

1 [442...

```
# Box plot for comparing the ph with other column and finding the outliers

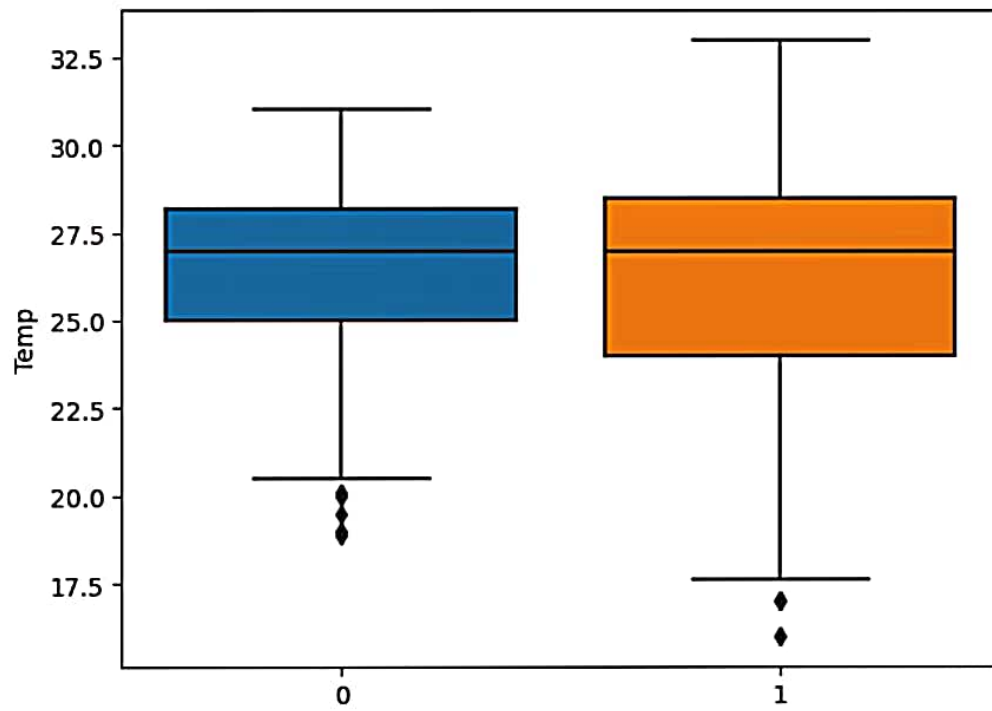
col_pruning=['Temp','D.O. (mg/l)','CONDUCTIVITY (µmhos/cm)','B.O.D. (mg/l)','NITRATENAN N+ NITRITENANN

for col in col_pruning:
    print("\n\n")
    coldesc=df[col].describe()
    col_IQR=coldesc[6]-coldesc[4]
    col_Lower=coldesc[4]-(1.5*col_IQR)

    col_Higher=coldesc[6]+(1.5*col_IQR)

#     print(col_Lower,col_Higher)
#     df.drop(df.index[(df[col]>col_Higher)],inplace=True,axis=0)
df.drop(df.index[(df[col]>col_Higher)],inplace=True,axis=0)
sns.boxplot(x='Water Qu',y=df[col],data=df)
plt.show()

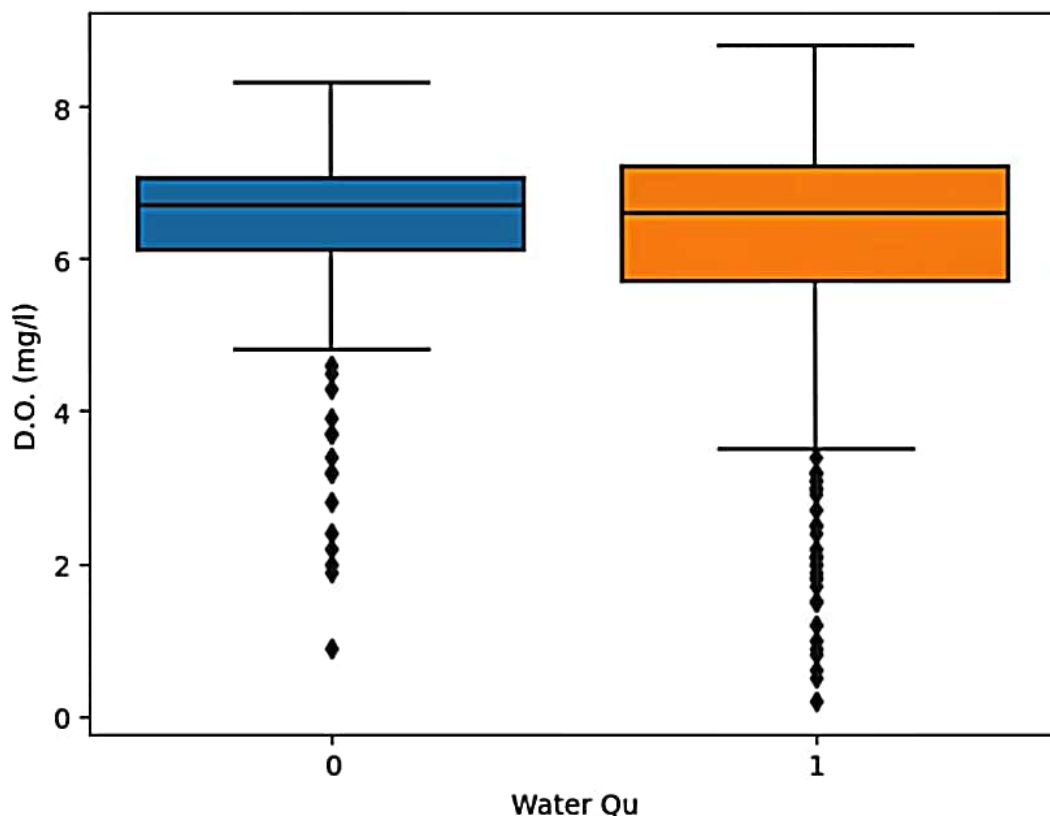
print(df[col].describe())
```



