

# **파이썬 프로그래밍 기말프로젝트**

조장 : 2017150003 권순호

조원 : 2017152037 정우영

**주제 : 우리나라의 미세먼지 농도**

# 목차

1. 주제, 데이터, 쿼리문에 대해 설명
2. 쿼리문 수행 과정 및 데이터 해석
3. 결론
4. 후기
5. 참고문헌 및 출처

## 주제, 데이터, 쿼리문에 대해 설명

우리 조는 최근 발병하게 된 코로나19사태로 인해 마스크를 착용하고 다니는 것에 많은 불편함을 느끼는 것에서 영감을 받았다. 코로나 이전에 우리는 주로 미세먼지 농도 때문에 마스크를 사용하였다. 이에 우리나라의 미세먼지 농도가 매년 혹은 매월 어떻게 변하게 되었는지에 대해 조사를 하기로 계획을 세우게 되었다.

먼저 우리나라는 어떤 기준으로 어떻게 나누는지에 대해 알아야 한다고 생각했고, 그에 따라 조사한 결과  $151\mu\text{g}/\text{m}^3$  이상은 매우 나쁨,  $81 \sim 150\mu\text{g}/\text{m}^3$ 은 나쁨,  $31\sim80\mu\text{g}/\text{m}^3$ 을 보통  $0\sim30\mu\text{g}/\text{m}^3$ 을 좋음으로 나누어 표현한다는 것을 알게 되었고, 미세먼지도 초미세먼지(PM2.5)와 일반 미세먼지(PM10)로 나뉜다는 것을 알게 되었다.

그 이후, 미세먼지 농도를 바탕으로 무엇을 추출할지 고민하였고, 우리 조는 우리나라의 미세먼지가 주로 어떤 계절에 심한지, 어느 도시에서 가장 심한지, 중국발 황사가 우리나라 미세먼지에 얼마나 큰 영향을 미치는지, 코로나 전과 후를 비교하여 코로나가 미세먼지의 변화에 얼마나 영향을 미쳤는지에 대해 데이터를 통해 알아보고 이유까지 분석해보기로 하였다.

위의 내용을 바탕으로 우리 조는 공공데이터 csv와 한국 환경관리 공단 (에어코리아) 사이트에서 자료를 수집하여, 대기 환경 연간 보고서를 통해 우리가 조사하고자 하는 내용을 조사하였다.

아래부터는 편의를 위해 초미세먼지를 PM2.5, 미세먼지를 PM10으로 표현하겠다.

## 쿼리문 수행 과정 및 데이터 해석

우선 우리는 19년도를 기준으로 우리나라의 평균 PM10과 PM2.5의 농도를 살펴보았다.

더 정확한 데이터를 한눈에 보기 위해, 그래프를 출력시키도록 해보았다.

```
In [21]: import csv
import matplotlib.pyplot as plt
f = open('우리나라 날짜별 미세먼지 비교.csv', encoding='cp949')
data = csv.reader(f)
header = next(data)

min_25 = 999
max_25 = -999
min_10 = 999
max_10 = -999

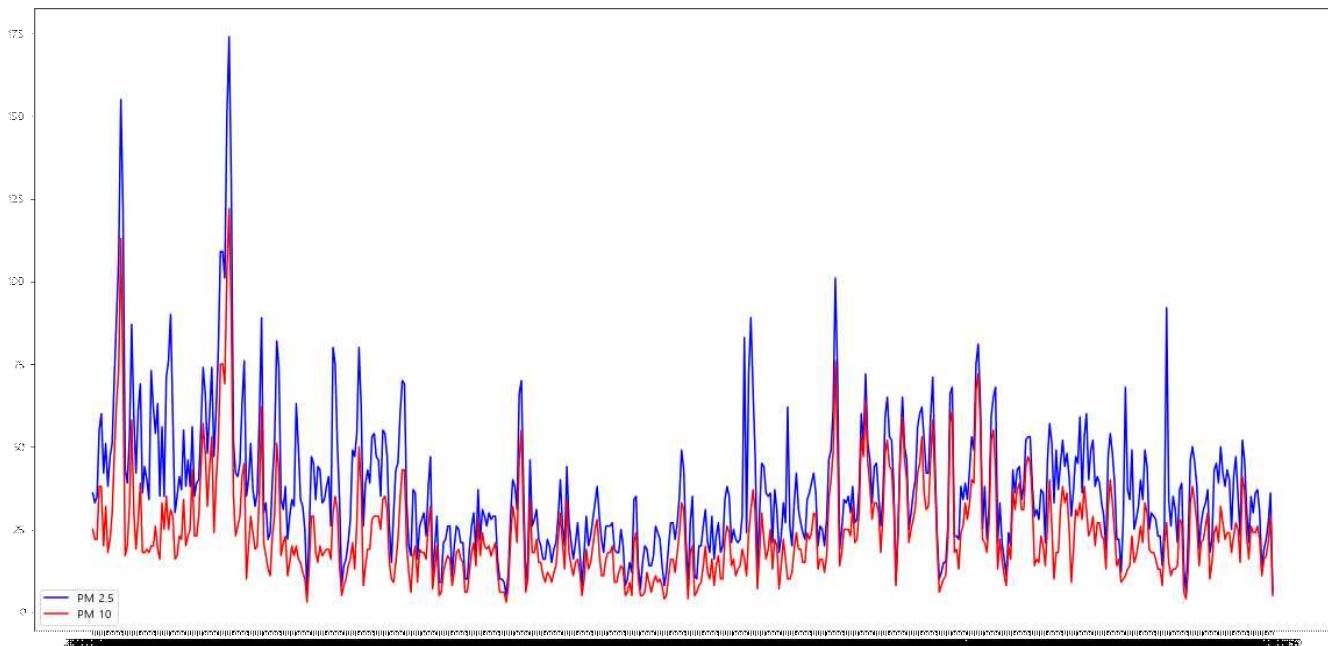
min_25_date = ""
max_25_date = ""
min_10_date = ""
max_10_date = ""

_25_arr = []
_10_arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    _10_arr.append(int(line[1]))
    _25_arr.append(int(line[2]))
    if(int(line[1]) > max_10):
        max_10 = int(line[1])
        max_10_date = line[0]
    if(int(line[1]) < min_10):
        min_10 = int(line[1])
        min_10_date = line[0]
    if(int(line[2]) > max_25):
        max_25 = int(line[2])
        max_25_date = line[0]
    if(int(line[2]) < min_25):
        min_25 = int(line[2])
        min_25_date = line[0]

plt.figure(figsize=(20, 10))
plt.plot(date_arr,_10_arr,'b',label='PM 2.5')
plt.plot(date_arr,_25_arr,'r',label='PM 10')
plt.legend(loc=3)
plt.show()

print("우리나라에서 PM2.5(초미세먼지)가 가장 많았던 날은 ", max_25_date, "이며, 정도는 ",max_25,"㎍/㎥.")
print("우리나라에서 PM2.5(초미세먼지)가 가장 적었던 날은 ", min_25_date, "이며, 정도는 ",min_25,"㎍/㎥.")
print("우리나라에서 PM10(미세먼지)가 가장 많았던 날은 ", max_10_date, "이며, 정도는 ",max_10,"㎍/㎥.")
print("우리나라에서 PM10(미세먼지)가 가장 적었던 날은 ", min_10_date, "이며, 정도는 ",min_10,"㎍/㎥.")
```



1년 365일이 모두 들어가 있어서 아래 X축의 내용은 거의 보이지 않는다.  
 파란색이 PM2.5, 빨간색이 PM10에 대한 그래프다.  
 그리고 따로 미세먼지의 종류에 따라 미세먼지가 가장 많았던 날과 가장 적었던 날을 아래처럼 출력시켜보았다.

