Wine Quality: To Booze or not to Booze

A Study By LeRoy, Rob, and Uche

Why we chose wine

Our group wanted to look at data that would lead to clear differences and conclusions in questions that are not always easy to answer. For this task we decided to study the effect of different variables like acidity and alcohol content on people's perception of wine quality.

Data Source

Wine Quality dataset is public 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 Jose Reis (CVRVV) in 2009. This repository has two datasets of red and white wine samples which consists of inputs includes objective tests (e.g. PH values) and the output is based on sensory data (median of at least 3 evaluations made by wine experts). Each expert graded the wine quality between 0 (very bad) and 10 (very excellent).

Questions

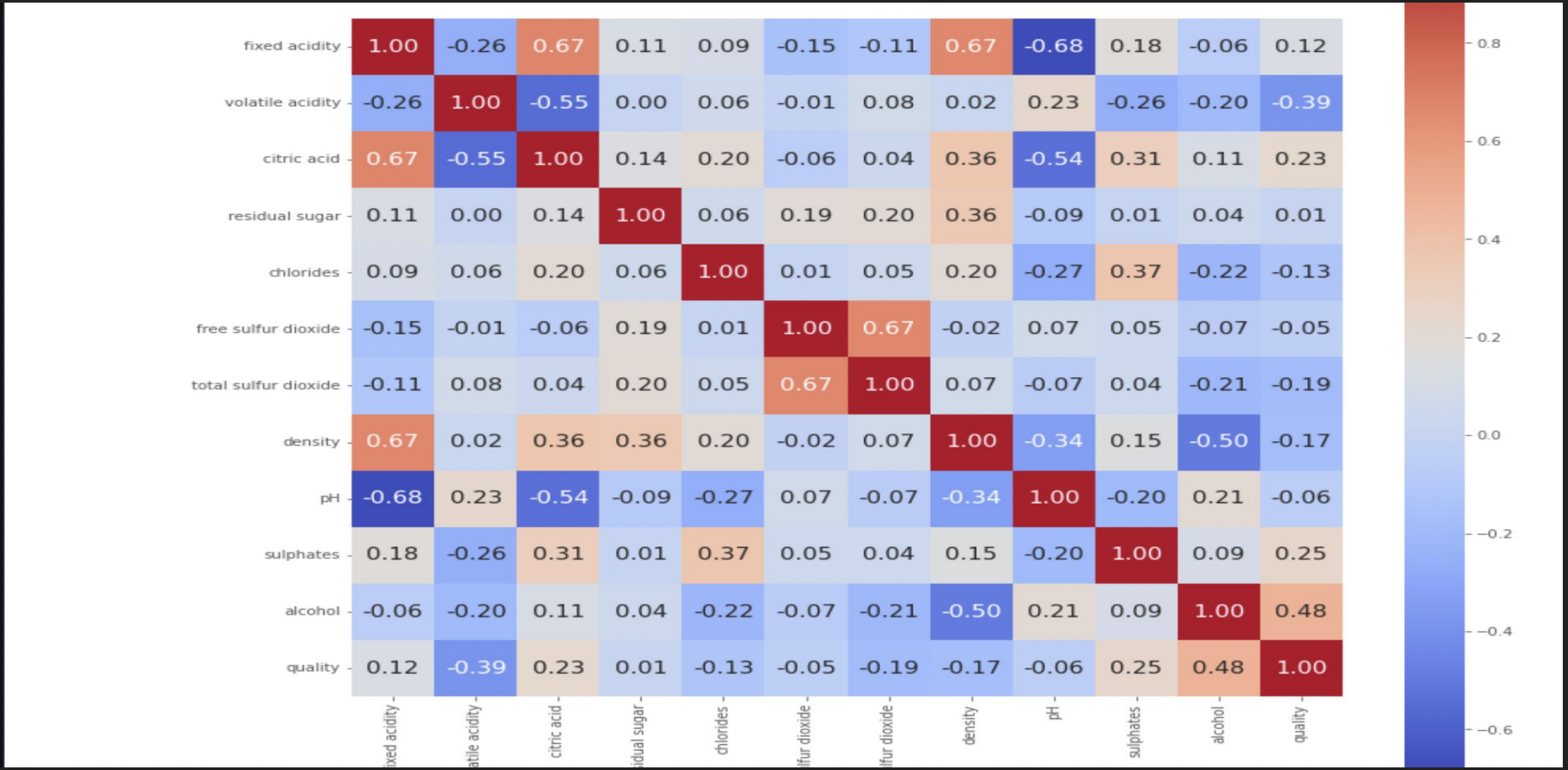
- ➔ Is it possible to predict the quality of a wine before we taste it?
- ➔ Can we use pH to predict quality?
- ➔ What other variables might have an effect on quality?

