

Project Development Phase

Project Development – Delivery of Sprint-2

Team ID	PNT2022TMID17967
Project Name	Project – Efficient Water Quality Analysis and Prediction using Machine Learning

Water Quality Index Calculation

```
In [18]: data['npH']=data.ph.apply(lambda x: (100 if (8.5>=x>=7)
                                         else(80 if (8.6>=x>=8.5) or (6.9>=x>=6.8)
                                         else(60 if (8.8>=x>=8.6) or (6.8>=x>=6.7)
                                         else(40 if (9>=x>=8.8) or (6.7>=x>=6.5)
                                         else 0))))))
```

```
In [19]: data['noo']=data.do.apply(lambda x: (100 if (x>=6)
                                             else(80 if (6>=x>=5.1)
                                             else(60 if (5>=x>=4.1)
                                             else(40 if (4>=x>=3)
                                             else 0))))
```

```
In [21]: data['nco']=data.tc.apply(lambda x: (100 if (5>=x>=0)
                                             else(80 if (50>=x>=5)
                                             else(60 if (500>=x>=50)
                                             else(40 if (10000>=x>=500)
                                             else 0))))))
```

```
In [22]: data['nbdo']=data.do.apply(lambda x: (100 if (3>=x>=0)
                                             else(80 if (6>=x>=3)
                                             else(60 if (80>=x>=6)
                                             else(40 if (125>=x>=80)
                                             else 0))))
```

```
In [23]: data['nec']=data.co.apply(lambda x: (100 if (75>=x>=0)
                                             else(80 if (150>=x>=75)
                                             else(60 if (225>=x>=150)
                                             else(40 if (300>=x>=225)
                                             else 0))))))
```

Water Quality Index Calculation 2

```
In [23]: data['nec']=data.co.apply(lambda x: (100 if (75>=x>=0)
                                             else(80 if (150>=x>=75)
                                             else(60 if (225>=x>=150)
                                             else(40 if (300>=x>=225)
                                             else 0))))))
```

```
In [24]: data['nna']=data.na.apply(lambda x: (100 if (20>=x>=0)
                                         else(80 if (50>=x>=20)
                                         else(60 if (100>=x>=50)
                                         else(40 if (200>=x>=100)
                                         else 0))))
```

Water Quality Index Calculation 3

```
In [25]: data['wph']=data.nph * 0.165
data['wdo']=data.ndo * 0.281
data['wbdo']=data.nbdo * 0.234
data['wec']=data.nec * 0.009
data['wna']=data.nna * 0.028
data['wco']=data.nco * 0.281
data['wqi']=data.wph+data.wdo+data.wbdo+data.wec+data.wna+data.wco
data
```

```
Out[25]:
```

	station	location	state	Temp	do	ph	co	bod	na	tc	...	nbdo	nec	nna	wph	wdo	wbdo	wec	wna	wco
0	1393	DAMANGANGA AT D/S OF MADHUBAN, DAMAN	DAMAN & DIU	30.600000	6.7	7.5	203.0	6.940049	0.100000	27.0	...	60	60	100	16.5	28.10	14.04	0.54	2.8	22.48
1	1399	ZUARI AT D/S OF PT. WHERE KUMBARJRIA CANAL JOI...	GOA	29.800000	5.7	7.2	189.0	2.000000	0.200000	8391.0	...	80	80	100	16.5	22.48	18.72	0.54	2.8	11.24
2	1475	ZUARI AT PANCHAWADI	GOA	29.500000	6.3	6.9	179.0	1.700000	0.100000	5330.0	...	60	60	100	13.2	28.10	14.04	0.54	2.8	11.24
3	3181	RIVER ZUARI AT BORIM BRIDGE	GOA	29.700000	5.8	6.9	64.0	3.800000	0.500000	8443.0	...	80	100	100	13.2	22.48	18.72	0.90	2.8	11.24
4	3182	RIVER ZUARI AT MARCAIM JETTY	GOA	29.500000	5.8	7.3	83.0	1.900000	0.400000	5500.0	...	80	80	100	16.5	22.48	18.72	0.72	2.8	11.24
...
1986	1330	TAMBIRAPARANI AT ARUMUGANERI, TAMILNADU	NAN	26.209814	7.9	738.0	7.2	2.700000	0.518000	202.0	...	60	100	100	0.0	28.10	14.04	0.90	2.8	16.88
1987	1450	PALAR AT VANIYAMBADI WATER SUPPLY HEAD WORK, T...	NAN	29.000000	7.5	585.0	6.3	2.600000	0.155000	315.0	...	60	100	100	0.0	28.10	14.04	0.90	2.8	16.88
1988	1403	GUMTI AT U/S SOUTH TRIPURA,TRIPURA	NAN	28.000000	7.6	98.0	6.2	1.200000	1.623079	570.0	...	60	100	100	0.0	28.10	14.04	0.90	2.8	11.24

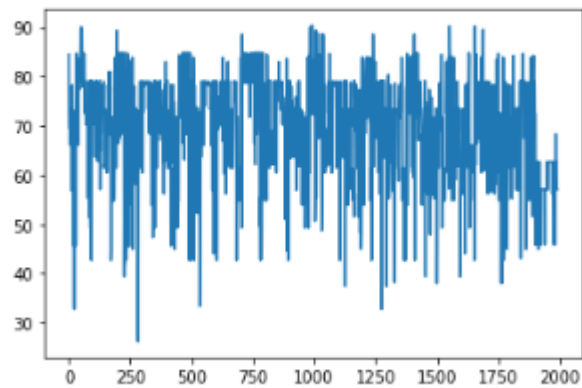
```
In [26]: average=data.groupby('year')['wqi'].mean()
average.head()
```

```
Out[26]: year
2003    58.900455
2004    54.270000
2005    69.043361
2006    68.363429
2007    68.773000
Name: wqi, dtype: float64
```

Data Visualization:

```
In [32]: plt.plot(data['wqi'])
```

```
Out[32]: [<matplotlib.lines.Line2D at 0x212ac4a3640>]
```



Splitting dependent and independent columns:

```
In [35]: x=data.iloc[:,0:7].values  
         y=data.iloc[:,7:].values
```

```
In [36]: x.shape
```

```
Out[36]: (1991, 7)
```

```
In [37]: y.shape
```

```
Out[37]: (1991, 1)
```

Splitting data into train and test:

```
In [38]: from sklearn.model_selection import train_test_split  
         X_train,X_test,Y_train,Y_test=train_test_split(x,y,test_size=0.3,random_state=10)  
         Y_trainId=np.ravel(Y_train)
```