

미세먼지, 초미세먼지 현황에 대한 다양한 분석

20201219 김인서

선정한 데이터 & 선정 이유

- <https://data.seoul.go.kr/dataList/OA-2220/S/1/datasetView.do#>
- 서울 열린데이터 광장 (서울시 기간별 일평균 대기환경 정보)
- 2014~2023년까지(즉, 최근 10년 동안)의 데이터를 수집함.

| | A | B | C | D | E | F | G | H | I |
|----|----------|-----|------|------|-------|---------|-------|-------|-------|
| 1 | 측정일자 | 권역명 | 측정소명 | 미세먼지 | 초미세먼지 | 오존(ppb) | 이산화질소 | 일산화탄소 | 아황산가스 |
| 2 | 20140101 | 서남권 | 강서구 | 116 | 21 | 0.015 | 0.035 | 0.7 | 0.012 |
| 3 | 20140101 | 서남권 | 금천구 | 113 | 22 | 0.013 | 0.026 | 0.6 | 0.009 |
| 4 | 20140101 | 서남권 | 양천구 | 119 | 24 | 0.011 | 0.034 | 0.7 | 0.011 |
| 5 | 20140101 | 서북권 | 은평구 | 119 | 28 | 0.011 | 0.042 | 0.6 | 0.009 |
| 6 | 20140101 | 동북권 | 도봉구 | 119 | 22 | 0.026 | 0.024 | 0.6 | 0.01 |
| 7 | 20140101 | 동북권 | 성북구 | 107 | 22 | 0.014 | 0.041 | 0.7 | 0.01 |
| 8 | 20140101 | 동북권 | 광진구 | 132 | 27 | 0.012 | 0.031 | 0.8 | 0.007 |
| 9 | 20140101 | 동북권 | 강북구 | 117 | 21 | 0.017 | 0.027 | 0.4 | 0.01 |
| 10 | 20140101 | 동남권 | 강남구 | 117 | 29 | 0.016 | 0.031 | 0.5 | 0.01 |

| | | | | | | | | | | | | | |
|----|----------|-----|------|-------|-----------|-------|-------|------|-------|---------|-------|-------|-------|
| 1 | 측정일자 | 권역명 | 측정소명 | 이산화질소 | 오존농도(ppb) | 이산화탄소 | 아황산가스 | 미세먼지 | 초미세먼지 | 오존(ppb) | 이산화질소 | 일산화탄소 | 아황산가스 |
| 2 | 20180101 | 동남권 | 강남구 | 0.034 | 0.009 | 0.6 | 0.006 | 34 | 22 | 16 | 0.025 | 0.5 | 0.009 |
| 3 | 20180101 | 동남권 | 강동구 | 0.039 | 0.01 | 0.7 | 0.005 | 49 | 24 | 09 | 0.032 | 0.4 | 0.007 |
| 4 | 20180101 | 동북권 | 강북구 | 0.026 | 0.018 | 0.6 | 0.004 | 38 | 18 | ... | ... | ... | ... |
| 5 | 20180101 | 서남권 | 강서구 | 0.038 | 0.011 | 0.7 | 0.004 | | 13 | | | | |
| 6 | 20180101 | 서남권 | 관악구 | 0.037 | 0.008 | 0.5 | 0.008 | 38 | 24 | | | | |
| 7 | 20180101 | 동북권 | 광진구 | 0.038 | 0.012 | 0.7 | 0.005 | 42 | 22 | | | | |
| 8 | 20180101 | 서남권 | 구로구 | 0.025 | 0.008 | 0.6 | 0.007 | 48 | 28 | | | | |
| 9 | 20180101 | 서남권 | 금천구 | 0.037 | 0.01 | 0.6 | 0.005 | 39 | 21 | | | | |
| 10 | 20180101 | 동북권 | 노원구 | 0.034 | 0.011 | 0.8 | 0.007 | 42 | 21 | | | | |
| 11 | 20180101 | 동북권 | 도봉구 | 0.03 | 0.016 | 0.6 | 0.006 | 45 | 22 | | | | |
| 12 | 20180101 | 동북권 | 동대문구 | 0.032 | 0.013 | 0.6 | 0.006 | 46 | 21 | | | | |
| 13 | 20180101 | 서남권 | 동작구 | 0.036 | 0.012 | 0.6 | 0.005 | 40 | 19 | | | | |
| 14 | 20180101 | 서북권 | 마포구 | 0.034 | 0.011 | 0.6 | 0.006 | 39 | 21 | | | | |
| 15 | 20180101 | 서북권 | 서대문구 | | | | | | | | | | |
| 16 | 20180101 | 동남권 | 서초구 | 0.031 | 0.01 | 0.6 | 0.005 | 49 | 24 | | | | |
| 17 | 20180101 | 동북권 | 성동구 | 0.033 | 0.009 | | 0.006 | 48 | 23 | | | | |
| 18 | 20180101 | 동북권 | 성북구 | 0.041 | 0.01 | 0.7 | 0.007 | 43 | 19 | | | | |
| 19 | 20180101 | 동남권 | 송파구 | 0.028 | 0.012 | 0.7 | 0.005 | 53 | 23 | | | | |
| 20 | 20180101 | 서남권 | 양천구 | 0.037 | 0.008 | 0.7 | 0.005 | 37 | 21 | | | | |
| 21 | 20180101 | 서남권 | 영등포구 | 0.031 | 0.01 | 0.7 | 0.007 | 48 | 24 | | | | |
| 22 | 20180101 | 도심권 | 용산구 | 0.029 | 0.011 | 0.3 | 0.005 | 39 | 18 | | | | |
| 23 | 20180101 | 서북권 | 은평구 | 0.015 | 0.014 | 0.7 | 0.003 | 41 | 21 | | | | |

- 최근 들어 미세먼지에 대한 관심이 높아지고 있는 상황
- 이에 따라 과거와 대비해 미세먼지와 초미세먼지 현황이 어떻게 변화했는지 궁금하여 선정하게 됨.
- 전국 데이터를 다 보는 것보다는 서울만 선정하여 진행하였음.

데이터 준비

• 데이터 병합

2014년부터 2017년까지의 데이터는 xlsx 파일로 되어 있고,
2018년부터 2023년까지의 데이터는 csv 파일로 되어 있다.

```
dust_2014 = pd.read_excel('기간별_일평균_대기환경_정보_2014년.xlsx') # xlsx 파일은 read_excel() 이용
dust_2015 = pd.read_excel('기간별_일평균_대기환경_정보_2015년.xlsx')
dust_2016 = pd.read_excel('기간별_일평균_대기환경_정보_2016년.xlsx')
dust_2017 = pd.read_excel('기간별_일평균_대기환경_정보_2017년.xlsx')
dust_2018 = pd.read_csv('기간별_일평균_대기환경_정보_2018년.csv', encoding='cp949') # csv 파일은 read_csv() 이용
dust_2019 = pd.read_csv('기간별_일평균_대기환경_정보_2019년.csv', encoding='cp949')
dust_2020 = pd.read_csv('기간별_일평균_대기환경_정보_2020년.csv', encoding='cp949')
dust_2021 = pd.read_csv('기간별_일평균_대기환경_정보_2021년.csv', encoding='cp949')
dust_2022 = pd.read_csv('기간별_일평균_대기환경_정보_2022년.csv', encoding='cp949')
dust_2023 = pd.read_csv('기간별_일평균_대기환경_정보_2023년.csv', encoding='cp949')
```

이름이 다른 열에 대해서 합치는 방법 중에는 merge도 있으나,
전체 컬럼 수가 많지 않으므로 전체에 대해서 대응되는 컬럼의 이름이 일치하도록 바꾼 후
concat를 사용하기로 하였다.

2014~2017년과 2018~2023년은 컬럼의 순서가 다르므로 이를 고려하여 바꾼다.

```
dust_2014.columns = ['측정일자', '권역명', '측정소명', '미세먼지', '초미세먼지', '오존', '이산화질소', '일산화탄소', '아황산가스']
dust_2015.columns = ['측정일자', '권역명', '측정소명', '미세먼지', '초미세먼지', '오존', '이산화질소', '일산화탄소', '아황산가스']
dust_2016.columns = ['측정일자', '권역명', '측정소명', '미세먼지', '초미세먼지', '오존', '이산화질소', '일산화탄소', '아황산가스']
dust_2017.columns = ['측정일자', '권역명', '측정소명', '미세먼지', '초미세먼지', '오존', '이산화질소', '일산화탄소', '아황산가스']
```

```
dust_2018.columns = ['측정일자', '권역명', '측정소명', '이산화질소', '오존', '일산화탄소', '아황산가스', '미세먼지', '초미세먼지']
dust_2019.columns = ['측정일자', '권역명', '측정소명', '이산화질소', '오존', '일산화탄소', '아황산가스', '미세먼지', '초미세먼지']
dust_2020.columns = ['측정일자', '권역명', '측정소명', '이산화질소', '오존', '일산화탄소', '아황산가스', '미세먼지', '초미세먼지']
dust_2021.columns = ['측정일자', '권역명', '측정소명', '이산화질소', '오존', '일산화탄소', '아황산가스', '미세먼지', '초미세먼지']
dust_2022.columns = ['측정일자', '권역명', '측정소명', '이산화질소', '오존', '일산화탄소', '아황산가스', '미세먼지', '초미세먼지']
dust_2023.columns = ['측정일자', '권역명', '측정소명', '이산화질소', '오존', '일산화탄소', '아황산가스', '미세먼지', '초미세먼지']
```

