DELIVERY TIME PREDICTION

USING
MACHINE LEARNING

Muhammad Fadhil Asyam

Table of Contents

Background

In the business world, especially in the e-commerce sector like Amazon, on-time delivery is critical to customer satisfaction. With increasing transaction volumes and customer expectations, companies need to understand and predict delivery times accurately.

Delivery time prediction analysis aims to identify factors that affect delivery duration and provide a better estimate of the time it takes to deliver a product to a customer.

What factors influence delivery time?

Data Preparation

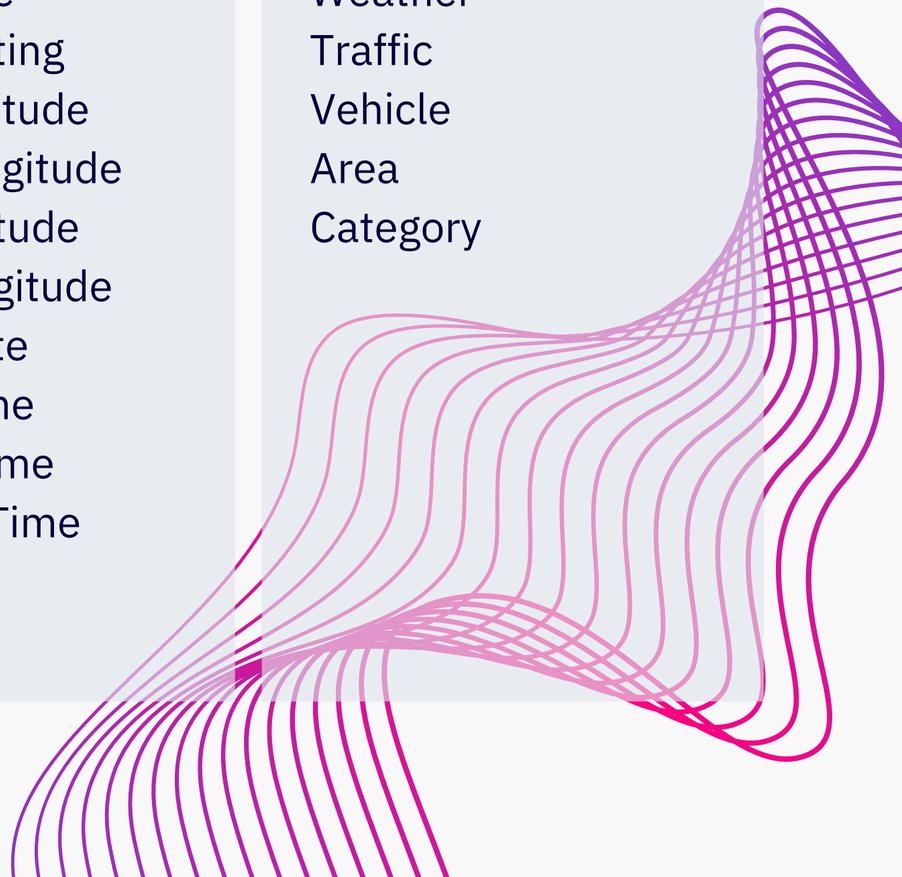
Dataset : amazon_delivery.csv

Data contains 43739 rows and 16 columns and with Delivery Time as target.

All of these features are influencing elements that occur on that day at a particular location which are used to predict delivery time.

Features consist 5 categorical features, and 10 numerical features. (exclude Order_ID)

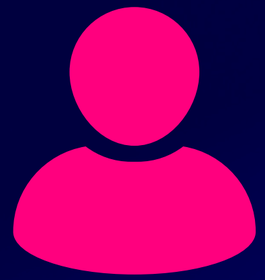
Numerical Features	Categorical Features
Agent_Age	Weather
Agent_Rating	Traffic
Store_Latitude	Vehicle
Store_Longitude	Area
Drop_Latitude	Category
Drop_Longitude	
Order_Date	
Order_Time	
Pickup_Time	
Delivery_Time	



