

wine

February 2024

EXPLORATORY DATA ANALYSIS

Presented by: Francesca Roefaro, Cesar Rojas, Jessica Hernandez

sip-by
-sip



The Story Behind Wine Ratings, Prices, and Locations.



BLIND TASTE TEST

130k participants | 512 variety

sip-by
-sip



WineEnthusiast.com

About The Scores

*Products deemed unacceptable (receiving a rating below 80 points) are not reviewed.

| | |
|--------|--|
| 98-100 | Classic The Pinnacle of quality |
| 94-97 | Superb A great achievement |
| 90-93 | Excellent Highly recommended |
| 87-89 | Very Good Often good value; well recommended |
| 83-86 | Good Suitable for everyday consumption, often good value |
| 80-82 | Acceptable Can be employed |

O1. HOW DOES PRICE AND RATINGS CORRELATE?

H0: Prices have no relation to ratings.

H1: If the wine is higher priced then the rating will be higher.

O2. DOES VARIETY AFFECT SCORE?

H0: Variety and score have no relation.

H1: We think that red wines are higher rated than white wines.

O3. ARE HIGHER RATED WINES MORE COMMON IN A SPECIFIC PLACE?

H0: Ratings and origin have no relation.

H1: If the wine is higher rated, it comes from a place that have similar weather conditions

01.

HOW DOES PRICE AND RATINGS CORRELATE?

Normality tests and graphs of data set

O1.

H0: Prices have no relation to ratings.

H1: If the wine is higher priced then the rating will be higher.

Alpha: 0.05

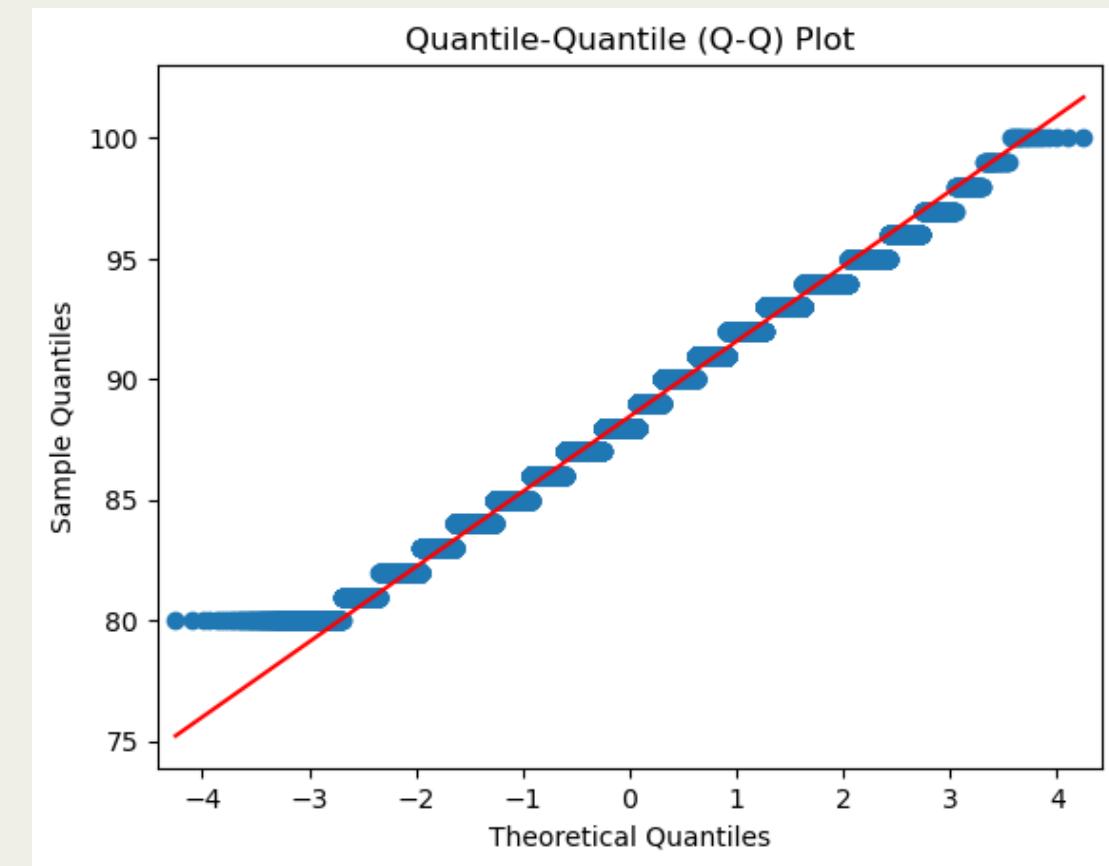
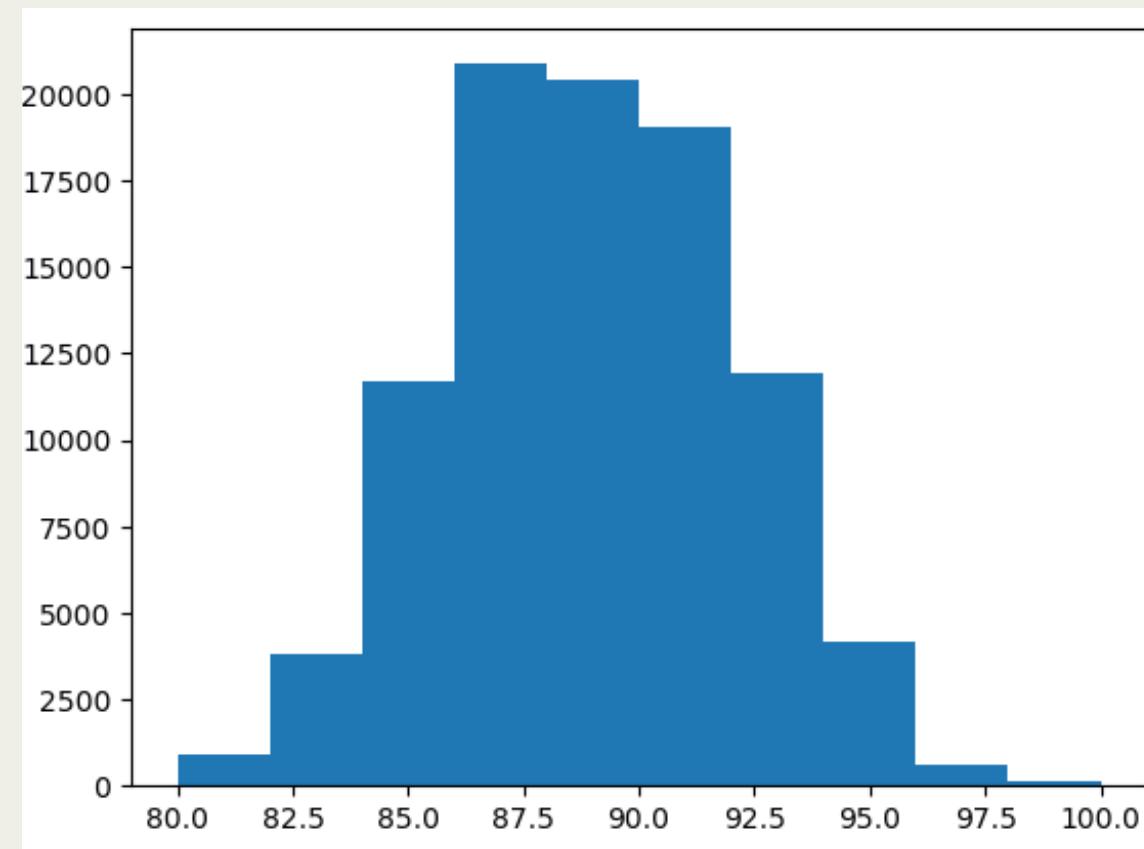
PLOTTING UPPER POINTS NORMALITY

Shapiro-Wilk test

ShapiroResult(statistic=0.9893147945404053,
pvalue=0.0)

Kolmogorov-Smirnov Test

(statistic=1.0, pvalue=0.0, statistic_location=80.0,
statistic_sign=-1)



