Wine Clustering Analysis Report

1. Main Objective

The primary goal of this analysis is to cluster wine data based on its quality, enabling wineries or distributors to better understand product characteristics. The analysis employs clustering techniques to group wines effectively, aiding quality control and market segmentation strategies.

2. Dataset Description

The dataset includes 12 attributes related to physicochemical properties and quality assessments of red and white wines. Key attributes are acidity, sugar content, pH, and alcohol. The target variable, quality, ranges from 3 to 9. Clustering the data can provide actionable insights for wineries to optimize processes and target specific markets.

3. Data Exploration and Preprocessing

• The dataset was clean with no missing values.

• color (categorical) was label-encoded.

• All numerical features were standardized using StandardScaler.

• Dimensionality reduction via PCA was applied for visualization and computational efficiency, reducing features to two principal components.

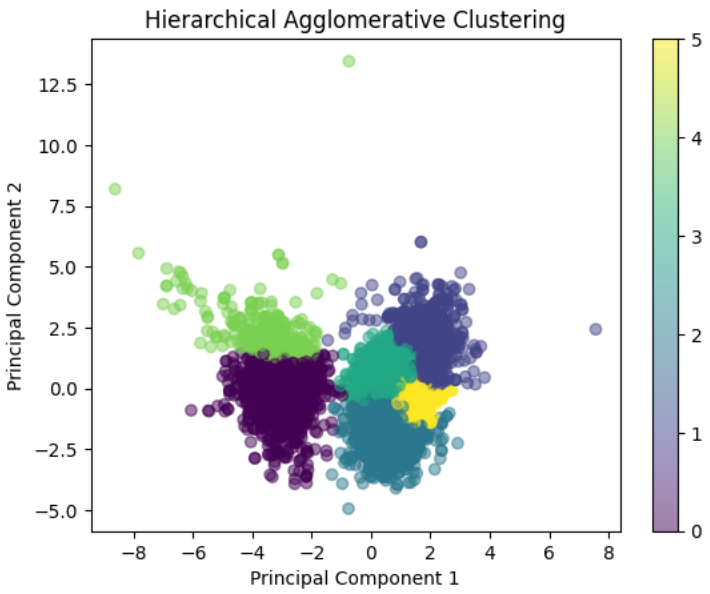
4. Clustering Models

Three clustering methods were applied:

1. **Hierarchical Agglomerative Clustering (HAC)**:

   • Used Ward linkage.

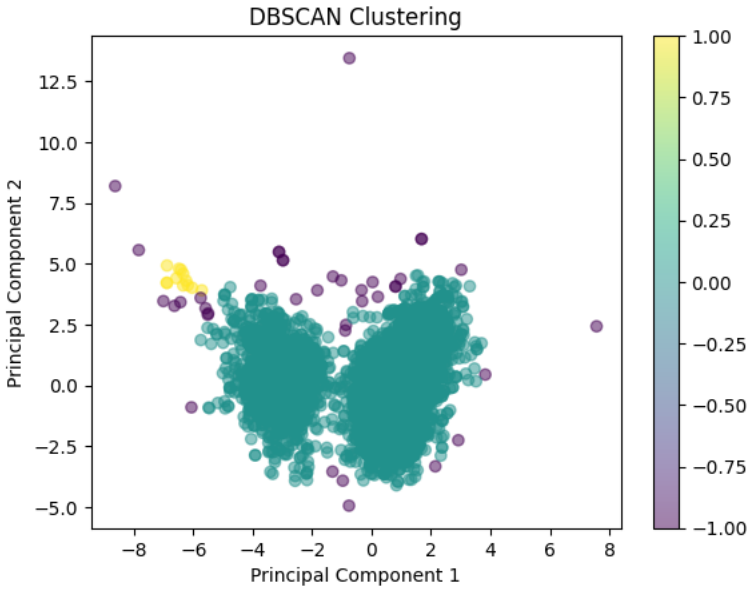
   • Identified 6 clusters with a silhouette score of **0.32**.



2. **DBSCAN**:

   • Tuned eps=0.5 and min_samples=5.

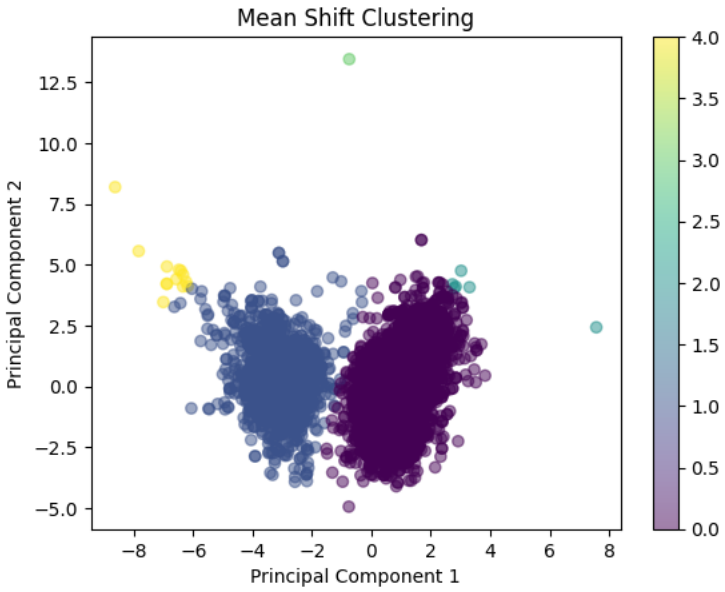
   • Produced 3 clusters but was less effective, with noise points dominating.



3. **Mean Shift**:

   • Automatically determined the optimal number of clusters.

   • Achieved a silhouette score of **0.46**, outperforming HAC.



5. Recommended Model

Mean Shift clustering is recommended due to its ability to identify clusters with minimal noise and the highest silhouette score, making it robust for this dataset.

6. Key Findings and Insights

• The clustering revealed distinct groupings of wine based on quality and physicochemical properties.

• High-quality wines (quality ≥ 7) tend to cluster separately, driven by features like higher alcohol content and lower residual sugar.

7. Suggestions for Next Steps

1. Incorporate additional features like grape variety or fermentation techniques for improved clustering.
2. Experiment with advanced clustering methods like Gaussian Mixture Models for comparison.
3. Use clustering results to guide supervised learning, predicting wine quality based on cluster membership.