**Customer Segmentation Using Clustering Techniques Report**

**Data Analysis and Data Science Task – 3**

**1. Introduction**

Customer segmentation is a strategic approach that enables businesses to group customers based on similar characteristics, such as purchasing behavior. It helps in personalizing marketing strategies, improving customer service, and ultimately boosting sales. This report outlines the methodology and findings of a clustering analysis conducted using Python to segment customers.

**2. Objective**

The primary objective is to segment customers based on their age, annual income, and spending score using clustering algorithms. By identifying distinct groups within the customer base, businesses can tailor marketing efforts and optimize product offerings to different segments.

**3. Dataset Overview**

* **Dataset Name:** customer\_data.csv
* **Columns:**
  + Customer ID – Unique identifier for each customer.
  + Age – Age of the customer.
  + Annual Income – Customer’s yearly income (in currency units).
  + Spending Score – Score based on customer spending behavior.

**4. Methodology**

**Step 1: Loading and Inspecting the Data**

* Imported the dataset using Pandas.
* Checked dataset shape, missing values, duplicates, and data types.
* Performed descriptive statistics to understand data distribution.

**Step 2: Data Preprocessing**

* Used StandardScaler from sklearn.preprocessing to scale features.
* Ensured all numerical variables (age, income, score) are on a comparable scale.
* This step improves the performance of clustering algorithms.

**Step 3: Determining Optimal Number of Clusters**

* **Employed Elbow Method:**
  + Calculated WCSS (Within-Cluster Sum of Squares) for a range of clusters.
  + Plotted number of clusters vs. WCSS to find the 'elbow point'.
* Used Silhouette Score for further validation.

**Step 4: Applying K-Means Clustering**

* Based on the Elbow Method, selected optimal k value.
* Applied KMeans from sklearn.cluster.
* Added a new column to the dataset with assigned cluster labels.

**5. Data Visualization**

**2D Scatter Plot**

* Used PCA to reduce dimensions for visualization.
* Created a scatter plot with different colors for each cluster.
* Helped in observing how clearly the customer segments were formed.

**Pair Plot**

* Visualized the pairwise relationships between age, income, and spending score within clusters.

**Cluster Centroids**

* Visualized centroids of each cluster to understand the characteristics of each customer group.

**Elbow Curve**

* Presented the Elbow Method graph showing optimal cluster number.

**6. Deliverables**

* **Clustered Dataset:**
  + Includes cluster labels for each customer.
* **Visualizations:**
  + Scatter plots, Elbow Curve, and Pair Plots.
* **Recommendation Report:**
  + Actionable insights for business strategies based on clusters.

