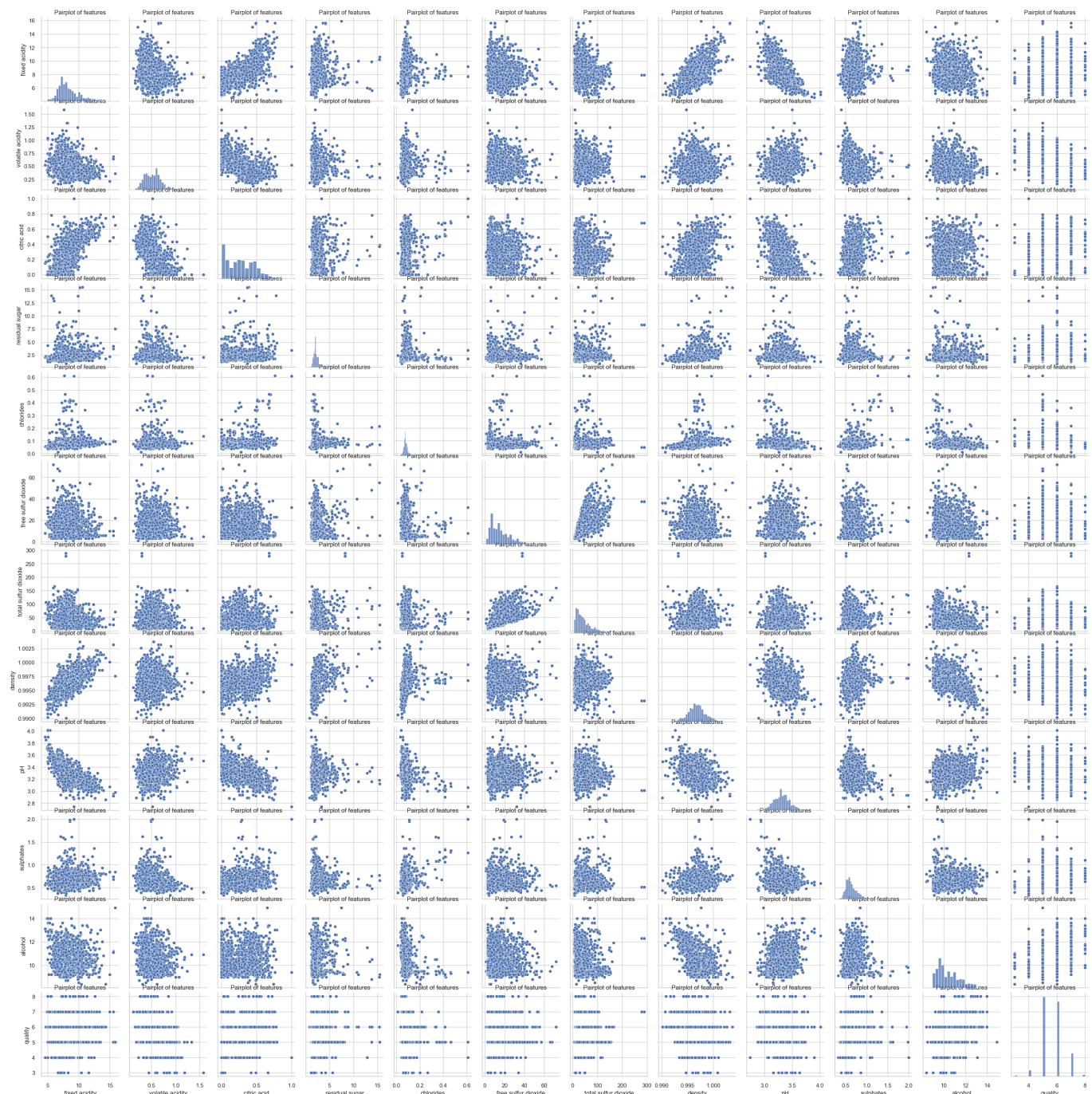# IntelliML Report

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 this dataset are chemical and sensory properties of red wine.

Fixed acidity is the amount of tartaric acid in the wine. Volatile acidity is the amount of acetic acid in the wine. Citric acid is a type of acid that occurs naturally in grapes. Residual sugar is the amount of sugar that remains in the wine after fermentation. Chlorides are salts that occur naturally in wine. Free sulfur dioxide is a type of preservative that is added to wine to prevent the growth of bacteria. Total sulfur dioxide is the amount of free sulfur dioxide plus the amount of bound sulfur dioxide in the wine. Density is the weight of a volume of wine compared to the weight of an equal volume of water. pH is a measure of the acidity or alkalinity of a solution. Sulphates are salts of sulfuric acid that occur naturally in wine. Alcohol is the percentage of alcohol by volume in the wine. Quality is a subjective measure of the overall taste of the wine.

