

머신러닝 프로젝트 - 최종 발표 -



ML 25

권태운 김윤서 배선용
송영석 이민지



001 주제 선정

- 주제
- 주제 선정 이유

002 데이터 소개

003 EDA

- 행정동 전처리 과정
- 행정동별 생활인구 현황
- 행정동별 시설 현황
- 강남구 시설별 분포 지도

004 머신러닝 모델링

- 랜덤좌표 추출
- 시설별 개수 추출
- 선형회귀, 군집화
- 밀도레벨 구분
- 정확도 계산

005 결과 및 시각화

- 밀도별 특성
- 인구대비 시설
- 특정좌표 분석
- 지하철역 분석

006 활용방안

1.

프로젝트 주제



프로젝트 주제

강남구 위치 기반 시설 밀집도 분석

- 비지도 학습을 통한 강남구 내
특정 위치의 시설 밀도 분석

프로젝트 주제

주제 선정 이유

강남구

서울에서 가장 시설 밀집도가 높은 곳.
가게들의 밀집도가 높은 지역은 상업적으로 활성화된 지역일 가능성이 큼.

강남구라고 모두 밀집도가 높을까?
강남구의 시설 밀집도가 가장 높은 곳은 어디고
밀집도가 낮은 곳은 어디일까?

2.

데이터 소개



데이터 소개

공공 데이터 포털

- 행정동 코드
- 행정동별 생활인구 통계
- 서울시 공중화장실
- 강남구 지하철역

편의점 크롤링



매장명, 행정동, 주소, 위경도 추출

출처: 편의점 빅5 브랜드별 매장검색

인허가 시설 데이터

LOCALDATA

지역 다운로드

전체 데이터를 지역(시도)별로 제공 받을 수 있습니다.



지역선택 ▼

카테고리



건강



동물



문화





생활





식품

데이터 소개

인 허가 시설 데이터 소개

 ARR_FOOD.csv 



음식점: '관광식당', '일반음식점'

 ARR_HOSPITAL.csv 



의료시설: '병원', '의원', '부속의료기관', '약국', '의료법인', '동물병원', '동물약국'

 ARR_SPORTS.csv 

체육시설: '골프연습장업', '골프장', '당구장업', '무도장업', '무도학원업', '수영장업',
'종합체육시설업', '체육도장업', '체력 단련장업'

 ARR_TOURISM.csv 

숙박시설: '관광숙박업', '숙박업', '외국인 관광도시 민박업'

 ARR_CULTURE.CSV 

문화시설: '공연장', '관광 공연장업', '영화상영관'



모든 인 허가 시설 데이터 확인 후 필요한 데이터만 추출, 크롤링

3.

EDA



WHAT ?

무엇이 밀집도에
영향을 줄까?