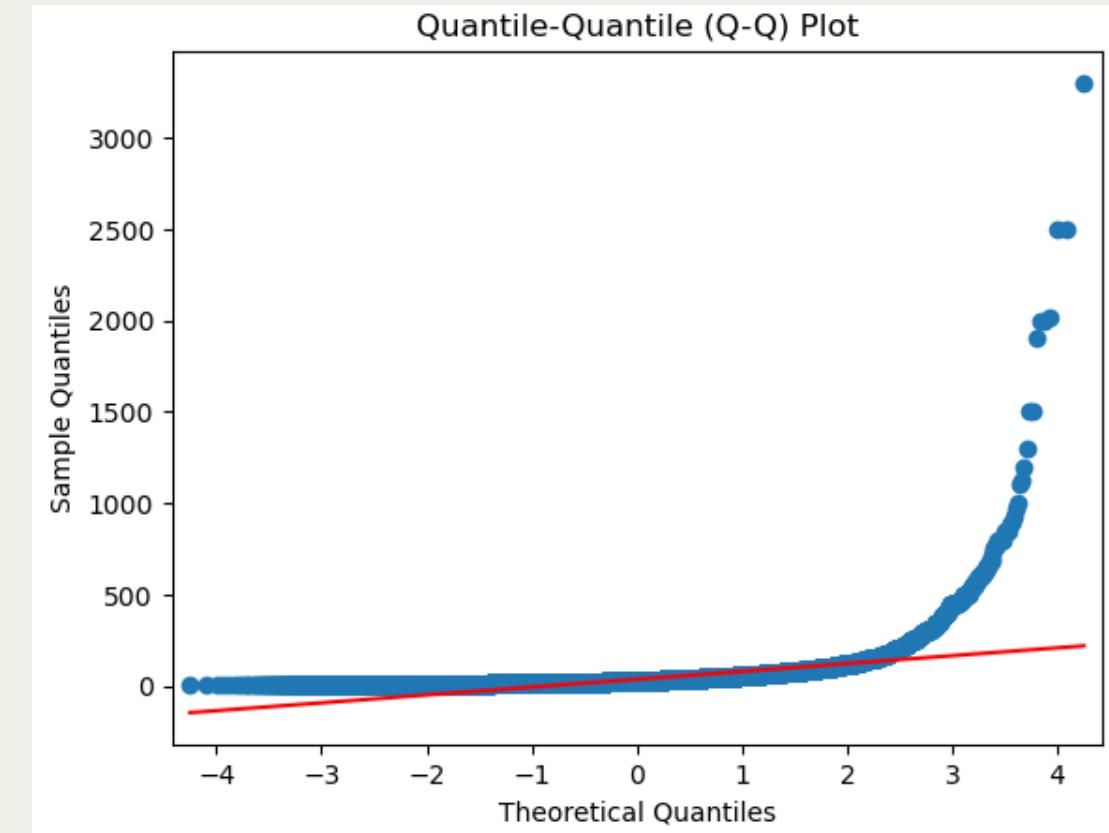
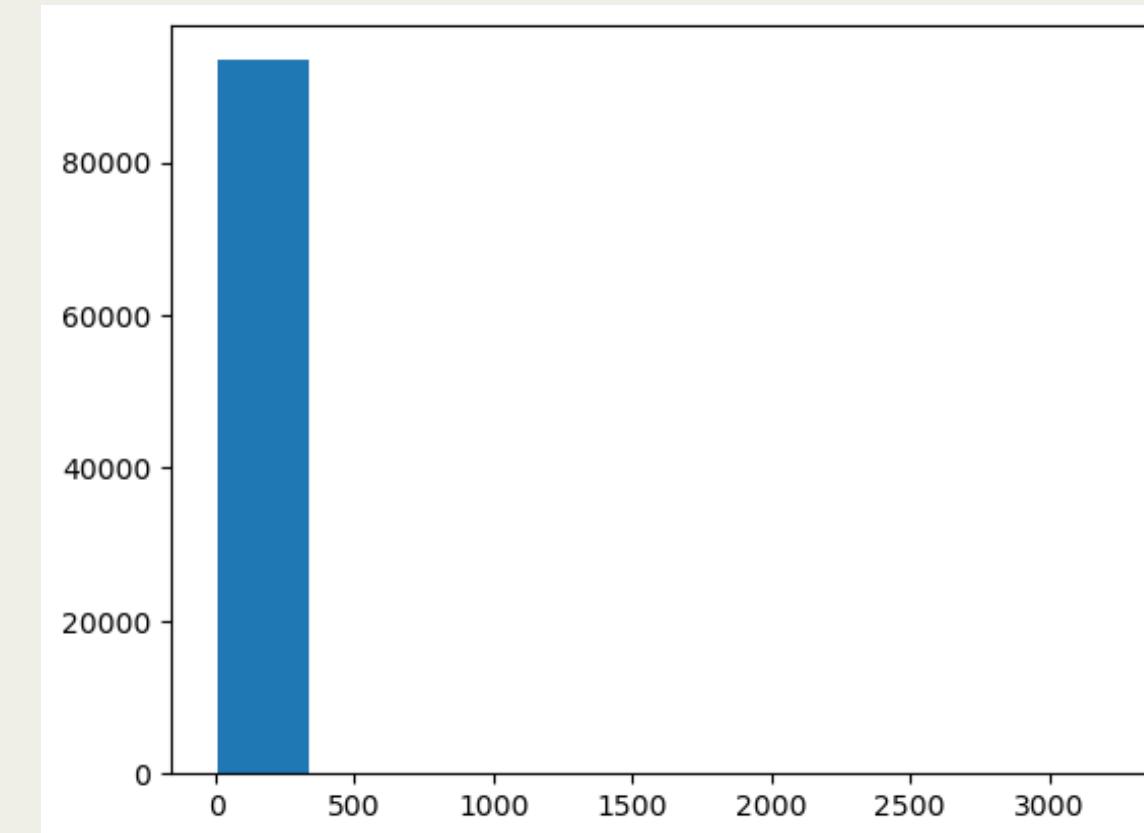
NORMALITY TEST FOR THE PRICE

Shapiro Wilk Test

ShapiroResult(statistic=0.429475128
6506653, pvalue=0.0)

Kolmogorov-Smirnov Test

KstestResult(statistic=0.9999683287581669,
pvalue=0.0, statistic_location=4.0, statistic_sign=-1)



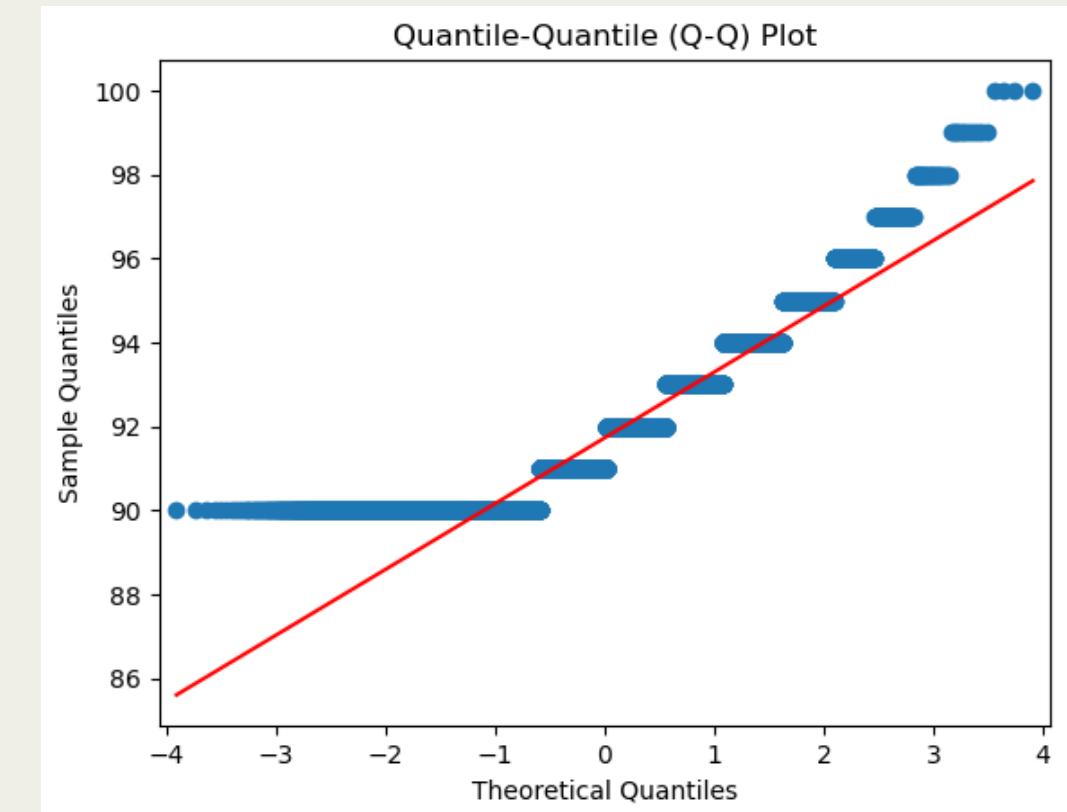
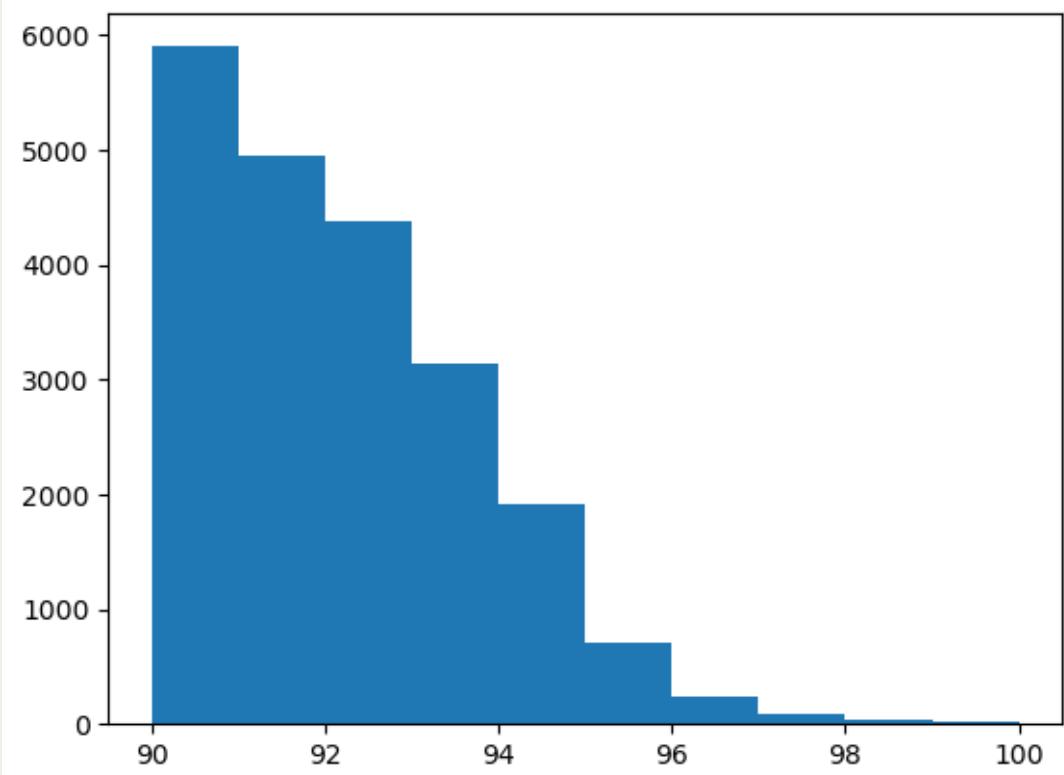
PLOTTING UPPER POINTS NORMALITY