Missing Values

Weather

91 Missing Values on Weather Column



54

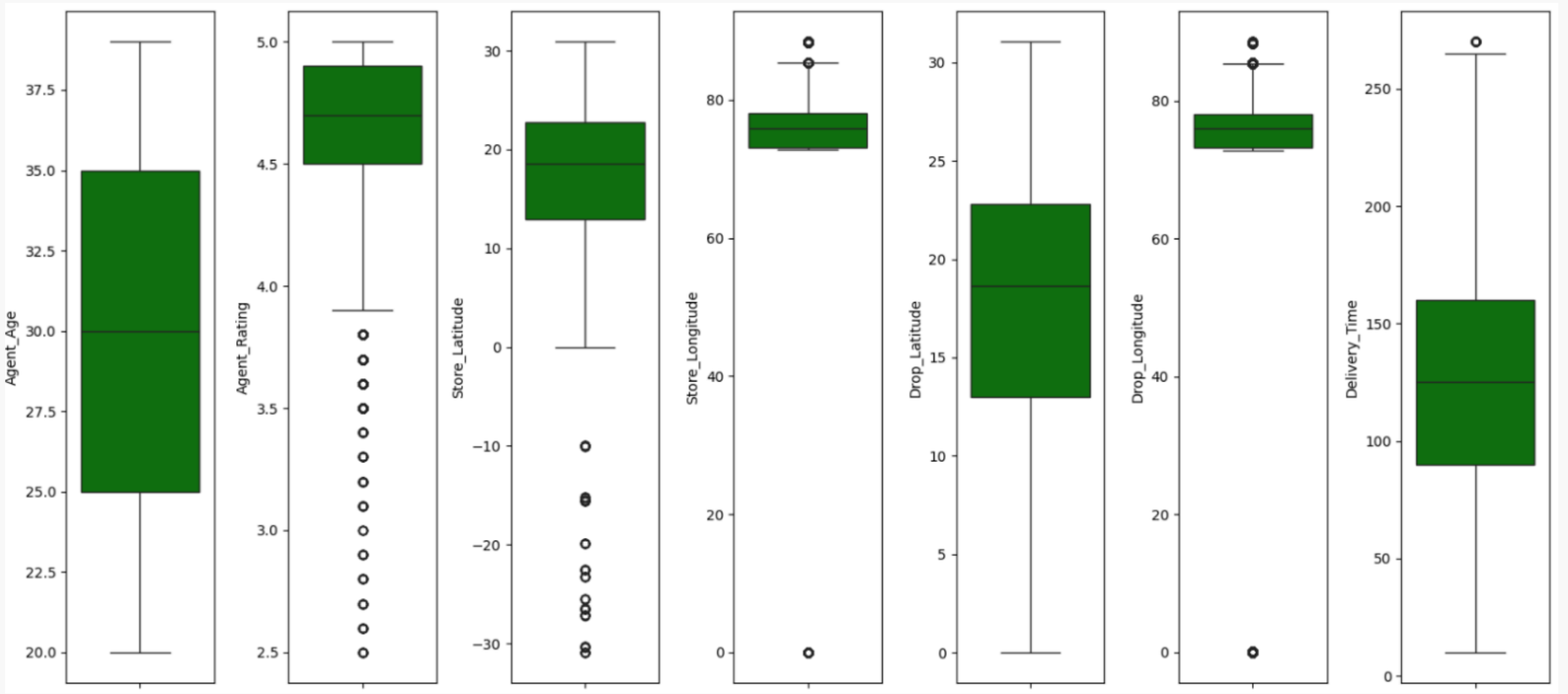
Agent Rating

54 Missing Values on Agent Rating Column

Dropping rows containing missing values from those two columns.

