



2021

교육도시 오산, 어린이를 지켜라

새로운 어린이 보호구역의 제시부터 교통안전 시설물 우선 설치 지역 선정까지



팀 데린이보호구역

CONTENTS

01 과제 개요

02 데이터 전처리

인구 및 어린이 교통사고, 어린이 보호구역 등 기본적인 특성
도로망과 도로의 통행량
특정 목적을 위해 건설된 시설 (건물)
도로 시설물(횡단보도, 과속 방지턱 etc)과 기타 정보

03 EDA

04 MODELING

학습데이터 설정
독립변수와 종속변수 설정
선형회귀 모델을 통한 위험도 측정
최종 결과 및 시각화

05 결론



01

과제 개요

교육도시 오산 어린이를 지켜라

과제 개요

데이터 분석 목적

어린이 교통사고 예방을 위한 어린이 보호구역 외 어린이 교통사고 위험지역 20개소 제시 & 기존 어린이 보호구역 중 교통안전시설물 우선 설치 지역 20개소 제시

분석 프로세스

01 Exploratory Data Analysis

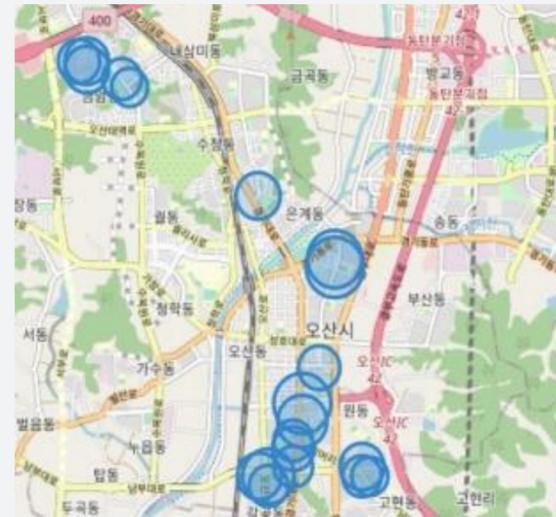
제공 받은 데이터를
격자 내 정보로 표현하여 시사점 제시

02 Data Processing

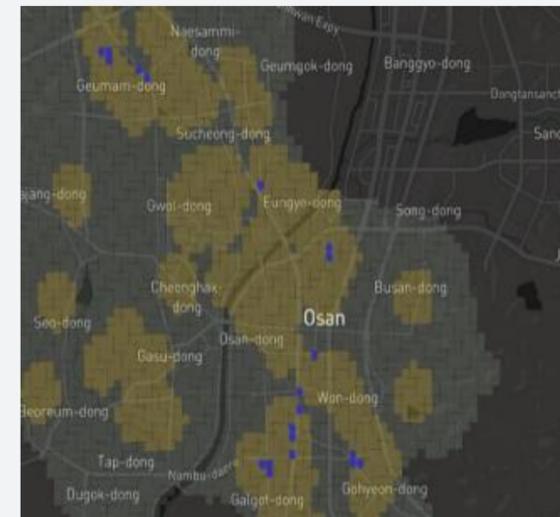
인구 및 데이터를 분석의 목적에 맞게 가공
중심점에 대한 좌표 정보 저장
격자에 데이터가 속하는지 여부를 판단

03 Modeling

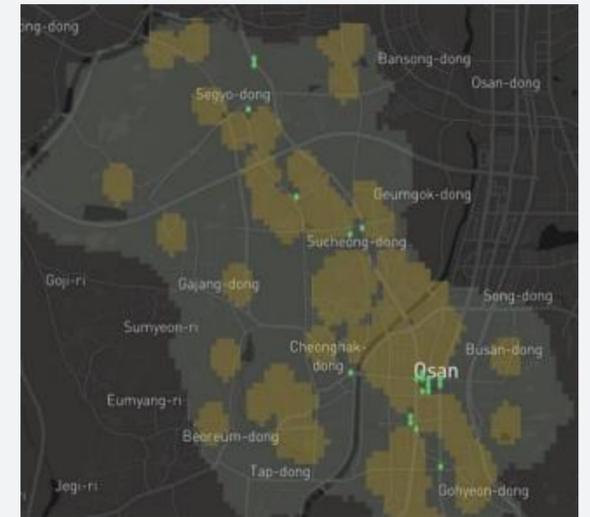
학습데이터 설정
독립변수와 종속변수 설정
다중선형회귀모델을 통한 위험도 측정



다중선형회귀 모델을
통한 위험도 측정



어린이 보호구역 외
교통사고 위험지역 20곳



어린이 보호구역 중
교통사고 위험지역 20곳

02

데이터 전처리

교육도시 오산 어린이를 지켜라

데이터 전처리

전체적인 분석 방법

각 시설물들이 격자의 중심점으로부터 일정 반경 내에 있는지 판단한 후, 그 개수를 세어 정보를 종합한다.

사용한 데이터 파일군



어린이 교통사고,
어린이 보호구역 등
기본적인 특성



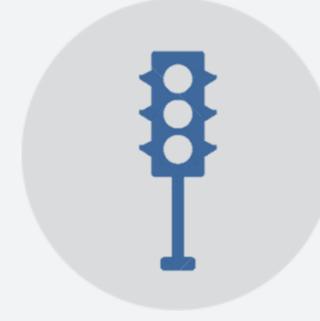
총인구 자료,
지역의
기본적인 특성



도로망, 도로의
교통 특성



학교 등 건물,
특정 목적을 위해
건설된 시설



도로 시설물과
기타 여러 정보

데이터 전처리

4.1.1 격자별 기본 정보(인구, 교통사고 발생 건수 결합하기)

사용한 데이터

- 오산시_어린이교통사고_격자.geojson
- 오산시_연령별_거주인구격자(총인구).geojson
- 오산시_연령별_거주인구격자(유소년).geojson
- 오산시_연령별_거주인구격자(생산가능인구).geojson
- 오산시_연령별_거주인구격자(고령).geojson

전처리 방법

- 모든 자료에 있는 격자의 고유번호에 대한 열인 'gid'를 중심으로 어린이 보호구역을 제외한 모든 자료를 결합한다.
- 기존 'geometry' 열에 있는 좌표 정보를 활용하기 쉽도록 'coordinates' 라는 열을 생성해 nested list로 분리해준다.
- 새롭게 생성된 'coordinates' 정보를 이용해 중심점의 좌표와 해당 격자의 위도, 경도, 최대/최소값을 구한다.
(각각 coord_cent / lon_min, max / lat_min, max)

val_total	val_senior	val_work	val_junior	accident_cnt	grid_id	coordinates	lon_min	lon_max	lat_min	lat_max	coord_cent
0.0	0.0	0.0	0.0	0	00000	[[126.99421564681425, 37.17418235770403], [126...	126.994210	126.995342	37.174182	37.175089	[126.99477586013964, 37.174635450576154]
0.0	0.0	0.0	0.0	0	00001	[[126.99420963816323, 37.17508373885349], [126...	126.994204	126.995336	37.175084	37.175990	[126.9947698580453, 37.17553683173449]
0.0	0.0	0.0	0.0	0	00002	[[126.99420362924478, 37.17598511986466], [126...	126.994198	126.995330	37.175985	37.176891	[126.99476385568379, 37.17643821275453]
0.0	0.0	0.0	0.0	0	00003	[[126.9941976200589, 37.17688650073755], [126....	126.994192	126.995324	37.176887	37.177793	[126.99475785305518, 37.17733959363628]
0.0	0.0	0.0	0.0	0	00004	[[126.99534207541716, 37.174187160436155], [12...	126.995336	126.996469	37.174187	37.175093	[126.9959022955662, 37.174640248032546]
...
0.0	0.0	0.0	0.0	0	04505	[[127.10138369086417, 37.14484471529977], [127...	127.101379	127.102510	37.144845	37.145750	[127.10194433397315, 37.14529730643224]
0.0	0.0	0.0	0.0	0	04506	[[127.10137896035867, 37.145746114265314], [12...	127.101374	127.102505	37.145746	37.146651	[127.10193961004353, 37.146198705390276]
0.0	0.0	0.0	0.0	0	04507	[[127.10251913455733, 37.14304569816435], [127...	127.102514	127.103645	37.143046	37.143951	[127.10307977130084, 37.143498284025306]
0.0	0.0	0.0	0.0	0	04508	[[127.10251441783353, 37.14394709752859], [127...	127.102510	127.103640	37.143947	37.144852	[127.10307506115258, 37.144399683381856]
0.0	0.0	0.0	0.0	0	04509	[[127.10250970089983, 37.144848496754896], [12...	127.102505	127.103636	37.144848	37.145754	[127.1030703507947, 37.145301082600476]

데이터 전처리

4.1.2 격자별 어린이 보호구역 여부 판단하기

사용한 데이터

격자별정보종합

=> (4.1.1 에서 새롭게 생성된 DataFrame)

9. 오산시_어린이보호구역.csv

전처리 방법

1. 축척을 설정한다.
[총인구(100M X 100M)의 격자 정보 사용]
2. 어린이 보호구역 중심으로부터 300M 반경 내부에 격자 중심점이 존재하는 지 판단
3. 격자 각각의 어린이보호구역 포함 여부를 '격자별정보종합'에 반영한다.

```
In [21]: # 거주인구 격자의 중심과 어린이보호구역 중심지점의 array 자료형 생성,
거주인구_경위도 = np.stack([오산시연령별거주인구격자총인구['geometry'].centroid.x,
                             오산시연령별거주인구격자총인구['geometry'].centroid.y], axis = 1).reshape([-1,1,2]) # 격자의 중심점에 대한 행렬
어린이보호구역_경위도 = np.stack([오산시어린이보호구역['보호구역_경도'], 오산시어린이보호구역['보호구역_위도']], axis=1) # 어린이보호구역의

# 어린이보호구역에 격자 중심점이 속하는지 여부 파악,
어린이보호구역_조건 = np.any(np.sum((거주인구_경위도 - 어린이보호구역_경위도)**2, axis = 2)**0.5 < 300 * 축척, axis = 1)

# 조건문 실행 결과 데이터프레임으로 저장,
어린이보호구역데이터 = pd.DataFrame()
어린이보호구역데이터 = 어린이보호구역데이터.append(pd.DataFrame(
    columns=['gid', '어린이보호구역 여부'],
    data=np.stack([오산시연령별거주인구격자총인구['gid'], 어린이보호구역_조건], axis=1),
    ), ignore_index = True)

# Bull 데이터 0/1로 전환
어린이보호구역데이터['어린이보호구역 여부'] = 어린이보호구역데이터['어린이보호구역 여부'].astype(str)
어린이보호구역데이터['children_protection'] = ''
어린이보호구역데이터.loc[어린이보호구역데이터['어린이보호구역 여부'].str.contains('True'), 'children_protection'] = '1'
어린이보호구역데이터.loc[어린이보호구역데이터['어린이보호구역 여부'].str.contains('False'), 'children_protection'] = '0'

# 격자 자료에 결합
격자별정보종합 = pd.merge(격자별정보종합, 어린이보호구역데이터[['gid', 'children_protection']])
격자별정보종합
```

