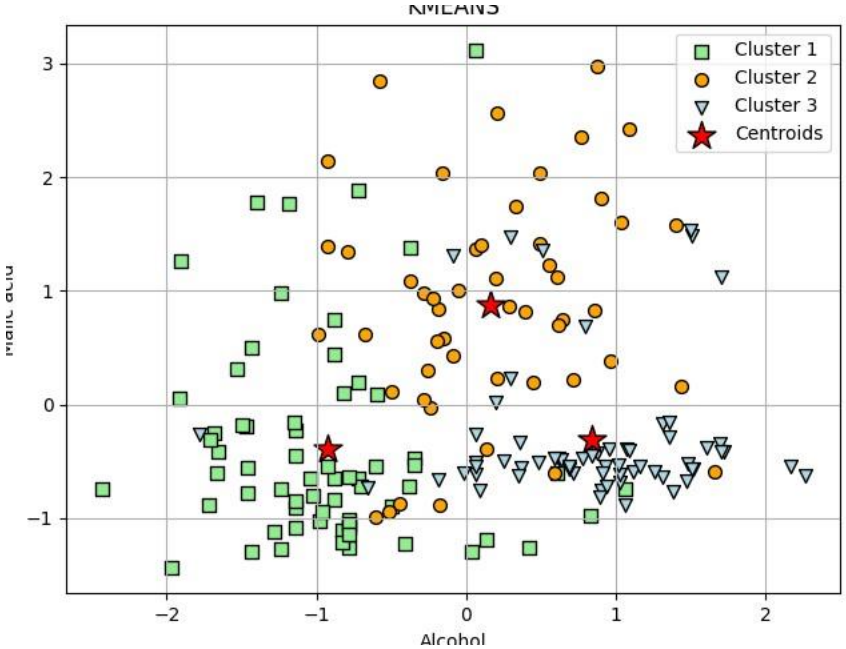
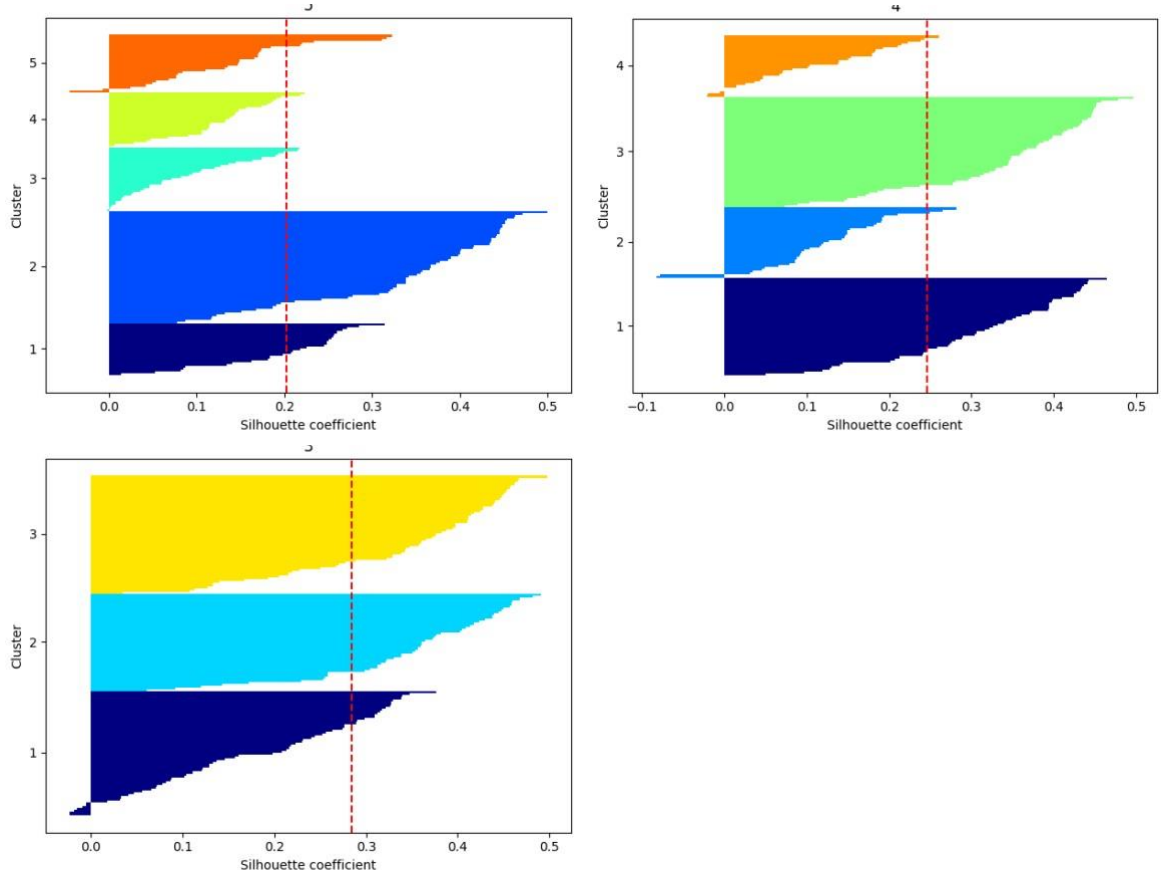
Daniel Rivera SM2



