Report

Introduction

This report presents the results of a customer segmentation analysis performed using clustering techniques on customer profile and transaction data. The goal of this analysis is to identify meaningful customer segments that can drive targeted marketing strategies and enhance customer engagement.

Data Overview

Datasets Used

- **Customers.csv**: Contains demographic and profile information about customers.
- **Transactions.csv**: Includes transaction history, such as purchase amounts and dates.
- **Products.csv**: Contains information about products purchased, which can be used for further analysis if needed.

Data Preparation

The datasets were merged to form a comprehensive view of customer transactions. The following features were engineered for analysis:

- **TotalSpent**: The total amount spent by each customer.
- **NumTransactions**: The total number of transactions made by each customer.
- **DaysSinceLastPurchase**: The number of days since the last transaction by each customer.

Methodology

Clustering Algorithm

We chose the **Agglomerative Clustering** algorithm, a hierarchical clustering method. This algorithm does not require the number of clusters to be predefined, making it flexible for identifying clusters in the data.

Feature Standardization

To ensure equal contribution of each feature during the clustering process, we standardized the features (TotalSpent, NumTransactions, and DaysSinceLastPurchase) using **StandardScaler**.

Optimal Number of Clusters

The optimal number of clusters was determined using two clustering evaluation metrics:

- **Silhouette Score**: Measures the cohesion and separation of the clusters.
- **Davies-Bouldin Index (DB Index)**: Measures the average similarity ratio of each cluster with the cluster that is most similar to it.

The optimal number of clusters was chosen as **4** based on these metrics.

Clustering Results

Number of Clusters Formed

- **Total Clusters**: 4

Davies-Bouldin Index

- **DB Index Value**: 1.0084
  A lower DB index suggests well-separated clusters. The value of **1.0084** indicates reasonable cluster separation.

Additional Clustering Metrics

- **Silhouette Scores**: Evaluated for a range of cluster counts (2 to 10), with the optimal cluster count selected based on the highest silhouette score.

Cluster Characteristics

The following table summarizes the average characteristics of each customer cluster:

Cluster	Average Total Spent	Average Number of Transactions	Average Days Since Last Purchase
0	2447.05	3.88	94.08
1	4569.27	6.13	109.27
2	712.30	1.31	293.69
3	6005.06	8.40	51.17

Marketing Strategies Based on Clusters

Based on the characteristics of each cluster, the following marketing strategies have been proposed:

Cluster 0 (High-value customers):
Strategy: Focus on exclusive offers and loyalty programs to retain and further engage these customers.

Cluster 1 (Moderate-value customers):
Strategy: Promote upselling and cross-selling opportunities to increase their average transaction value.

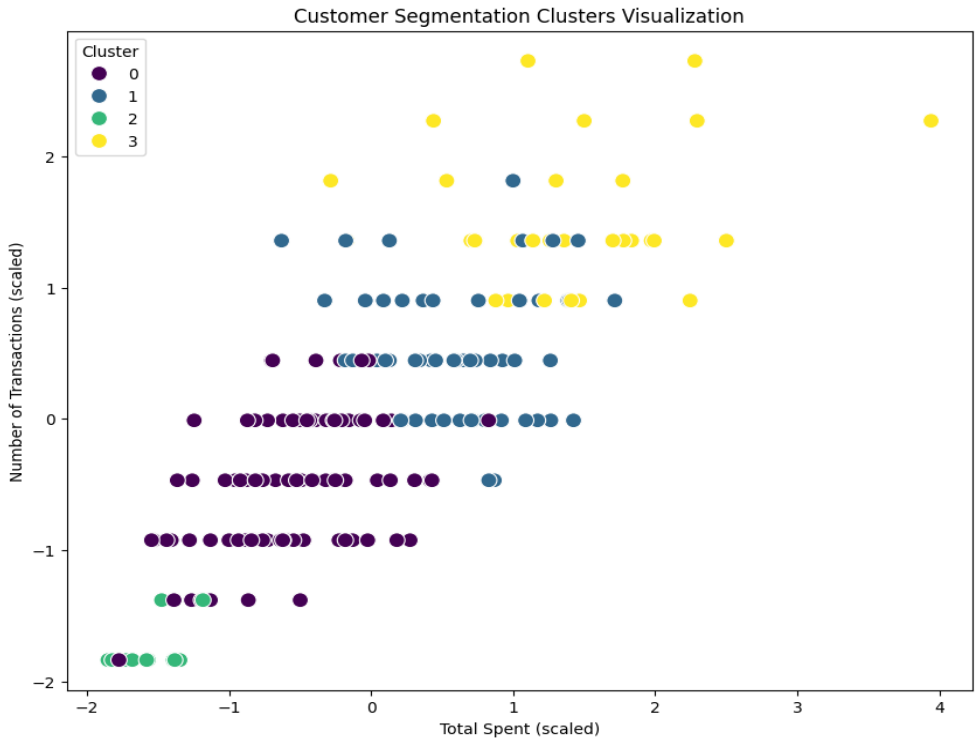
Cluster 2 (Infrequent low-value customers):
Strategy: Re-engagement strategies needed. Focus on bringing these customers back with personalized offers and incentives.

