

School of Computer Science and Engineering J Component report

Programme : M. Tech Integrated CSE with Business Analytics

Course Title : Exploratory Data Analytics

Course Code : CSE3040

Slot : F1

: Missing data handling and outlier analysis Title

Team Members: Keerthana M | 20MIA1082

Varsini SR | 20MIA1087

Sonalika P | 20MIA1089

Faculty : Sweetlin Hemalatha C

Sign: Sp. 29 4 22

MISSING DATA HANDLING AND OUTLIER ANALYSIS

Data preprocessing is the first and foremost step in data analysis. Data preprocessing converts the data in its raw form into a more readable format (graphs, documents, etc.), which can be interpreted and analyzed in the further stages.

The data has been made publicly available by the Central Pollution Control Board: https://cpcb.nic.in/ which is the official portal of Government of India. They also have a real-time monitoring app: https://app.cpcbccr.com/AQI_India/

Why is handling missing data important?

- 1. Missing data reduces statistical power, which refers to the probability that the test will reject the null hypothesis when it is false.
- 2. The lost data can cause bias in the estimation of parameters.
- 3. It can reduce the representativeness of the samples.
- 4. It may complicate the analysis of the study.

Each of these distortions may threaten the validity of the trials and can lead to invalid conclusions.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
df = pd.read excel (r'/content/Data 2017-2022.xlsx',
sheet name='Chennai')
df.head()
             From Date
                                     To Date PM2.5 (ug/m3)
                                                             PM10
(ug/m3) \setminus
0 01-Jan-2017 - 00:00 02-Jan-2017 - 00:00
                                                      32.61
NaN
1 02-Jan-2017 - 00:00
                        03-Jan-2017 - 00:00
                                                      22.93
NaN
                        04-Jan-2017 - 00:00
2 03-Jan-2017 - 00:00
                                                      24.19
NaN
                        05-Jan-2017 - 00:00
3 04-Jan-2017 - 00:00
                                                      33.61
NaN
                        06-Jan-2017 - 00:00
4 05-Jan-2017 - 00:00
                                                     129.38
NaN
   NO (ug/m3)
               NO2 (ug/m3) NOx (ppb)
                                       NH3 (ug/m3)
                                                     S02 (uq/m3)
                                                                  C0
(mq/m3)
0
         2.36
                      9.78
                                   NaN
                                                            2.11
                                                NaN
NaN
         2.33
                      8.21
                                   NaN
                                                NaN
                                                            2.86
1
NaN
```

```
2
        11.39
                       17.28
                                     NaN
                                                    NaN
                                                                 7.73
NaN
3
          6.06
                       12.32
                                     NaN
                                                    NaN
                                                                 2.72
NaN
                       12.67
                                     NaN
                                                                 2.65
4
          5.58
                                                    NaN
NaN
   Ozone (ug/m3)
                   Benzene (ug/m3)
                                      Toluene (ug/m3)
0
            24.63
                                0.52
                                                   2.95
            20.49
                                0.13
                                                   2.01
1
2
            13.04
                                0.45
                                                   3.52
3
                                0.65
                                                   3.98
            19.42
4
            25.89
                                0.60
                                                   3.53
df.shape
(1885, 13)
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884
Data columns (total 13 columns):
 #
     Column
                        Non-Null Count
                                          Dtype
- - -
     -----
                        _ _ _ _ _ _ _ _ _ _ _ _ _ _ _
 0
     From Date
                        1885 non-null
                                          object
 1
     To Date
                        1885 non-null
                                          object
 2
     PM2.5 (ug/m3)
                        1841 non-null
                                          float64
 3
     PM10 (uq/m3)
                        306 non-null
                                          float64
 4
     NO (ug/m3)
                        1844 non-null
                                          float64
 5
     N02 (ug/m3)
                        1843 non-null
                                          float64
 6
     NOx (ppb)
                        1598 non-null
                                          float64
 7
     NH3 (ug/m3)
                        306 non-null
                                          float64
     S02 \left( uq/m3 \right)
 8
                        1830 non-null
                                          float64
 9
     CO (mq/m3)
                                          float64
                        1687 non-null
                                          float64
 10
     Ozone (ug/m3)
                        1833 non-null
 11
     Benzene (ug/m3)
                        1859 non-null
                                          float64
     Toluene (ug/m3)
                        1859 non-null
                                          float64
 12
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
#no of null values in each column
df.isna().sum()
From Date
                        0
To Date
                        0
PM2.5 (ug/m3)
                       44
PM10 (uq/m3)
                     1579
NO (ug/m3)
                       41
N02 (ug/m3)
                       42
NOx (ppb)
                      287
```

```
NH3 (uq/m3)
                    1579
S02 (ug/m3)
                      55
CO (mg/m3)
                     198
Ozone (uq/m3)
                      52
Benzene (ug/m3)
                      26
Toluene (ug/m3)
                      26
dtype: int64
df.mean(numeric only=True)
PM2.5 (ug/m3)
                    30.500435
PM10 (ug/m3)
                    58.150719
NO (uq/m3)
                     6.952950
N02 (uq/m3)
                    12.557347
NOx (ppb)
                    16.026608
NH3 (ug/m3)
                    12.271961
S02 (ug/m3)
                     6.551770
CO (mg/m3)
                     0.827825
Ozone (ug/m3)
                    27.580562
Benzene (ug/m3)
                     0.583018
Toluene (ug/m3)
                     1.940172
dtype: float64
df.median(numeric only=True)
PM2.5 (ug/m3)
                    27.280
PM10 (ug/m3)
                    67.250
NO (ug/m3)
                     5.460
N02 (ug/m3)
                    11.280
NOx (ppb)
                    14.615
NH3 (ug/m3)
                    14.455
S02 (ug/m3)
                     4.880
CO (mq/m3)
                     0.740
Ozone (ug/m3)
                    24.680
Benzene (ug/m3)
                     0.000
Toluene (ug/m3)
                     0.190
dtype: float64
df.var(numeric only=True)
PM2.5 (uq/m3)
                    411.678015
PM10 (ug/m3)
                    672.376420
NO (ug/m3)
                     31.483505
N02 (ug/m3)
                     75.819259
NOx (ppb)
                     77.394611
NH3 (ug/m3)
                     30.887305
S02 (ug/m3)
                     25.583531
CO (mg/m3)
                      2.167166
```

Ozone (ug/m3)

Benzene (ug/m3)

282.994183

4.968048

```
Toluene (ug/m3)
                    19.185895
dtype: float64
df.std(numeric only=True)
PM2.5 (ug/m3)
                    20.289850
PM10 (ug/m3)
                    25.930222
NO (ug/m3)
                    5.611016
N02 (ug/m3)
                    8.707426
NOx (ppb)
                    8.797421
                    5.557635
NH3 (ug/m3)
S02 (ug/m3)
                    5.058016
CO (mq/m3)
                    1.472130
Ozone (ug/m3)
                    16.822431
Benzene (ug/m3)
                    2.228912
Toluene (ug/m3)
                    4.380171
dtype: float64
df.cov()
                 PM2.5 (ug/m3)
                                 PM10 (ug/m3)
                                                NO (ug/m3)
                                                            N02 (ug/m3)
PM2.5 (ug/m3)
                    411.678015
                                    30.890050
                                                 16.066220
                                                              51.241957
PM10 (uq/m3)
                     30.890050
                                   672.376420
                                                -44.806987
                                                             -24.387463
NO (ug/m3)
                      16.066220
                                   -44.806987
                                                 31.483505
                                                              22.519439
N02 (ug/m3)
                                   -24.387463
                                                 22.519439
                                                              75.819259
                     51.241957
                                   -61.393595
                                                 36.589019
NOx (ppb)
                     40.344852
                                                              55.207391
NH3 (ug/m3)
                      13.933097
                                   -18.579278
                                                 19.546962
                                                              22.055663
S02 (ug/m3)
                      -0.923446
                                     0.531256
                                                 -1.680204
                                                              -3.889626
CO (mg/m3)
                                     0.001428
                                                 -0.332261
                                                              -0.253260
                      -1.522403
Ozone (ug/m3)
                     60.713479
                                    38.184933
                                                 -0.932329
                                                              14.139986
Benzene (ug/m3)
                                     0.034594
                      0.625041
                                                  1.625605
                                                               1.354755
Toluene (ug/m3)
                      10.737620
                                     0.074053
                                                  4.639590
                                                               9.701272
                 NOx (ppb)
                             NH3 (ug/m3)
                                          S02 (ug/m3)
                                                        CO (mg/m3)
                               13.933097
                                                         -1.522403
PM2.5 (ug/m3)
                 40.344852
                                             -0.923446
PM10 (ug/m3)
                 -61.393595
                              -18.579278
                                             0.531256
                                                          0.001428
NO (uq/m3)
                 36.589019
                               19.546962
                                             -1.680204
                                                         -0.332261
```

22.055663

-3.889626

-0.253260

N02 (ug/m3)

55.207391

NOx (ppb) NH3 (ug/m3) S02 (ug/m3) C0 (mg/m3) Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3)	39.637911 -3.141972 0.158700	30.887305 8 8.791461 25 0.259349 -0 23.173855 2 0.043199 -1	3.791461 0 5.583531 -0 0.472042 2 2.729160 -0 098227 -0	.158700 .259349 .472042 .167166 .128536 .030333 .073975
PM2.5 (ug/m3) PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO (mg/m3) Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3)	Ozone (ug/m3) 60.713479 38.184933 -0.932329 14.139986 5.818156 23.173855 2.729160 -0.128536 282.994183 -0.550818 -0.989454	Benzene (ug/m 0.6250 0.0345 1.6256 1.3547 2.3267 0.0431 -1.0982 -0.0303 -0.5508 4.9680 4.8770	041 10 694 0 605 4 755 9 780 11 .99 0 227 -3 333 -0 318 -0 048 4	(ug/m3) .737620 .074053 .639590 .701272 .059783 .036882 .746441 .073975 .989454 .877070
df.corr()				
\	PM2.5 (ug/m3)	PM10 (ug/m3)	NO (ug/m3)	NO2 (ug/m3)
PM2.5 (ug/m3)	1.000000	0.077045	0.140907	0.290013
PM10 (ug/m3)	0.077045	1.000000	-0.342711	-0.205761
NO (ug/m3)	0.140907	-0.342711	1.000000	0.460980
NO2 (ug/m3)	0.290013	-0.205761	0.460980	1.000000
NOx (ppb)	0.217120	-0.272646	0.714484	0.724070
NH3 (ug/m3)	0.161947	-0.128751	0.702132	0.871728
S02 (ug/m3)	-0.009008	0.004149	-0.059458	-0.089198
CO (mg/m3)	-0.049465	0.000288	-0.039316	-0.019531
Ozone (ug/m3)	0.178225	0.058454	-0.009927	0.097121
Benzene (ug/m3)	0.013715	0.011016	0.129351	0.069390
Toluene (ug/m3)	0.120031	0.022940	0.188078	0.253154

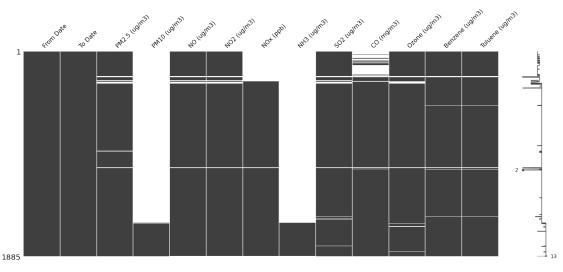
NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO (mg/m3) \setminus

PM2.5 (ug/m3)	0.217120	0.161947	-0.009008	-0.049465
PM10 (ug/m3)	-0.272646	-0.128751	0.004149	0.000288
NO (ug/m3)	0.714484	0.702132	-0.059458	-0.039316
NO2 (ug/m3)	0.724070	0.871728	-0.089198	-0.019531
NOx (ppb)	1.000000	0.825683	-0.068077	0.011973
NH3 (ug/m3)	0.825683	1.000000	0.323777	0.242040
S02 (ug/m3)	-0.068077	0.323777	1.000000	-0.061407
CO (mg/m3)	0.011973	0.242040	-0.061407	1.000000
Ozone (ug/m3)	0.038354	0.170788	0.032032	-0.008149
Benzene (ug/m3)	0.122707	0.064346	-0.096417	-0.008974
Toluene (ug/m3)	0.278826	0.053442	-0.167630	-0.010942
	Ozone (ug/m3)	Benzene ((ug/m3) Tolu	uene (ug/m3)
PM2.5 (ug/m3)	Ozone (ug/m3) 0.178225		(ug/m3) Tolu .013715	uene (ug/m3) 0.120031
PM2.5 (ug/m3) PM10 (ug/m3)		0.	-	_
	0.178225	0. 0.	013715	0.120031
PM10 (ug/m3) NO (ug/m3)	0.178225 0.058454	0. 0. 0.	013715 011016	0.120031 0.022940
PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3)	0.178225 0.058454 -0.009927	0. 0. 0. 0.	013715 011016 129351	0.120031 0.022940 0.188078
PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb)	0.178225 0.058454 -0.009927 0.097121 0.038354	0. 0. 0. 0.	.013715 .011016 .129351 .069390 .122707	0.120031 0.022940 0.188078 0.253154 0.278826
PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3)	0.178225 0.058454 -0.009927 0.097121 0.038354 0.170788	0. 0. 0. 0. 0.	.013715 .011016 .129351 .069390 .122707 .064346	0.120031 0.022940 0.188078 0.253154 0.278826 0.053442
PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3)	0.178225 0.058454 -0.009927 0.097121 0.038354 0.170788 0.032032	0. 0. 0. 0. 0.	013715 011016 129351 069390 122707 064346 096417	0.120031 0.022940 0.188078 0.253154 0.278826 0.053442 -0.167630
PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO (mg/m3)	0.178225 0.058454 -0.009927 0.097121 0.038354 0.170788 0.032032 -0.008149	0. 0. 0. 0. 0. -0.	.013715 .011016 .129351 .069390 .122707 .064346 .096417	0.120031 0.022940 0.188078 0.253154 0.278826 0.053442 -0.167630 -0.010942
PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO (mg/m3) Ozone (ug/m3)	0.178225 0.058454 -0.009927 0.097121 0.038354 0.170788 0.032032 -0.008149 1.000000	0. 0. 0. 0. 0. -0. -0.	.013715 .011016 .129351 .069390 .122707 .064346 .096417 .008974	0.120031 0.022940 0.188078 0.253154 0.278826 0.053442 -0.167630 -0.010942 -0.013333
PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO (mg/m3)	0.178225 0.058454 -0.009927 0.097121 0.038354 0.170788 0.032032 -0.008149	0. 0. 0. 0. 0. -0. -0.	.013715 .011016 .129351 .069390 .122707 .064346 .096417	0.120031 0.022940 0.188078 0.253154 0.278826 0.053442 -0.167630 -0.010942

#this is plot shows how amuch missing values in the dataset, and its clearly visible that almost PM10 and NH3 has no data.

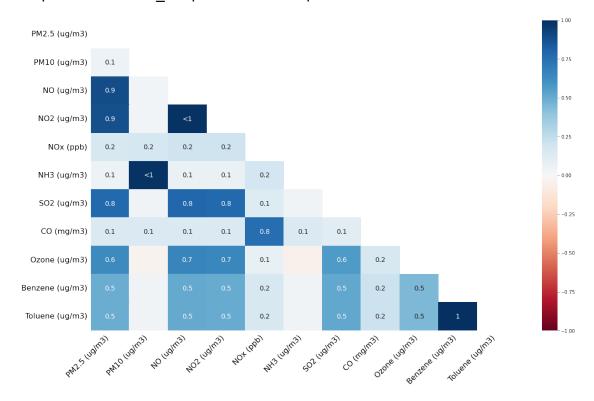
import missingno as msno
msno.matrix(df)

<matplotlib.axes._subplots.AxesSubplot at 0x7fad886b27d0>



#helps in visualizing the correlation between all the columns # Both NO and NO2 has strong positive correlation with PM2.5 msno.heatmap(df)

<matplotlib.axes. subplots.AxesSubplot at 0x7fad85c1aa90>



MISSING DATA HANDLING

#Listwise deletion

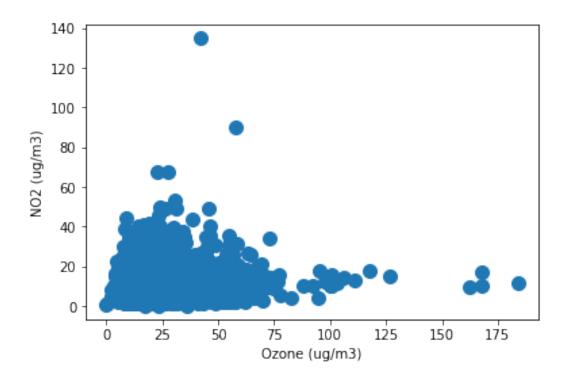
The process of deleting data for any case that has one or more missing values. It is also known as complete case analysis.

```
df1 = pd.read excel (r'/content/Data 2017-2022.xlsx',
sheet_name='Chennai')
df1.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884
Data columns (total 13 columns):
 #
     Column
                       Non-Null Count
                                        Dtype
- - -
 0
     From Date
                       1885 non-null
                                        object
                       1885 non-null
                                        object
 1
     To Date
 2
                                        float64
     PM2.5 (ug/m3)
                       1841 non-null
 3
     PM10 (ug/m3)
                       306 non-null
                                        float64
 4
     NO (ug/m3)
                       1844 non-null
                                        float64
 5
     N02 (ug/m3)
                       1843 non-null
                                        float64
                                        float64
 6
     NOx (ppb)
                       1598 non-null
 7
     NH3 (ug/m3)
                       306 non-null
                                        float64
```

```
8
     S02 (uq/m3)
                      1830 non-null
                                       float64
 9
     CO (mg/m3)
                                       float64
                      1687 non-null
 10
    Ozone (ug/m3)
                      1833 non-null
                                       float64
 11
    Benzene (ug/m3)
                      1859 non-null
                                       float64
     Toluene (ug/m3)
                     1859 non-null
                                       float64
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
dfl.isnull().sum() #finding the count of null values
From Date
                      0
To Date
                      0
PM2.5 (ug/m3)
                     44
PM10 (ug/m3)
                   1579
NO (ug/m3)
                     41
N02 (ug/m3)
                     42
NOx (ppb)
                    287
NH3 (ug/m3)
                   1579
S02 (uq/m3)
                     55
CO (mg/m3)
                    198
Ozone (ug/m3)
                     52
Benzene (ug/m3)
                     26
Toluene (ug/m3)
                     26
dtype: int64
df1.shape
(1885, 13)
#deleting all the Rows which have missing values
df1 new =df1.dropna()
df1 new.shape
(286, 13)
```

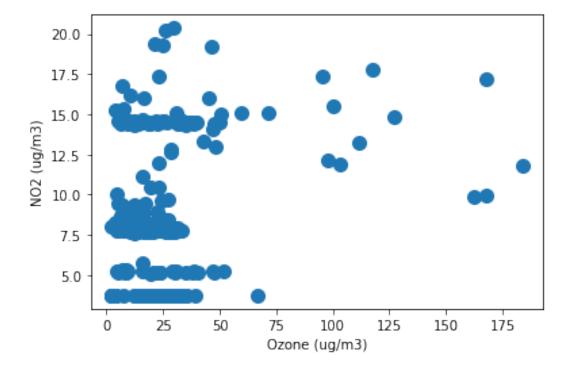
The size of the data has drastically reduced after deleting the missing values. If we delete all the rows where we have NaN value the sample size is 286, which is very less value to evaluate the model.

```
df.plot.scatter(x = "0zone (ug/m3)", y = 'N02 (ug/m3)', s = 100);
```



Scatter plot of Ozone and No2 levels before deletion.

$$dfl_new.plot.scatter(x = "0zone (ug/m3)", y = 'N02 (ug/m3)', s = 100);$$



Listwise deletion scatter plot of Ozone and No2 levels.

Pairwise deletion

A method in which data for a variable pertinent to a specific assessment are included, even if values for the same individual on other variables are missing. Its is also known as available case analysis.

```
df2 = pd.read excel (r'/content/Data 2017-2022.xlsx',
sheet name='Chennai')
#finding percentage of misssing vale in each column
Percent Missing Value = df2.isnull().sum()*100/len(df2)
Percent_Missing_Value
From Date
                    0.00000
To Date
                    0.000000
PM2.5 (ug/m3)
                    2.334218
PM10 (ug/m3)
                   83.766578
NO (uq/m3)
                   2.175066
N02 (ug/m3)
                    2.228117
NOx (ppb)
                   15.225464
NH3 (ug/m3)
                   83.766578
S02 (ug/m3)
                    2.917772
CO (mg/m3)
                   10.503979
Ozone (ug/m3)
                    2.758621
Benzene (ug/m3)
                    1.379310
Toluene (ug/m3)
                    1.379310
dtype: float64
```

If columns have more than half of rows as null then the entire column can be dropped.In PM10 and Nh3 most of the columns almost 84% of the data is missing,so dropping those columns .

df2.head()

F	rom Date	To Date F	PM2.5 (ug/m3)	PM10
(ug/m3) \			. 3, ,	
0 01-Jan-2017	7 - 00:00 02-Jan-2017	- 00:00	32.61	
NaN				
1 02-Jan-2017	7 - 00:00 03-Jan-2017	- 00:00	22.93	
NaN				
2 03-Jan-2017	7 - 00:00 04-Jan-2017	- 00:00	24.19	
NaN				
3 04-Jan-2017	7 - 00:00 05-Jan-2017	- 00:00	33.61	
NaN				
4 05-Jan-2017	7 - 00:00 06-Jan-2017	- 00:00	129.38	
NaN				
NO (ug/m3)	NO2 (ug/m3) NOx (ppl	o) NH3 (ug	g/m3) SO2 (u	g/m3) CO
$(mg/m3) \setminus$				
0 2.36	9.78 Na	aΝ	NaN	2.11
NaN				

```
2.33
                      8.21
                                   NaN
                                                 NaN
                                                             2.86
1
NaN
                      17.28
                                                             7.73
        11.39
                                   NaN
                                                 NaN
NaN
                      12.32
                                   NaN
                                                             2.72
3
         6.06
                                                 NaN
NaN
                     12.67
4
         5.58
                                   NaN
                                                 NaN
                                                             2.65
NaN
   Ozone (ug/m3)
                  Benzene (ug/m3)
                                    Toluene (ug/m3)
           24.63
0
                              0.52
                                                2.95
                              0.13
1
           20.49
                                                2.01
2
           13.04
                              0.45
                                                3.52
3
           19.42
                              0.65
                                                3.98
4
           25.89
                              0.60
                                                3.53
#droping PM10 column
df 2 = df2.drop("PM10 (ug/m3)",axis=1)
#droping Nh3 column
df 2 = df2.drop("NH3 (ug/m3)",axis=1)
df 2.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884
Data columns (total 12 columns):
     Column
                      Non-Null Count
                                       Dtype
- - -
     -----
                       _____
                                       - - - - -
 0
     From Date
                       1885 non-null
                                       object
 1
     To Date
                       1885 non-null
                                       object
 2
     PM2.5 (ug/m3)
                       1841 non-null
                                       float64
 3
     PM10 (ug/m3)
                                       float64
                       306 non-null
 4
                       1844 non-null
                                       float64
     NO (uq/m3)
 5
     N02 (ug/m3)
                       1843 non-null
                                       float64
 6
     NOx (ppb)
                       1598 non-null
                                       float64
 7
     S02 (uq/m3)
                       1830 non-null
                                       float64
 8
     CO (mq/m3)
                       1687 non-null
                                       float64
 9
     Ozone (ug/m3)
                       1833 non-null
                                       float64
 10
    Benzene (ug/m3)
                                       float64
                      1859 non-null
     Toluene (ug/m3)
                      1859 non-null
 11
                                       float64
dtypes: float64(10), object(2)
memory usage: 176.8+ KB
```

To find out the relation between Ozone and No2, deleting the missing values pertinent to the columns.

