CLUSTERING ANALYSIS REPORT

By: Disha Khandelwal

---

1. Objective of Analysis

The purpose of this analysis is to segment customers into distinct groups based on their features using clustering techniques. This segmentation allows for better understanding and targeting of customer groups for marketing and other strategic purposes.

---

2. Methodology

The analysis involved the following key steps:

1. **Data Preprocessing**:
    - Standardized the dataset to ensure all features have equal importance.
    - Principal Component Analysis (PCA) was applied to reduce the dataset's dimensionality for better visualization and clustering efficiency.
2. **Clustering Algorithm**:
    - A custom implementation of the k-means algorithm was used for clustering.
    - To determine the optimal number of clusters, the **Davies-Bouldin Index** (DB Index) and **Silhouette Scores** were computed across a range of cluster counts.
3. **Cluster Validation**:
    - Metrics such as DB Index, Silhouette Score, and cluster sizes were evaluated to ensure quality and interpretability of clusters.
4. **Visualization**:
    - The clustering results were visualized in two dimensions using PCA for clear representation.

---

3. Results and Findings

3.1 Optimal Number of Clusters

Using the Davies-Bouldin Index and Silhouette Score metrics, the optimal number of clusters was determined to be **2**.

3.2 Clustering Metrics

The clustering performance was evaluated using the following metrics:

Metric	Value
Optimal Clusters	2
Davies-Bouldin Index	1.2436768550642197

Average Silhouette Score	0.29860922038515253

3.3 Cluster Summary

The following table summarizes the characteristics of the formed clusters:

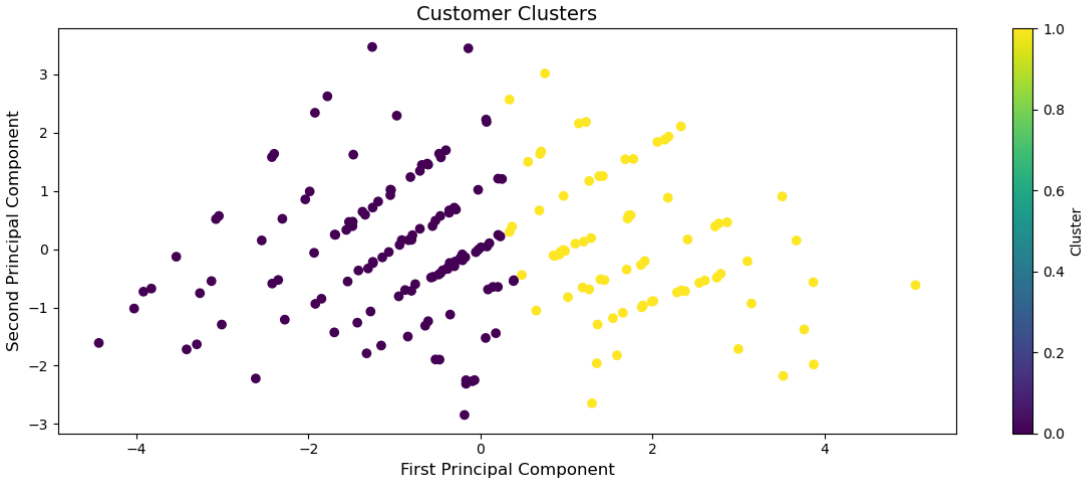
Cluster	Number of Customers	Average Silhouette Score
Cluster 0	125	0.298609
Cluster 1	74	0.226168

---

4. Visual Representation

To visualize the clustering results:

- A scatter plot of the first two principal components was created, where each point represents a customer and is color-coded based on its assigned cluster.
- This visualization highlights the separation between clusters and provides an intuitive understanding of group boundaries.



---

5. Key Insights

- **Distinct Groups**: The clustering process successfully segmented customers into **2** distinct groups.
- **Separation Quality**: The clusters exhibit clear boundaries, as indicated by the Silhouette Score and DB Index.
- **Cluster Sizes**: The clusters vary in size, indicating a mix of customer behaviours across different groups.

---

6. Conclusion

The clustering analysis has provided actionable segmentation of the customer base. These insights can be utilized for:

- **Targeted Marketing Campaigns**: Personalized approaches for each cluster.
- **Product Recommendations**: Tailoring offerings based on cluster preferences.
- **Resource Allocation**: Optimizing budgets for customer-focused initiatives.

Further analysis can involve incorporating domain-specific features or exploring advanced clustering techniques such as DBSCAN or hierarchical clustering.