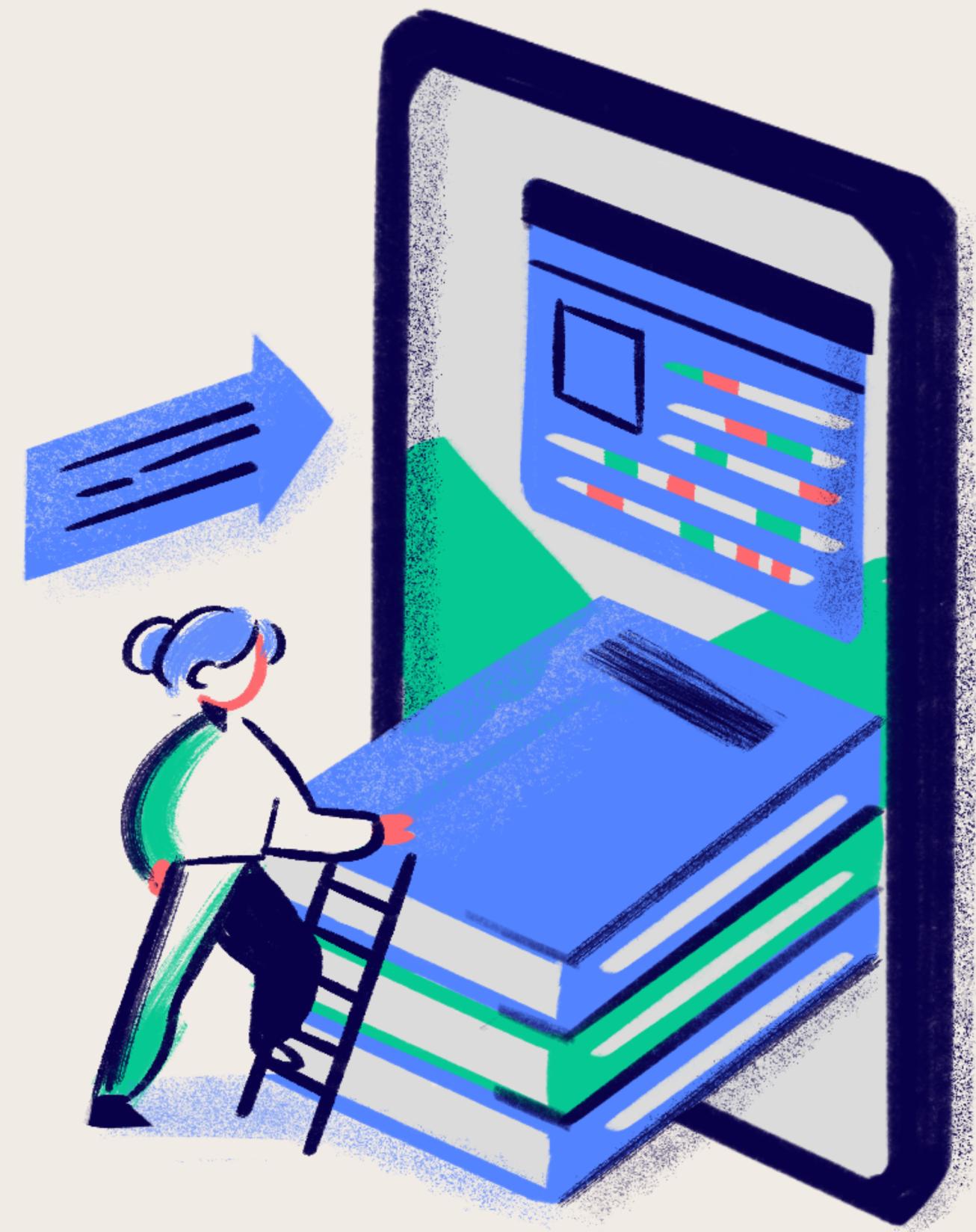


PRESENTED BY MALAK ESSAM

# FASTAG FRAUD DETECTION

MACHINE LEARNING FOR  
REAL-TIME SECURITY



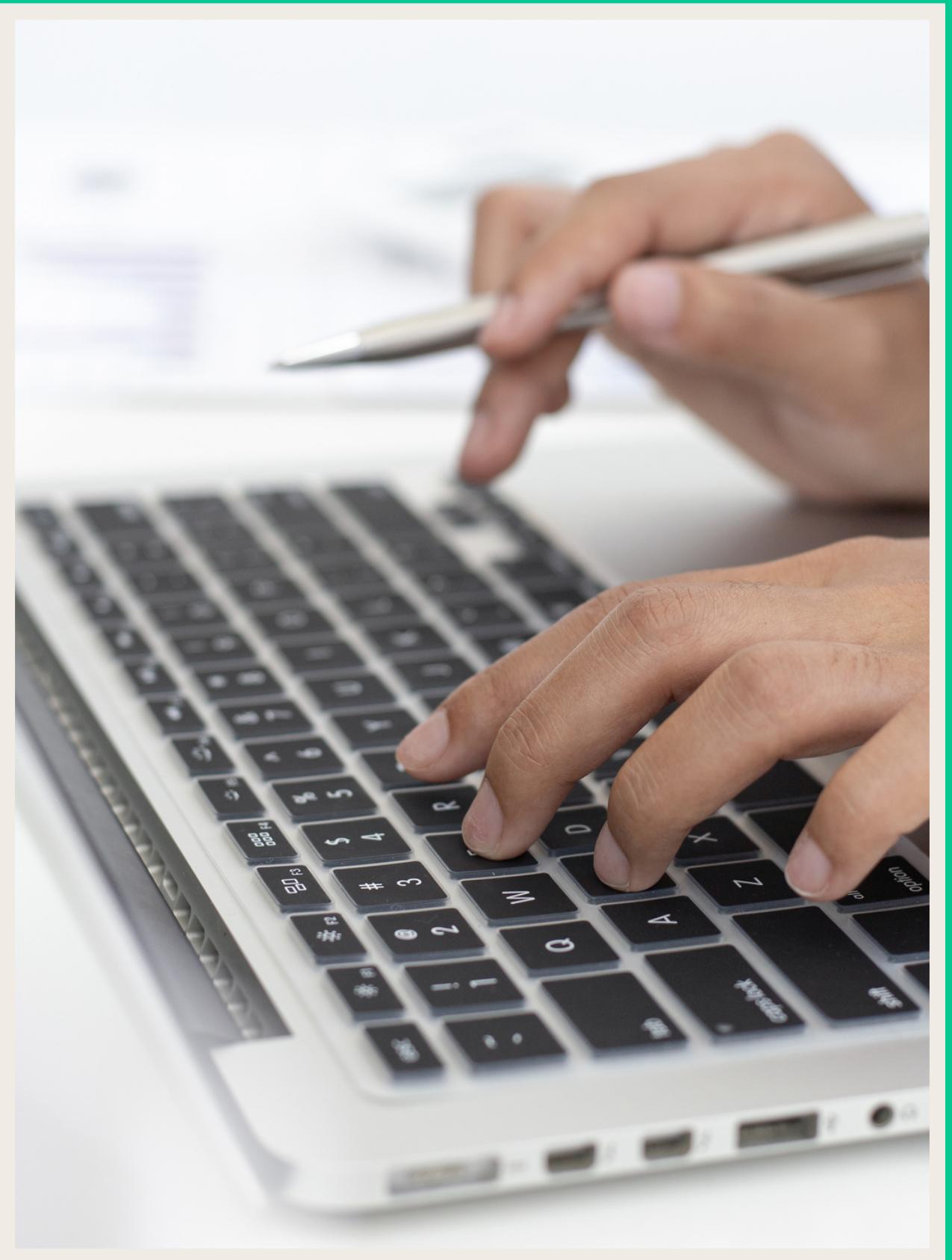
# INTRODUCTION

- Fastag is a vital electronic toll collection system used on highways, ensuring seamless travel and reduced congestion.
- fraudulent activities within the Fastag system pose significant challenges, leading to financial losses and compromised security.
- The project aims to develop a machine learning model for real-time detection of fraudulent Fastag transactions.
- The ultimate goal is to enhance security measures and mitigate financial risks associated with fraudulent activities in the Fastag system.



# DATASET

1. Transaction\_ID: Unique identifier for each transaction.
2. Timestamp: Date and time of the transaction.
3. Vehicle\_Type: Type of vehicle involved in the transaction.
4. FastagID: Unique identifier for Fastag.
5. TollBoothID: Identifier for the toll booth.
6. Lane\_Type: Type of lane used for the transaction.
7. Vehicle\_Dimensions: Dimensions of the vehicle.
8. Transaction\_Amount: Amount associated with the transaction.
9. Amount\_paid: Amount paid for the transaction.
10. Geographical\_Location: Location details of the transaction.
11. Vehicle\_Speed: Speed of the vehicle during the transaction.
12. Vehicle\_Plate\_Number: License plate number of the vehicle.
13. Fraud\_indicator: Binary indicator of fraudulent activity (target variable).

