

# REPORT ON WINES

## **Intro:**

Once upon a time, in Spain, there was this amazing, gorgeous, young, and smart empress named Naiz Rusconanz. She ruled over vast territories with a lot of authority, bringing an innate sense of style and sophistication to every corner. To celebrate the discovery of America, which brought loads of richness to folks in the Iberian Peninsula, she wanted to throw a massive party with guests from all over the globe. Because the empress really wanted to show off her people's prosperity and knew all about the trends of the time, she dreamed of creating the biggest and most epic feast ever with the world's best wine.

Back in those days, Spain had some serious trade connections with Portugal. So, it was there that Empress Naiz Rusconanz sent her loyal crew on a quest. The crew, flanked by an old and wise Wine Master by using statistics calculations, set out to figure out what factors influenced wine quality the most. They broke down Portuguese wines into two categories: quality 8 [subqual8] (which represent just 1.13% of the sample analyzed by them) and below [subqualmeno8]. The royal crew have analysed different variables and the outcomes showed that the most influencing elements are: ***sulphates, alcohol, citric acid and pH.***

## ***A deeper look into the dataset...***

## ***DESCRIPTIVE STATISTIC***

The royal crew computed the mean of the different variables according to quality.

Variable	Quality lower than 8	Quality 8
Sulphates	0.66	0.77
Alcohol	10.04	12.1
Fixed Acidity	8.3	8.6

Citric Acid	0.2	0.4
pH	3.31	3.26

## *The quality factors:*

### Sulphates:

*range : 0.33-2*

The royal crew discovered that on average, wine of quality 8 presents a greater quantity of sulphates.

The means:

- **Subqual8:** 0.77
- **Subqualmeno8:** 0.66

Both distributions are **right skewed** where the mean is greater than the median. Anyway, data is presented through two histograms, one for wines with a quality score of 8 and another for wines below that threshold. Notably, wines of inferior quality exhibit a broader range of sulphate levels, likely due to more influencing variables.

The emphasis shifts to density, indicating the frequency of values within a range. Analysing quality 8 wines, a noticeable pattern emerges – higher sulphate content correlates with a greater density, implying these wines occur more frequently. Specifically, in the sulphate range of 0.9 to 1.1, the density is 2, indicating a higher occurrence compared to wines below quality 8 in the same range.

In contrast, the analysis of wines below quality 8 reveals a different trend. Here, a higher density is observed in values with lower sulphate content. For instance, in the sulphate range of 0 to 0.9, the density is 3.5, indicating a higher occurrence for wines of inferior quality with lower sulphate content compared to wines with quality 8 in the same range.

### Alcohol:

*Range: 8.4-14.9*

The royal crew discovered that on average, wine of quality 8 presents a greater level of alcohol.

The means:

- **Subqual8:** 12.1
- **Subqualmeno8:** 10.04

After many days of hard work, the servitudes discovered that wines with a quality rating of 8 generally have more alcohol than lower-quality wines. The distributions of these two types of wines are different: quality 8 wines show a left-skewed distribution, while lower-quality wines have a right-skewed distribution.

This discovery is supported by looking at graphs. When comparing the alcohol levels, quality 8 wines have a higher density, especially in the range of 13 to 14. For instance, the density of better quality wines is 0.5 in this range, while it's close to 0 for other wines. On the flip side, at values like 9 or 10, the density of quality 8 wines is 0.1, whereas for lower-quality wines, it's close to 0.5.

In simple terms, the graphs confirm that better quality wines generally have more alcohol, and the servitudes' hard work paid off in making this discovery.

### **Fixed acidity:**

*Range: 4.6 -15.9*

The royal crew discovered that on average, wine of quality 8 presents slightly a greater level of fixed acidity .

The means:

- ***Subqual8***: 8.6
- ***Subqualmeno8***: 8.3

With regards to fixed acidity we firstly noticed with the mean comparison that there is not a huge difference in wines with quality 8 and others, considering that the values range from 15.9 to 4.6, the servants were disappointed:

Upon reviewing the histograms, it's evident that both distributions seem right skewed, indicating a higher mean compared to the median. The servants also noticed that the range of values in the two samples is quite similar, suggesting that fixed acidity may not directly impact wine quality. Most values cluster in the 4-12 range, and both distributions have comparable density.

This implies that the slightly higher mean fixed acidity in quality 8 wines doesn't stem from a real difference in wine characteristics. Instead, it's likely due to the larger variation and differing sample sizes.

