

(안양시) 도시데이터 수집을 위한 IoT센서 설치 위치 선정

2020. 11. 03

조수경

분석 목적

- 미세먼지 저감 정책 효과 검증 및 도로변 미세먼지 발생량 모니터링을 위한 미세먼지 측정 IoT센서 설치 위치 최적화
- 도시 정책에 활용될 수 있는 IoT복합센서 설치 위치 최적화

분석 프로세스

1. Data Preprocessing

- 제공받은 데이터 중 분석 목적에 맞는 데이터 선택
- 분석을 위한 기준 Point 생성
- 기준 Point에 데이터 통합

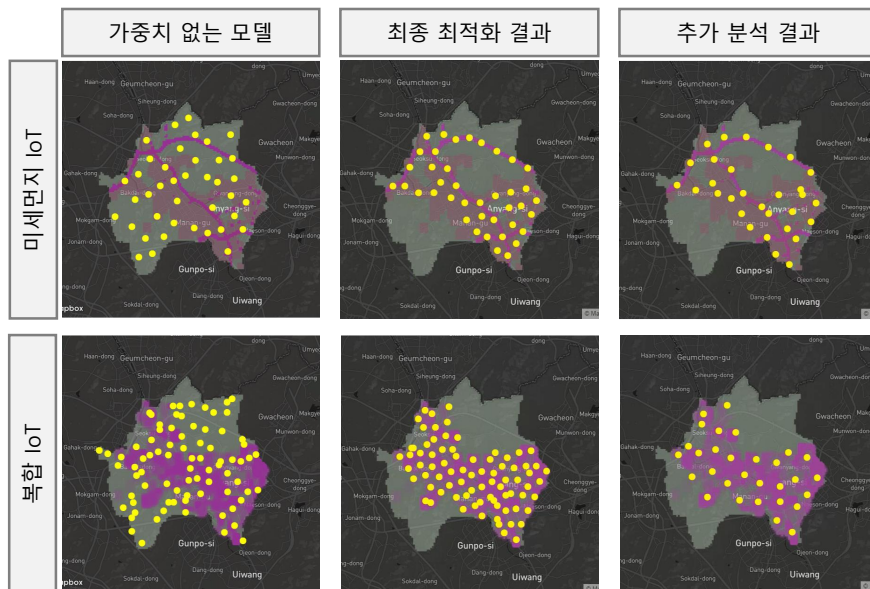
2. EDA

- 기준 Point에 할당된 정보 시각화
- 시사점 도출

3. 최적화 문제 정의

- 최적화 모델 선정
 - ✓ Maximal Covering Location Problem¹⁾
- 최종 최적화 모델 정의
 - ✓ 목적함수: 특정 지역을 주어진 IoT 기기 대수로 최대 cover
 - ✓ 결정 변수 분석 : covered distance 결정

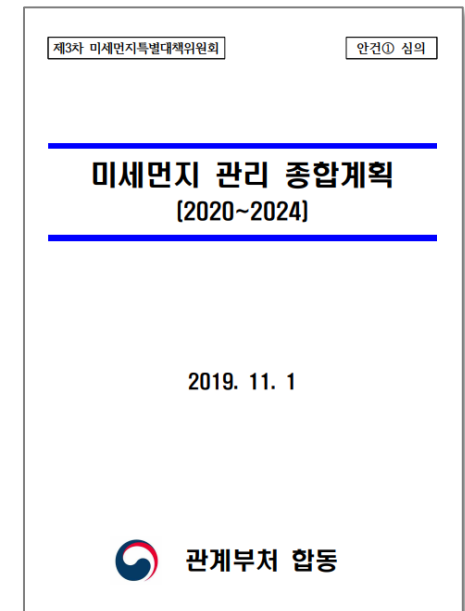
분석 결과



- 주어진 IoT 기기로 가장 효율적인 설치 위치 제안
- 추가 분석에서 임의의 covered distance를 가정하여 효율적인 IoT 대수 제안
 - ✓ 한 대의 IoT 기기가 700m를 cover 할 수 있다면 각각 30대만으로 타깃 지역을 90% cover 가능
- 활용 방안 제시
 - ✓ IoT 기기의 성능과 가격, 예산에 따른 최적화 위치 도출
 - ✓ 미세먼지나 복합 IoT 이외 특수 목적 기기 설치에 활용 가능 (밤길 안심 귀가 CCTV, 방범용 CCTV 추가 설치 등)

국가 정책미세먼지 관리 종합계획

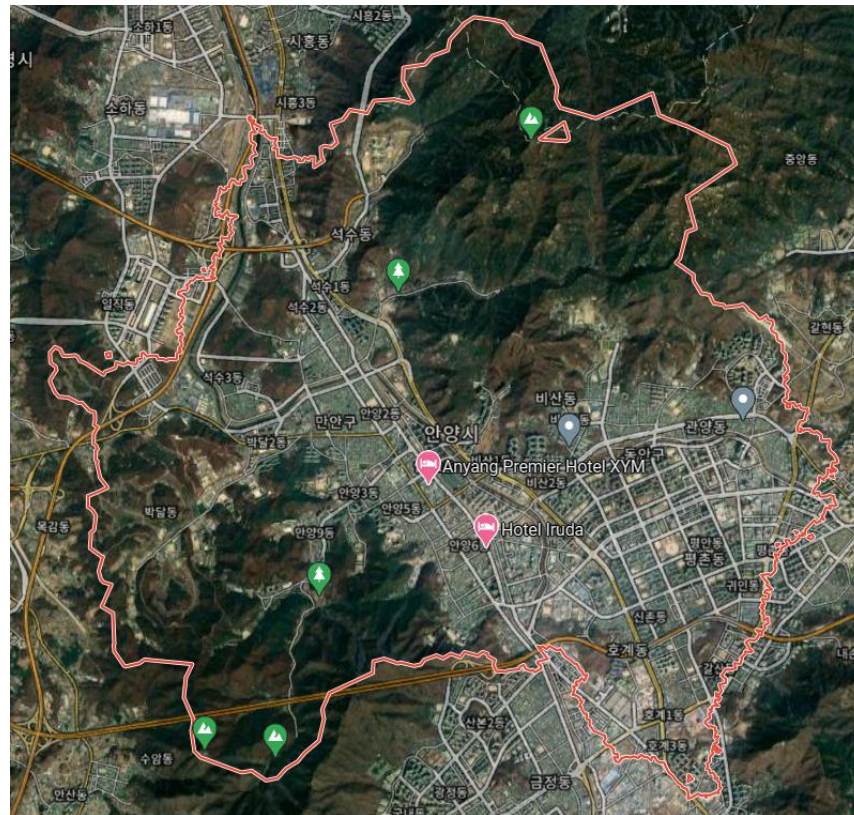
- ☑ 정부는 미세먼지 저감 정책을 위해 관계부처와 함께 미세먼지 관리 종합계획 수립
 - ☑ 국내 미세먼지는 초미세먼지·황산화물은 사업장, 질소산화물은 수송 부문에서 다량 배출
 - 효과적인 저감 정책을 위해서는 미세먼지가 가장 많이 발생하는 산업단지와 도로변의 면밀한 모니터링이 필요
 - 경기도는 특히 질소산화물이 다량 배출되며, 이는 자동차 수송에 의해 발생됨
- ➔ 미세먼지 IoT 센서는 공장 주변과 교통량이 많은 도로변에 설치
- ☑ 미세먼지는 호흡기에 침착 또는 체내에 흡수되어 건강 영향 발생
 - 실내 공기질 관리와 취약계층에 대한 집중 보호 조치 마련이 요구
- ➔ 복합 IoT센서 또한 미세먼지 측정이 가능하므로 취약계층이 많은 곳에 우선 설치



