어린이 교통사고 위험지역 도출

Team - Apdo

INTRODUCTION

분석 목적

› 어린이 보호 구역 외 교통사고 위험지역 20개소 제시

› 기존 어린이 보호구역 중 교통안전시설물(과속단속카메라, 교통안전시설물 등) 우선 설치 지역 20개소 제시

분석 프로세스

EDA 및 변수 생성

- 오산시 내 4510개의 격자(grid)마다 지리, 인구, 도로, 시설물 등 관련 정보 삽입 (지역 특성 변수 생성)
- 지역 특성 변수 시각화. 오산시 교통사고 발생 현황 및 문제점 파악
- 어린이 보호구역 외 교통사고 다발 지역, 어린이 보호구역 내 시설물 설치 미흡 지역 시각화

어린이 교통사고 위험구역 도출

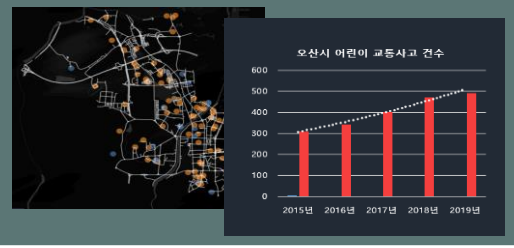
- 어린이 교통사고와 지역 특성 간 관계 파악
  1) 다중공선성 문제 해결 및 1차 변수 선정
  2) 변수선택법을 통한 2차 변수 선정
  3) 음이항 회귀 모형 적합, 결과 확인
- 모형 coefficient를 사용하여 도로 별 교통사고 위험지수 계산
- MCLP 모형 활용하여 최종 위험구역 도출

안전시설물 우선 설치지역 도출

- 각 안전시설물과 지역 특성 간 관계 파악
  1) 교통사고 위험구역 도출 단계에서의 1차 선정된 변수 사용
  2) 변수선택법을 통한 시설물 별 중요 변수 2차 선정
  3) 다중 회귀모형 적합, 시설물 별 설치 필요지수 계산
- 필요 설치지수 총합과 실제 설치 정도의 차를 이용하여 MCLP 모형 활용. 우선 설치지역 최종 도출

분석 결과

- 생성된 지역 특성 변수 확인
- 오산시 지역 특성 시각화 및 시사점 제시



- 도로 별 교통사고 위험지수 정의 (w_hazard)
- 위험 구역 도출 결과 확인 후 Kmeans clustering 을 활용한 조정 진행

- 어린이 보호구역 별 시설물 설치 필요지수 정의 (w_facility)
- 설치 필요 지역 도출 후 Kmeans clustering을 활용한 조정진행

INDEX

01

1. EDA 및 변수 생성

- 오산시 교통사고 현황 및 문제점
- 어린이 교통사고 및 교통안전 시설물 관련 변수 생성

02

2. 어린이 교통사고 위험지역 제시

- 교통사고 관련 설명 변수 선택
- 음이항 회귀모형 및 MCLP을 통한 위험지역 도출

03

3. 교통안전시설물 우선 설치 지역 제시

- 시설물 별 관련 변수 선택 및 모형 적합
- 시설물 부족 지수를 통한 우선 설치 지역 도출

04

4. 최종 결과 및 평가

- 선정 지역 확인 및 기대효과 제시
- 한계점 및 의의

1. EDA 및 분석 데이터 생성

01

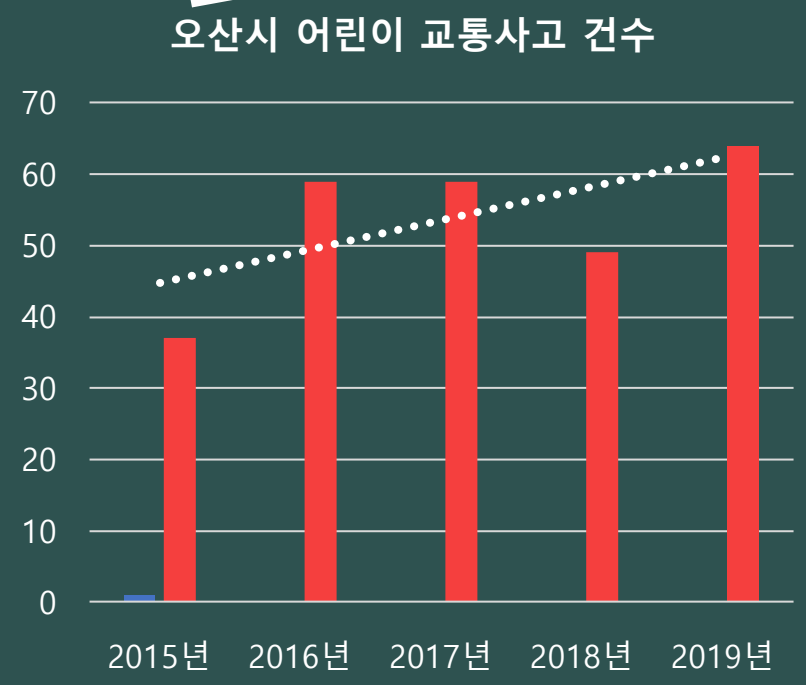
02

03

04

1-1. EDA

오산시 어린이 교통사고 증가 추이



	연도	사건종류	12세 이하
0	2015	사고건수	37
3	2016	사고건수	59
6	2017	사고건수	59
9	2018	사고건수	49
12	2019	사고건수	64

