New Delhi - Air Pollution Analysis - Diwali 2022

In [43]:
```
# Project by - Prof. Nirmal Gaud
# Contact - ds.ml.projects.sessions.1@gmail.com
```

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
import warnings
warnings.filterwarnings('ignore')
```

In [3]:
```
df = pd.read_csv('new_delhi_air_pollution.csv')
```

In [4]:
```
df.head()
```

Out[4]:

	Unnamed: 0	AQI	CO	datetime	no2	o3	pm10	pm25	so2	timestamp_local	timestamp_utc	ts
0	0	30.2	198.60268	2022-10-21:18	0.046857	55.789948	10.486722	5.637410	0.387430	2022-10-21T23:00:00	2022-10-21T18:00:00	1666375200
1	1	28.2	197.60132	2022-10-21:19	0.046456	54.931640	10.719325	4.618169	0.409782	2022-10-22T00:00:00	2022-10-21T19:00:00	1666378800
2	2	26.6	198.60268	2022-10-21:20	0.046857	54.645540	11.155578	3.520902	0.402331	2022-10-22T01:00:00	2022-10-21T20:00:00	1666382400
3	3	25.0	201.94054	2022-10-21:21	0.048196	55.074690	11.116206	2.225919	0.376254	2022-10-22T02:00:00	2022-10-21T21:00:00	1666386000
4	4	26.0	205.27840	2022-10-21:22	0.048865	55.789948	10.405250	1.979471	0.339001	2022-10-22T03:00:00	2022-10-21T22:00:00	1666389600

In [5]:
```
df.tail()
```

Out[5]:

	Unnamed: 0	AQI	CO	datetime	no2	o3	pm10	pm25	so2	timestamp_local	timestamp_utc	ts
67	67	22.0	193.59589	2022-10-24:13	0.035478	47.206880	10.423121	2.720472	0.391155	2022-10-24T18:00:00	2022-10-24T13:00:00	1666616400
68	68	24.0	195.26482	2022-10-24:14	0.039159	52.213670	11.240661	2.713109	0.383705	2022-10-24T19:00:00	2022-10-24T14:00:00	1666620000
69	69	26.0	196.93375	2022-10-24:15	0.044180	56.505203	12.098125	2.743044	0.376254	2022-10-24T20:00:00	2022-10-24T15:00:00	1666623600
70	70	27.0	198.60268	2022-10-24:16	0.046857	59.366226	12.845977	2.780022	0.372529	2022-10-24T21:00:00	2022-10-24T16:00:00	1666627200
71	71	28.0	198.60268	2022-10-24:17	0.047527	60.796738	12.583639	2.745875	0.368804	2022-10-24T22:00:00	2022-10-24T17:00:00	1666630800

In [6]:
```
df = df.drop('Unnamed: 0', axis = 1)
```

In [7]:
```
df.shape
```

Out[7]:
```
(72, 11)
```

In [8]:
```
df.columns
```

Out[8]:
```
Index(['AQI', 'CO', 'datetime', 'no2', 'o3', 'pm10', 'pm25', 'so2',
       'timestamp_local', 'timestamp_utc', 'ts'],
      dtype='object')
```

In [9]:
```
df.duplicated().sum()
```

Out[9]:
```
0
```

In [10]:
```
df.isnull().sum()
```

Out[10]:
```
AQI               0
CO                0
datetime          0
no2               0
o3                0
pm10              0
pm25              0
so2               0
timestamp_local   0
timestamp_utc     0
ts                0
dtype: int64
```

In [11]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72 entries, 0 to 71
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   AQI              72 non-null     float64
 1   CO               72 non-null     float64
 2   datetime         72 non-null     object 
 3   no2              72 non-null     float64
 4   o3               72 non-null     float64
 5   pm10             72 non-null     float64
 6   pm25             72 non-null     float64
 7   so2              72 non-null     float64
 8   timestamp_local  72 non-null     object 
 9   timestamp_utc    72 non-null     object 
 10  ts               72 non-null     int64  
dtypes: float64(7), int64(1), object(3)
memory usage: 6.3+ KB
```

In [12]:
```
df.describe()
```

Out[12]:

	AQI	CO	no2	o3	pm10	pm25	so2	ts
count	72.000000	72.000000	72.000000	72.000000	72.000000	72.000000	72.000000	7.200000e+01
mean	26.180556	200.095443	0.041996	56.571762	9.115038	2.295406	0.363423	1.666503e+09
std	3.054650	4.515537	0.010940	6.431549	1.531785	0.656507	0.044664	7.534242e+04
min	19.000000	191.926960	0.015563	41.484833	6.686746	1.633373	0.283122	1.666375e+09
25%	25.000000	197.434427	0.037988	53.644180	7.864763	1.828729	0.320375	1.666439e+09
50%	27.000000	200.271600	0.045518	57.220460	8.914176	2.201811	0.372529	1.666503e+09
75%	28.000000	203.609470	0.050204	60.081482	10.274836	2.522210	0.395812	1.666567e+09
max	32.000000	208.616260	0.060914	68.664550	12.845977	5.637410	0.454485	1.666631e+09

In [13]:
```
df.nunique()
```

In [13]:
```
df.nunique()
```

Out[13]:
```
AQI                17
CO                 12
datetime           72
no2                43
o3                 35
pm10               72
pm25               72
so2                35
timestamp_local    72
timestamp_utc      72
ts                 72
dtype: int64
```

In [14]:
```
object_columns = df.select_dtypes(include=['object']).columns
print("Object type columns:")
print(object_columns)

numerical_columns = df.select_dtypes(include=['int64', 'float64']).columns
print("\nNumerical type columns:")
print(numerical_columns)
```

```
Object type columns:
Index(['datetime', 'timestamp_local', 'timestamp_utc'], dtype='object')

Numerical type columns:
Index(['AQI', 'CO', 'no2', 'o3', 'pm10', 'pm25', 'so2', 'ts'], dtype='object')
```

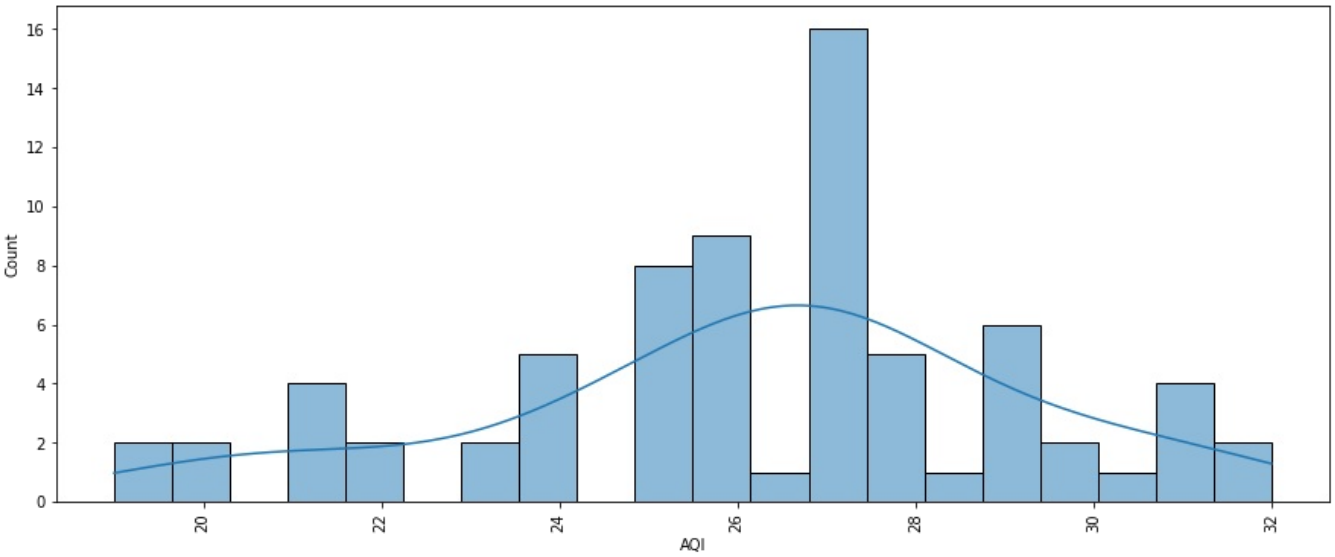
In [15]:
```
def classify_features(df):
    categorical_features = []
    non_categorical_features = []
    discrete_features = []
    continuous_features = []

    for column in df.columns:
        if df[column].dtype == 'object':
            if df[column].nunique() < 10:
                categorical_features.append(column)
            else:
                non_categorical_features.append(column)
        elif df[column].dtype in ['int64', 'float64']:
            if df[column].nunique() < 10:
                discrete_features.append(column)
            else:
                continuous_features.append(column)

    return categorical_features, non_categorical_features, discrete_features, continuous_features
```

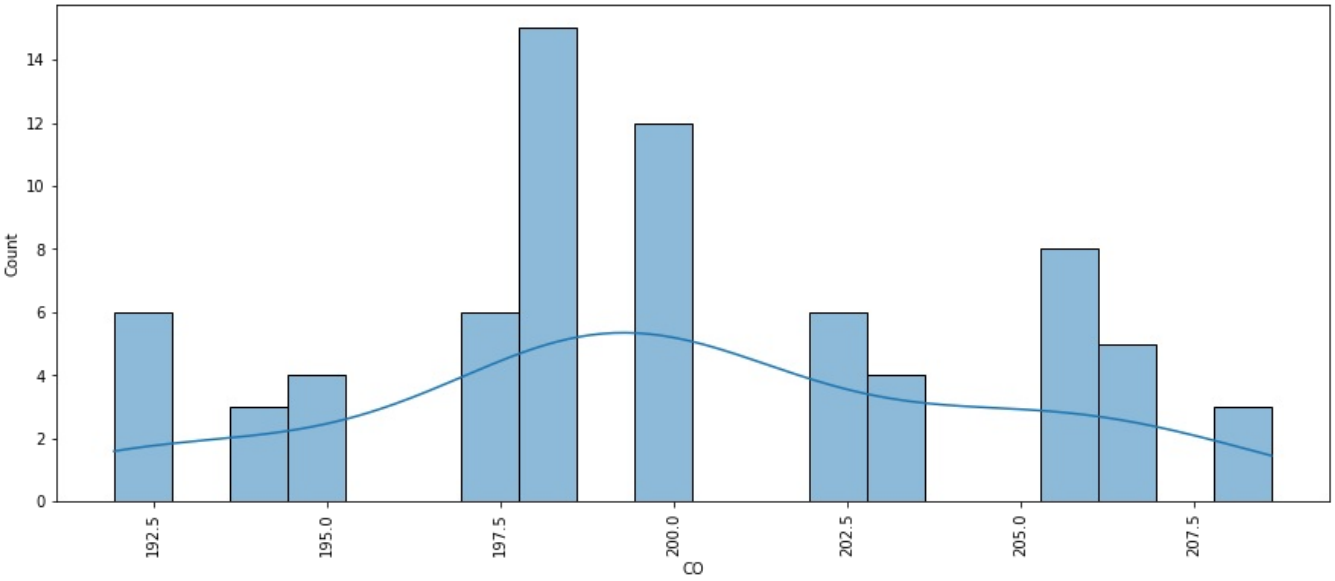
In [16]:
```
categorical, non_categorical, discrete, continuous = classify_features(df)
```

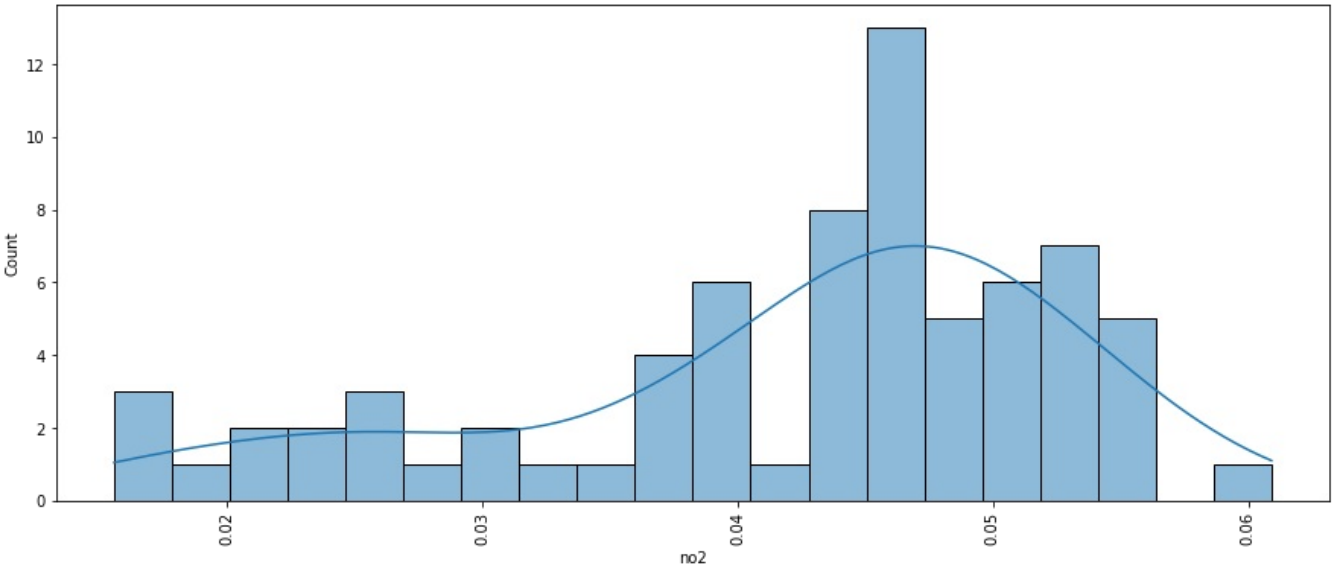
In [17]:
```
print("Categorical Features:", categorical)
print("Non-Categorical Features:", non_categorical)
print("Discrete Features:", discrete)
print("Continuous Features:", continuous)
```

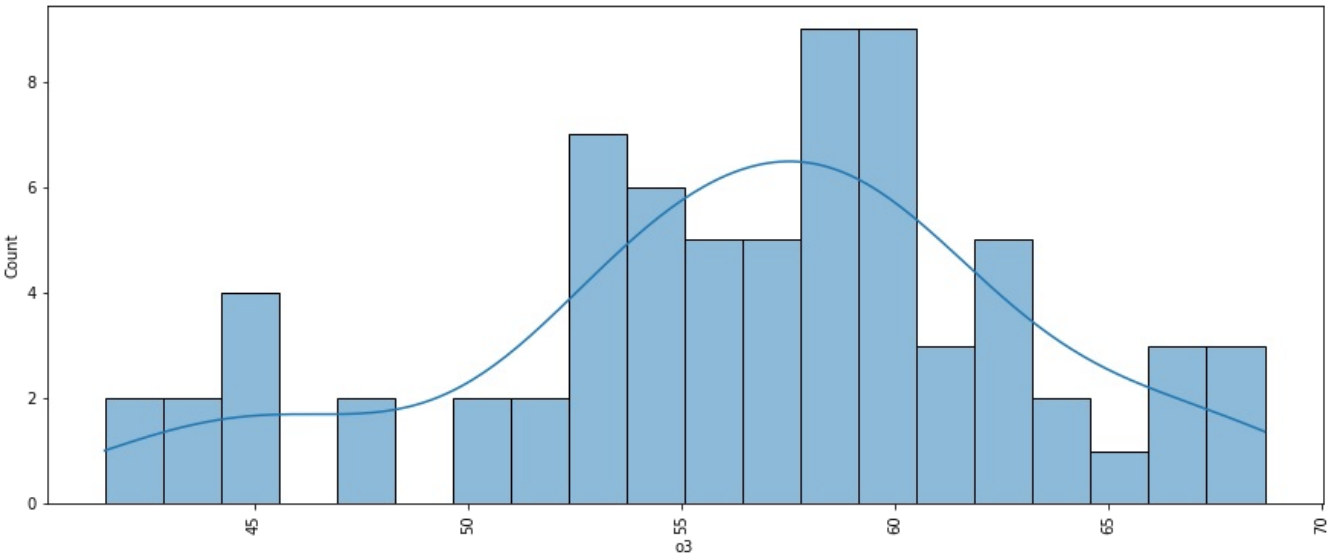
```
Categorical Features: []
Non-Categorical Features: ['datetime', 'timestamp_local', 'timestamp_utc']
Discrete Features: []
Continuous Features: ['AQI', 'CO', 'no2', 'o3', 'pm10', 'pm25', 'so2', 'ts']
```

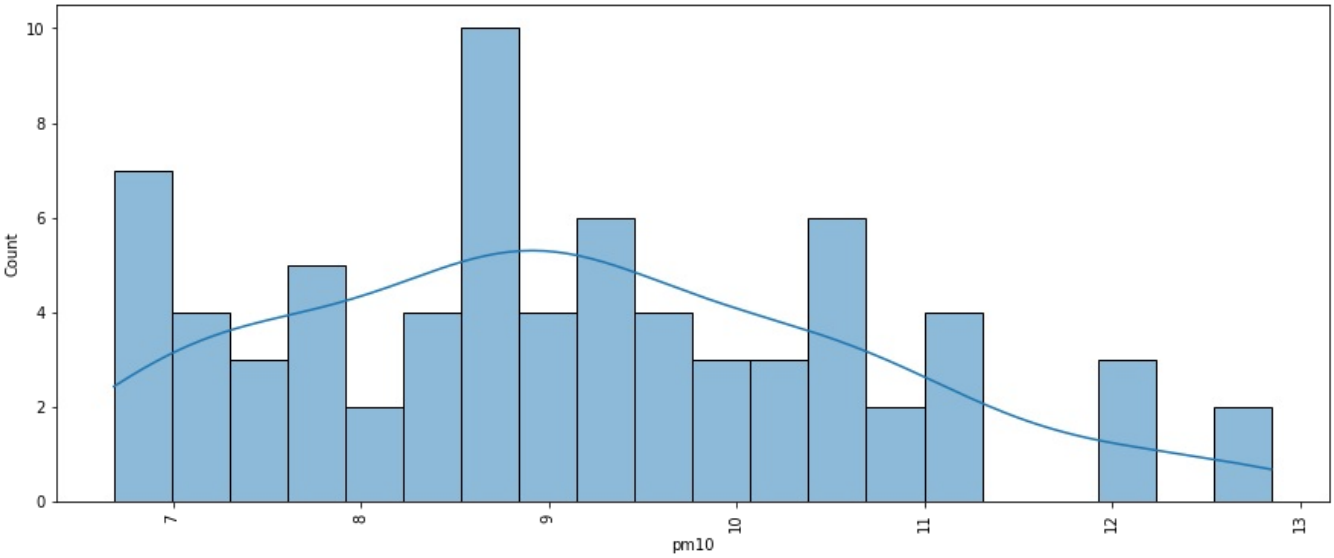
In [18]:
```
for i in numerical_columns:
    plt.figure(figsize=(15,6))
    sns.histplot(df[i], kde = True, bins = 20, palette = 'hls')
    plt.xticks(rotation = 90)
    plt.show()
```

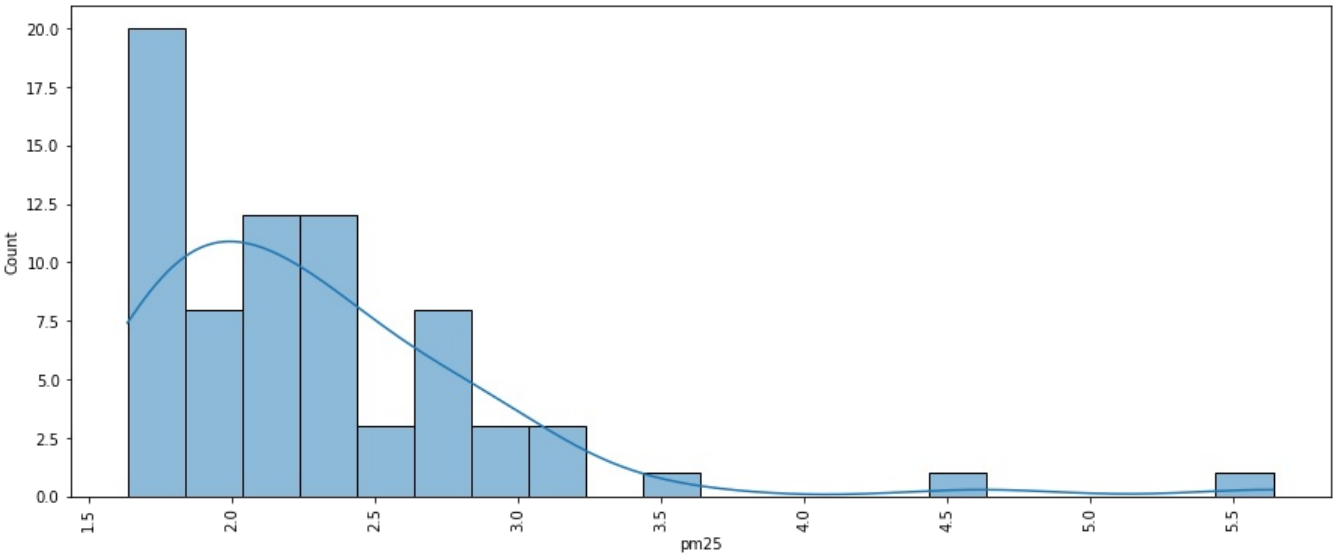

Count
AQI

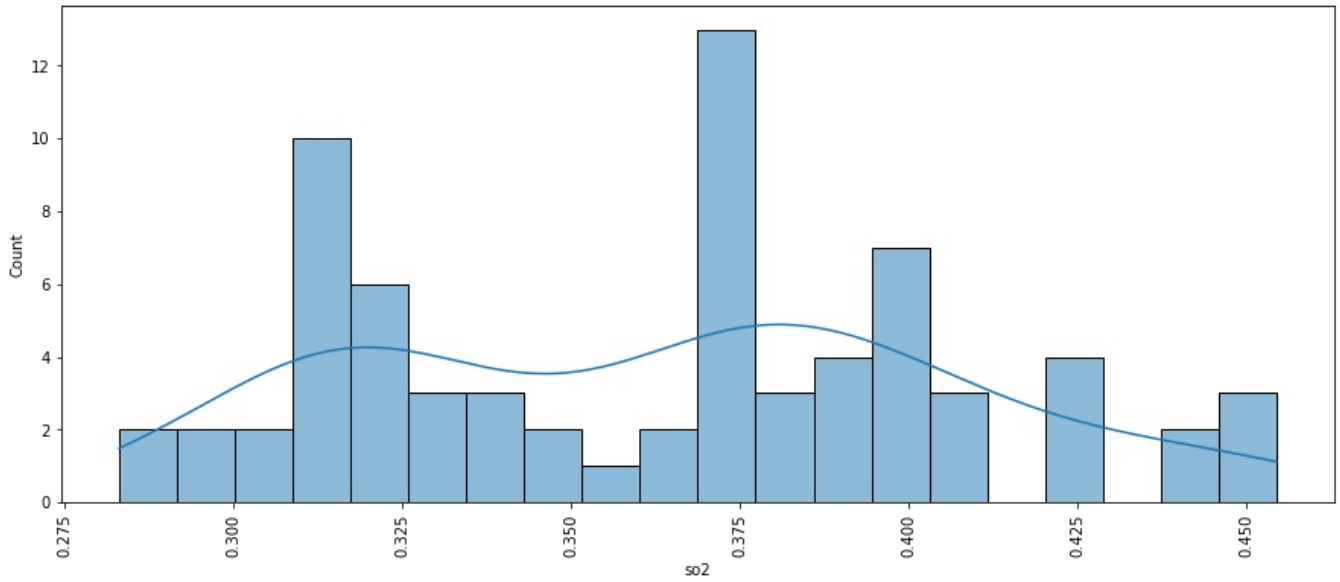

Count
CO

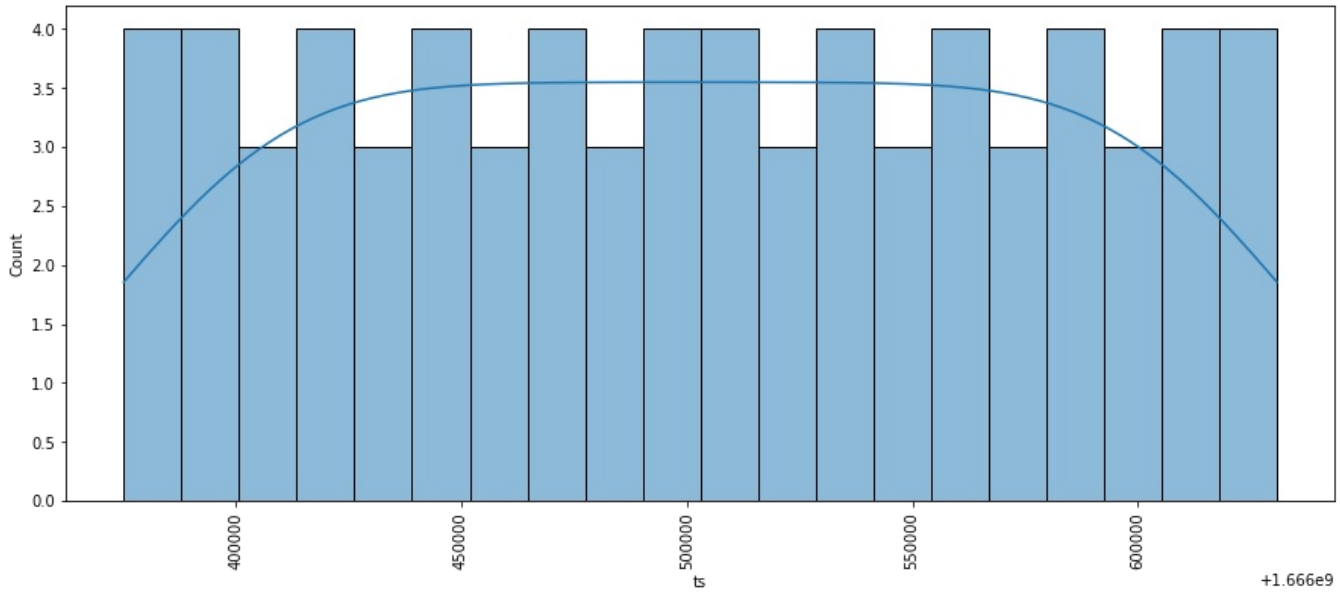

Count
no2


Count
o3

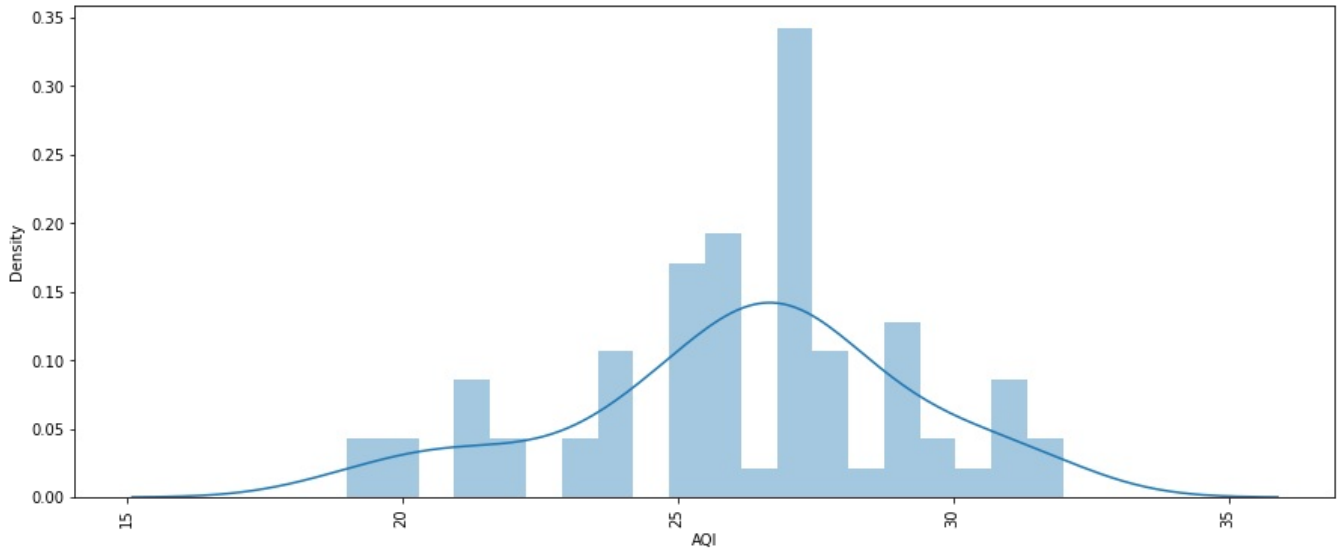

Count
pm10

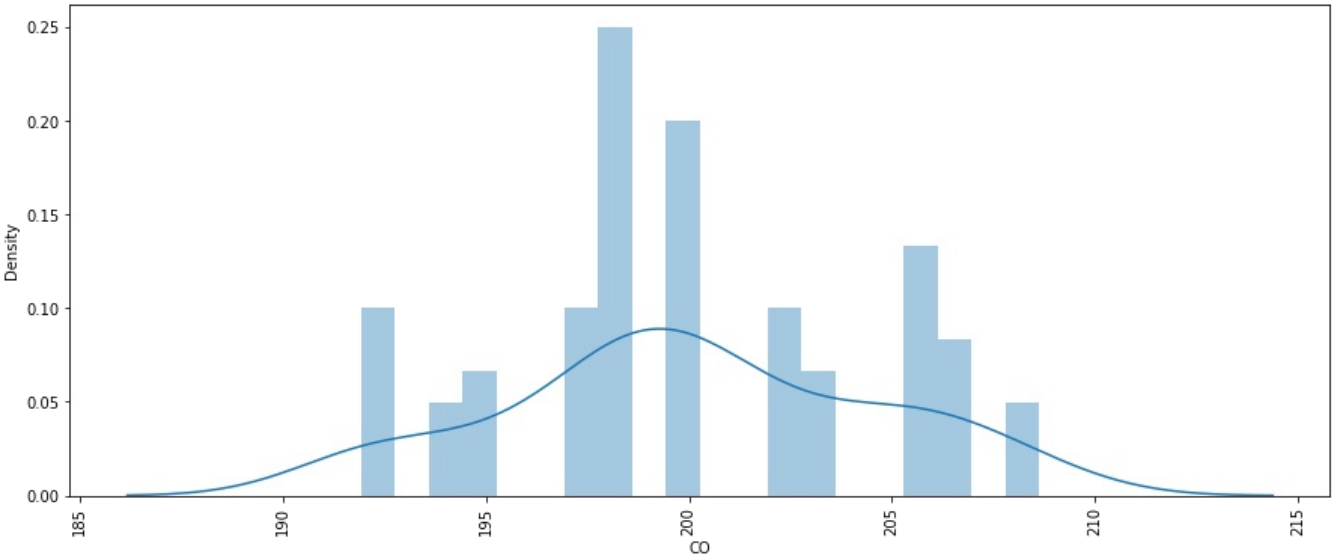

Count
pm25

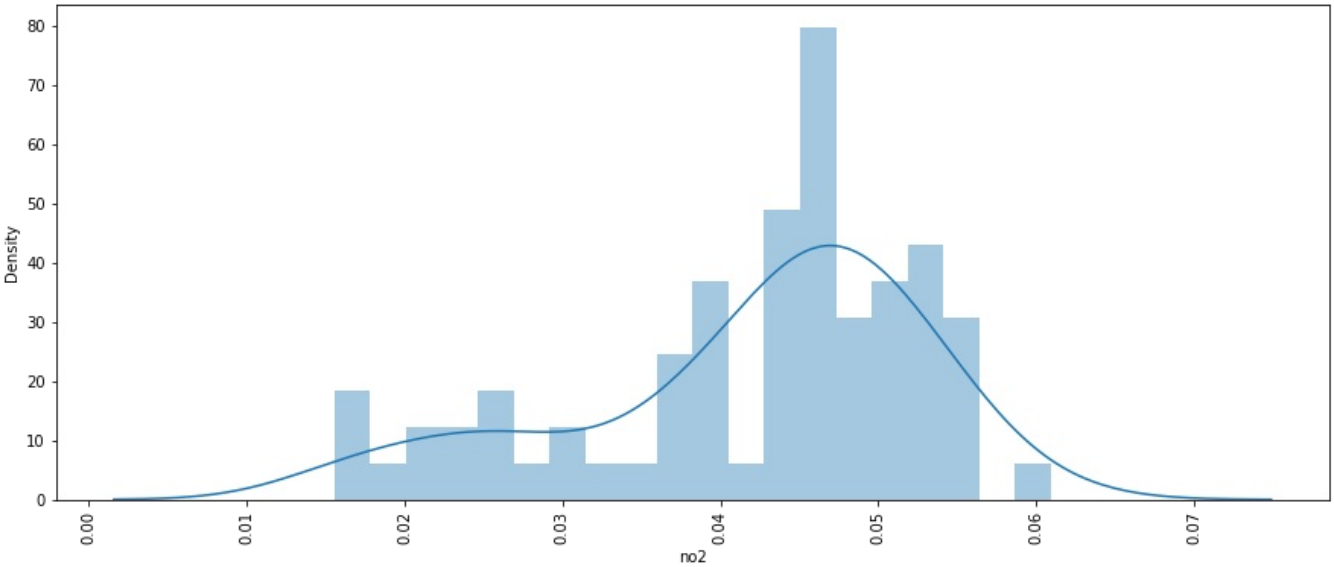


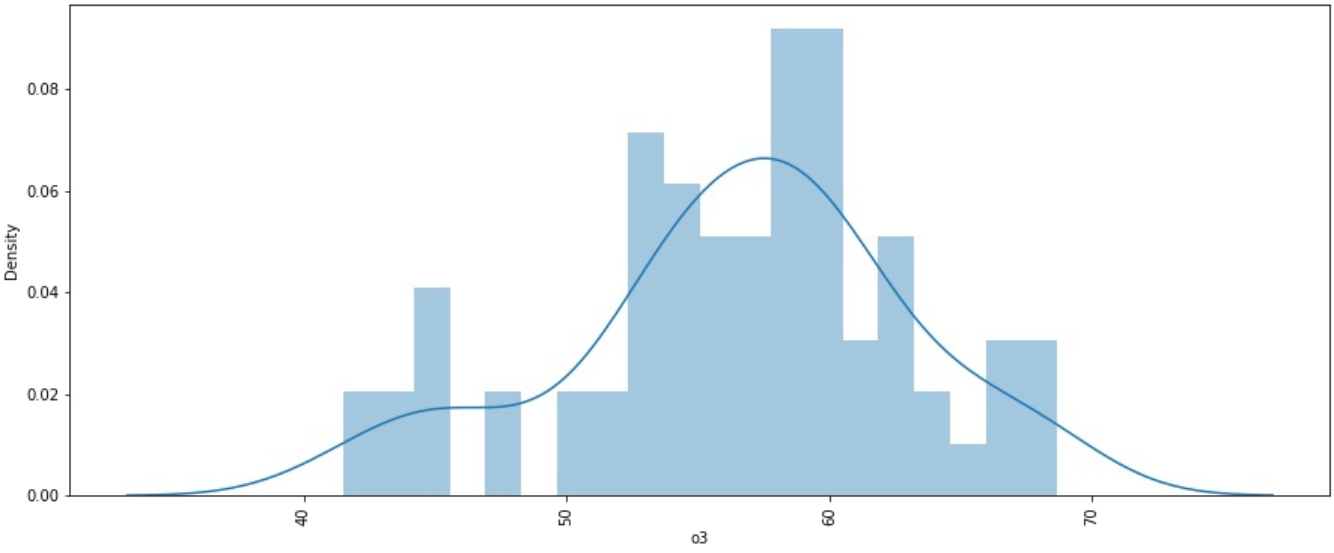


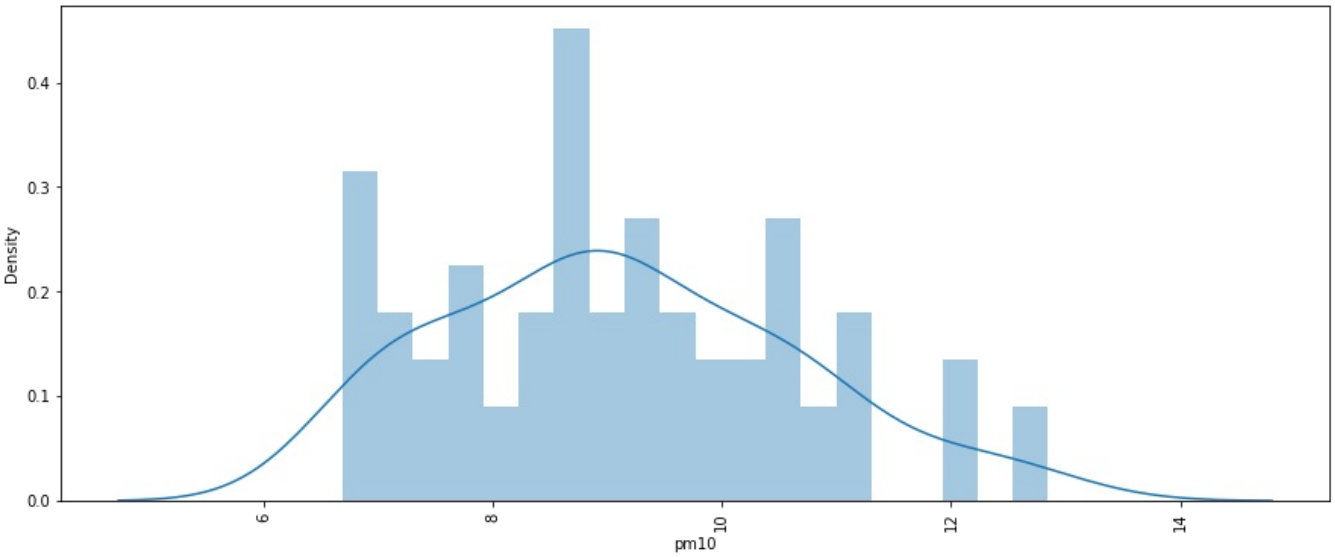
In [19]:
```
for i in numerical_columns:
    plt.figure(figsize=(15,6))
    sns.distplot(df[i], kde = True, bins = 20)
    plt.xticks(rotation = 90)
    plt.show()
```

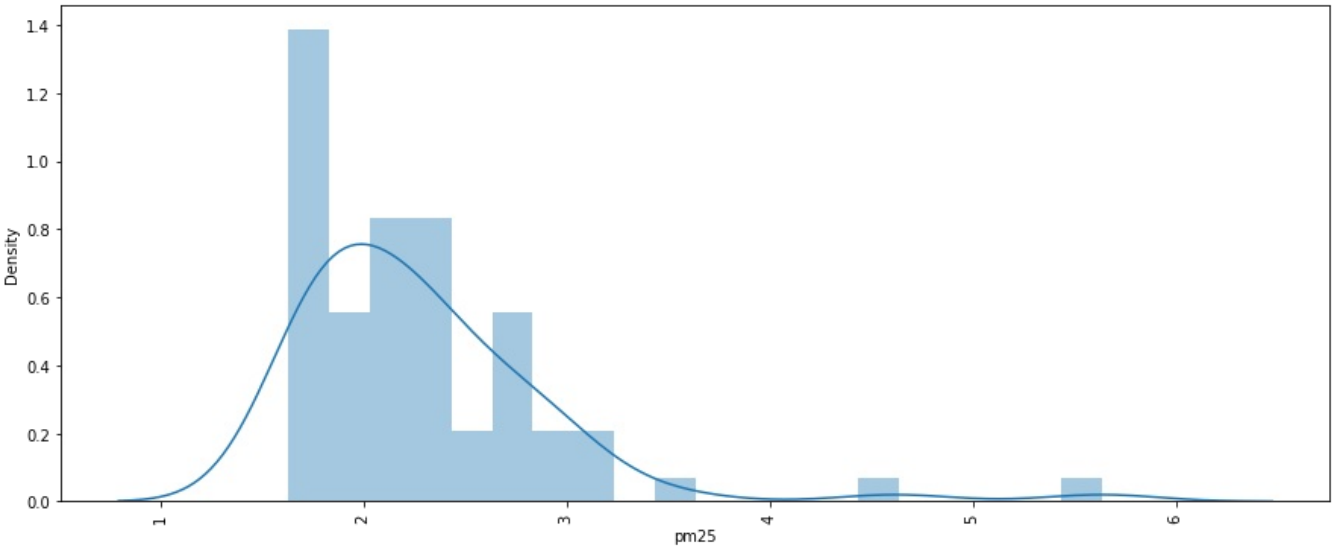


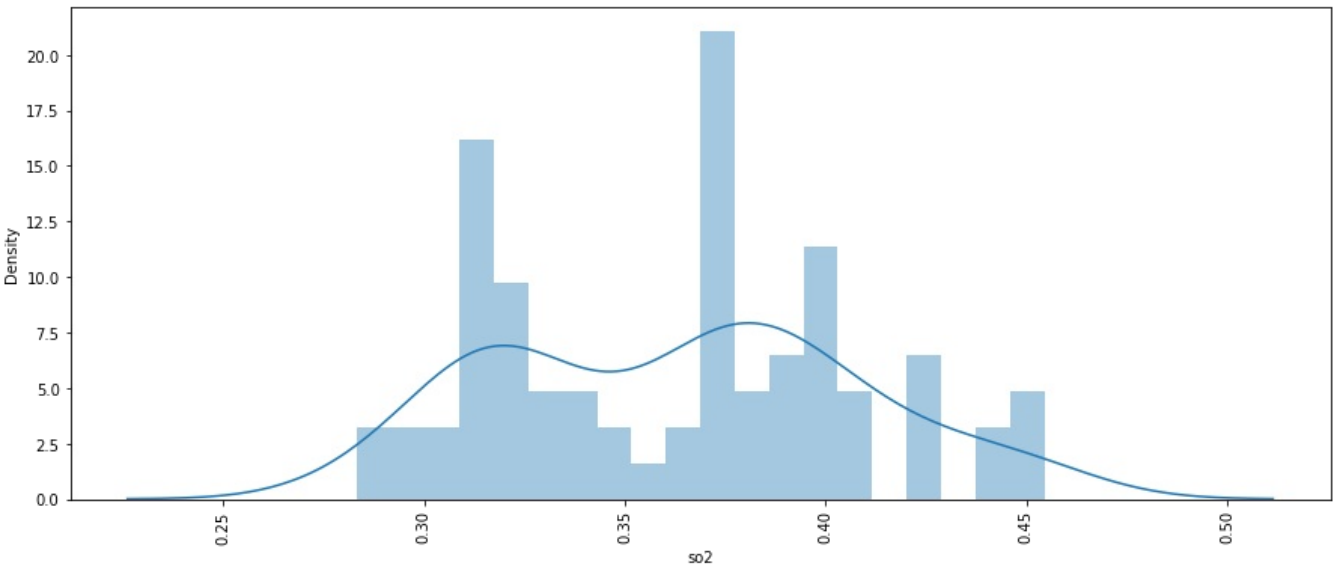

Density
0.25
0.20
0.15
0.10
0.05
0.00
185
190
195
200
205
210
215
CO

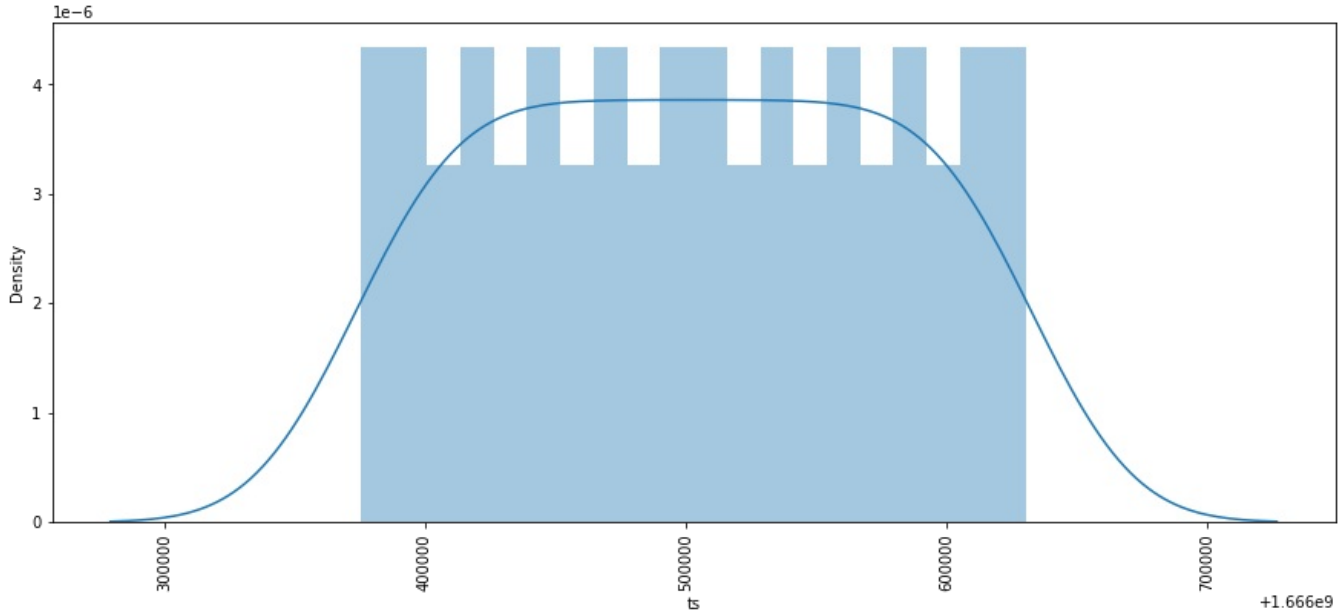

Density
80
70
60
50
40
30
20
10
0
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
no2


Density
0.08
0.06
0.04
0.02
0.00
40
50
60
70
o3

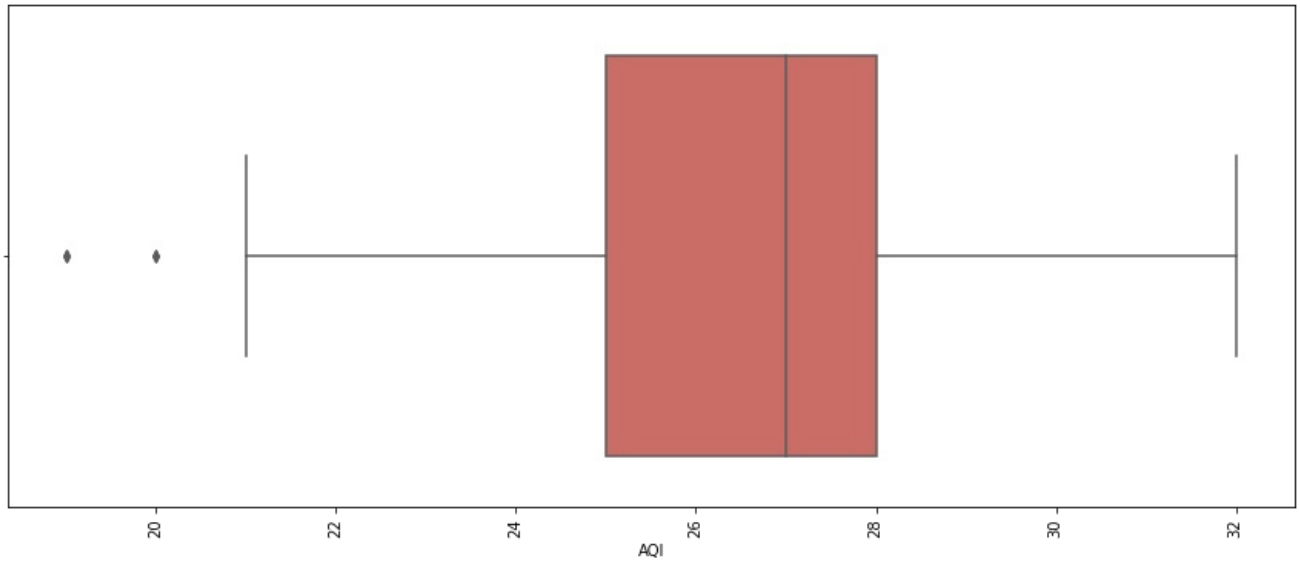

Density
0.0
0.1
0.2
0.3
0.4
6
8
10
12
14
pm10

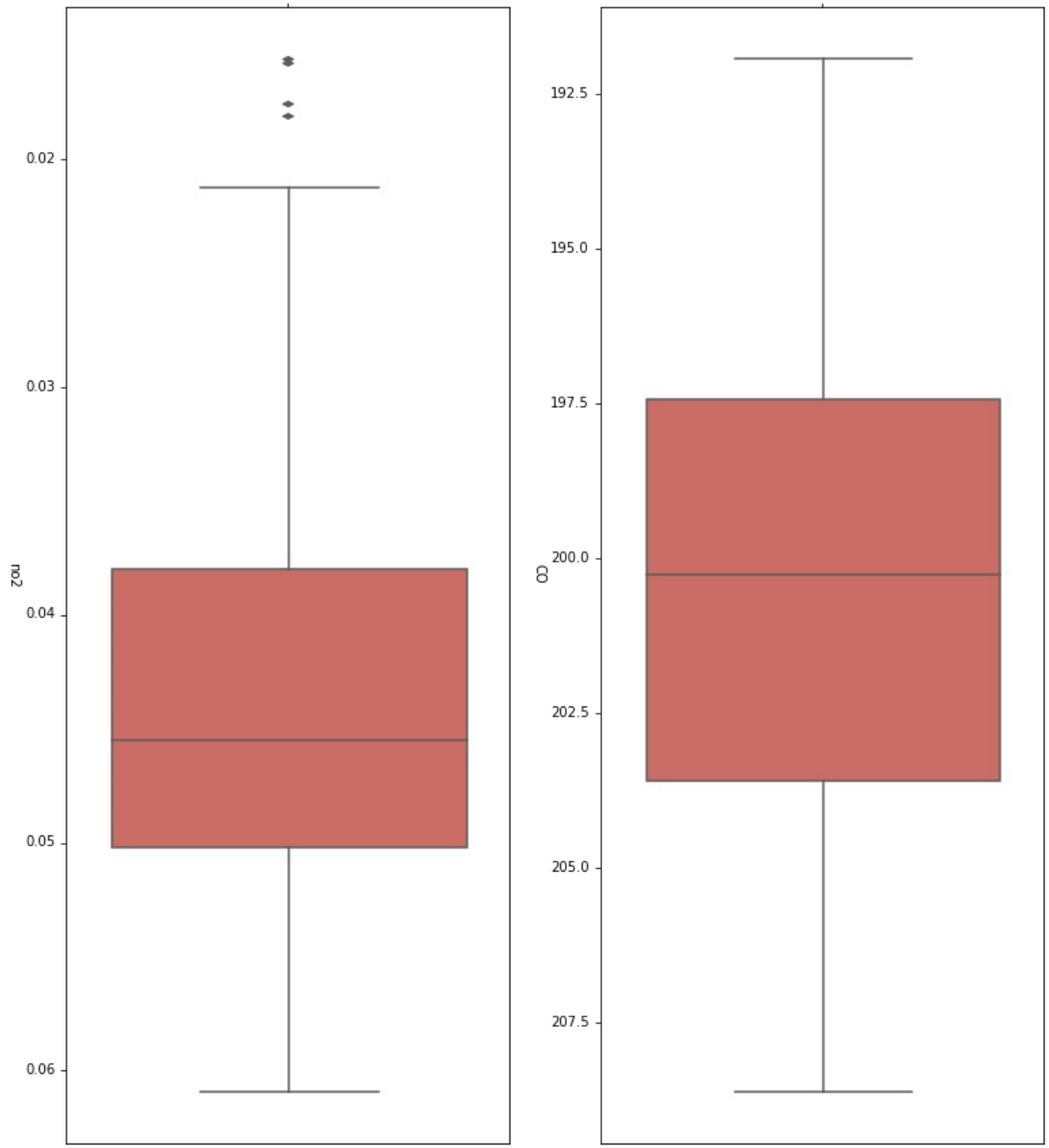

Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1
2
3
4
5
6
pm25

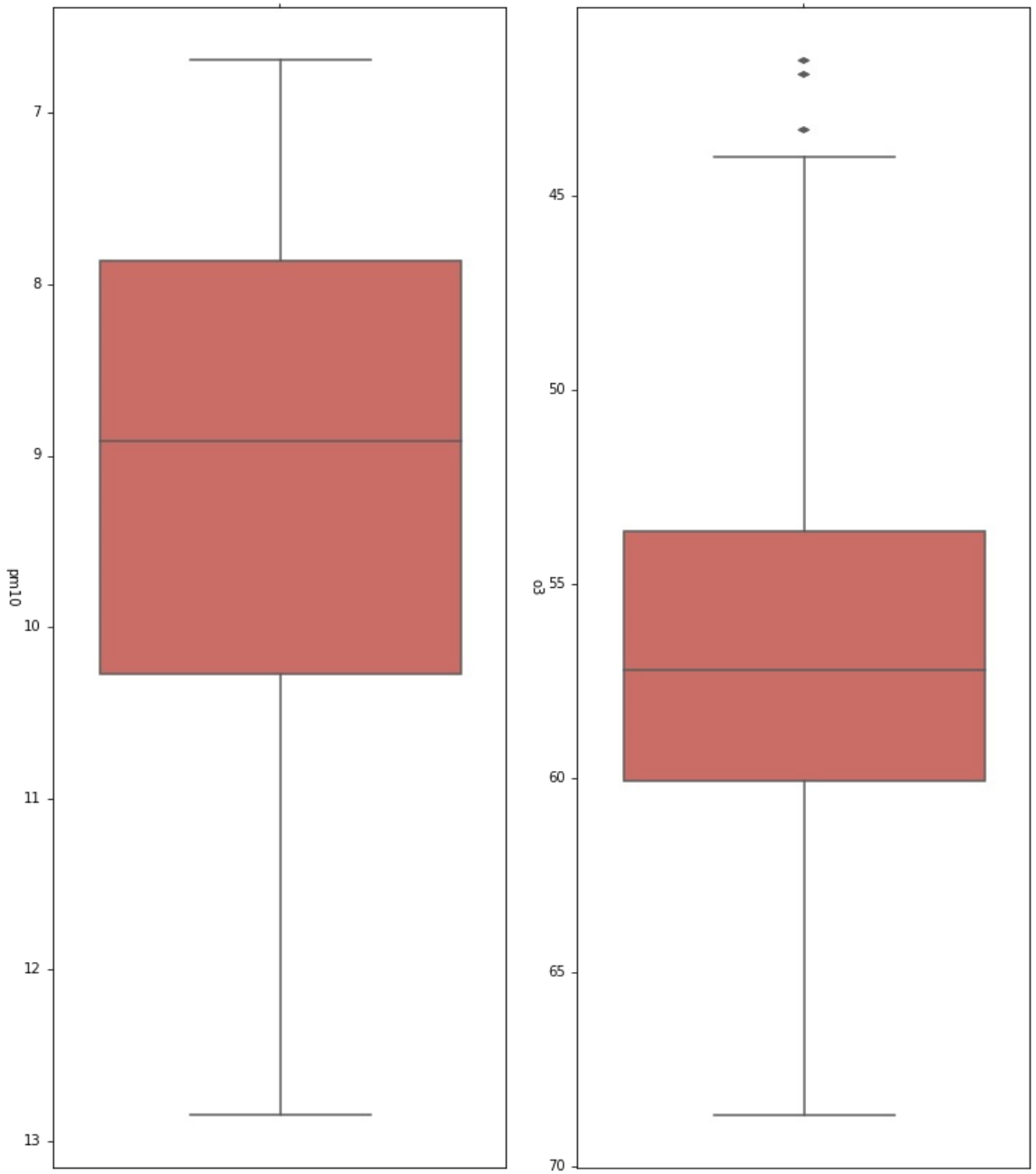

Density
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
0.25
0.30
0.35
0.40
0.45
0.50
so2