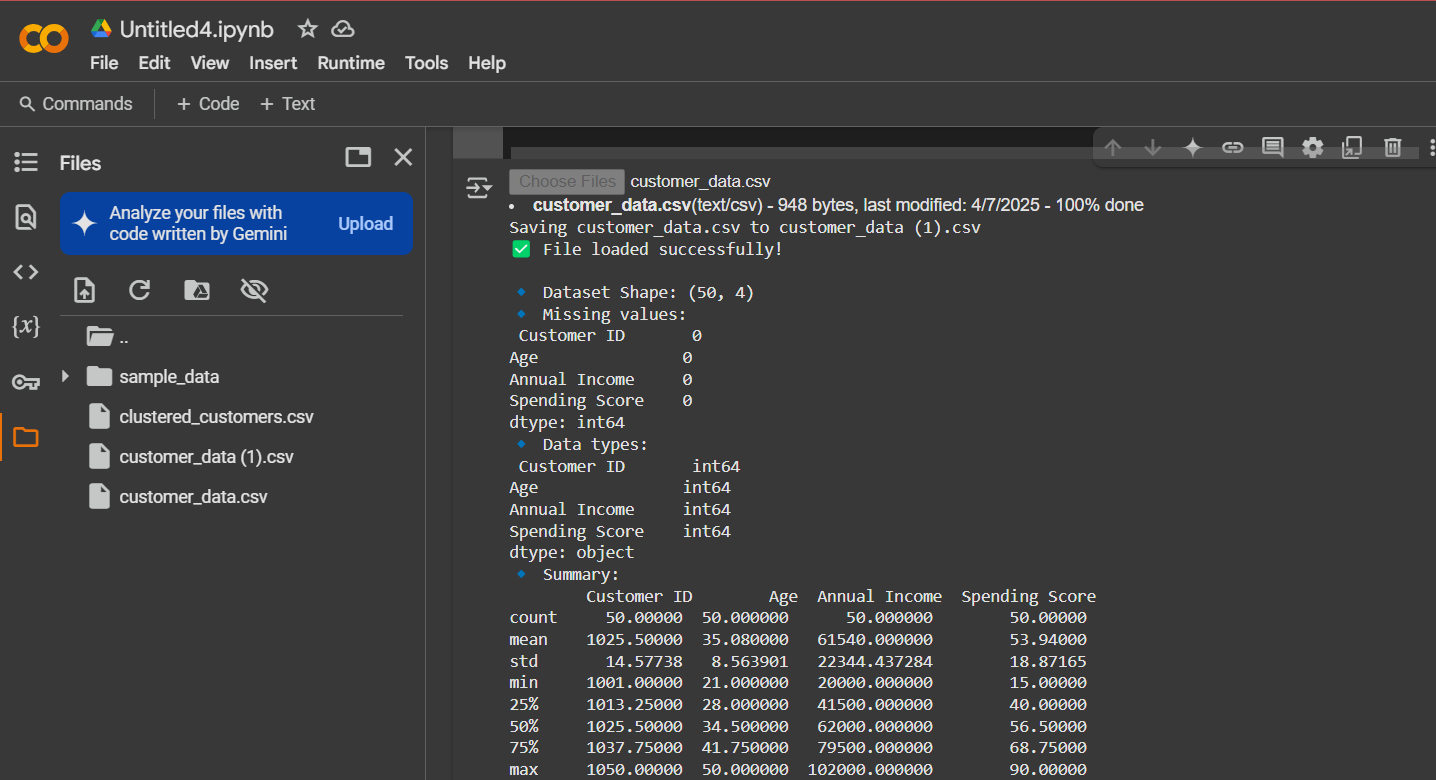
**7. Recommendations & Insights**

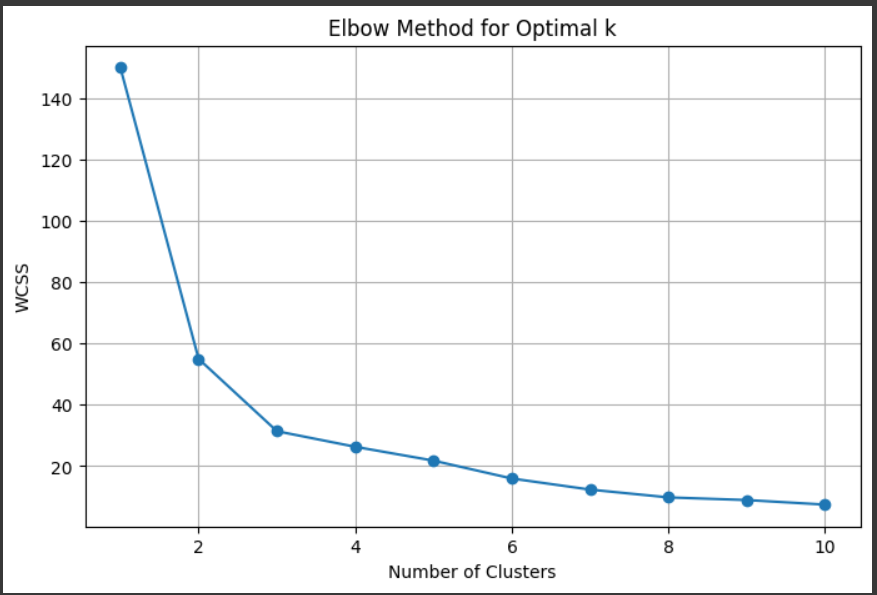
1. **High-Income, High-Spending Cluster:**
   * Target this group for premium products and loyalty programs.
2. **Young Customers with Moderate Income and Spending:**
   * Focus on trendy, budget-friendly options and digital campaigns.
3. **Low-Income, Low-Spending Cluster:**
   * Promote discounts or entry-level products to encourage spending.
4. **Older Age Group with Steady Income:**
   * Provide value-focused offerings and personalized services.

