Assignment Task: Basic Customer Segmentation Using Clustering

# Dataset:

Use the Mall Customers Dataset  
File: Mall\_Customers.csv  
Link: https://www.kaggle.com/vjchoudhary7/customer-segmentation

# Objective:

You will perform basic clustering analysis on mall customer data to understand spending habits and group similar customers together. You are required to perform preprocessing, data visualization, clustering, and cluster analysis.

## Part 1: Data Preprocessing (Minimum 5 Steps)

1. Load the dataset using pandas.

2. Display the first 10 rows using .head(10) and print shape of dataset.

3. Use .info() to check column data types and memory usage.

4. Check for missing values and handle (if any).

5. Convert the Gender column into numbers (e.g., Male=0, Female=1).

6. Drop the CustomerID column as it's not useful for clustering.

7. Create a new dataset with only relevant columns: Gender, Age, Annual Income, Spending Score.

## Part 2: Data Visualization (Minimum 5 Plots)

1. Count plot of Gender (How many male/female customers).

2. Histogram of Age (Distribution of customer ages).

3. Scatter plot of Age vs Spending Score.

4. Scatter plot of Annual Income vs Spending Score.

5. Boxplot of Spending Score by Gender.

6. Add at least 1 personal observation about a pattern you noticed (e.g., young customers spend more).

## Part 3: Apply K-Means Clustering (Minimum 5 Steps)

1. Choose the following features for clustering: ['Annual Income (k$)', 'Spending Score (1-100)'].

2. Apply KMeans with n\_clusters=3.

3. Use .fit() and .predict() to assign each data point to a cluster.

4. Store the cluster labels in a new column in your dataset called 'Cluster'.

5. Print how many customers are in each cluster.

6. Optional: Print the cluster centers (.cluster\_centers\_).

## Part 4: Cluster Visualization & Basic Analysis (Minimum 5 Steps)

1. Create a scatter plot with:  
 - X-axis = Annual Income  
 - Y-axis = Spending Score  
 - Color = Cluster Label

2. Plot cluster centers using red 'X' markers.

3. Add titles, labels, and a legend.

4. Write a short analysis: Which cluster seems to be high income & high spenders?

5. Write another insight: Which cluster could be low income & low spending?

6. (Optional) Plot same clusters using Age and Spending Score for another perspective.