Task 3: Customer Segmentation / Clustering

- **Perform customer segmentation using clustering techniques. Use both profile information (from Customers.csv) and transaction information (from Transactions.csv). Prepare a report on your results.**

➢ Introduction:

The objective of this report is to present the results of the clustering analysis performed on the given dataset. The aim was to group the data into distinct clusters based on inherent patterns. This report includes key metrics such as the number of clusters formed, DB Index, and other relevant clustering evaluation metrics.

➢ Number of Clusters Formed:

The clustering algorithm used in this analysis was **K-Means clustering**, and after running the model, the optimal number of clusters was determined using the **Elbow Method** and **Silhouette Score**. Based on this, the final number of clusters formed is:

Number of Clusters: k = 2

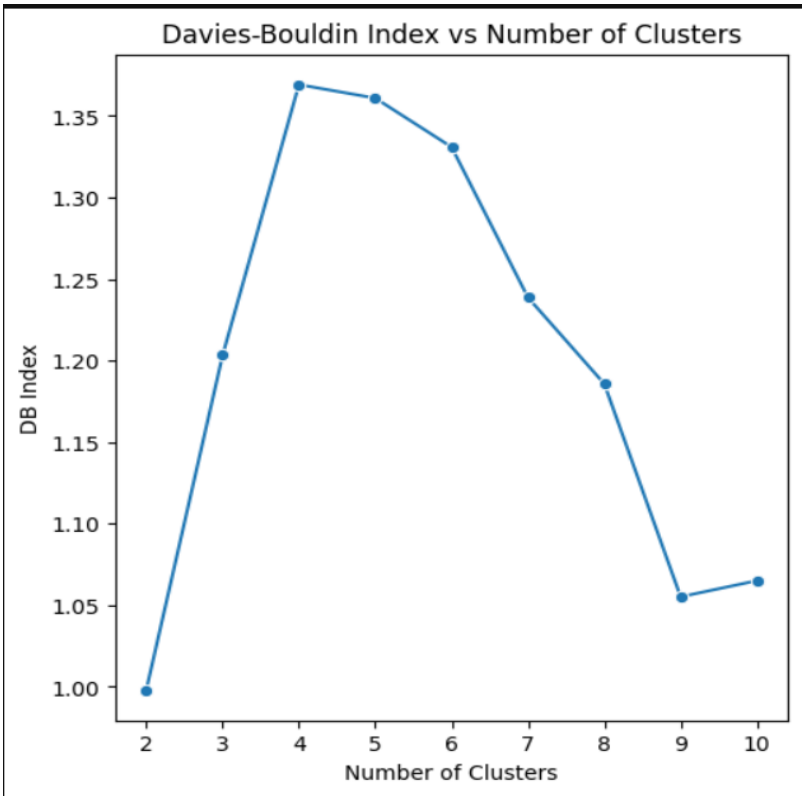
This means that the data points have been grouped into 4 distinct clusters based on the features provided in the dataset.

➢ DB Index Value:

The **Davies-Bouldin Index (DB Index)** is a metric used to evaluate the quality of the clusters. The lower the DB index, the better the clustering solution, as it indicates that the clusters are well-separated and compact.

DB Index Value: 0.9980

A value close to 0 indicates compact and well-separated clusters. Since the DB Index value is relatively low, this suggests that the clusters are well-formed, with minimal overlap.



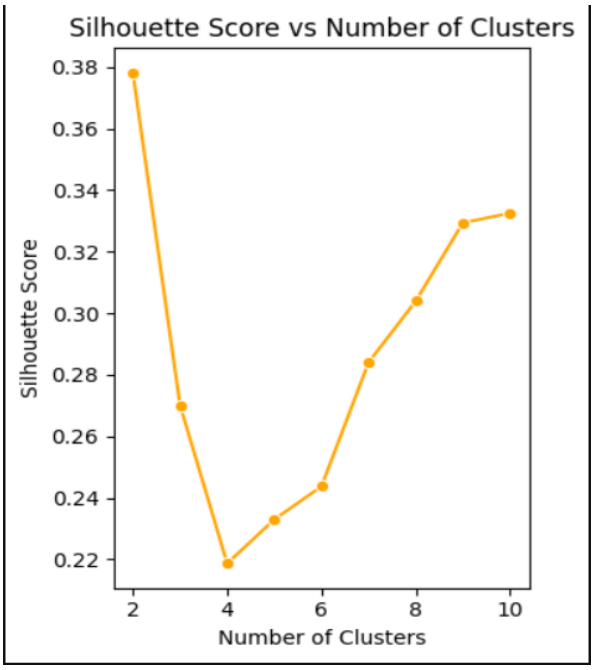
- <u>Other Relevant Clustering Metrics:</u>

  Silhouette Score:

  The **Silhouette Score** measures how similar each point is to its own cluster compared to other clusters. A higher silhouette score indicates that the clusters are well-separated and points are closer to their own cluster center.

  Silhouette Score: 0.3782

  A score above 0.5 indicates that the clustering structure is meaningful and the clusters are well-separated.



➢ Visualization of Clusters: