

물리환경이 112 신고 발생에 미치는 영향 연구

Effects of Physical Environment on 112 Emergency Reporting

안의순 Ahn, Euisoon
김은희 Kim, Eunhee
김영우 Kim, Youngwoo
임리사 Lim, Lisa
김유진 Kim, Yujin
이정현 Lee, Joungyun

(a u r . i

물리환경이 112신고 발생에 미치는 영향 연구

Effects of Physical Environment on 112 Emergency Reporting

지은이 안의순, 김영우, 김은희, 임리사, 김유진, 이정현
펴낸곳 건축공간연구원
출판등록 제2015-41호 (등록일 '08. 02. 18.)
인쇄 2025년 12월 27일, 발행: 2025년 12월 31일
주소 세종특별자치시 가림로 143, 8층
전화 044-417-9600
팩스 044-417-9608

<http://www.auri.re.kr>

가격: 10,000원, ISBN: 979-11-5659-539-7

연구진

연구책임 안의순 부연구위원

연구진 김영우 부연구위원
 김은희 연구위원

외부연구진 임리사 교수(한국과학기술원)
 김유진
 이정현

연구심의위원 오성훈 선임연구위원
 염철호 선임연구위원
 김상호 선임연구위원
 서수정 선임연구위원
 강석진 경상국립대학교 건축학과 교수
 이경환 공주대학교 도시·교통공학과 교수

제1장 서론

범죄두려움은 범죄 피해에 대한 개인의 주관적 불안감과 우려 등 심리적 현상으로 범죄 발생과 별개 문제로 관리할 필요가 있다. 범죄두려움은 특히 개인의 사회적 활동 참여를 감소시키면서 CPTED 원칙 중 자연적 감시, 접근 통제, 영역성 강화, 활동 지원 등의 효과적 적용을 저해하는 요인으로 작용하고 안전한 도시 환경 조성에 부정적 영향을 미칠 수 있다.

112신고는 범죄가 발생했다고 인식한 시민이 긴급 출동을 요청하는 행위로 구체적 행동을 수반하기 때문에 범죄 발생에 대한 행동으로 나타난 우려를 의미한다. 따라서 112신고 데이터는 범죄두려움과 관련된 정책 수립에 필수적인 실증적 근거를 제공할 수 있다.

본 연구의 목적은 건축·도시 공간의 물리적 환경요소와 112신고 발생 사이의 연계성을 파악하고 물리환경의 특성이 112신고 발생에 미치는 영향을 데이터 기반으로 실증하는 데에 있다. 112신고 데이터를 활용한 연구를 통하여 범죄예방뿐만 아니라 범죄두려움 저감에 대해서도 행태적 근거를 제공하고자 하였다.

본 연구의 내용적 범위는 물리환경이 112신고 발생에 미치는 영향을 실증하기 위하여 112신고 제도 및 활용 분석, 112신고 데이터 기술통계 분석, 마지막으로 물리환경 영향 실증 분석의 단계로 진행된다.

공간적 범위는 물리환경 특성의 영향을 다양한 환경에서 실증하기 위하여 본 연구의 공간적 범위를 서울특별시 전역으로 설정하였다. 전년도 연구에서 서울 내 6개 구 대상으로 물리환경이 범죄 발생에 미치는 영향을 연구하였는데 이를 계승하고 확장할 수 있도록 하였다. 연구의 시간적 범위는 2023년 1년간으로 설정하였다.

본 연구에서는 국내·외 긴급 신고 데이터 활용 방안을 검토하기 위하여 법·제도, 국내·외 사례, 연구문헌, 언론보도 등을 대상으로 문헌연구를 진행하였다. 또한 112신고에 대한 영향 분석을 위한 물리환경 변수 선정에도 선행연구와 CPTED 체크리스트 등 다양한 문헌연구 결과가 반영되었다.

건축물과 도시공간의 물리환경 특성, 112신고 발생 등에 대하여 격자 단위 공간정보화, 기초 통계 분석 및 GIS를 활용한 공간분석을 수행하였다. 112신고 데이터와 도시건축환경 요소 간의 관계를 분석하기 위하여 공간회귀모형과 기계학습 모형을 다수 검토하여 최종적으로 공간회귀에서는 다중 범위 지리 가중 회귀(MGWR), 기계학습에서는 XGBoost 등 두 가지 주요 예측 모델을 활용하였다.

제2장 112신고 제도 및 활용

2024년 「112신고의 운영 및 처리에 관한 법률」(약칭 「112신고처리법」)이 제정되면서 경찰청 예규에 의해 운영되어 온 종전보다 한층 통합적이고 신속한 대응체계가 확립되었다. 또한 신고자 보호와 정보 관리 강화, 전문 인력의 양성 및 첨단 정보시스템 구축 등 세부적인 운영 체계를 마련함으로써 국민의 생명과 재산 보호라는 국가의 책무를 실질적으로 이행할 수 있는 여건을 마련하였다.

「112신고처리법」은 112 시스템을 구축 및 운영하고 112 시스템에 입력된 신고 접수·처리 상황은 3년간 기록·보존하도록 규정하고 있다. 112신고 처리를 위해 필요한 경우 관계 기관의 정보시스템과 112 시스템의 연계를 요청할 수 있다. 또한, 112신고 통계 현황을 정기적으로 분석하고 범죄 예방 대책 수립 등 치안 활동에 반영하여야 한다. 112신고자 정보는 해당 112 신고의 처리를 위해 활용하거나 신고자가 동의하는 경우 등에 한하여 수집·이용 또는 제공될 수 있고 그 외 목적의 이용은 금지된다.

112신고 코드 분류 체계는 긴급성과 현장 출동 필요성에 따라 코드 0부터 코드 4까지 다섯 단계 코드로 분류한다. 코드 0은 가장 긴급한 상황으로 실시간 전파 및 즉각적인 현장 출동이 필수적이며 반대로 코드 4는 긴급성이 없는 민원·상담 신고를 의미한다.

국내·외 긴급 신고 데이터는 범죄예방 정책 수립을 위한 근거로 폭넓게 활용되고 있다. 미국 CompStat 시스템 등 국외 신고 데이터 활용 사례는 데이터 기반의 범죄 예방 정책이 범죄 억제와 효율적인 경찰력 배치에 크게 기여하는 결과를 보였다. 국내에서도 프리카스(Pre-CAS) 시스템을 통하여 112신고 데이터를 기반으로 범죄 위험도 예측과 전략적 순찰 배치를 수행하고 있다.

그러나 「112신고처리법」은 신고자 개인정보 보호와 정보 이용의 최소화 원칙을 강력히 규정하면서 연구 목적의 데이터 제공 근거를 마련하지 않아 112신고 데이터를 범죄예방 정책 수립과 연구 목적으로 적극 활용하는 데 어려움이 있는 현실이다. 향후 범죄예방 정책 및 연구 목적으로 112신고 데이터를 효과적으로 활용하기 위해서는 개인정보 보호라는 법적 가치를 유지하면서도 연구적 활용의 길을 열어주는 제도적 개선이 필요하다.

제3장 112신고 발생 기술통계 분석

분석 대상은 2023년 1년간 서울특별시 경계 내에 속한 112신고 데이터이다. 공간정보는 격자 단위 집계를 통하여 익명화된 데이터를 사용하여 신고자 정보가 보호될 수 있도록 하였다. 500m 단위 정사각형 격자를 설정하고 각 격자 셀의 중심점이 서울시 경계 내에 들어오는지 여부를 검사하여 최종 분석 대상을 선별하였다. 신고 데이터는 최종적으로 발생 일자 및 시각, 긴급코드(코드 0~코드 4), 500m 격자 단위 위치로 정제하였다. 이상치 탐지 분석을 통하여 데이터 내에 특별한 이상치가 없음을 확인하고 전체 데이터를 분석에 사용하였다.

서울 지역 112신고 데이터를 긴급코드별로 분석한 결과, 대다수가 코드 2(잠재적 위험)와 코드 1(출동 필요)에 해당함을 알 수 있었다. 시간대별 분석 결과 전체 112신고는 야간으로 갈수록 증가하다가 새벽에 감소하는 패턴을 보였다. 그러나 각 긴급코드별로 시간대별 분포를 분석한 결과, 코드별로 시계열적인 패턴이 구분된다는 점을 확인할 수 있었다. 긴급 신고에 해당하는 코드 0, 코드 1 그리고 정의상 비긴급 신고이기는 하나 즉시 출동이 요구되는 코드 2 신고는 주말과 야간에 신고 발생이 집중되는 패턴이 유사하게 나타났다. 반면 비긴급 신고인 코드 3, 민원 상담에 해당하는 코드 4 신고는 평일 업무 시간대에 집중되는 패턴을 보였다. 본 연구에서는 긴급 신고라는 행태적 지표를 통해 파악할 수 있는 범죄 발생 및 범죄두려움에 초점을 맞추기 위해 발생 패턴을 기준으로 긴급 신고에 해당하는 코드 0에서 코드 2를 실증 분석의 대상으로 설정하였다.

112신고 발생 패턴의 성격을 추가로 확인하기 위하여 112신고 발생(코드 0~코드 2)과 실제 범죄 발생(5대 범죄)의 공간적 패턴을 서울시 강남·서초·송파·강동 4개 구를 대상으로 분석한 결과 상관계수가 0.854로 나타나 112신고 발생과 범죄 발생 사이에 공간적 패턴이 일치함을 확인하였다.

제4장 물리환경이 112신고 발생에 미치는 영향 실증 분석

실제 건축 및 도시 환경에서 물리환경의 특성이 112신고 발생에 미치는 영향을 분석하기 위하여 기계학습과 공간회귀 모델을 활용한 공간데이터 분석을 수행하였다. 분석 대상은 2023년 1년간 서울특별시에서 발생한 112신고 중 코드 0에서 코드 2로 분류된 긴급 및 즉시 출동 신고로 설정하였다. 앞 장에서 112신고 데이터를 500m 격자 단위로 집계한 것과 동일한 방법을 적용하여 건축도시 물리환경 변수를 격자 단위로 전처리하고 탐색적 공간 데이터 분석을 통하여 112신고 데이터의 시공간적 패턴을 시각화하고 후속 연구 방향을 설정하였다. 마지막으로 기계학습 및 공간회귀 분석 방법론을 적용하여 물리환경 요인 중 주요 설명변수를 도출하고 이들 변수가 112신고 발생에 미치는 영향을 여러 관점에서 확인하였다.

본 연구는 물리환경이 범죄 발생에 미치는 영향을 분석한 전년도 연구의 연속으로 동일한 물리환경 요소를 적용하였다. 전년도 연구의 건축·도시 물리환경 변수는 국내외의 다양한 연구와 CPTED(범죄예방환경설계) 체크리스트를 참조하여 도출되었으며 공공데이터 구득이 가능한 변수를 분석에 사용하였다. 최종적으로 선정된 변수는 인구 특성, 건물 및 토지 특성, 생활 편의 및 상권 시설, 교통 및 기반 시설, 도로 공간 구조 등 분야의 변수가 포함되었다.

변수 전처리는 먼저 변수 스케일 조정, 이상치 처리, 결측치 제거를 거쳐 변수를 정제하였다. 다음으로 변수 간 다중공선성과 상관계수를 분석하여 VIF 값이 15 이상으로 나타나 다중공선성이 있다고 판단되는 변수와 상관계수가 0.7 이상인 변수 쌍을 대상으로 분석 대상에서 제외하였다. 면적 기반 변수와 상응하는 비율 간에는 상관계수가 1에 가깝게 나타나 두 변수 중 해석에 유리한 비율 변수만 분석에 적용하였다. 인구 관련 변수에서도 총인구와 연령대별 변수 사이에 강한 상관관계가 나타나 총인구를 분석 변수에서 제외하였다. VIF가 15 이상으로 나타난 도로 통합성, 보도 통합성, 보도 연결성, 도로 연결성, 노후도 등도 제외하였다.

전역적 공간 자기상관 분석(Global Moran's I)을 통하여 전체 공간에서의 평균적인 공간 자기상관 경향을 파악하고 공간회귀분석의 필요성을 확인하였다. 분석 결과 Moran's I 값이 0.485로 나타나 유의미한 양의 공간 자기상관성이 나타났다. 즉, 112신고가 많이 발생하는 지역은 인접한 다른 지역에서도 112신고가 많이 발생하는 경향을 보였다.

다음으로 국지적 공간 자기상관 분석(LISA)을 통하여 지역별 클러스터를 도출하였다. 특히 112신고가 집중되어 주변 지역보다 높게 나타나는 핫스팟(High-High)의 특성을 검토하였다. 핫스팟은 서울시 전역에 나타났으나 특히 도심 및 강남 권역에 상대적으로 많이 분포하였다. 특히 중구, 영등포구, 금천구 등에서는 핫스팟이 콜드스팟보다 우세하게 나타났다.

LISA 분석을 통하여 도출된 핫스팟 격자 지역의 특성을 콜드스팟 및 배경 지역과 비교하여 분석하고 물리환경이 112신고 발생에 미치는 영향을 비모수 통계 기법인 Mann-Whitney U 검정을 통하여 분석하였다. 먼저 핫스팟과 콜드스팟을 대상으로 물리환경 변수의 분포 차이를 독립적으로 검정한 결과, 모든 변수에서 강한 유의성을 보여 물리환경과 112신고 발생 사이에 유의한 영향이 있음을 확인하였다. 핫스팟은 콜드스팟에 비하여 인구가 밀집되고 토지가 격이 높고 건축 밀도가 높은 특징을 보였다.

다음으로 물리환경 변수별, 자치구별로 자치구 내 핫스팟과 해당 자치구 및 서울시 전역 등 배경지역 간 분포 차이를 검정한 결과, 대부분 변수에서 유의한 차이를 보였으나 일부 변수의 경우 유의한 평균 차이를 보이지 않았다. 인구 특성에서는 유아, 학생, 청년, 중장년, 고령 등 모든 연령대별 인구 변수에서 유의한 차이를 보였다. 건물 특성에서는 공시지가, 용적률, 건폐율, 건물 높이, 건물 수 등 변수에서 토지이용 특성에서는 주거, 상업, 녹지 등 용도지역 비율 변수에서 유의한 차이가 나타났다. 반면 농지, 나지, 공업, 공원, 수변 등 비율은 유의한 차

이를 보이지 않았다. 생활편의 및 상권 시설에서는 술집, 음식점·카페, 숙박, 편의점, 병·의원, 교육·공공시설, 세탁소, 소매점 등이 교통 및 기반 시설에서는 대중교통시설, 주차장, CCTV 등이 유의한 차이를 보였다.

기계학습 분석을 위하여 부스팅(Boosting) 계열 머신러닝 모델인 XGBoost와 LightGBM을 각각 학습시켜 예측 성능을 비교하였다. 두 모델 모두 본 연구 데이터셋에 대해 높은 예측력을 보였으나 XGBoost의 결정계수(Test R^2)가 0.85로 LightGBM(0.83)보다 근소하게 높게 나타나 최종 예측모형으로 채택하였다. 기계학습 모형의 해석을 위해서는 SHAP (Shapley Additive exPlanations) 기법을 적용하였다. 전역적 변수 중요도(Global Feature Importance)는 평균 SHAP 값($\text{mean}(|\text{SHAP value}|)$)이며 각 변수가 모형 전반에서 예측값 변동에 얼마나 기여했는지를 상대적으로 보여준다. 상위 10개 변수는 술집, 음식점·카페, 소매점, 대중교통시설, 건물 수, 상업지역 비율, 편의점, 녹지 비율, 공시지가, 교육·공공시설 등으로 나타났다.

상위 10개 변수에 대하여 SHAP 의존도 플롯(SHAP dependence plot)을 통하여 각 변수의 값에 따른 예측값의 변동(SHAP 값)의 관계를 분석하였다. 술집, 음식점·카페, 소매점, 대중교통시설, 편의점, 공시지가 등은 신고 증가 영향을 보였으며 변수 각각에 따라 구체적인 영향 패턴은 단순히 선형 관계가 아니라 다양한 형태를 보였다. 녹지 비율, 교육·공공시설 등은 변수값이 증가할수록 신고가 감소하는 영향을 보였다. 특히 녹지 비율의 경우 녹지 비율이 매우 낮은 경우 112신고 발생이 높게 나타나고 녹지 비율이 약 40%에 달할 때까지 신고 발생이 감소하다가 그 이상 구간에서는 추가 상승 영향이 나타나지 않았다. 건물 수, 상업지역 비율 등은 복합적 영향을 보였다.

기계학습 분석에서 도출된 상위 10개 변수를 대상으로 다중선형회귀(MLR), 지리 가중 회귀 모형(GWR), 다중 범위 지리 가중 회귀모형(MGWR) 등 여러 회귀 및 공간회귀 모형을 적용한 분석 결과를 비교하였다. GWR과 MGWR 등 공간회귀 모형 모두 MLR 회귀 모형 대비 성능이 개선되었다. GWR은 회귀계수가 가장 높았지만 MGWR은 회귀계수가 유사한 수준이면서 모형의 적합도를 나타내는 AICc(Akaike Information Criterion corrected)가 더 낮게 나타나 최종 모형으로 채택하였다.

분석 결과, 술집, 음식점·카페, 소매점, 대중교통시설, 편의점, 녹지 비율, 공시지가 등은 서울 전역에서 일정한 관계를 보였다. 그러나 교육·공공시설은 일부 지역에서만 유의한 영향이 나타났고 건물 수, 상업지역 비율 등은 지역별로 증가 영향, 감소 영향, 영향 없음 등이 혼합되어 나타나는 복합적인 관계를 보였다.

제5장 결론

본 연구는 112신고가 집중되는 지역의 물리적·사회적 특성을 체계적으로 규명하고자 물리환경 요인이 112신고 발생에 미치는 영향을 500m 격자 단위로 공간통계 및 기계학습 기법을 사용하여 정량적으로 분석하였다. 이를 통하여 112신고가 복합적 요인과 체계적으로 연결됨을 데이터를 기반으로 확인하고 그 구조를 시·공간적으로 설명하였다.

탐색적 공간 데이터 분석 중 LISA 분석을 통하여 핫스팟과 콜드스팟을 도출하였다. 핫스팟은 도심 및 한강 이남 중심지에 집중되어 분포하였고 핫스팟과 콜드스팟은 물리적·사회적 환경 변수 대부분에서 유의한 차이를 보였다. 특히 인구 및 건축물 밀도에서 큰 차이를 보였는데 신고가 집중된 공간이 인구가 밀집한 공간과 대체로 일치하는 경향으로 해석할 수 있었다.

기계학습 분석 결과 112신고 발생 영향 상위 10대 변수가 도출되었고 술집, 편의점, 대중교통시설, 소매점, 건물 수 등이 신고 증가 영향 요인으로 녹지 비율, 교육·공공시설 등이 신고 감소 요인으로 나타났다. 이들 변수에 대한 공간회귀 분석 결과 대부분 서울 전역에서 일정한 관계를 보였으나 교육·공공시설은 일부 지역에서만 유의한 영향이 나타나고 건물 수, 상업지역 비율 등은 복합적 영향을 보였다.

본 연구는 112신고를 안전 지표로 확장하여 물리환경이 범죄 발생 및 무질서에 미치는 영향에 대한 기존 주장을 신고 데이터 기반으로 실증하였다는 점에서 의의가 있다. 술집 수 등 각종 상업시설 관련 변수에서 신고 발생 증가 영향이 나타난 것은 선행연구의 주장과 일관된 결과였다. 반면 녹지 비율의 경우 선행연구의 주장이 엇갈리고 있는 상황이며 서울 지역을 대상으로 도심에서 녹지가 신고 발생 감소 영향이 나타난다는 결과를 새롭게 제시하였다.

본 연구는 112신고 데이터를 기반으로 범죄 발생 및 범죄두려움에 대한 물리환경의 영향을 실증하였으나 112신고 데이터 분석과 관련하여 여러 한계를 보였다. 첫째, 「112신고처리법」에 따라 112신고자 정보 보호를 위하여 신고 유형의 구분 없이 격자 단위로 집계한 통계치만 분석에 사용하였다. 이에 따라 신고 유형별 분석 등 세밀한 분석에 어려움이 있었다. 둘째, 격자 단위 분석으로 인하여 지점 또는 격자 단위 데이터를 제공하지 않는 물리환경 변수를 분석에 반영하지 못한 것도 연구의 한계로 남는다.

향후 격자 단위, 행정구역/집계구 단위, 전국 단위 데이터를 모두 고려한 분석이 이루어질 필요가 있으며 서울 외 지역에 대한 분석을 통하여 지역 간 차이를 확인할 필요가 있다. 이러한 후속 연구를 통하여 각 지역에서 이루어지고 있는 CPTED 정책 및 사업이 실증 근거에 기반하여 실질적인 효과를 거둘 수 있는 방향으로 이루어질 수 있을 것이다.

주제어

112신고, 범죄두려움, 범죄예방환경설계, 물리환경 요소, 기계학습

제1장 서론	1
1. 연구의 배경 및 목적	2
1) 연구의 배경	2
2) 연구의 목적	4
2. 연구의 범위 및 방법	5
1) 연구의 범위	5
2) 연구의 방법	6
3. 연구의 차별성	8
1) 선행연구 동향	8
2) 선행연구의 한계점	10
3) 본 연구의 차별성	11
제2장 112신고 제도 및 활용	15
1. 112 신고 제도	16
1) 112 신고 관련 제도 현황	16
2) 「112 신고처리법」 주요 내용	23
2. 112 신고 체계	26
1) 112 신고 처리 절차	26
2) 112 신고 코드 분류 체계	27
3) 112 신고 내용 분류 및 기록	28
4) 112 신고 데이터 활용	30
3. 112 신고 데이터 활용	32
1) 112 신고 데이터 활용 사례	32
2) 112 신고 데이터 활용의 가능성	40
3) 112 신고 데이터 활용의 한계	41
4. 소결	43

제3장 112신고 발생 기술통계 분석	45
1. 분석 개요	46
1) 분석 대상	46
2) 분석 방법	46
2. 분석 결과	52
1) 월별, 요일별 112 신고 발생	52
2) 시간대별 112 신고 발생	53
3) 긴급코드별 112 신고 발생	54
4) 112 신고와 범죄 발생 패턴 관계 분석	58
3. 소결	59
제4장 물리환경이 112신고 발생에 미치는 영향 실증 분석	61
1. 분석 개요	62
1) 분석 범위 및 방법	62
2) 건축도시환경 변수 선정 및 전처리	66
3) 공간정보 분석 및 기계학습 방법론	86
2. 분석 결과	96
1) 탐색적 공간 데이터 분석	96
2) 물리환경과 112 신고 발생 관계 분석	105
3) 공간회귀 기반 분석	126
제5장 결론	141
1. 연구의 성과	142
2. 연구의 한계 및 향후 과제	149
참고문헌	151
SUMMARY	161
부록	169

[표 1-1]	선행연구 검토 및 본 연구의 차별성	12
[표 2-1]	112 신고 관련 법령 현황	17
[표 2-2]	「112 치안종합상황실 운영 및 신고처리 규칙」의 제·개정	18
[표 2-3]	신고 대응 Code 분류	28
[표 2-4]	112 신고자료의 체계	29
[표 3-1]	112 신고 데이터 구조	47
[표 3-2]	월별, 요일별 112 신고건수 추이	52
[표 3-3]	계절별 112 신고건수 추이	53
[표 3-4]	시간대별 112 신고건수 추이	53
[표 3-5]	코드별 112 신고건수 추이	55
[표 4-1]	물리환경 요소 분류	65
[표 4-2]	기계학습 기반 112 신고데이터와 건축환경 분석을 위한 종속변수와 독립변수	68
[표 4-3]	활용가능한 독립변수 공공데이터 출처	69
[표 4-4]	변수별 데이터 프로세싱 과정	73
[표 4-5]	LISA 분석 결과 분류	87
[표 4-6]	상관관계 0.7 이상 변수쌍 목록	97
[표 4-7]	LISA 군집 유형별 의미	102
[표 4-8]	행정구별 LISA 군집 유형 수	104
[표 4-9]	변수별 HH/LL 그룹별 평균 차이 분석 결과	107
[표 4-10]	서울시-행정구-핫스팟(HH) 지역의 변수별 평균 차이 분석 결과	118
[표 4-11]	MLR, GWR, MGWR 에 대한 모델 평가(전체시간)	126
[표 4-12]	MLR, GWR, MGWR 에 대한 모델 평가(낮시간)	127
[표 4-13]	MLR, GWR, MGWR 에 대한 모델 평가(밤시간)	127
[표 4-14]	MLR 매개변수에 대한 요약통계(전체시간)	128
[표 4-15]	MLR 매개변수에 대한 요약통계(낮시간)	129
[표 4-16]	MLR 매개변수에 대한 요약통계(밤시간)	129
[표 4-17]	GWR 매개변수에 대한 요약통계(전체시간)	131
[표 4-18]	GWR 매개변수에 대한 요약통계(낮시간)	132
[표 4-19]	GWR 매개변수에 대한 요약통계(밤시간)	133
[표 4-20]	MGWR 매개변수에 대한 요약통계(전체시간)	135
[표 4-21]	MGWR 매개변수에 대한 요약통계(낮시간)	136
[표 4-22]	MGWR 매개변수에 대한 요약통계(밤시간)	138
[표 5-1]	주요 환경요인별 주요 결과 및 이론적 근거와 정책적 함의	148

[그림 1-1]	범죄피해두려움 통계 추이	3
[그림 1-2]	112 신고 접수 및 출동 관련 통계 추이	3
[그림 1-3]	연구의 흐름	7
[그림 2-1]	112 신고처리 흐름도	26
[그림 2-2]	NYPD CompStat 시스템	32
[그림 2-3]	PredPol	34
[그림 2-4]	Police.uk 에서 공개된 범죄 지도 및 통계	36
[그림 2-5]	Police.uk 에서 범죄 신고 및 처리 타임라인	37
[그림 2-6]	런던 캠프 차량 절도 핫스팟 분석	38
[그림 2-7]	프리카스(Pre-CAS)	39
[그림 3-1]	112 신고데이터 시간대별 분포 이상치 탐지 시각화	48
[그림 3-2]	112 신고데이터 IQR 이상치 탐지	48
[그림 3-3]	112 신고데이터 표준화 점수 이상치 탐지	49
[그림 3-4]	112 신고데이터 요일·시간대 그룹 Z-score 와 변동률 간 관계	50
[그림 3-5]	112 신고데이터 요일·시간대별 신고 추이 및 이상치 탐지	51
[그림 3-6]	요일·시간대별 112 신고패턴 히트맵	54
[그림 3-7]	112 신고건수 전반적 추세 시각화	55
[그림 3-8]	긴급코드별 112 신고건수 전반적 추세 시각화	56
[그림 3-9]	긴급코드별 112 신고건수 상세 추이 시각화	57
[그림 3-10]	긴급코드별 요일·시간대별 112 신고패턴 히트맵	57
[그림 3-11]	격자기반 공간 상관성 분석 결과 시각화	58
[그림 4-1]	연구 프레임워크	64
[그림 4-2]	지역 간 가로등 위치 데이터의 격차가 확인됨	66
[그림 4-3]	독립변수의 공간적 분포 시각화 (1)	78
[그림 4-4]	독립변수의 공간적 분포 시각화 (2)	79
[그림 4-5]	독립변수의 공간적 분포 시각화 (3)	80
[그림 4-6]	독립변수의 공간적 분포 시각화 (4)	81
[그림 4-7]	독립변수의 공간적 분포 시각화 (5)	82
[그림 4-8]	독립변수의 공간적 분포 시각화 (6)	83
[그림 4-9]	독립변수의 공간적 분포 시각화 (7)	84
[그림 4-10]	SHAP 개념 설명 및 예시	94
[그림 4-11]	피어슨 상관계수 값 시각화	97
[그림 4-12]	분산팽창지수(VIF) 값 시각화	98

[그림 4-13] Moran's I 산점도 시각화	100
[그림 4-14] LISA 국지적 군집 시각화	102
[그림 4-15] SHAP 기반 전역적 변수 중요도 시각화	121
[그림 4-16] SHAP 기반 요약 플롯 시각화	123
[그림 4-17] 변수별 SHAP 의존도 플롯	125
[그림 4-18] 변수별 MGWR 계수 분포 (전체시간)	134
[그림 4-19] 변수별 MGWR 계수 분포 (낮시간)	137
[그림 4-20] 변수별 MGWR 계수 분포 (밤시간)	139

제1장

서론

1. 연구의 배경 및 목적
2. 연구의 범위 및 방법
3. 연구의 차별성

1. 연구의 배경 및 목적

1) 연구의 배경

범죄두려움은 현대 사회에서 삶의 질을 결정하는 핵심적인 요인 중 하나로 부상하고 있다. 범죄가 개인과 사회에 미치는 영향은 단순히 실제 범죄 발생에 국한되지 않으며 범죄에 대한 우려와 두려움 자체가 독립적으로 광범위한 사회적 파급효과를 발생시키고 있다. 특히 코로나 19 팬데믹 이후 사회 전반에 걸친 불안감이 증대되면서 범죄두려움의 사회적 중요성은 더욱 부각되고 있는 상황이다. 범죄두려움은 범죄 피해에 대한 개인의 주관적 불안감과 우려를 의미하며 이는 실제 범죄 발생률과 반드시 일치하지 않는 심리적 현상이기 때문에 범죄 발생과 별개의 문제로 관리할 필요가 있다.

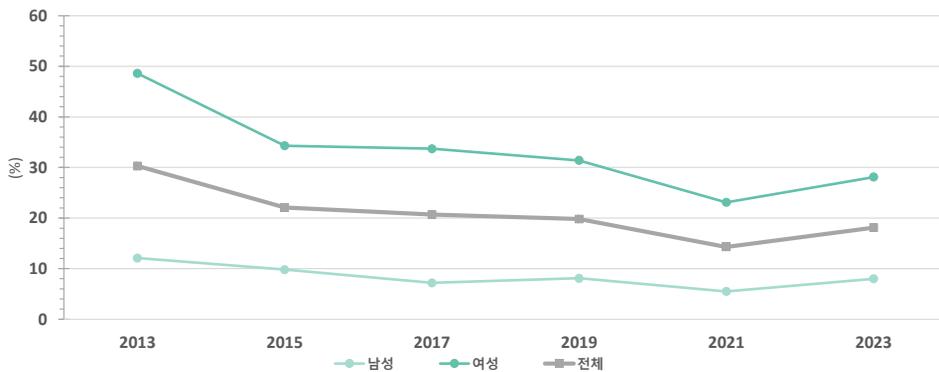
범죄두려움은 개인의 사회적 활동 참여를 현저히 감소시킨다. 야간 외출 기피, 대중교통 이용 기피, 문화·여가 활동 참여 감소 등이 대표적인 사례이다. 이러한 활동 제약은 개인의 사회적 관계 형성과 유지를 어렵게 할 뿐만 아니라 자연적 감시, 접근통제, 영역성 강화, 활동 지원 등의 범죄예방 환경설계 원칙들의 효과적 적용을 저해하는 요인으로 작용할 수 있다. 범죄두려움은 종종 과도한 물리적 보안 장치 설치에 대한 요구로 이어지며 CCTV, 철조망, 보안 게이트 등의 물리적 장벽은 단기적으로는 안전감을 제공할 수 있지만 장기적으로는 공간의 개방성과 접근성을 저해하여 자연적으로 범죄로부터 안전한 도시 환경 조성에 부정적 영향을 미칠 수 있다.

코로나19 팬데믹은 사회 전반에 걸쳐 불확실성과 불안감을 증대시켰으며 범죄두려움 수준도 상승하였다. 범죄피해두려움에 대한 조사에서 장기간 감소 추세를 보이던 범죄두려움 수준이 코로나19 이후 반등하는 양상을 보이고 있다. 이는 팬데믹이 사회 안전에 대한 인식과 태도에 구조적 변화를 가져왔음을 시사한다.

이러한 범죄두려움은 112신고 발생에도 영향을 미친다. 112신고는 범죄가 발생했다고 인식한 시민이 긴급 출동을 요청하는 행위로 실제 범죄 발생과 범죄에 대한 인식 및 두려움을 동시에 반영하는 중요한 행태적 지표이다. 설문조사나 면접을 통한 주관적 인식 조사와 달리, 112

신고는 실제로 전화를 걸어 신고하는 구체적 행동을 수반한다. 이는 단순한 의견 표명이 아닌 '행동으로 나타난 우려'를 의미하므로 실제로 범죄가 발생한 것이 아닌 경우에도 범죄두려움이 개인의 일상생활에 실질적으로 영향을 미치는 정도를 보다 정확히 측정할 수 있게 한다. 이는 객관적 현상과 주관적 인식을 연결하는 교량 역할을 한다. 코로나19 이후 범죄두려움이 증가한 시기에 112신고 접수도 함께 증가하였다.

112신고 데이터는 범죄두려움과 관련된 정책 수립에 필수적인 실증적 근거를 제공할 수 있다. 지역별, 시간대별, 유형별 신고 패턴 분석을 통해 범죄두려움이 높은 특정 지역이나 상황을 식별할 수 있으며 이는 선제적 안전 정책 수립의 기초 자료가 된다. 또한 정책 시행 전후의 112신고 패턴 변화를 추적함으로써 정책 효과를 실시간으로 평가할 수 있는 도구로 활용 가능하다.



[그림 1-1] 범죄피해두려움 통계 추이

출처: 한국형사정책연구원. (2024). 범죄피해두려움. 지표누리. <https://www.index.go.kr/unity/potal/indicator/indexInfo.do?clasCd=2&idxCd=4264> (검색일: 2025.2.28.)를 참고하여 연구진 작성



[그림 1-2] 112신고 접수 및 출동 관련 통계 추이

출처: 경찰청. (2023). 112신고접수 현황. 지표누리. https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx_cd=1609 (검색일: 2025.2.27.)를 참고하여 연구진 작성

2) 연구의 목적

본 연구의 목적은 건축·도시 공간의 물리적 환경요소와 112신고 발생 사이의 연계성을 파악하고 물리환경의 특성이 112신고 발생에 미치는 영향을 데이터 기반으로 실증하는 데에 있다.

112신고는 범죄 발생과 범죄에 대한 긴급한 우려를 포괄하는 지표로서 범죄예방 정책 수립 시 범죄두려움 저감 목표를 구체화하는 데 기여할 수 있다. 특히 기존 범죄두려움 연구는 주로 설문조사를 통한 자가보고(self-report) 방식에 의존해 왔으나 이는 응답자의 주관적 인식에 따른 편향과 사회적 바람직성 편향 등의 한계를 내포하고 있다. 112신고 데이터를 활용한 연구를 통하여 범죄예방뿐만 아니라 범죄두려움 저감에 대해서도 행태적 근거를 제공하는 것을 목적으로 한다. 물리환경이 112신고 발생에 미치는 영향을 실증하여 CPTED 관점에서 효과성 높은 물리환경 개입 전략에 대한 근거를 도출하고자 한다.

2. 연구의 범위 및 방법

1) 연구의 범위

■ 내용적 범위

본 연구는 물리환경이 112신고 발생에 미치는 영향을 실증하기 위하여 112신고 제도 및 활용 분석, 112신고 데이터 기술통계 분석, 마지막으로 물리환경 영향 실증 분석의 단계로 진행하였다. 112신고에 미치는 영향의 의미를 파악하기 위하여 가장 먼저 112신고의 성격을 분석하고 112신고 데이터에 대한 기술통계를 통해 물리환경 영향 분석의 조건을 설정하였다. 이를 통하여 범죄 발생과 범죄두려움 등 112신고가 범죄예방 정책 목표와 연관된 지점을 파악하고 112신고 발생에 미치는 물리환경의 영향과 범죄예방 환경설계 정책 추진방향을 제시하고자 하였다.

■ 공간적 범위

물리환경 특성의 영향을 다양한 환경에서 실증하기 위하여 본 연구의 공간적 범위는 서울특별시 전역으로 설정하였다. 서울특별시는 다양한 물리환경 및 인구사회학적 특성이 나타나는 거대도시로 높은 인구밀도로 인하여 범죄 발생 및 112신고 발생 수준이 높은 지역이다. 전년도에 수행한 물리환경이 범죄 발생에 미치는 영향 연구에서는 서울 내 6개 구를 연구의 공간적 범위로 설정하였는데 본 연구에서는 이를 계승하고 확장하여 연구의 연속성을 확보하는 동시에 해석의 폭이 확장될 수 있도록 하였다.

■ 시간적 범위

연구의 시간적 범위는 2023년 1년간으로 설정하였다. 2024년 7월 「112신고의 운영 및 처리에 관한 법률」(약칭 「112신고처리법」)이 시행되었다. 이로써 112신고 데이터에 대한 접근이 어려워짐에 따라 다년간에 대한 연구를 통한 특정 연도에 발생한 사건의 영향을 통제하는 것이 불가능하게 되었다. 본 연구에서는 이러한 한계를 반영하여 2023년 1년간을 시간적 범위

로 설정하였으나 해당 연도만의 특수성을 고려하여 해석에 유의하였다.

2) 연구의 방법

■ 문헌연구

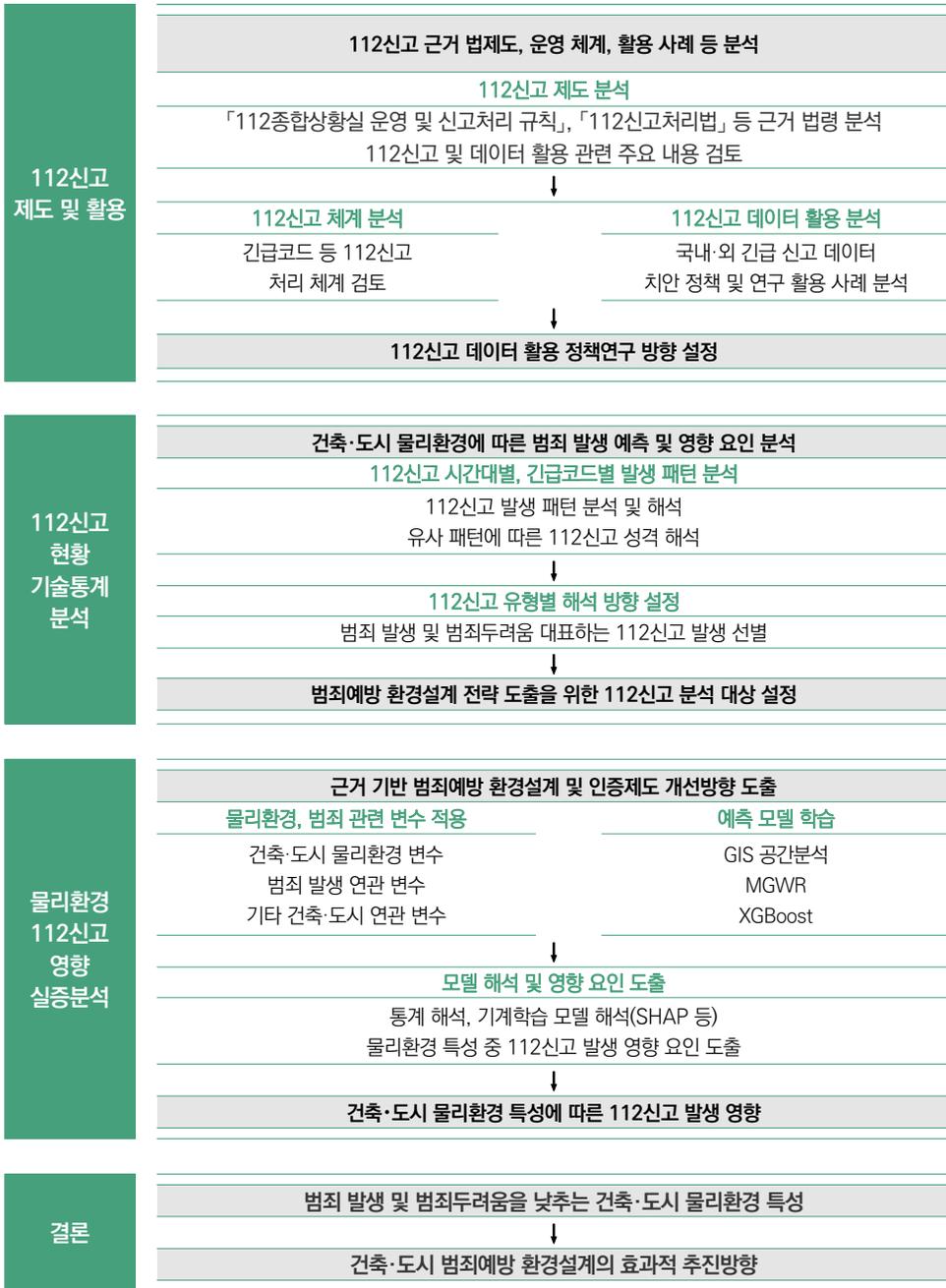
국내·외 긴급 신고 데이터를 활용하여 치안 정책 방향을 설정하고 범죄예방 환경설계에 환류하는 방안을 검토하기 위하여 법·제도, 국내·외 사례, 연구문헌, 언론보도 등을 검토하였다. 물리환경 변수 선정에도 선행연구와 CPTED 체크리스트 등 다양한 문헌연구 결과가 반영되었다.

■ GIS 공간분석 및 통계분석

건축물과 도시공간의 물리환경 특성, 112신고 발생 등에 대하여 격자 단위 공간정보화, 기초 통계 분석 및 GIS를 활용한 공간분석을 수행하였다. GIS 분석에는 핫스팟 분석, 공간 자기상관 분석, 버퍼 분석 등 다양한 공간분석 기법이 적용되었으며 이를 통하여 단순 통계분석에서 파악할 수 없는 시공간적 패턴을 탐색하고 물리환경의 비선형적 영향을 반영하였다.

■ 공간회귀 및 기계학습

본 연구는 전년도 물리환경의 범죄 발생 영향 연구의 후속 연구로서 112신고 데이터와 도시 건축환경 요소 간의 관계를 분석하는 것을 목적으로 한다. 이러한 연속성을 고려하여 전년도 연구에서 효과성이 검증된 기계학습 방법론을 동일하게 활용하고자 하였다. 물리환경 영향 분석에는 공간 회귀모형인 다중 범위 지리 가중 회귀(MGWR), 머신러닝 기반 회귀모형인 XGBoost 두 가지 주요 예측 모델을 활용하였다.



[그림 1-3] 연구의 흐름

출처: 연구진 작성

3. 연구의 차별성

1) 선행연구 동향

최근 사회 전반에서 범죄에 대한 대응과 예방의 중요성이 강조되면서 일선 현장의 가장 초기 반응을 포착할 수 있는 112(한국) 및 911(미국) 긴급 신고 데이터가 다양한 학술 및 정책 연구에 활용되고 있다. 이들 신고 데이터는 범죄 발생 당시의 시간·공간·상황 정보를 풍부하게 담고 있어 도시 안전정책 수립, 지역 위험요인 분석, 예측 모델 기반의 선제적 대응체계 구축 등에 매우 유용한 자료로 간주된다. 특히 공간정보 및 도시 물리환경 데이터와 결합할 경우, 도시 구조와 사회적 혼란 간의 상관성을 정량적으로 분석할 수 있어 최근에는 다양한 융복합 연구로 확장되고 있다.

국내에서의 112신고 데이터 활용 연구는 주로 경찰청 제공 자료를 기반으로 하며 공간정보 분석과 통계모형을 접목한 실증연구들이 이루어지고 있다. 대표적인 연구 유형은 다음과 같다.

첫째, 도시 물리적 특성과 범죄 간의 관계를 분석한 연구가 활발히 진행되고 있다. 정진도(2023)는 서울시 골목길의 공간 구조와 112신고 데이터를 연계하여 골목길의 물리적 특성과 지역 낙인 효과가 범죄 발생에 미치는 영향을 분석하였다. 이 연구는 골목길의 청결 상태, 이동성, 이동성, 노후도 등의 인식적 특성과 CCTV, 방범등, 비상벨 등의 방범 시설 배치가 범죄 발생과 어떤 관련을 가지는지를 음이항 회귀분석을 통해 규명하였다. 분석 결과, 미시적 도시 공간이 특정 유형의 범죄 신고와 통계적으로 유의미한 상관관계를 가지는 것으로 나타났으며 이는 도시 설계와 지역 낙인의 상호작용을 고려한 범죄 예방 전략 수립의 근거로 제시되었다.

둘째, 시간대, 기후, 인구 등 외부 변수와 결합한 예측 모델 개발이 시도되고 있다. 최재훈(2018)은 충남지방경찰청의 1년간 수집된 112신고 데이터 중 코드 C0~C2로 분류된 데이터를 기반으로 계절, 요일, 기온, 풍속, 인구, 주택 비율 등 다양한 독립변수를 반영한 음이항 회귀 모델과 신경망 기반 예측 모델을 개발하여 그 성능을 비교하였다. 분석 결과, 여름, 휴일 전날, 휴일, 높은 기온 등에서 112신고가 증가하는 경향을 보였으며 이는 계절성과 기후 조건이 신고량에 영향을 미침을 보여준다. 특히 신경망 기반 예측 모델이 회귀 모델보다 높은 예측

성능을 보였고 이를 통해 고정된 인력 운용 방식을 개선하여 특정 시간과 지역에서의 신고 집중도를 예측하고 효율적인 경찰 인력 배치를 도모할 수 있는 데이터 기반 전략을 제시하였다.

셋째, 신고 텍스트 데이터를 활용한 비정형 데이터 분석이 본격적으로 도입되고 있다. 홍세은 외(2020)는 112신고 내용을 벡터화하고 사건 유형 간 군집화를 수행하는 텍스트 마이닝 기반 연구를 통해 사건 간의 구조적 유사성과 유형별 공간 집중 현상을 분석하였다. 이는 기존의 범주형 코드 중심 분석을 넘어서 신고 데이터의 질적 내용을 정량적으로 해석하려는 시도로 평가된다. 또한 정재훈 외(2024)는 Transformer 기반 대형 언어 모델(LLM)을 활용하여 허위·오인 신고를 판별하는 예측 모델을 개발하였다. 이 연구는 신고자의 표면적 진술만으로는 구분하기 어려운 악의적 허위 신고를 조기에 식별하고자 하였으며 개발된 모델은 경찰의 의사결정을 지원하는 데 기여할 수 있을 것으로 기대된다.

넷째, 신고 내용과 긴급성 판단의 정합성 문제도 분석되고 있다. 노성훈 외(2016)는 112신고 데이터에 서술된 사건 내용을 기반으로 신고에 부여된 긴급성 코드와 실제 사건의 긴급성 간의 일치 여부를 검토하였다. 연구진은 ‘생명 및 신체에 대한 위협’을 중심으로 긴급성을 정의하였으며 분석 결과 신고 유형, 접수 시간, 관할 경찰서에 따라 긴급성 코드의 부여에 통계적으로 유의미한 차이가 있음을 확인하였다. 특히 긴급 신고로 분류된 사건 중 40% 이상이 실제로는 비긴급에 해당하였고 반대로 경찰이 출동한 비긴급 신고 중 약 16.5%는 출동이 불필요한 신고로 판단되었다. 이 연구는 효율적인 긴급신고 대응 체계 마련을 위해 명확한 기준과 운영 규정의 개선 필요성을 제안하였다.

해외에서는 미국의 911 신고 데이터를 중심으로 보다 다양한 연구가 축적되고 있으며 특히 공간정보기반 예측모델링과 기후 변수와 범죄의 상관성 분석, 도시 안전 정책 수립을 위한 응용연구가 활발하다.

Cramer 외(2012)는 911 신고 데이터를 활용하여 핫스팟 분석(Hot Spot Analysis), 일반 선형회귀(OLS), 그리고 지리 가중 회귀(GWR)를 수행하였다. 핫스팟 분석을 통해 911 신고가 집중되는 지역을 도출하였으며 이는 제한된 경찰 자원을 효과적으로 배치하기 위한 근거로 활용될 수 있음을 제안하였다. 이어서 환경적 요소와 신고 데이터 간의 연관성을 분석하기 위해 단계적 회귀분석(Stepwise Regression)을 실시하였고 이 과정에서 임차인 비율(renters), 도심으로부터의 거리(distance to urban center), 사업체 수(businesses), 직업 수(jobs), 비경제활동인구(not in labor force), 대졸자 비율(college graduates) 등이 유의미한 독립변수로 선정되었다. 이후 OLS와 GWR 모델을 통해 공간적 연관성을 분석하고 두 모델 간 예측 성능을 비교하였다.

911 신고 예측을 위한 시계열 기반 접근도 활발히 이루어지고 있다. Robles 외(2020)는 Gaussian Process 모델과 ARIMA 모델을 사용하여 날씨 조건과 휴일 여부에 따른 911 신고

량을 예측하였으며 도시 환경 요인은 제외한 채 두 시계열 모델 간의 성능을 비교하였다. Concoran(2019)은 공간적 인구통계학 특성, 시간적 요인, 날씨 등을 독립변수로 설정하고 폭력 사건 발생 여부를 종속변수로 하여 로지스틱 회귀(Logistic Regression), 랜덤 포레스트(Random Forest), 서포트 벡터 머신(SVM) 등 다양한 예측모형을 비교 분석하였다.

기온과 계절성과 범죄 발생 간의 관계를 밝히려는 연구도 주목받고 있다. Gamble 외(2012)의 연구에서는 평균 기온이 상승함에 따라 살인, 강간, 폭행 등 폭력 범죄 발생이 증가하는 경향을 보였으며 특히 약 30°C(90°F)를 전후한 시점에서 범죄 발생량이 증가에서 감소로 전환되는 ‘역U자형 관계’가 확인되었다. Shen et al.(2020)는 상대 습도나 안개 등 대기 오염 요소도 특정 범죄 유형에 영향을 미칠 수 있음을 제시하였다. 계절별 비교에서도 여름철이 겨울보다 범죄 발생과 더 높은 상관관계를 보였으며(Stevens et al., 2019), 특히 폭력 범죄는 날씨에 민감한 반면, 절도 범죄는 날씨의 영향을 거의 받지 않는 것으로 나타났다(Field, 1992).

2) 선행연구의 한계점

한국의 112신고 및 미국 911 신고 데이터를 활용한 선행연구들은 도시 내 범죄와 위험 요소를 조기 탐지하고 공간적 취약지역을 진단하는 데 있어 유의미한 기여를 해왔다. 특히, 공간 통계, 기후 변수, 물리적 환경 요인 등을 통합적으로 고려한 연구들은 범죄 발생의 다층적 원인을 규명하고자 하는 시도라는 점에서 중요한 의미를 가진다. 그러나 이러한 선행연구들에는 몇 가지 한계점이 존재하며 본 연구는 이를 보완하고 확장하고자 한다.

첫째, 공간정보에 치우친 국내 연구에서는 시계열 요소의 통합이 미흡하다. 예를 들어, 정진도(2023)의 연구는 골목길의 물리적 특성과 방범 시설, 인식적 요소 등을 반영한 공간 회귀모형을 통해 범죄 발생과의 연관성을 도출하였지만 기온이나 요일, 시간대와 같은 시간 기반 외부 요인은 반영되지 않았다. 이로 인해 특정 시기나 기상 조건에 따른 범죄 패턴의 변화를 설명하는 데 한계가 있다.

둘째, 시계열 요소를 반영한 연구는 환경적 요인과의 복합적인 상호작용을 설명하기 어렵다. 최재훈(2018)의 연구처럼 기온, 휴일 여부, 요일 등의 변수를 활용하여 112신고량 예측을 시도한 연구들은 대부분 음이항 회귀나 ARIMA와 같은 선형 모델을 사용하였다. 그러나 이러한 접근은 환경적 요인이 범죄에 미치는 비선형적 영향력이나 임계값 효과(예: Gamble 외, 2012에서 제시한 30도를 전후한 역U자형 패턴)를 포착하는 데 한계가 있다.

셋째, 비선형 모델이 도입된 일부 연구에서도 해석 가능성과 설명력의 부족이 문제로 지적된다. 예를 들어 Concoran(2019)은 Random Forest와 SVM을 포함한 여러 비선형 예측 모델을 통해 폭력 범죄 발생 여부를 예측하였으나 모델 성능 비교에 집중되어 있을 뿐, 각 요인이

범죄에 미치는 구체적인 영향이나 공간적·시간적 요인의 해석은 충분히 이루어지지 않았다. Robles et al.(2020) 또한 Gaussian Process를 활용해 시계열적 예측 성능을 개선하였으나 도시 환경 요인은 제외되어 통합적 분석이 어려운 구조였다.

넷째, 신고 데이터의 전면적 활용 부족이다. 국내 대부분의 연구는 코드 분류 체계에 대한 고려 없이 데이터를 일괄적으로 처리하거나 신고 유형 간의 이질성을 분석하지 않은 채 전체 평균값을 기준으로 해석하는 경향이 있다. 이에 따라 신고 데이터의 구조적 특성과 맥락을 충분히 반영하지 못했다는 한계가 있다.

3) 본 연구의 차별성

이러한 한계를 극복하기 위해 본 연구는 다음과 같은 차별적인 접근을 통해 기존 연구의 범위를 확장하고자 한다.

첫째, 112신고의 긴급코드 분류(코드 0~4)를 고려한 통합적 분석을 통해 신고 유형 간 구조적 차이를 반영하고 패턴 분석을 기반으로 차등적인 해석을 수행한다. 이를 통해 신고 데이터의 전체 스펙트럼을 반영하면서도 데이터의 내부 구조를 고려한 정밀 분석이 가능하다.

둘째, 시간-공간적 분석 단위의 정교화를 통해 시간대별·요일별 신고 패턴, 계절성 효과, 그리고 공간 그리드 단위에서의 신고 밀도 차이를 모두 통합 분석한다. 특히 시간대 기반 공간 자기상관 분석과 공간적 핫스팟 탐색을 병행함으로써 특정 시기·공간의 위험 집중 현상을 실증적으로 도출할 수 있다.

셋째, 물리환경 변수의 정량화 및 스케일링 체계 도입을 통해 환경 요인의 공간적 영향력을 체계적으로 측정한다. 토지이용, 건축물 특성, 인구밀도, CCTV 분포 등 다양한 공간 변수들을 정규화하고 결측 처리 및 스케일링을 수행함으로써 공간 간 비교 가능성을 확보하며 통계적 해석의 신뢰성을 제고한다.

넷째, 기계학습 기반 해석 모델 도입 및 변수 상호작용 분석을 통해 기존 회귀모형의 한계를 극복하고 설명력과 예측력을 동시에 강화한다. MGWR과 같은 공간가중 회귀모형뿐만 아니라 XGBoost 및 SHAP 분석을 통해 변수의 상대적 중요도, 비선형 효과, 상호작용 구조를 시각화하고 해석함으로써 보다 다각적인 해석을 제공한다.

이와 같은 접근을 통해 본 연구는 단순한 신고량 예측을 넘어 신고와 공간환경의 관계 구조를 정량화하고 시간-공간-환경 요인이 결합된 복합적 위험 메커니즘을 규명함으로써 실질적이고 정밀한 도시 안전 정책 설계에 기여하고자 한다.

[표 1-1] 선행연구 검토 및 본 연구의 차별성

구분	연구명 및 목적	연구 방법	주요 연구내용 및 결과
국내 연구	<ul style="list-style-type: none"> 연구명: 112 긴급신고시스템 운용상의 문제점 실증분석 및 개선방안 연구자(연도): 노성훈, 조준택 (2016) 연구목적: 112 긴급신고시스템의 운용상 문제점을 실증적으로 분석하고 개선방안을 제시 	<ul style="list-style-type: none"> 서울경찰청의 112신고 통계자료 중 무작위로 추출된 3,212건을 분석 	<ul style="list-style-type: none"> 신고 유형, 접수 시간, 관할 경찰서에 따라 긴급성 코드 범포에 통계적으로 유의한 차이가 있음 긴급성 코드의 분류 기준이 명확하지 않아 경찰력 낭비와 업무 부담 증가 초래 긴급성과 비긴급성 사건 구분 기준을 명확히 하고 코드 체계 개선 필요성 제시
	<ul style="list-style-type: none"> 연구명: Transformer 기반 LLM의 학습을 이용한 112 허위·오인 신고분류·예측 모델 개발 연구자(연도): 정재훈, 박현호 (2024) 연구목적: 딥러닝 기반 모델을 활용하여 112 허위·오인 신고를 분류하고 예측하는 모델 개발 	<ul style="list-style-type: none"> BERT, KoBERT, ELECTRA 등 5가지 트랜스포머 기반 모델을 동일한 하이퍼파라미터로 훈련 	<ul style="list-style-type: none"> 112 상황실 접수요원이 요약한 신고내용 텍스트를 입력으로 사용 악의적 허위신고와 오인신고를 구분하여 실험 진행 모델 성능 비교를 통해 효과적인 허위·오인 신고 분류 가능성 확인
	<ul style="list-style-type: none"> 연구명: 골목길 특성 및 지역 낙인 효과를 고려한 서울시 도시 범죄 영향 요인 연구 연구자(연도): 정진도 (2023) 연구목적: 골목길 특성과 지역 낙인 효과를 고려하여 서울시 도시 범죄에 영향을 미치는 요인 분석 	<ul style="list-style-type: none"> 112신고 데이터와 서울시 골목길 조사 데이터를 활용한 마이크로 스케일 분석 	<ul style="list-style-type: none"> 골목길의 물리적 특성과 지역의 사회적 낙인이 범죄 발생에 미치는 영향 분석 특정 지역의 골목길 특성이 범죄 발생률과 상관관계 있음 도시 설계 및 정책 수립 시 골목길 특성과 지역 낙인 효과 고려 필요성 제시
	<ul style="list-style-type: none"> 연구명: 기계학습을 활용한 데이터 기반 경찰신고건수 예측 연구자(연도): 최재훈 (2018) 연구목적: 기계학습 기법을 활용하여 경찰 신고 건수를 예측하고 최적의 경찰력 배치 방안 모색 	<ul style="list-style-type: none"> 신경망 분석과 음이항 회귀분석을 활용하여 예측 모델 개발 	<ul style="list-style-type: none"> 2016. 6월부터 2017.5까지 총남지방경찰청의 112신고 데이터를 활용 시간, 휴일, 계절, 기온 등 다양한 변수를 고려하여 예측 모델 구축 신경망 분석이 음이항 회귀분석보다 예측력이 뛰어남을 확인
	<ul style="list-style-type: none"> 연구명: 112신고 벡터화를 통한 텍스트 마이닝 기반의 군집분석과 이를 활용한 사건유형 간 유사성 탐색 연구자(연도): 홍세은, 방준성, 김용진, 장광호 (2020) 연구목적: 112신고 내용을 벡터화하여 사건 유형 간 유사성을 탐색하고 군집화 	<ul style="list-style-type: none"> 약 178만 건의 112신고 데이터를 텍스트 마이닝 기법으로 분석 	<ul style="list-style-type: none"> 신고 내용을 벡터화하여 유사성 기반의 군집 분석 실시 아동학대, 스토킹 등이 살인과 같은 중대범죄와 동일 군집으로 분류됨 사건 유형 간 유사성 분석을 통해 경찰 대응 전략 수립에 기여 가능성 제시
해외 연구	<ul style="list-style-type: none"> 연구명: Temporal Analysis of 911 Emergency Calls Through Time Series Modeling 연구자(연도): Robles, P., Tello, A., Solano-Quinde, L., & Zúñiga-Prieto, M. (2020) 연구목적: 시계열 모델링을 통해 911 긴급전화 데이터의 시간적 패턴 분석 	<ul style="list-style-type: none"> Gaussian Process와 ARIMA 모델을 활용한 시계열 예측 	<ul style="list-style-type: none"> 에과도르 쿠엥카시의 2015년 1월부터 2016년 12월까지의 911 긴급전화 데이터를 분석 두 가지 시계열 모델의 장단점을 비교하여 긴급사건 예측에의 활용 가능성 평가

구분	연구명 및 목적	연구 방법	주요 연구내용 및 결과
	<ul style="list-style-type: none"> • 연구명: Predicting 911 Calls Using Spatial Analysis • 연구자(연도): Cramer, D., Brown, A.A., & Hu, G. (2012) • 연구목적: 공간 분석을 통해 911 긴급 전화 발생을 예측하고 자원 배분에 활용 	<ul style="list-style-type: none"> • 핫스팟 분석 등 공간 분석 기법을 활용한 예측 모델 개발 	<ul style="list-style-type: none"> • 공간적 데이터 분석을 통해 긴급전화 발생 패턴 파악 • 정책 결정자들이 자원을 효율적으로 배분할 수 있도록 지원하는 모델 제시
	<ul style="list-style-type: none"> • 연구명: Temperature and Violent Crime in Dallas, Texas: Relationships and Implications of Climate Change • 연구자(연도): Gamble, J. L., & Hesses, J. J. (2012) • 연구목적: 텍사스주 댈러스에서의 기온과 폭력 범죄 간의 관계 분석 및 기후 변화의 영향 평가 	<ul style="list-style-type: none"> • 폭력 범죄 발생률과 일일 평균 기온 간의 상관관계 분석 	<ul style="list-style-type: none"> • 기온이 약 27°C까지는 상승할수록 폭력 범죄가 증가했으나 32°C 이상에서는 오히려 감소하는 곡선형 관계를 보임. • 연구자들은 폭염 시 사람들이 실내에 머무르며 사회적 접촉이 줄어들어 거리 범죄가 감소할 수 있다고 해석함.
	<ul style="list-style-type: none"> • 연구명: Predicting Violent Crime Reports from Geospatial and Temporal Attributes of US 911 Emergency Call Data • 연구자(연도): Corcoran, V. (2019) • 연구목적: 미국 911 긴급전화 데이터의 지리적 및 시간적 속성을 활용하여 폭력 범죄 보고를 예측 	<ul style="list-style-type: none"> • 지리적 및 시간적 데이터 분석을 통한 예측 모델 개발 	<ul style="list-style-type: none"> • 특정 지역과 시간대에서의 긴급전화 데이터를 분석하여 폭력 범죄 발생 예측 • 예측 모델을 통해 경찰 자원 배분 및 범죄 예방 전략 수립에 기여 가능성 제시
	<ul style="list-style-type: none"> • 연구명: The Effect of Temperature on Crime • 연구자(연도): Field, S. (1992) • 연구목적: 기온이 범죄 발생률에 미치는 영향 분석 	<ul style="list-style-type: none"> • 기온과 범죄율 간의 상관관계 분석 	<ul style="list-style-type: none"> • 높은 기온이 공격성과 범죄 발생률 증가와 관련이 있을 수 있음 • 기후 변화로 인한 기온 상승이 범죄율에 미치는 잠재적 영향 고려 필요
	<ul style="list-style-type: none"> • 연구명: Impacts of Climate Variations on Crime Rates in Beijing, China • 연구자(연도): Shen, B., Hu, X., & Wu, H. (2020) • 연구목적: 중국 베이징에서의 기후 변화가 범죄율에 미치는 영향 분석 	<ul style="list-style-type: none"> • 기후 변수와 범죄율 간의 상관관계 분석 	<ul style="list-style-type: none"> • 기온 상승이 특정 범죄 유형의 발생률 증가와 관련이 있을 수 있음 • 기후 변화가 범죄 패턴에 미치는 영향을 고려한 정책 수립 필요
기후 연구	<ul style="list-style-type: none"> • 연구명: 물리 환경이 112신고 발생에 미치는 영향 분석 • 연구목적: 112신고데이터의 구조적 특성과 도시물리환경 간의 관계를 정밀하게 분석하여 시간-공간-환경 요인이 복합적으로 작용하는 도시 안전 메커니즘을 규명하고자 함. 	<ul style="list-style-type: none"> • MGWR, XGBoost, SHAP 기반 통합 분석 • 복합적 위험 메커니즘을 정량화하여 정밀한 도시 안전 정책 설계에 기초자료 제공 	<ul style="list-style-type: none"> • 신고 유형 간 구조적 차이를 반영한 그룹 분석 수행 • 시공간별 신고 집중 현상 및 위험 지역 도출 • 물리환경 변수의 공간적 영향력 정량화 • 기계학습 기반 해석을 통해 변수 중요도 및 상호작용 시각화

제2장

112신고 제도 및 활용

1. 112신고 제도
2. 112신고 체계
3. 112신고 데이터 활용
4. 소결

1. 112신고 제도

112신고는 국민의 생명, 신체 및 재산을 보호하고 공공의 안녕과 질서를 유지하는 데 매우 중요한 제도다(류지웅, 김호, 2023). 그러나 기존에는 112신고 시스템이 명확한 법적 근거 없이 경찰청 예규인 「112종합상황실 운영 및 신고처리 규칙」에 의해서만 운영되어 온 문제점이 존재했다. 이러한 법적 기반 부재는 시스템의 체계적인 운영을 어렵게 하고 사건 현장에서 경찰관의 권한이 명시되지 않아 적극적인 초동 조치에 한계를 초래했다. 이러한 문제를 해결하고 예측 불가능한 범죄 및 재난 상황에 효과적으로 대응하기 위해 새로운 법제도 체계가 마련되었다. 현재 112신고와 관련된 법제도 체계는 최근 제정된 「112신고의 운영 및 처리에 관한 법률」(약칭 「112신고처리법」)을 중심으로 한 상위 법령과 그 내용을 구체화하는 하위 법령 및 규칙으로 구성된다.

