In [1]:

**import** **pandas** **as** **pd**  
**import** **numpy** **as** **np**  
**import** **matplotlib.pyplot** **as** **plt**

In [2]:

**import** **glob**

In [3]:

**import** **os**

In [4]:

os.chdir('C:/Users/JAEMIN/Desktop/air')

In [5]:

a = glob.glob('\*.\*')  
a

Out[5]:

['co.xlsx', 'dust.xlsx', 'no2.xlsx', 'o3.xlsx', 'so2.xlsx']

In [6]:

co\_1 = pd.read\_excel(a[0])  
dust\_1 = pd.read\_excel(a[1])  
no2\_1 = pd.read\_excel(a[2])  
o3\_1 = pd.read\_excel(a[3])  
so2\_1 = pd.read\_excel(a[4])

In [7]:

co\_1 = co\_1.drop(['시도별(2)'], axis =1)  
dust\_1 = dust\_1.drop(['시도별(2)'], axis =1)  
no2\_1 = no2\_1.drop(['시도별(2)'], axis =1)  
o3\_1 = o3\_1.drop(['시도별(2)'], axis =1)  
so2\_1 = so2\_1.drop(['시도별(2)'], axis =1)

In [8]:

co\_1 = co\_1.rename(columns = {'시도별(1)' : '지역'})  
dust\_1 = dust\_1.rename(columns = {'시도별(1)' : '지역'})  
no2\_1 = no2\_1.rename(columns = {'시도별(1)' : '지역'})  
o3\_1 = o3\_1.rename(columns = {'시도별(1)' : '지역'})  
so2\_1 = so2\_1.rename(columns = {'시도별(1)' : '지역'})

In [9]:

**for** i **in** range(7, 10):  
 co\_1['2019. 0**{}**'.format(i)] = 'NaN'  
 dust\_1['2019. 0**{}**'.format(i)] = 'NaN'  
 no2\_1['2019. 0**{}**'.format(i)] = 'NaN'  
 o3\_1['2019. 0**{}**'.format(i)] = 'NaN'  
 so2\_1['2019. 0**{}**'.format(i)] = 'NaN'  
**for** i **in** range(10, 13):  
 co\_1['2019. **{}**'.format(i)] = 'NaN'  
 dust\_1['2019. **{}**'.format(i)] = 'NaN'  
 no2\_1['2019. **{}**'.format(i)] = 'NaN'  
 o3\_1['2019. **{}**'.format(i)] = 'NaN'  
 so2\_1['2019. **{}**'.format(i)] = 'NaN'

In [10]:

var\_list = [co\_1, dust\_1, no2\_1, o3\_1, so2\_1]  
region\_list = ['계', '부산광역시','대구광역시', '인천광역시',  
 '광주광역시' ,'대전광역시','울산광역시', '경기도',   
 '강원도', '충청북도', '충청남도','전라북도',   
 '전라남도', '경상북도', '경상남도', '제주도']  
**for** x **in** var\_list:  
 i = 0  
 x.drop([1, 8], inplace = **True**)  
 x.reset\_index(drop = **True**, inplace = **True**)  
 **for** y **in** region\_list:  
 x['지역'][i] = y  
 i+=1

C:\Users\JAEMIN\Anaconda3\lib\site-packages\ipykernel\_launcher.py:11: SettingWithCopyWarning:   
A value is trying to be set on a copy of a slice from a DataFrame  
  
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
 # This is added back by InteractiveShellApp.init\_path()