안양시 지형 특징

▼ 많은 산지

- 석수동과 비산동, 박달동, 안양9동 일대
- 산지는 IoT 설치 시 후순위가 되어 할 것으로 판단



1. Data Preprocessing


인구 현황과 공장 등록 현황

데이터

- df_01: 01.안양시_격자별인구현황(전체), df_02: 02.안양시_격자별인구현황(고령), df_03: 03.안양시_격자별인구현황(생산가능인구), df_04: 04.안양시_격자별인구현황(유소년), df_10: 10.안양시_공장등록현황

전처리 방법

- 1. 격자 데이터의 geometry 에서 위도 경도 최대 최소 값을 컬럼으로 분리



df_01			df_pop						
gid	geometry	Val_total	gid	geometry	Val_total	lon_min	lon_max	lat_min	lat_max
다사 448339	MULTIPOLYGON (((126.87631 37.40257, 126.87630 ...	0	다사 448339	MULTIPOLYGON (((126.87631 37.40257, 126.87630 ...	0	126.8463	126.8763	37.4025	37.4325

- 2. “gid” 컬럼을 공통으로 가지는 데이터 셋을 모두 결합

- Val_total: 전체인구, val_senior: 고령인구, val_work: 생산가능 인구, val_junior: 유소년, fac_cnt: 공장등록 수

df_pop										
gid	geometry	Val_total	Val_senior	Val_work	Val_junior	Fac_cnt	lon_min	lon_max	lat_min	lat_max
다사 448339	MULTIPOLYGON (((126.87631 37.40257, 126.87630 ...	150	123	10	12	5	126.8463	126.8763	37.4025	37.4325

- 3. 모든 값 정규화 하여 시각화

- 정규화 값: 값 / 최대값

1. Data Preprocessing

주요시설 위치

☑ 데이터

- df_06: 06.안양시_학교현황, df_07: 07.안양시_유치원현황, df_08: 08.안양시_복지회관_보건소현황, df_09: 09.안양시_병원정보, df_10: 10.안양시_공장등록현황, df_11: 11.안양시_공장등록현황_격자별, df_13: 13.안양시_버스정류장 위치정보, df_14: 14.안양시_버스정류장별 이용객수, df_15: 15.안양시_지하철역 위치정보

☑ 전처리 방법

- 1. 주요 위치 데이터의 위도 경도 $\pm 300m$ 에 해당하는 행을 df_pop에서 추출, 1 부여
 - (WGS84 좌표계)약 $1m = (1/88.74/1000)$

df_06

지역명	학교명	Lon	Lat
경기도 안양시 동안구	관양고등학교	126.960063	37.408497

```
# 학교 근방 300m 체크
df_pop['school'] = 0
for i in range(len(df_06)):
    location = df_pop[(df_pop['lon_max'] <= df_06.loc[i, 'lon'] + coord_lm * 300) &
                      (df_pop['lat_max'] <= df_06.loc[i, 'lat'] + coord_lm * 300) &
                      (-df_pop['lon_min'] <= -df_06.loc[i, 'lon'] + coord_lm * 300) &
                      (-df_pop['lat_min'] <= -df_06.loc[i, 'lat'] + coord_lm * 300)
                      ].index
    df_pop.loc[location, 'school'] = 1
```

df_pop

gid	geometry	Val_total	...	School
다사448339	MULTIPOLYGON (((126.87631 37.40257, 126.87630 ...	0	...	1
다사448340	MULTIPOLYGON (((126.87630 37.40347, 126.87629 ...	0	...	1
...

1. Data Preprocessing

승하차 수

데이터

- df_13: 13.안양시_버스정류장 위치정보, df_14: 14.안양시_버스정류장별 이용객수

전처리 방법

1. df_14에서 특정 날짜 (2019. 10. 22) 선택
2. df_14에서 정류소 ID별로 승차/ 하차 결과 합 (Python cod: groubby)
3. df_14의 “전체 승차 건수” 열과 “전체 하차 건수” 열을 합하여 “승하차” 열 생성
4. df_14와 df_13을 “정류소ID”로 통합하여 정류소의 위도 경도 정보 결합
5. df_pop에서 위도 경도 $\pm 300m$ 에 해당하는 행 추출, 승하차 값 할당
6. 승하차 값 정규화 컬럼 생성

df_13

정류소ID	STATION_NM	Lon	Lat
209000225	친목마을	126.960063	37.408497

df_14

정류소ID	일자	전체 승차 건수	전체 하차 건수
209000225	20191022	0	1

df_pop

gid	geometry	Val_total	...	승하차
다사448339	MULTIPOLYGON (((126.87631 37.40257, 126.87630 ...	0	...	127
다사448340	MULTIPOLYGON (((126.87630 37.40347, 126.87629 ...	0	...	123
...

1. Data Preprocessing

교통량

☑ 데이터

- df_17: 17.안양시_링크정보, df_18: 18.안양시_교통량정보

☑ 전처리 방법

- 1. df_18에서 특정 날짜 (2019. 10. 22) 추출
- 2. df_18에서 출근 시간 대와 업무시간 대 교통량 추출 (각각 7시, 15시로 가정)
- 3. df_18과 df_17을 "LINK_ID"로 통합하여 도로 geometry와 교통량 결합

df_17

LINK_ID	ROAD_NAME	geometry
2080000700	서울외곽순환고속도로	MULTILINESTRING ((126.93914 37.39807, 126.9438...

df_18

dt	tm	LINK_ID	Sum_probecount
2091022	7	2080000700	136

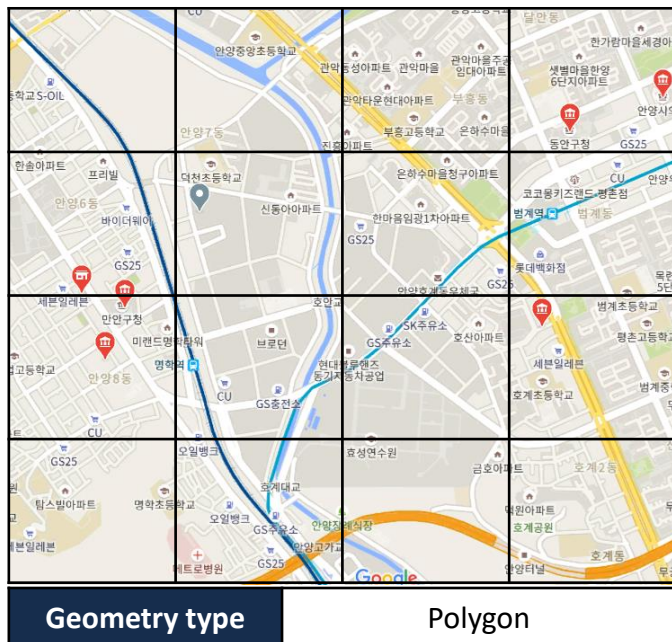
df_traffic_7

LINK_ID	ROAD_NAME	geometry	교통량07
2080000700	서울외곽순환고속도로	MULTILINESTRING ((126.93914 37.39807, 126.9438...	136
2080062700	제2경인고속도로	MULTILINESTRING ((126.93914 37.39807, 126.9438...	123
...

