Relation of Wine Quality to Physicochemical Variables

Capstone 1 Presentation

Presenter: Hannah-Lee Grothaus
Date: 08.24.2020

The Data

fixed acidity

most acids involved with wine or fixed or nonvolatile (do not evaporate readily)

volatile acidity

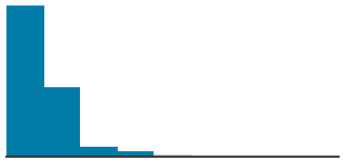
the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste

citric acid

found in small quantities, citric acid can add 'freshness' and flavor to wines

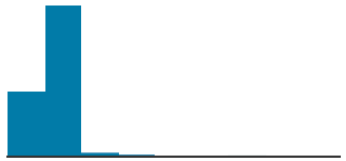
residual sugar

the amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and



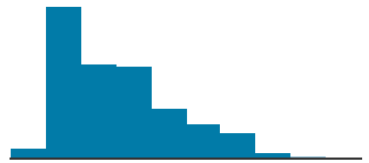
chlorides

the amount of salt in the wine



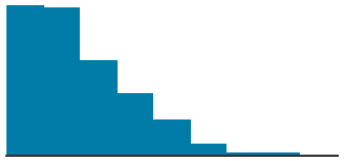
alcohol

the percent alcohol content of the wine



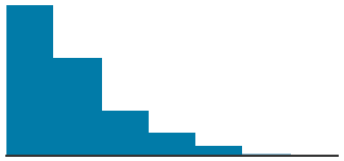
free sulfur dioxide

the free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulfite ion; it prevents



total sulfur dioxide

amount of free and bound forms of S02; in low concentrations, SO2 is mostly undetectable in wine, but at free SO2



density

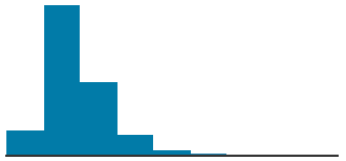
the density of water is close to that of water depending on the percent alcohol and sugar content

pH

describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the

sulphates

a wine additive which can contribute to sulfur dioxide gas (S02) levels, wich acts as an antimicrobial and



