MENTOR NESS

BATCH: MIP-ML-07

TASK_2

MACHINE LEARNING

FAST TAG FRAUD DETECTION

By : IRFAN ULLAH KHAN

Follow me for more daily update about Data Science & Machine Learning

----

LinkedIn: IRFAN ULLAH KHAN
Tutorial: https://youtu.be/hwEb65ljUL0
Github: IRFAN UALLH KHAN
Portfolio: https://flowcv.me/ikm
Website: https://datasciencetoyou.odoo.com/
Group: IRFAN ULLAH KHAN

TABLE OF CONTENTS

PROBLEM STATEMENT

This project focuses on leveraging machine learning classification techniques to develop an effective fraud detection system for Fast tag transactions. The dataset comprises key features such as transaction details, vehicle information, geographical location, and transaction amounts. The goal is to create a robust model that can accurately identify instances of fraudulent activity, ensuring the integrity and security of Fast tag transactions

DATASET DESCRIPTION

1. TRANSACTION_ID
2. TIMESTAMP
3. VEHICLE_TYP
4. FASTAGID E
5. TOLLBOOTHID
6. LANE_TYPE
7. VEHICLE_DIMENSIONS
8. TRANSACTION_AMOUNT
9. AMOUNT_PAID
10. GEOGRAPHICAL_LOCATION
11. VEHICLE_SPEED
12. VEHICLE_PLATE_NUMBE R
13. FRAUD_INDICATOR

DATASET DESCRIPTION

	Transaction_ID	Timestamp	Vehicle_Type	FastagID	TollBoothID	Lane_Type	Vehicle_Dimensions	Transaction_Amount	Amount_paid	Geographical_Location	Vehicle_Speed	Vehicle_Plate_Number	Fraud_indicator
0	1	2023-01-06 11:20:00	Bus	FTG-001-ABC-121	A-101	Express	Large	350	120	13.059816123454882, 77.77068662374292	65	KA11AB1234	Fraud
1	2	2023-01-07 14:55:00	Car	FTG-002-XYZ-451	B-102	Regular	Small	120	100	13.059816123454882, 77.77068662374292	78	KA66CD5678	Fraud
3	4	2023-01-09 02:05:00	Truck	FTG-044-LMN-322	C-103	Regular	Large	350	120	13.059816123454882, 77.77068662374292	92	KA11GH3456	Fraud
4	5	2023-01-10 06:35:00	Van	FTG-505-DEF-652	B-102	Express	Medium	140	100	13.059816123454882, 77.77068662374292	60	KA44IJ6789	Fraud
5	6	2023-01-11 10:00:00	Sedan	FTG-066-GHI-987	A-101	Regular	Medium	160	100	13.059816123454882, 77.77068662374292	105	KA77KL0123	Fraud
6	7	2023-01-12 15:40:00	SUV	FTG-707-JKL-210	B-102	Express	Large	180	160	13.059816123454882, 77.77068662374292	70	KA22MN4567	Fraud
7	8	2023-01-13 20:15:00	Bus	FTG-088-UVW-543	C-103	Regular	Large	350	90	13.059816123454882, 77.77068662374292	88	KA21OP8901	Fraud
8	9	2023-01-14 01:55:00	Car	FTG-909-RST-876	A-101	Express	Small	120	0	13.059816123454882, 77.77068662374292	45	KA16QR2345	Fraud
10	11	2023-01-16 12:10:00	Truck	FTG-021-QWE-765	C-103	Express	Large	350	120	13.059816123454882, 77.77068662374292	58	KA12UV0123	Fraud
11	12	2023-01-17 17:45:00	Van	FTG-011-ZXC-431	B-102	Regular	Medium	140	120	13.059816123454882, 77.77068662374292	81	KA35WX3454	Fraud
12	13	2023-01-18 22:20:00	Sedan	FTG-013-POI-104	A-101	Express	Medium	160	120	13.059816123454882, 77.77068662374292	67	KA38YZ6785	Fraud
13	14	2023-01-19 04:00:00	SUV	FTG-014-KJH-872	B-102	Regular	Large	180	180	13.059816123454882, 77.77068662374292	98	KA14AB0123	Not Fraud
14	15	2023-01-20 08:30:00	Bus	FTG-055-DCV-543	A-101	Express	Large	350	120	13.059816123454882, 77.77068662374292	50	KA40CD4557	Fraud
15	16	2023-01-21 13:10:00	Car	FTG-066-NBH-210	B-102	Regular	Small	120	90	13.059816123454882, 77.77068662374292	75	KA17EF8801	Fraud
17	18	2023-01-23 22:25:00	Truck	FTG-088-EYT-654	C-103	Regular	Large	350	120	13.059816123454882, 77.77068662374292	84	KA53IJ5789	Fraud
18	19	2023-01-24 02:55:00	Van	FTG-099-FTD-321	B-102	Express	Medium	140	120	13.059816123454882, 77.77068662374292	55	KA56KL0223	Fraud
19	20	2023-01-25 07:35:00	Sedan	FTG-120-TTU-098	A-101	Regular	Medium	160	160	13.059816123454882, 77.77068662374292	80	KA59MN4557	Not Fraud
20	21	2023-01-26 12:10:00	SUV	FTG-121-BTC-765	B-102	Express	Large	180	100	13.059816123454882, 77.77068662374292	68	KA62OP7901	Fraud
21	22	2023-01-27 15:45:00	Bus	FTG-122-XRW-432	C-103	Regular	Large	350	350	13.059816123454882, 77.77068662374292	90	KA65QR1345	Not Fraud
22	23	2023-01-28 21:25:00	Car	FTG-123-NMK-101	A-101	Express	Small	120	90	13.059816123454882, 77.77068662374292	48	KA88ST6784	Fraud

PROJECT OBJECTIVES

01

DATA EXPLORATION

Explore the dataset to understand the distribution of features and the prevalence of fraud indicators

02

FEATURE ENGINEERING

Identify and engineer relevant features that contribute to fraud detection accuracy

03

MODEL DEVELOPMENT

Build a machine learning classification model to predict and detect Fastag transaction fraud.
- Evaluate and fine-tune model performance using appropriate metrics

04

REAL-TIME FRAUD DETECTION

Explore the feasibility of implementing the model for real-time Fastag fraud detection

05

EXPLANATORY ANALYSIS

Provide insights into the factors contributing to fraudulent transactions

CHALLENGES

01

- Imbalanced dataset issues due to the likely low occurrence of fraud.

