Deliverables: Clustering Results Report

1. Number of Clusters Formed

Using the **elbow method** to determine the optimal number of clusters, we evaluated the inertia for different values of kk (ranging from 2 to 10). Based on the elbow curve, we chose **4 clusters** as the optimal number. This was determined by observing the point where the inertia reduction starts to slow down.

2. Davies-Bouldin Index (DB Index)

The **Davies-Bouldin Index (DB Index)** measures the average similarity ratio of each cluster with the one most similar to it. A lower DB Index indicates better clustering. For our dataset, the DB Index value is:

- **Davies-Bouldin Index**: **1.0673**

A DB Index value closer to 0 suggests that the clusters are well-separated and compact. The value of 1.0673 indicates a moderate separation between the clusters, but there is room for improvement.

3. Silhouette Score

The **Silhouette Score** measures how similar each point is to its own cluster compared to other clusters. The score ranges from -1 (bad clustering) to +1 (good clustering). A value close to 1 indicates that the points are well-clustered, while a value close to 0 suggests that the clusters are overlapping.

- **Silhouette Score**: **0.3211**

This score indicates that the clustering is somewhat effective, but there is still potential for improvement in the separation of the clusters.

4. Clustering Metrics

In addition to the DB Index and Silhouette Score, we can also look at the **cluster visualization** using **Principal Component Analysis (PCA)**. PCA was used to reduce the dimensionality of the dataset, allowing us to plot the clusters in a 2D space. The clusters are visually separated, although the separation is not perfect, which aligns with the moderate DB Index and Silhouette Score values.

5. Summary of Clustering Results

- **Optimal Number of Clusters**: 4
- **Davies-Bouldin Index**: 1.0673 (moderate cluster separation)
- **Silhouette Score**: 0.3211 (moderate clustering quality)

These results suggest that while the clustering algorithm has successfully identified 4 distinct customer groups, there is still potential to improve the cluster separation and overall quality of the clustering.