Zomato Delivery Operations Predictive Analytics

Presented By Dinda Raraswati

TABLE OF CONTENT

ABOUT ME

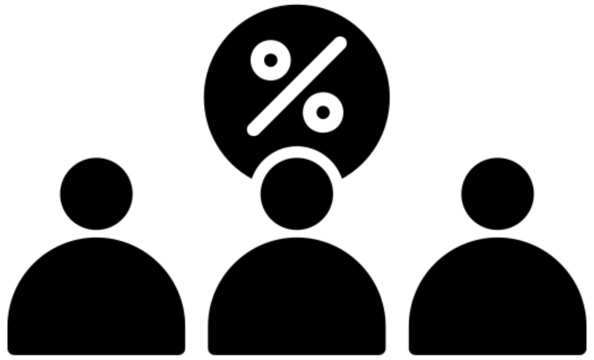
SELF-OVERVIEW
A data enthusiast with a background in Agricultural Engineering who is currently transitioning from academia to industry

EDUCATION
- **Bachelor of Science in Agricultural Engineering (2016 - 2020)**
  Bandung Institute of Technology (ITB)
- **Master of Agricultural Science (2021 - 2023)**
  Kyoto University
- **Data Science Bootcamp (Apr 2025 - present)**
  dibimbing.id

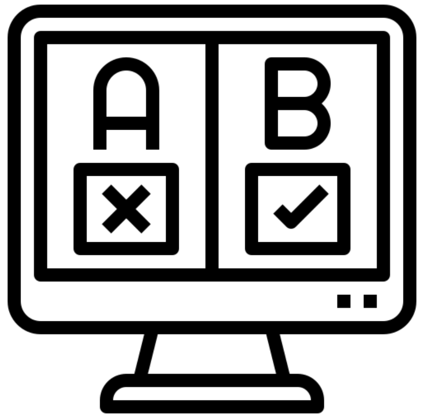
WORKING EXPERIENCE
- **Wageningen Food Safety Research (WFSR) (Nov 2023 - Aug 2025)**
  Researcher
- **Climate Change Center ITB (PPI-ITB) (Dec 2020 - Apr 2021)**
  Project Assistant

PREVIOUS PROJECTS

E-commerce Transaction Analytics
Analyze sales pattern at e-commerce



Bank Customer Churn Prediction
Develop customer churn prediction model using classification algorithms



A/B Testing on Landing Page Designs
Conduct A/B testing to evaluate the effectiveness of different landing page designs on speaker sales

Customer Segmentation of Airline Passengers
Segment airline passengers using K-Means clustering

EXECUTIVE SUMMARY

Problem Statement

Zomato is **India's #1 food delivery app** with market share in India reaching 58% in Q1 2025. While its competitor offer quick delivery, this company has been struggling with delivery time.

Objectives

- Identify key risk factors influencing delivery time
- Develop ML-based models to predict delivery time
- Derive Actionable Insights

Methodology

- Used Zomato transactional data (45.6K orders)
- Implemented data preprocessing on dataset
- Developed five ML models **(Linear Regression, Random Forest, Decision Tree, XGBoost, LightGBM)**
- Experimented on different types of data preprocessing (Scaled dataset vs unscaled dataset)
- Tuned chosen model using optuna
- Conducted SHAP analysis for model interpretability

Key Findings

- Overall, **delivery person riding a motorcyle took longer time to deliver food regardless road traffic conditions**
- Adult dominated delivery person demographics, but had the longest delivery duration
- The peak ordering hours and traffic jam occur between 19.00 and 22.00.
- **LightGBM is the best model with the root mean square error (RMSE) of 3.83**
- **Road traffic density, delivery person ratings, delivery person age, distance, and vehicle conditions** are top 3 key drivers of late deliver risk

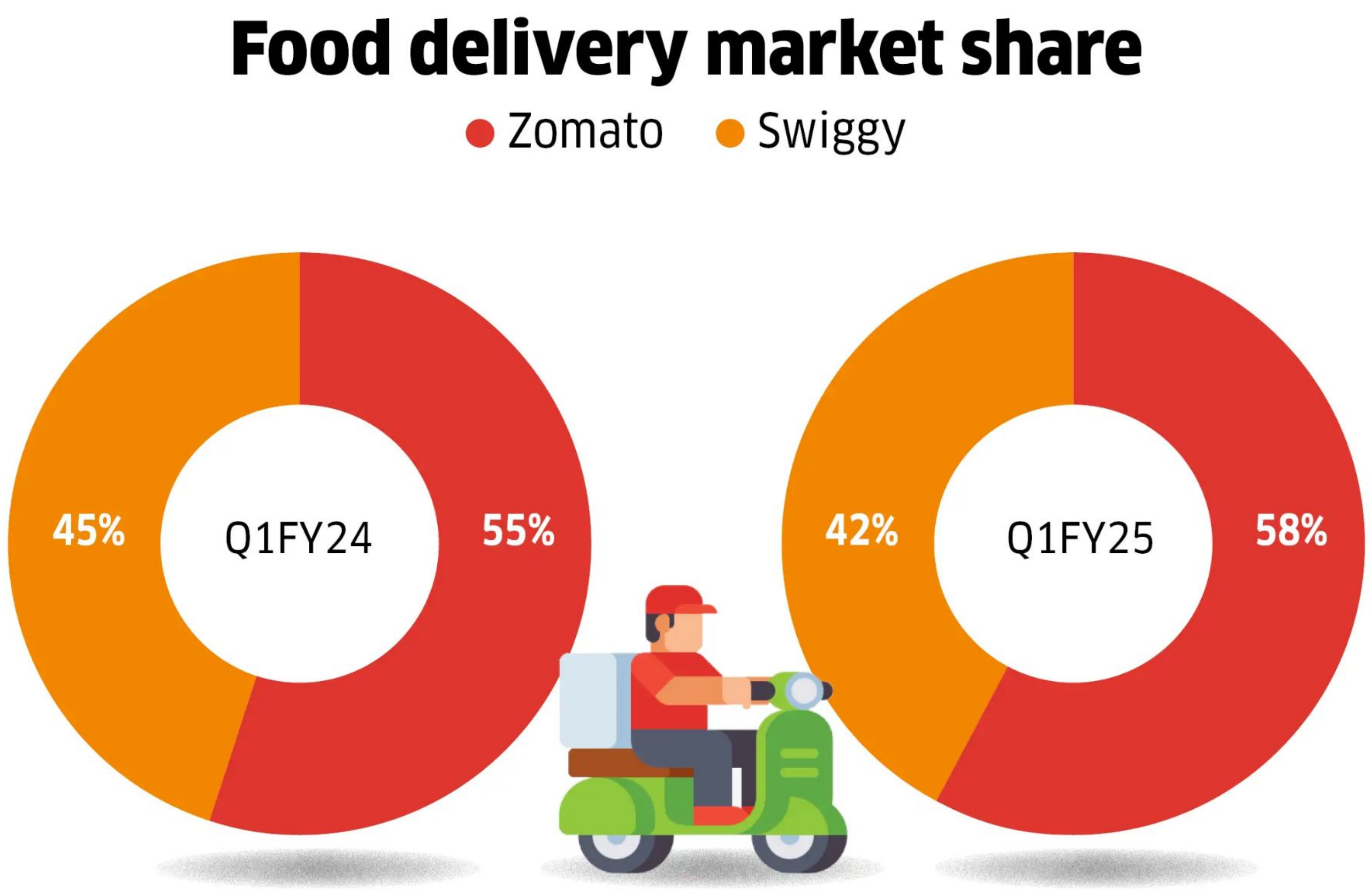
Business Recommendations

- **Route Optimization :** optimize routes during high traffic density/traffic jam, reduce the number of deliveries at the same time
- **Leverage High-Rated Delivery Person :** Prioritize high-rated delivery person for time-sensitive orders
- **Manage Delivery Person by Age**
- **Give Real-Time Expected ETA :** Update the expected ETA to customers based on real-time road and delivery person conditions
- **Maintain Vehicle Conditions :** Do regular checking and maintenance on all the vehicles, partner with vehicle service center
- **Distance-based Personnel Assignment**

