Customer Segmentation Clustering Report

Overview

This report presents the results of the customer segmentation analysis conducted using clustering techniques on the eCommerce Transactions dataset. The analysis aims to group customers based on their transaction behaviors and profile information to derive actionable insights for targeted marketing and customer relationship management.

Data Preparation

The analysis utilized two primary datasets:

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

Transactions.csv: Contains transaction details, including TransactionID, CustomerID, ProductID, TransactionDate, Quantity, TotalValue, and Price.

The datasets were merged to create a comprehensive view of customer profiles alongside their transaction history. Key features selected for clustering included:

Total Value: The total monetary value of transactions for each customer.

Quantity: The total quantity of products purchased by each customer.

Clustering Methodology

Clustering Algorithm

The K-Means clustering algorithm was chosen for this analysis due to its efficiency and effectiveness in partitioning data into distinct groups.

Optimal Number of Clusters

To determine the optimal number of clusters, the following methods were employed:

Elbow Method: This method involved plotting the inertia (within-cluster sum of squares) against the number of clusters. The "elbow" point in the plot indicates the optimal number of clusters.

Davies-Bouldin Index (DB Index): This index measures the average similarity ratio of each cluster with its most similar cluster. A lower DB Index indicates better clustering quality.

Results

Number of Clusters Formed: After analyzing the Elbow Method and DB Index, the optimal number of clusters was determined to be 4.

Davies-Bouldin Index Value: The DB Index value for the optimal clustering solution was calculated to be 0.7103. This value indicates a good separation between the clusters, with lower values suggesting better clustering quality.