

서울의 밤, 여성의 안전 : 데이터 기반 인프라 개선 시급도 분석

팀명: 지금, 잠이 옵니까?

by. DS 최유희, DS홍현경, DS 김효진, DE 김재현

여성 안전 어디가 가장 시급할까요?

서울의 밤길, 여성에게 이 도시는 안전한 지역입니까?
체감 안전도와 현실의 괴리



문제 정의: 안전 사각지대와 잠재적 고위험 지역

체감 안전도와 실제 위험 사이에는 종종 큰 차이가 있습니다.

우리 프로젝트는 이런 괴리를 데이터로 밝히고,

여성 안전을 위해 어디에, 어떤 대응이 시급한지 구체적으로 보여주기 위해 시작되었습니다.

불균형한 안전 인프라

- 범죄 예측
단순 통계가 아닌, 언제 어디서 위험이 높아지는지 예측하는 기술 필요성 증대
- 사각지대
CCTV, 안심벨, 가로등 등 방어 인프라가 부족한 지역에 대한 선제적 탐지

여성 노인 범죄 취약성

- 범죄에 더욱 취약한 여성 독거노인 등 특정 계층의 거주 특성 고려
- 잠재적 고위험 지역을 미리 찾아 경찰/지자체의 선제적 개입을 유도



Seoul Risk Analytics

for Women & Aged
(서울 여성 및 노인 위험 분석)

여성 안심 안전 지수 도출 서비스

서울시 25개 지역구의
인프라 개선 시급도 지수를 과학적으로 도출합니다

여성 안전, 어디가 가장 취약할까? 인프라 개선 시급도 지수 도출



분석 목표

서울시 25개 지역구의 '[인프라 개선 시급도 지수](#)'
를 도출합니다.

"지금 당장 CCTV를 더 설치하고 경찰 인력을 늘려야 하는 지역"을 과학적으로 찾아내는 거예요.



핵심 모델

지역별 **취약성(위험)** 요인과 **방어력(안전)** 요인
을 종합하여 지수를 산출합니다.

위험한 정도(취약성)에서 안전한 정도(방어력)를 빼서,
지금 얼마나 위험한 상태인지를 점수로 나타냅니다.



최종 활용

지자체 및 경찰의 **치안 인프라(CCTV, 안심벨, 순찰)** 투자 우선순위 결정에 활용됩니다.

한정된 예산을 가장 필요한 곳에 쓰는, '[효율적인 안전 예산 집행](#)'의 근거가 됩니다.

취약성(X)

성범죄, 성폭력 위험, 여성 독거노인

방어력(Y)

CCTV, 안심벨, 경찰 인프라

시급도 지수

인프라 개선 우선순위

1단계: 여성 안전을 결정하는 두 개의 축 위험 요소 vs 안전 시설

취약성(X)과 방어력(Y)이 여성 안전에 미치는 영향 분석

 **취약성 (X)**

여성 안전에 위협이 되는 요소들로, 높을수록 위험 합니다.

-  **성범죄 건수**
성범죄 발생이 많을수록 치안에 취약함
-  **성폭력 위험 지수**
잠재적인 위험 요소가 높을수록 취약함
-  **여성 독거노인**
스스로 방어하기 어려운 인구가 많을수록 취약함

 **방어력 (Y)**

-  **CCTV 설치 수**
감시 및 증거 확보를 통한 범죄 예방력
-  **안심벨 설치 수**
비상 상황 시 즉각적인 도움 요청 가능성
-  **경찰 인프라**
즉각적인 출동 및 상시 순찰 능력

 도시의 안전도 위험 요소(X)는 줄이고,
방어 시설(Y)은 늘려야 한다는 논리를 수치로 분석하는 모델

데이터 준비: 정확한 분석을 위한 재료 모으기!

데이터 통합

각 지역구별 11개 지표 데이터를 '지역구' 기준으로 하나의 표 (DataFrame)로 합쳤습니다.

```
</> pd.merge(..., how='left')
```

여러 데이터 파일을 '지역구'라는 공통 키로 정확하게 연결합니다.

최신성 확보

연도별 데이터 중 '[가장 최신 연도](#)'의 데이터를 선택하여 현재 시점의 '시급도'를 평가하는 데 사용했습니다.

▣ 시급도를 판단하는 것은 '현재의 상태'가 중요하므로,
시점이 통일되지 않은 데이터는 [가장 최신 데이터](#)로 기준을 맞춥니다.

데이터 보완

데이터가 없는 지역구는 [결측치\(NaN\)](#)를 0으로 처리했습니다.