우리나라에서 PM2.5(초미세먼지)가 가장 많았던 날은 2019-03-05이며, 정도는  $122 \mu\text{g}/\text{m}^3$ .  
 우리나라에서 PM2.5(초미세먼지)가 가장 적었던 날은 2019-04-10이며, 정도는  $3 \mu\text{g}/\text{m}^3$ .  
 우리나라에서 PM10(미세먼지)가 가장 많았던 날은 2019-03-05이며, 정도는  $174 \mu\text{g}/\text{m}^3$ .  
 우리나라에서 PM10(미세먼지)가 가장 적었던 날은 2019-07-11이며, 정도는  $5 \mu\text{g}/\text{m}^3$ .

이를 통해 보통 3월이나 4월인 봄쯤에 미세먼지가 가장 높게 나온다는 것을 확인할 수 있다.

미세먼지 중에서도 초미세먼지(PM10)와 일반 미세먼지(PM10)가 있는데 초미세먼지가 가장 높았던 날과 일반 미세먼지(PM10)가 가장 높았던 날이 3월 5일로 동일함을 알 수 있다.

그래서 초미세먼지(PM2.5)가 가장 낮았던 날과 일반 미세먼지(PM10)가 가장 낮았던 날도 서로 동일하지 않을까 생각했지만, 초미세먼지(PM2.5)가 가장 낮았던 날은 4월 10일이고 미세먼지(PM10)가 가장 낮았던 날은 7월 11일인 것으로 보아 꼭 그런 것은 아니라고 할 수 있겠다.

이 쿼리문을 통해 우리가 알 수 있었던 점은

1. 초미세먼지(PM2.5)와 일반 미세먼지(PM10)의 양은 대부분은 정비례한다는 것을 알 수 있다.
2. 일 년 중 가장 미세먼지가 높은 날은 3월쯤임을 알 수 있다.  
(황사가 3월 초쯤에 일어나므로 황사의 영향이라고 추측된다)

이를 통해 우리 조는 중국 미세먼지의 영향으로 미세먼지가 봄에 미세먼지가 가장 많다고 측정되는지와 우리나라 도시별로 미세먼지는 얼마나 다르게 측정되는지 궁금해졌다.

그리고 19년과 20년의 큰 차이는 코로나가 터졌는지 안 터졌는지므로 19년과 20년의 미세먼지 농도를 비교하여 코로나가 미세먼지에 얼마나 영향을 주었는지 알아보았다.

도시별로 측정하기에는 도시가 너무 많아서 광역시를 범위로 한정 짓기로 하였다. 우선 광역시 별로 얼마나 차이가 나는지 측정을 해보았다.

광역시에는 부산광역시, 인천광역시, 대구광역시, 대전광역시, 광주광역시, 울산광역시가 있다.

일단, 20년 기준으로 PM10과 PM2.5를 측정해보았다.

그리고 가독성을 위해 그래프로 출력하였다.

20년 12월은 아직 진행 중이므로 12월은 일부만 담았다.

-> 부산광역시에 대한 데이터(PM2.5) 추출

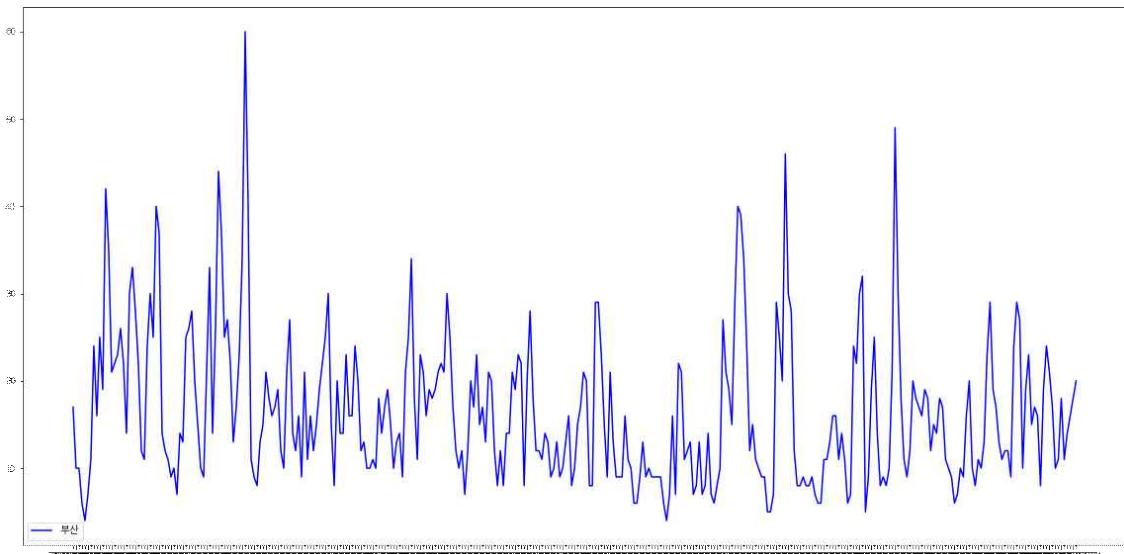
```
In [5]: import csv
import matplotlib.pyplot as plt

f = open('우리나라 2020년 도시별 PM2.5(초미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[2]))
    if(int(line[2]) > max):
        max = int(line[2])
        max_date = line[0]
    if(int(line[2]) < min):
        min = int(line[2])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr,'b',label='인천')
plt.legend(loc=3)
plt.show()

print("부산광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max," $\mu\text{g}/\text{m}^3$ 다.")
print("부산광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min," $\mu\text{g}/\text{m}^3$ 다.")
```



부산광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020-02-02이며, 정도는 60  $\mu\text{g}/\text{m}^3$ 다.  
부산광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020-01-27이며, 정도는 4  $\mu\text{g}/\text{m}^3$ 다.