quality

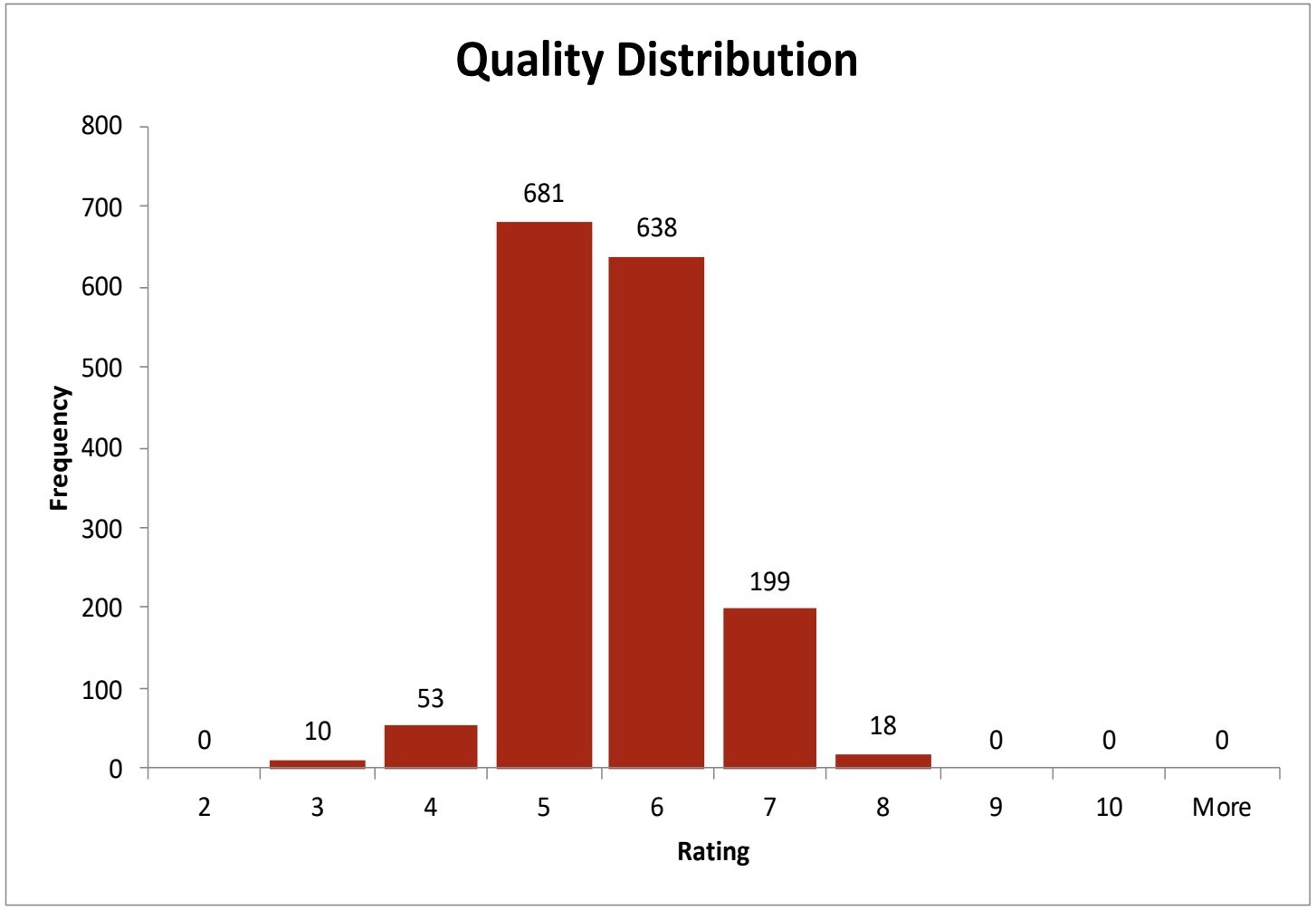
output variable (based on sensory data, score between 0 and 10)