BUSINESS UNDERSTANDING

Company Overview

Zomato is **India's #1 food delivery app** with 3 million+ restaurants and 3 billion+ orders delivered



Source : India Times

Problem Statement

- While Swiggy successfully launched 10-minute food delivery service [1], Zomato failed to sustain their 15-minute service and shut it down after only 4 months [2]
- In January 2022, customers experienced **average delivery times above 25 minutes**, which often leads to customer dissatisfaction and increases the risk of **customer churn**
- Without accurate delivery time prediction, Zomato may **lose competitive advantage** to Swiggy and other competitors

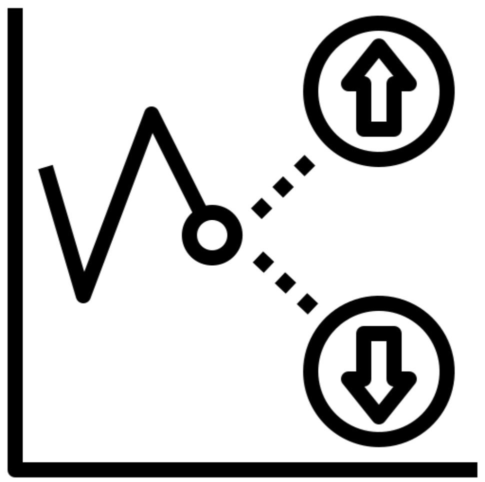
BUSINESS UNDERSTANDING

Key Challenge

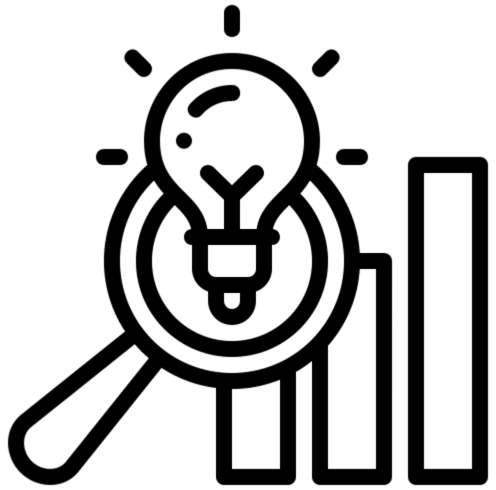
How can we leverage data-driven analysis and prediction model to formulate actionable recommendations for dealing with delivery time?

Project Objectives

Identify Key Risk Factors
Determine variables such as road traffic, weather conditions, distance, etc that influence delivery time



Develop Prediction Model
Develop ML-based models to predict delivery time



Derive Actionable Insights
Gain insights into factors influencing delivery time and formulate recommendations for effective strategies

DATA UNDERSTANDING

- Dataset can be downloaded on Kaggle
- Dataset provides a comprehensive view of delivery operations, including delivery person details, order timestamps, weather conditions, and road traffic density, and more
- Dataset contains **45,584 rows and 20 columns**
- Dataset was collected from 1st of January 2022 to 31st of January 2022
- **Target Feature : Time_taken (min)**

DATA PREPROCESSING

Convert Data Types
Convert column timestamp
Object -> Date

Check and Handle Missing Values
Some missing values found

Check and Handle Duplicates
No duplicates found

Check and Handle Outliers
Rows with outliers in restaurant coordinates were removed

Feature Engineering
Add some columns : geospatial and temporal for further analysis

MISSING VALUES

<5%

Missing values in two (out of 7 columns) columns are **imputed before train-test split**

0

No missing values in remaining columns

DUPLICATES

0

No duplicates found across all the columns

FINAL DATASET

25

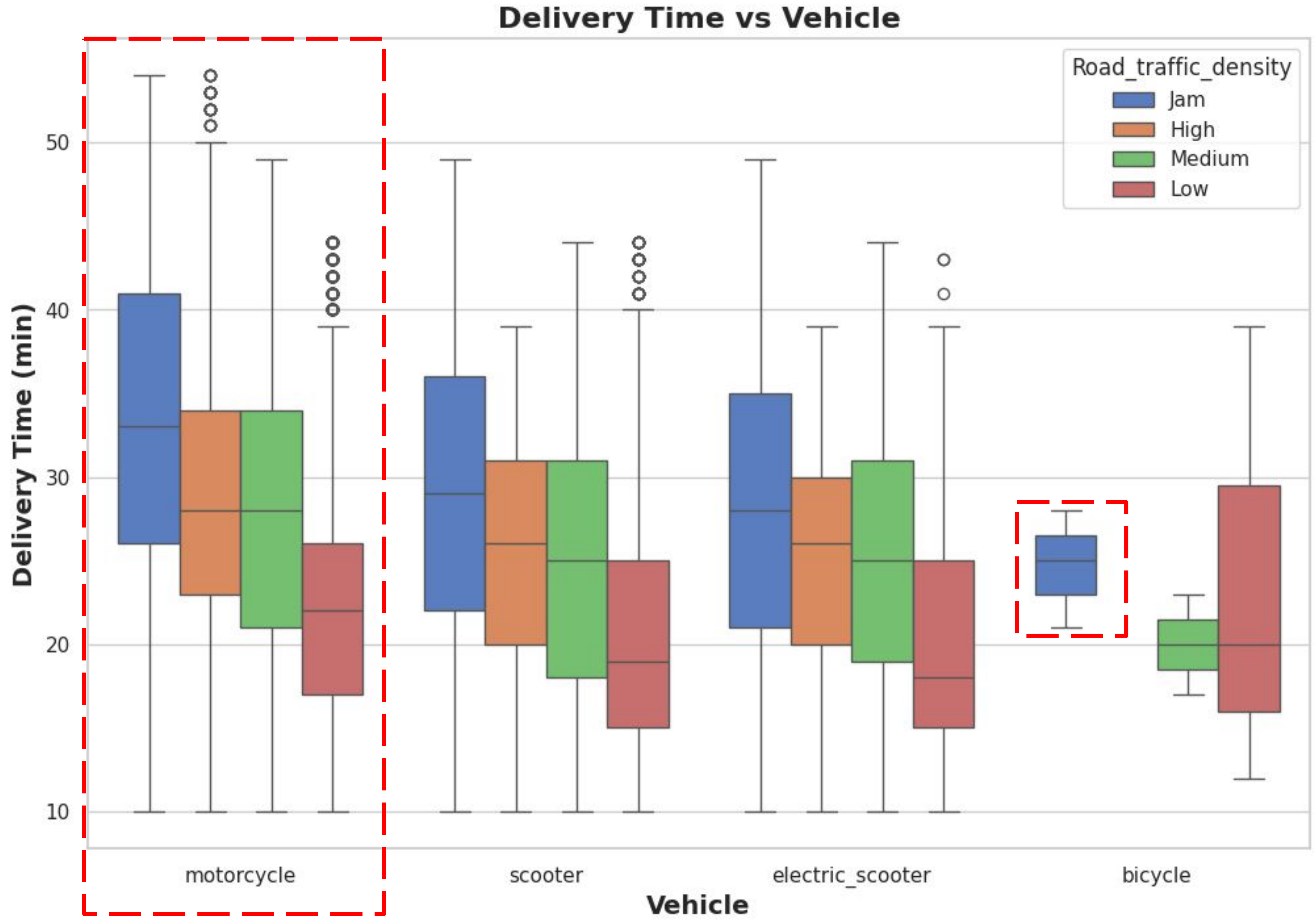
FEATURES

41,470

ROWS

EXPLORATORY DATA ANALYSIS

Delivery Time by Vehicle and Road Traffic



Insights :

