Wine Quality
Analysis Report

Antonio Baričević

December 2020

Table of Contents

1. Introduction

This report analyses the publicly available dataset of the quality of red and white wine.

The Python programming language with various libraries was used for the analysis, and creation of Machine Learning models to predict the outcomes of the marketing campaign.

The goal of this analysis is to find the parameters that are the most important parameters that determine the quality of wine and creating an ML model that can predict the quality score of the wine.

2. Data Information

Contains two data sets because there is a significant difference in properties between red and white wines which means that models for white and red wine are separately created. The data sets are related to the red and white variants of Portuguese wine.

This dataset is publicly available. The details are described in *P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modelling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.*

1. fixed acidity
2. volatile acidity
3. citric acid
4. residual sugar
5. chlorides
6. free sulfur dioxide
7. total sulfur dioxide
8. density
9. pH
10. sulphates
11. alcohol
12. quality (score between 0 and 10

3. Data Analysis

General Information

The data sets contain 6.527 wines - 1.599 red wines, 4.898 white wines and there are no missing values. The differences that occur among wines are as follows:

		Red wine	White wine
Residual sugar	**Mean**	2.54	6.39
	Median	2.2	5.2
	Max	15.5	65.8
Free sulfur dioxide	**Mean**	15.87	35.31
	Median	14.0	34.0
	Max	72.0	289.0
Total sulfur dioxide	**Mean**	46.47	138.36
	Median	38.0	134.0
	Max	289.0	440.0

In other wine properties, there is no significant difference in the intervals.

The influence of red wine properties on the final quality
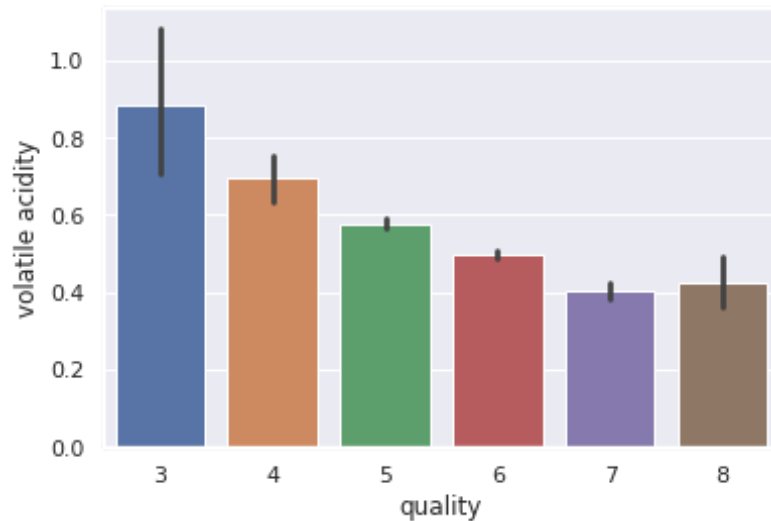
Fixed Acidity

We can see that there is no difference in the amount of fixed acidity that affects the quality of the wine.
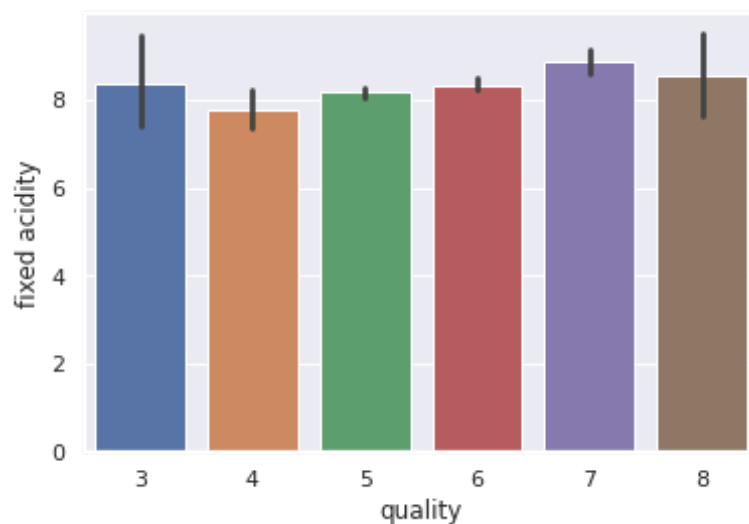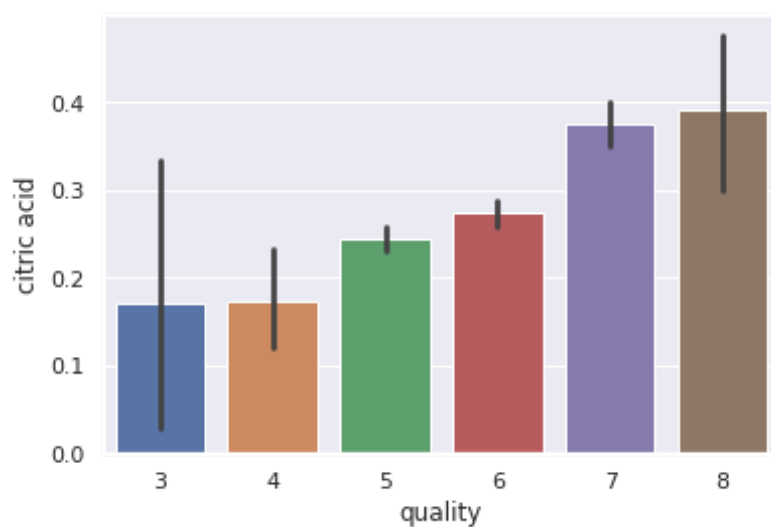
Volatile Acidity