Data's Limitations

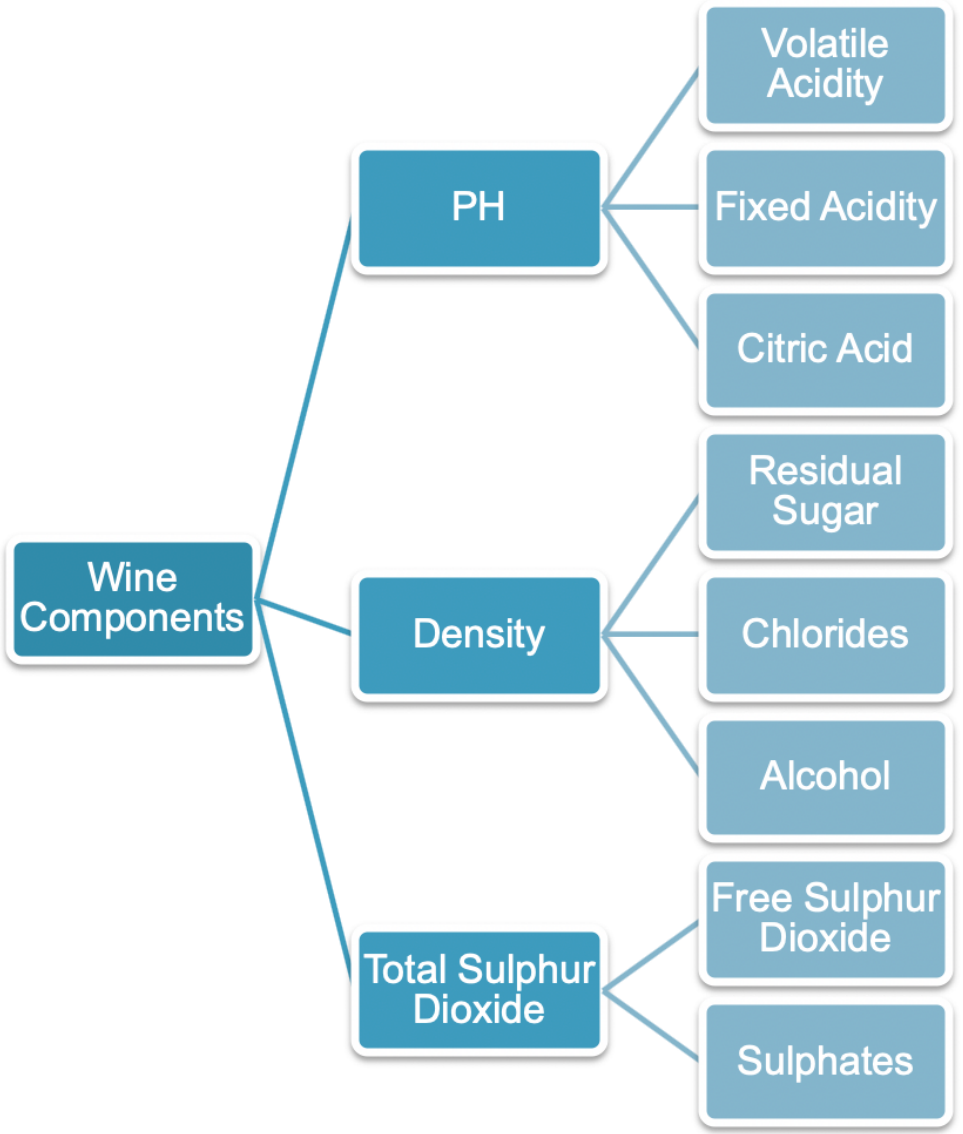
-The distribution of data focuses on ratings of 5 and 6.

-The data is limited to red wine grown in the Vinho Verde region of Portugal.

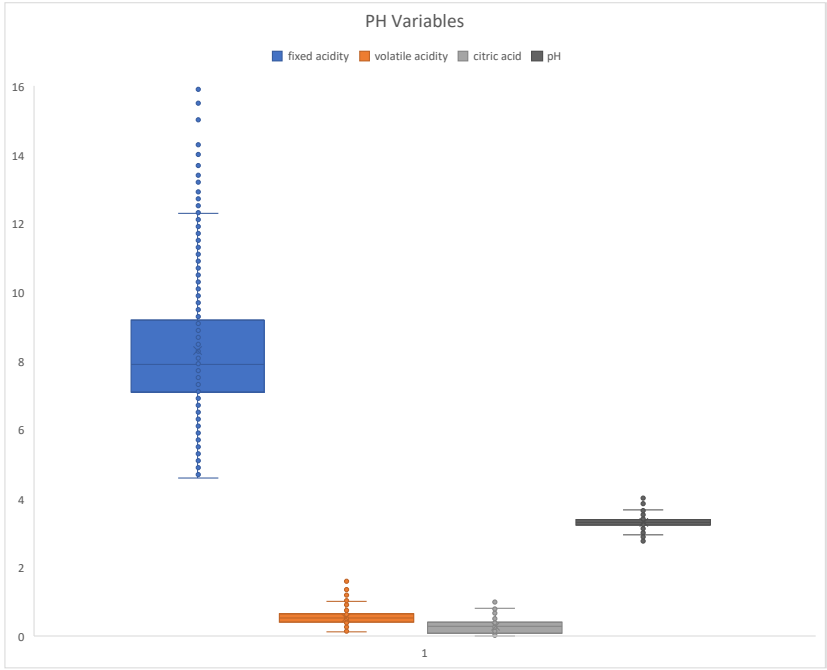
-We do not know how quality is being determined.