-> 인천광역시에 대한 데이터(PM2.5) 추출

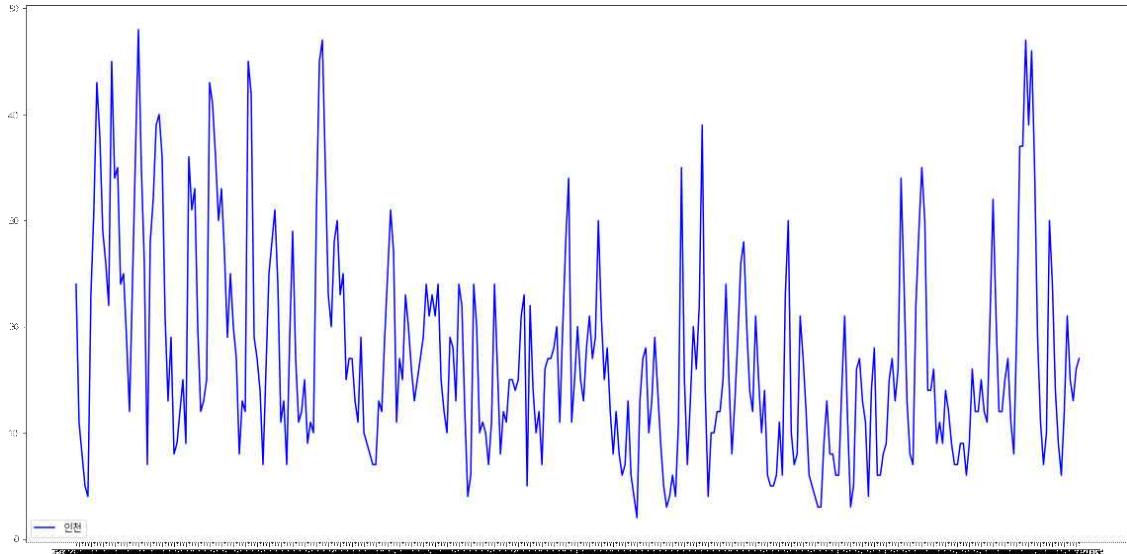
```
In [3]: import csv
import matplotlib.pyplot as plt

f = open('우리나라 2020년 도시별 PM2.5(초미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[4]))
    if(int(line[4]) > max):
        max = int(line[4])
        max_date = line[0]
    if(int(line[4]) < min):
        min = int(line[4])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr,'b',label='인천')
plt.legend(loc=8)
plt.show()

print("인천광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max," $\mu\text{g}/\text{m}^3$ 다.")
print("인천광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min," $\mu\text{g}/\text{m}^3$ 다.")
```



인천광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020-01-10 이며, 정도는 48  $\mu\text{g}/\text{m}^3$ 다.  
인천광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020-01-27 이며, 정도는 4  $\mu\text{g}/\text{m}^3$ 다.

-> 광주광역시에 대한 데이터(PM2.5) 추출

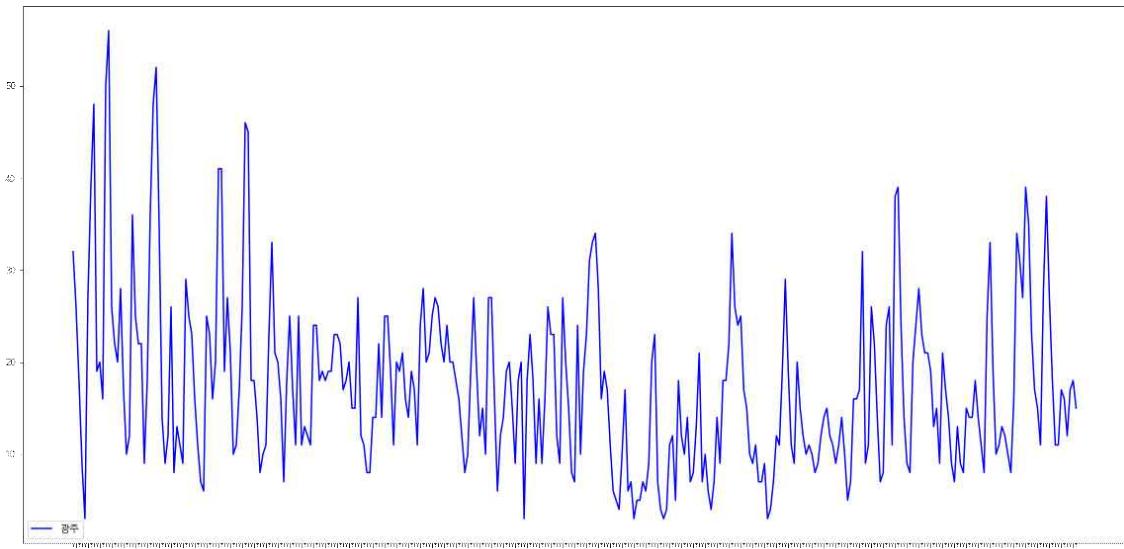
```
In [8]: import csv
import matplotlib.pyplot as plt

f = open('우리나라 2020년 도시별 PM2.5(초미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    if(line[5]== ''):
        line[5] ='13'
    arr.append(int(line[5]))
    if(int(line[5]) > max):
        max = int(line[5])
        max_date = line[0]
    if(int(line[5]) < min):
        min = int(line[5])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr, 'b', label='광주')
plt.legend(loc=3)
plt.show()
print("광주광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max,"µg/m³다.")
print("광주광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min,"µg/m³다.")
```

NAN 값이 존재하여서 빈자리에 중간값을 넣어주었다.



광주광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020/01/19이며, 정도는 56 µg/m<sup>3</sup>다.  
광주광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020/01/27이며, 정도는 3 µg/m<sup>3</sup>다.

-> 대전광역시에 대해 데이터(PM2.5) 추출

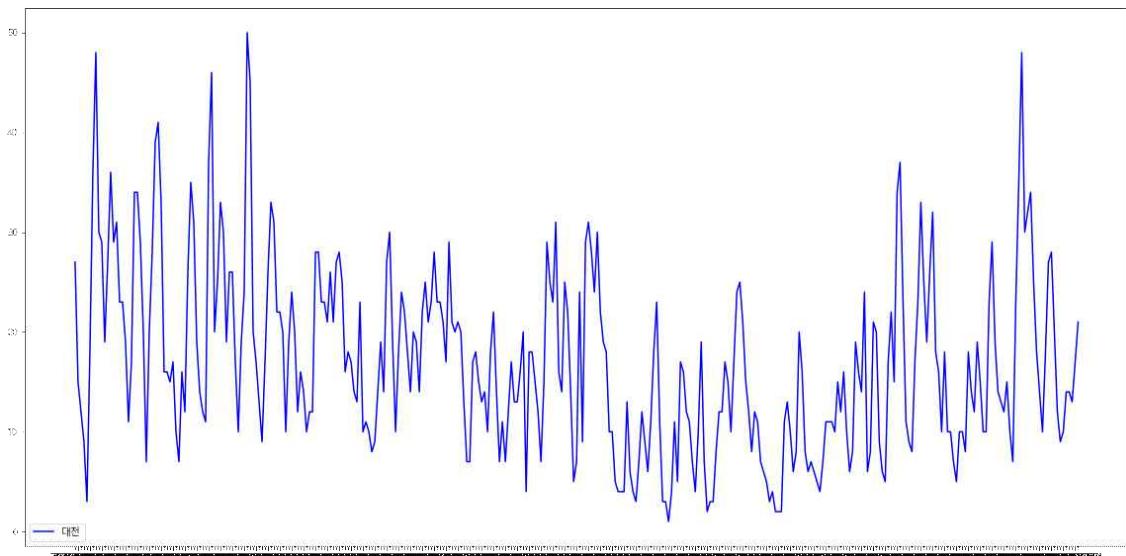
```
In [7]: import csv
import matplotlib.pyplot as plt

f = open('우리나라 2020년 도시별 PM2.5(초미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_date = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[6]))
    if(int(line[6]) > max):
        max = int(line[6])
        max_date = line[0]
    if(int(line[6]) < min):
        min = int(line[6])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr,'b',label='대전')
plt.legend(loc=3)
plt.show()

print("대전광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max,"ug/m다.")
print("대전광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min,"ug/m다.")
```



대전광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020/02/02이며, 정도는 50  $\mu\text{g}/\text{m}^3$ 이다.  
대전광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020/07/13이며, 정도는 1  $\mu\text{g}/\text{m}^3$ 이다.