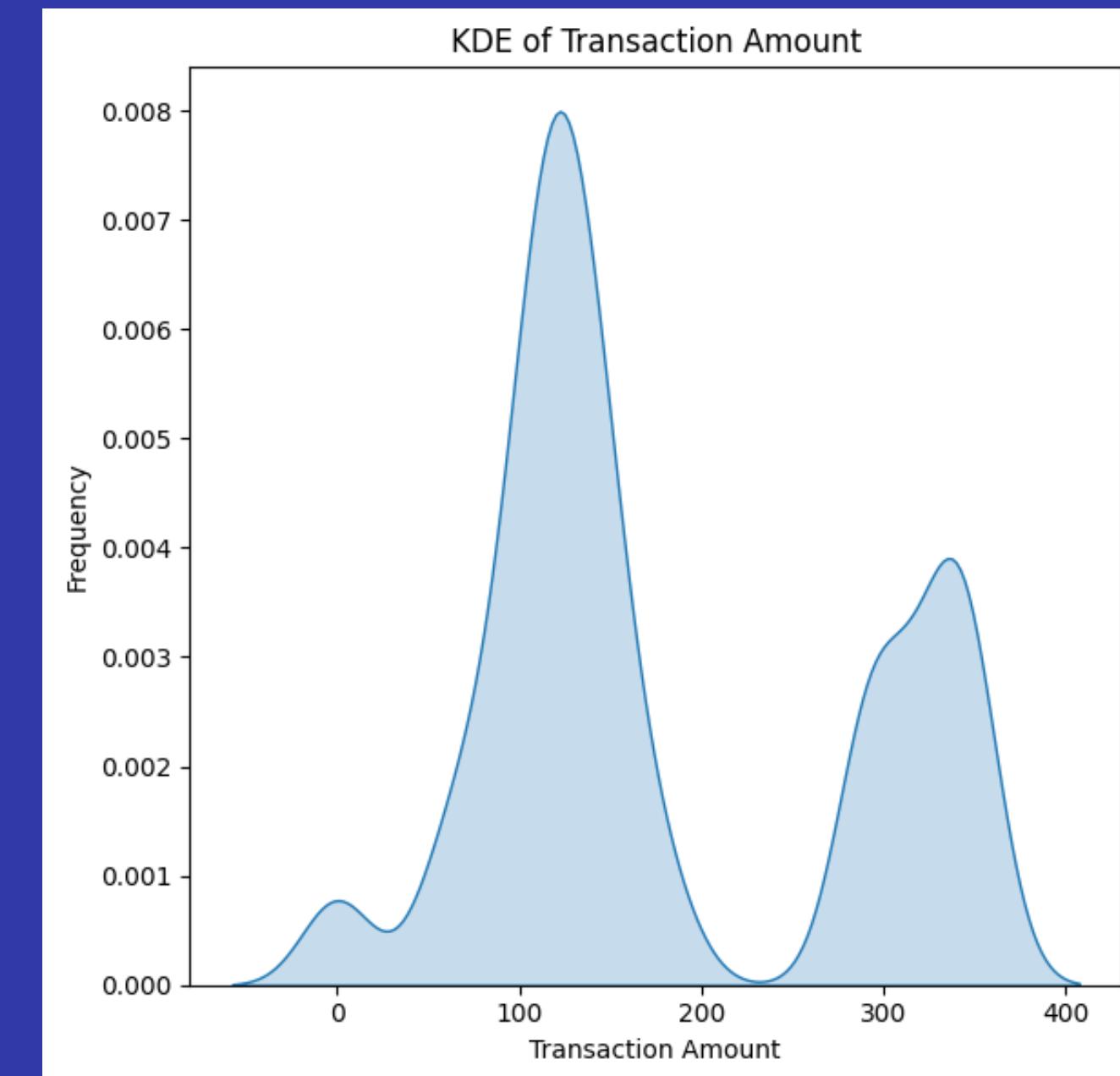


# PREPROCESSING

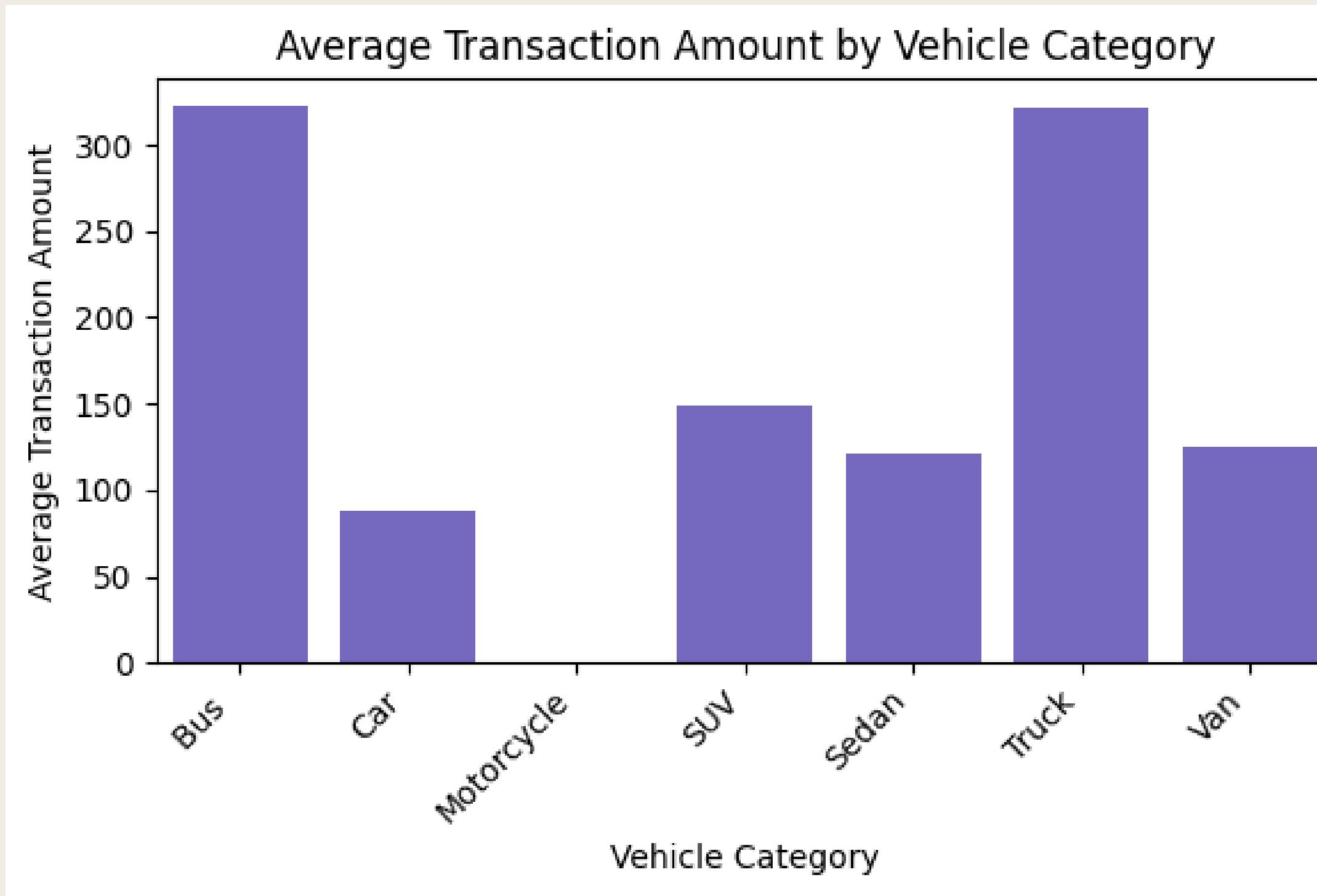
```
Transaction_ID          0  
Timestamp               0  
Vehicle_Type             0  
FastagID                549  
TollBoothID              0  
Lane_Type                 0  
Vehicle_Dimensions        0  
Transaction_Amount         0  
Amount_paid                0  
Geographical_Location       0  
Vehicle_Speed                0  
Vehicle_Plate_Number        0  
Fraud_indicator              0  
dtype: int64
```

- Removed Null FastagID

- There is no outliers in transaction amount but the bimodal shape might be because of relation with other features, let's explore!



