# COSC2670: Practical Data Science

# Assignment 2: Modelling wine quality based on physicochemical tests

Due midday on Monday, 22 May 2017

Submitted by:

Casey-Ann Charlesworth (3132392)

# Table of Contents

# Table of Tables



# Table of Figures

## 1. Executive summary

The aim of this report was to investigate whether wine quality (both red and white varieties) can be modelled based on a number of physicochemical tests. The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. [1][2] Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.). [2] Overall, due to the unbalanced nature of the class variable (there were many more “average” wines compared to “bad” or “excellent” wines), the classification models employed (Decision tree, K Nearest Neighbour, and Naïve Bayes) struggled to model and predict the quality of wine with greater than 60% accuracy.

The report concludes that due to the unbalanced datasets, and that the datasets likely represent wine ratings the world over (given that there are many more “average” wines compared to “bad” or “excellent” wines), no classifying algorithm can accurately predict the quality rating of a given wine at a sufficient rate. It is therefore recommended that the datasets be modified to balance the ratings more evenly and thus retested to see if, in fact, classification algorithms can model wine quality based on physicochemical tests with higher accuracy.

## 2. Introduction

Wine quality is subjective. Personal preferences and sensory physiology of each individual can lead to one person’s preferences of wine being different to another’s. However, any wine drinker would surely welcome a little guidance in which brand and/or type may be considered better quality over another when it comes to guiding their purchases, whether or not price range is factor. But are there factors that can help or hinder the measure of wine quality, and do any individual factors dominate in determining whether a wine should be considered excellent (or not so excellent)?

This report will discuss whether physicochemical levels of a number of variables present in wine production can reveal a preference for some brands/types over another, and thus lead to a higher quality rating.

## 3. Methodology

Two datasets, being red wine and white wine were obtained from the UCI Machine Learning Repository. [2] These datasets were supplied by researchers who completed a similar paper entitled *Modeling wine preferences by data mining from physicochemical properties*. [1]

The datasets had identical variables, however, the red wine dataset contained 1599 observations while the white wine dataset contained 4898 observations. A breakdown of the summary statistics of both datasets can be found in section 8 (Appendix).

The physicochemical variables included the following measurements:

- fixed acidity (g(tartaric acid)/$dm^3$)
- volatile acidity (g(acetic acid)/$dm^3$)
- citric acid (g/$dm^3$)
- residual sugar (g/$dm^3$)
- chlorides (g(sodium chloride)/$dm^3$)
- free sulfur dioxide (mg/$dm^3$)
- total sulfur dioxide (mg/$dm^3$)
- density (g/$cm^3$)
- pH
- sulphates (g(potassium sulphate)/$dm^3$)
- alcohol (% vol.)

The quality variable (being the designated “class” variable) was a sensory rating between 0-10.

Both datasets were unbalanced, with the value counts of each quality rating noted below in Table 1.

*Table 1: Value counts of each quality rating (both red and white wines)*

| Red wine | | White wine | |
|---|---|---|---|
| **Quality** | **Count** | **Quality** | **Count** |
| 3 | 10 | 3 | 20 |
| 4 | 53 | 4 | 163 |
| 5 | 681 | 5 | 1457 |
| 6 | 638 | 6 | 2198 |
| 7 | 199 | 7 | 880 |
| 8 | 18 | 8 | 175 |
| | | 9 | 5 |

The proportions of each quality rating are shown in Figure 1 below.

*Figure 1: Proportion of wine quality rating broken down by dataset*


The following analysis and modelling was completed using iPython and includes a number of classification models from *sklearn*, explained as follows:

- **K Nearest Neighbour**

  K-Nearest Neighbors, or simply kNN, belongs to the class of instance-based learning, also known as lazy classifiers. It's one of the simplest classification methods because the classification is done by just looking at the K closest examples in the training set … [3]

- **Decision Tree**

  A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node. [4]

- **Naïve Bayes**

> Naïve Bayes is a very common classifier used for probabilistic multiclass classification. Given the feature vector, it uses the Bayes rule to predict the probability of each class … it's very effective with large and fat data (with many features) with a consistent a priori[1] probability. [3]

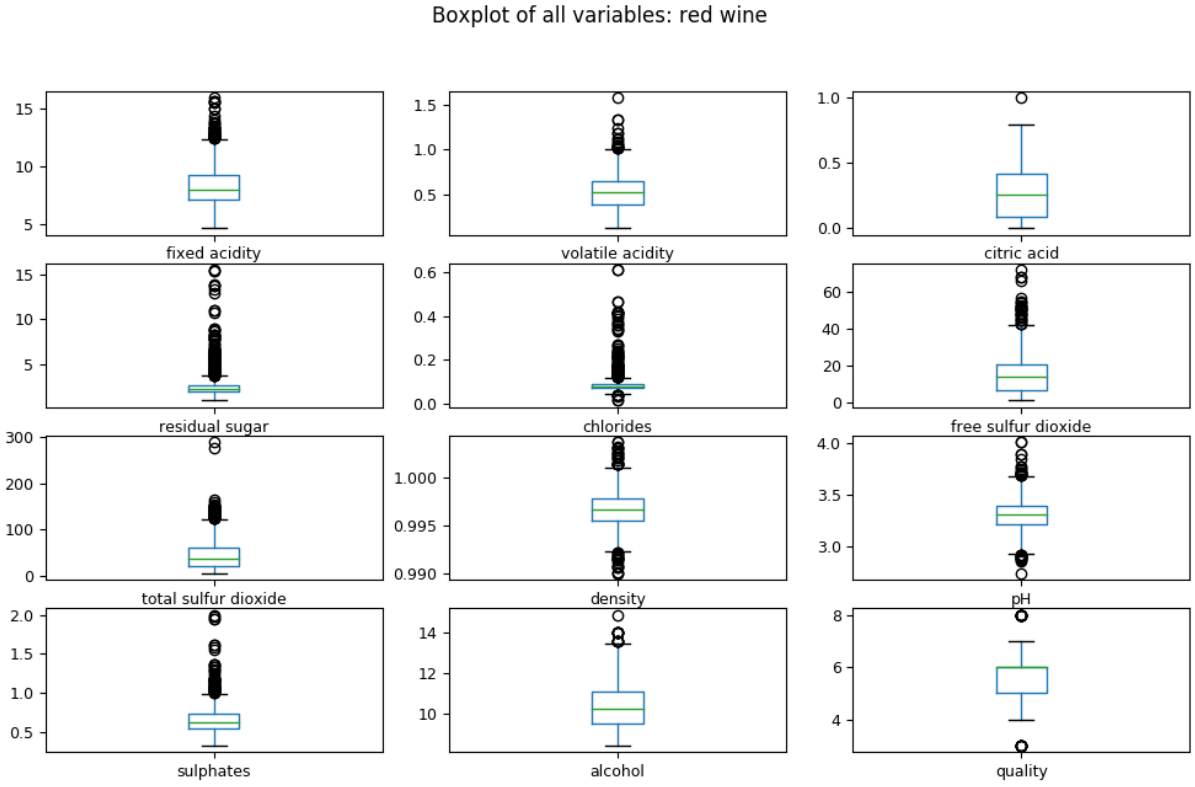
It should be noted that although some outliers were detected, due to the dataset being previously cleaned by P. Cortez, et al [1], and the author of this paper not being knowledgeable in the area nor structure of the physicochemical tests involved in the datasets, these outliers were retained for the analysis.