1) 112신고 관련 제도 현황

■ 112신고 관련 법령 연혁

112신고시스템 관련 법 제도는 2024년을 기점으로 제도적 기반이 크게 강화되었다. 법령 제정 이전 운영되던 「112치안종합상황실 운영 및 신고처리 규칙」은 112신고센터에 근무하는 모든 경찰공무원과 전·의경이 긴급신고를 적절히 접수·처리하기 위해 따라야 할 구체적인 근무 기준과 운영 절차를 상세히 규정하고 있었다. 2024년 1월에는 이처럼 기존 행정규칙 수준에 머물던 112신고체계를 법률로 격상한 「112신고의 운영 및 처리에 관한 법률」(약칭 「112신고처리법」)이 제정되었다. 이 법률은 국민의 긴급상황 대응 요청권을 국가가 보장해야 할 기본권으로 명시하고 신고의 관할 구분 없이 신속하게 접수·처리하도록 하며 다기관 간 공동 대응을 의무화함으로써 종전보다 한층 통합적이고 신속한 대응체계를 확립하였다.

이어 같은 해 7월에는 상위법의 위임사항을 구체화하기 위한 「112신고의 운영 및 처리에 관한 법률 시행령」이 제정되고 112신고처리법이 시행되었다. 시행령은 긴급성 기준에 따른 대응 단계, 처리결과 통보 절차, 기관 간 협력 체계, 영상촬영 고지 의무, 자료의 보존기간 차등화, 신고 공로자에 대한 포상금 등 세부 실행기준을 규정함으로써 법률의 실효성을 높이는 역

할을 한다. 이와 함께 기존 행정규칙도 법령 시행에 맞추어 개정되었다.

또한 2024년 12월에는 「112신고 공로자 포상 등에 관한 규정」이 제정되어 112신고를 통해 범죄 예방이나 타인 보호에 기여한 시민에게 포상 및 포상금을 지급하기 위한 구체적 기준과 절차를 마련하였다. 이 규정은 심사위원회 구성, 심의 절차, 지급 방식 등을 명시함으로써 시민의 적극적인 신고 참여를 유도하고 공동체 안전 강화에 기여한 개인의 공로를 제도적으로 인정하는 기반을 마련하였다.

[표 2-1] 112신고 관련 법령 현황

관계법령	제정 및 개정연월	주요 내용
112치안종합상황실 운영 및 신고처리 규칙*	2003년 12월 (일부개정) 2024년 7월 (최근개정)	112신고센터에 근무하는 모든 경찰공무원 및 전·의경이 112신고를 적절하게 접수하고 처리하기 위해 준수해야 할 근무 기준과 운영에 필요한 전반적인 사항을 상세히 규정
112신고의 운영 및 처리에 관한 법률	2024년 1월(제정)	그동안 행정규칙으로만 운영되던 112신고체계를 법률로 격상시켜 국민의 긴급상황 대응 요청권을 기본권으로 보장하고 관할 구분 없는 신속 처리와 다기관 공동대응을 의무화한 법률
112신고의 운영 및 처리에 관한 법률 시행령	2024년 7월 (제정)	상위법에서 위임된 사항을 구체화하여 긴급성 기준 대응체계, 처리결과 통보, 기관 간 협력, 영상촬영 고지, 자료 보존기간 차등화, 포상금 등 112신고 처리의 세부 실행기준을 명확히 규정한 하위법령
112신고 공로자 포상 등에 관한 규정	2024년 12월 (제정)	「112신고법」의 하위 행정규칙으로 112신고를 통해 범죄 예방이나 타인 보호에 기여한 신고자에 대한 포상 및 포상금 지급의 구체적 기준, 심사위원회 구성, 지급 절차 등을 규정하여 시민 참여를 장려하는 포상제도의 세부 운영기준 마련

출처: 연구진 작성

주: * [시행 2010.1.1.]까지는 「112신고센터 운영규칙」

■ 「112치안종합상황실 운영 및 신고처리 규칙」

「112치안종합상황실 운영 및 신고처리 규칙」은 지난 20여 년간 4차례의 주요 혁신을 통해 단순한 신고 접수 체계에서 법률 기반의 종합적 치안 행정 시스템으로 발전해 왔다.

2003년 12월 18일 일부개정을 통해 법적 체계의 기초를 마련한 후, 2009년 8월 26일 전면 재편되어 운영상의 문제점을 보완하고 새로운 제도적 틀을 구축했다.

2010년 1월 1일 전부개정에서는 규칙명을 '운영 및 신고처리 규칙'으로 변경하면서 112신고를 긴급성에 따라 코드 1~3으로 분류하는 과학적 대응체계를 도입했다. 이와 함께 기존의 신

외도 낮은 도착률 통계를 폐지하고 평균출동시간 중심의 실효성 있는 통계체계로 전환함으로써 효율적 경찰력 운용의 토대를 마련했다.

2015년 2월 6일에는 조직이 과 단위 직제로 격상되면서 '종합상황실'로 명칭을 변경하고 기능과 관할에 관계없이 총력대응할 수 있는 통합 운영체계를 완성했다. 교통순찰차 등 다양한 출동요소를 포함시키고 경찰관서 간 공조체계를 구체화하는 등 종합적 치안 대응 기능을 갖추게 되었다.

2021년 1월 22일에는 자치경찰제 도입에 따른 「경찰법」 개정을 반영하여 '지방경찰청'을 '시·도경찰청'으로 변경하는 등 법적 용어의 정합성을 확보했다.

2023년 10월 27일에는 '치안종합상황실'로 재명명하면서 다기관 협력체계를 본격 구축했다. 코드 4(민원·상담) 유형을 신설하여 0~4단계 분류체계를 완성하고 경찰청, 소방청, 해양경찰청을 포괄하는 '긴급신고대응기관' 개념을 도입하여 신고 이관 및 공동대응 절차를 체계화했다.

2024년 7월 24일에는 「112신고의 운영 및 처리에 관한 법률」 제정에 따라 전부개정을 단행했다. 영상정보 관리체계, 데이터 통계 분석 및 활용, 자료 보존기간 차등화, 전문인력 관리, 과태료 부과 등 포괄적 운영체계를 구축함으로써 단순한 행정 지침을 넘어 법률에 기반한 공적 행정의 핵심 도구로 완전히 변모했다.

이러한 발전 과정을 통해 112 대응체계는 기초 정립에서 과학적 분류, 통합 운영, 다기관 협력을 거쳐 법률 기반 데이터 행정에 이르는 단계적 진화를 이루며 국민 안전을 위한 범정부적 긴급대응 시스템의 중추적 역할을 담당하게 되었다.

[표 2-2] 「112치안종합상황실 운영 및 신고처리 규칙」의 제·개정

규칙명	시행일	개정형태	주요 변화 내용
112신고센터 운영규칙 예규 제414호	2003.12.18	일부개정	112신고 접수·처리 체계의 부분적 개선 필요 • 기존 규칙의 부분적 개선 • 운영체계 보완
112신고센터 운영규칙 예규 제408호	2009.08.26	폐지제정	제도 정비 및 운영체계 전면 개편 필요성 • 기존 규칙 완전 폐지 후 재제정 • 동일 명칭으로 새로운 체계 구축
112신고센터 운영 및 신고 처리 규칙 예규 제416호	2010.01.01	전부개정	112신고 효율적 대응 및 경찰력 낭비 방지, 치안서비스 수준 제고 • 규칙 제명 변경(운영 → 운영 및 신고처리) • 112신고 유형 분류체계 신설(code 1~3) • 긴급성·출동 필요성 기준 우선순위 도입 • 통계체계 개편: '5분(10분) 도착률' 폐지 → '평균출동시간' 도입 • 신고처리 절차 세부화

112종합상황실 운영 및 신고처리 규칙 예규 제496호	2015.02.06	전부개정	112종합상황실의 과 단위 직제 편성에 따른 환경 변화 반영 및 총력대응 체계 구축 <ul style="list-style-type: none"> • 제명 변경: '신고센터' → '종합상황실' • 과(課) 단위 직제 반영 • 출동요소 다양화: 교통순찰차 등 포함 • 총력대응 원칙: 기능·관할 무관 대응체계 • 경찰관서 간 공조 체계 구체화 • 전산화 반영: 사문화된 서식 정비
112종합상황실 운영 및 신고처리 규칙 예규 제582호	2021.01.22	타법개정	「경찰법」 전부개정 및 자치경찰제 도입에 따른 법체계 정비 <ul style="list-style-type: none"> • 자치경찰제 관련 용어 일괄 수정 • '지방경찰청' → '시·도경찰청' • '지방경찰청장' → '시·도경찰청장' • 관련 조항 전체 정비
112치안종합상황실 운영 및 신고처리 규칙 예규 제617호	2023.10.27	일부개정	신속·효율적 대응체계 구축 및 긴급신고의 공동대응체계 마련 <ul style="list-style-type: none"> • 제명 변경: '종합상황실' → '치안종합상황실' • code 4 신설: 긴급성 없는 민원·상담 신고 • 긴급신고대응기관 정의 신설(경찰청, 소방청, 해양경찰청 등) • 신고 이관·공동대응 절차 명문화 • code 0 신고 개념 도입 (최우선 출동)
112치안종합상황실 운영 및 신고처리 규칙 예규 제633호	2024.07.24	전부개정	<ul style="list-style-type: none"> • 「112신고의 운영 및 처리에 관한 법률」 제정 반영을 위한 정비 • 영상정보 수집·관리체계 신설 • 112신고 데이터 통계 분석·활용 규정 • 자료 보존기간 차등화(긴급성·출동 필요성 기준) • 112시스템 운영·개선 체계 구축 • 전문인력 관리 규정 강화 • 과태료 부과 고지서 발급 규정 신설

출처: 연구진 작성

■ 「112신고의 운영 및 처리에 관한 법률」

「112신고의 운영 및 처리에 관한 법률」(법률 제19870호, 2024년 1월 2일 제정, 7월 3일 시행)은 그동안 경찰청 내부 행정규칙으로만 운영되던 112신고체계를 법률 차원으로 격상시킨 법률이다. 이 법은 국민의 생명·신체·재산 보호라는 국가의 기본 책무를 법적으로 명문화하면서 장애인이나 언어 소통이 어려운 외국인 등 사회적 약자층이 112신고를 이용하는 데 불이익을 받지 않도록 접근성을 보장하도록 규정했다. 또한 누구든지 위급한 상황에서 112신고를 통해 국가로부터 신속한 대응을 요청할 권리를 기본권으로 규정하는 한편, 허위신고나 목적 외 사용을 금지하여 권리와 의무의 균형을 맞췄다.

법률의 핵심은 경찰 사무 구분이나 관할에 관계없이 신속한 접수·처리를 보장하고 다른 기관과의 공동대응 체계를 의무화하여 범정부적 긴급대응 시스템의 법적 근거를 마련한 것이다. 아울러 112치안종합상황실의 실시간 현장 확인을 위한 영상촬영장치 설치, 정보시스템 구

축·운영, 전문인력 양성을 위한 교육·훈련 프로그램 운영 등 기술 기반의 현대적 대응체계를 규정했다. 특히 신고자 보호와 개인정보 보안을 강화하고 범죄 예방이나 타인 보호에 기여한 신고자에 대한 포상 및 포상금 지급 근거를 마련함으로써 112 대응체계가 행정편의 중심에서 국민 권리 중심으로 개별 기관 대응에서 범정부적 통합 대응으로 전환하는 법적 토대를 완성했다.

■ 「112신고처리법」 제정 배경

- 상위 법령 부재

과거 112신고 시스템은 연간 약 2천만 건에 달하는 막대한 신고가 접수됨에도 불구하고 법적 근거 없이 주로 경찰청 예규로만 운영되어 왔다. 이로 인해 시스템의 체계적인 운영에 한계가 있었으며 경찰력 배치를 최적화하고 신고 대응의 효율성을 높이는 데 충분히 활용되지 못했다.

- 체계적인 운영의 필요성

(경찰관의 현장 조치 권한 명확화) 현장 경찰관이 신고 현장에 도착했을 때 인기척이 없거나 문이 잠겨있는 상황에서 실내로 출입할 법적 권한이 명시되지 않아 적극적인 조치를 취하는데 어려움을 겪었다. 실제 2021년 전자발찌 훼손 후 여성 2명을 살해한 가해자의 주거지를 수색하지 못했던 사례가 이러한 법적 권한 부재의 한계로 지적된다. 이러한 문제 해결을 위해 경찰의 권한을 법에 명시할 필요성이 있었다.

(긴급상황 대응 역량 강화 및 골든타임 확보) 112신고는 대부분의 사건이 경찰에게 가장 처음으로 인지되는 통로이며 범죄의 진압과 예방 측면 모두에서 매우 중요한 치안 활동 단계이다. 그러나 신고 내용만으로는 사건 유형을 정확히 판단하기 어렵고 긴급성 및 위험성 판단이 정확히 이루어지지 않으면 상황 해결의 골든타임을 놓칠 수 있다는 문제점이 지속적으로 제기되었다. 또한 급증하는 치안 수요에 효율적으로 대응하고 신고 접수 단계에서 정확한 정보 파악에 실패하는 문제점을 개선하기 위한 법적 토대 마련이 시급했다.

■ 제정 시기 및 목적

112신고 관련 법제의 핵심인 「112신고의 운영 및 처리에 관한 법률」(「112신고처리법」)은 2023년 12월 8일 국회 본회의를 통과하여 2024년 1월 2일 제정되었다.

이 법률은 범죄나 각종 사건·사고 등 위급한 상황으로부터 국민의 생명·신체 및 재산을 보호하고 공공의 안녕과 질서를 유지하는 것을 목적으로 한다. 또한, 긴급 상황에서의 경찰 권한을 명확히 하여 초동 조치의 신속성과 명확성을 확보하는 것을 목표로 한다.

■ 법률 제정 주요 특징

• 상위법 제정

법률-시행령-예규 간 위계 확립을 통해서 종전 관행적 절차들을 법적 위임 기반의 규칙 체계로 정비하였다. 이를 통해서 112신고 시스템이 단순히 경찰 내부의 행정 효율성을 위한 운영 규칙을 넘어 국민 안전 보장이라는 국가의 헌법적 책무를 법률로서 명확히 하였다.

• 신속하고 포괄적인 현장 대응 의무화

경찰청장 등이 112신고를 접수하면 경찰 사무의 구분이나 현장 출동이 필요한 지역의 관할에 관계없이 해당 112신고를 신속하게 접수하여 처리하도록 규정하였다. 이에 대응하여 2024년 7월 24일 전부 개정된 「112치안종합상황실 운영 및 신고처리 규칙」은 112신고법에서 위임된 사항과 그 시행에 필요한 사항을 정하기 위해 마련되었다. 현 실정에 맞는 112치안종합상황실의 기능, 112근무요원의 업무, 112신고의 대응체계 등을 상세히 규정하였다.

• 유기적인 기관 간 공동대응 및 협력 체계의 법제화

「112신고처리법」은 112신고 처리에 있어 다른 기관과의 공동대응 또는 협력이 필요한 경우 경찰청장 등이 관계 기관에 이를 요청할 수 있도록 하고 요청을 받은 기관의 장은 특별한 사유가 없으면 이에 따르도록 법적 근거를 명확히 하였다.

또한 시행령을 통해 공동대응 또는 협력을 요청받은 재난관리책임기관 등이 필요한 조치를 한 후 그 내용과 결과를 회신하도록 하고 수사기관에 신고를 인계하는 경우 지체 없이 관련 기록 등을 통보하도록 구체화하였다. 이러한 공동대응 체계는 2023년 「112치안종합상황실 운영 및 신고처리 규칙」 개정 시 긴급신고대응기관 간 신고 이관 및 공동대응에 관한 규정이 신설되었다.

• 신고자 보호 및 정보 관리 강화 및 공로자 포상 제도 신설

위급 상황에서 구조를 요청한 신고자를 국가가 보호하도록 명시하고 신고자 정보(이름, 전화번호 등)를 112신고 접수·처리 외의 목적으로 이용하는 것을 금지하도록 하였다. 또한, 출동 현장 촬영 사실 공지, 수집 영상정보 보호 및 관리를 위한 영상 정보관리체계 구축 및 운영의 무화, 112신고 관련 자료 보존 기간 및 연장 규정을 명시하였다.

생명·신체·재산 보호에 기여한 신고자에게 포상금을 지급할 수 있는 근거가 「112신고처리법」에 신설되었으며 이를 바탕으로 2024년 12월 「112신고 공로자 포상 등에 관한 규정」이 제정되어 포상금 지급 절차와 기준을 명확히 하였다.

• 첨단 정보 시스템 구축 및 전문 인력 양성 의무화

112신고의 접수·처리 및 정보 공유 등에 필요한 정보 시스템을 구축·운영하도록 명시하였

다. 또한, 112시스템 운영과 관련하여 전문 인력의 양성 및 기술 향상을 위한 교육·훈련 프로그램을 운영하고 112신고 데이터 통계 분석 및 활용, 112신고 서비스의 편의성 개선을 위한 대국민 홍보를 실시하도록 규정하였다.

■ 「112신고의 운영 및 처리에 관한 법률 시행령」

「112신고의 운영 및 처리에 관한 법률 시행령」(대통령령 제34629호, 2024년 7월 2일 제정, 7월 3일 시행)은 상위법인 「112신고의 운영 및 처리에 관한 법률」에서 위임된 사항과 그 시행에 필요한 구체적 사항을 정한 하위법령이다. 이 시행령은 112신고의 접수 및 처리에 관해 경찰청장이 긴급성과 현장 출동 필요성을 고려한 대응체계를 마련하도록 하고 신고자가 요청하는 경우 처리 결과를 통보하도록 규정했다. 또한 다른 기관과의 공동대응 시 필요한 조치 후 그 내용과 결과를 회신하도록 하고 수사기관 인계 시 관련 기록을 지체 없이 통보하도록 하여 기관 간 협력체계를 구체화했다.

이 시행령은 출동 현장 영상촬영 시 촬영 사실 고지 의무, 자료 보존기간 차등화(입력자료 3년, 녹음·녹화 3개월), 최고 5천만원 포상금 지급 등 법률의 추상적 규정을 구체적인 실행 기준으로 명확히 하여 112신고 처리의 실효성을 확보했다.

■ 「112신고 공로자 포상 등에 관한 규정」

이 규정은 「112신고의 운영 및 처리에 관한 법률」 제16조 및 같은 법 시행령에 따라 112신고를 통해 범죄 예방이나 타인의 생명·신체·재산 보호에 기여한 공이 큰 신고자에 대한 포상 및 포상금 지급을 위한 세부 사항을 규정한 하위규정이다.

이 규정은 112신고 공로자의 정의부터 포상 분야, 보상금심사위원회 설치·구성, 포상금 지급 기준과 제한사항, 중복·이중 지급 제한, 배분 지급 기준, 신청 및 지급 방법, 환수 절차, 비밀누설 금지 등 포상제도 운영의 전 과정을 체계적으로 규정하여 112신고를 통한 시민 참여를 실질적으로 장려하는 제도적 기반을 완성했다.

2) 「112신고처리법」 주요 내용

「112신고처리법」은 국민의 안전을 위한 112신고체계를 체계적으로 운영하기 위해 제정된 근거법률로서 기존 경찰청 예규 수준에 머물렀던 관련 규정들을 법률로 격상시켜 법적 기반을 확립했다는 점에서 중요한 의미를 갖는다.

이 법률은 112신고의 접수 및 처리 절차, 긴급성에 따른 5단계 코드 분류체계, 112시스템 구축 및 운영, 관련 기관과의 연계 및 협력, 신고자 보호 및 정보관리, 전문인력 양성을 위한 교육훈련, 공로자에 대한 포상제도, 그리고 법 위반 시 벌칙 및 과태료 등을 종합적으로 규정하고 있다. 특히 신고자의 신원보호와 개인정보 보안을 강화하여 국민이 안심하고 신고할 수 있는 환경을 조성하는 한편, 허위신고에 대한 제재를 통해 제도의 건전성과 신뢰성 확보에도 중점을 두고 있다.

■ 112신고 접수 및 처리

• 112신고의 정의 및 접수 원칙

"112"는 「전기통신사업법」에 따라 부여되는 특수번호이며 "112신고"는 범죄나 각종 사건·사고 등 위급한 상황 발생 시 피해자 또는 인지한 사람이 112를 이용한 음성, 문자, 인터넷, 영상, 스마트기기 등 다양한 통신수단으로 신고하는 것을 말한다. 경찰청장 등은 관할 구역이나 경찰 사무의 구분과 관계없이 112신고를 신속하게 접수하고 처리해야 한다.

• 대응 체계 및 코드 분류

112신고 접수 및 처리 업무의 효율성을 위해 긴급성과 현장 출동의 필요성을 고려한 대응 체계를 마련한다. 112신고는 긴급성에 따라 다음 5단계 코드로 분류된다.

• 현장 조치 및 보고

신고가 접수되면 경찰관은 현장에 신속하게 출동하여 위험 방지, 범죄 예방·진압, 구호 대상자 구조 등 필요한 조치를 해야 한다. 조치 후에는 112치안종합상황실에 내용을 보고한다. 신고자가 요청하면 처리 결과를 통보해야 한다.

• 타 기관 이관 및 공동 대응

다른 기관의 소관 업무에 해당하는 112신고는 해당 기관에 지체 없이 이관하며 재난관리책임기관 등 112신고 관계 기관과의 공동 대응 또는 협력이 필요한 경우 이를 요청할 수 있다.

■ 112 시스템 구축 및 정보 연계

• 112시스템 구축 및 운영

경찰청장은 112신고의 접수·처리, 정보 공유·이관 및 공동 대응 등에 필요한 정보시스템(112 시스템)을 구축하고 운영해야 한다. 112시스템은 신고 접수, 지령, 전파 및 순찰차 배치에 활용되는 전산 시스템이다.

• 정보 기록 및 보존

112신고 접수·처리 상황은 112시스템에 입력, 녹음, 녹화 등의 방법으로 기록하고 보존한다. 112시스템 입력 자료 중 코드 0, 1, 2로 분류된 자료는 3년간, 코드 3, 4로 분류된 자료는 1년간 보존하며 범죄 수사를 위해 필요한 경우 보존 기간을 연장할 수 있다. 녹음·녹화 자료는 3개월간 보존하며 필요시 연장할 수 있다.

• 다른 정보시스템과의 연계

경찰청장 및 시·도경찰청장은 급박한 사람의 생명·신체·재산 보호를 위한 112신고 처리를 위해 필요한 경우 관계 기관의 정보시스템과 112시스템의 연계를 요청할 수 있다. 이 경우 해당 기관의 장과 사전에 협의해야 한다.

■ 112신고자 보호 및 정보 관리

• 신고자 보호 의무

국가는 112신고를 처리할 때 범죄 피해자, 범죄 목격자, 위급 상황에서 구조를 요청한 사람 등 112신고자를 보호해야 한다.

• 신고자 정보 보호

경찰청장 등은 112신고 처리 외의 목적, 신고자 동의 없는 경우 등을 제외하고 112신고자 정보(전화번호, 이름, 주소, 성별, 나이, 음성 등)를 수집·이용 또는 제공해서는 안 된다. 정보를 활용할 때는 해당 업무 수행에 필요한 최소한의 범위에 그쳐야 한다.

• 비밀 누설 금지

112신고 포상 등 관련 업무를 수행하는 공무원이나 위원회 위원은 직무상 취득한 112신고 공로자에 대한 비밀을 누설해서는 안 된다.

• 영상정보 관리

경찰청장 등은 112신고 처리 시 출동 현장 상황 확인 및 지휘를 위해 순찰차 등에 영상촬영장치 설치하여 현장을 촬영할 수 있으며 수집된 영상 정보가 외부에 누설되거나 권한 없는 사람이 이용하지 않도록 관리책임자를 지정하여 관리한다.

■ 교육 훈련 및 포상

• 교육·훈련

경찰청장은 112시스템 운영과 관련하여 전문인력의 양성과 기술 향상에 필요한 교육·훈련 프로그램을 운영해야 한다. 이 프로그램에는 112신고 관련 법령, 112시스템 제원 및 조작 방법, 점검 및 관리 등이 포함된다. 경찰청장 등은 소속 경찰관에게 교육·훈련을 실시하며 필요 시 국내·외 교육기관에 위탁할 수 있다. 112근무요원의 근무기간은 2년 이상으로 하며 전문성 제고를 위해 112근무요원 전문인증제를 운영할 수 있다.

• 신고자 포상

경찰청장 등은 112신고를 통해 범죄 예방 및 다른 사람의 생명·신체·재산 보호에 기여한 공이 큰 112신고자에게 포상을 하거나 예산 범위 내에서 포상금을 지급할 수 있다. 포상금의 최고액은 5천만원이며 구체적인 지급 기준은 경찰청장이 고시한다. 포상 분야는 범죄 예방, 국민 안전 보호 및 위원회가 인정하는 분야로 나뉜다.

• 포상금 지급 제한 및 환수

신고 내용이 사실이 아니거나 이미 신고된 사항, 언론에 이미 공개된 사항, 익명 또는 가명으로 신고한 경우, 법령상 신고 의무가 있거나 직무로 규정된 경우, 공직자가 직무와 관련하여 신고한 경우, 불법 행위 관련자 등은 포상금을 지급하지 않거나 감액할 수 있다. 동일한 원인으로 다른 법령에 따른 포상금을 받거나 받을 예정인 경우 중복 지급을 제한하며 이미 지급된 포상금이 거짓이나 착오로 지급된 경우 환수할 수 있다.

■ 벌칙 및 과태료

• 벌칙

112신고자 정보를 목적 외 용도로 이용한 자는 5년 이하의 징역 또는 5천만원 이하의 벌금에 처해진다. 또한, 정당한 사유 없이 경찰관의 긴급 조치(토지·물건 일시 사용, 사용 제한, 처분, 출입)를 거부하거나 방해한 자는 300만원 이하의 벌금에 처해진다. 정당한 사유 없이 피난 명령을 위반한 자는 100만원 이하의 벌금에 처해진다.

• 과태료

범죄나 각종 사건·사고 등 위급한 상황을 거짓으로 꾸며 112신고를 한 사람에게는 500만원 이하의 과태료를 부과한다. 과태료의 구체적인 부과 기준은 112신고처리법 시행령 별표에 따른다.

2. 112신고 체계

1) 112신고 처리 절차

112신고 체계는 접수, 코드 분류, 현장 출동 및 조치, 기록 및 종결의 흐름으로 운영되며 그 절차는 다음과 같다.



[그림 2-1] 112신고처리 흐름도

출처: 경찰청. (2020). 112신고접수 및 지령 매뉴얼. 경찰청 참고하여 연구진 수정

- 신고 접수

먼저 신고자가 범죄나 각종 사건·사고 등 위급한 상황을 112로 신고하면 해당 지역을 관할하는 지방경찰청 112치안종합상황실의 접수 요원에게 연결되어 신고가 접수된다. 접수 요원은 신고 내용과 함께 신고자의 위치, 피해 상황, 가해자 정보 등을 파악하고 이를 112시스템에 입력한다.

- 코드 분류

접수된 신고는 긴급성 및 현장 출동 필요성에 따라 5단계의 코드로 분류된다.

- 현장 출동 및 조치

신고 내용이 코드 0부터 코드 3 유형에 해당하는 경우, 112근무요원은 출동 경찰관에게 출동할 장소, 신고 내용, 신고 유형 등을 고지하고 현장 출동, 조치 및 종결을 지령한다. 출동한 경찰관은 위험 발생 방지, 범죄 예방·진압, 구호 대상자 구조 등 필요한 조치를 취해야 한다.

- 기록 및 종결

필요한 조치를 완료한 경찰관은 112치안종합상황실에 조치 내용을 보고해야 한다. 112근무요원은 사건이 해결되거나 신고자가 취소하거나 허위/오인 신고임이 판명되거나 경찰 소관 업무가 아님이 확인된 경우 등 특정 상황에서 112신고 처리를 종결할 수 있다.

2) 112신고 코드 분류 체계

112신고는 긴급성과 현장 출동 필요성에 따라 다섯 단계의 코드(코드 0부터 코드 4까지)로 분류된다. 이 코드는 신고의 우선순위를 결정하고 경찰력을 효율적으로 배분하는 데 사용된다. 각 코드의 정의와 대응 방식은 다음과 같다.

- 코드 0 신고

가장 긴급한 상황으로 코드 1 신고 중 이등성 범죄, 강력범죄 현행범인 등 생명이 위급하거나 실시간 전파가 필요한 경우에 해당한다. 즉각적인 현장 출동이 필수적이다.

- 코드 1 신고

긴급 출동이 필요한 상황으로 생명·신체에 대한 위험 발생이 임박, 진행 중, 직후인 경우 또는 현행범인 경우에 부여된다. 다른 사건보다 우선적으로 출동해야 한다.¹⁾²⁾

- 1) 생명 신체의 위험의 기준: 살인 강도 납치강금 성폭력 주거침입 집단사태 등 중요범죄로 생명 신체에 대한 중대한 위험이 임박하거나 진행 중이거나 직후인 경우를 의미(다소 신빙성이 떨어지더라도 Code 1로 분류)(경찰청, 2020)
- 2) 현행범의 기준: 생명 신체에 대한 위험을 수반하지 않더라도 범죄행위가 진행 중이거나 직후인 경우를 의미(경찰청, 2020)

- 코드 2 신고

잠재적 위험이 있어 신속한 조치가 필요하지만 코드 0, 코드 1보다는 긴급성이 낮은 경우로써 상황에 따라 출동 여부를 판단하도록 한다.

- 코드 3 신고

즉각적인 현장 조치는 불필요하나 수사, 전문 상담 등이 필요한 경우에 해당한다. 상황에 따라 출동 여부를 판단하며 전화 상담이나 다른 방법으로 해결될 수 있다.

- 코드 4 신고

긴급성이 없는 민원·상담 신고를 의미하며 타 기관으로 인계될 수 있다.

[표 2-3] 신고 대응 Code 분류

구분	코드	코드 분류 판단 기준 예시	가능 코드					비고
			0	1	2	3	4	
긴급	0	'code1' 중 이동범죄, 강력범죄 현행범 등 실시간 전파가 필요한 경우 → 선지령 및 제반 출동 요소 공조 출동						비긴급 신고 및 다른 업무의 처리의 우선하여 출동
	1	생명·신체에 대한 위험이 임박, 진행 중, 직후인 경우 또는 현행범인 경우						
비긴급	2	생명·신체에 대한 잠재적 위험이 있는 경우 또는 범죄예방 등을 위해 필요한 경우						긴급신고에 지장 없는 범위 내 출동
	3	즉각적인 현장조치는 불필요하나 수사, 전문상담 등이 필요한 경우 → 즉시 출동 불필요, 먼저 신고자와 통화하여 약속 등을 통해 출동 또는 타부서 통보 필요						당일 근무 시간 내 출동
상담	4	긴급성이 없는 민원·상담 신고						

출처: 경찰청 내부자료 참고하여 연구진 작성

3) 112신고 내용 분류 및 기록

112신고 데이터에는 신고 처리와 분석을 위한 다양한 정보가 기록된다. 주요 기록 항목은 다음과 같다.

- 종별 코드

가정폭력, 성폭력, 주취자 등 46가지의 세부적인 신고 유형으로 분류된다.

- 긴급 코드

신고의 긴급성을 나타내는 코드 0부터 코드 4까지의 5단계 분류 체계이다.

- 접수 매체 유형

112 전화 신고, SMS 신고, 지역 경찰서로 직접 접수된 신고, 소방 등 유관기관의 공동 대응 요청 신고 등이 포함된다.

- 접수 완료 일자 및 시간

지방경찰청 112치안종합상황실에서 신고 접수가 완료된 정확한 일시이다.

- 관할 부서

해당 112신고를 처리할 관할 지구대 또는 파출소를 명시한다.

- 신고 주소 및 발생지점

신고자가 언급한 주소, 지번 주소, 도로명 주소 등 신고 발생 장소의 상세 정보가 포함되며 발생지 좌표(X, Y 좌표)도 기록된다.

- 신고 내용

신고자가 진술한 내용을 접수 요원이 요약하여 기록한 텍스트 정보이다.

- 종결 일시

현장 출동 경찰관이 112신고에 대한 조치를 완료한 시간이다.

- 종결 내용

현장 경찰관이 신고 종결 시 보고한 내용을 경찰서 112치안종합상황실 요원이 입력한 텍스트 정보이다.

- 종결 코드

체포, 보호 조치 등 경찰관이 112신고를 종결하면서 취한 조치를 상세하게 분류한 37가지 코드이다.

기타 포함되는 변수로는 출동 소요 시간, 자동 종착 여부, 신고자 성별 등이 있다.

[표 2-4] 112신고자료의 체계

분류항목	세부내용	분류기준_또는_코드	특징_및_활용
종별코드 (사건종별 분류)	신고 내용의 사건종별을 분류하여 코드 부여	총 46가지 신고 유형 코드 (주요범죄, 여성청소년 관련, 질서유지, 교통, 서비스 요청 등)	가정폭력, 성폭력, 주취자 등 포함하여 체계적 분류

분류항목	세부내용	분류기준_또는_코드	특징_및_활용
긴급코드 (신고 대응 코드)	긴급성과 현장 출동 필요성에 따른 분류	5단계 (코드 0~4): 코드0(최우선), 코드1(생명위험 임박), 코드2(잠재적 위험), 코드3(수사·상담 필요), 코드4(민원·상담)	경찰관이 신고내용 확인 후 코드 부여, 출동 우선순위 결정
접수매체 유형	신고 접수 경로 및 방법	유·무선 전화, 문자메시지, 인터넷, 영상, 스마트기기, 112신고 앱, 보이는 112	다양한 통신수단 활용, 직접 방문 및 유관기관 공동대응 요청 포함
접수완료일자 및 시간	지방청 112종합상황실 최종 접수 완료 일시	년월일시분초 단위 기록	22~23시 최다, 5~6시 최소 접수, 금·토요일 집중
관할 부서	신고 접수 및 처리 담당 부서	지방경찰청 112종합상황실 → 경찰서 112종합상황실 → 지구대·파출소	관할과 관계없이 신속 접수·처리 원칙
신고 주소 및 발생지점(발생지)	신고 장소의 위치 정보	지번주소, 도로명주소, X좌표, Y좌표	정확한 위치 파악으로 신속 출동 지원
신고 내용	신고자가 육성으로 말하는 내용을 요약한 텍스트	키워드 위주 단어형 작성	위치, 현장상황, 피해자 상태, 출동 주의사항 포함
종결 일시	현장 출동 경찰관의 조치 완료 시간	년월일시분초 단위 기록	경찰력 운영 효율성 측정 지표
종결 내용	사건 종결 후 현장 경찰관으로부터 청취한 내용	비정형 텍스트 정보	신고 내용 주요 단어와 유사하게 현출되는 경향
종결 코드	사건 처리 결과 분류 코드	37가지 세부 코드 (체포, 현장조치, 계속조사 등)	접수 시 사건종별과 다를 경우 수정 입력 가능

출처: 조준택, 김강일, 박현호. (2018, p266). 경찰청 치안상황관리관실. (2020)을 참고하여 연구진 수정

4) 112신고 데이터 활용

112신고 데이터는 「112신고의 운영 및 처리에 관한 법률」 및 하위 법령에 따라 국민의 생명·신체·재산 보호 및 공공의 안녕과 질서 유지를 위한 경찰 활동의 핵심 기반으로 다양한 목적으로 활용 가능하다. 주요 법령상 활용 내용은 다음과 같다.

- 기록 및 보존

112신고의 접수·처리 상황은 112시스템에 입력, 녹음, 녹화 등의 방법으로 기록하고 보존되어야 한다. 112시스템 입력 자료는 3년간, 녹음·녹화 자료는 3개월간 보존하며 범죄 수사의 필요성이 인정될 경우 보존 기간을 연장할 수 있다.

- 시스템 구축 및 운영

경찰청장은 112신고의 접수·처리, 112신고 정보의 공유·이관 및 공동 대응 등에 필요한 112 시스템을 구축·운영해야 한다.

- 정보 공유 및 연계

경찰청장 및 시·도경찰청장은 사람의 생명·신체·재산 보호를 위한 112신고 처리를 위해 필요한 경우, 관계 기관의 정보시스템과 112시스템의 연계를 요청할 수 있으며 이는 사전에 해당 기관의 장과 협의해야 한다.

- 통계 분석 및 치안 활동 반영

경찰청장 등은 112신고 통계 현황을 정기적으로 분석하고 이를 범죄 예방 대책 수립 등 치안 활동에 반영해야 한다. 또한, 112시스템과 연계된 통계 및 정보를 저장·관리·분석하는 체계를 구축·운영할 수 있다.

- 신고자 보호

112신고자 정보는 해당 112신고의 처리를 위해 활용되거나 신고자가 동의하는 경우 등에 한하여 수집·이용 또는 제공될 수 있으며 112신고 접수·처리 이외의 목적으로 이용하는 것은 금지된다.

- 현장 촬영 및 관리

112치안종합상황실에서 출동 현장 상황을 실시간으로 확인하고 지휘하기 위해 순찰차 등에 영상 촬영 장치를 설치하여 출동 현장을 촬영할 수 있다.

- 공로자 포상

경찰청장 등은 112신고를 통해 범죄를 예방하고 다른 사람의 생명·신체 및 재산을 보호하는데 크게 기여한 112신고자에게 포상을 하거나 포상금을 지급할 수 있으며 이와 관련된 정보 처리(범죄 경력 자료 포함)는 포상 대상자 선정에 한정된다.

- 전문 인력 교육 및 훈련

경찰청장은 112시스템의 운영과 관련하여 전문 인력 양성과 기술 향상에 필요한 교육·훈련 프로그램을 운영해야 한다.

3. 112신고 데이터 활용

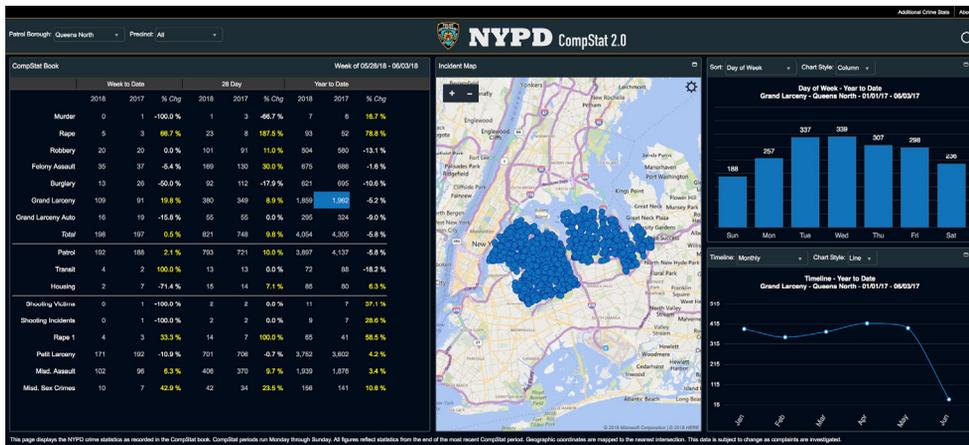
1) 112신고 데이터 활용 사례

■ 미국 911 범죄신고 데이터 활용의 치안 정책 사례

- 뉴욕시의 CompStat 시스템과 Operation IMPACT 전략

뉴욕경찰청(NYPD)은 CompStat 시스템을 통한 범죄 데이터 분석 결과를 반영하여 실제 경찰력을 집중 배치를 통해서 현장에서의 범죄 억제 및 예방을 위한 Operation IMPACT 전략을 시행하고 있다.

CompStat(Computer Statistics) 시스템은 1994년 도입되어 현재까지 전 세계 경찰 조직의 벤치마킹 모델로 활용되고 있다. CompStat 시스템은 911 신고 데이터를 포함한 범죄 정보를 실시간으로 수집하고 분석하는 4단계 프로세스를 운영한다. 1) 정확하고 신속한 정보 수집, 2) 신속한 자원 배치, 3) 효과적인 전술 개발, 4) 지속적인 추적 관리 (Police Executive Research Forum, 2013). 911 신고 데이터는 컴퓨터 기반 맵핑 프로그램을 통해 시각화되어, 범죄 발생 위치와 패턴이 지도상에 실시간으로 표현된다.



[그림 2-2] NYPD CompStat 시스템

출처: <https://urbanomnibus.net/2018/06/policing-is-an-information-business/> (검색일: 2025.6.13.)

CompStat 시스템은 911 신고 데이터를 다음과 같은 방식으로 활용한다. 첫째, 모든 911 신고는 실시간으로 전산 시스템에 입력되어 지리정보시스템(GIS)을 통해 지도상에 표시된다. 둘째, 이 데이터는 시간대별, 지역별, 범죄 유형별로 분류되어 패턴 분석이 이루어진다. 셋째, 주 2회 실시되는 CompStat 회의에서 각 관서별 지휘관들이 담당 지역의 범죄 동향을 보고하고 대응 전략을 논의한다(Weisburd et al., 2004).

Operation IMPACT는 2003년 레이먼드 켈리(Raymond Kelly) 청장에 의해 시작된 데이터 기반 핫스팟 폴리싱 전략이다. 이 정책은 1994년 도입된 CompStat 시스템의 4대 핵심 원칙(정확하고 시의적절한 정보, 신속한 자원 배치, 효과적인 전술, 지속적인 추적관리)을 기반으로 설계되었다(Bureau of Justice Assistance, 2013). Operation IMPACT는 범죄 신고 데이터와 지리정보시스템(GIS)을 활용하여 고범죄 지역을 'Impact Zone'으로 지정하고 경찰 아카데미 졸업생을 우선 배치하는 방식으로 운영되었다. Smith & Purtell(2007)은 이를 "CompStat on steroids"라고 표현하며 기존 CompStat보다 더욱 집중적이고 정밀한 데이터 기반 치안 전략이라고 평가했다.

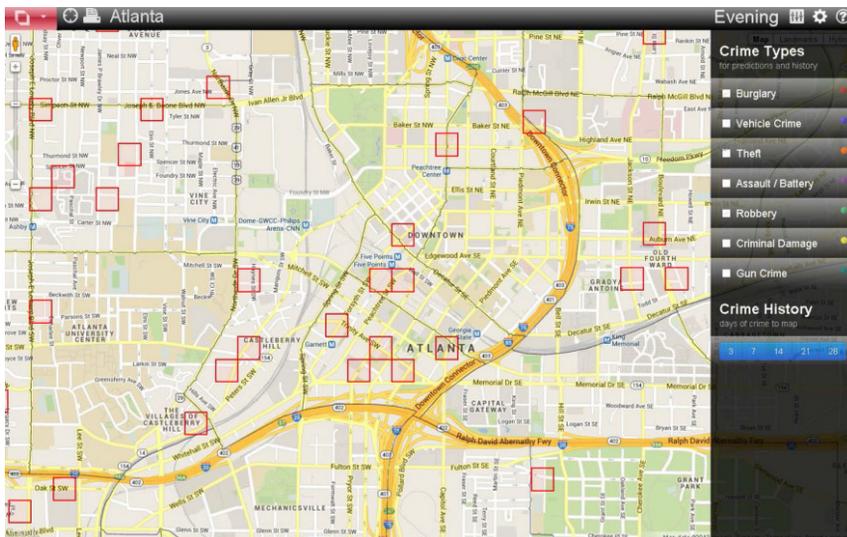
정책 실행 과정은 3단계로 구성되었다. 먼저, 각 관할서별 범죄 데이터 분석을 통한 핫스팟 후보지를 선정 후, 본청 범죄 분석관과 현장 지휘관 간 협의를 통한 최종 Impact Zone 확정 후, 마지막으로 신입 경찰관 집중 배치 및 일일 성과 모니터링하도록 하였다(NYC Global Partners, 2010). 초기 24개 지역에서 시작하여 최대 30개 관할서로 확대되었으며 각 Impact Zone은 6개월마다 재평가되어 범죄 패턴 변화에 따라 조정되었다. 특히 갠단 활동, 마약 거래, 공공주택 고범죄 지역을 우선 타깃으로 하였으며 제75관할서(75th Precinct)의 경우 Operation Trident라는 변형 모델을 통해 관할서 전체를 3개 구역으로 분할하여 운영했다.

CompStat과 Operation IMPACT는 증거 기반 치안 정책 결정의 패러다임 전환을 실현했다는 점에서 의미가 있다. 경험과 직감에 의존하던 전통적 경찰 활동을 객관적 데이터 기반으로 전환시켰으며 범죄 대응에서 범죄 예방으로의 전략적 변화를 이끌어냈다(Bureau of Justice Assistance, 2013). 특히 Operation IMPACT는 뉴욕시 전체 관할서의 25%에 불과한 Impact Zone이 주요 범죄의 30-40%를 차지한다는 분석 결과를 바탕으로 제한된 자원의 전략적 배치를 통해 최대 효과를 달성하는 모범을 제시했다(Smith & Purtell, 2007). 또한 조직 혁신 측면에서 정보 공유 개선, 의사결정 분권화, 실시간 성과 관리 체계를 구축하여 계층적 관료제에서 반응성 높은 조직으로의 변화를 이끌어냈다(Bureau of Justice Assistance, 2013). 이는 치안 분야를 넘어 볼티모어 CitiStat, 메릴랜드 StateStat 등 일부 선도적 지방정부의 성과관리시스템으로 확산되어 데이터 기반 공공 성과 관리의 주요 모델 중 하나로 평가 받고 있다(Henderson, 2003; Behn, 2008).

■ 로스앤젤레스 PredPol 예측적 치안 시스템

로스앤젤레스 경찰청(LAPD)은 2010년 예측적 치안(Predictive Policing) 시스템인 PredPol을 도입했다.³⁾ 이는 제한된 경찰 인력으로 1,302km²에 달하는 광대한 관할 지역을 효과적으로 관리해야 하는 현실적 필요에서 비롯되었다(Beck & McCue, 2009). 이 시스템은 지진학의 여진 예측 모델을 범죄 예측에 적용한 것으로 범죄의 시공간적 클러스터링 현상을 활용하였다(Mohler et al., 2011).

PredPol은 911 신고와 범죄 사건 보고서를 기반으로 특정 시간대와 공간에서 발생한 절도, 차량 절도, 강도 등의 범죄 패턴을 분석한다. 시스템은 과거 범죄가 발생한 150m × 150m 격자 구역을 중심으로 향후 12시간 이내의 재범 가능성을 산출하며 그 중 위험도가 높은 격자 20곳을 순찰 우선 대상으로 지정한다(Brantingham et al., 2018). 이러한 방식은 "가까운 시간과 장소에서 유사 범죄가 재발할 가능성이 높다"는 범죄학 이론에 기반하며 이 알고리즘의 예측 정확도는 기존 분석기 기반 시스템보다 약 1.4배에서 2.2배 높았으며 실제 순찰에 적용했을 때 평균 7.4%의 범죄 감소 효과를 거두었다(Brantingham et al., 2018; Mohler et al., 2015; Ratcliffe, 2004).



[그림 2-3] PredPol

출처: Degeling, M., & Berendt, B. (2018). What is wrong about Robocops as consultants? A technology-centric critique of predictive policing. *Ai & Society*, 33(3), p349.

LAPD는 PredPol이 생성한 예측 지도를 매일 아침 관서별로 확인하여 당일 순찰 계획을 수립하였다. 순찰 경찰관은 모바일 기기를 통해 해당 예측 정보를 실시간으로 확인하고 지정된 격

3) PredPol 시스템은 UCLA의 Jeff Brantingham 교수, Santa Clara University의 George Mohler 교수가 개발한 알고리즘을 기반으로 구축되었다.

자 구역에서 최소 15분 이상 순찰 활동을 수행해야 했다(Perry et al., 2013). GPS 추적을 통해 순찰 실적과 예측 지역의 범죄 발생 간 상관관계를 측정한 결과, 예측 지역에서 순찰 시간이 10% 증가할 때마다 해당 지역 범죄가 약 13% 감소하는 경향이 확인되었다(Hunt et al., 2014).

LAPD 공식 보고에 따르면, PredPol 도입 후 2013~2014년 사이에 전체 범죄율이 약 20% 감소했으며 특히 차량절도는 20%, 강도는 32% 감소하는 등 통계적으로 유의미한 성과가 확인되었다(Los Angeles Police Department, 2015). 그러나 독립적인 학술 연구들은 보다 신중한 해석을 제시하였다. Ratcliffe (2019)는 예측적 치안 시스템이 통계적으로는 효과가 있지만 범죄 유형과 지역별로 효과 편차가 크다고 지적하였다. 더불어 Richardson et al. (2019)는 PredPol이 과거 경찰 활동의 편향된 데이터를 학습하여 인종적·지역적 편향을 재생산하는 문제가 있다고 비판하였다. 흑인 및 히스패닉 거주 지역에 대한 과잉 순찰은 시민권 침해 우려로 이어졌다.

이러한 문제로 인해 LAPD는 2020년 PredPol의 사용을 중단하였다. Lum & Isaac (2016)은 예측적 치안의 성공은 단지 기술적 정확성에 있지 않으며 시민의 신뢰, 공정한 법집행, 알고리즘의 투명성 등이 동반되어야 한다고 강조하였다. LAPD는 이후 데이터 기반 치안 전략을 유지하되 공정성과 지역사회 참여를 기반으로 한 새로운 시스템 개발에 주력하고 있다(Los Angeles Police Department, 2021).

■ 영국의 지리정보 기반 범죄예방 정책

영국의 지리정보 기반 범죄예방 정책은 범죄무질서법(Crime and Disorder Act 1998)을 근간으로 발전되었다. 이 법은 영국 경찰의 범죄예방 체계를 근본적으로 변화시킨 핵심 법률로서 특히 section 115는 지방자치단체, 경찰, 건강서비스 등 공공기관 간의 정보 공유를 법적으로 보장함으로써 지리정보시스템(GIS) 기반 협력의 법적 토대를 마련하였다.

영국의 지리정보 기반 범죄예방 정책에서 범죄 신고 데이터 활용은 다층적 법적 체계 위에 구축되었다. Police Act 1996 Section 44는 내무장관이 경찰청장에게 치안 활동과 관련된 통계 데이터 형태의 정보를 명시할 수 있으며 이를 적절하다고 보는 방식으로 공표하도록 요구할 수 있다는 권한을 부여함으로써 범죄 신고 데이터 수집과 공개의 직접적 법적 근거를 마련했다. 이러한 법적 권한을 바탕으로 내무부는 National Crime Recording Standards (NCRS)와 Home Office Counting Rules (HOOCR)를 제정하여 전국적으로 일관된 범죄 신고 데이터 수집 체계를 확립했다.

또한, Crime and Courts Act 2013은 국가범죄청(NCA) 설립을 통해 범죄 신고 데이터의 국가적 통합 활용 체계를 구축했다. 특히 Section 7은 NCA 기능 수행 목적으로 정보를 NCA에

공개할 수 있으며 NCA는 다른 NCA 기능과 관련하여 획득한 정보를 사용할 수 있다고 규정하여 다기관 범죄 신고 데이터의 통합 분석 법적 근거를 제공했다. 이는 영국의 분산된 데이터베이스를 국가적 차원에서 통합하려는 움직임의 구체적 법적 토대가 되었다(정제용, 2019).

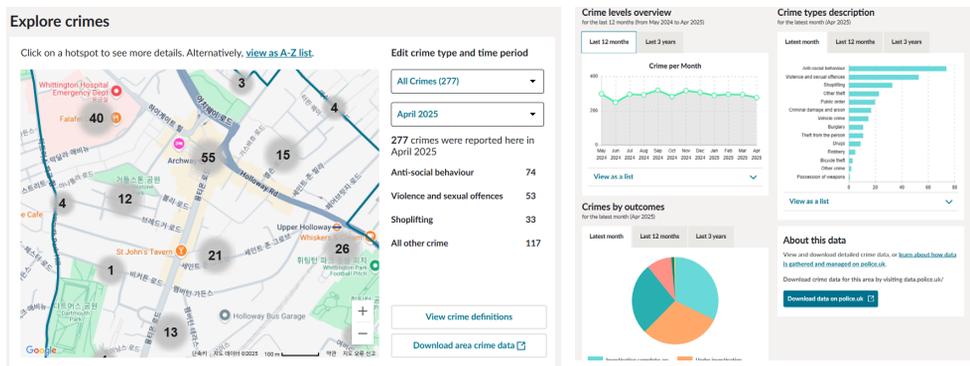
- 국가정보모델(NIM)과 범죄 신고 데이터의 체계적 활용

2000년대 초 도입된 국가정보모델(National Intelligence Model, NIM)은 범죄 신고 데이터의 체계적 분석과 활용을 위한 표준화된 프레임워크를 제공했다. NIM은 범죄 신고 데이터를 지역(Level 1), 광역(Level 2), 국가-국제(Level 3)의 3단계로 구분하여 각 단계에 적합한 분석 기법을 적용하도록 체계화하였다(Weir & Bangs, 2007).

Level 1에서는 개별 경찰서 관할 구역 내 범죄 신고 데이터를 기반으로 한 지역 핫스팟 분석이 수행되며, Level 2에서는 여러 경찰서 관할 구역을 넘나드는 범죄 신고 패턴을 분석하여 광역 범죄자의 이동 경로와 활동 패턴을 파악한다. Level 3에서는 국가 차원의 범죄 신고 데이터베이스를 활용하여 조직범죄의 광역적 네트워크를 분석한다.

- Police.uk: 범죄 신고 데이터 공개 사례

2011년 시작된 Police.uk는 범죄 신고 데이터를 지리정보시스템과 결합한 세계 최초의 영국 전역에 대한 범죄 지도 서비스이다. 이 서비스는 2010년 연립정부의 투명성 정책에 따라 시작되었으며 매월 갱신되는 범죄 신고 데이터를 거리 단위로 시각화하여 제공한다. 이 서비스는 "범죄 통계의 신뢰성 향상, 지역사회 중심의 경찰서비스 제공, 시민 정보 제공"이라는 세 가지 목표를 동시에 추구한다(Tompson et al., 2015).



[그림 2-4] Police.uk에서 공개된 범죄 지도 및 통계

출처: <https://www.police.uk/pu/your-area/metropolitan-police-service/junction/?tab=crimemap> (검색일: 2025.6.13.)

<https://www.police.uk/pu/your-area/metropolitan-police-service/junction/?tab=statistics> (검색일: 2025.6.13.)

Police.uk의 혁신성은 단순한 범죄 발생 지점 표시를 넘어서 범죄 신고 데이터의 맥락적 정보를 함께 제공한다는 점에 있다. 각 범죄 신고에 대해 범죄 유형, 발생 시기, 수사 진행 상황, 최종 처리 결과까지 포함하여 시민들이 지역의 치안 상황을 종합적으로 파악할 수 있게 했다. 이러한 접근방식은 범죄 신고 데이터를 단순한 통계 자료가 아닌 지역사회 안전 관리의 실질적 도구로 전환시켰다. 데이터 품질 관리 측면에서 Police.uk는 프라이버시 보호와 정확성 확보를 위한 정교한 시스템을 구축했다. 실제 범죄 발생 지점에서 약간 이동된 위치에 표시하여 개인정보를 보호하면서도 National Crime Recording Standards에 따라 기록된 범죄만을 포함하여 데이터 신뢰성을 확보했다.

Current status:	Timeline:
Investigation complete; no suspect identified	Case timeline
Investigation complete; no suspect identified	Case timeline
Investigation complete; no suspect identified	Case timeline

Violence and sexual offences (3)

Includes offences against the person such as common assaults, Grievous Bodily Harm and sexual offences.

