In [2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [3]:
```
data=pd.read_excel("C:/Users/Shivani Sharma/Desktop/data.xlsx")
data
```

Out[3]:

	country	state	city	station	lastupdate	pollutant_avg
0	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	167.0
1	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	105.0
2	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	21.0
3	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	5.0
4	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	28.0
5	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	38.0
6	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	56.0
7	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	195.0
8	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	124.0
9	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	26.0
10	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	4.0
11	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	12.0
12	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	44.0
13	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	87.0

	country	state	city	station	lastupdate	pollutant_avg
14	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	104.0
15	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	86.0
16	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	47.0
17	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	5.0
18	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	5.0
19	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	38.0
20	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	17.0
21	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	149.0
22	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	117.0
23	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	41.0
24	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	3.0
25	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	13.0
26	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	59.0
27	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	89.0
28	India	Assam	Guwahati	Railway Colony, Guwahati - APCB	13-01-2020 12:00:00	265.0
29	India	Assam	Guwahati	Railway Colony, Guwahati - APCB	13-01-2020 12:00:00	193.0
...	...	...	...	...	...	..

	country	state	city	station	lastupdate	pollutant_avg
1185	India	West_Bengal	Kolkata	Jadavpur, Kolkata - WBPCB	13-01-2020 12:00:00	66.0
1186	India	West_Bengal	Kolkata	Jadavpur, Kolkata - WBPCB	13-01-2020 12:00:00	8.0
1187	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	NaN
1188	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	NaN
1189	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	NaN
1190	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	NaN
1191	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	NaN
1192	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	67.0
1193	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	26.0
1194	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	129.0
1195	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	138.0
1196	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	53.0
1197	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	4.0
1198	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	19.0
1199	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	78.0
1200	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	22.0

	country	state	city	station	lastupdate	pollutant_avg
1201	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	197.0
1202	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	140.0
1203	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	74.0
1204	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	3.0
1205	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	9.0
1206	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	71.0
1207	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	17.0
1208	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	NaN
1209	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	106.0
1210	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	32.0
1211	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	7.0
1212	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	5.0
1213	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	39.0
1214	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	23.0

1215 rows × 11 columns

In [65]:
```
data.head(10) #print first 10 rows
```

Out[65]:

	country	state	city	station	lastupdate	pollutant_avg	p
0	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	167.0	
1	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	105.0	
2	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	21.0	
3	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	5.0	
4	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	28.0	
5	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	38.0	
6	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	56.0	
7	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	195.0	
8	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	124.0	
9	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	26.0	

In [66]:
```
data.tail(10) #printing last 10 rows
```

Out[66]:

	country	state	city	station	lastupdate	pollutant_avg	pollutant_max	polluta
1205	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	9.0	31.0	
1206	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	71.0	91.0	
1207	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	17.0	106.0	
1208	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	NaN	NaN	
1209	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	106.0	151.0	
1210	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	32.0	50.0	
1211	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	7.0	8.0	
1212	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	5.0	5.0	
1213	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	39.0	49.0	
1214	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	23.0	63.0	

In [67]:
```
data.columns #print the columns/features of the data
```

Out[67]:
```
Index(['country', 'state', 'city', 'station', 'lastupdate', 'pollutant_avg',
       'pollutant_max', 'pollutant_min', 'pollutant_id',
       'Predominant_Parameter', 'air quality index value'],
      dtype='object')
```

In [68]:
```
data.describe() #basic info of the dataset
```

Out[68]:

	pollutant_avg	pollutant_max	pollutant_min	air quality index value
count	1113.000000	1113.000000	1113.000000	1089.000000
mean	81.567835	132.238095	42.154537	220.348026
std	96.004389	140.312882	57.143337	112.401481
min	1.000000	1.000000	1.000000	29.000000
25%	16.000000	27.000000	6.000000	119.000000
50%	43.000000	87.000000	18.000000	205.000000
75%	105.000000	163.000000	55.000000	334.000000
max	416.000000	500.000000	338.000000	416.000000

In [69]:
```
data.shape #dimensions of the data
```

Out[69]:
```
(1215, 11)
```

In [103]:
```
data.isna().sum() #print the sum of null values for each columns
```

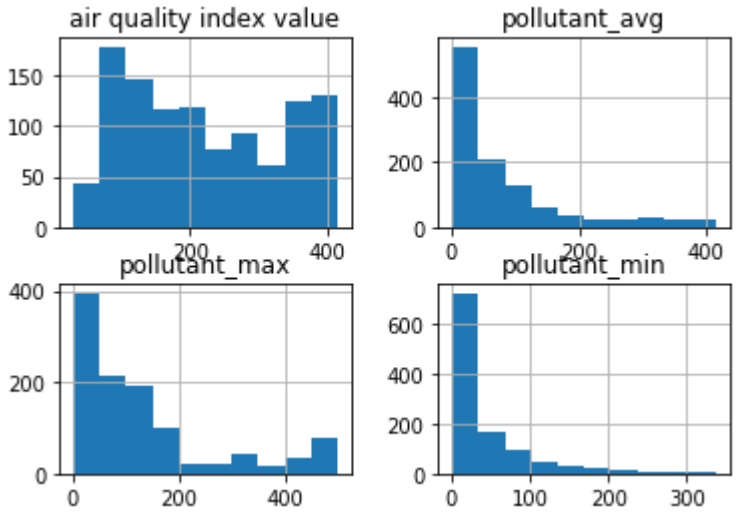
Out[103]:
```
country                      0
state                        0
city                         0
station                      0
lastupdate                   0
pollutant_avg              102
pollutant_max              102
pollutant_min              102
pollutant_id                 0
Predominant_Parameter      126
air quality index value    126
dtype: int64
```

In [109]:
```
#drop the columns that are not in use to us.
data.drop(['country','lastupdate'],axis=1,inplace=True)
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-109-f69d13ad0909> in <module>()
----> 1 data.drop(['country','lastupdate'],axis=1,inplace=True)
      2 data

G:\Anaconda3\lib\site-packages\pandas\core\frame.py in drop(self, labels, axis, index, columns, level, inplace, errors)
   3695                                           index=index, columns=columns,
   3696                                           level=level, inplace=inplace,
-> 3697                                           errors=errors)
   3698
   3699     @rewrite_axis_style_signature('mapper', [('copy', True),

G:\Anaconda3\lib\site-packages\pandas\core\generic.py in drop(self, labels, axis, index, columns, level, inplace, errors)
   3109         for axis, labels in axes.items():
   3110             if labels is not None:
-> 3111                 obj = obj._drop_axis(labels, axis, level=level, errors=errors)
   3112
   3113         if inplace:

G:\Anaconda3\lib\site-packages\pandas\core\generic.py in _drop_axis(self, labels, axis, level, errors)
   3141                 new_axis = axis.drop(labels, level=level, errors=errors)
   3142             else:
-> 3143                 new_axis = axis.drop(labels, errors=errors)
   3144             result = self.reindex(**{axis_name: new_axis})
   3145

G:\Anaconda3\lib\site-packages\pandas\core\indexes\base.py in drop(self, labels, errors)
   4402             if errors != 'ignore':
   4403                 raise KeyError(
-> 4404                     '{} not found in axis'.format(labels[mask]))
   4405             indexer = indexer[~mask]
   4406         return self.delete(indexer)

KeyError: "['country' 'lastupdate'] not found in axis"
```