# 4. Results

## 4.1 Exploration

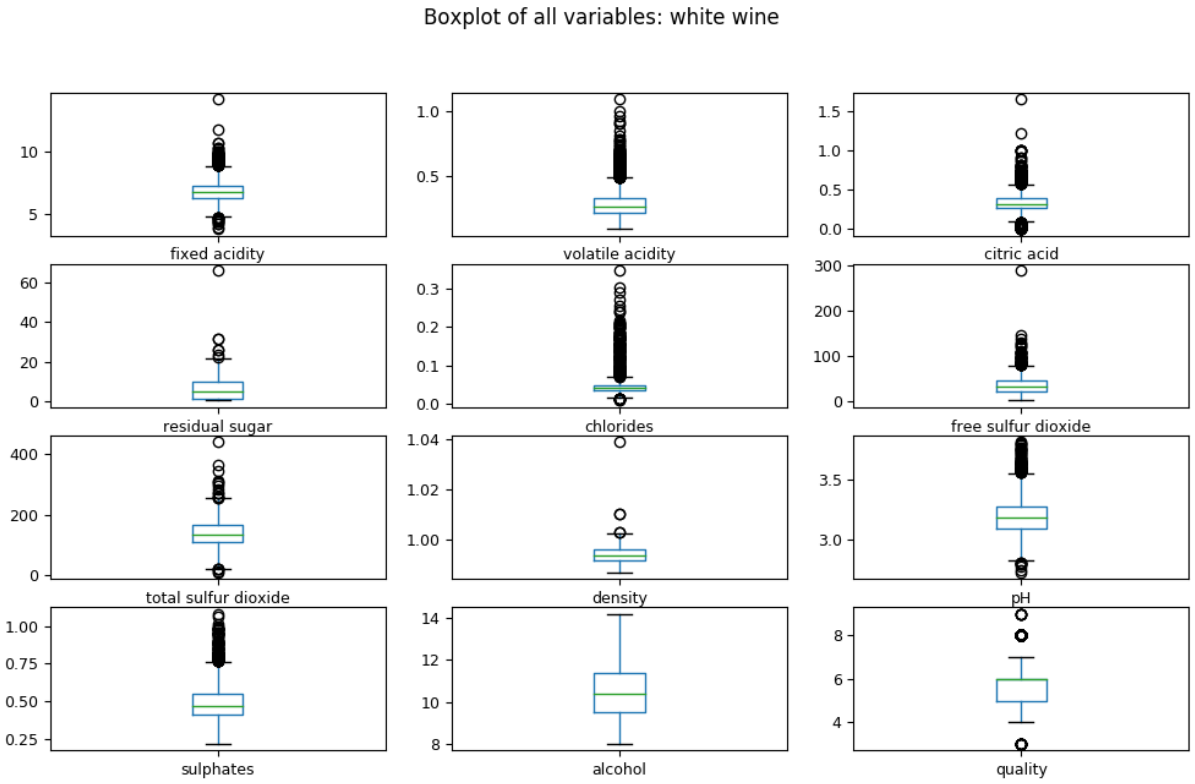
As this study looks to report on two discrete datasets, it was necessary to combine some functions. Therefore, each variable has been analysed below, for both datasets, however the analyses has been grouped into combined graphs. Both box plots and histograms were created for each variable per dataset in Figure 2 to Figure 5 below.

*Figure 2: Boxplot of all variables (red wine)*




[1] *A priori* meaning "from before" translates to say if the training data has balanced class value counts, then Naïve Bayes should work well. However, it is already known that the datasets in this study are unbalanced.

*Figure 3: Boxplot of all variables (white wine)*




A side by side comparison of the red and white boxplots reveals little difference (apart from outliers). Therefore, the combined histograms in Figure 4 and Figure 5 below should also be relatively similar in structure, with little variation between red and white wine.

*Figure 4: Histogram of all variables (red wine)*

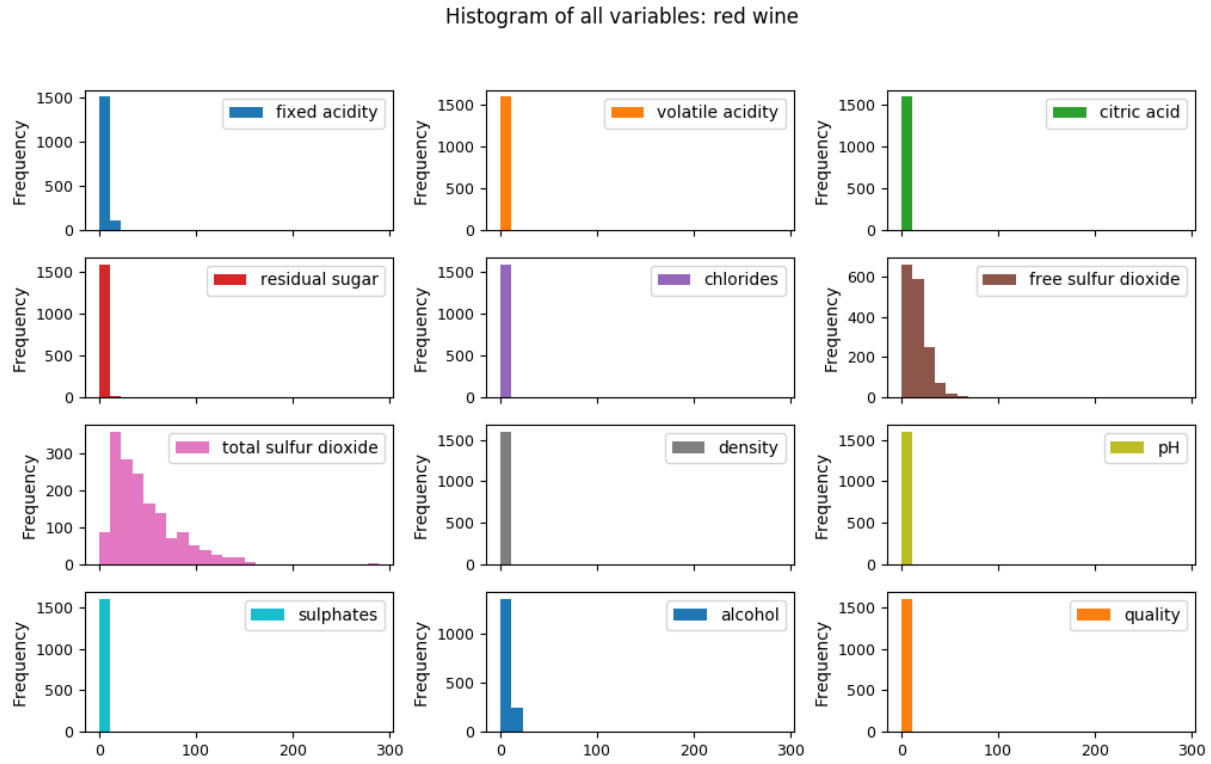
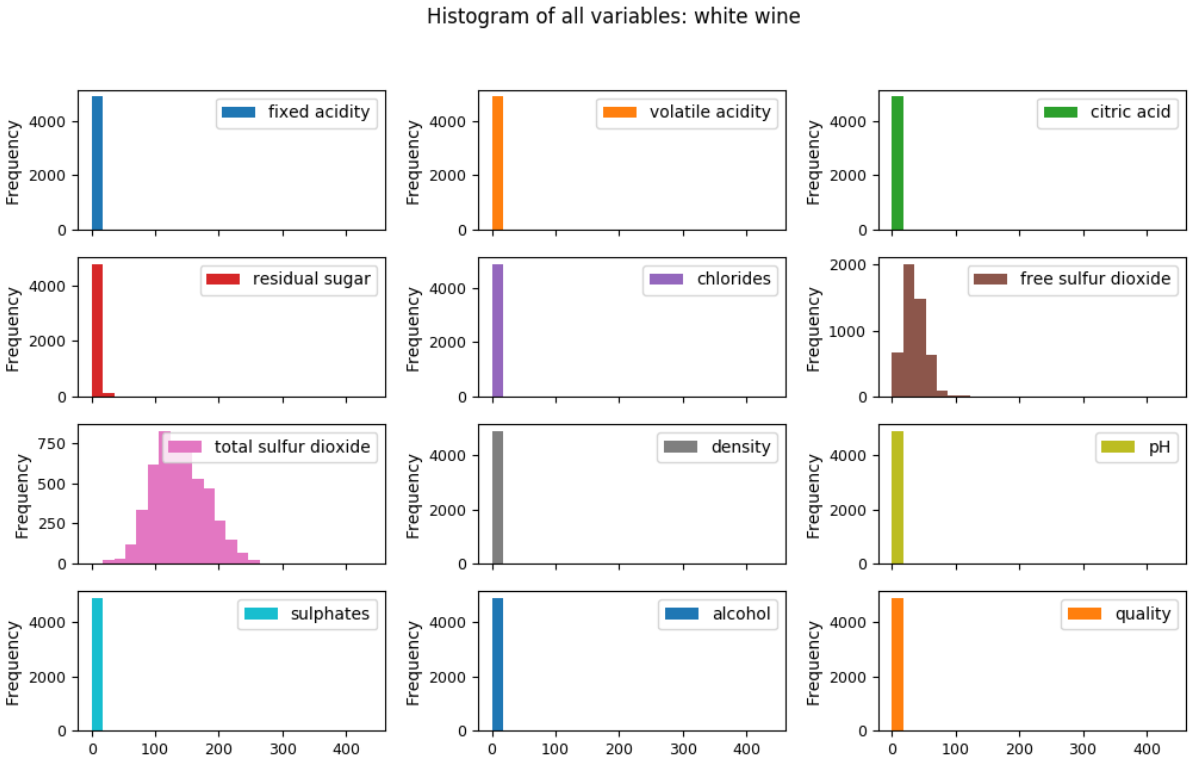


*Figure 5: Histogram of all variables (white wine)*




The only notable differences here were in the two sulfur dioxide variables.