Current status:	Timeline:
Under investigation	Case timeline
Investigation complete; no suspect identified	Case timeline
Under investigation	Case timeline

Theft from the person

Investigation complete; no suspect identified

Action:	Date:	Status:
Crime reported	April 2025	
Police actions	April 2025	Under investigation This crime is currently being investigated by the police.
	April 2025	Investigation complete; no suspect identified There is insufficient evidence to bring anyone to justice at this time. The investigation may be reopened if more information becomes available.

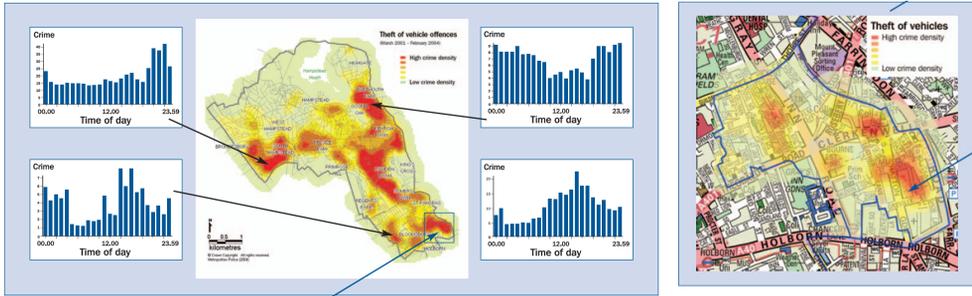
[그림 2-5] Police.uk에서 범죄 신고 및 처리 타임라인

출처: <https://www.police.uk/your-area/metropolitan-police-service/junction/crimedetails/?sid=bf9eafb6-e380-45c8-ad92-06f9908e6226> (검색일: 2025.6.13.)

• 런던 캠프의 차량범죄 신고 데이터 시공간 분석

런던 캠프 자치구는 범죄신고 데이터를 정밀하게 분석하여 지역 내 범죄유형 및 발생 패턴을 식별하고 이를 기반으로 구체적 공간 개선 및 자원 배치 전략을 수립한 대표적 사례를 제시하였다. 이 사례는 단순히 데이터를 시각화하는 수준을 넘어서 시간-공간 프로파일링을 통한 맞춤형 대책 수립과 물리적 환경 개선을 통한 범죄 예방의 실제 효과를 입증했다는 점에서 정책적 의미가 크다.

2000년대 초 캠프에서는 이륜차 절도가 차량 범죄의 주된 유형으로 파악되었다. 범죄신고 데이터를 활용한 분석 결과, 전체 차량 절도 중 약 75%가 오토바이 절도에 해당하며 특히 오후 시간대에 집중된다는 시간적 특성과 함께, 해당 범죄가 집중적으로 발생한 지역에는 이륜차 전용 주차장이 밀집되어 있었음이 확인되었다. 그러나 이들 주차장에는 차량을 고정시킬 수 있는 앵커포인트가 설치되어 있지 않아 범죄 발생을 용이하게 하는 물리적 환경적 취약 요인이 존재하였다(Home Office, 2005).



[그림 2-6] 런던 캠든 차량 절도 핫스팟 분석

출처: Home Office. (2005). Crime mapping: Improving performance – A good practice guide for front line officers. Home Office Communications Directorate. p8

이에 따라 캠든 자치구는 다음과 같은 데이터 기반 대응정책을 실행하였다. 첫째, 범죄신고 데이터를 GIS 기반 범죄 지도화하여 핫스팟 지역과 시간대를 식별하였다. 둘째, 시간-공간 분석 결과를 바탕으로 문제 프로파일링을 수행하고 범죄발생 요인이 물리적 환경과 밀접한 관련이 있음을 인식하였다. 셋째, 해당 지역 모터사이클 주차구역에 앵커포인트 설치와 같은 설계 개선을 시행하였다.

그 결과, 해당 지역에서의 이륜차 절도 발생 건수는 눈에 띄게 감소하였으며 이는 범죄신고 데이터를 활용한 공간정보 기반 개입이 저비용·고효율 범죄예방 정책으로 작동할 수 있음을 보여준다.

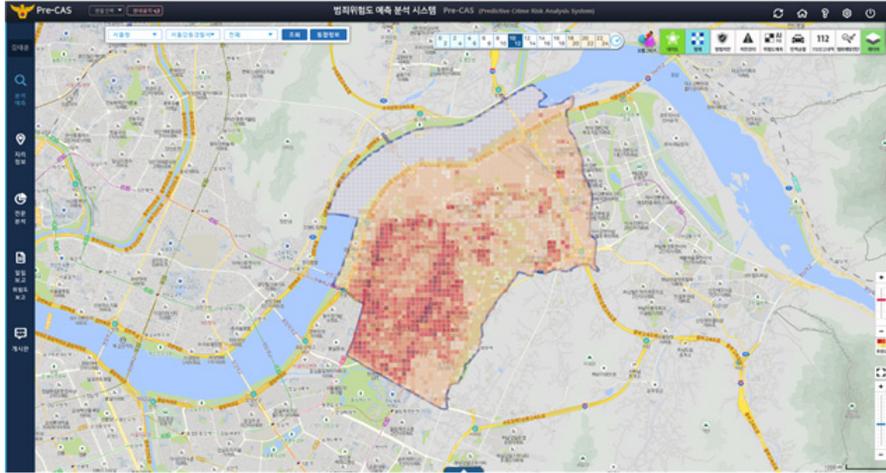
또한, 본 사례는 경찰 및 지역 파트너들이 실제로 인식하고 있던 범죄 핫스팟이 실제 신고 기반으로 나타난 범죄 분포와 상당한 차이를 보였다는 점도 드러냈다. 이로 인해 순찰 경로와 자원 배치 방식을 경험 기반에서 데이터 기반으로 전환되었으며 이는 이후 캠든뿐만 아니라 런던 전역의 지역 치안전략 수립 방식에 영향을 미쳤다(Chainey & Ratcliffe, 2005).

■ 한국 경찰청 프리카스(Pre-CAS, Predictive Crime Risk Analysis System)

경찰청이 2021년 5월부터 전국적으로 운영하고 있는 범죄위험도 예측·분석시스템인 프리카스(Pre-CAS, Predictive Crime Risk Analysis System)는 112신고 데이터를 핵심 구성 요소로 활용하는 대표적인 예측적 치안 시스템이다. 프리카스는 인공지능(AI)과 빅데이터를 활용하여 우범지역을 식별하고 범죄 발생 위험도를 예측하는 시스템으로 치안·공공데이터를 통합한 빅데이터를 최신 알고리즘을 적용한 인공지능으로 분석한다(경찰청, 2021).

이 시스템은 치안데이터와 공공데이터를 통합적으로 활용한다. 치안데이터에는 112신고 건수, 범죄 통계(KICS), 경찰관 수, 유흥시설 수, 방범시설 현황 등이 포함되며, 공공데이터는 인구 이동, 기상, 요일, 교통사고 건수, 고용지표, 공시지가, 학교·공원·소상공인 업소 수 등 다양한 도시사회 환경 정보를 망라한다(경찰청, 2021). 이러한 데이터들은 100m × 100m

단위의 격자망과 2시간 간격의 시간대 분석 단위로 세분화되어 범죄위험도 등급(1~10단계)과 예측 범죄 발생 건수를 산출하는 데 활용된다.



[그림 2-7] 프리카스(Pre-CAS)

출처: 이유지(2021) 오늘밤 당신 동네 범죄... AI는 이미 알고 있다, 9월18일 기사.

<https://www.hankookilbo.com/News/Read/A2021091617580002935> (검색일: 2025.6.13.)

예측 결과는 지구대·파출소에 배치된 순찰차의 112 내비게이션 시스템과 연동되어 고위험 격자 지역에 대해 사전 순찰을 유도한다. 실제로 경찰관은 해당 지역에서 일정 시간 이상 순찰하도록 지시받으며 순찰 이행 여부는 시스템을 통해 실시간으로 확인할 수 있다. 또한 프리카스는 지역의 도시적 특성을 반영하여 도시형, 주거형, 관광형 등 6개 유형으로 구분된 맞춤형 알고리즘을 적용하고 있다. 경찰청 시범운영 결과에 따르면 범죄 예측 정확도는 유형별로 73~89% 수준을 기록했으며 전체 평균 정확도는 83.1%로 나타났다(경찰청, 2021).

이러한 체계는 예측 정확도 향상뿐 아니라 실제 순찰 정책에도 긍정적 영향을 미쳤다. 2022년 기준 프리카스를 통한 순찰요청 건수 중 93.6%가 이행되었으며 이는 전년 대비 3.8% 증가한 수치이다(경찰청, 2023). 범죄 발생률 및 112신고 건수의 동시 감소도 확인되어 프리카스가 실질적인 범죄 억제 효과를 나타냈다는 평가가 제시되었다(경찰청, 2021).

프리카스는 지역 자치경찰의 자율성 제한, 알고리즘 편향에 따른 감시 불균형, 장기적 효과 검증의 미비 등의 한계를 동시에 안고 있는 것으로 평가되고 있다(박성미 & 안정민, 2022). 그럼에도 불구하고 프리카스는 112신고 데이터를 포함한 빅데이터 기반 범죄예측 시스템의 성공 가능성을 입증한 사례로서 의의가 있다. 이는 국내 치안 정책이 데이터 기반 예방 중심으로 전환되는 데 실질적 기여를 한 모델로 향후 알고리즘의 지역 맞춤형 정교화, 편향 통제, 투명성 강화 등의 보안을 통해 더욱 고도화될 필요가 있다. 더불어 프리카스의 성과는 공공데이터 치안 정책의 향후 방향성에 대해 중요한 시사점을 제공한다.

2) 112신고 데이터 활용의 가능성

112신고 데이터는 현장성과 시의성, 공간성을 동시에 갖춘 대표적인 공공안전 데이터로서 치안 활동의 정밀화와 안전정책의 과학화를 이끄는 핵심 자원으로 평가받고 있다. 특히 최근 국내·외 주요 도시에서 시행되고 있는 데이터 기반 치안 시스템은 112신고 데이터를 중심으로 실질적 효과를 입증하고 있으며 이를 통해 예방 중심의 선제적 치안 체계로의 전환을 가속화하고 있다. 이처럼 112신고 데이터는 정밀한 현장 대응과 과학적 정책 수립을 동시에 가능케 하는 전략적 공공 데이터로서의 가능성을 보여주고 있다.

■ 정밀한 현장 대응과 자원 배치의 최적화

112신고 데이터는 시계열과 공간정보를 기반으로 범죄 발생의 밀집도를 분석하고 고위험 지역을 조기 식별함으로써 치안 자원의 선제적 배치와 대응을 가능하게 한다. 뉴욕시의 CompStat 시스템과 Operation IMPACT 전략은 이러한 데이터를 기반으로 경찰력을 집중 배치하고 성과를 체계적으로 관리함으로써 범죄율 감소에 기여하였다. 한국의 Pre-CAS 시스템은 112신고 건수와 도시환경 데이터를 통합 분석하여 시간대별·지역별 범죄위험도를 예측하고 실제 순찰 경로에 반영함으로써 현장 대응의 실효성을 높이고 있다. 이는 제한된 인력과 자원을 효율적으로 운용하면서도 치안 활동의 품질을 높이는 대표적인 스마트 치안의 사례로 평가된다.

■ 공간적 위험 요인의 조기 발견과 환경 개선의 기초

112신고 데이터는 단순한 사건 기록을 넘어서 도시 공간 내 위험 요인을 구조적으로 파악할 수 있는 귀중한 단서 역할을 한다. 영국 캠든 자치구의 사례처럼 특정 시간대와 공간에 집중된 범죄 신고 데이터를 바탕으로 지역 내 이륜차 절도 핫스팟을 식별하고 주차장의 물리적 개선을 통해 범죄 발생을 실질적으로 줄일 수 있었다. 이처럼 신고 데이터는 공간계획, 설계, 시설개선 등 도시안전 정책 수립에 있어 과학적 근거를 제공하며 CPTED(범죄예방환경설계) 정책과도 효과적으로 연계될 수 있다.

■ 데이터 기반 정책 설계와 공공성과 신뢰성 강화

112신고 데이터는 객관적이고 반복 측정 가능한 특성을 바탕으로 치안 정책 전반의 설계와 성과 평가에 있어서 핵심 지표로 활용된다. LAPD의 PredPol, NYPD의 CompStat, 한국 경찰청의 Pre-CAS 등은 모두 신고 데이터를 활용하여 치안 예측 알고리즘을 개발하고 그 결과를 바탕으로 자원 투입의 우선순위를 합리화하였다. 특히 한국의 Pre-CAS는 범죄 유형별·지역 유형별 맞춤형 알고리즘을 적용하여 83% 이상의 예측 정확도를 기록하며 전국적인 데이터 기반 치안 모델로 자리매김하고 있다. 나아가 신고 데이터를 기반으로 한 정책은 성과 측

정이 가능하고 공공 정보로의 환원 또한 용이하여 시민과의 신뢰를 강화하고 참여 기반의 치안 거버넌스 구축에도 기여할 수 있다.

112신고 데이터는 실시간 대응부터 공간 분석, 정책 기획까지 전 주기적 치안 정책에 연계 가능한 고부가가치 공공데이터이며 범죄 예방과 안전한 도시 환경 조성을 위한 핵심 인프라로 기능할 수 있다. 향후 이러한 데이터를 활용한 치안 전략은 보다 정밀하고 맞춤형 대응체계, 도시 공간 내 위험요소의 선제적 제거, 성과 기반의 투명한 치안 정책 운영이라는 세 가지 방향에서 더욱 확대되고 고도화될 것으로 기대된다.

3) 112신고 데이터 활용의 한계

이처럼 112신고 데이터를 활용한 연구를 통해서 범죄 발생 패턴, 시공간적 집중 현상, 도시환경과 범죄의 상관관계, 효과적인 경찰력 배치 전략 등 범죄와 관련한 다양한 통찰적인 정보를 도출할 수 있다.

하지만 실제로 112신고 데이터를 연구에 활용하기에는 최근 제정된 「112신고처리법」의 엄격한 규제, 복잡한 데이터 접근 절차, 신고자 정보보호 강화 등으로 인해 현재 현실적 제약이 큰 실정이다.

■ 신고자 정보보호 규정의 강화

「112신고처리법」 제10조제2항은 "경찰청장 등은 다음 각 호의 어느 하나에 해당하는 경우를 제외하고는 112신고자 정보를 수집·이용 또는 제공해서는 아니 된다"라고 규정하고 있으며 그 예외 사유로 "1. 해당 112신고의 처리를 위하여 활용하는 경우, 2. 112신고자가 동의하는 경우, 3. 다른 법률에 특별한 규정이 있는 경우"만을 명시하고 있다. 이는 연구 목적의 데이터 활용에 대한 명시적 근거를 제공하지 않아 사실상 연구 활용을 제한하는 효과를 가져온다.

또한 동 법 제10조제4항은 "경찰청장 등은 제1항에 따라 112신고자 정보를 활용하는 경우에도 해당 업무 수행에 필요한 최소한의 범위에 그쳐야 한다"고 규정하여 연구 목적의 포괄적 데이터 활용을 더욱 제약하고 있다. 동법 제17조는 "제10조제2항을 위반하여 112신고자 정보를 목적 외의 용도로 이용한 자는 5년 이하의 징역 또는 5천만원 이하의 벌금에 처한다"는 강력한 처벌 조항을 두어 공공기관과 연구기관의 데이터 활용에 대한 소극적 접근을 유도하고 있다.

「112신고의 운영 및 처리에 관한 법률」 제10조(112신고자에 대한 보호 등)

- ① 국가는 112 신고를 처리할 때 112 신고를 한 사람(이하 "112 신고자"라 한다)이 범죄(이미 행하여졌거나 진행 중인 범죄와 눈앞에서 행하여지려고 하고 있다고 인정되는 범죄를 포함한다. 이하 같다) 피해자, 범죄를 목격한 사람, 그 밖에 각종 사건·사고 등 위급한 상황에서 구조를 요청한 사람에 해당하는 경우 그 신고자를 보호하여야 한다.
- ② 경찰청장등은 다음 각 호의 어느 하나에 해당하는 경우를 제외하고 112 신고에 사용된 전화번호, 112 신고자의 이름·주소·성별·나이·음성과 그 밖에 112 신고자를 특정하거나 유추하는 데 사용될 수 있는 일체의 정보(이하 "112 신고자 정보"라 한다)를 수집·이용 또는 제공하여서는 아니 된다.
 1. 112 신고의 처리를 위하여 112 신고자 정보를 활용하는 경우
 2. 112 신고자가 동의하는 경우
 3. 이 법 또는 다른 법률에 특별한 규정이 있는 경우
- ③ 누구든지 제2항에 따른 112 신고자 정보를 112 신고 접수·처리 이외의 목적에 이용하여서는 아니 된다.
- ④ 제2항 각 호에 따라 수집·이용 또는 제공하는 112 신고자 정보는 해당 업무를 수행하기 위하여 필요한 최소한의 범위에 그쳐야 한다.

「112신고의 운영 및 처리에 관한 법률」 제17조(벌칙)

제10조제3항을 위반하여 112 신고자 정보를 목적 외의 용도로 이용한 자는 5년 이하의 징역 또는 5천만원 이하의 벌금에 처한다.

■ 데이터 보존 및 활용 규정의 제약

동 법 시행령 제7조는 "112 시스템 입력 자료 중 긴급성 및 현장 출동의 필요성을 고려하여 코드 0·1·2로 분류된 자료는 3년간, 코드 3·4로 분류된 자료는 1년간 보존한다"고 규정하여 연구에 필요한 장기간 데이터 축적을 제한하고 있다. 또한 동 법 제15조는 "경찰청장 등은 112 신고 통계 현황을 정기적으로 분석하고 이를 범죄예방 대책 수립 등 치안 활동에 반영하여야 한다"고 규정하고 있지만 외부 연구기관과의 협력적 데이터 분석에 대해서는 명확한 법적 근거를 제시하지 않아 연구 활용의 법적 불확실성을 증대시키고 있다.

112 신고 데이터의 연구 및 정책 활용 활성화를 위해서는 개인정보 보호와 공익적 활용 간의 균형을 고려한 법제도 개선이 필요하다. 연구 및 정책 활용 목적 차원에서의 데이터 활용에 대한 명확한 기준과 절차를 마련하고 국내 실정에 맞는 데이터 거버넌스 체계 구축이 필요하며 데이터 기반 치안 연구의 기반을 조성하고 궁극적으로는 국민 안전 증진과 치안 정책 발전에 기여할 수 있는 환경을 마련해야 할 것이다.

「112신고의 운영 및 처리에 관한 법률 시행령」 제7조(112시스템의 구축·운영)

- ① 경찰청장은 법 제13조제1항에 따라 다음 각 호의 사항이 포함된 112 시스템을 구축·운영해야 한다.
 1. 112 신고의 접수·처리에 관한 사항
 2. 112 신고 정보의 공유·이관에 관한 사항
 3. 112 신고 관계 기관과의 공동 대응 또는 협력에 관한 사항
 4. 법 제14조제1항에 따른 다른 정보시스템과의 연계에 관한 사항
 5. 그 밖에 112 시스템의 구축·운영에 필요한 사항
- ② 경찰청장은 112 시스템의 효율적 운영을 위해 필요한 경우에는 112 시스템의 운영 상황 등을 점검하여 112 시스템 보완·개선 계획을 수립·시행할 수 있다.
- ③ 제1항 및 제2항에서 규정한 사항 외에 112 시스템의 구축·운영에 필요한 사항은 경찰청장이 정한다.

4. 소결

최근 제정된 「112신고의 운영 및 처리에 관한 법률」(이하 「112신고처리법」)은 그동안 경찰청 내부 예규로만 운영되었던 112신고 시스템을 법률로 격상하여 신고 처리 절차의 명확성, 경찰관 현장 대응 권한 확보, 다기관 공동대응의 법적 근거 등을 명확히 하였다. 이에 따라 112신고 체계는 보다 신속하고 효과적으로 운영될 수 있는 제도적 기반을 갖추게 되었으며 이는 경찰력 운용의 효율성 향상과 현장 대응의 실효성 확보라는 측면에서 큰 진전이라고 평가할 수 있다. 또한 신고자 보호와 정보 관리 강화, 전문 인력의 양성 및 첨단 정보시스템 구축 등 세부적인 운영체계를 마련함으로써 국민의 생명과 재산 보호라는 국가의 책무를 실질적으로 이행할 수 있는 여건을 마련하였다.

범죄 예방 정책 수립이라는 측면에서 112신고 데이터는 매우 중요한 의미를 가진다. 미국 뉴욕의 CompStat 시스템과 Operation IMPACT 전략, 로스앤젤레스의 PredPol 예측적 치안 시스템, 영국의 Police.uk와 같은 범죄 신고 데이터 활용 사례는 모두 데이터 기반의 범죄 예방 정책이 범죄 억제와 효율적인 경찰력 배치에 크게 기여했음을 증명하고 있다. 국내에서도 프리카스(Pre-CAS) 시스템을 통해 112신고 데이터를 기반으로 한 범죄 위험도 예측과 전략적 순찰 배치가 실질적인 치안 성과를 가져온 사례가 존재한다. 즉, 112신고 데이터는 범죄 발생 패턴 분석과 위험 지역 예측, 효율적인 자원 배치 등 증거 기반의 과학적 정책을 수립하는데 필수적인 공공데이터이며 도시 안전관리 정책 전반에서의 활용 가능성도 크다고 할 수 있다.

그러나 112신고 데이터를 범죄 예방 정책 수립과 연구 목적으로 적극 활용하는 데는 현실적인 어려움이 있다. 특히, 「112신고처리법」은 신고자 개인정보 보호와 정보 이용의 최소화 원칙을 강력히 규정하고 있으며 연구 목적의 데이터 제공 근거가 명확히 제시되지 않은 상태이다. 이로 인해 정부부처 및 지자체, 연구기관이 112 신고 데이터에 접근하는 과정에서 법적 불확실성과 실무적 어려움이 존재하며 데이터의 보존 기간이 코드별로 최대 3년으로 제한된 점도 장기적·대규모 연구 수행을 어렵게 하고 있다.

향후 범죄 예방 정책 및 연구 목적으로 112신고 데이터를 효과적으로 활용하기 위해서는 개인정보 보호라는 법적 가치를 유지하면서도 연구적 활용의 길을 열어주는 제도적 개선이 필요하다. 정책 수립 및 연구 목적으로 데이터를 제공할 때 신고자의 개인 정보를 비식별화하고 엄격한 데이터 접근 통제 시스템을 마련하는 등 구체적이고 명확한 기준과 절차를 법률이나 시행령에 명시할 필요가 있다. 또한, 데이터의 축적 및 활용을 위해 코드별 보존 기간을 유연하게 조정하거나 연구 목적으로 활용 가능한 장기 보존 데이터 세트를 별도로 구축하는 방안도 검토할 수 있을 것이다.

결론적으로 112신고 제도의 법적 기반이 마련된 것은 매우 긍정적인 변화이며 범죄 예방 정책의 과학적이고 증거 기반 접근을 위한 중요한 토대가 될 수 있다. 하지만 현재 법적·제도적 제약을 극복하고 데이터의 연구 및 정책적 활용성을 높이기 위한 추가적인 법적·제도적 개선 노력이 필요하다. 이를 통해 112신고 데이터가 보다 적극적으로 범죄 예방과 도시 안전정책 수립에 기여할 수 있는 여건을 마련할 수 있을 것이다.

제3장

112신고 발생 기술통계 분석

1. 분석 개요
2. 분석 결과
3. 소결

1. 분석 개요

1) 분석 대상

분석 대상은 2023년 1년간 발생한 112신고 데이터이며 실제 신고 발생지역이 서울특별시 경계 내에 속하는 신고를 대상으로 심층 통계분석을 수행하였다. 분석에 사용한 데이터는 2023년 1월 1일부터 12월 31일까지 전국 18개 경찰청(서울·경기남부·경기북부·인천·대전·충남·충북·광주·강원·전북·전남·대구·경남·경북·부산·제주·울산·세종)에 접수된 총 13,071,309건의 112신고 데이터이며 신고자 개인정보에 해당할 수 있는 공간정보는 격자 단위 집계를 통하여 익명화하는 과정을 거쳐 신고자 정보가 보호될 수 있도록 분석을 진행하였다.

2) 분석 방법

■ 서울시 대상 데이터 추출 및 전처리

전국을 대상으로 하는 112신고 원데이터를 바탕으로 신고 발생 위치가 서울특별시 경계 내에 있는 기록을 공간 필터링하였다. 이 과정에서는 신고가 접수된 관할청이 서울시 외부에 있더라도 실제 발생지가 서울시로 확인되면 모두 포함하도록 하여 총 2,732,589건을 최종 데이터셋으로 확보했다. 격자 단위 집계를 위하여 서울특별시 전역을 가로·세로 500m × 500m 크기의 정사각형 격자망으로 분할하였다. 각 격자 셀은 GIS 도구를 이용해 격자 중심점이 행정구역 내부에 들어오는지 여부를 검사하여 최종 분석 대상으로 선별하였다.

■ 주요 변수 및 데이터 구조

본 연구에서 활용한 112 긴급신고 데이터는 2023년 한 해 동안 접수된 모든 신고에 대한 정보를 포함한다. 각 레코드는 신고가 접수된 시점, 신고 유형, 우선순위, 담당 관서, 현장 주소 등 시간적, 공간적, 운영적 맥락을 아우르는 주요 속성들로 구성되어 있으며 이를 통해 112신고 발생의 시공간적 분포를 다른 정보와 연계하여 다각도로 분석할 수 있다. 특히 긴급코드는 112 상황실이 신고 내용을 바탕으로 긴급성과 현장 출동의 필요성에 따라 총 다섯 단계의 코

드로 분류한 것으로 긴급코드별 112신고 분포 비교를 통하여 신고의 성격에 따른 신고 발생 경향의 차이를 파악할 수 있다.

[표 3-1] 112신고 데이터 구조

변수명	설명
접수일자	112신고가 접수된 날짜 (YYYY-MM-DD)
접수시간	신고 접수 시각 (HH:MM:SS)
종결일시	사건 처리 완료 시각 (YYYY-MM-DD HH:MM:SS)
긴급코드	112신고 대응 우선순위 코드 (0-4)
접수유형	112 직접 신고 또는 이관 접수 여부
접수관서·관할관서	신고 접수 및 현장 출동을 담당하는 경찰서·지구대·파출소명
발생주소(도로·지번·상세)	신고 발생지의 주소 정보
X좌표·Y좌표	WGS84 기준 위도·경도 좌표

출처: 연구진 작성

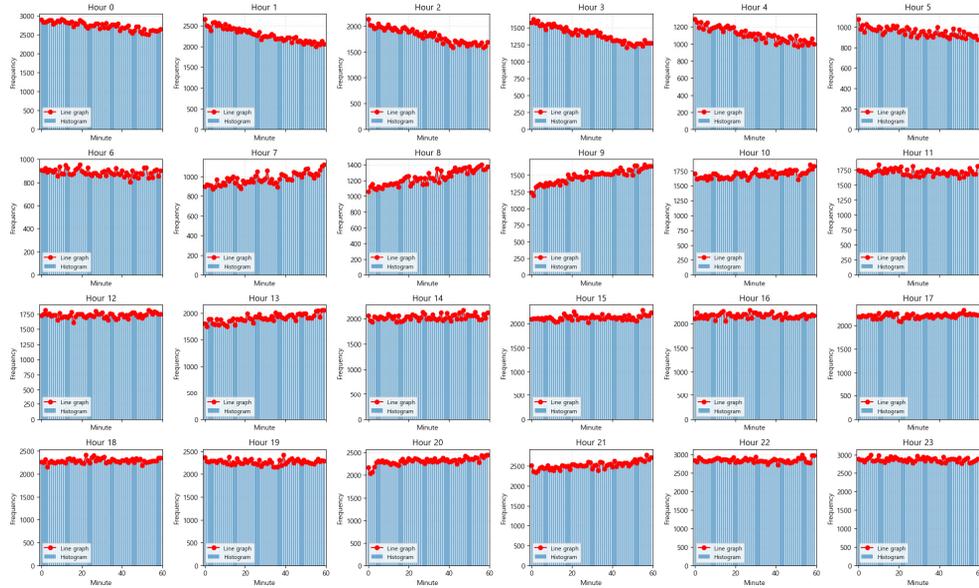
■ 시간대별 분포 이상치 확인 및 전처리

본 연구에서는 112신고 데이터의 시·공간 분석 및 예측 모델의 신뢰성을 확보하기 위해 먼저 데이터 자체의 유효성을 검증하고 시간 정보의 정밀도를 점검하였다. 첫째로 신고가 기록된 시각의 분·초 단위가 비정상적으로 특정 시점에 몰려 있는 현상이 있는지를 분석하였다. 전년도에 수행한 “범죄예방 환경설계(CPTED) 고도화 및 인증제도 개선 방향(조영진 외, 2024)” 연구에서는 범죄 발생 데이터가 매시 00분 00초에 집중 발생한 것으로 기록된 경향이 발견되었다. 특히 0시 00분 00초에 해당하는 자정에 범죄 발생이 가장 많은 것으로 나타났는데 이는 시스템 자동입력 오류나 단순 시간 미상에 따른 일괄 처리 가능성이 제기된 바 있다. 따라서 본 연구에서는 112신고 데이터 분석 전 동일한 이상 패턴이 112신고 데이터에서도 발생하는지를 점검하였다.

이를 위해 시간대별로 신고 발생 시각의 분 단위 분포를 히스토그램과 꺾은선 그래프로 시각화하여 특정 분·초에 집중되는 양상이 있는지를 검토하였다. 그 결과, 112신고 데이터에서는 전반적으로 시간대별 분포가 고르게 나타났으며 특정 시점에 비정상적으로 몰린 흔적은 발견되지 않았다. 이는 실제 시스템을 통한 신고 입력이 대부분 정규적으로 이루어졌음을 의미하며 시간 정보의 신뢰도가 높다는 것을 시사한다. 따라서 본 데이터셋은 시간대별 오류가 발견되지 않아 별도의 타임스탬프 수정 없이 후속 분석에 활용 가능한 것으로 판단하였다.

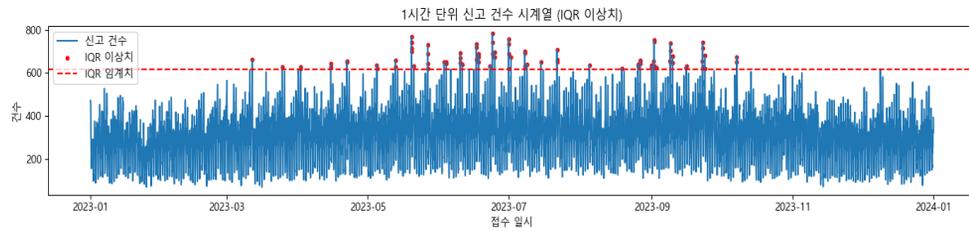
다음으로는 2023년 한 해 동안 집계된 112 긴급 신고 건수를 1시간 단위로 묶어 시계열 형태로 분석하였다. 먼저 각 시간대별 신고 건수 분포의 제1사분위수(Q1)와 제3사분위수(Q3)를 계산하고 사분위 범위(IQR=Q3-Q1)를 구하였다(Tukey, 1977). 이상치 탐지 기준인 상한 임계치($Q3+1.5 \times IQR$)를 적용하여 일반적인 “정상 분포 범위”를 벗어나 극단적으로 신고가 몰린 시간대를 식별하였다. 그 결과, 총 8,760개(24시간 × 365일)의 시간대 중 85개가 임계치

를 초과하는 극단값으로 나타났다. 이들 시간대는 차트상에서 빨간색 점으로 강조하여 표시하였다.



[그림 3-1] 112신고데이터 시간대별 분포 이상치 탐지 시각화

출처: 연구진 작성



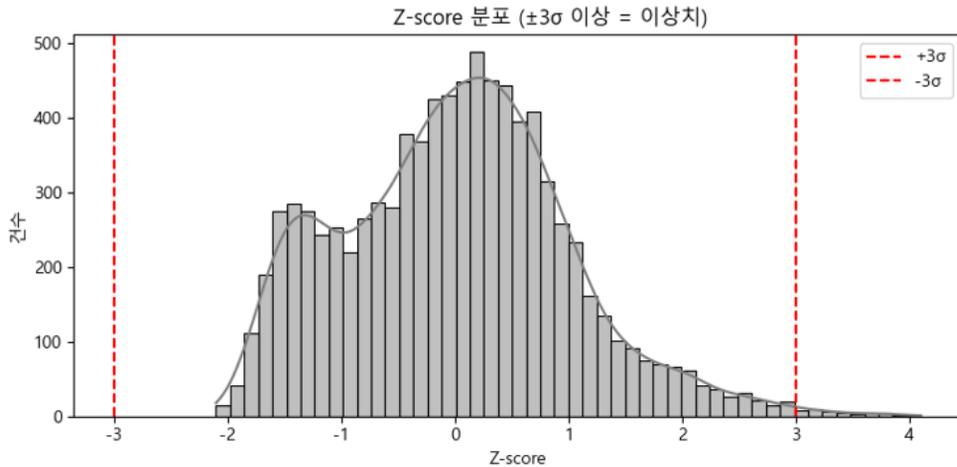
[그림 3-2] 112신고데이터 IQR 이상치 탐지

출처: 연구진 작성

IQR 분석 결과 극단값의 존재 가능성이 있다고 판단되어 시간대별 신고 건수의 이상 여부를 탐지하기 위해 표준화 점수(Z-score) 방법을 적용하였다. Z-score(표준화 점수)는 관측값이 전체 분포에서 평균으로부터 얼마나 떨어져 있는지를 표준편차 단위로 측정하는 통계적 지표이다. 우선, 전체 1시간 단위로 집계된 112신고 건수 데이터를 대상으로 평균(μ)과 표준편차(σ)를 산출하였다. 그런 다음 각 시간대의 신고 건수에서 전체 평균을 뺀 값을 표준편차로 나누어 Z-score를 계산하였다.

$$Z\text{-score} = (X - \mu) / \sigma$$

여기서 X 는 개별 시간대의 신고 건수, μ 는 전체 시간대 평균, σ 는 전체 시간대 표준편차를 의미한다. Z-스코어는 일반적으로 정규분포를 따른다고 가정할 때, ± 3 을 초과하거나 미만인 경우 극단치(outlier)로 간주할 수 있다(Iglewicz, B., & Hoaglin, D. C., 1993). 본 연구에서도 이러한 기준을 적용하여 Z-스코어가 +3보다 크거나 -3보다 작은 시간대를 모두 이상치로 분류하였으며 총 41개 시간대가 이에 해당하였다.



[그림 3-3] 112신고데이터 표준화 점수 이상치 탐지
출처: 연구진 작성

본 연구에서는 또한 주기적인 시간 패턴을 고려한 이상 탐지를 위해 요일-시간대 기준 Z-score(Conditional Z-score by weekday & hour group)를 적용하였다. 이는 전체 평균 기준의 Z-score와 달리 요일과 시간대가 동일한 관측값들만을 묶어 별도로 평균과 표준편차를 산출하고 그 내부에서 상대적으로 극단적인 값을 판별하는 방식이다. 즉, 전체 데이터를 하나의 기준으로 보는 것이 아니라 "같은 요일-같은 시간대"라는 세부적인 그룹 단위로 데이터를 분할하고 각 그룹별로 개별적인 평균과 표준편차를 산출한다. 예를 들어 모든 "화요일 오전 10시" 신고 건수를 하나의 집단으로 묶어 그 평균과 분산을 계산하고 이후 특정 일자의 해당 시간대 신고 건수가 이 범위를 벗어나는지를 판단하는 것이다.

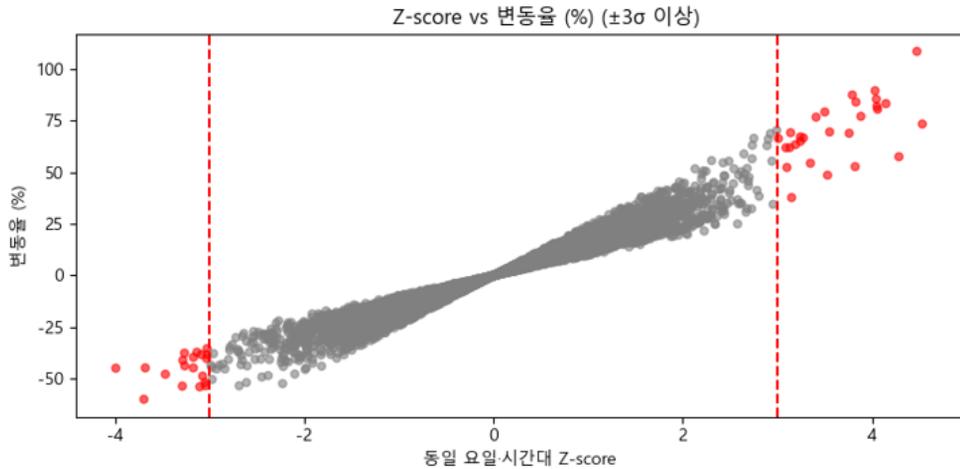
$$Z = (X - \mu_s) / \sigma$$

(X : 해당 조건의 관측값, μ_s : 동일 요일-시간대의 평균, σ_s : 동일 그룹의 표준편차)

해당 분석에서 또한 Z-score가 +3보다 크거나 -3보다 작은 경우를 이상치로 간주한 결과, 총 47개 시점이 조건부 그룹 기준 이상치로 탐지되었다. 이 방식은 단순히 전체 평균에서 벗어난

값이 아니라 시간대별 주기적 패턴 내에서도 유독 예외적인 상황을 찾아내는 데 적합하며 신고량 급증 등 특수 상황 탐지에 보다 정밀한 해석을 제공할 수 있다. 이러한 분석 결과는 산점도와 시계열 그래프를 통해 시각화되었다. 먼저 산점도에서는 같은 요일-시간대 그룹 내에서 계산된 Z-score를 기준으로 해당 시점의 신고 건수가 그 그룹 평균 대비 얼마나 비율적으로 벗어났는지(변동률)를 함께 나타냈다. 이때 변동률(%)은 “해당 시간대 신고 건수가 같은 요일-시간대 평균 대비 몇 % 더 많거나 적은가”를 정량화한 값이다.

이상치는 Z-score가 ± 3 을 초과하는 구간에 해당하며 이들 점은 붉은색으로 강조되어 시각적으로 쉽게 식별된다. 특히 이 시각화는 단순히 수치가 큰 값을 찾는 것이 아니라 정규성 기반 통계 기준을 벗어난 특이 사례를 명확히 보여준다. 한편, 시계열 그래프에서는 동일 요일-시간대 그룹 간의 전체 추이를 선으로 나타내고 그 중에서 조건부 Z-score(± 3 이상) 기준에 따라 탐지된 이상치를 붉은 점으로 표시하였다.



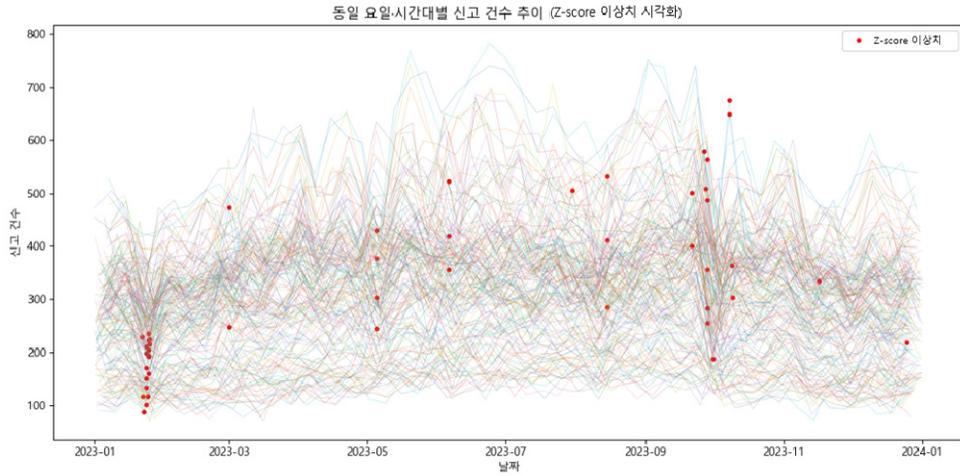
[그림 3-4] 112신고데이터 요일-시간대 그룹 Z-score와 변동률 간 관계

출처: 연구진 작성

112신고 데이터의 시계열 및 공간적 이상 여부도 연구자가 직접 종합적으로 점검하였다. 특히 공간적 측면에서는 특정 지역에 신고가 과도하게 밀집된 경우나 좌표가 부정확하게 기록된 경우를 파악하기 위해 시각화를 수행하여 연구자가 수작업으로 확인하였다. 확인 결과, 이러한 공간적 밀집 현상은 단순한 오류가 아닌 실제 사건에 따른 결과일 가능성이 큰 것으로 판단하였다.

112신고 데이터를 다양한 측면에서 점검하고 이상치를 탐색한 결과, 소수의 집중 사례가 존재하나 실제로 그러한 집중이 발생한 것으로 보이며 분석 결과를 왜곡할 것으로 판단되지는 않아 기계적으로 제거하는 것은 부적절한 것으로 판단하였다. 이에 따라 본 연구에서는 공간

적 밀집 또는 시간적 집중 현상을 제거 기준으로 사용하지 않고 해당 이상치 탐지 결과를 시각화하여 분석자가 수작업으로 확인하고 판단할 수 있는 보조 도구로만 활용하였다.



[그림 3-5] 112신고데이터 요일·시간대별 신고 추이 및 이상치 탐지

출처: 연구진 작성

2. 분석 결과

1) 월별, 요일별 112신고 발생

서울 지역 112신고 데이터는 하루 평균 약 7,400건 정도로 집계되며 시기에 따라 신고 건수가 다소 달라지는 양상을 보인다. 계절별로는 여름(6~8월, 743,776건)에 가장 많은 신고가 접수되었으며 봄(3~5월, 700,319건), 가을(9~11월, 691,848건), 겨울(12~2월, 589,646건) 순으로 나타났다. 특히 여름철은 전체 신고의 약 27%를 차지하여 가장 높았다. 월별 집계 결과, 9월(253,489건), 7월(251,978건), 8월(244,277건) 순으로 신고 건수가 많으며 1월(193,088건)과 2월(185,424건)은 비교적 신고 건수가 낮은 편이었다. 요일별로는 금요일(409,860건)과 토요일(439,890건)의 신고 빈도가 높았으며 월요일(360,703건)이 가장 적었다. 금요일부터 주말 동안 활동량 증가에 따른 치안 수요 증가가 동반되는 것으로 해석될 수 있다.

[표 3-2] 월별, 요일별 112신고건수 추이

월	신고 건수	요일	신고 건수
1	193,088	월요일	360,703
2	185,424	화요일	371,204
3	226,017	수요일	380,138
4	231,268	목요일	385,304
5	250,034	금요일	409,860
6	247,521	토요일	439,890
7	251,978	일요일	385,490
8	244,277		
9	253,849		
10	231,000		
11	206,999		
12	211,134		

출처: 연구진 작성

[표 3-3] 계절별 112신고건수 추이

계절	월 범위	신고 건수
봄	3-5월	707,319
여름	6-8월	743,776
가을	9-11월	691,848
겨울	12-2월	589,646

출처: 연구진 작성

2) 시간대별 112신고 발생

112신고는 하루 24시간 동안 시간대별로 뚜렷한 변화 양상을 보인다. 신고가 가장 적은 시간대는 새벽 6시(53,185건)였으며 이어서 5시(57,158건), 7시(58,690건) 등 전반적인 새벽 시간대에 신고가 가장 적게 발생하는 것으로 나타났다. 이후 오전 10시를 기점으로 신고 건수가 10만 건을 넘기기 시작하며 전반적으로 낮 시간과 저녁 시간에 걸쳐 점차 증가하는 경향을 보인다. 특히 밤 11시(23시)에는 171,773건으로 하루 중 가장 많은 신고가 접수되었고 이후 새벽 2시까지도 10만 건 이상의 신고가 지속되는 양상을 보였다. 이러한 경향은 심야 시간대의 사회적 활동, 음주, 갈등 상황의 증가와 관련된 것으로 추정된다.

[표 3-4] 시간대별 112신고건수 추이

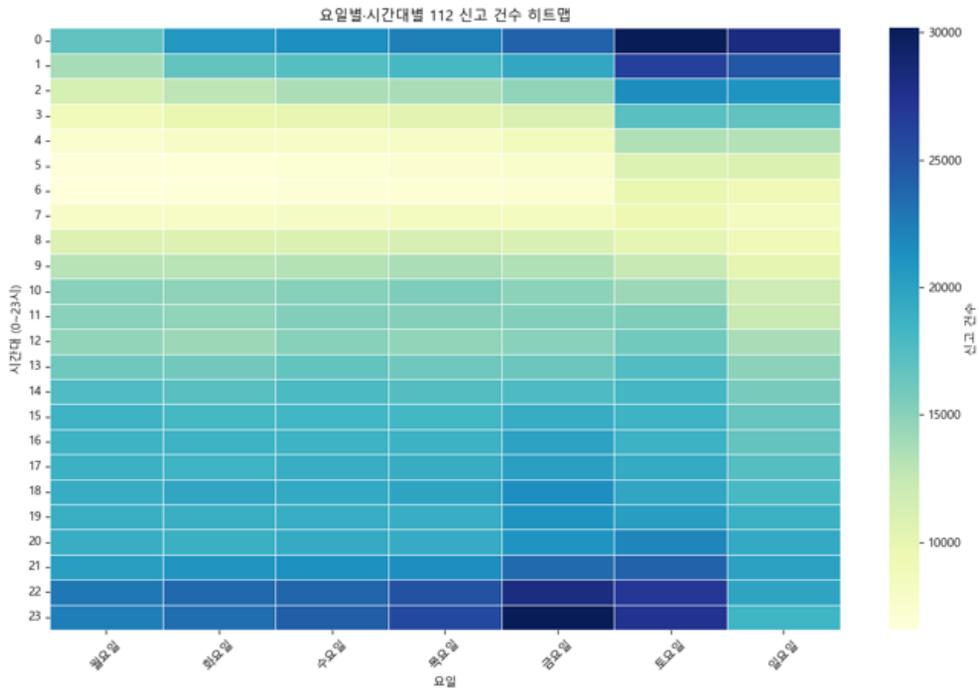
시간대	신고 건수	시간대	신고 건수
0	163,699	12	103,693
1	136,864	13	114,067
2	108,823	14	122,021
3	84,113	15	127,507
4	66,640	16	129,850
5	57,158	17	132,423
6	53,185	18	137,330
7	58,609	19	136,190
8	74,181	20	138,816
9	89,087	21	151,681
10	101,441	22	170,309
11	103,129	23	171,773

출처: 연구진 작성

요일과 시간대별 112신고 건수의 분포를 직관적으로 파악하기 위해, 전체 데이터를 기준으로 요일(월요일~일요일)과 시간대(0~23시) 단위로 집계하여 히트맵으로 시각화하였다. 분

석 결과, 전반적으로 신고는 새벽 시간대에 가장 적게 발생하며 오후부터 밤 시간대에 집중되는 경향을 보였다.

특히 금요일과 토요일 심야 시간대(21시~2시)는 다른 요일 대비 신고 밀도가 뚜렷하게 높게 나타났으며 이는 주말 전후의 야외 활동 증가, 유흥 활동, 음주 등과 관련된 사건·사고의 빈도 상승과 관련이 있는 것으로 해석된다. 반면, 평일의 새벽 시간대는 모든 요일에서 신고량이 상대적으로 낮아 일관된 패턴을 보였다.



[그림 3-6] 요일·시간대별 112신고패턴 히트맵

출처: 연구진 작성

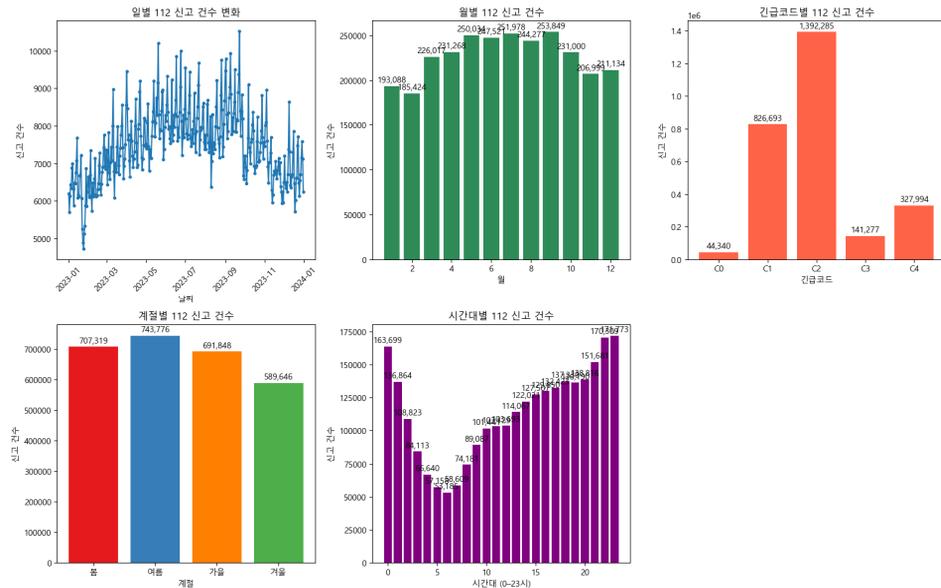
3) 긴급코드별 112신고 발생

긴급코드 분류별 112신고 중 가장 많은 비중을 차지한 유형은 코드 2로 총 1,392,285건(전체의 약 50%)에 달했다. 이는 생명·신체에 대한 잠재적 위험이나 범죄 예방을 위한 출동 요청에 해당하는 유형으로 시민의 치안 민감도를 보여주는 주요 지표 중 하나로 해석될 수 있다. 다음으로 코드 1(826,693건), 코드 4(327,994건), 코드 3(141,277건), 마지막으로 가장 긴급한 신고를 나타내는 코드 0(44,340건) 순으로 나타났다.

[표 3-5] 코드별 112신고건수 추이

긴급코드	신고 건수
코드 0	44,340
코드 1	826,693
코드 2	1,392,285
코드 3	141,277
코드 4	327,994

출처: 연구진 작성

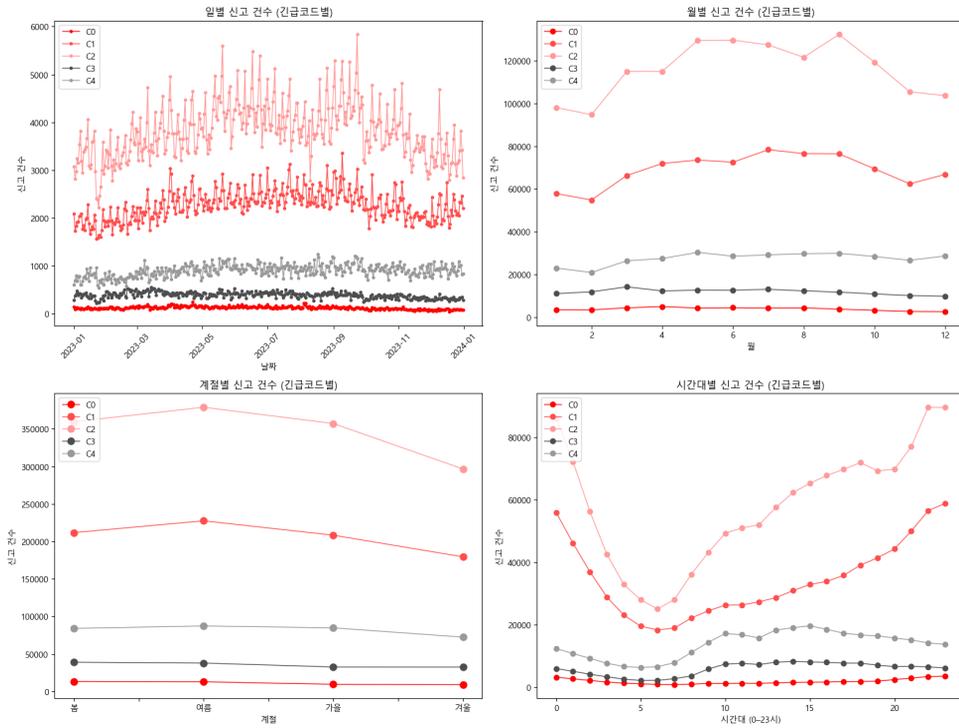


[그림 3-7] 112신고건수 전방적 추세 시각화

출처: 연구진 작성

이러한 긴급코드별 신고 건수의 일별 시계열, 월별·계절별 집계, 그리고 시간대별 분포를 다각적으로 분석하였다. 연간 일별 시계열 분석에서는 신고 발생이 많은 코드 1, 2에서는 연중 뚜렷한 추세가 나타나는 반면, 상대적으로 신고 건수가 적은 코드의 경우 장기적 추세보다는 일별 변동의 비중이 높아 불규칙한 흐름이 나타났다. 월별 및 계절별 분석에서도 이러한 추세가 확인되었는데 전반적으로 봄과 여름에 신고가 가장 활발하게 이루어졌으며 특히 가장 신고 건수가 많은 코드 1, 2는 여름철에 특히 집중되었다. 이는 사회적 활동량 증가나 기온 상승에 따른 사건 빈도 상승 등 계절성과 관련된 외생 요인의 영향을 추정해볼 수 있다.

시간대별로는 긴급코드에 따라 신고의 집중 시점이 상이하게 나타났다. 코드 0, 1, 2는 야간 시간대에 신고가 집중되는 양상을 보였으며 낮에는 저녁 시간대로 갈수록 발생량이 완만하게 상승하는 모습을 보였다. 반면, 코드 3, 4의 경우 업무 시간대 집중되는 패턴을 드러냈다.

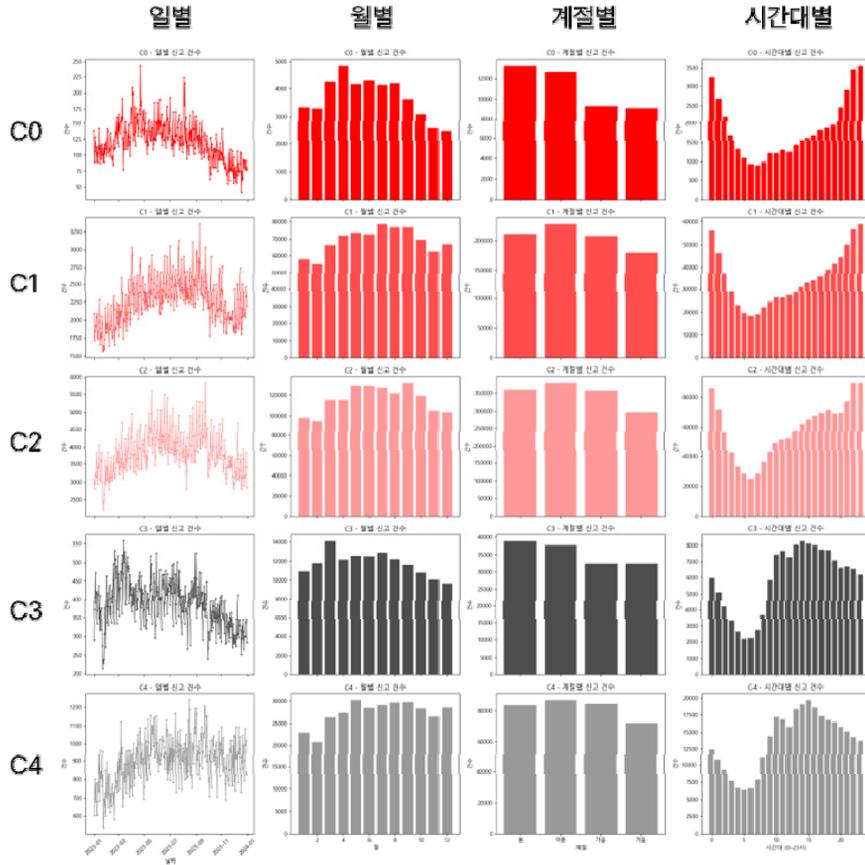


[그림 3-8] 긴급코드별 112신고건수 전반적 추세 시각화

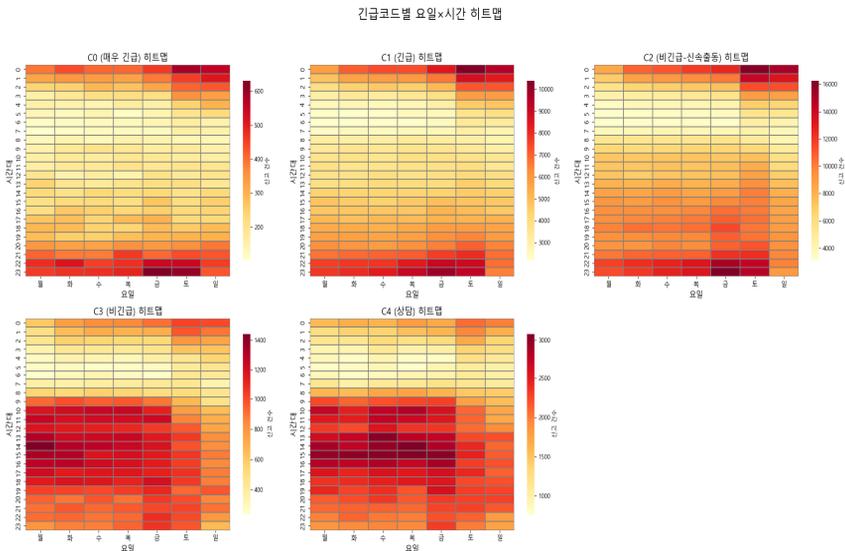
출처: 연구진 작성

마지막으로 긴급코드별, 요일별, 시간대별 히트맵 분석을 통해 긴급코드별 신고 패턴의 특성의 차이를 종합적으로 확인하였다. 히트맵은 시간대와 요일 간의 교차 분포를 직관적으로 보여주며 특정 시간대와 요일에 집중되는 신고 밀도를 색의 진하기로 구분할 수 있다는 점에서 유용하다. 분석 결과, 코드 0(매우 긴급), 코드 1(긴급), 코드 2(비긴급-신속출동) 모두 야간 시간대, 특히 21시~23시 사이에 신고 발생이 집중되며 금요일과 토요일에 상대적으로 높은 밀도를 보이는 등 유사한 분포 패턴을 공유하였다. 반면, 코드 3(비긴급)와 코드 4(상담)는 뚜렷하게 다른 양상을 보였다. 두 코드 모두 히트맵 상 가장 집중된 시간대는 평일 오전 9시부터 오후 6시 사이로 나타났다.

이러한 분포는 긴급한 범죄 신고와 일상적 민원, 생활 불편, 행정 상담과 같은 비긴급 신고가 서로 다른 상황에서 발생하는 것으로 해석할 수 있다. 긴급 신고의 경우 사회적 활동량이 증가하고 한편으로는 비일상적인 사건이 일어날 수 있는 야간과 주말에 집중되는 반면, 비긴급 신고는 주로 평일 업무 시간 중에 접수된다는 것을 예상해 볼 수 있다. 특히 코드 4는 상담 목적의 신고이며 코드 3 역시 비교적 비긴급한 민원 신고가 포함되어 있어 두 코드 모두 범죄 발생 가능성이나 물리적 환경요소와의 직접적인 연관성이 낮을 가능성이 있다.



[그림 3-9] 긴급코드별 112신고건수 상세 추이 시각화
출처: 연구진 작성



[그림 3-10] 긴급코드별 요일·시간대별 112신고패턴 히트맵
출처: 연구진 작성

3. 소결

이 장에서는 112신고의 성격을 실제 데이터 기반으로 파악하고 이에 기반한 물리환경 영향 실증 분석의 조건을 설정하기 위하여 112신고 데이터에 대한 기술통계 분석을 수행하였다. 112신고는 실제로 발생한 범죄에 대한 신고, 범죄 발생에 대한 긴급한 우려, 단순 민원 등이 복합적으로 나타나기 때문에 112신고의 성격을 규명하고 실제 범죄 발생 및 범죄두려움을 대표할 수 있는 지표로 정제할 필요가 있었다.

