

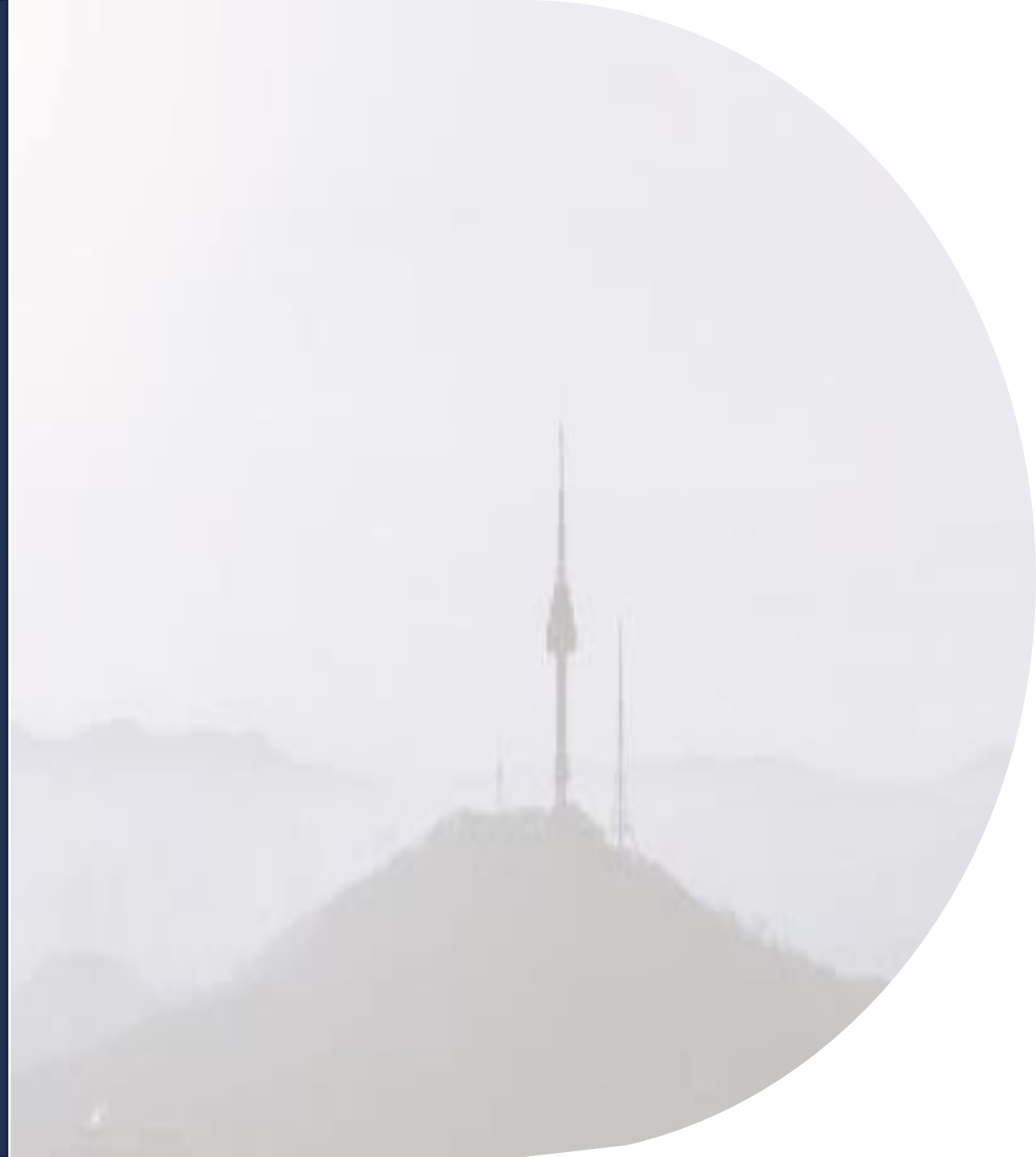
지하역사 미세먼지와 외부 환경의 상관관계 탐색

Contents

- 1** 분석 개요
- 2** 데이터 구축 및 분석 방법
- 3** 분석 결과 및 시각화
- 4** 결론

Part 1

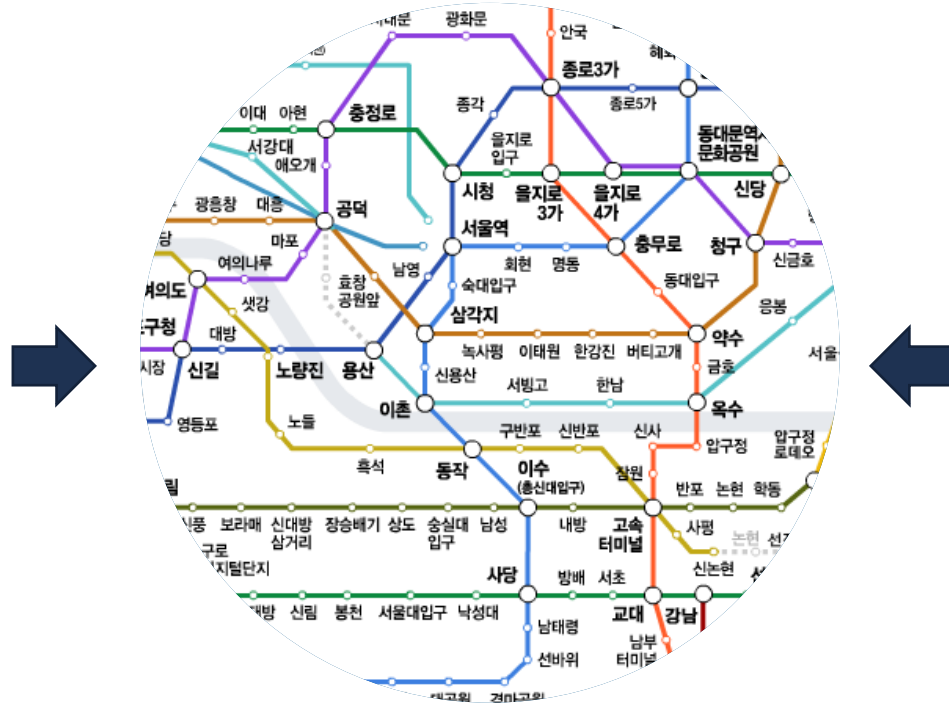
분석개요



분석의 배경 및 목적



외부 미세먼지



지하역사 미세먼지



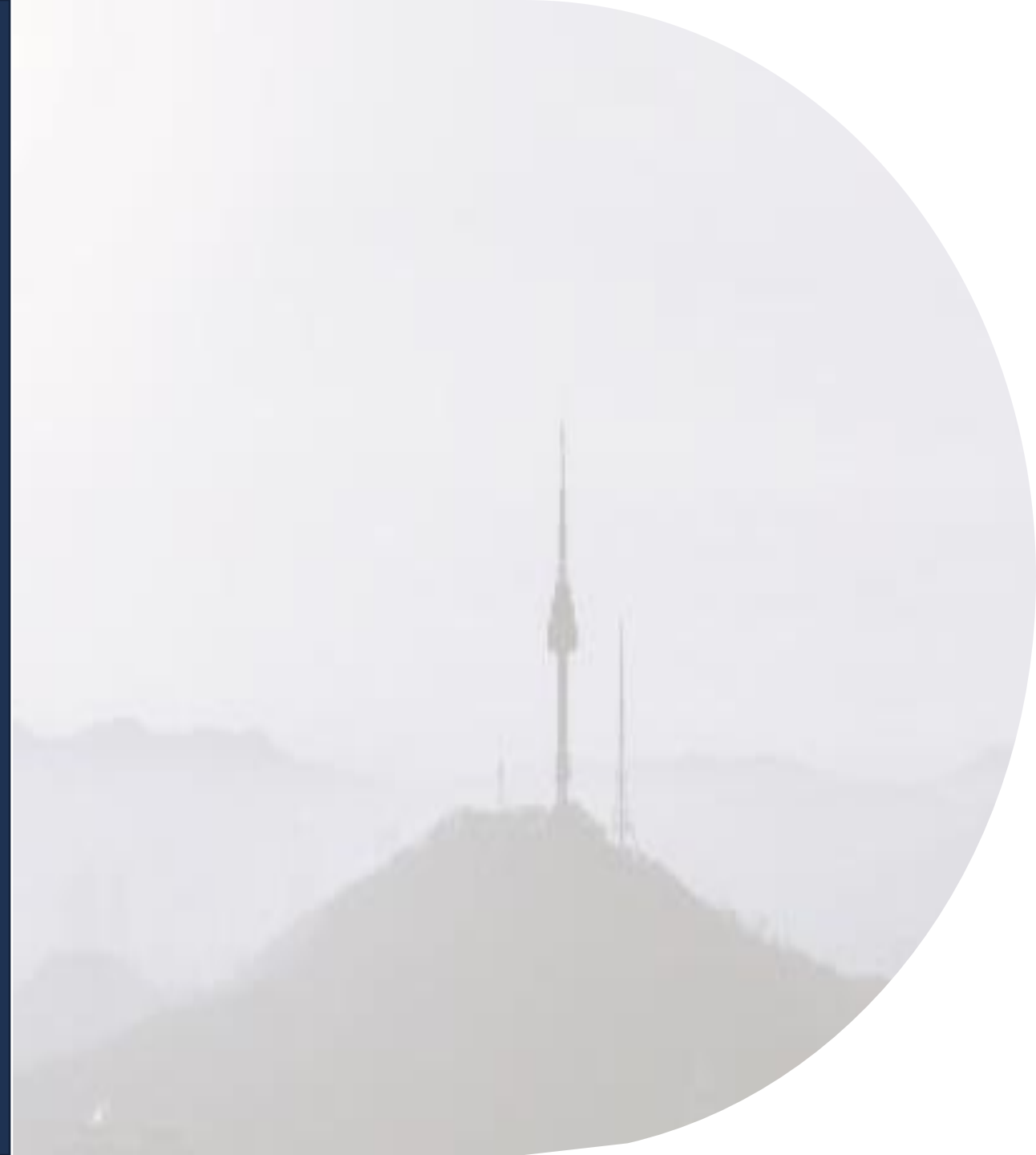
지하철 유동인구

EDA를 통한 외부 미세먼지, 지하철 유동인구와 지하역사 미세먼지 상관관계 탐색

분석대상	시간	공간	분석1 (외부미세먼지)	분석2 (유동인구)
서울 자치구별 미세먼지	2022년 8월 ~ 2023년 8월	서울 지하철 1~9호선	독립변수	통제변수
지하철 유동인구			통제변수	독립변수
지하역사 미세먼지			종속변수	종속변수

