Report on Clustering Results

1. Number of Clusters Formed

1. Optimal Number of Clusters: On the basis of evaluation metrics and clustering algorithm, 5 clusters were formed for customer segmentation. This number was selected so that we can get minimum DB Index.

2. DB Index (Davies-Bouldin Index)

1. DB Index Value: The Davies-Bouldin Index for the final clustering model is 0.65.

3. Other Clustering Metrics

1. **Silhouette Score:** Measures how similar an object is to its own cluster compared to other clusters. The value ranges from -1 to 1.

- 1: Perfect clustering (points are far from the decision boundary).
- 0: The point is on or very close to the decision boundary.
- -1: The point is likely misclassified (assigned to the wrong cluster).
- Interpretation: A higher silhouette score indicates better clustering.

2. **Dunn Index:** Measures the compactness and separation of clusters. It calculates the ratio of the minimum distance between points in different clusters to the maximum cluster diameter.

- Formula = min distance between clusters/ max diameter of clusters
- Interpretation: Higher values indicate better clustering (i.e., clusters are more compact and separated).

3. **Rand Index:** Measures the similarity between two data clusterings. It compares the agreement between pairs of elements in the clustering, considering whether they are in the same or different clusters in both clustering results.

    1. Formula = (a+b)/(a+b+c+d)

    - a = Pairs of points correctly clustered together.
    - b = Pairs of points correctly clustered apart.
    - c = Pairs of points wrongly clustered together.
    - d = Pairs of points wrongly clustered apart.

    2. Interpretation: A Rand Index closer to 1 indicates high similarity between clusterings.

4. Homogeneity, Completeness, and V-Measure:

- Homogeneity: Measures if each cluster contains only members of a single class.
- Completeness: Measures if all members of a single class are assigned to the same cluster.
- V-Measure: The harmonic mean of homogeneity and completeness, balancing both metrics.