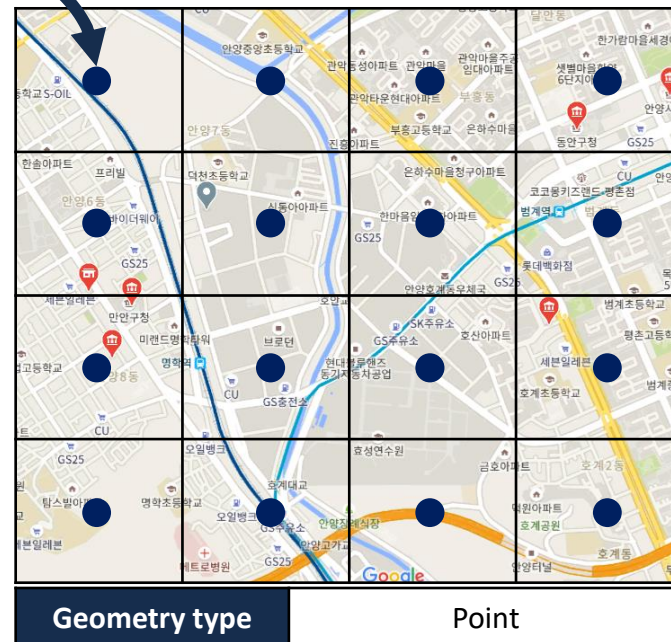
1. Data Preprocessing

인구 관련 요소와 공장 수 할당

✓ 기준 Point 생성



Point 할당 정보
인구 현황 수
고령 인구 수
생산가능인구 수
유소년 수
공장 수

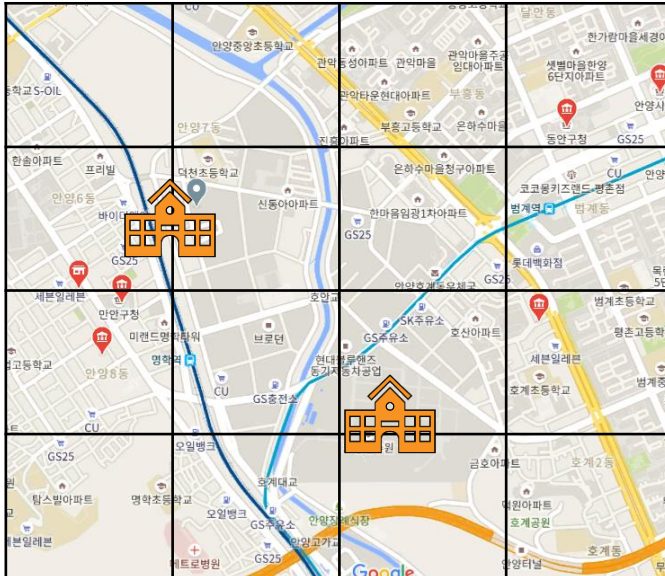


- 격자별 인구현황 관련 01~04 데이터와 공장 현황 관련 15 데이터는 'gid'이 공통이므로 한번에 할당
- 격자별 인구 현황의 grid 기준의 중심 (central point)를 **기준 Point**로 명명
- 기준 Point 마다 인구 현황과 공장 수 할당

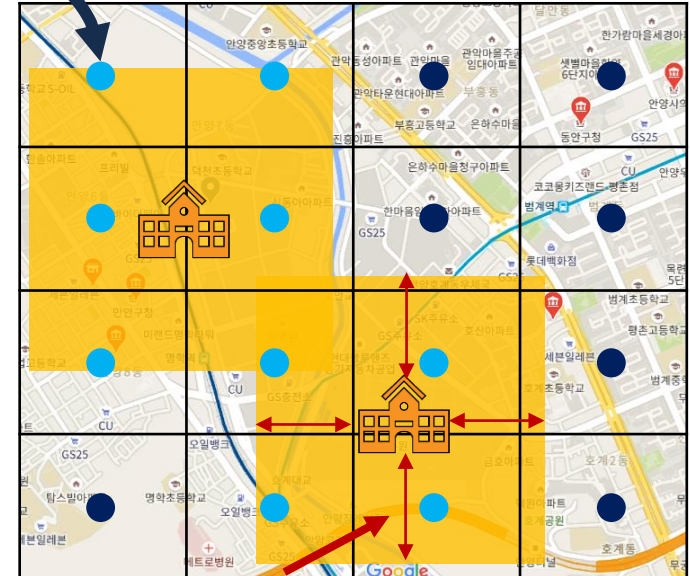
1. Data Preprocessing

주요 시설 요소 할당

☑ 주요시설 Point 관련 요소, 기준 Point에 할당



Point 할당 정보
인구 현황 수
고령 인구 수
생산가능인구 수
유소년 수
공장 수
학교
유치원
병원
보건소
지하철 역
승하차 수
공장



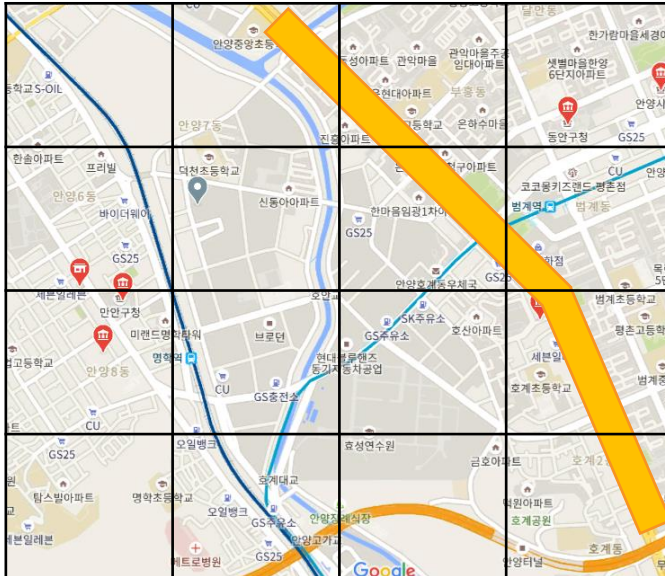
300m

- 주요 시설 Point의 300m 주변에 해당하는 기준 Point에 1 부여

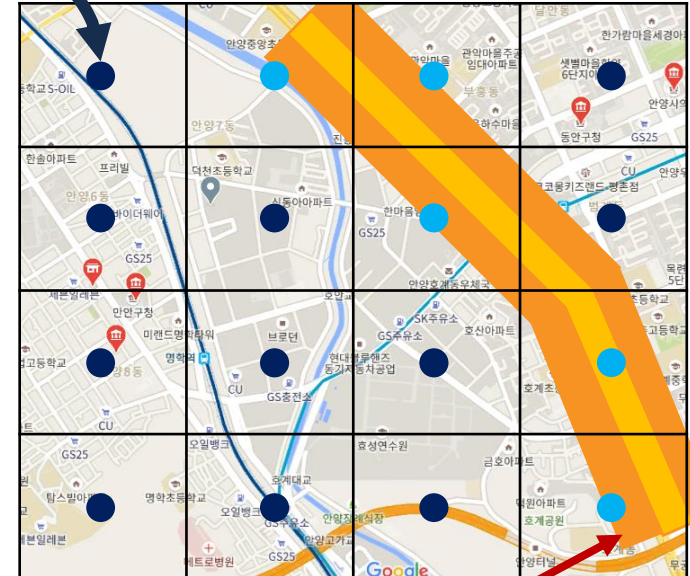
1. Data Preprocessing

교통량 할당

MultiLineString 관련 요소, 기준 Point에 할당



Point 할당 정보
인구 현황 수
고령 인구 수
생산가능인구 수
유소년 수
공장 수
학교
유치원
병원
보건소
지하철 역
승하차 수
공장
07시 교통량
15시 교통량



