**Task 2: Customer Segmentation**

**1. Objective**

The goal of this task is to perform **customer segmentation** using unsupervised learning.  
By analyzing spending patterns, income levels, and other customer attributes, we aim to group similar customers together.  
This helps businesses design **personalized marketing strategies** and improve customer relationship management.

**2. Dataset**

* **Source**: Mall Customers Dataset (Kaggle).
* **Dataset Description**:  
  The dataset contains information about customers visiting a shopping mall.  
  Features include:
  + **CustomerID** → Unique identifier for each customer
  + **Gender** → Male/Female
  + **Age** → Age of the customer
  + **Annual Income (k$)** → Customer’s yearly income
  + **Spending Score (1–100)** → A score assigned by the mall based on customer’s spending behavior and loyalty
* **Dataset Size**: 200 rows × 5 columns

**3. Problem Type**

* **Unsupervised Learning**
* **Clustering Problem** (no target variable).
* Algorithm used: **K-Means Clustering**

**4. Approach**

**(a) Data Preprocessing**

1. Load dataset and check missing values (none in this dataset).
2. Encode **Gender** column (Male/Female → numeric).
3. Select features for clustering (Age, Annual Income, Spending Score).

**(b) Feature Selection**

* Key features used:
  + **Annual Income**
  + **Spending Score**
  + (sometimes Age is also included for 3D visualization)

**(c) Optimal Number of Clusters**

* Use the **Elbow Method** to find the best number of clusters (k).
* Plot **WCSS (Within Cluster Sum of Squares)** vs **Number of Clusters**.
* From the curve, choose optimal k = 5.

**(d) Model Building**

* Apply **K-Means Clustering** with n\_clusters=5.
* Assign each customer to a cluster.

**(e) Visualization**

* 2D scatter plots for **Annual Income vs Spending Score**.
* Clusters are shown in different colors.
* Helps visualize distinct groups of customers.

**5. Tools & Libraries**

* **Python**
* **Pandas, NumPy** → Data preprocessing
* **Matplotlib, Seaborn** → Visualization
* **Scikit-learn** → KMeans clustering

**6. Workflow**

1. Import libraries and load dataset.
2. Explore dataset (shape, columns, summary statistics).
3. Encode categorical data (Gender).
4. Select features for clustering.
5. Use Elbow Method to determine optimal number of clusters.
6. Train K-Means model.
7. Visualize clusters and interpret results.
8. **Results**

* K-Means forms **5 distinct customer clusters** such as:
  1. Low income, low spending
  2. High income, high spending
  3. Average income, high spending
  4. High income, low spending
  5. Average customers
* Visualization shows clear separation of customer segments.

**8. Applications**

* Helps businesses:
  + Identify **high-value customers**.
  + Design **targeted marketing campaigns**.
  + Improve **customer loyalty programs**.
  + Optimize product recommendations.

