Wine Quality Analysis

This report provides an analysis of the Wine Quality dataset. The dataset is divided into two parts: red wine and white wine.

1. Load the Dataset

We began by loading the datasets for red and white wine.

2. Data Cleaning and Preparation

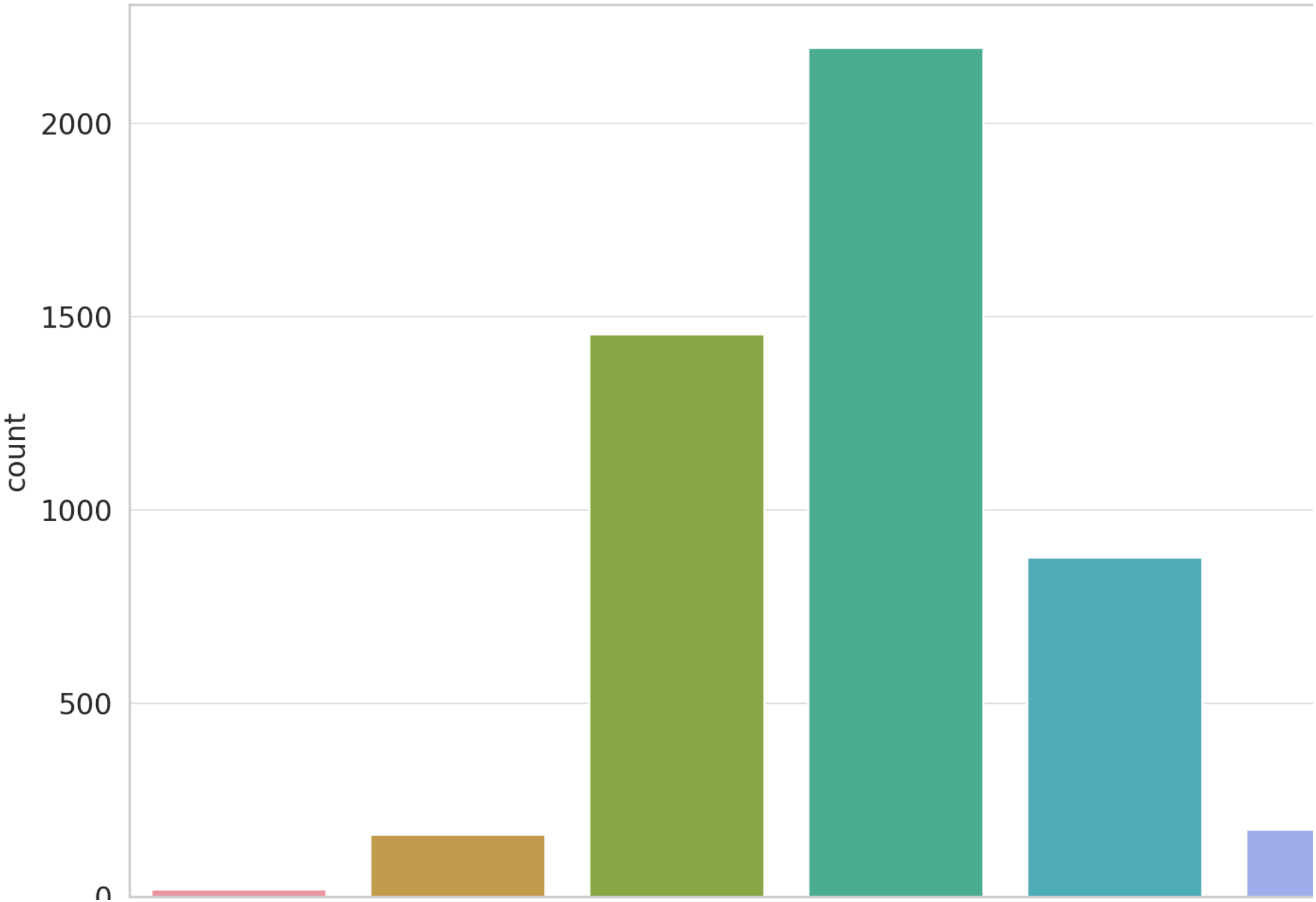
After loading the data, we examined it for any missing values, outliers, or inconsistencies and handled them accordingly. Both the red and white wine datasets did not have any missing values.

3. Data Exploration

Using visualizations and statistics, we got a sense of the data's distributions, relationships, and potential patterns.