Buffer

- MultiLineString 정보에 buffer를 적용, 도로 근방 약 50 m에 포함되는 기준 point에 MultiLineString 정보 할당
- 할당된 정보: 07시 교통량, 15시 교통량

2. Exploratory Data Analysis

인구 현황 분석

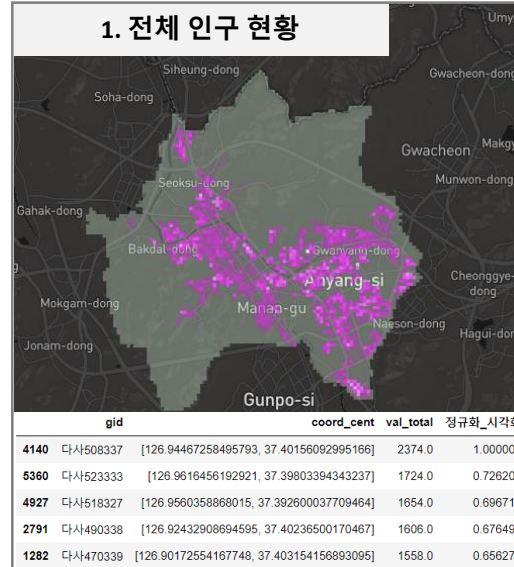
시각화

- 인구 적음  많음

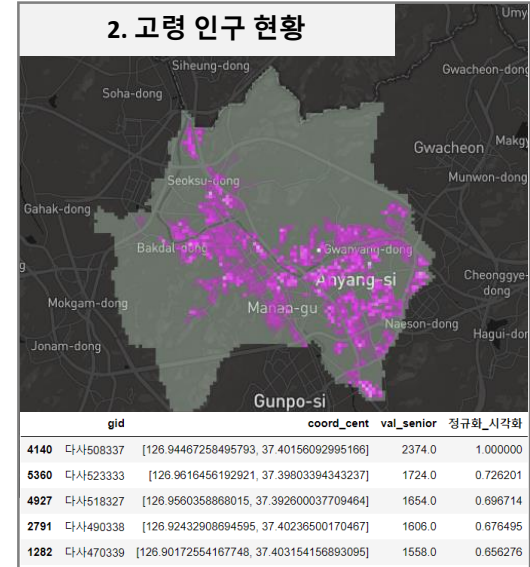
시사점

- 생산가능인구와 고령인구 패턴이 비슷함
 - 비산동에 가장 많은 인구가 있음
- 유소년인구는 호계동, 비산동 순으로 많음
- 산지를 제외하고 분포함을 확인,
인구 현황이 1명 이상 있는 곳을 추출,
복합 IoT 설치 후보지로 고려

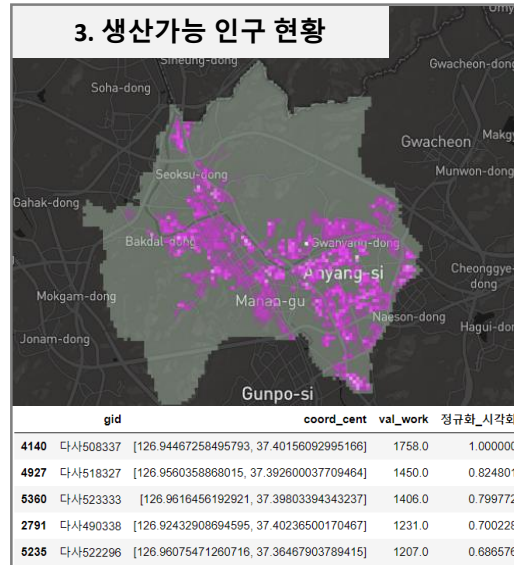
1. 전체 인구 현황



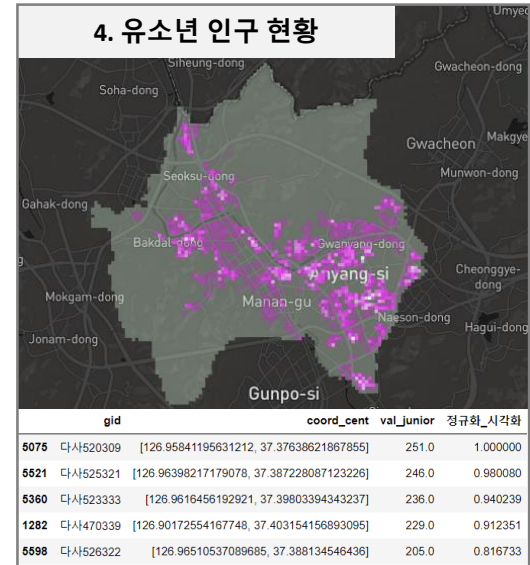
2. 고령 인구 현황



3. 생산가능 인구 현황



4. 유소년 인구 현황



2. Exploratory Data Analysis

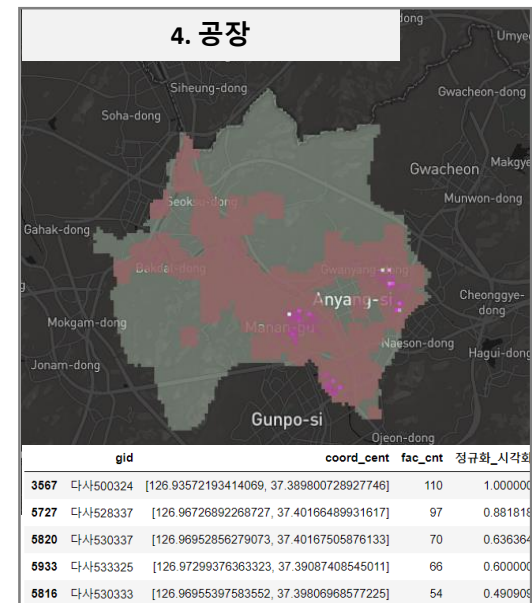
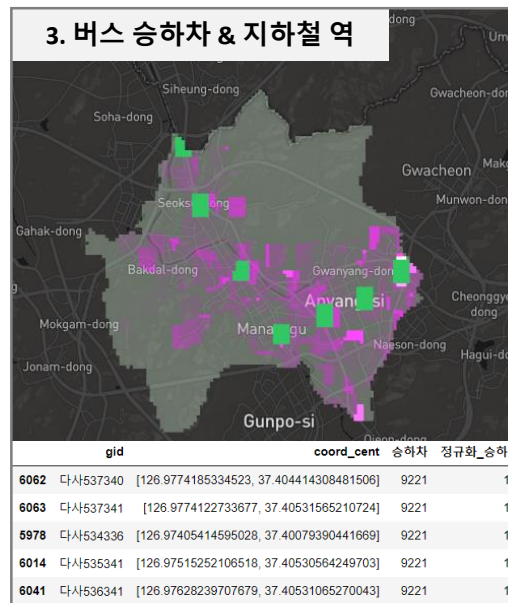
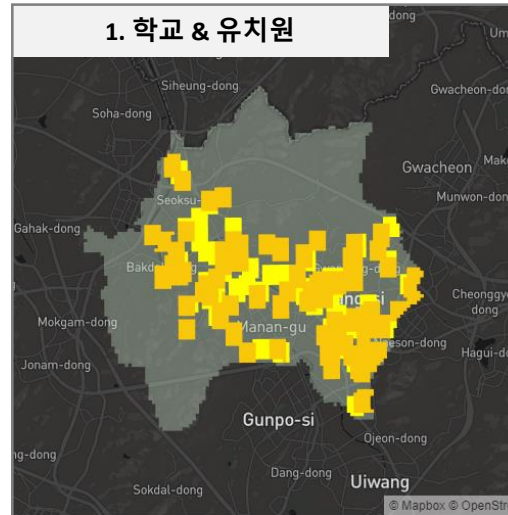
주요 시설 위치