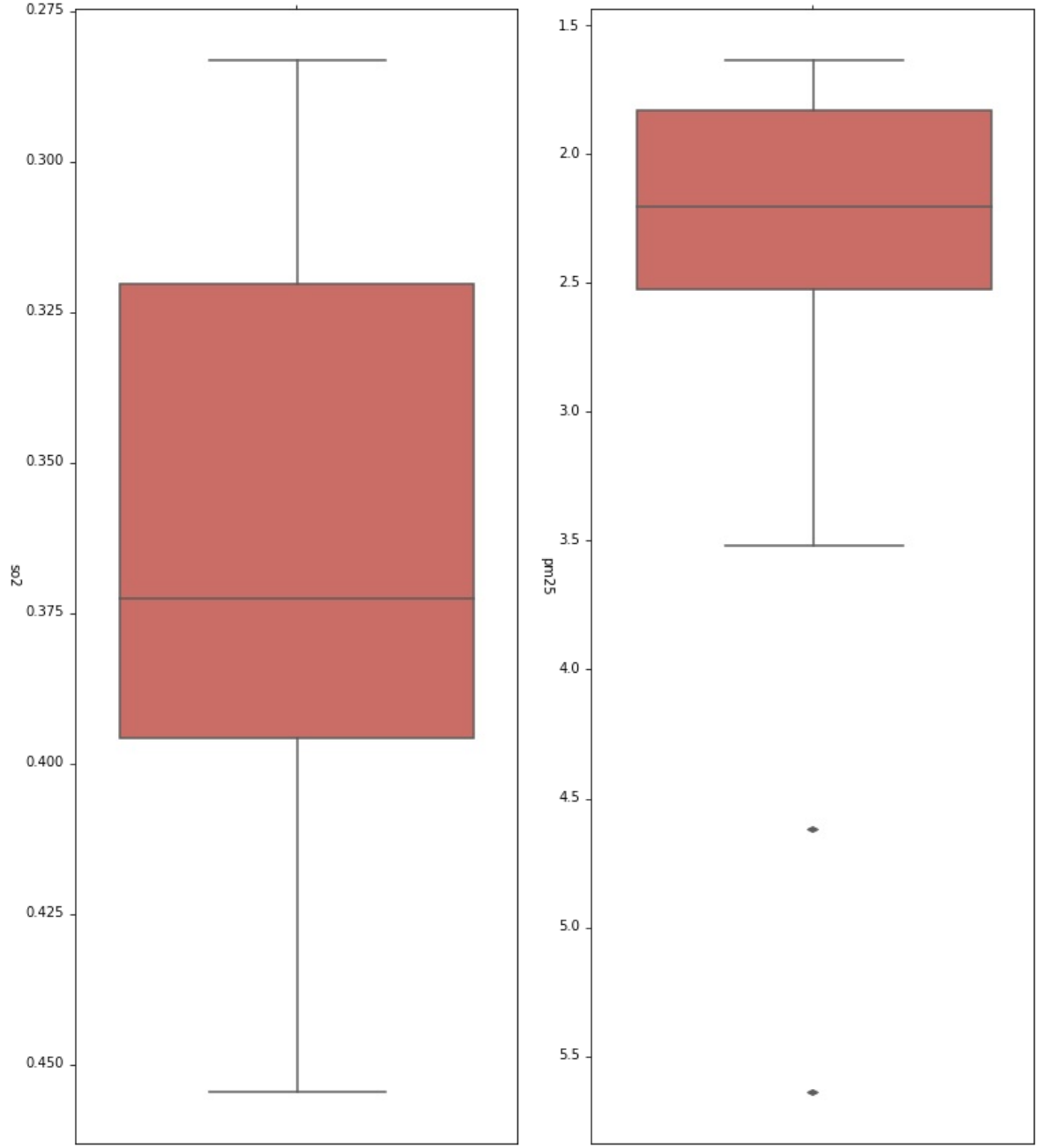


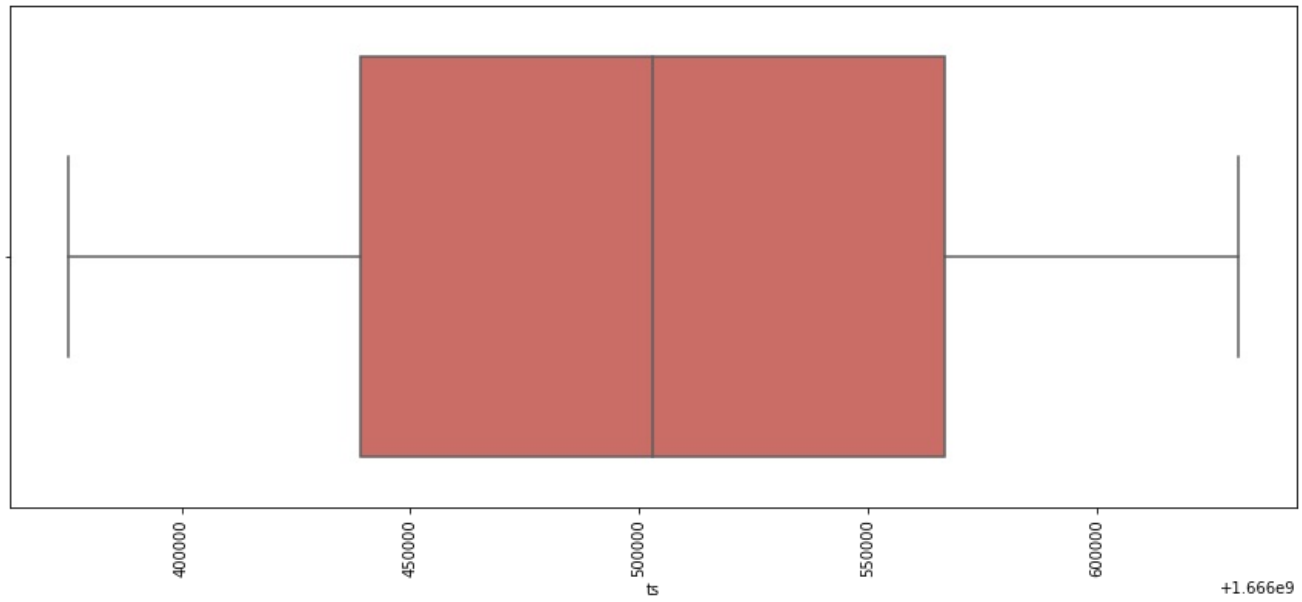
In [20]:
```
for i in numerical_columns:
    plt.figure(figsize=(15,6))
    sns.boxplot(x = df[i], data = df, palette = 'hls')
    plt.xticks(rotation = 90)
    plt.show()
```



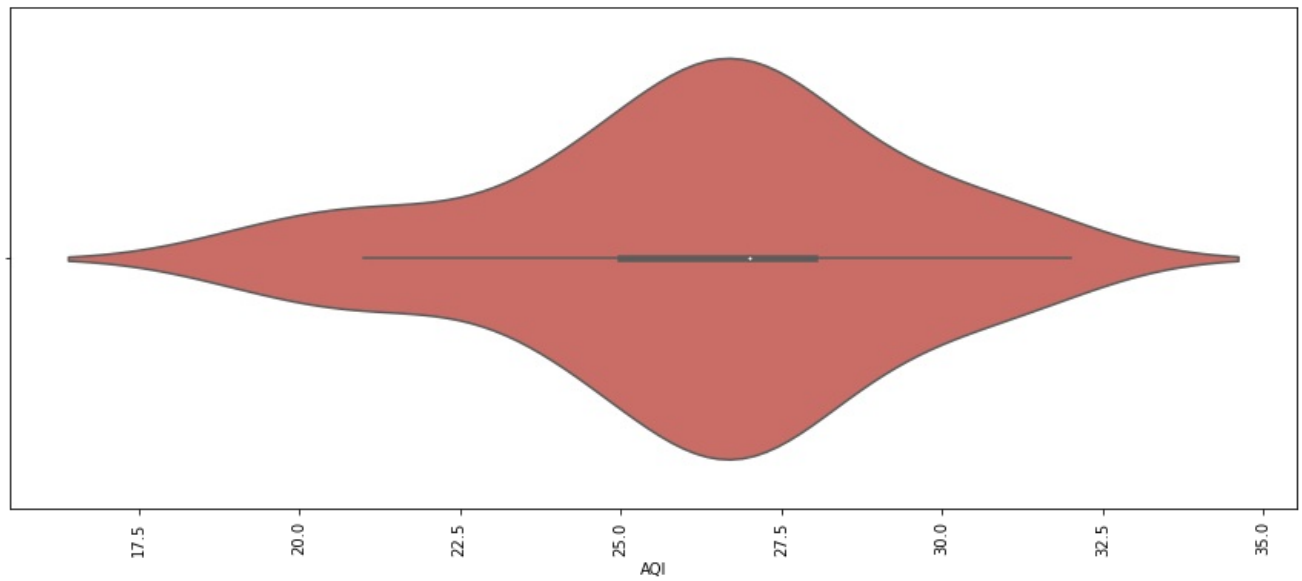

no2
0.02
0.03
0.04
0.05
0.06
CO
192.5
195.0
197.5
200.0
202.5
205.0
207.5

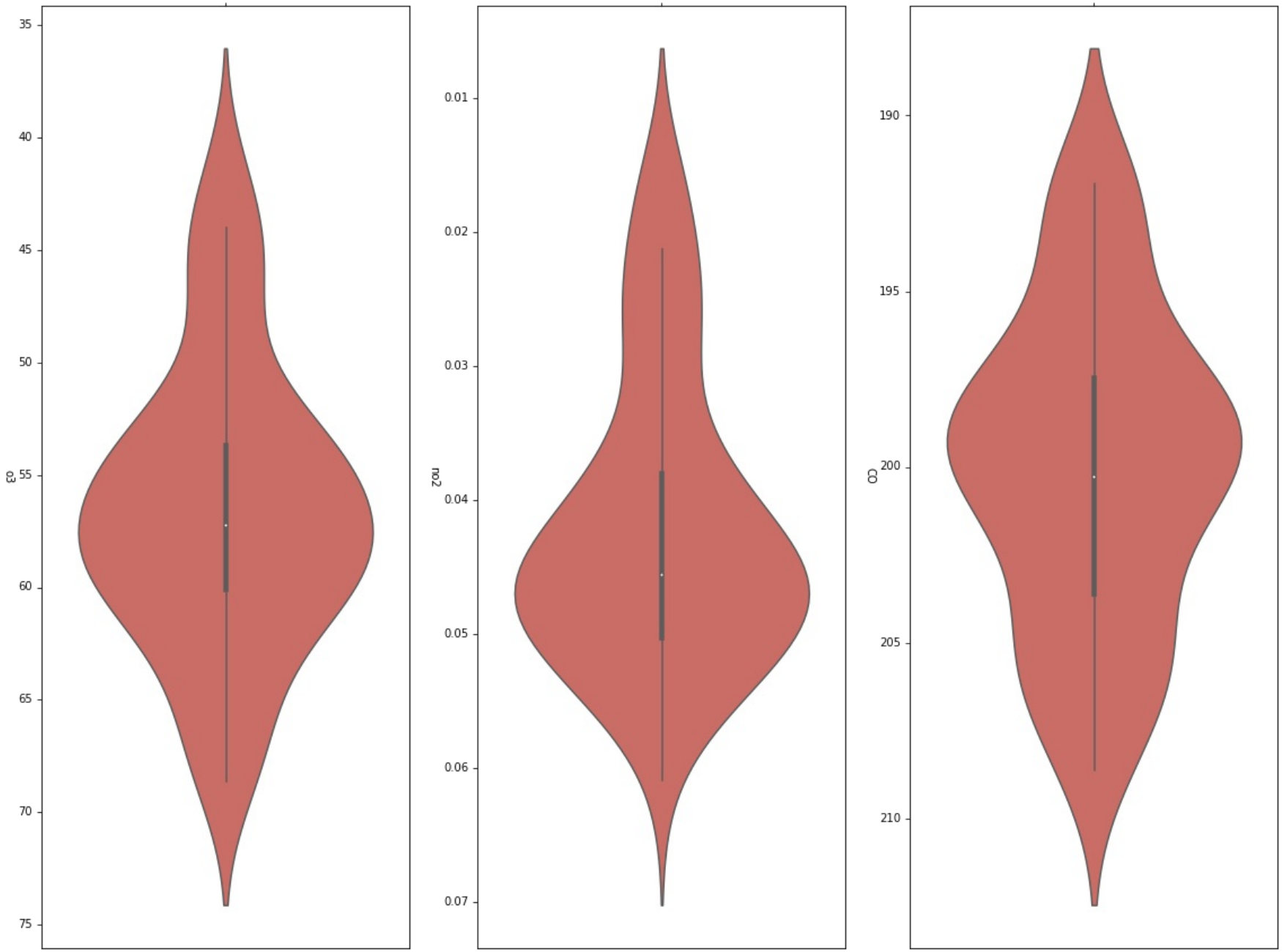

pm10
7
8
9
10
11
12
13
o3
45
50
55
60
65
70

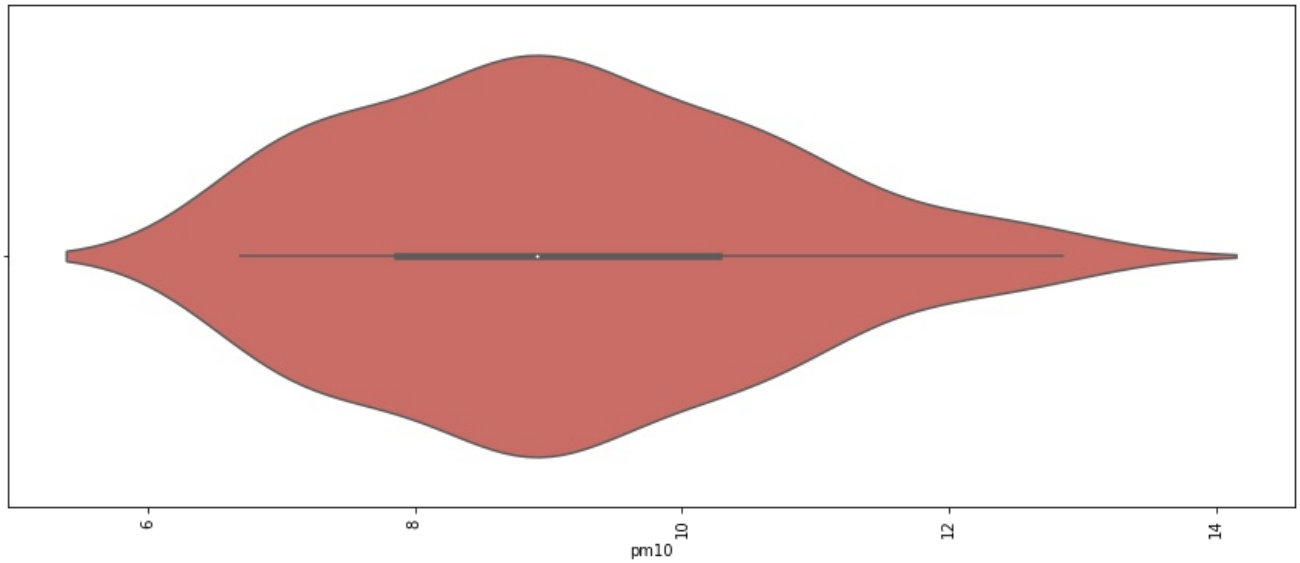

0.275
0.300
0.325
0.350
0.375
0.400
0.425
0.450
so2
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
pm25

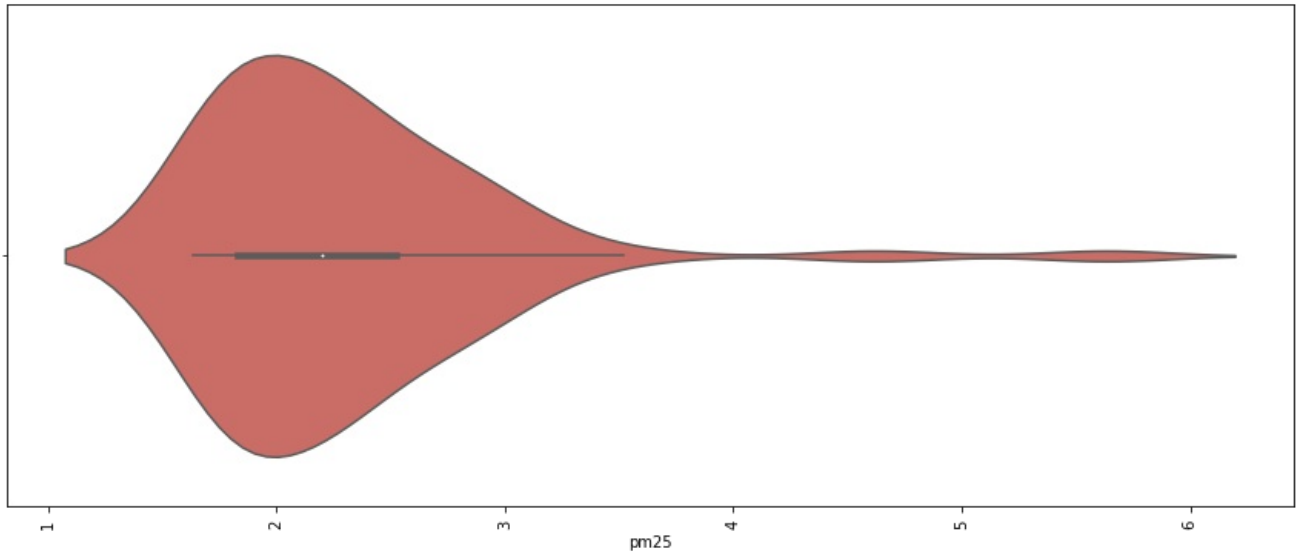


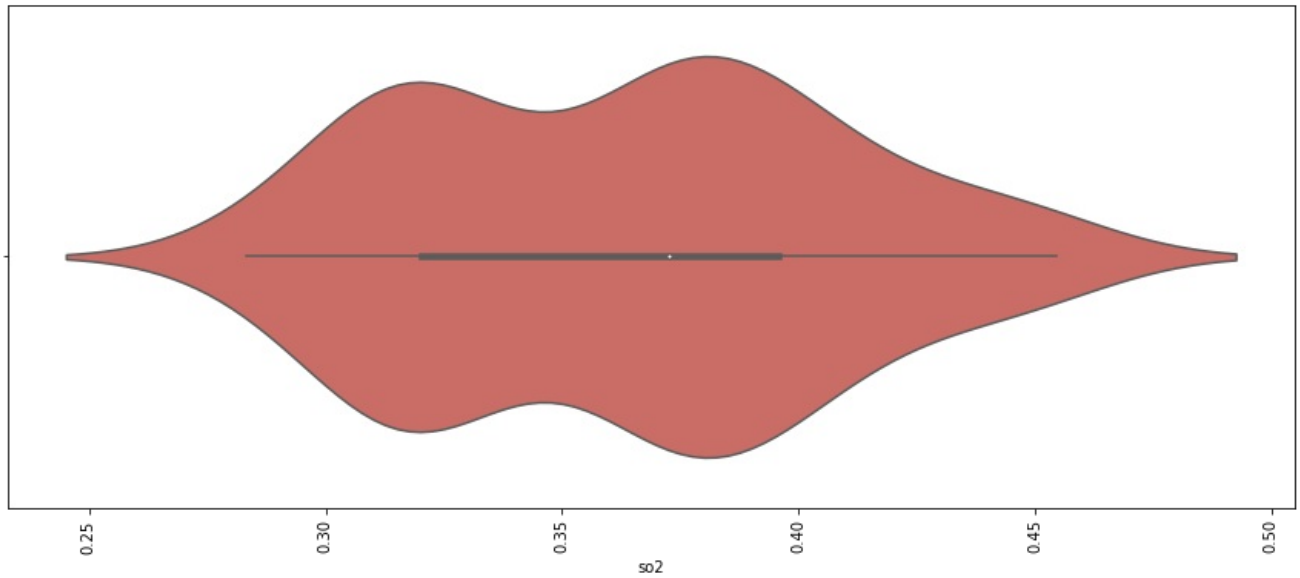
In [21]:
```
for i in numerical_columns:
    plt.figure(figsize=(15,6))
    sns.violinplot(x = df[i], data = df, palette = 'hls')
    plt.xticks(rotation = 90)
    plt.show()
```

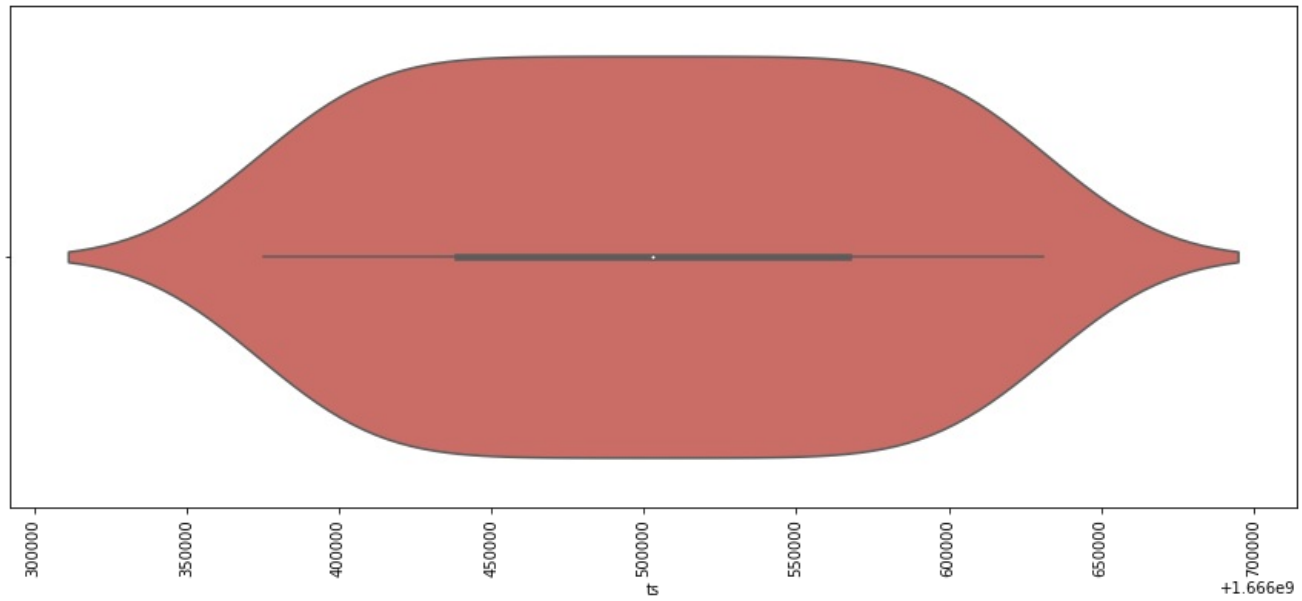



o3
35
40
45
50
55
60
65
70
75
no2
0.01
0.02
0.03
0.04
0.05
0.06
0.07
CO
190
195
200
205
210


6
8
10
12
14
pm10


1
2
3
4
5
6
pm25


0.25
0.30
0.35
0.40
0.45
0.50
so2



In [22]:
```
df['datetime'] = pd.to_datetime(df['datetime'], format='%Y-%m-%d:%H')
```

In [23]:
```
df['hour'] = df['datetime'].dt.hour
hourly_avg = df.groupby('hour').mean()
```

In [24]:
```
hourly_avg
```

Out[24]:

hour	AQI	CO	no2	o3	pm10	pm25	so2	ts
0	28.333333	204.165780	0.048865	61.035156	8.070345	1.873549	0.335276	1.666483e+09
1	27.000000	203.609470	0.050873	59.127808	8.195315	1.923857	0.347694	1.666487e+09
2	26.333333	201.940537	0.051097	56.982039	8.568985	2.088051	0.372529	1.666490e+09
3	25.666667	201.384230	0.045630	56.028367	8.528347	2.266779	0.392397	1.666494e+09
4	25.666667	200.827920	0.038825	55.313111	8.097564	2.385226	0.398606	1.666498e+09
5	25.333333	199.715297	0.033302	54.478646	8.034808	2.401974	0.393639	1.666501e+09
6	24.666667	200.271610	0.029509	54.240227	8.231443	2.360532	0.381221	1.666505e+09
7	24.666667	199.158987	0.026776	53.524971	8.518874	2.368372	0.372529	1.666508e+09
8	24.333333	199.158987	0.025939	53.167343	8.551964	2.359520	0.365078	1.666512e+09
9	24.666667	198.602680	0.026887	53.286552	8.557379	2.306958	0.357628	1.666516e+09
10	24.666667	197.490060	0.028784	53.524972	8.570801	2.245200	0.355144	1.666519e+09
11	24.333333	196.377440	0.033246	53.048134	9.016897	2.227348	0.358870	1.666523e+09
12	24.333333	195.821127	0.039605	52.809715	9.490121	2.234195	0.358870	1.666526e+09
13	24.666667	196.933750	0.045072	53.167343	10.073372	2.250574	0.356386	1.666530e+09
14	25.333333	198.046370	0.046299	55.074692	10.705149	2.256721	0.355144	1.666534e+09
15	26.333333	199.158990	0.047973	56.743622	10.994544	2.272842	0.353903	1.666537e+09
16	26.666667	199.158987	0.048642	58.174134	10.555905	2.284107	0.357628	1.666541e+09
17	27.333333	199.158987	0.048419	59.127808	9.946347	2.280922	0.361353	1.666544e+09
18	28.066667	199.158987	0.047750	57.935715	9.304162	3.262985	0.371287	1.666462e+09
19	27.733333	199.381513	0.047839	58.364868	9.618883	2.942286	0.378738	1.666465e+09
20	27.533333	200.827917	0.048642	59.223175	9.896310	2.565869	0.370046	1.666469e+09
21	28.000000	202.496847	0.049312	60.558317	9.711418	2.093676	0.356386	1.666472e+09
22	28.333333	204.165777	0.049312	61.273575	9.112165	1.958272	0.339001	1.666476e+09
23	28.333333	205.278397	0.049312	61.511993	8.409818	1.879931	0.332793	1.666480e+09

In [25]:
```
df.set_index('datetime', inplace=True)
```

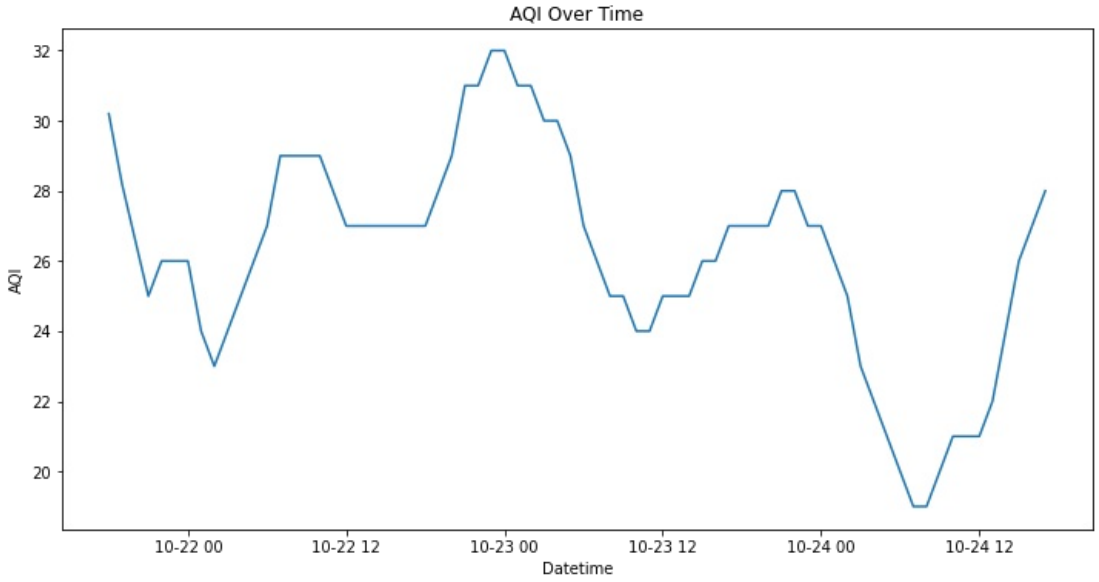
In [26]:
```
df
```

Out[26]:

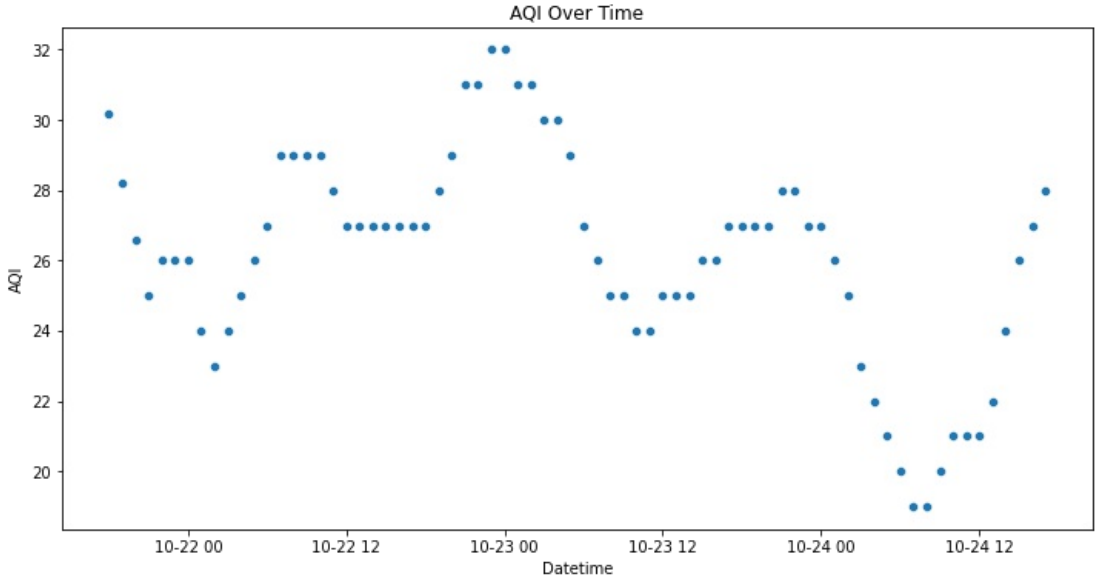
datetime	AQI	CO	no2	o3	pm10	pm25	so2	timestamp_local	timestamp_utc	ts	hour
2022-10-21 18:00:00	30.2	198.60268	0.046857	55.789948	10.486722	5.637410	0.387430	2022-10-21T23:00:00	2022-10-21T18:00:00	1666375200	18
2022-10-21 19:00:00	28.2	197.60132	0.046456	54.931640	10.719325	4.618169	0.409782	2022-10-22T00:00:00	2022-10-21T19:00:00	1666378800	19
2022-10-21 20:00:00	26.6	198.60268	0.046857	54.645540	11.155578	3.520902	0.402331	2022-10-22T01:00:00	2022-10-21T20:00:00	1666382400	20
2022-10-21 21:00:00	25.0	201.94054	0.048196	55.074690	11.116206	2.225919	0.376254	2022-10-22T02:00:00	2022-10-21T21:00:00	1666386000	21
2022-10-21 22:00:00	26.0	205.27840	0.048865	55.789948	10.405250	1.979471	0.339001	2022-10-22T03:00:00	2022-10-21T22:00:00	1666389600	22
...	...	...	...	...	...	...	...	...	...	...	...
2022-10-24 13:00:00	22.0	193.59589	0.035478	47.206880	10.423121	2.720472	0.391155	2022-10-24T18:00:00	2022-10-24T13:00:00	1666616400	13
2022-10-24 14:00:00	24.0	195.26482	0.039159	52.213670	11.240661	2.713109	0.383705	2022-10-24T19:00:00	2022-10-24T14:00:00	1666620000	14
2022-10-24 15:00:00	26.0	196.93375	0.044180	56.505203	12.098125	2.743044	0.376254	2022-10-24T20:00:00	2022-10-24T15:00:00	1666623600	15
2022-10-24 16:00:00	27.0	198.60268	0.046857	59.366226	12.845977	2.780022	0.372529	2022-10-24T21:00:00	2022-10-24T16:00:00	1666627200	16
2022-10-24 17:00:00	28.0	198.60268	0.047527	60.796738	12.583639	2.745875	0.368804	2022-10-24T22:00:00	2022-10-24T17:00:00	1666630800	17

72 rows × 11 columns

In [27]:
```
plt.figure(figsize=(12, 6))
sns.lineplot(x=df.index, y=df['AQI'])
plt.title('AQI Over Time')
plt.xlabel('Datetime')
plt.ylabel('AQI')
plt.show()
```

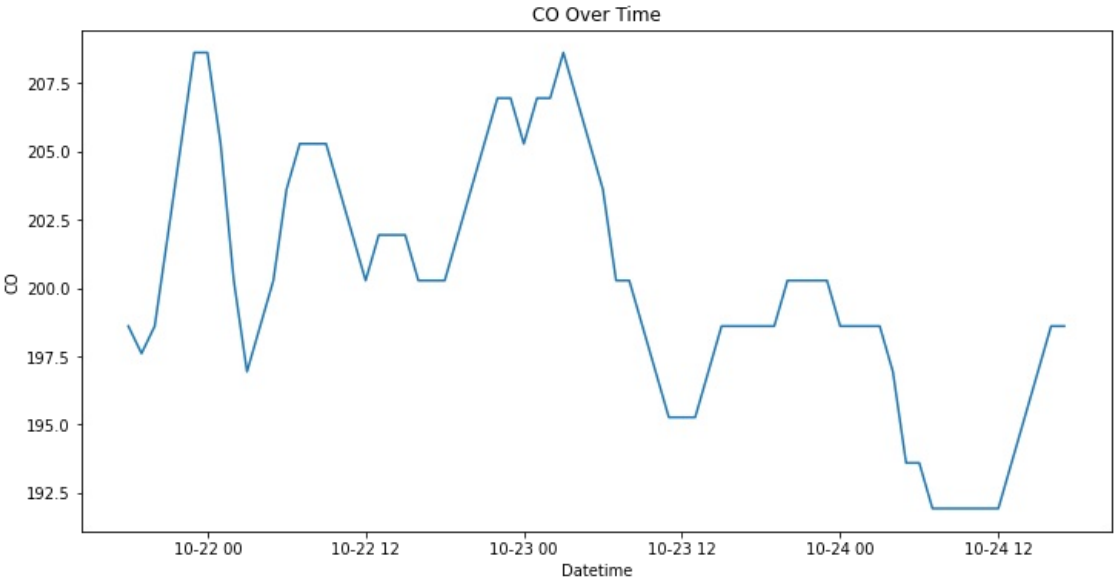


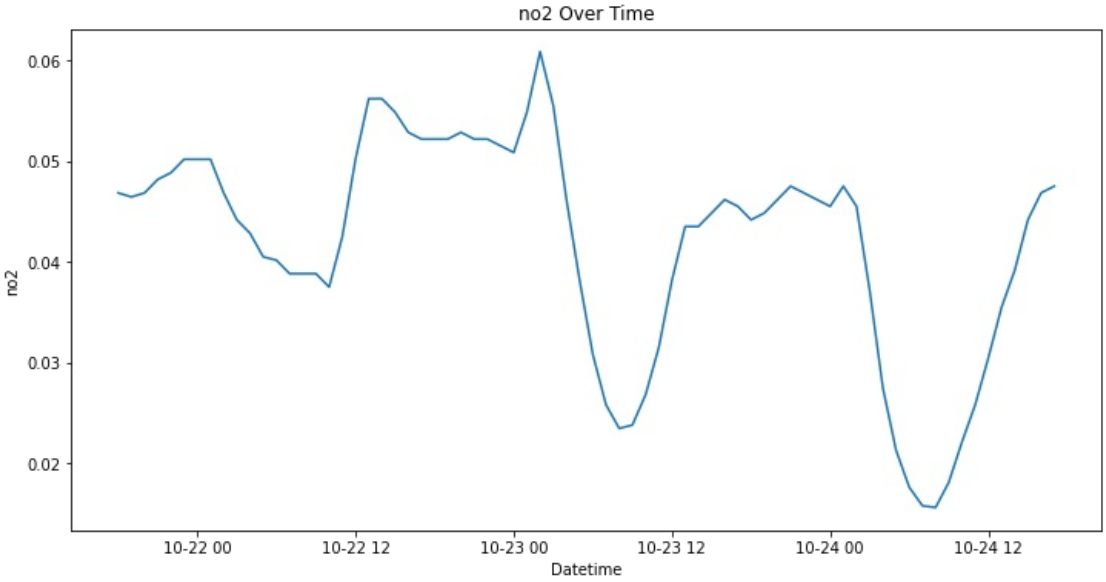
In [28]:
```
plt.figure(figsize=(12, 6))
sns.scatterplot(x=df.index, y=df['AQI'])
plt.title('AQI Over Time')
plt.xlabel('Datetime')
plt.ylabel('AQI')
plt.show()
```

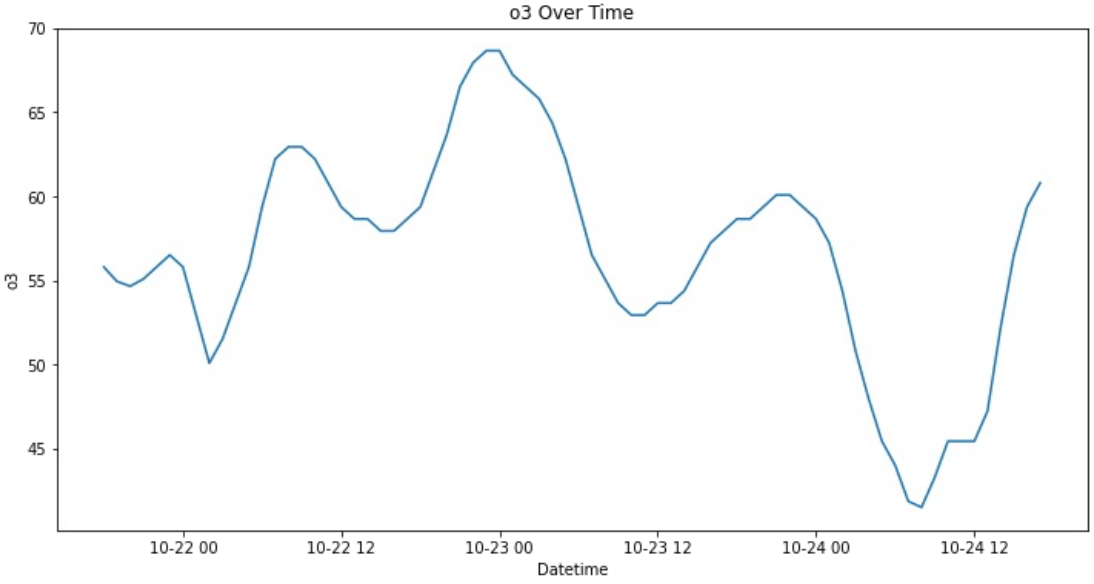


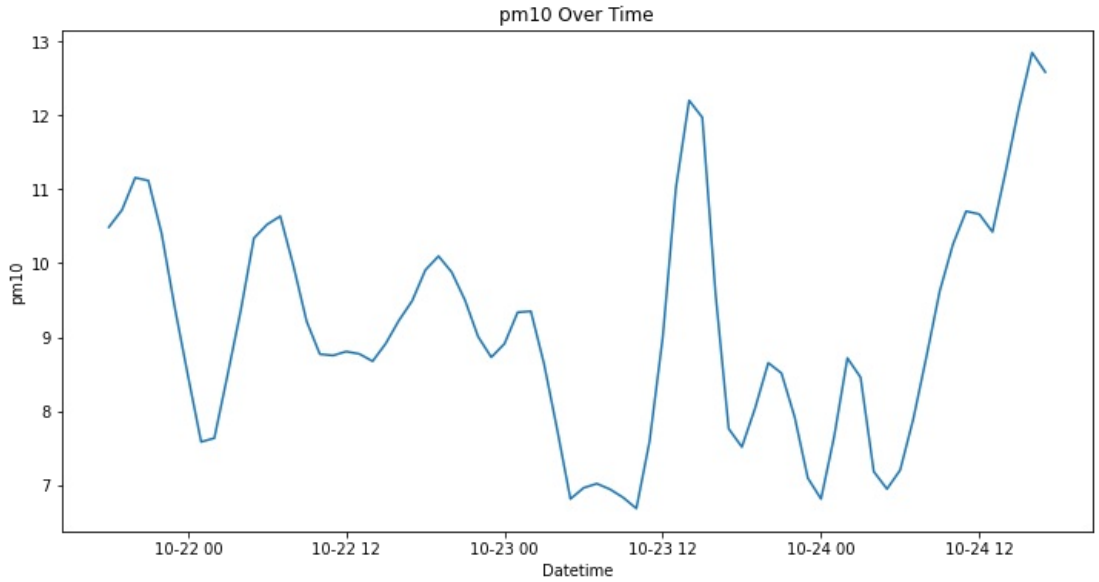
In [29]:
```
chemical_compounds = ['CO', 'no2', 'o3', 'pm10', 'pm25', 'so2']
```

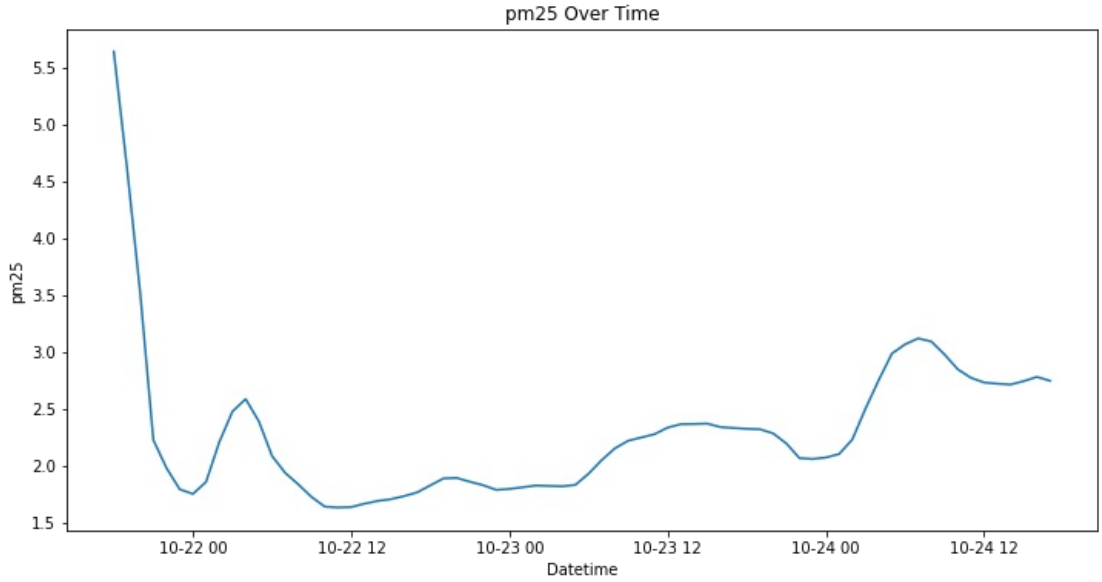
In [30]:
```
for i in chemical_compounds:
    plt.figure(figsize=(12, 6))
    sns.lineplot(x=df.index, y=df[i])
    plt.title(f'{i} Over Time')
    plt.xlabel('Datetime')
    plt.ylabel(i)
    plt.show()
```

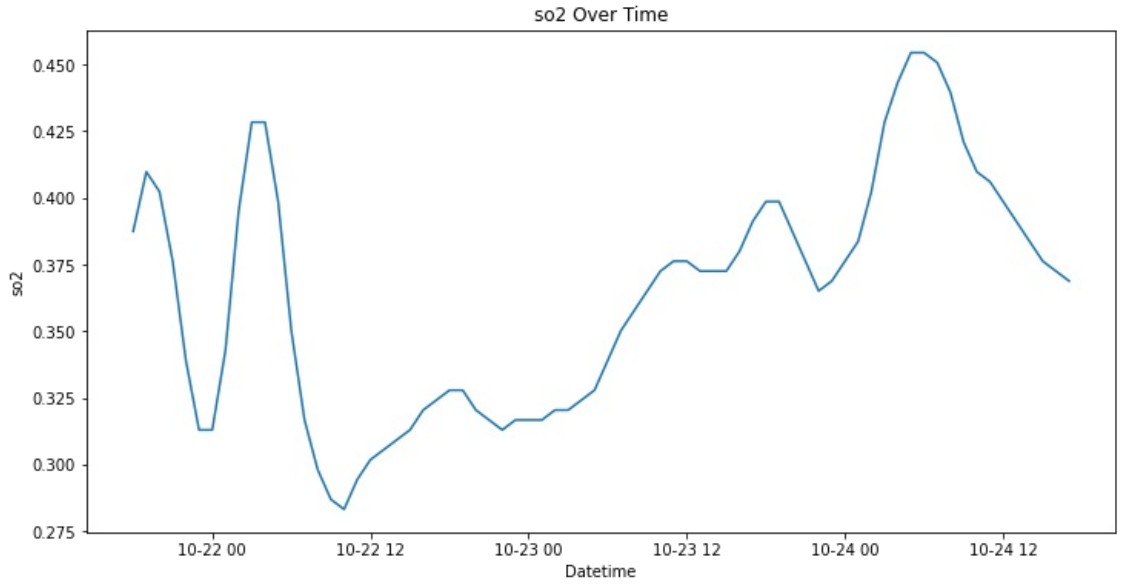

CO Over Time
CO
207.5
205.0
202.5
200.0
197.5
195.0
192.5
10-22 00
10-22 12
10-23 00
10-23 12
10-24 00
10-24 12
Datetime


no2 Over Time
no2
0.06
0.05
0.04
0.03
0.02
10-22 00
10-22 12
10-23 00
10-23 12
10-24 00
10-24 12
Datetime

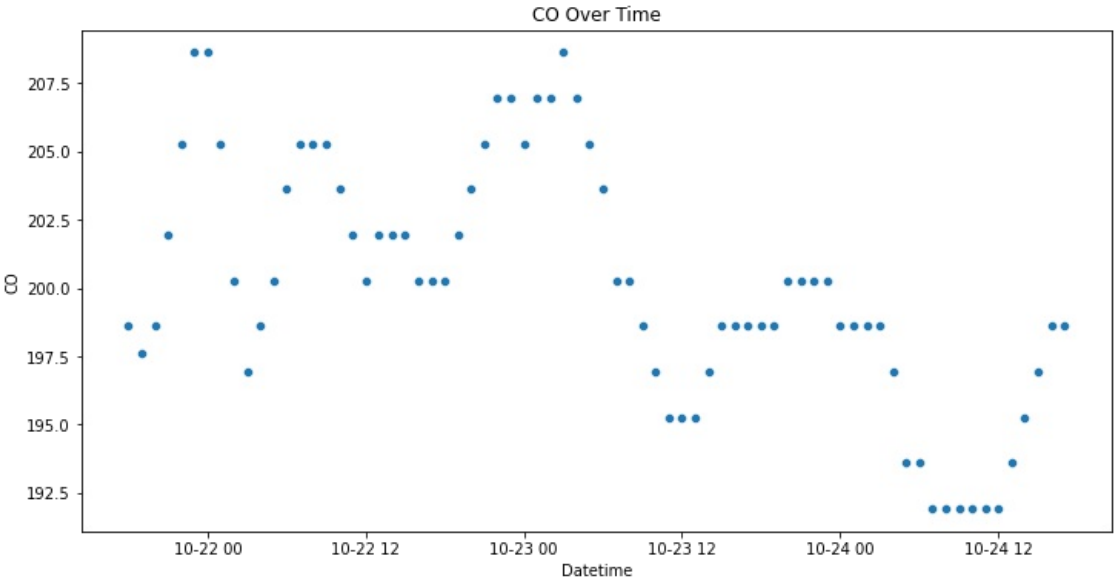

o3 Over Time
o3
70
65
60
55
50
45
10-22 00
10-22 12
10-23 00
10-23 12
10-24 00
10-24 12
Datetime

