

서울시 도시 데이터 센서(S-DoT) 위치 선정

팀명 : Spatial_Analysis



목차

#1, 분석 개요

#2, 기본 센서 설치

#3, 추가 센서 설치

#4, 활용 방안

Part 1,

분석 개요

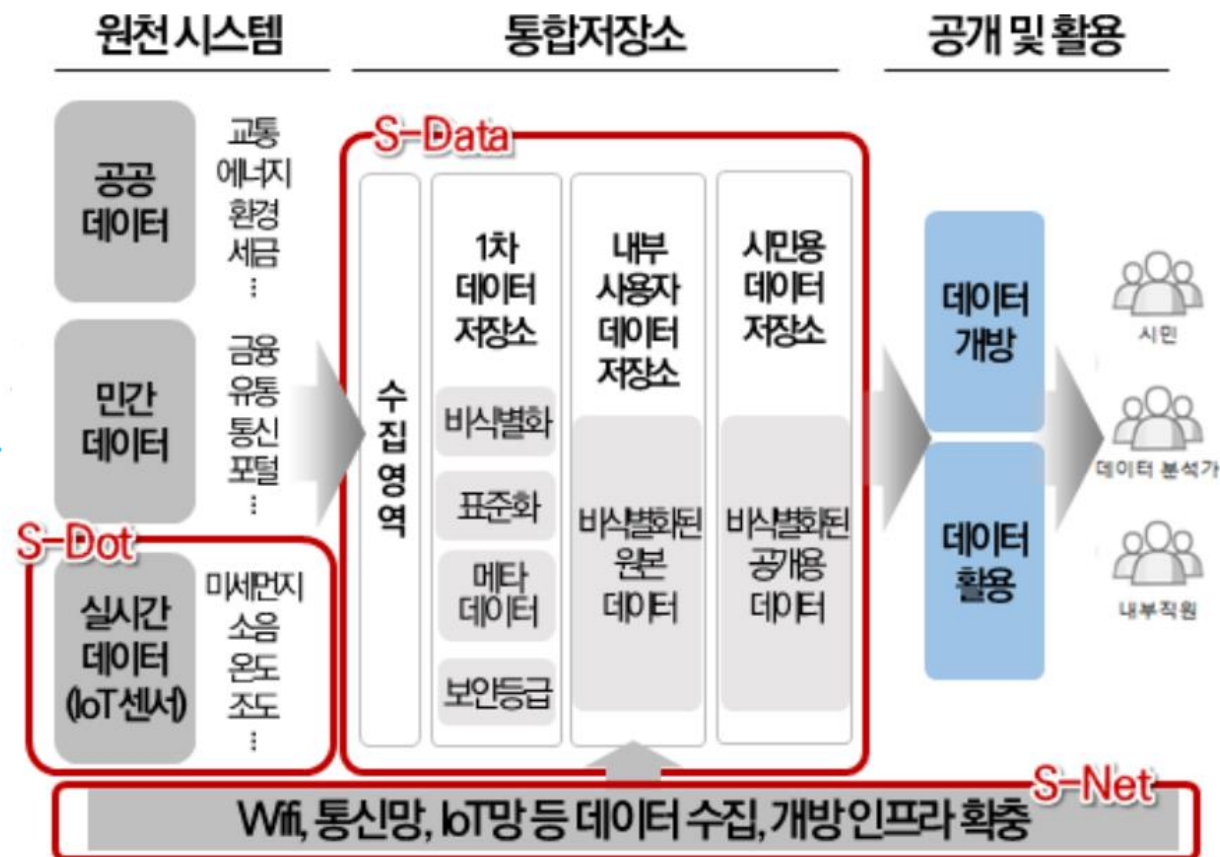


1. 분석 개요

서울시 스마트 시티 사업

- 안전, 복지를 제외한 총 6개 분야에서 주요 사업을 선정하여 2018년부터 추진 중인 사업

빅데이터	• 공공데이터를 업데이트하면서 부문간 연결 및 개방·활용하는 사업에 집중
신산업거점 일자리	• Bio, AI 등 신산업 거점지구를 조성하고, 특히 스마트시티의 핵심인 AI분야 를 양재 지역에 핵심사업으로 조성
교통	• 데이터 통합 및 통행 관리 시스템을 통한 플랫폼 서비스를 구축하면서, 상암 중심으로 자율주행 테스트베드 조성
환경 에너지	• 인프라를 확충하면서, 핵심사업인 친환경에너지 확대 및 저장 활용사업 모델 발굴 추진
재생	• 세운상가, Y밸리에 메이커시티를 조성하여 창업을 지원하고 있으나, 재래산업단지 와 연계는 미흡
시범사업	• 각 분야를 도시공간에 적용하는 실제적 종합시범사업으로 마곡지구 역점 추진



1. 분석 개요

S-DoT 사업



도시데이터 센서(S-DoT)를 기반으로 수집된 데이터는 데이터 개방 및 활용을 넘어 스마트 시티를 만들어 나갈 다양한 분야에서 활용될 예정

<고려해야 할 점>

1. 서울시 커버 범위 최대화
2. 측정값 불확실성 최소화

1. 분석 개요

과제 소개

목적

설치, 운영 중인 도시 데이터 센서 현황 정보 등을 활용하여 추가 설치가 필요한 센서의 위치와 측정 항목 도출

제안할 사항 (과제)

- 1) 대기환경 정보 측정기 설치위치 제시 (기본 센서)
- 2) 측정기 설치 위치별 측정 항목 제시 (추가 센서)
- 3) 추가된 측정 항목별 활용 서비스 예시

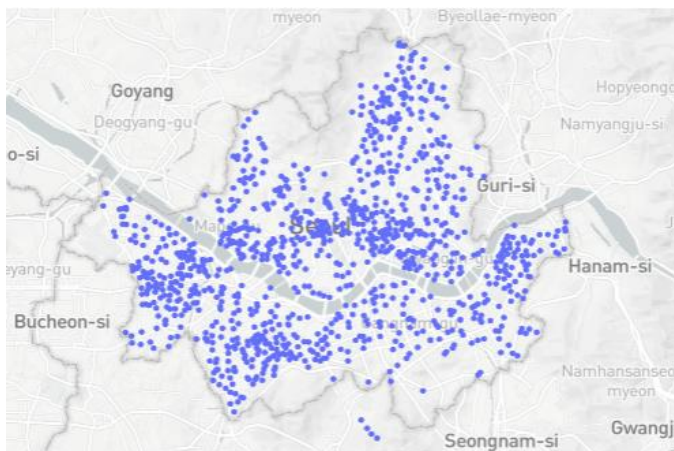
분석 시 고려사항

- 1) 측정기에는 기본 센서와 추가 센서 5종으로 구성되며, 각 센서별 측정 항목 및 설치 수량이 정해져 있음
- 2) 대기환경 정보 측정기는 CCTV 주변, 서울시 공공기관 건물(옥상 및 벽면) 및 시설물에만 설치 가능
- 3) 추가 센서는 기본 센서가 설치된 대기환경 정보 측정기에 추가되는 형식
- 4) 향후 측정기에서 수집되는 데이터는 서울시 대기환경 지도, 바람 길 지도 등에 활용 가능성 존재

1. 분석 개요

각 과제별 현황 및 분석 목표

기본 센서 위치 Map Box

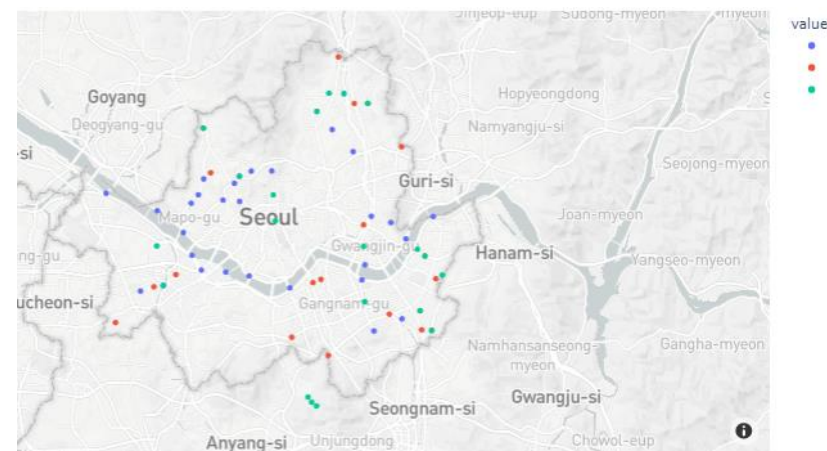


2019년	2020년	2021년	2022년
850대 설치	250대 설치	800대 계획	600대 계획



분석 목표 : 측정값 불확실성 최소화

추가 센서 위치 Map Box (1번 센서 제외)



