



# How to chose a good wine ?

## Group V - Timothée Darmaillacq - Tom Peypoudat

Upon receiving this assignment, we both immediately gravitated towards exploring the world of wine, rooted in our Bordeaux heritage. Wine, with its eight-millennia history, holds cultural significance, yet navigating its complexities – from regions to grape varieties – poses challenges, especially for those lacking oenological expertise. Our project aims to leverage data science to enhance the wine selection process.

In our research, we found studies on wine consumption, ranging from analyzing statistical outcomes to exploring the correlation between wine composition and quality. While some studies focused on the chemical structure of wine, our interest lies in the human experience. Notably, we discovered connections between textual wine reviews and prices, affirming the reliability of professional reviews.

The human-centric, sensory approach of wine reviews captivates us, aligning with consumer experiences. Our goal is to derive easily applicable insights for selecting good wines, focusing on factual quality differences. Data from reviews by wine experts, with their comprehensive and nuanced perspectives, prove pertinent for our work. We ensured these reviews solely considered wine quality, excluding external factors like packaging or price. Reviewers follow strict protocols, such as blind tastings in quiet rooms, ensuring evaluation integrity.

To commence our work, we gathered data from two sources: 2,000 wine reviews from "Wine Spectator" and 130,000 from "Wine Enthusiast." With this valuable data, our journey began to answer the question, "Which information is relevant when trying to buy good wine?"

## 1. What is the question ?

### Hypothesis :

**Is there a link between the price and the quality of a wine ?**

We have formulated the intriguing hypothesis that a relationship, albeit imperfect, exists between the price and quality of wine. This conjecture is based on the notion that the production of high-quality wine incurs higher costs, from the meticulous selection of grapes to sophisticated vinification processes. Vineyards, in their pursuit of excellence, often invest in sustainable agricultural practices, innovative winemaking techniques, and high-quality aging barrels all contributing elements to an exceptional tasting experience.

However, we are fully aware that the price of a bottle of wine can be influenced by factors beyond the intrinsic quality of the product. Vineyard reputation, vintage rarity, and marketing strategies can also play a significant role in price determination, complicating the correlation between cost and quality.

**Which region and which type of wine is the most liked ?**

From our perspective as French individuals from the Bordeaux region, we expect to find many French wines in these top 100, especially from the Bordeaux region, as they are of high quality and appreciated worldwide. We also anticipate the presence of numerous Italian wines, as Italy is among the major wine producers globally and boasts wines of excellent quality that are enjoyed internationally.

**Witch variety of grapes is most liked ?**

This question is intricate due to the plethora of grape varieties available. In France, Cabernet Sauvignon and Pinot Noir stand out as the most renowned grape varieties, not only locally but also potentially on a global scale. Therefore, we might assume that one of these two varieties is the most liked. However, it's crucial to note that individual preferences vary, and lesser-known varieties such as Syrah or Chardonnay might also be highly appreciated. Furthermore, the region of production and winemaking methods significantly influence taste. In summary, while Cabernet Sauvignon and Pinot Noir are widely recognized in France, preferences may vary based on wine regions and individual tastes, both nationally and internationally.

## 2. How we proceeded :

We ensured to obtain our two databases in Excel format for easier data manipulation. Then, using Python, we were able to perform the following analyses:

- Use of the `corr()` function to obtain the correlation between multiple variables, such as the price and the rating of wines, or the price and the age of a bottle.
- `mean()` to obtain the average price based on a certain rating. In our case, the average price for wines rated 98/100 or higher.

With this, we obtained everything we wanted related to the price of wines.

Then, thanks to the 130,000 rated bottles, we were able to extract the most appreciated regions, countries, grape varieties, and wineries. With all this information, we can pinpoint with great accuracy the most favored province, country, grape variety, and winery in the world of wine.

Our result :		
	Best variety :	<b>Pinot Noir</b>
	Best terroir :	
	◦ province :	<b>Tuscany</b> (Italy)
	◦ country :	<b>France</b>
	Best winery :	<b>Williams Selyem</b> (USA)

## 3. Is price a good indicator ?

We found a strong correlation between the price and the quality of a wine. However, the average price within the top 100 of each year is 48 euros. This means that the correlation has a limit in terms of price; in our case, out of 2100 wines, only 19 are priced above 200 euros.

## 4. Conclusion :

You may not satisfy every criterion given in our results. For example, Pinot Noir is not grown in Tuscany, and you won't find French wine produced by the Williams Selyem winery. However, this provides you with various indications to consider when making your choice, keeping in mind that price can be helpful, but spending large sums is not always necessary; exceptional wines can sometimes be found at 30 euros.