Part 2

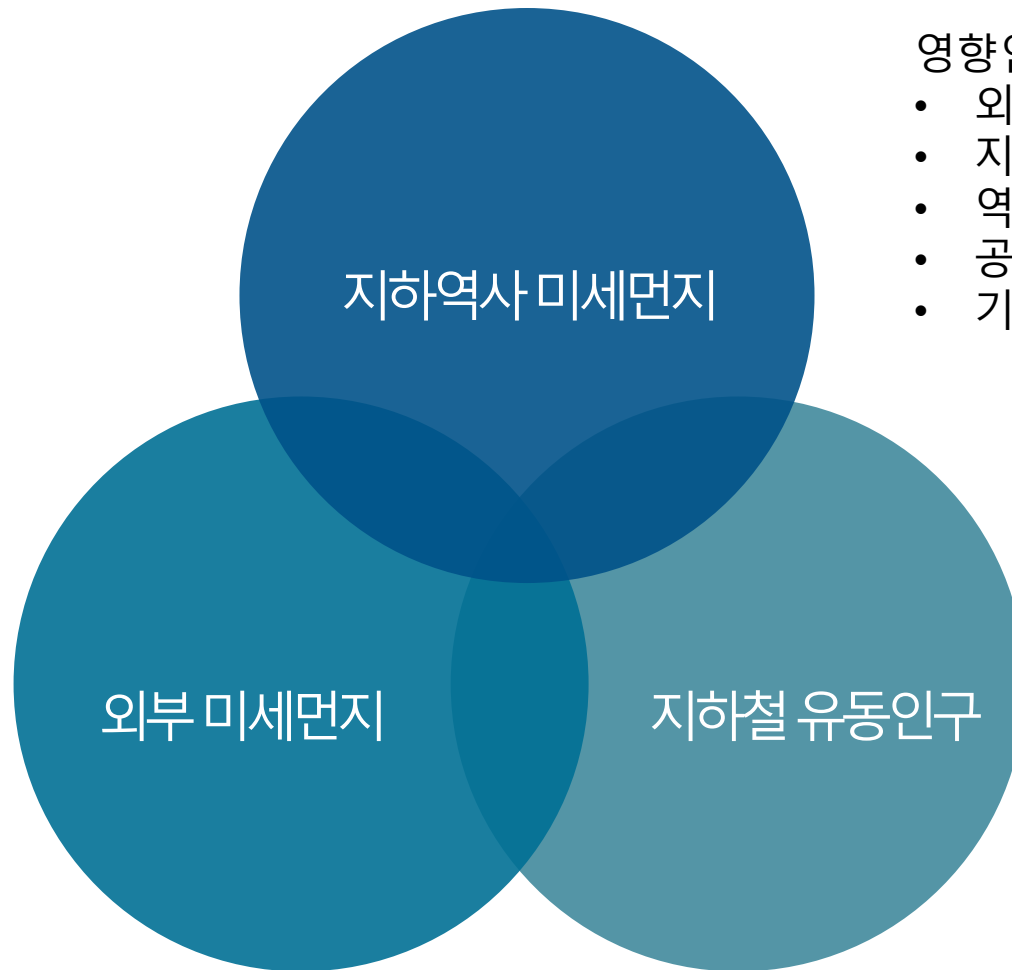
데이터 구축 및 분석



가설설정

통제요인

- 지하철 유동인구



영향인자

- 외부미세먼지
- 지하철유동인구
- 역사 주변 오염 배출원 유무
- 공기정화장치 설치 유무 및 성능
- 기타

통제요인

- 외부 미세먼지(동일 날짜)
- 날씨

STEP 1

데이터 수집

- 미세먼지
- 지하철 이용객 수

>>

STEP 2

데이터 전처리

- 결측치, 이상치
- 데이터 병합
- 변수 타입 변환

>>

STEP 3

데이터 분석

- 기초통계량
- 시각화

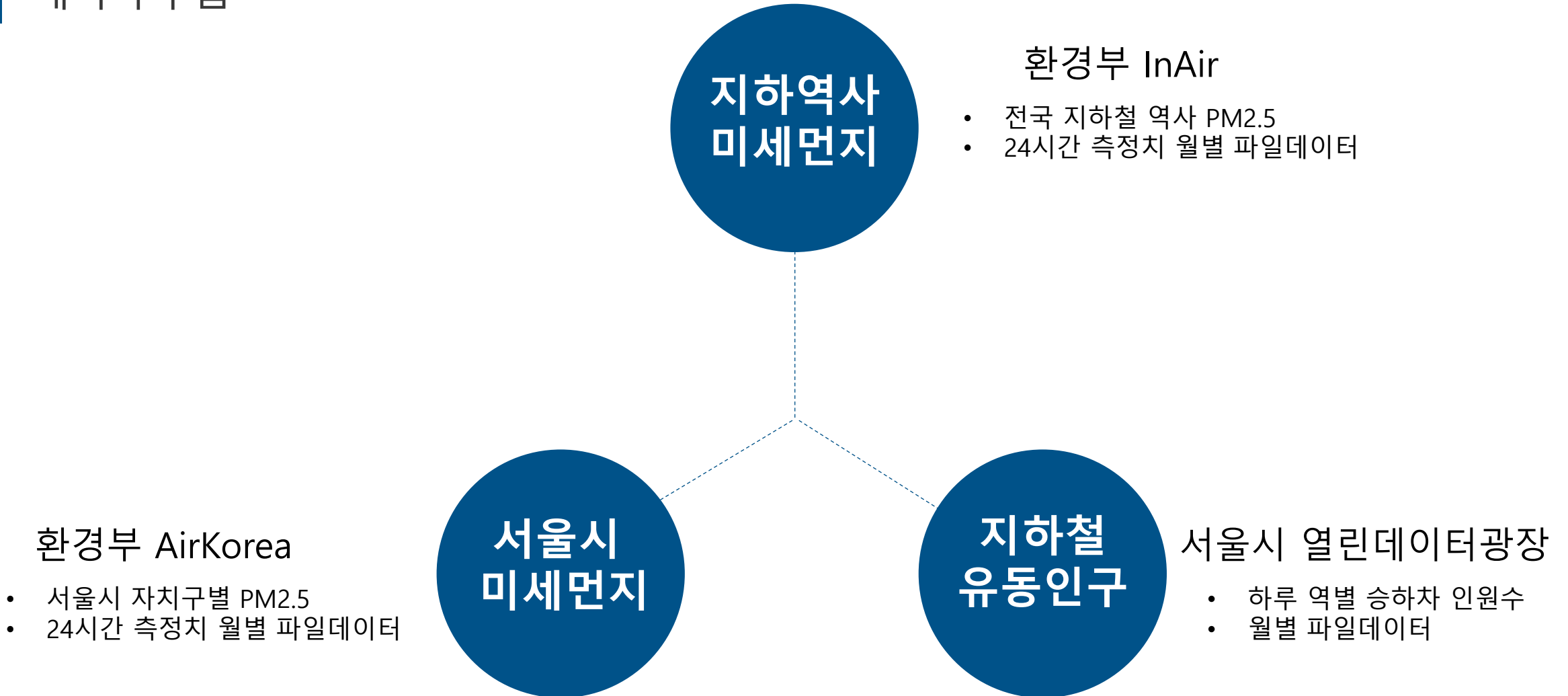
>>

STEP 4

결론

- 인사이트
- 한계점

데이터 수집



데이터 전처리

01 결측치, 이상치

```
지점명      0
호선        0
측정일자    0
일평균      38
dtype: int64
```

지하역사 미세먼지 데이터에서 38개 행의 결측치 존재

-> 총 126,228개 데이터 중 38개 제거해도 큰 영향 없을 것이라 생각하여 제거

02 파생변수

```
df6['일별평균'] = df6.iloc[:, :24].mean(axis=1)
df6.to_excel('서울미세먼지외부_2210.xlsx')
```

외부미세먼지 데이터는 시간당 측정치이기 때문에 일별평균 파생변수 생성

03 형식변환 및 통일

```
# 날짜 변환 함수 정의
def convert_date_string(date_str):
    # 마지막 두 글자가 '24'일 경우에만 처리
```

날짜 형식으로 변환하고 지하역사 이름 통일시키기

04 데이터 병합

- 전처리 데이터 병합

- 1) 서울지역 외부미세먼지
- 2) 서울 지하철사 미세먼지
- 3) 서울 지하철 유동인구

기초통계량

외부 미세먼지

일별평균	
count	10149.000000
mean	18.869839
std	12.815298
min	1.000000
25%	10.000000
50%	16.000000
75%	24.000000
max	102.000000

