DATASET (MALL CUSTOMERS) DESCRIPTION

Feature	Description	Type / Notes
CustomerID	A unique identifier for each customer.	Numeric (but not useful for clustering).
Gender	The gender of the customer: Male or Female.	Categorical; can be used for analysis, but often encoded numerically if included in clustering.
Age	The age of the customer in years.	Numeric; can help identify age-based segments.
Annual Income (k$)	The yearly income of the customer in thousands of dollars.	Numeric; often a key feature for clustering to segment by spending power.
Spending Score (1–100)	A score assigned by the mall based on customer behavior and spending patterns. Higher means more frequent or higher spending.	Numeric; usually combined with income to identify customer types.

Questions

1. **Load the Dataset**

   o Import the dataset into Python using pandas.

   o Print the name of each feature.

   o Display the first 5 rows.

2. **Select Features for Clustering**

   o Choose **Annual Income** and **Spending Score** as the two features.
      Why do you think these two features are commonly used for customer segmentation?

3. **Visualize the Data**

    o Create a scatter plot of Annual Income vs Spending Score.
    Do you observe any natural grouping in the scatter plot?

4. **Determine the Optimal Number of Clusters (k)**

    o Use the **Elbow Method** (compute WCSS for k = 1 to 10).
    Based on the elbow graph, what is the best value of k and why?

5. **Apply K-Means Clustering**

    o Fit the K-Means model using the chosen k.

    o Predict the cluster for each customer.
    How many customers belong to each cluster?

6. **Visualize the Clusters**

    o Plot the clusters along with their centroids.
    Describe the characteristics of each cluster (e.g., high income & high spending, etc.).

7. **Interpretation**
    If you were the mall manager, how would you use these clusters to design better marketing strategies?