Wine Quality Analysis

# 1. Introduction and Dataset Loading

We started by loading the provided datasets for red and white wines. The datasets contain the following columns:

fixed acidity

volatile acidity

citric acid

residual sugar

chlorides

free sulfur dioxide

total sulfur dioxide

density

pH

sulphates

alcohol

quality

Below is a glimpse of the loaded data:

fixed acidity volatile acidity citric acid residual sugar chlorides \  
0 7.4 0.70 0.00 1.9 0.076   
1 7.8 0.88 0.00 2.6 0.098   
2 7.8 0.76 0.04 2.3 0.092   
3 11.2 0.28 0.56 1.9 0.075   
5 7.4 0.66 0.00 1.8 0.075   
  
 free sulfur dioxide total sulfur dioxide density pH sulphates \  
0 11.0 34.0 0.9978 3.51 0.56   
1 25.0 67.0 0.9968 3.20 0.68   
2 15.0 54.0 0.9970 3.26 0.65   
3 17.0 60.0 0.9980 3.16 0.58   
5 13.0 40.0 0.9978 3.51 0.56   
  
 alcohol quality   
0 9.4 5   
1 9.8 5   
2 9.8 5   
3 9.8 6   
5 9.4 5

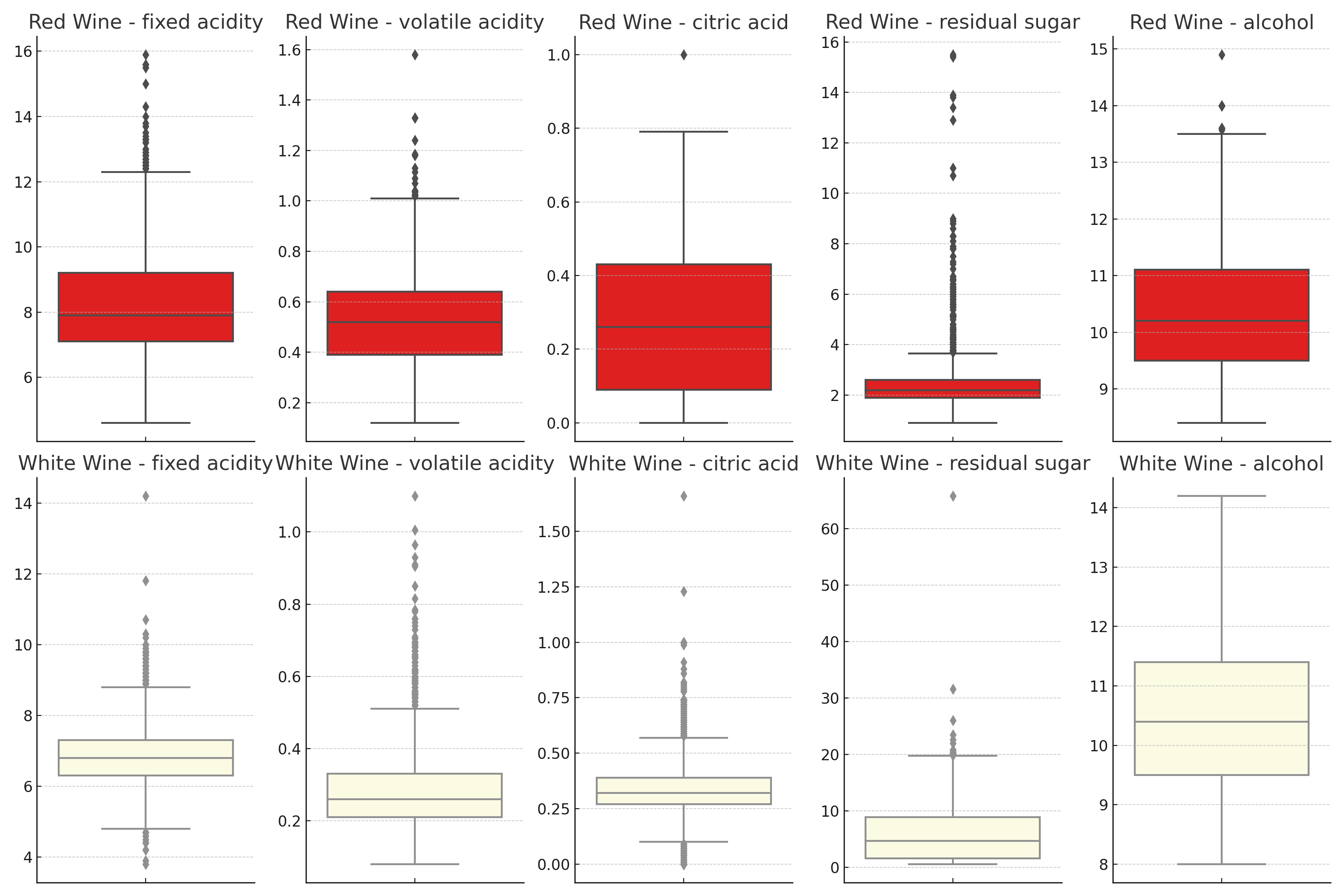
# 2. Data Cleaning

The data was inspected for duplicates, missing values, and outliers.

