Customer Segmentation Results

Introduction

This report presents the results of customer segmentation based on clustering techniques. The data used includes both customer profile information (from Customers.csv) and transaction details (from Transactions.csv). The task involved clustering customers based on their spending patterns, frequency of transactions, and recency of their latest purchase. Clustering was performed using KMeans, and the results were evaluated using clustering metrics, including the Davies-Bouldin Index (DBI).

Clustering Approach

1. Data Preprocessing:

- **Data Merging:** Customer and transaction data were merged based on the **CustomerID**.
- **Feature Engineering:**
  - **TotalSpending:** The total amount spent by each customer.
  - **Frequency:** The number of transactions made by each customer.
  - **Recency:** The number of days since the customer's last transaction.
  - **SignupYear:** The year the customer signed up.

2. Feature Scaling:

- The features were scaled using **StandardScale**r to ensure uniformity and prevent dominance of any single feature during clustering.

3. Clustering Algorithm:

- **KMeans Clustering** was applied with different numbers of clusters ranging from **2 to 10.**
- The optimal number of clusters was determined based on the **Davies-Bouldin Index (DBI),** which balances cluster separation and compactness.

Results

1. Number of Clusters Formed:

- **4 clusters** were formed based on the optimal balance of separation and compactness. This was determined through trial with different cluster numbers and the **DBI** evaluation.

2. Davies-Bouldin Index (DBI):

- **The Davies-Bouldin Index (DBI)** was calculated to assess the quality of the clusters. A lower DBI indicates better-defined clusters. The calculated DBI for the clustering model was 1.2528. This suggests that the clusters are reasonably well-separated but not perfectly distinct.

3. Other Relevant Clustering Metrics:

- **Cluster Summary: The following table presents key metrics for each cluster.**

Cluster	TotalSpending_mean	TotalSpending_std	Frequency_mean	Recency_mean	SignupYear_mean
0	4596.93	1461.74	6.61	47.25	2022
1	1742.44	1015.81	2.80	141.30	2022.58
2	5875.09	1310.11	7.55	45.97	2023.45
3	3043.83	1033.59	4.66	55.26	2023.71

- **Cluster 0:** Customers with higher total spending and moderate frequency. They tend to have signed up in 2022.
- **Cluster 1:** Customers with lower spending and frequency, and a relatively older customer base.
- **Cluster 2:** Customers with the highest spending and frequency, having signed up more recently in 2023.
- **Cluster 3:** Customers with moderate spending and frequency, and a relatively recent signup pattern.

4. Visual Representation:

The clusters were visualized using a scatter plot of **TotalSpending vs. SignupYear**. The different colors represent each cluster. The plot visually shows the distribution of customers in the 4 clusters, with each cluster showing distinct patterns in spending and signup year.

Conclusion

The customer segmentation using **KMeans clustering** resulted in 4 distinct clusters. The clusters were evaluated using the **Davies-Bouldin Index**, which yielded a value of **1.2528**. The clustering provided meaningful insights into customer behavior, distinguishing high-spending and frequent buyers from others, with variations based on the signup year.

Key Takeaways:

- **Number of clusters:** 4
- **DB Index:** 1.2528 (indicating reasonably well-separated clusters)
- **Cluster characteristics:** The clusters vary in terms of spending, frequency of transactions, recency, and signup year, allowing for targeted strategies in customer engagement and marketing.