

Customer Segmentation Analysis Report

Introduction

This report presents the results of a customer segmentation analysis performed using clustering techniques based on demographic and transaction data from an e-commerce dataset. The goal of this analysis is to identify distinct customer segments that can inform targeted marketing strategies and enhance customer engagement.

Data Overview

The analysis utilized two datasets:

- **Customers.csv:** Contains demographic information about customers, including CustomerID, CustomerName, Region, and SignupDate.
 - **Transactions.csv:** Contains transaction details, including TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, and Price.
-

Methodology

1. **Data Preparation:** The datasets were merged to create a comprehensive profile for each customer, aggregating transaction data to derive key features.
 2. **Feature Engineering:** Key features were created, include:
 - **Total Spend**
 - **Total Quantity Purchased**
 - **Number of Transactions**
 - **Average Transaction Value**
 - **Average Price of Products Purchased**
 3. **Clustering Algorithm:** K-Means clustering was employed to segment customers into distinct groups.
 4. **Evaluation Metrics:** The Davies-Bouldin Index (DB Index) was calculated to assess the quality of the clusters.
-

Clustering Results

1. Number of Clusters Formed

- Optimal Number of Clusters: 8
- The analysis determined that segmenting customers into 8 clusters provides the most meaningful differentiation among customer groups.

