Wine Quality Dataset Report

1.Introduction: The dataset is related to the red and white variants of the wine. It contains various physicochemical properties of the wines and a quality rating.

2. Data Overview: The dataset contains the following attributes:

1. Fixed acidity
2. Volatile acidity
3. Citric acid
4. Residual sugar
5. Chlorides
6. Free sulfur dioxide
7. Total sulfur dioxide
8. Density
9. pH
10. Sulphates
11. Alcohol
12. Quality (score between 0 and 10)

a. In the above reference, two datasets were created, using red and white wine samples.
b. The inputs include objective tests (e.g. PH values) and the output is based on sensory data (median of at least 3 evaluations made by wine experts). Each expert graded the wine quality between 0 (very bad) and 10 (very excellent).
c. Also, we plot the relative importance of the input variables vs quality.

3. Data Exploration:

from the above dataset using the [. describe ()] from pandas library we get basic statistics for each attribute, like mean, standard deviation, minimum, and maximum values.

the finding we get here is:

- Most wines have an alcohol percentage around 10%.
- The average quality rating is around 5.6.
-

4. Data Visualization:

Using libraries like Matplotlib and Seaborn, you could visualize the data:

Histograms: To see the distribution of each attribute.

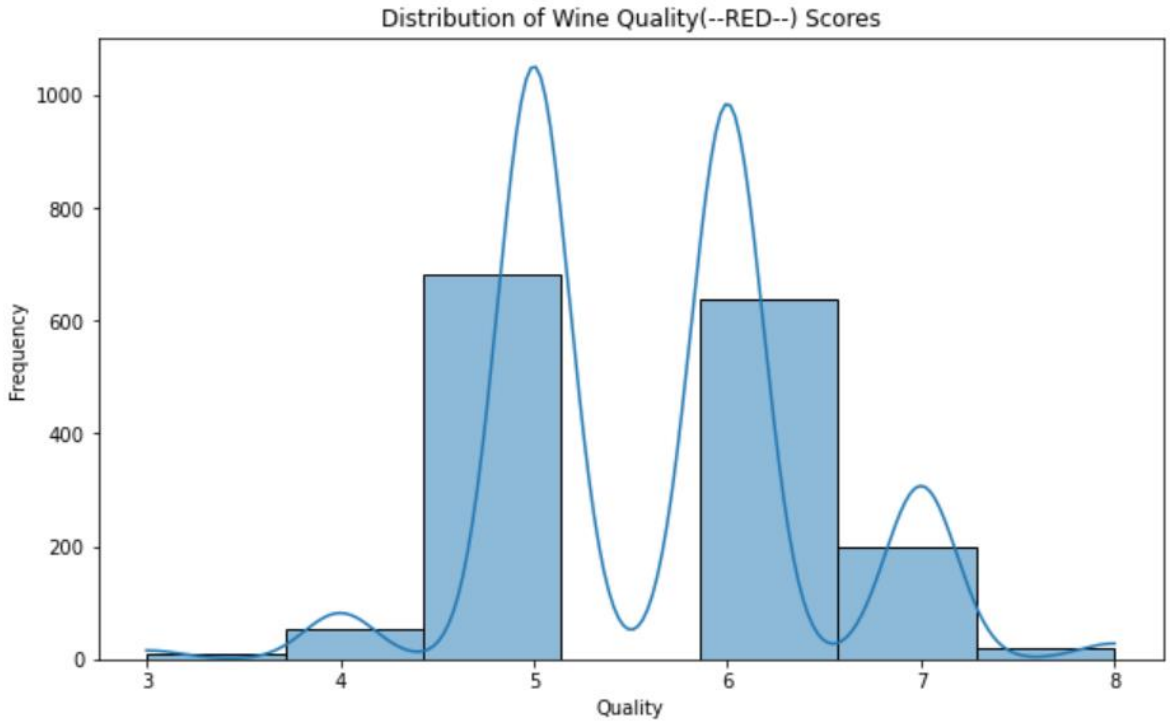
Correlation heatmap: To check the relationship between the attributes.

From visualizations, you might find:

- Most wines have a quality rating of 5 or 6.
- Attributes like alcohol, sulphates, and citric acid might have a positive correlation with wine quality.

The key findings of the dataset are as follows for red wine :

- There is a positive correlation between alcohol content and wine quality.
- There is a negative correlation between volatile acidity and wine quality.
- There is a negative correlation between residual sugar and wine quality.
- There is a negative correlation between pH and wine quality.
- There is a positive correlation between density and wine quality.
- There is a positive correlation between Sulphates and wine quality.




```
df.describe()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.467792	0.996747	3.311113	0.658149	10.422983
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324	0.001887	0.154386	0.169507	1.065668
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	0.990070	2.740000	0.330000	8.400000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000	0.995600	3.210000	0.550000	9.500000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000	0.996750	3.310000	0.620000	10.200000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000	0.997835	3.400000	0.730000	11.100000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000	1.003690	4.010000	2.000000	14.900000

The key findings of the dataset are as follows for white wine :

* There is a positive correlation between alcohol and wine quality.

* There is a negative correlation between volatile acidity and wine quality

* There is a positive correlation between residual sugar and wine quality.

* There is a positive correlation between pH and wine quality.

* There is a negative correlation between density and wine quality.

* There is a positive correlation between sulphates and wine quality.

* There is a positive correlation between fixed acidity and wine quality.

```
df1.describe()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	
count	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4
mean	6.854788	0.278241	0.334192	6.391415	0.045772	35.308085	138.360657	0.994027	3.188267	0.489847	10.514267	
std	0.843868	0.100795	0.121020	5.072058	0.021848	17.007137	42.498065	0.002991	0.151001	0.114126	1.230621	
min	3.800000	0.080000	0.000000	0.600000	0.009000	2.000000	9.000000	0.987110	2.720000	0.220000	8.000000	
25%	6.300000	0.210000	0.270000	1.700000	0.036000	23.000000	108.000000	0.991723	3.090000	0.410000	9.500000	
50%	6.800000	0.260000	0.320000	5.200000	0.043000	34.000000	134.000000	0.993740	3.180000	0.470000	10.400000	
75%	7.300000	0.320000	0.390000	9.900000	0.050000	46.000000	167.000000	0.996100	3.280000	0.550000	11.400000	
max	14.200000	1.100000	1.660000	65.800000	0.346000	289.000000	440.000000	1.038980	3.820000	1.080000	14.200000