BANK TRANSACTION FRAUD DETECTION USING GRAPH ANALYTICS

DATASET DESCRIPTION:

```
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   gT              54222 non-null  object
 1   sId             54222 non-null  object
 2   rId             54222 non-null  object
 3   sAcc            54222 non-null  object
 4   rAcc            54222 non-null  object
 5   TranAmount      54222 non-null  float64
 6   TranType        54222 non-null  object
 7   TranStatus      54222 non-null  object
 8   sBalbefore      54222 non-null  float64
 9   sBalAfter       54222 non-null  float64
 10  rBalBefore      54222 non-null  float64
 11  rBalAfter       54222 non-null  float64
 12  sf1             54222 non-null  bool
 13  sf2             54222 non-null  bool
 14  sf3             0 non-null      float64
 15  sf4             0 non-null      float64
 16  sTD             54222 non-null  object
 17  rTD             54222 non-null  object
 18  sAccID          54222 non-null  object
 19  ef1             0 non-null      float64
 20  ef2             0 non-null      float64
 21  NoDescription   54222 non-null  object
 22  TranTS          54222 non-null  object
 23  sType           54222 non-null  object
 24  rType           54222 non-null  object
dtypes: bool(2), float64(9), object(14)
memory usage: 9.6+ MB
```

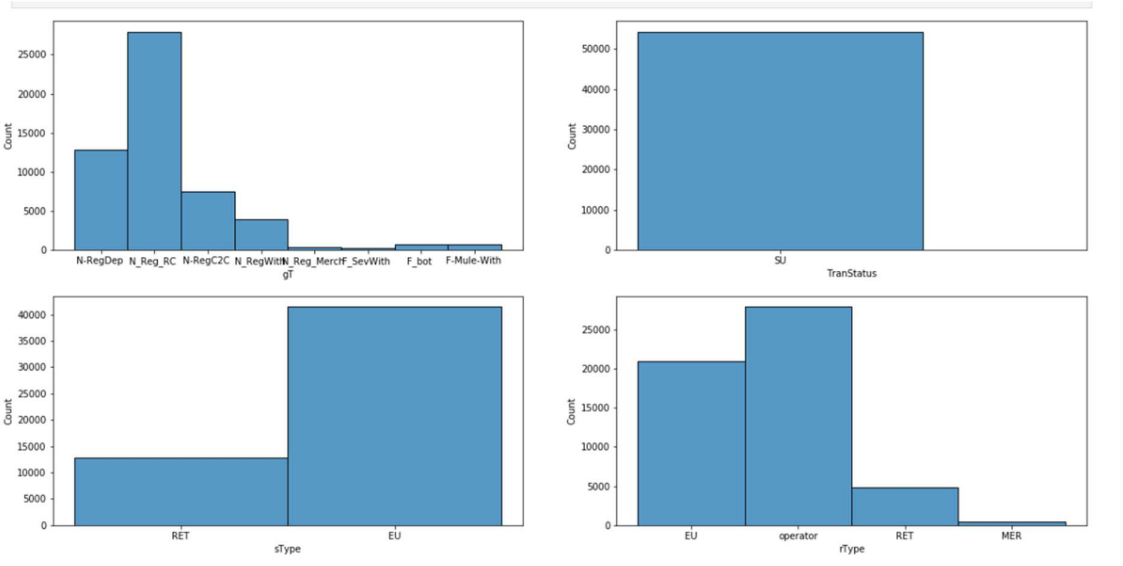
Column	Description
gT	General or global transaction identifier or group, potentially categorizing the transaction.
sId	Sender's unique identifier.
rId	Recipient's unique identifier.
sAcc	Source account number or identifier from which funds are transferred.
rAcc	Recipient account number or identifier to which funds are sent.
TranAmount	Amount of money involved in the transaction.
TranType	Type of transaction (e.g., deposit, withdrawal, transfer).
TranStatus	Status of the transaction (e.g., completed, pending, failed).