VS

02

- Feature engineering to capture nuanced patterns indicative of fraud

EVALUATION CRITERIA

MODEL PERFORMANCE ASSESSED USING METRICS

PRECISION

Imbalanced dataset issues due to the likely low occurrence of fraud.

RECALL

Feature engineering to capture nuanced patterns indicative of fraud

F1 SCORE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam vel euismod ipsum. Proin fermentum dolor vel est fermentum, at sagittis diam.

PROJECT SCOPE

LOAD THE DATASET

Load Dataset

```
[2] df = pd.read_csv('/content/FastagFraudDetection.csv')
```

```
[3] df.head()
```

	Transaction_ID	Timestamp	Vehicle_Type	FastagID	TollBoothID	Lane_Type	Vehicle_Dimensions	Transaction_Amount	Amount_paid	Geographical_Location	Vehicle_Speed	Vehicle_Plate_Number	Fraud_indicator
0	1	01-06-23 11:20	Bus	FTG-001-ABC-121	A-101	Express	Large	350	120	13.059816123454882, 77.77068662374292	65	KA11AB1234	Fraud
1	2	01-07-23 14:55	Car	FTG-002-XYZ-451	B-102	Regular	Small	120	100	13.059816123454882, 77.77068662374292	78	KA66CD5678	Fraud
2	3	01-08-23 18:25	Motorcycle	NaN	D-104	Regular	Small	0	0	13.059816123454882, 77.77068662374292	53	KA88EF9012	Not Fraud
3	4	01-09-23 2:05	Truck	FTG-044-LMN-322	C-103	Regular	Large	350	120	13.059816123454882, 77.77068662374292	92	KA11GH3456	Fraud
4	5	01-10-23 6:35	Van	FTG-505-DEF-652	B-102	Express	Medium	140	100	13.059816123454882, 77.77068662374292	60	KA44IJ6789	Fraud

loading...

DATA SUMMARY

```
[5] df.info()
```

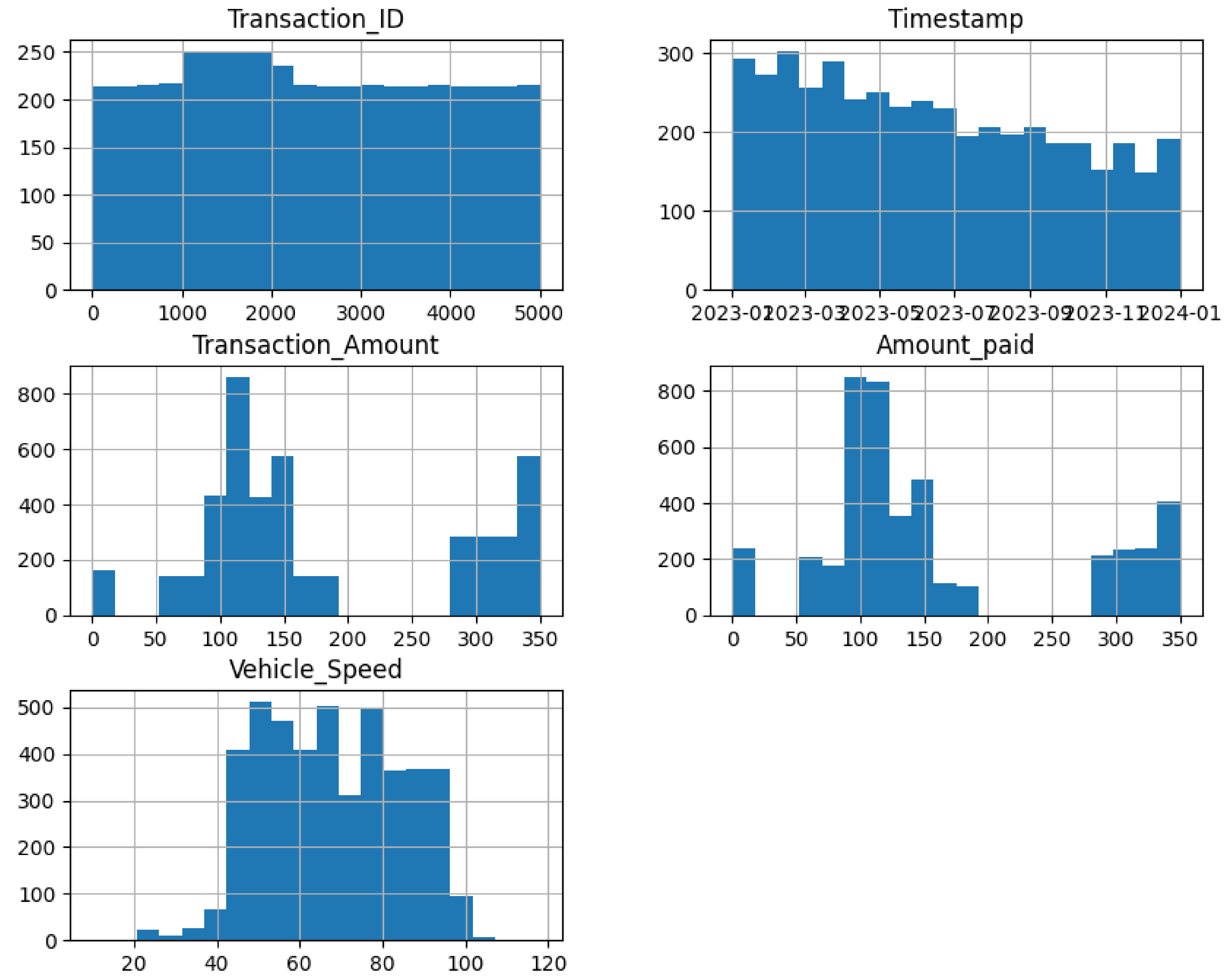
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Transaction_ID        5000 non-null   int64
 1   Timestamp             5000 non-null   object
 2   Vehicle_Type          5000 non-null   object
 3   FastagID              4451 non-null   object
 4   TollBoothID           5000 non-null   object
 5   Lane_Type             5000 non-null   object
 6   Vehicle_Dimensions    5000 non-null   object
 7   Transaction_Amount    5000 non-null   int64
 8   Amount_paid           5000 non-null   int64
 9   Geographical_Location 5000 non-null   object
 10  Vehicle_Speed         5000 non-null   int64
 11  Vehicle_Plate_Number  5000 non-null   object
 12  Fraud_indicator       5000 non-null   object
dtypes: int64(4), object(9)
memory usage: 507.9+ KB
```

