CODE Analytics

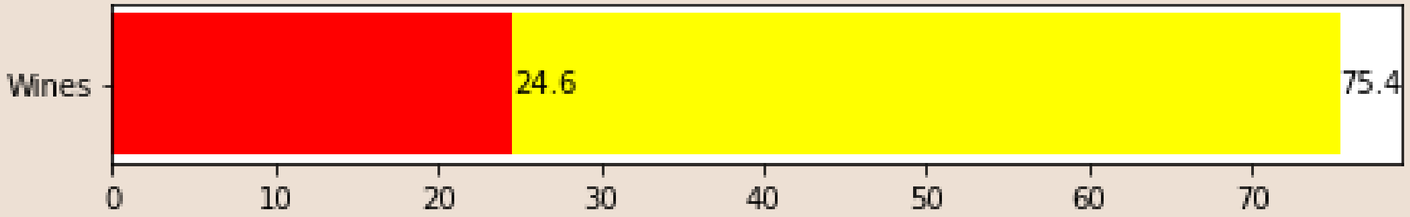
BlueBerry Winery Quality Analytics

Georgios Kosmaras

2022 May 18

Table of Contents

- We were given a dataset containing 6497 wines, of which 1599 were red



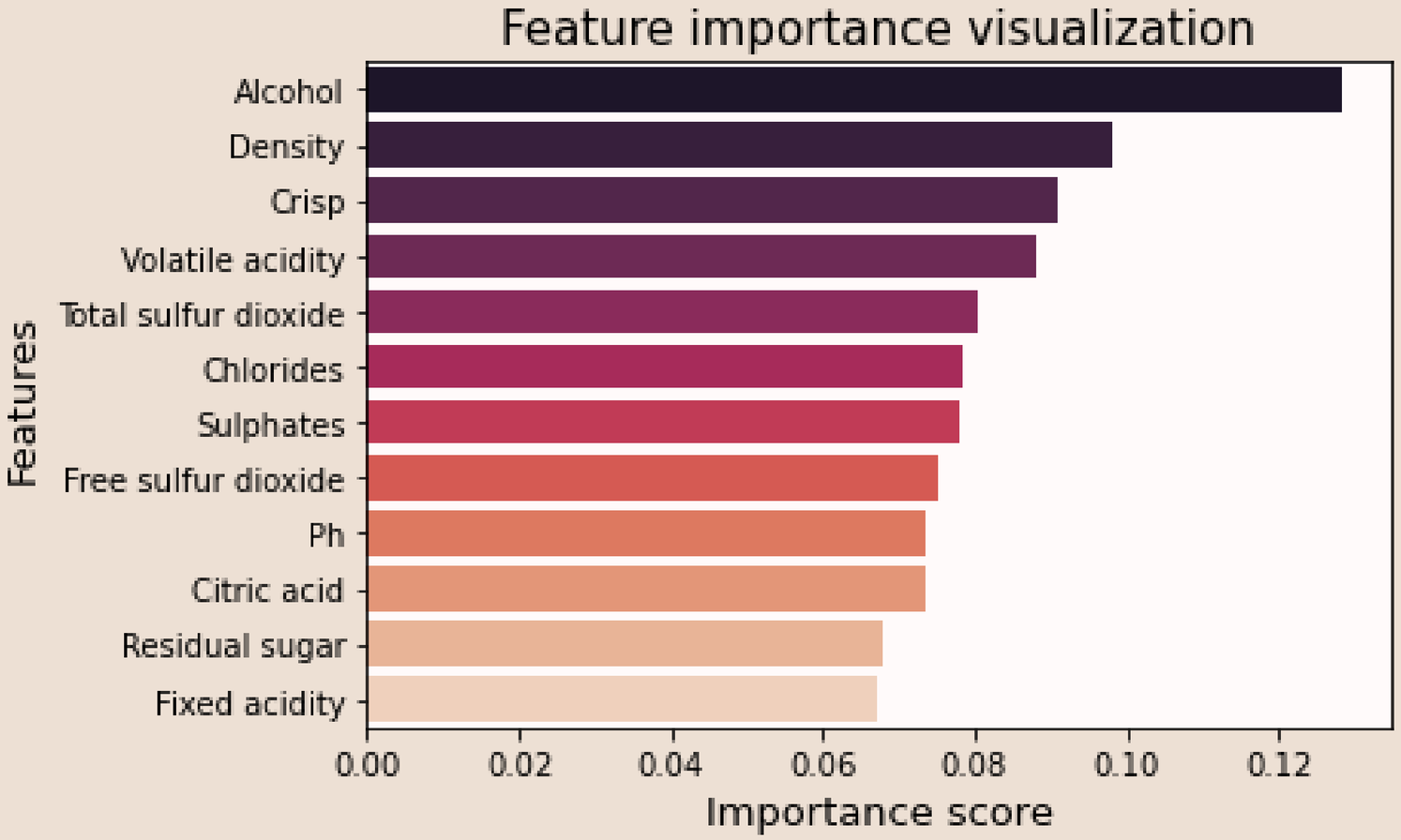
- Each wine has been measured for aciditic, sulphric and alcoholic content, as well as other chemical markers

