**Clustering Analysis Report**

**This report outlines the clustering analysis performed using the Wine dataset, focusing on the application of two clustering algorithms: KMeans and DBSCAN. The analysis compares their performance on both scaled and PCA-transformed datasets. The primary evaluation metrics include Silhouette Score and Davies-Bouldin Index, which assess the quality of clustering.**

**1. Dataset Overview**

* **Dataset: Wine dataset (sklearn)**
* **Features: 13 numerical features**
* **Target: 3 wine classes (0, 1, 2)**

**2. Dimensionality Reduction**

* **PCA was used to reduce dimensionality while retaining variance.**
* **Number of Components: Initially reduced to 5 components to capture ~80% variance. Further reduced to 2 components for visual analysis.**

**3. Clustering Algorithms**

**KMeans**

* **Objective: Partition data into clusters by minimizing intra-cluster variance.**
* **Best Number of Clusters (k): 3 (determined using Silhouette Score and Davies-Bouldin Index).**

| **Metric** | **Scaled Data** | **PCA Data** |
| --- | --- | --- |
| **Silhouette Score** | **0.284859** | **0.369076** |
| **Davies-Bouldin Index** | **1.329223** | **0.690010** |

* **Observation: PCA-transformed data outperforms scaled data in both Silhouette Score and Davies-Bouldin Index, indicating better clustering quality.**

**DBSCAN**

* **Objective: Identify clusters of varying densities based on neighborhood parameters (eps and min\_samples).**
* **Best Parameters:** 
  + **Scaled Data: No valid clusters found.**
  + **PCA Data: eps = 0.9, min\_samples = 6**

| **Metric** | **Scaled Data** | **PCA Data** |
| --- | --- | --- |
| **Silhouette Score** | **-1.000000** | **0.447933** |
| **Davies-Bouldin Index** | **inf** | **0.690010** |

* **Observation:** 
  + **DBSCAN fails to identify meaningful clusters on scaled data, as indicated by a Silhouette Score of -1 and an infinite Davies-Bouldin Index.**
  + **For PCA-transformed data, DBSCAN achieves better clustering performance with a Silhouette Score of 0.447933, surpassing KMeans.**

**4. Visual Analysis**

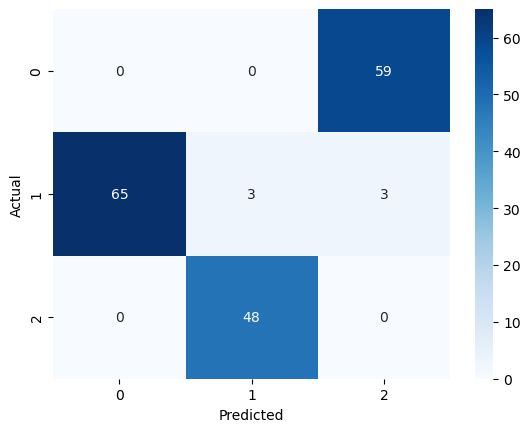
* **PCA with 2 Components:** 
  + **Clusters formed by KMeans and DBSCAN were plotted for visualization.**
  + **DBSCAN identified clusters with more flexibility compared to the rigid structure of KMeans.**

