Clustering Report

Number of Clusters: 5

The clustering algorithm identified 5 distinct clusters based on the input data. The number of clusters was optimized to ensure that the clustering structure best represents the inherent patterns within the data. The selection was guided by the evaluation of clustering metrics, particularly the Davies-Bouldin Index (DB Index).

Davies-Bouldin Index (DB Index): 0.5567

The DB Index is a metric that evaluates clustering quality based on two factors:

- **Compactness:** How tightly packed the data points are within a cluster.
- **Separation:** How distinct the clusters are from one another.

A lower DB Index indicates better clustering quality, with values closer to 0 signifying:

- Higher compactness of clusters.
- Greater separation between clusters.

In this case, the DB Index of 0.5567 suggests a strong clustering structure, with clusters being both well-separated and internally compact.

Silhouette Score: 0.6543

The Silhouette Score assesses how similar data points in a cluster are compared to points in other clusters. Its range is between -1 and 1:

- **1:** Perfectly defined clusters (high cohesion and separation).
- **0:** Overlapping clusters with minimal separation.
- **-1:** Misclassification of points in incorrect clusters.

A score of 0.6543 indicates that the clustering has a good structure; most points are closely related to their own cluster, and clear separation exists between clusters.

Cluster Sizes:

The distribution of points across the 5 clusters is as follows:

- **Cluster 2:** 190 points (largest cluster).
- **Cluster 1:** 59 points.
- **Cluster 0:** 54 points.
- **Cluster 3:** 50 points.
- **Cluster 4:** 46 points (smallest cluster).

The variation in cluster sizes suggests the data exhibits imbalanced groups, with some clusters being significantly larger than others.

Observations and Insights:

- **Cluster Quality:** The metrics (low DB Index and high Silhouette Score) confirm that the clustering model effectively identified meaningful groupings in the data.
- **Cluster Balance:** The presence of both large and small clusters indicates diversity in the dataset. Imbalanced clusters may signify that some groups in the data are naturally denser or more prevalent than others.
- **Application of Results:** The identified clusters can help uncover patterns or relationships within the data. These clusters may guide targeted interventions or deeper analysis.