-> 울산광역시에 대한 데이터(PM2.5) 추출

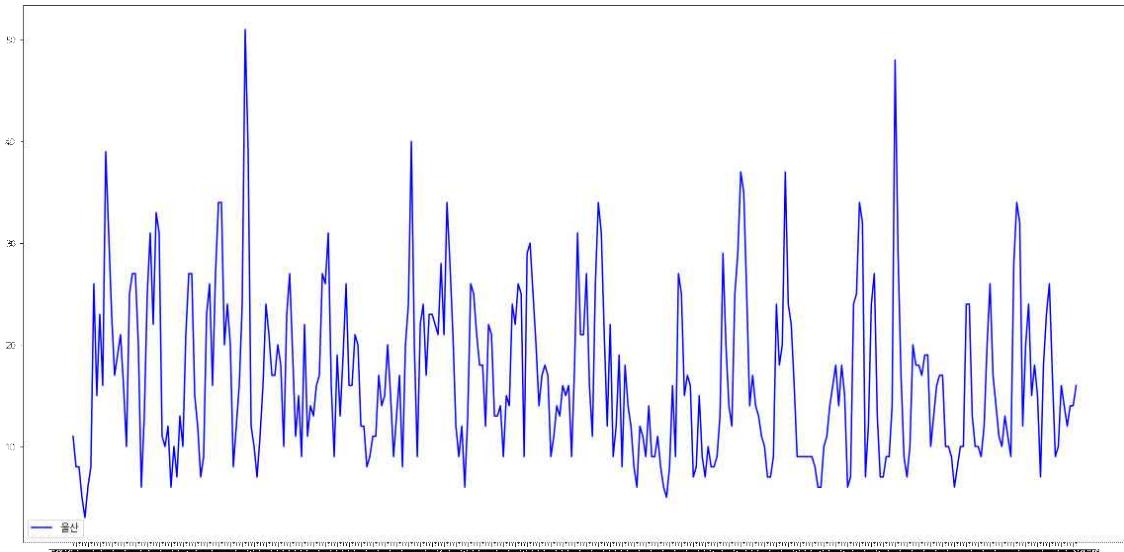
```
In [6]: import csv
import matplotlib.pyplot as plt

f = open('우리나라 2020년 도시별 PM2.5(초미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[7]))
    if(int(line[7]) > max):
        max = int(line[7])
        max_date = line[0]
    if(int(line[7]) < min):
        min = int(line[7])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr,'b',label='울산')
plt.legend(loc=3)
plt.show()

print("울산광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max," $\mu\text{g}/\text{m}^3$ 다.")
print("울산광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min," $\mu\text{g}/\text{m}^3$ 다.")
```



울산광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020/02/02이며, 정도는 51  $\mu\text{g}/\text{m}^3$ 다.  
울산광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020/01/27이며, 정도는 3  $\mu\text{g}/\text{m}^3$ 다.

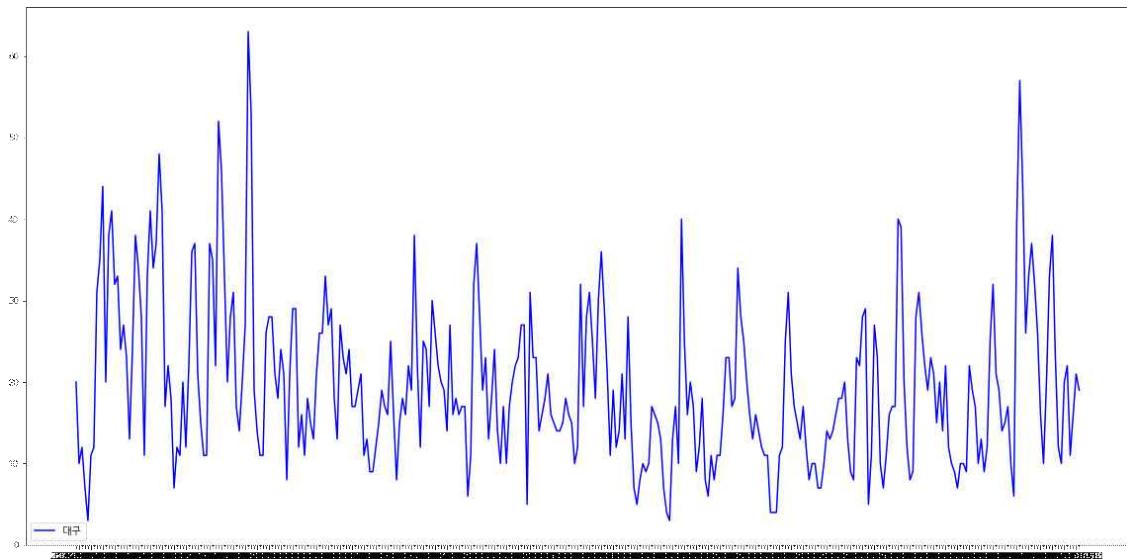
-> 대구광역시에 대한 데이터(PM2.5) 추출

```
In [9]: import csv
import matplotlib.pyplot as plt

f = open('우리나라_2020년_도시별_PM2.5(초미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[3]))
    if(int(line[3]) > max):
        max = int(line[3])
        max_date = line[0]
    if(int(line[3]) < min):
        min = int(line[3])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr,'b',label='대구')
plt.legend(loc=3)
plt.show()
print("대구광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max,"㎍/㎥다.")
print("대구광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min,"㎍/㎥다.")
```



대구광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020/02/02이며, 정도는 63  $\mu\text{g}/\text{m}^3$ .  
대구광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020/01/27이며, 정도는 3  $\mu\text{g}/\text{m}^3$ .