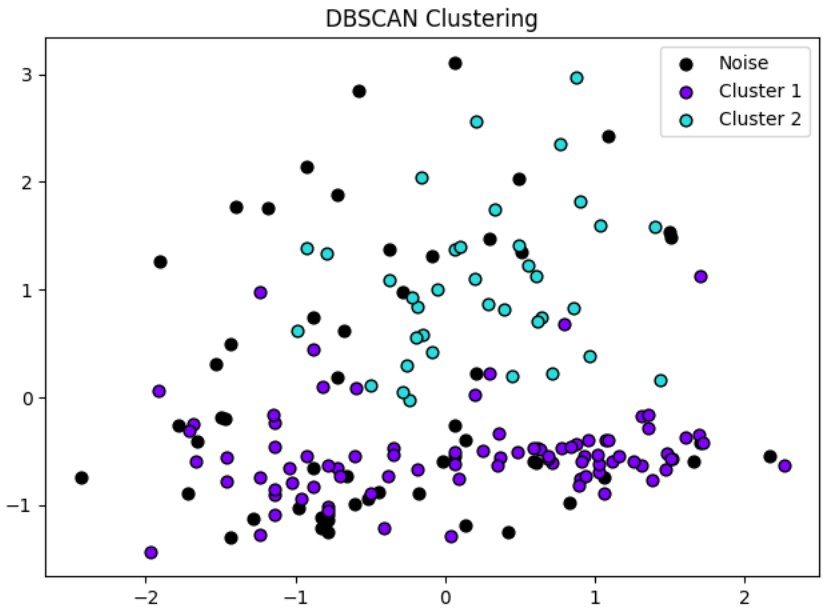
The clustering for K-Means was pretty simple. At first, I was just trying to guess at how many clusters just seeing what worked and what didn't. Soon after I ran the silhouette charts it confirmed the more correct cluster numbers should be around 3 except for one of them which showed 2 as the best clusters. The K-Means was a little strange I feel like the outliers pulled the centroids just a little bit further away from where I would've put them.

Daniel Rivera SM2



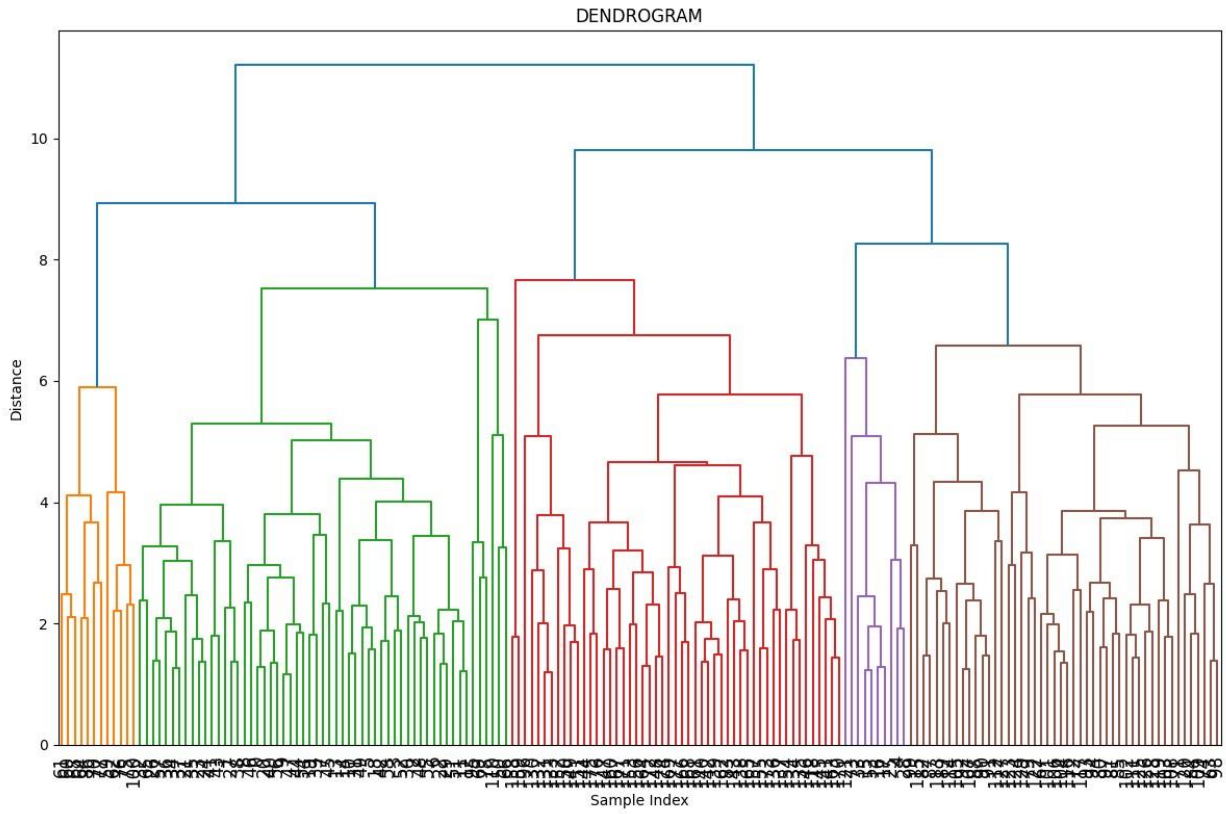
The Silhouette charts were pretty helpful after the K-Means in determining the most efficient clusters for the later algorithms. It is weird to see the coefficient become negative showing that some of the point assignments were wrong, however it isn't completely foreign to me how that can work since it is learning how where to assign them.

Daniel Rivera SM2



The DBSCAN clustering was by far the worst for me I spent a lot of time trying to fine turn the epsilon and the min_samples. I was going through different decimal values and higher/lower minimum samples. However, eventually I found a solid middle ground of eps=2.2 and min_samples=5. I still prefer the Kmeans for this particular dataset. I do see the advantages of the DBScan but in this case other algorithms were much better.

Daniel Rivera SM2



The Hierarchical clustering was by far the most confusing portion to actually write for me, however it also turned out to be the simplest algorithm. I found this to be the most confusing chart to try to interrupt. The 5 clusters where they are most specific is what I would want to use however 5 clusters seemed to also be the worst. And when I read the chart at the 3 blue higher clusters seemed to work the best, but my first thought is that it would be way to generic.