바꾼 데이터를 모두 합쳐서 하나의 데이터프레임으로 만든다.

```
total_data = [dust_2014, dust_2015, dust_2016, dust_2017, dust_2018, dust_2019, dust_2020, dust_2021, dust_2022, dust_2023]
```

```
dust = pd.concat(total_data, ignore_index=True) # 기존 인덱스를 무시하고 새로 부여한다.
dust
```

| | 측정일자 | 권역명 | 측정소명 | 미세먼지 | 초미세먼지 | 오존 | 이산화질소 | 일산화탄소 | 아황산가스 |
|-------|----------|-----|------|-------|-------|--------|--------|-------|--------|
| 0 | 20140101 | 서남권 | 강서구 | 116.0 | 21.0 | 0.0150 | 0.0350 | 0.70 | 0.0120 |
| 1 | 20140101 | 서남권 | 금천구 | 113.0 | 22.0 | 0.0130 | 0.0260 | 0.60 | 0.0090 |
| 2 | 20140101 | 서남권 | 양천구 | 119.0 | 24.0 | 0.0110 | 0.0340 | 0.70 | 0.0110 |
| 3 | 20140101 | 서북권 | 은평구 | 119.0 | 28.0 | 0.0110 | 0.0420 | 0.60 | 0.0090 |
| 4 | 20140101 | 동북권 | 도봉구 | 119.0 | 22.0 | 0.0260 | 0.0240 | 0.60 | 0.0100 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 91202 | 20231231 | 도심권 | 용산구 | 22.0 | 15.0 | 0.0146 | 0.0286 | 0.78 | 0.0025 |
| 91203 | 20231231 | 서북권 | 은평구 | 21.0 | 18.0 | 0.0178 | 0.0172 | 1.12 | 0.0025 |
| 91204 | 20231231 | 도심권 | 종로구 | 22.0 | 18.0 | 0.0180 | 0.0238 | 0.73 | 0.0029 |
| 91205 | 20231231 | 도심권 | 중구 | 20.0 | 18.0 | 0.0179 | 0.0261 | 0.67 | 0.0028 |
| 91206 | 20231231 | 동북권 | 중랑구 | 20.0 | 16.0 | 0.0130 | 0.0316 | 0.68 | 0.0024 |

91207 rows × 9 columns

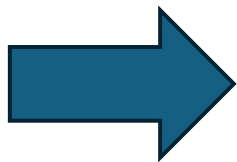
데이터 준비

• 타입 일부 변경 및 새로운 컬럼 추가

```
dust.info()
```

```
# 결측치가 일부 있다.
# 그러나 결측치를 채우는 것이 경우에 따라 왜곡을 더 심하게 할 수 있기 때문에
# 여기서는 결측치를 채우지 않는다.
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 91207 entries, 0 to 91206
Data columns (total 9 columns):
 #   Column  Non-Null Count  Dtype
---  -
0   측정일자  91207 non-null   int64
1   권역명    91207 non-null   object
2   측정소명  91207 non-null   object
3   미세먼지  90716 non-null   float64
4   초미세먼지  90759 non-null   float64
5   오존      90744 non-null   float64
6   이산화질소  90640 non-null   float64
7   일산화탄소  90312 non-null   float64
8   아황산가스  90573 non-null   float64
dtypes: float64(6), int64(1), object(2)
memory usage: 6.3+ MB
```



```
dust = dust.astype({'측정일자': 'str'})
# 측정일자를 이후에 연도와 월로도 분해할 것이기 때문에 문자열로 바꾸는 것이 슬라이싱에 용이하다.
```

```
# 측정일자 중에서 연도, 월, 연월에 해당하는 만큼을 슬라이싱해서 새로운 컬럼으로 할당한다.
```

```
dust['연도'] = dust['측정일자'].map(lambda x: x[:4])
dust['월'] = dust['측정일자'].map(lambda x: x[4:6])
dust['연월'] = dust['측정일자'].map(lambda x: x[:6])
```

```
# 미세먼지와 초미세먼지가 나쁨 이상(나쁨+매우 나쁨)인지, 매우 나쁨인지 여부를 확인하는 컬럼을 생성한다.
```

```
# 여기서 미세먼지는 나쁨 80 초과, 매우 나쁨 150 초과
# 초미세먼지는 나쁨 35 초과, 매우 나쁨 75 초과
```

```
dust['미세먼지_나쁨이상'] = dust['미세먼지'].map(lambda x: 1 if x > 80 else 0)
dust['미세먼지_매우나쁨'] = dust['미세먼지'].map(lambda x: 1 if x > 150 else 0)
dust['초미세먼지_나쁨이상'] = dust['초미세먼지'].map(lambda x: 1 if x > 35 else 0)
dust['초미세먼지_매우나쁨'] = dust['초미세먼지'].map(lambda x: 1 if x > 75 else 0)
```

| | 측정일자 | 권역명 | 측정소명 | 미세먼지 | 초미세먼지 | 오존 | 이산화질소 | 일산화탄소 | 아황산가스 | 연도 | 월 | 연월 | 미세먼지_나쁨이상 | 미세먼지_매우나쁨 | 초미세먼지_나쁨이상 | 초미세먼지_매우나쁨 |
|-------|----------|-----|------|-------|-------|--------|--------|-------|--------|------|-----|--------|-----------|-----------|------------|------------|
| 0 | 20140101 | 서남권 | 강서구 | 116.0 | 21.0 | 0.0150 | 0.0350 | 0.70 | 0.0120 | 2014 | 01 | 201401 | 1 | 0 | 0 | 0 |
| 1 | 20140101 | 서남권 | 금천구 | 113.0 | 22.0 | 0.0130 | 0.0260 | 0.60 | 0.0090 | 2014 | 01 | 201401 | 1 | 0 | 0 | 0 |
| 2 | 20140101 | 서남권 | 양천구 | 119.0 | 24.0 | 0.0110 | 0.0340 | 0.70 | 0.0110 | 2014 | 01 | 201401 | 1 | 0 | 0 | 0 |
| 3 | 20140101 | 서북권 | 은평구 | 119.0 | 28.0 | 0.0110 | 0.0420 | 0.60 | 0.0090 | 2014 | 01 | 201401 | 1 | 0 | 0 | 0 |
| 4 | 20140101 | 동북권 | 도봉구 | 119.0 | 22.0 | 0.0260 | 0.0240 | 0.60 | 0.0100 | 2014 | 01 | 201401 | 1 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 91202 | 20231231 | 도심권 | 용산구 | 22.0 | 15.0 | 0.0146 | 0.0286 | 0.78 | 0.0025 | 2023 | 12 | 202312 | 0 | 0 | 0 | 0 |
| 91203 | 20231231 | 서북권 | 은평구 | 21.0 | 18.0 | 0.0178 | 0.0172 | 1.12 | 0.0025 | 2023 | 12 | 202312 | 0 | 0 | 0 | 0 |
| 91204 | 20231231 | 도심권 | 종로구 | 22.0 | 18.0 | 0.0180 | 0.0238 | 0.73 | 0.0029 | 2023 | 12 | 202312 | 0 | 0 | 0 | 0 |
| 91205 | 20231231 | 도심권 | 중구 | 20.0 | 18.0 | 0.0179 | 0.0261 | 0.67 | 0.0028 | 2023 | 12 | 202312 | 0 | 0 | 0 | 0 |
| 91206 | 20231231 | 동북권 | 중랑구 | 20.0 | 16.0 | 0.0130 | 0.0316 | 0.68 | 0.0024 | 2023 | 12 | 202312 | 0 | 0 | 0 | 0 |

91207 rows × 16 columns

```
dust.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 91207 entries, 0 to 91206
Data columns (total 16 columns):
 #   Column  Non-Null Count  Dtype
---  -
0   측정일자  91207 non-null   object
1   권역명    91207 non-null   object
2   측정소명  91207 non-null   object
3   미세먼지  90716 non-null   float64
4   초미세먼지  90759 non-null   float64
5   오존      90744 non-null   float64
6   이산화질소  90640 non-null   float64
7   일산화탄소  90312 non-null   float64
8   아황산가스  90573 non-null   float64
9   연도      91207 non-null   object
10  월        91207 non-null   object
11  연월      91207 non-null   object
12  미세먼지_나쁨이상  91207 non-null   int64
13  미세먼지_매우나쁨  91207 non-null   int64
14  초미세먼지_나쁨이상  91207 non-null   int64
15  초미세먼지_매우나쁨  91207 non-null   int64
dtypes: float64(6), int64(4), object(6)
memory usage: 11.1+ MB
```

데이터 탐색

- 통계량 확인

```
dust.groupby('측정소명')['미세먼지'].describe()  
# 측정소명에 따른(또는 구별에 따른) 미세먼지의 통계량을 확인한다.
```

| | count | mean | std | min | 25% | 50% | 75% | max |
|------|--------|-----------|-----------|-----|------|------|------|-------|
| 측정소명 | | | | | | | | |
| 강남구 | 3641.0 | 40.024718 | 27.186190 | 0.0 | 24.0 | 35.0 | 50.0 | 565.0 |
| 강동구 | 3628.0 | 42.470232 | 28.184286 | 0.0 | 25.0 | 37.0 | 53.0 | 552.0 |
| 강북구 | 3627.0 | 39.780811 | 27.565320 | 0.0 | 24.0 | 34.0 | 49.0 | 584.0 |
| 강서구 | 3647.0 | 43.076501 | 28.627546 | 0.0 | 25.0 | 38.0 | 55.0 | 518.0 |
| 관악구 | 3596.0 | 42.973304 | 26.920226 | 0.0 | 26.0 | 38.0 | 53.0 | 528.0 |
| 광진구 | 3646.0 | 40.176632 | 27.849223 | 0.0 | 24.0 | 35.0 | 51.0 | 620.0 |
| 구로구 | 3622.0 | 42.332965 | 29.555776 | 0.0 | 24.0 | 37.0 | 54.0 | 522.0 |
| 금천구 | 3616.0 | 39.248894 | 24.988230 | 3.0 | 24.0 | 34.0 | 48.0 | 516.0 |
| 노원구 | 3645.0 | 39.304527 | 26.977549 | 0.0 | 24.0 | 34.0 | 49.0 | 609.0 |
| 도봉구 | 3647.0 | 38.790513 | 27.908979 | 0.0 | 22.0 | 34.0 | 49.0 | 629.0 |
| 동대문구 | 3617.0 | 39.928117 | 27.638085 | 0.0 | 23.0 | 35.0 | 50.0 | 558.0 |
| 동작구 | 3648.0 | 40.271382 | 26.468928 | 0.0 | 25.0 | 35.0 | 49.0 | 538.0 |
| 마포구 | 3601.0 | 40.560400 | 27.506682 | 0.0 | 24.0 | 35.0 | 50.0 | 589.0 |

```
dust.groupby('측정소명')['초미세먼지'].describe()  
# 측정소명에 따른(또는 구별에 따른) 초미세먼지의 통계량을 확인한다.
```