```
df.describe()
```

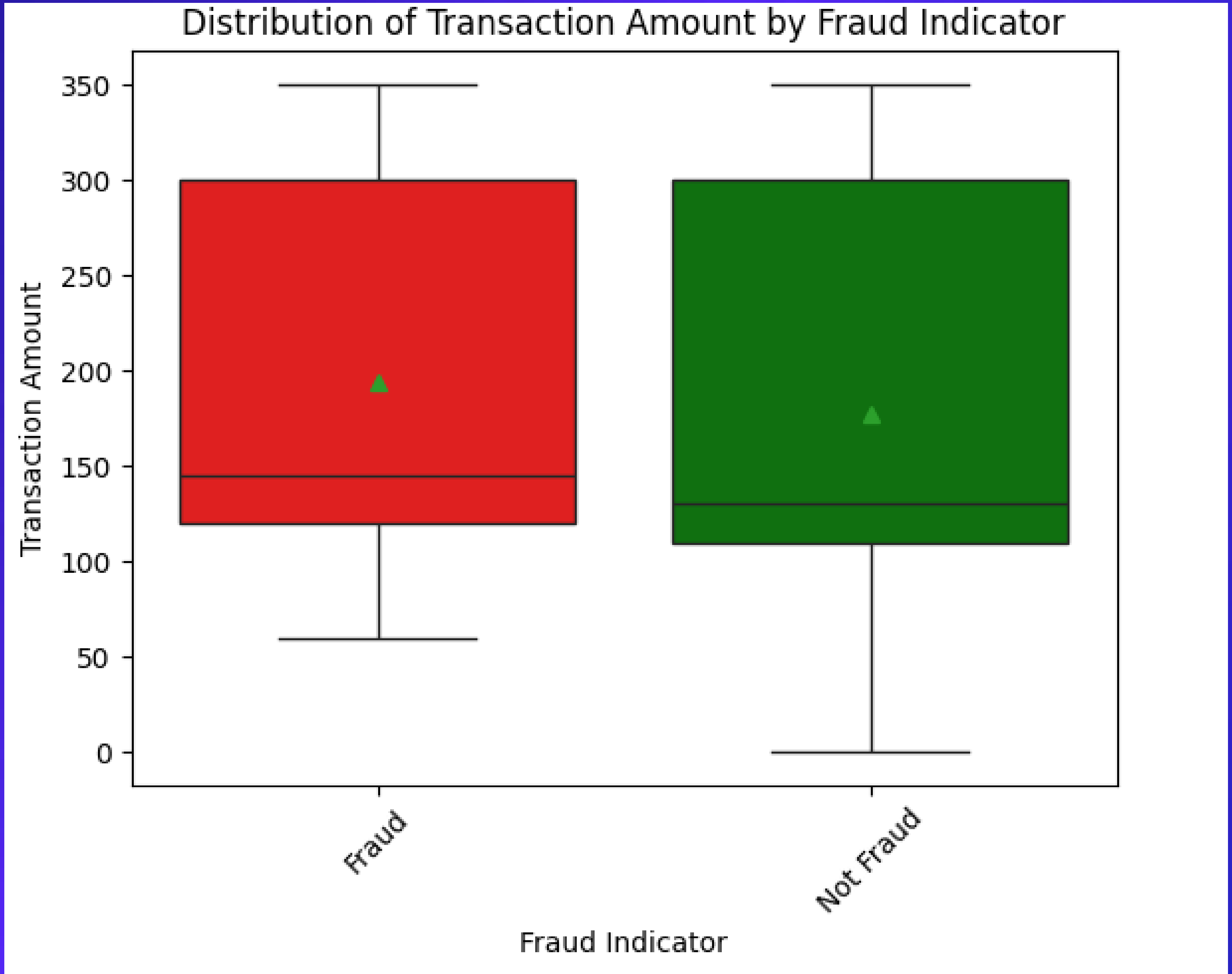
	Transaction_ID	Transaction_Amount	Amount_paid	Vehicle_Speed
count	5000.000000	5000.00000	5000.000000	5000.000000
mean	2500.500000	161.06200	141.261000	67.851200
std	1443.520003	112.44995	106.480996	16.597547
min	1.000000	0.00000	0.000000	10.000000