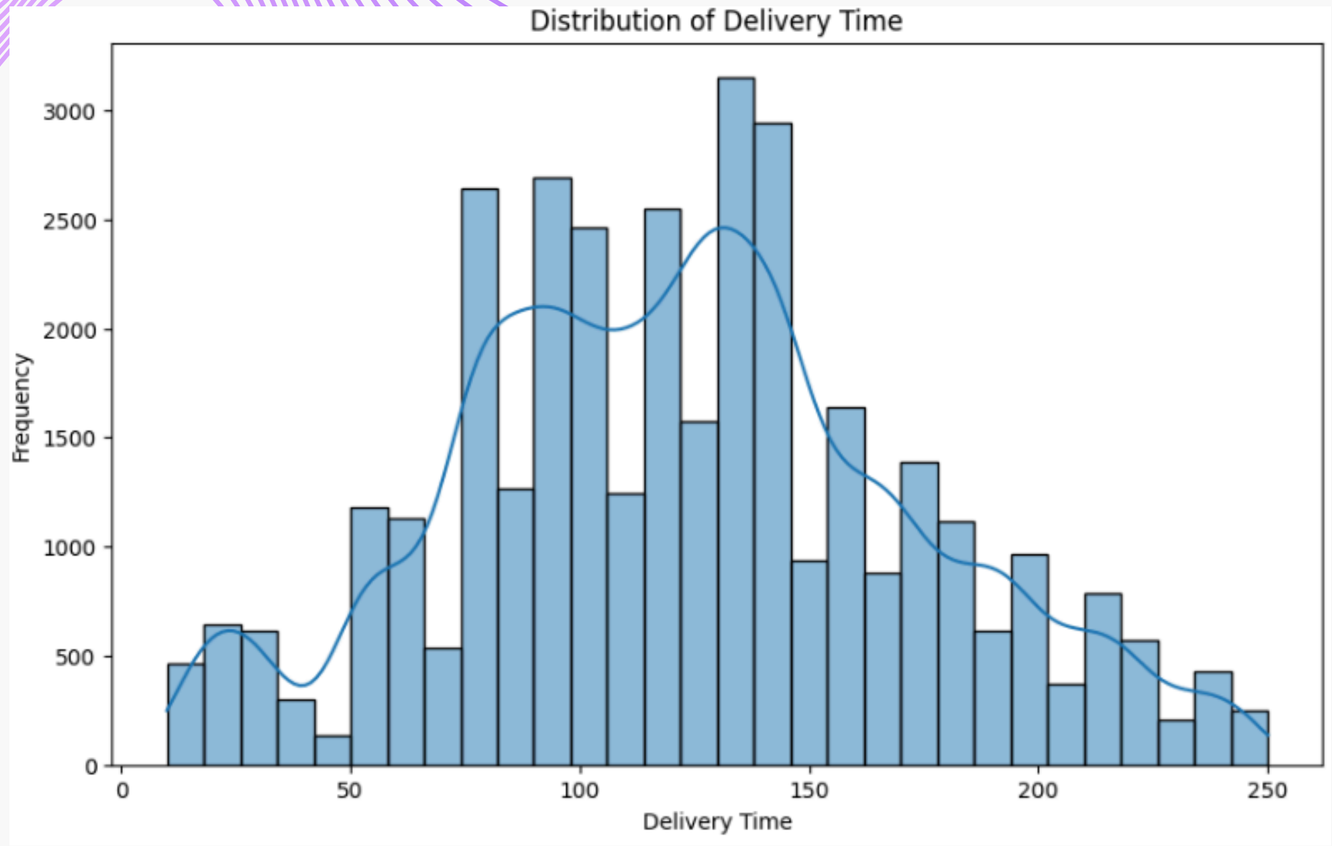
Exploratory Data Analysis

Handling Outliers



Six features have outliers and need to removed based IQR (Interquartile Range) upper and lower limit.

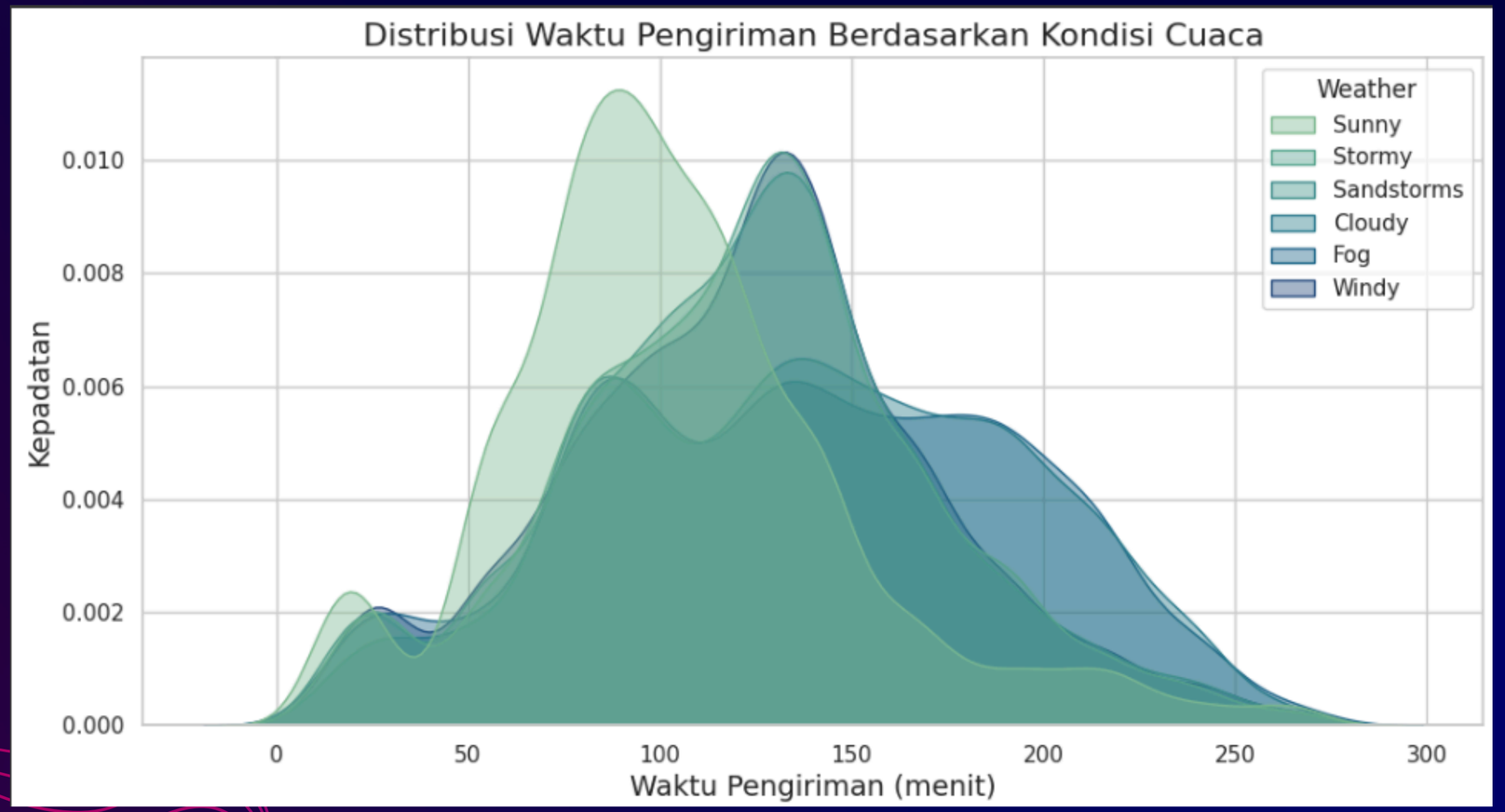
Distribution of Delivery Time



	Delivery_Time
count	35654.000000
mean	122.806193
std	50.320654
min	10.000000
25%	90.000000
50%	120.000000
75%	155.000000
max	250.000000

