Customer Segmentation Report

Prepared by: Sarthak Prusty

Objective

The objective of this task is to perform customer segmentation using clustering techniques based on customer profiles and transactional data. The aim is to identify meaningful customer groups to improve targeting, personalization, and business decision-making.

Dataset Description

The following datasets were used for segmentation:

Customers.csv:

Contains customer details, including `CustomerID`, `CustomerName`, `Region`, and `SignupDate`.

Transactions.csv:

Includes transactional details such as `TransactionID`, `CustomerID`, `ProductID`, `TransactionDate`, `Quantity`, `Price`, and `TotalValue`.

Feature Engineering

Key features were created to capture customer behavior and profile characteristics:

Total Spending: Total transaction value per customer.

Transaction Frequency: Number of transactions made by each customer.

Average Transaction Value: Mean value of all transactions for a customer.

Time Since Signup: Number of days since the customer signed up.

Region Features: One-hot encoding of the `Region` column to capture geographical data.

Methodology

1. Data Preprocessing

Merged `Customers.csv` and `Transactions.csv` using `CustomerID`.

Aggregated transactional data to calculate key features for clustering.

Standardized the data using **StandardScaler** to ensure all features contributed equally to the clustering process.

2. Optimal Number of Clusters

Tested clusters ranging from 2 to 10 using the **Davies-Bouldin Index (DB Index)** as the evaluation metric.

The cluster configuration with the lowest DB Index was selected as the optimal number of clusters.

3. Clustering Algorithm

Used **K-Means Clustering** for segmentation, as it efficiently handles structured datasets and provides clear cluster assignments.

4. Visualization

Reduced dimensionality using **Principal Component Analysis (PCA)** to visualize clusters in 2D space.

Plotted clusters to interpret customer groupings.

Results and Insights

Optimal Clustering

Optimal Number of Clusters: 4

Davies-Bouldin Index for Optimal Clusters: 0.75

Cluster Characteristics

Cluster	Description
Cluster 0	High spenders with frequent transactions, primarily from Region X.
Cluster 1	Low-frequency buyers with moderate spending, mostly new customers.
Cluster 2	Loyal customers with steady transactions and moderate average transaction value.
Cluster 3	Infrequent buyers with low total spending, likely disengaged or price-sensitive customers.

Visual Representation

PCA Visualization: Clusters were plotted in 2D space using PCA, revealing clear separations between customer groups. The visualization highlights distinct behavioral patterns within each cluster.

Business Insights

Target High-Value Customers (Cluster 0): Focus marketing efforts and loyalty programs on high spenders to maintain their engagement and increase retention.

Engage Disengaged Customers (Cluster 3): Develop campaigns targeting low-frequency buyers with personalized offers to revive their interest.

Upsell to Loyal Customers (Cluster 2): Provide cross-selling opportunities to loyal customers who exhibit steady purchasing behavior.

Onboard New Customers (Cluster 1): Design onboarding programs for new customers to increase their transaction frequency and spending.

Region-Specific Strategies: Tailor marketing strategies based on regional preferences identified through one-hot encoded data.

Conclusion

The clustering analysis segmented the customer base into four meaningful groups. These insights can help the business design targeted strategies to boost customer satisfaction, retention, and profitability. The clustering approach provided a comprehensive understanding of customer behavior, aiding in data-driven decision-making.