추가 센서1	추가 센서2	추가 센서3	추가 센서4	추가 센서5
1059대	32대	0대	20대	20대



분석 목표 : 서울시 커버 범위 최대화

1. 분석 개요

1번 과제(기본 센서 설치) 분석 과정

- 1 설치, 운영 중인 도시 데이터 센서 현황 정보 등을 활용하여 추가 설치가 필요한 센서의 위치와 측정 항목 도출
- 2 해당 날짜 기준 지점별 평균 속성값 생성 (미세먼지, 온도, 습도, 조도, 자외선)
- 3 후보 지점의 불확실성을 측정할 수 있는 모델 선정 : Kriging Model(크리깅 모델) -> 각 후보 지점의 속성별 분산 산출
- 4 각 후보 지점의 CV(Coefficient of Variation; 변동계수)값 계산 및 가중치를 통한 불확실성 생성
- 5 최적화 모델 선정 : 지점별 불확실성을 가중치로 한 MCLP(Maximal Covering Location Problem) 수행

1. 분석 개요

2번 과제(추가 센서 설치) 분석 과정

1

각 추가 센서별 연관 있는 변수에 대한 데이터 수집

2

해당 데이터에 대한 EDA 수행

3

연관 있는 변수에 대한 적절한 가중치 생성

4

최적화 모델 선정 : 해당 가중치를 반영한 MCLP(Maximal Covering Location Problem) 수행

1. 분석 개요

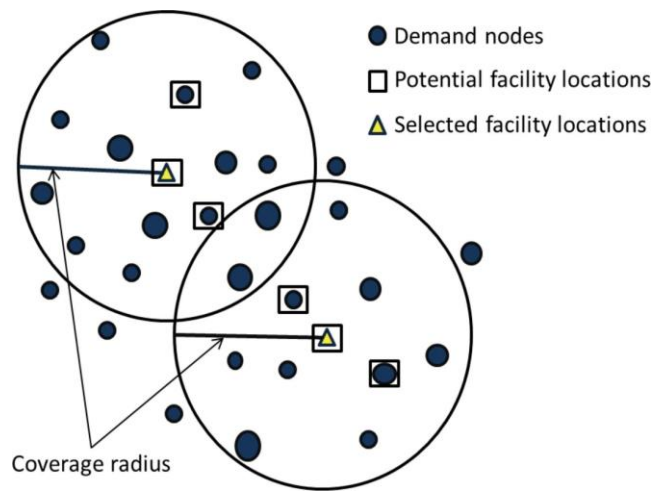
Kriging Model

- 미지의 관측지점 속성값을 이미 알고 있는 주위의 확률 변수들의 선형조합으로 모델링한 방법론
 - 분리 거리에 따른 관측치의 비유사성을 이용하고 이는 variogram을 통해 적합
 - 이때, variogram은 Auto-Fit 을 이용하여 적합
- 관심있는 지점의 속성에 대한 예측값과 분산을 산출
 - 이때, 관심있는 지점의 분산은 해당 지점의 불확실성과 연관
 - 기본 센서의 경우, 다수의 속성값이 존재하므로 이들의 분산을 적절히 결합하여 하나의 불확실성으로 만들 필요 존재
- $CV(\text{변동계수}) = \frac{\text{표준편차}(\delta)}{\text{평균}(\mu)}$
 - 표본 사이의 표준편차를 비교할 때, 평균의 영향을 없애기 위하여 사용
 - 상대표준편차(RSD; relative standard deviation)라고도 불림
 - 이를 이용하여 각 속성간의 평균의 영향을 없앤 후, 가중치를 이용한 가중 평균을 통해 불확실성 생성

1. 분석 개요

MCLP (Maximal Covering Location Problem)

- 설비가 커버하는 영역의 합을 최대화 하는 K개 설비를 세울 위치를 선정하는 방법론



- 센서 위치 선정에서 MCLP 활용
 - 1번 과제의 경우, 10,000개의 후보지에서 1,000개의 장소를 선정
 - 2번 과제의 경우, 1,000개의 센서에서 각 추가 센서별 200개의 센서를 선정

1. 분석 개요

MCLP (Maximal Covering Location Problem)

- 최대 지역 커버 문제로 Mixed Integer Linear Programming 문제

$$\sum_{i \in I} w_i y_i \dots (1)$$

$$\text{s.t. } y_i \leq \sum_{j \in N_i} x_j \quad \text{for all } i \in I \dots (2)$$

$$\sum_{j \in J} x_j = K \dots (3)$$

$$x_j, y_i \in \{0, 1\} \quad \text{for all } i \in I, j \in J$$

- 결정 변수

- y : 설치 후보 기준 Point (건물 좌표와 CCTV 좌표 중 랜덤하게 선택한 10000개)
- K : IoT 설치 대수 (1000개)
- w : 가중치로서 앞서 살펴 본 불확실성 지표값
- Radius(Cover 영역) : 500m
- M (Weight 고려 정도) : 다음 장에서 조절하며 살펴보기

<문자 DESCRIBE>

- i : 수요 포인트 index
- j : 설비 후보지역 index
- I : 수요 포인트 집합
- J : 설비 후보지역 집합
- K : 총 설치해야 하는 설비 개수
- x : 설비 후보 지역 중 위치 j 에 설비가 설치되면 1, 그렇지 않으면 0
- y : 적어도 하나의 설비로 그 포인트가 커버되면 1, 그렇지 않으면 0
- w : 가중치

<괄호 DESCRIBE>

- (1) : 목적함수로 가중치 w 인 수요 지점을 최대한 많이 커버하도록 Maximize
- (2) : 수요 지점 i 는 설비 후보 지역이 커버하는 거리 안에 적어도 하나 이상의 설비로부터 커버되도록 Constraint
- (3) : 총 설치할 설비 개수는 K 가 되도록 Constraint

A close-up, shallow depth-of-field photograph of a person's hands typing on a silver laptop keyboard. The person is wearing a white smartwatch with a black screen on their left wrist. The laptop is an Apple MacBook, with the logo visible on the lid. The background is blurred, showing a desk and a cup. A dark blue rectangular overlay is positioned on the left side of the image, containing white text.

