Python for DS - Learners Space

Final Assignment: Credit Card Fraud Detection and Customer Profiling

Objective:
Your task is to analyze two datasets of credit card transactions to build models for fraud detection, customer profiling, and transaction amount prediction. This project will involve the use of anomaly detection, k-means clustering, and linear regression.

1) Fraud Detection - Credit Card Dataset for Fraud Detection This dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions. The data has been anonymized by the result of a PCA transformation to protect the cardholders' identities.

   You have to implement an Anomaly Detection Algorithm over the dataset. Use statistical methods or algorithms like Isolation Forest to detect anomalous transactions that might indicate fraud. Visualise these anomalies and identify the patterns and characteristics that differentiate them from normal transactions.

2) Customer Profiling - Credit Card Customer Data - This different dataset of Credit Card Transactions is provided with various features. Firstly preprocess the data and then implement the various algorithms -

   a) Apply k-means clustering without the use of in-built libraries like scikit-learn to segment customers based on their transaction behaviour and profiles and also determine the optimal number of clusters using methods like the Elbow method or Silhouette score.
   b) Build a linear regression model to predict the transaction amount (Avg. Credit limit) for a potential customer based on the given features. Identify the key features to be used.