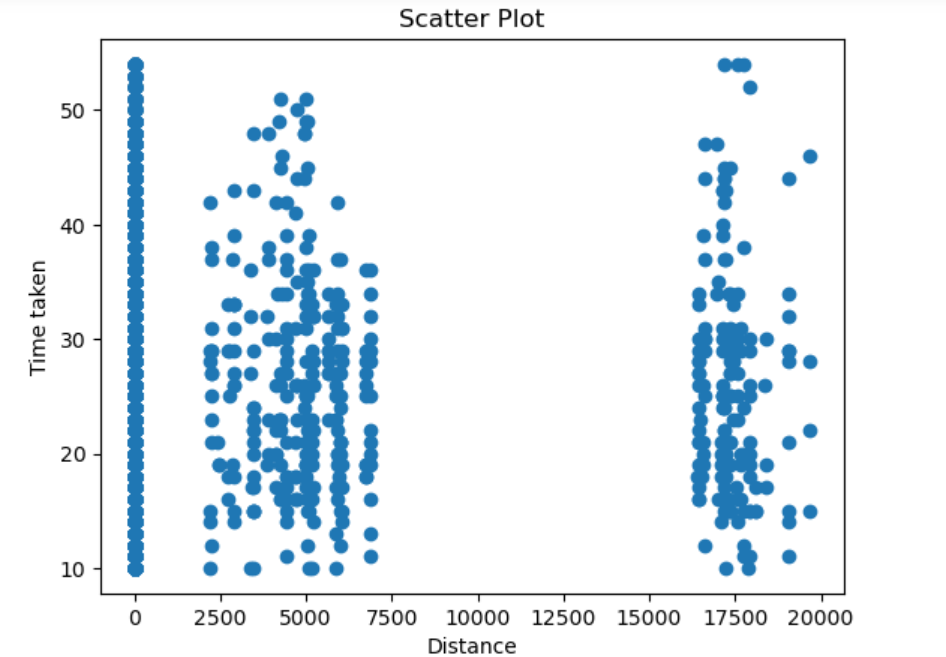
2.3 Spatial Analysis of Order Distribution

The histogram shows that most orders come from places within 0-2300 km, suggesting a lot of customers or a focus on nearby areas. As you go beyond 2300 km, orders drop. In particular, places between 2400-7600 km and 15500-19500 km see fewer orders. This spatial insight informs potential geographical preferences and aids in optimizing delivery strategies.



2.4 Scatterplot Analysis

The Scatterplot visualizes the relationship between the distance of delivery locations and the corresponding time taken for delivery. The data reveals a range of delivery times, spanning from a minimum of 10 minutes to a maximum of 54 minutes.



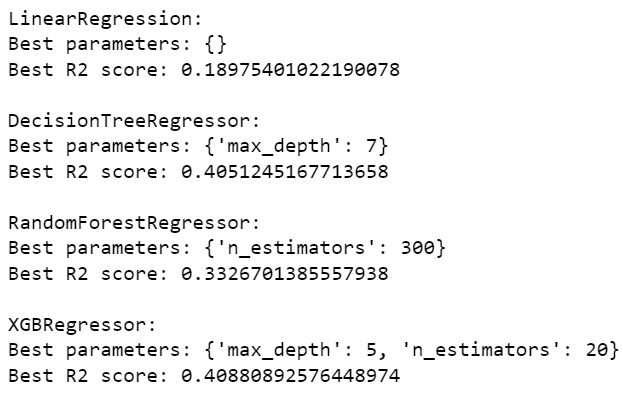
**3.Regression Model**

3.1 Splitting and Standardizing

We're preparing our data for making predictions on delivery times. First, we separate our features (like distance, location, etc.) from what we want to predict (the time taken for delivery).Next, we split our data into two parts: one for training our prediction model and the other for testing how well our model works.To make sure our model understands the data well, we use a process called standardization.

3.2 Regression Model Comparison and Optimization

In this analysis, we employed various regression models to predict the time taken for delivery based on input features. The models considered are Linear Regression, Decision Tree Regressor, RandomForest Regressor, and XGBoost Regressor. To enhance their performance, we utilized GridSearchCV to optimize hyper parameters.



|  |  |  |  |
| --- | --- | --- | --- |
| Model | Best Parameters | Best R2 Score | Analysis |
| Linear Regression | None (default settings) | 0.1898 | A basic model but doesn't predict very well; might benefit from using more advanced models. |
| Decision Tree Regressor | {'max\_depth': 7} | 0.4051 | Does better than Linear Regression by considering more factors; a good balance of simplicity and accuracy. |
| Random  Forest Regressor | {'n\_estimators': 300} | 0.3327 | Shows promise, but there's room to make it better; might need some adjustments to reach its full potential. |
| XGBoost Regressor | {'max\_depth': 5, 'n\_estimators': 20} | 0.4088 | Performs the best among all models; strikes a good balance between complexity and accuracy. |

**4.Conclusion**

In conclusion, the regression model evaluation highlighted different strengths and weaknesses. Linear Regression offered a basic understanding but lacked predictive power, suggesting the need for more advanced models. The Decision Tree Regressor improved accuracy by considering tree depth. While Random Forest showed promise, further optimization is possible. XGBoost Regressor stood out as the top performer, achieving a favorable balance between model complexity and predictive accuracy.