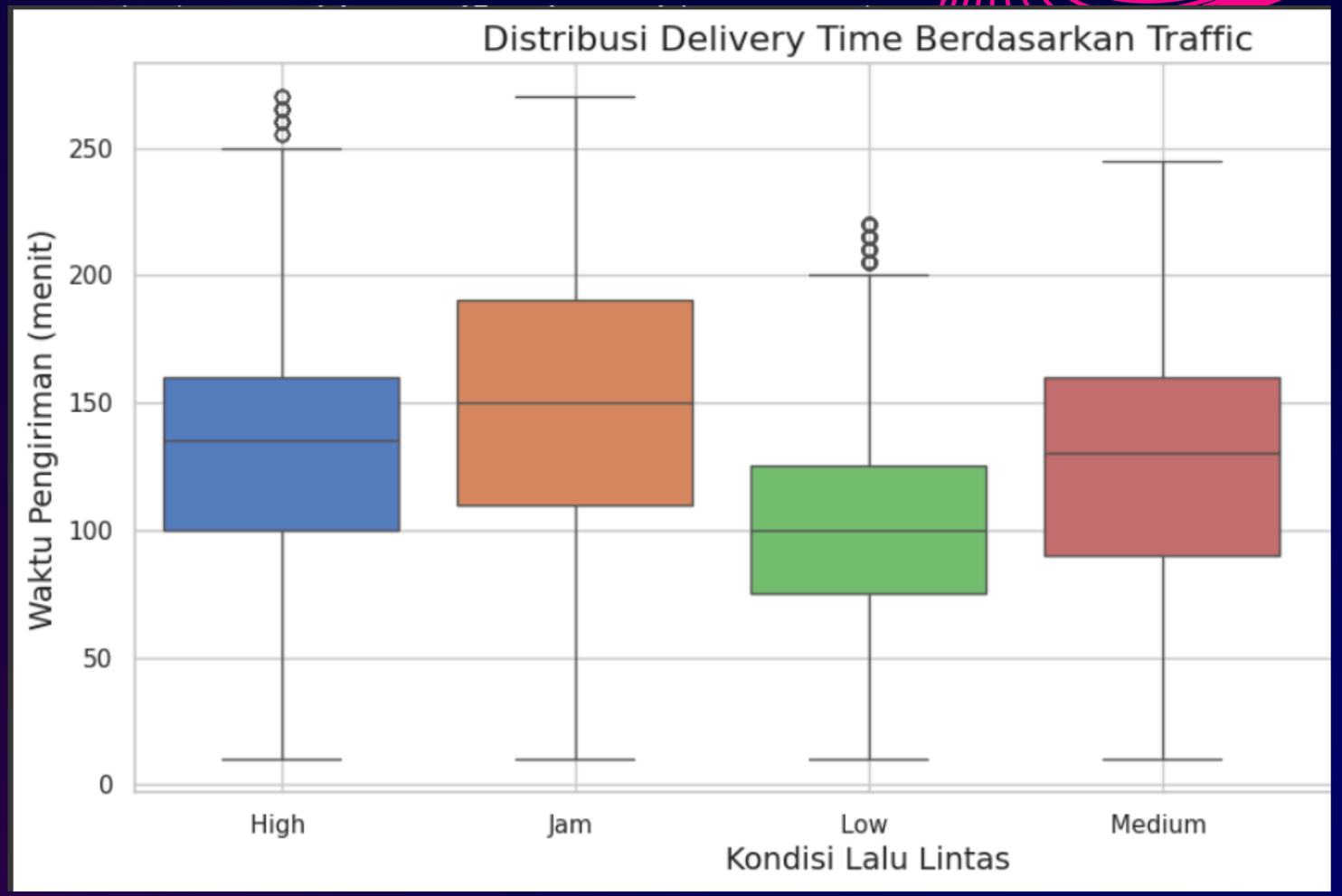
Mean Delivery Time: The average delivery time is approximately 122.8 minutes

Distribution of Weather



Sunny conditions (light green) have the highest density around 75 to 100 minutes, indicating faster deliveries. Stormy, sandstorms, and windy conditions show more spread-out distributions with longer delivery times.

Overall, there appears to be no strong correlation between agent ratings and delivery times, indicating that other factors might play a more significant role in determining delivery time.

