**SAMPLE PROJECT 5**

**FRAUD DETECTION SYSTEM**

**BY**

**GOTLA SHIVANI YADAV**

**AIM**

The aim of the product is to determine

* **Customer Demographics and Fraud:** the number of frauds that occur in a period of time as per customer demographics such as customer age, customer location are more prone to fraud.
* **Merchant Categories and Fraud:** Examine the merchant category column to identify if specific types of merchants are more frequently targeted by fraud.
* **Trends Over Time:** Use the transaction date column to identify trends in fraud incidents over time. This can help in understanding if certain periods have higher fraud rates.
* **Transaction Amount and Fraud:** Analyze the transaction amount column to see if higher-value transactions are more likely to be fraudulent.

**FEATURES**

**Transaction ID**: A unique identifier assigned to each transaction. This ensures every transaction can be individually tracked and referenced.

**Customer id:** A unique identifier for each customer involved in the transactions. It helps link multiple transactions to the same customer.

**Transaction date:** The date on which the transaction was conducted. This is essential for chronological analysis of transaction data.

**Transaction amount:** The total amount of money involved in the transaction. This helps in understanding the monetary value of each transaction.

**Merchant id:** A unique identifier for the merchant where the transaction took place. It allows tracking transactions back to specific merchants.

**Transaction type:** The category of the financial transaction, such as purchase, transfer, or withdrawal. This helps in categorizing the types of financial activities.

**Fraud flag:** An indicator showing if a transaction was flagged as fraudulent (1) or not (0). This is used for identifying potential fraudulent transactions.

**Customer age:** The age of the customer at the time of the transaction. It provides demographic information for customer analysis.

**Customer location:** The location of the customer, categorized as urban, rural, or suburban. This helps in geographic analysis of customer behaviour.

**Merchant category:** The type of business the merchant operates, like electronics or groceries. It helps in understanding the nature of the transaction.

**Fraud detected:** An indicator showing whether fraud was detected for the transaction (1) or not (0). This is critical for evaluating the performance of fraud detection systems.

**Accuracy rate:** The accuracy rate of the fraud detection system for that transaction. It measures how accurately the system identifies fraud.

**Transaction volume:** The number of transactions conducted by the customer. This helps in assessing the transaction frequency of each customer.

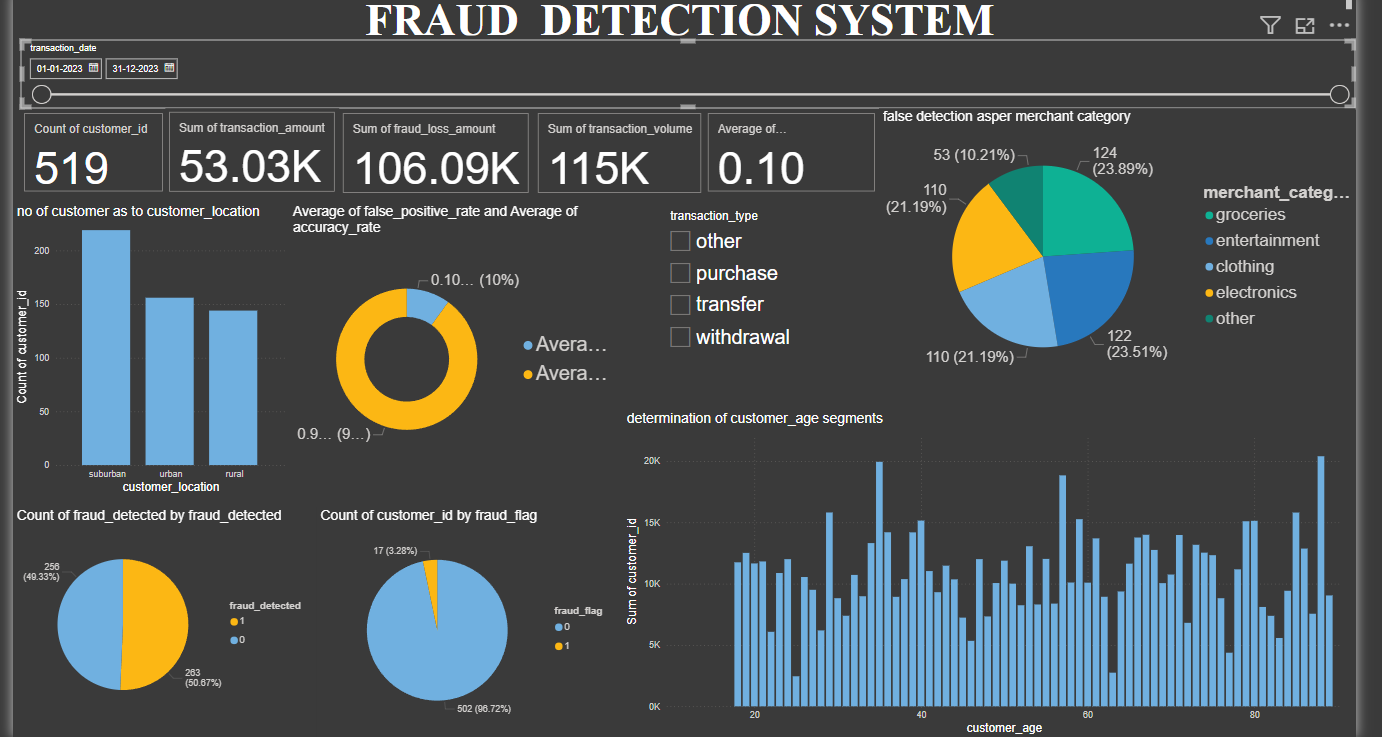
**False positive rate:** The rate at which the fraud detection system incorrectly flagged legitimate transactions as fraudulent. It is important for evaluating the efficiency of the system.

