CUNY Baruch College

Project Proposal

Group Submission

Yula Ko

Farzana Manjra

Natalia Gomez

Victoria Wayda

JiaRui (Jessica) Shao

STA 3920 NFA

Yuanfeng Cai

23 October 2018

Team Name: Data Diggers

Data Set: Red Wine Quality

Source:

The dataset has 12 variables and 1,599 observations. The predictor variables include fixed acidity, volatile acidity (amount of acetic acid - too much acetic acid can have an unpleasant taste), citric acid (can add a fresh taste), residual sugars (amount of sugar after fermentation stops), chlorides (amount of salt), free sulfur dioxide (prevents oxidation), total sulfur dioxide (can affect the taste), density (depending on alcohol and sugar content), pH (how acidic or basic), sulphates (contributes to SO2 levels), and alcohol (the percentage of alcoholic content). We will use these variables to predict the quality of wine using logistic regression. The main question we intend to address is "To determine which physicochemical properties make a wine 'good'." Our target audience could be anyone else who is looking to determine the quality of wine based on these variables.