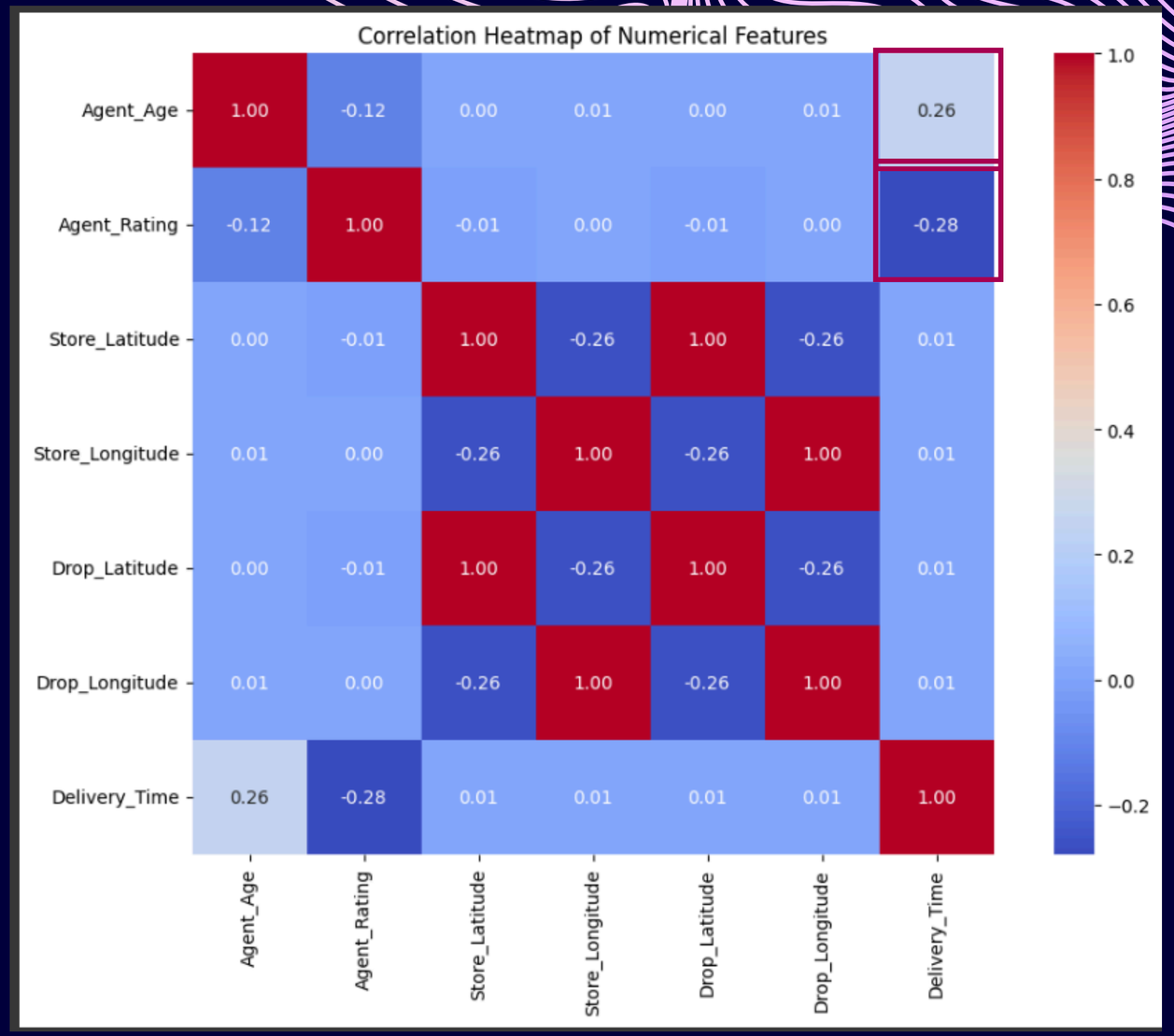


Delivery time with low traffic looks faster, while the Jam traffic condition has a longer delivery time distribution.

Corellation Heatmap of Numerical Features

Agent_Age vs Delivery_Time: There is a moderate positive correlation, indicating that older agents might have longer delivery times.

Agent_Rating vs Delivery_Time: A moderate negative correlation (-0.28) suggests that agents with higher ratings tend to have shorter delivery times.



Feature Engineering

Categorical Features

Weather, Vehicle, Area → Label Encoder

Traffic → Ordinal Encoder

Category → Label Encoder

New Features

Pickup_Time

[Pickup_Time] - [Order_Time]

