DATA VISUALIZATION

SET 4

Data: Transaction.csv

Name: Chirag Rao KV

Reg no: 240962180

Sec: P C

Roll no: 66

**Data overview and Exploration**

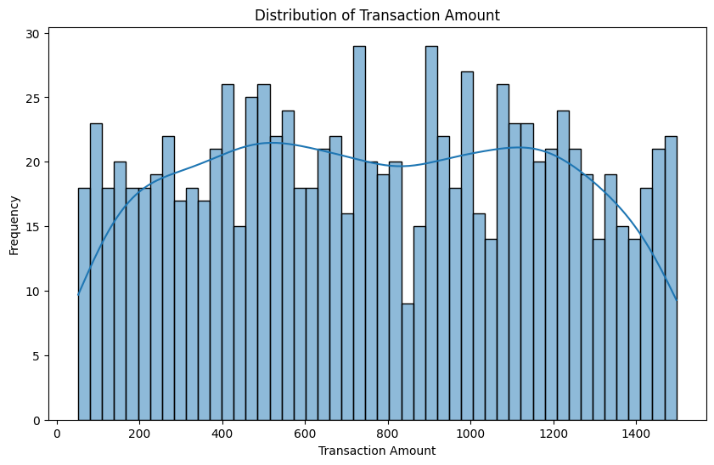
**Columns**

* + Transaction ID
  + Sender Account ID
  + Receiver Account ID
  + Transaction Amount
  + Transaction Type
  + Timestamp
  + Transaction Status
  + Fraud Flag
  + Geolocation (Latitude/Longitude)
  + Device Used
  + Amount Range

**Duplicates and NAN values**

* As there is no nan values in the data set no need to fill nan values.
* As there are no duplicates in dataset no need to remove any.

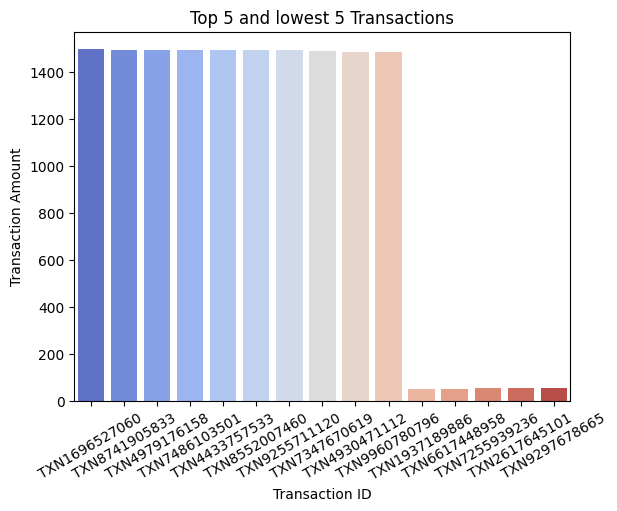
**Data visualization and analysis**

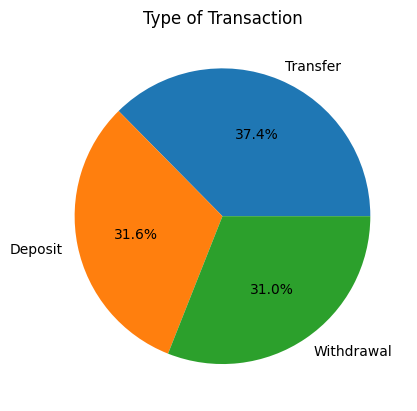
**Transaction amount distribution**

This histogram shows very equal distribution of Transaction amount across the range. This means transactions of large amount and smaller amount are equally expected

