Stephanos, Dembe Koi

Project 4: write up

# Part 1

Overall, the clusters for both the red and white wine datasets do not in anyway reflect the different numerical categories quality. This is because the datasets are hardly equally distributed for red or white wines, and in fact, the vast majority of all ratings were either a 5 or 6 for both datasets, with the mode and mean scores of 5, 5.6 and 6, 5.9 respectively. As a result, some clusters do appear to represent the outliers, and do have a notably different mean quality score, but the vast majority of identified clusters across both datasets have very similar quality scores, reflective of the averages within their given samples. A summary of the data and clusters can be found below:

Table

|  |  |  |
| --- | --- | --- |
|  | Red Wine | White Wine |
| Min Rating | 3 | 3 |
| Max Rating | 8 | 9 |
| Mode Rating | 5 | 6 |
| Average Rating | 5.6 | 5.9 |

Figure

Figure

# Part 2