Distance_km

Haversine formula from store lat&long and drop lat&long

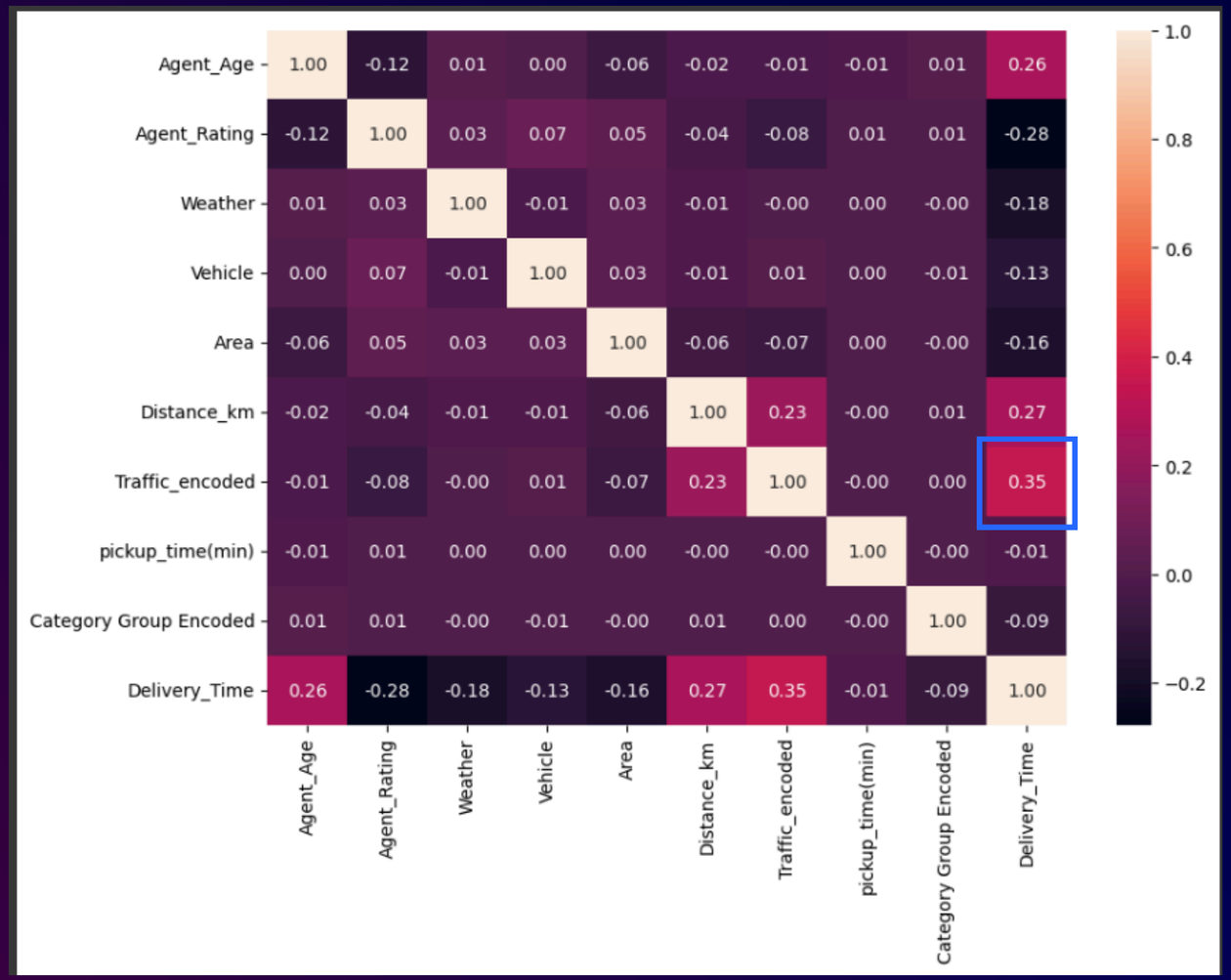
Others

Drop Feature

Order_date

Scaling

Min Max Scaler



Agent rating, agent age, traffic, and delivery distance appear to be the features that have a greater impact on delivery time than the others.

Traffic_encoded has the strongest effect (correlation 0.35) on delivery time, indicating that heavy traffic is a major factor in extending delivery time.

Modelling
And Evaluation

Modelling

Model	RMSE	MAE	MAPE (%)	Mean Error	Mean Percentage Error (%)
Ridge	40.51	31.02	44.64	0.95	0.77
Lasso	41.35	31.87	46.05	0.98	0.79
Linear	40.51	31.02	44.64	0.95	0.77
Random Forest	33.72	24.62	35.16	1.51	1.22
Gradient Boosting	32.89	24.29	35.97	0.85	0.69

Data splitted by 80:20 proportion for train data and test data

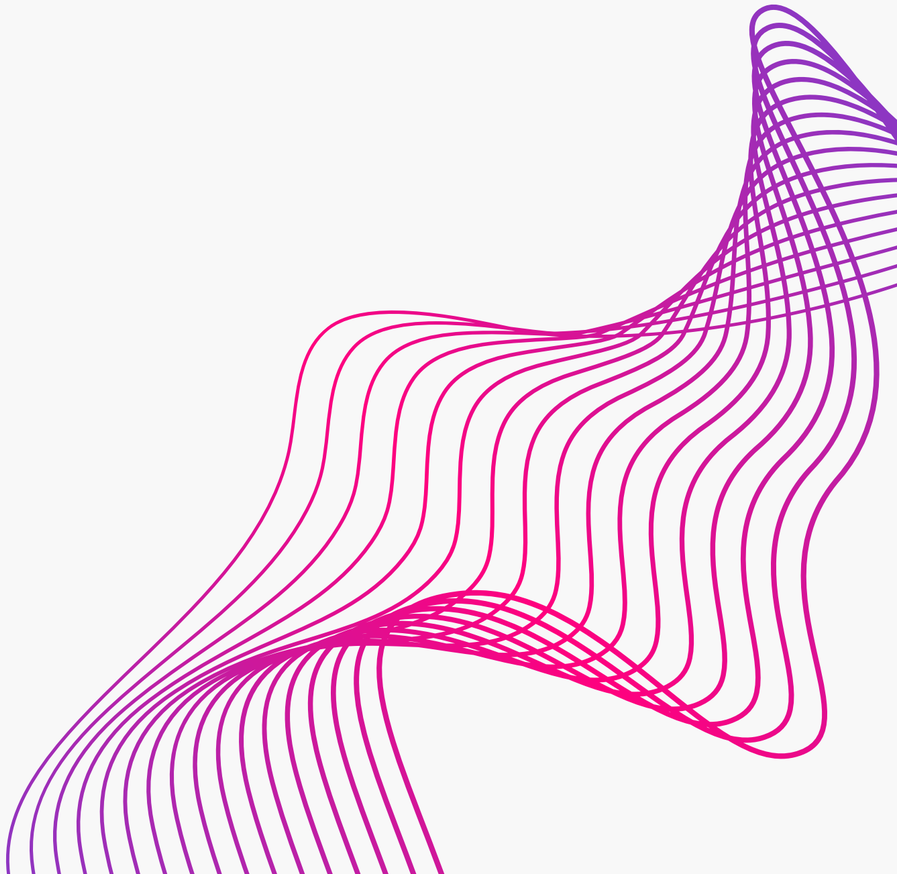
Gradient Boosting has the lowest MAE (Mean Absolute Error) & RMSE (Root Mean Square Error).