To drill down further, the variables were split into pairs for analysis. This time, red and white graphs for the same pairs will be side by side for easy reference.

*Figure 6: Free v total sulfur dioxide (red wine)*

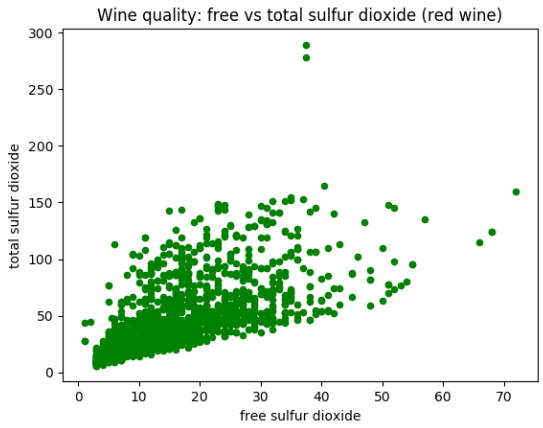
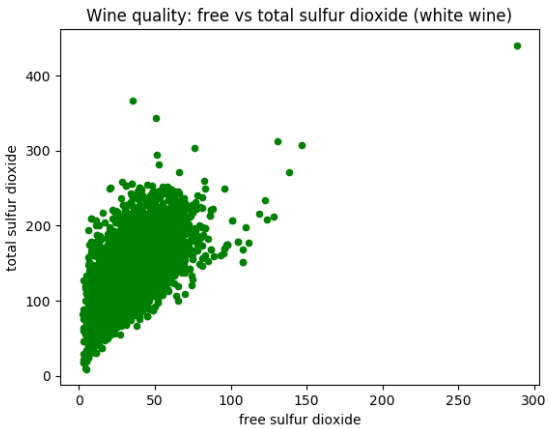



*Figure 7: Free v total sulfur dioxide (white wine)*




At first glance, there appears to be a significant difference between the combined variables in the two datasets, however, the singular outlier at >250 (free) and >400 (total) in Figure 7 above proves this to be misleading. Differences do exist, however, as the white wine has consistently higher values of both variables.

*Figure 8: Fixed v volatile acidity (red wine)*

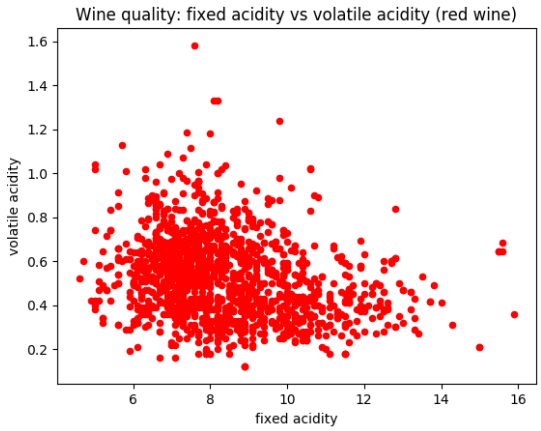
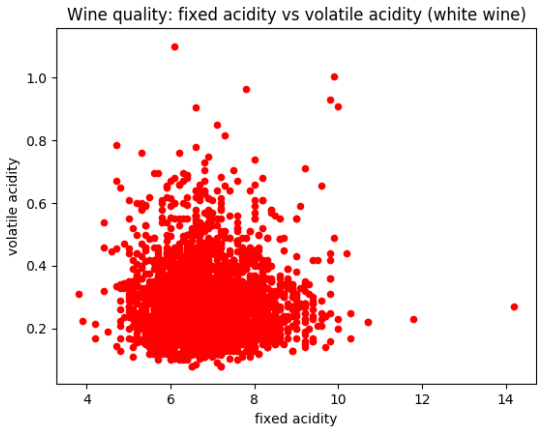



*Figure 9: Fixed v volatile acidity (white wine)*




Again, there is little difference between the two datasets when it comes to fixed v volatile acidity levels. Remembering that the white wine dataset has 50% more observations would account for the higher level of clustering.

*Figure 10: Citric acid v residual sugar (red wine)*

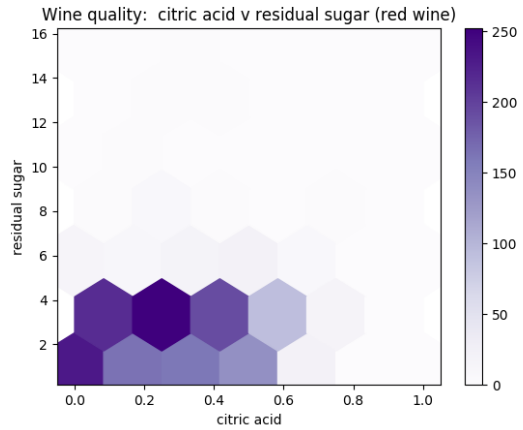
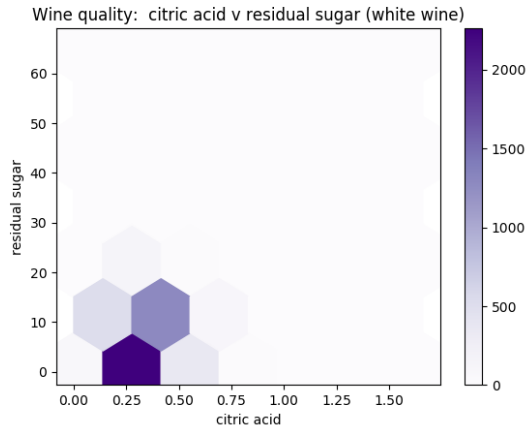



*Figure 11: Citric acid v residual sugar (white wine)*




Due to the larger number of observations and outliers in the white wine dataset for these variables, the hexbin graphs in Figure 10 and Figure 11 above appear different, however the clustering in the white wine graph appears to be a clustered version comparable to that of the red wine graph.

*Figure 12: Chlorides v sulphates (red wine)*

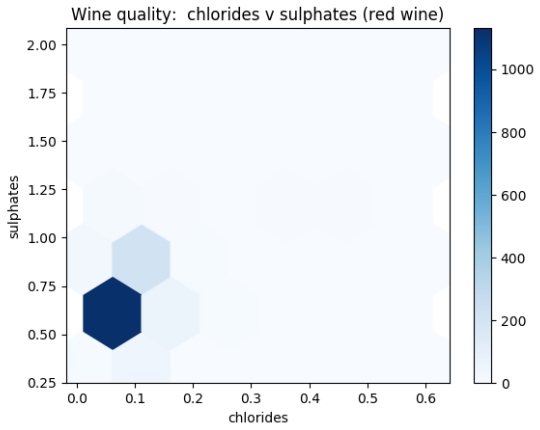
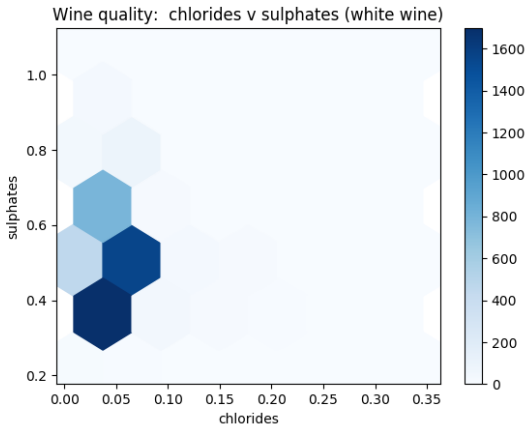


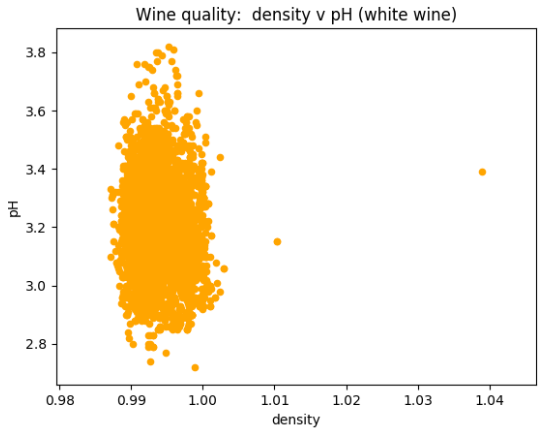

*Figure 13: Chlorides v sulphates (white wine)*




There is variation in Figure 12 and Figure 13 above in that red wine is clearly clustered in one area of the chart, whilst white wine appears to have lower sulphate levels and, to some degree, lower chloride levels.

