Customer Segmentation and Clustering Report

1. Introduction

Customer segmentation is a crucial technique for businesses to understand customer behaviors and tailor marketing strategies. This report focuses on clustering customers based on their transaction behaviors such as total transactions, average transaction value, and total spend. The data analysis was conducted using the **K-Means clustering** algorithm, with the **Davies-Bouldin Index** (DBI) used to evaluate the quality of the clustering.

2. Data Preprocessing

The initial step in this analysis involved loading two datasets:

- **Customers.csv**: Contains customer information.
- **Transactions.csv**: Records of transactions, including Transaction ID, Customer ID, and Total Value of the transaction.

The datasets were then merged based on the **CustomerID** field, and transaction data was summarized:

- **Total Transactions**: The total number of transactions for each customer.
- **Average Transaction Value**: The average value of each transaction made by a customer.
- **Total Spend**: The total monetary value spent by each customer.

3. Feature Selection and Scaling

The features selected for clustering were:

- **total_transactions**
- **avg_transaction_value**
- **total_value**

These features were scaled using the **StandardScaler** to normalize the data, ensuring that each feature contributed equally to the clustering model. This step is crucial since K-Means clustering is sensitive to the scale of the data.

4. Elbow Method for Optimal Clusters

The **Elbow Method** was employed to identify the optimal number of clusters (k) for the K-Means algorithm. The **distortion (inertia)** for different values of k (from 1 to 10) was calculated, and a plot was generated to visualize the results. The optimal value of k was identified as **k=4** based on the Elbow curve, where the distortion starts to decrease at a diminishing rate beyond this point.

5. <u>K-Means Clustering</u>

Using **k=4**, the **K-Means clustering algorithm** was applied to segment the customers. Each customer was assigned to one of four clusters, based on their transaction behaviors. These clusters represent different customer profiles, which can be used for targeted marketing strategies and business analysis.

6. <u>Davies-Bouldin Index (DBI) Evaluation</u>

The **Davies-Bouldin Index (DBI)** was calculated to evaluate the quality of the clusters. The DBI value for this clustering model was found to be **1.074**, which suggests that the clusters are moderately well-separated, but there is room for improvement in terms of cohesion and separation.

- **<u>Interpretation</u>**: A DBI value of **1.074** is slightly above 1, indicating that the clustering model could potentially benefit from further tuning (e.g., trying different values of k or using additional features).

7. <u>Cluster Visualization</u>

A scatter plot was created to visualize the clusters based on the first two features (`total_transactions` and `avg_transaction_value`). Each cluster is represented in a different color, which helps in understanding the distribution of customers across different segments.

8. <u>Output</u>

A CSV file named **Customer_Segments.csv** was generated, containing each customer's **CustomerID** and their corresponding cluster number. This file can be used for further analysis or integration with marketing efforts.

9. <u>Business Implications</u>

The clustering results provide meaningful insights that can be leveraged for business decisions:

- **<u>Targeted Marketing</u>**: By analyzing the characteristics of each cluster, businesses can design personalized marketing campaigns tailored to specific customer segments.
- **<u>Customer Retention</u>**: High-value or frequent customers can be identified for loyalty programs, while low-value customers may be targeted with re-engagement strategies.
- **<u>Product Recommendations</u>**: Understanding customer purchase patterns allows for effective upselling and cross-selling strategies, increasing the average transaction value.

10. <u>Conclusion</u>

The customer segmentation process using K-Means clustering has successfully identified distinct customer groups. Although the DBI value of **1.074** suggests moderate cluster separation, the clusters are still meaningful for practical business use. The next steps would include refining the model by experimenting with additional features or advanced clustering algorithms to improve the clustering quality.