In [15]:
```
data.hist()
```

Out[15]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x00000256AC738E10>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x00000256ADFEAE48
>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x00000256AE01C470>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x00000256AE044B00
>]],
      dtype=object)
```



In [110]:
```
data
```

Out[110]:

	state	city	station	pollutant_avg	pollutant_max	poll
0	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	167.0	272.0	
1	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	105.0	132.0	
2	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	21.0	51.0	
3	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	5.0	5.0	
4	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	28.0	111.0	
5	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	38.0	60.0	
6	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	56.0	78.0	
7	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	195.0	272.0	
8	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	124.0	160.0	
9	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	26.0	46.0	
10	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	4.0	4.0	
11	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	12.0	21.0	
12	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	44.0	64.0	
13	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	87.0	132.0	
14	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	104.0	201.0	
15	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	86.0	127.0	
16	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	47.0	87.0	

	state	city	station	pollutant_avg	pollutant_max	poll
17	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	5.0	6.0	
18	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	5.0	11.0	
19	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	38.0	50.0	
20	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	17.0	31.0	
21	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	149.0	223.0	
22	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	117.0	144.0	
23	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	41.0	127.0	
24	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	3.0	4.0	
25	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13.0	15.0	
26	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	59.0	92.0	
27	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	89.0	118.0	
28	Assam	Guwahati	Railway Colony, Guwahati - APCB	265.0	444.0	
29	Assam	Guwahati	Railway Colony, Guwahati - APCB	193.0	428.0	
...	...	...	...	...	...	
1185	West_Bengal	Kolkata	Jadavpur, Kolkata - WBPCB	66.0	110.0	
1186	West_Bengal	Kolkata	Jadavpur, Kolkata - WBPCB	8.0	21.0	
1187	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	NaN	NaN	
1188	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	NaN	NaN	
1189	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	NaN	NaN	

	state	city	station	pollutant_avg	pollutant_max	poll
1190	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	NaN	NaN	
1191	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	NaN	NaN	
1192	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	67.0	109.0	
1193	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	26.0	90.0	
1194	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	129.0	302.0	
1195	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	138.0	202.0	
1196	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	53.0	100.0	
1197	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	4.0	4.0	
1198	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	19.0	45.0	
1199	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	78.0	114.0	
1200	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	22.0	86.0	
1201	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	197.0	327.0	
1202	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	140.0	213.0	
1203	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	74.0	145.0	
1204	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	3.0	4.0	
1205	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	9.0	31.0	
1206	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	71.0	91.0	
1207	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	17.0	106.0	
1208	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	NaN	NaN	
1209	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	106.0	151.0	
1210	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	32.0	50.0	

	state	city	station	pollutant_avg	pollutant_max	poll
1211	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	7.0	8.0	
1212	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	5.0	5.0	
1213	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	39.0	49.0	
1214	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	23.0	63.0	

1215 rows × 9 columns

In [111]:
```
data1= pd.read_excel("C:/Users/Shivani Sharma/Desktop/data.xlsx",index_col='po
llutant_id')
data1
```

Out[111]:

pollutant_id	country	state	city	station	lastupdate	pollut
PM2.5	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	
PM10	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	
NO2	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	
NH3	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	
SO2	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	
CO	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	
OZONE	India	Andhra_Pradesh	Amaravati	Secretariat, Amaravati - APPCB	13-01-2020 12:00:00	
PM2.5	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	
PM10	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	
NO2	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	
NH3	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	
SO2	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	
CO	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	
OZONE	India	Andhra_Pradesh	Rajamahendravaram	Anand Kala Kshetram, Rajamahendravaram - APPCB	13-01-2020 12:00:00	

pollutant_id	country	state	city	station	lastupdate	pollut
PM2.5	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	
PM10	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	
NO2	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	
NH3	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	
SO2	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	
CO	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	
OZONE	India	Andhra_Pradesh	Tirupati	Tirumala, Tirupati - APPCB	13-01-2020 12:00:00	
PM2.5	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	
PM10	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	
NO2	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	
NH3	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	
SO2	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	
CO	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	
OZONE	India	Andhra_Pradesh	Visakhapatnam	GVM Corporation, Visakhapatnam - APPCB	13-01-2020 12:00:00	
PM2.5	India	Assam	Guwahati	Railway Colony, Guwahati - APCB	13-01-2020 12:00:00	
PM10	India	Assam	Guwahati	Railway Colony, Guwahati - APCB	13-01-2020 12:00:00	

pollutant_id	country	state	city	station	lastupdate	pollut
...	...	...	...	...	...	
CO	India	West_Bengal	Kolkata	Jadavpur, Kolkata - WBPCB	13-01-2020 12:00:00	
OZONE	India	West_Bengal	Kolkata	Jadavpur, Kolkata - WBPCB	13-01-2020 12:00:00	
PM2.5	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	
PM10	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	
NO2	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	
NH3	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	
SO2	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	
CO	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	
OZONE	India	West_Bengal	Kolkata	Rabindra Bharati University, Kolkata - WBPCB	13-01-2020 12:00:00	
PM2.5	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	
PM10	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	
NO2	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	
NH3	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	
SO2	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	
CO	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	

pollutant_id	country	state	city	station	lastupdate	pollut
OZONE	India	West_Bengal	Kolkata	Rabindra Sarobar, Kolkata - WBPCB	13-01-2020 12:00:00	
PM2.5	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	
PM10	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	
NO2	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	
NH3	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	
SO2	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	
CO	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	
OZONE	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	
PM2.5	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	
PM10	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	
NO2	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	
NH3	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	
SO2	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	
CO	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	
OZONE	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	

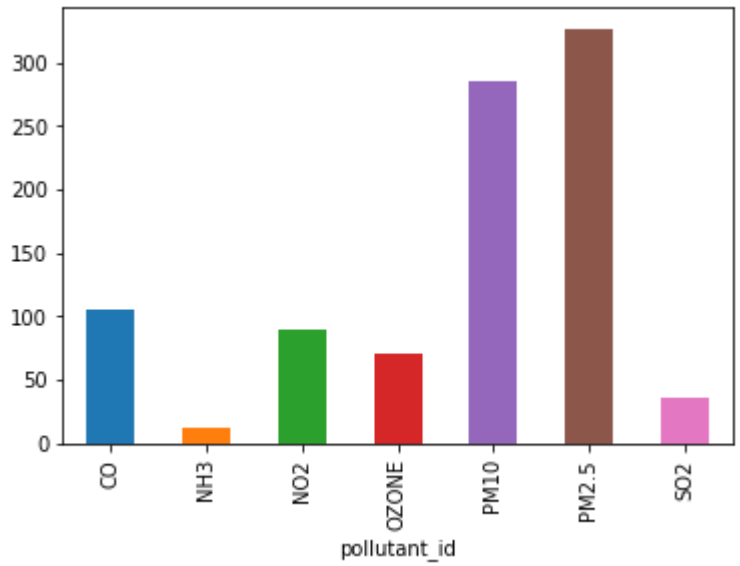
1215 rows × 10 columns

In [112]:
```
data.groupby("pollutant_id")['pollutant_max'].mean()
```

Out[112]:
```
pollutant_id
CO       104.796512
NH3       12.838235
NO2       89.170732
OZONE     70.188235
PM10     285.868056
PM2.5    326.624242
SO2       35.777778
Name: pollutant_max, dtype: float64
```

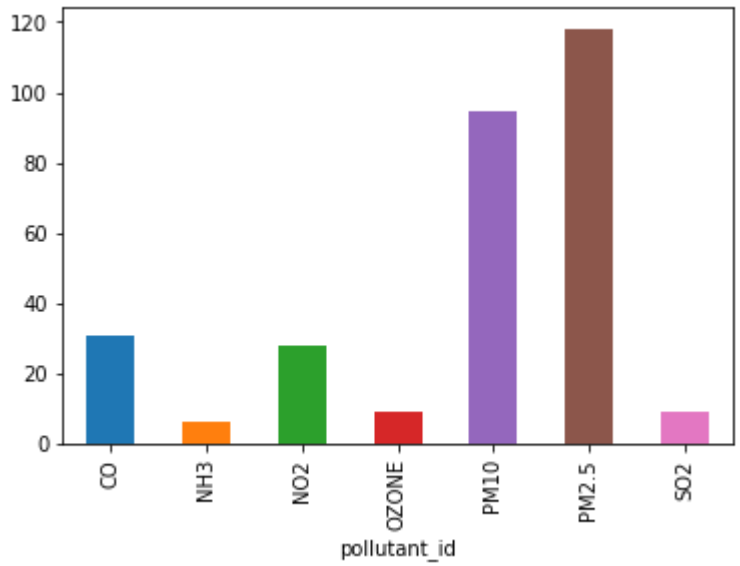
In [113]:
```
data.groupby("pollutant_id")['pollutant_max'].mean().plot(kind='bar')
```

Out[113]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1d655b0bc18>
```



