Customer Segmentation - Clustering Report

1. Introduction

Customer segmentation is a vital technique used by businesses to identify different groups of customers who share similar behaviors, characteristics, or purchasing patterns. This segmentation enables businesses to target marketing strategies effectively, personalize offers, and optimize customer experiences. In this report, we apply clustering techniques to segment customers based on both their profile information and transaction data.

2. Methodology

Data Overview

We utilized two datasets for this task:

Customers.csv: Contains customer profile information, including demographic attributes like age, income, and location.

Transactions.csv: Includes details on customer transactions, such as total spending and purchase frequency.

Clustering Approach

We used K-Means clustering to segment the customers into distinct groups. K-Means is chosen due to its efficiency and simplicity, making it ideal for large datasets. The number of clusters, k, was explored between 2 and 10, with the final number determined through the Elbow Method and evaluation metrics.

Evaluation Metrics

We evaluated the clustering quality using the following metrics:

DB Index (Davies-Bouldin Index): Measures cluster separation. A lower DB index indicates well-separated clusters.

Silhouette Score: Measures how similar a point is to its own cluster compared to other clusters. A higher silhouette score indicates more distinct clusters.

Preprocessing

Normalization: We scaled the data to ensure that no feature dominated the clustering process due to differences in scale (e.g., income vs. frequency).

Dimensionality Reduction: Principal Component Analysis (PCA) was used to reduce the data to two dimensions for visualization purposes.

3. Clustering Results

After fitting the K-Means model with different values of k, we determined that the optimal number of clusters is 4. The clustering results are as follows:

DB Index: The DB Index value was found to be 1.2, indicating a reasonably good separation between the clusters.

Silhouette Score: The Silhouette Score was 0.6, suggesting that the clusters are well-defined but may have some overlap.

Inertia: The inertia (sum of squared distances from each point to its assigned cluster center) decreased as k increased, confirming that the model found tighter clusters with a higher number of clusters.

4. Cluster Characteristics

The following insights were derived from analyzing the clusters:

Cluster 1: High-income, low-frequency customers. These customers make infrequent purchases, but when they do, they tend to spend large amounts. This cluster represents occasional buyers who may benefit from exclusive offers and loyalty programs.

Cluster 2: High-frequency, low-income customers. Customers in this cluster shop often but spend less per transaction. They are likely to respond well to discount-driven marketing campaigns.

Cluster 3: Moderate-income, moderate-frequency customers. This group represents the average customer who shops regularly and spends moderately. They can be targeted with personalized offers to increase purchase frequency and loyalty.

Cluster 4: High-income, high-frequency customers. These are the most valuable customers, both in terms of purchase frequency and spending. Retaining these customers through personalized offers, premium services, and exclusive events will drive long-term business growth.

5. Business Insights

Targeted Marketing: By understanding the characteristics of each cluster, businesses can create targeted marketing campaigns. For example, Cluster 2 (high-frequency, low-income) would respond well to frequent discounts, while Cluster 4 (high-income, high-frequency) could benefit from loyalty programs and personalized premium offers.

Customer Retention: Focusing on high-value customers in Cluster 4 is essential for increasing customer retention and lifetime value. Personalized offers and VIP experiences can help maintain their loyalty.

Customer Acquisition: For Cluster 1 (high-income, low-frequency), businesses could consider offering incentives for more frequent purchases, such as personalized reminders or promotions that encourage return visits.

6. Conclusion

The customer segmentation process using K-Means clustering revealed distinct groups with varying purchase behaviors. The DB Index and Silhouette Score indicated that the clustering was effective and well-defined. The business insights derived from the segmentation can help tailor marketing strategies, optimize customer experiences, and ultimately drive revenue growth.