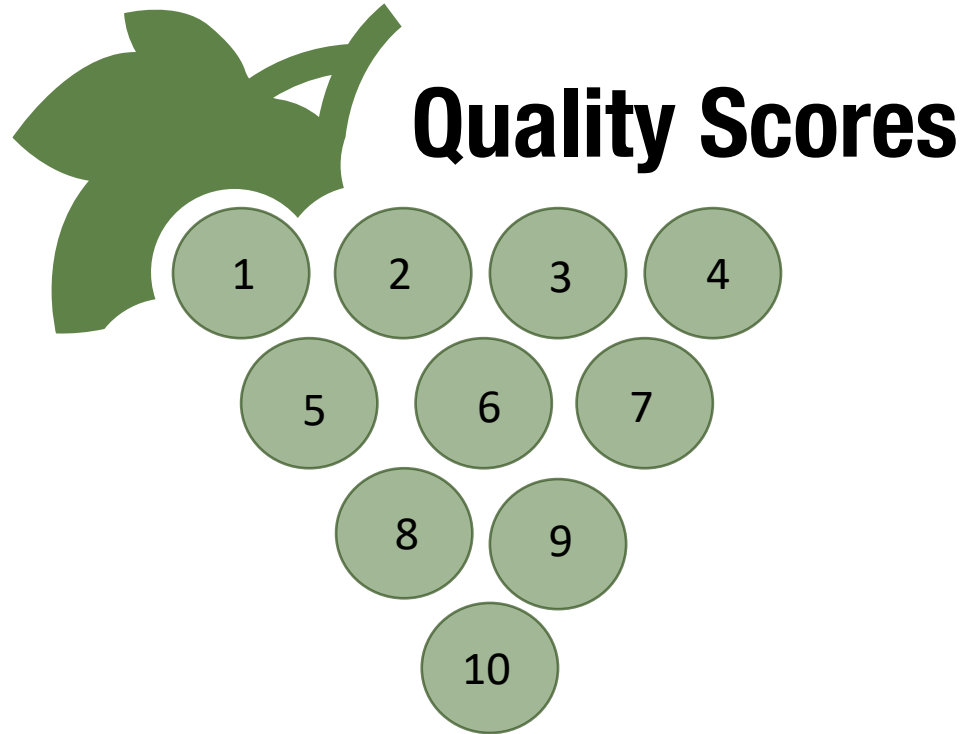


**What makes a  
good white wine?**

Can you predict the basic  
quality of a wine?

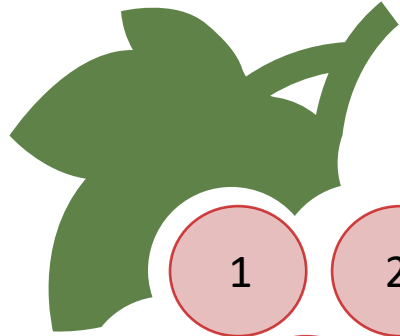
# **What makes a good white wine?**

Do certain properties  
correspond with  
higher quality wines?

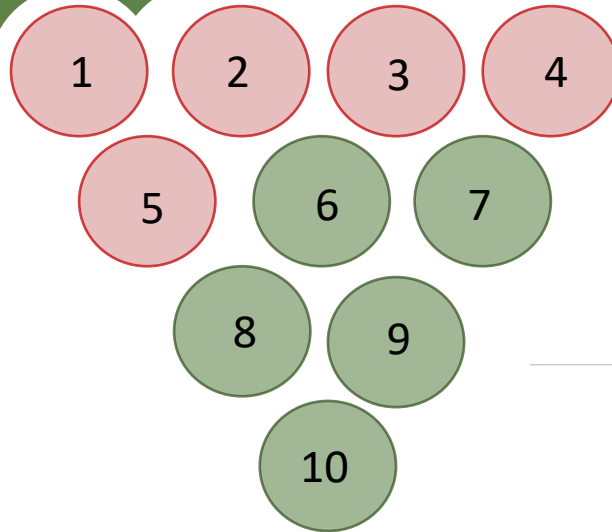


## Quality Scores

Three or more wine experts evaluated the white wines and gave them scores from 1 to 10.