On the other hand, the quality of volatile acidity has an impact on quality. More volatile acidity = less quality. The best-rated wines (score 7 and 8) have an average of 0.4 volatile acidity.
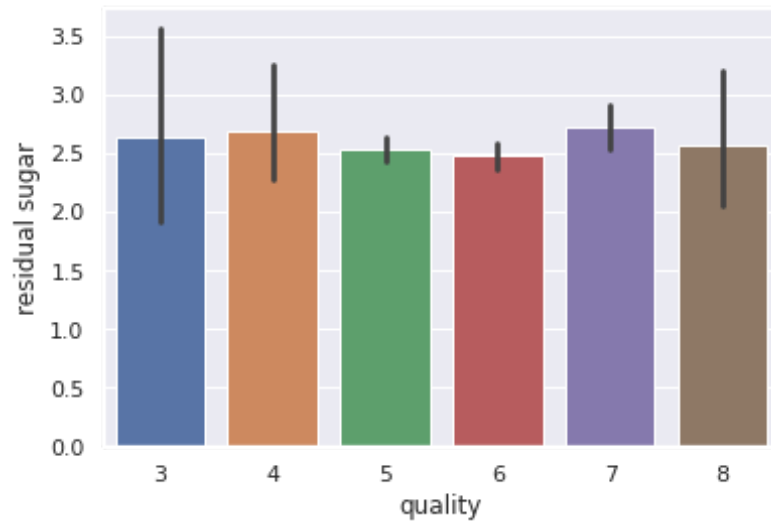

Citric Acid

When we check citric acid, then we have the opposite situation from than before. In this case, a better quality of wine has more citric acid.
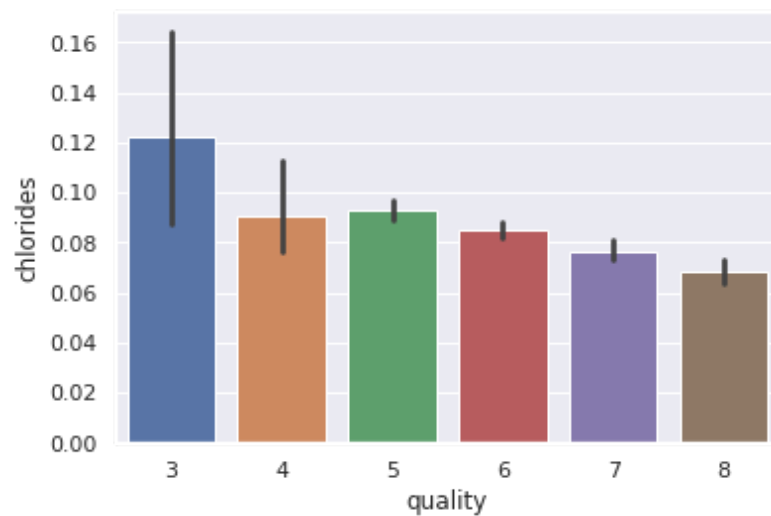
Residual Sugar