gid	geometry	children_protection
326	다사 569086 MULTIPOLYGON (((127.01449 37.17517, 127.01448 ...	1
327	다사 569087 MULTIPOLYGON (((127.01448 37.17607, 127.01447 ...	1
328	다사 569088 MULTIPOLYGON (((127.01447 37.17697, 127.01447 ...	1
329	다사 569089 MULTIPOLYGON (((127.01447 37.17787, 127.01446 ...	1
357	다사 570085 MULTIPOLYGON (((127.01562 37.17427, 127.01561 ...	1
...
4350	다사 638058 MULTIPOLYGON (((127.09235 37.15022, 127.09234 ...	1
4351	다사 638059 MULTIPOLYGON (((127.09234 37.15112, 127.09234 ...	1
4352	다사 638060 MULTIPOLYGON (((127.09234 37.15203, 127.09233 ...	1
4353	다사 638061 MULTIPOLYGON (((127.09233 37.15293, 127.09233 ...	1
4354	다사 638062 MULTIPOLYGON (((127.09233 37.15383, 127.09232 ...	1

1471 rows x 17 columns

데이터 전처리

4.2 도로시설물(과속방지턱, 도로안전표지, 횡단보도, 신호등)

사용한 데이터

- 16.오산시_도로안전표지표준데이터.csv
- 17.오산시_횡단보도.geojson
- 18.오산시_과속방지턱표준데이터.csv
- 19.오산시_신호등.geojson

	과속방지턱_개수	도로안전표지_개수
count	4510.000000	4510.000000
mean	0.018182	0.546785
std	0.156552	1.139390
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.000000	1.000000
max	3.000000	10.000000

	횡단보도_개수	신호등_개수
count	4510.000000	4510.000000
mean	0.783370	0.962528
std	1.814981	2.445624
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.000000	0.000000
max	22.000000	26.000000

전처리 방법

1. 어린이 도보 속도 기준으로 중심점으로부터 반경을 설정한다.
2. 격자 중심점으로부터 반경 내에 도로시설물이 존재하는지 판단한다.
3. 각 격자에 속하는 도로시설물의 개수를 센다.
4. 위 데이터를 '도로시설물_데이터'로 종합, 격자별정보종합 데이터에 부여한다.

gid	geometry	횡단보도_개수	신호등_개수	과속방지턱_개수	도로안전표지_개수
5	다사552086 (POLYGON ((126.9953360801466 37.17508854174157...	0	0	0	1
6	다사552087 (POLYGON ((126.9953300846092 37.17598992290871...	0	0	0	1
11	다사553086 (POLYGON ((126.9964625223967 37.17509333392205...	0	0	0	1
12	다사553087 (POLYGON ((126.9964565402404 37.1759947152448,...	0	0	0	2
13	다사553088 (POLYGON ((126.9964505578178 37.17689609642928...	0	0	0	1
...
4358	다사638066 (POLYGON ((127.0923078750135 37.15743363761779...	1	2	0	1
4359	다사638067 (POLYGON ((127.0923030345987 37.15833503365665...	1	4	0	2
4385	다사639066 (POLYGON ((127.0934340704674 37.15743750640318...	1	0	0	3
4386	다사639067 (POLYGON ((127.0934292434227 37.1583389025677,...	1	0	0	3
4387	다사639068 (POLYGON ((127.0934244161632 37.15924029859425...	1	0	0	3

데이터 전처리

4.3.1 어린이집, 유치원, 초등학교

사용한 데이터

- 10. 오산시_학교위치정보.csv
 - 13. 오산시_어린이집_유치원현황.csv
- 격자별정보종합(격자별 데이터를 기록한 Dataframe)

	gid	geometry
0	다사551085	(POLYGON ((126.9942156468143 37.17418235770403...
1	다사551086	(POLYGON ((126.9942096381632 37.17508373885349...
2	다사551087	(POLYGON ((126.9942036292448 37.17598511986466...
3	다사551088	(POLYGON ((126.9941976200589 37.17688650073755...
4	다사552085	(POLYGON ((126.9953420754172 37.17418716043615...

전처리 방법

1. 어린이집과 유치원의 특성을 고려하여 각 시설의 데이터를 구분한다.
2. '오산시_학교위치정보'에서 어린이(만 13세 미만)에 관련된 초등학교 데이터만을 추출한다.
3. 각각 중심점에 대한 좌표정보를 저장하여 격자에 해당하는지 여부를 판단한다.
4. 한 격자에 속하는 각 시설물의 개수를 계산한다.
5. 위 데이터를 '교육시설_데이터'로 종합, 격자별정보종합 데이터에 부여한다.

	gid	geometry	어린이집_개수	유치원_개수	초등학교_개수
997	다사584109	(POLYGON ((127.0312542497649 37.19596842295436...	0	1	1
998	다사584110	(POLYGON ((127.0312486767251 37.19686980575049...	0	1	1
1067	다사585109	(POLYGON ((127.0323810081625 37.19597287581157...	2	1	1
1068	다사585110	(POLYGON ((127.0323754485179 37.19687425875226...	1	1	1
1502	다사591094	(POLYGON ((127.0392237235712 37.18247859432989...	0	1	1

데이터 전처리

4.3.2 체육시설

사용한 데이터

29.오산시_체육시설현황.csv
격자별정보종합(격자별 데이터들을 기록한 Dataframe)

	시설구분명	설치위치_경도	설치위치_위도
0	체육도장업	127.087530	37.129453
1	체육도장업	127.086589	37.143006
2	체육도장업	127.068675	37.135502
3	체육도장업	127.044759	37.176358
4	체육도장업	127.078059	37.156365

전처리 방법

1. 어린이가 자주 이용하는 체육도장업 외의 체육시설 데이터를 제거한다.
2. 중심점에 대한 좌표 정보를 저장하여 격자에 해당하는지 여부를 판단한다.
3. 한 격자에 속하는 시설물의 개수를 계산한다.
4. 위 데이터를 격자별정보종합 데이터에 부여한다.

	gid	geometry	체육도장_개수
672	다사579078	(POLYGON (((127.0257951775385 37.16800308527249...	1
719	다사580077	(POLYGON (((127.0269271383509 37.16710620053858...	1
720	다사580078	(POLYGON (((127.0269215197383 37.16800758717508...	1
721	다사580079	(POLYGON (((127.0269159008756 37.16890897367342...	1
836	다사582080	(POLYGON (((127.0291630204195 37.16981933230637...	1

데이터 전처리

4.3.3 학원 및 교습소

사용한 데이터

30.오산시_학원_및_교습소_현황.csv
격자별정보종합(격자별 데이터들을 기록한 Dataframe)

전처리 방법

1. 교습과정명에 포함되어 있는 결측치를 제거한다.
2. '고등'이라는 단어가 포함된 교습과정명을 제거한다.
3. 중심점에 대한 좌표 정보를 저장하여 격자에 해당하는지 여부를 판단한다.
4. 한 격자에 속하는 시설물의 개수를 계산한다.
5. 위 데이터를 격자별정보종합 데이터에 부여한다.

	업종구분명	교습과정명	시설위치_경도	시설위치_위도
0	교습소	음악	127.048919	37.146279
1	교습소	미술	127.048919	37.146279
2	교습소	보습	127.052419	37.145809
3	학교교과교습학원	종합(국,수,사,과,논술)	127.048919	37.146279
4	학교교과교습학원	피아노(체르니40이상)	127.048919	37.146279

	gid	geometry	학원및교습소_개수
586	다사576109	(POLYGON ((127.0222401915633 37.19593241432686...	13
623	다사577109	(POLYGON ((127.02336694796 37.19593695291063, ...	13
863	다사582107	(POLYGON ((127.0290119326268 37.19415671966592...	24
864	다사582108	(POLYGON ((127.0290063332952 37.19505810244826...	24
865	다사582109	(POLYGON ((127.0290007337144 37.19595948509242...	8

데이터 전처리

4.4.3 격자별 자료와 도로망 자료 결합하기

사용한 데이터

격자별정보종합

=> 격자별 데이터들을 기록한 DataFrame

교통정보종합

=> 4.4.1과 4.4.2 결과 생성된 도로망 관련 DataFrame

전처리 방법

1. "Buffer"(도로에 추가적인 두께를 부여하는 수단)를 부여한 도로가 격자의 중심을 지나는지 판단한다.
2. 도로가 격자의 중심을 지나갈 경우 해당 격자에 그 도로와 관련한 데이터(교통량, 혼잡 빈도/시간 강도)를 가져온다.

```
In [70]: # 1) 8시 추정 교통량

# grid 마다 도로 관련 수치 부여
df_grid = []
정제된_도로정보 = 교통정보종합[교통정보종합['추정교통량_8시'] > 0]

정제된_도로정보 = gpd.GeoDataFrame(정제된_도로정보, geometry = "geometry_buff")

# 처리에 시간 소요
for i in range(len(정제된_도로정보)):
    try:
        grid_ids = point_cent[point_cent.within(정제된_도로정보.loc[i, 'geometry_buff'])]['grid_id']
        if len(grid_ids) != 0:
            df_grid.append([i, str(tuple(grid_ids))])
    except:
        pass

print('Point와 관련된 grid 개수: ', len(df_grid))

정제된_도로정보['grid_ids'] = 0
for i in range(len(df_grid)):
    id_idx = df_grid[i][0]
    grids = df_grid[i][1]
    정제된_도로정보['grid_ids'][id_idx] = grids

grid_list = []
for i in 격자별정보종합['grid_id']:
    try:
        grid_list.append([i, sum(정제된_도로정보[정제된_도로정보['grid_ids'].str.contains(i)==True]['추정교통량_8시'])])
    except:
        pass
```

gid	geometry	val_total	val_senior	val_work	val_junior	accident_cnt	grid_id	coordinates	추정교통량_8시
5	다사 552086 MULTIPOLYGON (((126.99534 37.17509, 126.99533 ...	0.0	0.0	0.0	0.0	0	00005	[[126.99533608014656, 37.175088541741566], [12...	261.44
6	다사 552087 MULTIPOLYGON (((126.99533 37.17599, 126.99532 ...	0.0	0.0	0.0	0.0	0	00006	[[126.99533008460917, 37.17598992290871], [126...	403.05
9	다사 553084 MULTIPOLYGON (((126.99647 37.17329, 126.99647 ...	0.0	0.0	0.0	0.0	0	00009	[[126.99647448591081, 37.173290570861724], [12...	2224.39
10	다사 553085 MULTIPOLYGON (((126.99647 37.17419, 126.99646 ...	0.0	0.0	0.0	0.0	0	00010	[[126.99646850428688, 37.17419195246102], [126...	1330.87
11	다사 553086 MULTIPOLYGON (((126.99646 37.17509, 126.99646 ...	0.0	0.0	0.0	0.0	0	00011	[[126.99646252239674, 37.17509333392205], [126...	907.96

