Wine Quality Analytics System for BlueBerry Winery

Exploratory Data Analysis of Vinho Verde

Analytics & Research, Michal Pasternak

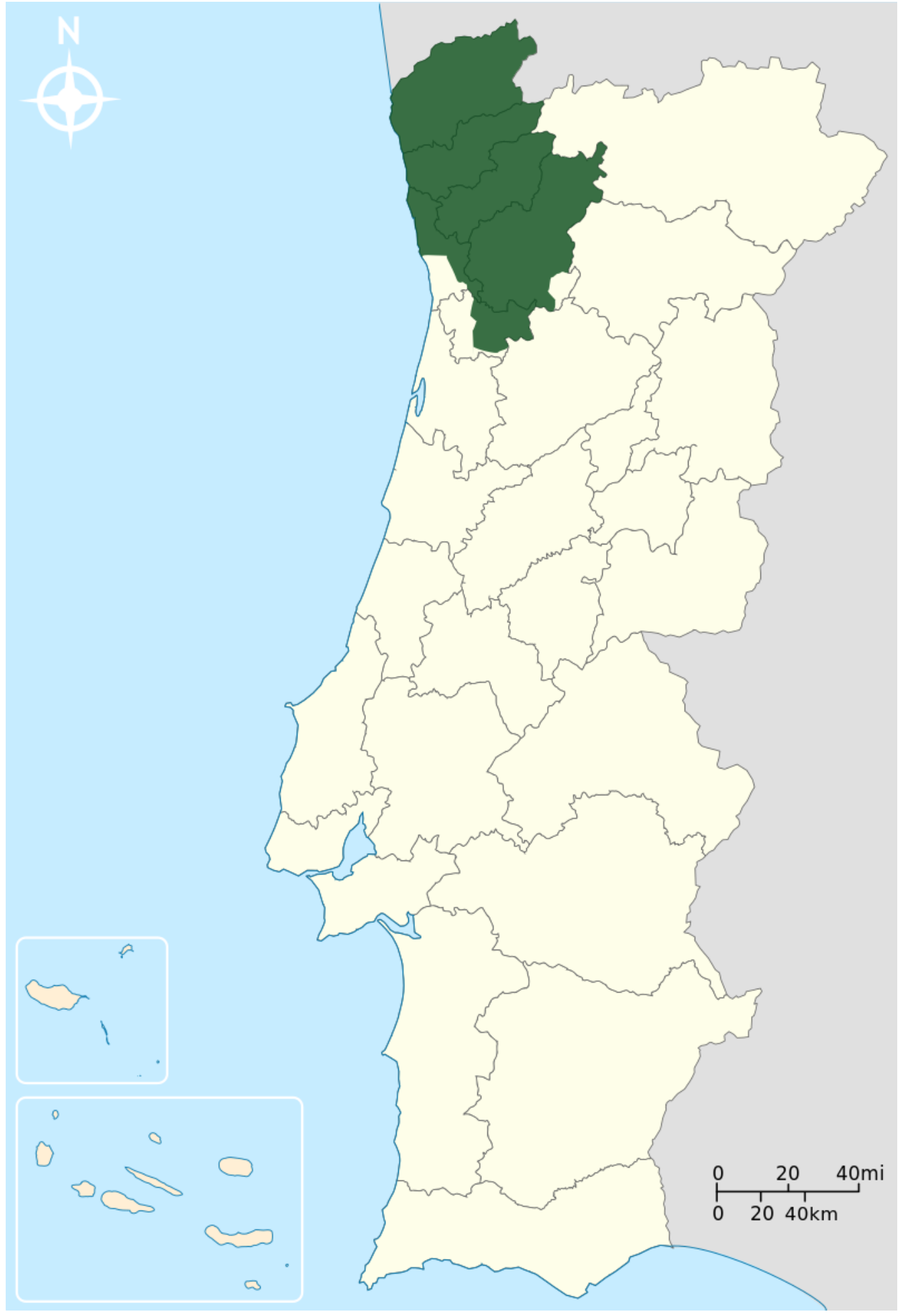
Agenda

1. Vinho Verde
2. Dataset
3. Goal
4. Research Questions
5. Exploratory Data Analysis
6. Findings

Vinho Verde

Vinho Verde is a unique type of wine that comes from the Vinho Verde region in **Portugal**. It is known for its crisp and **refreshing** qualities, making it a popular choice, especially in warmer weather. Typically, Vinho Verde wines are consumed when they are **young**. They are known for their **light and lively** nature. Many of the wines exhibit **fruity** flavors, such as citrus, green apple, and tropical fruit notes.

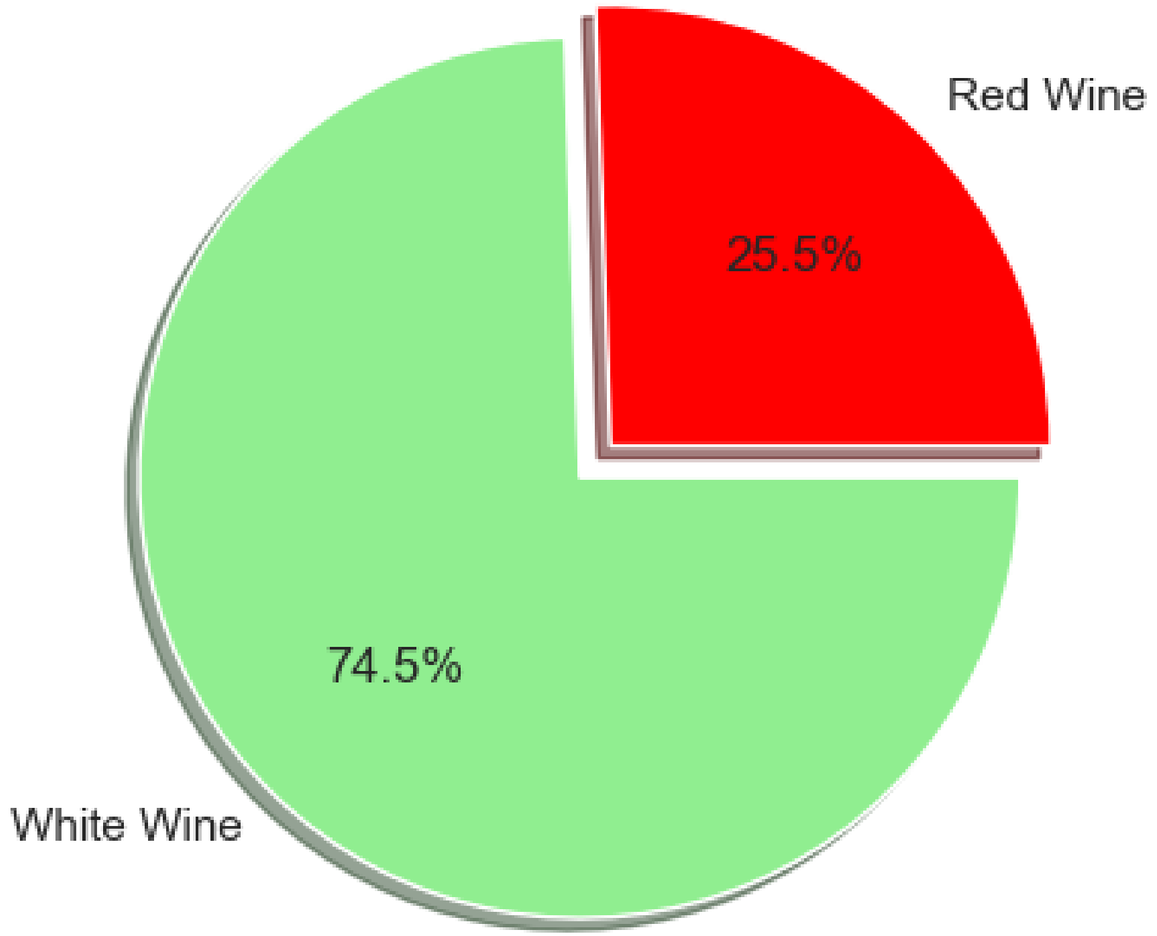
The Vinho Verde region's unique terroir, with its Atlantic Ocean influence, contributes to the wine's distinct characteristics. The maritime climate and soil composition play a role in shaping the flavor profile.



Dataset

Number of Instances:

- red wine - 1359
- white wine - 3961



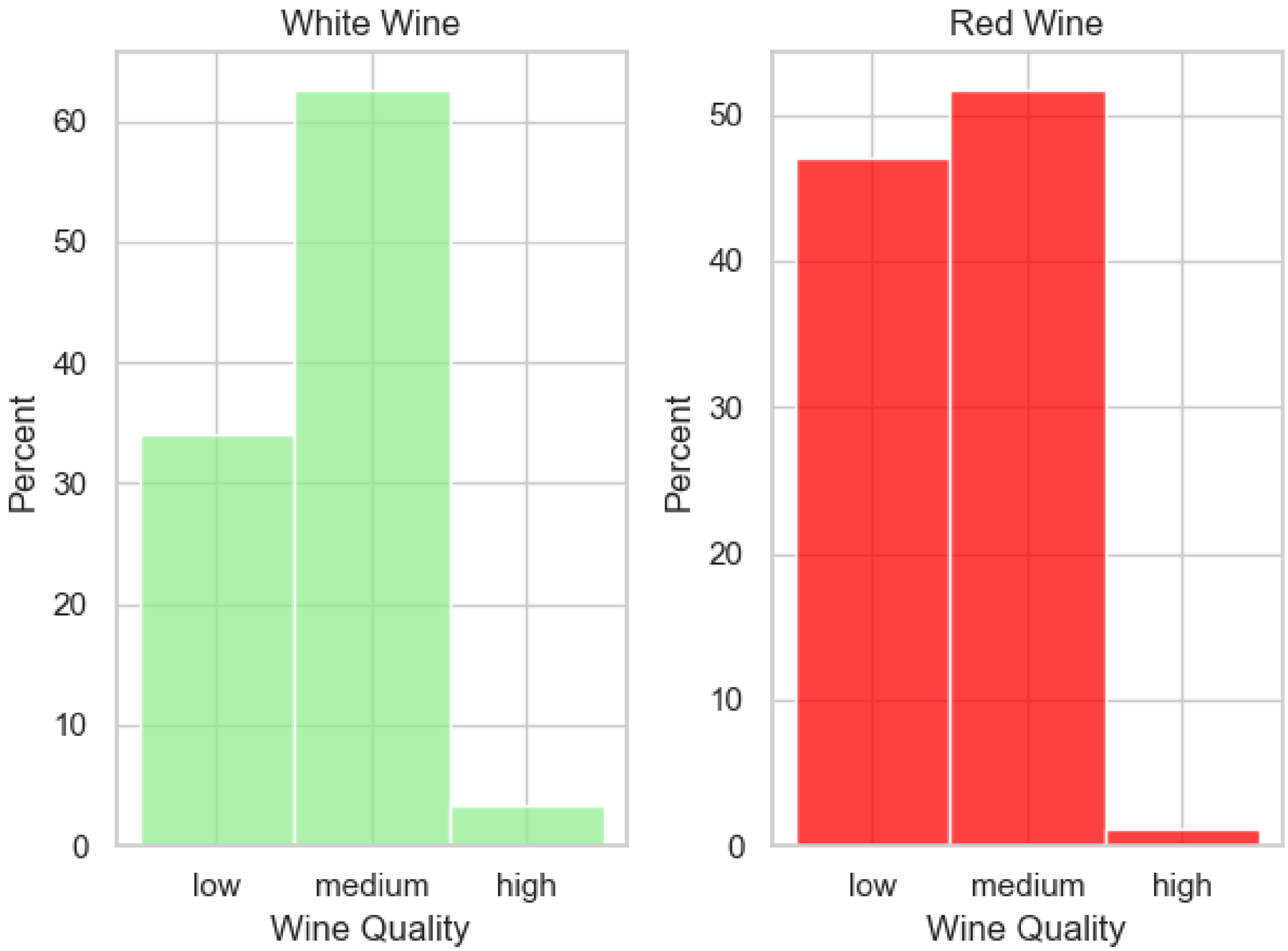
Dataset

Attributes:

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

Dataset

Quality



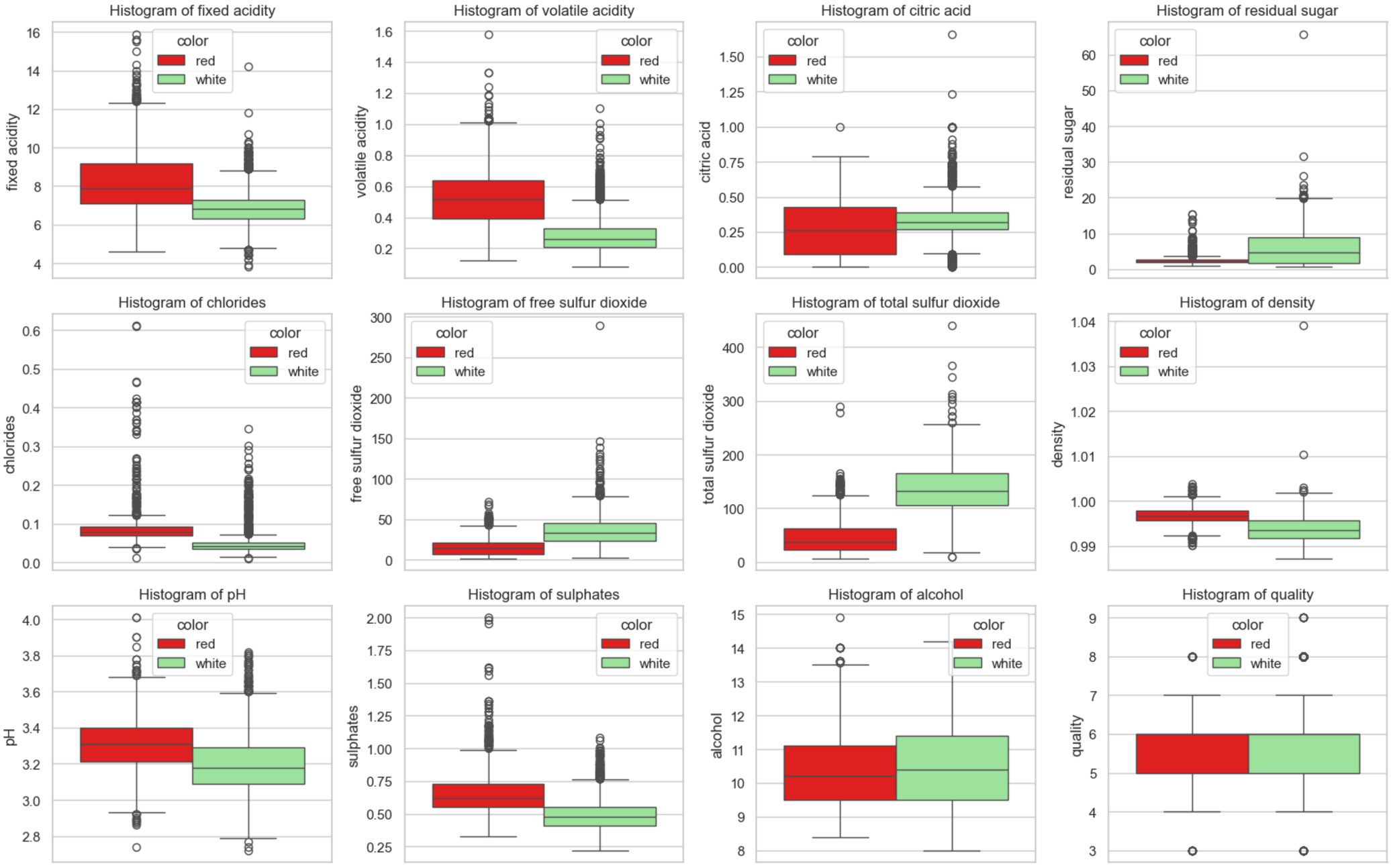
How can we produce more good quality wine?

Research Questions

1. Is the chemical composition of red and white wines comparable? Is it a good idea to analyse them together?
2. Does the chemical composition have any impact on the perceived quality of the wine?
3. What should we pay attention to, when making good quality wine?
4. What chemical composition would a good wine from the region have?

Research Questions

Is the chemical composition of red and white wines comparable? Is it a good idea to analyse them together?


Histogram of fixed acidity
fixed acidity
Histogram of volatile acidity
volatile acidity
Histogram of citric acid
citric acid
Histogram of residual sugar
residual sugar
Histogram of chlorides
chlorides
Histogram of free sulfur dioxide
free sulfur dioxide
Histogram of total sulfur dioxide
total sulfur dioxide
Histogram of density
density
Histogram of pH
pH
Histogram of sulphates
sulphates
Histogram of alcohol
alcohol
Histogram of quality
quality
color
red
white