As with fixed acidity, residual sugar also does not affect on quality since the value is equal among all scores.
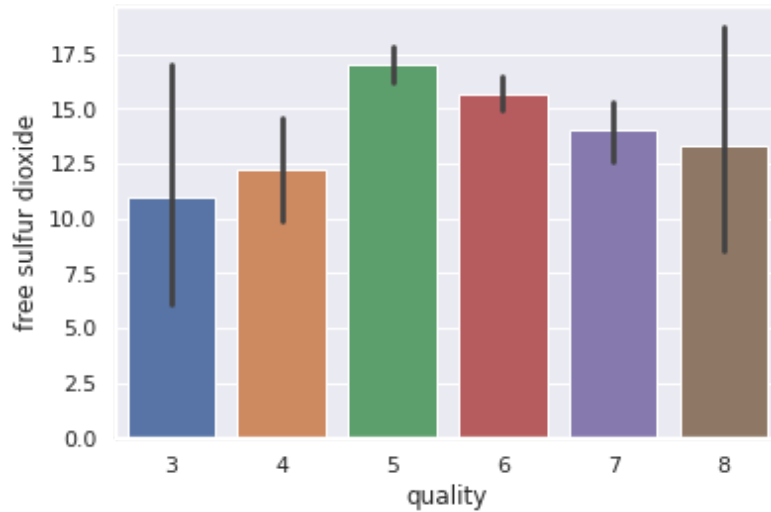

Chlorides

A high amount of chlorides (0.12) can bring poorer wine quality. If the chloride level is between 0.06 and 0.08, there is a greater probability that the wine will be of better quality.
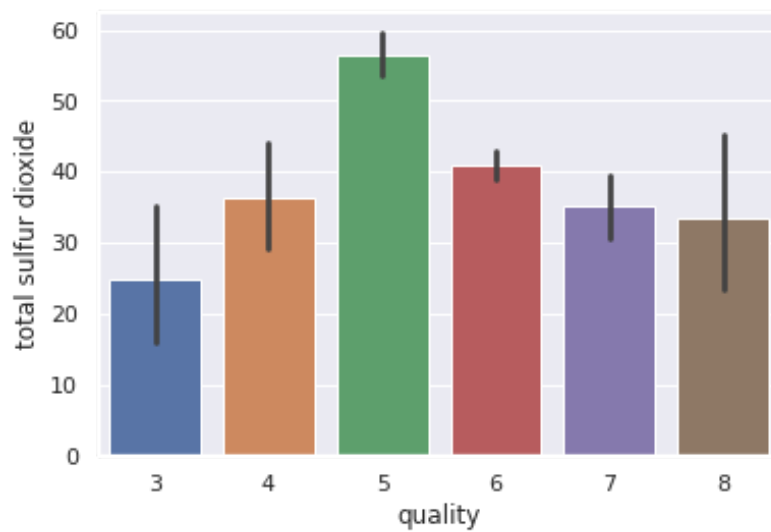
Free sulfur dioxide

Not too much or too little. If there is too much free sulfur dioxide then we have the average wine, and if is too little then the wine is bad. The balance is the key.
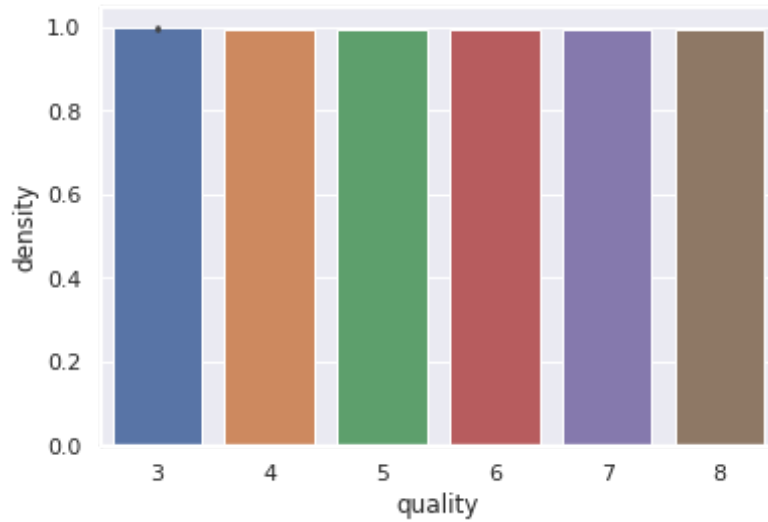

