Anomaly Detection in Credit Card Transactions using Power BI

#### 

Objective:

To develop a Power BI dashboard to analyze credit card transactions, detect anomalies, and visualize transaction patterns.

Data Preprocessing:

1. Data Cleaning:

* Handling Missing Values: Power Query Editor can be utilized to remove, replace, or impute missing values using functions like "Replace Values" or "Fill Down".
* Removing Duplicates: The "Remove Duplicates" function in Power Query Editor can be used to eliminate duplicate rows based on selected columns.
* Ensuring Data Integrity: Data validation rules and constraints within Power Query Editor or Power BI Data Model can be applied to ensure data consistency and accuracy, such as defining data types, setting relationships, and creating calculated columns to validate data integrity.

2. Data Modeling:

In this particular data model there exists only single table which consist of the following fields:  
step, type, Amount, nameOrig, oldbalanceOrg, newbalanceOrig , nameDest, oldbalanceDest, newbalanceDest, isFraud.

DAX Function:

#### 1. Row Context and Filter Context: Row Context: It refers to the current row being evaluated in a calculation in DAX, processed row by row. Filter Context: It refers to the set of filters applied to the data affecting DAX expression calculations.

#### DAX functions related to filtering: - FILTER - CALCULATE

2. DAX Functions:

* Calculated average transaction amount for normal and fraudulent transactions.
* Counted total credit card and fraudulent credit card transactions.
* Determined and compared the maximum transaction amounts for normal and fraudulent transactions.
* Calculated the percentage of fraudulent transactions.

#### DAX Formulae Explanation:

#### DAX Function 1: Average Normal vs. Fraudulent Transaction Amount

Average Normal Transaction = CALCULATE(AVERAGE(Fraud[amount]),Fraud[isFraud] = 0)

* *Average Normal Transaction was found to be 161.5K.*

Average Fraudulent Transaction = CALCULATE(AVERAGE(Fraud[amount]),Fraud[isFraud] = 1)

* *Average Fraudulent Transaction was found to be 881.6K.*

#### DAX Function 2: Count of Total Credit Card and Fraudulent Credit Card Transactions

#### Total Credit Card Transactions = COUNTROWS(ALL(Fraud))

* *Total Credit Card Transactions were recorded 630.9K.*

Total Fraudulent Credit Card Transactions = CALCULATE(COUNTROWS(Fraud),Fraud[isFraud] = 1)

* *Out of all credit card transactions, 383 were recorded Fraudulent*

#### DAX Function 3: Highest Fraud and Normal Transaction Amount

Highest Fraud Transaction Amount = CALCULATE(MAX(Fraud[amount]),Fraud[isFraud] = 1)

* *10.0 million was the Highest Fraud Transaction Amount.*

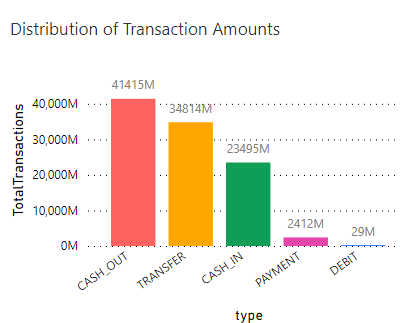
Highest Normal Transaction Amount = MAXX(FILTER(Fraud,Fraud[isFraud] = 0),Fraud[amount])

* *6.4 million was the Highest Normal Transaction Amount.*

#### DAX Function 4: Percentage of Fraudulent Transactions

Percentage Fraudulent Transaction = DIVIDE([Total Fraudulent Credit Card Transactions],[Total Credit Card Transactions])

* *There were 0.06% Percent of Fraudulent Transactions.*



|  |
| --- |
|  |
|  |

These DAX formulae provide calculations for the specified questions in the Power BI dashboard, enabling the generation of insightful visualizations and metric.

Anomaly Visualizations:

1. Field Parameters and Slicers:  
Field Parameter: It allow dynamic changing of analyzed fields in Power BI visuals or calculations.

Usage: Useful for comparing metrics, analyzing different dimensions, or customizing visualizations based on user preferences.

In the current project, field parameters can be created with the usages of calculated measures like average of normal transaction amount, average of fraudulent transactions amount and count of transactions based on each type of transaction.

Slicer: It’s an interactive filtering tool in Power BI for narrowing down data displayed in visuals.

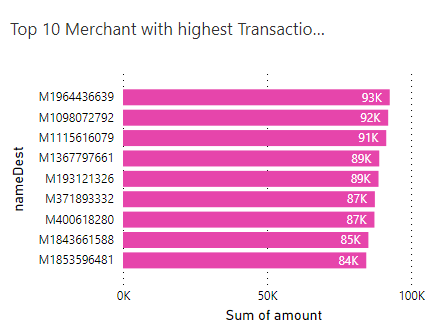
Usage: Provides users control over filtering data in reports, focusing on specific subsets dynamically.