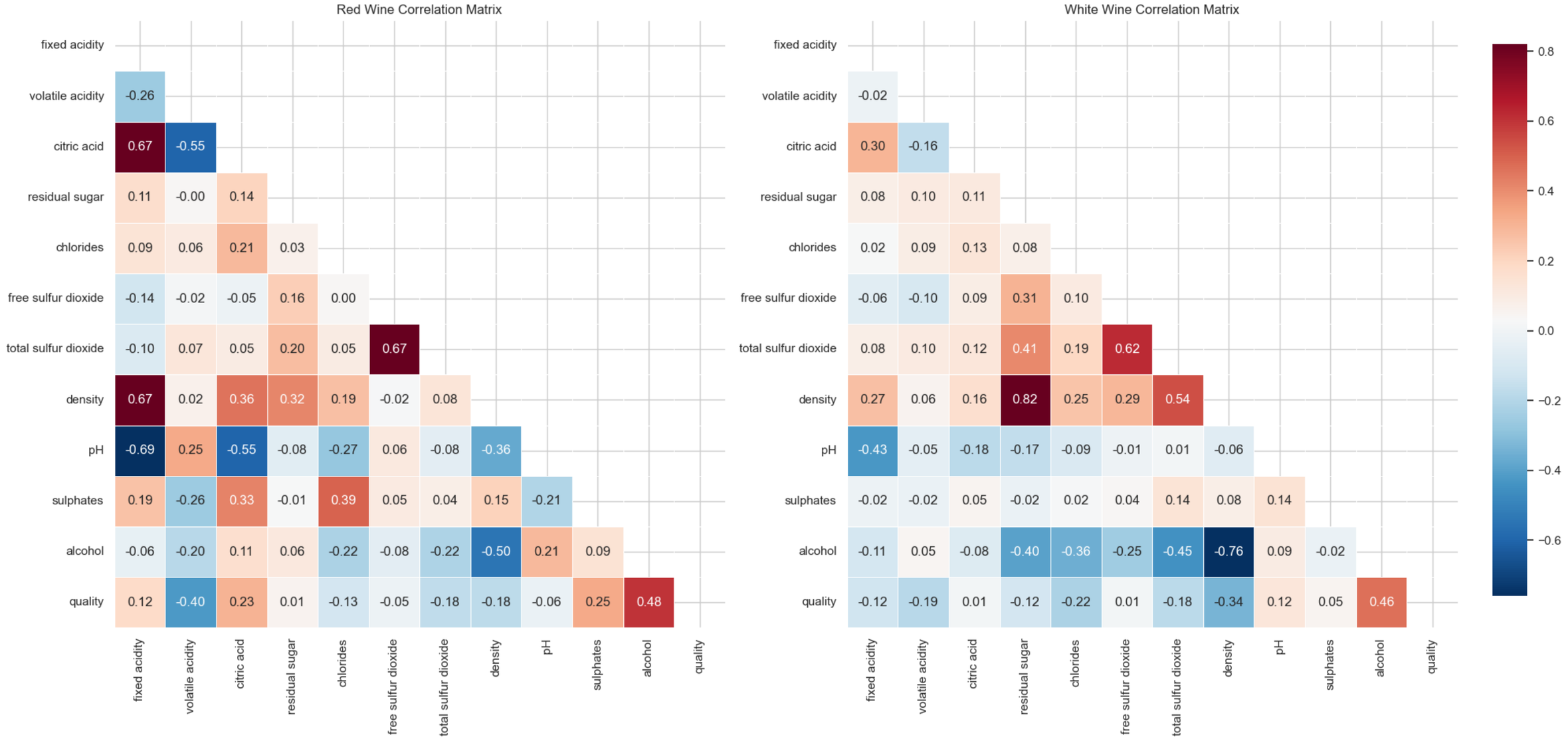
Findings

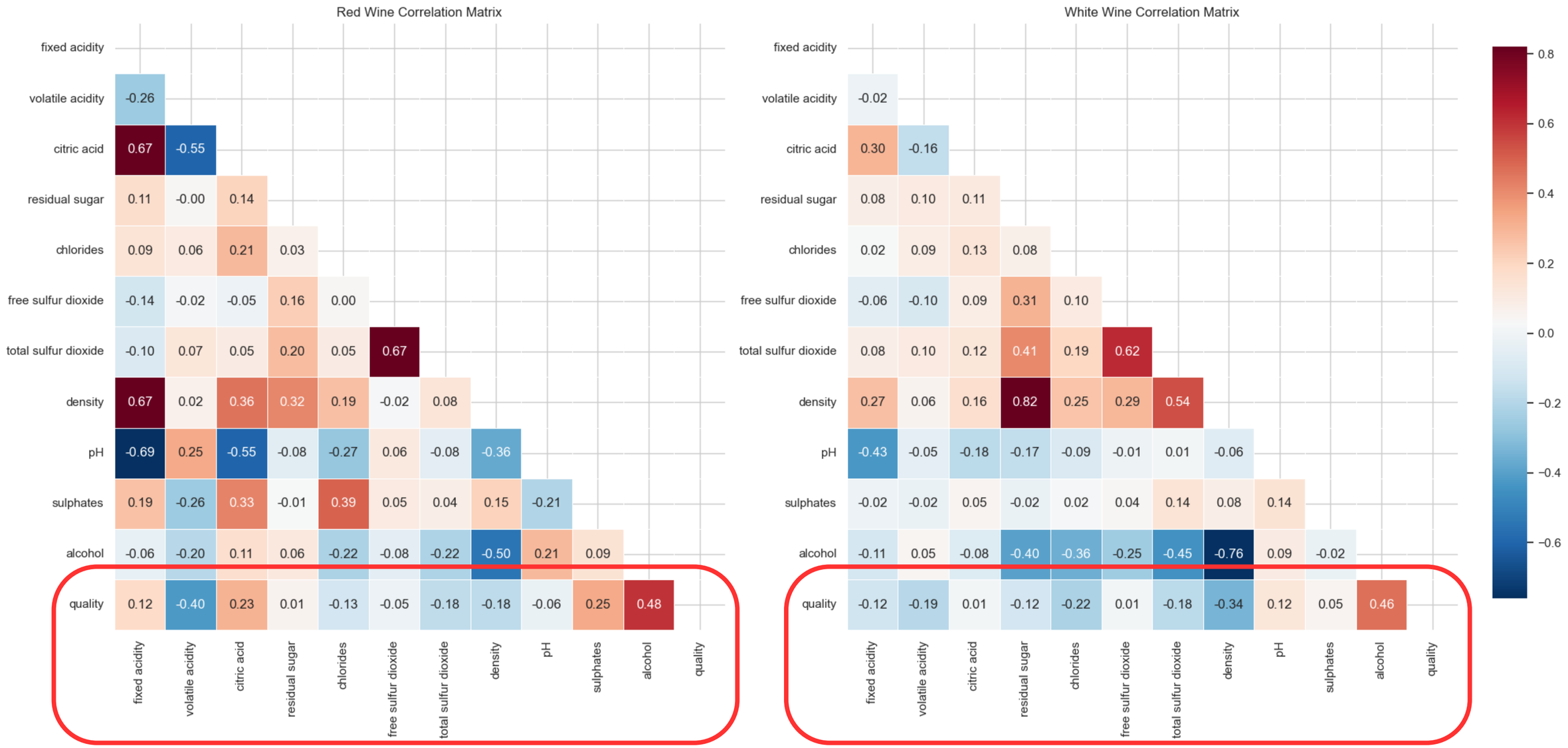
- The average values for the chemical properties of the two different wine types rarely overlap which indicates significant differences between the chemical compositions of red and white wine.

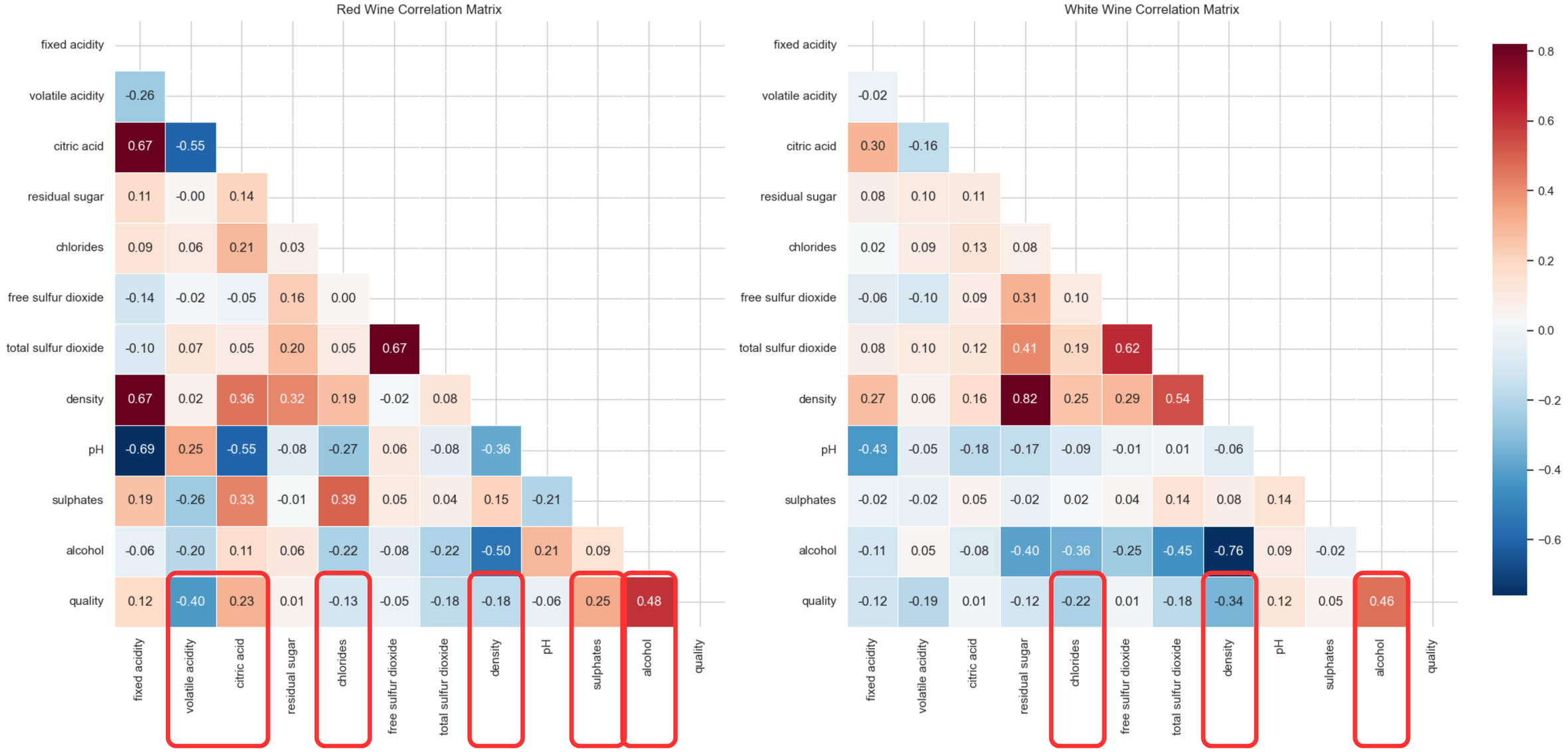
- This is further confirmed when we perform a statistical tests (t-test).

Research Questions

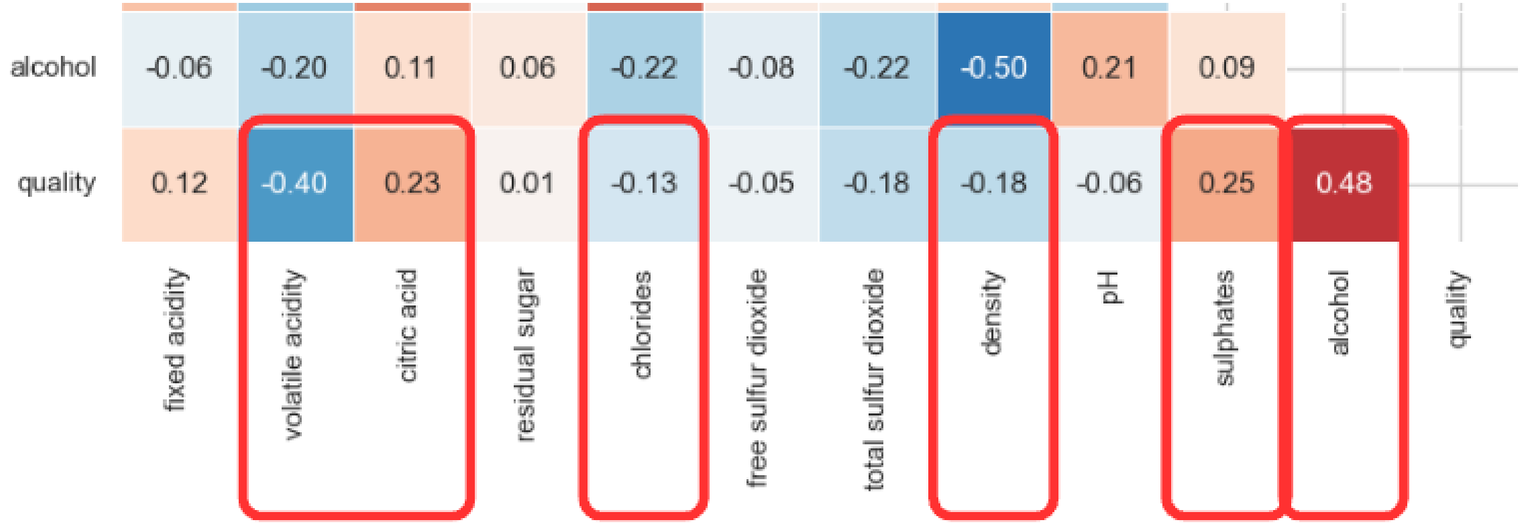
Does the chemical composition have any impact on the perceived quality of the wine?