| | count | mean | std | min | 25% | 50% | 75% | max |
|------|--------|-----------|-----------|-----|------|------|------|-------|
| 측정소명 | | | | | | | | |
| 강남구 | 3637.0 | 22.647237 | 14.531011 | 0.0 | 13.0 | 20.0 | 29.0 | 148.0 |
| 강동구 | 3626.0 | 22.863486 | 14.289590 | 0.0 | 13.0 | 20.0 | 29.0 | 136.0 |
| 강북구 | 3628.0 | 21.448181 | 13.718780 | 0.0 | 12.0 | 19.0 | 27.0 | 147.0 |
| 강서구 | 3649.0 | 22.399287 | 14.141876 | 0.0 | 13.0 | 20.0 | 28.0 | 138.0 |
| 관악구 | 3595.0 | 23.631989 | 14.767632 | 0.0 | 13.0 | 21.0 | 30.0 | 139.0 |
| 광진구 | 3646.0 | 22.150576 | 14.675870 | 0.0 | 12.0 | 19.0 | 28.0 | 139.0 |
| 구로구 | 3633.0 | 22.037710 | 15.350062 | 0.0 | 11.0 | 19.0 | 29.0 | 133.0 |
| 금천구 | 3618.0 | 23.523217 | 14.170811 | 1.0 | 14.0 | 21.0 | 30.0 | 124.0 |
| 노원구 | 3645.0 | 22.381070 | 14.247765 | 0.0 | 13.0 | 19.0 | 28.0 | 143.0 |
| 도봉구 | 3647.0 | 21.110228 | 13.950403 | 1.0 | 11.0 | 18.0 | 27.0 | 137.0 |
| 동대문구 | 3614.0 | 21.902048 | 14.091977 | 0.0 | 12.0 | 19.0 | 28.0 | 146.0 |
| 동작구 | 3648.0 | 22.673333 | 14.073333 | 0.0 | 13.0 | 20.0 | 28.0 | 137.0 |
| 마포구 | 3601.0 | 22.673333 | 14.073333 | 0.0 | 13.0 | 20.0 | 28.0 | 137.0 |

(일부 행만 첨부)

데이터 탐색

- 통계량 확인

```
dust.groupby('연도')['미세먼지'].describe()  
# 연도에 따른 미세먼지의 통계량을 확인한다.
```

| | count | mean | std | min | 25% | 50% | 75% | max |
|------|--------|-----------|-----------|-----|-------|------|------|-------|
| 연도 | | | | | | | | |
| 2014 | 9120.0 | 45.332127 | 27.681283 | 0.0 | 27.00 | 39.0 | 58.0 | 187.0 |
| 2015 | 9039.0 | 45.343511 | 38.024276 | 0.0 | 29.00 | 39.0 | 53.0 | 709.0 |
| 2016 | 9148.0 | 47.294819 | 23.046267 | 0.0 | 33.00 | 44.0 | 59.0 | 237.0 |
| 2017 | 9125.0 | 43.668603 | 24.184661 | 0.0 | 27.00 | 39.0 | 55.0 | 243.0 |
| 2018 | 9028.0 | 39.664931 | 23.818312 | 3.0 | 22.75 | 36.0 | 51.0 | 204.0 |
| 2019 | 8912.0 | 41.764026 | 26.785521 | 3.0 | 24.00 | 36.0 | 52.0 | 228.0 |
| 2020 | 9105.0 | 35.119824 | 17.962304 | 3.0 | 22.00 | 33.0 | 46.0 | 146.0 |
| 2021 | 9078.0 | 37.899647 | 37.411182 | 3.0 | 20.00 | 30.0 | 44.0 | 503.0 |
| 2022 | 9125.0 | 32.884493 | 20.516683 | 0.0 | 21.00 | 29.0 | 40.0 | 404.0 |
| 2023 | 9036.0 | 37.734949 | 26.944100 | 3.0 | 21.00 | 32.0 | 45.0 | 319.0 |

```
dust.groupby('연도')['초미세먼지'].describe()  
# 연도에 따른 초미세먼지의 통계량을 확인한다.
```

| | count | mean | std | min | 25% | 50% | 75% | max |
|------|--------|-----------|-----------|-----|------|------|------|-------|
| 연도 | | | | | | | | |
| 2014 | 9120.0 | 23.623136 | 13.611789 | 0.0 | 14.0 | 20.0 | 30.0 | 104.0 |
| 2015 | 9039.0 | 23.113951 | 12.219640 | 0.0 | 15.0 | 21.0 | 29.0 | 104.0 |
| 2016 | 9148.0 | 25.882816 | 12.405314 | 0.0 | 18.0 | 24.0 | 33.0 | 93.0 |
| 2017 | 9125.0 | 24.615342 | 14.641222 | 0.0 | 14.0 | 22.0 | 31.0 | 110.0 |
| 2018 | 9027.0 | 22.745430 | 15.734794 | 1.0 | 11.0 | 20.0 | 30.0 | 124.0 |
| 2019 | 8922.0 | 24.934656 | 18.444453 | 1.0 | 14.0 | 21.0 | 30.0 | 153.0 |
| 2020 | 9120.0 | 20.698246 | 12.371154 | 1.0 | 12.0 | 18.0 | 27.0 | 91.0 |
| 2021 | 9082.0 | 19.774499 | 15.017943 | 1.0 | 10.0 | 16.0 | 24.0 | 123.0 |
| 2022 | 9125.0 | 18.265096 | 12.179525 | 0.0 | 10.0 | 16.0 | 23.0 | 105.0 |
| 2023 | 9051.0 | 19.660259 | 13.013382 | 1.0 | 11.0 | 17.0 | 25.0 | 116.0 |

데이터 탐색

- 통계량 확인

```
dust.groupby('월')['미세먼지'].describe()  
# 월에 따른 미세먼지의 통계량을 확인한다.
```

| | count | mean | std | min | 25% | 50% | 75% | max |
|----|--------|-----------|-----------|-----|------|------|------|-------|
| 월 | | | | | | | | |
| 01 | 7703.0 | 49.794366 | 24.397274 | 0.0 | 33.0 | 45.0 | 60.0 | 199.0 |
| 02 | 7005.0 | 51.521199 | 40.923220 | 0.0 | 32.0 | 44.0 | 63.0 | 709.0 |
| 03 | 7716.0 | 59.574650 | 32.055905 | 0.0 | 38.0 | 53.0 | 73.0 | 330.0 |
| 04 | 7495.0 | 51.381321 | 29.637865 | 0.0 | 33.0 | 48.0 | 64.0 | 319.0 |
| 05 | 7738.0 | 48.694236 | 39.180342 | 0.0 | 30.0 | 41.0 | 55.0 | 503.0 |
| 06 | 7492.0 | 34.423919 | 15.093805 | 0.0 | 24.0 | 33.0 | 43.0 | 404.0 |
| 07 | 7723.0 | 28.018128 | 14.530588 | 0.0 | 18.0 | 26.0 | 35.0 | 97.0 |
| 08 | 7693.0 | 25.013779 | 12.622713 | 0.0 | 16.0 | 24.0 | 32.0 | 99.0 |
| 09 | 7340.0 | 24.363896 | 13.196870 | 0.0 | 15.0 | 22.0 | 31.0 | 89.0 |
| 10 | 7611.0 | 31.600184 | 17.035529 | 0.0 | 21.0 | 28.0 | 38.0 | 151.0 |
| 11 | 7458.0 | 41.769777 | 21.475097 | 0.0 | 26.0 | 39.0 | 53.0 | 204.0 |
| 12 | 7742.0 | 42.101524 | 22.437391 | 0.0 | 28.0 | 38.0 | 52.0 | 255.0 |

```
dust.groupby('월')['초미세먼지'].describe()  
# 월에 따른 초미세먼지의 통계량을 확인한다.
```

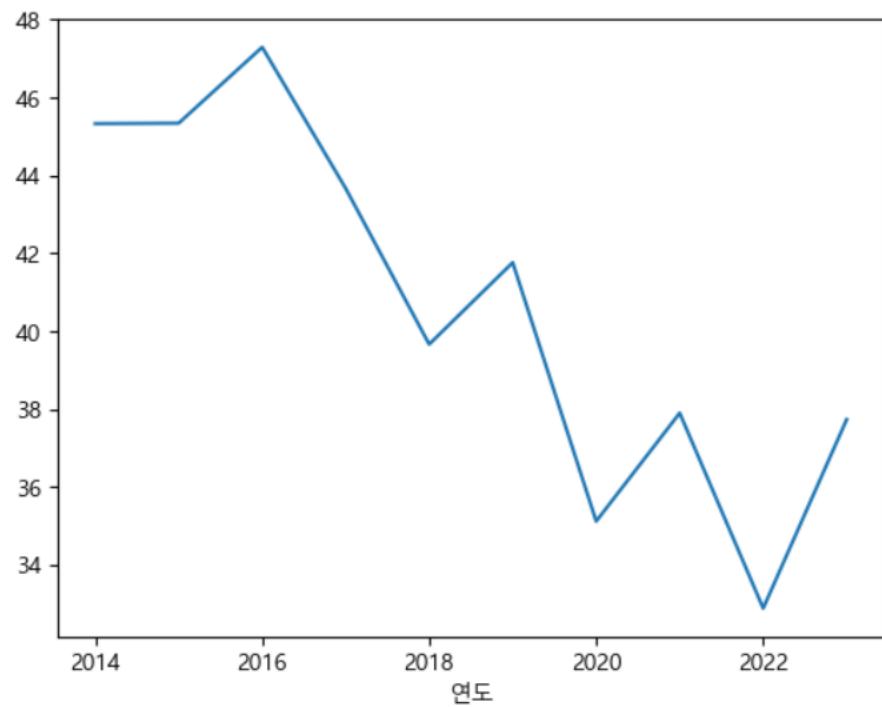
| | count | mean | std | min | 25% | 50% | 75% | max |
|----|--------|-----------|-----------|-----|------|------|------|-------|
| 월 | | | | | | | | |
| 01 | 7709.0 | 28.794396 | 17.410596 | 0.0 | 17.0 | 24.0 | 36.0 | 148.0 |
| 02 | 7013.0 | 29.035505 | 16.234555 | 0.0 | 17.0 | 25.0 | 37.0 | 104.0 |
| 03 | 7714.0 | 31.731009 | 18.857786 | 0.0 | 19.0 | 28.0 | 40.0 | 153.0 |
| 04 | 7489.0 | 23.448658 | 11.818362 | 0.0 | 15.0 | 22.0 | 30.0 | 116.0 |
| 05 | 7740.0 | 22.740956 | 12.841362 | 0.0 | 14.0 | 20.0 | 28.0 | 114.0 |
| 06 | 7499.0 | 20.494466 | 9.435738 | 0.0 | 14.0 | 20.0 | 26.0 | 68.0 |
| 07 | 7735.0 | 18.162896 | 10.503738 | 0.0 | 11.0 | 16.0 | 24.0 | 69.0 |
| 08 | 7698.0 | 15.442582 | 9.084472 | 0.0 | 8.0 | 14.0 | 20.0 | 73.0 |
| 09 | 7343.0 | 13.583549 | 9.101627 | 0.0 | 7.0 | 11.0 | 18.0 | 62.0 |
| 10 | 7616.0 | 16.515362 | 10.292672 | 0.0 | 10.0 | 14.0 | 20.0 | 81.0 |
| 11 | 7465.0 | 22.948426 | 13.480123 | 0.0 | 14.0 | 21.0 | 29.0 | 123.0 |
| 12 | 7738.0 | 25.153528 | 14.522122 | 0.0 | 15.0 | 22.0 | 33.0 | 110.0 |

데이터 모델링 & 시각화

- 앞선 데이터 탐색 파트와, 그 이외 내용 중 일부를 시각화

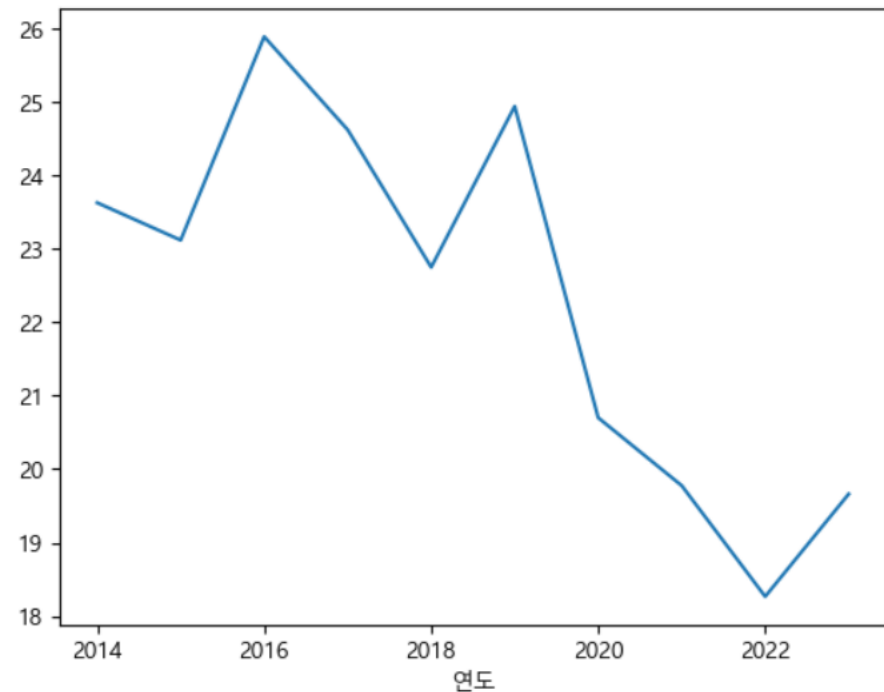
```
dust.groupby('연도')['미세먼지'].mean().plot()  
# 연도별 미세먼지의 평균을 그래프로 시각화한다.  
# 표로만 보면 차이를 한눈에 확인하기 어렵다.
```

<AxesSubplot: xlabel='연도'>



```
: dust.groupby('연도')['초미세먼지'].mean().plot()  
# 연도별 초미세먼지의 평균을 그래프로 시각화한다.
```

<AxesSubplot: xlabel='연도'>

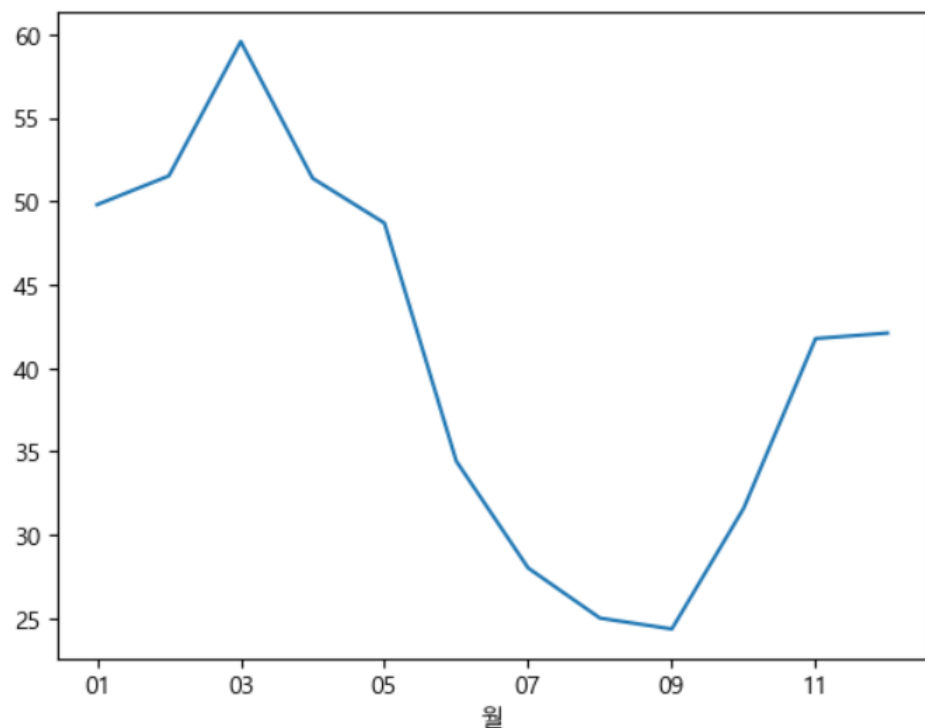


데이터 모델링 & 시각화

- 앞선 데이터 탐색 파트와, 그 이외 내용 중 일부를 시각화

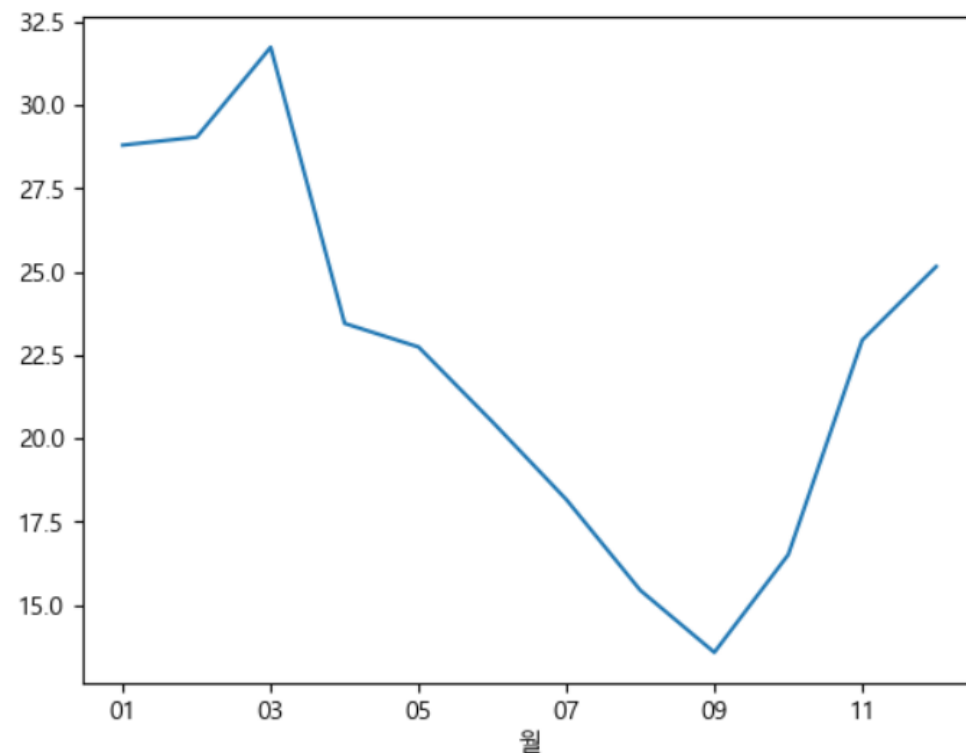
```
dust.groupby('월')['미세먼지'].mean().plot()  
# 월별 미세먼지의 평균을 그래프로 시각화한다.
```

<AxesSubplot: xlabel='월'>



```
dust.groupby('월')['초미세먼지'].mean().plot()  
# 월별 초미세먼지의 평균을 그래프로 시각화한다.
```