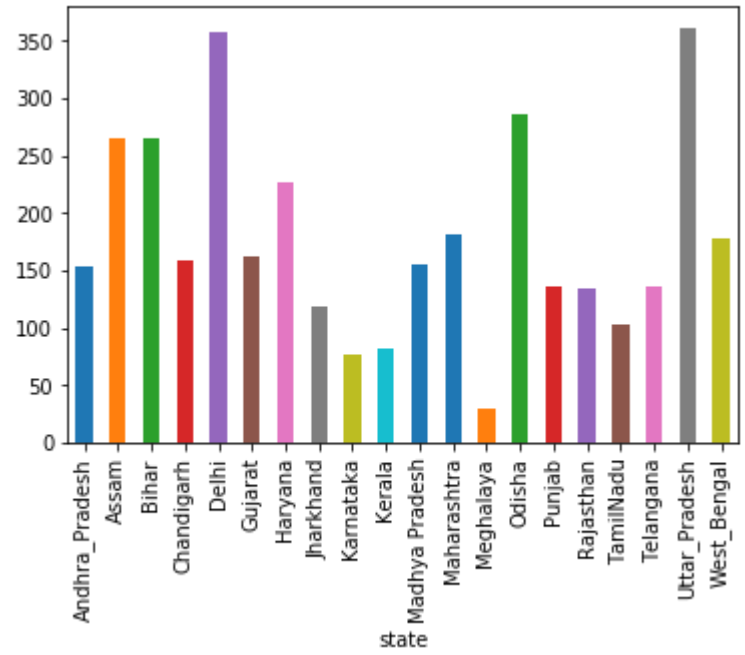
In [114]:
```
data.groupby("pollutant_id")['pollutant_min'].mean().plot(kind='bar')
```

Out[114]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1d655bbac18>
```



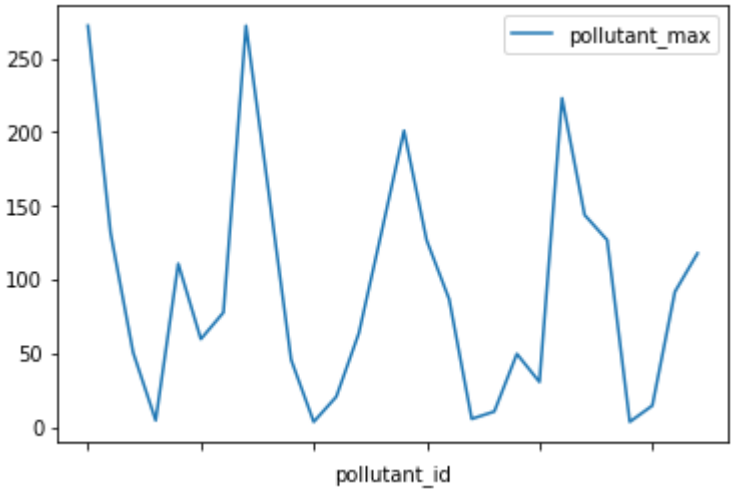
In [117]:
```
data.groupby("state")['air quality index value'].mean().plot(kind='bar')
```

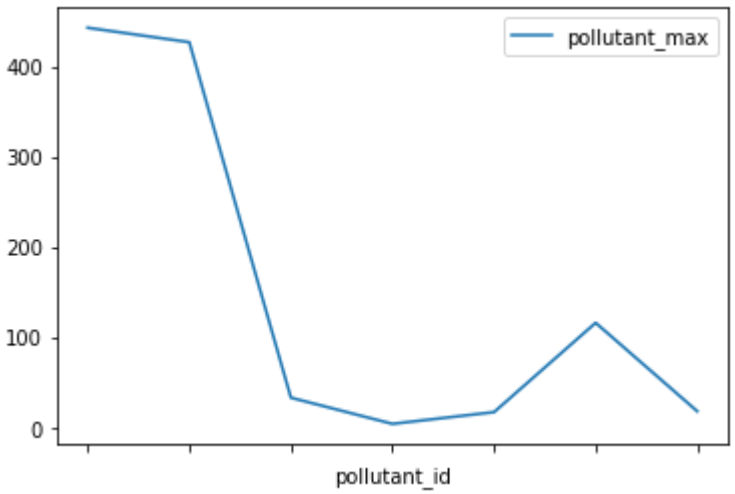
Out[117]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1d65779c7f0>
```

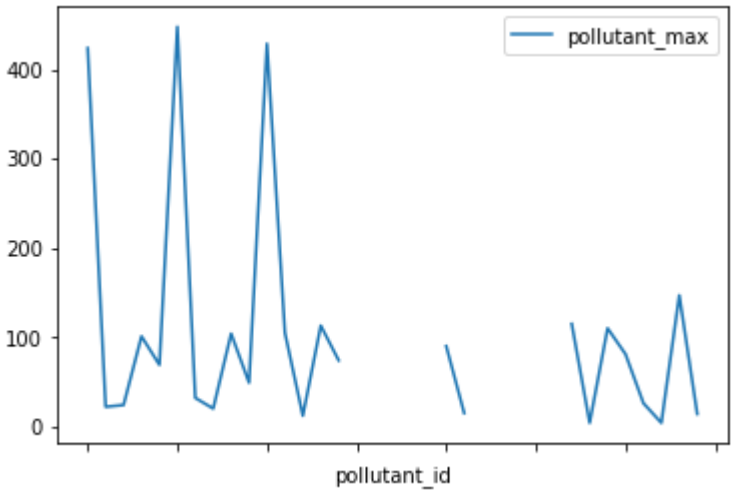


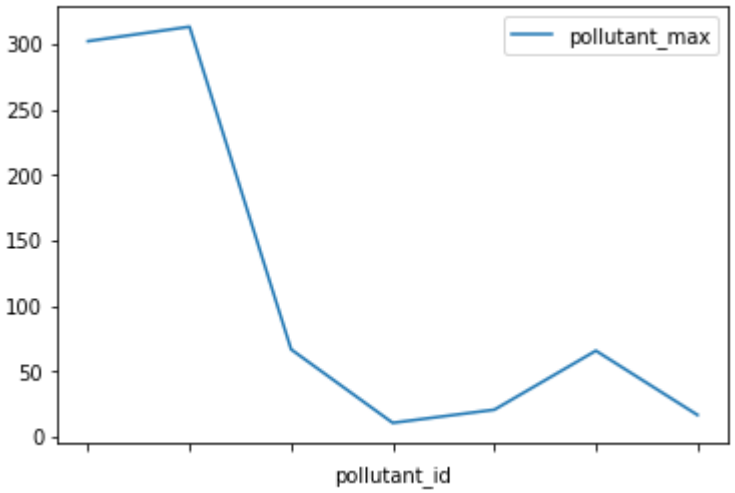
In [115]:
```
data.groupby("state").plot(x='pollutant_id',y='pollutant_max')
```

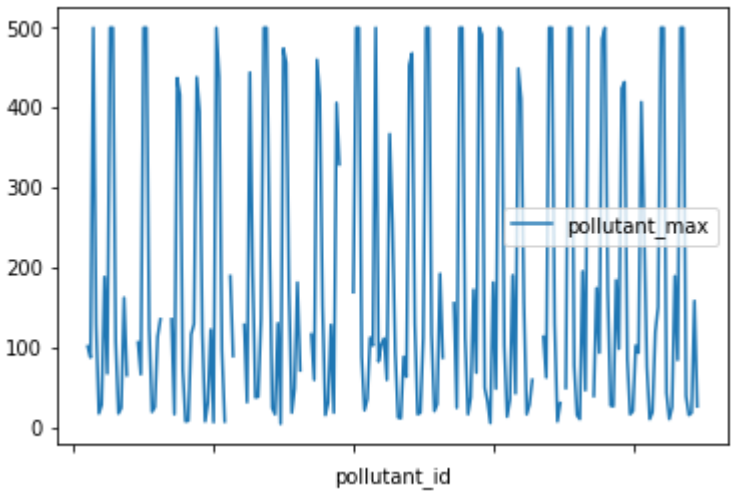
Out[115]:
```
state
Andhra_Pradesh    AxesSubplot(0.125,0.125;0.775x0.755)
Assam             AxesSubplot(0.125,0.125;0.775x0.755)
Bihar             AxesSubplot(0.125,0.125;0.775x0.755)
Chandigarh        AxesSubplot(0.125,0.125;0.775x0.755)
Delhi             AxesSubplot(0.125,0.125;0.775x0.755)
Gujarat           AxesSubplot(0.125,0.125;0.775x0.755)
Haryana           AxesSubplot(0.125,0.125;0.775x0.755)
Jharkhand         AxesSubplot(0.125,0.125;0.775x0.755)
Karnataka         AxesSubplot(0.125,0.125;0.775x0.755)
Kerala            AxesSubplot(0.125,0.125;0.775x0.755)
Madhya Pradesh    AxesSubplot(0.125,0.125;0.775x0.755)
Maharashtra       AxesSubplot(0.125,0.125;0.775x0.755)
Meghalaya         AxesSubplot(0.125,0.125;0.775x0.755)
Odisha            AxesSubplot(0.125,0.125;0.775x0.755)
Punjab            AxesSubplot(0.125,0.125;0.775x0.755)
Rajasthan         AxesSubplot(0.125,0.125;0.775x0.755)
TamilNadu         AxesSubplot(0.125,0.125;0.775x0.755)
Telangana         AxesSubplot(0.125,0.125;0.775x0.755)
Uttar_Pradesh     AxesSubplot(0.125,0.125;0.775x0.755)
West_Bengal       AxesSubplot(0.125,0.125;0.775x0.755)
dtype: object
```