5 rows x 26 columns

데이터 전처리

4.5 기타 관련 정보(주정차단속, 무인교통단속카메라, cctv설치현황, 인도, 버스정류장)

사용한 데이터

1. 오산시_주정차단속(2018~2020).csv

전처리 방법

gid	geometry	주정차단속_여부	주정차단속_여부	주정차단속_여부	주정차단속_횟수
0	다사551085 (POLYGON ((126.9942156468143 37.17418235770403...	False	True	1	39
1	다사551086 (POLYGON ((126.9942096381632 37.17508373885349...	False	True	1	209
2	다사551087 (POLYGON ((126.9942036292448 37.17598511986466...	False	True	1	170
3	다사551088 (POLYGON ((126.9941976200589 37.17688650073755...	False	True	1	17
4	다사552085 (POLYGON ((126.9953420754172 37.17418716043615...	False	True	1	43
...
4505	다사646052 (POLYGON ((127.1013836908642 37.14484471529977...	False	True	1	33
4506	다사646053 (POLYGON ((127.1013789603587 37.14574611426531...	False	True	1	15
4507	다사647050 (POLYGON ((127.1025191345573 37.14304569816435...	False	True	1	19
4508	다사647051 (POLYGON ((127.1025144178335 37.14394709752859...	False	True	1	46
4509	다사647052 (POLYGON ((127.1025097008998 37.1448484967549,...	False	True	1	76

1. 데이터를 불러와 결측치가 존재하는지 확인한다.
2. 단속위치_경도와 단속위치_위도에서 총 188개의 결측치가 확인됐다.
각 결측치에서는 단속 위치를 파악할 수 없으므로 제외한다.
3. 주정차단속 데이터의 경우 개별 데이터가 총 60000개가 넘어 처리에 오랜 시간이 걸린다.
4. 주정차 단속이 발생한 격자를 우선 선별하고, 그 개수를 세어 '격자별정보종합' 데이터에 부여한다.

데이터 전처리

4.5 기타 관련 정보(주정차단속, 무인교통단속카메라, cctv설치현황, 인도, 버스정류장)

사용한 데이터

20.오산시_CCTV 설치현황.csv

전처리 방법

1. 데이터를 불러와 결측치가 존재하는지 확인한다.
2. 설치위치_도로명 주소에서 결측치가 존재하지만, 설치위치는 제시되어 있기 때문에 빼지 않는다.
3. CCTV 유형코드 중 어린이 교통사고와 관련 있는 자료는 C(어린이보호), E(과속단속)이므로, 나머지는 제외한다.

	CCTV 유형코드	설치위치_도로명주소
0	A	성호대로55번길 31-2
1	A	성호대로93번길 45
2	C	성호대로93번길 26
3	A	운천로165번길 52-1
4	A	운천로165번길 5
...
562	E	성호초
563	E	대원초
564	E	원일초
565	E	운천초
566	E	운산초

	CCTV 유형코드	설치위치_도로명주소
0	C	성호대로93번길 26
1	C	성산새싹길 13
2	C	현충로72번길 24
3	C	성호새싹길 8
4	C	대원로29번길 16
...
130	E	성호초
131	E	대원초
132	E	원일초
133	E	운천초
134	E	운산초

데이터 전처리

4.5 기타 관련 정보(주정차단속, 무인교통단속카메라, cctv설치현황, 인도, 버스정류장)

사용한 데이터

- 15.오산시_무인교통단속카메라.csv
- 21.오산시_인도.csv
- 22.오산시_버스정류장.csv

전처리 방법

1. 각 시설물들이 격자의 중심점으로부터 일정 반경 내에 있는지를 판단한 후, 그 개수를 세어 '격자별정보종합' 데이터에 부여한다.

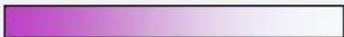
	gid	geometry	무인교통단속카메라_개수	CCTV_개수	인도_개수	버스정류장_개수
12	다사553087	(POLYGON (((126.9964565402404 37.1759947152448,...	0	0	0	1
20	다사554087	(POLYGON (((126.9975829961378 37.17599949687294...	0	0	0	2
21	다사554088	(POLYGON (((126.9975770270969 37.17690087821268...	0	0	0	1
28	다사555087	(POLYGON (((126.9987094523009 37.17600426779312...	0	0	0	1
104	다사561083	(POLYGON (((127.0054916936862 37.17242713739633...	0	0	0	2
...
4358	다사638066	(POLYGON (((127.0923078750135 37.15743363761779...	0	0	3	1
4359	다사638067	(POLYGON (((127.0923030345987 37.15833503365665...	0	0	12	3
4360	다사638068	(POLYGON (((127.0922981939685 37.15923642955755...	0	0	6	1
4385	다사639066	(POLYGON (((127.0934340704674 37.15743750640318...	0	0	2	1
4386	다사639067	(POLYGON (((127.0934292434227 37.1583389025677,...	0	0	8	2

03

EDA

교육도시 오산 어린이를 지켜라

인구 분포 현황 파악

인구 : 적음  많음

시사점

생산가능인구와 고령인구는 그 분포가 유사하다.

유소년 인구의 경우 세교동, 수청동에 주로 분포한다.

고령, 생산가능인구와 달리, 유소년 인구는 남촌동, 가수동, 중앙동에서 매우 낮은 인구 분포도를 보이고 있다.

#1. 총인구



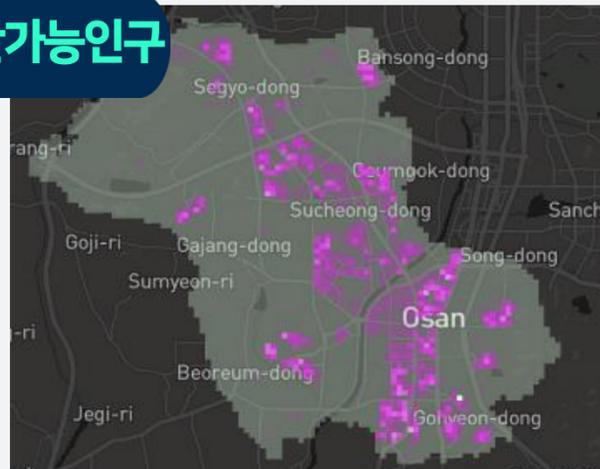
gid	coord_cent	val_total	정규화_시각화
4012	다사629042 [127.08285157938758, 37.13621749920832]	3991.0	1.000000
3611	다사620036 [127.07274908293027, 37.13077302244187]	2365.0	0.592583
1726	다사594081 [127.04323452122895, 37.17122652074008]	2304.0	0.577299
3955	다사627065 [127.08048532618662, 37.15694167141951]	2249.0	0.563518
3787	다사623062 [127.07599568365859, 37.15422145650601]	2185.0	0.547482

#2. 고령인구



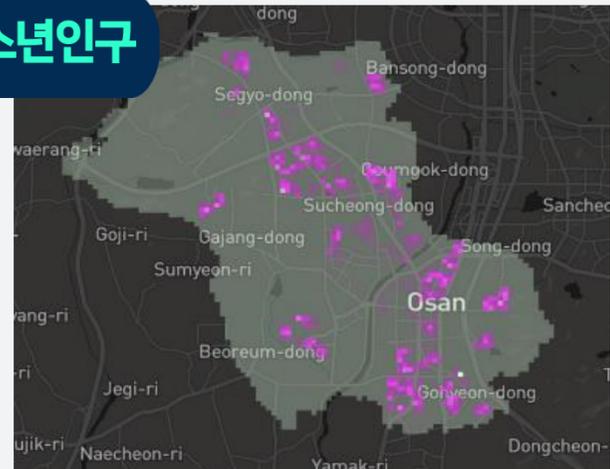
gid	coord_cent	val_senior	정규화_시각화
4012	다사629042 [127.08285157938758, 37.13621749920832]	2069.0	1.000000
3611	다사620036 [127.07274908293027, 37.13077302244187]	1240.0	0.599323
1726	다사594081 [127.04323452122895, 37.17122652074008]	1197.0	0.578540
3955	다사627065 [127.08048532618662, 37.15694167141951]	1168.0	0.564524
3787	다사623062 [127.07599568365859, 37.15422145650601]	1137.0	0.549541

#3. 생산가능인구



gid	coord_cent	val_work	정규화_시각화
4012	다사629042 [127.08285157938758, 37.13621749920832]	1500.0	1.000000
3611	다사620036 [127.07274908293027, 37.13077302244187]	963.0	0.642000
3955	다사627065 [127.08048532618662, 37.15694167141951]	899.0	0.599333
3787	다사623062 [127.07599568365859, 37.15422145650601]	886.0	0.590667
1726	다사594081 [127.04323452122895, 37.17122652074008]	830.0	0.553333

#4. 유소년인구



gid	coord_cent	val_junior	정규화_시각화
4012	다사629042 [127.08285157938758, 37.13621749920832]	422.0	1.000000
1726	다사594081 [127.04323452122895, 37.17122652074008]	277.0	0.656398
1505	다사591097 [127.03976786087675, 37.18563563326318]	258.0	0.611374
4319	다사637057 [127.09178596901121, 37.14976982738334]	229.0	0.542654
1651	다사593086 [127.04208093059681, 37.175729111958454]	210.0	0.497630

EDA

어린이 교통사고, 어린이 보호구역 현황 파악

어린이 교통사고 : 적음  많음

어린이 보호구역 : 

