

Customer Segmentation Report

1. Number of Clusters Formed:

*In this customer segmentation analysis, we applied a **kmeans** clustering algorithm to group customers into distinct clusters. After experimenting with different numbers of clusters, we decided on **4 clusters** as the optimal number.*

2. DB Index Value:

*The **Davies-Bouldin Index (DB Index)** for the clustering solution was calculated as **1.008**. The DB Index is used to evaluate the separation and compactness of clusters:*

- **DB Index \approx 1** indicates that the clusters are well-separated, but there is still some room for improvement.
- A lower value would indicate more compact and distinct clusters, whereas a higher value suggests that some clusters might be overlapping or not well-separated.

Given that the DB Index is near 1, it suggests a good balance between compactness and separation, but there is potential for further optimization (e.g., by adjusting the number of clusters or experimenting with other clustering algorithms).

3. Cluster Distribution (Cluster Counts):

The number of customers in each cluster is as follows:

- **Cluster 1:** 78 customers
- **Cluster 3:** 67 customers
- **Cluster 0:** 37 customers
- **Cluster 2:** 17 customers

These clusters vary in size, indicating different levels of segmentation within the customer base. It is important to note that some clusters are more densely populated than others, which could be indicative of varying customer behaviors.

4. Cluster Analysis:

Based on the features used for clustering (total spending, frequency of purchases, and recency), the clusters have been differentiated as follows:

- **Cluster 0** (Small group): Likely represents **recent customers** who have made only a few purchases but are highly active.
- **Cluster 1** (Large group): Appears to represent **high-frequency and high-spending customers**.
- **Cluster 2** (Small group): Consists of **low-frequency and low-spending customers**.
- **Cluster 3** (Large group): Includes customers who might be **mid-range in spending and frequency** but could have different product preferences.

5. Other Relevant Clustering Metrics:

- **Silhouette Score** (optional metric): A silhouette score helps assess how well-separated the clusters are, with values ranging from -1 to 1. A higher score indicates better clustering.
- **Cluster Characteristics**: By inspecting the average values for each cluster in terms of total spending, frequency, and recency, you can gain deeper insights into the specific characteristics of each customer segment.

6. Visual Representation:

The following visualizations were used to interpret the clusters:

- **Scatter Plot:** A scatter plot was created to visualize the clusters based on the features **total spending** and **frequency of purchases**. Different colors represent each cluster, giving a clear view of how the customers are grouped.
- **Cluster Distribution Bar Chart:** A bar chart was used to visualize the distribution of customers across the four clusters, showing the number of customers in each cluster.

7. Conclusion:

The customer segmentation analysis using KMeans with 4 clusters has resulted in reasonably well-separated clusters, as indicated by the DB Index value of 1.008. While the clusters are distinct, further improvements could be made by:

- Experimenting with different clustering algorithms (e.g., DBSCAN).
- Exploring more granular features that could improve cluster separation.

