Project Update #1

BY John Christman, Laura Larregui, and Angela Rivera

# Dataset

Wine Quality dataset is publicly available for research in the University of California, Irvine Machine Learning repository created by Paulo Cortez (Univ. Minho), Antonio Cerdeira, Fernando Almeida, Telmo Matos and José Reis (CVRVV) in 2009.

**Citation**: P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.

**Available at** https://archive.ics.uci.edu/ml/datasets/Wine+Quality

Goals

Main Goal: To create a model that predicts the rating of a wine.

Secondary Goals:

* Understand which variables are responsible for the quality of the wine.
  + What is the minimum set of properties and their values that define a high-quality wine?
  + What are properties which make wine second-rate?

# Plans for Next Update

* Create a table with the variables and their description
* Finding out if there are any missing values/ “Clean up” dataset
* Look if there is a relationship between Wine Quality and Alcohol content in the wine
* Create a correlation matrix