1. Overall, **delivery person riding a motorcyle took longer time to deliver food regardless road traffic conditions**

2. **During traffic jam, bicycle is the best option as vehicle for delivery.** On average, bicycle deliveries took 24.7 minutes in traffic jams compared to 28–32 minutes for other vehicles. This is likely due to the flexibility of bicycle to navigate tight spaces and its ability to access certain streets that may be restricted for motorcycles (e.g. bike paths)

3. In contrast, **scooter and electric scooter emerged as the most efficient vehicles during low traffic.**

EXPLORATORY DATA ANALYSIS

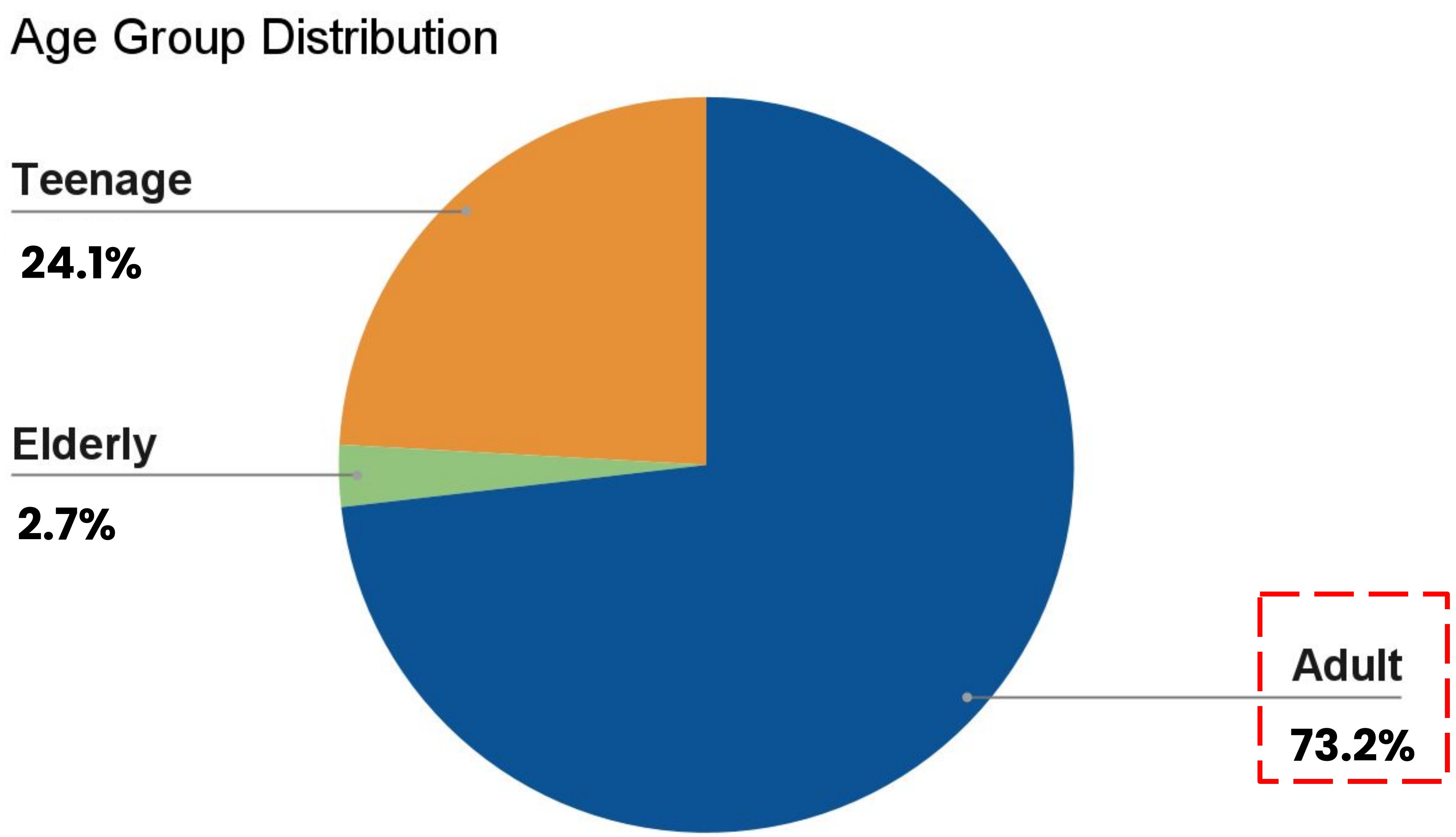
Delivery Time vs Delivery Ratings

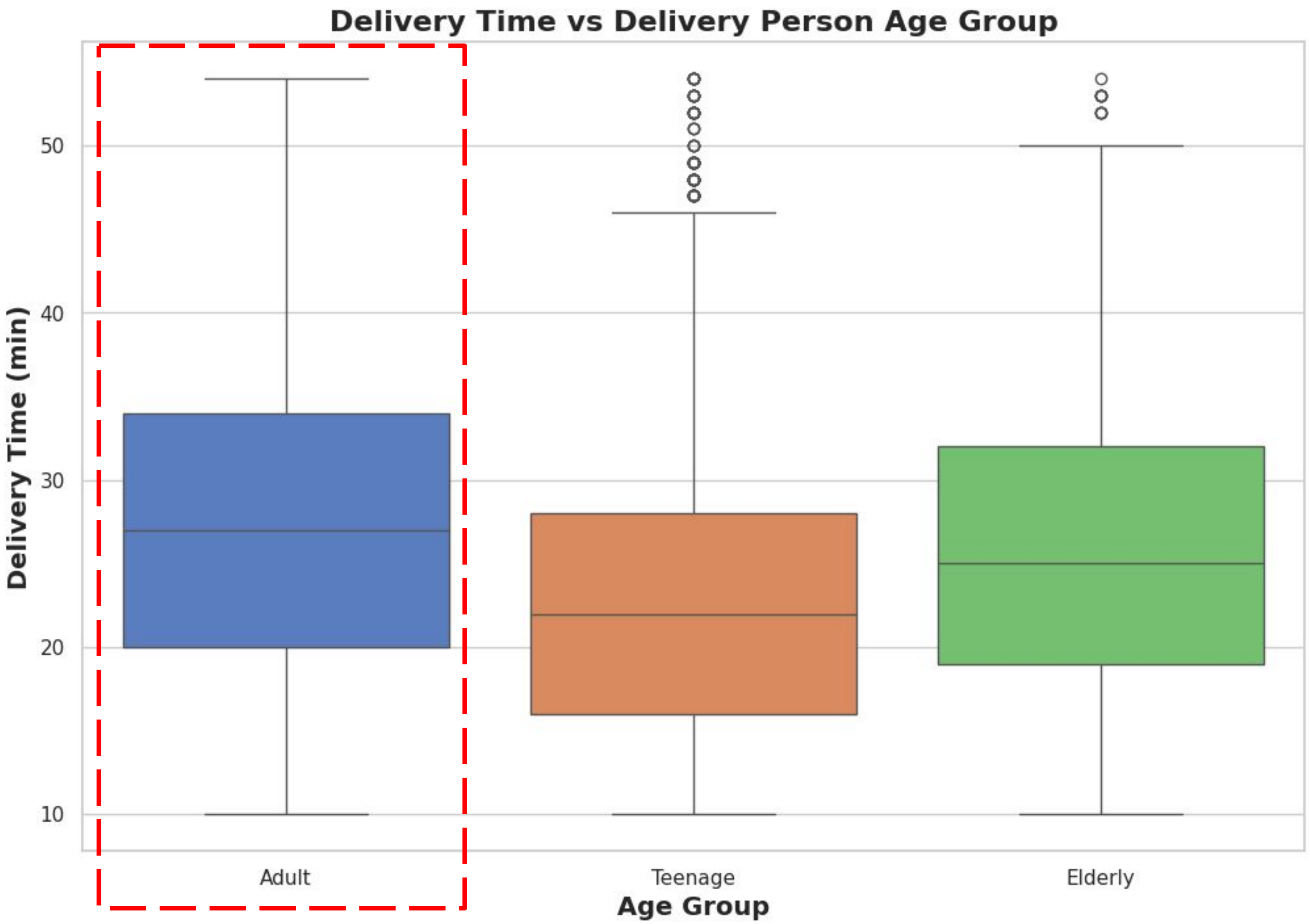
Insights :

