```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from imblearn.over_sampling import SMOTE

import pickle

# import warnings
# %matplotlib inline
# warnings.filterwarnings('ignore')
```

Loading the dataset

```
df = pd.read_csv('G:\Project\Wine Quality Predictor\winequality.csv')
df.head()
```

	type	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	den
0	white	7.0	0.27	0.36	20.7	0.045	45.0	170.0	1
1	white	6.3	0.30	0.34	1.6	0.049	14.0	132.0	0
2	white	8.1	0.28	0.40	6.9	0.050	30.0	97.0	0
3	white	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0

Statistical info

```
df.describe()
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	
count	6487.000000	6489.000000	6494.000000	6495.000000	6495.000000	6497.000000	6
mean	7.216579	0.339691	0.318722	5.444326	0.056042	30.525319	
std	1.296750	0.164649	0.145265	4.758125	0.035036	17.749400	

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6497 entries, 0 to 6496
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   type                  6497 non-null   object 
 1   fixed acidity         6487 non-null   float64
 2   volatile acidity      6489 non-null   float64
 3   citric acid           6494 non-null   float64
 4   residual sugar        6495 non-null   float64
 5   chlorides             6495 non-null   float64
 6   free sulfur dioxide   6497 non-null   float64
 7   total sulfur dioxide  6497 non-null   float64
 8   density               6497 non-null   float64
 9   pH                    6488 non-null   float64
 10  sulphates             6493 non-null   float64
 11  alcohol               6497 non-null   float64
 12  quality               6497 non-null   int64  
dtypes: float64(11), int64(1), object(1)
memory usage: 660.0+ KB
```

Filling Missing Values

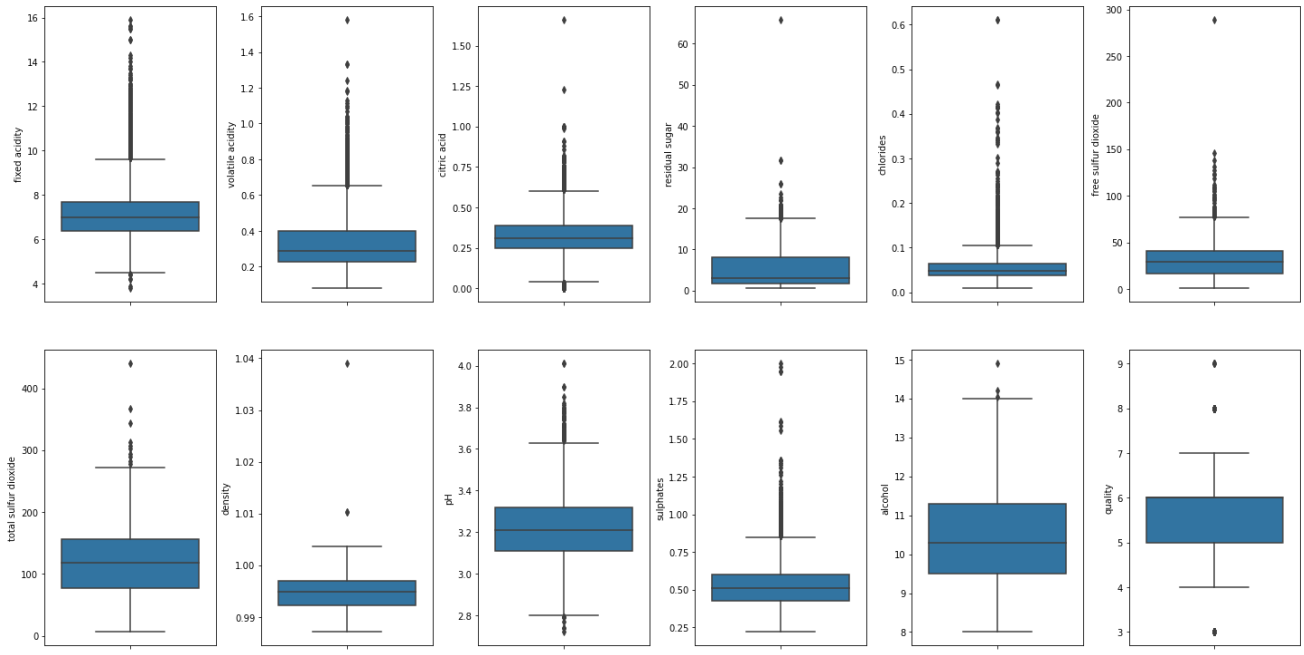
```
#Checking Missing values
df.isnull().sum()
```

```
type                    0
fixed acidity          10
volatile acidity        8
citric acid             3
residual sugar          2
chlorides               2
free sulfur dioxide     0
total sulfur dioxide    0
density                 0
pH                      9
sulphates               4
alcohol                 0
quality                 0
dtype: int64
```

```
volatile acidity        0
citric acid             0
residual sugar          0
chlorides               0
free sulfur dioxide     0
total sulfur dioxide    0
density                 0
pH                      0
sulphates               0
alcohol                 0
quality                 0
dtype: int64
```