Part 1 행정동 전처리 과정

Part 2 행정동 별 생활인구 &
시설 현황

Part 3 강남구 시설별 분포 지도



개포동

개포1동
개포2동
개포4동
개포동

논현동

논현1동
논현2동
논현동

대치동

대치1동
대치2동
대치4동
대치동

도곡동

도곡1동
도곡2동
도곡동

삼성동

삼성1동
삼성2동
삼성동

세곡동

세곡동
자곡동
율현동

수서동

신사동

압구정동

역삼동

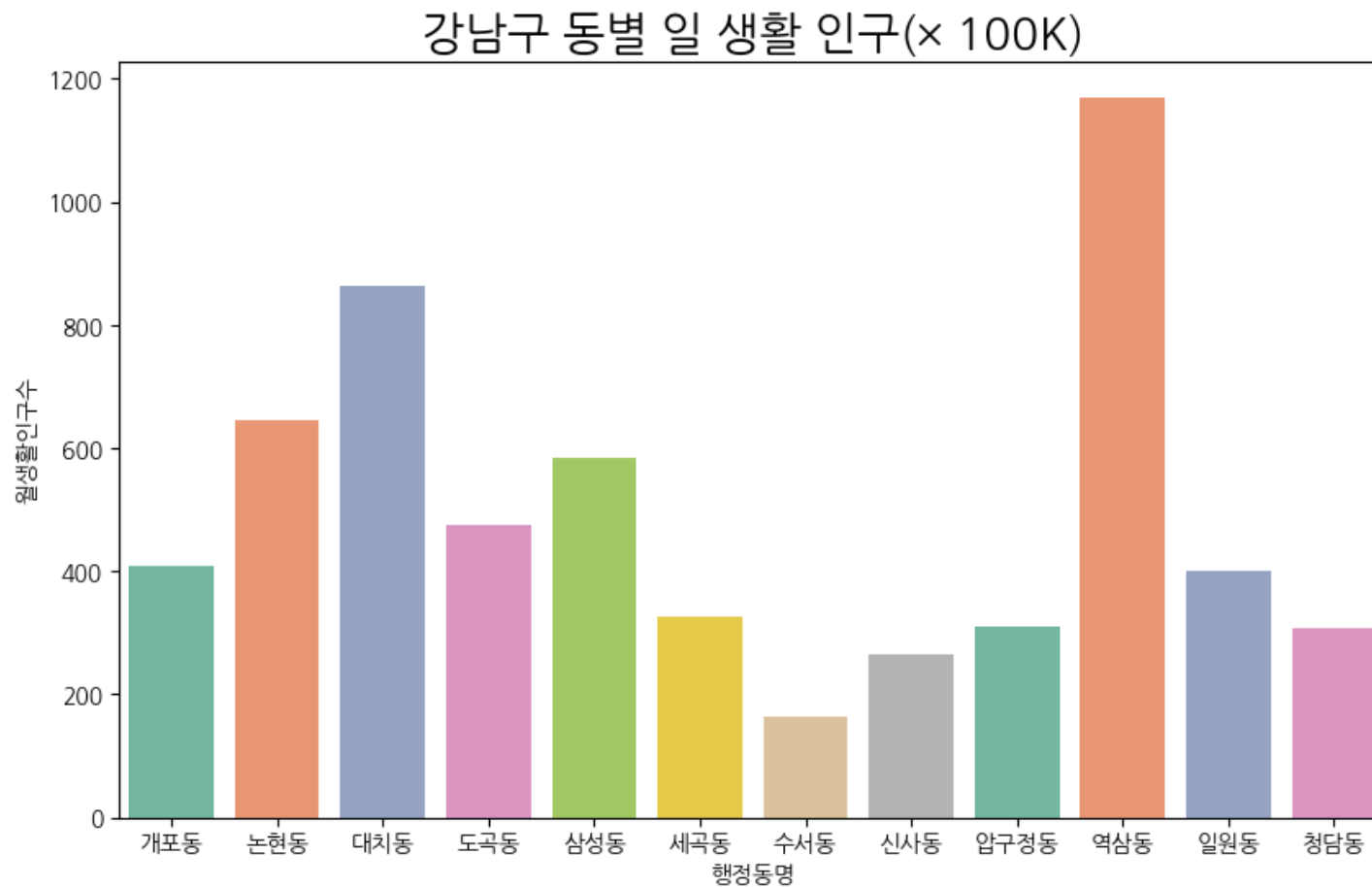
역삼1동
역삼2동
역삼동

일원동

일원1동
일원2동
일원본동
일원동

청담동

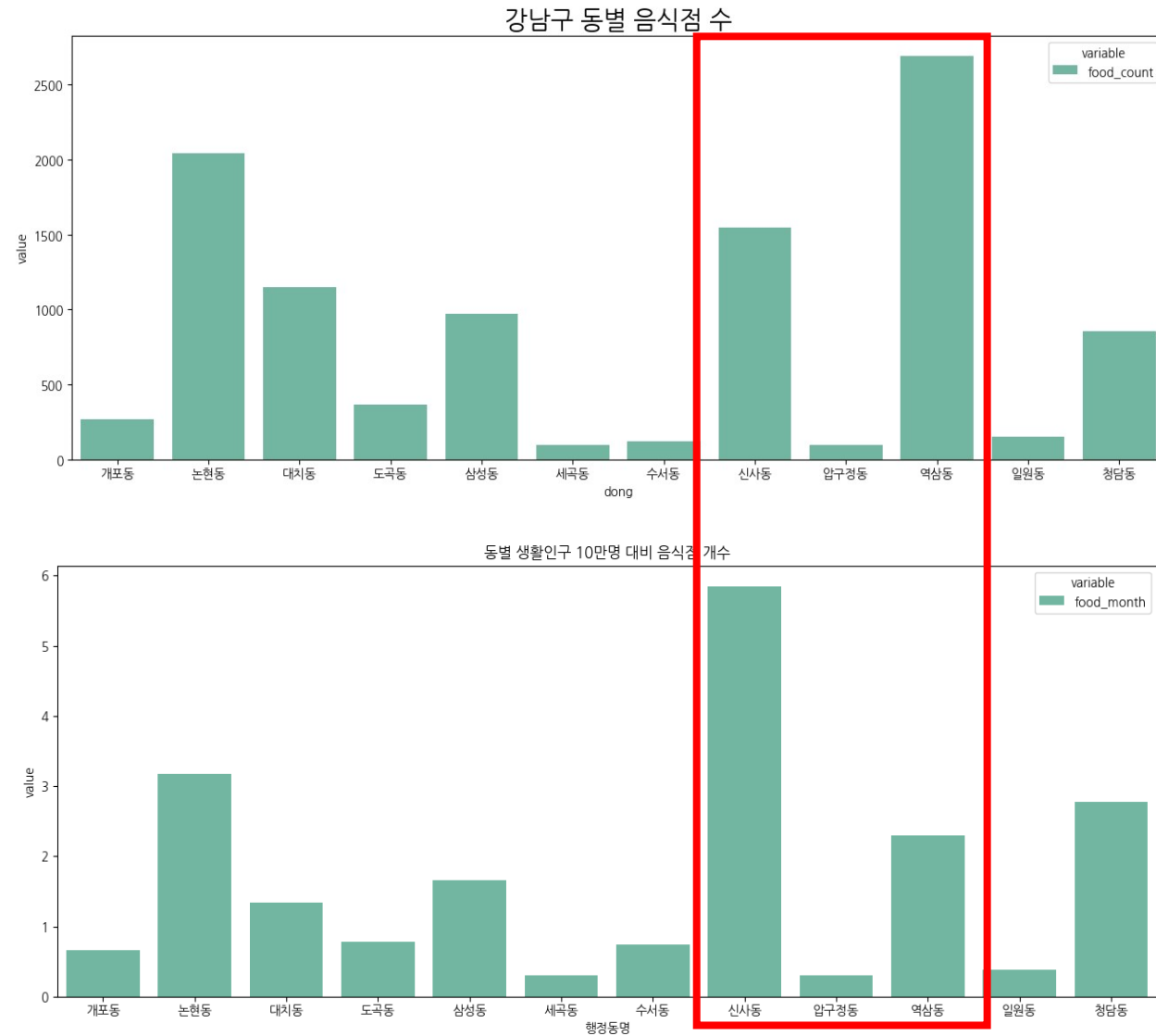
생활인구: 주거 인구 + 체류 인구



2023년 6월 기준, 월 생활 인구 바탕으로 일 평균값 추출

EDA Part 2

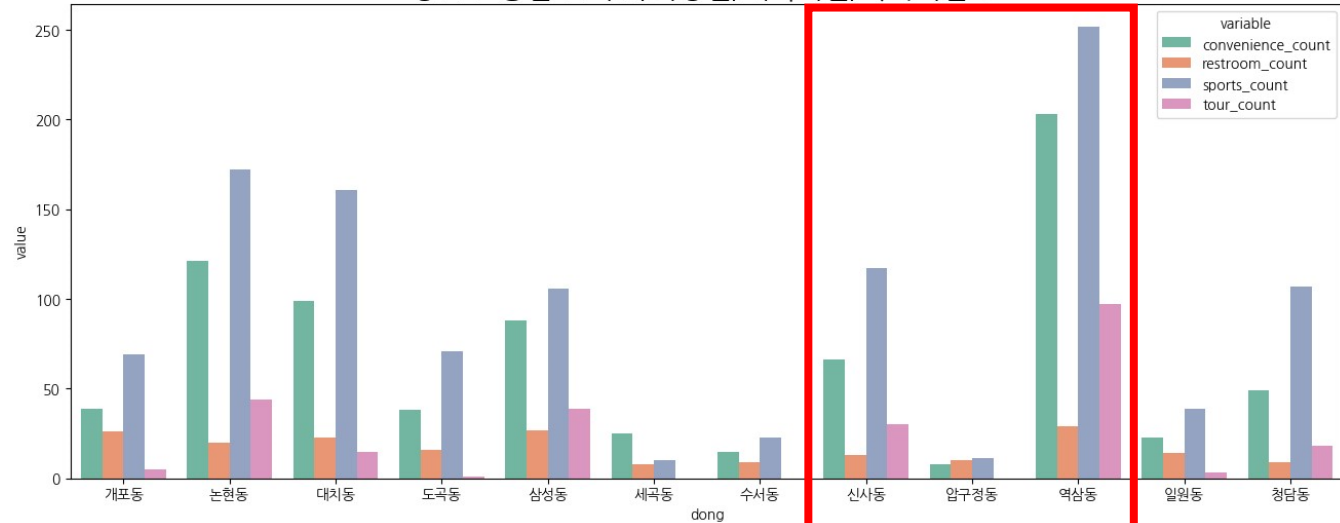
행정동별 시설 현황 - 음식점