기술통계 분석을 통하여 112신고 데이터에 대한 긴급코드 계절, 요일, 시간대별 발생 패턴을 분석한 결과, 긴급코드별로 시계열적인 패턴이 구분된다는 점을 확인할 수 있었다. 긴급코드는 신고를 접수한 112 상황실에서 신고 내용을 바탕으로 분류한 것으로 즉시 출동 필요성 등 긴급도를 평가한 것이다. 긴급 신고에 해당하는 코드 0, 코드 1은 주말과 야간에 신고 발생이 많은 것으로 나타났으며 비긴급 신고인 코드 3, 민원 상담에 해당하는 코드 4는 평일 업무 시간대에 집중되어 긴급 신고와는 다른 패턴을 보였다. 이는 112신고를 크게 긴급 신고와 비긴급 민원 두 가지 군집으로 나누어 해석할 필요성을 제기하고 있다. 다만, 정의상 비긴급 신고 이기는 하나 즉시 출동이 요구되는 코드 2의 경우, 발생 패턴이 긴급 신고인 코드 0, 코드 1과 유사하게 주말과 야간에 집중된 것으로 나타났다.

이처럼 112신고에 대한 기술통계 분석 결과는 각 긴급코드별 신고 패턴이 시간대, 계절, 요일 등 다양한 요인과 상호작용하며 구조적으로 구분된다는 점을 명확히 보여주었다. 따라서 긴급 신고라는 행태적 지표를 통해 파악할 수 있는 범죄 발생 및 범죄두려움에 초점을 맞추기 위해 비긴급 신고에 해당하는 코드 3과 코드 4를 분석 대상에서 제외하고 긴급 대응 및 물리적 위험성과의 관련성이 높을 것으로 사료되는 코드 0, 코드 1과 출동이 요구되고 긴급 신고와 발생 분포가 유사하게 나타나는 코드 2를 실증 분석의 대상으로 설정하였다.

제4장

물리환경이 112신고 발생에 미치는 영향 실증 분석

1. 분석 개요
2. 분석 결과

1. 분석 개요

1) 분석 범위 및 방법

본 장은 실제 건축 및 도시 환경에서 물리환경의 특성이 112신고 발생에 미치는 영향을 분석하는 것을 목표로 한다. 건축·도시 물리환경 대상 범죄예방 환경설계를 통한 범죄 발생과 범죄두려움 저감 달성 전략을 제시하는 데 기여하고자 하였다. 기계학습 모델을 활용하여 물리환경 특성과 112신고 발생 간의 관계를 분석하고 긴급 신고 발생에 영향을 미치는 주요 물리환경 요인을 식별하였다. 이러한 분석 결과를 활용하여 범죄 발생뿐만 아니라 범죄두려움 저감 측면에서도 실효성이 있는 CPTED 접근을 가능하게 하는 것을 목표로 하였다.

■ 분석 범위

분석 대상의 범위는 2023년 서울특별시에서 발생한 112신고 중 코드 0, 1, 2로 분류된 긴급 신고 및 즉시 출동 신고이다. 코드 3, 4로 분류된 비긴급 신고의 경우 신고 분류의 의미와 신고 발생 패턴 모두 긴급 신고와 크게 달라 제외하였다.

• 내용적 범위

본 연구는 “물리환경이 112신고 발생에 미치는 영향”을 분석하는 데 중점을 둔다. 이를 위해 2023년 한 해 동안 서울시 전역에서 접수된 112신고 데이터를 기반으로 신고 건수와 물리적 환경 변수 간의 관계를 정량적으로 규명하고자 하였다. 먼저, 112신고 데이터에 포함된 접수 일자, 접수시간, 긴급코드, 발생주소(도로명 및 지번) 등의 주요 변수에 대해 전처리를 수행하고 이를 바탕으로 통계적 집계를 진행하였다. 이어서 CCTV 수, 도로 폭, 건물의 용도 및 층수, 노후 건축물의 분포 등 다양한 물리적 환경요소를 GIS 데이터와 공공데이터를 통해 수집하고 이를 500m × 500m 격자 단위로 공간 정보화하였다. 이후 일별, 시간대별, 요일별, 계절별로 나타나는 112신고 건수의 시공간적 패턴을 시계열 분석 및 히트맵 등을 활용해 시각화하며 물리환경 변수와의 관계를 탐색하였다. 마지막으로 랜덤포레스트와 XGBoost 등의 머신러

닝 기반 예측 모형, 그리고 MGWR 등 공간회귀 분석 기법을 적용하여 물리적 환경 요인이 신고 건수에 미치는 영향을 추정하고 그중에서도 주요 설명변수를 도출하였다.

- 공간적 범위

본 연구는 서울특별시 전역을 공간적 분석 범위로 설정하였다. 기본 분석 단위는 500m × 500m 크기의 정사각형 격자(grid cell)이며 각 격자별로 112신고 건수 및 주요 물리환경 변수(예: 토지이용, 인프라 밀도 등)를 집계하여 공간 분석을 수행하였다. 이러한 고해상도 격자 기반 접근은 도시 전역에서 발생하는 미시적 공간 패턴과 이상 분포를 보다 정밀하게 파악하는 데 효과적이다. 아울러 분석 결과는 필요에 따라 자치구·법정동·행정동 등 행정구역 단위로 재집계하여 지역 간 비교 및 정책적 해석에 활용 가능하다. 최종적으로는 격자 기반 분석을 통해 서울시 내 지역별 112신고 패턴의 이질성을 반영하여 범죄예방 및 대응 전략 수립에 기여하고자 하였다.

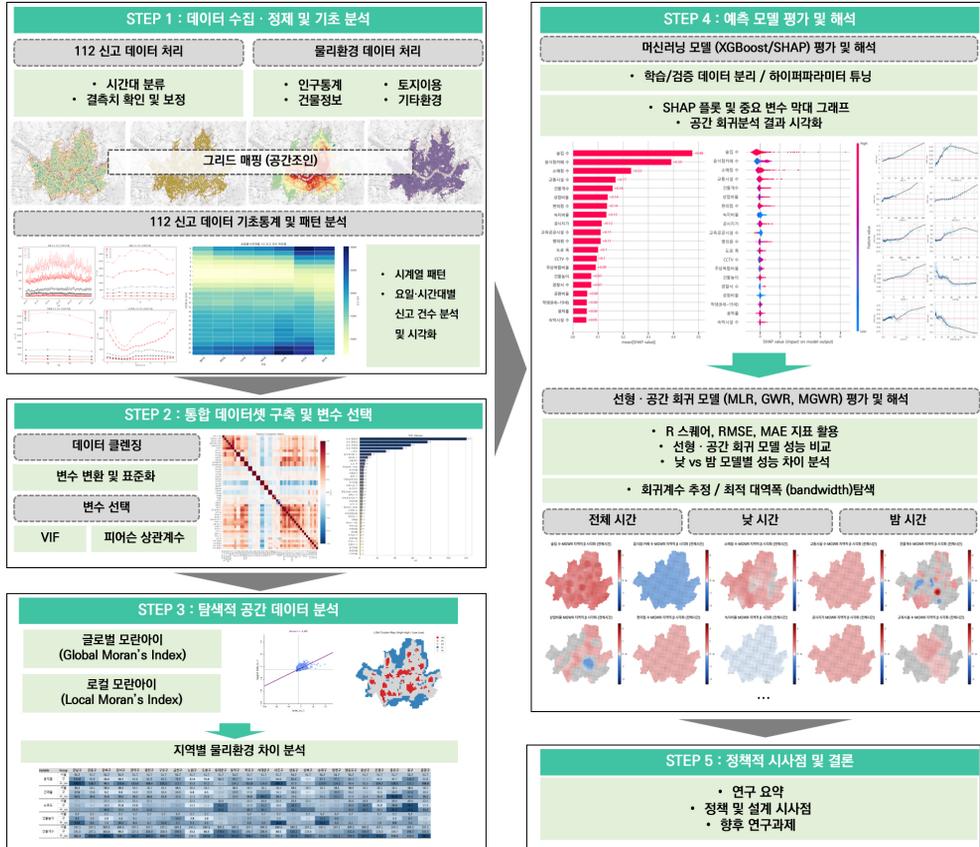
- 시간적 범위

본 연구의 시간적 범위는 2023년 1월 1일 00시부터 2023년 12월 31일 23시 59분까지이다. 본 연구에서는 계절별(봄·여름·가을·겨울), 월별, 요일별, 시간대별 신고 패턴을 종합적으로 분석하였다. 분석 기간 동안 발생한 112신고 데이터 중 결측치 및 이상치를 사전에 확인 및 처리하고 동일 기간의 물리환경 데이터(CCTV 설치 현황, 건물 용도 및 특성 등)를 수집 및 정제하여 통합 분석에 활용하였다. 본 연구는 내용적·공간적·시간적 범위를 명확히 설정하여 서울시 전역의 112신고와 물리환경 요인 간의 시·공간적 상호작용을 다각도로 규명함으로써 객관적이고 실증적인 범죄예방 정책 수립을 지원하고자 하였다.

■ 분석 방법

기계학습을 활용하여 건축·도시의 물리환경 특성이 112신고 발생에 미치는 영향을 실증 분석하는 과정은 물리환경이 범죄 발생에 미치는 영향을 분석하였던 직전년도 연구와의 연속성을 유지하면서 분석 대상 범위를 확대하여 다양한 환경에서 물리환경의 영향을 실증하고자 하였다. 서울특별시 전역을 대상으로 물리환경 공간정보를 구축하고 112신고 데이터와 연계하여 공간회귀 및 기계학습 분석을 적용하여 전년도 연구에서 강남 4개 구 분석으로 지리적 다양성이 충분하지 않았던 점을 보완하였다.

본 연구에서는 112신고 데이터를 기반으로 도시 물리환경과의 관계를 체계적으로 규명하고자 총 6단계의 분석 절차를 수행하였다. ① 먼저, 국내·외 112 및 911 신고 데이터를 활용한 선행연구를 검토하고 범죄예방이론(CPTED) 및 물리환경 요인의 역할에 대한 이론적 배경을 정립하였다. ② 다음으로 112신고 데이터를 수집한 후 이상값 및 결측치를 정제하고 1시간 단위의 시계열 및 격자 기반의 공간 단위로 집계·분석하여 데이터의 기초 구조를 점검하였다.



[그림 4-1] 연구 프레임워크

출처: 연구진 작성

③ 이를 통해 신고 건수의 분포, 코드별 특성, 시간대·계절별 추이를 시각화하고 물리환경 변수 또한 토지이용, 건물, 인구, 도로, CCTV 등 다양한 도시공간 요소를 기반으로 정제하여 격자 단위 변수로 구성하고자 하였다. ④ 이후, 112신고 데이터와 물리환경 데이터를 동일한 시공간 기준으로 통합하고 공간 자기상관 분석(Moran's I, LISA)을 통해 신고 밀도 및 공간적 집중성을 탐색적으로 분석하였다. ⑤ 이러한 공간분석 기반을 바탕으로 도시 내 특정 시점과 장소에서 물리환경이 신고 발생에 미치는 영향을 평가하고자 다중 스케일 지리 가중 회귀(MGWR) 모델과 머신러닝 기반 예측모형(XGBoost)등을 활용하여 시간대·계절별 영향 요인을 비교 분석하였다. 더불어 SHAP 기반 변수 중요도 해석을 통해 각 변수의 기여도와 영향 방향성을 시각적으로 제시하였다. ⑥ 마지막으로 분석 결과에 대한 모델 성능을 교차 검증하고 해석 결과를 바탕으로 시공간 맞춤형 도시 안전 환경 개선을 위한 정책적 시사점을 도출하고자 하였다.

본 연구에서는 조영진 외(2024)에서 제시한 범죄예방 건축·도시 물리환경 특성 체계를 기반으로 일부 보완·확장하여 적용하였다. 조영진 외(2024)는 범죄예방 환경설계 관련 주요 선행

연구를 분석하여 범죄예방과 관련된 건축·도시 물리환경 요소를 접근통제, 감시, 영역성 강화의 원칙 등과 대응시켜 정리하였다. 범죄예방을 위한 환경설계는 크게 건축적 요소와 공간 환경 요소로 구분되며 건축적 요소로는 명확한 출입구, 잠금장치, 인터폰, CCTV, 센서, 조명 등의 접근통제 장치와 투명한 재료를 활용한 개방적 건물 배치를 통한 감시 강화, 그리고 개인 공간과 공용 공간의 명확한 구분 및 주민 참여 디자인을 통한 영역성 강화 등이 제시되었다. 공간환경 요소로는 좁고 긴 골목길을 지양하고 넓고 개방적인 공간을 조성하여 자연 감시를 강화하며 투시형 담장이거나 낮은 울타리 설치를 통한 시야 확보, 충분한 조도를 위한 가로등과 보안등 설치, 그리고 쉼터, 놀이터, 커뮤니티 시설 등의 공동체 공간 조성을 통해 영역성을 강화하는 등이 있었다.

[표 4-1] 물리환경 요소 분류

구분	건축적 요소	공간환경 요소
접근통제	명확한 출입구, 보안 시스템(출입 통제 시스템, 인터폰, CCTV), 안전한 조명 설계	좁고 긴 골목길 지양, 안전 지대 조성
감시	투명한 재료 사용, 개방적인 건물 배치, CCTV 설치	가로등, 보안등 설치, 거리 활성화, 주민 참여형 감시 활동
영역성 강화	개인 공간과 공용 공간 구분, 주민 참여형 디자인, 지역 특성 반영 디자인	공동체 공간 조성, 지역 주민 활동 지원, 범죄 취약 지역 환경 개선

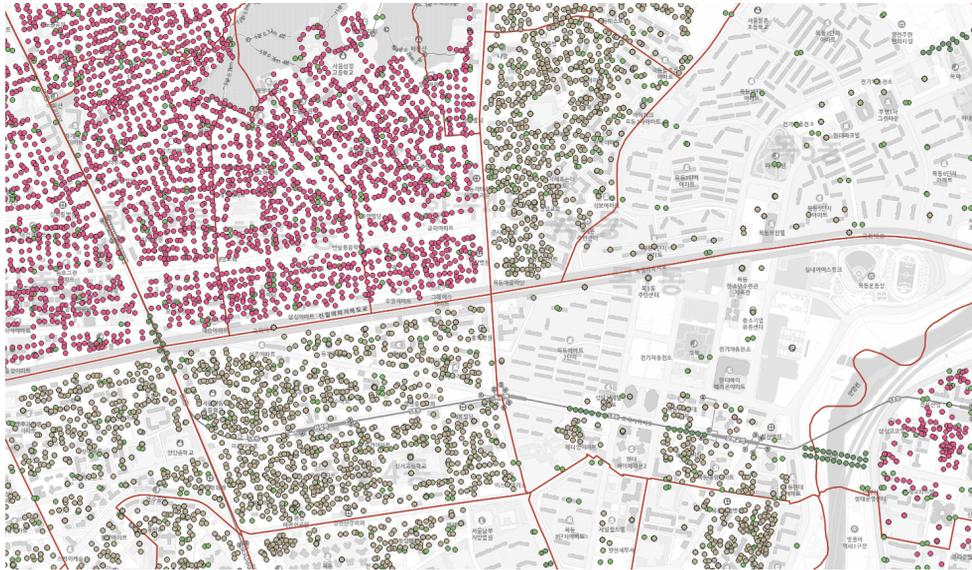
출처: 조영진, 안의순, 박성남, 고영호, 권오규, 임보영, 임리사, 김유진, 이정현. (2024). 범죄예방 환경설계(CPTED) 고도화 및 인증제도 개선 방향. 건축공간연구원.

2) 건축도시환경 변수 선정 및 전처리

■ 건축도시환경 변수 선정

본 연구는 전년도 연구의 연속으로 기존에 활용된 물리환경 요소들을 본 연구에도 동일하게 적용하고자 하였다. 전년도 연구의 건축도시 물리환경 변수는 국내의 다양한 연구와 CPTED(범죄예방환경설계) 체크리스트를 참조하여 도출되었으며 공공데이터 구득이 가능한 변수를 분석에 사용하였다. 이러한 물리환경 변수 선정 기준은 본 연구에서도 동일하게 적용될 수 있으므로 전년도 물리환경 변수 선정 결과는 본 연구의 예측 모델 개발과 실증 분석에도 기초로 활용되었다.

우선, 선행연구와 국내 CPTED 체크리스트 모두에서 공통적으로 제시된 변수로는 물리적 시설 요소인 건물 창문 수, CCTV 수, 가로등 수, 벽 또는 울타리 길이, 진출입구 수 등이 있다. 도로 및 보행 환경과 관련해서는 도로 수, 도로 폭, 도로 연결성, 보행로 면적, 도로 포장 상태, 표지판 등이 포함되며 환경 정비 상태를 나타내는 쓰레기, 불법 주정차, 깨진 창문, 빈집 등의 변수도 함께 제시되었다. 이 외에도 가판대 수, 나지 면적, 녹지 면적 등이 공통 변수로 도출되었다.



[그림 4-2] 지역 간 가로등 위치 데이터의 격차가 확인됨

출처: 연구진 작성

반면, 선행연구에서는 활용되었지만 CPTED 체크리스트에는 포함되지 않았던 변수들도 있다. 이들은 건축 및 용도 관련 변수로서 건물 노후도, 건물 수, 건물 높이, 건폐율, 용적률, 혼합용도 면적, 공업·농업·상업·주거 면적, 건축 후퇴선 등이 있으며 시설 분포 측면에서는 업무 시설, 병원, 소매시설, 교육시설, 여가시설, 숙박시설, 세탁소, 편의점, 술집, 주유소, 음식점,

약국, 경찰서, 주차장, 지하철역, 기차역 등의 변수가 포함된다. 또한 인구 및 상업 정보에 해당하는 인구 수, 은행 수, 카페 수와 도로 및 보행 네트워크와 관련된 도로 통합도, 도로 ANG 연결성, 도로 레인 수, 교차로 수, 보도 연결성, 보도 통합도, 보도 ANG 연결성 등의 변수도 선행연구를 통해 확인되었다. 환경 및 녹지 공간에서는 공원 면적과 수변 면적이 도출되었다. 한편, 국내 CPTED 체크리스트에만 등장하고 선행연구에서는 다루어지지 않았던 변수들도 있다. 대표적으로는 벤치 수, 자전거 보관소 수, 옥외 운동시설 수가 있으며 이는 지역 커뮤니티 공간의 안전성과 관련된 요소로 간주된다. 그러나 공공데이터의 부재 혹은 활용 한계로 인해 본 연구에서 제외된 변수들도 존재한다. 창문 수, 벤치 수, 벽 또는 울타리, 진출입구 수, 도로 포장 상태, 표지판 수, 쓰레기, 불법 주정차, 깨진 창문 수, 빈집 수, 은행 수, 가판대 수 등은 현재 공공데이터로 접근이 어려워 활용이 불가능하였다. 기차역 수는 데이터는 존재하지만 실제 수가 극히 적어 통계 분석에는 적합하지 않다고 판단되어 제외하였다. 또한 자전거 보관소 수, 실외 운동기구 수, 가로등 수는 공공데이터가 존재하더라도 지역 간 편차가 커 해석의 일반화에 어려움이 있어 분석 변수에서 제외하였다.

[표 4-2] 기계학습 기반 112신고데이터와 건축환경 분석을 위한 종속변수와 독립변수

변수 구분	데이터 명	설명	CPTED 체크 리스트	선행연구 변수	사용계획	
종속 변수	112신고	112신고 데이터	2023년에 수집된 112신고 데이터	-	-	0
독립 변수	건물 및 토지 특성	건물 노후도	면적 내 건물 사용승인일로부터 경과한 평균 연수		0	0
		건물 개수	면적 내 건물 개수		0	0
		건물 높이	면적 내 평균 건물 높이		0	0
		창문 수	도로에 면하는 창문 수	0	0	
		건폐율	면적 내 평균 건폐율		0	0
		용적률	면적 내 평균 용적률		0	0
		업무시설 수	면적 내 업무시설 수		0	0
		혼합용도면적	면적 내 혼합용도 토지 비율		0	0
		공업면적	면적 내 공업용도 토지 비율		0	0
		농업면적	면적 내 농업용도 토지 비율		0	0
		상업면적	면적 내 상업용도 토지 비율		0	0
		주거면적	면적 내 주거용도 토지 비율		0	0
		녹지면적	면적 내 나무, 풀, 숲 등의 녹지 비율	0	0	0
		수변면적	면적 내 습지, 강 등의 수변 비율		0	0
		나지면적	면적 내 운동장, 공터 등 개방형 공간 비율	0	0	0
공원면적	공원면적			0	0	
인구 밀도	인구 수	면적 내 인구 수		0	0	
생활 편의	병·의원시설수	면적 내 병·의원시설 수		0	0	

및 상권 시설 변수	소매시설 수	면적 내 소매시설 수		0	0
	교육시설 수	면적 내 교육시설 수		0	0
	여가시설 수	면적 내 여가시설 수		0	0
	숙박시설 수	면적 내 숙박시설		0	0
	세탁소 수	면적 내 세탁소 수		0	0
	편의점 수	면적 내 편의점 수		0	0
	술집 수	면적 내 술집 수		0	0
	주유소 수	면적 내 주유소 수		0	0
	은행 수	면적 내 은행 및 ATM기 수		0	
	경찰서 수	면적 내 경찰서 수		0	0
	카페 수	면적 내 카페 수		0	0
	음식점 수	면적 내 음식점 수		0	0
	약국 수	면적 내 약국 수		0	0
교통 및 기반 시설	경찰서 수	면적 내 경찰서 수		0	0
	CCTV 수	면적 내 CCTV 수	0	0	0
	가로등 수	면적 내 가로등 수 혹은 밝기	0	0	
	벤치 수	면적 내 벤치 수	0		
	자전거 보관소 수	면적 내 자전거 보관소 수	0		
	주차장 수	면적 내 주차장 수		0	0
	버스정류장 수	면적 내 버스정류장 수	0	0	0
	지하철역 수	면적 내 지하철역 수		0	0
기차역 수	면적 내 기차역 수		0		
거리 경계	벽 혹은 울타리	면적 내 벽 혹은 울타리 길이	0	0	
	진출입구 수	진출입구 수	0	0	
도로 공간 구조	도로연결성	Space syntax로 계산한 도로 연결성	0	0	0
	도로 통합도	Space syntax로 계산한 도로 통합도		0	0
	도로 ANG 연결성	Space syntax로 계산한 도로 ANG 연결성		0	0
	보도 연결성	Space syntax로 계산한 보도 연결성		0	0
	보도 통합도	Space syntax로 계산한 보도 통합도		0	0
	보도 ANG 연결성	Space syntax로 계산한 보도 ANG 연결성		0	0
	도로 레인수	면적 내 도로 레인 수		0	0
	도로 수	면적 내 도로 수	0	0	
	도로 폭	면적 내 평균 도로 폭	0	0	0
	교차로	면적 내 교차로 밀도		0	
	보행로 면적	면적 내 보행로 밀도, 면적, 혹은 길이	0	0	0
도로포장	면적 내 도로 포장 상태	0	0		
거리 개체	표지판 수	면적 내 표지판 수	0	0	
건물 위치	건축 후퇴선	면적 내 건축 후퇴선 길이		0	
거리 유지 관리	쓰레기	면적 내 거리 쓰레기 비율 혹은 불만 접수 건수	0	0	
	불법주정차	면적 내 거리 내 불법 주정차 비율	0	0	
건물 유지 관리	빈집 수	면적 내 빈집 수	0	0	
	깨진 창문	면적 내 도로를 면하는 깨진 창문 수	0	0	
거리활동	가판대 수	면적 내 가판대 수	0	0	
	옥외 운동시설 수	면적 내 옥외 휴게공간 혹은 운동 시설 수	0		

출처: 연구진 작성

■ 사용 변수 출처

선행 연구와 국내 CPTED 체크리스트를 바탕으로 선정된 독립변수들은 V-WORLD, 지방행정 인허가 데이터, 공공데이터포털, 국토정보맵, 서울 열린 데이터 광장, 환경공간정보서비스 등에서 제공하는 공공데이터를 통해 확보되었으며 세부 항목별로는 다음과 같다(표 4-3 참조).

V-WORLD 제공 공공데이터에서는 GIS 건물통합정보를 활용하여 건물 노후도, 건물 개수, 건물 높이, 건폐율, 용적률, 업무시설 수 등의 변수를 추출하였다. 또한 토지특성도를 이용해 혼합용도면적, 공업면적, 농업면적, 상업면적, 주거면적을 도출하였다. 교통 관련 변수로는 도로중심선 데이터를 통해 도로 라인 수, 도로 폭, 도로 연결성, 도로 통합도, 도로 ANG 연결성 등을 확보하였으며 보행로 면적은 인도(보도) 데이터를 기반으로 추출하였다.

지방행정 인허가 데이터에서는 CCTV 정보 데이터를 이용해 각 격자 내 CCTV 수를 산정하였다. 공공데이터포털에서는 다양한 생활편의 및 상권 관련 변수를 추출하였다. 소상공인시장진흥공단의 상가(상권)정보 데이터를 통해 술집 수, 교육시설 수, 음식점 수, 숙박시설 수, 소매시설 수, 여가시설 수, 카페 수, 편의점 수, 병·의원시설 수, 약국 수, 주유소 수 등을 파악하였으며, 주차장 수는 전국 주차장 정보 표준 데이터를, 버스정류장 수는 전국 버스정류장 위치 정보 데이터를 경찰서 수는 경찰관서 위치 주소 현황 데이터를 기반으로 추출하였다.

국토정보맵에서는 국토통계지도를 통해 격자 단위 인구 수 데이터를 확보하여 인구 밀도 분석에 활용하였다.

서울 열린 데이터 광장은 보행환경과 공공시설 관련 변수를 제공하였다. 보도 연결성, 보도 통합도, 보도 ANG 연결성을 계산하기 위해 서울시 도보 네트워크 데이터를 활용하였고 공원 면적은 서울시 생활권계획 시설(공원) 공간정보, 용도구역(도시자연공원구역) 공간정보, 도시계획시설(공간시설) 데이터 등을 통합하여 산정하였다. 또한, 서울시 자치회관 정보 및 사회복지시설 목록 데이터를 활용해 공공시설 수를 추출하였다.

마지막으로, 환경공간정보서비스에서는 토지피복지도(세분류) 데이터를 통해 나지 면적, 녹지 면적, 수변 면적을 추출하였다. 이러한 변수들은 연구의 예측 모델링과 공간적 해석에 핵심적으로 활용될 계획이다.

[표 4-3] 활용가능한 독립변수 공공데이터 출처

변수 구분		데이터 명	설명	공공데이터 출처	데이터 명
독립 변수	건물 및 토지 특성	건물 노후도	그리드별 사용승인일로부터 경과한 평균 년수	V-WORLD	GIS건물 통합정보
		건물 개수	그리드별 건물 개수		
		건물 높이	그리드별 평균 건물 높이		

변수 구분	데이터 명	설명	공공데이터 출처	데이터 명
	건폐율	그리드별 평균 건폐율		토지 특성도
	용적률	그리드별 평균 용적률		
	업무시설 수	그리드별 업무시설 수		
	혼합용도면적	그리드별 혼합용 토지 면적		
	공업면적	그리드별 공업용 토지 면적		
	농업면적	그리드별 농업용 토지 면적		
	상업면적	그리드별 상업용 토지 면적		
	주거면적	그리드별 주거용 토지 면적	환경공간 정보서비스	토지피복지도 세분류
	녹지면적	그리드별 나무, 풀 등의 녹지 비율		
	수변면적	그리드별 습지, 강 등의 수변 비율		
	나지면적	그리드별 운동장, 공터 등 개방형 공간 비율	서울 열린 데이터 광장	서울시 생활권계획 시설(공원) 공간정보
	공원 면적	그리드별 공원 면적		서울시 생활권계획 시설(공원) 공간정보
				서울시 도시계획시설(공간 시설) 공간정보
인구	인구 수	그리드별 인구 수	국토정보맵	국토통계지도
생활편의 및 상권 시설 변수	병·의원 시설 수	그리드별 병·의원시설 수	공공데이터포털	소상공인시장진흥공단상가(상권)정보
	소매시설 수	그리드별 소매시설 수		
	숙박시설 수	그리드별 숙박시설 수		
	여가시설 수	그리드별 여가시설 수		
	교육시설 수	그리드별 교육시설 수	소상공인시장진흥공단 상가(상권)정보	기획재정부 공공기관 일반 현황
				V-WORLD
				학교기본정보(중)
				학교기본정보(고)
				학교기본정보(특수)
				학교기본정보(그 외)
	공공시설 수	그리드별 공공시설 수	국가도서관체계	서울시 자치화관 정보

변수 구분	데이터 명	설명	공공데이터 출처	데이터 명
			시스템 공공데이터포털	서울시 사회복지시설 목록
		세탁소 수	그리드별 세탁소 수	소상공인시장진흥공단상 가(상권)정보
		편의점 수	그리드별 편의점 수	
		술집 수	그리드별 술집 수	
		카페 수	그리드별 카페 수	
		음식점 수	그리드별 음식점 수	
		약국 수	그리드별 약국 수	
		주유소 수	그리드별 주유소 수	
		경찰서 수	그리드별 경찰서 수	
		CCTV 수	그리드별 CCTV 수	지방행정 인허 가 데이터
교통 및 기반 시설	주차장 수	그리드별 주차장 수	공공데이터포털	전국주차장정보표준
	버스 정류장 수	그리드별 버스정류장 수		전국 버스정류장 위치정보
	지하철역 수	그리드별 지하철역 수	카카오맵 기반 연구진 제작	
도로 공간 구조	도로 연결성	Space syntax로 계산한 그리 드별 평균 도로 연결성	V-WORLD	도로 중심선
	도로 통합도	Space syntax로 계산한 그리 드별 평균 도로 통합도		
	도로 각도 연결성	Space syntax로 계산한 그리 드별 평균 도로 각도 연결성		
	도로 레인 수	그리드별 평균 도로 레인 수		
	도로 폭	그리드별 평균 도로 폭		
	보도 연결성	Space syntax로 계산한 그리 드별 평균 보도 연결성	서울열린데이터 광장	보도 네트워크
	보도 통합도	Space syntax로 계산한 그리 드별 평균 보도 통합도		
	보도 각도 연결성	Space syntax로 계산한 그리 드별 평균 보도 통합도		
보행로 면적	그리드별 보행로 면적	V-WORLD	인도(보도)	

출처: 연구진 작성

■ 선정된 건축환경 변수 프로세스

• 인구 특성 변수

인구 변수로는 국토정보맵에서 제공하는 250m 그리드 단위의 인구 SHP 데이터를 기반으로 각 격자 내 나이대별 인구 수를 활용하였다. 나이대는 유아, 학생, 청년, 중장년, 고령으로 나누어 사용하였다. 인구는 물리환경 해석에 있어 기반 수요 요인을 설명하는 핵심 변수로 기능하며 공간 단위별 신고 건수 분석 시 주요 제어변수로 활용된다.

• 건물, 토지 특성 변수

건물 및 토지 특성 변수에는 건물 노후도, 건물 수, 건물 높이, 건폐율, 용적률, 업무시설 수 등이 포함된다. 해당 변수들은 GIS 기반의 건물통합정보 데이터를 활용하여 산출하였다. 건물 노후도는 각 건물의 사용승인일과 데이터 수집 시점 간의 차이를 계산하여 도출하였고 건물 수는 각 500m×500m 그리드 내 존재하는 건물의 개수로 산정하였다. 건물 높이, 건폐율, 용적률은 건물별 속성값을 기반으로 평균값을 계산하였으며 업무시설 수는 용도가 '업무시설'로 분류된 건물 수를 집계하여 산정하였다.

토지 특성 변수로는 혼합용도면적, 공업면적, 농업면적, 상업면적, 주거면적, 나지 면적, 녹지면적, 수변 면적, 공원면적 등이 있으며 이들 대부분은 토지특성도 및 토지피복도 SHP파일을 기반으로 추출하였다. 혼합용도면적은 주상용, 주상나지, 주상기타로 분류된 지역을 공업면적은 공업용, 공업나지, 공업기타로 분류된 지역을 포함한다. 농업면적은 전(밭), 과수원, 답(논)으로 상업면적은 상업용, 업무용, 상업나지, 상업기타로 분류된 지역을 통해 산정하였으며 주거면적은 단독, 연립, 다세대, 아파트, 주거나지, 주거기타 지역을 기준으로 추출하였다. 나지 면적, 녹지 면적, 수변 면적은 각각 나지, 산림·초지, 습지·수역으로 분류된 지역을 QGIS에서 추출하여 활용하였다. 공원면적은 서울열린데이터광장에서 제공하는 생활권계획 시설(공원), 도시자연공원구역, 도시계획시설(공간시설) 관련 공간정보 데이터를 기반으로 하며 이 중 '공원'이라는 키워드가 포함된 구역을 추출하고, AutoCAD DXF 및 라이노(Rhino)를 통해 도형 오류를 수정한 후 최종 활용하였다.

• 생활 편의 및 상권 시설 변수

생활 편의 및 상권 관련 변수로는 숙박시설 수, 병·의원 수, 여가시설 수, 음식점 수, 술집 수, 카페 수, 세탁소 수, 편의점 수, 약국 수, 주유소 수가 포함된다. 이들 모두는 2023년 소상공인시장진흥공단 상가(상권)정보 데이터를 기반으로 산정되었으며 대부분명이 '숙박업', '보건 의료업', '예술·스포츠 및 여가관련 서비스업'에 해당하는 데이터를 각각 집계하였다. 음식점 수는 대부분이 '음식점업'인 데이터 중 '주점업'과 '비알코올 음료점업'을 제외하여 산정하였고 해당 항목들은 각각 술집 수와 카페 수로 분리하여 추출하였다. 세탁소 수는 중분류가 '세

탁업'인 데이터를 편의점 수와 약국 수는 소분류가 각각 '편의점', '약국'인 항목을 기준으로 필터링하였다. 주유소 수는 '주유소' 및 '가스충전소' 소분류 항목을 결합하여 집계하였다.

교육시설 수는 소상공인시장진흥공단의 교육서비스업 분류 데이터와 함께 V-WORLD로부터 제공받은 초·중·고·특수학교 및 기타 학교에 대한 기본정보를 결합하여 추출하였다. 공공 시설 수는 서울열린데이터광장에서 제공하는 서울시 자치회관 정보 및 사회복지시설 목록 데이터를 주소 기반으로 좌표를 추출한 뒤 QGIS를 활용하여 결합하였다.

- 교통 및 기반 시설 변수

교통 및 기반 시설 변수에는 경찰서 수, 주차장 수, 지하철역 수, CCTV 수, 버스정류장 수, 가로등 수 등이 포함된다. 경찰서 및 주차장 데이터는 공공데이터포털에서 제공한 주소 정보를 파이썬을 이용해 지오코딩한 뒤, 이를 SHP 파일로 변환하여 QGIS에서 활용하였다. 지하철역 출구 데이터는 별도 공공데이터가 부재한 관계로 연구진이 카카오맵을 기반으로 QGIS에서 직접 포인트를 생성하여 제작하였다. CCTV와 버스정류장 수는 각각 지방행정 인허가 데이터와 공공데이터포털에서 제공된 좌표 데이터를 CSV로 받아 SHP 파일로 변환한 후 사용하였다. 가로등 수 또한 동일한 방식으로 CSV를 SHP로 변환하고 병합하여 활용하였다.

- 도로 공간 구조 변수

도로 공간 구조 변수는 도로 레인 수, 도로 폭, 보행로 면적, 그리고 도로 및 보도에 대한 연결성 및 통합도(ANG 연결성 포함) 등을 포함한다. 도로 레인 수와 도로 폭은 V-WORLD의 도로 중심선 속성값에서 직접 추출하였으며 보행로 면적은 V-WORLD의 도보 라인 데이터를 기반으로 자전거도로 항목을 제거한 뒤, 선형 데이터를 폴리곤으로 변환하고 오류를 수정한 후 서울시 경계 내 보행로만을 분리하여 계산하였다.

공간 연결성 관련 변수들은 DepthMapX 프로그램을 이용한 Space Syntax 분석을 통해 도출하였다. 도로 연결성, 도로 통합도, 도로 ANG 연결성은 서울시 및 인접 지역(경계 기준 반경 10km 이내)의 도로중심선 데이터를 활용하였고 보도 연결성, 보도 통합도, 보도 ANG 연결성은 서울시 도보네트워크 데이터를 동일한 방식으로 분석하여 추출하였다. 이들 Space Syntax 기반 지표들은 112신고 패턴이나 보행 패턴 분석 시 공간적 맥락을 반영하는 주요 분석 변수로 활용된다.

[표 4-4] 변수별 데이터 프로세싱 과정

변수 구분	데이터 명	설명	데이터 프로세싱 과정
인구 특성	총인구	면적 내 전체 연령 인구 수	국토정보맵의 국토통계지도에서 제공하는 250m 그리드 단위의 인구 데이터 shp파일 다운 및 사용

변수 구분	데이터 명	설명	데이터 프로세싱 과정
	유아 (0-7세)	면적 내 유아 인구 수	국토정보맵의 국토통계지도에서 제공하는 250m 그리드 단위의 인구 데이터 shp파일 다운 및 사용
	학생 (8-19세)	면적 내 학생 인구 수	국토정보맵의 국토통계지도에서 제공하는 250m 그리드 단위의 인구 데이터 shp파일 다운 및 사용
	청년 (20-30대)	면적 내 청년 인구 수	국토정보맵의 국토통계지도에서 제공하는 250m 그리드 단위의 인구 데이터 shp파일 다운 및 사용
	중년 (40-50대)	면적 내 중년 인구 수	국토정보맵의 국토통계지도에서 제공하는 250m 그리드 단위의 인구 데이터 shp파일 다운 및 사용
	고령 (60세 이상)	면적 내 고령 인구 수	국토정보맵의 국토통계지도에서 제공하는 250m 그리드 단위의 인구 데이터 shp파일 다운 및 사용
건물 및 토지 특성	건물 노후도	그리드별 사용승인일로부터 경과한 평균 년수	1. V-WORLD 디지털트윈국도 웹사이트에서 GIS건물통합 정보 SHP파일을 다운로드 2. 사용승인 날짜와 현재 날짜의 차이 계산
	건물 개수	그리드별 건물 개수	1. V-WORLD 디지털트윈국도 웹사이트에서 GIS건물통합 정보 SHP파일을 다운로드 2. 그리드 내 건물 수 추출
	건물 높이	그리드별 평균 건물 높이	1. V-WORLD 디지털트윈국도 웹사이트에서 GIS건물통합 정보 SHP파일을 다운로드 2. 건물 층수를 사용하여 높이 계산
	건폐율	그리드별 평균 건폐율	1. V-WORLD 디지털트윈국도 웹사이트에서 GIS건물통합 정보 SHP파일을 다운로드 2. 제공된 건폐율 데이터 사용
	용적률	그리드별 평균 용적률	1. V-WORLD 디지털트윈국도 웹사이트에서 GIS건물통합 정보 SHP파일을 다운로드 2. 제공된 용적률 데이터 사용
	업무시설 수	그리드별 업무시설 수	1. V-WORLD 디지털트윈국도 웹사이트에서 GIS건물통합 정보 SHP파일을 다운로드 2. 업무시설로 지정되어있는 건물 추출
	혼합용도 면적	그리드별 혼합용 토지 면적	토지 특성도 SHP파일에서 주상용, 주상나지, 주상기타로 분류되어있는 지역을 QGIS에서 추출함
	공업 면적	그리드별 공업용 토지 면적	토지 특성도 SHP파일에서 공업용, 공업나지, 공업기타로 분류되어있는 지역을 QGIS에서 추출함
	농업 면적	그리드별 농업용 토지 면적	토지 특성도 SHP파일에서 전, 과수원, 답으로 분류되어있는 지역을 QGIS에서 추출함
	상업 면적	그리드별 상업용 토지 면적	토지 특성도 SHP파일에서 상업용, 업무용, 상업나지, 상업기타로 분류되어있는 지역을 QGIS에서 추출함
	주거 면적	그리드별 주거용 토지 면적	토지 특성도 SHP파일에서 단독, 연립, 다세대, 아파트, 주거나지, 주거기타로 분류되어있는 지역을 QGIS에서 추출함
	녹지 면적	그리드별 나무, 풀 등의 녹지 비율	토지피복도 SHP파일에서 산림지역, 초지로 분류되어있는 지역을 QGIS에서 추출함

변수 구분	데이터 명	설명	데이터 프로세싱 과정
	수변 면적	그리드별 습지, 강 등의 수변 비율	토지피복도 SHP파일에서 습지, 수역으로 분류되어있는 지역을 QGIS에서 추출함
	나지 면적	그리드별 운동장, 공터 등 개방형 공간 비율	토지피복도 SHP파일에서 나지로 분류되어있는 지역을 QGIS에서 추출함
	공원 면적	그리드별 공원 면적	1. 서울 열린데이터 광장에서 서울시 생활권계획 시설(공원) 공간정보, 서울시 용도구역(도시자연공원구역) 공간정보, 서울시 도시계획시설 (공간시설) 공간정보 다운로드 2. 서울시 도시계획시설 (공간시설) 공간정보 데이터에서 "공원"이라는 단어가 들어간 구역 추출 3. 추출된 shp파일을 QGIS에서 결합
생활편의 및 상권 시 설 변수	병·의원 시설 수	그리드별 병·의원시설 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종대분류명이 '보건의료업'인 데이터 추출
	소매시설 수	그리드별 소매시설 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종대분류명이 '소매업'인 데이터 추출 3. 소분류명이 '주유소', '가스충전소', '약국' 인 데이터 제외
	숙박시설 수	그리드별 숙박시설 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종대분류명이 '숙박업'인 데이터 추출
	여가시설 수	여가시설 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종대분류명이 '숙박업'인 데이터 추출
	공공시설 수	그리드별 공공시설 수	1. 서울시 자치회관 정보와 서울시 사회복지시설 목록의 데이터를 주소를 기반으로 좌표 추출 2. CSV 형태인 모든 데이터를 QGIS에서 SHP파일로 전환 3. 전환한 모든 레이어를 QGIS에서 병합
	교육시설 수	그리드별 교육시설 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. V-WORLD에서 학교기본정보(고), 학교기본정보(중), 학교기본정보(초), 학교기본정보(그 외), 학교기본정보(특수) 다운로드 3. 상권업종대분류명이 '교육 서비스업'인 데이터 추출 4. QGIS에서 모든 레이어를 병합하여 사용
	세탁소 수	그리드별 세탁소 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종중분류명이 '세탁업'인 데이터 추출
	편의점 수	그리드별 편의점 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종소분류명이 '편의점'인 데이터 추출
	술집 수	그리드별 술집 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종중분류명이 '주점업'인 데이터 추출

변수 구분	데이터 명	설명	데이터 프로세싱 과정
교통 및 기반 시설	카페 수	그리드별 카페 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종소분류명이 '카페'인 데이터 추출
	음식점 수	그리드별 음식점 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종대분류명이 '음식점업'인 데이터 추출 3. 추출된 데이터중 중분류명이 '주점업', '비알코올 음료점업'인 곳은 제외.
	약국 수	그리드별 약국 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드
	주유소 수	그리드별 주유소 수	1. 공공데이터포털에서 소상공인시장진흥공단 상가(상권)정보 데이터 다운로드 2. 상권업종소분류명이 '주유소'와 '가스충전소'인 데이터를 결합하여 사용
	경찰서 수	그리드별 경찰서 수	1. 행정안전부의 공공데이터포털에서 '경찰관서 위치 주소 현황' 데이터 CSV 파일 다운 2. Python에서 경찰관서 데이터의 주소를 지오코딩하여 좌표를 CSV로 추출 3. 데이터셋의 좌표를 기반으로 QGIS에서 CSV파일을 SHP 파일로 변환
	CCTV 수	그리드별 CCTV 수	1. 행정안전부의 지방행정 인허가 데이터로부터 'CCTV' 데이터 CSV 파일 다운 2. 데이터셋의 좌표를 기반으로 QGIS에서 CSV파일을 SHP 파일로 변환
	주차장 수	그리드별 주차장 수	1. 행정안전부의 공공데이터포털에서 '전국주차장정보표준' 데이터 CSV 파일 다운 2. Python에서 '주차장' 데이터의 주소를 지오코딩하여 좌표를 CSV로 추출 3. 데이터셋의 좌표를 기반으로 QGIS에서 CSV파일을 SHP 파일로 변환
	버스정류장 수	그리드별 버스정류장 수	1. 행정안전부의 공공데이터포털에서 '전국 버스정류장 위치 정보' 데이터 CSV 파일 다운 2. 데이터셋의 좌표를 기반으로 QGIS에서 CSV파일을 SHP 파일로 변환
도로 공간 구조	지하철역 수	그리드별 지하철역 수	QGIS에 카카오맵 연동 지하철 출구 위치에 맞춰 포인트 레이어 생성
	도로 연결성	Space syntax로 계산한 그리드별 평균 도로 연결성	1. 국토교통부의 V-WORLD 디지털트윈국도 웹사이트에서 '도로중심선' SHP파일 다운 2. 서울 경계의 10km버퍼 레이어를 생성하여 서울지역 도로 중심선 데이터 추출 3. 추출한 데이터를 DepthmapX를 통해 도로 연결성 추출
도로 통합도	Space syntax로 계산한 그리드별 평균 도로 통합도	1. 국토교통부의 V-WORLD 디지털트윈국도 웹사이트에서 '도로중심선' SHP파일 다운 2. 서울 경계의 10km버퍼 레이어를 생성하여 서울지역 도로 중심선 데이터 추출 3. 추출한 데이터를 DepthmapX를 통해 도로 통합도 추출	

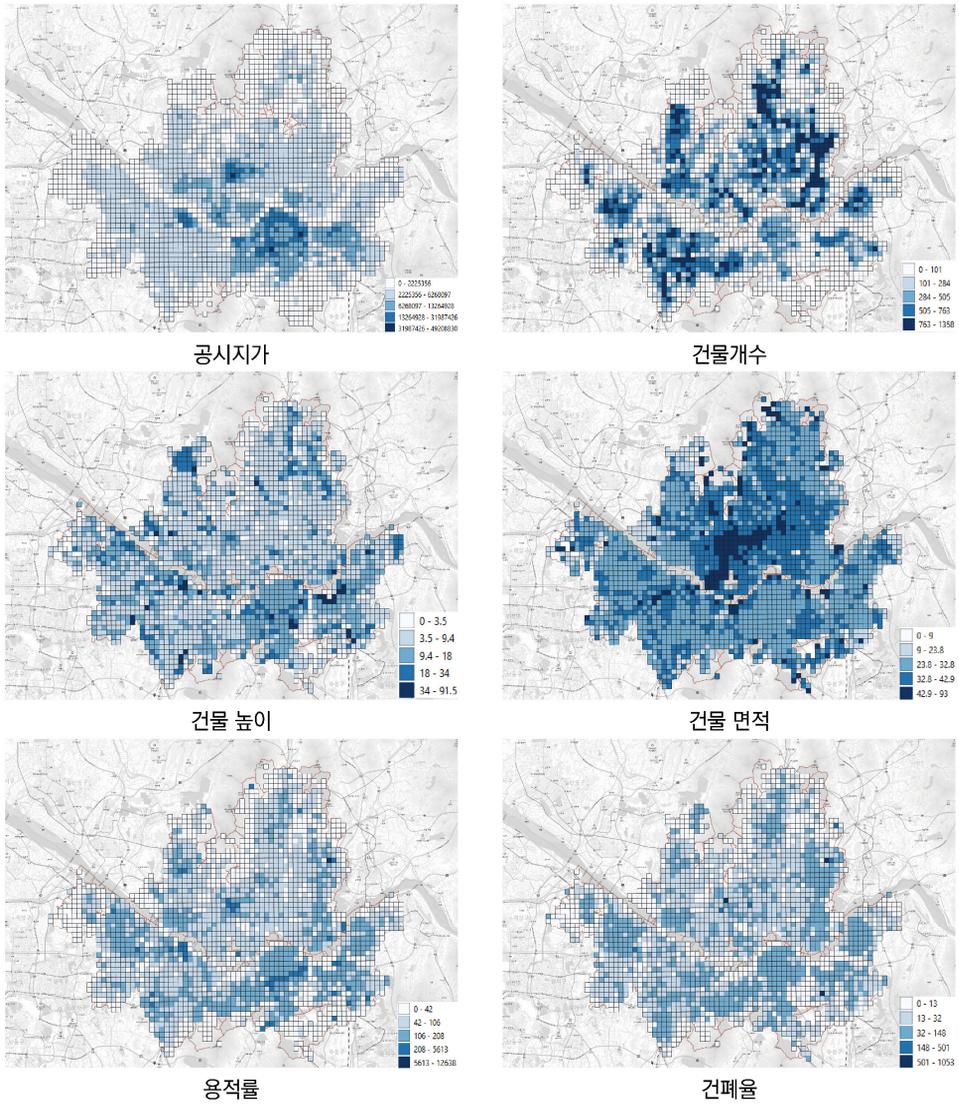
변수 구분	데이터 명	설명	데이터 프로세싱 과정
	도로 ANG 연결성	Space syntax로 계산한 그리드별 평균 도로 각도 연결성	1. 국토교통부의 V-WORLD 디지털트윈국도 웹사이트에서 '도로중심선' SHP파일 다운 2. 서울 경계의 10km버퍼 레이어를 생성하여 서울지역 도로 중심선 데이터 추출 3. 추출한 데이터를 도로 ANG 연결성 추출 (DepthmapX)
	보도 연결성	Space syntax로 계산한 그리드별 평균 보도 연결성	1. 서울열린데이터 광장에서 도보 네트워크 데이터 다운 2. DepthmapX를 통해 도보네트워크의 보도 연결성 추출
	보도 통합도	Space syntax로 계산한 그리드별 평균 보도 통합도	1. 서울열린데이터 광장에서 도보 네트워크 데이터 다운 2. DepthmapX를 통해 도보네트워크의 보도 통합도 추출
	보도 ANG 연결성	Space syntax로 계산한 그리드별 평균 보도 각도 연결성	1. 서울열린데이터 광장에서 도보 네트워크 데이터 다운 2. 도보네트워크의 보도 ANG 연결성 추출 (DepthmapX)
	도로 레인 수	그리드별 평균 도로 레인 수	1. 국토교통부의 V-WORLD 디지털트윈국도 웹사이트에서 '도로중심선' SHP파일 다운 2. 제공된 도로 레인 수 활용
	도로 폭	그리드별 평균 도로 폭	1. 국토교통부의 V-WORLD 디지털트윈국도 웹사이트에서 '도로중심선' SHP파일 다운 2. 제공된 도로 폭 데이터 활용
	보행로 면적	그리드별 보행로 면적	1. 국토교통부의 V-WORLD 디지털트윈국도 웹사이트에서 '인도(도보)' SHP파일 다운 2. 자전거 도로 삭제 (QGIS) 3. 선으로 이루어진 파일을 폴리곤으로 전환 (QGIS) 4. 잘못 전환된 폴리곤 삭제 5. 대상지에 맞게 추출하여 사용

출처: 연구진 작성



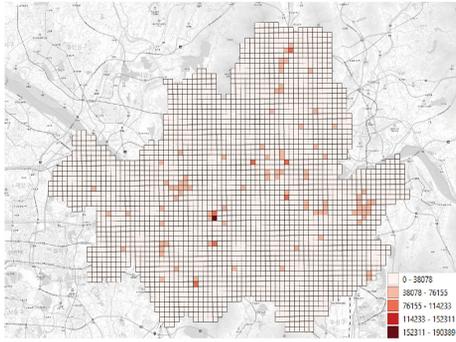
[그림 4-3] 독립변수의 공간적 분포 시각화 (1)

출처: 연구진 작성

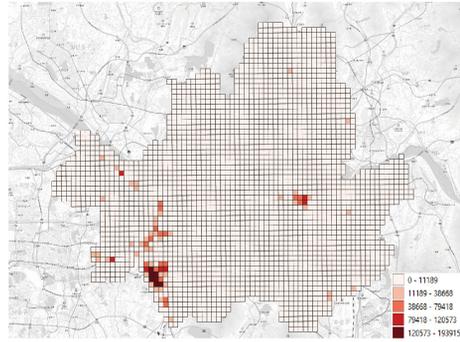


[그림 4-4] 독립변수의 공간적 분포 시각화 (2)

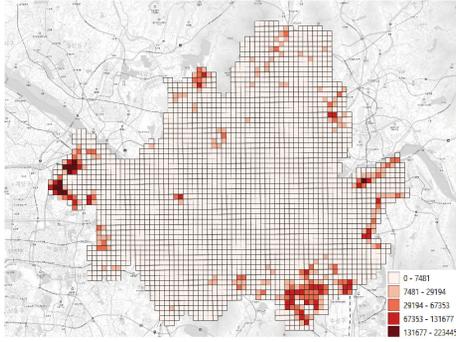
출처: 연구진 작성



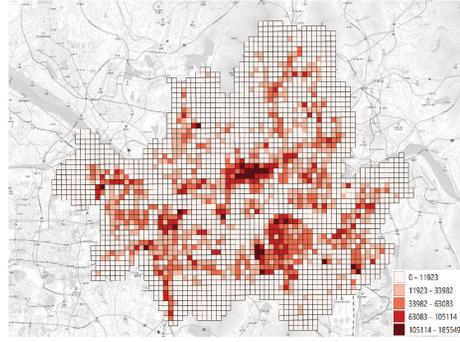
주상복합면적



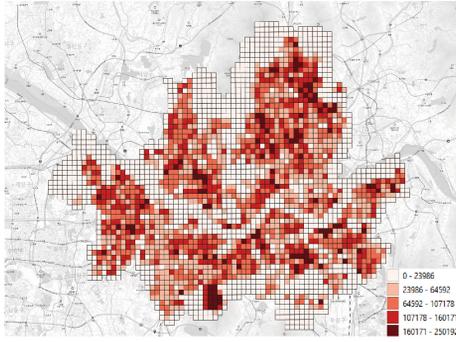
공업면적



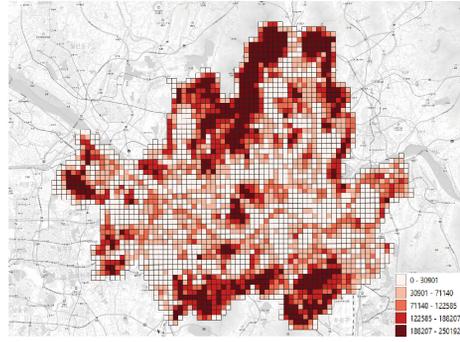
농지면적



상업면적



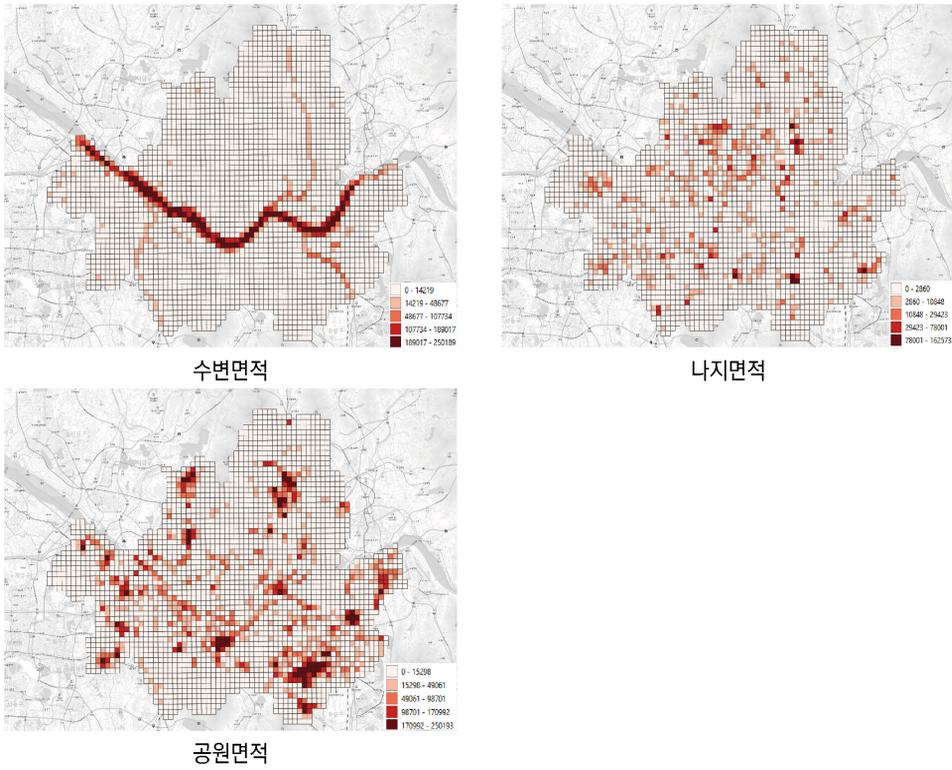
주거면적



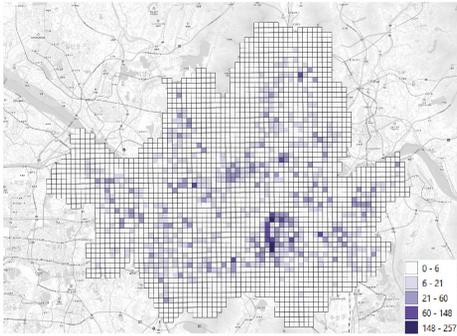
녹지면적

[그림 4-5] 독립변수의 공간적 분포 시각화 (3)

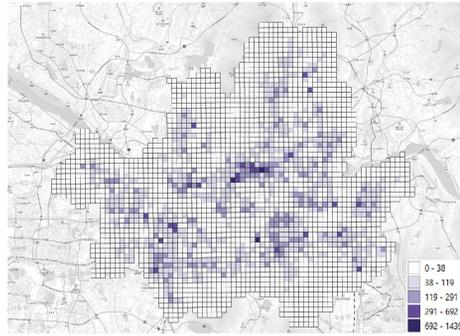
출처: 연구진 작성



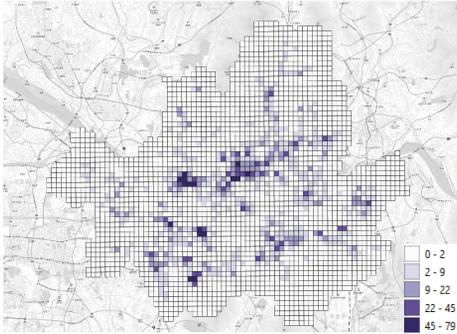
[그림 4-6] 독립변수의 공간적 분포 시각화 (4)
출처: 연구진 작성



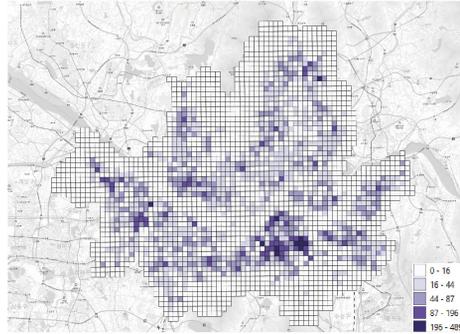
병의원 수



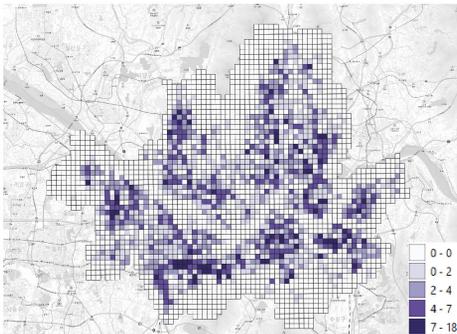
소매시설 수



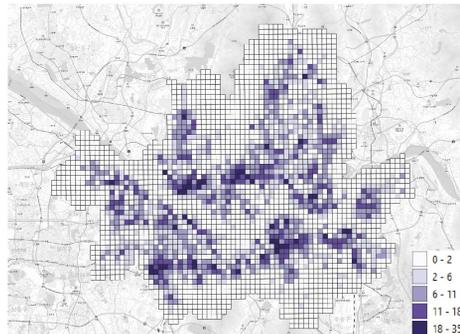
숙박시설 수



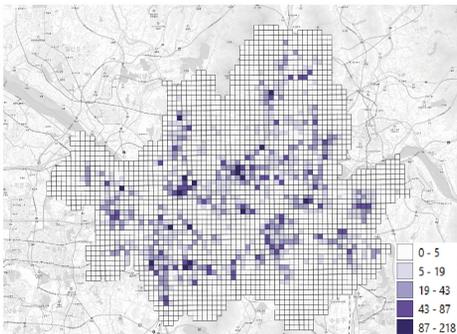
공공·교육시설 수



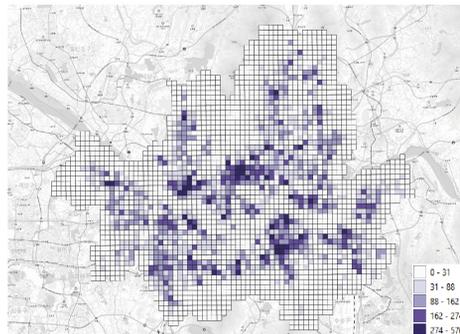
세탁소 수



편의점 수



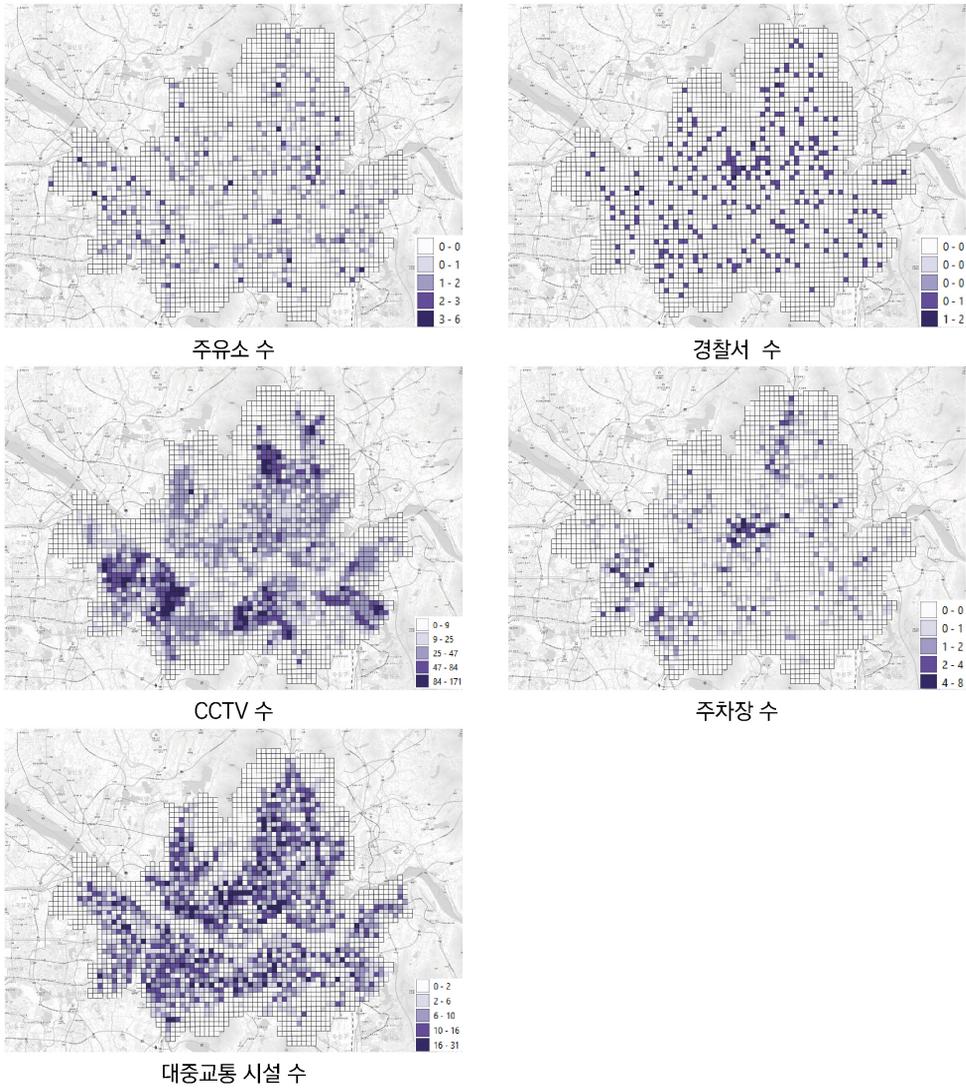
술집 수



음식점·카페 수

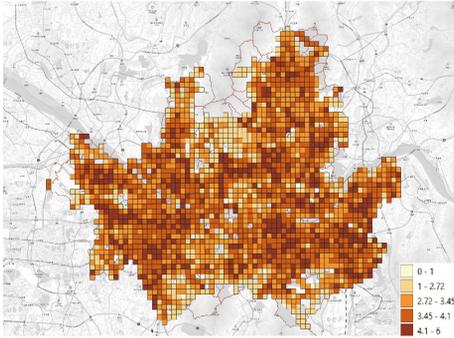
[그림 4-7] 독립변수의 공간적 분포 시각화 (5)

출처: 연구진 작성

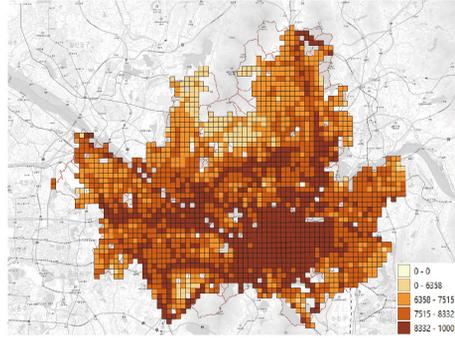


[그림 4-8] 독립변수의 공간적 분포 시각화 (6)

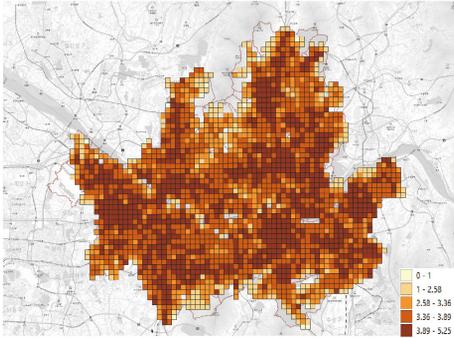
출처: 연구진 작성



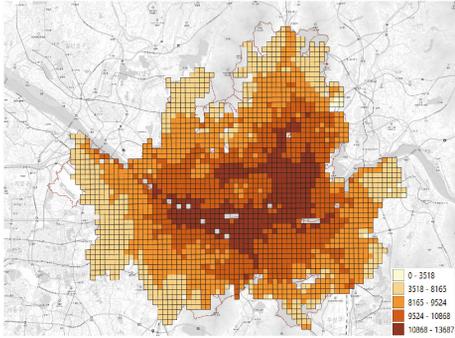
도로 연결성



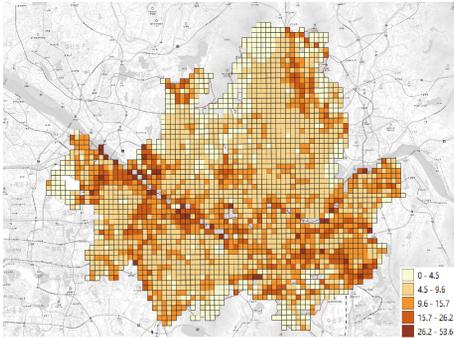
도로 통합도



도로 연결성



도로 통합도



도로 폭

[그림 4-9] 독립변수의 공간적 분포 시각화 (7)

출처: 연구진 작성

■ 데이터 전처리 및 변수 처리 전략

본 연구에서 수행한 데이터 전처리는 112신고 발생 건축환경 예측 모델 개발의 핵심 단계로, 모델 학습에 적합한 형태로 데이터를 정제하고 구조화하는 과정이다. 이 과정은 예측 성능 향상뿐만 아니라 모델이 변수 간의 관계를 보다 정확하게 학습할 수 있도록 도와준다. 주요 처리 단계로는 변수 스케일 조정, 이상치 처리, 결측치 제거, 다중공선성 검토 및 변수 선택 등이 포함된다.

