Customer Segmentation Using Clustering

Overview

We performed customer segmentation using clustering techniques to group customers based on their transaction behavior and regional information. This analysis provides actionable insights into customer patterns, enabling better-targeted marketing strategies and customer retention efforts.

Key Steps in Analysis

1. Data Preparation

- Merged the Customers.csv and Transactions.csv datasets.

- Engineered features for each customer:

    - **TotalSpent**: Total revenue generated by the customer.

    - **AvgTransactionValue**: Average value of transactions.

    - **TotalTransactions**: Number of transactions completed by the customer.

- Added one-hot encoded region information (e.g., Region_Asia, Region_Europe).

2. Clustering Analysis

- **StandardScaler**: Normalized all features to ensure equal weighting during clustering.

- **KMeans Clustering**: Used the KMeans algorithm to cluster customers.

3. Cluster Evaluation

- Evaluated clusters for optimal performance using:

    - **Elbow Method**: To determine the point of diminishing returns in inertia (cluster compactness).

    - **Davies-Bouldin Index (DB Index)**: Lower values indicate better-defined clusters.

    - **Silhouette Score**: Higher scores indicate better-defined and separated clusters.

Key Results

Optimal Number of Clusters

Based on the metrics, the optimal number of clusters was determined to be **4**.

Cluster Metrics

- **Davies-Bouldin Index:** 1.24 (lower values indicate well-separated clusters).

- **Silhouette Score:** 0.29 (higher values indicate better-defined clusters).

Cluster Characteristics

1. **Cluster 0**: Customers with high total spending but fewer transactions (e.g., bulk purchases).

2. **Cluster 1**: Frequent buyers with moderate average transaction values.

3. **Cluster 2**: Low spenders with minimal transaction activity.

4. **Cluster 3**: High-value customers with frequent and high-value transactions.
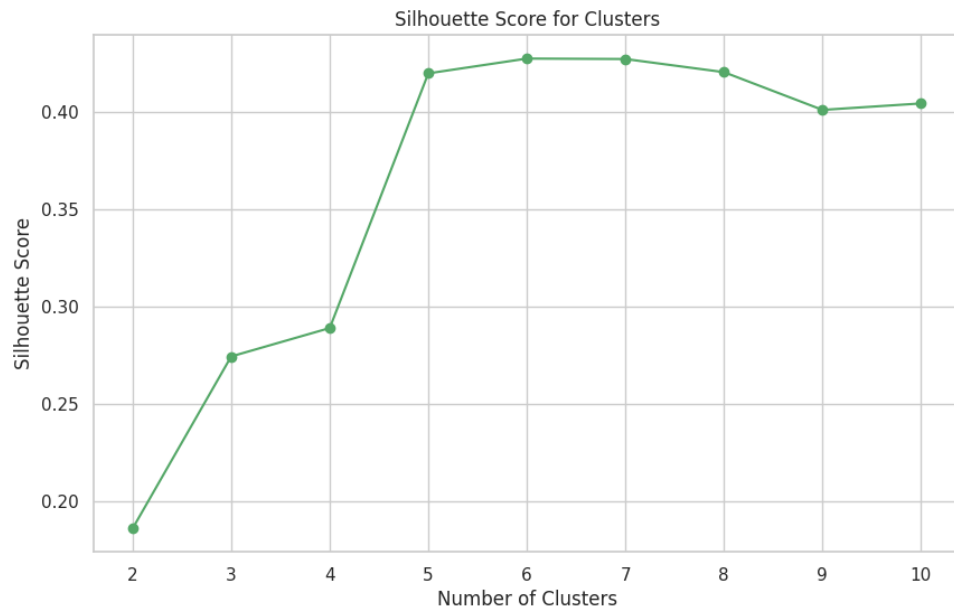
Visualizations

