Clustering Analysis Report

In this analysis, we performed customer segmentation using two clustering algorithms: **K-Means** and **Agglomerative Clustering**. We aimed to determine the optimal number of clusters (K) using the **Elbow Method** and evaluated the performance of each algorithm based on various clustering metrics.

Optimal Number of Clusters:

Based on the Elbow Method, the optimal number of clusters (K) was found to be **5**.

K-Means Clustering Results:

- **Number of Clusters Formed**: 5
- **Davies-Bouldin Index (DB Index)**: 0.8525
  - A lower DB Index indicates better clustering, with well-separated and cohesive clusters. K-Means shows a reasonably good separation with this DB value.
- **Silhouette Score**: 0.3535
  - The Silhouette Score measures the consistency of clusters. A value closer to 1 indicates well-defined clusters, but here the score suggests moderate cluster cohesion.
- **Calinski-Harabasz Score**: 143.70
  - This score measures the ratio of within-cluster dispersion to between-cluster dispersion. Higher values indicate better-defined clusters, and the score suggests relatively good clustering.

Agglomerative Clustering Results:

- **Number of Clusters Formed**: 5
- **Davies-Bouldin Index (DB Index)**: 0.8851
  - The DB Index for Agglomerative Clustering is slightly higher than K-Means, indicating that the clusters may not be as well-separated or cohesive as those produced by K-Means.
- **Silhouette Score**: 0.3427
  - The Silhouette Score for Agglomerative Clustering is close to K-Means, indicating similarly moderate cluster cohesion.
- **Calinski-Harabasz Score**: 130.61
  - The Calinski-Harabasz Score for Agglomerative Clustering is slightly lower than K-Means, suggesting that the clustering might not be as well-defined.

Based on these metrics, **K-Means** with 5 clusters is recommended for customer segmentation.