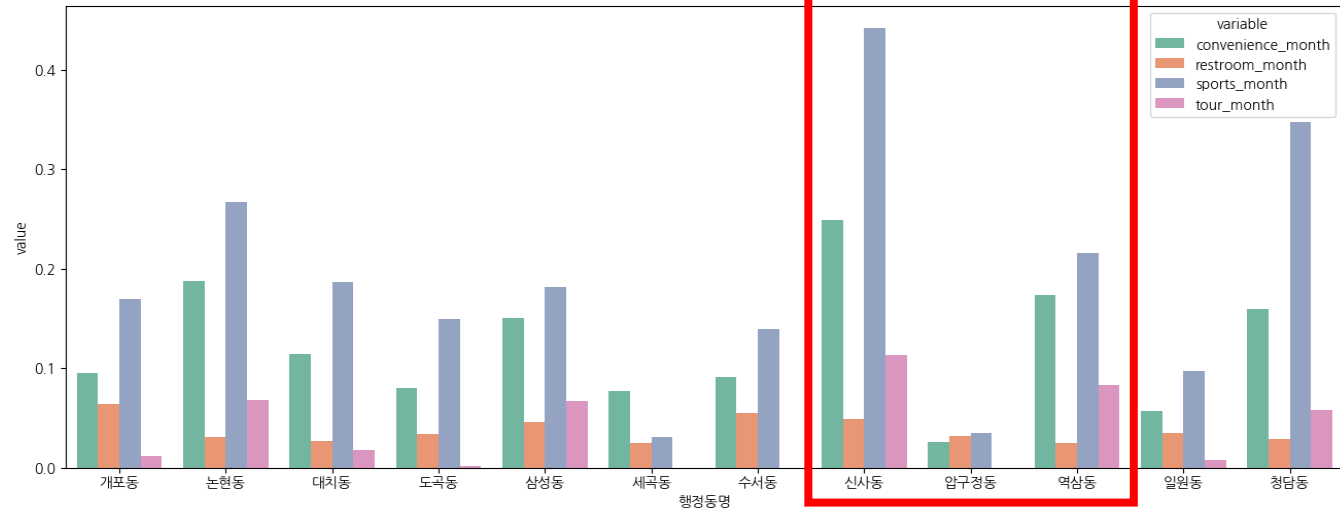
EDA Part 2

행정동별 시설 현황 - 편의점, 화장실, 체육시설, 숙박시설

강남구 동별 편의점/화장실/체육시설/숙박시설 수

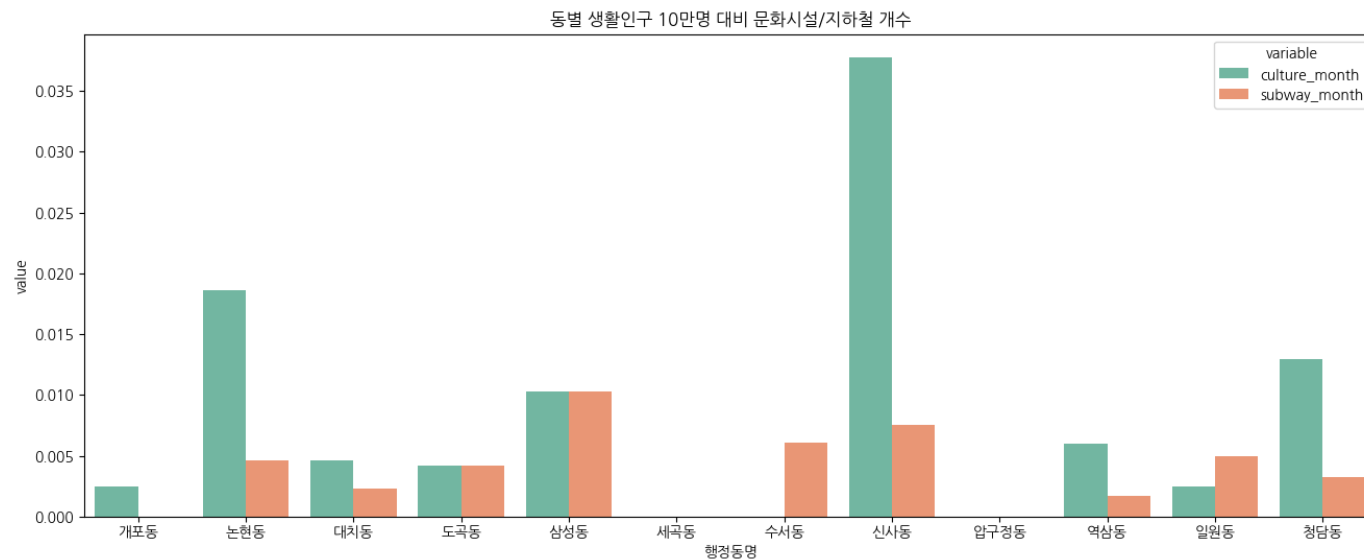
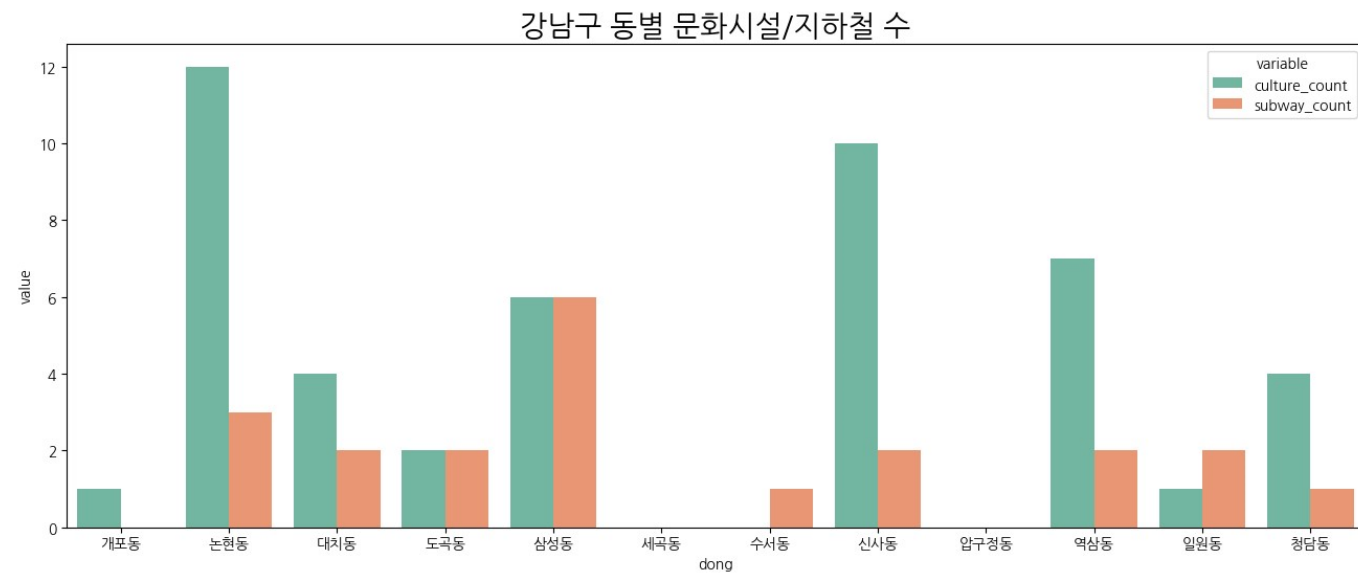


동별 생활인구 10만명 대비 편의점/화장실/체육시설/숙박시설 개수



EDA Part 2

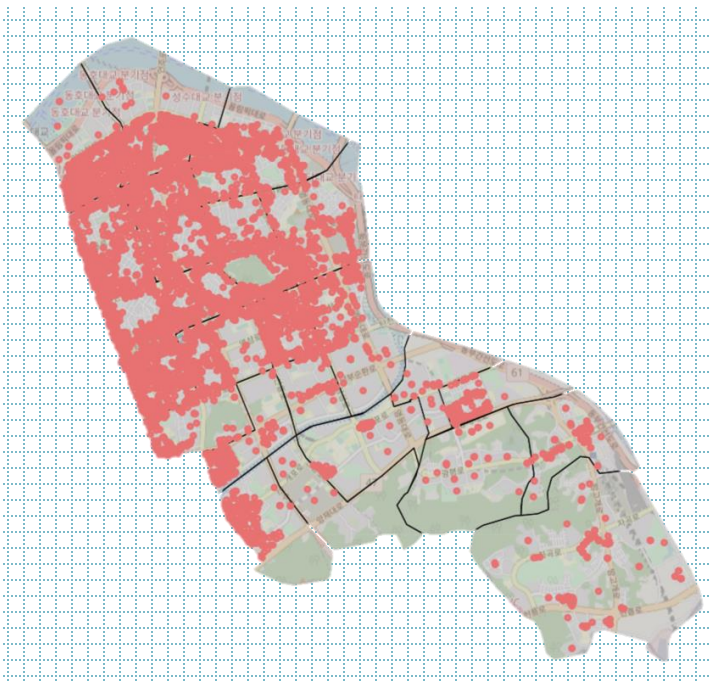
행정동별 시설 현황 - 문화시설, 지하철



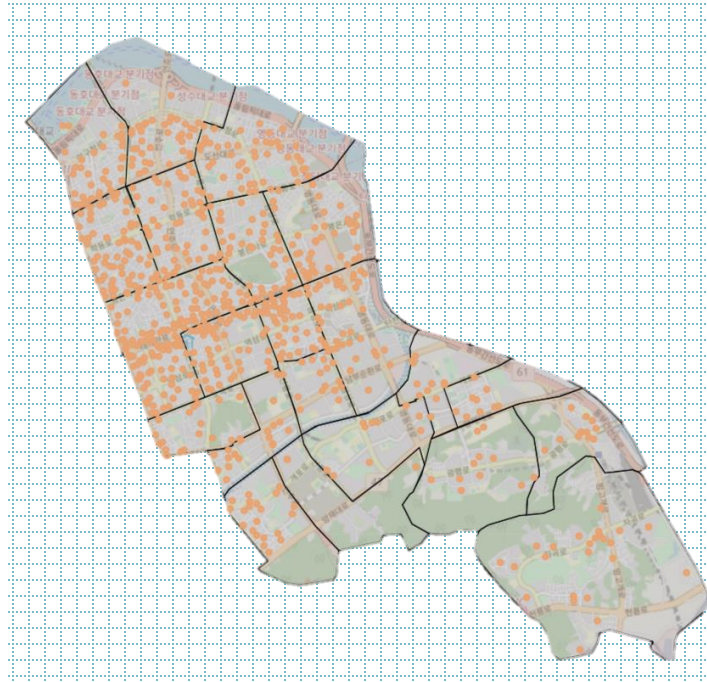
EDA Part 3

강남구 시설별 분포 지도 - 음식점, 편의점, 의료시설

음식점



편의점



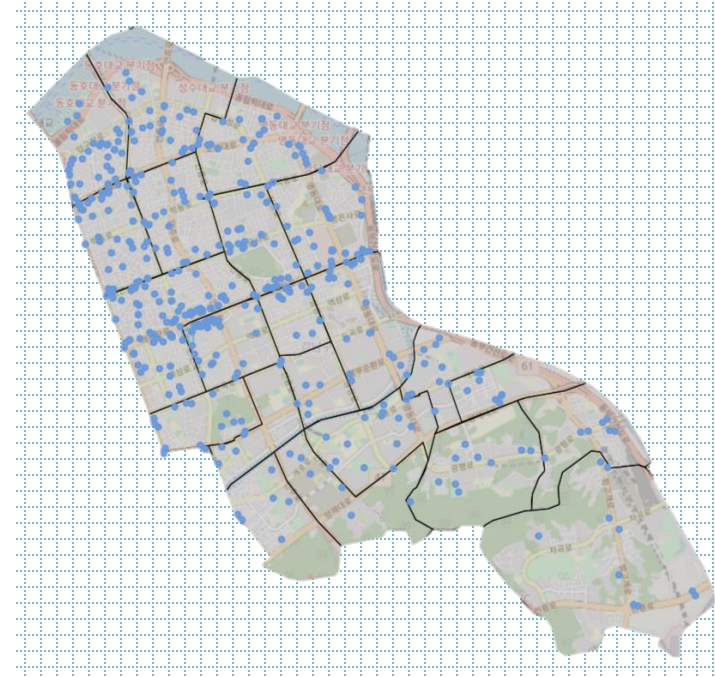
의료시설



체 육 시 설



숙 박 시 설 , 문 화 시 설 , 화 장 실



4.