사용 Data: (외부데이터) TAAS_연령별 어린이 사상자.xls

- 교육 도시로 성장하고 있는 오산시, 어린이 교통사고는 증가추세 및 2018년에 비해 큰 폭으로 상승

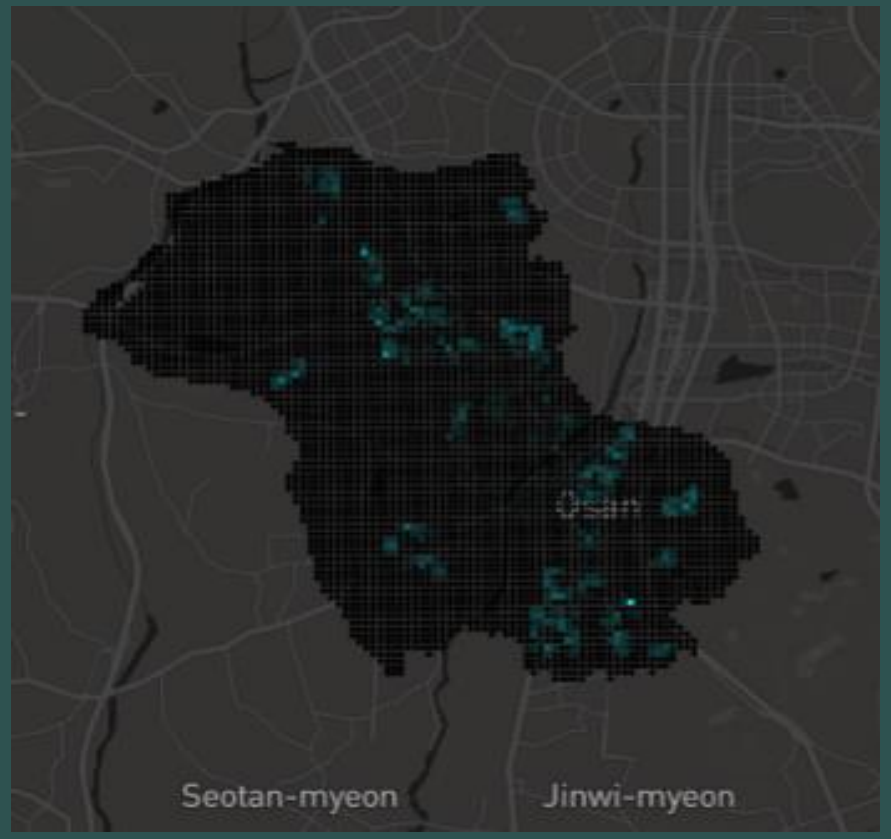
- 어린이 보호구역 외에 추가적인 위험 지역 선정 및 관리 필요

- 어린이 보호구역 내 부족한 시설물 보강 또한 필요

1-1. EDA

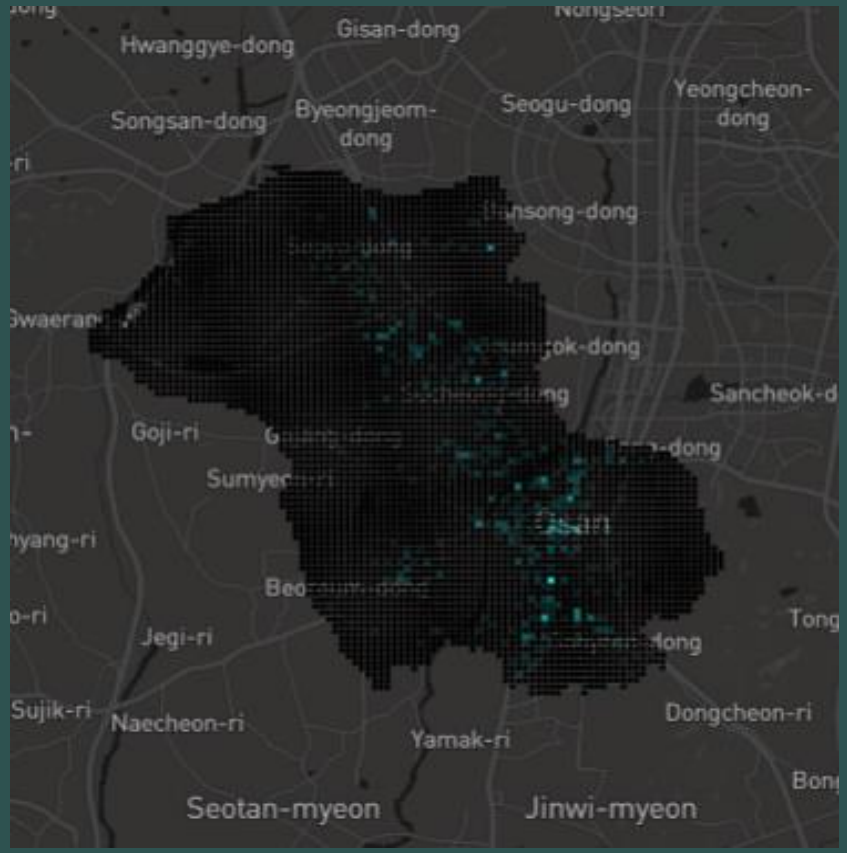
오산시 유소년인구 및 격자별 어린이교통사고 현황

<오산시 유소년인구 >



사용 Data: 5.오산시_연령별_거주인구격자(유소년).geojson

<100 X 100 격자별 어린이 교통사고 현황>



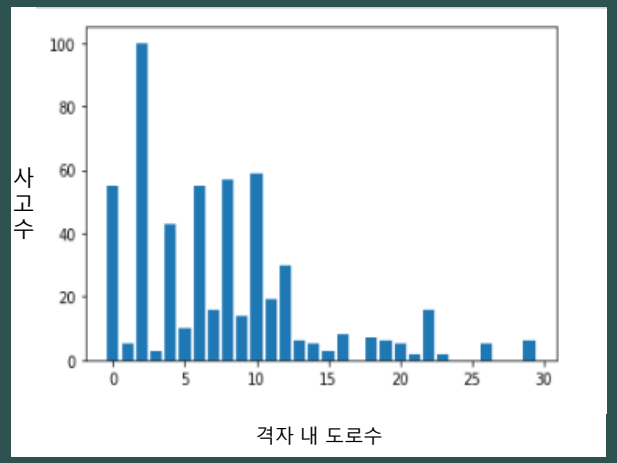
사용 Data:2.오산시_어린이교통사고_격자.geojson

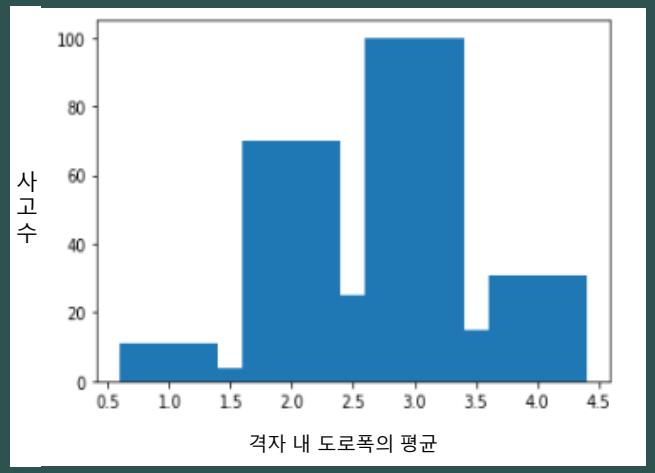
[고현동] [금암동] [내삼미동] 등등의 **고급아파트 단지에 유소년 인구가 특히 밀집**되어 있으며
교통사고 현황Plot이 이와 **비슷한 양상**을 보임 (유소년인구가 **밀집**될수록 **진한 청록색**을 띄고있음)

1-1. EDA

오산시 어린이교통사고 최다 발생지역 격자 _ 도로관련변수

	gid	accident_cnt	geometry
3681	다사621048	10	MULTIPOLYGON (((127.07325 37.14114, 127.07325 ...
3617	다사620042	7	MULTIPOLYGON (((127.07216 37.13573, 127.07215 ...
3828	다사624061	6	MULTIPOLYGON (((127.07657 37.15287, 127.07656 ...
3863	다사625056	6	MULTIPOLYGON (((127.07772 37.14837, 127.07771 ...
3098	다사612101	6	MULTIPOLYGON (((127.06285 37.18888, 127.06284 ...





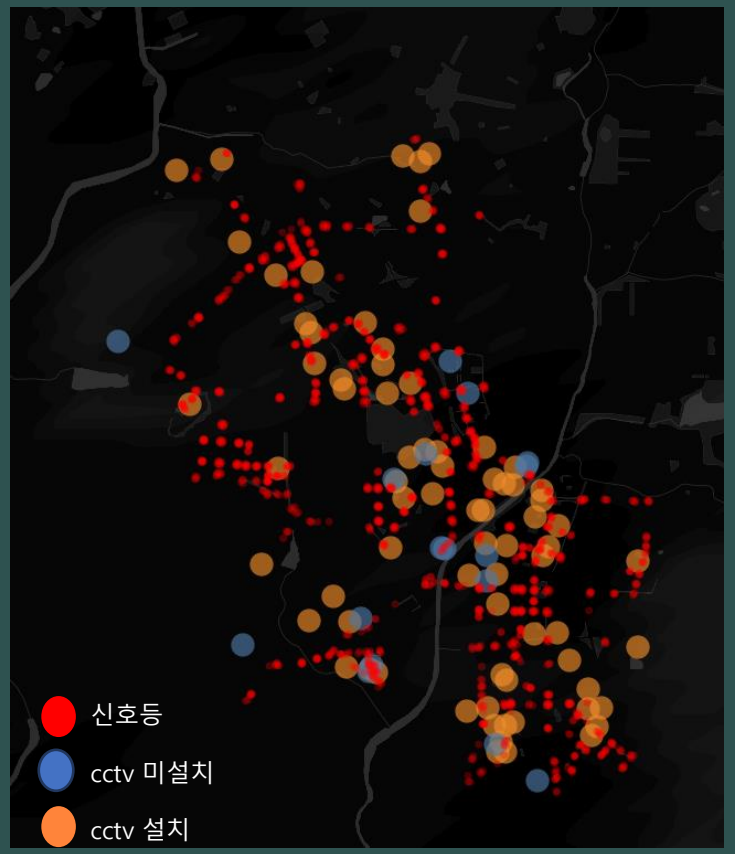
최다빈도 사고발생지역 격자 (100X100)에서의 어린이 교통사고 수는 **10번**이며, 아래그림에서 알 수 있듯이 여러 개의 서로 다른 도로가 만나는 지점에서 상당수의 사고가 발생하기 때문에 도로의 수, 넓이와 같은 **도로관련 변수**들의 영향 역시 확인할 필요성이 있음을 알 수 있음

사용 Data:2.오산시_어린이교통사고_격자.geojson

1-1. EDA

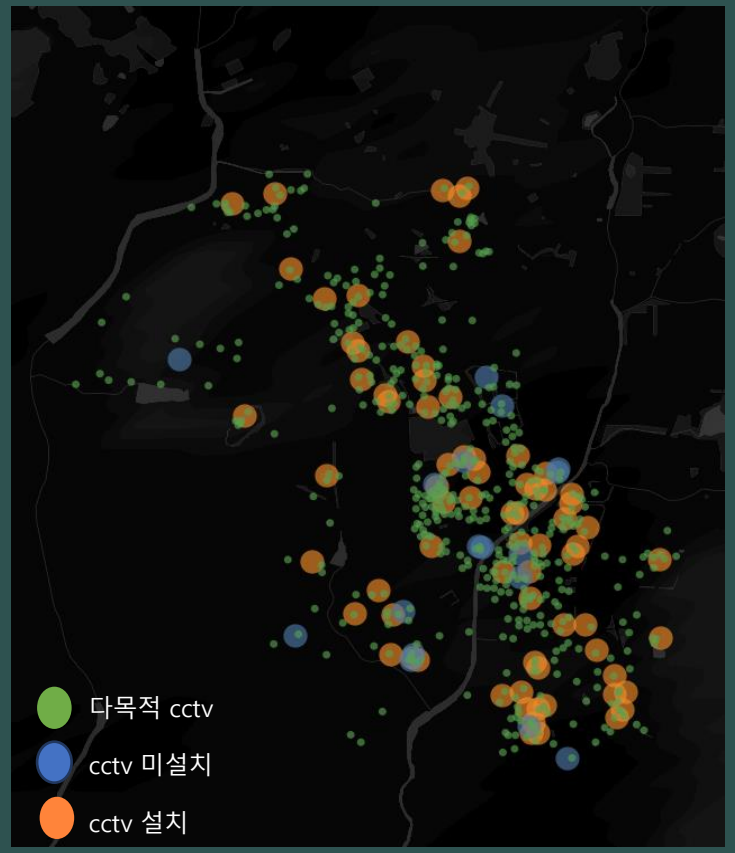
현 어린이보호구역 교통안전시설물 현황

<신호등 설치 현황 >



사용 Data: 19.오산시_신호등.geojson

<CCTV 설치현황>



사용 Data:20.오산시_CCTV설치현황.csv

신호등	어린이보호구역
0	79
1	3
2	3
3	2
4	2
5	0
6	2
7	0
8	2

어린이 보호구역 격자 내
신호등 설치율

15%

CCTV	어린이보호구역
0	41
1	39
2	12
3	1
4	0

어린이 보호구역 격자 내
CCTV 설치율

56%

1-1. EDA

주변 시설 변수 활용

데이터는 격자 별로 분류되어 있으나 교통사고의 발생은 '주변 교육시설로의 이동' 과 같은 이유로 주변 격자의 영향을 받을 수 밖에 없음.

주변 격자의 특성에 의한 영향에 받을 가능성을 고려한 변수가 필요

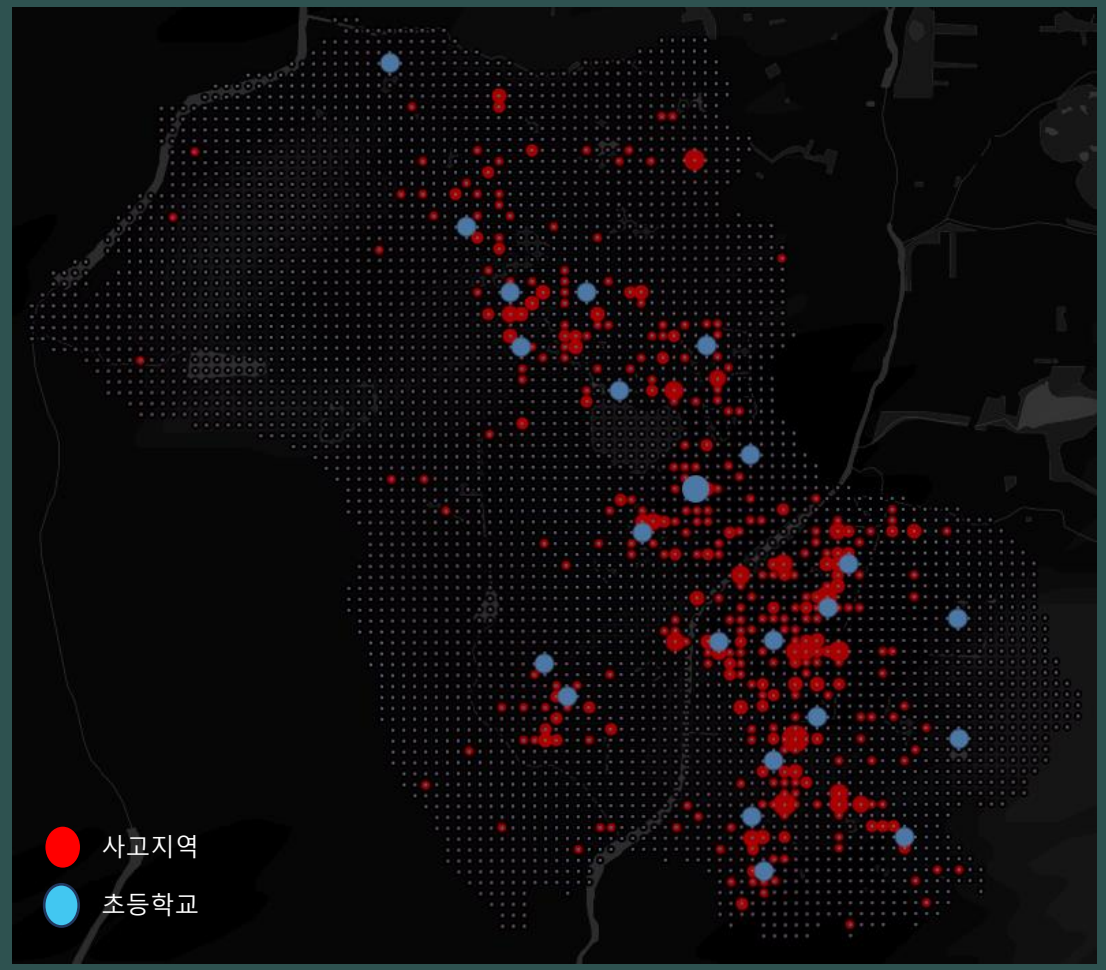
주변 8방향의 격자에 존재하는 건물,시설,교통정보 등을 주변 변수(_around)로 추가

