

불법촬영 등 범죄위험도 분석을 통한

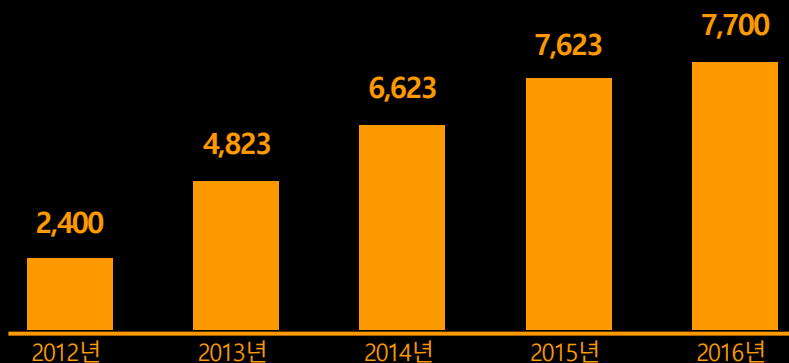
디지털 성범죄 예방서비스



경찰청
KOREAN NATIONAL POLICE AGENCY

스마트폰의 보급, 디지털 촬영기기 발달 등의 영향으로 몰래카메라 범죄는 매년 증가하고 있습니다

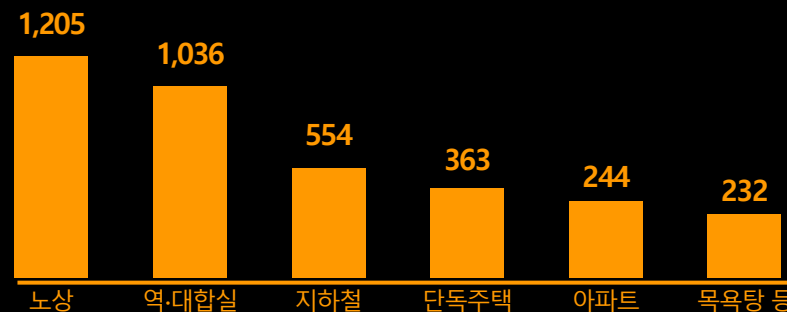
연간 디지털 성범죄 발생건수



※2016년은 추정치

자료출처 : 경찰청

불법촬영 주요 발생장소(2014년)



단속
어려움

피해구제
어려움

도덕적
해이초래

디지털 성범죄는 불법촬영, 제작, 유포 및 시청행위를 의미하며 이러한 성범죄 근절을 위해 정부는 종합 대책을 추진하고 있습니다

디지털 성범죄 유형 및 정부 종합대책

불법 영상물 촬영



본 과제 대상 범죄

불법 영상물 제작



불법 영상물 유포



불법 영상물 시청



몰카처벌 강화

SOS 성범죄(대응센터)

1. 변형카메라 수입과 판매업 등록제 도입, 스마트폰 무입촬영앱 다운 시 법정 경고 메시지 고지
2. 불법 영상물 신속 삭제 및 차단 : 기존 평균 10.8일에서 3일 이내로 긴급 차단 가능
3. **몰카 단속 수사 강화 및 일제 점검 실시 : 본 과제와 연계**
4. 몰카 가해자 처벌 강화 : 벌금형을 삭제하고 징역형으로만 처벌
5. 디지털 성범죄 피해자 지원 확대