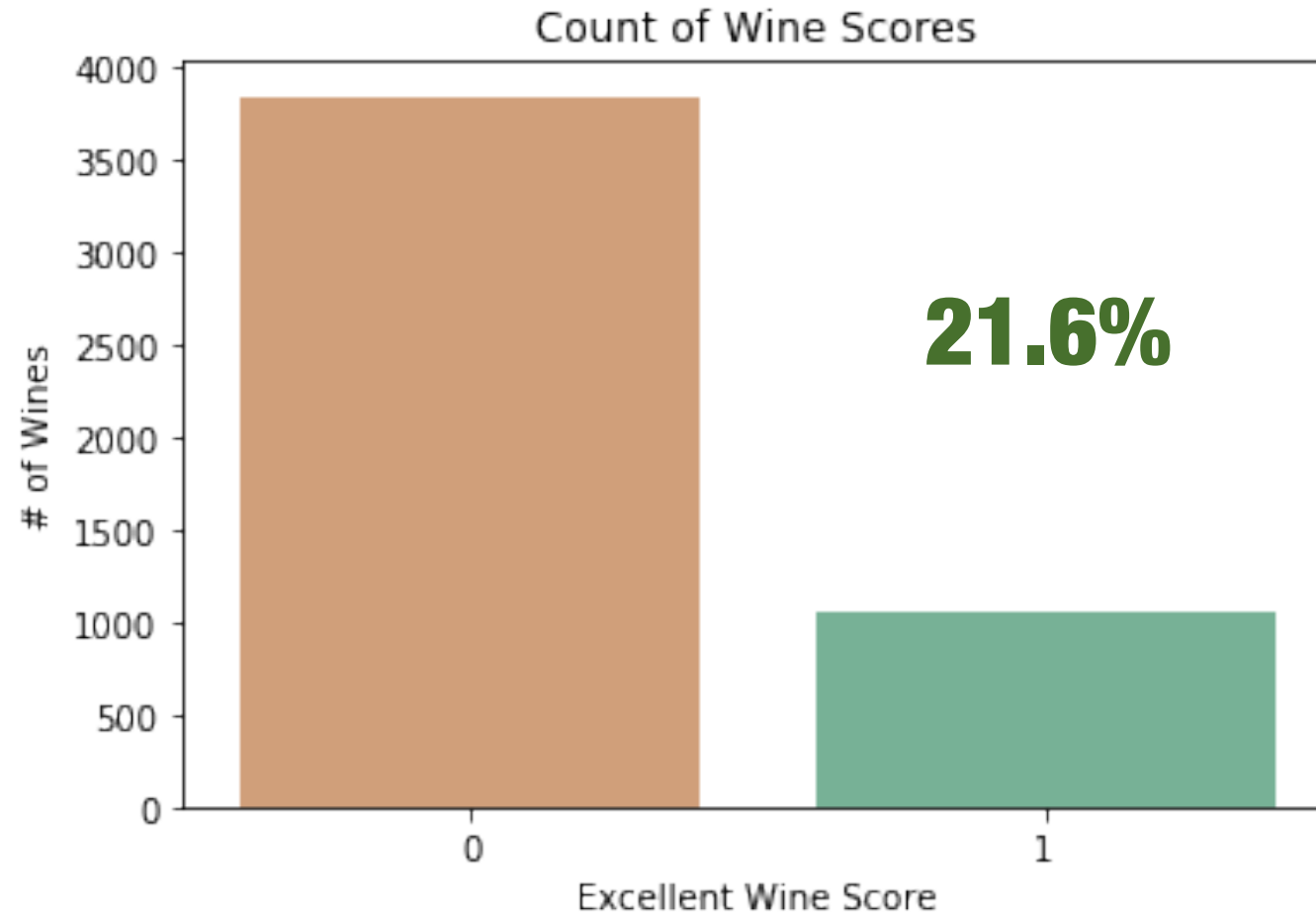
# White Wine Scoring

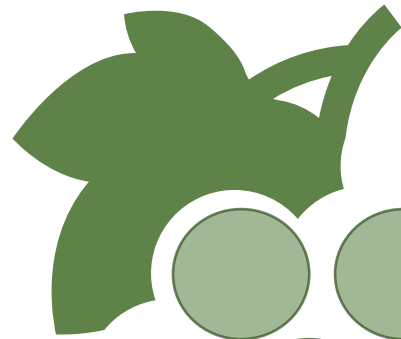


Non-Excellent Wines

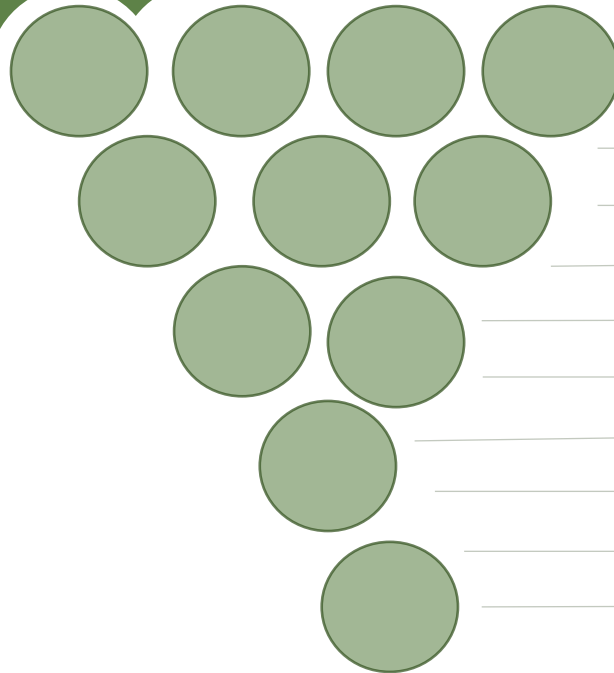
Excellent Wines

The majority of the 4000+ white wines were not excellent.





