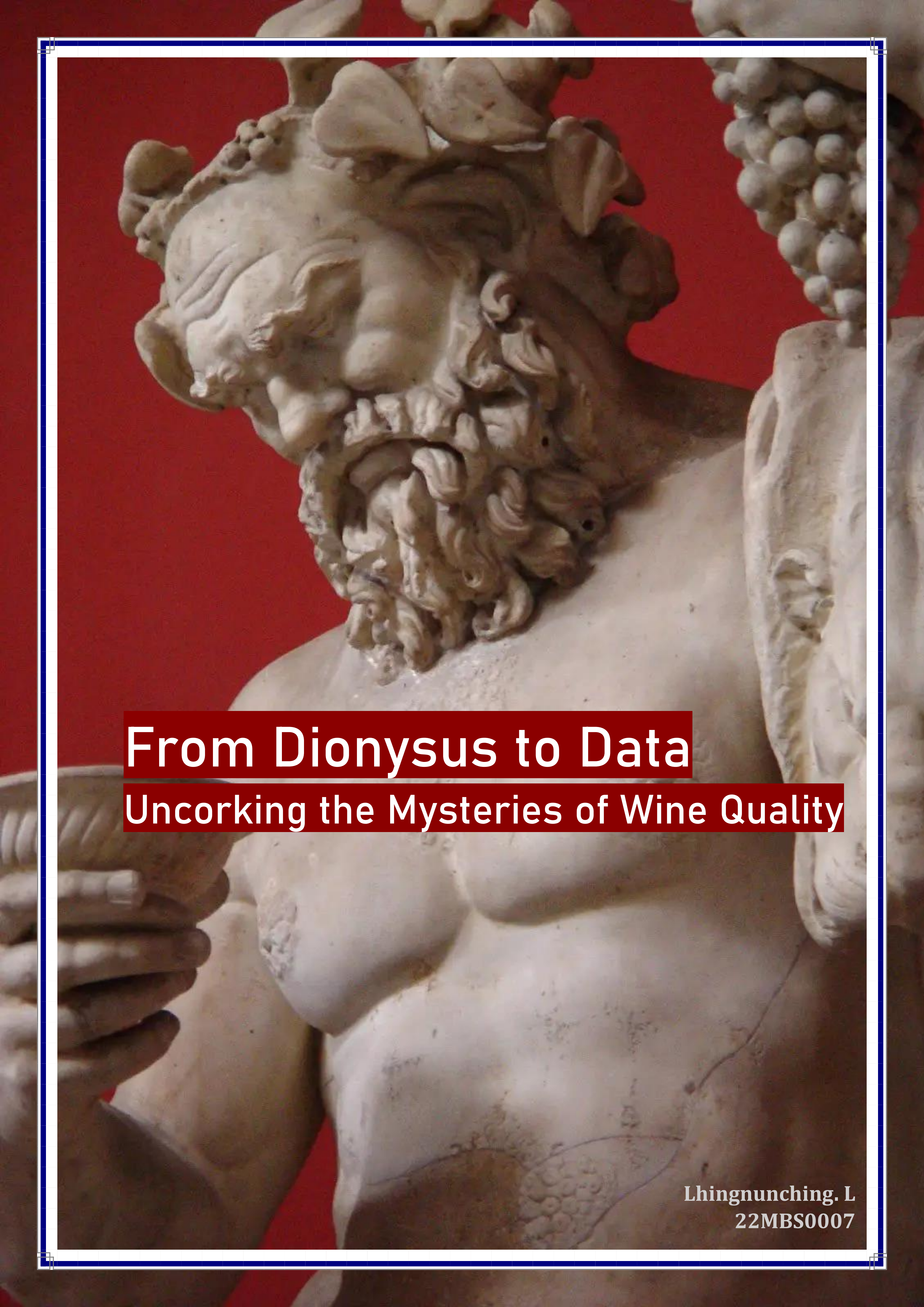


From Dionysus to Data

Uncorking the Mysteries of Wine Quality

Lhingnunching. L
22MBS0007

History of wine

The history of wine is deeply intertwined with human civilization, spanning thousands of years and cultures around the world. Wine's origins can be traced back to ancient times, where it played a significant role in various mythologies and religious practices. In Greek mythology, Dionysus was considered the patron deity of viticulture and winemaking. Legends depict him as the bringer of vine cultivation and the secrets of winemaking to mortals. He possessed a dual character that reflected the duality of wine; on the one hand, he provided joy and divine ecstasy; on the other, he would bring terrible and blinding rage.