### **Citric acid:**

*Range: 0-1*

The royal crew discovered that on average, wine of quality 8 presents a higher level of citric acid.

The means:

- **subqualmeno8**: 0.2
- **subqual8**: 0.4

Taking into account that the range of Citric acid is 0-1 the servants come to the conclusion that citric acid influence quality since:

Also the histograms accurately drawn by the servants demonstrated that the distribution for wines of quality 8 is highly right skewed if compared to the distribution of wines with quality 8. Looking at the shape of the graphs it can be noticed that the citric acid level in high quality wines is centred mainly around 0.4 and 0.6, instead lower quality wines have a citric acid concentration that is predominantly in the 0-0.3 range. Having said that, the royal crew in selecting the high quality wine will also consider the citric acid level that is predominant in high quality wines.

**pH:**

Range: 2.74 - 4.01

The royal crew discovered that on average, wine of quality 8 presents a lower level of pH.

The means:

- **subqual8**: 3.26
- **subqualmeno8**: 3.31

The royal crew started to analyse the pH distributions of the two different subsets, firstable it noticed that on average there was a greater pH on wine of quality less than 8 rather than in wine quality of quality 8. They immediately started to draw two different graphs which resulted to be both very similar to normal distributions. Even though at the beginning they looked very similar one with the other there are some tiny differences which justify the two different means.

In fact the wine with a pH less than 3 are more frequently in wine at quality 8 (density = 0.5) rather than in wine with quality less than 8 (density = 0.1)

## **REAL ANALYSIS OF DATA**

Even though the royal crew was exhausted they did not stop working. After they have analysed all the types of Portugal's wine, together with the Wine Master they travel all around the world in order to find the best wine in all the world. They found their answers not too far away, in Italy, which at the time and even now is recognised as the best wine producer in the world.

## Sulphates:

Sulphur dioxide is a molecule that serves three functions in wine: it acts as an antioxidant (preventing oxygen from degrading and ruining the product), an anti-oxidative agent (preventing enzymes in the must from oxidising the present molecules), and an antiseptic (hindering the proliferation of microorganisms in the wine).

Its use in wine is crucial because it is the only substance that can halt numerous processes that would lead to wines with terrible colour, taste, and aroma. Additionally, sulphur dioxide prevents the formation of acetic acid by bacteria and the presence of an extremely toxic molecule like *ochratoxin A*, produced by moulds that could develop in the wine.

According to Italian legislation, added sulphates are not permitted in the latter, and the sulphur dioxide levels must be below 15 mg/L. In our dataset, the only red wines that have sulphates exceeding 15 mg/L are those with a quality rating of 5 (3 wines) and 7 (2 wines), respectively. However, concerning wines with a quality rating of 8, all of them meet these specified characteristics.

## Alcohol:

Any wine with an alcohol content below 11% cannot be considered wine. This designation is reserved for specific wines guaranteed by the state, such as *Moscato rosa* and *Brachetto*.

Examining our data on red wines with a quality rating of 8, 22.22% of the sample falls below this threshold. For the 2/18 subset, 11.11% of the wines have an alcohol content exceeding 14%. Unlike the previous category, these wines can only be considered wine if no sugars are added; otherwise, they would be classified as high-alcohol beverages. In the analysis of red wines with a quality rating below 8, 75% have an alcohol content below 11%, while 0.4% exceed 14%.

Additionally, the perception of alcohol is minimised when drinking due to higher acidity, masking the warmth typically associated with alcohol. Alcohol also acts as a preservative, emphasising the need to consume wines with less than 11% alcohol content sooner rather than later.

## Fixed acidity:

Wine is determined by the presence of fixed acids found in wine grapes, with concentrations varying based on grape variety, vineyard soil content, and the type of wine, be it white or red. In the context of red wines, specific thresholds ensure the normal values for highly rated wines, typically ranging between 4.5 mg/L and 9 mg/L.

Our analysis reveals that no wines with a quality rating lower than 8 have a fixed acidity below the minimum allowed threshold of 4.5. Interestingly, when considering red wines within

this range, 72% fall within the specified limits, calculated as  $(1581 - 433 \text{ wines exceeding the max threshold}) / 1581$ .