Shapiro Wilk Test

ShapiroResult(statistic=0.8893659710884094,
pvalue=0.0)

Kolmogorov-Smirnov Test

KstestResult(statistic=1.0, pvalue=0.0,
statistic_location=90.0, statistic_sign=-1)



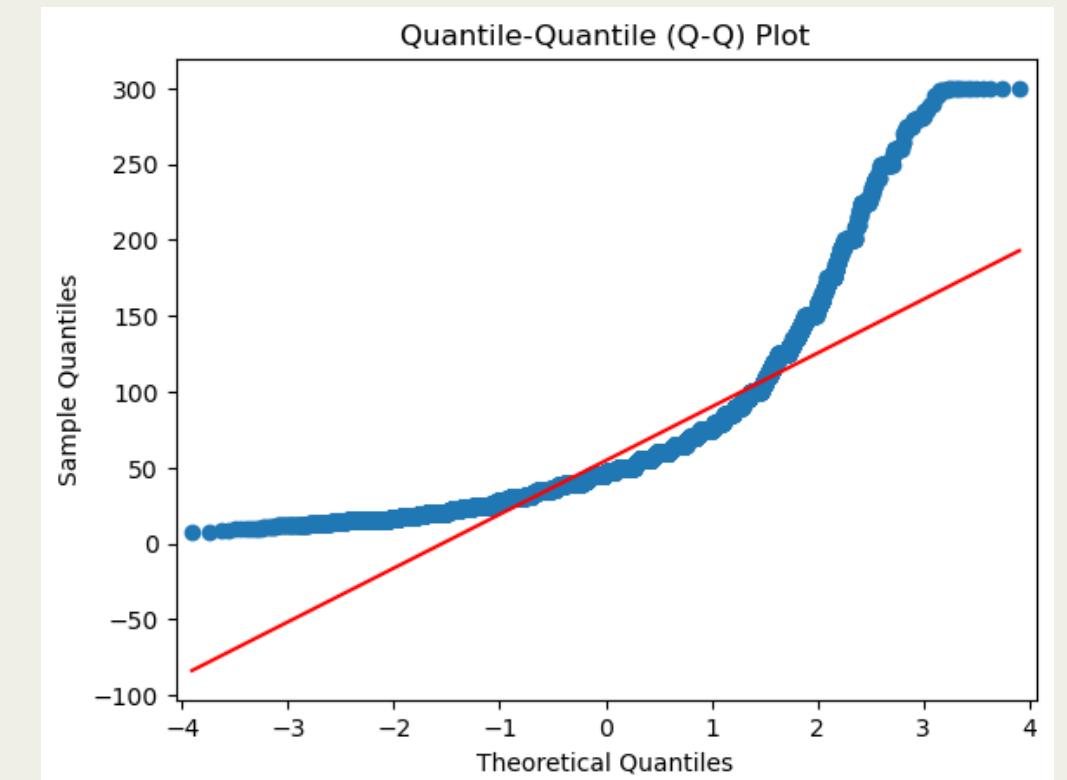
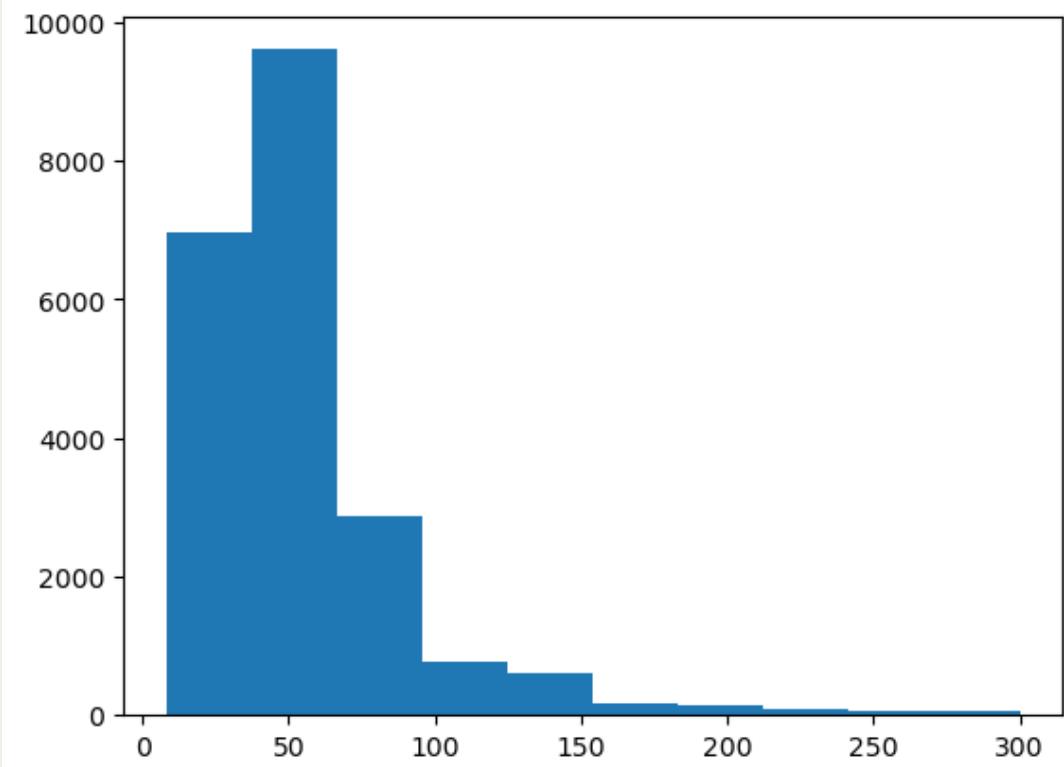
NORMALITY TEST FOR THE PRICE

Shapiro Wilk Test

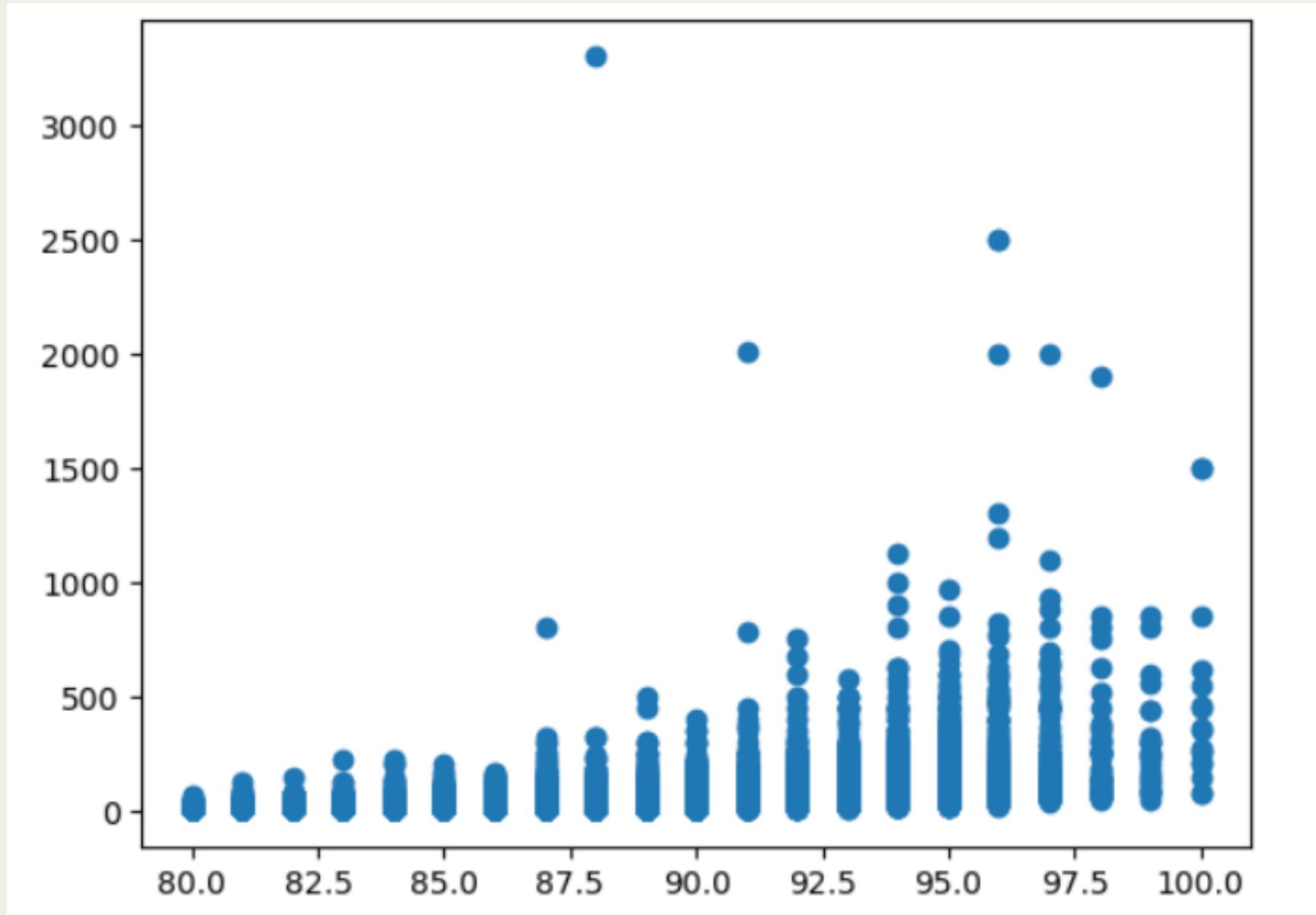
ShapiroResult(statistic=0.7734923958778381,
pvalue=0.0)

