Part 1: BlueBerry Winery: New Name in VihnoVerdo, Portugal

Yurii Novachynskyi • 10.10.2023

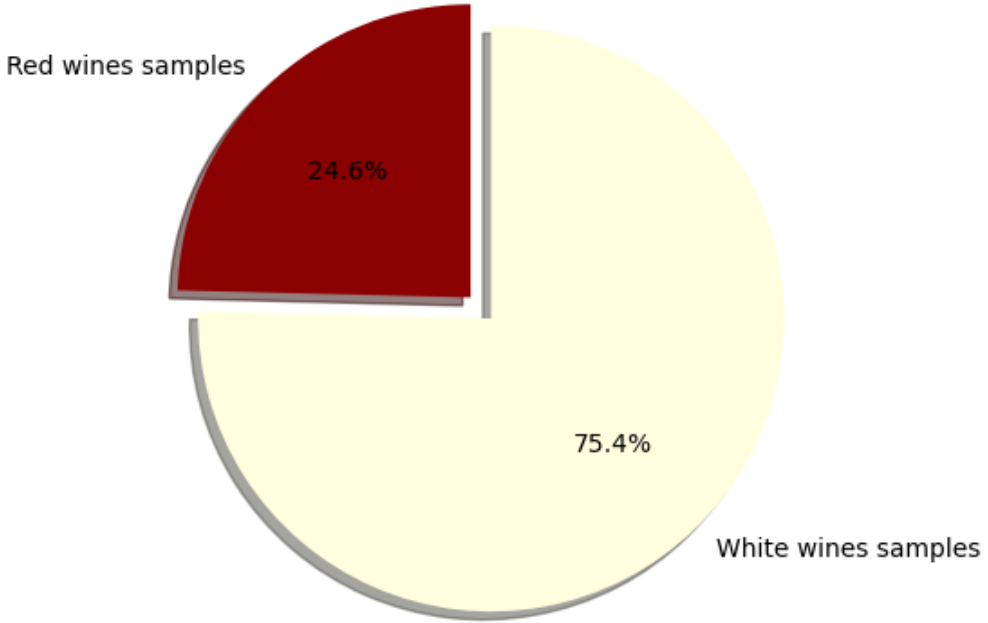
Out database:
Vihno Verde wine
from North Portugal

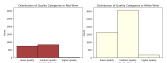
Samples:
Red wine – 1599
White wine – 4898

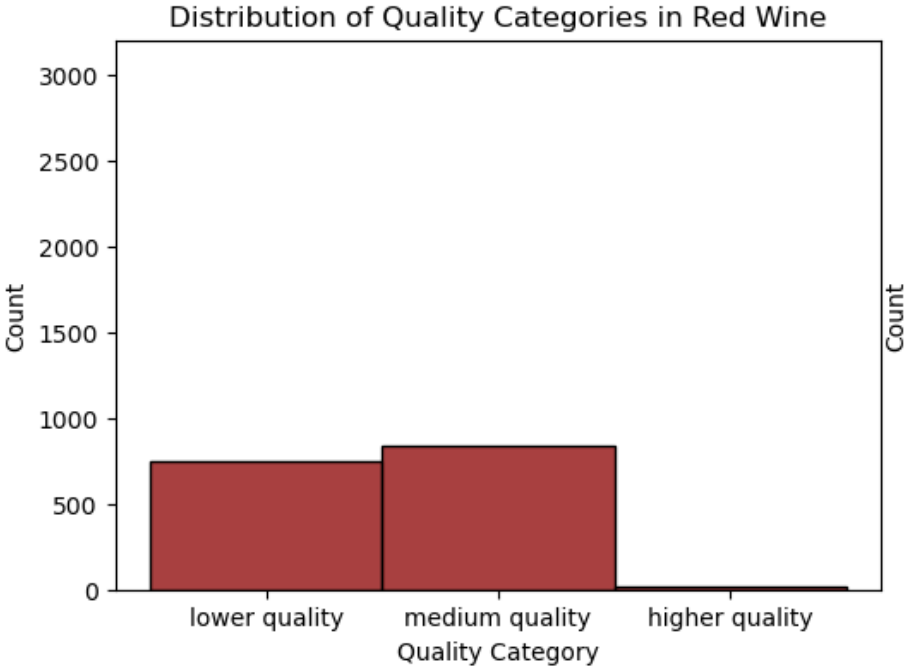
Wine characteristics:

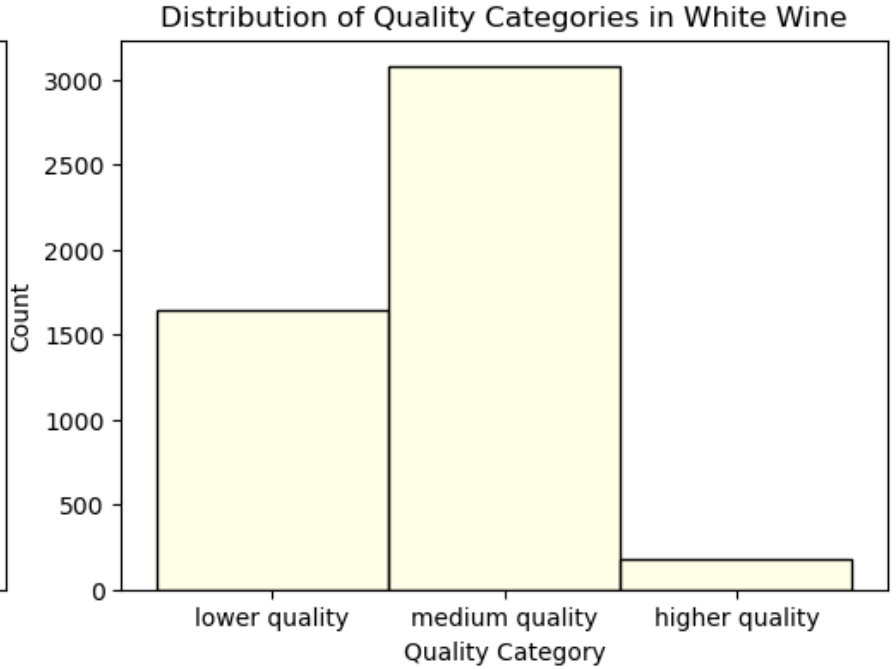
Physicochemical
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
Alcohol

Sencory: Quality
(from 0 to 10, grouped into 3 categories)






Distribution of Quality Categories in Red Wine
Count
3000
2500
2000
1500
1000
500
0
lower quality
medium quality
higher quality
Quality Category


Distribution of Quality Categories in White Wine
Count
3000
2500
2000
1500
1000
500
0
lower quality
medium quality
higher quality
Quality Category

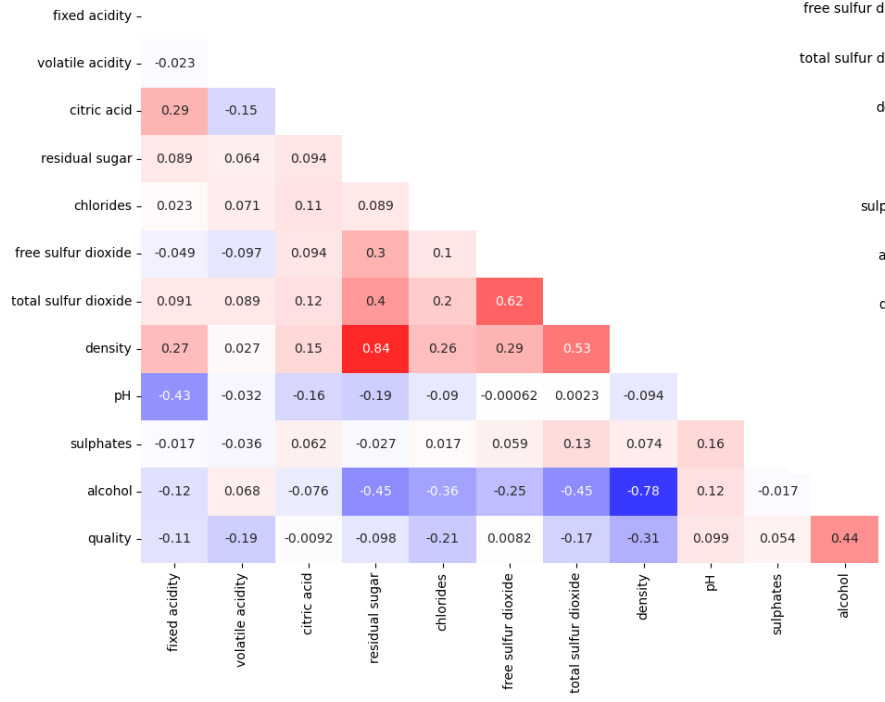
Vihno Verde region is famous for it`s white wine and it`s very fresh, due to its natural acidity, with fruity and floral aromas that depend on the grape variety.

