Clustering Report on Customer Segmentation

1. Number of Clusters

The optimal number of clusters formed for customer segmentation is:

Optimal Number of Clusters: 6

2. DB Index

The Davies-Bouldin Index (DB Index) measures the average similarity of each cluster with its most similar cluster. A lower value indicates better clustering performance. In this case:

DB Index Value: 0.8378

This value indicates that the clusters have a reasonable level of separation, though improvements could still be made in terms of intra-cluster cohesion and inter-cluster separation.

3. Other Clustering Metrics

- **Silhouette Score**: A value close to +1 suggests that samples are well matched to their own cluster and poorly matched to other clusters. Here, the silhouette score is:

Silhouette Score: 0.371

This score is relatively low, indicating that while the clusters are somewhat distinguishable, there might be some overlap or unclear boundaries between the clusters.

- **Calinski-Harabasz Index**: This index measures the ratio of between-cluster dispersion to within-cluster dispersion. A higher value is preferable. For this segmentation, the index is:

Calinski-Harabasz Index: 146.29

This suggests that the clusters have a moderate level of dispersion between them and relatively tight internal clusters, which is a positive indicator of cluster validity.

4. Visualization

The customer segmentation has been visualized using PCA to reduce the dimensionality of the data. The scatter plot shows the customers grouped into 6 clusters, colored according to their cluster label. This visualization provides insight into how the customers are distributed across the segments.

5. Final Thoughts

The clustering results show that 6 clusters offer a reasonable level of segmentation for the customers. Although the silhouette score indicates some potential overlap, the DB Index and Calinski-Harabasz Index suggest that the clusters are relatively well-formed. Further refinements, such as exploring different clustering algorithms or tuning hyperparameters, may yield even more distinct segments.

This segmentation can be useful for customer targeting, understanding buying behavior, and tailoring marketing strategies to specific customer groups.