Hyperparameter Tuning

GradientBoosting	RMSE
Before Tuned	32.89
After Tuned	32.05

by RandomizedSearchCV

The lower RMSE (32.05) after tuning indicates that the model has improved in terms of prediction accuracy. Although this decrease is not very large, it shows that tuning has had a positive impact.



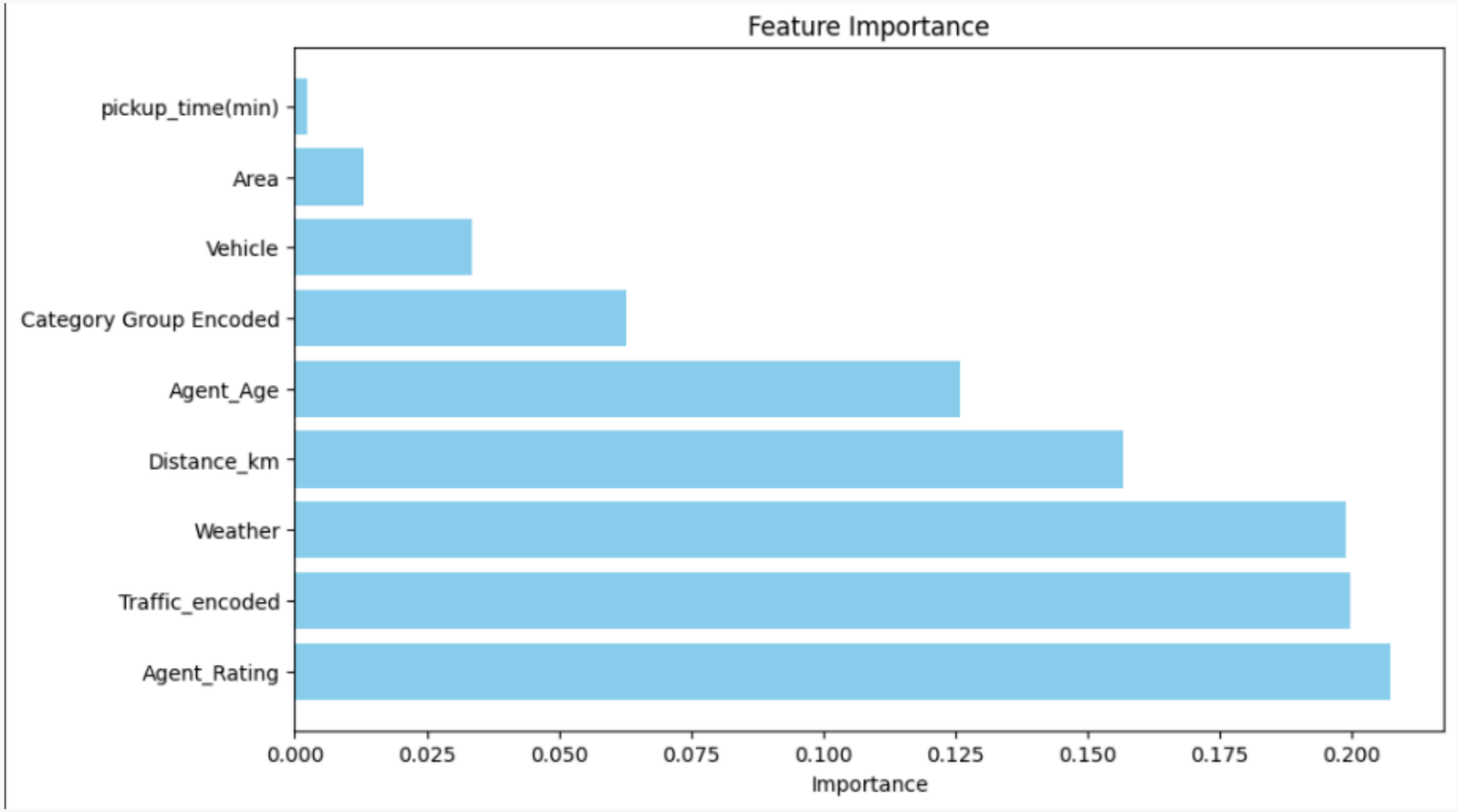
Evaluation

Dataset	MSE (Mean Squared Error)	RMSE (Root Mean Squared Error)
Train Data	1010.55	31.79
Test Data	1027.52	32.05