*Figure 14: Density v pH (red wine)*


*Figure 15: Density v pH (white wine)*




Similar to the comparative scenario in Figure 6 and Figure 7 above, the outliers in the white wine dataset cause the graphs to look uneven, however, density of the main cluster in Figure 15 appears comparable to that in Figure 14.

*Figure 16: Alcohol (%) v quality (red wine)*

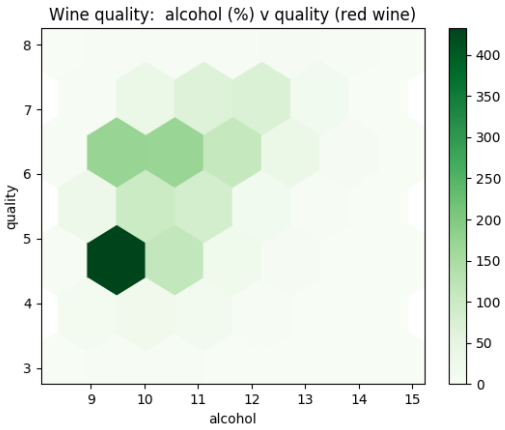
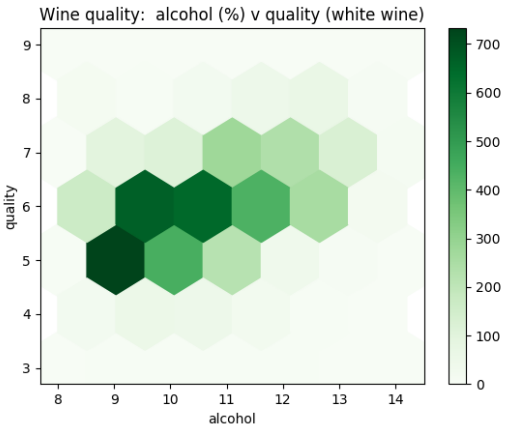



*Figure 17: Alcohol (%) v quality (white wine)*

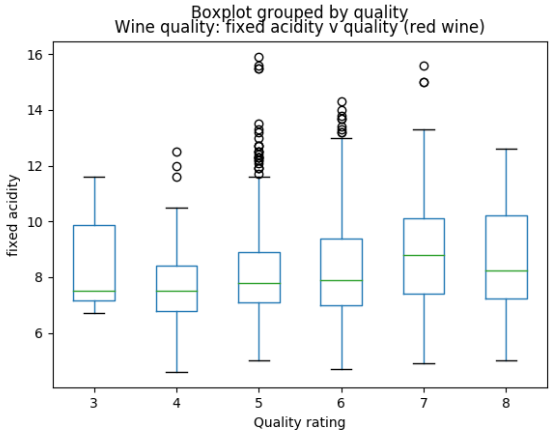


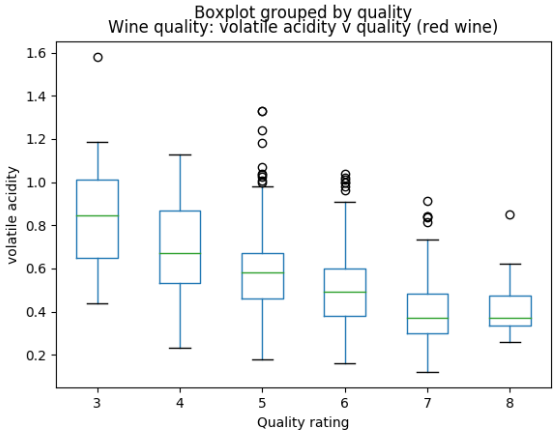

Although alcohol v quality rating will also be explored below (alongside all other variables v the class variable of quality), it was interesting to create the above hexbin. Both red and white wines are heavily clustered around the 9% alcohol/5 rating, with the white wine demonstrating a progression towards higher quality when alcohol content is also higher.

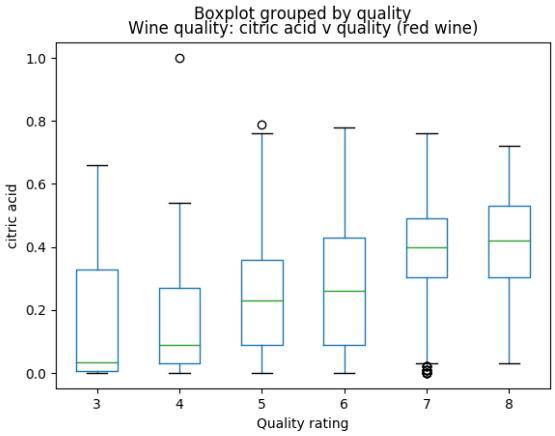
Next, each variable will be broken down into boxplots in a comparison against the class variable (quality).

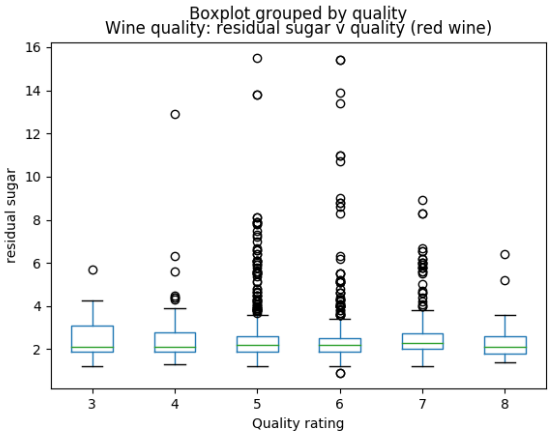
Rather than compare red to white wines side by side, all the red wines will be in the first section, followed by the white wines grouped together in their own section. This will facilitate the ability to discern any specific patterns that may lead to one or more variables affecting the overall quality of the wine.

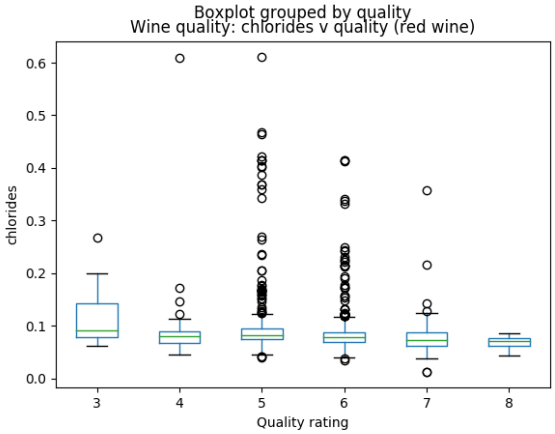
*Figure 18: Red wine variables compared to quality rating (class variable)*

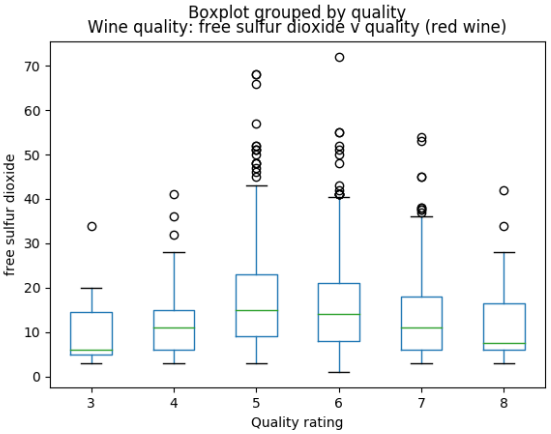


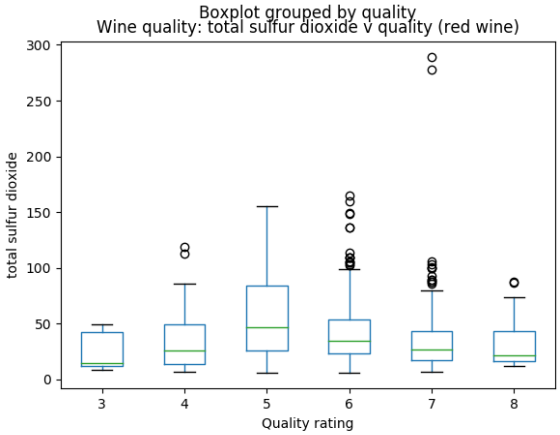



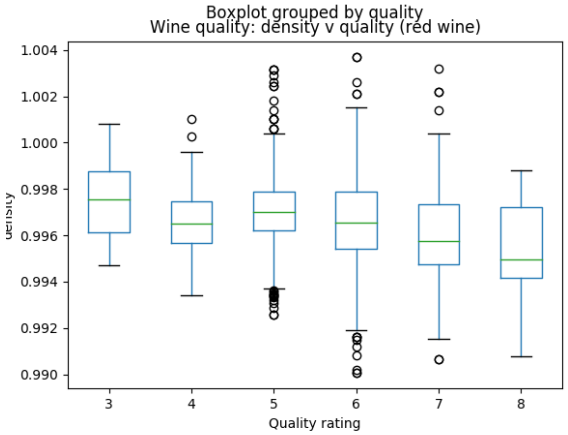

Boxplot grouped by quality
Wine quality: citric acid v quality (red wine)
citric acid
Quality rating