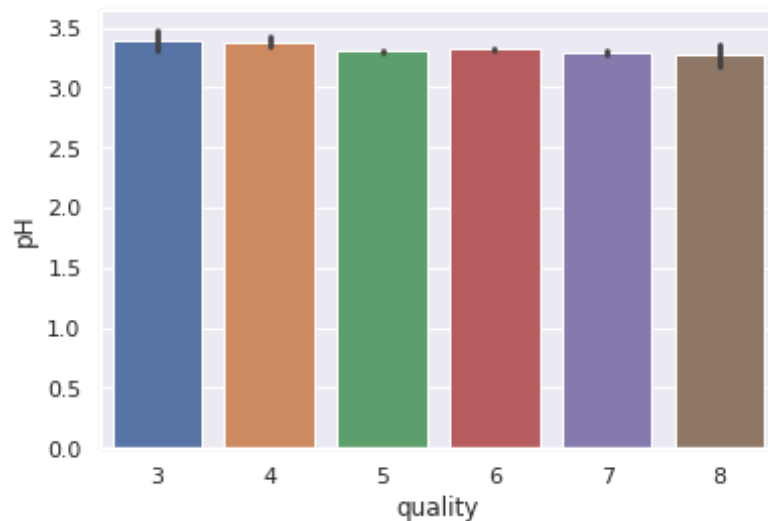
Total sulfur dioxide

Same as for the free sulfur dioxide.

Density

No impact. All wines have identical values. The minimum value is 0.990 and the maximum 0.977 so it doesn't matter.
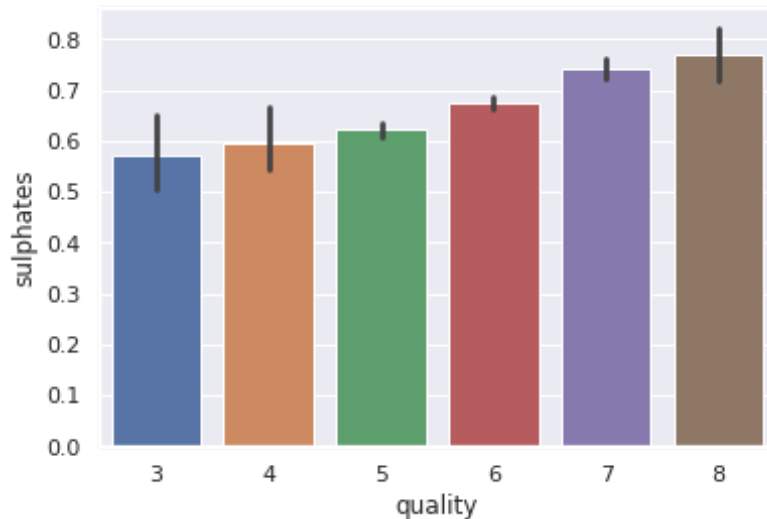

pH

Same as density. The wines have small differences. The minimum value is 2.74 and the maximum is 4.01. The mean and median are the same 3.31.
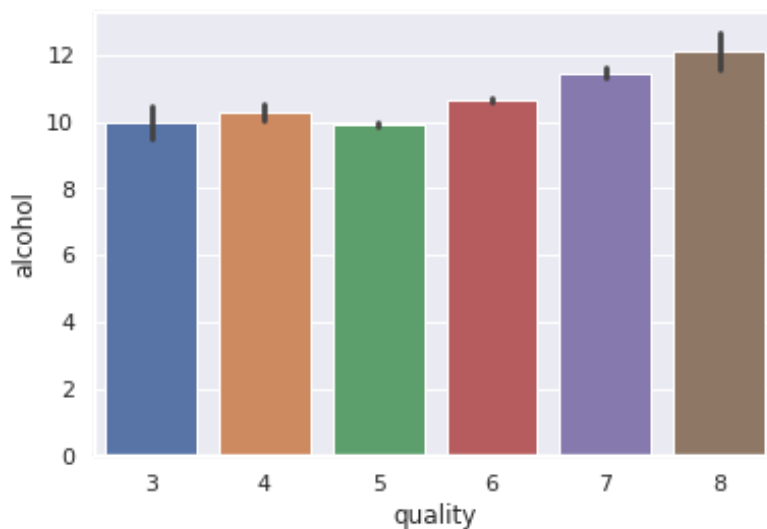
Sulphates

Another property where higher value brings better quality. Better wines have more than 0.7 sulphates. The averages are between 0.6 and 0.7, and all below 0.6 brings low-quality score.