sBalBefore	Sender's balance before the transaction.
sBalAfter	Sender's balance after the transaction.
rBalBefore	Recipient's balance before the transaction.
rBalAfter	Recipient's balance after the transaction.
sf1	A flag or special condition for the transaction (likely boolean).
sf2	Another flag or special condition for the transaction (likely boolean).
sf3	A third flag or special condition (likely boolean).
sf4	A fourth flag or special condition (likely boolean).
sTD	Sender's transaction detail or timestamp.
rTD	Recipient's transaction detail or timestamp.
sAccID	Identifier for the source account, likely for distinguishing different accounts.
ef1	External factor or error flag related to the transaction (numerical value).
ef2	Another external factor or error flag (numerical value).
NoDescription	Placeholder for transactions without a specific description.
TranTS	Timestamp when the transaction occurred.
sType	Type or category of the sender's account (e.g., individual, corporate).
rType	Type or category of the recipient's account (e.g., individual, corporate).

AFTER DROPPING THE NULL VALUES

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54222 entries, 0 to 54221
Data columns (total 19 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   gT          54222 non-null  object
 1   sId         54222 non-null  object
 2   rId         54222 non-null  object
 3   sAcc        54222 non-null  object
 4   rAcc        54222 non-null  object
 5   TranAmount  54222 non-null  float64
 6   TranType    54222 non-null  object
 7   TranStatus  54222 non-null  object