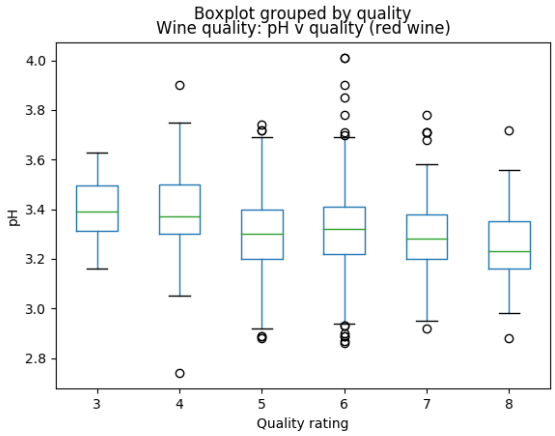

Boxplot grouped by quality
Wine quality: residual sugar v quality (red wine)
residual sugar
Quality rating

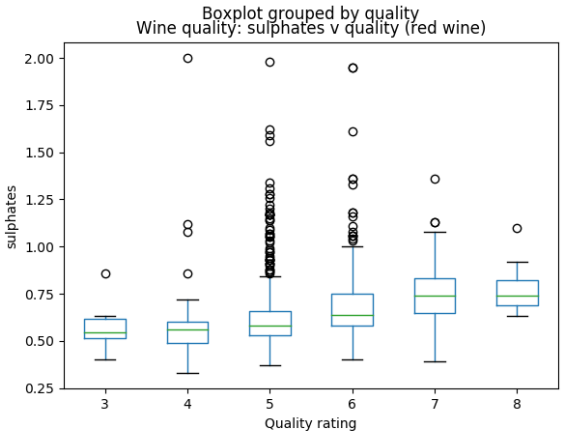

Boxplot grouped by quality
Wine quality: chlorides v quality (red wine)
chlorides
Quality rating

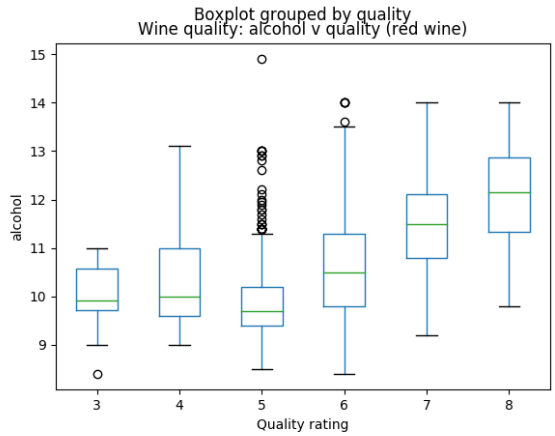

Boxplot grouped by quality
Wine quality: free sulfur dioxide v quality (red wine)
free sulfur dioxide
Quality rating


Boxplot grouped by quality
Wine quality: total sulfur dioxide v quality (red wine)
total sulfur dioxide
Quality rating


Boxplot grouped by quality
Wine quality: density v quality (red wine)
Quality rating

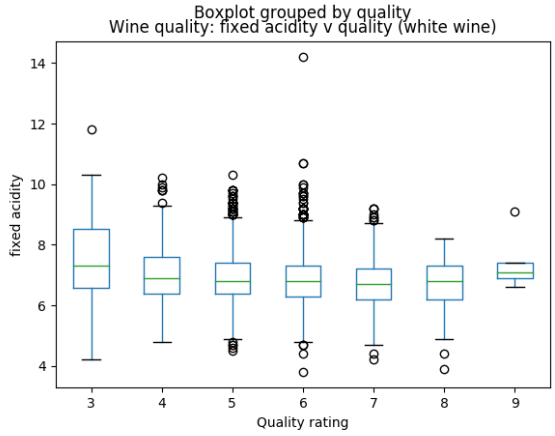


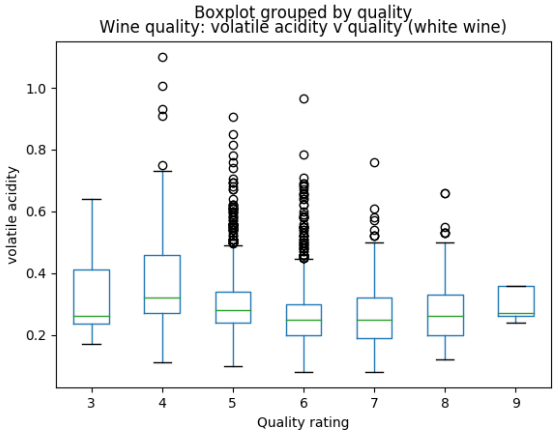



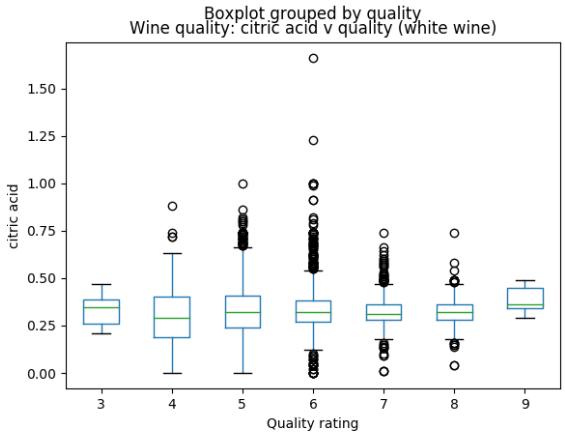



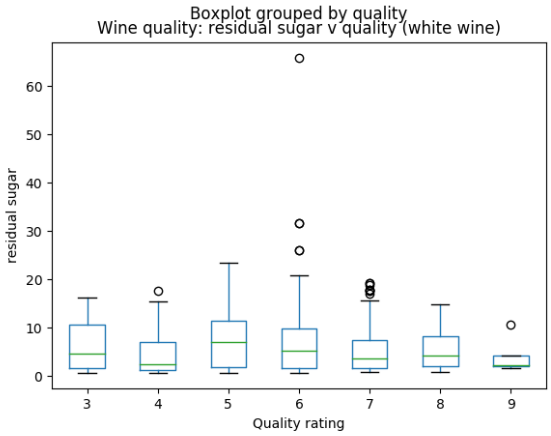

In the above graphs as part of Figure 18, it was notable that the following variables contributed to a higher quality rating when the volume was lower: volatile acidity; density; and pH, while the following variables contributed to a higher quality rating when the volume was higher: citric acid; sulfates; and alcohol %.