시사점

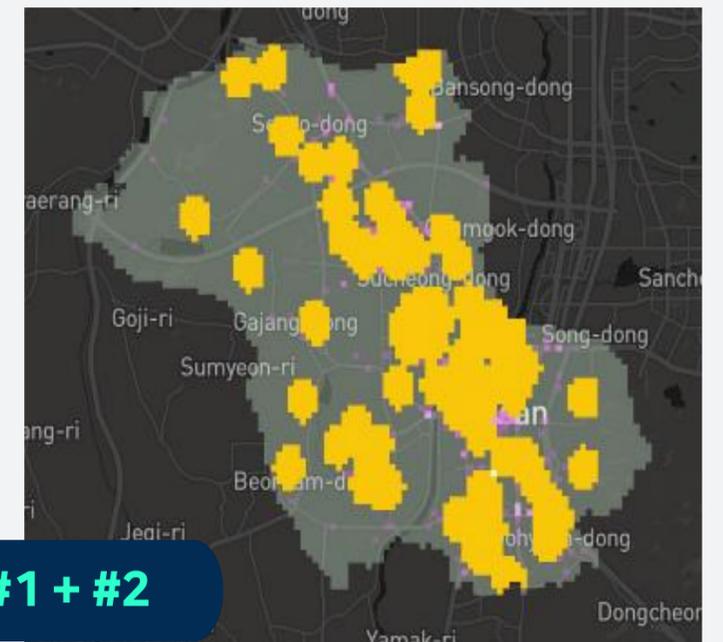
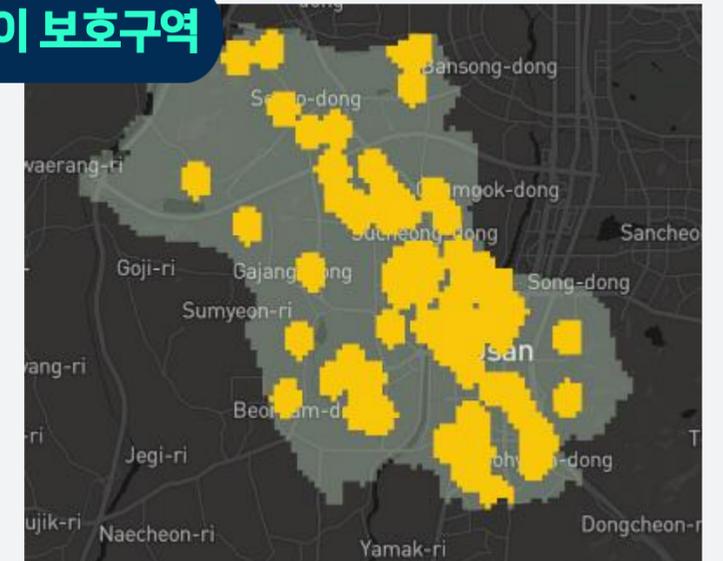
어린이교통사고는 유소년 인구가 많았던 지역 뿐만 아니라, 낮은 분포도를 보였던 남촌동, 가수동 등에서도 많이 발생했다. 어린이 보호구역 시각화 결과, 남촌동, 가수동, 중앙동의 경우 실제 유소년 거주 인구는 적지만 학교, 어린이집 등이 많이 분포되어 있어 대부분이 보호구역으로 지정되어 있다. 하지만 두 데이터를 겹친 결과, 대부분의 교통사고가 보호구역인 곳에서 발생했음을 알 수 있으며, 어린이 보호구역 내의 안전 시설물 설치의 시급성을 확인할 수 있다. 또한 보호구역 주위에도 교통사고가 다수 발생하고 있어 교통사고 위험 지역의 지정이 필요해 보인다.

#1. 어린이 교통사고



gid	coord_cent	accident_cnt	정규화_시각화
3681	다사621048 [127.07381421297224, 37.141593841344836]	10	1.0
3617	다사620042 [127.07271867315109, 37.136181407501454]	7	0.7
3863	다사625056 [127.07827801072962, 37.148821120367984]	6	0.6
3828	다사624061 [127.07712684904553, 37.15332408507212]	6	0.6
3098	다사612101 [127.0634058861932, 37.18933072437019]	6	0.6

#2. 어린이 보호구역



#1 + #2

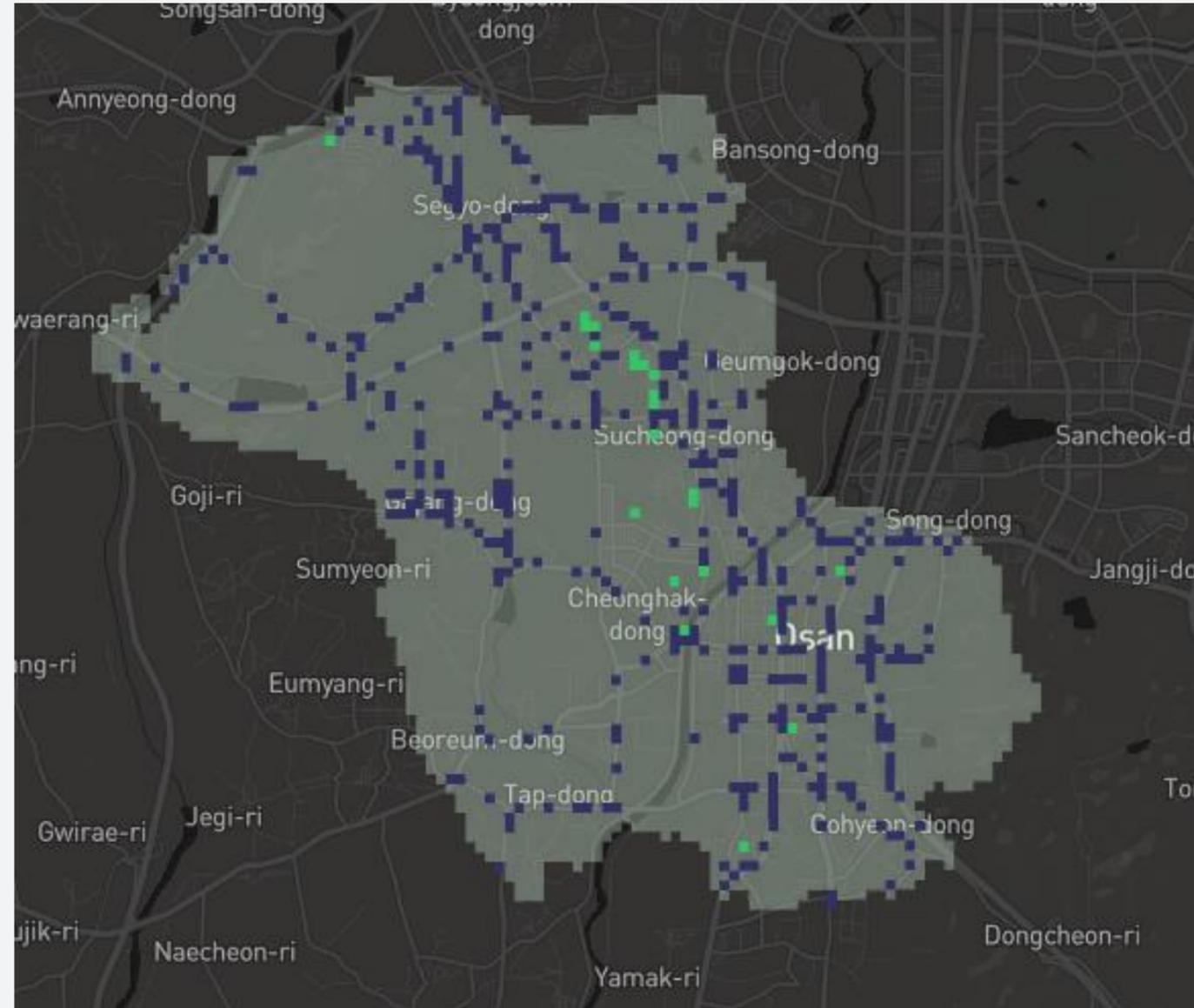
도로시설물(과속방지턱, 도로안전표지, 횡단보도, 신호등)

도로안전표지 : ■

과속방지턱 : ■

시사점

과속방지턱은 차량의 감속과 안전운전 주의 지역 상기 등 역할을 하며, 도로안전표지는 운전자에게 정보를 안내하거나 위험을 상기시키는 도로시설물을 의미한다. 이에 두 데이터와 교통사고 수는 반비례하는 경향을 가진다고 볼 수 있다.



도로시설물(횡단보도, 신호등, 과속방지턱, 도로안전표지)