Insights on dataset

The dataset contains 1599 rows and 11 columns. The features are: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates and alcohol. The quality of the wine is also included.

The mean of fixed acidity is 8.319637, the median is 7.9, the minimum is 4.6 and the maximum is 15.9. The mean of volatile acidity is 0.527821, the median is 0.52, the minimum is 0.12 and the maximum is 1.58. The mean of citric acid is 0.270976, the median is 0.26, the minimum is 0.0 and the maximum is 1.0. The mean of residual sugar is 2.538806, the median is 2.2, the minimum is 0.9 and the maximum is 15.5. The mean of chlorides is 0.087467, the median is 0.08, the minimum is 0.012 and the maximum is 0.611. The mean of free sulfur dioxide is 15.874922, the median is 14, the minimum is 1 and the maximum is 72. The mean of total sulfur dioxide is 46.467792, the median is 38, the minimum is 6 and the maximum is 289. The mean of density is 0.996747, the median is 0.99675, the minimum is 0.99007 and the maximum is 1.00369. The mean of pH is 3.311113, the median is 3.31, the minimum is 2.74 and the maximum is 4.01. The mean of sulphates is 0.658149, the median is 0.62, the minimum is 0.33 and the maximum is 2.0. The mean of alcohol is 10.422983, the median is 10.2, the minimum is 8.4 and the maximum is 14.9.

The distribution of the data is approximately normal.

Insights on Null Values in the dataset

The dataset contains no missing values for any of the features. 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. It is still possible that some of the data is incorrect or incomplete, even if it is not missing. Therefore, it is important to carefully examine the data before making any conclusions.

Feature Distribution

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

Fixed acidity: slightly left skewed. This indicates that the data is more concentrated towards the left side of the distribution, with a long tail on the right. This could be due to a number of factors, such as the fact that most wines have a relatively low level of fixed acidity.

Volatile acidity: slightly left skewed. This indicates that the data is more concentrated towards the left side of the distribution, with a long tail on the right. This could be due to a number of factors, such as the fact that most wines have a relatively low level of volatile acidity.

Citric acid: slightly left skewed. This indicates that the data is more concentrated towards the left side of the distribution, with a long tail on the right. This could be due to a number of factors, such as the fact that most wines have a relatively low level of citric acid.

Residual sugar: moderately right skewed. This indicates that the data is more concentrated towards the right side of the distribution, with a long tail on the left. This could be due to a number of factors, such as the fact that some wines have a very high level of residual sugar.

Chlorides: moderately right skewed. This indicates that the data is more concentrated towards the right side of the distribution, with a long tail on the left. This could be due to a number of factors, such as the fact that some wines have a very high level of chlorides.

Free sulfur dioxide: moderately right skewed. This indicates that the data is more concentrated towards the right side of the distribution, with a long tail on the left. This could be due to a number of factors, such as the fact that some wines have a very high level of free sulfur dioxide.

Total sulfur dioxide: moderately right skewed. This indicates that the data is more concentrated towards the right side of the distribution, with a long tail on the left. This could be due to a number of factors, such as the fact that some wines have a very high level of total sulfur dioxide.

Density: slightly right skewed. This indicates that the data is more concentrated towards the right side of the distribution, with a long tail on the left. This could be due to a number of factors, such as the fact that some wines have a very high density.

pH: slightly left skewed. This indicates that the data is more concentrated towards the left side of the distribution, with a long tail on the right. This could be due to a number of factors, such as the fact that most wines have a relatively low pH.

Sulphates: moderately right skewed. This indicates that the data is more concentrated towards the right side of the distribution, with a long tail on the left. This could be due to a number of factors, such as the fact that some wines have a very high level of sulphates.

Alcohol: slightly right skewed. This indicates that the data is more concentrated towards the right side of the distribution, with a long tail on the left. This could be due to a number of factors, such as the fact that some wines have a very high level of alcohol.

Quality: slightly left skewed. This indicates that the data is more concentrated towards the left side of the distribution, with a long tail on the right. This could be due to a number of factors, such as the fact that most wines are rated as being of good quality.

The skewness of the data could have a number of consequences. For example, it could make it more difficult to perform certain statistical analyses on the data, such as linear regression. It could also make it more difficult to interpret the results of such analyses.

fixed acidity

residual sugar