**Case 1: Wine Board of Portugal**

**BAIS:9110:0700 Spr21 Advanced Analytics**

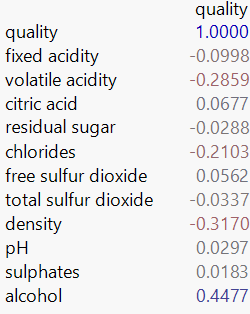
**Group 10 – Brooke Arnold, Canny Li, Ellie Paul, Jeffrey Chan**

**Due March 3, 2021**

**Evaluation of Wine Quality**

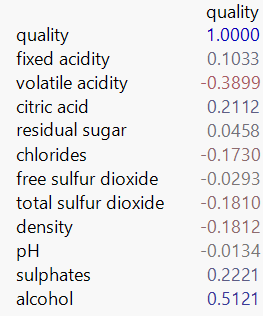
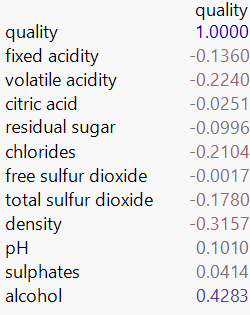
In order to determine what makes a good wine, we must look into what goes into each wine. In what ways are wines different from each other and how do those differences impact their quality rating? In this report, we will take a deep dive into the different aspects of each wine, in hopes of gaining more knowledge about how to create a high-quality wine.

The first thing that we did to see what features of a wine impact their rating is create a correlation matrix between each feature and quality. A high correlation (typically closer to 1 or –1) indicates that the feature is related to how the wine is rated. A positive value indicates that the feature has a positive effect on quality while a negative value has a negative effect. We began by finding the correlations between quality and all features. The matrix is shown below.



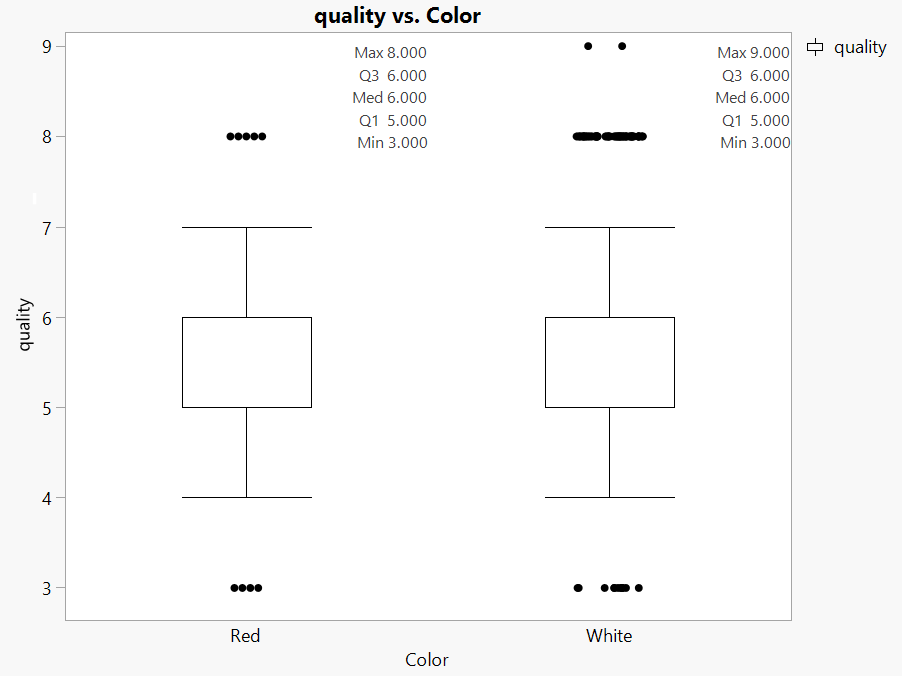
What this image shows us is that the features with the highest values are the best predictors of quality. Alcohol is the highest correlation, followed by density, and volatile acidity. This indicates that these features may be the best to use to create a model for wine quality.

We then saw that there is a difference in important features in red wines versus white wines. You wouldn’t expect the same taste/features from a white wine as you would a red wine. We then created a correlation matrix of red wines and a correlation of white wines. The left matrix is for red and the right is for white.

