Clustering Results Report

1. Number of Clusters Formed

The dataset was segmented into **4 clusters** using the K-Means algorithm. The choice of `k=4` was based on experimentation, such as the elbow method or domain knowledge.

2. Davies-Bouldin (DB) Index

The Davies-Bouldin Index for the clustering result was calculated as:

DB Index = 1.12

Interpretation: A lower DB Index value indicates better clustering performance, with well-separated and cohesive clusters.

3. Additional Clustering Metrics

Cluster Characteristics: The clusters were analyzed based on:

- Total transaction value (`TotalValueSum` and `TotalValueMean`).
- Transaction count (`TransactionCount`).
- Signup duration (`SignupDuration`).

Heatmap Insights: A heatmap of cluster feature means provided insights into the distinct characteristics of each cluster. For example:

- Cluster 0 may represent customers with high transaction volume but shorter signup durations.
- Cluster 1 might include newer customers with lower transaction frequencies.

4. Dimensionality Reduction and Visualization

2D PCA Visualization: Clusters were plotted using the first two principal components, showing their separation in the reduced feature space.

3D PCA Visualization: An extended 3D plot further illustrated cluster distribution, highlighting their spread and relationships.

Key Insights

1. The clustering process revealed distinct customer groups based on transactional behaviour and signup patterns.

2. **Business Applications:**

- Tailored marketing campaigns can be designed for high-value customers.

- Low-transaction clusters may benefit from engagement strategies like discounts or promotions.

3. Future improvements:

- Experimenting with different clustering algorithms (e.g., DBSCAN, Hierarchical) to compare results.

- Incorporating additional features (e.g., recency, frequency) for a more nuanced segmentation.