# PREPROCESSING

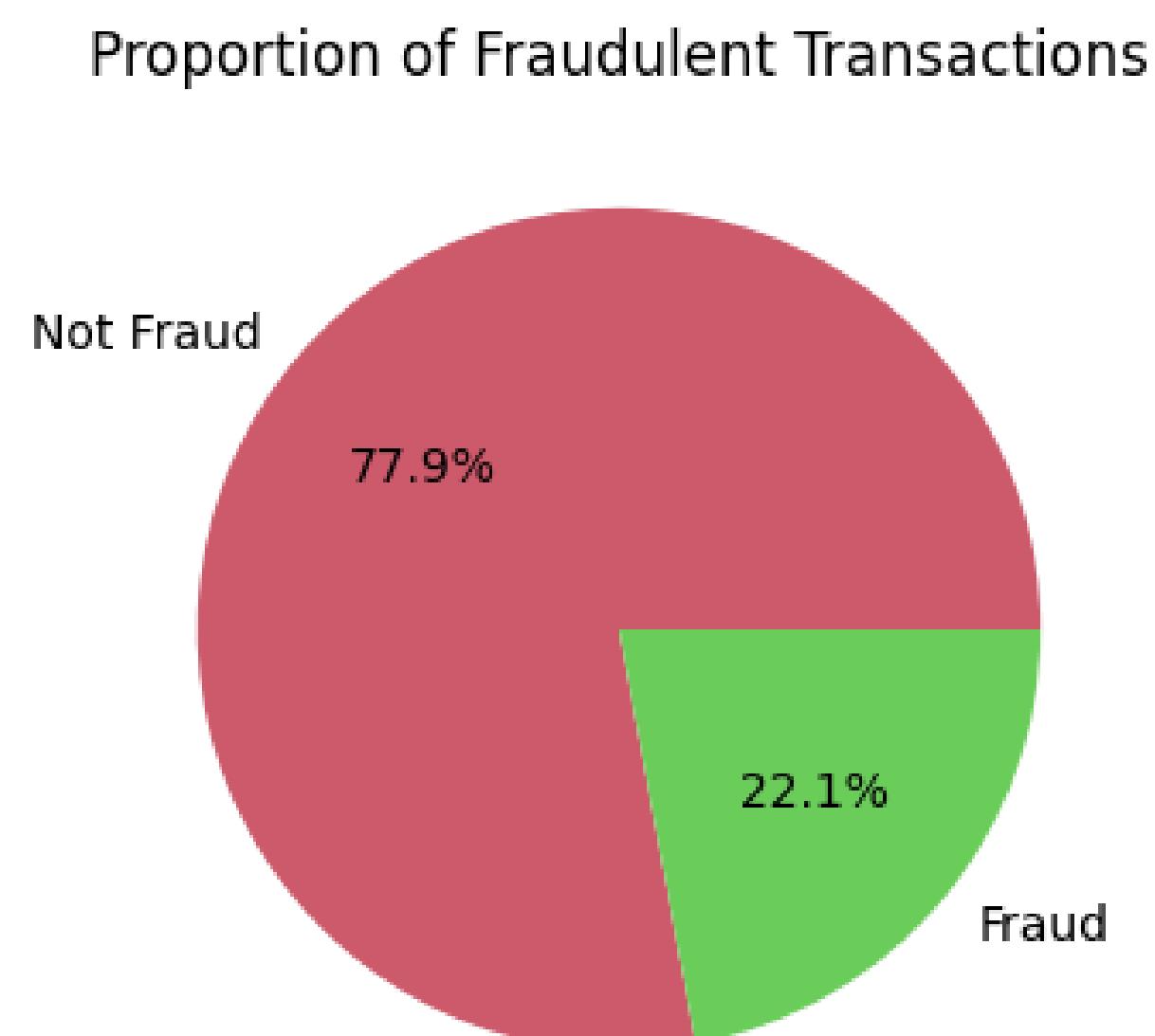


## COMMERCIAL VS. PERSONAL USAGE:

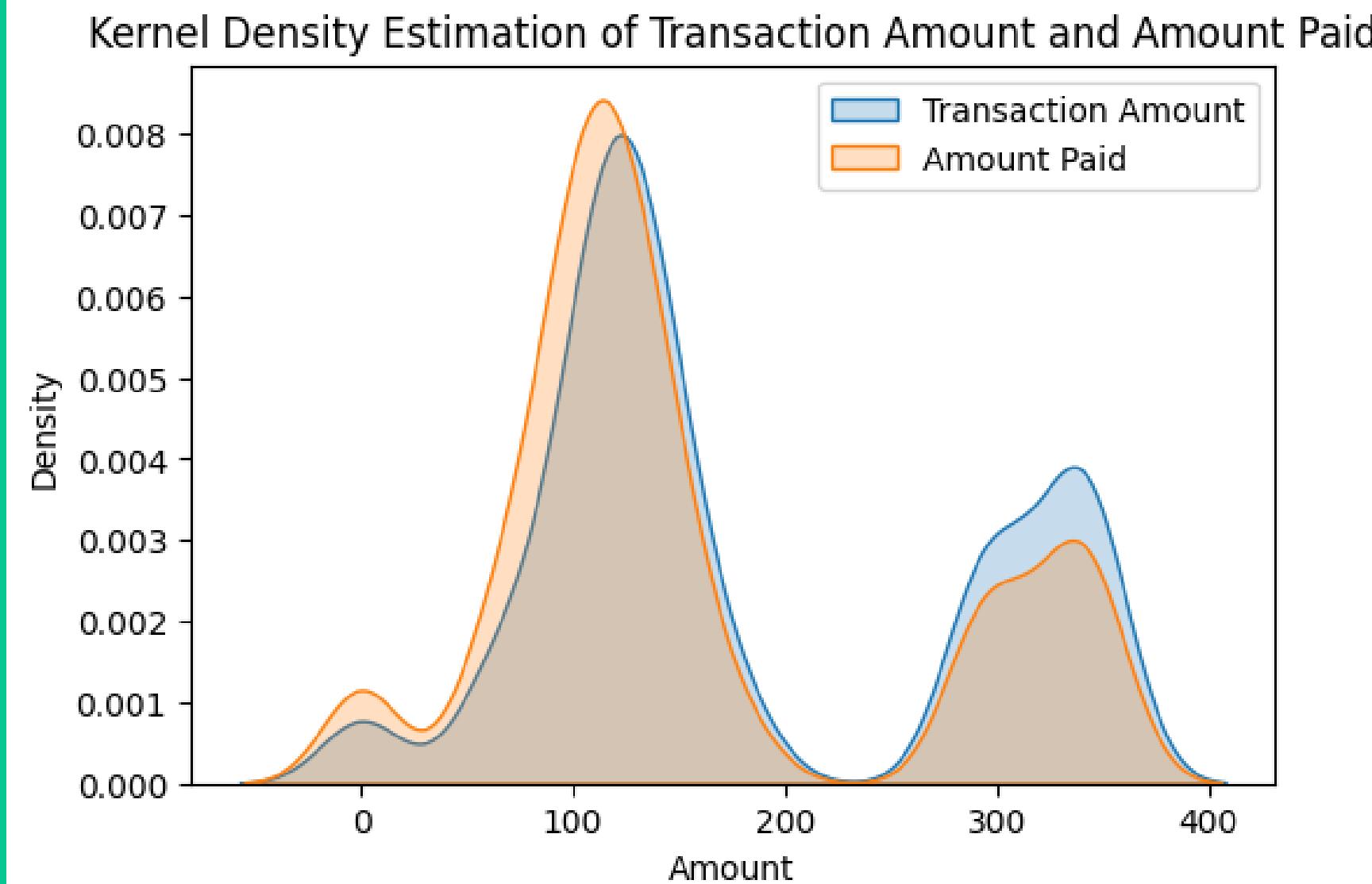
- BUSES AND TRUCKS ARE OFTEN USED FOR COMMERCIAL PURPOSES, WHILE CARS AND MOTORCYCLES ARE TYPICALLY USED FOR PERSONAL TRANSPORTATION.
- THE HIGHER TRANSACTION AMOUNTS FOR BUSES AND TRUCKS COULD INDICATE INCREASED TOLL CHARGES FOR COMMERCIAL VEHICLES COMPARED TO PERSONAL VEHICLES.

