ANOMALY DETECTION IN CREDIT CARD TRANSACTIONS

Overview: Anomaly detection in credit card transactions refers to the process of identifying unusual or fraudulent activities in credit card transactions. It involves applying statistical, machine learning and Power BI techniques to detect patterns and deviations from normal behaviour, helping to identify potential fraudulent transactions in real-time.

Objective: The objective of this project is to develop a Power BI dashboard for anomaly detection in credit card transactions. Anomaly detection is crucial for detecting fraudulent activities and ensuring the security of credit card transactions. By leveraging Power BI's data visualisation and analytical capabilities, we can create an interactive dashboard that provides insights into transaction patterns and identifies potential anomalies.

Data Source:

We will use a dataset that includes 1 tables namely Fraud. Key Columns are:

Step - maps a unit of time in the real world. In this case 1 step is 1 hour of time. Total steps 744 (30 days simulation).

Type - CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER.

Amount - amount of the transaction in local currency.

NameOrig - customer who started the transaction

OldbalanceOrg - initial balance before the transaction

NewbalanceOrig - new balance after the transaction

NameDest - customer who is the recipient of the transaction

OldbalanceDest - initial balance recipient before the transaction. Note that there is no information for customers that start with M (Merchants).

NewbalanceDest - new balance recipient after the transaction. Note that there is no information for customers that start with M (Merchants).

IsFraud - This is the transactions made by the fraudulent agents inside the simulation. In this specific dataset the fraudulent behaviour of the agents aims to profit by taking control of customers accounts and try to empty the funds by transferring to another account and then cashing out of the system.

Project Steps:

a. **Data Loading**
   Data is in .CSV Format. Imported the dataset into Power BI.

b. **Data Transformation**
   Performed data cleaning and transformation which includes removing null values, removing errors, removing duplicates, checking for data types.

c. **Data Modelling**
   There is only 1 Fact Tables - Fraud

d. **Data Analysis using DAX Functions:**

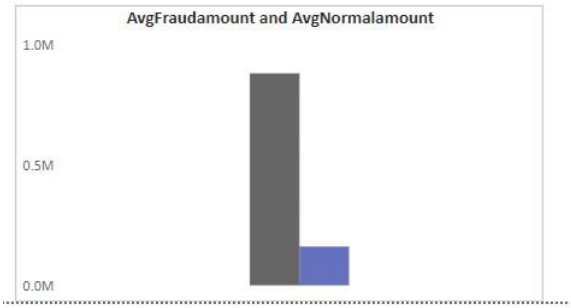
What is the average transaction amount for normal transactions versus fraudulent transactions?

```
Fraud_transaction_avg = CALCULATE(AVERAGE(Fraud[amount]),
                          FILTER(Fraud, Fraud[isFraud] = 1)
                        )
```

// This DAX function returns the Average amount for fraud transactions.

```
Normal_transactions_avg = CALCULATE(AVERAGE(Fraud[amount]),
                            FILTER(Fraud, Fraud[isFraud] = 0)
                            )
```

// This DAX function returns the Average amount for normal transactions.



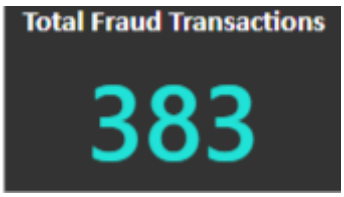
How many credit card transactions were recorded in the dataset? And **How many fraudulent credit card transactions were recorded in the dataset?**

```
Creditcardcount = COUNTA(Fraud[type])
```

//This DAX function returns the no. of total transactions recorded in a dataset.


```
Fraudcount = CALCULATE([Creditcardcount], FILTER(Fraud, Fraud[isFraud] = 1))
```

// This DAX function returns the fraud no. of transactions recorded in a dataset



What is the highest Fraud transaction amount recorded?

```
Highest_fraud_transaction = CALCULATE(MAX(Fraud[amount]),
                                FILTER(Fraud, Fraud[isFraud] = 1)
                                )
```

// This DAX function returns the highest fraud transaction amount recorded in a dataset

Is there a significant difference in the maximum transaction amount for normal transactions compared to fraudulent transactions?

