

SPRINT 1

CODE AND TEST CASES

Date	18 October 2022
Team ID	PNT2022TMID52665
Project Name	Project – Efficient Water Quality Analysis and Prediction using Machine Learning
Maximum Marks	8 Marks

Data Collection & Data pre-processing:

Indian water quality data

Data Code (8) Discussion (0)

79 New Notebook Download (46 kB)

Business Earth and Nature Energy India History Pollution

water_dataX.csv (176.68 kB)

10 of 12 columns

STATION ...	LOCATIONS	STATE	Temp	D.O. (mg/l)	PH	CONDUCT..
1393	DAMANGANGA AT D/S OF MADHUBAN, DAMAN	DAMAN & DIU	30.6	6.7	7.5	203
1399	ZUARI AT D/S OF PT. WHERE KUMBARJRIA CANAL JOINS, GOA	GOA	29.8	5.7	7.2	189
1475	ZUARI AT PANCHAWADI	GOA	29.5	6.3	6.9	179
3181	RIVER ZUARI AT BORIM BRIDGE	GOA	29.7	5.8	6.9	64
3182	RIVER ZUARI AT MARCAIM JETTY	GOA	29.5	5.8	7.3	83
1400	MANDOVI AT	GOA	30	5.5	7.4	81

Data Explorer

176.68 kB

water_dataX.csv

```
File Edit Selection View Go Run Terminal Help
water_dataX.csv - Visual Studio Code

D:\> study > item > water_dataX.csv

1 STATION CODE,LOCATIONS,STATE,Temp,D.O. (mg/l),PH,CONDUCTIVITY (µmhos/cm),B.O.D. (mg/l),NITRATENAN N+ NITRITENANN (mg/l),FECAL COLIFORM (MPN/100ml),
2 1393,"DAMANGANGA AT D/S OF MADHUBAN, DAMAN",DAMAN & DIU,30.6,6.7,7.5,203,NAN,0.1,11,27,2014
3 1399,"ZUARI AT D/S OF PT. WHERE KUMBARJRIA CANAL JOINS, GOA",GOA,29.8,5.7,7.2,189,2.0,2,4953,8391,2014
4 1475,ZUARI AT PANCHAWADI,GOA,29.5,6.3,6.9,179,1.7,0.1,3243,5330,2014
5 3181,RIVER ZUARI AT BORIM BRIDGE,GOA,29.7,5.8,6.9,64,3.8,0.5,5382,8443,2014
6 3182,RIVER ZUARI AT MARCAIM JETTY,GOA,29.5,5.8,7.3,83,1.9,0.4,3428,5500,2014
7 1400,"MANDOVI AT NEIGHBOURHOOD OF PANAJI, GOA",GOA,30.5,5.7,7.4,81,1.5,0.1,2853,4049,2014
8 1476,"MANDOVI AT TONCA, MARCELA, GOA",GOA,29.2,6.1,6.7,308,1.4,0.3,3355,5672,2014
9 3185,RIVER MANDOVII AT AMONA BRIDGE,GOA,29.6,6.4,6.7,414,1.0,2,6073,9423,2014
10 3186,RIVER MANDOVII AT IFFI JETTY,GOA,30.6,4.7,6.305,2.2,0.1,3478,4990,2014
11 3187,RIVER MANDOVII NEAR HOTEL MARRIOTT,GOA,30.1,6.3,7.6,77,2.3,0.1,2606,4301,2014
12 1543,"RIVER KALHA AT CHANDELMAN PERHEM, GOA",GOA,27.8,7.1,7.1,176,1.2,0.1,4573,7817,2014
13 1548,"RIVER ASSONORA AT ASSONORA, GOA",GOA,27.9,6.7,6.4,93,1.4,0.1,2147,3433,2014
14 2276,"RIVER BICHOLIM VARAZAN NAGAR , BICHOLIM",GOA,29.3,7.4,6.8,121,1.7,0.4,11633,18125,2014
15 2275,"RIVER CHAPORA NEAR ALORNA FORT ,PERNEM",GOA,29.2,6.9,7.620,1.1,0.1,3500,6300,2014
16 3189,RIVER CHAPORA AT SIOLIM,GOA,30.6,7.5,72,1.6,0.2,4995,9517,2014
17 1546,"RIVER KHANDEPAR AT OPA NAN PONDA, GOA",GOA,29.7,3.7,7.247,1.5,0.2,1095,2453,2014
18 2270,"RIVER KHANDEPAR AT CODLI NEAR BRIDGE ,U/S OPA WATERWORKS, SANGUEM",GOA,29.1,7.3,7.188,1.0,1.1,1286,3048,2014
19 2272,"RIVER KUSHAWATI NEAR BUND AT KEVONA,RIVON,SANGUEM",GOA,28.7,7.6,9.224,1.2,0.3,3896,6742,2014
20 1545,"RIVER MADAI AT DABOS NAN VALPOI, GOA",GOA,28.7,7.3,6.7,144,1.5,0.1,1940,3052,2014
21 2274,RIVER MAPUSA ON CULVERT ON HIGHWAY MAPUSANANPANAJI,GOA,29.5,5.3,6.8,319,1.8,0.3,6458,10250,2014
22 2271,"RIVER SAL PAZORKHONT,CUNCOLIM(NEAR CULVERT MARGAONAN CANACONA NATIONAL HIGHWAY)",GOA,29.6,3.6,4.79,1.6,1.4,7592,12842,2014
23 2273,"RIVER SAL NEAR HOTEL LEELA MOBOR,CAVELOSSIM",GOA,29.4,5.4,7.6,39,1.4,0.1,3176,6367,2014
24 3183,"RIVER SAL AT KHAREBAND, MARGAO",GOA,28.3,2.2,6.5,322,4.7,1.2,11210,14920,2014
25 184,"RIVER SAL AT ORLIM BRIDGE, ORLIM",GOA,30.1,5.2,7.1,192,2.6,0.3,5073,8925,2014
26 3190,RIVER SINQUERIM (CANDOLIM SIDE NEAR BRIDGE),GOA,30.3,5.6,7.5,282,1.8,0.1,3205,5082,2014
27 3191,RIVER SINQUERIM NEAR NERUL TEMPLE,GOA,30.5,5.5,7.4,275,1.5,0.1,4698,8625,2014
28 1547,"RIVER TALPONA AT CANACONA, GOA",GOA,29.1,7.3,6.7,55,1.4,0.1,2638,4003,2014
29 3188,RIVER TIRACOL AT TIRACOL,GOA,30.1,6.5,7.5,415,2.0,1.864,1538,2014
30 1544,"RIVER VALVANT AT SANKLI NAN BICHOLIM, GOA",GOA,29.2,7.2,6.3,100,1.5,0.1,7942,13575,2014
31 2651,"AMBA RIVER AT D/S OF WAKEN BRIDGENANU/S OF JS PETROCHEMICALS, WAKENPHATA, ROHA",MAHARASHTRA,25.1,6.6,7.8,95,4.9,0.2,16,36,2014
32 1461,"BHATSA AT D/S OF PISE DAM NEAR PISE VILLAGE (ULHAS), MAHARASHTRA",MAHARASHTRA,29.6,6.9,7.9,99,5.0,4,15,34,2014
```