The red and rose Vinho Verde wines are much less common than the white ones what is caused by the region's relatively cool temperatures and high level of rainfall.

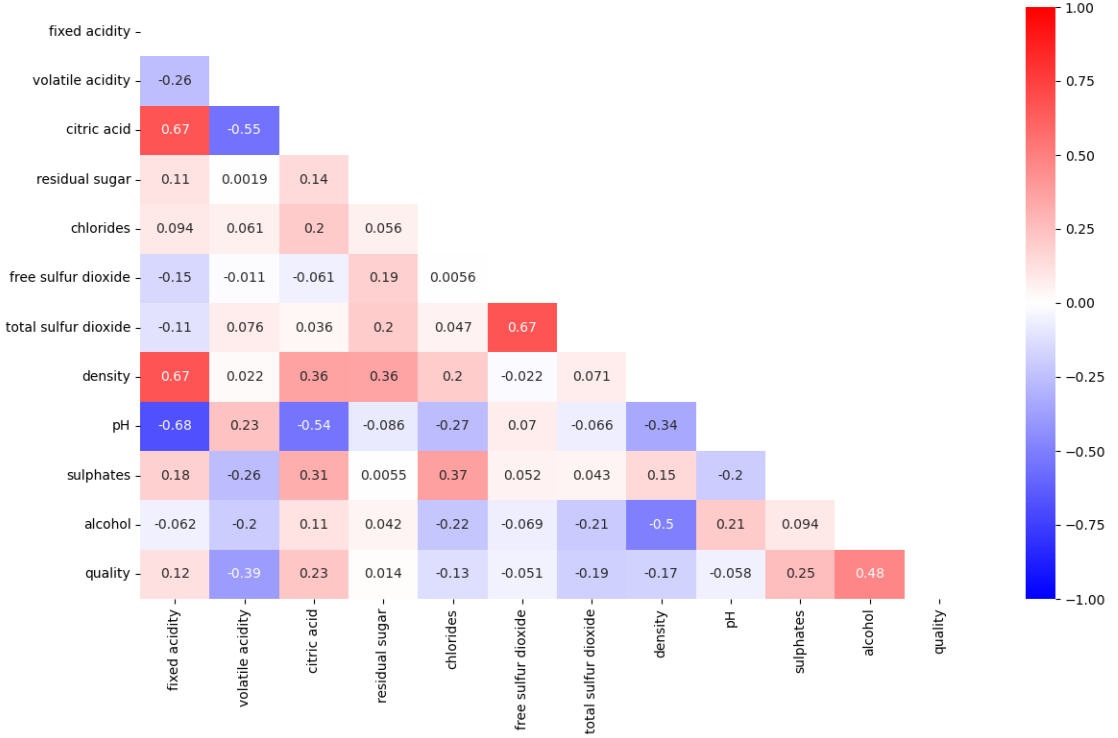
@ wikipedia.org/Vinho_Verde

Correlations coefficients

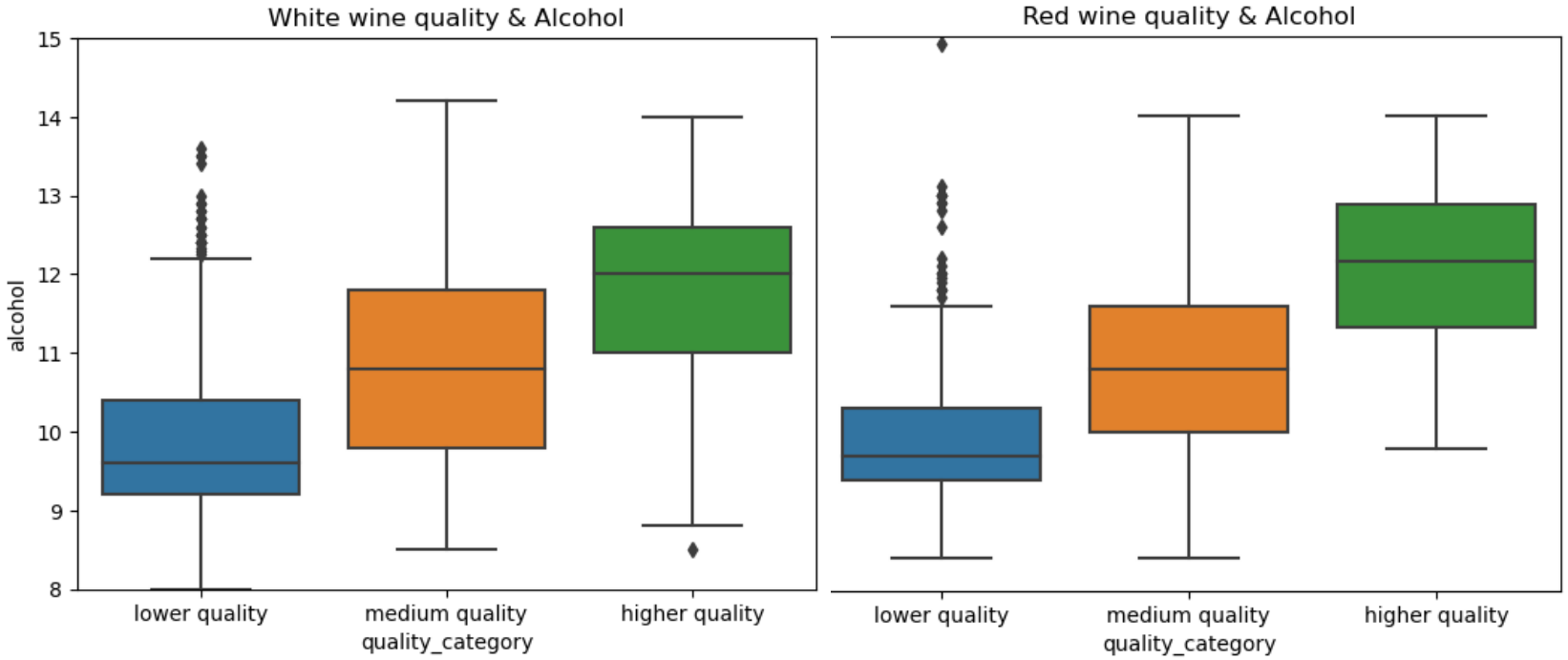
Correlation coefficients between features in the White wine samples