몰래카메라 범죄 근절을 위해 케이티와 경찰청이 협력해 빅데이터 기반 범죄위험도를 개발해 경찰 업무에 적용합니다



범죄위험도 지표를 개발하기 위해서 경찰청 과학수사대의 범죄 프로파일러와 긴밀한 협조를 통해 과제를 수행했습니다

요구사항 도출

빅데이터 수집·설계

분석모델링

GIS 시각화 개발

본 연구과제에 진행을 위한 요구사항 도출 및 지하철역 단위 위험도 개발에 필요한 데이터 항목 및 분석 시나리오 정의

경찰청의 비정형 범죄데이터에 대한 텍스트마이닝 및 유동인구와의 융합 데이터 마트 구축

탐색적 데이터 분석을 통해 범죄데이터와 유동인구의 상관도를 파악하고 도출된 분석시나리오 기반의 분석모델링 수행

경찰청 원천에서 나오는 범죄데이터와 지하철역 범죄위험도 지수 등에 대한 GIS 시각화 설계를 통해 GeoPros 시각화 개발

분석 환경 구축

빅데이터 분석 서버 납품

탐색적 데이터 현황 분석

시각화 디멘전 정의

데이터 현황 분석

범죄데이터 텍스트마이닝

분석 가설 및 모델링 설계

화면 설계

요구사항 정의

유동인구 및 외부데이터 수집

지하철 위험도 모형 구축

시스템 개발

분석 시나리오 정의

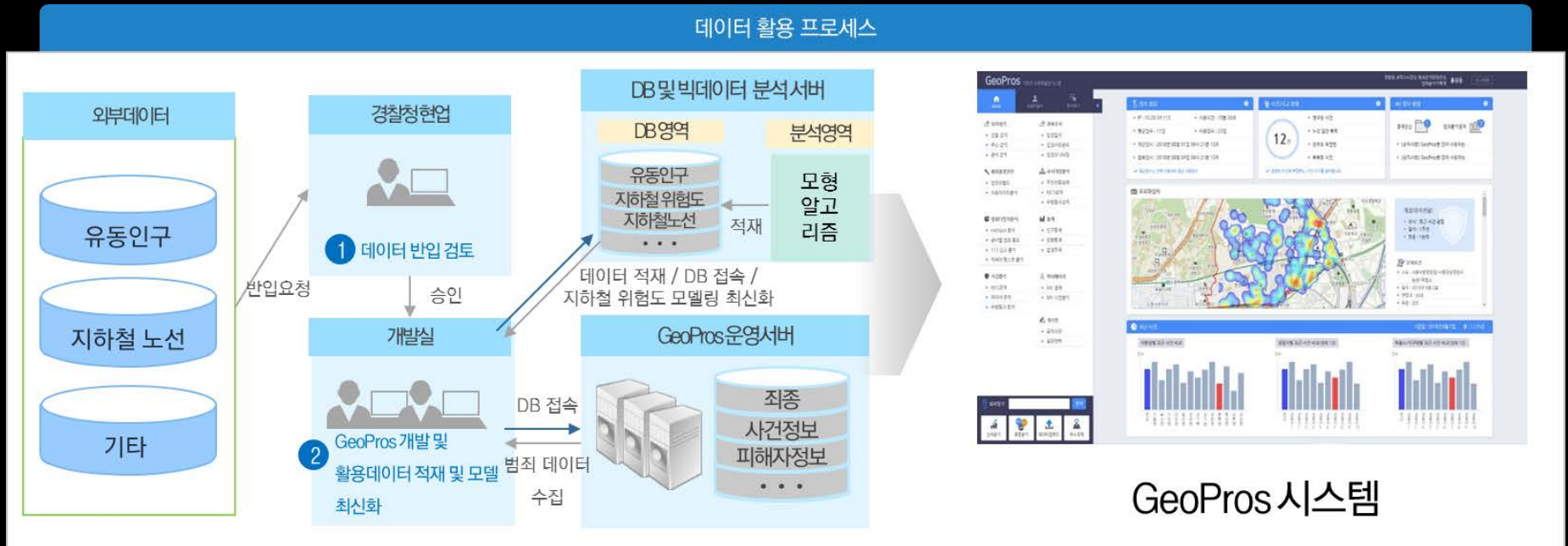
데이터 융합 MART 구성

가설 및 모델 검증

테스트 및 시험

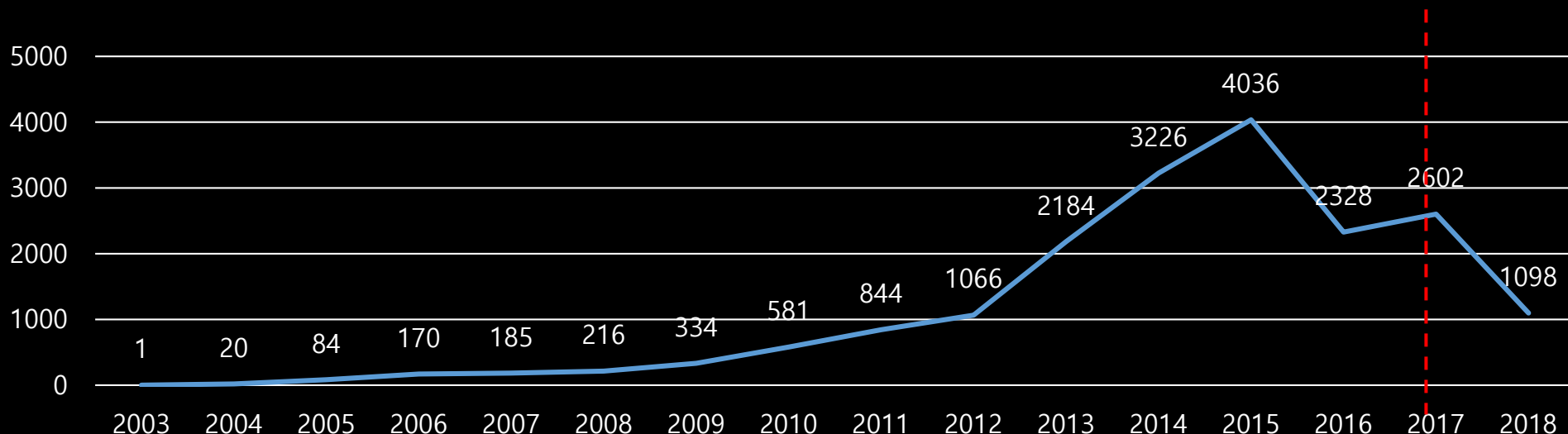
지표 개발의 공간적 범위는 수도권 지하철을 대상으로 데이터를 수집 분석해 경찰청 지리 프로파일링 시스템에 탑재합니다

구분	데이터명	데이터 설명	제공기관	수집방식	데이터유형	수급 상태
민간데이터	유동인구	지하철역/ 출구 기반의 유동인구 데이터	KT	오프라인	FILE(TXT)	수집완료
민간데이터	지하철역 배후지 유형	지하철역 반경 300M	오픈메이트	오프라인	FILE(TXT)	수집완료
공공 공개데이터	지하철 노선 데이터	수도권 지역의 지하철 노선 데이터	행정안전부	오프라인	FILE(SHP)	수집완료
공공기관데이터	사건정보 데이터	디지털 성범죄 사건정보 데이터	경찰청	온라인	FILE(TXT)	수집완료



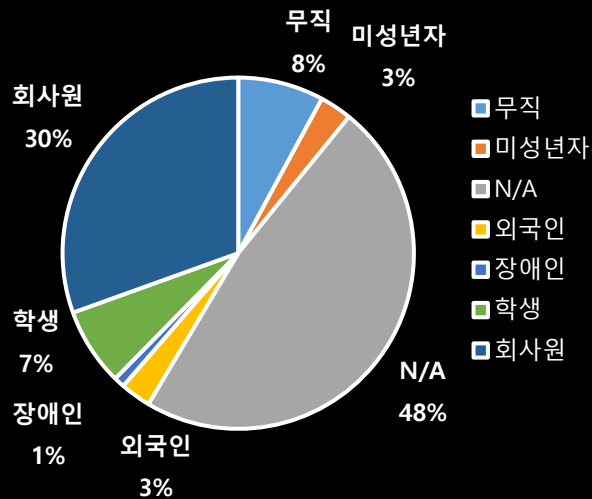
2003년부터 2018년까지 성범죄 데이터 약 18만건을 수집·정제해 지하철역 단위의 불법촬영 범죄 데이터 마트를 생성했습니다

성범죄 CASE 추출	사건내용 유무 확인	디지털 성범죄 CASE 추출	발생장소 매핑	지하철 유관데이터 추출	지하철역 구분	범죄발생시간 확인	결측치 제거
<ul style="list-style-type: none"> 대표접수죄명_대표검거죄명 중 하나라도 성폭력범죄로 구분되는 경우 정제 데이터 활용비율 : 0.72% 	<ul style="list-style-type: none"> 성범죄 CASE별 사건내용 없는 데이터 정제 정제데이터 활용비율 : 99.89% 	<ul style="list-style-type: none"> 대표접수죄명_대표검거죄명에서 통신매체이용음란카메라등이용촬영에 해당하는 데이터 추출 정제데이터 활용비율 : 73.61% 	<ul style="list-style-type: none"> 사건정보 중 건축사사무소, 검문소, 고속도로 등 지하철 범죄와 관련 없는 장소 제외 정제데이터 활용비율 : 60.43% 	<ul style="list-style-type: none"> 지하철에서 발생한 범죄 구분결과 지하철과 무관한 범죄 CASE 제외 정제데이터 활용비율 : 20% 	<ul style="list-style-type: none"> 지하철역명으로 매핑되는 데이터 추출 정제데이터 활용비율 : 66% 	<ul style="list-style-type: none"> 지하철역 단위 디지털 성범죄 사건들의 범죄발생시간 확인 정제데이터 활용비율 100% 	<ul style="list-style-type: none"> 추출된 지하철 데이터의 사건정보 점검을 통한 부적합 데이터 정제