Data

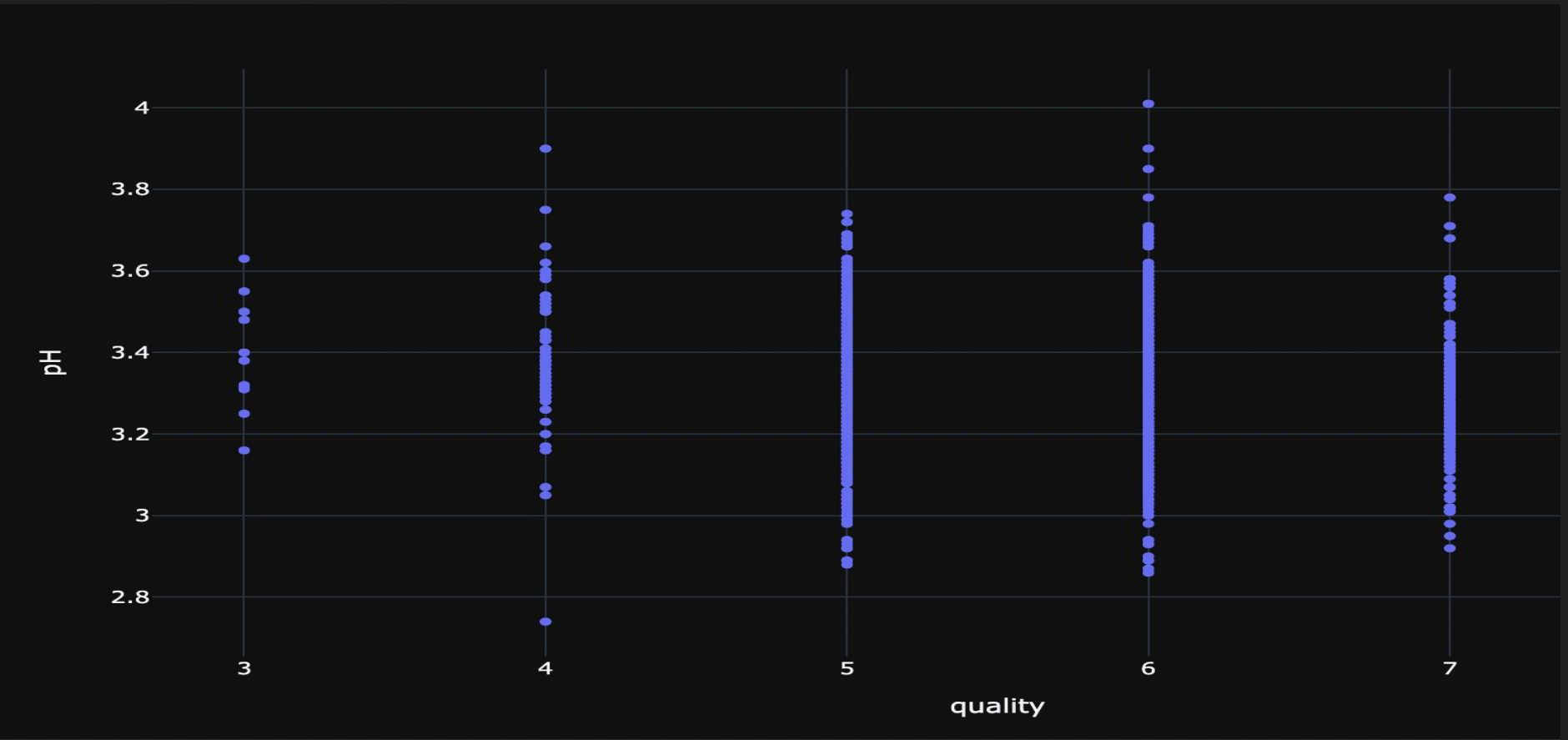
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

