Anomaly Detection in Credit Card Transactions using Power BI

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

Project Overview:
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.

Project Steps:
—----------

- **Dataset Info:**

  - ***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.

Imported the Dataset given and followed the steps to accomplish the task

- **Power Query:**

  Cleaned the data by removing duplicate rows, blank rows and errors.
  The dataset comprised 6,30,894 rows and 10 columns.

- **DAX Function:**

  Solved the following using DAX functions:

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

```
Average Normal Amount = CALCULATE(AVERAGE(Fraud[amount]),Fraud[isFraud]=0)
Average Fraud Amount = CALCULATE(AVERAGE(Fraud[amount]),Fraud[isFraud]=1)
```

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

```
Total Transactions = COUNTROWS('Fraud')
Fraudulent Transactions = COUNTROWS(FILTER('Fraud', 'Fraud'[isFraud] = 1))
```

  - What is the highest Fraud transaction amount recorded?

```
Max Fraud Amount = CALCULATE(MAX(Fraud[amount]),Fraud[isFraud]=1)
```

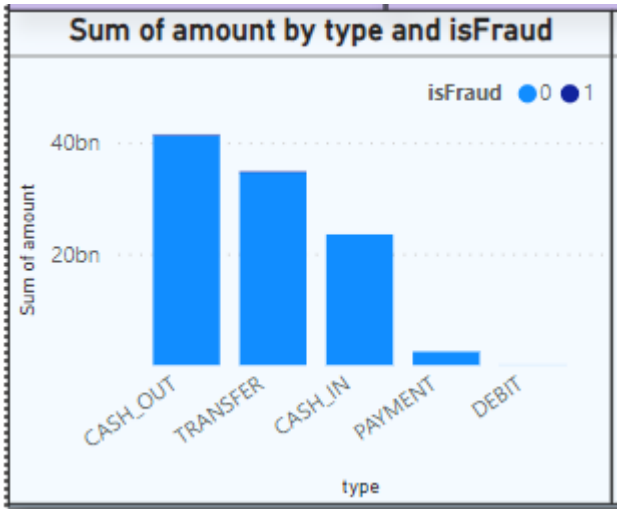
  - Is there a significant difference in the maximum transaction amount for normal transactions compared to fraudulent transactions?

```
Max Fraud Amount = CALCULATE(MAX(Fraud[amount]),Fraud[isFraud]=1)
Max Normal Amount = CALCULATE(MAX(Fraud[amount]),Fraud[isFraud]=0)
Max Diff Normal and Fraud = CALCULATE([Max Fraud Amount]-[Max Normal Amount])
```

  - What is the percentage of fraudulent transactions in the dataset?

```
Fraud Transactions Percentage = DIVIDE([Fraudulent Transactions], [Total
Transactions], 0)*100
```

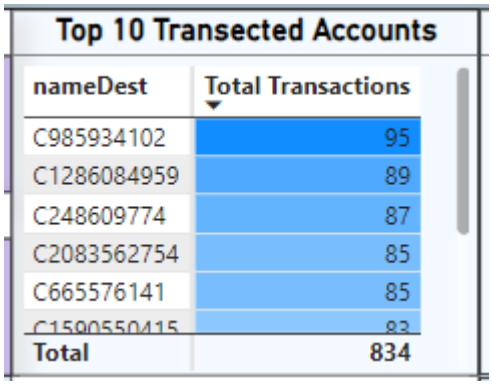
- What is the distribution of transaction amounts? (using Clustered column)



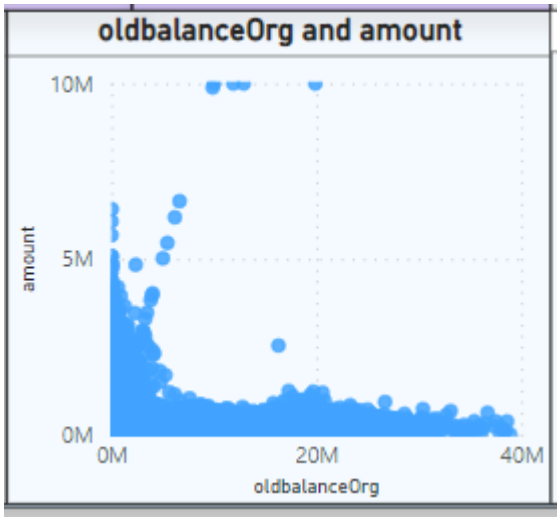
- **Anomaly Visualisation:**

Developed visualisations that highlight potential anomalies in the credit card transactions.
Used line charts, scatter plots, and tables to display transaction patterns and identify outliers.

- ***Merchants with highest number of transactions.(Only Top 10)***
  The tile comprises two columns(name, count of highest transaction) in a table. A table is used here so that the data is compared easily among different rows.

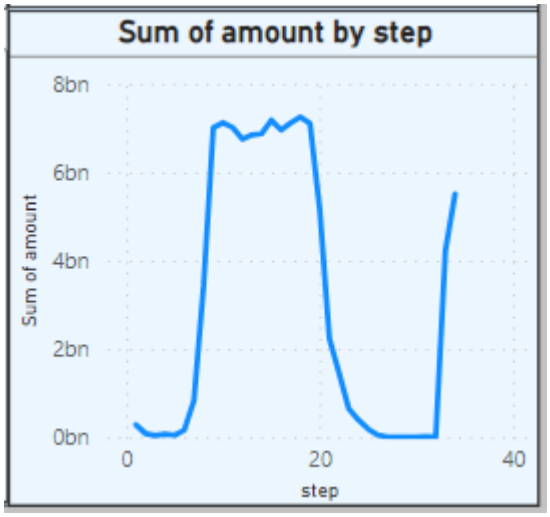


- ***Create a scatter plot to visualise the relationship between 'oldbalanceOrg' and 'amount' columns***.
  The x-axis has oldbalanceorg and y-axis has amount column. The chart reveals how both numerical values are related to each other. It also helps in finding the outliers.

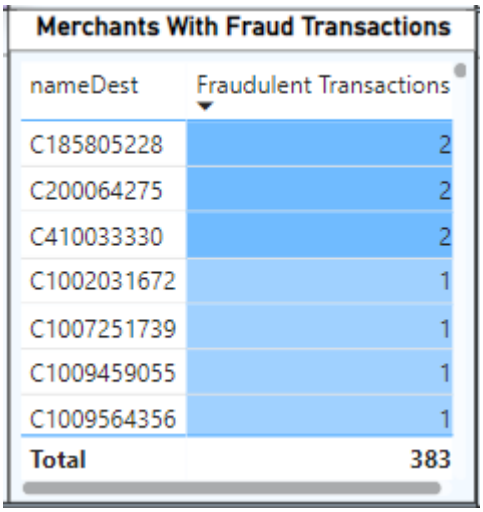


- ***Use a line chart to plot the transaction amount over time (step) to identify any unusual spikes or drops in transaction amounts***.

  A line chart represents series of data points for a span of time, over here it's showing the fluctuation in sum of amount with steps(hours).



Merchants with a high occurrence of fraudulent transactions.



Merchants With Fraud Transactions

nameDest	Fraudulent Transactions
C185805228	2
C200064275	2
C410033330	2
C1002031672	1
C1007251739	1
C1009459055	1
C1009564356	1
Total	**383**

- **Importance of using above charts for visualization.**

1. **Diversity of Insights:** Different chart types offer diverse insights. Bar charts help compare counts, scatter plots reveal relationships, and line charts show temporal trends.

2. **Pattern Recognition:** Scatter plots and line charts are excellent for pattern recognition. Anomalies may be more easily identified when examining relationships and trends.

3. **Focus on Top Contributors:** Limiting visualizations to the top 10 merchants or focusing on fraudulent transactions helps prioritize and identify the most critical areas for investigation.

4. **User-Friendly Interpretation:** These visualizations are user-friendly, allowing stakeholders to easily grasp and interpret the data, which is crucial for timely decision-making in fraud detection.