Customer Segmentation Using Clustering

Introduction

Customer segmentation is a fundamental task in marketing and data analytics that helps organizations understand customer behavior and tailor strategies to specific groups. Using clustering techniques, we can divide customers into distinct groups based on their purchasing behavior and demographic profiles. This report outlines the methodology, results, and metrics obtained from applying clustering on the customer dataset.

Data Preparation

The datasets used for this analysis included Customers.csv and Transactions.csv. The customer profile data contained details such as CustomerID, CustomerName, Region, and SignupDate, while the transaction data included TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, and Price.

We merged these datasets on CustomerID to create a unified dataset containing both profile and transaction information. Relevant features like total spend, purchase frequency, average price, and total quantity purchased were aggregated and engineered from the transaction data. The Region column was encoded into numeric values for clustering purposes, while unnecessary columns, such as CustomerName and SignupDate, were excluded. Finally, the features were standardized using StandardScaler to ensure all variables were on the same scale.

Clustering Technique

The K-Means algorithm was used for customer segmentation due to its efficiency and interpretability. We experimented with different numbers of clusters, ranging from 2 to 10, and settled on 5 clusters based on performance metrics. The clustering process grouped

customers into segments with similar purchasing patterns and demographic attributes.

Evaluation Metrics

To evaluate the quality of clustering, we calculated the Davies-Bouldin (DB) Index and Silhouette Score:

- Davies-Bouldin Index: A lower DB index indicates better cluster separation. For this segmentation, the DB Index was approximately 0.976.
- Silhouette Score: The average silhouette score of 0.615 indicated reasonably well-defined clusters.

Cluster Visualization

To visualize clusters, we applied Principal Component Analysis (PCA) to reduce the feature dimensions to two components. A scatter plot of the clusters in PCA space revealed distinct groups, highlighting meaningful segmentation.

Insights

1. Customers in specific clusters exhibited high purchase frequency and average spending, representing potential targets for premium offerings.
2. Other clusters revealed low engagement or single-purchase behavior, suggesting the need for retention strategies.
3. Region-specific segmentation provided insights into localized preferences.