1. **If the delivery person delivered food more than 30 minutes during low traffic, customers would not be happy and gave ratings of around 2.5 - 3.5.** This indicates how unsatisfied customers with the long delivery time if there is no any traffic
2. **Meanwhile, even the delivery took almost an hour during traffic jam, customers tend to give ratings more than 4.** This indicates that customers could understand the condition and be more patient with their orders being late
3. Interestingly, some orders received the lowest rating (1) despite being delivered in under 20 minutes, indicating that factors beyond delivery time influence customer satisfaction

EXPLORATORY DATA ANALYSIS

Delivery Time vs Delivery Person Age



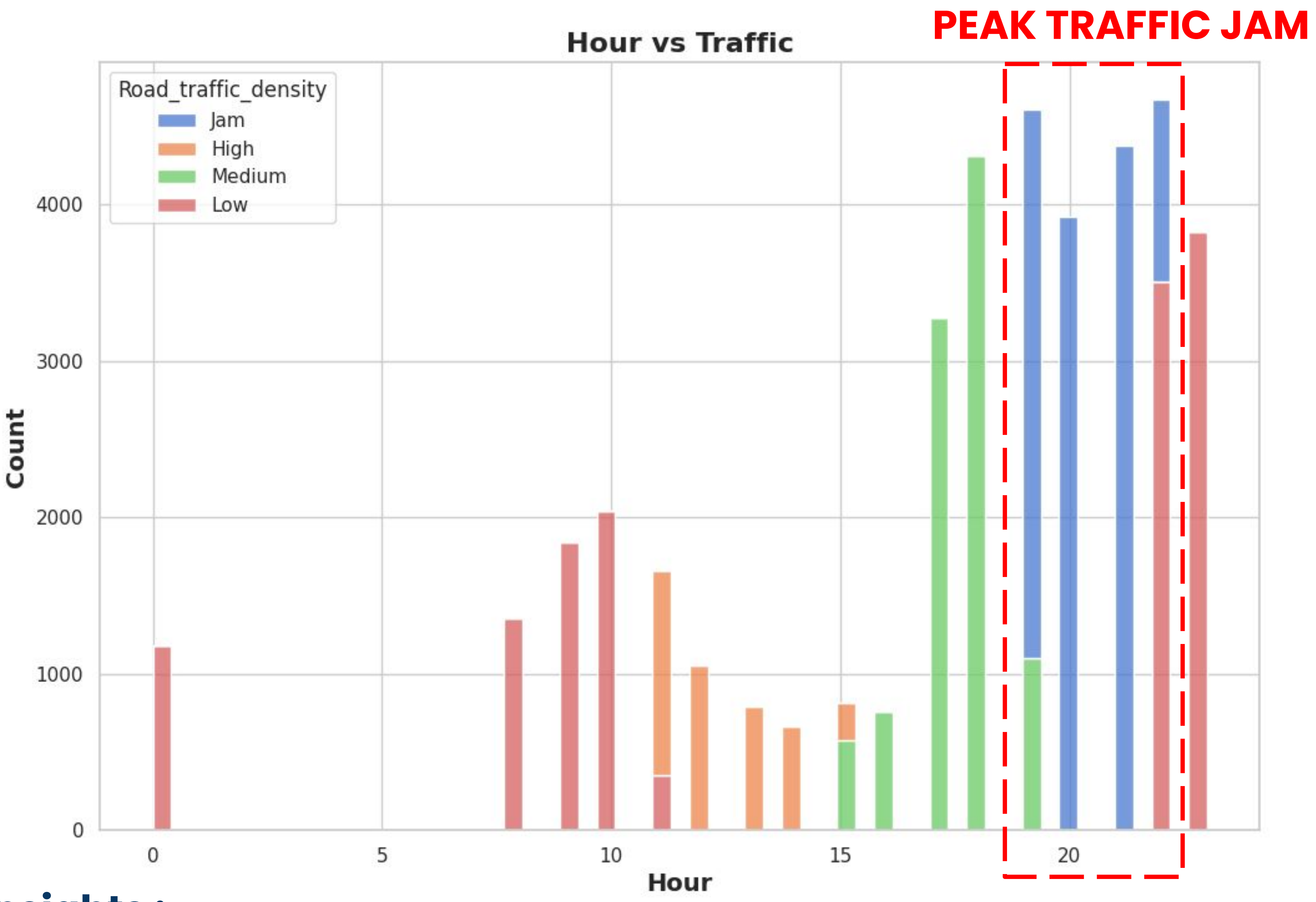


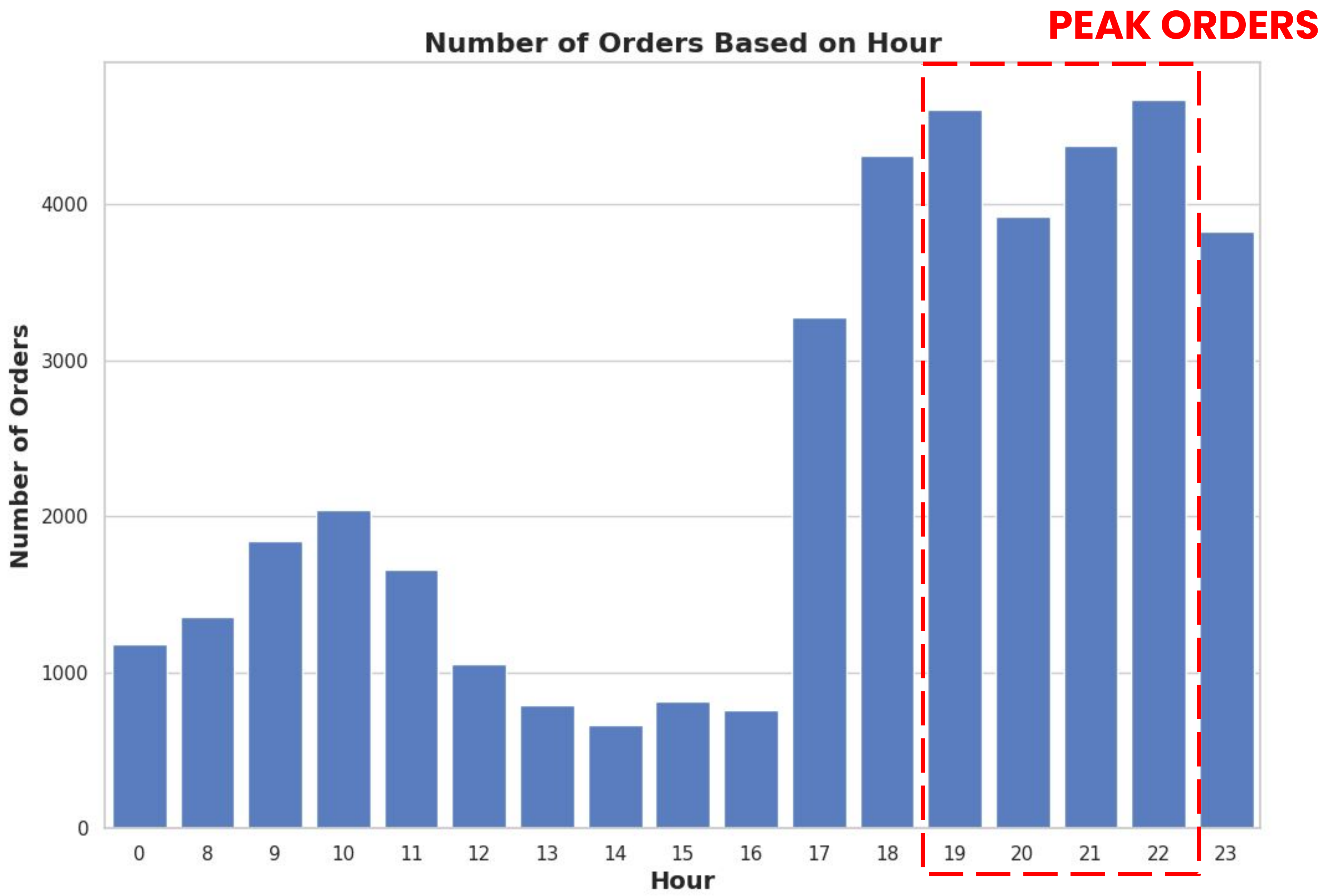
Insights :

1. **Adult dominated delivery person demographics** (30.1K out of 41K), **but this age group had the longest delivery duration** (median and mean of 27 minutes). This can be worrying since majority of delivery person aged from 25 years old to 64 years old
2. Teenage spent average of 23 minutes for food delivery