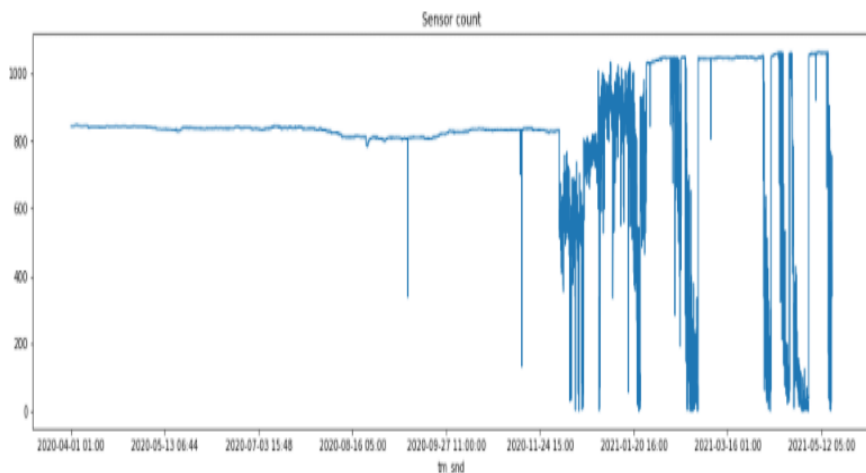
Part 2,

기본 센서 설치

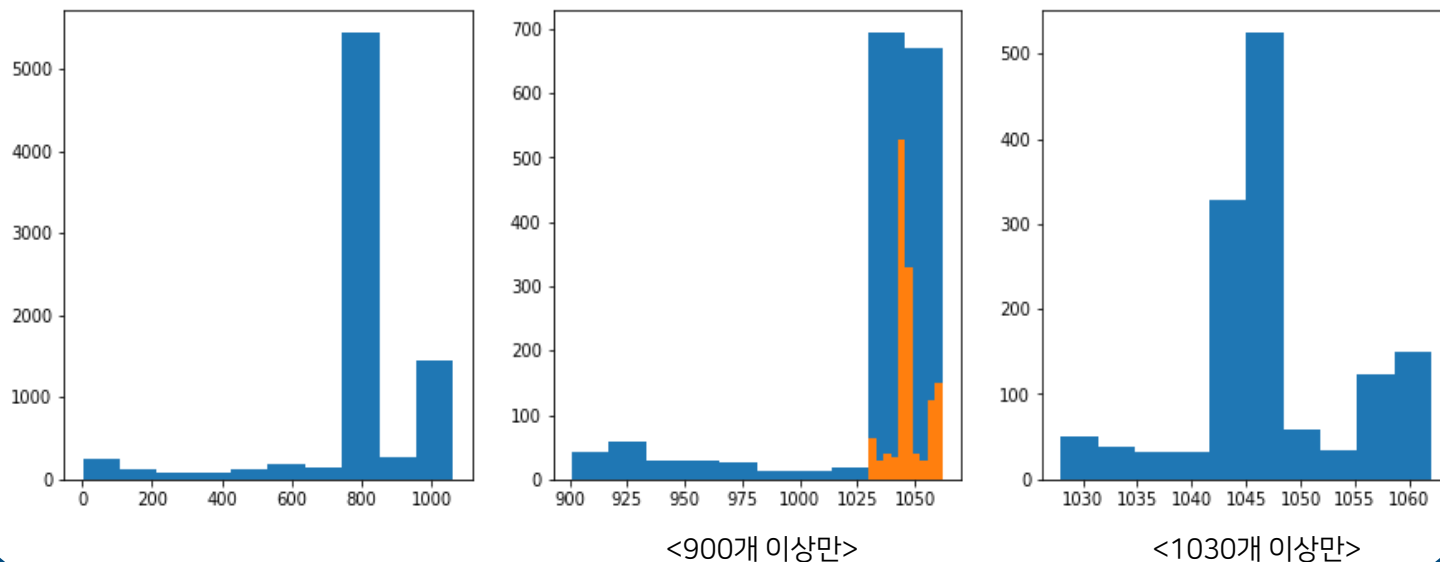
2. 기본 센서 설치

시간별 센서 개수 확인

시간별 측정된 센서 개수



센서 측정 기준 날짜 개수



- 크리깅 모델을 수행하기 위해서는 동일한 시간대에 측정된 속성값이 존재해야 함
- 하지만, 모든 센서가 동시에 측정된 시간대는 존재하지 않으므로 측정된 센서 개수가 1040개 이상인 시간대 선정
(총 센서 개수 : 1064개)

2. 기본 센서 설치

지점별 평균 측정값 생성

지점	시간	미세먼지	...	자외선
A1	2021-02-10 00:00	평균		평균
	2021-02-10 01:00			
	...			
	2021-02-10 23:00			
A2	2021-02-10 00:00	평균		평균
	2021-02-10 01:00			
	...			
	2021-02-10 23:00			
⋮	⋮	⋮	⋮	⋮



지점	미세먼지	...	자외선
A1			
A2			
⋮	⋮	⋮	⋮
A1063			

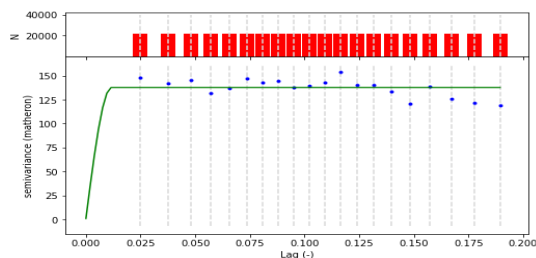
<일별 평균 측정값>

- 각 지점별 일별 평균 측정값 생성 : 미세먼지(PM10, PM25), 온도, 습도, 조도, 자외선
- 지점별 속성값을 통하여 각 속성들에 대하여 크리깅 모델 적합

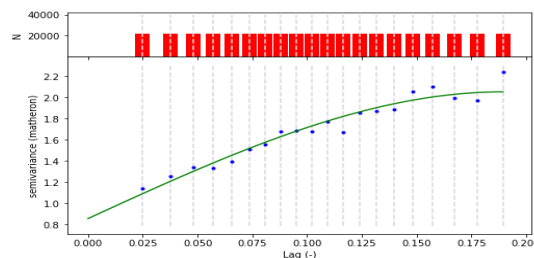
2. 기본 센서 설치

크리깅 모델 적합

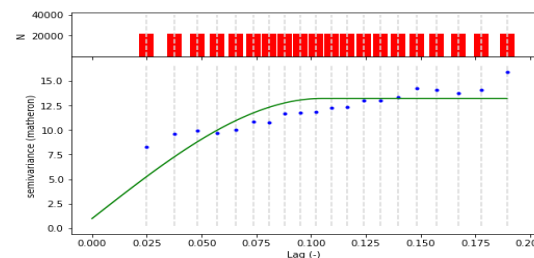
- 앞서 본 센서별 평균 측정값을 토대로 크리깅 모델 적합
 - Variogram의 경우 데이터에 맞게 Auto-Fit



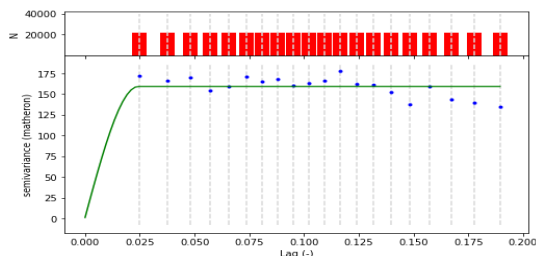
-PM10-



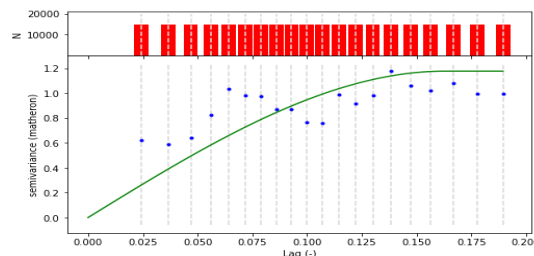
-온도-



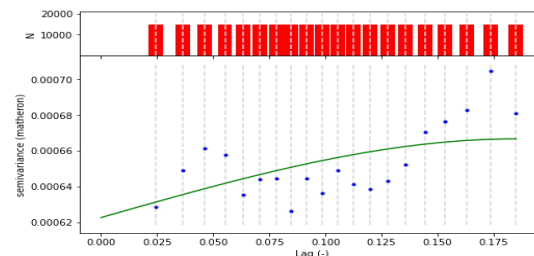
-습도-



-PM25-



-조도-



-자외선-

- 건물 좌표와 CCTV 좌표들에 대하여 예측값 및 분산 산출
 - 후보지 : 대략 60만 개 (서울시 전체에 골고루 분포)
 - 불확실성 지표 산출을 위해 후보 좌표에 대한 표준편차 계산

※ 예측이 진행되며 서버가 터지는 문제

→ 추후, MCLP 최적화를 할 때에도 서버가 터지는 문제

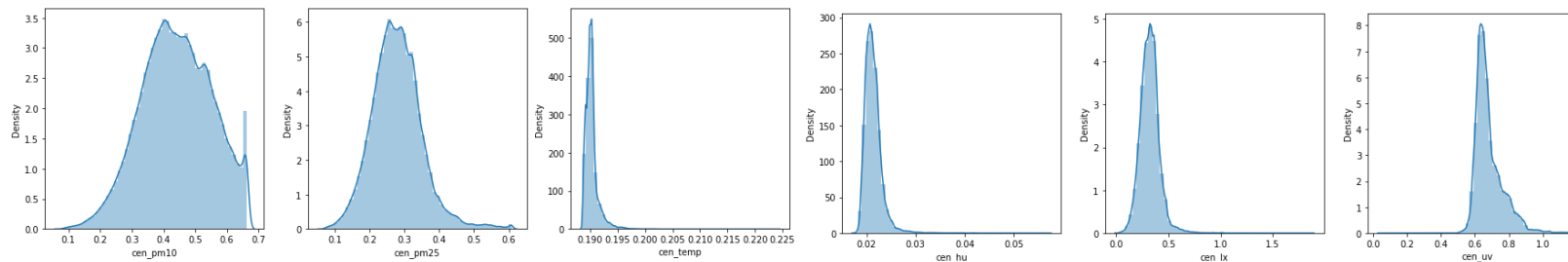
→ 일부 데이터만 뽑아서 후보지로 선정 (10000개)

