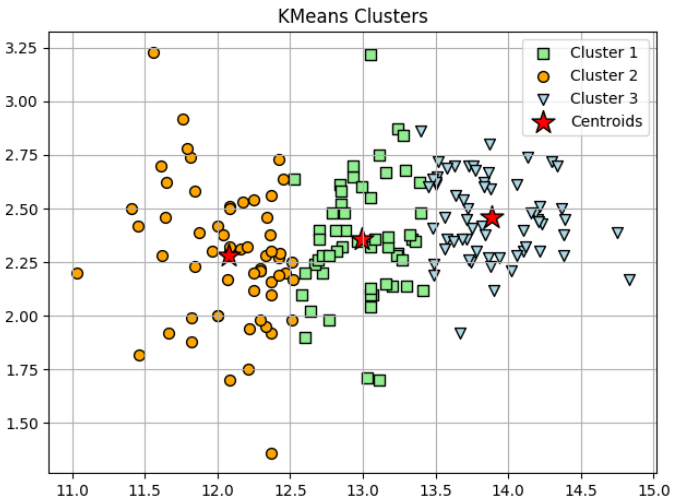
Small Project 2

CS4275- Machine Learning Foundations

KMeans Clusters:

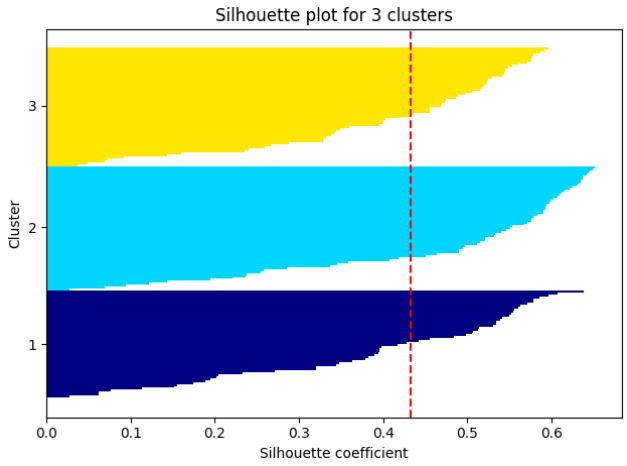
The picture beside shows the cluster formation and the centroids with KMeans algorithm. This forms three clusters meaning we get three different types of wine.



Quantifying the quality of clusters via silhouette plots:

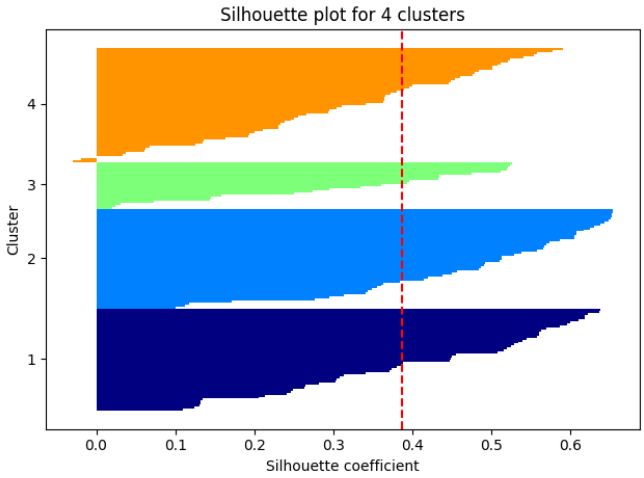
For 3 clusters:

The picture beside shows the silhouette plots for the three clusters. The picture shows good quality of clusters as it doesn't have any negative values graphed.



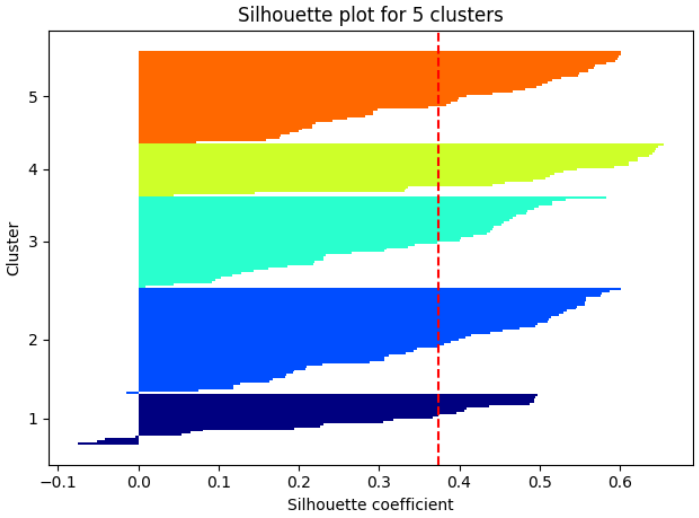
For 4 clusters:

In this picture, we can see there is a negative value graphed, which means it was assigned to a wrong cluster. We can also see the red dotted line shifting closer to 0. It is still good quality however; it could be better.



