Red Wine Quality

By: Jordan Chong

Introduction

01

PROBLEMS AND STAKEHOLDERS

02

DATA ACQUISITION AND DATA CLEANING

03

DATA ANALYSIS AND PREDICTIVE MODELING

Problems and Stakeholders

- Knowing what does the wine contain to predict the quality of wine
  - For sommeliers, wine makers, and wine aficionados

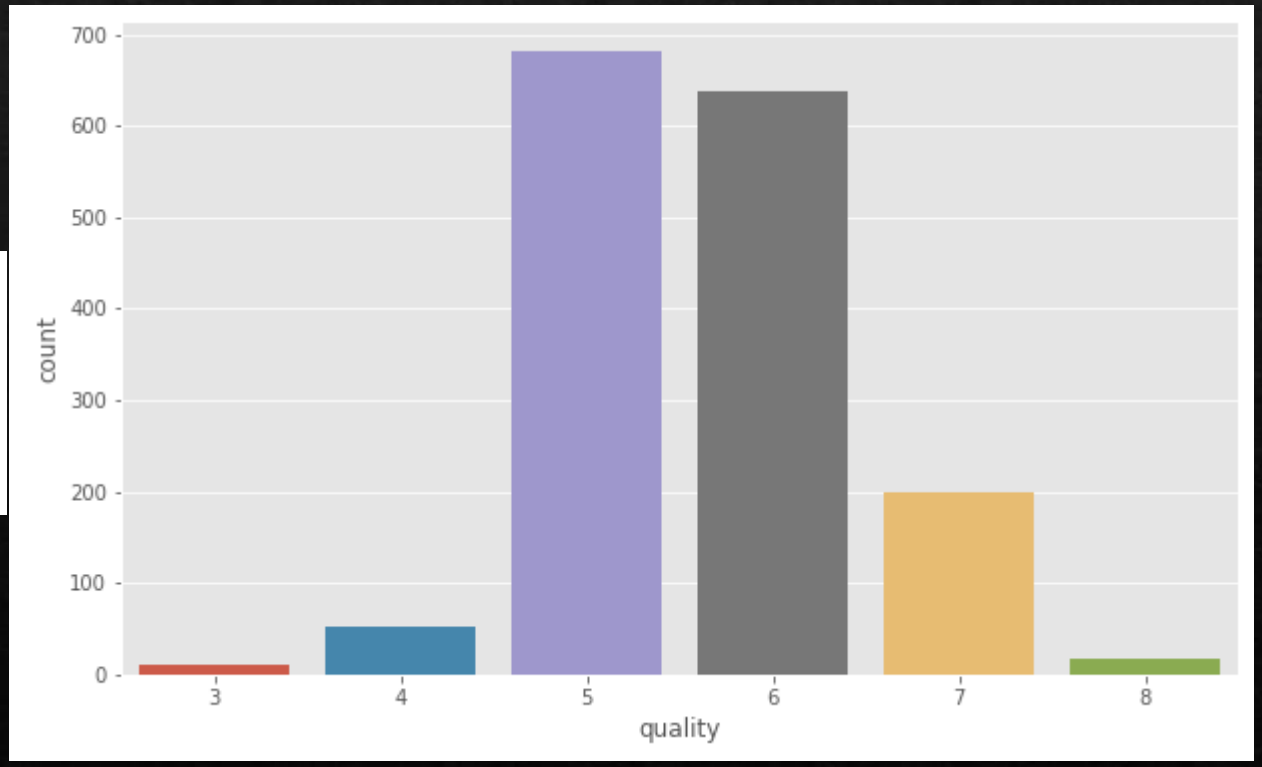
Data Acquisition and Data Cleaning

- Red wine dataset from Kaggle, named “Red Wine Quality” from UCI Machine Learning
- Not much to clean or wrangle since the data has been cleaned for machine learning

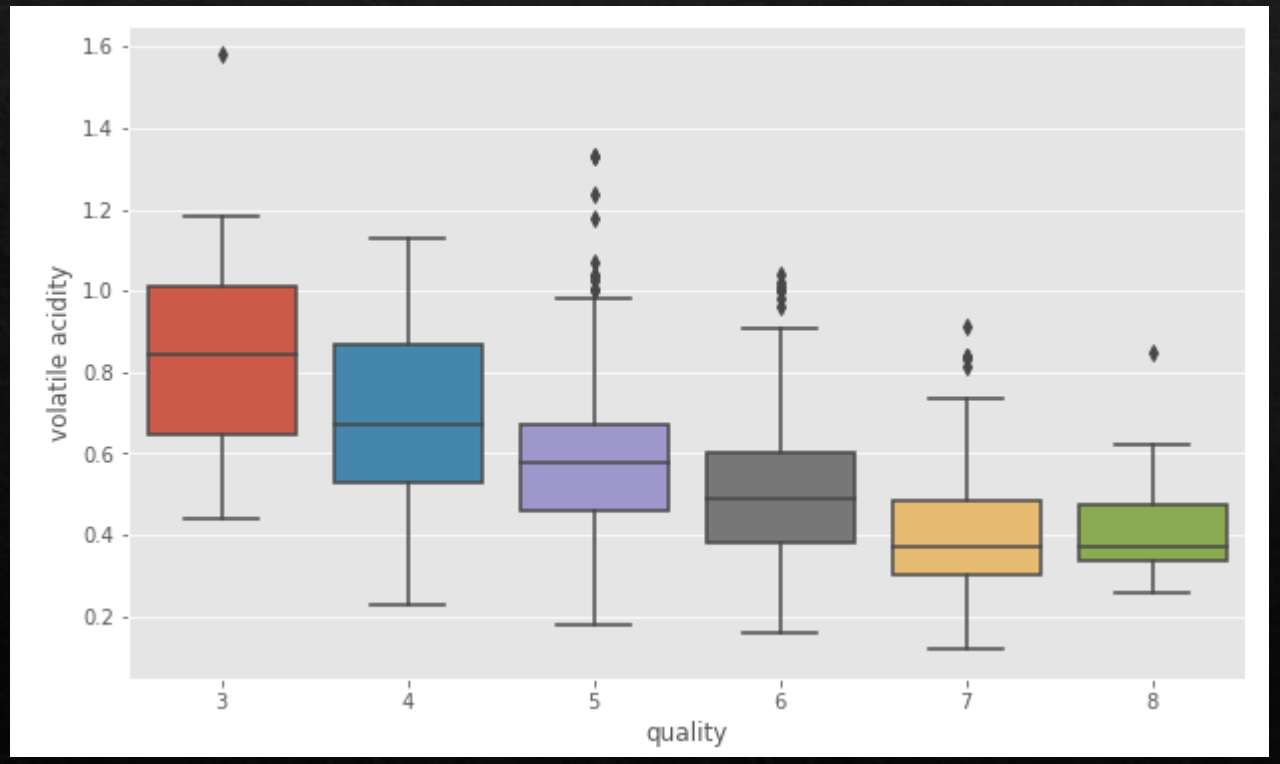
```
No of empty cells:
 fixed acidity           0
volatile acidity         0
citric acid              0
residual sugar           0
chlorides                0
free sulfur dioxide      0
total sulfur dioxide     0
density                  0
pH                       0
sulphates                0
alcohol                  0
quality                  0
dtype: int64
```

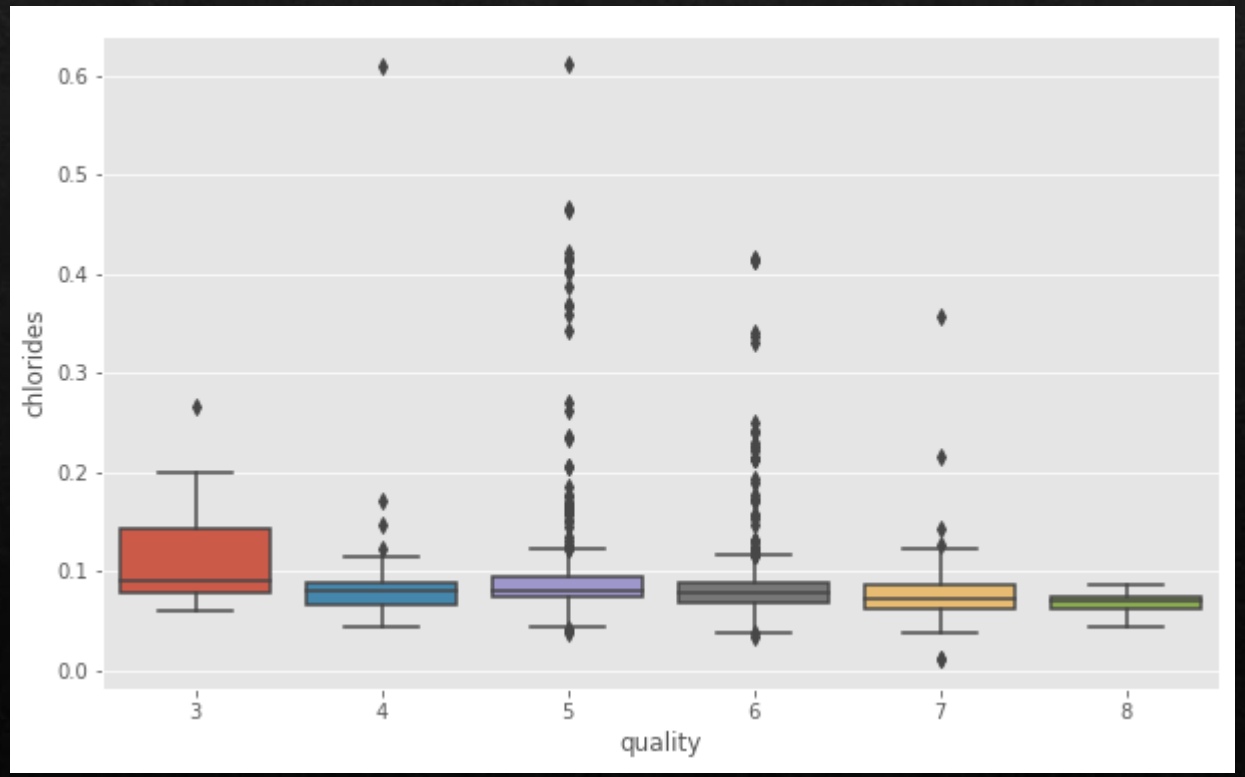
Data Analysis – Wine Amount and Wine Quality

```
3      10
8      18
4      53
7     199
6     638
5     681
Name: quality, dtype: int64
```

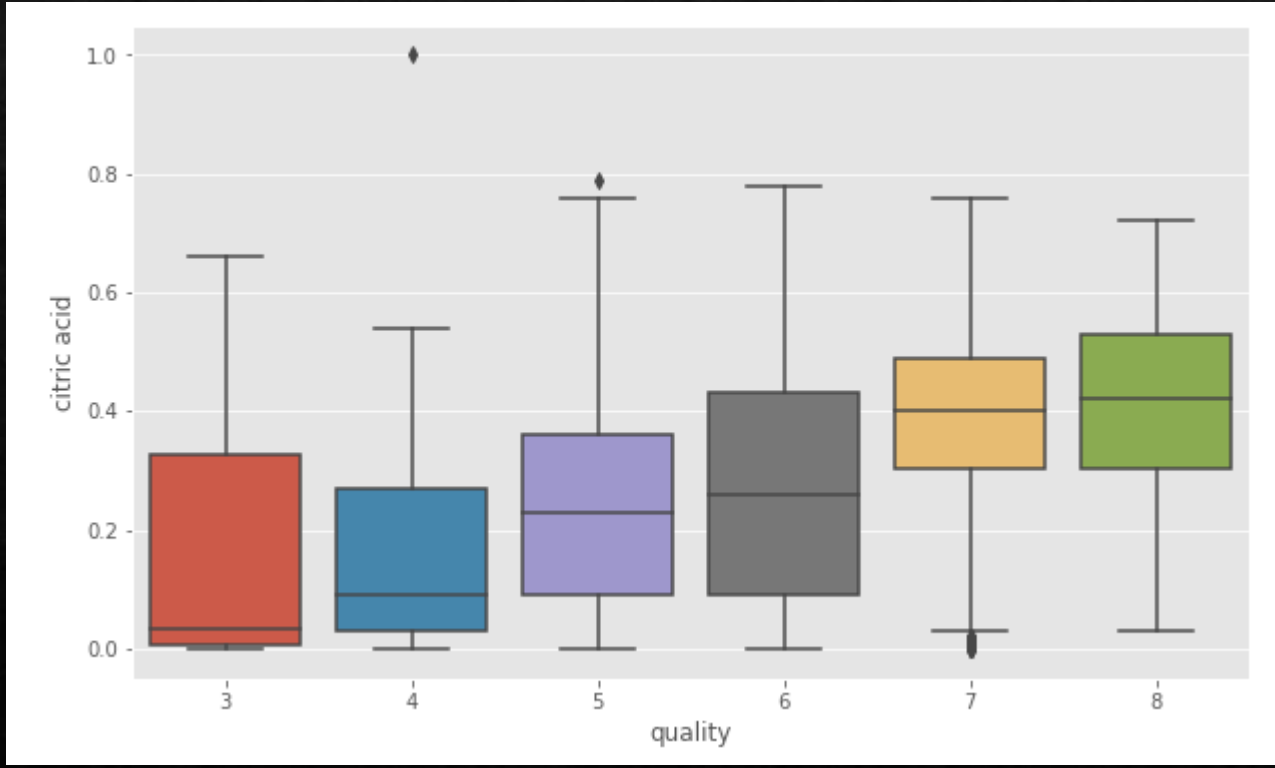


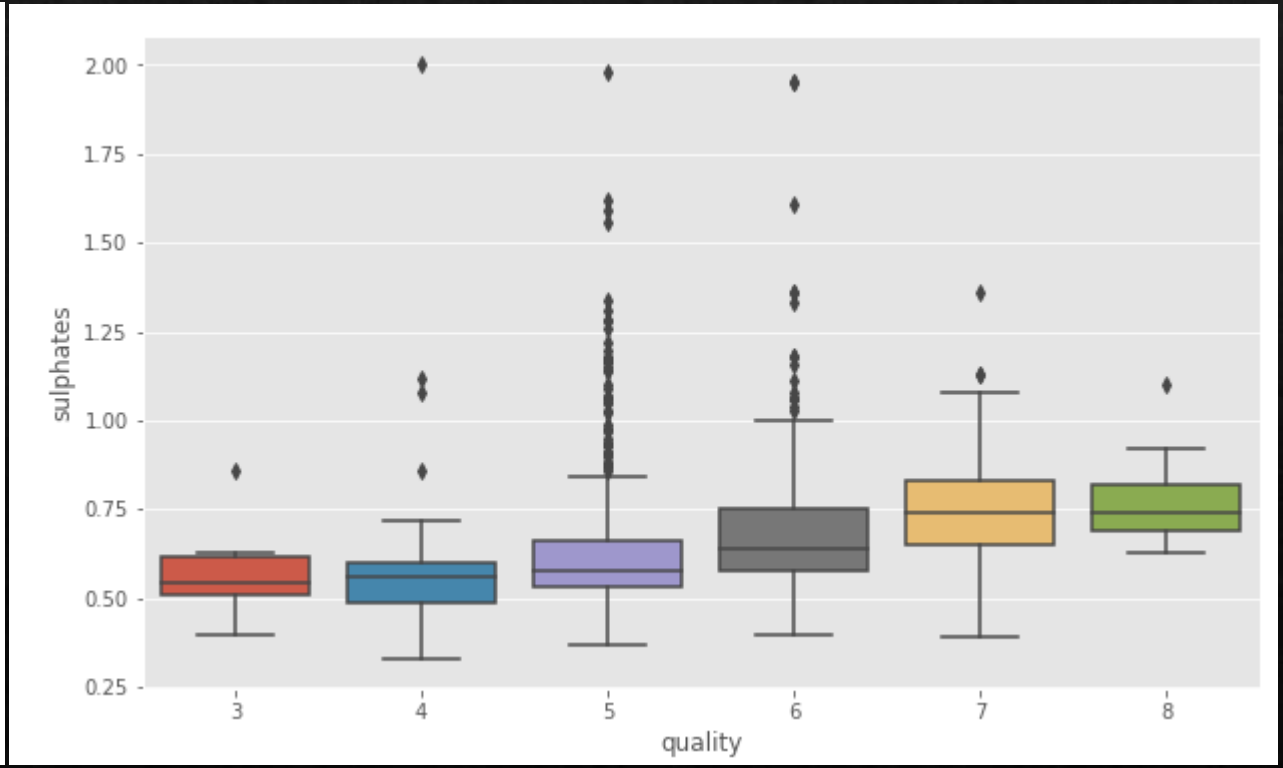
Data Analysis - Negative Relations



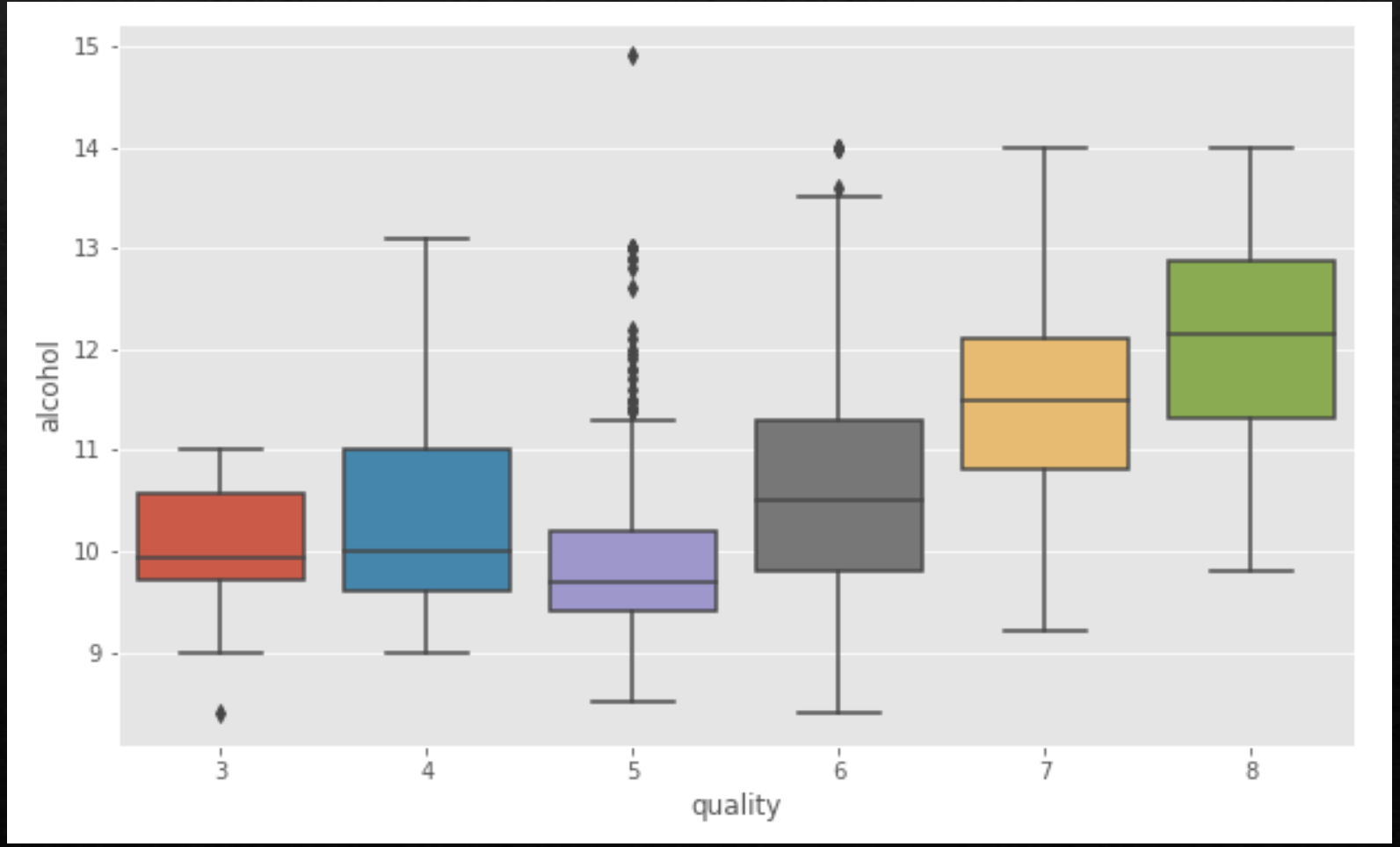


Data Analysis – Positive Relations (Part 1)





Data Analysis – Positive Relations (Part 2)

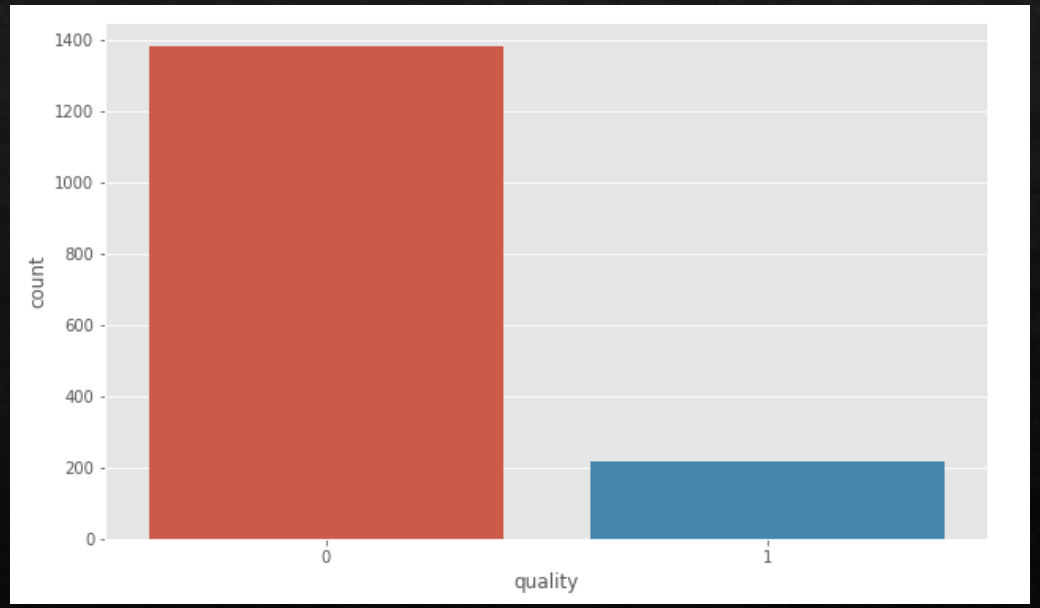


Data Analysis – General Description of Dataset

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.467792	0.996747	3.311113	0.658149	10.422983
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324	0.001887	0.154386	0.169507	1.065668
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	0.990070	2.740000	0.330000	8.400000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000	0.995600	3.210000	0.550000	9.500000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000	0.996750	3.310000	0.620000	10.200000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000	0.997835	3.400000	0.730000	11.100000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000	1.003690	4.010000	2.000000	14.900000

