Clustering Report

This report summarizes the results from the clustering analysis performed on customer data. The analysis utilized K-Means clustering to group customers into distinct clusters based on various transaction and product-related features. The performance of the clustering was evaluated using several relevant metrics, including the number of clusters, the Davies-Bouldin Index, and the Silhouette Score.

1. **Number of Clusters Formed** *(Optimal)***:** Based on analysis and experimentation with different cluster numbers, the optimal number of clusters identified for this customer segmentation task was **3**. This choice was made after considering the balance between simplicity and meaningful differentiation among the customer groups.

2. **Davies-Bouldin Index (DBI):** The Davies-Bouldin Index (DBI) is a metric used to evaluate the quality of clustering. It measures the average similarity ratio of each cluster with its most similar cluster, where a lower value indicates better clustering. In this case, the final DBI score was **0.6929.**

- *Interpretation*: A DBI score of 0.6929 suggests that the clustering structure is relatively well-defined, with clusters being distinct enough from one another. A DBI value closer to 0 indicates better clustering result, while 1< indicates poor clustering.

3. Silhouette Score:

The Silhouette Score is a metric used to assess the quality of clustering by measuring how similar an object is to its own cluster compared to other clusters. It ranges from -1 (incorrect clustering) to +1 (well-clustered). A score near 0 suggests that the clusters are overlapping or not well-defined. The Silhouette Score for this clustering task was 0.4637.

- *Interpretation*: A Silhouette Score of 0.4637 indicates that the clustering quality is moderate. It suggests that while the clusters are not perfectly separated, they are reasonably well-formed with some degree of separation.

4. Additional Considerations:

- *Cluster Interpretation*: Based on the features used for clustering, such as transaction value, product IDs, and transaction dates, the three clusters likely represent different customer segments with distinct purchasing behaviors.

- *Scalability*: As the dataset grows, further tuning of the clustering parameters, such as the number of clusters or the features used, may be necessary to maintain meaningful segmentation.

Conclusion:

The clustering analysis yielded 3 clusters, with a Davies-Bouldin Index of 0.6929, which suggests a reasonably well-separated clustering structure. The Silhouette Score of 0.4637 indicates that the clusters are moderately distinct. Overall, the results show promising segmentation, with potential for further refinement to achieve stronger cluster separation.