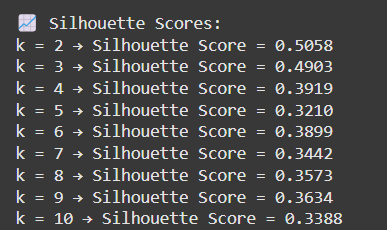
**8. Expected Outcomes**

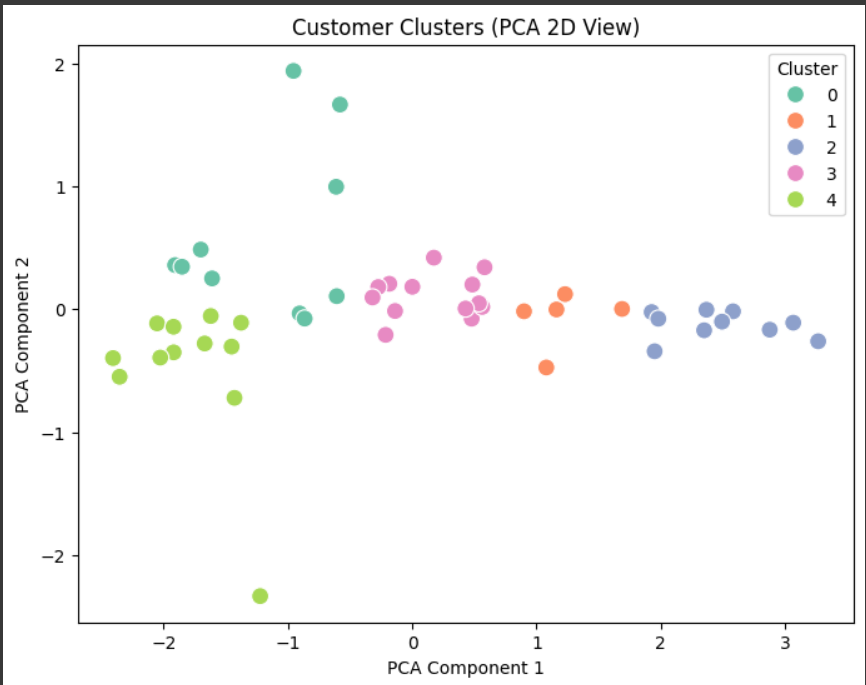
* Clear segmentation of customers into meaningful groups.
* Visual representation of customer clusters.
* Enhanced decision-making through data-driven customer insights.
* Better targeting, marketing, and resource allocation.

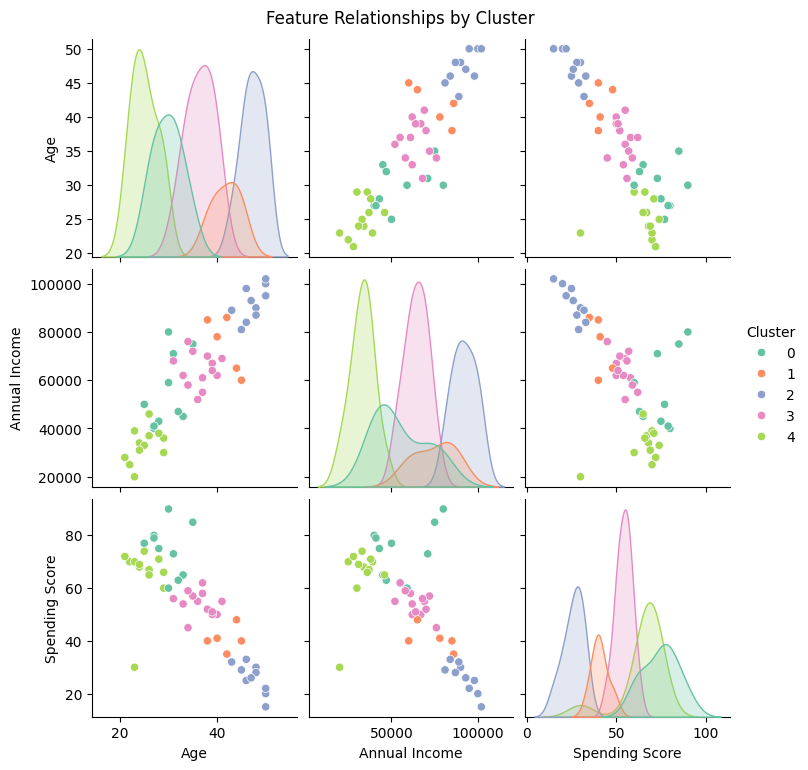
**RESULTS:**

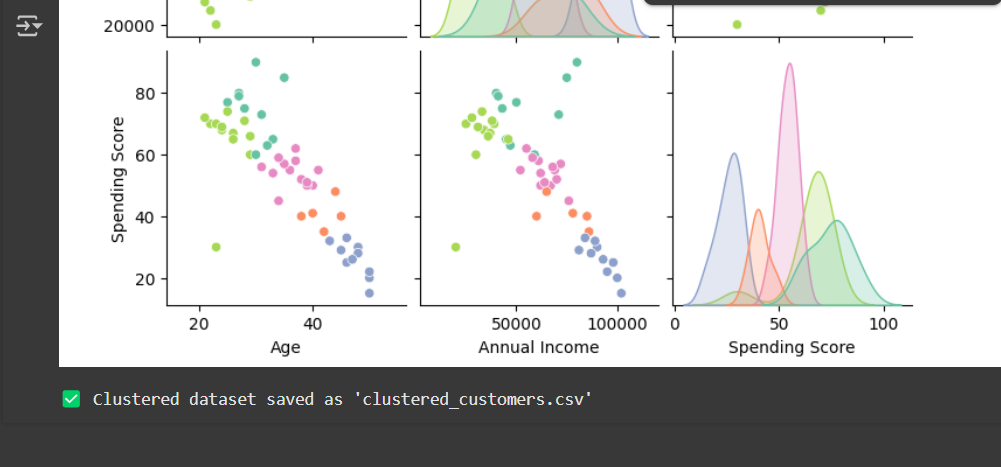
****

****

****

****



****