<AxesSubplot: xlabel='월'>

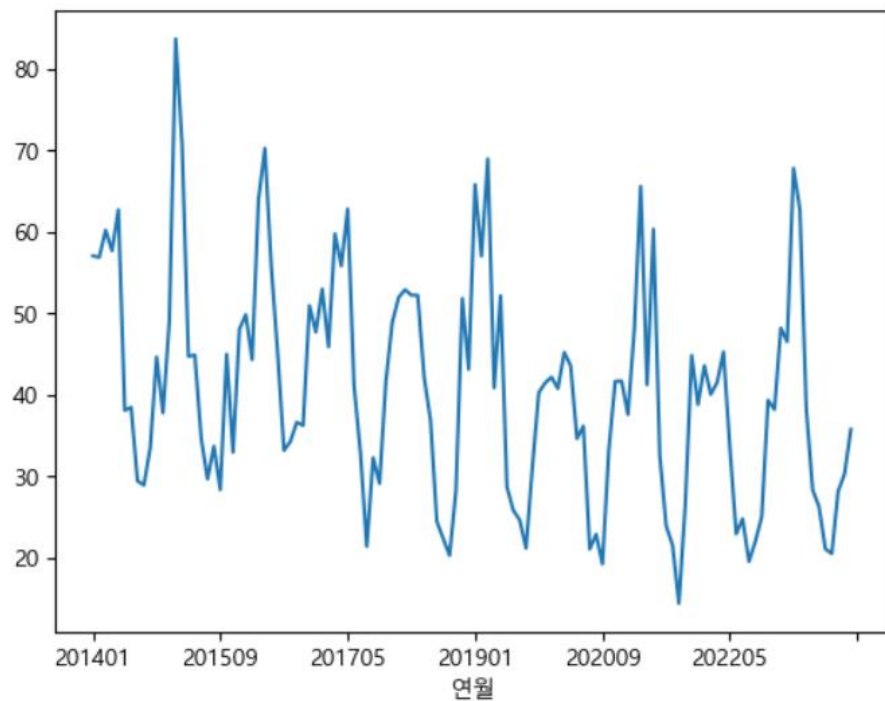


데이터 모델링 & 시각화

- 앞선 데이터 탐색 파트와, 그 이외 내용 중 일부를 시각화

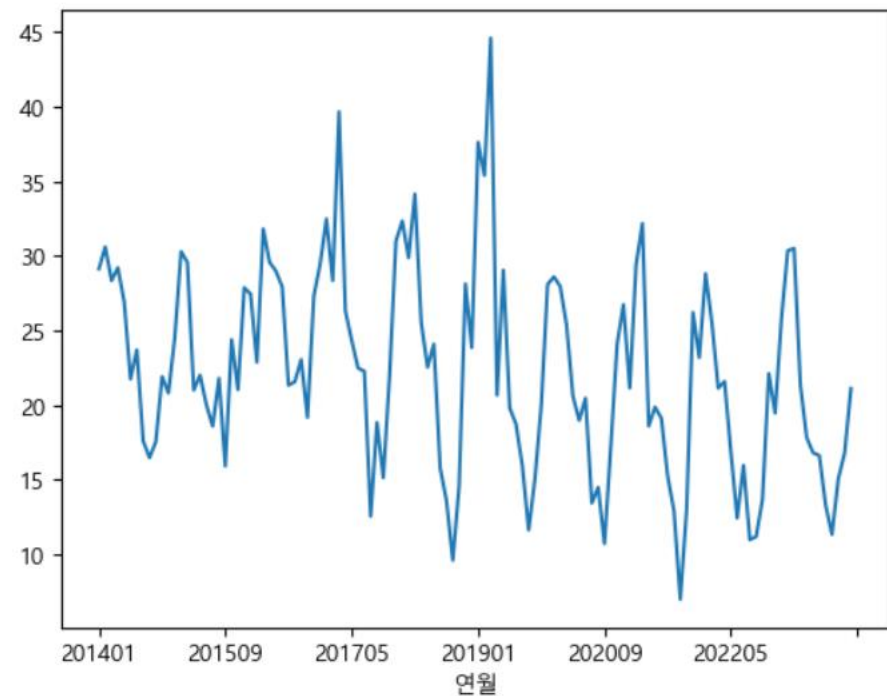
```
dust.groupby('연월')['미세먼지'].mean().plot()  
# 2014~2023년까지의 월별 미세먼지의 평균의 시간적 흐름을 그래프로 시각화한다.
```

```
<AxesSubplot: xlabel='연월'>
```



```
dust.groupby('연월')['초미세먼지'].mean().plot()  
# 2014~2023년까지의 월별 초미세먼지의 평균의 시간적 흐름을 그래프로 시각화한다.
```

```
<AxesSubplot: xlabel='연월'>
```



데이터 모델링 & 시각화

- t-검정

최근 몇 년간 우하향 추세를 보이다가 2023년에 상승하는 경향이 있다.

- Q. 2022년과 2023년 사이에는 농도에 대한 유의미한 차이가 있는가?

```
from scipy import stats
```

```
mise_2022 = dust[(dust['연도']=='2022') & (dust['미세먼지'].notnull())]['미세먼지']
```

```
# 2022년 미세먼지 데이터 중 결측치가 아닌 것만 추출
```

```
mise_2023 = dust[(dust['연도']=='2023') & (dust['미세먼지'].notnull())]['미세먼지']
```

```
# 2023년 미세먼지 데이터 중 결측치가 아닌 것만 추출
```

```
stats.ttest_ind(mise_2022, mise_2023, equal_var=False) # 등분산성이 있는지 보장하기 어려우므로 다르다고 생각하고 진행
```

```
Ttest_indResult(statistic=-13.639016126163611, pvalue=3.9253009046519093e-42)
```

```
chomise_2022 = dust[(dust['연도']=='2022') & (dust['초미세먼지'].notnull())]['초미세먼지']
```

```
# 2022년 초미세먼지 데이터 중 결측치가 아닌 것만 추출
```

```
chomise_2023 = dust[(dust['연도']=='2023') & (dust['초미세먼지'].notnull())]['초미세먼지']
```

```
# 2023년 초미세먼지 데이터 중 결측치가 아닌 것만 추출
```

```
stats.ttest_ind(chomise_2022, chomise_2023, equal_var=False) # 등분산성이 있는지 보장하기 어려우므로 다르다고 생각하고 진행
```

```
Ttest_indResult(statistic=-7.460980544034909, pvalue=8.977085479058821e-14)
```

데이터 모델링 & 시각화

- t-검정

월별로 보았을 때, 3월에 가장 높은 경향성을 보인다.

- Q. 2022년 3월과 2023년 3월 사이에는 농도에 대한 유의미한 차이가 있는가?

```
mise_202203 = dust[(dust['연도']=='2022') & (dust['미세먼지'].notnull())]['미세먼지']
mise_202303 = dust[(dust['연도']=='2023') & (dust['미세먼지'].notnull())]['미세먼지']

stats.ttest_ind(mise_202203, mise_202303, equal_var=False) # 등분산성이 있는지 보장하기 어려우므로 다르다고 생각하고 진행

Ttest_indResult(statistic=-13.639016126163611, pvalue=3.9253009046519093e-42)
```

```
chomise_202203 = dust[(dust['연도']=='2022') & (dust['초미세먼지'].notnull())]['초미세먼지']
chomise_202303 = dust[(dust['연도']=='2023') & (dust['초미세먼지'].notnull())]['초미세먼지']

stats.ttest_ind(chomise_202203, chomise_202303, equal_var=False) # 등분산성이 있는지 보장하기 어려우므로 다르다고 생각하고 진행

Ttest_indResult(statistic=-7.460980544034909, pvalue=8.977085479058821e-14)
```

데이터 모델링 & 시각화

- 구별, 연도별 미세먼지, 초미세먼지 평균 비교

```
mise_place_year = dust.groupby(['측정소명', '연도']).mean()['미세먼지'].unstack()
# 각 측정소마다 연도별 미세먼지 농도의 평균을 나타낸다.
mise_place_year
```

