

목 차

행정으로 국민의 삶의질을 개선하라!

해커톤

분석결과 보고서

서울시 공중화장실 위험도 분석 및 위험 지도 개발

참여자:

송인용, 송**, 안**, 오**, 오**, 이**, 이**

씨에스리 컨소시엄

CSLEE 한국생산성본부

1. 분석 개요	4
가. 분석 배경 및 개요	4
나. 분석 목적 및 방향	5
다. 분석 결과 활용 방안	5
2. 분석 데이터	6
가. 분석 데이터 목록	6
나. 데이터 상세 설명	6
다. 데이터 정제 방안	13
3. 분석 프로세스	26
가. 분석 프로세스	26
나. 분석 내용 및 방법	26
4. 공중화장실 위험도 분석결과	33
가. QGIS 분석 프로세스	33
1) 격자 데이터 준비	
2) QGIS 좌표계 설정	
3) 위험인자 필드 계산	
4) 가중치 부여	
나. 분석 결과 산출	44
다. 시각화 구현	44
5. 활용 방안	44

가. 문제점 개선 방안	44
나. 업무 활용 방안	45

1. 분석 개요

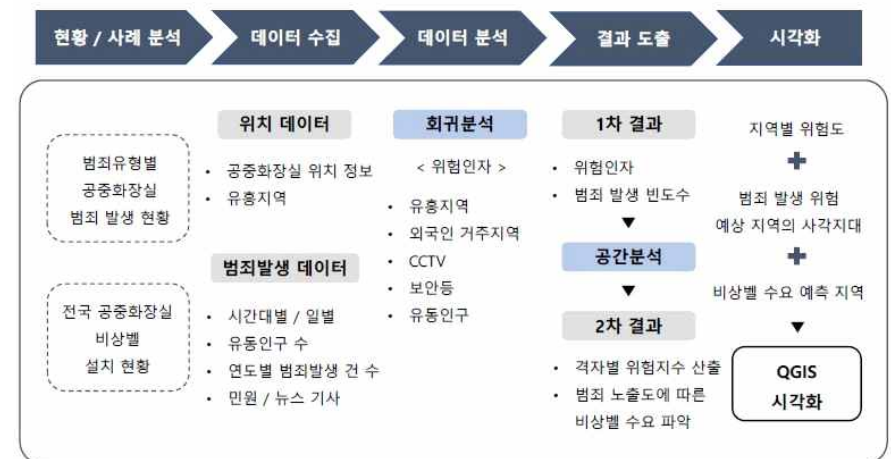
가. 분석 배경 및 개요

1) 서울시 공중화장실 위험도 분석 배경

- ☐ 서울시 공중화장실의 범죄 발생 다발 지역이나 발생 가능성이 높은 지역을 바탕으로 위험도를 분석하여 위험 지도 개발
- ☐ 개발된 지도는 향후 긴급 비상벨 설치 시 효율적인 정책 의사결정 지원

2) 서울시 공중화장실 위험도 분석 개요

서울시 공중화장실 위험도 분석은 서울시 5대 강력범죄 발생 현황과 비상벨 설치 현황을 분석하고 지역별 위험도 예측 및 위험 지도를 개발하여 범죄 취약지역과 비상벨 우선 설치 지역을 선정, 범죄발생 이전에 선제적인 비상벨 설치로 사회 안전망을 강화하고 제한된 예산을 효율적으로 집행할 수 있는 기반 마련



[그림 1-1]

나. 분석 목적 및 방향

1) 서울시 공중화장실 위험도 예측 분석

- ☐ 범죄 발생 현황 및 주요 위험인자들을 기반으로 하여 위험도를 예측 분석
- ☐ 범죄 취약 대상으로 선정된 지역을 기반으로 비상벨 설치 우선 지역을 선정하여 시각화

2) 분석결과의 정책의사 결정 지원

- ☐ 빅데이터를 기반으로 공중화장실 및 범죄 사각지대의 범죄 위험도를 분석하여 비상벨 우선 설치 지역을 선정하여 효율적인 정책 의사결정 지원
- ☐ 비상벨 설치 요청 민원이나 기존 지자체 기관 담당자의 경험과 직관에 의존하던 설치지역 결정 방식을 빅데이터 분석 기반의 우선 설치지역 선정 방식으로 전환하여 한정된 예산을 효율적으로 집행

다. 분석 결과 활용 방안

1) 지자체 업무 활용의 측면

- ☐ ‘공중화장실 등에 관한 법률 일부개정법률안’이 국회를 통과할 경우, 공중화장실의 긴급 비상벨 설치가 의무화되며, 이에 따른 빅데이터 기반의 범죄 위험도 분석 결과를 바탕으로 비상벨 우선 설치 지역을 선정하는 합리적 의사결정 가능
- ☐ 객관적인 통계자료 및 빅데이터 분석을 통해 최우선 설치 지역을 선정하여 우선적으로 설치함으로써 한정된 예산으로 최대의 효과 창출

2) 국민 생활 안전의 측면

- ☐ 범죄 취약 지역에 대한 사전 예방적 비상벨 설치로 사회 안전망 구축
- ☐ 비상벨 호출 즉시 자동으로 중앙관제센터로 위치가 전송되어 각종 긴급 상황 시에 빠른 대응 가능

2. 분석 데이터

가. 분석 데이터 목록

서울시 공중화장실 위험도 분석을 위해 필요한 데이터는 아래와 같습니다.

[표 1-1]

No.	데이터	출처	형태
1	서울시 공중화장실 위치 정보	공공데이터포털	CSV
2	자치구별 외국인 인구 비율	공공데이터포털	CSV
3	시간대별 유동인구 수	서울시 상권분석서비스	CSV
4	서울시 총 인구 수	국토통계지도	SHP
5	연령대별 여성 인구 비율	국토통계지도	SHP
6	자치구별 CCTV 수	공공데이터포털	CSV
7	유흥주점영업 인허가 정보	서울시 열린데이터광장	CSV
8	보안등 위치 표준데이터	공공데이터포털	CSV
9	노후화 건물 비율	국가공간정보포털	SHP
10	면적별 1인당 경찰관 수	서울지방경찰청	CSV
11	서울시 도로노선 정보	서울시 열린데이터광장	CSV
12	숙박시설 등록 정보	서울시 열린데이터광장	CSV
13	공시지가 격자 정보	국토통계지도	SHP
14	건폐율 격자 정보	국토통계지도	SHP
15	서울시 단독주택 수	국토통계지도	SHP
16	범죄 관련 뉴스	네이버 포털	CSV

나. 데이터 상세 설명

1) 서울시 공중화장실 위치 정보

서울시 공중화장실 위험도 분석을 하기 위한 핵심 데이터로 공중화장실 위치 정보를 파악할 수 있으며, 건물명칭, 좌표정보 등의 속성 데이터를 포함하고 있습니다.

	A	B	C	D	E	F	G	H	I	J
1	POI ID	대명칭	중명칭	소명칭	중앙좌표X1	중앙좌표Y1	WSG84X좌표	WSG84Y좌표	등록일자	수정일자
2	102423	우성스포츠허브	민간개발화장실		192026.0773	443662.9039	126.9098324	37.49238615	20100712	20100712
3	102424	프레곤빌딩	민간개발화장실		191560.4489	442968.6992	126.9045751	37.48612718	20100712	20100712
4	102425	하림빌딩	민간개발화장실		201472.0427	443869.7965	127.016646	37.4942835	20100712	20100712
5	102426	크레신타워	민간개발화장실		201633.5404	446523.7166	127.0184781	37.51819518	20100712	20100712
6	102483	한주실업빌딩	민간개발화장실		195045.521	446004.0241	126.9439599	37.51350089	20100712	20100712
7	102484	백광빌딩	민간개발화장실	무	195528.255	442509.7842	126.9494413	37.48202004	20100712	20100712
8	102485	서원빌딩	민간개발화장실		195228.9802	442534.5135	126.9460575	37.48224135	20100712	20100712
9	102434	SC원보건설/귀족	개발		210662.4457	447613.636	127.1206262	37.52795523	20100712	20100712
10	102435	브라운스톤전호	개발		210943.4685	448733.15	127.1238221	37.53803875	20100712	20100712

[그림 2-1] 서울시 공중화장실 위치 데이터

2) 자치구별 외국인 인구 비율

서울 생활인구 장기/단기 체류 외국인 정보로 중국, 기타 외국인 체류자로 구분하여 서울시내 거주하는 외국인 수를 행정 동 단위로 확인할 수 있습니다.

기준일ID	시간대구분	자치구코드	총생활인구수	중국인체류인구수	중국외외국인체류인구수
20200917	00	11560	36779.3003	33136.2938	3643.0092
20200917	00	11590	15352.8472	12121.4504	3231.3921
20200917	00	11620	24072.893	18173.7999	5899.093
20200917	00	11650	10707.7533	4362.5224	6345.2272
20200917	00	11680	15144.3152	6104.3483	9039.9671
20200917	00	11710	14871.8151	9325.2687	5546.547
20200917	00	11740	9400.0909	6998.0861	2402.003
20200917	01	11110	14942.8101	9948.8155	4993.9922
20200917	01	11140	17032.727	11696.314	5336.4136
20200917	01	11170	18635.2573	4184.6302	14450.6269
20200917	01	11200	14911.6985	10014.9816	4896.7184
20200917	01	11215	23719.4383	18723.2825	4996.1523
20200917	01	11230	24083.8237	17800.1235	6283.703
20200917	01	11260	6717.3709	4061.8411	2655.5315

[그림 2-2] 자치구별 외국인 인구 데이터

3) 시간대별 유동인구 수

서울시 우리마을 상권서비스에서 제공하는 데이터로써, 연령대, 성별, 시간대별 유동인구 데이터를 구축하는데 활용할 수 있습니다.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	기준 년코드	기준 분기 코드	상권 구분 코드	상권 구분 코드	상권 코드 명	총 유동인구 수	남성 유동인구 수	여성 유동인구 수	연령대_10_유동인	연령대_20_유동인	연령대_30_유동인	연령대_40_유동인	
2	2014	1 A	골목상권	1000001	계동길	147154	85726	61430	3250	23914	44733	41050	
3	2014	1 A	골목상권	1000002	난계로27길	189584	115933	73649	2531	23886	45672	46017	
4	2014	1 A	골목상권	1000003	돈계로11가	97007	60604	36403	666	12116	23662	25507	
5	2014	1 A	골목상권	1000004	명륜길	95160	51702	43460	6185	37165	18976	16599	
6	2014	1 A	골목상권	1000005	백석동길	81885	41317	40567	5421	19591	14637	18395	
7	2014	1 A	골목상권	1000006	북촌로11길	71933	42874	29060	1737	11464	21531	19610	
8	2014	1 A	골목상권	1000007	북촌로5길	88588	52873	35717	1786	16053	27858	23838	
9	2014	1 A	골목상권	1000008	북촌로5나길	37217	22177	15039	711	5190	11988	10976	
10	2014	1 A	골목상권	1000009	삼청로5길	15623	9259	6365	191	2486	5265	4734	

[그림 2-3] 유동인구 데이터

4) 서울시 총 인구 수

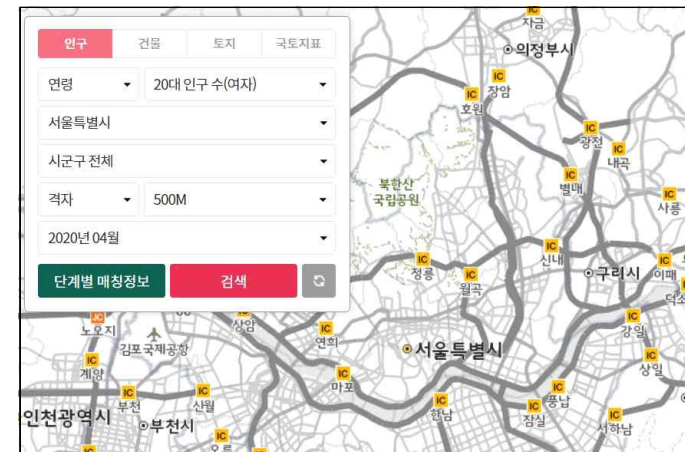
행정 동별 서울 총 생활인구 수로 전체 인구 수 대비 여성 인구 비율을 나타내는 데 활용될 수 있습니다.