사고정규화	주정차단속	총인구	유소년인구	생산가능인구	등교	하교	이외유동	어린이보호구역	초등학교	...	추정교통량_승용차_around	추정교통량_버스_around	추정교통량_화물차_around	도로평균제한속도_around	버스정류장_around	중앙분리대_around	상행차로_around	하행차로_around	일방통행_around
0.0	0	0.0	0.0	0.0	210.56	372.26	218.59	0	0	...	29336.184444	1276.695556	8550.655556	44.444444	0.222222	0.888889	0.888889	0.0	0.444444
0.0	0	0.0	0.0	0.0	214.15	345.98	194.78	0	0	...	21206.144444	919.818889	6183.433333	33.333333	0.222222	0.666667	0.666667	0.0	0.333333
0.0	0	0.0	0.0	0.0	160.35	275.74	163.82	0	0	...	4946.064444	206.065556	1448.988889	11.111111	0.000000	0.222222	0.222222	0.0	0.111111
0.0	0	0.0	0.0	0.0	9.08	13.84	8.20	0	0	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000
0.0	0	0.0	0.0	0.0	0.00	0.00	0.00	0	0	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000

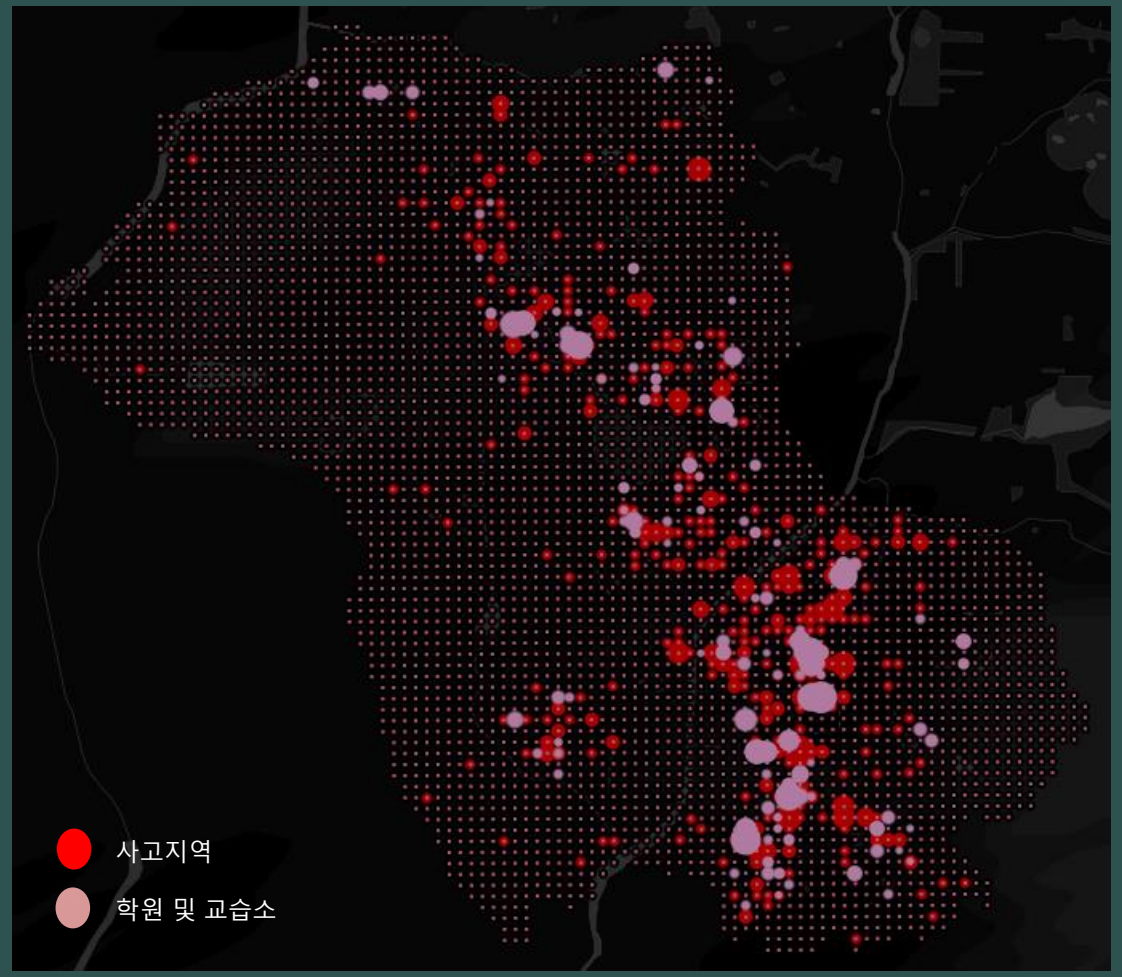
1-1. EDA

주변 시설 변수 활용

교육시설이 있는 격자 주변에 사고가 밀집되어 있는 모습으로 미루어 보아 **교육시설**이 **어린이 교통사고**와 유의미한 연관이 있을 것이며 교육시설이 위치한 격자 대비 **교육시설 인접 지역**에서의 사고 비율이 높다는 점에서 교육시설의 인접정도 역시 유의미한 연관이 있을 것임을 추측 가능



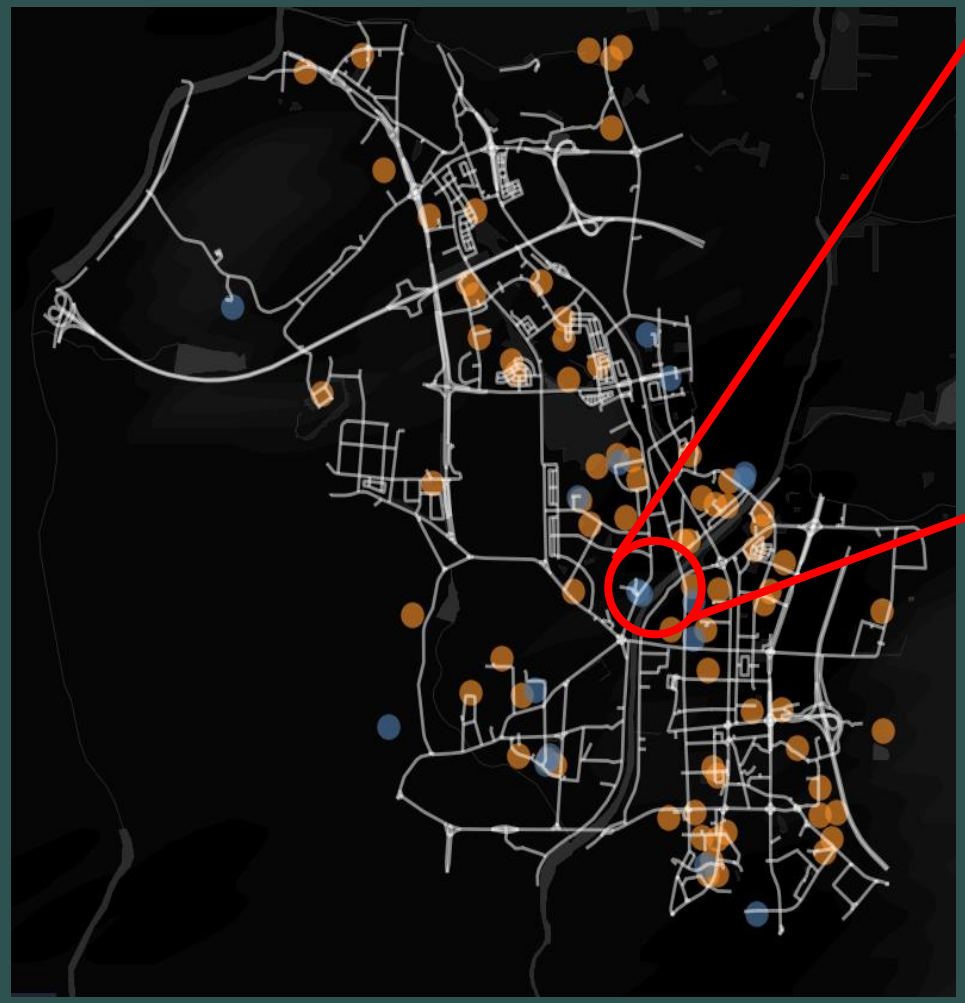
<사고 발생 대비 초등학교 위치>

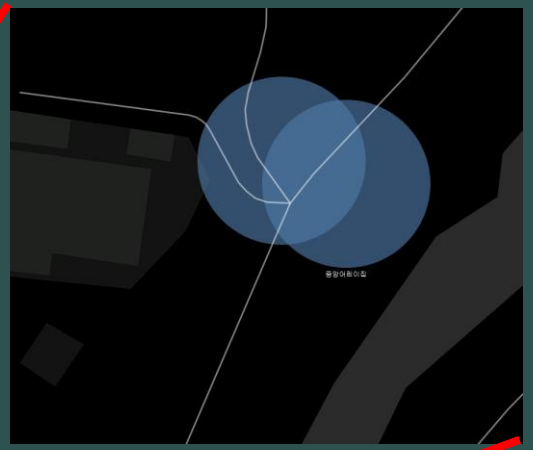


<사고 발생 대비 학원,교습소 위치>

1-1. EDA

오산시 어린이 교통사고 위험 지역 탐색





- 어린이 보호구역
- CCTV 설치 無
- 신호등 설치 無
- 도로수 多

위와 같은 지역의 탐지를 목적으로 분석을 시행.