EXPLORATORY DATA ANALYSIS

Hours vs Traffic & Total Orders





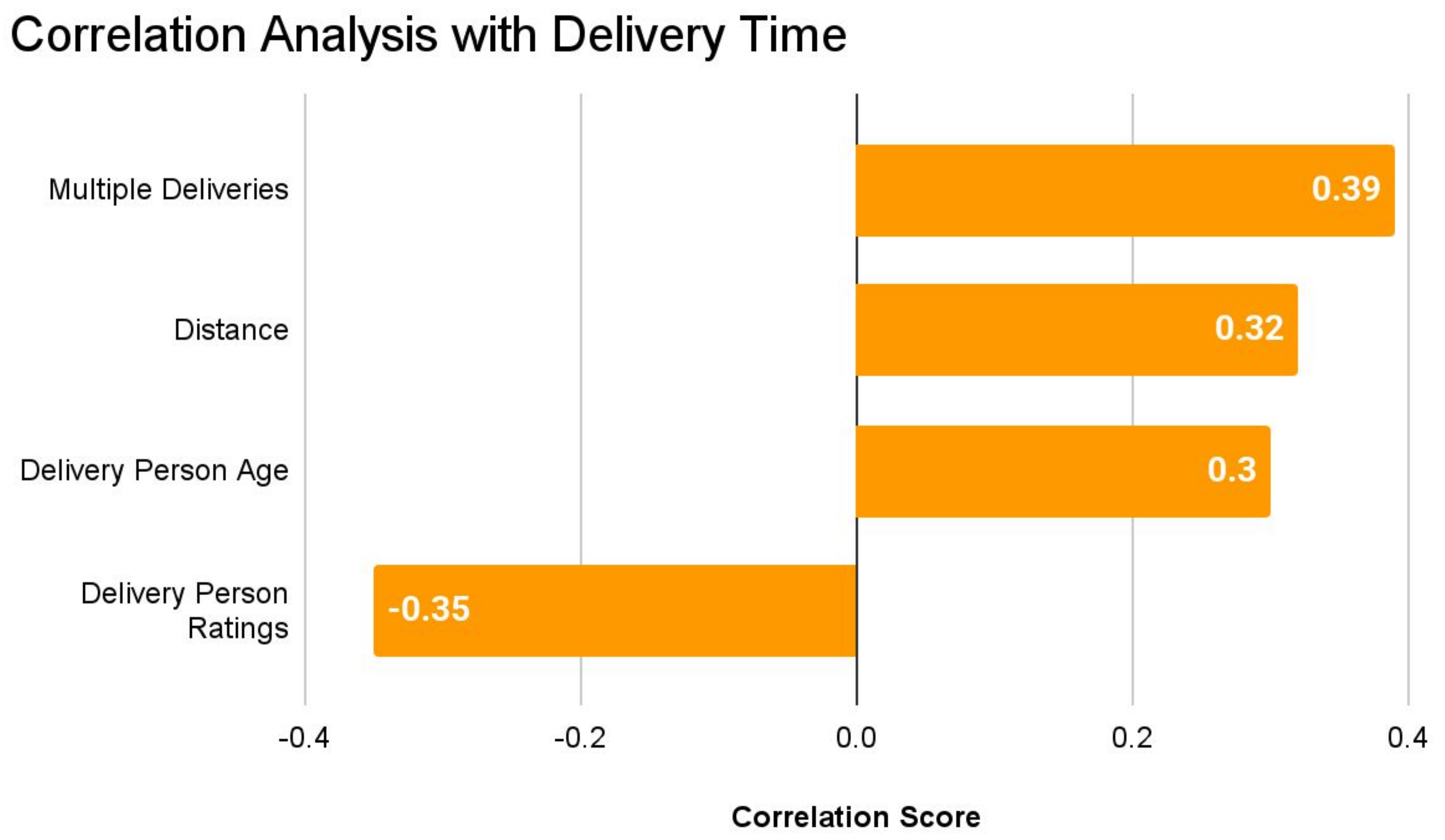
Insights :

1. **The peak ordering hours occur at 19.00 and 22.00**. Since the high probability of traffic jam also occurred at 19.00, it could lead to delivery issues
2. Number of orders increases in the evening. This makes sense because most of people have finished their day and spend their time with family or friends. Ordering food for dinner becomes a natural part of this routine

EXPLORATORY DATA ANALYSIS

Correlation Analysis

Out of 13 numerical features, **only 4 features have moderate correlation** (>=0.3) with 'Time_taken (min)'



POSITIVE CORRELATION

Number of Deliveries

Distance

Delivery Person Age

LONGER DELIVERY TIME

NEGATIVE CORRELATION

Delivery Person Ratings

Delivery Time

DATA PREPARATION FOR MODELING

Train-Test Split

80% Training Data
20% Testing Data

Missing Values Handling

5 out of 7 columns with missing values were imputed

Feature Encoding

- Categorical data with ordered values : **Ordinal Encoding**
- Categorical data with unordered values : **One Hot Encoding**

