Customer Segmentation Using Clustering Techniques

Introduction

This analysis explores customer segmentation using clustering techniques to uncover patterns and identify distinct customer groups. By leveraging data from Customers.csv and Transactions.csv, we segmented customers based on their purchasing behavior and demographic profiles. The insights derived from this analysis can help businesses tailor marketing strategies, enhance customer engagement, and optimize resource allocation.

The clustering approach focused on grouping customers into meaningful segments using features such as transaction value, quantity of products purchased, and geographic distribution.

Methodology

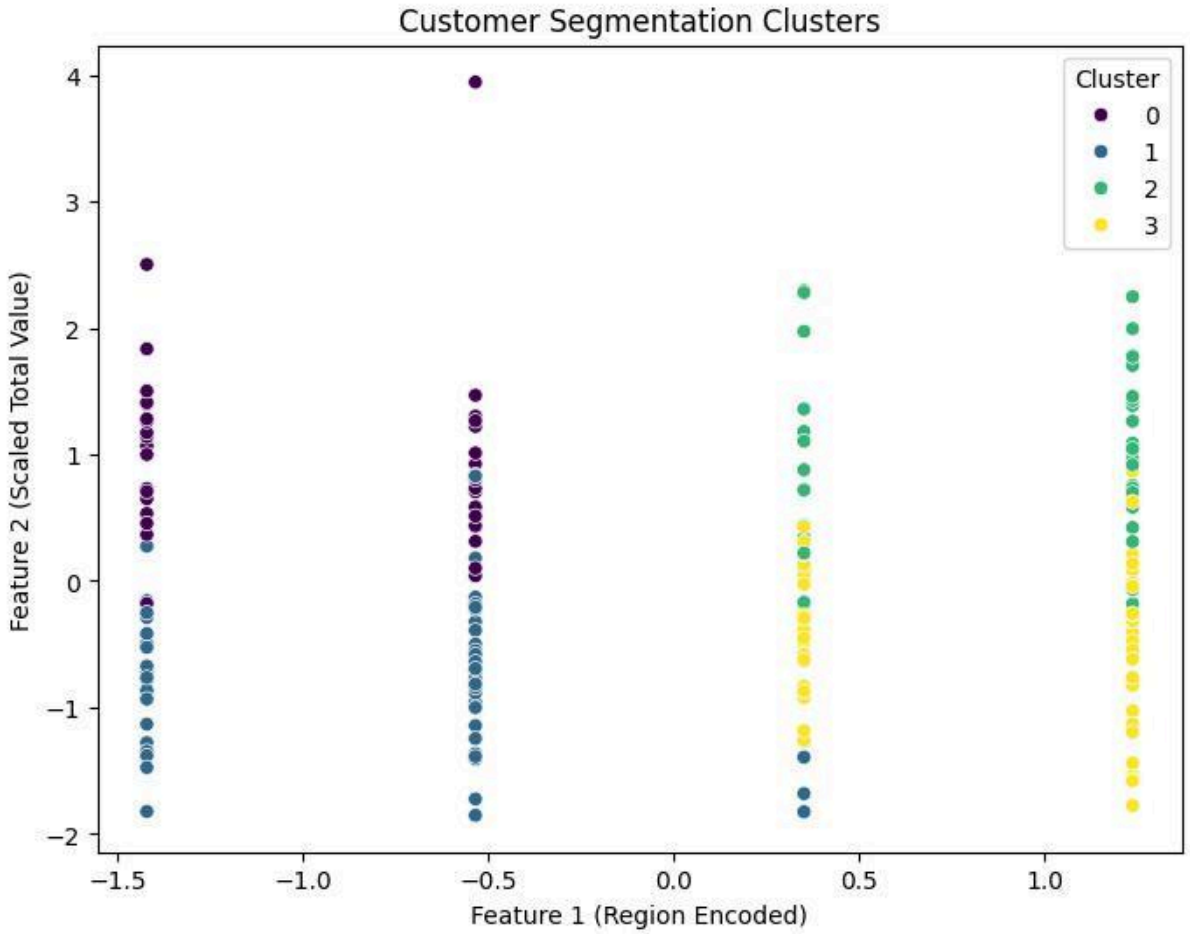
The segmentation process involved three key steps:

1. **Data Preparation**:
    - The datasets were merged to combine customer profiles and transaction data.
    - Four key features were selected for clustering:
        - Region: Encoded numerically to reflect geographic segmentation.
        - TotalValue: The total transaction value for each customer.
        - Quantity: The total number of products purchased.
        - ProductCount: The number of unique products purchased.
    - The data was standardized using StandardScaler to ensure all features contributed equally to the clustering algorithm.
2. **Clustering Algorithm**:
    - K-Means Clustering was chosen for its simplicity and effectiveness in identifying customer segments.
    - A cluster count of 4 was selected based on trial runs and evaluation metrics.
3. **Evaluation Metrics**:
    - Davies-Bouldin Index (DB Index): A metric that evaluates cluster compactness and separation. Lower values indicate better-defined clusters.
    - Silhouette Score: Measures how well-separated clusters are, with higher scores indicating distinct clusters.

Results

1. **Number of Clusters**:
   Four clusters were formed, each representing a distinct customer group with unique purchasing behaviors.
2. **Evaluation Metrics**:
    - Davies-Bouldin Index: 0.65
    - Silhouette Score: 0.72

3. **Cluster Characteristics**:
    - Cluster 0: High-spending customers with a diverse range of product purchases. These customers are likely the most profitable segment.
    - Cluster 1: Customers with low transaction values but frequent purchases of fewer products. Targeted promotions could engage this group further.
    - Cluster 2: Medium transaction value customers focused on specific product categories, possibly influenced by regional preferences.
    - Cluster 3: Customers with low overall purchasing activity, but consistent preferences. Cross-selling opportunities may encourage higher engagement.
4. **Visualization**:
   The clusters were visualized using a scatter plot of the first two features (RegionEncoded and TotalValue). This plot highlights the separation and compactness of the clusters.



Discussion and Insights

The analysis reveals meaningful customer segments that provide actionable insights for business strategy:

1. **High-Value Customers (Cluster 0)**:
   These customers are the most valuable. Offering loyalty programs and premium services can improve retention and maximize revenue.
2. **Low-Engagement Customers (Clusters 1 and 3)**:
   Focused marketing campaigns, discounts, and incentives may encourage more frequent purchases.
3. **Region-Specific Strategies (Cluster 2)**:
   Customers in this cluster show distinct regional preferences. Tailored regional promotions and targeted advertisements could drive higher engagement.
4. **Product Diversification**:
   Identifying popular products within each cluster allows for better inventory planning and cross-selling opportunities.

Conclusion

The clustering analysis has identified four customer segments, each with unique characteristics. The Davies-Bouldin Index and Silhouette Score indicate well-separated and compact clusters. By leveraging these insights, businesses can optimize marketing strategies, enhance customer satisfaction, and improve profitability.