Kolmogorov-Smirnov Test

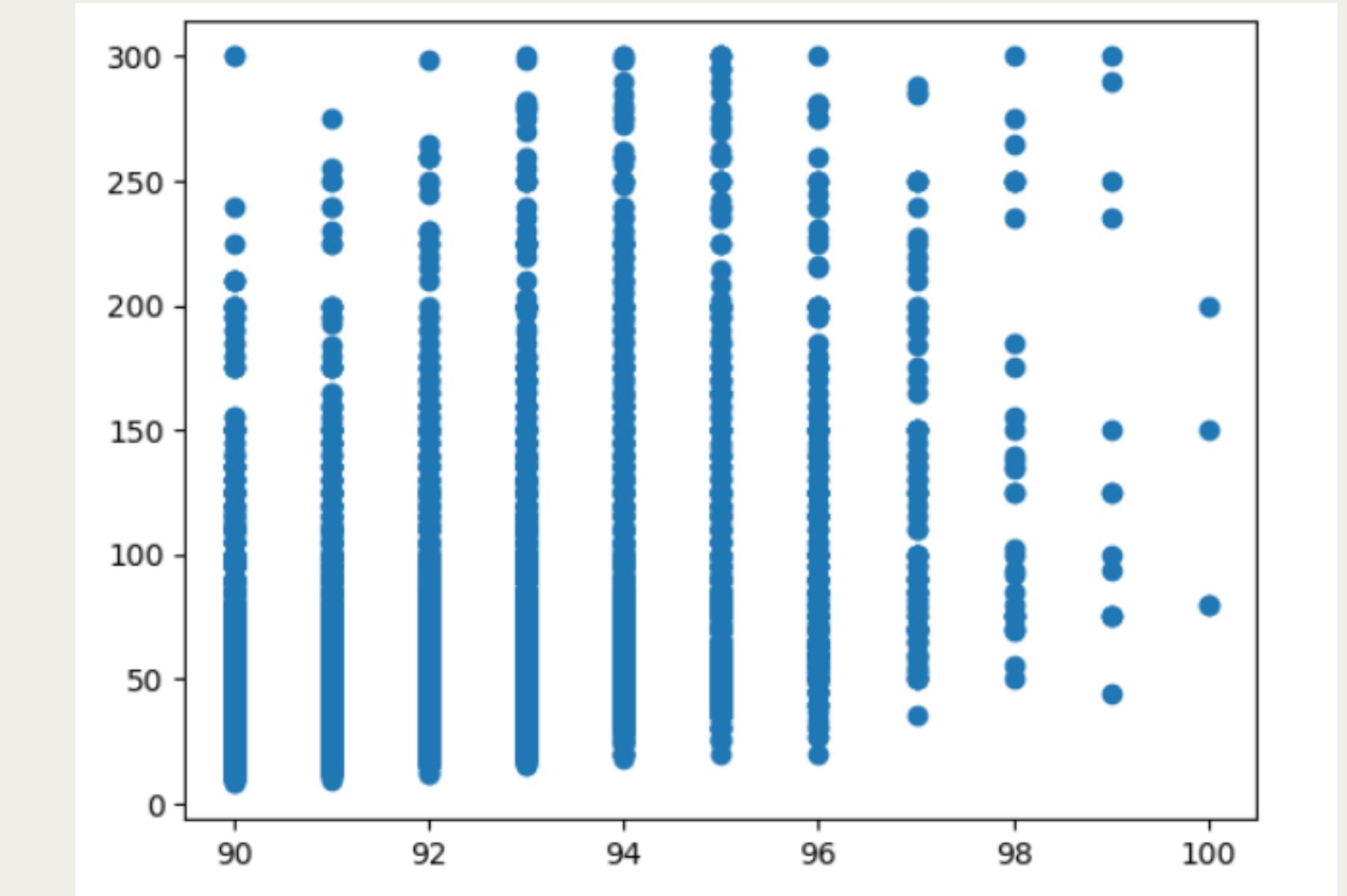
KstestResult(statistic=0.9999999999999993,
pvalue=0.0, statistic_location=8.0, statistic_sign=-1)



SPEARMAN'S R RANK CORRELATION



SignificanceResult(statistic=0.6091152033991962,
pvalue=0.0)



SignificanceResult(statistic=0.4487728524484996,
pvalue=0.0)



FINAL FINDINGS

- The original dataset was not normal
- Spearman's rank correlation needed to be used to find the correlation
- There's a strong positive monotonic relationship between the variables between prices and ratings
- Further testing, we need to normalize the data and use Pearson to find the linear relation and regression between price and rating

H1: If the wine is higher priced then the rating will be higher.

02.

DOES VARIETY AFFECT SCORE?