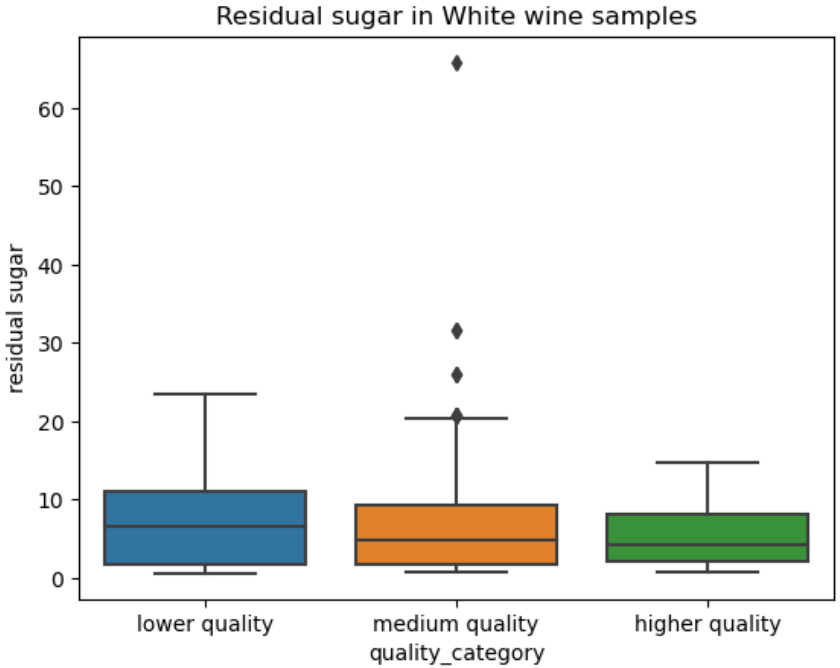
Correlation coefficients between features in the Red wine samples

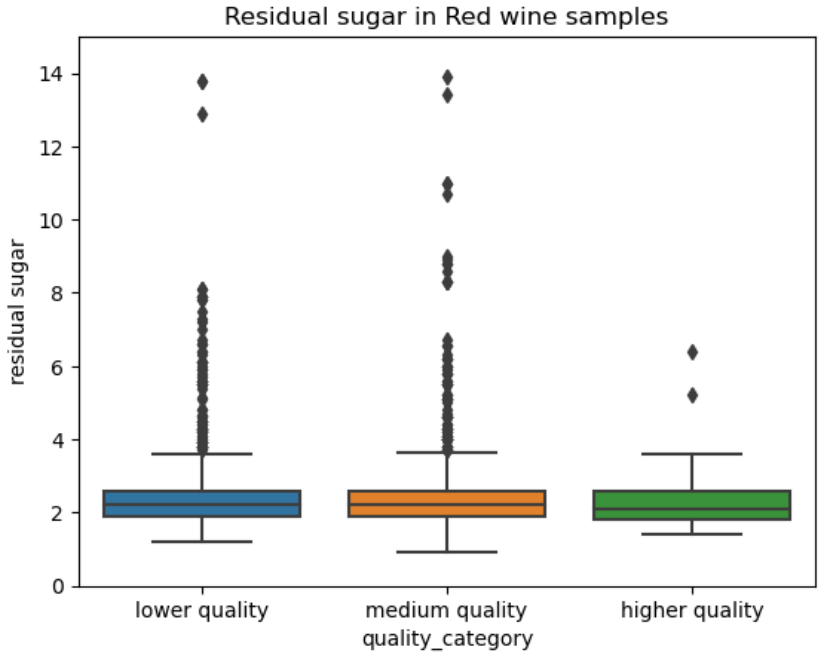


Correlations of Wine quality and % of Alcohol

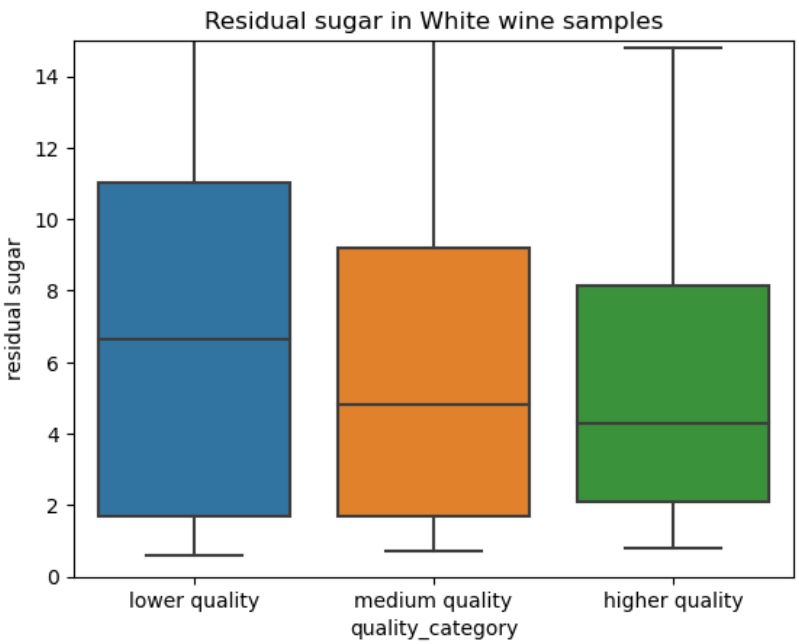


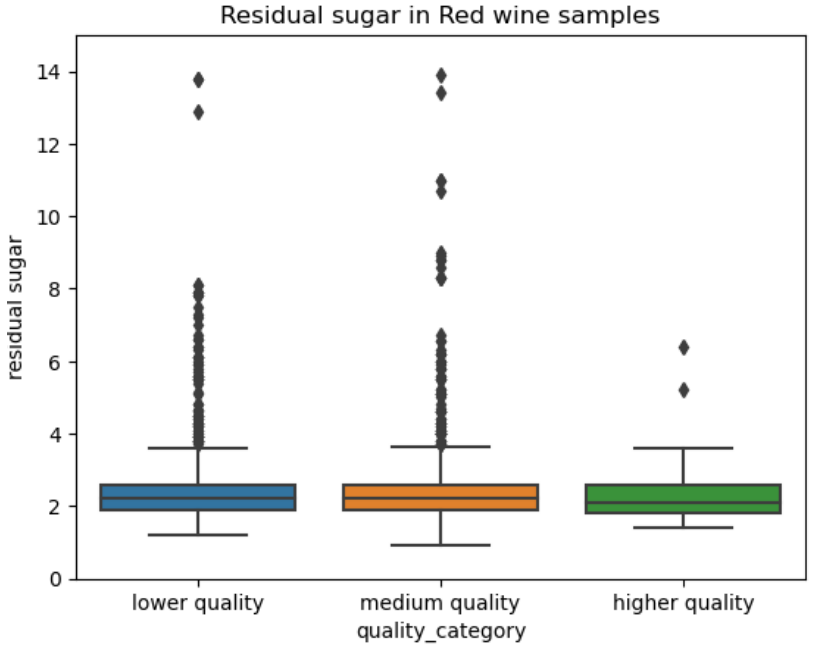
Correlations of Wine quality and Residual Sugar