지하역사 미세먼지

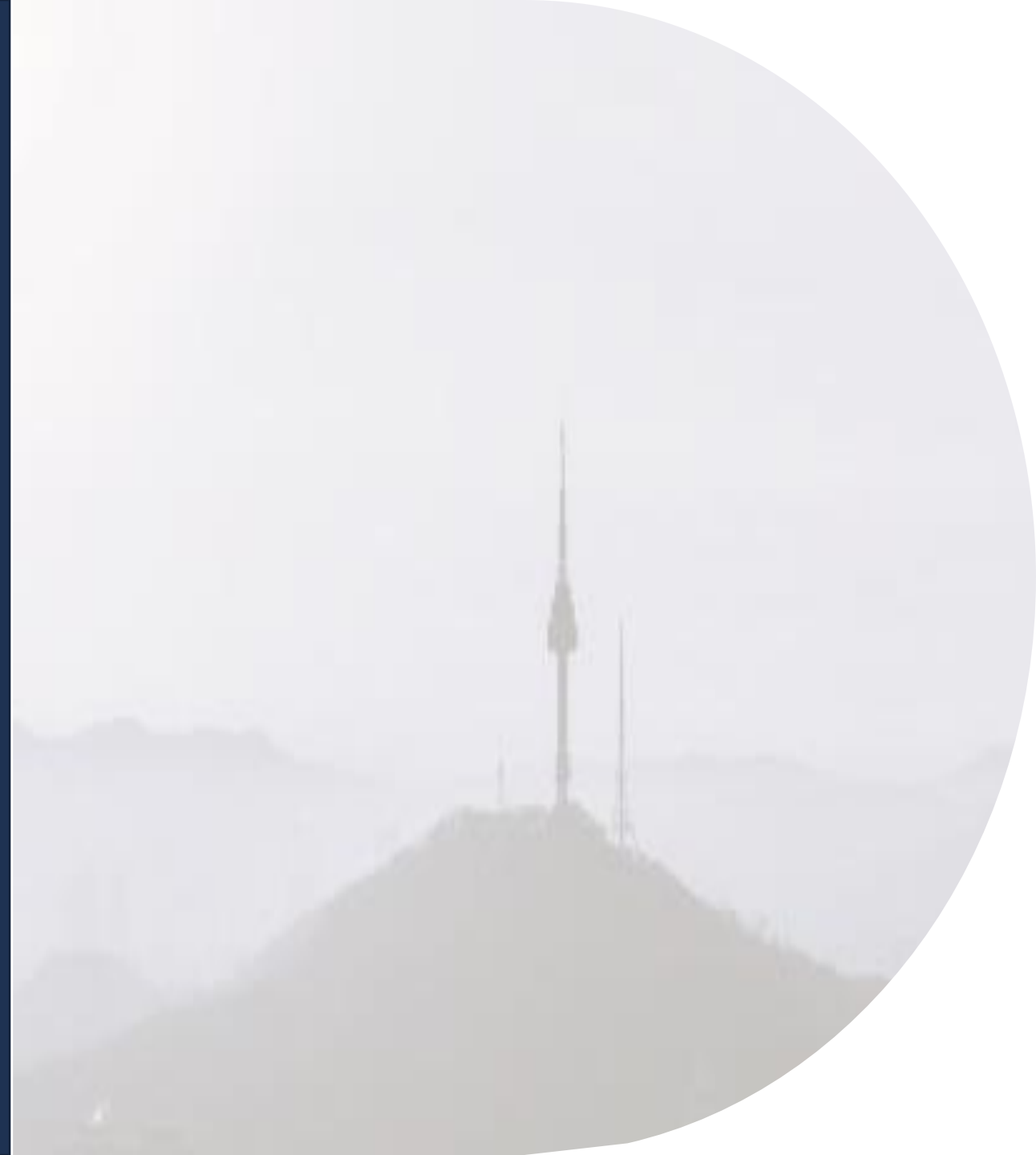
일평균	
count	126228.000000
mean	35.704089
std	21.300820
min	0.000000
25%	21.400000
50%	31.200000
75%	44.600000
max	386.500000

지하철 이용객수

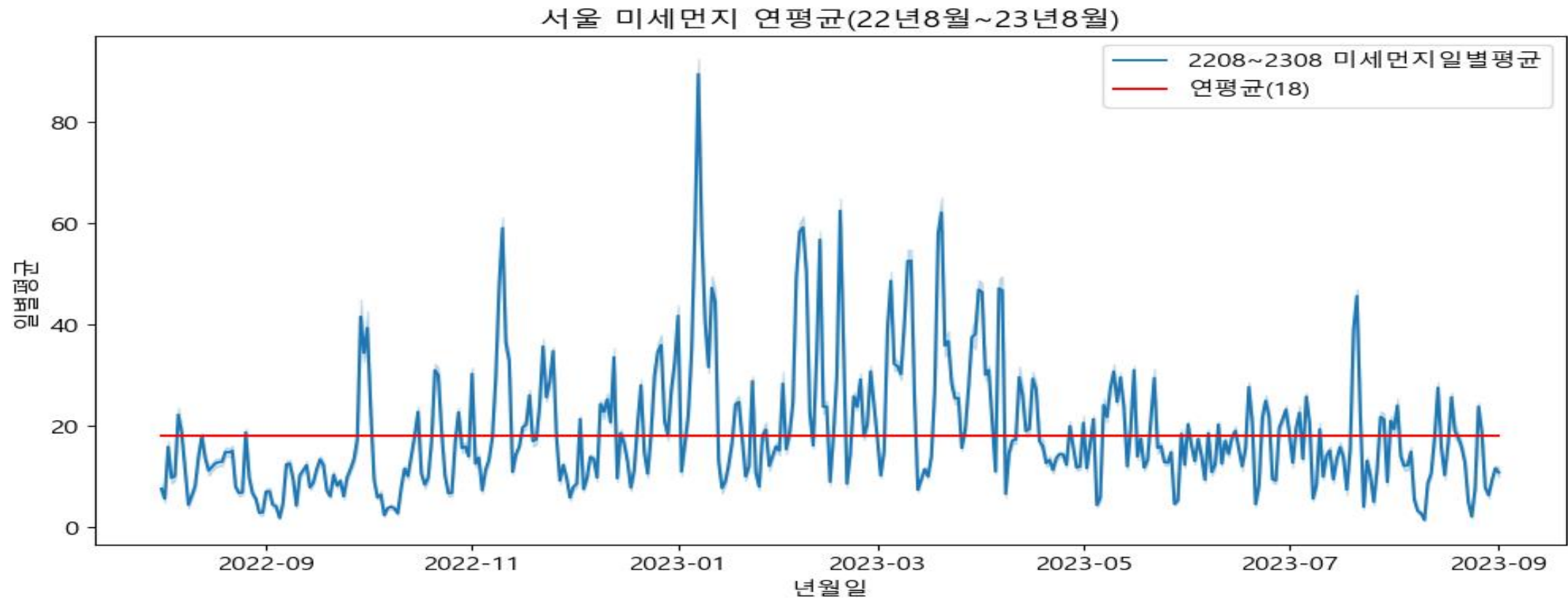
사용일자변경			하차총승객수
count	298		298.000000
mean	2023-01-07 00:00:00		11528.895973
min	2023-01-07 00:00:00		0.000000
25%	2023-01-07 00:00:00		4965.750000
50%	2023-01-07 00:00:00		8529.000000
75%	2023-01-07 00:00:00		13989.750000
max	2023-01-07 00:00:00		85772.000000
std		NaN	10666.613846