```
#deleting the missing values from ozone
df2.dropna(subset=['Ozone (ug/m3)'],how='any',inplace=True)
df2['Ozone (ug/m3)'].isnull().sum()
```

```
#deleting the missing values from Nh3
df2.dropna(subset=['N02 (ug/m3)'],how='any',inplace=True)
df2['N02 (ug/m3)'].isnull().sum()
0
df.shape #intial size of the data
(1885, 13)
df2.shape #size after deletion
(1822, 13)
df.describe()
       PM2.5 (ug/m3)
                       PM10 (uq/m3)
                                       NO (ug/m3)
                                                    N02 (uq/m3)
                                                                    N0x
(ppb)
count
         1841.000000
                         306,000000
                                      1844.000000
                                                    1843.000000
1598,000000
           30.500435
                          58.150719
                                         6.952950
                                                      12.557347
mean
16.026608
           20.289850
                          25.930222
                                                       8.707426
std
                                         5.611016
8.797421
            0.410000
                          21.600000
                                         0.010000
                                                       0.020000
min
0.000000
25%
           16.540000
                          36.850000
                                         3.295000
                                                       6.510000
9.860000
50%
                          67.250000
           27.280000
                                         5.460000
                                                      11.280000
14.615000
                          69.415000
                                         9,400000
75%
           39.530000
                                                      16.470000
20.872500
          278,970000
                         371.610000
                                        98.620000
                                                     134.760000
max
106.740000
       NH3 (ug/m3)
                     S02 (ug/m3)
                                    CO (mq/m3)
                                                 Ozone (ug/m3)
                                                                 Benzene
(ug/m3)
                     1830.000000
                                   1687.000000
                                                   1833.000000
count
        306.000000
1859.000000
mean
         12.271961
                        6.551770
                                      0.827825
                                                     27.580562
0.583018
          5.557635
                        5.058016
                                      1.472130
                                                     16.822431
std
2.228912
min
          5.360000
                        0.090000
                                      0.000000
                                                      0.100000
0.000000
          5.360000
25%
                        3.890000
                                      0.600000
                                                     16.770000
0.000000
50%
         14.455000
                        4.880000
                                      0.740000
                                                     24.680000
0.000000
                        7.205000
                                      0.910000
                                                     34.770000
75%
         16.675000
```

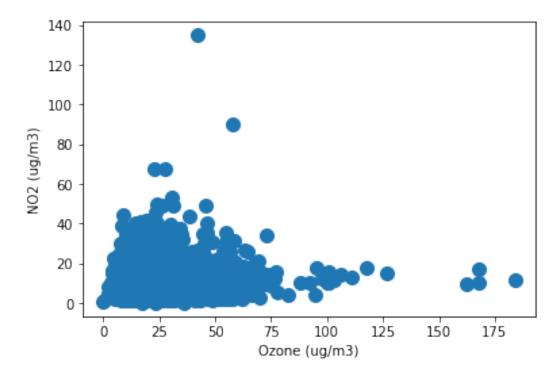
```
0.365000
         33.680000
                       37.180000
                                     48.020000
                                                    183.990000
max
46.230000
       Toluene (ug/m3)
           1859.000000
count
mean
               1.940172
               4.380171
std
               0.000000
min
25%
               0.00000
50%
               0.190000
75%
               2.755000
            121.150000
max
Computing the statistical variables after deletion
df['N02 (ug/m3)'].describe()
         1843.000000
count
           12.557347
mean
            8.707426
std
min
            0.020000
25%
            6.510000
50%
           11.280000
75%
           16.470000
          134.760000
max
Name: NO2 (ug/m3), dtype: float64
df2['NO2 (ug/m3)'].describe()
count
         1822.000000
mean
           12.634682
            8,704530
std
min
            0.020000
25%
            6.672500
50%
           11.360000
           16.527500
75%
          134.760000
max
Name: NO2 (ug/m3), dtype: float64
df['Ozone (ug/m3)'].describe()
         1833.000000
count
           27.580562
mean
           16.822431
std
            0.100000
min
25%
           16.770000
50%
           24.680000
           34.770000
75%
          183.990000
max
Name: Ozone (ug/m3), dtype: float64
```

```
df2['Ozone (ug/m3)'].describe()
```

```
1822.000000
count
           27.633079
mean
           16.725983
std
min
            0.100000
25%
           16.832500
           24.760000
50%
75%
           34.792500
          183.990000
max
```

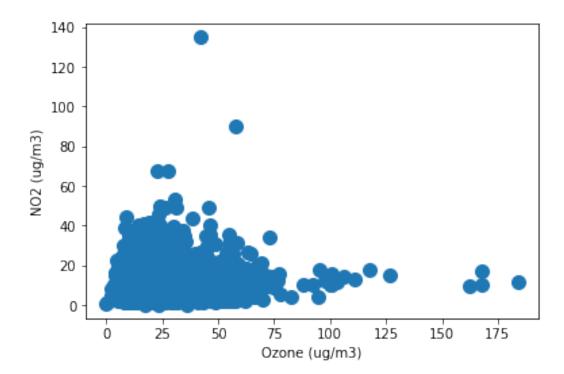
Name: Ozone (ug/m3), dtype: float64

```
df.plot.scatter(x = "0zone (ug/m3)", y = 'N02 (ug/m3)', s = 100);
```



Scatter plot of Ozone and No2 levels before deletion.

```
df2.plot.scatter(x = "0zone (ug/m3)", y = 'N02 (ug/m3)', s = 100);
```



Pairwise deletion scatter plot of Ozone and No2 levels. After deletion this resembles similar to the original dataset whereas in listwise the data is more dispersed and scattered.

#Mean imputation

Mean imputation is a method in which the missing values are replaced with the mean value of the entire feature column.

```
df3 = pd.read excel (r'/content/Data 2017-2022.xlsx',
sheet name='Chennai')
#all the missing values in the columns are replaced withe the mean
values of the same column
x=['PM2.5 (ug/m3)','PM10 (ug/m3)','N0 (ug/m3)','N02 (ug/m3)','N0x (ppb)','NH3 (ug/m3)','S02 (ug/m3)','C0 (mg/m3)','Ozone
(ug/m3)', 'Benzene (ug/m3)', 'Toluene (ug/m3)']
df3[x] = df3[x].fillna(df3[x].mean())
df3.head(100)
               From Date
                                                   PM2.5 (ug/m3)
                                         To Date
                                                                    PM10
(ug/m3) \setminus
    01-Jan-2017 - 00:00 02-Jan-2017 - 00:00
                                                            32.61
58.150719
    02-Jan-2017 - 00:00 03-Jan-2017 - 00:00
                                                            22.93
58.150719
    03-Jan-2017 - 00:00 04-Jan-2017 - 00:00
                                                            24.19
58.150719
    04-Jan-2017 - 00:00 05-Jan-2017 - 00:00
                                                            33.61
58.150719
```

4 05-Ja 58.150719		00:00	06-Jan-20	17 - 00:0	Θ	129.38	
 95 06-Ap 58.150719		00:00	07-Apr-20	17 - 00:0	0	42.08	
96 07-Ap 58.150719	r-2017 -	00:00	08-Apr-20	17 - 00:0	0	37.94	
	r-2017 -	00:00	09-Apr-20	17 - 00:0	0	35.41	
	r-2017 -	00:00	10-Apr-20	17 - 00:0	0	34.46	
	r-2017 -	00:00	11-Apr-20	17 - 00:0	Θ	40.33	
	g/m3) N(02 (ug/r	m3) NOx (ppb) NH3	(ug/m3)	S02 (ug/m3)	CC
0.827825	2.36	9	.78 16.02	6608 1	2.271961	2.11	
0.827825 1 0.827825	2.33	8	.21 16.02	6608 1	2.271961	2.86	
	11.39	17	.28 16.02	6608 1	2.271961	7.73	
3	6.06	12	.32 16.02	6608 1	2.271961	2.72	
0.827825 4 0.827825	5.58	12	.67 16.02	6608 1	2.271961	2.65	
 95 0.827825	4.36	7	.03 16.02	6608 1	2.271961	3.16	
96 1.230000	4.06	8	.86 16.02	6608 1	2.271961	3.73	
97 1.270000	4.75	13	.42 16.02	6608 1	2.271961	3.85	
98 1.150000	3.42	9	.20 16.02	6608 1	2.271961	3.63	
99 1.240000	4.31	9	.92 16.02	6608 1	2.271961	3.93	
0zone 0 1 2 3	(ug/m3) 24.63 20.49 13.04 19.42 25.89	Benzer	ne (ug/m3) 0.52 0.13 0.45 0.65 0.60		(ug/m3) 2.95 2.01 3.52 3.98 3.53		
95 96 97	25.22 26.72 26.48		1.41 1.40 1.41		1.78 1.77 1.79		

98 99	26.56 31.01	1.4		1.78 1.83	
[100 rows	x 13 columns]			
df.describ	e()				
	.5 (ug/m3)	PM10 (ug/m3	3) NO (ug/m	3) NO2 (ug/m3)	NOx
	841.000000	306.00000	90 1844.0000	00 1843.000000	1
1598.00000 mean	30.500435	58.1507	19 6.9529	50 12.557347	,
16.026608 std	20.289850	25.93022	22 5.6110	16 8.707426	i
8.797421 min	0.410000	21.6000	0.0100	00 0.020000)
0.000000 25%	16.540000	36.85000	3.2950	00 6.510000	1
9.860000 50%	27.280000	67.25000	90 5.4600	00 11.280000	1
14.615000 75%	39.530000	69.4150	9.4000	00 16.470000)
20.872500 max	278.970000	371.61000	98.6200	00 134.760000	1
106.740000					
NH3 (ug/m3) \	(ug/m3) S0	2 (ug/m3)	CO (mg/m3)	Ozone (ug/m3)	Benzene
-		30.000000	1687.000000	1833.000000	
mean 1	2.271961	6.551770	0.827825	27.580562	
	5.557635	5.058016	1.472130	16.822431	
	5.360000	0.090000	0.000000	0.100000	
	5.360000	3.890000	0.600000	16.770000	
	4.455000	4.880000	0.740000	24.680000	
	6.675000	7.205000	0.910000	34.770000	
0.365000 max 3	3.680000	37.180000	48.020000	183.990000	
46.230000					
Tol count mean std	uene (ug/m3) 1859.000000 1.940172 4.380171				

0.000000

min

```
25% 0.000000
50% 0.190000
75% 2.755000
max 121.150000
```

df3.mean() #finding the mean after imputation

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

"""Entry point for launching an IPython kernel.

```
PM2.5 (uq/m3)
                   30.500435
PM10 (ug/m3)
                   58.150719
NO (ug/m3)
                    6.952950
N02 (uq/m3)
                   12.557347
NOx (ppb)
                   16.026608
                   12.271961
NH3 (ug/m3)
S02 (uq/m3)
                   6.551770
CO (mq/m3)
                    0.827825
Ozone (ug/m3)
                   27.580562
Benzene (ug/m3)
                   0.583018
Toluene (ug/m3)
                    1.940172
dtype: float64
```

df3.median() #finding the median after imputation

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

"""Entry point for launching an IPython kernel.

```
PM2.5 (ug/m3)
                   27.590000
PM10 (uq/m3)
                   58.150719
NO (ug/m3)
                    5.580000
N02 (ug/m3)
                   11.510000
NOx (ppb)
                   16.026608
NH3 (ug/m3)
                   12.271961
S02 (ug/m3)
                   4.950000
CO (ma/m3)
                    0.770000
                   25.150000
Ozone (ug/m3)
Benzene (ug/m3)
                    0.000000
Toluene (ug/m3)
                    0.240000
dtype: float64
```

df3.std() #finding the standard deviation after imputation

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

"""Entry point for launching an IPython kernel.

PM2.5 (ug/m3)	20.051520
PM10 (ug/m3)	10.433156
NO (ug/m3)	5.549627
NO2 (ug/m3)	8.609821
NOx (ppb)	8.099671
NH3 (ug/m3)	2.236143
S02 (ug/m3)	4.983640
CO (mg/m3)	1.392626
Ozone (ug/m3)	16.588650
Benzene (ug/m3)	2.213479
Toluene (ug/m3)	4.349842
dtype: float64	

PM10 (ug/m3)

df3.cov() #finding the covariance after imputation

-9.804390

u15.cov() #1111u1	ing the covarian	ce areer impac	acion	
	PM2.5 (ug/m3)	PM10 (ug/m3)	NO (ug/m3)	NO2 (ug/m3)
\ PM2.5 (ug/m3)	402.063454	4.790271	15.665399	49.936239
PM10 (ug/m3)	4.790271	108.850747	-7.162718	-3.856727
NO (ug/m3)	15.665399	-7.162718	30.798354	22.017413
NO2 (ug/m3)	49.936239	-3.856727	22.017413	74.129021
NOx (ppb)	33.779861	-9.804390	30.740263	46.359606
NH3 (ug/m3)	2.281140	-2.979065	3.164450	3.570582
S02 (ug/m3)	-0.894573	0.284655	-1.628816	-3.769204
CO (mg/m3)	-1.341713	0.000231	-0.291711	-0.226026
Ozone (ug/m3)	58.637250	5.934287	-0.898932	13.671079
Benzene (ug/m3)	0.607667	0.004987	1.582438	1.318097
Toluene (ug/m3)	10.436053	0.009948	4.516390	9.438699
PM2.5 (ug/m3)		(ug/m3) S02 2.281140 -0		(mg/m3) \ .341713

-2.979065

0.284655

0.000231

NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO (mg/m3) Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3)	30.740263 46.359606 65.604667 6.416965 -2.588604 0.133991 4.819938 1.960402 9.320662	3.570582 -: 6.416965 -: 5.000333 1.378188 24 0.041986 -: 3.464105 0.004861 -:	1.628816 3.769204 2.588604 1.378188 4.836665 9.409550 2.621455 1.062622 3.624998	-0.291711 -0.226026 0.133991 0.041986 -0.409550 1.939407 -0.107126 -0.026996 -0.065842
PM2.5 (ug/m3) PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO (mg/m3) Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3)	Ozone (ug/m3) 58.637250 5.934287 -0.898932 13.671079 4.819938 3.464105 2.621455 -0.107126 275.183304 -0.532692 -0.956679	Benzene (ug/r 0.6070 0.0049 1.5824 1.3180 1.9604 0.0044 -1.0620 -0.0269 -0.5320 4.8994	667 987 438 997 402 861 622 996 692 487	e (ug/m3) 10.436053 0.009948 4.516390 9.438699 9.320662 -0.001140 -3.624998 -0.065842 -0.956679 4.809764 18.921121
df3.corr() #fina				
\ PM2.5 (ug/m3)	PM2.5 (ug/m3) 1.000000	PM10 (ug/m3) 0.022898	NO (ug/m3 0.14077	_
PM2.5 (ug/m3) PM10 (ug/m3)	0.022898	1.000000	-0.12370	
NO (ug/m3)	0.140777	-0.123708	1.00000	
NO2 (ug/m3)	0.289251	-0.042935	0.46079	
NOx (ppb)	0.207990	-0.116021	0.68387	
NH3 (ug/m3)	0.050875	-0.127692		
S02 (ug/m3)	-0.008952	0.005475		
CO (mg/m3)	-0.048048	0.000475		
_				
Ozone (ug/m3)	0.176285	0.034288		
Benzene (ug/m3)	0.013691	0.000216	0.12882	1 0.069164
Toluene (ug/m3)	0.119651	0.000219	0.18709	2 0.252026

```
NOx (ppb)
                             NH3 (ug/m3)
                                           S02 (uq/m3)
                                                         CO (mq/m3)
PM2.5 (uq/m3)
                   0.207990
                                 0.050875
                                              -0.008952
                                                          -0.048048
PM10 (uq/m3)
                  -0.116021
                                -0.127692
                                              0.005475
                                                           0.000016
NO (ug/m3)
                   0.683875
                                 0.254997
                                              -0.058893
                                                          -0.037745
N02 (ug/m3)
                   0.664780
                                 0.185458
                                              -0.087843
                                                          -0.018851
NOx (ppb)
                   1.000000
                                 0.354293
                                              -0.064129
                                                           0.011879
NH3 (ug/m3)
                   0.354293
                                 1.000000
                                              0.123669
                                                           0.013482
S02 (uq/m3)
                  -0.064129
                                 0.123669
                                              1.000000
                                                          -0.059010
CO (mq/m3)
                   0.011879
                                 0.013482
                                              -0.059010
                                                           1.000000
Ozone (ug/m3)
                   0.035873
                                              0.031709
                                                          -0.004637
                                 0.093386
Benzene (ug/m3)
                   0.109346
                                 0.000982
                                              -0.096329
                                                          -0.008758
Toluene (ug/m3)
                   0.264549
                                -0.000117
                                              -0.167220
                                                          -0.010869
                  Ozone (uq/m3)
                                  Benzene (ug/m3)
                                                    Toluene (ug/m3)
PM2.5 (ug/m3)
                       0.176285
                                         0.013691
                                                           0.119651
PM10 (ug/m3)
                       0.034288
                                         0.000216
                                                           0.000219
NO (uq/m3)
                      -0.009765
                                                           0.187092
                                         0.128821
N02 (ug/m3)
                       0.095719
                                         0.069164
                                                           0.252026
NOx (ppb)
                       0.035873
                                         0.109346
                                                           0.264549
NH3 (ug/m3)
                       0.093386
                                         0.000982
                                                          -0.000117
S02 (ug/m3)
                       0.031709
                                        -0.096329
                                                          -0.167220
CO (mg/m3)
                      -0.004637
                                        -0.008758
                                                          -0.010869
Ozone (uq/m3)
                       1.000000
                                        -0.014507
                                                          -0.013258
Benzene (ug/m3)
                      -0.014507
                                         1.000000
                                                           0.499545
                      -0.013258
Toluene (ug/m3)
                                         0.499545
                                                           1.000000
```

This method reduces the variance of the imputed variables.It doesnt preserve the relationship between variables such as correlation. This method is ineffective if the data has outliers since mean gets affected by outliers.

Hockdeck imputation

Hot deck imputation involves replacing missing values of one or more variables for a non-respondent (called the recipient) with observed values from a respondent (the donor) that is similar to the non-respondent with respect to characteristics observed by both cases.

```
df6= pd.read_excel (r'/content/Data 2017-2022.xlsx')
from sklearn.impute import KNNImputer
imputer = KNNImputer(n_neighbors=2, weights="uniform")
df6['PM2.5 (ug/m3)'] = imputer.fit_transform(df6[['PM2.5 (ug/m3)','PM10 (ug/m3)','N0 (ug/m3)','N02 (ug/m3)','N0x (ppb)','NH3 (ug/m3)','S02 (ug/m3)','C0 (mg/m3)','Ozone (ug/m3)','Benzene (ug/m3)','Toluene (ug/m3)']])
print(df6['PM2.5 (ug/m3)'])
df6.head() #after imputation
```

```
2
         24.19
3
         33.61
        129.38
         . . .
1880
         32.05
1881
         38.95
1882
         38.40
1883
         27.51
1884
         34.58
Name: PM2.5 (ug/m3), Length: 1885, dtype: float64
             From Date
                                     To Date PM2.5 (ug/m3)
                                                              PM10
(uq/m3)
0 01-Jan-2017 - 00:00 02-Jan-2017 - 00:00
                                                      32.61
NaN
1 02-Jan-2017 - 00:00
                        03-Jan-2017 - 00:00
                                                      22.93
NaN
2 03-Jan-2017 - 00:00 04-Jan-2017 - 00:00
                                                      24.19
NaN
3 04-Jan-2017 - 00:00
                        05-Jan-2017 - 00:00
                                                      33.61
NaN
4 05-Jan-2017 - 00:00 06-Jan-2017 - 00:00
                                                     129.38
NaN
   NO (ug/m3)
               NO2 (ug/m3) NOx (ppb)
                                                     S02 (ug/m3)
                                        NH3 (ug/m3)
                                                                   C0
(mg/m3)
         2.36
                      9.78
                                                             2.11
                                   NaN
                                                NaN
NaN
1
         2.33
                      8.21
                                   NaN
                                                NaN
                                                             2.86
NaN
2
        11.39
                     17.28
                                   NaN
                                                NaN
                                                             7.73
NaN
         6.06
                     12.32
                                                NaN
                                                             2.72
3
                                   NaN
NaN
         5.58
                     12.67
                                   NaN
                                                NaN
                                                             2.65
4
NaN
   Ozone (ug/m3)
                  Benzene (ug/m3)
                                    Toluene (ug/m3)
0
           24.63
                             0.52
                                               2.95
1
           20.49
                             0.13
                                               2.01
                              0.45
                                               3.52
2
           13.04
3
           19.42
                              0.65
                                               3.98
           25.89
                             0.60
                                               3.53
```

Here it is not very effective because there are still many NaNs as some values in the row may not have similar characteristics to any other for imputing.

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884

