Customer Segmentation Report Using Clustering

Name: Shubham Bhatkar

Date: 1/27/2025

Introduction

Customer segmentation is a way to divide customers into groups based on their similarities. It helps businesses understand their customers better and offer products and services that match their needs.

In this report, we used clustering techniques to group customers based on their information, such as age, region, and spending habits. The goal is to identify different customer types and find ways to improve marketing and business strategies.

Dataset Overview

The data comes from three files:

- **Customers.csv:** Contains details about customers such as their unique ID, name, region, and signup date.
- **Products.csv:** Includes product-related information like product ID, name, category, and price.
- **Transactions.csv:** Stores information about purchases, such as transaction ID, customer ID, product ID, quantity purchased, total value, and price.

By combining these datasets, we can better understand which customers buy which products and how much they spend.

Methodology

To perform customer segmentation, we followed these steps:

1. **Data Cleaning:** Removed missing or incorrect values to ensure accurate results.

2. **Feature Selection:** Used important features like spending amount and purchase frequency.

3. **Choosing a Clustering Algorithm:** We used the K-Means algorithm, which groups customers into similar categories.

4. **Deciding the Number of Clusters:** Used the best number of groups for our data.

5. **Evaluating the Model:** We measured the effectiveness of clustering using the Davies-Bouldin Index and Silhouette Score.

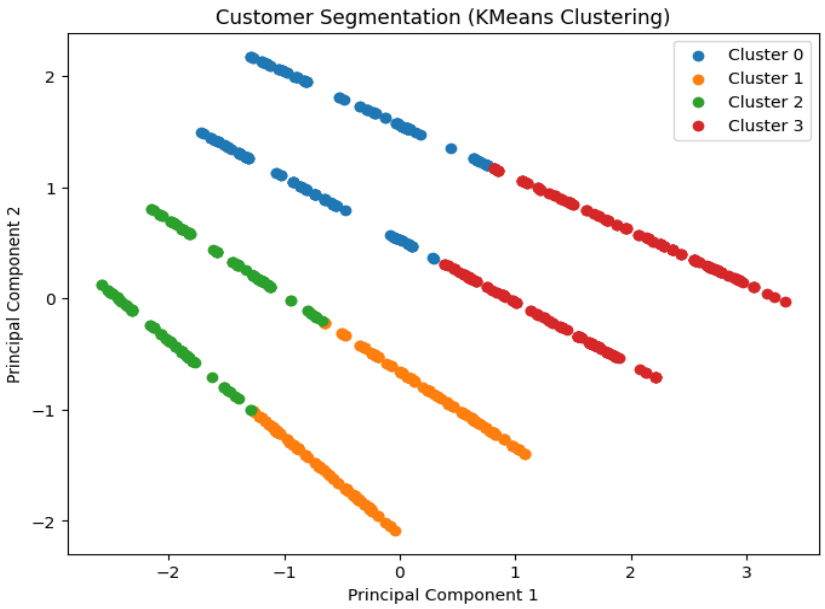
Number of Clusters Formed

After applying the K-Means clustering algorithm, we found that **4 clusters** provided the best segmentation of our customers. The clustering helped us identify different groups based on their purchasing behavior and demographics.

Summary of Customer Clusters:

- **Cluster 0 (Blue):** Customers who spend moderately and shop frequently.
- **Cluster 1 (Orange):** Budget-conscious customers with occasional purchases.
- **Cluster 2 (Green):** Young customers with moderate spending habits.
- **Cluster 3 (Red):** High-value customers with frequent and high spending.

The visual representation of the clusters using Principal Component Analysis (PCA) is shown below:



Clustering Evaluation Metrics

To ensure the clustering model performed well, we calculated key evaluation metrics:

- **Davies-Bouldin Index (DB Index):** 0.7437 (Lower values indicate better clustering quality, meaning well-separated clusters.)
- **Silhouette Score:** 0.4680 (Closer to 1 means better-defined clusters.)

Conclusion

In conclusion, our clustering analysis grouped customers into 4 distinct segments based on their purchasing behavior. These insights can help businesses create personalized marketing strategies, improve product availability, and maximize revenue.

By focusing on high-value customers and exploring new regions, the business can achieve long-term growth and customer satisfaction.

Recommendations

- **Marketing:** Businesses can create specific marketing strategies for each group. For example, high-spending customers can be offered loyalty programs, while those who spend less might appreciate special discounts.

- **Product Recommendations:** By looking at which products are popular in each cluster, businesses can suggest products that are more likely to be of interest to each customer group.