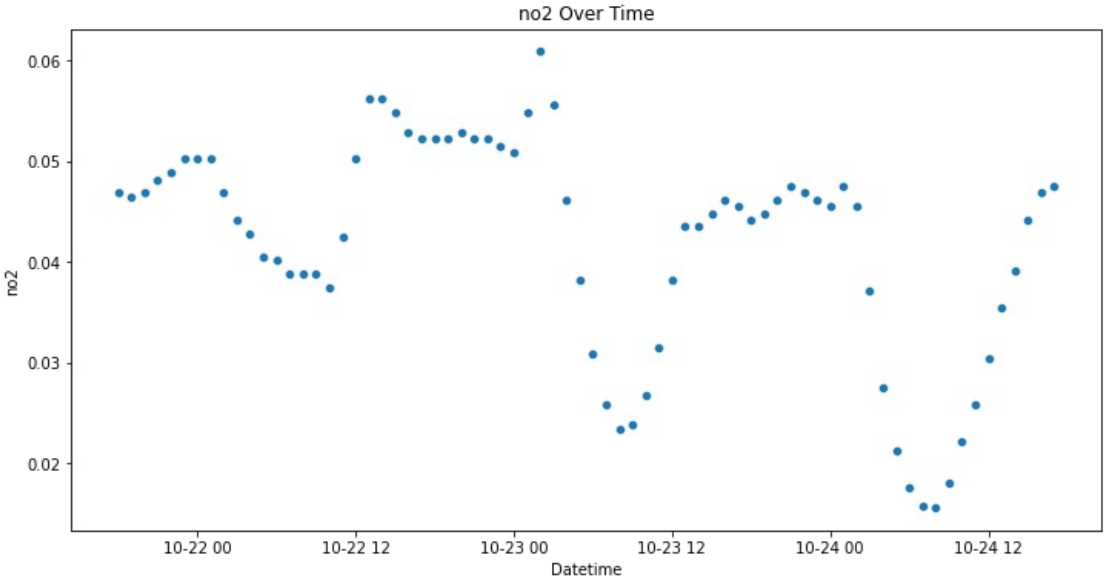


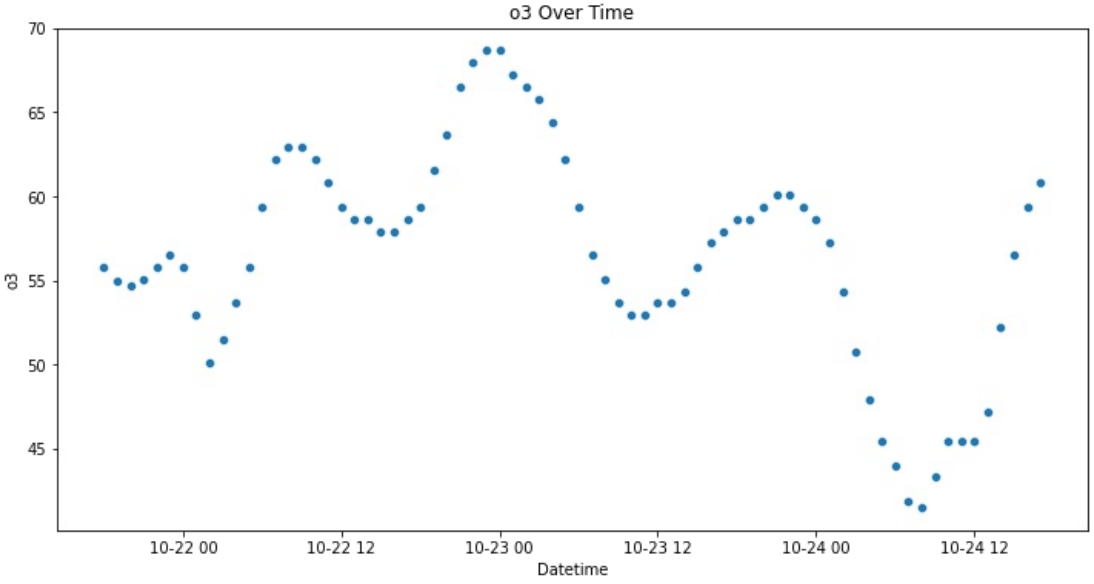


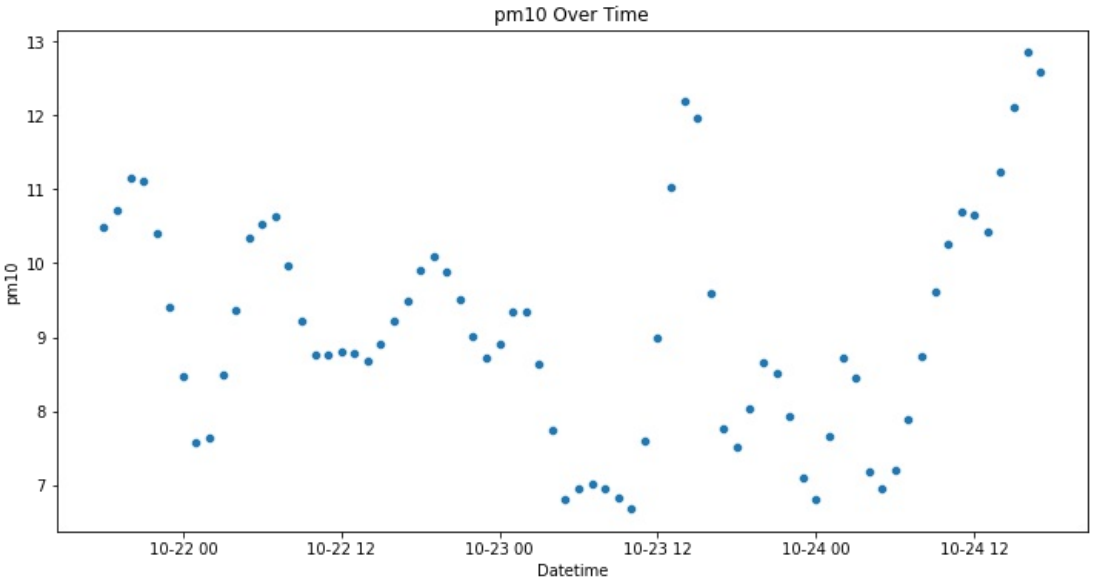


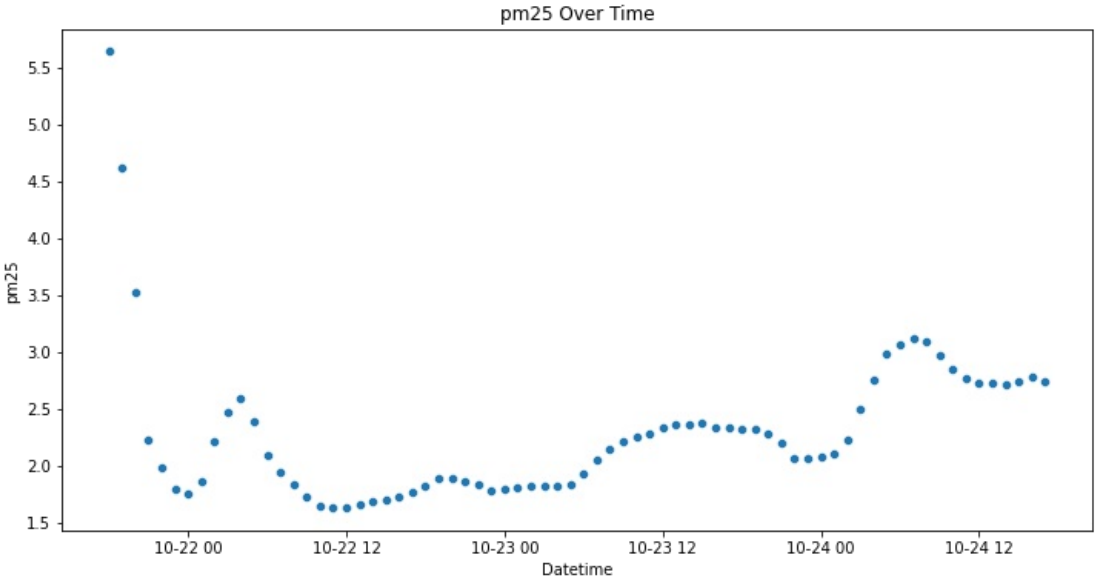
In [31]:
```
for i in chemical_compounds:
    plt.figure(figsize=(12, 6))
    sns.scatterplot(x=df.index, y=df[i])
    plt.title(f'{i} Over Time')
    plt.xlabel('Datetime')
    plt.ylabel(i)
    plt.show()
```

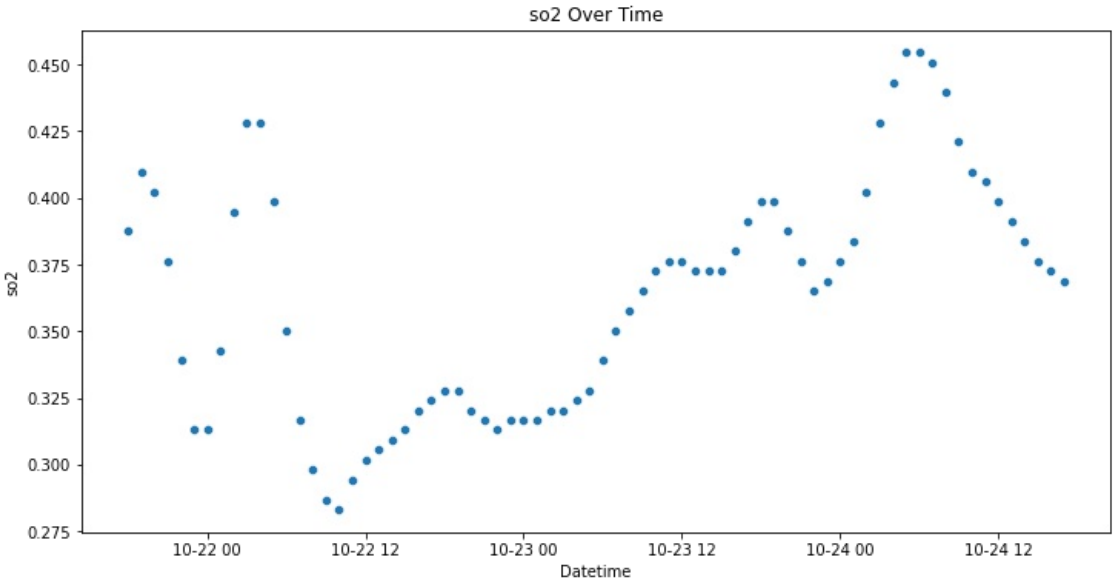

CO Over Time
CO
207.5
205.0
202.5
200.0
197.5
195.0
192.5
10-22 00
10-22 12
10-23 00
10-23 12
10-24 00
10-24 12
Datetime


no2 Over Time
no2
0.06
0.05
0.04
0.03
0.02
10-22 00
10-22 12
10-23 00
10-23 12
10-24 00
10-24 12
Datetime


o3 Over Time
o3
70
65
60
55
50
45
10-22 00
10-22 12
10-23 00
10-23 12
10-24 00
10-24 12
Datetime

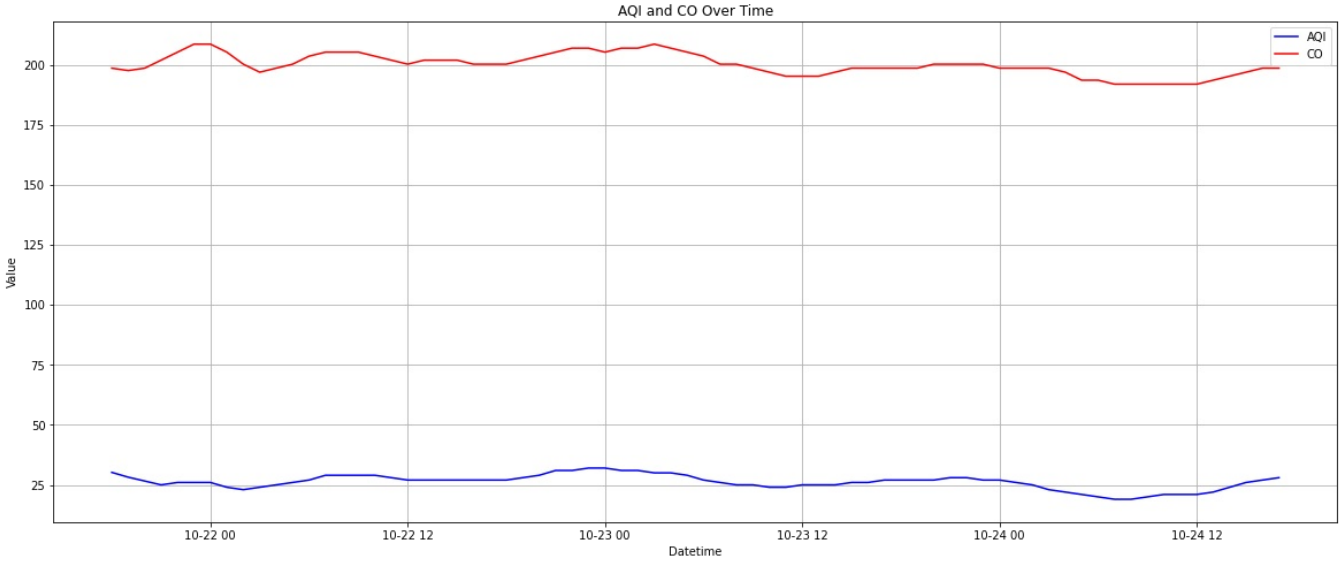


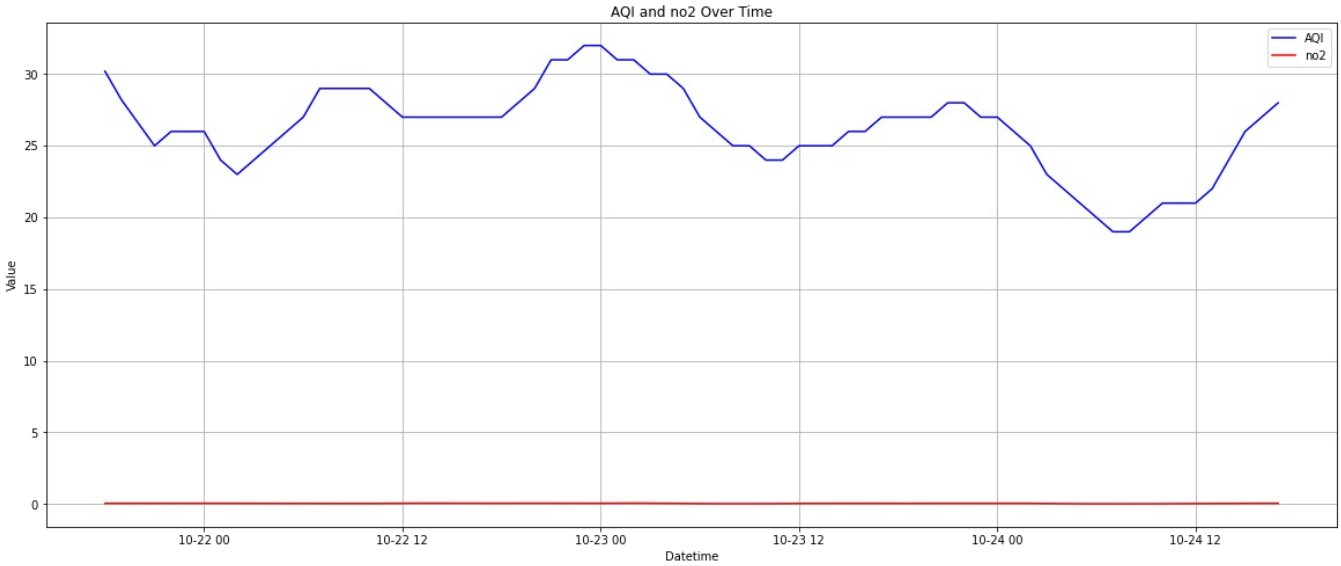


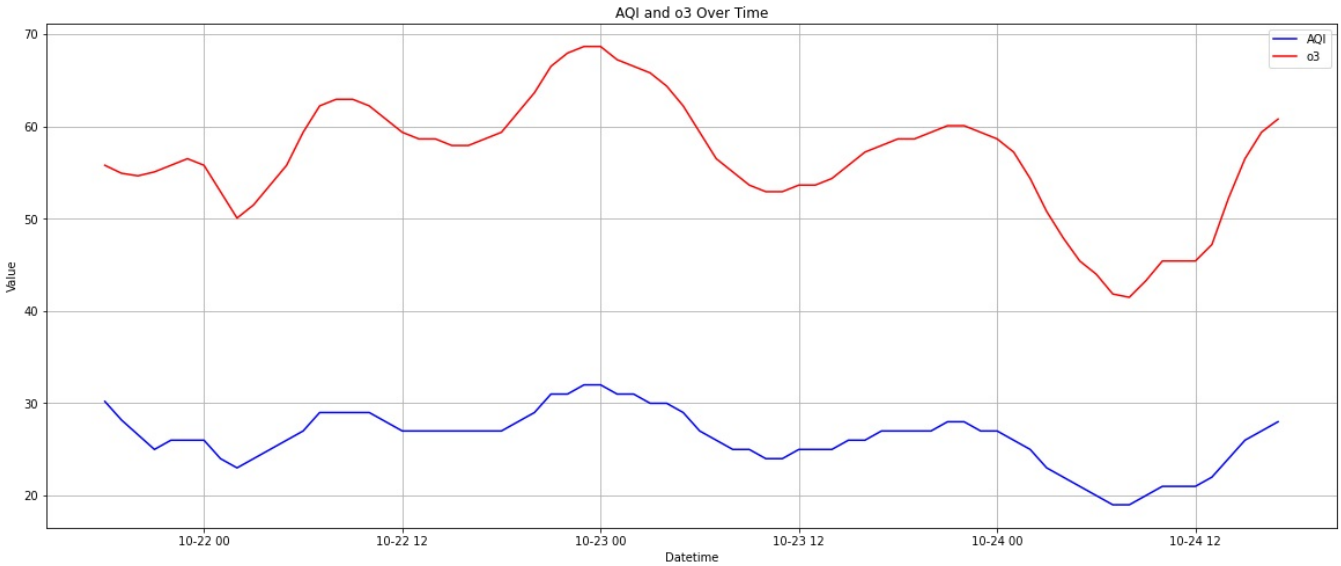


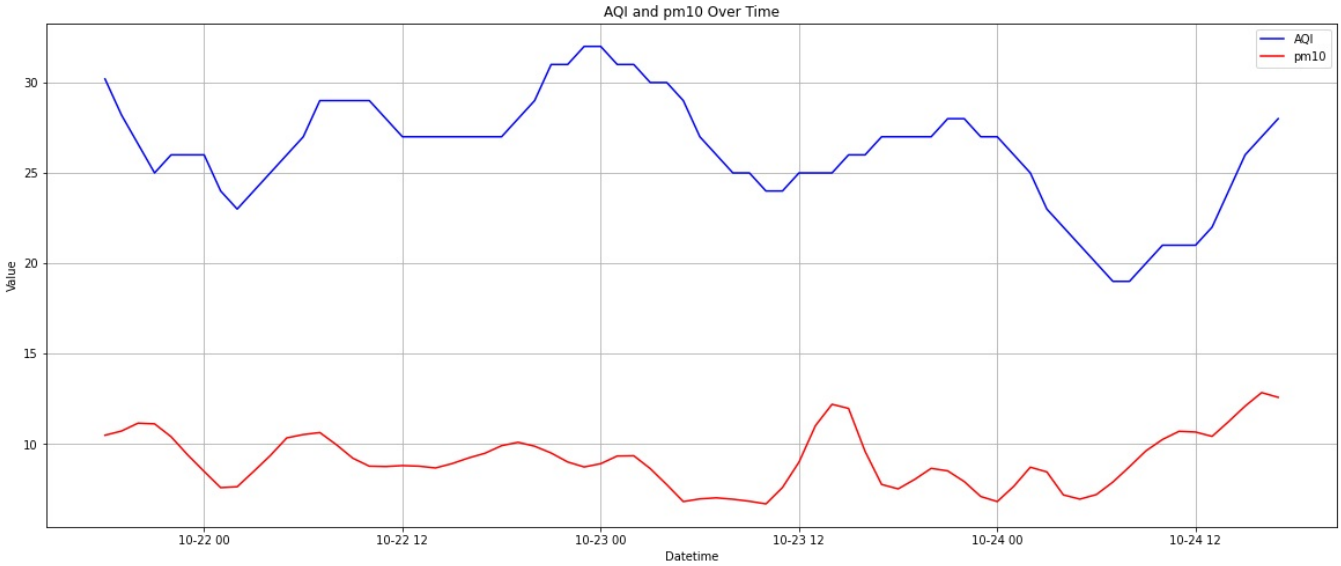
In [32]:

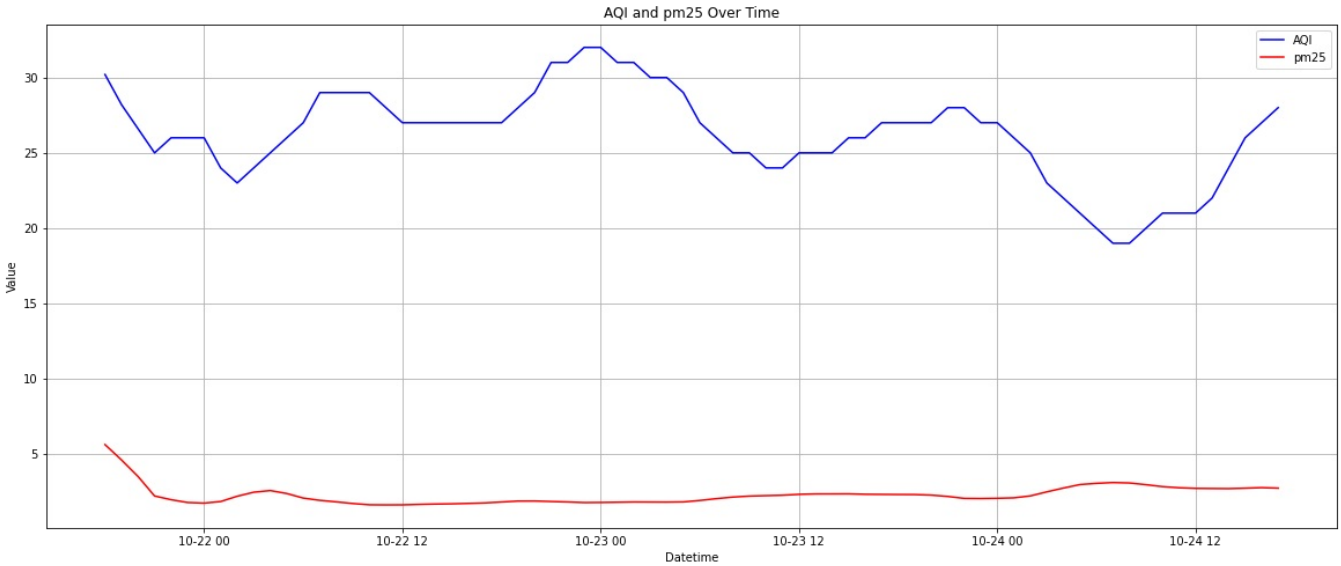
```
for i in chemical_compounds:
    plt.figure(figsize=(20, 8))
    sns.lineplot(x=df.index, y='AQI', data=df, label='AQI', color='blue')
    sns.lineplot(x=df.index, y=df[i], data=df, label=i, color='red')
    plt.xlabel('Datetime')
    plt.ylabel('Value')
    plt.title(f'AQI and {i} Over Time')
    plt.legend()
    plt.grid()
    plt.show()
```

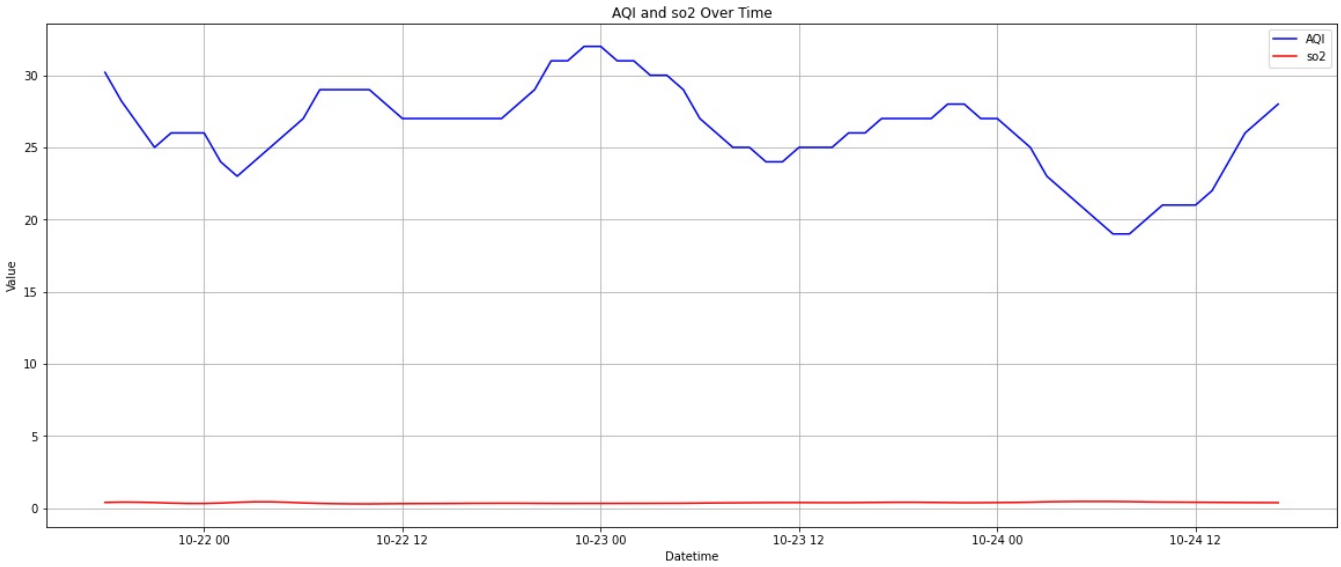

AQI and CO Over Time
AQI
CO
Value
Datetime
10-22 00
10-22 12
10-23 00
10-23 12
10-24 00
10-24 12


AQI and no2 Over Time
AQI
no2
Value
Datetime
10-22 00
10-22 12
10-23 00
10-23 12
10-24 00
10-24 12

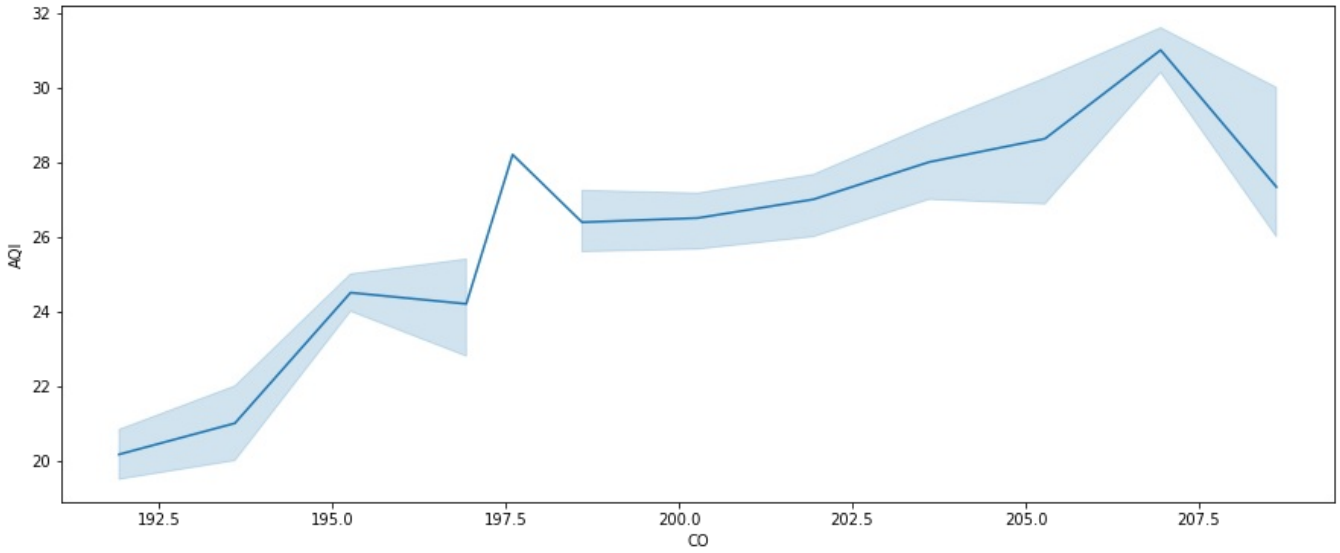

AQI and o3 Over Time
AQI
o3
Value
Datetime
10-22 00
10-22 12
10-23 00
10-23 12
10-24 00
10-24 12

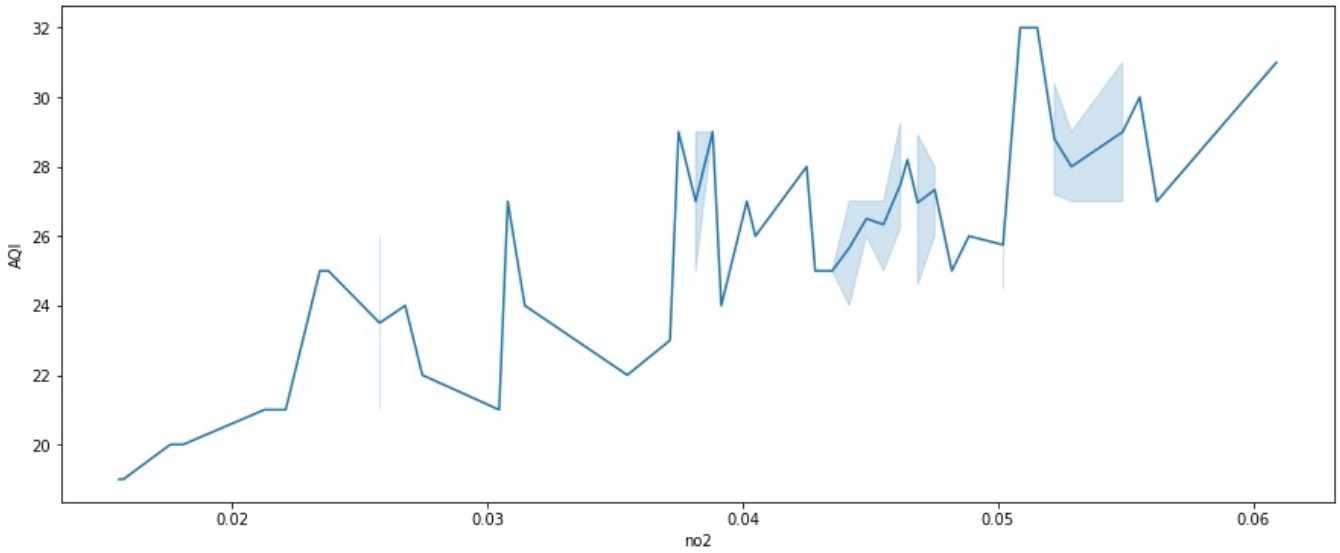


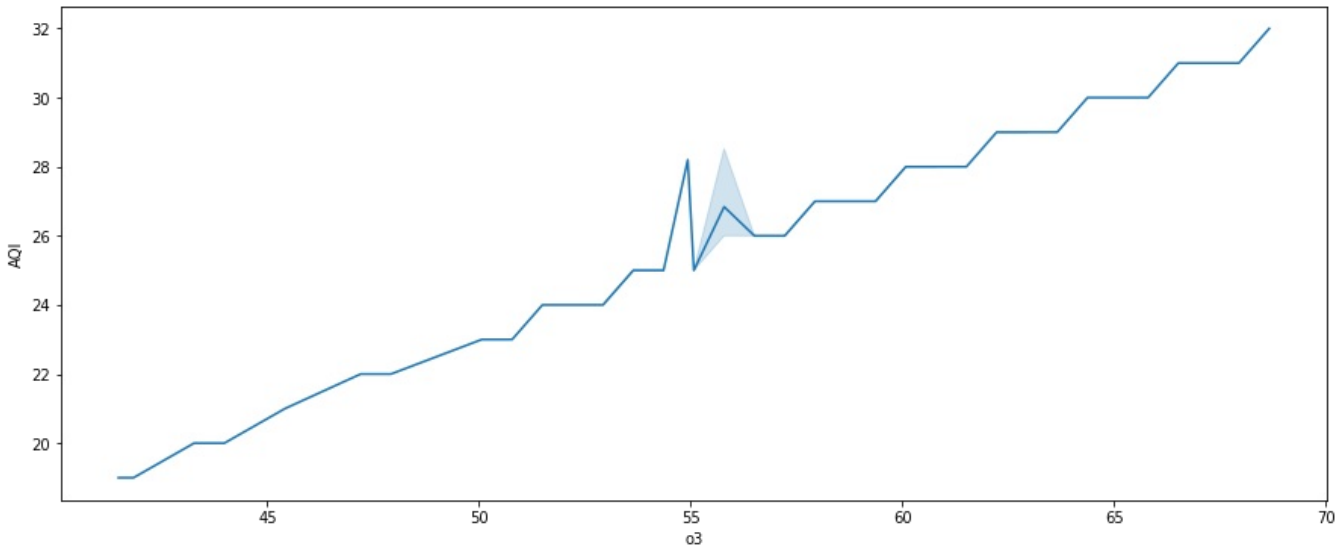


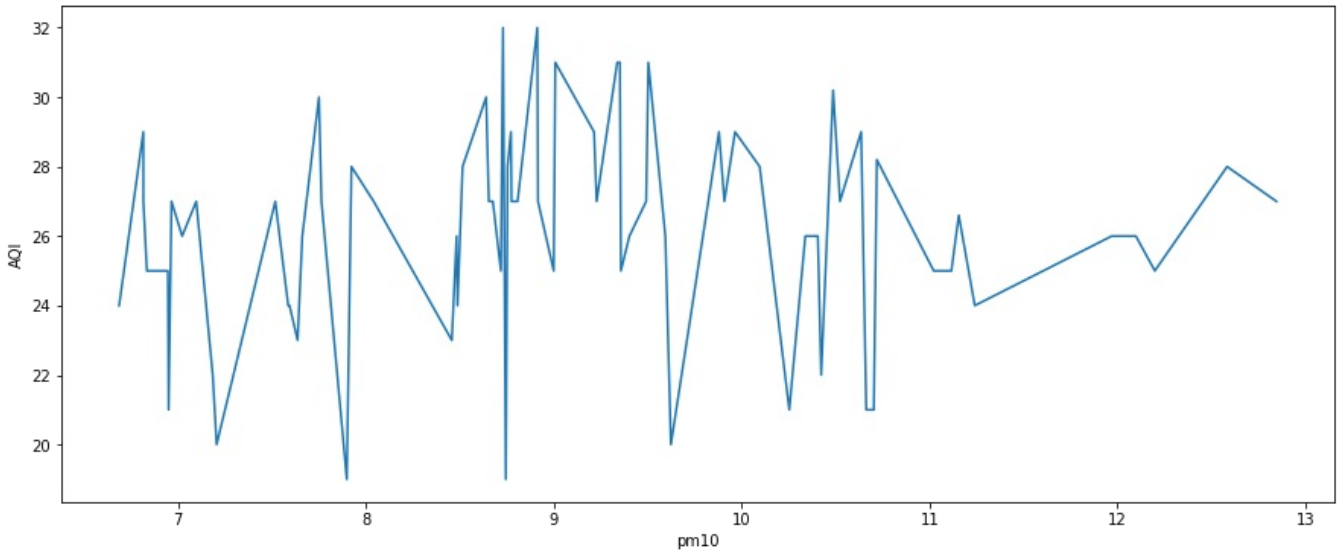


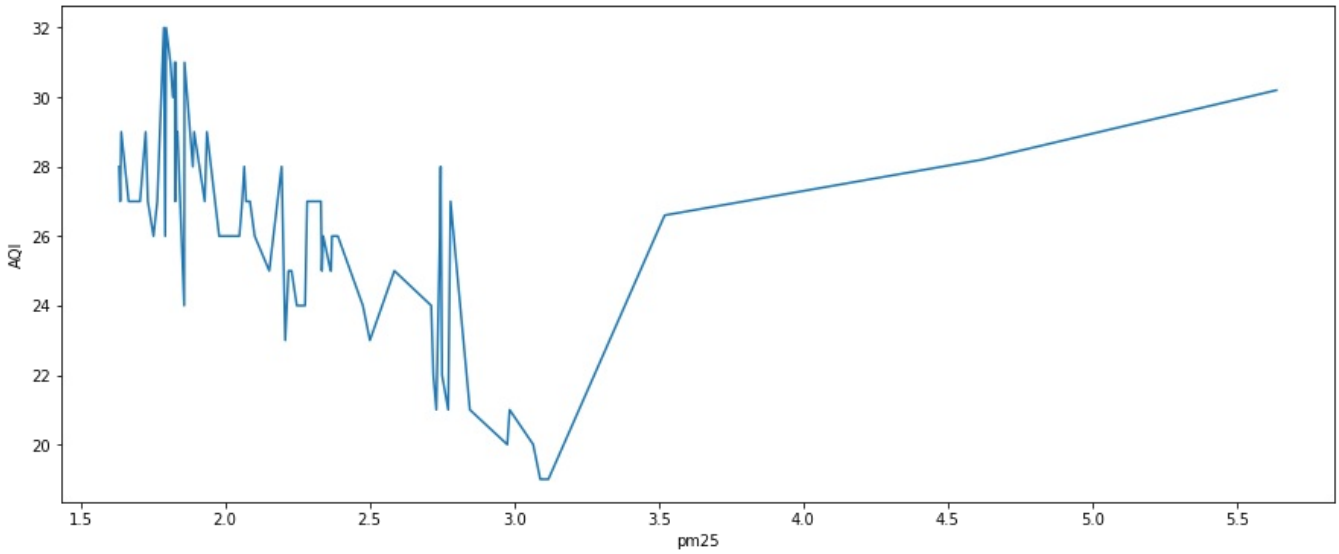
In [33]:
```
for i in numerical_columns:
    for j in numerical_columns:
        if i != j:
            plt.figure(figsize=(15,6))
            sns.lineplot(x = df[j], y = df[i], data = df, palette = 'hls')
            plt.show()
```