# PREPROCESSING

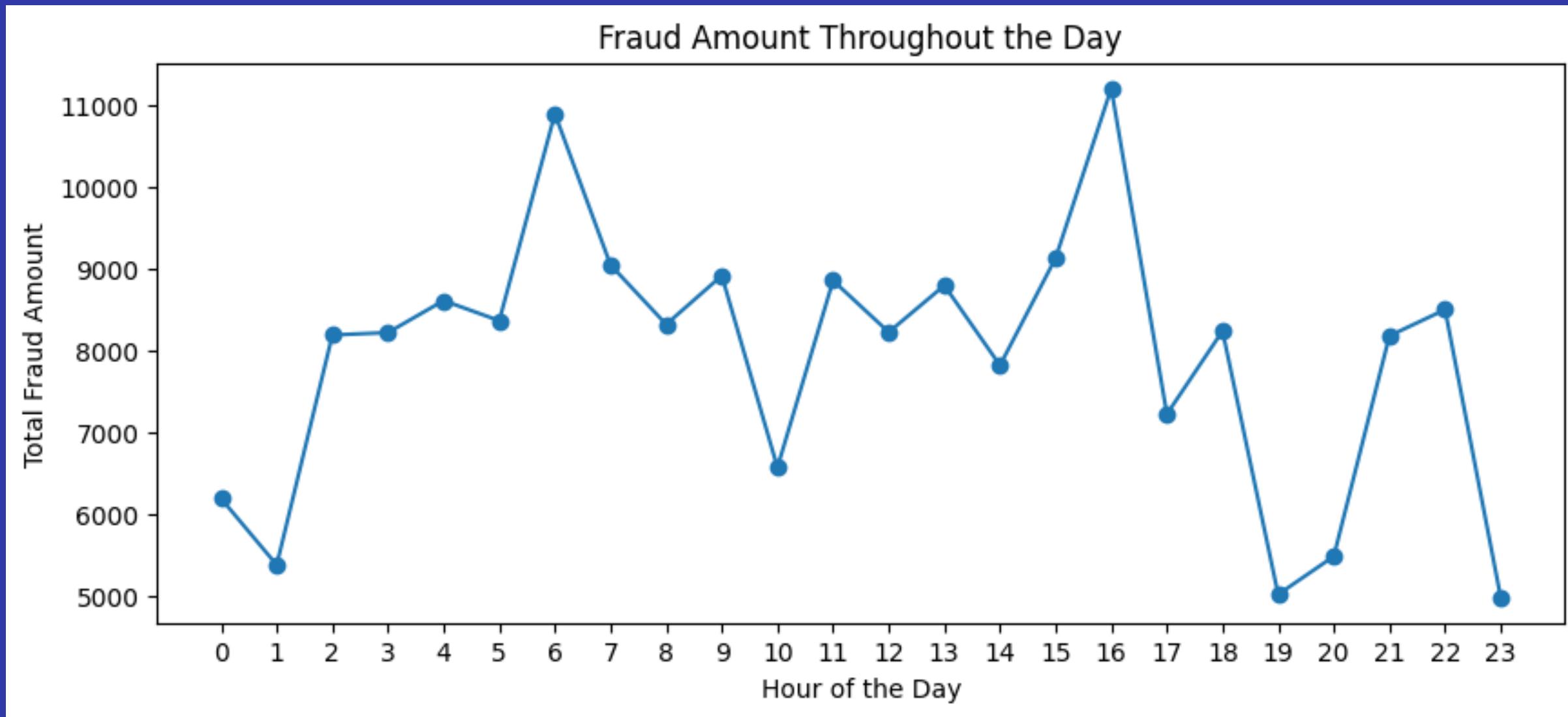
Imbalanced Dataset



Slight difference between Transaction amount and amount Paid



# PREPROCESSING



1. Fraudulent activity peaks around 6 AM, possibly due to the start of the day or commuting hours.
2. Another peak is observed around 4 PM, likely linked to the end of the workday or rush hour traffic.

