Wine Quality Analytics

based on the composition.

ℹ Automatic video captions available.

Agenda

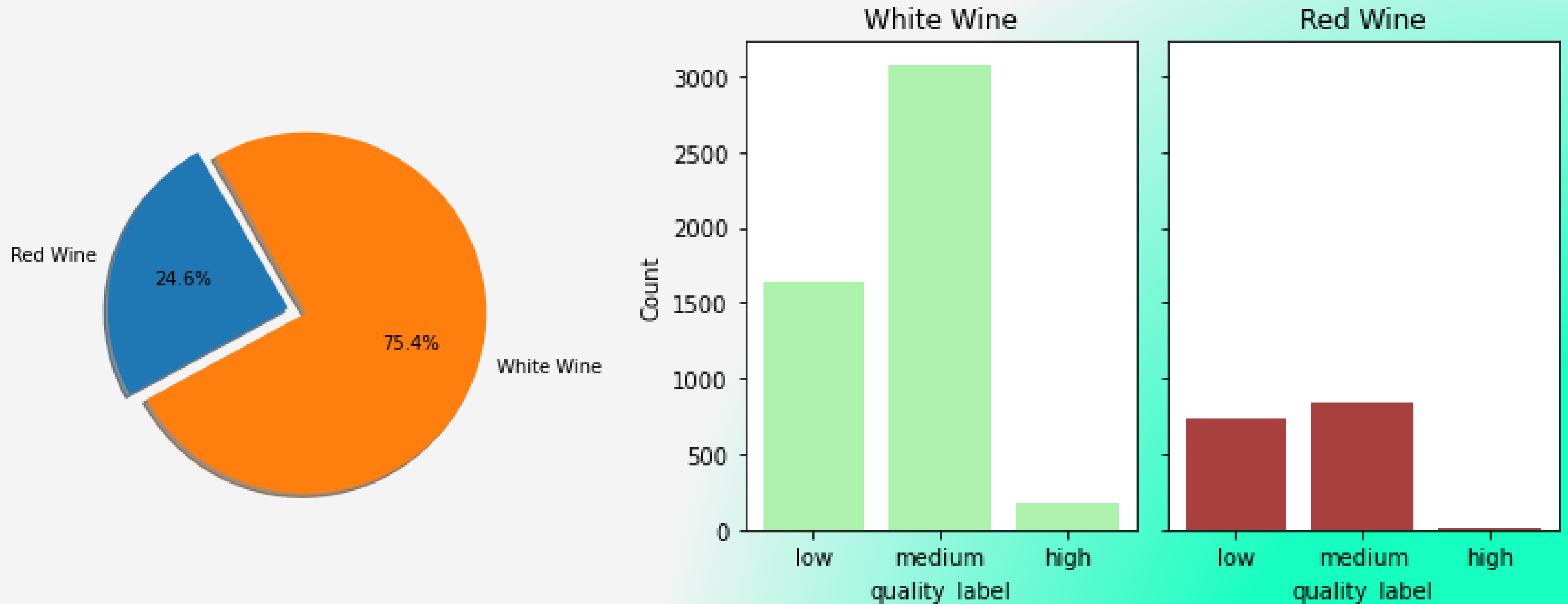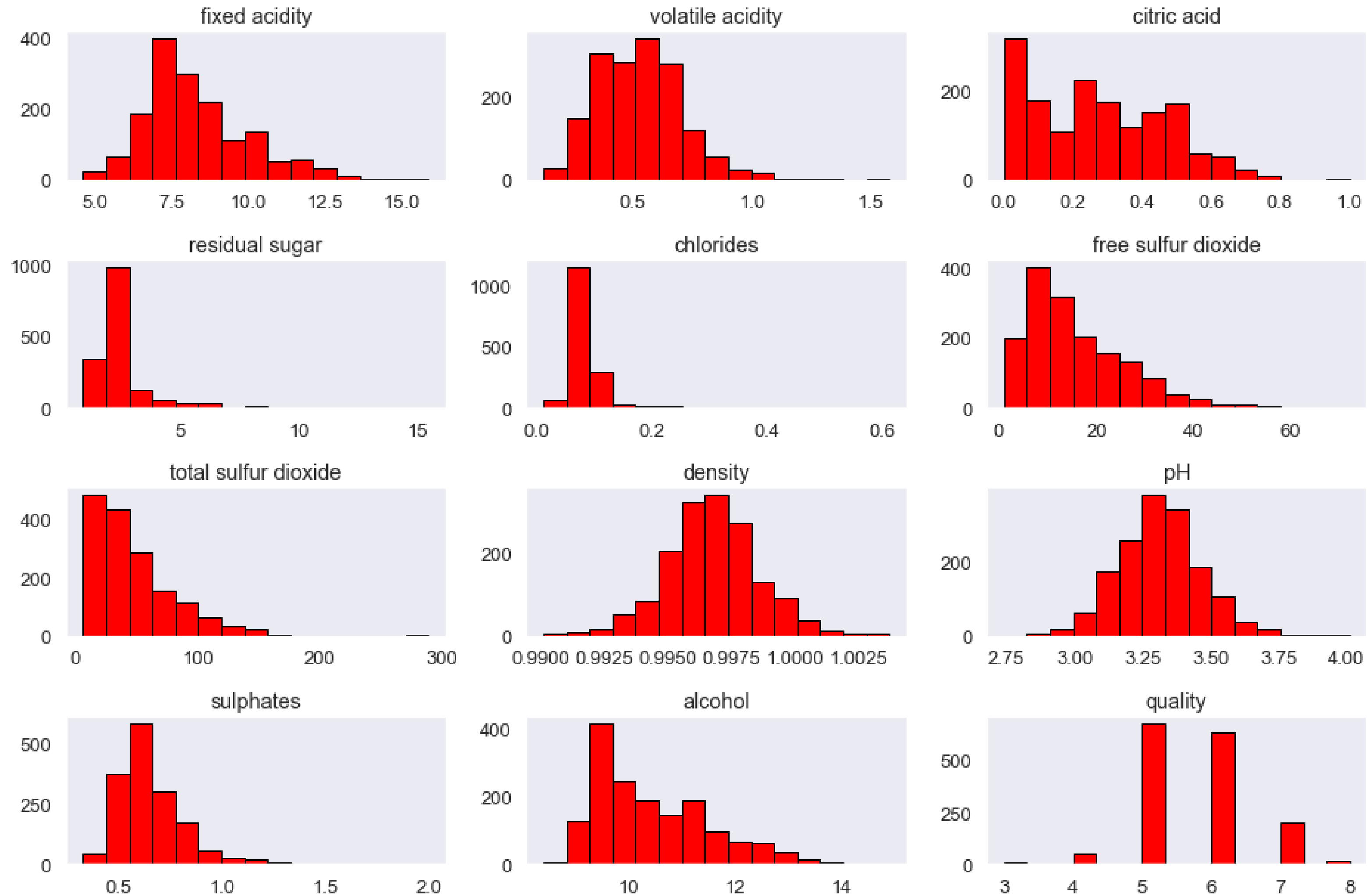- <u>About Data Set</u>

- <u>Where We Are Now</u>

- <u>Quality Analysis</u>

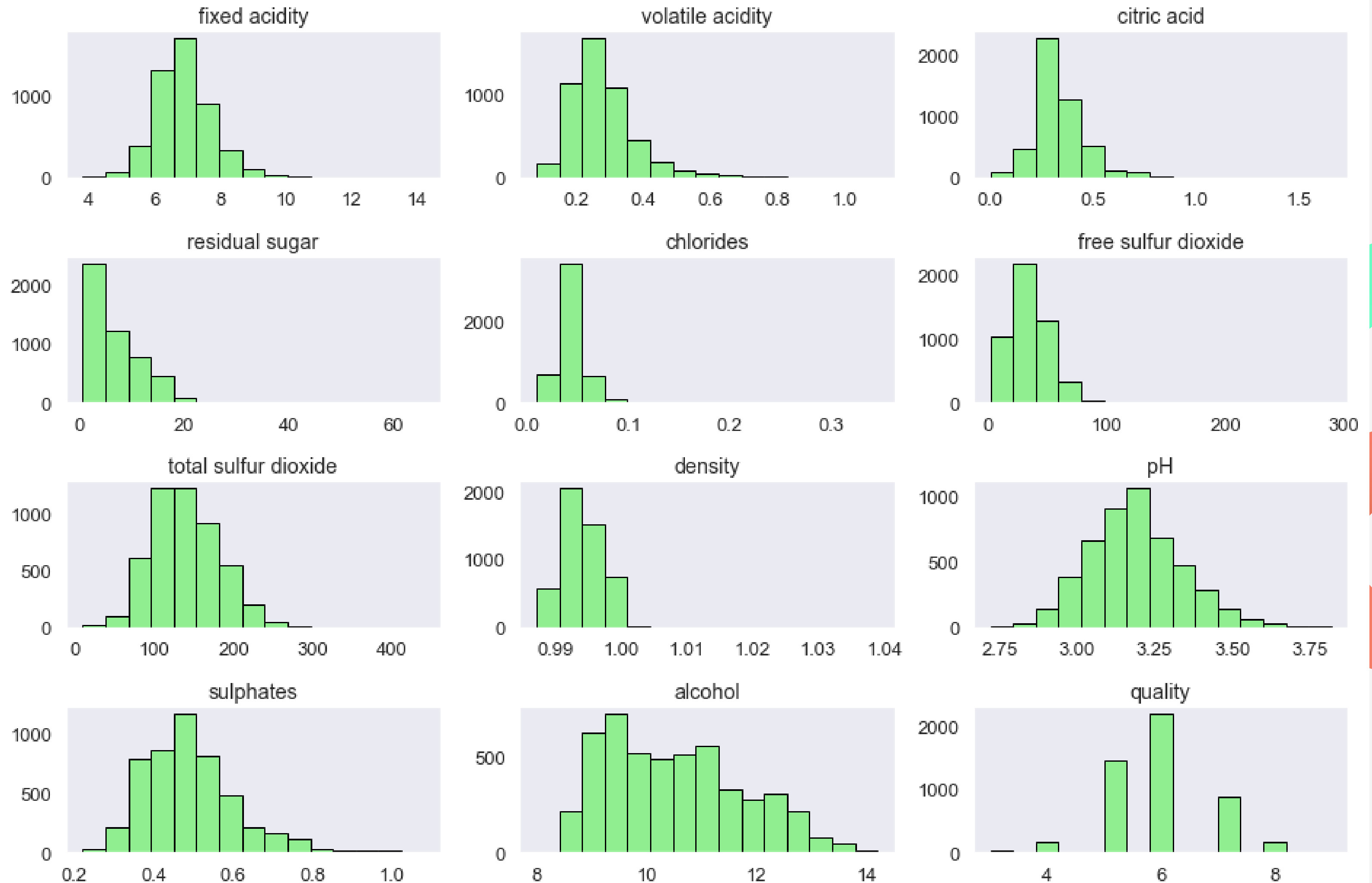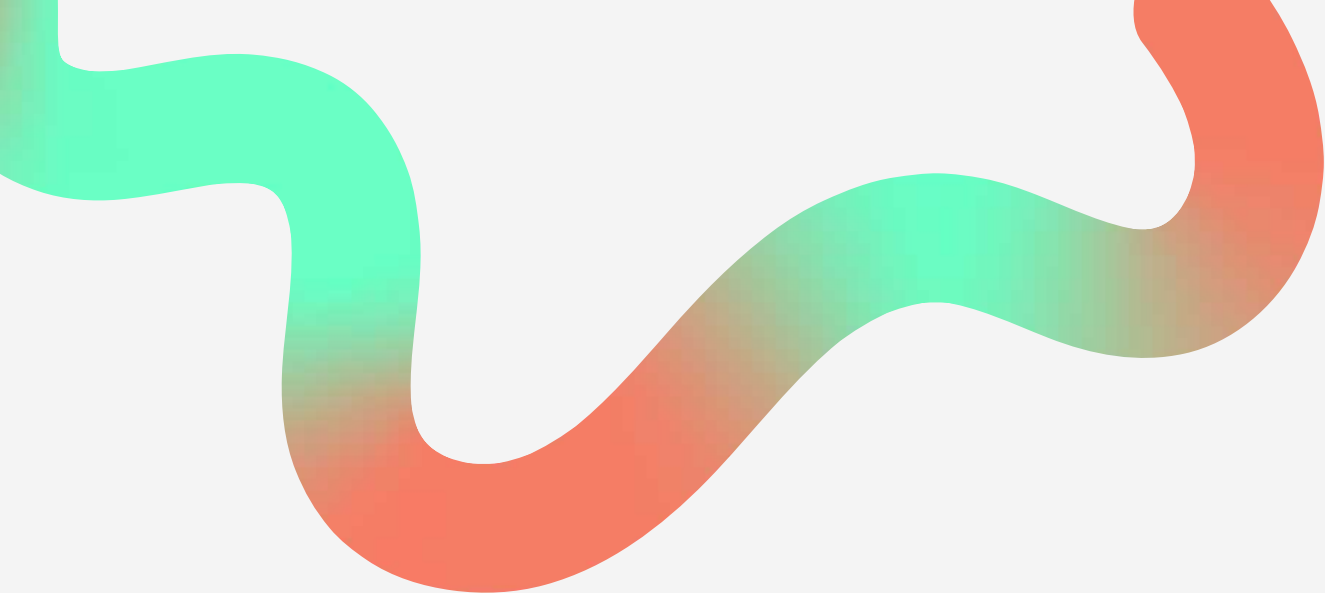- <u>Composition Analysis</u>

2
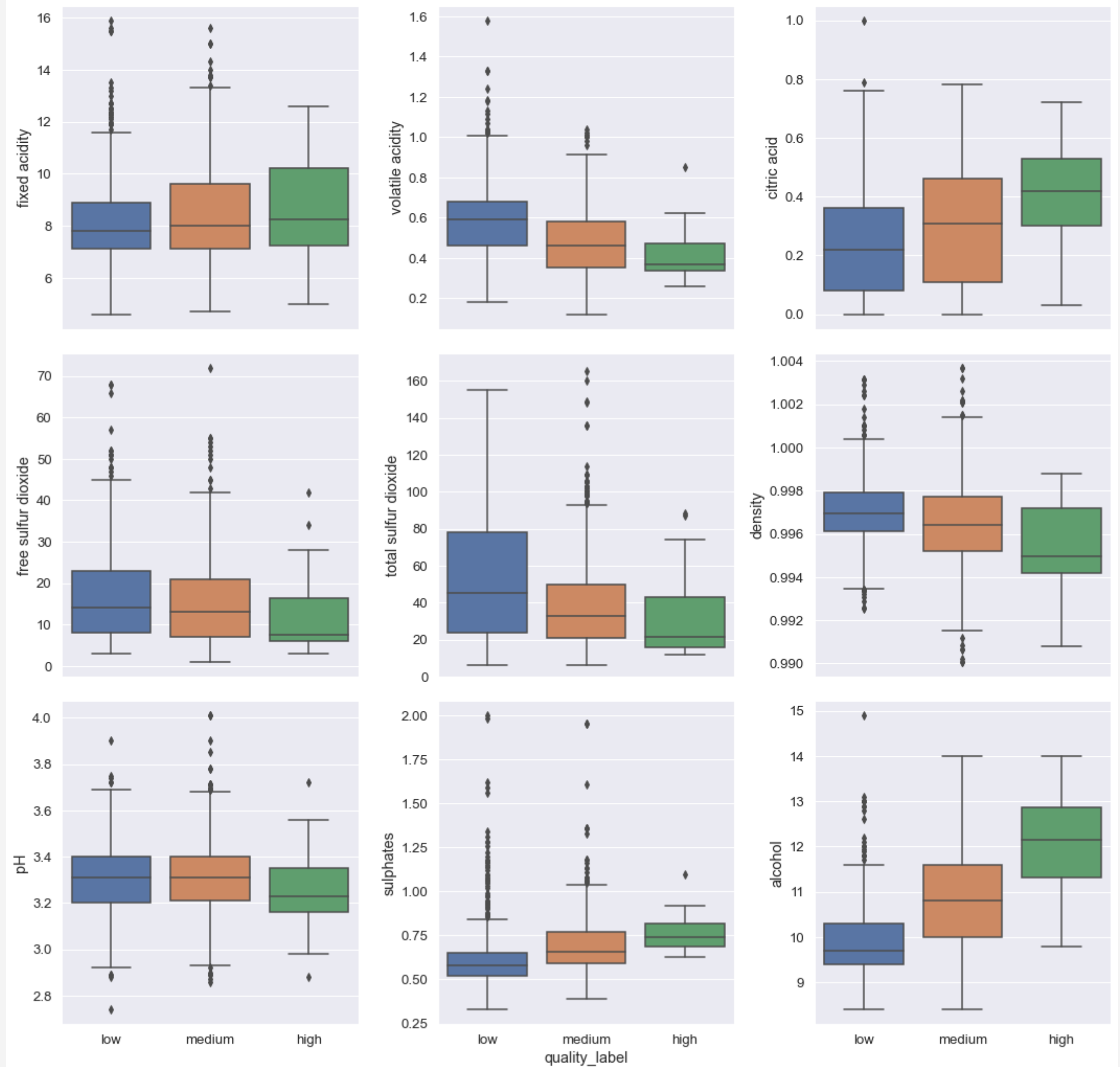
Production Summary

Red Wine Univariate Plots

White Wine Univariate Plots

Significant Points
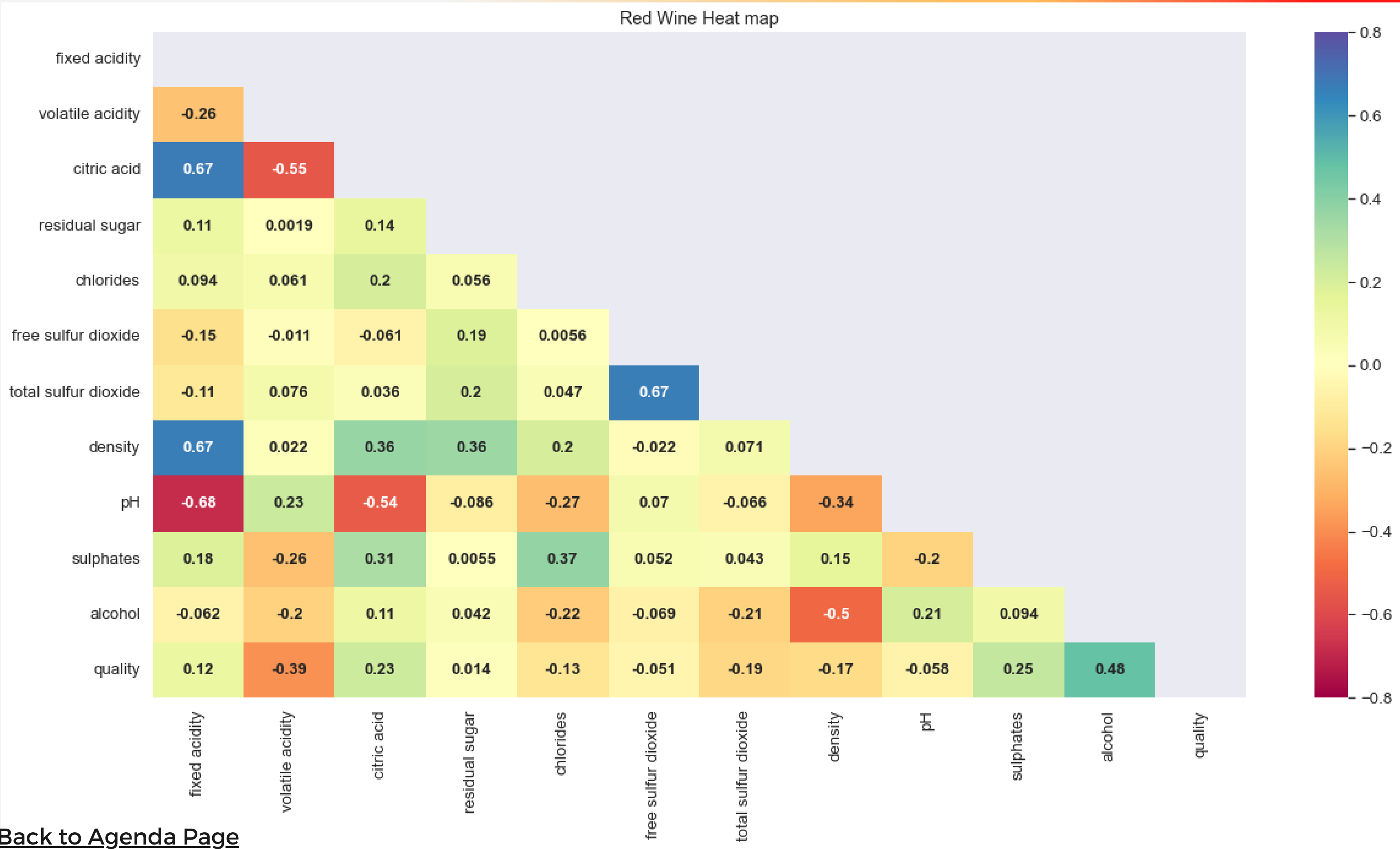
Correlations to focus on :

- Citiric Acids
- Density
- Alcohol

Red Wine Multivariate Quality Analysis
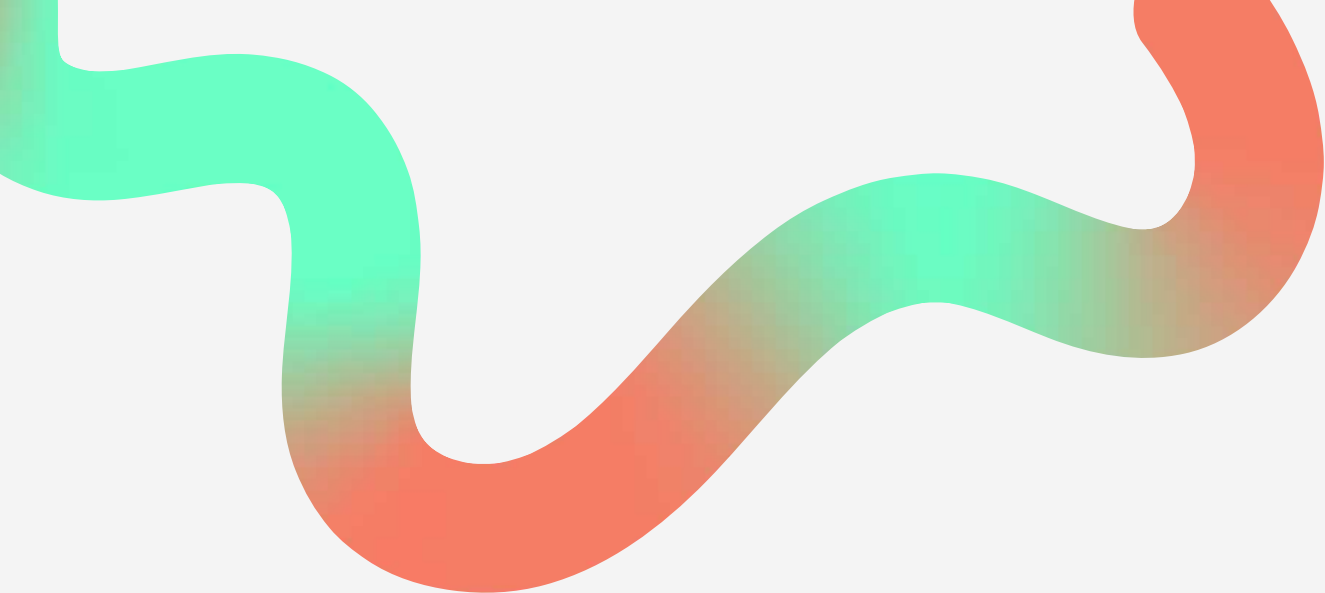
Red Wine Heat map

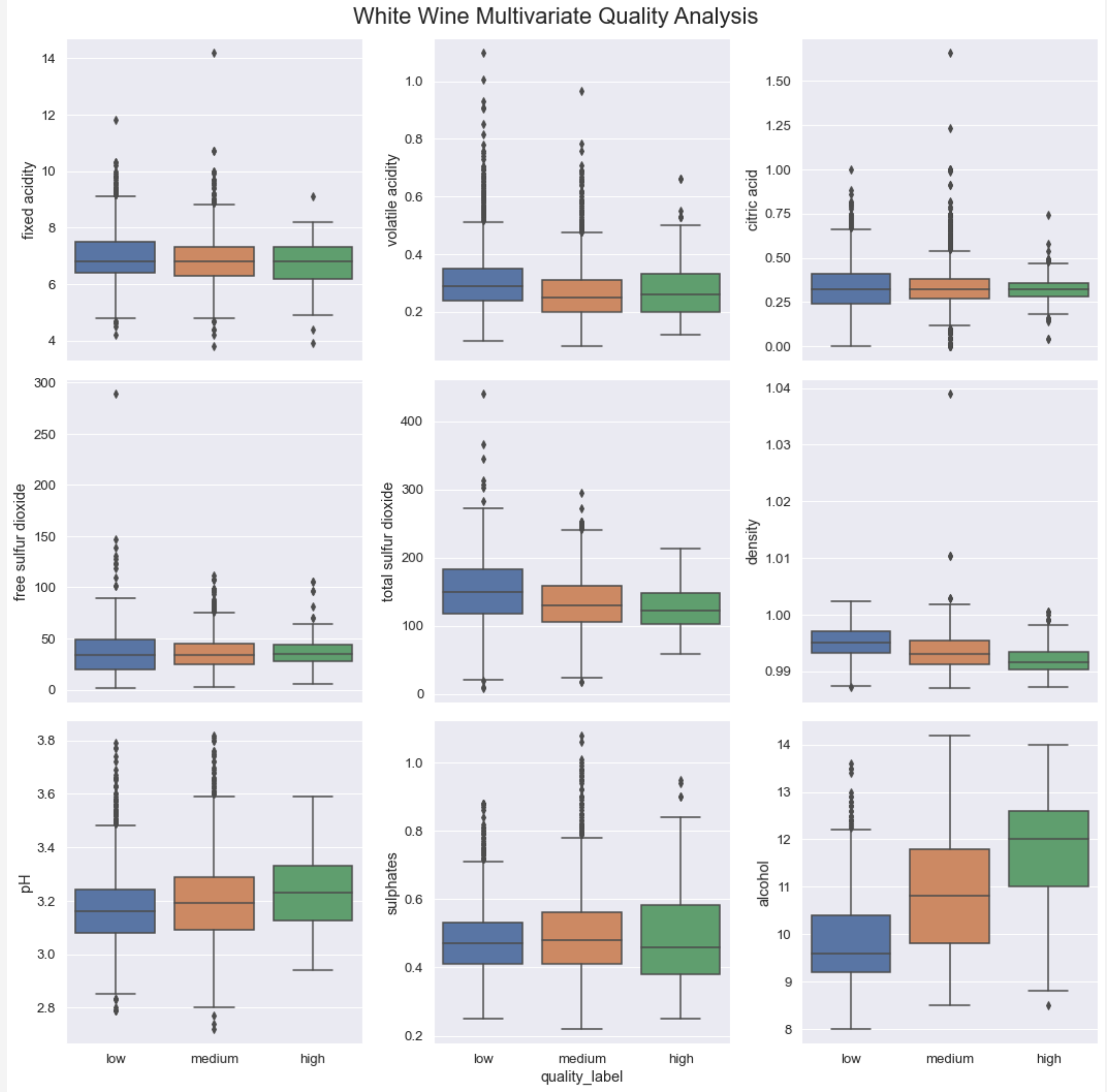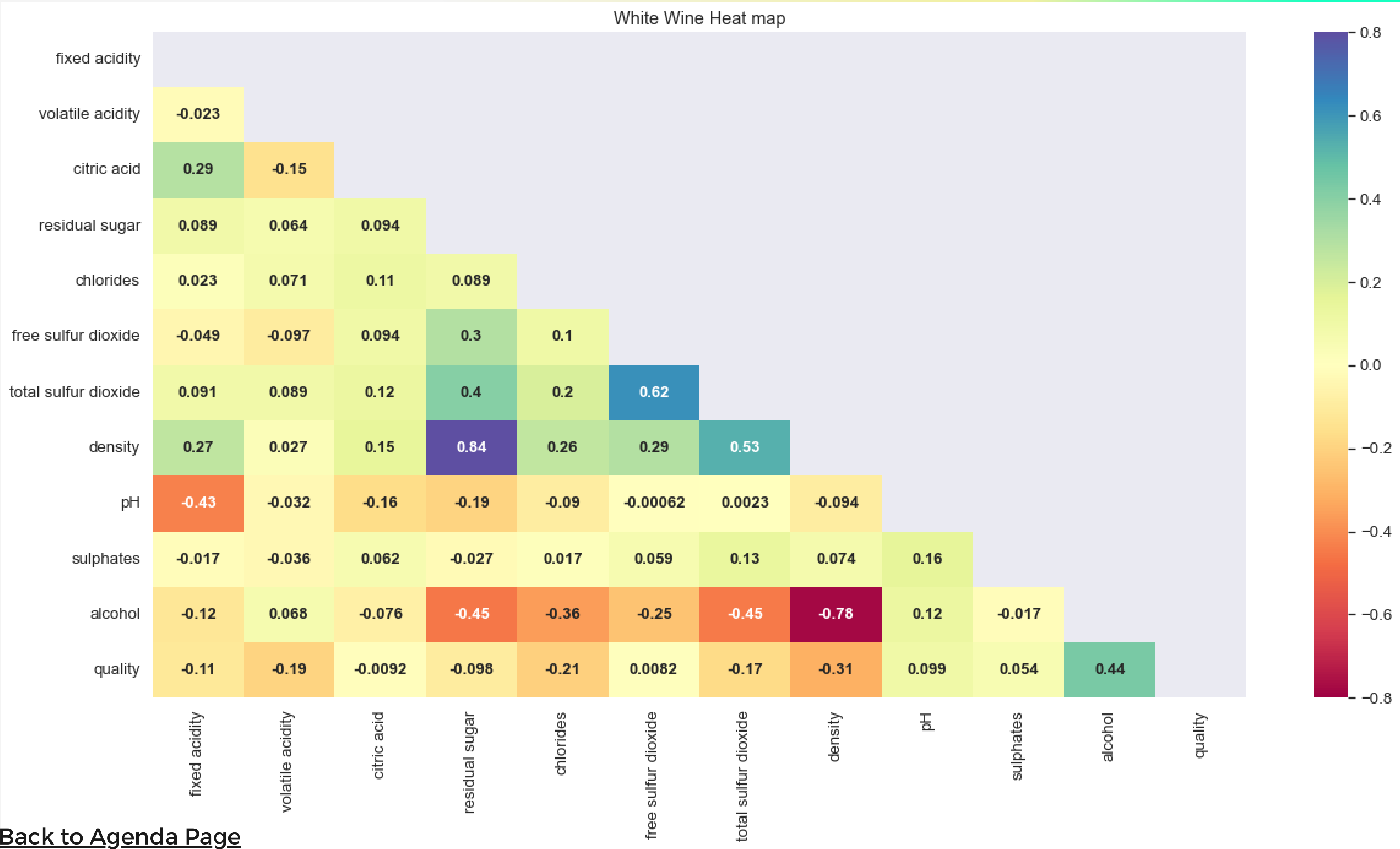# Significant Points

Correlations to focus on :

- Citiric Acids
- Density
- pH

White Wine Multivariate Quality Analysis

8

White Wine Heat map

Red Wine

White Wine

Red Wines- Alcohol & Quality

White Wine- Alcohol & Quality

Red Wine

White Wine

12
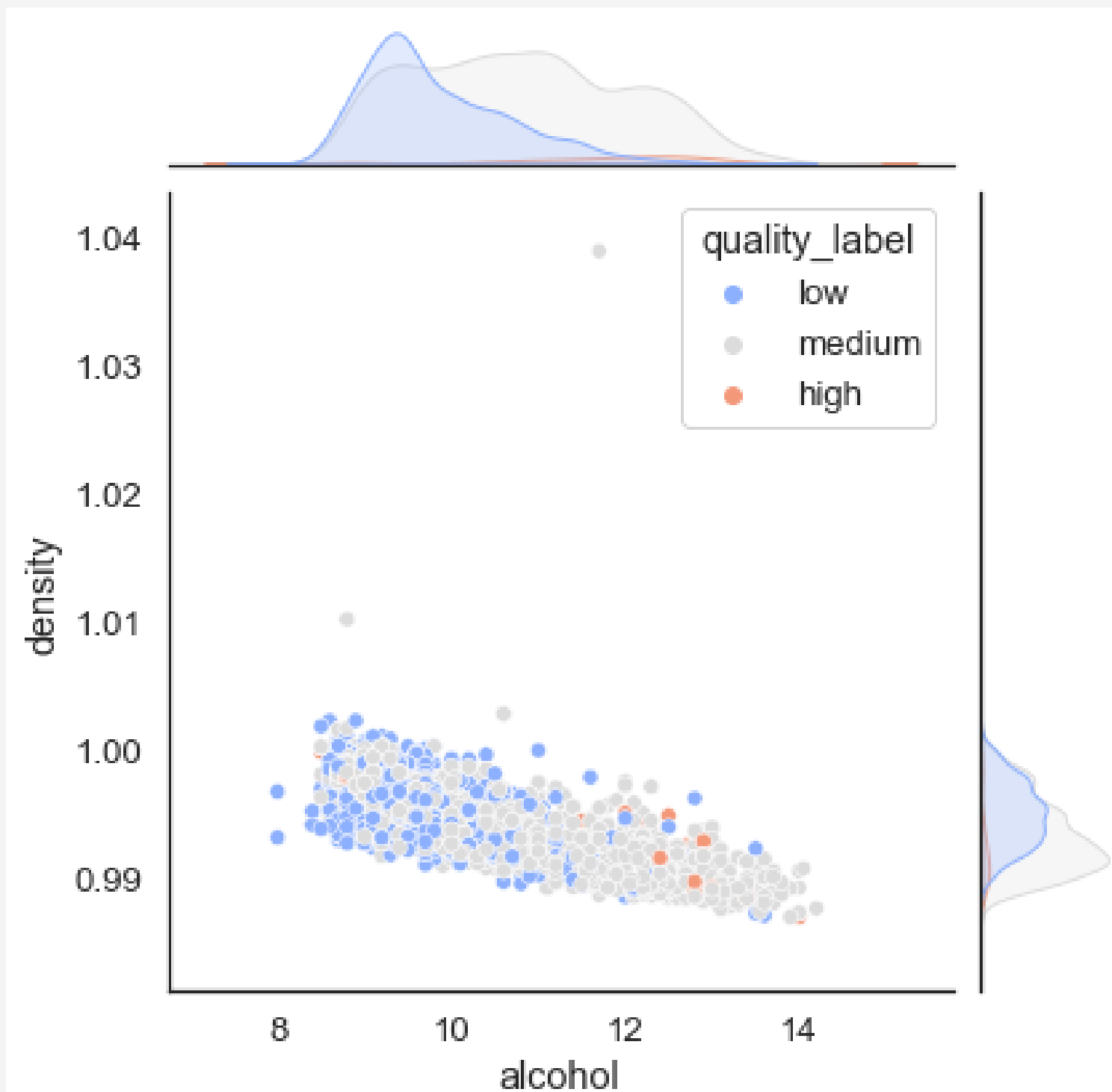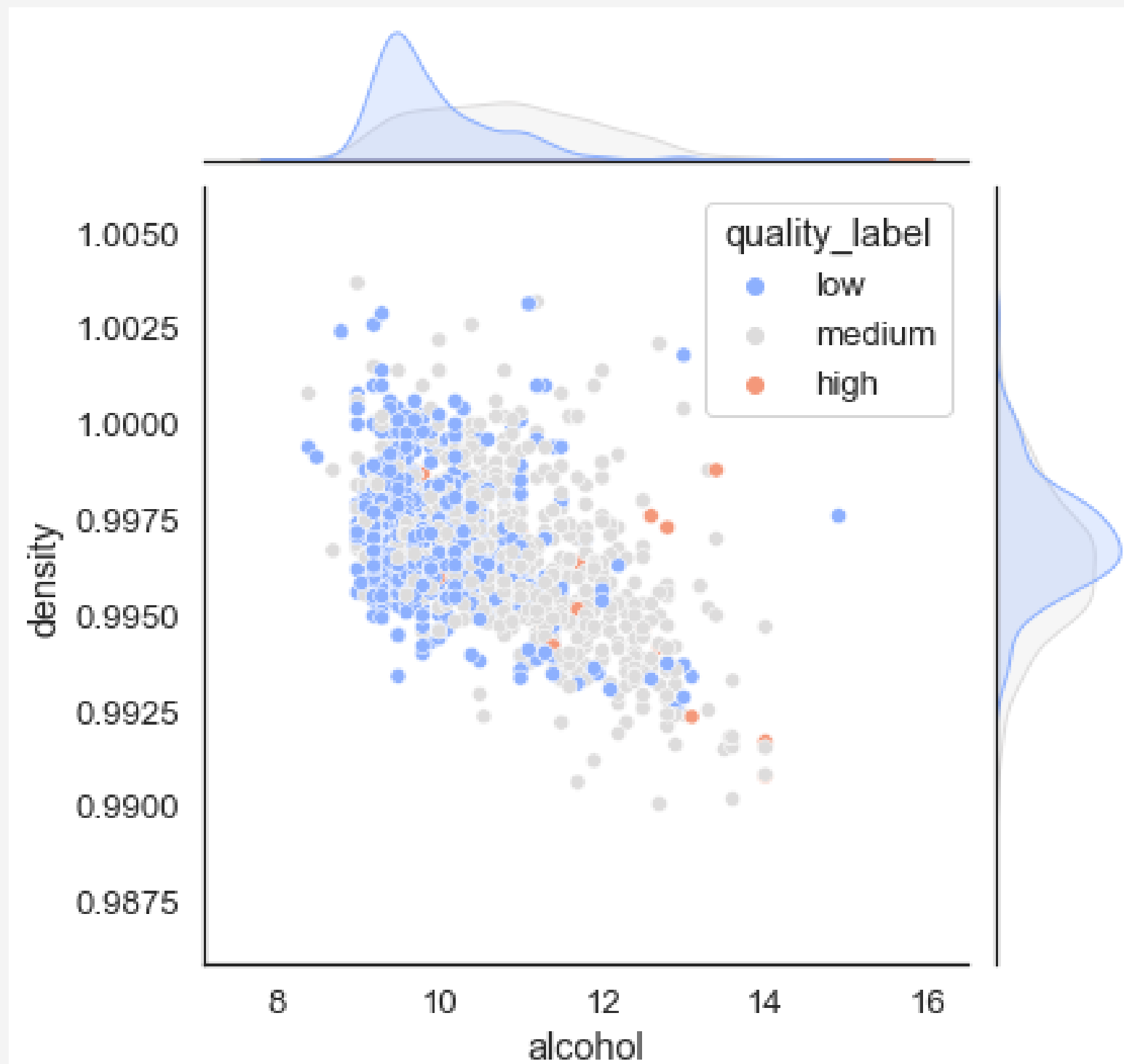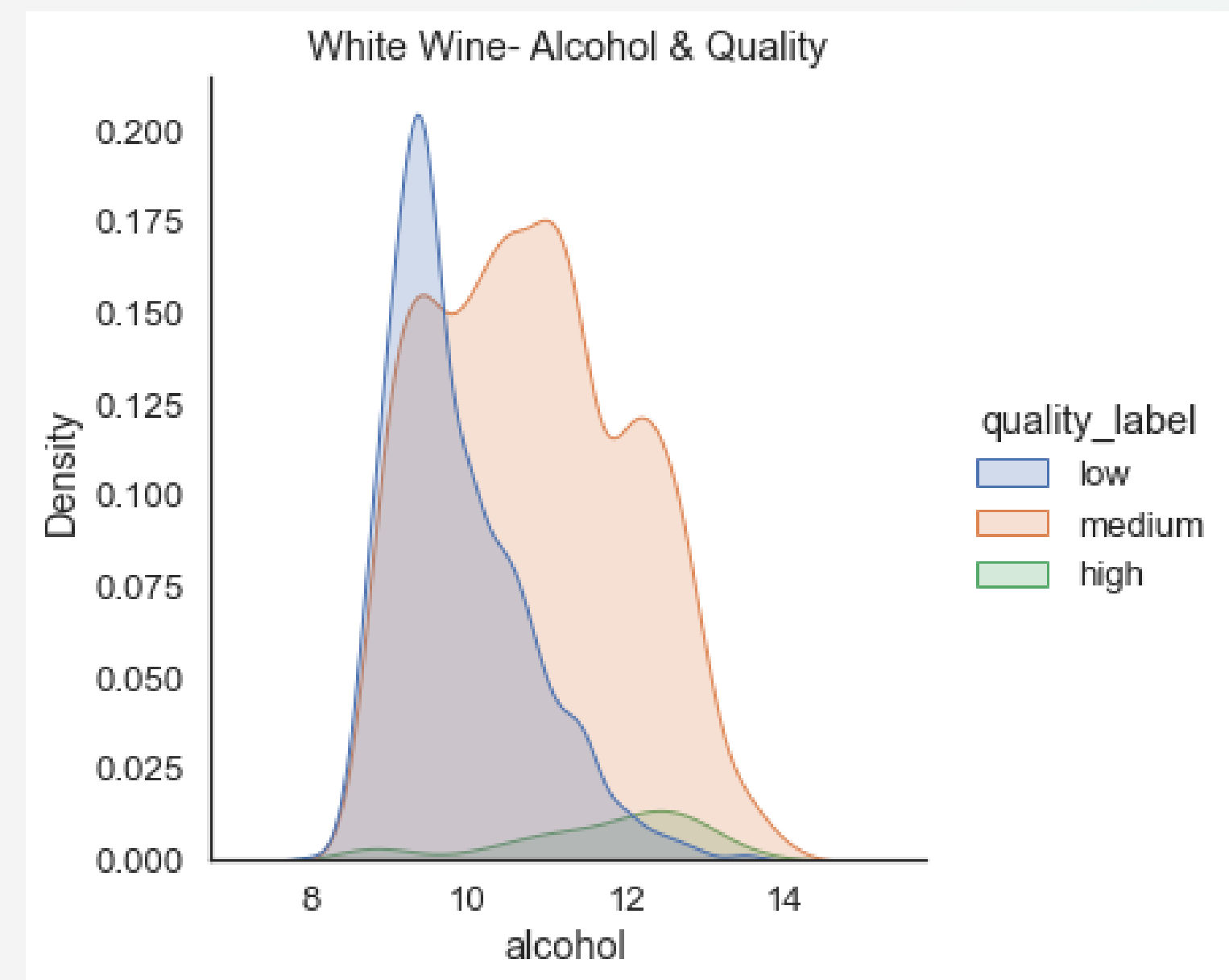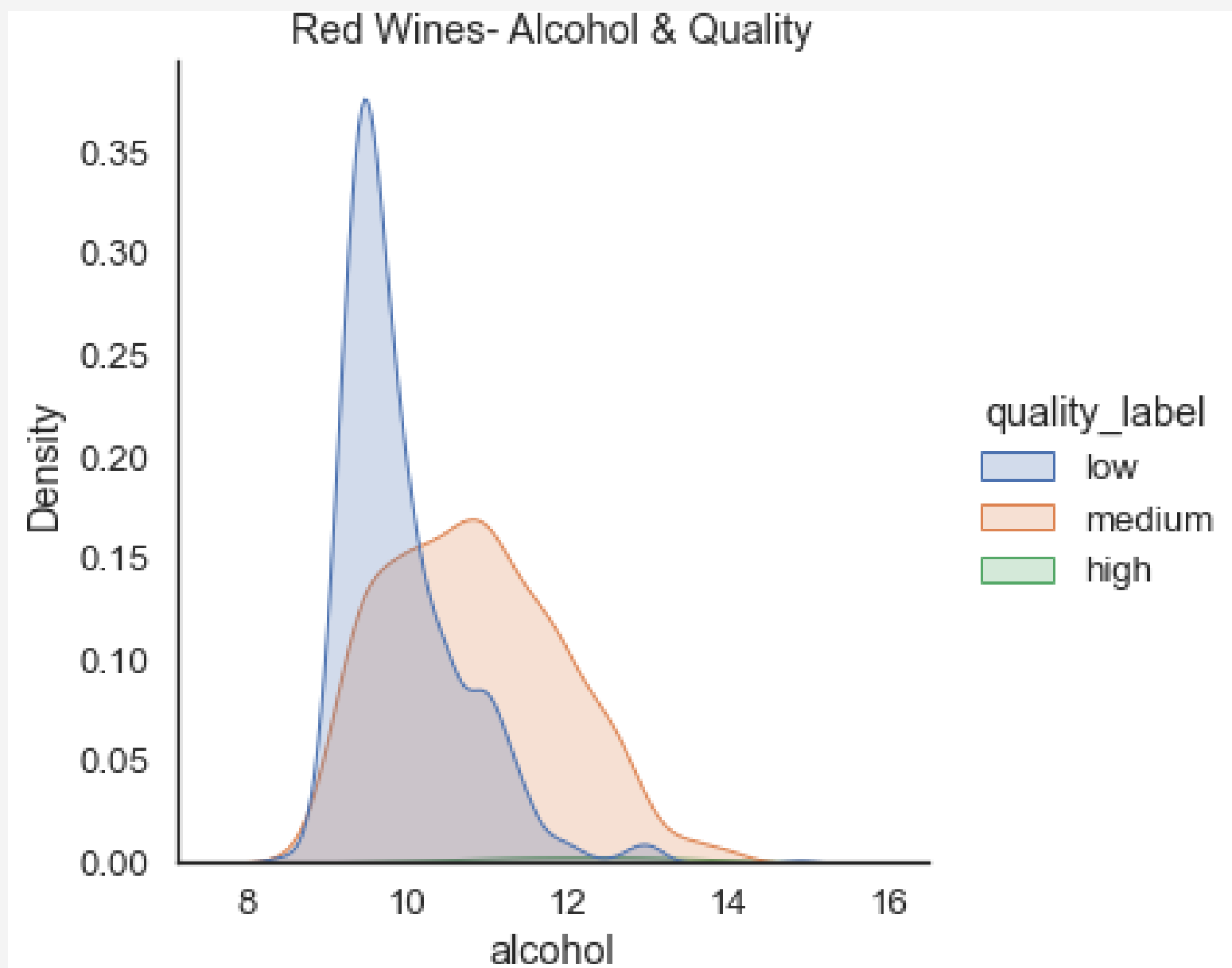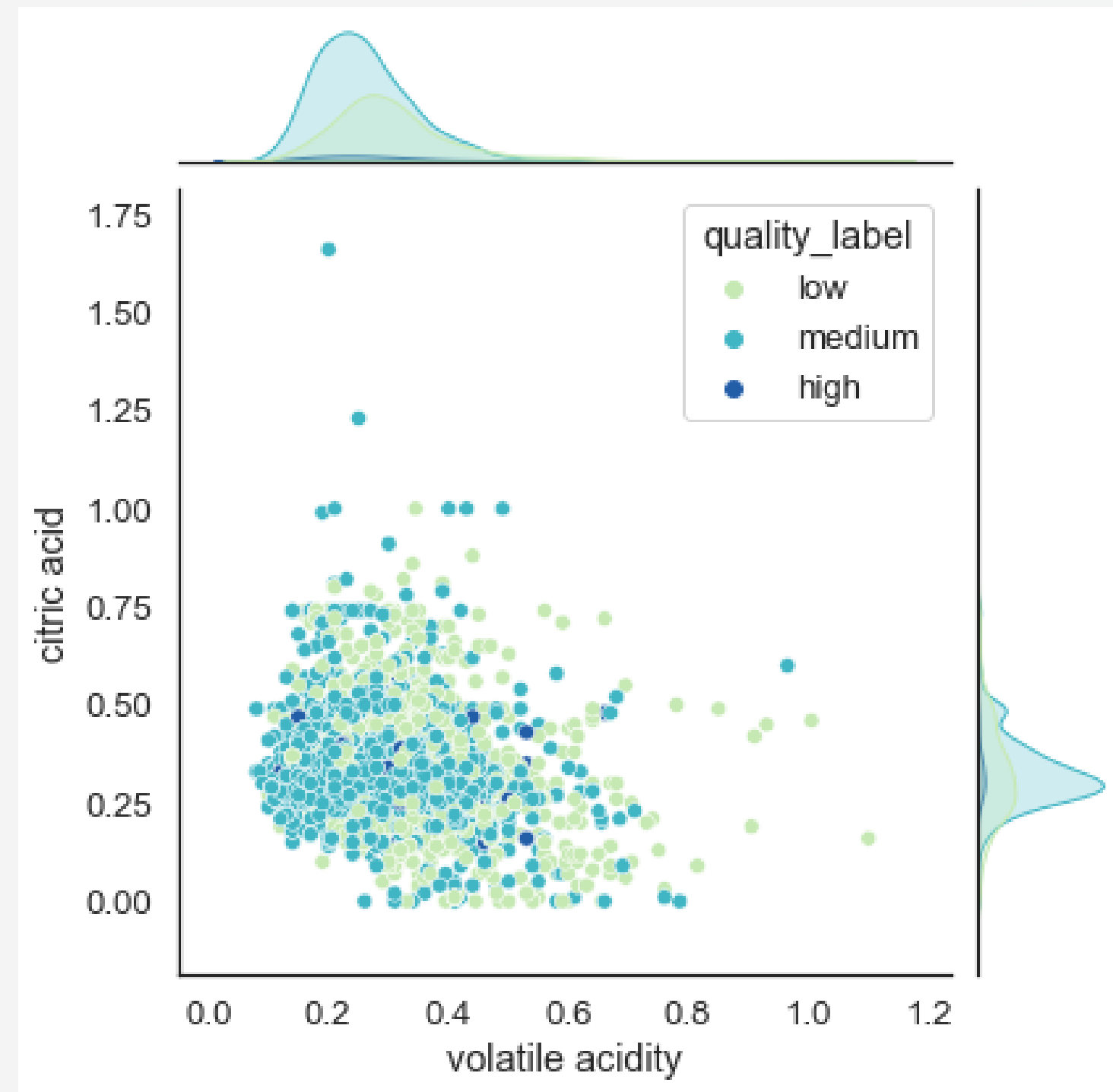
White Wine- Density & Residual Sugar

13
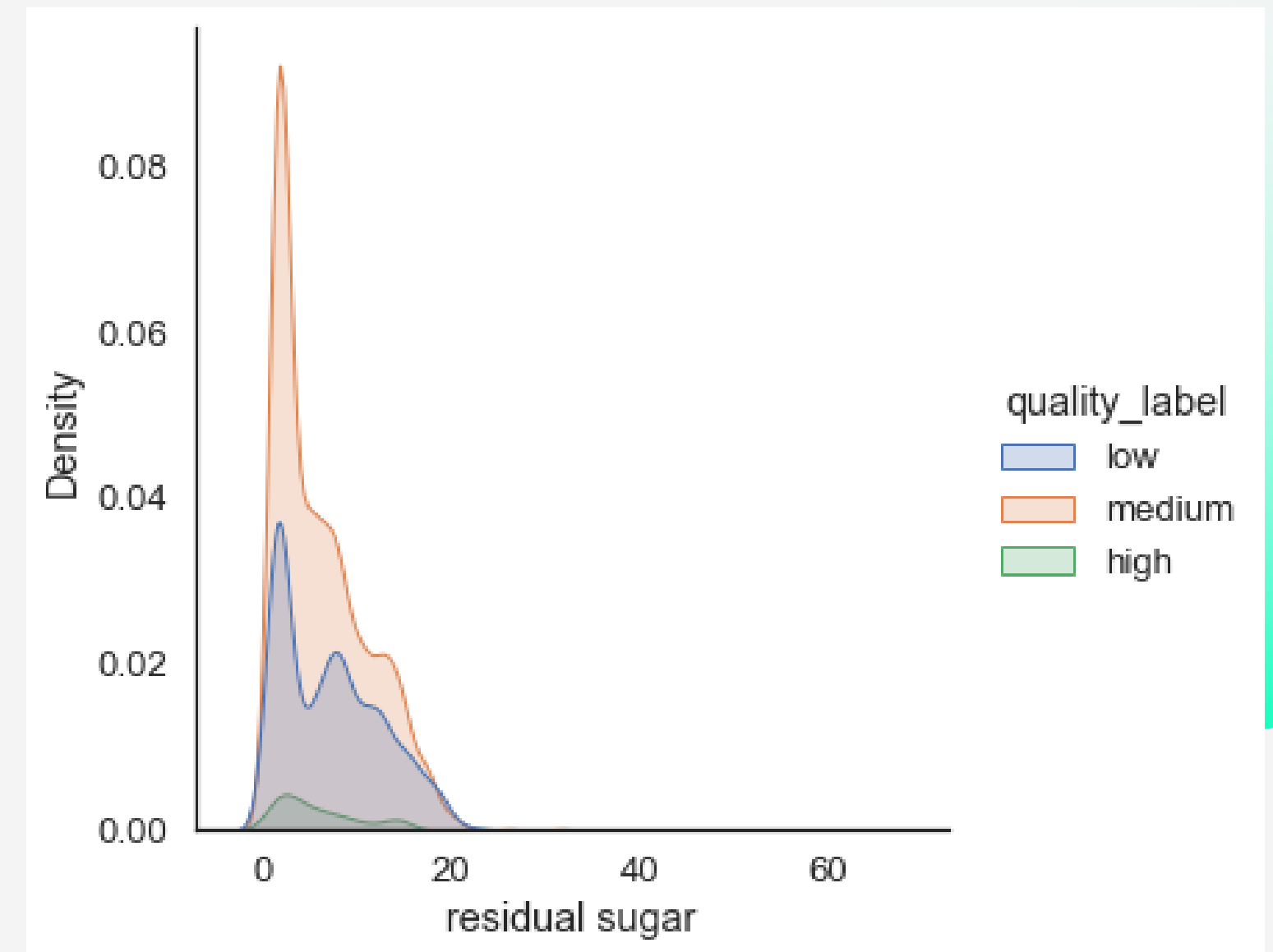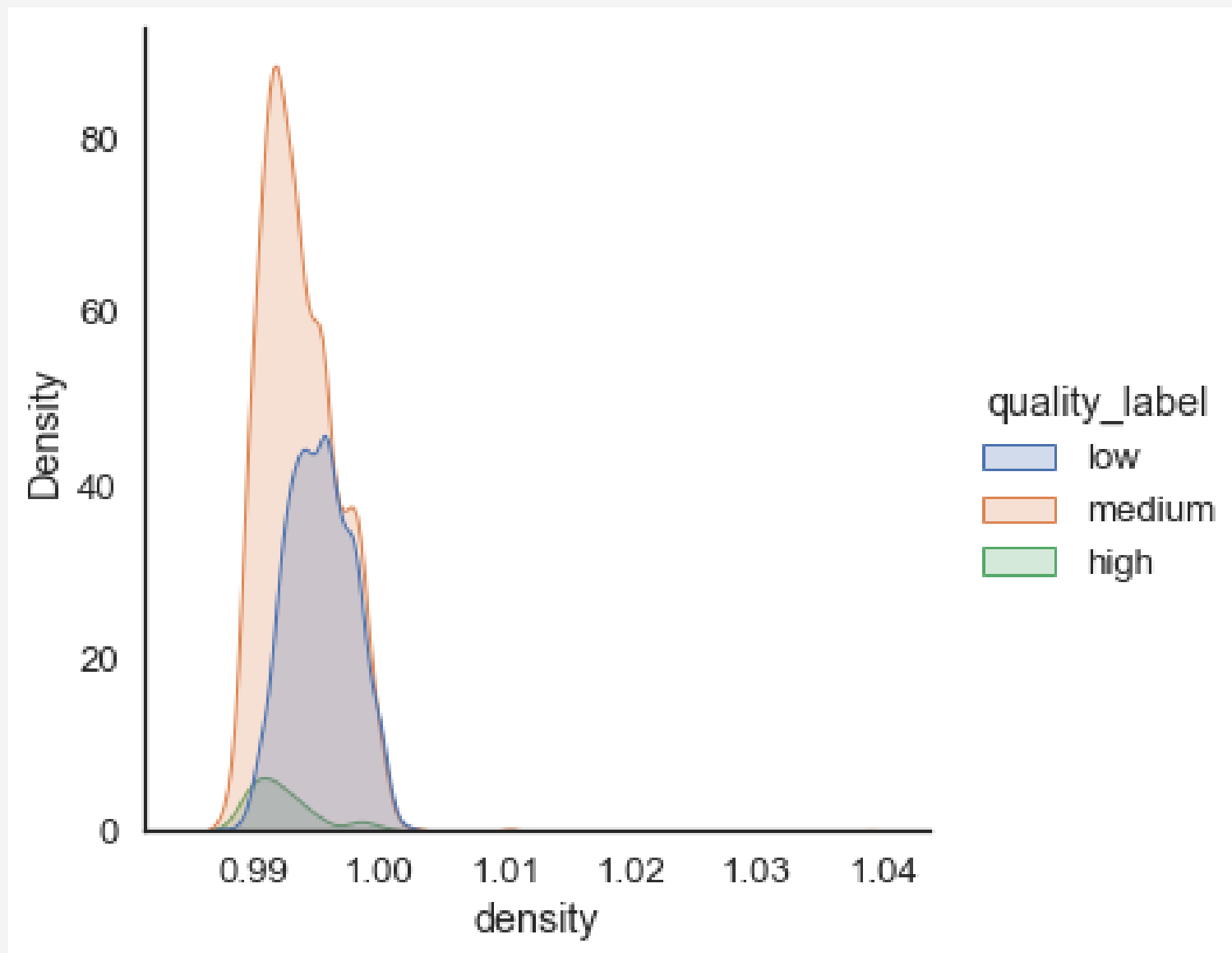
Apply Machine Learning models

Agenda

Back to Agenda Page
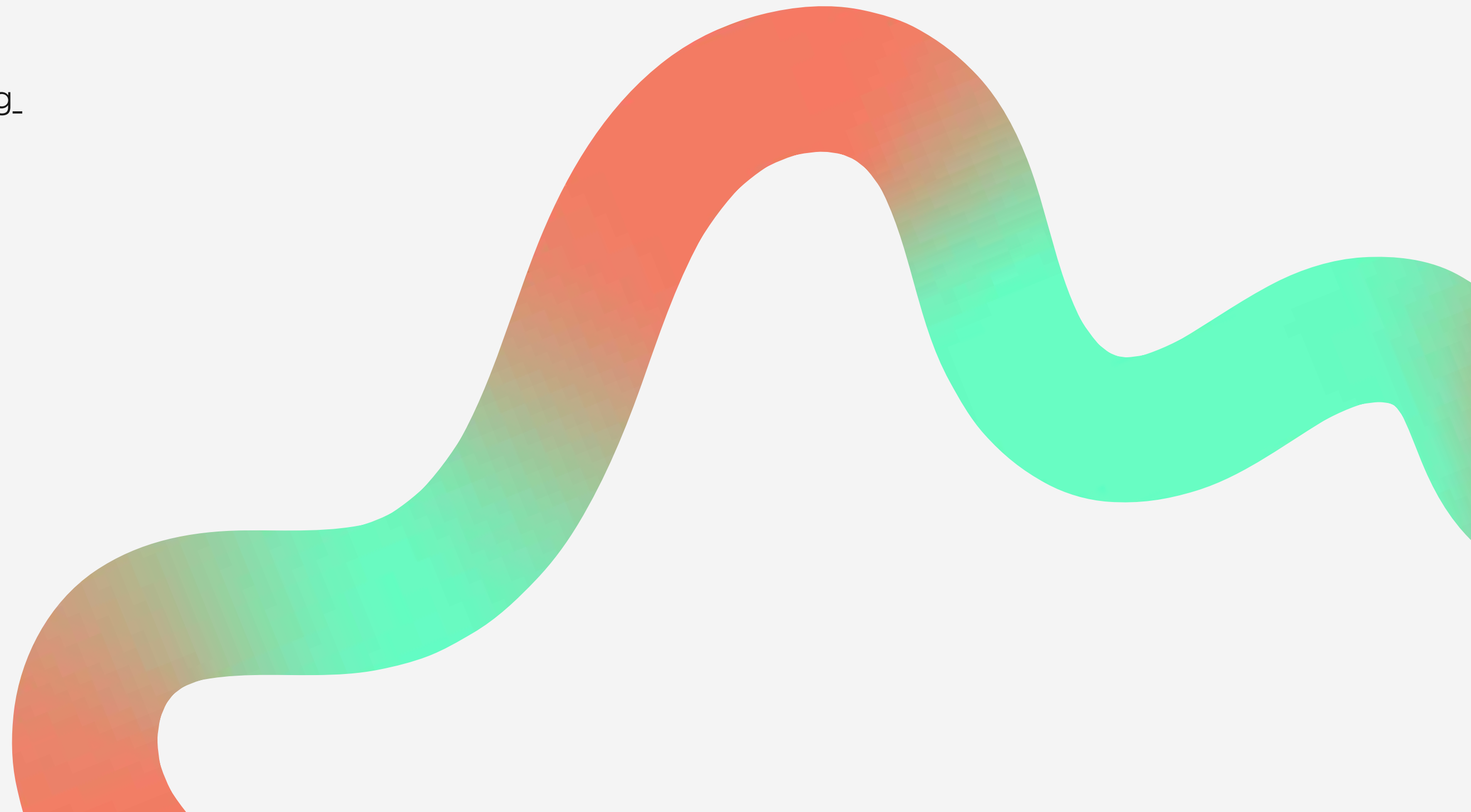
White Wine Accuracy Score

The accuracy for Random Forest is
- The training is: 100.0 %
- The testing is: 79.28%
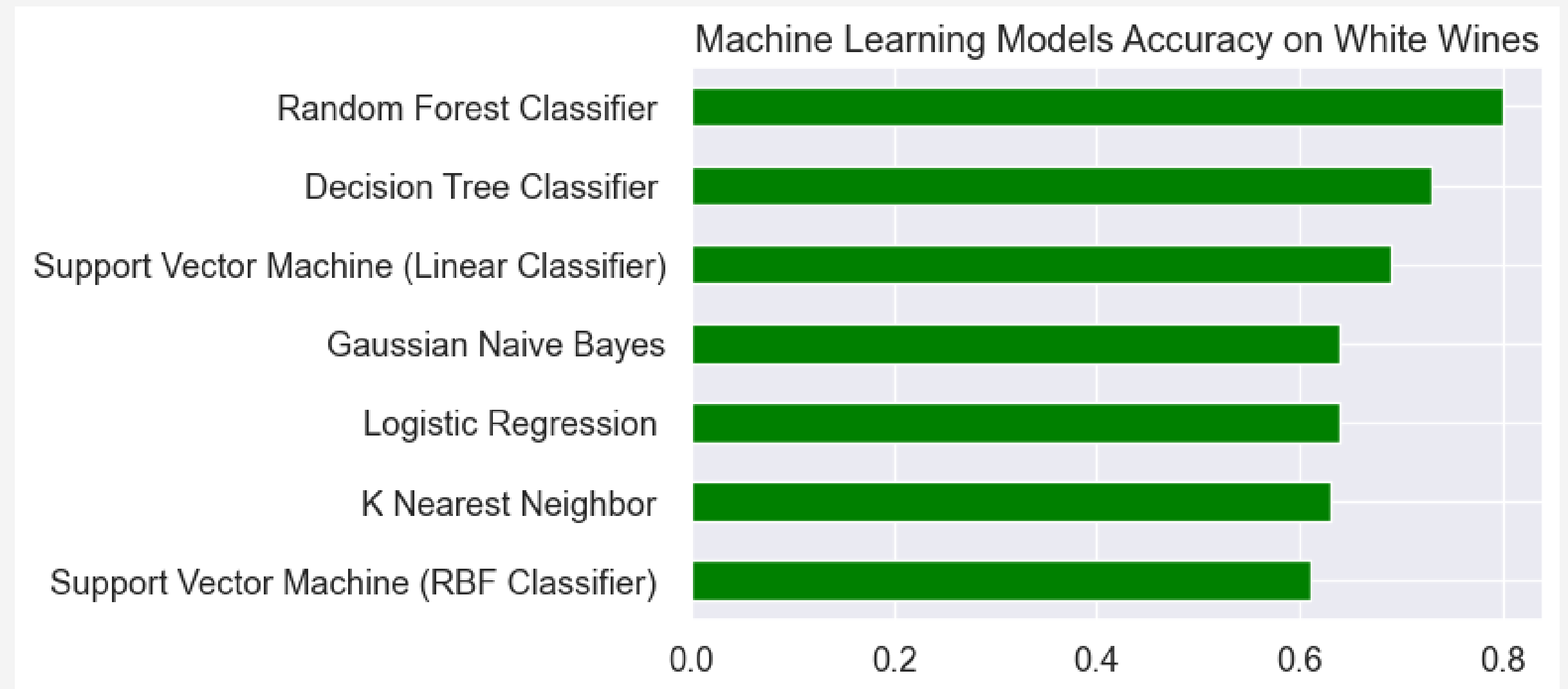
Machine Learning Models Accuracy on White Wines