In [11]:

co\_1

Out[11]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **지역** | **2010. 01** | **2010. 02** | **2010. 03** | **2010. 04** | **2010. 05** | **2010. 06** | **2010. 07** | **2010. 08** | **2010. 09** | **...** | **2019. 03** | **2019. 04** | **2019. 05** | **2019. 06** | **2019. 07** | **2019. 08** | **2019. 09** | **2019. 10** | **2019. 11** | **2019. 12** |
| 0 | 계 | 0.7 | 0.6 | 0.5 | 0.5 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | ... | 0.5 | 0.4 | 0.4 | 0.4 | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | 부산광역시 | 0.5 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.3 | 0.3 | 0.3 | ... | 0.4 | 0.4 | 0.4 | 0.4 | NaN | NaN | NaN | NaN | NaN | NaN |
| 2 | 대구광역시 | 0.7 | 0.6 | 0.5 | 0.5 | 0.4 | 0.5 | 0.4 | 0.4 | 0.4 | ... | 0.5 | 0.4 | 0.4 | 0.4 | NaN | NaN | NaN | NaN | NaN | NaN |
| 3 | 인천광역시 | 0.9 | 0.7 | 0.5 | 0.5 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | ... | 0.6 | 0.5 | 0.5 | 0.4 | NaN | NaN | NaN | NaN | NaN | NaN |
| 4 | 광주광역시 | 0.7 | 0.6 | 0.5 | 0.5 | 0.5 | 0.4 | 0.4 | 0.3 | 0.4 | ... | 0.5 | 0.4 | 0.4 | 0.4 | NaN | NaN | NaN | NaN | NaN | NaN |
| 5 | 대전광역시 | 0.8 | 0.7 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | ... | 0.5 | 0.3 | 0.4 | 0.3 | NaN | NaN | NaN | NaN | NaN | NaN |
| 6 | 울산광역시 | 0.6 | 0.6 | 0.4 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | ... | 0.5 | 0.5 | 0.5 | 0.5 | NaN | NaN | NaN | NaN | NaN | NaN |
| 7 | 경기도 | 0.9 | 0.7 | 0.6 | 0.5 | 0.5 | 0.5 | 0.4 | 0.4 | 0.4 | ... | 0.6 | 0.4 | 0.5 | 0.4 | NaN | NaN | NaN | NaN | NaN | NaN |
| 8 | 강원도 | 0.9 | 0.7 | 0.6 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | ... | 0.5 | 0.4 | 0.4 | 0.3 | NaN | NaN | NaN | NaN | NaN | NaN |
| 9 | 충청북도 | 1.1 | 0.9 | 0.6 | 0.6 | 0.4 | 0.4 | 0.3 | 0.3 | 0.3 | ... | 0.5 | 0.4 | 0.4 | 0.3 | NaN | NaN | NaN | NaN | NaN | NaN |
| 10 | 충청남도 | 0.7 | 0.6 | 0.5 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.5 | ... | 0.6 | 0.5 | 0.5 | 0.5 | NaN | NaN | NaN | NaN | NaN | NaN |
| 11 | 전라북도 | 0.6 | 0.6 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | ... | 0.5 | 0.4 | 0.4 | 0.3 | NaN | NaN | NaN | NaN | NaN | NaN |
| 12 | 전라남도 | 0.6 | 0.6 | 0.5 | 0.5 | 0.4 | 0.4 | 0.3 | 0.4 | 0.4 | ... | 0.5 | 0.5 | 0.5 | 0.5 | NaN | NaN | NaN | NaN | NaN | NaN |
| 13 | 경상북도 | 0.7 | 0.6 | 0.5 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.3 | ... | 0.4 | 0.4 | 0.4 | 0.4 | NaN | NaN | NaN | NaN | NaN | NaN |
| 14 | 경상남도 | 0.6 | 0.6 | 0.4 | 0.4 | 0.4 | 0.3 | 0.3 | 0.3 | 0.4 | ... | 0.5 | 0.4 | 0.4 | 0.4 | NaN | NaN | NaN | NaN | NaN | NaN |
| 15 | 제주도 | 0.3 | 0.4 | 0.4 | 0.4 | 0.4 | 0.3 | 0.3 | 0.3 | 0.2 | ... | 0.3 | 0.3 | 0.3 | 0.2 | NaN | NaN | NaN | NaN | NaN | NaN |

16 rows × 121 columns

In [12]:

pol\_list = ['일산화탄소', '미세먼지PM10', '이산화질소', '오존', '아황산가스']

In [13]:

*'''*  
*x = 2010*  
*co = co\_1.iloc[:1, 1:13].rename(columns={'{}. 01'.format(x) : '1월 {}'.format(pol\_list[0]), '{}. 02'.format(x) : '2월 {}'.format(pol\_list[0]),*  
 *'{}. 03'.format(x) : '3월 {}'.format(pol\_list[0]), '{}. 04'.format(x) : '4월 {}'.format(pol\_list[0]),*   
 *'{}. 05'.format(x) : '5월 {}'.format(pol\_list[0]), '{}. 06'.format(x) : '6월 {}'.format(pol\_list[0]),*  
 *'{}. 07'.format(x) : '7월 {}'.format(pol\_list[0]), '{}. 08'.format(x) : '8월 {}'.format(pol\_list[0]),*  
 *'{}. 09'.format(x) : '9월 {}'.format(pol\_list[0]), '{}. 10'.format(x) : '10월 {}'.format(pol\_list[0]),*  
 *'{}. 11'.format(x) : '11월 {}'.format(pol\_list[0]), '{}. 12'.format(x) : '12월 {}'.format(pol\_list[0])})*  
*co*  
*'''*

Out[13]:

"\nx = 2010\nco = co\_1.iloc[:1, 1:13].rename(columns={'{}. 01'.format(x) : '1월 {}'.format(pol\_list[0]), '{}. 02'.format(x) : '2월 {}'.format(pol\_list[0]),\n '{}. 03'.format(x) : '3월 {}'.format(pol\_list[0]), '{}. 04'.format(x) : '4월 {}'.format(pol\_list[0]), \n '{}. 05'.format(x) : '5월 {}'.format(pol\_list[0]), '{}. 06'.format(x) : '6월 {}'.format(pol\_list[0]),\n '{}. 07'.format(x) : '7월 {}'.format(pol\_list[0]), '{}. 08'.format(x) : '8월 {}'.format(pol\_list[0]),\n '{}. 09'.format(x) : '9월 {}'.format(pol\_list[0]), '{}. 10'.format(x) : '10월 {}'.format(pol\_list[0]),\n '{}. 11'.format(x) : '11월 {}'.format(pol\_list[0]), '{}. 12'.format(x) : '12월 {}'.format(pol\_list[0])})\nco\n"

In [14]:

co = pd.DataFrame()  
dust = pd.DataFrame()  
no2 = pd.DataFrame()  
o3 = pd.DataFrame()  
so2 = pd.DataFrame()

In [15]:

**for** i **in** range(0, 17):  
 x = 2010  
 **for** j **in** range(1, 121, 12):  
 *# 일산화탄소*  
 tmp\_c = co\_1.iloc[i:i+1, j:j+12].rename(columns={'**{}**. 01'.format(x) : '1월 **{}**'.format(pol\_list[0]), '**{}**. 02'.format(x) : '2월 **{}**'.format(pol\_list[0]),  
 '**{}**. 03'.format(x) : '3월 **{}**'.format(pol\_list[0]), '**{}**. 04'.format(x) : '4월 **{}**'.format(pol\_list[0]),  
 '**{}**. 05'.format(x) : '5월 **{}**'.format(pol\_list[0]), '**{}**. 06'.format(x) : '6월 **{}**'.format(pol\_list[0]),  
 '**{}**. 07'.format(x) : '7월 **{}**'.format(pol\_list[0]), '**{}**. 08'.format(x) : '8월 **{}**'.format(pol\_list[0]),  
 '**{}**. 09'.format(x) : '9월 **{}**'.format(pol\_list[0]), '**{}**. 10'.format(x) : '10월 **{}**'.format(pol\_list[0]),  
 '**{}**. 11'.format(x) : '11월 **{}**'.format(pol\_list[0]), '**{}**. 12'.format(x) : '12월 **{}**'.format(pol\_list[0])})  
 co = pd.concat([co,tmp\_c], axis = 0, sort = **False**)  
 *# 미세먼지*  
 tmp\_d = dust\_1.iloc[i:i+1, j:j+12].rename(columns={'**{}**. 01'.format(x) : '1월 **{}**'.format(pol\_list[1]), '**{}**. 02'.format(x) : '2월 **{}**'.format(pol\_list[1]),  
 '**{}**. 03'.format(x) : '3월 **{}**'.format(pol\_list[1]), '**{}**. 04'.format(x) : '4월 **{}**'.format(pol\_list[1]),  
 '**{}**. 05'.format(x) : '5월 **{}**'.format(pol\_list[1]), '**{}**. 06'.format(x) : '6월 **{}**'.format(pol\_list[1]),  
 '**{}**. 07'.format(x) : '7월 **{}**'.format(pol\_list[1]), '**{}**. 08'.format(x) : '8월 **{}**'.format(pol\_list[1]),  
 '**{}**. 09'.format(x) : '9월 **{}**'.format(pol\_list[1]), '**{}**. 10'.format(x) : '10월 **{}**'.format(pol\_list[1]),  
 '**{}**. 11'.format(x) : '11월 **{}**'.format(pol\_list[1]), '**{}**. 12'.format(x) : '12월 **{}**'.format(pol\_list[1])})  
 dust = pd.concat([dust,tmp\_d], axis = 0, sort = **False**)  
 *#이산화질소*  
 tmp\_n = no2\_1.iloc[i:i+1, j:j+12].rename(columns={'**{}**. 01'.format(x) : '1월 **{}**'.format(pol\_list[2]), '**{}**. 02'.format(x) : '2월 **{}**'.format(pol\_list[2]),  
 '**{}**. 03'.format(x) : '3월 **{}**'.format(pol\_list[2]), '**{}**. 04'.format(x) : '4월 **{}**'.format(pol\_list[2]),  
 '**{}**. 05'.format(x) : '5월 **{}**'.format(pol\_list[2]), '**{}**. 06'.format(x) : '6월 **{}**'.format(pol\_list[2]),  
 '**{}**. 07'.format(x) : '7월 **{}**'.format(pol\_list[2]), '**{}**. 08'.format(x) : '8월 **{}**'.format(pol\_list[2]),  
 '**{}**. 09'.format(x) : '9월 **{}**'.format(pol\_list[2]), '**{}**. 10'.format(x) : '10월 **{}**'.format(pol\_list[2]),  
 '**{}**. 11'.format(x) : '11월 **{}**'.format(pol\_list[2]), '**{}**. 12'.format(x) : '12월 **{}**'.format(pol\_list[2])})  
 no2 = pd.concat([no2,tmp\_n], axis = 0, sort = **False**)  
 *#오존*  
 tmp\_o = o3\_1.iloc[i:i+1, j:j+12].rename(columns={'**{}**. 01'.format(x) : '1월 **{}**'.format(pol\_list[3]), '**{}**. 02'.format(x) : '2월 **{}**'.format(pol\_list[3]),  
 '**{}**. 03'.format(x) : '3월 **{}**'.format(pol\_list[3]), '**{}**. 04'.format(x) : '4월 **{}**'.format(pol\_list[3]),  
 '**{}**. 05'.format(x) : '5월 **{}**'.format(pol\_list[3]), '**{}**. 06'.format(x) : '6월 **{}**'.format(pol\_list[3]),  
 '**{}**. 07'.format(x) : '7월 **{}**'.format(pol\_list[3]), '**{}**. 08'.format(x) : '8월 **{}**'.format(pol\_list[3]),  
 '**{}**. 09'.format(x) : '9월 **{}**'.format(pol\_list[3]), '**{}**. 10'.format(x) : '10월 **{}**'.format(pol\_list[3]),  
 '**{}**. 11'.format(x) : '11월 **{}**'.format(pol\_list[3]), '**{}**. 12'.format(x) : '12월 **{}**'.format(pol\_list[3])})  
 o3 = pd.concat([o3,tmp\_o], axis = 0, sort = **False**)  
 *#아황산가스*  
 tmp\_s = so2\_1.iloc[i:i+1, j:j+12].rename(columns={'**{}**. 01'.format(x) : '1월 **{}**'.format(pol\_list[4]), '**{}**. 02'.format(x) : '2월 **{}**'.format(pol\_list[4]),  
 '**{}**. 03'.format(x) : '3월 **{}**'.format(pol\_list[4]), '**{}**. 04'.format(x) : '4월 **{}**'.format(pol\_list[4]),  
 '**{}**. 05'.format(x) : '5월 **{}**'.format(pol\_list[4]), '**{}**. 06'.format(x) : '6월 **{}**'.format(pol\_list[4]),  
 '**{}**. 07'.format(x) : '7월 **{}**'.format(pol\_list[4]), '**{}**. 08'.format(x) : '8월 **{}**'.format(pol\_list[4]),  
 '**{}**. 09'.format(x) : '9월 **{}**'.format(pol\_list[4]), '**{}**. 10'.format(x) : '10월 **{}**'.format(pol\_list[4]),  
 '**{}**. 11'.format(x) : '11월 **{}**'.format(pol\_list[4]), '**{}**. 12'.format(x) : '12월 **{}**'.format(pol\_list[4])})  
 so2 = pd.concat([so2,tmp\_s], axis = 0, sort = **False**)  
 x += 1  
len(co), len(dust), len(no2), len(o3), len(so2)

Out[15]:

(160, 160, 160, 160, 160)

In [16]:

data\_list = [co, dust, no2, o3, so2]  
**for** item **in** data\_list:  
 item.reset\_index(drop = **True**, inplace = **True**)

In [17]:

etc = pd.DataFrame(index = range(0, 160), columns = ['year\_local', 'area', 'year'])

In [18]:

**import** **numpy** **as** **np**  
local = co\_1['지역']  
local = local.tolist()  
local

Out[18]:

['계',  
 '부산광역시',  
 '대구광역시',  
 '인천광역시',  
 '광주광역시',  
 '대전광역시',  
 '울산광역시',  
 '경기도',  
 '강원도',  
 '충청북도',  
 '충청남도',  
 '전라북도',  
 '전라남도',  
 '경상북도',  
 '경상남도',  
 '제주도']

In [19]:

i = 0  
**for** item **in** local:  
 **for** j **in** range (0, 10):  
 etc['year\_local'][i] = '201**{}**\_**{}**'.format(j, item)  
 etc['area'][i] = item  
 etc['year'][i] = '201**{}**'.format(j)  
 i += 1  
etc

Out[19]:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **year\_local** | **area** | **year** |
| 0 | 2010\_계 | 계 | 2010 |
| 1 | 2011\_계 | 계 | 2011 |
| 2 | 2012\_계 | 계 | 2012 |
| 3 | 2013\_계 | 계 | 2013 |
| 4 | 2014\_계 | 계 | 2014 |
| ... | ... | ... | ... |
| 155 | 2015\_제주도 | 제주도 | 2015 |
| 156 | 2016\_제주도 | 제주도 | 2016 |
| 157 | 2017\_제주도 | 제주도 | 2017 |
| 158 | 2018\_제주도 | 제주도 | 2018 |
| 159 | 2019\_제주도 | 제주도 | 2019 |

160 rows × 3 columns

In [20]:

co = pd.concat([etc, co], axis = 1, sort = **False**)  
dust = pd.concat([etc, dust], axis = 1, sort = **False**)  
no2 = pd.concat([etc, no2], axis = 1, sort = **False**)  
o3 = pd.concat([etc, o3], axis = 1, sort = **False**)  
so2 = pd.concat([etc, so2], axis = 1, sort = **False**)

In [21]:

co.to\_csv('refined\_co.csv', encoding = 'CP949', index=**False**)  
dust.to\_csv('refined\_dust.csv', encoding = 'CP949', index=**False**)  
no2.to\_csv('refined\_no2.csv', encoding = 'CP949', index=**False**)  
o3.to\_csv('refined\_o3.csv', encoding = 'CP949', index=**False**)  
so2.to\_csv('refined\_so2.csv', encoding = 'CP949', index=**False**)

In [ ]:

In [ ]: