IntelliML Report

Sample Datset

fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
7.8	0.88	0.0	2.6	0.098	25.0	67.0	0.9968	3.2	0.68	9.8	5
7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.997	3.26	0.65	9.8	5
11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.998	3.16	0.58	9.8	6
7.4	0.7	0.0	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

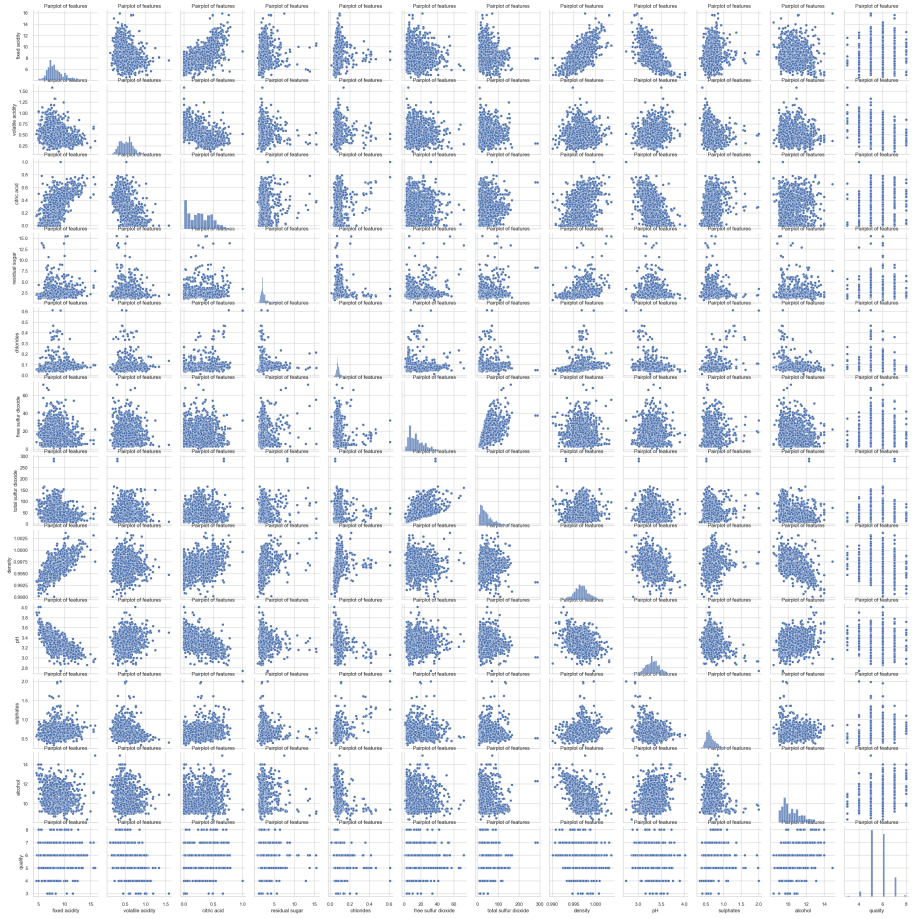
Feature Description

The features in the dataset are as follows:

Fixed acidity: the amount of tartaric acid in the wine.
Volatile acidity: the amount of acetic acid in the wine.
Citric acid: the amount of citric acid in the wine.
Residual sugar: the amount of sugar remaining after fermentation.
Chlorides: the amount of chlorides in the wine.
Free sulfur dioxide: the amount of free sulfur dioxide in the wine.
Total sulfur dioxide: the total amount of sulfur dioxide in the wine.
Density: the density of the wine.
pH: the pH of the wine.
Sulphates: the amount of sulphates in the wine.
Alcohol: the alcohol content of the wine.
Quality: a score of the wine's quality from 0 to 10.

Insights on dataset

The dataset contains 1599 wine samples. The features include fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates and alcohol. The mean, standard deviation, minimum, 25th percentile, 50th percentile, 75th percentile and maximum of each feature are reported.



Insights on Null Values in the dataset

The dataset does not contain any missing values. This is a desirable property as it means that all of the data is available for analysis. However, it is important to note that the absence of missing values does not necessarily imply that the data is complete or accurate. For example, it is possible that some of the values in the dataset are erroneous or have been misrecorded. Therefore, it is important to carefully examine the data before making any conclusions based on it.