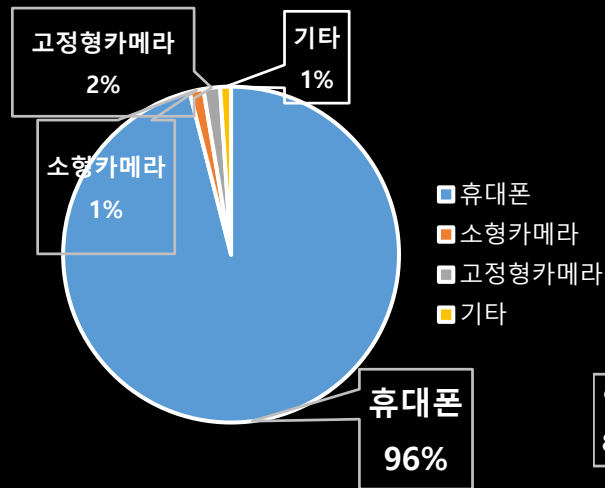


2017~2018년 기준으로 데이터를 추가로 정제하여 피의자 직업, 범행도구 및 범죄장소 데이터를 분류해 모델링에 반영했습니다

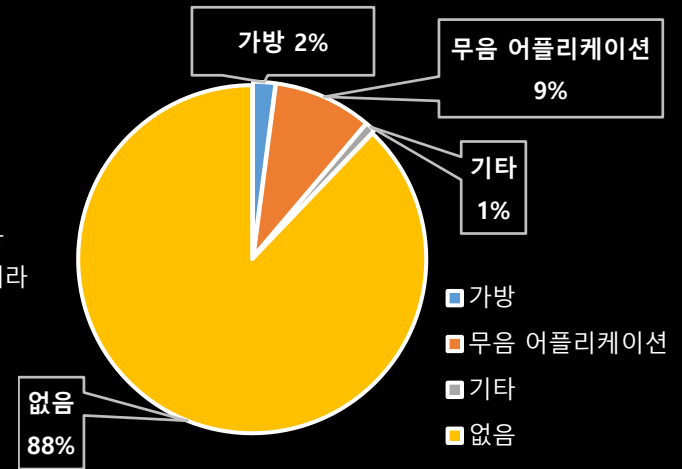
구분	AS-IS	TO-BE	구분
피의자 직업	-	- 무직,미성년자,외국인,장애인,대학생,회사원,N/A	GIS 시각화 부문 활용
피해자 도구	-	- 핸드폰, 소형카메라, 가방, 어플리케이션, 기타	GIS 시각화 부문 활용
피해장소	-	- 계단,노상,대합실,승강장,에스칼레이터,엘리베이터,화장실	GIS 시각화 부문 활용
피해출구	-	-지하철역 별 피해출구	출구단위 위험도 및 GIS 시각화 부문 활용