Cluster 3 (Frequent moderate-value customers):
Strategy: Implement onboarding campaigns and promotions to enhance their experience and encourage higher spending.

Visualizations

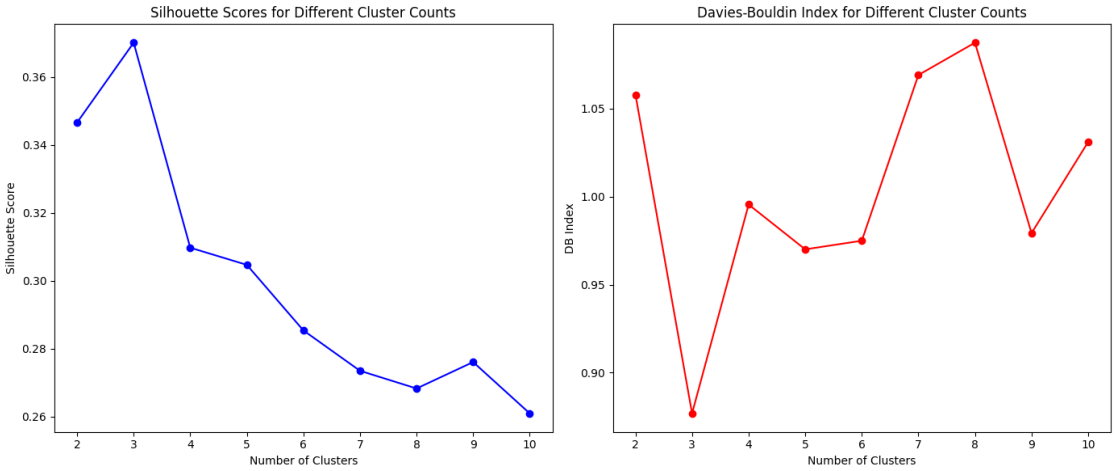
Cluster Visualization

A scatter plot was created to visualize the customer segments based on **TotalSpent** and **NumTransactions**, providing an intuitive understanding of the distribution across different clusters.



Silhouette Scores and DB Index

Plots of the **Silhouette Scores** and **Davies-Bouldin Index** for different cluster counts were generated, aiding in the selection of the optimal number of clusters.



Conclusion

The customer segmentation analysis successfully identified four distinct customer segments. The **Davies-Bouldin Index** indicates that the clusters are well-separated, and the proposed marketing strategies based on cluster characteristics can help improve customer engagement and drive sales.

Key Insights:

- Customers can be grouped into four distinct segments based on their spending behavior, transaction frequency, and recency.
- Marketing strategies have been tailored for each cluster to maximize engagement and retention.