Clustering Results Report

1. Number of Clusters Formed

After applying the **Elbow Method** to determine the optimal number of clusters, the optimal number of clusters for customer segmentation was found to be **4 clusters**. This is the point where the inertia (sum of squared distances from each point to its assigned cluster center) levels off, indicating a good balance between model complexity and data fit.

2. Davies-Bouldin Index (DB Index)

The **Davies-Bouldin Index** is a metric that evaluates the quality of clustering by comparing the average similarity ratio of each cluster with the one that is most similar. A lower Davies-Bouldin index indicates better-defined and well-separated clusters.

- **DB Index value**: **0.7219**

A **DB Index** of **0.7219** suggests that the clusters are reasonably well-separated and compact. However, further optimization could still be made to improve the cluster separation.

3. Silhouette Score

The **Silhouette Score** is another metric used to evaluate the clustering quality. It measures how similar an object is to its own cluster compared to other clusters. The score ranges from -1 (incorrect clustering) to +1 (well-clustered), with a score close to 0 indicating overlapping clusters.

- **Silhouette Score**: **0.6014**

A **Silhouette Score** of **0.6014** indicates that the clusters are fairly well-defined, with most customers being clearly assigned to their respective clusters. This score suggests that the KMeans algorithm has produced a reasonable segmentation.

4. PCA Visualization

To visualize the clustering results, a **Principal Component Analysis (PCA)** was applied to reduce the data dimensions to 2D, which allows for easier visual inspection of the clusters. The scatter plot shows the customers in 2D space with different colors corresponding to their assigned clusters.

This visualization helps in understanding the distribution and separation of the clusters in the feature space, further confirming that 4 clusters provide a meaningful segmentation.

Summary of Clustering Performance

- **Optimal Clusters**: 4
- **DB Index**: 0.7219 (Good separation between clusters)
- **Silhouette Score**: 0.6014 (Clusters are well-defined with some overlap)
- **PCA Visualization**: Shows distinct customer segments with clear boundaries between the clusters.

Recommendations

- The clustering results suggest that customer segmentation is meaningful and that the 4 clusters provide a clear distinction between different customer groups based on their spending behavior and region.
- While the clustering results are promising, further analysis and optimization (such as testing different clustering algorithms or adjusting hyperparameters) could improve the overall clustering quality.