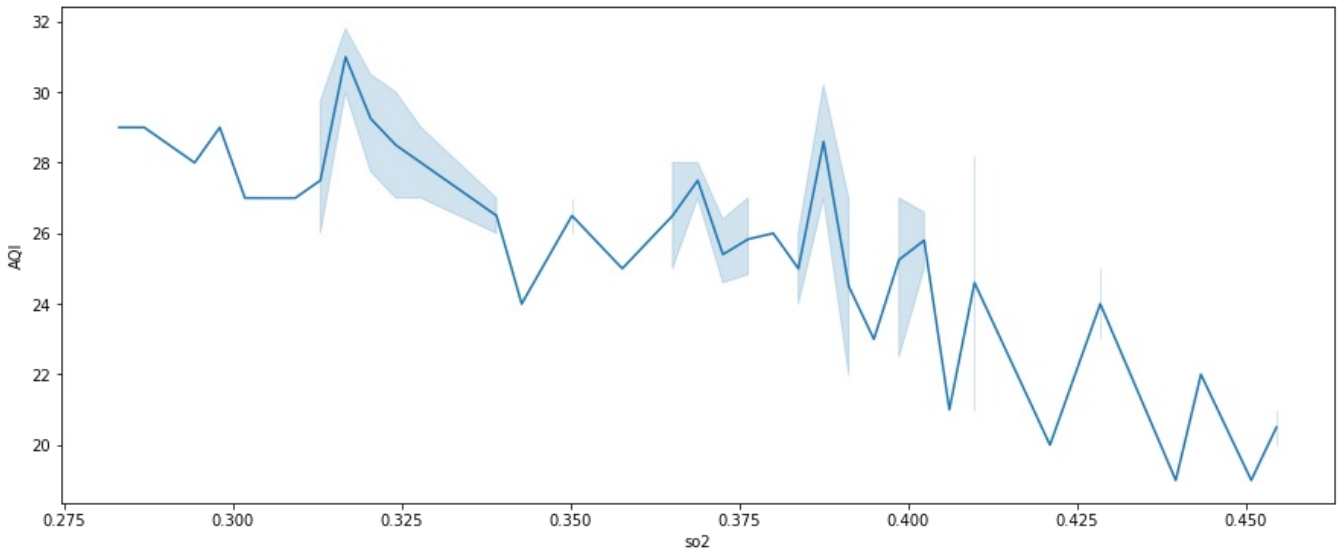

AQI
CO
32
30
28
26
24
22
20
192.5
195.0
197.5
200.0
202.5
205.0
207.5

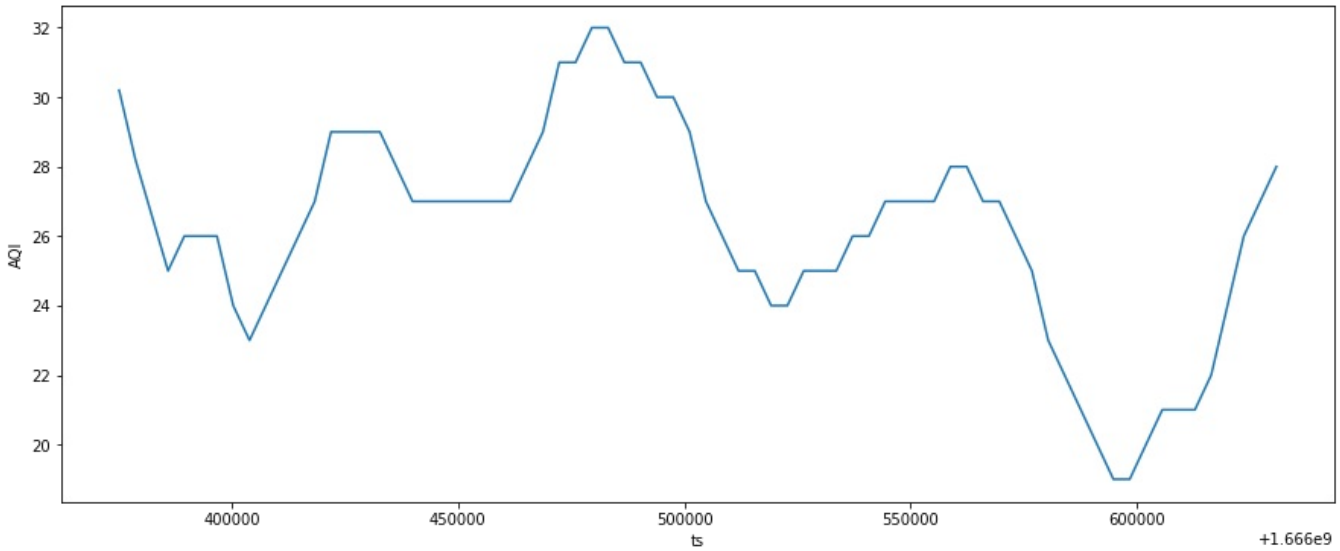

AQI
no2
32
30
28
26
24
22
20
0.02
0.03
0.04
0.05
0.06

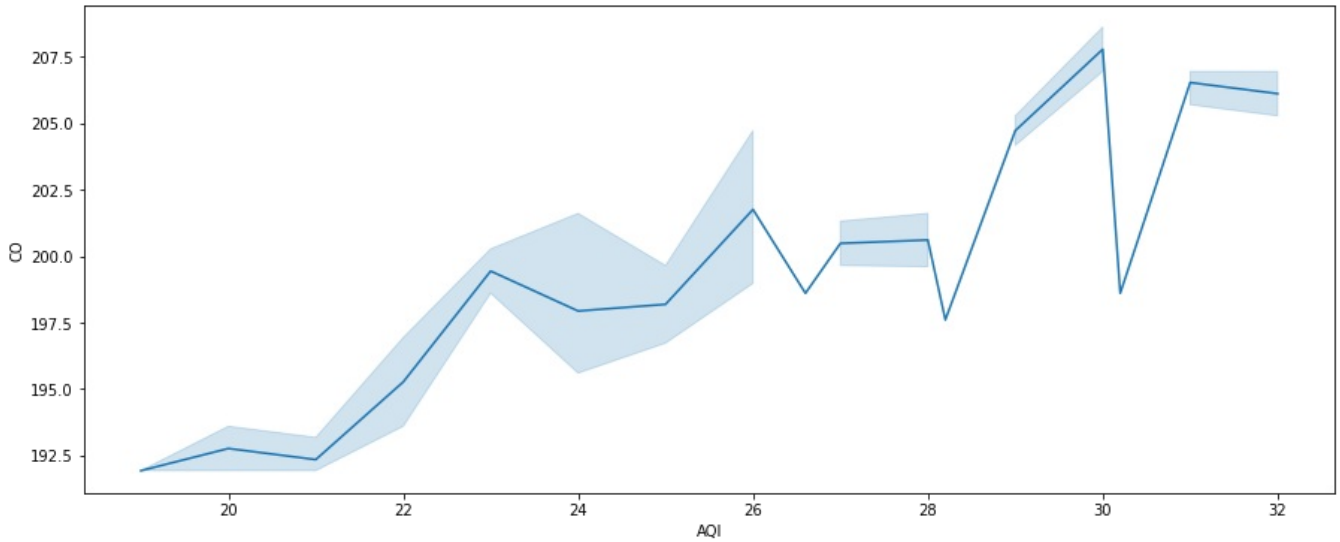

AQI
o3
32
30
28
26
24
22
20
45
50
55
60
65
70

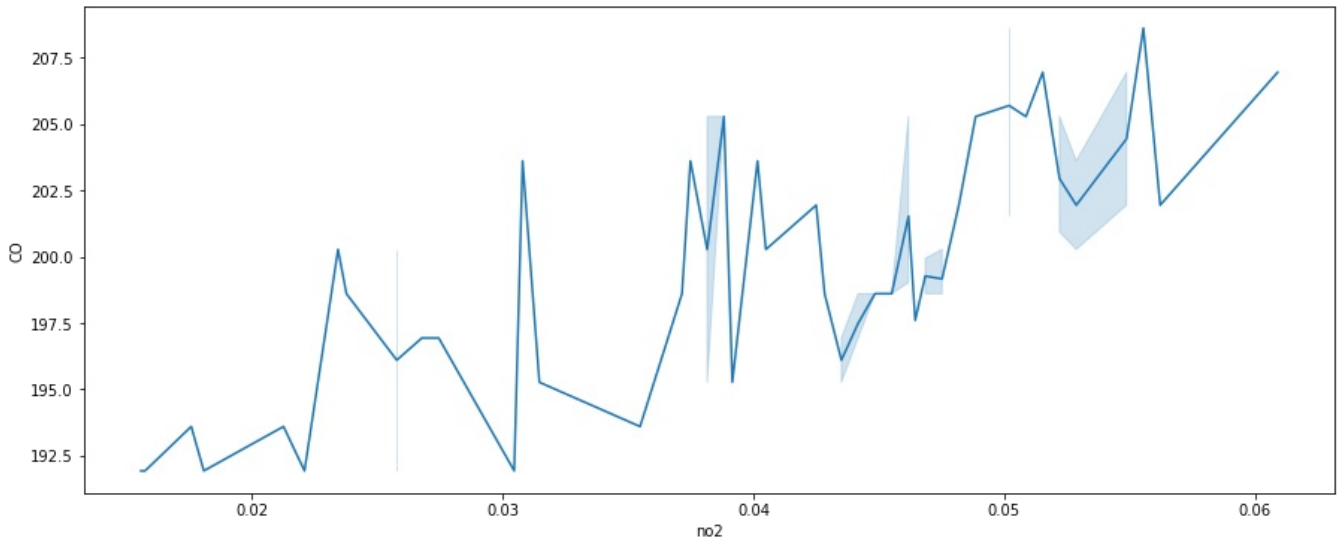

AQI
pm10

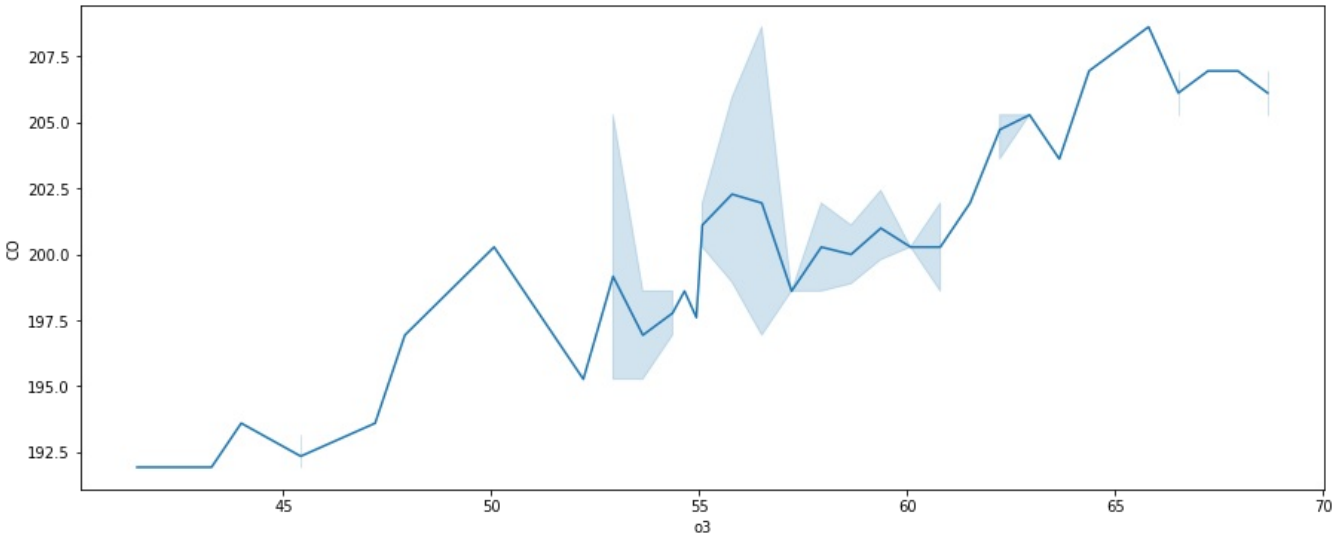

AQI
pm25

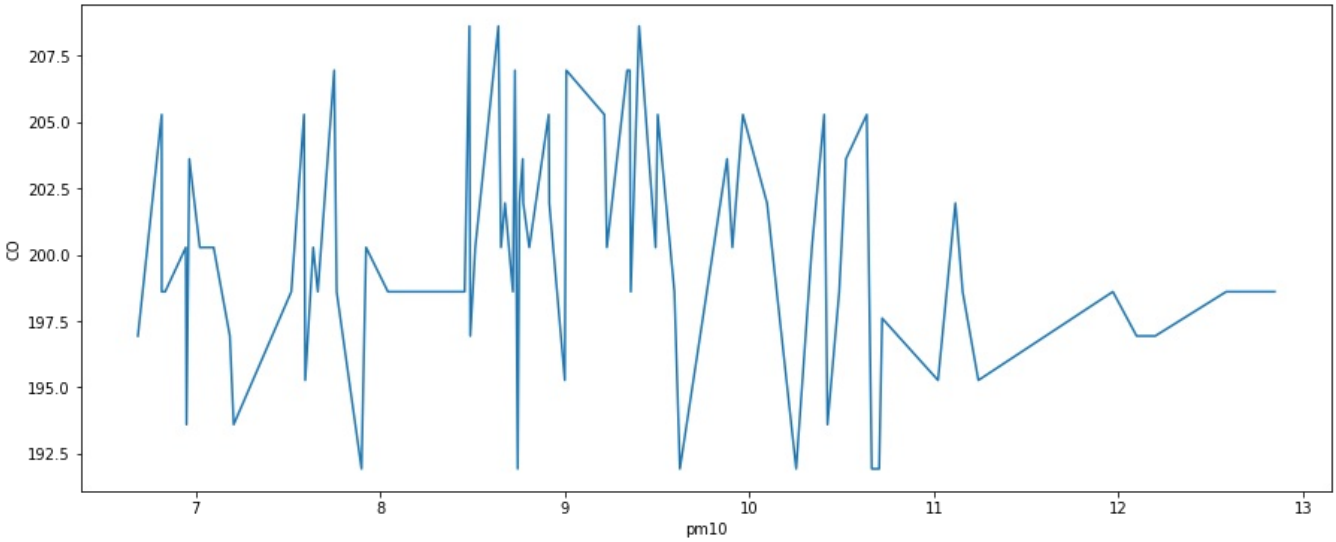

AQI
so2

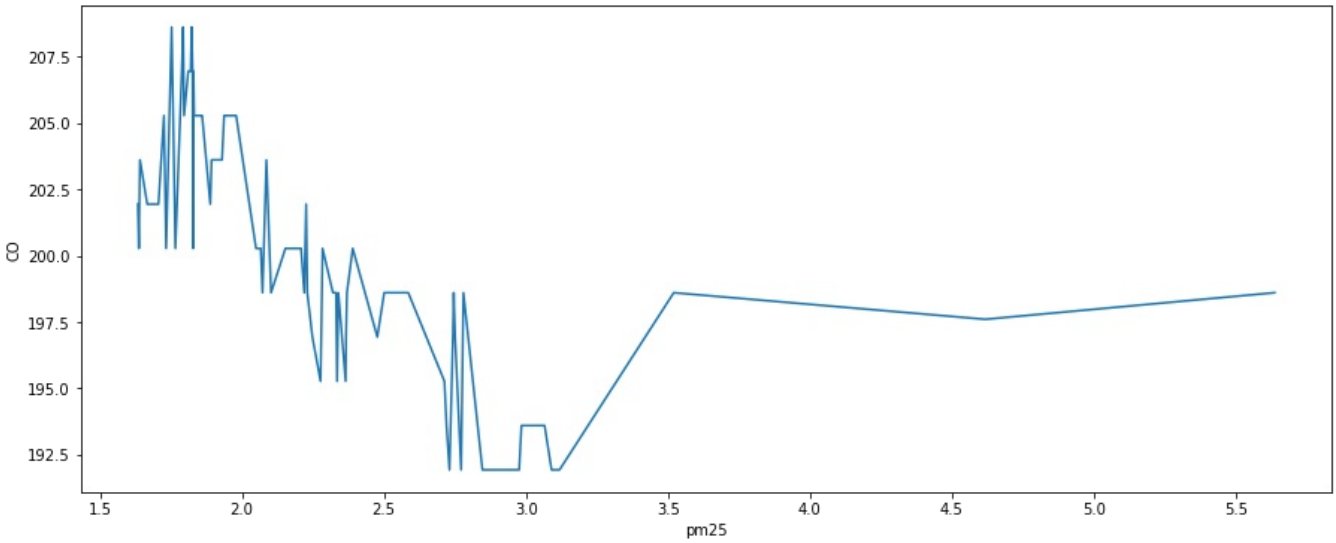

AQI
ts
+1.666e9

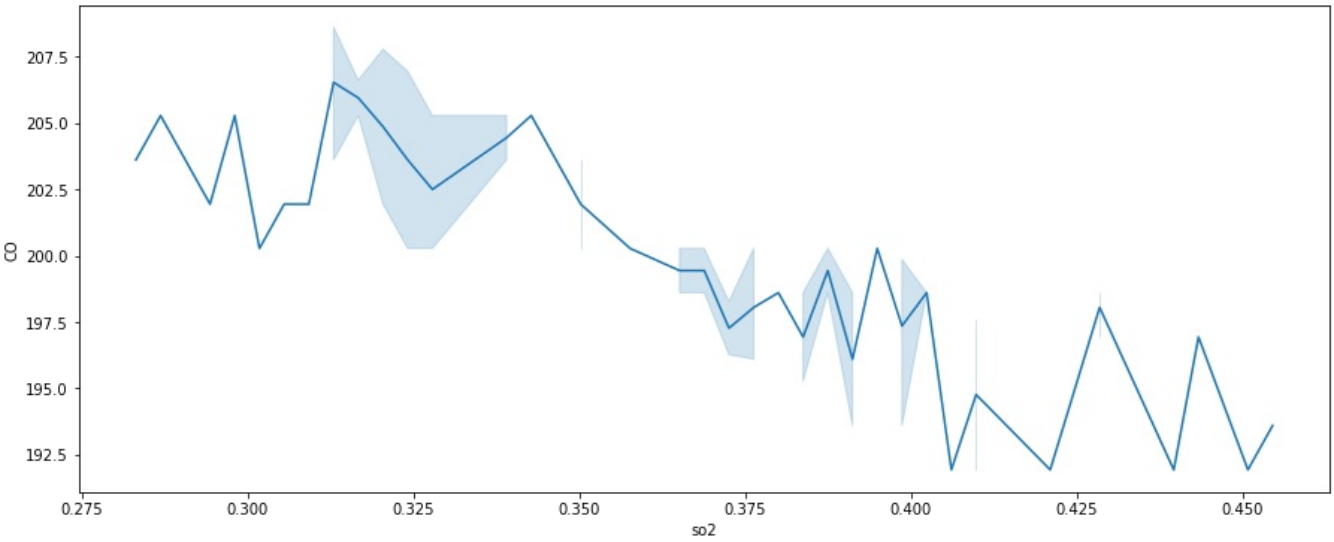

CO
AQI

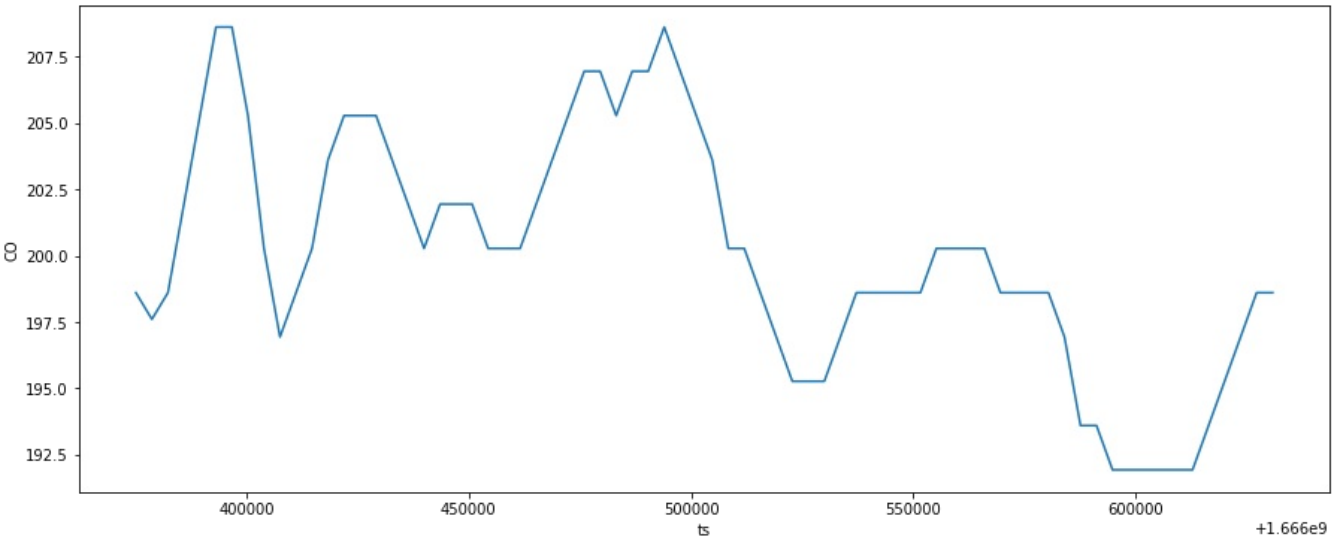

CO
no2

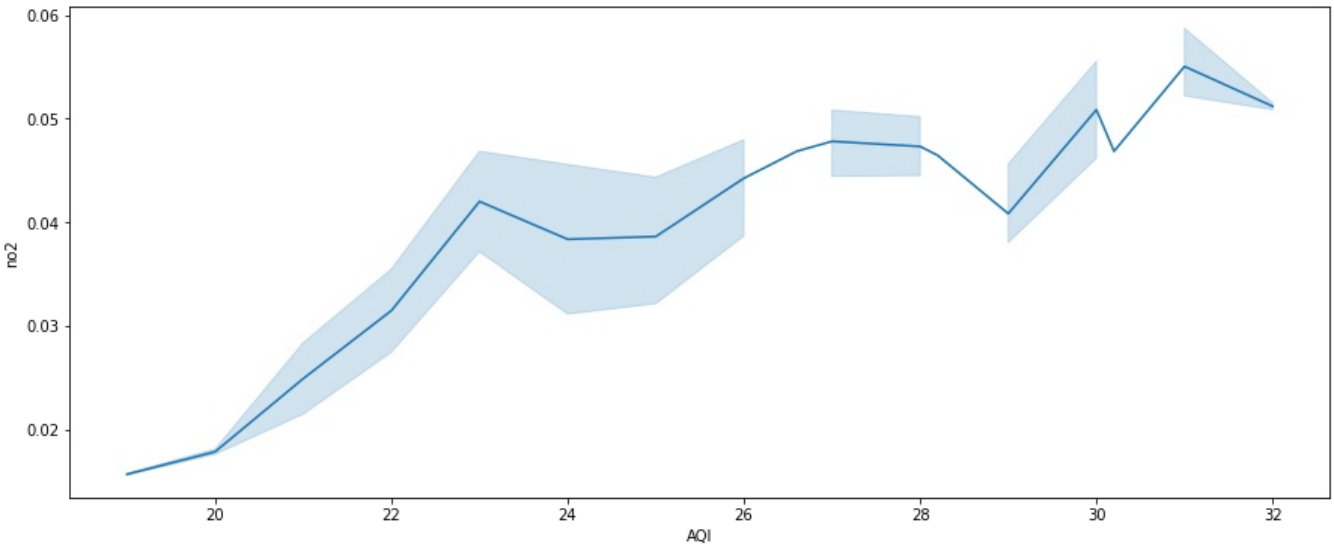

CO
207.5
205.0
202.5
200.0
197.5
195.0
192.5
45
50
55
60
65
70
o3

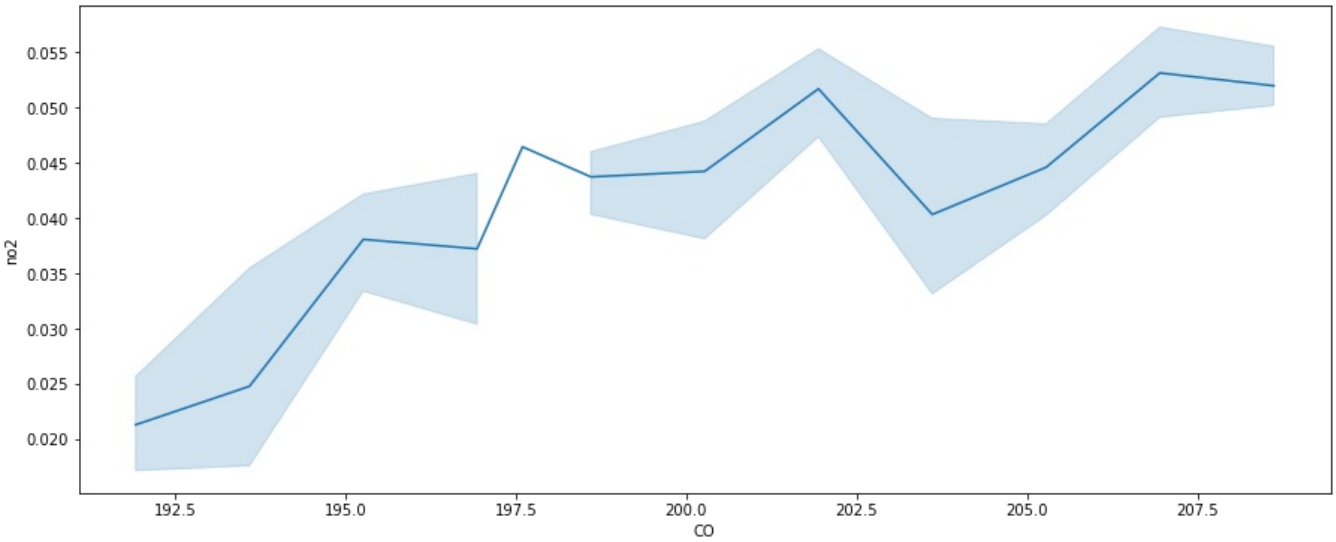

CO
207.5
205.0
202.5
200.0
197.5
195.0
192.5
7
8
9
10
11
12
13
pm10

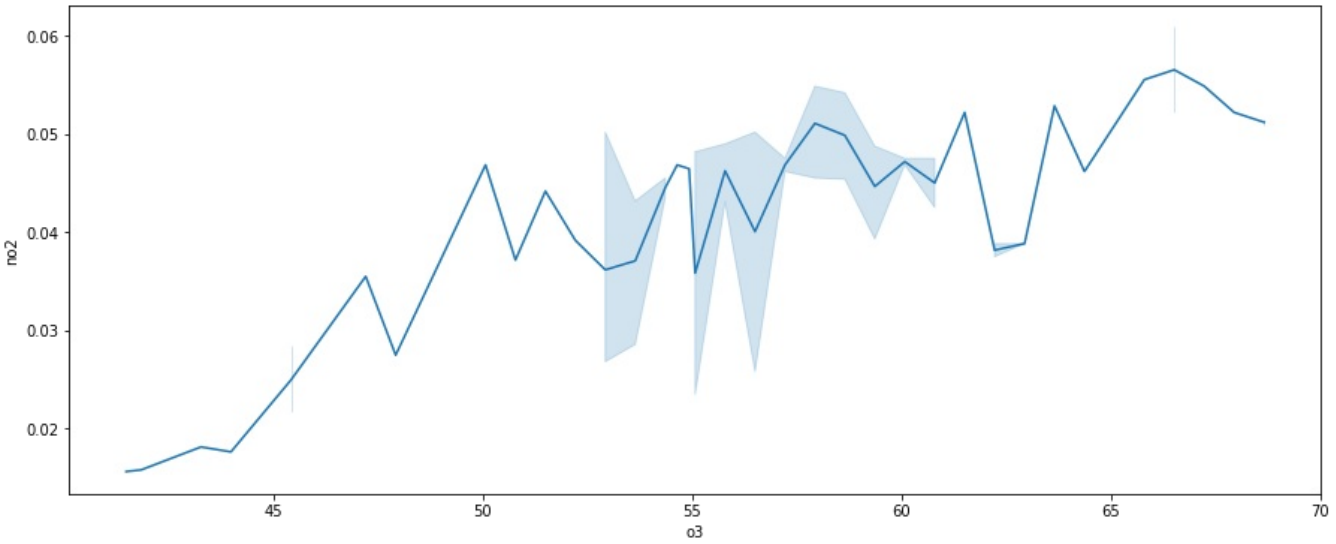

CO
207.5
205.0
202.5
200.0
197.5
195.0
192.5
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
pm25

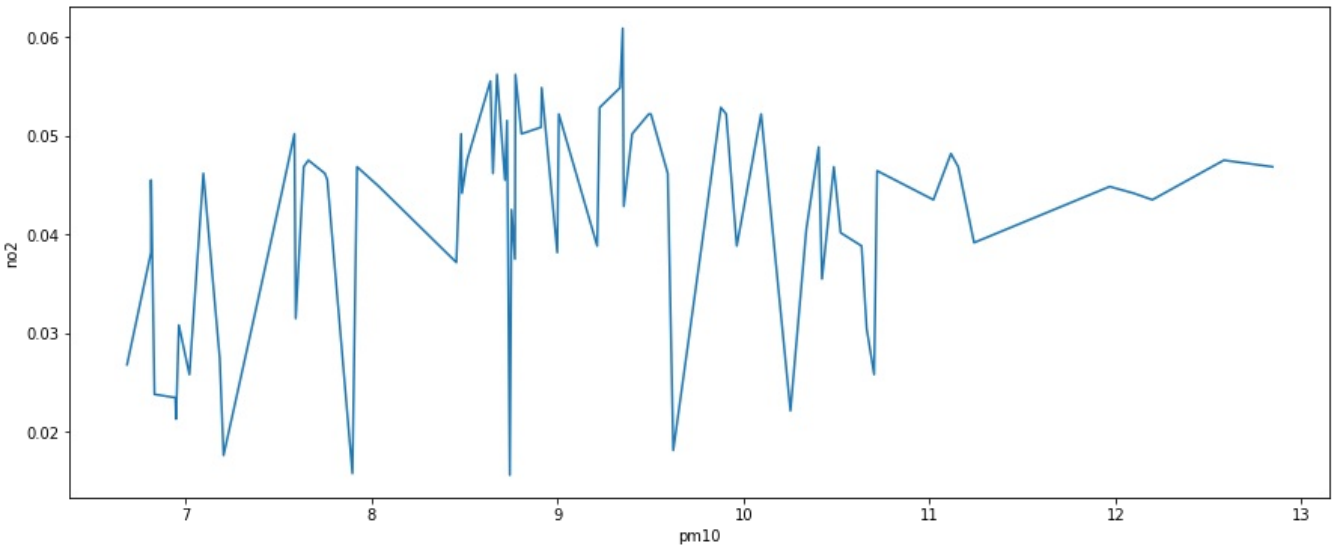

CO
so2
207.5
205.0
202.5
200.0
197.5
195.0
192.5
0.275
0.300
0.325
0.350
0.375
0.400
0.425
0.450

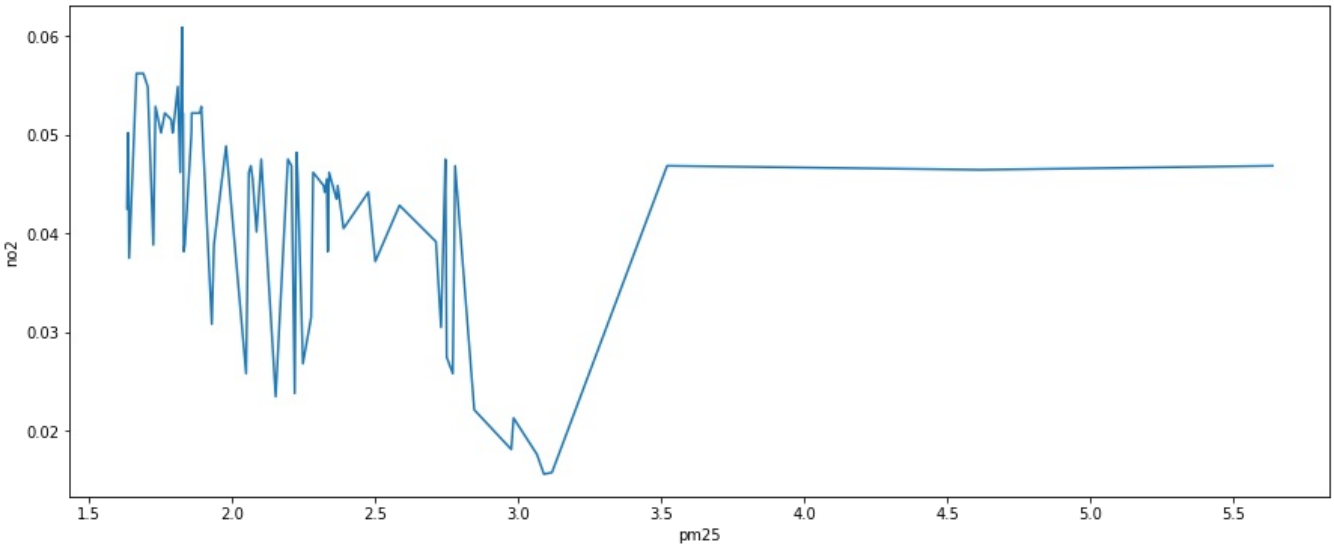

CO
ts
+1.666e9
207.5
205.0
202.5
200.0
197.5
195.0
192.5
400000
450000
500000
550000
600000

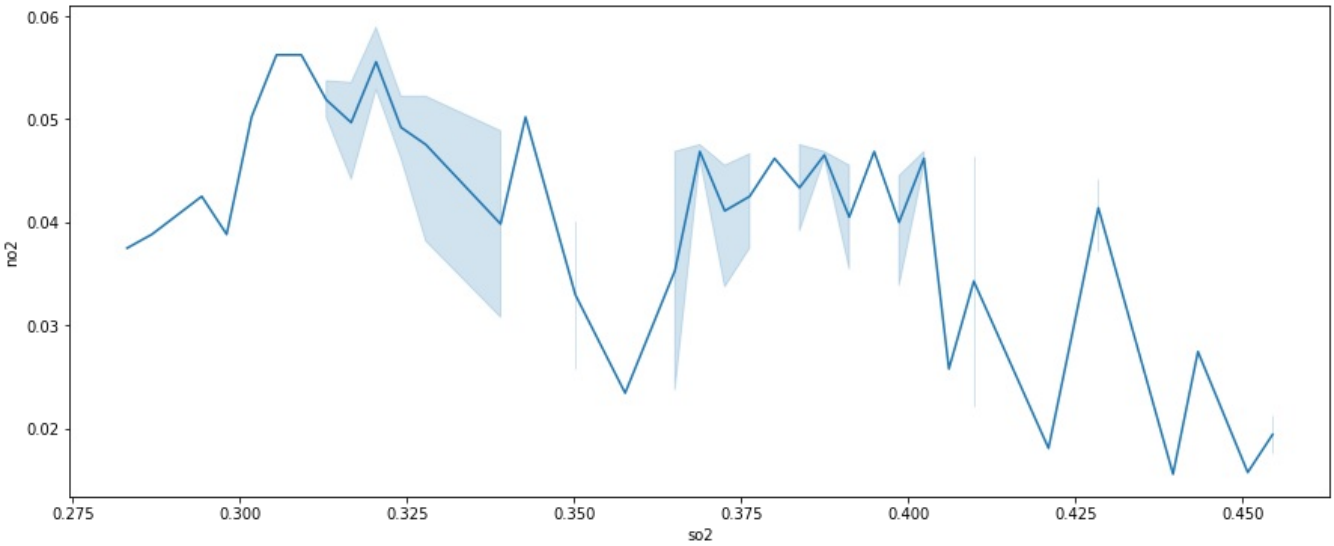

no2
AQI
0.06
0.05
0.04
0.03
0.02
20
22
24
26
28
30
32

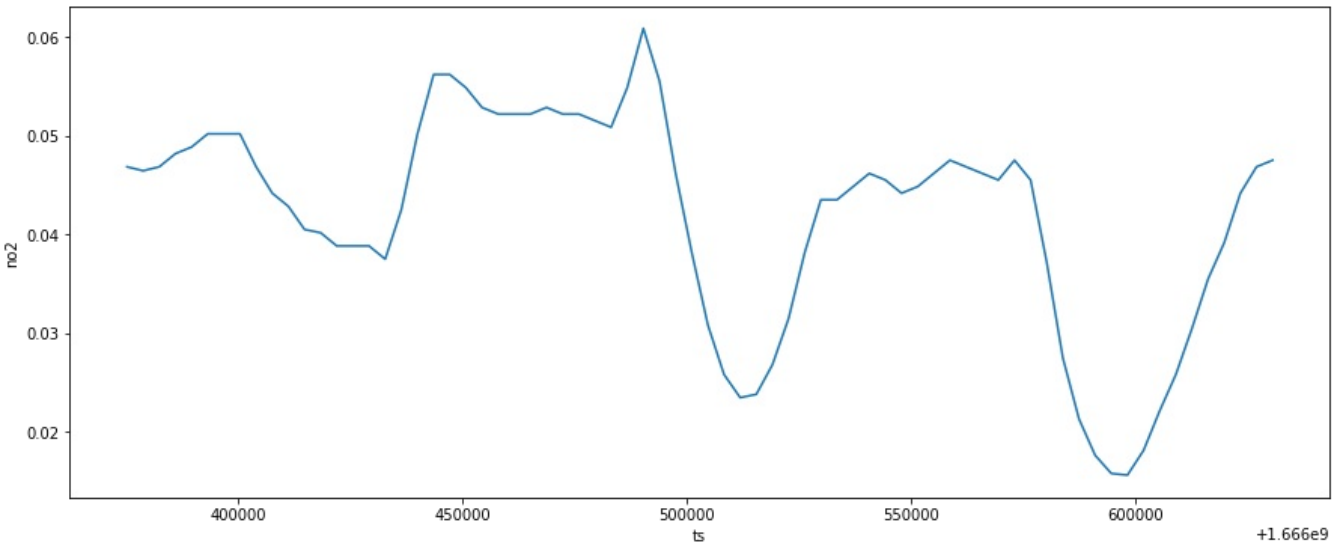

no2
0.055
0.050
0.045
0.040
0.035
0.030
0.025
0.020
192.5
195.0
197.5
200.0
202.5
205.0
207.5
CO

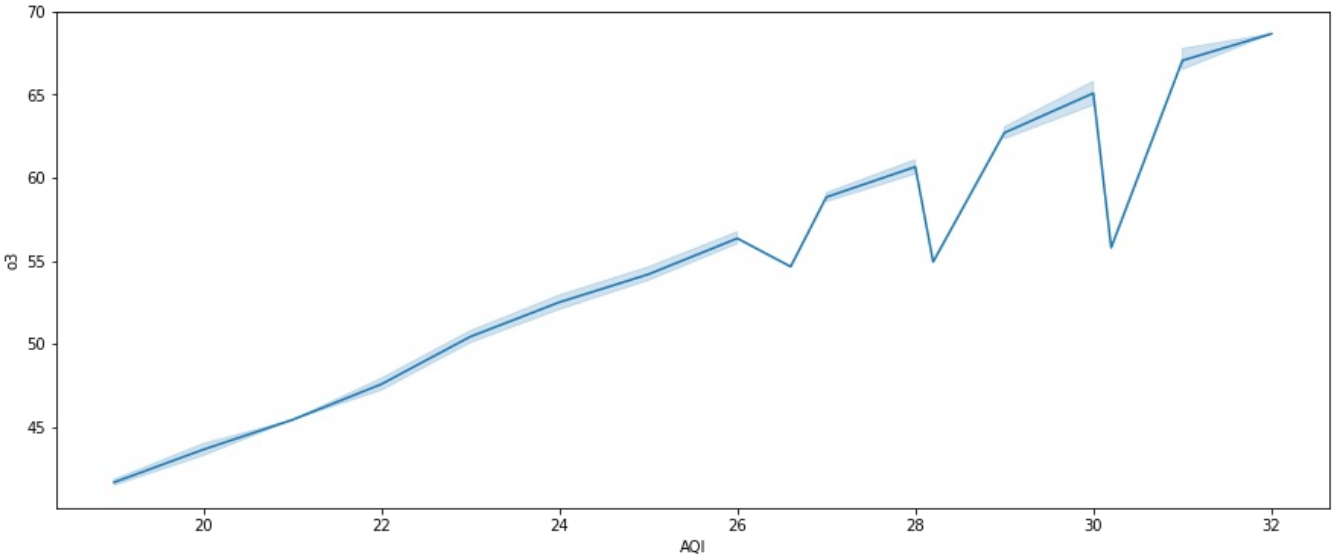

no2
0.06
0.05
0.04
0.03
0.02
45
50
55
60
65
70
o3

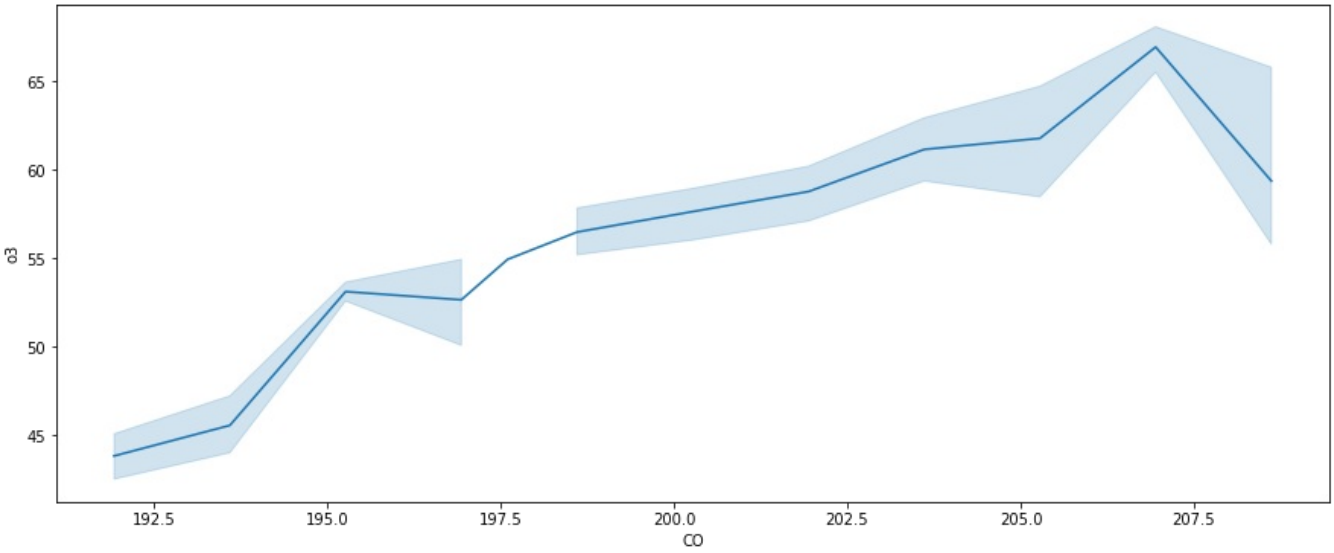

no2
0.06
0.05
0.04
0.03
0.02
7
8
9
10
11
12
13
pm10

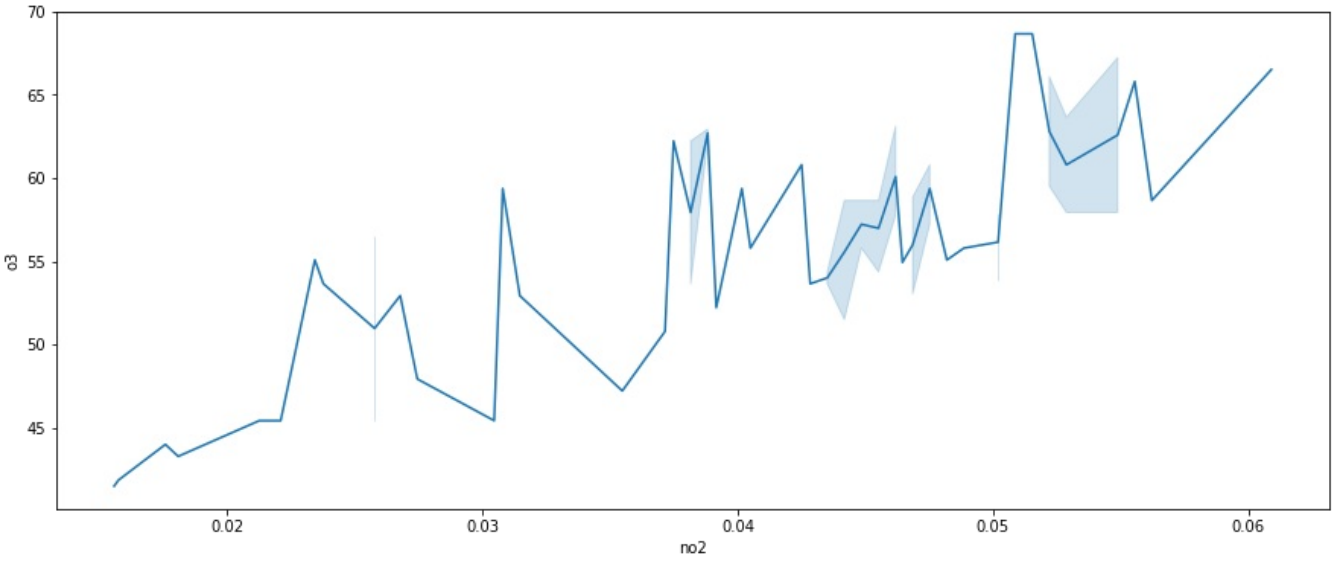

no2
0.06
0.05
0.04
0.03
0.02
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
pm25

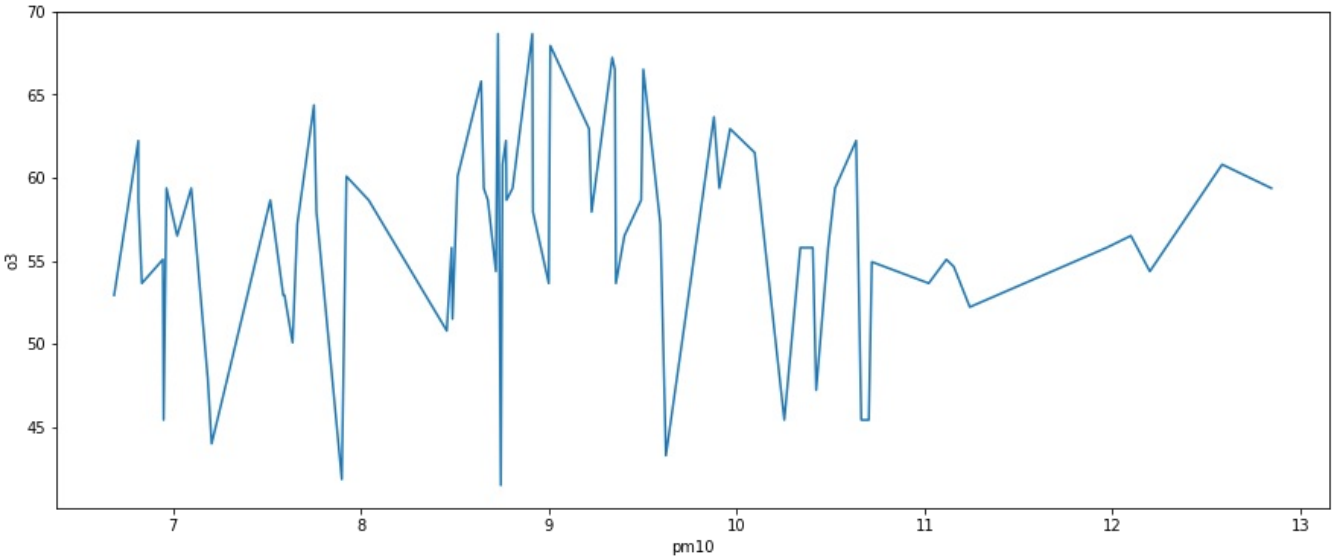

no2
0.06
0.05
0.04
0.03
0.02
0.275
0.300
0.325
0.350
0.375
0.400
0.425
0.450
so2

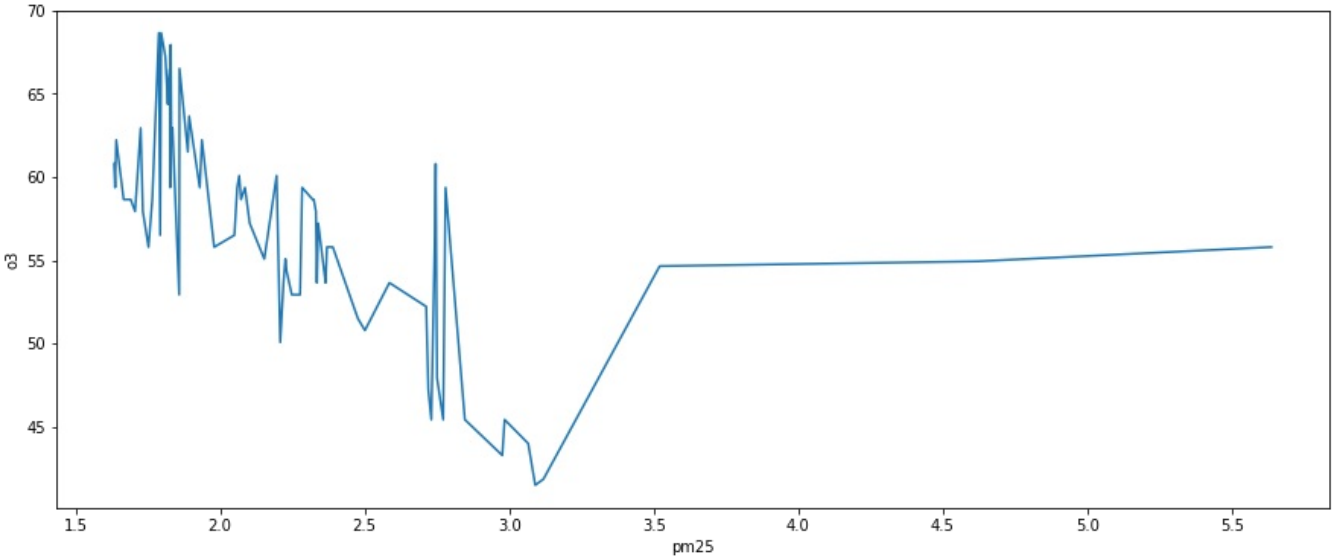

no2
0.06
0.05
0.04
0.03
0.02
400000
450000
500000
550000
600000
ts
+1.666e9

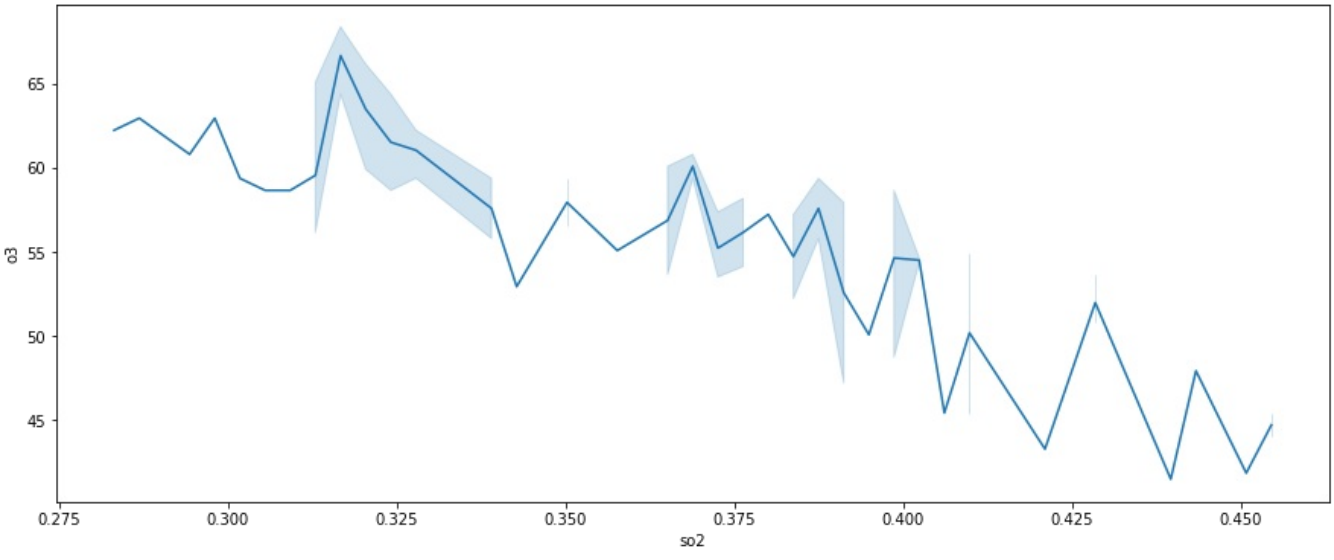

70
65
60
55
50
45
o3
20
22
24
26
28
30
32
AQI

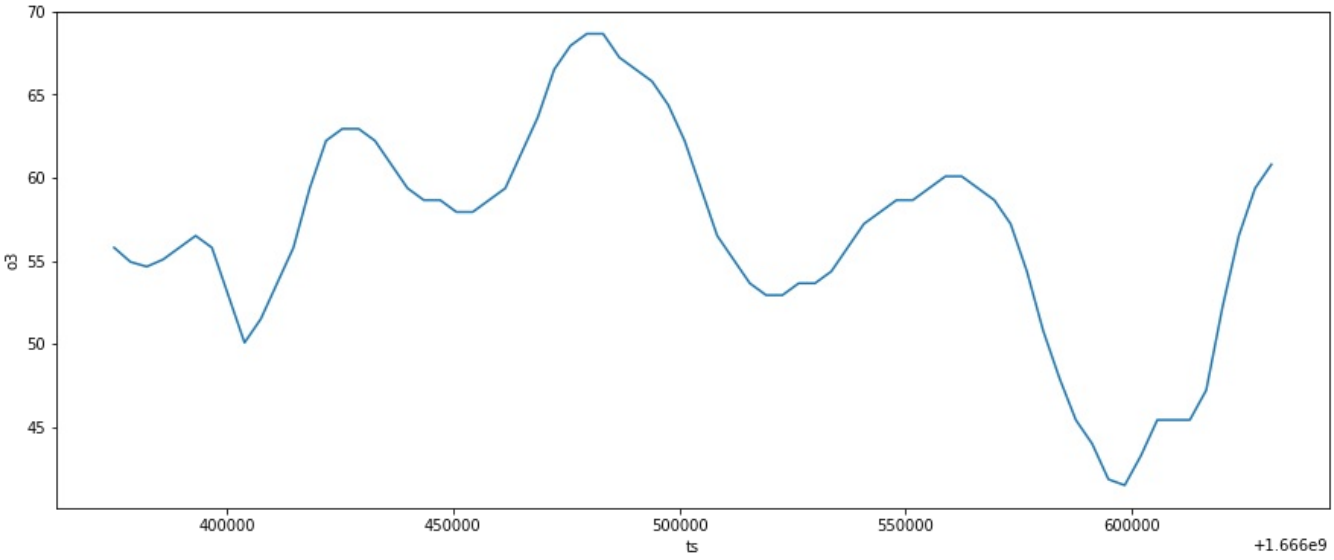

65
60
55
50
45
o3
192.5
195.0
197.5
200.0
202.5
205.0
207.5
CO

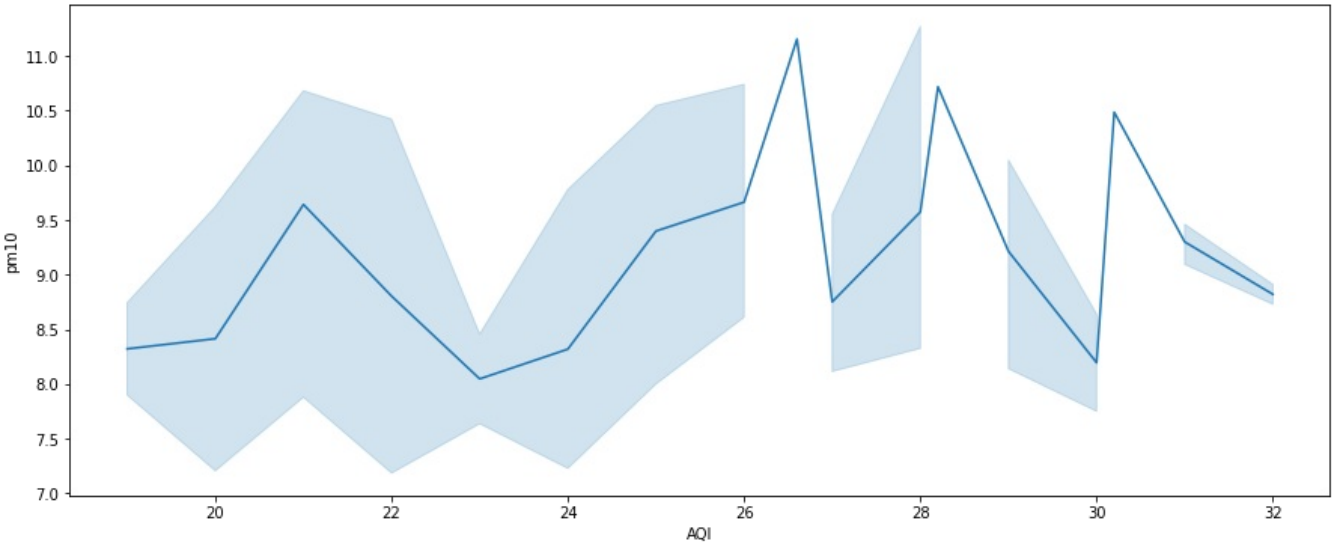

70
65
60
55
50
45
o3
0.02
0.03
0.04
0.05
0.06
no2

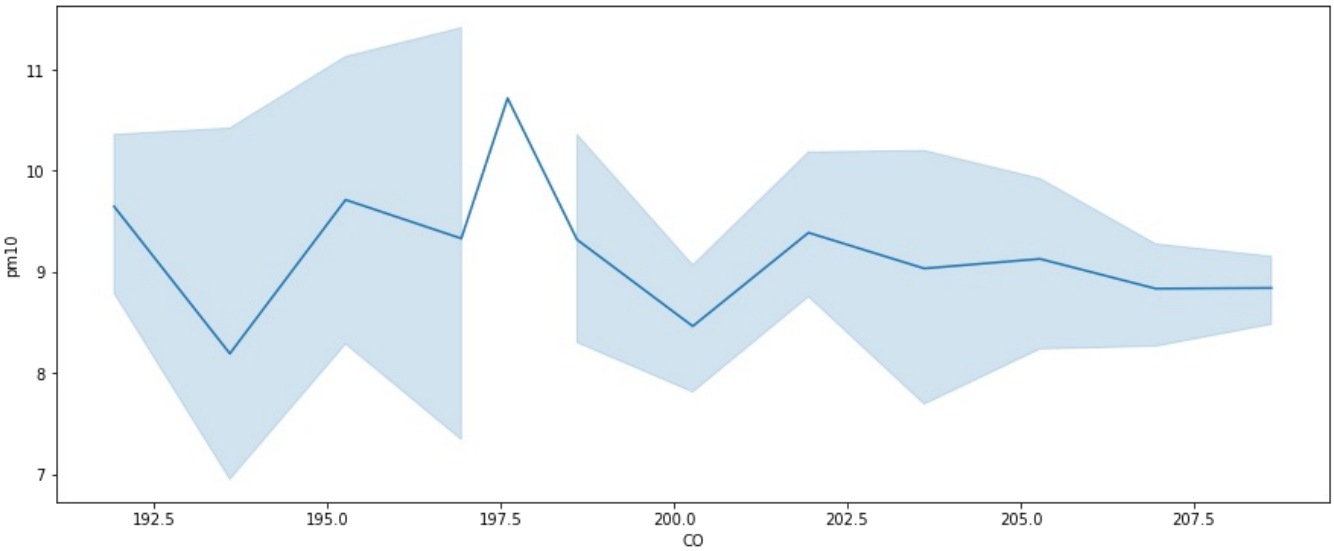

o3
pm10

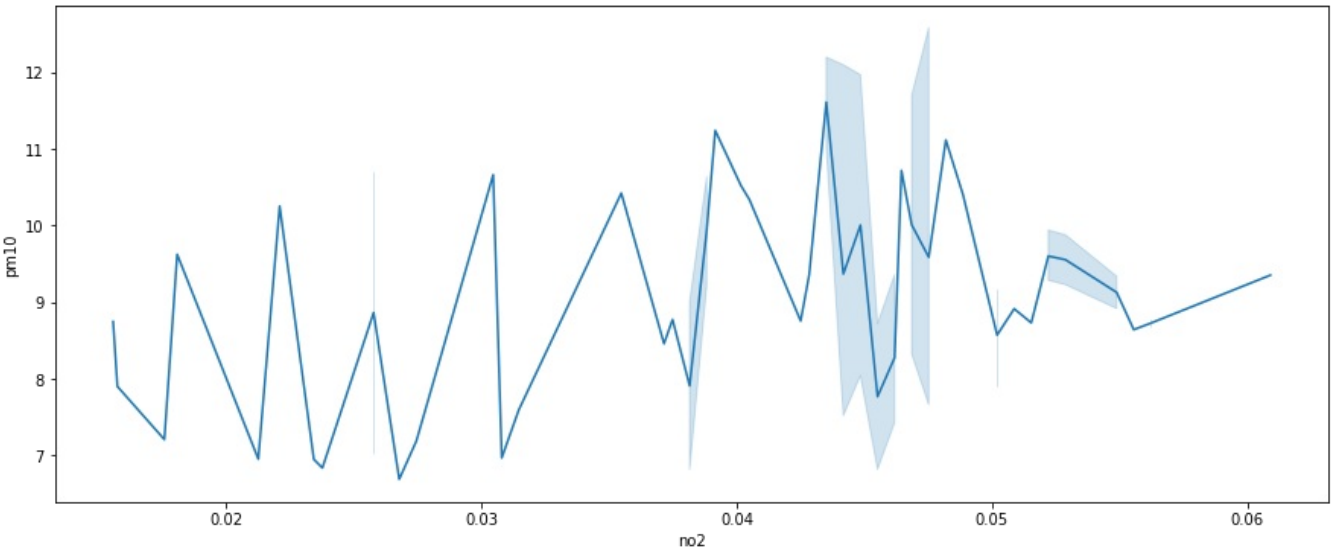

o3
pm25

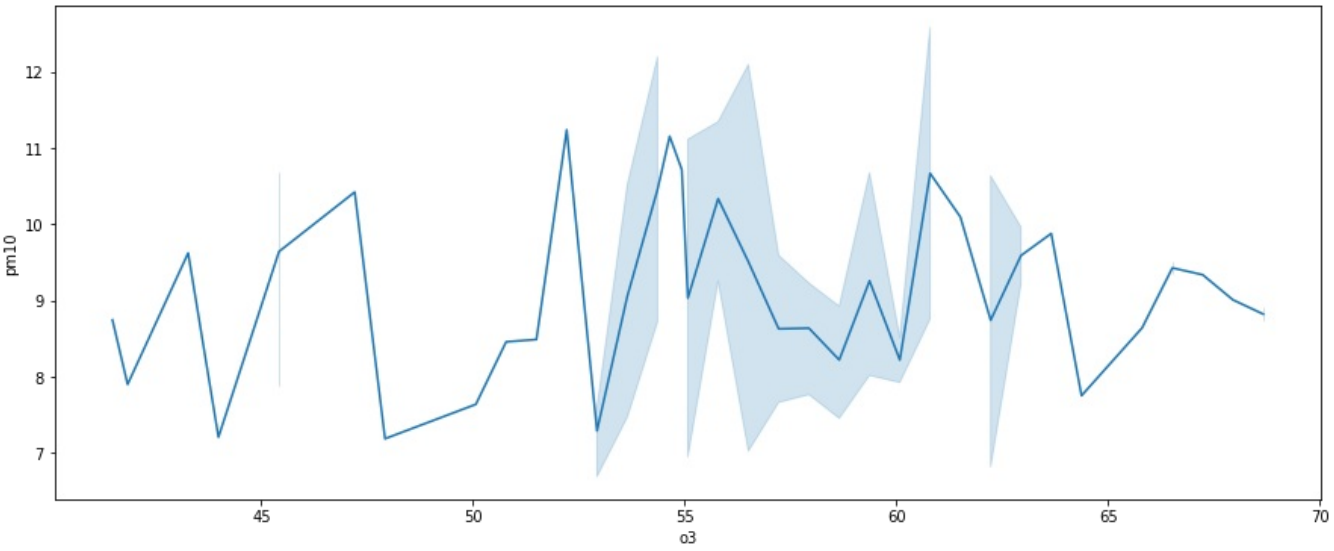

o3
so2

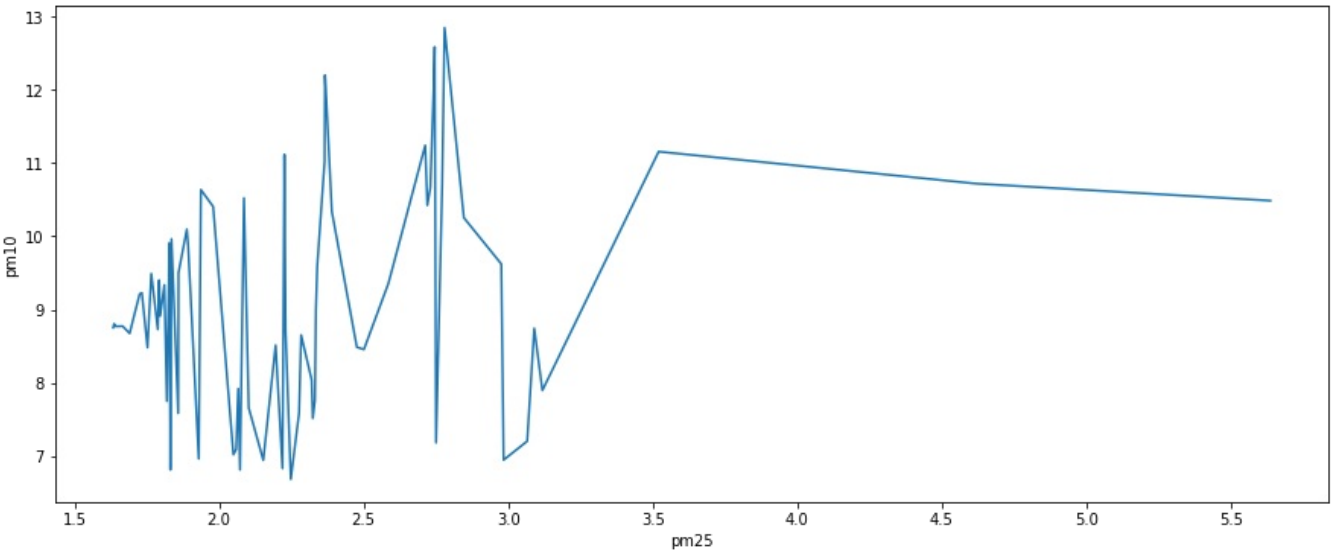

70
65
60
55
50
45
o3
400000
450000
500000
550000
600000
ts
+1.666e9

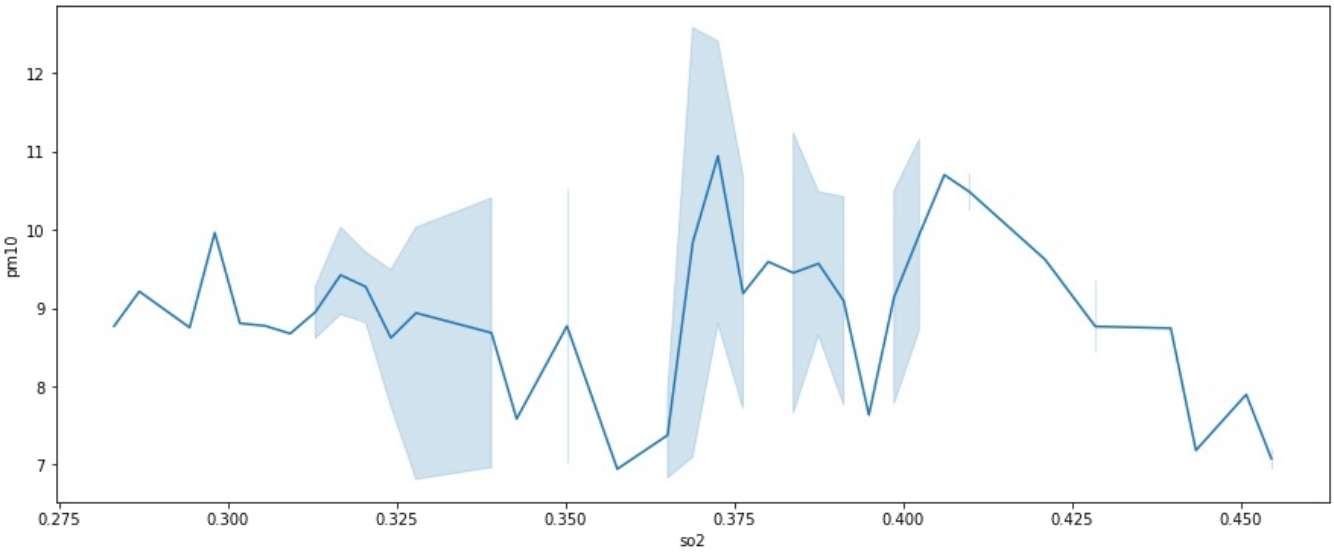

11.0
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
pm10
20
22
24
26
28
30
32
AQI