-> 부산광역시에 대한 데이터(PM10) 추출

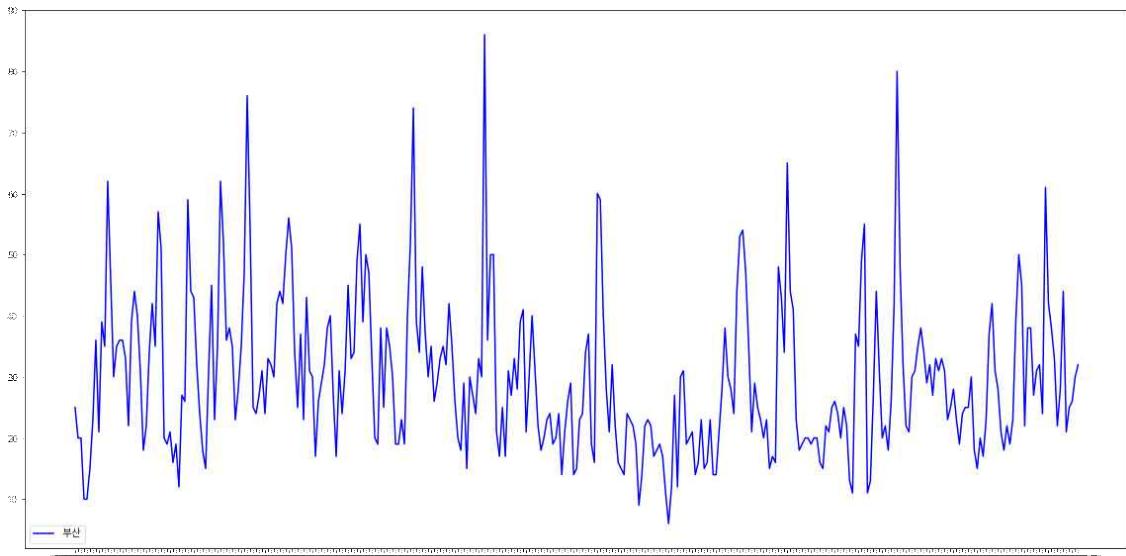
```
In [10]: import csv
import matplotlib.pyplot as plt

f = open('우리나라 2020년 도시별 PM10(미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[2]))
    if(int(line[2]) > max):
        max = int(line[2])
        max_date = line[0]
    if(int(line[2]) < min):
        min = int(line[2])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr,'b',label='부산')
plt.legend(loc=3)
plt.show()

print("부산광역시에서 PM10(미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max,"㎍/㎥다.")
print("부산광역시에서 PM10(미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min,"㎍/㎥다.")
```



부산광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/05/14 이며, 정도는 86  $\mu\text{g}/\text{m}^3$ 이다.  
부산광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/07/13 이며, 정도는 6  $\mu\text{g}/\text{m}^3$ 이다.

-> 인천광역시에 대한 데이터(PM10) 추출

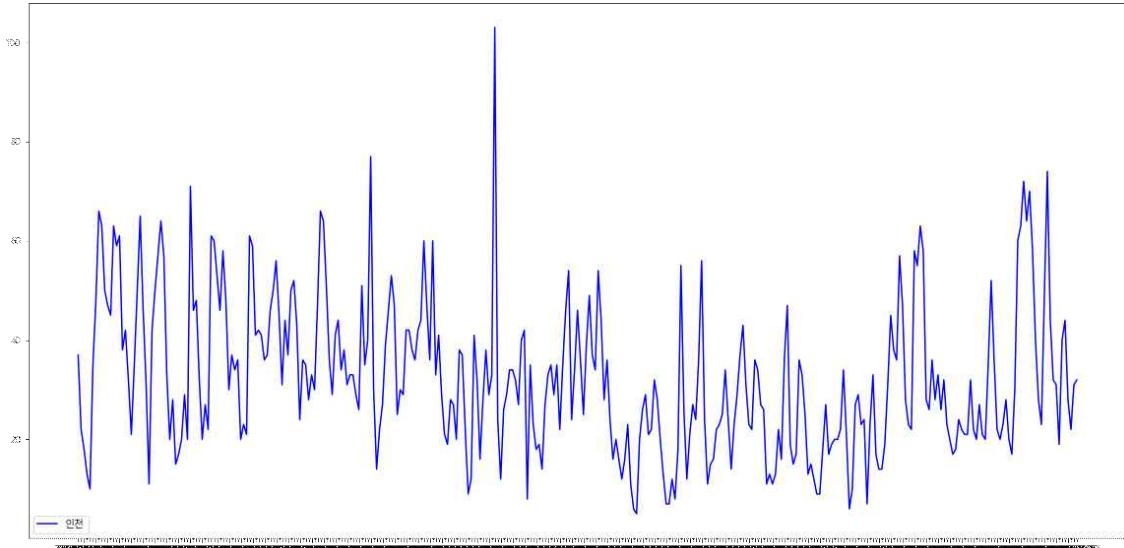
```
In [11]: import csv
import matplotlib.pyplot as plt

f = open('우리나라 2020년 도시별 PM10(미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[4]))
    if(int(line[4]) > max):
        max = int(line[4])
        max_date = line[0]
    if(int(line[4]) < min):
        min = int(line[4])
        min_date = line[0]

plt.rcParams['font', family = 'Malgun Gothic']
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr,'b',label='인천')
plt.legend(loc=3)
plt.show()

print("인천광역시에서 PM10(미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max,"㎍/㎥다.")
print("인천광역시에서 PM10(미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min,"㎍/㎥다.")
```



인천광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/05/11이며, 정도는 103  $\mu\text{g}/\text{m}^3$ .  
인천광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/07/24이며, 정도는 5  $\mu\text{g}/\text{m}^3$ .

-> 대구광역시에 대한 데이터(PM10) 추출

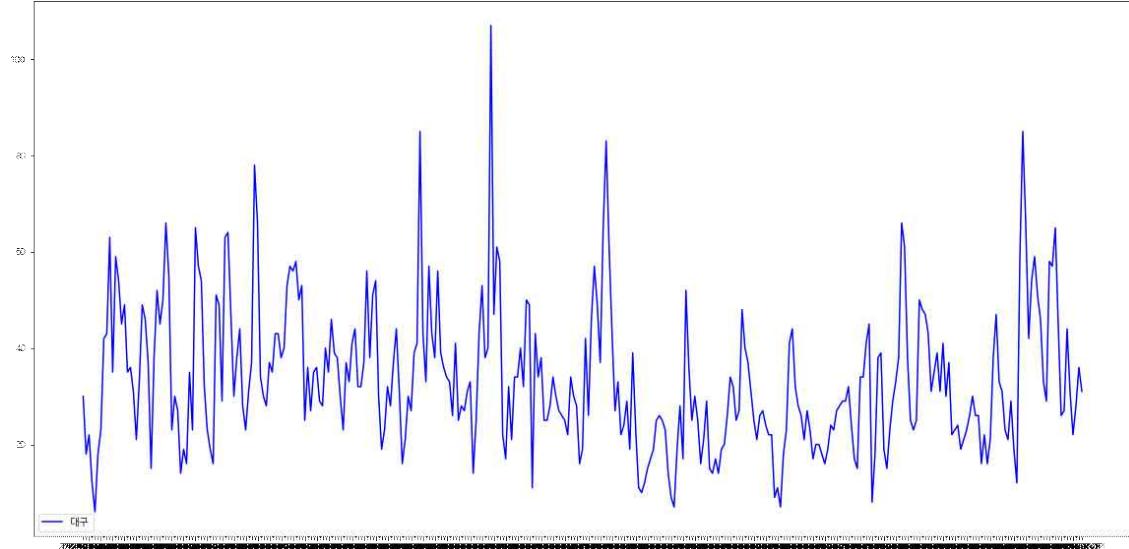
```
In [12]: import csv
import matplotlib.pyplot as plt

f = open('우리나라 2020년 도시별 PM10(미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[3]))
    if(int(line[3]) > max):
        max = int(line[3])
        max_date = line[0]
    if(int(line[3]) < min):
        min = int(line[3])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr, 'b', label='대구')
plt.legend(loc=3)
plt.show()

print("대구광역시에서 PM10(미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max,"ug/m다.")
print("대구광역시에서 PM10(미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min,"ug/m다.")
```