Alcohol

And the last property is alcohol. The wines rated with the highest rating in this data set have 12% alcohol. Wines with 10% alcohol are below average quality.

The influence of white wine properties on the final quality

Fixed Acidity

The worst-rated wine has an approximate value of fixed acidity as well as the best-rated wine. Other wines are somewhere in between so it cannot be confirmed that this property affects the quality of the white wine.
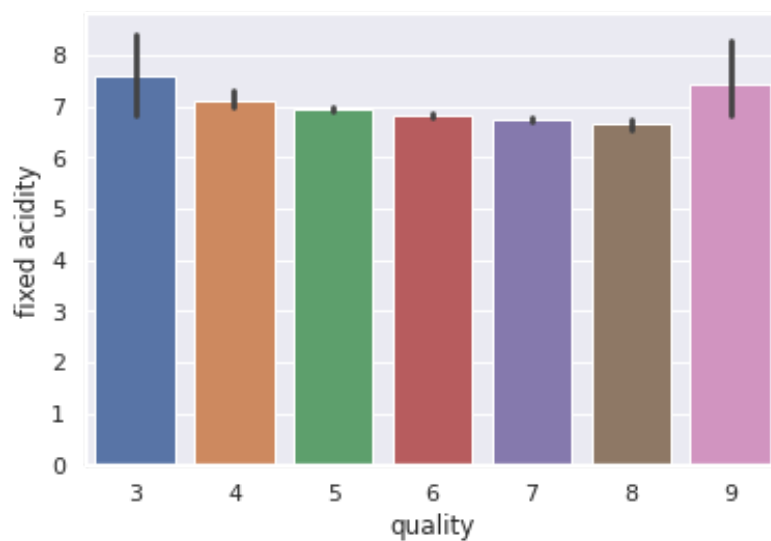

Volatile Acidity

More volatile acidity means lower quality scores (scores 3 and 4). All other scores have volatile acidity between 0.25 and 0.3.
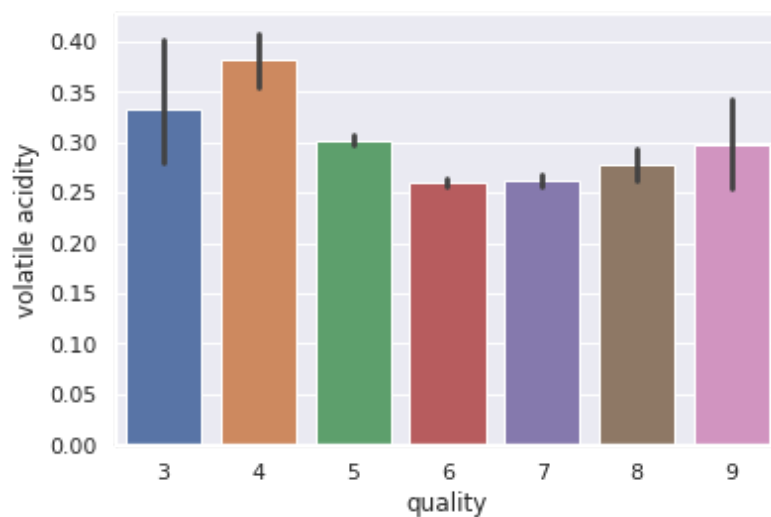
Citric Acid