횡단보도 : ■

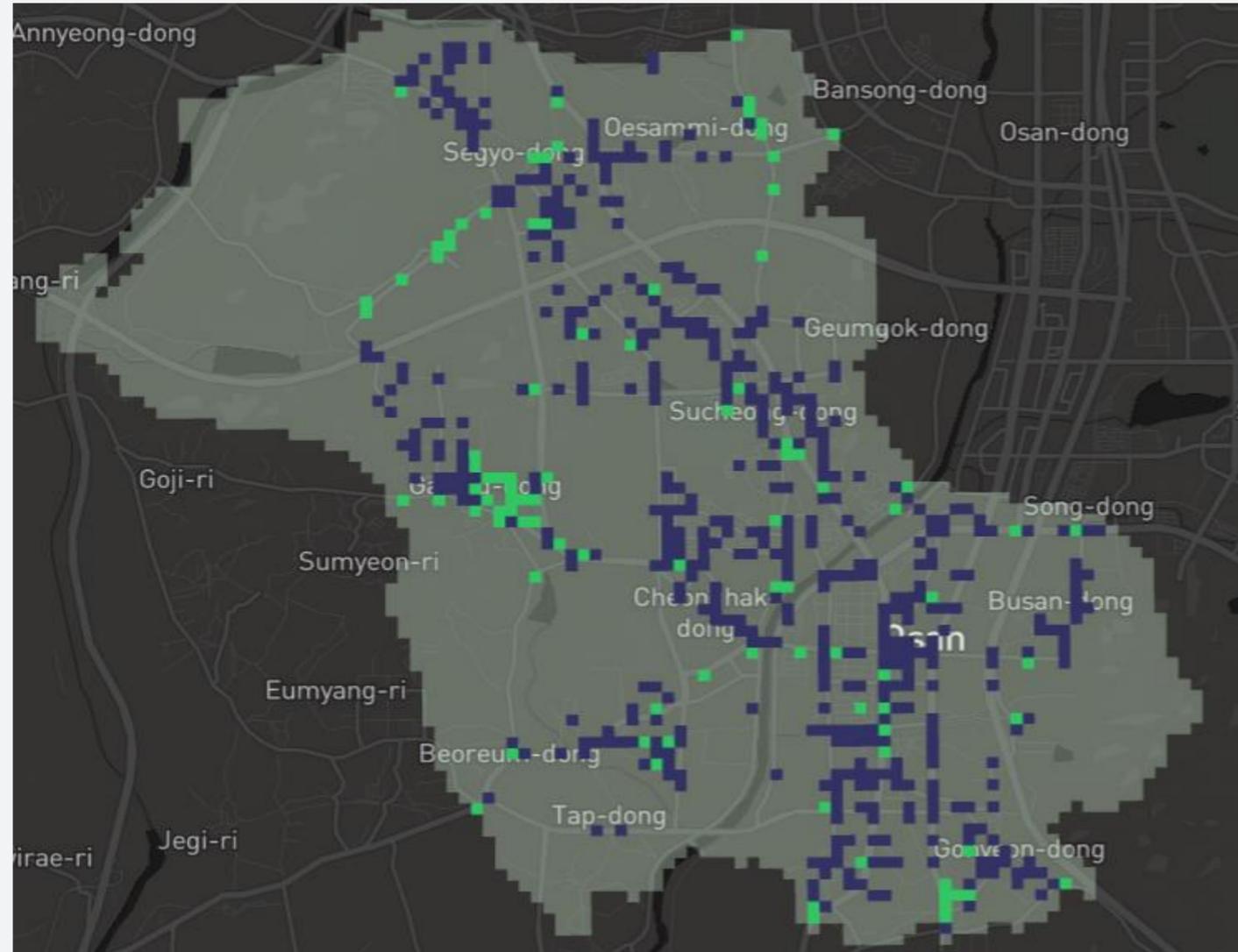
신호등 : ■

시사점

횡단보도와 신호등은 보행자가 도로를 가로질러 통행 가능하기 때문에 교통사고와 관련이 많은 데이터이다.

이에 두 데이터와 교통사고 수는 큰 상관관계를 가진다고 볼 수 있다.

따라서 횡단보도와 신호등 데이터를 중심으로 어린이 보호구역을 선정해야 한다.



주요 시설 위치

어린이집, 유치원 : ■

체육시설 : ■

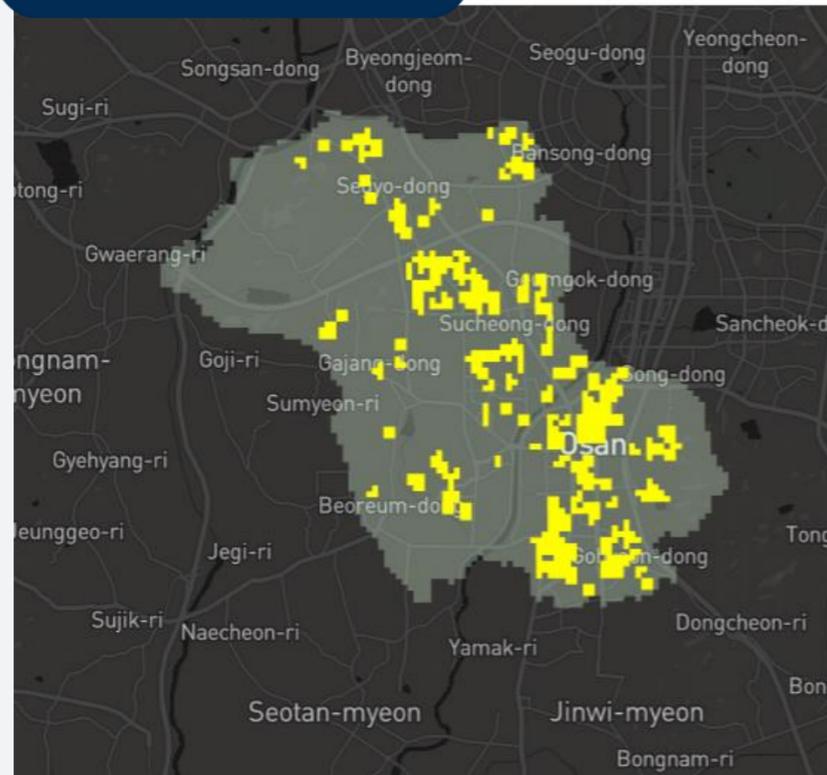
학원 및 교습소 : ■

시사점

대부분의 어린이들은 방과 후 체육시설, 학원 및 교습소에 방문하기 때문에 두 데이터는 연관성이 높다.

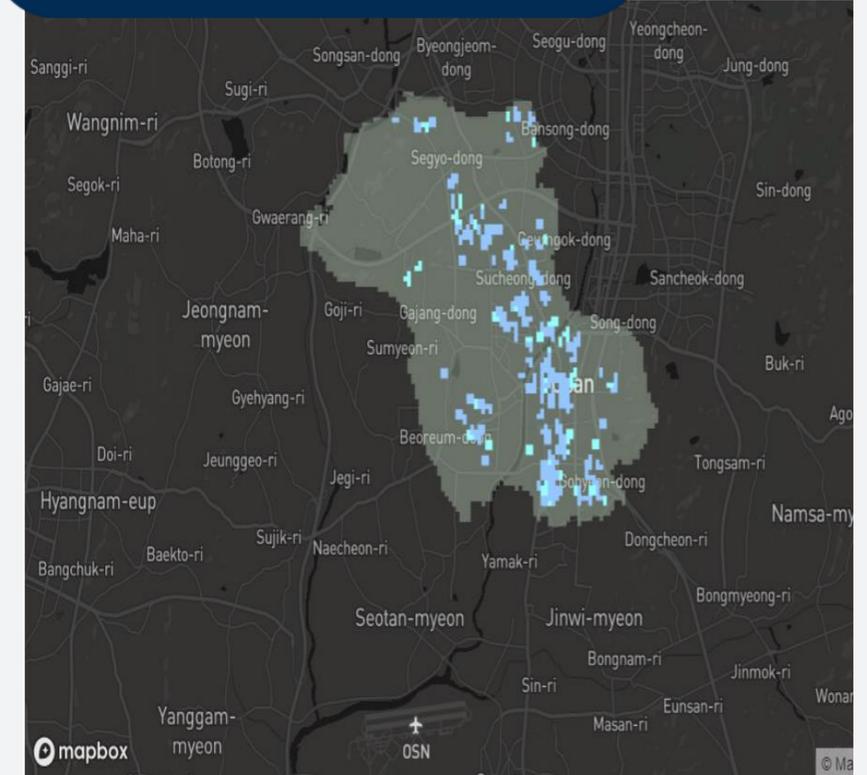
어린이 교통사고 발생과 이 같은 주요시설 간의 상관관계를 분석한다면 기존 어린이 보호구역 외 어린이 교통사고 위험지역 20개를 선정할 수 있을 것이다.

#1. 어린이집, 유치원



gid	geometry	val_total	val_senior	val_work	val_junior	accident_cnt	grid_id	coordinates	lon_min	lon_max	lat_min	lat_max
0	MULTIPOLYGON (((126.99422 37.17418, 126.99421 ...	0.0	0.0	0.0	0.0	0	00000	[[126.99421564681425, 37.17418235770403], [126...	126.994210	126.995342	37.174182	37.174182
1	MULTIPOLYGON (((126.99421 37.17508, 126.99420 ...	0.0	0.0	0.0	0.0	0	00001	[[126.99420963816323, 37.17508373885349], [126...	126.994204	126.995336	37.175084	37.175084
2	MULTIPOLYGON (((126.99420 37.17599, 126.99420 ...	0.0	0.0	0.0	0.0	0	00002	[[126.99420362924478, 37.17598511986466], [126...	126.994198	126.995330	37.175985	37.175985
3	MULTIPOLYGON (((126.99420 37.17689, 126.99419 ...	0.0	0.0	0.0	0.0	0	00003	[[126.9941976200589, 37.1768850073755], [126...	126.994192	126.995324	37.176887	37.176887
4	MULTIPOLYGON (((126.99534 37.17419, 126.99534 ...	0.0	0.0	0.0	0.0	0	00004	[[126.99534207541716, 37.174187160436155], [12...	126.995336	126.996469	37.174187	37.174187

#2. 체육시설, 학원 및 교습소



gid	geometry	val_total	val_senior	val_work	val_junior	accident_cnt	grid_id	coordinates	lon_min	lon_max	lat_min	lat_max
0	MULTIPOLYGON (((126.99422 37.17418, 126.99421 ...	0.0	0.0	0.0	0.0	0	00000	[[126.99421564681425, 37.17418235770403], [126...	126.994210	126.995342	37.174182	37.174182
1	MULTIPOLYGON (((126.99421 37.17508, 126.99420 ...	0.0	0.0	0.0	0.0	0	00001	[[126.99420963816323, 37.17508373885349], [126...	126.994204	126.995336	37.175084	37.175084
2	MULTIPOLYGON (((126.99420 37.17599, 126.99420 ...	0.0	0.0	0.0	0.0	0	00002	[[126.99420362924478, 37.17598511986466], [126...	126.994198	126.995330	37.175985	37.175985
3	MULTIPOLYGON (((126.99420 37.17689, 126.99419 ...	0.0	0.0	0.0	0.0	0	00003	[[126.9941976200589, 37.1768850073755], [126...	126.994192	126.995324	37.176887	37.176887
4	MULTIPOLYGON (((126.99534 37.17419, 126.99534 ...	0.0	0.0	0.0	0.0	0	00004	[[126.99534207541716, 37.174187160436155], [12...	126.995336	126.996469	37.174187	37.174187

실제 도로망 및 9시 / 14시 교통량 분석

오산시 도로망 

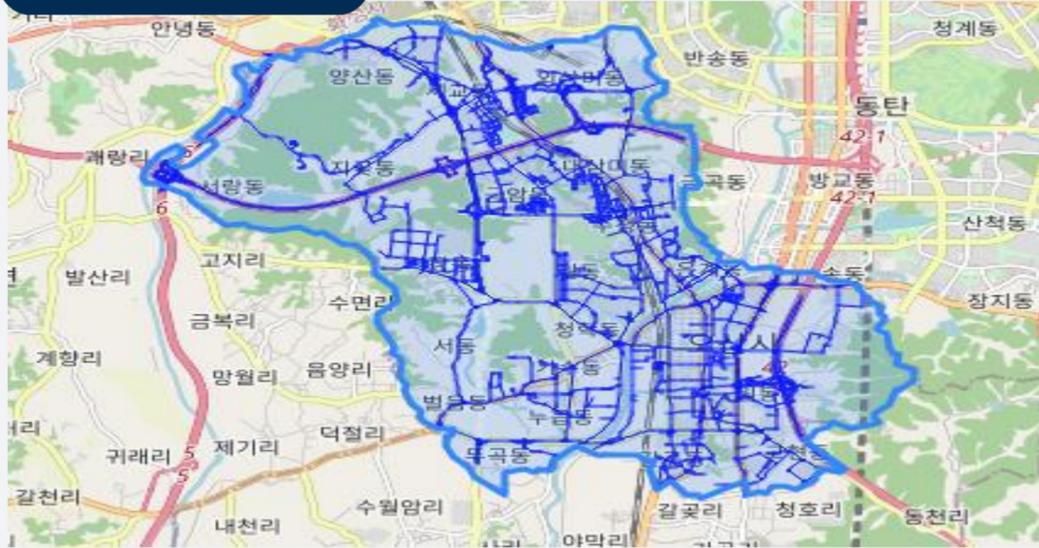
교통량 : 적음  많음 (두께 : 교통량)

시사점

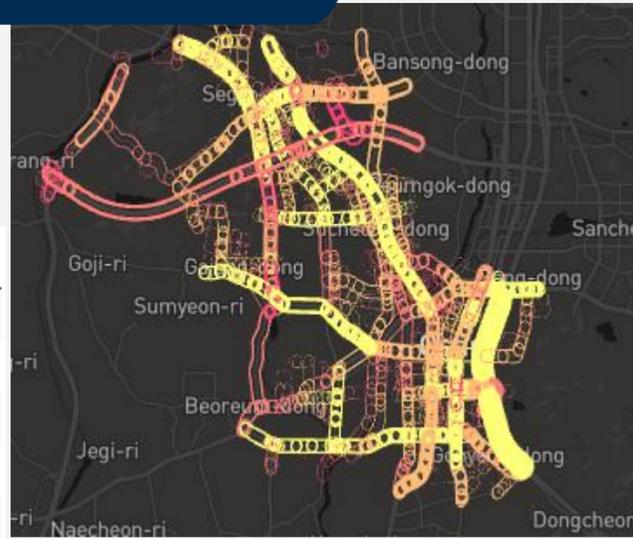
등교(등원) 시간대와 하교(하원) 시간대의 교통량 분포는 비슷한 경향을 보인다.
 특히 경부고속도로 오산IC 구간과 수청로, 경기대로의 경우 차선이 좁음에도 불구하고 교통량이 많은 모습을 보인다.
 경기대로의 경우 오산시의 중심과 대부분의 어린이 보호구역을 지나가는데, 교통량이 많은 모습을 보여 교통사고와의 상관관계를 보이고 있다.

	link_id	추정교통량_9시	정규화_도로폭
0	478344684	1117.07	0.25
1	478344686	543.58	0.50
2	478344690	179.63	0.25
3	478344696	248.64	0.25
4	478344700	521.28	0.25

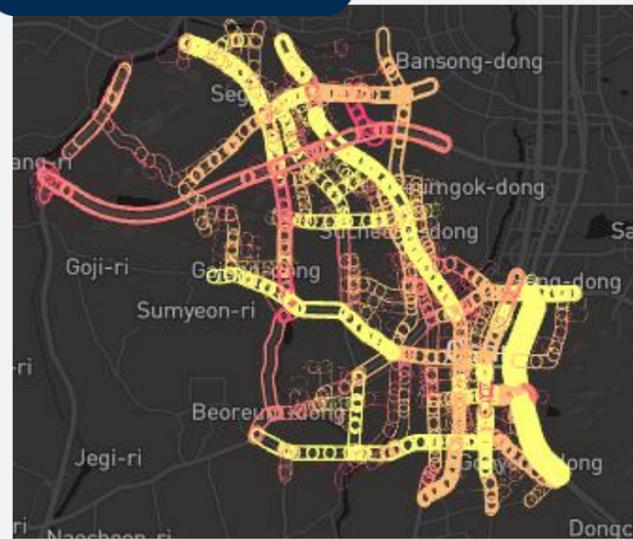
#1. 오산시 도로망



#2. 9시 교통량



#3. 14시 교통량



	link_id	추정교통량_14시	정규화_도로폭
0	478344684	1112.77	0.25
1	478344686	665.88	0.50
2	478344690	163.52	0.25
3	478344696	205.77	0.25
4	478344700	463.38	0.25

혼잡시간강도 / 혼잡빈도강도 분석

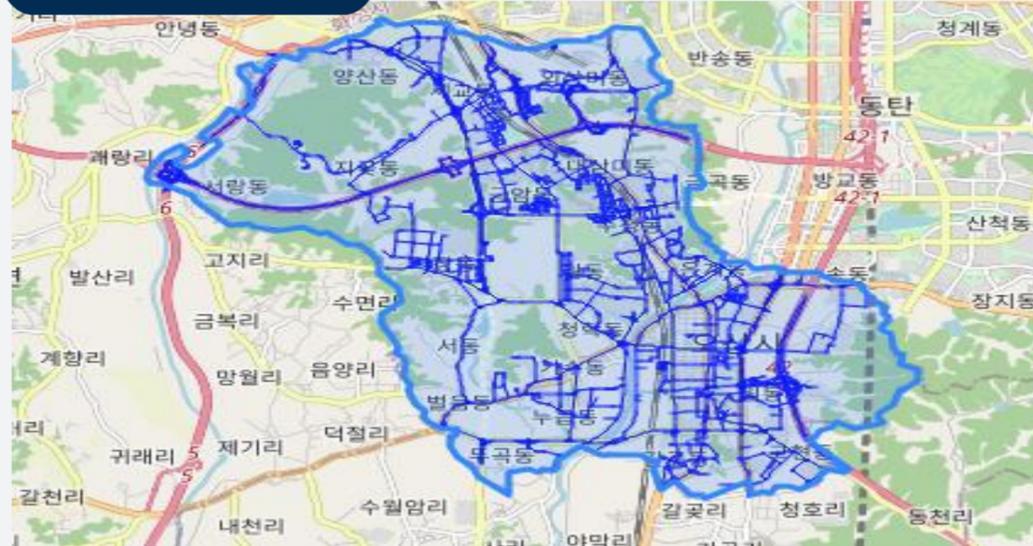
오산시 도로망 

혼잡강도 : 적음  많음 (두께 : 교통량)

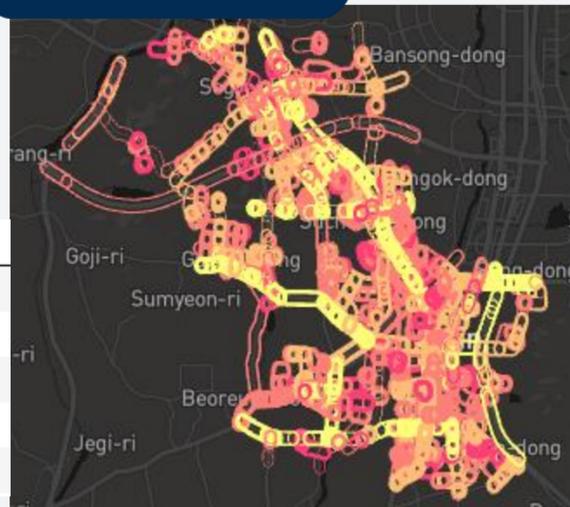
시사점

경부고속도로 오산IC 구간과 수청로, 경기대로의 경우 차량의 이동량이 특히 많다. 이는 교통량이 많은 동시에 도로폭도 좁은 점에서 기인한 것으로 보인다.

#1. 오산시 도로망

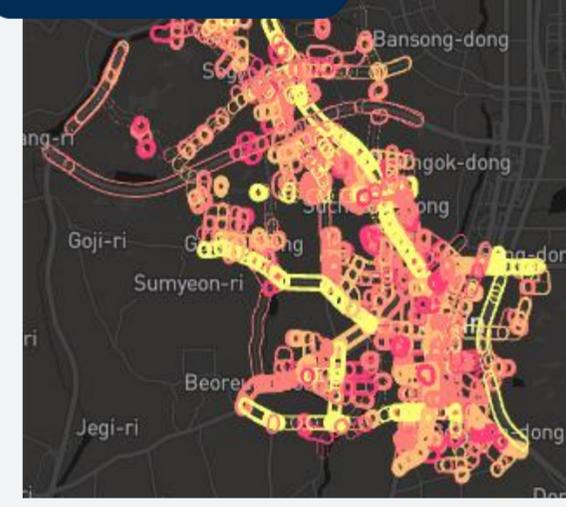


#2. 혼잡시간강도



	link_id	혼잡시간강도	정규화_도로폭
0	478344684	0.0	0.25
1	478344686	0.0	0.50
2	478344690	0.0	0.25
3	478344696	0.0	0.25
4	478344700	0.0	0.25

#3. 혼잡빈도강도



	link_id	혼잡빈도강도	정규화_도로폭
0	478344684	0.0	0.25
1	478344686	0.0	0.50
2	478344690	0.0	0.25
3	478344696	0.0	0.25
4	478344700	0.0	0.25

04

Modeling

교육도시 오산 어린이를 지켜라

Modeling

학습 데이터 설정

특정 요소에 편향된 영향력을 최소화하기 위해 전처리 된 데이터 정규화

도로 데이터를 가진 행만 학습 데이터로 설정

HOW?

특정 요소 정규화값 = 특정 요소 데이터 / 특정 요소 데이터 최대값

'정규화_총인구', '정규화_유소년인구', '정규화_추정교통량_8시', '정규화_추정교통량_9시',
'정규화_추정교통량_13시', '정규화_추정교통량_14시', '정규화_추정교통량_15시',
'정규화_추정교통량_16시', '정규화_추정교통량_17시', '정규화_추정교통량_18시',
'정규화_혼잡빈도강도', '정규화_혼잡시간강도', '정규화_횡단보도_개수', '정규화_신호등_개수',
'정규화_과속방지턱_개수', '정규화_도로안전표지_개수', '정규화_어린이집_개수', '정규화_유치원_개수',
'정규화_초등학교_개수', '정규화_체육도장_개수', '정규화_학원및교습소_개수',
'정규화_무인교통단속카메라_개수', '정규화_CCTV_개수', '정규화_인도_개수', '정규화_버스정류장_개수',
'정규화_주정차단속_횟수', '정규화_어린이교통사고'

```
#교통량이 없는 행을 뺍니다.  
df3 = df_result.copy()  
df3.loc[df3['추정교통량_8시'] == 0] = np.nan  
df3.loc[df3['추정교통량_8시'] == '0'] = np.nan  
df3.dropna(subset = ['추정교통량_8시'], inplace = True)
```

**어린이 교통사고가 발생할 수 있는
도로 데이터만 남김**

Modeling

독립변수와 종속변수 설정

적절한 독립변수 설정 필요성 인식

독립변수

적절한 독립변수만
설정하는 작업이 필요함

종속변수

```
독립변수_후보군=  
df3[["정규화_총인구",  
"정규화_유소년인구",  
"정규화_추정 교통량_8시",  
"정규화_추정 교통량_9시",  
"정규화_추정 교통량_14시",  
"정규화_추정 교통량_15시",  
'정규화_추정 교통량_16시',  
'정규화_추정 교통량_17시',  
'정규화_추정 교통량_18시',  
'정규화_혼잡빈도강도',  
'정규화_혼잡시간강도',  
'정규화_횡단보도_개수',  
'정규화_신호등_개수',  
'정규화_과속방지턱_개수',  
'정규화_도로안전표지_개수',  
'정규화_어린이집_개수',  
'정규화_유치원_개수',  
'정규화_초등학교_개수',  
'정규화_체육도장_개수',  
'정규화_학원및교습소_개수',  
'정규화_무인교통단속카메라_개수',  
'정규화_CCTV_개수',  
'정규화_인도_개수',  
'정규화_버스정류장_개수',  
'정규화_주정차단속_횟수']]
```

```
df_result['정규화_어린이교통사고']
```

Modeling

독립변수와 종속변수 설정

독립변수 설정하기

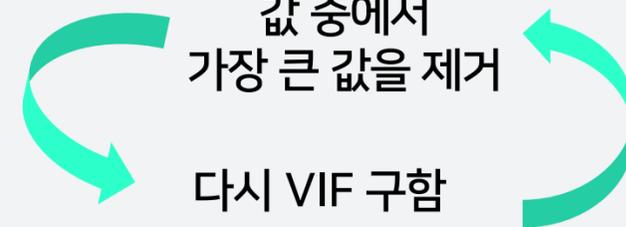
분산팽창요인

분산팽창요인
(Variance Inflation Factor)을 구하여
이 값이 10을 넘는다면 보통
다중공선성의 문제가 있으므로
vif가 10인 독립변수들을 하나씩 제거

	VIF Factor	features
0	11.65682	정규화_총인구
1	10.16679	정규화_유소년인구
2	1274.72099	정규화_추정교통량_8시
3	2668.46735	정규화_추정교통량_9시
4	18863.91910	정규화_추정교통량_14시
5	1239564.84452	정규화_추정교통량_15시
6	1218420.62787	정규화_추정교통량_16시
7	4871.38354	정규화_추정교통량_17시
8	1939.90555	정규화_추정교통량_18시
9	81.47756	정규화_혼잡빈도강도
10	98.32872	정규화_혼잡시간강도
11	3.53445	정규화_횡단보도_개수
12	3.05467	정규화_신호등_개수
13	1.13538	정규화_과속방지턱_개수
14	2.52320	정규화_도로안전표지_개수
15	1.78003	정규화_어린이집_개수
16	2.30197	정규화_유치원_개수
17	2.17152	정규화_초등학교_개수
18	1.81525	정규화_체육도장_개수
19	1.80490	정규화_학원및교습소_개수
20	1.59302	정규화_무인교통단속카메라_개수
21	1.45018	정규화_CCTV_개수
22	1.98283	정규화_인도_개수
23	1.47879	정규화_버스정류장_개수
24	1.47324	정규화_주정차단속_횟수

vif가 10이상인
값 중에서
가장 큰 값을 제거

다시 VIF 구함



Modeling

다중선형회귀모델을 통한 위험도 측정

다중선형회귀분석

Formulation

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i$$

-X = '정규화_유소년인구', '정규화_혼잡빈도강도', '정규화_횡단보도_개수', '정규화_신호등_개수',
'정규화_과속방지턱_개수', '정규화_도로안전표지_개수', '정규화_어린이집_개수', '정규화_유치원_개수',
'정규화_초등학교_개수', '정규화_체육도장_개수', '정규화_학원및교습소_개수', '정규화_무인교통단속카메라_개수',
'정규화_CCTV_개수', '정규화_인도_개수', '정규화_버스정류장_개수', '정규화_주정차단속_횟수',
'출퇴근 추정 교통량'=('정규화_추정교통량_8시'+ '정규화_추정교통량_8시')/2)

-Y = "정규화_어린이교통사고"

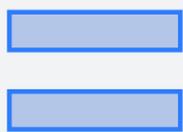
정규화	
유소년인구	0.01229007
혼잡빈도강도	0.06842072
횡단보도_개수	-0.03790445
신호등_개수	0.10468282
과속방지턱_개수	-0.02650778
도로안전표지_개수	0.05031891
어린이집_개수	-0.00567461
유치원_개수	-0.00885286
초등학교_개수	0.03600286
체육도장_개수	0.05638707
학원및교습소_개수	0.09553492
무인교통단속카메라_개수	0.00034711
CCTV_개수	0.04165476
인도_개수	0.01515244
버스정류장_개수	0.00575875
주정차단속_횟수	0.04527366
출퇴근 추정 교통량	0.0577561

Modeling

다중선형회귀모델을 통한 위험도 측정

다중선형회귀분석

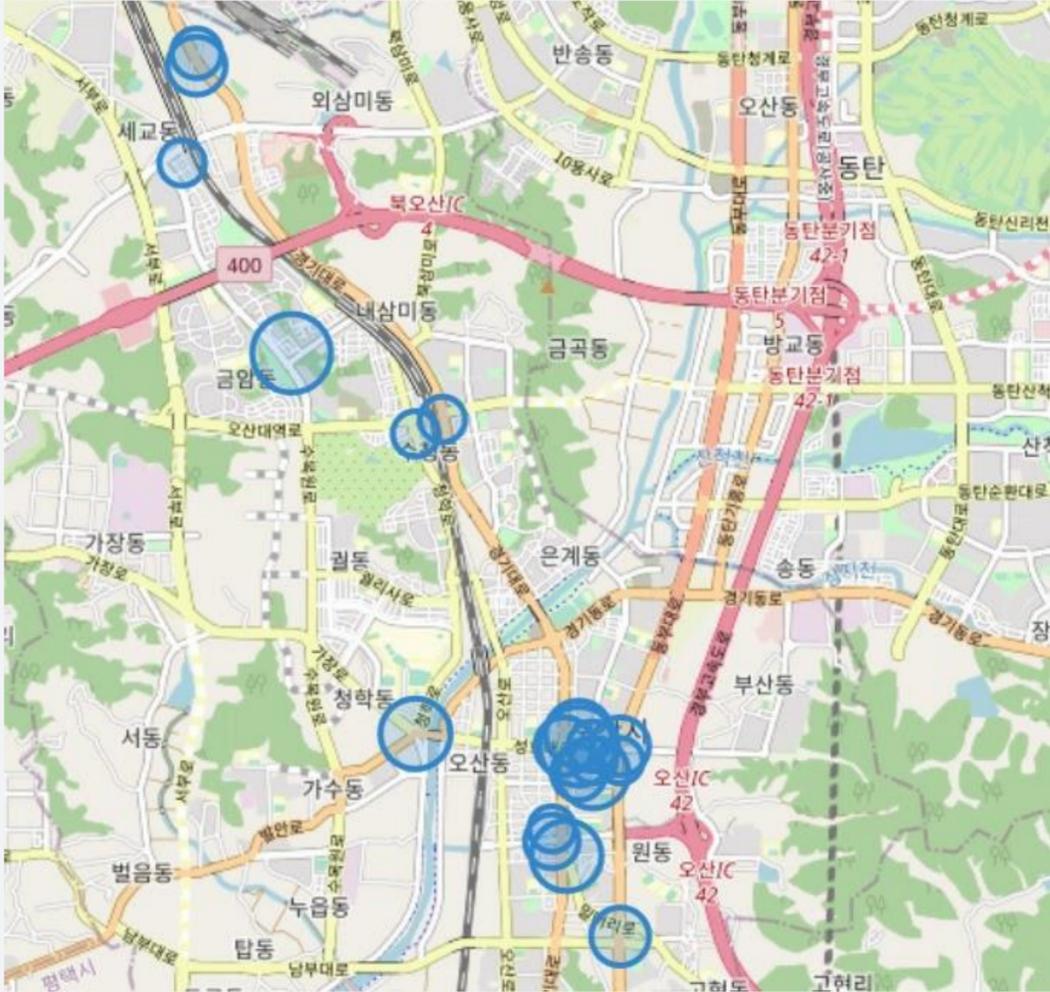
정규화			
유소년인구		0.01229007	
혼잡빈도강도		0.06842072	
횡단보도_개수		-0.03790445	
신호등_개수		0.10468282	
과속방지턱_개수		-0.02650778	
도로안전표지_개수		0.05031891	
어린이집_개수		-0.00567461	
유치원_개수		-0.00885286	
초등학교_개수		0.03600286	
체육도장_개수		0.05638707	
학원및교습소_개수		0.09553492	
무인교통단속카메라_개수		0.00034711	
CCTV_개수		0.04165476	
인도_개수		0.01515244	
버스정류장_개수		0.00575875	
주정차단속_횟수		0.04527366	
출퇴근 추정 교통량		0.0577561	

