Food Delivery Time Prediction Project Report

Objective:
To predict food delivery times using customer and restaurant locations, traffic, weather, and other contextual factors. The goal is to assist delivery operations by accurately forecasting delivery durations and identifying factors that cause delays.

Phase 1: Data Preprocessing and Exploratory Data Analysis (EDA)

1. **Data Overview:**
2. Total Records: 200
3. Columns: 15 (including Order ID, Location coordinates, Weather, Traffic, Ratings, Cost, Tip, etc.)
4. **Missing Values:**
5. No missing values were found in any column.
6. **Categorical Encoding:**
7. Label Encoding applied to categorical variables:
8. Weather_Conditions
9. Traffic_Conditions
10. Order_Priority
11. Order_Time
12. Vehicle_Type
13. **Numerical Feature Scaling:**
14. Standardized using `StandardScaler`:
15. Distance
16. Delivery_Person_Experience
17. Restaurant_Rating
18. Customer_Rating
19. Order_Cost
20. Tip_Amount
21. **Outlier Detection:**
22. Checked using IQR method.
23. No significant outliers found in Distance, Delivery_Time, or Order_Cost.

24. **Feature Engineering:**

25. **Rush_Hour** binary feature was added based on time of order (Evening/Night = 1).
26. **Delivery_Status** binary target variable created:
27. 1 = Delayed
28. 0 = Fast
    (Based on median Delivery_Time)

Phase 2: Predictive Modeling

Linear Regression

- **Goal:** Predict actual Delivery Time
- **Train-Test Split:** 80/20
- **Features Used:** All numeric + encoded categorical + Rush_Hour

Evaluation: - Mean Squared Error (MSE): 962.66 - Mean Absolute Error (MAE): 26.49 - R-squared (R^2): -0.04 *(Model is not performing well)*

Logistic Regression

- **Goal:** Classify deliveries as Fast or Delayed
- **Train-Test Split:** 80/20

Evaluation: - Accuracy: 45% - Precision: 48% - Recall: 57.1% - F1 Score: 52.2% - AUC Score: 0.46

Visualizations: - Confusion Matrix displayed delivery classification - ROC Curve plotted to assess classification threshold tradeoffs

---

Phase 3: Recommendations & Insights

- Linear regression performs poorly, indicating a weak linear relationship. Future models should consider non-linear methods.
- Logistic regression shows moderate potential but requires improvements.

Suggestions: 1. Use **Haversine distance** from lat/lon instead of provided Distance column. 2. Try **tree-based models** like Random Forest or XGBoost. 3. Collect more features like time taken by delivery personnel, historical delivery logs. 4. Use **external APIs** (like Google Maps or Weather) to enrich real-time data. 5. Consider adding **weekend vs weekday**, **holiday**, and **city tier** information.

Deliverables: - Jupyter Notebook: Includes full data cleaning, modeling, evaluation, and plots - Visuals: Confusion Matrix, ROC Curve - This report

---

Conclusion: The models highlight some of the challenges in predicting food delivery time with limited features. With better feature engineering and advanced models, prediction accuracy can be significantly improved.