Comparing wine and Residual Sugar

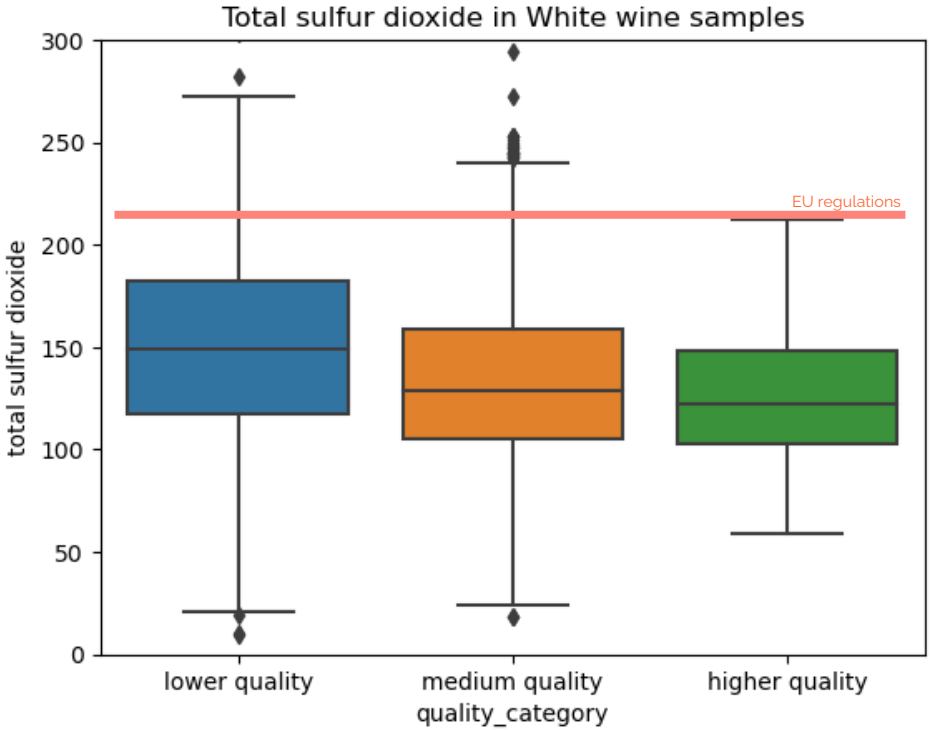


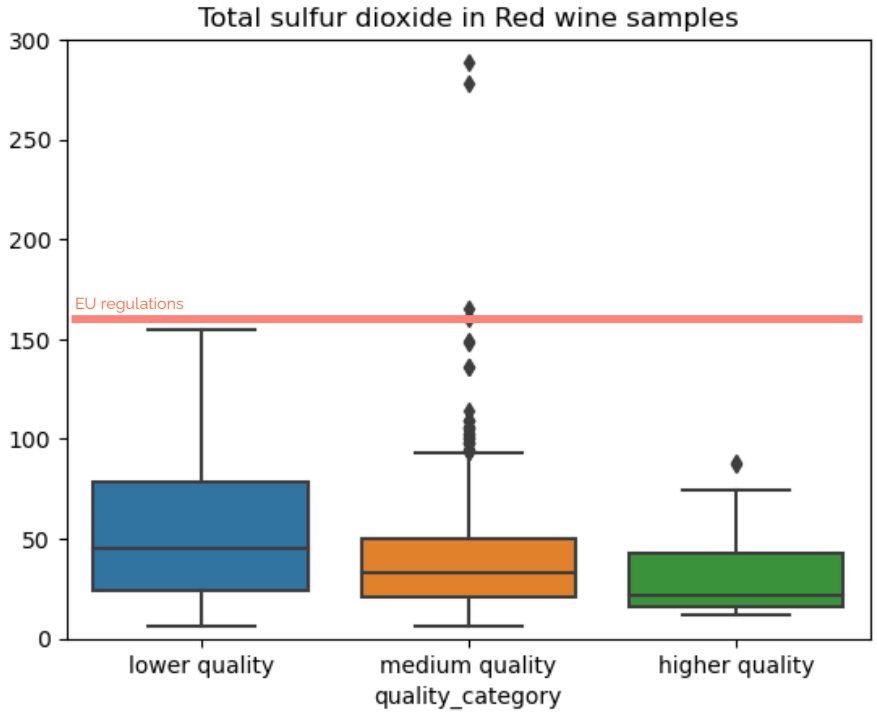


Wine with a higher level of Alcohol (11-13%) received a higher quality score.

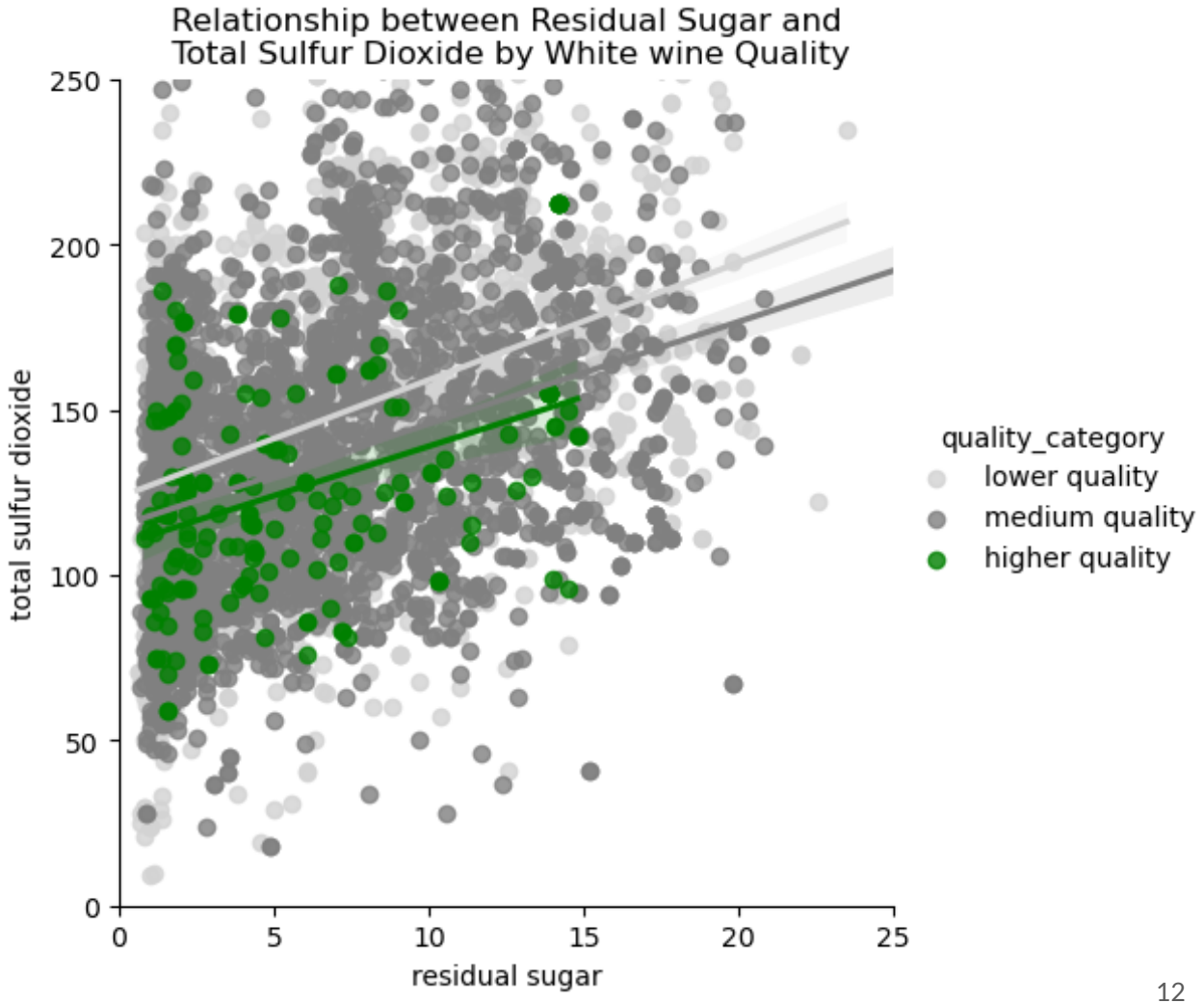
In the white wine category, residual sugar is partly related to the quality of the wine. Wines with an RS level of 2.4 to 8.2 form the main group of high quality wines.