1. Elbow Curve
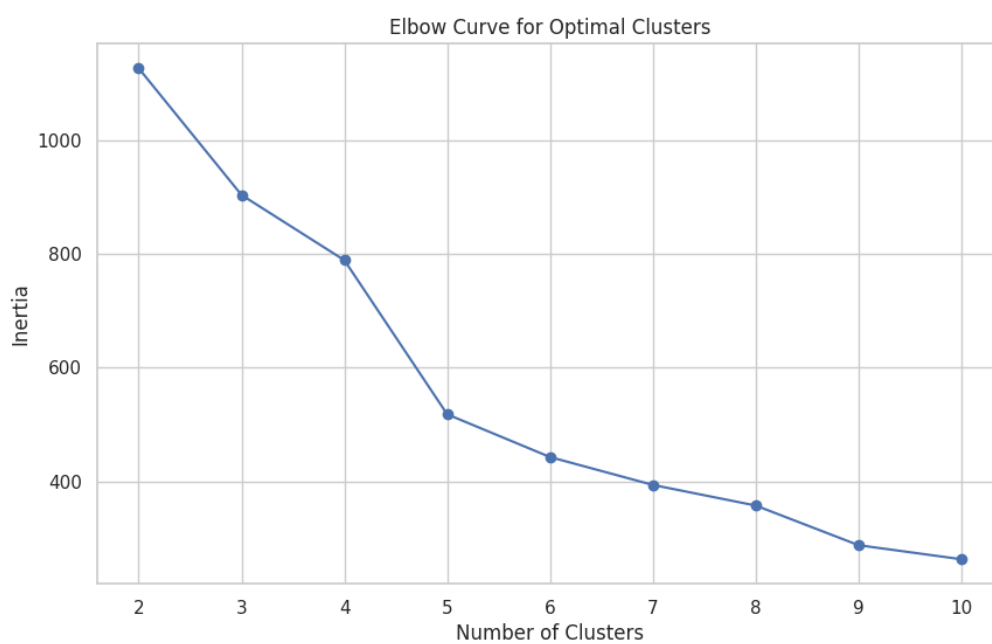
2. Davies-Bouldin Index

3. Silhouette Score
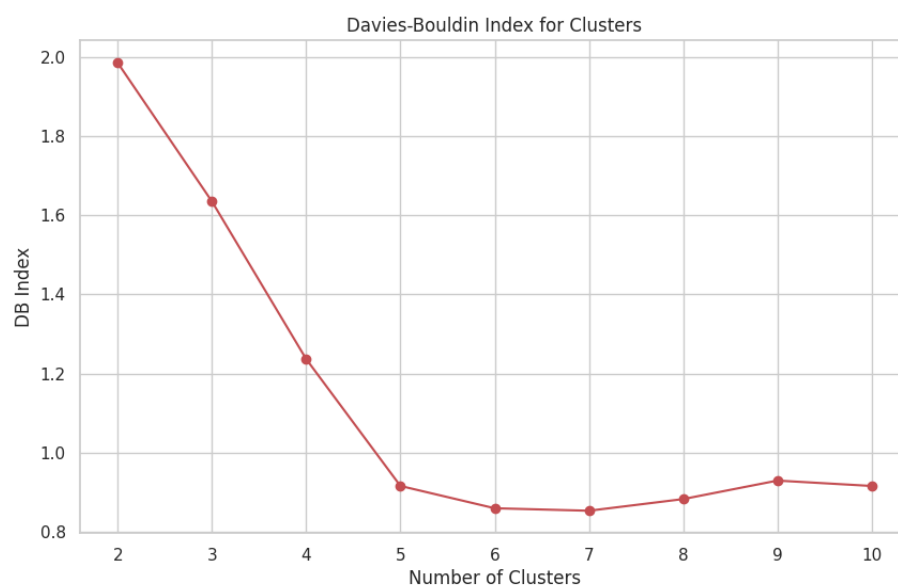
Silhouette Score for Clusters

4. Cluster Visualization

The scatter plot below shows customer segmentation based on TotalSpent and AvgTransactionValue:

Customer Segmentation Clusters (Total Spent vs. Avg Transaction Value)