The Red wine dataset has 1599 number of observations and 12 variables

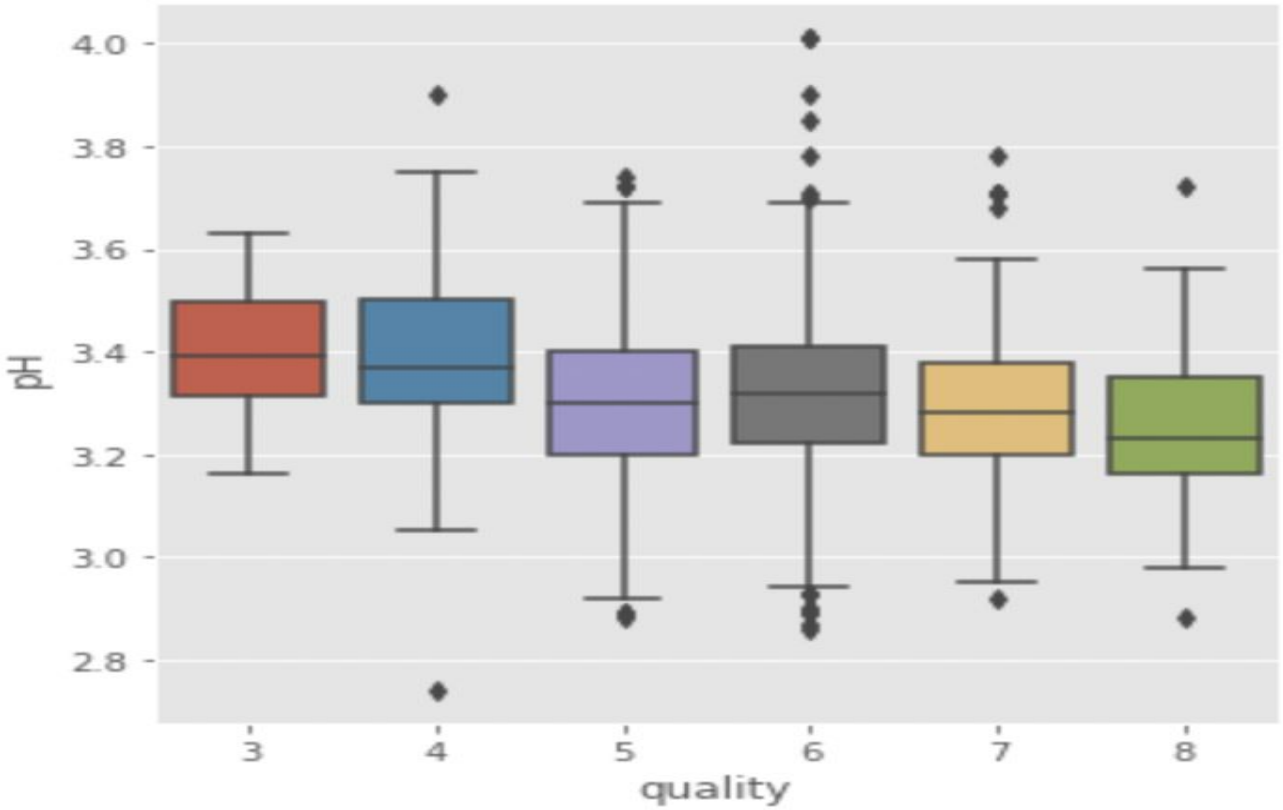
Heat Map


fixed acidity 1.00 -0.26 0.67 0.11 0.09 -0.15 -0.11 0.67 -0.68 0.18 -0.06 0.12
volatile acidity -0.26 1.00 -0.55 0.00 0.06 -0.01 0.08 0.02 0.23 -0.26 -0.20 -0.39
citric acid 0.67 -0.55 1.00 0.14 0.20 -0.06 0.04 0.36 -0.54 0.31 0.11 0.23
residual sugar 0.11 0.00 0.14 1.00 0.06 0.19 0.20 0.36 -0.09 0.01 0.04 0.01
chlorides 0.09 0.06 0.20 0.06 1.00 0.01 0.05 0.20 -0.27 0.37 -0.22 -0.13
free sulfur dioxide -0.15 -0.01 -0.06 0.19 0.01 1.00 0.67 -0.02 0.07 0.05 -0.07 -0.05
total sulfur dioxide -0.11 0.08 0.04 0.20 0.05 0.67 1.00 0.07 -0.07 0.04 -0.21 -0.19
density 0.67 0.02 0.36 0.36 0.20 -0.02 0.07 1.00 -0.34 0.15 -0.50 -0.17
pH -0.68 0.23 -0.54 -0.09 -0.27 0.07 -0.07 -0.34 1.00 -0.20 0.21 -0.06
sulphates 0.18 -0.26 0.31 0.01 0.37 0.05 0.04 0.15 -0.20 1.00 0.09 0.25
alcohol -0.06 -0.20 0.11 0.04 -0.22 -0.07 -0.21 -0.50 0.21 0.09 1.00 0.48
quality 0.12 -0.39 0.23 0.01 -0.13 -0.05 -0.19 -0.17 -0.06 0.25 0.48 1.00
citric acid
density
pH
sulphates
alcohol
quality
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