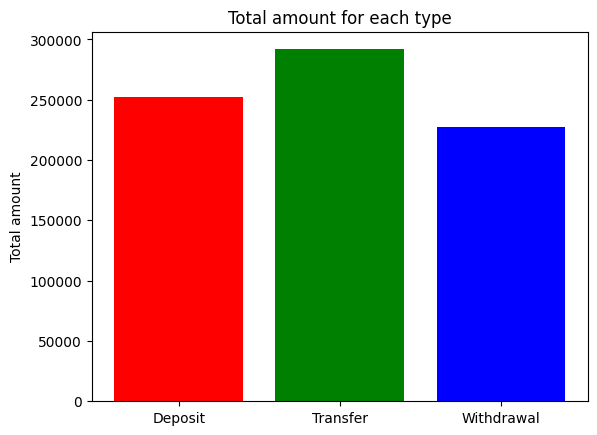
Transaction extremes

The top 5 transaction amount and bottom 5 transaction can be visualized here

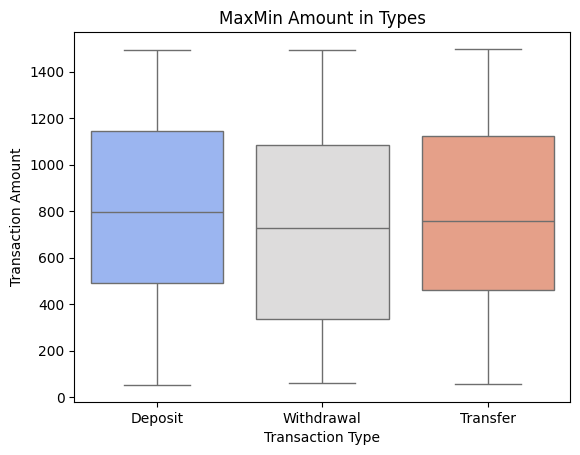


**Type of transaction**

The pie plot shows the distribution of transactions types. This shows transfer has the highest number of transactions.

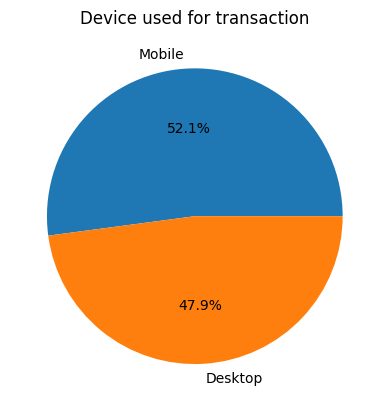
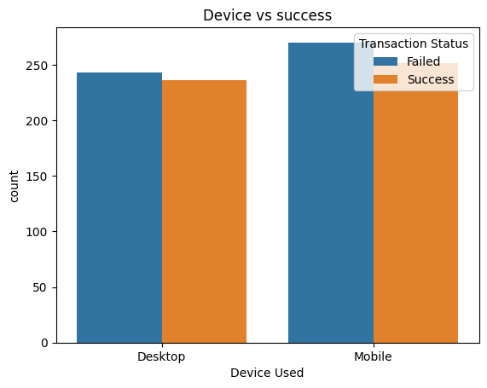
**Gross transaction across types**

This graph shows transfer has the most total amount of data, followed by deposit and withdrawal.

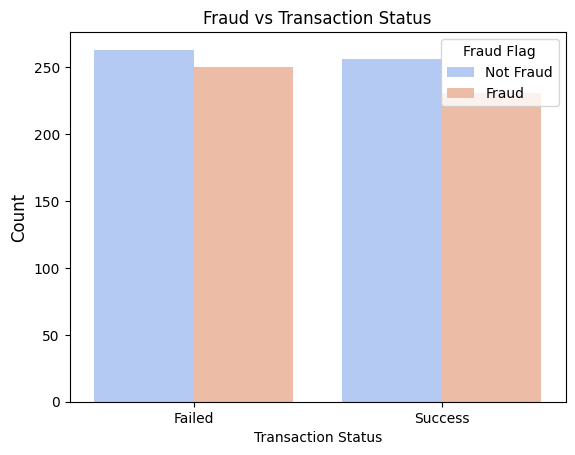
**Max and min across type of transactions**

This histogram plot shows deposit has highest single transaction amount and also the highest average across all the types.

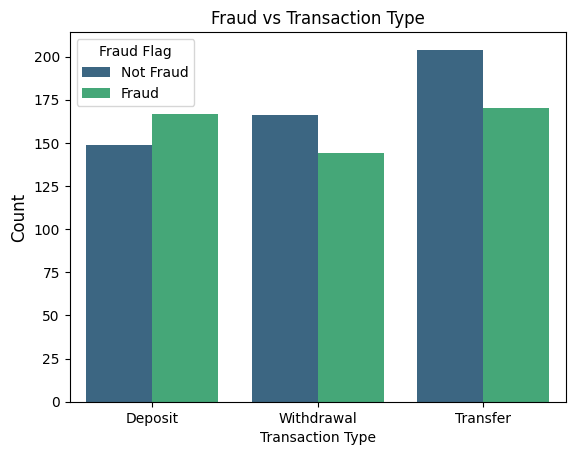
**Type of Device and fraud transactions**



