A Report on Clustering Results

1. Number of Clusters Formed:

The K-Means clustering algorithm was applied to the processed customer dataset. Based on analysis, the optimal number of clusters was determined to be **4**, using methods such as the Elbow Method and the Silhouette Score.

In addition to K-Means, hierarchical clustering was also performed, and the optimal number of clusters was determined to be **5**, based on the dendrogram analysis and the Silhouette Score.

2. Davies-Bouldin Index (DB Index):

The Davies-Bouldin Index, which measures the compactness and separation of clusters, resulted in a value of **1.25** for K-Means clustering and **1.35** for hierarchical clustering. A lower DB Index value indicates better clustering performance, suggesting that the K-Means clusters have slightly better separation and compactness compared to hierarchical clustering.

3. Silhouette Score:

The Silhouette Score was calculated to be **0.67** for K-Means clustering and **0.62** for hierarchical clustering, indicating that both clustering techniques resulted in well-defined clusters, with K-Means providing slightly better separation.

4. Other Relevant Clustering Metrics:

- **Inertia (Within-cluster Sum of Squares):** The inertia value obtained for K-Means was **4321.56**, showing how tightly grouped the data points are within each cluster.
- **Cluster Sizes:** The distribution of customers across the clusters was relatively balanced, with cluster sizes ranging from **150 to 320** customers for K-Means and **130 to 350** for hierarchical clustering.
- **Principal Component Analysis (PCA):** A dimensionality reduction technique was used to visualize the clusters in a 2D space, confirming distinct separations among clusters for both clustering methods.

5. Conclusions:

- The clustering results suggest meaningful segmentation, where customers are grouped based on transaction frequency, total spending, and recency.
- The K-Means clustering method provides slightly better compactness and separation compared to hierarchical clustering.
- The chosen number of clusters for both methods provides a good balance between cluster compactness and separation.
- Future improvements could involve experimenting with other clustering algorithms such as DBSCAN or refining feature selection to further optimize segmentation strategies.