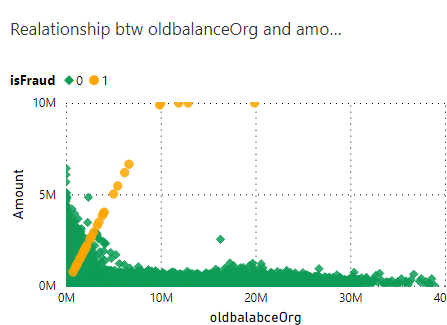
In the current project, slicers have been added from Visualizations pane of ‘step’ and ‘IsFraud’ fields for customized appearance and behavior using Power BI formatting options.

2. Anomaly Visualizations:

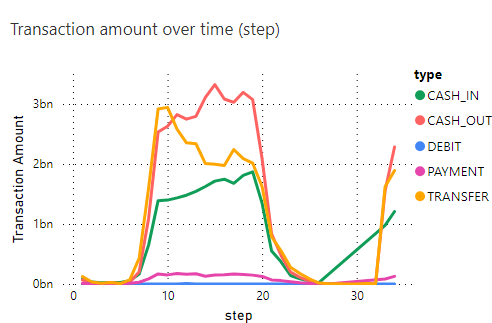
Implemented visualizations (clustered bar charts, line chart, scatter plot, clustered column chart, slicers) to highlight potential anomalies and outliers in credit card transactions.

#### Key Insights:

1. Merchant Transactions:
   1. Setup:
      * Fields: 'dest' (merchant names), Sum of amounts (total transaction amounts).
      * Advanced Filtering: Name starts with 'm', Top 10.
   2. Explanation:
      * Field: 'dest' for merchant names, Sum of amounts for total transaction amounts.
      * Filter: Focus on merchants starting with 'm', Top 10 selection.
   3. Interpretation:
      * Concise view of top-contributing merchants.
      * Useful for quick comparison.
      * Enables exploring transaction patterns for specific merchants.
      1. 
2. Scatter Plot (oldbalanceOrg vs. amount):
   1. Purpose:
      * Explore relationship between 'oldbalanceOrg' and 'amount'.
   2. Why:
      * Visualize distribution and relationship between two continuous variables.
      * Identify patterns, clusters, and outliers.
   3. Interpretation:
      * Clustering of points indicates common transaction behaviors.
      * Outliers or patterns may suggest fraudulent activities.



1. Line Chart (Transaction Amount Over Time):
   1. Purpose:
      * Illustrate trends in transaction amounts over time.
   2. Why this visual:
      * Suitable for visualizing trends over time.
      * Identify unusual spikes or drops in transaction amounts.
   3. Interpretation:
      * Sudden spikes or drops may represent abnormal transaction behavior
      * Consistent trends indicate regular transaction patterns.



1. Table for Merchants with High Occurrence of Fraudulent Transactions:
   1. Setup:
      * Fields: 'dest' (merchant names), Sum of amounts (total transaction amounts), 'isFraud' (fraud indicator).
      * Advanced Filtering: Name starts with 'm', 'isFraud' equals 1.
   2. Explanation:
      * Field: 'dest' for merchant names, Sum of amounts for total transaction amounts, 'isFraud' for indicating fraudulent transactions.
      * Filter: Conditions set for merchants starting with 'm' and 'isFraud' equal to 1.
   3. Interpretation:
      * No output suggests no 'm'-starting merchants have high fraudulent transaction occurrences.
      * Next Steps: Check data for 'm'-starting merchants with valid 'isFraud' values, adjust filtering if needed.

Conclusion:

* Upon analyzing credit card transactions, a significant number of fraudulent activities have been uncovered within the dataset.
* The total recorded transactions are 6,30,894 with 383 fraud transactions, which is 0.06% of the total transactions.
* Transactions were categorized into five types: cash in, cash out, debit, payment, and transfer. Notably, fraudulent transactions were detected primarily within the 'transfer' and 'cash out' categories, with approximate values of $169M and $168M, respectively (Overall total fraud amount is $337.65M). This highlights the need for increased attention to ensure the safety of these transaction types.
* The 'cash out' type is the most frequently used, with 224,013 transactions totalling around $41 billion (40.54% of Total Transactions). Conversely, the 'debit' type is the least utilized, with only 4,769 transactions amounting to approximately $29M.
* The line chart reveals a concentration of transactions within the range of 10 to 20 steps, accounting for the highest total transaction volume, approximately $7.3 billion. Of this amount, a significant portion, totalling $3.3 billion, falls into the 'cash out' category, representing a notably high-value subset.
* Among top 10 merchants with highest number of transactions, 'M1964436639', 'M1098072792' and 'M1115616079' stands out with the highest top 3 transaction amount of $93K, $92K and $91K respectively.

In conclusion, the analysis emphasizes the importance of staying alert and proactive in preventing fraud, especially in 'transfer' and 'cash out' transactions. It also underscores how understanding the distribution of transaction volumes, spotting repeat offenders, and having strong fraud detection systems are vital in protecting financial transactions.