

# **Financial Fraud Detection: Anomaly Detection in Credit Card Transactions**

## **Business Introduction:**

Welcome to SecureGuard Financial Solutions, a leading provider of innovative fraud detection analysis for the financial industry. At SecureGuard, we specialize in developing cutting-edge solutions to safeguard our clients' assets and protect against fraudulent activities in real-time.

Our mission at SecureGuard is to provide comprehensive fraud detection services that empower financial institutions to maintain the trust and confidence of their customers while minimizing financial losses due to fraudulent transactions.

With the increasing prevalence of online transactions and digital payments, the risk of fraudulent activity has become a significant concern for financial institutions worldwide. From unauthorized credit card transactions to identity theft, the threat of fraud poses a substantial risk to both financial institutions and their customers.

That's where SecureGuard comes in. By analyzing patterns and anomalies in transaction data, our system can swiftly flag potentially fraudulent transactions, enabling financial institutions to take immediate action to mitigate the risk.

## **Problem Statement:**

Design a fraud detection system to identify fraudulent credit card transactions in real-time. The system should detect suspicious activities such as unauthorized transactions, unusual spending patterns, and fraudulent card usage to prevent financial losses for both cardholders and financial institutions.

## **Excel Tasks:**

### **1. Data Exploration:**

- Create a statistical summary on amt and city\_pop
- Plot a histogram on amt
- Create a report showing number of Frauds by gender and category
- Show the top 3 states by the highest number of transactions

### **2. Data Analysis in Excel:**

- Is there a correlation between transaction amount and city population?
- How many fraudulent transactions occurred in each category?
- How does the average transaction amount vary between different job roles?

## **SQL Tasks:**

### **3. Data Loading:**

- Create a schema named **finance**, set **finance** as the default schema, and create tables with cc\_data.csv and location\_data.csv

#### 4. Data Exploration with SQL:

- Calculate the total number of transactions in the cc\_data table
- Identify the top 10 most frequent merchants in the cc\_data table
- Find the average transaction amount for each category of transactions in the cc\_data table
- Determine the number of fraudulent transactions and the percentage of total transactions that they represent
- Join the cc\_data and location\_data tables to identify the latitude and longitude of each transaction
- Identify the city with the highest population in the location\_data table
- Find the earliest and latest transaction dates in the cc\_data table

#### 5. Using Data Aggregation with SQL:

- What is the total amount spent across all transactions in the cc\_data table?
- How many transactions occurred in each category in the cc\_data table?
- What is the average transaction amount for each gender in the cc\_data table?
- Which day of the week has the highest average transaction amount in the cc\_data table?

### Python Tasks (including EDA):

#### 6. Exploratory Data Analysis (EDA) with Python:

- What are the dimensions (number of rows and columns) of the dataset?
- How many unique values are there in each categorical variable?
- What is the distribution of numerical variables in the dataset?
- Are there any missing values in the dataset? If so, how should they be handled?
- What are the summary statistics (mean, median, min, max, etc.) for numerical variables?
- Is there any correlation between numerical variables? If so, how strong is the correlation?
- How does the distribution of an amt differ across is\_fraud categories?
- Are there any outliers in the city\_pop and amt?
- Are there any trends or patterns in the data over time (if applicable)?
- How does the target variable (if available) distribute across different categories?

- Are there any unusual or unexpected values in the dataset that require further investigation?
- Are there any potential data entry errors or inconsistencies in the dataset?
- How does the distribution of numerical variables vary between different groups or segments of the dataset?
- What are the top factors that influence the target variable, if applicable?
- Write an analysis report on performing exploratory data analysis (EDA) using Python in the context of building a fraud detection system for the financial industry.

## Tableau Tasks:

### 7. Interactive Dashboard Design in Tableau:

- Show the amount spent by different genders on different categories through a box and whisker plot
  - Choose **Text file** in Connect pane and upload the **cc\_data.csv** dataset
  - Open a new sheet
  - Drag **Amt** to **Columns**
  - Drag **Category** to **Rows**
  - Drag **Gender** to **Rows**
  - Drag **Category** to **Color** and **Amt** to **Label**
  - Choose **box and whisker plot** from **Show Me**
  - Drag **Category** into **Color** and **Amt** into **Label**
- Show the geographical distribution of transactions using latitude and longitude coordinates
  - Open a new worksheet and drag **Long** to **Columns** and **Lat** to **Rows**
  - Drag **Amt** to **Color** and **Size**
  - Right-click on the **Long** pill and click on **Dimension**
  - Right-click on the **Lat** pill and click on **Dimension**
- Use Tableau's built-in maps to visualize the locations of fraudulent transactions on a map
  - Open a new worksheet and drag **Long** to **Columns** and **Lat** to **Rows**
  - Right-click on the **Long** pill and click on **Dimension**
  - Right-click on the **Lat** pill and click on **Dimension**
  - Drag **Is Fraud** to **Color** and **Label**
- Create a time series analysis line chart showing the trend of monthly transaction amounts
  - Open a new worksheet and drag **Trans Date Trans Time** to **Columns**
  - Right-click on **Trans Date Trans Time** pill and select **Month**
  - Drag **Amt** to **Rows**, right-click on it, and choose the **Measure** as **Count**
  - Drag **Amt** to **Label** and change its measure to count as you have done in the previous step

- Create a Calculated Field for Inflation-Adjusted Transaction Amounts
  - Open a new sheet, click on **Analysis** and choose **Create Calculated Field**
  - Name the calculated field as **Inflation-Adjusted Transaction Amounts**, use the formula:  $[Amt] * (1 + 0.02)^{(YEAR(TODAY()) - YEAR([Trans Date Trans Time]))}$  and click on **Apply** and **OK**
  - Drag **Trans Date Trans Time** to **Columns** and **Inflation-Adjusted Transaction Amounts** to **Rows**
  - Right click on **Trans Date Trans Time** and select **Week Number**
  - Drag **Amt** to **Label**
  
- Create a dashboard with all the visualizations
  - Click on the new dashboard icon at the bottom
  - Drag **Box and whisker plot**, **Map 1**, **Fraud Map**, **Time Series** and **Inflation-Adjusted Transaction Amounts** to the dashboard area