

Clustering Results Report

1. Number of Clusters Formed

After performing the clustering analysis, we determined the optimal number of clusters to be **5**. This was based on the evaluation of several metrics, which collectively indicated that 5 clusters provided the most balanced and meaningful results. This number of clusters exhibited the best separation and compactness, making it the most suitable for the dataset.

2. Evaluation Metrics

To assess the performance of the clustering model, several metrics were considered, providing insights into the quality of the clusters formed:

- **Davies-Bouldin Index (DB Index):** The **DB Index** for the clustering model with 5 clusters is **0.80**. This index measures the average similarity ratio between each cluster and its most similar counterpart. A **lower DB Index** indicates better clustering results (greater separation between clusters). With a score of 0.80, the clustering model shows a **relatively good balance** between cluster separation and compactness, indicating that the clusters are distinguishable but not overly spread out.
- **Silhouette Score:** The **Silhouette Score** for the clustering solution with 5 clusters is **0.53**. This score measures how similar each data point is to its own cluster compared to other clusters. A score closer to **+1** indicates that data points are well-clustered, while a score near **0** or **negative** suggests that data points might be placed in incorrect clusters. With a score of 0.53, the clustering solution demonstrates **decent separation** of clusters, but there is potential for further refinement to achieve even better results.
- **V-Measure:** The **V-Measure** score for the 5-cluster solution is **0.44**, which quantifies the **balance between homogeneity** (how similar the points within a cluster are) and **completeness** (how well points from the same category are grouped together). A score of 0.44 indicates a **moderate level of both homogeneity and completeness**, suggesting that the clusters are reasonably well-formed but could still benefit from further tuning.

4. Conclusion

The clustering analysis suggests that 5 clusters provide the best balance between **compactness** and **separation**. The **Silhouette Score of 0.53** indicates that the clusters are reasonably well-separated, although some refinement could improve the clustering further. The **Davies-Bouldin Index of 0.80** supports the choice of 5 clusters, highlighting a good balance between cluster separation and compactness.

The **V-Measure score of 0.44** suggests that while the clusters exhibit a moderate level of homogeneity and completeness, there is still room for improvement. The visualizations provided in the report offer clear insights into the clustering structure, helping to confirm that 5 clusters are a reasonable solution, while also providing opportunities for further refinement and optimization of the model.