- ➊ 데이터가 없는 경우 실제로 인프라가 없는 것일 가능성성이 높으므로, 0으로 채워 분석 오류를 방지했습니다

데이터 전처리 흐름



성범죄
데이터



성폭력 위험



인구 통계



CCTV 설치



안심벨 설치

▣ 통합 데이터프레임

지역구 기준 11개 지표 포함

✓ 최신 데이터 & 결측치 처리

모든 지역구에 동일한 시점의 데이터 확보

위험 등급 매기기: '시급도 지수'의 탄생 공식

1. 정규화 (Normalization)

수치가 큰 변수와 작은 변수의 영향력을 동일하게 조정합니다.

- Min-Max Scaling을 적용하여 모든 데이터를 0과 1 사이의 값으로 변환합니다.

2. 종합 점수 산출

정규화된 각 지표를 합산하여 점수를 계산합니다.

- 취약성 점수 (Vulnerability_Score) : 높을수록 위험
- 방어력 점수 (Defense_Score) : 높을수록 안전

3. 최종 시급도 계산

마지막으로 인프라 개선 시급도 지수를 도출합니다.

시급도 지수 공식

취약성 점수
(X)

방어력 점수
(Y)

시급도
지수

해석 방법

- 시급도 지수가 높을수록 : 취약성(위험)은 높은데 방어력(시설)이 부족하다는 뜻입니다.
- 인프라 개선이 가장 시급한 지역으로 판단됩니다.
- 시급도 지수는 절대적 순위 보다는 상대적 비교에 초점을 맞춥니다.

이 지수는 지역 간 비교를 통해 자원 배분의 우선순위를 정하는데 도움을 줍니다

우리 지수는 믿을 수 있을까?

검증: '최종 시급도 지수' vs '외부 위험 지표' 상관분석

▣ 검증 결과

112 신고 건수 (Calls_112)

상관계수 **0.618** (양의 관계)

'최종 시급도 지수'가 높을수록 실제 112 신고 건수(Calls_112)가 높은 경향을 보였습니다.

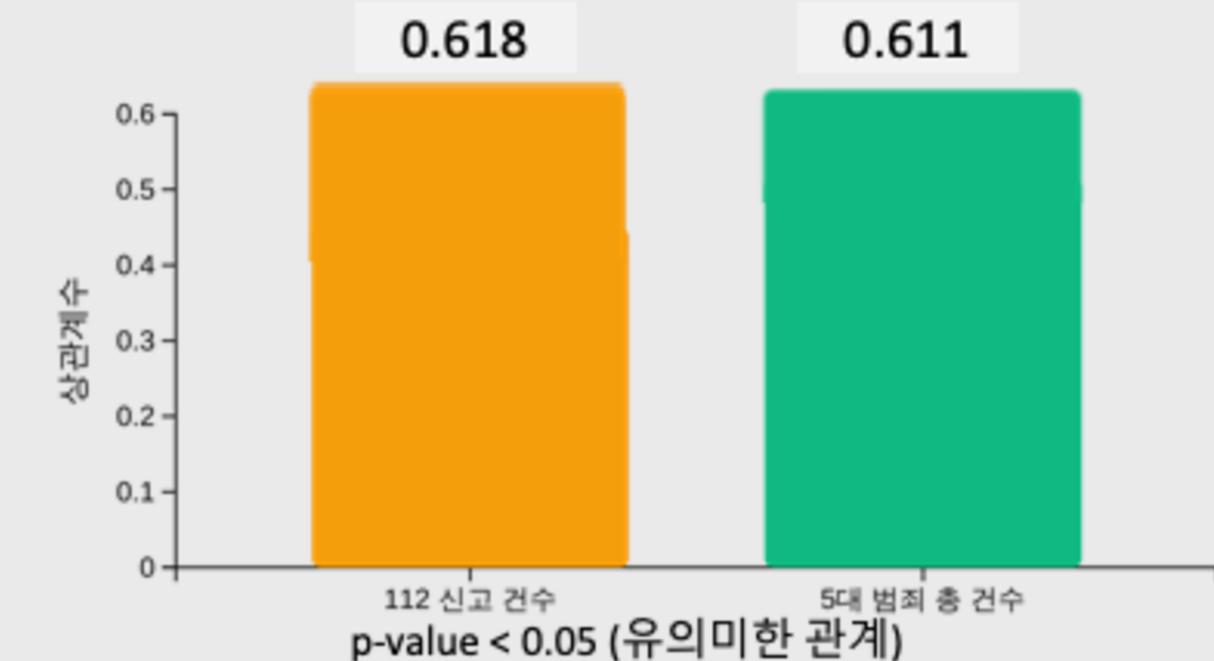
5대 범죄 총 건수 (Total_Crime_5)

상관계수 **0.611** (뚜렷한 양의 관계)

'최종 시급도 지수'가 높을수록 5대 범죄 총 건수(Total_Crime_5)가 높은 경향을 보였습니다.