Part 3

분석 결과 및 시각화

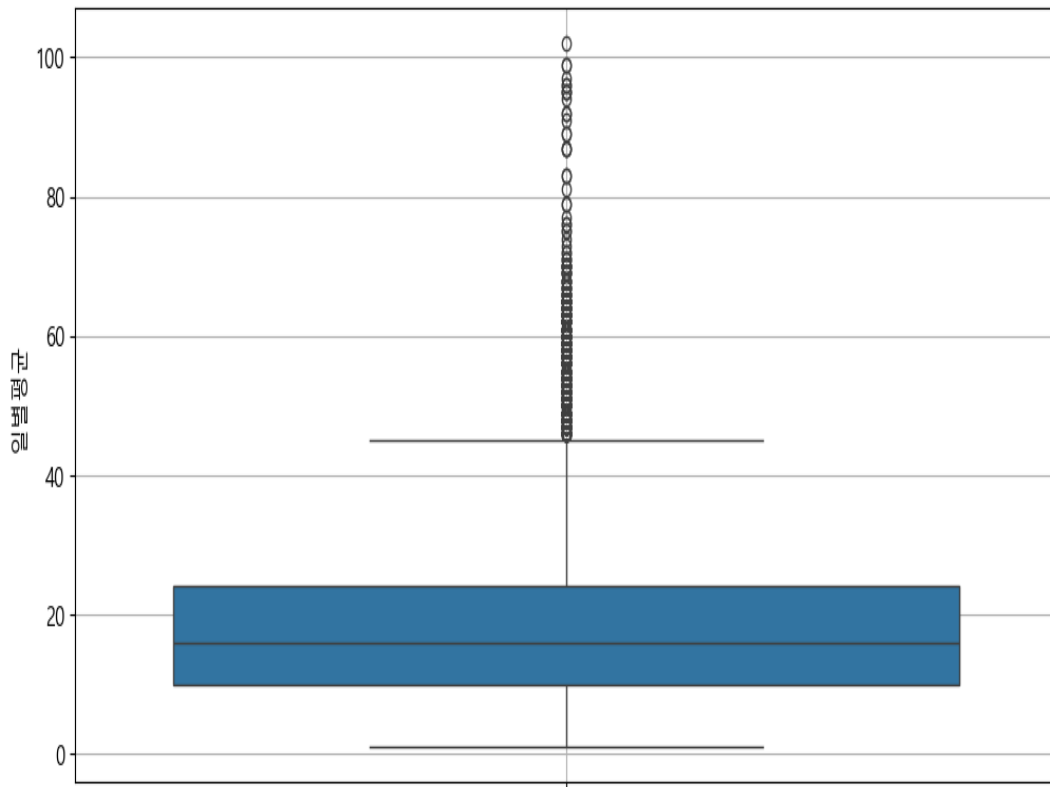


<22년 8월 ~23년 8월 서울 초미세먼지 농도>

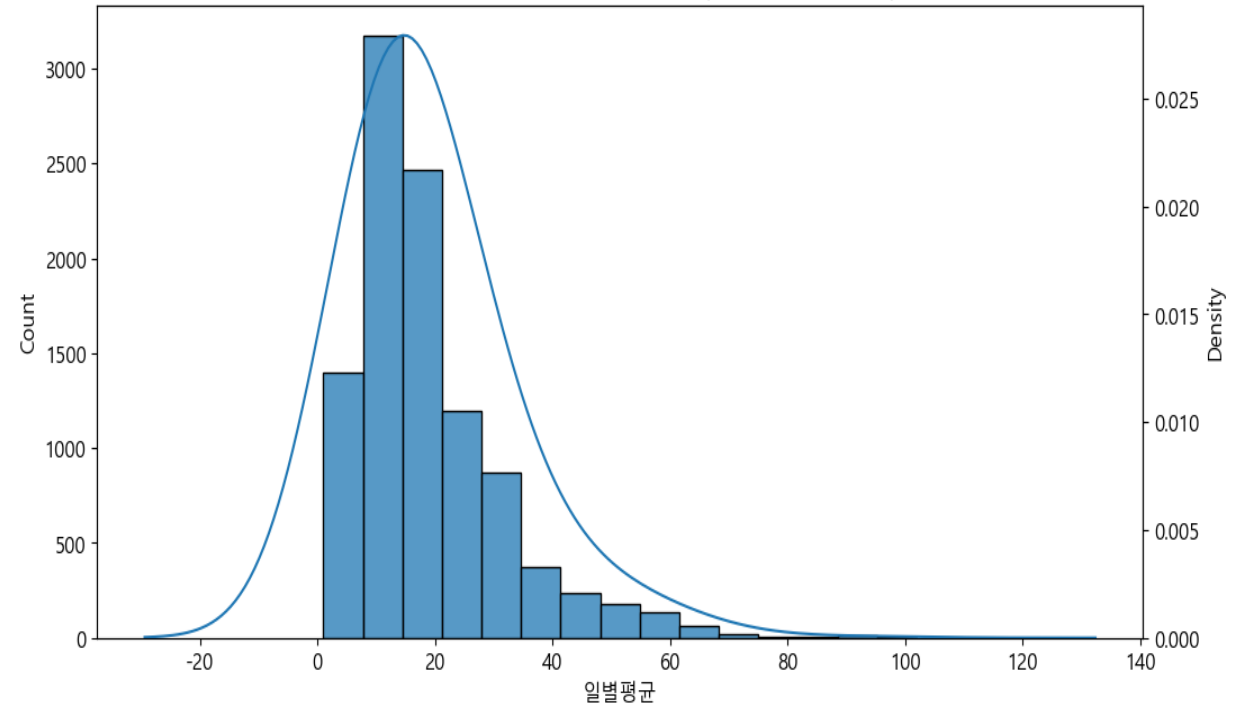


<22년 8월 ~23년 8월 서울 초미세먼지 농도 상자그림과 히스토그램>

서울 외부 미세먼지 일별평균 상자그림

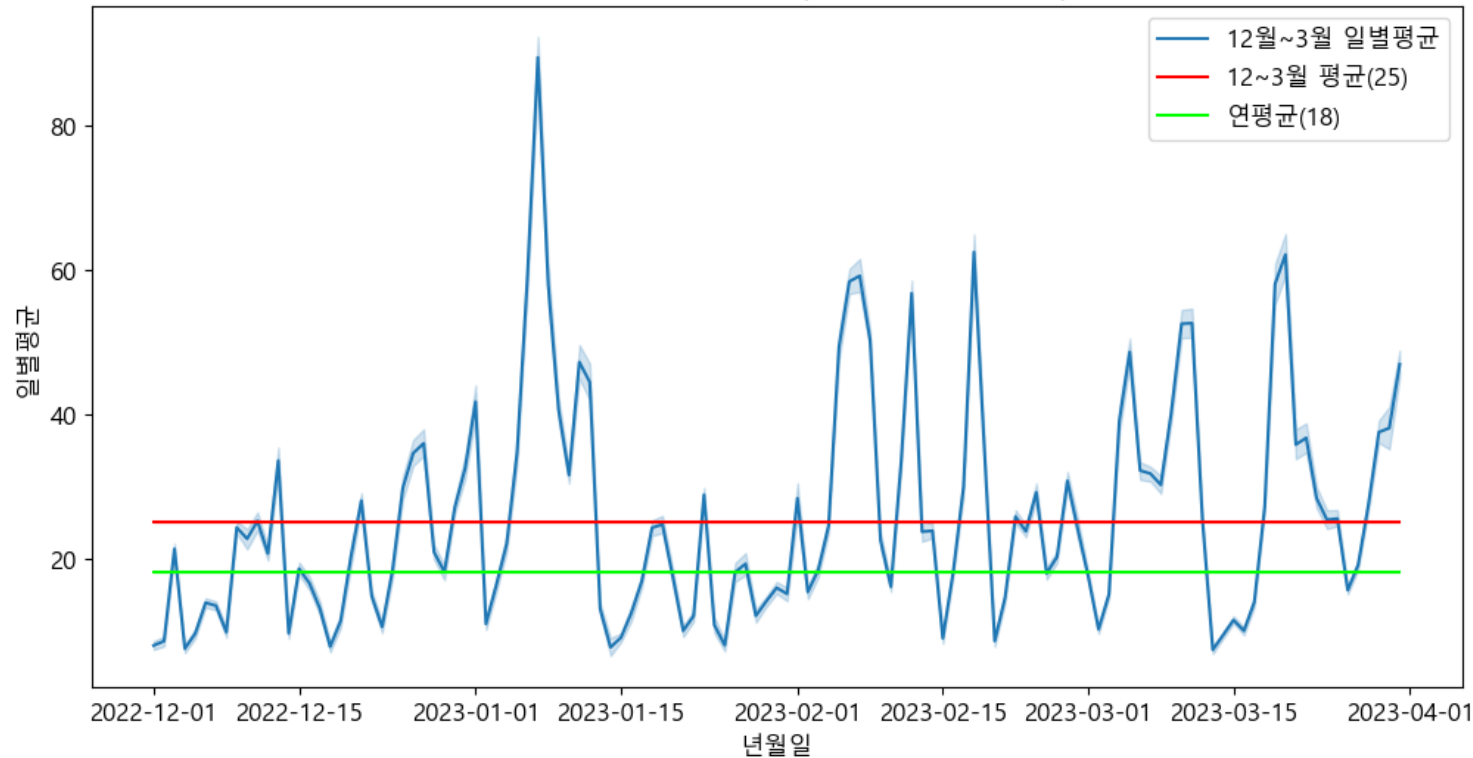


서울 미세먼지 일별평균 히스토그램(22년8월~23년8월)