행정동코드	총생활인구수	남자0세부터9...	남자10세부터1...	남자15세부터1...	남자20세부터
11110515	14772.8816	533.6249	340.027	496.6113	513.6963
11110530	14475.7955	358.5768	159.3673	232.3019	407.5177
11110540	4361.0046	71.2009	54.5567	126.638	427.659
11110550	15544.0872	428.6551	332.373	445.7122	542.8405
11110560	20313.4157	536.6221	388.1563	633.5162	668.093
11110570	8658.0486	256.4764	164.0252	286.2242	376.853
11110580	6267.2533	299.0466	118.9616	156.2381	168.2058
11110600	5417.9332	120.8073	85.2755	135.5139	172.3745
11110615	20337.2659	375.7191	118.3774	205.0757	866.3341

[그림 2-4] 총 인구 데이터

5) 연령대별 여성 인구 비율

국토통계지도에서 제공하는 단계별 매칭정보로써, 서울시 시군구 전체의 성별 및 연령별 인구 데이터를 격자 데이터로 다운받아 활용할 수 있습니다.



[그림 2-5] 여성 인구 비율 데이터

6) 자치구별 CCTV 수

서울시 자치구별 CCTV 설치 현황으로 각 25개구의 CCTV 설치 개수와 설치 목적 및 좌표정보(위도, 경도)를 포함하고 있습니다.

	A	B	C	D	E	F	G	H	I	J	K	L
1	관리기관	소재지도로명주소	소재지지번주소	설치목적구	카메라대수	카메라화소	촬영방면정	보관일수	설치년월	관리기관전	위도	경도
4	서울특별시	서울특별시 강남구 서울특별시 강남 생활방면			4	200	360전방면	30		02-3423-6772	37.4854108	127.054705
5	서울특별시	서울특별시 강남구 서울특별시 강남 생활방면			4	200	360전방면	30		02-3423-6772	37.4861526	127.056121
13	서울특별시	서울특별시 강남구 서울특별시 강남 생활방면			3	200	360전방면	30		02-3423-6772	37.4832803	127.054393
14	서울특별시	서울특별시 강남구 서울특별시 강남 생활방면			4	200	360전방면	30		02-3423-6772	37.4807876	127.061627
15	서울특별시	서울특별시 강남구 서울특별시 강남 생활방면			3	200	360전방면	30		02-3423-6772	37.4832803	127.054393
16	서울특별시	서울특별시 강남구 서울특별시 강남 생활방면			4	200	360전방면	30		02-3423-6772	37.4816184	127.058736
17	서울특별시	서울특별시 강남구 서울특별시 강남 생활방면			4	200	360전방면	30		02-3423-6772	37.4773494	127.058573
18	서울특별시	서울특별시 강남구 서울특별시 강남 생활방면			3	200	360전방면	30		02-3423-6772	37.4816184	127.058736
20	서울특별시	서울특별시 강남구 서울특별시 강남 생활방면			3	200	360전방면	30		02-3423-6772	37.4760653	127.067286

[그림 2-6] 자치구별 CCTV 데이터

7) 서울시 유흥주점 영업 인허가 정보

서울시의 유흥주점 영업 인허가 정보로써, 유흥종사자를 두거나 유흥시설을 설치하여 운영하는 업소의 영업상태와 업태 구분, 좌표정보 및 소재지 등을 파악할 수 있습니다.

개방자치단체명	관리번호	인허가일자	인허가취소일자	영업상태코드	영업상태명	상세영'
3000000	3000000-...	20181219		03	폐업	02
3050000	3050000-...	20000105		01	영업/정상	01
3040000	3040000-...	19800408		03	폐업	02
3240000	3240000-...	20000623		01	영업/정상	01
3240000	3240000-...	20110303		01	영업/정상	01
3200000	3200000-...	20130724		01	영업/정상	01
3010000	3010000-...	20061124		03	폐업	02
3110000	3110000-...	20020218		01	영업/정상	01
3000000	3000000-...	20010921		01	영업/정상	01

[그림 2-7] 유흥지역 데이터

8) 가로등/보안등 위치 표준데이터

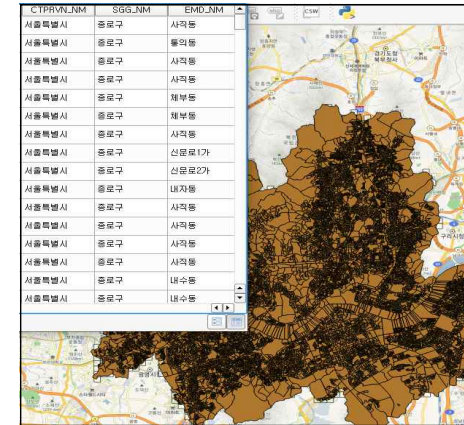
전국 보안등정보(보안등 위치, 설치개수 등), 소재지, 좌표 정보(위도, 경도), 관리기관명을 알 수 있습니다.

	A	B	C	D	E	F	G	H	I	J	K
1	보안등위치명	설치개수	소재지도로명주소	소재지지번주소	위도	경도	관리기관전화번호	관리기관명	데이터기준일자	제공기관코드	제공기관명
2	영화사로7길 40	40	서울특별시 관진서울특별시 관진	37.55689466	127.0931344	02-450-7873	서울특별시 관진	2019-01-01	3040000	서울특별시 관진	
3	영화사로7길 40	40	서울특별시 관진서울특별시 관진	37.55675416	127.0930528	02-450-7873	서울특별시 관진	2019-01-01	3040000	서울특별시 관진	
4	영화사로3길 20	20	서울특별시 관진서울특별시 관진	37.55785442	127.0917232	02-450-7873	서울특별시 관진	2019-01-01	3040000	서울특별시 관진	
5	영화사로3길 20	20	서울특별시 관진서울특별시 관진	37.55786542	127.0914832	02-450-7873	서울특별시 관진	2019-01-01	3040000	서울특별시 관진	
6	영화사로3길 20	20	서울특별시 관진서울특별시 관진	37.55757012	127.0911886	02-450-7873	서울특별시 관진	2019-01-01	3040000	서울특별시 관진	
7	영화사로3길 18	18	서울특별시 관진서울특별시 관진	37.55765671	127.0910665	02-450-7873	서울특별시 관진	2019-01-01	3040000	서울특별시 관진	
8	영화사로3길 19	19	서울특별시 관진서울특별시 관진	37.55784057	127.0909988	02-450-7873	서울특별시 관진	2019-01-01	3040000	서울특별시 관진	
9	영화사로3길 15	15	서울특별시 관진서울특별시 관진	37.55767136	127.0907587	02-450-7873	서울특별시 관진	2019-01-01	3040000	서울특별시 관진	
10	영화사로3길 14	14	서울특별시 관진서울특별시 관진	37.55750544	127.0909441	02-450-7873	서울특별시 관진	2019-01-01	3040000	서울특별시 관진	

[그림 2-8] 보안등 위치 데이터

9) 노후 건물 비율

집계구별 노후주택비율로써 행정동코드, 시군구명, 전체주택 수, 20년이상 주택 수, 20년이상 주택 비율, 좌표 등의 속성들을 포함하고 있습니다.



[그림 2-9] 노후건물 데이터

10) 면적별 1인당 경찰관 수

서울지방경찰청에서 제공하는 면적별 1인당 경찰관 수로써, 지역별 면적 및 인구 밀도 등의 속성을 포함하며, 서울시 총 인구 수 대비 1인당 경찰관 수를 산출할 수 있습니다.

	A	B	C	D	E
1	기간	지역	인구	면적	인구밀도
2	2015	강남구	581,760	39.5	14,728
3	2015	강동구	463,321	24.59	18,842
4	2015	강북구	334,426	23.6	14,172
5	2015	강서구	595,691	41.44	14,376
6	2015	관악구	529,031	29.57	17,891
7	2015	광진구	375,180	17.06	21,987
8	2015	구로구	454,604	20.12	22,596
9	2015	금천구	256,167	13.02	19,675
10	2015	노원구	578,221	35.44	16,317
11	2015	도봉구	353,241	20.71	17,056
12	2015	동대문구	373,824	14.21	26,301
13	2015	동작구	412,774	16.35	25,240
14	2015	마포구	398,351	23.84	16,706
15	2015	서대문구	323,105	17.61	18,351
16	2015	서초구	451,258	47	9,601
17	2015	성동구	305,065	16.86	18,093
18	2015	성북구	469,560	24.58	19,104
19	2015	송파구	667,480	33.88	19,704
20	2015	양천구	489,010	17.4	28,098

[그림 2-10] 면적별 경찰관 데이터

11) 서울시 도로노선 정보

서울시의 도로노선에 도로명, 도로종류, 도로기능, 도로규모, 도로폭 등의 정보를 알 수 있습니다.

순번	노선명(도로명)	도로종류	도로기능	도로규모	도로폭	시도구도구분	
1	가락로	일반도로	국지도로	소로3류	폭6-8m	구도	
2	가락로11길	일반도로	국지도로	소로3류	폭6-8m	구도	
3	가락로12길	일반도로	국지도로	소로2류	폭8-10m	구도	
4	가락로13길	일반도로	국지도로	소로3류	폭6-8m	구도	
5	가락로15길	일반도로	국지도로	소로	6m미만	구도	
4338	등촌로39라길	일반도로	국지도로	소로	6m미만	구도	
4339	등촌로39마길	일반도로	국지도로	소로	6m미만	구도	
4340	등촌로39바길	일반도로	국지도로	소로	6m미만	구도	
4341	등촌로47길	일반도로	국지도로	소로	6m미만	구도	
6	가락로16길	일반도로	국지도로	소로2류	폭8-10m	구도	
7	가락로17길	일반도로	국지도로	소로	6m미만	구도	
8	가락로18길	일반도로	국지도로	소로	6m미만	구도	

[그림 2-11] 도로노선 데이터

12) 숙박시설 등록 정보

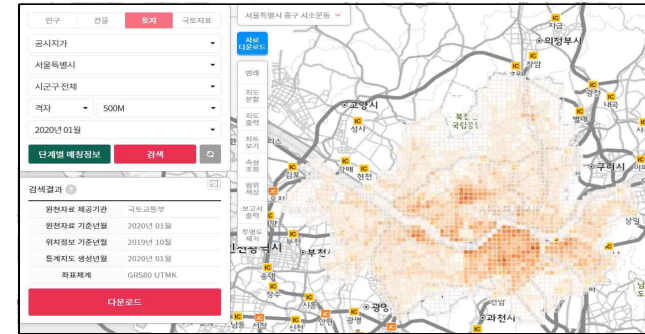
서울시에 있는 숙박업소 등록 현황으로 업소명, 업종, 소재지 등을 파악할 수 있습니다.