Red Wine Correlation Matrix
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality
-0.26
0.67 -0.55
0.11 -0.00 0.14
0.09 0.06 0.21 0.03
-0.14 -0.02 -0.05 0.16 0.00
-0.10 0.07 0.05 0.20 0.05 0.67
0.67 0.02 0.36 0.32 0.19 -0.02 0.08
-0.69 0.25 -0.55 -0.08 -0.27 0.06 -0.08 -0.36
0.19 -0.26 0.33 -0.01 0.39 0.05 0.04 0.15 -0.21
-0.06 -0.20 0.11 0.06 -0.22 -0.08 -0.22 -0.50 0.21 0.09
0.12 -0.40 0.23 0.01 -0.13 -0.05 -0.18 -0.18 -0.06 0.25 0.48
White Wine Correlation Matrix
-0.02
0.30 -0.16
0.08 0.10 0.11
0.02 0.09 0.13 0.08
-0.06 -0.10 0.09 0.31 0.10
0.08 0.10 0.12 0.41 0.19 0.62
0.27 0.06 0.16 0.82 0.25 0.29 0.54
-0.43 -0.05 -0.18 -0.17 -0.09 -0.01 0.01 -0.06
-0.02 -0.02 0.05 -0.02 0.02 0.04 0.14 0.08 0.14
-0.11 0.05 -0.08 -0.40 -0.36 -0.25 -0.45 -0.76 0.09 -0.02
-0.12 -0.19 0.01 -0.12 -0.22 0.01 -0.18 -0.34 0.12 0.05 0.46
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

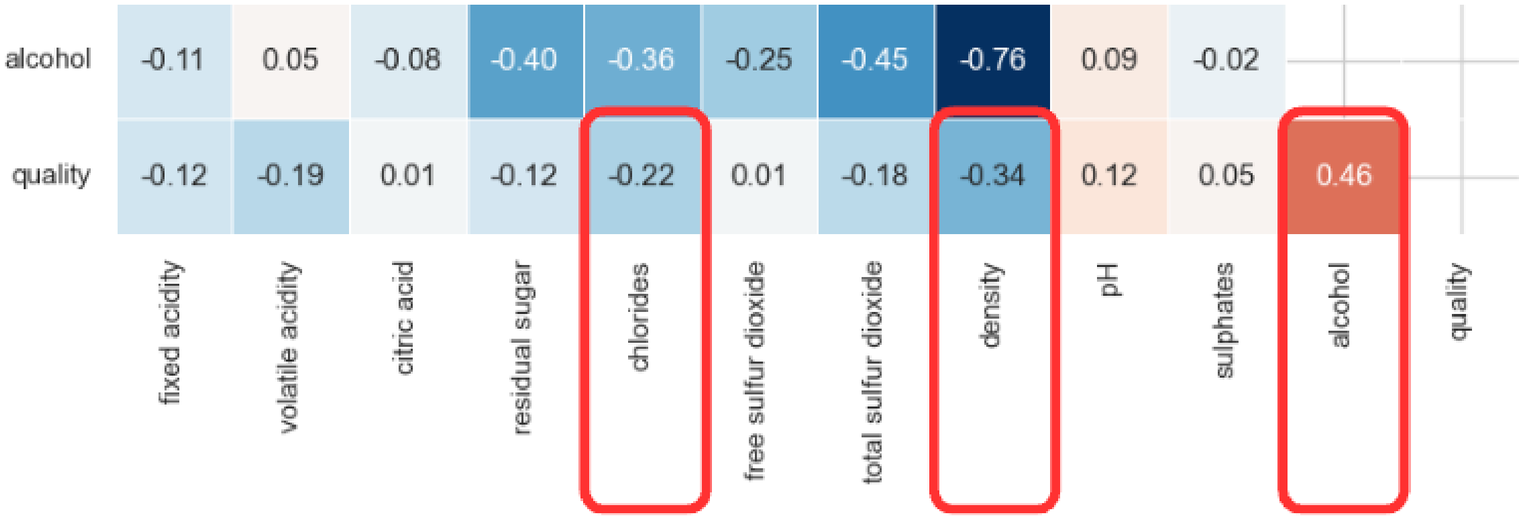

Red Wine Correlation Matrix
fixed acidity
volatile acidity -0.26
citric acid 0.67 -0.55
residual sugar 0.11 -0.00 0.14
chlorides 0.09 0.06 0.21 0.03
free sulfur dioxide -0.14 -0.02 -0.05 0.16 0.00
total sulfur dioxide -0.10 0.07 0.05 0.20 0.05 0.67
density 0.67 0.02 0.36 0.32 0.19 -0.02 0.08
pH -0.69 0.25 -0.55 -0.08 -0.27 0.06 -0.08 -0.36
sulphates 0.19 -0.26 0.33 -0.01 0.39 0.05 0.04 0.15 -0.21
alcohol -0.06 -0.20 0.11 0.06 -0.22 -0.08 -0.22 -0.50 0.21 0.09
quality 0.12 -0.40 0.23 0.01 -0.13 -0.05 -0.18 -0.18 -0.06 0.25 0.48
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality
White Wine Correlation Matrix
fixed acidity
volatile acidity -0.02
citric acid 0.30 -0.16
residual sugar 0.08 0.10 0.11
chlorides 0.02 0.09 0.13 0.08
free sulfur dioxide -0.06 -0.10 0.09 0.31 0.10
total sulfur dioxide 0.08 0.10 0.12 0.41 0.19 0.62
density 0.27 0.06 0.16 0.82 0.25 0.29 0.54
pH -0.43 -0.05 -0.18 -0.17 -0.09 -0.01 0.01 -0.06
sulphates -0.02 -0.02 0.05 -0.02 0.02 0.04 0.14 0.08 0.14
alcohol -0.11 0.05 -0.08 -0.40 -0.36 -0.25 -0.45 -0.76 0.09 -0.02
quality -0.12 -0.19 0.01 -0.12 -0.22 0.01 -0.18 -0.34 0.12 0.05 0.46
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


Red Wine Correlation Matrix
fixed acidity
volatile acidity -0.26
citric acid 0.67 -0.55
residual sugar 0.11 -0.00 0.14
chlorides 0.09 0.06 0.21 0.03
free sulfur dioxide -0.14 -0.02 -0.05 0.16 0.00
total sulfur dioxide -0.10 0.07 0.05 0.20 0.05 0.67
density 0.67 0.02 0.36 0.32 0.19 -0.02 0.08
pH -0.69 0.25 -0.55 -0.08 -0.27 0.06 -0.08 -0.36
sulphates 0.19 -0.26 0.33 -0.01 0.39 0.05 0.04 0.15 -0.21
alcohol -0.06 -0.20 0.11 0.06 -0.22 -0.08 -0.22 -0.50 0.21 0.09
quality 0.12 -0.40 0.23 0.01 -0.13 -0.05 -0.18 -0.18 -0.06 0.25 0.48
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality
White Wine Correlation Matrix
fixed acidity
volatile acidity -0.02
citric acid 0.30 -0.16
residual sugar 0.08 0.10 0.11
chlorides 0.02 0.09 0.13 0.08
free sulfur dioxide -0.06 -0.10 0.09 0.31 0.10
total sulfur dioxide 0.08 0.10 0.12 0.41 0.19 0.62
density 0.27 0.06 0.16 0.82 0.25 0.29 0.54
pH -0.43 -0.05 -0.18 -0.17 -0.09 -0.01 0.01 -0.06
sulphates -0.02 -0.02 0.05 -0.02 0.02 0.04 0.14 0.08 0.14
alcohol -0.11 0.05 -0.08 -0.40 -0.36 -0.25 -0.45 -0.76 0.09 -0.02
quality -0.12 -0.19 0.01 -0.12 -0.22 0.01 -0.18 -0.34 0.12 0.05 0.46
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

Red Wine



White Wine



Findings

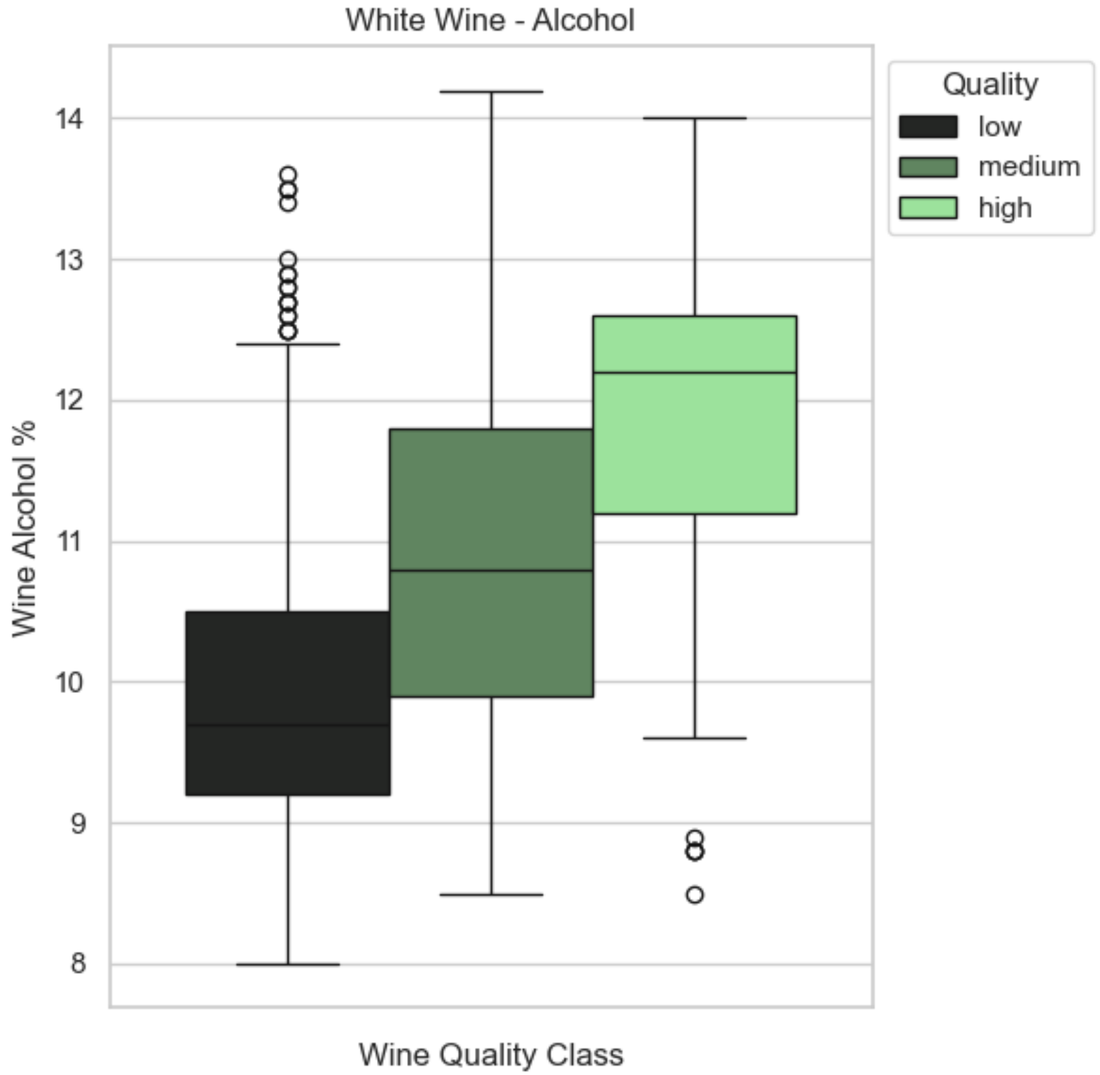
- A lot of the single chemical properties of wine are correlated with it’s percieved quality