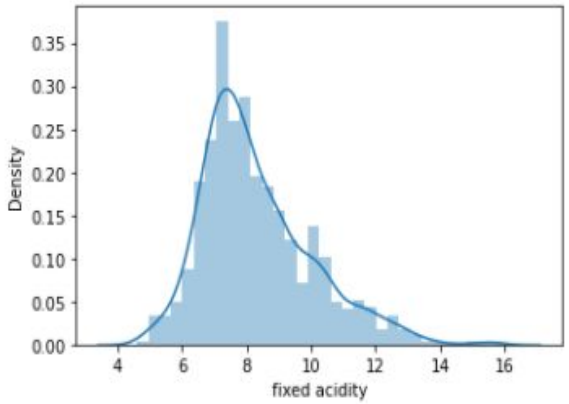
pH Against Quality

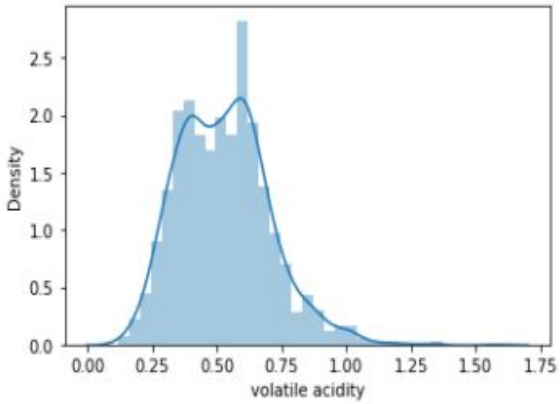


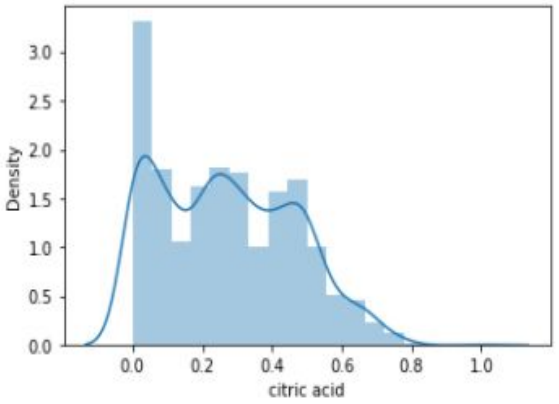
Box Plot

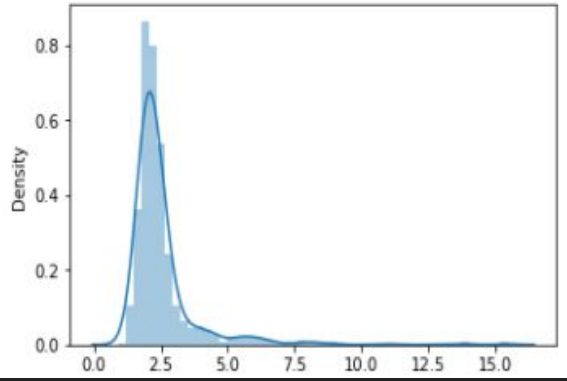


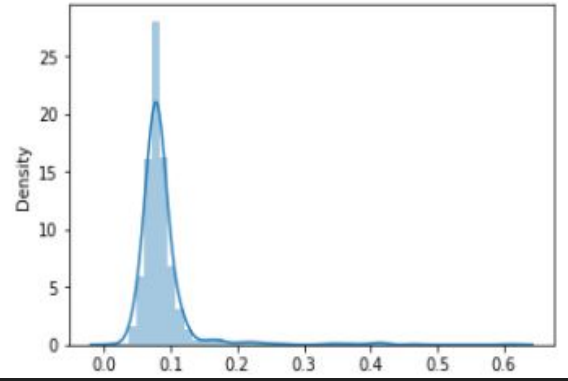