피의자 직업

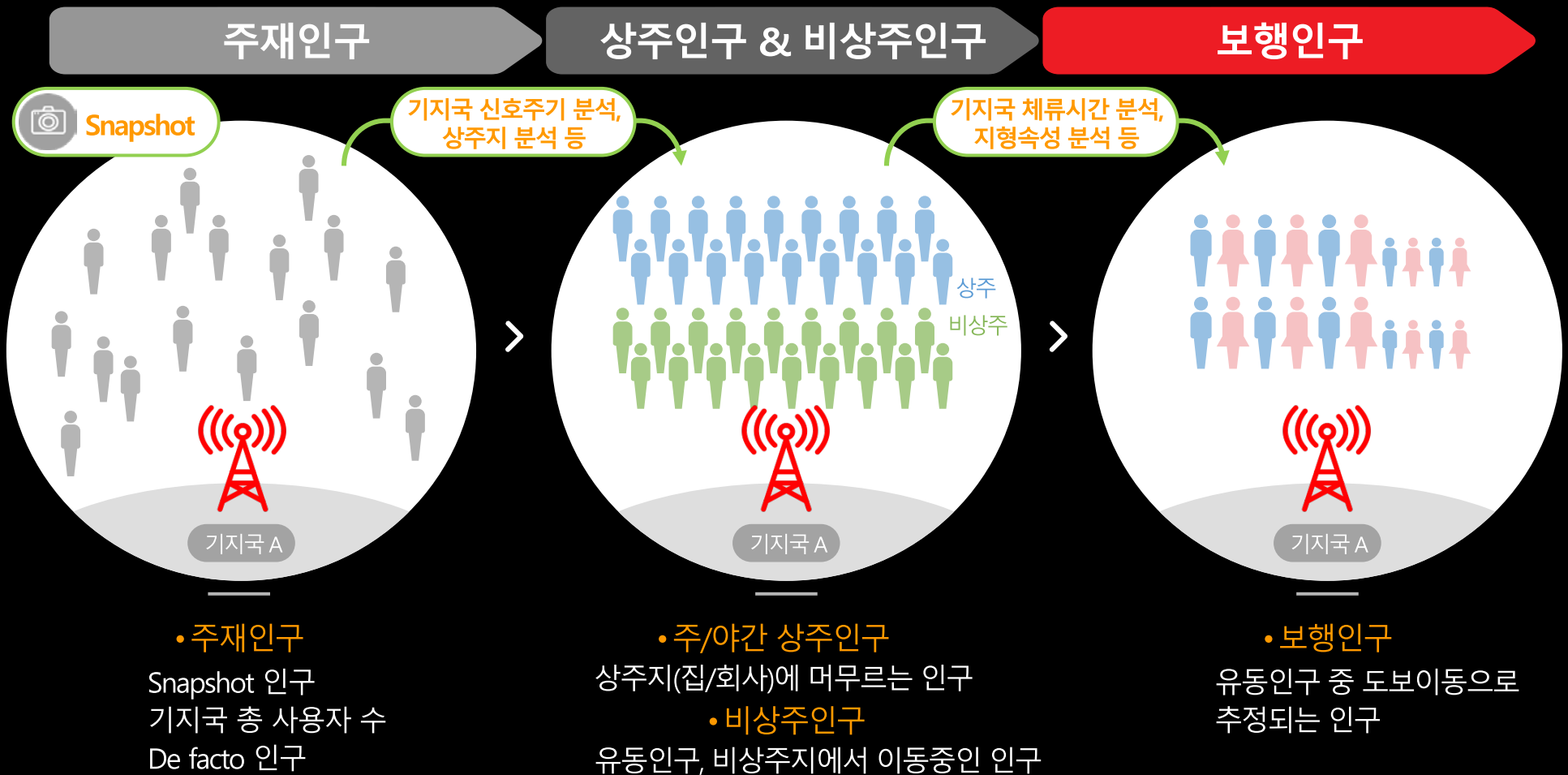


주범행도구



부범행도구

KT의 유동인구 데이터는 기지국에 수집되는 공간데이터를 기반으로 분석 목적에 따라 상주/비상주, 보행/비보행인구 등을 도출합니다

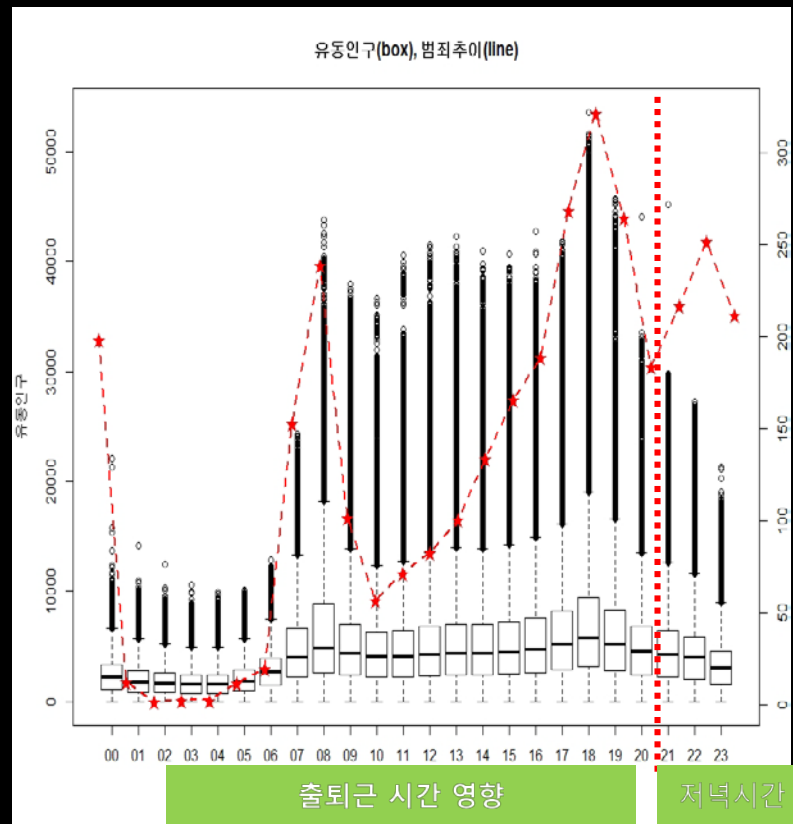


(기초분석 1/3) 범죄발생과 유동인구 간의 상관도 분석 결과

유동인구 데이터 분석

■ 지하철 / 지하철 출구별 유동인구 수집

구분	컬럼정보
지하철 별	지하철명
	지하철 출구명
	일
	시간
	성별
	연령별 (10세단위)
	인구수
	노선정보



분석결과

- 지하철역 주변 500m에 포함된 기지국을 통해 집계한 유동인구 데이터 확보

활용계획

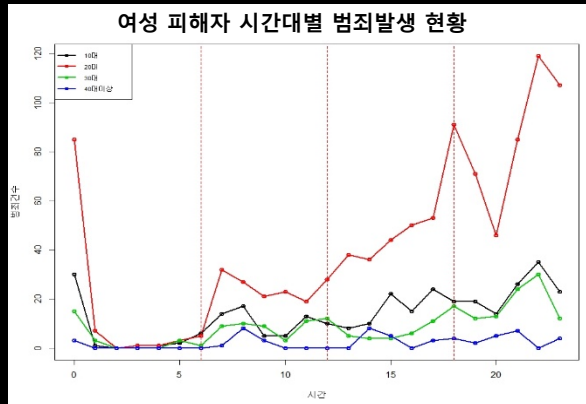
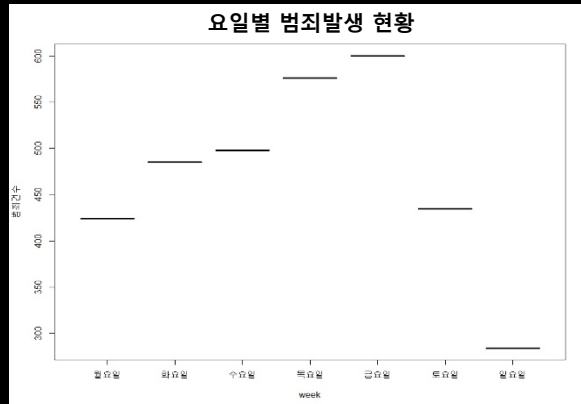
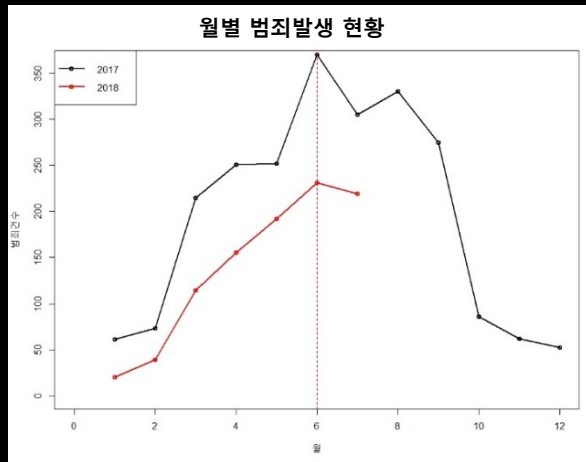
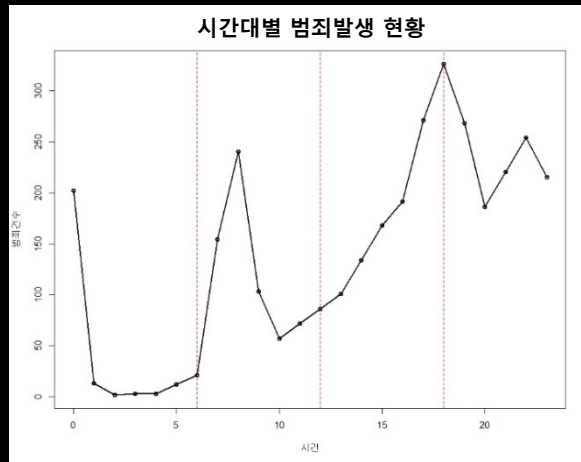
- 디지털 성범죄 데이터를 유동인구 데이터의 지하철명을 기준으로 융합 Mart를 구성하여 범죄발생 확률과 유동인구와의 상관분석을 진행함
- 상관계수가 높은 조건의 유동인구를 빅데이터 분석모형의 설명변수로 하여 지하철 위험도 결과 산출