- A quality mark determined by the average of three seperate tastings was provided as well

- There were no obviously incorrect nor missing values in the dataset

- Find whether any relationship exists between chemical composition and wine quality
- Identification of indicators that could help with determing quality
- Popular perception of Vinho Verde wines and ways to improve it
- Comparative analysis with similarly scaled wine producers, so that improvements and advantages can be recognized and implemented

- **From the numerous features and markers of each wine, we determine which are more likely to impact the quality**



Main indicators for a wine's quality seem to be the following:

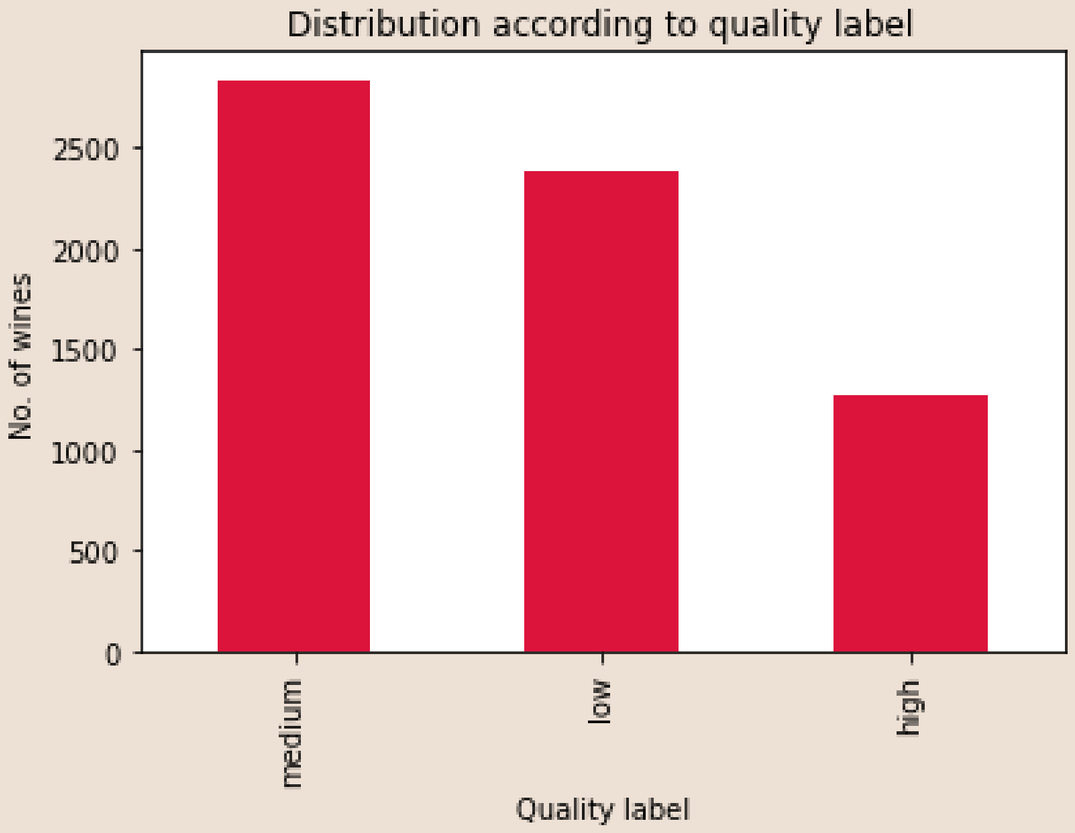
1. Alcohol
2. Density
3. Crisp
4. Volatile acidity
5. Sulfuric content