```
Data columns (total 13 columns):
                        Non-Null Count
#
     Column
                                         Dtype
     _ _ _ _ _ _
                        _ _ _ _ _ _ _ _ _ _ _ _ _ _ _
 0
     From Date
                        1885 non-null
                                         obiect
 1
     To Date
                        1885 non-null
                                         object
 2
     PM2.5 (ug/m3)
                        1841 non-null
                                         float64
 3
     PM10 (ug/m3)
                                         float64
                        306 non-null
 4
     NO (ug/m3)
                        1844 non-null
                                         float64
     NO2 (ug/m3)
 5
                        1843 non-null
                                         float64
 6
     NOx (ppb)
                        1598 non-null
                                         float64
 7
     NH3 (ug/m3)
                        306 non-null
                                         float64
 8
     S02 (ug/m3)
                        1830 non-null
                                         float64
 9
     CO (mq/m3)
                                         float64
                        1687 non-null
 10
     Ozone (ug/m3)
                        1833 non-null
                                         float64
 11
     Benzene (ug/m3)
                       1859 non-null
                                         float64
 12
     Toluene (ug/m3)
                        1859 non-null
                                         float64
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
df6.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884
Data columns (total 13 columns):
#
                       Non-Null Count
     Column
                                         Dtype
- - -
 0
     From Date
                        1885 non-null
                                         object
 1
     To Date
                        1885 non-null
                                         object
 2
     PM2.5 (ug/m3)
                        1885 non-null
                                         float64
 3
     PM10 (ug/m3)
                        306 non-null
                                         float64
 4
     NO (ug/m3)
                                         float64
                        1844 non-null
 5
     N02 (uq/m3)
                        1843 non-null
                                         float64
 6
     (dqq) xON
                        1598 non-null
                                         float64
 7
     NH3 (ug/m3)
                        306 non-null
                                         float64
 8
     S02 (ug/m3)
                        1830 non-null
                                         float64
 9
     CO (mg/m3)
                        1687 non-null
                                         float64
 10
     Ozone (ug/m3)
                        1833 non-null
                                         float64
 11
     Benzene (ug/m3)
                        1859 non-null
                                         float64
 12
     Toluene (ug/m3)
                        1859 non-null
                                         float64
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
df.describe()
                       PM10 (ug/m3)
                                        NO (ug/m3)
                                                     N02 (ug/m3)
                                                                     N0x
       PM2.5 (ug/m3)
(ppb)
                          306.000000
                                       1844.000000
count
          1841.000000
                                                     1843.000000
1598,000000
            30.500435
                           58.150719
                                          6.952950
                                                       12.557347
mean
16.026608
           20.289850
                           25.930222
                                                        8.707426
                                          5.611016
std
```

8.797421					
min 0.000000	0.410000	21.60000	0.010000	0.020000	
25%	16.540000	36.85000	3.295000	6.510000	
	27.280000	67.25000	5.460000	11.280000	
14.615000 75%	39.530000	69.41500	9.400000	16.470000	
20.872500 max 2 106.740000	78.970000	371.610000	98.620000	134.760000	
	(ug/m3) S02	(ug/m3)	CO (mg/m3) O:	zone (ug/m3)	Benzene
		0.000000	1687.000000	1833.000000	
		6.551770	0.827825	27.580562	
	.557635	5.058016	1.472130	16.822431	
	.360000	0.090000	0.000000	0.100000	
	.360000	3.890000	0.600000	16.770000	
0.000000 50% 14	.455000	4.880000	0.740000	24.680000	
0.000000 75% 16	6.675000	7.205000	0.910000	34.770000	
0.365000 max 33	3.680000 3	7.180000	48.020000	183.990000	
46.230000					
Tolu count mean std min 25% 50% 75% max	ene (ug/m3) 1859.000000 1.940172 4.380171 0.000000 0.000000 0.190000 2.755000 121.150000				
df6.describ	e()				
	5 (ug/m3) P	M10 (ug/m3) NO (ug/m3)	NO2 (ug/m3)	N0x
	85.000000	306.000000	9 1844.000000	1843.000000	
1598.000000 mean	30.467786	58.150719	9 6.952950	12.557347	
16.026608 std	20.103691	25.93022	5.611016	8.707426	

8.797421					
min 0.000000	0.41000	0 21.6000	0.0100	00 0.020	900
25%	16.73000	0 36.8500	00 3.2950	00 6.510	900
9.860000 50%	27.52000	0 67.2500	00 5.4600	00 11.280	900
14.615000 75%	39.25000	0 69.4150	9.4000	00 16.470	900
20.872500 max 106.740000	278.97000)	0 371.6100	98.6200	00 134.760	900
	3 (ug/m3)	S02 (ug/m3)	CO (mg/m3)	Ozone (ug/mi	3) Benzene
	6.000000	1830.000000	1687.000000	1833.0000	90
	12.271961	6.551770	0.827825	27.5805	62
0.583018 std	5.557635	5.058016	1.472130	16.8224	31
2.228912 min	5.360000	0.090000	0.000000	0.1000	90
0.000000 25%	5.360000	3.890000	0.600000	16.7700	90
	14.455000	4.880000	0.740000	24.6800	90
0.000000 75%	16.675000	7.205000	0.910000	34.7700	90
0.365000 max 3 46.230000	33.680000	37.180000	48.020000	183.9900	90
To1	Luono (ua/	m 2			
count	luene (ug/l 1859.000	000			
mean std	1.940 4.380				
min	0.000	900			
25% 50%	0.000 0.190				
75%	2.755	900			
max	121.150	000			
df.cov()					
\	PM	2.5 (ug/m3)	PM10 (ug/m3)	NO (ug/m3)	NO2 (ug/m3)
\ PM2.5 (ug/	/m3)	411.678015	30.890050	16.066220	51.241957
PM10 (ug/n	n3)	30.890050	672.376420	-44.806987	-24.387463
NO (ug/m3))	16.066220	-44.806987	31.483505	22.519439

```
N02 (ug/m3)
                      51.241957
                                    -24.387463
                                                 22.519439
                                                               75.819259
NOx (ppb)
                                    -61.393595
                                                 36.589019
                                                               55.207391
                      40.344852
NH3 (uq/m3)
                      13.933097
                                    -18.579278
                                                 19.546962
                                                               22.055663
S02 (ug/m3)
                      -0.923446
                                      0.531256
                                                 -1.680204
                                                               -3.889626
CO (mq/m3)
                      -1.522403
                                      0.001428
                                                 -0.332261
                                                               -0.253260
Ozone (ug/m3)
                      60.713479
                                     38.184933
                                                 -0.932329
                                                               14.139986
Benzene (ug/m3)
                       0.625041
                                      0.034594
                                                  1.625605
                                                                1.354755
Toluene (ug/m3)
                      10.737620
                                      0.074053
                                                  4.639590
                                                                9.701272
                                           S02 (uq/m3)
                 (dqq) xON
                             NH3 (uq/m3)
                                                         CO (mq/m3)
                                                          -1.522403
PM2.5 (ug/m3)
                  40.344852
                               13.933097
                                             -0.923446
                              -18.579278
PM10 (ug/m3)
                 -61.393595
                                              0.531256
                                                           0.001428
NO (ug/m3)
                  36.589019
                               19.546962
                                             -1.680204
                                                          -0.332261
N02 (ug/m3)
                  55.207391
                               22.055663
                                             -3.889626
                                                          -0.253260
NOx (ppb)
                 77.394611
                               39.637911
                                             -3.141972
                                                           0.158700
NH3 (uq/m3)
                 39.637911
                               30.887305
                                              8.791461
                                                           0.259349
S02 (uq/m3)
                  -3.141972
                                8.791461
                                             25.583531
                                                          -0.472042
                                             -0.472042
CO (mg/m3)
                  0.158700
                                0.259349
                                                           2.167166
Ozone (ug/m3)
                  5.818156
                               23.173855
                                              2.729160
                                                          -0.128536
Benzene (ug/m3)
                                0.043199
                  2.326780
                                             -1.098227
                                                          -0.030333
Toluene (ug/m3)
                 11.059783
                                0.036882
                                             -3.746441
                                                          -0.073975
                  Ozone (ug/m3)
                                 Benzene (ug/m3)
                                                   Toluene (ug/m3)
                                                          10.737620
                                         0.625041
PM2.5 (uq/m3)
                      60.713479
PM10 (uq/m3)
                                         0.034594
                                                           0.074053
                      38.184933
NO (uq/m3)
                      -0.932329
                                         1.625605
                                                           4.639590
                                                           9.701272
N02 (uq/m3)
                      14.139986
                                         1.354755
                                                          11.059783
NOx (ppb)
                       5.818156
                                         2.326780
NH3 (ug/m3)
                      23.173855
                                         0.043199
                                                           0.036882
S02 (uq/m3)
                       2.729160
                                        -1.098227
                                                          -3.746441
CO (mg/m3)
                      -0.128536
                                        -0.030333
                                                          -0.073975
Ozone (uq/m3)
                     282.994183
                                        -0.550818
                                                          -0.989454
Benzene (ug/m3)
                      -0.550818
                                         4.968048
                                                           4.877070
Toluene (ug/m3)
                      -0.989454
                                         4.877070
                                                          19.185895
```

df6.cov() #finding covariance after imputing missing data using
hotdeck

 $PM2.5 \hspace{0.1cm} (ug/m3) \hspace{0.1cm} PM10 \hspace{0.1cm} (ug/m3) \hspace{0.1cm} N0 \hspace{0.1cm} (ug/m3) \hspace{0.1cm} N02 \hspace{0.1cm} (ug/m3)$

PM2.5 (ug/m3)	404.158399	31.362541	15.908389	51.252191
PM10 (ug/m3)	31.362541	672.376420	-44.806987	-24.387463
NO (ug/m3)	15.908389	-44.806987	31.483505	22.519439
NO2 (ug/m3)	51.252191	-24.387463	22.519439	75.819259
NOx (ppb)	40.121315	-61.393595	36.589019	55.207391
NH3 (ug/m3)	13.795534	-18.579278	19.546962	22.055663
S02 (ug/m3)	-1.005195	0.531256	-1.680204	-3.889626
CO (mg/m3)	-1.451634	0.001428	-0.332261	-0.253260
Ozone (ug/m3)	61.966990	38.184933	-0.932329	14.139986
Benzene (ug/m3)	0.623286	0.034594	1.625605	1.354755
Toluene (ug/m3)	10.641059	0.074053	4.639590	9.701272
PM2.5 (ug/m3) PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO (mg/m3) Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3)	40.121315 13 -61.393595 -18 36.589019 19 55.207391 22 77.394611 39 39.637911 30 -3.141972 8 0.158700 0 5.818156 23 2.326780 0	.795534 -1579278 0546962 -1055663 -3637911 -3887305 8791461 25259349 -0173855 2.	.005195 -1. .531256 0. .680204 -0. .889626 -0. .141972 0. .791461 0. .583531 -0. .472042 2. .729160 -0.	mg/m3) \ 451634 001428 332261 253260 158700 259349 472042 167166 128536 030333 073975
PM2.5 (ug/m3) PM10 (ug/m3) NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO (mg/m3) Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3)	Ozone (ug/m3) 61.966990 38.184933 -0.932329 14.139986 5.818156 23.173855 2.729160 -0.128536 282.994183 -0.550818 -0.989454	Benzene (ug/m3 0.62328 0.03459 1.62560 1.35475 2.32678 0.04319 -1.09822 -0.03033 -0.55081 4.96804 4.87707	36 10. 94 0. 95 4. 55 9. 30 11. 99 0. 27 -3. 33 -0. 18 -0.	ug/m3) 641059 074053 639590 701272 059783 036882 746441 073975 989454 877070 185895

df.corr()

ui.coii()				
\	PM2.5 (ug/m3)	PM10 (ug/m3)	NO (ug/m3)	NO2 (ug/m3)
\ PM2.5 (ug/m3)	1.000000	0.077045	0.140907	0.290013
PM10 (ug/m3)	0.077045	1.000000	-0.342711	-0.205761
NO (ug/m3)	0.140907	-0.342711	1.000000	0.460980
NO2 (ug/m3)	0.290013	-0.205761	0.460980	1.000000
NOx (ppb)	0.217120	-0.272646	0.714484	0.724070
NH3 (ug/m3)	0.161947	-0.128751	0.702132	0.871728
S02 (ug/m3)	-0.009008	0.004149	-0.059458	-0.089198
CO (mg/m3)	-0.049465	0.000288	-0.039316	-0.019531
Ozone (ug/m3)	0.178225	0.058454	-0.009927	0.097121
Benzene (ug/m3)	0.013715	0.011016	0.129351	0.069390
Toluene (ug/m3)	0.120031	0.022940	0.188078	0.253154
PM2.5 (ug/m3) PM10 (ug/m3) N0 (ug/m3) N02 (ug/m3) N0x (ppb) NH3 (ug/m3) S02 (ug/m3) C0 (mg/m3) Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3)	0.217120 -0.272646 -0.714484 0.724070 1.000000 0.825683 -0.068077 0.011973 0.038354 0.122707	0.161947 -0 0.128751 0 0.702132 -0 0.871728 -0 0.825683 -0 1.000000 0 0.323777 1 0.242040 -0 0.170788 0 0.064346 -0	.009008 -0 .004149 0 .059458 -0 .089198 -0 .068077 0 .323777 0 .000000 -0 .061407 1 .032032 -0 .096417 -0	(mg/m3) \ .049465 .000288 .039316 .019531 .011973 .242040 .061407 .000000 .008149 .008974 .010942
PM2.5 (ug/m3) PM10 (ug/m3) N0 (ug/m3) N02 (ug/m3) N0x (ppb) NH3 (ug/m3) S02 (ug/m3) C0 (mg/m3)	Ozone (ug/m3) 0.178225 0.058454 -0.009927 0.097121 0.038354 0.170788 0.032032 -0.008149	Benzene (ug/m 0.0137 0.0110 0.1293 0.0693 0.1227 0.0643 -0.0964 -0.0089	15 0 16 0 51 0 90 0 07 0 46 0 17 -0	(ug/m3) .120031 .022940 .188078 .253154 .278826 .053442 .167630 .010942

Ozone (ug/m3)	1.000000	-0.014557	-0.013333
Benzene (ug/m3)	-0.014557	1.000000	0.499545
Toluene (ug/m3)	-0.013333	0.499545	1.000000

df6.corr() #finding correlation after imputing missing data using hotdeck

	PM2.5 (ug/m3)	PM10 (ug/m3)	NO (ug/m3)	NO2 (ug/m3)
\ PM2.5 (ug/m3)	1.000000	0.078255	0.139757	0.290068
PM10 (ug/m3)	0.078255	1.000000	-0.342711	-0.205761
NO (ug/m3)	0.139757	-0.342711	1.000000	0.460980
NO2 (ug/m3)	0.290068	-0.205761	0.460980	1.000000
NOx (ppb)	0.217071	-0.272646	0.714484	0.724070
NH3 (ug/m3)	0.160586	-0.128751	0.702132	0.871728
S02 (ug/m3)	-0.009821	0.004149	-0.059458	-0.089198
CO (mg/m3)	-0.047706	0.000288	-0.039316	-0.019531
Ozone (ug/m3)	0.181929	0.058454	-0.009927	0.097121
Benzene (ug/m3)	0.013835	0.011016	0.129351	0.069390
Toluene (ug/m3)	0.120195	0.022940	0.188078	0.253154
PM2.5 (ug/m3) PM10 (ug/m3) N0 (ug/m3) N02 (ug/m3) N0x (ppb) NH3 (ug/m3) S02 (ug/m3) C0 (mg/m3) Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3)	0.217071 -0.272646 -0.714484 0.724070 1.000000 0.825683 -0.068077 0.011973 0.038354 0.122707	0.160586 -0 0.128751 0 0.702132 -0 0.871728 -0 0.825683 -0 1.000000 0 0.323777 1 0.242040 -0 0.170788 0 0.064346 -0	0.009821 -6 0.004149 6 0.059458 -6 0.089198 -6 0.068077 6 0.323777 6 0.000000 -6 0.061407 1 0.032032 -6 0.096417 -6	(mg/m3) \ 0.047706 0.000288 0.039316 0.019531 0.011973 0.242040 0.061407 1.000000 0.008149 0.008974 0.010942
PM2.5 (ug/m3) PM10 (ug/m3) NO (ug/m3)	Ozone (ug/m3) 0.181929 0.058454 -0.009927	Benzene (ug/m 0.0138 0.0110 0.1293	335 6 916 6	(ug/m3)).120195).022940).188078

NO2 (ug/m3)	0.097121	0.069390	0.253154
NOx (ppb)	0.038354	0.122707	0.278826
NH3 (ug/m3)	0.170788	0.064346	0.053442
S02 (ug/m3)	0.032032	-0.096417	-0.167630
CO (mg/m3)	-0.008149	-0.008974	-0.010942
Ozone (ug/m3)	1.000000	-0.014557	-0.013333
Benzene (ug/m3)	-0.014557	1.000000	0.499545
Toluene (ug/m3)	-0.013333	0.499545	1.000000

Filling with before observation

This method would fill the missing values with first non-missing value that occurs before it. This will be carried forward until another non-null value is encountered.

df.head(50) #before the impuation is done

		rom Date		To	Date	PM2.5	(ug/m3)	PM10
	/m3) \	00.00	02 lan 2017		00.00		22 61	
0 NaN	01-Jan-201/	- 00:00	02-Jan-2017	-	00:00		32.61	
1 NaN	02-Jan-2017	- 00:00	03-Jan-2017	-	00:00		22.93	
2 NaN	03-Jan-2017	- 00:00	04-Jan-2017	-	00:00		24.19	
3	04-Jan-2017	- 00:00	05-Jan-2017	-	00:00		33.61	
NaN 4	05-Jan-2017	- 00:00	06-Jan-2017	-	00:00		129.38	
NaN 5	06-Jan-2017	- 00:00	07-Jan-2017	-	00:00		64.52	
NaN 6	07-Jan-2017	- 00:00	08-Jan-2017	-	00:00		45.01	
NaN 7	08-Jan-2017	- 00:00	09-Jan-2017	-	00:00		32.71	
NaN 8	09-Jan-2017	- 00:00	10-Jan-2017	-	00:00		33.66	
NaN 9	10-Jan-2017	- 00:00	11-Jan-2017	-	00:00		34.82	
NaN 10	11-Jan-2017	- 00:00	12-Jan-2017	-	00:00		38.33	
NaN 11	12-Jan-2017	- 00:00	13-Jan-2017	-	00:00		42.32	
NaN 12	13-Jan-2017	- 00:00	14-Jan-2017	-	00:00		103.51	
NaN 13	14-Jan-2017	- 00:00	15-Jan-2017	-	00:00		33.96	
NaN 14	15-Jan-2017	- 00:00	16-Jan-2017	-	00:00		54.93	
NaN 15	16-Jan-2017	- 00:00	17-Jan-2017	-	00:00		46.01	

NaN							
16	17-Jan-2017	-	00:00	18-Jan-2017	-	00:00	47.76
NaN 17	18-Jan-2017		00.00	19-Jan-2017		00.00	36.59
NaN	10-Jan-2017	-	00.00	19-3411-2017	-	00.00	30.39
18 NaN	19-Jan-2017	-	00:00	20-Jan-2017	-	00:00	42.82
NaN 19	20-Jan-2017	-	00:00	21-Jan-2017	-	00:00	68.26
NaN 20	21-Jan-2017	-	00:00	22-Jan-2017	-	00:00	38.96
NaN 21	22-Jan-2017	_	00:00	23-Jan-2017	_	00:00	26.28
NaN 22	23-Jan-2017	_	00:00	24-Jan-2017	_	00:00	33.59
NaN							
23 NaN	24-Jan-2017	-	00:00	25-Jan-2017	-	00:00	29.59
24	25-Jan-2017	-	00:00	26-Jan-2017	-	00:00	40.49
NaN 25	26-Jan-2017		00.00	27-Jan-2017		00.00	29.95
NaN	20-Jan-2017	-	00:00	27-Jan-2017	-	00.00	29.93
26	27-Jan-2017	-	00:00	28-Jan-2017	-	00:00	18.93
NaN 27	28-Jan-2017	_	00:00	29-Jan-2017	_	00:00	7.64
NaN	20 3411 2017		00100	23 3411 2017		00.00	7101
28	29-Jan-2017	-	00:00	30-Jan-2017	-	00:00	8.65
NaN 29	30-Jan-2017	_	00:00	31-Jan-2017	_	00:00	19.69
NaN							
30 NaN	31-Jan-2017	-	00:00	01-Feb-2017	-	00:00	35.80
31	01-Feb-2017	_	00:00	02-Feb-2017	_	00:00	37.48
NaN	02 5 1 2017		00.00	02 5 1 2017		00.00	24.04
32 NaN	02-Feb-201/	-	00:00	03-Feb-2017	-	00:00	34.84
33	03-Feb-2017	-	00:00	04-Feb-2017	-	00:00	36.89
NaN	04 Feb 2017		00.00	05 Fab 2017		00.00	40.02
34 NaN	04-Feb-2017	-	00:00	05-Feb-2017	-	00:00	40.92
35 NaN	05-Feb-2017	-	00:00	06-Feb-2017	-	00:00	43.07
36	06-Feb-2017	-	00:00	07-Feb-2017	-	00:00	42.05
NaN 37	07-Feb-2017	-	00:00	08-Feb-2017	-	00:00	47.71
NaN 38	08-Feb-2017	-	00:00	09-Feb-2017	-	00:00	61.60
NaN 39	09-Feb-2017	_	00:00	10-Feb-2017	_	00:00	54.49
NaN							
40	10-Feb-2017	-	00:00	11-Feb-2017	-	00:00	50.29

NaN 41	11-Feb-2017	- 00:00	12-Feb-	-2017 -	00:00		44.20	
NaN 42 NaN	12-Feb-2017	- 00:00	13-Feb-	-2017 -	00:00		26.51	
43 NaN	13-Feb-2017	- 00:00	14-Feb-	-2017 -	00:00		17.22	
44 NaN	14-Feb-2017	- 00:00	15-Feb-	-2017 -	00:00		22.04	
45 NaN	15-Feb-2017	- 00:00	16-Feb-	-2017 -	00:00		30.00	
46 NaN	16-Feb-2017	- 00:00	17-Feb-	-2017 -	00:00		36.92	
47 NaN	17-Feb-2017	- 00:00	18-Feb-	-2017 -	00:00		38.03	
48	18-Feb-2017	- 00:00	19-Feb-	-2017 -	00:00		29.63	
NaN 49 NaN	19-Feb-2017	- 00:00	20-Feb-	-2017 -	00:00		34.19	
/ mm m	_	N02 (ug/r	m3) NO>	(ppb)	NH3 (ug	/m3)	S02 (ug/m3)	CO
0	′m3) \ 2.36	9	. 78	NaN		NaN	2.11	
NaN 1	2.33	8	.21	NaN		NaN	2.86	
NaN 2 NaN	11.39	17	. 28	NaN		NaN	7.73	
3 NaN	6.06	12	.32	NaN		NaN	2.72	
4 NaN	5.58	12	. 67	NaN		NaN	2.65	
5 NaN	6.91	14	. 38	NaN		NaN	2.28	
6 NaN	5.72	9	. 66	NaN		NaN	2.19	
7 NaN	4.33	10	. 57	NaN		NaN	2.23	
8 NaN	6.02	15	.74	NaN		NaN	3.76	
9 NaN	6.96	16	. 31	NaN		NaN	2.71	
10 NaN	6.57	18	. 45	NaN		NaN	2.75	
11 NaN	22.63		.31	NaN		NaN	2.57	
12 NaN	32.73		. 26	NaN		NaN	2.65	
13 NaN	24.83	27	. 27	NaN		NaN	2.48	

14 NaN	16.29	20.86	NaN	NaN	3.54
NaN 15 NaN	4.18	14.54	NaN	NaN	3.23
16 NaN	4.08	20.86	NaN	NaN	4.49
17 NaN	4.76	20.39	NaN	NaN	3.52
18 NaN	2.71	15.88	NaN	NaN	2.55
19 NaN	2.86	18.64	NaN	NaN	4.17
20 NaN	2.37	15.49	NaN	NaN	2.53
21 NaN	2.47	15.81	NaN	NaN	2.99
22 NaN	3.12	15.40	NaN	NaN	4.70
23 NaN	2.83	14.93	NaN	NaN	3.77
24 NaN	2.37	12.49	NaN	NaN	2.63
25 NaN	2.44	12.68	NaN	NaN	2.60
26 NaN	2.53	12.44	NaN	NaN	2.46
27 NaN	2.59	11.63	NaN	NaN	2.35
28 NaN	6.07	13.59	NaN	NaN	2.90
29 NaN 30	6.07 6.47	18.00 17.91	NaN NaN	NaN NaN	2.77 3.97
NaN 31	3.63	16.19	NaN	NaN	2.27
1.91 32	6.89	18.19	NaN	NaN	2.41
1.01 33	6.19	18.27	NaN	NaN	3.09
NaN 34	6.62	17.96	NaN	NaN	2.64
NaN 35	5.41	18.31	NaN	NaN	2.90
NaN 36	6.82	21.76	NaN	NaN	2.83
NaN 37	9.01	25.18	NaN	NaN	4.03
NaN 38 NaN	9.12	23.48	NaN	NaN	2.85

39 NaN	5.63	22.22	NaN	NaN	2.97
40	6.47	21.84	NaN	NaN	2.74
NaN 41	10.70	22.33	NaN	NaN	2.32
NaN 42	5.93	17.56	NaN	NaN	2.20
NaN 43	5.46	15.74	NaN	NaN	2.54
NaN 44	2.54	12.35	NaN	NaN	2.17
NaN 45	3.77	14.66	NaN	NaN	2.57
NaN 46	3.26	14.88	NaN	NaN	2.20
NaN 47	5.55	18.64	NaN	NaN	2.10
1.21 48	5.58	17.27	NaN	NaN	1.99
1.21 49	5.52	18.13	NaN	NaN	3.60
NaN					
0zone 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25	e (ug/m3) 24.63 20.49 13.04 19.42 25.89 25.16 35.32 31.22 24.32 21.71 24.67 30.78 30.94 44.53 45.14 50.70 52.13 50.51 63.89 62.39 55.96 62.37 54.17 65.30 71.66 51.52	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	m3) Toluene .52 .13 .45 .65 .60 .74 .36 .34 .51 .55 .74 .60 .23 .00 .44 .55 .88 .00 .26 .00 .26 .00 .00 .42 .31 .00 .00	e (ug/m3) 2.95 2.01 3.52 3.98 3.53 3.80 3.20 1.94 3.27 3.88 3.79 3.52 4.35 0.77 2.82 1.41 2.52 2.55 1.10 2.05 1.59 2.37 2.69 2.15 0.84 0.77	