(기초분석 2/3) 범죄 데이터 기초분석 결과

경찰청 사건정보 데이터 분석

분석결과

■ 지하철 노선기반 디지털 성범죄 범죄 추이



- 비정형데이터의 텍스트 마이닝을 통한 지하철역 단위 범죄데이터 추출
- 경찰청 내부 데이터 정제 결과 검토 후 추출 정확성 입증
- 시간대는 주로 평일 출퇴근 시간에 범죄가 급증하며, 여름에 디지털 성범죄가 많이 발생함

활용계획

- 유동인구 데이터와의 상관분석을 통한 위험도 개발
- GeoPros 지하철 성범죄 대시보드의 범죄 현황 차트 개발

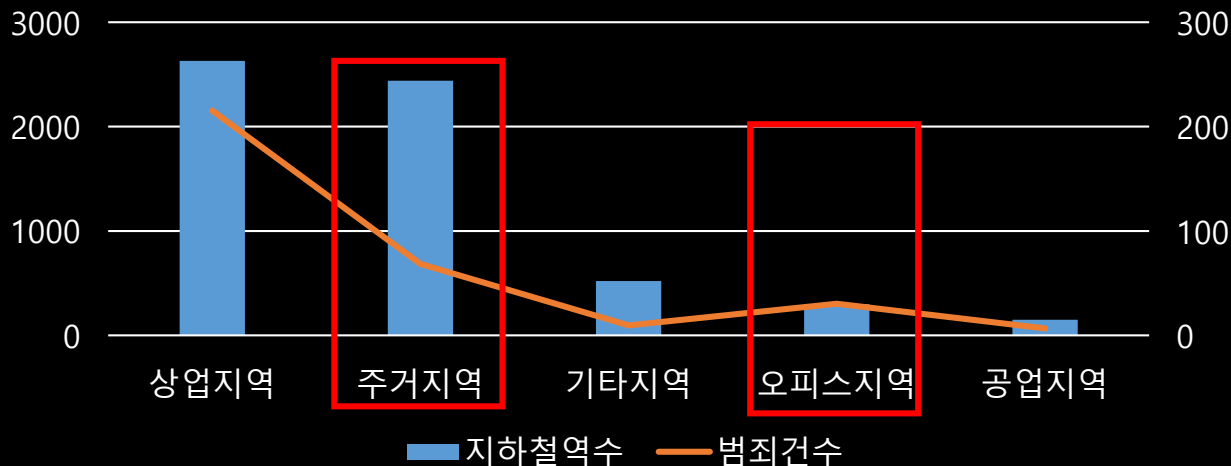
(기초분석 3/3) 지하철역 배후지 유형 기초분석 결과

배후지 유형 분석

■ 5개의 지역구분을 통한 유형분류

순위	지역구분	지하철역 수
1	상업지역	263
2	주거지역	244
3	기타지역	52
4	오피스지역	30
5	공업지역	15

순위	지역구분	범죄건수
1	상업지역	2,153
2	주거지역	685
3	오피스지역	302
4	기타지역	95
5	공업지역	67



분석결과

- 지하철역 단위 유형분류를 통해 범죄데이터를 분석한 결과 상업지역과 주거지역에 포함된 지하철역 수가 가장 많음
- 유형별 범죄건수 순위를 확인하면 주거지역이 지하철역 수를 포함하는 것 대비 범죄건수가 낮게 나타나며, 오피스지역은 지하철역 수를 포함하는 것 대비 범죄건수가 매우 높게 나타남

활용 계획

- 지하철역 유형화를 통해 범죄가 상업지역과 오피스지역에 집중되는 현황을 연령대별 피해자 데이터를 결합하여 향후 분석모형의 주요변수로 활용할 계획임

배후지 유형 설명

- 기타지역 : 일반적으로 사람이 주로 생활하는 지역이 아닌 학교, 하천, 산지, 문화시설 등 특수한 성격을 갖는 지역
- 배후지 구분 기준 : 지하철역 반경 300M

몰래카메라 범죄가 발생하는 상황에 대한 가설을 수립한 후 분석을 통해 가설을 검증하여 모델링에 필요한 변수를 확정합니다

번호	관점	가설	결과	IV
1	피해자	평소보다 20대 여성이 많은 경우 발생가능성이 높을 것이다.	기각	0.08
2	피해자	평소보다 30대 여성이 많은 경우 발생가능성이 높을 것이다.	기각	0.10
3	피해자	평소보다 20/30대 여성이 많은 경우 발생가능성이 높을 것이다.	기각	0.09
4	피해자	다른 연령대보다 20대 여성이 많은 역에서 발생가능성이 높을 것이다.	채택	0.97
5	피해자	다른 연령대보다 30대 여성이 많은 역에서 발생가능성이 높을 것이다.	채택	0.22
6	피해자	다른 연령대보다 20/30대 여성이 많은 역에서 발생가능성이 높을 것이다.	채택	0.12
7	피의자	평소보다 10대 남성이 많은 경우 발생가능성이 높을 것이다.	기각	0.05
8	피의자	평소보다 20대 남성이 많은 경우 발생가능성이 높을 것이다.	기각	0.07
9	피의자	평소보다 30대 남성이 많은 경우 발생가능성이 높을 것이다.	기각	0.12
10	피의자	평소보다 40대 남성이 많은 경우 발생가능성이 높을 것이다.	기각	0.10
11	피의자	평소보다 10/20/30/40대 남성이 많은 경우 발생가능성이 높을 것이다.	기각	0.08
12	피의자	다른 연령대보다 10대 남성이 많은 역에서 발생가능성이 높을 것이다.	기각	0.34
13	피의자	다른 연령대보다 20대 남성이 많은 역에서 발생가능성이 높을 것이다.	채택	0.09
14	피의자	다른 연령대보다 30대 남성이 많은 역에서 발생가능성이 높을 것이다.	채택	0.13
15	피의자	다른 연령대보다 40대 남성이 많은 역에서 발생가능성이 높을 것이다.	기각	0.91
16	피의자	다른 연령대보다 10/20/30/40대 남성이 많은 역에서 발생가능성이 높을 것이다.	채택	0.15
17	복합	평소보다 20/30대 여성, 10/20/30/40대 남성이 많은 경우 발생가능성이 높을 것이다.	기각	0.08
18	복합	다른 연령대보다 20/30대 여성, 10/20/30/40대 남성이 많은 역에서 발생가능성이 높을 것이다.	채택	0.67
19	복합	피해연령대(20/30대 여성) / 가해연령대(10/20/30/40대 남성) 비율에 따라 발생가능성이 다를 것이다.	채택	0.62
20	혼잡도	시간대별 혼잡도에 따라 범죄 발생가능성이 다를 것이다.	채택	1.55
21	시간/요일	시간대 / 요일에 따른 발생 가능성이 다를 것이다.	채택	1.05
22	지하철역	지하철역 유형별 범죄 발생가능성이 다를 것이다.	채택	0.63
23	배후지	배후지 지역에 따라 범죄 발생 가능성이 다를 것이다.	채택	0.14
24	유동인구	인구수 증감에 따른 범죄 발생 가능성이 다를 것이다.	채택	1.10
25	호선	지하철 호선에 따라 범죄 발생 가능성이 다를 것이다.	채택	0.25
26	시간대	시간 / 지하철 호선에 따라 범죄 발생 가능성이 다를 것이다.	채택	1.15
27	계절	월별 범죄 발생 가능성이 다를 것이다.	채택	0.21
28	환승역	환승 속성에 따라 범죄 발생 가능성이 다를 것이다.	채택	0.70

