Customer Segmentation - Clustering Analysis Report

Objective :

The goal of this analysis is to segment customers into different clusters based on their profile and transaction history. This segmentation will provide insights into customer behavior and enable tailored marketing, recommendations, and services.

Clustering Algorithm

For this task, i have used the **K-Means Clustering** algorithm, which is one of the most widely used unsupervised machine learning techniques for customer segmentation. The K-Means algorithm partitions customers into groups, or clusters, where each cluster consists of customers who are similar to each other in terms of the features provided.

Based on the **Silhouette Score**, the optimal number of clusters is **2**.

Number of Clusters (k)

We tested multiple values of **k** (the number of clusters) ranging from 2 to 10 to find the optimal number of clusters. The evaluation was based on clustering metrics such as the **Davies-Bouldin Index (DB Index)** and **Silhouette Score**.

Evaluation Metrics

1. **Davies-Bouldin Index (DB Index)**:
   The DB Index is used to evaluate the quality of clusters. It is a ratio of the sum of the cluster dispersions to the separation between clusters. A **lower DB Index** indicates better clustering with compact and well-separated clusters.
2. **Silhouette Score**:
   The Silhouette Score is a measure of how similar an object is to its own cluster compared to other clusters. A **higher Silhouette Score** indicates that the clusters are more distinct and well-defined.

Clustering Metrics

Number of Clusters	DB Index	Silhouette Score
2	0.7327	0.4859

3	0.7550	0.4330
4	0.8082	0.3944
5	0.8466	0.4054
6	0.8593	0.4032
7	0.8298	0.4092
8	0.8630	0.3999
9	0.8134	0.4041
10	0.7950	0.4027

Optimal Number of Clusters

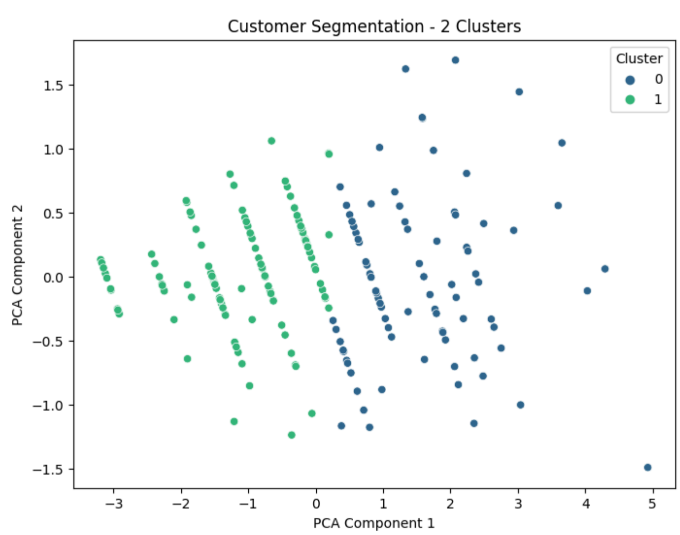
Based on the **Silhouette Score**, the optimal number of clusters was found to be **2 clusters** with the highest Silhouette Score of **0.4859**.

Cluster Visualization

To visualize the clustering results, we used **Principal Component Analysis (PCA)** to reduce the dimensionality of the customer data to two components. This allowed us to plot the clusters on a 2D plane, making it easier to interpret the results.

Cluster Scatterplot with Boundaries

Here is the 2D visualization of the clusters after applying PCA. The decision boundaries are drawn to show the separation between clusters.



The scatterplot indicates that the customers have been grouped into two main clusters based on their transaction history and profile data. The decision boundaries show clear distinctions between the two clusters.

Insights and Recommendations :

- The segmentation reveals two distinct customer groups:
  - **Cluster 0**
  - **Cluster 1**
- These insights can be used for:
  - Targeted marketing strategies.
  - Personalized product recommendations.
  - Improved customer service.

Conclusion

The clustering analysis has successfully segmented customers into distinct groups, providing valuable insights into customer behavior. This information can be used for targeted strategies in marketing, recommendations, and customer engagement. The results also provide a foundation for future enhancements and more detailed analysis.