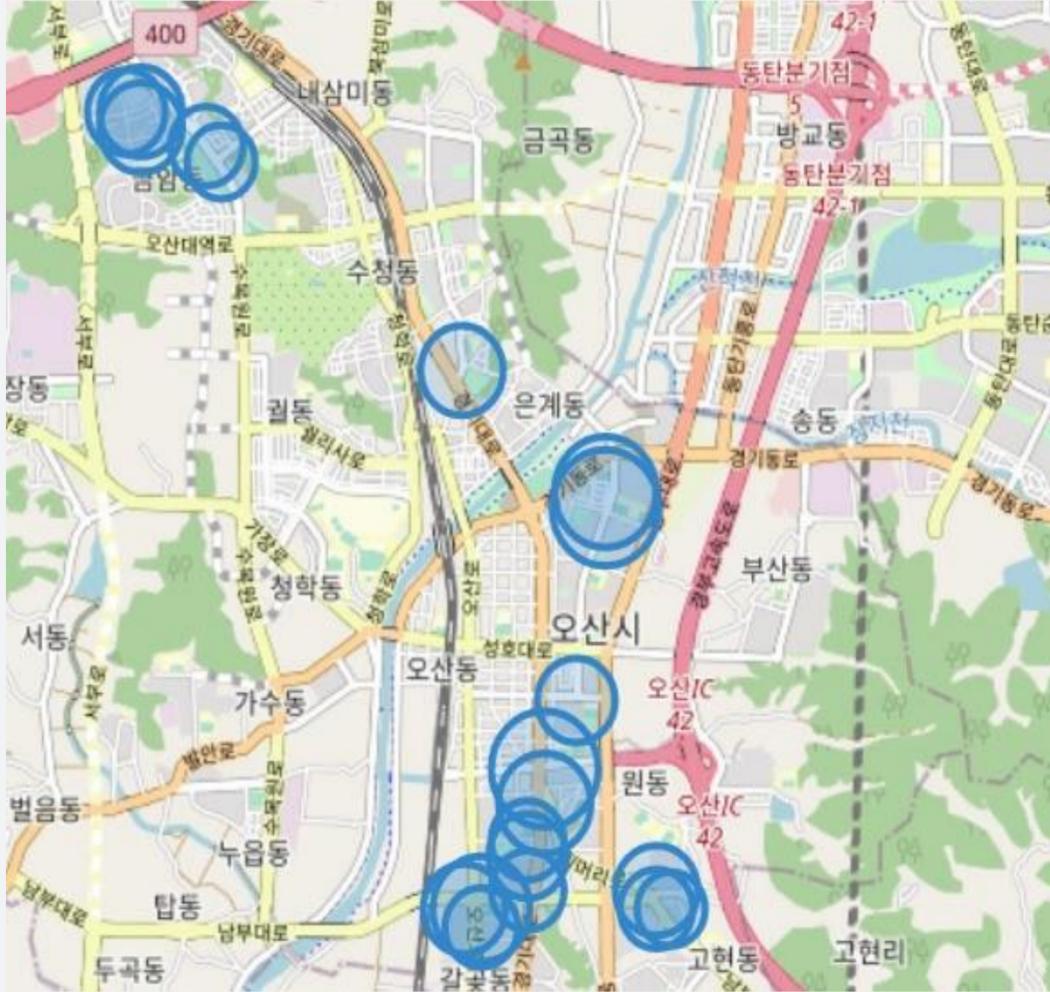
Modeling

다중선형회귀모델을 통한 위험도 측정

원의 반경은 위험도에 비례함



어린이보호구역이 아닌 위험도 상위 20곳

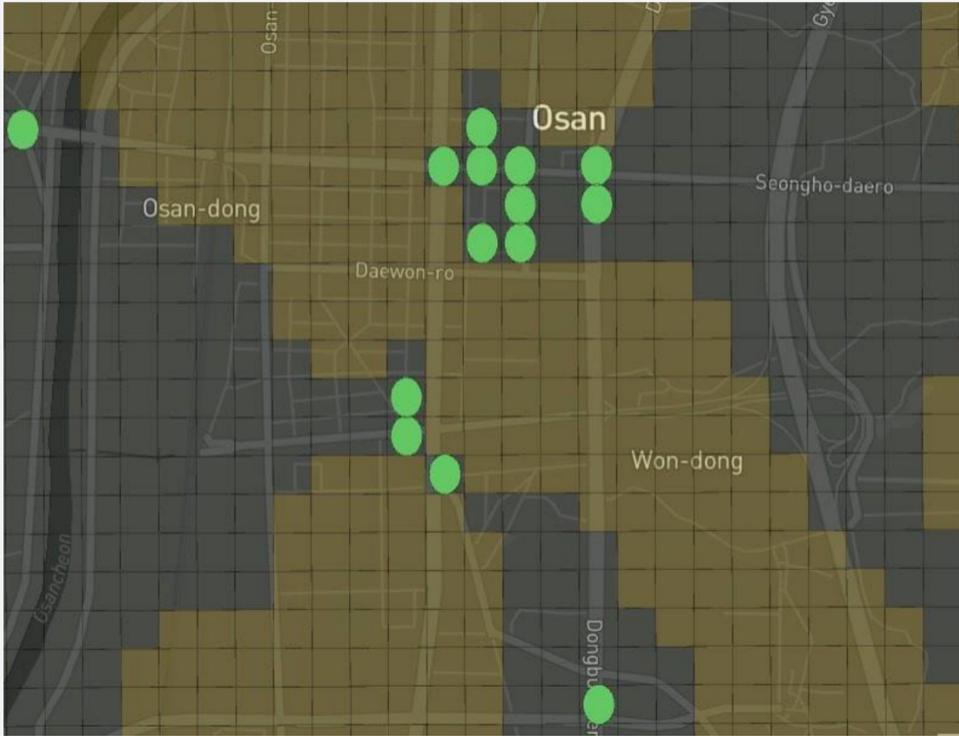


어린이보호구역 위험도 상위 20곳

Modeling

최종 결과 및 시각화

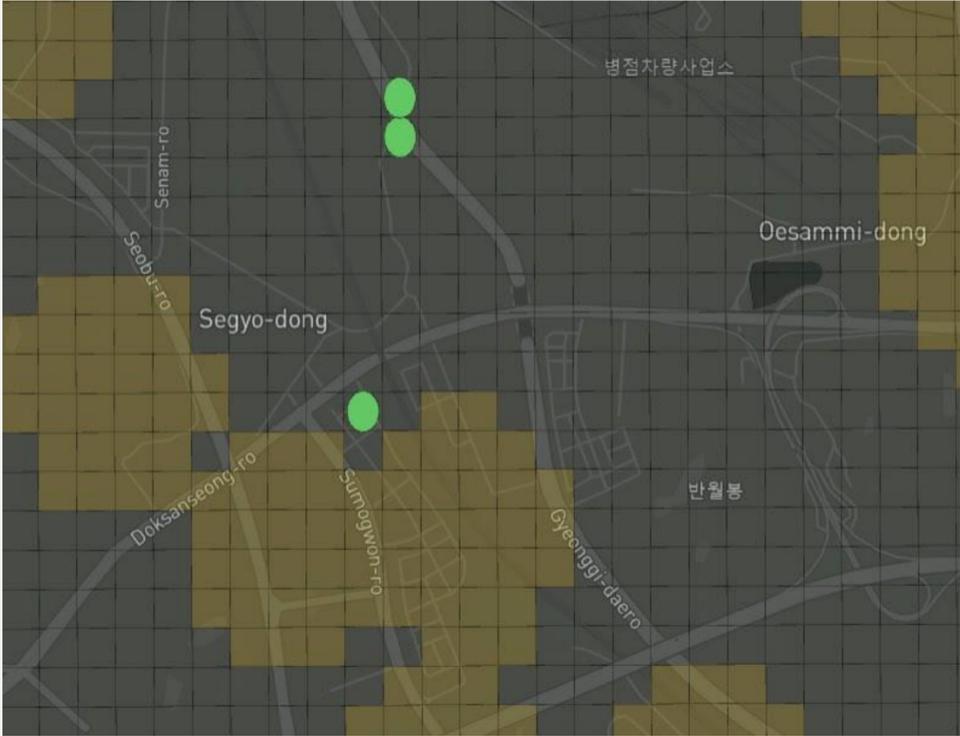
어린이 보호구역 외 어린이 교통사고 위험지역 20개소



원동, 오산동 위험지역 분포
특히 오산 시청 근처에 위험지역이 몰려있음
남촌오거리도 위험지역으로 분류된 모습 확인 가능



수청동, 금암동의 위험지역 분포
물향기공원 앞 수청로, 경기대로
고인돌공원 앞 수목원로에 위험지역 확인 가능



세교동의 위험지역 분포
수목원로와 독산성로 연결지점 근처와
경기대로 부근 위험지역 확인 가능

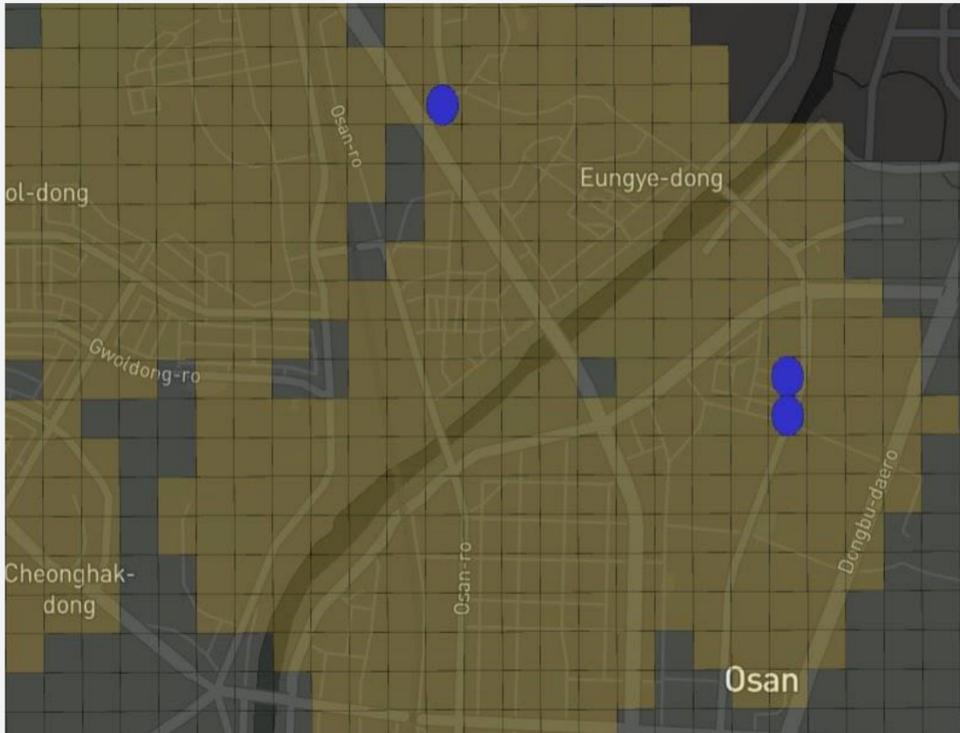
Modeling

최종 결과 및 시각화

기존 어린이 보호구역 중 교통안전시설물 우선 설치 지역 20개소



원동, 오산동 교통안전시설물 우선 설치 지역
원동사거리, 한전사거리와 기타 삼거리 근처에
위험지역 몰려있음을 확인 가능



은계동 교통안전시설물 우선 설치 지역
은암주공단지, 수청공원 근처 위험지역 확인 가능



금암동 교통안전시설물 우선 설치 지역
고인돌공원 앞, 금암마을휴먼시아5단지 근처 위험지역 확인 가능

교육도시 오산, 어린이를 지켜라

THANK YOU

팀 데린이보호구역