Confusion Matrix

based on Random Forest Classifier model
before dropping any feature
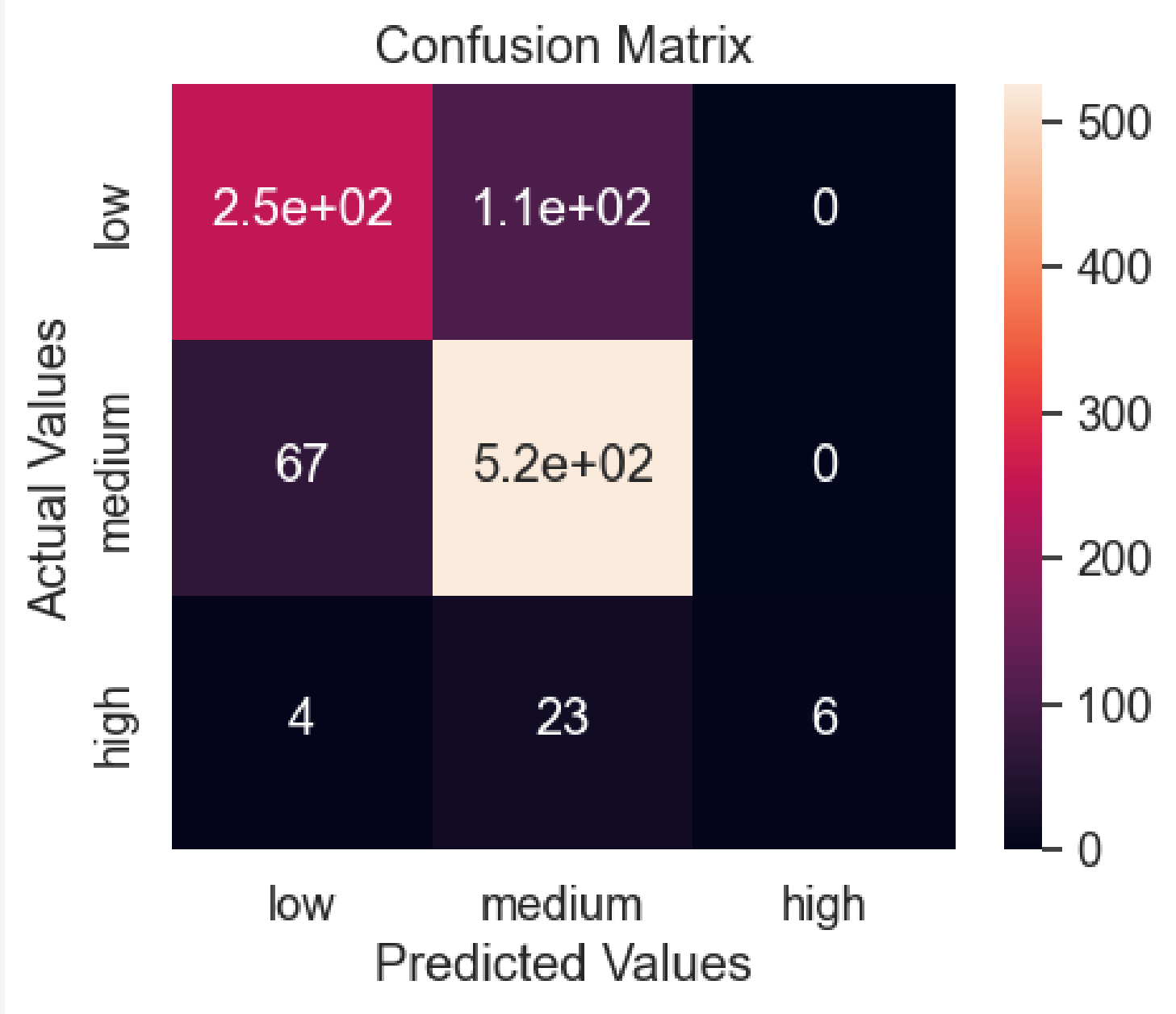

```
confusion_matrix:
 [[249 106   0]
 [ 67 525   0]
 [  4  23   6]]
Accuracy score: 0.7959183673469388
              precision    recall  f1-score   support

         low       0.78      0.70      0.74       355
      medium       0.80      0.89      0.84       592
        high       1.00      0.18      0.31        33

    accuracy                           0.80       980
   macro avg       0.86      0.59      0.63       980
weighted avg       0.80      0.80      0.79       980


Cohen-Kappa score: 0.5733881253890656
```
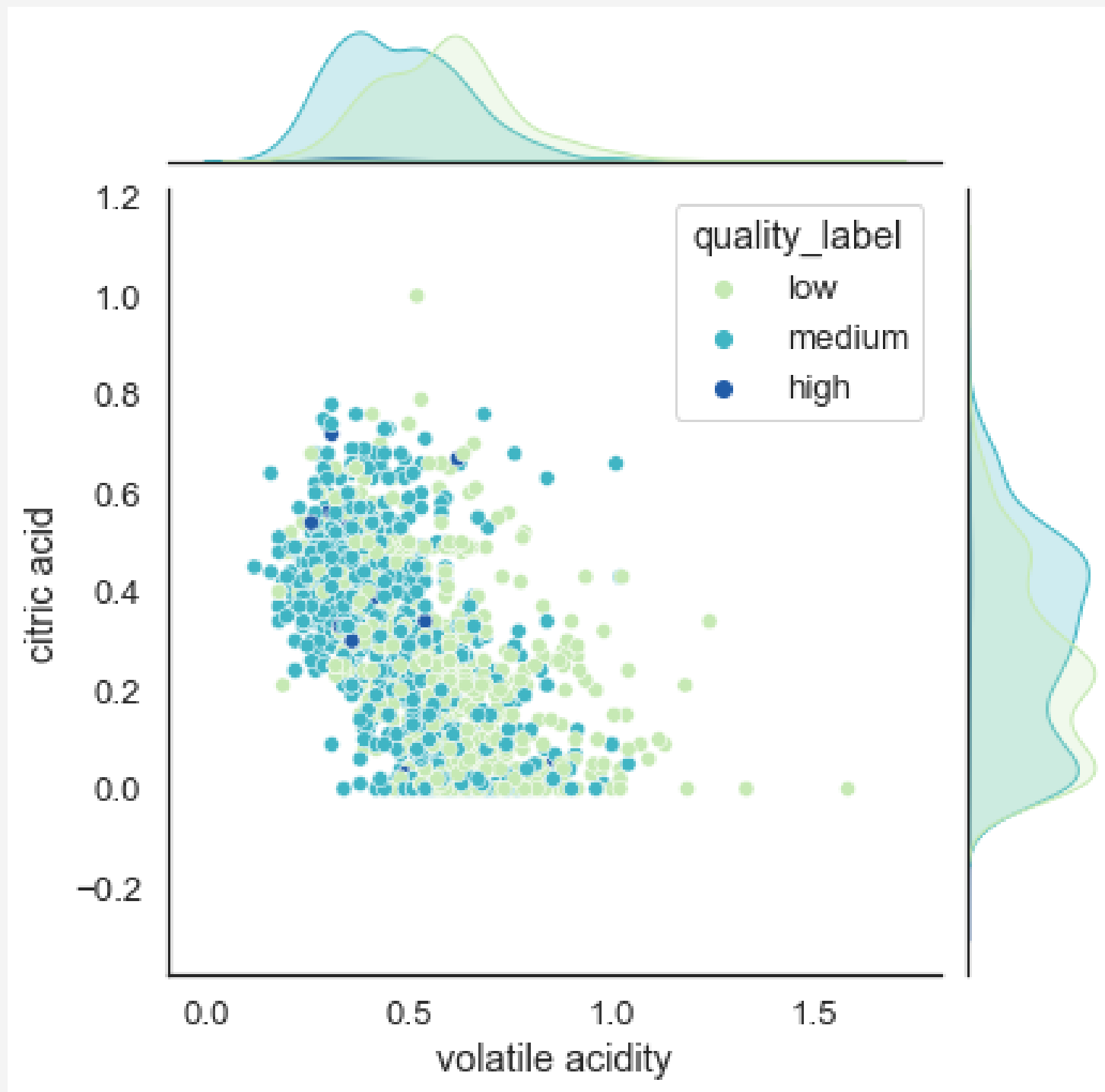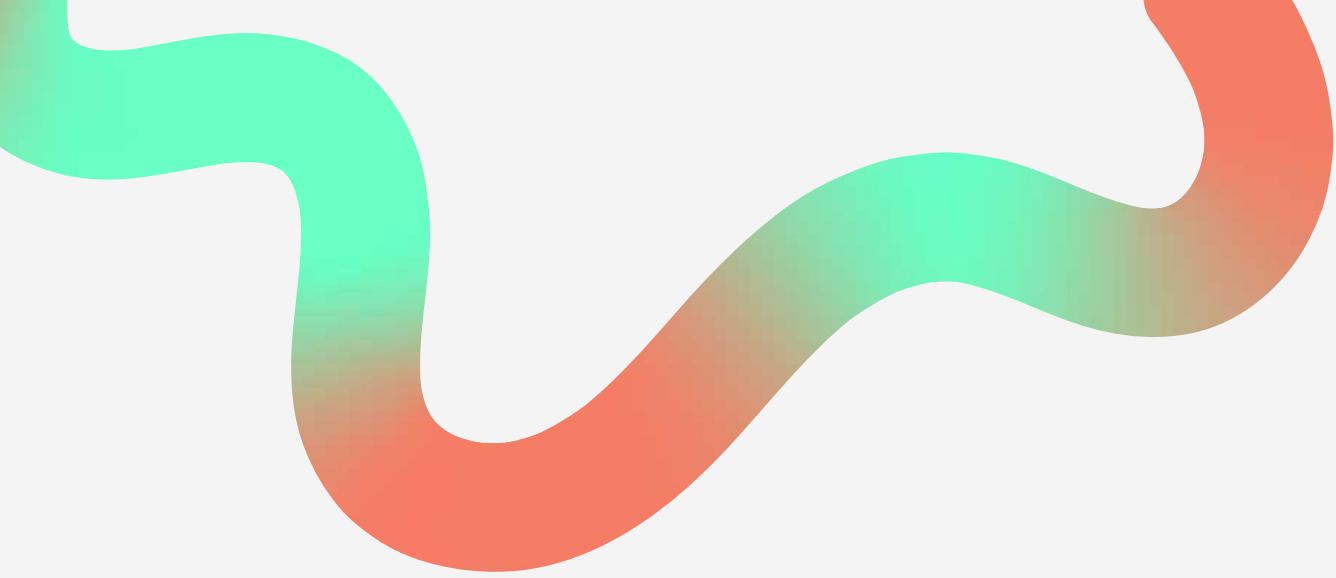
Feature Importance plot

Legend:

- [0]: fixed acidity
- [1]: volatile acidity
- [2]: citric acid
- [3]: residual sugar
- [4]: chlorides
- [5]: free sulfur dioxide
- [6]: total sulfur dioxide
- [7]: density
- [8]: pH
- [9]: sulphates
- [10]: alcohol
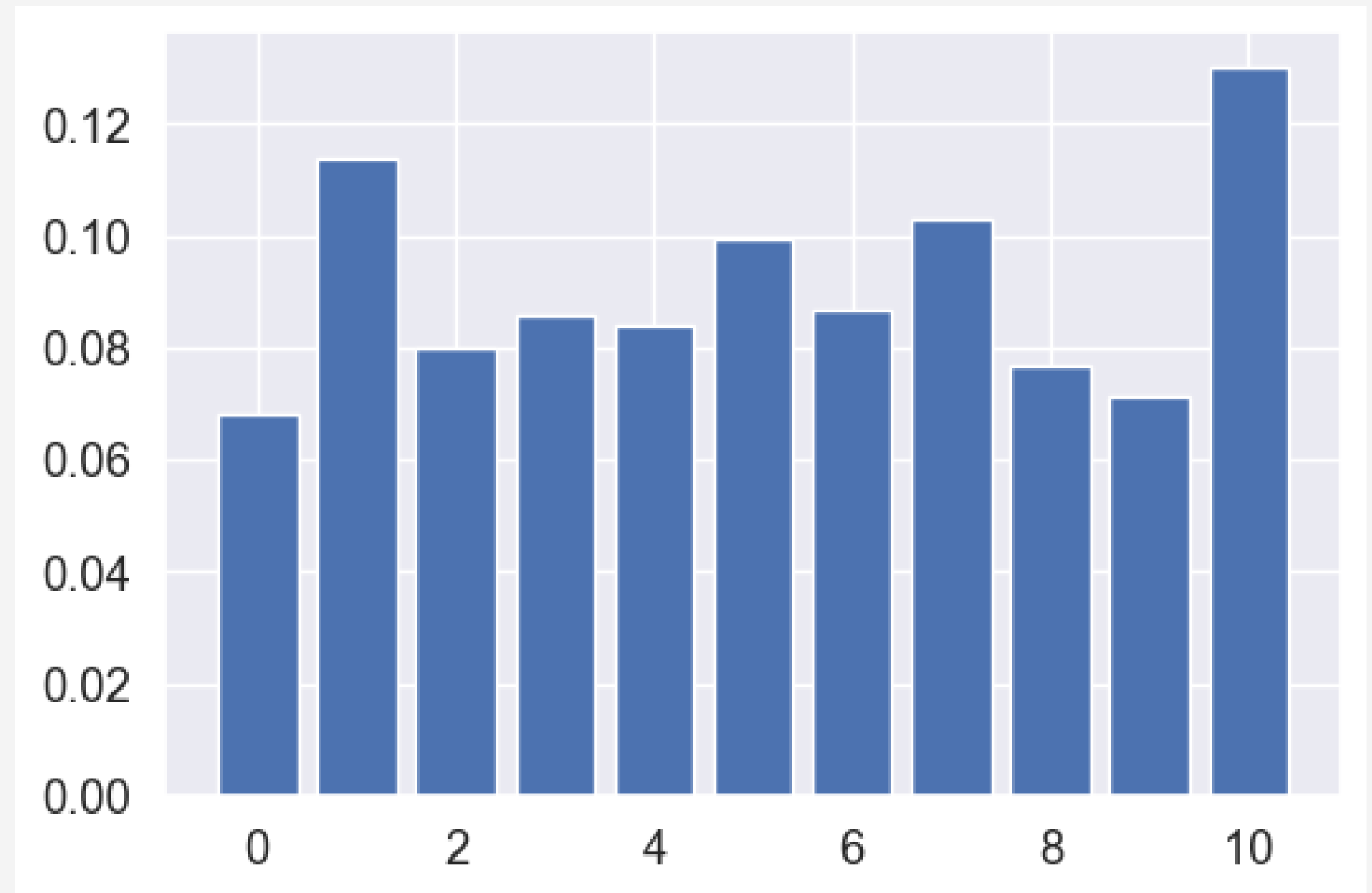
Confusion Matrix

based on Random Forest Classifier model
after dropping any feature
[free sulfur dioxide, chlorides, sulphates, fixed acidity, density]
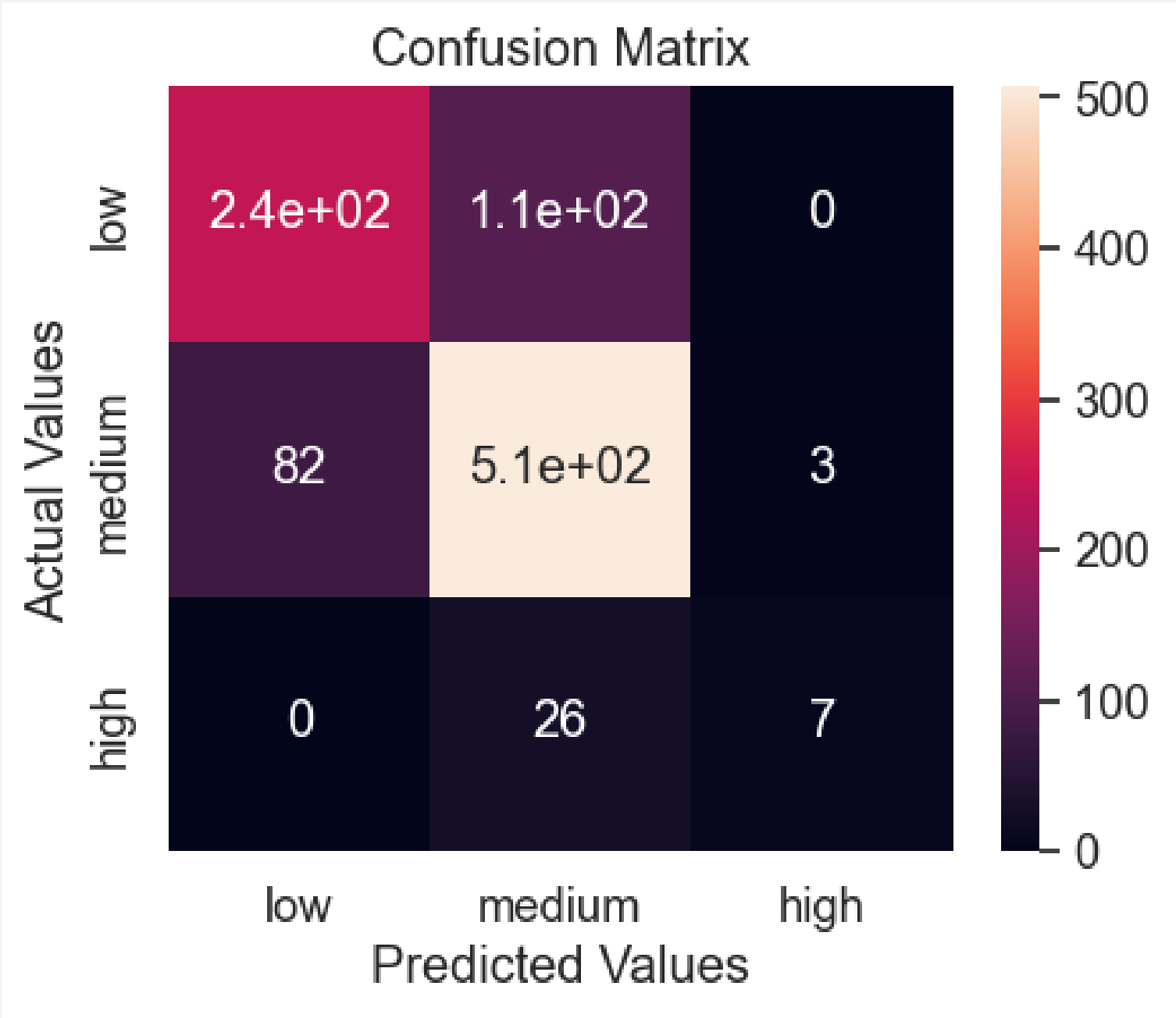

```
confusion_matrix:
 [[243 112   0]
 [ 82 507   3]
 [  0  26   7]]
Accuracy score: 0.7724489795918368
              precision    recall  f1-score   support

         low       0.75      0.68      0.71       355
      medium       0.79      0.86      0.82       592
        high       0.70      0.21      0.33        33

    accuracy                           0.77       980
   macro avg       0.74      0.58      0.62       980
weighted avg       0.77      0.77      0.77       980


Cohen-Kappa score: 0.5278434931025915
k-fold cross-validator: 0.6911218229623138
```

17

Accuracy & Confusion Matrix

based on Random Forest Classifier model
applied to Randomized Search CV, Grid Search CV

```
Based on RandomizedSearchCV :


Accouracy:  0.789795918367347
Confusion matrix:
  [[229 126    0]
 [ 51 538    3]
 [  0  26    7]]
```

```
Based on GridSearchCV :


Accouracy:  0.767346938775102
Confusion matrix:
  [[216 139    0]
 [ 55 536    1]
 [  0  33    0]]
```

18

Red Wine Accuracy Score
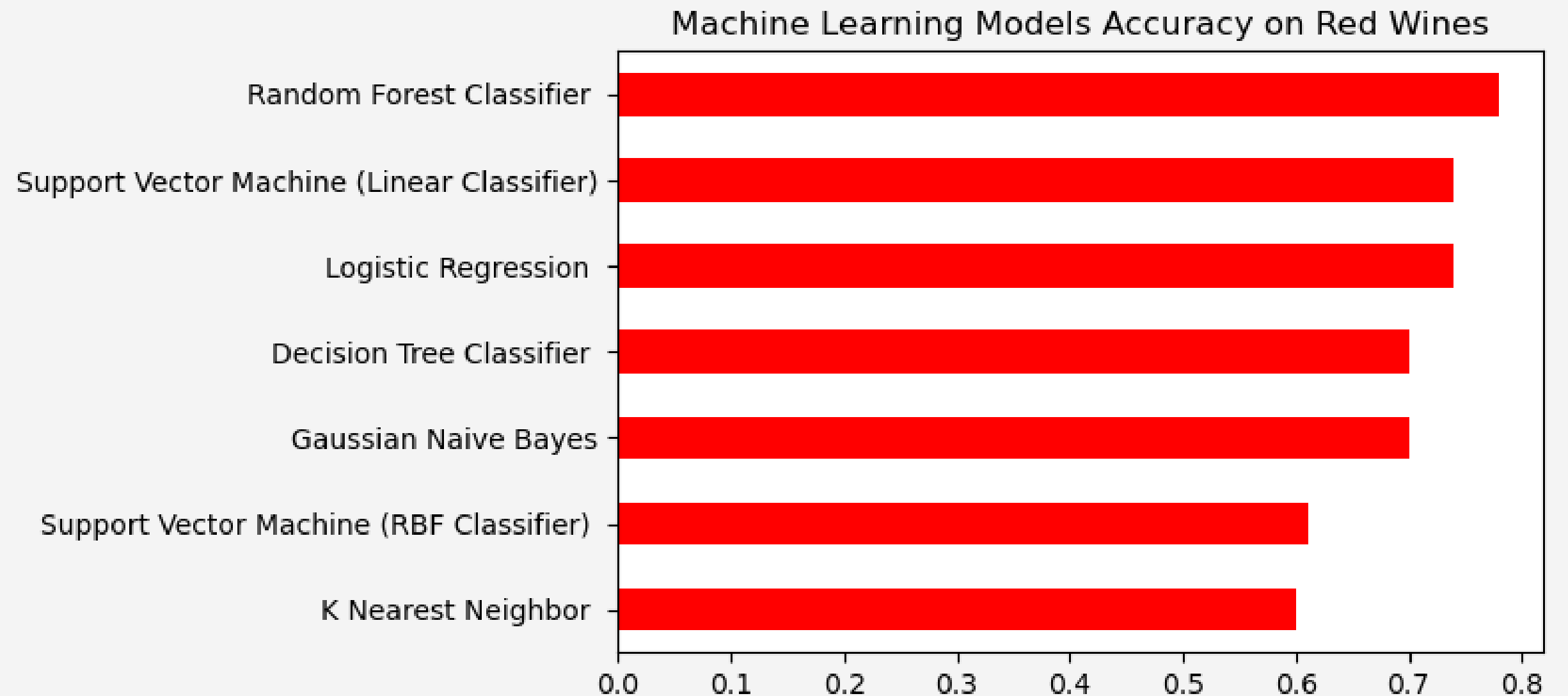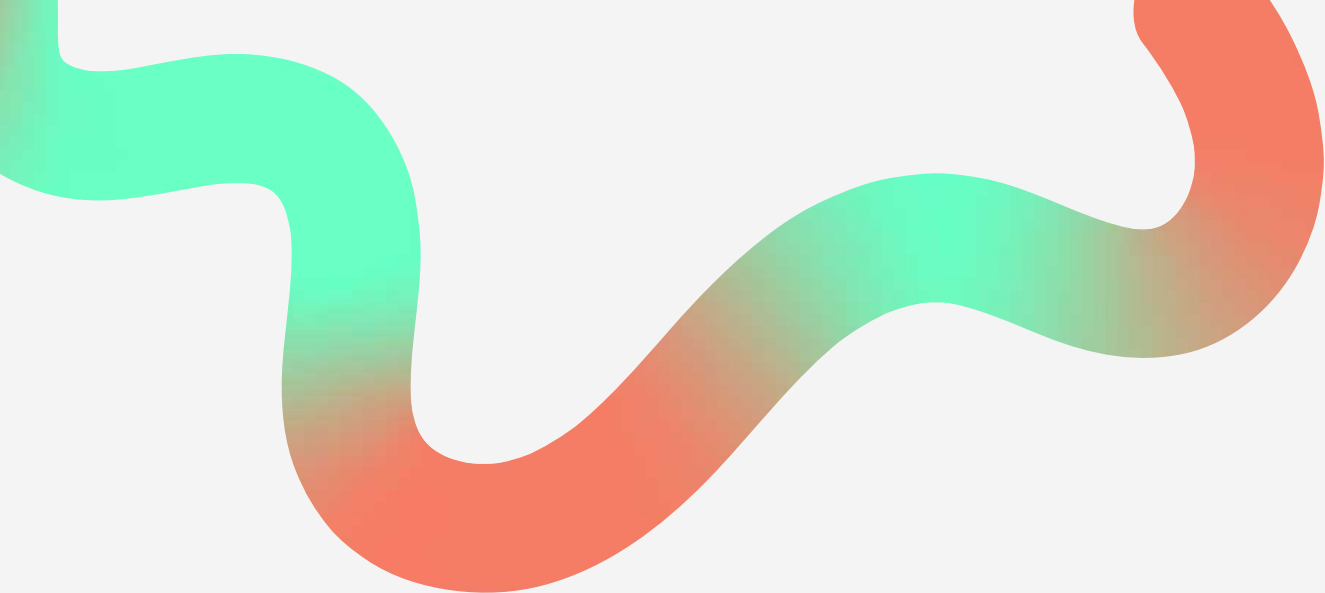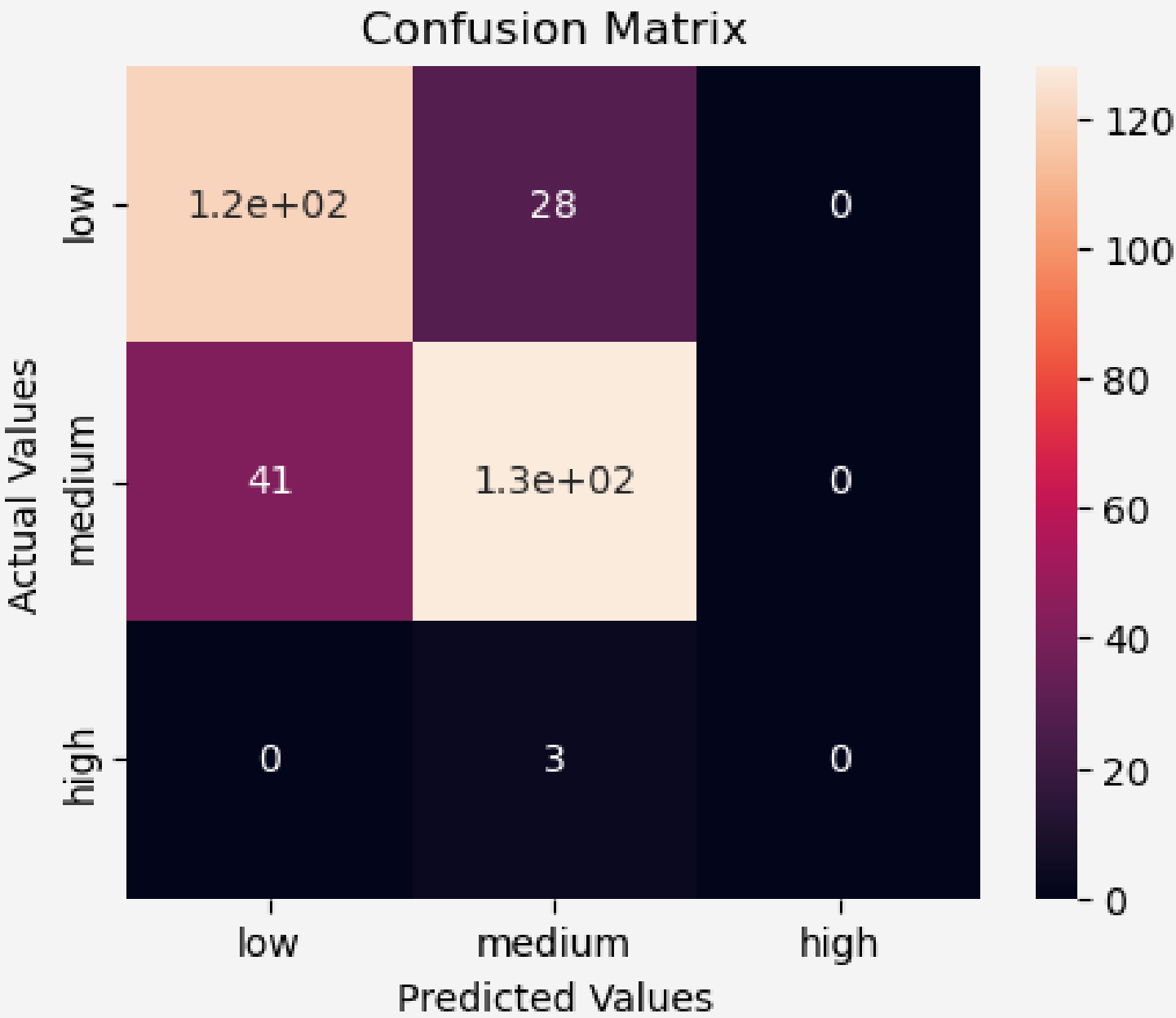
The accuracy for Random Forest is