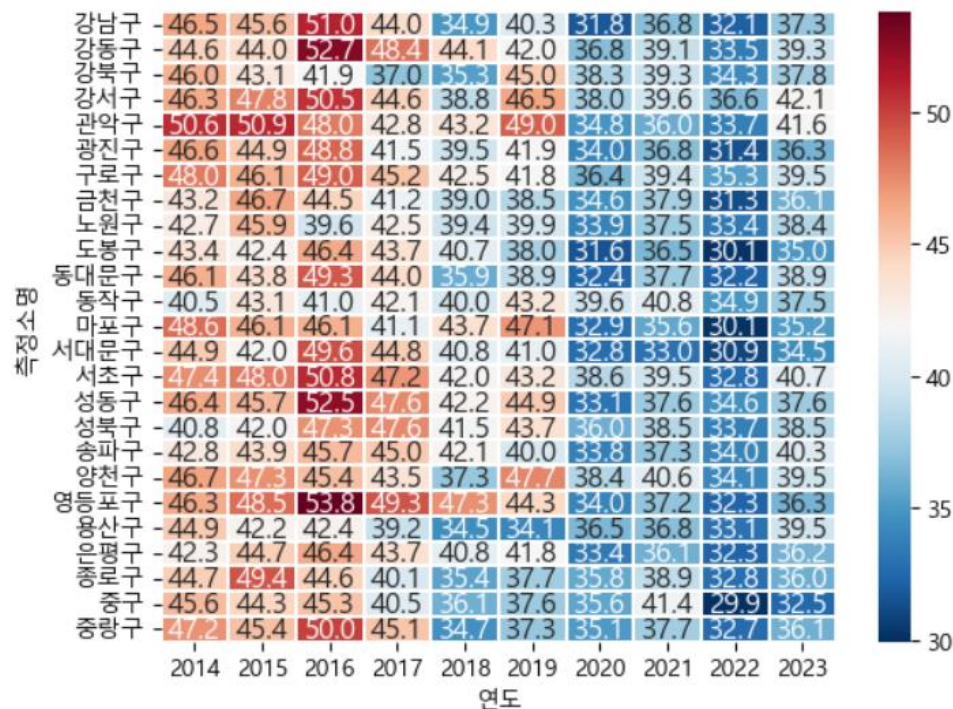
| | 연도 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
|------|----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 측정소명 | | | | | | | | | | | |
| 강남구 | | 46.506849 | 45.575342 | 51.019126 | 43.958904 | 34.947945 | 40.292958 | 31.778689 | 36.750000 | 32.147945 | 37.260274 |
| 강동구 | | 44.627397 | 44.049315 | 52.672131 | 48.380822 | 44.057534 | 42.013850 | 36.811475 | 39.106849 | 33.523288 | 39.266667 |
| 강북구 | | 46.019231 | 43.125731 | 41.855191 | 37.032877 | 35.328767 | 44.950685 | 38.309589 | 39.265753 | 34.320548 | 37.821918 |
| 강서구 | | 46.315068 | 47.786301 | 50.467213 | 44.589041 | 38.799451 | 46.504132 | 37.975410 | 39.556474 | 36.619178 | 42.134247 |
| 관악구 | | 50.646575 | 50.887671 | 48.030055 | 42.802740 | 43.232877 | 48.974359 | 34.786885 | 35.950549 | 33.663014 | 41.611570 |
| 광진구 | | 46.594521 | 44.871233 | 48.819672 | 41.495890 | 39.541667 | 41.879452 | 34.008197 | 36.789041 | 31.430137 | 36.310440 |
| 구로구 | | 48.038356 | 46.079452 | 48.969945 | 45.238356 | 42.509859 | 41.810585 | 36.366197 | 39.365385 | 35.260274 | 39.482094 |
| 금천구 | | 43.224658 | 46.684524 | 44.513661 | 41.249315 | 39.046575 | 38.454545 | 34.592287 | 37.853994 | 31.317808 | 36.090411 |
| 노원구 | | 42.709589 | 45.851240 | 39.625683 | 42.452055 | 39.353425 | 39.887363 | 33.853591 | 37.509589 | 33.400000 | 38.394521 |
| 도봉구 | | 43.416438 | 42.380822 | 46.418033 | 43.690411 | 40.660274 | 37.986301 | 31.603306 | 36.509589 | 30.142466 | 35.016529 |
| 동대문구 | | 46.052055 | 43.835616 | 49.273224 | 44.005479 | 35.928767 | 38.924471 | 32.358904 | 37.660274 | 32.191781 | 38.931507 |
| 동작구 | | 40.539726 | 43.073973 | 40.975410 | 42.147945 | 39.972603 | 43.195055 | 39.648352 | 40.788462 | 34.868493 | 37.509589 |
| 마포구 | | 48.553425 | 46.068493 | 46.073770 | 41.095890 | 43.717877 | 47.086154 | 32.868852 | 35.622590 | 30.071233 | 35.170799 |
| 서대문구 | | 44.928767 | 42.030137 | 49.565574 | 44.783562 | 40.806647 | 40.958904 | 32.759689 | 33.030641 | 30.917808 | 34.491667 |

표로는 한눈에 식별하기 어려우므로 시각화를 통해 나타낸다.

```
colormap = plt.cm.RdBu_r
# 빨간색이 높은 수치, 파란색이 낮은 수치가 되도록
```

```
import seaborn as sns
```

```
sns.heatmap(mise_place_year, linewidths=0.1, cmap=colormap, # 미세먼지
            linecolor='white', annot=True, fmt='.1f', annot_kws={'size':12}) # 소수점 아래 1자리
plt.figure(figsize=(30, 20))
plt.show()
```



데이터 모델링 & 시각화

- 구별, 연도별 미세먼지, 초미세먼지 평균 비교

```
chomise_place_year = dust.groupby(['측정소명', '연도']).mean()['초미세먼지'].unstack()
# 각 측정소마다 연도별 초미세먼지 농도의 평균을 나타낸다.
chomise_place_year
```

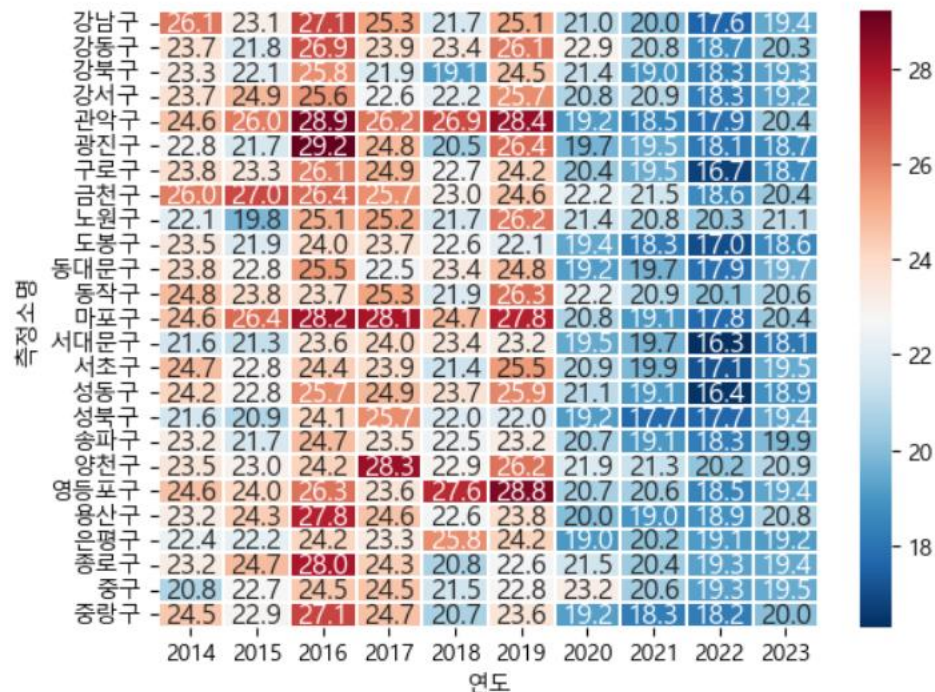
| 연도 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
|------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 측정소명 | | | | | | | | | | |
| 강남구 | 26.139726 | 23.106849 | 27.120219 | 25.293151 | 21.715068 | 25.098592 | 20.994490 | 20.038567 | 17.605479 | 19.391781 |
| 강동구 | 23.717808 | 21.753425 | 26.937158 | 23.893151 | 23.380822 | 26.136872 | 22.948087 | 20.820937 | 18.704110 | 20.261494 |
| 강북구 | 23.258242 | 22.055556 | 25.751366 | 21.865753 | 19.123288 | 24.498630 | 21.371585 | 18.991781 | 18.331507 | 19.265753 |
| 강서구 | 23.742466 | 24.857534 | 25.581967 | 22.649315 | 22.197260 | 25.738292 | 20.811475 | 20.917582 | 18.271233 | 19.235616 |
| 관악구 | 24.589041 | 26.049315 | 28.939891 | 26.197260 | 26.945205 | 28.360129 | 19.153005 | 18.483516 | 17.915068 | 20.352617 |
| 광진구 | 22.786301 | 21.720548 | 29.207650 | 24.800000 | 20.469444 | 26.425824 | 19.748634 | 19.501370 | 18.142466 | 18.679452 |
| 구로구 | 23.819178 | 23.320548 | 26.071038 | 24.923288 | 22.687324 | 24.206044 | 20.367403 | 19.537396 | 16.742466 | 18.673973 |
| 금천구 | 26.043836 | 27.020833 | 26.423497 | 25.706849 | 22.972603 | 24.579670 | 22.214286 | 21.542700 | 18.632877 | 20.353425 |
| 노원구 | 22.101370 | 19.790634 | 25.122951 | 25.227397 | 21.666667 | 26.170330 | 21.441096 | 20.849315 | 20.309589 | 21.112637 |
| 도봉구 | 23.484932 | 21.939726 | 23.959016 | 23.652055 | 22.641096 | 22.120548 | 19.397790 | 18.260274 | 17.010959 | 18.607143 |
| 동대문구 | 23.821918 | 22.830137 | 25.502732 | 22.531507 | 23.353425 | 24.788520 | 19.170330 | 19.741047 | 17.863014 | 19.657534 |
| 동작구 | 24.805479 | 23.800000 | 23.726776 | 25.290411 | 21.936986 | 26.344353 | 22.238356 | 20.873973 | 20.117808 | 20.646409 |
| 마포구 | 24.594521 | 26.386301 | 28.150273 | 28.065753 | 24.737430 | 27.778116 | 20.781421 | 19.060440 | 17.750685 | 20.355372 |
| 서대문구 | 21.589041 | 21.279452 | 23.581967 | 23.978082 | 23.383686 | 23.167123 | 19.465565 | 19.731302 | 16.298630 | 18.117808 |

표로는 한눈에 식별하기 어려우므로 시각화를 통해 나타낸다.

```
colormap = plt.cm.RdBu_r
# 빨간색이 높은 수치, 파란색이 낮은 수치가 되도록
```

```
import seaborn as sns
```

```
sns.heatmap(chomise_place_year, linewidths=0.1, cmap=colormap, # 초미세먼지
            linecolor='white', annot=True, fmt='.1f', annot_kws={'size':12}) # 소수점 아래 1자리
plt.figure(figsize=(30, 20))
plt.show()
```



