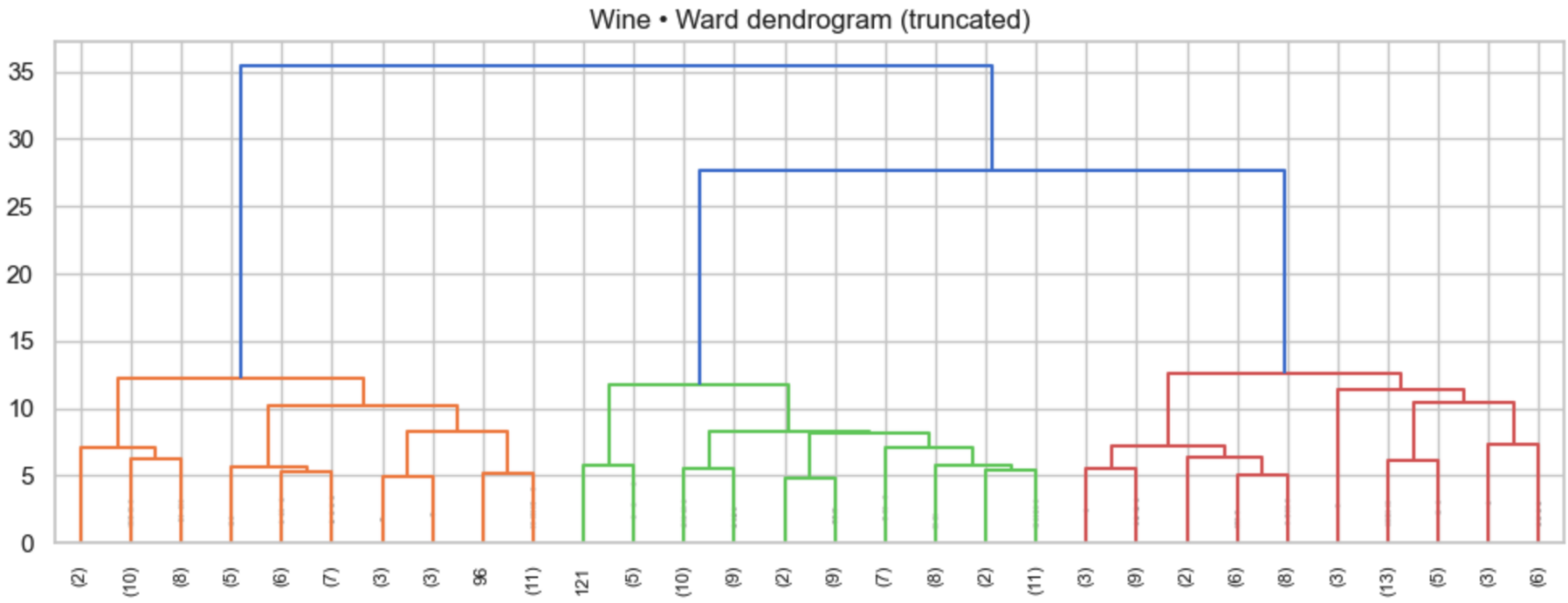
Clustering Wine Data with Agglomerative Clustering

Dataset & Preparation

- Dataset: 178 wines, 13 chemical features, 3 true cultivars
- StandardScaler: normalize features to equalize scale which ensures fair distance-based clustering

Ward Dendrogram



Ward Dendrogram

- Built hierarchical tree using Ward linkage
- Visualization: Dendrogram (truncated for clarity)
- Shows 3 major branches

Silhouette vs n_clusters

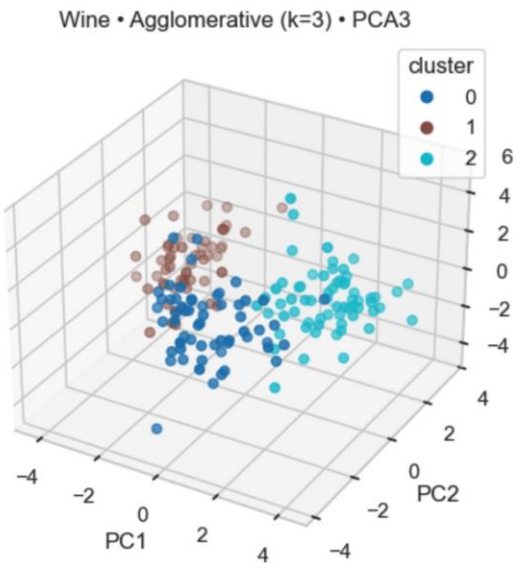
Silhouette vs n_clusters

- Tested 2–9 clusters with Agglomerative (Ward)
- Visualization: Silhouette plot
- Silhouette peaked at k=3 (~0.28)
- Confirms 3 clusters as best choice

Agglomerative Clustering (k=3)

- Ran AgglomerativeClustering with k=3 (Ward)
- Assigned cluster labels to all wines
- Matches dendrogram + silhouette insights

Visualization with PCA (3D)



- Reduced 13 features data into 3 PCA
- Visualization: 3D scatter plot (PC1, PC2, PC3)
- One cluster well separated, two overlapping

Evaluation

- Silhouette score: 0.277 (moderate separation)
- Adjusted Rand Index (ARI): 0.79 (strong match to true cultivars)
- Confirms meaningful clustering

Cluster Profiling

- Profiled clusters by mean z-scores of features
- **Cluster 0 →** Lighter wines (low alcohol, low color intensity)
- **Cluster 1 →** Acidic wines (high malic acid, low flavanoids)
- **Cluster 2 →** Rich wines (high alcohol, high flavanoids, high proline)
- Visualization: Heatmap of feature means

Takeaways

- Used dendrogram + silhouette to pick cluster count
- Scaling + PCA are critical for clustering
- Visualizations (dendrogram, silhouette, PCA 3D, heatmap) tell the story
- Clustering revealed distinct chemical profiles of wines