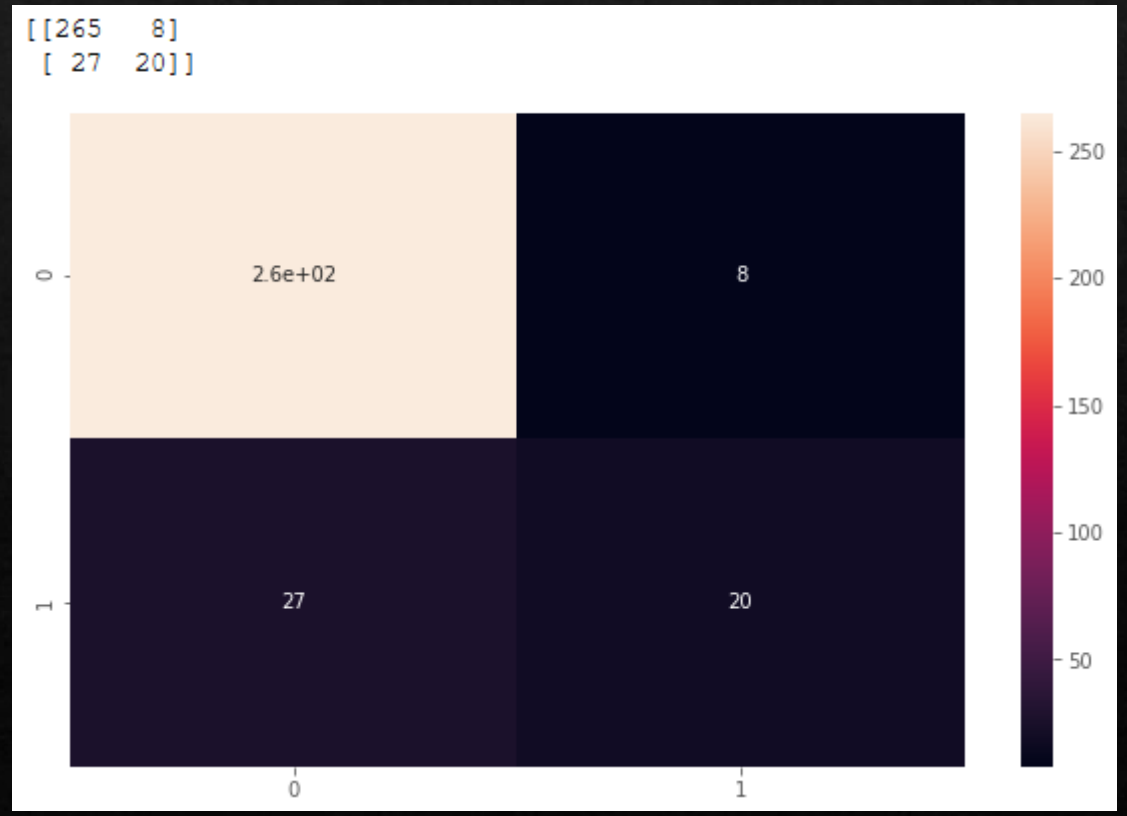
Predictive Modeling – Preparation of Dataset

- Separating the wines into “bad” quality wine and “good” quality wine
  - If 6 and below, “bad”
  - Else “good”



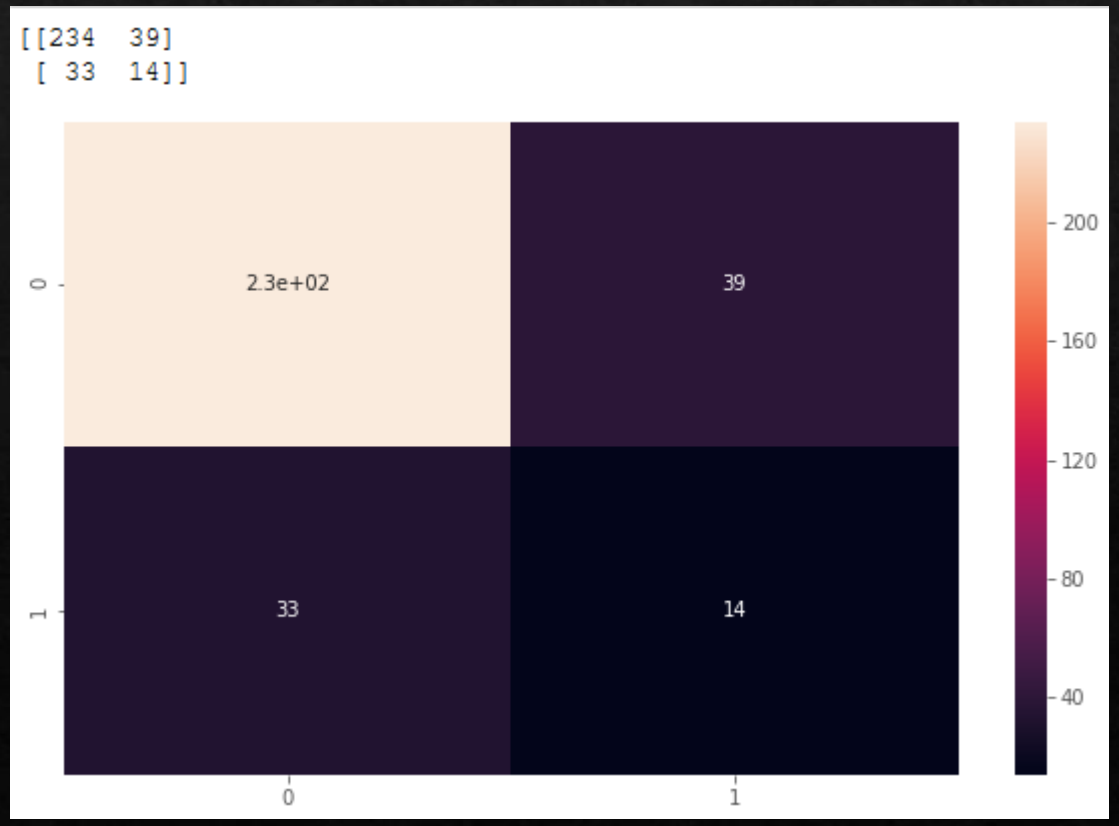
Predictive Modeling – Using Random Forest Classifier

	precision	recall	f1-score	support
0	0.91	0.97	0.94	273
1	0.71	0.43	0.53	47
micro avg	0.89	0.89	0.89	320
macro avg	0.81	0.70	0.74	320
weighted avg	0.88	0.89	0.88	320



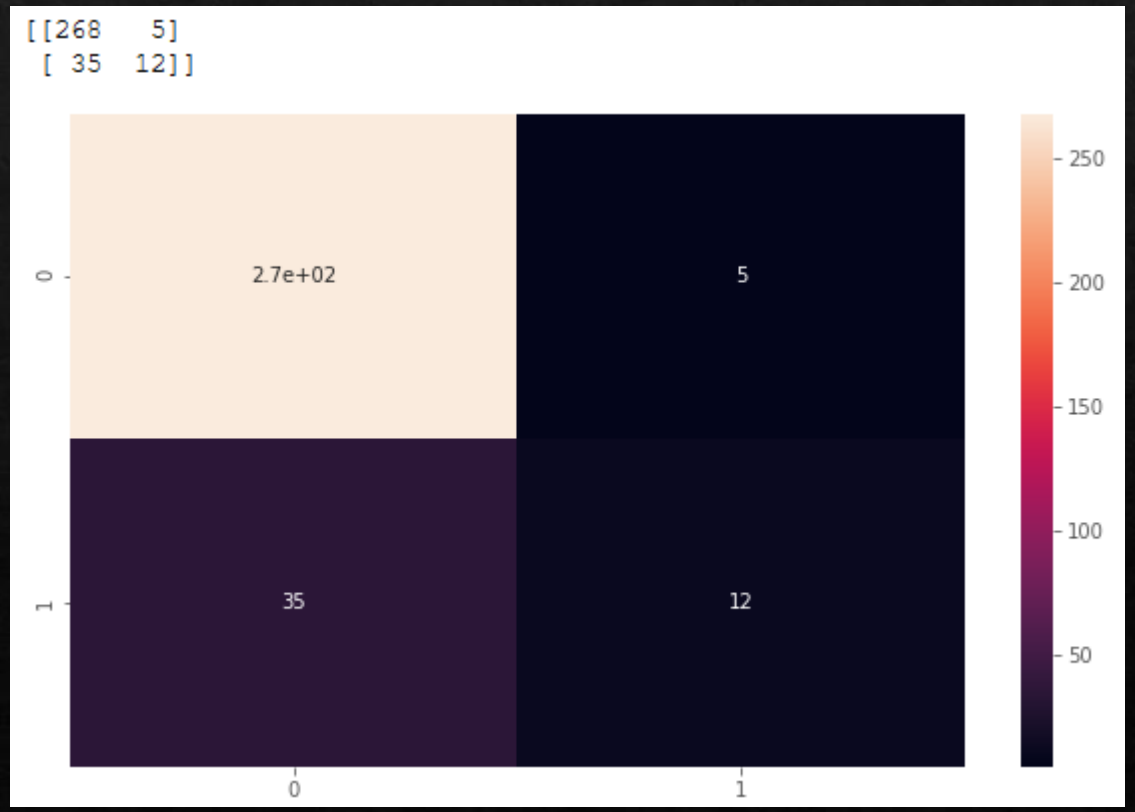
Predictive Modeling – Using Stochastic Gradient Decent Classifier