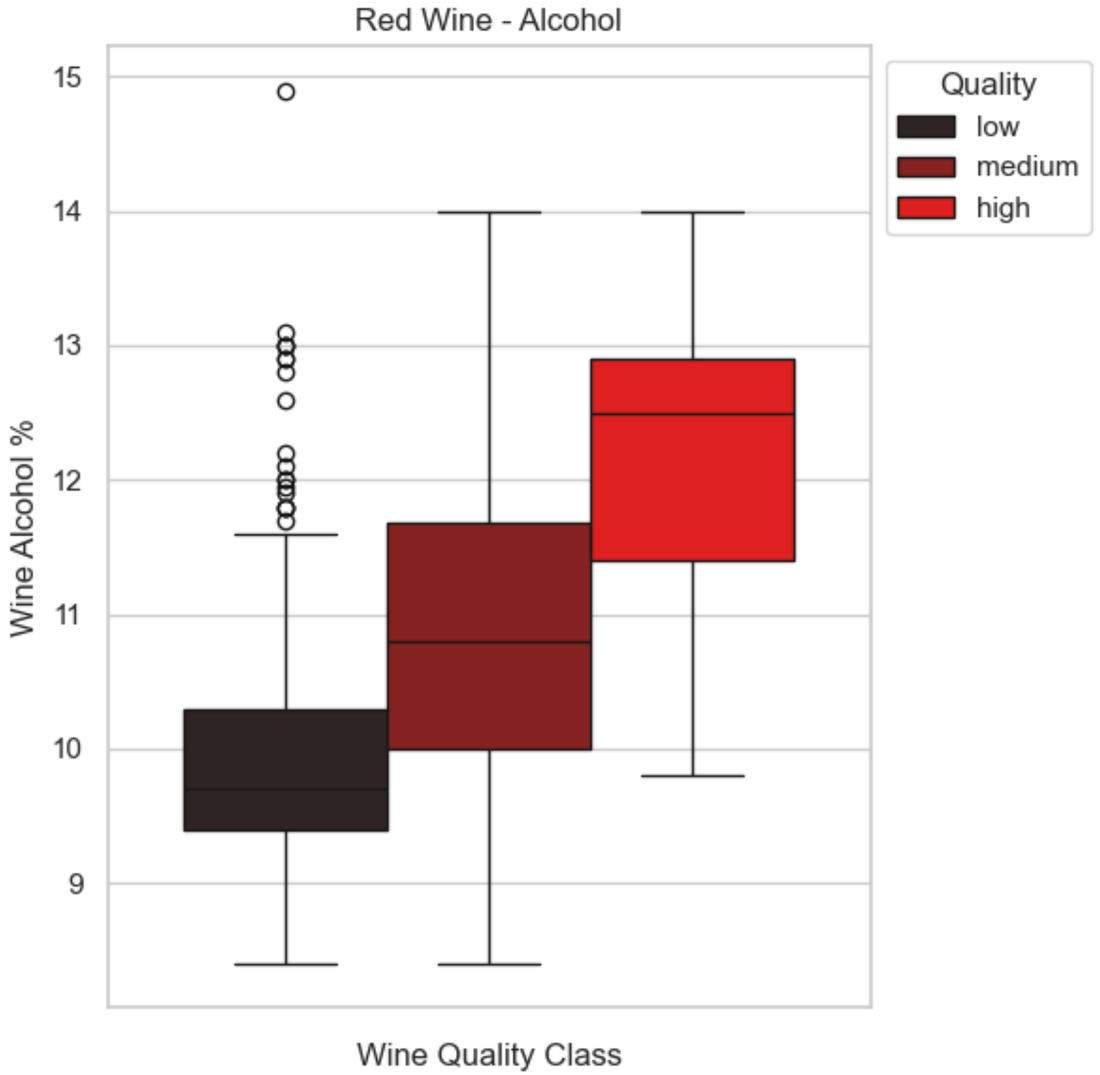
- This is further confirmed when we perform a statistical tests (ANOVA).

Research Questions

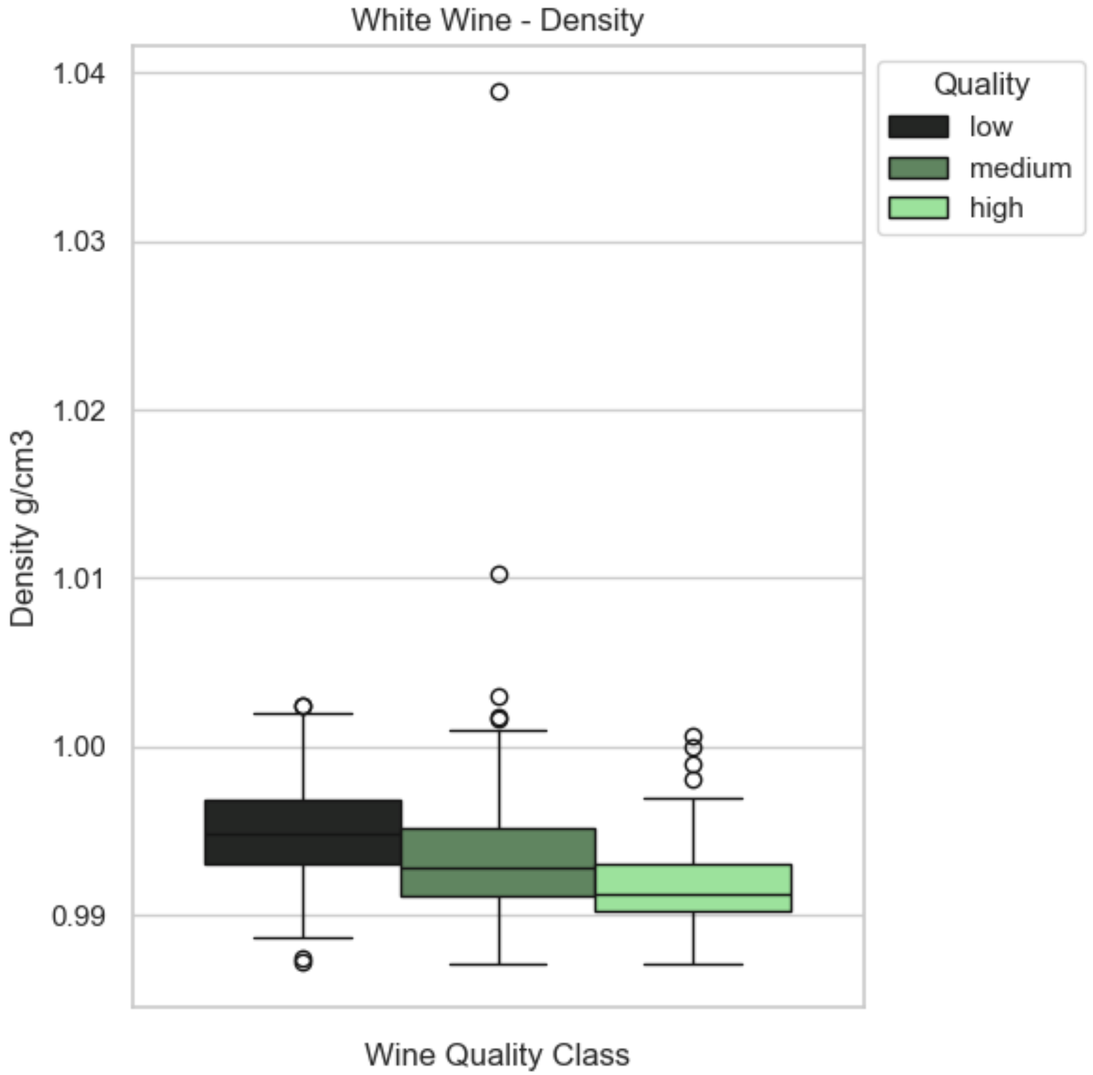
What should we pay attention to, when making good quality wine?

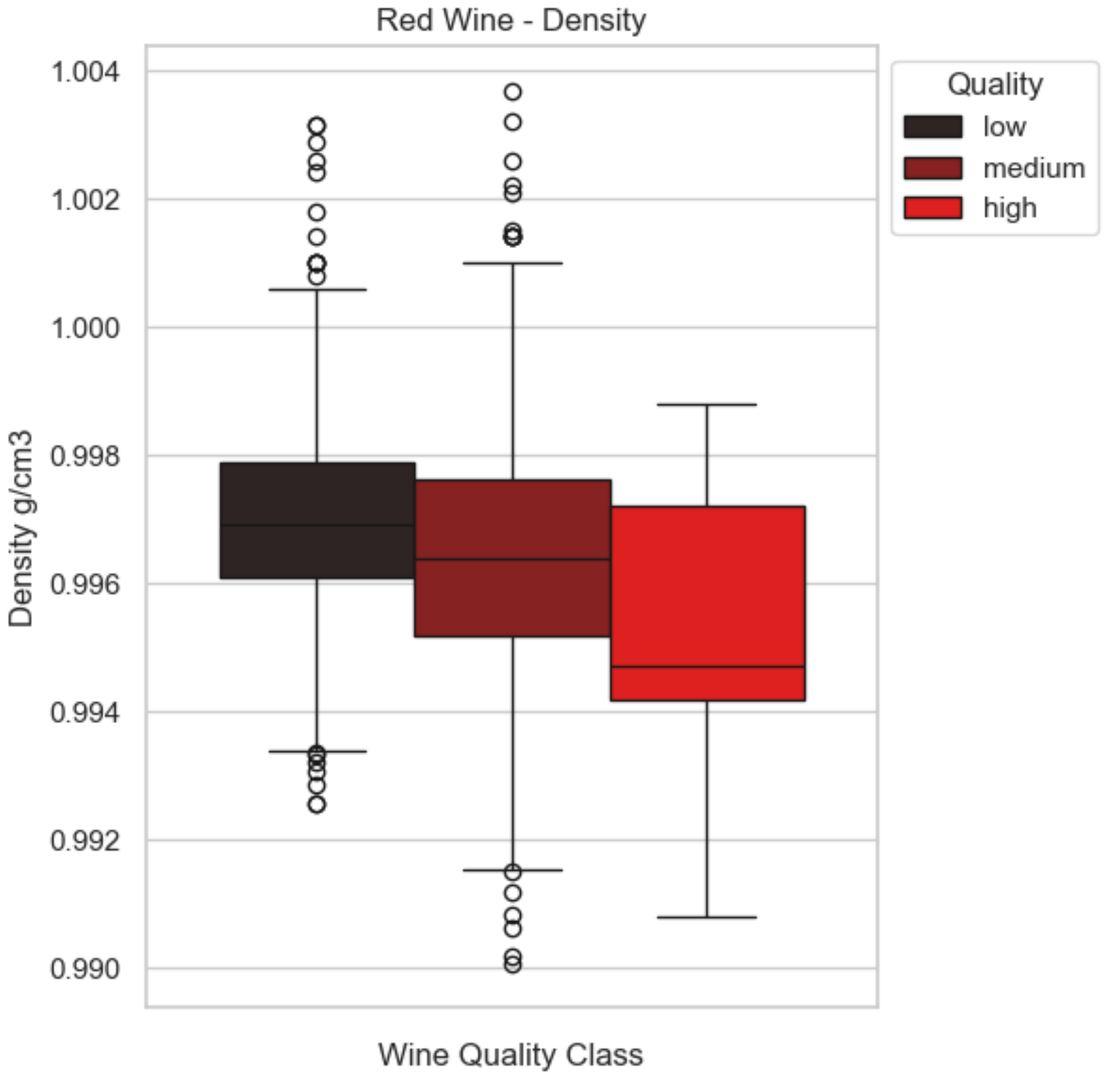
Alcohol


White Wine - Alcohol
Wine Alcohol %
Wine Quality Class
Quality
low
medium
high
14
13
12
11
10
9
8

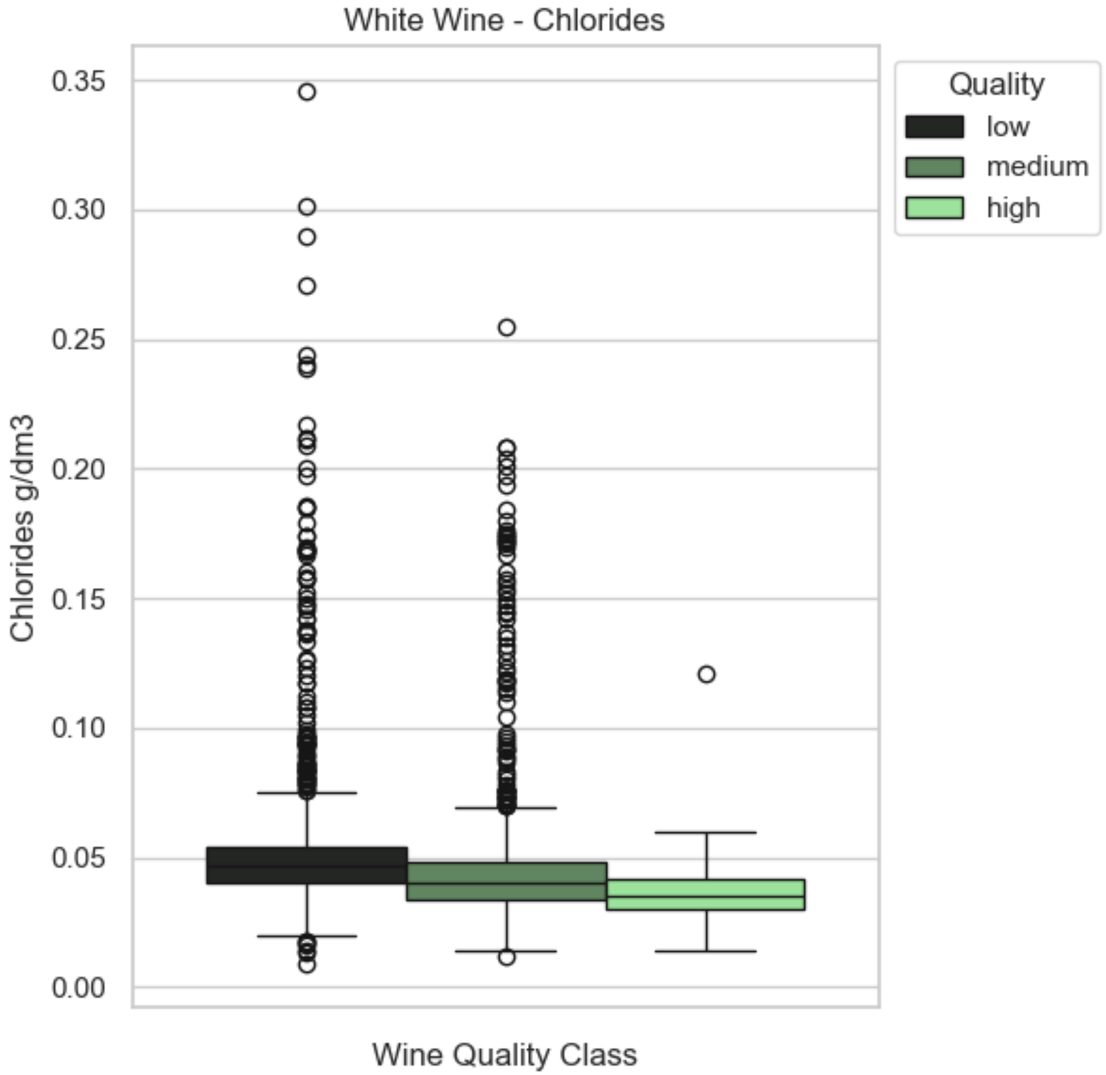

Red Wine - Alcohol
Wine Alcohol %
Wine Quality Class
Quality
low
medium
high
15
14
13
12
11
10
9

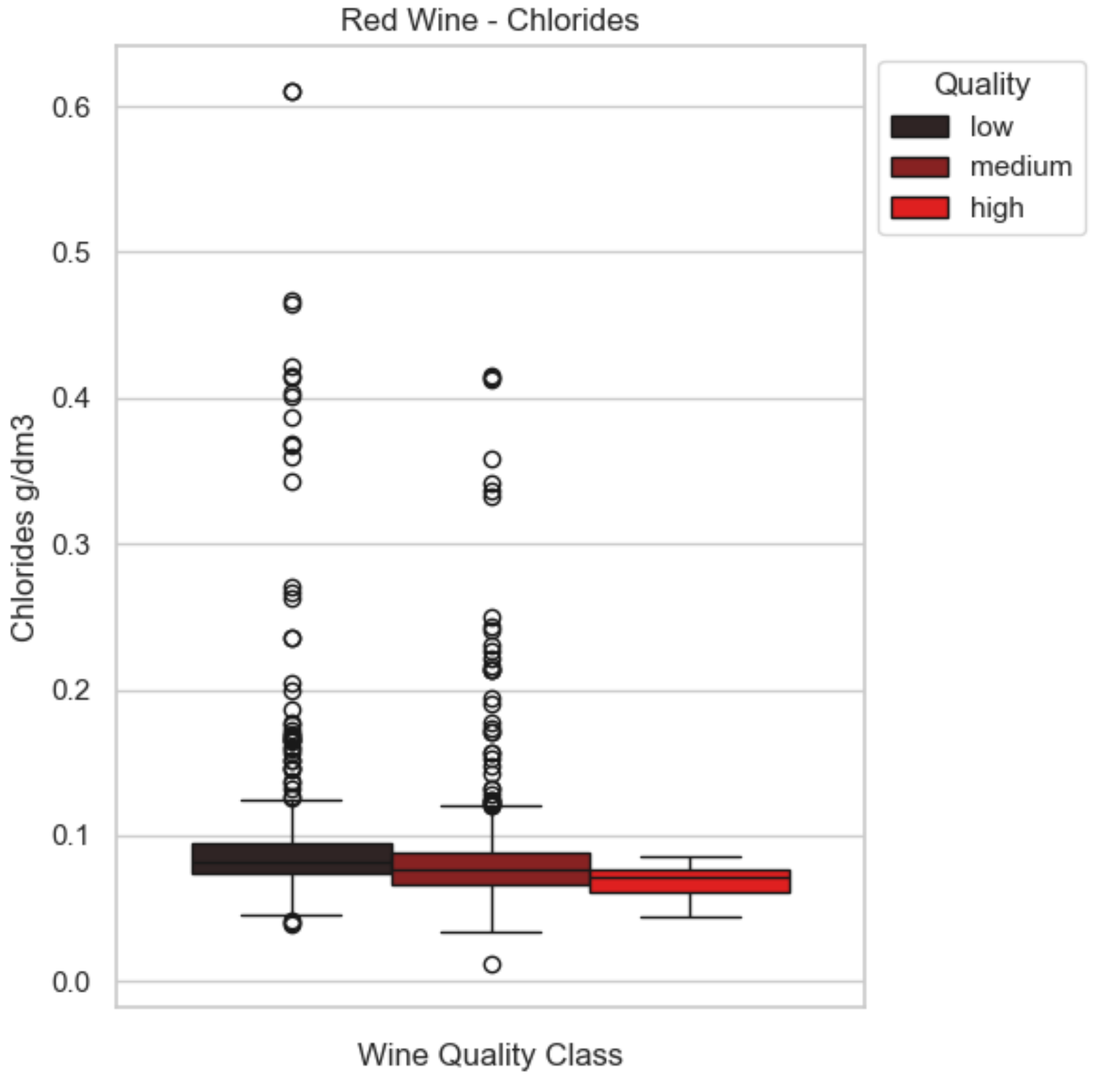
Density


White Wine - Density
1.04
1.03
1.02
1.01
1.00
0.99
Density g/cm3
Quality
low
medium
high
Wine Quality Class