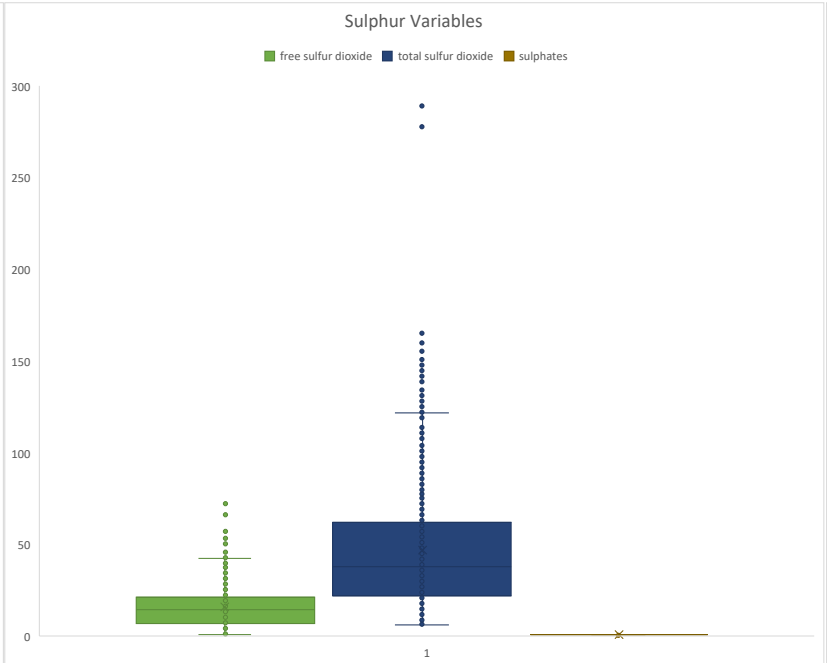


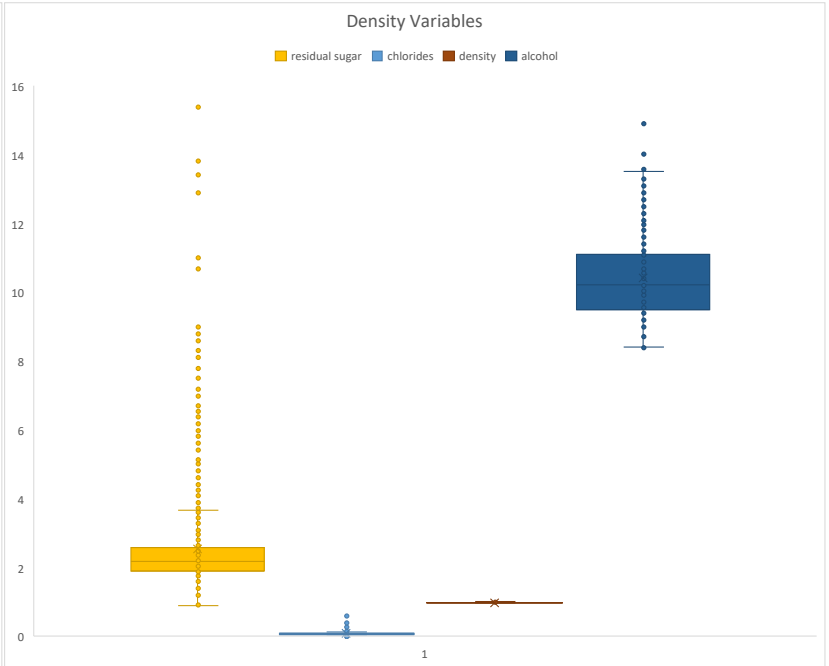
Using descriptive and inferential statistics, we identified key variables with the strongest correlation to quality, enabling the development of a multivariate equation capable of predicting quality.