	precision	recall	f1-score	support
0	0.88	0.86	0.87	273
1	0.26	0.30	0.28	47
micro avg	0.78	0.78	0.78	320
macro avg	0.57	0.58	0.57	320
weighted avg	0.79	0.78	0.78	320



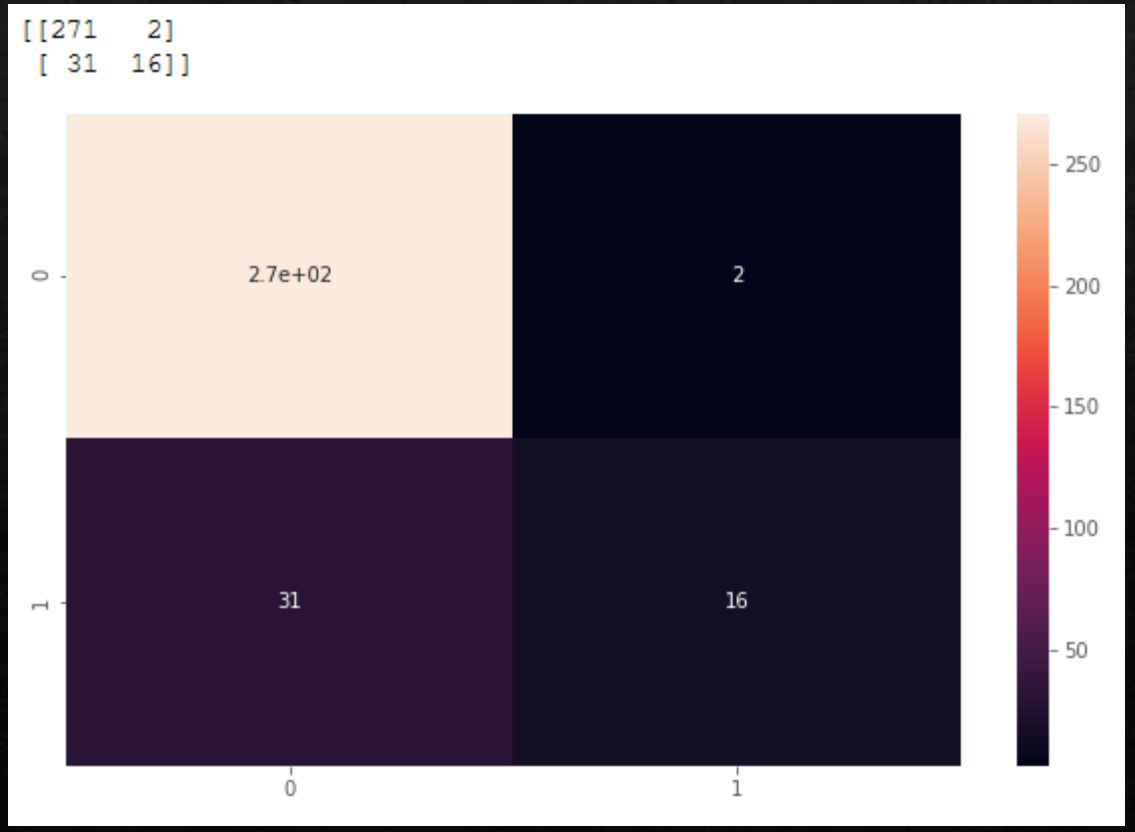
Predictive Modeling – Using Support Vector Classifier

	precision	recall	f1-score	support
0	0.88	0.98	0.93	273
1	0.71	0.26	0.37	47
micro avg	0.88	0.88	0.88	320
macro avg	0.80	0.62	0.65	320
weighted avg	0.86	0.88	0.85	320



Predictive Modeling – Using Support Vector Classifier (Weighted)

	precision	recall	f1-score	support
0	0.90	0.99	0.94	273
1	0.89	0.34	0.49	47
micro avg	0.90	0.90	0.90	320
macro avg	0.89	0.67	0.72	320
weighted avg	0.90	0.90	0.88	320



Conclusion

- Volatile Acidity and Chlorides reduces Quality of wine
- Citric Acids, Sulphates, and Alcohol increases Quality of wine
- Using Classification Models to predict the classification of a product