(두 지표 모두 통계적으로 유의미한 양의 관계, $p<0.05$)

▣ 상관계수 시각화



▣ 최종 결론

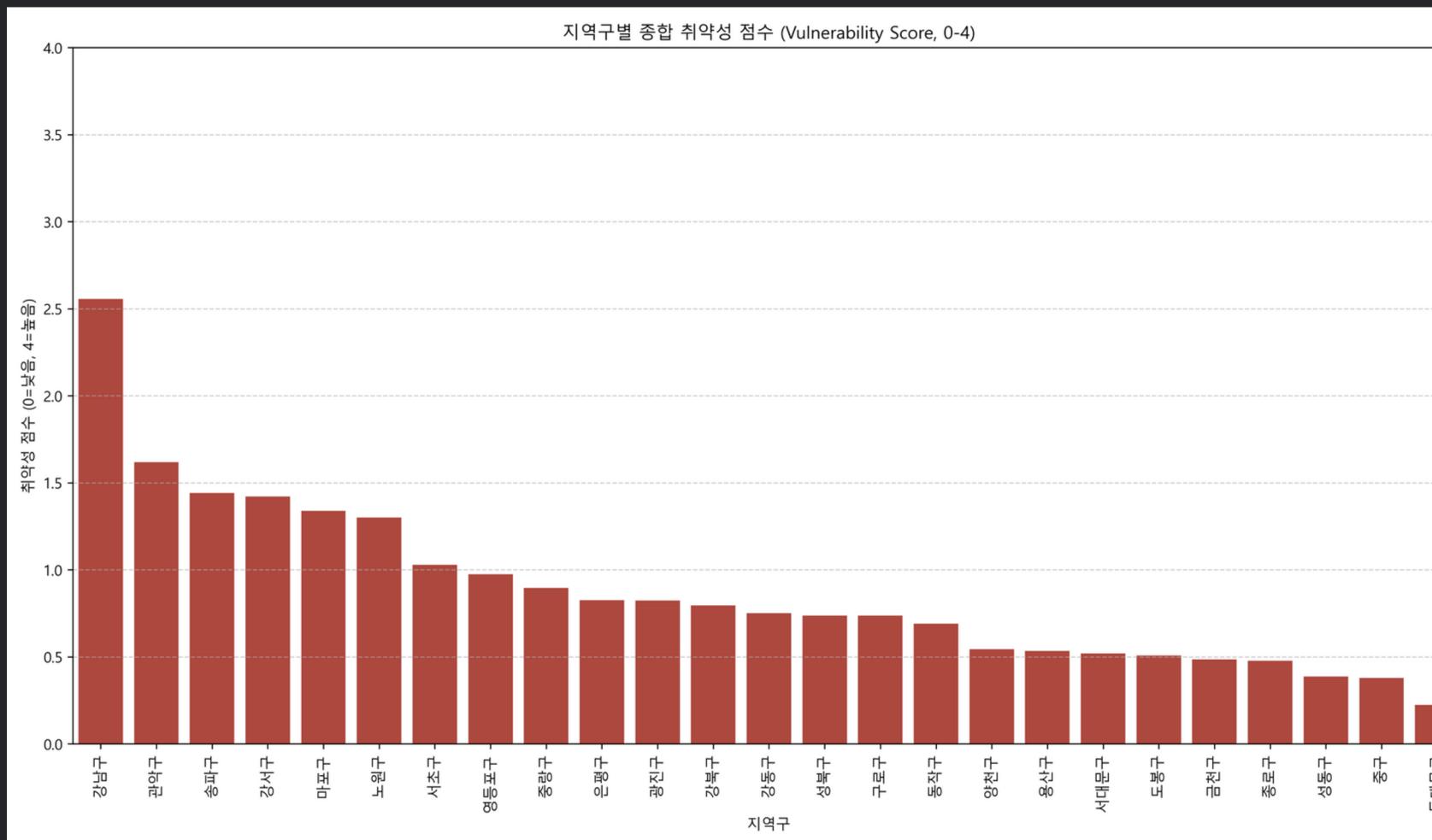
'최종 시급도 지수'는 실제 치안 현실(범죄 발생, 112 신고 건수)과 통계적으로 유의미한 양의 상관관계를 보였습니다($p<0.05$).

즉, 시급도가 높을수록 외부 위험 지표도 높아지는 경향이 확인되었습니다.