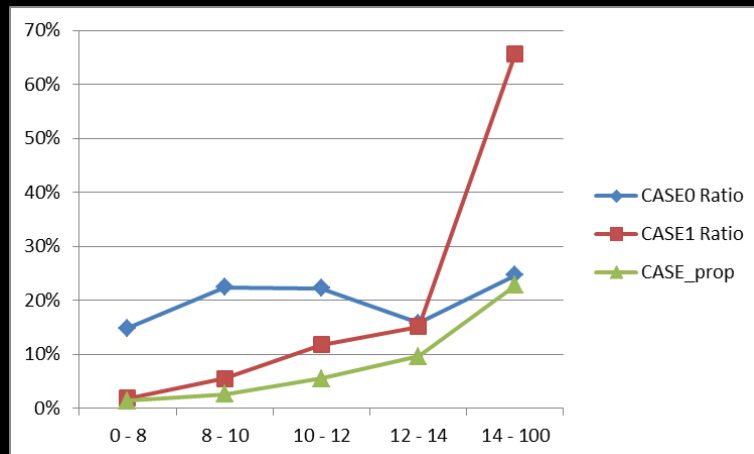
(검증결과 예시 1/4) 다른 연령대보다 20대 여성이 많은 역에서 발생가능성이 높을 것이다 ▶ 가설에 부합

20대여성 유동인구/전체 유동인구 % 반환

총인구 성연령 비율

F00	F10	F20	F30	F40	F50	F60	F70	M00	M10	M20	M30	M40	M50	M60	M70
0.09	2.15	13.32	10.12	8.01	8.09	4.22	1.73	0.06	2.00	10.60	11.96	10.31	9.29	5.74	2.30

유동인구		CASE_0	CASE_1	Total_row	CASE0 Ratio	CASE1 Ratio	CASE_prop	WOE	IV
lower	upper								
0	8	3,521	49	3,570	15%	2%	1.37%	207.746	0.270
8	10	5,298	145	5,443	22%	6%	2.66%	140.113	0.236
10	12	5,264	310	5,574	22%	12%	5.56%	63.485	0.066
12	14	3,736	398	4,134	16%	15%	9.63%	4.209	0.000
14	100	5,842	1,727	7,569	25%	66%	22.82%	-97.854	0.401
TOTAL									0.975



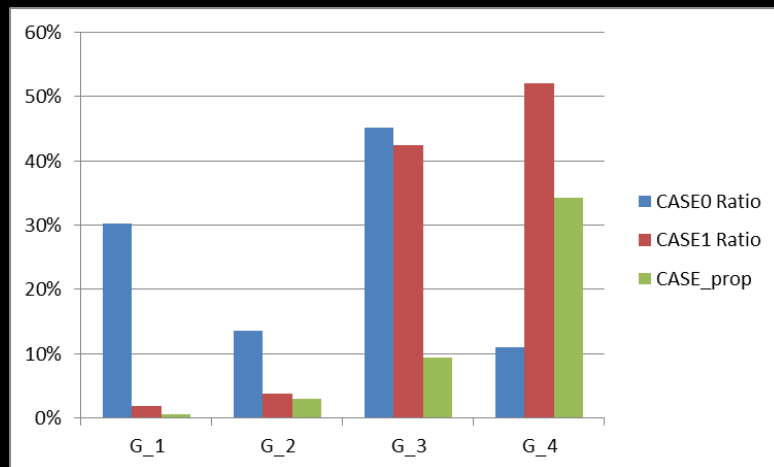
검증 결과

1. 20대 여성이 많을수록 범죄 발생률이 단조 증가
2. 20대 여성 유동인구의 비율이 경우 14% 이상 구간에서 66%의 범죄가 발생
3. 14% 이상 구간의 발생률은 8%미만 지역의 20배.
4. 20대 여성이 시간과 장소에서 범죄 발생 비율이 증가하는 것으로 나타나며, 해당 가설에 부합한다고 판단됨

(검증결과 예시 2/4) 시간대별 혼잡도에 따라 범죄 발생률의 차이가 있을 것이다 ▶ 가설에 부합

혼잡도 : 유동인구 / 시간대별 평균 유동인구(시간대별 평균 유동인구 대비 변화량)

그룹	CASE_0	CASE_1	Total_row	CASE0 Ratio	CASE1 Ratio	CASE_prop	WOE	IV
G_1	7,156	49	7,205	30%	2%	0.68%	278.6662	0.7909
G_2	3,212	100	3,312	14%	4%	3.02%	127.2254	0.1243
G_3	10,670	1,113	11,783	45%	42%	9.45%	6.315242	0.0017
G_4	2,623	1,367	3,990	11%	52%	34.26%	-154.552	0.6323
TOTAL							1.5492	



검증 방법

1. 혼잡도와 시간대의 cross table에서 범죄 발생률을 기준으로 4개의 그룹으로 그룹화
2. $G_1(0, 0.02] < G_2(0.02, 0.04] < G_3(0.04, 0.20] < G_4(0.20, 0.48]$ 의 순서로 범죄 발생 비율, 혼잡도가 증가하는 형태로 그룹화됨.