 8   sBalbefore  54222 non-null  float64
 9   sBalAfter   54222 non-null  float64
 10  rBalBefore  54222 non-null  float64
 11  rBalAfter   54222 non-null  float64
 12  sf1         54222 non-null  bool
 13  sf2         54222 non-null  bool
 14  sTD         54222 non-null  object
 15  rTD         54222 non-null  object
 16  TranTS      54222 non-null  object
 17  sType       54222 non-null  object
 18  rType       54222 non-null  object
dtypes: bool(2), float64(5), object(12)
memory usage: 7.1+ MB
```

VISUALISING THE CATEGORICAL VARIABLES IN THE DATASET



LOOKING FOR DUPLICATE VALUES IN THE DATASET AND CHECKING RELATIVE SIZE FACTORS >2 AND DIGGING IN FURTHER

```
--- Duplicates Analysis ---
Number of duplicate rows found: 0

--- Relative Size Factor (RSF) Analysis ---
Number of senders with a high Relative Size Factor (RSF > 2): 33
Senders with the highest RSF values:
                sId  largest_amount  second_largest_amount       RSF
272    PN_EU_0_1251         5692.42                1222.28  4.657214
707     PN_EU_0_617        23824.30                6221.89  3.829110
583     PN_EU_0_502         1968.88                 518.82  3.794919
1507    PN_EU_1_445        17010.86                4767.95  3.567751
188    PN_EU_0_1173         9045.19                2912.36  3.105794
```

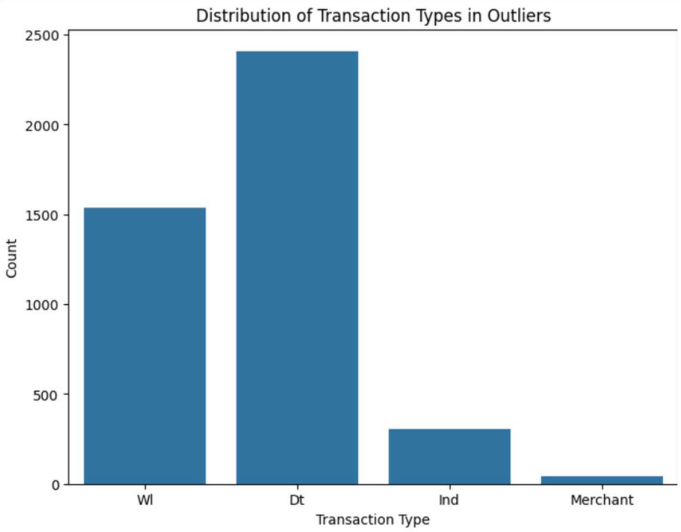
DETERMINING THE MAXIUM AND MINIMUM MOST POSSIBLE PERMISIBLE AMOUNT THROUGH IQR

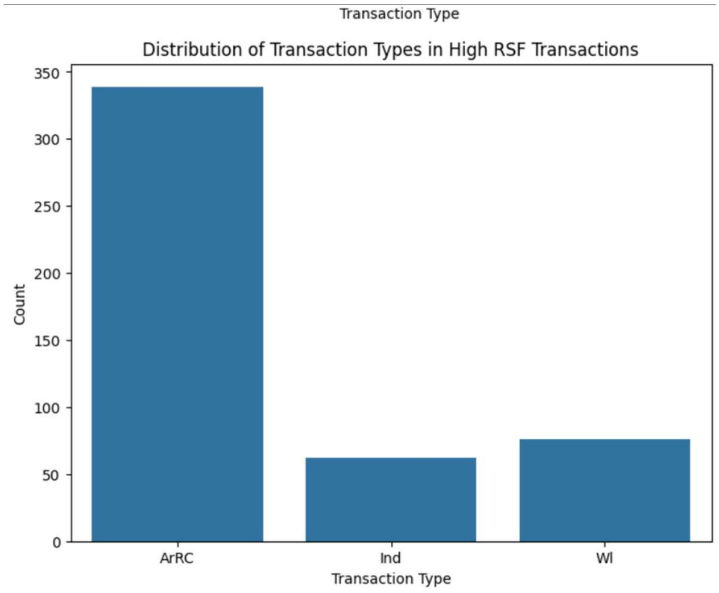
```
--- Outlier Detection on TranAmount ---
Number of outliers detected in TranAmount: 4289