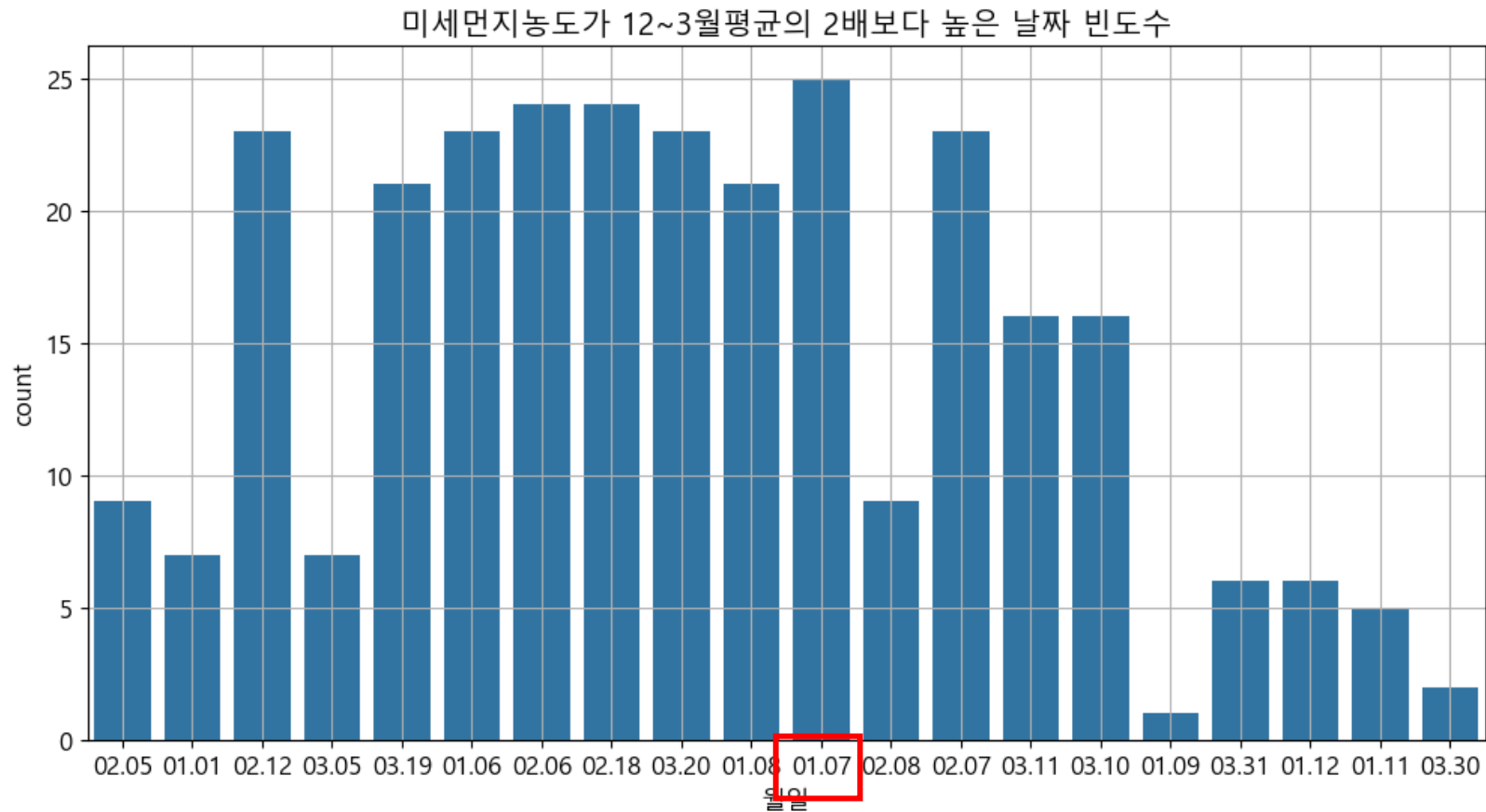
<22년 12월 ~23년 3월 서울 초미세먼지 농도 lineplot>

서울 미세먼지 일별평균(22년12월~23년3월)



분석 기준을 잡기 위
해 미세먼지 농도가
연중 가장 높은 12월
~3월을
시간기준으로 잡음

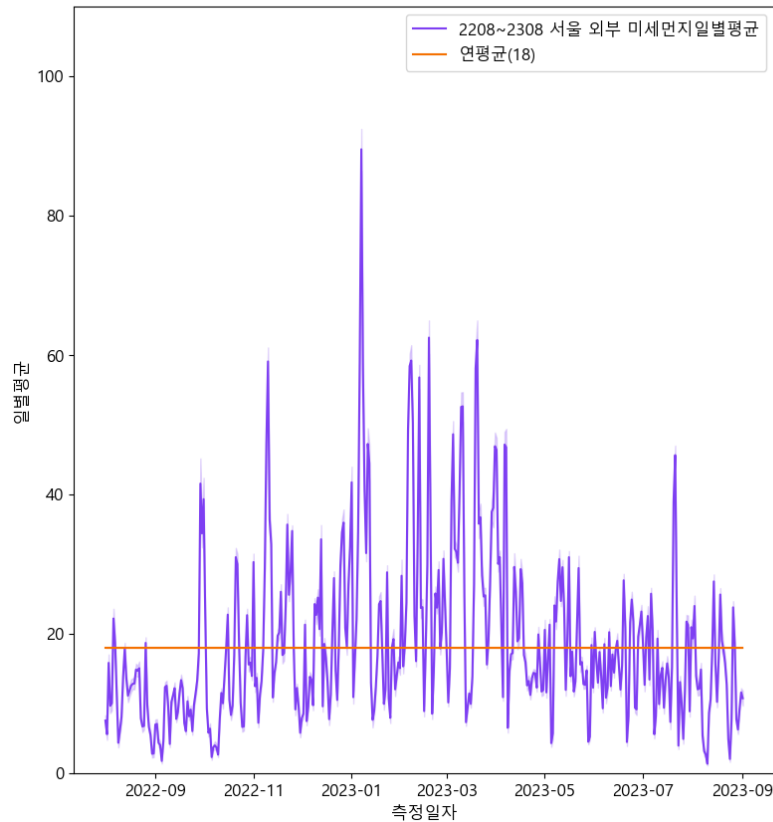
<22년 12월 ~23년 3월 서울 초미세먼지 평균농도의 2배 초과한 자치구들의 날짜 빈도



