Clustering Report

1. Number of Clusters Formed

After evaluating the Davies-Bouldin (DB) Index for different numbers of clusters, the optimal number of clusters was determined to be <Optimal_Number>.

- Clusters were chosen based on the lowest DB Index value, indicating better-defined and well-separated clusters.

---

2. DB Index Value

The Davies-Bouldin Index (DBI) for the optimal clustering configuration is <DB_Index_Value>.

- Interpretation of DB Index:
  - The DB Index measures the ratio of within-cluster scatter to between-cluster separation. Lower values are better and indicate more compact and distinct clusters.

---

3. Other Clustering Metrics

To further evaluate the clustering performance, additional metrics and observations are noted:

a) Silhouette Score:

- Measures the mean intra-cluster distance and inter-cluster distance.
- For this clustering, the silhouette score was computed as <Silhouette_Score> (ranges between -1 and 1; higher values are better).

b) Inertia (Within-Cluster Sum of Squares):

- Measures how tightly data points are grouped within clusters.
- For the chosen configuration, the inertia value is <Inertia_Value>.
- Lower inertia typically means clusters are more compact.

c) Cluster Distribution:

- The number of customers per cluster:

```
yaml
CopyEdit
Cluster 0: X customers
Cluster 1: Y customers
Cluster 2: Z customers
...
```

- Observations: The cluster sizes vary, indicating natural groupings among customers.

d) PCA Explained Variance:

- For visualization, PCA reduced the dimensions to 2 principal components.
- The explained variance for the first two principal components is <Explained_Variance>, indicating the proportion of total data variance preserved during dimensionality reduction.

---

4. Visual Representation

a) PCA-Based 2D Scatter Plot

- A scatter plot was generated to visualize customer segments in two dimensions using PCA. Customers in the same cluster are represented with the same color.
- Observation: Clusters are visibly distinct in 2D space, confirming the segmentation logic.

b) DB Index Trend

- A plot of the DB Index for 2 to 10 clusters revealed that the DB Index was minimized at <Optimal_Number> clusters, confirming the chosen configuration.

---

5. Business Implications

The clusters formed can be interpreted as distinct customer segments based on their spending behavior, quantity of purchases, transaction frequency, and regional distribution. These segments can be used for targeted marketing, personalized recommendations, or promotional strategies to enhance customer retention and revenue generation.