A	B	C	D	E	F
연번	업소명	업종	업태	도로명 주소	연락처
1	이지스테이	숙박업(일반)	여관업	서울특별시 중로구 창신길 9-8, (창신동)	764-5537
2	호스텔 바닐라 1	숙박업(일반)	여관업	서울특별시 중로구 창신길 28-7, (창신동)	6459-5113
3	영모텔	숙박업(일반)	여관업	서울특별시 중로구 보문로7길 5-1, (송인동)	927-6393
4	호텔더디자인즈 종로	숙박업(일반)	일반호텔	서울특별시 중로구 수표로 89-8, (관수동)	2267-7474
5	대진	숙박업(일반)	여관업	서울특별시 중로구 창신1길 6, (창신동)	743-5242
6	대호	숙박업(일반)	여관업	서울특별시 중로구 성관로1길 6-2, (명륜3가)	745-5656
7	크리스하우스	숙박업(일반)	여관업	서울특별시 중로구 보문로9길 9, (송인동)	8967-6888
8	동미	숙박업(일반)	여관업	서울특별시 중로구 종로65길 12-28, (송인동)	763-3638
9	삼오	숙박업(일반)	여관업	서울특별시 중로구 종로60길 9, (송인동)	2232-7122

[그림 2-12] 숙박시설 데이터

13) 공시지가 격자 정보

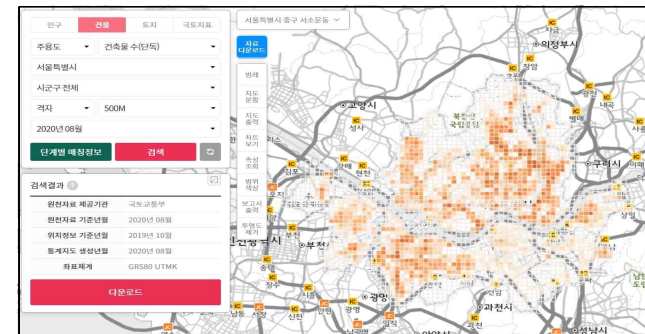
국토통계지도에서 서울시 전체의 공시지가 데이터를 격자 데이터로 다운받아 활용할 수 있습니다.



[그림 2-13] 공시지가 데이터

14) 서울시 단독주택 수

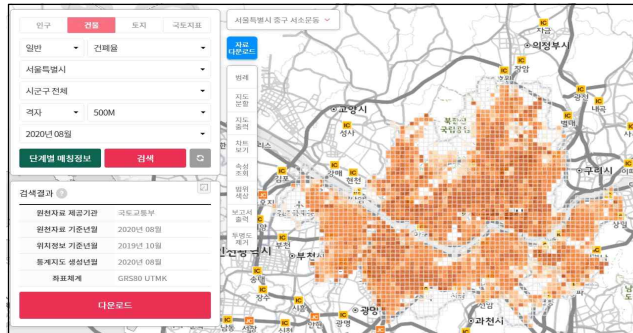
국토통계지도에서 서울시 전체의 단독주택 수를 격자 데이터로 다운받아 활용할 수 있습니다.



[그림 2-14] 단독주택 데이터

15) 건폐율 격자 정보

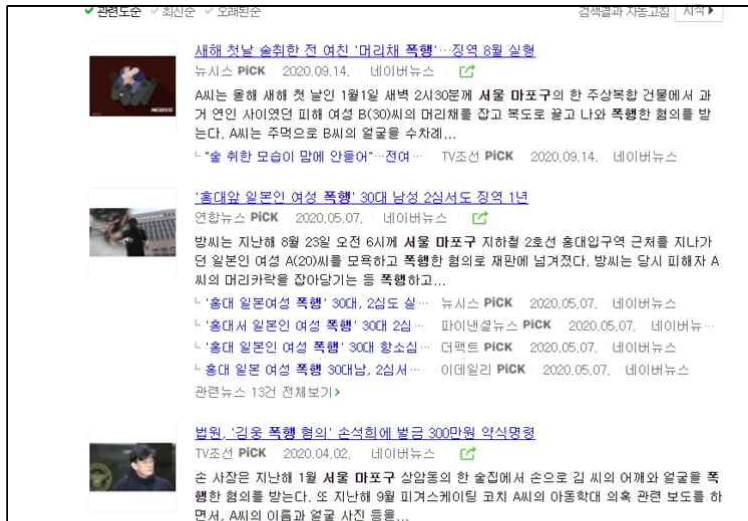
국토통계지도에서 서울시 전체 대지면적 당 건축 면적 비율을 건폐율 데이터로 다운받아 활용할 수 있습니다.



[그림 2-15] 건폐율 데이터

16) 범죄 관련 뉴스

네이버 포털사이트 뉴스탭을 이용해 서울시 범죄 관련 뉴스를 크롤링하여 가져올 수 있습니다.



[그림 2-16] 범죄관련 뉴스 화면

다. 데이터 정제 방안

BigZami, Python, Excel 프로그램을 이용해 데이터 셋 확인 및 결측값 처리, 이상값 처리, 지오코딩 등의 데이터 전처리 과정을 거칩니다.

□ 유동인구 데이터

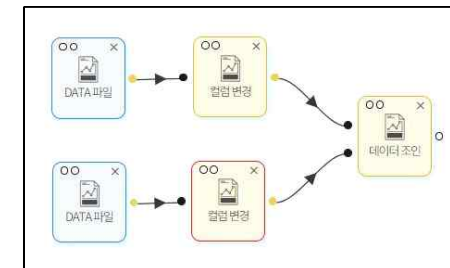
유동인구 데이터는 행정동 코드로 분류되어 있어, 쉽게 알아볼 수 있는 구와 동의 분류로 변환하기 위해 빅재미를 활용해 데이터 조인 처리합니다.

A	B	C	D
행정동코드	기준일ID	시간대구분	총생활인구
11110515	20200801	0	13475.26
11110515	20200801	1	14103.55
11110515	20200801	2	14132.28
11110515	20200801	3	14152.17
11110515	20200801	4	14175.26
11110515	20200801	5	14098.2
11110515	20200801	6	14344.84
11110515	20200801	7	14463.36
11110515	20200801	8	14567.08
11110515	20200801	9	14449.92
11110515	20200801	10	14368.17
11110515	20200801	11	14566.45
11110515	20200801	12	14468.77
11110515	20200801	13	14744.24
11110515	20200801	14	14963.58

[그림 2-17] 유동인구 데이터

자치구	행정동코드	행정동
종로구	11110515	청운효자동
종로구	11110515	청운효자동
종로구	11110515	청운효자동
종로구	11110515	청운효자동
종로구	11110515	청운효자동
종로구	11110515	청운효자동
종로구	11110515	청운효자동
종로구	11110515	청운효자동
종로구	11110515	청운효자동
종로구	11110530	사직동
종로구	11110530	사직동
종로구	11110530	사직동
종로구	11110530	사직동
종로구	11110530	사직동
종로구	11110530	사직동

[그림 2-18] 변환표



[그림 2-19] 빅재미 데이터 조인

데이터 영역					
행정동코드	가동일	시간대구분	총생활인구수	자치구	행정동
1110515	20200801	0	13475.2594	종로구	한문여자동
1110515	20200801	1	14100.9541	종로구	한문여자동
1110515	20200801	2	14182.2032	종로구	한문여자동
1110515	20200801	3	14152.1696	종로구	한문여자동
1110515	20200801	4	14175.2033	종로구	한문여자동
1110515	20200801	5	14096.0969	종로구	한문여자동
1110515	20200801	6	14344.0033	종로구	한문여자동
1110515	20200801	7	14463.3817	종로구	한문여자동
1110515	20200801	8	14567.0792	종로구	한문여자동
1110515	20200801	9	14446.9152	종로구	한문여자동
1110515	20200801	10	14368.2715	종로구	한문여자동
1110515	20200801	11	14566.4531	종로구	한문여자동
1110515	20200801	12	14469.7732	종로구	한문여자동

[그림 2-20] 조인 데이터

□ 조인한 유동인구 데이터는 피벗테이블을 이용해 구별 데이터로 변환합니다.

2020 8월 기준	
행 레이블	평균 : 총생활인구수
강남구	34110.75
강동구	28516.09
강북구	22784.37
강서구	28529.85
관악구	22968.43
광진구	24119.41
구로구	28163.22
금천구	22039.19
노원구	25962.31
도봉구	19974.1
동대문구	25966.77
동작구	28132.8
마포구	29488.12

[그림 2-21] 피벗테이블

구별 시간대별 유동인구(8월 전월 기준)															
행 레이블	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
강남구	29487.83	29274.36	29128.88	29038.1	29034.33	29283.9	30075.38	31512.47	34266.38	36907.53	38124.09	38827.69	39365.74	39771.1	39925.4
강동구	30011.73	30064.38	30100.35	30138.23	30148.02	30087.25	29798.8	29303.08	28339.12	27741.08	27382.54	27036.08	26770.24	26608.12	26596.97
강북구	24379.83	24386.12	24438.13	24442.27	24358.65	24119.28	23720.74	23081.35	22298.04	21773.78	21469.79	21318	21240.45	21148.81	21124.63
강서구	28513.48	28531.27	28574.04	28599.99	28712.85	28968.26	29199.41	29302.08	28985.16	28761.86	28745.92	28453.24	28370	28386.89	28380.44
관악구	24699.15	24762.92	24800.89	24818.37	24792.94	24667.26	24348.19	23832.24	22731.27	21859.23	21453.66	21184.89	20988.65	20828.11	20777.26
광진구	24953.28	24931.59	24920.82	24902.13	24885.2	24838.12	24724.89	24503.54	23658.69	23101.35	22923.65	22776.15	22740.16	22753.84	22830.02
구로구	28964.57	28985.26	29019.51	29033.84	29047.54	28946.97	28781.05	28812.37	28313.26	27862.93	27600.45	27307.34	27107.47	26954.09	26919.7
금천구	20449.78	20481.91	20501.92	20519.14	20491.37	20437.08	20545.91	21176.37	22680.66	23558.79	23776.26	23832.28	23894.28	23934.82	23944.14
노원구	27584.12	27607.32	27612.48	27614.09	27552.41	27413.71	27013.83	26434.59	25500.24	25025.23	24688.89	24390.78	24224.15	24178.67	24202.11
도봉구	20937.12	21015.13	21065.07	21103.71	21057.27	21009.92	20779.38	20510.24	19816.77	19257.67	18987.84	18788.47	18652.3	18537.84	18539.52
동대문구	25823.77	25837.36	25875.85	25910.02	25992.61	26024.69	26091.72	26365.12	26389.2	26307.61	26260.44	26246.1	26115.94	26041	25992.12
동작구	29637.96	29645.98	29635.57	29648.71	29604.24	29688.26	29550.08	29195.48	28116.9	27273.22	26821.78	26439.06	26257.66	26102.2	26073.62
마포구	28326.41	28176.58	28043.78	27959.79	27898.51	27933.84	28040.53	28125.29	28742.82	29439.06	29790.45	30055.34	30336.59	30585.03	30778.78

[그림 2-22] 구별 데이터로 변환한 최종 결과 데이터

□ 유흥주점 데이터

유흥주점 데이터는 엑셀 필터를 이용해 현재 영업중인 구별 데이터만을 필터링 합니다.

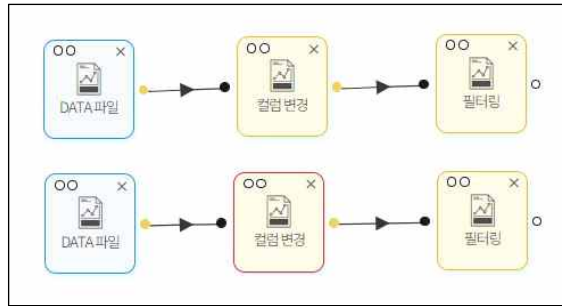
구	유흥주점 개수
강남구	208
강동구	124
강북구	68
강서구	127
관악구	192
광진구	16
구로구	34
금천구	49
노원구	21
도봉구	23
동대문구	72
동작구	47
마포구	55

[그림 2-23] 유흥주점 필터링 데이터

영업상태명	상세영업상세영업상
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업
영업/정상	TRUE 1 영업

[그림 2-24] 필터링 결과 데이터

- 영업중인 데이터를 추출한 후, 빅재미를 활용해 컬럼을 원하는 형태로 변경합니다.



[그림 2-25] 빅재미 컬럼 변경

- 년도별 유흥주점 통계를 내기 위해 엑셀을 이용해 데이터 이상 유무를 확인한 후, 파이썬(Python) 판다스 라이브러리를 이용해 년도별 구간을 나누어 전처리 과정을 거칩니다.