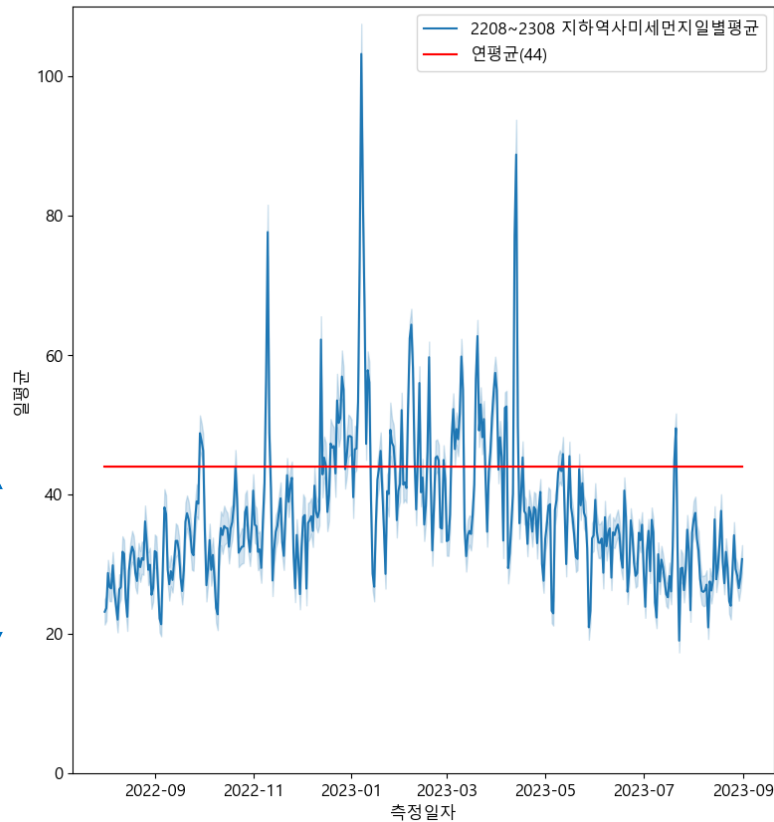
분석 기준 날짜
23년 1월 7일

날짜를 고정해서
외부변수 고정

서울 외부와 지하역사 미세먼지 비교



26



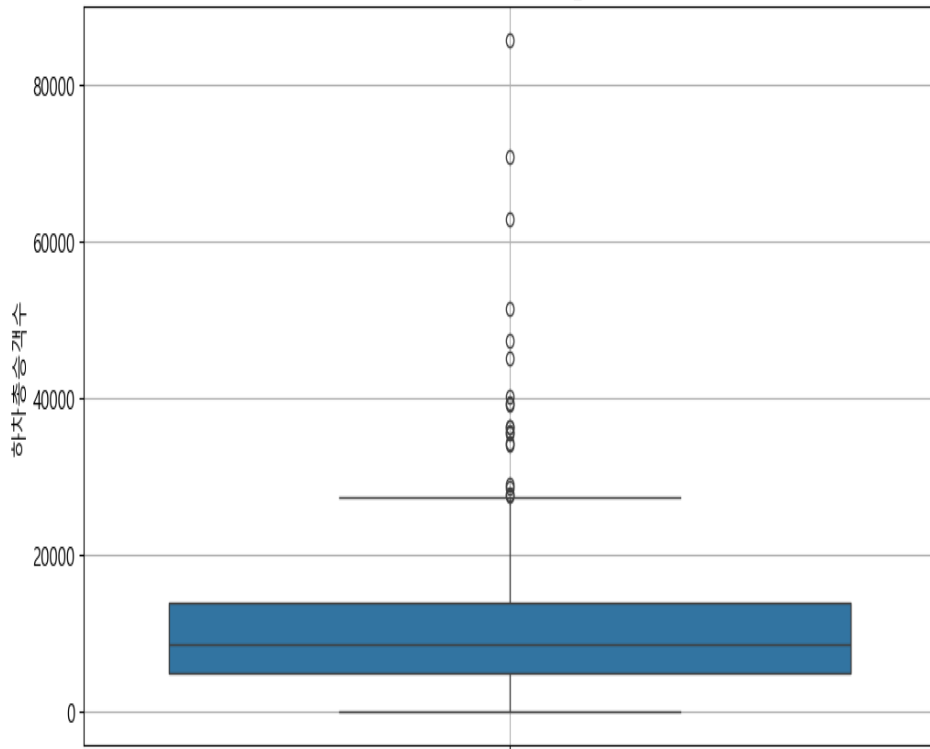
외부 연평균 $18 \mu\text{g}/\text{m}^3$

지하역사 연평균 $44 \mu\text{g}/\text{m}^3$

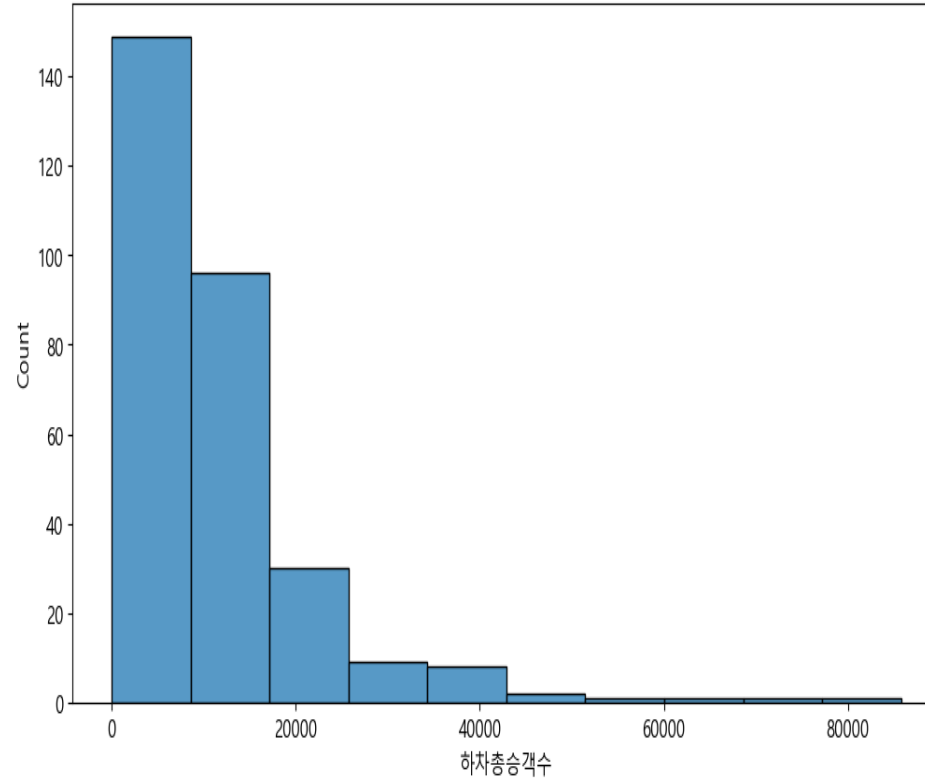
대기환경 기준
연평균 $15 \mu\text{g}/\text{m}^3$

<23년 1월 7일 서울 지하철역사 유동인구 상자그림과 히스토그램_하차인원>

서울 지하철 하차승객수_230107



서울 지하철 하차승객수_230107



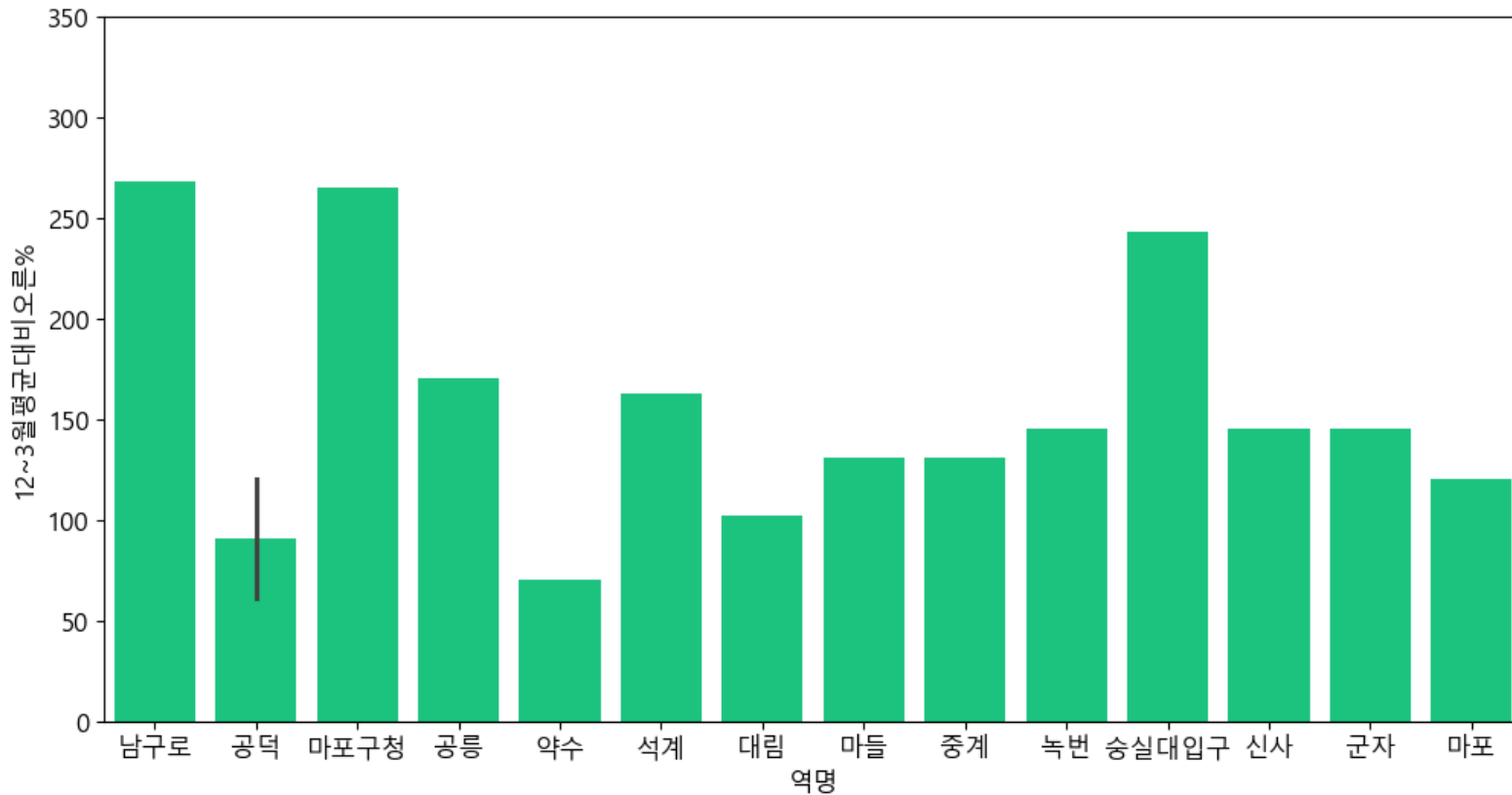
이상치가 높게 나
와 평균과
중간값
차이가 많이 나서
기준점을
중간값 $\pm 5\%$ 잡음

데이터 시각화

역명	호선	측정일자	일평균	월평균	12~3월평균대비%
남구로	7호선	2023-01-07	162	44	268.863636
공덕	5호선	2023-01-07	97	44	121.136364
마포구청	6호선	2023-01-07	161	44	267.727273
공릉	7호선	2023-01-07	119	44	172.045455
약수	3호선	2023-01-07	75	44	71.136364
석계	6호선	2023-01-07	116	44	164.772727
대림	7호선	2023-01-07	89	44	103.636364
마들	7호선	2023-01-07	102	44	132.954545
공덕	6호선	2023-01-07	71	44	62.727273
중계	7호선	2023-01-07	102	44	131.818182
녹번	3호선	2023-01-07	108	44	145.681818
송실대입구	7호선	2023-01-07	151	44	243.409091
신사	3호선	2023-01-07	108	44	147.272727
군자	7호선	2023-01-07	108	44	147.045455
마포	5호선	2023-01-07	97	44	121.363636

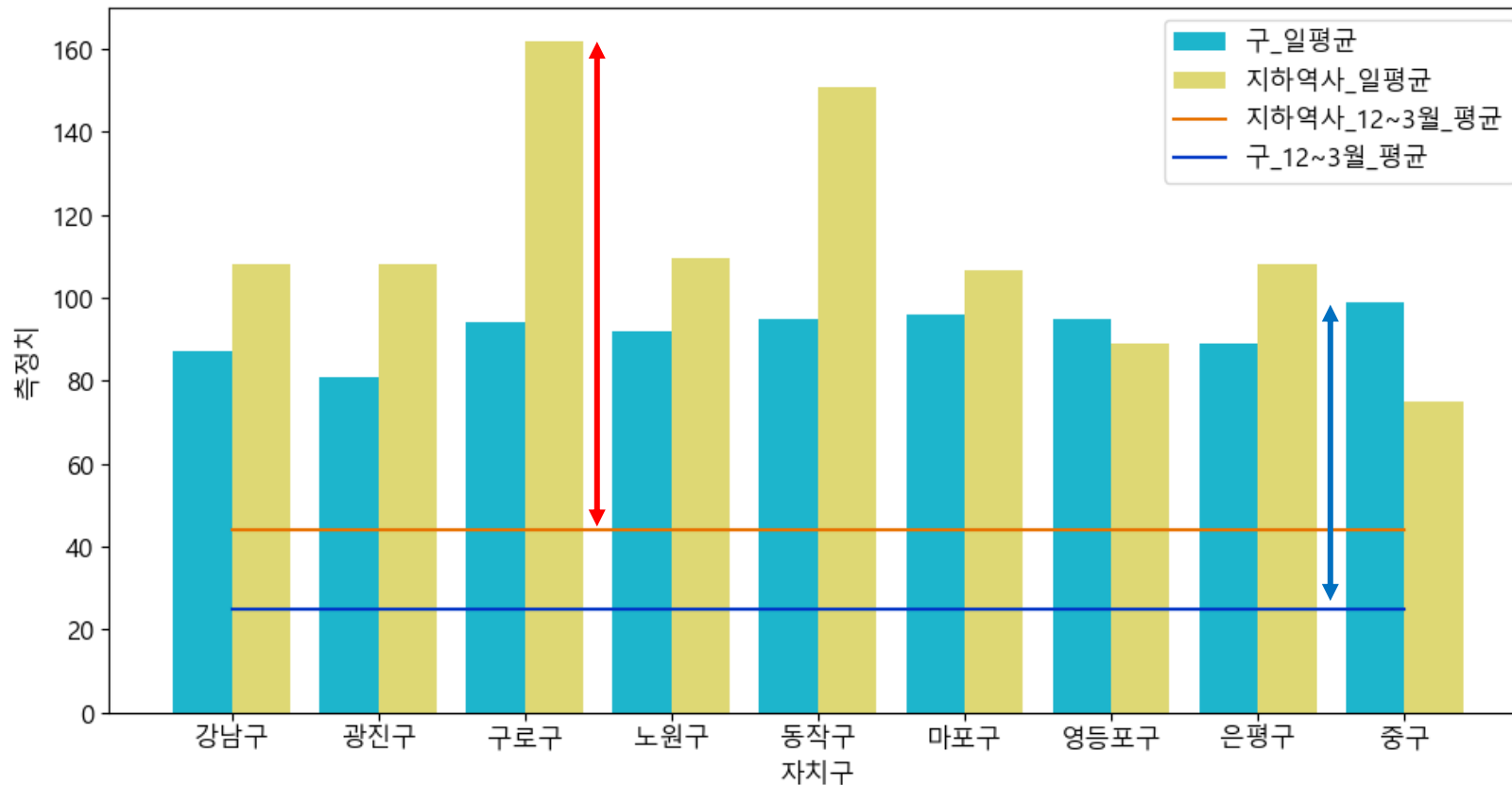
1월 7일 유동인구 중간값 8500
 명 $\pm 5\%$ 범위의
 지하철역 추출 후 지하철역사의
 12~3월 평균 대비 오른 % 계산