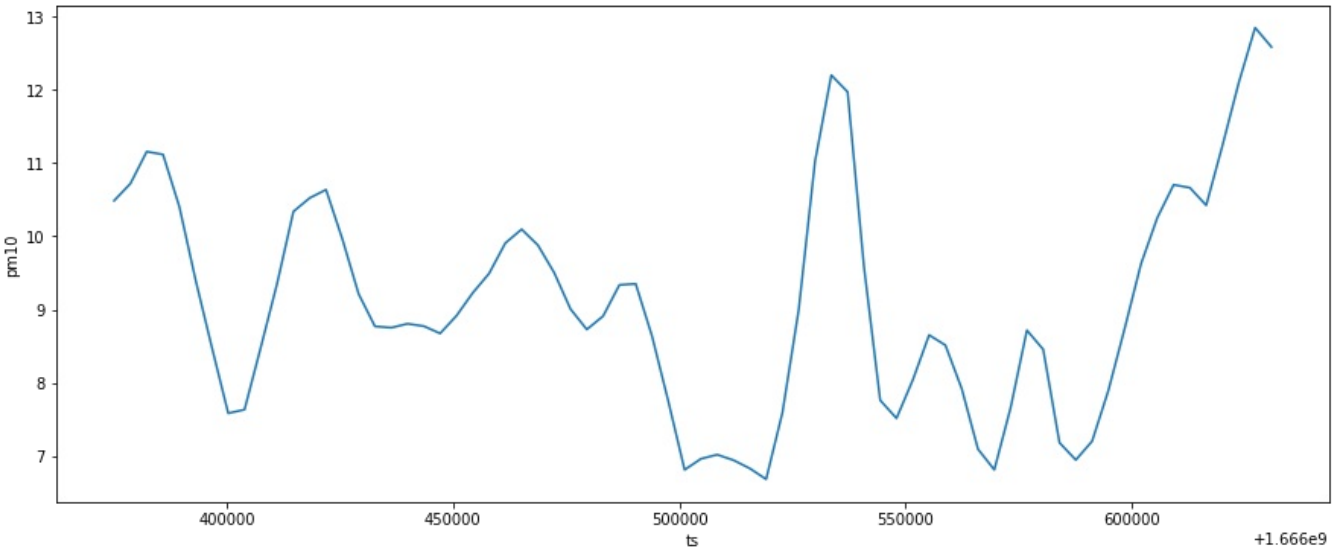

11
10
9
8
7
pm10
192.5
195.0
197.5
200.0
202.5
205.0
207.5
CO

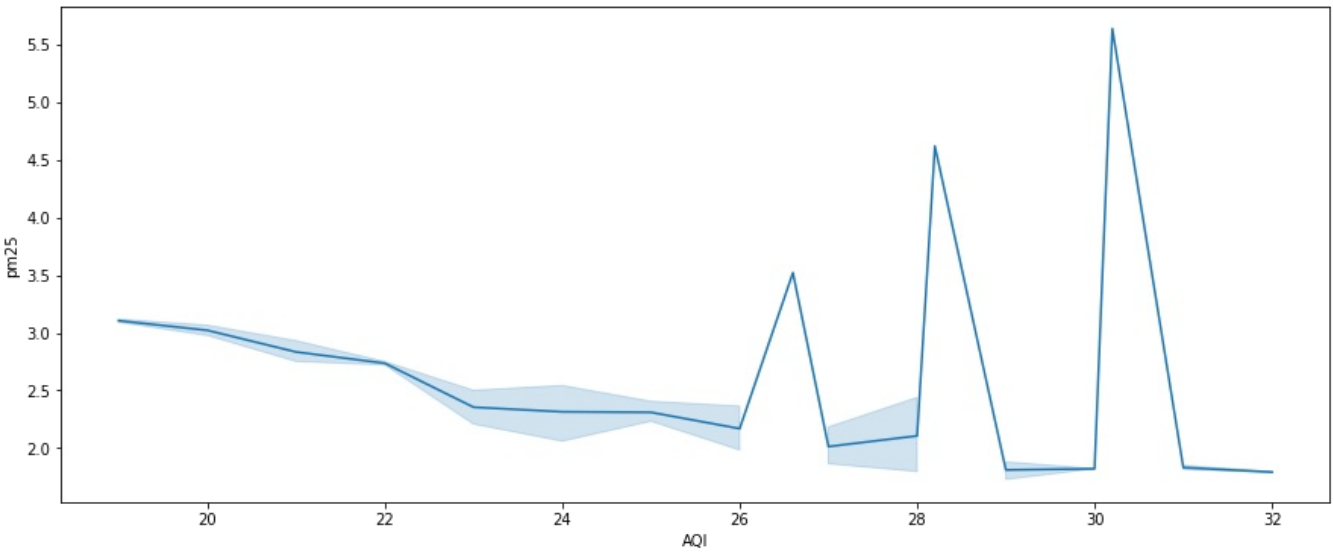

pm10
12
11
10
9
8
7
0.02
0.03
0.04
0.05
0.06
no2

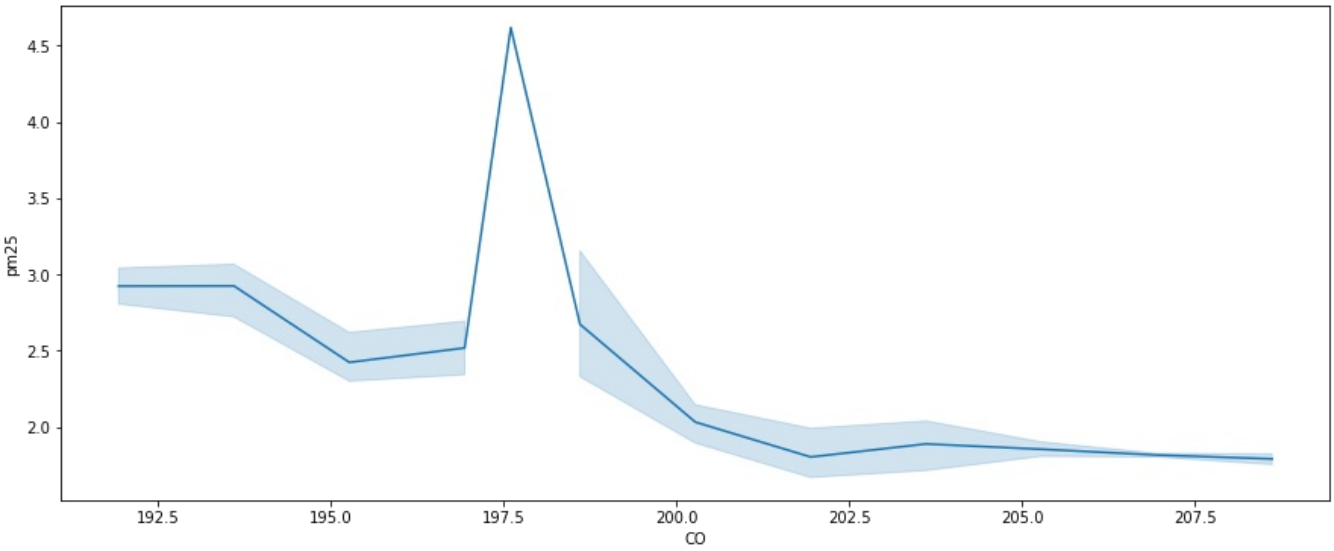

pm10
12
11
10
9
8
7
45
50
55
60
65
70
o3

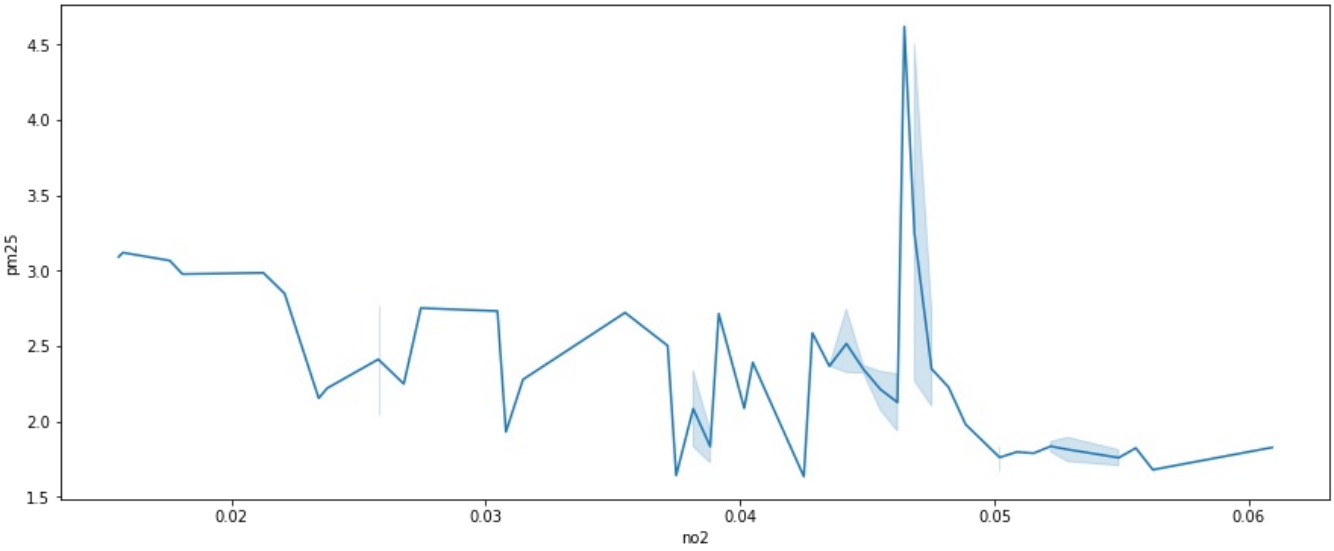

pm10
13
12
11
10
9
8
7
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
pm25

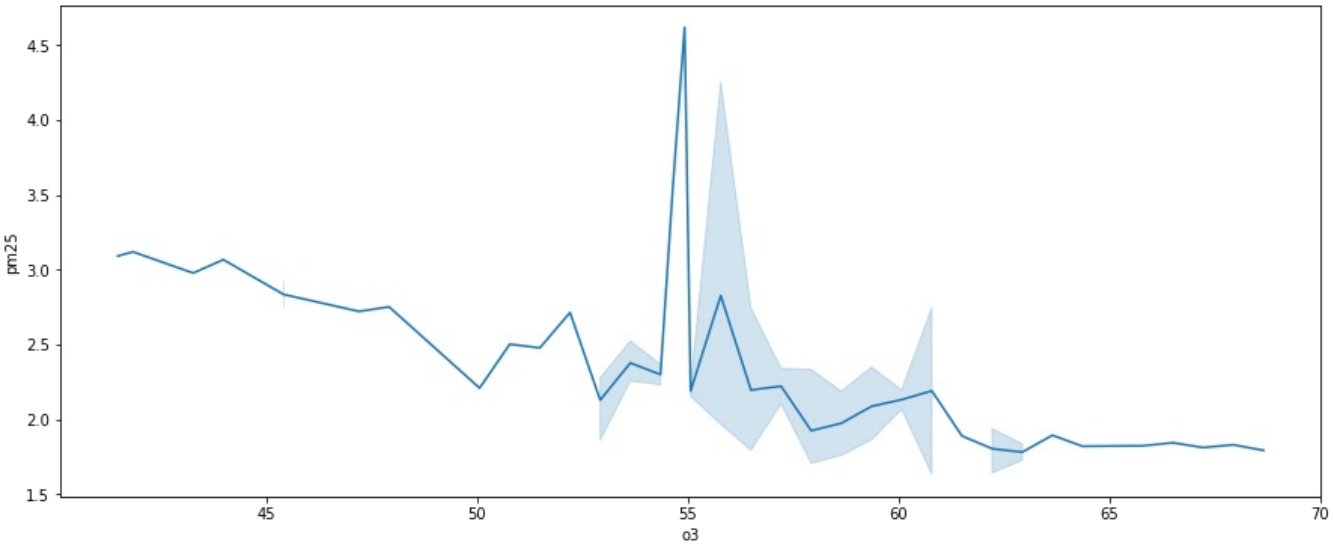

pm10
12
11
10
9
8
7
0.275
0.300
0.325
0.350
0.375
0.400
0.425
0.450
so2

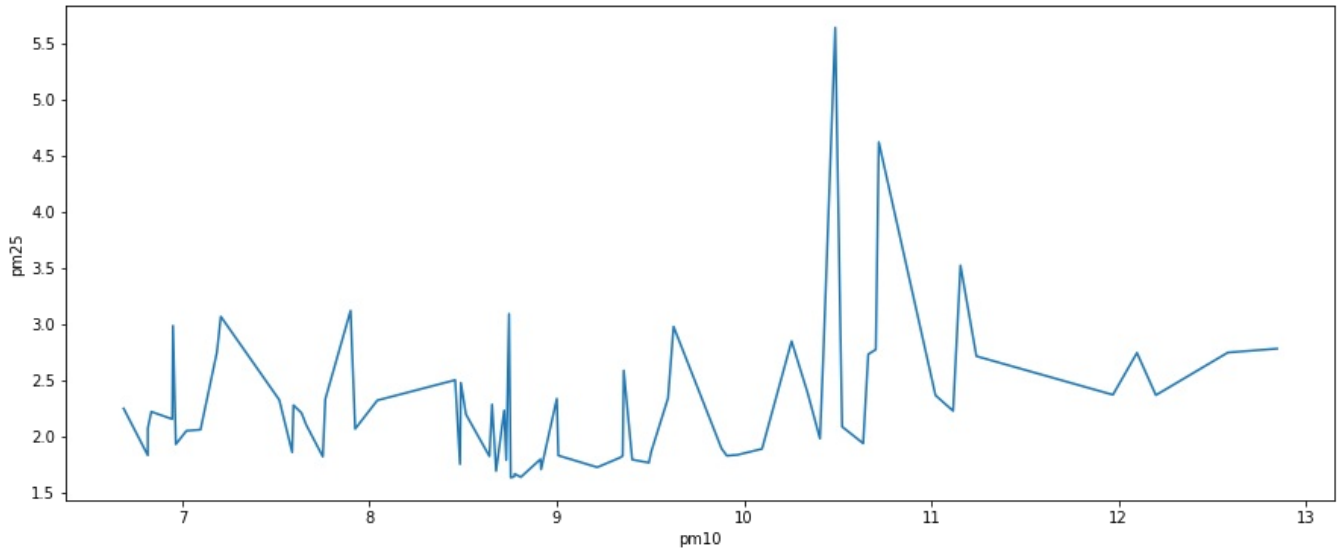

pm10
13
12
11
10
9
8
7
400000
450000
500000
550000
600000
ts
+1.666e9

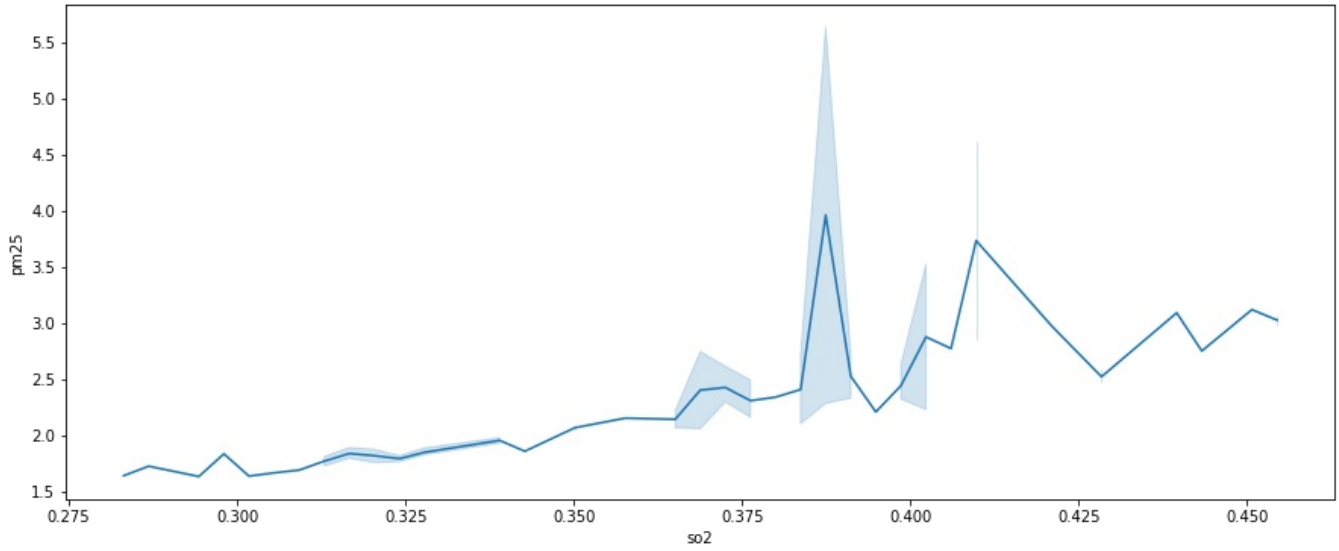

pm25
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
20
22
24
26
28
30
32
AQI

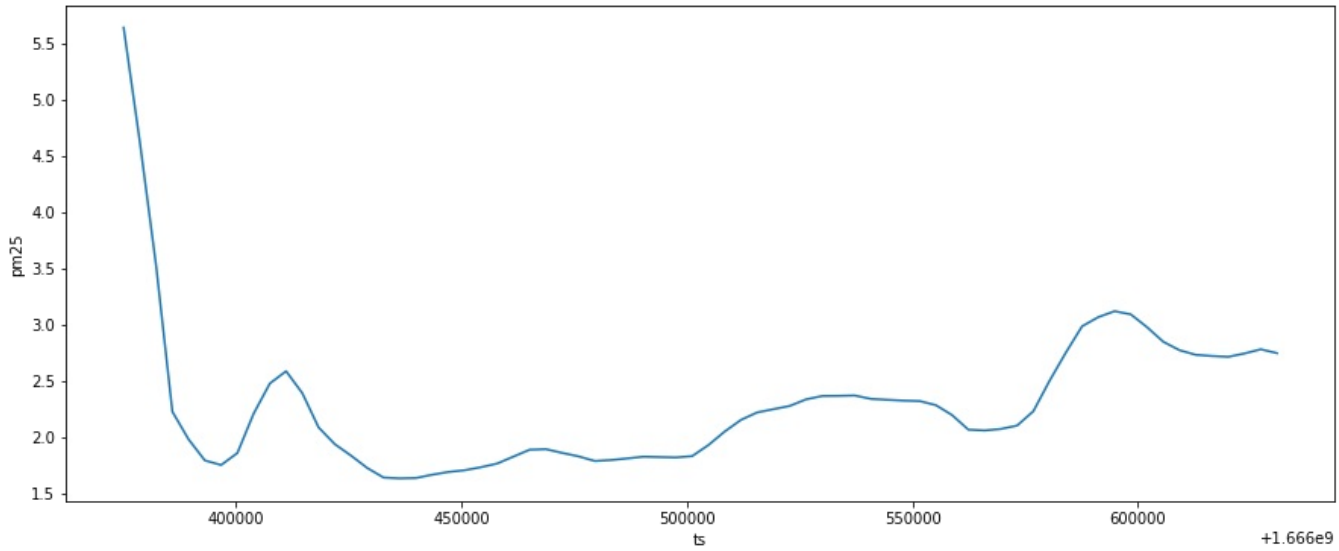

pm25
4.5
4.0
3.5
3.0
2.5
2.0
192.5
195.0
197.5
200.0
202.5
205.0
207.5
CO

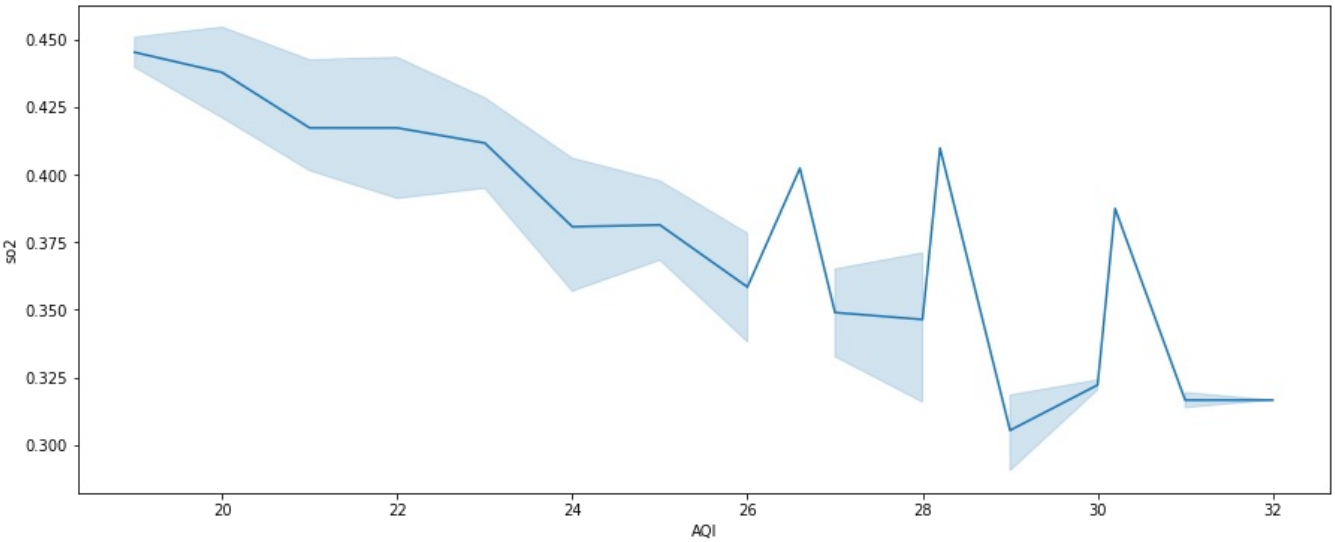

pm25
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0.02
0.03
0.04
0.05
0.06
no2

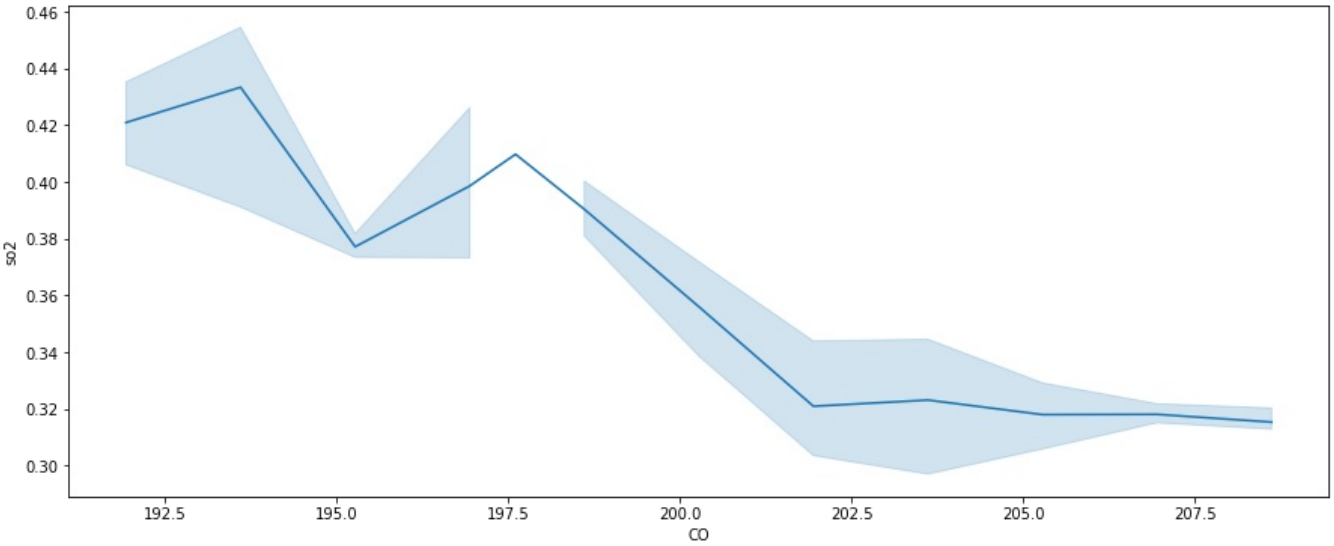

pm25
4.5
4.0
3.5
3.0
2.5
2.0
1.5
45
50
55
60
65
70
o3

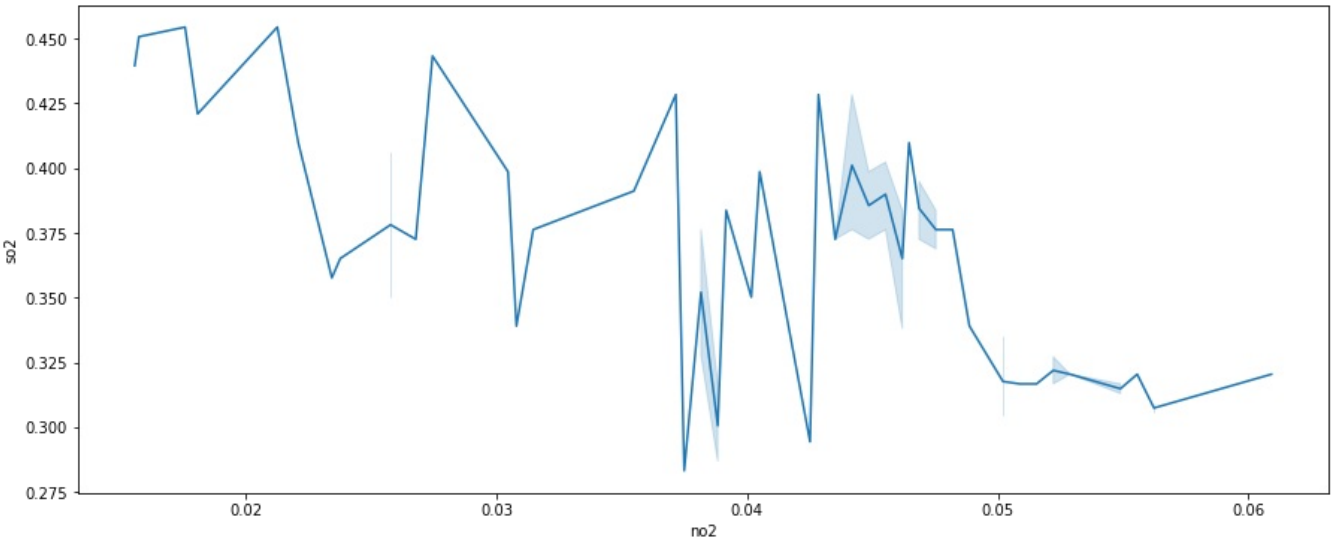

pm25
pm10
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
7
8
9
10
11
12
13

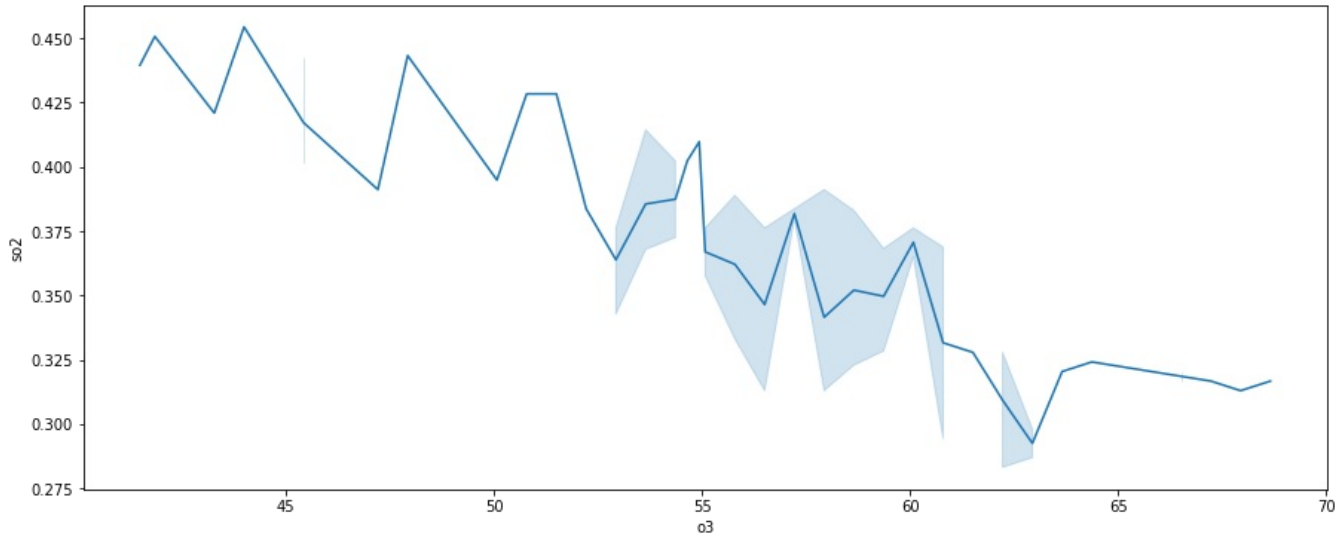

pm25
so2
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0.275
0.300
0.325
0.350
0.375
0.400
0.425
0.450

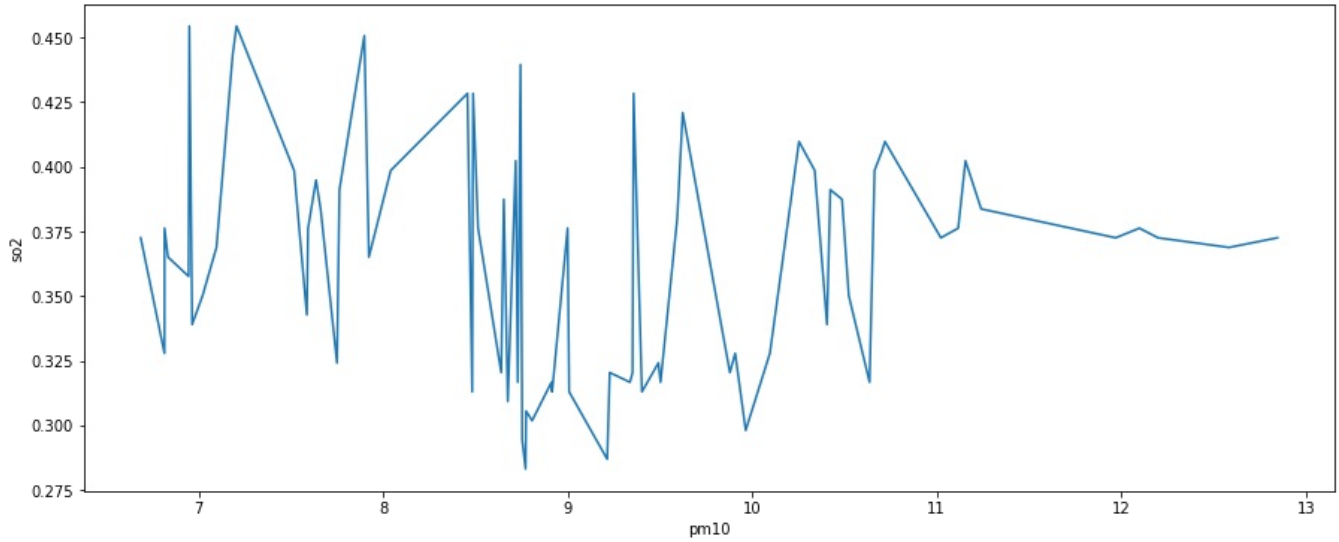

pm25
ts
+1.666e9
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
400000
450000
500000
550000
600000

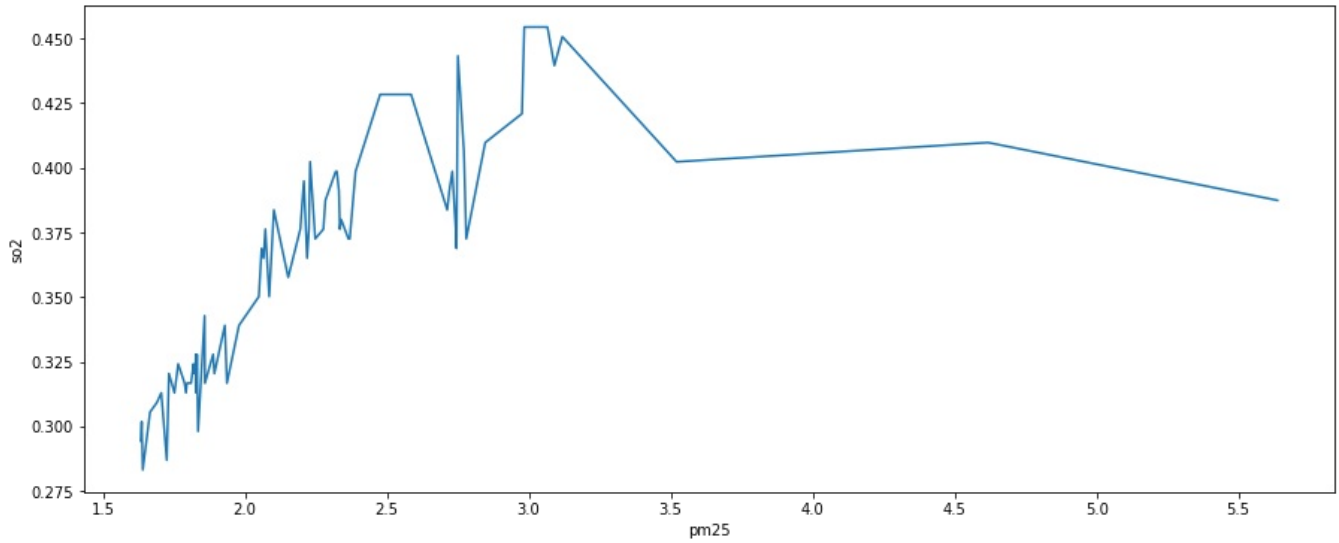

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
20
22
24
26
28
30
32
AQI

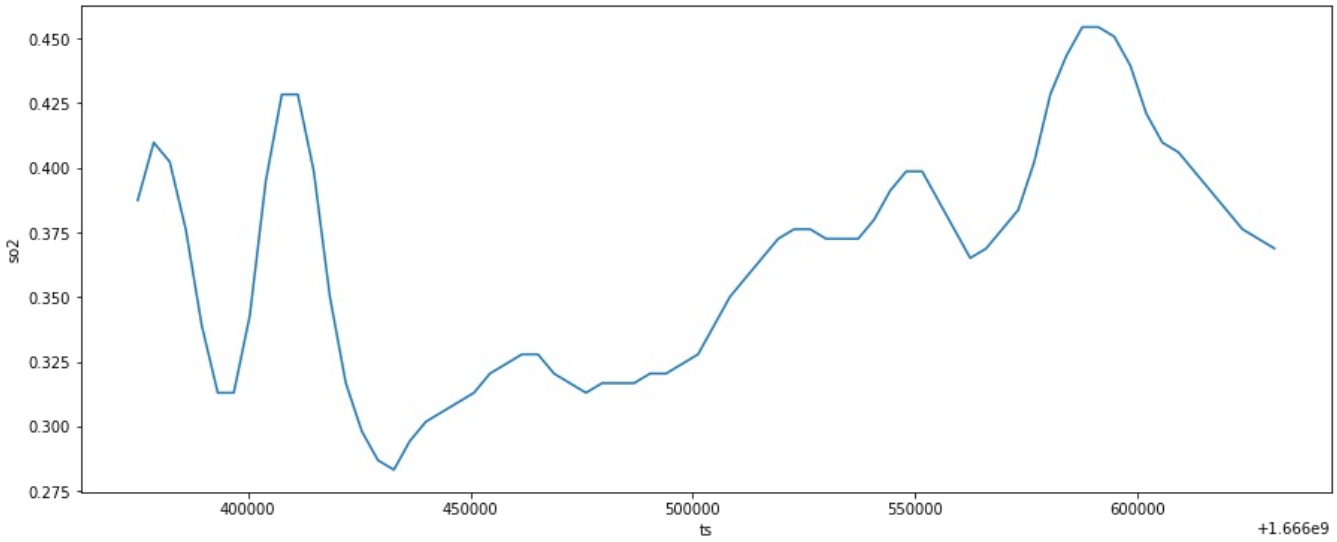

so2
0.46
0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
192.5
195.0
197.5
200.0
202.5
205.0
207.5
CO

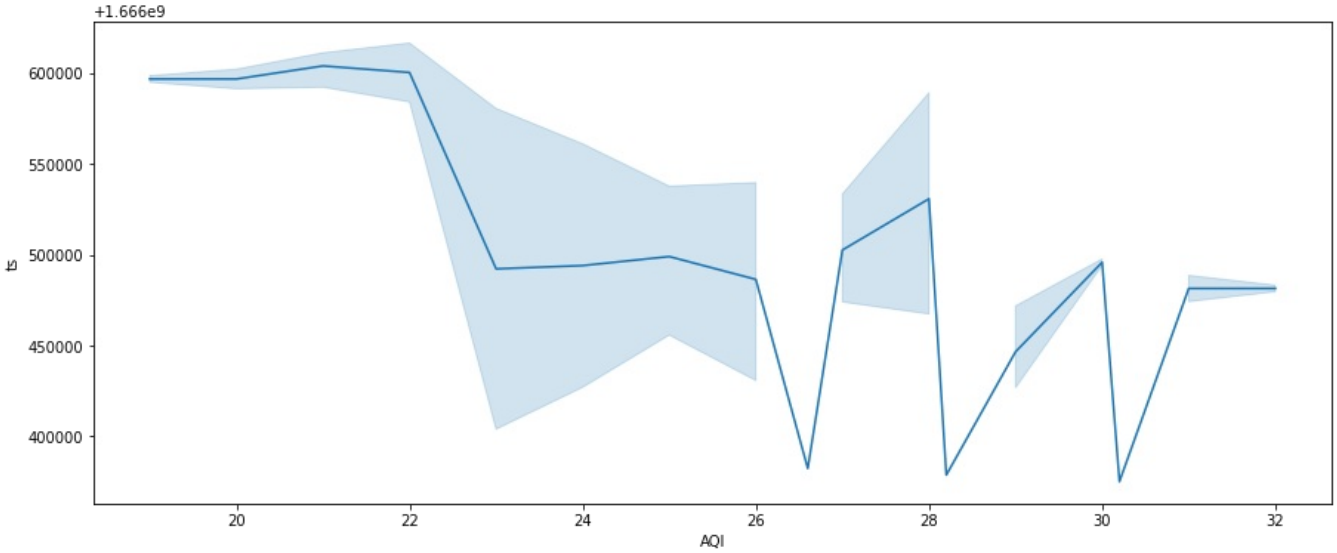

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
0.02
0.03
0.04
0.05
0.06
no2

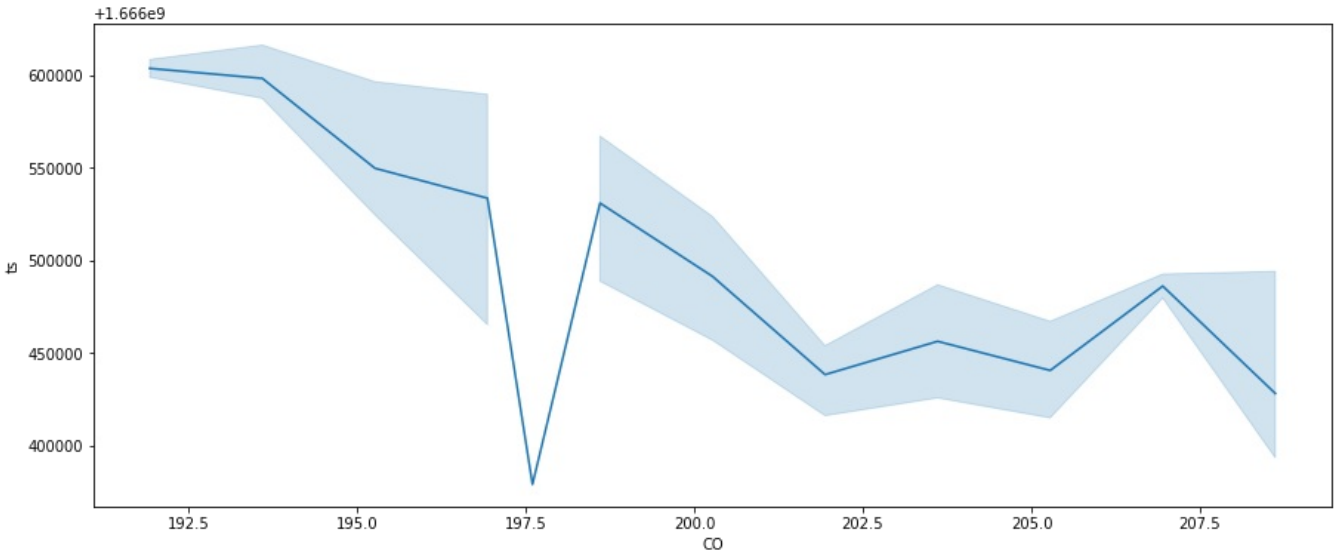

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
45
50
55
60
65
70
o3

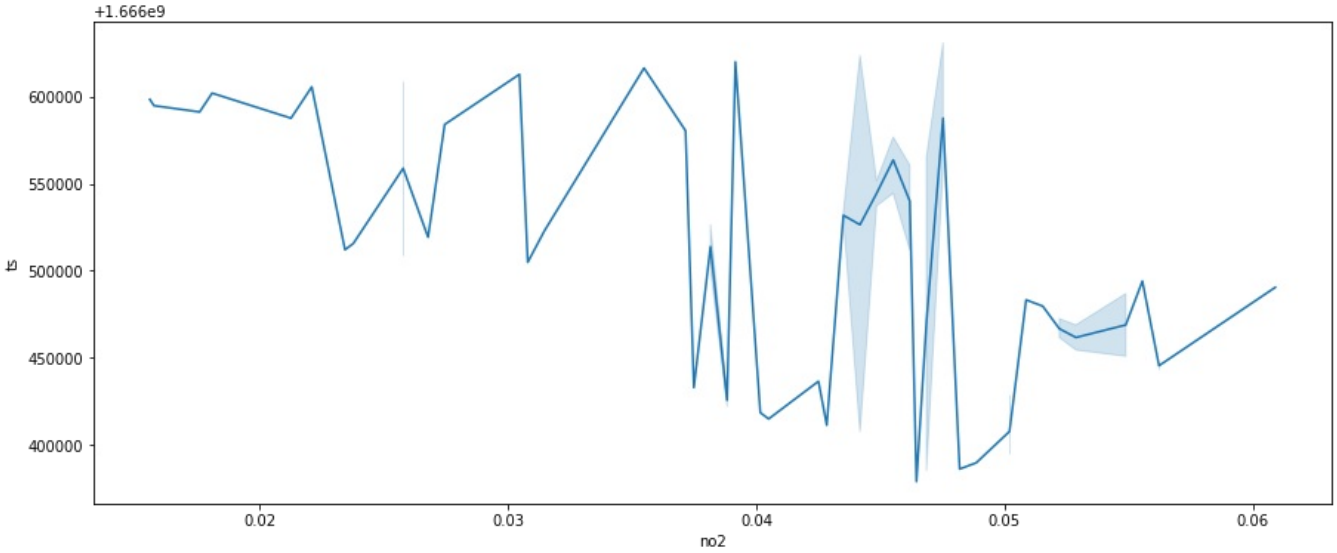

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
7
8
9
10
11
12
13
pm10

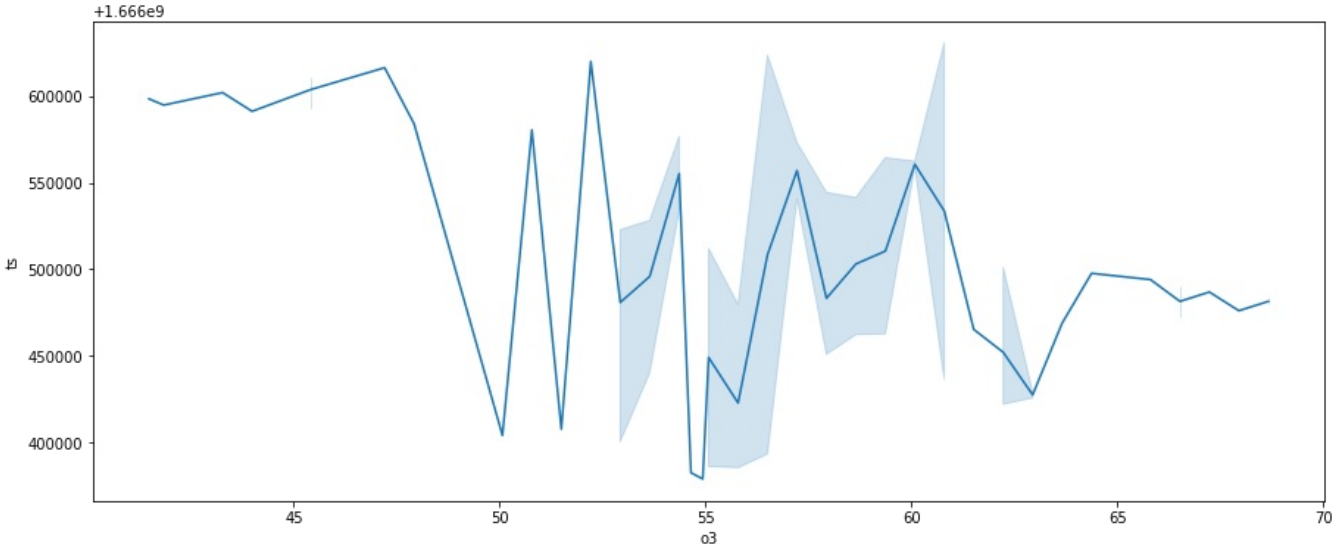

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
pm25

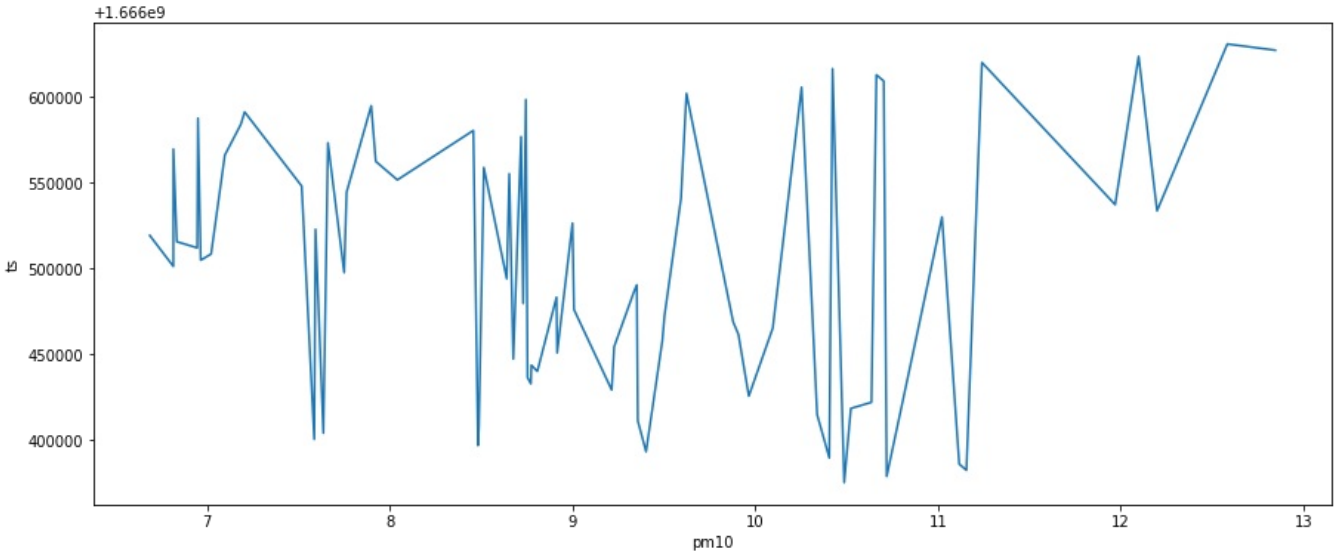

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
400000
450000
500000
550000
600000
ts
+1.666e9