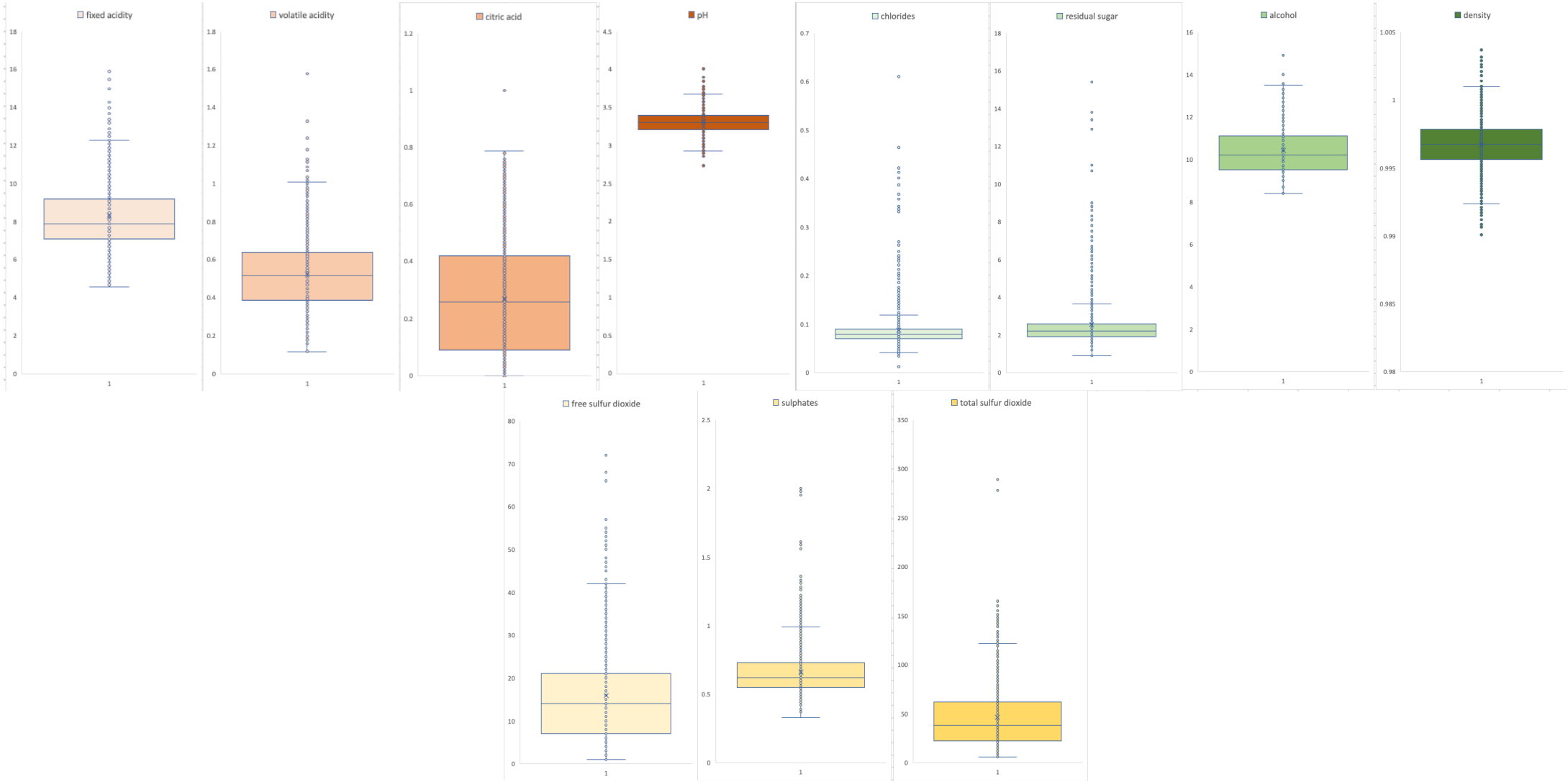
Grouped with their related variables, the boxplots showed high level of variation for fixed acidity, total Sulphur dioxide, alcohol, and residual sugar variables.