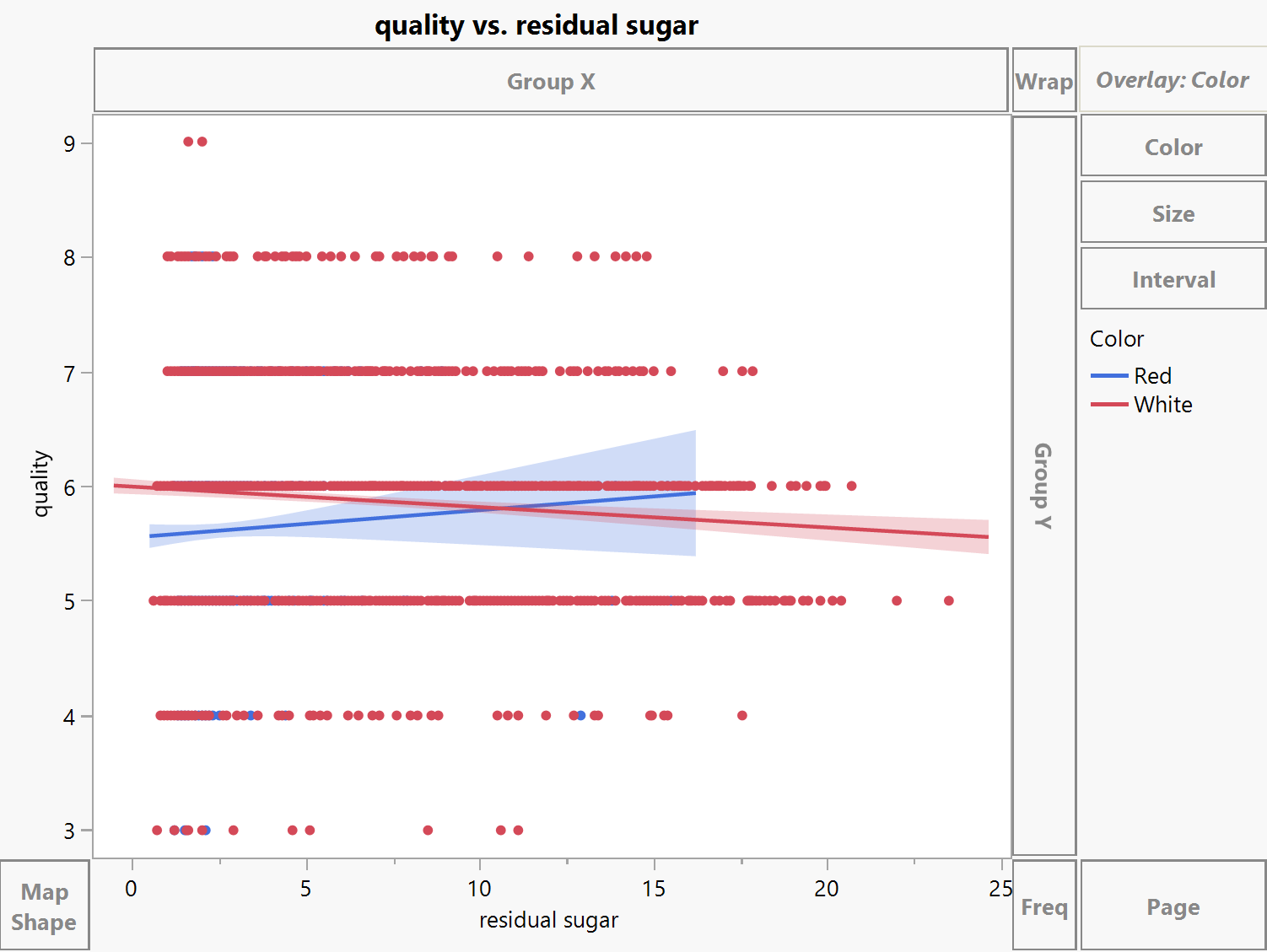
What these new matrices show us are that alcohol is the feature most closely associated with quality rating in both red and white wine. The next best feature for red wine is volatile acidity while the next best for white wine is density. As you can see from these two tables, different aspects are associated with a high rating when comparing red and white. This indicates that the best model should be broken down by color.

After determining that the color of the wine determines what features are best predictors of wine quality, we wanted to look at the quality scores by color. To do this, we created a box plot to show the distribution of the quality scores, the average, and any outliers based on color. The box plot is shown below.



What can be seen here is that overall, the quality scores between white and red wines are consistent with the same minimum and median scores. However, you can see that white wines have higher quality scores than red wines do.