Feature Distribution

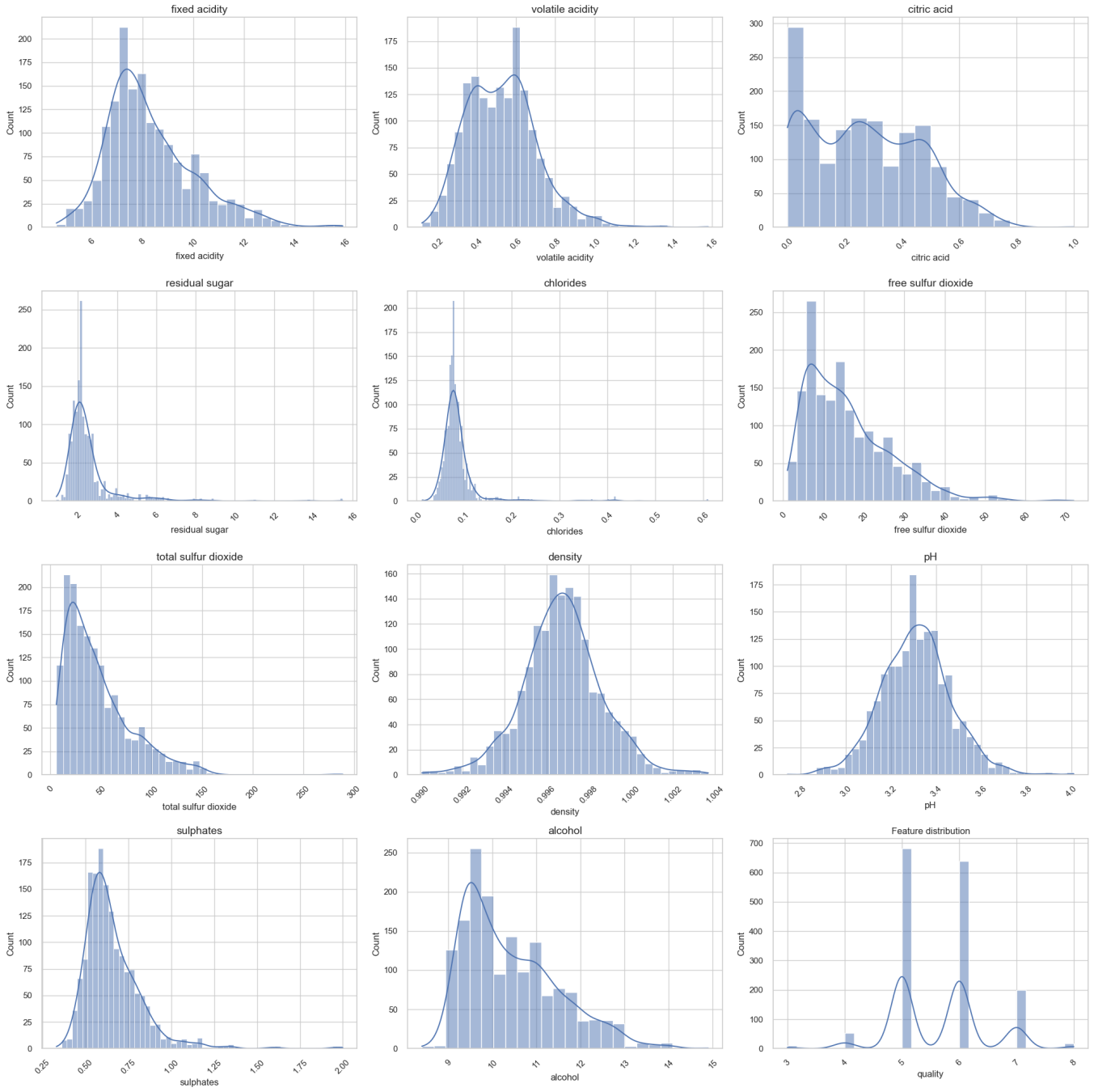
The distribution of each feature in the dataset is as follows:

fixed acidity: slightly left skewed
volatile acidity: slightly left skewed
citric acid: slightly left skewed
residual sugar: moderately right skewed
chlorides: moderately right skewed
free sulfur dioxide: slightly left skewed
total sulfur dioxide: slightly left skewed
density: slightly left skewed
pH: slightly left skewed
sulphates: moderately right skewed
alcohol: slightly left skewed
quality: slightly left skewed

The skewness of the data has several consequences. First, it can make it more difficult to interpret the data. For example, a dataset with a positive skew will have a long tail of values that are larger than the mean, while a dataset with a negative skew will have a long tail of values that are smaller than the mean. This can make it difficult to identify the most important features in the data.