시각화

- 적음 많음

시사점


- 학교와 유치원은 비슷한 곳에 분포
- 보건소와 병원은 비슷한 곳에 분포
- 버스 승하차가 가장 많은 곳은 **관양 1동**
- 지하철 역 근처는 승하차 수가 보통 많으므로, 유동인구가 많은 곳으로 기대
- 공장은 안양에 전반적으로 고르게 분포
 - 안양 7동, 관양 2동, 평촌동에 많이 분포
- 복합 IoT는 환경 데이터 뿐만 아니라 도시 데이터 수집 목적도 있으므로 인구 현황이 높은 곳 뿐만 아니라 유동인구가 높은 곳과 주요 시설 근처에 설치되어야 한다고 판단



2. Exploratory Data Analysis

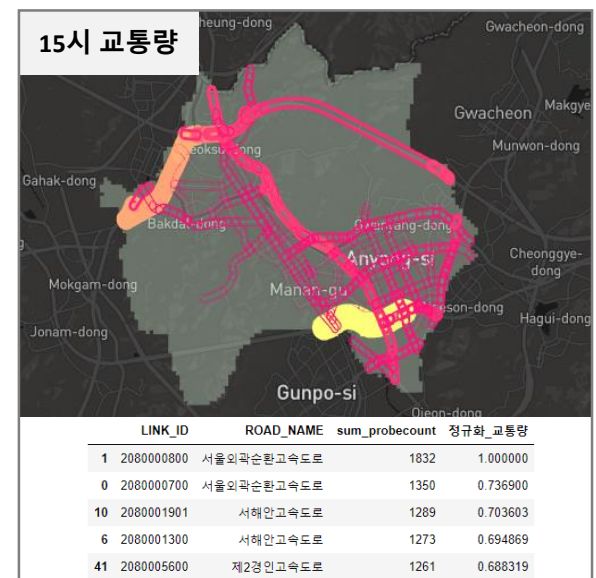
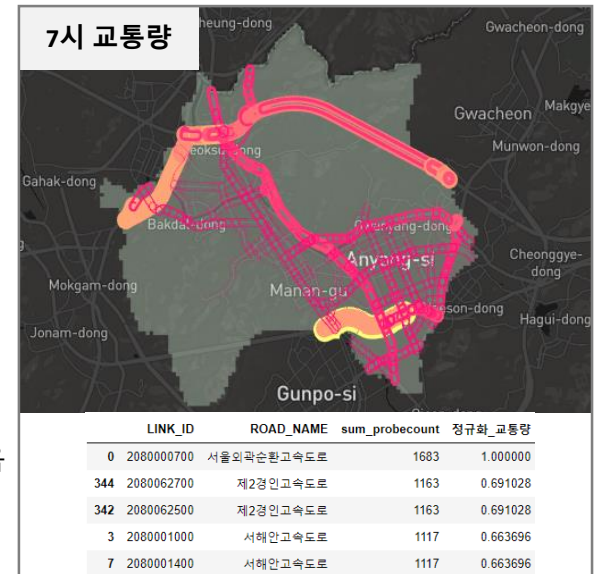
교통량 분석

시각화

- 교통량 많음  교통량 적음
- 선의 굵기: 교통량

시사점

- 출근 시간과 업무시간 패턴이 비슷함
 - 서울외곽순환고속도로와 제2 경인고속도로, 서해안고속도로에 교통량 많음
- 석수동, 비산동, 호계동, 평촌동 일대 교통량이 많음
- 미세먼지 모니터링은 도로 주변에 설치 되어야 효과적이라 생각됨

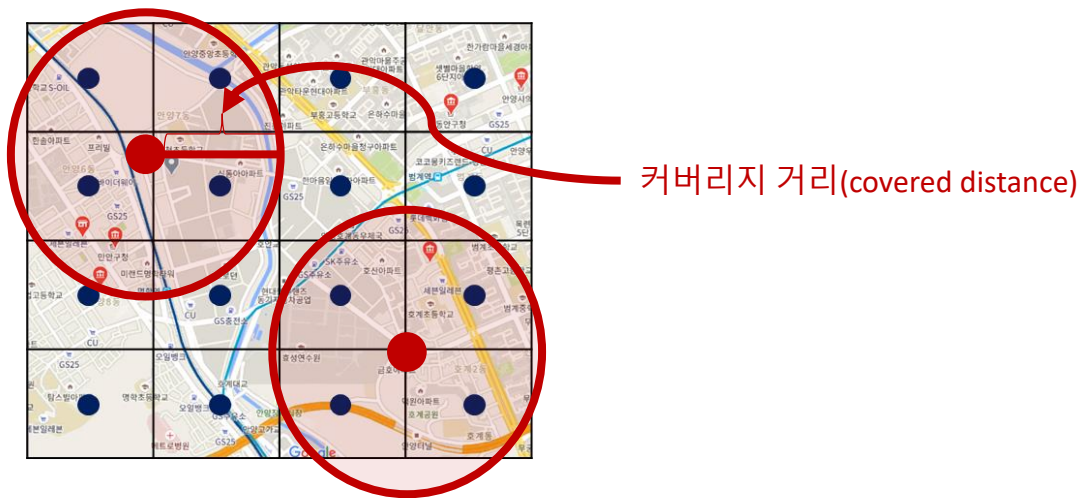


3. 최적화 문제 정의

최적화 모델 설명

✓ Maximal Covering Location Problem (MCLP)

- MCLP는 최대지역커버문제로 Mixed Integer Linear Programming 문제
- 설비가 cover하는 수요 (covered demand)의 합을 최대화 하면서 κ 개 설비를 세울 위치를 선정하는 문제
- 가정
 - 설비의 위치가 수요 발생 지점으로부터 일정 거리 Residual 이내에 수요를 커버함.
 - 이때 거리 Residual은 커버리지 거리(covered distance) 라고 함



3. 최적화 문제 정의

최적화 모델 문제 정의

✓ Maximal Covering Location Problem (MCLP)

- Formulation

Formulation	Mathematical statement
$\sum_{i \in I} w_i y_i \dots (1)$ $\text{s.t. } y_i \leq \sum_{j \in N_i} x_j \quad \text{for all } i \in I \dots (2)$ $\sum_{j \in J} x_j = K \dots (3)$ $x_j, y_i \in \{0, 1\} \quad \text{for all } i \in I, j \in J$	<ul style="list-style-type: none"> - 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 j에 설비가 설치되면 1, 그렇지 않으면 0 - y : 적어도 하나의 설비로 그 포인트가 커버가 되면 1, 그렇지 않으면 0 - w : 가중치
	Describe
	<ul style="list-style-type: none"> -(1) : 목적함수, 가중치 w인 수요 포인트를 최대한 많이 커버하게 해라 -(2) : 수요포인트 i는 설비 후보 지역이 커버하는 거리 안에서 적어도 하나 이상의 설비로 부터 커버가 된다. -(3) : 총 설치할 설비는 K개 이다.