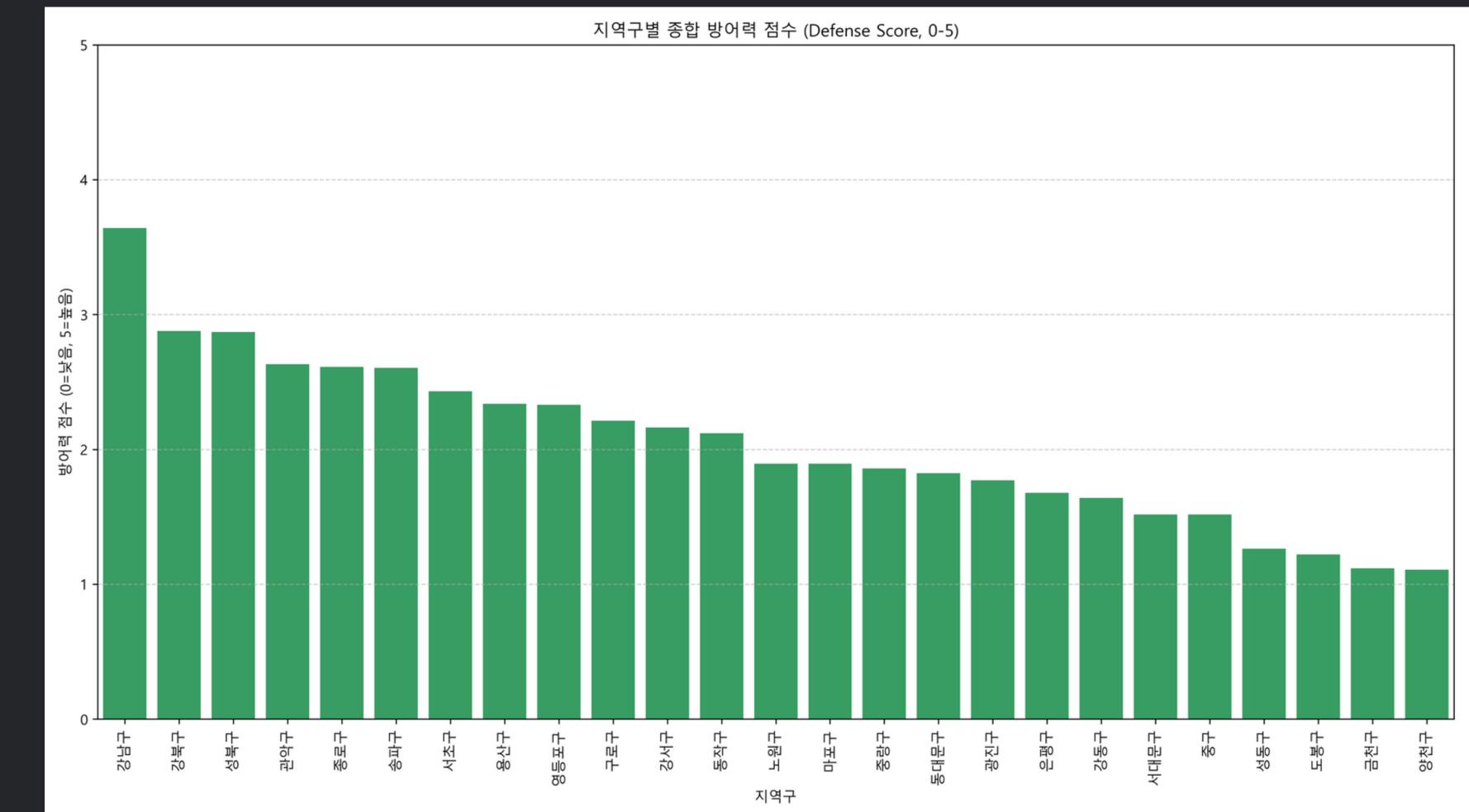
💡 검증 결과, '최종 시급도 지수'가 높을수록
실제 외부 위험 지표도 높아지는 경향을 보임

2단계 : 개선 시급도 지수 산출

- 취약성(X): 강남구가 가장 높고, 도봉구가 가장 낮음 → 위험 요인 분포의 양극화 뚜렷
- 방어력(Y): 강남구가 가장 높고, 강북구가 가장 낮음 → 안전 인프라의 지역별 불균형 존재