1-2. 전처리

오산시 내 격자 (grid) 정보 생성

- 오산시 내 총 4510개의 100x100 격자마다 격자 내 정보가 반영된 데이터 프레임 생성

ex) 격자 별 지리 정보 + accident_cnt

	gid	accident_cnt	geometry	coordinates	coord_cent	geo_cent	grid_id	min_lon	max_lon	min_lat
0	다사 551085	0	MULTIPOLYGON (((126.99422 37.17418, 126.99421 ...	[[126.99421564681425, 37.17418235770403], [126...	[126.99477586013964, 37.174635450576154]	POINT (126.9947758601396 37.17463545057615)	00000	126.994210	126.995342	37.174182
1	다사 551086	0	MULTIPOLYGON (((126.99421 37.17508, 126.99420 ...	[[126.99420963816323, 37.17508373885349], [126...	[126.9947698580453, 37.17553683173449]	POINT (126.9947698580453 37.17553683173449)	00001	126.994204	126.995336	37.175084
2	다사 551087	0	MULTIPOLYGON (((126.99420 37.17599, 126.99420 ...	[[126.99420362924478, 37.17598511986466], [126...	[126.99476385568379, 37.17643821275453]	POINT (126.9947638556838 37.17643821275453)	00002	126.994198	126.995330	37.175985
3	다사 551088	0	MULTIPOLYGON (((126.99420 37.17689, 126.99419 ...	[[126.9941976200589, 37.17688650073755], [126....	[126.99475785305518, 37.17733959363628]	POINT (126.9947578530552 37.17733959363628)	00003	126.994192	126.995324	37.176887
4	다사 552085	0	MULTIPOLYGON (((126.99534 37.17419, 126.99534 ...	[[126.99534207541716, 37.174187160436155], [12...	[126.9959022955662, 37.174640248032546]	POINT (126.9959022955662 37.17464024803255)	00004	126.995336	126.996469	37.174187

###_cent

: 격자 별 centroid

(coord = float type)

(geo = geometry type)

accident_cnt

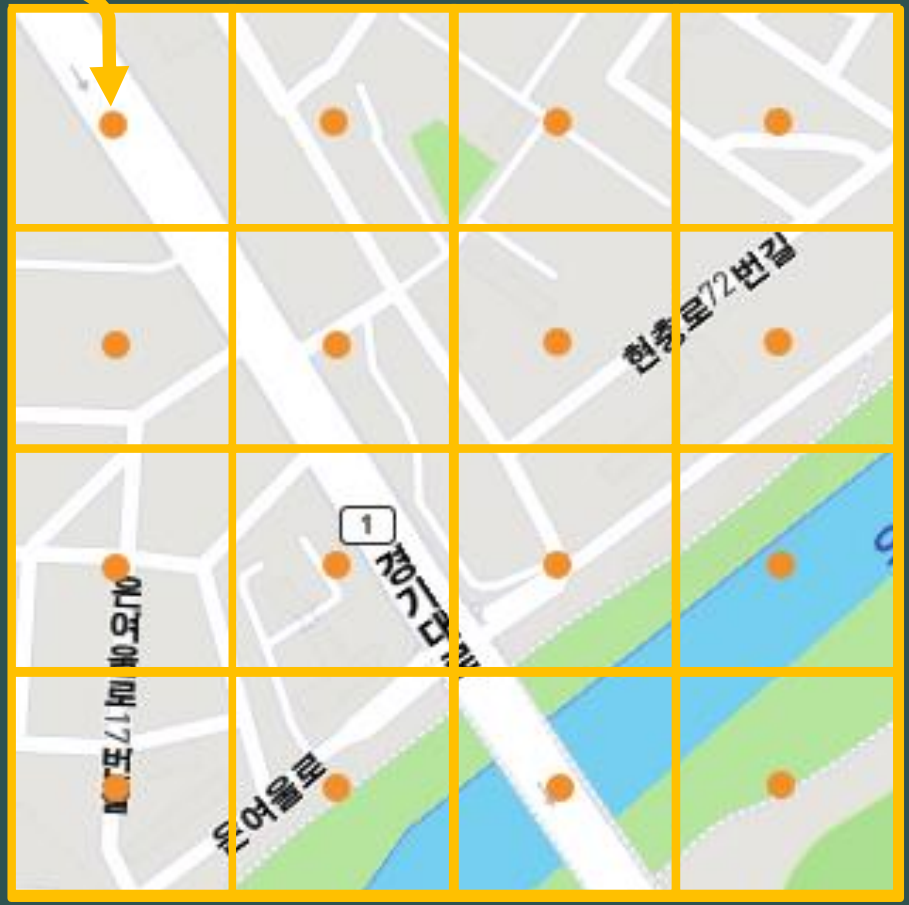
: 격자 별 어린이 교통사고 발생 횟수

1-2. 전처리

오산시 내 격자 (grid) 정보 생성

격자 내 정보 할당

- 100m X 100m 격자마다 인구, 도로, 유동성, 시설물 등 지역 특성 할당
- 격자마다 centroid point 계산하여 이후 교통사고 위험지역 및 시설물 우선 설치 지역 도출 시 사용
- Min/Max Lon, Min/Max Lat 변수는 격자 별 최대/소 위경도를 반영



1-2. 전처리

지역 특성 변수 생성

안전 시설물

- 무인 교통 단속
- 도로 안전 표지판
- 과속 방지턱
- 신호등
- CCTV

인구

- 총인구
- 생산가능인구
- 유소년인구
- 유동인구 (2019년)
  - 등교 시간 (08시-10시)
  - 하교 시간 (14시-18시)
  - 기타 유동 (18시-22시)

기타 위치 정보

- 초등학교
- 어린이집 / 유치원
- 학원 / 교습소
- 버스 정류장

1-2. 전처리

지역 특성 변수 생성

도로 & 교통량

격자 당 종합

- 도로수
- 중앙분리대
- 상/하행 차로
- 일방통행 도로
- 추정 교통량

(전체, 승용차, 버스, 화물차)

격자 내 도로들의 특성 합 / 도로 수

- 격자 내 도로 평균 특성

(제한속도, 길이, 폭, 혼잡빈도강도)

인접 지역 정보

- 격자마다 인접 지역(격자)의 정보를 변수로 추가 (###_around)

(등,하교 등으로 인한 인접 지역으로의 유동성 및 도로망, 지역 특성 간접 파악 가능)

1-2. 전처리

최종 Input data 생성

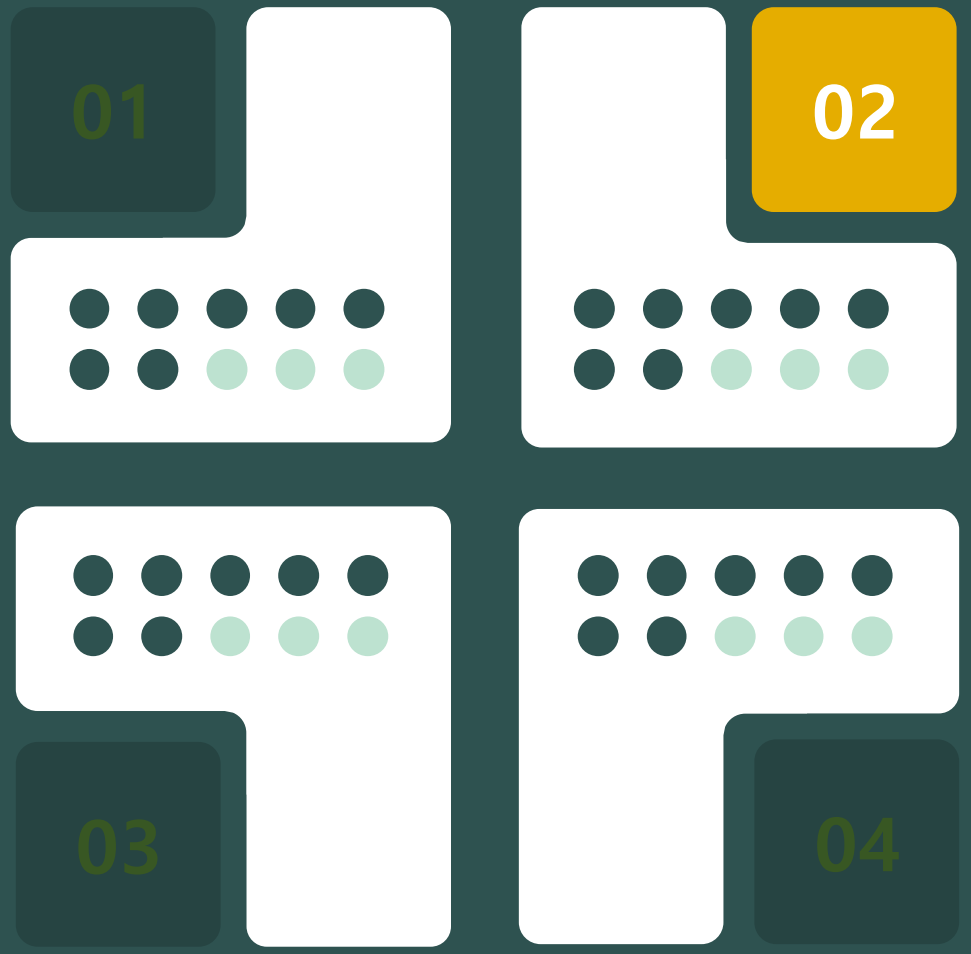
- 격자 별 지리 정보와 어린이 교통사고 횟수, 추가적으로 생성한 설명 변수들을 결합
- 교통사고 횟수를 제외한 타 변수들은 **변수 별 최대값으로 나누어 정규화** 진행
- 추후 분석을 위한 **최종 데이터(final_df)** 완성

ex)

```
final_df.columns
```

```
Index(['gid', 'accident_cnt', 'geometry', 'coordinates', 'coord_cent',
       'geo_cent', 'grid_id', 'min_lon', 'max_lon', 'min_lat', 'max_lat',
       '사고정규화', '주정차단속', '총인구', '유소년인구', '생산가능인구', '등교', '하교', '이외유동',
       '어린이보호구역', '초등학교', '인접초등학교통학구', '어린이집_유치원', '학원_교습소', '무인교통단속',
       '도로안전표지판', '횡단보도', '과속방지턱', '신호등', 'CCTV', '버스정류장', '추정교통량_전체',
       '추정교통량_승용차', '추정교통량_버스', '추정교통량_화물차', '혼잡빈도강도', '도로수', '도로평균제한속도',
       '도로길이', '도로폭', '중앙분리대', '일방통행', '상행차로', '하행차로', '주정차단속_around',
       '총인구_around', '유소년인구_around', '생산가능인구_around', '등교_around', '하교_around',
       '이외유동_around', '어린이보호구역_around', '초등학교_around', '인접초등학교통학구_around',
       '어린이집_유치원_around', '학원_교습소_around', '무인교통단속_around', '도로안전표지판_around',
       '횡단보도_around', '과속방지턱_around', '신호등_around', 'CCTV_around',
       '혼잡빈도강도_around', '도로수_around', '도로길이_around', '도로폭_around',
       '추정교통량_전체_around', '추정교통량_승용차_around', '추정교통량_버스_around',
       '추정교통량_화물차_around', '도로평균제한속도_around', '버스정류장_around', '중앙분리대_around',
       '상행차로_around', '하행차로_around', '일방통행_around'],
      dtype='object')
```

2. 어린이 교통사고 위험구역 제시



2-1. 다중공선성 문제 해결

변수간 상관성에 기반한 1차 변수 선정

1) 주변 정보 변수 중 격자 내 지역 특성과 상관성이 높은 변수 제거

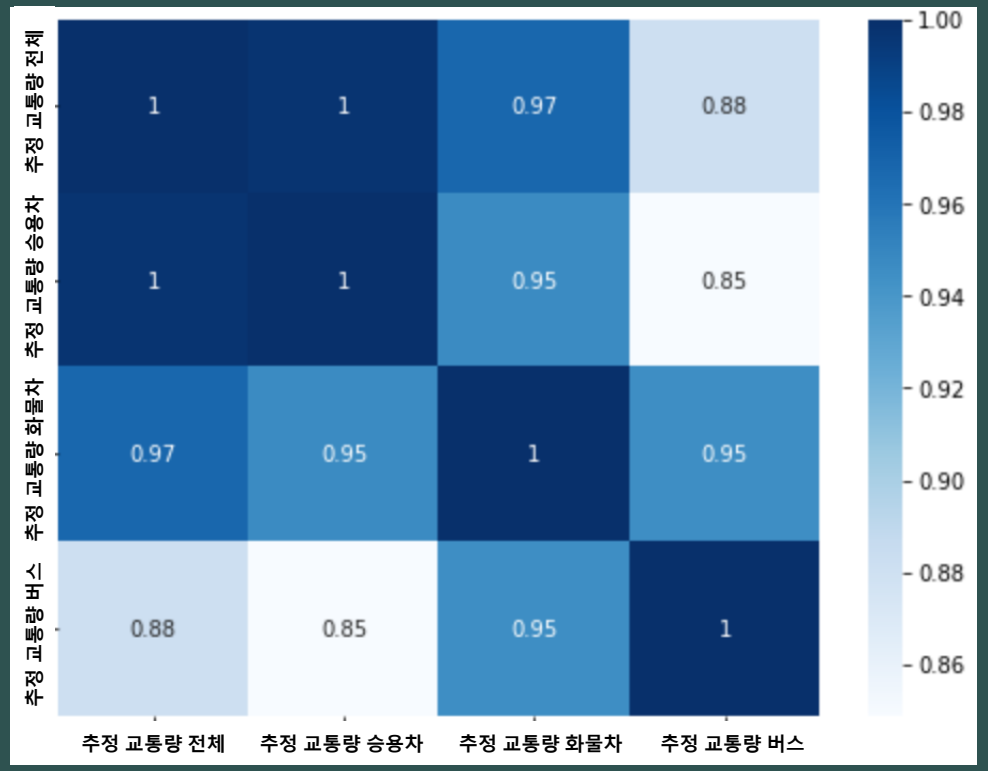
2) VIF (variance inflation factor) 값이 10 이상인 변수들 확인. 해당 변수 중 correlation이 0.7 이상이 변수 쌍 출력