File Edit Selection View Go Run Terminal Help Sprint1.ipynb - Visual Studio Code

Sprint1.ipynb X Sprint2.ipynb X

D: > study > ibm > Sprint1.ipynb > import numpy as np

+ Code + Markdown Run All Clear Outputs of All Cells Restart Variables Outline ... Python 3.8.7 64-bit

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

[1] Python

```
data = pd.read_csv('water_dataX.csv', encoding='ISO-8859-1', low_memory=False)
```

[2] Python

```
data.head()
```

[3] Python

	IN:STATION CODE	LOCATIONS	STATE	Temp	D.O. (mg/l)	PH	CONDUCTIVITY (Åumhos/cm)	B.O.D. (mg/l)	NITRATENAN N+ NITRITENANN (mg/l)	FECAL COLIFORM (MPN/100ml)	TOTAL COLIFORM (MPN/100ml)	Mean	year
0	1393	DAMANGANGA AT D/S OF MADHUBAN, DAMAN	DAMAN & DIU	30.6	6.7	7.5	203	NAN	0.1	11		27	2014
1	1399	ZUARI AT D/S OF PT. WHERE KUMBARJURIA CANAL JOIN	GOA	29.8	5.7	7.2	189	2	0.2	4953		8391	2014
2	1475	ZUARI AT PANCHAWADI	GOA	29.5	6.3	6.9	179	1.7	0.1	3243		5330	2014
3	3181	RIVER ZUARI AT BORIM BRIDGE	GOA	29.7	5.8	6.9	64	3.8	0.5	5382		8443	2014
4	3182	RIVER ZUARI AT MARCAIM JETTY	GOA	29.5	5.8	7.3	83	1.9	0.4	3428		5500	2014

Discovering Python Interpreters Jupyter Server: Local Cell 1 of 15 Go Live

File Edit Selection View Go Run Terminal Help Sprint1.ipynb - Visual Studio Code

Sprint1.ipynb X Sprint2.ipynb

D: > study > ibm > Sprint1.ipynb > import numpy as np

+ Code + Markdown Run All Clear Outputs of All Cells Restart Variables Outline ... Python 3.8.7 64-bit

```
data.describe()
```

[4] Python

	year
count	1991.000000
mean	2010.038172
std	3.057333
min	2003.000000
25%	2008.000000
50%	2011.000000
75%	2013.000000
max	2014.000000

```
data.info()
```