<23년 1월 7일 유동인구 유사한 지하철역 미세먼지의 12~3월 평균대비 증가비율>



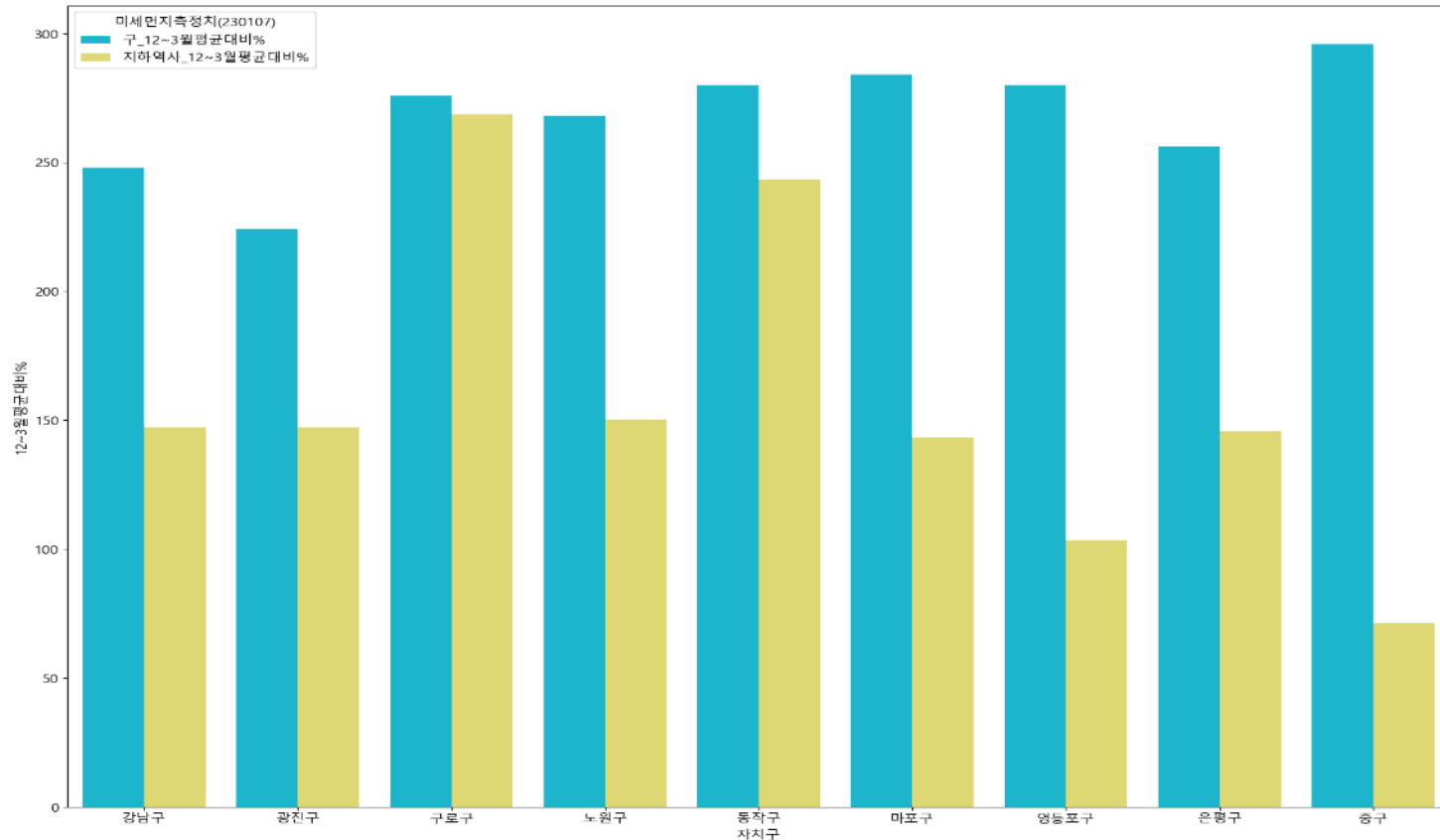
외부 미세먼지와 비교하기
위해서
역사가 위치한 자치구로
변환 필요

<23년 1월 7일 유동인구가 유사한 지하역사 미세먼지와 외부 미세먼지 측정치>



지하역사가 속한 자치구로
변환
자치구의 외부와 지하역사
비교

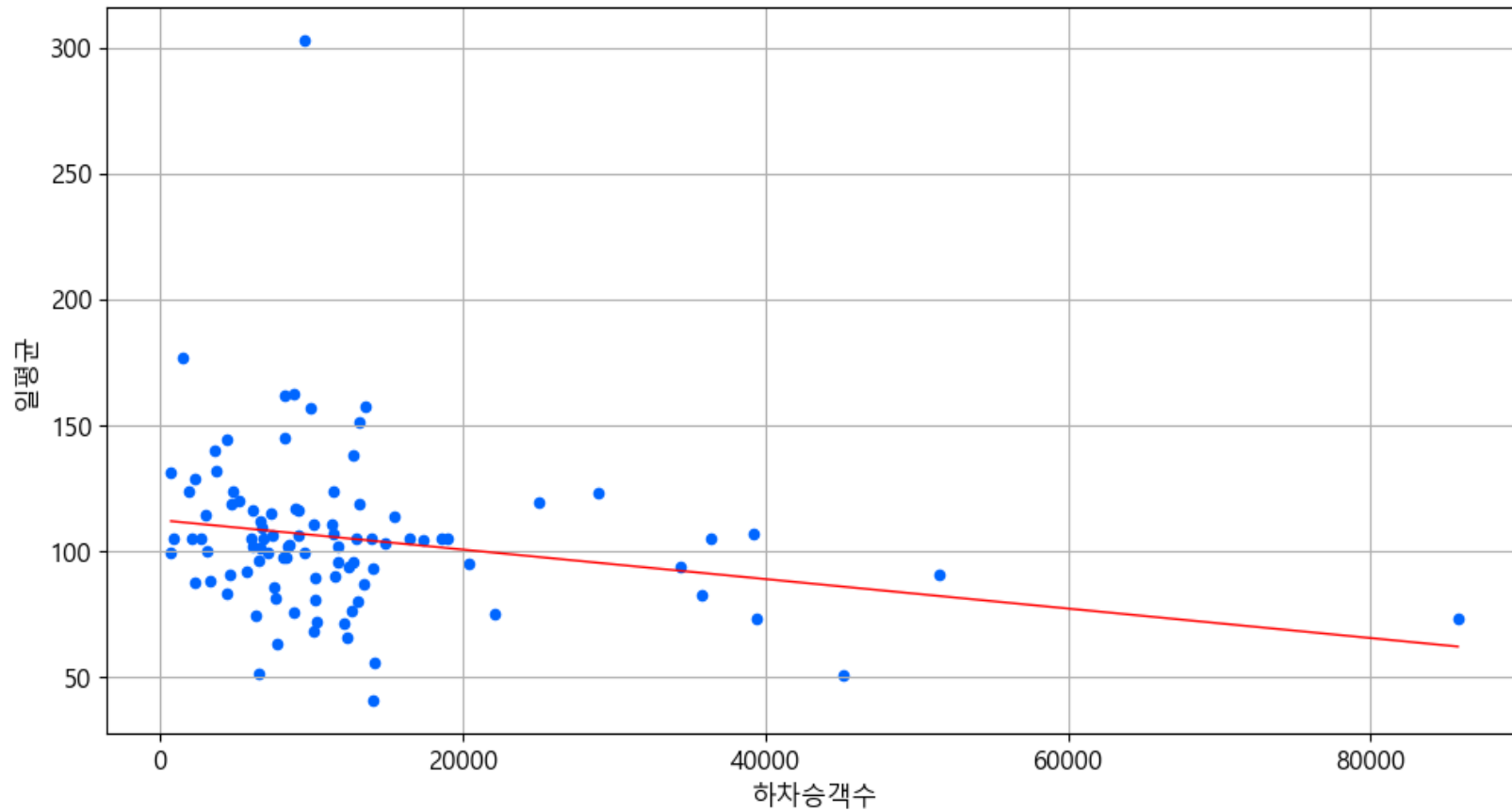
<23년 1월 7일 유동인구 유사한 지하역사 미세먼지와 외부 미세먼지 측정치>