# White Wine Properties

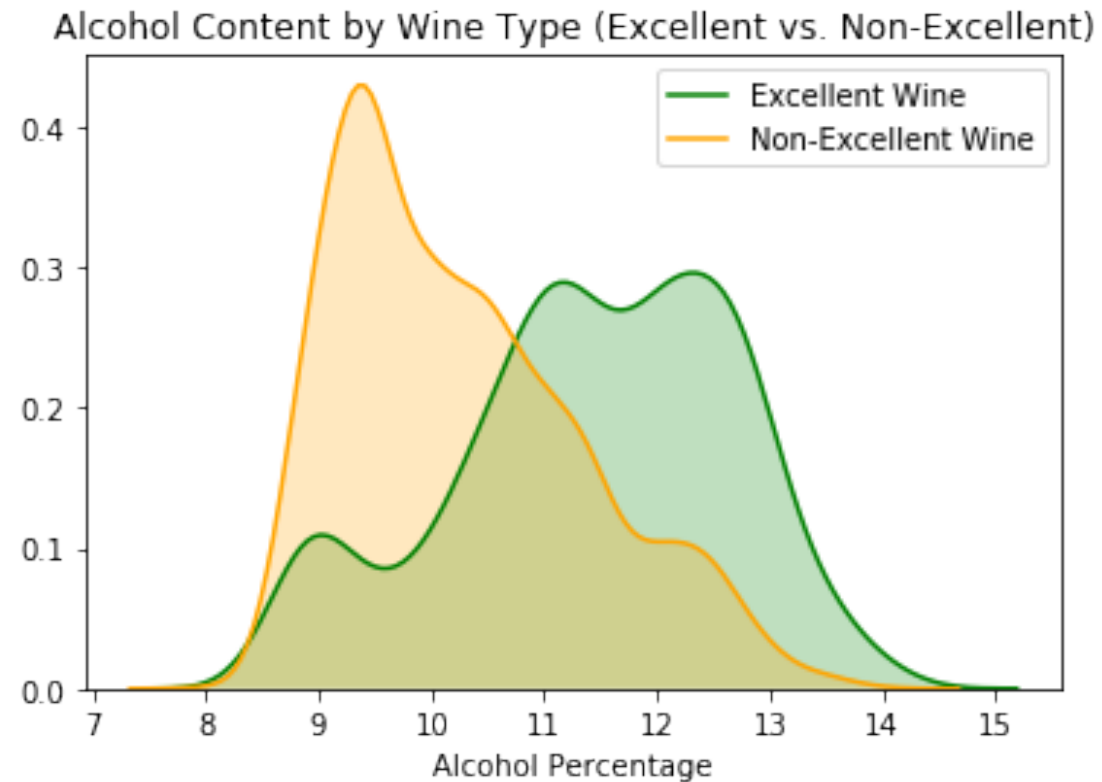


fixed acidity  
volatile acidity  
citric acid  
residual sugar  
chlorides  
free sulfur dioxide  
total sulfur dioxide  
density  
pH  
sulphates  
alcohol

The data on the Portuguese "Vinho Verde" wine and the properties of its varieties.