Histogram of the Red wine Predictors

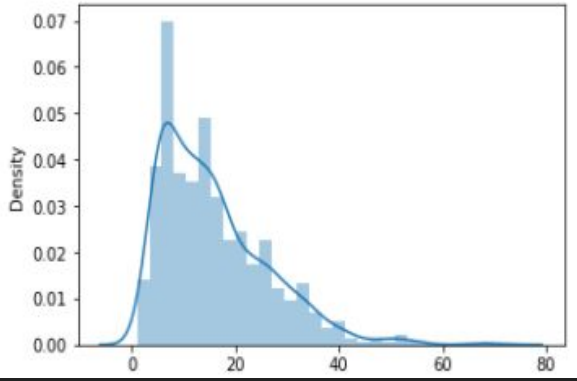


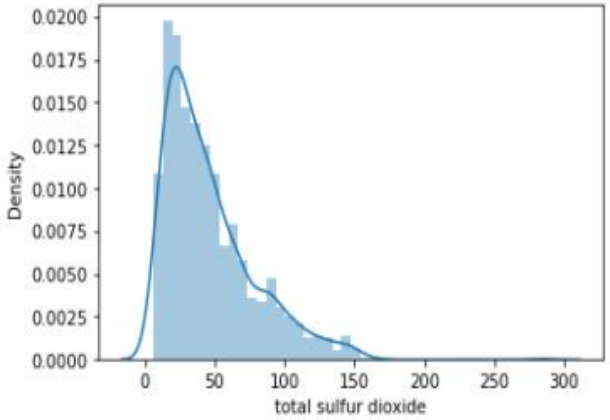


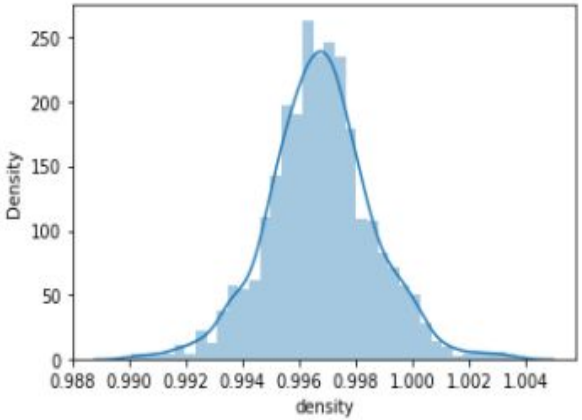


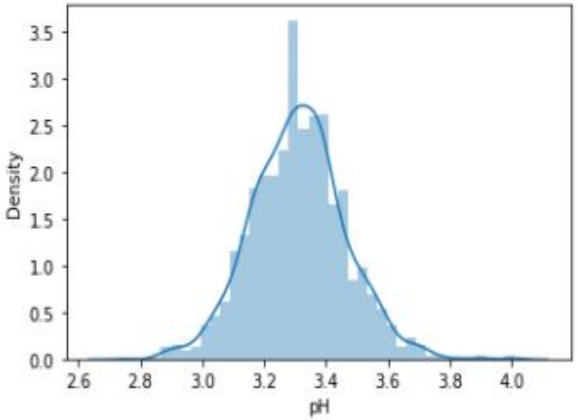


