Approach Used

- **Age**
- **Annual Income (k$)**
- **Spending Score (1-100)**

Steps Followed:

1. **Data Preprocessing:**
    - Standardized the feature values using StandardScaler to normalize data and ensure all features contribute equally to clustering.
2. **Elbow Method for Optimal K:**
    - Calculated the **Within-Cluster Sum of Squares (WCSS)** for different values of **K (1 to 10)**.
    - Used the **elbow method** to determine the best number of clusters.
3. **Clustering with K-Means:**
    - Chose **K = 5** as the optimal number of clusters.
    - Applied K-Means clustering with the **k-means++** initialization to improve cluster convergence.
    - Assigned each customer to one of the five clusters.

Challenges Faced

- **Choosing the Right K:** The elbow method helped, but choosing the optimal K can sometimes be ambiguous.
- **Feature Scaling Impact:** Without scaling, the model was biased towards features with larger numeric values.
- **Cluster Interpretation:** Defining meaningful customer segments from the clusters required additional analysis.

Model Performance & Improvements

1. **Cluster Insights:**
    - The model segmented customers into five distinct groups based on spending habits and income levels.
    - Potential customer categories could include **high spenders, budget-conscious customers, young spenders, and affluent customers**.
2. **Possible Improvements:**
    - **Silhouette Score Analysis:** Use silhouette score to validate clustering effectiveness.
    - **Feature Engineering:** Introduce additional features (e.g., purchasing frequency) for better segmentation.
    - **Hierarchical Clustering Comparison:** Compare results with hierarchical clustering for validation.