머신러닝 모델링



머신러닝 모델링

모델링 과정

(1) 폴리곤 바운더리 활용하여 강남구 내 랜덤 좌표 추출



(2) 랜덤 좌표 기준 반경 500m 내 시설별 개수 추출, 전처리



(3) 선형회귀(Linear Regression), 군집화(KMeans)



(4) 클러스터별 밀도 레벨 구분 - 평균값, 중위값 활용



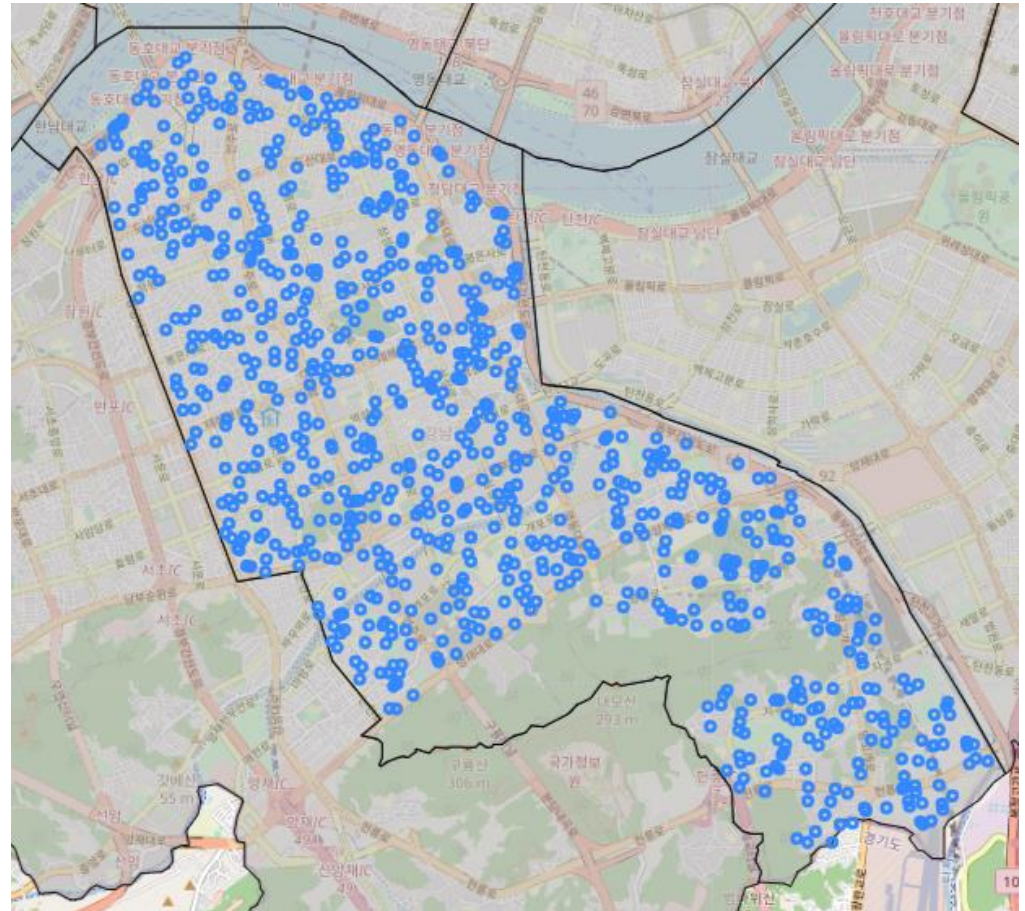
(5) 비지도 학습한 것을 학습이 잘 됐는지 지도학습으로 정확도 계산

머신러닝 모델링

(1) 폴리곤 바운더리 활용하여 강남구 내 랜덤 좌표 추출

```
polylist = [  
    (127.0158715456462, 37.52395552514407), \  
    (127.03404427570877, 37.48455972739361), \  
    (127.04138283634332, 37.48556651426365), \  
    (127.04501866584928, 37.47730743834274), \  
    (127.05109704553787, 37.47143038122402), \  
    (127.08181938718701, 37.488766534231814), \  
    (127.0829699570822, 37.486260930171355), \  
    ...  
    (127.09986054922331, 37.49332977890127), \  
    (127.06956401375487, 37.50279404849988), \  
    (127.06595858545498, 37.51822131133216), \  
    (127.05383896278687, 37.52730956318118), \  
    (127.02705198082393, 37.53395010704315)]  
polygon = Polygon(polylist)  
polygon
```

✓ 0.0s



머신러닝 모델링

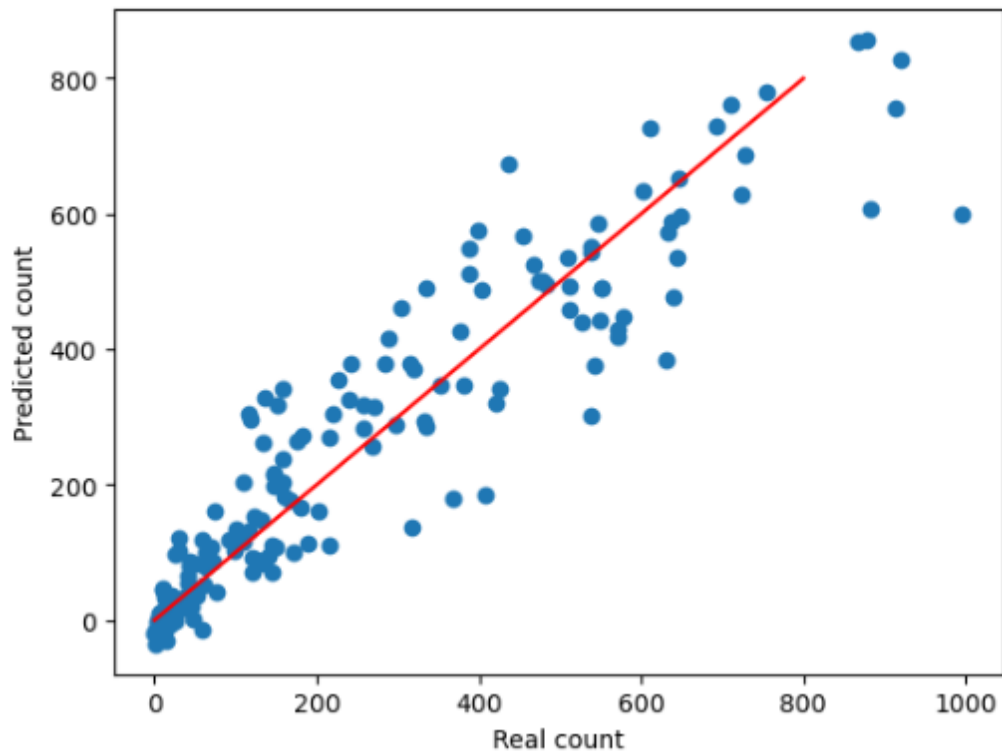
(2) 랜덤 좌표 기준 반경 500m 내 시설별 개수 추출, 전처리

	lat	lng	food_500	convenience_500	hospital_500	sports_500	tour_500	restroom_500	culture_500
0	37.474002	127.099472	42	9	13	4	1	0	0
1	37.488512	127.061428	13	3	20	7	0	5	0
2	37.490649	127.081547	139	11	43	30	1	4	0
3	37.514602	127.048614	190	26	85	45	14	8	1
4	37.509825	127.062266	511	41	120	31	13	9	4
...
869	37.464886	127.095948	33	3	20	4	0	0	0
870	37.482109	127.104837	8	1	1	1	0	2	0
871	37.485941	127.105473	111	8	51	17	0	8	0
872	37.472615	127.051909	52	8	9	8	0	0	1
873	37.481765	127.087633	15	7	9	8	0	3	0

머신러닝 모델링

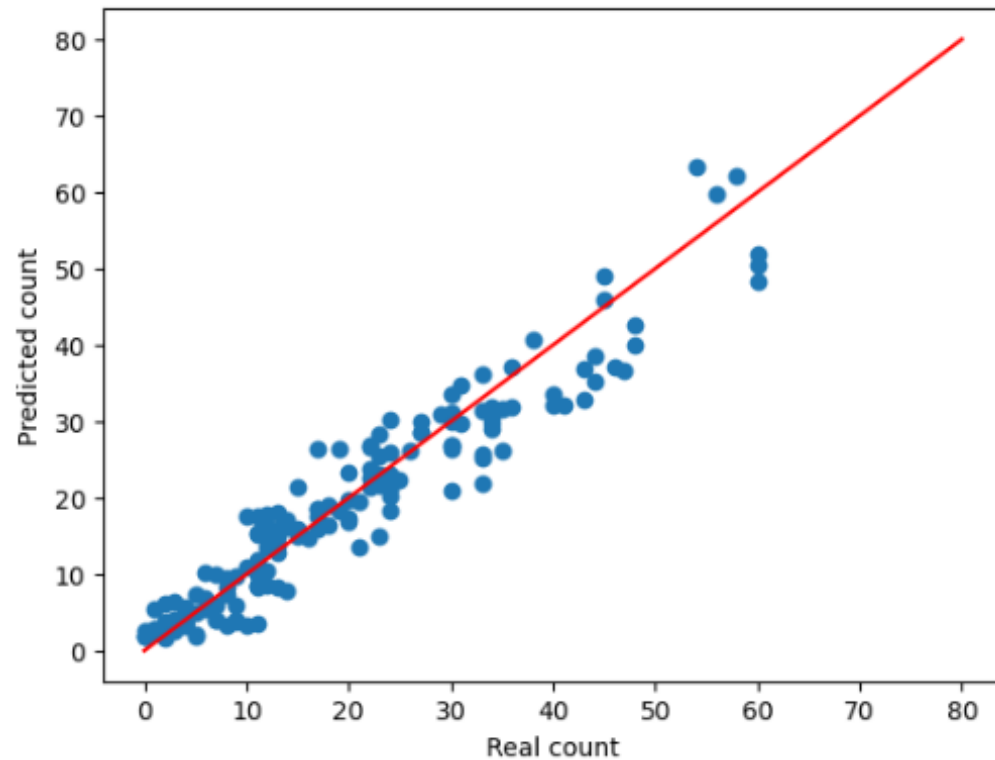
(3) 선형회귀(Linear Regression)