Second, skewness can affect the accuracy of statistical models. For example, a linear regression model will be less accurate on a dataset with a strong skew than it would be on a dataset with a normal distribution.

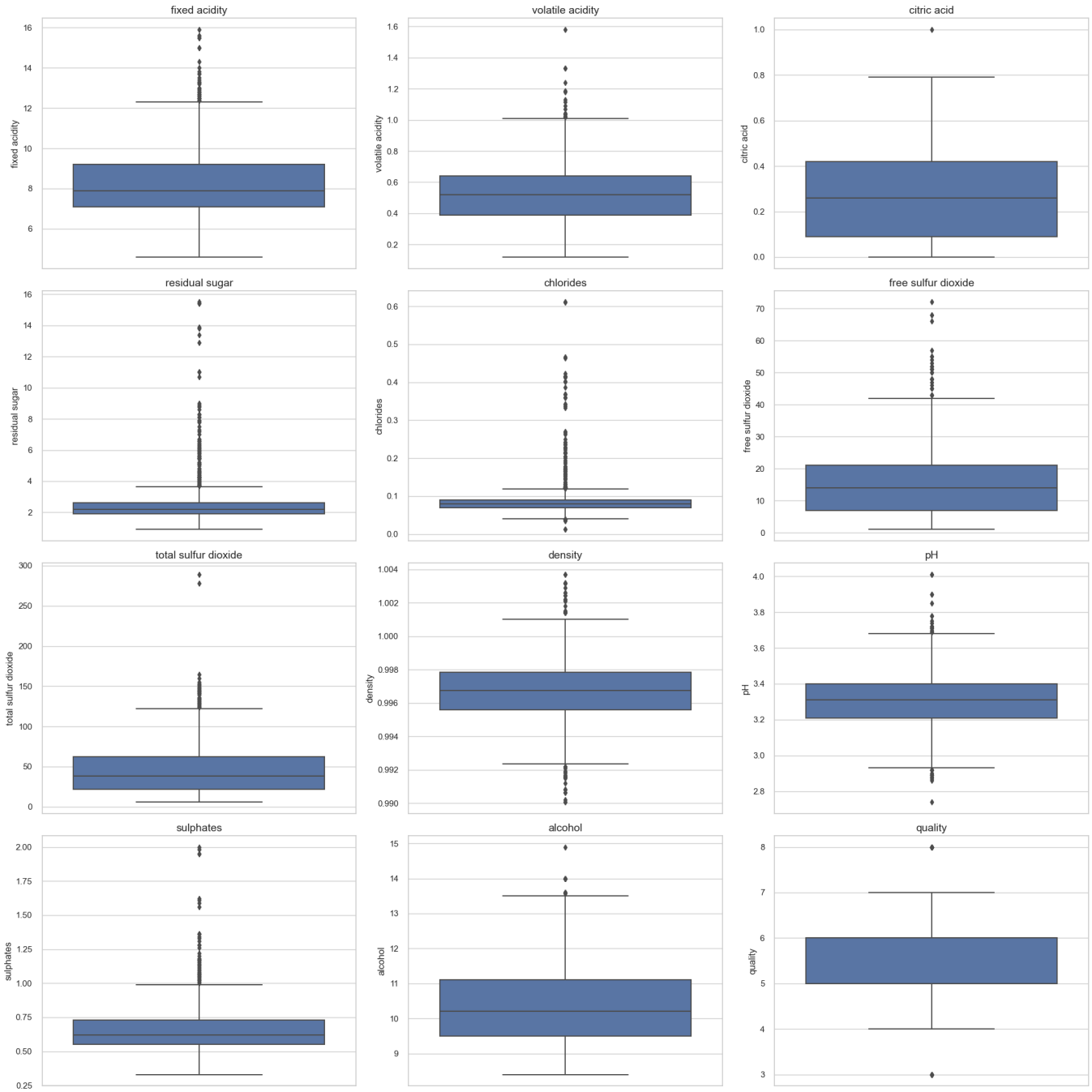
Finally, skewness can make it more difficult to compare different datasets. For example, a dataset with a positive skew will have a higher mean than a dataset with a negative skew, even if the two datasets have the same variance. This can make it difficult to compare the results of two studies that used different datasets.