[그림 2-26] 파이썬을 이용한 전처리 과정

	A	B	C	D	E	F	G
구		2020	2019	2018	2017	2016	2015
강남구		207	218	222	233	255	285
강동구		124	131	133	143	148	153
강북구		68	70	74	74	72	74
강서구		127	130	133	133	118	120
관악구		192	211	212	220	218	232
광진구		16	17	18	18	18	19
구로구		34	35	35	33	37	40
금천구		49	50	51	53	54	60
노원구		21	21	23	24	25	28

[그림 2-27] 파이썬을 이용한 전처리 결과 데이터

- 보안등 엑셀 데이터는 개별 파일별로 분리되어 있어 하나의 파일로 통합한 후, 엑셀과 빅재미를 활용하여 데이터를 정제합니다.

이름	수정된 날짜	유형	크기
서울특별시 동대문구_보안등정보_20200427	2020-09-18 오후 2:11	Microsoft Excel 열...	1
서울특별시 강남구_보안등정보_20190816	2020-09-18 오전 11:43	Microsoft Excel 열...	1
서울특별시 강동구_보안등정보_20200911_1600068487538_1161530	2020-09-18 오전 11:37	Microsoft Excel 열...	1
서울특별시 강북구_보안등정보_20190521	2020-09-18 오전 11:43	Microsoft Excel 열...	3
서울특별시 강서구_보안등정보_20191218	2020-09-18 오전 11:43	Microsoft Excel 열...	1
서울특별시 관악구_보안등정보_20190605	2020-09-18 오전 11:43	Microsoft Excel 열...	1
서울특별시 광진구_보안등정보_20200708_1594361968419_1027782	2020-09-18 오전 11:42	Microsoft Excel 열...	1
서울특별시 구로구_보안등정보_20200624_1592982508940_1339676	2020-09-18 오전 11:42	Microsoft Excel 열...	1
서울특별시 금천구_보안등정보_20200813_1599205227765_900904	2020-09-18 오전 11:41	Microsoft Excel 열...	1

[그림 2-28] 개별 파일로 분리되어 있는 보안등 데이터

- 엑셀 필터를 이용해 위도 경도의 좌표값이 없는 주소를 추출합니다.

A	B	C	D	E	F	G	H	I	J	K
보안등	설치개수	소재지도로명주소	소재지지번주소	위도	경도	설치년도	설치형태	관리기관	관리기관	데이터구분
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	
개포동660	1	서울특별시 강남구 개포동660-37						02-3423-6 서울특별시	*****	

[그림 2-29] Null 값 추출

□ 지오코딩은 검색 가능량이 제한되어 있어 카카오 지도 API를 이용해 위도, 경도 좌표값을 추출합니다.

```

In [3]: import requests; from urllib.parse import urlparse
import pandas as pd

In [63]: df = pd.read_csv('0803.csv')
df['주소']

Out[63]:
0      서울특별시 강남구 개포동660-37
1      서울특별시 강남구 개포동141
2      서울특별시 강남구 개포동660-12
3      서울특별시 강남구 개포동660-21
4      서울특별시 강남구 개포동660-20

30395      서울특별시 중랑구 묵동 3-5
30396      서울특별시 중랑구 원목초교
30397      서울특별시 중랑구 망우동 폐기장 진입로
30398      서울특별시 중랑구 두원사당
30399      서울특별시 중랑구 신내동 신흥초교
Name: 주소, Length: 30400, dtype: object

In [44]: def coordinate(address):
url = "https://dapi.kakao.com/v2/local/search/address.json?query=" + address
result = requests.get(urlparse(url).geturl(), headers={"Authorization": "KakaoAK [...]})
json_obj = result.json()
val = []
for document in json_obj['documents']:
    val = [document['address_name'], document['x'], document['y']]

In [52]: coordinate('서울특별시 은평구 진관동 352-3')
Out[52]: ['서울 은평구 진관동 352-3', '126.943252927789', '37.6376433121982']

In [64]: list = []
for address in df['주소']:
    list.append(coordinate(address))
df1 = pd.DataFrame(list, columns = ['address_name', 'x', 'y'])
df1.head()

Out[64]:
   address_name      x      y
0  서울 강남구 개포동 660-37  127.062079083808  37.4804729709299
1  서울 강남구 개포동 141  127.061614360667  37.4784893722905
2  서울 강남구 개포동 660-12  127.060454620403  37.4793317842759
3  서울 강남구 개포동 660-21  127.058573366497  37.4773534846907
4  서울 강남구 개포동 660-20  127.058668163998  37.4774264194266

In [66]: df1 = pd.DataFrame(list, columns = ['address_name', 'x', 'y'])
df1.head()

```

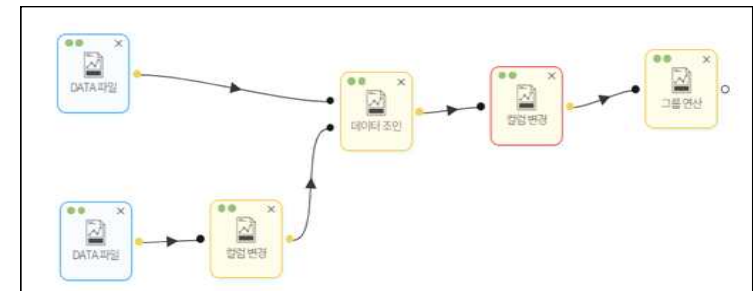
[그림 2-30] 좌표값 추출 화면

□ 경찰관 수 데이터에서 엑셀을 이용해 31개의 관서로 구분된 데이터에서 25개 자치구 데이터로 통합하여 변환합니다.

C	D	E	F	G	H	I	J	K	L
지역	인구	면적	인구밀도		2015	2016	2017	2018	2019
강남구	581,760	39.5	14,728	1365	10.78974	10.61099	10.40513	10.15311	10.2044
강동구	463,321	24.59	18,842	700	26.91714	26.05429	25.58429	25.09143	25.58429
강북구	334,426	23.6	14,172	617	22.96921	22.71475	22.52512	22.17666	21.81848
강서구	595,691	41.44	14,376	828	17.36232	17.54952	17.72826	17.593	17.4372
관악구	529,031	29.57	17,891	799	22.39174	22.24781	22.04881	22.01126	21.89737
광진구	375,180	17.06	21,987	647	33.983	33.70479	33.72334	33.61051	33.24111
구로구	454,604	20.12	22,596	730	30.95342	30.61233	30.06301	29.85342	29.9137
금천구	256,167	13.02	19,675	535	36.7757	36.56075	36.39065	36.46729	36.1514
노원구	578,221	35.44	16,317	756	21.58333	21.32143	20.83069	20.46032	20.05556
도봉구	353,241	20.71	17,056	529	32.24197	32.03781	31.6673	31.24953	30.70888
동대문구	373,824	14.21	26,301	758	34.69789	34.36675	33.96834	33.81266	33.68997
동작구	412,774	16.35	25,240	637	39.62323	39.66876	39.21193	39.2967	39.25118
마포구	398,351	23.84	16,706	787	21.22745	20.831	20.55146	20.58323	20.55909

[그림 2-31] 구별로 통합한 경찰관 수 데이터

□ 외국인 체류자 데이터가 행정동 코드로 분류되어 있으므로, 외국인 구/동별 체류 평균을 구하기 위해 빅재미를 활용해 행정동 코드표와 외국인 데이터 컬럼을 변경한 후 데이터 조인합니다.



[그림 2-32] 빅재미로 컬럼변경 후 조인

□ 파이썬과 카카오 API를 이용해 서울시 구/동 별 위도, 경도 좌표값을 추출합니다.

```
def coordinate(address):
    url = "https://dapi.kakao.com/v2/local/search/address.json?query="+address
    result = requests.get(uriparse(url).geturl(), headers={"Authorization": "KakaoAK"})
    json_obj = result.json()
    val = []
    for document in json_obj['documents']:
        val = [document['address_name'], document['x'], document['y']]

    return val

In [52]: coordinate('서울특별시 은평구 진관동 352-3')
Out[52]: ['서울 은평구 진관동 352-3', '126.943252927789', '37.6376433121992']

In [4]: list = []
for addres in df['구 동']:
    list.append(coordinate(addres))
df1 = pd.DataFrame(list, columns = ['address_name', 'x', 'y'])
df1.head()
```

```
In [4]: list = []
for addres in df['구 동']:
    list.append(coordinate(addres))
df1 = pd.DataFrame(list, columns = ['address_name', 'x', 'y'])
df1.head()
```

```
Out[4]:
```

	address_name	x	y
0	서울 강남구 개포1동	127.057489994874	37.48197916687
1	서울 강남구 개포2동	127.069133611378	37.4898261143433
2	서울 강남구 개포4동	127.051622162647	37.4788499689324
3	서울 강남구 논현1동	127.028552712008	37.511470639272
4	서울 강남구 논현2동	127.037350857638	37.5173931863142

```
In [66]: df1 = pd.DataFrame(list, columns = ['address_name', 'x', 'y'])
df1.head()
```

```
Out[66]:
```

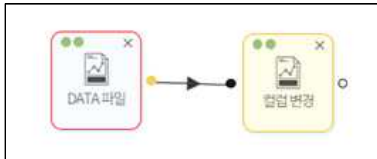
	address_name	x	y
0	서울 강남구 개포동 660-37	127.062079083808	37.4804729709299
1	서울 강남구 개포동 141	127.061614360667	37.4784893722905
2	서울 강남구 개포동 660-12	127.060454620403	37.4793317842759
3	서울 강남구 개포동 660-21	127.058573366497	37.4773534846907
4	서울 강남구 개포동 660-20	127.058668163998	37.4774264194256

[그림 2-33] 좌표값 추출 과정

□ 카카오 API를 활용해 추출한 좌표값과 현재 체류중인 외국인 데이터 엑셀 파일을 결합해 원하는 데이터로 가공합니다.

	A	B	C	D	E		A	B	C	D	E	F
								address_name	x	y		총생합관구수평균
17	서도구	목야면	1266.245	서도구	목야면	1	0	서도 관남구 개포초	57.48197948994874	57.48197916687		120.481329
18	서도구	신촌면	7337.72	서도구	신촌면	3	1	서도 관남구 개포초	57.0699136113378	57.4898261143433		244.74029
19	서도구	신촌면	5547.246	서도구	신촌면	4	2	서도 관남구 개포초	57.062162162447	57.4780496893924		245.91767
20	서도구	전원면	6993.9429	서도구	전원면	5	3	서도 관남구 논현초	57.208552712008	57.5114706369173		177.688335
21	서도구	전원면	1243.976	서도구	전원면	6	4	서도 관남구 논현초	57.20733791861427	57.513971861427		124.00028
22	서도구	훈촌계1	254.4767	서도구	훈촌계1	7	5	서도 관남구 대지초	57.05682016381	57.493260109643		358.36057
23	서도구	훈촌계2	1009.111	서도구	훈촌계2	8	6	서도 관남구 대지초	57.064564557307	57.502250908328		919.45012
24	서도구	훈촌계3	461.7111	서도구	훈촌계3	9	7	서도 관남구 대지초	57.057147782925	57.499728886978		85.25734
25	서도구	훈촌계4	158.4364	서도구	훈촌계4	10	8	서도 관남구 대지초	57.038896912922	57.4883045975414		805.10036
26	서도구	훈촌계5	33.147	서도구	훈촌계5	11	9	서도 관남구 대지초	57.038896912922	57.4883045975414		63.38097
27	서도구	반포면	1520.961	서도구	반포면	12	10	서도 관남구 삼정초	57.062523947115	57.514391209485		156.82735
28	서도구	반포면	403.9138	서도구	반포면	13	11	서도 관남구 삼정초	57.0625270584216	57.5112632807882		949.52512
29	서도구	반포면	324.2885	서도구	반포면	14	12	서도 관남구 새곡초	57.106760267266	57.4644437055763		528.62516
30	서도구	반포면	404.6674	서도구	반포면	15	13	서도 관남구 우제초	57.120530937213	57.488878954615		355.59050
31	서도구	반포면	242.326	서도구	반포면	16	14	서도 관남구 신사초	57.120737448907	57.523680298463		387.71063
32	서도구	반포면	153.9293	서도구	반포면	17	15	서도 관남구 양곡초	57.1300829170421	57.5306864414269		143.86705
	<div>배치도</div>											
	<div>대상을 선택한 후, Enter 키를 눌러서 해당 지역을 선택합니다.</div>											
	<div>50.9360767, 126.415: 현재 위치</div>											

□ 빅재미를 이용해 원하는 CSV 파일 형식으로 정제합니다.