**Fraud loss amount:** The amount of financial loss due to fraudulent transactions. This metric is crucial for understanding the financial impact of fraud.

**METHODOLOGY**

* **DATA COLLECTION** – firstly the data of fraud detection system is collected
* **DATA CLEANING** –the various aspects of data is processed and cleaned use SQL. this process includes removal of null values for various primary key aspects, and for amount aspects the null values are replaced with the average of other values in the column is done, for the null values of categorical is replaced with a new category named other, and outlier detection is done to the data
* **DATA PROCESSING** – the cleaned data is then processed to the excel for more processing of values converted to rounded off and other aspects of date is also altered here.
* **DATA ANALYSIS –** an analytical dashboard is created to the transformed and cleaned data through power bi and different graphs for different metrics of data is represented

**OUTPUT**



**OUTPUT ANALYSIS**

* The dashboard starts off with the topic introducing fraud detection system
* There by give the range of date of extracted data to which the analysis is done
* A flash boards were kept to show various aspects of data firstly the no of frauds observed for a given period of time, then the total amount of transaction amounted to fraud, then the total amount of loss accounted through fraud transactions, the transaction volume is determined and finally a rate fraud positive rate is determined.
* A slicer is introduced for transaction type to determine the change in the data aspects as per those slicers applied categories include purchase, transfer or withdrawal
* A pie charts was introduced to the merchant categories which determines the division of categories to which the fraud occur. the analysis explains that the frauds more happen at the category of groceries, entertainment, then clothing and electronics hold a same percentage there is a null category which is termed as other
* A bar chart was introduced to know the location of customers who are more prone to frauds here suburban is more prone to frauds and next is urban there by rural is less prone.
* A 3rd pie chart is introduced to know the percentage contribution towards the average fraud positive rate and false accuracy rate
* A 2 pie charts were introduced in order to compare the fraud detected rate and fraud flag as the equal percentage of fraud come under detected showed as 1 and not detected as 0. fraud flag rate more of red flag rate is occurred through unusual transactions made, different logins from unfamiliar locations etc.
* A bar chat is introduced to identify the age of customers are more prone to fraud here it is observed as age of 35 ,55 80 and 29 is observed to be customer of those age groups are more prone to fraud.

**SUMMARY AND ANALYSIS**

* **Customer Demographics and Fraud:** the number of frauds that occur in a period of time as per customer demographics such as customer age of 35,55,80 and 29 are more prone to fraud and customer location od suburban are more prone to fraud. so, people of that location and area are must be alerted and properly observed so that fraud can be reduced.
* **Merchant Categories and Fraud:** the category of merchants who buy groceries are more frequently targeted by fraud. Educate the customers about the website which are fraudulent and buy from the trusted websites
* **Trends Over Time:** Use the transaction date column to identify trends in fraud incidents over time. This can help in understanding if certain periods have higher fraud rates.
* **Transaction Amount and Fraud:** Analyze the transaction amount column to see if higher-value transactions are more likely to be fraudulent.

**THANK YOU**