• 변수 스케일 조정

예측 모델에 사용된 변수들은 서로 다른 단위와 범위를 가지므로, 스케일의 불일치를 해소하지 않으면 분산이 큰 변수가 모델에 과도한 영향을 미칠 수 있다(James et al., 2013). 이를 방지하기 위해 변수의 값을 표준화(Standardization)하여 평균이 0, 표준편차가 1이 되도록 조정하였다. 이러한 스케일 통일은 변수 간 비교를 가능하게 하고 모델 수렴 속도와 예측 정확도를 향상시키는데 기여한다.

• 이상치 처리

이상치는 데이터 수집 과정의 오류, 입력 실수, 또는 실제 극단값으로 인해 발생할 수 있으며 모델 학습에 부정적인 영향을 미친다(James et al., 2013). 본 연구에서는 정규분포를 가정한 Z-score 방법을 활용하여 독립변수의 이상치를 탐지하였으며, Z-score 값이 ± 3 을 초과하는 데이터를 제거하였다(Iglewicz & Hoaglin, 1993). 이를 통해 전체적인 데이터 분포를 왜곡시키는 극단값의 영향을 최소화하였다.

• 결측치 제거

결측치는 분석의 정확도와 모델의 일반화 능력을 저하시킬 수 있는 요인이며 적절한 처리 전략이 필요하다(Little & Rubin, 2019). 본 연구에서는 평균, 중앙값, 최빈값 대체 방법 대신, 분석 단위인 3,456개의 그리드 중 인구가 0명인 그리드, 건물이 전무한 그리드, 중복된 그리드를 제거하여 최종적으로 2,713개의 유효 그리드를 확보하였다. 이는 결측 및 분석 불가능 지역을 사전 제거함으로써 모델의 학습 효율성과 신뢰성을 높이기 위한 조치였다.

• 다중공선성과 상관계수 낮은 변수 선택

다중공선성은 두 개의 변수가 아닌 세 개 이상의 변수 간의 상관관계가 있는 것을 말한다(James et al., 2013). 변수 간 다중공선성(Multicollinearity)은 회귀 계수의 불안정성과 해석력 저하를 초래할 수 있으며 이는 모델의 예측 성능에도 부정적인 영향을 미친다. 본 연구에서는 de Jongh et al.(2015)의 기준에 따라 VIF(Variance Inflation Factor) 값이 15 이상인 변수를 다중공선성이 높다고 판단하여 제거하였다. 이어서 Gregorutti et al.(2017)의 제안에 따라 상관계수(Correlation Coefficient) 값이 0.7 이상인 변수 쌍 중 하나를 추가적으

로 제외하였다. 이 과정을 통해 변수 간 중복성과 예측 모델의 과적합(overfitting) 가능성을 낮추고, 해석이 가능한 모델 구조를 확보하였다.

3) 공간정보 분석 및 기계학습 방법론

■ 공간 자기상관 분석 (Moran's I, LISA)

공간 자기상관 분석은 연구 대상 현상이 지리적으로 인접한 단위들 사이에서 얼마나 유사하게 분포하는지를 정량적으로 평가하는 기법으로, 본 연구에서는 서울시 내 112신고 건수의 공간 분포 특성을 심층적으로 탐색하고자 한다. 이를 위해 먼저 전역적 지표(Global Moran's I)를 활용하여 전체적 패턴을 식별하고, 이어서 국지적 지표(LISA: Local Indicators of Spatial Association)를 적용하여 세부 클러스터링 양상을 파악하고자 한다. 이와 같은 다단계 분석 절차를 바탕으로 고밀도 지역과 저밀도 지역을 명확히 구분하고, 이후 기후·환경·인구 사회경제적 변수와의 관계 분석을 위한 기초 자료를 구축할 예정이다.

• 전역적 공간 자기상관 (Global Moran's I)

전역 Moran's I는 전체 연구 지역 내에서 변수(112신고 건수)가 공간적으로 유사한 경향을 가지는지를 단일 지표로 나타낸다(Moran, 1950). 해당 지표는 -1부터 +1 사이의 값을 가지며, +1에 가까울수록 유사한 값끼리 인접하여서 모여 있는 경향(positive spatial autocorrelation), -1에 가까울수록 서로 다른 값들이 인접한 경향(negative spatial autocorrelation)을 의미한다. 값이 0에 가까운 경우 공간적 무작위성(randomness)을 시사한다. 전역 Moran's I는 전체 연구 영역에서 변수 간 공간적 자기상관의 전반적인 강도를 파악하기 위해 다음의 식에 따라 계산된다:

$$I = \frac{n \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_i (x_i - \bar{x})^2}$$

여기서 n 은 관측 격자 수, x_i 는 i 격자의 112신고 건수, \bar{x} 는 격자별 신고 건수의 평균, w_{ij} 는 이웃 관계를 표현하는 공간 가중치, W 는 모든 w_{ij} 의 합을 의미한다. 산출된 I 값은 -1에서 +1 사이를 가지며 +값일수록 유사한 신고 빈도 간 군집(positive autocorrelation), -값일수록 이질적 값 간 군집(negative autocorrelation), 0 근처는 무작위 분포를 나타낸다. 본 연구에서는 GeoDa 소프트웨어를 사용하여 공간 가중치 행렬을 구축하고 Moran's I를 계산할 예정이다.

- 국지적 공간 자기상관 (LISA)

전역 Moran's I는 전체 공간 단위에서의 평균적인 공간 자기상관 경향을 파악하는 데 유용한 반면, LISA(Local Moran's I)는 개별 격자 수준에서의 국지적 공간 자기상관을 평가하여 지역별 클러스터 유형을 도출하는 데 활용된다(Anselin, 1995). 특히, LISA 분석은 전역 Moran's I가 포착하지 못하는 개별 격자 차원의 국지적 패턴을 검출하기 위해 고안되었으며 LISA 분석은 각 격자의 값이 주변 격자의 값들과 얼마나 유사하거나 이질적인지를 비교함으로써 공간적 패턴을 네 가지 유형으로 분류한다. 구체적으로 높은 값이 주변의 높은 값들과 인접한 경우는 High-High(Hotspot) 낮은 값이 주변의 낮은 값들과 인접한 경우는 Low-Low(Coldspot)으로 정의된다. 반면, 높은 값이 낮은 값들에 둘러싸인 경우는 High-Low, 낮은 값이 높은 값들에 둘러싸인 경우는 Low-High로 분류된다.

[표 4-5] LISA 분석 결과 분류

	주변격자 : 높은 값	주변격자 : 낮은 값
해당 격자: 높은 값	High-High (Hotspot)	High-Low
해당 격자: 낮은 값	Low-High	Low-Low (Coldspot)

출처: 연구진 작성

LISA 분석 결과는 ArcGIS 또는 GeoDa에서 제공하는 Moran Scatterplot과 LISA cluster map 기능을 활용하여 시각화할 계획이다. 이를 통해 서울시 내 고위험 지역(Hotspot)과 위험이 상대적으로 낮은 지역(Coldspot)의 공간적 경계를 정확히 식별하고 각 클러스터 유형별로 시간 요인(낮, 밤), 녹지 비율, 인구밀도, 건물용도 밀도 등의 환경 변수 값을 비교·분석함으로써 112신고 발생 메커니즘에 대한 종합적인 이해를 위해 활용될 예정이다.

■ 통계 및 머신러닝 기반 예측 모형

본 연구는 2024년 경찰청의 범죄 예방 우수시설 인증제도 관련 연구의 후속 연구로 112신고 데이터와 도시건축환경 요소 간의 관계를 분석하는 것을 목적으로 한다. 이를 위해 시간대 및 계절 요인을 고려하여 112신고 건수에 영향을 미치는 도시건축환경 요소를 분석하며 머신러닝 기반 회귀모형과 공간회귀모형을 주요 예측 모델로 활용하였다.

- XGBoost
 - XGBoost 회귀모형(XGBRegressor)

XGBoost는 기존의 결정트리 기반 부스팅 알고리즘을 개선한 고성능 머신러닝 모델로, 회귀 문제에 적용될 때 여러 개의 간단한 결정트리를 순차적으로 학습시키며 점점 더 정교한 예측

을 만들어낸다(Chen, et al., 2016). 각 단계에서 모델은 이전 예측의 오류를 보완하는 방향으로 학습되며 이러한 방식으로 전체 예측 성능을 지속적으로 향상시킨다. XGBoost의 가장 큰 특징 중 하나는 과적합을 방지하기 위한 정규화 기능으로 모델이 지나치게 복잡해지는 것을 막기 위해 트리의 크기나 구조에 패널티를 부여한다. 이를 통해 학습 데이터에만 과하게 맞춘 모델이 되는 것을 방지하고 새로운 데이터에 대해서도 일반화 성능을 유지한다. 또한 XGBoost는 대규모 데이터에도 효과적으로 작동할 수 있도록 병렬 처리, 희소 데이터 최적화, 캐시 활용, 디스크 기반 학습 등 다양한 시스템 최적화 기술을 포함하고 있어 실제 환경에서 빠르고 안정적인 예측이 가능한 모델로 널리 활용되고 있다.

- XGBoost 모델 예측 성능 평가

XGBoost 모델의 성능 평가는 결정계수(R^2), 평균제곱오차(MSE), 평균절대오차(MAE), 그리고 RMSE(Root Mean Squared Error)와 같은 다양한 지표를 기반으로 수행된다. 결정계수는 종속 변수의 변동성 중 모델이 설명하는 비율을 나타내며 1에 가까울수록 예측력이 높다고 해석된다. 그러나 이 지표만으로는 모델이 과적합되었는지 여부를 판단하기 어렵기 때문에 다른 오류 지표들과 병행하여 해석하는 것이 필요하다(Wooldridge, 2016).

RMSE는 예측 오차의 제곱 평균의 제곱근으로 계산되며 오차가 클수록 더 큰 패널티를 부여하는 특징이 있다. 값이 작을수록 예측이 실제값에 가까움을 의미하지만 이상치에 민감하다는 단점이 있다. MAE는 예측값과 실제값 간의 절대 오차의 평균으로 계산되며 모든 오차를 동일한 비중으로 반영하기 때문에 해석이 직관적이고 이상치의 영향을 덜 받는다(Wooldridge, 2015).

• LightGBM

- LightGBM 주요 해석 지표

LightGBM은 예측 정확도가 높은 부스팅 기반 머신러닝 기법으로 성능뿐 아니라 모델의 해석 가능성도 중요하게 다루어진다. 특히 공공정책 의사결정에 활용되는 예측모형의 경우, 단순한 예측 결과 이상으로 각 변수의 영향력을 구조적으로 해석할 수 있는 지표가 요구된다. 이에 따라 LightGBM에서는 변수 중요도(feature importance), SHAP(SHapley Additive exPlanations), 그리고 잔차 분석(residual analysis) 등이 주요 해석 지표로 활용된다.

첫째, 변수 중요도는 모델이 예측을 수행하는 과정에서 각 입력변수가 얼마나 자주 분할 기준으로 사용되었는지를 기준으로 산출된다. 주로 split 기준(변수가 사용된 빈도)과 gain 기준(변수가 모델 성능 개선에 기여한 정도)으로 구분되며 이를 통해 각 변수의 상대적 영향력을 확인할 수 있다. 변수 중요도는 모델의 전반적 구조를 요약하고 주요 결정 요인을 파악하는데 유용하나 상호작용 효과나 비선형 관계까지는 반영하지 못하는 한계가 있다(Lundberg et al., 2020).

둘째, SHAP 값은 게임이론 기반의 기여도 분석 방법으로 각 예측값에 대해 모든 변수의 기여도를 정량화한다. SHAP는 각 입력값이 예측값을 기준값(base value)에서 얼마나 변화시켰는지를 나타내며 모델 해석의 정밀성과 직관성을 동시에 제공한다. 특히 SHAP summary plot, dependence plot 등 시각화 기법을 활용하면 변수 간 상호작용 및 예측 민감도를 명확히 파악할 수 있다(Molnar, 2022). SHAP 값은 모델 구조에 관계없이 계산할 수 있는 장점이 있으나 연산 복잡도가 높은 단점도 존재한다.

셋째, 잔차 분석은 예측값과 실제값 간의 차이를 분석하여 모델의 적합성 및 편향 여부를 진단하는 기법이다. 특히 잔차의 분포, 오차 패턴, 특정 변수 수준에서의 체계적 과소 또는 과대 예측 여부 등을 통해 모델의 구조적 한계나 데이터 특성에 대한 부적합 가능성을 점검할 수 있다. 이는 모델 성능 개선뿐만 아니라 정책적 해석의 타당성을 확보하는 데 필수적인 절차이다 (James et al., 2013).

- LightGBM 예측 성능 평가

LightGBM 모델의 예측 성능을 평가하기 위해 주로 사용되는 지표는 정확도, 정밀도, 재현율, F1 점수, 그리고 AUC-ROC 등으로 구분된다. 이들 지표는 각각 예측 결과의 다양한 측면을 정량화하여 모델의 분류 성능을 종합적으로 판단하는 데 기여한다(Yadav et al., 2022; Ke et al., 2017). 먼저 정확도는 전체 예측 중에서 실제 값과 일치한 비율을 의미하며 모델이 얼마나 많은 데이터를 정확히 예측했는지를 보여준다. 그러나 클래스 불균형 문제가 존재하는 경우, 정확도는 왜곡된 성능 평가를 초래할 수 있다(Zhang et al., 2021).

정밀도는 모델이 양성으로 예측한 사례 중 실제로 양성인 비율을 나타내며 오탐(false positive)을 얼마나 줄였는지를 평가하는 데 유용하다. 반면, 재현율은 실제 양성 중에서 모델이 양성으로 올바르게 예측한 비율로서 누락(false negative)을 얼마나 줄였는지를 판단하는 기준이 된다. 이 두 지표는 상충하는 경향이 있으므로 이를 균형 있게 고려하기 위해 F1 점수가 사용된다. F1 점수는 정밀도와 재현율의 조화평균으로 두 지표의 불균형을 조정하고 모델의 종합적인 분류 성능을 평가한다(Yadav et al., 2022).

또한 AUC-ROC는 모델의 분류 경계값 변화에 따른 민감도(재현율)와 특이도 간의 trade-off를 시각화한 곡선이며 그 아래 면적(AUC)을 통해 모델의 전반적인 분류 능력을 수치화한다. AUC 값이 1에 가까울수록 이상적인 분류 성능을 갖춘 것으로 간주된다(Ke et al., 2017).

LightGBM 모델은 이러한 지표들을 기반으로 모델 학습 단계에서 성능을 검증하고 최적의 하이퍼파라미터를 탐색하며 모델의 일반화 능력을 평가하는 데 활용된다. 이러한 지표들은 특히 공공 정책 분석 및 예측 모델 개발 과정에서 모델의 신뢰성과 타당성을 확보하는 데 필수적인 도구로 기능한다(Zhang et al., 2021).

- 다중선형회귀(MLR)

- MLR 주요 해석 지표

다중선형회귀(Multiple Linear Regression, MLR)는 가장 기본적인면서도 널리 사용되는 통계 분석 기법으로 독립 변수들이 종속 변수에 미치는 영향을 동시에 평가할 수 있다는 장점이 있다. MLR 분석 결과를 해석할 때는 단순한 계수 해석을 넘어 모델의 설명력, 변수 유의성, 회귀 가정 만족 여부 등을 종합적으로 고려할 필요가 있다.

MLR의 대표적인 해석 지표는 결정계수(R^2)와 수정된 결정계수(Adjusted R^2)이다. 결정계수는 모델이 종속 변수의 분산을 얼마나 설명하는지를 나타내며 값이 1에 가까울수록 높은 설명력을 의미한다. 그러나 독립 변수가 많을수록 R^2 는 인위적으로 증가할 수 있기 때문에 실제 비교나 해석에는 Adjusted R^2 가 더 적합하다. 이는 변수 개수와 표본 수에 따라 조정된 값으로 과적합 위험을 일정 부분 반영하는 장점이 있다. 실제로 MLR을 사용할 때 두 지표를 동시에 고려하는 것이 바람직하다는 점은 여러 연구에서도 강조된 바 있다(Zhang et al., 2018). 개별 변수의 영향력은 회귀 계수(β)와 그에 대한 유의성 검정(p-value)을 통해 파악할 수 있다. 회귀 계수는 해당 독립 변수가 종속 변수에 미치는 효과의 크기와 방향을 나타내며 p-value는 그 효과가 통계적으로 유의한지를 판단하는 기준이다. 표준화된 회귀계수(Standardized β)를 사용하면 서로 다른 단위를 가진 변수들 간 영향력을 비교할 수 있다는 점에서 해석력과 직관성을 높일 수 있다.

- MLR 예측 성능 평가

MLR 모델의 해석 지표가 모형의 적합도를 평가하는 데 유용하다면 예측 성능 평가는 모델의 실질적인 활용 가능성을 검증하는 절차라 할 수 있다. 특히 학습 데이터에 대한 높은 설명력이 반드시 새로운 데이터에서도 재현된다는 보장은 없기 때문에 별도의 평가 절차가 반드시 필요하다.

테스트 데이터 기반의 평가에서는 RMSE(Root Mean Squared Error), MAE(Mean Absolute Error) 등의 예측 오차 지표가 사용된다. 이 중 RMSE는 오차의 제곱 평균의 제곱근으로 계산되며 이상치에 민감하다는 특성이 있다. 반면 MAE는 절대값 기준이기 때문에 상대적으로 이상치의 영향을 덜 받는다. 이러한 다양한 지표를 병렬적으로 활용하여 모델의 예측력을 다면적으로 평가하는 것이 필요하다(Yadav et al., 2022).

또한, 모델의 일반화 가능성을 높이기 위해 AIC, BIC 등의 기준을 활용해 변수 선택을 수행할 수 있다. 이들 지표는 모델의 복잡도와 설명력을 동시에 고려하여 과도한 변수 투입에 따른 과적합을 방지하는 데 유용하다(Lee & Park, 2019).

■ 공간 회귀 기반 예측 모형

- 지리 가중 회귀모형(GWR)
 - GWR 주요 해석 지표

GWR(Geographically Weighted Regression, 지리적으로 가중된 회귀분석)은 공간적 비정상성(spatial non-stationarity)을 반영하기 위해 개발된 통계적 분석 기법으로 동일한 회귀계수를 전체 지역에 일률적으로 적용하는 전통적 회귀모형의 한계를 보완한다(Fotheringham et al., 2002). 즉, GWR은 공간상 각 지점에서 독립변수와 종속변수 간의 관계가 다르게 나타날 수 있음을 전제로 하며 이를 고려하여 각 위치별로 상이한 회귀계수를 추정한다. 이러한 접근은 지역적 이질성이 강한 사회경제적, 환경적, 혹은 도시공간적 현상을 분석할 때 매우 유용하게 활용된다(Chen et al., 2020).

GWR의 주요 해석 지표는 크게 세 가지로 구분된다. 첫째, 국소 회귀계수(local coefficients)는 각 관측지점별로 독립변수가 종속변수에 미치는 영향의 크기와 방향을 나타낸다. 이를 통해 특정 변수의 효과가 공간적으로 어디에서 강하고 약한지를 식별할 수 있으며 공간적 불균형이나 지역적 특성을 세밀하게 파악할 수 있다(Lu et al., 2014). 예를 들어 인구밀도나 접근성 변수의 영향력이 도시 중심부와 외곽 지역에서 상이하게 나타나는 경우, 국소 회귀계수의 공간적 분포를 통해 그 차이를 시각적으로 검증할 수 있다.

둘째, 대역폭(bandwidth)은 GWR의 핵심 매개변수로 공간적 가중함수(spatial weighting function)의 범위를 결정한다. GWR은 모든 지역의 대역폭이 동일하다는 것을 가정한다. 대역폭이 너무 작으면 모델이 과적합(overfitting)될 위험이 있고 반대로 너무 크면 공간적 차이를 반영하지 못하는 과도한 평활화(smoothing)가 발생한다. 따라서 대역폭은 모델의 정확성과 일반화 가능성 사이의 균형을 조절하는 중요한 인자이며 일반적으로 AICc 최소화 또는 교차검증(cross-validation) 절차를 통해 최적값이 산정된다(Paez et al., 2011).

마지막으로 AICc(corrected Akaike Information Criterion)는 모델의 복잡도와 적합도를 동시에 평가하는 지표로 값이 낮을수록 데이터에 더 적합한 모델로 간주된다. GWR에서는 모델의 자유도 및 지역적 회귀계수의 수가 일반 회귀모형보다 많기 때문에 AICc 보정값을 적용함으로써 과적합 가능성을 통제한다(Huang et al, 2010). 이를 통해 GWR의 모델 성능을 OLS(Ordinary Least Squares)나 SAR(Spatial Autoregressive Model) 등 다른 공간모형과 객관적으로 비교할 수 있다.

- MGWR 주요 해석 지표

MGWR 모델의 해석은 대역폭(Bandwidth), 회귀계수(Coefficient), 신뢰도(p-value)를 중심으로 수행된다(Fotheringham et al., 2017). MGWR은 각 독립변수가 서로 다른 공간적 영향 범위를 가질 수 있음을 전제로 하므로 변수별 회귀계수를 지역 단위에서 해석하는 것이 핵심이다. 이를 통해 어떤 지역에서 특정 변수가 강한 영향력을 발휘하는지를 파악하고 그 공간적 관계를 시각적으로 분석할 수 있다.

대역폭은 변수별로 공간적 영향 범위의 차이를 나타낸다. 대역폭이 작을수록 해당 변수는 국지적인 맥락에서 강하게 작용하며 대역폭이 클수록 보다 광역적인 영향을 갖는 것으로 해석된다. 이를 통해 변수별 공간 스케일의 차이를 명확히 파악할 수 있다. 회귀계수는 공간적으로 불균일하게 나타나며 특정 변수의 영향력이 지역에 따라 달라질 수 있음을 의미한다. 예를 들어 어떤 변수는 특정 지역에서는 양의 영향을 다른 지역에서는 음의 영향을 미칠 수 있다. 이러한 공간적 이질성은 MGWR의 주요 해석 요소이다.

p-value는 회귀계수의 통계적 유의성을 평가하는 지표로 지역별로 각 변수의 영향이 통계적으로 유의한지를 판단하는 데 사용된다. 일반적으로 p-value가 0.05 이하인 경우, 해당 변수의 효과가 통계적으로 의미 있다고 해석한다. MGWR의 또 다른 주요 해석 요소는 결과의 지리적 시각화이다. 회귀계수, 대역폭, p-value 등을 지도 위에 시각화함으로써 변수의 공간적 패턴을 직관적으로 이해할 수 있으며 이는 지역 맞춤형 112신고 대응 전략 수립에 유용한 근거 자료로 활용된다.

- MGWR 모델 예측 성능 평가

MGWR 모형의 예측 성능은 결정계수(R^2)와 AIC(Akaike Information Criterion)를 통해 평가된다. 결정계수는 모델이 종속 변수의 분산을 얼마나 설명하는지를 나타내는 지표로 값이 1에 가까울수록 모델의 설명력이 높다는 것을 의미한다(Wooldridge, 2016). 그러나 R^2 는 과적합 여부를 판단하기 어렵기 때문에 다른 성능 지표와 병행하여 해석하는 것이 바람직하다. AIC는 모델의 적합도와 복잡도를 동시에 고려하는 기준으로 값이 작을수록 좋은 모델로 간주된다(Qin et al., 2024). AIC는 변수 수가 많은 모델에 패널티를 부과하여 과적합을 방지하는 역할을 하며 다양한 모델 간 비교에서 유용하게 활용된다(Fotheringham et al., 2017).

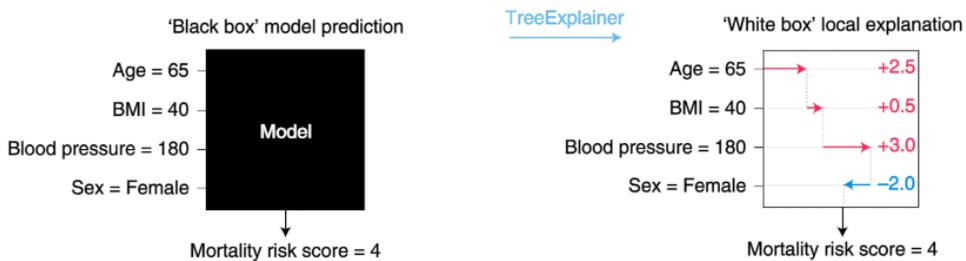
■ 변수 중요도 및 상호작용 해석

기계학습 예측 모델을 해석하는 방법론으로는 SHAP, PDP, LIME 등이 있다. 본 연구에서는 SHAP(Shapley Additive exPlanations)을 활용해 예측모형을 해석하였다. SHAP은 개별 예측부터 전역적 패턴까지 일관된 척도로 변수의 기여도를 정량화함으로써 복잡한 비선형 모

델에서도 투명성과 설명가능성을 높여준다.

- SHAP (Shapley Additive exPlanations)

SHAP은 게임 이론에 기반한 해석 기법으로 각 변수의 예측 기여도를 정량적으로 평가한다 (Lundberg & Lee, 2017). 이는 협력 게임에서의 기여도 산정 방식인 Shapley Value를 머신러닝 모델에 적용한 것으로 변수 조합별 영향력을 계산하여 각 변수의 공정한 기여도를 산정할 수 있다. 전통적인 변수 중요도(feature importance)는 모델 전체 차원에서 변수가 얼마나 중요한지에만 순위화하는 데 그친다면 SHAP은 개별 관측값 수준에서 “어떤 변수가 예측값에 얼마나 어떤 방향으로 작용했는지”를 정밀하게 정량화할 수 있다.



[그림 4-10] SHAP 개념 설명 및 예시

출처: Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), p57

SHAP의 주요 장점은 첫째, 개별 예측값에 대한 해석이 가능하다는 점이다. 즉, 각 관측값에서 어떤 변수가 얼마만큼 기여했는지를 시각화할 수 있다. 둘째, SHAP 값은 양의 방향 또는 음의 방향으로 나타나므로 특정 변수가 예측값을 증가시켰는지 또는 감소시켰는지를 명확히 알 수 있다. 셋째, 복잡한 비선형 모델에 대해서도 명확한 해석이 가능하다는 점에서 모델의 투명성과 설명 가능성을 높이는 데 기여한다.

특히 SHAP global importance(전역 중요도)는 보통 평균 절대 SHAP(mean |SHAP|) 막대로 제시하는데 이는 “해당 특성이 전체 표본에서 예측을 얼마나 강하게 흔들었는가”를 순위로 보여준다. 다만 이 지표는 방향성을 반영하지 않으므로 “증가시키는가/감소시키는가”는 SHAP summary plot (요약 플롯)으로 함께 봐야 한다. 요약 플롯에서 x축은 SHAP 값(오른쪽일수록 예측 증가, 왼쪽일수록 예측 감소), y축은 변수이며 점의 색은 그 관측치에서의 변수 값 크기를 의미한다(통상 빨강=값 큼, 파랑=값 작음). 예컨대 빨간 점들이 오른쪽에 몰리면 “그 변수의 큰 값이 예측을 증가시키는 경향”을 뜻한다. 반대로 파란 점들이 오른쪽에 모이면 “값이 작은 구간에서 예측 증가가 나타난다”는 비선형 또는 상호작용 신호로 해석할 수 있다.

점들의 퍼짐이 넓을수록(수평 분산이 클수록) 공간·집단 간 효과의 이질성이 크다는 뜻이며 한쪽으로 긴 꼬리가 생기면 국지적 임계값이 존재할 가능성을 시사한다.

특성별 작동 구간과 임계점, 상호작용 등은 SHAP dependence plot(디펜던스 플롯)으로 정밀하게 확인할 수 있다. 이 플롯은 x축에 해당 변수의 원값, y축에 그 관측치의 SHAP 값을 두어 “값이 커질수록 예측에 얼마나 더(혹은 덜) 영향을 주는가”의 기울기를 직접 보여준다. 곡선이 완만에서 급해지는 구간은 임계값(threshold) 후보가 된다. 이러한 분석 방법을 통해 “어떤 변수가 얼마나 중요한지(크기)”, “어느 방향으로 작동하는지(부호)”, “어떤 값의 범위에서 특히 강한지(임계·비선형)” 등의 패턴을 확인할 수 있다. 본 연구에서는 이러한 원칙에 따라 전역 중요도로 변수 순위를 확정하고 요약/디펜던스 플롯 등을 통해 변수의 방향성, 임계 구간, 상호작용을 단계적으로 점검하였다.

2. 분석 결과

1) 탐색적 공간 데이터 분석

① 변수 다중공선성 검사

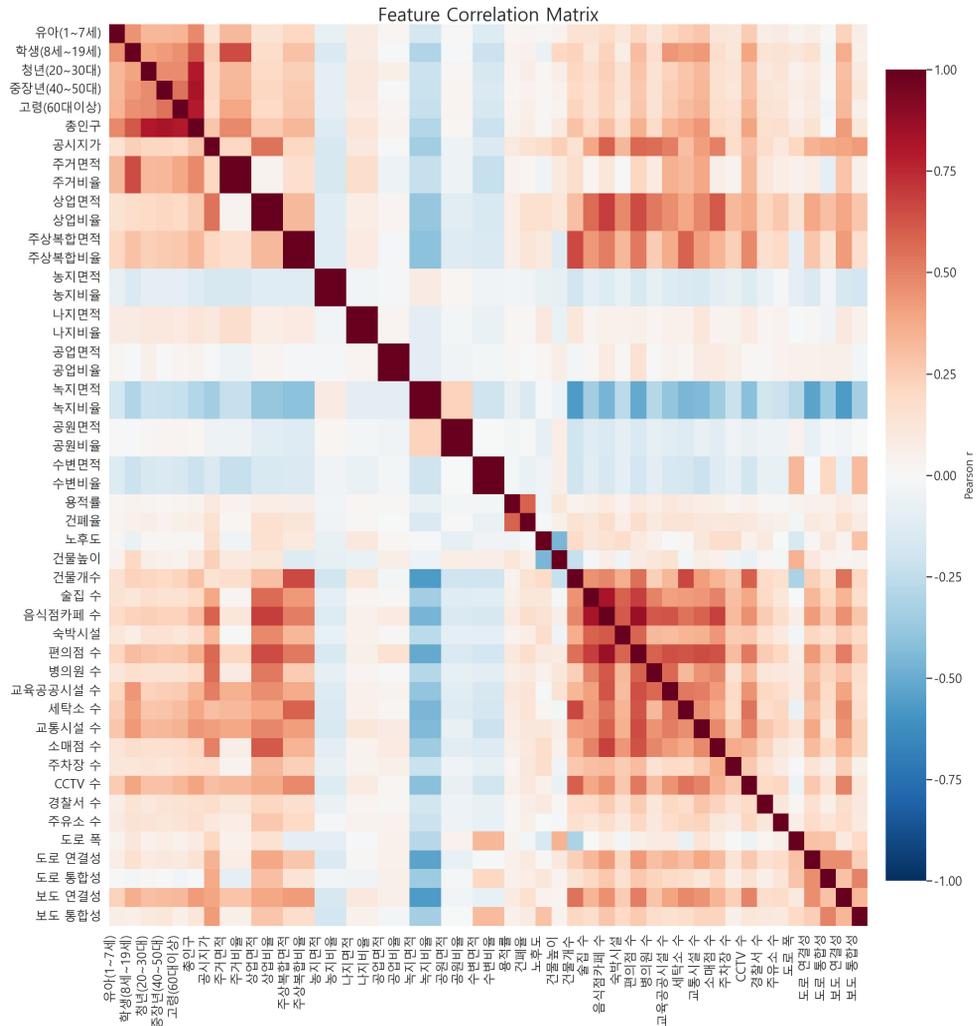
■ 상관관계 분석

효과적인 예측모형을 구축하기 위해서는 입력 변수 간 중복성과 다중공선성(multicollinearity)을 사전에 제거하는 것이 필수적이다. 변수 간 높은 상관관계는 모형의 안정성과 해석력을 저하시킬 수 있으며 특히 회귀계수의 신뢰도를 떨어뜨려 정책적 시사점 도출에도 혼선을 줄 수 있다. 이에 따라 본 연구에서는 먼저 피어슨 상관관계수(Pearson correlation coefficient)를 활용한 상관관계 분석을 실시하였다.

그 결과, 면적 기반 변수와 그에 상응하는 비율 변수들 간의 상관관계가 거의 1에 수렴하는 것으로 나타났다. 예를 들어, 나지면적과 나지비율, 공업면적과 공업비율은 모두 상관관계수 $r = 1.000$ 으로 나타나 동일한 정보를 이중으로 포함하고 있는 것으로 판단하였다. 따라서 절대적 물리량인 면적 변수는 제외하고 공간 단위 간 상대 비교가 가능한 격자 면적 대비 비율 변수를 유지하여 공간 단위별 상대적 특성을 보다 정밀하게 반영하도록 하였다.

또한, 인구 관련 변수에서도 유사한 패턴이 발견되었다. 총인구 변수는 청년 인구(20~30대), 중장년(40~50대), 고령층(60대 이상) 등 개별 연령대 변수들과 상관관계수가 0.79 이상으로 높게 나타났으며 다중공선성 문제를 야기할 수 있어 총인구는 최종 모형에서 제외되었으며 연령별 인구는 개별적으로 독립변수로 투입되었다.

한편, 편의점 수, 음식점·카페 수, 술집 수는 상호 상관성이 높은 변수들이었음에도 불구하고(예: 편의점 수 vs 음식점·카페 수: $r = 0.865$), 이들 변수는 실제 도시 상업환경과 치안 수요를 세분화하여 반영할 수 있는 특성이 있으므로 해석적 재량에 따라 유지하였다.



[그림 4-11] 피어슨 상관계수 값 시각화

출처: 연구진 작성

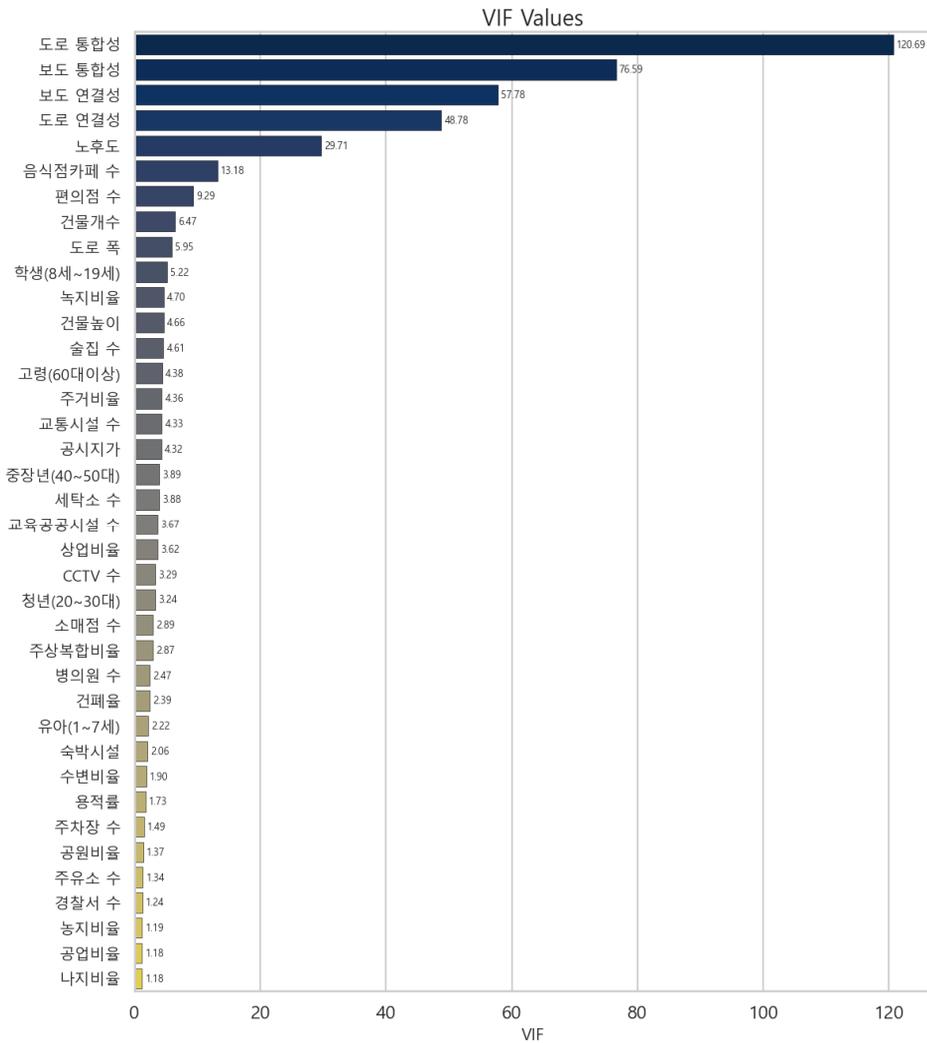
[표 4-6] 상관관계 0.7 이상 변수쌍 목록

var1	var2	corr
나지면적	나지비율	1.000
공업면적	공업비율	1.000
상업면적	상업비율	1.000
주상복합면적	주상복합비율	1.000
수변면적	수변비율	1.000
공원면적	공원비율	1.000
주거면적	주거비율	1.000

녹지면적	녹지비율	1.000
농지면적	농지비율	1.000
편의점 수	음식점 카페 수	0.865
중장년(40~50대)	총인구	0.825
술집 수	음식점 카페 수	0.825
청년(20~30대)	총인구	0.802
고령(60대이상)	총인구	0.797

출처: 연구진 작성

■ 분산팽창지수 분석



[그림 4-12] 분산팽창지수(VIF) 값 시각화

출처: 연구진 작성

상관관계 분석 이후에도 남아 있을 수 있는 다중공선성을 보다 엄밀하게 진단하기 위해 분산 팽창지수(Variance Inflation Factor, VIF)를 계산하였다. 일반적으로 VIF 값이 10 이상이면 다중공선성이 우려되는 것으로 간주되며 15 이상일 경우에는 모델 성능과 해석 신뢰성을 해칠 가능성이 매우 높다. 본 연구에서는 이러한 기준에 따라 $VIF \geq 15$ 를 변수 제거 기준으로 설정하였다. $VIF \geq 15$ 를 변수 제거 기준으로 설정하였다. 분석 결과, 다음 다섯 개의 변수가 해당 기준을 초과하여 최종 예측모형에서 제외되었다.

- 도로 통합성 ($VIF \approx 120.7$)
- 보도 통합성 ($VIF \approx 76.6$)
- 보도 연결성 ($VIF \approx 57.8$)
- 도로 연결성 ($VIF \approx 48.8$)
- 노후도 ($VIF \approx 29.8$)

이들 변수 또한 중요한 지표이나 상관관계 및 VIF 결과에서 다른 지표들과 설명력이 중복된다는 점이 확인되었다. 따라서 모델 단순화 및 해석 용이성을 고려하여 제외하였다.

② 전역적 공간자기상관 진단

■ 공간자기상관 진단의 필요성

공간 데이터 분석에서 가장 핵심적인 출발점은 공간적 자기상관성(spatial autocorrelation)의 존재 여부를 확인하는 일이다. 공간적 자기상관성이란 어떤 변수의 값이 인접한 지리적 위치에서도 유사하거나 비슷한 값을 보이는 경향을 의미한다. 예를 들어 특정 지역에서 112신고 건수가 높다면 이웃 지역에서도 비슷하게 높은 신고 건수를 보일 가능성이 있다는 의미이다. 이러한 공간 자기상관성을 무시하고 분석할 경우, 전통적 회귀모형의 기본 가정 중 하나인 오차의 독립성 가정이 위배될 수 있으며 이는 분석 결과의 신뢰성을 떨어뜨릴 수 있다.

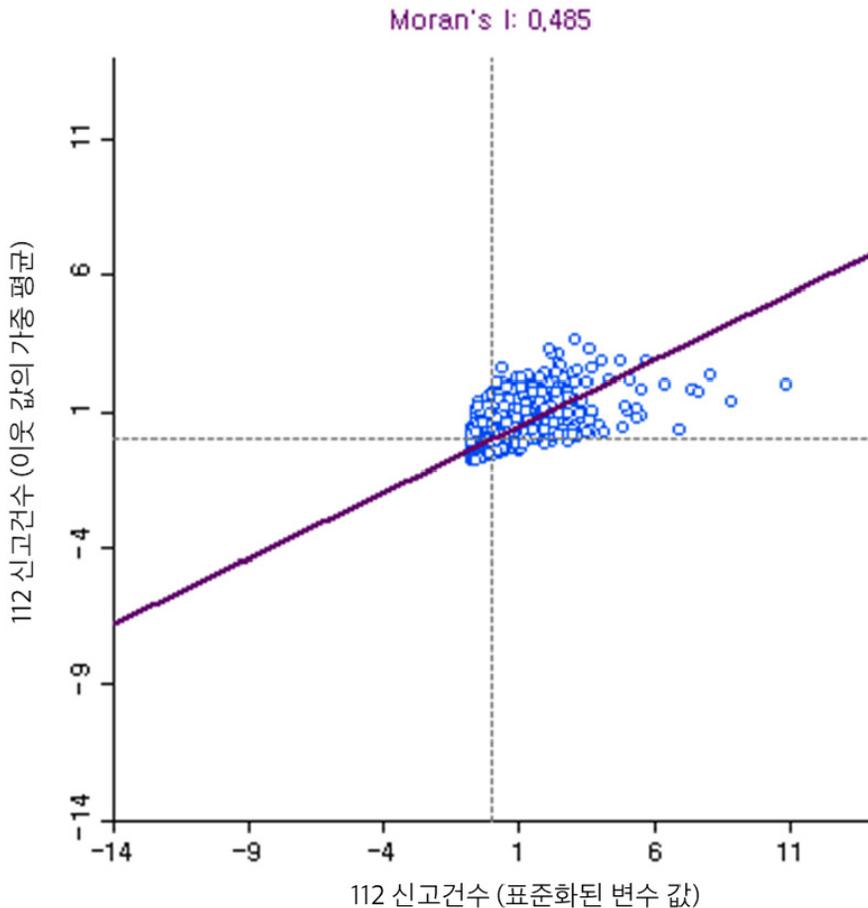
특히 112신고 발생 건수와 같은 사회적 지표는 지역 간 상호작용과 연쇄적인 영향을 통해 발생하는 경우가 많기 때문에 단순한 평균이나 분포 분석만으로는 현상을 완전히 설명할 수 없다. 이러한 공간적 상관 구조를 무시하고 분석을 수행할 경우, 회귀 분석의 기본 가정 중 하나인 독립성 가정(independence assumption)이 깨지며 이는 통계적 추론의 정확성을 떨어뜨릴 수 있다. 이에 따라 공간회귀모형(GWR, MGWR 등)을 적용하기 전 공간적 자기상관성 진단은 모형 적용의 사전 요건으로 매우 중요하다.

■ Global Moran's I 분석 및 해석

공간적 자기상관성을 정량적으로 측정하는 대표적인 지표로는 Moran's I (모란의 아이) 지수가 있다. 이는 전역(global) 분석 지표로 연구 대상 전체 지역의 평균적인 자기상관 정도를 한 숫자로 요약해준다. Moran's I는 -1에서 +1 사이의 값을 가지며 해석은 다음과 같다:

- +1에 가까울수록: 유사한 값끼리 공간적으로 모여 있음 (강한 양의 자기상관)
- 0에 가까울수록: 무작위 분포 (공간적 상관 없음)
- -1에 가까울수록: 이질적인 값끼리 이웃 (강한 음의 자기상관)

본 연구에서는 종속변수인 112신고 건수를 대상으로 Moran's I 분석을 실시하였으며 공간 가중치 행렬(W)은 Queen contiguity 방식을 사용하였다. 공간 가중치 행렬 W의 설정 방식은 분석 결과에 중요한 영향을 미치는데 본 연구에서 활용한 Queen contiguity 방식은 하나



[그림 4-13] Moran's I 산점도 시각화

출처: 연구진 작성

의 격자가 다른 격자와 꼭짓점 또는 변을 공유하면 인접한 것으로 간주하는 방법으로 근접 지역 간 영향력을 민감하게 반영할 수 있는 장점이 있다.

분석 결과, Moran's I 값은 0.485로 나타났으며 이는 종속변수인 112신고 건수가 지역 간 유의미한 양의 공간 자기상관성을 가지고 있음을 나타낸다. 즉, 112신고가 많이 발생하는 지역은 인접한 다른 지역에서도 112신고가 많이 발생하는 경향을 보이고 반대로 112신고가 적은 지역은 인접 지역도 112신고가 낮은 경향을 보인다고 해석할 수 있다. 이러한 결과는 아래 산점도(Moran's I scatter plot)를 통해 시각적으로 확인할 수 있다.

Moran's I 산점도에서 X축은 표준화된 112신고 건수, Y축은 공간적으로 인접한 지역의 평균 값(lagged value), 대각선 방향의 추세선은 전체적인 자기상관의 방향성을 나타낸다. 특히 분석 결과에서 다수의 격자가 1사분면(높은 값과 높은 이웃값)에 집중되어 있으며 이는 112신고 발생이 지리적으로 군집화되어 있다고 해석할 수 있다. 즉, 고위험 지역이 주변과 함께 클러스터를 형성하고 있으며 이는 공간적으로 정책적 개입의 타겟 지역이 존재함을 시사한다.

Moran's I 분석을 통해 확인된 양(+)의 공간적 자기상관은 112신고 건수가 공간적으로 독립적이지 않으며 지역 간 연관 구조가 존재함을 명확히 보여준다. 이는 단순 OLS 회귀모형이 아닌, 공간적 이질성을 반영할 수 있는 공간 회귀모형(GWR, MGWR 등)을 추가 적용할 필요가 있다는 통계적 근거가 된다.

특히, 신고 클러스터가 존재하는 지역은 지역사회 내 범죄 인지 감수성이 높거나 실제 사건이 잦은 공간일 가능성이 높으므로 공간적으로 우선 개입이 필요한 지역으로 판단할 수 있다. 따라서 국지적 자기상관 분석인 LISA(Local Indicators of Spatial Association)분석을 추가적으로 실시하여 Hot Spot 및 Cold Spot을 좀 더 정밀하게 구분하고 정책 개입의 공간적 우선 순위를 도출할 수 있을 것이다.

③ 국지적 공간자기상관 진단

■ LISA 군집 유형별 의미

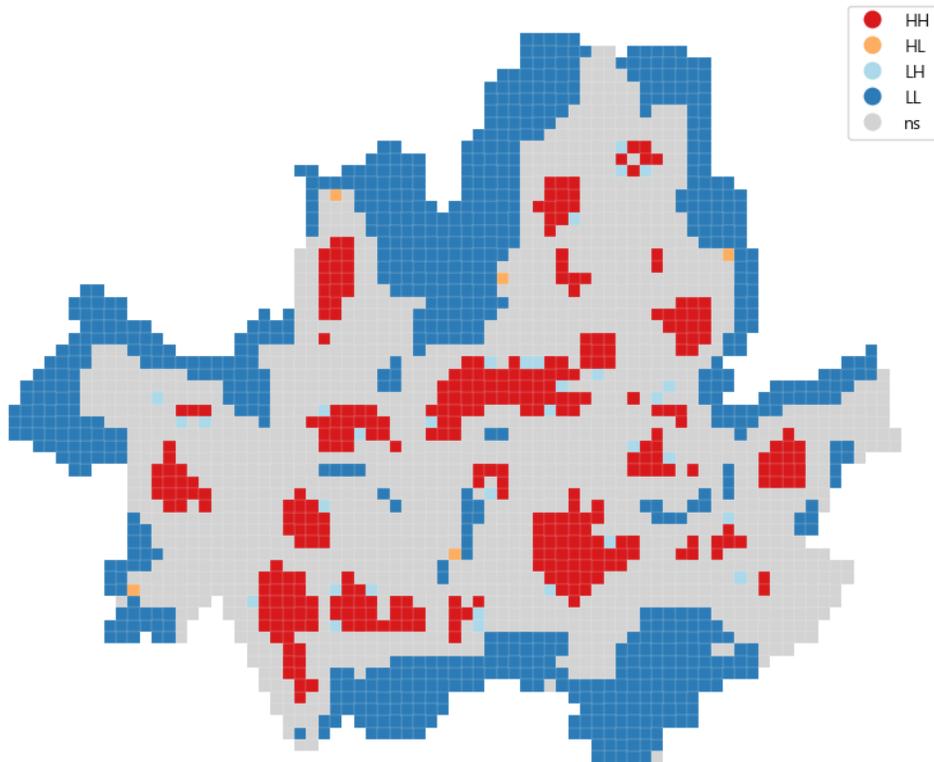
112신고는 단순한 사건 발생뿐 아니라 지역 사회의 불안정성과 사회적 위험을 반영하는 지표이다. 본 분석은 서울시 25개 자치구를 대상으로 112신고 데이터에 대해 공간 자기상관 분석(LISA: Local Indicators of Spatial Association)을 수행한 결과를 요약한 것이다. LISA 클러스터는 특정 지역의 신고 밀집도와 인접 지역 간의 공간적 상관관계를 기반으로 네 가지 유형(HH, LL, HL, LH)으로 분류되며 이를 통해 신고가 집중된 지역, 소외된 지역 등을 탐지할 수 있다.

[표 4-7] LISA 군집 유형별 의미

유형	정의 및 의미
HH (High-High)	높은 112신고 건수를 가진 격자가 인접 격자들과 함께 높은 신고 건수를 보이는 지역 (핫스팟)
LL (Low-Low)	낮은 112신고 건수를 가진 격자가 인접 격자들과 함께 낮은 신고 건수를 보이는 지역 (콜드스팟)
HL (High-Low)	높은 112신고 건수를 가진 격자가 인접 지역들과는 달리 혼자 높은 값을 보이는 지역
LH (Low-High)	낮은 112신고 건수를 보이지만 인접 격자들이 높은 신고 건수를 가진 지역

출처: 연구진 작성

■ 서울시 행정구별 군집 유형별 격자 수 분석

LISA Cluster Map of total_co_1 ($p < 0.05$)

[그림 4-14] LISA 국지적 군집 시각화

출처: 연구진 작성

• HH(핫스팟) 지역의 분포

HH 그리드는 서울시 전역에 고르게 분포하고 있으나 특히 도심 및 남부권역에 상대적으로 밀집되어 있다. 분석 결과, HH 그룹이 가장 많이 분포한 행정구는 강남구(44건)이며 이어서

중구(33건), 관악구(33건), 영등포구(27건), 마포구(21건), 강서구(20건), 서초구(19건), 구로구(18건), 성동구(17건), 강동구(16건), 강북구(16건) 순으로 나타났다. 대부분의 행정구에서 HH 그룹이 일정 수준 이상 존재하였으며 이는 서울시 전역에서 고밀도 국지적 군집이 폭넓게 형성되어 있음을 의미한다. HH 그룹이 20건 이상 나타난 지역은 총 8개 행정구로 확인되었으며 이는 112신고 집중 구역이 일부 행정구에 집중되어 있음을 보여준다.

- LL 지역의 분포

지도에서 LL 유형은 서울 북부와 외곽 지역을 중심으로 광범위하게 분포하고 있으며 가장 많이 분포한 행정구는 서초구(122건)이며 이어서 강서구(102건), 노원구(74건), 은평구(73건), 강북구(67건), 관악구(62건), 도봉구(37건), 강동구(37건), 성북구(40건), 마포구(31건), 구로구(29건), 종로구(69건) 등으로 나타났다. 특히 서초구와 강서구는 LL 유형이 100건 이상으로 압도적으로 높게 나타났으며 다수의 행정구에서 LL 유형이 HH 유형보다 우세하게 분포하고 있었다.

- HL과 LH지역의 희소성

HL(High-Low)과 LH(Low-High) 유형은 주변과 상반되는 값을 갖는 공간적 이상(outlier) 유형으로 희소하게 나타났다. HL 그룹과 LH 그룹 모두 전체적으로 매우 적은 수의 격자에서만 관찰되었으며 대부분의 행정구에서 0~3건 수준에 불과하였다. 특히 HL 유형은 1건 이하인 지역이 다수이며 LH 유형도 5건을 초과하는 경우는 존재하지 않았다. 이는 서울시의 대부분 지역이 주변 격자와 유사한 속성을 공유하고 있어 공간적 이질성이 급격히 발생하는 지역은 극히 제한적임을 보여준다.