The selection of the indicators was made with both wine types in mind

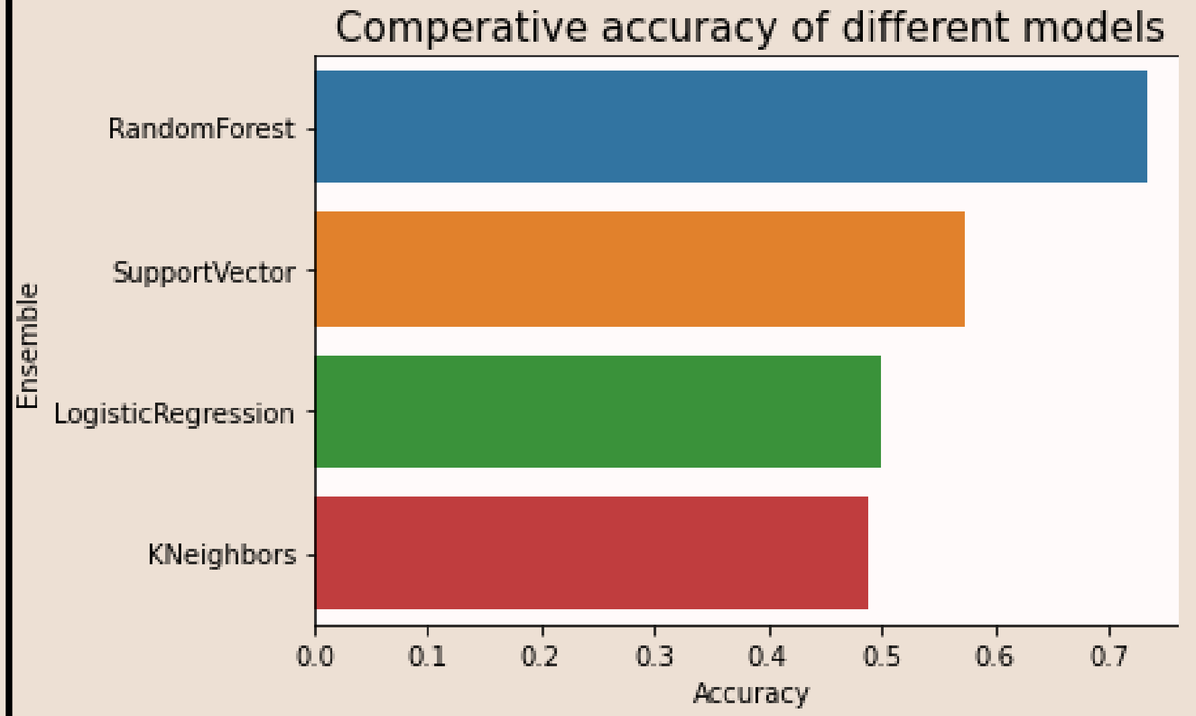
- **We plan to create a predictive model using Machine Learning to help us determine wine's quality**

- **As we are not interested in individual quality scores, we will divide them in categories:**