```
26
             24.48
                                0.00
                                                   1.33
27
             16.77
                                0.00
                                                   1.06
                                                   1.67
28
             15.22
                                0.21
29
             21.93
                                0.41
                                                   2.19
                                 1.05
             23.68
30
                                                   3.39
             33.67
31
                                0.06
                                                   2.74
             27.56
32
                                0.51
                                                   2.93
33
             33.36
                                0.54
                                                   3.29
34
             38.83
                                0.47
                                                   2.02
35
             32.93
                                0.65
                                                   0.78
36
             32.11
                                 0.84
                                                   0.54
37
             35.07
                                 1.32
                                                   1.34
38
             40.37
                                 1.50
                                                   2.12
             39.35
39
                                 1.32
                                                   1.89
40
             32.65
                                 1.34
                                                   1.90
41
             29.16
                                 1.34
                                                   1.90
42
             34.39
                                 1.41
                                                   1.96
             34.28
43
                                 1.90
                                                   2.33
             52.83
                                                   2.59
44
                                2.23
             49.97
                                                   2.37
45
                                1.94
             67.63
46
                                1.67
                                                   2.16
47
             47.98
                                 1.34
                                                   1.90
48
             39.44
                                 1.34
                                                   1.90
49
             32.65
                                 1.39
                                                   1.95
df7 = pd.read_excel (r'/content/Data 2017-2022.xlsx')
df7.ffill(inplace=True)
df7.head(50) #after imputation
               From Date
                                        To Date PM2.5 (ug/m3)
                                                                  PM10
(ug/m3) \setminus
    01-Jan-2017 - 00:00 02-Jan-2017 - 00:00
                                                           32.61
NaN
1
    02-Jan-2017 - 00:00
                           03-Jan-2017 - 00:00
                                                           22.93
NaN
2
    03-Jan-2017 - 00:00
                           04-Jan-2017 - 00:00
                                                           24.19
NaN
3
    04-Jan-2017 - 00:00
                           05-Jan-2017 - 00:00
                                                           33.61
NaN
    05-Jan-2017 - 00:00
                           06-Jan-2017 - 00:00
4
                                                          129.38
NaN
5
    06-Jan-2017 - 00:00
                           07-Jan-2017 - 00:00
                                                           64.52
NaN
    07-Jan-2017 - 00:00
                           08-Jan-2017 - 00:00
                                                           45.01
6
NaN
7
    08-Jan-2017 - 00:00
                           09-Jan-2017 - 00:00
                                                           32.71
NaN
    09-Jan-2017 - 00:00
                          10-Jan-2017 - 00:00
8
                                                           33.66
NaN
```

10-Jan-2017 - 00:00 11-Jan-2017 - 00:00

34.82

9

NaN

10 NaN	11-Jan-2017	-	00:00	12-Jan-2017	-	00:00	38.33
11 NaN	12-Jan-2017	-	00:00	13-Jan-2017	-	00:00	42.32
12 NaN	13-Jan-2017	-	00:00	14-Jan-2017	-	00:00	103.51
13 NaN	14-Jan-2017	-	00:00	15-Jan-2017	-	00:00	33.96
14 NaN	15-Jan-2017	-	00:00	16-Jan-2017	-	00:00	54.93
15 NaN	16-Jan-2017	-	00:00	17-Jan-2017	-	00:00	46.01
16 NaN	17-Jan-2017	-	00:00	18-Jan-2017	-	00:00	47.76
17 NaN	18-Jan-2017	-	00:00	19-Jan-2017	-	00:00	36.59
18 NaN	19-Jan-2017	-	00:00	20-Jan-2017	-	00:00	42.82
19 NaN	20-Jan-2017	-	00:00	21-Jan-2017	-	00:00	68.26
20 NaN	21-Jan-2017	-	00:00	22-Jan-2017			38.96
21 NaN	22-Jan-2017						26.28
22 NaN	23-Jan-2017						33.59
23 NaN	24-Jan-2017						29.59
24 NaN	25-Jan-2017			26-Jan-2017			40.49
25 NaN	26-Jan-2017						29.95
26 NaN				28-Jan-2017			18.93
27 NaN	28-Jan-2017						7.64
28 NaN 29				30-Jan-2017			8.65 19.69
NaN 30				31-Jan-2017 01-Feb-2017			35.80
NaN 31				02-Feb-2017			37.48
NaN 32				03-Feb-2017			34.84
NaN 33				04-Feb-2017			36.89
NaN 34				05-Feb-2017			40.92
NaN							

35 NaN	05-Feb-2017	- 00:00	06-Feb-20	917 -	00:00	43.07	
36	06-Feb-2017	- 00:00	07-Feb-20	917 -	00:00	42.05	
NaN 37	07-Feb-2017	- 00:00	08-Feb-20	917 -	00:00	47.71	
NaN 38	08-Feb-2017	- 00:00	09-Feb-20	917 -	00:00	61.60	
NaN 39	09-Feb-2017	- 00:00	10-Feb-20	917 -	00:00	54.49	
NaN 40	10-Feb-2017	- 00:00	11-Feb-20	917 -	00:00	50.29	
NaN 41	11-Feb-2017	- 00:00	12-Feb-20	917 -	00:00	44.20	
NaN 42	12-Feb-2017	- 00:00	13-Feb-20	917 -	00:00	26.51	
NaN 43	13-Feb-2017	- 00:00	14-Feb-20	917 -	00:00	17.22	
NaN 44	14-Feb-2017	- 00:00	15-Feb-20	917 -	00:00	22.04	
NaN 45	15-Feb-2017	- 00:00	16-Feb-20	917 -	00:00	30.00	
NaN 46	16-Feb-2017	- 00:00	17-Feb-20	917 -	00:00	36.92	
NaN 47	17-Feb-2017	- 00:00	18-Feb-20	917 -	00:00	38.03	
NaN 48	18-Feb-2017	- 00:00	19-Feb-20	917 -	00:00	29.63	
NaN 49	19-Feb-2017	- 00:00	20-Feb-20	917 -	00:00	34.19	
NaN							
(mg/		N02 (ug/n	m3) NOx	(ppb)	NH3 (ug/m3)	S02 (ug/m3)	CO
0 NaN	2.36	9	. 78	NaN	NaN	2.11	
1 NaN	2.33	8	.21	NaN	NaN	2.86	
2 NaN	11.39	17	. 28	NaN	NaN	7.73	
3 NaN	6.06	12	.32	NaN	NaN	2.72	
4 NaN	5.58	12	. 67	NaN	NaN	2.65	
5	6.91	14	. 38	NaN	NaN	2.28	
NaN 6	5.72	9	. 66	NaN	NaN	2.19	
NaN 7	4.33	10	. 57	NaN	NaN	2.23	
NaN 8	6.02	15	.74	NaN	NaN	3.76	

NaN 9	6.96	16.31	NaN	NaN	2.71
NaN 10	6.57	18.45	NaN	NaN	2.75
NaN 11	22.63	26.31	NaN	NaN	2.57
NaN 12					
NaN	32.73	38.26	NaN	NaN	2.65
13 NaN	24.83	27.27	NaN	NaN	2.48
14 NaN	16.29	20.86	NaN	NaN	3.54
15	4.18	14.54	NaN	NaN	3.23
NaN 16	4.08	20.86	NaN	NaN	4.49
NaN 17	4.76	20.39	NaN	NaN	3.52
NaN 18	2.71	15.88	NaN	NaN	2.55
NaN 19	2.86	18.64	NaN	NaN	4.17
NaN					
20 NaN	2.37	15.49	NaN	NaN	2.53
21	2.47	15.81	NaN	NaN	2.99
NaN 22	3.12	15.40	NaN	NaN	4.70
NaN 23	2.83	14.93	NaN	NaN	3.77
NaN 24	2.37	12.49	NaN	NaN	2.63
NaN 25					
NaN	2.44	12.68	NaN	NaN	2.60
26 NaN	2.53	12.44	NaN	NaN	2.46
27	2.59	11.63	NaN	NaN	2.35
NaN 28	6.07	13.59	NaN	NaN	2.90
NaN 29	6.07	18.00	NaN	NaN	2.77
NaN 30	6.47	17.91	NaN	NaN	3.97
NaN 31					
1.91	3.63	16.19	NaN	NaN	2.27
32 1.01	6.89	18.19	NaN	NaN	2.41
33	6.19	18.27	NaN	NaN	3.09

	1.01 34		6.62	17.96		NaN	NaN	2.64
	1.01 35	L	5.41	18.31		NaN	NaN	2.90
	1.01	L						
	36 1.01		6.82	21.76		NaN	NaN	2.83
	37 1.01		9.01	25.18		NaN	NaN	4.03
	38	9	9.12	23.48		NaN	NaN	2.85
	1.01 39	!	5.63	22.22		NaN	NaN	2.97
	1.01 40		6.47	21.84		NaN	NaN	2.74
	1.01 41		0.70	22.33		NaN	NaN	2.32
	1.01 42	L	5.93	17.56		NaN	NaN	2.20
	1.01 43	L	5.46	15.74		NaN	NaN	2.54
	1.01	L						
	44 1.01		2.54	12.35		NaN	NaN	2.17
	45 1.01	:	3.77	14.66		NaN	NaN	2.57
	46	:	3.26	14.88		NaN	NaN	2.20
,	1.01 47	!	5.55	18.64		NaN	NaN	2.10
	1.21 48	!	5.58	17.27		NaN	NaN	1.99
	1.21 49		5.52	18.13		NaN	NaN	3.60
	1.21	L						
	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14	Ozone	(ug/m3) 24.63 20.49 13.04 19.42 25.89 25.16 35.32 31.22 24.32 21.71 24.67 30.78 30.94 44.53 45.14	Benzene	(ug/m3) 0.52 0.13 0.45 0.65 0.60 0.74 0.36 0.51 0.55 0.74 0.74 6.60 0.23 1.00	Toluene	(ug/m3) 2.95 2.01 3.52 3.98 3.53 3.80 3.20 1.94 3.27 3.88 3.79 3.52 4.35 0.77 2.82	

This method is not very effective due to the large number of persisting NaN values even after the imputation, this is due to the reason that there are a lot of NaN values in a column before the first non-null value, which cannot be replace with any numerical.

```
To Date
                       1885 non-null
 1
                                         object
 2
     PM2.5 (ug/m3)
                       1841 non-null
                                         float64
 3
     PM10 (ug/m3)
                       306 non-null
                                         float64
 4
     NO (ug/m3)
                       1844 non-null
                                         float64
 5
     N02 (ug/m3)
                       1843 non-null
                                         float64
 6
     NOx (ppb)
                       1598 non-null
                                         float64
 7
     NH3 (uq/m3)
                       306 non-null
                                         float64
 8
     S02 (uq/m3)
                       1830 non-null
                                         float64
 9
     CO (mg/m3)
                       1687 non-null
                                         float64
 10
     Ozone (ug/m3)
                       1833 non-null
                                         float64
 11
     Benzene (ug/m3)
                       1859 non-null
                                         float64
 12
     Toluene (ug/m3)
                       1859 non-null
                                         float64
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884
Data columns (total 13 columns):
#
     Column
                       Non-Null Count
                                         Dtype
- - -
     _ _ _ _ _ _
 0
     From Date
                       1885 non-null
                                         object
 1
     To Date
                       1885 non-null
                                         object
 2
     PM2.5 (ug/m3)
                       1885 non-null
                                         float64
 3
     PM10 (uq/m3)
                       309 non-null
                                         float64
 4
     NO (ug/m3)
                       1885 non-null
                                         float64
 5
     N02 (uq/m3)
                       1885 non-null
                                         float64
 6
     NOx (ppb)
                       1609 non-null
                                         float64
 7
     NH3 (ug/m3)
                       309 non-null
                                         float64
 8
     S02 (ug/m3)
                       1885 non-null
                                         float64
 9
     CO (mg/m3)
                       1854 non-null
                                         float64
 10
     Ozone (ug/m3)
                       1885 non-null
                                         float64
 11
     Benzene (ug/m3)
                       1885 non-null
                                         float64
     Toluene (ug/m3)
 12
                       1885 non-null
                                         float64
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
```

Filling with next observation

This method would fill the missing values with first non-missing value that occurs after it. The imputation happens backwards and is continued until a nn value is encountered.

df.head(50) #dataset before handling the missing data

From Date (ug/m3) \	To Date	PM2.5 (ug/m3)	PM10
0 01-Jan-2017 - 00:00	02-Jan-2017 - 00:00	32.61	
NaN 1 02-Jan-2017 - 00:00	03-Jan-2017 - 00:00	22.93	
NaN 2 03-Jan-2017 - 00:00	04-Jan-2017 - 00:00	24.19	

NaN							
3 NaN	04-Jan-2017	-	00:00	05-Jan-2017	-	00:00	33.61
4	05-Jan-2017	-	00:00	06-Jan-2017	-	00:00	129.38
NaN 5	06-Jan-2017	-	00:00	07-Jan-2017	-	00:00	64.52
NaN 6	07-Jan-2017	-	00:00	08-Jan-2017	-	00:00	45.01
NaN 7	08-Jan-2017	-	00:00	09-Jan-2017	-	00:00	32.71
NaN 8	09-Jan-2017	_	00:00	10-Jan-2017	_	00:00	33.66
NaN 9	10-Jan-2017	_	00:00	11-Jan-2017	_	00:00	34.82
NaN 10	11-Jan-2017	_	00:00	12-Jan-2017	_	00:00	38.33
NaN 11	12-Jan-2017				_	00:00	42.32
NaN							
12 NaN	13-Jan-2017	-	00:00	14-Jan-2017	-	00:00	103.51
13	14-Jan-2017	-	00:00	15-Jan-2017	-	00:00	33.96
NaN 14	15-Jan-2017	-	00:00	16-Jan-2017	-	00:00	54.93
NaN 15	16-Jan-2017	-	00:00	17-Jan-2017	-	00:00	46.01
NaN 16	17-Jan-2017	-	00:00	18-Jan-2017	_	00:00	47.76
NaN 17	18-Jan-2017	_	00:00	19-Jan-2017	_	00:00	36.59
NaN 18	19-Jan-2017		00.00	20-Jan-2017		00.00	42.82
NaN	19-Jan-2017	-	00:00	20-Jan-2017	-	00:00	42.02
19 NaN	20-Jan-2017	-	00:00	21-Jan-2017	-	00:00	68.26
20	21-Jan-2017	-	00:00	22-Jan-2017	-	00:00	38.96
NaN 21	22-Jan-2017	-	00:00	23-Jan-2017	-	00:00	26.28
NaN 22	23-Jan-2017	-	00:00	24-Jan-2017	_	00:00	33.59
NaN 23	24-Jan-2017	_	00:00	25-Jan-2017	_	00:00	29.59
NaN 24	25-Jan-2017	_	00 · 00	26-Jan-2017	_	00.00	40.49
NaN							
25 NaN	26-Jan-2017	-	00:00	27-Jan-2017	-	00:00	29.95
26 NaN	27-Jan-2017	-	00:00	28-Jan-2017	-	00:00	18.93
27	28-Jan-2017	-	00:00	29-Jan-2017	-	00:00	7.64