(검증결과 예시 2/4) 시간대별 혼잡도에 따라 범죄 발생률의 차이가 있을 것이다 ► 가설에 부합

각 구간별 범죄발생률 (4개의 구간으로 그룹화)

시간대	(0,20]	(20,50]	(50,100]	(100,140]	(140,Inf]	(0,20]	(20,50]	(50,100]	(100,140]	(140,Inf]
time_00	3%	9%	11%	13%	25%	G_2	G_3	G_3	G_3	G_4
time_01	1%	1%	0%	1%	3%	G_1	G_1	G_1	G_1	G_2
time_02	0%	0%	0%	0%	0%	G_1	G_1	G_1	G_1	G_1
time_03	0%	0%	0%	0%	1%	G_1	G_1	G_1	G_1	G_1
time_04	0%	0%	0%	0%	1%	G_1	G_1	G_1	G_1	G_1
time_05	0%	0%	2%	1%	1%	G_1	G_1	G_1	G_1	G_1
time_06	0%	0%	1%	4%	2%	G_1	G_1	G_1	G_2	G_1
time_07	0%	3%	5%	8%	32%	G_1	G_2	G_3	G_3	G_4
time_08	3%	2%	9%	19%	45%	G_2	G_1	G_3	G_3	G_4
time_09	0%	2%	3%	12%	17%	G_1	G_2	G_2	G_3	G_3
time_10	0%	1%	2%	5%	10%	G_1	G_1	G_2	G_3	G_3
time_11	1%	3%	4%	5%	10%	G_1	G_2	G_3	G_3	G_3
time_12	3%	3%	5%	6%	14%	G_2	G_2	G_3	G_3	G_3
time_13	2%	2%	4%	2%	24%	G_2	G_1	G_3	G_2	G_4
time_14	3%	2%	6%	6%	26%	G_2	G_1	G_3	G_3	G_4
time_15	1%	4%	8%	11%	29%	G_1	G_2	G_3	G_3	G_4
time_16	7%	5%	6%	14%	35%	G_3	G_3	G_3	G_3	G_4
time_17	4%	5%	9%	16%	47%	G_3	G_3	G_3	G_3	G_4
time_18	2%	6%	13%	24%	47%	G_1	G_3	G_3	G_4	G_4
time_19	0%	8%	10%	20%	44%	G_1	G_3	G_3	G_3	G_4
time_20	3%	3%	8%	10%	33%	G_2	G_2	G_3	G_3	G_4
time_21	0%	6%	10%	17%	36%	G_1	G_3	G_3	G_3	G_4
time_22	2%	7%	12%	22%	37%	G_1	G_3	G_3	G_4	G_4
time_23	4%	5%	10%	15%	31%	G_2	G_3	G_3	G_3	G_4

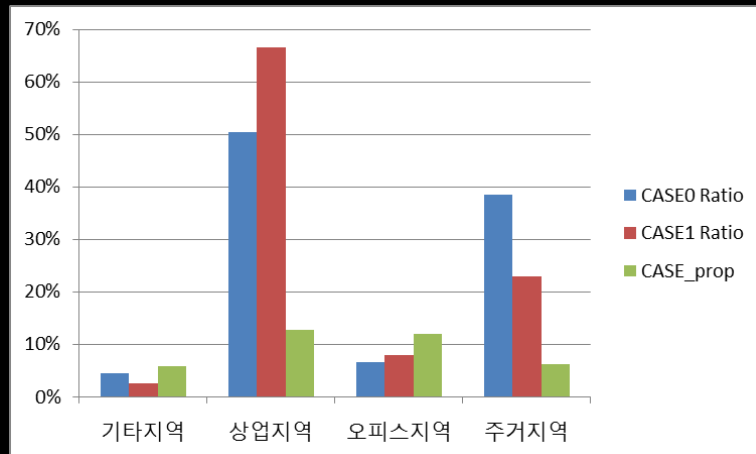
검증 결과

1. 지하철이 다니지 않는 새벽 시간대와 혼잡하지 않은 상황에서 발생 가능성이 적게 나타남.
2. 주로 출퇴근 시간이 몰리는 08,17,18,19시간대에 높은 혼잡도와 범죄 가능성이 나타남.
3. 01~06시간대를 제외하고 시간대 별 혼잡도에 따라 범죄 발생 가능성이 다를 것이라는 가설에 부합됨

(검증결과 예시 3/4) 지하철역의 배후지 속성에 따라 범죄 발생 가능성이 다를 것이다 ▶ 가설에 부합

※ 기타지역 = 기타지역 + 공업지역

배후지	CASE_0	CASE_1	Total_row	CASE0 Ratio	CASE1 Ratio	CASE_prop	WOE	IV
기타지역	1,053	66	1,119	4%	3%	5.90%	57.252	0.0111
상업지역	11,929	1,750	13,679	50%	67%	12.79%	-27.787	0.0449
오피스지역	1,550	209	1,759	7%	8%	11.88%	-19.355	0.0027
주거지역	9,129	604	9,733	39%	23%	6.21%	51.841	0.0809
TOTAL							0.1396	

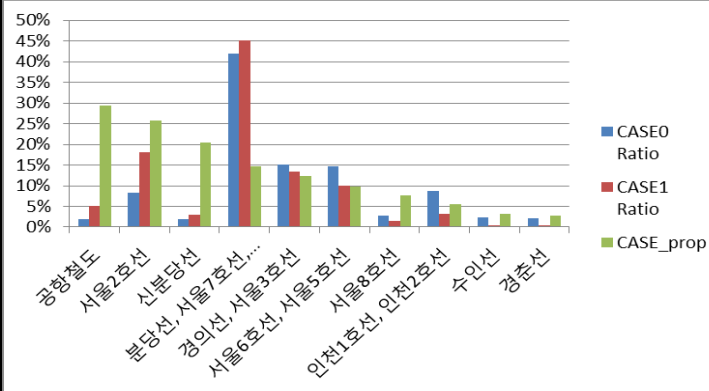


검증 결과

1. 유동인구가 많은 상업지, 오피스 지역에서 범죄 가능성이 높게 나타남.
2. 오피스 지역의 경우 적은 비율을 차지하지만 상업 지역과 비슷한 범죄 발생률이 나타남.
3. 배후지 별로 범죄 발생률이 차이가 있을 것 이라는 가설에 부합됨