2. 기본 센서 설치

불확실성 지표 산출

- 각 속성별 분산과 평균을 통하여 속성에 대한 CV(변동계수) 계산

$$CV(\text{변동계수}) = \frac{\text{표준편차}(\delta)}{\text{평균}(\mu)}$$



- 속성별 CV(변동계수) 값에 대한 가중 평균을 통해서 산출

	PM10	...	자외선
B1	4.113.	...	2.123
B2	5.124	...	2.457
...	3.143	...	3.147

<후보 지점별 속성 cv값>

X
행렬곱

가중치
w1
...
w6

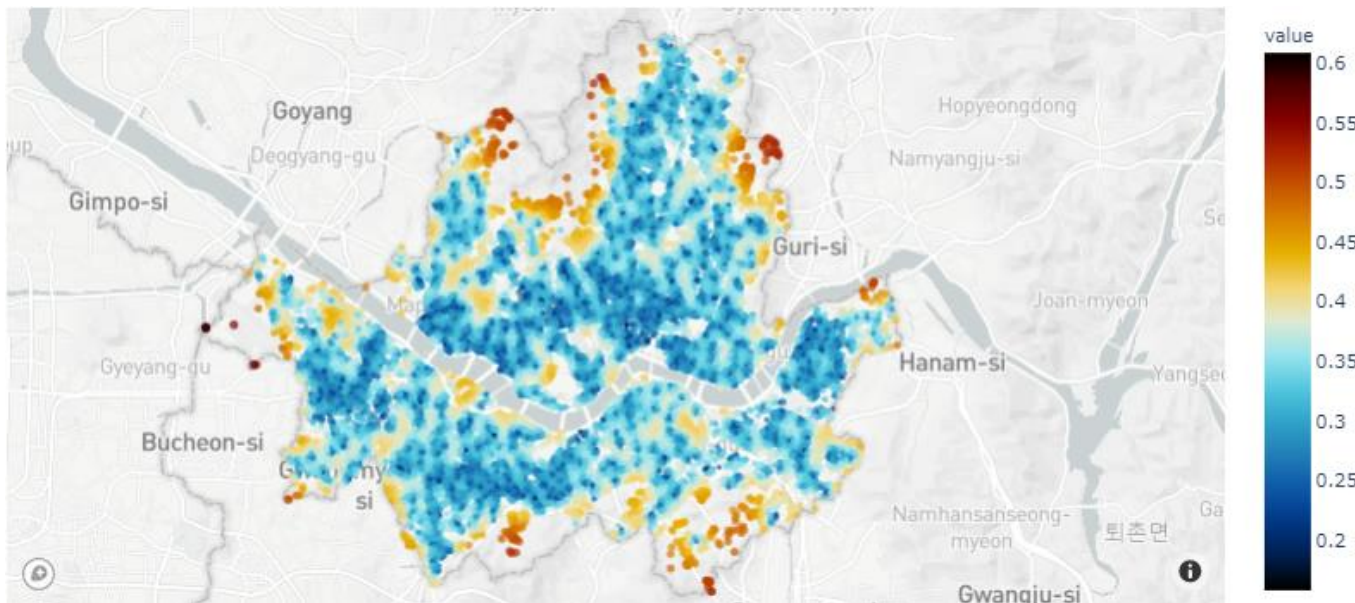
=

가중치
3.154
...
2.451

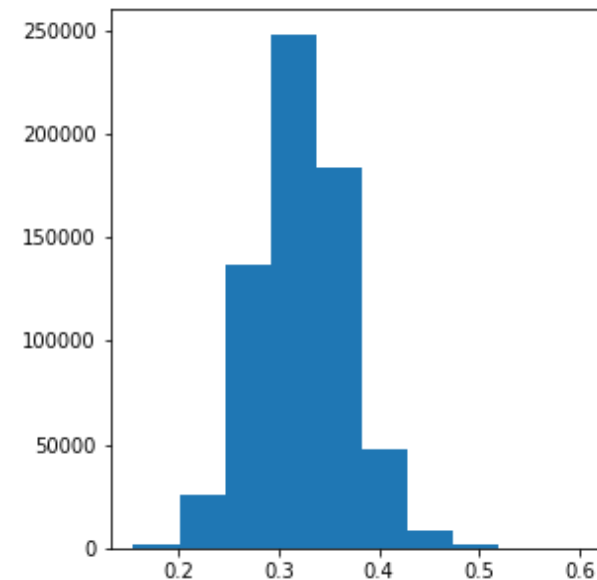
※ 단, 가중치는 추후 조정 가능

2. 기본 센서 설치

불확실성 지표 EDA



<불확실성 지표 HEAT MAP>

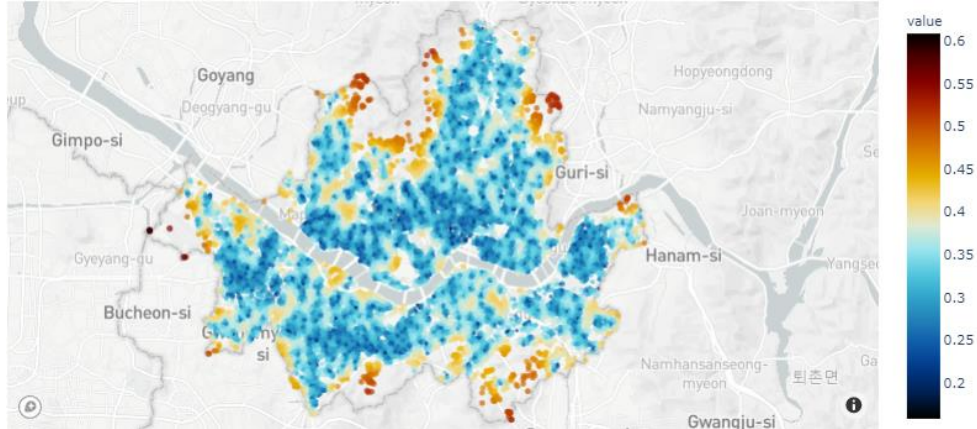


<불확실성 지표 히스토그램>

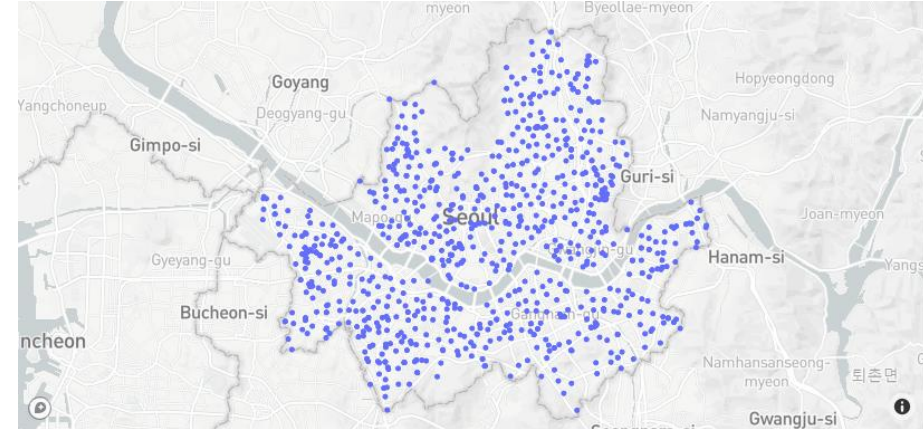
- 주요 도심의 경우 불확실성이 낮게 분포
- 주로 측정 센서가 부족한 외곽지역에서 불확실성이 높게 나타남
- 한강을 주변으로 불확실성이 낮지는 않은 것을 확인