# FEATURE EXTRATION

Hour	Day	Month			
11	4	1			
14	5	1			
2	0	1	Amount_Difference	Payment_Ratio	
			230	0.342857	
			20	0.833333	
			230	0.342857	
Latitude	Longitude				
13.059816	77.770687				
13.059816	77.770687				

01.

Extracting Time-based Features:

Derive day of the week and month from the timestamp.

02.

Extracting Transaction Behavior Features:

difference and payment ratio between transaction amount and amount paid.

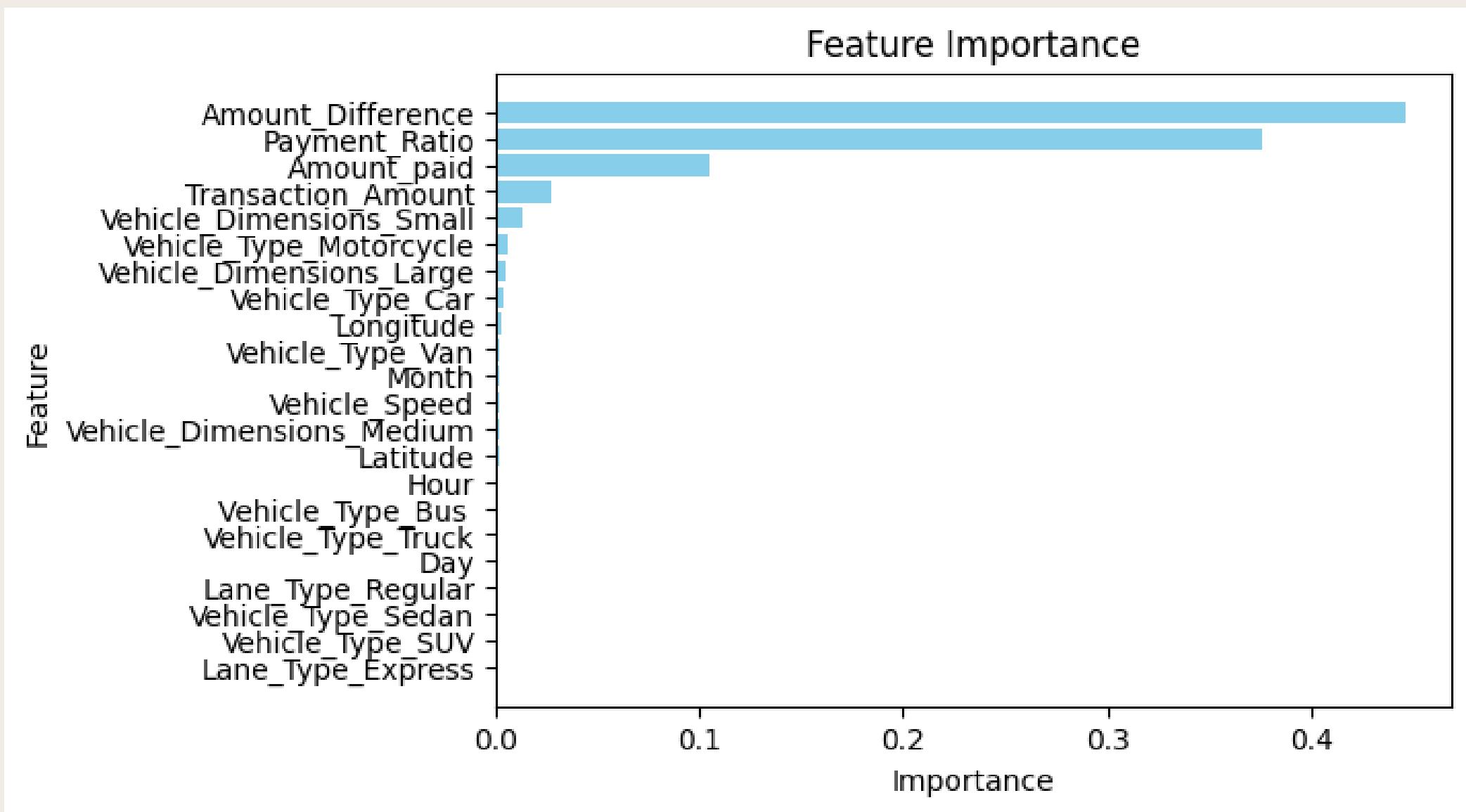
03.

Extracting Location Features:

- Extract latitude and longitude from the geographical location.