[5] Python

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1991 entries, 0 to 1990
Data columns (total 12 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   IN:STATION CODE      1991 non-null   object
1   LOCATIONS            1991 non-null   object
2   STATE                1991 non-null   object
3   Temp                1991 non-null   object
4   D.O. (mg/l)         1991 non-null   object
5   PH                  1991 non-null   object
```

Discovering Python Interpreters Jupyter Server: Local Cell 1 of 15 Go Live

```
File Edit Selection View Go Run Terminal Help Sprint1.ipynb - Visual Studio Code
Sprint1.ipynb Sprint2.ipynb
D: > study > itm > Sprint1.ipynb > import numpy as np
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Variables | Outline ... Python 3.8.7 64-bit

7 B.O.D. (mg/l) 1991 non-null object
8 NITRATENAN N+ NITRITENANN (mg/l) 1991 non-null object
9 FECAL COLIFORM (MPN/100ml) 1991 non-null object
10 TOTAL COLIFORM (MPN/100ml)Mean 1991 non-null object
11 year 1991 non-null int64
dtypes: int64(1), object(11)
memory usage: 186.8+ KB

data.shape
[6] Python
... (1991, 12)

data.isnull().any()
[7] Python
...
i=2STATION CODE False
LOCATIONS False
STATE False
Temp False
D.O. (mg/l) False
PH False
CONDUCTIVITY (µmhos/cm) False
B.O.D. (mg/l) False
NITRATENAN N+ NITRITENANN (mg/l) False
FECAL COLIFORM (MPN/100ml) False
TOTAL COLIFORM (MPN/100ml)Mean False
year False
dtype: bool
```

```
File Edit Selection View Go Run Terminal Help Sprint1.ipynb - Visual Studio Code
Sprint1.ipynb Sprint2.ipynb
D: > study > itm > Sprint1.ipynb > import numpy as np
+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Variables | Outline ... Python 3.8.7 64-bit

data.isnull().sum()
[8] Python
...
i=2STATION CODE 0
LOCATIONS 0
STATE 0
Temp 0
D.O. (mg/l) 0
PH 0
CONDUCTIVITY (µmhos/cm) 0
B.O.D. (mg/l) 0
NITRATENAN N+ NITRITENANN (mg/l) 0
FECAL COLIFORM (MPN/100ml) 0
TOTAL COLIFORM (MPN/100ml)Mean 0
year 0
dtype: int64

data.dtypes
[9] Python
...
i=2STATION CODE object
LOCATIONS object
STATE object
Temp object
D.O. (mg/l) object
PH object
CONDUCTIVITY (µmhos/cm) object
B.O.D. (mg/l) object
NITRATENAN N+ NITRITENANN (mg/l) object
FECAL COLIFORM (MPN/100ml) object
```

```
File Edit Selection View Go Run Terminal Help Sprint1.ipynb - Visual Studio Code

Sprint1.ipynb x Sprint2.ipynb x
D: > study > ibm > Sprint1.ipynb > import numpy as np
+ Code + Markdown Run All Clear Outputs of All Cells Restart Variables Outline ... Python 3.8.7 64-bit

B.O.D. (mg/l) object
NITRATENAN N+ NITRITENANN (mg/l) object
FECAL COLIFORM (MPN/100ml) object
TOTAL COLIFORM (MPN/100ml)Mean object
year int64
dtype: object

data['Temp']=pd.to_numeric(data['Temp'],errors='coerce')
data['D.O. (mg/l)']=pd.to_numeric(data['D.O. (mg/l)'],errors='coerce')
data['PH']=pd.to_numeric(data['PH'],errors='coerce')
data['B.O.D. (mg/l)']=pd.to_numeric(data['B.O.D. (mg/l)'],errors='coerce')
data['CONDUCTIVITY (µmhos/cm)']=pd.to_numeric(data['CONDUCTIVITY (µmhos/cm)'],errors='coerce')
data['NITRATENAN N+ NITRITENANN (mg/l)']=pd.to_numeric(data['NITRATENAN N+ NITRITENANN (mg/l)'],errors='coerce')
data['TOTAL COLIFORM (MPN/100ml)Mean']=pd.to_numeric(data['TOTAL COLIFORM (MPN/100ml)Mean'],errors='coerce')
data.dtypes

[10] Python Python

i_w_STATION CODE object
LOCATIONS object
STATE object
Temp float64
D.O. (mg/l) float64
PH float64
CONDUCTIVITY (µmhos/cm) float64
B.O.D. (mg/l) float64
NITRATENAN N+ NITRITENANN (mg/l) float64
FECAL COLIFORM (MPN/100ml) object
TOTAL COLIFORM (MPN/100ml)Mean float64
year int64
dtype: object

