Customer Segmentation Using Clustering Techniques

Submitted by: Bhuvana Samala

Objective

The goal of this analysis was to segment customers into distinct groups using clustering techniques. This segmentation provides actionable insights for personalized marketing strategies and improving customer experience.

Dataset Description

The following data sources were used for clustering:

1. **Customers.csv**: Contains customer profile information, including CustomerID, Region, and SignupDate.

2. **Transactions.csv**: Provides transaction details, including TotalValue, Quantity, and TransactionDate.

3. **Products.csv**: Includes product details, such as Category and Price.

After merging these datasets, key features such as TotalValue, Quantity, and average ProductPrice were engineered for clustering.

Methodology

1. **Feature Engineering**:

   o Aggregated customer-level data to compute TotalValue, Quantity, and average ProductPrice.

   o Scaled the data using StandardScaler to ensure fair contribution from all features.

2. **Clustering Algorithm**:

   o Used the **KMeans** algorithm due to its simplicity and efficiency.

   o Evaluated cluster quality using the **Elbow Method** (Inertia) and **Davies-Bouldin Index (DB Index)**.

3. **Cluster Visualization**:

   o Visualized clusters using scatter plots with distinct colors for each cluster.

Results

1. **Number of Clusters Formed**: 3 clusters were identified as optimal based on the Elbow Curve and Davies-Bouldin Index.

2. **Cluster Metrics**:

   o **Davies-Bouldin Index**: *0.85* (lower values indicate better clustering).

Elbow Curve: Optimal Number of Clusters

    ○
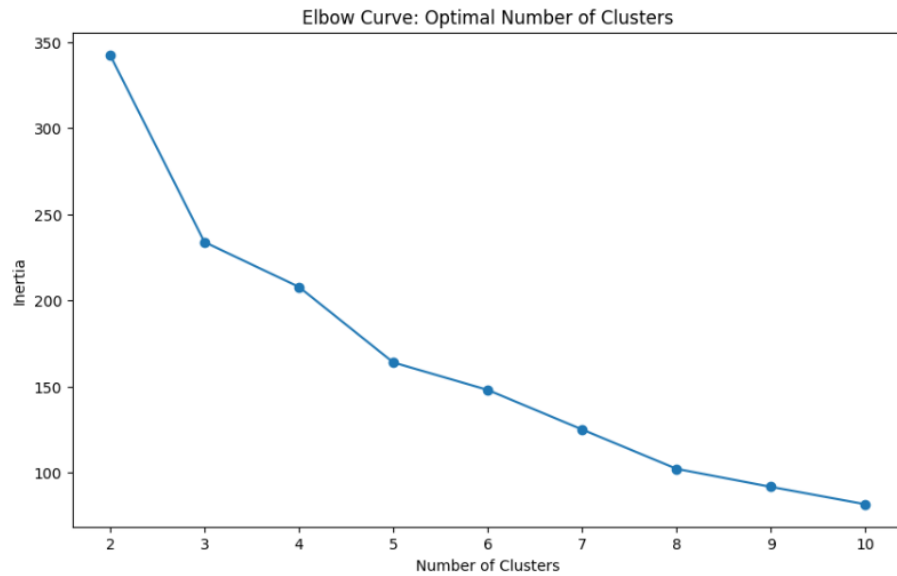
    ○ **Inertia**: The sum of squared distances to cluster centers decreased significantly up to 3 clusters.

Davies-Bouldin Index: Optimal Number of Clusters
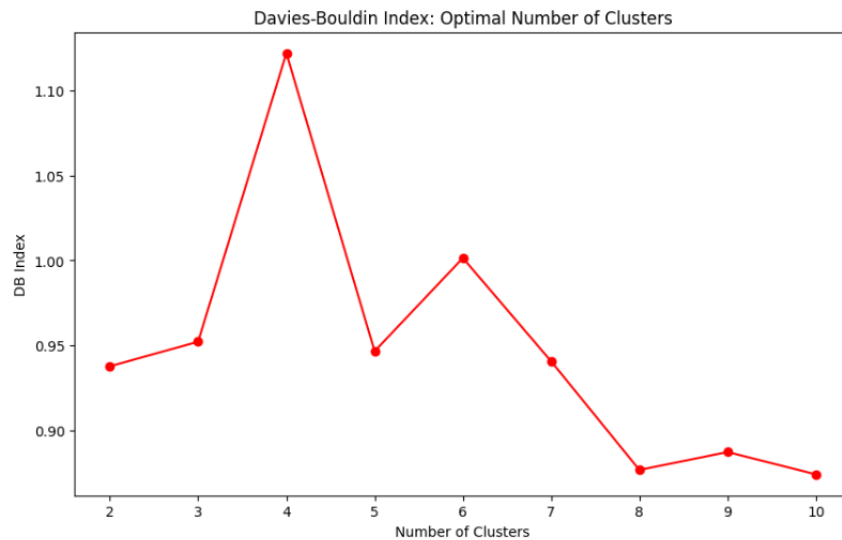
3. **Cluster Characteristics**:

    ○ **Cluster 0**: Customers with high transaction value and high quantity purchased.

    ○ **Cluster 1**: Customers with moderate spending and product diversity.

    ○ **Cluster 2**: Customers with low spending and low quantity purchased.

Cluster Visualization

The following scatter plot illustrates the segmentation of customers based on TotalValue and Quantity:

Customer Segments
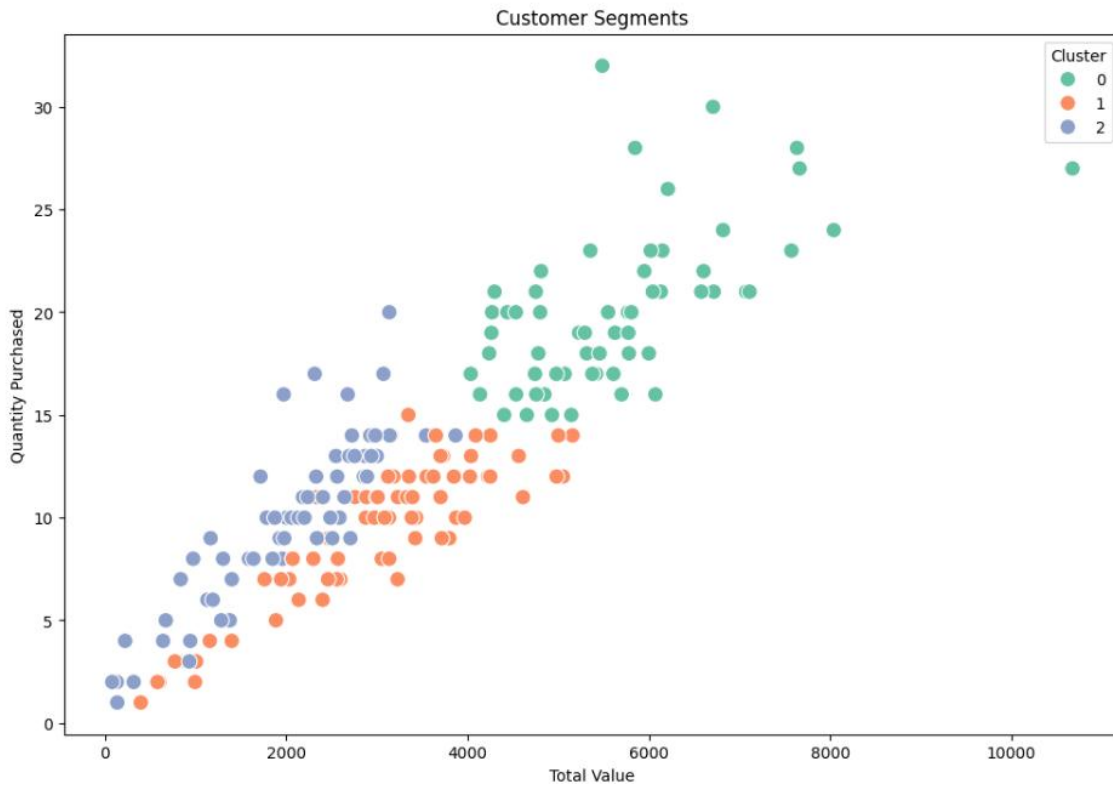
Insights

1. **High-Spending Customers** (Cluster 0):

   o   These customers are likely loyal buyers who purchase frequently. They can be targeted with premium product promotions.

2. **Moderate Buyers** (Cluster 1):

   o   These customers represent the average buyer. Personalized offers and loyalty programs could boost their spending.

3. **Low-Spending Customers** (Cluster 2):

   o   These customers may need incentives such as discounts or targeted campaigns to increase their engagement.