[그림 2-37] CSV로 정제

□ 최종적으로 얻은 경찰관 1인당 인구밀도 수치와 서울시 인구데이터가 txt 파일로 제공되므로, 파이썬의 판다스 라이브러리를 이용해 원하는 형태의 CSV 파일로 변환합니다.

```

In [2]: import pandas as pd
df = pd.read_csv("서울시 인구수.txt", sep = "\t")
df.to_excel('c:\Users\User\서울시 인구.xlsx')
    
```

구분	자치구	세대	이거	남자	여자	계	남자	여자	계	남자	여자	세대당인구	65세이상고령자
2015	합계	4,189,948	10,297,138	5,063,550	5,233,588	10,297,138	4,930,943	5,091,236	9,957,179	132,607	142,350		
2015	종로구	72,645	163,822	80,531	83,291	154,986	76,548	78,438	154,986	8,836	3,983	2.13	24,892
2015	중구	59,321	134,329	66,755	67,574	125,733	62,687	63,046	125,733	8,596	4,068	2.12	20,302
2015	용산구	107,512	247,909	121,027	126,882	233,342	113,177	120,165	233,342	14,567	7,850	2.17	35,715
2015	성동구	126,714	305,065	151,459	153,606	297,003	147,570	149,433	297,003	8,062	3,889	2.34	38,164
2015	광진구	158,353	375,180	183,436	191,744	360,369	176,884	183,485	360,369	14,811	6,552	2.28	40,368
2015	동대문구	159,941	373,824	185,827	187,997	360,153	180,485	179,668	360,153	13,671	5,342	2.25	52,490
2015	동양구	176,390	418,620	208,393	210,227	418,620	206,505	207,404	413,909	14,711	1,888	2.35	54,095
2015	성북구	190,973	469,560	229,183	240,377	469,560	224,980	234,295	459,275	10,285	4,203	2.4	63,281
2015	강북구	141,088	334,426	164,337	170,089	330,873	162,967	167,906	330,873	3,553	1,370	2.35	52,598
2015	도봉구	136,903	353,241	173,804	179,437	353,241	173,029	178,213	351,242	1,999	775	2.57	48,590
2015	노원구	219,768	578,221	281,538	296,683	578,221	279,987	294,596	574,583	3,638	1,551	2.61	68,909

[그림 2-38] 파이썬을 이용해 CSV 파일로 변환

□ 엑셀 계산 기능을 이용해 경찰관 1명이 담당하는 인구 수 데이터를 계산합니다.

	B	C	D	E	F	G	H	I	J	K	L	M	N
1	2015년	2016년	2017년	2018년	2019년	자치구	이거	구별	2015	2016	2017	2018	2019
2	1365	1431	1451	1482	1532	강남구	581,760	강남	426,1978	399,8183	386,6657	369,4015	359,1443
3	700	723	741	763	766	강동구	463,321	강동	661,8871	620,2918	594,2767	566,0813	574,9217
4	617	670	667	672	671	강북구	334,426	강북	542,0194	493,5881	491,7571	480,5283	473,465
5	828	881	902	944	941	강서구	595,691	강서	719,4336	683,4325	674,3404	639,4184	635,7843
6	799	849	861	885	918	관악구	529,031	관악	662,1164	619,0895	605,0279	587,6158	563,5447
7	647	679	697	712	724	광진구	375,180	광진	579,8764	548,0177	534,1435	521,1559	506,8674
8	730	763	749	768	752	구로구	454,604	구로	622,7452	589,2529	589,5314	570,9453	584,2699
9	535	563	580	583	575	금천구	256,167	금천	478,8168	452,3162	437,0534	435,7136	437,9478
10	756	798	811	833	809	노원구	578,221	노원	764,8426	715,8045	688,1319	658,0552	664,157
11	529	540	563	578	561	도봉구	353,241	도봉	667,7524	648,6519	614,9805	591,0882	598,2727
12	758	795	817	823	826	동대문구	373,824	동대문	493,1715	465,8013	447,9939	442,695	439,4952
13	637	672	670	678	674	동작구	412,774	동작	647,9969	614,9509	609,691	603,8127	606,6944
14	787	853	858	884	883	마포구	398,351	마포	506,1639	458,2497	449,6305	437,0577	437,0612
15	646	678	682	688	674	서대문구	323,105	서대문	500,1625	480,6357	476,5806	469,593	479,4822
16	996	1041	1071	1102	1115	서초구	451,258	서초	453,0703	433,6955	415,8739	397,6071	390,2305
17	627	656	651	651	642	성동구	305,065	성동	486,547	468,2332	480,3548	486,1183	481,2757

[그림 2-39] 경찰관 1인당 담당 인구 수 데이터

□ 빅재미를 활용해 원하는 컬럼을 따로 추출합니다.

구분	2015	2016	2017	2018	2019
강남	426,1978	399,8183	386,6657	369,4015	359,1443
강동	661,8871	620,2918	594,2767	566,0813	574,9217
강북	542,0194	493,5881	491,7571	480,5283	473,465
강서	719,4336	683,4325	674,3404	639,4184	635,7843
관악	662,1164	619,0895	605,0279	587,6158	563,5447
광진	579,8764	548,0177	534,1435	521,1559	506,8674
구로	622,7452	589,2529	589,5314	570,9453	584,2699
금천	478,8168	452,3162	437,0534	435,7136	437,9478
노원	764,8426	715,8045	688,1319	658,0552	664,157
도봉	667,7524	648,6519	614,9805	591,0882	598,2727
동대문	493,1715	465,8013	447,9939	442,695	439,4952
동작	647,9969	614,9509	609,691	603,8127	606,6944
마포	506,1639	458,2497	449,6305	437,0577	437,0612
서대문	500,1625	480,6357	476,5806	469,593	479,4822
서초	453,0703	433,6955	415,8739	397,6071	390,2305
성동	486,547	468,2332	480,3548	486,1183	481,2757

[그림 2-40] 빅재미를 이용해 원하는 컬럼 추출

□ 최종적으로 경찰관 1인이 담당하는 인구 수 데이터를 완성합니다.

	A	B	C	D	E	F
1	구별	2015	2016	2017	2018	2019
2	강남	426.1978	399.8183	386.6657	369.4015	359.1443
3	강동	661.8871	620.2918	594.2767	566.0813	574.9217
4	강북	542.0194	493.5881	491.7571	480.5283	473.465
5	강서	719.4336	683.4325	674.3404	639.4184	635.7843
6	관악	662.1164	619.0895	605.0279	587.6158	563.5447
7	광진	579.8764	548.0177	534.1435	521.1559	506.8674
8	구로	622.7452	589.2529	589.5314	570.9453	584.2699
9	금천	478.8168	452.3162	437.0534	435.7136	437.9478
10	노원	764.8426	715.8045	688.1319	658.0552	664.157
11	도봉	667.7524	648.6519	614.9805	591.0882	598.2727
12	동대문	493.1715	465.8013	447.9939	442.695	439.4952
13	동작	647.9969	614.9509	609.691	603.8127	606.6944
14	마포	506.1639	458.2497	449.6305	437.0577	437.0612
15	서대문	500.1625	480.6357	476.5806	469.593	479.4822
16	서초	453.0703	433.6955	415.8739	397.6071	390.2305
17	성동	486.547	468.2332	480.3548	486.1183	481.2757

[그림 2-41] 경찰관 1인 담당 인구 수 데이터

□ 서울시 숙박시설 등록 현황

파이썬으로 카카오 API 지오크딩을 사용해 서울시 숙박시설 등록 현황을 전처리하여 준비합니다.

```
import requests; from urllib.parse import urlparse
import pandas as pd

df = pd.read_csv('서울시숙박업소.csv', encoding='utf-8')
df = df.dropna(subset=['도로명 주소'])

def coordinate(address):
    url = "https://dapi.kakao.com/v2/local/search/address.json?query="
    result = requests.get(urlparse(url).geturl(), headers={"Authorization": "KakaoAK 2010-08-26", "Authorization": "KakaoAK 2010-08-26"})
    json_obj = result.json()
    val = []
    for document in json_obj['documents']:
        val = [document['address_name'], document['x'], document['y']]
    return val

list = []
for address in df['도로명 주소']:
    list.append(coordinate(address))
df1 = pd.DataFrame(list, columns=['address_name', 'x', 'y'])
df1.head()

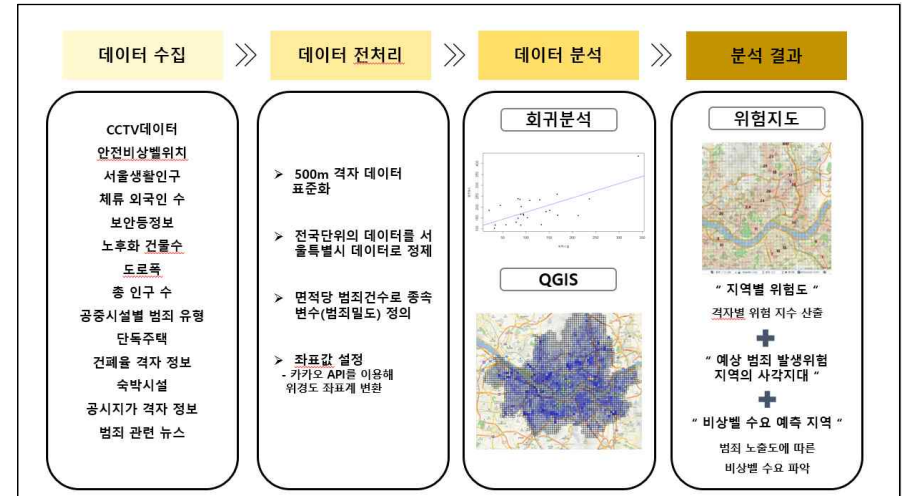
df1.to_csv('숙박업소좌표.csv', encoding='utf-8')
```

[그림 2-42] 서울시 숙박시설 등록 현황 전처리

3. 분석 프로세스

가. 분석 프로세스 개요

공중화장실 위험도 예측의 전체 분석 프로세스는 다음과 같습니다.