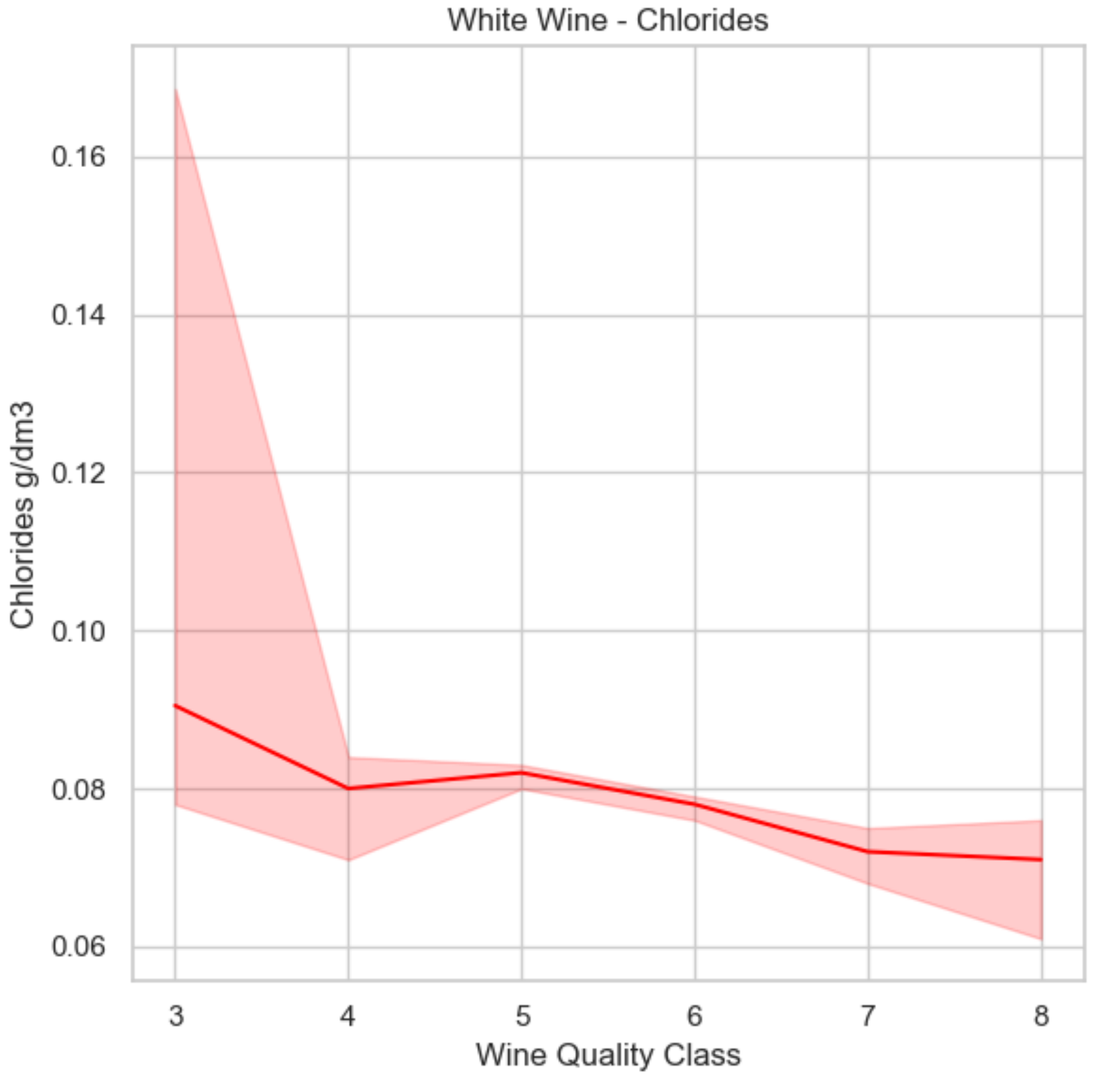

Red Wine - Density
1.004
1.002
1.000
0.998
0.996
0.994
0.992
0.990
Density g/cm3
Quality
low
medium
high
Wine Quality Class

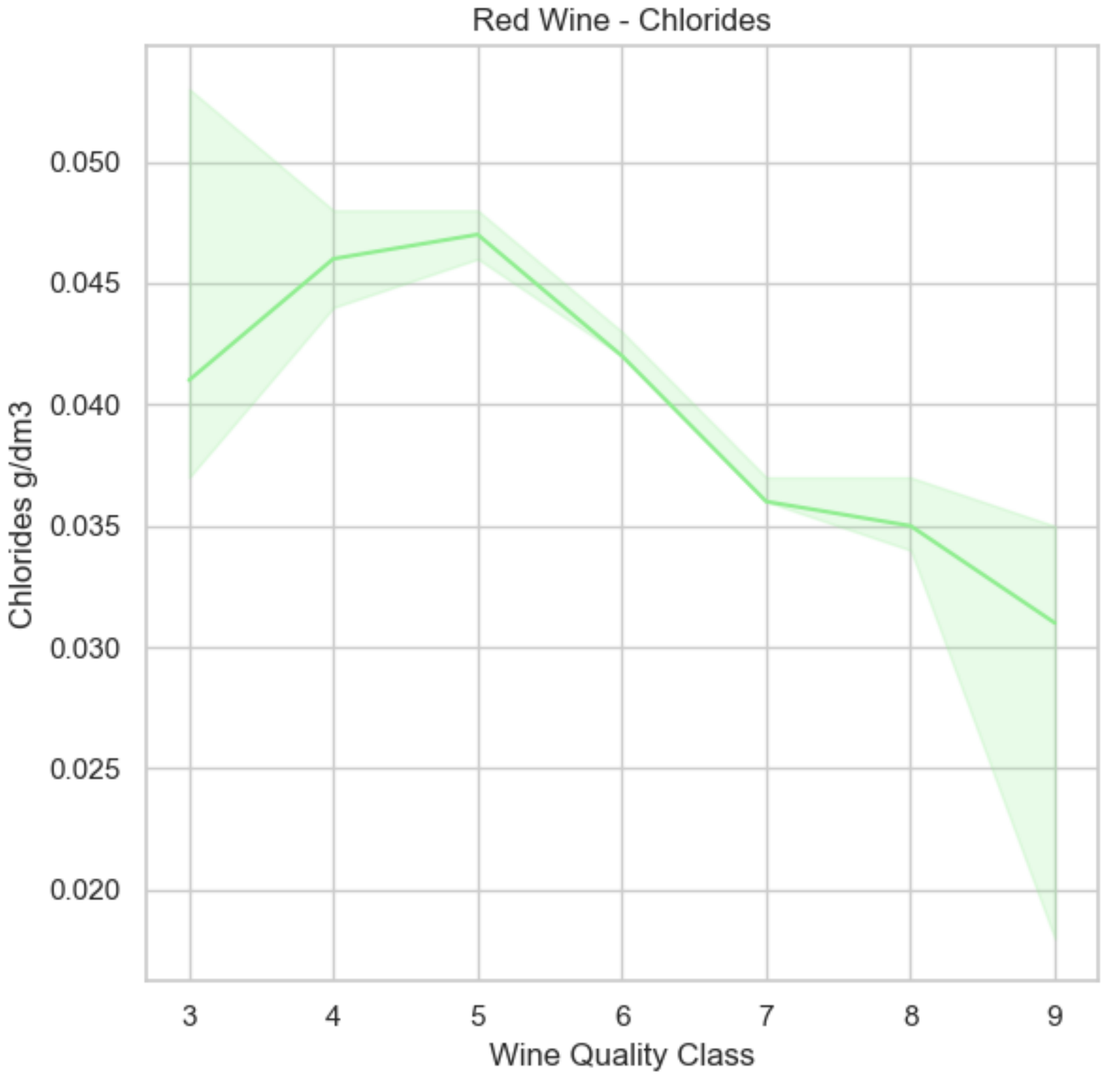
Chlorides


White Wine - Chlorides
Chlorides g/dm3
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Wine Quality Class
Quality
low
medium
high

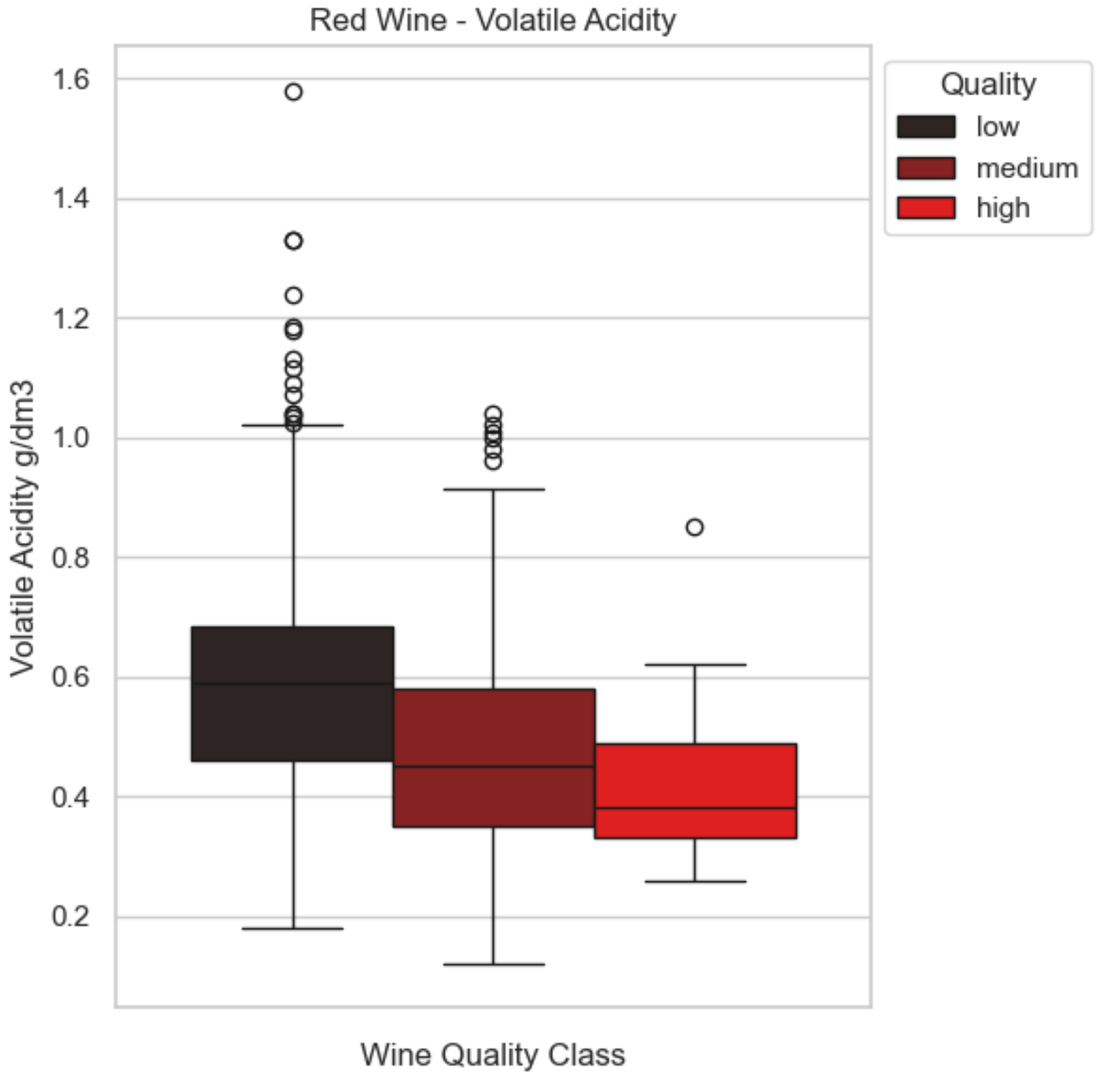

Red Wine - Chlorides
Chlorides g/dm3
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Wine Quality Class
Quality
low
medium
high

