Data Science Assignment

1. **Classification model:**
    a. **Loan Status Prediction:**
        i. This is a Property Loan Status Prediction dataset, we have the data of applicants who had previously applied for the loan based on the property and applicant-based metrics.
        ii. The bank will decide whether to give a loan to the applicant based on some factors such as Applicant Income, Loan Amount, previous Credit History, Co-applicant Income, etc.
        iii. The objective is to build a Machine Learning Model to predict the loan to be approved or to be rejected for an applicant.
        iv. You are free to use any classification model such Logistic Regression, Decision Tree, RF or XGBoost.
        v. Write the performance and key insights on the data in a word doc.

2. **Regression model:**
    a. **Sales Forecasting:**
        i. You are provided with historical sales data for 45 retail stores located in different regions. Each store contains a number of departments, and you are tasked with predicting the department-wide sales for each store.
        ii. In addition, the retail company runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labour Day, Thanksgiving, and Christmas.
        iii. Data:
            1. stores.csv: This file contains anonymized information about the 45 stores, indicating the type and size of the store.
            2. features.csv: Comprise of information on Markdown and macroeconomic factors.
            3. train.csv: This is the historical training data.
        iv. Write the performance and key insights on the data in a word doc.
3. **Clustering:**
    a. Credit Card Data: Marketing Strategy
        i. This case requires to develop a customer segmentation to define marketing strategy.
        ii. Sample Dataset summarizes the usage behaviour of about 9000 active credit card holders during the last 6 months. The file is at a customer level with 18 behavioural variables.

iii. Following is the Data Dictionary for Credit Card dataset:

CUST_ID: Identification of Credit Card holder (Categorical)

BALANCE: Balance amount left in their account to make purchases

BALANCE_FREQUENCY: How frequently the Balance is updated, score between 0 and 1 (1 = frequently updated, 0 = not frequently updated)

PURCHASES: Amount of purchases made from account

ONEOFF_PURCHASES: Maximum purchase amount done in one-go

INSTALLMENTS_PURCHASES: Amount of purchase done in installment

CASH_ADVANCE: Cash in advance given by the user

PURCHASES_FREQUENCY: How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased)

ONEOFFPURCHASESFREQUENCY: How frequently Purchases are happening in one-go (1 = frequently purchased, 0 = not frequently purchased)

PURCHASESINSTALLMENTSFREQUENCY: How frequently purchases in installments are being done (1 = frequently done, 0 = not frequently done)

CASHADVANCEFREQUENCY: How frequently the cash in advance being paid

CASHADVANCETRX: Number of Transactions made with "Cash in Advanced"

PURCHASES_TRX: Number of purchase transactions made

CREDIT_LIMIT: Limit of Credit Card for user

PAYMENTS: Amount of Payment done by user

MINIMUM_PAYMENTS: Minimum amount of payments made by user

PRCFULLPAYMENT: Percent of full payment paid by user

TENURE: Tenure of credit card service for user