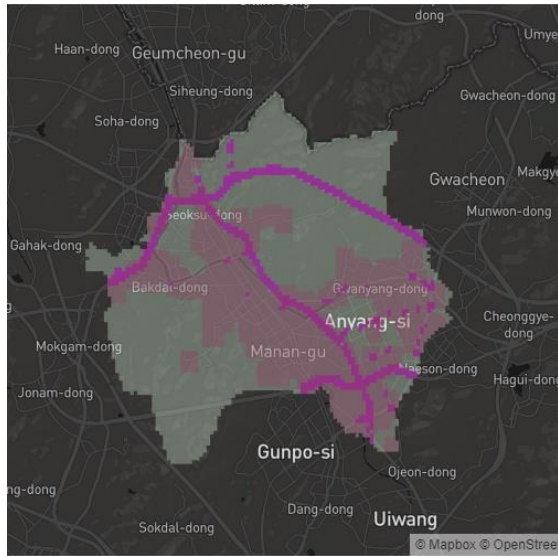
• 결정 변수

- y: 설치 후보 기준 Point
- K: IoT 설치 대수
- w : 가중치, 각 IoT 설치 목적에 맞는 가중치를 정의
- **Covered distance**: 한 설비가 cover 가능한 범위,
시나리오 분석을 통해 미세먼지 IoT 40개, 복합 IoT 80개를 가장 효과적으로 설치할 수 있는 범위 도출,
만약 실제 IoT가 cover 가능 한 범위가 주어진다는 가정하에 가장 효과적인 IoT 개수를 추가 분석 함

3. 최적화 문제 정의

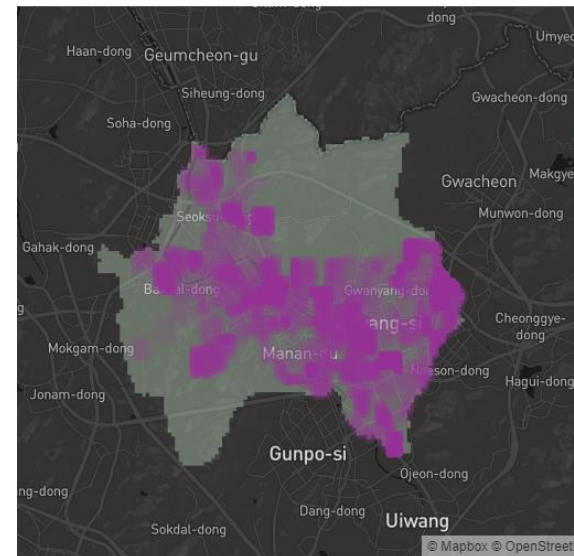
결정 변수 선정 (y, K, w)

미세먼지 IoT



- **y**: 도로 주변과 공장이 1개 이상 있는 Point
- **K**: 40
- **w**: 교통량과 공장등록 개수가 많을 수록 큼
 - 정규화된 07시 교통량 (정규화_교통량07)
 - 정규화된 15시 교통량 (정규화_교통량15)
 - 정규화된 공장등록 개수 (정규화_fac_cnt)

복합 IoT

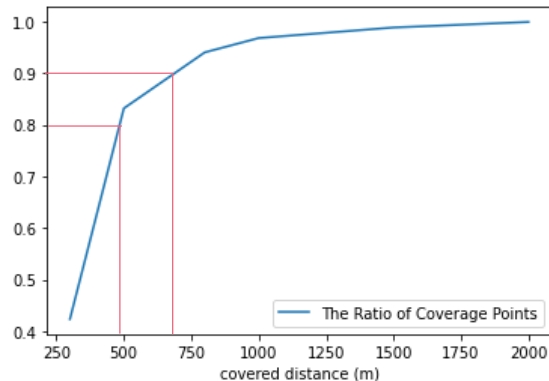


- **y**: 인구 현황과 승하차 기록이 1명 이상 있는 곳
- **K**: 80
- **w**: 다음과 같이 우선순위를 고려하여 가중치 개발
 - (공기 질 모니터링) 유소년 인구가 많은 곳 (정규화_junior*0.20)
 - (공기 질 모니터링) 고령 인구가 많은 곳 (정규화_senior*0.15)
 - (공기 질 모니터링) 학교 혹은 어린이집이 있는 곳 (kinder0.075, school0.075)
 - (공기 질 모니터링) 보건소 혹은 병원이 있는 곳 (pub_hos0.05, hos0.05)
 - (유동인구 모니터링) 승하차 인구가 많은 곳 (정규화_승하차*0.2)
 - (유동인구 모니터링) 지하철 역이 있는 곳 (sub*0.15)
 - (공기 질 모니터링) 인구가 많은 곳 (정규화_인구*0.05)

3. 최적화 문제 정의

미세먼지 IoT covered distance 분석

✓ (40대 일 때) Covered distance 에 따른 cover 가능성

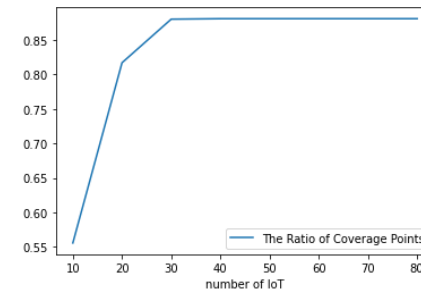


	Covered distance (m)	The Ratio of Coverage Points
0	300	0.423024
1	500	0.831831
2	800	0.940361
3	1000	0.968100
4	1500	0.988558
5	2000	0.999307

- 미세먼지IoT 40대를 설치할 경우, covered distance가 500m 이면 약 80%를 cover 할 수 있으며, 700m 이면 약 90%를 cover 할 수 있다.

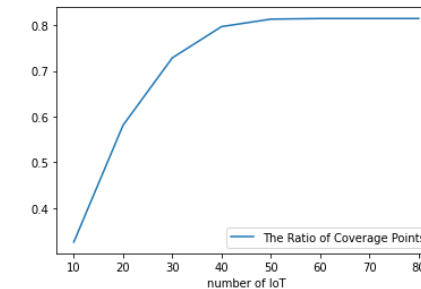
✓ IoT 대수에 따른 cover 가능성

- Covered distance 700m인 경우 90% 커버 가능



- 30대가 가장 효율적

- Covered distance 500m인 경우 80% 커버 가능



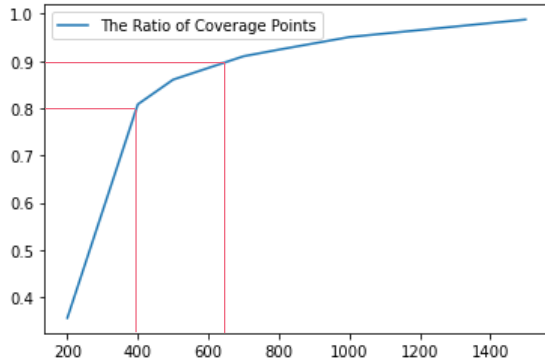
- 40대가 가장 효율적

covered distance가 500m일 때 40대로 80%를 cover 하는 것이 가장 효율적이라 예상됨

3. 최적화 문제 정의

복합 IoT covered distance 분석

☑ (80대 일 때) Covered distance 에 따른 cover 가능성

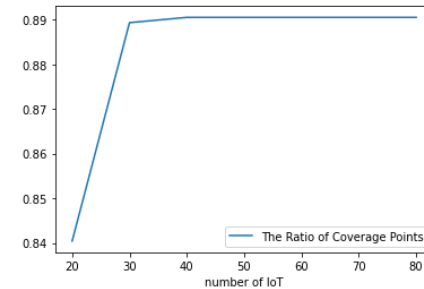


	covered distance (m)	The Ratio of Coverage Points
0	200	0.355977
1	400	0.808365
2	500	0.860575
3	700	0.910116
4	800	0.924058
5	1000	0.950756
6	1500	0.987837

- 복합IoT 80대를 설치할 경우, covered distance가 400m 이면 약 80%를 cover 할 수 있으며, 700m 이면 약 90%를 cover 할 수 있다.