25%	1250.750000	100.00000	90.000000	54.000000
50%	2500.500000	130.00000	120.000000	67.000000
75%	3750.250000	290.00000	160.000000	82.000000
max	5000.000000	350.00000	350.000000	118.000000

DATA VISUALIZATION

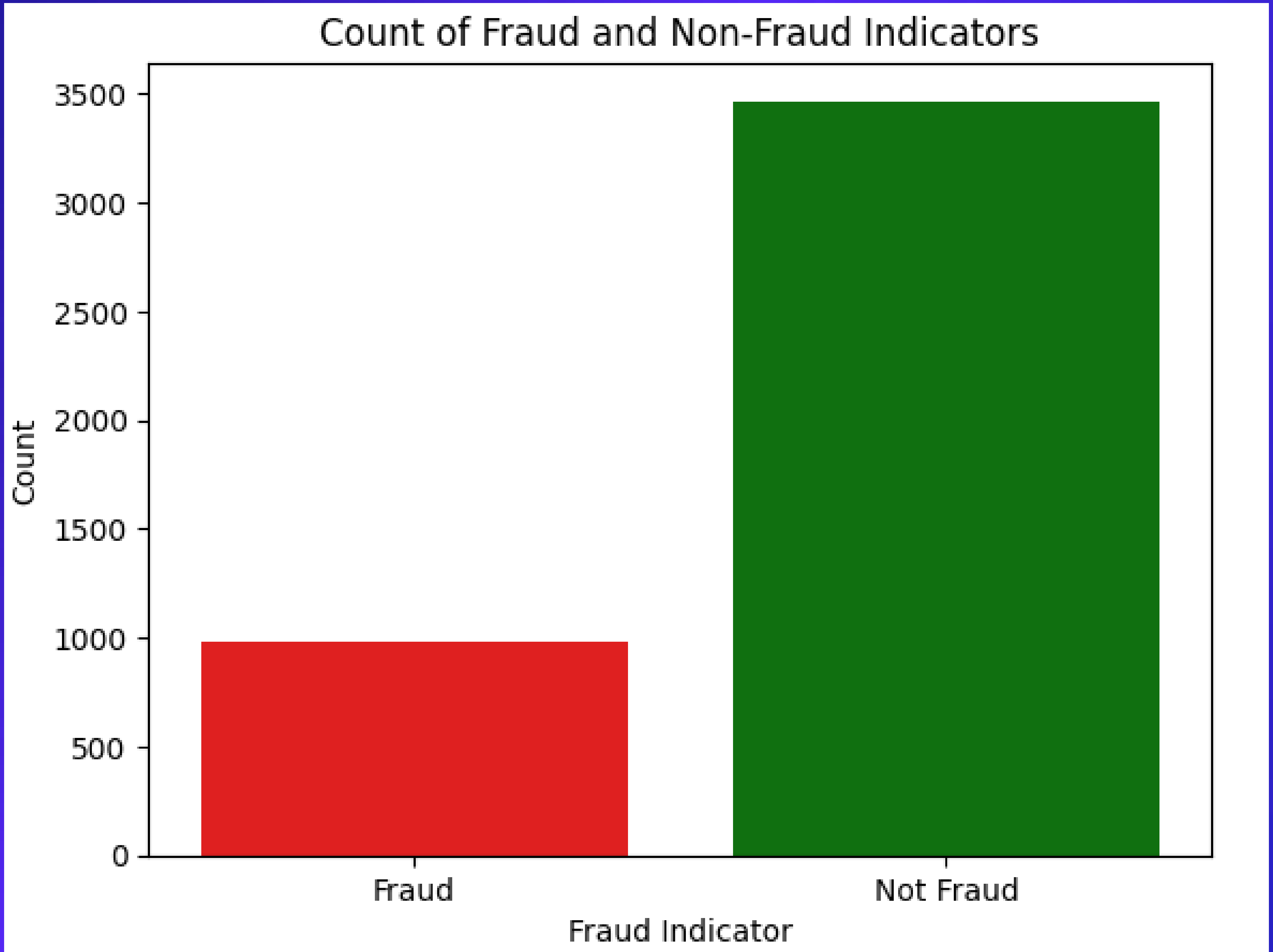
```
df.hist(figsize=(10, 8), bins=20)