For all quality scores the citric acid is approximately equal, around 0.32 and 0.35, except for the best rated which is close to 0.4. Since 5 out of 7 scores have approximately identical values, the significant importance of the property cannot be confirmed.
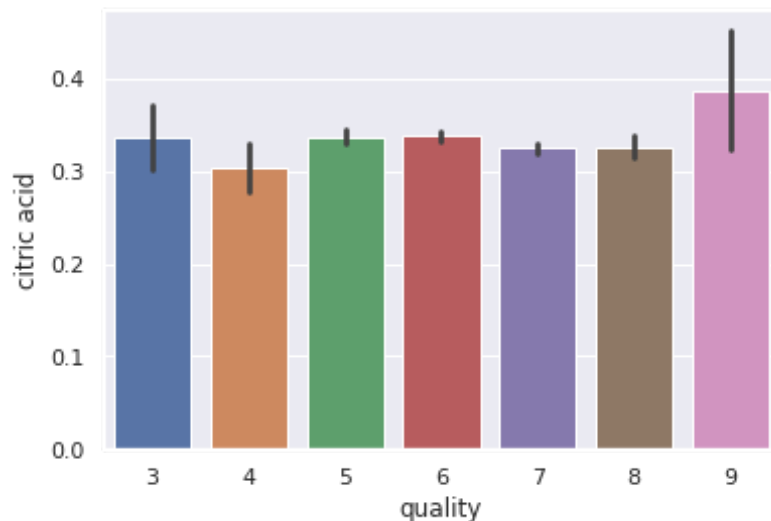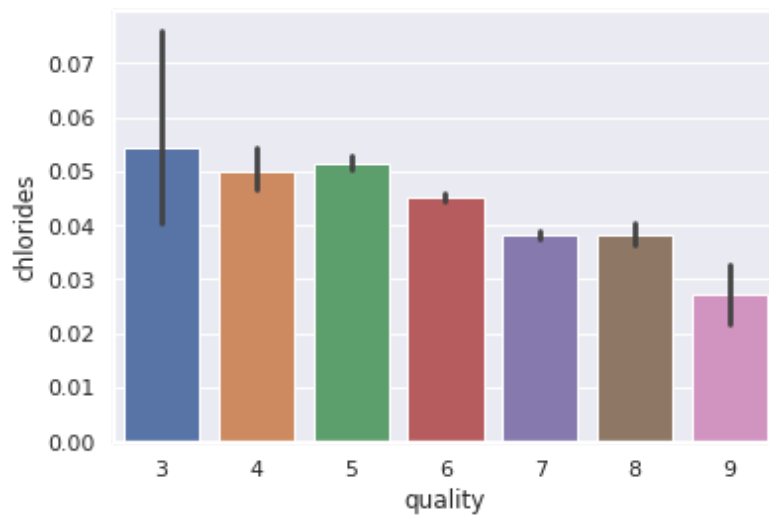

Residual Sugar

The residual sugar varies across all quality scores, so it cannot be confirmed as an extremely important property that contributes to the final quality score. There is not too much difference between wines rated 9 to those rated 4.
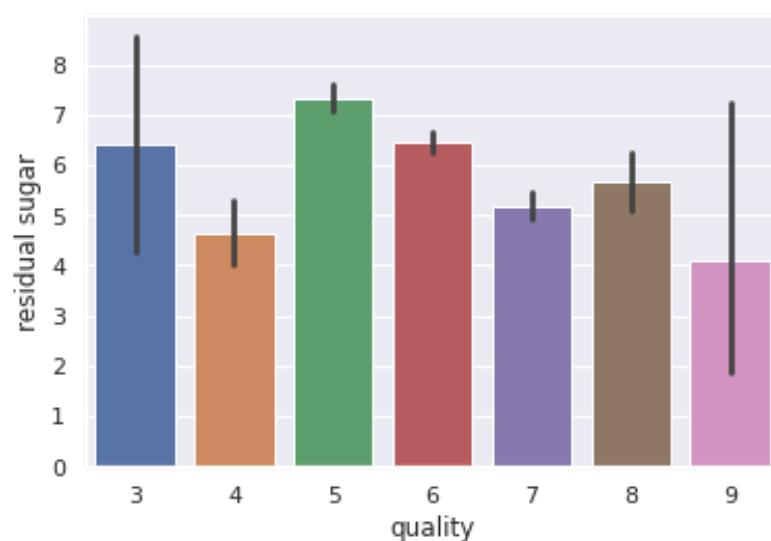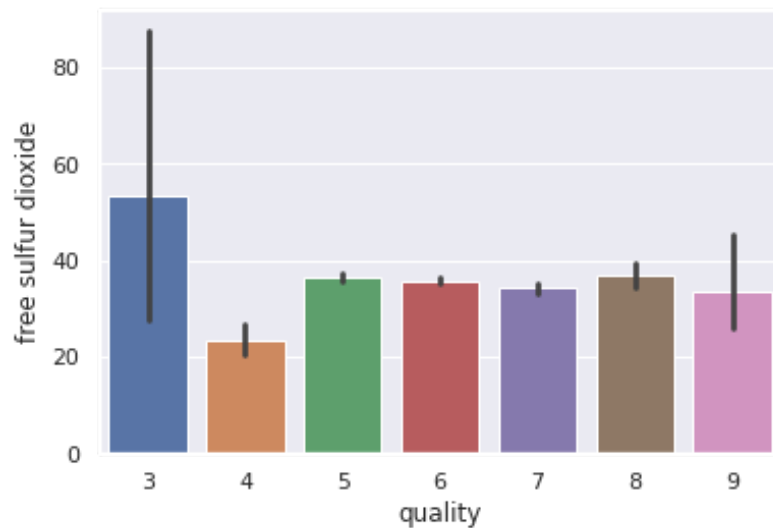
Chlorides