Activate Windows
Go to Settings to activate Windows.

Water Qu

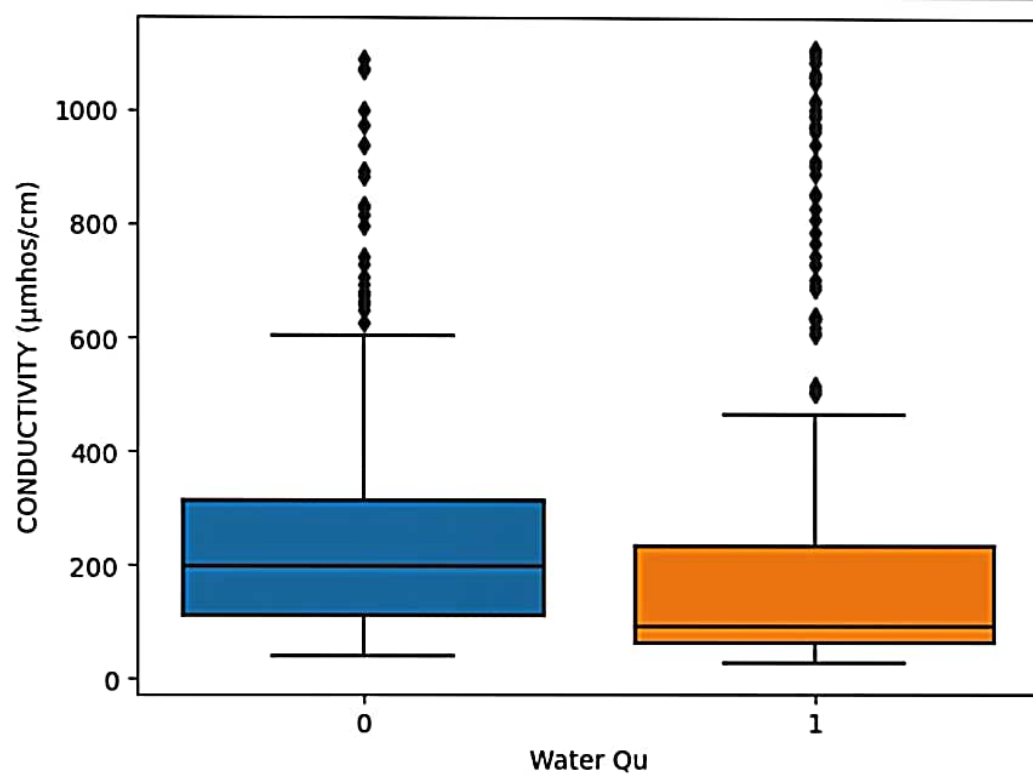
```
count    879.000000
mean      26.093743
std        3.261618
min       16.000000
25%       24.450000
50%       27.000000
75%       28.400000
max       33.000000
Name: Temp, dtype: float64
```



