K-Nearest Neighbor Classification

Import packages and data set

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
%matplotlib inline

df = pd.read_table("Classified_Data.txt",sep=',', index_col=0)
df.head()
```

	WTT	PTI	EQW	SBI	LQE	QWG	FDJ	PJF	HQE
0	0.913917	1.162073	0.567946	0.755464	0.780862	0.352608	0.759697	0.643798	0.879422
1	0.635632	1.003722	0.535342	0.825645	0.924109	0.648450	0.675334	1.013546	0.621552
2	0.721360	1.201493	0.921990	0.855595	1.526629	0.720781	1.626351	1.154483	0.957877
3	1.234204	1.386726	0.653046	0.825624	1.142504	0.875128	1.409708	1.380003	1.522692

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1000 entries, 0 to 999
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   WTT           1000 non-null   float64
 1   PTI           1000 non-null   float64
 2   EQW           1000 non-null   float64
 3   SBI           1000 non-null   float64
 4   LQE           1000 non-null   float64
 5   QWG           1000 non-null   float64
 6   FDJ           1000 non-null   float64
 7   PJF           1000 non-null   float64
 8   HQE           1000 non-null   float64
 9   NXJ           1000 non-null   float64
 10  TARGET CLASS  1000 non-null   int64
dtypes: float64(10), int64(1)
memory usage: 93.8 KB
```

```
df.describe()
```

	WTT	PTI	EQW	SBI	LQE	QWG	FDJ	PJF	HQE	NXJ	
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000
mean	0.949682	1.114303	0.834127	0.682099	1.032336	0.943534	0.963422	1.071960	1.158251	1.362725	
std	0.289635	0.257085	0.291554	0.229645	0.243413	0.256121	0.255118	0.288982	0.293738	0.204225	
min	0.174412	0.441398	0.170924	0.045027	0.315307	0.262389	0.295228	0.299476	0.365157	0.639693	
25%	0.742358	0.942071	0.615451	0.515010	0.870855	0.761064	0.784407	0.866306	0.934340	1.222623	
50%	0.940475	1.118486	0.813264	0.676835	1.035824	0.941502	0.945333	1.065500	1.165556	1.375368	
75%	1.163295	1.307904	1.028340	0.834317	1.198270	1.123060	1.134852	1.283156	1.383173	1.504832	

Check the spread of the features

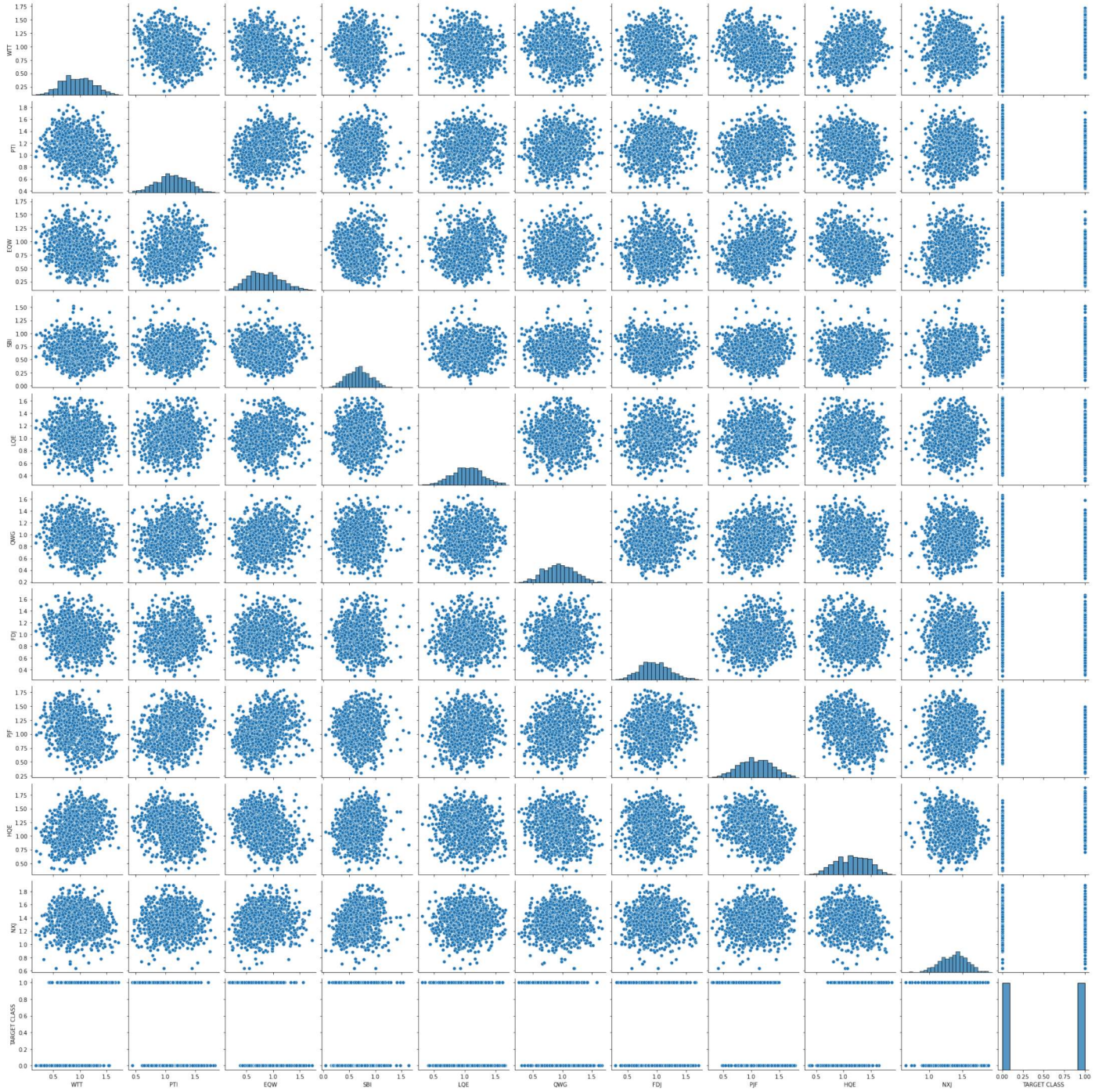
```
l=list(df.columns)
l[0:len(l)-2]
```

```
['WTT', 'PTI', 'EQW', 'SBI', 'LQE', 'QWG', 'FDJ', 'PJF', 'HQE']
```