```
NaN
   29-Jan-2017 - 00:00 30-Jan-2017 - 00:00
                                                        8.65
28
NaN
29
    30-Jan-2017 - 00:00
                         31-Jan-2017 - 00:00
                                                       19.69
NaN
30
    31-Jan-2017 - 00:00
                         01-Feb-2017 - 00:00
                                                       35.80
NaN
31 01-Feb-2017 - 00:00
                         02-Feb-2017 - 00:00
                                                       37.48
NaN
32
   02-Feb-2017 - 00:00
                         03-Feb-2017 - 00:00
                                                       34.84
NaN
   03-Feb-2017 - 00:00
                         04-Feb-2017 - 00:00
33
                                                       36.89
NaN
34
   04-Feb-2017 - 00:00
                         05-Feb-2017 - 00:00
                                                       40.92
NaN
35
   05-Feb-2017 - 00:00
                         06-Feb-2017 - 00:00
                                                       43.07
NaN
36 06-Feb-2017 - 00:00
                         07-Feb-2017 - 00:00
                                                       42.05
NaN
37 07-Feb-2017 - 00:00
                         08-Feb-2017 - 00:00
                                                       47.71
NaN
                         09-Feb-2017 - 00:00
38
   08-Feb-2017 - 00:00
                                                       61.60
NaN
39
    09-Feb-2017 - 00:00
                         10-Feb-2017 - 00:00
                                                       54.49
NaN
40
   10-Feb-2017 - 00:00
                         11-Feb-2017 - 00:00
                                                       50.29
NaN
41
    11-Feb-2017 - 00:00
                         12-Feb-2017 - 00:00
                                                       44.20
NaN
42
   12-Feb-2017 - 00:00
                         13-Feb-2017 - 00:00
                                                       26.51
NaN
43
   13-Feb-2017 - 00:00
                         14-Feb-2017 - 00:00
                                                       17.22
NaN
    14-Feb-2017 - 00:00
44
                         15-Feb-2017 - 00:00
                                                       22.04
NaN
   15-Feb-2017 - 00:00
                         16-Feb-2017 - 00:00
                                                       30.00
45
NaN
    16-Feb-2017 - 00:00
                         17-Feb-2017 - 00:00
46
                                                       36.92
NaN
47
   17-Feb-2017 - 00:00
                         18-Feb-2017 - 00:00
                                                       38.03
NaN
48
    18-Feb-2017 - 00:00
                         19-Feb-2017 - 00:00
                                                       29.63
NaN
49
   19-Feb-2017 - 00:00
                         20-Feb-2017 - 00:00
                                                       34.19
NaN
    NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3)
                                                      S02 (uq/m3)
                                                                   C0
(mg/m3) \setminus
          2.36
                       9.78
                                                             2.11
0
                                    NaN
                                                 NaN
NaN
```

1 NaN	2.33	8.21	NaN	NaN	2.86
NaN 2 NaN	11.39	17.28	NaN	NaN	7.73
NaN NaN	6.06	12.32	NaN	NaN	2.72
4 NaN	5.58	12.67	NaN	NaN	2.65
5 NaN	6.91	14.38	NaN	NaN	2.28
6 NaN	5.72	9.66	NaN	NaN	2.19
7 NaN	4.33	10.57	NaN	NaN	2.23
8 NaN	6.02	15.74	NaN	NaN	3.76
9 NaN	6.96	16.31	NaN	NaN	2.71
10 NaN	6.57	18.45	NaN	NaN	2.75
11 NaN 12	22.63 32.73	26.31 38.26	NaN NaN	NaN NaN	2.57
NaN 13	24.83	27.27	NaN	NaN	2.48
NaN 14	16.29	20.86	NaN	NaN	3.54
NaN 15	4.18	14.54	NaN	NaN	3.23
NaN 16	4.08	20.86	NaN	NaN	4.49
NaN 17	4.76	20.39	NaN	NaN	3.52
NaN 18	2.71	15.88	NaN	NaN	2.55
NaN 19 NaN	2.86	18.64	NaN	NaN	4.17
20 NaN	2.37	15.49	NaN	NaN	2.53
21 NaN	2.47	15.81	NaN	NaN	2.99
22 NaN	3.12	15.40	NaN	NaN	4.70
23 NaN	2.83	14.93	NaN	NaN	3.77
24 NaN	2.37	12.49	NaN	NaN	2.63
25 NaN	2.44	12.68	NaN	NaN	2.60

26 NaN	2.53	12.44	NaN	NaN	2.46
NaN 27	2.59	11.63	NaN	NaN	2.35
NaN 28	6.07	13.59	NaN	NaN	2.90
NaN 29	6.07	18.00	NaN	NaN	2.77
NaN 30	6.47	17.91	NaN	NaN	3.97
NaN 31	3.63	16.19	NaN	NaN	2.27
1.91	6.89	18.19	NaN	NaN	2.41
1.01 33	6.19	18.27	NaN	NaN	3.09
NaN 34	6.62	17.96	NaN	NaN	2.64
NaN 35	5.41	18.31	NaN	NaN	2.90
NaN 36	6.82	21.76	NaN	NaN	2.83
NaN 37	9.01	25.18	NaN	NaN	4.03
NaN 38	9.12	23.48	NaN	NaN	2.85
NaN 39	5.63	22.22	NaN	NaN	2.97
NaN 40	6.47	21.84	NaN	NaN	2.74
NaN 41	10.70	22.33	NaN	NaN	2.32
NaN 42	5.93	17.56	NaN	NaN	2.20
NaN 43	5.46	15.74	NaN	NaN	2.54
NaN 44	2.54	12.35	NaN	NaN	2.17
NaN 45	3.77	14.66	NaN	NaN	2.57
NaN 46	3.26	14.88	NaN	NaN	2.20
NaN 47	5.55	18.64	NaN	NaN	2.10
1.21 48	5.58	17.27	NaN	NaN	1.99
1.21 49 NaN	5.52	18.13	NaN	NaN	3.60

Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3)

0	24.63	0.52	2.95
1	20.49	0.13	2.01
2	13.04	0.45	3.52
3	19.42	0.65	3.98
4	25.89	0.60	3.53
5 6	25.16 35.32	0.74 0.36	3.80 3.20
7	31.22	0.34	1.94
8	24.32	0.51	3.27
9	21.71	0.55	3.88
10	24.67	0.74	3.79
11	30.78	0.74	3.52
12	30.94	6.60	4.35
13	44.53	0.23	0.77
14 15	45.14 50.70	1.00 0.44	2.82 1.41
16	52.13	0.55	2.52
17	50.51	0.88	2.55
18	63.89	0.00	1.10
19	62.39	0.26	2.05
20	55.96	0.00	1.59
21	62.37	0.00	2.37
22	54.17	0.42	2.69
23	65.30	0.31	2.15
24 25	71.66	0.00	0.84
25 26	51.52 24.48	0.00 0.00	0.77 1.33
27	16.77	0.00	1.06
28	15.22	0.21	1.67
29	21.93	0.41	2.19
30	23.68	1.05	3.39
31	33.67	0.06	2.74
32	27.56	0.51	2.93
33	33.36	0.54	3.29
34 35	38.83 32.93	0.47 0.65	2.02 0.78
36	32.11	0.84	0.54
37	35.07	1.32	1.34
38	40.37	1.50	2.12
39	39.35	1.32	1.89
40	32.65	1.34	1.90
41	29.16	1.34	1.90
42	34.39	1.41	1.96
43 44	34.28 52.83	1.90 2.23	2.33 2.59
44 45	49.97	1.94	2.39
46	67.63	1.67	2.16
47	47.98	1.34	1.90
48	39.44	1.34	1.90
49	32.65	1.39	1.95

df8 = pd.read_excel (r'/content/Data 2017-2022.xlsx')
df8.bfill(inplace=True)
df8.head(50) #after imputation

From Date	To Date	PM2.5 (ug/m3)	PM10
(ug/m3) \ 0 01-Jan-2017 - 00:00	02-Jan-2017 - 00:00	32.61	
60.71 1 02-Jan-2017 - 00:00	03-Jan-2017 - 00:00	22.93	
60.71 2 03-Jan-2017 - 00:00	04-Jan-2017 - 00:00	24.19	
60.71 3 04-Jan-2017 - 00:00	05-Jan-2017 - 00:00	33.61	
60.71 4 05-Jan-2017 - 00:00	06-Jan-2017 - 00:00	129.38	
60.71 5 06-Jan-2017 - 00:00	07-Jan-2017 - 00:00	64.52	
60.71 6 07-Jan-2017 - 00:00	08-Jan-2017 - 00:00	45.01	
60.71 7 08-Jan-2017 - 00:00	09-Jan-2017 - 00:00	32.71	
60.71 8 09-Jan-2017 - 00:00	10-Jan-2017 - 00:00	33.66	
60.71 9 10-Jan-2017 - 00:00	11-Jan-2017 - 00:00	34.82	
60.71 10 11-Jan-2017 - 00:00	12-Jan-2017 - 00:00	38.33	
60.71 11 12-Jan-2017 - 00:00	13-Jan-2017 - 00:00	42.32	
60.71 12 13-Jan-2017 - 00:00	14-Jan-2017 - 00:00	103.51	
60.71 13 14-Jan-2017 - 00:00	15-Jan-2017 - 00:00	33.96	
60.71 14 15-Jan-2017 - 00:00	16-Jan-2017 - 00:00	54.93	
60.71 15 16-Jan-2017 - 00:00	17-Jan-2017 - 00:00	46.01	
60.71 16 17-Jan-2017 - 00:00	18-Jan-2017 - 00:00	47.76	
60.71 17 18-Jan-2017 - 00:00	19-Jan-2017 - 00:00	36.59	
60.71 18 19-Jan-2017 - 00:00	20-Jan-2017 - 00:00	42.82	
	21-Jan-2017 - 00:00	68.26	
60.71 20 21-Jan-2017 - 00:00	22-Jan-2017 - 00:00	38.96	
60.71 21 22-Jan-2017 - 00:00 60.71	23-Jan-2017 - 00:00	26.28	

22 23-Jan-2017 60.71	-	00:00	24-Jan-2017	-	00:00	33.59
23 24-Jan-2017 60.71	-	00:00	25-Jan-2017	-	00:00	29.59
24 25-Jan-2017 60.71	-	00:00	26-Jan-2017	-	00:00	40.49
25 26-Jan-2017 60.71	-	00:00	27-Jan-2017	-	00:00	29.95
26 27-Jan-2017 60.71	-	00:00	28-Jan-2017	-	00:00	18.93
27 28-Jan-2017 60.71	-	00:00	29-Jan-2017	-	00:00	7.64
28 29-Jan-2017 60.71						8.65
29 30-Jan-2017 60.71						19.69
30 31-Jan-2017 60.71						35.80
31 01-Feb-2017 60.71 32 02-Feb-2017						37.48 34.84
60.71 33 03-Feb-2017						36.89
60.71 34 04-Feb-2017						40.92
60.71 35 05-Feb-2017	_	00:00	06-Feb-2017	_	00:00	43.07
60.71 36 06-Feb-2017	-	00:00	07-Feb-2017	-	00:00	42.05
60.71 37 07-Feb-2017	-	00:00	08-Feb-2017	-	00:00	47.71
60.71 38 08-Feb-2017	-	00:00	09-Feb-2017	-	00:00	61.60
60.71 39 09-Feb-2017 60.71	-	00:00	10-Feb-2017	-	00:00	54.49
40 10-Feb-2017 60.71	-	00:00	11-Feb-2017	-	00:00	50.29
41 11-Feb-2017 60.71	-	00:00	12-Feb-2017	-	00:00	44.20
42 12-Feb-2017 60.71	-	00:00	13-Feb-2017	-	00:00	26.51
43 13-Feb-2017 60.71	-	00:00	14-Feb-2017	-	00:00	17.22
44 14-Feb-2017 60.71						22.04
45 15-Feb-2017 60.71						30.00
46 16-Feb-2017 60.71	-	00:00	1/-Feb-2017	-	00:00	36.92

47 17 60.71	-Feb-2017	- 00:00 18-	Feb-2017 -	00:00	38.03	
48 18	-Feb-2017	- 00:00 19-	Feb-2017 -	00:00	29.63	
60.71 49 19 60.71	-Feb-2017	- 00:00 20-	Feb-2017 -	00:00	34.19	
NO (mg/m3)		NO2 (ug/m3)	NOx (ppb)	NH3 (ug/m3)	S02 (ug/m3)	
0	2.36	9.78	27.66	5.36	2.11	
1.91 1	2.33	8.21	27.66	5.36	2.86	
1.91 2	11.39	17.28	27.66	5.36	7.73	
1.91 3	6.06	12.32	27.66	5.36	2.72	
1.91 4	5.58	12.67	27.66	5.36	2.65	
1.91 5	6.91	14.38	27.66	5.36	2.28	
1.91 6	5.72	9.66	27.66	5.36	2.19	
1.91 7	4.33	10.57	27.66	5.36	2.23	
1.91 8	6.02	15.74	27.66	5.36	3.76	
1.91 9	6.96	16.31	27.66	5.36	2.71	
1.91 10	6.57	18.45	27.66	5.36	2.75	
1.91 11	22.63	26.31	27.66	5.36	2.57	
1.91 12	32.73	38.26	27.66	5.36	2.65	
1.91 13	24.83	27.27	27.66	5.36	2.48	
1.91 14	16.29	20.86	27.66	5.36	3.54	
1.91 15	4.18	14.54	27.66	5.36	3.23	
1.91 16	4.08	20.86	27.66	5.36	4.49	
1.91 17	4.76	20.39	27.66	5.36	3.52	
1.91 18	2.71	15.88	27.66	5.36	2.55	
1.91 19	2.86	18.64	27.66	5.36	4.17	
1.91 20	2.37	15.49	27.66	5.36	2.53	

1.91 21	2.47	15.81	27.66	5.36	2.99
1.91 22	3.12	15.40	27.66	5.36	4.70
1.91 23 1.91	2.83	14.93	27.66	5.36	3.77
24 1.91	2.37	12.49	27.66	5.36	2.63
25 1.91	2.44	12.68	27.66	5.36	2.60
26 1.91	2.53	12.44	27.66	5.36	2.46
27 1.91	2.59	11.63	27.66	5.36	2.35
28 1.91	6.07	13.59	27.66	5.36	2.90
29 1.91	6.07	18.00	27.66	5.36	2.77
30 1.91	6.47	17.91	27.66	5.36	3.97
31 1.91	3.63	16.19	27.66	5.36	2.27
32 1.01	6.89	18.19	27.66	5.36	2.41
33 1.21	6.19	18.27	27.66	5.36	3.09
34 1.21	6.62	17.96	27.66	5.36	2.64
35 1.21	5.41	18.31	27.66	5.36	2.90
36 1.21	6.82	21.76	27.66	5.36	2.83
37 1.21	9.01	25.18	27.66	5.36	4.03
38 1.21	9.12	23.48	27.66	5.36	2.85
39 1.21	5.63	22.22	27.66	5.36	2.97
40 1.21	6.47	21.84	27.66	5.36	2.74
41 1.21	10.70	22.33	27.66	5.36	2.32
42 1.21	5.93	17.56	27.66	5.36	2.20
43 1.21	5.46	15.74	27.66	5.36	2.54
44 1.21	2.54	12.35	27.66	5.36	2.17
45	3.77	14.66	27.66	5.36	2.57

1.21	-						
46 1.21	_	3.26	14.88	27.	66	5.36	2.20
47 1.21		5.55	18.64	27.	66	5.36	2.10
48 1.21		5.58	17.27	27.	66	5.36	1.99
49 1.16		5.52	18.13	27.	66	5.36	3.60
		((m2)	Danzana (.	(m2)	Taluana	//m2\	
0	0zone	(ug/m3) 24.63	Benzene (ı	0.52	Toluene	(ug/iii3) 2.95	
1		20.49		0.13		2.01	
2		13.04 19.42		0.45 0.65		3.52 3.98	
4		25.89		0.60		3.53	
5		25.16		0.74		3.80	
6 7		35.32 31.22		0.36 0.34		3.20 1.94	
8		24.32		0.51		3.27	
9		21.71		0.55		3.88	
10		24.67 30.78		0.74 0.74		3.79	
11 12		30.76		6.60		3.52 4.35	
13		44.53		0.23		0.77	
14		45.14		1.00		2.82	
15 16		50.70 52.13		0.44 0.55		1.41 2.52	
17		50.51		0.88		2.55	
18		63.89		0.00		1.10	
19 20		62.39 55.96		0.26 0.00		2.05 1.59	
21		62.37		0.00		2.37	
22		54.17		0.42		2.69	
23		65.30		0.31		2.15	
24 25		71.66 51.52		0.00 0.00		0.84 0.77	
26		24.48		0.00		1.33	
27		16.77		0.00		1.06	
28 29		15.22 21.93		0.21 0.41		1.67 2.19	
30		23.68		1.05		3.39	
31		33.67		0.06		2.74	
32 33		27.56 33.36		0.51 0.54		2.93 3.29	
34		38.83		0.34		2.02	
35		32.93		0.65		0.78	
36 27		32.11		0.84		0.54	
37 38		35.07 40.37		1.32 1.50		1.34 2.12	

```
39
             39.35
                                 1.32
40
             32.65
                                 1.34
41
             29.16
                                 1.34
42
             34.39
                                 1.41
43
             34.28
                                 1.90
44
             52.83
                                 2.23
45
             49.97
                                 1.94
46
             67.63
                                 1.67
47
             47.98
                                 1.34
48
             39.44
                                 1.34
49
             32.65
                                 1.39
df.info()
df8.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884
Data columns (total 13 columns):
 #
                        Non-Null Count
     Column
                                          Dtype
- - -
     _ _ _ _ _ _
                        _ _ _ _ _ _ _ _ _ _ _ _ _ _ _
 0
     From Date
                        1885 non-null
                                          object
 1
     To Date
                        1885 non-null
                                          object
 2
     PM2.5 (ug/m3)
                        1841 non-null
                                          float64
 3
     PM10 (ug/m3)
                        306 non-null
                                          float64
 4
     NO (ug/m3)
                        1844 non-null
                                          float64
 5
     N02 (uq/m3)
                        1843 non-null
                                          float64
     NOx (ppb)
                        1598 non-null
                                          float64
 6
 7
     NH3 (uq/m3)
                        306 non-null
                                          float64
     S02 (ug/m3)
 8
                        1830 non-null
                                          float64
 9
     CO (mg/m3)
                        1687 non-null
                                          float64
 10
                                          float64
     Ozone (ug/m3)
                        1833 non-null
     Benzene (ug/m3)
                        1859 non-null
                                          float64
 11
     Toluene (ug/m3)
 12
                        1859 non-null
                                          float64
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884
Data columns (total 13 columns):
 #
                        Non-Null Count
     Column
                                         Dtype
     _ _ _ _ _
- - -
 0
     From Date
                        1885 non-null
                                          object
 1
     To Date
                        1885 non-null
                                          object
 2
     PM2.5 (ug/m3)
                                          float64
                        1885 non-null
 3
     PM10 (uq/m3)
                        1885 non-null
                                          float64
 4
     NO (ug/m3)
                        1885 non-null
                                          float64
 5
     N02 (ug/m3)
                        1885 non-null
                                          float64
 6
                                          float64
     NOx (ppb)
                        1885 non-null
 7
     NH3 (ug/m3)
                        1885 non-null
                                          float64
 8
     S02 (ug/m3)
                                          float64
                        1885 non-null
 9
     CO (mg/m3)
                        1885 non-null
                                          float64
 10
     Ozone (ug/m3)
                        1885 non-null
                                          float64
```

1.89

1.90

1.90

1.96

2.33

2.59

2.37

2.16

1.90

1.90

1.95

```
11 Benzene (ug/m3) 1885 non-null float64
12 Toluene (ug/m3) 1885 non-null float64
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
```

As observed there are no null values in the dataset now (there are a total of 1885 nn values in all the colums) and this method is more effective than the previous as the null values right from the beginning of each column is imputed with some numerical.

Median imputation

Median imputation is a method in which the missing values are replaced with the median value of the entire feature column. When the data is skewed, it is good to consider using the median value for replacing the missing values.

df.head(50)

	Fr	om Date		To	Date	PM2.5	(ug/m3)	PM10
(ug,	/m3) \							
0 NaN	01-Jan-2017	- 00:00	02-Jan-2017	-	00:00		32.61	
1 NaN	02-Jan-2017	- 00:00	03-Jan-2017	-	00:00		22.93	
2 NaN	03-Jan-2017	- 00:00	04-Jan-2017	-	00:00		24.19	
3 NaN	04-Jan-2017	- 00:00	05-Jan-2017	-	00:00		33.61	
4 NaN	05-Jan-2017	- 00:00	06-Jan-2017	-	00:00		129.38	
5 NaN	06-Jan-2017	- 00:00	07-Jan-2017	-	00:00		64.52	
6 NaN	07-Jan-2017	- 00:00	08-Jan-2017	-	00:00		45.01	
7 NaN	08-Jan-2017	- 00:00	09-Jan-2017	-	00:00		32.71	
8 NaN	09-Jan-2017	- 00:00	10-Jan-2017	-	00:00		33.66	
9 NaN	10-Jan-2017	- 00:00	11-Jan-2017	-	00:00		34.82	
10 NaN	11-Jan-2017	- 00:00	12-Jan-2017	-	00:00		38.33	
11 NaN	12-Jan-2017	- 00:00	13-Jan-2017	-	00:00		42.32	
12 NaN	13-Jan-2017	- 00:00	14-Jan-2017	-	00:00		103.51	
NaN 13 NaN	14-Jan-2017	- 00:00	15-Jan-2017	-	00:00		33.96	
14 NaN	15-Jan-2017	- 00:00	16-Jan-2017	-	00:00		54.93	

15 N- N	16-Jan-2017	-	00:00	17-Jan-2017	-	00:00	46.01
NaN 16	17-Jan-2017	-	00:00	18-Jan-2017	-	00:00	47.76
NaN 17	18-Jan-2017	- (00:00	19-Jan-2017	_	00:00	36.59
NaN 18	19-Jan-2017	_	00:00	20-Jan-2017	_	00:00	42.82
NaN							
19 NaN	20-Jan-2017	-	00:00	21-Jan-2017	-	00:00	68.26
20 NaN	21-Jan-2017	-	00:00	22-Jan-2017	-	00:00	38.96
21	22-Jan-2017	-	00:00	23-Jan-2017	-	00:00	26.28
NaN 22	23-Jan-2017	-	00:00	24-Jan-2017	-	00:00	33.59
NaN 23	24-Jan-2017	_ (00:00	25-Jan-2017	_	00:00	29.59
NaN	25-Jan-2017		00.00				40.49
24 NaN				26-Jan-2017			
25 NaN	26-Jan-2017	-	00:00	27-Jan-2017	-	00:00	29.95
26 NaN	27-Jan-2017	-	00:00	28-Jan-2017	-	00:00	18.93
27	28-Jan-2017	-	00:00	29-Jan-2017	-	00:00	7.64
NaN 28	29-Jan-2017	_	00:00	30-Jan-2017	_	00:00	8.65
NaN 29	30-Jan-2017	_		31-Jan-2017			19.69
NaN							
30 NaN	31-Jan-2017	-	00:00	01-Feb-2017	-	00:00	35.80
31 NaN	01-Feb-2017	-	00:00	02-Feb-2017	-	00:00	37.48
32	02-Feb-2017	-	00:00	03-Feb-2017	-	00:00	34.84
NaN 33	03-Feb-2017	_	00:00	04-Feb-2017	_	00:00	36.89
NaN 34	04-Feb-2017	_	00:00	05-Feb-2017	_	00:00	40.92
NaN							
35 NaN	05-Feb-2017	-	00:00	06-Feb-2017	-	00:00	43.07
36 NaN	06-Feb-2017	-	00:00	07-Feb-2017	-	00:00	42.05
37 NaN	07-Feb-2017	-	00:00	08-Feb-2017	-	00:00	47.71
38	08-Feb-2017	-	00:00	09-Feb-2017	-	00:00	61.60
NaN 39	09-Feb-2017	_	00:00	10-Feb-2017	_	00:00	54.49
NaN							

40		-Feb-2017	-	00:00	11-	Feb-2	2017	-	00:00)		50.	. 29	
Na 41	l 11	-Feb-2017	-	00:00	12-	Feb-2	2017	-	00:00)		44.	. 20	
Na 42	2 12	-Feb-2017	-	00:00	13-	Feb-2	2017	-	00:00)		26.	.51	
Na 43	3 13	-Feb-2017	-	00:00	14-	Feb-2	2017	-	00:00)		17.	. 22	
Na 44	14	-Feb-2017	-	00:00	15-	Feb-2	2017	-	00:00)		22.	. 04	
Na 45	5 15	-Feb-2017	-	00:00	16-	Feb-2	2017	-	00:00)		30.	. 00	
Na 46	5 16	-Feb-2017	-	00:00	17-	Feb-2	2017	-	00:00)		36.	.92	
Na 47	7 17	-Feb-2017	-	00:00	18-	Feb-2	2017	-	00:00)		38.	. 03	
Na 48	3 18	-Feb-2017	-	00:00	19-	Feb-2	2017	-	00:00)		29.	. 63	
Na 49	19	-Feb-2017	-	00:00	20-	Feb-2	2017	-	00:00)		34.	. 19	
Na	aN													
		(ug/m3)	N	02 (ug/r	n3)	NOx	(ppb)	NH3	(ug/	m3)	S02	(ug/m3)	CO
0	ng/m3	2.36		9	.78		Na	N			NaN		2.11	
Na 1		2.33		8	.21		Na	N			NaN		2.86	
Na 2		11.39		17	. 28		Na	N			NaN		7.73	
Na 3		6.06		12	. 32		Na	N			NaN		2.72	
Na 4		5.58		12	. 67		Na	N			NaN		2.65	
Na 5		6.91		14	. 38		Na	N			NaN		2.28	
Na 6		5.72		9	.66		Na	N			NaN		2.19	
Na 7		4.33		10	.57		Na	N			NaN		2.23	
Na 8		6.02		15	.74		Na	N			NaN		3.76	
Na 9		6.96		16	.31		Na	N			NaN		2.71	
Na 10		6.57		18	. 45		Na	N			NaN		2.75	
Na 11		22.63		26	.31		Na	N			NaN		2.57	
Na 12	2	32.73		38	. 26		Na	N			NaN		2.65	
Na 13		24.83		27	. 27		Na	N			NaN		2.48	

NaN 14	16.29	20.86	NaN	NaN	3.54
NaN	10.29	20.00	IVAIV	INGIN	3.34
15	4.18	14.54	NaN	NaN	3.23
NaN					
16 NaN	4.08	20.86	NaN	NaN	4.49
17	4.76	20.39	NaN	NaN	3.52
NaN 18	2.71	15.88	NaN	NaN	2.55
NaN	21,1	13100	Hall	iidii	2.33
19	2.86	18.64	NaN	NaN	4.17
NaN 20	2.37	15.49	NaN	NaN	2.53
NaN	,				
21 NaN	2.47	15.81	NaN	NaN	2.99
22	3.12	15.40	NaN	NaN	4.70
NaN	2 22	14.00			2 77
23 NaN	2.83	14.93	NaN	NaN	3.77
24	2.37	12.49	NaN	NaN	2.63
NaN 25	2.44	12.68	NaN	NaN	2.60
NaN					
26 NaN	2.53	12.44	NaN	NaN	2.46
NaN 27	2.59	11.63	NaN	NaN	2.35
NaN					
28 NaN	6.07	13.59	NaN	NaN	2.90
29	6.07	18.00	NaN	NaN	2.77
NaN					
30 NaN	6.47	17.91	NaN	NaN	3.97
31	3.63	16.19	NaN	NaN	2.27
1.91 32	6.89	18.19	NaN	NaN	2.41
1.01	0.03	10.13	Han	nan	2
33 NaN	6.19	18.27	NaN	NaN	3.09
34	6.62	17.96	NaN	NaN	2.64
NaN 35	5 /11	18.31	NaN	NaN	2 00
NaN	5.41	10.31	NaN	NaN	2.90
36	6.82	21.76	NaN	NaN	2.83
NaN 37	9.01	25.18	NaN	NaN	4.03
NaN	0 12	22 40	NaN	NaN	7 OF
38	9.12	23.48	NaN	NaN	2.85

5.63 6.47 10.70 5.93 5.46 2.54 3.77 3.26	22.22 21.84 22.33 17.56 15.74 12.35 14.66 14.88	NaN NaN NaN NaN NaN NaN	NaN NaN NaN NaN NaN NaN	2.97 2.74 2.32 2.20 2.54 2.17
10.70 5.93 5.46 2.54 3.77 3.26	22.33 17.56 15.74 12.35 14.66	NaN NaN NaN NaN	NaN NaN NaN NaN	2.32 2.20 2.54
10.70 5.93 5.46 2.54 3.77 3.26	22.33 17.56 15.74 12.35 14.66	NaN NaN NaN NaN	NaN NaN NaN NaN	2.32 2.20 2.54
5.93 5.46 2.54 3.77 3.26	17.56 15.74 12.35 14.66	NaN NaN NaN NaN	NaN NaN NaN	2.20 2.54
5.46 2.54 3.77 3.26	15.74 12.35 14.66	NaN NaN NaN	NaN NaN	2.54
5.46 2.54 3.77 3.26	15.74 12.35 14.66	NaN NaN NaN	NaN NaN	2.54
2.54 3.77 3.26	12.35 14.66	NaN NaN	NaN	
3.77 3.26	14.66	NaN		2.17
3.77 3.26	14.66	NaN		2.17
3.26			NaN	
	14.88			2.57
	11.00	NaN	NaN	2.20
5 55		Nan	Nan	2.20
5.55	18.64	NaN	NaN	2.10
5.58	17.27	NaN	NaN	1.99
	1,.1,			2.55
5.52	18.13	NaN	NaN	3.60
zone (ug/m3) 24.63 20.49 13.04 19.42 25.89 25.16 35.32 31.22 24.32 21.71 24.67 30.78 30.94 44.53 45.14 50.70 52.13 50.51 63.89 62.39 55.96 62.37 54.17 65.30	6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	1.52 1.13 1.45 1.65 1.60 1.34 1.55 1.74 1.60 1.23 1.00 1.44 1.55 1.88 1.00 1.42 1.31	ne (ug/m3) 2.95 2.01 3.52 3.98 3.53 3.80 3.20 1.94 3.27 3.88 3.79 3.52 4.35 0.77 2.82 1.41 2.52 2.55 1.10 2.05 1.59 2.37 2.69 2.15	
	24.63 20.49 13.04 19.42 25.89 25.16 35.32 31.22 24.32 21.71 24.67 30.78 30.94 44.53 45.14 50.70 52.13 50.51 63.89 62.39 55.96 62.37 54.17	5.58 17.27 5.52 18.13 zone (ug/m3) Benzene (ug/24.63	5.58 17.27 NaN 5.52 18.13 NaN 2000 (ug/m3) Benzene (ug/m3) Toluer 24.63 0.52 20.49 0.13 13.04 0.45 19.42 0.65 25.89 0.60 25.16 0.74 35.32 0.36 31.22 0.34 24.32 0.51 21.71 0.55 24.67 0.74 30.78 0.74 30.78 0.74 30.94 6.60 44.53 0.23 45.14 1.00 50.70 0.44 52.13 0.55 50.51 0.88 63.89 0.00 62.39 0.26 55.96 0.00 62.37 0.00 54.17 0.42 65.30 0.31	5.58 17.27 NaN NaN 5.52 18.13 NaN NaN 20ne (ug/m3) Benzene (ug/m3) Toluene (ug/m3) 24.63 0.52 2.95 20.49 0.13 2.01 13.04 0.45 3.52 19.42 0.65 3.98 25.89 0.60 3.53 25.16 0.74 3.80 35.32 0.36 3.20 31.22 0.34 1.94 24.32 0.51 3.27 21.71 0.55 3.88 24.67 0.74 3.79 30.78 0.74 3.79 30.78 0.74 3.52 30.94 6.60 4.35 44.53 0.23 0.77 45.14 1.00 2.82 50.70 0.44 1.41 52.13 0.55 2.52 50.51 0.88 2.55 63.89 0.00 1.10 62.39 0.26 2.05 55.96 0.00 2.37 54.17 0.42 2.69 65.30 0.31 2.15