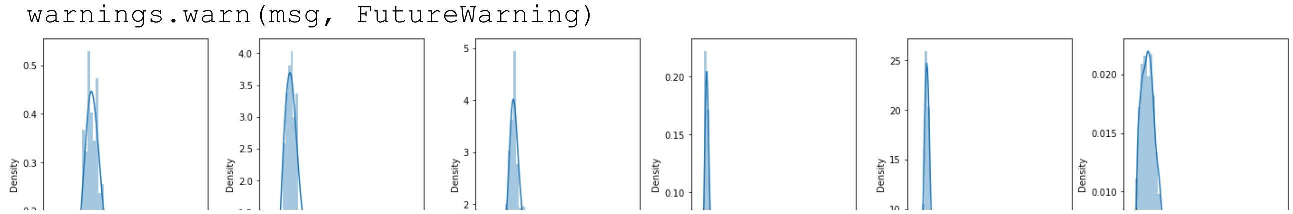
Exploratory Data Analysis

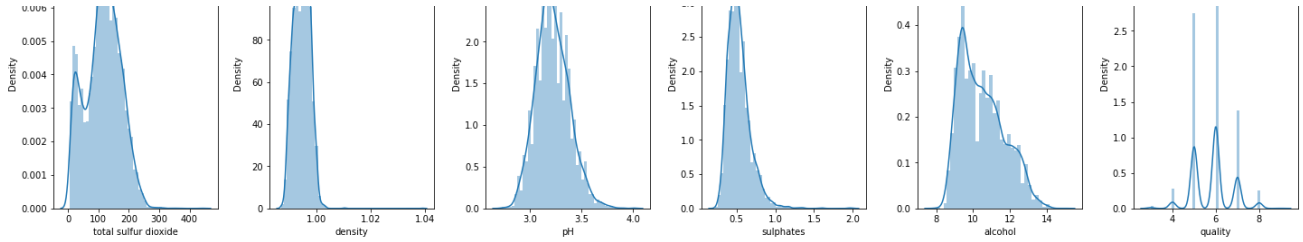
```
# create box plots
fig, ax = plt.subplots(ncols=6, nrows=2, figsize=(20,10))
index = 0
ax = ax.flatten()

for col, value in df.items():
    if col != 'type':
        sns.boxplot(y=col, data=df, ax=ax[index])
        index += 1
plt.tight_layout(pad=0.5, w_pad=0.7, h_pad=5.0)
```



```
# create dist plot
fig, ax = plt.subplots(ncols=6, nrows=2, figsize=(20,10))
index = 0
ax = ax.flatten()

for col, value in df.items():
    if col != 'type':
        sns.distplot(value, ax=ax[index])
        index += 1
plt.tight_layout(pad=0.5, w_pad=0.7, h_pad=5.0)
```

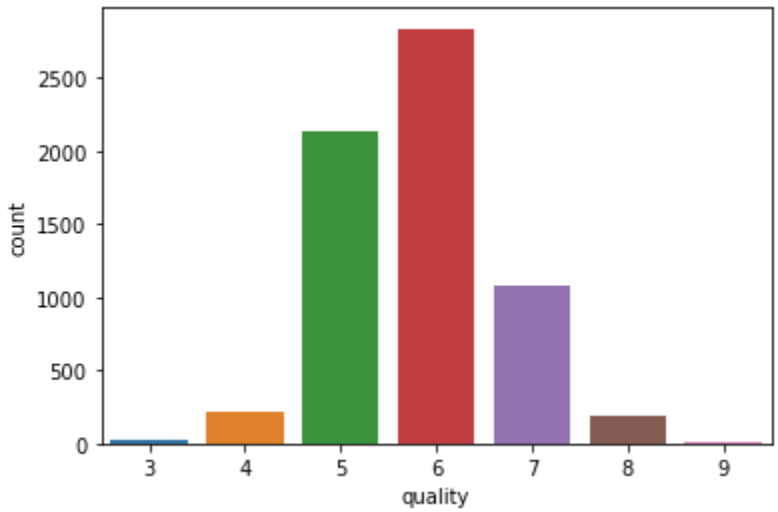
```
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
  warnings.warn(msg, FutureWarning)
```