plt.show()
```



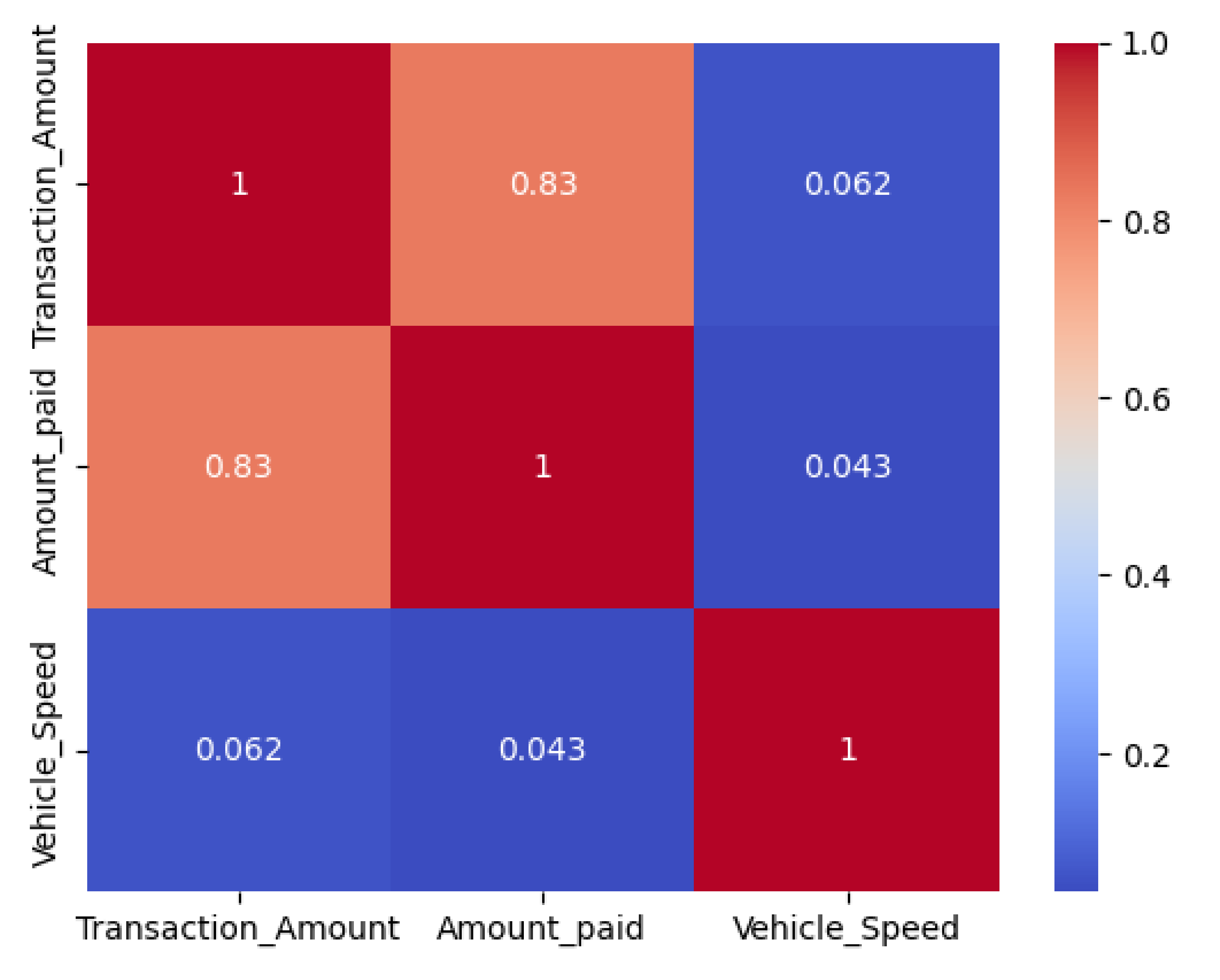
BOX PLOTS FOR NUMERICAL FEATURES



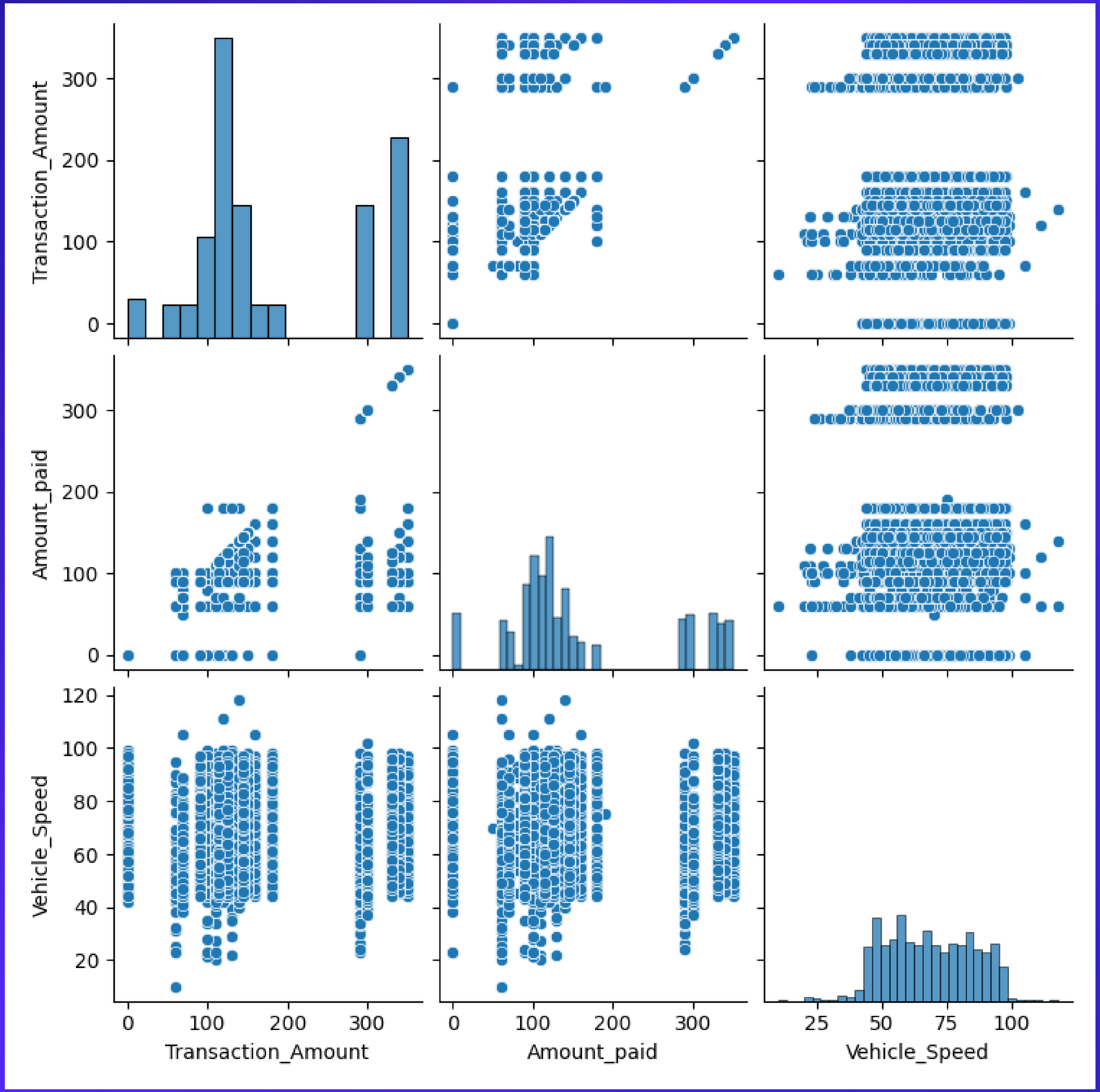
BAR PLOTS FOR CATEGORICAL FEATURES



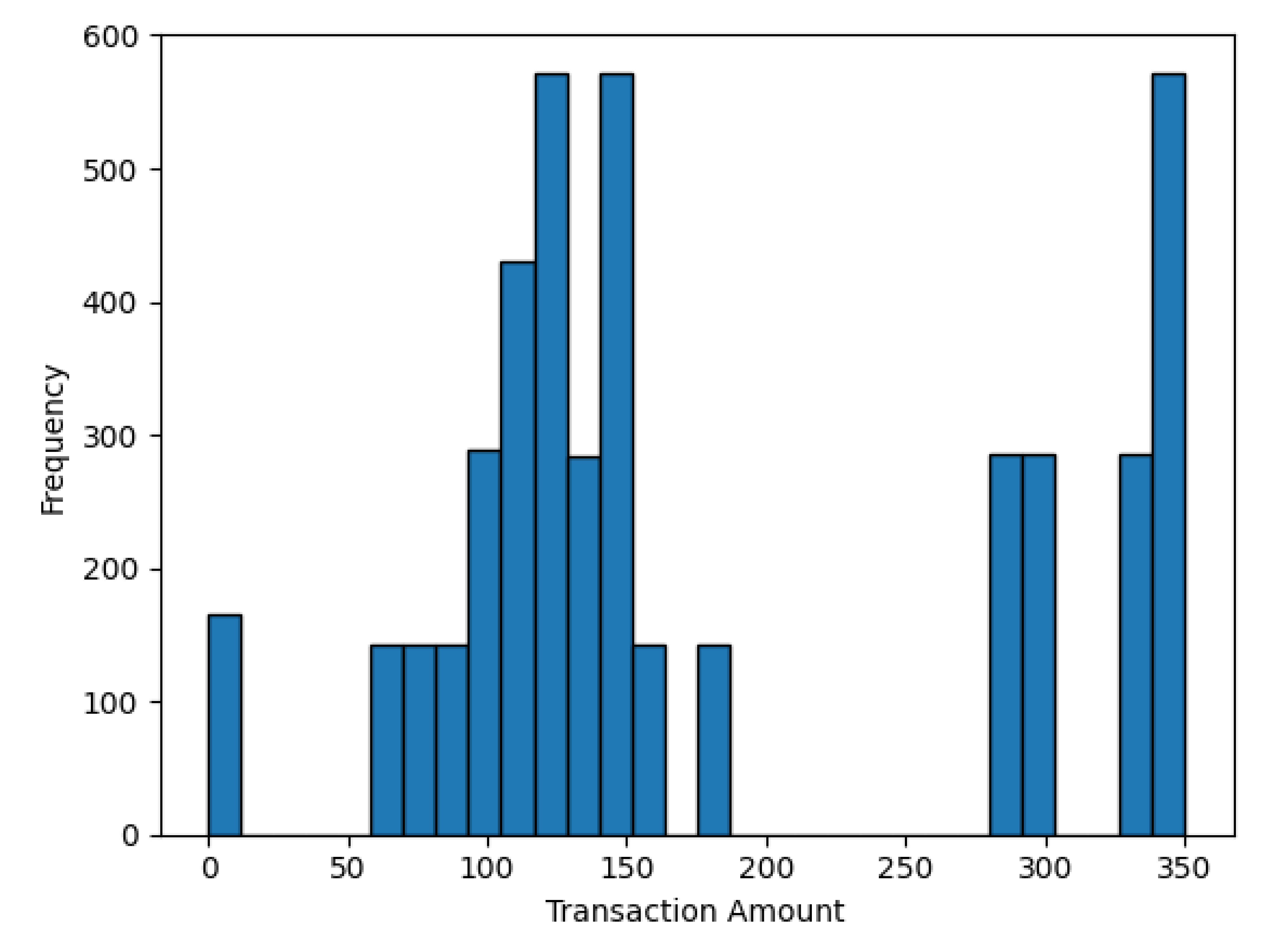
CORRELATION ANALYSIS



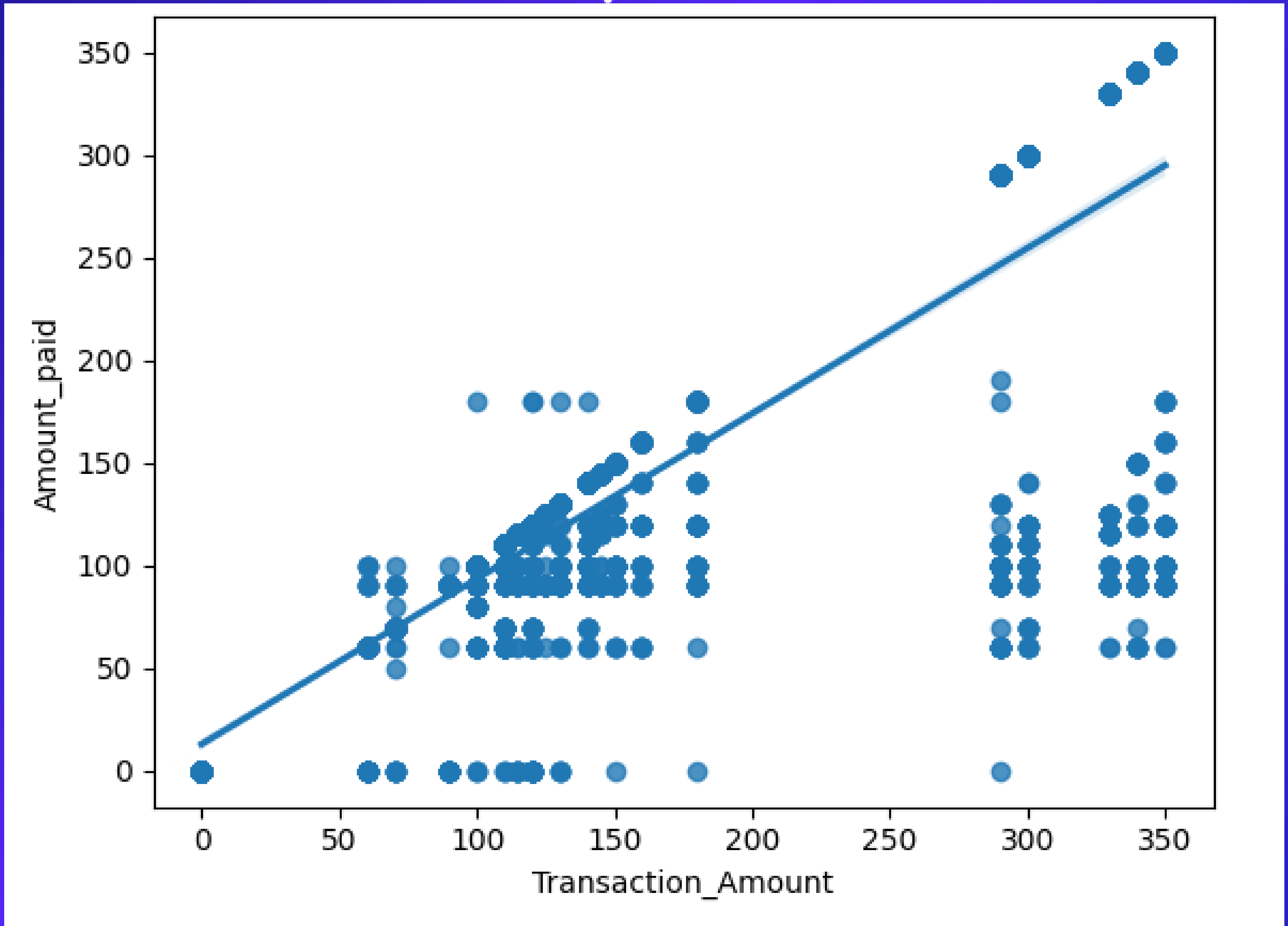