- 행정구별 HH(핫스팟)/LL(콜드스팟) 분포 비교

LISA 분석 결과를 기반으로 서울시 25개 행정구의 공간 자기상관 유형을 비교한 결과, HH(High-High)와 LL(Low-Low) 유형 간의 상대적 우세 정도에 따라 행정구의 공간 구조는 크게 세 가지로 구분된다. 각 유형은 고밀도 또는 저밀도 공간 군집의 집중 양상과 균형 수준에 따라 상이한 특성을 나타낸다.

첫째, HH 유형이 우세한 지역은 고값 격자들이 군집 형태로 집중된 고밀도 공간 구조가 뚜렷한 행정구에 해당한다. 중구는 33건(HH), 1건(LL)으로 HH가 절대적으로 우세하였으며 영등포구(27건, 4건), 금천구(19건, 10건), 성동구(17건, 0건), 동대문구(17건, 0건), 용산구(10건, 4건) 등에서도 HH 유형이 LL 유형을 상회하였다. 이들 지역은 행정구 내에서 고밀도 군집이 중심적인 공간 구조를 형성하고 있으며 저밀도 공간은 거의 나타나지 않거나 제한적으로 분포하였다.

둘째, LL 유형이 우세한 지역은 값이 적은 격자가 인접하여 저밀도 군집을 형성하는 공간 구

조가 주를 이루는 행정구로 분류된다. 서초구는 122건(LL), 19건(HH)으로 LL이 약 6배 이상 많았으며 강서구(102건, 20건), 노원구(74건, 7건), 은평구(73건, 20건), 강북구(67건, 16건), 성북구(40건, 6건), 도봉구(37건, 5건), 강동구(37건, 16건), 마포구(31건, 21건), 구로구(29건, 18건) 등에서도 LL이 우세하게 나타났다. 이러한 지역은 행정구 내 다수의 격자에서 저밀도 공간 특성이 우세하며 고밀도 군집은 일부에 국한되어 있다.

셋째, HH와 LL 유형이 유사한 수준으로 공존하는 지역은 고밀도와 저밀도 공간 구조가 동일한 행정구 내에서 병존하는 양상을 보인다. 강남구는 44건(HH), 43건(LL)으로 양자의 비중이 거의 동일하였으며 관악구는 33건, 62건, 종로구는 23건, 69건으로 두 유형이 모두 일정 수준 이상 존재하였다. 이처럼 HH와 LL이 공존하는 지역은 하나의 뚜렷한 공간 특성으로 정의되기 어렵고 내부 구역 간 밀도 격차와 신고양상의 이질성이 뚜렷하다는 점에서 세분화된 공간 분석 및 맞춤형 대응 전략 수립의 필요성이 크다고 할 수 있다.

[표 4-8] 행정구별 LISA 군집 유형 수

행정구	HH	HL	LH	LL
강남구	44	0	1	43
강동구	16	0	0	37
강북구	16	0	1	67
강서구	20	0	3	102
관악구	33	0	3	62
광진구	15	0	4	16
구로구	18	1	1	29
금천구	19	0	2	10
노원구	7	1	3	74
도봉구	5	0	3	37
동대문구	17	0	5	0
동작구	10	1	2	1
마포구	21	0	3	31
서대문구	12	0	0	12
서초구	19	0	2	122
성동구	17	0	4	0
성북구	6	1	2	40
송파구	13	0	2	9
양천구	12	0	0	8
영등포구	27	0	1	4
용산구	10	1	2	4
은평구	20	0	0	73
종로구	23	0	3	69
중구	33	0	2	1
중랑구	17	1	0	22

출처: 연구진 작성

2) 물리환경과 112신고 발생 관계 분석

① 핫스팟·콜드스팟 지역 간 물리환경 차이 분석

■ 핫스팟·콜드스팟 지역의 물리환경 차이 분석 목적

도시 내에서 112신고 건수는 단순한 사건 발생 빈도만을 의미하는 것이 아니라 지역 사회의 불안정성, 사회적 위험, 물리적 환경 요인의 복합적인 영향을 반영한다. 특히 112신고가 특정 지역에서 공간적으로 군집되어 나타난다는 점은 신고가 단순히 무작위적으로 발생하는 것이 아니라 특정 공간 구조 및 환경 요소와의 관계 속에서 형성됨을 시사한다. 이에 따라 본 연구는 공간통계기법인 LISA(Local Indicators of Spatial Association)를 활용하여 도시 공간을 High-High(HH) 및 Low-Low(LL) 군집으로 분류하고 이들 간의 물리적 환경 변수 값의 차이를 통계적으로 검증하는 것을 목적으로 한다.

보다 구체적으로는 HH로 분류된 격자(grid)들은 112신고 건수가 높고 주변 지역도 동일하게 높아 도시 내 핵심 위험 밀집지로 해석되며 LL 격자는 그 반대로 신고 건수와 주변 지역 모두 낮은 상대적으로 안전하거나 치안 수요가 낮은 지역으로 간주된다. 이 두 공간 군집 간에 어떠한 물리적 특성 차이가 존재하는지를 실증적으로 검토함으로써 도시 계획 및 치안 정책 수립에 실질적 인사이트를 제공하고자 한다.

■ 정규성 검정과 비모수 통계 적용

분석에 앞서 각 변수에 대한 정규성 검정을 수행하기 위해 Shapiro-Wilk 검정을 실시하였다. 이 검정은 각 변수의 값들이 정규분포를 따르는지를 평가하는 것이며 그 결과, 대부분의 변수에서 p-value가 0.05 이하로 나타나 정규성 가정이 위배됨을 확인할 수 있었다. 이에 따라, 두 그룹 간 평균값의 차이 검정에 비모수 통계 기법인 Mann-Whitney U 검정을 적용하였다. Mann-Whitney U 검정은 각 그룹 간의 분포 차이를 평가할 수 있는 검정으로 특히 데이터가 정규분포를 따르지 않거나 이상값의 영향을 받을 수 있는 상황에서 유용하게 활용된다. 본 연구에서는 각 물리적 환경 변수에 대해 HH 그룹과 LL 그룹 간의 분포 차이를 독립적으로 검정하였다.

■ 핫스팟·콜드스팟 지역의 물리환경 특성 차이 검정 결과

앞서 LISA 분석을 통해 112신고 건수가 공간적으로 높은 밀도를 이루는 HH(High-High) 군집 지역과 신고 건수가 낮고 주변 지역도 유사한 LL(Low-Low) 군집 지역을 분류한 후, 두 군집 간의 물리적 환경 특성의 유의미한 차이 여부를 확인하였다.

총 48개 변수에 대해 HH 격자와 LL 격자 간의 분포 차이를 독립적으로 검정한 결과, 모든 변수에서 p-value가 0.01 미만으로 유의 수준을 충족하였으며 이는 HH와 LL 지역 간 물리적 환경 요소가 전반적으로 통계적 유의한 차이를 보이고 있음을 의미한다.

- 인구통계

인구구성 측면에서 HH 지역은 유아(0-7세), 학생(8-19세), 청년(20-30대), 중년(40-50대), 고령(60세 이상) 모두 중앙값이 0인 LL 지역과 비교해 현저히 높은 값을 나타냈다. 이는 HH 지역이 전 연령대에서 인구가 밀집된 공간임을 보여주며 LL 지역은 상대적으로 인구가 거의 존재하지 않는 저밀도 지역임을 시사한다.

- 건물정보

건물정보 측면에서도 뚜렷한 차이가 확인되었다. HH 지역은 공시지가, 용적률, 건폐율, 건물 개수에서 모두 LL 지역보다 높은 중앙값을 보였다. 특히 공시지가와 용적률에서의 차이는 HH 지역이 고밀도의 개발이 집중된 환경임을 잘 보여준다. 반면 LL 지역은 건축 밀도가 낮아 상대적으로 저개발적 성격을 띠는 것으로 나타났다.

- 토지이용

토지이용 변수 역시 양 집단 간 차이가 극명했다. HH 지역은 주거비율, 상업비율, 주상복합비율에서 모두 LL 지역보다 압도적으로 높은 값을 보였으며 반대로 LL 지역은 녹지비율이 압도적으로 높았다. 이는 HH 지역이 복합적 토지이용 구조를 갖춘 도심적 공간인 반면, LL 지역은 녹지가 중심이 되는 저밀도 공간임을 보여준다.

- 생활편의 및 상권 시설

편의 인프라 변수에서도 HH 지역이 우세하였다. 음식점 수, 소매점 수, 병의원 수, 도시관·교육 시설 수, 세탁소 수, 대중교통(버스·지하철) 수, 주차장 수, CCTV 수, 술집 수 모두 HH 지역에서만 시설이 존재하였다. 주유소 수, 주차장 수, 경찰서 수는 그 수가 충분하지 않아 중앙값이 0이었다. 이러한 결과는 HH 지역이 생활편의와 안전 관련 시설이 집적된 공간임을 보여준다.

- 도로 네트워크

도로폭 변수는 HH 지역이 8.330, LL 지역이 3.569로 HH 지역이 더 큰 수치를 보였다. 이는 HH 지역이 LL 지역에 비해 상대적으로 더 넓은 도로가 구축되어있음을 시사한다.

[표 4-9] 변수별 HH/LL 그룹별 평균 차이 분석 결과

카테고리	변수	Statistic	p-value	HH 중앙값	LL 중앙값
인구통계	유아(0-7세)	327633	1.37×10^{-103} ***	120.000	0.000
	학생(8-19세)	351058	1.09×10^{-137} ***	384.500	0.000
	청년(20-30대)	328511	8.15×10^{-98} ***	1454.000	0.000
	중년(40-50대)	324418	4.84×10^{-91} ***	1477.000	0.000
	고령(60세 이상)	318065	1.78×10^{-81} ***	1233.500	0.000
건물정보	공시지가	387938	8.35×10^{-186} ***	4767249.000	211872.000
	용적률	382574	7.66×10^{-187} ***	109.885	0.000
	건폐율	371435	2.66×10^{-165} ***	35.023	0.000
	건물높이	330438	1.40×10^{-96} ***	8.266	0.000
	건물개수	390476	1.34×10^{-197} ***	541.500	0.000
토지이용	주거비율	334434	1.49×10^{-103} ***	23.291	0.000
	상업비율	385068	1.17×10^{-216} ***	15.144	0.000
	주상복합비율	374057	2.97×10^{-208} ***	7.568	0.000
	농지비율	122159	1.44×10^{-48} ***	0.000	0.000
	나지비율	346147	2.68×10^{-132} ***	0.615	0.000
	공업비율	207643	4.39×10^{-06} ***	0.000	0.000
	녹지비율	50852	1.83×10^{-108} ***	4.600	61.238
	공원비율	293186	2.60×10^{-55} ***	1.201	0.000
생활편의 및 상권 시설	수변비율	166240	2.98×10^{-08} ***	0.000	0.000
	술집 수	385233	1.17×10^{-242} ***	14.000	0.000
	음식점·카페 수	392308	2.55×10^{-223} ***	110.500	0.000
	숙박시설 수	372204	4.45×10^{-223} ***	6.000	0.000
	편의점 수	390747	3.62×10^{-240} ***	9.000	0.000
	병·의원시설 수	382835	4.87×10^{-234} ***	9.000	0.000
	공공·교육시설 수	388287	2.36×10^{-215} ***	35.000	0.000
	세탁소 수	379808	9.77×10^{-231} ***	3.500	0.000
	소매시설 수	392830	1.19×10^{-220} ***	82.500	0.000
	주유소 수	249582	1.71×10^{-43} ***	0.000	0.000
	대중교통(버스·지하철) 수	375834	7.01×10^{-188} ***	8.000	0.000
	주차장 수	278698	7.16×10^{-86} ***	0.000	0.000
	CCTV 수	388291	1.26×10^{-203} ***	30.000	0.000
경찰서 수	232893	8.89×10^{-35} ***	0.000	0.000	
도로 네트워크	도로 폭	299151	1.10×10^{-54} ***	8.330	3.569

*** p < 0.01

출처: 연구진 작성

② 핫스팟과 배경지역의 물리환경 차이 분석

핫스팟 지역(112신고가 집중된 High-High 군집)이 해당 자치구 전체 및 서울시 전체와 비교하여 물리적·사회적 환경에서 어떠한 특성을 보이는지를 분석하였다. 비교 대상은 (1) 서울시 평균, (2) 각 자치구 전체 평균, (3) 자치구 내 핫스팟 지역 평균으로 구분하였다. 이를 통해 핫스팟 지역의 특성이 자치구와 서울 내에서 어떤 점에서 구별되는지를 파악하고자 하였다.

[표4-10]에서는 셀의 색상 진하기를 통해 각 변수값의 크고 작음을 시각적으로 표현하였다. 또한 통계적 유의미성을 표시하기 위해 표에 기호를 병기하였다. *은 핫스팟 지역 값이 배경 지역인 자치구 및 서울시 중 하나와 유의한 차이를 보일 때, ***는 자치구 및 서울시 모두와 유의한 차이를 보일 때를 의미한다. 이를 통해 각 구별 핫스팟 지역이 어떤 변수에서, 어느 비교 그룹과 구별되는지를 직관적으로 확인할 수 있도록 구성하였다.

■ 인구통계

• 유아(1~7세)

서울시 전체 평균은 격자당 약 85.9명으로, 유아 인구는 전 연령대 중 가장 낮은 밀도를 보인다. 자치구 전체 평균은 서울시 평균보다 다소 높은 경향을 보이지만 여전히 격자당 40명대에서 150명대 수준으로 분포해 있다. 반면, 자치구별 핫스팟 지역은 대부분의 자치구에서 130명 이상, 일부 지역에서는 200명을 초과하여 상대적으로 높은 유아 밀집 현상이 특정 격자군에서 관찰된다.

다수의 자치구에서 핫스팟 지역은 자치구 전체 평균 및 서울시 평균과 모두 통계적으로 유의한 차이를 보였으며 이는 핫스팟 지역이 유아 인구가 특별히 집중된 공간이라는 점을 입증한다. 반면 일부 지역에서는 자치구 평균과 핫스팟 간 차이가 뚜렷하지 않아 비유의한 경우도 존재하였다.

서초구, 성동구, 중랑구, 강서구, 강북구 등은 핫스팟 지역의 유아 인구가 자치구 평균의 두 배 이상으로 공간적 밀도가 매우 높은 지역으로 나타났다. 특히 중랑구는 자치구 전체 평균이 146.1명인데 반해 핫스팟 지역은 210.3명으로 가장 높은 수준을 기록하였다.

• 학생(8세~19세)

서울시 전체 평균은 격자당 207.1명으로 유아보다 현저히 높은 수준이며 전 연령대 중에서도 높은 인구 밀도를 보인다. 자치구 전체 평균은 지역 간 편차가 매우 크게 나타나며 100명 이하(강북구, 강서구, 도봉구 등)에서 400명 이상(동대문구, 동작구 등)에 이르기까지 다양하게 분포한다. 특히 자치구별 핫스팟 지역은 대부분의 구에서 400~600명 수준을 보이며 일부 지역에서는 600명을 초과하는 경우도 확인된다.

학생 연령대는 대부분의 자치구에서 핫스팟 지역이 자치구 평균 및 서울시 평균과 모두 유의한 차이를 보였다. 이는 학령기 인구가 특정 격자에 집중되어 있다는 점을 통계적으로도 뒷받침해준다.

강서구, 노원구, 서초구, 성북구, 양천구 등은 구 전체 평균이 낮음에도 불구하고 핫스팟 지역의 학생 수가 500명 이상으로 매우 높아 내부적으로 인구 분포의 이질성이 큰 자치구에 해당한다. 예를 들어 노원구는 자치구 평균 116.3명, 핫스팟 지역 589.6명으로 약 5배 차이를 보여 가장 극단적인 편차를 보이는 지역 중 하나로 분석된다.

- 청년(20~30대)

서울시 전체 격자 평균은 835.7명이다. 자치구 전체 평균과 핫스팟 지역 간에는 여전히 큰 격차가 존재하며 특히 핫스팟 지역은 대부분 1,300~2,000명 이상으로 나타난다.

청년 인구 또한 대부분의 자치구에서 통계적으로 유의한 차이를 나타냈다. 특히 서울시 평균과 비교하여 핫스팟 지역의 인구 밀도는 일관되게 높은 수준을 보이며 자치구 평균과의 차이도 유의한 경우가 많았다. 이는 청년 인구가 특정 지역에 밀집되어 있음을 의미한다.

강북구, 노원구, 도봉구 등은 자치구 전체 평균이 600명 미만임에도 불구하고 핫스팟 지역에서 1,900명을 초과하는 수치를 보여 공간 집중성이 매우 높은 지역으로 분석된다.

- 중장년(40~50대)

서울시 전체 평균은 격자당 약 887.3명으로 청년 인구와 유사한 수준을 보이나 자치구 평균은 서울시보다 높은 경우가 많았다. 다수의 자치구에서 자치구 전체 평균은 1,000명을 상회하며 일부 지역에서는 1,700명 이상으로 나타난다. 반면 자치구별 핫스팟 지역에서는 대부분 1,400명에서 1,800명 이상으로 구 전체 평균보다 한층 높은 인구 밀도를 보인다. 이는 핫스팟 지역이 중장년 인구가 밀집된 생활권과 상당 부분 중첩됨을 시사한다.

핫스팟 지역은 거의 모든 자치구에서 서울시 전체 및 자치구 전체와 비교하여 통계적으로 유의한 차이를 나타냈다. 특히 핫스팟 지역의 평균 수치가 자치구 평균을 현저히 상회하는 경우가 많아 중장년 인구의 공간적 집중도가 높다는 점이 명확하게 드러났다.

동작구와 마포구는 자치구 전체 평균이 각각 1348.0명과 958.3명으로 높은 수준인데 핫스팟 지역에서는 각각 1872.0명, 1503.7명으로 상승폭이 크다. 강북구와 관악구처럼 자치구 전체 평균이 상대적으로 낮은 지역에서도 핫스팟 지역은 1,400명 이상으로 나타났다.

- 고령(60대 이상)

서울시 전체 평균은 약 788.3명으로 유아 및 학생보다 높고 청년·중장년보다는 낮은 수준이다. 자치구 전체 평균은 400명대에서 1,300명대까지 다양하게 분포하며 고령 인구의 지역 간

편차가 큰 특성을 보여준다. 자치구별 핫스팟 지역의 경우 대부분의 자치구에서 1,000명을 초과하며 일부 지역에서는 1,600명 이상으로 고령 인구가 매우 밀집된 격자가 확인된다.

핫스팟 지역은 서울시 평균과의 차이뿐 아니라 자치구 전체 평균과의 차이에서도 높은 통계적 유의성을 보였다. 특히 자치구 전체 평균이 낮은 지역에서도 핫스팟 지역의 고령 인구는 상대적으로 매우 높게 나타나며 이는 핫스팟 지역이 고령 인구가 구조적으로 밀집된 공간임을 시사한다.

강북구, 강서구, 도봉구 등은 자치구 전체 평균이 400~600명대로 낮은 편이나 핫스팟 지역에서의 평균은 1,400명 이상으로 극단적인 편차를 보였다. 중랑구, 중구, 관악구 등도 핫스팟 지역에서의 고령 인구가 1,300명 이상으로 고령 밀집 지역으로 나타났다.

■ 건물정보

• 공시지가

서울시 전체 평균 공시지가는 약 2,846,264.3원/㎡로 자치구 전체 평균은 이보다 편차가 크게 나타난다. 강남구, 서초구, 중구 등은 8백만 원에서 1천만 원을 상회하는 높은 수준을 보이며 기타 지역은 1백만 원~4백만 원 사이로 상대적으로 낮은 수치를 기록하였다. 자치구별 핫스팟 지역의 평균 공시지가는 대부분 자치구 전체 평균을 상회하며 일부 지역에서는 1,000만 원을 초과하기도 한다.

공시지가 항목은 대부분의 자치구에서 핫스팟 지역이 자치구 전체와 서울시 전체보다 유의하게 높은 값을 보여주며 신고 밀집 지역이 부동산 가치와도 일정 부분 연관되어 있음을 시사한다. 특히 강남구, 서초구, 종로구, 중구 등 주요 상업·고가 주거지역은 핫스팟 지역의 공시지가가 서울시 평균의 3배 이상으로 유의미한 공간 경제적 차이를 보였다.

서초구는 자치구 전체 평균이 약 2,936,239원임에도 핫스팟 지역은 약 15,942,105만 원으로 압도적으로 높은 수치를 나타냈다. 중구와 종로구도 1,130만 원 이상으로 핫스팟 지역이 고가 부동산 밀집 지역과 일치함을 보여준다. 반면 도봉구, 강북구 등은 자치구 전체 평균과 핫스팟 지역 간 차이는 있지만 절대값 자체는 비교적 낮아 상대적 이질성은 존재하나 절대적 부동산 가치는 낮은 것으로 판단된다.

• 용적률

서울시 전체 평균 용적률은 51.7%로 자치구 전체 평균은 대체로 이보다 높게 나타났다. 자치구별 핫스팟 지역에서는 이 수치가 더욱 높아지는 경향을 보이며 대부분 90% 이상, 일부 자치구에서는 130%를 넘는 격자들도 확인된다. 이는 핫스팟 지역이 상대적으로 고밀 개발이 이루어진 지역에 위치하고 있음을 시사한다.

용적률은 대부분의 자치구에서 핫스팟 지역과 서울시 평균 간 유의미한 차이를 보이며 자치구 전체 평균과도 차이가 나는 경우가 많다.

강남구(196.5%), 강서구(133.6%), 서초구(189.3%) 등은 핫스팟 지역의 용적률이 매우 높으며 개발 강도가 높은 주거·상업 복합지구일 가능성이 높다. 도봉구와 강북구처럼 자치구 평균이 15~26%대인 곳에서도 핫스팟 지역의 용적률은 90%를 넘겨 공간 내 밀도 차이가 상당함을 보여준다.

- 건폐율

서울시 전체 평균 건폐율은 18.1%이다. 이에 비해 자치구 전체 평균은 다양하게 나타나며 특히 핫스팟 지역에서는 대부분 자치구보다 높은 건폐율을 보여 30% 이상의 지역이 다수 확인된다.

대부분의 자치구에서 핫스팟 지역은 서울시 전체뿐 아니라 자치구 평균과도 뚜렷한 차이를 보인다. 이는 핫스팟 지역이 단위 면적당 건물 배치 밀도가 높고 도시 물리 환경이 집약적으로 구성되어 있음을 의미한다.

서대문구는 자치구 평균이 41.7%로 전체적으로도 높는데 핫스팟 지역에서는 69.3%로 매우 높은 건폐율을 보인다. 종로구는 자치구 평균이 26.9%, 핫스팟 지역은 82.8%로 극단적인 차이를 보여 해당 지역의 공간 밀도가 높은 수준임을 나타낸다.

- 건물 높이

서울시 전체 평균은 5.7층으로 나타나며 자치구 전체 평균은 지역별로 다소 편차가 있으나 대부분 이보다 약간 높은 수준이다. 핫스팟 지역의 경우 대체로 7~10층 수준으로 전체적으로 건물 높이가 높아지는 양상을 보인다.

건물 높이 역시 다수 자치구에서 핫스팟 지역이 유의미하게 높은 수치를 보이며 도시 구조상 밀도뿐 아니라 수직적 확장도 집중된 공간이라는 점을 뒷받침한다.

서초구의 핫스팟 지역은 평균 16.6층으로 전 자치구 중 가장 높은 건물 밀도를 보인다. 강남구, 강서구, 구로구, 송파구 등도 핫스팟 지역의 평균 건물 높이가 9층 이상으로 고층화된 도시 공간 내 신고 밀집지역과 중첩됨을 시사한다.

- 건물 개수

서울시 전체 평균은 197.1개이다. 자치구 전체 및 핫스팟 지역 간 수치는 큰 편차를 보인다. 특히 핫스팟 지역은 대부분의 자치구에서 500개 이상의 건물이 밀집되어 있으며 일부 지역에서는 700개를 초과한다.

핫스팟 지역과 서울시 전체 간에는 거의 모든 자치구에서 유의미한 차이가 나타나며 자치구

전체 평균과도 통계적으로 구별되는 경우가 많다. 이는 핫스팟 지역이 구조적으로 건물 밀도가 높은 지역에 해당함을 나타낸다.

강북구와 도봉구는 자치구 평균이 33.2개와 66.4개로 비교적 낮지만 핫스팟 지역은 각각 787.4개, 527.6개로 극단적인 증가를 보여 같은 자치구 내에서도 물리적 환경의 편차가 매우 크다는 것을 보여준다. 마포구, 성동구, 도봉구 등도 유사한 경향을 보인다.

■ 토지 이용

• 주거 비율

서울시 전체의 평균 주거비율은 13.7%이나 자치구별 전체 평균은 이보다 높은 경향을 보인다. 특히 자치구별 핫스팟 지역의 주거비율은 대부분의 자치구에서 자치구 평균 및 서울시 평균보다 높은 수준을 보이며 20% 이상인 지역이 다수다.

주거비율은 다수 자치구에서 핫스팟 지역과 서울시 전체, 자치구 전체 간 유의미한 차이를 보인다. 핫스팟 지역이 상대적으로 주거 기능이 밀집된 공간임을 시사한다.

노원구(44.0%), 성북구(40.9%), 은평구(35.7%) 등의 핫스팟 지역은 특히 높은 주거비율을 보이며 이는 이들 지역의 신고 밀집 지역이 주거 중심지에 위치함을 나타낸다. 중구(11.1%), 종로구(6.0%) 등은 상대적으로 낮은 주거비율을 보이며 도심 상업지와외의 중첩 가능성을 시사한다.

• 상업비율

서울시 전체 평균은 6.4%이다. 자치구 및 핫스팟 지역에서는 상당한 차이를 보인다. 특히 핫스팟 지역은 전반적으로 상업비율이 자치구 평균보다 높게 나타나는 경향이 있으며 일부 지역은 20%를 초과한다.

상업비율에서도 다수의 자치구에서 핫스팟 지역이 서울시 전체 및 자치구 전체와 유의미한 차이를 보인다. 이는 신고 집중된 지역이 상업 밀집 지역과도 관련이 깊다는 점을 시사한다.

종로구(33.7%), 중구(32.3%), 강남구(25.8%) 등의 핫스팟 지역은 매우 높은 상업비율을 보여 도심 상업지 및 중심업무지구(CBD) 내에 위치함을 시사한다. 서대문구(17.4%), 서초구(20.7%) 등도 비교적 높은 상업비율을 보여 외곽지에서도 지역 중심 상권과의 연관성을 보인다.

• 주상복합비율

서울시 평균은 2.8%이다. 자치구 전체와 핫스팟 지역에서는 지역별 차이가 뚜렷하다. 특히 핫스팟 지역에서는 대부분 자치구에서 평균보다 높은 주상복합비율을 보이며 복합적 기능이 혼재된 공간의 특성을 보여준다.

핫스팟 지역은 다수 자치구에서 자치구 전체보다 주상복합비율이 통계적으로 유의하게 높아 물리적 혼합 환경이 신고 집중에 영향을 줄 수 있음을 시사한다.

강동구(17.6%), 강북구(10.3%), 동대문구(12.4%), 성동구(10.7%) 등은 핫스팟 지역 내 주상복합비율이 두드러지게 높으며 혼합 용도의 복합 개발 환경과 밀접한 관련이 있을 가능성이 높다. 일부 자치구(예: 용산구, 중구)는 상대적으로 낮은 수준을 유지한다.

- 농지비율

서울시 전체 평균 농지비율은 2.8%이다. 자치구 전체 및 핫스팟 지역에서는 지역별로 농지비율에 차이를 보였다. 자치구 전체의 경우 강서구(10.9%)나 서초구(8.4%)처럼 상대적으로 농지비율이 높은 지역도 있었으나 핫스팟 지역에서는 대부분의 자치구에서 농지비율이 0%로 나타났다. 이는 핫스팟 지역이 농지보다는 도시화된 지역에 집중되어 있다는 점을 시사한다.

대부분의 자치구에서 핫스팟 지역과 자치구 전체 및 서울시 평균 간 농지비율 차이는 통계적으로 유의하지 않았다. 즉, 핫스팟 지역의 낮은 농지비율이 일관되게 다른 그룹과 차별화되었다고 보기 어렵다.

- 나지비율

서울시 평균 나지비율은 0.7%로 나타났다. 대부분의 자치구 전체와 핫스팟 지역에서도 이와 유사한 수준을 보인다. 구 전체 기준으로는 2~3% 수준에 도달한 경우도 있으나 대체로 1% 이내에 그친다.

대부분 자치구에서 나지비율에 대한 통계적 유의성은 확보되지 않아 핫스팟 지역이 자치구나 서울시 전체와 구별되는 특성을 보인다고 판단하긴 어렵다.

- 공업비율

서울시 평균 공업비율은 0.7%이다. 자치구 전체 및 핫스팟 지역에서는 자치구별로 큰 차이를 보였다. 특히 구로구, 금천구, 영등포구 등 일부 자치구는 핫스팟 지역에서 공업비율이 10% 이상으로 매우 높게 나타났다.

특정 자치구에서만 공업비율의 유의미한 차이가 확인되었고 대부분의 자치구에서는 차이가 통계적으로 유의하지 않았다. 이에 따라 핫스팟 지역의 공업비율 특성은 일부 자치구에 국한된 경향으로 해석된다.

- 녹지비율

서울시 평균 녹지비율은 39.2%로 높게 유지되나 자치구 전체와 핫스팟 지역은 뚜렷한 차이를 보인다. 구 전체 평균은 지역에 따라 10% 미만에서 70%를 상회하는 곳까지 다양하며 핫스팟 지역은 대부분 15% 이하의 낮은 수치를 보였다.

녹지비율은 전반적으로 핫스팟 지역에서 서울시 및 자치구 평균 대비 유의미하게 낮은 경향이 두드러졌으며 많은 자치구에서 통계적으로도 유의한 차이를 보였다.

특히 강북구(62.9%), 노원구(59.6%), 성북구(70.2%) 등은 자치구 전체 평균이 매우 높은 반면 이들 지역의 핫스팟 지역은 5~15% 수준으로 녹지비율이 크게 낮아 공간적 밀도 및 개발 정도의 차이를 강하게 시사한다.

- 공원비율

서울시 평균 공원비율은 5.7%이다. 자치구 전체에서는 최대 20.1%(강남구)까지 다양하게 나타난다. 반면 핫스팟 지역에서는 대부분 자치구에서 5% 이하의 낮은 공원비율을 보였다.

다수의 자치구에서 공원비율 차이는 통계적으로 유의하지 않았다. 따라서 공원비율이 핫스팟지역을 설명하는 특성이라고 이해하기는 어렵다.

- 수변비율

서울시 평균 수변비율은 3.6%이며 자치구 전체에서는 한강이나 하천을 끼고 있는 지역을 중심으로 10%를 상회하는 사례도 있다. 그러나 핫스팟 지역은 대부분 자치구에서 5% 미만, 일부는 0%로 매우 낮은 비율을 기록하였다.

대부분 자치구에서 수변비율 차이는 통계적으로 유의하지 않아 수변 접근성이 핫스팟 지역의 보편적 특성이라고 보기는 어렵다.

■ 생활편의 및 상권 시설

- 술집 수

서울시 평균은 7.7개이다. 자치구 전체와 핫스팟 지역에서는 큰 차이를 보인다. 자치구 전체 평균은 1~15개 수준에서 다양하게 나타나며 핫스팟 지역에서는 대부분 자치구에서 평균을 크게 상회한다. 특히 20개 이상인 자치구도 다수 존재한다.

대다수 자치구에서 핫스팟 지역은 자치구 전체 및 서울시 평균 대비 유의미하게 높은 값을 보이며 통계적으로도 유의하다. 이는 핫스팟 지역이 주취 관련 112신고가 집중된 상업적·야간 유흥 활동이 활발한 공간임을 시사한다.

구로구, 마포구, 서대문구, 종로구 등은 핫스팟 지역에서 30개 이상의 평균 술집 수를 기록해 타 지역 대비 확연히 높은 수치를 보인다.

- 음식점·카페 수

서울시 평균은 46.1개이며 자치구 전체에서는 지역별 편차가 크다(예: 은평구 7.5 vs 강남구 99.6). 핫스팟 지역에서는 거의 모든 자치구에서 평균보다 높은 수치를 보이며 100개 이상을

초과하는 경우도 상당수 존재한다.

핫스팟 지역은 자치구 전체 및 서울시 평균과 비교하여 통계적으로 유의하게 높은 값을 보이며 음식점 및 카페 밀집이 핫스팟 지역의 주요 공간 특성 중 하나임을 명확히 보여준다.

강남구, 마포구, 서대문구, 종로구, 중구 등은 핫스팟 지역에서 평균 170개 이상으로 특히 밀도가 높아 해당 지역 내 중심 상업지 기능과의 연결성이 강하게 나타난다.

- 숙박시설 수

서울시 평균은 3.4개이다. 자치구 전체에서는 강서구, 마포구, 중구, 용산구 등에서 8개 이상으로 높은 편이다. 핫스팟 지역에서는 이보다 훨씬 높은 10개 이상을 기록하는 자치구가 다수 존재한다.

여러 자치구에서 핫스팟 지역의 숙박시설 수는 자치구 전체 대비 유의미하게 높은 수치를 보이며 통계적으로 유의한 차이를 나타낸다.

마포구와 서대문구는 핫스팟 지역에서 21.3개로 서울시 전체 평균의 약 6배 이상 높은 값을 보여주며 용산구, 종로구, 중구 등도 15개 이상으로 높은 수치를 기록하고 있다.

- 편의점 수

서울시 평균은 3.8개이다. 자치구 전체에서는 1~7개 수준으로 나타난다. 반면 핫스팟 지역에서는 모든 자치구에서 평균을 초과하며 10개 이상인 자치구도 다수 존재한다.

핫스팟 지역은 자치구 전체 및 서울시 평균과 비교했을 때 통계적으로 유의하게 높은 편의점 수를 보였다.

강남구, 구로구, 마포구, 서대문구, 종로구, 중구 등에서 핫스팟 지역의 편의점 수는 12~14개 수준으로 높은 밀집도를 보인다.

- 병·의원시설 수

서울시 평균은 5.6개이며 자치구 전체는 평균 1~10개 수준으로 분포한다. 핫스팟 지역은 거의 대부분 자치구에서 평균보다 높으며 특히 10개 이상을 기록한 자치구도 많다.

핫스팟 지역의 병·의원시설 수는 서울시 및 자치구 전체 평균 대비 통계적으로 유의하게 높은 경우가 다수이다.

특히 강남구와 서초구는 핫스팟 지역에서 40개이상으로 매우 높은 병·의원시설 수를 기록하였으며 강동구, 송파구, 중구 등도 15개 이상으로 높은 수치를 나타낸다.

- 교육·공공시설 수

서울시 평균은 16.7개이다. 자치구 전체에서는 4~54개 수준이며 핫스팟 지역은 대부분 30개 이상으로 뚜렷하게 높다. 특히 일부 자치구의 핫스팟 지역은 평균의 5~6배를 초과한다.

핫스팟 지역의 교육시설 수는 자치구 전체와 비교해 통계적으로 유의미하게 높은 경향을 보이며 이는 교육시설이 밀집된 지역에서 112신고가 집중될 수 있음을 시사한다.

강남구, 마포구, 서초구, 은평구 등은 핫스팟 지역에서 50개 이상 교육시설이 밀집되어 있는 것으로 나타난다. 이들은 학원가나 학교가 집중된 지역과도 연결된다.

- 세탁소 수

서울시 평균은 1.4개이며, 자치구 전체는 0.2~2.8개 수준, 핫스팟 지역은 대부분 3~6개로 높은 편이다. 핫스팟 지역은 평균의 2~4배에 달하는 경우가 흔하다.

핫스팟 지역의 세탁소 수는 전반적으로 서울시 및 자치구 평균에 비해 유의하게 높은 경향을 보이며 주거 밀도와 생활 편의시설의 밀접한 관계를 반영한다.

강남구, 강동구, 강서구, 관악구, 광진구, 금천구, 서초구, 성동구, 송파구, 양천구, 은평구, 중랑구 등은 핫스팟 지역에서 평균 4개 이상 세탁소가 밀집되어 있어 생활 서비스의 집중 양상을 보여준다.

- 소매시설 수

서울시 평균은 41.3개이다. 자치구 전체는 5~100개 내외, 핫스팟 지역은 대부분 80개 이상, 많게는 260개를 초과하는 경우도 있다.

핫스팟 지역의 소매시설 수는 서울시 및 자치구 평균 대비 압도적으로 높고 통계적으로도 유의미하다. 이는 상업 활동과 소비 인프라가 112신고 집중과 연관되어 있음을 시사한다.

강남구, 구로구, 동대문구, 송파구, 종로구, 중구 등은 핫스팟 지역에서 120개 이상 소매점이 위치해 높은 상업 밀집도를 보인다. 특히 중구, 종로구는 260개 이상으로 타 지역 대비 독보적으로 높다.

- 주유소 수

서울시 전체 평균은 0.2개로 낮은 수준이다. 자치구 전체에서도 0.1~0.8 사이의 수치로 비교적 작고 일정한 범위를 보인다. 핫스팟 지역에서는 이보다 약간 높은 수치를 보이는 경우가 있으나 전반적인 분포 자체는 제한적이다. 일부 자치구 핫스팟 지역에서는 1.0을 넘는 경우가 있으나 대부분은 0.5~0.8 수준에 머문다.

통계적 유의성은 제한적이다. 핫스팟 지역이 자치구 전체 평균보다 다소 높은 값을 보이는 경

향이 있으나 서울시 평균과의 차이는 뚜렷하지 않으며 지역 간 유의한 분산도 적다. 전체적으로 볼 때 변수의 규모와 분포가 협소하여 유의미한 비교를 끌어내기 어렵다.

- (대중)교통시설 수

서울시 평균은 3.7개, 자치구 전체는 1~7개 수준이며 핫스팟 지역은 대부분 7개 이상, 많게는 12개 이상까지 나타난다.

핫스팟 지역은 대부분 자치구 대비 교통시설 수가 유의미하게 높으며 통계적으로도 유의하다. 이는 핫스팟 지역이 교통 접근성이 높은 중심지에 위치하는 경향이 있음을 보여준다.

구로구, 중구, 종로구, 마포구, 서초구 등은 핫스팟 지역에서 10개 이상의 교통시설이 밀집되어 있으며 전형적인 도심 중심지역의 특징을 반영한다.

- 주차장 수

서울시 전체 평균은 0.3개로 매우 낮은 수치를 유지하고 있으며 이는 자치구 전체에서도 유사하게 나타난다. 대부분 자치구 평균이 0.0~0.7 사이에 분포하고 있어 전반적으로 주차장 수 자체가 극히 적은 수준임을 보여준다. 반면 핫스팟 지역에서는 1.0 이상을 넘는 지역이 나타나며 특히 몇몇 자치구에서는 2.0~3.0 수준까지 확인된다. 이는 핫스팟 지역이 차량 기반 접근성과 관련된 수요가 상대적으로 높은 공간임을 시사한다.

그러나 통계적 유의성 면에서는 제한적이다. 값의 스케일 자체가 작고 지역 간 분산도 작아 유의한 차이로 해석되기엔 다소 부족하다. 일부 지역에서 구 전체와 유의한 차이를 보이는 하나 서울시 전체 평균과의 차이에서는 뚜렷한 경계가 확인되지 않는다.

- CCTV 수

서울시 전체 평균은 14.5대이다. 자치구 평균은 4.8~72.2까지 넓은 범위를 보인다. 전반적으로 자치구별 차이가 큰 편이며 특히 핫스팟 지역에서는 대체로 자치구 전체 평균을 상회하는 수준으로 나타난다. 다수의 핫스팟 지역에서 30대를 넘는 수치가 확인되며 85대를 초과하는 지역도 존재한다. 이는 핫스팟 지역이 CCTV 설치가 집중적으로 이루어진 공간, 즉 치안이 상대적으로 강화된 공간일 가능성을 보여준다.

통계적으로도 유의한 차이가 일관되게 나타난다. 대부분의 자치구에서 핫스팟 지역은 자치구 평균뿐 아니라 서울시 평균과도 뚜렷한 차이를 보이며 이 차이는 통계적으로 유의미한 것으로 판단된다.

강북구, 구로구, 서초구, 양천구, 영등포구 등 도심부 자치구에서는 핫스팟 지역의 CCTV 수가 50~80대를 넘는 경우가 다수 확인되며 그 외 강서구, 노원구, 은평구 등도 상대적으로 높은 수치를 보인다. 이러한 경향은 핫스팟 지역의 특성을 대표적으로 드러내는 지표 중 하나로

해석될 수 있다.

• 경찰서 수

서울시 전체 평균은 0.1개로 매우 낮고 자치구 평균 역시 대부분 0.0~0.2 수준으로 큰 차이를 보이지 않는다. 핫스팟 지역에서도 수치는 소폭 증가하는 경향이 있으나 전체적으로 0.0~0.5 수준으로 여전히 낮은 값을 유지하고 있다. 변수의 분포 자체가 전반적으로 정적인 양상을 보인다.

통계적 유의성은 매우 낮다. 대부분의 자치구에서 핫스팟 지역과 자치구 전체, 또는 서울시 평균 간의 차이가 명확하지 않으며 통계적으로도 유의미한 결과가 거의 도출되지 않았다. 이는 변수의 스케일 자체가 너무 작아 비교 분석에 한계가 있음을 의미한다.

■ 도로 네트워크

• 도로 폭

서울시 전체 평균 도로 폭은 7.1m로 자치구 간 큰 차이는 없지만 자치구 평균 및 핫스팟 지역 평균에서는 뚜렷한 지역 간 편차가 나타난다. 자치구 전체 평균은 6.8m(성북구)에서 13.3m(노원구)까지 다양하게 분포하며 핫스팟 지역의 평균 도로 폭은 대부분 7m 이상으로 비교적 넓은 편이다. 특히 핫스팟 지역에서의 도로 폭은 서울시 평균은 물론 해당 자치구 평균보다도 높은 경우가 다수 관찰된다.

대부분의 자치구 핫스팟 지역과 통계적으로 유의하지 않았다. 도로 폭은 핫스팟 지역의 구조적 특성을 설명하는 지표로 일반화하기에는 제한적일 수 있다. 다만, 노원구, 서대문구, 종로구 등 일부 자치구에서는 자치구 전체 대비 핫스팟 지역의 도로 폭이 통계적으로 유의하게 더 넓은 것으로 나타나 지역에 따라서는 일정한 경향이 존재함을 시사한다.

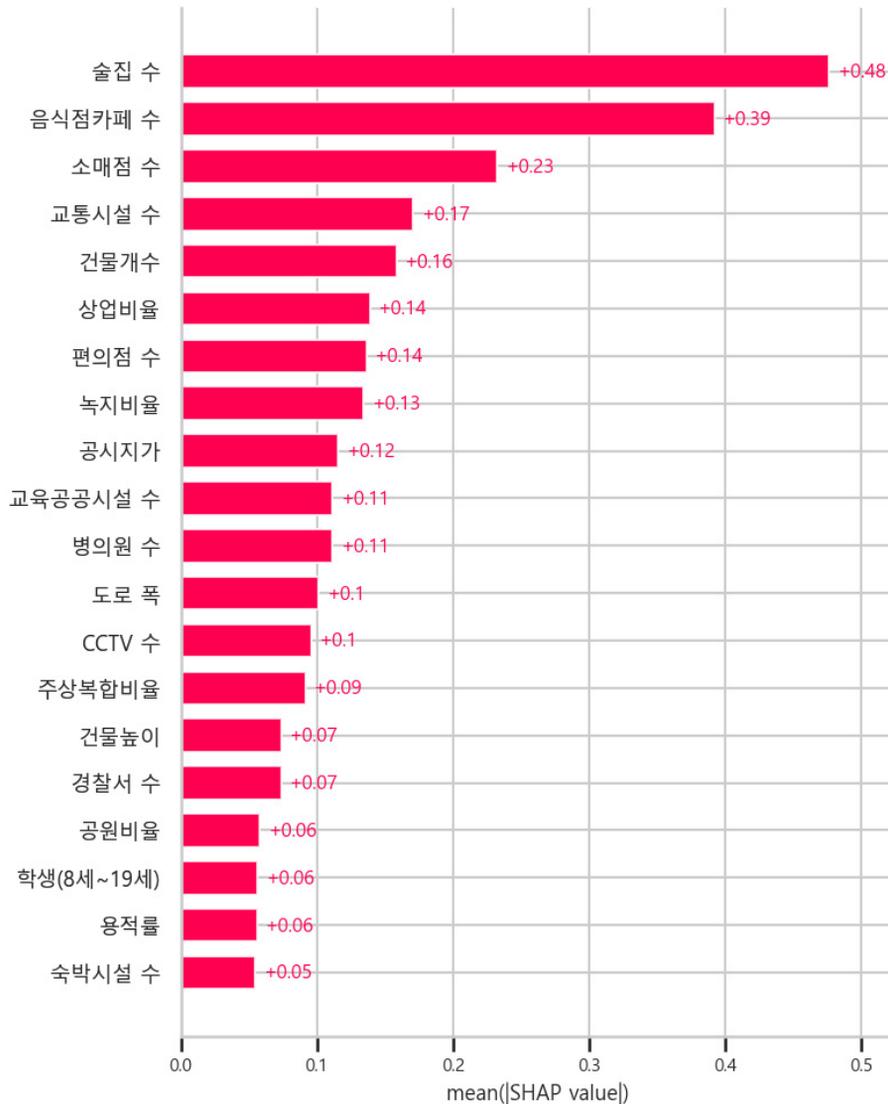
[표 4-10] 서울시-행정구-핫스팟(HH) 지역의 변수별 평균 차이 분석 결과

변수	그룹	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구	마포구	서대문구	서초구	성동구	성북구	송파구	양천구	영등포구	용산구	은평구	종로구	중구	중랑구
		유아	서울	**	*	**	**	*	*	*	*	**	*	**	*	*	**	*	**	*		*	**		*	*
	구	*	*	**	**	*				**	*	*	*	*	*	*	*	*		*	*		*	*	*	*
	HH	*	**	**	**	**	*	*	*	**	*	*	*	*	**	*	*	*	*	*	*	*	**	**	*	**
학생	서울	**	*	**	**	*	*	**	**	**	**	**	*	*	*	**	*	*	*	**	**	*	*	*	**	*
	구	*	*	**	**	*	*	*	*	**	**	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	HH	*	**	**	**	**	**	*	*	**	**	*	*	**	*	*	*	*	*	*	*	*	**	*	*	**

③ 기계학습 예측 분석 결과

■ 예측모델 성능 평가

본 연구에서는 최종 선정된 독립변수들을 활용하여 부스팅(Boosting) 계열 머신러닝 모델인 XGBoost와 LightGBM을 각각 학습시켜 예측 성능을 비교하였다. 두 모델 모두 다수의 결정 트리(Decision Tree)를 순차적으로 학습시키며 오차를 점진적으로 보정하는 방식으로 작동하지만 트리 성장 방식, 학습 속도, 메모리 효율성 등에서 차이가 존재한다.



[그림 4-15] SHAP 기반 전역적 변수 중요도 시각화

출처: 연구진 작성

실제 분석 결과, 두 모델 모두 본 연구 데이터셋에 대해 높은 예측력을 보였으나 XGBoost가 전반적으로 더 우수한 성능을 나타냈다. XGBoost의 검증 데이터 기준 결정계수(Test R²)는 0.85로 LightGBM의 0.83보다 높았다. 평균 제곱근 오차(Test RMSE) 또한 XGBoost가 1.14로 LightGBM의 1.18에 비해 더 낮은 값을 기록하였다.

따라서 본 연구에서는 XGBoost를 최종 예측모형으로 채택하였으며 이후 변수 중요도 해석 (Feature Importance), 변수별 기여도 및 영향 방향을 정량적으로 해석할 수 있는 SHAP (Shapley Additive exPlanations) 기법 적용 등 모든 후속 분석은 XGBoost 결과를 기준으로 수행하였다.

■ 전역적 변수 중요도 해석

XGBoost 예측모형의 학습 결과를 바탕으로 본 연구에서는 각 독립변수가 112신고 건수 예측값에 미치는 영향을 해석하기 위해 SHAP (SHapley Additive exPlanations) 값을 활용한 변수 중요도 분석을 수행하였다. SHAP 분석은 게임이론 기반의 공정한 기여도 분배 방식으로 복잡한 머신러닝 모델의 예측 결과에 대해 각 변수의 영향력을 정량적으로 설명할 수 있는 강력한 도구이다. 이를 통해 블랙박스 모델의 해석력을 높이고 변수 간 상대적 중요도를 시각화할 수 있다.

전역적 변수 중요도(Global Feature Importance)는 평균 SHAP 값(mean(|SHAP value|))을 기준으로 정렬한 것으로 모형 전반에서 예측값 변동에 얼마나 기여했는지를 상대적으로 보여준다. 가장 중요한 변수는 '술집 수'로 예측값에 가장 큰 영향을 미쳤으며 이는 유흥시설 밀도가 높은 지역에서 신고 건수가 많아지는 현상과 연결될 수 있다. 그 뒤를 이어 '음식점 및 카페 수', '소매시설 수', '교통시설 수'가 높은 중요도를 보였다. 이는 사람들의 활동량이 많고 유동인구가 집중되는 지역적 특성이 신고 발생과 밀접한 관련이 있음을 시사한다. 또한, '건물 개수', '상업비율', '편의점 수' 또한 상위권에 위치하여 밀집된 도시환경 및 상업 시설의 밀도 역시 중요한 설명 변수임을 보여준다.

■ 개별 예측값에 대한 변수 영향 해석

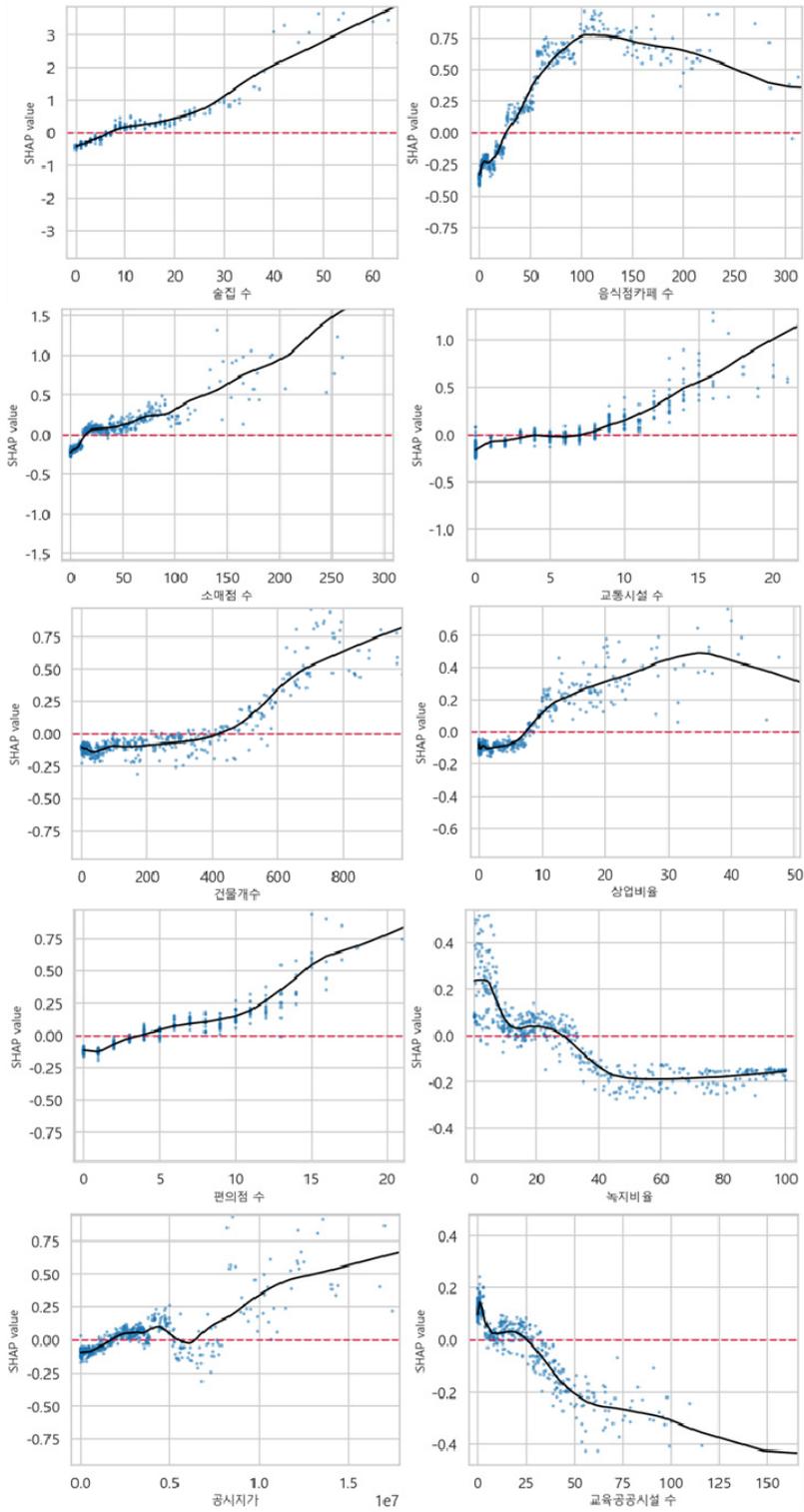
SHAP 요약 플롯(SHAP Summary Plot)은 변수의 개별 관측치 수준에서의 기여도를 시각화한다. 가로축의 SHAP 값은 해당 관측치에서 그 변수가 예측을 증가(+) 또는 감소(-)시키는 방향과 크기를 뜻하며 세로축에는 변수명이 배열된다. 각 점은 하나의 격자 단위 데이터 포인트를 나타내며 점의 색상은 해당 지점에서 해당 변수의 값이 높은지(진한 분홍색) 혹은 낮은지(파란색)를 나타낸다. 해석의 핵심은 점의 색상(값의 고/저)과 분포 방향(+/-) 간의 결합이다. 예를 들어 술집 수, 음식점 카페 수와 같은 변수들의 경우, 분홍색 점들이 전반적으로 오른쪽

■ 상위 10개 변수에 대한 SHAP 의존도 플롯 해석

SHAP 의존도 플롯(SHAP dependence plot)은 각 변수값(가로축)이 예측에 기여한 SHAP 값(세로축)과 어떤 관계를 갖는지 관측치 단위로 보여준다. 점 하나가 하나의 격자 관측치를 의미하며 세로축이 0보다 크면 해당 변수값이 112신고전수를 늘리는 방향으로 0보다 작으면 줄이는 방향으로 기여했음을 뜻한다. 검은 실선은 관측치 산점도의 국지적 추세(스무딩 곡선)을 나타낸다. 이 플롯은 평균화된 효과만 보여주는 요약 플롯과 달리 값이 같은 구간에서도 관측치별 기여가 얼마나 이질적인지까지 확인할 수 있다는 장점이 있다.

본 결과에서 술집 수는 전 구간에서 단조 증가하는 패턴을 보이며 약 10개를 기점으로 기여가 음(-)에서 양(+)로 전환되는 양상이 확인된다. 음식점·카페 수는 약 25개 전후에서 음(-)에서 양(+)으로 전환한 뒤 약 100개까지 급증하고 이후 증가 폭이 완만해지는 효과가 관찰된다. 소매시설 수는 약 10개를 기점으로 음(-)에서 양(+)으로 전환 후, 지속적으로 증가하는 패턴을 보이며 교통시설 수는 낮은 값의 구간 영향이 미미하다가 약 7개를 기점으로부터 양(+)의 기여가 빠르게 확대되는 양상을 보인다. 건물 개수는 S자형의 분포를 보이며 저밀 구간에서는 평탄하지만 약 400개를 기점으로 급증 구간이 형성되는 패턴을 보인다. 상업 비율은 약 8%를 기준으로 양(+) 연관성이 뚜렷해지고 35% 전후에서 증가세가 둔화되는 패턴이 관찰되며 편의점 수는 약 4개를 기점으로 음(-)에서 양(+)으로 전환되는 양상을 보인다. 녹지 비율은 음(-)의 비선형성 관계가 포착되며 약 20%를 기점으로 양(+)에서 음(-)으로 전환되는 패턴을 보인다. 이는 녹지가 20% 이상일 때 112신고 감소와 연관이 있다고 해석될 수 있다. 공시지가는 저가 구간에서는 영향이 작거나 약한 음(-)의 관계를 보이지만 중·고가로 갈수록 양(+)으로 바뀌는 패턴을 보인다. 교육·공공시설 수는 전반적으로 112신고와 음(-)의 관계를 보이며 약 25개를 기점으로 양(+)에서 음(-)으로 전환되는 패턴을 보인다.

정리하면 SHAP를 활용한 분석은 기존 머신러닝 예측 결과를 전역적 차원과 국지적 차원에서 동시에 설명하여 각 변수의 방향(±)과 기여 크기를 정량화하고 요약·의존도 플롯 등을 통해 임계값 및 비선형 구조 등을 포착할 수 있어 정책 임계값 후보와 표적화된 개입 우선순위 도출에 유용하다. 다만 SHAP는 예측모형 내 조건부 기여도를 제시할 뿐 인과 효과가 아니며 공간·시간 자기상관을 SHAP 자체가 교정하지 못하므로 혼합 패턴이 보이는 변수는 시간대·공간 이질성(GWR/MGWR) 및 상호작용을 함께 검토하고 제시된 임계값에 대해서는 현장 점검·과일터를 통한 실증 검증으로 보완하는 것이 바람직하다.



[그림 4-17] 변수별 SHAP 의존도 플롯
출처: 연구진 작성

3) 공간회귀 기반 분석

① 공간회귀 모델별 성능 비교

앞에서 SHAP 전역 중요도로 선별한 상위 10개 변수를 대상으로 MLR(Multiple Linear Regression), GWR(Geographically Weighted Regression), MGWR(Multiscale Geographically Weighted Regression) 모델을 구축하고 그 성능을 비교하였다. MLR은 전 지역에 하나의 계수를 적용하는 전역(baseline) 모형으로 결과 해석이 단순하고 기준선 마련에 적합하며 GWR은 위치별로 계수가 달라질 수 있다는 공간적 이질성을 반영해 지역별 패턴을 포착한다. MGWR은 여기에 더해 변수마다 작동하는 공간 범위가 다를 수 있음을 고려하여 변수별로 서로 다른 최적 대역폭을 추정함으로써 어떤 요인은 미시 단위에서 어떤 요인은 광역단위에서 영향이 두드러지는지를 동시에 밝혀낸다. 모든 모형은 동일 표본·동일 변수로 학습하여 Log-likelihood, AICc, R^2 , Adj. R^2 , RSS 등 공통 지표로 일관되게 평가했으며 GWR/MGWR에서는 Adaptive bisquare 커널과 AICc 기반 대역폭 최적화를 적용해 과도한 지역화나 과적합을 억제했다.⁴⁾

■ 전체시간 모델 성능평가

전체시간 자료에서 GWR과 MGWR 모두 MLR 대비 공간적 이질성을 반영하며 성능이 개선되었다. OLS는 $R^2=0.819$ 을 보인 반면, GWR은 $R^2=0.929$ 로 적합도가 크게 향상되었고 MGWR은 $R^2=0.906$ 으로 GWR보다는 약간 낮지만 AICc가 1946.930으로 가장 낮아(GWR 1970.126, OLS 2995.152) 모형 간 기준에서 MGWR의 균형적 우수성이 확인되었다.