Correlations of Wine quality and Total Sulfur Dioxide (mg/dm3)

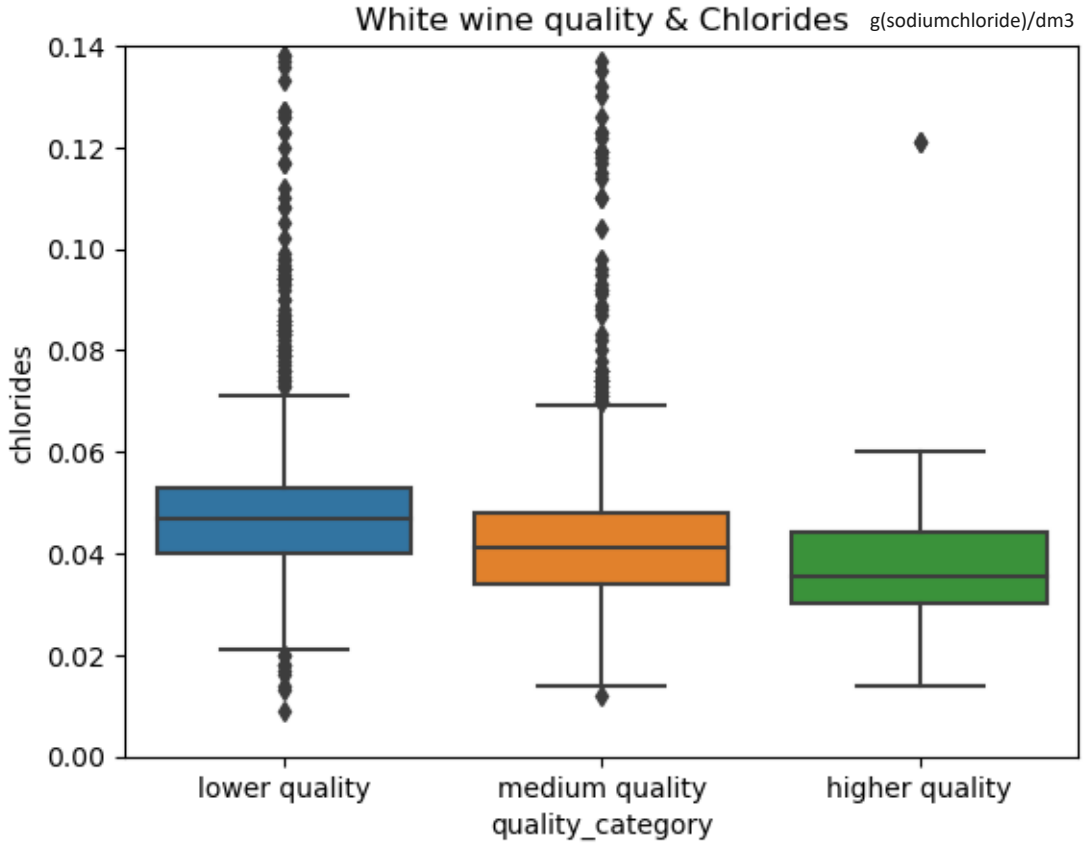




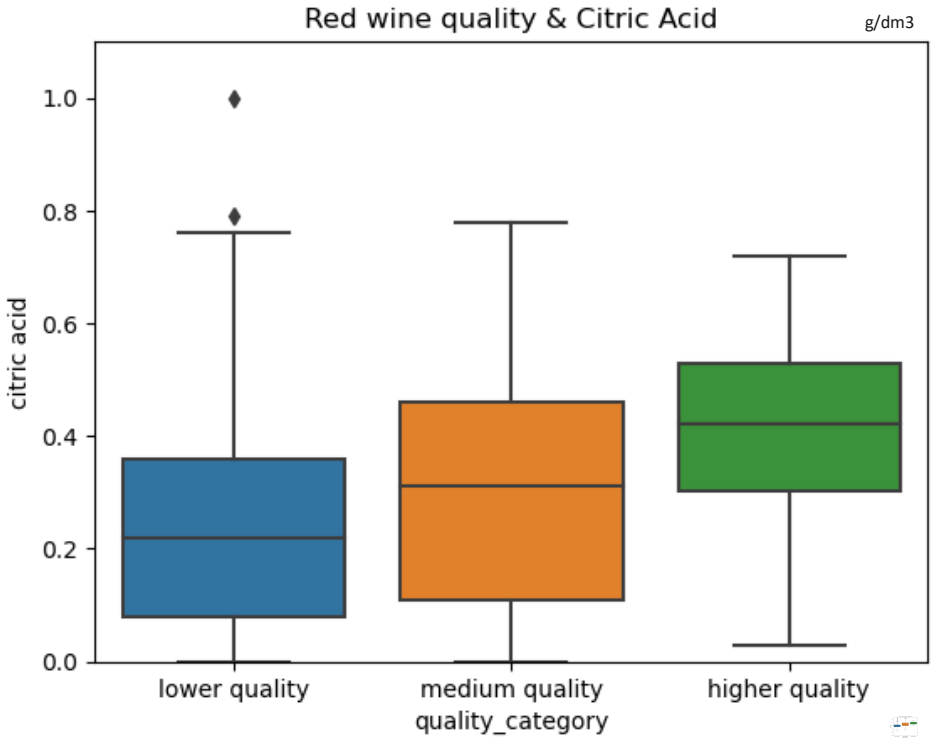
- **The balance between preserving and taste of wine**
- **Lower TSD - better original taste**