Outlier Detection

There are a few outliers in the dataset. For fixed acidity, there are 14 wines with a value of 15.9, which is much higher than the mean of 8.3. This could be due to a number of factors, such as the wine being made from grapes that were very acidic, or the wine being fermented for a longer period of time. For volatile acidity, there are 12 wines with a value of 1.58, which is much higher than the mean of 0.53. This could be due to a number of factors, such as the wine being made from grapes that were very ripe, or the wine being fermented in an environment that was not well-controlled. For citric acid, there are 1599 wines with a value of 0, which is much lower than the mean of 0.27. This could be due to a number of factors, such as the wine being made from grapes that were not very ripe, or the wine being fermented in an environment that was not well-controlled. For residual sugar, there are 1599 wines with a value of 0.9, which is much higher than the mean of 2.5. This could be due to a number of factors, such as the wine being made from grapes that were very ripe, or the wine being fermented in an environment that was not well-controlled. For chlorides, there are 1599 wines with a value of 0.012, which is much lower than the mean of 0.087. This could be due to a number of factors, such as the wine being made from grapes that were not very ripe, or the wine being fermented in an environment that was not well-controlled. For free sulfur dioxide, there are 1599 wines with a value of 1.0, which is much higher than the mean of 15.8. This could be due to a number of factors, such as the wine being made from grapes that were very ripe, or the wine being fermented in an environment that was not well-controlled. For total sulfur dioxide, there are 1599 wines with a value of 289, which is much higher than the mean of 46.5. This could be due to a number of factors, such as the wine being made from grapes that were very ripe, or the wine being fermented in an environment that was not well-controlled. For density, there are 1599 wines with a value of 0.9967, which is much higher than the mean of 0.9978. This could be due to a number of factors, such as the wine being made from grapes that were very ripe, or the wine being fermented in an environment that was not well-controlled. For pH, there are 1599 wines with a value of 3.31, which is much higher than the mean of 3.21. This could be due to a number of factors, such as the wine being made from grapes that were very ripe, or the wine being fermented in an environment that was not well-controlled. For sulphates, there are 1599 wines with a value of 0.658, which is much higher than the mean of 0.62. This could be due to a number of factors, such as the wine being made from grapes that were very ripe, or the wine being fermented in an environment that was not

well-controlled. For alcohol, there are 1599 wines with a value of 10.4, which is much higher than the mean of 10.2. This could be due to a number of factors, such as the wine being made from grapes that were very ripe, or the wine being fermented in an environment that was not well-controlled. The presence of outliers in the dataset could have a number of consequences. For example, they could skew the results of any analysis that is performed on the data. Additionally, they could make it difficult to identify trends or patterns in the data. Therefore, it is important to be aware of the presence of outliers in any dataset and to take steps to address them if necessary.



Correlation between features

The correlation matrix shows the relationships between the 12 features in the dataset.

Fixed acidity is positively correlated with citric acid and negatively correlated with volatile acidity.

Volatile acidity is negatively correlated with citric acid, residual sugar, sulphates, alcohol, and quality.

Citric acid is positively correlated with residual sugar, sulphates, alcohol, and quality.

Residual sugar is positively correlated with sulphates and alcohol.

Chlorides is positively correlated with free sulfur dioxide and total sulfur dioxide.

Free sulfur dioxide is negatively correlated with total sulfur dioxide.

Density is positively correlated with sulphates and negatively correlated with pH.

pH is negatively correlated with sulphates and alcohol.

Sulphates is positively correlated with alcohol and quality.

Alcohol is positively correlated with quality.

Quality is positively correlated with sulphates and alcohol.