대구광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/05/14 이며, 정도는 107  $\mu\text{g}/\text{m}^3$ 이다.  
대구광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/01/27 이며, 정도는 6  $\mu\text{g}/\text{m}^3$ 이다.

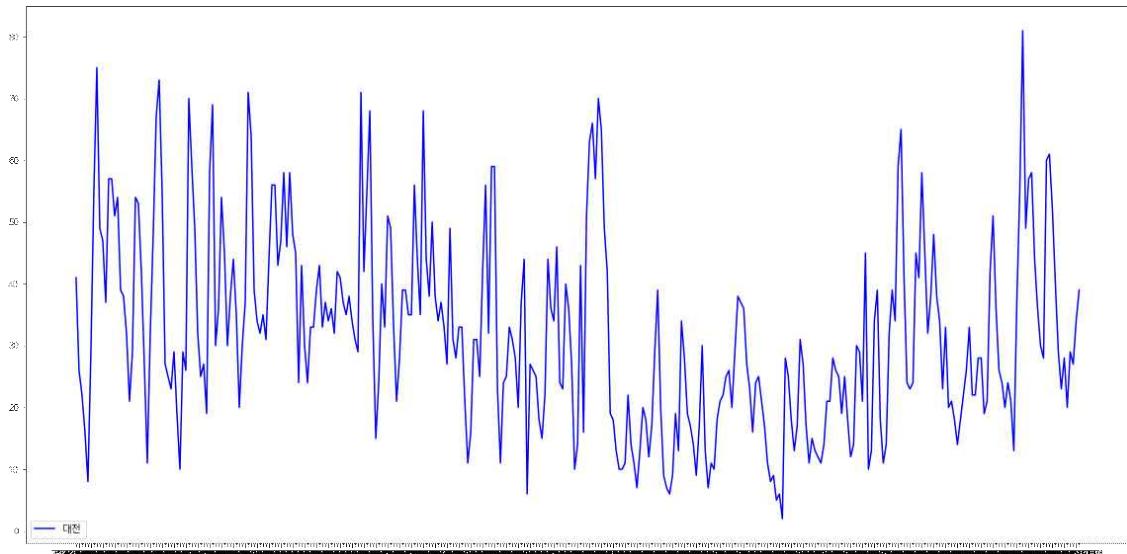
-> 대전광역시에 대한 데이터(PM10) 추출

```
In [15]: import csv
import matplotlib.pyplot as plt

f = open('우리나라 2020년 도시별 PM10(미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_date = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[6]))
    if(int(line[6]) > max):
        max = int(line[6])
        max_date = line[0]
    if(int(line[6]) < min):
        min = int(line[6])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr,'b',label='대전')
plt.legend(loc=3)
plt.show()
print("대전광역시에서 PM10(미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max," $\mu\text{g}/\text{m}^3$ 다.")
print("대전광역시에서 PM10(미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min," $\mu\text{g}/\text{m}^3$ 다.")
```



대전광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/11/16이며, 정도는 81  $\mu\text{g}/\text{m}^3$ 다.  
대전광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/08/06이며, 정도는 2  $\mu\text{g}/\text{m}^3$ 다.

-> 광주광역시에 대한 데이터(PM10) 추출

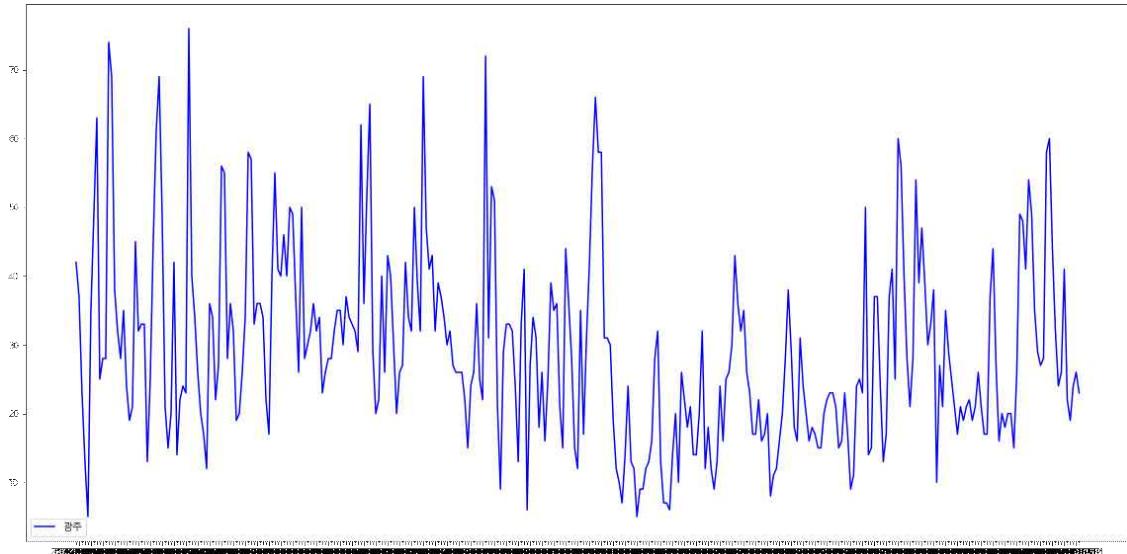
```
In [17]: import csv
import matplotlib.pyplot as plt

f = open('우리나라_2020년_도시별_PM10(미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    if(line[5]=='' or line[5]=='10'):
        arr.append(int(line[5]))
    if(int(line[5]) > max):
        max = int(line[5])
        max_date = line[0]
    if(int(line[5]) < min):
        min = int(line[5])
        min_date = line[0]

plt.rcParams['font.family'] = 'Malgun Gothic'
plt.figure(figsize=(20, 10))
plt.plot(date_arr, arr, 'b', label='광주')
plt.legend(loc=3)
plt.show()

print("광주광역시에서 PM10(미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ", max, "ug/m³다.")
print("광주광역시에서 PM10(미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ", min, "ug/m³다.")
```



광주광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/02/22이며, 정도는 76 ug/m<sup>3</sup>다.  
광주광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/01/27이며, 정도는 5 ug/m<sup>3</sup>다.

