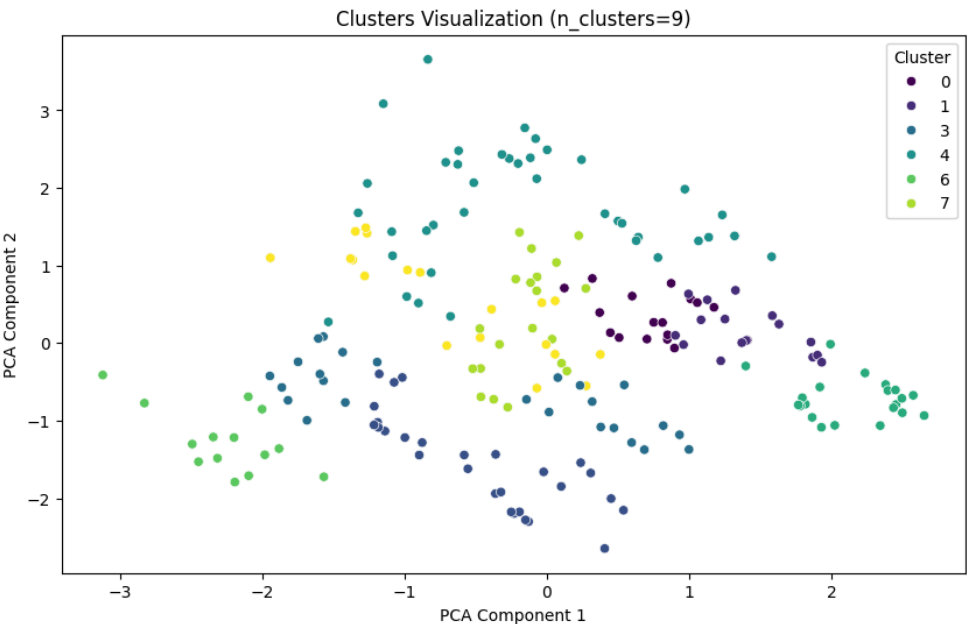


Clustering Analysis Report Using KMeans

Objective: The goal of this analysis is to evaluate the clustering performance for different numbers of clusters using KMeans and visualize the results.

1. KMeans Clustering Results

- Visualization of Clusters:**
The first plot shows the clustering results using KMeans with 9 clusters, visualized in a 2D PCA space. Each cluster is represented by different colors. The clusters are relatively well-separated, with some concentration of points around certain regions.



2. Clustering Evaluation Metrics

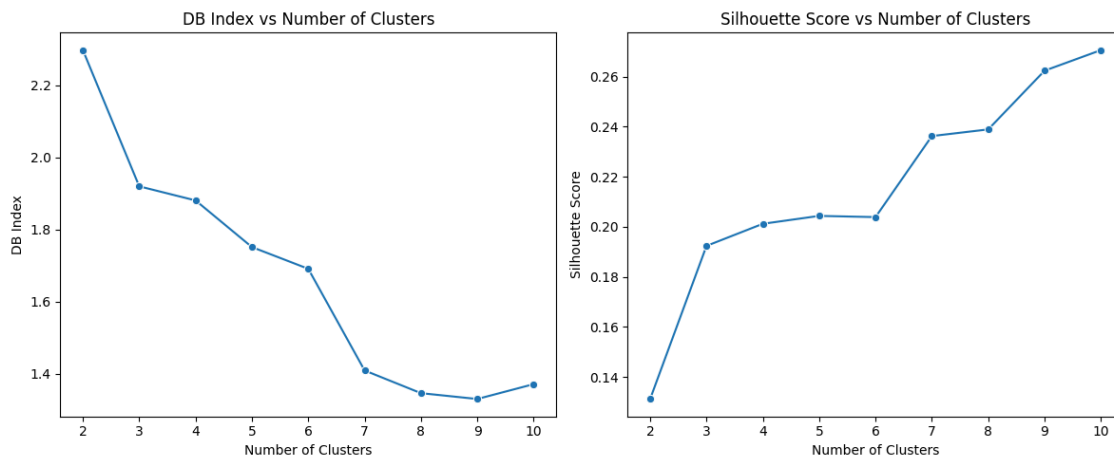
We evaluated the clustering performance based on two metrics:

- Davies-Bouldin Index (DB Index):** This index measures the compactness and separation of clusters. A lower value indicates better clustering.
- Silhouette Score:** A higher score indicates that the clusters are well-separated and well-formed.
- DB Index vs Number of Clusters:**
The DB Index decreases as the number of clusters increases. A sharp drop is observed

around 4 clusters, indicating that the clustering quality improves as more clusters are used, with the best performance around 9 clusters.

- **Silhouette Score vs Number of Clusters:**

The silhouette score increases as the number of clusters rises, suggesting that the clusters are becoming more distinct and better formed with a higher number of clusters.



3. Optimal Number of Clusters

Based on the DB Index and Silhouette Score, the optimal number of clusters appears to be 9, as the DB Index reaches its minimum and the silhouette score is higher in this range.

4. Next Steps

The results indicate that KMeans clustering with 9 clusters provides the most compact and well-separated clusters. Future work can involve further tuning of the KMeans parameters, such as initialization methods and convergence criteria, to improve clustering performance.