

## Clustering Results Report

### *Number of Clusters Formed:*

The clustering analysis was performed using **KMeans** with a range of cluster numbers from **2 to 10**. Based on the evaluation metrics, several cluster configurations were tested, and **5 clusters** emerged as an alternative clustering solution.

### *Davies-Bouldin Index (DB Index):*

The **Davies-Bouldin Index** measures the average similarity ratio between each cluster and the cluster most similar to it. A lower DB Index value indicates that the clusters are well-separated. For the **5-cluster solution**, the **Davies-Bouldin Index** was calculated as:

- **DB Index for 5 clusters: 1.2289**

This DB Index suggests that the 5 clusters are more distinctly separated compared to the 4-cluster solution, with a lower index value indicating improved cluster cohesion.

### *Other Clustering Metrics:*

1. **Silhouette Score:** The **Silhouette Score** measures how similar each data point is to its own cluster in comparison to other clusters. A higher score indicates well-separated clusters. For the **5-cluster solution**, the **Silhouette Score** was:
  - a. **Silhouette Score for 5 clusters: 0.2442**

A Silhouette Score of **0.2442** is relatively low, indicating that although the clusters are distinguishable, the data points within the clusters are not strongly cohesive. This suggests there is room for improvement in the cluster structure.

2. **Calinski-Harabasz Index (CH Index):** The **Calinski-Harabasz Index** evaluates the overall quality of the clustering by comparing the within-cluster variance to the between-cluster variance. A higher value indicates better clustering. For the **5-cluster solution**, the **Calinski-Harabasz Index** was:
  - a. **Calinski-Harabasz Index for 5 clusters: 83.35**

This relatively lower CH Index indicates that while the clusters are somewhat distinct, the overall clustering structure is not optimal and may require further refinement.

### ***Summary of DB Index for Different Cluster Numbers:***

Here is a summary of the **Davies-Bouldin Index** for cluster configurations from **2 to 10 clusters**:

- **2 clusters: 1.2236**
- **3 clusters: 1.4923**
- **4 clusters: 1.3811**
- **5 clusters: 1.2289**
- **6 clusters: 1.4584**
- **7 clusters: 1.4522**
- **8 clusters: 1.4416**
- **9 clusters: 1.4601**
- **10 clusters: 1.4147**

The **lowest DB Index** of **1.2236** was achieved with **2 clusters**, suggesting that 2 clusters provide the most distinct separation. However, as the number of clusters increases, the DB Index increases, indicating a slight deterioration in cluster separation, although it stabilizes after 5 clusters.

### ***Conclusion:***

- The **5-cluster solution** offers a better cluster separation with a DB Index of **1.2289** compared to other configurations, though the **Silhouette Score** and **Calinski-Harabasz Index** suggest that further improvements can be made.
- The **2-cluster solution** performed best in terms of the Davies-Bouldin Index, but it did not provide a satisfactory structure based on other metrics.
- **5 clusters** represent a reasonable balance between cluster coherence and separation. However, the clustering structure could be further optimized by considering alternative clustering algorithms or tuning hyperparameters.