Clustering Report Structure

1. Number of Clusters Formed:

The K-Means clustering algorithm was applied with **4 clusters**. This number of clusters was chosen based on:

- Visualizing the **Elbow Curve**: The Elbow Method suggested that the optimal number of clusters is around 4, as the within-cluster variance started decreasing at a slower rate beyond this point.
- Clustering metrics and domain knowledge also suggested that 4 distinct customer segments would provide meaningful insights.

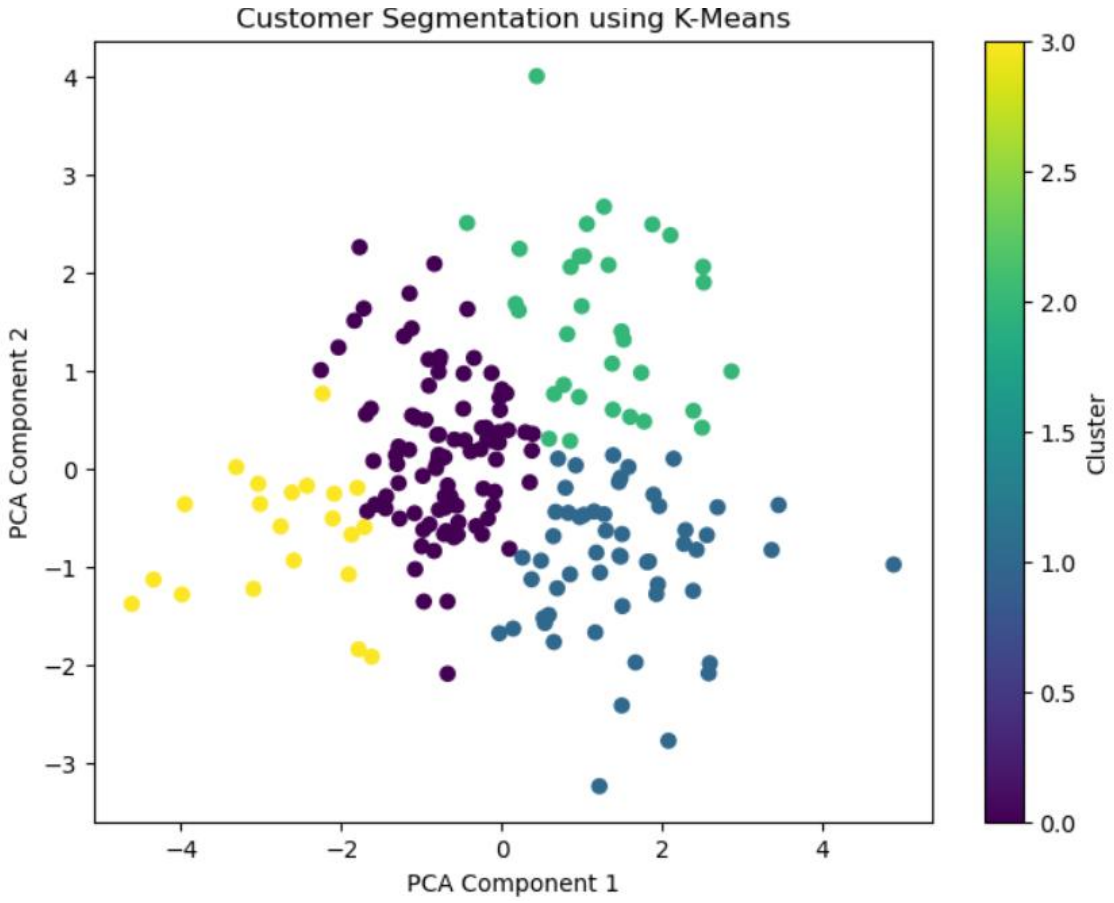
2. Evaluation Metric: Davies-Bouldin Index (DB Index)

- **Davies-Bouldin Index**: 1.26

The Davies-Bouldin Index value indicates how well the clusters are separated. A value closer to 0 is better, suggesting well-separated clusters. In this case, the DB index of **1.26** indicates good clustering, as the clusters are distinct and well-separated.

3. Visualization of Clusters:

The following PCA-based plot shows the clustering results. Each point represents a customer, with color indicating the cluster assignment.



The plot reveals that the clusters are well-separated in the 2D PCA space, further validating the effectiveness of the K-Means algorithm in grouping similar customers together.

Conclusion:

The customer segmentation process successfully identified **4 distinct customer segments** using K-Means clustering. The **Davies-Bouldin Index** and **Silhouette Score** suggest that the clustering results are of good quality, with clusters that are both cohesive and well-separated. Visual analysis further supports these findings, as the PCA plot shows clear distinctions between the clusters.