2. 기본 센서 설치

MCLP(Maximal Covering Location Problem)를 통한 최적화 해결



<불확실성 지표 HEAT MAP> → MCLP 가중치로 사용

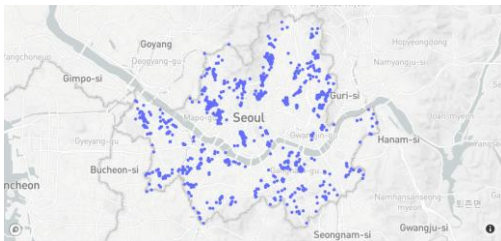


<가중치를 사용하지 않은 위치 선정>

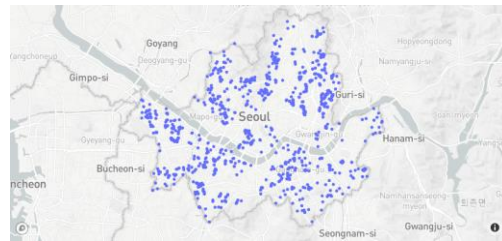
<가중치를 사용한 위치 선정>

가중치를 적당히 반영하며 coverage가 70%이상 되도록 위치 선정

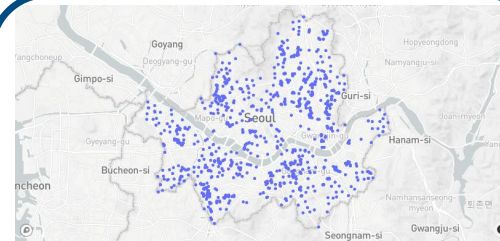
※ 추후 전문가에 의해 조절 가능



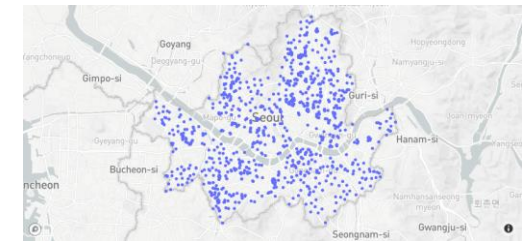
M : 1500
Coverage : 0.4184



M : 2000
Coverage : 0.5377



M : 3000
Coverage : 0.7245



M : 4000
Coverage : 0.8451

<최종 기본 센서 설치 위치 선정>

Part 3,

추가 센서 설치



3. 추가 센서 설치

추가 센서 위치 선정 프로세스

가중치 구하기

- 각 센서별 관련 데이터를 사용
- 데이터별 방법으로 후보지의 가중치 출력

MCLP 모델 적합

- MCLP 모델 적용
- 적용된 값은 아래와 같음
K (추출 개수) : 200
radius (Cover 영역) : 500m
M (Weight 고려 정도) : 700

추가 센서 위치 선정

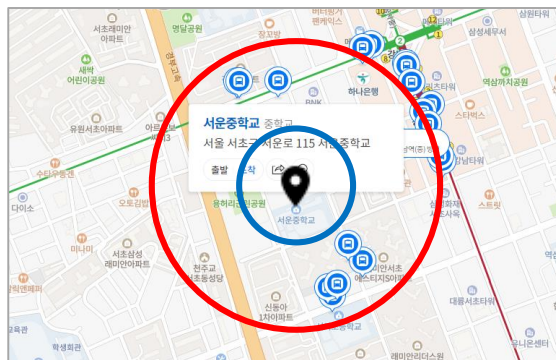
- MCLP 모델링 결과 사용

3. 추가 센서 설치

추가 센서 위치 선정 - 소음(#1)

- 가중치 구하기

- 소음은 혼잡도와 연관이 있고, 이를 알아내기 위해 버스, 지하철 승하차 인원 데이터를 사용
- 사용 데이터
 - ✓ 8. 서울시_버스정류소_공간정보.csv
 - ✓ 9. 서울시_버스정류소별_승하차이력.zip
 - ✓ 10. 서울시_지하철역_공간정보.geojson
 - ✓ 11. 서울시_지하철역별_승하차이력.csv
- 후보지를 기준으로 아래의 경우 승하차 인원의 합을 사용
 1. 1000m 내에 지하철 역이 있을 때
 2. 300m 내에 버스 정류장이 있을 때



주소	카테고리	거리	승하차 인원
강남역	지하철	859m	201,000
교대역	지하철	1,515m	137,000
롯데 칠성	버스	150m	500
레미안 에스티지	버스	160m	210

주소	가중치
서운중학교	201,710
...	...

3. 추가 센서 설치

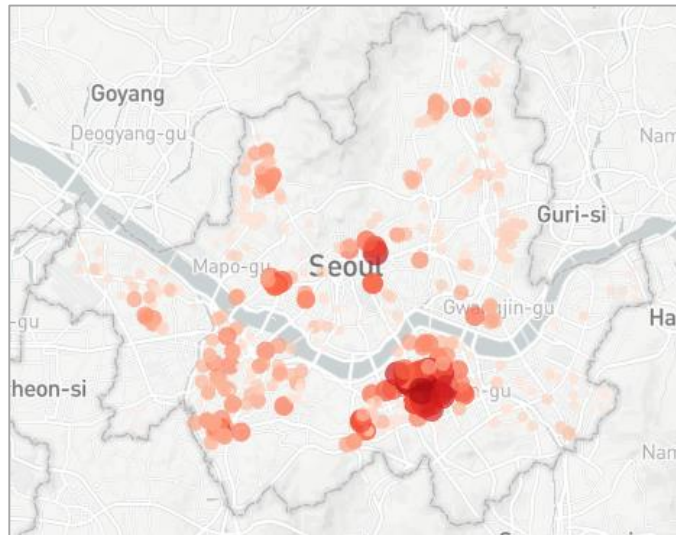
추가 센서 위치 선정 - 소음(#1)

- MCLP 모델 적합

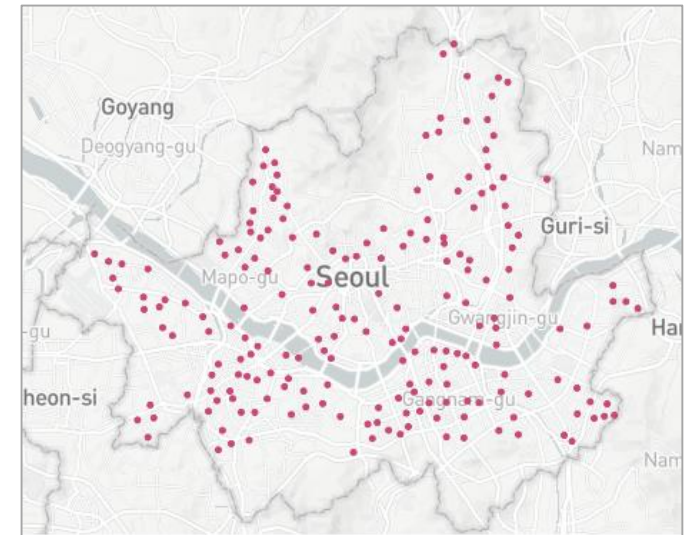
- 지하철 역과 지하철 역 근처 버스 정류장의 승하차 인원이 많음 → 지하철 역 부근의 소음이 높음
- 후보지 기준으로 소음의 가중치를 보았을 때, 강남역 부근과 종로 부근에서 가중치가 큰 값을 나타냄

이름	카테고리	승하차 인원
강남역	지하철	4,084,668
신림역	지하철	3,158,649
구로디지털단지역	지하철	2,940,770
홍대입구역	지하철	2,691,741
...	지하철	...
당산역	버스	4,295
신림역	버스	3,433
신림역4번출구	버스	3,416
가산디지털단지역	버스	3,082
...	버스	...

