Importing the libraries

In [ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Loading the dataset

In [6]:
```
import pandas as pd
data = pd.read_csv("https://raw.githubusercontent.com/dsrscientist/dataset1/master/winequality-red.csv")
data.head()
```

Out[6]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

In [ ]:
```

```

data.info()

This gives a concise summary of the data frame. We can see there are no null values in the dataframe,including 13 columns and 1599 entries.The dataset is already clean and tidy

"""Dropping the first columns containing the index values which is no use for us""

In [9]:
```
data = data.iloc[:,1:]
print(data.head())
```

```
   volatile acidity  citric acid  residual sugar  chlorides  \
0              0.70         0.00             1.9      0.076
1              0.88         0.00             2.6      0.098
2              0.76         0.04             2.3      0.092
3              0.28         0.56             1.9      0.075
4              0.70         0.00             1.9      0.076

   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
4                 11.0                  34.0   0.9978  3.51       0.56

   alcohol  quality
0      9.4        5
1      9.8        5
2      9.8        5
3      9.8        6
4      9.4        5
```

In [10]:
```
data.describe()
```

Out[10]:

	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000
mean	0.527821	0.270976	2.538806	0.087467	15.874922	46.467792	0.996747	3.311113	0.658149	10.422983	5.636023
std	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324	0.001887	0.154386	0.169507	1.065668	0.807569
min	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	0.990070	2.740000	0.330000	8.400000	3.000000
25%	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000	0.995600	3.210000	0.550000	9.500000	5.000000
50%	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000	0.996750	3.310000	0.620000	10.200000	6.000000
75%	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000	0.997835	3.400000	0.730000	11.100000	6.000000
max	1.580000	1.000000	15.500000	0.611000	72.000000	289.000000	1.003690	4.010000	2.000000	14.900000	8.000000

This is used to view some basic statistical details like percentile,mean,std etc of a dataframe or a series of numeric values

In [11]:
```
data.corr()['quality']
```

Out[11]:
```
volatile acidity       -0.390558
citric acid             0.226373
residual sugar          0.013732
chlorides              -0.128907
free sulfur dioxide    -0.050656
total sulfur dioxide   -0.185100
density                -0.174919
pH                     -0.057731
sulphates               0.251397
alcohol                 0.476166
quality                 1.000000
Name: quality, dtype: float64
```

we are trying figure out the correlation of every other feature w.r.t the quality of wine

In [14]:
```
import matplotlib.pyplot as plt
plt.figure(figsize=(10,10))
sns.heatmap(data.corr(),annot=True,linewidth=0.5,center=0,cmap='coolwarm')
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_1700/2892027293.py in <module>
      1 import matplotlib.pyplot as plt
      2 plt.figure(figsize=(10,10))
----> 3 sns.heatmap(data.corr(),annot=True,linewidth=0.5,center=0,cmap='coolwarm')
      4 plt.show()

NameError: name 'sns' is not defined
```


In [19]:
```
plt.hist(data.quality,bins=6,alpha=0.5,histtype='bar',ec='black')
plt.xlabel('Quality')
plt.ylabel('Count')
plt.show()
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_1700/3643494611.py in <module>
      1 plt.hist(data.quality,bins=6,alpha=0.5,histtype='bar',ec='black')
----> 2 plt.xlabel('Quality')
      3 plt.ylabel('Count')
      4 plt.show()

AttributeError: module 'matplotlib.pyplot' has no attribute 'xlable'
```


The above plot shows the distribution of the quality of the wine in the dataset and the represents that most of the wine is of average quality i.e. quality ranging from 5-7.

In [22]:
```
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(8,5))
sns.barplot(data['quality'],data['pH'],palette='Gnbu_d')
plt.show()
```

```
D:\ANACONDA\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```


From the barplot of quality vs pH we can visualise that there is a slight decrese in pH with the increase in quality of the wine

In [23]:
```
sns.boxplot(x='quality',y='pH',data=data,palette='Gnbu_d')
plt.title("Boxplot of Quality and pH")
plt.show()
```


In [ ]:
```

```

In [26]:
```
sns.boxplot(x='quality',y='residual sugar',data=data,palette='Gnbu_d')
plt.title("Boxplot of Quality and residual sugar")
plt.show()
```


In [ ]:
```

```

In [28]:
```
sns.boxplot(x='quality',y='density',data=data,palette='Gnbu_d')
plt.title("Boxplot of Quality and density")
plt.show()
```


In [ ]:
```

```

In [29]:
```
sns.boxplot(x='quality',y='sulphates',data=data,palette='Gnbu_d')
plt.title("Boxplot of Quality and sulphates")
plt.show()
```


In [ ]:
```

```

In [30]:
```
sns.boxplot(x='quality',y='chlorides',data=data,palette='Gnbu_d')
plt.title("Boxplot of Quality and chlorides")
plt.show()
```


From the above plot we can see that there is not even a slight change in the quality of wine for a particular amount of chloride mixed in them

Effect of acidity on the quality of wine 1 Citric Acid 2 Volatile Acidity 3 Fixed Acidity

In [33]:
```
sns.boxplot(x='quality',y='citric acid',data=data,palette="coolwarm")
plt.title("Boxplot of Quality and citric acid")
plt.show()
```


The quality of the wine increase with increase in the amount of citric acid in the wine

In [32]:
```
sns.boxplot(x='quality',y='volatile acidity',data=data,palette="coolwarm")
plt.title("Boxplot of Quality and volatile acidity")
plt.show()
```


From the above boxplot we can see that the quality of wine increase with the decrese in the amount of volatile acids

THANKYOU

In [ ]:
```

```