Run a 'for' loop to draw boxlots of all the features for '0' and '1' TARGET CLASS

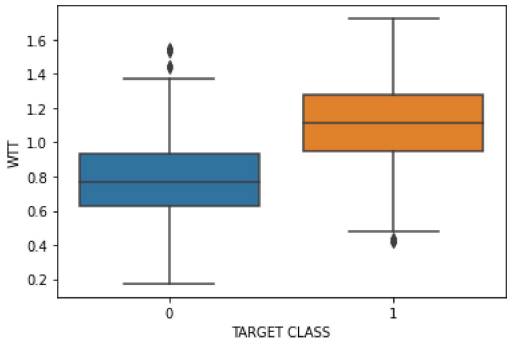
```
sns.pairplot(df)
```

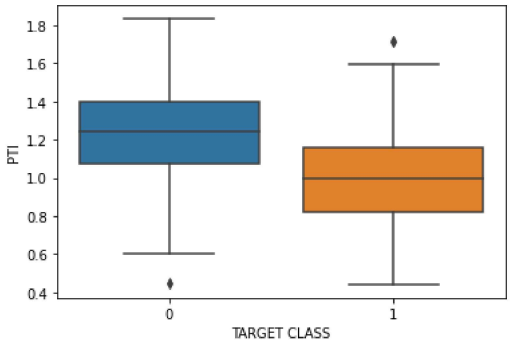
```
<seaborn.axisgrid.PairGrid at 0x121129070>
```

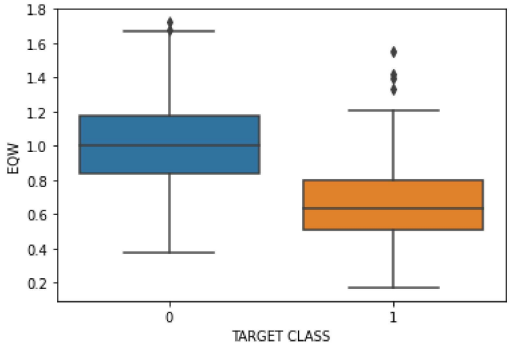


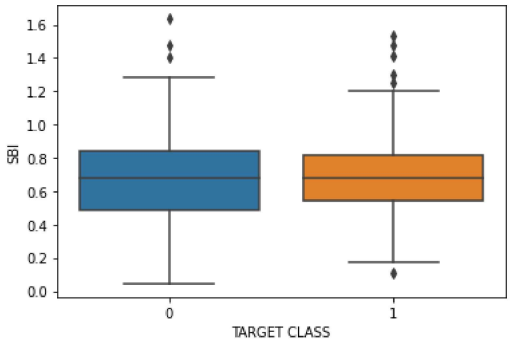
Identify the Target Classes from the Dataset and their Counts

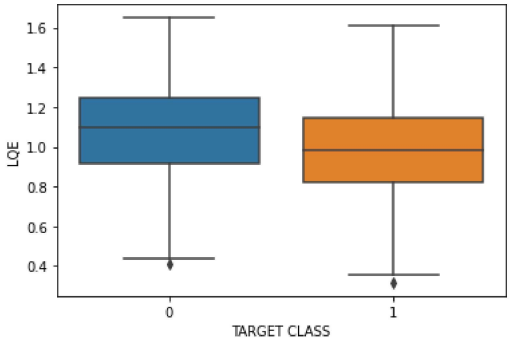
```
l=list(df.columns)
l[0:len(l)-2]

for i in range(len(l)-1):
    sns.boxplot(x='TARGET CLASS',y=l[i], data=df)
    plt.figure()
```

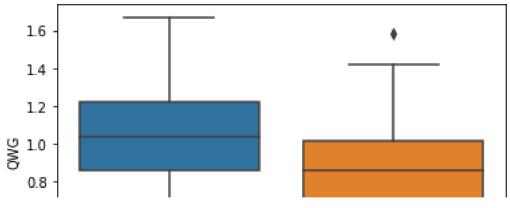

WTT
TARGET CLASS


PTI
TARGET CLASS


EQW
TARGET CLASS


SBI
TARGET CLASS


LQE
TARGET CLASS


QWG