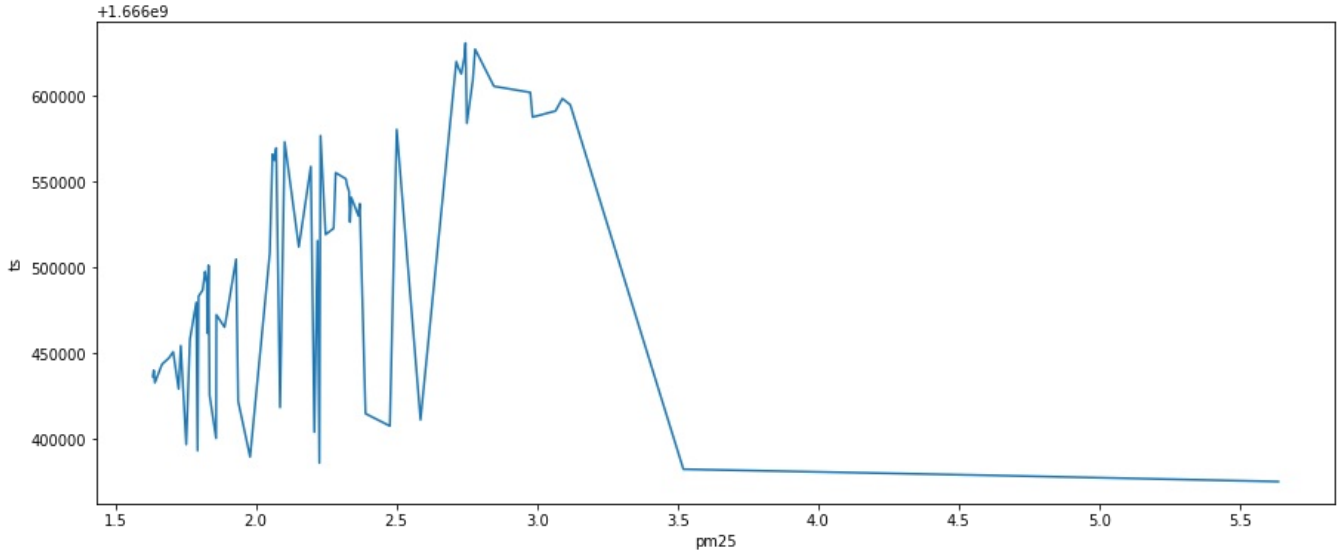

+1.666e9
ts
600000
550000
500000
450000
400000
20
22
24
26
28
30
32
AQI

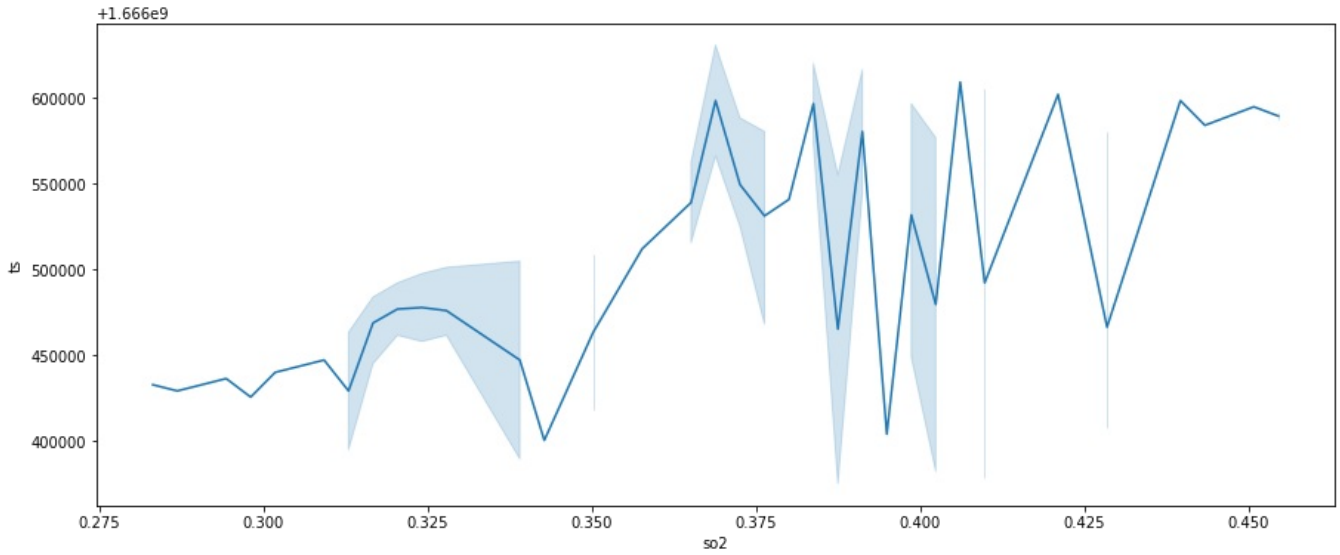

+1.666e9
ts
600000
550000
500000
450000
400000
192.5
195.0
197.5
200.0
202.5
205.0
207.5
CO


+1.666e9
600000
550000
500000
450000
400000
ts
0.02
0.03
0.04
0.05
0.06
no2

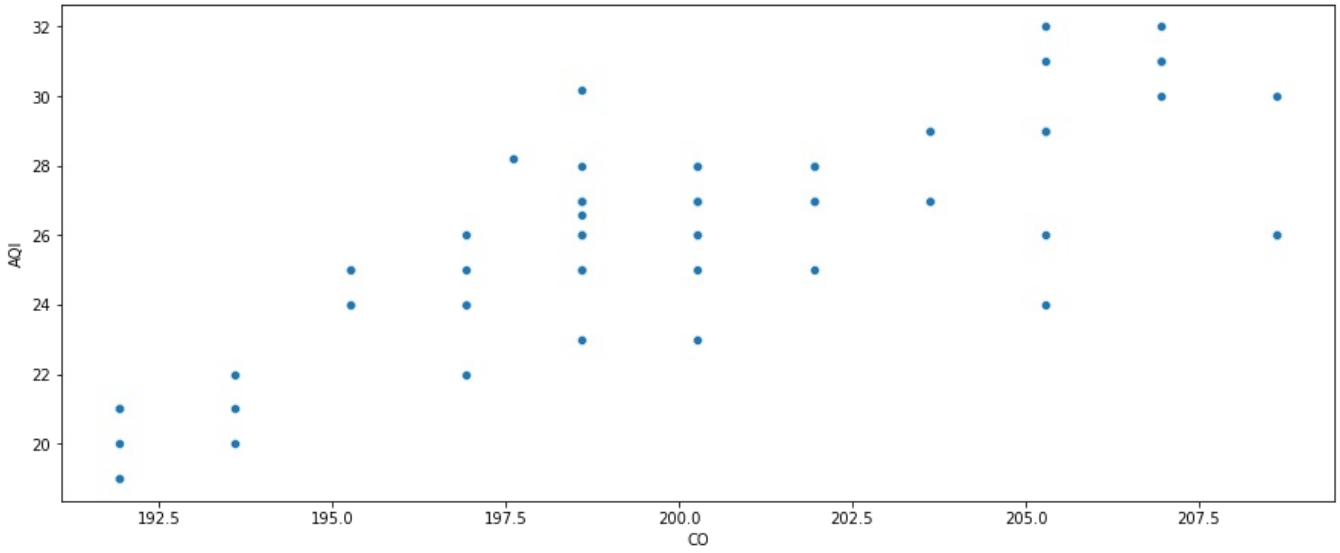

+1.666e9
600000
550000
500000
450000
400000
ts
45
50
55
60
65
70
o3

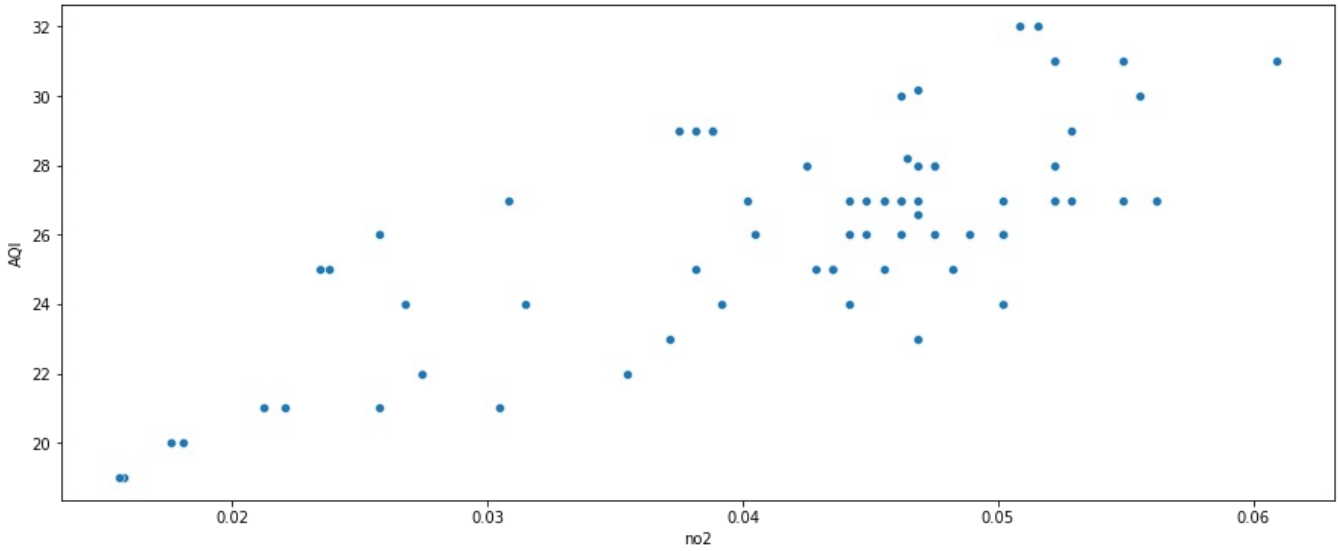

+1.666e9
600000
550000
500000
450000
400000
ts
7
8
9
10
11
12
13
pm10

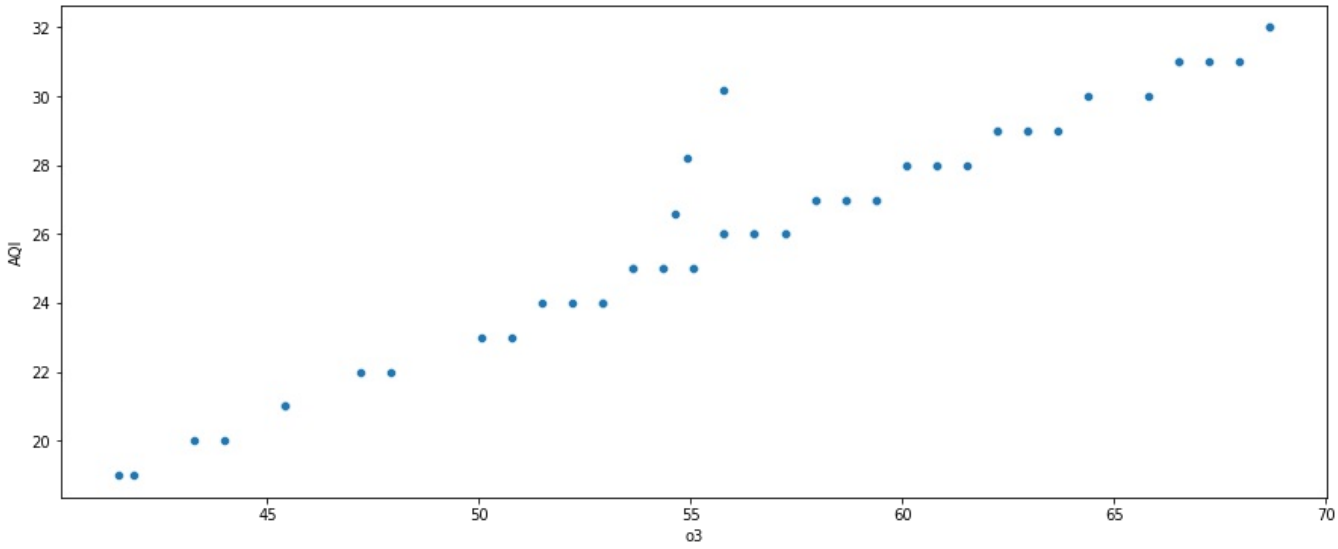


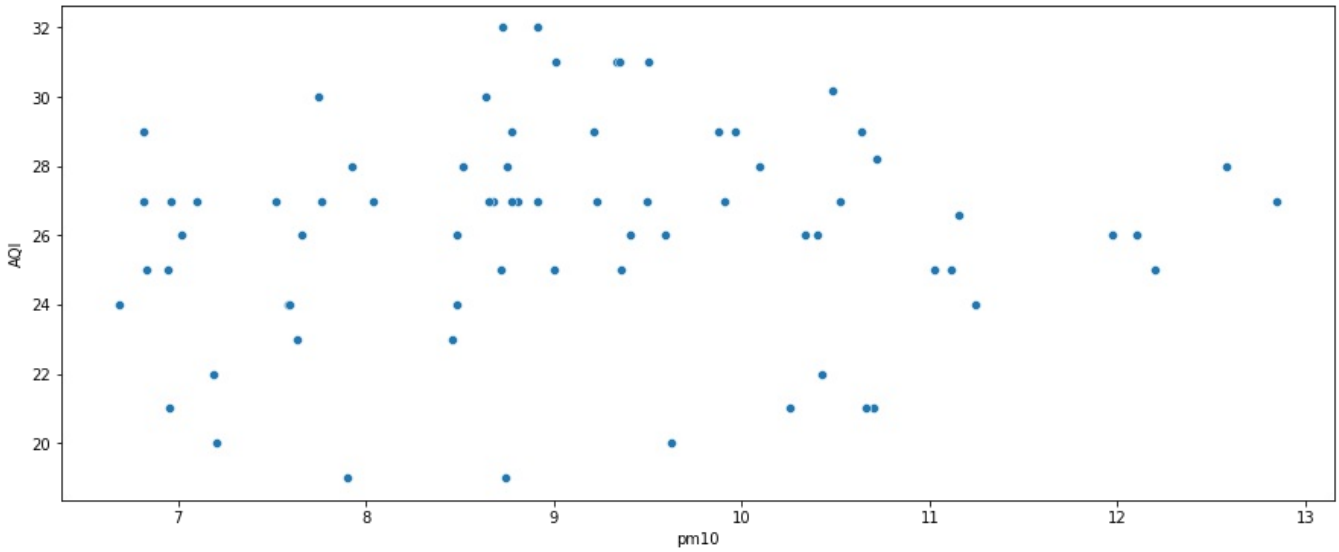


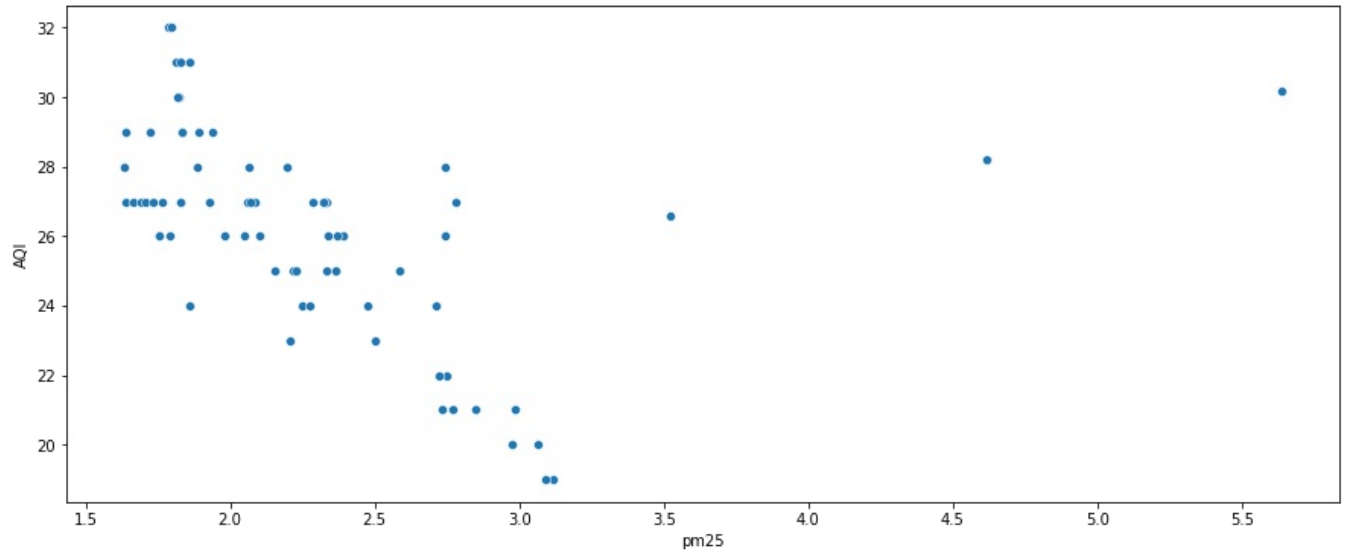
In [34]:
```
for i in numerical_columns:
    for j in numerical_columns:
        if i != j:
            plt.figure(figsize=(15,6))
            sns.scatterplot(x = df[j], y = df[i], data = df, palette = 'hls')
            plt.show()
```

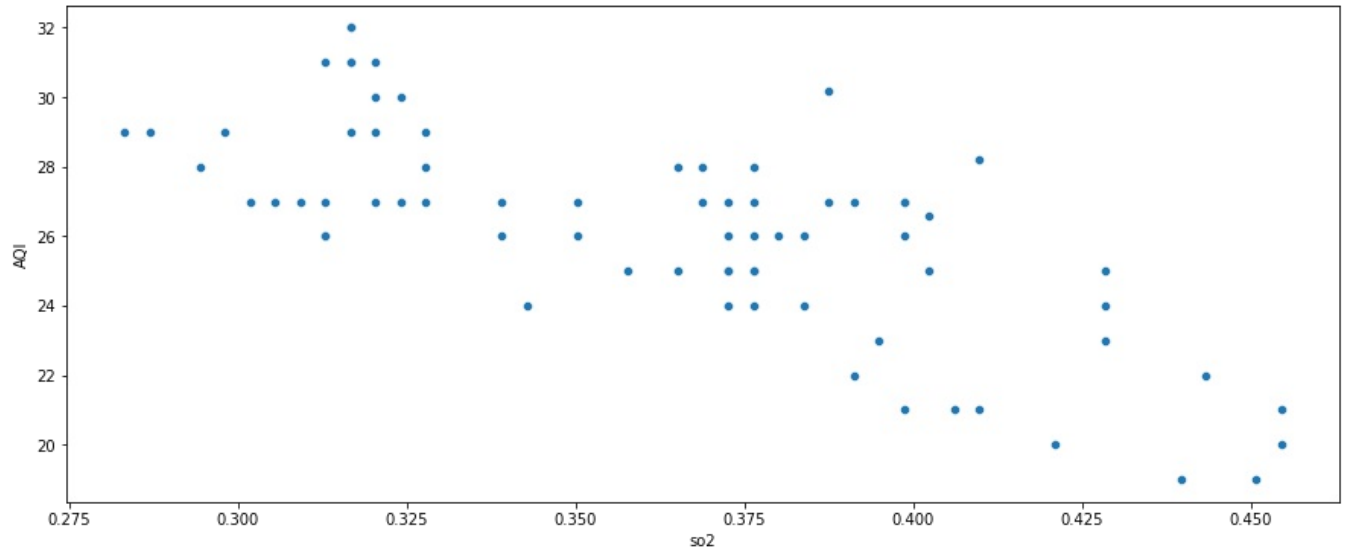


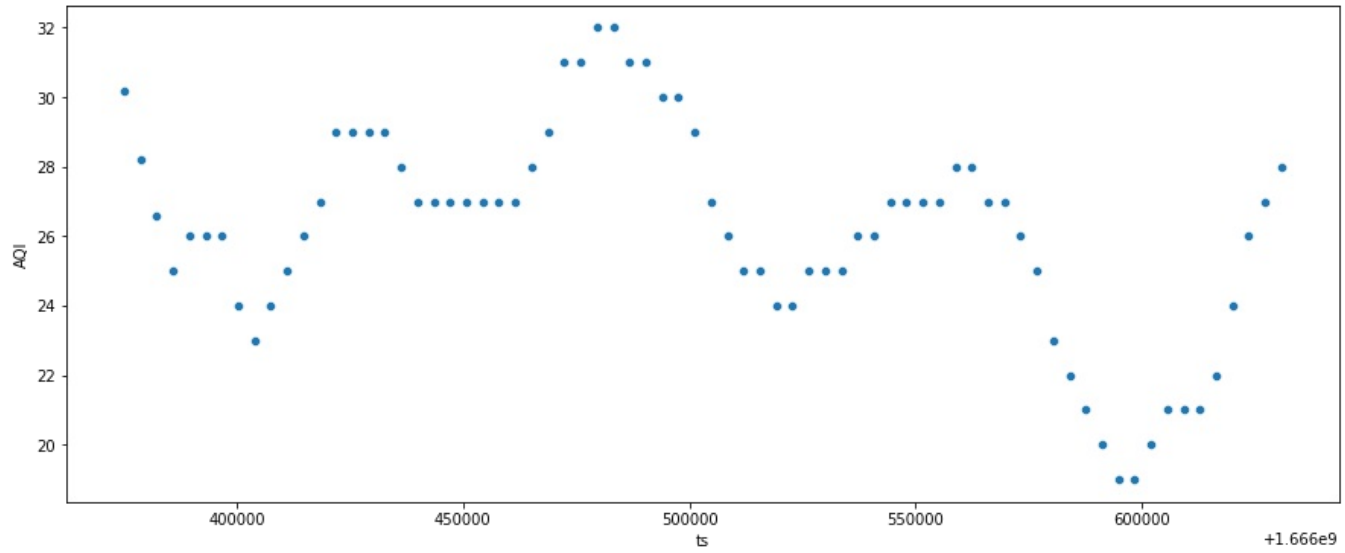

AQI
32
30
28
26
24
22
20
0.02
0.03
0.04
0.05
0.06
no2

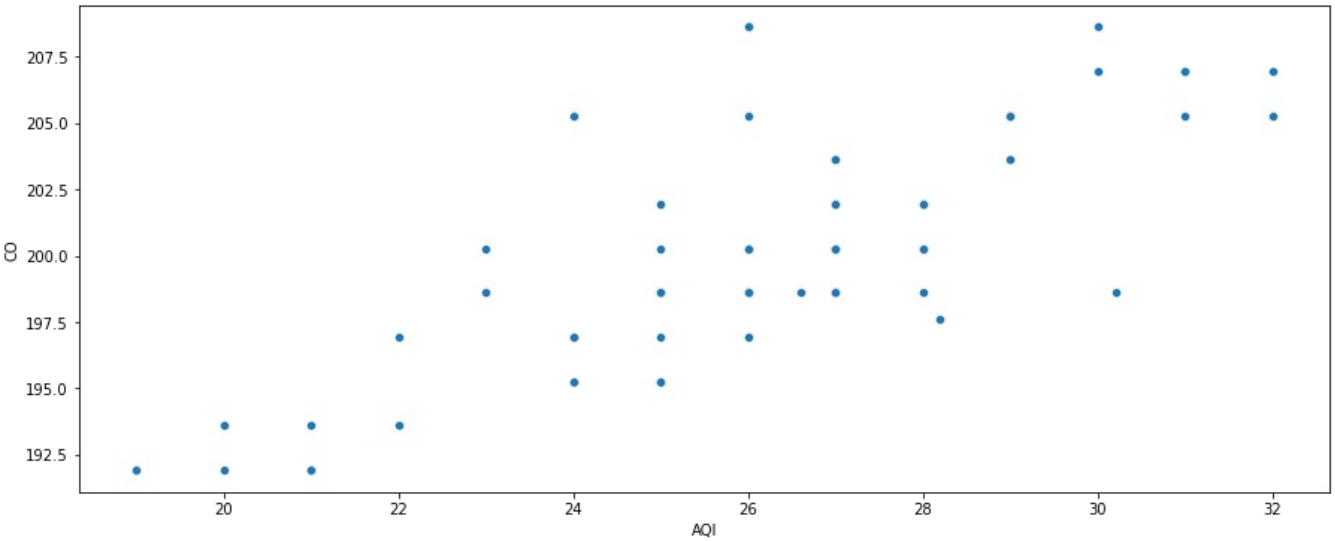

AQI
32
30
28
26
24
22
20
45
50
55
60
65
70
o3

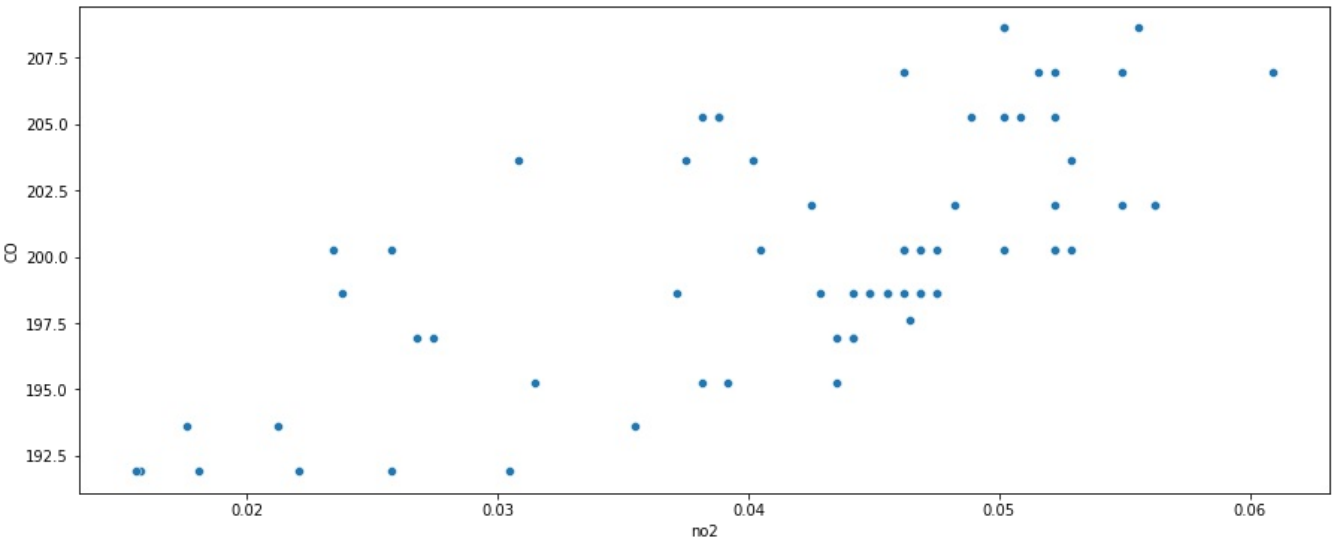

AQI
32
30
28
26
24
22
20
7
8
9
10
11
12
13
pm10

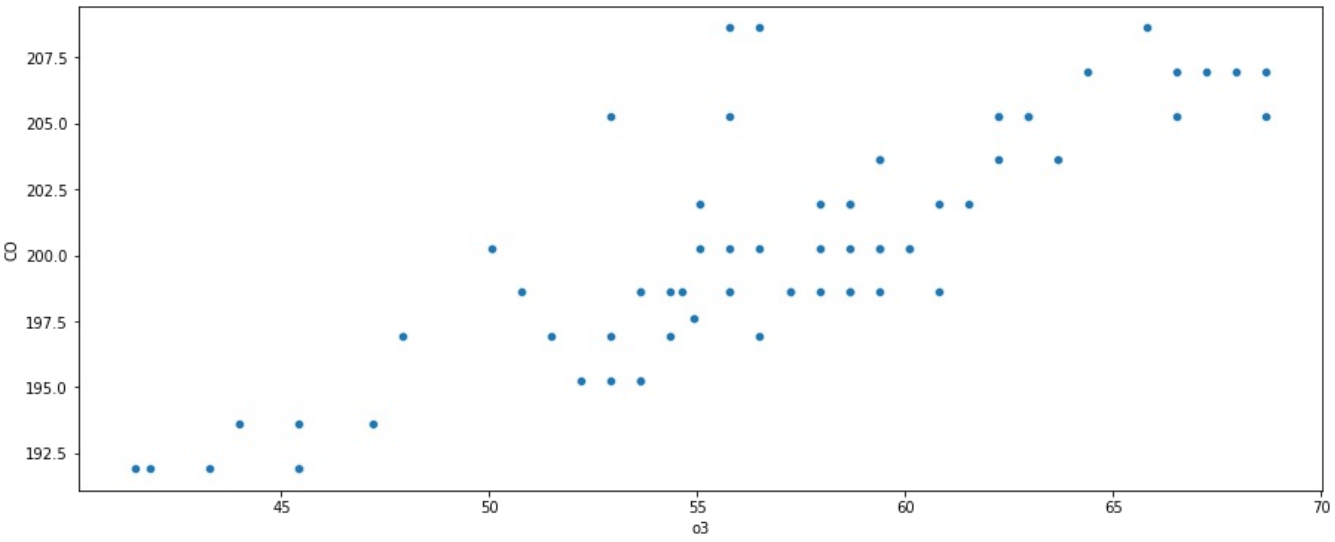

AQI
pm25

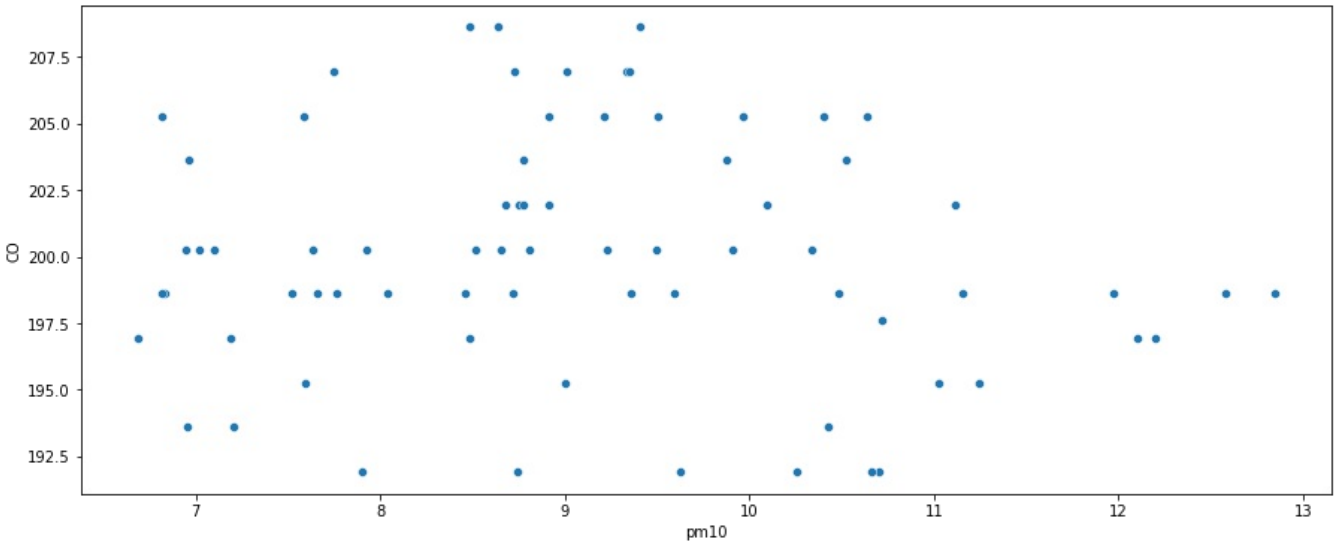

AQI
so2

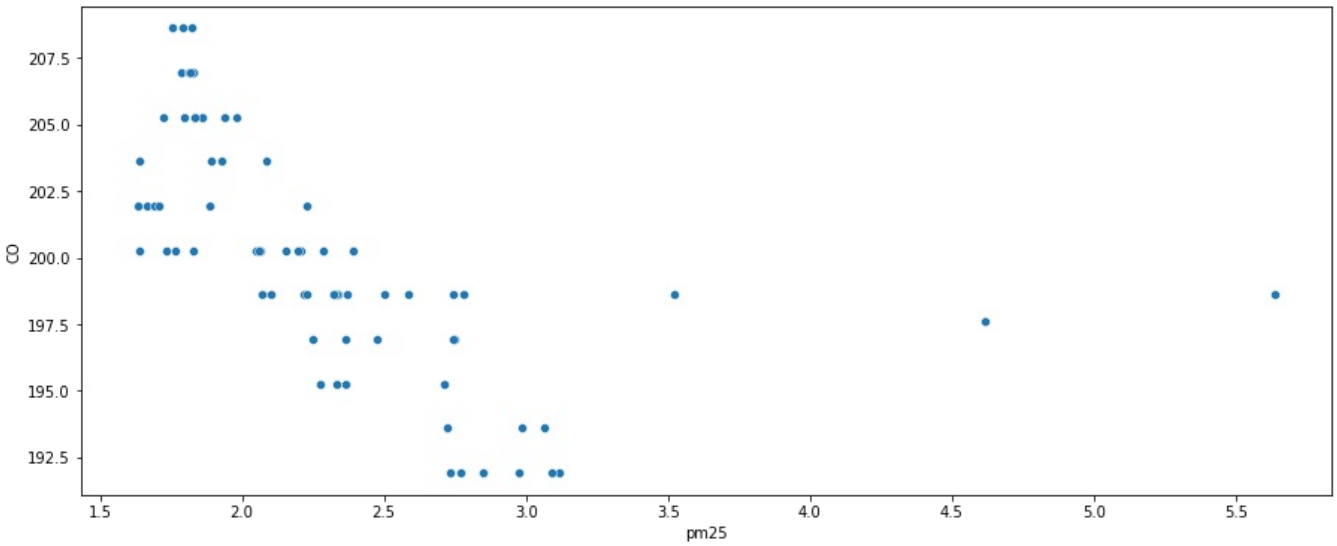

AQI
ts
+1.666e9

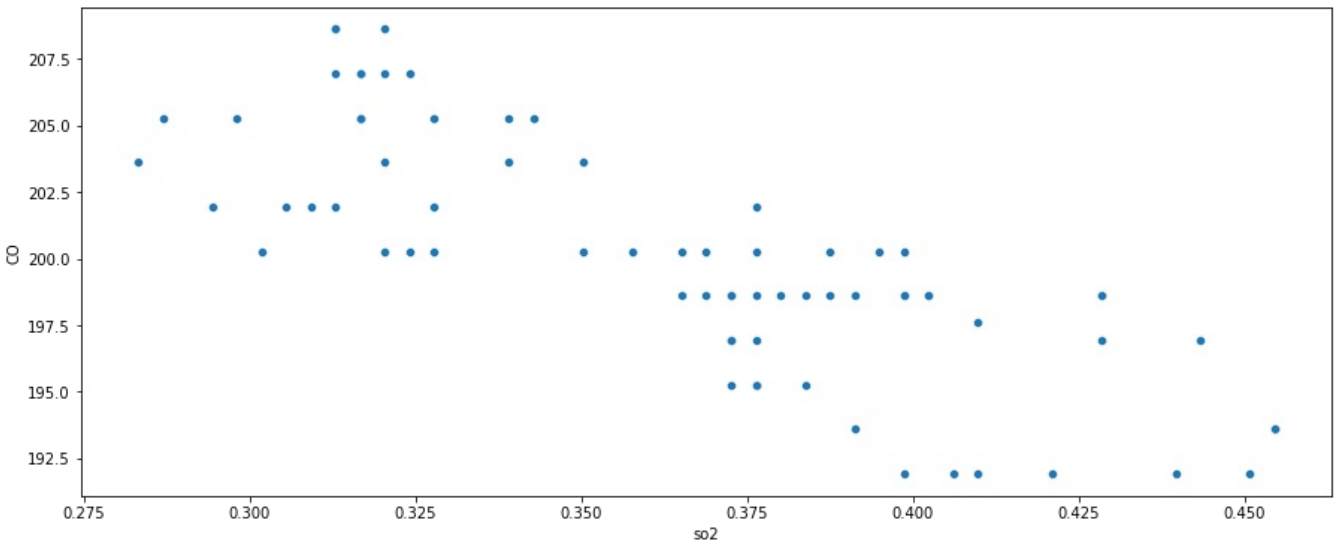

CO
AQI
207.5
205.0
202.5
200.0
197.5
195.0
192.5
20
22
24
26
28
30
32

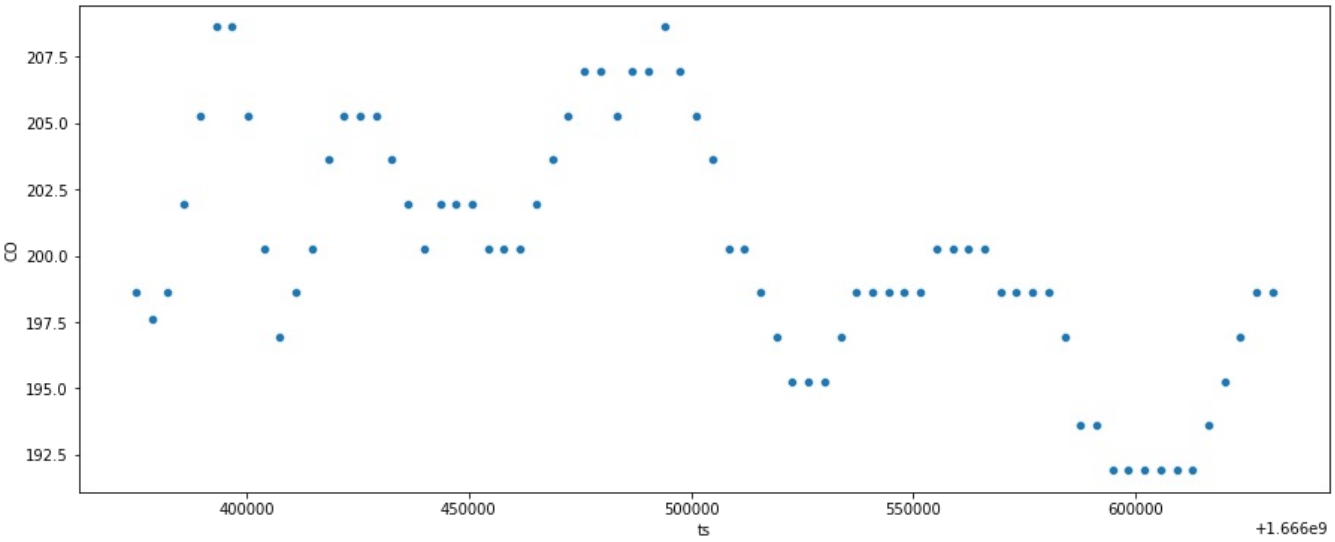

CO
no2
207.5
205.0
202.5
200.0
197.5
195.0
192.5
0.02
0.03
0.04
0.05
0.06

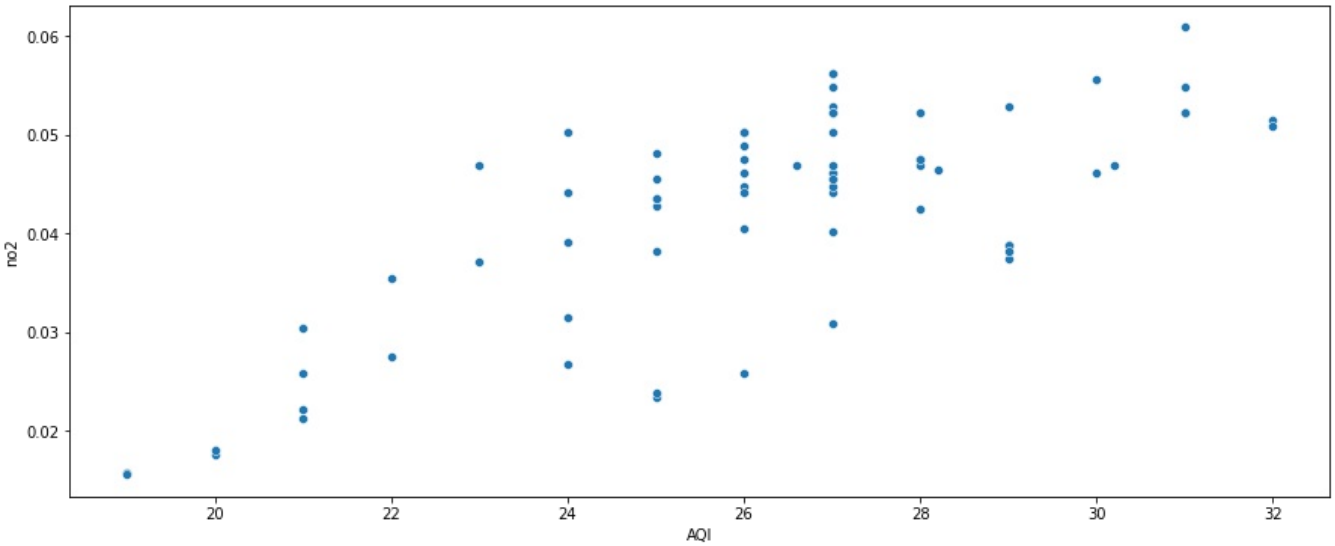

CO
o3
207.5
205.0
202.5
200.0
197.5
195.0
192.5
45
50
55
60
65
70

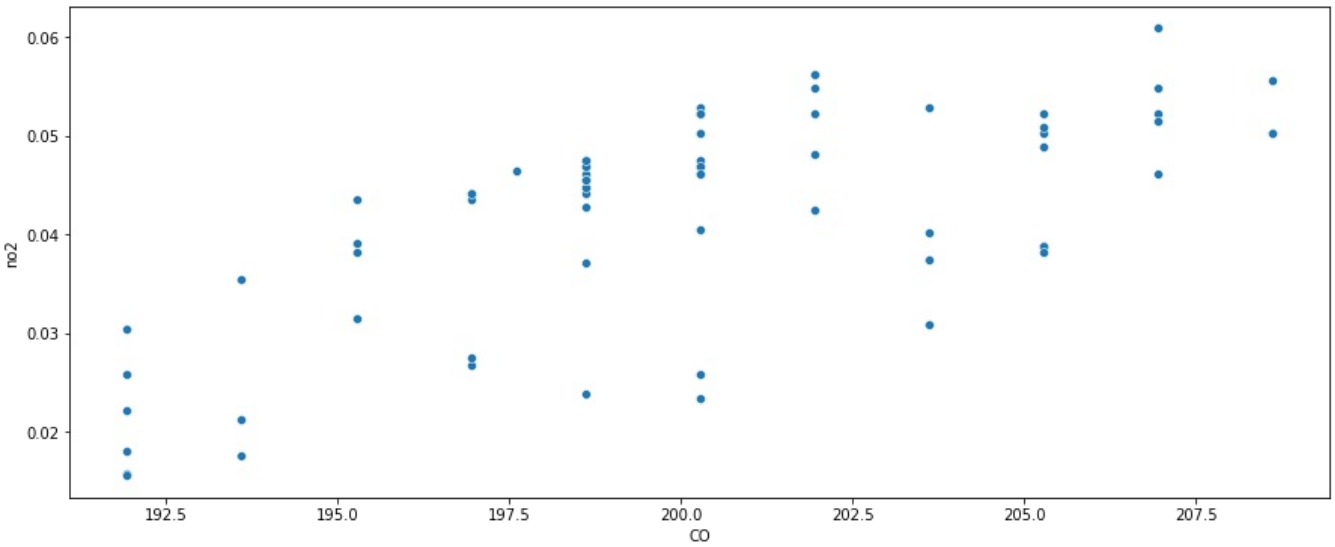

CO
207.5
205.0
202.5
200.0
197.5
195.0
192.5
7
8
9
10
11
12
13
pm10

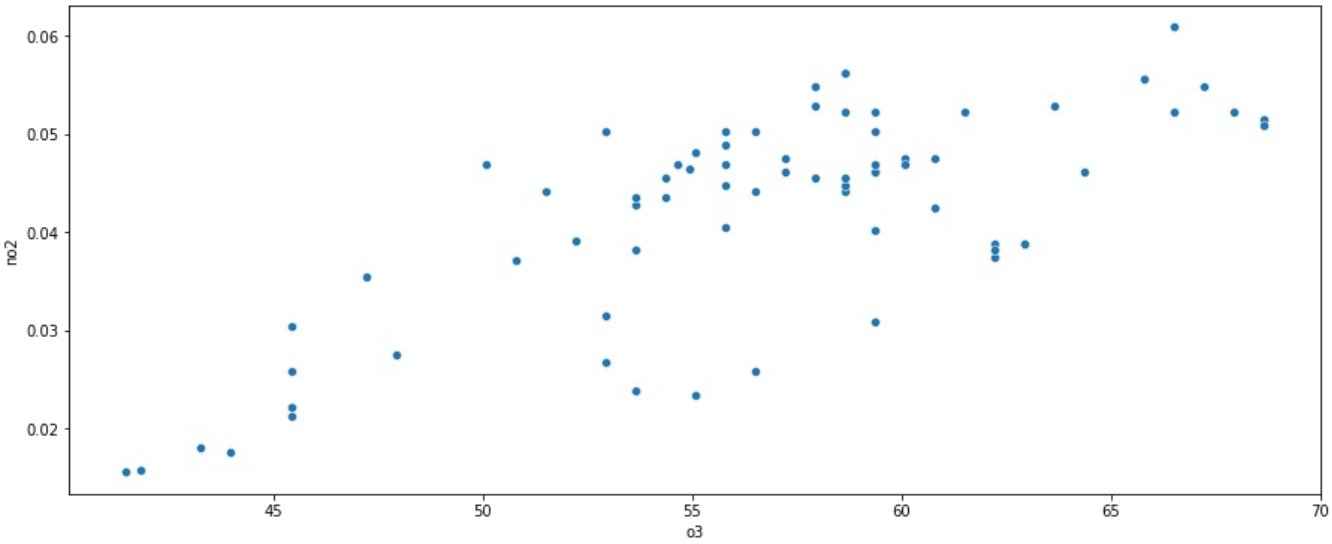

CO
207.5
205.0
202.5
200.0
197.5
195.0
192.5
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
pm25

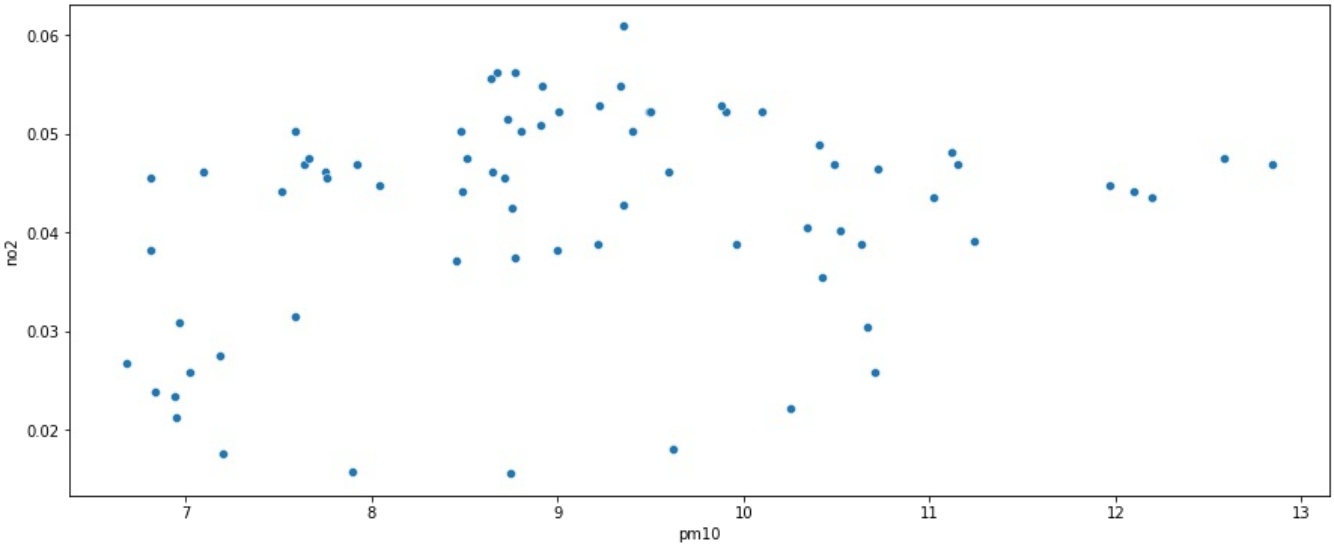

CO
207.5
205.0
202.5
200.0
197.5
195.0
192.5
0.275
0.300
0.325
0.350
0.375
0.400
0.425
0.450
so2