3) 2)에서 확인한 변수 쌍들을 근거로 correlation plot 시각화. 모형 적합에 사용할 후보 변수 1차 선정

ex)

- 추정교통량_전체, 총인구, 하교, 도로폭, 도로폭_around 변수만을 사용하고 이외 상관성 높은 변수들 삭제

```
X = X.drop(['추정교통량_승용차','추정교통량_화물차', '추정교통량_버스',
            '유소년인구', '생산가능인구', '등교', '이외유동',
            '중앙분리대', '상행차로', '하행차로',
            '중앙분리대_around', '상행차로_around', ], axis = 1)
```



<VIF 10 이상 변수 중 사용 변수>

- **인구) 총인구**
- **교통량) 추정교통량_전체**
- **도로) 도로폭, 도로폭_around**
- **유동인구) 하교(14-18)**

ref) 경기연구원 『오산시 어린이보호구역 보행안전 확보방안 연구』(2020.09)를 근거로 변수 선정

2-2. 최종 변수 선택

변수 선택법을 통한 2차 변수 선정

1) 1차 변수 선정 후, VIF 값 상위 변수 출력하여 다중공선성 문제가 해결되었음을 확인 (모든 VIF 값이 10 이하)

2) 음이항 회귀 모형을 이용한 전진선택법 및 후진제거법 진행

(목적변수: 격자 별 어린이 교통사고 발생횟수(accident_cnt) / 설명변수: 격자 별 지역 특성 반영 변수 (시설물 관련 변수 제외))

(모형 적합 데이터는 어린이보호구역인 곳 & 도로가 존재하는 곳으로 설정 – shape: (1678, 39))

	VIF Factor	features
0	41.50	Intercept
32	3.39	도로수_around
28	3.09	횡단보도_around
15	2.90	추정교통량_전체
33	2.73	도로폭_around
17	2.73	도로수
30	2.61	신호등_around
10	2.27	횡단보도
18	2.17	도로평균제한속도
12	2.16	신호등

Ex. 다중공선성 문제가 해결됨

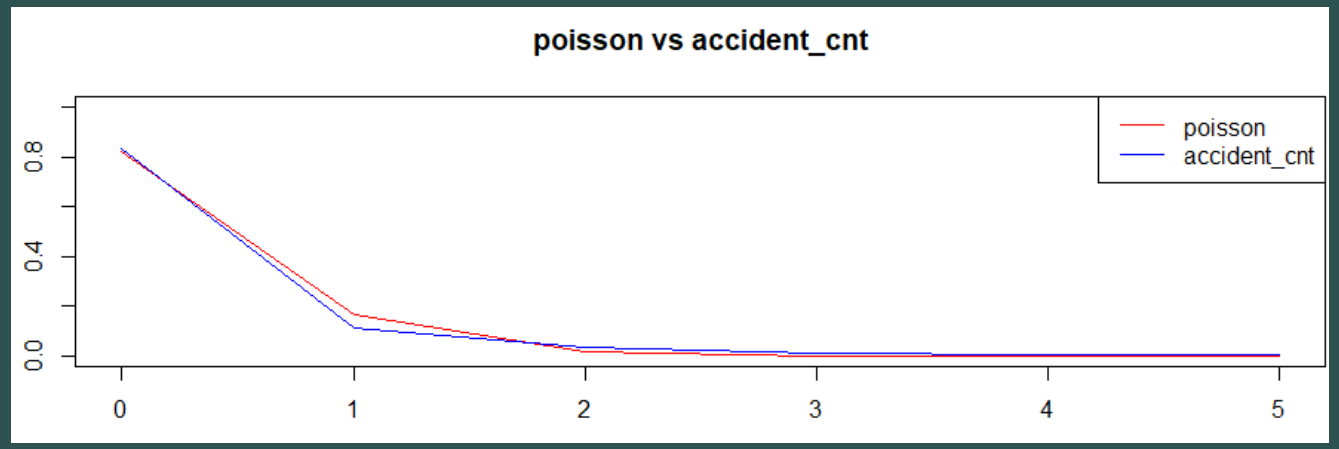
<어린이 교통사고 발생 횟수와 유의미한 관계를 보이는 지역특성 변수>

추정교통량_전체, 하교, 혼잡빈도강도,
버스정류장, 횡단보도, 초등학교, 학원_교습소,
도로수, 도로평균제한속도, 일방통행,
학원_교습소_around, 초등학교_around

- EDA를 통해 도출된 '인접 지역의 지역 특성이 중요'하다는 추론을 증명
- 도로 특성과 교통량 정도에 관련된 변수가 多 -> 여러 도로가 모이는 지점에서 사고가 다발 하는 등 EDA 과정에서 시각화를 통해 확인한 부분이 실제로 도로 위험도에 영향을 준다는 것을 확인

2-3. 음이항 회귀 모형 적합

주요 변수 coefficient 추출 및 결과 확인



Ex. 포아송 분포와 오산시 어린이 교통사고 발생횟수 분포가 근사하는 것을 확인

[음이항 회귀 모형 사용 이유]

- 일정 영역 내에서, 발생 확률이 낮은 사건의 분포는 일반적으로 포아송 분포를 따름. (표본 개수 N이 큰 경우)
- 하지만 포아송 분포는 평균과 분산이 같다는 가정(E(y) = μ, Var(y) = μ)이 필요한데, 실제 데이터의 분포가 해당 가정을 만족하는 경우는 드묾. (평균보다 분산이 더 큰 과대 산포(Overdispersion)의 문제가 발생)
- 때문에, 평균보다 분산이 크다고 가정하는 (E(y) = μ , Var(y) = $\mu + D\mu^2$) 음이항 분포를 고려
- 교통사고 분석 시, 포아송 회귀보다 음이항 회귀 모형이 더 적합하다는 연구결과가 있다.

ref) 정재풍, 최종후. (2014). 교통사고건수에 대한 포아송 회귀와 음이항 회귀모형 적합. , 16(1), 165-172.

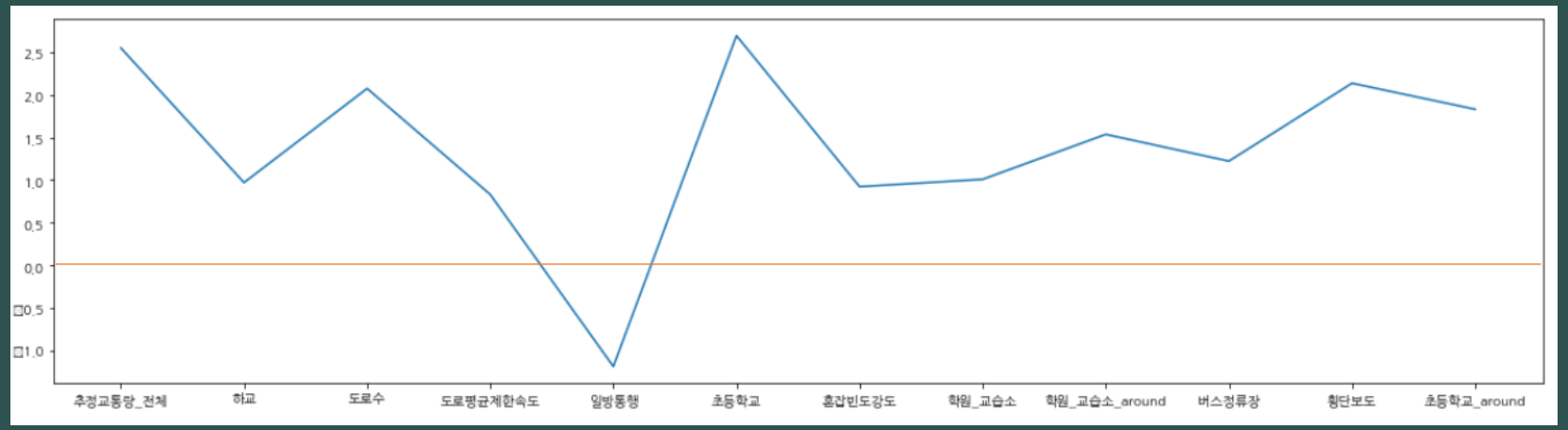
2-3. 음이항 회귀 모형 적합

주요 변수 coefficient 추출 및 결과 확인

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-3.5607	0.239	-14.920	0.000	-4.028	-3.093
추정교통량_전체	2.5475	0.677	3.765	0.000	1.221	3.873
하교	0.9708	0.384	2.529	0.011	0.218	1.723
도로수	2.0730	0.517	4.011	0.000	1.060	3.086
도로평균제한속도	0.8313	0.346	2.402	0.016	0.153	1.510
일방통행	-1.1835	0.318	-3.727	0.000	-1.806	-0.561
초등학교	2.6918	0.965	2.790	0.005	0.801	4.582
혼잡빈도강도	0.9229	0.245	3.765	0.000	0.442	1.403
학원_교습소	1.0090	0.500	2.017	0.044	0.028	1.990
학원_교습소_around	1.5357	0.378	4.058	0.000	0.794	2.277
버스정류장	1.2223	0.338	3.613	0.000	0.559	1.886
횡단보도	2.1346	0.341	6.254	0.000	1.466	2.804
초등학교_around	1.8297	0.335	5.460	0.000	1.173	2.486

〉 음이항 회귀 모형 적합 결과

〉 음이항 회귀 모형 적합 결과



- 모형 적합에 사용된 모든 지역 특성 변수의 p-value가 0.05 이하로 **유의**하다는 것을 확인할 수 있음.
- EDA에서 추측했던 결과와 같이 격자 내 혹은 격자 주위에 **초등학교나 학원**이 있을 때 **교통사고 위험지수**는 **증가**

  (교통사고의 타겟이 될 수 있는 어린이의 절대적 수치가 증가하기 때문)
- 반대로 **격자 내 도로가 일방통행**일 경우 **교통사고 위험지수**는 **감소**

  (어린이들이 주의 해야 할 '차가 오는 방향' 이 한 곳으로 제한되기 때문)

2-4. MCLP

교통사고 위험지수 계산 및 MCLP 적용

지역 특성 변수	영향도
추정교통량_전체	2.5475
하교	0.9708
도로수	2.0730
도로평균제한속도	0.8313
일방통행	-1.1835
초등학교	2.6918
혼잡빈도강도	0.9229
학원_교습소	1.0090
학원_교습소_around	1.5357
버스정류장	1.2223
횡단보도	2.1346
초등학교_around	1.8297

×

=

w_hazard
교통사고 위험지수

→

MCLP 적용

[교통사고 위험지수]

- 정규화 된 지역 특성 변수 값 X 모형 coefficient

[Maximal Covering Location Problem]

- 각 격자의 교통사고 위험지수를 이용하여 일정 범위 내의 지수 총합이 가장 커지는 최적 위치 (격자 centroid)를 결정

2-4. MCLP

교통사고 위험지수 계산 및 MCLP 적용

[MCLP?]

제한된 시설의 개수로 각 수요지점의 수요량을 통하여 가장 많은 수요를 커버할 수 있도록 시설의 적정 위치를 결정하는 문제

[Formulation]

$$
\begin{array}{llll}
I.\ Maximize & z = \sum_{i \in I} a_i y_i & & \cdots ① \\
S.T. & \sum_{i \in N_i} x_i \geq y_i & for\ all\ i\ \in I & \cdots ② \\
 & \sum_{j \in J} x_i = P & & \cdots ③ \\
 & x_j = (0,1) & for\ all\ j\ \in J & \cdots ④ \\
 & y_i = (0,1) & for\ all\ i\ \in I & \cdots ⑤
\end{array}
$$

[Description]

① : 목적함수로 커버리지 거리 내에서 수요지수를 최대한 커버하도록 설정
② : 시설물의 커버리지가 수요지점 i를 포함하는지 여부
③ : 제한된 시설물의 총합
④ : 커버리지가 수요지점을 포함하는지 여부
⑤ : 시설물이 j지역에 설치되었는지 여부

[Notation]

I = 수요지점 포인트
J = 시설 후보 포인트 셋
i = 수요지점 포인트 index
j = 시설 후보 포인트 인덱스
xj = 후보 포인트에 시설이 선정되면 1, 아니면 0
S = 수요지를 커버할 수 있는 거리
Ni = 사고지수를 커버할 수 있는 시설 후보지 집합
ai = i 포인트의 수요지수
P = 설치 예정인 시설물 개수

2-4. MCLP

교통사고 위험지수 계산 및 MCLP 적용

[어린이 보호구역]

Input

I = 모든 격자
J = 위험지수가 높은 격자
S = 100m
a = 위험지수(by 음이항회귀)
P = 50개

→ **MCLP 적용**

OutPut : 어린이교통사고 위험지역

	lon	lat	w_hazard
0	127.072607	37.156012	7.895149
1	127.071628	37.129868	4.961979
2	127.047654	37.185666	3.397441
	⋮		
17	127.029567	37.195511	0.573622
18	127.062508	37.149665	0.479079
19	127.083987	37.134419	0.435627

[시설물] (무인 교통단속, 도로안전표지판, 과속방지턱, 신호등, CCTV, 중앙분리대)

Input

I = 모든 격자
J = 시설물 필요지수가 높은 격자
S = 100m
a = 시설물 필요지수(by OLS)
P = 50개

→ **MCLP 적용**

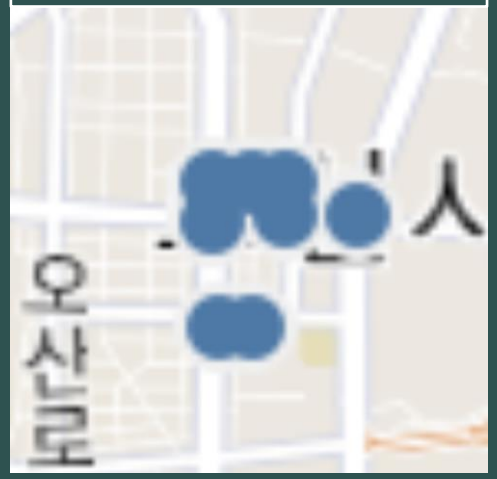
OutPut : 시설물 적정 입지

	lon	lat	w_facility
1	127.072703	37.138886	3.714985
19	127.071618	37.131670	1.218383
6	127.079364	37.156036	1.210674
	⋮		
9	127.048861	37.172149	0.100721
3	127.059979	37.197431	0.081990
13	127.039779	37.183833	0.075056

2-5. 최종 위험구역 도출

K-Means 알고리즘을 적용한 최종 결과 도출

최초 위험구역 도출 결과



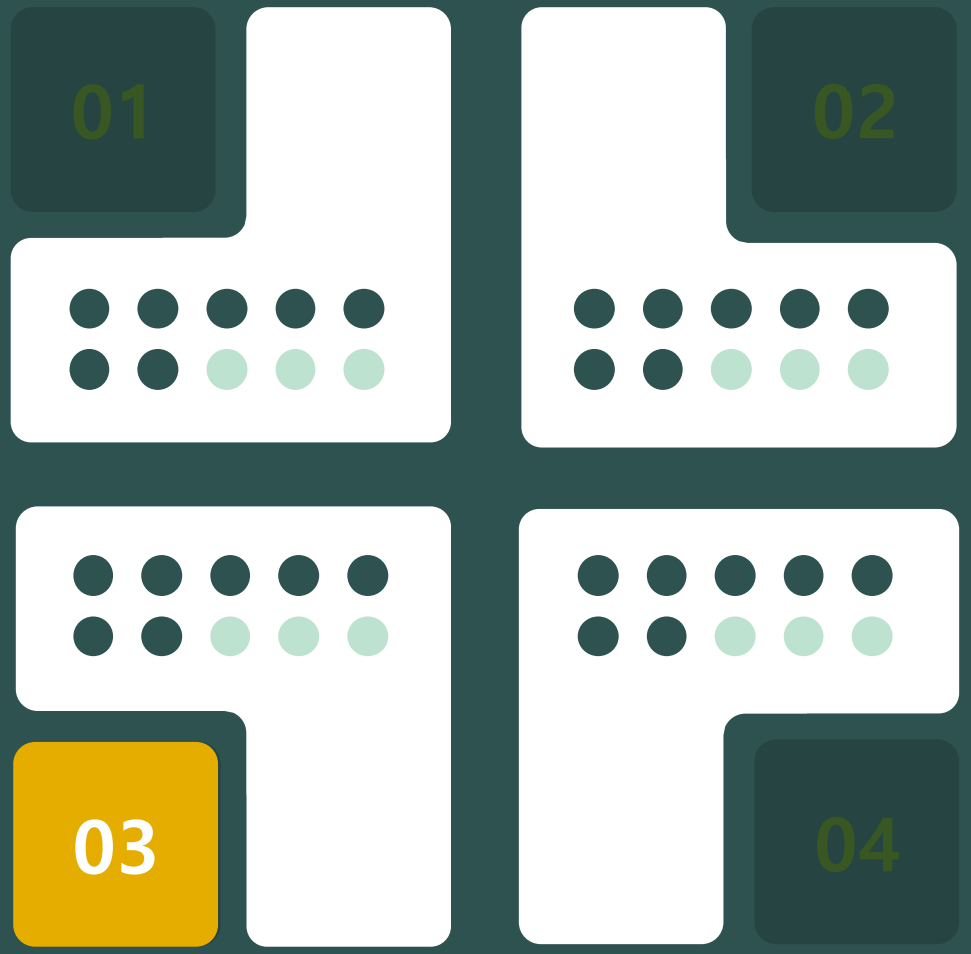
K-Means 적용

최종 위험구역 결과

	lon	lat	w_hazard	cluster
0	127.072607	37.156012	7.895149	13
1	127.071628	37.129868	4.961979	5
2	127.047654	37.185666	3.397441	17
3	127.059009	37.170385	3.090950	12
4	127.063561	37.162289	2.886692	4
5	127.073769	37.149706	2.869737	18
6	127.076041	37.146109	2.710286	1
7	127.051098	37.174862	2.453533	15
8	127.040041	37.140566	2.281091	6
9	127.072719	37.136181	2.270667	19
10	127.082728	37.158752	1.859151	7
11	127.090655	37.150667	1.387861	9
12	127.054562	37.160453	1.271679	2
13	127.045450	37.177545	1.147269	0
14	127.070436	37.141582	1.092610	16
15	127.040884	37.187443	0.737862	3
16	127.033152	37.162173	0.712795	14
17	127.029567	37.195511	0.573622	10
18	127.062508	37.149665	0.479079	11
19	127.083987	37.134419	0.435627	8

- MCLP를 통한 위험구역 최초 선정 시, 인접 위치의 여러 구역이 위험지역으로 나타남.
- 1차적으로 50개의 위험구역 선정 후, K-Means를 적용하여 위/경도 기준 20개의 클러스터 생성
- 각 클러스터 별로 어린이 교통사고 위험지수가 최대인 지점을 2차 선정하여 최종 결과 도출

3. 교통안전시설물 우선 설치 지역 제시



3-1. 변수 선택

변수선택법을 통한 시설물 별 모형 변수 선택

1) 각 교통안전시설물 (무인교통단속기, 도로안전표지판, 과속방지턱, 신호등, CCTV)에 대해, 그 외 지역 특성 변수를 후보 설명변수로 설정한 변수 선택 과정 진행 (전진선택법과 후진제거법 진행 후 중복되는 변수를 모형의 최종 설명변수로 설정)

2) 시설물 별로 결정된 최종 설명변수를 사용하여 다섯 번의 다중 회귀 모형 적합 진행

```
                                 OLS Regression Results
==========================================================================================
Dep. Variable:                    신호등   R-squared (uncentered):                  0.737
Model:                            OLS   Adj. R-squared (uncentered):              0.725
Method:                 Least Squares   F-statistic:                              58.91
Date:                Thu, 28 Jan 2021   Prob (F-statistic):                    1.34e-23
Time:                        16:58:05   Log-Likelihood:                          140.51
No. Observations:                  88   AIC:                                     -273.0
Df Residuals:                      84   BIC:                                     -263.1
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
도로평균제한속도      -0.0553      0.025     -2.255      0.027      -0.104      -0.007
횡단보도           0.4673      0.044     10.653      0.000       0.380       0.555
신호등_around      0.0816      0.025      3.219      0.002       0.031       0.132
추정교통량_전체       0.2643      0.113      2.347      0.021       0.040       0.488
==============================================================================
Omnibus:                       30.303   Durbin-Watson:                   2.471
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              122.735
Skew:                           0.948   Prob(JB):                     2.23e-27
Kurtosis:                       8.466   Cond. No.                         7.49
==============================================================================
```

Ex) '신호등'을 종속변수로 설정한 다중회귀모형 적합 결과

- '무인교통단속기'를 제외한 나머지 시설물에 대한 모형은 R-squared 값이 모두 약 0.6 이상으로 높은 적합도 보임.
- '무인교통단속기'는 격자 내/외의 지역적 특성으로부터 크게 영향을 받지 않는 것으로 판단. 시설물 필요지수 산출 과정에서 제외함.

3-2. 시설물 설치 필요지수 계산

시설물 별 설치 필요지수를 통해 격자마다의 최종 지수 계산

	w_신호등	w_과속방지턱	w_신호등	w_CCTV	w_facility
0	0.032991	0.0	0.032991	0.000000	0.167009
1	0.007840	0.0	0.007840	0.088632	0.303529
2	0.000000	0.0	0.000000	0.088632	0.311368
3	0.008248	0.0	0.008248	0.000000	0.241752
4	0.000000	0.0	0.000000	0.000000	0.200000
5	0.000000	0.0	0.000000	0.000000	0.200000
6	0.000000	0.0	0.000000	0.000000	0.200000
7	0.000000	0.0	0.000000	0.000000	0.200000
8	0.000000	0.0	0.000000	0.000000	0.250000
9	0.000000	0.0	0.000000	0.000000	0.200000
10	0.000000	0.0	0.000000	0.000000	0.250000
11	0.000000	0.0	0.000000	0.088632	0.161368
12	0.000000	0.0	0.000000	0.000000	0.400000
13	0.000000	0.0	0.000000	0.000000	0.800000
14	0.011122	0.0	0.011122	0.000000	0.438878

[시설물 설치 필요지수 (w_facility) 계산]

- 어린이 교통사고 위험지수 계산 과정과 동일하게 각 교통안전시설물의 설치 필요지수 (w_신호등, w_과속방지턱, w_도로안전표지판, w_CCTV,) 계산

- 시설물 별 설치 필요지수를 합하여 격자 별 시설물 설치 필요지수 산출

- 시설물 설치 필요지수 – 실제 시설물 개수로 최종 설치 필요지수 산출

3-3. MCLP & Kmeans

MCLP와 K-Means 적용. 시설물 우선 설치지역 도출

최종 우선설치구역 결과

lon	lat	w_facility	cluster
127.072703	37.138886	3.714985	16
127.071618	37.131670	1.218383	2
127.079364	37.156036	1.210674	7
127.063556	37.163190	0.556127	8
127.069177	37.165014	0.530492	9
127.044323	37.177541	0.486338	14
127.083997	37.132616	0.421752	19
127.052203	37.178472	0.377934	1
127.051268	37.146018	0.306624	3
127.031809	37.197323	0.265624	11
127.071511	37.150600	0.216074	0
127.053558	37.139716	0.206835	13
127.064615	37.174913	0.205174	18
127.055730	37.153246	0.140632	17
127.090650	37.151569	0.132413	15
127.079429	37.144318	0.107995	6
127.027476	37.169362	0.101540	10
127.048861	37.172149	0.100721	12
127.059979	37.197431	0.081990	5
127.039779	37.183833	0.075056	4

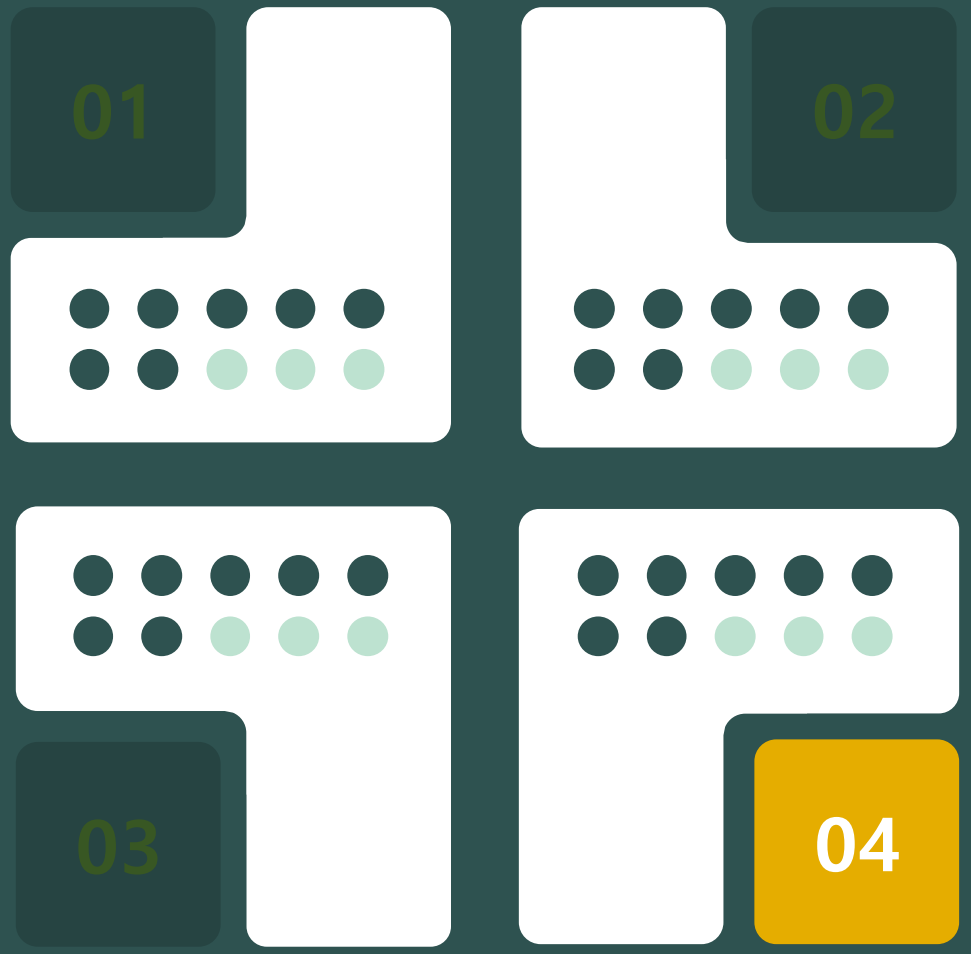
[MCLP 적용]

- 격자 별 설치 필요지수를 이용하여 MCLP 적용. 시설물 우선 설치지역 50곳을 1차적으로 선정

[K-Means 적용]

- K-Means를 통해 위/경도 기준 20개의 클러스터 생성하고, 각 클러스터 별로 설치 필요지수가 최대인 지점을 2차 선정하여 최종 우선 설치 지역 도출
- 유소년인구대비 사고수가 많은 구역에도 위험구역이 배치되게 함으로써 사고가 많이 일어나는 구역에 위험구역이 몰리지 않게 함

