Customer Segmentation Using Clustering

Objective

The goal of customer segmentation is to group customers into distinct clusters based on their purchasing behavior and profile information. This allows businesses to better understand their customers and develop targeted marketing strategies.

Approach

1. **Data Preparation**:

   o Combined data from Customers.csv and Transactions.csv.

   o Aggregated transaction data for each customer, including total transaction value (TotalValue), total quantity purchased (Quantity), and frequency of transactions.

   o Normalized features using **StandardScaler** to ensure all variables contribute equally to the clustering process.

2. **Clustering Algorithm**:

   o Applied the **K-Means Clustering** algorithm to segment customers.

   o Experimented with the number of clusters (k) ranging from 2 to 10 to find the optimal value of k.

3. **Evaluation Metrics**:

   o Used the **Davies-Bouldin Index (DBI)** to evaluate clustering performance. Lower DBI values indicate better-defined clusters.

   o Visualized clusters using 2D scatter plots to interpret their separation and structure.

Metrics and Results

1. **Optimal Number of Clusters**:

   o The elbow method was used to determine the initial range for k. A significant drop in inertia was observed at k = 4.

   o DBI was calculated for each k. The lowest DBI value (0.87) was observed for k = 4, indicating well-separated clusters.

2. **Davies-Bouldin Index (DBI)**:

   o For k = 4, DBI = 0.87.

   o For k = 3, DBI = 1.02.

   o For k = 5, DBI = 0.92.

   o Based on these values, k = 4 was selected as the optimal number of clusters.

3. **Cluster Characteristics**:

   o **Cluster 1**: High-value customers with frequent purchases. Likely VIP customers.

   o **Cluster 2**: Moderate-value customers with occasional purchases. Represents regular customers.

   o **Cluster 3**: Low-value customers with infrequent purchases. May require targeted engagement.

   o **Cluster 4**: New customers or those with very limited transaction history.

Insights

1. **High-Value Customers (Cluster 1)**:

   o Contribute significantly to overall revenue.

   o Should be targeted with loyalty programs and exclusive offers.

2. **Moderate-Value Customers (Cluster 2)**:

- Represent a steady source of income.

- Can be incentivized to increase purchase frequency through discounts or promotions.

3. **Low-Value Customers (Cluster 3)**:

   - Represent potential churn risk.

   - Require re-engagement campaigns to boost transactions.

4. **New or Limited-Activity Customers (Cluster 4)**:

   - Should be introduced to key products through onboarding campaigns and introductory offers.