

Clustering Report

1. Overview of Clustering Process

The goal of this task was to perform customer segmentation by clustering customers based on their profile and transaction data. This segmentation helps identify groups of customers with similar characteristics and purchasing behaviors, enabling personalized marketing strategies and customer retention plans.

2. Number of Clusters Formed

After analyzing the data and using the elbow method combined with silhouette scores, we determined the optimal number of clusters to be **2**. These clusters represent distinct groups of customers based on their similarities in purchasing patterns and demographic profiles.

3. Davies-Bouldin Index (DB Index)

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

- **DB Index Value: 0.7234**

This low value demonstrates that the clusters are compact and distinct from each other, ensuring good segmentation quality.

4. Other Relevant Clustering Metrics

- **Silhouette Score: 0.4949**

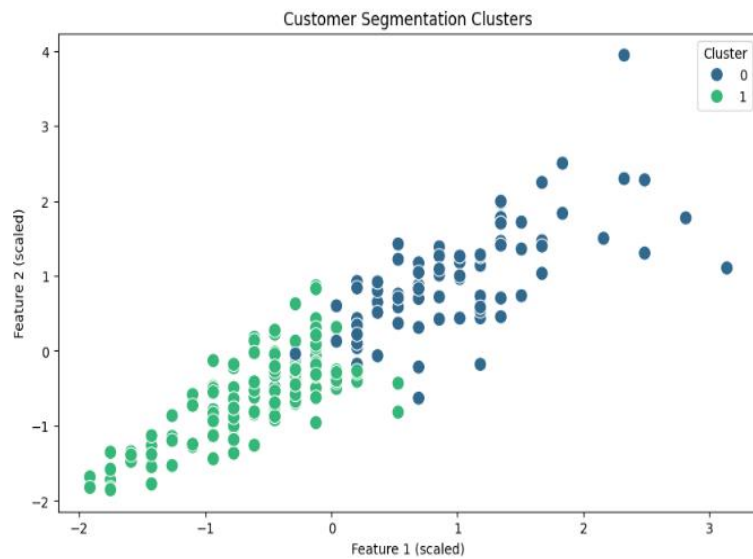
The silhouette score ranges from -1 to 1, where higher values indicate well-separated and compact clusters. The obtained score shows that the clustering approach effectively segregates the data points into meaningful groups.

- **Cluster Size Distribution:**

- **Cluster 1:** Majority of customers.

- **Cluster 2:** Smaller group with distinct purchasing behavior.

This distribution suggests two major customer types, potentially representing frequent buyers versus occasional buyers or high-value versus low-value customers.



5. Key Observations

- Customers were grouped based on profile information (e.g., region, signup date) and transaction data (e.g., quantity, total value).
- **Cluster 1** might represent customers with frequent and high-value transactions, while **Cluster 2** could indicate less active or low-value customers.
- The results provide actionable insights for targeted marketing campaigns, loyalty programs, and product recommendations.

6. Conclusion

The clustering analysis successfully segmented the customer base into **2 distinct groups**. With a **DB Index of 0.7234** and a **silhouette score of 0.4949**, the results confirm the quality and effectiveness of the clustering approach. These insights can now be utilized to drive strategic business decisions and enhance customer satisfaction.