데이터 모델링 & 시각화

- 구별 미세먼지, 초미세먼지 나쁨 이상 또는 매우 나쁨 비율 비교

```
mise_bad = dust.groupby('측정소명')['미세먼지_나쁨이상'].mean().sort_values(ascending=False) * 100
mise_verybad = dust.groupby('측정소명')['미세먼지_매우나쁨'].mean().sort_values(ascending=False) * 100
chomise_bad = dust.groupby('측정소명')['초미세먼지_나쁨이상'].mean().sort_values(ascending=False) * 100
chomise_verybad = dust.groupby('측정소명')['초미세먼지_매우나쁨'].mean().sort_values(ascending=False) * 100
# 조건에 부합하면 1, 아니면 0으로 입력되어 있다.
# 각각에 대해 평균을 구하고 100을 곱하면 백분율로 나타낼 수 있다.
# 내림차순으로 정렬한다.
```

```
mise_bad_df = mise_bad.to_frame().reset_index()
mise_verybad_df = mise_verybad.to_frame().reset_index()
chomise_bad_df = chomise_bad.to_frame().reset_index()
chomise_verybad_df = chomise_verybad.to_frame().reset_index()
# 추후 Folium Choropleth 지도에 이용하기 위해 reset_index()를 이용해 측정소명을 컬럼으로 올린다.
```

mise_bad_df

| | 측정소명 | 미세먼지_나쁨이상 |
|----|------|-----------|
| 0 | 서초구 | 7.447974 |
| 1 | 구로구 | 7.228916 |
| 2 | 영등포구 | 7.097835 |
| 3 | 성동구 | 7.037240 |
| 4 | 강동구 | 6.927711 |
| 5 | 강서구 | 6.845564 |
| 6 | 양천구 | 6.738470 |
| 7 | 관악구 | 6.380066 |
| 8 | 서대문구 | 6.133625 |
| 9 | 마포구 | 6.078861 |
| 10 | 성북구 | 6.078861 |

```
import geopandas as gpd
```

```
kr_sido = gpd.read_file('sig_20230729', encoding='cp949') # 시군구 shp 데이터를 받아서 사용
seoul_gu = kr_sido[:25] # 맨 앞 25행만 서울시 데이터이므로 이것만 사용
seoul_gu
```

| | SIG_CD | SIG_ENG_NM | SIG_KOR_NM | geometry |
|---|--------|---------------|------------|---|
| 0 | 11110 | Jongno-gu | 종로구 | POLYGON ((956615.453 1953567.199, 956621.579 1... |
| 1 | 11140 | Jung-gu | 중구 | POLYGON ((957890.386 1952616.746, 957909.908 1... |
| 2 | 11170 | Yongsan-gu | 용산구 | POLYGON ((953115.761 1950834.084, 953114.206 1... |
| 3 | 11200 | Seongdong-gu | 성동구 | POLYGON ((959681.109 1952649.605, 959842.412 1... |
| 4 | 11215 | Gwangjin-gu | 광진구 | POLYGON ((964825.058 1952633.250, 964875.565 1... |
| 5 | 11230 | Dongdaemun-gu | 동대문구 | POLYGON ((962141.919 1956519.461, 962149.853 1... |
| 6 | 11260 | Junngang-gu | 중랑구 | POLYGON ((965698.531 1957987.355, 965772.290 1... |
| 7 | 11290 | Seongbuk-gu | 성북구 | POLYGON ((956787.318 1953542.506, 956761.258 1... |
| 8 | 11305 | Gangbuk-gu | 강북구 | POLYGON ((956316.789 1965175.345, 956401.551 1... |

Folium과 좌표계를 맞추는 과정 진행

```
seoul_gu.crs='epsg:5179'
seoul_gu.to_crs(epsg=4326).head()
```

| | SIG_CD | SIG_ENG_NM | SIG_KOR_NM | geometry |
|---|--------|--------------|------------|---|
| 0 | 11110 | Jongno-gu | 종로구 | POLYGON ((127.00864 37.58047, 127.00871 37.580... |
| 1 | 11140 | Jung-gu | 중구 | POLYGON ((127.02314 37.57196, 127.02336 37.571... |
| 2 | 11170 | Yongsan-gu | 용산구 | POLYGON ((126.96918 37.55566, 126.96917 37.554... |
| 3 | 11200 | Seongdong-gu | 성동구 | POLYGON ((127.04341 37.57234, 127.04524 37.571... |
| 4 | 11215 | Gwangjin-gu | 광진구 | POLYGON ((127.10166 37.57240, 127.10224 37.572... |

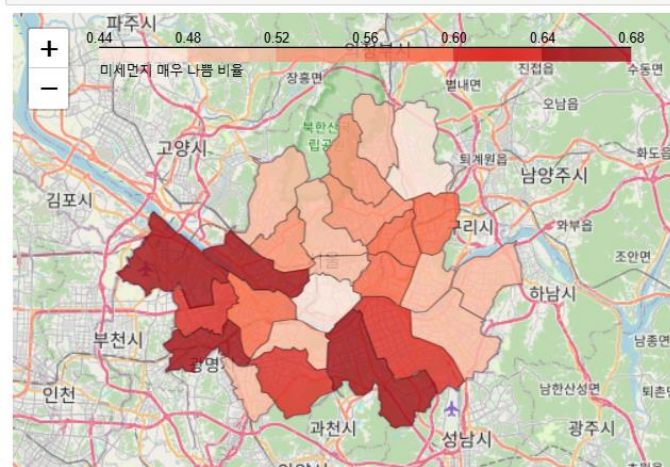
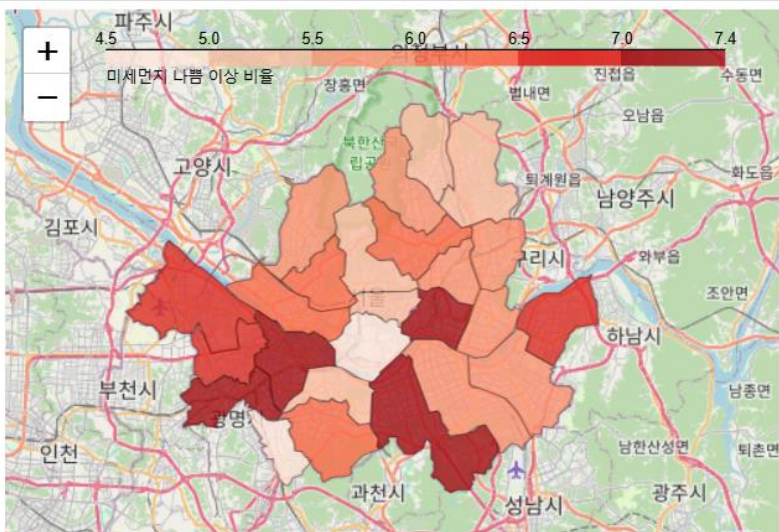
데이터 모델링 & 시각화

- 구별 미세먼지, 초미세먼지 나쁨 이상 또는 매우 나쁨 비율 비교

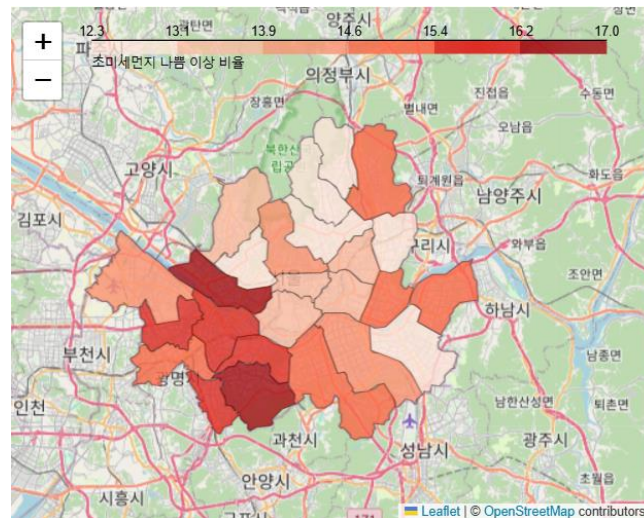
```
import folium

m = folium.Map(location=[37.55, 127], zoom_start=10, width=500, height=400)
folium.Choropleth(
    geo_data=seoul_gu,
    data=mise_bad_df,
    columns=['측정소명', "미세먼지_나쁨이상"],
    key_on="feature.properties.SIG_KOR_NM",
    fill_color="Reds", # 빨간색으로 표시
    fill_opacity=0.8,
    line_opacity=0.3,
    legend_name="미세먼지 나쁨 이상 비율"
).add_to(m)
m
```

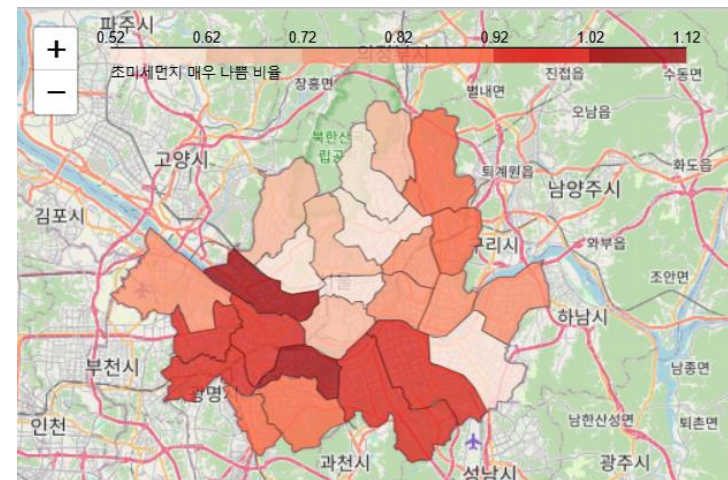
미세먼지 나쁨 이상 비율을 구별로 나타내기 위한 시각화 과정



미세먼지 매우 나쁨



초미세먼지 나쁨 이상



초미세먼지 매우 나쁨

데이터 모델링 & 시각화

- 구별(+연도별) 미세먼지, 초미세먼지 나쁨 이상 또는 매우 나쁨 비율 비교

```

mise_bad_total = dust.groupby(['측정소명', '연도']).mean()['미세먼지_나쁨이상'].unstack() * 100 # 미세먼지, 나쁨 이상
mise_verybad_total = dust.groupby(['측정소명', '연도']).mean()['미세먼지_매우나쁨'].unstack() * 100 # 미세먼지, 매우 나쁨
chomise_bad_total = dust.groupby(['측정소명', '연도']).mean()['초미세먼지_나쁨이상'].unstack() * 100 # 초미세먼지, 나쁨 이상
chomise_verybad_total = dust.groupby(['측정소명', '연도']).mean()['초미세먼지_매우나쁨'].unstack() * 100 # 초미세먼지, 매우 나쁨

# 결측치 없는 비율이 연도와 구마다 약간씩 다르므로 비율로 계산하는 게 정확하다고 판단하여 그렇게 계산함.

mise_bad_total|
    
```