```
25
                                0.00
                                                  0.77
            51.52
26
            24.48
                                0.00
                                                  1.33
                                                  1.06
27
            16.77
                                0.00
            15.22
28
                                0.21
                                                  1.67
29
            21.93
                                0.41
                                                  2.19
30
            23.68
                                1.05
                                                  3.39
31
            33.67
                                0.06
                                                  2.74
32
            27.56
                                0.51
                                                  2.93
33
            33.36
                                0.54
                                                  3.29
34
            38.83
                                0.47
                                                  2.02
35
            32.93
                                0.65
                                                  0.78
36
            32.11
                                0.84
                                                  0.54
37
            35.07
                                1.32
                                                  1.34
            40.37
38
                                1.50
                                                  2.12
39
            39.35
                                1.32
                                                  1.89
40
            32.65
                                1.34
                                                  1.90
41
            29.16
                                1.34
                                                  1.90
            34.39
42
                                1.41
                                                  1.96
43
            34.28
                                1.90
                                                  2.33
                                                  2.59
44
            52.83
                                2.23
45
            49.97
                                1.94
                                                  2.37
46
            67.63
                                1.67
                                                  2.16
            47.98
47
                                1.34
                                                  1.90
48
            39.44
                                1.34
                                                  1.90
            32.65
49
                                1.39
                                                  1.95
df9 = pd.read excel (r'/content/Data 2017-2022.xlsx')
c=['PM2.5 (ug/m3)','PM10 (ug/m3)','N0 (ug/m3)','N02 (ug/m3)','N0x]
(ppb)','NH3 (ug/m3)','S02 (ug/m3)','C0 (mg/m3)','Ozone
(ug/m3)', 'Benzene (ug/m3)', 'Toluene (ug/m3)']
df9[c] = df9[c].fillna(df9[c].median())
df9.head(50) #all the missing values in the columns are replaced withe
the median values of the same column
                                       To Date PM2.5 (ug/m3)
               From Date
                                                                 PM10
(ug/m3)
    01-Jan-2017 - 00:00 02-Jan-2017 - 00:00
                                                         32.61
67.25
    02-Jan-2017 - 00:00
                          03-Jan-2017 - 00:00
                                                         22.93
67.25
2
    03-Jan-2017 - 00:00
                          04-Jan-2017 - 00:00
                                                         24.19
67.25
    04-Jan-2017 - 00:00
                          05-Jan-2017 - 00:00
                                                         33.61
67.25
                          06-Jan-2017 - 00:00
    05-Jan-2017 - 00:00
                                                        129.38
67.25
                          07-Jan-2017 - 00:00
5
    06-Jan-2017 - 00:00
                                                         64.52
67.25
```

08-Jan-2017 - 00:00

45.01

32.71

07-Jan-2017 - 00:00

08-Jan-2017 - 00:00 09-Jan-2017 - 00:00

6

67.25

67	25						
	.25 09-Jan-2017	_	00.00	10-lan-2017	_	00.00	33.66
	.25		00.00	10 3411 2017		00.00	33.00
9	10-Jan-2017	-	00:00	11-Jan-2017	-	00:00	34.82
	.25 11-Jan-2017	_	00:00	12-Jan-2017	_	00:00	38.33
67	. 25						
	12-Jan-2017 .25	-	00:00	13-Jan-2017	-	00:00	42.32
12	13-Jan-2017	-	00:00	14-Jan-2017	-	00:00	103.51
	.25 14-Jan-2017	_	00:00	15-Jan-2017	_	00:00	33.96
67	. 25						
	15-Jan-2017 .25	-	00:00	16-Jan-2017	-	00:00	54.93
15	16-Jan-2017	-	00:00	17-Jan-2017	-	00:00	46.01
	.25		00.00	10 lan 2017		00.00	47.76
	17-Jan-2017 . 25	-	00:00	18-Jan-2017	-	00:00	47.70
	18-Jan-2017	-	00:00	19-Jan-2017	-	00:00	36.59
	.25 19-Jan-2017	_	00:00	20-Jan-2017	_	00:00	42.82
67	. 25						
	20-Jan-2017 .25	-	00:00	21-Jan-2017	-	00:00	68.26
	21-Jan-2017	-	00:00	22-Jan-2017	-	00:00	38.96
	.25		00.00	22 100 2017		00.00	26.20
	22-Jan-2017 .25	-	00:00	23-Jan-2017	-	00:00	26.28
	23-Jan-2017	-	00:00	24-Jan-2017	-	00:00	33.59
	.25 24-Jan-2017	_	00:00	25-Jan-2017	_	00:00	29.59
67	. 25						
	25-Jan-2017 .25	-	00:00	26-Jan-2017	-	00:00	40.49
	26-Jan-2017	-	00:00	27-Jan-2017	-	00:00	29.95
	.25 27-Jan-2017		00.00	20 lan 2017		00.00	18.93
	.25	-	00:00	20-Jan-2017	-	00:00	10.93
	28-Jan-2017	-	00:00	29-Jan-2017	-	00:00	7.64
	.25 29-Jan-2017	_	00:00	30-Jan-2017	_	00:00	8.65
67	. 25						
	30-Jan-2017 .25	-	00:00	31-Jan-2017	-	00:00	19.69
	31-Jan-2017	-	00:00	01-Feb-2017	-	00:00	35.80
	.25		00.00	02 Eab 2017		00.00	27 40
	01-Feb-2017 .25	-	00:00	uz-ren-zul/	-	טט: טט	37.48
	02-Feb-2017	-	00:00	03-Feb-2017	-	00:00	34.84

67.25												
	-Feb-2017	-	00:00	04-	Feb-2	2017	-	00:00)	36	. 89	
34 04-	-Feb-2017	-	00:00	05-	Feb-2	2017	-	00:00)	40	. 92	
67.25 35 05-	-Feb-2017	-	00:00	06-	-Feb-2	2017	-	00:00)	43	. 07	
67.25 36 06-	-Feb-2017	_	00:00	07-	-Feb-2	2017	_	00:00)	42	. 05	
67.25	-Feb-2017				.Feh-	2017	_	00.00)		.71	
67.25												
67.25	-Feb-2017									61		
39 09- 67.25	-Feb-2017	-	00:00	10-	·Feb-2	2017	-	00:00)	54	. 49	
	-Feb-2017	-	00:00	11-	Feb-2	2017	-	00:00)	50	. 29	
41 11-	-Feb-2017	-	00:00	12-	Feb-2	2017	-	00:00)	44	. 20	
	-Feb-2017	-	00:00	13-	Feb-2	2017	-	00:00)	26	.51	
67.25 43 13-	-Feb-2017	_	00:00	14-	-Feb-2	2017	_	00:00)	17	. 22	
67.25 44 14-	-Feb-2017	_	00:00	15-	Feb-	2017	_	00:00)	22	. 04	
67.25)			
67.25	-Feb-2017										. 00	
46 16- 67.25	-Feb-2017	-	00:00	17-	·Feb-2	2017	-	00:00)	36	. 92	
47 17- 67.25	-Feb-2017	-	00:00	18-	Feb-2	2017	-	00:00)	38	. 03	
	-Feb-2017	-	00:00	19-	Feb-2	2017	-	00:00)	29	. 63	
49 19-	-Feb-2017	-	00:00	20-	Feb-2	2017	-	00:00)	34	. 19	
67.25												
NO (mg/m3)	(ug/m3)) \	N	02 (ug/r	m3)	N0x	(ppb)	NH3	(ug/m3)	S02	(ug/m3)	C0
0	2.36		9	.78		14.61	.5		14.455		2.11	
0.74 1	2.33		8	.21		14.61	.5		14.455		2.86	
0.74 2	11.39		17	. 28	•	14.61	.5		14.455		7.73	
0.74 3	6.06		12	.32		14.61	.5		14.455		2.72	
0.74 4	5.58			. 67		14.61			14.455		2.65	
0.74 5	6.91			.38		14.61			14.455		2.28	
0.74	0.91		14	. 50	•	- - .01	ر.		14.477		2.20	

6 0.74	5.72	9.66	14.615	14.455	2.19
7 0.74	4.33	10.57	14.615	14.455	2.23
8 0.74	6.02	15.74	14.615	14.455	3.76
9 0.74	6.96	16.31	14.615	14.455	2.71
10 0.74	6.57	18.45	14.615	14.455	2.75
11 0.74	22.63	26.31	14.615	14.455	2.57
12 0.74	32.73	38.26	14.615	14.455	2.65
13 0.74	24.83	27.27	14.615	14.455	2.48
14 0.74	16.29	20.86	14.615	14.455	3.54
15 0.74	4.18	14.54	14.615	14.455	3.23
16 0.74	4.08	20.86	14.615	14.455	4.49
17 0.74	4.76	20.39	14.615	14.455	3.52
18 0.74	2.71	15.88	14.615	14.455	2.55
19 0.74	2.86	18.64	14.615	14.455	4.17
20 0.74	2.37	15.49	14.615	14.455	2.53
21 0.74	2.47	15.81	14.615	14.455	2.99
22 0.74	3.12	15.40	14.615	14.455	4.70
23 0.74	2.83	14.93	14.615	14.455	3.77
24 0.74	2.37	12.49	14.615	14.455	2.63
25 0.74		12.68	14.615	14.455	2.60
26 0.74 27	2.53	12.44	14.615	14.455	2.46
0.74	2.59	11.63	14.615	14.455	2.35
28 0.74 29	6.07 6.07	13.59 18.00	14.615 14.615	14.455 14.455	2.90
0.74 30	6.47	17.91	14.615	14.455	3.97
0.74	0.47	17.91	14.013	14.433	3.31

31	3.63	16.19	14.615	14.455	2.27
1.91					
32 1.01	6.89	18.19	14.615	14.455	2.41
33 0.74	6.19	18.27	14.615	14.455	3.09
34	6.62	17.96	14.615	14.455	2.64
0.74 35	5.41	18.31	14.615	14.455	2.90
0.74 36	6.82	21.76	14.615	14.455	2.83
0.74 37	9.01	25.18	14.615	14.455	4.03
0.74 38	9.12	23.48	14.615	14.455	2.85
0.74 39	5.63	22.22	14.615	14.455	2.97
0.74 40	6.47	21.84	14.615	14.455	2.74
0.74 41	10.70	22.33	14.615	14.455	2.32
0.74 42	5.93	17.56	14.615	14.455	2.20
0.74 43	5.46	15.74	14.615	14.455	2.54
0.74 44	2.54	12.35	14.615	14.455	2.17
0.74 45	3.77	14.66	14.615	14.455	2.57
0.74 46	3.26	14.88	14.615	14.455	2.20
0.74 47	5.55	18.64	14.615	14.455	2.10
1.21 48	5.58	17.27	14.615	14.455	1.99
1.21 49	5.52	18.13	14.615	14.455	3.60
0.74					
0:	zone (ug/m3)	Benzene (ug	g/m3) Tolue	ene (ug/m3)	
0 1	24.63 20.49		0.52 0.13	2.95 2.01	
1 2 3 4 5 6 7 8	13.04		0.45	3.52	
3	19.42		0.65	3.98	
4	25.89		0.60	3.53	
5 6	25.16 35.32		0.74 0.36	3.80 3.20	
7	31.22		0.34	1.94	
8	24.32		0.51	3.27	
9	21.71		0.55	3.88	

10 11 12 13 14 15 16 17 18 19 20 21 22 23	24.67 30.78 30.94 44.53 45.14 50.70 52.13 50.51 63.89 62.39 55.96 62.37 54.17	0.74 0.74 6.60 0.23 1.00 0.44 0.55 0.88 0.00 0.26 0.00 0.00 0.42 0.31	3.79 3.52 4.35 0.77 2.82 1.41 2.52 2.55 1.10 2.05 1.59 2.37 2.69
24 25	71.66 51.52	0.00 0.00	0.84 0.77
26	24.48	0.00	1.33
27	16.77	0.00	1.06
28 29	15.22 21.93	0.21 0.41	1.67 2.19
30	23.68	1.05	3.39
31	33.67	0.06	2.74
32	27.56	0.51	2.93
33	33.36	0.54	3.29
34	38.83	0.47	2.02
35 36	32.93 32.11	0.65	0.78
37	35.07	0.84 1.32	0.54 1.34
38	40.37	1.50	2.12
39	39.35	1.32	1.89
40	32.65	1.34	1.90
41	29.16	1.34	1.90
42	34.39	1.41	1.96
43 44	34.28	1.90 2.23	2.33
45	52.83 49.97	1.94	2.59 2.37
46	67.63	1.67	2.16
47	47.98	1.34	1.90
48	39.44	1.34	1.90
49	32.65	1.39	1.95
<pre>df.info() df9.info()</pre>			
-clace Inan	das core frame D	ataFrama'>	

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884
Data columns (total 13 columns):

#	Column	Non-Null Count	Dtype
0	From Date	1885 non-null	object
1	To Date	1885 non-null	obiect

```
1841 non-null
                                        float64
 2
     PM2.5 (uq/m3)
 3
     PM10 (ug/m3)
                       306 non-null
                                        float64
 4
     NO (ug/m3)
                       1844 non-null
                                        float64
 5
     N02 (uq/m3)
                       1843 non-null
                                        float64
 6
     NOx (ppb)
                       1598 non-null
                                        float64
 7
     NH3 (uq/m3)
                       306 non-null
                                        float64
 8
     S02 (uq/m3)
                       1830 non-null
                                        float64
 9
     CO (mq/m3)
                       1687 non-null
                                        float64
 10
     Ozone (ug/m3)
                       1833 non-null
                                        float64
 11
     Benzene (ug/m3)
                       1859 non-null
                                        float64
 12
     Toluene (ug/m3)
                       1859 non-null
                                        float64
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1885 entries, 0 to 1884
Data columns (total 13 columns):
#
     Column
                       Non-Null Count
                                        Dtype
     _ _ _ _ _
 0
     From Date
                       1885 non-null
                                        object
 1
     To Date
                       1885 non-null
                                        obiect
 2
     PM2.5 (ug/m3)
                                        float64
                       1885 non-null
 3
     PM10 (ug/m3)
                       1885 non-null
                                        float64
 4
                                        float64
     NO (ug/m3)
                       1885 non-null
 5
     N02 (ug/m3)
                       1885 non-null
                                        float64
                       1885 non-null
 6
     NOx (ppb)
                                        float64
 7
     NH3 (ug/m3)
                       1885 non-null
                                        float64
     S02 (ug/m3)
 8
                       1885 non-null
                                        float64
 9
     CO (mg/m3)
                       1885 non-null
                                        float64
     Ozone (ug/m3)
                                        float64
 10
                       1885 non-null
 11
     Benzene (ug/m3)
                       1885 non-null
                                        float64
                                        float64
     Toluene (ug/m3)
                       1885 non-null
dtypes: float64(11), object(2)
memory usage: 191.6+ KB
df9.describe() #this method is more precise than the mean imputation
       PM2.5 (ug/m3)
                       PM10 (ug/m3)
                                       NO (ug/m3)
                                                    N02 (ug/m3)
                                                                    N0x
(ppb)
count
         1885.000000
                        1885.000000
                                      1885.000000
                                                    1885.000000
1885.000000
           30,425263
                          65.772875
                                         6.920477
                                                      12.528886
mean
15.811684
           20.057418
                          10.959726
                                         5.553900
                                                       8.611886
std
8.115541
            0.410000
                          21,600000
                                         0.010000
                                                       0.020000
min
0.000000
25%
           16.850000
                          67.250000
                                         3.330000
                                                       6,630000
10.530000
           27.280000
                          67.250000
                                         5.460000
                                                      11.280000
50%
14.615000
```

67.250000

9.210000

16.310000

75%

39.010000

19.280000 max 106.74000	278.97000	9 371.6100	00 98.6200	00 134.760000	
NH (ug/m3)	H3 (ug/m3)	S02 (ug/m3)	CO (mg/m3)	Ozone (ug/m3)	Benzene
count 18	385.000000	1885.000000	1885.000000	1885.000000	
1885.0000 mean 0.574976	14.100618	6.502992	0.818599	27.500546	
std 2.214523	2.376704	4.991580	1.392886	16.595455	
min 0.000000	5.360000	0.090000	0.000000	0.100000	
25% 0.000000	14.455000	3.920000	0.620000	16.930000	
50% 0.000000	14.455000	4.880000	0.740000	24.680000	
75% 0.340000	14.455000	7.070000	0.880000	34.180000	
max 46.230000	33.680000 9	37.180000	48.020000	183.990000	
count mean std min 25% 50% 75% max	1885.000 1.916 4.354 0.000 0.190 2.730 121.150	000 032 631 000 000 000			

OUTLIER ANALYSIS

An outlier is an observation that appears to deviate markedly from other observations in the sample. Otherwise, Outliers are extreme values that stand out greatly from the overall pattern of values in a dataset or graph.

Outlier analysis is critical in analyzing the data for at least two reasons:

The outliers may negatively bias the entire outcome of an analysis. Detect points that are considered "abnormal," or which don't fit a particular pattern.

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

```
df17 = pd.read excel (r'/content/Data 2017-2022.xlsx',
sheet name='Chennai')
df17.head()
             From Date
                                    To Date PM2.5 (ug/m3)
                                                            PM10
(ua/m3) \
0 01-Jan-2017 - 00:00 02-Jan-2017 - 00:00
                                                      32.61
NaN
1 02-Jan-2017 - 00:00
                        03-Jan-2017 - 00:00
                                                      22.93
NaN
2 03-Jan-2017 - 00:00
                        04-Jan-2017 - 00:00
                                                      24.19
NaN
3 04-Jan-2017 - 00:00
                       05-Jan-2017 - 00:00
                                                      33.61
NaN
4 05-Jan-2017 - 00:00 06-Jan-2017 - 00:00
                                                     129.38
NaN
              NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) SO2 (ug/m3) CO
   NO (ug/m3)
(mg/m3)
                      9.78
         2.36
                                  NaN
                                                NaN
                                                            2.11
0
NaN
         2.33
                      8.21
                                  NaN
                                                NaN
                                                            2.86
1
NaN
2
        11.39
                     17.28
                                  NaN
                                                NaN
                                                            7.73
NaN
         6.06
                     12.32
                                  NaN
                                                            2.72
3
                                                NaN
NaN
         5.58
                     12.67
                                  NaN
                                                NaN
                                                            2.65
NaN
   Ozone (ug/m3)
                  Benzene (ug/m3)
                                   Toluene (ug/m3)
           24.63
                             0.52
                                               2.95
0
           20.49
                             0.13
                                               2.01
1
2
           13.04
                             0.45
                                               3.52
3
                             0.65
           19.42
                                               3.98
           25.89
                             0.60
                                               3.53
```

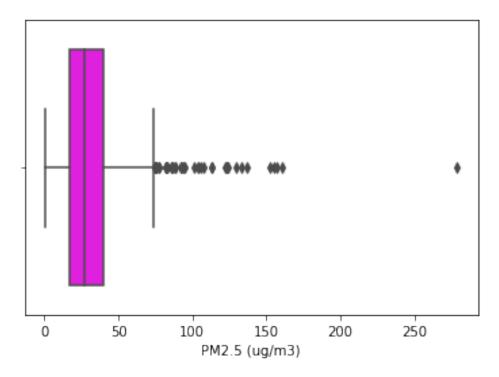
UNIVARIATE VISUALIZATION

- 1. Box plot
- 2. Histogram
- 3. Scatter plot

```
warnings.filterwarnings('ignore')
sns.boxplot(df17['PM2.5 (ug/m3)'],color='magenta')
```

we can clearly see that data points above 65 are considered to be outliers in PM2.5 column

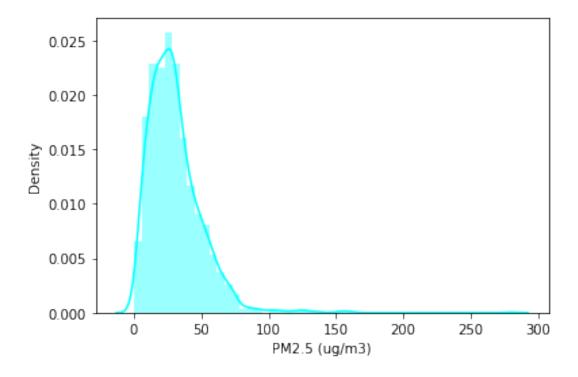
<matplotlib.axes. subplots.AxesSubplot at 0x7fad83a6dd90>



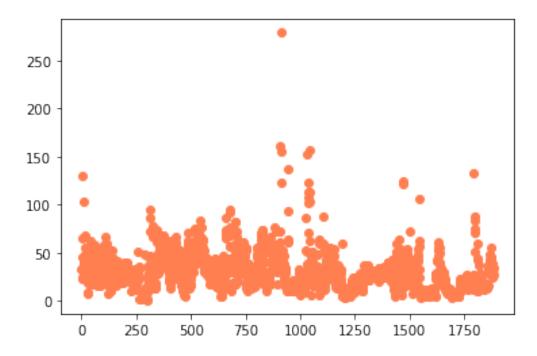
warnings.filterwarnings('ignore')
sns.distplot(df17['PM2.5 (ug/m3)'],color='cyan')

#This plots helps us to identify the nature of distribution. The PM 2.5 is highly skewed and diviates from mean.

<matplotlib.axes._subplots.AxesSubplot at 0x7fad82d0b110>



plt.scatter(df17.index,df17['PM2.5 (ug/m3)'],c='coral')
<matplotlib.collections.PathCollection at 0x7fad82d381d0>

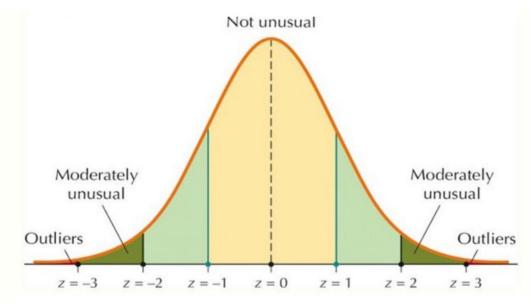


METHOD 1

Z-score treatment

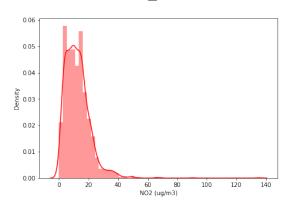
Assumption: Data normally distributed

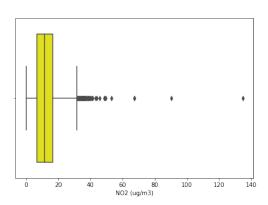
If the z score of a data point is more than 3, it indicates that the data point is quite different from the other data points. Such a data point can be a outlier.