Sample of detected outliers:
            gT           sId          rId        sAcc        rAcc  TranAmount
8   N_RegWith    PN_EU_1_131      PN_Ret6  EUAcc1_131       RAcc6   314643.58
10   N-RegDep        PN_Ret4  PN_EU_1_497       RAcc4  EUAcc1_497   232019.51
12   N-RegDep        PN_Ret5  PN_EU_0_143       RAcc5  EUAcc0_143   228085.86
47   N-RegDep        PN_Ret5  PN_EU_2_136       RAcc5  EUAcc2_136   275794.00
57   N-RegDep        PN_Ret3   PN_EU_0_15       RAcc3   EUAcc0_15   190012.85
```

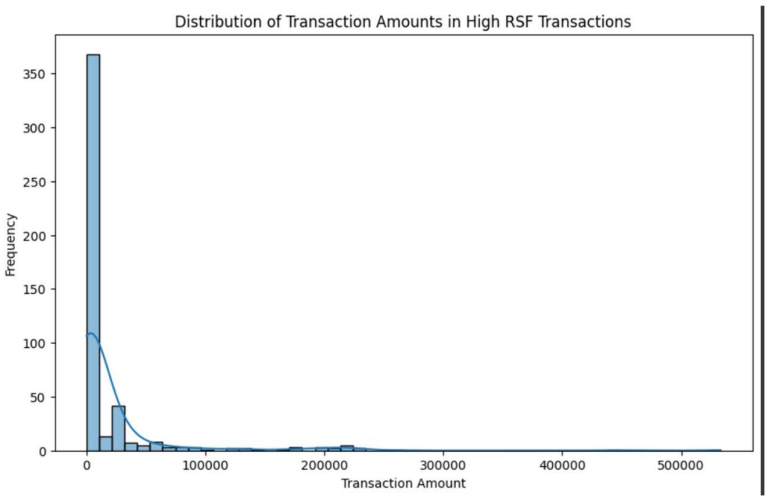
DETERMING THE EXACT OF FIGURES OF EACH CATEGORIES

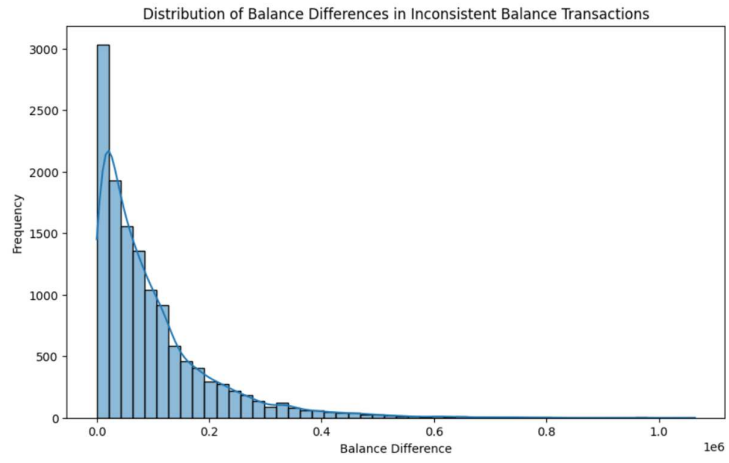
```
Shape of df_outliers: (4289, 27)
Shape of df_high_rsf: (477, 27)
Shape of df_inconsistent_balances: (13094, 27)
```

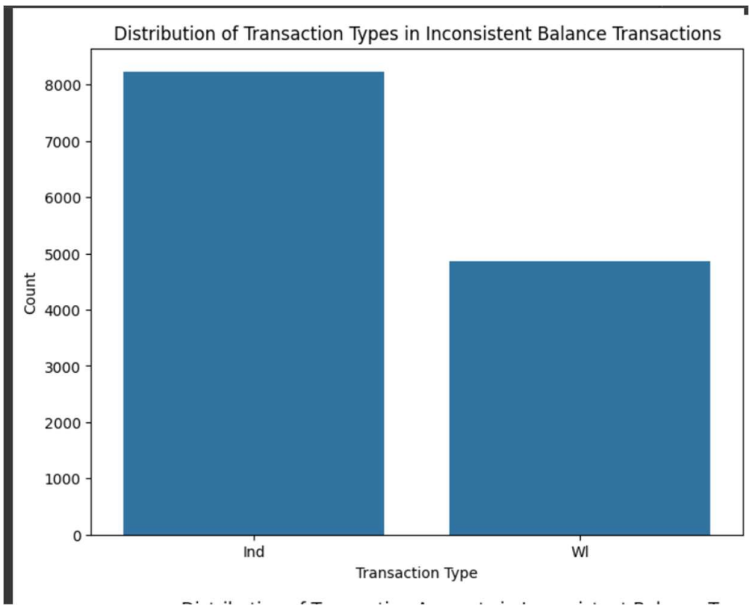
VISUALISING TO ANALYSE THE OUTLIERS

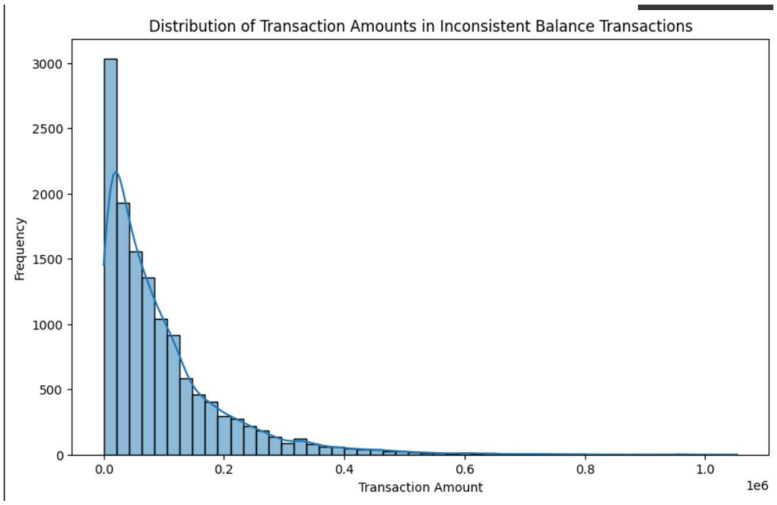



Transaction Type
Distribution of Transaction Types in High RSF Transactions
Count
0
50
100
150
200
250
300
350
ArRC
Ind
WI
Transaction Type


Distribution of Transaction Amounts in High RSF Transactions
Frequency
0
50
100
150
200
250
300
350
0
100000
200000
300000
400000
500000
Transaction Amount


Distribution of Balance Differences in Inconsistent Balance Transactions
Frequency
0
500
1000
1500
2000
2500
3000
0.0
0.2
0.4
0.6
0.8
1.0
1e6
Balance Difference





Summary of Flagged Transaction Characteristics

Outlier Transactions (based on TranAmount):

- Primarily characterized by unusually high transaction amounts compared to the rest of the dataset, as identified by the IQR method.