(검증결과 예시 4/4) 지하철 노선에 따라 범죄 발생률이 다를 것이다 ▶ 가설에 부합

노선	CASE_0	CASE_1	Total_row	CASE0 Ratio	CASE1 Ratio	CASE_prop	WOE	IV
공항철도	543	227	770	2%	5%	29.48%	-95.66671	0.02977
서울2호선	2,342	809	3,151	8%	18%	25.67%	-76.58647	0.07379
신분당선	540	138	678	2%	3%	20.35%	-46.45109	0.00530
분당선, 서울1호선(경인선), 서울4호선, 서울7호선, 서울9호선	11,756	2,024	13,780	42%	45%	14.69%	-6.95384	0.00210
경의선, 서울3호선	4,228	599	4,827	15%	13%	12.41%	12.53964	0.00223
서울6호선, 서울5호선	4,109	449	4,558	15%	10%	9.85%	38.50857	0.01808
서울8호선	800	67	867	3%	1%	7.73%	65.10927	0.00891
인천1호선, 인천2호선	2,418	142	2,560	9%	3%	5.55%	100.60426	0.05516
수인선	645	21	666	2%	0%	3.15%	159.59015	0.02934
경춘선	595	17	612	2%	0%	2.78%	172.65217	0.03019
TOTAL								0.25486



검증 결과

1. 분당선, 1호선, 4호선, 7호선, 9호선의 경우 가장 많은 지하철 역을 포함하고 범죄 발생률도 높게 나타남.
2. 공항철도, 신분당선 노선의 경우 지하철역을 많이 포함하고 있지 않지만, 홍대입구, 서울역, 강남역 등을 포함하고 있어 적은 지하철역을 포함하는 것 대비 범죄 발생률이 높게 나타남.
3. 지하철 호선에 따른 발생률이 차이가 있을 것 이라는 가설에 부합됨

가설 검증 결과를 기반으로 변수를 선정하고 노선별/지하철역별/출구별로 5단계의 위험도 지표를 개발합니다

디지털 성범죄
발생위험도

=

범죄 프로파일러와 공동으로 작업한 가설 검증 결과를
기반으로 최종 변수를 선정해 모델링에 반영

배후지	월/요일	유동인구	환승	지하철노선
피해/피의자	혼잡도/시간	F20/30	M1234	...



지하철 노선별/역별/출구별 위험도

1. 월별/요일별/시간대 별로 위험도 범죄위험도 지표 표시
2. 지하철 노선별/역별/출구별 위험도 표시
3. 위험도는 총 5단계로 구분
4. 위험도에 근거한 순찰 지역 추천 등 서비스 제공

경찰청 과학수사대에서 운영 중인 지리 프로파일링 시스템인 GeoPros에 디지털 성범죄 위험도를 탑재해 단속업무에 활용합니다

GeoPros GIS 시각화 기능

지하철 위험현황 대시보드

개요

지하철역
위험도 TOP5

주요 범죄발생
장소

요일/시간대별
범죄발생현황

피해자 연령대

피의자 연령대

범죄발생 추이

피의자 프로파일

피의자 연령대별
가해장소

범죄발생 미검거
지하철역 TOP10

범행도구

피의자 직업

내/외국인 범죄
비율

외국인 국적별
범죄비율

피해자 프로파일

연령대별
피해장소

연령대별
범죄발생 TOP5

시간대별 범죄발생
/유동인구 추이

연령대 별 요일/
시간대 범죄현황

피해장소 프로파일

장소별 시간대 범죄발생 현황

지하철역 위험도

지역/ 노선별 지하철역 위험도

지하철역명

지하철역 위험도

지하철역
위험출구

위험출구의
평균 유동인구

지하철역
범죄발생 건수

지하철역
범죄발생 다발
시간대

주요 범죄발생
장소

순찰지역 추천하기

지역/ 노선별 우선 순찰 지하철역 TOP5

우선 순찰 지하철역 상세현황

본 과제를 통해 개발된 범죄위험도 지표는 생활안전지도 및 KT의 BigSight를 통해 대국민 정보 서비스로서 제공될 예정입니다

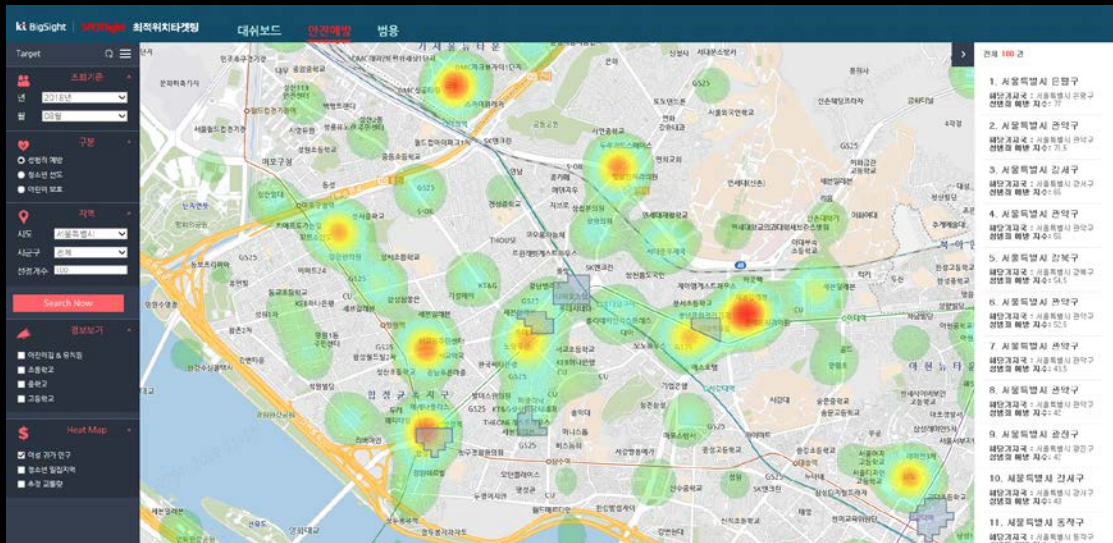


행정안전부 생활안전지도 서비스

- 8대 안전분야에 대한 정보
- 교통, 재난, 치안, 맞춤, 시설, 보건, 산업, 사고 안전분야
- 치안 안전 분야에 디지털 성범죄 카테고리 제공 예정

Kt BigSight 안심맵 서비스

- KT의 빅데이터 포털을 통해 범죄위험도 지표 공개
- 지하철 노선별/역별/출구별 위험도 지표 제공
- 청소년, 어린이 안심 서비스 추가



**KT와 경찰청은 빅데이터로
안전한 사회를 만드는데 기여하겠습니다.**



경찰청
KOREAN NATIONAL POLICE AGENCY