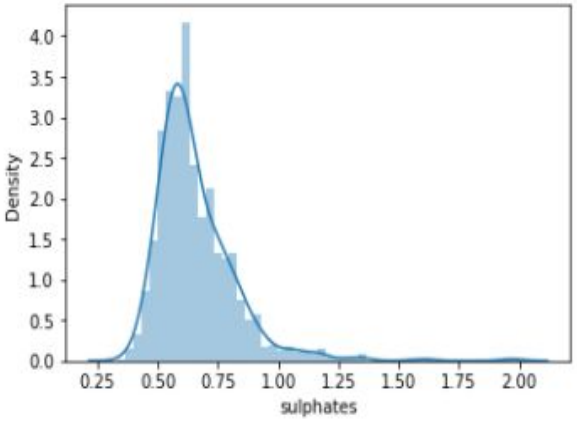


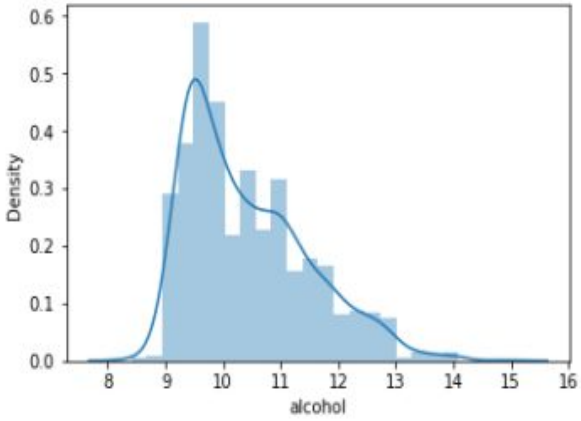


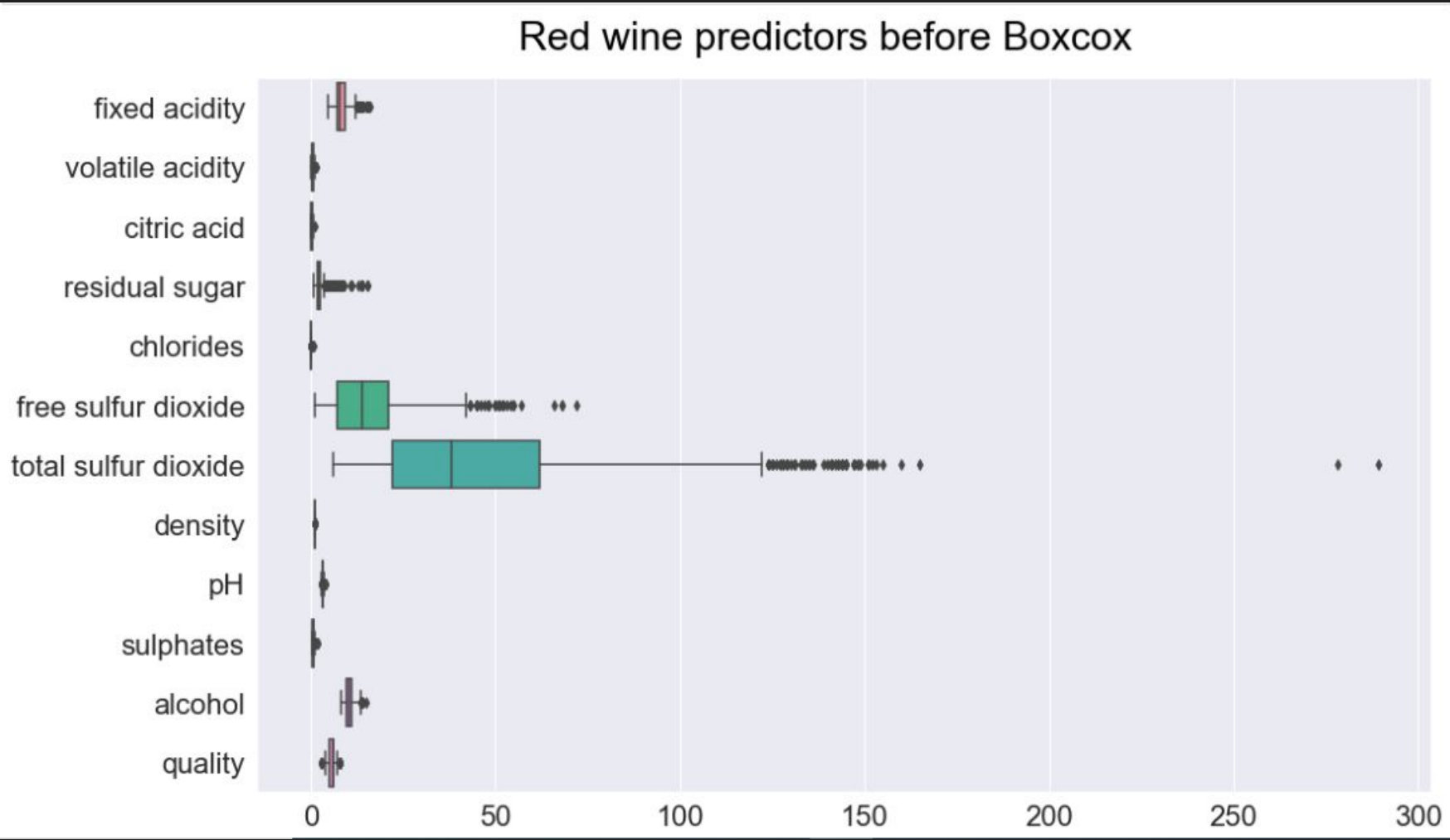

Density
0.0200
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
0
50
100
150
200
250
300
total sulfur dioxide