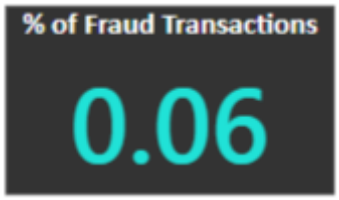
```
Highest_normal_transaction = CALCULATE(MAX(Fraud[amount]),
                                        FILTER(Fraud, Fraud[isFraud] = 0)
                             )
Highest_fraud_transaction = CALCULATE(MAX(Fraud[amount]),
                                FILTER(Fraud, Fraud[isFraud] = 1)
                                )
Diff = [Highest_fraud_transaction]-[Highest_normal_transaction]
```

// This DAX function returns the difference of transaction amount recorded in a dataset

What is the percentage of fraudulent transactions in the dataset?

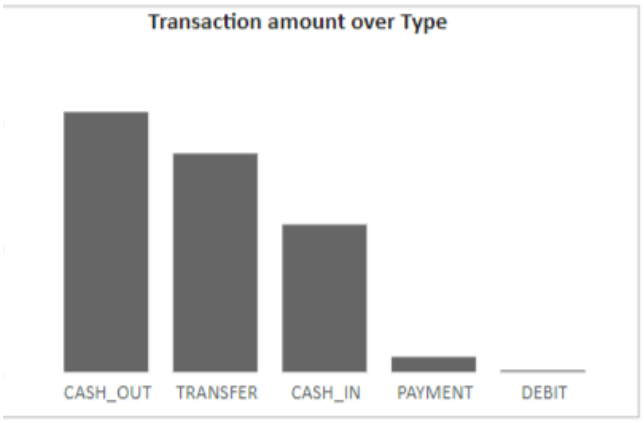
```
% of fraudalent transactions = DIVIDE([Fraudcount],[Creditcardcount])*100
```

// This DAX function returns the percentage of fraud transaction recorded in a dataset

