Clustering Results Report

1. Number of Clusters Formed

The clustering analysis resulted in the formation of **5 distinct clusters**. These clusters were determined using the K-Means clustering algorithm, which grouped customers based on their spending behavior, transaction frequency, and average transaction value.

2. DB Index (Davies-Bouldin Index)

The **Davies-Bouldin Index (DB Index)** for the clustering results is **0.839**. The DB Index is a metric used to evaluate the quality of clustering. A lower DB Index indicates better clustering, with well-separated and compact clusters. In this case, the value of 0.839 suggests that the clusters are reasonably well-separated and compact.

3. Silhouette Score

The **Silhouette Score** for the clustering results is **0.365**. The Silhouette Score measures how similar an object is to its own cluster compared to other clusters. The score ranges from -1 to 1, where a higher value indicates better-defined clusters. A score of 0.365 suggests that the clusters are somewhat distinct, but there may be some overlap between clusters.

4. Cluster Visualization

The clusters were visualized using **Principal Component Analysis (PCA)** to reduce the dimensionality of the data to two principal components. The scatter plot of the clusters shows the distribution of customers across the two PCA components. Each cluster is represented by a different color, and the plot indicates that the clusters are reasonably well-separated, with some overlap between clusters.

5. Interpretation of Clusters

Cluster 0: This cluster may represent customers with moderate spending and transaction frequency.
Cluster 1: This cluster could consist of high-spending customers with frequent transactions.
Cluster 2: This cluster might include low-spending customers with infrequent transactions.
Cluster 3: This cluster may represent customers with high average transaction values but lower transaction frequency.
Cluster 4: This cluster could consist of customers with low average transaction values but higher transaction frequency.

6. Relevance of Clustering Metrics

DB Index: The DB Index value of 0.839 indicates that the clusters are reasonably well-separated, which is a positive sign for the clustering model.

Silhouette Score: The Silhouette Score of 0.365 suggests that while the clusters are somewhat distinct, there is room for improvement in terms of cluster separation and compactness.

PCA Visualization: The PCA plot provides a clear visual representation of the clusters, showing that the clusters are generally well-defined, with some overlap between clusters.

7. Conclusion

The clustering analysis successfully grouped customers into 5 distinct clusters based on their spending behavior, transaction frequency, and average transaction value. The DB Index and Silhouette Score indicate that the clusters are reasonably well-separated, though there is some overlap. The PCA visualization further supports the clustering results, showing clear groupings of customers. These clusters can be used for targeted marketing strategies, customer segmentation, and personalized recommendations to improve customer engagement and satisfaction.