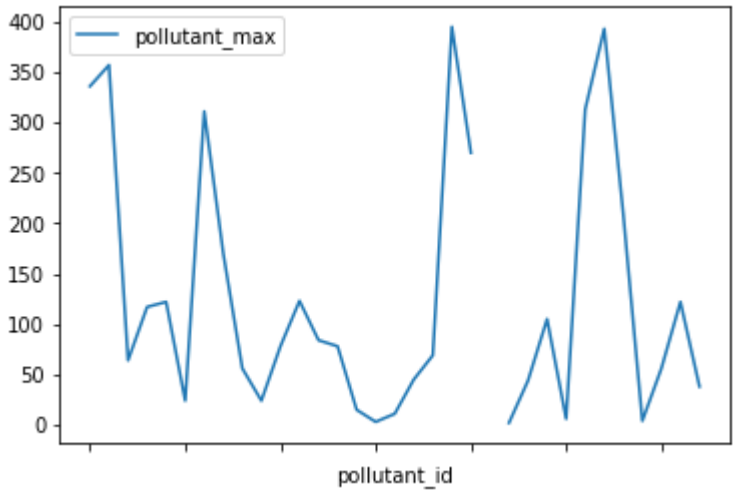


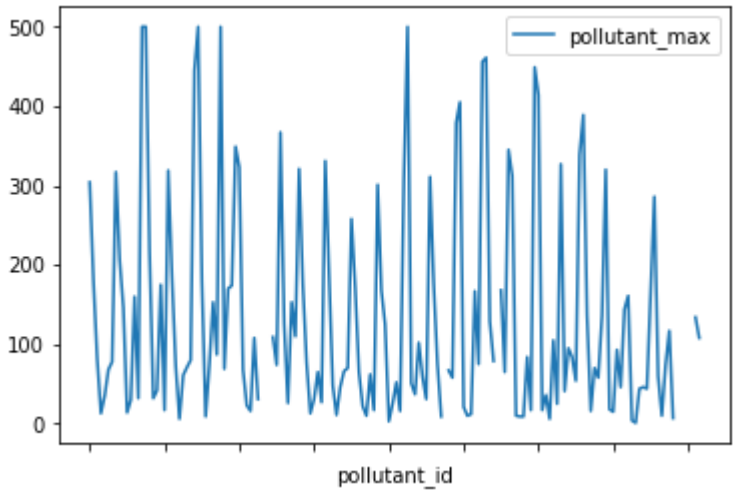


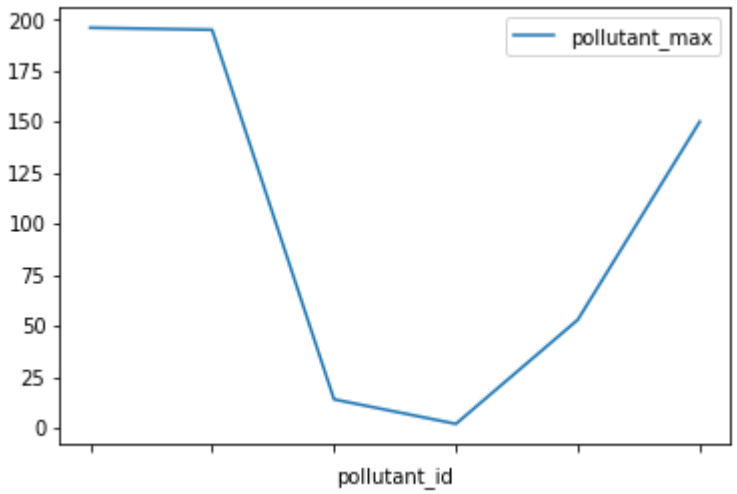

pollutant_max
400
300
200
100
0
pollutant_id

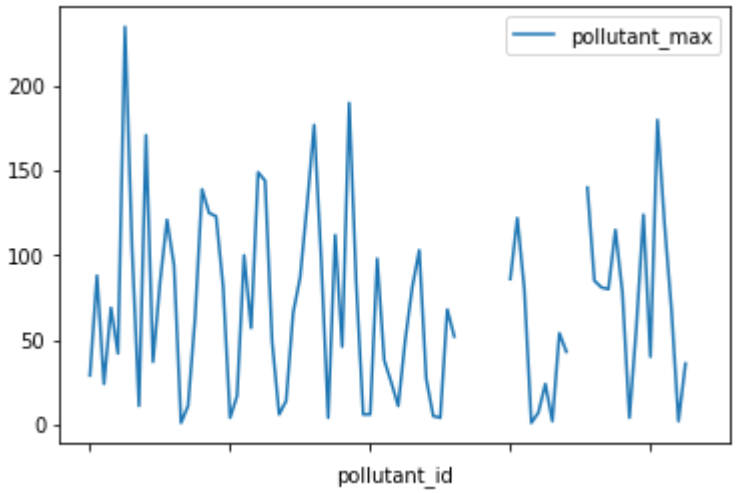

pollutant_max
300
250
200
150
100
50
0
pollutant_id

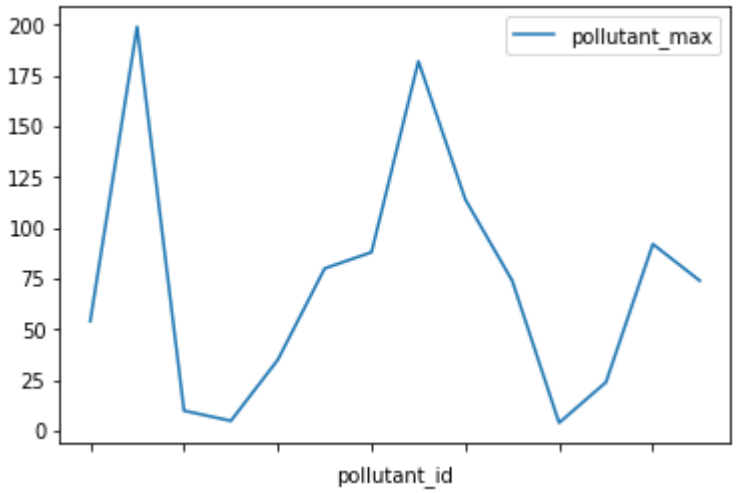

pollutant_max
500
400
300
200
100
0
pollutant_id

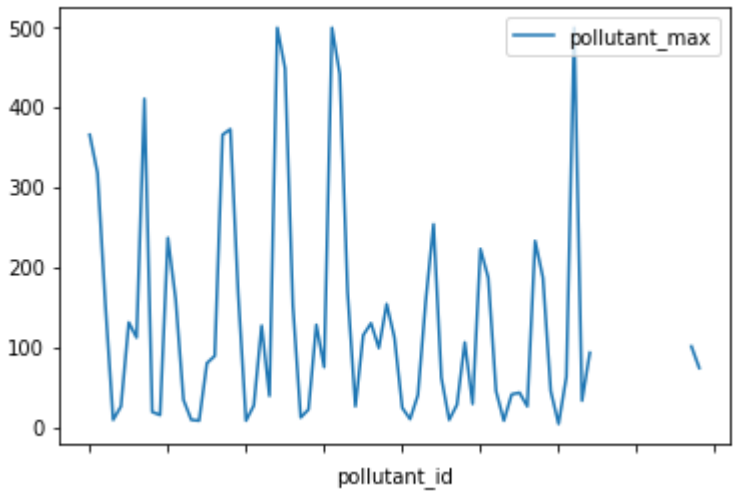

pollutant_max
400
350
300
250
200
150
100
50
0
pollutant_id

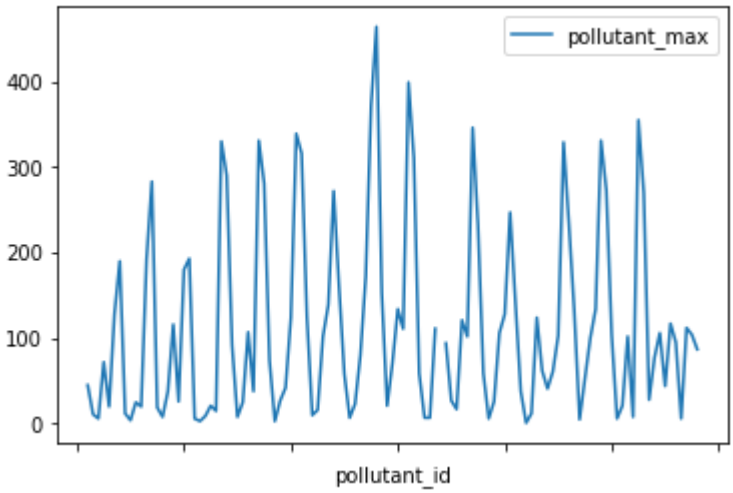

pollutant_max
500
400
300
200
100
0
pollutant_id

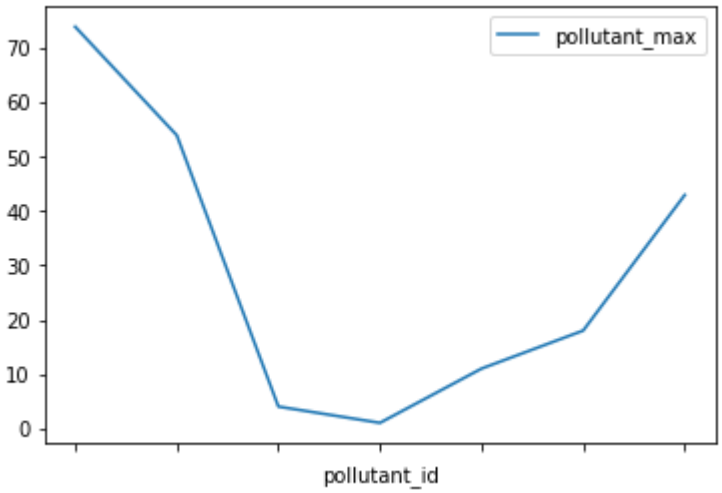

pollutant_max
200
175
150
125
100
75
50
25
0
pollutant_id

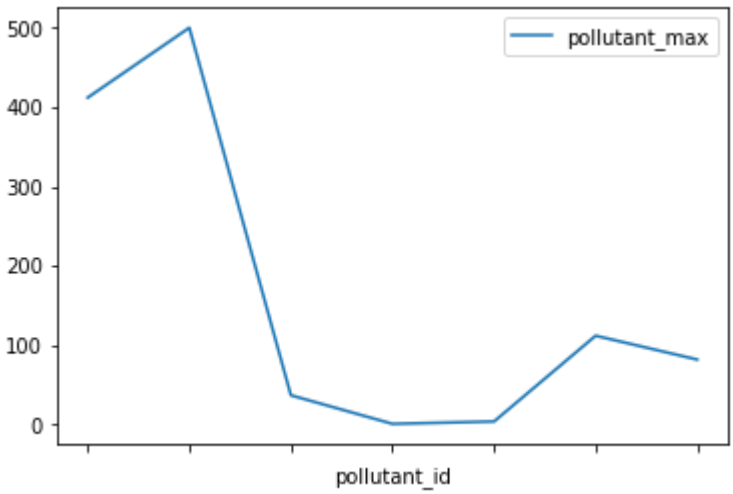

pollutant_max
200
150
100
50
0
pollutant_id

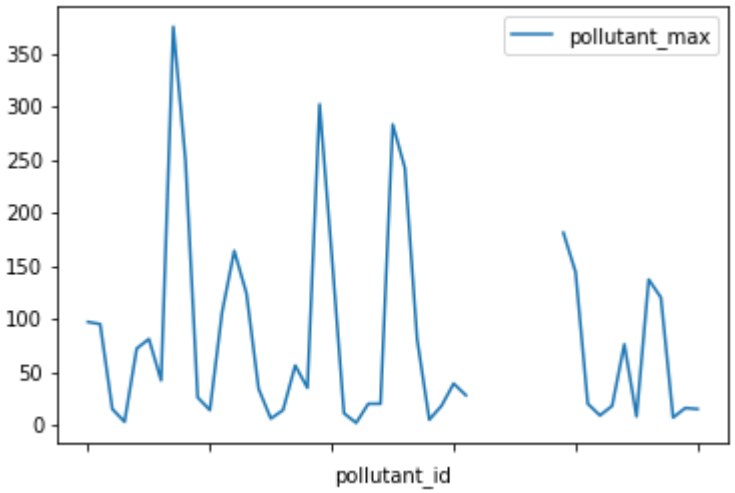

pollutant_max
200
175
150
125
100
75
50
25
0
pollutant_id

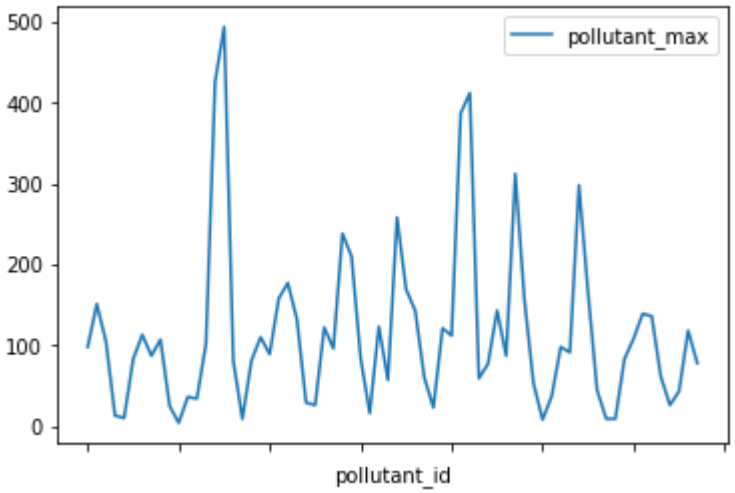

pollutant_max
500
400
300
200
100
0
pollutant_id

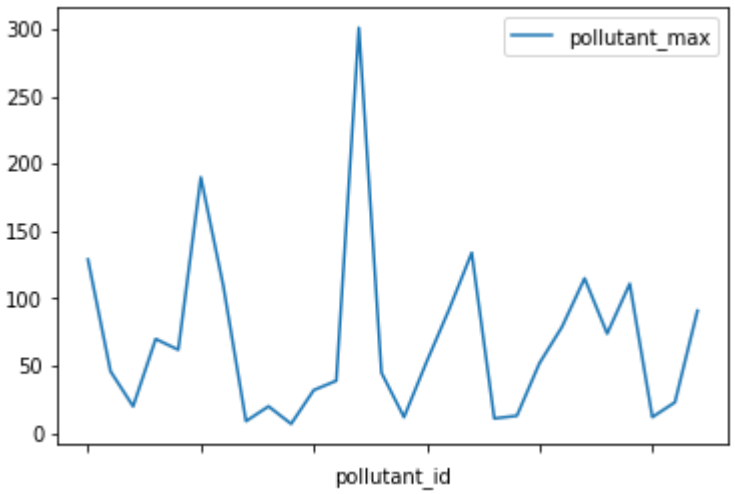

pollutant_max
400
300
200
100
0
pollutant_id

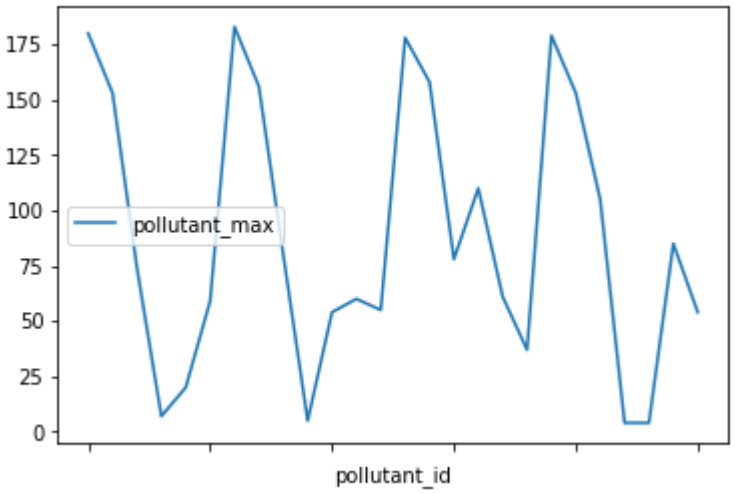

pollutant_max
70
