Task 1: Data analysis

Here is the background information on your task

As a cybersecurity generalist at Commonwealth Bank, it is important to be aware of the increasing rate and complexity of financial fraud and the need for effective defense solutions. Financial fraud poses a significant challenge for financial institutions, and it is important for Commonwealth Bank to stay up to date with the latest fraud detection technologies and strategies to minimize risk. Protecting against and responding to fraud is a major responsibility for you and your team. By detecting and stopping fraud, the bank can protect its customers, employees and reputation while also enhancing the resilience of its financial system.

To help with this task, you will be using a tool called Splunk to visually represent the given data. Representing data in a visual format, also known as data visualization, makes it easier for the data analytics team to understand and gain insights. Visual data is a universal, fast and effective way to communicate information.

You will be building a dashboard to make it easier to identify patterns and trends in the given dataset. The dashboard will provide crucial reporting and metrics information that can aid in identifying and detecting fraud. By using this dashboard, the team will be able to quickly identify any suspicious activity and take the necessary steps to prevent fraud from occurring. Overall, the goal of this task is to use data visualization and a dashboard to make it easier to detect fraud and protect Commonwealth Bank and its customers from financial loss.

About the dataset
Data was collected and structured by the Fraud team. This dataset consists of payments from various customers made in different periods and amounts. The feature columns include:

- **Step:** This feature represents the month from the start of the simulation. The steps represent four months that the simulation ran virtually.
    - 0: May
    - 1: June
    - 2: July
    - 3: August
- **Customer:** Customer ID
- **Age:** Categorized age
    - 0.0: <= 18
    - 1.0: 19 - 25
    - 2.0: 26 - 35
    - 3.0: 36 - 45
    - 4.0: 46 - 55
    - 5.0: 56 - 65
- **Gender:** Gender of the customer
    - F: Female
    - M: Male
- **Postcode Origin:** The postcode of origin/source.
- **Merchant:** The merchant's ID.
- **Category:** Category of the purchase.
- **Amount:** Amount of the purchase.
- **Fraud:** Target variable that shows if the transaction is fraudulent - 1 or non-fraudulent - 0.

Note: Dataset was randomly generated and created for your virtual experience.

Here is your task

Your first task is to analyze and visualize the data you have prepared in a software tool called Splunk. Here are the steps you need to take:

1. Download the 60-day free trial of Splunk Enterprise. (The link is provided in the Resources section below.)
2. Install Splunk Enterprise on your computer.
3. Using the **"prepared data"** file in the Resources section, import this file into Splunk.
4. Study the file using the "Interesting Fields" section in Splunk. This tells you about the data you're using.
5. Create a dashboard to include the following charts/tables:
    1. Count by Category, Fraudulent transactions, Age and Merchant.
    2. Fraud detected by Age, Category, Step (month) and Gender.
    3. Which gender performed the most fraudulent activities and in what category?
    4. Which age group performed the most fraudulent activities and to what merchant?
6. Export the dashboard as a PDF and upload it below as your submission for this task.

Please refer to the **"Task Guide"** in the Resources to help you with this task.

Here are some resources to help you

XLS

Prepared_data.csv

Click to download file →

PDF

Task Guide

Click to download file →

Upload your Dashboard in PDF format

Select a File or drag it here.