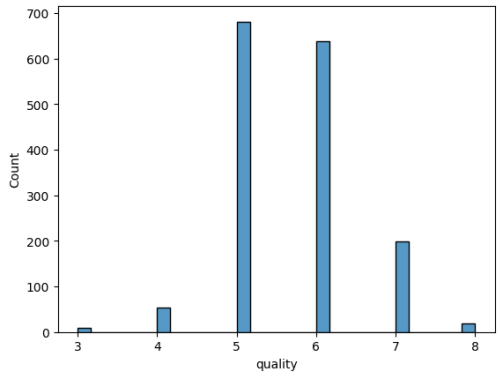
```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sns
import numpy as np
from sklearn.preprocessing import LabelEncoder
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
```

```
df=pd.read_csv('winequality-red.csv')
df.head()
```

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

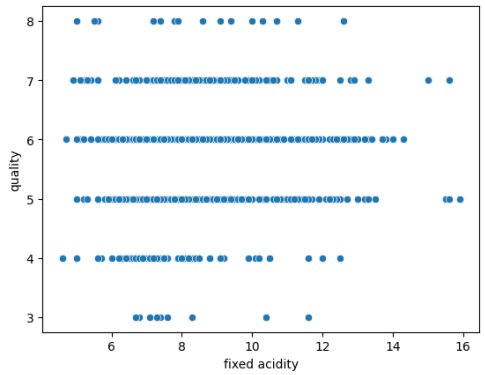
```
sns.histplot(df['quality'])
```

```
<Axes: xlabel='quality', ylabel='Count'>
```



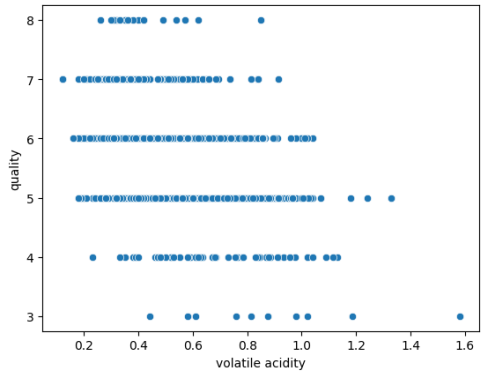

```
sns.scatterplot(x='fixed acidity',y='quality',data=df)
```

```
<Axes: xlabel='fixed acidity', ylabel='quality'>
```



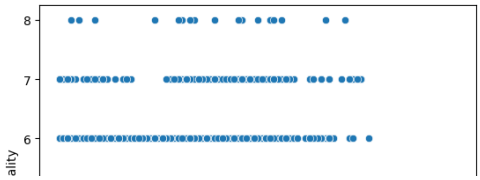

```
sns.scatterplot(x='volatile acidity',y='quality',data=df)
```

```
<Axes: xlabel='volatile acidity', ylabel='quality'>
```



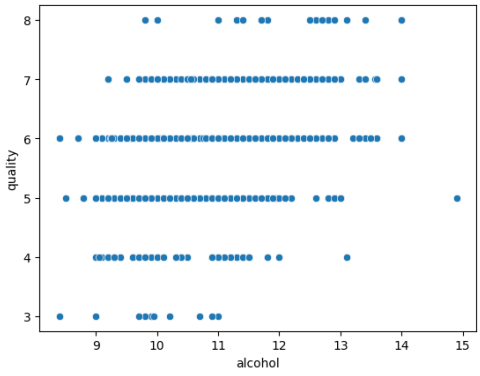

```
sns.scatterplot(x='citric acid',y='quality',data=df)
```

```
<Axes: xlabel='citric acid', ylabel='quality'>
```



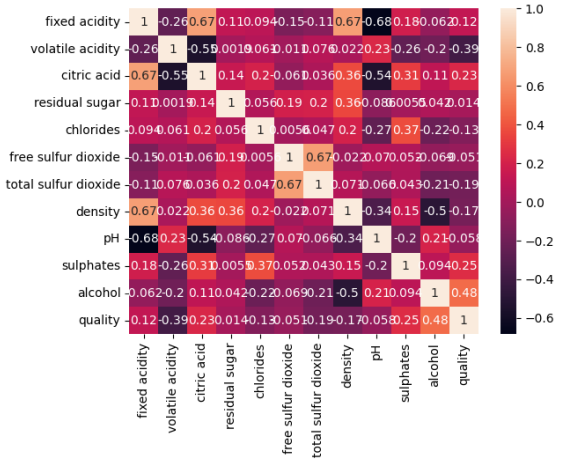

```
sns.scatterplot(x='alcohol',y='quality',data=df)
```

```
<Axes: xlabel='alcohol', ylabel='quality'>
```




```
sns.heatmap(df.corr(),annot=True)
```

```
<Axes: >
```




```
df.describe()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	dens
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.467792	0.996
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.895324	0.001
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.000000	0.990
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.000000	0.995
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.000000	0.996
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.000000	0.997

```
df.isnull().any()
```

```
fixed acidity           False
volatile acidity        False
citric acid             False
residual sugar          False
chlorides               False
free sulfur dioxide     False
total sulfur dioxide    False
density                 False
pH                      False
sulphates               False
alcohol                 False
quality                 False
dtype: bool
```

```
df.value_counts(df['quality'])
```

```
quality
5    681
6    638
7    199
4     53
8     18
3     10
dtype: int64
```

```
x=df.iloc[:,0:11]
y=df.iloc[:,11]

smt=SMOTE()
x_resampled,y_resampled=smt.fit_resample(x,y)

x_train,x_test,y_train,y_test=train_test_split(x_resampled,y_resampled,test_size=0.2,random_state=42)
```

```
model=DecisionTreeClassifier()
model.fit(x_train,y_train)
```

```
▾ DecisionTreeClassifier
DecisionTreeClassifier()
```

```
y_pred=model.predict(x_test)

acc=accuracy_score(y_pred,y_test)
cm=confusion_matrix(y_pred,y_test)

print(acc)
print(cm)
```

```
0.7836185819070904
[[130   0   6   5   1   0]
 [  1 107  11  12   0   0]
 [  0  19 105  37   2   0]
 [  1   5  25  73   9   5]
 [  0   0   3  25 101   4]
 [  0   0   0   3   3 125]]
```