Clustering Report

This report presents the results of clustering analysis using multiple metrics for evaluation. The dataset was clustered into different numbers of groups (k), ranging from 2 to 10 clusters. The clustering performance was evaluated using the following metrics:

- **Davies-Bouldin Index (DB Index)**: Measures the compactness and separation of clusters. Lower values indicate better-defined clusters.
- **Silhouette Score**: Indicates the compactness and separation of clusters. Higher values suggest better clustering quality.
- **Calinski-Harabasz Index**: Evaluates the variance ratio within and between clusters. Higher values are preferred for well-separated clusters.

Based on the evaluation, **10 clusters** were identified as the optimal number of clusters across all three metrics.

Optimal Number of Clusters

- **Based on DB Index**: 10 clusters (DB Index = 0.790706)
- **Based on Silhouette Score**: 10 clusters (Silhouette Score = 0.435032)
- **Based on Calinski-Harabasz Index**: 10 clusters (Calinski-Harabasz Index = 117.686208)

Clustering Metrics Summary

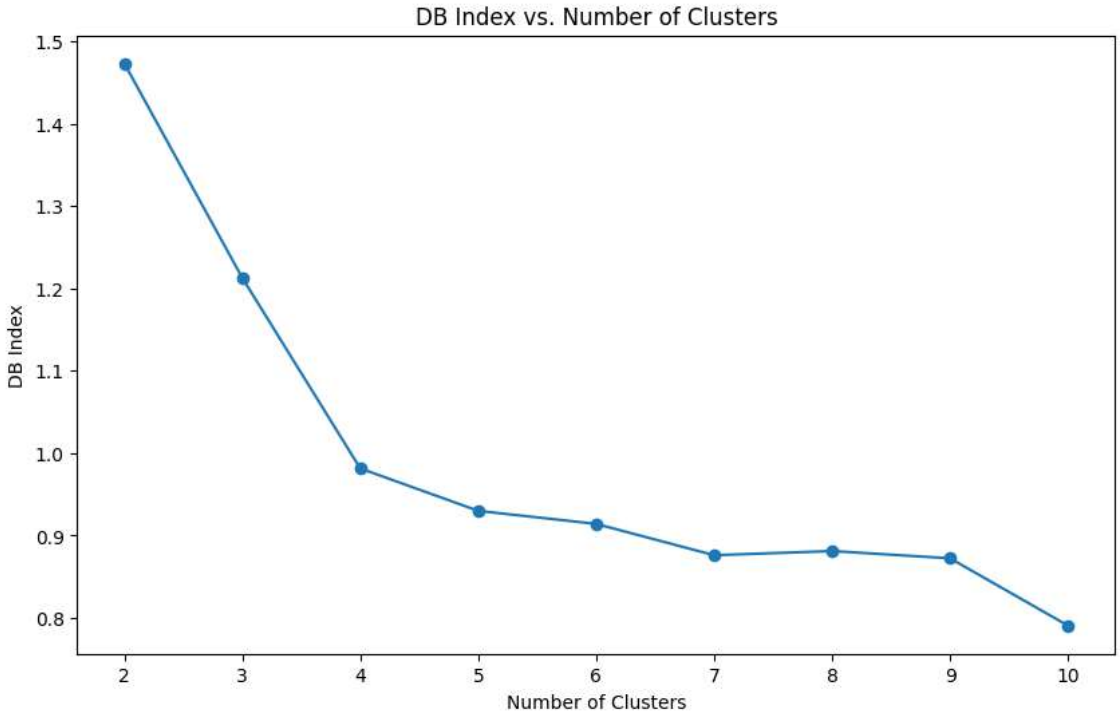
Number of Clusters (k)	DB Index	Silhouette Score	Calinski-Harabasz Index
2	1.473033	0.262692	83.885747
3	1.212701	0.315193	83.851244
4	0.981530	0.379724	95.655821
5	0.929999	0.405491	105.422756
6	0.914181	0.403884	101.165903
7	0.876148	0.407472	102.304994
8	0.881209	0.419258	104.067523
9	0.872415	0.414100	107.957894
10	0.790706	0.435032	117.686208

