**Anomaly Detection in Credit Card Transactions**

> OBJECTIVE:-

The objective of this project is to implement anomaly detection techniques to identify potentially fraudulent credit card transactions.

> WHAT IS ANOMALY DETECTION:-

Anomaly detection is detecting rare occurrences that seems suspicious because they’re different from the established pattern of behaviour.

> PROJECT DETAILS:-

Project contains a dataset where there is

1. “Step” column in which it maps unit of time in the real world. In this case 1 step is equal to 1 hour and there are total 744 steps.
2. “Type” column where there are 5 different payment methods which are CASH-IN, CASH-OUT, DEBIT, PAYMENT AND TRANSFER.
3. “Amount” column where there is amount of transactions in the local currency.
4. “Nameorg” column in which there is the customer who started the transition.
5. “Oldbalanceorg” column in which there is the initial balance before the transaction has been made.
6. “Newbalanceorig” column in which there is the new balance after the transactions have been made.
7. “namedest” column in which there is the customer who is the recipient of the transaction.
8. “OldbalanceDest” column in which there is the initial balance recipient before the transaction has been made. And there is no information of the customers that starts with ‘M’ (Merchants).
9. “newbalanceDest” column in which there is the new balance recipient after the transactions has been made. And there is no information of the customers that starts with ‘M’ (Merchants).
10. “isFraud” column in which there is the transactions made by the fraudulent agents inside the simulation. The number as 1 is the fraud transition and the number as 0 is the normal transaction.

> PREPROCESSING STEPS :-

* By utilising power query editor the process of data cleaning has been started under that the data type for each column has been changed as appropriate. For handing missing values all the rows has been deleted which contains null values.
* PERFORMING ALL THE BELOW QUESTIONS USING DAX FUNCTION
* What is the average transaction amount for normal transactions versus fraudulent transactions?
* To calculate this transactions I have created a new measure for normal transaction using AVERAGEX(FILTER(Fraud,Fraud[isFraud]=0),Fraud[amount]) where the value came as 161.50k as the average normal transactions.
* And to calculate the fraudulent transactions I have created a new measure using AVERAGEX(FILTER(Fraud,Fraud[isFraud]=1),Fraud[amount]) where the value came as 881.59k as the average fraudulent transactions.
* How many credit card transactions were recorded in the dataset? And How many fraudulent credit card transactions were recorded in the dataset?
* To calculate how many credit card transactions were recorded in the dataset I have created a measure using COUNT(Fraud[isFraud]) and the count of credit card transactions value came out as 631K.
* To calculate How many fraudulent transactions were recorded in the data set I have created a Measure using CALCULATE(COUNT(Fraud[isFraud]),Fraud[isFraud]=1) And the count of fraudulent transactions value came out as 383.
* What is the highest fraud transaction amount recorded?
* To calculate the highest fraud transactions amount recorded I have created a new measure using CALCULATE(MAX(Fraud[amount]),Fraud[isFraud]=1)

And the highest fraud transaction came as 10 million.

* Is there a significant difference in the maximum transaction amount for normal transactions compared to fraudulent transactions?
* To calculate the significant difference in the maximum transaction amount for normal transaction as compared to fraudulent transactions I have created a new measure using CALCULATE(MAX(Fraud[amount]),Fraud[isFraud]=0) and the highest normal transaction came as 6.42 million.

So as comparing the Significant difference between the maximum transaction amount of normal transaction and the fraudulent transaction the fraudulent transaction has the highest amount of transaction as compared to the normal transaction.

* What is the percentage of fraudulent transactions in the dataset?
* To calculate the percentage of fraudulent transactions in the data set I have created a new measure using CONCATENATE(ROUND(DIVIDE([Countoffraudtrans],[Countofcreditcard])\*100,2),"%") and The percentage of fraudulent transaction came as 0.06%.
* What is the distribution of transaction amounts? (using Clustered column chart)
* So using the clustered column chart, I have put the type column in the x-axis and sum of amount in the y-axis, where the Cash-out value came as 41.42 billion, Transfer value as 38.81 billion, Cash-in value as 23.5 billion, Payment value as 2.41 billion and Debit value as 0.03 billion.
* Creating Dashboard and develop visualisations that highlight potential anomalies in the credit card transactions.
* Which merchants have the highest number of transactions? (Only Top 10)
* To know which merchants have the highest number of transactions I have created a normal table in which added namedest and count of credit card in columns and filtered the top 10 merchants in the filter session.
* Create a scatter plot to visualise the relationship between 'oldbalanceOrg' and 'amount' columns.
* To visualise the relationship between oldbalanceorg and amount columns I have created a scatter plot in which added oldbalanceorg in X-axis, Amount in the Y-axis and isfraud in the Legend.
* Use a line chart to plot the transaction amount over time (step) to identify any unusual spikes or drops in transaction amounts.
* Created a line chart and added Step in the X-axis and sum of amount in the Y-axis. Where we can see the transactions amount overtime and identify any unusual spikes or drops in transaction amount.
* Are there any merchants with a high occurrence of fraudulent transactions?
* There are no merchants with high occurrence of fraudulent transactions.

> USING FIELD PARAMETER IN THE PROJECT.

* By using field parameter I have analysed total count of credit card transactions, count of fraudulent transactions, Normal amount transactions and fraud transactions.
* By analyzing this I have identified the usage that we can analyze the different fields in a single chart.

> USING SLICER IN THE PROJECT.

* We can filter out different categories based on different requirements.
* The data can be analyze with the various scenarios using different slicer.
* We can also see the difference in the fraud transactions and the normal transactions by using slicer.

> Summarization of the project in the following points.

* Problem statement :-

The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be a fraud. This model is then used to identify whether a new transaction is fraudulent or not.

* Identifying Factors, KPI:-

I have identified that fraud transactions are mainly in the transferred and cash-out type and the other transactions are more secured.

* Your insights after data analysis and dashboarding :-

- fraud transactions are in transferred and cash oot type and the other types are fraudulent free transactions and comparing the transferred and cashout type the transferred amount is greater.

- The highest amount of transactions in the types is of cash of types in the distributions of the transactions amount.

- Comparing the Normal transaction amount and the fraud transactions amount the fraud transaction is greater.

- I have found the top 10 merchants where all the merchants having 2 credit cards.

- the percentage of the fraud transactions are 0.06% where the percentage of normal transactions are 99.94% which is way too more then the fraud transactions but the amount of fraud transactions is greater as compared to normal transactions.

* What’s the conclusion about the problem statement :-

I have observed that the fraud transactions are increasing and it looks like the unexpected pattern or an outlier. The amount of the fraud transactions are more then the normal transactions.

Where the normal transactions count is greater and are as expected.