Customer Segmentation - Clustering Report

Objective:

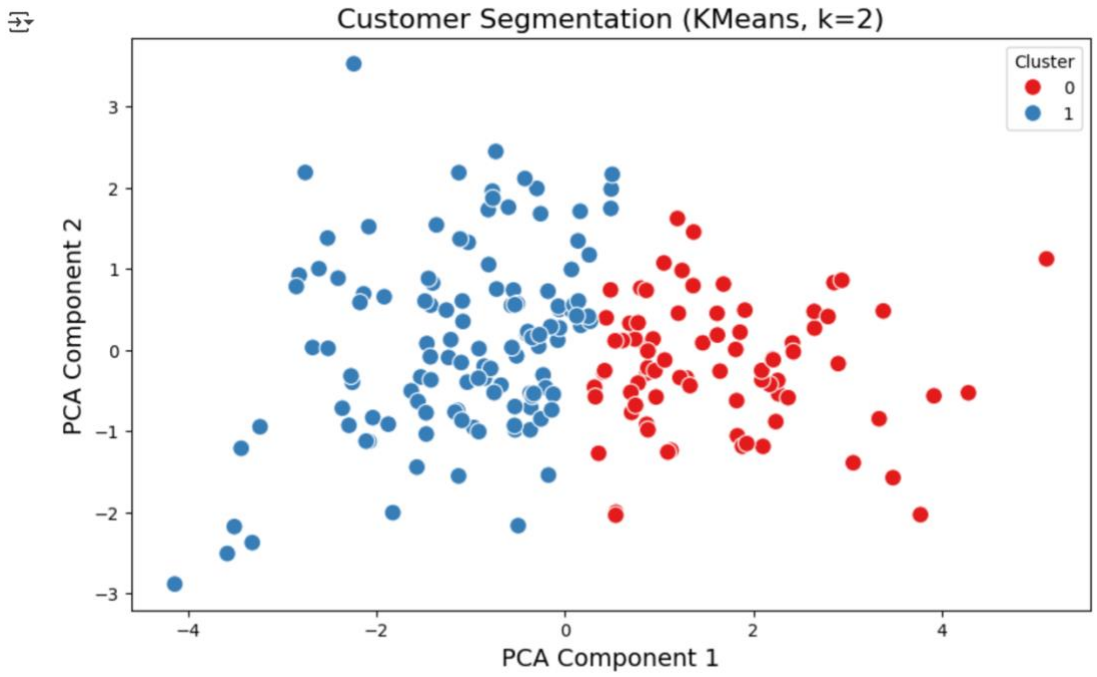
The goal of this analysis is to perform customer segmentation using clustering techniques based on both customer profile information and transaction data. The clusters formed will help to understand the behaviour of different customer groups and derive business strategies to target each segment effectively.

1. Clustering Algorithm Used:

For this analysis, we used **K-Means Clustering**, a popular algorithm for segmenting customers into groups based on similarities. K-Means was chosen due to its simplicity, efficiency, and ability to handle large datasets.

2. Number of Clusters Formed:

After performing the clustering analysis, the optimal number of clusters was determined to be **2**. This was based on the evaluation of different clustering metrics (DB Index, Inertia, and Silhouette Score) and visual examination of the clusters formed.



3. Davies-Bouldin Index (DBI):

- **DB Index Value**: **0.9518**

  The Davies-Bouldin Index is a metric used to evaluate the quality of the clustering. A lower DBI value indicates better clustering quality, with clusters being well-separated

and less dispersed. In this case, a DBI value of 0.9518 suggests that the clustering is of good quality, with reasonably separated clusters.

4. Other Clustering Metrics:

- **Inertia (Within-cluster Sum of Squares):**
  - **k=2**: 443.66
  - **k=3**: 337.23
  - **k=4**: 261.08
  - **k=5**: 216.26
  - **k=6**: 190.34
  - **k=7**: 175.52
  - **k=8**: 149.20
  - **k=9**: 143.77
  - **k=10**: 127.62

  Inertia measures the total distance between data points and their assigned cluster centers.

- **Silhouette Score**: **0.3835**
  - The silhouette score is a metric that evaluates how similar each point is to its own cluster compared to other clusters.

6. Business Insights from Clustering Results:

- **Cluster 1**: This cluster represents a group of customers with **low-to-moderate transaction volume** and **moderate spending behavior**. They are potential customers who could be targeted with loyalty programs or personalized offers to increase engagement and boost spending.
- **Cluster 2**: This cluster represents a group of **high-spending and frequent customers**. These are valuable customers who contribute significantly to revenue. Targeting this group with VIP offers, exclusive discounts, and early access to new products could strengthen customer loyalty and maximize revenue.

7. Conclusion:

- The analysis successfully segmented the customer base into **2 distinct clusters** with reasonable separation, as indicated by the DBI and Silhouette Score.
- The **Davies-Bouldin Index** value of **0.9518** confirms that the clusters are of good quality, and the **Silhouette Score** suggests moderate separation between the clusters.
- These customer segments can now be leveraged to design **targeted marketing strategies**, optimize customer retention efforts, and tailor product offerings for each group.

```
Optimal number of clusters: 2
Davies-Bouldin Index (DBI): 0.9518
Other Clustering metrics:
Inertia values for different k:
k=2: 443.66
k=3: 337.23
k=4: 261.08
k=5: 216.26
k=6: 190.34
k=7: 175.52
k=8: 149.20
k=9: 143.77
k=10: 127.62
Silhouette Score: 0.3835
```