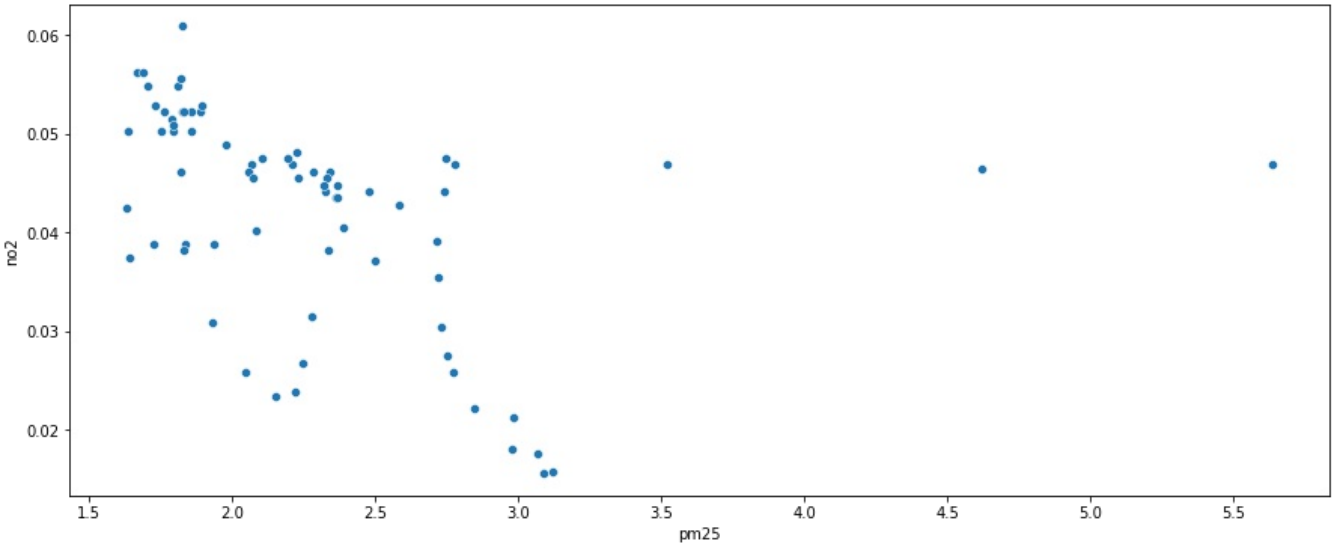

CO
207.5
205.0
202.5
200.0
197.5
195.0
192.5
400000
450000
500000
550000
600000
ts
+1.666e9

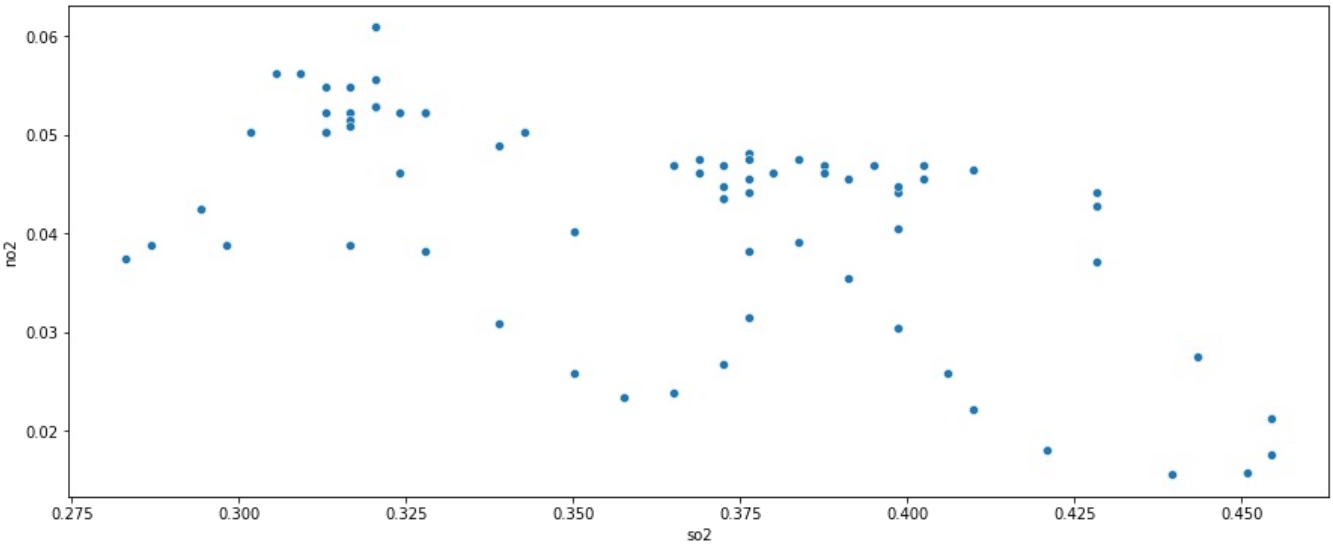

no2
0.06
0.05
0.04
0.03
0.02
20
22
24
26
28
30
32
AQI

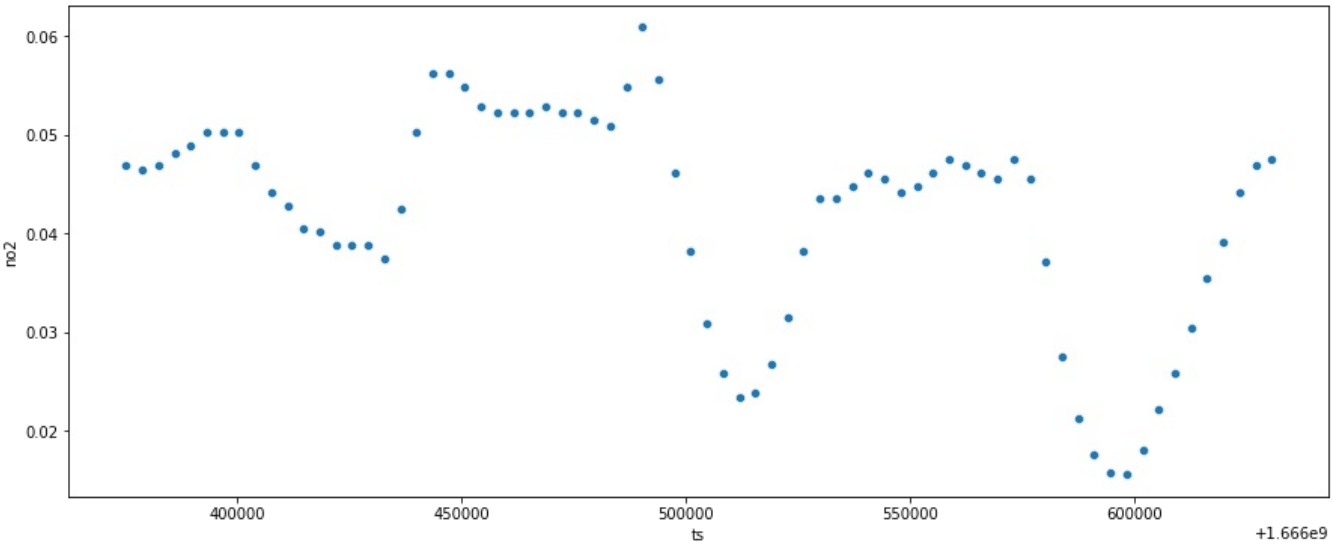

no2
0.06
0.05
0.04
0.03
0.02
192.5
195.0
197.5
200.0
202.5
205.0
207.5
CO

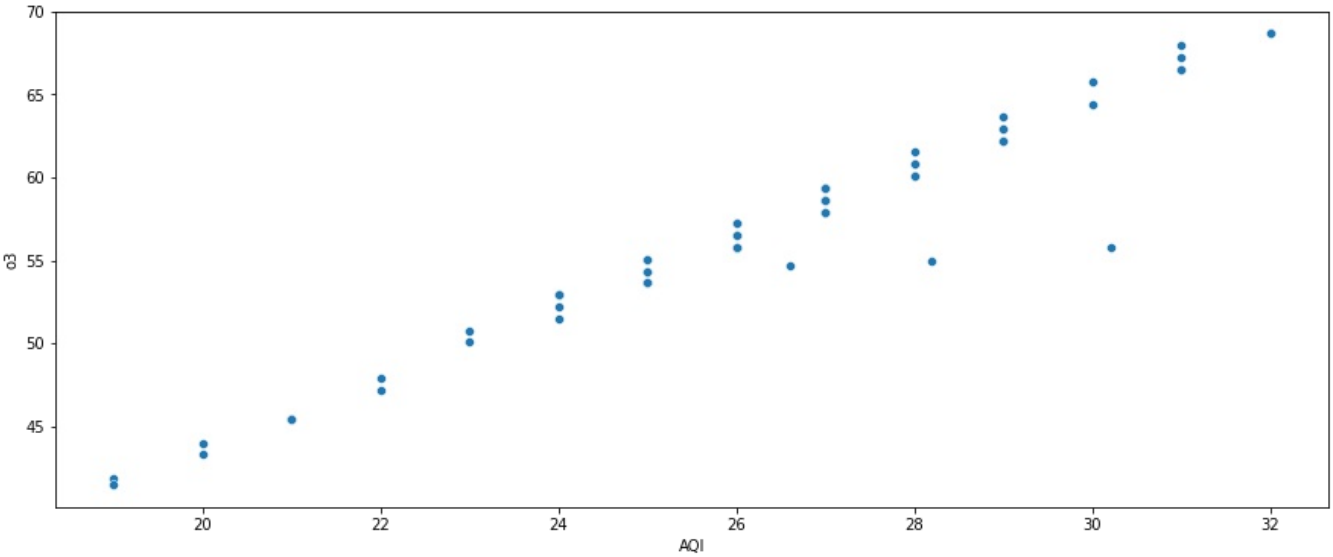

no2
0.06
0.05
0.04
0.03
0.02
45
50
55
60
65
70
o3

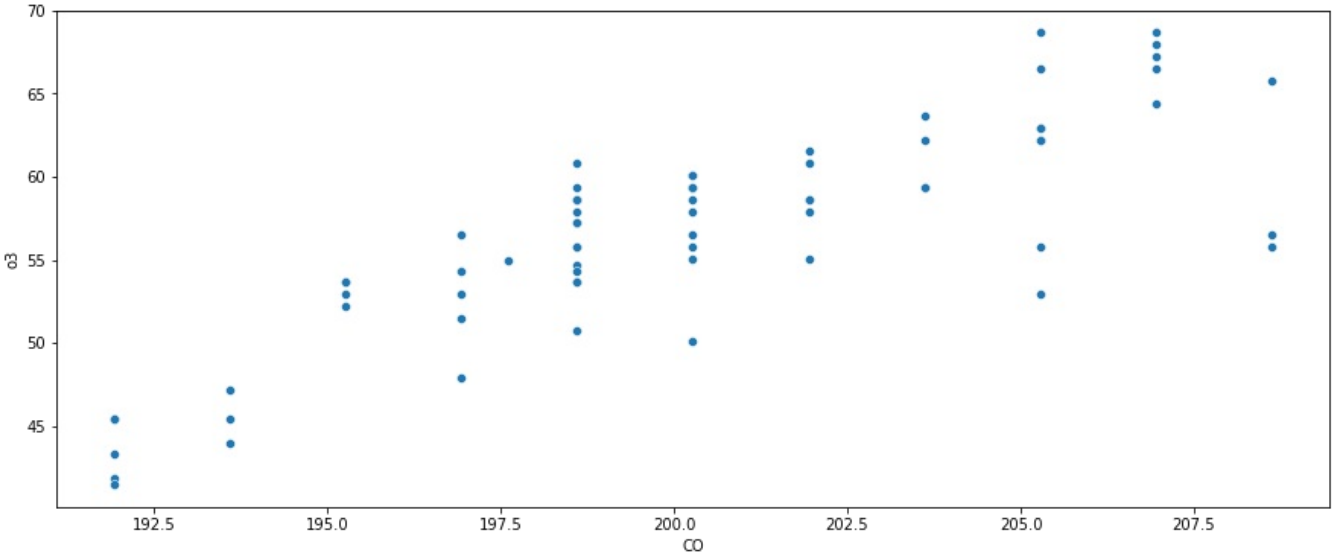

no2
0.06
0.05
0.04
0.03
0.02
7
8
9
10
11
12
13
pm10

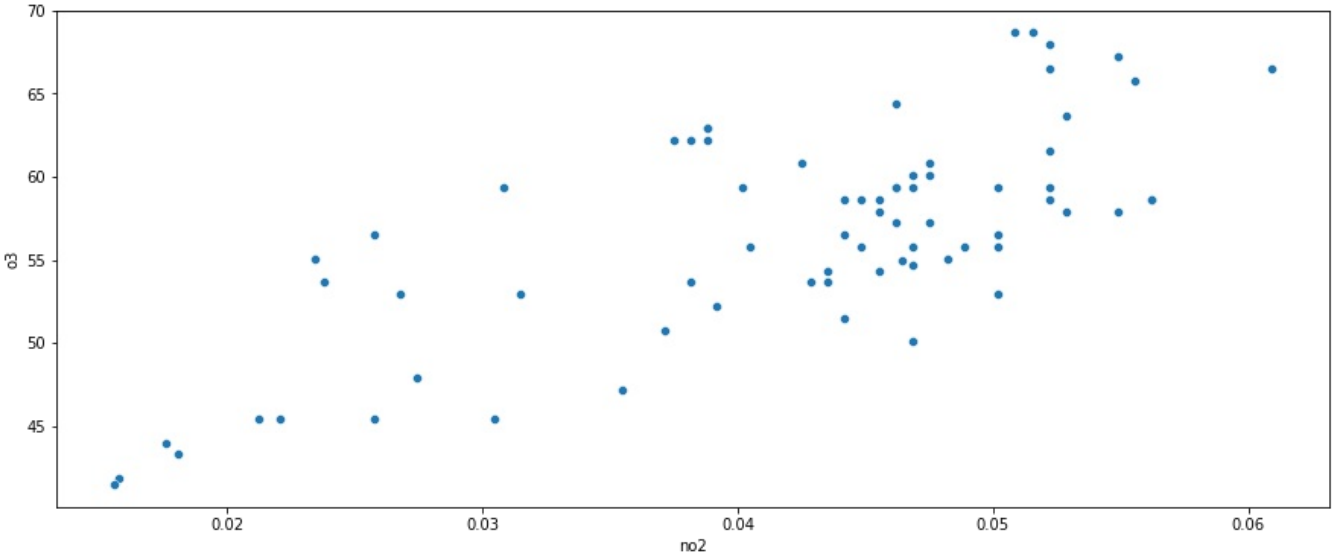

no2
0.06
0.05
0.04
0.03
0.02
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
pm25

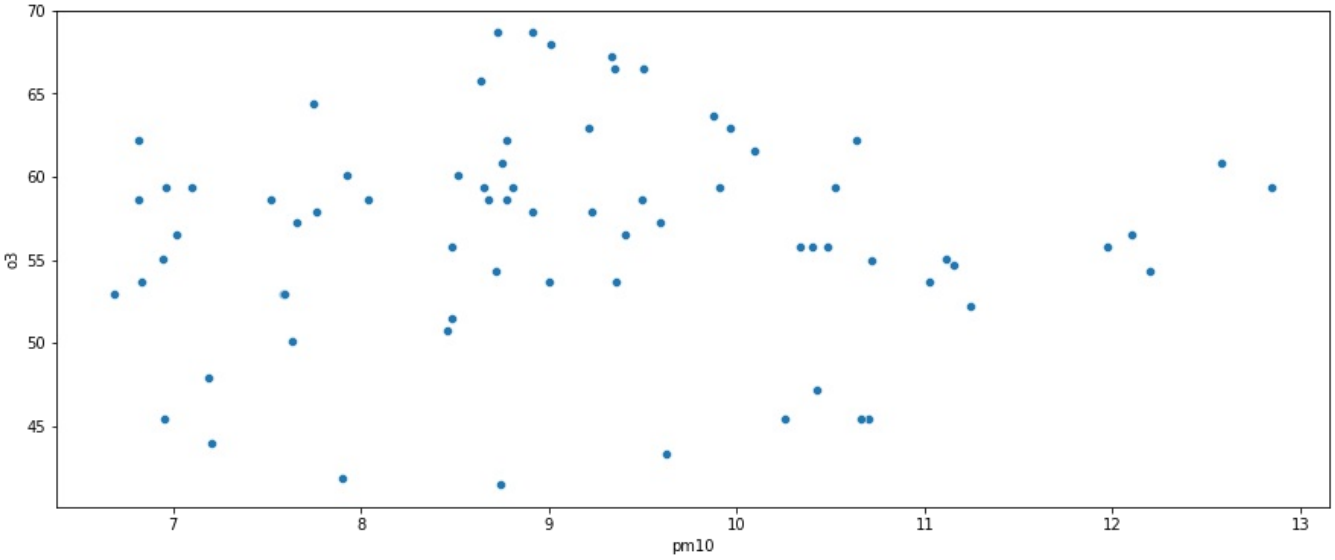

no2
0.06
0.05
0.04
0.03
0.02
0.275
0.300
0.325
0.350
0.375
0.400
0.425
0.450
so2

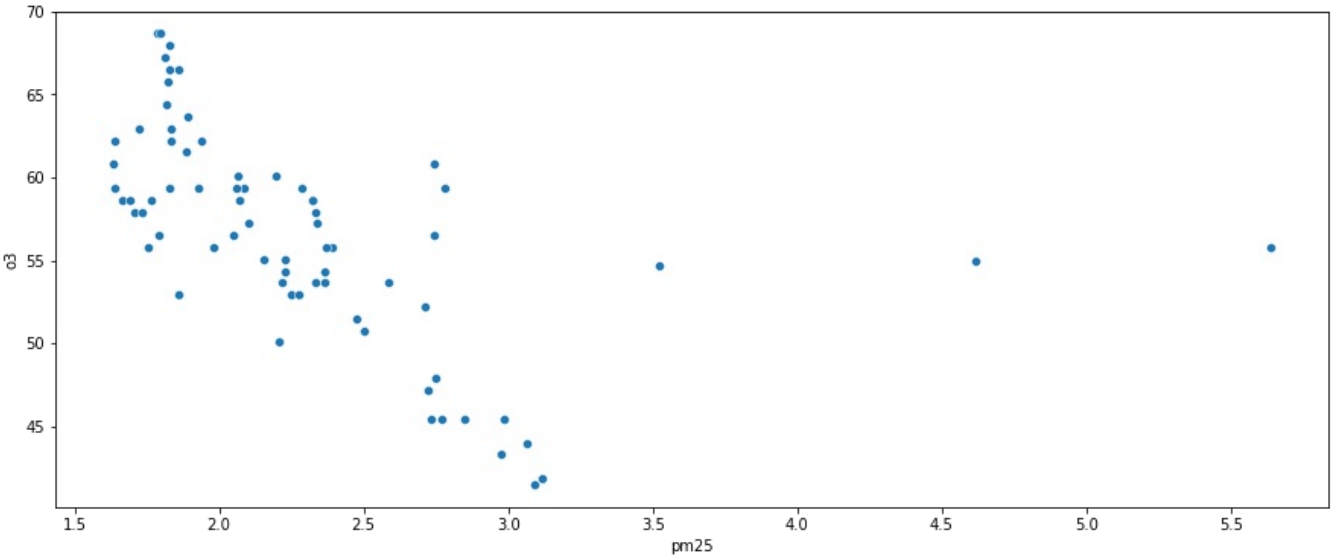

no2
0.06
0.05
0.04
0.03
0.02
400000
450000
500000
550000
600000
ts
+1.666e9

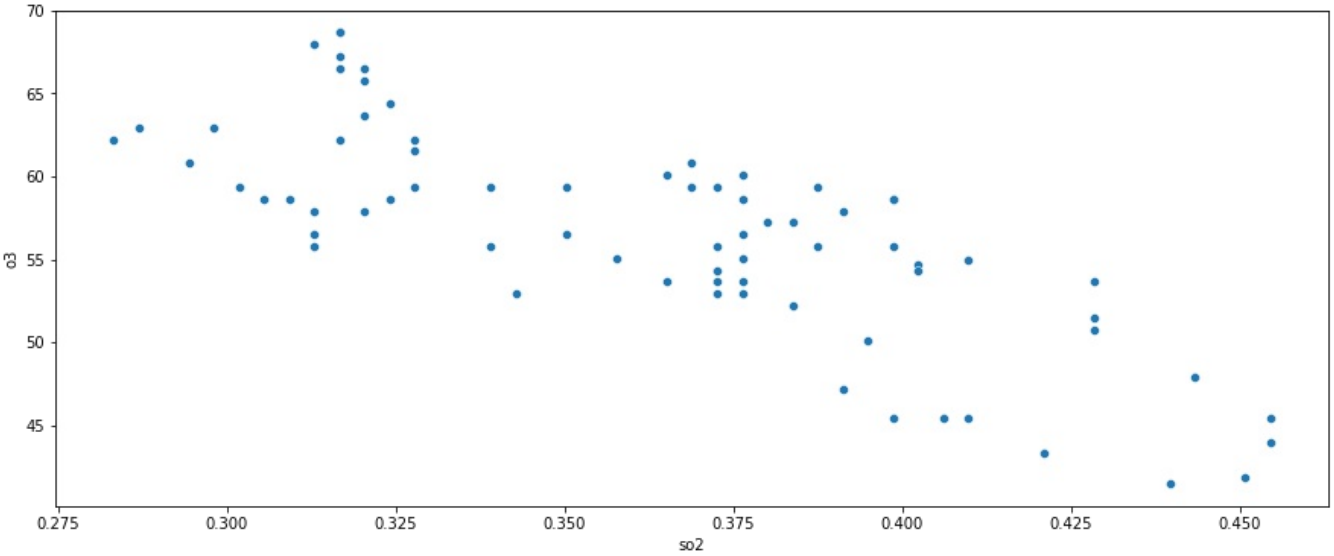

o3
70
65
60
55
50
45
20
22
24
26
28
30
32
AQI

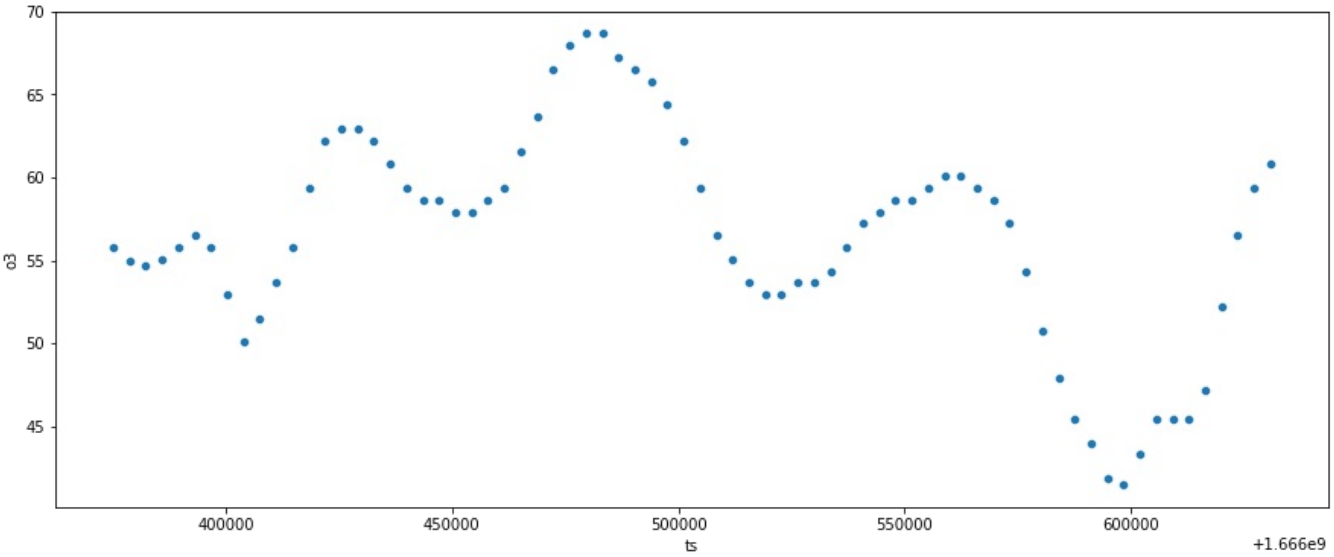

o3
CO
70
65
60
55
50
45
192.5
195.0
197.5
200.0
202.5
205.0
207.5

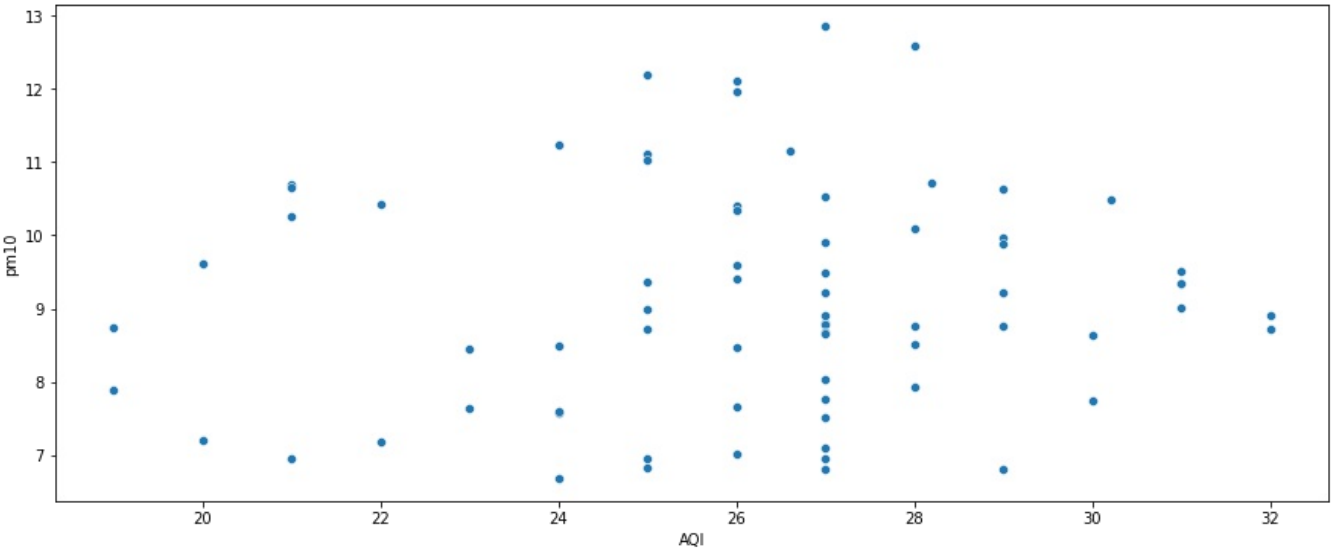

o3
no2
70
65
60
55
50
45
0.02
0.03
0.04
0.05
0.06

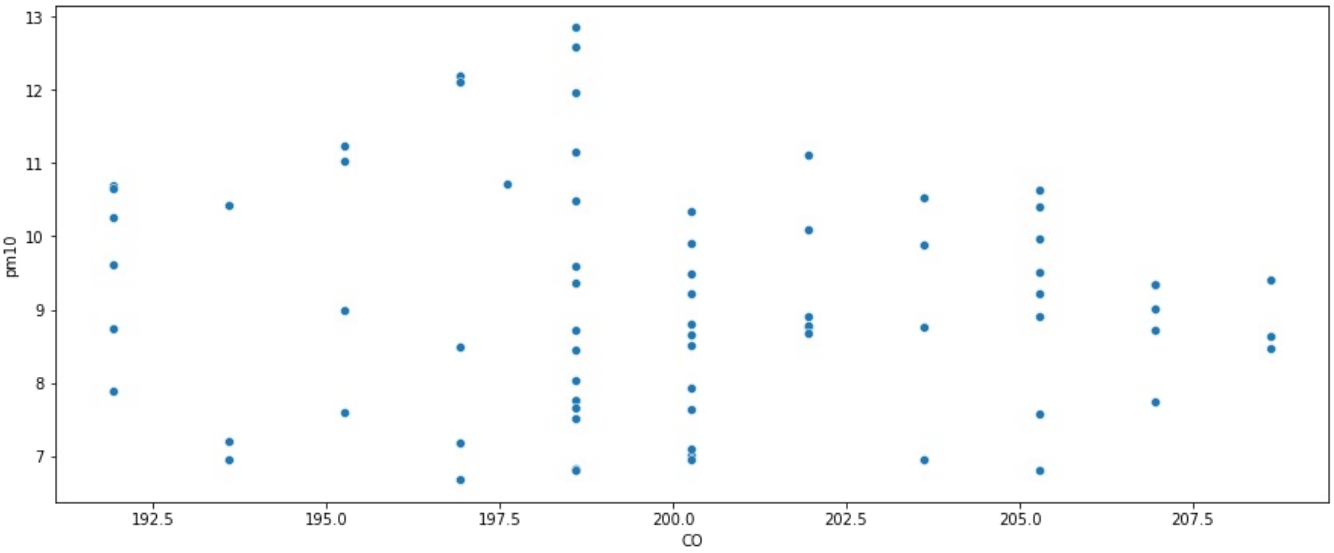

o3
pm10
70
65
60
55
50
45
7
8
9
10
11
12
13

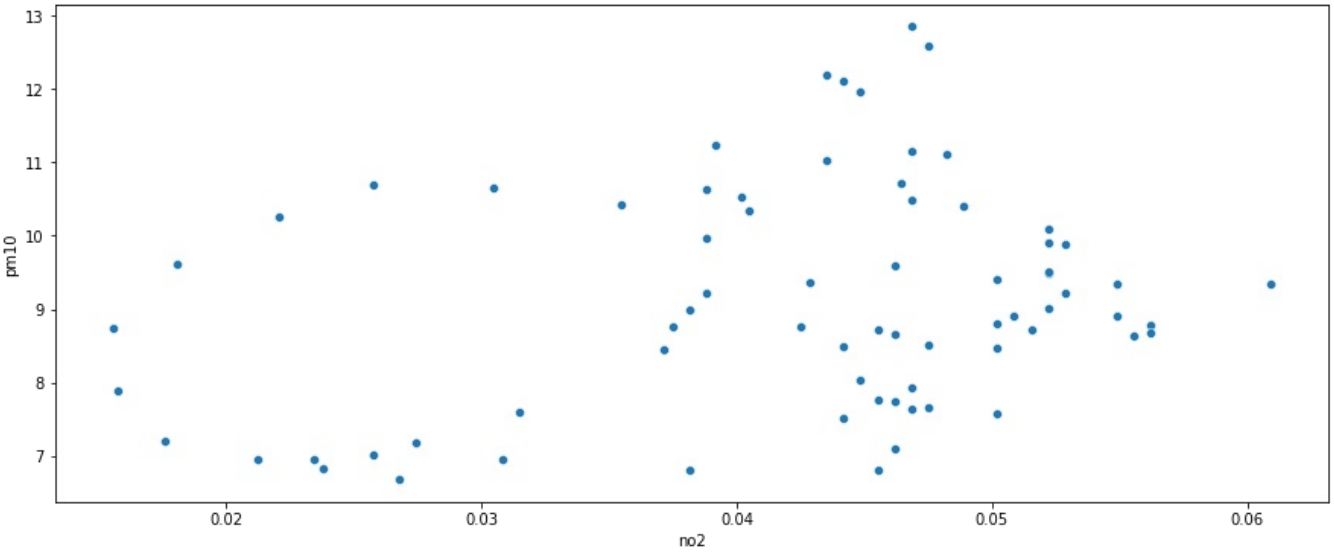

o3
pm25

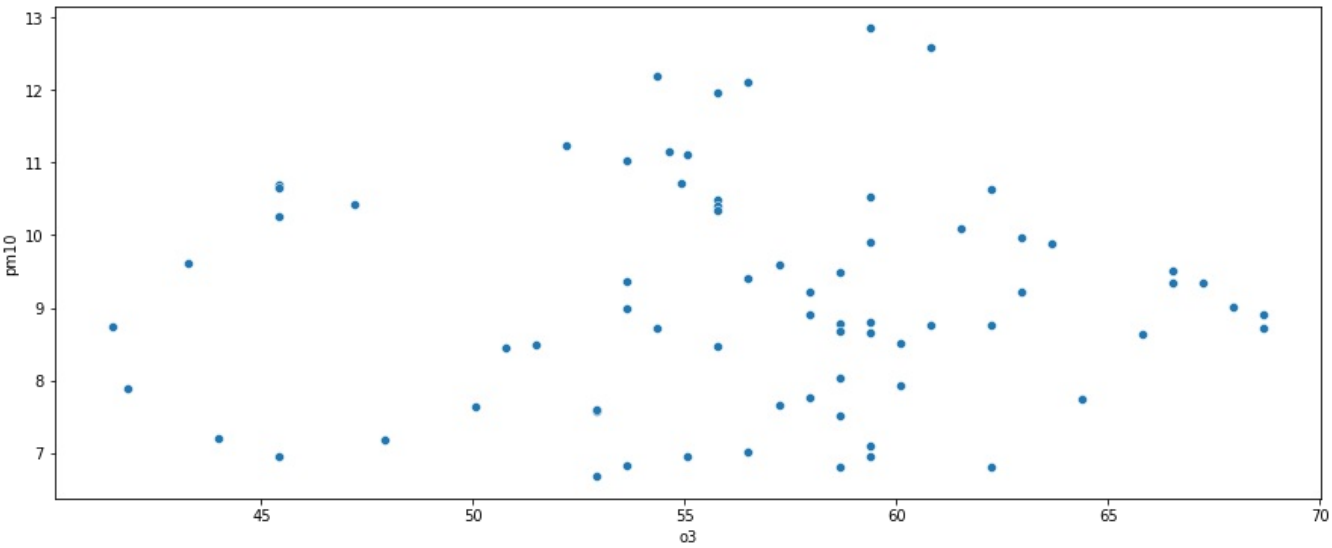

o3
so2

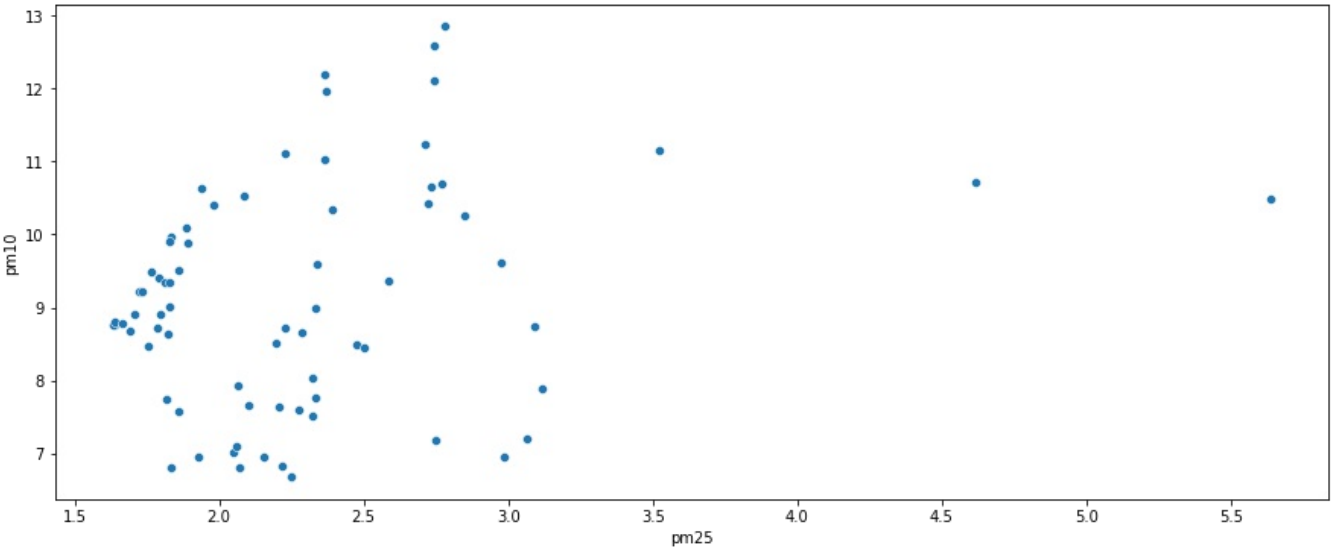

o3
ts
+1.666e9

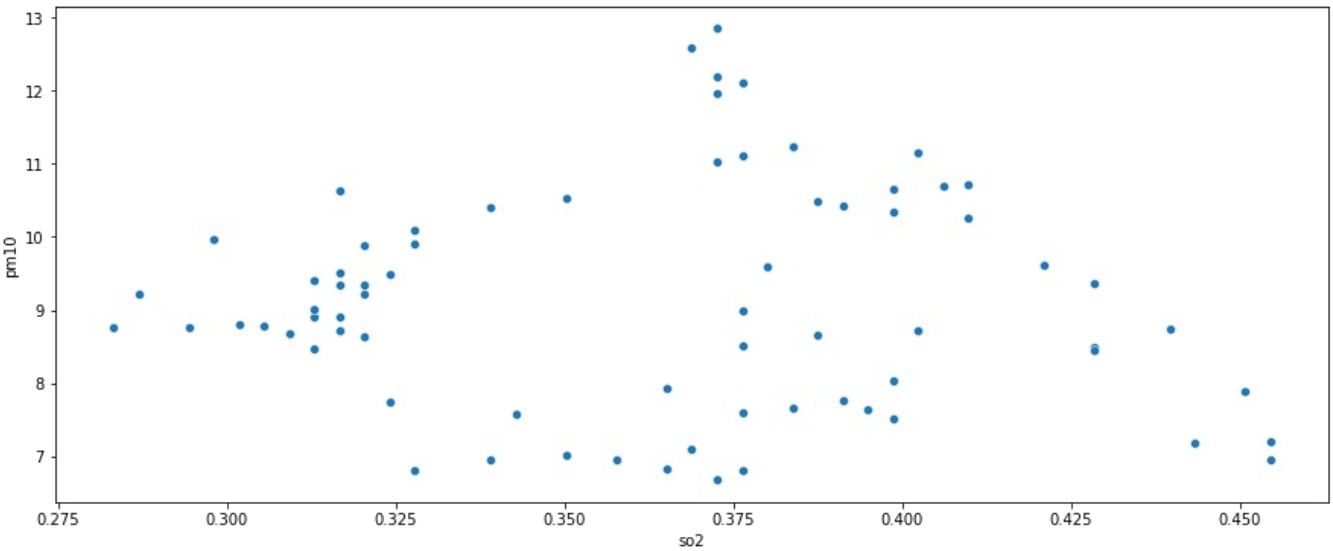

pm10
AQI
13
12
11
10
9
8
7
20
22
24
26
28
30
32

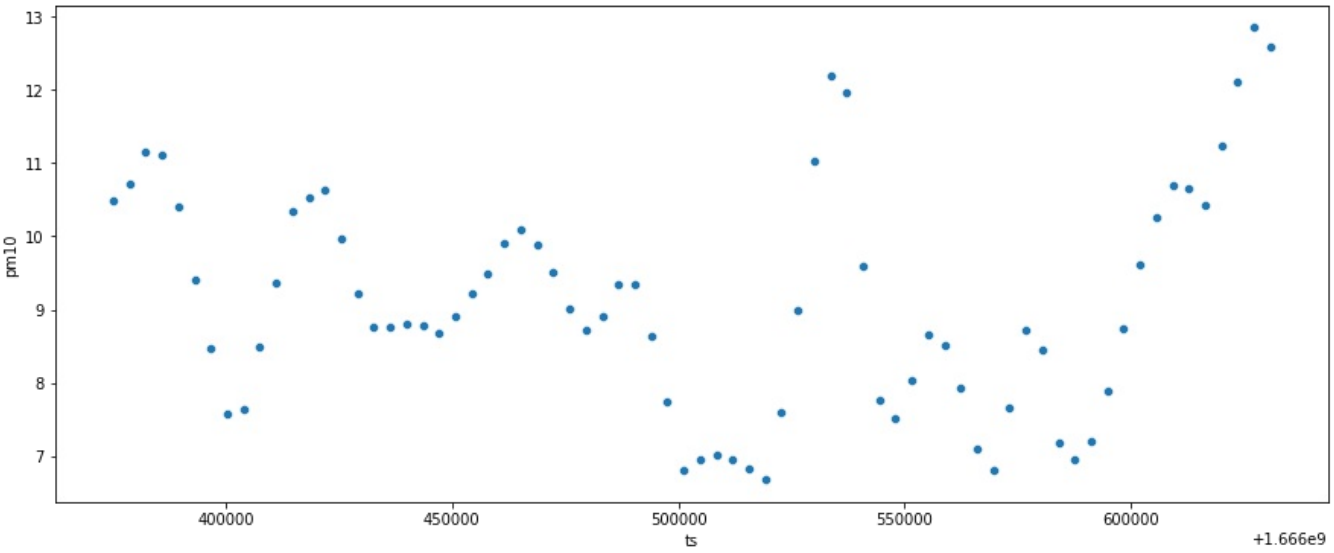

pm10
CO
13
12
11
10
9
8
7
192.5
195.0
197.5
200.0
202.5
205.0
207.5

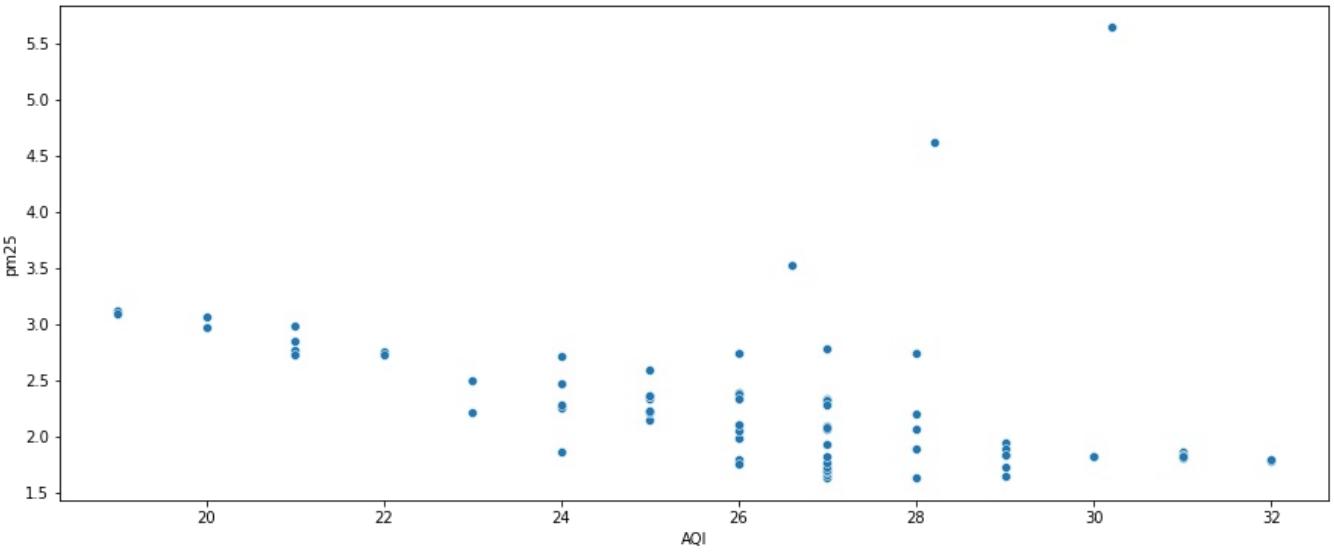

pm10
no2
13
12
11
10
9
8
7
0.02
0.03
0.04
0.05
0.06

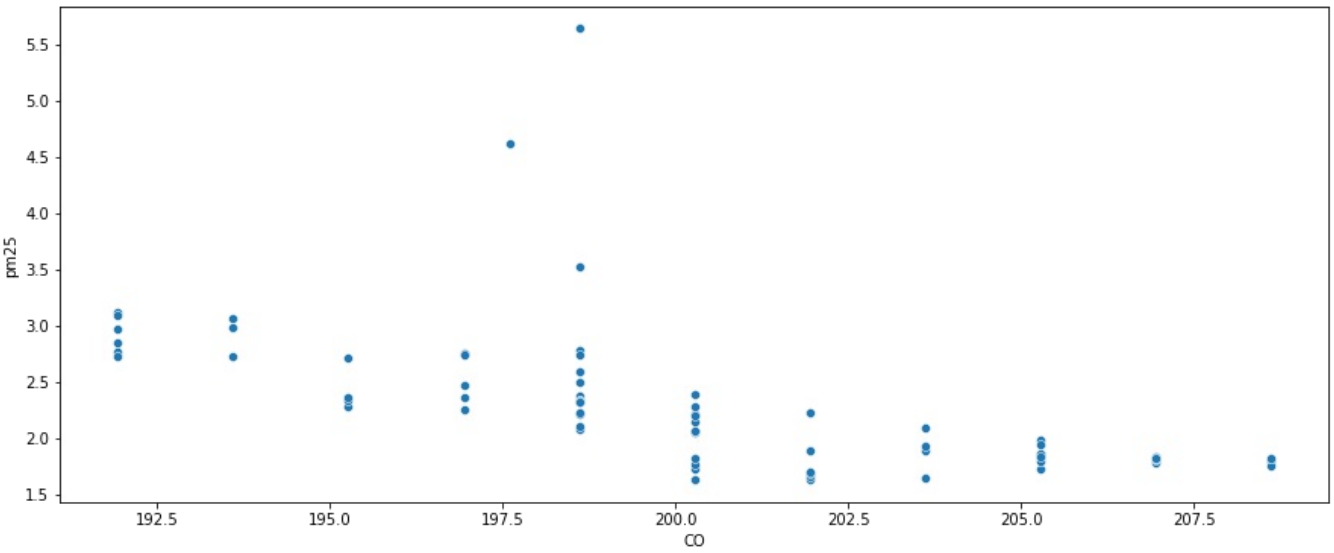

pm10
o3
13
12
11
10
9
8
7
45
50
55
60
65
70

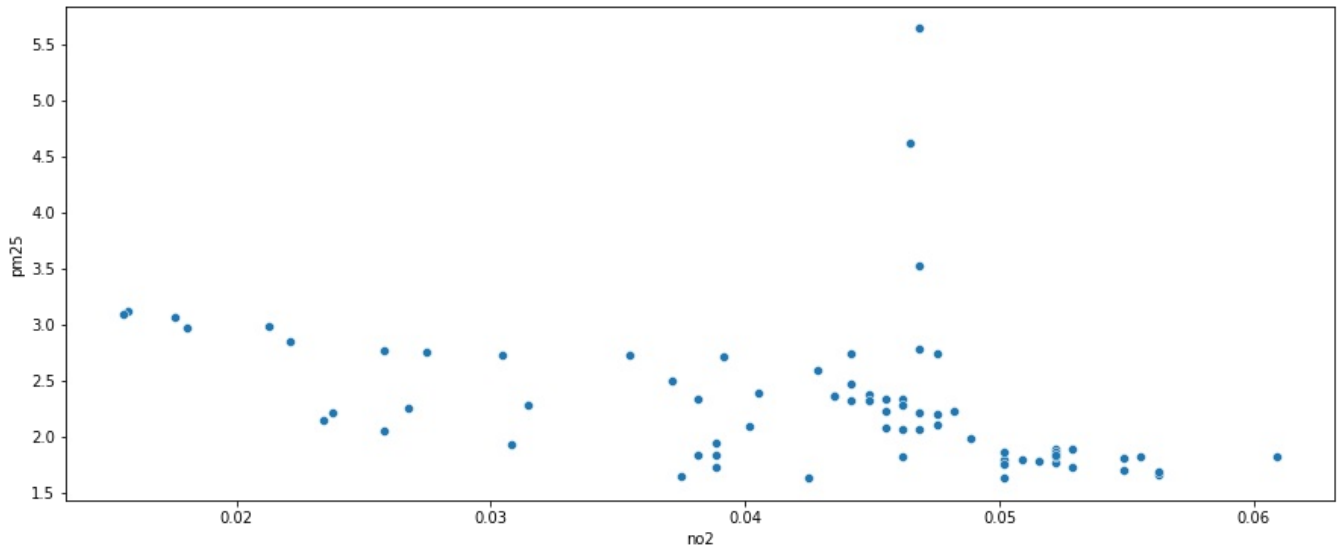

pm10
pm25
13
12
11
10
9
8
7
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5

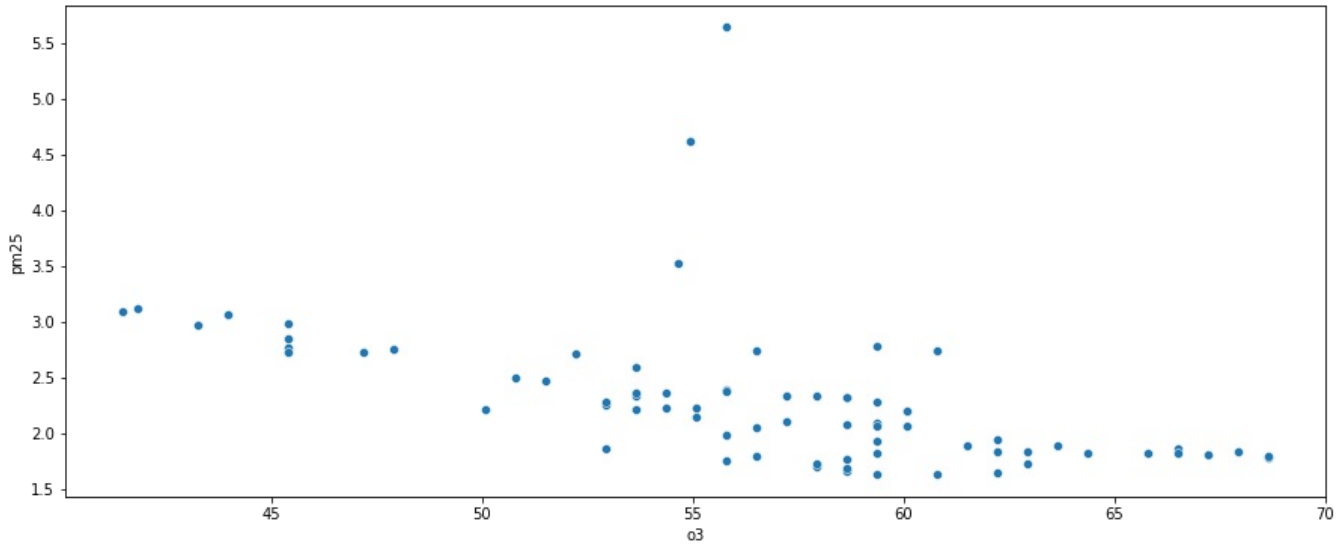

pm10
so2
13
12
11
10
9
8
7
0.275
0.300
0.325
0.350
0.375
0.400
0.425
0.450

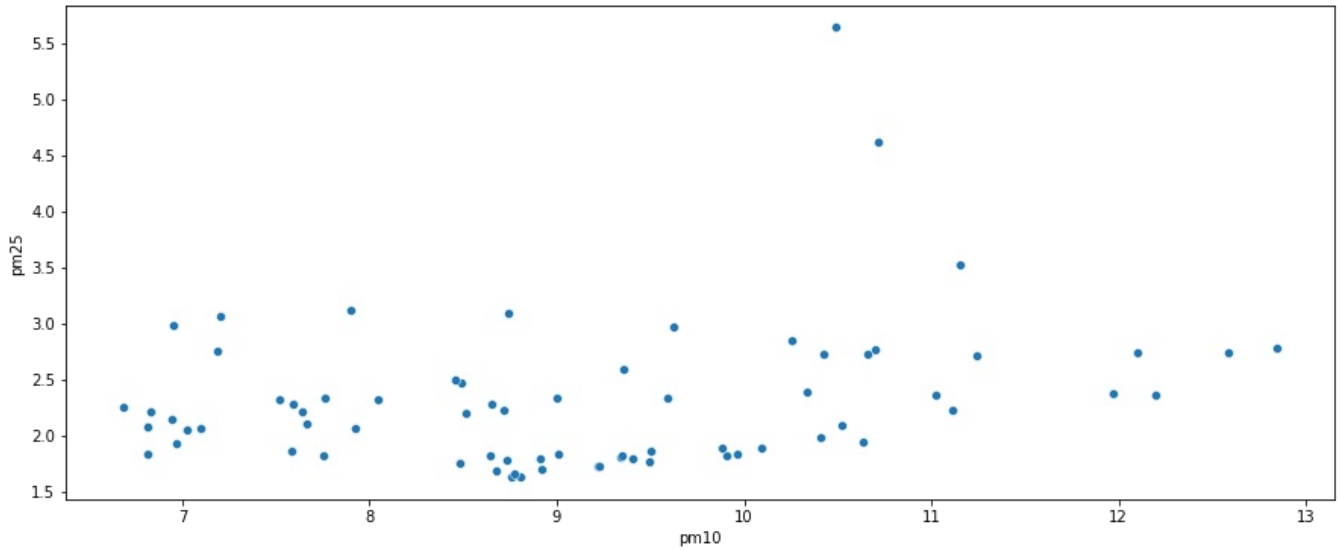

pm10
ts
+1.666e9

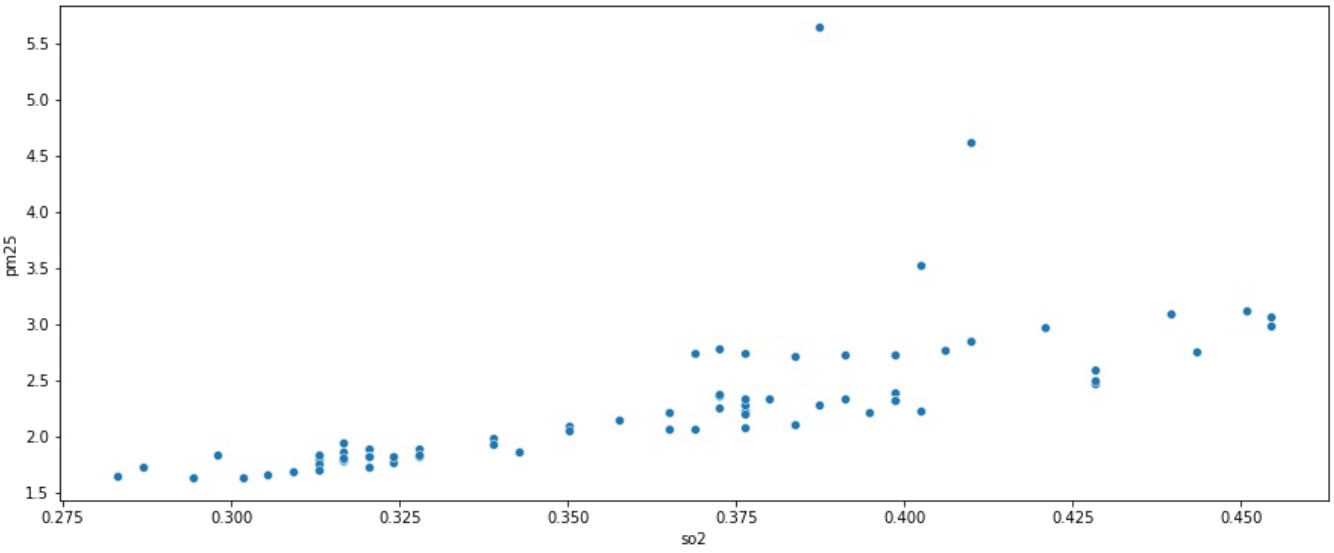

pm25
AQI

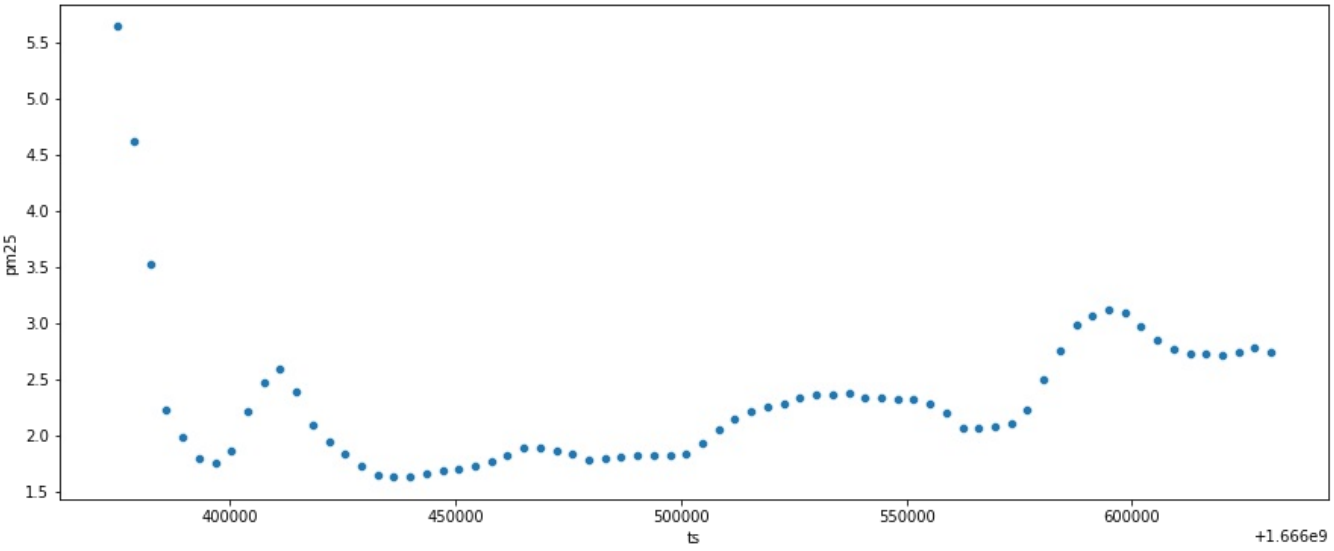

pm25
CO

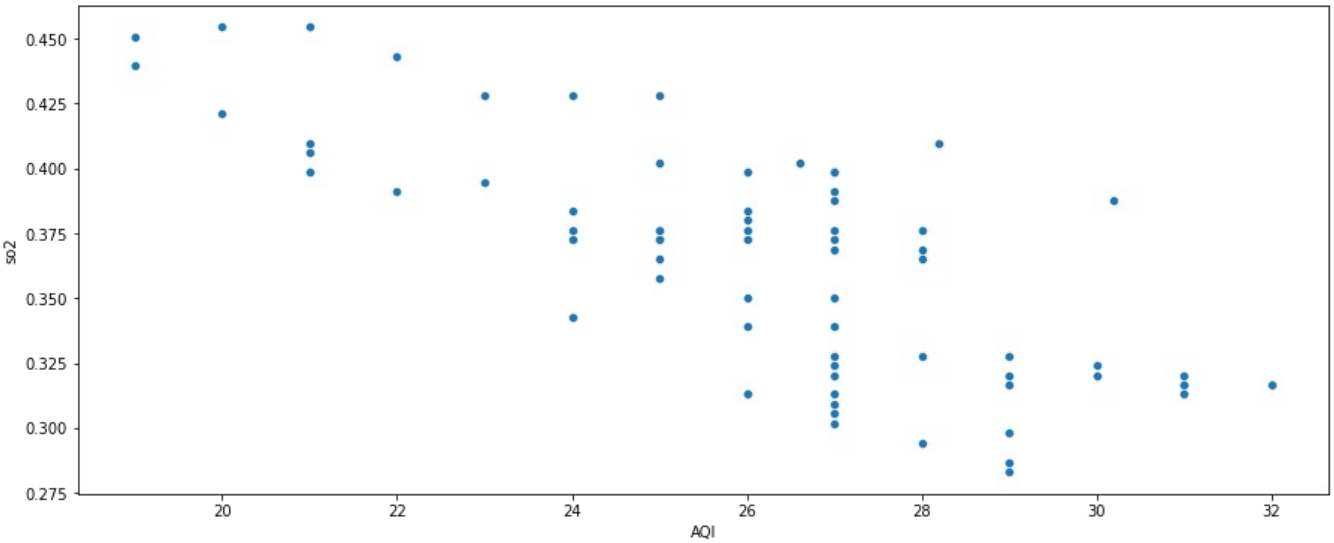

pm25
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0.02
0.03
0.04
0.05
0.06
no2

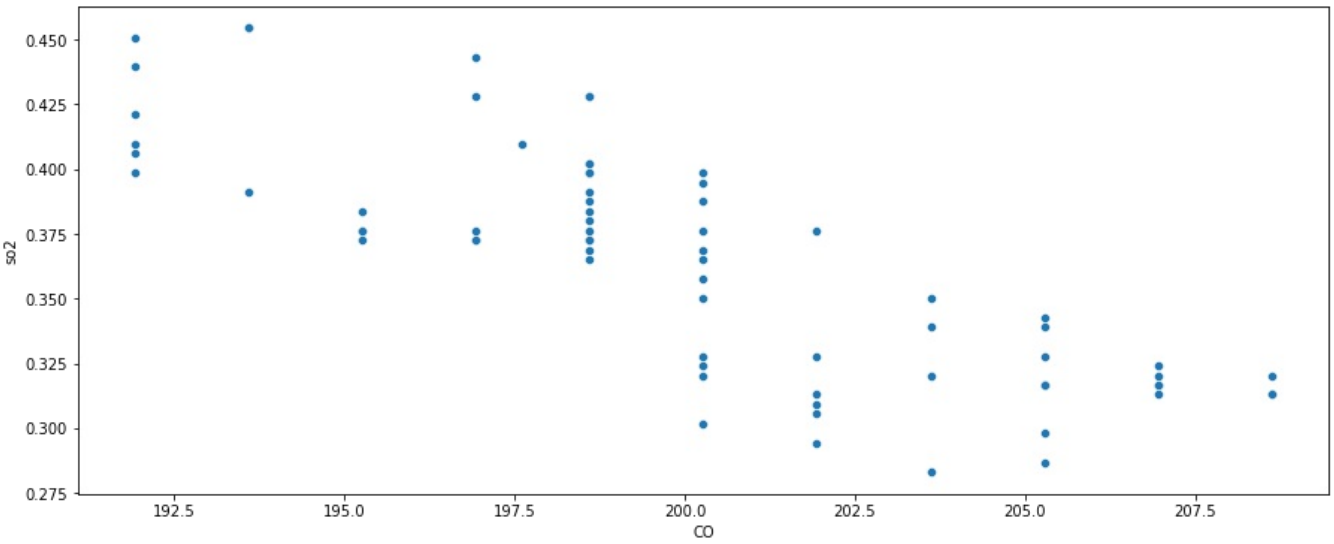

pm25
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
45
50
55
60
65
70
o3

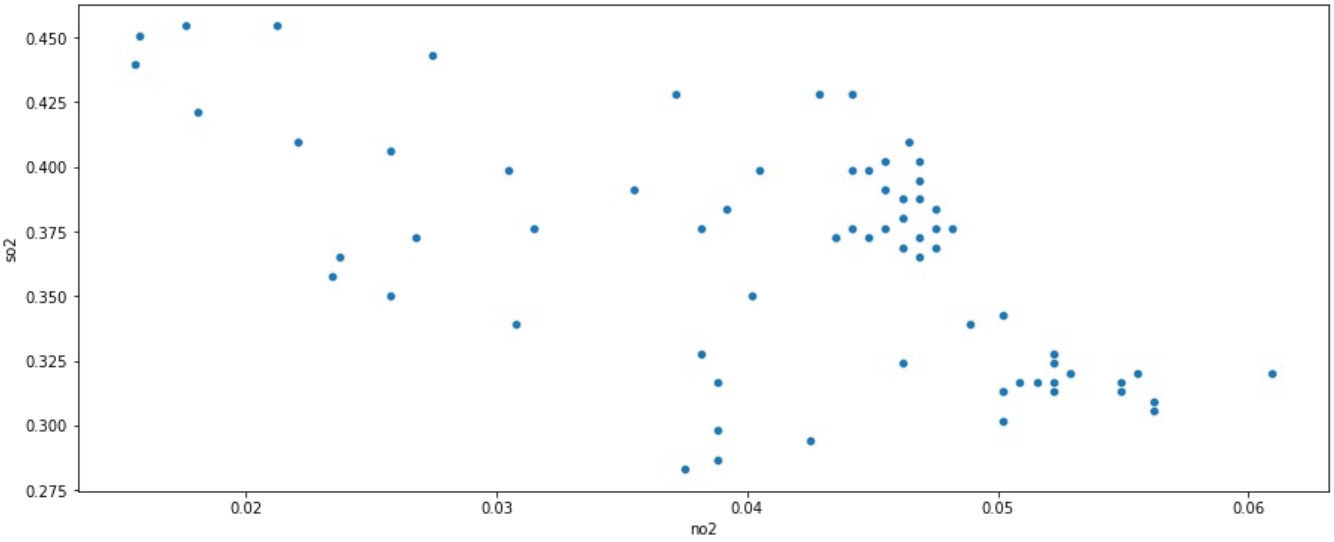

pm25
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
7
8
9
10
11
12
13
pm10

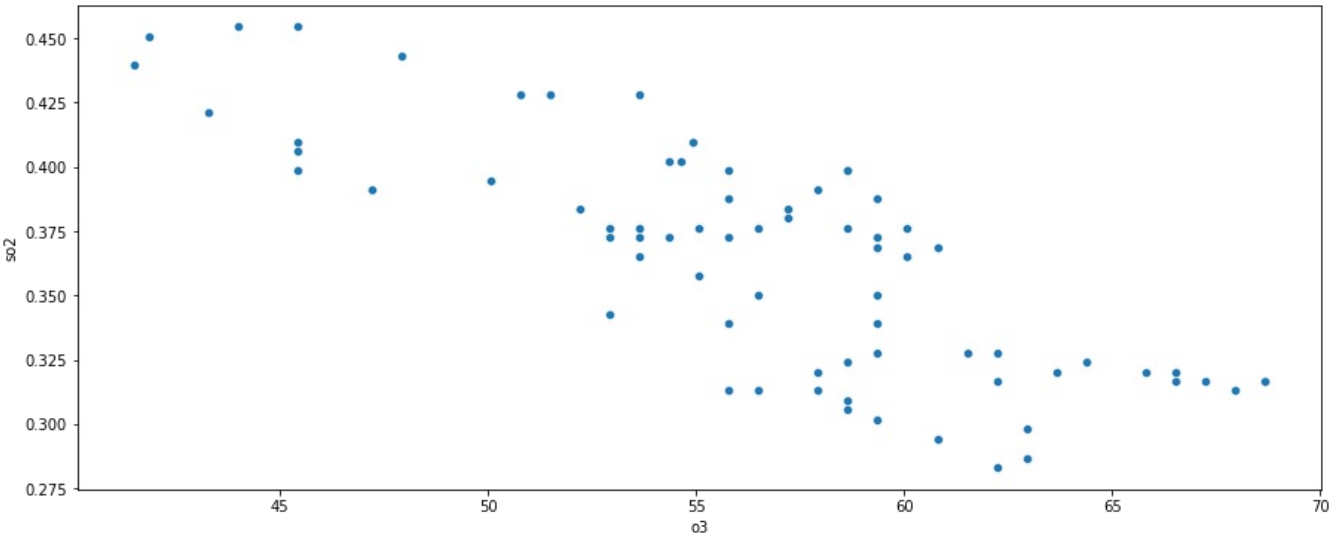

pm25
so2
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0.275
0.300
0.325
0.350
0.375
0.400
0.425
0.450

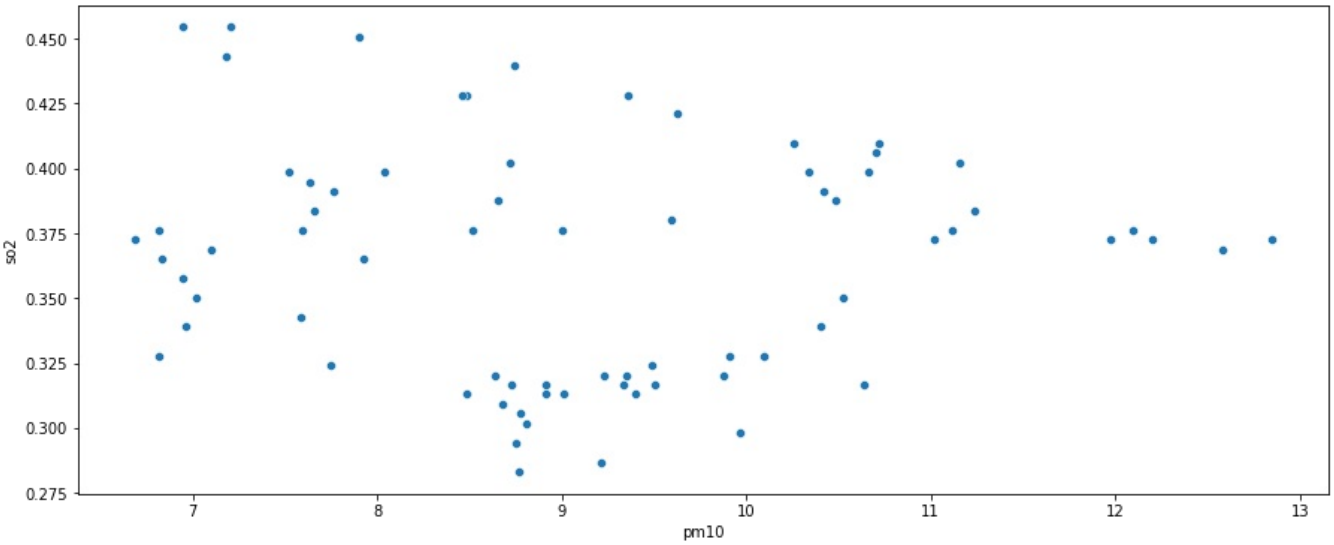

pm25
ts
+1.666e9
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
400000
450000
500000
550000
600000

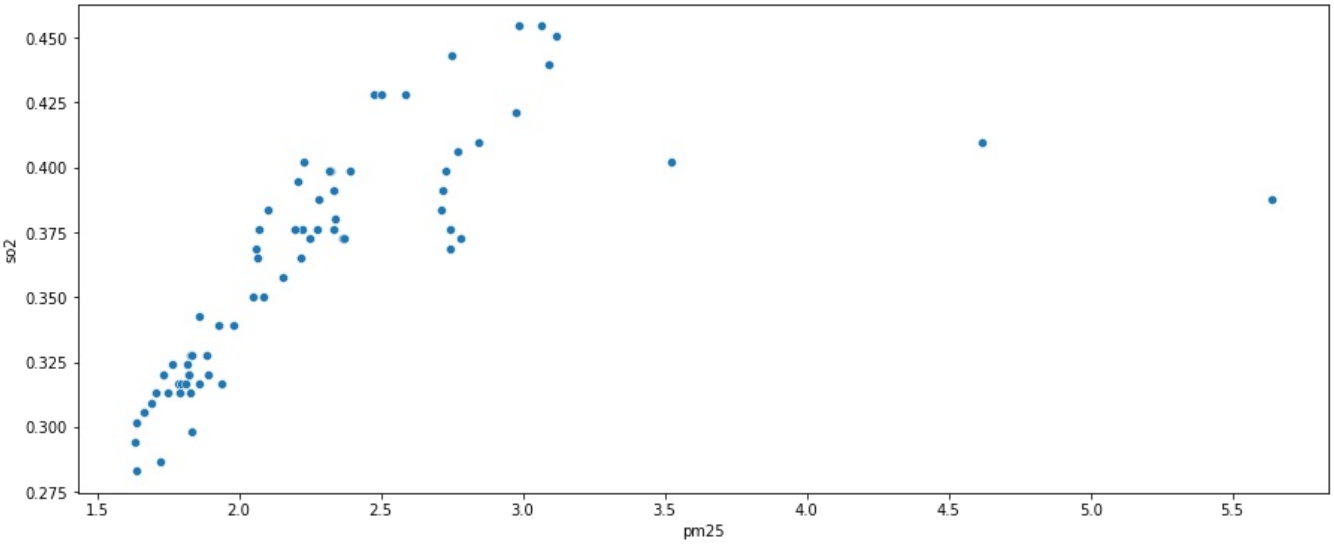

so2
AQI
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
20
22
24
26
28
30
32

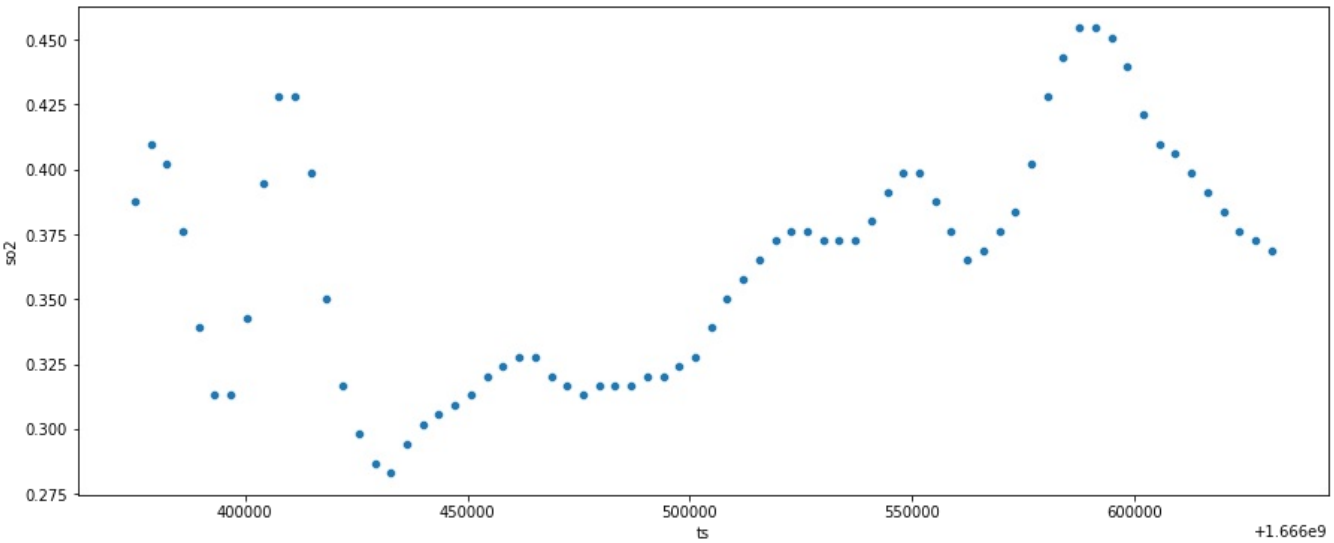

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
192.5
195.0
197.5
200.0
202.5
205.0
207.5
CO

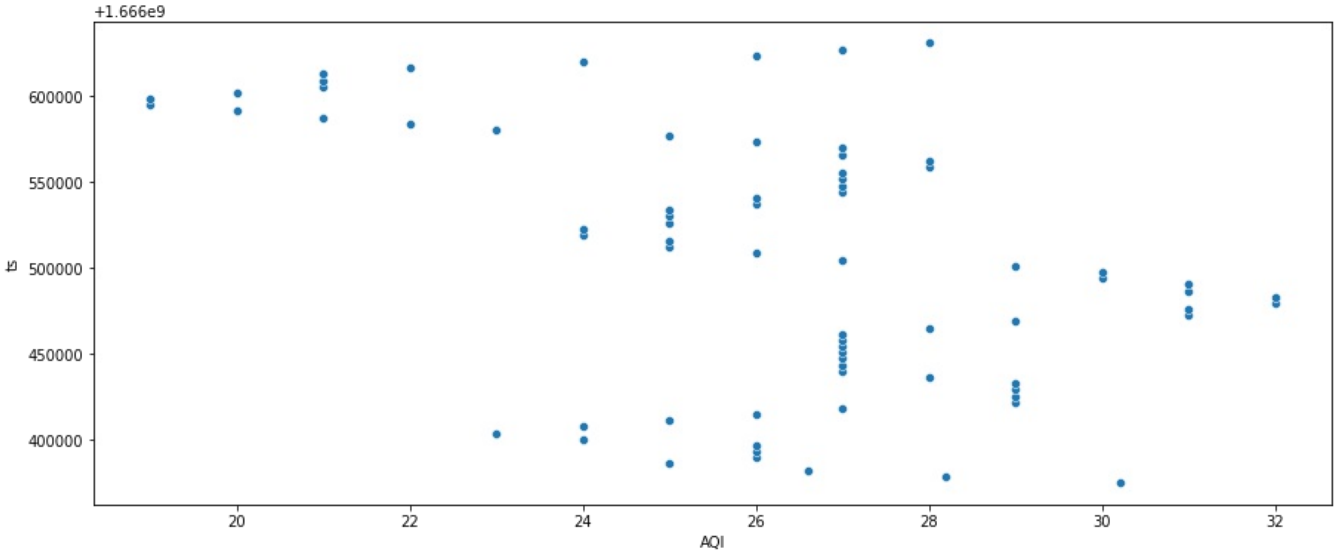

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
0.02
0.03
0.04
0.05
0.06
no2

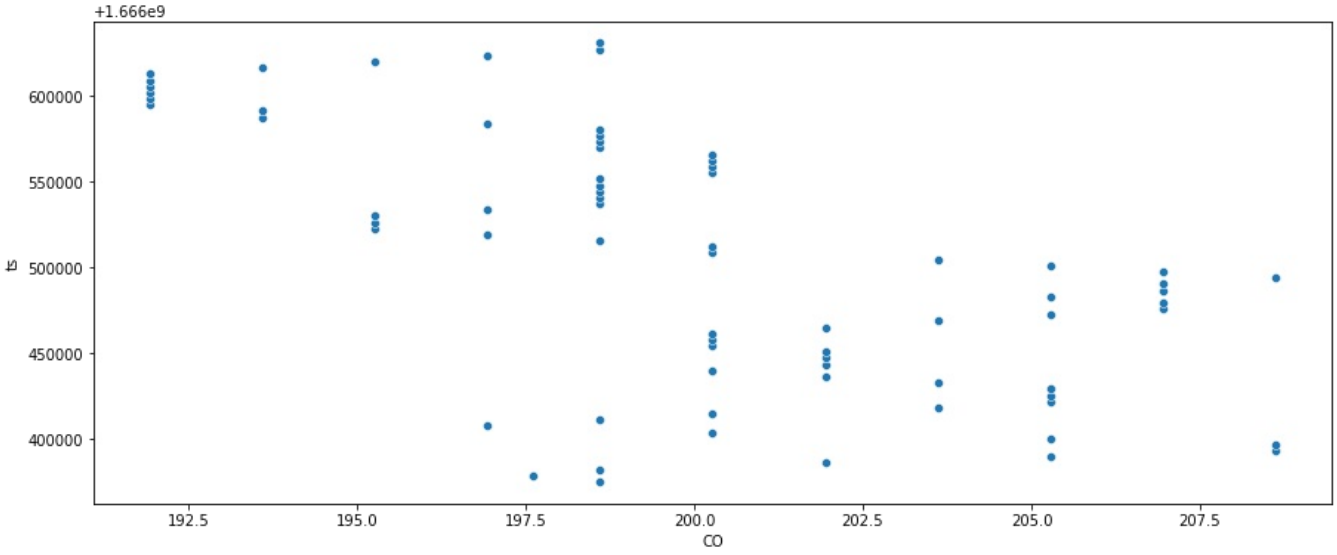

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
45
50
55
60
65
70
o3

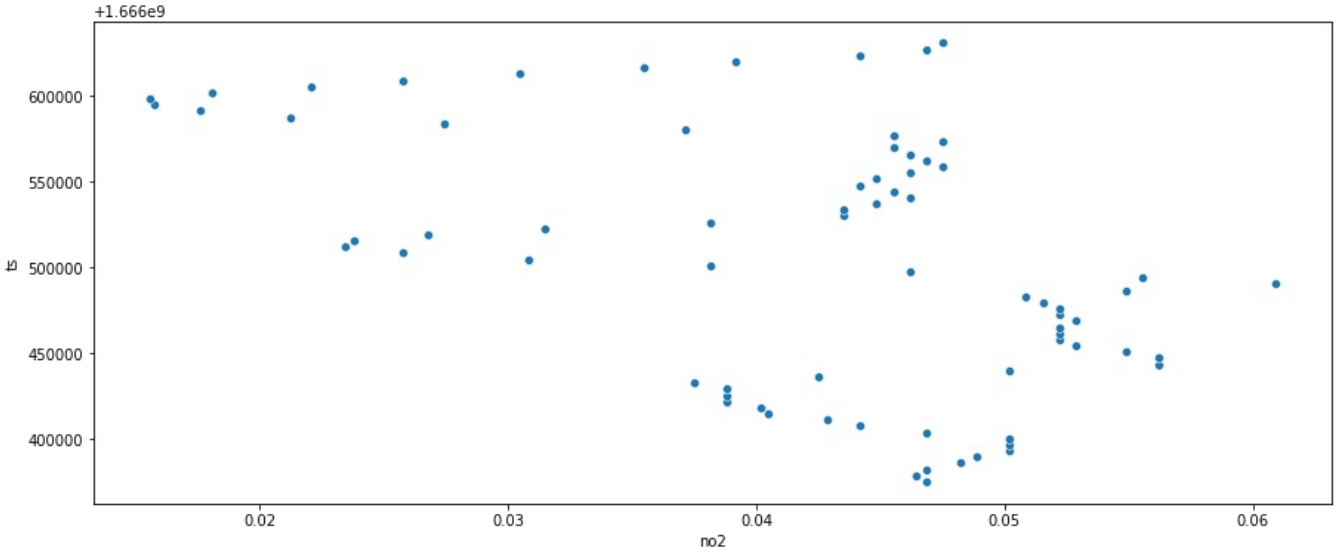

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
7
8
9
10
11
12
13
pm10

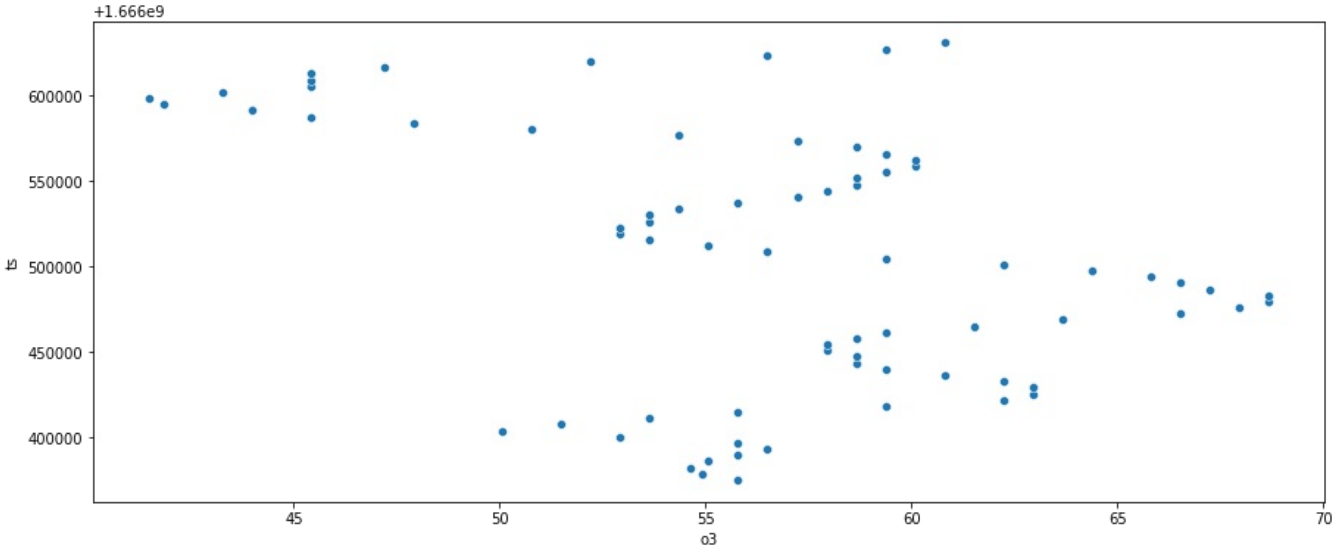

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
pm25

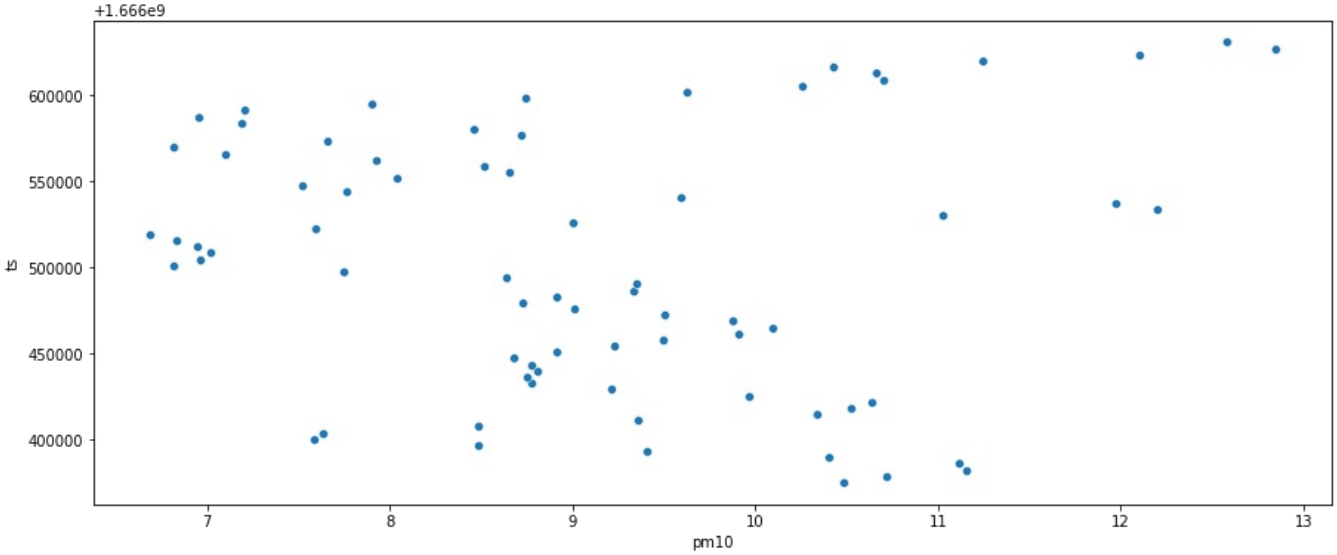

so2
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
400000
450000
500000
550000
600000
ts
+1.666e9

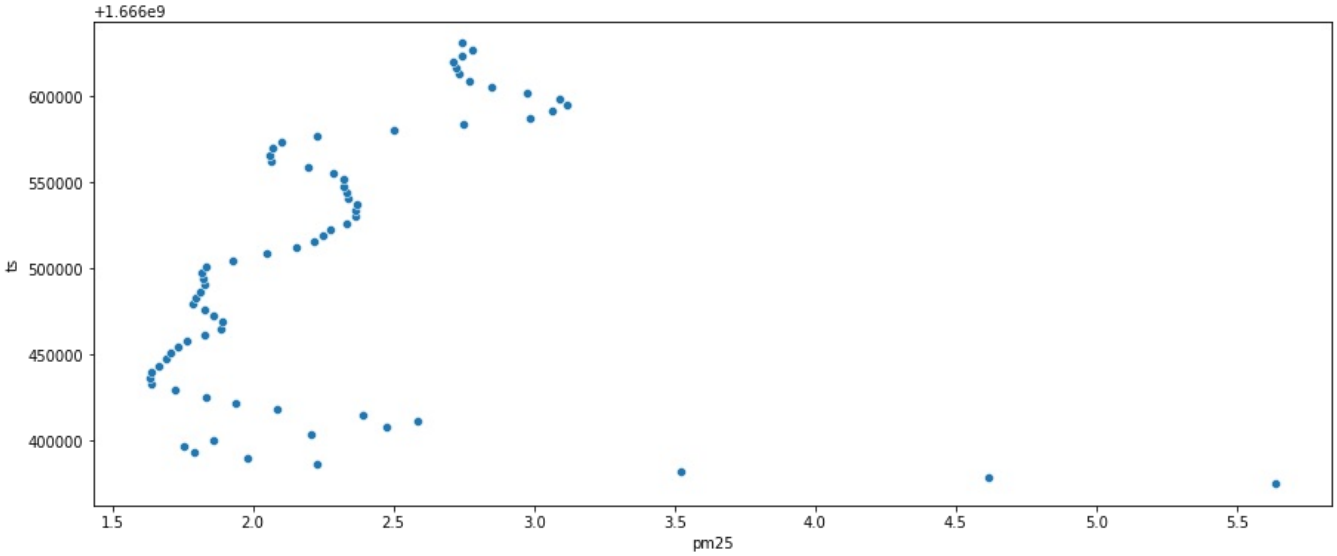

+1.666e9
600000
550000
500000
450000
400000
t5
20
22
24
26
28
30
32
AQI

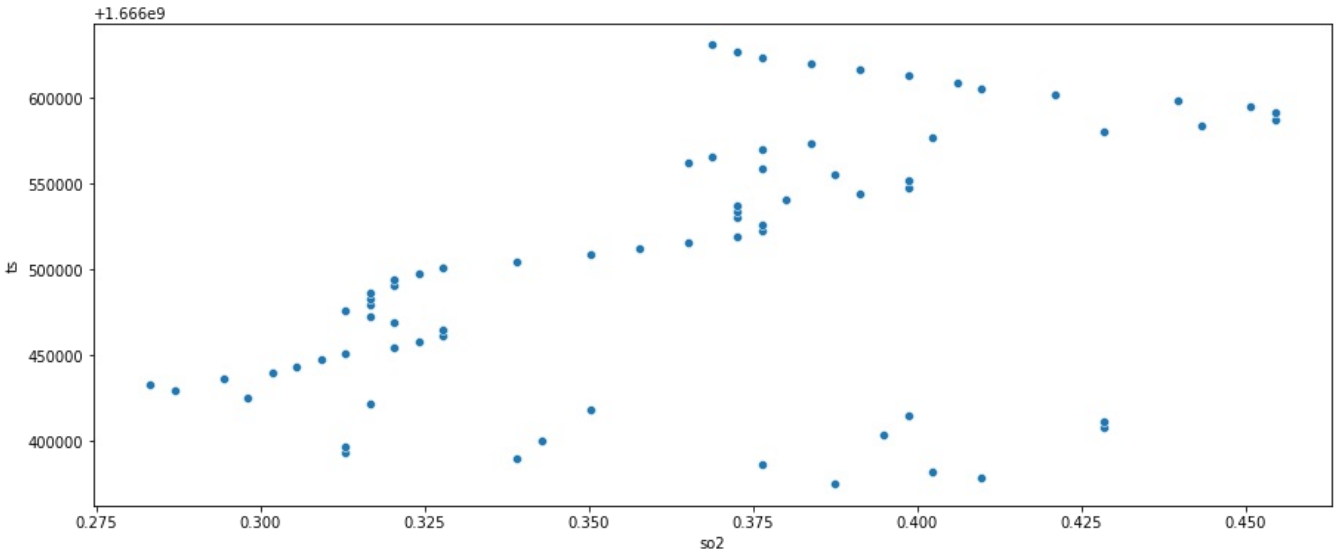

+1.666e9
600000
550000
500000
450000
400000
t5
192.5
195.0
197.5
200.0
202.5
205.0
207.5
CO


+1.666e9
600000
550000
500000
450000
400000
t5
0.02
0.03
0.04
0.05
0.06
no2


+1.666e9
600000
550000
500000
450000
400000
45
50
55
60
65
70
o3


+1.666e9
600000
550000
500000
450000
400000
7
8
9
10
11
12
13
pm10


+1.666e9
600000
550000
500000
450000
400000
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
pm25



In [35]:
```
df_new = df[numerical_columns]
```

In [36]:
```
df_new
```

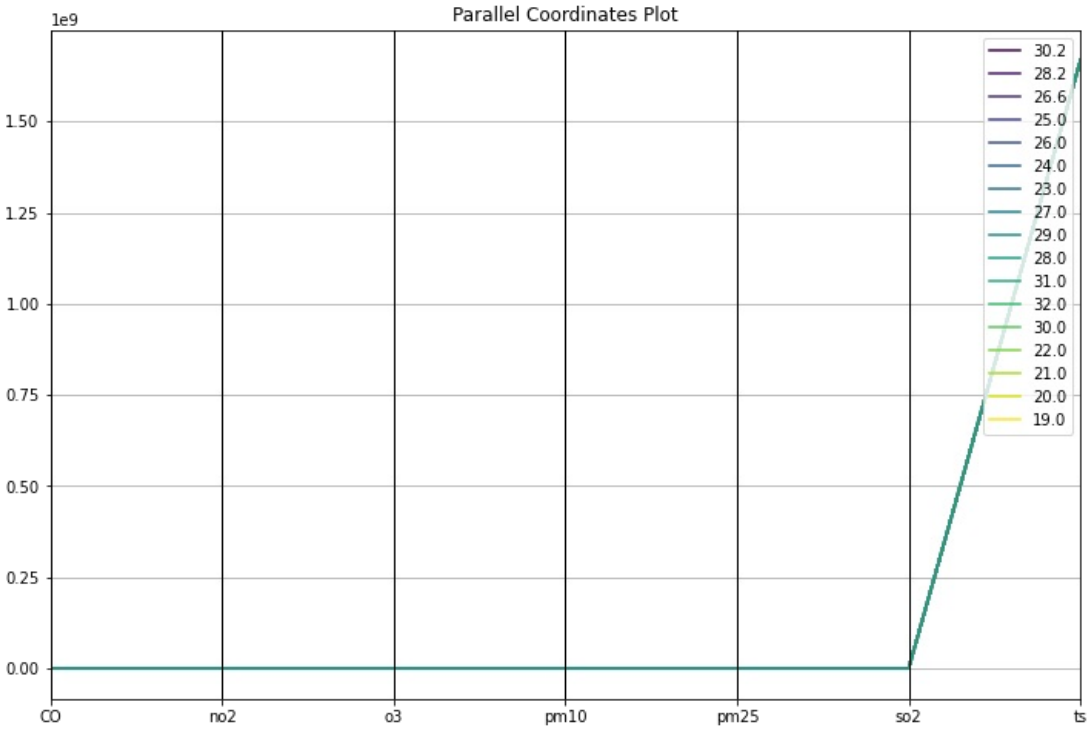
Out[36]:

datetime	AQI	CO	no2	o3	pm10	pm25	so2	ts
2022-10-21 18:00:00	30.2	198.60268	0.046857	55.789948	10.486722	5.637410	0.387430	1666375200
2022-10-21 19:00:00	28.2	197.60132	0.046456	54.931640	10.719325	4.618169	0.409782	1666378800
2022-10-21 20:00:00	26.6	198.60268	0.046857	54.645540	11.155578	3.520902	0.402331	1666382400
2022-10-21 21:00:00	25.0	201.94054	0.048196	55.074690	11.116206	2.225919	0.376254	1666386000
2022-10-21 22:00:00	26.0	205.27840	0.048865	55.789948	10.405250	1.979471	0.339001	1666389600
...	...	...	...	...	...	...	...	...
2022-10-24 13:00:00	22.0	193.59589	0.035478	47.206880	10.423121	2.720472	0.391155	1666616400
2022-10-24 14:00:00	24.0	195.26482	0.039159	52.213670	11.240661	2.713109	0.383705	1666620000
2022-10-24 15:00:00	26.0	196.93375	0.044180	56.505203	12.098125	2.743044	0.376254	1666623600
2022-10-24 16:00:00	27.0	198.60268	0.046857	59.366226	12.845977	2.780022	0.372529	1666627200
2022-10-24 17:00:00	28.0	198.60268	0.047527	60.796738	12.583639	2.745875	0.368804	1666630800

72 rows × 8 columns

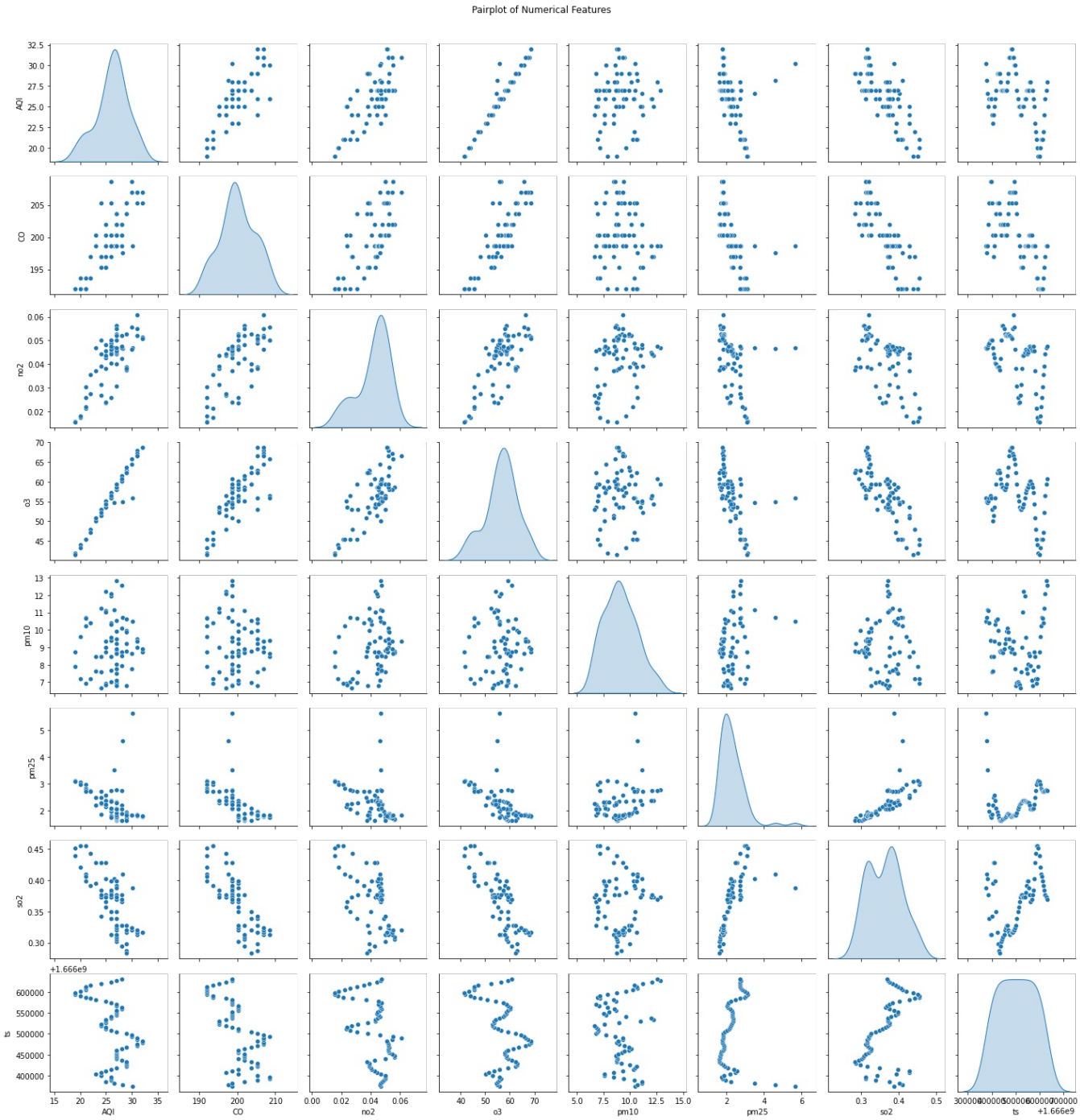
In [37]:
```
from pandas.plotting import parallel_coordinates
```

In [38]:
```
plt.figure(figsize=(12, 8))
parallel_coordinates(df_new, 'AQI', colormap='viridis')
plt.title('Parallel Coordinates Plot')
plt.show()
```



In [39]:
```
plt.figure(figsize=(12, 8))
sns.pairplot(df_new, diag_kind='kde')
plt.suptitle('Pairplot of Numerical Features', y=1.02)
plt.show()
```

```
<Figure size 864x576 with 0 Axes>
```



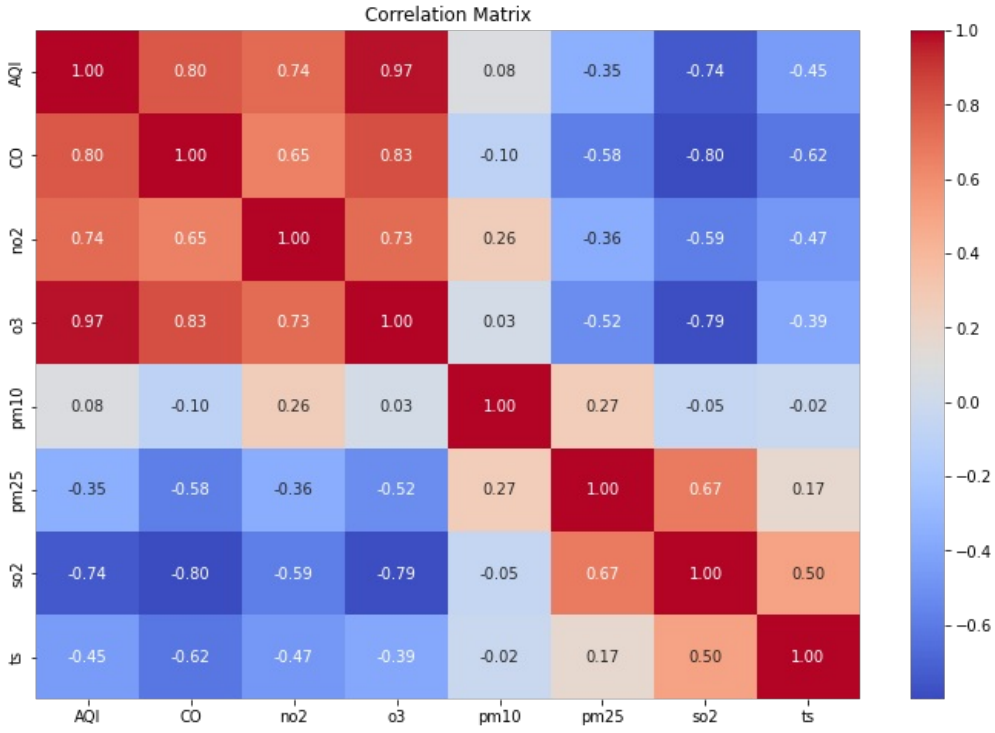
In [40]:
```
correlation_matrix = df_new.corr()
```

In [41]:
```
correlation_matrix
```

Out[41]:

	AQI	CO	no2	o3	pm10	pm25	so2	ts
AQI	1.000000	0.797633	0.737655	0.973599	0.082576	-0.346655	-0.742478	-0.453870
CO	0.797633	1.000000	0.652023	0.829198	-0.095986	-0.582915	-0.798720	-0.620788
no2	0.737655	0.652023	1.000000	0.732042	0.256524	-0.358809	-0.585913	-0.471094
o3	0.973599	0.829198	0.732042	1.000000	0.033811	-0.521978	-0.785625	-0.387106
pm10	0.082576	-0.095986	0.256524	0.033811	1.000000	0.268483	-0.047794	-0.019671
pm25	-0.346655	-0.582915	-0.358809	-0.521978	0.268483	1.000000	0.674343	0.165993
so2	-0.742478	-0.798720	-0.585913	-0.785625	-0.047794	0.674343	1.000000	0.504122
ts	-0.453870	-0.620788	-0.471094	-0.387106	-0.019671	0.165993	0.504122	1.000000

In [42]:
```
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix')
plt.show()
```



Thanks !!!