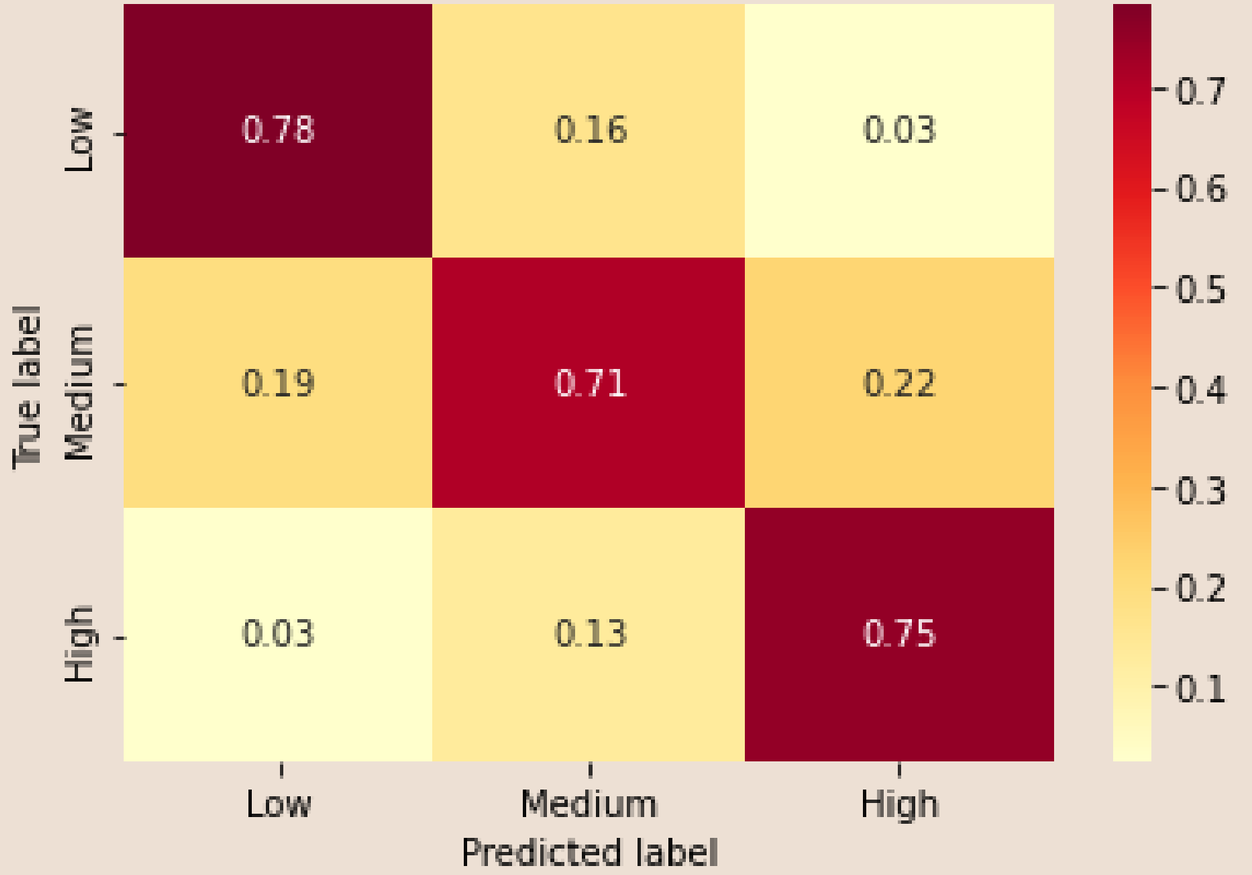
1. Low with score less or equal to 5
2. Medium with score equal to 6
3. High quality with score greater of equal to 7



- For the predictive model we had to choose between the various ML ensembles available
- We went with Random Forest, as it offered the best accuracy
- Determined the best hyperparameters for optimal results and tested it