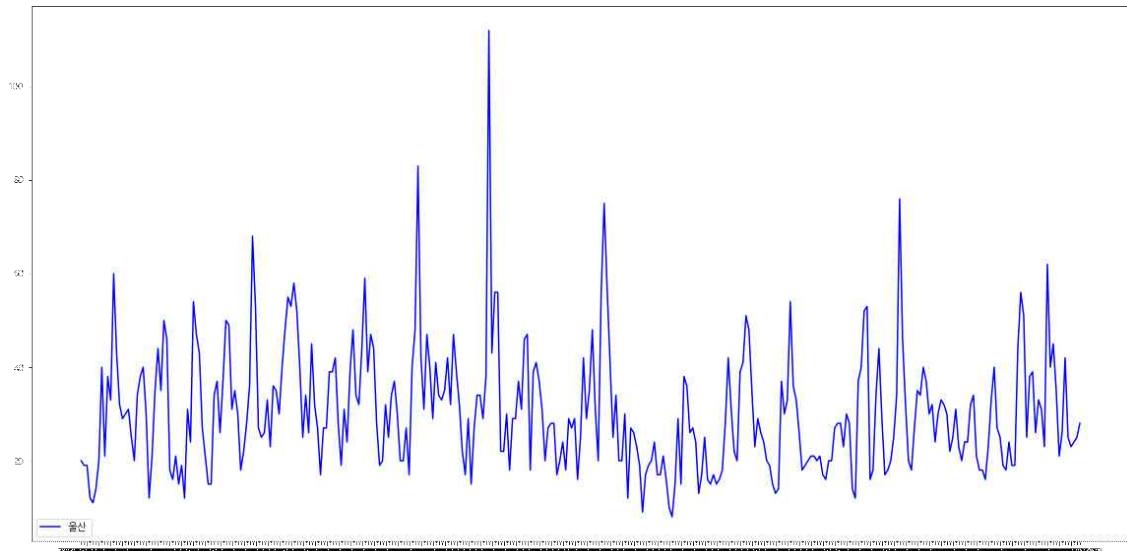
-> 울산광역시에 대한 데이터(PM10) 추출

```
In [61]: import csv
f = f = open('우리나라 2020년 도시별 PM10(미세먼지).csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
header = next(data)
min = 999
max = -999
min_date = ""
max_data = ""
arr = []
date_arr = []

for line in data:
    date_arr.append(line[0])
    arr.append(int(line[7]))
    if(int(line[7]) > max):
        max = int(line[7])
        max_date = line[0]
    if(int(line[7]) < min):
        min = int(line[7])
        min_date = line[0]

plt.rc('font', family = 'Malgun Gothic')
plt.figure(figsize=(20, 10))
plt.plot(date_arr,arr,'b',label='울산')
plt.legend(loc=3)
plt.show()

print("울산광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 ", max_date, "이며, 정도는 ",max,"ug/m다.")
print("울산광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 ", min_date, "이며, 정도는 ",min,"ug/m다.")
```



울산광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/05/14 이며, 정도는 112  $\mu\text{g}/\text{m}^3$ .  
울산광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/07/13 이며, 정도는 8  $\mu\text{g}/\text{m}^3$ .

종합해보면 아래와 같다.

부산광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020-02-02이며, 정도는 60  $\mu\text{g}/\text{m}^3$ 이다.  
부산광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020-01-27이며, 정도는 4  $\mu\text{g}/\text{m}^3$ 이다.

인천광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020-01-10이며, 정도는 48  $\mu\text{g}/\text{m}^3$ 이다.  
인천광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020-01-27이며, 정도는 4  $\mu\text{g}/\text{m}^3$ 이다.

광주광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020/01/19이며, 정도는 56  $\mu\text{g}/\text{m}^3$ 이다.  
광주광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020/01/27이며, 정도는 3  $\mu\text{g}/\text{m}^3$ 이다.

대전광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020/02/02이며, 정도는 50  $\mu\text{g}/\text{m}^3$ 이다.  
대전광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020/07/13이며, 정도는 1  $\mu\text{g}/\text{m}^3$ 이다.

울산광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020/02/02이며, 정도는 51  $\mu\text{g}/\text{m}^3$ 이다.  
울산광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020/01/27이며, 정도는 3  $\mu\text{g}/\text{m}^3$ 이다.

대구광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020/02/02이며, 정도는 63  $\mu\text{g}/\text{m}^3$ 이다.  
대구광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020/01/27이며, 정도는 3  $\mu\text{g}/\text{m}^3$ 이다.

대전광역시에서 PM2.5(초미세먼지)가 가장 많았던 날은 2020/02/02이며, 정도는 50  $\mu\text{g}/\text{m}^3$ 이다.  
대전광역시에서 PM2.5(초미세먼지)가 가장 적었던 날은 2020/07/13이며, 정도는 1  $\mu\text{g}/\text{m}^3$ 이다.

부산광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/05/14이며, 정도는 86  $\mu\text{g}/\text{m}^3$ 이다.  
부산광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/07/13이며, 정도는 6  $\mu\text{g}/\text{m}^3$ 이다.

인천광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/05/11이며, 정도는 103  $\mu\text{g}/\text{m}^3$ 이다.  
인천광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/07/24이며, 정도는 5  $\mu\text{g}/\text{m}^3$ 이다.

대구광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/05/14이며, 정도는 107  $\mu\text{g}/\text{m}^3$ 이다.  
대구광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/01/27이며, 정도는 6  $\mu\text{g}/\text{m}^3$ 이다.

광주광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/02/22이며, 정도는 76  $\mu\text{g}/\text{m}^3$ 이다.  
광주광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/01/27이며, 정도는 5  $\mu\text{g}/\text{m}^3$ 이다.

울산광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/05/14이며, 정도는 112  $\mu\text{g}/\text{m}^3$ 이다.  
울산광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/07/13이며, 정도는 8  $\mu\text{g}/\text{m}^3$ 이다.

대전광역시에서 PM10(미세먼지)가 가장 많았던 날은 2020/11/16이며, 정도는 81  $\mu\text{g}/\text{m}^3$ 이다.  
대전광역시에서 PM10(미세먼지)가 가장 적었던 날은 2020/08/06이며, 정도는 2  $\mu\text{g}/\text{m}^3$ 이다.

PM2.5을 비교하면, 대전광역시가 초미세먼지(PM2.5)가 63으로 가장 높게 올라갔음을 알 수 있다. 그리고 인천광역시가 48로 제일 낮음을 알 수 있다.

PM10을 비교하면 112로 울산광역시에서 가장 높았고 광주광역시에서 76으로 가장 낮았다.

(초미세먼지의 경우 거의 대전을 제외한 모든 광역시에서 20. 01. 27에 가장 적었음을 알 수 있다.)

이를 통해 PM2.5의 농도는 대전광역시에서 가장 큰 것을 알 수 있고, PM10의 농도는 울산광역시에서 가장 큰 것을 알 수 있다.

처음에는 중국과 상대적으로 가까운 도시가 미세먼지 농도가 높지 않을까 생각했지만, 꼭 그런 것은 아닌 것을 알 수 있다.