Conversely, among red wines with a quality rating of 8, only 55% of the analysed wines fall within this optimal acidity range. Moreover, wines within the specified acidity range are deemed more drinkable. Acidity plays a crucial role in the overall taste profile of wine, and when it hovers around 70%, as observed in wines with a quality rating lower than 8, there's a higher likelihood of customer satisfaction and a greater willingness to purchase additional bottles, as the taste remains fresh. Conversely, if soil quality is subpar, acidity decreases, resulting in less satisfied customers who are less likely to make additional purchases.

### Citric acid:

It serves as a common acid supplement in winemaking, particularly during fermentation, aiding winemakers in enhancing the acidity of their wines, especially those derived from grapes grown in warmer climates. Additionally, it functions as a stabiliser to prevent ferric hazes. Notably, 8.25% of the total analysed data in our dataset lacks citric acids, and this percentage is specific to red wines with a quality rating lower than 8. This absence suggests that these wines have naturally matured, requiring no artificial interventions. Consequently, they are likely to possess a more natural taste, rendering them more drinkable.

Conversely, in instances where the climate was excessively hot, wines may lack the desired organoleptic features. This prompts producers to incorporate citric acids to enhance drinkability. To delve further, qualities 6 and 5 exhibit higher percentages of red wines with zero citric acids, at 3.4% (54 out of 1599) and 3.6% (57 out of 1599), respectively.

Furthermore, it's worth noting that within the European Union, citric acid can only be used for stabilisation purposes, and the final citric acid content should not surpass 1g/L. According to our dataset, no wines exceed this mandatory threshold.

### pH:

For wines with a quality rating of 8, 33.3% (6 out of 18) fall within the recommended pH range. This indicates that only a third of red wines with a quality rating of 8 have incorrect acidity levels (lower than the allowed threshold). When such wines are tasted, a perception of bitterness may arise. Only 5.56% (1 out of 18) of these wines exhibit sweetness, surpassing the threshold in a positive direction. However, wines with higher acidity tend to age more quickly, requiring consumption within a shorter time frame.

In contrast, for red wines with a quality rating lower than 8, only 24.2% have incorrect acidity levels, and a mere 3.4% exhibit sweetness. This suggests that a significant portion of red wines with a quality rating lower than 8 falls within the recommended pH range, contributing to a balanced taste profile and allowing for longer shelf life.

## *Skewness and Kurtosis:*

The servants turned out to be expert data analysts and to have a higher degree of accuracy computed also the Skewness and Kurtosis test.

### Sulphates:

From the histogram analysis, it is evident that the distribution exhibits a right-skewed pattern, signifying a concentration of values towards the left side. This observation aligns with the calculated skewness value of 1.331612 and kurtosis value of 4.816021. The positive skewness value corroborates the right-skewed nature seen in the histogram. Additionally, the kurtosis exceeding 3 indicates a distribution with more values in the tails compared to a normal distribution.

To further assess the normality of the dataset, the moments library includes the `jarque.test()` function, conducting a goodness-of-fit test. The null hypothesis posits that the dataset's skewness and kurtosis match those of a normal distribution, while the alternative hypothesis suggests otherwise.

The ensuing code executes this test, yielding a p-value of 0.02031. Given that this value does fall below the significance level ( $\alpha = 0.05$ ), there is sufficient evidence to reject the null hypothesis. Consequently, it can be asserted that the dataset's skewness and kurtosis significantly deviate from those expected in a normal distribution.

### Alcohol:

From the histogram analysis, it is evident that the distribution exhibits a right-skewed pattern, signifying a concentration of values towards the left side. This observation aligns with the calculated skewness value of 0.8600211 and kurtosis value of 3.195654. The positive skewness value corroborates the right-skewed nature seen in the histogram. Additionally, the kurtosis exceeding 3 indicates a distribution with more values in the tails compared to a normal distribution.

To further assess the normality of the dataset, the moments library includes the `jarque.test()` function, conducting a goodness-of-fit test. The null hypothesis posits that the dataset's skewness and kurtosis match those of a normal distribution, while the alternative hypothesis suggests otherwise.

The ensuing code executes this test, yielding a p-value of  $2.2 \times 10^{-16}$ . Given that this value does fall below the significance level ( $\alpha = 0.05$ ), there is sufficient evidence to reject the null hypothesis. Consequently, it can be asserted that the dataset's skewness and kurtosis significantly deviate from those expected in a normal distribution.