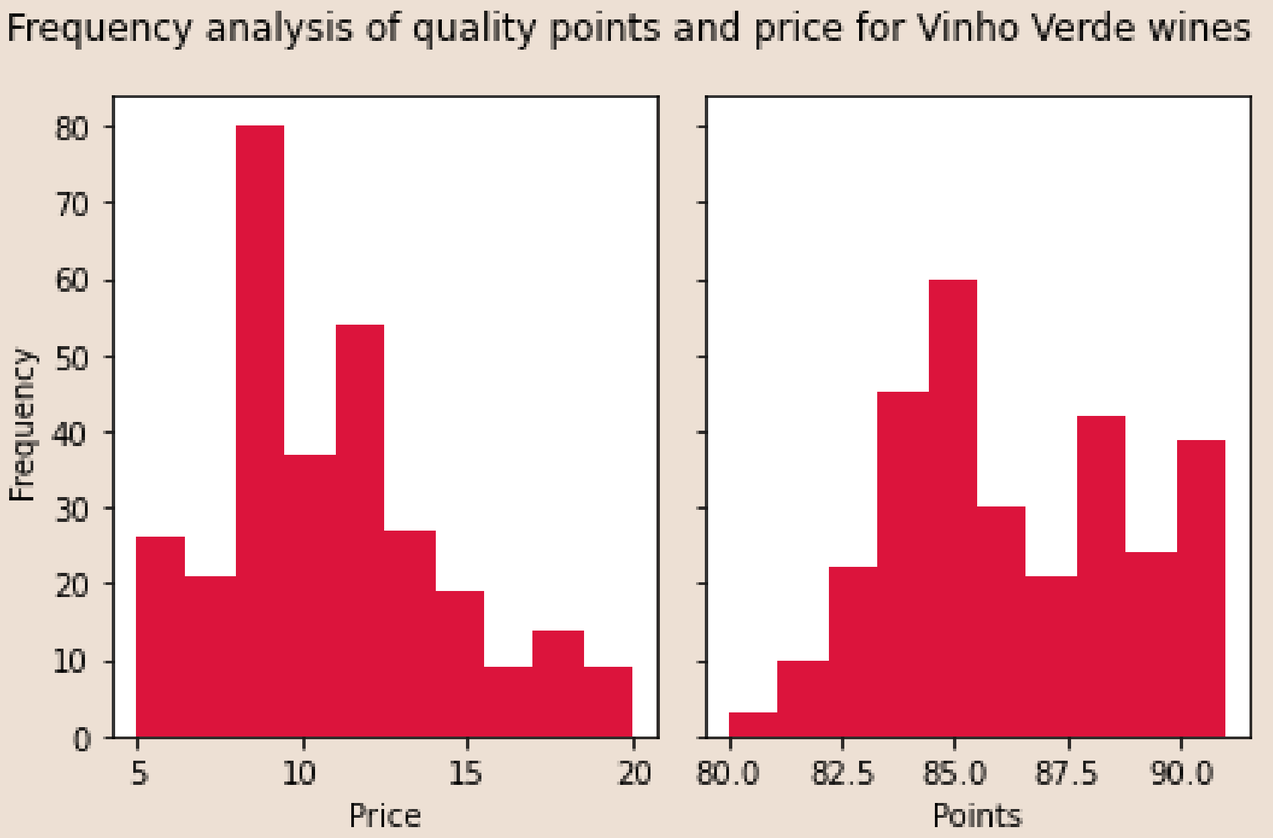


- Our Random Forest predictive model achieved an average of 74% accuracy rate when determining quality label
- 15%-20% in each category were misidentified in a neighboring category
- Gross misidentification was less than 10%