음 식 점



R^2 (test) :
0.86

편 의 점



R^2 (test) :
0.91

머신러닝 모델링

(3) 군집화(KMeans)

랜덤 좌표 기준 반경 500m 내 시설별 개수 데이터로
군집화 진행

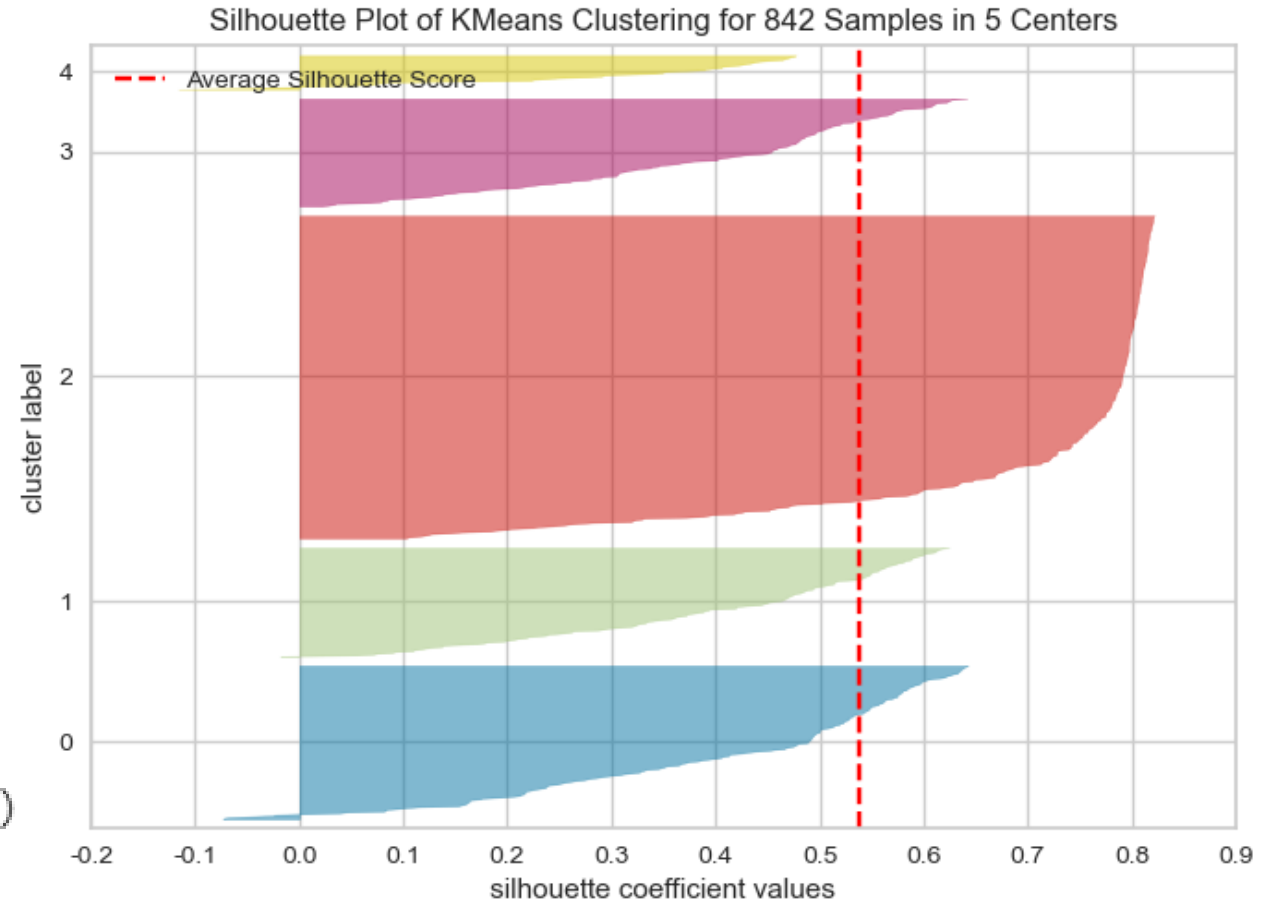
Cluster 2개 ~ 10개 시도

Cluster 2~ 3개가 실루엣계수가 잘 나옴

Cluster 4 ~ 7개는 실루엣계수가 비슷한 수준

➤ 밀도 레벨을 나누었을 때 최적의 기준이 5개라 판단

```
avg_value: 0.5372127551650233  
silhouette_samples() return 값의 shape: (842,)
```



머신러닝 모델링

(4) 클러스터별 밀도 레벨 구분

평 균 값		food_500	convenience_500	hospital_500	sports_500	tour_500	restroom_500	culture_500	cluster
	0	194.342697	18.808989	75.831461	33.106742	4.561798	4.752809	1.089888	0.0
	1	612.166667	39.031746	148.500000	49.468254	19.626984	6.111111	2.190476	1.0
	2	39.544236	5.774799	21.731903	8.756032	0.613941	3.434316	0.134048	2.0
	3	413.056000	29.912000	109.024000	45.872000	9.512000	5.472000	1.848000	3.0
	4	789.975000	36.375000	320.250000	54.750000	14.550000	7.450000	4.350000	4.0
중 위 값		food_500	convenience_500	hospital_500	sports_500	tour_500	restroom_500	culture_500	cluster
	0	186.5	18.0	61.0	32.0	2.0	5.0	1.0	0.0
	1	600.5	41.0	133.5	47.5	16.0	6.0	2.0	1.0
	2	35.0	5.0	15.0	7.0	0.0	3.0	0.0	2.0
	3	417.0	29.0	97.0	47.0	8.0	6.0	1.0	3.0
	4	796.0	31.5	288.0	54.0	16.5	7.5	4.5	4.0

평균값과 중위값을 기준으로 밀도가 높은 클러스터 선별 : $4 > 1 > 3 > 0 > 2$

머신러닝 모델링

(5) 비지도 학습한 것을 학습이 잘 됐는지 지도학습으로 정확도 계산

1. sklearn으로 train, test 분류

```
X_train.shape, X_test.shape
```

✓ 0.0s

((673, 7), (169, 7))

2. Decision Tree 학습

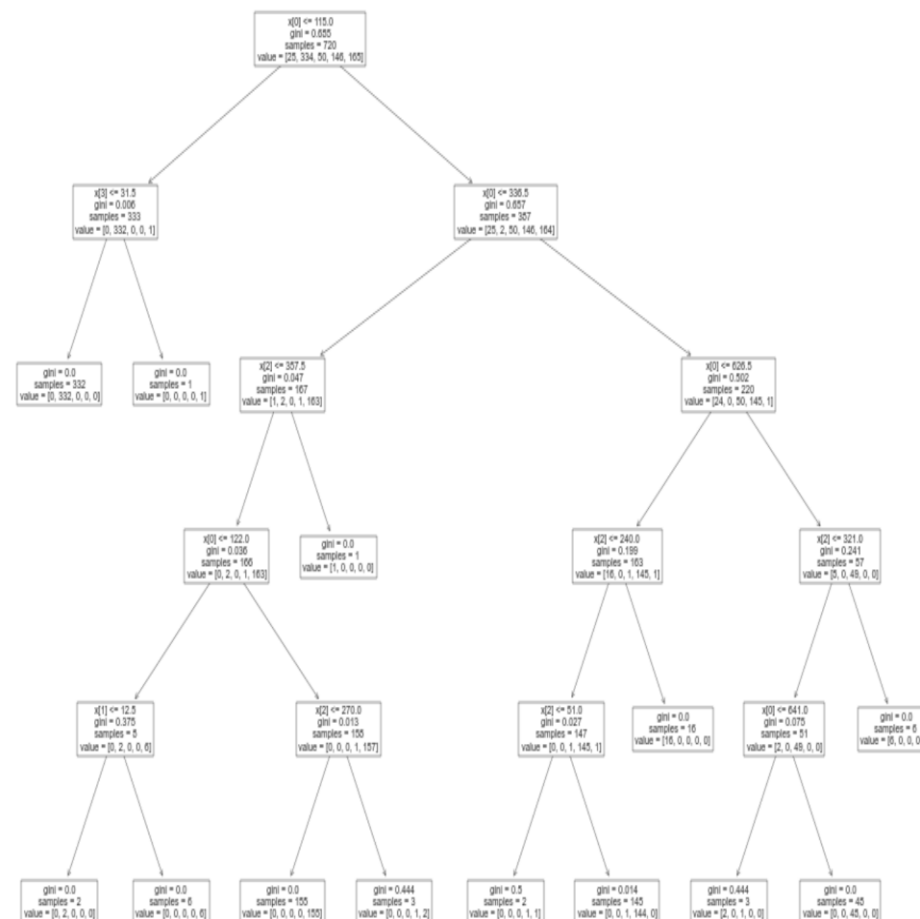
```
DecisionTreeClassifier  
DecisionTreeClassifier(max_depth=5, random_state=13)
```

3. 정확도 확인

0.9763313609467456



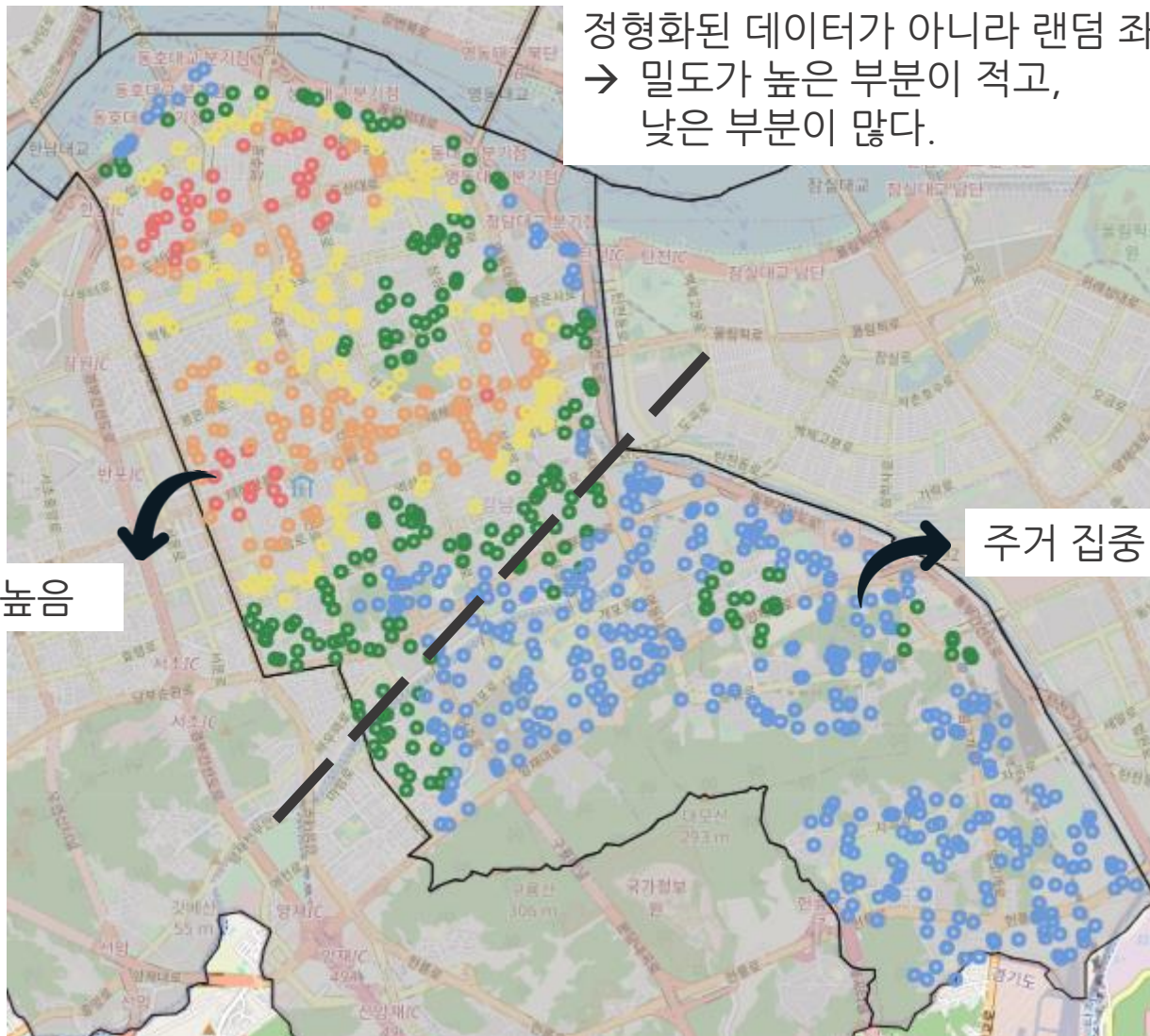
정확도 97%로 높은 수준으로
비지도 학습이 잘된 것으로 보임.



머신러닝 모델링

결과 및 시각화

- 밀도 레벨 4
- 밀도 레벨 3
- 밀도 레벨 2
- 밀도 레벨 1
- 밀도 레벨 0



5.