### Fixed acidity:

From the histogram analysis, it is evident that the distribution exhibits a right-skewed pattern, signifying a concentration of values towards the right side. This observation aligns with the calculated skewness value of 0.03989292 and kurtosis value of 2.199881. The positive skewness value corroborates the right-skewed nature seen in the histogram. Additionally, the kurtosis not exceeding 3 indicates a distribution with not many values in the tails compared to a normal distribution.

To further assess the normality of the dataset, the moments library includes the `jarque.test()` function, conducting a goodness-of-fit test. The null hypothesis posits that the dataset's skewness and kurtosis match those of a normal distribution, while the alternative hypothesis suggests otherwise.

The ensuing code executes this test, yielding a p-value of 0.7847. Given that this value does not fall below the significance level ( $\alpha = 0.05$ ), there is insufficient evidence to reject the null hypothesis. Consequently, it cannot be asserted that the dataset's skewness and kurtosis significantly deviate from those expected in a normal distribution.

### Citric acid:

From the histogram analysis, it is evident that the distribution exhibits a slightly right-skewed pattern, signifying a concentration of values towards the left side. This observation aligns with the calculated skewness value of 0.1935018 and kurtosis value of 3.800671. The positive skewness value corroborates the right-skewed nature seen in the histogram. Additionally, the kurtosis exceeding 3 indicates a distribution with more values in the tails compared to a normal distribution.

To further assess the normality of the dataset, the moments library includes the `jarque.test()` function, conducting a goodness-of-fit test. The null hypothesis posits that the dataset's skewness and kurtosis match those of a normal distribution, while the alternative hypothesis suggests otherwise.

The ensuing code executes this test, yielding a p-value of  $3.61 \times 10^{-12}$ . Given that this value does fall below the significance level ( $\alpha = 0.05$ ), there is sufficient evidence to reject the null hypothesis. Consequently, it can be asserted that the dataset's skewness and kurtosis significantly deviate from those expected in a normal distribution.



# INFERENCE STATISTICS

## *Confidence interval:*

In order to efficiently detect the high quality wines the servitude computed also the confidence interval of the variables with a 95% confidence level to promptly detect the best quality wines by looking at the mean of their chemical components

The outcomes showed that for quality 8 wines:

- **Sulphates**: level of sulphates is comprised in the 0.64-0.71 interval
- **Alcohol**: we are 95% sure that alcohol in best quality wines will be between 11.5-12.6
- **Fixed acidity**: there is a 95% probability that fixed acidity will be between 7.6 and 9.5
- **Citric acid**: quality 8 wines usually have a citric acid level between 0.3-0.5

## *Hypothesis testing:*

The servants weren't satisfied enough, so in order to be sure to have enough statistical proofs they computed the Hypothesis Testing on sulphates ,fixed acidity and alcohol.

- **Sulphates**: while checking if the mean of sulphates in high quality wines is higher with respect to the others, hypothesis testing confirmed our assumptions, showing that wines with a high degree of quality rely on high sulphates
- **Alcohol**: once again it has been proven that the higher the alcohol level the higher the quality is.
- **Fixed acidity**: Hypothesis testing confirmed the presumptions. Fixed acidity does not directly influence quality.

## *The final consideration:*

After many months, the humble servants managed to find the best wine, which turned out to be from Italy. The royal crew, exhausted and with some mates lost during non-stop work, threw a big party. Everyone got really drunk (just to test the wine before presenting it to the Empress). A week later, the massive feast began and lasted for three months. People from all over the world were impressed by the Empress Naiz Rusconanenz's magnificent rule. However, the Portuguese's careless actions, which ruined the taste of the wine, didn't go unnoticed. After hearing the slaves' report, the empress decided to attack Portugal and wiped

out everyone. The surviving humble servants were then promoted to Lords, each receiving 30 lands.

## Sources

<https://waterhouse.ucdavis.edu/whats-in-wine/fixed-acidity#:~:text=The%20predominant%20fixed%20acids%20found.2%2C000%20mg%2FL%20succinic%20acid.>

[https://en.wikipedia.org/wiki/Acids\\_in\\_wine](https://en.wikipedia.org/wiki/Acids_in_wine)

<https://sensorex.com/ph-improve-taste-color-wine/#:~:text=All%20wines%20are%20acidic.,falling%20between%203.3%20and%203.6.>

[https://www.awri.com.au/industry\\_support/winemaking\\_resources/frequently\\_asked\\_questions/acidity\\_and\\_ph/](https://www.awri.com.au/industry_support/winemaking_resources/frequently_asked_questions/acidity_and_ph/)

<https://www.intravino.com/>