A lower amount of chloride = higher quality of the wine.

Free sulfur dioxide

The medium-rated wines, as well as better quality wines are around 36-38 amount of free sulfur dioxide, while the wines with the worst quality score are not even close to that number.
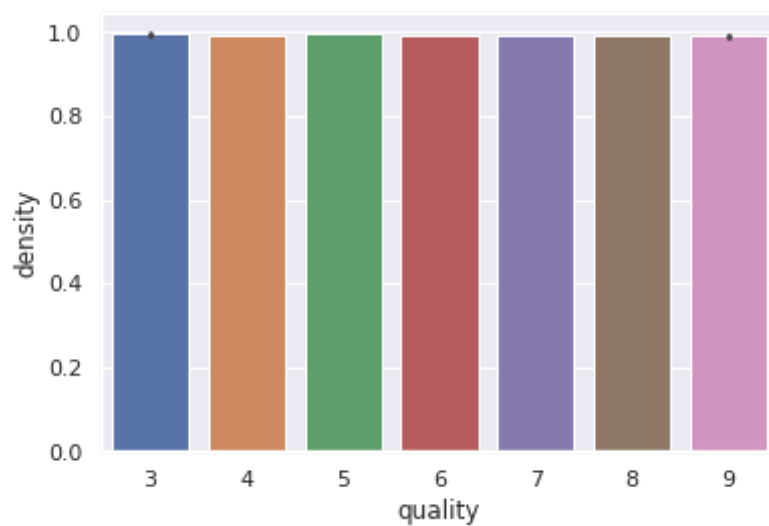
Total sulfur dioxide

Similar to free sulfur dioxide. With a smaller amount of total sulfur dioxide (110-130) the probability of a better-rated wine is higher.
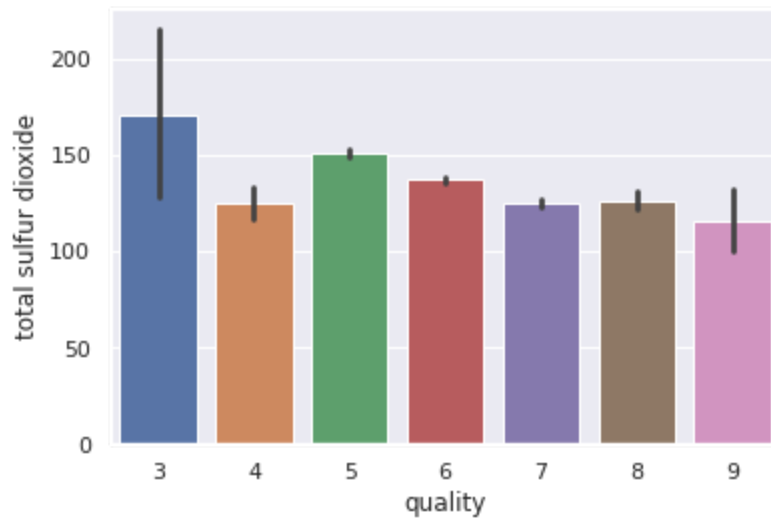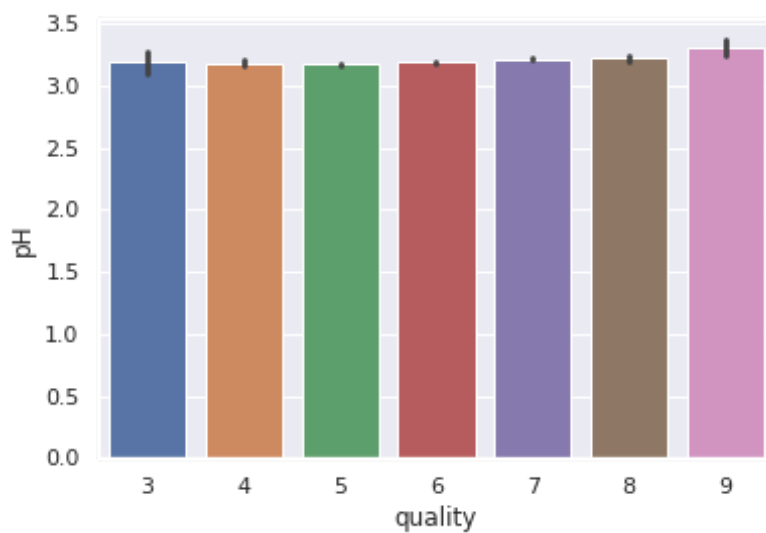