PAIRWISE SCATTER PLOTS FOR NUMERICAL VARIABLES



HISTOGRAM OF TRANSACTION_AMOUNT



SCATTER PLOT WITH REGRESSION LINE BETWEEN 'TRANSACTION_AMOUNT' AND 'AMOUNT_PAID'



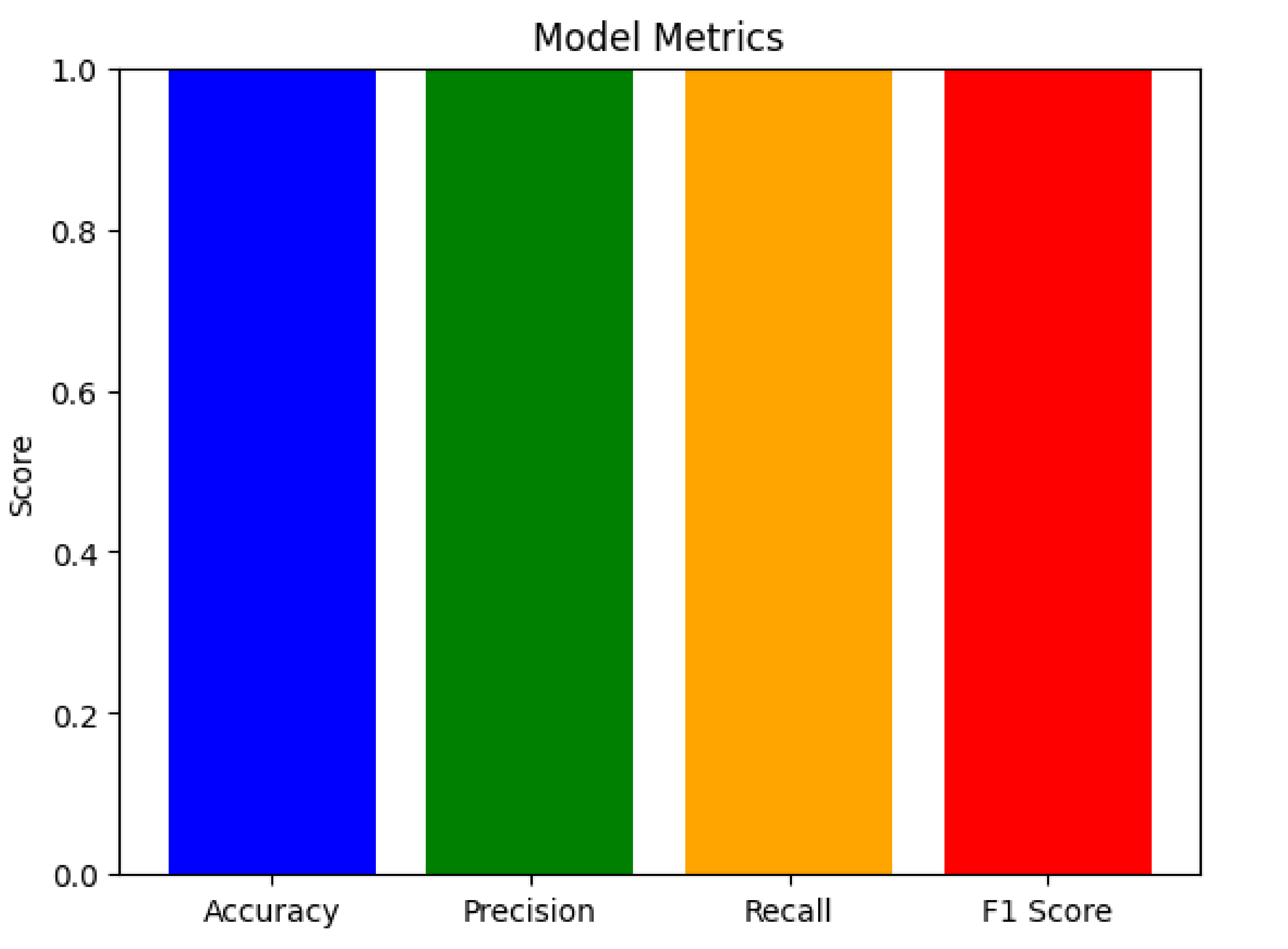
MODEL DEVELOPMENT

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 32)                96

 dense_1 (Dense)             (None, 16)                528

 dense_2 (Dense)             (None, 1)                 17

=================================================================
Total params: 641 (2.50 KB)
Trainable params: 641 (2.50 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

DISPLAY ACCURACY METRICS