# FEATURE EXTRACTION

Extracted transaction behavior features was the most informative features to the model



# MODEL DEVELOPMENT

## Logistic regression

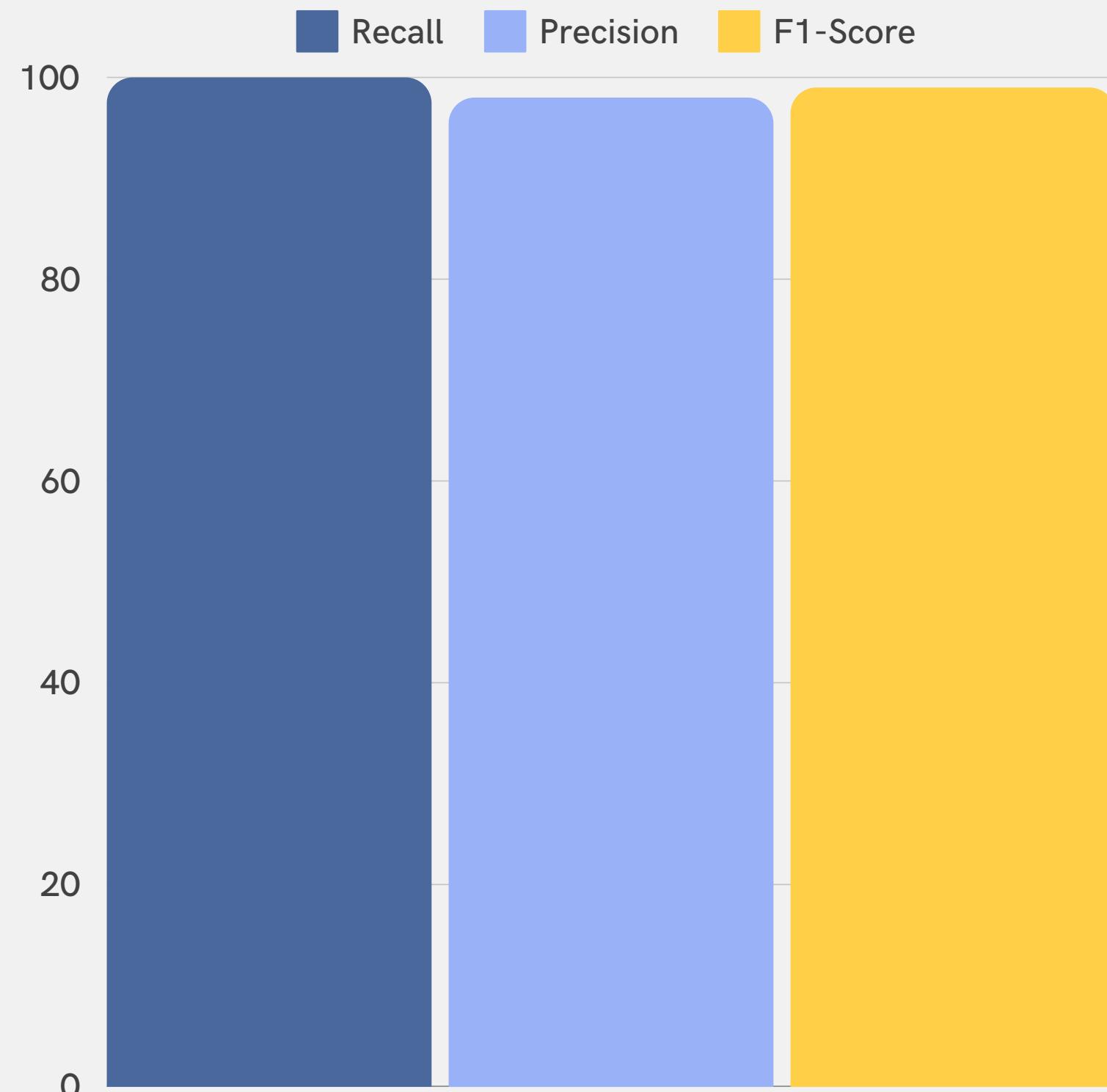
CV mean of

