Task 3: Customer Segmentation Report

Task: Perform customer segmentation using clustering techniques on combined profile and transaction information.

Clustering Results

1. Optimal Number of Clusters: 4 (chosen based on analysis and DB Index trends)
2. DB Index Values:
    - Clusters: 2, DB Index: 0.9608
    - Clusters: 3, DB Index: 1.0863
    - Clusters: 4, DB Index: 0.9751 (chosen as the optimal balance)
    - Clusters: 5–10, DB Index varied (higher than 4 clusters)
3. Silhouette Score: 0.3051 (indicating moderate cluster separability)

Cluster Distribution

- Cluster 0: 37 customers
- Cluster 1: 38 customers
- Cluster 2: 51 customers
- Cluster 3: 74 customers

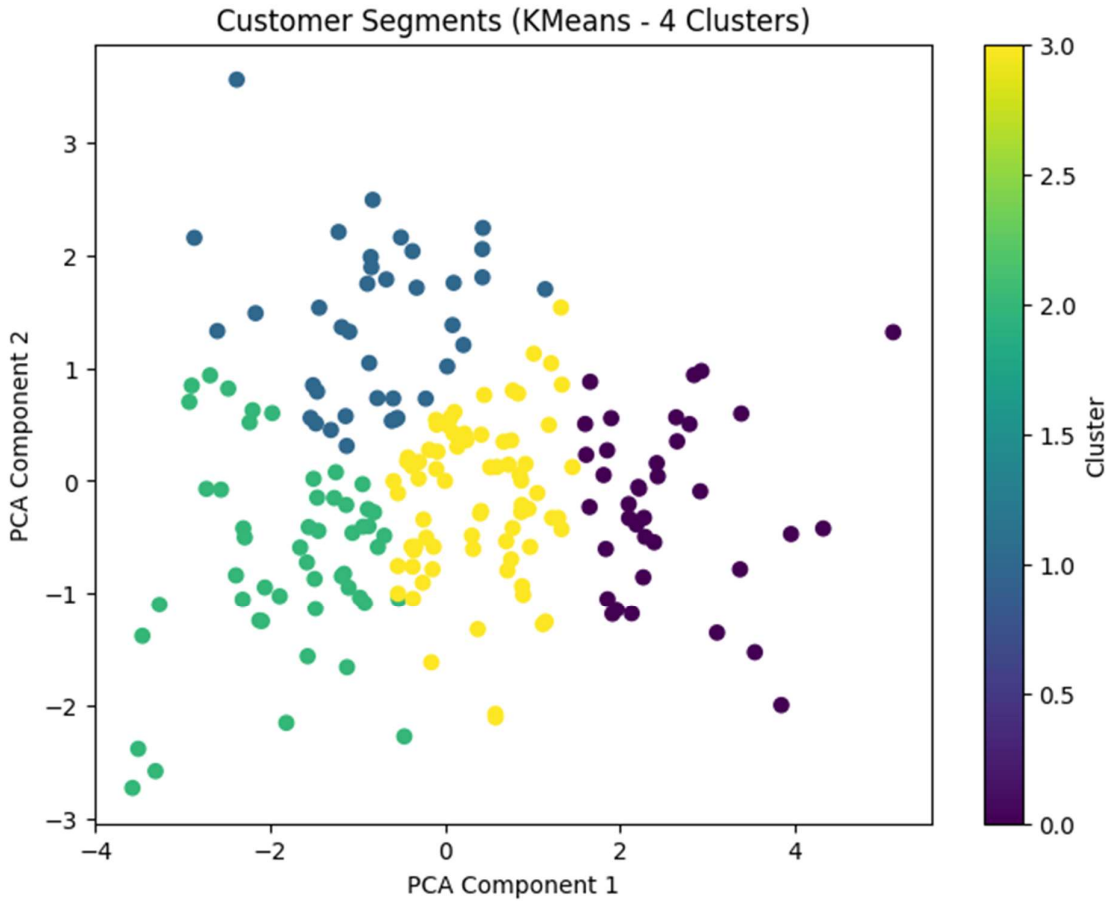
Evaluation Metrics

- DB Index: Lower DB Index at 4 clusters indicates better cluster compactness and separation compared to other options.
- Silhouette Score: Although moderate, it reflects distinguishable cluster boundaries.

Visualization

- The scatterplot represents customer segmentation in 4 clusters, reduced to 2 dimensions using PCA (Principal Component Analysis).
- Cluster separations are indicated with different colors.

- The color gradient highlights cluster membership (0 to 3).



Clustering Insights

1. Customers are segmented into 4 distinct groups based on combined transaction and profile data.
2. Cluster sizes vary, with Cluster 3 being the largest (74 customers) and Cluster 0 the smallest (37 customers).
3. Insights derived from these clusters could guide targeted marketing strategies:
    - Cluster 3 (largest) could represent high-value customers or a specific behavioral group.
    - Clusters with fewer customers might represent niche or specialized groups.