```
#Column used for Z scroe outlier detection: NO2
df10 = pd.read_excel (r'/content/Data 2017-2022.xlsx',
sheet_name='Chennai')
plt.figure(figsize=(16,5))
plt.subplot(1,2,1)
sns.distplot(df10['N02 (ug/m3)'],color='red')
plt.subplot(1,2,2)
sns.boxplot(df10['N02 (ug/m3)'],color='yellow')
#this is the histogram and box plot before oultier detection and handling. The data is positively skewed and contains ouliers.
```

<matplotlib.axes. subplots.AxesSubplot at 0x7fad82b4b590>





#any value above the upper limit and below lower limit are considered to be outliers.

#Upper Limit=mean+ 3*standard deviation #Lower Limit=Mean -3*standard deviation

38.679623230940415 -13.564929796322787

#rows with outliers in NO2 column
df10[(df10['NO2 (ug/m3)'] > upper_limit) | (df10['NO2 (ug/m3)'] <
lower_limit)]</pre>

From Date			To Date	PM2.5 (ug/m3)	PM10
(ug/m3	3) \			_	
180	30-Jun-2017 - 00:00	01-Jul-2017	- 00:00	31.75	
NaN					
249	07-Sep-2017 - 00:00	08-Sep-2017	- 00:00	19.21	
NaN					
250	08-Sep-2017 - 00:00	09-Sep-2017	- 00:00	22.40	
NaN					
318	15-Nov-2017 - 00:00	16-Nov-2017	- 00:00	37.34	
NaN					
341	08-Dec-2017 - 00:00	09-Dec-2017	- 00:00	40.15	

NaN 450	27-Mar-2018	- 00:00	28-Mar-20	918 - 00:00	36	. 68
NaN 451	28-Mar-2018	- 00:00	29-Mar-20	918 - 00:00	26	. 03
NaN 452	29-Mar-2018	- 00:00	30-Mar-20	918 - 00:00	26	. 92
NaN 485	01-May-2018	- 00:00	02-May-20	918 - 00:00	60	. 14
NaN 486	02-May-2018	- 00:00	03-May-20	918 - 00:00	70	.31
NaN 489	05-May-2018	- 00:00	06-May-20	918 - 00:00	41	.37
NaN 490	06-May-2018	- 00:00	07-May-20	918 - 00:00	41	.88
NaN 491	07-May-2018	- 00:00	08-May-20	918 - 00:00	50	.53
NaN 492	08-May-2018	- 00:00	09-May-20	918 - 00:00	24	.55
NaN 493	09-May-2018	- 00:00	10-May-20	918 - 00:00	13	. 16
NaN 907	27-Jun-2019	- 00:00	28-Jun-20	919 - 00:00	45	.49
NaN 940	30-Jul-2019	- 00:00	31-Jul-20	919 - 00:00	14	.59
NaN 944	03-Aug-2019	- 00:00	04 - Aug - 20	919 - 00:00	93	. 28
NaN 1031 NaN	29-0ct-2019	- 00:00	30-0ct-20	019 - 00:00	19	.56
180 249 250 318 341 450 451 452 485 486 490 491 492 493 907 940 944 1031	NO (ug/m3) 27.48 16.52 7.39 7.71 8.72 4.77 7.38 4.10 27.83 25.57 27.44 29.45 26.82 25.54 27.68 10.31 8.77 9.13 71.75	67 67 39 39 40 38 43 41 134 90 49 53 48 45 48 41 43 39	. 25 . 18 . 90 . 89 . 23 . 92 . 96 . 76 . 25 . 53 . 20 . 70 . 68 . 75 . 40 . 29 . 21	(ppb) NH3 NaN NaN 24.58 25.51 16.36 20.62 17.44 35.59 52.20 43.73 47.02 42.86 40.43 43.53 51.68 52.05 48.27	(ug/m3) S02 NaN NaN NaN NaN NaN NaN NaN NaN NaN Na	(ug/m3) \ 2.89 3.19 3.54 19.52 12.88 5.31 4.19 4.44 3.61 4.42 4.21 4.70 32.59 7.09 6.81 3.78 4.23 4.35 10.83
1031	, 1., 5	73	.50 10		HON	10.05

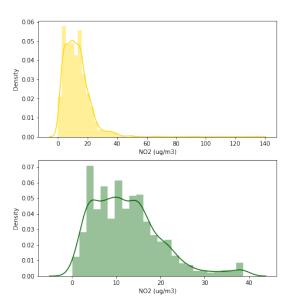
250 318 341 450 451 452 485 486 489 490 491 492 493 907 940 944	/m3) Ozone NaN 2.09 2.06 NaN 0.00 0.88 1.05 0.60 0.74 0.66 0.72 0.57 0.61 0.71 1.21 0.84 0.79 0.74 0.57	(ug/m3) 27.53 22.56 14.14 23.73 46.45 8.53 8.64 16.80 42.12 57.81 24.28 30.71 45.60 23.26 30.99 19.44 38.56 30.13 26.30	Benzene (ug/m3) 1.11 0.05 0.03 0.42 2.27 1.52 0.85 0.52 0.74 1.36 1.45 1.49 1.79 0.00 1.74 2.37 0.00	Toluene	(ug/m3) 3.06 2.47 1.93 3.51 5.31 1.58 0.68 0.03 1.18 2.22 2.49 2.35 2.26 2.40 2.79 6.53 3.30 5.00 2.04
<pre>#Rows excluding outliers values in NO2 column new_df = df10[(df10['N02 (ug/m3)'] < upper_limit) & (df10['N02 (ug/m3)'] > lower_limit)] new_df.head()</pre>						
((2))	From Date		To Dat	e PM2.	5 (ug/m3)	PM10
	17 - 00:00	02-Jan-20	17 - 00:0	0	32.61	
	17 - 00:00	03-Jan-20	17 - 00:0	0	22.93	
	17 - 00:00	04-Jan-20	17 - 00:0	0	24.19	
	17 - 00:00	05-Jan-20	17 - 00:0	0	33.61	
NaN	17 00.00	06 1 20	17 00.0	0	120 20	

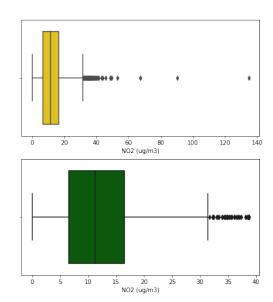
4 NaN		- 00:00 06-	Jan-2017 -	00:00	129.38	
,	NO (ug/m3)	NO2 (ug/m3)	NOx (ppb)	NH3 (ug/m3)	S02 (ug/m3)	CO
	g/m3) \	0.70	MaN	N - N	2 11	
0	2.36	9.78	NaN	NaN	2.11	
NaN 1	2.33	8.21	NaN	NaN	2.86	
NaN	J					
2	11.39	17.28	NaN	NaN	7.73	
NaN						
3	6.06	12.32	NaN	NaN	2.72	
NaN	J					

```
5.58
                     12.67
                                   NaN
                                                             2.65
4
                                                 NaN
NaN
   Ozone (uq/m3)
                  Benzene (ug/m3)
                                    Toluene (ug/m3)
0
           24.63
                              0.52
                                                2.95
1
           20.49
                              0.13
                                                2.01
2
           13.04
                              0.45
                                                3.52
3
           19.42
                              0.65
                                                3.98
4
           25.89
                              0.60
                                                3.53
#CAPPING --> In this technique, we cap our outliers data and make the
limit (we replace the oultier values with upper limit and lower limit
respectively)
new df cap = df10.copy()
new df cap['N02 (ug/m3)'] = np.where( new_df_cap['N02 (ug/m3)'] >
upper limit, upper limit,
                                      np.where(new df cap['N02
(ug/m3)'] < lower limit, lower limit, new df cap['NO2 (ug/m3)']))</pre>
df10['N02 (ug/m3)'].describe()
count
         1843.000000
           12.557347
mean
            8.707426
std
min
            0.020000
            6.510000
25%
50%
           11.280000
75%
           16.470000
          134.760000
max
Name: NO2 (ug/m3), dtype: float64
new df cap['N02 (ug/m3)'].describe()
#It is evident the maximum value has come nearer to the mean of the
data. and there are slight variations in statistical parameters after
handling outliers.
         1843.000000
count
           12.401293
mean
std
            7.738514
min
            0.020000
25%
            6.510000
50%
           11.280000
75%
           16.470000
           38.679623
max
Name: NO2 (ug/m3), dtype: float64
import warnings
warnings.filterwarnings('ignore')
plt.figure(figsize=(16,8))
plt.subplot(2,2,1)
sns.distplot(df10['N02 (ug/m3)'],color='gold' )
```

```
plt.subplot(2,2,2)
sns.boxplot(df10['N02 (ug/m3)'],color='gold' )
plt.subplot(2,2,3)
sns.distplot(new_df_cap['N02 (ug/m3)'],color='darkgreen')
plt.subplot(2,2,4)
sns.boxplot(new_df_cap['N02 (ug/m3)'],color='darkgreen')
plt.show()
```

from histogram we can see that after handling outliers using capping technique, the data points have become normally distributed.
The box plot before and after removal of ouliers, the range of values have reduced from 0-140 to 0-40.





METHOD 2

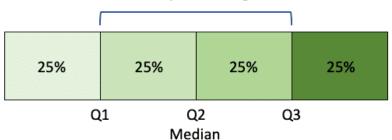
IQR based filtering

For Skewed data distribution.

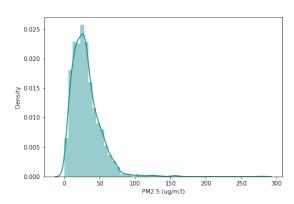
Ouliers are:

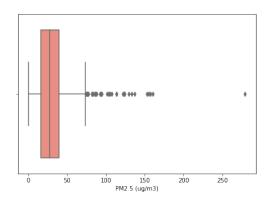
greater than 75th percentile + 1.5 IQR less than the 25th percentile – 1.5 IQR

Interquartile range



<matplotlib.axes._subplots.AxesSubplot at 0x7fad80564f90>





#Inter quatile range is an statistical parameter which helps in identifying the outliers in the dataset.
Q1+1.5*IQR and Q3-1.5*IQR are considered as outlier in this method, where Q1 and Q2 are lower quartile (25%) and upper quartile(75%).

```
Q1 = df11['PM2.5 (ug/m3)'].quantile(0.25)
Q3 = df11['PM2.5 (ug/m3)'].quantile(0.75)
IQR=Q3-Q1

upper_limit = Q1 + 1.5 * IQR
lower_limit = Q3 - 1.5 * IQR
print(upper_limit)
print(lower_limit)
```

12

14

NaN

NaN

30.94

45.14

6.60

1.00

4.35

2.82

#rows with outliers in PM 2.5 column df11[(df11['PM2.5 (ug/m3)'] > upper limit) | (df11['PM2.5 (ug/m3)'] < lower limit)] From Date To Date PM2.5 (ug/m3) PM10 (ug/m3)05-Jan-2017 - 00:00 06-Jan-2017 - 00:00 129.38 NaN 5 06-Jan-2017 - 00:00 07-Jan-2017 - 00:00 64.52 NaN 13-Jan-2017 - 00:00 14-Jan-2017 - 00:00 12 103.51 NaN 14 15-Jan-2017 - 00:00 16-Jan-2017 - 00:00 54.93 NaN 19 20-Jan-2017 - 00:00 21-Jan-2017 - 00:00 68.26 NaN . . . 1798 04-Dec-2021 - 00:00 05-Dec-2021 - 00:00 82.82 24.72 1799 05-Dec-2021 - 00:00 06-Dec-2021 - 00:00 88.01 24.56 1801 07-Dec-2021 - 00:00 08-Dec-2021 - 00:00 74.48 82.86 1812 18-Dec-2021 - 00:00 19-Dec-2021 - 00:00 59.56 43.90 1876 20-Feb-2022 - 00:00 21-Feb-2022 - 00:00 55.46 100.33 NO (ug/m3) NO2 (ug/m3) NOx (ppb) NH3 (ug/m3) S02 (ug/m3) 5.58 12.67 2.65 4 NaN NaN 5 6.91 14.38 NaN 2.28 NaN 12 32.73 38.26 NaN NaN 2.65 14 16.29 20.86 NaN NaN 3.54 19 2.86 18.64 NaN NaN 4.17 14.48 27.72 13.23 18.35 9.35 1798 14.46 9.39 1799 14.49 28.94 18.53 1801 15.32 14.44 29.76 18.71 9.41 1812 12.05 14.57 23.76 17.74 9.38 1876 4.36 15.48 19.85 17.17 19.95 CO (mg/m3) Ozone (ug/m3) Benzene (ug/m3) Toluene (ug/m3) 4 NaN 25.89 0.60 3.53 5 25.16 0.74 3.80 NaN

19	NaN	62.39	0.26	2.05
1798	0.82	39.26	0.00	0.00
1799	0.82	23.70	0.00	0.00
1801	0.87	9.26	0.00	0.00
1812	0.87	31.00	0.00	0.00
1876	0.94	100.45	0.00	0.00

[279 rows x 13 columns]

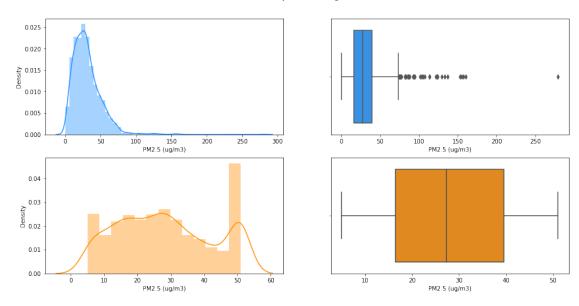
#dataset excluding the rows which have outliers in PM 2.5 column
df11[(df11['PM2.5 (ug/m3)'] < upper_limit) & (df11['PM2.5 (ug/m3)'] >
lower limit)]

lower_	_limit)]			
/ · · · · · / · · · ·		rom Date	To Date	PM2.5 (ug/m3) PM10
(ug/m3 0 NaN	01-Jan-2017	- 00:00 02-Ja	n-2017 - 00:00	32.61
NaN 1 NaN	02-Jan-2017	- 00:00 03-Ja	n-2017 - 00:00	22.93
NaN 2 NaN	03-Jan-2017	- 00:00 04-Ja	n-2017 - 00:00	24.19
3 NaN	04-Jan-2017	- 00:00 05-Ja	n-2017 - 00:00	33.61
6 NaN	07-Jan-2017	- 00:00 08-Ja	n-2017 - 00:00	45.01
1880 64.31	24-Feb-2022	- 00:00 25-Fe	b-2022 - 00:00	32.05
1881 74.92	25-Feb-2022	- 00:00 26-Fe	b-2022 - 00:00	38.95
1882 74.07	26-Feb-2022	- 00:00 27-Fe	b-2022 - 00:00	38.40
1883 57.33	27-Feb-2022	- 00:00 28-Fe	b-2022 - 00:00	27.51
1884 68.20	28-Feb-2022	- 00:00 01-Ma	r-2022 - 00:00	34.58
0 1 2 3 6	NO (ug/m3) 2.36 2.33 11.39 6.06 5.72	N02 (ug/m3) N 9.78 8.21 17.28 12.32 9.66	Ox (ppb) NH3 (NaN NaN NaN NaN NaN	ug/m3) S02 (ug/m3) \ NaN 2.11 NaN 2.86 NaN 7.73 NaN 2.72 NaN 2.19
1880 1881 1882 1883	6.21 11.21 NaN NaN	11.89 14.10 NaN NaN	18.10 25.32 NaN NaN	18.37 20.30 27.52 20.64 NaN 21.26 NaN 21.77

```
1884
            4.09
                         13.32
                                     17.40
                                                   17.40
                                                                 18.81
      CO (mq/m3)
                   Ozone (ug/m3)
                                   Benzene (ug/m3)
                                                     Toluene (ug/m3)
0
             NaN
                           24.63
                                              0.52
                                                                 2.95
1
             NaN
                           20.49
                                              0.13
                                                                 2.01
2
                           13.04
                                              0.45
                                                                 3.52
             NaN
3
             NaN
                           19.42
                                              0.65
                                                                 3.98
6
             NaN
                           35.32
                                              0.36
                                                                 3.20
              . . .
                                                . . .
                                                                  . . .
1880
            1.22
                          103.20
                                              0.00
                                                                 0.00
                                              0.00
1881
            1.13
                           46.97
                                                                 0.00
1882
            0.95
                           63.83
                                              0.00
                                                                 0.00
1883
            1.05
                           84.68
                                              0.00
                                                                 0.00
1884
            1.41
                           42.84
                                              2.08
                                                                 2.17
[1562 rows x 13 columns]
#capping the oultiers with the maximum and minimum
new df cap = df11.copy()
new df cap['PM2.5 (ug/m3)'] = np.where(new df cap['PM2.5 (ug/m3)'] >
upper limit, upper limit,
    np.where(new df cap['PM2.5 (ug/m3)'] <</pre>
lower limit,lower limit,new df cap['PM2.5 (ug/m3)']))
df11['PM2.5 (ug/m3)'].describe()
         1841.000000
count
           30.500435
mean
std
           20.289850
min
            0.410000
25%
           16.540000
50%
           27,280000
75%
           39.530000
max
          278.970000
Name: PM2.5 (ug/m3), dtype: float64
new df cap['PM2.5 (ug/m3)'].describe()
#The maximum value is deviated towards the maen unlike before capping.
count
         1841.000000
           28.310193
mean
           14.220360
std
min
            5.045000
25%
           16.540000
50%
           27.280000
75%
           39.530000
max
           51.025000
Name: PM2.5 (ug/m3), dtype: float64
plt.figure(figsize=(16,8))
plt.subplot(2,2,1)
```

```
sns.distplot(df11['PM2.5 (ug/m3)'],color='dodgerblue' )
plt.subplot(2,2,2)
sns.boxplot(df11['PM2.5 (ug/m3)'],color='dodgerblue' )
plt.subplot(2,2,3)
sns.distplot(new_df_cap['PM2.5 (ug/m3)'],color='darkorange')
plt.subplot(2,2,4)
sns.boxplot(new_df_cap['PM2.5 (ug/m3)'],color='darkorange')
plt.show()
```

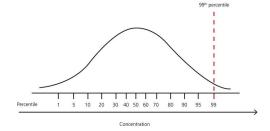
data distribution have change from positively skewed to normal distribution and ouliers are completely removed.



METHOD 3

Percentile

- This technique works by setting a particular threshold value(user defined, or domain specific).
- While capping, we use a method is known as Winsorization.
- Symmetry is maintained on both sides means if remove 1% from the right then in the left we also drop by 1%.
- --For Other distributions: Use percentile-based approach.

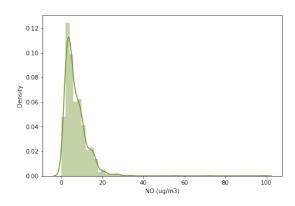


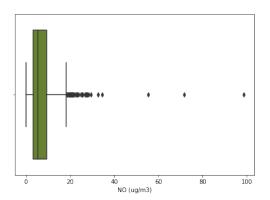
```
df12 = pd.read_excel (r'/content/Data 2017-2022.xlsx',
sheet_name='Chennai')

plt.figure(figsize=(16,5))
plt.subplot(1,2,1)
sns.distplot(df12['N0 (ug/m3)'],color='olivedrab')
plt.subplot(1,2,2)
sns.boxplot(df12['N0 (ug/m3)'],color='olivedrab')
```

#NO values are skewed and outliers are present

<matplotlib.axes._subplots.AxesSubplot at 0x7fad8051d7d0>





```
upper_limit = df12['N0 (ug/m3)'].quantile(0.90)
lower_limit = df12['N0 (ug/m3)'].quantile(0.10)
print(upper_limit,lower_limit)
```

unlike inter quartile range, the percentage value which is considered to be outlier is not fixed. # But same percent of data points are detected as ouliers in both extremes.

13.881000000000002 2.07

#rows with outliers
df12[(df12['N0 (ug/m3)'] > upper_limit) | (df12['N0 (ug/m3)'] <
lower_limit)]</pre>

	From	Date		To Date	PM2.5 (ug/m3)	PM10
(ug/mi	3) \				_	
11	12-Jan-2017 - 0	0:00 13	-Jan-2017	- 00:00	42.32	
NaN						
12	13-Jan-2017 - 0	0:00 14	-Jan-2017	- 00:00	103.51	
NaN						
13	14-Jan-2017 - 0	0:00 15	-Jan-2017	- 00:00	33.96	
NaN						
14	15-Jan-2017 - 0	0:00 16	-Jan-2017	- 00:00	54.93	
NaN						
54	24-Feb-2017 - 0	0:00 25	-Feb-2017	- 00:00	21.08	
NaN						

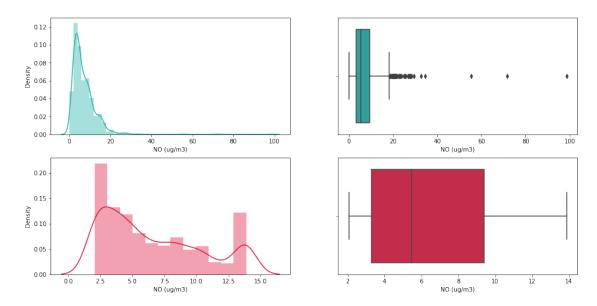
```
. . .
1820
      26-Dec-2021 - 00:00 27-Dec-2021 - 00:00
                                                            18.97
33.59
      27-Dec-2021 - 00:00 28-Dec-2021 - 00:00
1821
                                                            17.11
30.60
      28-Dec-2021 - 00:00 29-Dec-2021 - 00:00
1822
                                                            18.40
32.68
1826
      01-Jan-2022 - 00:00 02-Jan-2022 - 00:00
                                                            16.89
30.25
1853
      28-Jan-2022 - 00:00 29-Jan-2022 - 00:00
                                                            14.52
26.41
                                            NH3 (ug/m3)
      NO (ug/m3)
                   N02 (ug/m3)
                                 NOx (ppb)
                                                           S02 (ug/m3)
11
           22.63
                         26.31
                                       NaN
                                                     NaN
                                                                  2.57
12
           32.73
                         38.26
                                       NaN
                                                     NaN
                                                                  2.65
13
           24.83
                         27.27
                                                                  2.48
                                       NaN
                                                     NaN
14
           16.29
                         20.86
                                                                  3.54
                                       NaN
                                                     NaN
                         37.35
54
           26.66
                                       NaN
                                                                 22.53
                                                     NaN
                                        . . .
           16.36
                         20.26
                                     33.89
1820
                                                   21.71
                                                                  9.41
                         19.33
                                                   22.65
                                                                  9.41
1821
           23.86
                                     40.42
           16.55
                         19.39
                                                   21.86
1822
                                     35.94
                                                                  9.39
1826
           14.07
                         14.47
                                     28.54
                                                   18.48
                                                                  9.39
1853
                          5.11
                                      7.13
                                                    8.41
            2.03
                                                                 13.75
      CO (mq/m3)
                   Ozone (ug/m3)
                                   Benzene (ug/m3) Toluene (ug/m3)
11
              NaN
                            30.78
                                               0.74
                                                                 3.52
12
             NaN
                            30.94
                                               6.60
                                                                 4.35
13
             NaN
                           44.53
                                               0.23
                                                                 0.77
14
             NaN
                           45.14
                                               1.00
                                                                 2.82
54
                            34.30
                                               1.64
                                                                 2.09
             NaN
              . . .
. . .
                              . . .
                                                . . .
                                                                  . . .
            0.84
                            25.79
                                               0.00
                                                                 0.00
1820
            0.76
                                               0.00
1821
                            24.77
                                                                 0.00
1822
            0.78
                            21.34
                                               0.00
                                                                 0.00
1826
            0.78
                           36.92
                                               0.00
                                                                 0.00
1853
            0.77
                           47.45
                                               0.00
                                                                 0.00
[369 rows x 13 columns]
#rows excluding outliers
df12[(df12['N0 (ug/m3)'] \le upper limit) & (df12['N0 (ug/m3)'] >=
lower limit)]
                 From Date
                                         To Date PM2.5 (ug/m3)
                                                                   PM10
(ug/m3)
      01-Jan-2017 - 00:00 02-Jan-2017 - 00:00
                                                            32.61
NaN
      02-Jan-2017 - 00:00 03-Jan-2017 - 00:00
                                                            22.93
1
```

AL - AL					
NaN 2 NaN	03-Jan-2017	- 00:00	04-Jan-2017 -	00:00	24.19
3	04-Jan-2017	- 00:00	05-Jan-2017 -	00:00	33.61
NaN 4	05-Jan-2017	- 00:00	06-Jan-2017 -	00:00	129.38
NaN 					
1878	22-Feb-2022	- 00:00	23-Feb-2022 -	00:00	31.25
63.08 1879	23-Feb-2022	- 00:00	24-Feb-2022 -	00:00	24.57
52.81 1880	24-Feb-2022	- 00:00	25-Feb-2022 -	00:00	32.05
64.31 1881	25-Feb-2022	- 00:00	26-Feb-2022 -	00:00	38.95
74.92 1884	28-Feb-2022	- 00:00	01-Mar-2022 -	00:00	34.58
68.20					
0 1 2 3 4	2.36 2.33 11.39 6.06 5.58	9 8 17 12 12	.32 NaN .67 NaN	NaN NaN NaN NaN NaN	2.11 2.86 7.73 2.72 2.65
1878 1879 1880 1881 1884	5.02 4.44 6.21 11.21 4.09	9 9 11 14		14.92 13.56 18.37 27.52	20.20 20.15 20.30 20.64
0 1 2 3 4		-	g/m3) Benzene 24.63 20.49 13.04 19.42 25.89		
1878 1879 1880 1881 1884	1.18 1.33 1.22 1.13 1.41	16 10	67.65 62.59 03.20 46.97 42.84	0.00 0.00 0.00 0.00 2.08	0.00 0.00 0.00 0.00 2.17

[1475 rows x 13 columns]

#capping to handle outliers
new_df_cap = df12.copy()

```
new df cap['N0 (ug/m3)'] = np.where(new df cap['N0 (ug/m3)'] >
upper limit, upper limit,
    np.where(new df cap['N0 (ug/m3)'] <</pre>
lower limit,lower limit,new df cap['N0 (ug/m3)']))
df12['NO (ug/m3)'].describe()
         1844.000000
count
mean
            6.952950
std
            5.611016
min
            0.010000
25%
            3.295000
50%
            5.460000
75%
            9,400000
max
           98.620000
Name: NO (ug/m3), dtype: float64
new df cap['NO (ug/m3)'].describe()
#removal of outliers will affect all statistical parameters
considerably.
         1844.000000
count
mean
            6.557823
std
            3.856485
            2.070000
min
25%
            3.295000
50%
            5.460000
75%
            9.400000
           13.881000
max
Name: NO (ug/m3), dtype: float64
plt.figure(figsize=(16,8))
plt.subplot(2,2,1)
sns.distplot(df12['N0 (ug/m3)'],color='lightseagreen')
plt.subplot(2,2,2)
sns.boxplot(df12['NO (ug/m3)'],color='lightseagreen')
plt.subplot(2,2,3)
sns.distplot(new_df_cap['NO (ug/m3)'],color='crimson')
plt.subplot(2,2,4)
sns.boxplot(new df cap['N0 (ug/m3)'],color='crimson')
plt.show()
#graphically representation gives a clear view of how outlier analysis
can improve the distribution of data points.
```



There are many more advanced ways to handle outliers. But among these 3 methods, IQR seems to be the most effective one, though it changes from one analysis to analysis.

Outliers, unless they are data entry errors, are always an important part of a data set.

Figuring out why they are important is challenging and needs lot of analytical and lo. Then you have to figure out what to do about them.

REFERENCES:- https://www.analyticsvidhya.com/blog/2021/05/feature-engineering-how-to-detect-and-remove-outliers-with-python-code/

https://www.geeksforgeeks.org/detect-and-remove-the-outliers-using-python/

https://github.com/atulpatelDS/Youtube/blob/main/Data_Cleaning/Hands-on %20Handling%20missing%20value%20with%20List%20%26%20Pairwise%20Deletion %20with%20Python%20-%20Data%20Cleaning%20Tutorial%206.ipynb

https://www.analyticsvidhya.com/blog/2021/10/a-complete-guide-to-dealing-with-missing-values-in-python/

https://towardsdatascience.com/missing-data-and-imputation-89e9889268c8

https://www.geeksforgeeks.org/python-pandas-dataframe-ffill/https://www.geeksforgeeks.org/python-pandas-dataframe-ffill/



	mple(5) M2.5 (ug/m3) PN 27.41 55.18 65.12 29.34 24.08	NaN NaN NaN NaN NaN NaN 69.24	9.10 11.07 19.70 9.31 5.40	3.20 10.19 24.90 4.73 9.09	10x (ppb) N 11.72 21.07 25.75 13.39 14.35	NaN NaN NaN NaN NaN NaN 17.36	5.86 7.09 4.60 7.31 5.37	O(mg/m3) Ozo 0.29 1.10 0.68 0.97 0.70	37.80 53.32 25.59 23.58 9.69	0.01 0.87 0.00 0.11 0.00	e (ug/m3) 0.00 0.00 3.94 0.00 0.00		
	11)												
Data CC # CC O PN 1 PN 2 NC 3 NC 4 NC 5 NH 6 SC	olumns (total olumn M2.5 (ug/m3) M10 (ug/m3) O (ug/m3) O2 (ug/m3) Ox (ppb) H3 (ug/m3)	11 columns Non-Null 1841 non 306 non- 1844 non 1843 non 1598 non 306 non-): Count Dty	oat64 oat64 oat64 oat64 oat64 oat64 oat64									
8 02 9 Be 10 To dtypes memory df.isr) 1859 non KB 44	-null flo	oat64 oat64 oat64 oat64									
	/m3) g/m3) g/m3) g/m3) g/m3) m3) (ug/m3) e (ug/m3) e (ug/m3)	1579 41 42 287 1579 55 198 52 26 26											
#helps # Both import plt.fi msno.h	s in visualiza n NO and NO2 I missingno as igure(figsizes neatmap(df)	nas strong _l											
PM2	e size 1080x1 2.5 (ug/m3) 110 (ug/m3)	0.1	Axes>									- 1.00 - 0.75	
	NO (ug/m3) O2 (ug/m3) NOx (ppb)	0.9	0.2	<1 0.2	0.2							- 0.50 - 0.25	
S	H3 (ug/m3) O2 (ug/m3) CO(mg/m3)	0.1	0.1	0.1	0.1	0.2	0.1	0.1				- 0.00 0.25	
Benze	ne (ug/m3) ne (ug/m3) ne (ug/m3)	0.6 0.5 0.5		0.7 0.5 0.5	0.7 0.5 0.5	0.1		0.6 0.5 0.5	0.2 0).5).5 1		0.50 0.75 1.00	
1.Se	_{RM2.5} (v				alm3) MC	* (DDD) WH3	1g/m ³ 1	olmo Colmo	Ozone lugim3)	Zene lugim31	(ug/m ³)		
test_c)(mg/m3)"].	isna()]				O2 (ug/m3) C	O(mg/m3) Ozo NaN	ne (ug/m3) Benzer 24.63	ne (ug/m3) Toluen 0.52	e (ug/m3) 2.95		
1 2 3 4 	22.93 24.19 33.61 129.38 NaN	NaN NaN NaN NaN 	2.33 11.39 6.06 5.58 	8.21 17.28 12.32 12.67 NaN	NaN NaN NaN NaN 	NaN NaN NaN NaN 	2.86 7.73 2.72 2.65 NaN	NaN NaN NaN NaN 	20.49 13.04 19.42 25.89 NaN	0.13 0.45 0.65 0.60 	2.01 3.52 3.98 3.53 NaN		
	NaN 35.22 48.23 NaN 5 × 11 columns	NaN NaN NaN NaN	NaN 6.74 7.42 NaN	NaN 3.61 3.24 NaN	9.51 9.79 NaN	NaN NaN NaN NaN	NaN 4.27 4.48 NaN	NaN NaN NaN NaN	NaN 31.88 38.56 NaN	NaN 4.14 11.22 NaN	NaN 0.00 0.26 NaN		
#dropp df.dro	oping the poing all the popna (inplace=	rows with n	ull values	(in any coi	lumn) fron			O(mg/m3) Ozo	ne (ug/m3) Benzer	ne (ug/m3) Toluen	e (ug/m3)		
1576 1577 1578 1580 1584 	7.34 5.08 5.84 8.28 5.40 	60.71 56.41 56.20 56.20 56.20 	1.77 1.81 1.70 1.72 1.76 5.02	3.74 3.74 3.73 3.74 3.75 9.97	5.51 5.54 5.43 5.46 5.51 	5.36 5.36 5.36 5.36 5.36 	8.00 6.61 5.89 5.72 5.37 	0.50 0.51 0.57 0.53 0.54 	12.44 27.50 19.75 25.36 14.95 	0.00 0.00 0.00 0.00 0.00 0.00	0.00 0.00 0.00 0.00 0.00 0.00		
1879 1880 1881 1884 286 rows	24.57 32.05 38.95 34.58 3 × 11 columns	52.81 64.31 74.92 68.20	5.02 4.44 6.21 11.21 4.09	9.97 9.82 11.89 14.10 13.32	14.98 14.25 18.10 25.32 17.40	14.92 13.56 18.37 27.52 17.40	20.20 20.15 20.30 20.64 18.81	1.18 1.33 1.22 1.13 1.41	167.65 162.59 103.20 46.97 42.84	0.00 0.00 0.00 0.00 2.08	0.00 0.00 0.00 0.00 2.17		
PM2.5 (PM10 (UQN02 (UQN0X (PPN3 (UQS02 (UQS0	ug/m3) /m3) g/m3) ob) g/m3) g/m3)												
CO(mg/r Ozone (Benzene Toluene dtype:	m3) (ug/m3) e (ug/m3) e (ug/m3) int64 e has drastica)))	d from (188	85,11) to ((286, 11)								
3.Cr #y_tra # y_tr #Trair	eate "X_ ain is the col- rain contains n data will ha in=df["CO(mg/r	lumn which of completely ave only roo	we are choo filled dat	osing for a	imputatior		ıll values						
y_trai 1576 1577 1578 1580 1584 1878 1878	0.50 0.51 0.57 0.53 0.54 1.18 1.33												
	1.22 1.13 1.41 CO(mg/m3), Lei	ngth: 286,	dtype: floa	at64									
x_trai x_trai P1 1576	M2.5 (ug/m3) PM 7.34	(mg/m3)", a. 10 (ug/m3) N 60.71	xis=1) O (ug/m3) No	O2 (ug/m3) N 3.74	IOx (ppb) N 5.51	I H3 (ug/m3) S 6 5.36	O2 (ug/m3) O 8.00	zone (ug/m3) E 12.44	Benzene (ug/m3) To 0.00	0.00			
1577 1578 1580 1584 1878	5.08 5.84 8.28 5.40 31.25 24.57	56.41 56.20 56.20 56.20 63.08 52.81	1.81 1.70 1.72 1.76 5.02 4.44	3.74 3.73 3.74 3.75 9.97 9.82	5.54 5.43 5.46 5.51 14.98 14.25	5.36 5.36 5.36 5.36 14.92 13.56	6.61 5.89 5.72 5.37 20.20 20.15	27.50 19.75 25.36 14.95 167.65 162.59	0.00 0.00 0.00 0.00 0.00	0.00 0.00 0.00 0.00 0.00			
	32.05 38.95 34.58 5 × 10 columns	64.31 74.92 68.20	6.21 11.21 4.09	11.89 14.10 13.32	18.10 25.32 17.40	18.37 27.52 17.40	20.30 20.64 18.81	103.20 46.97 42.84	0.00 0.00 2.08	0.00 0.00 2.17			
#simpl	uild the regression sklearn.linear	model using			in sklearr	7							
lr.fit	the model on (x_train,y_train)	ain)											
#the w x_test x_test		vill have or	nly null va m <mark>3)"</mark> ,axis=1	1)	IOx (ppb) N	I H3 (ug/m3) S o	O2 (ug/m3) O	zone (ug/m3) E	Benzene (ug/m3) To 0.52	oluene (ug/m3) 2.95			
1 2 3 4 1071	22.93 24.19 33.61 129.38 NaN	NaN NaN NaN NaN NaN	2.33 11.39 6.06 5.58 NaN	8.21 17.28 12.32 12.67 NaN	NaN NaN NaN NaN NaN	NaN NaN NaN NaN NaN	2.86 7.73 2.72 2.65 NaN	20.49 13.04 19.42 25.89 NaN	0.13 0.45 0.65 0.60 NaN	2.01 3.52 3.98 3.53 NaN			
1082 1083 1088 198 rows	35.22 48.23 NaN s × 10 columns	NaN NaN NaN	6.74 7.42 NaN	3.61 3.24 NaN	9.51 9.79 NaN	NaN NaN NaN	4.27 4.48 NaN	31.88 38.56 NaN	4.14 11.22 NaN	0.00 0.26 NaN			
#remov x_test x_test	ving columns was a column of the column of t	els='PM10 (under the contract of the contract	ug/m3)', axg/m3)', axs	xis=1) is=1) et complete	ely filled	d data)		lie '		17-7			
x=['PM x_test #final x_test PI	M2.5 (ug/m3)', [[x] = x_test [X_test data [M2.5 (ug/m3) PN 7.34	'NO (ug/m3 [x].fillna(: 10 (ug/m3) N 60.71)','NO2 (ug x_test[x].r O (ug/m3) No	g/m3)','N0> mean()) O2 (ug/m3) N 3.74	((ppb) ' , ' IOx (ppb) S 5.51	SO2 (ug/m3) SO2 (ug/m3) C	O(mg/m3) Oz 0.50	one (ug/m3) Bo	zene (ug/m3)', enzene (ug/m3) To 0.00	oluene (ug/m3) 0.00	5)']		
1577 1578 1580 1584 1878	5.08 5.84 8.28 5.40 31.25 24.57	56.41 56.20 56.20 56.20 63.08 52.81	1.81 1.70 1.72 1.76 5.02 4.44	3.74 3.73 3.74 3.75 9.97 9.82	5.54 5.43 5.46 5.51 14.98 14.25	6.61 5.89 5.72 5.37 20.20 20.15	0.51 0.57 0.53 0.54 1.18	27.50 19.75 25.36 14.95 167.65 162.59	0.00 0.00 0.00 0.00 0.00	0.00 0.00 0.00 0.00 0.00			
1880 1881 1884 286 rows	32.05 38.95 34.58 3 × 10 columns	64.31 74.92 68.20	6.21 11.21 4.09	11.89 14.10 13.32	18.10 25.32 17.40	20.30 20.64 18.81	1.22 1.13 1.41	103.20 46.97 42.84	0.00 0.00 2.08	0.00 0.00 0.00 2.17			
#predi y_pred C:\User l be ra Feature - CO(mo	d=lr.predict(x rs\Murali\ana aised. e names unsee g/m3) e names seen	conda3\lib\	site-packa me:	ges\sklearı	·					tch those that	were passed duri	ing fit. Starting	version 1.2,
- NH3 (warn: y_prec array((ug/m3) ings.warn(mes (0.59228563, 0.64499524, 0.66193982, 0.65988706,	sage, Futur 0.61962348, 0.63190067, 0.63432952, 0.64634404,	eWarning) 0.62705290 0.63662440 0.6250001 0.6324636	2, 0.63774; 6, 0.640858 , 0.615889 , 0.658900	806, 0.633 959, 0.639 925, 0.612	112603, 541253, 271557,							
	0.63625866, 0.66206673, 0.67292409, 0.66071735, 0.70354965, 0.71564759, 0.66224837, 0.64063845, 0.66360482, 0.56702375, 0.57166757,	0.69105188, 0.68017852, 0.67175487, 0.72718457, 0.69366327, 0.67582786, 0.66990378, 0.60947098, 0.56123403,	0.6507554 0.685563 0.6605690 0.7171329 0.6894826 0.6741921 0.6260146 0.5885898 0.5845780	1, 0.646498 , 0.691450 9, 0.69386 2, 0.71177 7, 0.714299 3, 0.647829 6, 0.659920 5, 0.586330 3, 0.569860	899, 0.646 925, 0.695 773, 0.676 313, 0.728 937, 0.686 902, 0.636 933, 0.656 964, 0.568 624, 0.572	660843, 66958 , 807293, 836997, 812516, 380382, 961993, 861086, 259753,							
	0.56218718, 0.57258998, 0.69412242, 0.66468178, 0.70306086, 0.68192827, 0.66780663, 0.66768133, 0.75690584, 0.785243,	0.57161908, 0.62921369, 0.68798278, 0.70016527, 0.66541113, 0.67150521, 0.66741474, 0.69072277, 0.6531067, 0.73868954, 0.74760146,	0.5412958 0.7302264 0.6587610 0.6928315 0.6666538 0.6849362 0.6821524 0.6744366 0.7142273 0.7447510 0.7177569	4, 0.56989 3, 0.787146 6, 0.67313 5, 0.68997 3, 0.67459 , 0.66357 1, 0.66303 9, 0.67774 6, 0.77152 2, 0.76296 9, 0.72639	725, 0.563 637, 0.790 302, 0.673 38, 0.673 358, 0.678 579, 0.698 789, 0.660 774, 0.689 578, 0.769 776, 0.723	344829, 030559, 752268, 222971, 398217, 302801, 010201, 541922, 512787, 368957,							
	0.785243 , 0.73765223, 0.71908346, 0.79431546, 0.72952827, 0.75532554, 0.75273629, 0.81901071, 0.79682627, 0.75488768, 0.82847624,	0.74760146, 0.72679441, 0.7338085, 0.74531671, 0.78563538, 0.73343087, 0.69539106, 0.80165038, 0.75802258, 0.7217279, 0.76712955,	0.7177569 0.8335030 0.7258644 0.7428552 0.7593341 0.7533450 0.7178178 0.7364645 0.7817267 0.7088133	9, 0.72639 4, 0.749568 7, 0.747619 5, 0.742124 3, 0.722300 8, 0.760020 1, 0.70365 8, 0.718199 3, 0.757400 9, 0.802070 8, 0.57318	743, 0.704 892, 0.733 556, 0.779 471, 0.716 601, 0.783 99, 0.743 729, 0.719 528, 0.750 647, 0.793 954, 0.883 296, 0.498	486148, 310297, 919157, 678162, 343442, 168451, 964509, 967636, 1109 , 315102,							
	0.7451032, 0.73477416, 0.85657152, 0.87014873, 0.83286882, 0.94276975, 0.81705371, 0.81254696, 0.83911844, 0.95165272, 0.8569465,	0.69928489, 0.88650758, 0.90938751, 0.85307269, 1.01267904, 0.86350447, 0.81582667, 0.80351282, 0.83569963, 0.99423786, 0.84801477,	0.7239170 0.8798663 0.8487472 0.9725941 0.8365411 0.8929687 0.8211503 0.7855827 0.8305111 0.8673681 0.8285752	2, 1.083063 4, 0.88063 3, 0.909870 3, 0.785280 4, 0.938083 4, 0.825243 4, 0.819878 8, 0.932638 5, 0.830220 1, 0.954700 7, 0.857730	301, 0.892 784, 0.843 907, 0.919 941, 0.869 168, 0.963 174, 0.818 842, 0.823 829, 0.873 939, 0.898 693, 0.936	28276 , 382784, 939003, 595344, 194497, 348902, 353856, 12541 , 382398, 603636, 554856,							
	0.8569465, 0.63938144, 0.63232597, 0.57674705, 0.59943047, 0.57682188, 0.59255419, 0.70870395, 0.88561119, 0.97919073, 0.85178601, 1.11402078])	0.63970605, 0.60121669, 0.56173981, 0.59233343, 0.60087936, 0.58005114, 0.81379321, 1.00027378, 1.00239402,	0.634049 0.5708301 0.5914808 0.5837935 0.6090365 0.5764708 0.82386609 0.8563058 0.8013299	, 0.628160 8, 0.566810 8, 0.593603 , 0.57823 , 0.601889 6, 0.603063 5, 0.859769 , 0.860130 4, 0.822513	929, 0.630 616, 0.57 357, 0.589 732, 0.581 971, 0.590 195, 0.638 926, 1.01 643, 0.89 149, 0.94	961425, 786824, 997044, 180757, 64905 , 878237, 729138, 775106,							
y_pred (286,) 7.Re	eplace th				vith pr	edicted	d value	èS					
Attribu ~\AppDa ~\anacc 5485	,	_data.CO(mg 		Tracebacl 1623517.py pred array).isnull()	k (most re in <modul ,'CO(mg/m3</modul 	ecent call i le> 3)']=y_pred	·						
5486 -> 548 6 5486 Attrib	ret defs defs dataframe is		getattri	str, value) -> None		es using li	near regress	ion				
test_c	nk you!!!												
test_c													