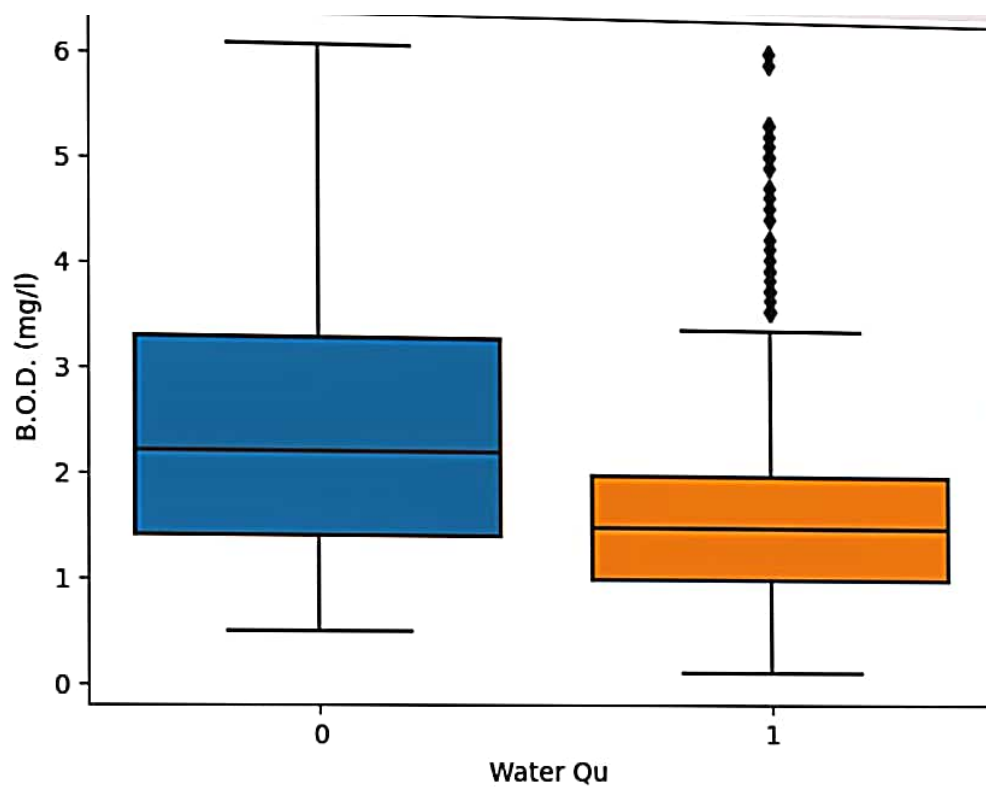
Activate Windows
Go to Settings to activate Windows.

water_qd

```
count    878.000000
mean      6.306640
std       1.295557
min       0.200000
25%       5.900000
50%       6.700000
75%       7.100000
max       8.800000
Name: D.O. (mg/l), dtype: float64
```

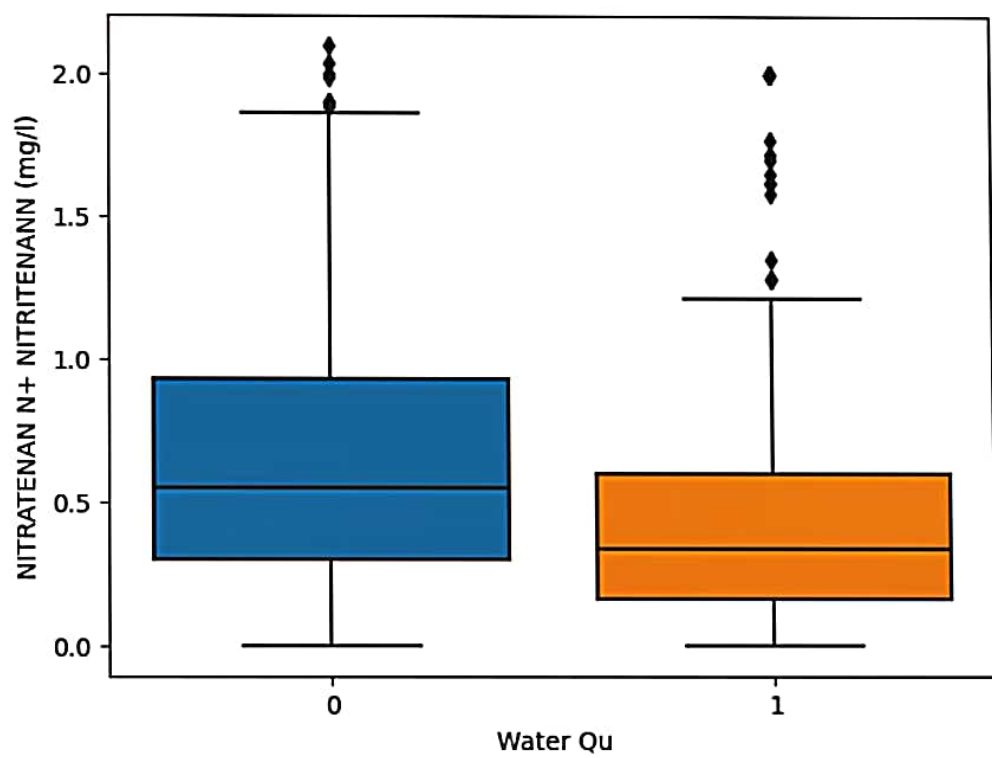


```
count      745.000000
mean       222.344966
std        243.275990
min         27.000000
25%        69.000000
50%       120.000000
75%       274.000000
max       1110.000000
Name: CONDUCTIVITY (μmhos/cm), dtype: float64
```



Activate Windows
Go to Settings to activate Windows.


```
count    675.000000
mean      1.939630
std       1.140444
min       0.100000
25%       1.100000
50%       1.600000
75%       2.500000
max       6.100000
Name: B.O.D. (mg/l), dtype: float64
```



Activate Windows
Go to Settings to activate Windows.

0

1

Water Qu

count	571.000000
mean	0.523135
std	0.451816
min	0.000000
25%	0.200000
50%	0.400000
75%	0.720000
max	2.100000

Name: NITRATENAN N+ NITRITENANN (mg/l), dtype: float64