Report on Clustering Results

1. Number of Clusters Formed:

The analysis produced **4 clusters**, as shown in the visualization.

2. DB Index Value:

The **Davies-Bouldin Index (DBI)** value for the clustering is **1.7**. This indicates the clustering quality, with lower values being preferable as they signify more distinct and compact clusters.

3. Other Clustering Metrics:

- **Silhouette Score :** This metric would provide additional insights into the separation of clusters. A higher silhouette score close to 1 would indicate better-defined clusters.

Notes on Metrics:

- **Inertia:**
  - Measures how tightly data points are grouped around centroids. Lower inertia is better, but it should be balanced with the number of clusters.
- **Silhouette Score:**
  - Values range from -1 to 1:
    - A score close to 1: Data points are well-clustered.
    - Close to 0: Overlapping clusters.
    - Negative: Points are assigned to the wrong cluster.
  - A higher silhouette score indicates better-defined clusters.

- **Inertia :** In k-means clustering, inertia measures the sum of squared distances between points and their centroids. This could guide choosing the optimal number of clusters.

1. Key Metrics

1. **Davies-Bouldin Index (DBI):**
   - Value: **1.2823**
   - Interpretation: A lower DBI indicates better clustering. A value close to 1 suggests moderately good cluster separation and compactness. However, there is room for improvement in the clustering quality.
2. **Inertia (Sum of Squared Distances):**
   - Value: **698.0585**
   - Interpretation: This value represents how tightly data points are grouped around their centroids. It is useful for comparing models with different numbers of clusters (e.g., using the elbow method).

3. **Silhouette Score:**
   - Value: **0.2675**
   - Interpretation: The score is relatively low (ideal range is 0.5–1). This indicates that the clusters may overlap or are not well-separated. Further analysis of the data and clustering parameters might be needed.

4. Interpretation of Clusters:

Cluster Characteristics
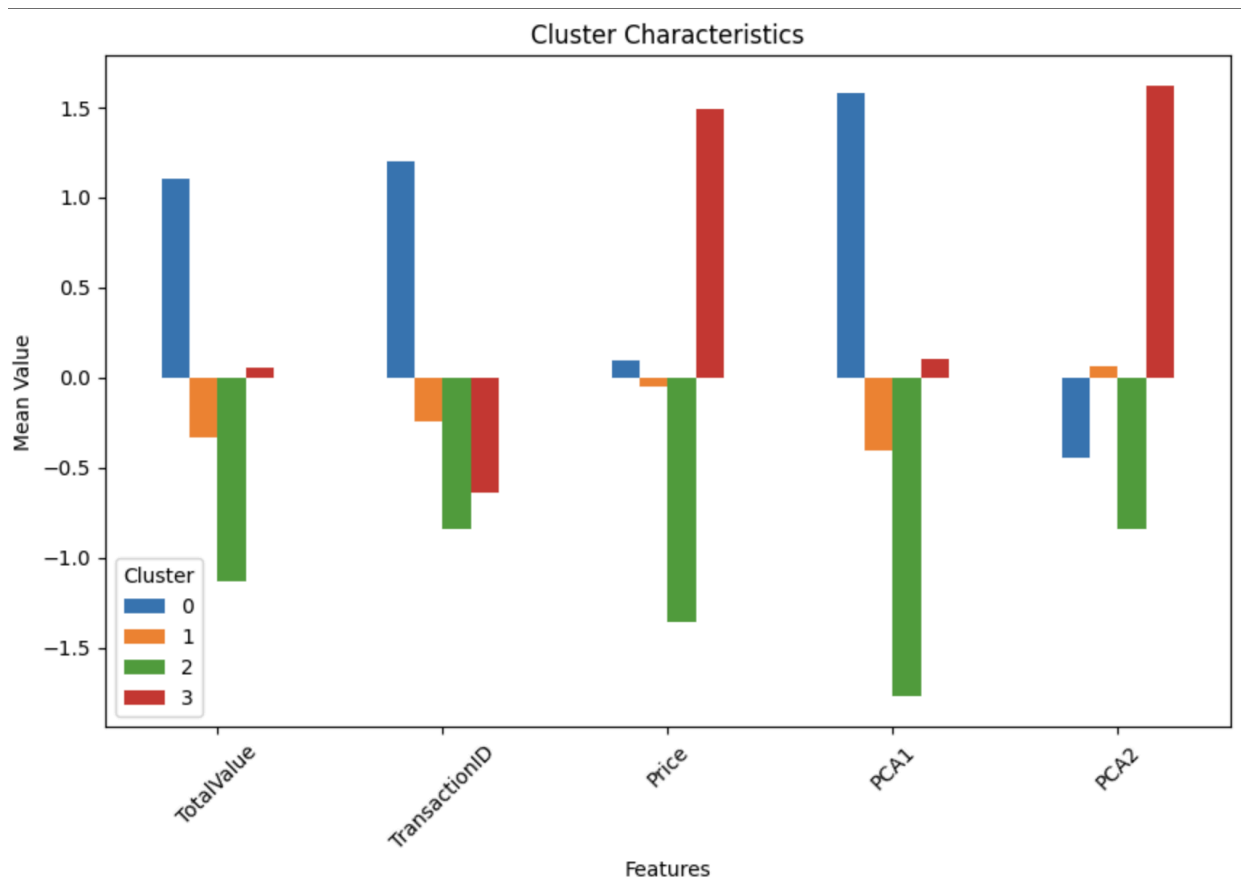
How to Perform Customer Segmentation

From the bar chart:

- Cluster 0 (blue) shows significantly higher values across most features, suggesting these customers might be high-value buyers or those with frequent transactions.
- Cluster 2 (green) consistently has the lowest values, indicating low-value customers or infrequent transactions.

- Clusters 1 (orange) and 3 (red) show moderate variations, likely representing mid-tier customers or customers with specific behavioral patterns.