Key Observations

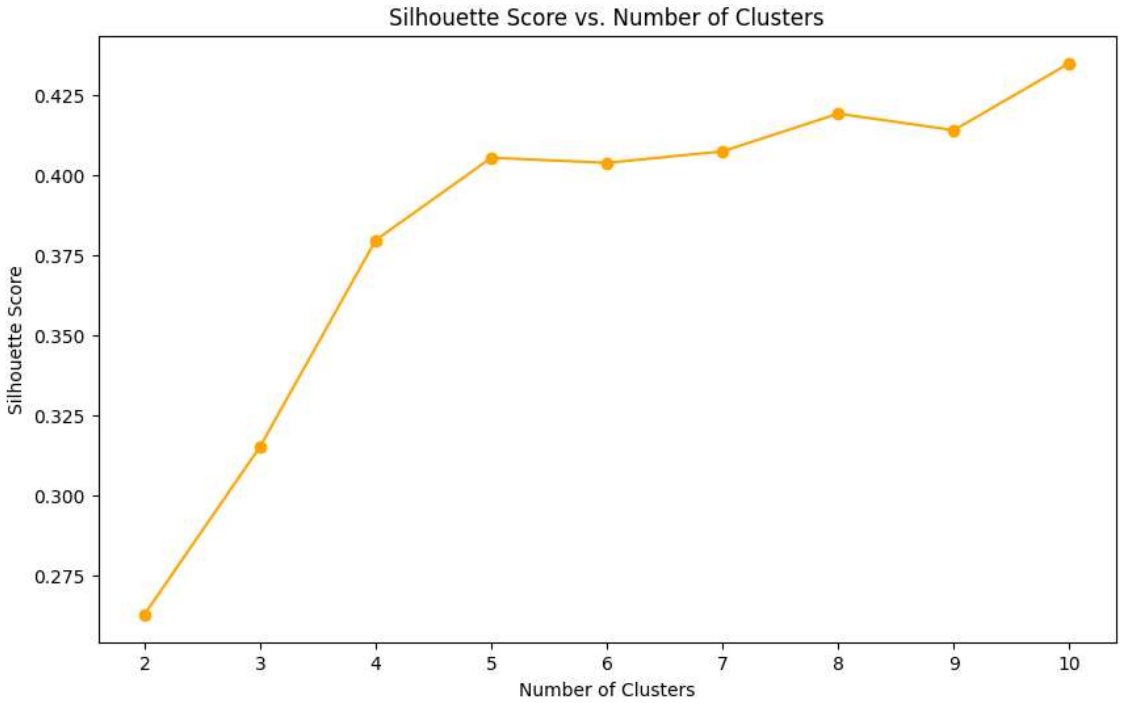
1. **Davies-Bouldin Index**:
    - The DB Index improves (decreases) as the number of clusters increases, reaching its lowest value at 10 clusters (0.790706). This indicates the best separation and compactness of clusters.
2. **Silhouette Score**:
    - The Silhouette Score increases steadily, peaking at 10 clusters (0.435032). This suggests that 10 clusters provide the highest intra-cluster similarity and inter-cluster separation.
3. **Calinski-Harabasz Index**:
    - The Calinski-Harabasz Index also achieves its maximum value with 10 clusters (117.686208), indicating well-separated clusters with minimal within-cluster variance.

Visualization of Clustering Metrics

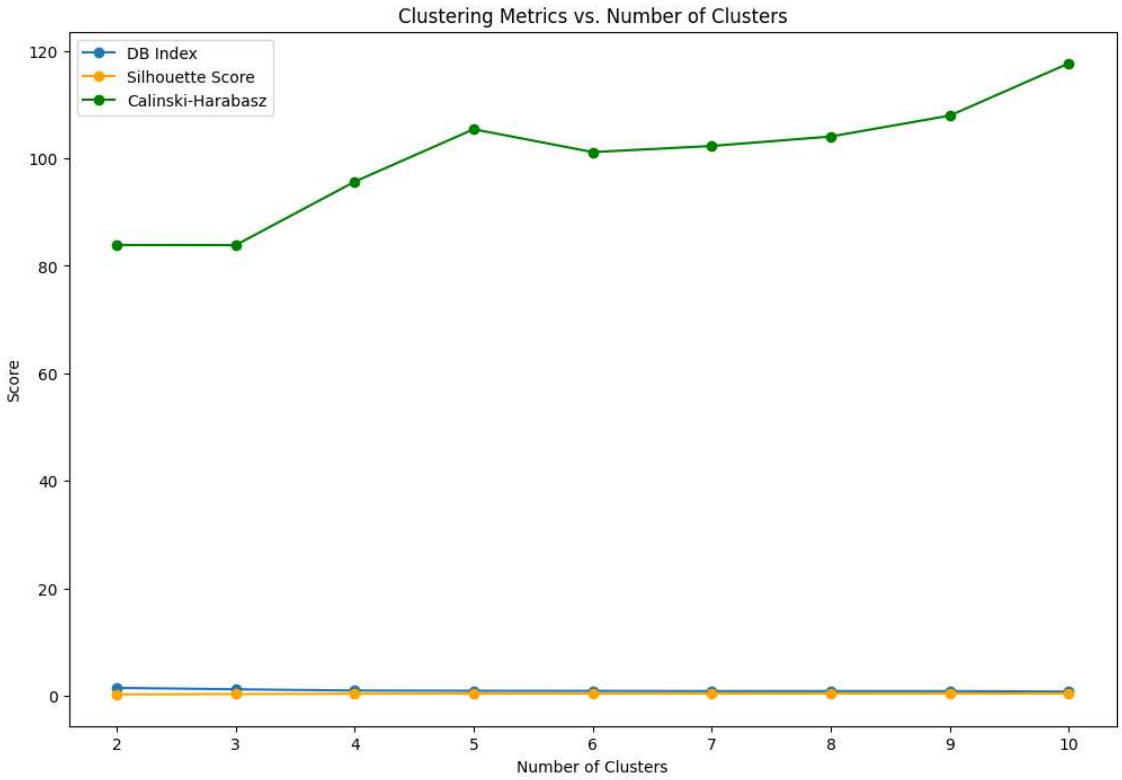
1. Davies-Bouldin Index vs. Number of Clusters



2. **Silhouette Score vs. Number of Clusters**



3. **Calinski-Harabasz Index vs. Number of Clusters**


Clustering Metrics vs. Number of Clusters
DB Index
Silhouette Score
Calinski-Harabasz
Score
Number of Clusters
120
100
80
60
40
20
0
2
3
4
5
6
7
8
9
10