February 1, 2025

Customer Segmentation / Clustering Report

Presented by

Om Wanjari

ZEOTAP

Table of Contents

1. Overview

This report presents the results of customer segmentation performed on an eCommerce transactions dataset. The goal was to group customers into distinct clusters based on their transaction behavior and profile information. The segmentation was performed using the K-Means clustering algorithm, and the results were evaluated using the Davies-Bouldin Index (DB Index). The analysis provides actionable insights to improve customer engagement and business strategies.

2. Dataset Description

The dataset consists of three files:

1. **Customers.csv:** Contains customer profile information (CustomerID, CustomerName, Region, SignupDate).
2. **Transactions.csv:** Contains transaction details (TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, Price).
3. **Products.csv:** Contains product information (ProductID, ProductName, Category, Price).

For this analysis, we focused on merging Customers.csv and Transactions.csv to create a customer-level dataset with features such as total spending, average transaction value, transaction count, and unique products purchased.

3. Methodology

1.Data Preprocessing:

- Merged Customers.csv and Transactions.csv on CustomerID.
- Engineered customer-level features:
  - total_spent: Total amount spent by the customer.
  - avg_transaction_value: Average transaction value.
  - transaction_count: Number of transactions made.
  - unique_products: Number of unique products purchased.
  - signup_year: Year the customer signed up.

2.Clustering:

- Applied K-Means clustering to segment customers.
- Used the Elbow Method and Davies-Bouldin Index to determine the optimal number of clusters (k=4).

3.Evaluation:

- Calculated the Davies-Bouldin Index to evaluate clustering performance.
- Visualized clusters using PCA for dimensionality reduction.

4. Results

1.Clustering Metrics

Number of Clusters: 4

Davies-Bouldin Index: 1.157

(A lower value indicates better clustering; this score suggests reasonably well-separated clusters.)

2. Cluster Characteristics

The customers were grouped into 4 clusters based on their transaction behavior. Below is a summary of each cluster:

Cluster	Total Spent (Mean)	Avg Transaction Value (Mean)	Transaction Count (Mean)	Unique Products (Mean)	Signup Year (Mode)
0	$5,599.40	$744.83	7.66	7.34	2022
1	$1,793.06	$518.51	3.31	3.28	2022
2	$3,232.40	$592.01	5.52	5.46	2024
3	$3,351.86	$971.55	3.43	3.33	2024

3. Cluster Interpretation

Cluster 0: High-Value Customers

- Characteristics: Highest total spending, highest average transaction value, most transactions, and most unique products purchased.
- Insight: These are your most loyal and valuable customers. Focus on retaining them through loyalty programs or exclusive offers.

Cluster 1: Low-Value Customers

- Characteristics: Lowest total spending, lowest average transaction value, fewest transactions, and fewest unique products purchased.
- Insight: These are likely occasional or new customers. Encourage them to make more frequent purchases through targeted promotions or discounts.

Cluster 2: Mid-Value Customers

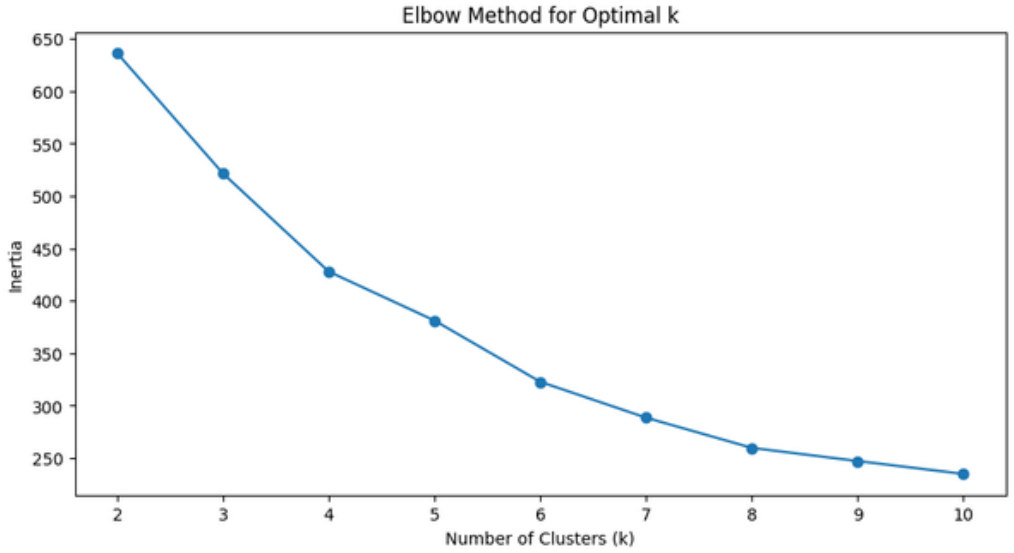
- Characteristics: Moderate total spending, moderate average transaction value, moderate transaction count, and moderate unique products purchased.
- Insight: These are regular but not highly engaged customers. Engage them with personalized recommendations to increase their spending.