1월 7일(날씨 고정)
지하철역 유동인구 고정

초미세먼지 12~3월 평균 대비 증
가율

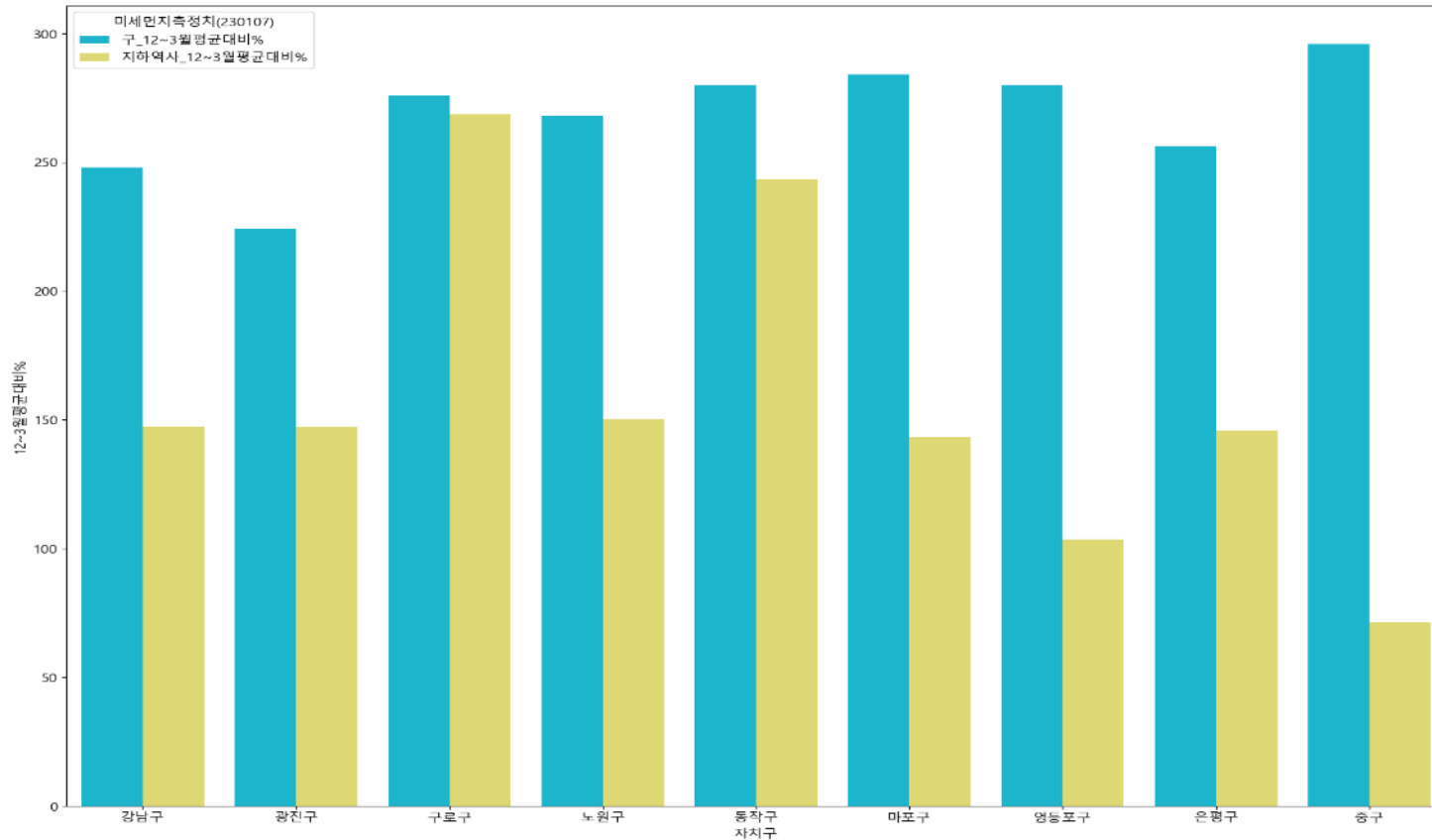
외부 : 268%
지하역사: 157%



1월 7일(외부대기 고정)

초미세먼지 수치 상위 10개
구의 역사 선정($\pm 5\%$ 범위)

<23년 1월 7일 유동인구 유사한 지하역사 미세먼지와 외부 미세먼지 측정치>



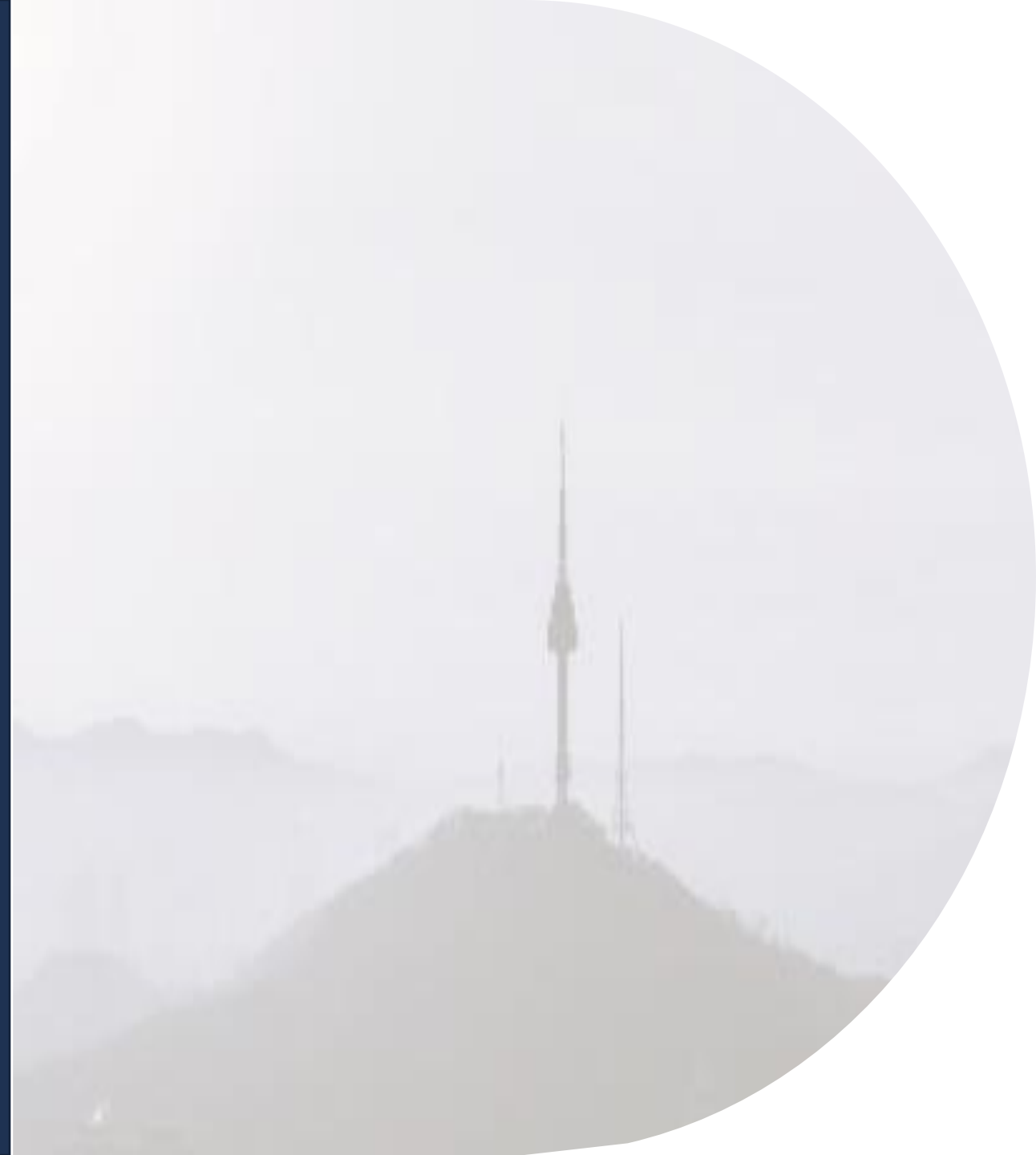
1월 7일(날씨 고정)
지하철역 유동인구 고정

초미세먼지 12~3월 평균 대비 증
가율

외부 : 268%
지하역사: 157%

Part 4

결론 및 인사이트



인사이트와 한계점

외부 미세먼지와 지하역사
미세먼지의 상관관계



외부가 268% 오를 때
지하역사는 157% 오름

지하철 유동인구와 지하역사
미세먼지의 상관관계



지하철 유동인구와 지하역사 미세
먼지의 산점도 그래프 확인 결과
유의미한 관계 없음

외부 미세먼지와 지하역사 미세먼
지는 양의 상관관계가 있고 유동
인구와는 상관관계가 적음



표본이 적고 오염배출원, 실내공
기정화장치 등 많은 변수 미고려
한 한계점

감사합니다