Individually, the boxplots for citric acid and free Sulphur dioxide had the most variation.



The correlation analysis shows that volatile acidity and alcohol are most closely correlated to a rating change.

Using step-wise regression, we determined that 7 of the 11 variables in our data had significant impact on the quality of the wine.

1st Round - all against quality	
alcohol	2.83E-91
volatile acidity	2.05E-59
sulphates	1.80E-24
citric acid	4.99E-20
total sulfur dioxide	8.62E-14
density	1.87E-12
chlorides	2.31E-07
fixed acidity	6.50E-07
pH	2.10E-02
free sulfur dioxide	4.28E-02
residual sugar	5.83E-01

Second Round - w/ Alcohol	
volatile acidity	5.56E-45
sulphates	1.05E-21
citric acid	9.41E-16
pH	3.23E-13
fixed acidity	1.97E-12
total sulfur dioxide	4.98E-05
density	1.31E-03
chlorides	2.71E-01
free sulfur dioxide	4.23E-01
residual sugar	7.74E-01

3rd - w/ Alch & Volatile Acidity	
sulphates	2.26E-11
total sulfur dioxide	1.53E-04
pH	2.53E-04
fixed acidity	3.40E-04
density	3.69E-02
free sulfur dioxide	2.23E-01
chlorides	3.59E-01
citric acid	5.10E-01
residual sugar	8.81E-01

4th - w/ Alch, VAcid, & Sulphates	
total sulfur dioxide	1.28E-05
chlorides	3.12E-05
fixed acidity	5.58E-03
pH	8.03E-03
free sulfur dioxide	1.09E-01
citric acid	4.46E-01
density	4.74E-01
residual sugar	8.56E-01

5th - w/ Alch, VAcid, Sulphates, & Total SD	
chlorides	1.43E-05
pH	5.45E-03
fixed acidity	2.68E-02
free sulfur dioxide	7.55E-02
residual sugar	4.33E-01
density	6.16E-01
citric acid	6.80E-01

6th - w/ Alch, VAcid, Sulphates, Ttl SD, chlorides	
pH	1.83E-04
fixed acidity	1.72E-02
free sulfur dioxide	8.73E-02
residual sugar	2.54E-01
citric acid	5.37E-01
density	5.49E-01

7th - w/ Alch, VAcid, Sulphates, Ttl SD, chlrds, ph	
free sulfur dioxide	1.70E-02
citric acid	1.67E-01
residual sugar	4.12E-01
density	7.45E-01
fixed acidity	8.75E-01

8th - w/ Alch, VAcid, Sulphates, Ttl SD, chlrds,	
citric acid	2.88E-01
residual sugar	5.29E-01
density	7.87E-01
fixed acidity	9.00E-01

SUMMARY OUTPUT		
Regression Statistics		
Multiple R	**0.599558934**	
R Square	0.359470916	
Adjusted R Square	**0.356652749**	
Standard Error	0.64774281	
Observations	1599	
	Coefficients	*P-value*
alcohol	0.289302753	4.2364E-61
Intercept	4.430098698	3.72673E-27
volatile acidity	-1.0127527	4.72108E-23
sulphates	0.882665133	1.86484E-15
chlorides	-2.017813817	4.31372E-07
total sulfur dioxide	-0.003482245	4.43483E-07
pH	-0.482661444	4.23496E-05
free sulfur dioxide	0.00507737	0.017021558

Leveraging the variables from our step-wise and correlation analyses, we created a multivariable regressive equation to predict quality outcomes.

Conclusions:

- The variables we can use to determine quality are:
  - Volatile acidity, alcohol, and sulphates
  - Others of significance are chlorides, free sulfur dioxide, total sulfur dioxide, and pH
- By analyzing these variable levels in the wine, we can better predict wine ratings and construct a suitable marketing plan to ensure better profits.