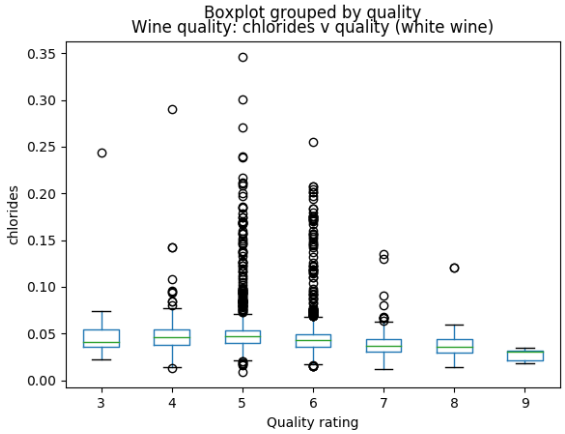
*Figure 19: White wine variables compared to quality rating (class variable)*

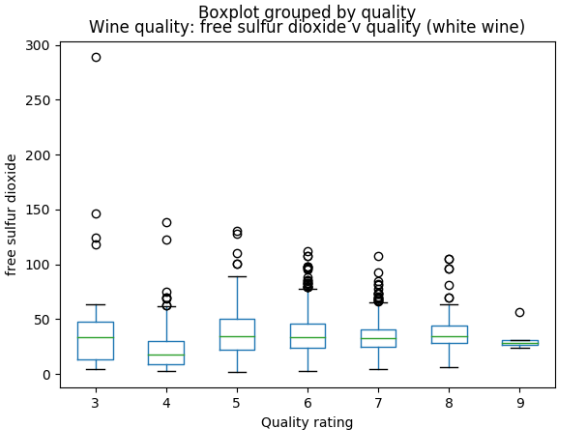


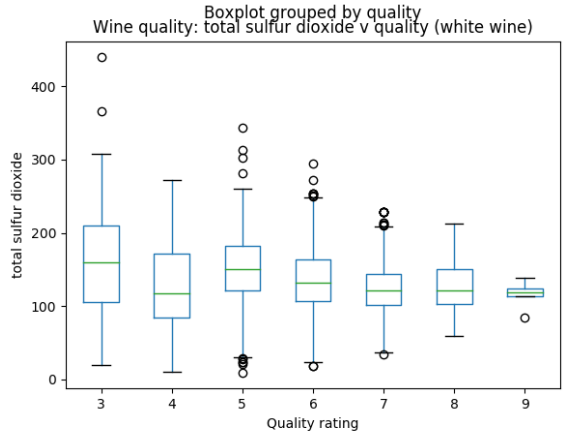



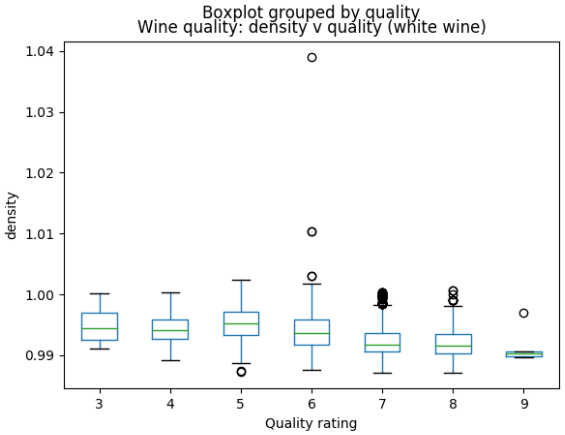

Boxplot grouped by quality
Wine quality: citric acid v quality (white wine)
citric acid
Quality rating

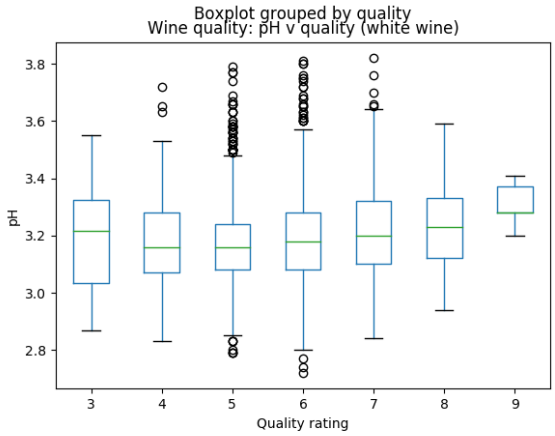

Boxplot grouped by quality
Wine quality: residual sugar v quality (white wine)
residual sugar
Quality rating

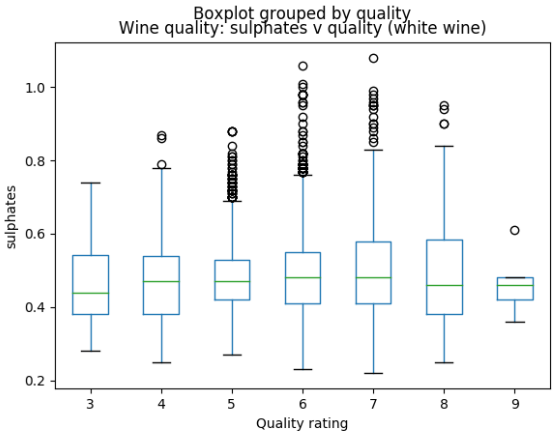

Boxplot grouped by quality
Wine quality: chlorides v quality (white wine)
chlorides
Quality rating

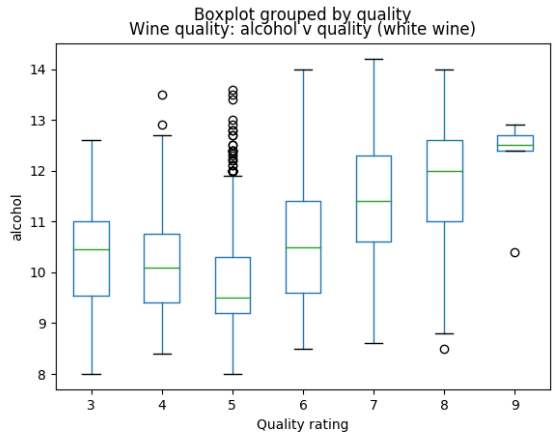

Boxplot grouped by quality
Wine quality: free sulfur dioxide v quality (white wine)
free sulfur dioxide
Quality rating


Boxplot grouped by quality
Wine quality: total sulfur dioxide v quality (white wine)
total sulfur dioxide
Quality rating


Boxplot grouped by quality
Wine quality: density v quality (white wine)
density
Quality rating




In contrast, the white wine variables that had an effect on the quality rating were: density (lower); pH (higher – this is in opposition to the red wine graph); and alcohol % (also higher). This may be due to the larger number of observations in the white wine dataset compared to the red wine dataset, as, referring back to Figure 14 and Figure 15, there was little difference in the pH levels of red and white wine.

## 4.2 Modelling

The modelling section will include analysis based on three separate classification algorithms, namely, K Nearest Neighbour, Decision Tree and Naïve Bayes (as described in section 3 above).

### *4.2.1 K Nearest Neighbour*

The K Nearest Neighbour (**KNN**) algorithm was run using a number of different scenarios to ascertain the best value for *k*. Test set was set to 25% and random state = 4.

The following table details classification error rates for various values of *k*.

*Table 2: Comparing CER for wine red/white with KKN where k=n*

| k | Classification error rate | |
|---|---|---|
| | Red | White |
| 1 | 0.39 | 0.44 |
| 2 | 0.478 | 0.505 |
| 3 | 0.493 | 0.511 |
| 4 | 0.478 | 0.527 |
| 5 | 0.483 | 0.522 |

| k | Classification error rate | |
|---|---|---|
| | Red | White |
| 10 | 0.47 | 0.54 |
| 20 | 0.448 | 0.538 |
| 50 | 0.468 | 0.546 |
| 100 | 0.48 | 0.56 |

The algorithm provided the best result with *k=1*, being 0.39 error rate for red and 0.44 error rate for white. Once *k≥2*, the results did not vary greatly and had no discernible pattern, such as continuously getting worse.

The confusion matrix for *k=1* looks like this:

*Figure 20: Confusion matrix (k=1) (red wine)*

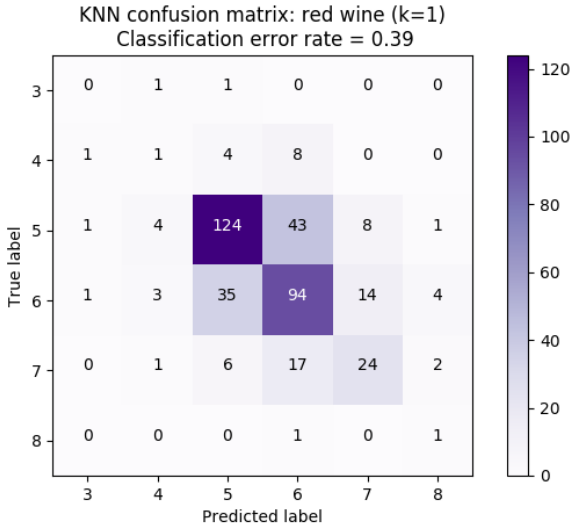
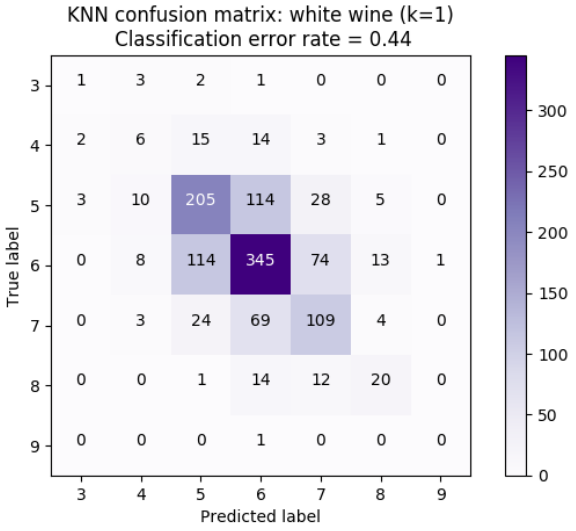



*Figure 21: Confusion matrix (k=1) (white wine)*




While there are good matches in both datasets, the algorithm struggled to differentiate between wines with a quality rating between 5-7.