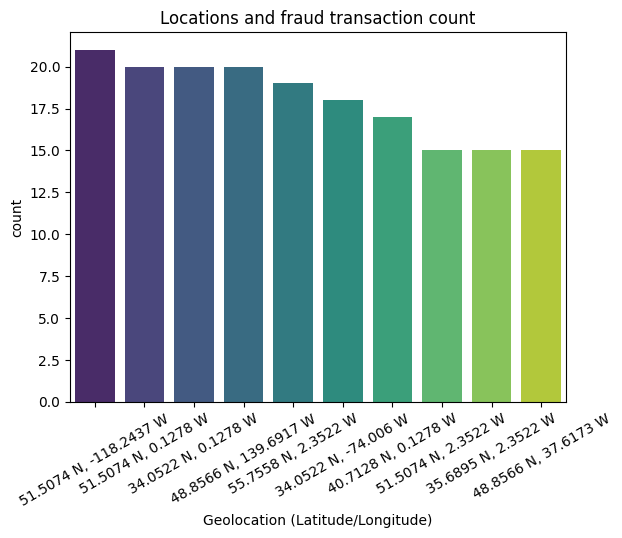
* + Mobile users are dominant among the transactions.
  + Mobile transactions are more prone to failed transactions than desktop.

**Transaction status and relation with fraud**

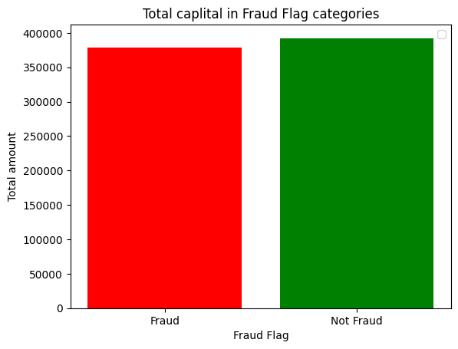
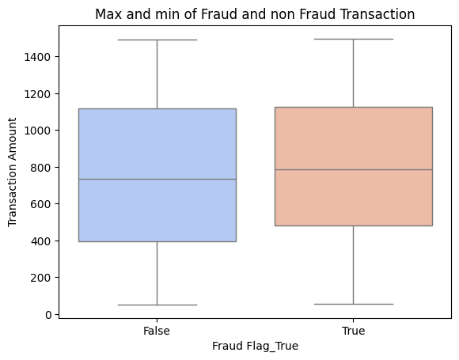
* The transactions that are fraud are more likely successful than non fraud transaction.
* However failed fraud transactions are also more than failed non fraud transactions

**Relation between fraud and Transaction type**

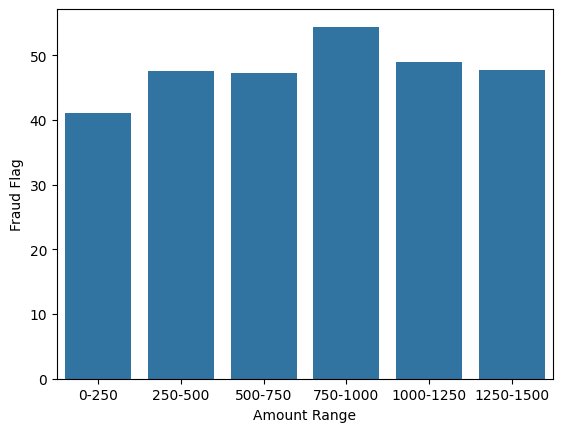
* Transactions of type deposit are more fraudulent.
* Transfer is generally likely to be non-fraudulent.
* However withdraw has the least amount of fraud transactions.

**Location and fraud transaction relation**

This shows that there are some locations that produce a greater number of fraudulent transactions. Some locations reaching up to 20 fraud transactions also.

**Boxplot of fraud and non-fraud Transactions**

This boxplot shows the fraudulent transactions have higher average than non- fraud transactions. But the gross sum of amount across types shows more capital lies in the non-fraud transactions however there is not much difference.

**Fraud and Transaction amount relation**

This bar graph shows the middle range of rs.750 to rs.1000 amount has the highest probability of fraud than any other range. The lower range transactions have smaller probability of being fraud.

**Conclusion**

**Fraud Analysis**:

* Fraudulent transactions are more likely to succeed.
* Deposits are the most fraudulent, while withdrawals have the least fraud.
* Certain locations have high fraud rates.
* Transactions between ₹750–₹1000 have the highest fraud probability.

Mobile transactions are more common but prone to failures, while deposits have the highest fraud rates. Fraudulent transactions are often successful, especially in certain locations and within the ₹750–₹1000 range. Overall, transfers dominate in volume, and deposits show the highest single transaction amounts.

To decrease fraud and improve transaction security, several measures can be implemented. Improving fraud detection with machine learning models can help identify high-risk transactions, especially in the ₹750–₹1000 range. Enhancing security for mobile transactions, such as multi-factor authentication, can reduce failures and fraud. Monitoring high-risk locations and transaction types, like deposits, can help detect suspicious activity early. Additionally, real-time alerts for unusual transaction patterns can improve overall security and prevent fraudulent activities.