Density

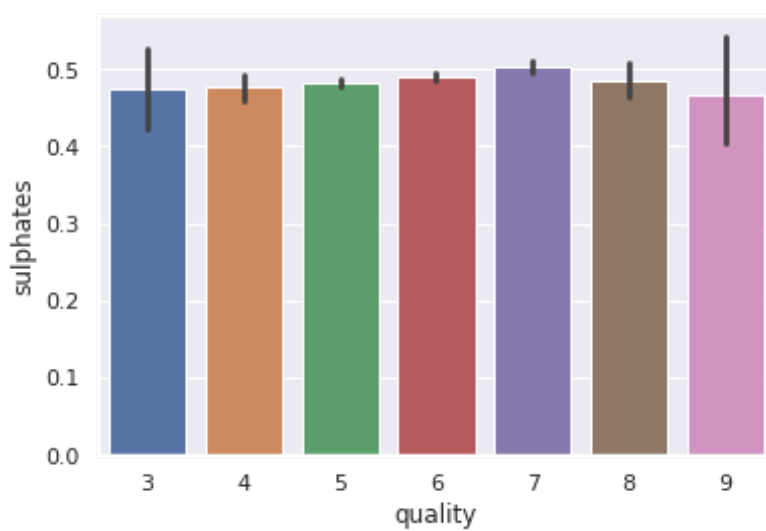As was the case with red wine, density value does not effect.

pH

Wines with better quality scores have a higher pH value, but the difference is minimal.
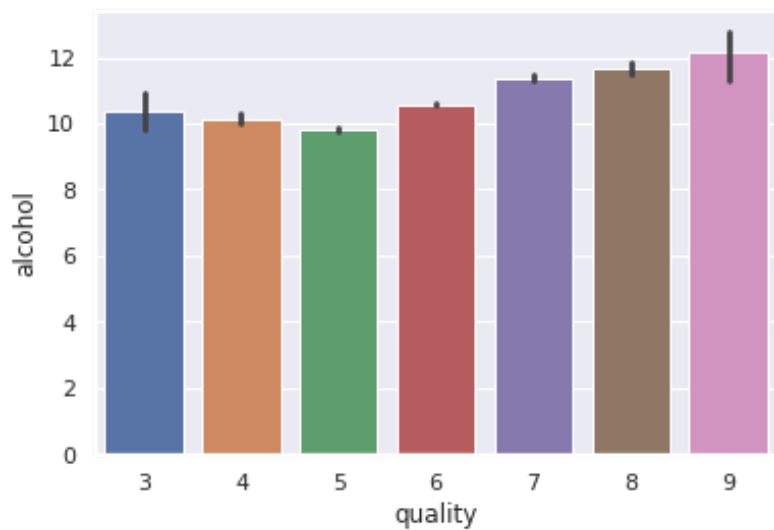

Sulphates

The values are identical among all quality scores so there is no impact.

Alcohol

Sane as with red wine. The wines rated with the highest rating in this data set have 12% alcohol. Wines with around 10% alcohol are below average quality.

4. Conclusions

The red wines that have the majority of the following properties are more likely to get a better quality score:

- less volatile acidity - around 0.4

- more citric acid - between 0.35 and 0.4

- chlorides between 0.06 and 0.08

- total sulfur dioxide between 30 and 35

- more sulphates - higher than 0.7

- more alcohol - around 12%

The white wines that have the majority of the following properties are more likely to get a better quality score:

- volatile acidity between 0.25 and 0.3

- a lower amount of chloride - between 0.26 and 0.38

- total sulfur dioxide between 110 and 130

- more alcohol - around 12%