- The dominant transaction types are 'Dt' and 'WI'.

- Sender types are mainly 'RET' and 'EU', while receiver types are primarily 'EU' and 'RET'.

High Relative Size Factor (RSF) Transactions:

- Identified by a large ratio between the largest and second-largest transaction amounts for a given sender, suggesting a significantly larger transaction event for that sender.

- The most frequent transaction type is 'ArRC'.

- Sender types are exclusively 'EU'.

- Receiver types are predominantly 'operator', 'RET', and 'EU'.

- **While flagged by RSF, the absolute transaction amounts are generally lower (Mean: 19114.41, Median: 2819.34) compared to transactions flagged as outliers based purely on amount.**

Inconsistent Balance Transactions:

- These transactions show a discrepancy between the transaction amount and the change in the sender's balance.

- The primary transaction types are 'Ind' and 'Wl'.

- Sender types are exclusively 'EU'.

- Receiver types are mainly 'EU' and 'RET'.

- These transactions involve a wide range of amounts (Mean TranAmount: 95503.20, Median TranAmount: 63674.45).

- **The 'balance_diff' (sBalbefore - sBalAfter) in this group has a mean of 96458.23 and a median of 64311.20, indicating the magnitude of the inconsistency.**

Comparative Analysis:

- Outliers based on amount tend to involve large sums regardless of sender/receiver types or transaction type distribution, although certain types are more frequent.

- High RSF transactions highlight senders with disproportionately large individual transactions relative to their history, often involving specific transaction types ('ArRC') and 'EU' senders interacting with 'operator' or 'RET' receivers. Their absolute amounts are lower than general outliers.