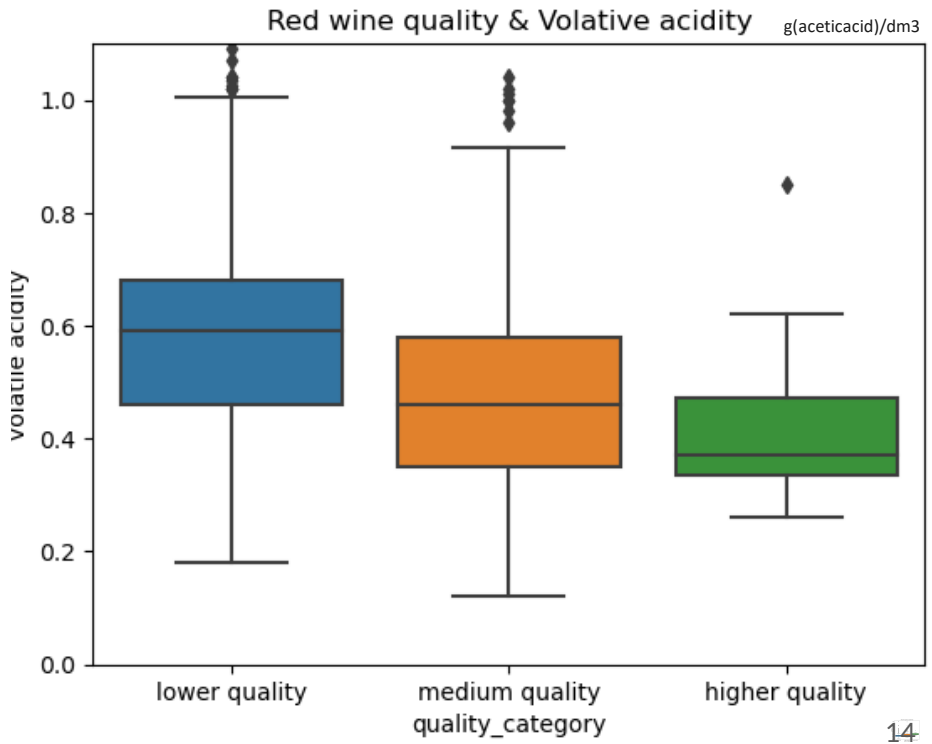


More correlations between white wine quality and chemical characteristics

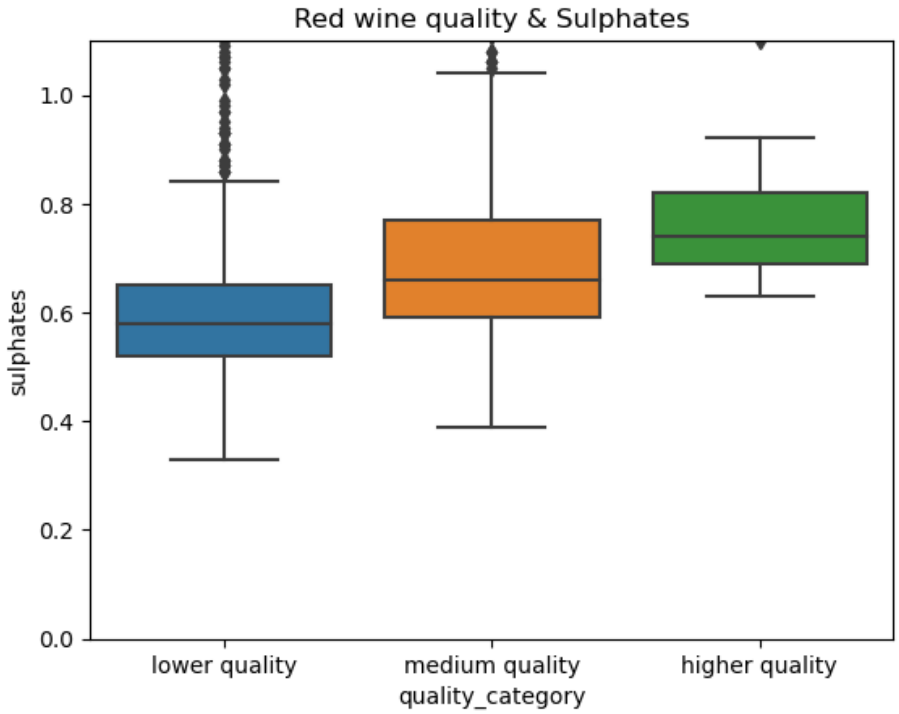


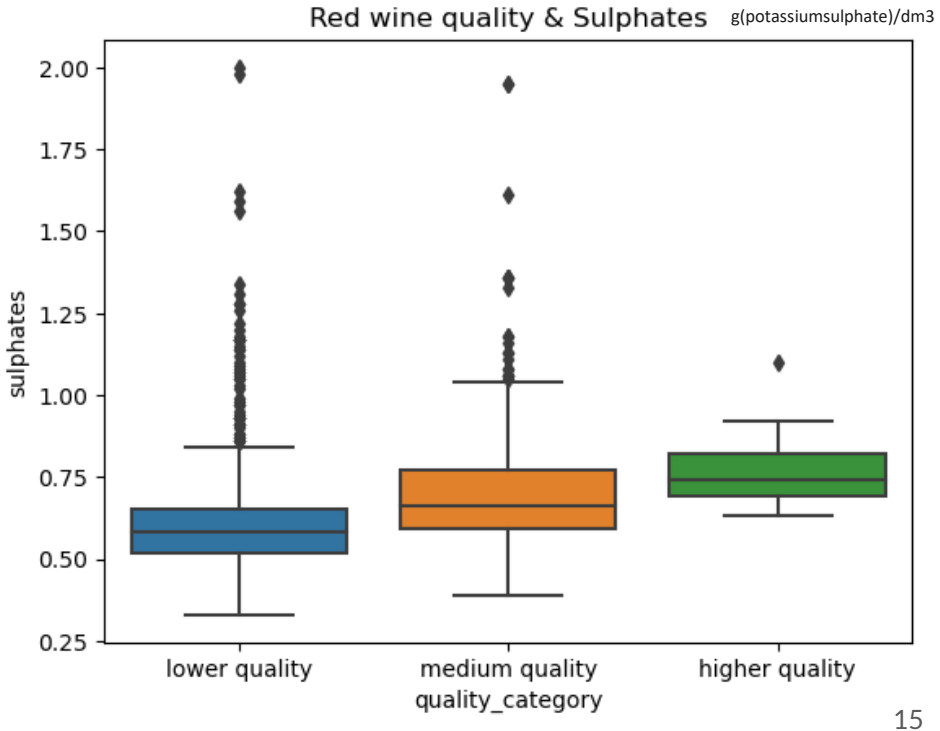
More correlations between red wine quality and chemical characteristics





More correlations between red wine quality and chemical characteristics





RECOMENDATIONS

Given the current popularity of young white wines from the region, it can be costly to promote certain red wines. Specializing in white wine for the global market and selling red wine locally to ensure a diversified product portfolio can be a good solution for the new winery.

Higher is better. Controlling the alcohol content (in combination with other characteristics) will allow to better predict the quality of wine.

RECOMENDATIONS

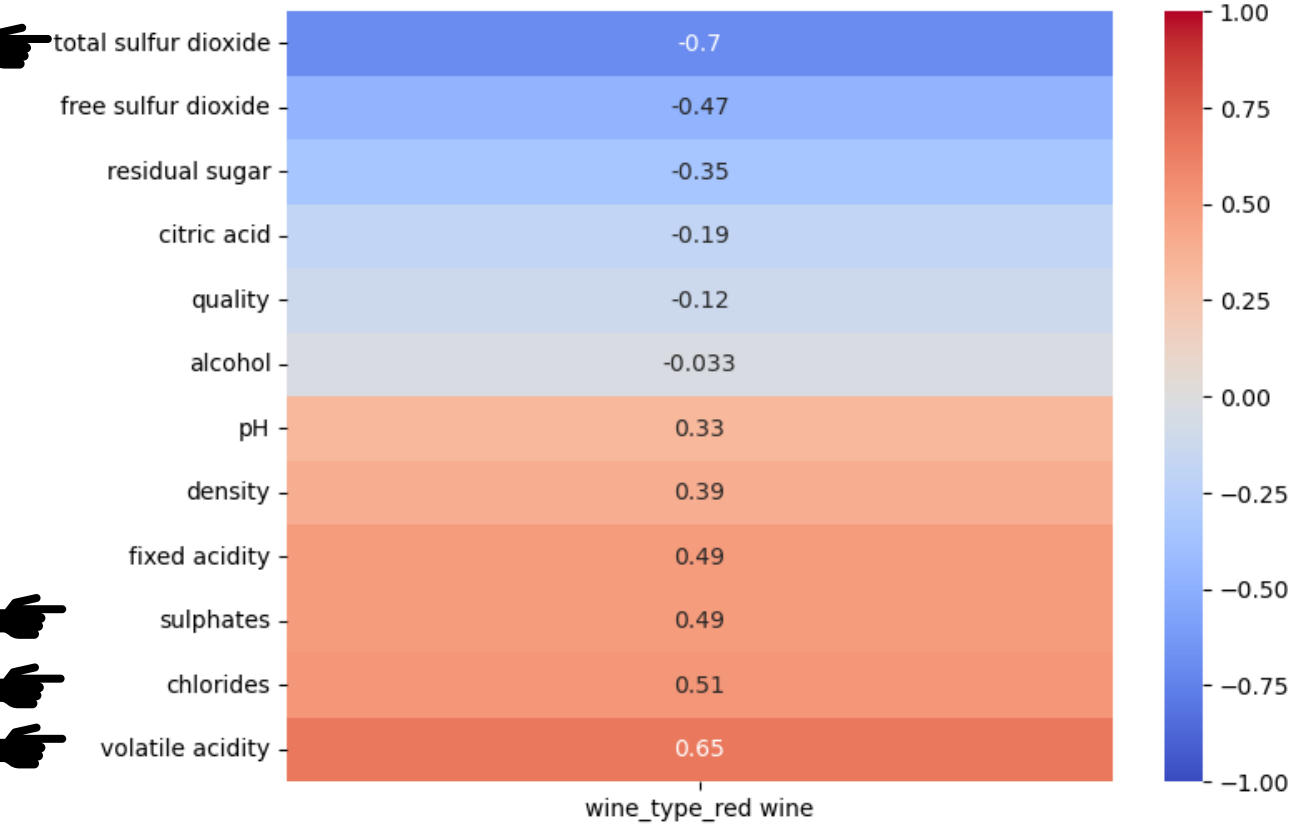
Reducing the amount of sugar in white wine can have a positive impact on wine quality.