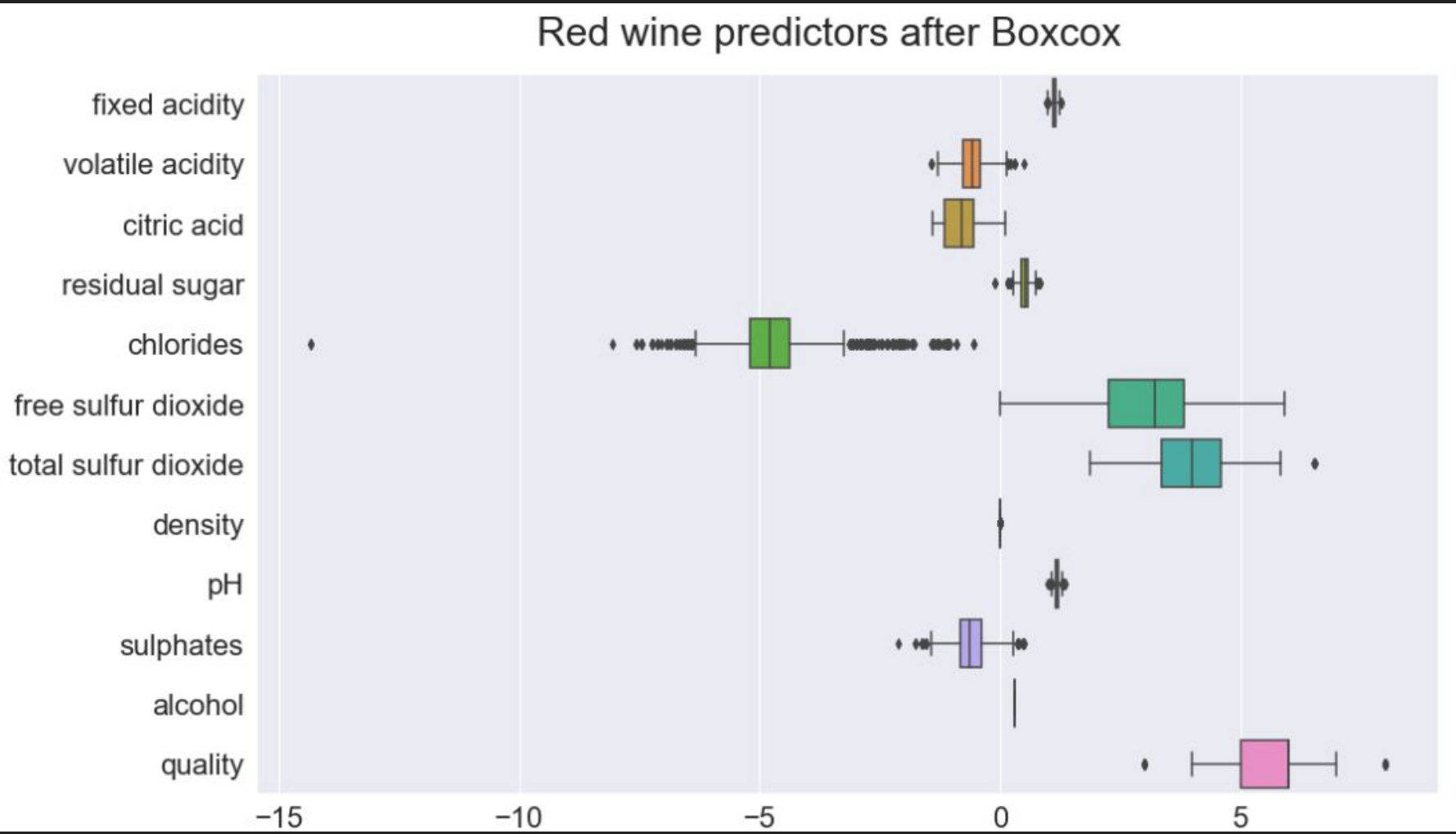

Density
250
200
150
100
50
0
0.988
0.990
0.992
0.994
0.996
0.998
1.000
1.002
1.004
density


Density
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2.6
2.8
3.0
3.2
3.4
3.6
3.8
4.0
pH


Density
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
sulphates


Density
0.6
0.5
0.4
0.3
0.2
0.1
0.0
8
9
10
11
12
13
14
15
16
alcohol


Red wine predictors before Boxcox
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality
0
50
100
150
200
250
300


Red wine predictors after Boxcox
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
free sulfur dioxide
total sulfur dioxide
density
pH
sulphates
alcohol
quality
−15
−10
−5
0
5

Results

- ➔ Through testing we found that certain variables can help predict quality.
- ➔ pH can be used to predict quality in that there is an inverse relationship.
- ➔ Best variable to measure quality with and why its sulfur dioxide.

Recommendations and Regrets