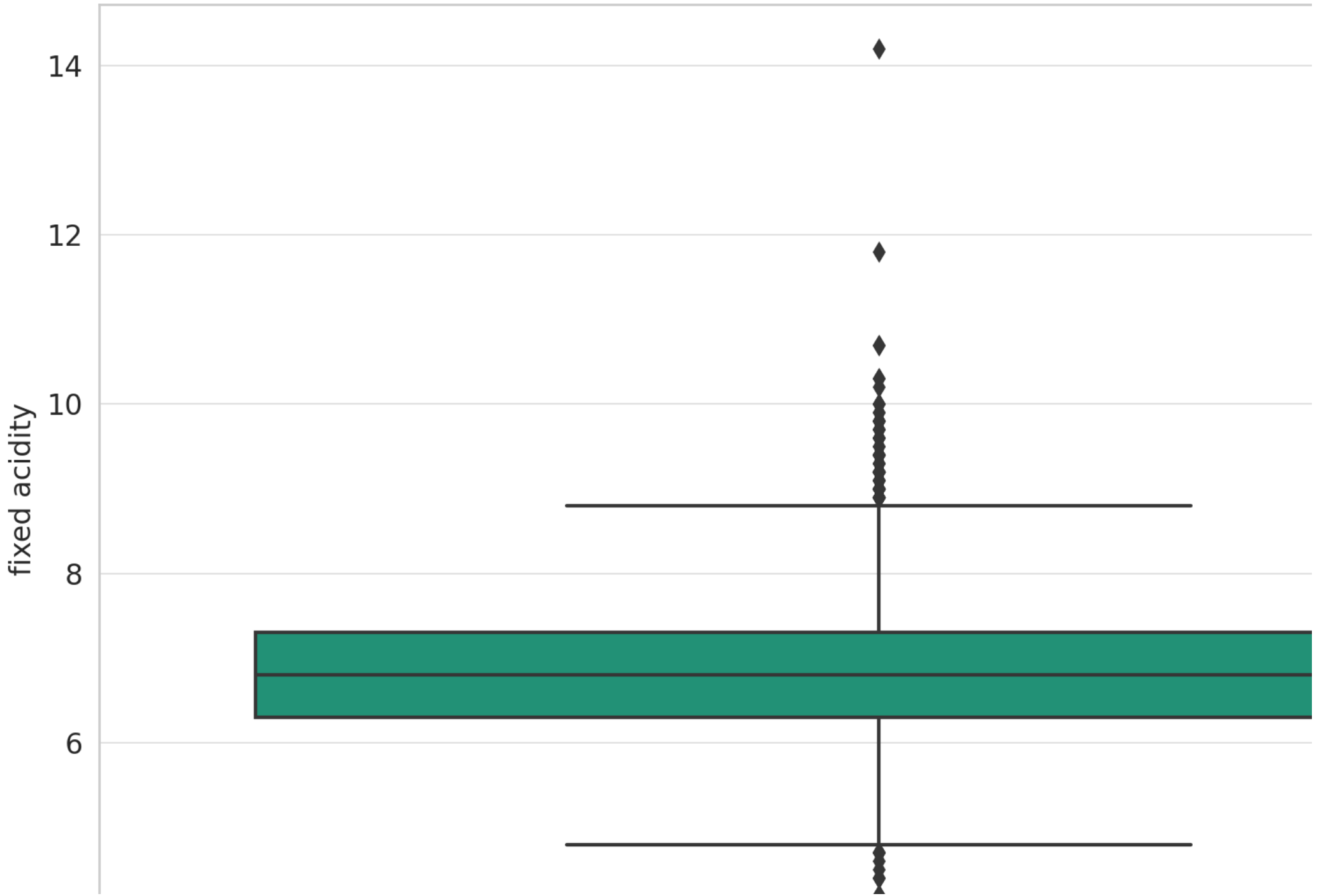
3.1 Distribution of Wine Quality Scores

The majority of wines (both red and white) have quality scores between 5 and 6. Few wines have scores of 3, 8, or 9.