Beyond mythology, the historical significance of wine is evident in the development of ancient civilizations. Wine production and trade flourished in regions such as Mesopotamia, Egypt, Greece, and Rome. Wine became an integral part of religious ceremonies, social gatherings, and everyday life in various cultures and was considered a symbol of abundance, wealth, and sophistication. Wine was also used as a currency, a diplomatic gift, and an essential commodity in trade networks.

Over time, wine production techniques evolved, and the art of winemaking spread to different parts of the world through trade and exploration. Each culture and region developed its unique traditions, grape varieties, and winemaking methods, resulting in a rich tapestry of wine diversity. Today, wine continues to be celebrated globally. It has become an emblem of culture, a source of pleasure, and a subject of study and analysis.

"The Triumph of Bacchus", by Diego Velázquez, 1628
(The Romans referred to Dionysus as *Bacchus*)

About this project

The exploration of wine quality through data analysis allows us to uncover patterns, understand consumer preferences, and refine the winemaking process. In this pursuit, the connection to Dionysus serves as a reminder of wine's mythical origins, its enduring allure, and its ability to bring people together in celebration and appreciation of this remarkable beverage.

By analysing this dataset, we can gain insights into the relationship between different chemical attributes and the perceived quality of the wine. It allows for the exploration of patterns and correlations to understand which factors contribute to higher or lower quality ratings.

What makes a good wine?

Several chemical characteristics are associated with a good quality wine, including:

1. Fixed Acidity
Fixed acidity refers to the total concentration of acids in wine, primarily tartaric, malic, and citric acids. The ideal range for fixed acidity is typically between 4 to 8 grams per litre (g/L), although this can vary depending on the wine style.

2. Volatile Acidity
Volatile acidity represents the concentration of acetic acid in wine. In general, a good wine has a low volatile acidity level, typically below 0.6 g/L, as excessive levels can result in undesirable vinegar-like flavours.

3. Citric Acid
Citric acid contributes to the overall acidity and freshness of a wine. The optimal range for citric acid varies depending on the wine style and grape variety but is typically between 0.2 to 1 g/L.

4. Residual Sugar
Residual sugar refers to the amount of sugar remaining in the wine after fermentation. The desired level of residual sugar varies depending on the wine style, ranging from dry wines with very low residual sugar (less than 4 g/L) to sweet wines with higher levels (above 20 g/L).

5. Chlorides
Chlorides in wine contribute to its saltiness and can impact the overall flavour profile. The recommended range for chlorides in wine is generally below 0.2 g/L.

6. Free Sulphur Dioxide

Free sulphur dioxide (SO2) acts as a preservative and antioxidant in wine. The appropriate range for free SO2 is typically between 10 to 40 parts per million (ppm), with the specific amount depending on factors such as the wine's pH and sugar content.

7. Total Sulphur Dioxide

Total sulphur dioxide represents the sum of free and bound SO2 in wine. The total SO2 levels in a good wine are typically around 100 ppm, but they can vary depending on the wine style and winemaking practices.

8. Density

Density measures the mass of the wine per unit volume and can provide an indication of its body and richness. The density of a good wine typically ranges between 0.98 to 1.03 g/cm^3.

9. pH

pH measures the level of acidity or basicity in wine on a logarithmic scale from 0 to 14. A good wine generally has a pH value between 3 and 4, with variations depending on the wine style.

10. Sulphates

Sulphates come from sulphur-containing compounds such as potassium or sodium sulphate. They can contribute to the wine's aroma and act as antioxidants. The ideal range for sulphates is typically around 0.4 to 0.8 g/L.

11. Alcohol

The alcohol content in wine affects its body, flavour, and perceived warmth. The alcohol level in good wines can vary depending on the style, ranging from around 11% to 15% by volume.

The specific optimal ranges for these chemical characteristics can vary depending on the wine style, grape variety, and regional preferences. Winemakers carefully balance these components to achieve the desired style, flavour profile, and overall quality of the wine. Overall, a good wine should have a balance of these chemical components, with no single element overpowering the others.

About the Dataset

The wine quality dataset from the UCI Machine Learning Repository is a collection of chemical and sensory attributes of red and white wines. However, for our purpose we will be using only the red wine quality data. The dataset provides information on various physicochemical properties, such as acidity levels, residual sugar, alcohol content, pH, and more.

The dataset contains 11 input variables, which are used to predict the quality of the wine, or the output variable, which is represented by a rating score ranging from 0 to 10.

Input variables:
1 - fixed acidity
2 - volatile acidity
3 - citric acid
4 - residual sugar
5 - chlorides
6 - free sulphur dioxide
7 - total sulphur dioxide
8 - density
9 - pH
10 - sulphates
11 – alcohol

Output variable (based on sensory data):
12 - quality (score between 0 and 10)

The dataset consists of two separate files for red and white wines, with 1,599 and 4,898 instances, respectively. Each instance represents a unique wine sample. The data is gathered from Portuguese Vinho Verde wines and is based on sensory evaluations performed by experts.

Dataset taken from: https://archive.ics.uci.edu/ml/datasets/Wine+Quality

Visualizing the Wine Quality Dataset and drawing inferences

```
#importing the required libraries

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import plotly.express as px
import seaborn as sns
import warnings

#reading the data
data = pd.read_excel('wq.xlsx')
```

[1] Snippet of the Dataset

```
#Display the data frame
data
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68	9.8	5
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	0.65	9.8	5
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	0.58	9.8	6
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5
...	...	...	...	...	...	...	...	...	...	...	...	...
1594	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	10.5	5
1595	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76	11.2	6
1596	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	11.0	6
1597	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	10.2	5
1598	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	0.66	11.0	6

1599 rows × 12 columns

[2] Summary Statistics of the Dataset

```
#Generating the summary statistics
data.describe()
```

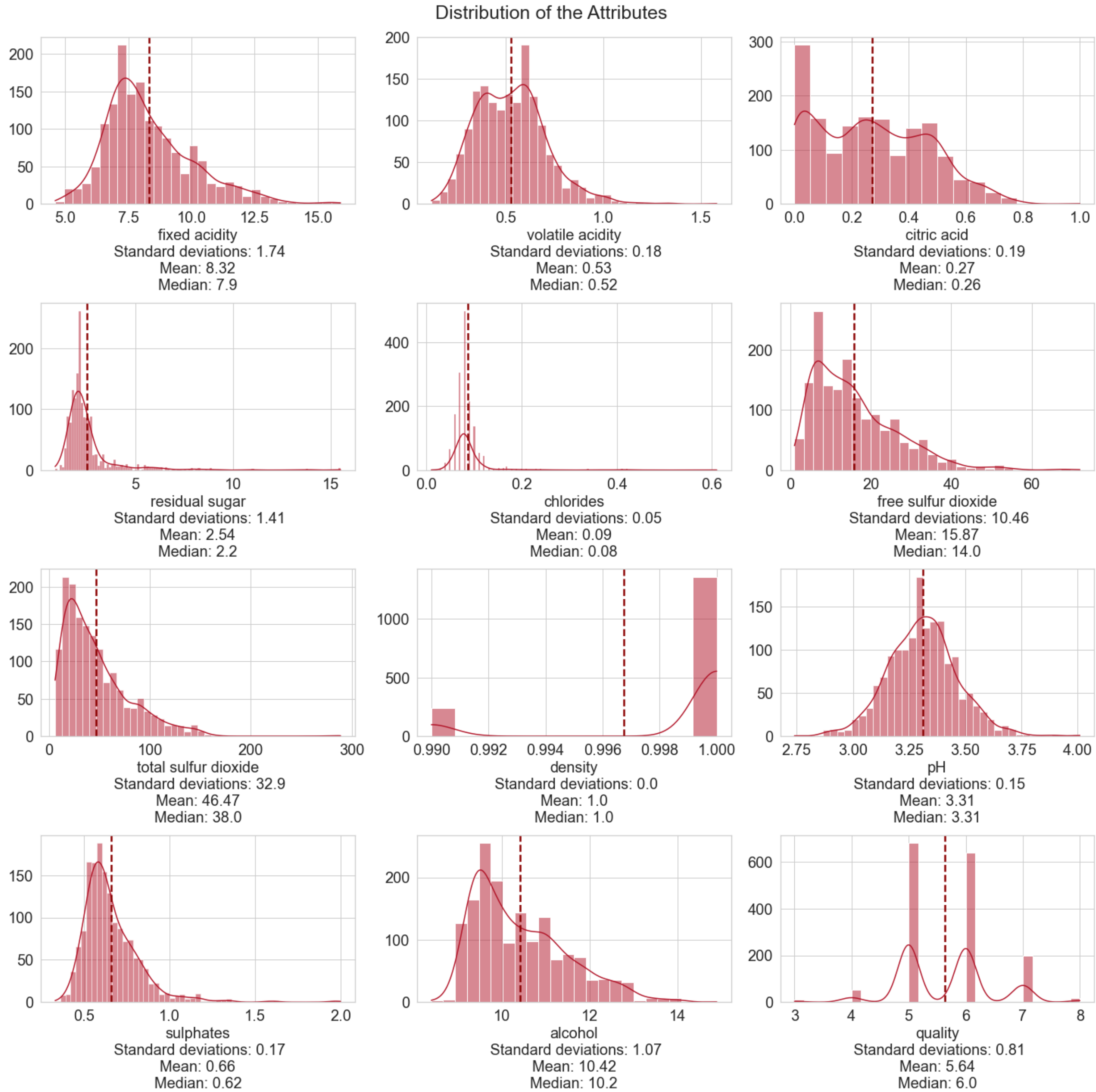
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.467792	0.996747	3.311113	0.658149	10.422983	5.636023
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324	0.001887	0.154386	0.169507	1.065668	0.807569
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	0.990070	2.740000	0.330000	8.400000	3.000000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000	0.995600	3.210000	0.550000	9.500000	5.000000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000	0.996750	3.310000	0.620000	10.200000	6.000000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000	0.997835	3.400000	0.730000	11.100000	6.000000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000	1.003690	4.010000	2.000000	14.900000	8.000000

Inference 1: We can see that the attributes have varying ranges, indicating that the measurement scales are different for each of them.

[3] Distribution of the Attributes

```
fig, ax = plt.subplots(4,3, constrained_layout= True)
fig.set_size_inches(20,20)
ax_list =[]
for n in range(4):
    for m in range(3):
        ax_list.append(ax[n,m])
for i in range(len(list(data.columns[:12]))):
    col = data.columns[i]
    sns.histplot(round(data[col],2),color='#b11226', ax = ax_list[i], kde =True)
    ax_list[i].set_xlabel(col + '\n Standard deviations: ' + str(round(data[col].std(),2)) +
'\nMean: ' + str(round(data[col].mean(),2)) +
                          '\nMedian: ' + str(round(data[col].median(),2)), size =20)
    ax_list[i].set_ylabel('')
    ax_list[i].tick_params(axis ='y', labelsize =20)
    ax_list[i].tick_params(axis ='x', labelsize = 20)
```

```
    ax_list[i].axvline(data[col].mean(), c= '#8B0000', ls ='--', lw=2.5)
    fig.suptitle('Distribution of the Attributes', size =25)
plt.show()
```



Inference 2: pH values in the dataset are approximately normally distributed without any significant skew.

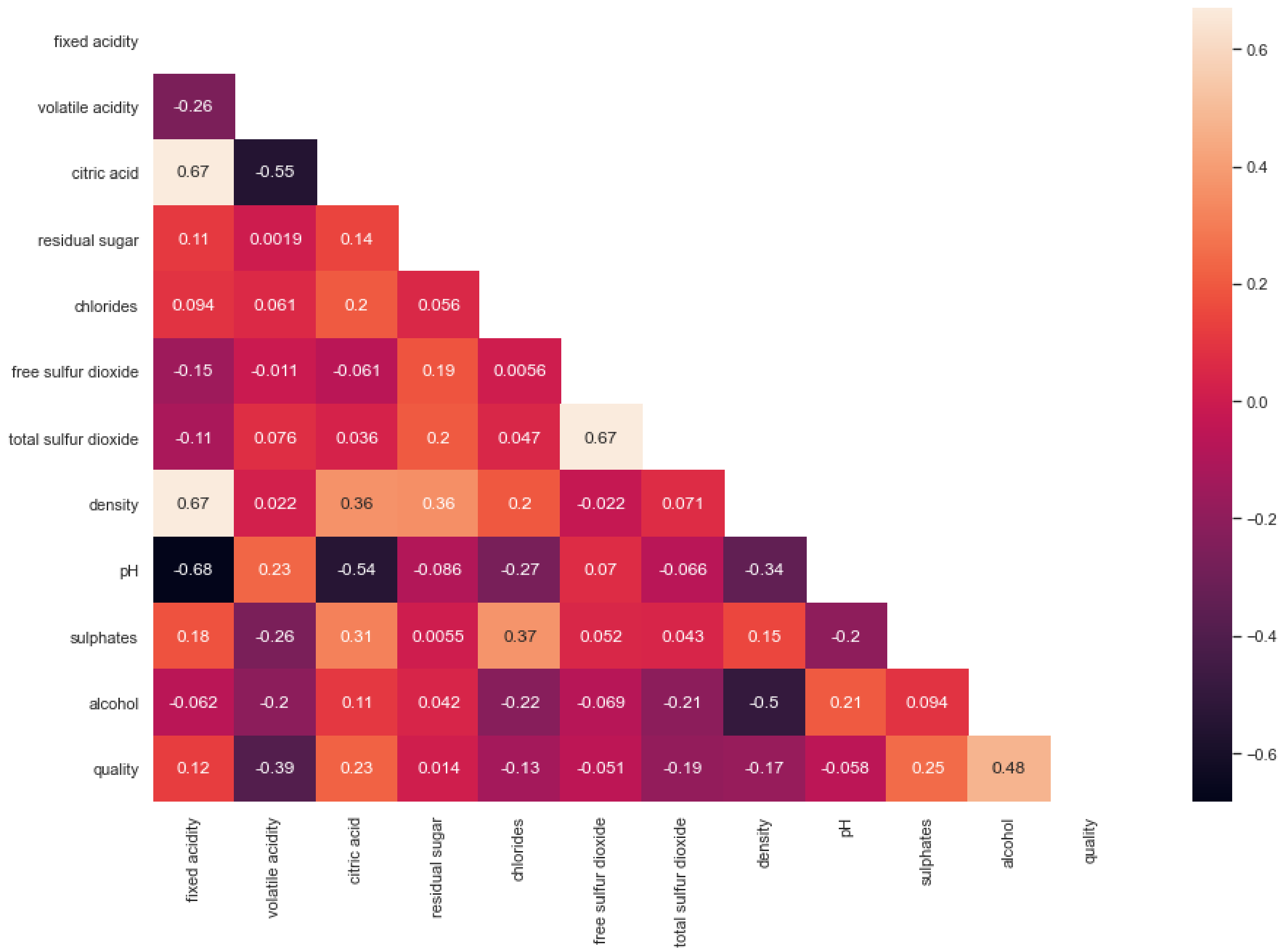
Inference 3: The distribution of alcohol content suggests that most wines in the dataset have alcohol percentages ranging from approximately 8% to 14%.

Inference 4: The distribution of residual sugar shows that most wines in the dataset contain relatively low levels of residual sugar.

[4] Correlation Heatmap between the Attributes

```
heatmap_data = data.corr()
matrix = np.triu(data.corr())

# using the upper triangle matrix as mask
sns.heatmap(heatmap_data, annot=True, mask=matrix)
```



Inference 5: The correlation heatmap indicates some positive and negative correlations between different variables. For example, alcohol content has a positive correlation with wine quality, while volatile acidity has a negative correlation.

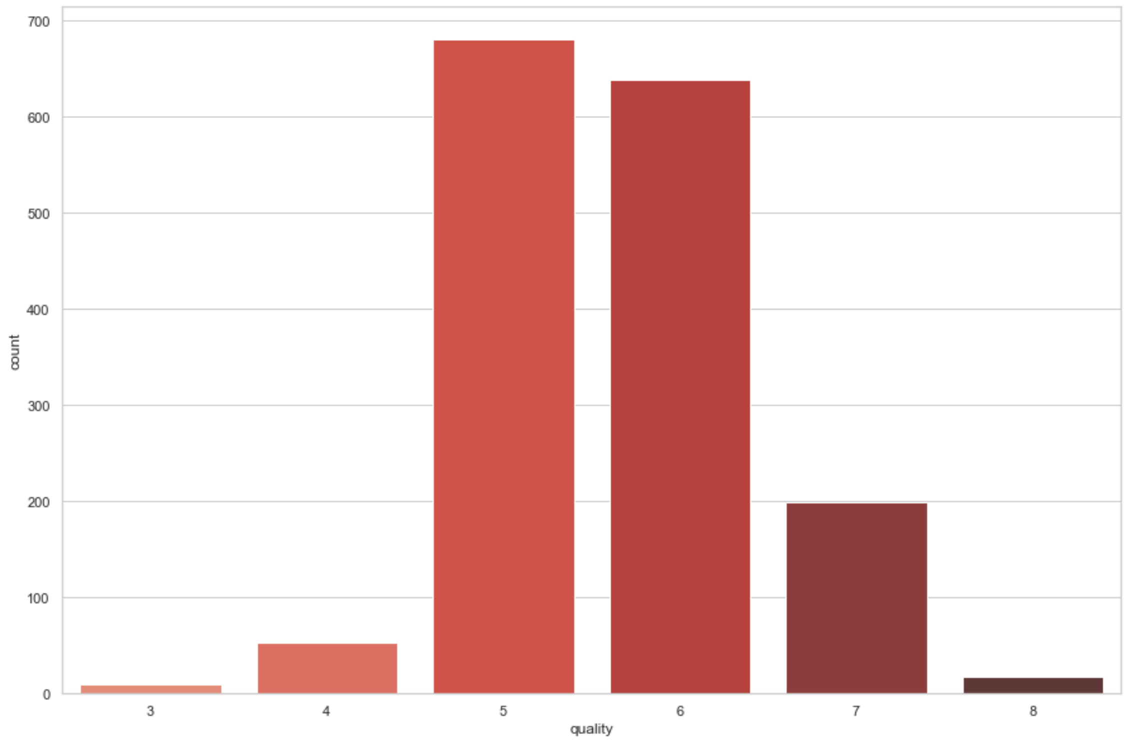
- Positive relationships ($0 < r \leq 1$):
  - Weak positive relationship ($0 < r \leq 0.5$):
    - Citric acid - Density
    - Citric acid - Sulphates
    - Residual sugar - Density
    - Chlorides - Sulphates
    - Sulphates - Quality

    - Alcohol – Quality
  - Strong positive relationship (0.5 < r ≤1):
    - Fixed acidity - Citric acid
    - Fixed acidity - Density
    - Free sulphur dioxide - Total sulphur dioxide

- Negative relationships (-1 ≤ r < 0):
  - Weak negative relationship (-0.5 ≤ r < 0):
    - Fixed acidity - Volatile acidity
    - Volatile acidity - Sulphates
    - Volatile acidity - Quality
    - Chlorides - pH
    - Density - pH
    - Density – Alcohol
  - Strong negative relationship (-1 ≤ r < -0.5):
    - Fixed acidity – pH
    - Volatile acidity - Citric acid
    - Citric acid - pH

[5] Barplot of Ratings for the Quality of the wine

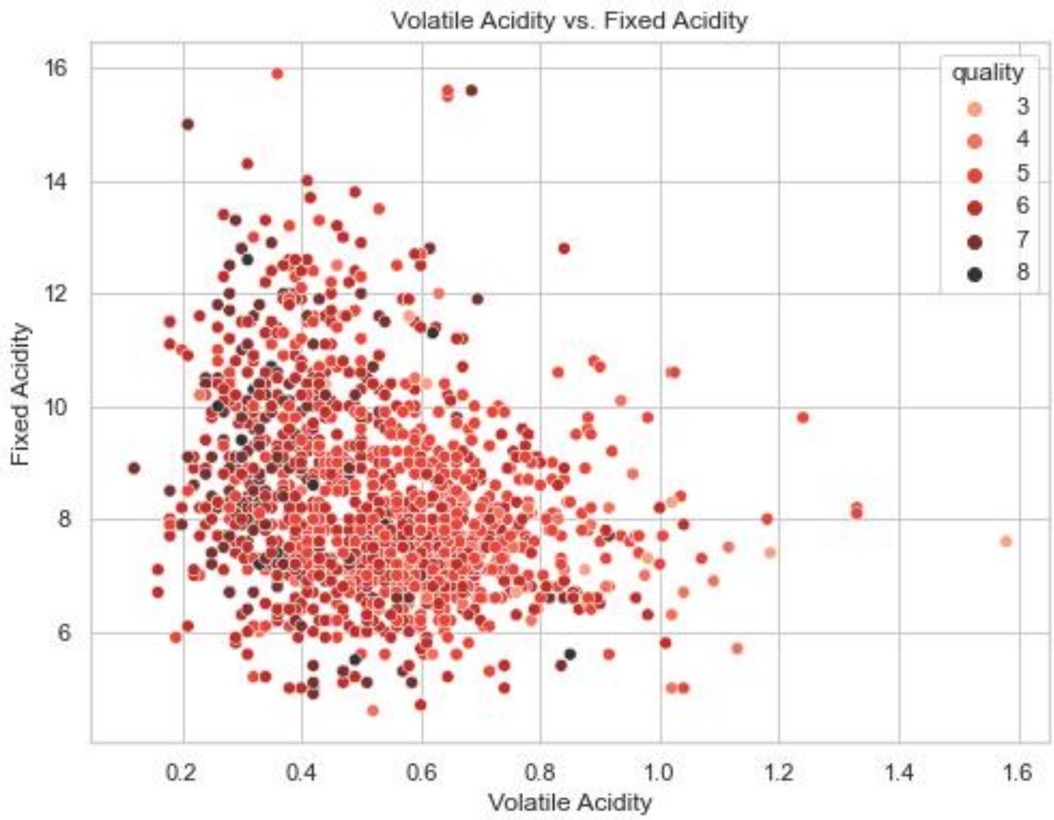
```
sns.countplot(data =data, x="quality", palette = 'Reds_d')
```



Inference 6: The distribution of wine quality shows that the dataset is imbalanced, with a higher number of average quality wines (Rating of 5,6). Wines with a quality rating of 5 and 6 are the most common in the dataset, while wines with ratings of 3 and 9 are the least common.

[6] Relationship of Volatile Acidity and Fixed Acidity with Quality

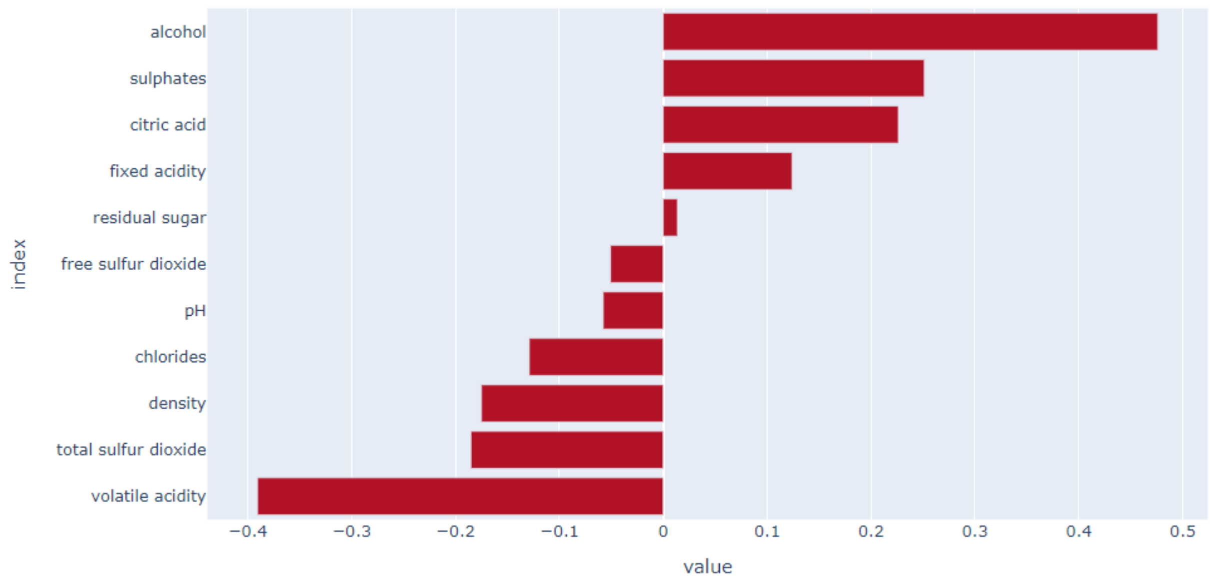
```
plt.figure(figsize=(8, 6))
sns.scatterplot(x='volatile acidity', y='fixed acidity', data = data, hue='quality',
palette = 'Reds_d')
plt.title('Volatile Acidity vs. Fixed Acidity')
plt.xlabel('Volatile Acidity')
plt.ylabel('Fixed Acidity')
plt.show()
```



Inference 7: The scatter plot of Volatile Acidity vs. Fixed Acidity suggests that good quality wines (Rating of 7,8) have higher fixed acidity and tend to have lower volatile acidity levels.

[7] Correlation of the attributes with respect to Quality