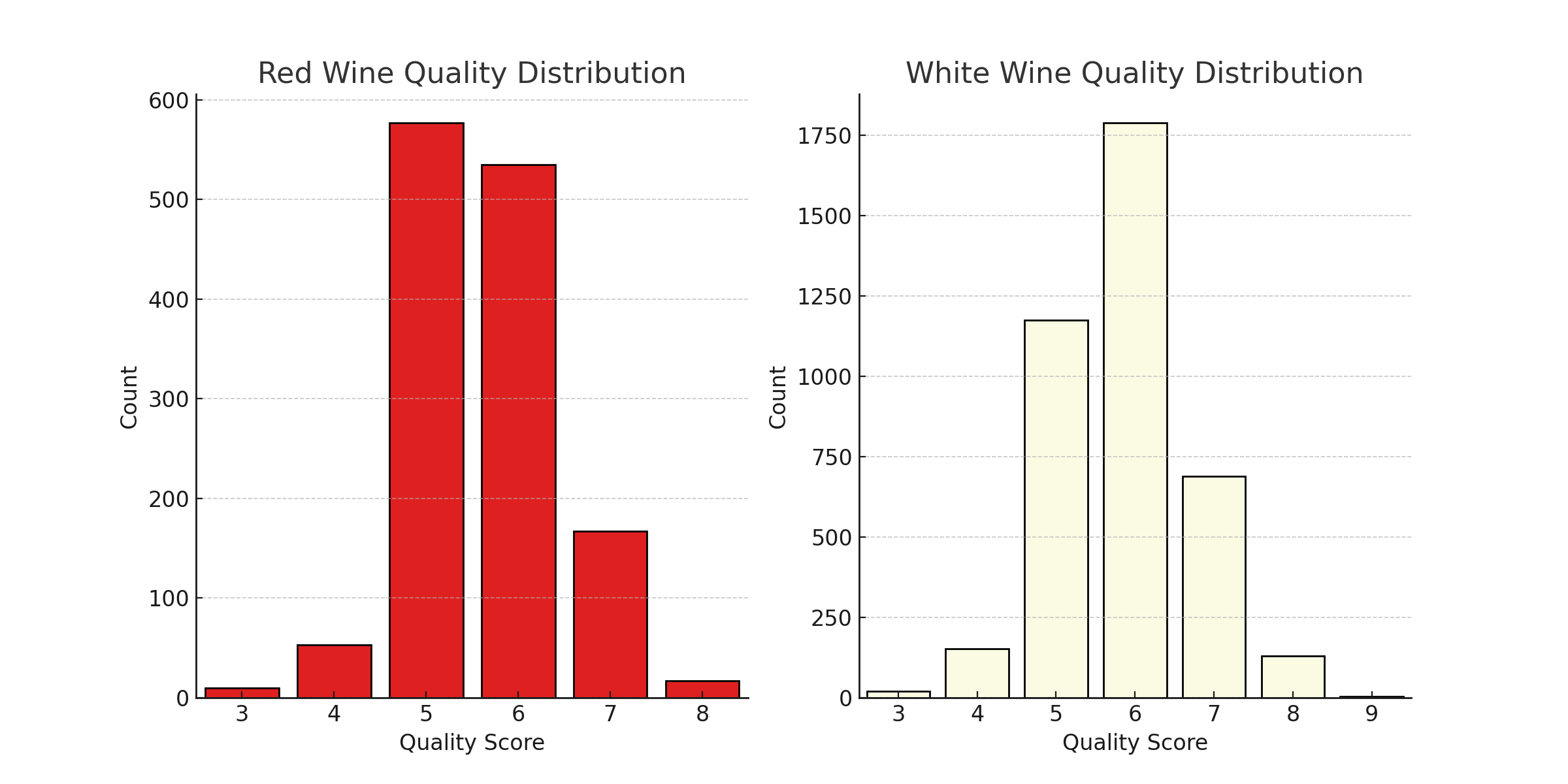
Duplicates detected: Red wine - 240, White wine - 937

Missing values: None detected in either dataset.

Outliers: Were visually identified using box plots.



The distribution of wine quality scores was visualized for both red and white wines.



The relationship between different features and wine quality was explored using correlation analysis.

