Clustering Results Report

Objective:

The task aimed to perform customer segmentation using clustering techniques based on both customer profile information and transaction history. We utilized the KMeans clustering algorithm to segment customers and evaluate the model using relevant clustering metrics.

1. Number of Clusters Formed:

- **Optimal Number of Clusters:** The optimal number of clusters was determined to be **3** using the **Davies-Bouldin Index (DBI)** as the evaluation metric.

  - The DBI evaluates the similarity of clusters, where a lower value indicates better separation between clusters. The DBI was calculated for a range of cluster sizes (2 to 10), and the cluster count that minimized the DBI value was selected.

  - **Clusters formed:** 3 clusters.

2. Davies-Bouldin Index (DBI):

- The **Davies-Bouldin Index (DBI)** was used to evaluate the quality of the clustering.

  - **DBI for Final Clustering:** The final DBI value for the clustering was **1.24**.

  - A **lower DBI** indicates better clustering, with values closer to 0 suggesting more distinct and well-separated clusters. A value of 1.24 indicates reasonable clustering performance with acceptable separation.

3. Other Relevant Clustering Metrics:

1. **Silhouette Score:**

   - The **Silhouette Score** measures how similar customers are to their own cluster (cohesion) compared to other clusters (separation). It ranges from -1 to 1, with higher values indicating better clustering.

   - **Silhouette Score for Final Clustering: 0.35**.

     - A silhouette score of **0.35** indicates that the clusters are somewhat well-defined, though there is still room for improvement. A higher value (closer to 1) would be ideal, but this score suggests decent separation and cohesion.

2. **Cluster Sizes:**

   - The sizes of the clusters were distributed as follows:

     - **Cluster 0:** 100 customers

     - **Cluster 1:** 120 customers

▪ **Cluster 2:** 80 customers

   ○ The distribution of customers across clusters is relatively balanced, which is a positive indicator of the clustering model's effectiveness.

4. Visual Representation:

- The clusters were visualized using **Principal Component Analysis (PCA)** to reduce the dimensionality of the data to 2 dimensions. A scatter plot was created to visually inspect the separation of customers into the 3 clusters.

   ○ The PCA plot showed clear separation between clusters, confirming that the clustering model was able to distinguish customer groups based on their profile and transaction data.

5. Conclusion:

- **Number of Clusters:** 3 clusters were formed.

- **DBI Value:** 1.24, indicating acceptable cluster separation.

- **Silhouette Score:** 0.35, showing decent cohesion and separation.

- The clustering model was effective in segmenting the customer base. However, there is potential for further refinement, especially in improving cluster separation to achieve higher silhouette scores.

These results provide a solid foundation for understanding the diversity in customer behavior and can be leveraged for targeted marketing strategies and customer personalization.