그 다음은 중국발 황사가 우리나라 미세먼지에 얼마나 영향을 미치는지 알아보도록 한다.

```
In [31]: import csv
import matplotlib.pyplot as plt
f = open('20년 중국발 미세먼지 비교.csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)

maxPM10_1 = ''
maxPM10_2 = ''
maxPM25_1 = ''
maxPM25_2 = ''
max_wherePM10 = ""
max_wherePM25 = ""
max_datePM10 = ""
max_datePM25 = ""
max_differencePM10 = -999
max_differencePM25 = -999
differencePM10 = 0
differencePM25 = 0

for line in data:
    differencePM10 = int(line[2]) - int(line[3])
    if(differencePM10 > max_differencePM10):
        max_differencePM10 = differencePM10
        maxPM10_1 = line[2]
        maxPM10_2 = line[3]
        max_wherePM10 = line[1]
        max_datePM10 = line[0]
    if(differencePM25 > max_differencePM25):
        max_differencePM25 = differencePM25
        maxPM25_1 = line[4]
        maxPM25_2 = line[5]
        max_wherePM25 = line[1]
        max_datePM25 = line[0]
print("PM10기준, 월평균과 중국발 황사제외 월평균이 가장 컸던 날은",max_datePM10, "0")
print("PM2.5기준, 월평균과 중국발 황사제외 월평균이 가장 컸던 날은",max_datePM25, "0")
```

PM10기준, 월평균과 중국발 황사제외 월평균이 가장 컸던 날은 20년 3월 이고 인천광역시에서 월평균은  $50 \mu\text{g}/\text{m}^3$ , 중국발 황사 제외 월평균은  $36 \mu\text{g}/\text{m}^3$ 으로 차이는  $14 \mu\text{g}/\text{m}^3$ 이다  
PM2.5기준, 월평균과 중국발 황사제외 월평균이 가장 컸던 날은 20년 3월 이고 인천광역시에서 월평균은  $30 \mu\text{g}/\text{m}^3$ , 중국발 황사 제외 월평균은  $2 \mu\text{g}/\text{m}^3$ 으로 차이는  $8 \mu\text{g}/\text{m}^3$ 이다

처음에는 보고 놀랐다. 생각보다 월평균과 중국발 황사 제외 월평균의 차이가 그렇게 크지 않았다.

이를 통해 중국발 황사의 영향이 아예 없는 것은 아니지만, 생각보다 그렇게 크지 않음을 알 수 있다.

그 다음은 코로나 발생 전(19년), 코로나 발생 후(20년)을 비교하여 코로나가 미세먼지에 끼친 영향을 알아보기로 하였다.

코로나는 20년 1월 말부터 유행하였으므로 2월부터 6월까지의 범위에서 비교하기로 하였다.

```
In [55]: import csv
import matplotlib.pyplot as plt
f = open('19년과 20년 상반기 비교.csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
header = next(data)
count1 = 0
count2 = 0

for line in data:
    if(int(line[3]) - int(line[1]) > 0):
        count1 = count1 + 1
    if(int(line[4]) - int(line[2]) > 0):
        count2 = count2 + 1
print("2월초부터 6월말까지 20년도가 19년도보다 PM 10 농도가 큰 횟수는", count1,"번이다.")
print("2월초부터 6월말까지 20년도가 19년도보다 PM 2.5 농도가 큰 횟수는", count2,"번이다.")
```

2월초부터 6월말까지 20년도가 19년도보다 PM 10 농도가 큰 횟수는 69 번이다.

2월초부터 6월말까지 20년도가 19년도보다 PM 2.5 농도가 큰 횟수는 74 번이다.

<결과>

2월 초부터 6월 말까지는 총 150일 정도가 있는데 그중 대략 70일 정도가 20년이 19년보다 미세먼지 농도가 높았던 것을 알 수 있다.

즉, 20년과 19년의 상반기를 비교해보았을 때, 20년보다는 19년이 미세먼지 농도가 더 높은 날이 더 많은 것을 보아, 코로나가 미세먼지에 그렇게 큰 영향을 준 것 같지는 않다고 추측할 수 있었다.

# 결론

- 초미세먼지(PM2.5)와 일반 미세먼지(PM10)의 양은 대부분은 정비례한다는 것을 알 수 있다.
- 일 년 중 가장 미세먼지가 높은 날은 3월~4월 임을 알 수 있다.  
(황사가 3월 초쯤에 일어나므로 황사의 영향이라고 추측된다.)
- 중국과 상대적으로 가까운 도시가 미세먼지 농도가 높지 않을까 생각했지만, 꼭 그런 것은 아닌 것을 알 수 있다.
- 중국에서 시작된 황사의 영향이 아예 없는 것은 아니지만, 생각보다 그렇게 크지 않음을 알 수 있다.
- 코로나가 미세먼지에 그렇게 큰 영향을 준 것 같지는 않다고 추측할 수 있었다.

## 후기

권순호 : 생각보다 의외의 결과가 나와서 놀랐다. 코로나가 미세먼지에 영향이 많이 없을 것 같긴 했는데 중국발 미세먼지가 우리나라에 영향이 지대할 줄 알았는데 그 정도는 되지 않아서 놀랐다. 공공데이터를 찾아 내가 원하는 데이터를 추출할 수 있도록 다시 가공하는 것과 그 가공한 데이터에서 의미 있는 추출 및 분석을 하는 게 쉽지 않았지만 데이터 분석을 통해 새로운 것을 알게 되는 기회가 된 것 같아 뿌듯했다.

정우영 : 코로나와 미세먼지의 영향이 없을 것이라고는 생각했지만, 중국에서 시작된 황사로 인한 미세먼지가 우리나라에 영향을 많이 줄 것이라고 생각했으나, 그것은 아니라는 것에 신기함을 느낄 수 있었고, 공공데이터를 바탕으로 우리가 찾고자 하고 궁금해 던 것을 알아내는 기회가 된 것에 있어서 신기함을 느낀다. 이를 바탕으로 더욱 많은 데이터들을 바탕으로 궁금증을 해결해 나갈 수 있을 것이라 생각이 들었다.

<참고 문헌 및 출처>

- 출처

1. PM10의 위치별 일자별 모음

-> 한국환경공단 - 에어코리아 환경부 대기환경 정보/ 실시간 자료 조회

[https://www.airkorea.or.kr/web/pmRelay?itemCode=11008&pMENU\\_NO=109](https://www.airkorea.or.kr/web/pmRelay?itemCode=11008&pMENU_NO=109)

2. 시도별 대기정보(pm2.5)의 모음 - 에어코리아 환경부 대기 환경 정보/시도별 대기정보 조회

[https://www.airkorea.or.kr/web/sidoQualityCompare?itemCode=10008&pMENU\\_NO=102](https://www.airkorea.or.kr/web/sidoQualityCompare?itemCode=10008&pMENU_NO=102)

-참고 문헌

3. [파이썬 머신러닝 완벽 가이드, 권철민 위키북스 2020]