Anomaly Detection in Credit Card Transactions

Context

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 behavior, helping to identify potential fraudulent transactions in real-time

Objectives:

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 visualization and analytical capabilities, we can create an interactive dashboard that provides insights into transaction patterns and identifies potential anomalies.

About Dataset:

- ***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.

Preliminary Data Analysis:

Extraction:

1. We did some preliminary analysis and found a sample dataset for fraudulent credit card transactions on Kaggle
2. The dataset was in csv format, the dataset was imported on power BI through **Get Data.**

Transformation:

1) For data cleaning tasks, we did not find any missing value in the data set. There were certain duplicates in old balance and new balance columns, but we did not remove them as two customers can have same balance.
2) The data types of all the columns were appropriate so no transformation was required. Similarly, since there was just one table no relationship modelling was performed. We changed the names of columns to make them easy to understand.

DAX Calculations:

- ***What is the average transaction amount for normal transactions versus fraudulent transactions?***
  - The Average Fraud Transactions and Normal transactions were plotted against payment type and the highest difference was in cash out payment type, that is $669,996.32.
- ***How many credit card transactions were recorded in the dataset? And how many fraudulent credit card transactions were recorded in the dataset?***
  - Total credit card transactions recorded were 683 k out of which 383 were using credit card.
- ***What is the highest Fraud transaction amount recorded?***
  - *The highest Fraud transaction amount recorded was 10 million.*
- ***Is there a significant difference in the maximum transaction amount for normal transactions compared to fraudulent transactions?***
  - *Yes, this is visualized using a column chart which shows a total difference of 3.58 million.*
- ***What is the percentage of fraudulent transactions in the dataset?***
  - *The percentage fraudulent transaction recorded was 0.06%.*
- ***What is the distribution of transaction amounts? (Using Clustered column chart)***
  - A clustered column chart showcases the distribution of transaction amounts, enabling users to identify trends and anomalies at a glance.

Anomaly Visualization:

Visualizations were developed using line charts and scatter plots, to display transaction patterns and identify outliers, that highlight potential anomalies in the credit card transactions.

- ***Which merchants have the highest number of transactions? (Only Top 10)***
  - *This was estimated using a table chart, which shows that all the Top 10 merchants have 2 highest transactions.*
- ***Create a scatter plot to visualize the relationship between 'oldbalanceOrg' and 'amount' columns.***
  - *The outliers in the scatter plot represent the fraud transactions with amount around 10 million.*
- ***Use a line chart to plot the transaction amount over time (step) to identify any unusual spikes or drops in transaction amounts.***
  - Line chart plot depicts that highest transaction were carried out at step 19 followed by 18 and 15. There was drop observed after 19 step and again gradual increase was observed from step 32***.***
- ***Are there any merchants with a high occurrence of fraudulent transactions.***
  - *No such high occurrence of fraudulent transactions was observed in accordance with merchants as the fraud transactions were carried out by taking control of customer accounts.*

Conclusion:

In this Power BI project, we effectively showcased the detection of credit card transaction anomalies. We processed the data by extracting and transforming it using Power BI Desktop and Power Query Editor, applying DAX transformations. Through stacked column charts, line charts, bar graphs, and scatter plots, we identified anomalies in payment types, particularly focusing on fraud. Our analysis revealed that fraudulent transactions were most associated with transfer and cash-out payment modes, emphasizing the need for caution. Normal transactions ranged from $6,117.85 to $674,086.23, while fraudulent ones ranged from $854,286.86 to $910,504.14. Step 19 had the highest transaction count at 51,352 instances. The real-time data updating feature in the dashboard allows users to promptly respond to anomalies. In conclusion, this analysis highlights the importance of customers avoiding sharing OTPs and passwords to prevent fraud, while advocating for transparent transaction practices. Banks should enhance security measures for both customers and merchants to combat fraudulent activities effectively.