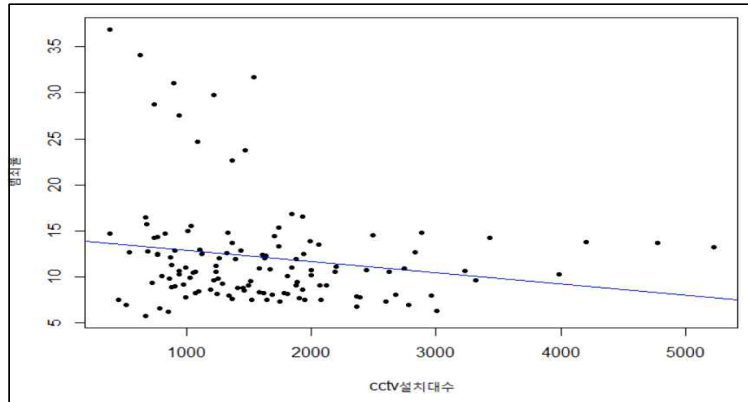
나. 분석 내용 및 방법

서울시 공중화장실 위험도 분석은 아래와 같이 예측한 가설에 대해 선형 회귀분석을 사용하여 적합도를 검증하고 이를 활용한 지역별 위험도 분석 및 시각화로 위험도를 도출

1) 회귀분석을 통한 가설 검증 (틀린 것도 다 포함시키기)

□ CCTV 설치대수와 범죄율

유의확률 0.04163을 가지므로 CCTV 설치대수는 범죄율에 유의한 영향을 미치며, 설치대수가 많아질수록 범죄율이 낮아지기는 하나, 영향은 미미하다는 것을 알 수 있습니다.



```

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.118146   1.077746  13.100  <2e-16 ***
cctv설치대수 -0.001217   0.000591  -2.059   0.0416 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

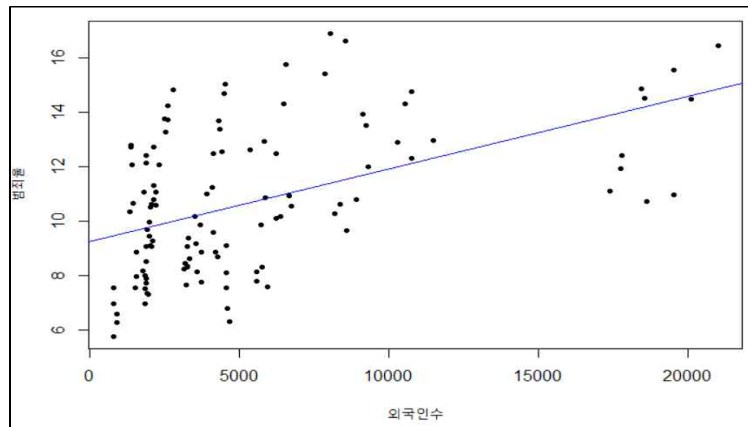
Residual standard error: 5.682 on 123 degrees of freedom
Multiple R-squared:  0.03331, Adjusted R-squared:  0.02545
F-statistic: 4.239 on 1 and 123 DF, p-value: 0.04163

```

[그림 3-1] CCTV와 범죄율 회귀모형 그래프

□ 외국인 수와 범죄율

유의확률 3.01e-08을 가지므로 외국인수는 범죄율에 유의한 영향을 미친다는 것을 알 수 있습니다.



```

> m2 <- lm(범죄율 ~ 외국인수)
> plot(외국인수, 범죄율, pch=20)
> abline(m2, col="blue")
> summary(m2)

Call:
lm(formula = 범죄율 ~ 외국인수)

Residuals:
    Min       1Q   Median       3Q      Max
-4.1805 -1.8382 -0.4988  1.6996  5.4840

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.265e+00  3.234e-01  28.646  < 2e-16 ***
외국인수      2.651e-04  4.452e-05   5.954  3.01e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

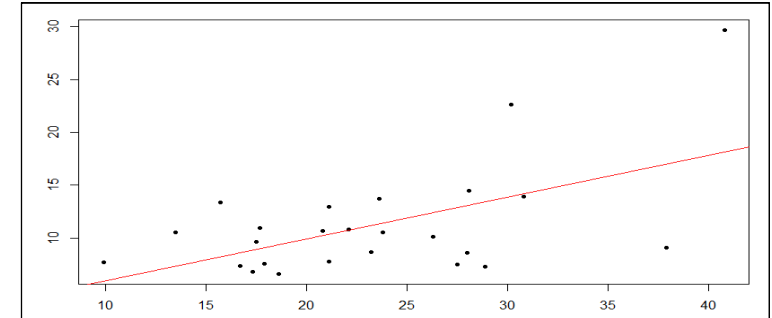
Residual standard error: 2.332 on 113 degrees of freedom
Multiple R-squared:  0.2388, Adjusted R-squared:  0.2321
F-statistic: 35.45 on 1 and 113 DF, p-value: 3.01e-08

```

[그림 3-2] 외국인 수와 범죄율 회귀모형 그래프

□ 노후건물 수와 범죄율

유의확률 0.004227을 가지므로 노후건물 수는 범죄율에 유의한 영향을 미친다는 것을 알 수 있습니다.



```

Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-7.8990 -2.5014 -0.2546  1.9648 11.5962

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.0231     3.0111   0.672  0.50835
x            0.3944     0.1242   3.175  0.00423 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

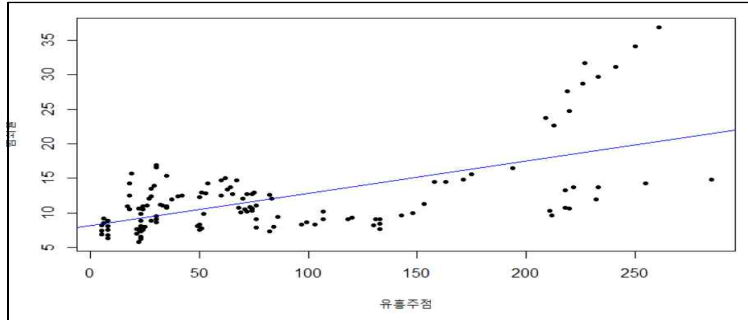
Residual standard error: 4.433 on 23 degrees of freedom
Multiple R-squared:  0.3047, Adjusted R-squared:  0.2744
F-statistic: 10.08 on 1 and 23 DF, p-value: 0.004227

```

[그림 3-3] 노후건물 수와 범죄율 회귀모형 그래프

□ 유흥주점 수와 범죄율

유의확률 2.581e-14를 가지므로 유흥주점 수는 범죄율에 유의한 영향을 미친다는 것을 알 수 있습니다.



```
Residuals:
    Min       1Q   Median       3Q      Max
-8.3688 -2.7064 -0.6801  2.0028 16.5572

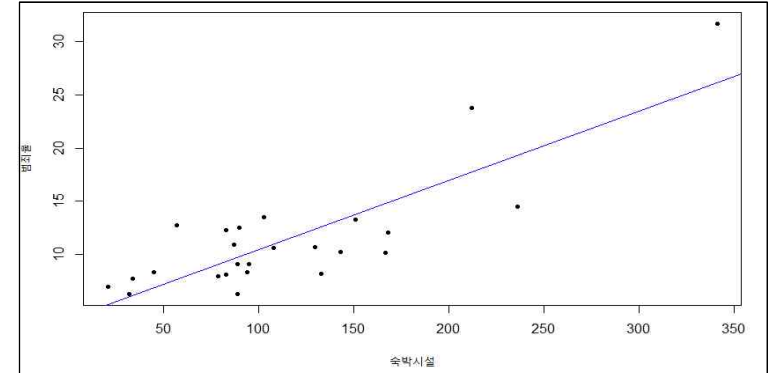
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.130394   0.620055  13.112 < 2e-16 ***
유흥주점      0.046713   0.005411   8.632 2.58e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.561 on 123 degrees of freedom
Multiple R-squared:  0.3773,    Adjusted R-squared:  0.3722
F-statistic: 74.52 on 1 and 123 DF,  p-value: 2.581e-14
```

[그림 3-4] 유흥주점 수와 범죄율 회귀모형 그래프

□ 숙박시설 수와 범죄율

유의확률 2.758e-08를 가지므로 숙박시설 수는 범죄율에 유의한 영향을 미친다는 것을 알 수 있습니다. 단, 표본의 수가 적어 신뢰도가 낮아질 수 있다는 가능성이 있습니다.



```
Call:
lm(formula = 범죄율 ~ 숙박시설)

Residuals:
    Min       1Q   Median       3Q      Max
-4.7766 -1.7326 -0.5066  1.6129  6.0208

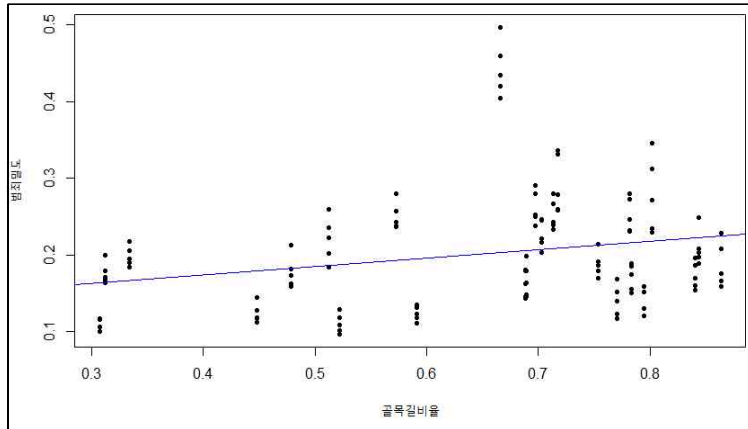
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.977597   1.216789   3.269 0.00337 **
숙박시설      0.064855   0.009065   7.154 2.76e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.153 on 23 degrees of freedom
Multiple R-squared:  0.69,    Adjusted R-squared:  0.6765
F-statistic: 51.19 on 1 and 23 DF,  p-value: 2.758e-07
```

[그림 3-5] 숙박시설 수와 범죄율 회귀모형 그래프

□ 골목길 비율과 범죄율

유의확률 0.006323을 가지므로 골목길 비율은 범죄율에 유의한 영향을 미친다는 것을 알 수 있습니다.



```
Call:
lm(formula = 범죄밀도 ~ 골목길비율)

Residuals:
    Min       1Q   Median       3Q      Max
-0.09821 -0.05847 -0.01523  0.03766  0.29366

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.13110    0.02625   4.995 1.97e-06 ***
골목길비율   0.10909    0.03927   2.778  0.00632 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

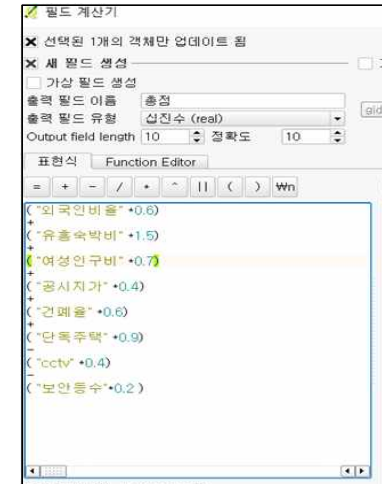
Residual standard error: 0.07279 on 123 degrees of freedom
Multiple R-squared:  0.05905, Adjusted R-squared:  0.0514
F-statistic: 7.719 on 1 and 123 DF, p-value: 0.006323
```

[그림 3-6] 골목길 비율과 범죄율 회귀모형 그래프

2) 지역별 위험도 분석

논문들을 참조하여 각각의 변수에 가중치를 곱한 후 연산한다.