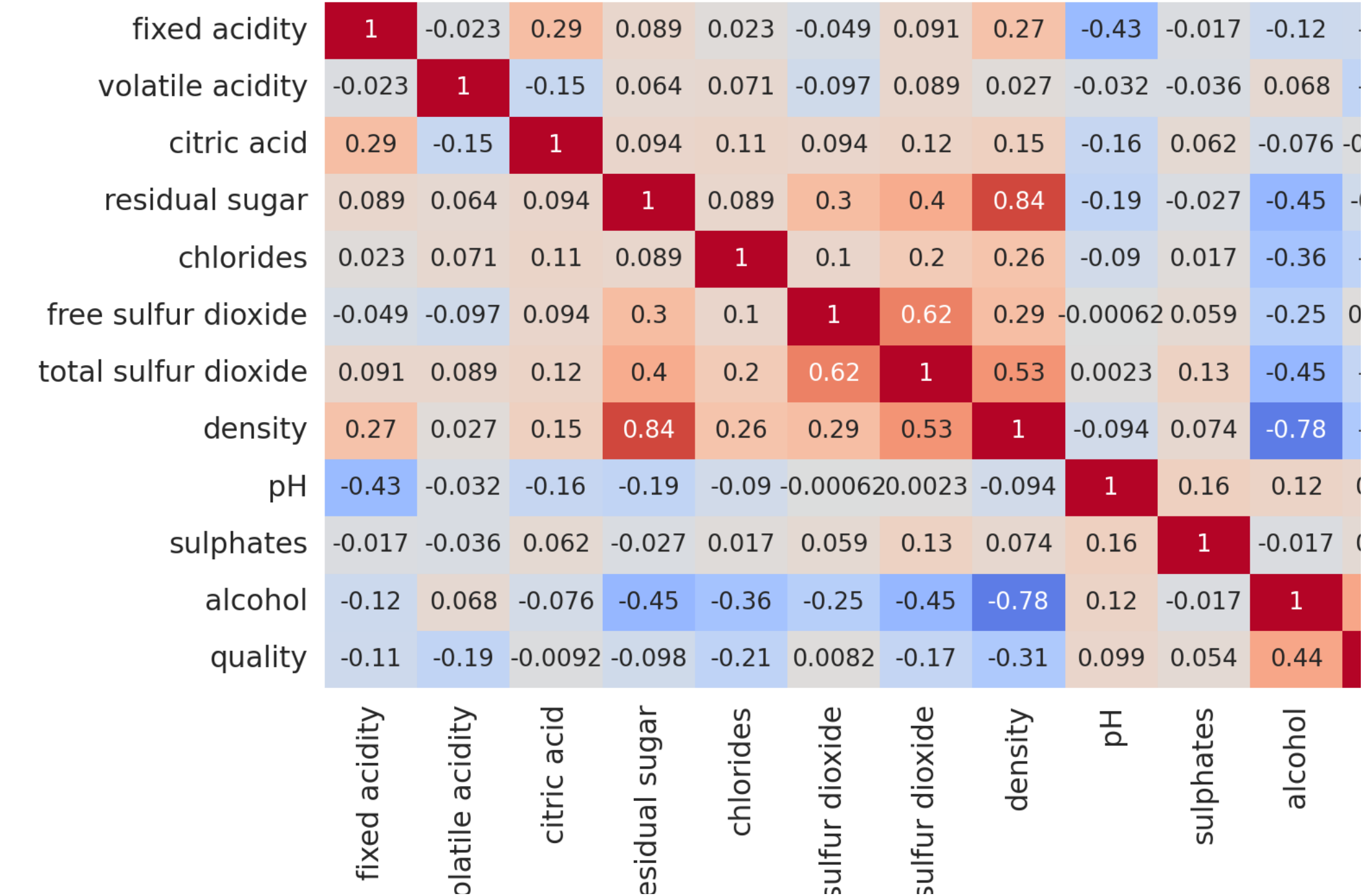

count
2000
1500
1000
500
0

3.2 Relationships Between Features and Wine Quality

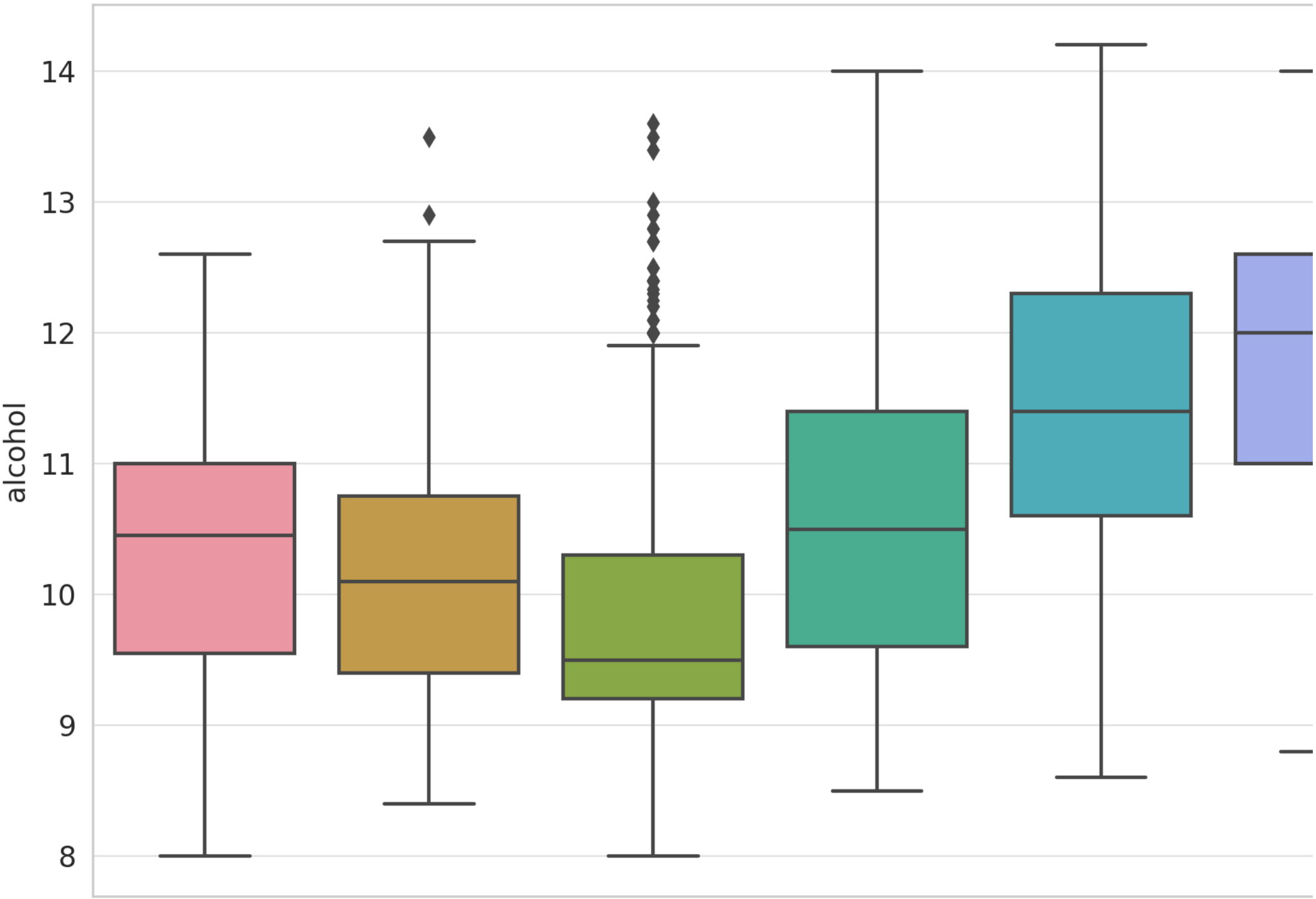
We visualized the distributions of various features to get a sense of their spread and potential outliers.


14
12
10
8
6
fixed acidity

Furthermore, we investigated the relationships between features and wine quality using correlation matrices and visualizations.