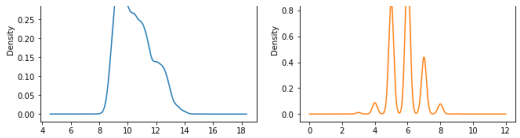





```
sns.countplot(df['quality'])
```

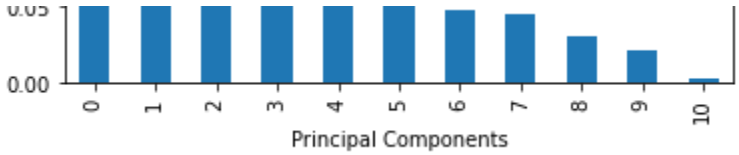
```
c:\Users\sudha\anaconda3\lib\site-packages\seaborn\_decorators.py:36: Futur
  warnings.warn(
<AxesSubplot:xlabel='quality', ylabel='count'>
```

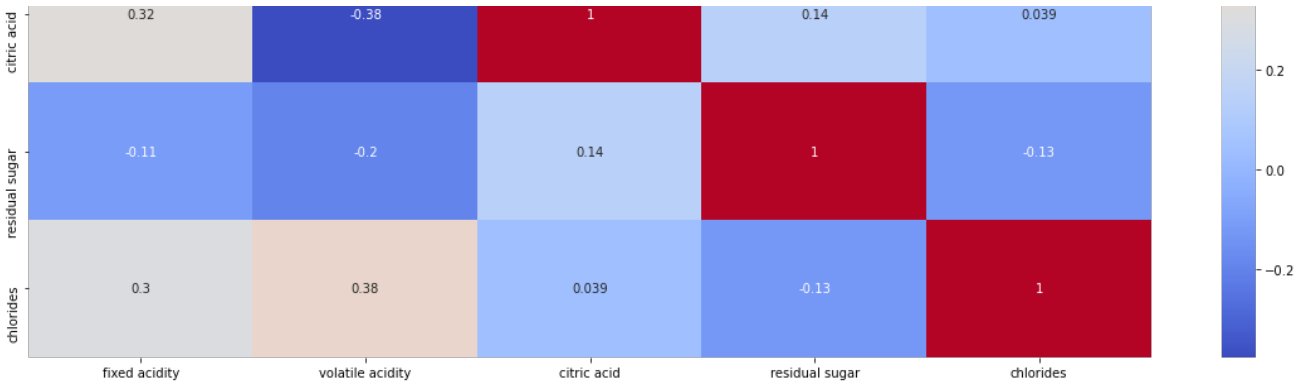



Density
0.25
0.20
0.15
0.10
0.05
0.00
4
6
8
10
12
14
16
18
Density
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
12

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	densit
0	7.0	0.270	0.36	20.7	0.045	45.0	170.0	1.0010
1	6.3	0.300	0.34	1.6	0.049	14.0	132.0	0.9940
2	8.1	0.280	0.40	6.9	0.050	30.0	97.0	0.9951
3	7.2	0.230	0.32	8.5	0.058	47.0	186.0	0.9956
4	7.2	0.230	0.32	8.5	0.058	47.0	186.0	0.9956
...	...	...	...	...	...	...	...	
6492	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.9949
6493	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.9951
6494	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.9957
6495	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.9954
6496	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.9954

count	6.497000e+03	6.497000e+03	6.497000e+03	6.497000e+03	6.497000e+03	6.497
mean	-1.381738e-15	3.086397e-15	2.119284e-16	-4.138345e-16	2.648763e-16	-5.71
std	1.000077e+00	1.000077e+00	1.000077e+00	1.000077e+00	1.000077e+00	1.000
min	-2.636958e+00	-1.578333e+00	-2.194751e+00	-1.018352e+00	-1.342973e+00	-1.663
25%	-6.302460e-01	-6.666730e-01	-4.732259e-01	-7.660934e-01	-5.150623e-01	-7.62(
50%	-1.671586e-01	-3.020089e-01	-6.005986e-02	-5.138350e-01	-2.581246e-01	-8.59
75%	3.731101e-01	3.665421e-01	4.908282e-01	5.582634e-01	2.557510e-01	5.90


0.05
0.00
0
1
2
3
4
5
6
7
8
9
10
Principal Components


citric acid
residual sugar
chlorides
0.32
-0.38
1
0.14
0.039
-0.11
-0.2
0.14
1
-0.13
0.3
0.38
0.039
-0.13
1
fixed acidity
volatile acidity
citric acid
residual sugar
chlorides
0.2
0.0
−0.2