Code Academy

Data Science and AI Stream

Homework 2 - Week3: Real-Life Fraud Detection Analysis

Due Date: April 30th. 2025 @11:59 pm

Dataset link: https://www.kaggle.com/datasets/sriharshaeedala/financial-fraud-detection-dataset?resource=download

Objective:

Act as a data analyst hired by a financial institution to assess and analyze mobile money transactions for potential fraud. Your task is to extract meaningful insights, identify suspicious behaviors, and propose actionable recommendations for fraud mitigation.

Scenario

You have been provided with a synthetic dataset that closely mimics real-world mobile money transactions. Your role is to analyze this dataset using tools like NumPy and pandas. The company expects actionable insights to enhance their fraud detection systems. Time is limited, and accuracy is crucial, as the company faces significant financial risks from fraud.

Phase 1: Understanding the Dataset

1. **Business Understanding**
    - Write a brief explanation (50-100 words) of the business context: why fraud detection is crucial for financial institutions.
    - Identify key questions your analysis should answer, such as:
        - Which transaction types are most prone to fraud?
        - Are there patterns in transaction amounts linked to fraud?
2. **Data Preparation**
    - Load the dataset into a pandas DataFrame.
    - Display the first few rows to familiarize yourself with the structure and contents.

- List the columns and explain the importance of each in the context of fraud detection.

Phase 2: Data Cleaning and Exploration

1. **Initial Review**
    - Check for missing or inconsistent values.
    - Remove or handle columns that are irrelevant for fraud detection. Justify your choice.
2. **Exploratory Analysis**
    - Calculate the total number of transactions and categorize them by type.
    - Identify the percentage of fraudulent transactions and compare them across transaction types.
    - Examine the distribution of transaction amounts (mean, median, standard deviation) for both fraudulent and non-fraudulent transactions.

Phase 3: Real-Life Fraud Detection Analysis

1. **Detecting Suspicious Patterns**
    - Identify and flag transactions exceeding the legal limit (amount > 200,000) as potentially fraudulent (isFlaggedFraud).
    - Find patterns in fraud-related transactions, such as the time step, type, or transaction amount.
2. **Group Analysis**
    - Group transactions by type and identify which types have the highest volume and value.
    - Examine whether certain customers (nameOrig or nameDest) are repeatedly involved in flagged or fraudulent transactions.
3. **Critical Thinking Task**

    - Write a short note to the company describing one scenario where a legitimate transaction might appear fraudulent. Suggest ways to improve fraud detection without flagging such cases incorrectly.

Phase 4: Insights and Recommendations

1. **Actionable Insights**
    - Summarize your key findings (250-300 words).
    - Highlight which transaction types and patterns are most indicative of fraud.
2. **Recommendations**
    - Propose three practical steps the company could implement to reduce fraud risks. Base these on your analysis.

Submission

Submit a well-documented Python file and pdf report that includes:

- Code for all analyses, with comments explaining the logic.
- A short summary and actionable recommendations.