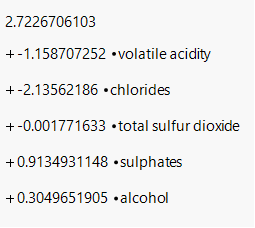
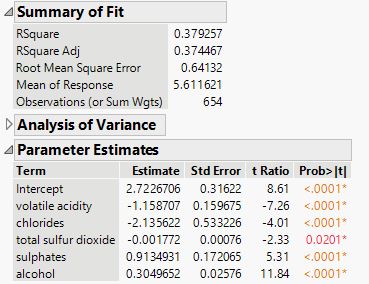
Taking a deeper dive into how wine color impacts its rating, we also investigated how the color of a wine combined with its sweetness impacts the quality. Looking at the picture below, the sweetness is plotted on the x-axis. This means that the farther to the right a point is, the sweeter the wine. The rating is plotted on the y-axis, meaning that the further up a point is, the higher the rating of the wine.



Most interesting about this plot is the two lines that cross in the center. The blue line represents the trend in red wine, showing that the sweeter a red wine is, the higher the quality it is. The red line represents the trend in white wine, showing that the sweeter a white wine is, the lower the quality is. It is even more evident from this analysis that red and white wines are evaluated differently, having different criteria for what makes a good wine. Our next step was to create two separate models to evaluate the different colored wines.

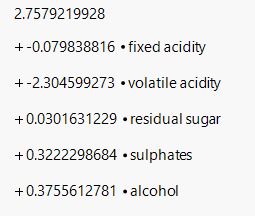
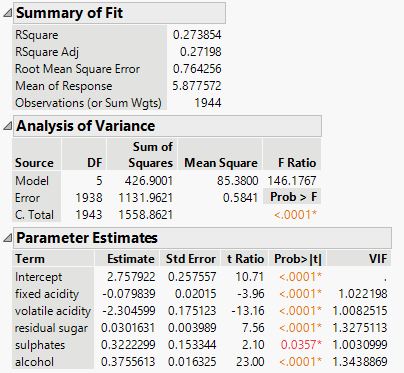
**Red Wine Analysis**

After deciding to do the regression analysis by color, we ran a regression for red wine where we removed properties that are insignificant towards predicting quality rating for red wine. The features that were removed because of a low significance were fixed acidity, citric acid, residual sugar, free sulfur dioxide, density and PH value. The formula for the final model is indicated in the screenshot below. From the formula we can conclude that, on one hand, sulphate and alcohol amount in red wine positively impacts the quality rating of the wine. On the other hand, the volatile acidity, the chloride amount and the total sulfur dioxide in red wine negatively impacts the quality rating of red wine. The model that was produced can accurately predict approximately 37.45% of the outcome.

**White Wine Analysis**

Similarly, we created best model to predict the quality of white wine using the given predictors. Again, we set the significance threshold as a p-value of 0.05 and repeated what we did in the red wine model. If one feature has a p-value greater than 0.05, it means that the feature is not significant to predicting white wine quality. Following that rule, we removed features in the order of chlorides, citric acid, free sulfur dioxide, and total sulfur dioxide. We then checked the Variance Inflation Factor (VIF) which shows if any features have a large relationship to another feature. If the VIF value is over 5, that feature should be removed. We removed the density as it has a VIF over 5, which then caused pH’s p-value to rise greater than 0.05, so it was removed as well. Eventually, we built a model using variables: volatile acidity, residual sugar, sulphates, fixed acidity, and alcohol. The final model can accurately predict approximately 27.2% of the outcome.

Finally, we tested our model by determining the rating of a wine with the following characteristics:

fixed acidity = 8.3

volatile acidity = 0.55

citric acid = 0.41

residual sugar = 4.3

chlorides = 0.040

free sulfur dioxide = 22

total sulfur dioxide = 129

density = 0.9971

pH = 3.21

sulphates = 0.44

alcohol = 11.1

color = Red

Plugging these numbers into our model for red wine, we get the following equation:

Quality Rating = 2.72 -1.159 (0.55) - 2.136 (0.040) - 0.002 (129) + 0.913 (0.44) + 0.305 (11.1)

= 5.558

Therefore, according to our final model, the wine quality of rating of given characteristics should be between 5 and 6.

Based on our analysis, we conclude that the qualities of red and white wine are not determined in the same way as they both have unique characteristics that determine their ratings. The features that impact the rating for Red wine are Volatile Acidity, Chlorides, Total Sulfur Dioxide, Sulphate, and Alcohol. The features that impact the rating for White wine are Fixed Acidity, Volatile Acidity, Residual Sugar, Sulphates, and Alcohol. By using the models created above, the Wine Board of Portugal can use wine characteristics to determine how to make the best quality wine.