<사용 데이터>



<후보지별 가중치 Map Box>



<추가 센서 위치 선정 결과>

3. 추가 센서 설치

추가 센서 위치 선정 - 풍향 풍속(#2)

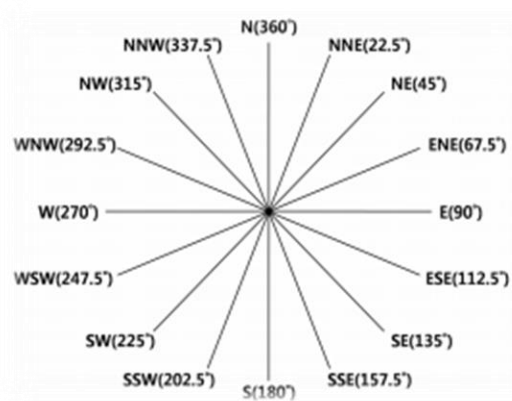
- 가중치 구하기

• 사용 데이터

- ✓ 5. 서울시_일별평균_대기오염도이력.csv
- ✓ 6. 서울시_대기오염도_측정지점.csv

• 풍향, 풍속의 경우, 해당 값의 변동량이 큰 곳에 센서를 설치하는 것이 바람직하다고 판단

- ✓ 풍속 : 일별 평균 풍속의 표준편차를 가중치로 사용
- ✓ 풍향 : 일별 데이터를 기준으로 풍향이 변하는 각도 계산 > 각도의 표준편차를 가중치로 사용



일자	위치	WD	WD_rad	WD_diff
20200501	중랑구	N	360.0	0.0
20200502	중랑구	WNW	292.5	45.0
20200503	중랑구	ENE	67.5	135.0
20200504	중랑구	NW	315.0	112.5
20200505	중랑구	SE	135.0	180.0

1. WD_rad : WD를 기준으로 각도 정의
2. WD_diff : WD_rad를 기준으로 각도의 변화량 계산

위치	가중치
중랑구	48.714

3. 추가 센서 설치

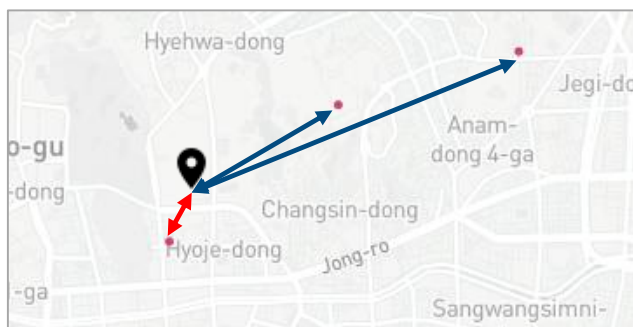
추가 센서 위치 선정 - 풍향 풍속(#2)

- 가중치 구하기

- 풍속과 풍향의 값을 각 속성별 가중치로 곱하여 MCLP 모델에 이용할 하나의 가중치를 생성



- 후보지 별 가장 가까운 관측치의 값을 가중치로 사용
 - ✓ 아래의 경우 빨간색 화살표로 표시된 관측소의 가중치를 가짐

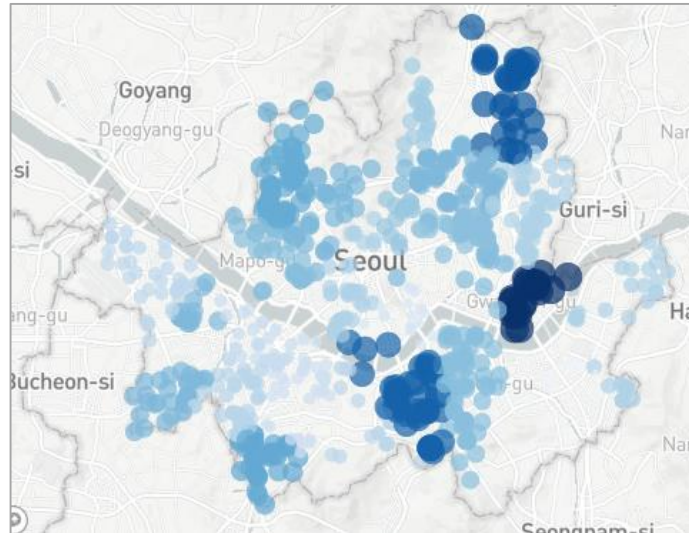


추가 센서 위치 선정 - 풍향 풍속(#2)

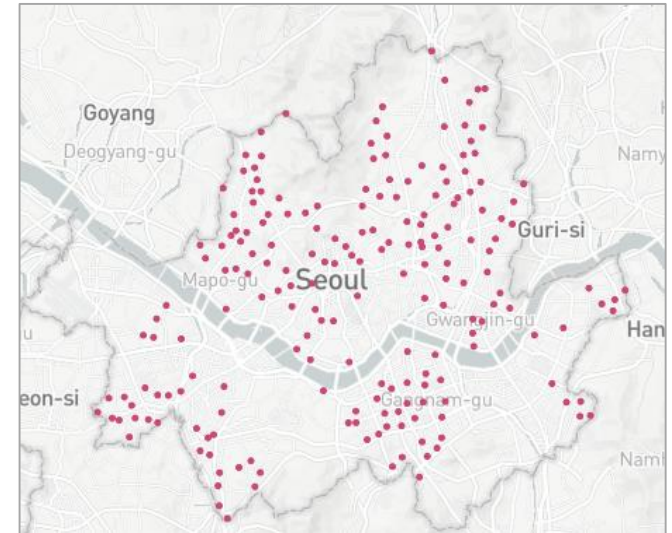
- MCLP 모델 적합
 - 후보지 기준으로 풍향, 풍속의 가중치를 보았을 때, 광진구, 노원구, 서초구의 가중치가 큰 값을 나타냄

이름	가중치
광진구	0.864436
노원구	0.760012
서초구	0.7356
금천구	0.5494
은평구	0.54294

<가중치가 높은 5개 관측소>



<후보지별 가중치 Map Box>

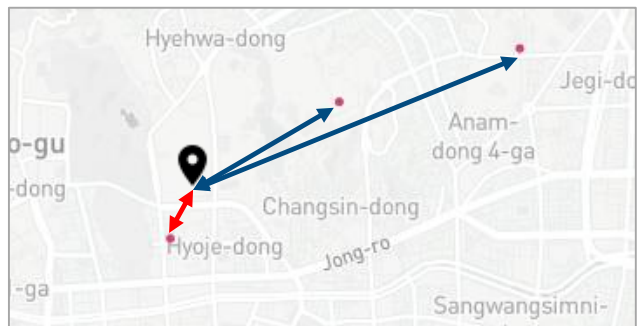


<추가 센서 위치 선정 결과>

3. 추가 센서 설치

추가 센서 위치 선정 - 오존(#3)

- 가중치 구하기
 - 사용 데이터
 - ✓ 5. 서울시_일별평균_대기오염도이력.csv
 - ✓ 6. 서울시_대기오염도_측정지점.csv
 - 후보지 별 가장 가까운 관측치의 값을 가중치로 사용
 - ✓ 아래의 경우 빨간색 화살표로 표시된 관측소의 가중치를 가짐



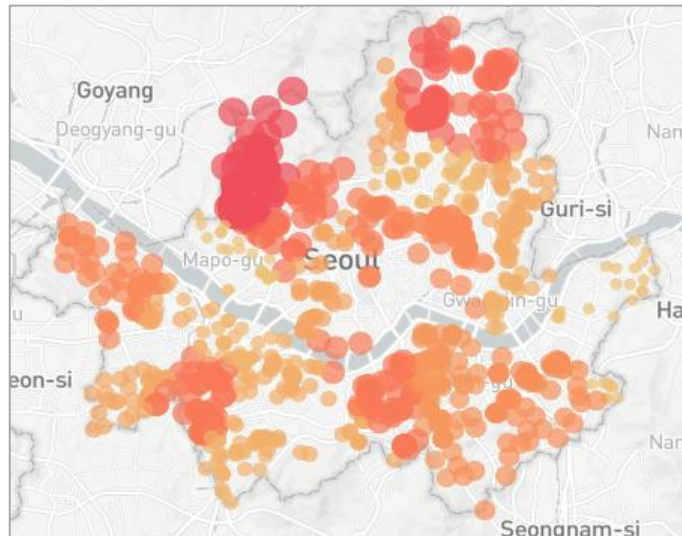
3. 추가 센서 설치

추가 센서 위치 선정 - 오존(#3)

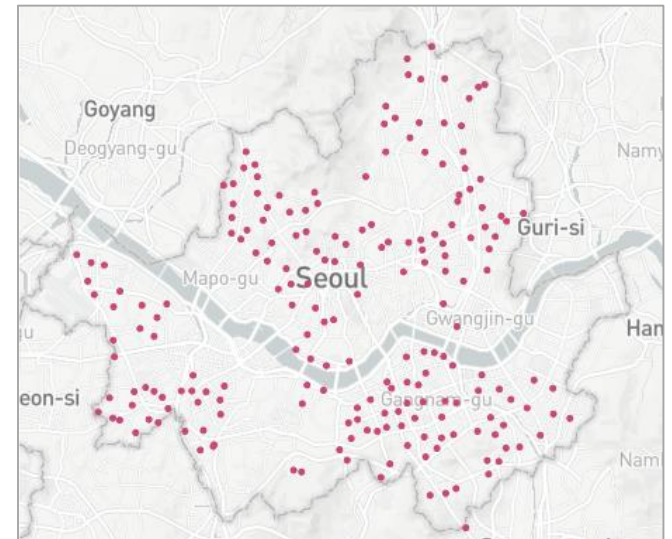
- MCLP 모델 적합
 - 후보지 기준으로 오존의 가중치를 보았을 때, 은평구, 도봉구, 노원구의 가중치가 큰 값을 나타냄