*Table 3: Comparing classification reports for wine red/white with KKN where k=1*

| Classification report (red) | | | | | Classification report (white) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *precision* | *recall* | *f1-score* | *support* | | *precision* | *recall* | *f1-score* | *support* |
| 3 | 0.00 | 0.00 | 0.00 | 2 | 3 | 0.17 | 0.14 | 0.15 | 7 |
| 4 | 0.10 | 0.07 | 0.08 | 14 | 4 | 0.20 | 0.15 | 0.17 | 41 |
| 5 | 0.73 | 0.69 | 0.71 | 181 | 5 | 0.57 | 0.56 | 0.56 | 365 |
| 6 | 0.58 | 0.62 | 0.60 | 151 | 6 | 0.62 | 0.62 | 0.62 | 555 |
| 7 | 0.52 | 0.48 | 0.50 | 50 | 7 | 0.48 | 0.52 | 0.50 | 209 |
| 8 | 0.12 | 0.50 | 0.20 | 2 | 8 | 0.47 | 0.43 | 0.44 | 47 |
| **avg/total** | **0.62** | **0.61** | **0.61** | **400** | 9 | 0.00 | 0.00 | 0.00 | 1 |
| | | | | | **avg/total** | **0.56** | **0.56** | **0.56** | **1225** |

For KNN, red wine had a precision rate of 62%, while white wine had a precision rate of just 56%.

### *4.2.2 Decision tree*

The Decision tree algorithm was run using a number of different scenarios to ascertain pruning. However, the only pruning that appeared to affect the results was limiting the number of variables used. Again, test set was set to 25% and random state = 4.

The confusion matrix for *max_features=3* looks like this:

*Figure 22: Confusion matrix (max_features=3) (red wine)*

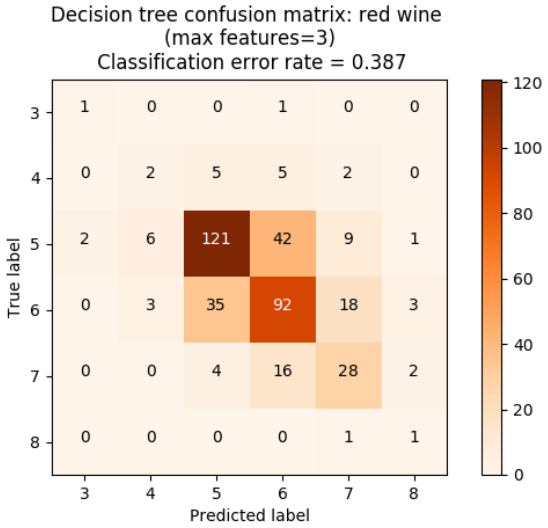
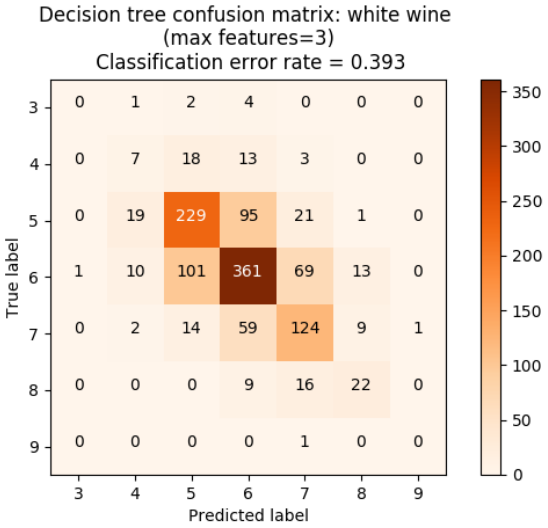



*Figure 23: Confusion matrix (max_features=3) (white wine)*




Similar to KKN above, the Decision tree algorithm made strong matches, however, struggled to differentiate between quality ratings between 5-7.

*Table 4: Comparing classification reports for wine red/white with Decision tree where max_features=3*

| Classification report (red) | | | | | Classification report (white) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *precision* | *recall* | *f1-score* | *support* | | *precision* | *recall* | *f1-score* | *support* |
| 3 | 0.00 | 0.00 | 0.00 | 2 | 3 | 0.00 | 0.00 | 0.00 | 7 |
| 4 | 0.12 | 0.07 | 0.09 | 14 | 4 | 0.18 | 0.17 | 0.18 | 41 |
| 5 | 0.72 | 0.65 | 0.68 | 181 | 5 | 0.63 | 0.63 | 0.63 | 365 |
| 6 | 0.60 | 0.66 | 0.63 | 151 | 6 | 0.67 | 0.65 | 0.66 | 555 |
| 7 | 0.56 | 0.54 | 0.55 | 50 | 7 | 0.53 | 0.59 | 0.56 | 209 |
| 8 | 0.12 | 0.50 | 0.20 | 2 | 8 | 0.49 | 0.47 | 0.48 | 47 |
| **avg/total** | **0.63** | **0.61** | **0.62** | **400** | 9 | 0.00 | 0.00 | 0.00 | 1 |
| | | | | | **avg/total** | **0.60** | **0.61** | **0.61** | **1225** |

For Decision tree, red wine had a precision rate of 63%, while white wine had a precision rate of 60%. Both rates are increases (slight for red) on KKN above.

### *4.2.3 Naïve Bayes*

The Naïve Bayes algorithm was run with the test set set to 25% and random state = 4[2].

The confusion matrix for Naïve Bayes looks like this:

*Figure 24: Confusion matrix (red wine)*

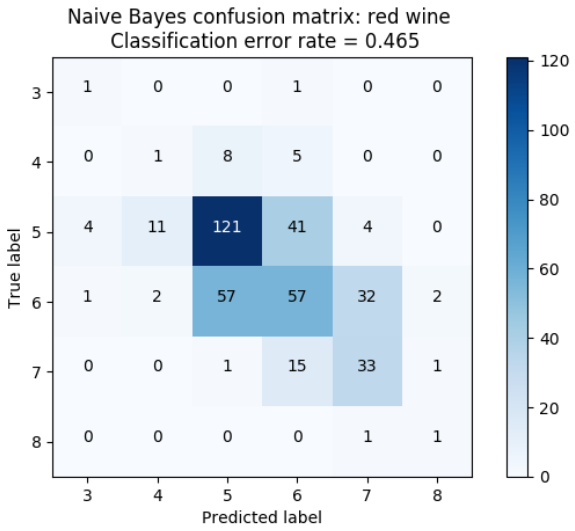
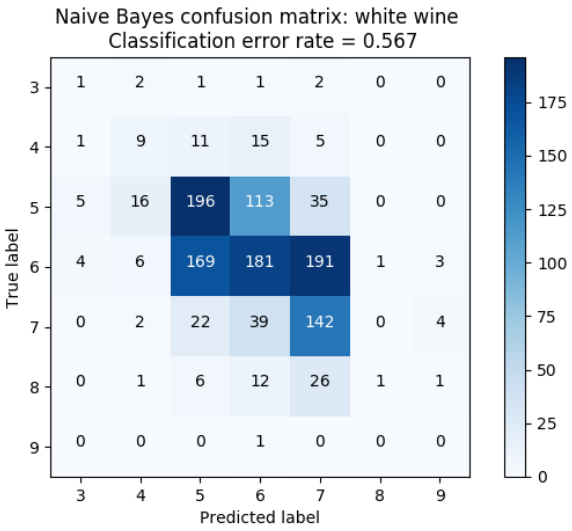



*Figure 25: Confusion matrix (white wine)*




Similar to KKN and Decision tree above, the Naïve Bayes algorithm made a strong showing in matches, however, really struggled to differentiate between quality ratings between 5-7 (especially obvious in the white wine confusion matrix in Figure 25 above).