Discovering Python Interpreters Jupyter Server: Local Cell 1 of 15 Go Live
```

```
File Edit Selection View Go Run Terminal Help Sprint1.ipynb - Visual Studio Code

Sprint1.ipynb x Sprint2.ipynb
D: > study > ibm > Sprint1.ipynb > import numpy as np
+ Code + Markdown Run All Clear Outputs of All Cells Restart Variables Outline ... Python 3.8.7 64-bit

data.isnull().sum()

[11] Python Python

i_w_STATION CODE 0
LOCATIONS 0
STATE 0
Temp 92
D.O. (mg/l) 31
PH 8
CONDUCTIVITY (µmhos/cm) 25
B.O.D. (mg/l) 43
NITRATENAN N+ NITRITENANN (mg/l) 225
FECAL COLIFORM (MPN/100ml) 0
TOTAL COLIFORM (MPN/100ml)Mean 132
year 0
dtype: int64

data['Temp'].fillna(data['Temp'].mean(),inplace=True)
data['D.O. (mg/l)'].fillna(data['D.O. (mg/l)'].mean(),inplace=True)
data['PH'].fillna(data['PH'].mean(),inplace=True)
data['CONDUCTIVITY (µmhos/cm)'].fillna(data['CONDUCTIVITY (µmhos/cm)'].mean(),inplace=True)
data['B.O.D. (mg/l)'].fillna(data['B.O.D. (mg/l)'].mean(),inplace=True)
data['NITRATENAN N+ NITRITENANN (mg/l)'].fillna(data['NITRATENAN N+ NITRITENANN (mg/l)'].mean(),inplace=True)
data['TOTAL COLIFORM (MPN/100ml)Mean'].fillna(data['TOTAL COLIFORM (MPN/100ml)Mean'].mean(),inplace=True)

[12] Python

data.drop(["FECAL COLIFORM (MPN/100ml)",axis=1,inplace=True)

[13] Python

Discovering Python Interpreters Jupyter Server: Local Cell 1 of 15 Go Live
```

FileEditSelectionViewGoRunTerminalHelpSprint1.ipynb - Visual Studio Code

Sprint1.ipynb xSprint2.ipynb

D: > study > ibm > Sprint1.ipynb > import numpy as np

+ Code + Markdown | Run All | Clear Outputs of All Cells | Restart | Variables | Outline ...

Python 3.8.7 64-bit

```
data=data.rename(columns = {'D.O. (mg/l)': 'do'})
data=data.rename(columns = {'CONDUCTIVITY (µmhos/cm)': 'co'})
data=data.rename(columns = {'B.O.D. (mg/l)': 'bod'})
data=data.rename(columns = {'NITRATE+N+ NITRITE+NAN (mg/l)': 'na'})
data=data.rename(columns = {'TOTAL COLIFORM (MPN/100ml)Mean': 'tc'})
data=data.rename(columns = {'STATION CODE': 'station'})
data=data.rename(columns = {'LOCATIONS': 'location'})
data=data.rename(columns = {'STATE': 'state'})
data=data.rename(columns = {'PH': 'ph'})
data
```

[14]

Python

	STATION CODE	location	state	Temp	do	ph	co	bod	na	tc	year
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	2014
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	2014
2	1475	ZUARI AT PANCHAWADI	GOA	29.500000	6.3	6.9	179.0	1.700000	0.100000	5330.0	2014
3	3181	RIVER ZUARI AT BORIM BRIDGE	GOA	29.700000	5.8	6.9	64.0	3.800000	0.500000	8443.0	2014
4	3182	RIVER ZUARI AT MARCAIM JETTY	GOA	29.500000	5.8	7.3	83.0	1.900000	0.400000	5500.0	2014
...
1986	1330	TAMBIRAPARANI AT ARUMUGANERI, TAMILNADU	NAN	26.209814	7.9	738.0	7.2	2.700000	0.518000	202.0	2003
1987	1450	PALAR AT VANİYAMBADI WATER SUPPLY HEAD WORK, T...	NAN	29.000000	7.5	585.0	6.3	2.600000	0.155000	315.0	2003
1988	1403	GUMTI AT U/S SOUTH TRIPURA,TRIPURA	NAN	28.000000	7.6	98.0	6.2	1.200000	1.623079	570.0	2003
1989	1404	GUMTI AT D/S SOUTH TRIPURA, TRIPURA	NAN	28.000000	7.7	91.0	6.5	1.300000	1.623079	562.0	2003
1990	1726	CHANDRAPUR, AGARTALA D/S OF HAORA RIVER, TRIPURA	NAN	29.000000	7.6	110.0	5.7	1.100000	1.623079	546.0	2003

1991 rows x 11 columns

0 6

Discovering Python Interpreters

Jupyter Server: Local

Cell 1 of 15

Go Live