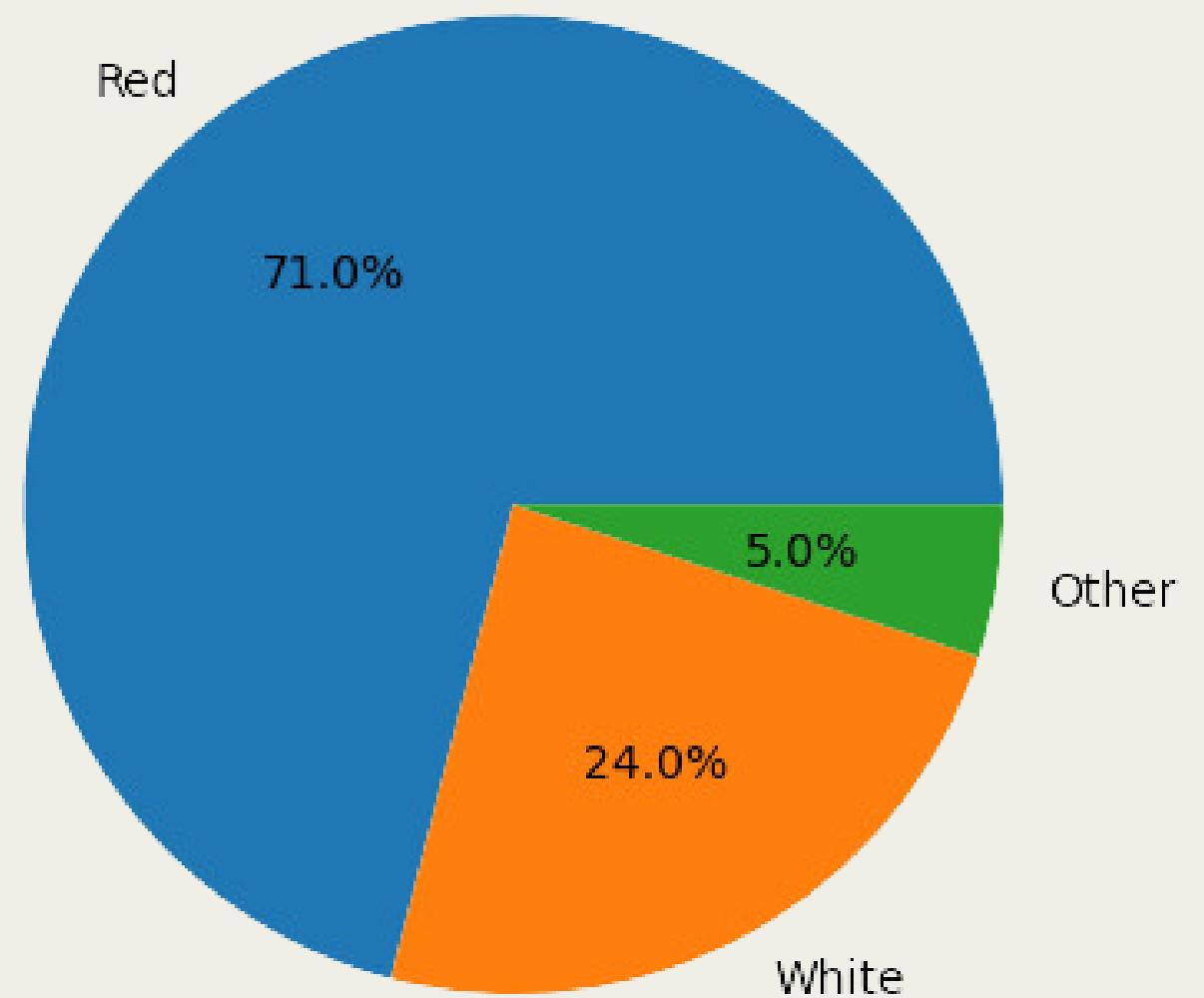
Analyze wine varieties



CREATING A NEW DF

```
COLOR_CW_DF = CLEAN_WINE_DF.COPY()
```

Wine color percentage distribution among
the 9 most reviewed wine variety



GROUPBY WINE COLOR

```
Wine Color
Other      2557
Red        36141
White     12223
Name: Wine Color, dtype: int64
```

02.

H0: The mean scores of red and white wines are equal.

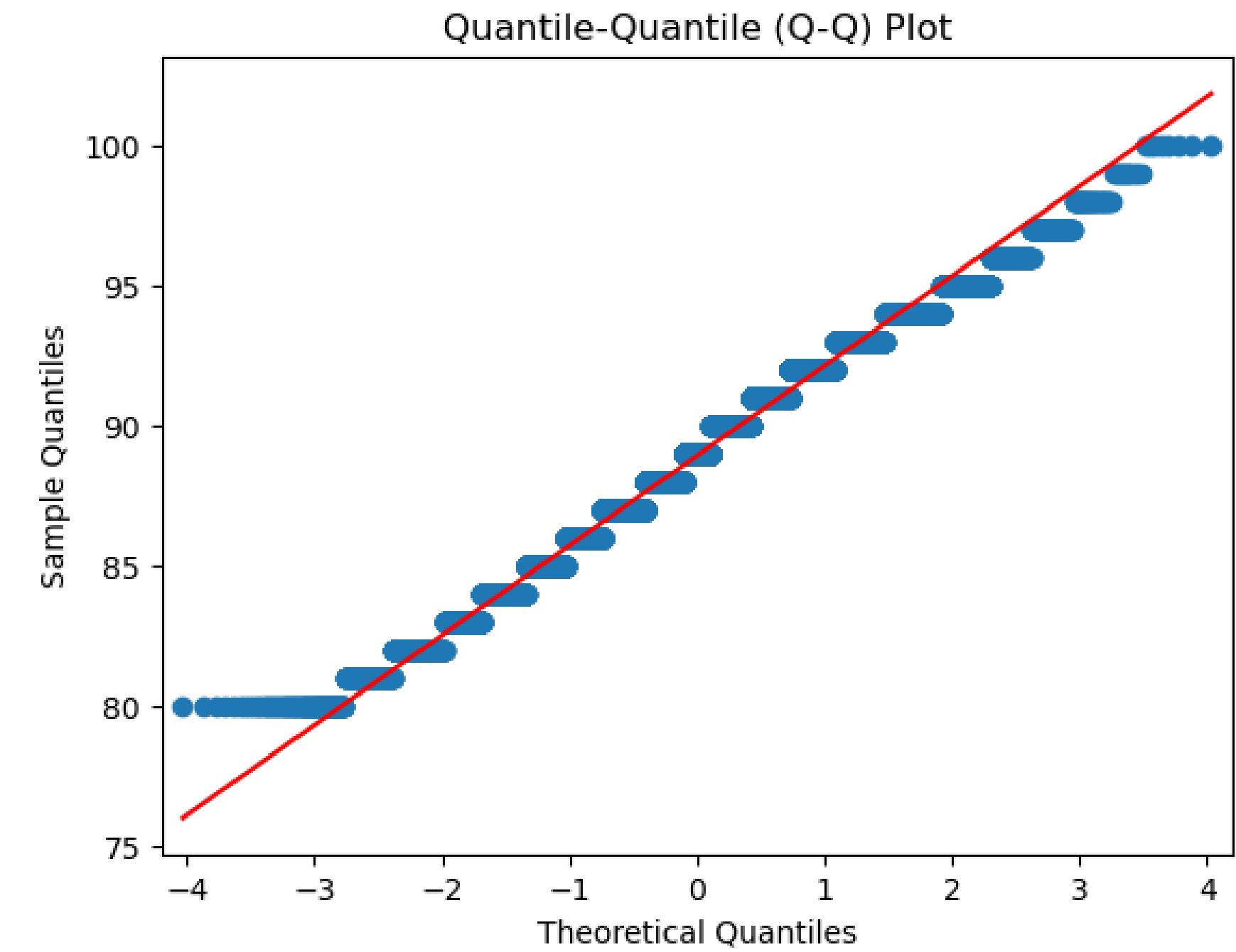
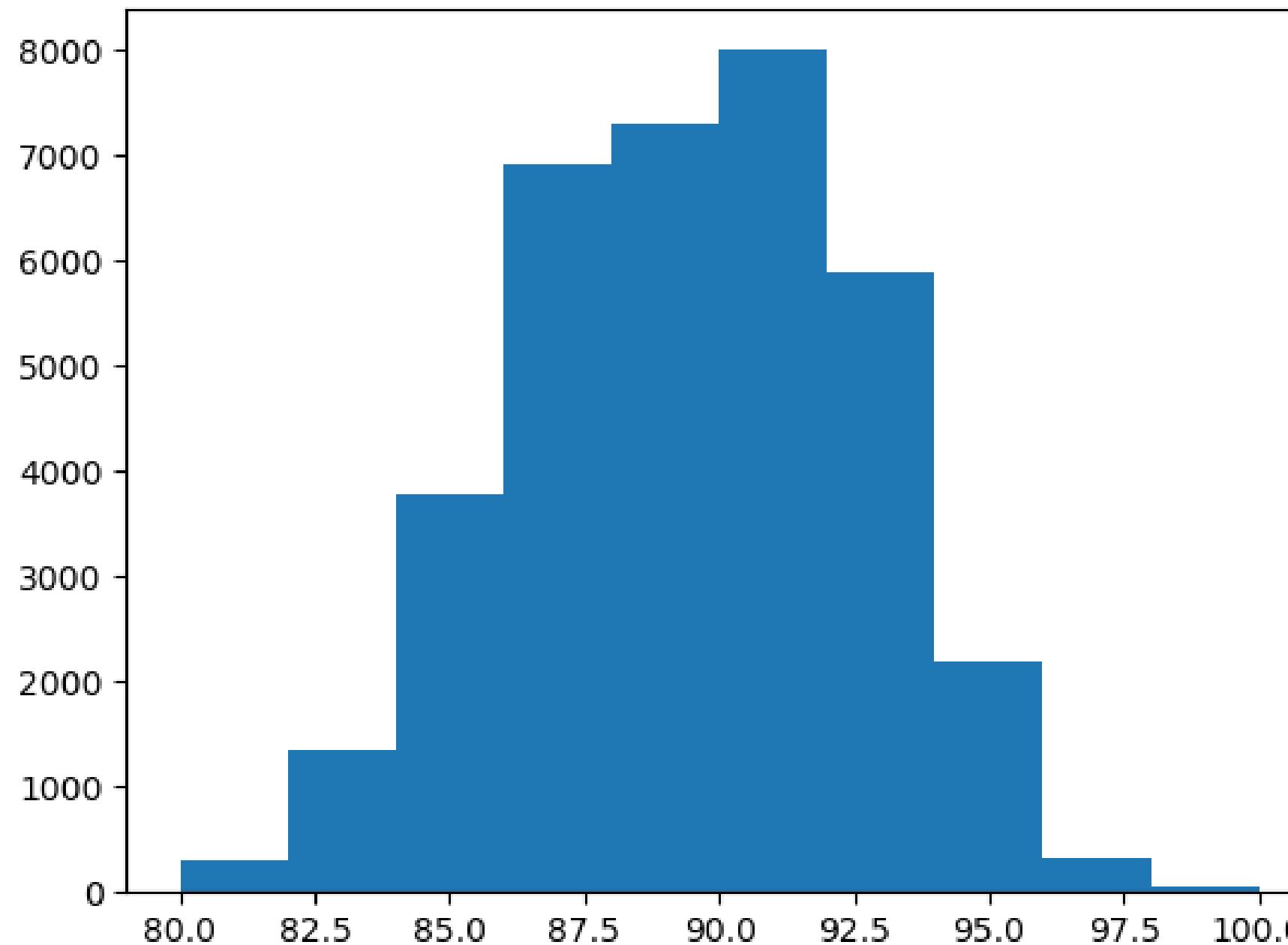
H1: The mean score of red wines is higher than the mean score of white wines.

