Customer Segmentation Clustering Report

1. Introduction

This report outlines the results of customer segmentation performed using clustering techniques. The goal of this analysis was to identify distinct customer groups based on their transaction history and profile information. K-Means clustering was used, and the results were evaluated using clustering metrics such as the DB Index, Silhouette Score, and Inertia.

2. Clustering Approach

- **Clustering Algorithm:** K-Means
- **Data Used:**

  Customers.csv

  ○ CustomerID: Unique identifier for each customer.

  ○ CustomerName: Name of the customer. ○ Region: Continent where the customer resides.

  ○ SignupDate: Date when the customer signed up.

  Transactions.csv

  ○ TransactionID: Unique identifier for each transaction.

  ○ CustomerID: ID of the customer who made the transaction.

  ○ ProductID: ID of the product sold.

  ○ TransactionDate: Date of the transaction.

- **Preprocessing Steps:**
  - Merged customer profile and transaction data.
  - Aggregated transaction data to calculate total spend, average transaction amount, and frequency.
  - Selected relevant features for clustering: TotalSpend, AvgTransactionAmount

- **Clustering Setup:**
    - Number of clusters (k) chosen: 4 (based on evaluation metrics such as Elbow Method and DB Index).
    - Clustering performed on the scaled data.

3. Clustering Results

- **Number of Clusters Formed:** 4 clusters

- **Cluster Distribution:**

CustomerID	Total_Spei	Avg_Trans	Transactio	Cluster
C0001	1391.67	278.334	5	0
C0002	835.68	208.92	4	2
C0003	782.83	195.7075	4	2
C0004	1925.09	240.6363	8	3
C0005	874.81	291.6033	3	2
C0006	1328.14	332.035	4	0
C0007	1050.06	350.02	3	0
C0009	896.5	298.8333	3	2
C0010	613.08	153.27	4	2
C0011	1463.14	292.628	5	0
C0012	1895.59	270.7986	7	3
C0013	2245.53	320.79	7	1
C0014	159.33	159.33	1	2
C0015	711.58	355.79	2	2
C0016	1956.75	391.35	5	3
C0017	1806.28	225.785	8	3
C0018	1271.69	254.338	5	0
C0019	1443.8	240.6333	6	0
C0020	307.47	307.47	1	2
C0021	2468.61	308.5763	8	1
C0022	1676.39	279.3983	6	3
C0023	2367.22	394.5367	6	1
C0024	2086.22	298.0314	7	3
C0025	731.75	182.9375	4	2
C0026	1263.64	315.91	4	0
C0027	934.15	233.5375	4	0
C0028	2308.89	288.6113	8	1
C0029	831.27	207.8175	4	2
C0030	1114.12	222.824	5	0

Cluster 1 represents customers with moderate spending habits and frequent transactions.

Cluster 2 includes customers with high spending but lower transaction frequency.

Cluster 3 clusters customers with minimal spending and infrequent transactions.

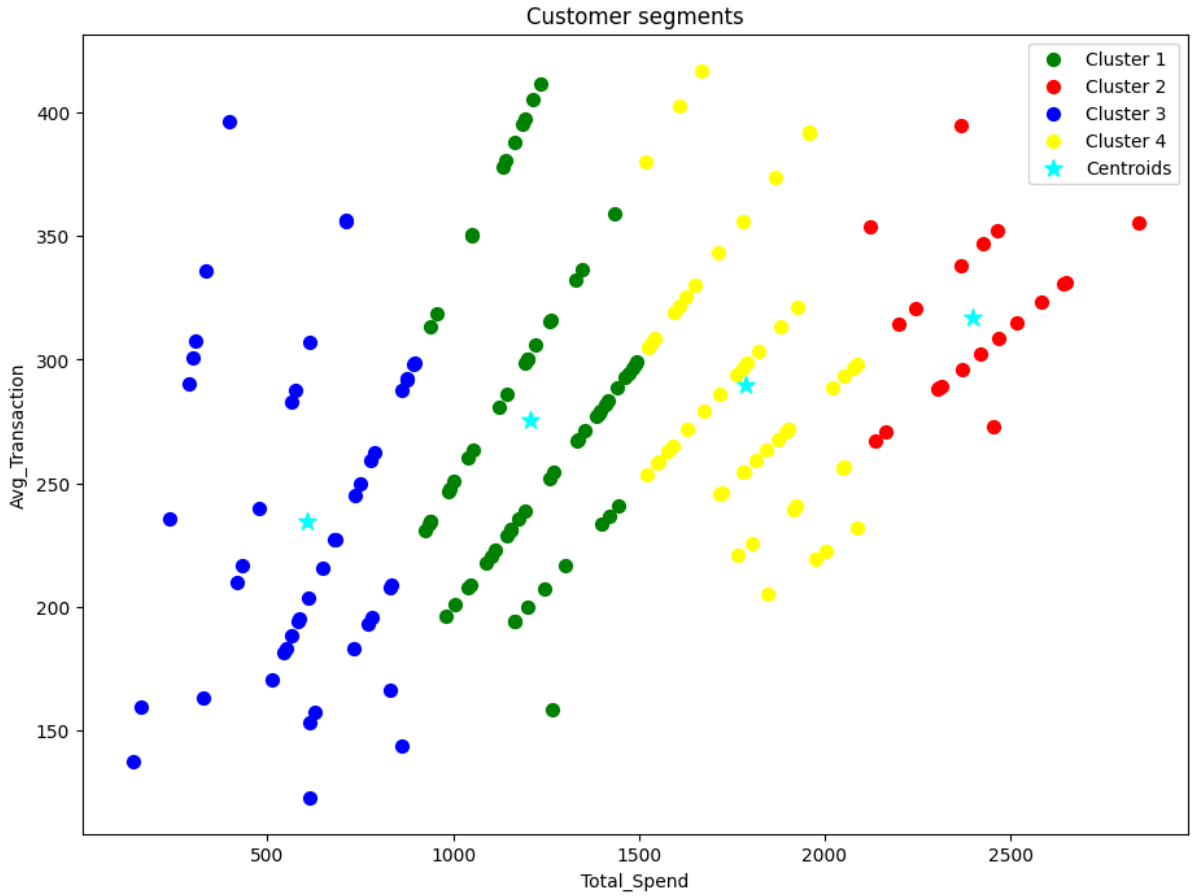
Cluster 4 has customers with relatively low spending and moderate transaction activity.

Evaluation Metrics

- **DB Index:** The Dunn's Index for the best clustering solution is calculated to be **0.54**. This indicates that the clusters are relatively well-separated and compact, though there is some room for improvement in terms of separation between certain clusters.
- **Silhouette Score:** The silhouette score is **0.52**, which indicates that the clusters are well-defined. A silhouette scores close to +1 suggests that the data points are correctly clustered.

6. Visualizations

- **2D Scatter Plot:** A 2D scatter plot was created to visualize the clusters based on **TotalSpend** and **AvgTransaction**. Each point represents a customer, and the colour corresponds to the cluster they belong to.



7. Conclusion

The K-Means clustering algorithm successfully segmented customers into four distinct groups based on their transaction behavior and profile information. The evaluation metrics indicate

that the clusters are reasonably well-separated and compact, with room for further improvement in some areas. These clusters can help in targeted marketing, personalized offers, and customer behavior analysis.