Customer Segmentation Report

Objective:

This task focused on customer segmentation using both profile and transaction data. The aim was to group customers into distinct segments based on similarities in their behaviours.

Clustering Algorithm:

- **Algorithm Used:** KMeans Clustering
- **Number of Clusters:** 4 (selected based on the Elbow Method)

Clustering Metrics:

1. **Davies-Bouldin Index (DB Index):** 0.939
   This metric indicates the separation between clusters. A lower value reflects better separation, with 0.939 being a favorable result.
2. **Silhouette Score:** 0.452

The silhouette score measures how similar each point is to its own cluster versus other clusters. A score of 0.452 suggests moderate separation between clusters, indicating that the clustering is somewhat effective.

3. **Calinski-Harabasz Index:** 87.86

   This index evaluates the density and separation of clusters. Higher values indicate well-separated and distinct clusters, and 87.86 suggests that the chosen clustering solution is quite effective.

Visualizations:

- **PCA-based Scatter Plot:** A 2D projection of the data showing clear separation between the 4 clusters, confirming the results of the KMeans algorithm.

Conclusion:

- The KMeans algorithm with 4 clusters effectively grouped customers based on their profiles and transaction data. The DB Index, Silhouette Score, and Calinski-Harabasz Index all suggest that the segmentation is meaningful and that the clusters are reasonably well-separated.