

Customer Segmentation Clustering Report

Introduction

The objective of this analysis is to perform customer segmentation using clustering techniques, based on both customer profile information and transaction behavior. By clustering customers into distinct segments, businesses can tailor marketing strategies and optimize resources for better customer engagement and retention.

Clustering Algorithm

We used the KMeans clustering algorithm to segment the customers based on their transaction behavior (TotalValue, Quantity, and number of unique TransactionDates). KMeans is a widely used algorithm that groups customers into clusters based on the similarity of their features.

Number of Clusters

The number of clusters, k , was determined by evaluating clustering metrics, specifically the Davies-Bouldin Index, to ensure optimal segmentation.

Feature Engineering

To prepare the data for clustering:

1. Merged the customer and transaction data on the CustomerID.
2. Aggregated the transaction data to create a summary per customer, including the total transaction value, total quantity of items purchased, and the number of unique transaction dates.
3. Standardized the features to bring them to a similar scale, ensuring the clustering algorithm treats them equally.

Evaluation Metric

We used the Davies-Bouldin Index (DB Index) to evaluate the clustering performance. The DB Index measures the compactness and separation of clusters. A lower DB Index indicates better clustering performance.

Clustering Results

Final Davies-Bouldin Index

- Value: 0.865
- The Davies-Bouldin Index for $k=4$ is 0.865, indicating good clustering with well-separated groups.

Number of Clusters

- Value: 4
- The algorithm formed 4 distinct clusters based on customer transaction behavior and profile information.

Cluster Summary

The following table summarizes the key characteristics of the 4 clusters formed:

Cluster	Avg Total Value	Avg Quantity	Avg Transaction Date
0	6263.45	23.00	8.43
1	1273.37	5.27	2.36
2	2982.41	10.87	4.39
3	4477.57	16.10	6.31

- Cluster 0: High-value customers who make frequent and high-volume purchases. These customers represent the top tier of spending.
- Cluster 1: Low-value customers who make fewer purchases and have lower overall transaction values.
- Cluster 2: Moderate-value customers with a reasonable number of transactions and moderate total value.
- Cluster 3: Moderate-value customers who make frequent transactions but with lower transaction values compared to Cluster 0.

Visualization

To visualize the clusters, we reduced the data to two dimensions using Principal Component Analysis (PCA). The resulting 2D scatter plot provides a clear view of how the customers are distributed across the clusters:

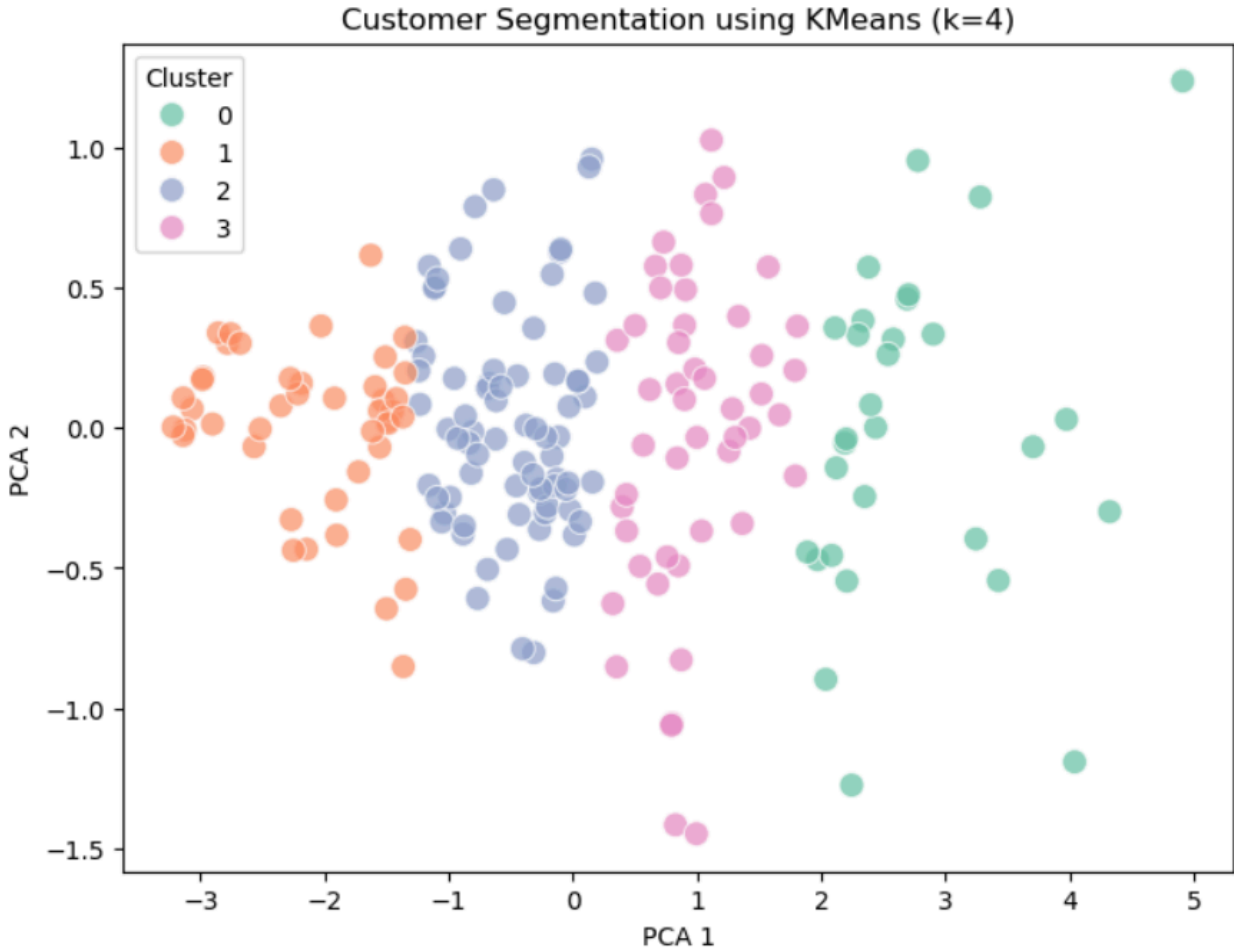


Figure 1: Scatter plot showing customer segmentation after dimensionality reduction using PCA.

Conclusion

Clustering Insights:

1. Cluster 0 represents high-value, high-frequency customers who contribute significantly to the revenue.
2. Cluster 1 represents low-value customers, who make fewer transactions and have lower total spending.
3. Cluster 2 includes moderate-value customers who display a balance between total spending and transaction frequency.
4. Cluster 3 consists of customers with moderate total spending, though they have a higher frequency of transactions compared to Cluster 1 and Cluster 2.

Evaluation Metrics:

- Davies-Bouldin Index: 0.865, which indicates good separation and compactness between the clusters.

Recommendations:

1. Cluster 0 (High-Value Customers):
 - Develop loyalty programs to retain and nurture these customers.
 - Offer personalized promotions and rewards based on transaction history.
2. Cluster 1 (Low-Value Customers):
 - Consider targeted campaigns to encourage more frequent purchases.
 - Offer discounts or incentives to increase engagement.
3. Cluster 2 & 3 (Moderate-Value Customers):
 - Implement loyalty schemes to increase their spending and frequency.
 - Design marketing strategies to move these customers toward higher spending behaviors.