Alpha: 0.05



variance of **Red wine** is 10.288635418551847

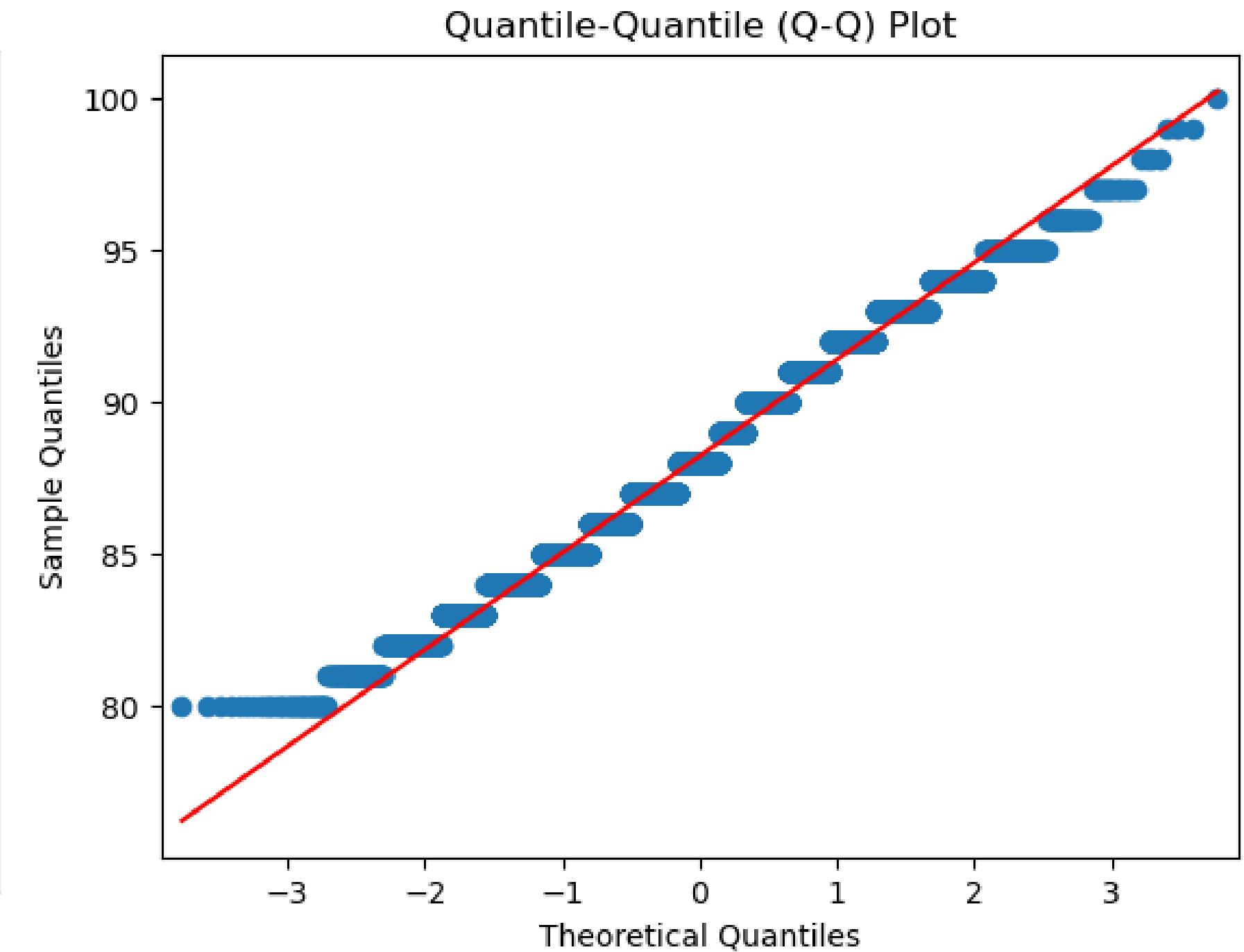
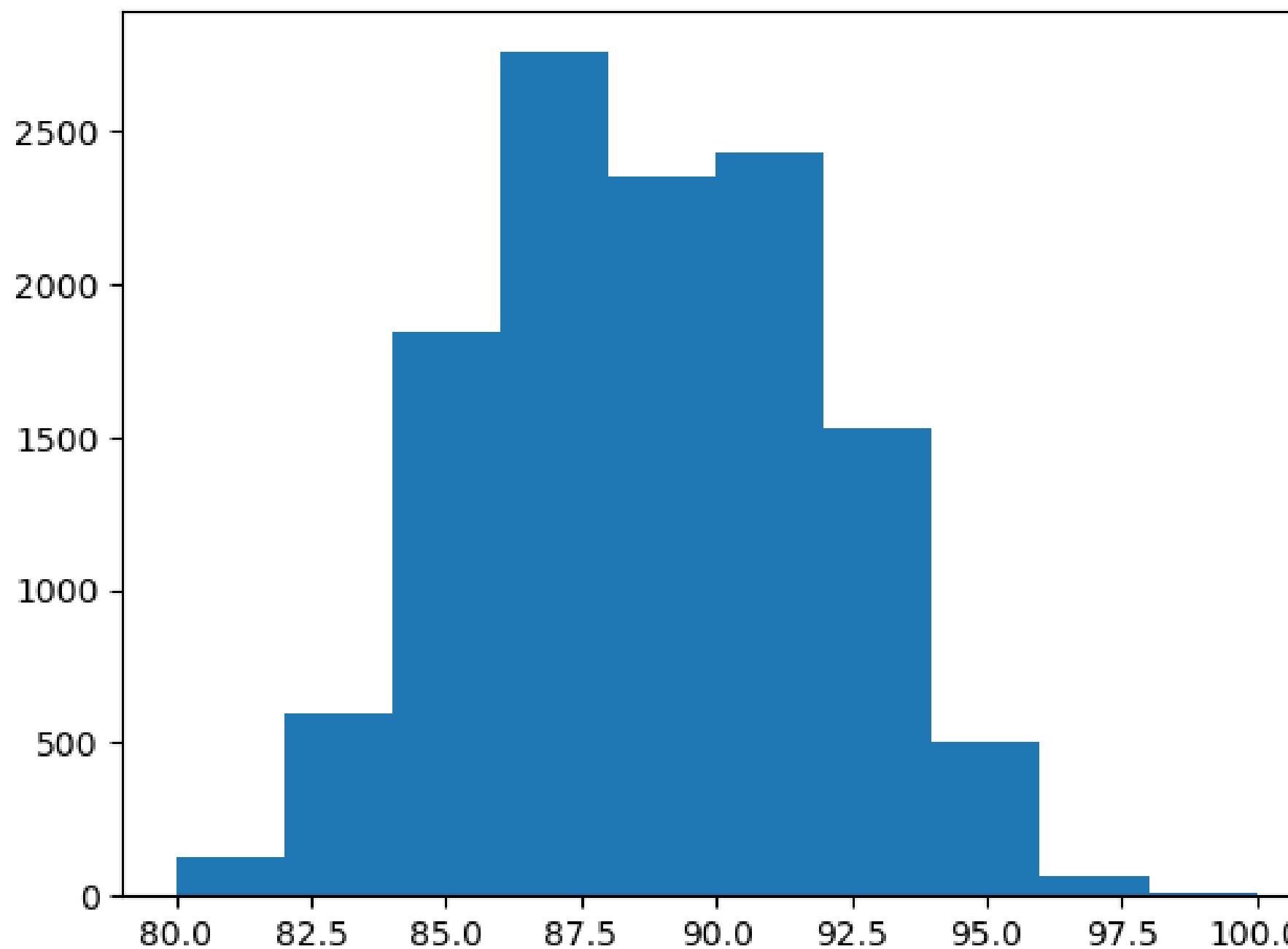
Shapiro-Wilk: W:0.986842691898346 p:0.0





variance of **White wine** is 10.137882897879102

Shapiro-Wilk: W:0.9858100414276123 p:4.533615388182705e-33





FINAL FINDINGS

- On average red wines are higher rated than white wines in the upper point analysis
- Both groups have normality and the variance is similar
- The difference between means is statistically significant
- P-value is less than alpha so we can reject the null hypothesis

H1: The mean score of red wines is higher than the mean score of white wines.

PAIRED T-TEST

TtestResult(statistic=20.741037276519087,
pvalue=3.840109180432591e-95, df=48362.0)

03.

ARE HIGHER RATED WINES MORE COMMON IN A SPECIFIC PLACE?