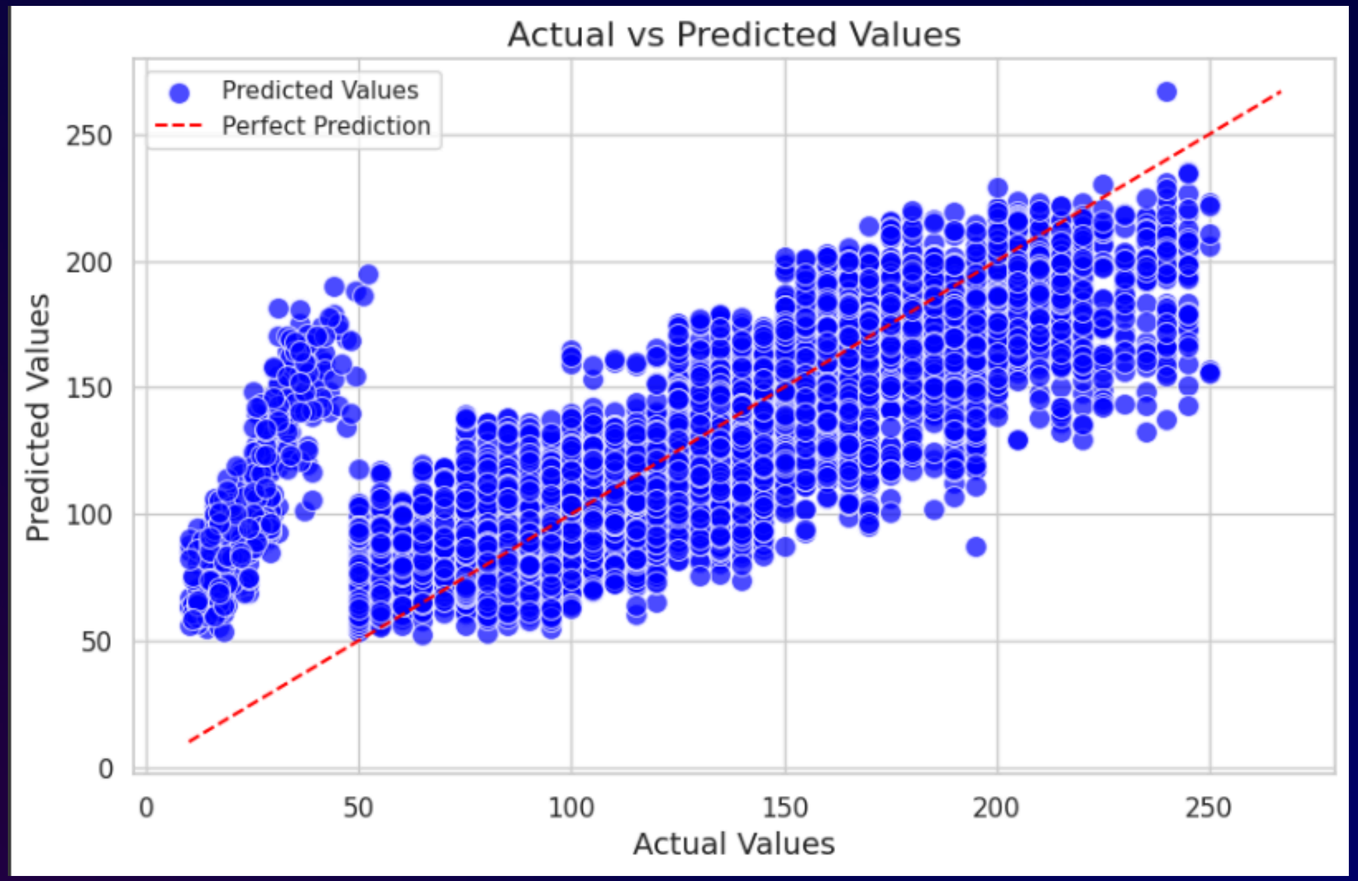
An RMSE of about 32 indicates that, on average, the model predictions deviate by about 32 minutes from the actual value.

The error metric for the test data is slightly higher compared to the training data, indicating that the model may be experiencing slight overfitting. However, this difference is not very significant, indicating that the model is quite stable

Feature Importance



Agent rating, Traffic, and Weather have the biggest influence among other features.



Overall, despite some outliers, the model seems to do a pretty good job of predicting the values of this data overall.

Conclusion

Gradient Boosting shows relatively good performance in predicting data values, with an RMSE of 32 minutes. This shows that the model can capture data patterns quite well overall.

Most Influential Features: Agent rating, Weather, and Traffic have been identified as important factors.

Things to Consider

Outliers: Although the model may capture the general pattern, the presence of several outliers (predicted values that are far from the actual values) indicates that there may be other variables or factors that are not detected by the model.

ContinuousE valuation: It is important to continuously monitor the performance of the model over time and make updates when necessary.

Business Recommendation

Companies can improve customer service by:

Inform customers about factors such as weather and traffic that can affect delivery times. This can increase customer satisfaction by providing more realistic expectations.
Improve agent rating systems and provide training or incentives to improve their ratings, which in turn can improve delivery efficiency.

Companies must anticipate deliveries during rush hours or holidays when traffic significantly increases, and also provide real-time information if there are weather issues that may delay delivery times.

Thank You !