Feature Scaling

MinMax Scaler

Feature Selection

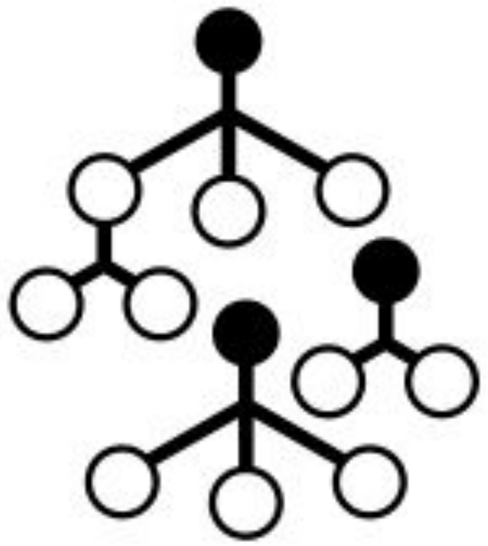
Drop columns that are possibly leakage to target feature (speed) and redundant features (Restaurant's longitude, Restaurant's latitude, Delivery location longitude, delivery location latitude

Training Data : 32,923 rows and 34 predictors
Testing Data : 8,231 rows and 34 predictors
Target Feature : Time_taken (min)

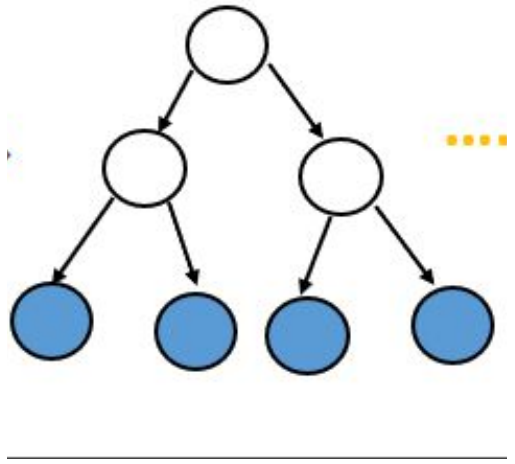
MODEL BUILDING

Five ML models were developed to compare their performances