0.9836002976077882

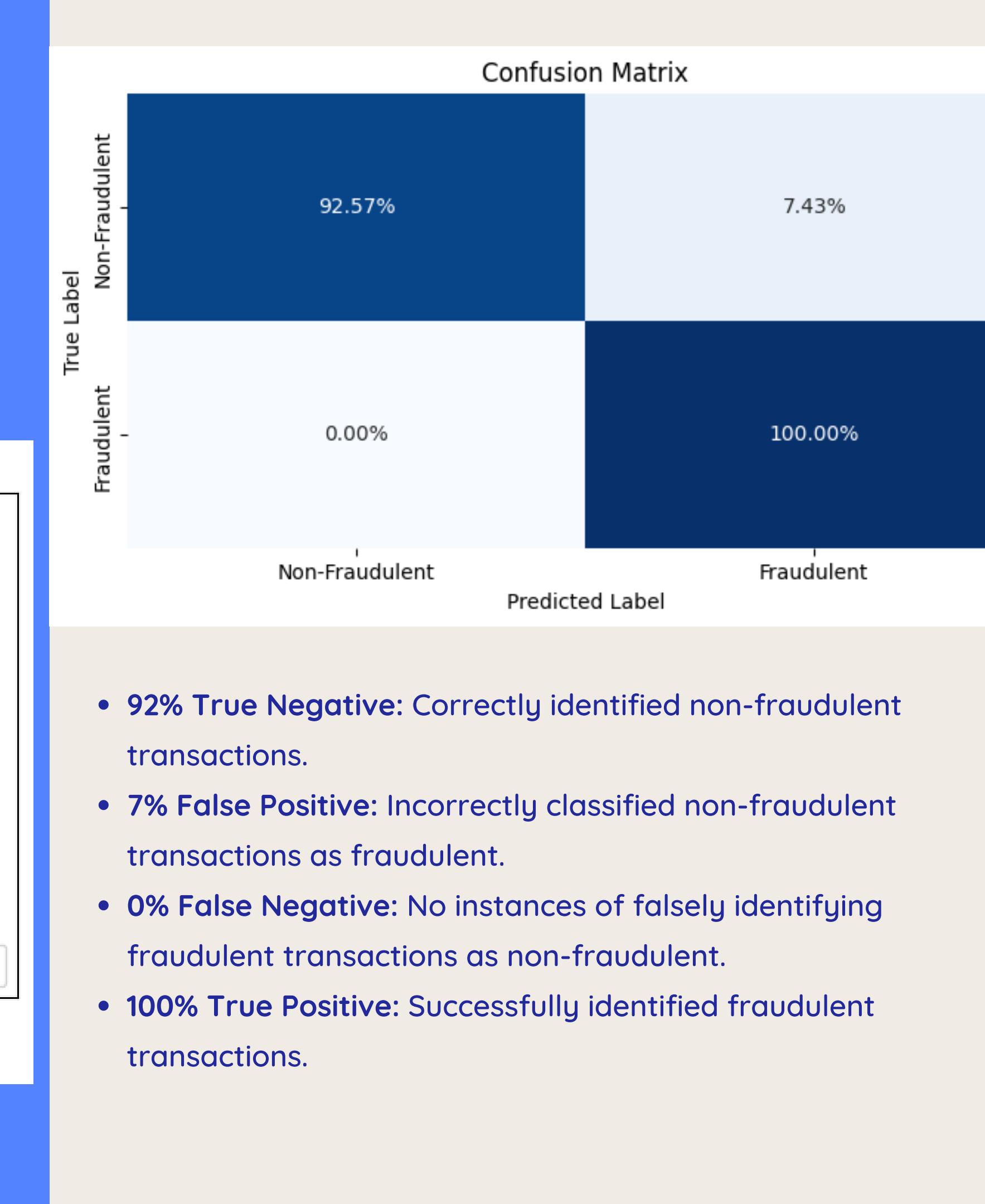
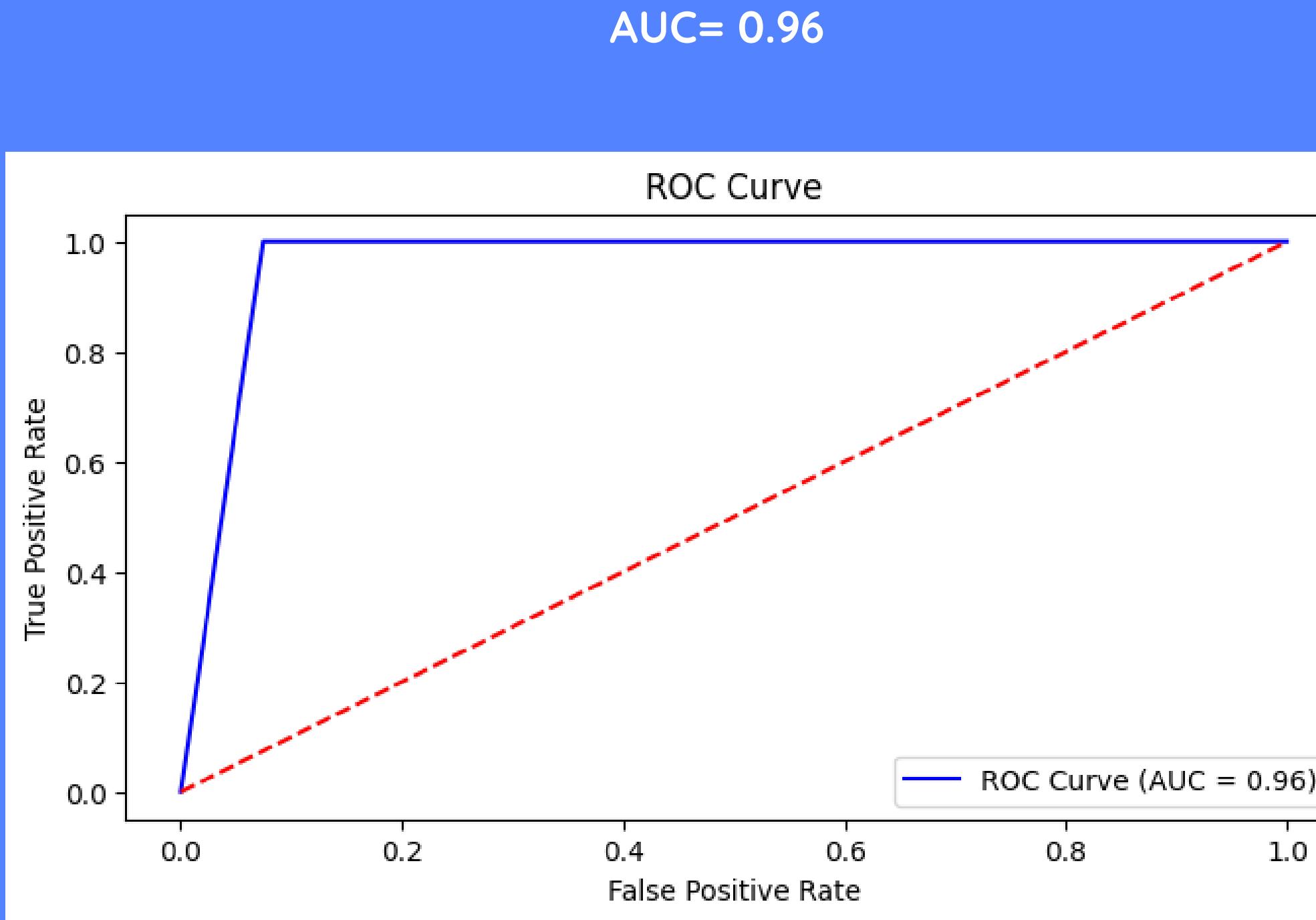
Recall: 100%

F1-score: 99%

Precision: 98%



# MODEL DEVELOPMENT



PRESENTED BY MALAK ESSAM

**THANK  
YOU VERY  
MUCH!**