- Inconsistent balance transactions point to data integrity issues or potential manipulation. They involve a significant portion of the dataset and cover a broader range of transaction amounts than high RSF transactions, although generally lower than the most extreme outliers. The inconsistency is focused on 'EU' senders and 'EU' or 'RET' receivers, with 'Ind' and 'Wl' transaction types being most common.

ANALYSING THE POSSIBILITY OF SMURFING AND ROUNDTRIPS

filtering the graph based on nodes and the number of transactions if there is a minimum of 10 transaction even done in the account we consider it

get those nodes whose degree is higher, if no of transactions is high it can be a poosible fraud

```
nds=list(G2.nodes()) #list of all nodes
print(type(nds))
degrees = [val for (node, val) in G2.degree()]
l2=len(degrees)  # degree of respective nodes

repeated_nodes=[]
for i in range(l2):
    val=degrees[i]
    if val>=10:        # If it has done more than 10 transactions
        ind = nds[i]
        repeated_nodes.append(ind)
print(len(repeated_nodes))
```

```
<class 'list'>
206
```

NOW WE ONLY CONSIDER THE INDIVIDUAL AND WITHDRAWAL TYPE AND FILTER THE DATASET BASED ON THE TRANSACTION AMOUNT ,THAT IF THE TRANSACTION AMOUNT IS LESS THAN THE AVERAGE OF THE AMOUNT TRANSACTED WE HAVE VISUALISED IT

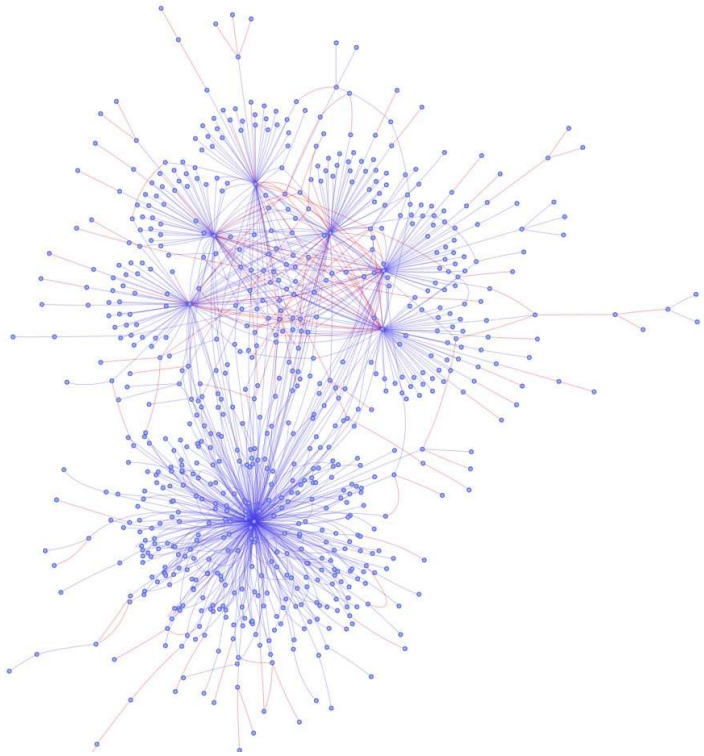
node of the graph = Account Id(Account holder)

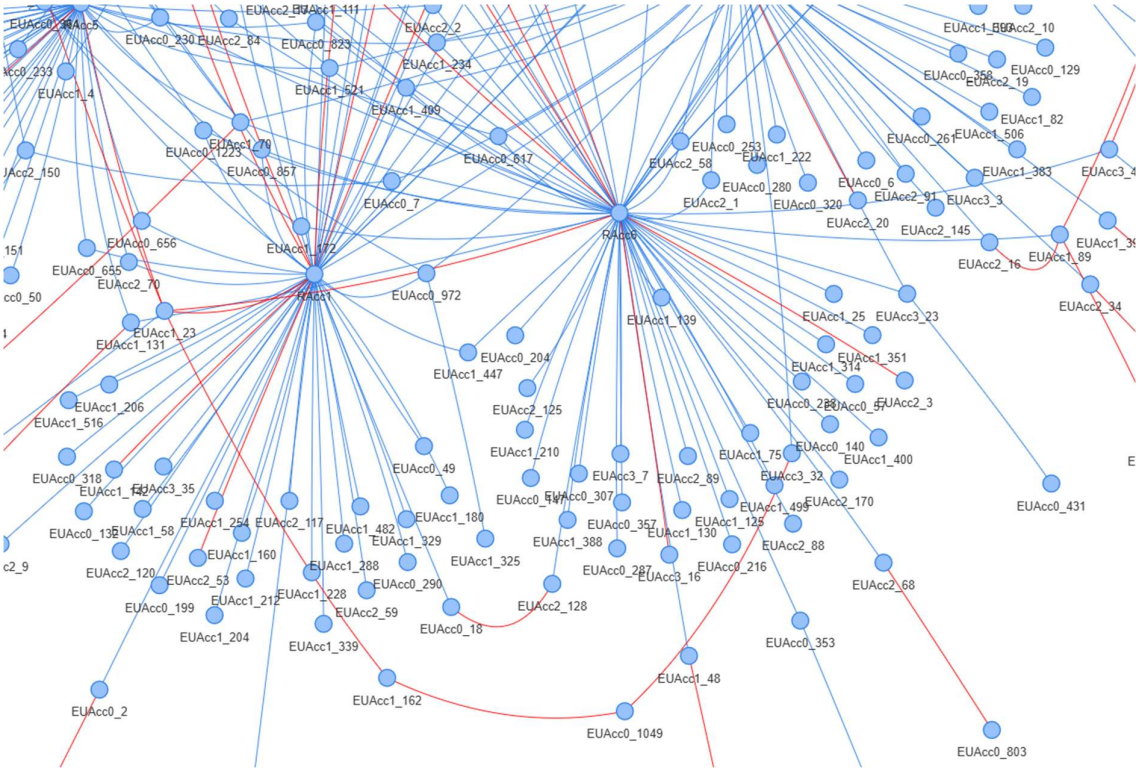
edge = A transaction.

edge length = Amount of transaction.

An edge connects the sender and receiver of a transaction.

Normal transactions are visualised in blue color and possible fraud transactions are visualized in red color in the final graph network.





Conclusion

Out of total 54222 transactions 3614 transactions are suspected to be fraud transactions.