Clustering Report

1. Number of Clusters Formed:

The number of clusters formed for customer segmentation was **4** using the K-Means clustering algorithm. This was chosen after experimenting with different values of n_clusters within the range of 2 to 10. We observed that 4 clusters provided a good balance of distinct groups while minimizing overlap between them.

2. DB Index Value:

The **Davies-Bouldin Index (DB Index)** value for the clustering is 1.0604240399923033. The DB Index is a measure of how well-separated and compact the clusters are. A lower DB Index indicates better clustering performance. This value suggests that the clusters are relatively well-separated and compact, but further tuning of the clustering model might reduce the DB Index even more.

3. Other Relevant Clustering Metrics:

- **Silhouette Score**: The silhouette score for the clustering was 0.3135106549790539. The silhouette score measures how similar an object is to its own cluster compared to other clusters. A value close to +1 indicates that the object is well matched to its own cluster and poorly matched to neighbouring clusters. This score reflects the overall quality of the clustering.
- **Cluster Characteristics**:
    - **Cluster 1**: Customers with high total spend and high frequency of purchases.
    - **Cluster 2**: Customers with moderate total spend but low frequency of purchases.
    - **Cluster 3**: Customers with low total spend but frequent small transactions.
    - **Cluster 4**: Customers with low total spend and low frequency of purchases.

These metrics help assess the effectiveness of the clustering and provide insight into customer behaviour based on transaction history.

Conclusion:

The customer segmentation using K-Means clustering has successfully grouped customers into 4 distinct segments based on their transaction patterns and demographic information. The clustering results, including the DB Index and silhouette score, indicate that the model is performing reasonably well, with further tuning potentially improving the quality of the segmentation.