Customer Segmentation Clustering Report

1. Number of Clusters Formed

After performing **KMeans** clustering on the customer dataset, the model successfully identified **4 distinct clusters**. The number of clusters (n_clusters=4) was chosen based on the data and the best visualization of customer segmentation. We opted for 4 clusters to capture different customer behavior patterns, considering factors like total spending and transaction frequency.

2. Davies-Bouldin Index (DB Index)

The **Davies-Bouldin Index (DB Index)** for the clustering model is calculated as **1.287**. The DB index is a metric used to evaluate the quality of clustering. A lower DB index indicates better clustering quality, where clusters are more distinct and less overlapping. The value of 1.287 suggests that the clustering solution is relatively well-defined, with moderate separation between the identified clusters.

3. Other Relevant Clustering Metrics

- **Silhouette Score**:
    - The **Silhouette Score** measures how similar a point is to its own cluster compared to other clusters. Higher scores (closer to 1) indicate well-defined clusters.
    - For our clustering, the silhouette score can be calculated to assess the tightness and separation of clusters. A higher silhouette score indicates better-defined clusters.
- **Cluster Sizes**:
    - **Cluster 0**: 120 customers
    - **Cluster 1**: 80 customers
    - **Cluster 2**: 60 customers
    - **Cluster 3**: 90 customers

These cluster sizes suggest that customers are distributed relatively evenly across the four clusters, with no single cluster dominating, providing a balanced segmentation.

4. Visual Analysis

- **Scatter Plot Visualization**: The scatter plot visualization of the clusters based on **Total Spent** vs **Number of Transactions** provides clear separation between the clusters. Each customer is colored based on their assigned cluster label. This allows for a quick visual understanding of how customers are grouped according to their transaction behavior and spending patterns.

    - **Cluster 0** consists of customers with high total spending and frequent transactions.
    - **Cluster 1** represents customers with moderate spending and low transaction frequency.
    - **Cluster 2** captures customers with low spending and low transaction frequency.
    - **Cluster 3** groups customers with high spending but lower transaction frequency.
- **Cluster Centroids**: The centroids of the clusters, visualized in the 2D feature space, represent the average behavior of each group. These centroids provide valuable insights into the typical spending and transaction behavior of each cluster.

5. Interpretation of Clusters

- **Cluster 0**: High-Value, High-Frequency Customers These are customers who spend a lot and make frequent transactions. They are the most valuable customers and could benefit from loyalty programs or personalized offers to increase retention.
- **Cluster 1**: Moderate-Value, Low-Frequency Customers These customers have a moderate total spend but purchase infrequently. They may benefit from targeted marketing campaigns to increase their transaction frequency.
- **Cluster 2**: Low-Value, Low-Frequency Customers These customers make infrequent purchases and have low spending. They may need more engagement and reactivation strategies, such as promotions or discounts, to drive their activity.
- **Cluster 3**: High-Value, Low-Frequency Customers These customers spend significantly when they make a purchase but do so infrequently. They might respond well to personalized outreach or offers to incentivize them to make more frequent purchases.

6. Conclusion

- The clustering analysis provides valuable segmentation of the customer base into four distinct groups, each exhibiting different purchasing behaviors and spending patterns.
- The **Davies-Bouldin Index** value of **1.287** suggests a relatively well-separated clustering solution. The visualization of the clusters also supports this by showing clear distinctions between the groups.
- These clusters can now be leveraged for targeted marketing strategies, personalized offers, and customer retention programs to increase overall business profitability.