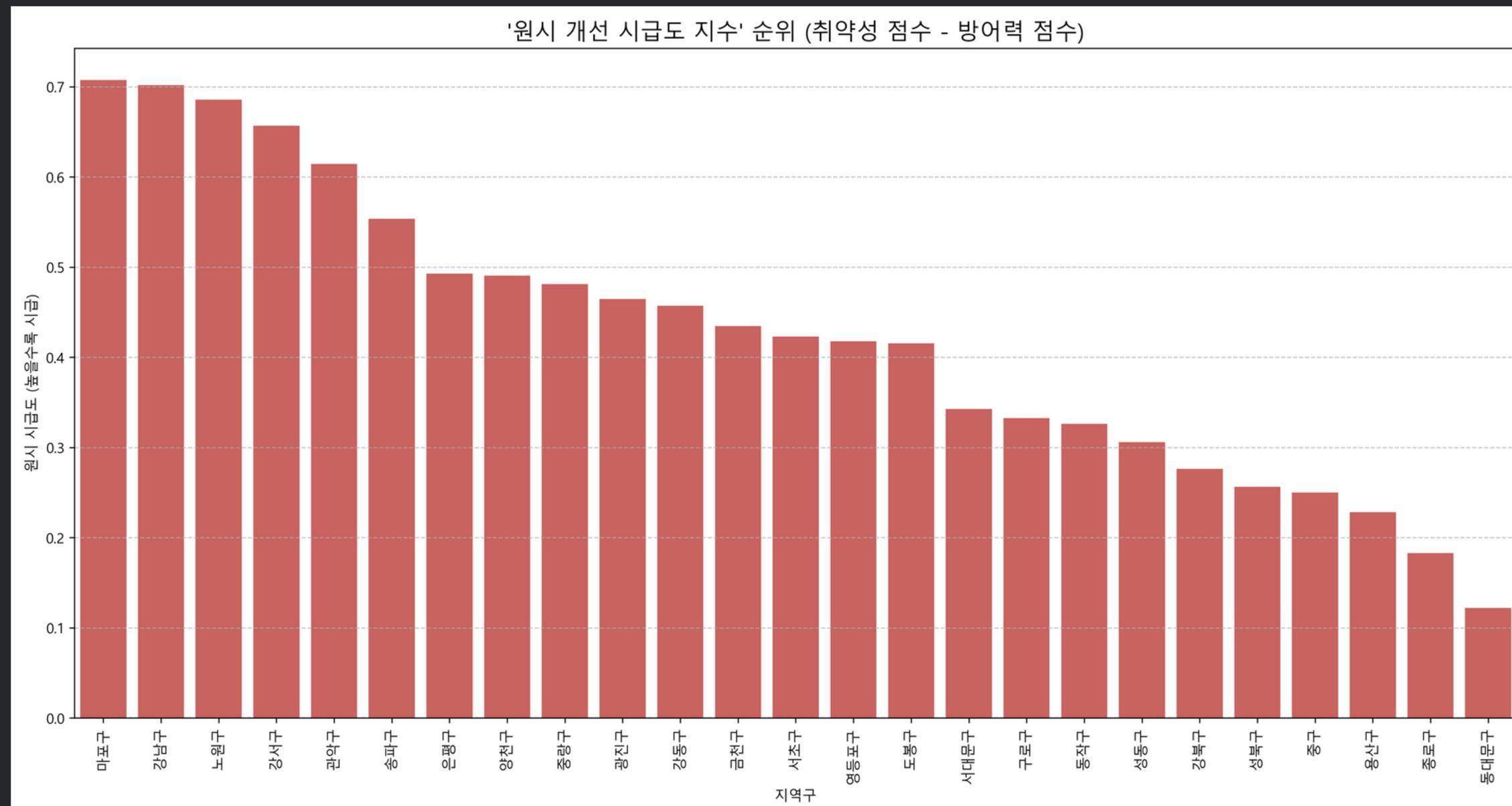
취약성(X) 개별 그래프



방어력(Y) 개별 그래프

3단계 : 개선 시급도 지수 산출

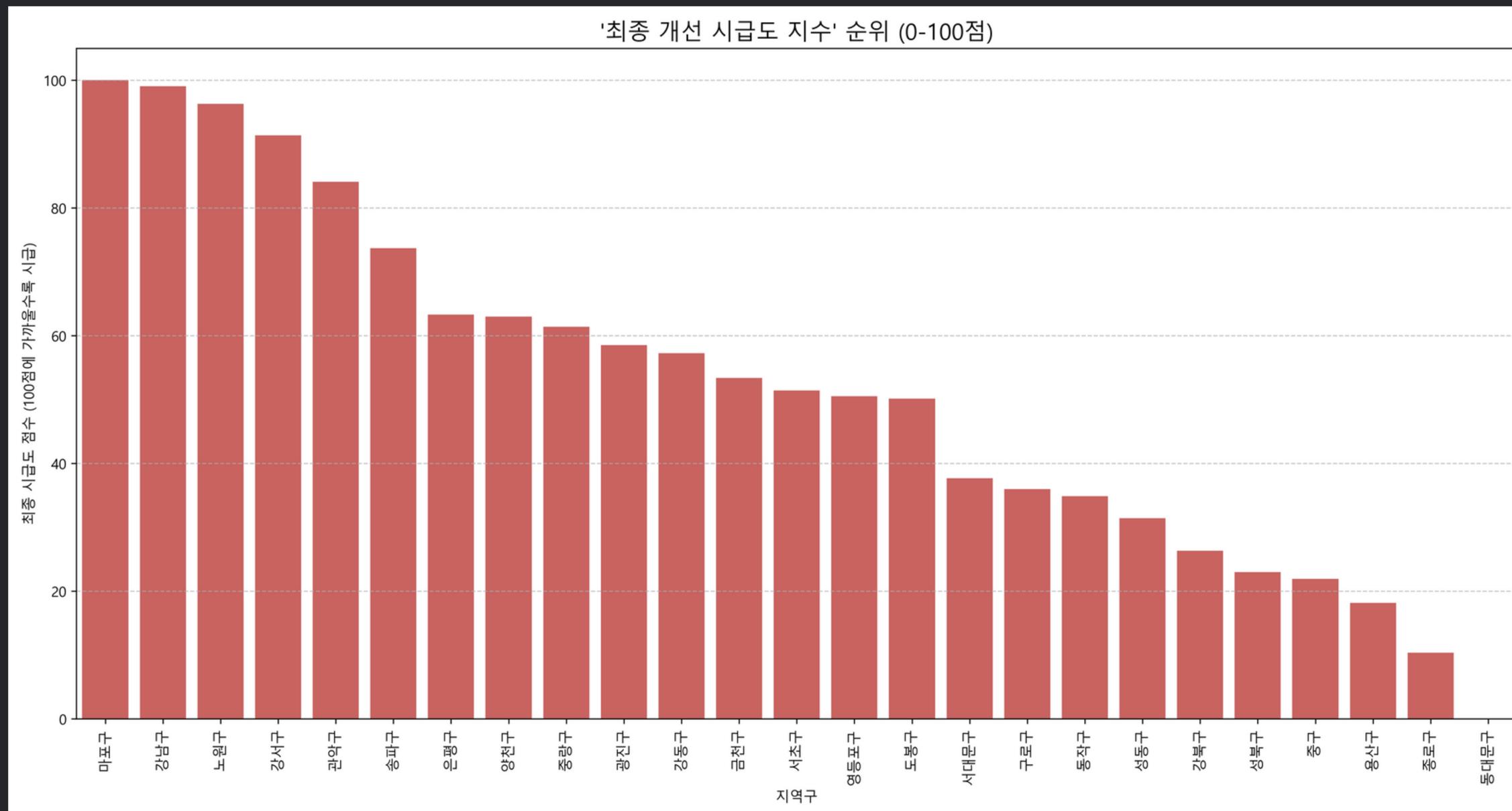
- 취약성(X): 지역의 위험 노출 정도
- 방어력(Y): 위험을 완화하는 인프라 수준
- 개선 시급도 지수($Z = X / Y$): 방어력 대비 취약성이 높을수록 인프라 개선이 시급한 지역



- 최우선 개선 지역 (상위권) :
→ 취약성 높고 방어력 낮아 인프라 보강 시급
- 시급도 낮은 지역 (하위권) :
→ 취약성 낮고 방어력 높아 안전 인프라 작동
→ 취약성, 방어력 함께 높아 위험 대비 대응 효율 양호

4단계 : 최종 결과 확인 (0~100점 스케일링)

- 3단계 지수를 0~100점으로 Min-Max 스케일링하여 시급도가 직관적으로 표현되도록 변환
- 가장 시급한 지역은 100점, 가장 양호한 지역은 0점으로 설정해 막대 차트로 시각화



- 주요 결과
→ 마포구, 강남구, 노원구순으로 시급도가 가장 높게 나타남
- 결과 해석
→ 인프라 투자 불균형이 뚜렷
→ 지도를 통해 개선 우선순위를 한눈에 파악 가능

5단계 : 상관관계 시각화 (Heatmap)

목표: 시급도(Z) 모델의 타당성을 검증하기 위한 변수 간 상관관계 시각화

- 취약성(X) 요인 간의 관계 (모델 구성의 타당성)
→ 위험 요인 간 양(+)의 상관관계 예상
- 방어력(Y) 요인 간의 관계 (인프라 투입의 연관성)
→ 안전 인프라 간 양(+)의 상관관계 예상

