This dataset comes from the University of California, Irvine's massive archive of Machine Learning Datasets.

The Wine Quality Dataset that focuses on the Portuguese "Vinho Verde" Wine. The datasets for white wine has 4898 entries over 12 features. The features are:

* 1 - fixed acidity
* 2 - volatile acidity
* 3 - citric acid
* 4 - residual sugar
* 5 - chlorides
* 6 - free sulfur dioxide
* 7 - total sulfur dioxide
* 8 - density
* 9 - pH
* 10 - sulphates
* 11 - alcohol
* 12 - quality (score between 0 and 10)

The quality of the wine, the dependent variable, is a classification marking on the label (range between 1 -10) and is identified by patrons.

`From our data set, we are interested in determining the following:

* What can you comment on the variables?
* Do we need to do outlier treatment? What is the most apt method for this?
* Does it make a business case to predict the quality of the wine?
* Are we able to classify the type of wine given the predictors?
* Which among the below methods will be your recommendation? Does this make business sense? how will this help in the real world?

We will use the following classification methods for this:

1. KNN
2. Decision trees - CART
3. Naïve Bayes