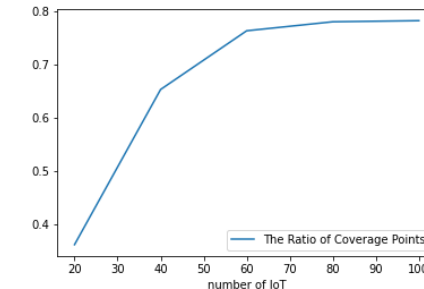
☑ IoT 대수에 따른 cover 가능성

- Covered distance 700m인 경우 90% 커버 가능



- 30대가 가장 효율적

- Covered distance 400m인 경우 80% 커버 가능



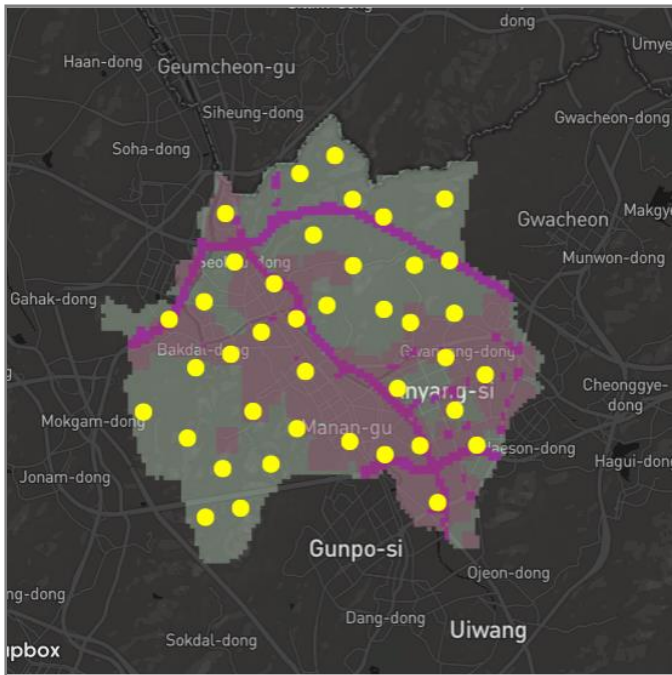
- 80대가 가장 효율적

covered distance가 400m일 때 80대로 80%를 cover 하는 것이 가장 효율적이라 예상됨

4. IoT 센서 설치 최적화 결론

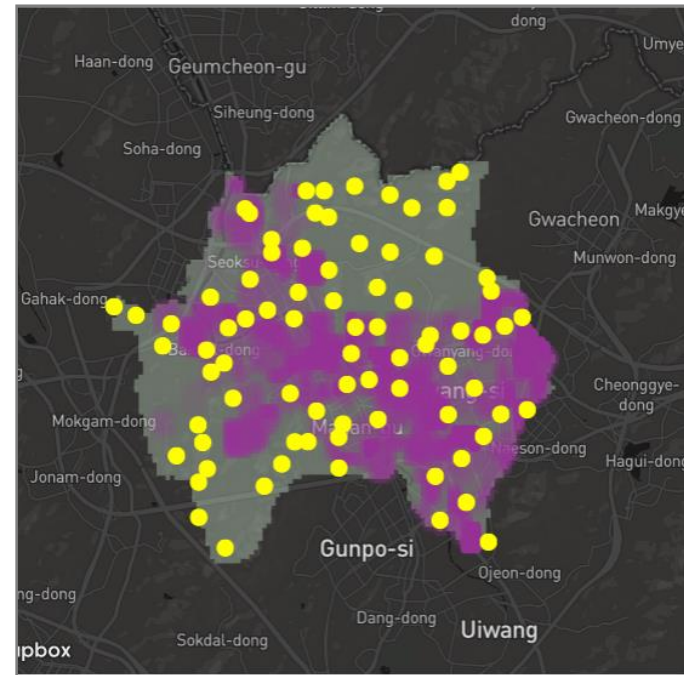
가중치를 고려하지 않을 때 최적화 결론

미세먼지 IoT



- 도로변과 공장주변이 아닌 안양시 전체에 고르게 분포

복합 IoT



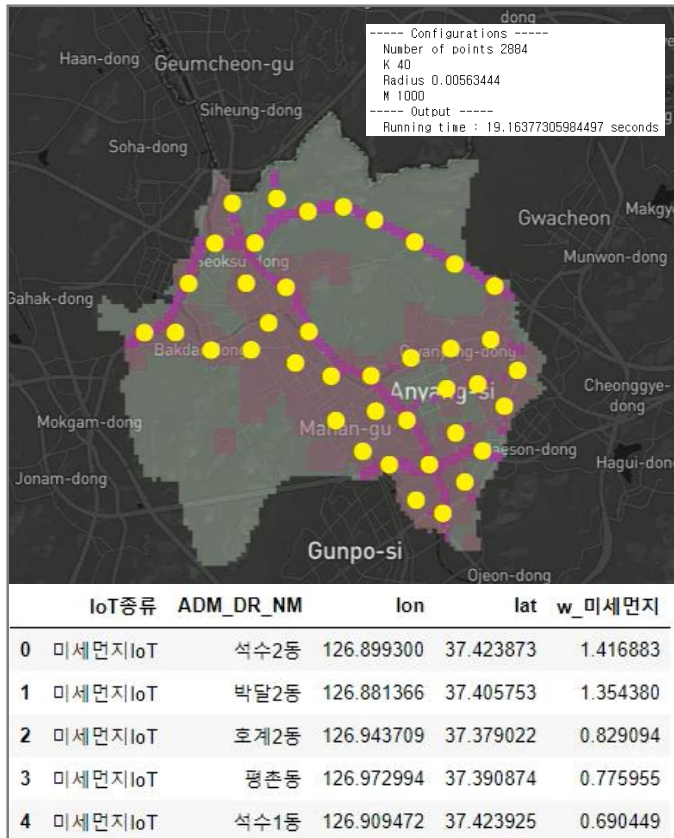
- 인구 현황과 유동인구 분포와는 상관 없이 고르게 분포

산지를 구분하지 않고, 목적에 맞지 않게 임의로 분포 됨

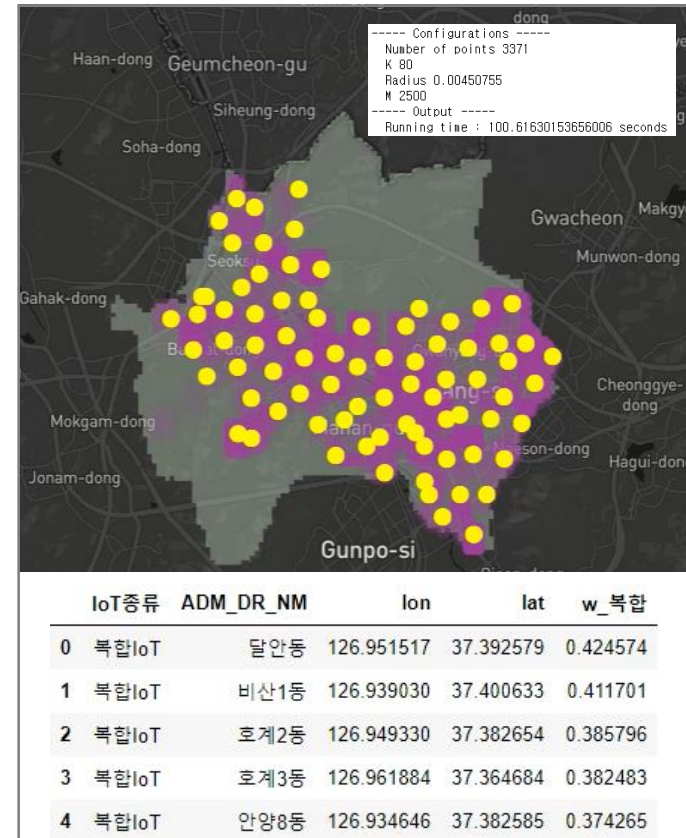
4. IoT 센서 설치 최적화 결론

최종 선정 지역

미세먼지 IoT



복합 IoT

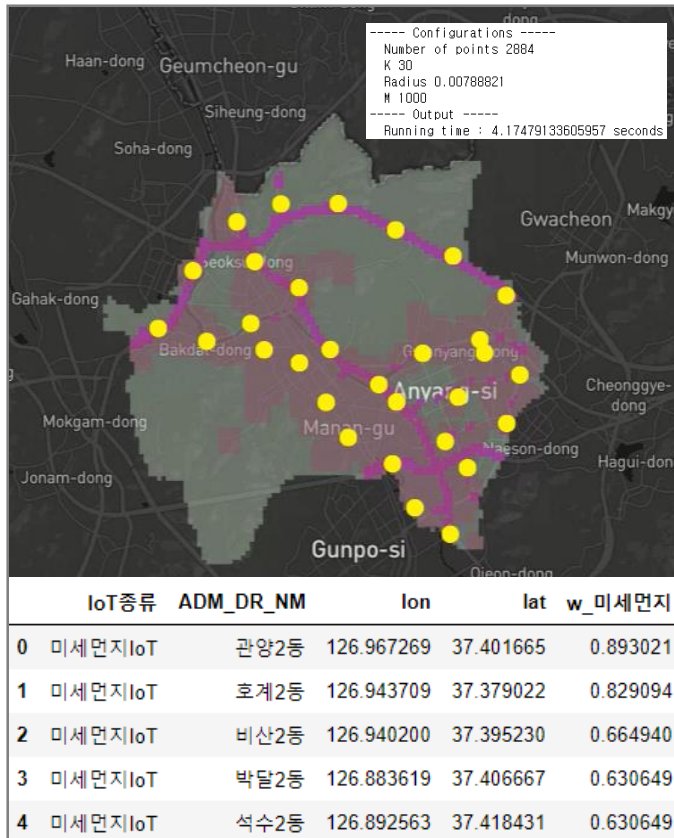


미세먼지 IoT 기기는 도로변과 공장 주변으로, 복합 IoT는 주요 시설과 유동인구 주변에 고르게 분포

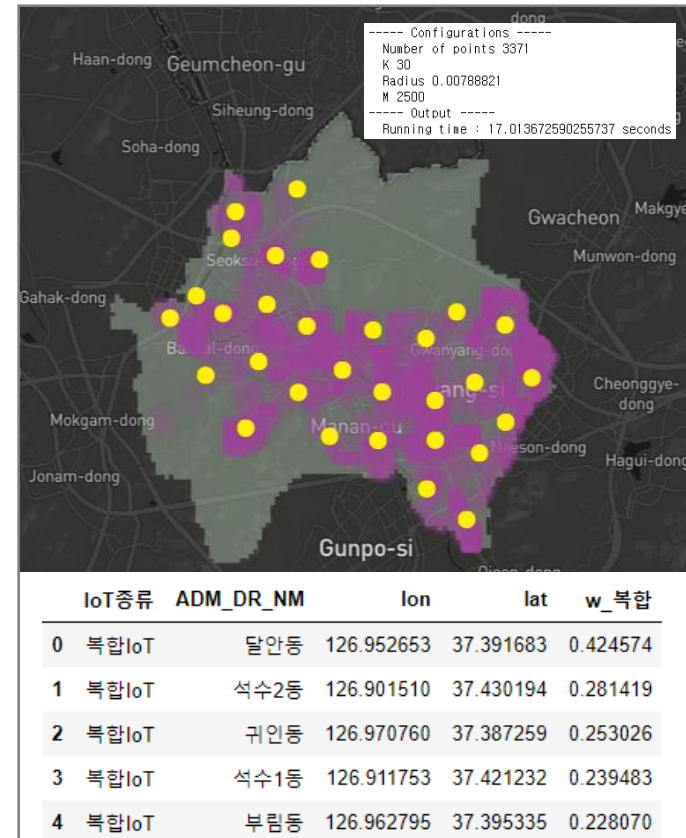
4. IoT 센서 설치 최적화 결론

(추가) Covered distance 700m 가정

미세먼지 IoT



복합 IoT



700m를 cover 할 수 있는 IoT가 있다면 미세먼지와 복합 IoT 기기는 30대 만으로 90% cover 가능

결과 유의성	<ul style="list-style-type: none"> 미세먼지 관리 종합계획을 반영해 미세먼지 IoT와 복합 IoT 최적화 위치가 목적에 맞게 잘 배치됨 <ul style="list-style-type: none"> 미세먼지 IoT는 도로변과 공장주변에, 복합 IoT는 주요 위치 주변과 유동인구가 많은 곳에 적절하게 배치됨