Analyze points, regions, weather

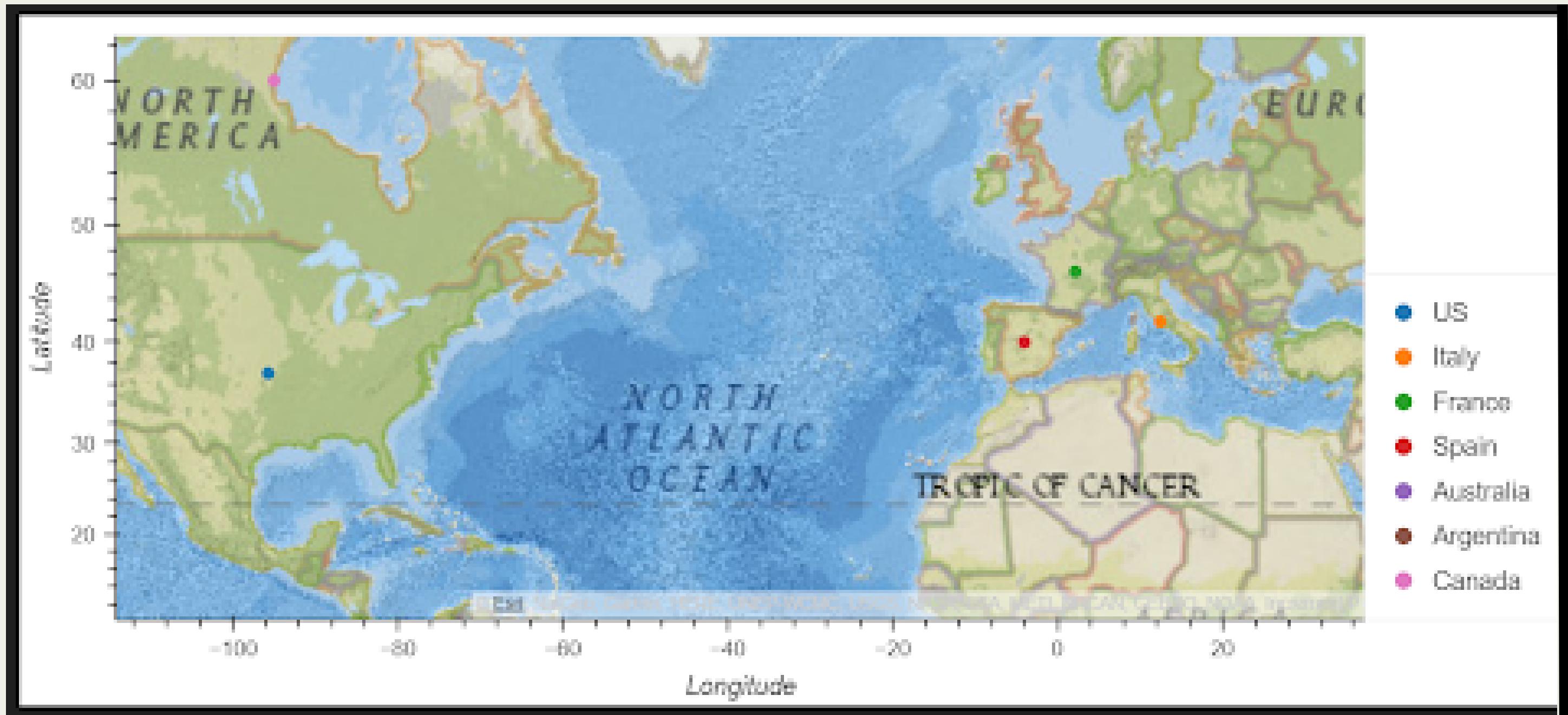
03.

H0: Ratings and origin have no relation.

H1: If the wine is higher rated, its places of origin have similar weather conditions



USING THE COORDINATES_DF AND MAP THE POINTS





HIGH_REGION_COORDINATES_DF

PLOT THE POINTS ON A MAP

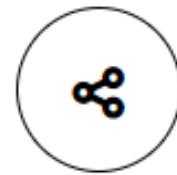




Which factors do we analyze?



WINE ENTHUSIAST

[SHOP](#)[RATINGS](#)[CULTURE](#)[BLOG](#)

from unusually hot weather during a particular harvest to a winemaker that intentionally shoots for opulence. Ultimately, however, climate plays the most pivotal role in a wine's expression.

Winemakers know that wine grapes grow best in climates that aren't too tropical, too arid or too reminiscent of arctic tundra. Most of the suitable climates are found between 30° – 50° latitude, both north and south.

Climate is also a function of elevation. Vineyards perched in the mountains of Trentino, [Italy](#), for instance, enjoy cooler temperatures than fruit on the valley floor.

Other factors can include the amount of rainfall, humidity, and cooling or warming forces like cloud cover and wind patterns. Sonoma's Petaluma Gap enjoys cooling gusts, while North Africa's dry scirocco winds can cause Sicily's temperatures to soar.





Open Weather Statistical API

sip-by
-sip



This product is useful for analysis of climate indicators and statistical approach to weather forecasting. Statistical Weather Data API is calculated based on our [Historical weather data](#).