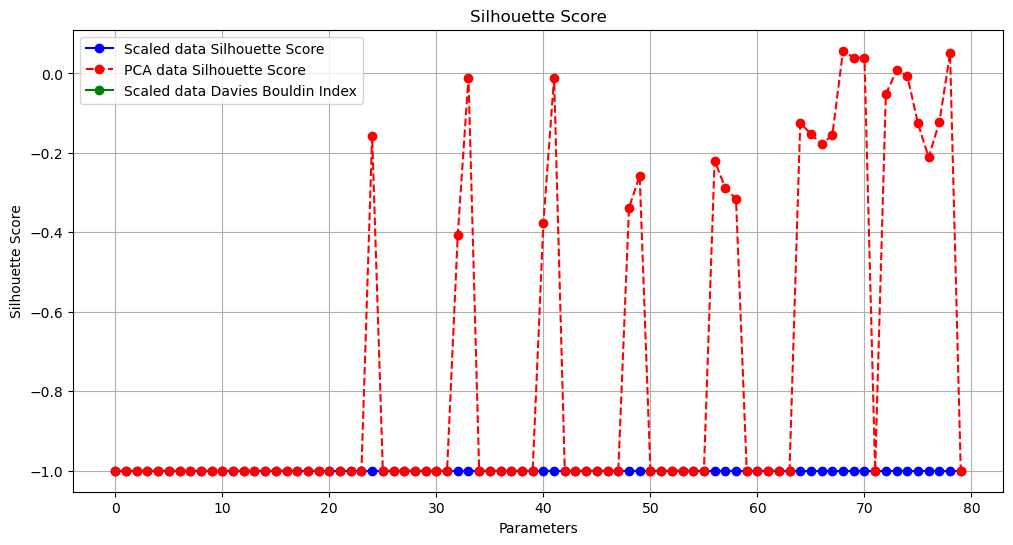
**5. Key Findings**

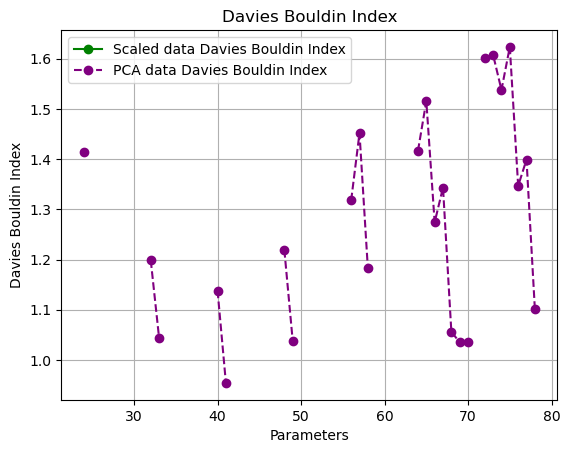
1. **KMeans Results:**
   * **Achieves reasonable clustering with the best performance on PCA data.**
   * **PCA significantly improves cluster quality, as reflected by a lower Davies-Bouldin Index (0.690010) and higher Silhouette Score (0.369076).**
2. **DBSCAN Results:**
   * **Scaled data does not support meaningful clustering with DBSCAN.**
   * **PCA-transformed data, using eps=0.9 and min\_samples=6, yields the best DBSCAN clustering with a Silhouette Score of 0.447933.**
3. **Overall Comparison:**
   * **PCA-transformed data performs better across all metrics and algorithms.**
   * **DBSCAN on PCA data surpasses KMeans in clustering quality (higher Silhouette Score).**

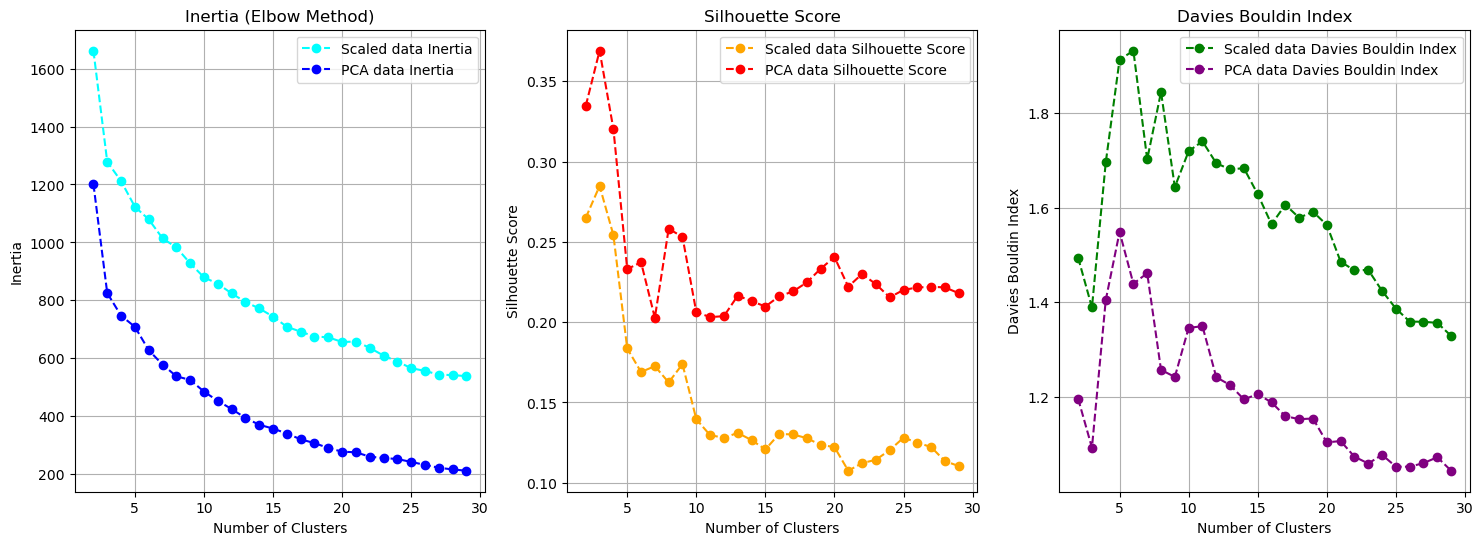
**6. Conclusion**

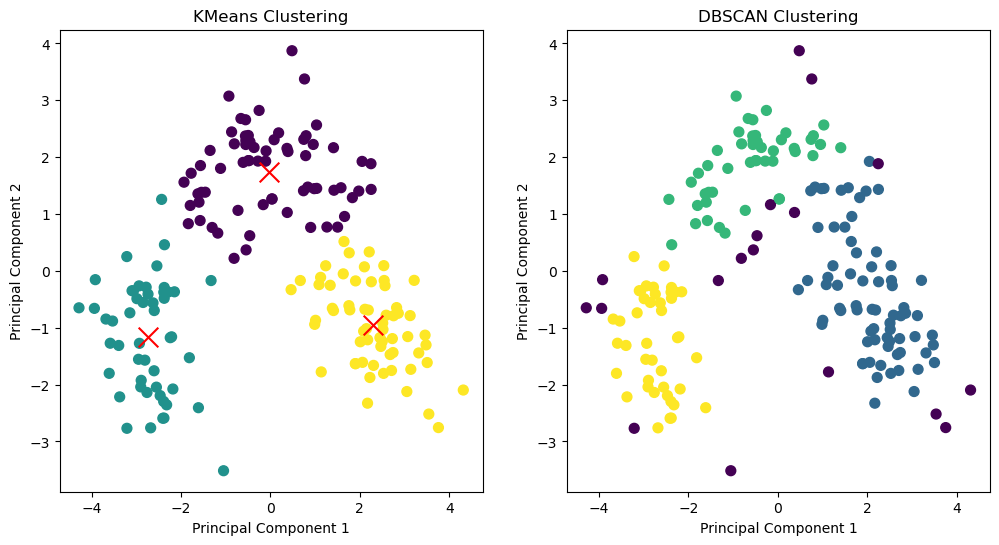
* **PCA transformation is crucial for improving clustering performance in this analysis.**
* **While KMeans performs reliably, DBSCAN excels in identifying more flexible clusters when applied to PCA data.**
* **For the Wine dataset, the best clustering approach is DBSCAN with PCA-transformed data, achieving the highest Silhouette Score of 0.447933.**

****

****

****

****

****