이름	가중치
은평구	0.030378
도봉구	0.029068
노원구	0.027932
서대문구	0.027759
구로구	0.027485

<가중치가 높은 5개 관측소>



<후보지별 가중치 Map Box>

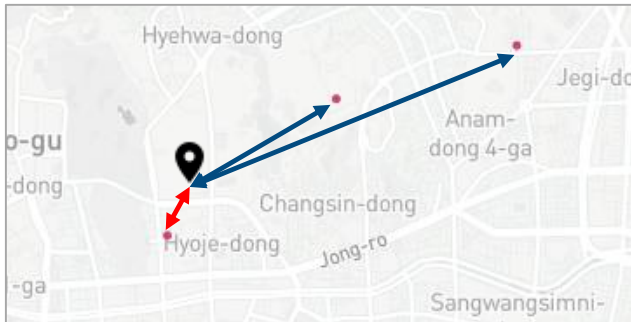


<추가 센서 위치 선정 결과>

3. 추가 센서 설치

추가 센서 위치 선정 - 탄소산화물, 질소산화물, 이산화황(#4)

- 가중치 구하기
 - 사용 데이터
 - ✓ 5. 서울시_일별평균_대기오염도이력.csv
 - ✓ 6. 서울시_대기오염도_측정지점.csv
 - 각 속성에 정규화 및 가중치를 곱해 MCLP에 사용할 하나의 가중치를 생성
 - ✓ 대기오염도 이력 데이터 내 NO, NO2, NOx, CO, SO2 5개의 속성이 존재
 - ✓ 풍향, 풍속 가중치를 생성하는 방법과 같음
 - 후보지 별 가장 가까운 관측치의 값을 가중치로 사용
 - ✓ 아래의 경우 빨간색 화살표로 표시된 관측소의 가중치를 가짐



3. 추가 센서 설치

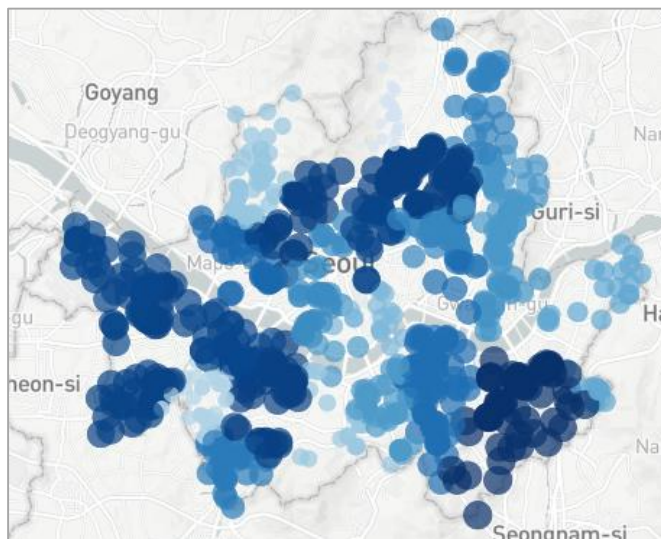
추가 센서 위치 선정 - 탄소산화물, 질소산화물, 이산화황(#4)

- MCLP 모델 적합

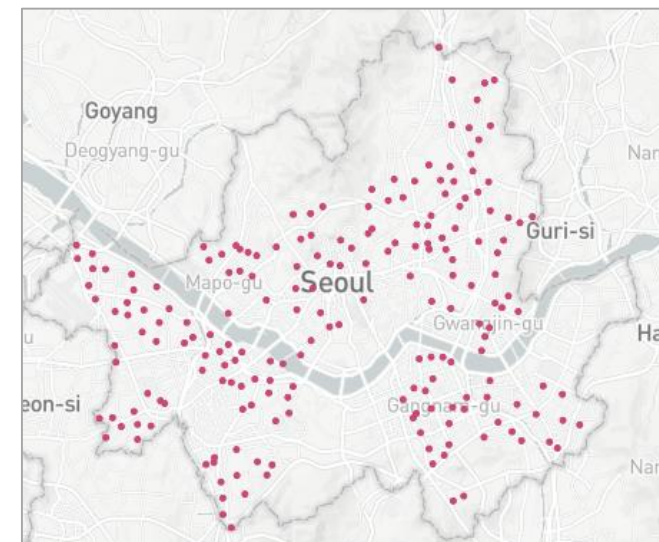
- 후보지 기준으로 오존의 가중치를 보았을 때, 송파구, 관악구, 서대문구의 가중치가 큰 값을 나타냄

이름	가중치
송파구	0.327652
관악구	0.322246
서대문구	0.322156
성북구	0.321908
양천구	0.321334

<가중치가 높은 5개 관측소>



<후보지별 가중치 Map Box>

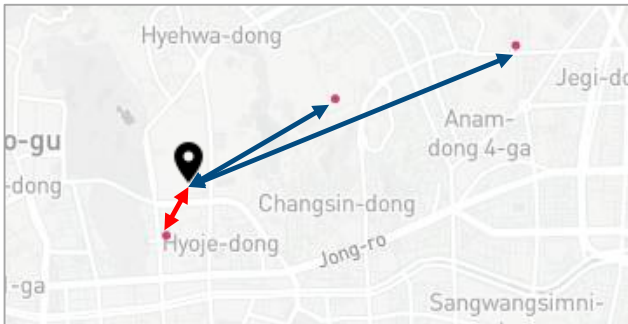


<추가 센서 위치 선정 결과>

3. 추가 센서 설치

추가 센서 위치 선정 - 암모니아, 황화수소가스(#5)

- 가중치 구하기
 - 암모니아, 황화수소가스는 해당 데이터가 없어서 주어진 데이터 중 가장 관련이 있는 종합 가스 소비량을 가중치로 적용
 - 사용 데이터
 - ✓ 14. 서울시_법정경계(읍면동).geojson
 - ✓ 28. 서울시_건물에너지(가스).zip
 - 후보지 별 가장 가까운 관측치의 값을 가중치로 사용
 - ✓ 아래의 경우 빨간색 화살표로 표시된 관측소의 가중치를 가짐



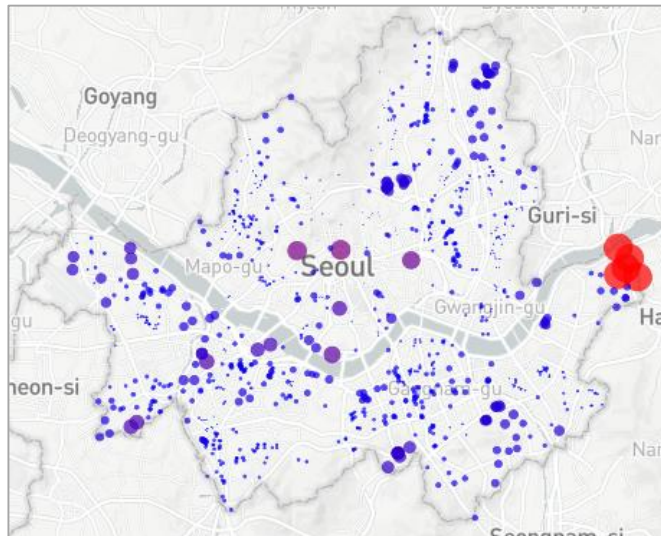
3. 추가 센서 설치

추가 센서 위치 선정 - 암모니아, 황화수소가스(#5)