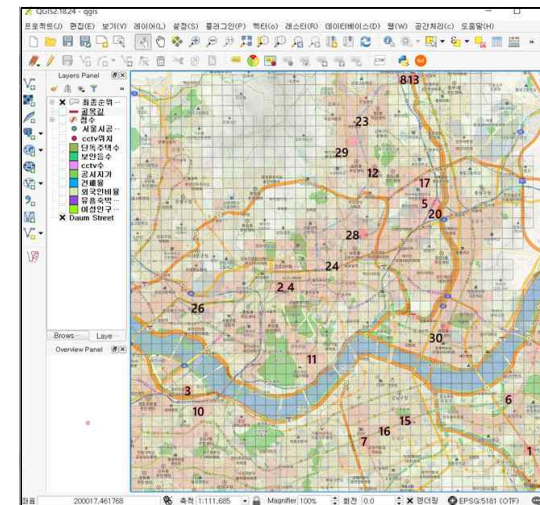
$$\begin{aligned}
 & ("외국인비율" * 0.6) + ("유형숙박비" * 1.5) + ("여성인구비" * 0.7) \\
 & + ("공시지가" * 0.4) + ("건폐율" * 0.6) + ("단독주택" * 0.9) \\
 & - ("cctv" * 0.4) - ("보안등수" * 0.2)
 \end{aligned}$$



[그림 3-7] 변수에 가중치 연산

3) 분석 결과 시각화

- 변수에 가중치를 곱해 모두 더하여 산출된 총 위험지수의 순위 필드 생성,
1위 ~ 30위 레이어의 순위를 출력



[그림 3-8] 최종 순위 설정된 시각화 화면

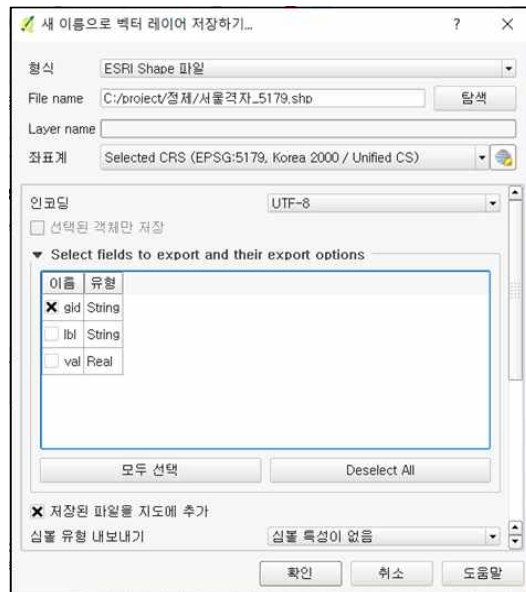
4. 공중화장실 위험도 분석결과

가. QGIS 분석 프로세스

1) 격자 데이터 준비

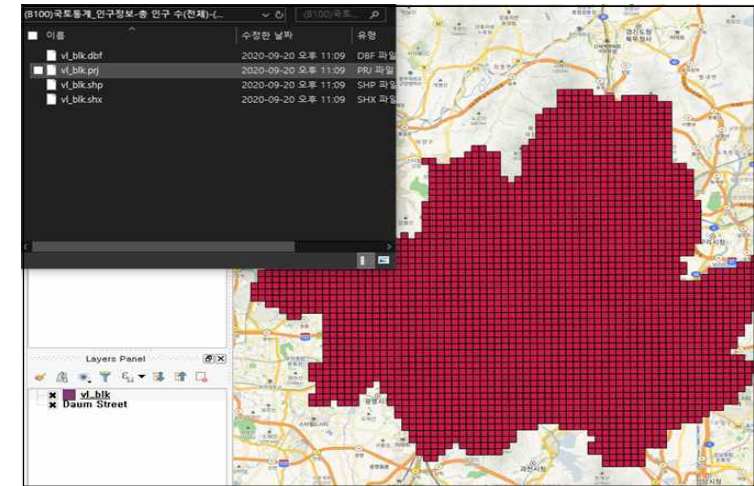
□ 빈 격자 생성

데이터를 병합하여 가공할 목적의 빈 격자 레이어를 생성합니다.



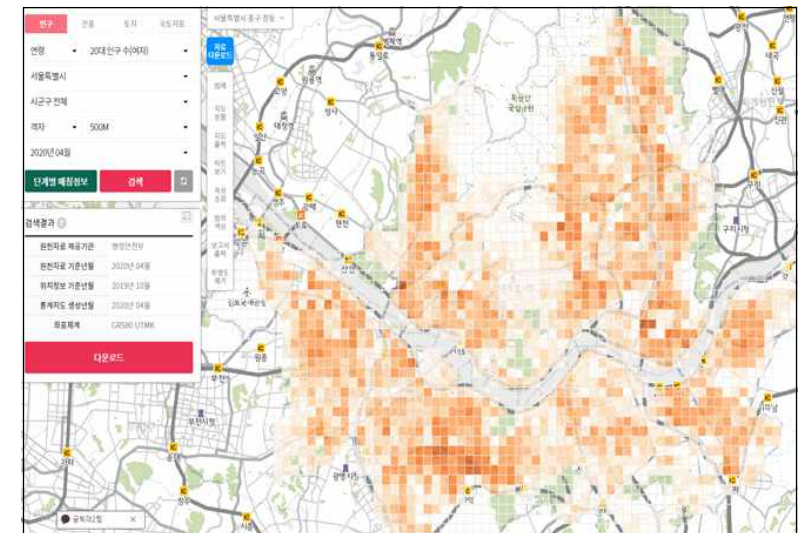
[그림 4-1] 빈 격자 레이어 생성

□ 국토 통계 인구정보 500M 격자 데이터를 준비합니다.



[그림 4-2] 국토통계 인구 격자 지도

□ 국토통계지도를 활용하여 서울시 여성 인구수를 격자에 표시합니다.



[그림 4-4] 색상 클래스 적용

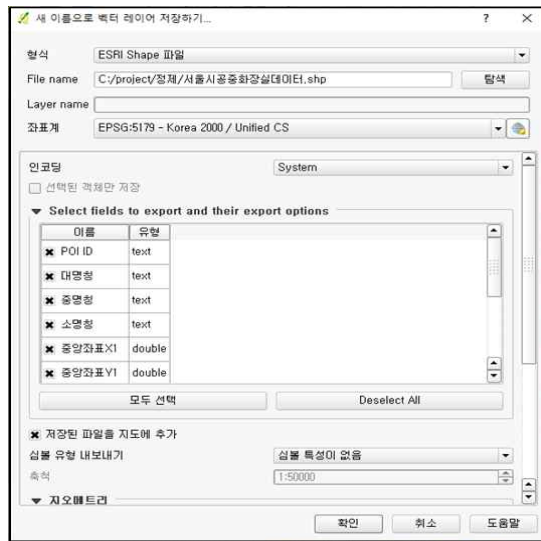
3) QGIS 좌표계 설정

□ 서울시 인구 격자 파일을 5179 좌표계로 변환 후 SHP 파일로 저장



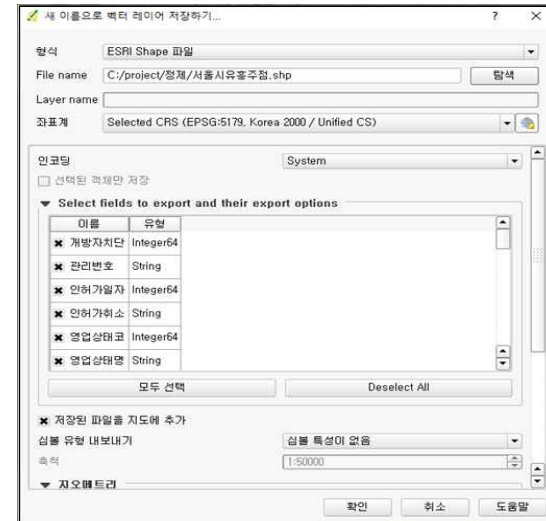
[그림 4-5] 인구 격자 좌표계 변환

□ 서울시 공중화장실 데이터를 CSV 파일로 불러온 후 좌표계 5179로 설정해 SHP 파일로 다른 이름 저장



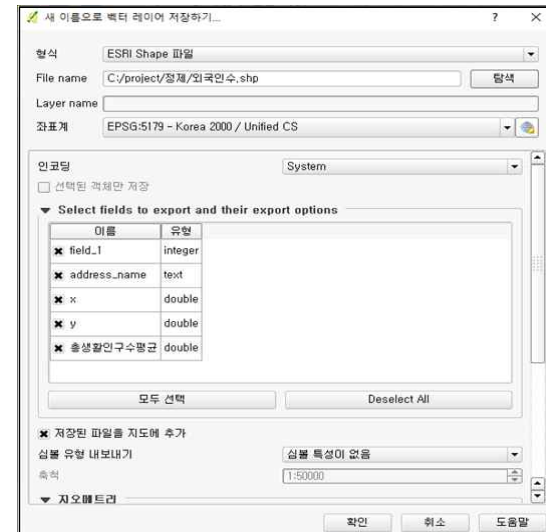
[그림 4-6] 공중화장실 좌표계 변환

□ 서울시 유흥주점 데이터를 5179 좌표계로 변환



[그림 4-7] 유흥주점 좌표계 변환

□ 구/동별 외국인 평균 체류 수 데이터를 5179 좌표계로 변환



[그림 4-8] 외국인 체류 수 좌표계 변환

□ 서울시 건물 통합정보 데이터를 5179 좌표계로 변환

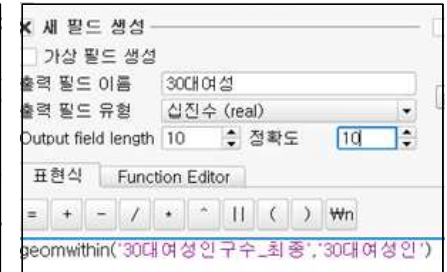
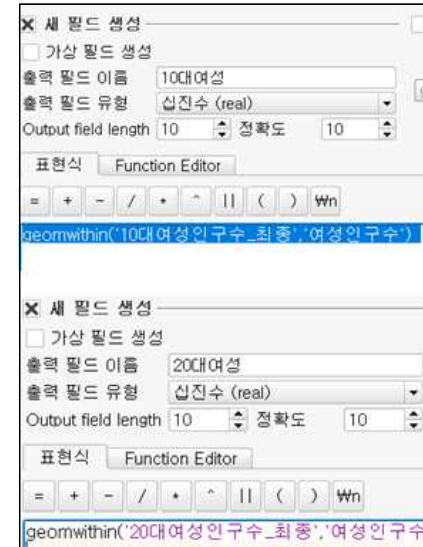
□ 서울시 여성인구 수 데이터를 5179 좌표계로 변환



[그림 4-10] 여성 인구 수 좌표계 변환

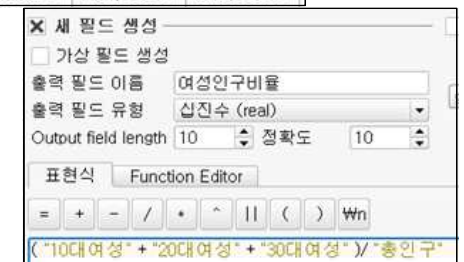
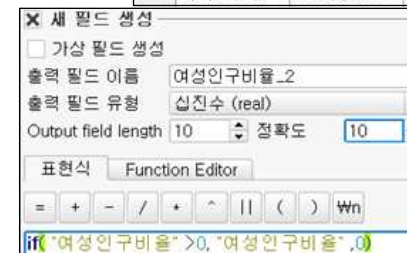
4) 위험인자 필드 계산

□ 여성인구 수 필드를 생성한 후 저장



[그림 4-11] 여성 인구 수 필드 생성

	gid	총인구	10대여성	20대여성	30대여성
1	다사66b42a	10536.0000...	69.0000000...	941.000000...	2108.00000...
2	다사49a43a	13021.0000...	41.0000000...	2164.00000...	1794.00000...
3	다사65a44a	14929.0000...	125.000000...	854.000000...	1686.00000...
4	다사67b48b	13063.0000...	109.000000...	1194.00000...	1619.00000...