[표 4-11] MLR, GWR, MGWR에 대한 모델 평가(전체시간)

Diagnostic Information	MLR	GWR	MGWR
Log-likelihood	-1485.516	-254.812	-627.152
AICc	2995.152	1970.126	1946.930
R^2	0.819	0.929	0.906
Adj. R^2	0.818	0.909	0.893
RSS	476.438	187.141	248.286

출처: 연구진 작성

4) ① Log-likelihood는 통계 모델이 주어진 데이터를 얼마나 잘 설명하는지를 나타내는 지표이며, 모델 적합도로 0에 가까울수록(값이 클수록) 우수함. ② AICc는 모델의 복잡도와 적합도를 동시에 고려하여 모델 간의 우수성을 비교하는 지표로, 모델 적합도와 복잡도의 균형 지표로 값이 작을수록 우수함. ③ R^2 은 모델이 종속변수의 전체 변동성을 설명하는 비율(0~1)을 의미하며, 1에 가까울수록 모델의 설명력(예측력)이 높음. ④ Adjusted R^2 은 독립변수의 개수가 늘어남에 따라 결정계수가 인위적으로 높아지는 단점을 보완한 값으로, 과적합 위험을 반영하여 모델의 설명력을 더 정확하게 평가함. ⑤ RSS는 모델에 의한 예측값과 실제 관측값의 차이(잔차)를 제곱하여 합산한 결과이며, 모델이 설명하지 못한 변동량을 나타냄

■ 낮시간 모델 성능평가

낮시간 모델에서도 공간 모형의 이점이 뚜렷했다. MLR의 설명력은 $R^2=0.785$ 로 기준선을 제공했고 GWR은 $R^2=0.909$ 로 적합도가 크게 향상되어 공간적 이질성 반영의 효과를 보여준다. 한편 MGWR은 $R^2=0.883$ 으로 GWR보다 다소 낮지만 $AICc=2565.129$ 로 세 모형 중 가장 작은 값을 기록(OLS 3452.861, GWR 2612.644)하여 예측력-복잡도 균형에서 최적임을 시사한다.

[표 4-12] MLR, GWR, MGWR에 대한 모델 평가(낮시간)

Diagnostic Information	MLR	GWR	MGWR
Log-likelihood	-1714.371	-576.071	-914.036
AICc	3452.861	2612.644	2565.129
R^2	0.785	0.909	0.883
Adj. R^2	0.784	0.884	0.866
RSS	566.857	238.840	308.712

출처: 연구진 작성

■ 밤시간 모델 성능평가

밤시간 모델에서는 공간 효과가 더 강하게 나타났다. OLS의 $R^2=0.821$ 에 비해 GWR은 $R^2=0.938$ 로 가장 높은 설명력을 보이며 야간의 지역별 패턴을 정밀하게 포착한다. MGWR은 $R^2=0.915$ 로 GWR보다는 낮지만 $AICc=1722.243$ 으로 세 모형 중 최저(OLS 2973.842, GWR 1804.949)를 기록하여 모형 간 비교 기준에서는 MGWR이 가장 우수하다. 종합하면 야간은 낮시간 대비 전반 지표가 더 개선되었으며 이는 야간에 특히 강한 공간적 이질성과 변수별 작동 스케일 차이가 존재함을 의미하며 이후 해석·시각화에서 제시할 시간대별 개입 단위 설정의 근거가 된다.

[표 4-13] MLR, GWR, MGWR에 대한 모델 평가(밤시간)

Diagnostic Information	MLR	GWR	MGWR
Log-likelihood	-1474.861	-80.887	-497.784
AICc	2973.842	1804.949	1722.243
R^2	0.821	0.938	0.915
Adj. R^2	0.820	0.918	0.903
RSS	472.598	163.989	225.056

출처: 연구진 작성

② 공간회귀 모델별 분석 결과

■ MLR 모델 해석

• MLR 모델-전체시간

MLR 전체시간 모델에서는 대부분의 변수가 112신고건수에 유의미하게 기여했다($p < 0.001$). 특히 술집 수(Est.=0.533, $t=34.06$)가 가장 강한 양(+) 효과를 보이며 그다음 편의점 수(Est.=0.201, $t=10.62$), 교통시설 수(Est.=0.132, $t=11.52$), 소매시설 수(Est.=0.110, $t=9.04$), 건물개수(Est.=0.106, $t=9.35$), 공시지가(Est.=0.096, $t=8.34$), 상업비율(Est.=0.063, $t=5.08$) 순으로 112신고 건수에 대해 양(+) 기여가 확인된다. 반대로 녹지비율(Est.=-0.040, $t=-3.94$)은 억제(음) 효과가 유의하며 음식점·카페 수(Est.=-0.144, $t=-5.88$) 또한 음(-)으로 추정되었다. 교육 공공시설 수(Est.=0.027, $p=0.024$) 변수는 효과 크기는 작지만 112신고건수와 양(+)의 연관성을 보인다($p < 0.05$).

[표 4-14] MLR 매개변수에 대한 요약통계(전체시간)

Variable	Est.	SE	t-value	p-value
Intercept	0.000	0.008	0.000	1.000
술집 수	0.533	0.016	34.060	<0.001
음식점 카페 수	-0.144	0.024	-5.877	<0.001
소매시설 수	0.110	0.012	9.039	<0.001
교통시설 수	0.132	0.011	11.522	<0.001
건물개수	0.106	0.011	9.349	<0.001
상업비율	0.063	0.012	5.079	<0.001
편의점 수	0.201	0.019	10.622	<0.001
녹지비율	-0.040	0.010	-3.936	<0.001
공시지가	0.096	0.011	8.335	<0.001
교육 공공시설 수	0.027	0.012	2.258	0.024

출처: 연구진 작성

• MLR 모델-낮시간

MLR 낮시간 모델에서 112신고건수와 유의한 양(+)의 관계를 나타내는 변수는 술집 수(Est.=0.352, $t=20.630$), 소매시설 수(Est.=0.150, $t=11.361$), 교통시설 수(Est.=0.170, $t=13.535$), 건물개수(Est.=0.108, $t=8.741$), 상업비율(Est.=0.106, $t=7.873$), 편의점 수(Est.=0.178, $t=8.639$), 공시지가(Est.=0.147, $t=11.756$) 변수로 모두 $p < 0.001$ 로 유의한 결과가 도출되었다. 반면 음식점·카페 수(Est.=-0.089, $t=-3.323$, $p=0.001$)와 녹지비율(Est.=-0.047, $t=-4.174$, $p < 0.001$)은 음(-)의 관계로 유의하게 도출되었으며, 교육 공공시설 수

(Est.=-0.002, t=-0.139, p=0.889)는 음의 영향을 보이지만 낮시간에서는 유의하지 않은 변수로 도출되었다.

[표 4-15] MLR 매개변수에 대한 요약통계(낮시간)

Variable	Est.	SE	t-value	p-value
Intercept	0.000	0.009	0.000	1.000
술집 수	0.352	0.017	20.630	<0.001
음식점 카페 수	-0.089	0.027	-3.323	0.001
소매시설 수	0.150	0.013	11.361	<0.001
교통시설 수	0.170	0.013	13.535	<0.001
건물개수	0.108	0.012	8.741	<0.001
상업비율	0.106	0.014	7.873	<0.001
편의점 수	0.178	0.021	8.639	<0.001
녹지비율	-0.047	0.011	-4.174	<0.001
공시지가	0.147	0.013	11.756	<0.001
교육 공공시설 수	-0.002	0.013	-0.139	0.889

출처: 연구진 작성

• MLR 모델-밤시간

MLR 밤시간 모델에서 112신고 건수와 가장 큰 양(+) 효과를 보이는 변수는 술집 수 (Est.=0.633, t=40.603)이며 이어 편의점 수 (Est.=0.210, t=11.160), 건물개수 (Est.=0.102, t=9.046), 교통시설 수 (Est.=0.106, t=9.267), 소매시설 수 (Est.=0.082, t=6.763), 공시지가 (Est.=0.061, t=5.366), 교육 공공시설 수 (Est.=0.043, t=3.720, p<0.001), 상업비율 (Est.=0.034, t=2.785, p=0.005)이 모두 양(+)으로 유의한 변수로 도출되었다. 녹지비율 (Est.=-0.035, t=-3.472, p=0.001)과 음식점·카페 수 (Est.=-0.174, t=-7.167, p<0.001)는 음(-)으로 유의하며 영향의 크기가 크지는 않으나 양(+)으로 유의하다. 낮시간과 비교하면 술집 수·편의점 수의 양(+)의 효과가 확대되었으며 교통시설·소매점의 영향은 다소 축소되었고 음식점·카페 수의 음(-) 효과는 확대된 것으로 보인다.

[표 4-16] MLR 매개변수에 대한 요약통계(밤시간)

Variable	Est.	SE	t-value	p-value
Intercept	0.000	0.008	0.000	1.000
술집 수	0.633	0.016	40.603	<0.001
음식점 카페 수	-0.174	0.024	-7.167	<0.001
소매시설 수	0.082	0.012	6.763	<0.001

교통시설 수	0.106	0.011	9.267	<0.001
건물개수	0.102	0.011	9.046	<0.001
상업비율	0.034	0.012	2.785	0.005
편의점 수	0.210	0.019	11.160	<0.001
녹지비율	-0.035	0.010	-3.472	0.001
공시지가	0.061	0.011	5.366	<0.001
교육 공공시설 수	0.043	0.012	3.720	<0.001

출처: 연구진 작성

■ GWR 모델 해석

GWR 모델에서는 각 격자 위치별 국지계수를 추정해 전역 모형이 놓치지 쉬운 공간적 차이를 드러낸다. Mean 값은 격자마다 추정된 ‘국지 회귀계수’들의 평균값이며 STD는 그 계수들의 공간적 변동성(지역별 분산 정도)을 뜻한다. 또한, 모든 변수가 같은 국지 범위(Bandwidth: 대역폭)로 추정되었기 때문에 변수 간 Mean·STD를 상호 비교할 수 있다. 따라서 GWR 결과는 OLS에서 확인된 방향성 및 효과 크기의 지역별 편차를 확인시켜 준다.

• GWR 모델-전체시간 (bandwidth 97)

전체시간 GWR 모델에서는 모든 변수가 동일 대역폭(97)에서 추정되었다. 변수별로 살펴보면 술집 수(Mean: 0.383, STD: 0.345)는 112신고 건수와의 양(+)의 효과가 나타나며 변동성도 큰 편으로 보여진다. 이는 112신고건수와 증가와 큰 관련이 있으며 지역별 차이가 클 것으로 해석될 수 있다. 음식점·카페 수(Mean: 0.058, STD: 0.391)는 양(+)의 영향을 보이나 평균 영향력은 작으며 변동성이 큰 경향을 보여 지역별로 영향의 차이가 상이할 것으로 보여진다. 소매시설 수(Mean: 0.168, STD: 0.329)와 교통시설 수(Mean: 0.120, STD: 0.104)는 모두 평균 양(+)의 경향을 보이지만 전자는 변동성이 비교적 크고 후자는 변동성이 작은 편으로 나타나 소매시설 수는 지역별 차이가 크고 교통시설 수는 비교적 일관되게 나타난다고 요약할 수 있다. 건물개수(Mean: 0.081, STD: 0.227)는 평균적으로 양(+)의 영향을 가지면서 변동성이 중간 이상으로 효과 크기 차가 눈에 띄는 편이다. 상업비율(Mean: 0.038, STD: 0.155)은 평균적으로 약한 양(+)의 영향과 중간 수준 변동성을 보이며 편의점 수(Mean: 0.124, STD: 0.192)는 상업비율에 비해서는 강한 양(+)의 영향과 더 높은 변동성을 보이는 것으로 나타났다. 녹지비율(Mean: -0.039, STD: 0.134)은 평균적으로 음(-)의 효과가 나타나며 변동성은 비교적 작은 편이라 전반적으로 음의 경향이 나타난다고 해석할 수 있다. 공시지가(Mean: 0.164, STD: 0.222)는 평균 양(+)이면서 변동성이 중간 이상으로 효과 크기 차가 꽤 큰 축에 속한다. 마지막으로 교육 공공시설 수(Mean: -0.011, STD: 0.226)는 평균이 0에 가까워 전역적 방향성이 약하지만 음(-)의 방향으로 나타나며 변동성은 큰 편으로 효과의 지역별 이질성이 상대적으로 크게 나타난다는 것을 시사한다.

[표 4-17] GWR 매개변수에 대한 요약통계(전체시간)

Variable	Mean	STD	Min	Median	Max	Band width
Intercept	0.032	0.157	-0.488	0.045	1.012	97
술집 수	0.383	0.345	-1.481	0.377	2.074	97
음식점 카페 수	0.058	0.391	-1.472	0.045	1.629	97
소매시설 수	0.168	0.329	-0.850	0.132	1.440	97
교통시설 수	0.120	0.104	-0.185	0.113	0.575	97
건물개수	0.081	0.227	-0.517	0.064	2.311	97
상업비율	0.038	0.155	-0.895	0.035	0.570	97
편의점 수	0.124	0.192	-0.546	0.117	1.418	97
녹지비율	-0.039	0.134	-0.596	-0.026	1.442	97
공시지가	0.164	0.222	-0.744	0.137	1.127	97
교육 공공시설 수	-0.011	0.226	-0.689	-0.026	1.363	97

출처: 연구진 작성

- GWR 모델-낮시간 (bandwidth 97)

낮시간 GWR 모델에서도 전체시간 GWR 모델과 마찬가지로 모든 변수가 동일 대역폭(97)에서 추정되었다. 변수별로 보면 술집 수(Mean: 0.205, STD: 0.420)는 112신고건수와 평균 양(+)의 연관성을 보이며 변동성이 매우 큰 편이다. 음식점·카페 수(Mean: 0.109, STD: 0.461)도 평균적으로 양(+)의 연관을 보이지만 술집 수보다는 적은 영향력을 가지고 변동성은 더 크게 확인된다. 소매시설 수(Mean: 0.233, STD: 0.383)는 평균 양(+)의 영향이며 변동성도 큰 축에 속한다. 교통시설 수(Mean: 0.152, STD: 0.127) 또한 평균 양(+)의 영향력을 가지며 변동성은 비교적 작은 편으로 일관성이 높다. 건물개수(Mean: 0.069, STD: 0.225), 상업비율(Mean: 0.079, STD: 0.163), 편의점 수(Mean: 0.105, STD: 0.238)는 평균 양(+)의 영향을 가지나 비교적 약한 효과를 보이며 변동성은 중간 정도에 속한다. 녹지비율(Mean: -0.053, STD: 0.146)은 평균이 112신고건수와 음(-)의 영향을 가지며 이는 112신고건수 감소에 영향을 주는 변수로 해석될 수 있고 변동성은 중간 수준이다. 공시지가(Mean: 0.191, STD: 0.236)는 평균 양(+)의 영향, 변동성은 중간 이상으로 요약된다. 마지막으로 교육 공공시설 수(Mean: -0.041, STD: 0.246) 또한 녹지비율과 마찬가지로 평균적으로 음(-)의 영향을 보이며 변동성은 녹지비율보다 더 크게 보여진다.

[표 4-18] GWR 매개변수에 대한 요약통계(낮시간)

Variable	Mean	STD	Min	Median	Max	Bandwidth
Intercept	0.032	0.168	-0.549	0.039	0.819	97
술집 수	0.205	0.420	-2.763	0.224	2.369	97
음식점 카페 수	0.109	0.461	-1.457	0.057	3.369	97
소매시설 수	0.233	0.383	-0.969	0.162	1.878	97
교통시설 수	0.152	0.127	-0.231	0.132	0.618	97
건물개수	0.069	0.225	-0.999	0.077	1.781	97
상업비율	0.079	0.163	-0.662	0.063	0.735	97
편의점 수	0.105	0.238	-0.960	0.100	1.199	97
녹지비율	-0.053	0.146	-0.846	-0.028	1.100	97
공시지가	0.191	0.236	-0.691	0.176	1.257	97
교육 공공시설 수	-0.041	0.246	-0.802	-0.047	1.326	97

출처: 연구진 작성

- GWR 모델-밤시간 (bandwidth 89)

밤시간 GWR 모델은 모든 변수를 동일 대역폭(89)에서 추정했으며 변수별로 낮시간과 비교하면, 술집 수(Mean: 0.481, STD: 0.357)는 평균 영향이 더 커지고 변동성은 작아져 야간에 상대적으로 강하고 일관된 양(+) 경향이 나타난다고 해석가능하다. 음식점·카페 수(Mean: 0.040, STD: 0.403)는 평균 영향이 작아지고 변동성도 완화되어 낮보다 영향력이 전반적으로 축소되는 것으로 보여진다. 소매시설 수(Mean: 0.116, STD: 0.333), 교통시설 수(Mean: 0.098, STD: 0.103) 역시 평균과 변동성이 동시에 감소해 야간에는 낮보다 효과가 약하고 공간적 차이도 줄어드는 것으로 확인된다. 건물개수(Mean: 0.086, STD: 0.256)는 평균과 변동성이 모두 증가하여 밤에는 영향력도 커지고 지역별 차이도 확대되는 양상을 보인다. 상업비율(Mean: 0.010, STD: 0.174)은 평균이 크게 낮아지고 변동성은 다소 커져 야간에는 전반 영향이 약해지는 대신 지역별 편차는 커진다고 해석될 수 있다. 편의점 수(Mean: 0.129, STD: 0.184)는 평균이 증가하고 변동성은 감소하여 밤에 더 강하고 비교적 고른 양(+)의 효과가 관찰된다. 녹지비율(Mean: -0.031, STD: 0.144)은 음(-) 크기가 완화되고 변동성도 유사하거나 소폭 감소하여 야간에는 억제 효과가 낮보다 다소 약해지는 것으로 보인다. 공시지가(Mean: 0.143, STD: 0.239)는 평균이 작아지고 변동성은 거의 비슷하여 야간에는 낮보다 영향이 다소 줄지만 지역별 차이 수준은 유사하다. 교육 공공시설 수(Mean: 0.007, STD: 0.228)는 평균이 0에 더 가까워지며 양(+)의 효과로 전환되었고 변동성은 감소하여 야간에는 전반적으로 효과가 미미해지고 공간적 차이도 일부 축소된 것으로 해석된다.

[표 4-19] GWR 매개변수에 대한 요약통계(밤시간)

Variable	Mean	STD	Min	Median	Max	Bandwidth
Intercept	0.032	0.167	-1.114	0.041	1.176	89
술집 수	0.481	0.357	-1.195	0.468	2.580	89
음식점·카페 수	0.040	0.403	-1.502	0.025	1.348	89
소매시설 수	0.116	0.333	-1.325	0.097	1.636	89
교통시설 수	0.098	0.103	-0.252	0.096	0.582	89
건물개수	0.086	0.256	-0.373	0.060	2.836	89
상업비율	0.010	0.174	-2.382	0.009	0.492	89
편의점 수	0.129	0.184	-0.582	0.118	1.578	89
녹지비율	-0.031	0.144	-0.584	-0.022	1.954	89
공시지가	0.143	0.239	-0.844	0.101	1.367	89
교육·공공시설 수	0.007	0.228	-0.726	-0.011	1.438	89

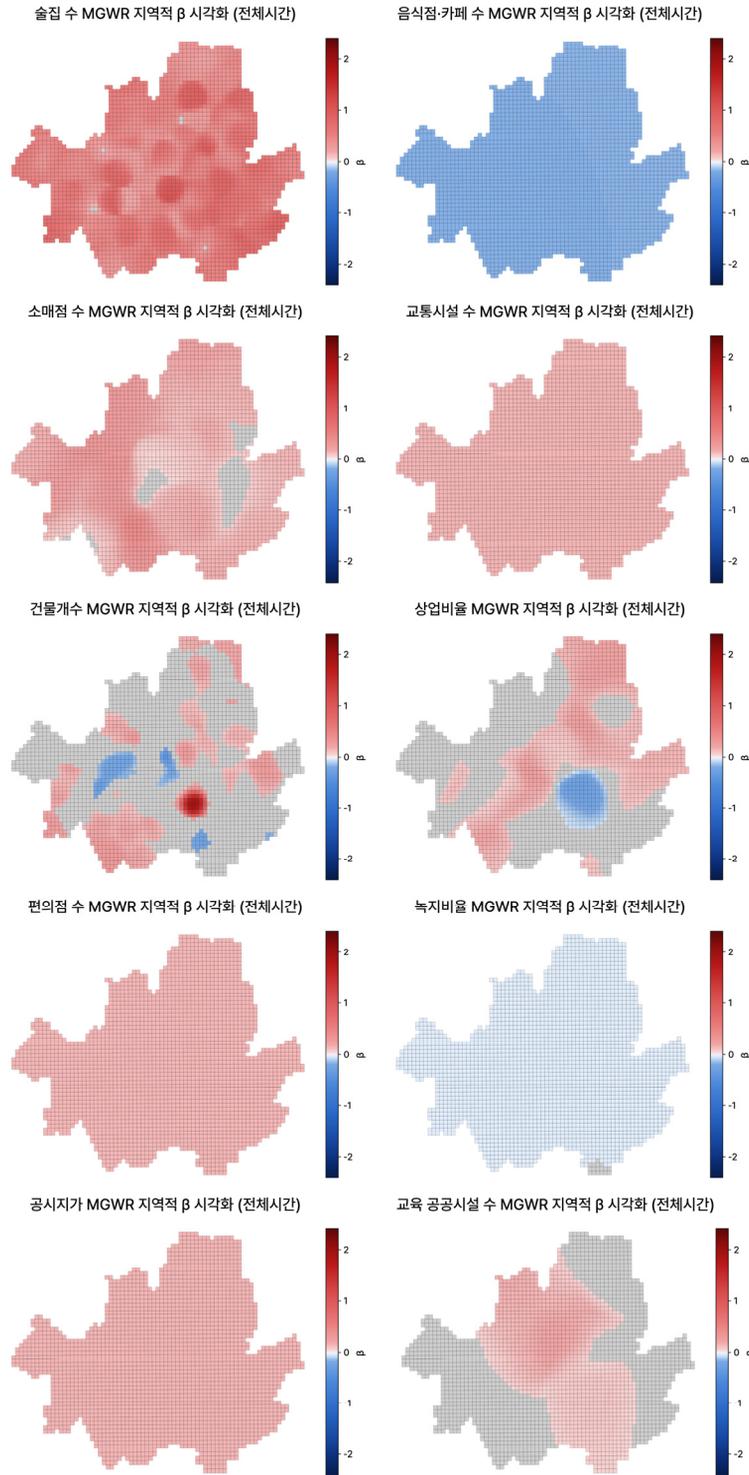
출처: 연구진 작성

■ MGWR 모델 해석

MGWR 모델은 변수별로 대역폭(Bandwidth)을 별도로 추정하여 각 계수를 계산할 때 적응형 커널(adaptive bisquare) 기준으로 얼마나 많은 주변 이웃을 포함해 가중 회귀를 수행했는지를 수치로 제시한다. 대역폭이 작을수록 소수의 가까운 관측치만 반영하는 국지적 패턴, 중간값이면 생활권 단위의 중간 패턴, 아주 클수록 광범위한 관측치를 포함하는 도시 전반(광역) 패턴으로 읽을 수 있다. 이 값은 AICc 최소화로 자동 결정되며 따라서 MGWR 출력은 부호(방향)·크기(Mean)·변동성(STD)에 더해 작동 범위(대역폭)까지 함께 해석할 수 있다.

• MGWR 모델-전체시간

MGWR 전체시간 모델 결과를 살펴보면, 술집 수(Mean: 0.505, STD: 0.147, Bandwidth: 57)는 평균적으로 112신고 건수와 양(+)의 관계를 가지며 매우 작은 대역폭으로 국지적 패턴이 강조된다. 음식점·카페 수(Mean: -0.170, STD: 0.001, Bandwidth: 2626)는 평균적으로 음(-)의 관계를 보이고 대역폭이 매우 커 공간 전반에서 거의 일정한 효과가 나타난다. 소매시설 수(Mean: 0.163, STD: 0.078, Bandwidth: 221)는 양(+)의 관계를 가지며 중간 대역폭에서 추정되었다. 교통시설 수(Mean: 0.146, STD: 0.002, Bandwidth: 2633)는 양(+)의 관계를 보이고 아주 큰 대역폭과 매우 작은 변동성으로 광역 수준의 일관된 패턴을 보인다. 건물개수(Mean: 0.073, STD: 0.195, Bandwidth: 57)는 양(+)의 관계이고 작은 대역폭에서 변동성이 비교적 큰 편으로 추정되었다. 상업비율(Mean: 0.065, STD: 0.092, Bandwidth: 176)은 양(+)이며 중간 대역폭을 갖는다. 편의점 수(Mean: 0.140, STD: 0.001, Bandwidth:



[그림 4-18] 변수별 MGWR 계수 분포 (전체시간)
 빨간색:(+) 영향/파란색:(-) 영향/회색 : 비유의지역($p>0.05$)
 출처: 연구진 작성

2633)는 양(+), 아주 큰 대역폭과 극히 작은 변동성으로 광역 패턴이 두드러진다. 녹지비율 (Mean: -0.025, STD: 0.002, Bandwidth: 2633)은 음(-)의 효과를 가지며 아주 큰 대역폭과 작은 변동성을 보여 전반적으로 안정적인 패턴을 보인다. 공시지가(Mean: 0.146, STD: 0.002, Bandwidth: 2626)는 양(+), 큰 대역폭과 매우 작은 변동성으로 광역 수준의 일관성이 확인된다. 마지막으로 교육 공공시설 수(Mean: 0.057, STD: 0.055, Bandwidth: 399)는 약한 양(+),의 관계를 보이며 중간보다 큰 대역폭과 중간 변동성으로 요약된다. 정리하면, MGWR 결과는 술집 수·건물개수처럼 작은 대역폭에서 추정된 변수와 교통시설 수·편의점 수·녹지비율·공시지가·음식점·카페 수처럼 아주 큰 대역폭에서 거의 일정하게 나타나는 변수를 구분해 주며 소매시설 수·상업비율·교육 공공시설 수는 그 사이의 중간 범위에서 추정되어 변수별 작동 스케일 차이를 명확히 보여준다.

[표 4-20] MGWR 매개변수에 대한 요약통계(전체시간)

Variable	Mean	STD	Min	Median	Max	Bandwidth
Intercept	-0.018	0.095	-0.277	-0.010	0.261	89
술집 수	0.505	0.147	0.100	0.495	1.013	57
음식점 카페 수	-0.170	0.001	-0.172	-0.170	-0.167	2626
소매시설 수	0.163	0.078	-0.005	0.152	0.439	221
교통시설 수	0.146	0.002	0.143	0.146	0.148	2633
건물개수	0.073	0.195	-0.436	0.072	1.972	57
상업비율	0.065	0.092	-0.289	0.069	0.255	176
편의점 수	0.140	0.001	0.138	0.140	0.141	2633
녹지비율	-0.025	0.002	-0.028	-0.024	-0.022	2633
공시지가	0.146	0.002	0.143	0.146	0.149	2626
교육 공공시설 수	0.057	0.055	-0.034	0.044	0.211	399

출처: 연구진 작성

• MGWR 모델-낮시간

MGWR 낮시간 모델 결과를 살펴보면 술집 수 (Mean: 0.304, STD: 0.157, Bandwidth: 57)는 평균적으로 112신고건수와 양(+),의 영향을 보이며 작은 대역폭으로 국지적 차이가 비교적 강하게 반영된다. 음식점·카페 수 (Mean: -0.074, STD: 0.000, Bandwidth: 2633)는 음(-)의 영향을 보이고 변동성이 사실상 0으로 매우 큰 대역폭에서 공간 전반에 걸쳐 일관된 방향이 나타난다. 소매시설 수 (Mean: 0.196, STD: 0.114, Bandwidth: 177), 교통시설 수 (Mean: 0.168, STD: 0.071, Bandwidth: 175)는 양(+),의 효과를 가지며 중간 대역폭에서 추정되어 생활권 수준에서의 차이가 관찰됨을 알 수 있다. 건물개수 (Mean: 0.077, STD:

0.160, Bandwidth: 89)는 양(+의 효과를 보이면서 변동성이 중간 수준 이상이며 작은 대역폭으로 국지적인 이질성이 함께 관찰된다고 볼 수 있다. 상업비율 (Mean: 0.112, STD: 0.123, Bandwidth: 89)은 양(+의 효과를 보이고 변동성이 중간 수준이며 작은 대역폭으로 지역별 차이가 비교적 뚜렷하게 나타남을 확인 할 수 있다. 편의점 수 (Mean: 0.127, STD: 0.002, Bandwidth: 2633)는 양(+의 효과를 보이고 변동성이 매우 작으며 광역 대역폭에서 전반적으로 균일한 효과를 보인다고 해석할 수 있다. 녹지비율 (Mean: -0.036, STD: 0.002, Bandwidth: 2633)은 음(-)의 효과를 보이며 변동성이 거의 없고 대역폭은 광역적으로 나타나며 이는 전반적으로 음의 효과가 나타남을 알 수 있다. 공시지가 (Mean: 0.174, STD: 0.001, Bandwidth: 2631)는 양(+의 효과와 함께 변동성이 극히 작아 매우 큰 대역폭에서 일관된 양의 효과가 관찰된다고 볼 수 있다. 교육 공공시설 수 (Mean: 0.023, STD: 0.053, Bandwidth: 412)는 양(+의 효과가 관찰되고 변동성은 중간 수준이며 비교적 큰 중간 대역폭에서 추정되어 생활권을 넘어서는 범위에서 작동하는 모습을 확인할 수 있다.

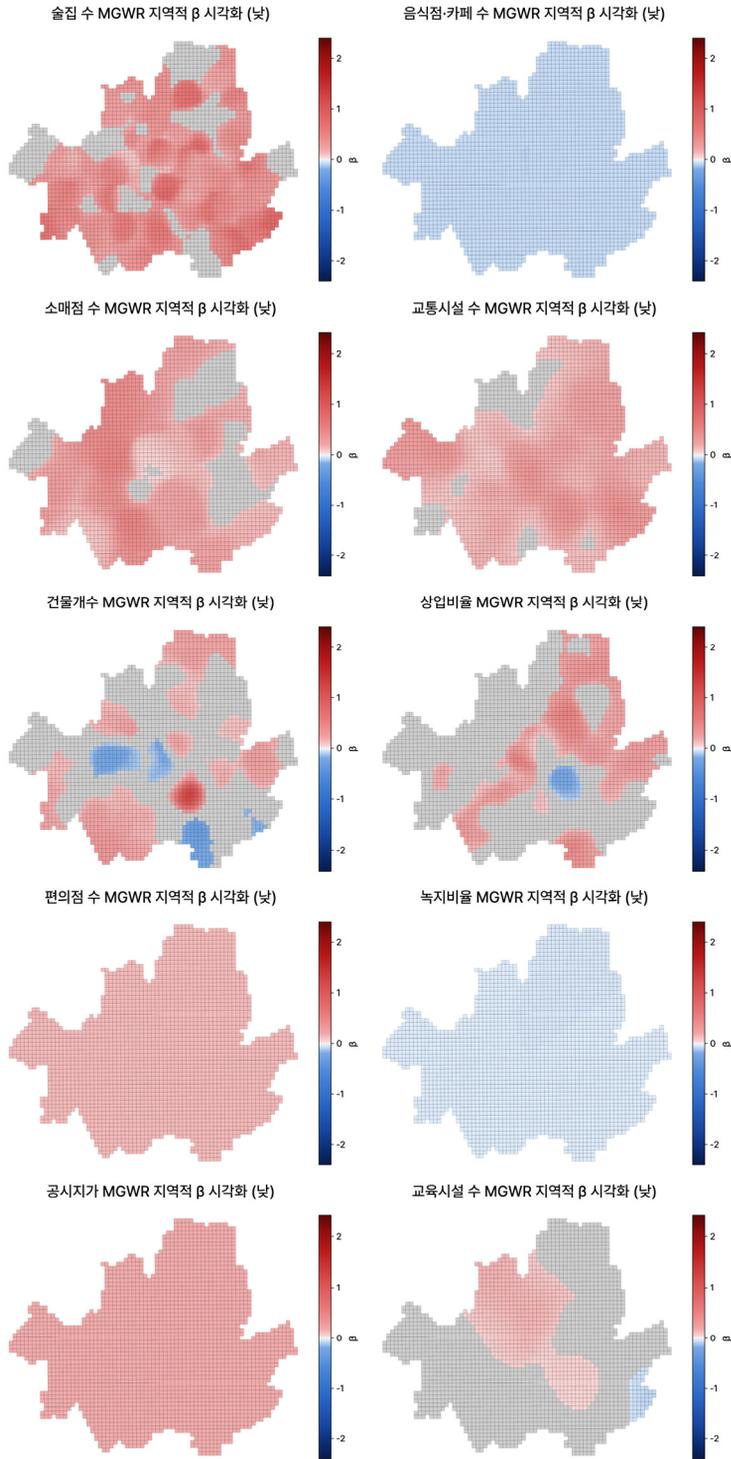
[표 4-21] MGWR 매개변수에 대한 요약통계(낮시간)

Variable	Mean	STD	Min	Median	Max	Bandwidth
Intercept	-0.013	0.099	-0.319	-0.008	0.278	89
술집 수	0.304	0.157	-0.106	0.289	0.776	57
음식점 카페 수	-0.074	0.000	-0.075	-0.074	-0.073	2633
소매시설 수	0.196	0.114	0.003	0.168	0.506	177
교통시설 수	0.168	0.071	0.043	0.155	0.388	175
건물개수	0.077	0.160	-0.331	0.088	1.239	89
상업비율	0.112	0.123	-0.224	0.101	0.514	89
편의점 수	0.127	0.002	0.124	0.127	0.130	2633
녹지비율	-0.036	0.002	-0.039	-0.036	-0.032	2633
공시지가	0.174	0.001	0.172	0.174	0.176	2631
교육 공공시설 수	0.023	0.053	-0.085	0.010	0.156	412

출처: 연구진 작성

• MGWR 모델-밤시간

MGWR 밤시간 모델 결과를 살펴보면 술집 수 (Mean: 0.611, STD: 0.156, Bandwidth: 57)는 낮에 비해 평균 양(+) 효과가 약 2배 가량 강화되었으며 작은 대역폭에서 국지적 차이를 강하게 반영하는 것을 확인할 수 있다. 음식점·카페 수 (Mean: -0.221, STD: 0.001, Bandwidth: 2633)는 음(-)의 효과가 낮보다 더 뚜렷하고 매우 큰 대역폭(광역)과 거의 0에 가까운 변동성으로 공간 전반의 일관된 감소 효과가 유지된다. 소매시설 수 (Mean: 0.137, STD: 0.077, Bandwidth: 221)는 양(+의 효과를 가지지만 낮 대비 평균 효과가 다소 축소되어 중간 스케



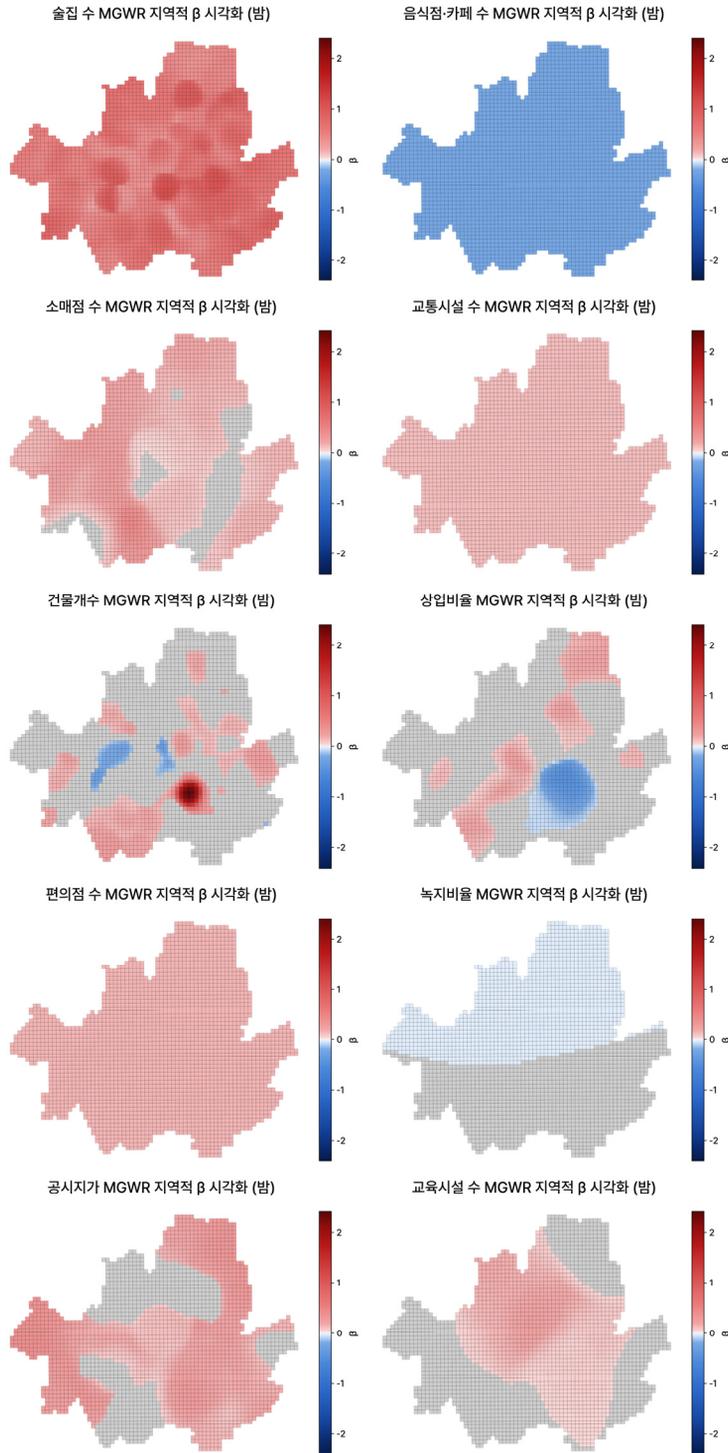
[그림 4-19] 변수별 MGWR 계수 분포 (낮시간)
 빨간색:(+) 영향/파란색:(-) 영향/회색 : 비유의지역(p>0.05)
 출처: 연구진 작성

일에서 완만하게 작동한다. 교통시설 수 (Mean: 0.118, STD: 0.000, Bandwidth: 2633)는 광역 스케일에서 변동성이 거의 0으로 균일한 양(+)의 효과가 유지되며 낮보다 평균 크기는 작다. 건물개수 (Mean: 0.083, STD: 0.228, Bandwidth: 57)는 미시적 스케일에서 변동성이 더 커져 특히 밤에 지역별 이질성이 확대되는 모습을 보인다. 상업비율 (Mean: 0.030, STD: 0.115, Bandwidth: 137)은 평균 양(+)의 효과가 낮보다 완화되고 중간 이하 대역폭에서 공간적 차이는 유지된다. 편의점 수 (Mean: 0.143, STD: 0.001, Bandwidth: 2633)는 광역적 스케일에서 매우 작은 변동성과 함께 균일한 양(+)의 효과가 지속되며 낮보다 평균은 소폭 커진다. 녹지비율 (Mean: -0.020, STD: 0.002, Bandwidth: 2633)은 광역적 스케일에서 변동성이 거의 없고 음(-) 효과는 낮보다 약간 완화된다. 공시지가 (Mean: 0.163, STD: 0.096, Bandwidth: 177)는 낮의 광역-안정적 패턴에서 밤에는 중간 대역폭으로 이동하고 변동성이 커져 야간에 지역별 차이가 보다 크게 나타나는 모습으로 전환된다. 교육 공공시설 수 (Mean: 0.065, STD: 0.062, Bandwidth: 399)는 평균이 소폭 양(+)의 방향으로 커지고 변동성은 비슷한 수준을 유지하며 중간 대역폭에서 작동하는 모습을 보인다.

[표 4-22] MGWR 매개변수에 대한 요약통계(밤시간)

Variable	Mean	STD	Min	Median	Max	Bandwidth
Intercept	-0.017	0.107	-0.281	-0.005	0.275	97
술집 수	0.611	0.156	0.216	0.600	1.130	57
음식점 카페 수	-0.221	0.001	-0.223	-0.221	-0.218	2633
소매시설 수	0.137	0.077	-0.007	0.131	0.448	221
교통시설 수	0.118	0.000	0.117	0.118	0.119	2633
건물개수	0.083	0.228	-0.343	0.064	2.409	57
상업비율	0.030	0.115	-0.479	0.042	0.236	137
편의점 수	0.143	0.001	0.142	0.143	0.144	2633
녹지비율	-0.020	0.002	-0.023	-0.020	-0.017	2633
공시지가	0.163	0.096	-0.074	0.157	0.382	177
교육 공공시설 수	0.065	0.062	-0.021	0.043	0.250	399

출처: 연구진 작성



[그림 4-20] 변수별 MGWR 계수 분포 (밤시간)
 빨간색:(+) 영향/파란색:(-) 영향/회색 : 비유의지역(p>0.05)
 출처: 연구진 작성

제5장

결론

1. 연구의 성과
2. 연구의 한계 및 향후 과제

1. 연구의 성과

■ 물리환경 요인이 112신고 발생에 미치는 영향을 500m 격자 단위로 공간통계 및 기계학습 기법을 사용하여 정량 분석

본 연구는 112신고가 집중되는 지역의 물리적·사회적 특성을 체계적으로 규명하고자 서울시 전역을 균일한 500m × 500m 격자로 세분화하여 112신고 발생 건수의 시·공간 분포를 정량적으로 분석하였다. 아울러 도시 물리환경 요인이 신고 변동에 어떻게 관여하는지 예측·해석 가능한 근거를 마련함으로써 공간 단위의 위험 요인 파악과 정책적 개입 지점을 도출하였다. 구체적으로 서울시 전역(25개 자치구)을 대상으로 112신고 데이터를 공간통계기법(LISA)을 통해 분석하고 핫스팟(High-High) 및 콜드스팟(Low-Low) 지역을 중심으로 물리적·사회적 환경을 비교·검증하였다. 아울러 실제 신고 데이터에 다양한 물리환경 변수를 결합한 뒤, 부스팅 계열 트리 기반 예측모형을 학습·검증하고 모델 내부 의사결정을 설명 가능한 AI 기법인 SHAP으로 투명화하여 변수의 기여 방향(증가/감소)과 크기, 임계구간 및 상호작용을 해석하였다. 더 나아가 GWR 및 MGWR 공간회귀모형을 적용하여 지역별로 상이한 계수(공간적 이질성)와 변수별 영향 범위의 다중 스케일성을 포착함으로써 동일한 요인이라도 위치와 스케일에 따라 효과가 달라질 수 있음을 확인하였다. 종합적으로 본 연구의 주된 목적은 112신고가 단순한 사건 빈도 지표에 그치지 않고 지역 사회의 구조적 속성, 기반시설, 토지이용, 인구 구성 등 복합적 요인과 체계적으로 연결되어 있다는 가정을 실증적으로 검증하고 그 메커니즘을 시·공간적으로 설명하는 데 있다.

■ LISA 분석 결과, 핫스팟(High-High)은 도심 및 한강 이남 중심지 집중, 콜드스팟(Low-Low)은 외곽 지역 광범위 분포

먼저 LISA 클러스터링 결과를 보면 핫스팟(High-High) 유형 격자는 서울시 전역에 산재하나 특히 도심 및 남부권 중심지에 집중되는 양상을 보였다. 강남구는 44건으로 가장 많은 핫스팟 격자를 보였고 중구·관악구·영등포구 등도 핫스팟 격자 수가 높게 나타났다. 반면 콜드스팟(Low-Low) 유형은 북부 및 외곽 지역을 중심으로 넓게 분포하였으며 서초구 및 강서구 등은 콜드스팟 격자가 핫스팟 격자보다 훨씬 더 많았다. HL / LH 유형은 전반적으로 매우 희소했

으며 대부분의 행정구에서 0~3건 수준에 불과했다. 이러한 패턴은 서울시 내 112신고 밀집 현상이 공간적으로 비교적 안정적이며 주변 지역과 유사한 성향을 갖는 격자군이 주를 이룬다는 점을 암시한다.

공간 군집 유형 분포를 자치구별로 비교하면 일부 행정구는 핫스팟이 우세하고 일부는 콜드스팟이 우세하며 일부는 두 유형이 병존하는 양상을 보였다. 예컨대 중구는 핫스팟 33건, 콜드스팟 1건으로 핫스팟이 절대적으로 우세한 반면, 서초구는 콜드스팟 122건, 핫스팟 19건으로 콜드스팟이 우세하다. 강남구는 핫스팟 44건, 콜드스팟 43건으로 매우 균형적인 분포를 보였고 관악구나 종로구 등도 핫스팟과 콜드스팟이 모두 일정 수준 존재해 복합적 공간 구조를 드러냈다.

■ 핫스팟과 콜드스팟은 물리적·사회적 환경 변수 대부분에서 유의한 차이 보였으며 특히 인구 및 건축물 밀도에서 큰 차이

다음으로 중요한 것은 핫스팟 지역과 콜드스팟 지역 간 물리적·사회적 환경 변수 차이이다. 이를 위해 핫스팟 격자와 콜드스팟 격자 간 Mann-Whitney U 비모수 검정을 실시하였고 대부분의 변수에서 유의수준 $p < 0.05$ 를 충족하여 통계적으로 유의한 차이가 확인되었다.

인구 통계 측면에서는 핫스팟 지역이 전 연령대에서 콜드스팟 지역을 훨씬 상회했다. 유아부터 고령까지 모든 계층에서 핫스팟 지역의 중앙값이 콜드스팟 지역의 중앙값을 압도적으로 초과했으며 콜드스팟 지역의 중앙값은 대체로 0에 가까운 경우가 많았다. 이는 신고가 집중된 공간이 인구가 밀집한 공간과 대체로 일치함을 보여준다.

건물 관련 정보에서도 핫스팟 지역은 콜드스팟 지역보다 높은 값을 보였다. 공시지가는 핫스팟 지역이 평균적으로 매우 높았으며 용적률, 건폐율, 건물 높이, 건물 개수 등 밀도 관련 지표에서도 유의미한 차이를 보였다. 예를 들어 용적률은 핫스팟 지역이 콜드스팟 지역에 비해 훨씬 높았고 건물 개수나 층 수 역시 비약적으로 많았다.

토지이용 분야에서는 핫스팟 지역이 주거지, 상업지, 복합용도지로 구성되는 반면 콜드스팟 지역은 녹지, 공터, 낮은 밀도용지 중심이었다. 특히 핫스팟 지역은 상업비율이 높고 주거비율이 높은 경향을 같이 보였으며 콜드스팟 지역은 녹지비율과 공터 비율(오픈스페이스)이 높게 나타났다.

생활 편의 및 상권 시설 측면에서도 핫스팟 지역은 음식점, 술집, 편의점, 병원, 교육시설, 대중교통시설, CCTV 등 다양한 시설이 집중되어 있었고 콜드스팟 지역은 이들 시설이 거의 없거나 매우 낮은 수준이었다. 핫스팟 지역 내에서는 시설이 없는 경우가 드물었고 복합 인프라가 집적된 공간 특성이 강하게 나타났다.

마지막으로 도로 네트워크 측면에서 도로 폭은 핫스팟 지역이 평균적으로 더 넓었지만 통계적으로 유의성은 제한적이었다. 일부 자치구에서만 유의미한 차이를 보였고 전체적으로 일관된 패턴이라고 보기엔 무리가 있었다.

■ 기계학습 및 공간회귀 분석 결과, 술집 수·편의점·교통시설·소매시설·건물 수 등이 신고 증가 영향 요인으로 녹지비율·교육공공시설 수는 신고 감소 요인으로 나타남

HH-LL 비교는 112신고가 어떤 토지이용·시설·네트워크 조건과 함께 나타나는지의 ‘동반 패턴’을 파악하는데 매우 용이하다. 그러나 이러한 대비만으로는 변수들의 상대적 중요도, 영향 방향(±), 비선형성·임계구간을 정량화하거나 미관측 격자에 대한 예측 가능성을 확보하기 어렵다. 따라서 다음 단계에서는 예측 가능성과 해석 가능성을 동시에 달성할 수 있는 기계학습 기반 접근을 통해 112신고와 도시 물리환경 간의 관계를 모형적으로 규명하였다.

예측모델로는 여러 머신러닝 모델 후보 중 성능 비교를 거쳐 XGBoost를 기준 모형으로 채택했으며 과적합 억제를 위한 정규화와 일관된 데이터 분할 전략을 적용해 재현성과 신뢰성을 확보하였다. 변수 해석 측면에서 전역적 중요도(평균 절대 SHAP 값 기준)는 ‘술집 수’가 일관되게 최상위 영향 변수로 드러났고 뒤이어 ‘편의점 수’, ‘교통시설 수’, ‘소매시설 수’, ‘건물개수’ 등의 변수가 112신고 건수 예측값을 키우는 데 크게 기여했다. 반대로 ‘녹지비율’, ‘교육공공시설 수’는 전반적으로 신고를 낮추는 방향으로 작동했다. SHAP 요약 플롯과 의존도 플롯을 결합하면 단순한 “중요/비중요”를 넘어 각 변수의 방향(±), 기여 크기, 값의 범위에 따른 비선형 반응, 전환점(임계구간)을 함께 파악할 수 있다. 예를 들어 술집 수는 전 구간에서 일관된 단조 증가 패턴으로 112신고건수 증가와 관련된 경향을 보이고 음식점·카페 수는 특정 임계값까지 급격히 상승한 뒤 감소로 전환되는 양상이 나타난다. 반면 녹지비율은 일정 수준을 넘어서면 112신고건수와의 감소 연관성이 뚜렷해지는 비선형 패턴이 관찰된다.

추가적으로 실시한 공간회귀 비교(MLR/GWR/MGWR)는 어떤 요인이 어느 공간 스케일에서 작동하는지를 구조적으로 보여준다. 전역모형(MLR)은 전체 방향성과 평균 효과를 간결히 제시해 기준선을 제공한다. GWR은 위치마다 계수가 달라지는 공간적 이질성을 반영해 지역별 편차를 수치로 드러낸다. MGWR은 더 나아가 변수별로 최적 대역폭을 추정함으로써 어떤 변수는 국지적으로 작동하고 어떤 변수는 광역적으로 작동하는지 구분해 준다. 실무적으로는 국지 스케일 변수에 대해 미시 개선(가로 환경·조도·시야 확보·골목 관리 등)을 광역 스케일 변수에는 네트워크·용도구성·거점 관리 등 구조적 처방을 우선 검토할 수 있다. 따라서 예측모형이 “무엇이 중요했는가”를 말해 준다면 공간회귀 결과를 통해서도 “어디에서 어떤 스케일로 개입할 것인가”를 판단할 수 있다.

■ 본 연구는 112신고를 안전 지표로 확장하여 물리환경이 범죄 발생 및 무질서에 미치는 영향에 대한 기존 주장을 500m 격자 수준에서 실증하였다는 점에서 기여

주요 변수별로 본 연구에서 발견한 내용을 선행연구의 맥락에서 해석하면 다음과 같다. 본 연구 결과에서 전역 중요도 1위로 관찰된 '술집 수'는 112신고 건수 증가와 전 구간에서 일관된 양(+)의 관계를 보였으며 특히 야간에 그 효과가 주간보다 뚜렷하게 강화되는 것으로 나타났다. 이는 알코올 판매시설 밀도가 폭력·강력범죄 및 무질서 사건과 유의한 양(+)의 상관관계를 보인다는 다수의 국제 연구 결과와 일관된다(Toomey et al, 2009; Gruenewald et al, 2006). 선행 연구들은 알코올 시설 밀도가 폭력 범죄와 긍정적으로 연관되어 있음을 발견했으며(Toomey et al, 2009), 특히 불안정한 빈곤층 소수민족 지역과 농촌 중산층 지역의 폭력과 유의미한 상관관계가 있음을 확인하였다(Gruenewald et al, 2006). 이를 통해 알코올 업소를 단순한 상업 밀집지가 아니라 구조적 위험 환경의 핵심 요소로 규정해 왔다. 이러한 맥락에서 '술집 수'의 영향 관련 본 연구의 기여는 이러한 국제적 논의를 112신고 데이터를 통하여 500m×500m 격자 단위의 미시 공간에서 실증하였다는 점에 있다.

다른 환경 변수들에서도 본 연구 결과를 통하여 미시 공간 수준의 위험 패턴을 확인할 수 있었다. 특히 음식점·카페 수, 소매점 수, 편의점 수 등 상업시설 관련 변수들 역시 112신고 건수 증가와 유의한 양(+)의 관계를 보이는 것으로 나타났다. 이는 상업시설들의 집적이 인구 유입과 밀집, 익명적 보행 인구의 증가를 초래하여 잠재적 범죄 기회와 신고 발생 가능성을 동시에 높인다는 점에서 상업 활동의 공간적 집중이 신고 고위험 지역을 형성하는 구조적 요인으로 작동할 수 있음을 시사한다(안영환 외, 2024). 본 연구에서는 더 나아가 음식점·카페 수가 500m×500m 격자 내 약 25개 부근에서 112신고건수와의 관계가 음(-)에서 양(+)의 효과로 전환되며 편의점 수가 약 4개 수준에서 양(+) 방향으로 전환되는 등 비선형적 관계를 발견하였다는 점에서 의의가 있다.

지하철역·버스정류장 등 교통시설 수 역시 112신고 건수와 유의한 양(+)의 관련성을 보였다. 이러한 양상은 다수의 잠재적 표적, 가해자, 방관자가 특정 시점과 공간에 고밀도로 응집되는 범죄 생성자로 기능한다는 일상활동이론 및 해외 연구 결과와 부합한다. Loukaitou-Sideris(1999)는 로스앤젤레스에서 특정 버스정류장이 강력범죄와 무질서의 '핫스팟'으로 반복되는 것을 보였고 Stucky & Smith(2014)는 인디애나폴리스에서 버스정류장 밀도가 강도·절도 등의 범죄와 유의하게 연관되며 특히 상업·공업 용도지역에서 그 효과가 강화된다는 점을 확인하였다. 본 연구의 결과는 이러한 선행연구 결과에서 더 나아가 500m×500m 격자 내 약 7개의 교통시설 수 내외에서 효과가 음(-)에서 양(+)으로 전환되는 비선형적 관계를 확인하였다. MGWR 분석 결과 지역적으로는 공간적 변동성은 크지 않은 변수로 확인되었다.

녹지비율 변수와 교육 공공시설 수는 전반적으로 음(-)의 효과가 나타났다. 특히 녹지의 경

우, 약 20% 수준까지는 완충효과, 일부 구간에서는 효과 약화 또는 반전 가능성이 관찰되었다. 관련 선행 연구에서도 도시 녹지는 잘 관리·가시화된 녹지의 경우 폭력·재산범죄를 유의하게 낮추지만 방치·차폐된 녹지는 은신·비가시 영역을 제공해 위험을 높일 수 있는 양면성이 반복 보고된다(Wolfe et al, 2012; Wo et al, 2024). 서울을 대상으로 한 CPTED 선행연구에서는 수목활력 (NDVI)이 심리적 안정을 통해 폭력성을 낮추고 영역성 강화와 이미지 개선 효과가 있는 것으로 밝혀졌다(조민균, 2018). 이는 본 연구에서 도심 격자 단위에서 높은 녹지 비율이 신고를 억제하는 결과와 일치한다.

또한 교육·공공시설 수가 증가할수록 112신고가 감소하는 경향은 학교·공공기관·복지·문화 시설 등이 정기적 활동과 공식적 감시를 통해 집합효능성을 강화하고 주변 무질서를 흡수·완화할 수 있다는 기존 논의와 맥을 같이 한다(Ferreira et al., 2025; Sampson et al., 1997). 그러나 일부 연구에서는 특정 학교가 인접 블록의 절도·침입 등 범죄가 비정상적으로 집중되는 ‘핫스팟’으로 기능할 수 있음을 지적하고 있으며(Kautt et al., 2007), 이는 본 연구의 MGWR 분석에서 일부 구간에서 교육·공공시설 주변에서 국지적인 양(+) 효과가 관찰된 결과와 상응할 수 있다. 본 연구의 성과는 이렇게 엇갈린 선행연구의 발견에 대하여 비선형적 관계를 밝힌 것으로 향후 이를 고려한 후속 연구가 이어질 필요가 있다.

종합하면 본 연구는 112신고 데이터를 범죄·불안 및 위험 징후를 포괄하는 안전 지표로 확장하여 해석하고 500m 격자 단위의 미시 공간 분석, 머신러닝 예측(XGBoost), 설명가능 AI(SHAP), 공간회귀(MGWR)를 통합함으로써 서울 전역에서 ‘어떤 환경 요인이 어느 수준과 스케일에서 어떤 방식으로 신고 위험을 증폭 또는 완충하는가’를 체계적으로 제시하였다. 이러한 결과는 향후 시설 단위 CPTED 설계, 야간 운영 관리, 용도·밀도 규제 등 구체적인 예방 전략을 수립하는 데 직접 활용할 수 있는 실증적 근거를 제공할 수 있다.

