Customer Segmentation - Clustering Report

1. Introduction

The task involves performing customer segmentation using clustering techniques based on both customer profile information (from Customers.csv) and transaction data (from Transactions.csv). The goal is to group similar customers together to identify patterns in spending behavior and transaction patterns.

2. Clustering Algorithm

- Algorithm Used: KMeans Clustering
- Optimal Number of Clusters: 4 (Determined using the Elbow Method)

3. Evaluation Metrics

- **Davies-Bouldin Index (DB Index): 0.9079**
    - The DB Index is a measure of clustering quality. A lower value indicates better clustering results, with values closer to zero indicating well-separated and compact clusters.
- **Silhouette Score: 0.3463**
    - The Silhouette Score provides insight into how well each data point fits into its assigned cluster compared to other clusters. A score closer to +1 indicates well-defined clusters, while scores closer to 0 suggest overlapping clusters.

4. Cluster Analysis

- The clustering analysis was performed on the following features:
    - TotalSpending: The total amount spent by the customer.
    - NumTransactions: The total number of transactions made by the customer.
    - AvgTransactionValue: The average transaction value.

5. Sample Data Points

Below is a table displaying sample data points from the customer segmentation analysis, showing their respective features after clustering:

CustomerID	Region	Total Spending	Num Transactions	AvgTransactionValue
C0001	South America	-0.061701	-0.011458	-0.070263
C0002	Asia	-0.877744	-0.467494	-0.934933
C0003	South America	-0.405857	-0.467494	-0.026271

6. Key Insights

- Cluster Distribution: The customers have been segmented into 4 distinct clusters based on their spending and transaction behaviors.
- Cluster Variability: The clusters show variability in spending patterns, with customers from South America exhibiting higher spending and transaction frequencies compared to customers from Asia.
- Silhouette Score Analysis: While the clusters are moderately well-defined, there is some overlap, as indicated by the Silhouette Score of 0.3463.

7. Conclusion

The customer segmentation performed using KMeans clustering provides valuable insights into customer behavior. The segmentation is based on significant variables like TotalSpending, NumTransactions, and AvgTransactionValue. The DB Index value of 0.9079 and the Silhouette Score of 0.3463 suggest that the clustering is useful but could benefit from further refinement. Future improvements could include exploring other clustering algorithms or adding more features for a more granular segmentation.