Clustering Analysis Report

Vishal Singh   Email: vishal21575@iiitd.ac.in   Phone: +91 7889182803

Introduction

This report explains how customers were grouped into different clusters using the KMeans algorithm. The goal of this analysis was to divide customers into groups based on how much they spend and how often they shop. The best number of clusters was found using the Davies-Bouldin Index (DBI), and we used PCA to show the groups visually.

Evaluation Criteria

- **Clustering Logic and Metrics:** The clusters were judged based on the DBI and key statistics, such as how much customers spend on average and how often they make purchases.

- **Visual Representation of Clusters:** We used PCA to reduce the data to two or three dimensions, so it is easier to see the groups and how they are different from each other.

Clustering Results

1. Number of Clusters Formed

We found that dividing customers into **7 clusters** worked best. We tested different numbers of clusters, from 2 to 10, and picked 7 because it gave the lowest Davies-Bouldin Index value.

2. Davies-Bouldin Index

The final DBI value for 7 clusters is **0.8535**. A lower DBI means the groups are clearly separated and well-formed.

3. Cluster Metrics

The table below shows the details for each group, including the average spending, total spending, number of purchases, and average value per purchase:

Cluster	Avg. Total Spend ($)	Total Spend ($)	Avg. Transactions	Avg. Purchase Value ($)
0	6,606.11	59,454.96	9.11	734.49
1	3,717.84	219,352.56	5.15	700.67
2	1,857.15	44,571.53	3.79	489.03
3	2,124.87	55,246.60	3.19	691.65
4	2,724.53	103,532.31	4.53	628.82
5	5,379.35	96,828.37	7.50	736.29
6	4,440.37	111,009.23	5.32	880.44

Table 1: Cluster Metrics Summary

4. Visual Representation of Clusters

We used PCA to reduce the data to three dimensions. The plot below shows the 7 groups, with each cluster shown in a different color:
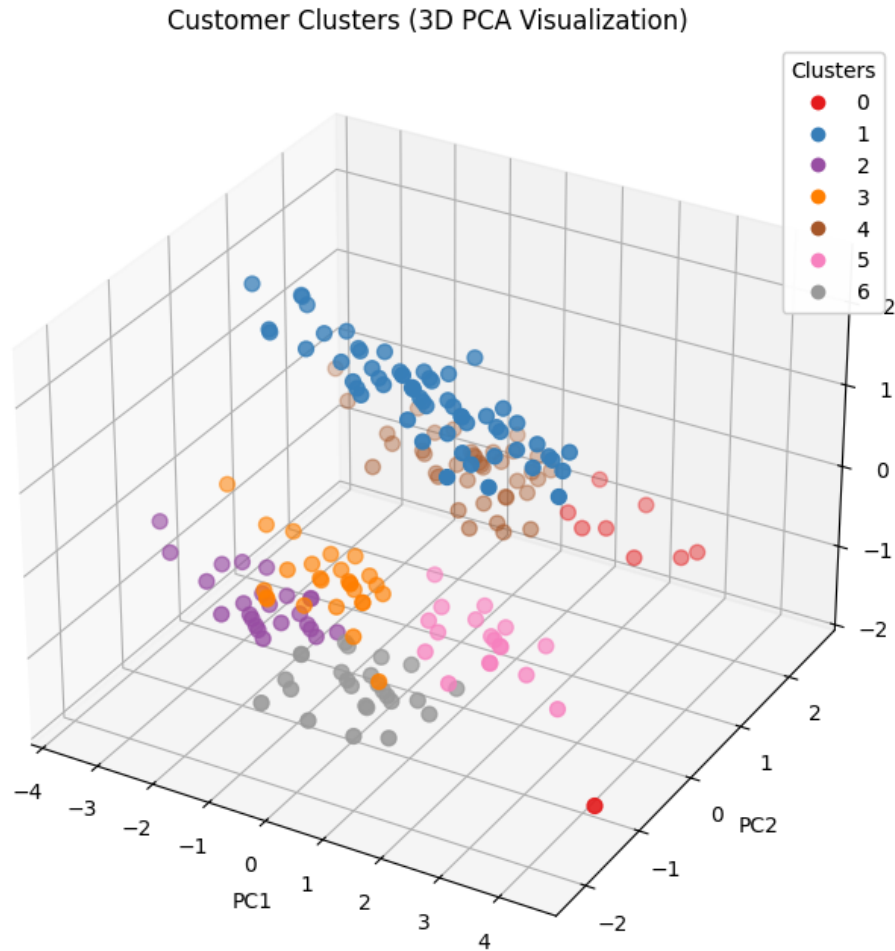
Figure 1: Customer Clusters (PCA Visualization)

Key Insights

1. **High-Spending Customers (Clusters 0 and 6):** These groups include customers who spend a lot on average. Offering them loyalty programs or personalized discounts could help keep them coming back.

2. **Low-Spending Customers (Cluster 2):** Customers in this group spend the least and shop less often. Discounts or bundle deals might encourage them to shop more.

3. **Moderate-Spending Customers (Clusters 1, 4, and 5):** These groups make up most of the customer base. They shop regularly and contribute a lot to total sales. Cross-selling and upselling can work well for these customers.

4. **General Strategy:** Looking deeper into the behavior of each group can help make targeted marketing campaigns. Each group has different spending habits and preferences, so tailor offers accordingly.

Conclusion

The clustering analysis shows clear differences between customer groups. By understanding the spending and shopping patterns of each group, businesses can create targeted strategies to increase sales and keep customers happy. The Davies-Bouldin Index shows the clusters are well-formed, and the PCA visualization helps to see how the groups are separated.