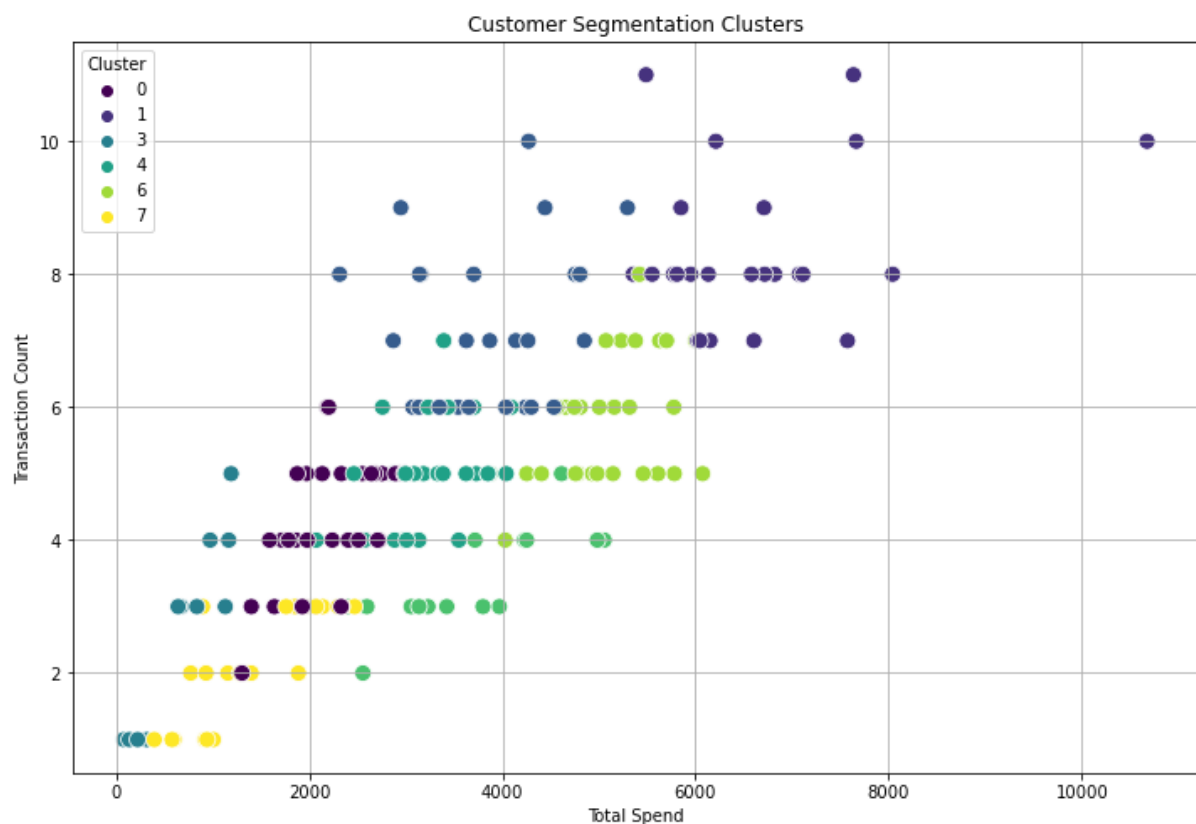
2. Davies-Bouldin Index

- DB Index Value for 8 Clusters: 0.9856
 - The DB Index is a measure of cluster separation; lower values indicate better-defined clusters. A value of approximately 0.9856 suggests that while the clusters are reasonably well-separated, there may still be some overlap among them.
-

Visualizations

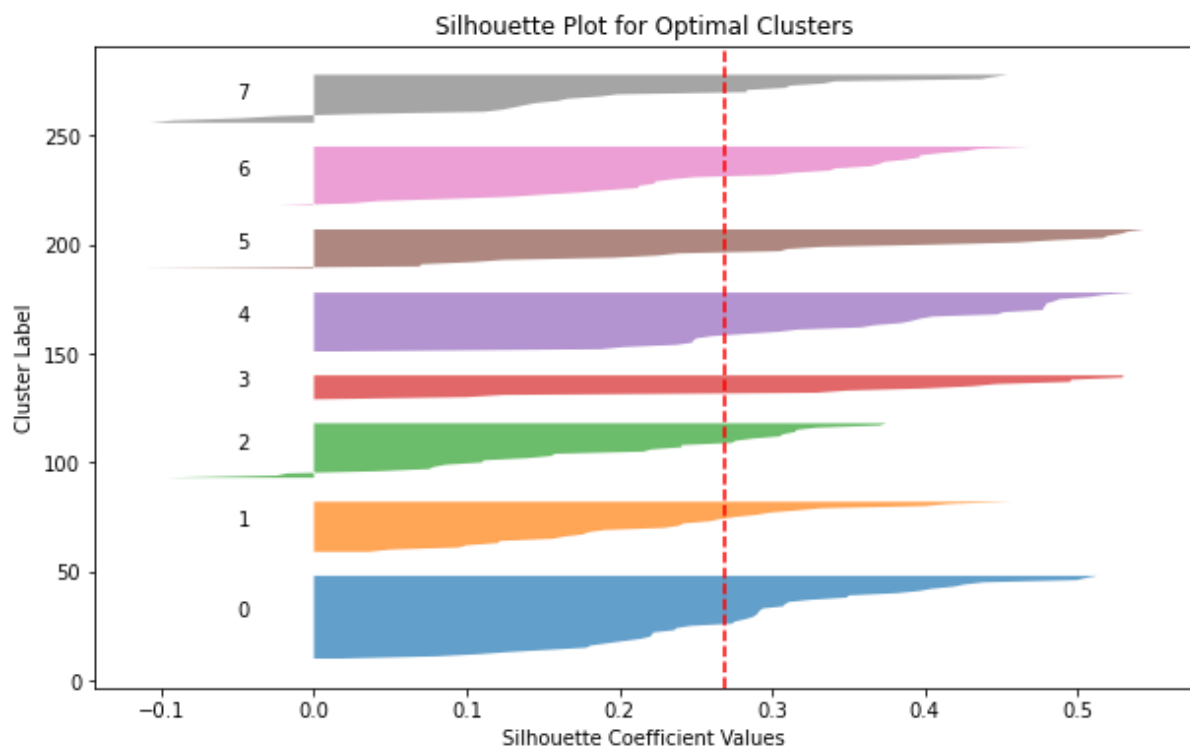
1. Customer Segmentation Clusters Plot

- This plot visualizes the distribution of customers across the identified clusters based on Transaction Count and Total Spend.



2. Silhouette Plot for Optimal Clusters

- The silhouette plot illustrates the silhouette scores for each cluster, providing insight into how well-separated the clusters are.



Conclusion

The clustering analysis has successfully identified 8 distinct customer segments within the dataset. The Davies-Bouldin Index value of approximately 0.986 suggests that while the clusters are reasonably well-separated.

This report provides a foundational understanding of customer segments that can drive more effective marketing strategies and enhance customer satisfaction through personalized experiences.