결과 및 시각화



결과 및 시각화

Part 1 밀도 레벨별 특성

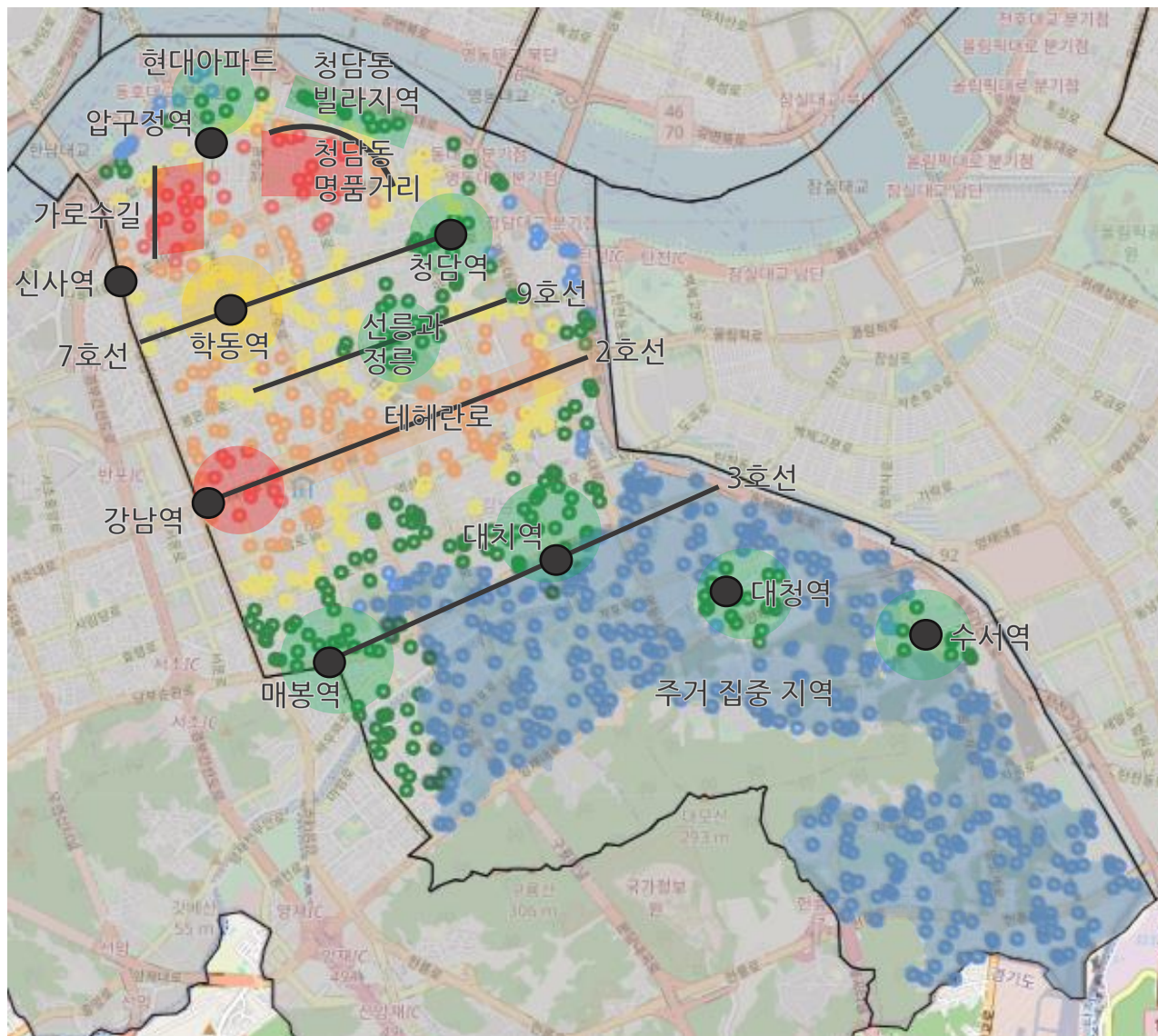
Part 2 강남구 인구 대비 시설 수의 군집화

Part 3 특정 좌표의 밀도 레벨과 주변 시설

Part 4 지하철역의 밀도 레벨과 주변 시설, 승객 수

결과 및 시각화 Part 1

밀도 레벨별 특성

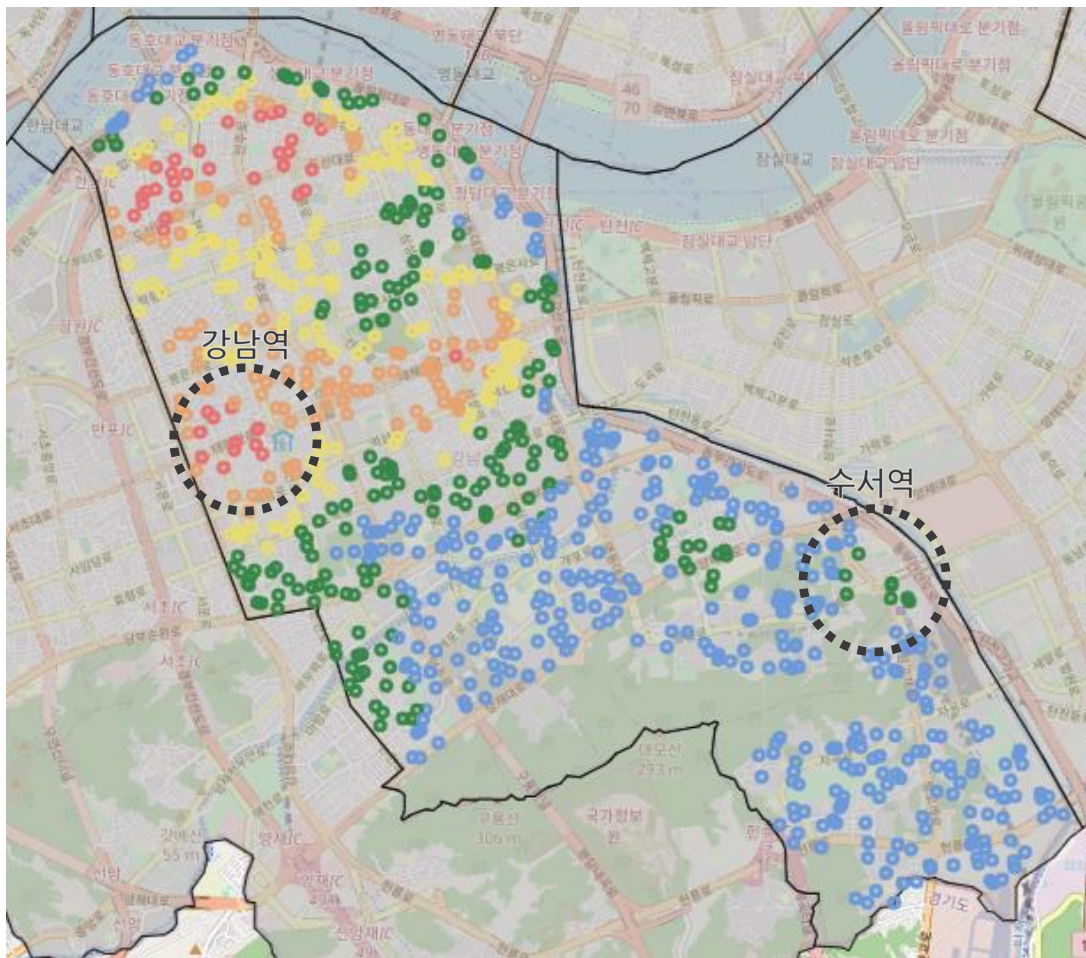


- 밀도 레벨 4 - 상업 집중 지역
- 밀도 레벨 3 - 상권과 가까운 지하철역, 대로변
- 밀도 레벨 2 - 학동역 주변
상권과 조금 떨어진 지역
- 밀도 레벨 1 - 압구정과 청담의 주거 지역
청담역, 선릉과 정릉 주변
주거 집중 지역 근처 지하철역
- 밀도 레벨 0 - 주거 집중 지역

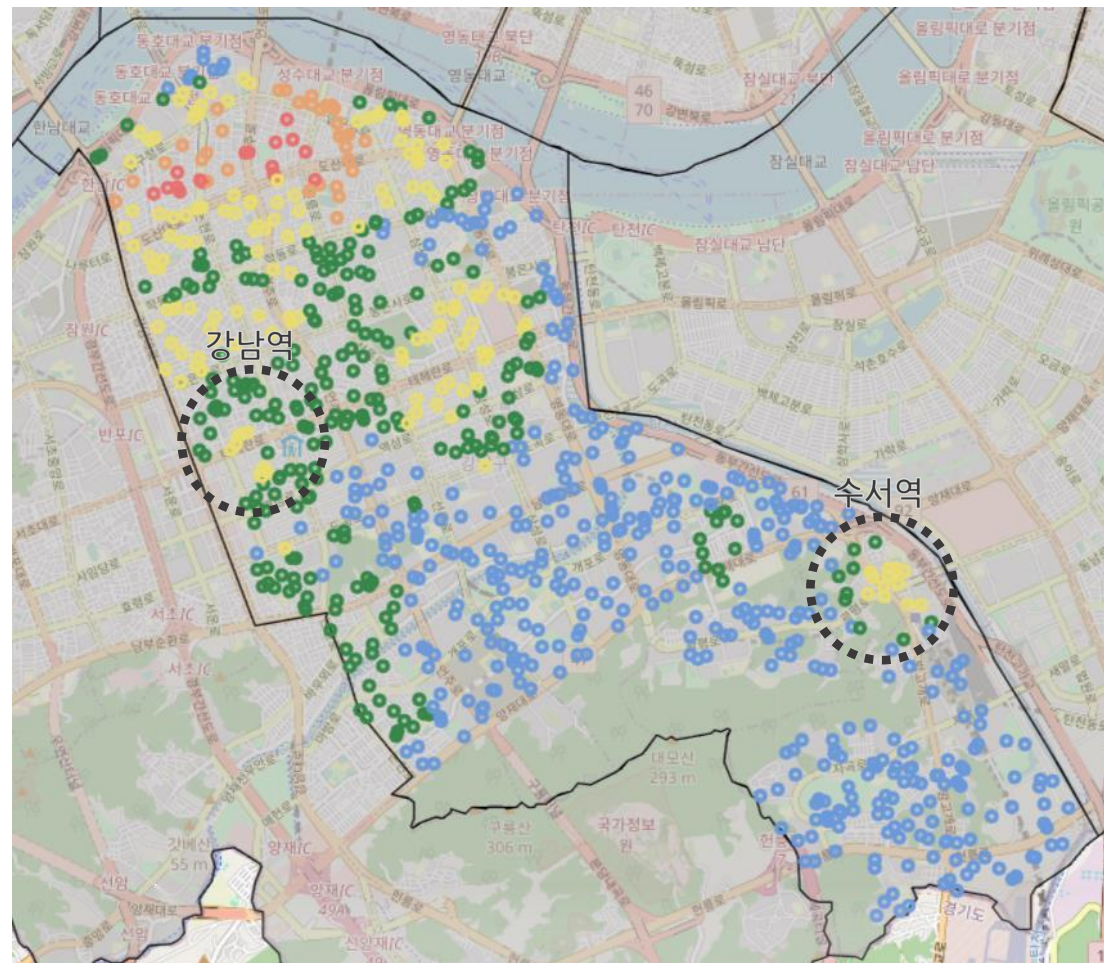
결과 및 시각화 Part 2

동별 생활인구 대비 시설 수 군집화 특성

시설 수 밀도



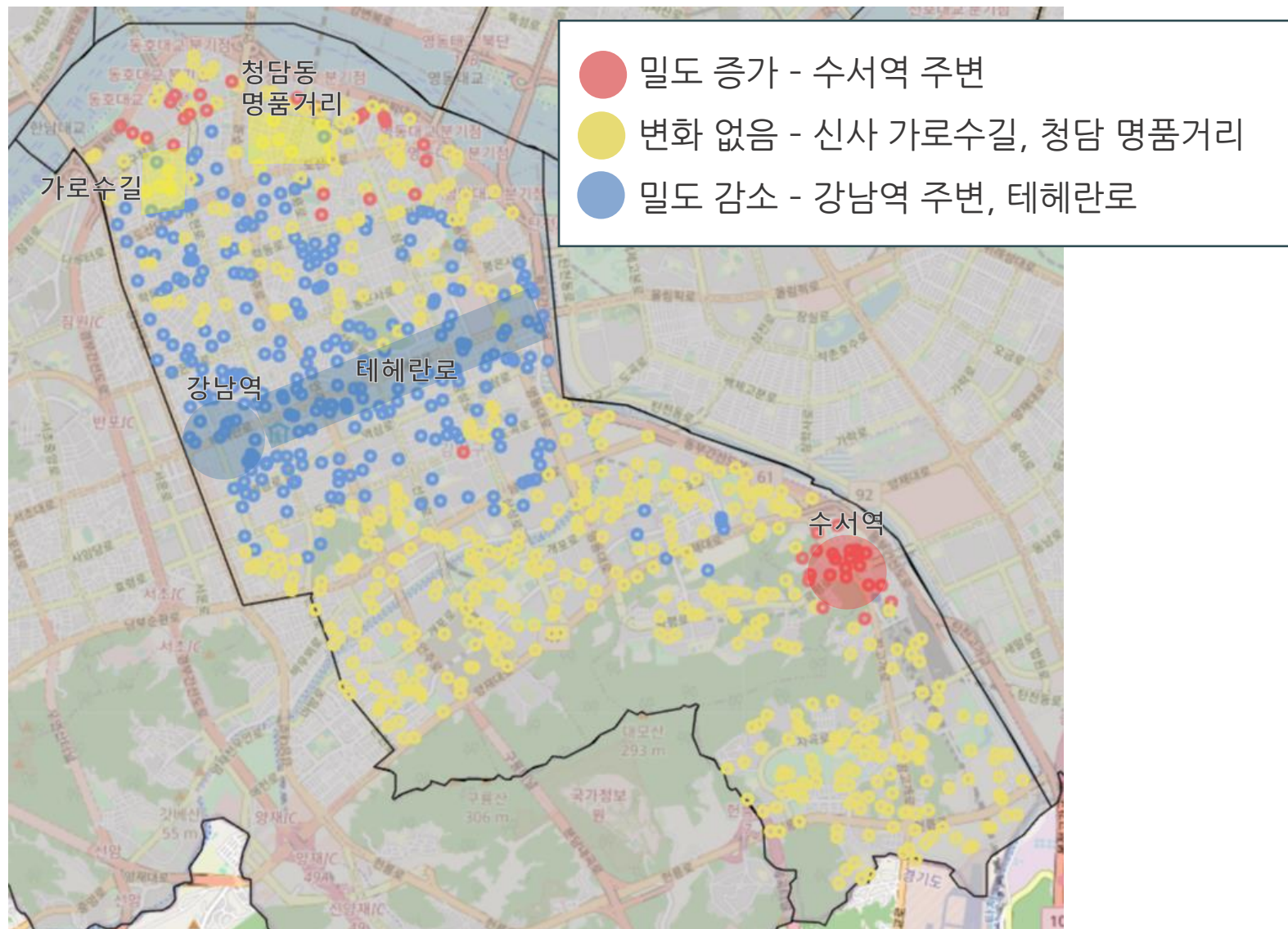
인구 대비 시설 수 밀도
(생활인구 10만명당)