- The training is: 100.0 %
- The testing is: 80.0 %

Machine Learning Models Accuracy on Red Wines

Confusion Matrix

based on Random Forest Classifier model before dropping any feature

Confusion Matrix

```
confusion_matrix:
  [[115  33   0]
  [ 36 133   0]
  [  0   3   0]]
Accuracy score: 0.775
                precision    recall  f1-score   support

         low       0.76      0.78      0.77       148
      medium       0.79      0.79      0.79       169
        high       0.00      0.00      0.00         3

    accuracy                           0.78       320
   macro avg       0.52      0.52      0.52       320
weighted avg       0.77      0.78      0.77       320


Cohen-Kappa score: 0.5525431628828339
```
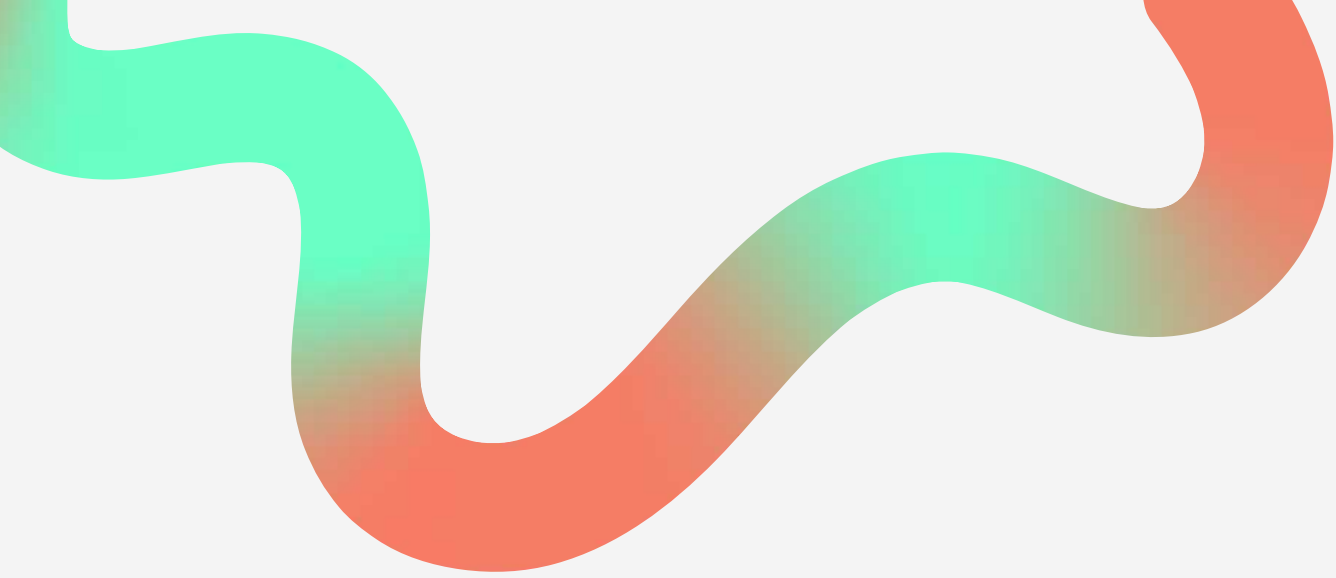
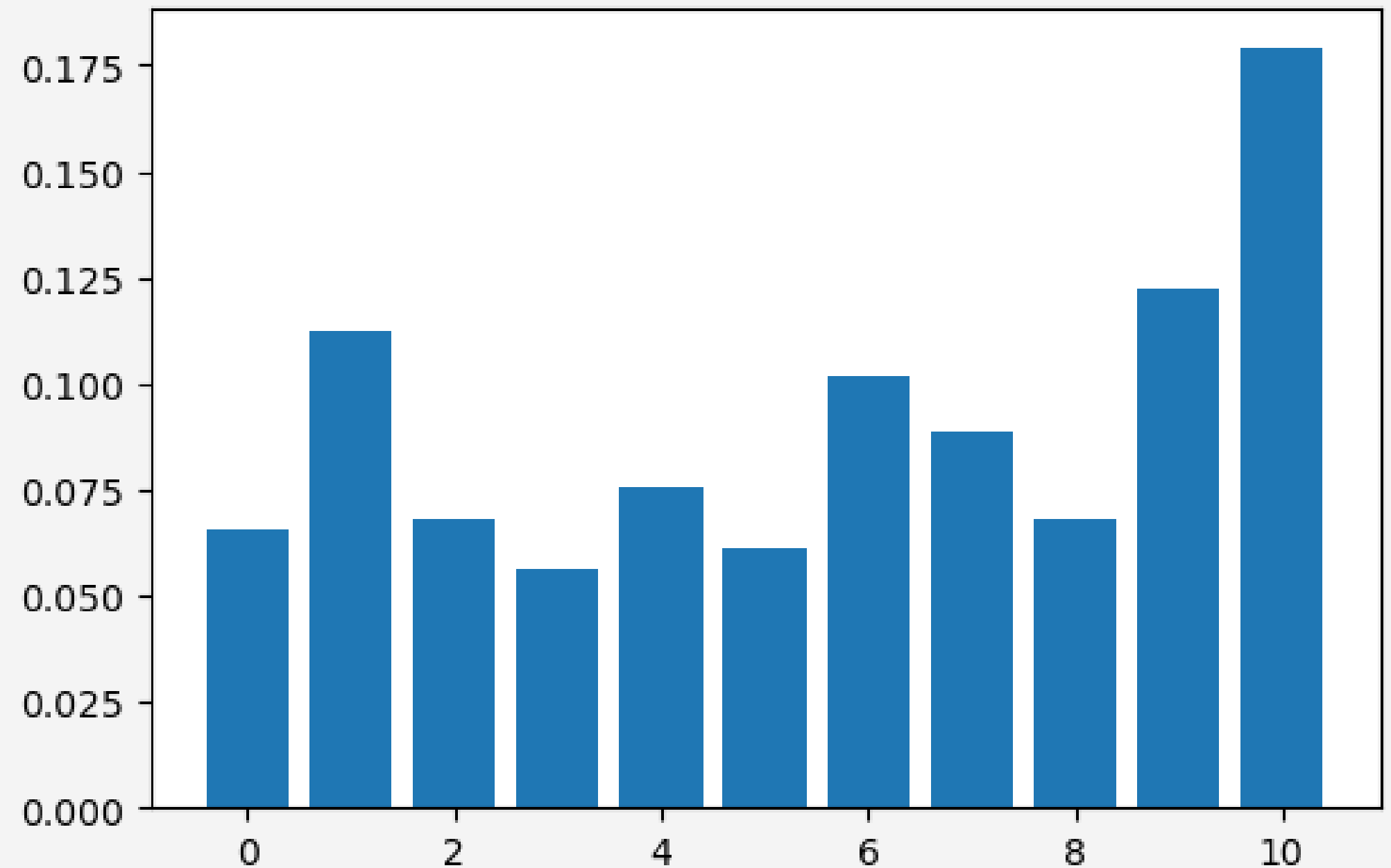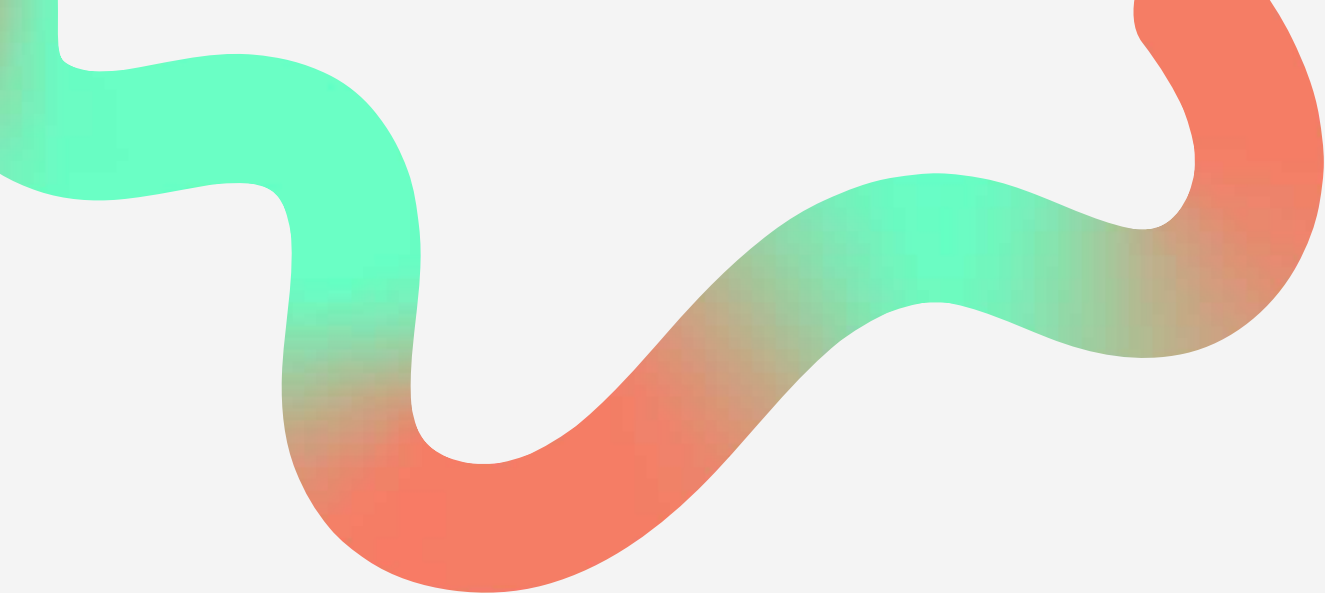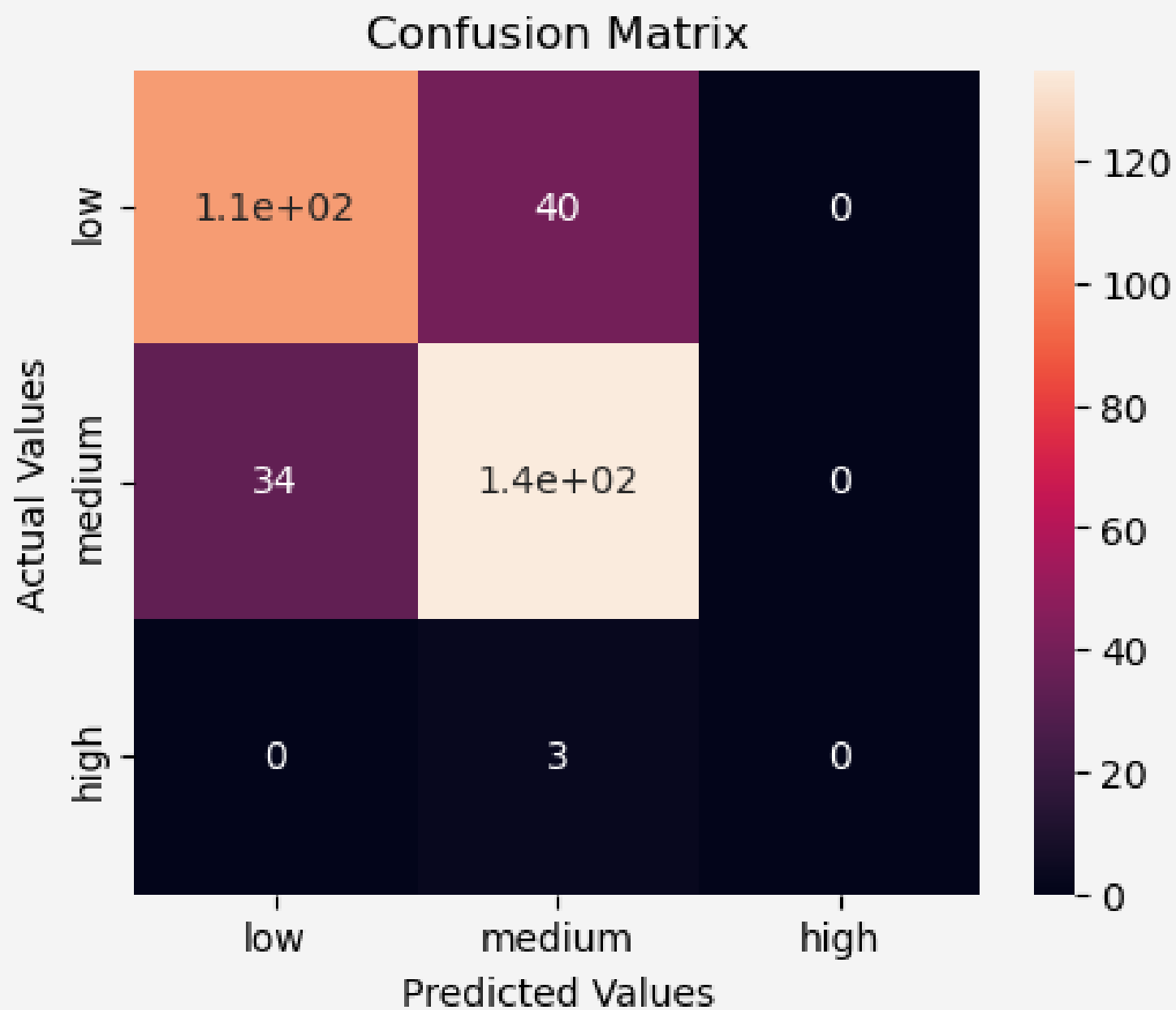# Feature Importance plot

Legend:

- [0]: fixed acidity
- [1]: volatile acidity
- [2]: citric acid
- [3]: residual sugar
- [4]: chlorides
- [5]: free sulfur dioxide
- [6]: total sulfur dioxide
- [7]: density
- [8]: pH
- [9]: sulphates
- [10]: alcohol

Confusion Matrix

based on Random Forest Classifier model
after dropping any feature
[free sulfur dioxide, citric acid, pH, fixed acidity, residual sugar]

Confusion Matrix

```
confusion_matrix:
 [[108  40   0]
 [ 34 135   0]
 [  0   3   0]]
Accuracy score: 0.759375
              precision    recall  f1-score   support

         low       0.76      0.73      0.74       148
      medium       0.76      0.80      0.78       169
        high       0.00      0.00      0.00         3

    accuracy                           0.76       320
   macro avg       0.51      0.51      0.51       320
weighted avg       0.75      0.76      0.76       320


Cohen-Kappa score: 0.5197068340415578
```

22

Thank You!

Cheers

Muayad Sirafy

Junior Data Analyst