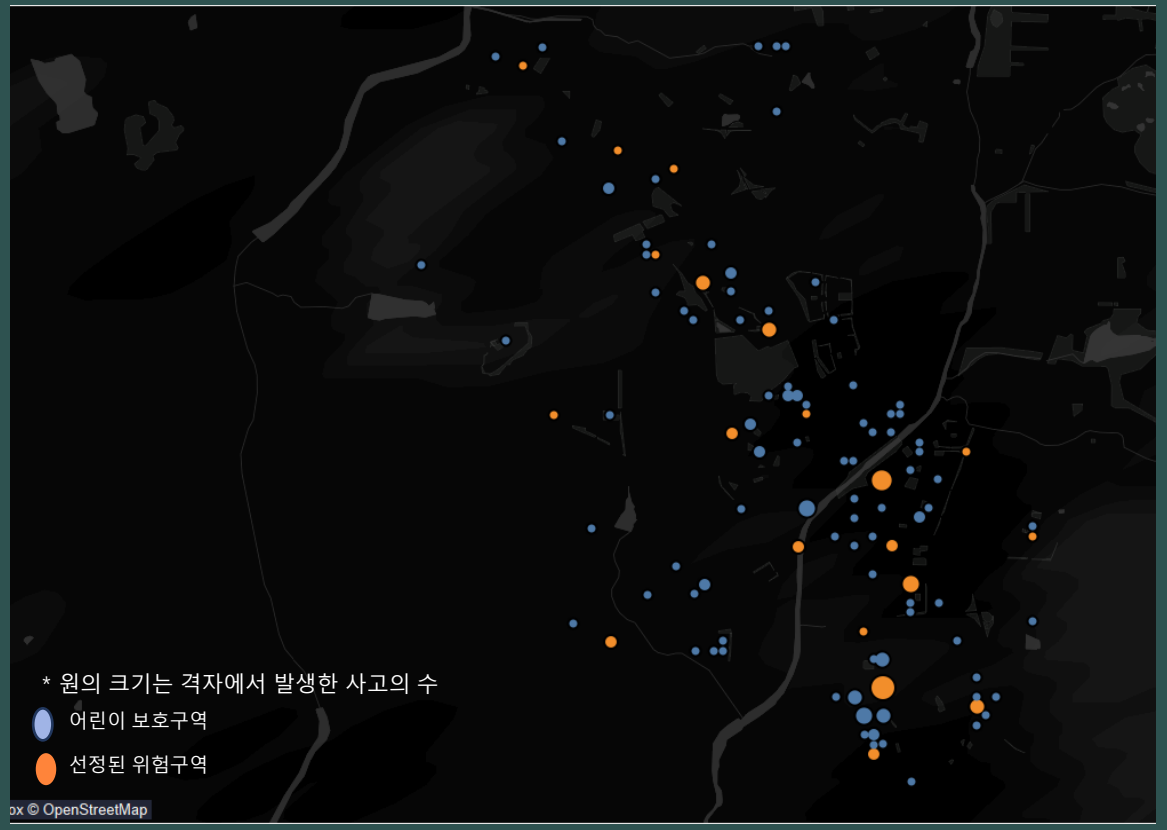
4. 최종 결과 및 평가



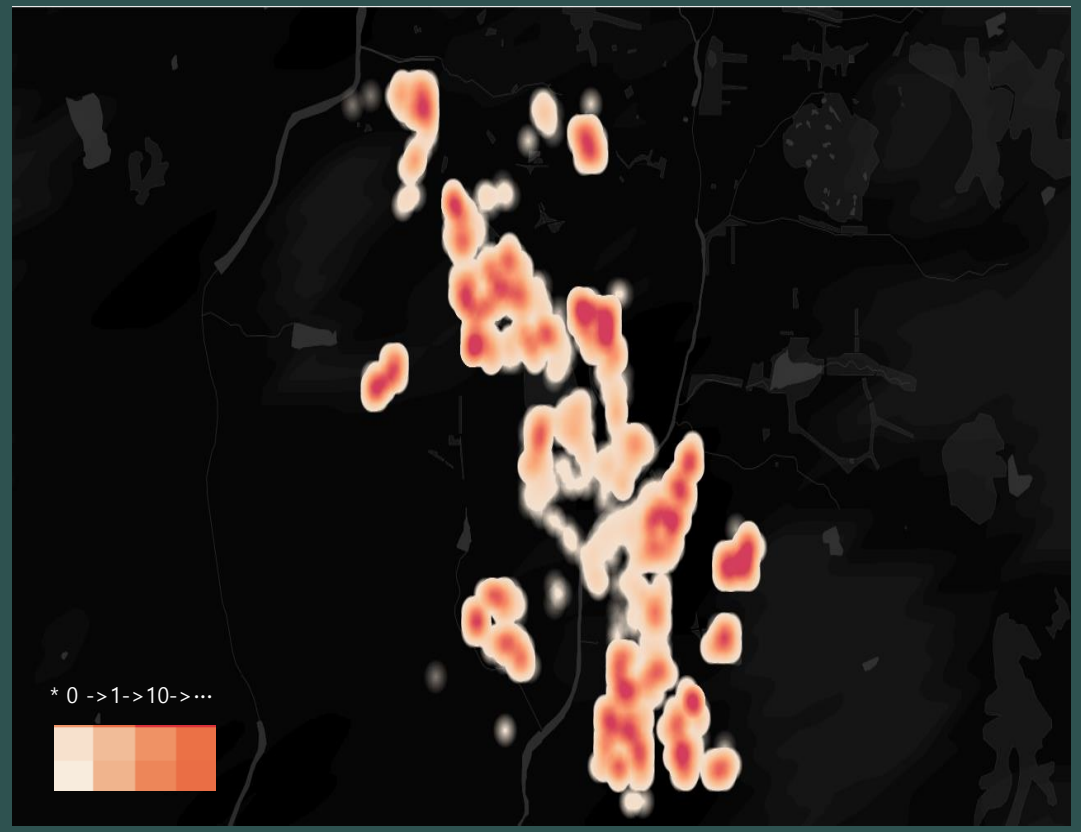
4-1. 어린이 교통사고 위험구역 확인

어린이 보호구역 외 어린이 교통사고 위험지역 20개소 제시

<어린이 교통사고 위험구역 20개소 제시>



<유소년 인구 Heatmap>



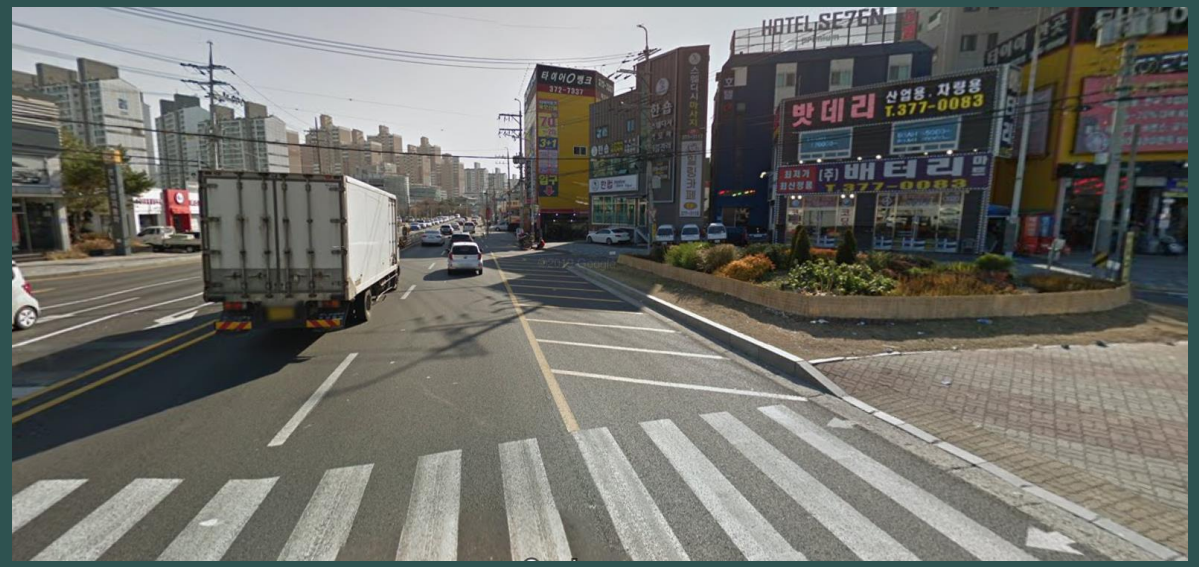
- 선정된 위험구역은 **실제 사고 수**에서 어린이 보호구역과 동등 혹은 그 이상의 수치를 보여줌
- **유소년인구 Plot**과 비교해 봄으로써 K-means의 과정을 통해 위험구역 몰림 현상을 방지했음을 확인
- 위험구역 선정 과정 및 결과의 실효성을 증명

4-1. 어린이 교통사고 위험구역 확인

어린이 보호구역 외 어린이 교통사고 위험지역 20개소 제시

위험순위	시설명/주소지	중심점 위치		반경범위
		X좌표(경도)	Y좌표(위도)	
1	희망주택아파트 일대	127.0726	37.15601	100m
~	~	~	~	~
20	오산빌리지 일대	127.084	37.13442	100m

예시) 희망주택아파트 일대



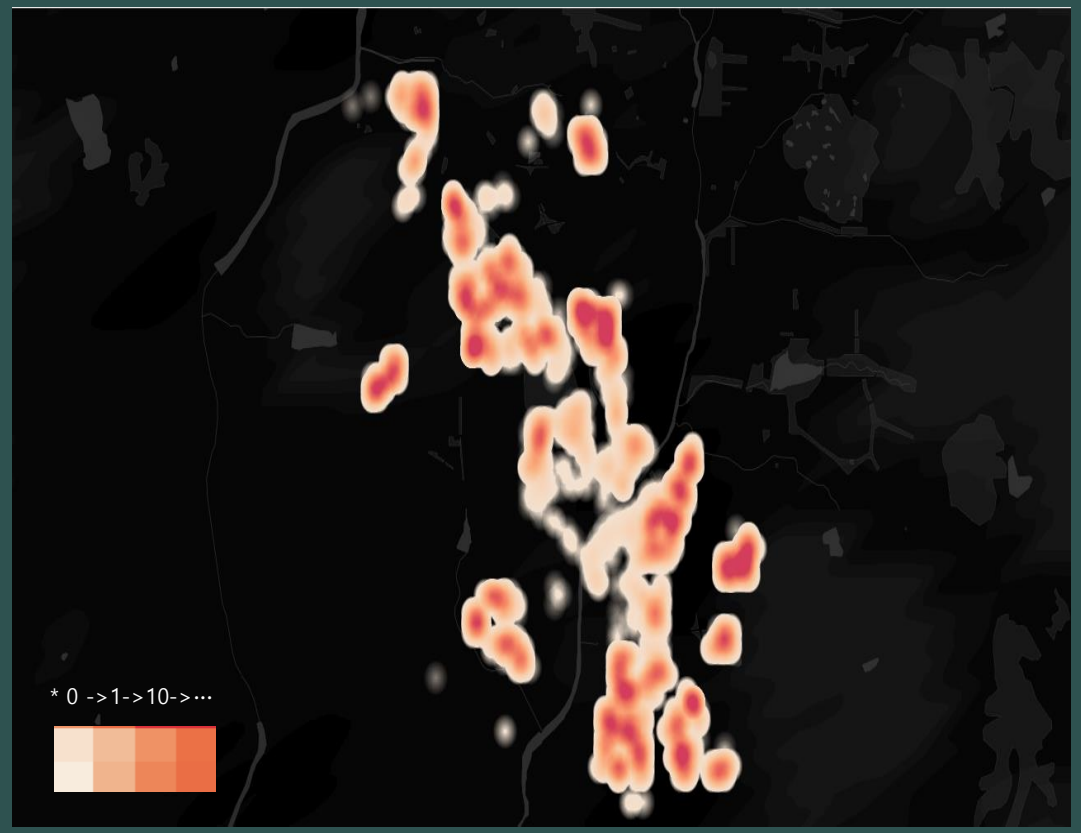
선정 지역 中 하나인 **희망주택아파트일대**를 구글 Road View를 통해 확인해본 결과 **도로수,신호등 수**, **주변 교육시설** 등 앞에서 확인했던 어린이 교통사고 위험에 **유의미한 영향을 주는 요소**들이 많이 존재함을 확인할 수 있음

4-2. 교통안전시설물 우선 설치 지역 확인

교통안전시설물(과속단속카메라, 교통안전시설물 등) 우선 설치 지역 20개소 제시

<교통안전시설물 우선 설치지역 20개소 제시>

<유소년 인구 Heatmap>



- 선정된 우선설치 구역은 어린이 보호구역으로 지정되었음에도 불구하고 시설물이 **적정수에 도달하지 못한** 구역
- **유소년인구 Plot**과 비교해 봄으로써 K-means의 과정을 통해 우선 설치구역 몰림 현상을 방지했음을 확인

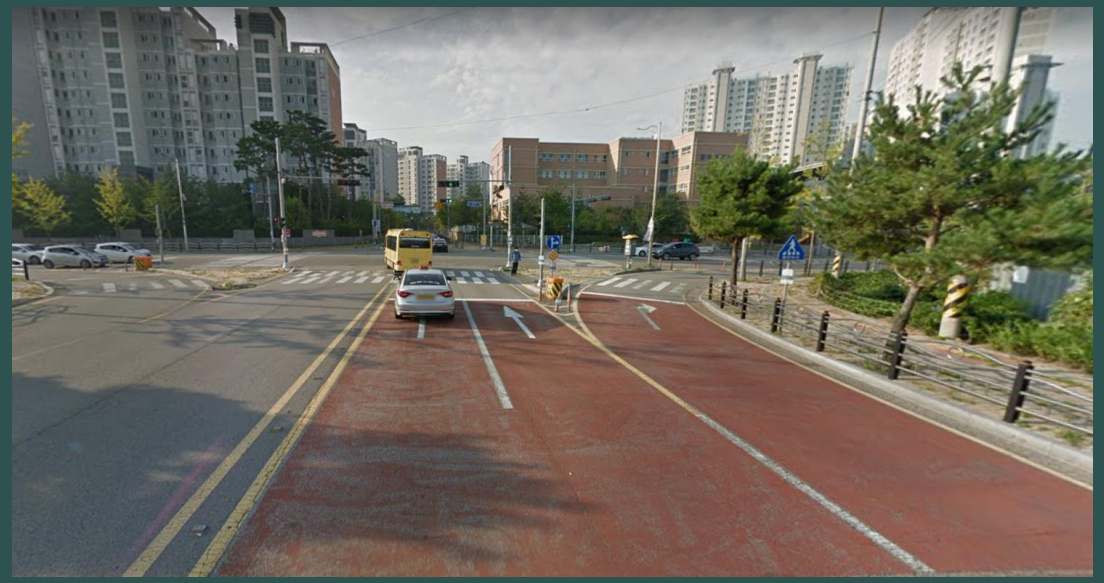
4-2. 교통안전시설물 우선 설치 지역 확인

교통안전시설물(과속단속카메라, 교통안전시설물 등) 우선 설치 지역 20개소 제시

위험순위	시설명/주소지	중심점 위치		비고
		X좌표(경도)	Y좌표(위도)	
1	양산초등학교 일대	127.0318	37.1973	도로안전 표지판
~	~	~	~	~
20	오산대원초등학교 일대	127.0716	37.1317	신호등

선정 지역 中 하나인 양산초등학교일대를 구글 Road View를 통해 확인해본 결과 CCTV, 과속방지턱, 도로안전표지판과 같은 시설물 설치가 적정 수에 도달하지 못할 만큼 부족하다는 것을 확인

예시) 양산초등학교 일대



5. 의의 및 한계점

[의의]

MCLP로 1차 선정구역을 추린 후 K-means를 이용하여 최종적으로는 위험지역 혹은 필요지역에 과도하게 몰리지 않고 외곽지역에도 선정될 수 있는 모델

[한계점]

어린이교통사고건수와 같은 주요변수들이 Point가 아닌 격자(100x100) 데이터로 주어져 데이터 통합과정에서 오차가 발생할 가능성 존재

감사합니다.