60
50
40
30
20
10
0
pollutant_id

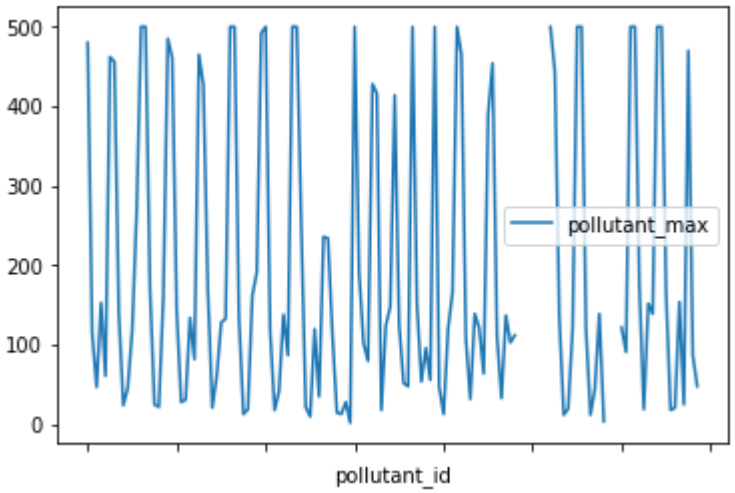

pollutant_max
500
400
300
200
100
0
pollutant_id

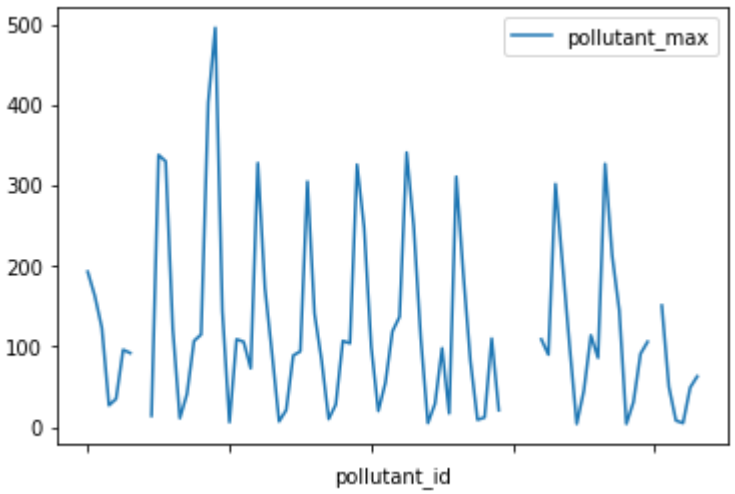

pollutant_max
350
300
250
200
150
100
50
0
pollutant_id


pollutant_max
500
400
300
200
100
0
pollutant_id


pollutant_max
300
250
200
150
100
50
0
pollutant_id

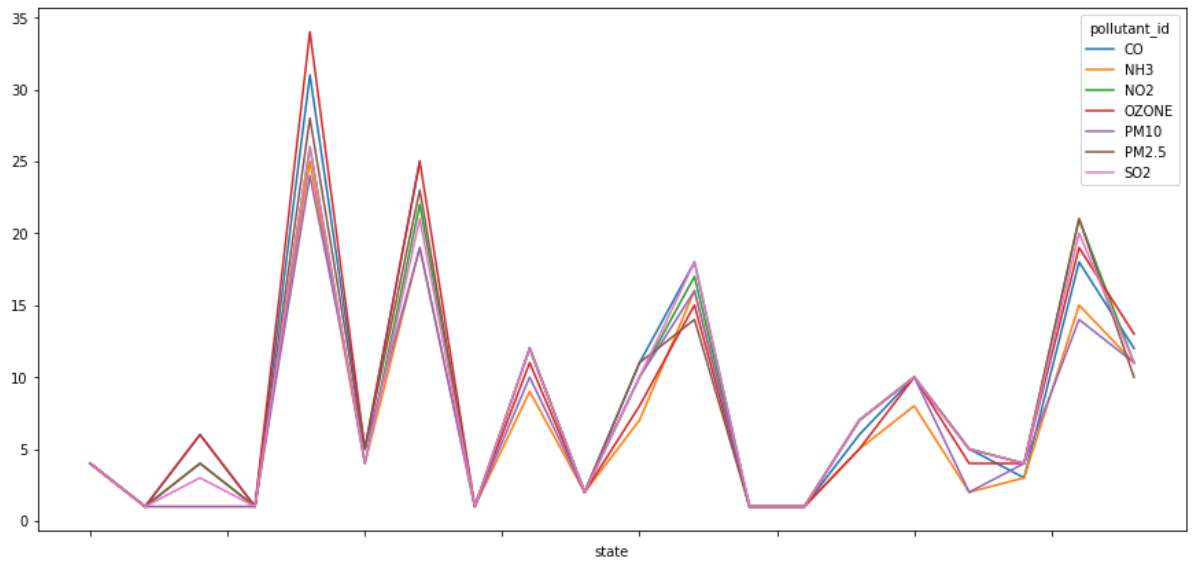

175
150
125
100
75
50
25
0
pollutant_max
pollutant_id


500
400
300
200
100
0
pollutant_max
pollutant_id


500
400
300
200
100
0
pollutant_max
pollutant_id

In [116]:
```
fig, ax = plt.subplots(figsize=(15,7))
data.groupby(['state','pollutant_id']).count()['pollutant_max'].unstack().plot
(ax=ax)
```

Out[116]: <matplotlib.axes._subplots.AxesSubplot at 0x1d6574f1080>



In [126]:
```
#CALCULATE TOTAL MISSING VALUES AND THEIR PERCENTAGE

total = data.isnull().sum().sort_values(ascending=False)
total.head()
```

Out[126]:
```
air quality index value    126
Predominant_Parameter      126
pollutant_min              102
pollutant_max              102
pollutant_avg              102
dtype: int64
```

In [127]:
```
#Calculate the percent of null values for each columns (sum of null values / t
otal non-null value) *100

percent = (data.isnull().sum()/data.isnull().count()*100).sort_values(ascendin
g=False)  #count(returns Non-NAN value)
```

In [128]:
```
missing_data = pd.concat([total, percent], axis=1, keys=['Total', 'Percent'])
```

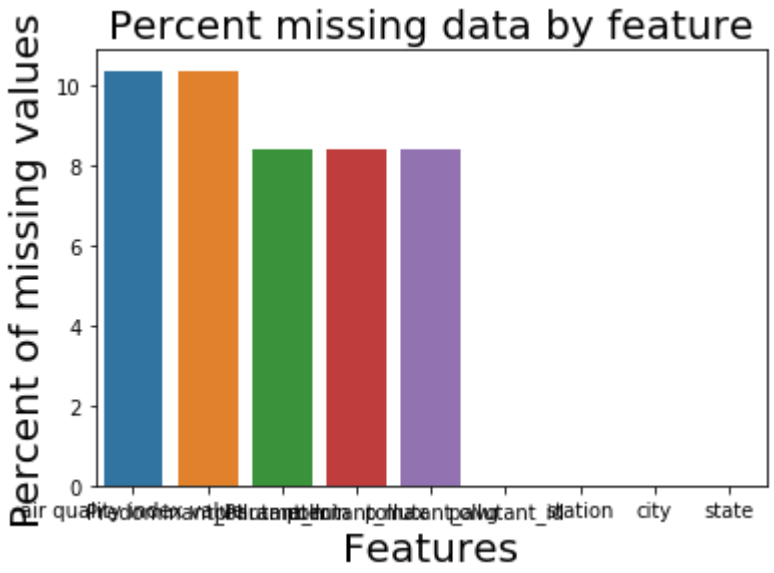
In [129]:
```
missing_data.head()
```

Out[129]:

	Total	Percent
air quality index value	126	10.370370
Predominant_Parameter	126	10.370370
pollutant_min	102	8.395062
pollutant_max	102	8.395062
pollutant_avg	102	8.395062

In [137]:
```
#PERCENT OF MISSING VALUE (BAR PLOT)

sns.barplot(x=missing_data.index, y=missing_data['Percent'])
plt.xlabel('Features', fontsize=20)
plt.ylabel('Percent of missing values', fontsize=20)
plt.title('Percent missing data by feature', fontsize=20)
```

Out[137]: Text(0.5,1,'Percent missing data by feature')



In [138]:
```
#MEAN DISTRIBUTION BY STATE
data.groupby('state')[['pollutant_id','pollutant_avg','pollutant_max','pollutant_min','air quality index value']].mean()
```

Out[138]:

state	pollutant_avg	pollutant_max	pollutant_min	air quality index value
Andhra_Pradesh	60.178571	94.428571	33.214286	153.750000
Assam	83.428571	152.142857	24.285714	265.000000
Bihar	76.120000	105.280000	52.400000	265.333333
Chandigarh	58.000000	113.857143	36.142857	158.000000
Delhi	122.087629	180.226804	60.582474	358.449198
Gujarat	70.437500	126.375000	38.031250	161.909091
Haryana	81.850649	134.051948	43.935065	227.052632
Jharkhand	53.833333	101.666667	34.166667	118.000000
Karnataka	39.320513	70.179487	20.602564	76.740741
Kerala	35.285714	73.928571	12.142857	81.500000
Madhya Pradesh	65.731343	134.268657	25.895522	155.538462
Maharashtra	64.561404	112.122807	32.938596	180.759615
Meghalaya	14.857143	29.285714	6.428571	29.000000
Odisha	99.857143	164.000000	32.428571	287.000000
Punjab	52.659091	81.136364	36.454545	135.863636
Rajasthan	61.485294	115.441176	24.058824	133.735294
TamilNadu	42.892857	69.678571	23.571429	102.214286
Telangana	61.615385	88.153846	32.423077	136.192308
Uttar_Pradesh	129.179688	194.281250	73.210938	361.779528
West_Bengal	68.050633	121.645570	33.683544	177.000000

In [10]:
```
#FILL MISSING VALUES BY MEAN (GROUP BY STATE)

grp_state = data.groupby('state')
```

In [12]:
```
def impute_mean_by_state(series):
    return series.fillna(series.mean())
```

In [21]:
```
data['pollutant_avg']=grp_state['pollutant_avg'].transform(impute_mean_by_state)  #fill value with mean value group by state
data['pollutant_max']=grp_state['pollutant_max'].transform(impute_mean_by_state)
data['pollutant_min']=grp_state['pollutant_min'].transform(impute_mean_by_state)
data.fillna(data['Predominant_Parameter'].mode()[0], inplace=True) #fill value with median value group by state
data['air quality index value']=grp_state['air quality index value'].transform(impute_mean_by_state)
```

In [22]:
```
data.describe()
```

Out[22]:

	pollutant_avg	pollutant_max	pollutant_min	air quality index value
count	1215.000000	1215.000000	1215.000000	1215.000000
mean	82.146036	132.868394	42.536967	223.693759
std	92.344411	134.831531	54.919986	111.153724
min	1.000000	1.000000	1.000000	29.000000
25%	17.000000	32.000000	6.000000	123.000000
50%	50.000000	95.000000	21.000000	209.000000
75%	110.000000	170.000000	58.000000	338.500000
max	416.000000	500.000000	338.000000	416.000000

In [23]:
```
data.isna().sum()
```

Out[23]:
```
country                    0
state                      0
city                       0
station                    0
lastupdate                 0
pollutant_avg              0
pollutant_max              0
pollutant_min              0
pollutant_id               0
Predominant_Parameter      0
air quality index value    0
dtype: int64
```

In [27]:
```
data.tail(10)
```

Out[27]:

	country	state	city	station	lastupdate	pollutant_avg	pollutant_max	polluta
1205	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	9.000000	31.00000	4.
1206	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	71.000000	91.00000	33.
1207	India	West_Bengal	Kolkata	Victoria, Kolkata - WBPCB	13-01-2020 12:00:00	17.000000	106.00000	1.
1208	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	68.050633	121.64557	33.
1209	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	106.000000	151.00000	52.
1210	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	32.000000	50.00000	19.
1211	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	7.000000	8.00000	6.
1212	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	5.000000	5.00000	4.
1213	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	39.000000	49.00000	15.
1214	India	West_Bengal	Siliguri	Ward-32 Bapupara, Siliguri - WBPCB	13-01-2020 12:00:00	23.000000	63.00000	8.

In [35]:
```
data.pollutant_id.unique() #identifying the pollutants that are mostly affecti
ng the air quality
```

Out[35]:
```
array(['PM2.5', 'PM10', 'NO2', 'NH3', 'SO2', 'CO', 'OZONE'], dtype=object)
```

In [66]:
```
is_pm2 = data['Predominant_Parameter']=='PM2.5'
print(is_pm2)
```

```
0        True
1        True
2        True
3        True
4        True
5        True
6        True
7        True
8        True
9        True
10       True
11       True
12       True
13       True
14       True
15       True
16       True
17       True
18       True
19       True
20       True
21       True
22       True
23       True
24       True
25       True
26       True
27       True
28       True
29       True
        ...
1185     True
1186     True
1187     True
1188     True
1189     True
1190     True
1191     True
1192     True
1193     True
1194    False
1195    False
1196    False
1197    False
1198    False
1199    False
1200    False
1201     True
1202     True
1203     True
1204     True
1205     True
1206     True
1207     True
1208    False
1209    False
1210    False
```

```
1211    False
1212    False
1213    False
1214    False
Name: Predominant_Parameter, Length: 1215, dtype: bool
```

In [71]:
```
pid_pm2 = data[is_pm2]
print(pid_pm2.shape)
```

```
(957, 11)
```

In [72]:
```
print(pid_pm2.head())
```

```
  country           state       city                         station  \
0   India  Andhra_Pradesh  Amaravati  Secretariat, Amaravati - APPCB
1   India  Andhra_Pradesh  Amaravati  Secretariat, Amaravati - APPCB
2   India  Andhra_Pradesh  Amaravati  Secretariat, Amaravati - APPCB
3   India  Andhra_Pradesh  Amaravati  Secretariat, Amaravati - APPCB
4   India  Andhra_Pradesh  Amaravati  Secretariat, Amaravati - APPCB

            lastupdate  pollutant_avg  pollutant_max  pollutant_min  \
0  13-01-2020 12:00:00          167.0          272.0           68.0
1  13-01-2020 12:00:00          105.0          132.0           65.0
2  13-01-2020 12:00:00           21.0           51.0            8.0
3  13-01-2020 12:00:00            5.0            5.0            4.0
4  13-01-2020 12:00:00           28.0          111.0            3.0

  pollutant_id Predominant_Parameter  air quality index value
0        PM2.5                 PM2.5                   167.0
1         PM10                 PM2.5                   167.0
2          NO2                 PM2.5                   167.0
3          NH3                 PM2.5                   167.0
4          SO2                 PM2.5                   167.0
```

In [73]:
```
is_pm10 = data['Predominant_Parameter']=='PM10'
print(is_pm10)
```

```
0       False
1       False
2       False
3       False
4       False
5       False
6       False
7       False
8       False
9       False
10      False
11      False
12      False
13      False
14      False
15      False
16      False
17      False
18      False
19      False
20      False
21      False
22      False
23      False
24      False
25      False
26      False
27      False
28      False
29      False
        ...
1185    False
1186    False
1187    False
1188    False
1189    False
1190    False
1191    False
1192    False
1193    False
1194     True
1195     True
1196     True
1197     True
1198     True
1199     True
1200     True
1201    False
1202    False
1203    False
1204    False
1205    False
1206    False
1207    False
1208     True
1209     True
1210     True
```

```
1211     True
1212     True
1213     True
1214     True
Name: Predominant_Parameter, Length: 1215, dtype: bool
```

In [74]:
```
pid_pm10 = data[is_pm10]
print(pid_pm10.shape)
```

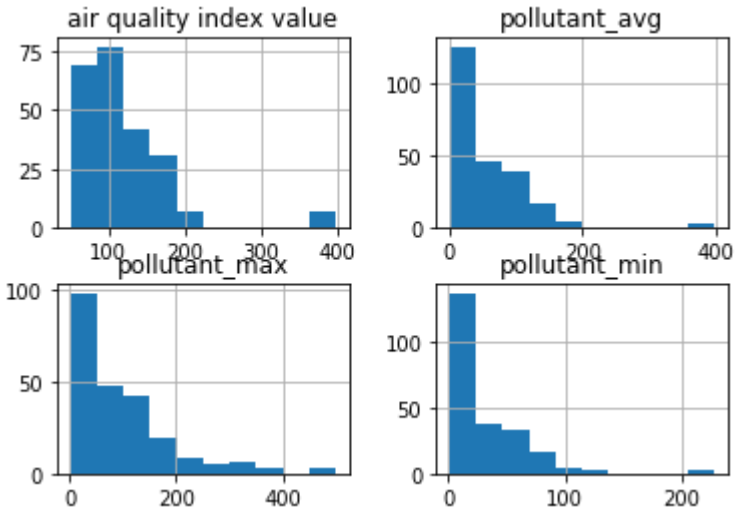
```
(233, 11)
```

In [75]:
```
print(pid_pm10.head())
```

```
    country    state         city                         station  \
313   India  Gujarat  GandhiNagar  Sector-10, Gandhinagar - GPCB
314   India  Gujarat  GandhiNagar  Sector-10, Gandhinagar - GPCB
315   India  Gujarat  GandhiNagar  Sector-10, Gandhinagar - GPCB
316   India  Gujarat  GandhiNagar  Sector-10, Gandhinagar - GPCB
317   India  Gujarat  GandhiNagar  Sector-10, Gandhinagar - GPCB

              lastupdate  pollutant_avg  pollutant_max  pollutant_min  \
313  13-01-2020 12:00:00           48.0           84.0           29.0
314  13-01-2020 12:00:00           50.0           78.0           30.0
315  13-01-2020 12:00:00           12.0           15.0            9.0
316  13-01-2020 12:00:00            2.0            3.0            2.0
317  13-01-2020 12:00:00            8.0           11.0            6.0

    pollutant_id Predominant_Parameter  air quality index value
313        PM2.5                  PM10                    50.0
314         PM10                  PM10                    50.0
315          NO2                  PM10                    50.0
316          NH3                  PM10                    50.0
317          SO2                  PM10                    50.0
```

In [86]:
```
pid_pm10.hist()
```

Out[86]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x00000256AE69DCF8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x00000256AF6F7390
>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x00000256AF7206A0>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x00000256AF7489B0
>]],
      dtype=object)
```



In [76]:
```
is_CO = data['Predominant_Parameter']=='CO'
print(is_CO)
```

```
0       False
1       False
2       False
3       False
4       False
5       False
6       False
7       False
8       False
9       False
10      False
11      False
12      False
13      False
14      False
15      False
16      False
17      False
18      False
19      False
20      False
21      False
22      False
23      False
24      False
25      False
26      False
27      False
28      False
29      False
        ...
1185    False
1186    False
1187    False
1188    False
1189    False
1190    False
1191    False
1192    False
1193    False
1194    False
1195    False
1196    False
1197    False
1198    False
1199    False
1200    False
1201    False
1202    False
1203    False
1204    False
1205    False
1206    False
1207    False
1208    False
1209    False
1210    False
```

```
1211    False
1212    False
1213    False
1214    False
Name: Predominant_Parameter, Length: 1215, dtype: bool
```

In [77]:
```
pid_CO = data[is_CO]
print(pid_CO.shape)
```

```
(13, 11)
```

In [78]:
```
print(pid_CO.head())
```

```
   country  state   city                    station           lastupdate  \
64   India  Bihar  Patna  Samanpura, Patna - BSPCB  13-01-2020 12:00:00
65   India  Bihar  Patna  Samanpura, Patna - BSPCB  13-01-2020 12:00:00
66   India  Bihar  Patna  Samanpura, Patna - BSPCB  13-01-2020 12:00:00
67   India  Bihar  Patna  Samanpura, Patna - BSPCB  13-01-2020 12:00:00
68   India  Bihar  Patna  Samanpura, Patna - BSPCB  13-01-2020 12:00:00

    pollutant_avg  pollutant_max  pollutant_min pollutant_id  \
64          110.0          110.0          110.0        PM2.5
65           81.0           81.0           81.0         PM10
66           26.0           26.0           26.0          NO2
67            4.0            4.0            4.0          NH3
68          147.0          147.0          147.0           CO

   Predominant_Parameter  air quality index value
64                     CO                    147.0
65                     CO                    147.0
66                     CO                    147.0
67                     CO                    147.0
68                     CO                    147.0
```

In [79]:
```
is_NO2 = data['Predominant_Parameter']=='NO2'
print(is_NO2)
```

```
0       False
1       False
2       False
3       False
4       False
5       False
6       False
7       False
8       False
9       False
10      False
11      False
12      False
13      False
14      False
15      False
16      False
17      False
18      False
19      False
20      False
21      False
22      False
23      False
24      False
25      False
26      False
27      False
28      False
29      False
        ...
1185    False
1186    False
1187    False
1188    False
1189    False
1190    False
1191    False
1192    False
1193    False
1194    False
1195    False
1196    False
1197    False
1198    False
1199    False
1200    False
1201    False
1202    False
1203    False
1204    False
1205    False
1206    False
1207    False
1208    False
1209    False
1210    False
```

```
1211    False
1212    False
1213    False
1214    False
Name: Predominant_Parameter, Length: 1215, dtype: bool
```

In [80]:
```
pid_NO2 = data[is_NO2]
print(pid_NO2.shape)
```

```
(5, 11)
```

In [81]:
```
print(pid_NO2.head())
```

```
    country      state       city                        station  \
504   India  Karnataka  Bengaluru  BTM Layout, Bengaluru - CPCB
505   India  Karnataka  Bengaluru  BTM Layout, Bengaluru - CPCB
506   India  Karnataka  Bengaluru  BTM Layout, Bengaluru - CPCB
507   India  Karnataka  Bengaluru  BTM Layout, Bengaluru - CPCB
508   India  Karnataka  Bengaluru  BTM Layout, Bengaluru - CPCB

              lastupdate  pollutant_avg  pollutant_max  pollutant_min  \
504  13-01-2020 12:00:00           21.0           29.0           15.0
505  13-01-2020 12:00:00           36.0           88.0           19.0
506  13-01-2020 12:00:00            9.0           24.0            4.0
507  13-01-2020 12:00:00           30.0           69.0            5.0
508  13-01-2020 12:00:00           24.0           42.0            6.0

    pollutant_id Predominant_Parameter  air quality index value
504        PM2.5                   NO2                    36.0
505          NO2                   NO2                    36.0
506          SO2                   NO2                    36.0
507           CO                   NO2                    36.0
508        OZONE                   NO2                    36.0
```

In [82]:
```
is_ozone = data['Predominant_Parameter']=='OZONE'
print(is_ozone)
```

```
0       False
1       False
2       False
3       False
4       False
5       False
6       False
7       False
8       False
9       False
10      False
11      False
12      False
13      False
14      False
15      False
16      False
17      False
18      False
19      False
20      False
21      False
22      False
23      False
24      False
25      False
26      False
27      False
28      False
29      False
        ...
1185    False
1186    False
1187    False
1188    False
1189    False
1190    False
1191    False
1192    False
1193    False
1194    False
1195    False
1196    False
1197    False
1198    False
1199    False
1200    False
1201    False
1202    False
1203    False
1204    False
1205    False
1206    False
1207    False
1208    False
1209    False
1210    False
```

```
1211    False
1212    False
1213    False
1214    False
Name: Predominant_Parameter, Length: 1215, dtype: bool
```

In [83]:
```
pid_ozone = data[is_ozone]
print(pid_ozone.shape)
```

```
(7, 11)
```

In [84]:
```
print(pid_ozone.head())
```

```
    country      state    city                             station  \
571   India  Karnataka  Mysuru  Hebbal 1st Stage, Mysuru - KSPCB
572   India  Karnataka  Mysuru  Hebbal 1st Stage, Mysuru - KSPCB
573   India  Karnataka  Mysuru  Hebbal 1st Stage, Mysuru - KSPCB
574   India  Karnataka  Mysuru  Hebbal 1st Stage, Mysuru - KSPCB
575   India  Karnataka  Mysuru  Hebbal 1st Stage, Mysuru - KSPCB

              lastupdate  pollutant_avg  pollutant_max  pollutant_min  \
571  13-01-2020 12:00:00      42.000000      54.000000      30.000000
572  13-01-2020 12:00:00      34.000000      43.000000      24.000000
573  13-01-2020 12:00:00      39.320513      70.179487      20.602564
574  13-01-2020 12:00:00      39.320513      70.179487      20.602564
575  13-01-2020 12:00:00      26.000000     140.000000       7.000000

    pollutant_id Predominant_Parameter  air quality index value
571        PM2.5                 OZONE                    45.0
572         PM10                 OZONE                    45.0
573          NO2                 OZONE                    45.0
574          NH3                 OZONE                    45.0
575          SO2                 OZONE                    45.0
```

In [85]:
```
data.Predominant_Parameter.unique()
```

Out[85]:
```
array(['PM2.5', 'CO', 'PM10', 'NO2', 'OZONE'], dtype=object)
```

In [ ]:
```

```