| 연도 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
|------|-----------|----------|-----------|----------|----------|-----------|----------|----------|----------|----------|
| 측정소명 | | | | | | | | | | |
| 강남구 | 10.136986 | 7.123288 | 10.928962 | 6.027397 | 4.931507 | 6.301370 | 0.546448 | 5.479452 | 1.917808 | 5.479452 |
| 강동구 | 8.219178 | 6.849315 | 10.928962 | 8.493151 | 8.493151 | 7.397260 | 2.459016 | 7.397260 | 1.917808 | 7.123288 |
| 강북구 | 9.615385 | 7.309942 | 4.918033 | 3.835616 | 5.205479 | 9.315068 | 1.366120 | 6.027397 | 2.739726 | 6.027397 |
| 강서구 | 9.041096 | 8.219178 | 7.650273 | 5.753425 | 5.479452 | 10.136986 | 4.371585 | 7.123288 | 2.465753 | 8.219178 |
| 관악구 | 13.150685 | 9.315068 | 4.644809 | 4.383562 | 7.123288 | 9.315068 | 0.819672 | 5.205479 | 1.643836 | 8.219178 |
| 광진구 | 12.054795 | 6.575342 | 8.196721 | 5.479452 | 5.479452 | 7.945205 | 0.819672 | 5.479452 | 1.643836 | 5.205479 |
| 구로구 | 10.684932 | 7.671233 | 9.289617 | 6.575342 | 7.671233 | 7.945205 | 2.732240 | 7.123288 | 3.835616 | 8.767123 |
| 금천구 | 6.849315 | 7.440476 | 4.371585 | 4.383562 | 4.931507 | 4.931507 | 1.092896 | 5.753425 | 1.917808 | 4.931507 |
| 노원구 | 9.041096 | 8.264463 | 3.551913 | 5.753425 | 5.479452 | 5.479452 | 1.366120 | 5.753425 | 2.191781 | 6.301370 |

표로는 한눈에 식별하기 어려우므로 시각화를 통해 나타낸다.

```

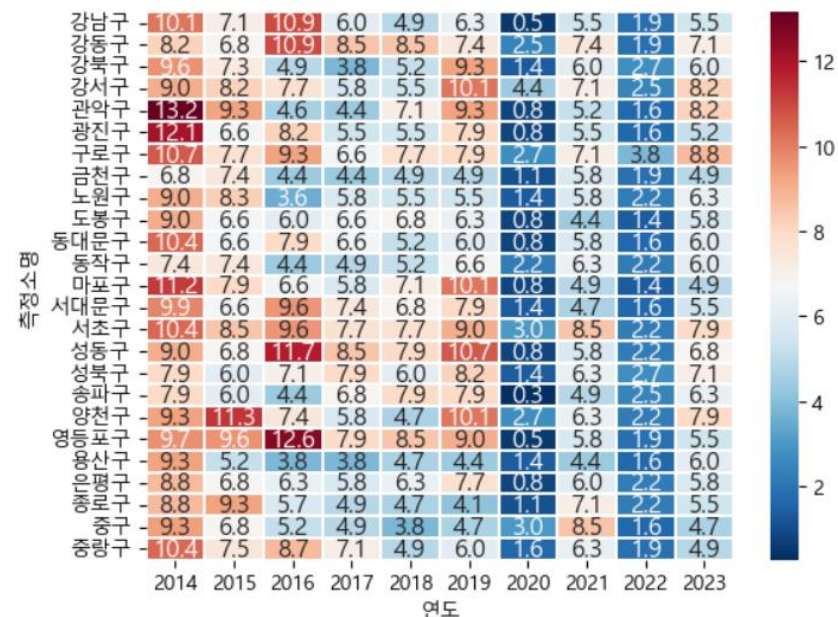
colormap = plt.cm.RdBu_r
# 빨간색이 높은 수치, 파란색이 낮은 수치가 되도록
    
```

```

import seaborn as sns
    
```

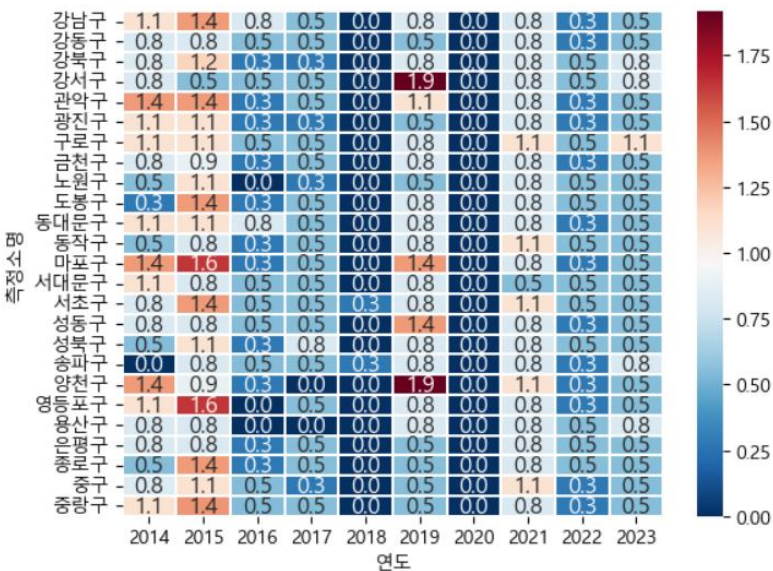
```

sns.heatmap(mise_bad_total, linewidths=0.1, cmap=colormap, # 미세먼지, 나쁨 이상
            linecolor='white', annot=True, fmt='.1f', annot_kws={'size':12}) # 소수점 이하 1자리.
plt.figure(figsize=(30, 20))
plt.show()
    
```

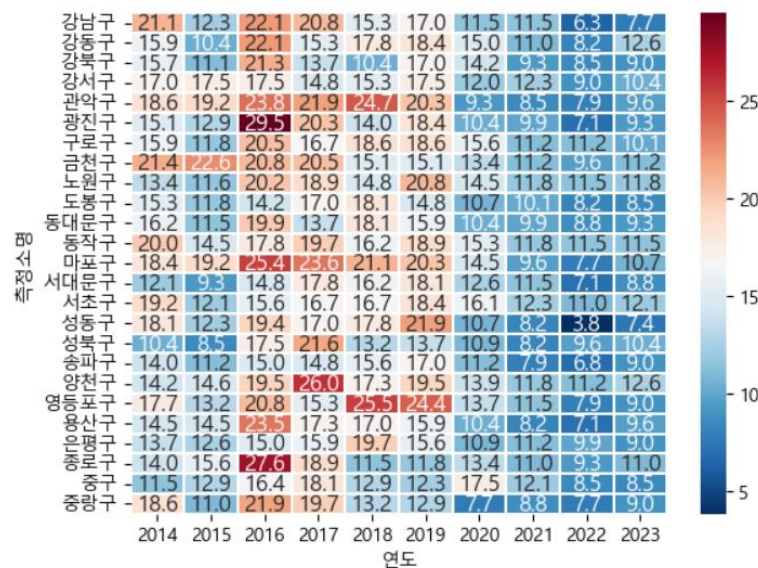


데이터 모델링 & 시각화

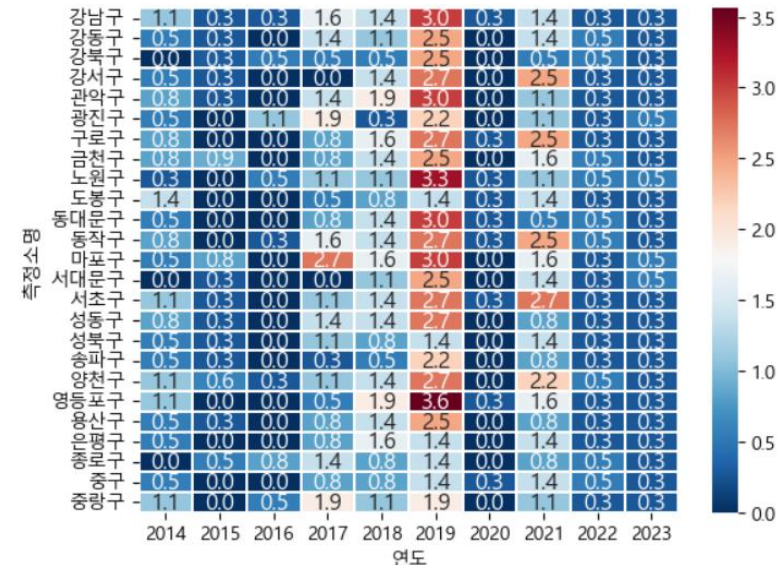
- 구별(+연도별) 미세먼지, 초미세먼지 나쁨 이상 또는 매우 나쁨 비율 비교



미세먼지 매우 나쁨



초미세먼지 나쁨 이상



초미세먼지 매우 나쁨

데이터 모델링 & 시각화

- 상관분석

```
# 미세먼지와 초미세먼지, 그리고 다른 대기 오염 물질의 상관관계를 파악하기 위한 히트맵 작성

colormap = plt.cm.RdBu
# 색이 진할수록 절대값이 크며, 빨간색은 음수, 파란색은 양수
heatmap_data = dust[['미세먼지', '초미세먼지', '오존', '이산화질소', '일산화탄소', '아황산가스']]

import seaborn as sns

sns.heatmap(heatmap_data.corr(), linewidths=0.1, vmax=1.0, cmap=colormap,
            linecolor='white', annot=True, annot_kws={'size':12})
plt.figure(figsize=(30, 20))
plt.show()
```

