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, machine learning 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 customer accounts and try to empty the funds by transferring to another account and then cashing out of the system.

- ### Power BI Dashboard Creation:

  1. Launch Power BI and connect to the preprocessed credit card transaction dataset.
  2. Design an intuitive and visually appealing dashboard layout with appropriate charts, tables, and filters.
  3. Create visualisations that provide an overview of transaction statistics, such as total transactions, average transaction amount, and transaction frequency.

Preliminary data analysis steps:

- ### 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 Function:

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

Ans-Average Fraudulent Amount and Average Normal Amount diverged the most when the Type of Transaction was CASH_OUT, when Average Fraudulent Amount were 6,69,996.32 higher than Average Normal Amount.

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

Ans- There are a total 6,30,894 credit card transactions were recorded in the dataset.

The total number of many fraudulent credit card transactions were recorded in the dataset are 383 and 6,30,511.

3. What is the highest Fraud transaction amount recorded?

Ans- The highest Fraud transaction amount recorded is 10M.

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

Ans- Yes, this is visualized with the help of a column chart which shows a total difference of 3.58 million.

5. What is the percentage of fraudulent transactions in the dataset?

Ans- The percentage of fraudulent transactions in the dataset is 0.06%.

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

Ans- With the help of chart we can find the distribution of transaction amounts, enabling users to identify trends and anomalies at a glance.

• Anomaly Visualisation:

Visualizations were developed by using line chart and scatter plot to display transaction patterns and identify outliers, that highlight potential anomalies in the credit card transaction.

• Which merchants have the highest number of transactions? (Only Top 10)

Ans-This was estimated by using a table chart which shows the list of top 10 merchants with highest number of transactions.

• Create a scatter plot to visualise the relationship between 'oldbalanceOrg' and 'amount' columns.

Ans- 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.

Ans-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?

Ans-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 account

• Conclusion:

Anomaly detection in credit card transactions was effectively demonstrated using Power BI. The dataset underwent extraction, transformation, and DAX transformations within Power BI Desktop and Power Query Editor. Visualizations including stacked column charts, line charts, bar graphs, and scatter plots were employed to discern anomalies in transactions, differentiating between normal and fraudulent transactions and payment types. Notably, transfer and cash-out payment modes were frequently linked to fraudulent activities. Across all payment types, normal transactions ranged from 6,117.85 to 674,086.23, while fraudulent transactions ranged from 854,286.86 to 910,504.14. The highest transaction count occurred at step 19, with 51,352 transactions. The real-time dashboard underscores the urgency for customers to safeguard their OTPs and passwords and for banks to enhance security measures for both customers and merchants.