For sulphur dioxide in white wine, a balance must be struck between preservation and taste of the wine. Wines that are sold locally (e.g. at winery tastings) do not require high levels. Globally marketed wines require higher levels for transport and to preserve the quality of the wine.

Part 2: Predicting wine, wine quality and wine price

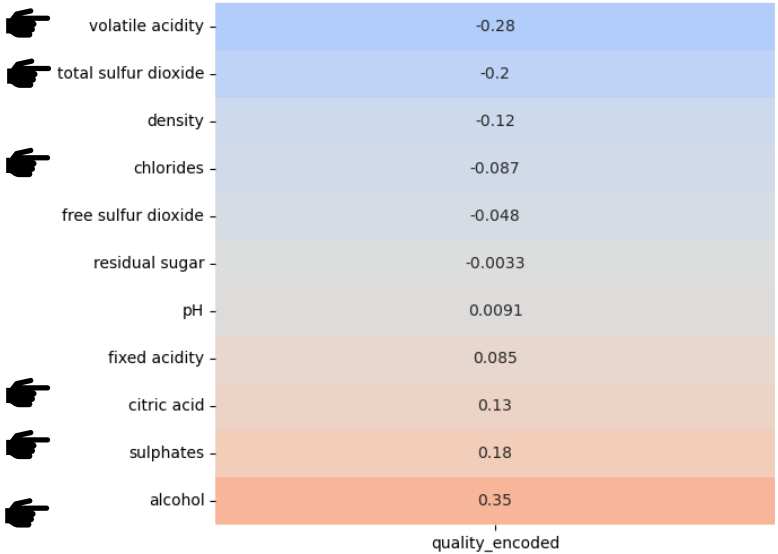
Yurii Novachynskyi • 25.10.2023

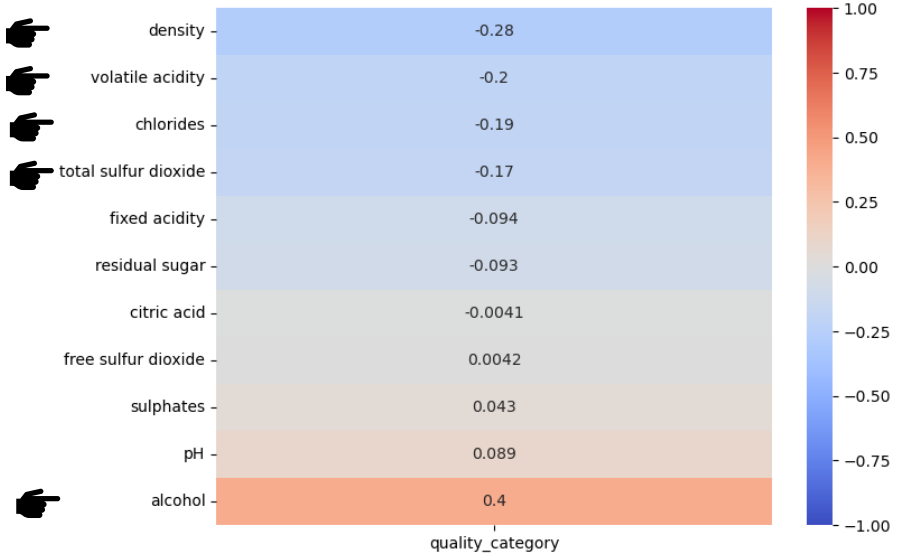
Choosing values for machine learning model



Total sulfur dioxide, volatile acidity, chlorides and sulphates were choosen for predicting type of wine

Choosing values for machine learning model

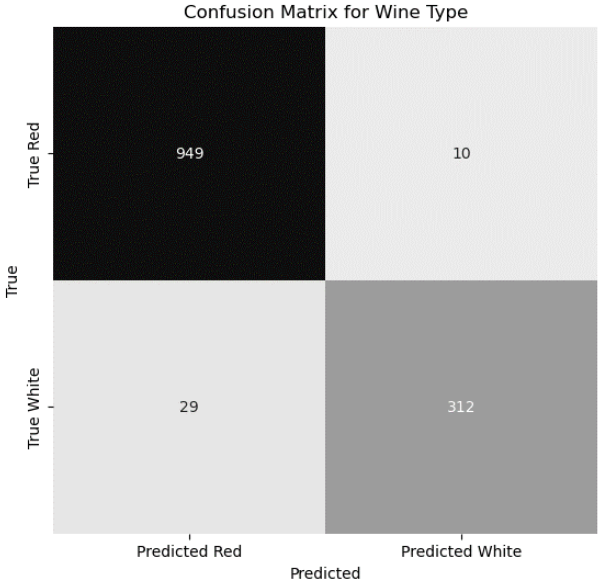


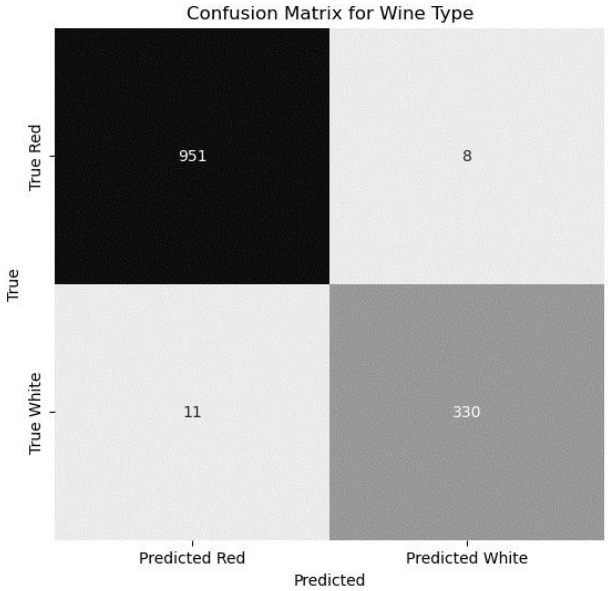


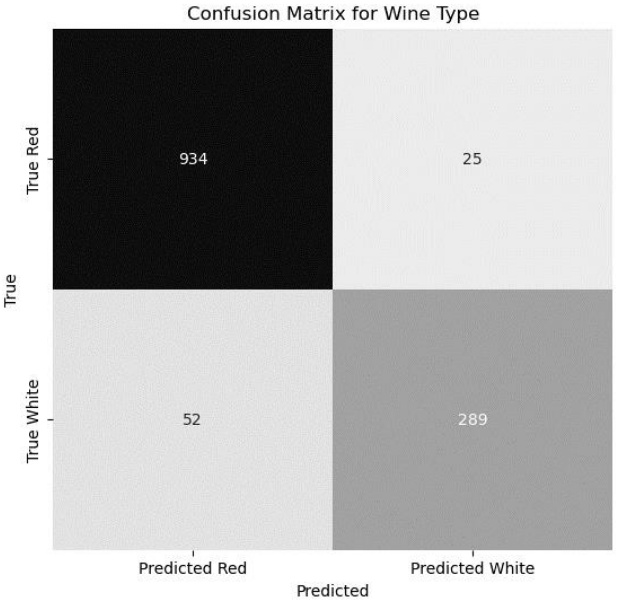
Red wine: alcohol ,volatile acidity, total sulfur dioxide, sulphates , citric acid, chlorides