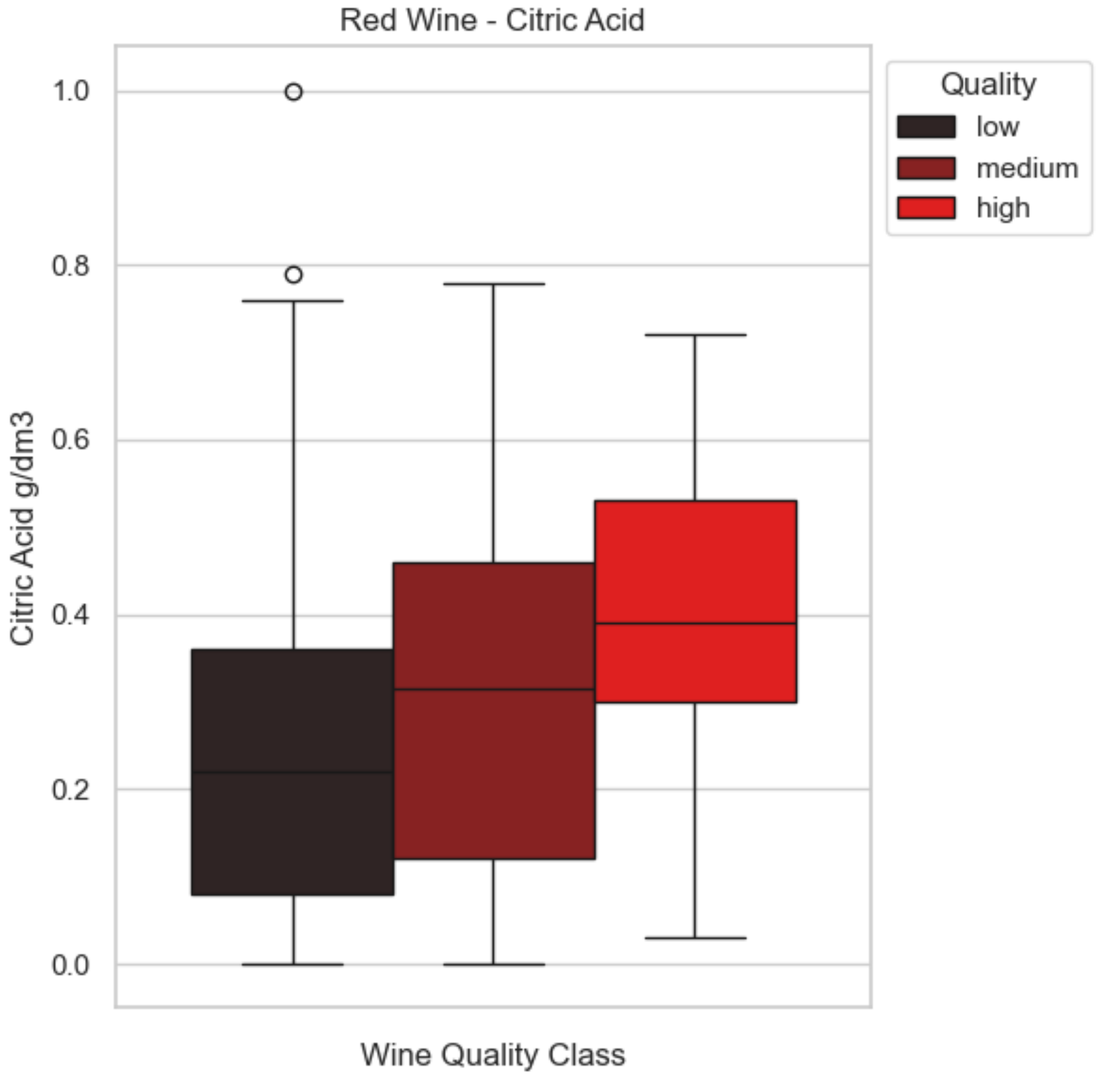
Chlorides


White Wine - Chlorides
Chlorides g/dm3
0.16
0.14
0.12
0.10
0.08
0.06
3
4
5
6
7
8
Wine Quality Class

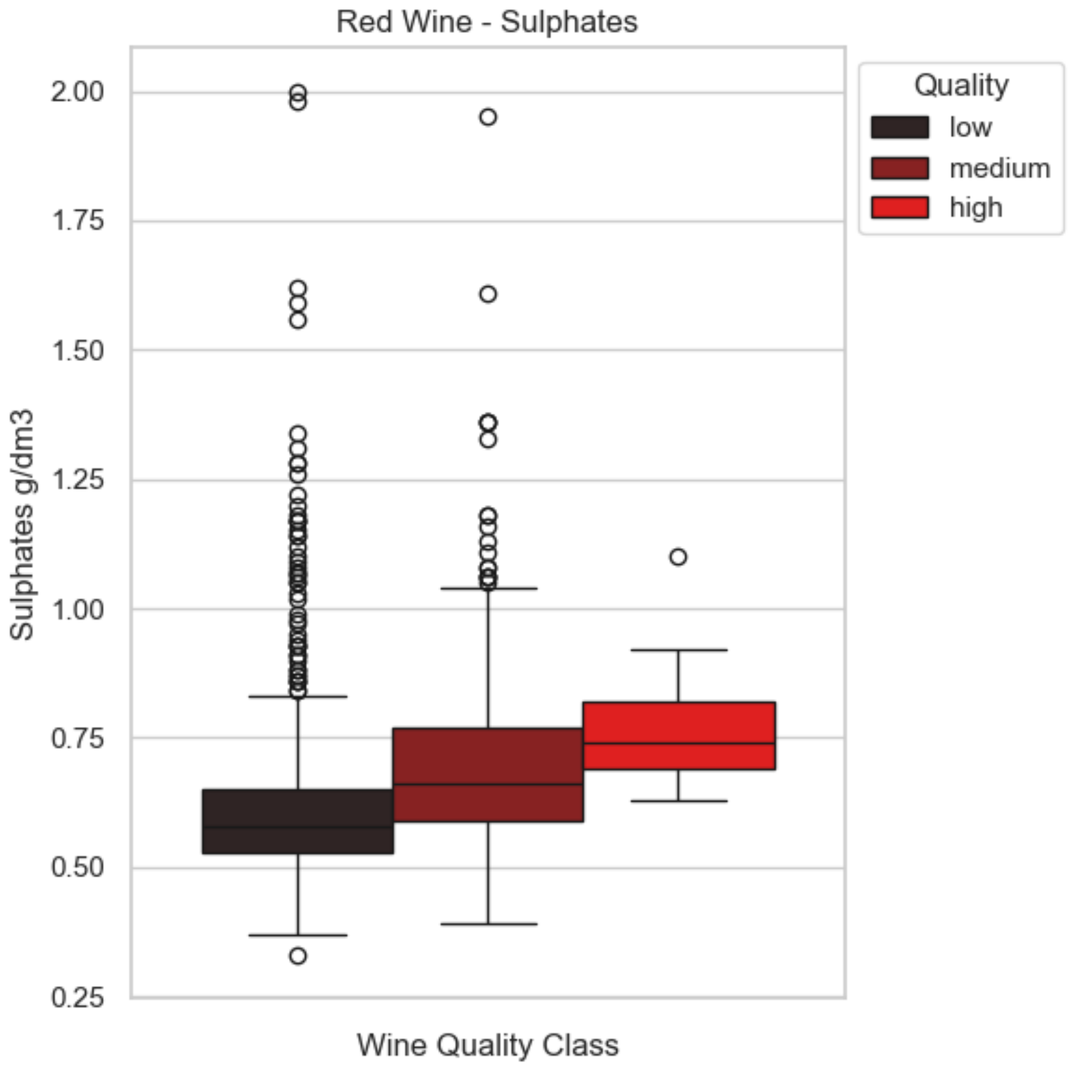

Red Wine - Chlorides
Chlorides g/dm3
0.050
0.045
0.040
0.035
0.030
0.025
0.020
3
4
5
6
7
8
9
Wine Quality Class

Acidity


Red Wine - Volatile Acidity
Volatile Acidity g/dm3
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
Quality
low
medium
high
Wine Quality Class


Red Wine - Citric Acid
Citric Acid g/dm3
1.0
0.8
0.6
0.4
0.2
0.0
Quality
low
medium
high
Wine Quality Class

Sulphates


Red Wine - Sulphates
Quality
low
medium
high
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
Sulphates g/dm3
Wine Quality Class

Findings Overall

- Wine that contains more alcohol tends to be of better quality.
- Good wine has usualy a lower density, which is to a big extend caused by the alcohol.
- There is a lot of outlier data in chlorides, which makes the analysis dificult, but we can still say that avoiding high values (like in the outliers) should make for a better wine.

Findings Red Wine

- It’s good to keep the volatile acidity low.
- An on average high level of citric acid is appreciated in red wine.
- A rather high sulfates level can help increase the quality of red wine, but a too high level might ruin it, as seen based on the outliers.

Research Questions

What chemical composition would a good wine from the region have?

What chemical composition would a good wine from the region have?

White

alcohol:	12.2 %
density:	0.99123 g/cm3
chlorides:	0.035 g/dm3

Red

alcohol:	12.5 %
density:	0.99472 g/cm3
chlorides:	0.071 g/dm3
volatile acidity:	0.38 g/dm3
citric acid:	0.39 g/dm3
sulphates:	0.74 g/dm3