결과 및 시각화

Part 2

시설 수와 인구 대비 시설 수의 밀도 증감 시각화

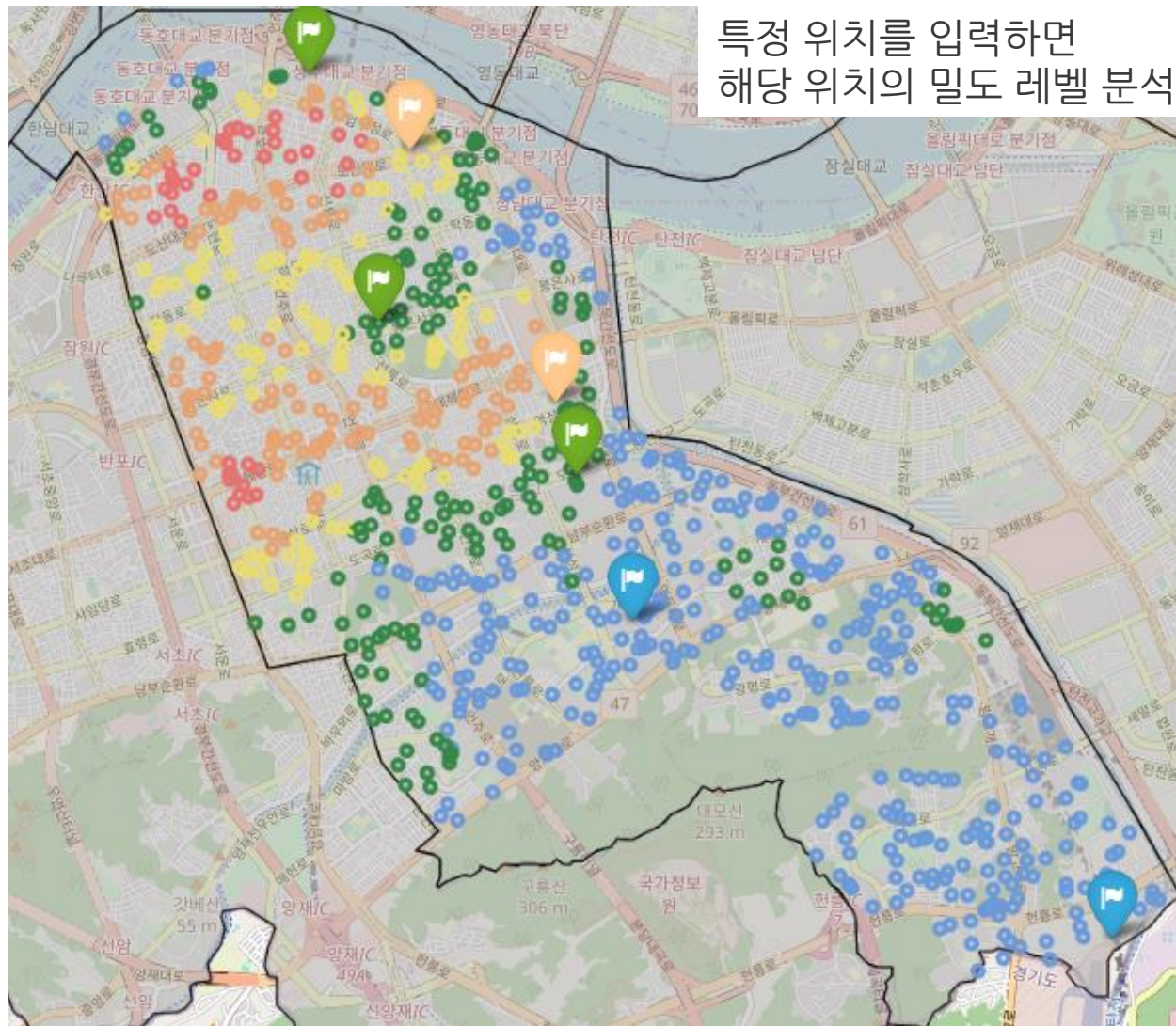


결과 및 시각화

Part 3

특정 좌표의 밀도 레벨 추출






- 밀도 레벨 4
- 밀도 레벨 3
- 밀도 레벨 2
- 밀도 레벨 1
- 밀도 레벨 0



결과 및 시각화

Part 3

특정 좌표의 주변 시설 추출

-  밀도 레벨 4
-  밀도 레벨 3
-  밀도 레벨 2
-  밀도 레벨 1
-  밀도 레벨 0



-  음식점
-  편의점
-  의료 시설
-  체육 시설
-  기타

결과 및 시각화 Part 4 지하철역의 밀도 레벨과 주변 시설 추출



결과 및 시각화 Part 4

지하철역의 밀도 레벨과 승객 수

역명	호선	승차총승객수	하차총승객수	총승객수	level
강남	2	73163	70646	143809	3
선릉	2	60605	60712	121317	3
삼성(무역센터)	2	51274	51472	102745	2
역삼	2	43630	49584	93214	3
수서	3	31191	31852	63043	1
압구정	3	30006	32288	62294	2
신사	3	26004	27263	53266	2
강남구청	7	23292	26229	49521	2
학동	7	20489	21651	42140	2
봉은사	9	20225	21227	41452	1
청담	7	17312	18953	36265	1
선정릉	9	15779	16865	32645	1
논현	7	15966	16378	32344	2
도곡	3	11665	11415	23080	0
대치	3	11165	10930	22094	1
매봉	3	11091	10712	21803	1
언주	9	9922	10049	19971	2
일원	3	9985	9601	19586	0
대청	3	9662	9088	18750	0
삼성중앙	9	6703	6935	13638	2
학여울	3	2218	2267	4484	0

➤ 지하철역 승객 수가 많을수록
밀도 레벨이 높게 나오는 경향

➤ 강남역 = 강남구 + 서초구

강남구의 강남역만 포함하면 밀도 레벨이 3이긴 하나
실제 강남역 주변은 밀도 레벨이 더 높을 것으로 예상

6.

활용방안



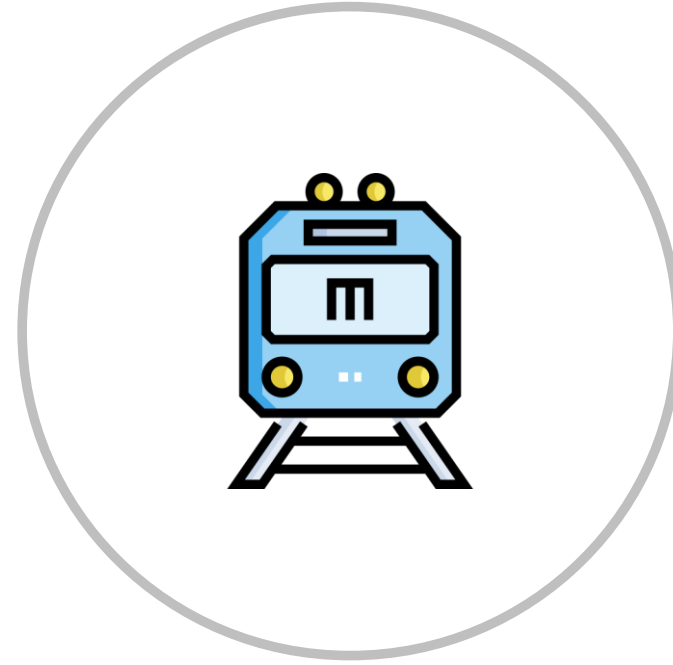
활용방안



특정 좌표 데이터

상권 분석

부동산 분석



역 주변 좌표 데이터

도시계획개발

공공시설 증설

감사합니다