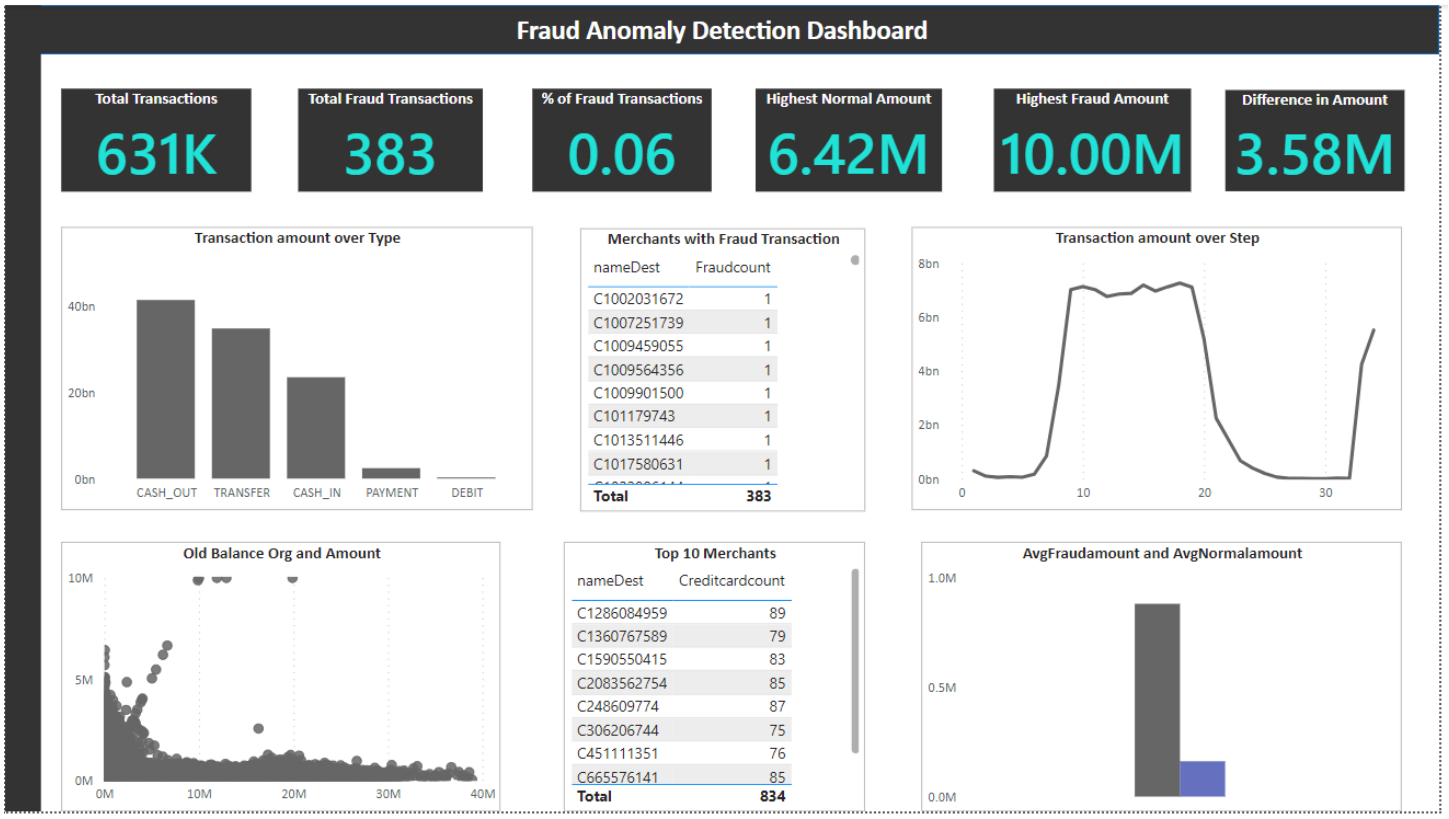


What is the distribution of transaction amounts? (using Clustered column chart)

This chart is a column chart between type of payment and amount. This chart suggests that the cash-out had the highest total payment amount, i.e., 41.415 billion and the debit had the lowest total payment amount, i.e., 29,18 million.



e. **Data Visualization**



- **Scatter plot between Old Balance Org and Amount -** This Scatter plot gives the relationship between the initial balance amount before the transaction (Old Balance Amount) and the transaction amount. This chart is also used to find outliers.
- **Line Chart between Transaction Amount by Step -** This chart gives the trend of the total transaction amount with steps. In this dataset, one step is equal to one hour. This chart suggests that the 18th step records the highest total transaction amount i.e. 7.25 billion.
- **Top 10 Merchants -** This table gives a list of top 10 merchants. The highest number of transactions recorded by a single merchant is 95.

- **Merchants with Fraud Transaction -** This table suggests that 383 merchants have committed fraud transactions.

f. **Insights**

- The transaction done by credit cards are analyzed and further discovered, there are numerous fraudulent in the transactions.
- The total recorded transactions are 630894 with 383 fraud transaction. The percentage of fraudulent transaction is 0.06%.
- The fraud transaction comprises 197cash_out transaction and 186 transfer type, It is found that these 2 transaction type has highest no. of fraud transactions, which means that proper safety is to be ensured for these two type of transactions.
- The max fraud amount was 10M and with 6 individuals each which is a very big number in itself.
- There are 3 different customers who have been targeted twice for the fraud with the max amount of 828720 at once.
- The total fraud amount is approximately 102 billion.
- The most used payment type is cash_ out type with 224013 transactions and approximately amount of 41 billion.
- The least used payment type is debit type with only 4769 transactions and approximately amount of 29 million.
- Steps between 15-20 consist of max amount of transactions with approximately 7.3 billion.

g. **Conclusion**

- The fraud transaction comprises 197 cash_out transaction and 186 transfer type, It is found that these 2 transaction type has fraud transactions, which means that proper safety is to be ensured for these two type of transactions. The max fraud amount was 10M and with 6 individuals each which is a very big number in itself. Data Privacy and security steps must be taken to prevent fraud transactions and ensure safe transactions.