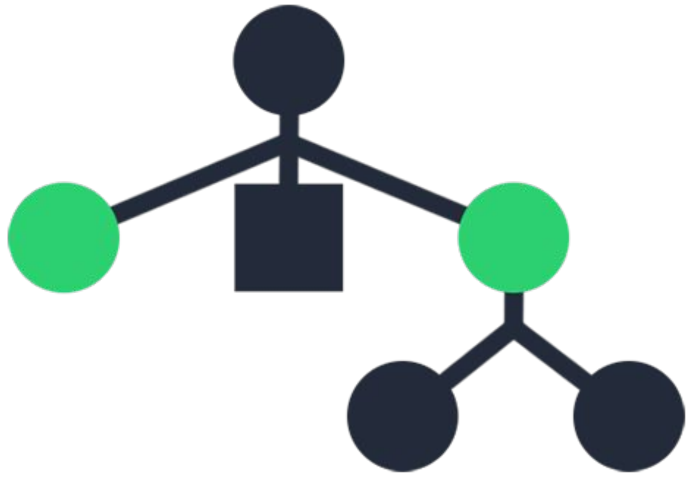
Linear Regression



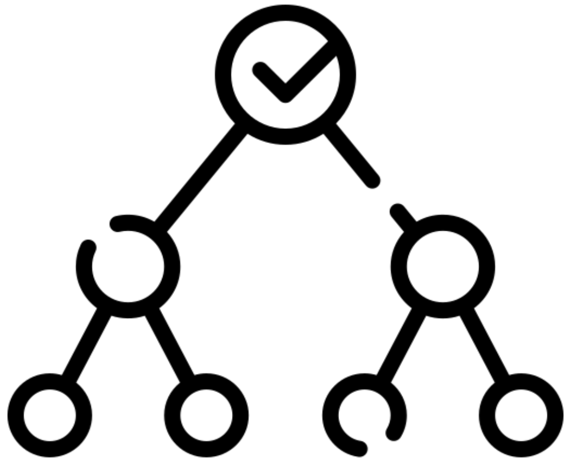
Random Forest



LightGBM



XGBoost

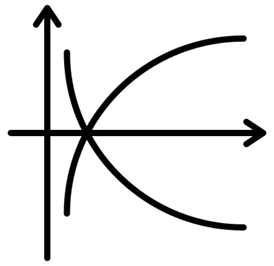


Decision Tree

EXPERIMENT 1

Baseline Dataset

EXPERIMENT 2

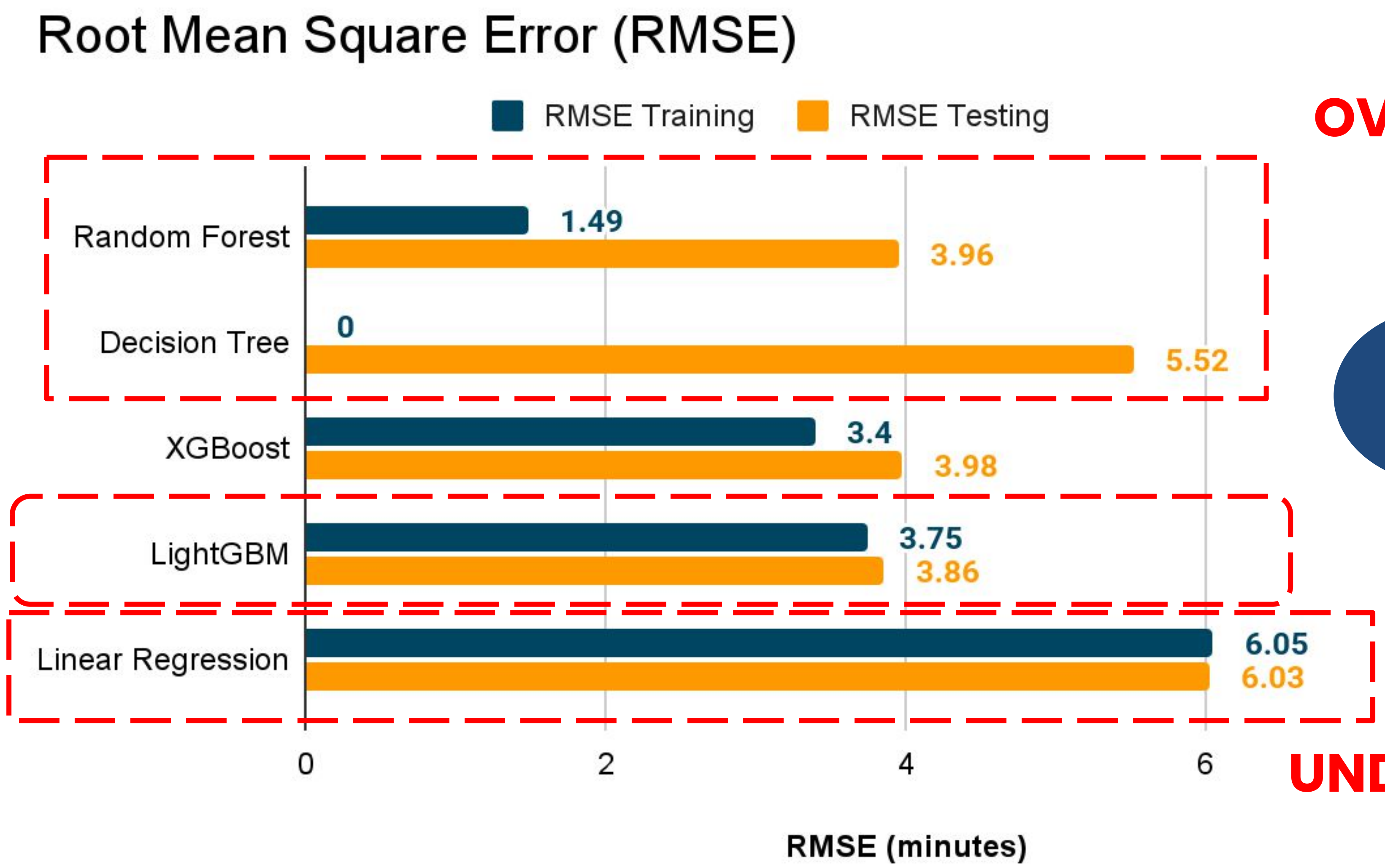


Scaled Data

MinMax Scaler

MODEL EVALUATION

Unscaled Dataset

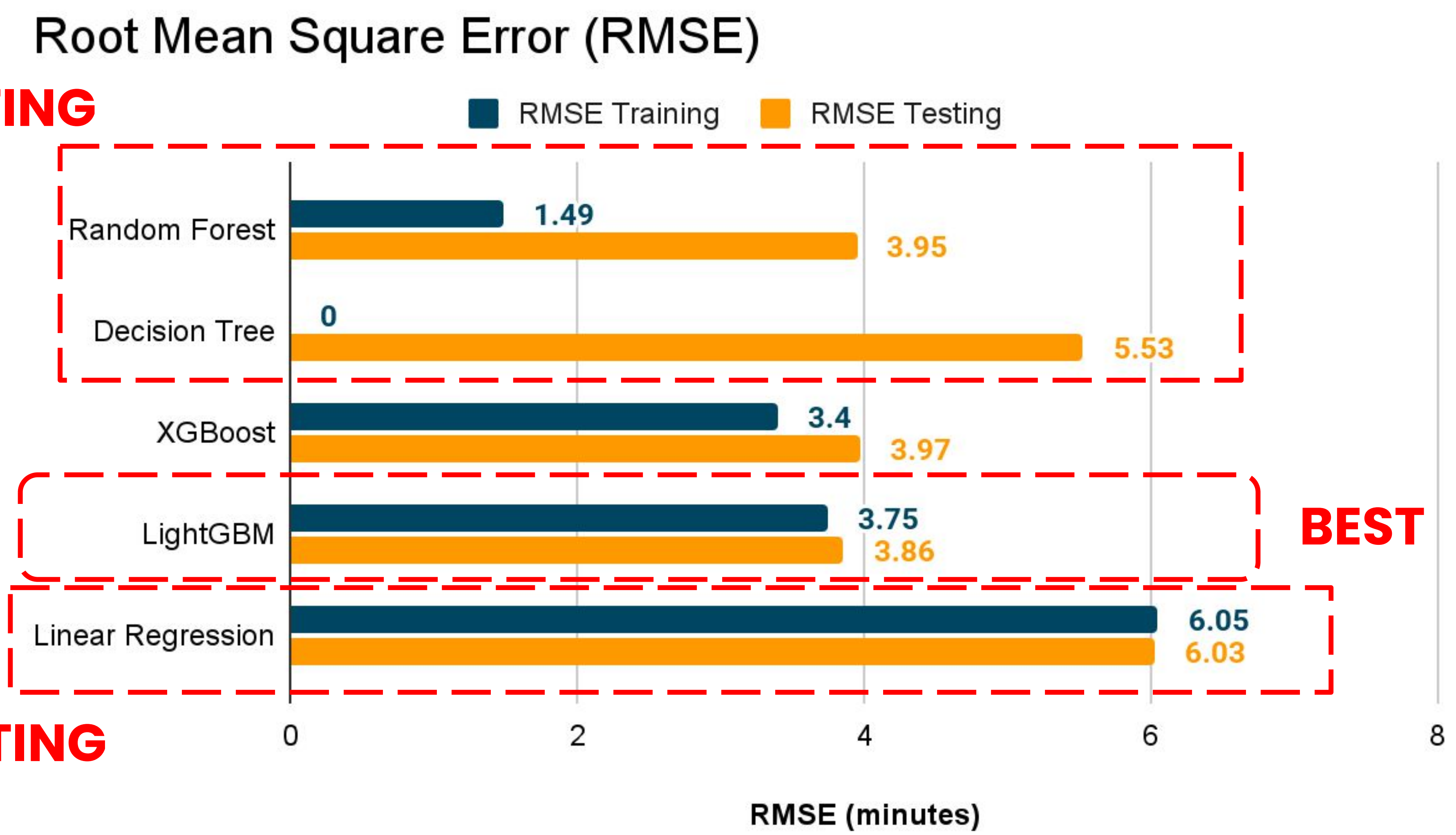


OVERFITTING

VS

UNDERFITTING

Scaled Dataset



BEST

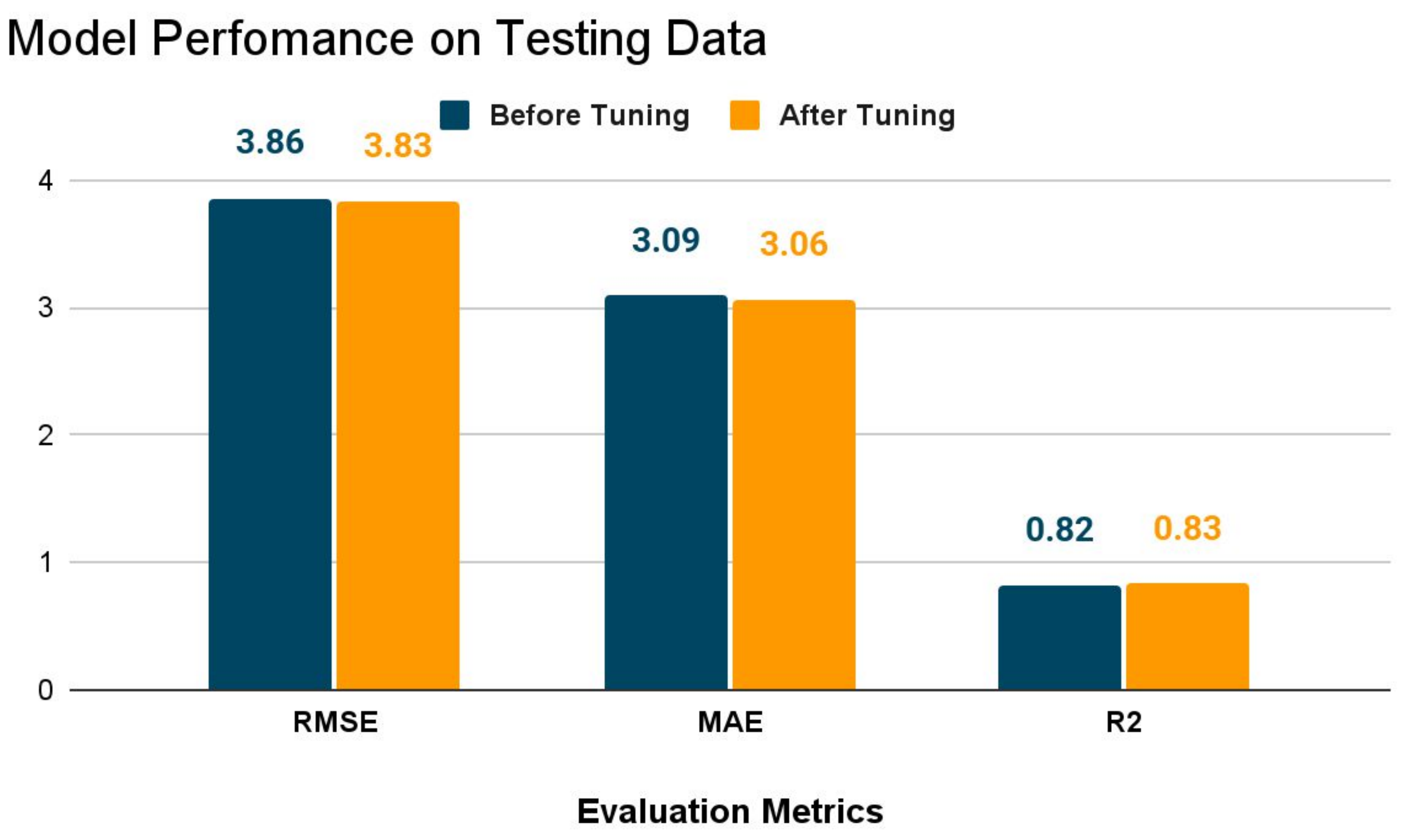
Insights :