```
data_corr_bar = data.corr().quality.sort_values()[:-1]
fig = px.bar(data_corr_bar, orientation="h",
color_discrete_sequence=["#b11226"])
fig.update_layout(showlegend=False)
fig.show()
```



Inference 8: Higher-quality wines tend to have more alcohol percentages as compared to lower-quality wines.

Inference 9: Higher-quality wines tend to have higher sulphate levels as compared to lower-quality wines.

Inference 10: Higher-quality wines tend to have higher citric acid levels as compared to lower-quality wines.

Inference 11: Higher-quality wines tend to have lower volatile acidity as compared to lower-quality wines.

Inference 12: Higher-quality wines tend to have lower total sulphur dioxide levels as compared to lower-quality wines.

Inference 13: Higher-quality wines tend to have lower density levels as compared to lower-quality wines.

Inference 14: Higher-quality wines tend to have lower chlorides levels as compared to lower-quality wines.

References

- Understanding the five basic characteristics of wine: https://www.wineinvestment.com/learn/magazine/2019/05/understanding-the-five-basic-characteristics-of-wine/

- Discover The 5 Basic Wine Characteristics https://winefolly.com/deep-dive/wine-characteristics/

- How to Choose a Good Wine: Understanding the 5 Characteristics of Wine https://winery.ph/blogs/wineryph-blog/five-characteristics-wine

- Modeling wine preferences by data mining from physicochemical properties https://www.sciencedirect.com/science/article/abs/pii/S0167923609001377?via%3Dihub

- Dionysus: Greek God of fertility and wine https://www.greekmythology.com/Other_Gods/Dionysus/dionysus.html

- History of wine https://www.arenaflowers.com/pages/history-of-wine/#:~:text=Carbon%20dating%20confirmed%20the%20wine,domestication%20of%20animals%20and%20plants.

Cheers!