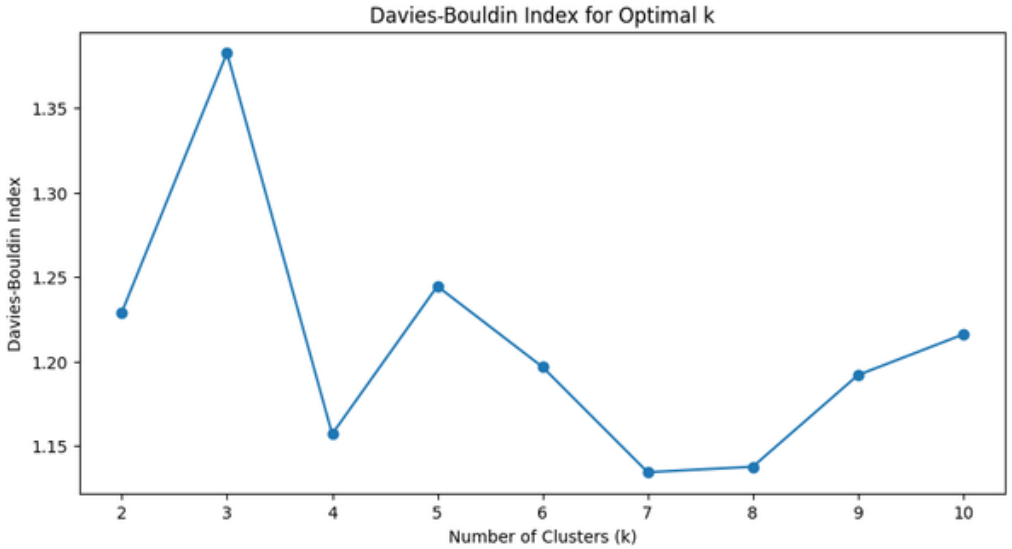
Cluster 3: High-Transaction Customers

- Characteristics: High average transaction value but moderate total spending, fewer transactions, and moderate unique products purchased.
- Insight: These customers make fewer but high-value purchases. Offer bundled deals or upsell opportunities to increase their transaction frequency.

4. Visualization

The clusters were visualized using PCA to reduce the data to 2 dimensions. The scatter plot below shows the separation of clusters:




Davies-Bouldin Index for Optimal k
Davies-Bouldin Index
1.35
1.30
1.25
1.20
1.15
2
3
4
5
6
7
8
9
10
Number of Clusters (k)

5. Actionable Insights

Retain High-Value Customers (Cluster 0):
Implement loyalty programs or exclusive offers to keep these customers engaged.
Provide personalized recommendations to encourage repeat purchases.

Engage Low-Value Customers (Cluster 1):
Offer discounts or promotions to increase their transaction frequency.
Use email marketing to highlight popular products or limited-time deals.

Increase Spending of Mid-Value Customers (Cluster 2):
Send personalized product recommendations based on their purchase history.
Introduce tiered pricing or volume discounts to encourage larger purchases.

Upsell to High-Transaction Customers (Cluster 3):
Offer bundled deals or complementary products to increase transaction frequency.
Highlight premium products or services to encourage higher spending.

6. Conclusion

The customer segmentation analysis successfully grouped customers into 4 distinct clusters based on their transaction behavior. The clusters provide valuable insights into customer preferences and spending patterns, enabling targeted marketing strategies to improve customer engagement and drive revenue growth.