Customer Clustering Report

Overview

This report outlines the clustering analysis performed on the eCommerce dataset. Using transaction and customer profile data, customers were segmented into distinct groups to identify patterns and support targeted marketing strategies. Clustering was conducted using various techniques, with KMeans providing the optimal results.

Data Preparation

1. Datasets Used:

- Customers.csv: Contains customer demographic information.

- Transactions.csv: Includes transaction details such as quantity, total value, and transaction date.

2. Features Created:

- Total quantity purchased.

- Total transaction value.

- Number of transactions per customer.

- Regional data encoded using one-hot encoding.

3. Preprocessing Steps:

- Standardized numerical features (e.g., Quantity, TotalValue, TransactionCount).

- Handled categorical data by one-hot encoding for the `Region` column.

Methodology

Clustering Techniques Explored

1. KMeans Clustering:

- Determined the optimal number of clusters using the Elbow Method.

- Experimented with cluster counts ranging from 2 to 10.

2. Evaluation Metrics:

- Davies-Bouldin Index (DBI): Measures the average similarity ratio within clusters and across clusters.

- Silhouette Score: Evaluates how well each data point fits within its cluster compared to other clusters.

Results

Optimal Number of Clusters

Using the Elbow Method, the optimal number of clusters was determined to be **4**. This selection balanced within-cluster compactness and separation between clusters.

Clustering Metrics

- Davies-Bouldin Index: 1.22

- Silhouette Score: 0.25

Cluster Characteristics

Cluster	Average Quantity	Average Total Value	Average Transactions	Customer Count
0	Low	Low	Low	26
1	Moderate	Moderate	Moderate	61
2	Below Average	Below Average	Below Average	79

| 3 | High | High | High | 33 |

Visualizations

1. Elbow Curve

The Elbow Curve displayed a significant bend at 4 clusters, indicating the optimal number of clusters.

2. Pairplot of Clusters

Clusters were visualized using a pairplot to show the distribution of key features across the groups.

Business Insights

1. High-Value Customers (Cluster 3):

- These customers make frequent and high-value transactions.

- Action: Implement exclusive loyalty programs to retain these customers.

2. Moderate Customers (Cluster 1):

- Regular customers with average spending patterns.

- **Action:** Target them with promotional offers to encourage upselling.

3. Low-Value Customers (Clusters 0 & 2):

- Customers with infrequent transactions and lower spending.

- **Action:** Use retention strategies like discounts or personalized recommendations to engage them.

Recommendations

1. Marketing Strategies:

- Focus on Cluster 3 for premium campaigns.
- Design general promotions for Cluster 1.

2. Inventory Optimization:

- Align inventory with Cluster 3's purchase patterns to avoid stockouts.

3. Data-Driven Campaigns:

- Use clustering insights to create targeted email and advertising campaigns.

Conclusion

The clustering analysis provides a clear segmentation of customers, enabling targeted strategies to maximize customer engagement and revenue. Implementing these recommendations will help the business achieve higher customer satisfaction and profitability.