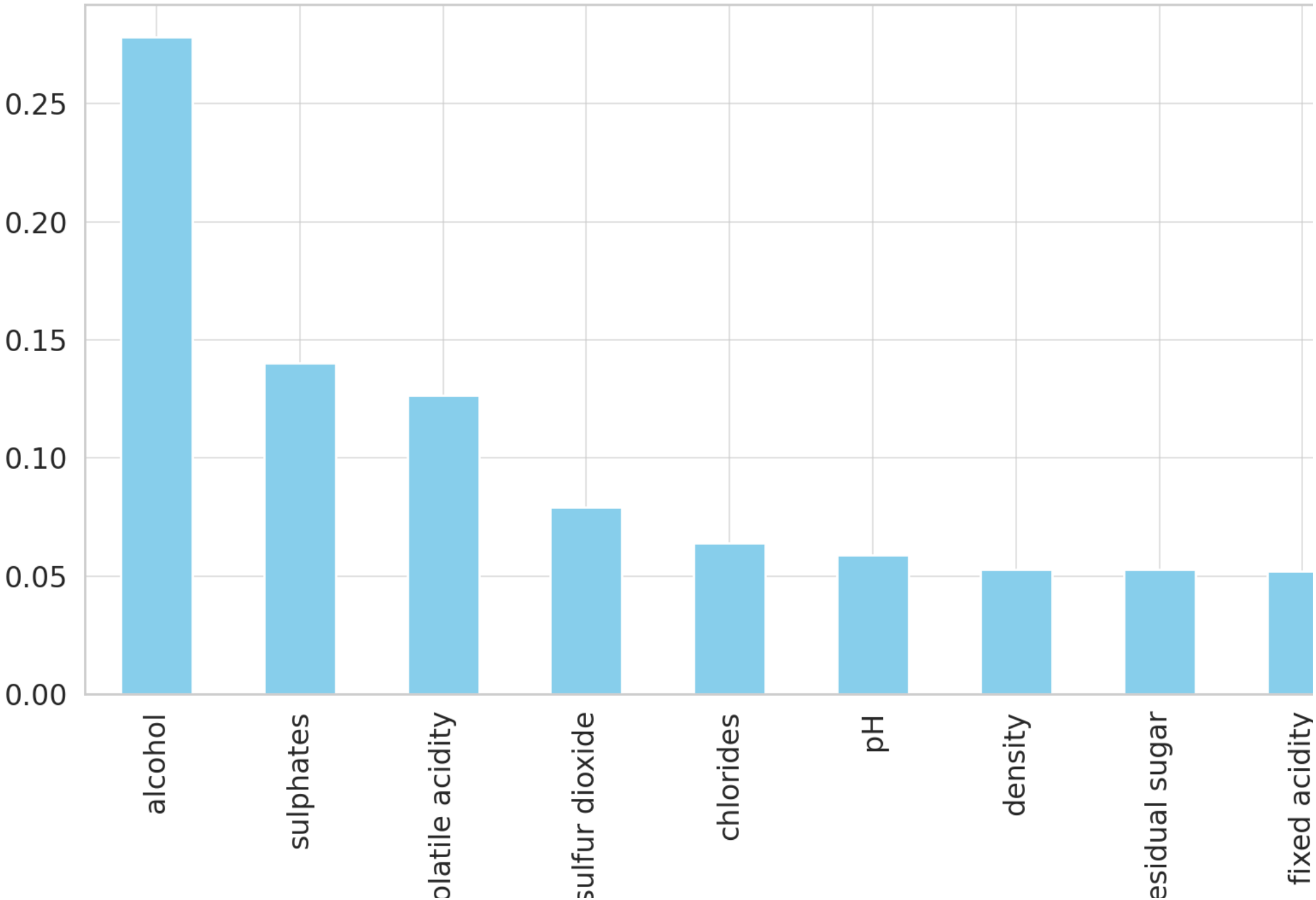

fixed acidity 1 -0.023 0.29 0.089 0.023 -0.049 0.091 0.27 -0.43 -0.017 -0.12
volatile acidity -0.023 1 -0.15 0.064 0.071 -0.097 0.089 0.027 -0.032 -0.036 0.068
citric acid 0.29 -0.15 1 0.094 0.11 0.094 0.12 0.15 -0.16 0.062 -0.076
residual sugar 0.089 0.064 0.094 1 0.089 0.3 0.4 0.84 -0.19 -0.027 -0.45
chlorides 0.023 0.071 0.11 0.089 1 0.1 0.2 0.26 -0.09 0.017 -0.36
free sulfur dioxide -0.049 -0.097 0.094 0.3 0.1 1 0.62 0.29 -0.00062 0.059 -0.25
total sulfur dioxide 0.091 0.089 0.12 0.4 0.2 0.62 1 0.53 0.0023 0.13 -0.45
density 0.27 0.027 0.15 0.84 0.26 0.29 0.53 1 -0.094 0.074 -0.78
pH -0.43 -0.032 -0.16 -0.19 -0.09 -0.00062 0.0023 -0.094 1 0.16 0.12
sulphates -0.017 -0.036 0.062 -0.027 0.017 0.059 0.13 0.074 0.16 1 -0.017
alcohol -0.12 0.068 -0.076 -0.45 -0.36 -0.25 -0.45 -0.78 0.12 -0.017 1
quality -0.11 -0.19 -0.0092 -0.098 -0.21 0.0082 -0.17 -0.31 0.099 0.054 0.44
fixed acidity
citric acid
chlorides
density
pH
sulphates
alcohol

free
total


alcohol
14
13
12
11
10
9
8

3.3 Important Factors Influencing Wine Quality

Alcohol, volatile acidity, and sulphates are among the most influential factors determining wine quality for red wines. This was derived using a Random Forest regressor to extract feature importances.


0.25
0.20
0.15
0.10
0.05
0.00
alcohol
sulphates
olatile acidity
ulfur dioxide
chlorides
pH
density
esidual sugar
fixed acidity

End of analysis.