[그림 4-12] 여성 인구 수 필드 계산 화면

[그림 4-14] 타레이어 참조

□ 서울시 총인구에 20대 여성인구 테이블 생성 완료

20대여성인구수_최종 :: Features total: 2634, filtered: 2634, selected: 1

	gid	여성인구수
1	다사60b48a	28
2	다사58b48a	217
3	다사44b44b	17

서울시인구500격자 :: Features total: 2634, filtered: 2634, selected: 1

	gid	lbl	val	여성인구
	다사60b48a	468,00	468,000000000...	28
	다사58b48a	3877,00	3877,000000000...	217
	다사44b44b	286,00	286,000000000...	17
	다사47a39a	5592,00	5592,000000000...	410
	다사53b54a	11,00	11,000000000...	
	다사53b48b	133,00	133,000000000...	
	다사51a54b	7509,00	7509,000000000...	527
	다사61a59a	8137,00	8137,000000000...	427

	gid	여성인구수
1	다사60b48a	28
2	다사58b48a	217
3	다사44b44b	17
4	다사47a39a	410
5	다사51a54b	527
6	다사61a59a	427
7	다사47b42b	721
8	다사58a45b	834
9	다사58a47b	303

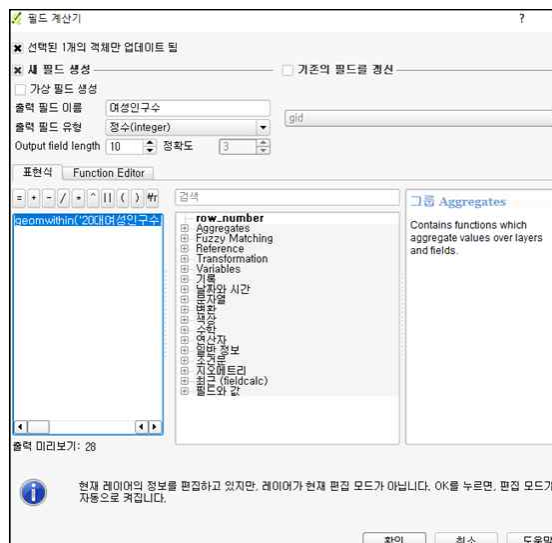
[그림 4-15] 여성인구 테이블 생성

□ refFunctions 플러그인을 설치

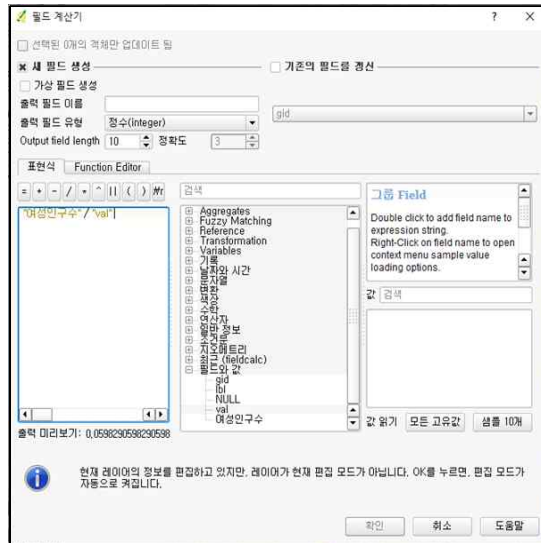


[그림 4-13] 플러그인 설치 화면

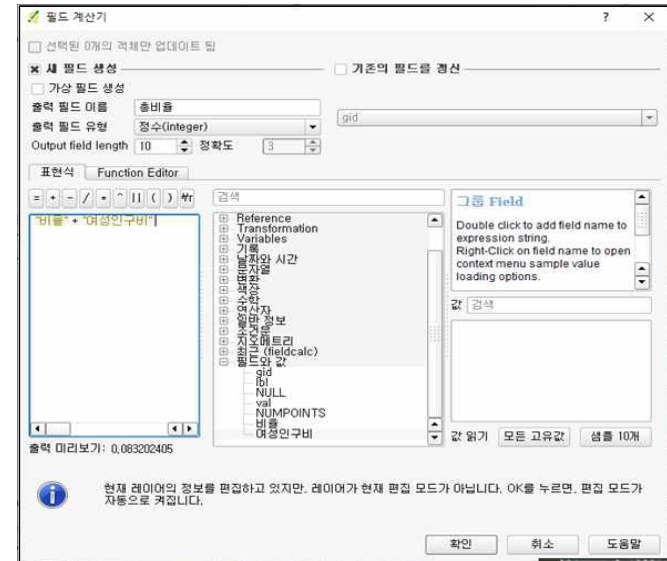
□ refFunctions 플러그인을 이용해 타 레이어 필드를 참조하여 필드 생성
- geomwithin('10/20/30대 여성인구수_최종', '여성인구수')]



☐ 총 위험도 비율을 설정 [비율 + 여성인구비]



[그림 4-17] 결과 계산 화면



[그림 4-19] 총 위험도 비율

서울시인구500억자 : Features total: 1813, filtered: 1813, selected: 0

sec gid |< >

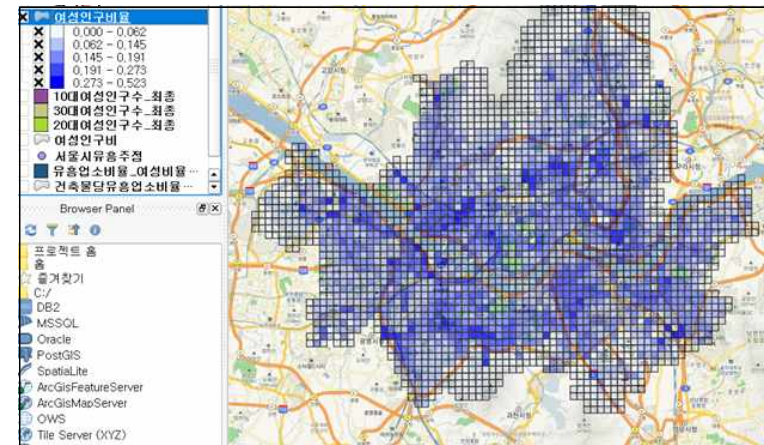
	gid	lbi	val	여경인구수	여경인구비율
	다사5a51b	123.00	123.0000000000	51,000,000,000	0.4146341463
2	다사62b59a	22.00	22.0000000000	9,000,000,000	0.4090909091
3	다사50b51a	3771.00	3771.0000000000	1143,000,000,000	0.3031026253
4	다사40b52a	2162.00	2162.0000000000	640,000,000,000	0.2960222017
5	다사58a54a	1534.00	1534.0000000000	430,000,000,000	0.2803129074
6	다사41a51a	1263.00	1263.0000000000	325,000,000,000	0.2573238321
7	다사45b42a	246.00	246.0000000000	63,000,000,000	0.2560975610
8	다사52a40a	133.00	133.0000000000	34,000,000,000	0.2556390977
9	다사55a54a	1218.00	1218.0000000000	295,000,000,000	0.2422003284
10	다사57b54b	4829.00	4829.0000000000	1111,000,000,000	0.2300683371
11	다사51a52a	1657.00	1657.0000000000	377,000,000,000	0.2275191138
12	다사62a49b	7114.00	7114.0000000000	1607,000,000,000	0.2258926061
13	다사46a53b	1332.00	1332.0000000000	298,000,000,000	0.2237237237

모든 객체 보기

[그림 4-18] 결과 산출 화면

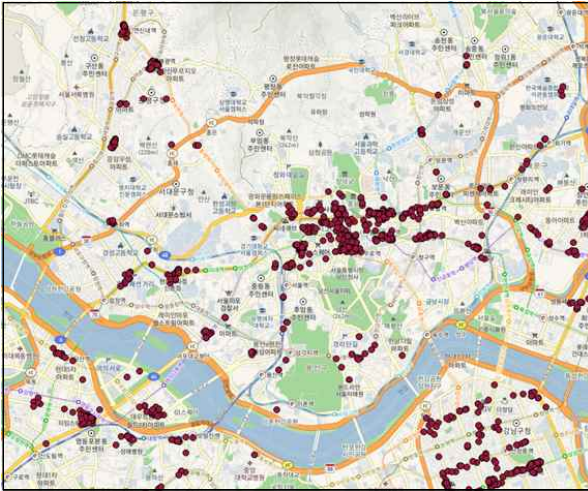
나. 분석 결과 산출

1) 총 인구당 여성 인구 비율



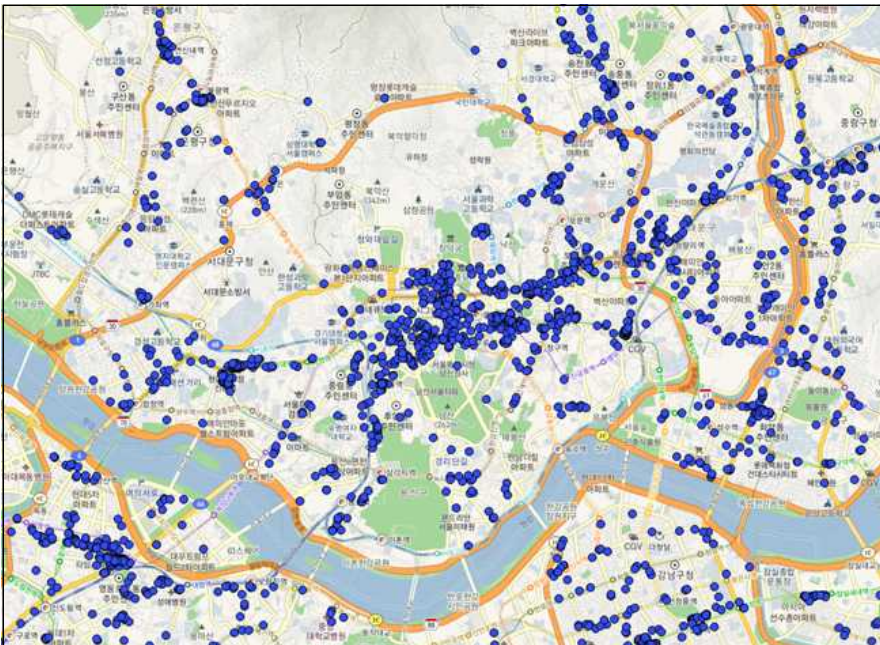
[그림 4-20] 총 인구당 여성 인구 비율 표시

2) 서울 유흥주점 인허가 건물 비율



[그림 4-21] 유흥주점 건물 표시 화면

3) 서울 숙박시설 비율



[그림 4-22] 숙박시설 건물 표시 화면

다. 시각화 구현

□ 위험도 순위 안에 위치한 공중화장실 시각화(아래 초록색 포인트가 공중화장실 위치를 의미합니다.)



[그림 4-23] 공중화장실 위험도 표시 화면

5. 활용 방안

가. 문제점 개선 방안

- 1) 빅데이터를 기반으로 범죄 취약 지역을 발굴했습니다.
- 2) 비상벨 우선 설치지역을 선정하여 예산 집행 효율성을 높일 수 있습니다.
- 3) 범죄에 노출된 지역에 거주하는 시민들의 불안감을 감소시킬 수 있습니다.

나. 업무 활용 방안

- 1) 개발된 위험 지도는 공중화장실 뿐 만 아니라, 기타 범죄 사각지대 및 취약

지역 골목길 등의 비상벨 설치 시에도 활용될 수 있습니다.

2) 치안과 관련된 국가 정책 개발 및 의사결정 과정에서 범죄 취약 지역을 제시하는 데에 활용될 수 있습니다.

3) 위험도가 높은 공중화장실에 긴급 비상벨을 우선적으로 설치함으로써, 효율적인 예산 집행과 더불어 사회 안전망 증대로 시민들의 높은 만족도를 기대할 수 있습니다.

[부록] -선택사항

1. 사례조사 한 내용

가. 참고 문헌

- ☐ 논문 - 「토지이용에 따른 도시범죄에 대한 연구」
- ☐ 논문 - 「서울의 범죄발생 특성과 안심도시 추진방안」
- ☐ 발의안 - 공중화장실 등에 관한 법률 일부개정법률안(김용판의원 발의: 2020. 9. 16.)

나. 전문가와의 인터뷰

- ☐ 대전경찰청 생활안전과 현직 전문가와의 인터뷰를 통해 범죄 관련 미디어 노출도와 비상벨 설치 요구도 간의 관계가 양의 상관관계에 있다는 사실을 알게 되어, 인터넷이나 뉴스, SNS 등 간접적으로 범죄 관련 미디어에 노출되었던 시민들은 그렇지 않은 시민들에 비해 비상벨 설치 민원 비율이 높아진다는 사실을 알게 되었습니다.

2. 주제설계를 위한 마인드맵 - 첨부 참조

3. 분석 상세코드 등