White wine: alcohol , density, volatile acidity, total sulfur dioxide, chlorides

Prediction type of wine

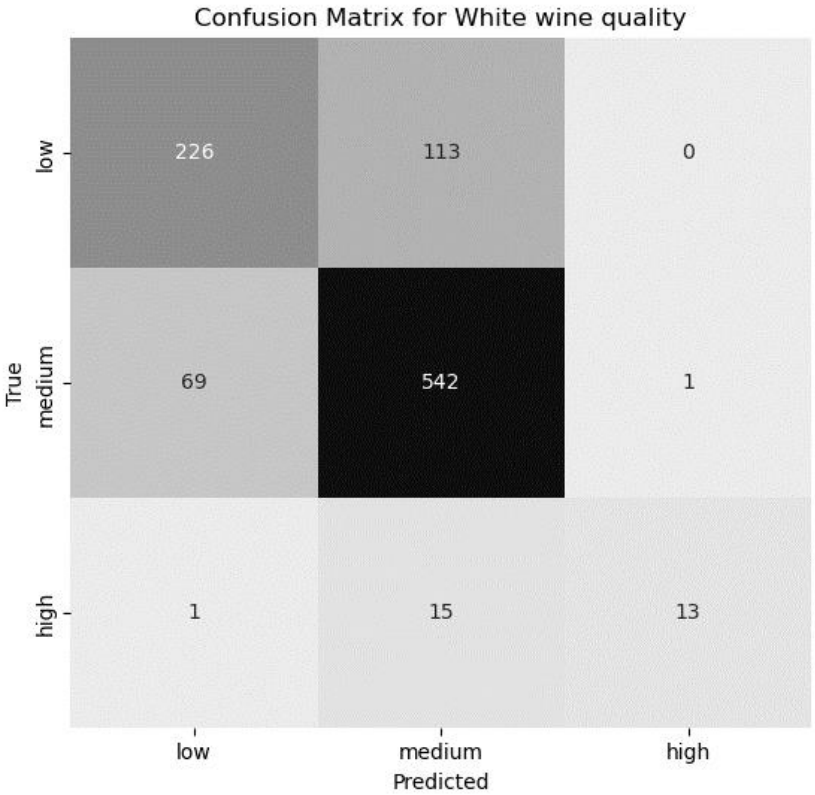






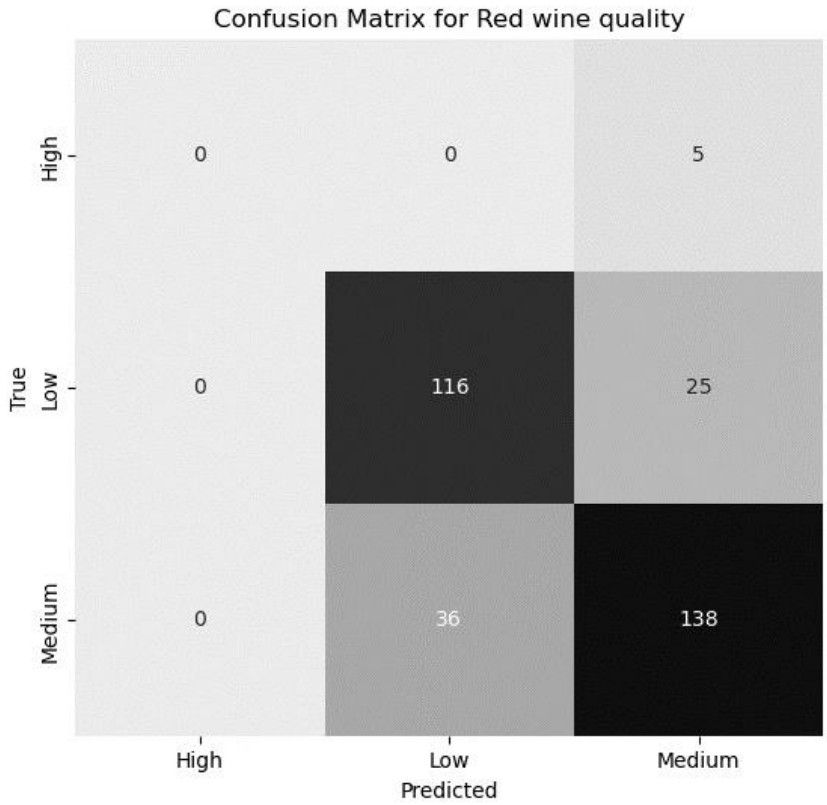
LogisticRegression	**DecisionTreeClassifier**	**KNeighborsClassifier**
Accuracy: 97.69	Accuracy: 99.96	Accuracy: 96.73
Cohen-Kappa score: 0.92	Cohen-Kappa score: 0.96	Cohen-Kappa score: 0.84

Predicting white wine quality



Random Forest Classifier
Accuracy: 79.69
Cohen-Kappa score: 0,57

Predicting red wine quality



Random Forest Classifier
Accuracy: 79.375
Cohen-Kappa score: 0,59

Take-away for DA: depents from encoding methode, we can get different order. Make sure, it`s same

Predicting the prices

***Wine Spectator's* 100-Point Scale**

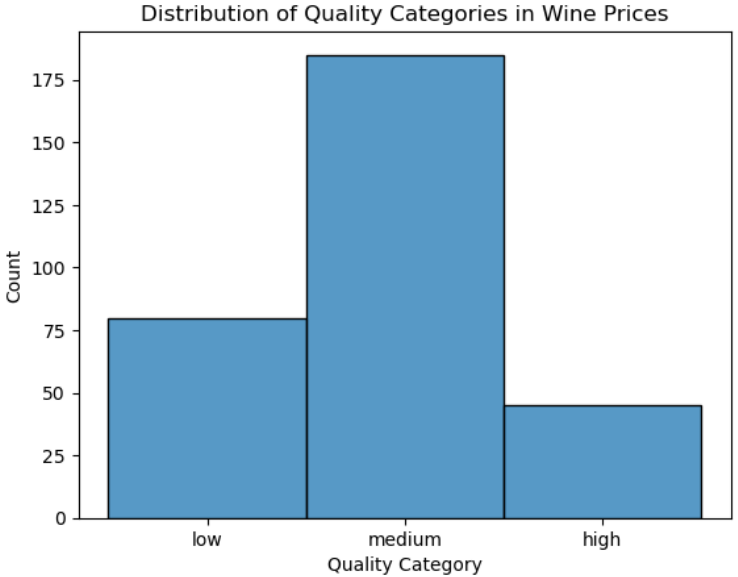
Wine Spectator tasters review wines on the following 100-point scale:

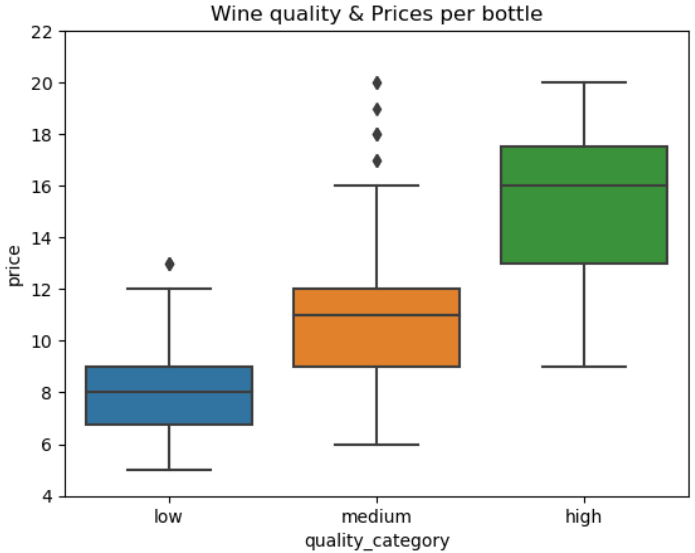
- **95-100** Classic: a great wine
- **90-94** Outstanding: a wine of superior character and style
- **85-89** Very good: a wine with special qualities
- **80-84** Good: a solid, well-made wine
- **75-79** Mediocre: a drinkable wine that may have minor flaws
- **50-74** Not recommended

In out database Vihno Verde wines received **from 80 to 92.**

Source: https://www.winespectator.com/articles/scoring-scale

Predicting the prices





Prices are given in USD, as the database originates from the US market.

The project was implemented as part of the Data Analysis & Machine Learning Bootcamp at Code Academy Berlin
6-8/21 week of the program