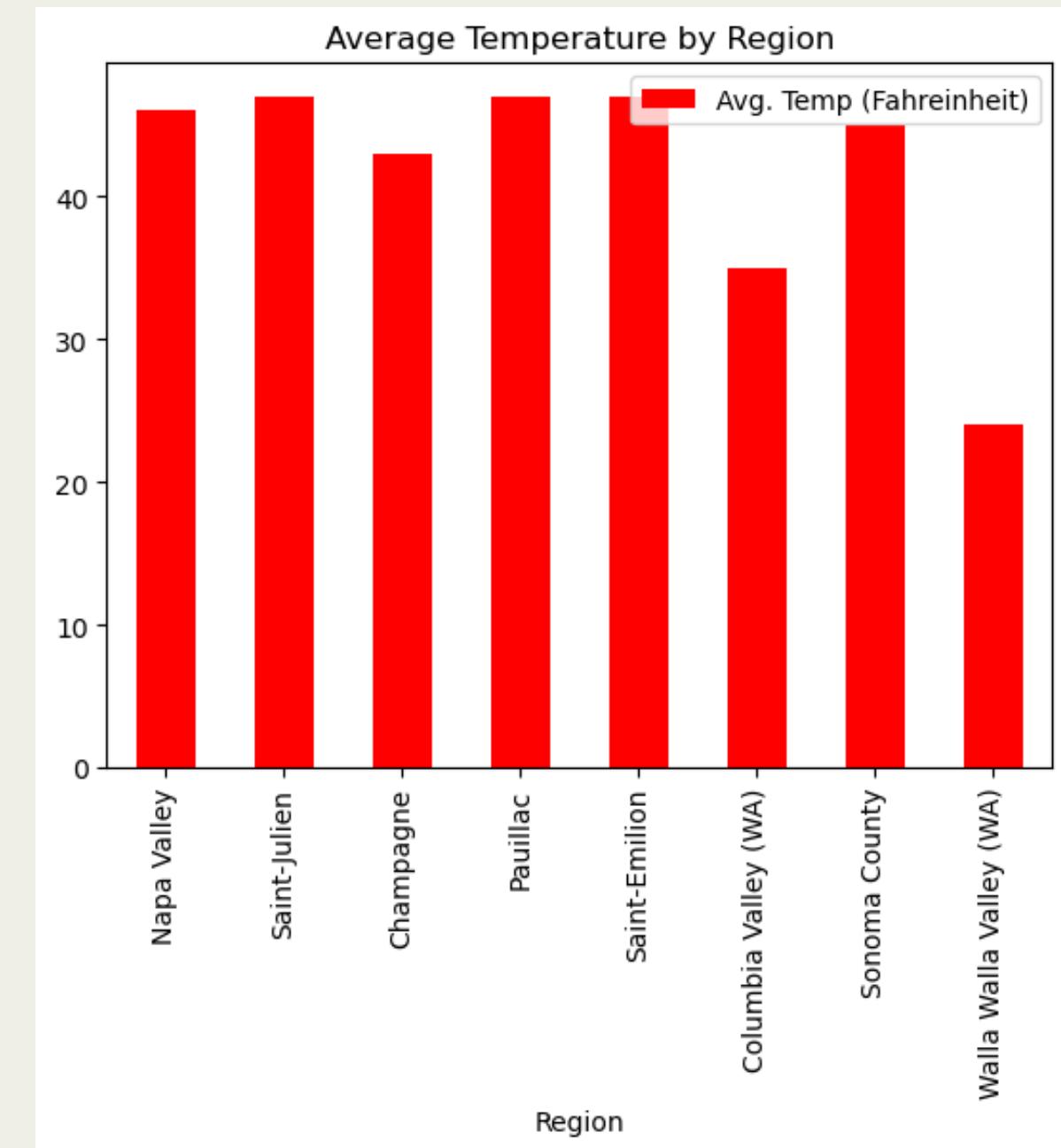
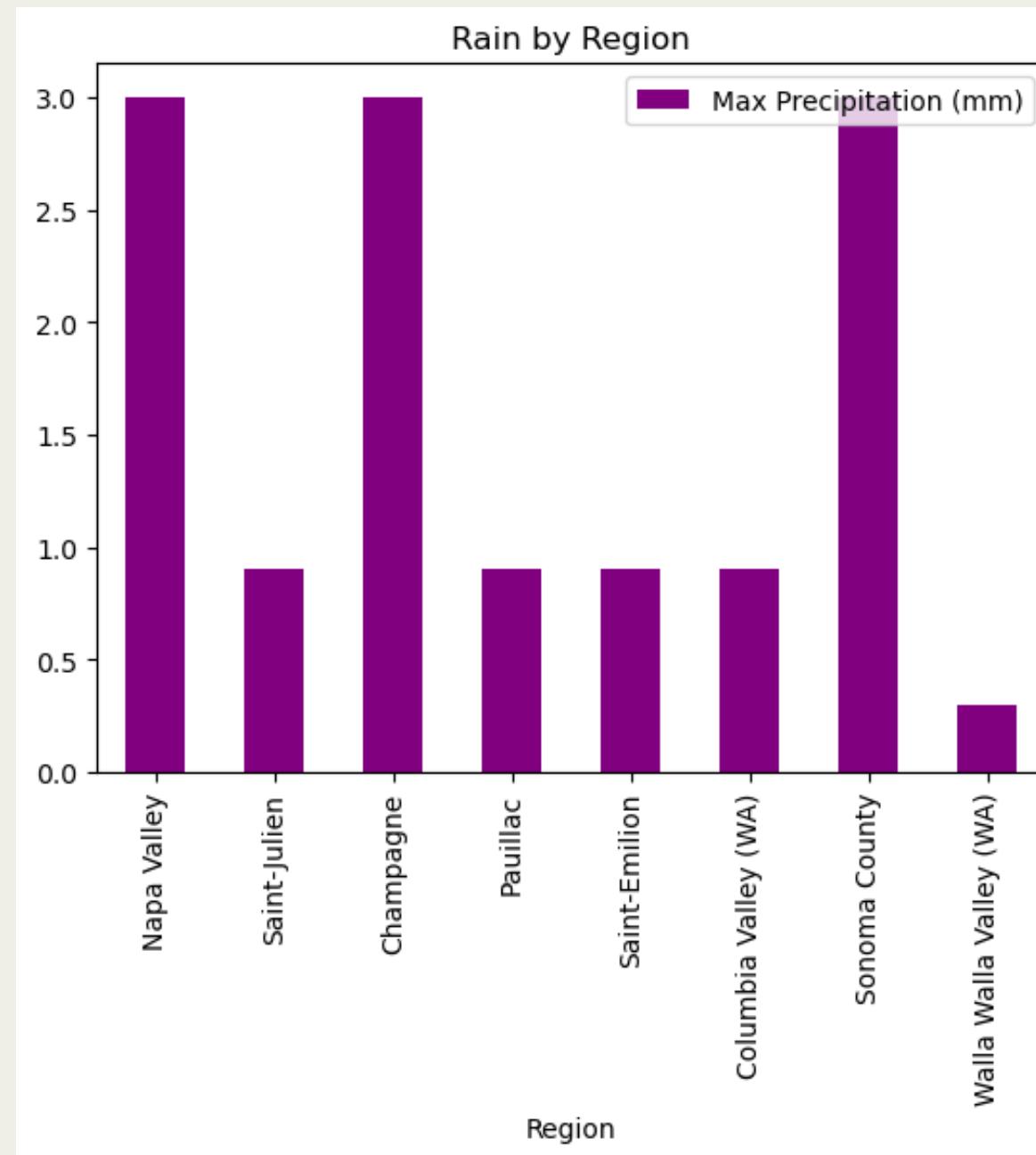
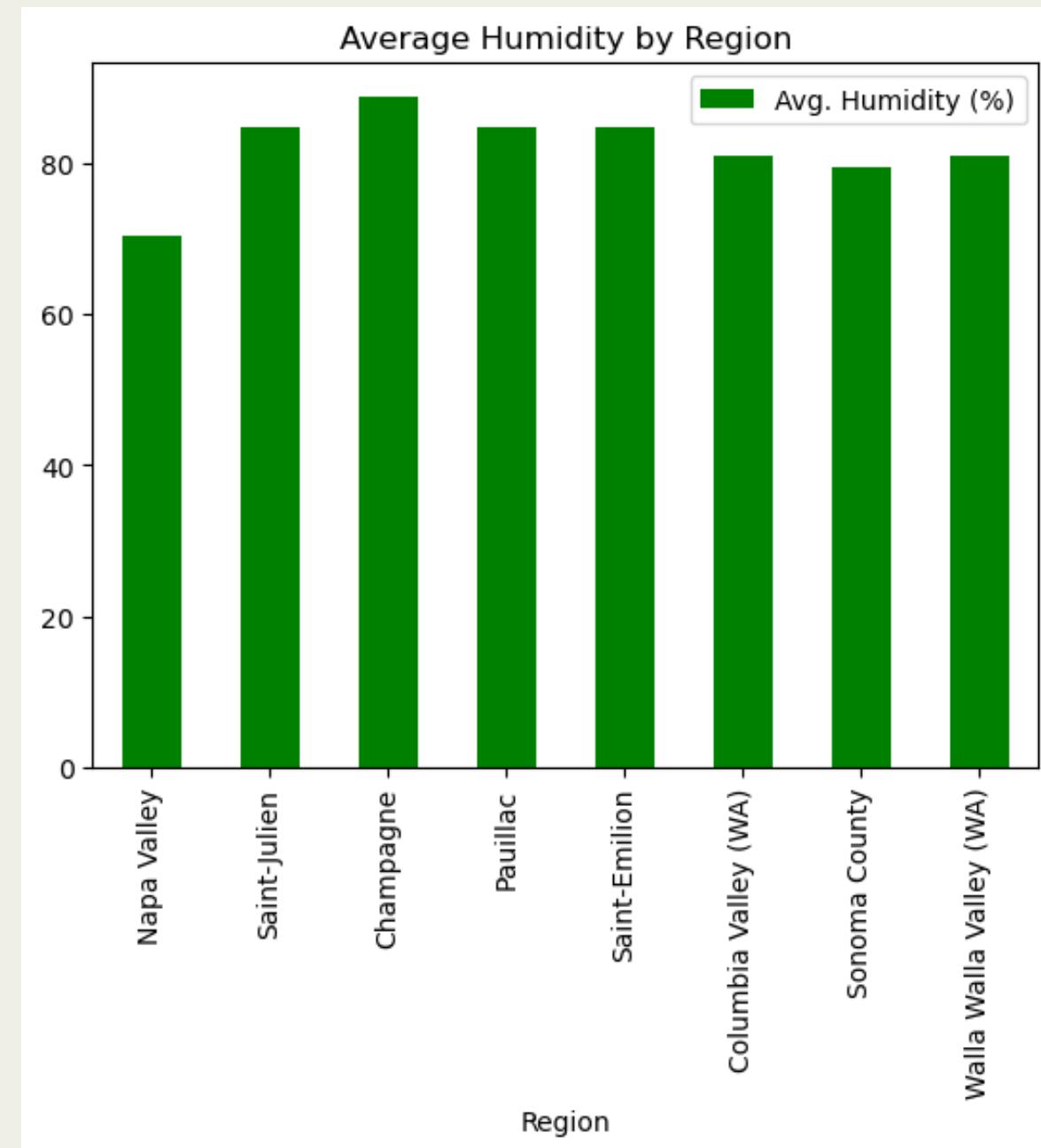
In this product, you will be able to request basic statistical data for the following weather characteristics:

- Temperature
- Pressure
- Humidity
- Wind
- Precipitation
- Clouds



FINAL DATAFRAME

| | Region | latitude | longitude | Average Temperature (kelvin) | Max Precipitation (mm) | Avg. Humidity (%) | Avg. Temp (Fahreinheit) |
|---|-------------------------|-----------------|------------------|-------------------------------------|-------------------------------|--------------------------|--------------------------------|
| 0 | Napa Valley | 38.297539 | -122.286865 | 281.37 | 3.0 | 70.22 | 46 |
| 1 | Saint-Julien | 45.157199 | -0.739182 | 281.51 | 0.9 | 84.69 | 47 |
| 2 | Champagne | 48.957500 | 4.365000 | 279.50 | 3.0 | 88.74 | 43 |
| 3 | Pauillac | 45.196440 | -0.748730 | 281.51 | 0.9 | 84.69 | 47 |
| 4 | Saint-Emilion | 44.883330 | -0.150000 | 281.51 | 0.9 | 84.69 | 47 |
| 5 | Columbia Valley (WA) | 49.000000 | -122.083300 | 275.26 | 0.9 | 80.81 | 35 |
| 6 | Sonoma County | 38.525290 | -122.922540 | 280.44 | 3.0 | 79.42 | 45 |
| 7 | Walla Walla Valley (WA) | 46.064583 | -118.343018 | 268.94 | 0.3 | 80.89 | 24 |



FINAL FINDINGS:

- Wine ratings that are high scored (90 and above) come from specific countries
- The US and France produced wines rated 100
- Regions with higher-rated wines have similar weather conditions.
- Further testing, average all the points in these specific regions and find if there is a statistical correlation between the points and their weather conditions.



04.

HOW DO YOU KNOW
HOW MUCH WINE TO
DRINK?



Thank you!

FROM OUR TEAM



Data Analysis
Jessica Hernandez



Data Analysis
Cesar Rojas



Data Analysis
Francesca Roefaro

sip-by
-sip





Sources

Kaggle

<https://www.kaggle.com/datasets/zyncide/wine-reviews>

Open Weather API

<https://openweathermap.org/api/statistics-api>

Wine Enthusiast

<https://www.wineenthusiast.com/basics/cool-vs-warm-climate-wine/>

Documentation

<https://pandas.pydata.org/docs/index.html>

<https://www.statology.org/pandas-t-test/>

<https://matplotlib.org/stable/>

sip-by
-sip