```
y_pred_prob = model.predict(X_test_scaled)
# Convert probabilities to binary predictions
y_pred = np.round(y_pred_prob)
# Print accuracy metrics
accuracy = accuracy_score(y_test_encoded, y_pred)
precision = precision_score(y_test_encoded, y_pred)
recall = recall_score(y_test_encoded, y_pred)
f1 = f1_score(y_test_encoded, y_pred)
# Print accuracy metrics
print("Accuracy: {:.2f}%".format(accuracy * 100))
print("Precision: {:.2f}%".format(precision * 100))
print("Recall: {:.2f}%".format(recall * 100))
print("F1 Score: {:.2f}%".format(f1 * 100))
```

```
28/28 [==============================] - 0s 1ms/step
Accuracy: 98.43%
Precision: 98.04%
Recall: 100.00%
F1 Score: 99.01%
```



DELIVERABLES

(1)

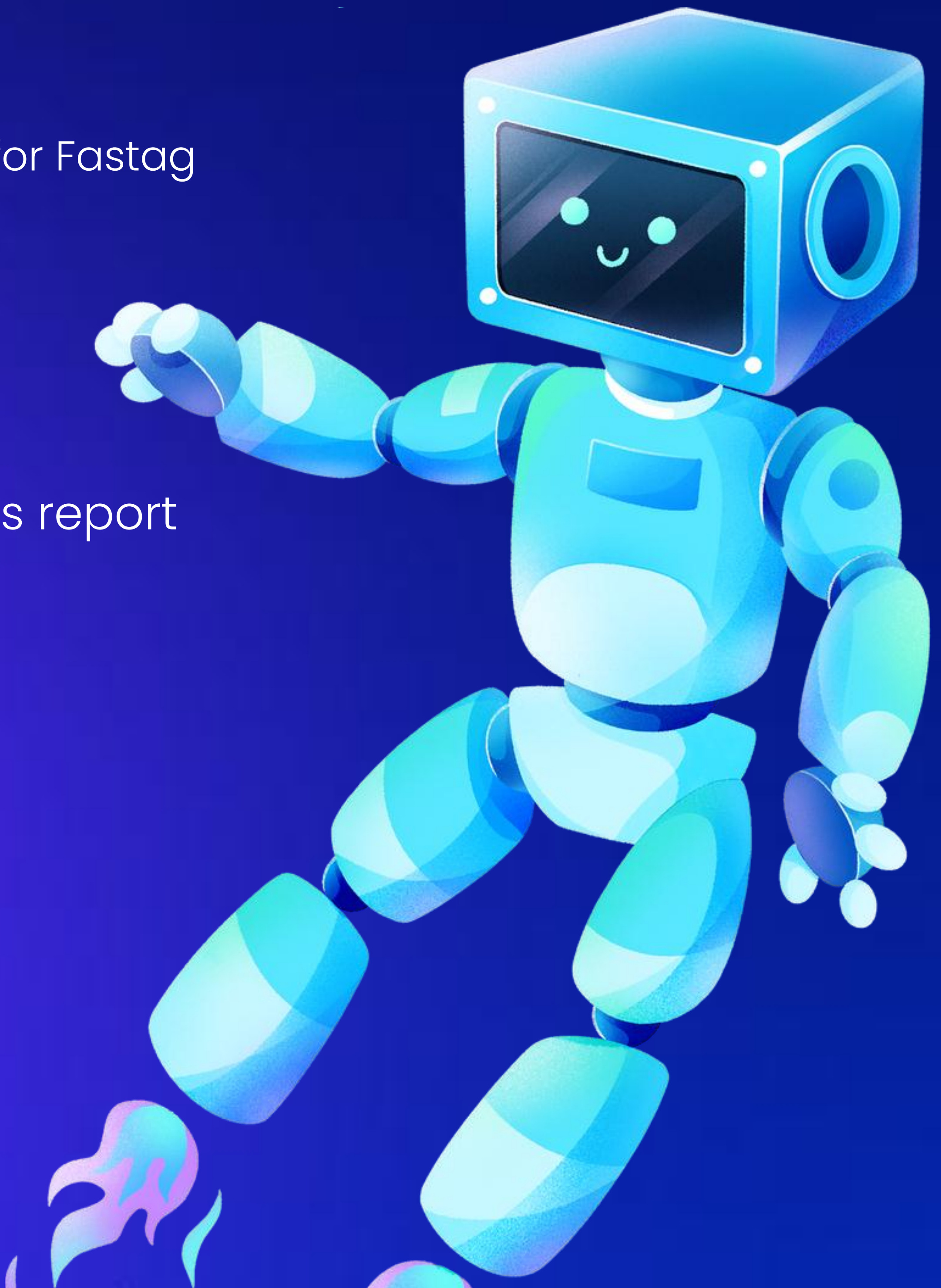
Trained machine learning model for Fastag fraud detection

(2)

Evaluation metrics and analysis report

(3)

Documentation on relevant features and their impact on fraud detection



EXPECTED OUTCOME

An effective and scalable Fastag fraud detection system capable of minimizing financial losses and ensuring the security of digital toll transactions

THANK YOU!

Follow me for more daily update about Data Science & Machine Learning

----

LinkedIn: IRFAN ULLAH KHAN
Tutorial: https://youtu.be/hwEb65ljUL0
Github: IRFAN UALLH KHAN
Portfolio: https://flowcv.me/ikm
Website: https://datasciencetoyou.odoo.com/
Group: IRFAN ULLAH KHAN