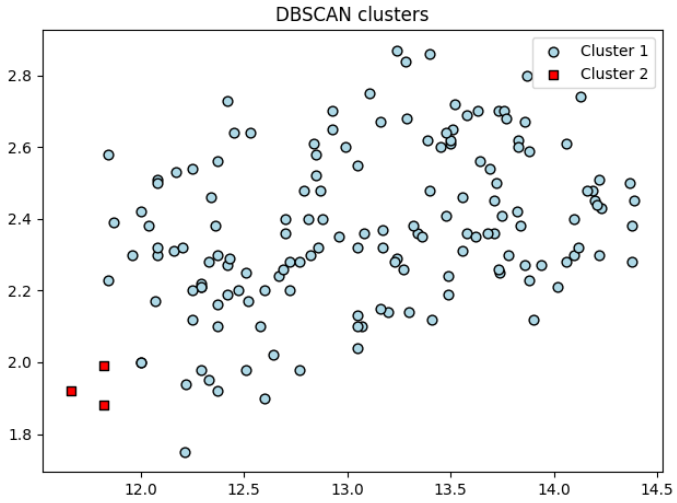
For 5 clusters:

The picture has more clusters on the negative side. And the has slightly less quality than 4 clusters. Adding clusters leads to worsening the quality of clusters could be because of the overlap or misclassification as noted by the clusters in the negative. I think 3 clusters are of the best quality.



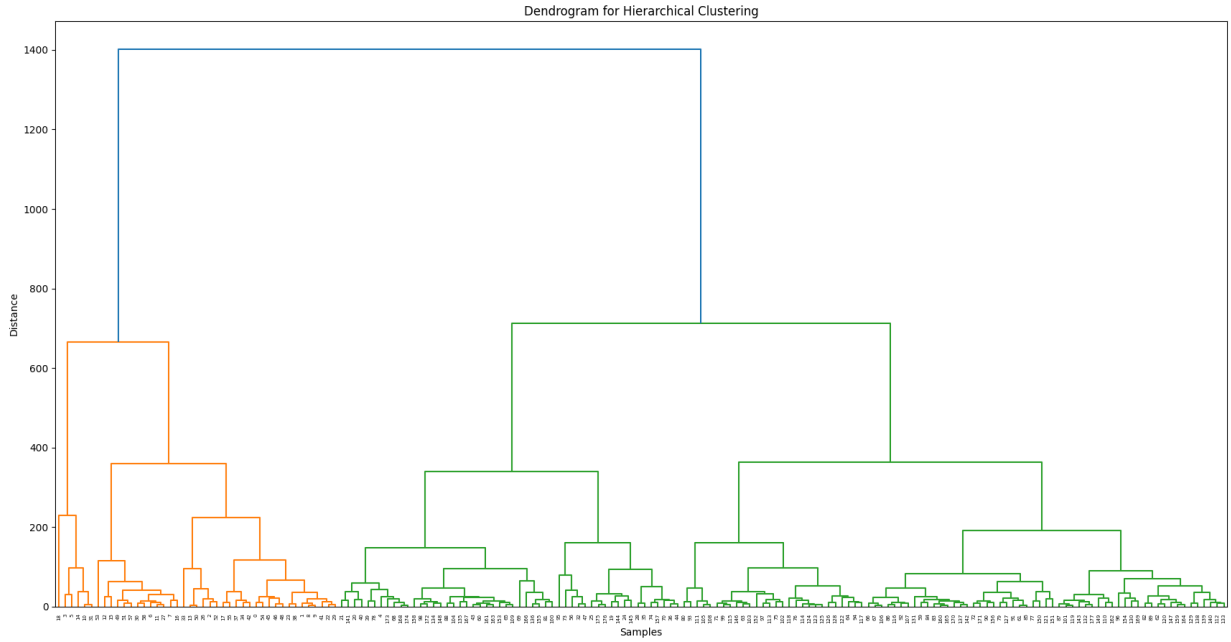
DBSCAN Clusters:

DBSCAN only finds two clusters probably because the density isn't very different. It forms only two clusters. I have eps set at 0.2 and min_samples at 5. Even then it only has three points in the cluster 2, which could mean that these points are too sparse to meet the density criteria. When eps is decreased to 0.1, it does form two clusters with some points each however, the distance is larger. Whereas increasing eps merges two clusters to cluster1. The distance metric used is Euclidean.



Hierarchical Clustering Dendrogram:

The picture below is the dendrogram for hierarchical clustering. The x-axis has the samples, and the y-axis is the distance between samples. The larger distance means a lot of dissimilarity hence the clusters are too far apart.



Agglomerative Clustering:

For the agglomerative clustering, it gives three cluster labels,

Cluster labels: [1 0 0 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 1 0 0 0 0 0 1 1 1 1 1 1 0 1 0 0 0 1 1 0 1 0 0 1 1 1 1 0 1 1 1 1 1 1 0 1 0 0 0 1 0 2 2 0 0 0 0 2 1 0 0 2 2 0 2 0 0 0 0 2 0 2 0 2 2 2 2 2 2 2 0 2 2 0 2 0 0 2 0 0 2 2 0 0 2 0 0 2 2 0 2 2 2 2 2 0 0 0 2 2 2 0 2 2 0 2 0 2 0 0 0 0 0 0 2 2 0 0 0 0 0 1 2 0 1 0 0 0 0 0 0 0 0 0 1 2 1 0 2 1 0 0 1 1 0 0 0 0 2 0 1 1 0 0 0 1]