Link: <https://archive.ics.uci.edu/ml/index.php>

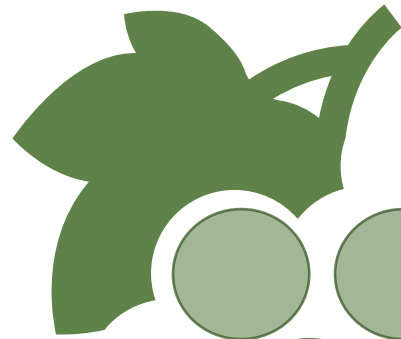
Alcohol content seems to have  
some effect on basic quality score  
of a white wine



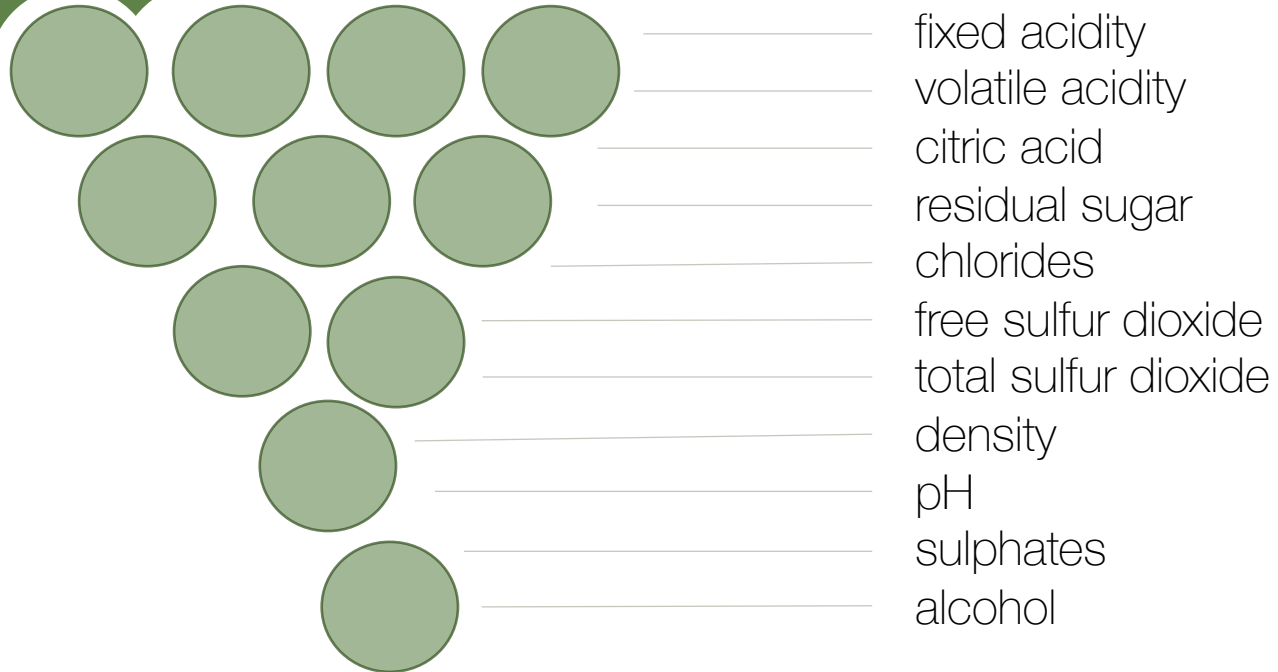
The Random Forest Classifier was most adept at predicting wine quality

Model Type	Score
Random Forest	0.868707
Decision Tree	0.821088
Logistic Regression	0.797959





# White Wine Properties



The data on the Portuguese "Vinho Verde" wine and the properties of its varieties.

Link: <https://archive.ics.uci.edu/ml/index.php>

The data was not balanced

**21.6%**

You can use a different scoring metric:

- **Confusion Matrix:** a table that shows correct predictions and incorrect predictions and determines accuracy. Of the total predictions made, how many excellent and non-excellent wines did the model correctly predict
- **Precision Scoring:** what proportion of the predicted excellent wines were truly excellent
- **Recall Scoring:** what proportion of predicted excellent wines were truly excellent.

You can transform the data:

- Under-sample the non-excellent wines
- Over-sample the excellent wines
- Build synthetic samples