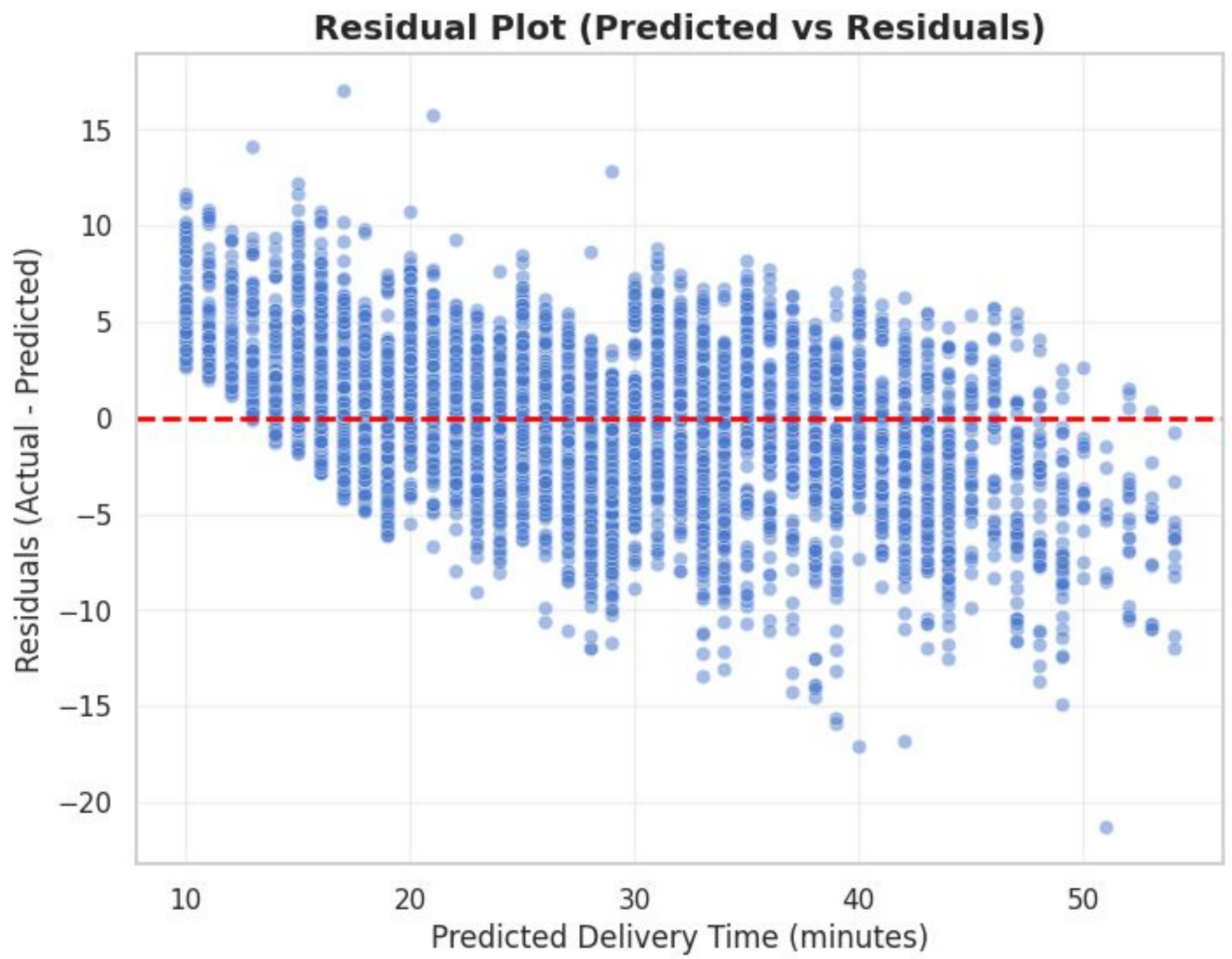
1. Random Forest and Decision Tree are overfitting to the training data, shown by large gap between RMSE training and testing data
2. Although Linear Regression good for data generalization, this model is possibly underfitting
3. **There is a very slight difference in model performance between scaled and unscaled datasets**

MODEL EVALUATION

Hyperparameter Tuning

Optuna optimization was implemented to choose the best parameters for LightGBM model with unscaled data





avg of 1% increase across all the evaluation metrics

Best model : **LightGBM after tuning**

Insights:

- **Hyperparameter tuning has improved LightGBM performance** while making it not overfitting
- Residual plot shows slight bias : **underestimate short times and overestimate long times**

MODEL INTERPRETATION

Top 5 Key Drivers of Delivery Time

1. **Road Traffic Density :** as higher traffic density, longer delivery time
2. **Delivery Person Ratings :** as higher delivery person ratings, the faster delivery time
3. **Delivery Person Age :** as older delivery person, the slower delivery time
4. **Distance :** as far distance between restaurant and delivery location, the slower delivery time
5. **Vehicle Condition :** as better condition, the faster delivery time

BUSINESS RECOMMENDATIONS

Route Optimization

- Optimize routes during high traffic density/traffic jam
- Reduce the number of deliveries at the same time
- Inform real-time road conditions to delivery person

Leverage High-Rated Delivery Person

- Prioritize high-rated delivery person for time-sensitive orders
- Provide incentives to high-rated delivery person

Manage Delivery Person by Age

- Assign shorter distance or less traffic to old delivery person
- Provide training for adults to elderly people to speed up their delivery time

Give Real-Time ETA

- Update the expected ETA to customers based on real-time road and delivery person conditions

Maintain Vehicle Conditions

- Do regular checking and maintenance on all the vehicles
- Partner with vehicle service center

Distance-based Personnel Allocation

- Allocate delivery person who is close to restaurant to the orders

ROOM FOR IMPROVEMENT

Experiment with Different Techniques for Data Preprocessing

Try different ways for missing value imputation; different techniques for feature encoding, etc

Monitor Performance

Track error regularly and refine the model if the error increased

Deploy Model

Implement model in the company system either via web or app

STREAMLIT DEPLOYMENT

[LINK STREAMLIT]

Thank you!

[dindararas]

tableau [Dinda Raraswati]

[Dinda Raraswati]