Group Members:

* Stephen DeFerrari
* Patrick Marino
* Kavita Kalsy
* Hatice Erdogan

Dataset: <https://archive.ics.uci.edu/dataset/186/wine+quality>

Our group’s intention is to use the wine quality dataset from the UCI machine learning repository. This dataset has 4898 datapoints for us to use and has 12 features and 1 target variable. The target variable is the “quality” of the wine and is an integer from 0 to 10. We want to convert this column into a category by splitting it up. Our initial split will be:

* 0-2:Poor
* 3-5:Fair
* 6-8:Good
* 9-10:Excellent.

We will perform EDA to figure out the best way to split this column up based on the distribution of scores but our goal is to end up with 3-4 categories. Our EDA will also focus on some preliminary feature selection because our initial goal is to utilize every feature present but may narrow them down based on what our EDA says. We will ultimately look to use the following models:

* QDA/LDA
* Logistic Regression
* Decision Trees
* Random Forest

With each of these models, we intend to utilize techniques such as cross-validation to ensure we are properly testing each model before deciding which is the “best” one.