**Problem Statement**

As a wine lover, wouldn't it be ideal to know which property affect wine quality? What about knowing how else you could increase the "quality" of your wine? As a wine drinker, wouldn't it be insightful to know which characteristics to look out for in a wine when you wish to buy a good wine or as an investment? Multiple machine learning algorithms are used to determine which physiochemical properties have impact on a wine’s quality.

Based on the above dataset, a predictive model will be created. It will then conclude with recommendations on which attribute affect wine quality. (Fixed acidity, alcohol etc..)

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. For more details, consult: [Web Link] or the reference [Cortez et al., 2009].

Dataset from: https://archive.ics.uci.edu/ml/datasets/wine+quality

**Executive Summary**

During the EDA, the relationship between each variable and wine quality was examined closely. The dataset collected was rather extensive and was useful to help identify general trends

**Findings:**

From correlation plot we can see there are some highly correlated variables. Example like fixed acidity, citric acid will affect the pH value of the wine.

There is a positive correlation of 0.67 between the “fixed acidity” variable and the variables “citric acid” and “density”.

Hence, “fixed acidity” variable increases the “citric acid” will also increase. This will also apply to the relationship between “free sulfur dioxide” and “total sulfur dioxide” variables.

Variables like “fixed acidity” and “pH” have a negative linear correlation of -0.68. This relationship indicates that when the fixed acidity of the wine increases, the pH value of the wine decreases.

**Model Evaluation and Selection**

We used a Stochastic Gradient Decent, Support Vector Classifier, Logistic Regression, Random Forest to help in our prediction.

**Results, conclusion and recommendations**

From the Random Forest Classifier, we can see which feature is important in the prediction. It will help us in making decision and for wine makers to take into consideration when making wines.

Top 3 variables that affect wine quality score.

1. Alcohol 2) Sulphates 3) Volatile acidity

We also the key factors that determine the quality of the wine score. While wine quality is a very subjective measure as everyone’s taste is different. We can use some of the variables to help tune or consider when making wine. This is by no means a sure win way to get high “wine quality” score, but rather shows us which factor affects the wine quality. Ultimately, wine is a mix of all the features. However, if the goal is to predict wine with more accuracy, more years of data could be used to train the model.