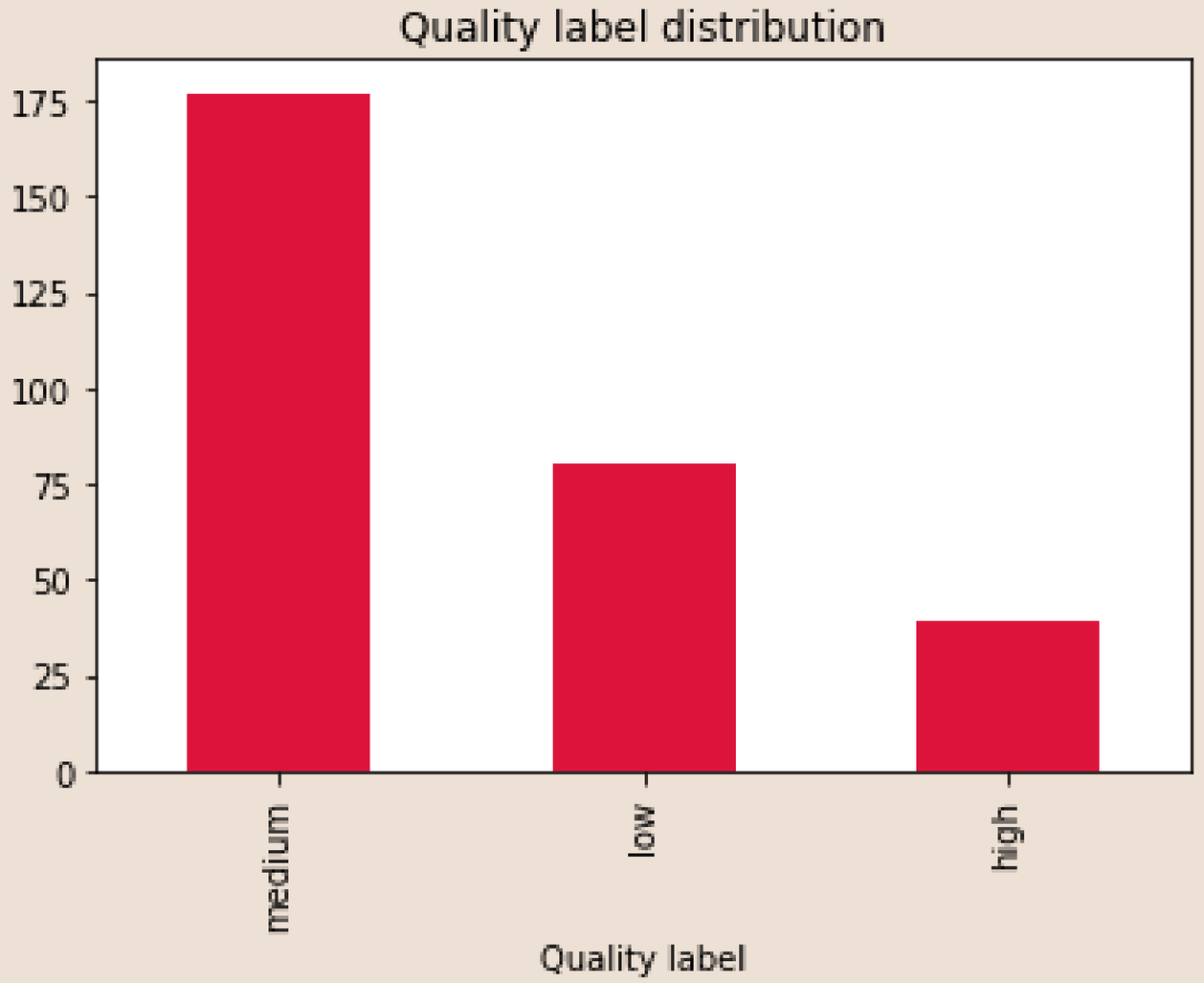


- **For determing the best price for the wines we had another dataset available, with the price and a quality ranking (of a different scale)**

- **As before, most of the wines were in the middle range, with a few exceptions of above or below average.**



- **Despite the diffent quality scales in used in the datasets, we can still assign quality labels**
- **We managed to achieve a similar distribution to the original dataset**
- **This can help in determining the best price for a wine based on its quality**



- **Vinho Verde wines are marketed and sold as affordable wines, with an average price per bottle at 11€ and a low at 8€. The higher end ones are prices at ~15€ per bottle.**
- **Recommended for quick consuption, not ideal for aging and storing**
- **Secong biggest wine sales of Portugal, with annual growth around 4%.**
- **Untapped potential markets in high income Asiatic economies for premium quality wines.**
- **Positive consumer attitude towards Vinho Verde, assosiating them with refreshing, fuzzy and crisp wines.**

Source(s):
https://www.vinhoverde.pt/pt/estatisticas#!
Viticulture Commission of the Vinho Verde Region
https://www.theportugalnews.com/news/2022-01-25/record-year-for-vinho-verde/64832#:~:text=The%20production%20of%20%E2%80%9Cvinho%20verde,(CVRVV)%2C%20Manuel%20Pinheiro.

- Most indicators affecting negatively the quality, such as sulphates and chlorides, are a result of the area and soil salinity, so there is not much to be done there.

- Striving for production of high end wines is not the best approach, as neither the consumers assossiate the particular wines with high end quality, and the price difference is not big enough to support it

- Main concerns are alcoholic content and the "crispness" of the wine, a function of acidity and residual sugar

- Due to short life cycle sulfuric content is not a big concern, as it does not remain in bottle long enough to become an issue

- Main indicators that can be improved in the production line is alcoholic content and viscosity of wine

- Based on the aforementioned findings we can recommended a few courses of action for BlueBerry Winery.
- According to past trends, Vinho Verde wines can be sold at 6.5-9€, 9-12.5€ and 13-18€ for low, medium and high quality respectively
- The main quality indicator that can be influenced is alcoholic content
- High end wines have higher demand in countries like Japan and Singapore
- Medium and low range ones are consumed at the national market and exported
- The quality label can be deduced with reasonable certainty using our predictive model, making classification and designation for different sale point of wines much easier