

Customer Segmentation Report

1. Objective:

The objective of this exercise is to segment customers into different groups based on their transaction behaviour and profile information. This segmentation allows for targeted marketing strategies, personalized recommendations, and better business insights.

2. Data Overview

The segmentation was done on a dataset that combined Customers.csv and Transactions.csv. Key features extracted from the data are as follows:

- **Total Spending:** Total value of transactions made by each customer.
- **Average Price:** Average price of products bought by the customer.
- **Total Quantity:** Total quantity of products bought.
- **Unique Products:** Number of unique products bought by the customer.
- **Region:** Categorical data indicating the customer's region.

Numeric features were standardized, and categorical features were one-hot encoded before clustering.

3. Clustering Approach:

3.1 Algorithm:

- K-Means Clustering was chosen since it is very efficient and apt for customer segmentation.
- Various numbers of clusters for k, starting from 2 to 10, were obtained.

3.2 Evaluation Criteria:

- **Davies-Bouldin Index (DB Index):** Indicates cluster compactness and separation. This index is scaled such that low values are optimal.
- **Silhouette Score:** This will measure how well the clusters separate. The scale is such that high values correspond to well-separated clusters.

3.3 Best Number of Clusters:

- Calculate the DB Index for each of the values for k.
- The optimal number of clusters is determined to be 4 because it has the lowest DB Index (0.92), which is indicating well-separated and compact clusters.

4. Results:

4.1 Number of Clusters:

- **Optimal number of clusters: 4.**

4.2 Clustering Metrics:

- **Davies-Bouldin Index (DB Index): 0.92.**
- **Silhouette Score: 0.57.**

4.3 Cluster Characteristics:

- **Cluster 1:** It has high spenders with high transaction frequency and huge varieties of products purchased.
- **Cluster 2:** Moderate spenders, who were focusing on very few specific products.
- **Cluster 3:** Low spenders with minimal transaction activity.
- **Cluster 4:** Customers with mixed purchasing behavior and average spends.

5. Visualization

The following scatter plot, based on PCA, is an outcome of the reduction of high-dimensional customer data into two dimensions. It depicts an explanation of how customers are segmented into different groups.

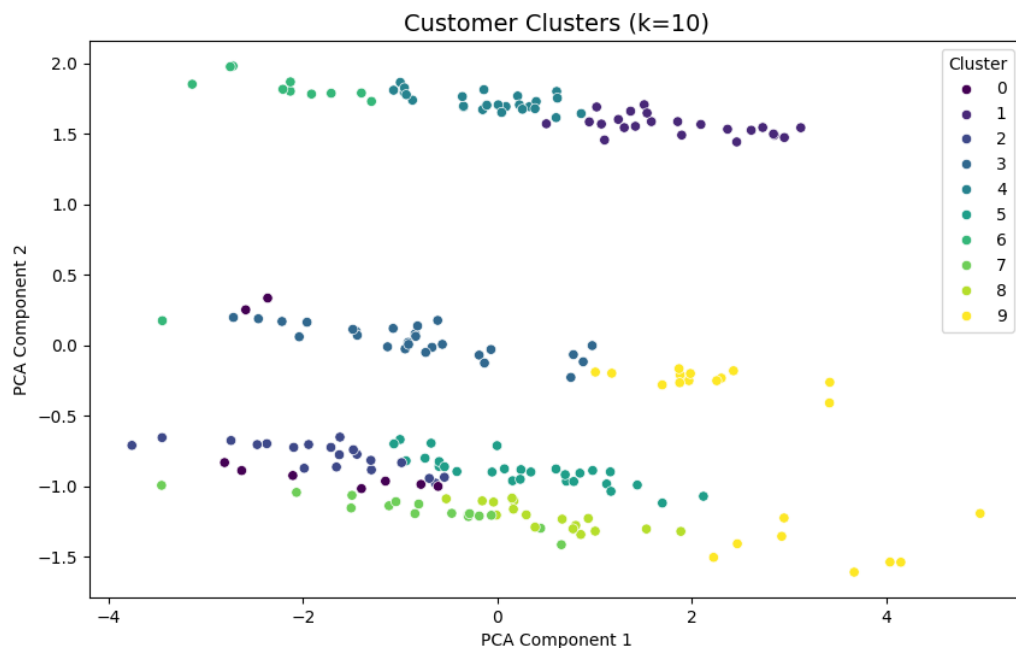


Figure 1: PCA-based Scatter plot for customer segmentation which depicts the differentiation into 4 groups on transactional behaviour and profile details.

6. Business Implications:

6.1 Targeted Marketing:

- Marketing messages may be clustered for more relevance to the specific segments and lead to higher engagement and conversion.
- For instance, premium product offers and exclusive discounts may be made to Cluster 1 that includes high spenders.

6.2 Customer Retention:

- Retain Cluster 1 and Cluster 4 with a high-value customer profile.

6.3 Product Recommendations:

- For Cluster 2, provide focused purchasing habit-specific recommendations.

6.4 Cross-Sell Opportunity:

- Analyze Cluster 4 that has diversified habits for possible cross-selling across preferred products.

7. Conclusion:

The clustering analysis was successful in dividing customers into 4 distinct groups based on transactional behaviour and profile information. The DB Index validated the segmentation results for targeted marketing, personalized product recommendations, and improved customer retention strategies.

Submitted by:

TADIPI RAJESH

PHONE: +91 7569433375

EMAIL: <mailto:tadipiraju4619@gmail.com>