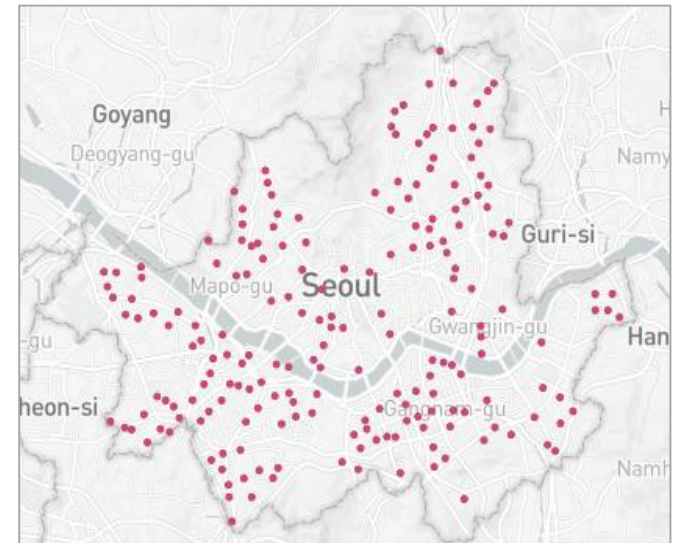
- MCLP 모델 적합
 - 후보지 기준으로 오존의 가중치를 보았을 때, 종로1가, 장교동, 오쇠동의 가중치가 큰 값을 나타냄

이름	가중치
종로1가	682959.6
장교동	413965.1
오쇠동	396169.8
강일동	388418.6
홍파동	335528.5

<가중치가 높은 5개 관측소>



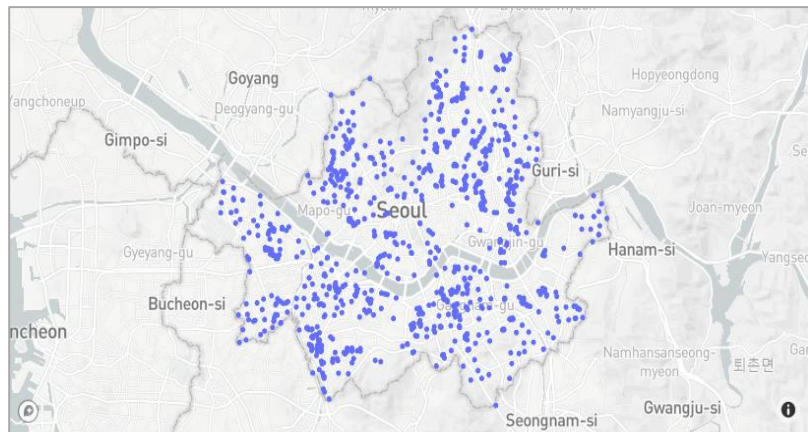
<후보지별 가중치 Map Box>



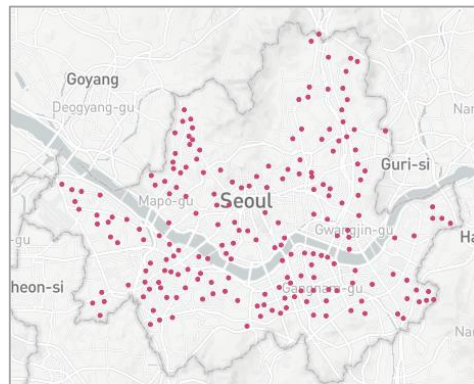
<추가 센서 위치 선정 결과>

최종 결과

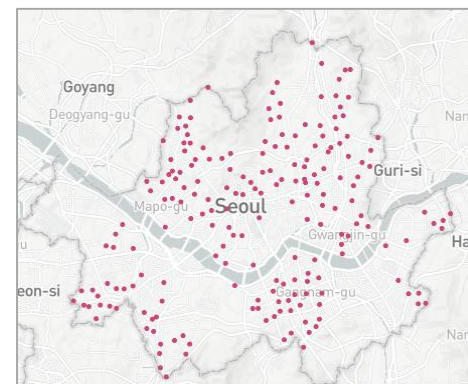
순번	설치위치 주소	위도(X)	경도(Y)	추가 센서#1	추가 센서#2	추가 센서#3	추가 센서#4	추가 센서#5
0	서울특별시 서초구 신원동	37.43053543	127.0711442	0	0	1	0	0
1	서울특별시 은평구 진관동	37.65799687	126.9469561	0	1	0	0	0
2	서울특별시 강동구 강일동	37.57658282	127.1676644	0	0	0	0	0
3	서울특별시 구로구 항동	37.47433833	126.818004	0	0	0	0	0
...



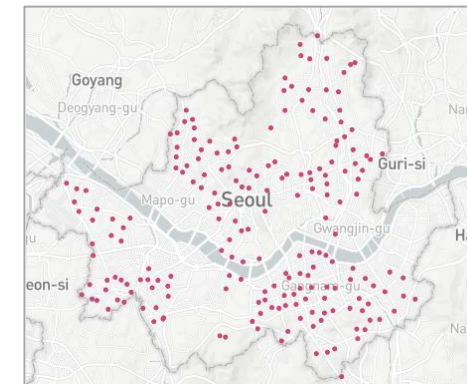
기본센서



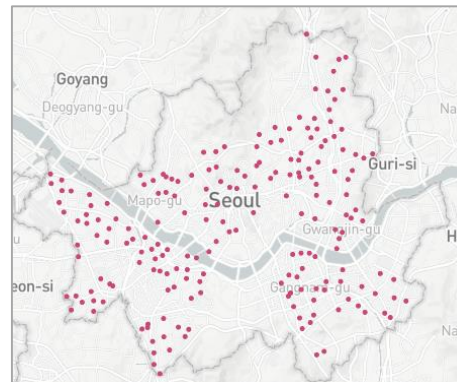
추가 센서#1



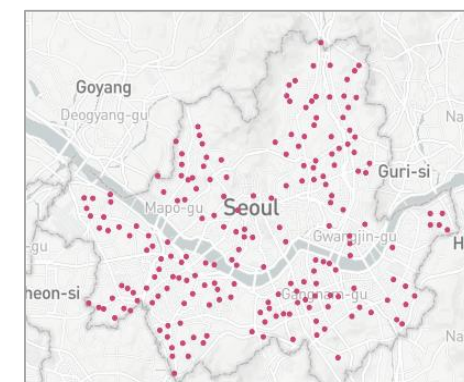
추가 센서#2



추가 센서#3



추가 센서#4



추가 센서#5

A close-up photograph of a brown leather bag with a strap, resting on a dark blue textured surface. Inside the bag, a black smartphone, a brown notebook with a red elastic band, a red and silver pen, and a pair of black-rimmed glasses are visible.

Part 4,

활용 방안

4. 활용방안

추가된 측정 항목별 활용 서비스 예시

- S-DoT 사업을 통한 실시간 측정 항목 데이터는 다른 공공데이터 및 민간데이터와 합쳐질 때, 시너지 효과를 낼 수 있을 것으로 판단

센서종류	측정항목	활용방안
기본 센서	미세먼지, 온도, 습도, 조도, 자외선	기초 자료로서 다른 데이터와 합하여 시너지를 내거나, 다양한 연구에서 올바르게 활용될 수 있도록 QC 작업 프로세스 개발과 추가 필요한 지점의 최신화
추가 센서 #1	소음	인접 건설 공사로 인한 (소음 및 진동) 손해배상 근거 확보 및 모니터링
추가 센서 #2	풍향, 풍속	바람길 지도(기준)를 이용한 산책로 및 자전거길 생성에 옵션으로 활용가능
추가 센서 #3	오존	노약자 및 영유아 대상, 특히 기관지염, 심장병, 폐기종, 천식 등의 환자의 경우 오존 위험 지수 PUSH 알림
추가 센서 #4	탄소산화물, 질소산화물, 이산화황	앞으로 있을 가능성이 높은 ESG의 '탄소 배출권 거래제' 실시 방안 및 '탄소세' 도입 방안의 정책기초활용 및 모니터링에 활용
추가 센서 #5	암모니아, 황화수소가스	악취와 관련된 데이터로 산책로 및 자전거길 생성에 옵션으로 활용 가능, Virtual Seoul System(기준)

The background of the slide is a deep blue night sky filled with stars. A bright, hazy band of the Milky Way galaxy stretches diagonally from the upper right towards the center. In the lower third of the image, there is a dark, silhouetted mountain range. A semi-transparent dark blue rectangle is centered horizontally, containing the text 'Q & A'.

Q & A

Spatial_Analysis