*Table 5: Comparing classification reports for wine red/white with Naïve Bayes*

| Classification report (red) | | | | | Classification report (white) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *precision* | *recall* | *f1-score* | *support* | | *precision* | *recall* | *f1-score* | *support* |
| 3 | 0.17 | 0.50 | 0.25 | 2 | 3 | 0.09 | 0.14 | 0.11 | 7 |
| 4 | 0.07 | 0.07 | 0.07 | 14 | 4 | 0.25 | 0.22 | 0.23 | 41 |
| 5 | 0.65 | 0.67 | 0.66 | 181 | 5 | 0.48 | 0.54 | 0.51 | 365 |
| 6 | 0.48 | 0.38 | 0.42 | 151 | 6 | 0.50 | 0.33 | 0.39 | 555 |
| 7 | 0.47 | 0.66 | 0.55 | 50 | 7 | 0.35 | 0.68 | 0.47 | 209 |
| 8 | 0.25 | 0.50 | 0.33 | 2 | 8 | 0.50 | 0.02 | 0.04 | 47 |
| **avg/total** | **0.54** | **0.54** | **0.53** | **400** | 9 | 0.00 | 0.00 | 0.00 | 1 |
| | | | | | **avg/total** | **0.46** | **0.43** | **0.42** | **1225** |

As expected, Naïve Bayes had a significantly lower precision rate of 54% for red wine and 46% for white wine. This is due to Naïve Bayes being strongest on balanced datasets.

## 5. Discussion

The question of whether wine quality (both red and white varieties) can be modelled based on a number of physicochemical tests is an interesting one when it comes to classification algorithms. As both these datasets demonstrate, wine quality is usually considered “average” with very few “bad” or “excellent” wines. If this is representative of the world of wine (which I believe it is), modelling such an unbalanced set of data may always prove problematic.

[2] These settings were maintained across all three algorithms to assist in comparison.

As demonstrated in the results section, it is possible to take a set of physicochemical values and predict a quality rating, however, as the alcohol % variable was the only one that consistently showed a higher rating based on higher alcohol content, it is possible that all other variable values did not determine a rating setting at all.

It was interesting to note that red and white wine appeared to contain very similar chemical markings and hence both sets of comparative results reflected one another.

Based on the results above, KNN with *k=1*, even though it is part of the *lazy* set of classifiers, achieved results that were, in fact, relatively good in the scheme of this report. As demonstrated in the confusion matrices, KNN predicted the "average" range of 5-7 quality quite well with precision of 62% and 56% for red and white respectively.

If the KNN results are compared with that of the Decision tree (*max_features=3*) with precision of 63% (red) and 60% (white), this is a relatively good result for KNN. The Decision tree algorithm should have been able to be pruned to achieve better results. However, as mentioned above, the variables themselves do not appear to be capable of indicating a higher rating (except for alcohol content). However, when the tree was pruned to *max_features=1* in the hope that alcohol rating would guide the classifier, interestingly these results were not better than the chosen *max_features=3*.

Other pruning restrictions were also tested, including *max_depth*, *min_samples_split*, *min_samples_leaf*, *max_leaf_nodes* (all at varying values), but none of these prunes had any effect on the outcome. Therefore, the Decision tree was constantly overfitting on the training set and hence the results were unbalanced.

The Naïve Bayes algorithm did not perform well either. However, that was to be expected due to it's a *priori* structure of working best with balanced datasets. Its precision rates of 54% for red wine and 46% for white wine gave it the worst results of all three classification algorithms tested.

## 6. Conclusion

Overall, while the algorithms all succeeded in classifying the "average" wines, none of them predicted the "bad" or "excellent" rated wines at a sufficient level. Therefore, and based on these datasets likely representing the quality of wine around the world, it would be hard to say that any classification algorithm could predict such unbalanced data with any confidence.

Wine rating is sensory and therefore the makeup of physicochemicals accurately affecting tastebuds the same way in every customer is unlikely. If certain physicochemicals could "guarantee" an excellent rating every time, then the wine producers of the world would likely have already cornered this area of manufacturing. However, there are other factors that go into the making of wine – including available crops (which may be adversely affected by the weather), price, soil and location of vines. Therefore, wines around the world are likely to maintain that "average" rating for 70-80% of their stock (as demonstrated in Figure 1).

While this may be disappointing from a data science/classification point of view, it would be interesting to run some further tests.

The possibilities could include:

- Combining both red and wine datasets (given the differences in variable values were negligible) so that a larger dataset could be analysed, however, remove the outliers and trim the volume of "average" wines to balance the dataset more; and
- Reduce the dataset to a smaller number of variables (using a select attributes algorithm) to determine if some variables are more likely to classify ratings better than the whole set.

## 7. References

[1] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. *Modeling wine preferences by data mining from physicochemical properties.* In Decision Support Systems, Elsevier, 47(4):547-553, 2009.

[2] UCI Machine Learning Repository (2017). *Wine Quality Data Set*. Retrieved from http://archive.ics.uci.edu/ml/datasets/Wine+Quality. Date retrieved: 4 May 2017.

[3] Boschetti, A and Massaron, L, 2015, *Python Data Science Essentials*, Packt Publishing Ltd, Birmingham, UK.

[4] Tutorials Point (2017). *Data Mining Decision Tree Induction.* Retrieved from https://www.tutorialspoint.com/data_mining/dm_dti.htm. Date retrieved: 14 May 2017.

# 8. Appendix

*Table 6: Summary statistics for red wine*

| | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | pH | sulphates | alcohol | quality |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 1599 | 1599 | 1599 | 1599 | 1599 | 1599 | 1599 | 1599 | 1599 | 1599 | 1599 | 1599 |
| mean | 8.319637273 | 0.527820513 | 0.27097561 | 2.538805503 | 0.087466542 | 15.87492183 | 46.46779237 | 0.996746679 | 3.311113196 | 0.658148843 | 10.42298311 | 5.636022514 |
| std | 1.741096318 | 0.179059704 | 0.194801137 | 1.40992806 | 0.047065302 | 10.46015697 | 32.89532448 | 0.001887334 | 0.154386465 | 0.16950698 | 1.065667582 | 0.80756944 |
| min | 4.6 | 0.12 | 0 | 0.9 | 0.012 | 1 | 6 | 0.99007 | 2.74 | 0.33 | 8.4 | 3 |
| 25% | 7.1 | 0.39 | 0.09 | 1.9 | 0.07 | 7 | 22 | 0.9956 | 3.21 | 0.55 | 9.5 | 5 |
| 50% | 7.9 | 0.52 | 0.26 | 2.2 | 0.079 | 14 | 38 | 0.99675 | 3.31 | 0.62 | 10.2 | 6 |
| 75% | 9.2 | 0.64 | 0.42 | 2.6 | 0.09 | 21 | 62 | 0.997835 | 3.4 | 0.73 | 11.1 | 6 |
| max | 15.9 | 1.58 | 1 | 15.5 | 0.611 | 72 | 289 | 1.00369 | 4.01 | 2 | 14.9 | 8 |

*Table 7: Summary statistics for white wine*

| | fixed acidity | volatile acidity | citric acid | residual sugar | chlorides | free sulfur dioxide | total sulfur dioxide | density | pH | sulphates | alcohol | quality |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 4898 | 4898 | 4898 | 4898 | 4898 | 4898 | 4898 | 4898 | 4898 | 4898 | 4898 | 4898 |
| mean | 6.854787668 | 0.278241119 | 0.334191507 | 6.391414863 | 0.045772356 | 35.30808493 | 138.3606574 | 0.994027376 | 3.188266639 | 0.489846876 | 10.51426705 | 5.877909351 |
| std | 0.843868228 | 0.100794548 | 0.121019804 | 5.072057784 | 0.021847968 | 17.00713733 | 42.49806455 | 0.002990907 | 0.1510006 | 0.114125834 | 1.230620568 | 0.885638575 |
| min | 3.8 | 0.08 | 0 | 0.6 | 0.009 | 2 | 9 | 0.98711 | 2.72 | 0.22 | 8 | 3 |
| 25% | 6.3 | 0.21 | 0.27 | 1.7 | 0.036 | 23 | 108 | 0.9917225 | 3.09 | 0.41 | 9.5 | 5 |
| 50% | 6.8 | 0.26 | 0.32 | 5.2 | 0.043 | 34 | 134 | 0.99374 | 3.18 | 0.47 | 10.4 | 6 |
| 75% | 7.3 | 0.32 | 0.39 | 9.9 | 0.05 | 46 | 167 | 0.9961 | 3.28 | 0.55 | 11.4 | 6 |
| max | 14.2 | 1.1 | 1.66 | 65.8 | 0.346 | 289 | 440 | 1.03898 | 3.82 | 1.08 | 14.2 | 9 |