데이터 활용성	<ul style="list-style-type: none"> IoT 목적에 맞게 데이터를 활용함 <ul style="list-style-type: none"> 미세먼지 IoT 설치를 위해 교통량과 공장 현황 데이터를 활용했으며, 이는 미세먼지 관리 종합계획을 반영 복합 IoT 설치를 위해 학교·유치원·병원 위치, 고령·유소년 인구현황 등 데이터를 활용하여 취약계층을 고려 했으며, 전체 인구 현황, 지하철 역 위치, 버스 승하차 등 데이터를 활용하여 유동인구를 고려했음
분석 창의성	<ul style="list-style-type: none"> IoT 목적을 반영한 최적화 모델 정의로 해결 <ul style="list-style-type: none"> 미세먼지 IoT와 복합 IoT 목적을 반영한 적절한 가중치 부여 최적의 결정 변수 도출을 위한 다양한 시나리오 분석 <ul style="list-style-type: none"> 주어진 covered distance가 없어 시나리오 분석을 통해 주어진 IoT 설치에서 가장 효율적인 covered distance 도출 임의의 covered distance를 가정하여 효율적인 IoT 대수 제안
공공 활용성	<ul style="list-style-type: none"> 모든 패키지는 오픈소스를 사용 <ul style="list-style-type: none"> Mixed Integer Programming을 풀기 위한 solver는 COIN-OR CBC로, 오픈소스 solver 중 가장 성능이 뛰어남
경제성	<ul style="list-style-type: none"> 추가 모델 구현 없이 적은 결정변수 (IoT 대수, discovered distance) 수정으로 상황에 맞는 최적화 결과 도출 가능 한 IoT의 covered distance가 주어지면 타깃 범위를 cover하는 가장 효율적인 IoT 대 수를 유추, 예산 절약
활용 방안 및 제안	<ul style="list-style-type: none"> IoT 기기의 성능과 가격이 주어지면 예산에 맞춰 가장 효율적인 IoT 대수와 위치 도출 가능 미세먼지나 복합 IoT 이외 특수 목적 기기 설치에 활용 가능 <ul style="list-style-type: none"> 밤길 안심 귀가를 위한 CCTV 설치, 범죄 빈도 높은 지역 방범용 CCTV 추가 설치 등

첨부1.

분석 소요 시간

- ✓ 분석에는 총 1일 5시간 씩 7일 소요
- ✓ 제출 코드의 running time은 약 10분 가량 소요됨

감사합니다.