[표 5-1] 주요 환경요인별 주요 결과 및 이론적 근거와 정책적 함의

변수	분석 결과 요약 (본 연구)	이론·선행 근거 (핵심 인용)	함의 및 한계·향후과제
술집 수	① 112신고 증가에 양의(+) 영향 변수/가장 큰 기여도 가짐 ② 변수 값이 약 10개 부근에서 음(-)에서 양(+)으로 전환 후 증가 ③ 밤에 양(+) 효과가 낮의 2배 이상 강화됨 ④ 미시 단위의 영향력 ⑤ 공간적 변동성에 따른 지역별 편차 있음	- 고립된 술집과 바 스트리트 근처에서 범죄율이 더 높으며 수년에 걸쳐 술집과 범죄 사이의 공간적 연결이 확인됨 (Huang et al, 2025) - 주류 판매점 주변의 해로운 사회적, 물리적 조건(예: 깨진 병, 어슬렁거리는 사람 등)은 주민과 방문객에게 잠재적으로 위협적인 것으로 여겨짐(Nesoff et al., 2018).	① 그리드당 술집 개수 허가 총량/제한 등 밀도 상한 설정 ② 심야(야간) 시간대 술집 밀집 부근 순찰·단속 집중 ③ 집 주변 가시성 증대 및 CPTED 기준 의무화

변수	분석 결과 요약 (본 연구)	이론·선행 근거 (핵심 인용)	함의 및 한계·향후과제
음식점·카페 수	<ul style="list-style-type: none"> ① 112신고 증가에 양의(+) 영향 변수 ② 변수 값이 약 25개 부근에서 음(-)에서 양(+)으로 전환 후 증가 ③ 밤에 음(-) 효과가 낮보다 강화됨 ④ 거시 단위의 영향력 ⑤ 광역 단위에서 거의 일정한 패턴 	<ul style="list-style-type: none"> - 패스트 푸드 음식점의 수가 높은 곳에서 폭력 범죄율이 더 높은 것으로 확인됨 (Singleton, et al., 2022) - 몇몇 연구에서는 일반음식점 수와 범죄 사이에 음의 관계가 있는 것으로 나타나는 반면 다른 연구에서는 양의 관계가 있는 것으로 나타나 상충되는 결과가 있는 것으로 보임(Sohn, D., 2016; Rosenthal et al., 2010) - 카페는 사회적 행동 유도 등을 통해 범죄율을 감소하는 것으로 여겨짐(Wo, J., 2016) 	<ul style="list-style-type: none"> ① 고위험 외식업소 유형(패스트푸드 등)에 대한 허가 제한 및 위치 기준 마련 ② 저위험 업소(카페 등)의 사회적 연결 기능 촉진을 위한 디자인 및 관리 기준 도입 ③ 식음시설 유형별 차등 관리 방안 마련: 추후 단순 업소 수보다 업종별 성격 반영 필요
소매 시설 수	<ul style="list-style-type: none"> ① 112신고 증가에 양의(+) 영향 변수 ② 변수 값이 약 10개 부근에서 음(-)에서 양(+)으로 전환 후 증가 ③ 낮과 밤 모두 양(+) 효과를 보이며 밤에 영향력이 소폭 감소 ④ 미시~생활권(중간) 단위 영향력 ⑤ 공간적 변동성에 따른 지역별 편차 있음 	<ul style="list-style-type: none"> - 소매 시설 수가 높은 곳이 범죄율이 높은 경향이 보이는 것으로 나타나며 이는 도시 내 혼잡도와 익명성 증가 등의 가능성 때문일수 있음(Rosenthal et al., 2010) - 모든 유형의 소매점은 범죄와 상관관계가 있으나 주류 판매점이 특히 범죄와 높은 관련성을 보이는 것으로 확인됨(Cameron, M., 2022) 	<ul style="list-style-type: none"> ① 고위험 업종(주류 판매점 등) 밀집 제한 및 관리 강화 ② 혼잡도 및 익명성 완화 위한 CPTED 기반 설계 도입 ③ 생활권 단위의 유동인구 기반 치안 배치 최적화
교통 시설 수	<ul style="list-style-type: none"> ① 112신고 증가에 양의(+) 영향 변수 ② 변수 값이 약 7개 부근에서 음(-)에서 양(+)으로 전환 후 증가 ③ 낮과 밤 모두 양(+) 효과를 보이며 밤에 영향력이 소폭 감소 ④ 거시 단위에서 영향력 ⑤ 공간적 변동성 거의 없음 	<ul style="list-style-type: none"> - 지하철 역은 범죄 유발 또는 범죄 유인 장소로 기능할수있음이 확인되었으며 공간적, 시간적 패턴 모두 중요한 것이 강조됨(Irvin-Erickson et al. 2015) - 교통시설 그 자체보다는 그 주변의 도시 디자인 특성이 범죄와 더 밀접한 관련이 있는 것으로 확인됨(Su et al., 2023) 	<ul style="list-style-type: none"> ① 교통 요충지(터미널 등) 주변 상시 순찰 강화 ② 교통거점 내 범죄예방 디자인 적용: 동선 가시성, CCTV 시각지대 최소화
건물 개수	<ul style="list-style-type: none"> ① 112신고 증가에 양의(+) 영향 변수 ② 변수 값이 약 400개 부근에서 음(-)에서 양(+)으로 전환 후 증가 ③ 낮과 밤 모두 양(+) 효과를 보이며 밤에 영향력이 소폭 증가 ④ 미시 단위의 영향력 ⑤ 야간에 지역 간 이질성 증가 	<ul style="list-style-type: none"> - 더 많은 수의 건물은 절도와 관련된 범죄와 연관성을 갖은것으로 확인됨(Silva et al., 2020) - 건물 수 증가에 따른 범죄율은 부유한 지역에서는 상승 추세를 보인 반면, 저소득 지역에서는 이미 높은 범죄율로 인해 일정 수준 이후 더 이상 증가하지 않는 경향을 나타냄(Hipp et al., 2022) 	<ul style="list-style-type: none"> ① 건물 밀집지역 대상 맞춤형 방범 인프라 확충 ② 사회경제적 조건을 고려한 차등적 대응 전략 수립 ③ 고위험 건물군(다세대·공실률 높은 노후건물 등) 중심의 CPTED 적용 확대
상업 비율	<ul style="list-style-type: none"> ① 112신고 증가에 양의(+) 영향 변수 ② 변수 값이 약 8% 부근에서 음(-)에서 양(+)으로 전환 후 증가 ③ 낮에 양(+) 효과가 밤의 약 4배 강화됨 ④ 미시 단위의 영향력 ⑤ 전반적으로 공간적 변동성이 있으며 야간에 지역 간 이질성 증가 	<ul style="list-style-type: none"> - 상업 비율이 더 높은 지역은 더 높은 범죄율과 연관이 있는 것으로 발견되었으며 이는 증가하는 낮선 방문객 수에 따른 것으로 해석됨(Twinam et al., 2017) - 상업비율과 범죄는 비선형적인 관계를 보여주는 경향이 있으며 이는 일정 수준 이상에서는 자연 감시가 강화되기 때문인 것으로 해석됨(Browning et al., 2010) 	<ul style="list-style-type: none"> ① 상업용도 비율 기반 '방범 취약 경계선' 설정 ② 상업밀집도 변화 감시를 위한 공간지표 개발 및 시계열 모니터링 체계 구축 ③ 비선형 관계 고려한 상업 비율 밀도 관리전략 수립

변수	분석 결과 요약 (본 연구)	이론·선행 근거 (핵심 인용)	함의 및 한계 향후과제
편의점 수	<ul style="list-style-type: none"> ① 112신고 증가에 양의(+) 영향 변수 ② 변수 값이 약 4개 부근에서 음(-)에서 양(+)으로 전환 후 증가 ③ 낮과 밤 모두 양(+) 효과를 보이며 밤에 영향력이 소폭 증가 ④ 거시 단위의 영향력 ⑤ 공간적 변동성 거의 없음 	<ul style="list-style-type: none"> - 편의점 수의 증가는 폭력 및 절도 범죄율과 양의 관계가 있는 것으로 확인됨 (White et al., 2015) - 폭력범죄는 편의점으로부터 거리가 가까울수록 높은 것으로 나타남 (Furr-Holden et al., 2016) 	<ul style="list-style-type: none"> ① 편의점 수 밀집도 기준 및 거리 제한 도입 검토 ② 심야영업점의 출입구 개방 방향, 조도, CCTV 기준 강화 ③ 무인편의점 등의 조명 및 보안 관리 규제 신설
녹지 비율	<ul style="list-style-type: none"> ① 112신고 증가에 음의(-) 영향 변수 ② 변수 값이 약 20% 부근에서 음(-)에서 음(-)으로 전환 ③ 낮과 밤 모두 음(-) 효과를 보이며 밤에 영향력이 조금 감소 ④ 거시 단위의 영향력 ⑤ 공간적 변동성 거의 없음 	<ul style="list-style-type: none"> - 도심 속 녹지 공간은 낮은 범죄율과 연관 있으며 이는 자연 감시, 사회적 교류, 스트레스 완화 등을 통한 것으로 해석됨 (Shepley et al., 2019) - 잘 계획된 녹지는 공공성, 가시성, 접근성이 좋아 범죄 발생을 억제하는 반면 버려진 녹지는 오히려 범죄 은신처가 될 수 있음 (Venter et al. 2022) 	<ul style="list-style-type: none"> ① 일정 비율 이상의 녹지 확보를 위한 도시계획 기준 강화 ② 범죄예방형 녹지 디자인 가이드라인 개발(개방성, 조명 등 포함) ③ 야간에도 안전하게 이용 가능한 녹지 조성 및 관리 방안 마련
공시지가	<ul style="list-style-type: none"> ① 112신고 증가에 양의(+) 영향 변수 ② 변수 값이 약 200만원인 부근에서 양(+)으로 전환 후 증가하다 급격히 떨어져 약 550만원 부근에서 음(-)으로 전환 후 다시 약 600만원 부근에서 양(+)으로 전환 ③ 낮과 밤 모두 양(+) 효과를 보이며 밤에 영향력이 조금 감소 ④ 미시~생활권(중간) 단위의 영향력 ⑤ 공간적 변동성은 낮에는 거의 없었으나 밤에 지역별 차이 증가 	<ul style="list-style-type: none"> - 절도 범죄율은 더 낮은 소득의 지역과 인접해 있는 고소득 지역에서 더 높은 것으로 나타남 (Metz et al., 2018) - 그에 반해 폭력 범죄율 등은 공시지가가 낮은 지역에서 더 높은 것으로 나타남 (McIlhatton et al., 2016) 	<ul style="list-style-type: none"> ① 고가 주거지에서의 야간 순찰강도 강화 ② 저가 지역에 대한 범죄예방 설계지원(조도, 감시체계 등) 강화 ③ 부동산 가치와 방범자원 투입 간의 불균형 해소 전략 필요
교육 공공 시설 수	<ul style="list-style-type: none"> ① 112신고 증가에 음의(-) 영향 변수 ② 변수 값이 약 25개 부근에서 양(+)에서 음(-)으로 전환 ③ 밤에 양(+) 효과가 낮의 약 3배 강화됨 ④ 생활권(중간)~거시 단위의 영향력 ⑤ 약간의 공간적 변동성 있음 	<ul style="list-style-type: none"> - 중학교~대학교 근처에서 폭력 범죄 및 절도 범죄율이 더 높은 것을 나타남 (Cundiff K., 2020) - 그에 반해 더 나은 환경의 학교 주변 혹은 초등학교 근처는 절도 범죄율이 적은 것으로 나타남 (Willits et al., 2013; Limbos et al., 2008) 	<ul style="list-style-type: none"> ① 학교급별 환경수준별 맞춤 방범 전략 도입 ② 야간 개방형 공공시설(학교·도서관 등)의 안전관리 강화 ③ 학교 주변 입지 가이드라인 수립 ④ 학교 환경의 질적 요인 반영한 CPTED 인증제 도입

2. 연구의 한계 및 향후 과제

- 112신고 데이터 기반 범죄 발생·범죄두려움 영향 실증하였으나 112신고 데이터 활용 제약은 구조적 한계이며 격자 단위 물리환경 외 영향 요인에 대한 연구범위 확장 필요

본 연구는 격자 단위 미시 공간에 대한 112신고 발생에 미치는 물리환경 변수의 영향을 실증하여 범죄 발생 및 범죄두려움에 미치는 물리환경의 영향을 체계적으로 확인하였다. 그러나 본 연구는 112신고 데이터 활용과 관련하여 다양한 한계를 보였으며 물리환경 외 영향 요인에 대한 고려를 확장할 필요성도 제기되었다.

먼저 「112신고처리법」은 112신고 데이터의 연구 목적 활용을 고려하지 않고 있으며 “112신고 통계 현황을 정기적으로 분석”할 것을 규정하면서도 112신고자 정보가 드러날 수 있는 데이터 제공에 대한 처벌을 강하게 규정하고 있어 연구 목적 활용에 현실적으로 제약이 되고 있다. 본 연구에서도 112신고 유형의 구분이 되지 않은 격자 단위 집계치만 분석에 활용할 수 있어 신고 유형 등 세밀한 분석에 어려움이 있었다.

한편, 본 연구에서 범죄 발생 및 범죄두려움에 영향을 미치는 물리환경 변수에 대하여 폭넓게 고려하였으나 격자 단위 공공데이터로 제공하지 않는 물리환경 변수나 인구 및 사회학적 변수 등을 분석에 반영하지 못한 점은 연구의 한계로 남는다. 특히 유동인구는 모든 사회현상의 양적 특성에 절대적 영향을 미칠 것으로 예상되는 변수이나 휴대전화 기반 유동인구 데이터를 격자 단위로 구득하려면 높은 비용이 발생하여 본 연구에서는 적용하지 못하였다. 주민등록인구는 독립변수에 포함되었으나 주·야간 유동인구를 고려한 분석은 후속 연구에서 우선적으로 적용할 필요가 있다.

이러한 한계를 고려할 때 향후 연구 방향은 범죄 및 범죄두려움에 대한 영향 요인을 격자 단위 데이터로 제공하고 있는 변수로 한정하지 않고 격자 단위, 행정구역/집계구 단위, 전국 단위 데이터를 모두 고려할 필요가 있다. 또한 서울 외 지역에 대한 분석을 통하여 지역 간 차이를 확인할 필요가 있다.

이러한 후속 연구를 통하여 각 지역에서 이루어지고 있는 CPTED 정책 및 사업이 실증 근거에 기반하여 실질적인 효과를 거둘 수 있는 방향으로 이루어질 수 있을 것이다.

- 강용길. (2015). 지역경찰의 112신고 대응 실태에 관한 연구. 치안정책연구소.
- 경찰청. (2023). 2022년 자체평가결과보고서.
- 경찰청. (2020). 112신고접수 및 지령 매뉴얼. 경찰청.
- 경찰청. (2021). 빅데이터기반인공지능(AI) 범죄예측시스템(Pre-CAS) 전국운영. 4월 30일 보도자료.
- 경찰청. (2023). 112신고접수 현황. 지표누리. https://www.index.go.kr/unity/potal/main/EachDt/1PageDetail.do?idx_cd=1609 (검색일: 2025.2.27.).
- 김상호. (2009). 대구경북 지역의 치안수요 분석: 112신고 접수 자료를 중심으로. 한국지방자치연구, 11(3), 185-203.
- 김중곤, 이태현. (2020). 빅데이터 분석을 통한 순찰 및 112신고 대응 효율화 모델: 대구시의 사례를 중심으로. 한국치안행정논집, 17(3), 91-106.
- 김지은, 김지현, 유안식, 이성용. (2023). 야간 시간대 보행자의 범죄 두려움 감소를 위한 CPTED 적용 방안: 광진구 자양2동을 중심으로. 한국경찰연구, 23(2), 179-204.
- 김현중, 이성우. (2011). 수도권 5 대 범죄의 결정요인: Mixed GWR 모형의 적용: Mixed GWR 모형의 적용. 서울도시연구, 12(4), 137-155.
- 김현호, 정제용. (2020). 112신고 벡터화를 통한 텍스트 마이닝 기반의 군집분석과 이를 활용한 사건 유형 간 유사성 탐색. 한국치안행정논집, 17(3), 63-90.
- 노성훈, 조준택. (2016). 112 긴급신고시스템 운용상의 문제점 실증분석 및 개선방안. 경찰학논총, 11(4), 9-38.
- 류지용, 김호. (2023). 국가긴급신고제도의 개선을 위한 법적 연구. 입법학연구, 21(2), 31-62.
- 박성미, & 안정민. (2022). 스마트치안기술활성화를위한형사사법정보활용방안: 개인정보보호와의균형을중심으로. 범죄수사학연구, 205-224.
- 송상현 (2024) '현장 경찰만 아는' 유흥주점·폐가 위치 수시 공유... '집단지성' 활용, 2월15일 기사, <https://www.news1.kr/society/incident-accident/5320655> (검색일: 2025.6.12.)
- 신승재. (2023). 112치안종합상황실 공간표준화를 위한 기초연구. 한국공간디자인학회 논문집, 18(4), 139-148.
- 안영환, 김동현, & 박진아. (2024). 생활업종별 밀도가 주거침입 범죄 발생에 미치는 영향 연구: 범죄 신고 데이터를 활용하여. Journal of Korea Planning Association-Vol, 59(2), 57-72.
- 염윤호, 김혁. (2023). 아동학대 112신고의 시공간적 군집 분석: 서울시 관악구를 중심으로. 경찰학연구, 23(3), 117-135.
- 이상훈, 김재홍. (2023). 112신고 건수에 영향을 미치는 주요 변수 분석. 한국데이터정보과학회지, 34(6), 1147-1160.

- 이유지(2021) 오늘밤 당신 동네 범죄... AI는 이미 알고 있다, 9월18일 기사. <https://www.hankookilbo.com/News/Read/A2021091617580002935> (검색일: 2025.6.13.).
- 이정, 정준호. (2022). 골목길 특성 및 지역 낙인 효과를 고려한 서울시 도시 범죄 영향 요인 연구: 112신고 데이터를 중심으로. *도시행정학보*, 35(4), 1-27.
- 이지은, 최훈. (2023). 국민 시각을 반영한 112신고 코드 체계에 관한 연구. *치안정책연구*, 37(2), 5-38.
- 이희정, 김주환, 김성준. (2021). 112 범죄 신고 데이터 분석을 통한 캠퍼스 내에서의 범죄취약요소 연구. *한국공간정보학회지*, 29(4), 87-98.
- 장일식, 장석태. (2024). 세종특별자치시 자살 감소 방안에 관한 탐색적 연구. *한국경찰연구*, 24(1), 153-170.
- 정민주, 김현우. (2023). 인지적 지수를 고려한 골목상권 범죄 영향요인 분석 및 공간적 변이 탐색: COVID-19 전후를 대상으로. *도시행정학보*, 36(2), 143-167.
- 정연대, 이윤호, 김광주. (2023). 사회적 거리두기로 인한 지역 치안수요의 변화에 관한 연구: 월별 112신고 자료를 중심으로. *한국경찰연구*, 23(1), 177-200.
- 정재훈, 박현호. (2024). Transformer 기반 LLM의 학습을 이용한 112 허위오인신고분류-예측모델 개발. *한국통신학회논문지*, 49(4), 489-502. <https://doi.org/10.7840/kics.2024.49.4.489>
- 정제용. (2019). 영국 경찰기관의 빅 데이터 활용 관리와 정책 시사점에 관한 연구. *경찰학연구*, 19(1), 85-114.
- 정제용. (2021). 112신고자료에 대한 빅데이터 분석: EDA 기법을 중심으로. *시큐리티연구*, 66, 74-92.
- 정진도. (2023). 골목길 특성 및 지역 낙인 효과를 고려한 서울시 도시 범죄 영향 요인 연구: 112신고 데이터 및 서울시 골목길 조사 데이터를 활용한 마이크로 스케일 영향 연구. *한국도시설계학회지 도시설계*, 24(5), 19-36. <https://doi.org/10.38195/judik.2023.10.24.5.19>.
- 조민균, 박찬 and 장정인. (2018). CPTED 요소로써 서울시 공원·녹지의 효과와 계획적 함의. *한국조경학회지*, 46(3), 27-35.
- 조영진, 안의순, 박성남, 고영호, 권오규, 임보영, 임리사, 김유진, 이정현. (2024). 범죄예방 환경설계(CPTED) 고도화 및 인증제도 개선 방향. *건축공간연구원*.
- 조영진, 한수경, 김승남, 박유나. (2021). 저층주거지 내 범죄예방 환경조성 시설·기법의 효과성 실증 연구. *대한건축학회논문집*, 37(10), 23-33.
- 조준택, 김강일, 박현호. (2018). 범죄예방을 위한 112신고 자료의 활용방안 연구. *한국치안행정논집*, 15(1), 257-282.
- 최대만. (2023, 12월 10일). 범죄 피해자 보호 강화를 위한 112신고법 국회 통과. *충북일보*.
- 최재훈. (2018). 기계학습을 활용한 데이터 기반 경찰신고건수 예측. *한국빅데이터학회지*, 3(2), 101-112.
- 한국형사정책연구원. (2024). 범죄피해두려움. 지표누리. <https://www.index.go.kr/unity/potal/indicator/IndexInfo.do?clasCd=2&idxCd=4264> (검색일: 2025.2.28.).
- 홍세은, 방준성, 김용진, 장광호. (2020). 112신고 벡터화를 통한 텍스트 마이닝 기반의 군집분석과 이를 활용한 사건유형 간 유사성 탐색. *경찰학연구*, 20(3), 63-86.
- 황정용. (2024). AI의 112 상황실 적용 및 기대효과: 신고 접수와 경찰 언어모델 개발을 중심으로. *한국치안행정논집*, 21(2), 179-198.
- Ahmed, M. U., Bari, M. W., & Hossain, M. A. (2018). Cross-validation for reliable performance estimation in regression modeling. *IEEE Access*, 6, 63992-64002.
- Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93-115.

- Armitage, R. (2013). *Crime Prevention Through Housing Design: Policy and Practice*. Palgrave Macmillan.
- Beck, C., & McCue, C. (2009). Predictive policing: What can we learn from Wal-Mart and Amazon about fighting crime in a recession? *The Police Chief*, 76(11), 18-24.
- Behn, R. D. (2008). *The seven big errors of PerformanceStat*. Harvard Kennedy School, Rappaport Institute for Greater Boston.
- Brantingham, P. J., Valasik, M., & Mohler, G. O. (2018). Does predictive policing lead to biased arrests? Results from a randomized controlled trial. *Statistics and Public Policy*, 5(1), 1-6.
- Brantingham, P. L., & Brantingham, P. J. (1993). Environment, routine and situation: Toward a pattern theory of crime. In R. V. Clarke & M. Felson (Eds.), *Routine Activity and Rational Choice* (pp. 259-294). Transaction Publishers.
- Browning, C. R., Byron, R. A., Calder, C. A., Krivo, L. J., Kwan, M.-P., Lee, J.-Y., & Peterson, R. D. (2010). Commercial density, residential concentration, and crime: Land use patterns and violence in neighborhood context. *Journal of Research in Crime and Delinquency*, 47(3), 329-357. <https://doi.org/10.1177/0022427810365906>
- Bureau of Justice Assistance. (2013). *CompStat: Its origins, evolution, and future in law enforcement agencies*. Police Executive Research Forum.
- Cameron, M. P. (2022). The relationship between alcohol outlets and crime is not an artefact of retail geography. *Addiction*, 117(8), 2215-2224. <https://doi.org/10.1111/add.15880>
- Chainey, S., & Ratcliffe, J. (2005). *GIS and crime mapping*. John Wiley & Sons.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Chen, X., Li, Y., & Zhao, J. (2022). Explaining gradient boosting decision tree models: A SHAP approach for environmental and social sciences. *Environmental Modelling & Software*, 157, 105-132.
- Chen, Y., Liu, X., Li, X., & Liu, T. (2020). Spatial heterogeneity and its implications in urban studies: A review of geographically weighted regression. *Sustainability*, 12(1), 1-19.
- Corcoran, V. (2019). *Predicting violent crime reports from geospatial and temporal attributes of US 911 emergency call data*. Master's thesis, Technological University Dublin.
- Cramer, D., Brown, A. A., & Hu, G. (2012). Predicting 911 calls using spatial analysis. In R. Lee (Ed.), *Software Engineering Research, Management and Applications 2011. Studies in Computational Intelligence* (Vol. 377). Springer. https://doi.org/10.1007/978-3-642-23202-2_2.
- Cundiff, K. (2020). Colleges and community crime: An analysis of campus proximity and neighborhood crime rates. *Crime & Delinquency*, 67(3), 431-448. <https://doi.org/10.1177/0011128720974312>
- Czarnota, J., Gibbons, R. D., & Delucchi, K. (2015). Geographically weighted regression for investigating spatial heterogeneity of associations between neighborhood environment and health. *Annals of Epidemiology*, 25(10), 755-761.
- Data Collaborative for Justice. (2020). *Tracking enforcement trends in New York City: 2003-2018*. John Jay College of Criminal Justice.
- Degeling, M., & Berendt, B. (2018). What is wrong about Robocops as consultants? A technology-centric critique of predictive policing. *Ai & Society*, 33(3), p347-356.

- Ferreira Neto, A. B., Nowicki, J., & Shakya, S. (2025). Do public libraries help mitigate crime? Evidence from Kansas City, MO. *Journal of cultural economics*, 49(1), 141-161.
- Field, S. (1992). The effect of temperature on crime. *The British Journal of Criminology*, 32(3), 340-351. <https://doi.org/10.1093/oxfordjournals.bjc.a048222>
- Fotheringham, A. S., Brunson, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Wiley.
- Fotheringham, A. S., Charlton, M. E., & Brunson, C. (2003). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 35(11), 1905-1927.
- Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, 107(6), 1247-1265.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>
- Furr-Holden, C. D. M., Milam, A. J., Nesoff, E. D., Johnson, R. M., Fakunle, D. O., Jennings, J. M., & Thorpe, R. J. Jr. (2016). Not in my back yard: A comparative analysis of crime around publicly funded drug treatment centers, liquor stores, convenience stores, and corner stores in one mid-Atlantic city. *Journal of Studies on Alcohol and Drugs*, 77(1), 17-24. <https://doi.org/10.15288/jsad.2016.77.17>
- Gamble, J. L., & Hess, J. J. (2012). Temperature and violent crime in Dallas, Texas: Relationships and implications of climate change. *Western Journal of Emergency Medicine*, 13(3), 239-246. <https://doi.org/10.5811/westjem.2012.3.11746>
- Goh, C. F., Yusof, Z. M., & San, O. T. (2019). A predictive model using multiple linear regression for household electricity consumption. *Sustainability*, 11(3), 939.
- Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing*, 27(3), 659-678. <https://doi.org/10.1007/s11222-016-9646-1>
- Gruenewald, P. J., Freisthler, B., Remer, L., LaScala, E. A., & Treno, A. (2006). Ecological models of alcohol outlets and violent assaults: crime potentials and geospatial analysis. *Addiction*, 101(5), 666-677.
- Henderson, L. J. (2003). *The Baltimore CitiStat program: Performance and accountability*. IBM Center for The Business of Government.
- Hipp, J., Lee, S., Ki, D., & Kim, J. (2022). How concentrated disadvantage moderates the built environment and crime relationship on street segments in Los Angeles. *Criminology & Criminal Justice*, 25, 501-529. <https://doi.org/10.1177/17488958221132764>
- Home Office. (2005). *Crime mapping: Improving performance - A good practice guide for front line officers*. Home Office Communications Directorate.
- Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383-401.
- Huang, S., Qin, J., & Li, Y. (2021). Comparative study of MLR and ML-based methods for hydrological prediction. *Journal of Hydrology*, 597, 126332.
- Huang, Z., Xu, Y., Li, Y., Jiang, S., & Chen, R. (2025). Where drinks and danger meet: analyzing the spatial link between bars and crime in Detroit. *Applied Geography*, 174, 103480.
- Hunt, P., Saunders, J., & Hollywood, J. S. (2014). Evaluation of the Shreveport predictive polic

- ing experiment. RAND Corporation.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.
- Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers* (Vol. 16). ASQC Quality Press.
- Irvin-Erickson, Y., & La Vigne, N. G. (2015). A spatio-temporal analysis of crime at Washington, DC Metro Rail: Stations'crime-generating and crime-attracting characteristics as transportation nodes and places. *Crime Science*, 4(1), Article 14. <https://doi.org/10.1186/s40163-015-0026-5>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.
- Jeffery, C. R. (1971). *Crime prevention through environmental design*. Sage Publications.
- de Jongh, P. J., Pienaar, M., & Gordon-Grant, H. (2015). The impact of pre-selected variance inflation factor thresholds on the stability and predictive power of logistic regression models in credit scoring. *ORiON*, 31(1), 17-37.
- Kautt, P. M., & Roncek, D. W. (2007). Schools as criminal "hot spots" primary, secondary, and beyond. *Criminal Justice Review*, 32(4), 339-357.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146-3154.
- Kim, J. (2020). Multicollinearity in regression: Problems, detection, and solutions. *Journal of Urban Planning and Development*, 146(2), 04020014.
- Lee, S., & Park, Y. (2019). Model selection using Mallows' Cp, AIC, and BIC in environmental regression. *Environmental Modelling & Software*, 118, 124-132.
- Limbos, M., & Casteel, C. (2008). Schools and neighborhoods: Organizational and environmental factors associated with crime in secondary schools. *The Journal of School Health*, 78(10), 539-544. <https://doi.org/10.1111/j.1746-1561.2008.00341.x>
- Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Wiley.
- Liu, Y., Li, Z., Xiong, H., & Chen, Y. (2019). Geographically weighted regression: Principles, applications, and perspectives. *Computers, Environment and Urban Systems*, 75, 84-97.
- Los Angeles Police Department. (2015). *2014 Annual report*. City of Los Angeles.
- Los Angeles Police Department. (2021). *Strategic plan 2021-2023*. City of Los Angeles.
- Loukaitou-Sideris, A. (1999). Hot spots of bus stop crime: The importance of environmental attributes. *Journal of the American Planning association*, 65(4), 395-411.
- Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science*, 28(4), 660-681.
- Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, 13(5), 14-19.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- Lundberg, S. M., Erion, G. G., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), 252-259.

- Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review Online*, 94, 15-55.
- Robles, P., Tello, A., Solano-Quinde, L., & Zúñiga-Prieto, M. (2020). Temporal analysis of 911 emergency calls through time series modeling. In M. Botto-Tobar et al. (Eds.), *Advances in Emerging Trends and Technologies. ICAETT 2019. Advances in Intelligent Systems and Computing*, 1066. Springer. https://doi.org/10.1007/978-3-030-32022-5_13
- Rosenthal, S. S., & Ross, A. (2010). Violent crime, entrepreneurship, and cities. *Journal of Urban Economics*, 67(1), 135-149. <https://doi.org/10.1016/j.jue.2009.09.001>
- Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *science*, 277(5328), 918-924.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Shen, B., Hu, X., & Wu, H. (2020). Impacts of climate variations on crime rates in Beijing, China. *Science of The Total Environment*, 725, 138190. <https://doi.org/10.1016/j.scitotenv.2020.138190>
- Shepley, M., Sachs, N., Sadatsafavi, H., Fournier, C., & Peditto, K. (2019). The impact of green space on violent crime in urban environments: An evidence synthesis. *International Journal of Environmental Research and Public Health*, 16(24), Article 5119. <https://doi.org/10.3390/ijerph16245119>
- Sherman, L. W. (2013). The rise of evidence-based policing: Targeting, testing, and tracking. *Crime and Justice*, 42(1), 377-451.
- Silva, P., & Li, L. (2020). Urban crime occurrences in association with built environment characteristics: An African case with implications for urban design. *Sustainability*, 12(7), Article 3056. <https://doi.org/10.3390/su12073056>
- Singleton, C., Winata, F., Adams, A., McLafferty, S., Sheehan, K., & Zenk, S. (2022). County-level associations between food retailer availability and violent crime rate. *BMC Public Health*, 22, Article 1294. <https://doi.org/10.1186/s12889-022-14415-y>
- Smith, D. C., & Purtell, R. (2007). An empirical assessment of NYPD's "Operation Impact": A targeted zone crime reduction strategy. NYU Marron Institute.
- Sohn, D. (2016). Do all commercial land uses deteriorate neighborhood safety? Examining the relationship between commercial land-use mix and residential burglary. *Habitat International*, 55, 148-158. <https://doi.org/10.1016/j.habitatint.2016.03.007>
- Stevens, H. R., Beggs, P. J., Graham, P. L., et al. (2019). Hot and bothered? Associations between temperature and crime in Australia. *International Journal of Biometeorology*, 63, 747-762. <https://doi.org/10.1007/s00484-019-01689-y>
- Stucky, T. D., & Smith, S. L. (2017). Exploring the conditional effects of bus stops on crime. *Security Journal*, 30(1), 290-309.
- Su, N., Li, W., & Qiu, W. (2023). Measuring the associations between eye-level urban design quality and on-street crime density around New York subway entrances. *Habitat International*, 134, Article 102728. <https://doi.org/10.1016/j.habitatint.2022.102728>
- Tompson, L., Johnson, S., Ashby, M., Perkins, C., & Edwards, P. (2015). UK open source crime data: accuracy and possibilities for research. *Cartography and geographic information science*, 42(2), 97-111.

<법률>

- 112신고 공로자 포상 등에 관한 규정 (경찰청고시 제2024-9호).
- 112신고센터 운영 및 신고처리 규칙 (예규 제416호)
- 112신고센터 운영규칙 (예규 제408호)
- 112신고센터 운영규칙 (예규 제414호)
- 112신고의 운영 및 처리에 관한 법률 (법률 제19927호).
- 112신고의 운영 및 처리에 관한 법률 시행령 (대통령령 제34629호).
- 112종합상황실 운영 및 신고처리 규칙 (경찰청예규 제496호).
- 112종합상황실 운영 및 신고처리 규칙 (예규 제496호)
- 112종합상황실 운영 및 신고처리 규칙 (예규 제582호)
- 112치안종합상황실 운영 및 신고처리 규칙 (경찰청예규 제633호).
- 112치안종합상황실 운영 및 신고처리 규칙 (예규 제617호)
- 112치안종합상황실 운영 및 신고처리 규칙 (예규 제633호)
- 119긴급신고의 관리 및 운영에 관한 법률 (법률 제19928호).

Effects of Physical Environment on 112 Emergency Reporting

Ahn, Euisoon Kim, Youngwoo Kim, Eunhee Lim, Lisa Kim, Yujin Lee, Jounghyun

Chapter 1. Introduction

Fear of crime is a psychological phenomenon that refers to subjective anxiety and concern about becoming victims of crime. Fear of crime, in particular, reduces individuals' participation in social activities and acts as a factor that undermines the effective application of CPTED principles such as natural surveillance, access control, territorial reinforcement, and activity support, thereby exerting a negative influence on the creation of safe urban environments.

A 112 emergency reporting call involves a concrete action, a 112 call represents concern about crime that is manifested behaviorally. 112 call data can provide essential empirical evidence for establishing policies related to fear of crime.

The purpose of this study is to identify the relationship between physical elements of architectural and urban spaces and the occurrence of 112 calls.

This study proceeds in three stages in order to empirically analyze the impact of physical environments on 112 call occurrences: (1) analysis of the 112 call system and its utilization, (2) descriptive statistical analysis of 112 call data, and (3) empirical analysis of the effects of physical environmental factors.

The spatial scope of this study is set as the entirety of Seoul Metropolitan City, and the temporal scope of the study is set to the one-year period of 2023.

This study conducted a literature review covering laws and institutions, domestic and

international cases, academic research, and media reports. Prior studies and CPTED checklists were also extensively reviewed to inform the selection of physical environmental variables for analyzing their effects on 112 calls.

GIS-based spatial analysis were conducted at the 500 m grid level for physical characteristics of buildings and urban spaces as well as 112 call occurrences. To analyze the relationship between 112 call data and urban and architectural environmental factors, multiple spatial regression and machine learning models were examined. Ultimately, MGWR was adopted for spatial regression analysis, and XGBoost was selected as the primary machine learning prediction model.

Chapter 2. Legal Framework Governing the 112 Call System

With the enactment of the Act on the Operation and Processing of 112 Emergency Reports (“112 Emergency Reports Act” for short) in 2024, a more integrated and rapid response system was established compared to the previous framework that had been operated under National Police Agency regulations. The Act also introduced detailed operational systems such as strengthened protection of callers, enhanced information management, training of specialized personnel, and the construction of advanced information systems.

The 112 Emergency Reports Act mandates the establishment and operation of the 112 system and requires that records of call reception and handling entered into the system be stored and preserved for three years. When necessary for handling 112 calls, linkages with information systems of relevant agencies may be requested. In addition, statistics on 112 calls must be regularly analyzed and reflected in public safety activities, including the formulation of crime prevention measures. Information on 112 callers may be collected, used, or provided only for the purpose of handling the specific call or with the caller’s consent; use for any other purposes is prohibited.

The 112 call classification system consists of five levels, from Code 0 to Code 4, based on urgency and the need for on-site response. Code 0 represents the most urgent situations requiring real-time dissemination and immediate dispatch, whereas Code 4 refers to non-urgent civil complaints or consultation calls.

Emergency call data have been widely used both domestically and internationally as

evidence for establishing crime prevention policies. International cases, such as the CompStat system in the United States, demonstrate that data-driven crime prevention policies contribute significantly to crime deterrence and efficient police deployment. In Korea, the Pre-CAS system utilizes 112 call data to predict crime risk and support strategic patrol allocation.

However, while the 112 Emergency Reports Act strongly emphasizes the protection of callers' personal information and the principle of minimizing data use, it does not provide a clear legal basis for supplying data for research purposes. As a result, there are practical difficulties in actively utilizing 112 call data for crime prevention policy development and academic research.

Chapter 3. Descriptive Statistical Analysis of 112 Call Occurrence

The subject of analysis consisted of 112 call data recorded within the boundaries of Seoul Metropolitan City during the one-year period of 2023. To ensure anonymity and protect caller information, spatial data were aggregated at the grid level.

Call data were refined to include the date and time of occurrence, emergency codes (Codes 0-4), and grid-level location information. Outlier detection analysis confirmed the absence of significant anomalies, and the full dataset was used for analysis.

Analysis of Seoul's 112 call data by emergency code showed that the majority of calls fell under Code 2 (potential risk) and Code 1 (dispatch required). When examining temporal distributions by emergency code, distinct time-series patterns emerged.

Emergency calls corresponding to Codes 0 and 1, as well as Code 2, which is formally classified as non-emergency but requires immediate dispatch, showed similar patterns of concentration during weekends and nighttime hours. In contrast, non-emergency calls under Code 3 and civil consultation calls under Code 4 were concentrated during weekday working hours.

To focus on crime occurrence and fear of crime as captured through behavioral indicators of emergency reporting, this study set Codes 0 through 2 as the focus of the empirical analysis.

To further examine the nature of 112 call patterns, a spatial comparison was

conducted between 112 calls (Codes 0-2) and actual crime occurrences (the five major crimes) in four districts that are Gangnam, Seocho, Songpa, and Gangdong. The resulting correlation coefficient of 0.854 indicated a strong spatial correspondence between 112 call occurrences and crime occurrences.

Chapter 4. Empirical Analysis of the Effects of Physical Environments on 112 Call Occurrence

To analyze how physical characteristics of architectural and urban environments affect the occurrence of 112 calls, spatial data analysis using machine learning and spatial regression models was conducted. The analysis focused on 112 calls classified as Codes 0-2 that occurred in Seoul during 2023.

Physical environmental variables related to buildings and urban spaces were preprocessed at the grid level using the same 500-meter aggregation method applied to the 112 call data. Exploratory spatial data analysis was then conducted to visualize spatiotemporal patterns of 112 calls and to establish directions for subsequent analyses. Finally, machine learning and spatial regression methodologies were applied to identify key explanatory variables and to examine their effects on 112 call occurrence from multiple perspectives.

This study applied the same physical environmental factors used in the previous year's research, which analyzed the effects of physical environments on crime occurrence. These variables were derived from a wide range of domestic studies and CPTED checklists, and limited to those for which public data were available.

The final set of variables included factors related to population characteristics, building and land attributes, neighborhood convenience and commercial facilities, transportation and infrastructure, and road network structure.

Preprocessing variables involved scale adjustment, outlier handling, and removal of missing values. Multicollinearity and correlation analyses were then conducted. Variables with variance inflation factors (VIF) exceeding 15 and variable pairs with correlation coefficients above 0.7 were excluded. For pairs of area-based and corresponding ratio variables that showed near-perfect correlation, only the ratio variables were retained for ease of interpretation. In population-related variables,

strong correlations between total population and age-group variables led to the exclusion of total population. Variables such as road integration, sidewalk integration, sidewalk connectivity, road connectivity, and building age, which exhibited VIF values above 15, were excluded.

Global Moran's I was used to assess overall spatial autocorrelation and to confirm the necessity of spatial regression analysis. The Moran's I value of 0.485 indicated significant positive spatial autocorrelation.

Local Indicators of Spatial Association (LISA) were then applied to identify regional clusters. High-High hotspots, areas with high 112 call occurrences surrounded by similarly high areas, were particularly concentrated in urban centers and the Gangnam region.

The effects of physical environments on 112 call occurrence were examined upon hotspots compared with cold spots and background areas using the nonparametric Mann-Whitney U test. Comparisons between hotspots and cold spots showed strong statistical significance across all variables, confirming a significant relationship between physical environments and 112 call occurrence. Hotspots were characterized by higher population density, higher land prices, and greater building density than cold spots.

Further analyses compared hotspots within each district to district-level and citywide background areas. While most variables showed significant differences, some did not exhibit statistically significant mean differences.

For machine learning analysis, XGBoost and LightGBM were trained for predictive performance comparison. Both models showed high predictive accuracy but XGBoost exhibited a slightly higher test R^2 (0.85) than LightGBM (0.83) and was therefore selected as the final model.

To interpret the machine learning model, SHAP (Shapley Additive exPlanations) was applied. Global feature importance, measured by mean absolute SHAP values, indicated the relative contribution of each variable to prediction variability. The top ten variables were bars, restaurants and cafés, retail stores, public transportation facilities, number of buildings, proportion of commercial zones, convenience stores, proportion of green areas, officially assessed land value, and educational and public facilities.

SHAP dependence plots were used to analyze the relationship between variable values and changes in predicted outcomes. Bars, restaurants and cafés, retail stores, public transportation facilities, convenience stores, and land value showed increasing effects on 112 call occurrence, with relationships that were often nonlinear. Green area proportion and educational/public facilities showed decreasing effects on call occurrence. In particular, 112 calls were highest when green area proportions were very low, decreased until approximately 40% green coverage, and showed no further reduction beyond that threshold. Variables such as number of buildings and proportion of commercial zones exhibited mixed effects.

Using the top ten variables identified through machine learning analysis, multiple regression models—including multiple linear regression (MLR), geographically weighted regression (GWR), and multiscale geographically weighted regression (MGWR)—were applied and compared.

Both GWR and MGWR outperformed MLR in terms of model performance. While GWR exhibited the highest coefficients, MGWR showed similar coefficient levels with a lower Akaike Information Criterion corrected (AICc), indicating better overall model fit. Accordingly, MGWR was selected as the final spatial regression model.

Results showed that bars, restaurants and cafés, retail stores, public transportation facilities, convenience stores, green area proportion, and land value exhibited consistent relationships across Seoul. In contrast, educational and public facilities were significant only in certain areas, and variables such as number of buildings and proportion of commercial zones showed mixed effects, including positive, negative, or insignificant relationships depending on location.

Chapter 5. Conclusion

This study quantitatively analyzed the effects of physical environmental factors on 112 call occurrence using spatial statistical and machine learning techniques. We confirmed that 112 call occurrence is systematically linked to complex factors and explained its structure across space and time using data-driven methods.

Exploratory spatial data analysis using LISA identified hotspots and cold spots of 112 calls. Hotspots were concentrated in urban centers mostly in south of the Han River,

and showed significant differences from cold spots across most physical and social environmental variables. In particular, large differences were observed in population and building density, suggesting that areas with concentrated calls generally correspond to densely populated spaces.

Machine learning analysis identified the top ten variables influencing 112 call occurrence. Bars, convenience stores, public transportation facilities, retail stores, and number of buildings were associated with increased call occurrence, while green area proportion and educational/public facilities were associated with decreased call occurrence. Spatial regression analysis showed that most variables exhibited consistent relationships across Seoul, whereas educational/public facilities were significant only in certain areas, and variables such as number of buildings and commercial zoning showed complex, location-dependent effects.

This study is significant in that it extends the use of 112 calls as a safety indicator and empirically validates, using call data, existing claims regarding the influence of physical environments on crime occurrence and disorder. The finding that commercial facility-related variables, such as the number of bars, increase 112 call occurrence is consistent with prior research. In contrast, findings regarding green areas have been mixed in the literature, and this study newly demonstrates that green spaces in urban Seoul are associated with reduced 112 call occurrence.

Despite its contributions, the study has several limitations. First, due to the requirements of the 112 Call Handling Act, only grid-level aggregated statistics were used without distinguishing call types, limiting more granular analyses. Second, the grid-based approach prevented the inclusion of physical environmental variables that are not available at point or grid levels.

Future research should incorporate analyses across multiple spatial units—grid, administrative district, and national levels—and extend the analysis beyond Seoul to examine regional differences. Such follow-up studies could support CPTED policies and programs that are grounded in empirical evidence and yield substantive effectiveness across different regions.

Keywords :
112 Emergency Reporting, Fear of Crime, Crime Prevention Through Environmental Design (CPTED), Physical Environment Factors, Machine Learning

햇스팟 지역-자치구-서울시 물리환경 세부 수치

[부록] 햇스팟 지역-자치구-서울시 물리환경 세부 수치 (강남구~동작구)

변수	그룹	강남구	강동구	강북구	강서구	관악구	광진구	구로구	금천구	노원구	도봉구	동대문구	동작구
유아	서울	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9
	구	131.6	74.3	40.5	38.0	57.3	66.6	113.1	117.0	57.8	42.0	141.1	153.1
	HH	180.0	136.4	166.9	192.9	125.5	129.3	180.9	140.8	169.1	73.8	140.5	138.4
학생	서울	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1
	구	296.5	197.5	101.1	108.7	154.3	217.3	283.2	329.8	116.3	102.4	434.9	365.9
	HH	363.1	463.8	433.5	545.2	387.8	423.3	377.9	405.8	589.6	443.6	407.0	477.7
청년	서울	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7
	구	1104.1	872.9	517.9	433.9	707.4	1163.4	1091.8	1178.2	491.4	579.5	1280.1	1486.6
	HH	1379.6	1493.3	1925.3	1399.2	1373.6	1716.5	1797.1	1484.6	1939.3	2174.6	1188.1	1778.6
중장년	서울	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3
	구	1252.2	754.9	505.8	462.7	784.8	1039.3	1251.6	1155.9	550.5	390.0	1701.5	1348.0
	HH	1558.3	1263.9	1427.9	1196.6	1648.9	1591.3	1650.8	1463.2	1497.0	936.8	1655.3	1872.0
고령	서울	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3
	구	1134.9	673.7	600.0	455.4	742.6	885.5	1023.9	1050.7	460.4	427.4	1324.5	1192.5
	HH	1254.2	1221.3	1698.2	1404.2	1546.9	1142.4	1357.9	1252.2	1175.7	1655.6	1205.4	1323.1
공시지가	서울	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3
	구	8871865.1	1966676.2	761341.1	908475.6	1660543.2	2917697.8	2164991.3	2453040.1	711149.2	577633.1	4032011.4	3701970.3
	HH	15983488.0	4940002.3	3111279.4	3747105.3	4319911.2	5234199.1	3869947.7	3497797.1	4234303.8	3021993.9	4169434.9	4394108.0

용적률	서울	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7
	구	113.0	55.9	26.4	28.1	42.6	61.9	65.5	78.9	12.4	15.8	66.3	90.7
	HH	196.5	128.7	96.5	133.6	113.6	108.4	131.5	115.7	83.8	97.2	65.6	104.2
건폐율	서울	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1
	구	27.6	19.8	9.2	9.9	14.0	19.9	18.4	24.0	6.8	6.5	22.4	29.9
	HH	44.5	39.4	35.0	39.5	34.1	36.6	32.9	35.0	22.3	32.6	22.4	35.0
건물 높이	서울	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
	구	9.2	5.9	2.4	4.8	6.1	6.2	7.0	10.0	2.8	2.0	5.3	6.7
	HH	14.8	9.5	7.0	10.2	8.3	8.2	10.6	9.7	9.3	8.0	4.8	7.0
건물 개수	서울	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1
	구	191.5	207.1	161.6	99.7	227.3	304.6	208.4	389.6	33.1	66.4	578.0	463.3
	HH	362.3	659.0	787.4	548.3	642.7	647.3	494.1	576.5	228.0	527.6	653.6	632.6
주거 비율	서울	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7
	구	16.6	9.2	6.8	7.9	25.7	16.0	17.2	15.5	13.2	7.4	24.0	23.9
	HH	22.6	24.6	27.9	36.7	26.2	26.6	19.5	19.5	44.0	29.3	19.7	31.6
상업 비율	서울	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4
	구	13.7	4.8	3.3	3.0	4.0	7.5	5.8	10.6	0.9	1.5	16.6	7.8
	HH	25.8	13.7	13.8	13.8	11.8	14.1	13.9	16.4	10.1	11.9	19.9	10.8
주상 복합 비율	서울	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8
	구	3.1	5.5	2.2	1.2	2.8	4.6	3.2	5.7	0.4	1.2	10.4	5.7
	HH	5.9	17.6	10.3	7.2	8.2	10.0	8.2	9.0	4.7	9.6	12.4	7.9
농지 비율	서울	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8
	구	4.2	5.1	0.2	10.9	0.2	0.0	1.8	0.1	0.9	0.8	0.0	0.0
	HH	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
나지 비율	서울	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	구	0.5	0.5	0.2	0.2	0.2	0.2	0.7	1.3	0.2	0.1	1.7	0.5
	HH	0.8	1.5	0.5	0.4	0.6	0.4	0.6	1.9	0.5	0.5	2.2	0.7
공업 비율	서울	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	구	0.3	0.0	0.0	0.2	0.6	1.8	6.0	8.3	0.0	0.2	0.0	0.0
	HH	0.0	0.0	0.0	0.4	1.7	4.3	13.6	9.1	0.0	0.1	0.0	0.0
녹지 비율	서울	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2
	구	35.3	17.5	62.9	27.9	44.5	20.4	22.7	22.2	59.6	65.5	9.7	22.2
	HH	7.4	2.6	5.9	6.2	8.9	6.2	4.5	6.0	12.9	7.6	7.3	7.9

공원 비율	서울	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
	구	20.1	7.7	1.7	3.9	1.8	1.0	6.7	0.8	0.6	0.3	3.2	16.6
	HH	2.8	1.7	3.1	5.8	3.2	1.5	1.0	1.1	2.3	1.6	2.9	4.2
수변 비율	서울	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6
	구	1.4	10.6	0.3	11.7	0.3	12.8	0.6	0.0	0.6	0.8	3.0	3.0
	HH	0.0	0.0	0.4	0.0	0.4	1.4	0.8	0.0	2.8	2.7	1.7	0.3
술집 수	서울	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7
	구	13.0	8.4	5.1	3.0	8.3	9.8	11.5	14.6	1.0	0.8	9.6	6.8
	HH	25.5	27.8	26.3	18.4	24.4	22.7	30.6	23.7	11.7	6.4	11.7	9.4
음식점· 카페 수	서울	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1
	구	99.6	37.6	22.8	13.8	39.6	56.4	56.1	68.5	7.5	8.4	72.5	69.2
	HH	193.0	119.1	112.6	81.7	116.9	129.0	141.4	107.9	80.9	65.6	84.1	96.6
숙박 시설	서울	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
	구	4.4	2.2	1.6	0.8	2.3	3.4	4.0	4.0	0.2	0.4	8.1	2.6
	HH	8.6	7.4	7.8	4.8	6.8	7.9	10.7	6.3	2.4	2.8	9.9	3.7
편의점 수	서울	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8
	구	7.7	3.4	1.9	1.4	3.6	4.9	5.2	6.6	0.6	1.0	6.2	5.7
	HH	14.6	10.7	9.6	8.0	10.5	11.1	12.3	9.9	5.6	6.4	7.1	7.8
병·의원 시설 수	서울	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6
	구	24.6	5.4	2.1	1.5	3.5	4.1	4.7	6.0	0.9	0.6	9.7	6.4
	HH	48.2	16.3	10.9	8.9	10.4	9.5	11.8	9.7	11.0	4.6	12.2	8.9
공공· 교육 시설 수	서울	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7
	구	54.5	14.1	6.5	6.8	14.6	14.2	17.8	20.3	4.7	3.7	23.3	30.8
	HH	103.0	40.9	33.0	39.3	41.3	30.7	37.4	30.7	46.1	22.6	23.8	42.8
세탁소 수	서울	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
	구	2.6	1.5	0.7	0.8	2.1	2.2	1.3	2.9	0.2	0.4	2.7	2.8
	HH	4.7	4.8	3.8	4.8	6.0	4.9	3.2	4.4	2.1	2.4	2.9	3.9
교통 시설 수	서울	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7
	구	4.9	3.1	2.1	1.7	3.4	3.4	5.5	6.0	1.1	1.2	6.9	7.1
	HH	7.9	8.4	9.1	7.8	7.9	7.1	10.4	8.4	8.4	7.4	7.1	7.8
소매 시설 수	서울	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3
	구	67.7	36.9	17.5	12.7	28.3	39.9	55.1	72.9	4.7	5.8	99.2	51.4
	HH	130.2	117.4	89.3	76.0	83.2	87.8	140.4	114.6	47.7	44.0	120.5	71.6

주차장 수	서울	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	구	0.0	0.2	0.2	0.2	0.2	0.2	0.6	0.7	0.0	0.2	0.7	0.6
	HH	0.0	0.6	1.1	1.3	0.6	0.4	1.3	0.9	0.3	0.8	0.8	0.7
CCTV 수	서울	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5	14.5
	구	17.8	11.3	14.3	7.5	12.0	11.8	23.8	22.5	4.9	4.8	18.8	24.1
	HH	31.3	34.5	65.3	41.8	32.5	23.9	52.3	32.3	37.6	39.2	20.4	32.3
경찰서 수	서울	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	구	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.1	0.2	0.1
	HH	0.2	0.0	0.3	0.1	0.2	0.3	0.3	0.2	0.1	0.8	0.2	0.2
주유소 수	서울	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	구	0.3	0.2	0.1	0.2	0.2	0.3	0.3	0.4	0.1	0.2	0.8	0.4
	HH	0.5	0.5	0.8	0.6	0.5	0.5	0.6	0.6	0.6	1.2	0.8	0.6
도로 폭	서울	7.1	7.1	7.1	7.1	7.1	7.1	7.1	7.1	7.1	7.1	7.1	7.1
	구	8.5	9.6	3.2	7.6	5.5	6.2	7.7	9.2	6.1	3.5	8.4	8.0
	HH	9.8	7.8	7.0	9.0	8.1	8.3	9.7	8.8	13.3	9.7	8.2	7.6

출처: 연구진 작성

[부록] 핫스팟 지역-자치구-서울시 물리환경 세부 수치 (마포구~중랑구)

변수	그룹	마포구	서대문구	서초구	성동구	성북구	송파구	양천구	영등포구	용산구	은평구	종로구	중구	중랑구
유아	서울	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9	85.9
	구	103.6	118.0	68.1	222.8	49.3	88.1	151.8	137.9	111.3	88.6	100.1	146.6	146.1
	HH	170.1	129.7	190.7	195.3	167.2	157.2	183.6	160.9	112.8	170.5	158.8	128.8	210.3
학생	서울	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1	207.1
	구	217.1	229.5	175.7	492.1	184.9	327.2	490.1	332.6	262.9	205.1	147.7	351.0	261.7
	HH	337.5	299.8	551.2	484.1	674.8	465.2	645.3	384.1	342.4	485.2	218.0	340.0	463.5
청년	서울	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7	835.7
	구	888.3	1171.8	764.4	1550.7	573.7	959.8	1106.0	1353.6	1309.1	690.8	783.1	1514.2	1089.3
	HH	1499.2	1509.7	1654.4	1514.8	1469.7	1327.2	1255.1	1567.6	1813.0	1253.8	1242.6	1481.7	1525.3
중장년	서울	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3	887.3
	구	958.3	1474.2	783.3	1585.6	615.9	779.7	1025.4	1605.6	1221.6	916.6	933.8	1259.0	1193.1
	HH	1503.8	1849.2	1730.2	1681.5	1876.3	1209.0	1146.8	1679.7	1435.9	1658.7	1407.0	1226.4	1737.6
고령	서울	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3	788.3
	구	909.1	1005.5	661.0	1288.1	595.4	943.5	1206.4	1214.6	990.5	783.1	742.6	1289.9	907.3
	HH	1408.6	1163.4	1180.9	1187.0	1309.7	1089.6	1353.3	1373.0	1268.9	1176.8	1137.6	1245.7	1379.0
공시지가	서울	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3	2846264.3
	구	3526830.7	3960525.1	2936239.5	4876743.3	1193755.9	5028616.6	2345718.1	3971548.2	5691334.9	1284497.6	3573870.1	11308268.2	1531020.0
	HH	7066274.0	6588987.1	15942106.0	4947073.2	4292403.1	7842560.3	3515421.7	4452778.1	7308618.7	3654116.5	11321607.5	12055271.5	3254775.5
용적률	서울	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7	51.7
	구	54.4	74.3	44.6	85.4	18.9	87.1	97.1	82.1	65.4	42.0	45.7	118.3	51.8
	HH	112.8	114.0	189.3	87.9	68.9	129.9	124.0	94.4	92.3	114.7	128.5	123.2	108.8
건폐율	서울	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1	18.1
	구	17.0	41.7	14.1	28.0	8.2	22.4	39.9	24.2	23.6	15.5	26.9	47.6	20.3
	HH	34.8	69.3	39.1	29.6	24.3	35.3	52.2	27.9	33.4	33.4	82.8	50.0	39.2
건물높이	서울	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
	구	7.7	7.2	5.3	7.0	3.2	12.4	9.0	7.7	5.9	6.6	4.4	8.7	3.5
	HH	8.8	9.0	16.6	6.8	7.7	16.5	8.8	8.1	6.7	7.8	8.8	9.0	5.6

건물 개수	서울	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1	197.1
	구	244.7	286.0	60.5	528.2	119.6	249.3	322.6	431.4	439.9	173.0	149.7	398.7	335.5
	HH	558.7	510.8	315.3	597.8	548.0	453.8	531.1	508.9	697.4	721.0	374.4	416.6	780.1
주거 비율	서울	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7	13.7
	구	17.0	19.8	9.9	27.4	10.2	19.0	25.5	18.2	19.2	11.8	8.6	13.1	12.3
	HH	30.2	25.2	30.0	23.6	40.9	24.8	34.4	21.6	22.1	35.7	6.0	11.1	23.5
상업 비율	서울	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4
	구	5.4	8.8	3.5	15.0	1.8	9.9	7.7	13.9	9.1	2.8	9.2	29.9	5.8
	HH	12.6	17.4	20.7	17.2	11.8	17.9	12.6	16.1	13.9	11.5	33.7	32.3	13.5
주상 복합 비율	서울	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8
	구	4.9	3.5	0.9	8.6	1.2	5.1	5.2	6.3	2.9	2.2	1.8	3.7	4.0
	HH	10.7	7.0	5.9	10.4	5.9	9.4	8.6	7.5	4.1	9.0	5.8	3.8	9.1
농지 비율	서울	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8
	구	0.9	0.2	8.4	0.0	0.1	2.5	2.7	0.0	0.0	2.0	0.5	0.0	0.3
	HH	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
나지 비율	서울	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	구	0.7	1.1	0.8	0.9	0.7	0.5	0.3	1.1	0.8	0.3	2.3	1.5	0.3
	HH	1.3	1.2	1.6	0.9	1.5	0.9	0.5	1.3	1.3	0.4	2.8	1.6	0.7
공업 비율	서울	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	구	0.0	0.0	0.2	5.4	0.0	0.0	0.0	1.9	0.0	0.0	0.0	0.0	0.2
	HH	0.0	0.0	0.0	3.8	0.0	0.0	0.0	2.3	0.0	0.0	0.0	0.0	0.0
녹지 비율	서울	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2	39.2
	구	22.2	41.1	56.4	9.9	70.2	11.7	19.4	7.9	29.4	39.1	56.3	11.3	26.6
	HH	11.9	9.0	11.7	7.7	14.6	8.9	3.2	7.9	11.1	3.2	8.6	7.5	4.1
공원 비율	서울	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
	구	7.0	2.1	13.3	2.6	1.6	8.8	2.7	3.5	6.9	6.0	2.3	4.7	3.5
	HH	3.4	4.2	2.5	2.1	6.9	11.2	3.8	2.7	3.0	1.5	4.7	4.3	1.2
수변 비율	서울	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6
	구	17.0	0.2	0.4	1.7	0.3	21.7	0.0	12.0	5.3	0.5	0.3	0.4	0.3
	HH	0.4	0.4	0.0	1.6	0.2	4.5	0.0	0.7	0.0	0.6	0.6	0.4	0.7

술집 수	서울	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7	7.7
	구	13.7	15.0	3.3	17.0	1.5	11.0	14.1	15.4	12.8	4.4	8.4	22.4	9.2
	HH	34.9	30.1	24.5	20.8	9.0	20.2	23.4	18.1	20.2	19.9	32.8	24.1	21.4
음식점·카페 수	서울	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1	46.1
	구	78.1	89.4	24.3	114.9	12.8	79.9	46.8	84.2	71.4	25.7	51.9	167.3	44.7
	HH	194.1	177.9	169.3	129.8	79.2	142.2	77.2	99.4	114.6	110.2	199.9	180.4	103.1
숙박 시설	서울	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
	구	8.7	10.8	1.0	8.1	1.3	4.5	3.5	7.2	10.2	1.1	4.3	18.6	2.4
	HH	21.3	21.3	7.1	9.9	8.5	8.3	5.8	8.6	15.3	4.0	16.3	19.7	5.6
편의점 수	서울	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8
	구	5.3	6.9	1.9	9.4	1.2	6.8	4.7	6.8	4.7	2.4	3.2	13.0	3.9
	HH	12.3	13.3	12.5	10.6	5.8	11.7	7.6	8.0	7.2	10.3	12.1	13.9	8.9
병·의원 시설 수	서울	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6	5.6
	구	4.5	4.5	8.1	8.5	1.3	10.0	2.8	7.8	3.4	2.3	3.7	14.3	4.5
	HH	10.9	8.8	58.8	10.2	9.2	17.5	4.5	9.2	5.4	10.5	14.7	15.5	10.4
공공·교육 시설 수	서울	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7	16.7
	구	28.0	25.0	15.7	29.8	6.5	28.2	29.3	22.9	15.2	11.9	10.4	26.6	14.1
	HH	63.8	48.8	102.2	31.9	37.0	49.2	46.6	27.1	22.1	51.3	33.3	27.7	31.1
세탁소 수	서울	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
	구	1.5	1.8	0.7	3.9	0.4	2.5	2.7	2.4	2.0	1.1	0.6	2.3	2.0
	HH	3.0	3.4	4.1	4.3	2.2	4.2	4.3	2.9	3.0	4.9	2.0	2.3	4.6
교통 시설 수	서울	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7
	구	6.0	5.5	2.2	7.5	1.7	5.1	3.4	7.9	6.3	2.8	4.1	11.9	4.3
	HH	11.6	9.7	10.1	7.6	7.5	8.5	3.9	9.3	8.7	9.6	12.5	12.6	9.2
소매 시설 수	서울	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3	41.3
	구	41.9	46.4	17.5	92.7	14.0	75.7	43.5	85.5	51.4	19.6	68.1	239.8	37.6
	HH	101.8	91.1	119.3	106.3	92.2	136.7	71.8	101.0	81.8	83.7	265.5	260.0	86.0
주차장 수	서울	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	구	0.0	0.4	0.0	0.3	0.1	0.3	0.9	0.5	0.4	0.1	0.7	1.9	0.3
	HH	0.0	0.7	0.2	0.4	0.5	0.5	1.4	0.6	0.6	0.4	3.0	2.0	0.6