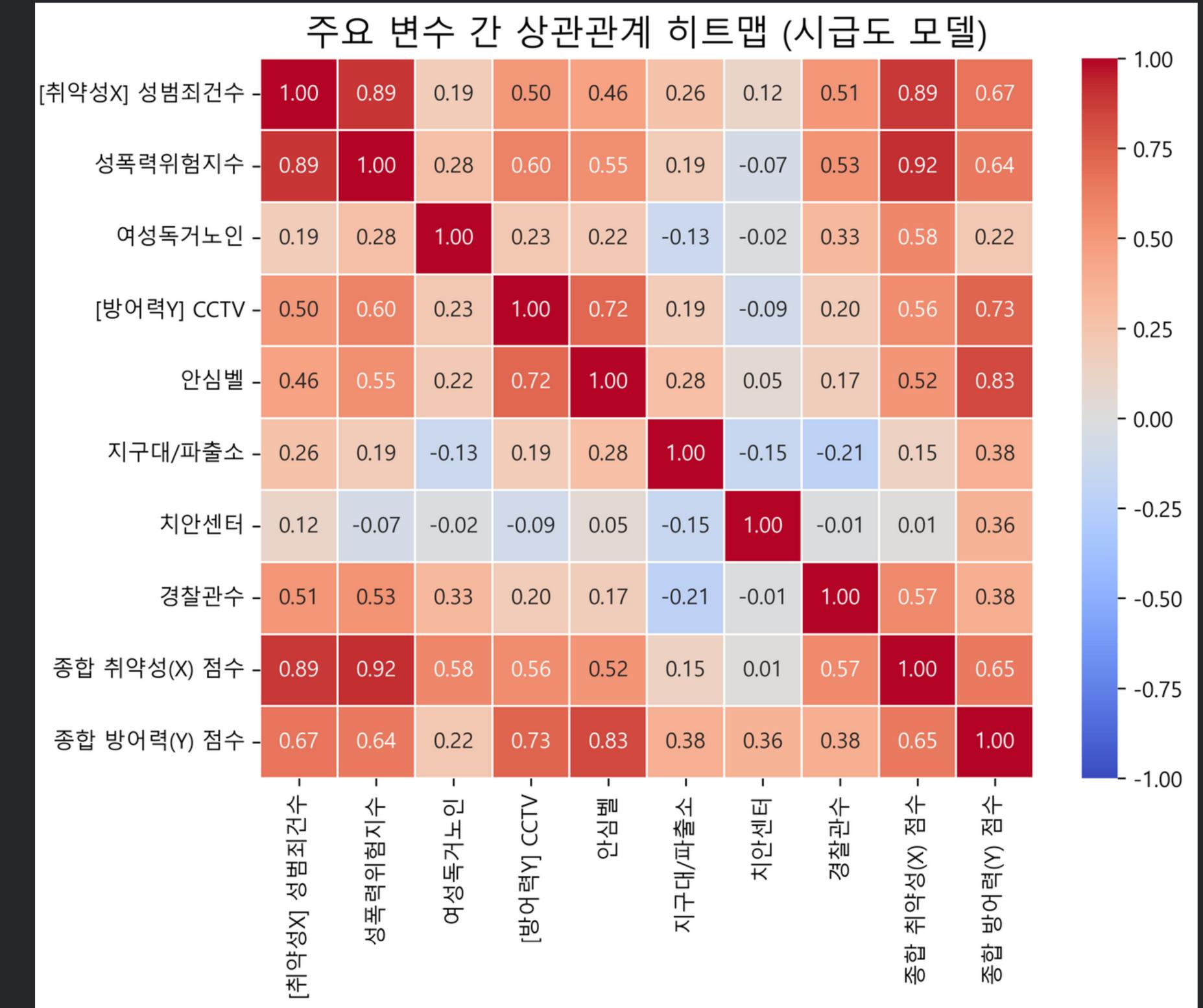
<참고>

X와 Y의 관계: → 양(+)이면 위험 지역에 인프라가 잘 구축된 것
→ 음(-)이면 위험 지역에 인프라가 부족한 것

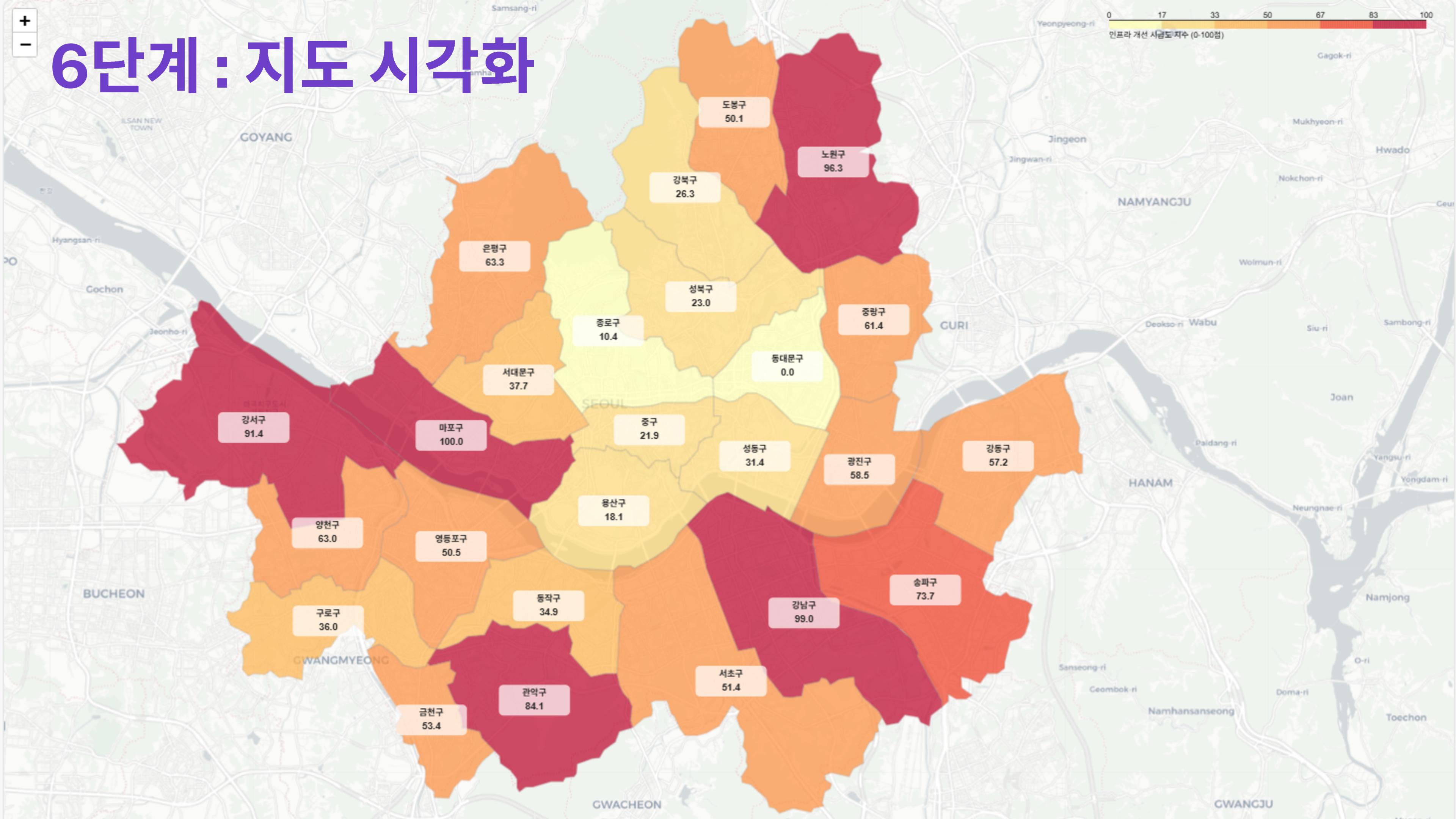
p-value(별표) 의미: (p<0.05): 유의미 / (p<0.01): 매우 유의미 — 통계적으로 신뢰 가능한 상관관계

- 주요 분석 결과:
종합 취약성(X) 점수 vs. 종합 방어력(Y) 점수
→ 0.65(강한 양의 관계)

취약성이 높은 지역일수록 방어력 인프라도 많이 갖춰져 있다는 것을 의미



6단계: 지도 시각화



더 안전한 도시를 위한 한 걸음: 서비스의 효과

여성 안심 안전 지수 도출 서비스의 기대 효과



효율적인 자원 배분

시급도 지수를 통해 **CCTV, 안심벨, 경찰 인력 등 안전 인프라** 를 가장 필요한 지역에 집중 투자할 수 있습니다.



시민의 안전 체감도 향상

데이터 기반의 예측적 투자로 **잠재적인 범죄 위험** 을 미리 낯춰, 여성 및 노약자의 **실질적인 안전 체감도** 를 높일 수 있습니다.



미래 확장성

이 모델은 서울뿐만 아니라, **전국 모든 지자체** 에 적용하여 지역별 맞춤형 **치안 정책을 수립** 하는 데 활용될 수 있습니다.

여성 안심 안전 지수는 단순한 분석을 넘어,
더 안전하고 편안한 도시 환경 조성을 위한 실질적인 기여를 합니다.

미래를 위한 약속: 더 완벽한 지수를 향해!

한계점 분석 및 지속적인 개선 방안



데이터 수집의 한계



최신성 부족

한계점: 최신 데이터(2024년 등) 수급에 한계가 있었습니다.

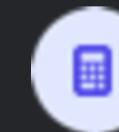
개선 방안: 기관과의 MOU 및 데이터 공유 시스템 구축을 통해 실시간 데이터 반영을 추진할 예정입니다.



분석 단위의 한계

한계점: 동 단위 세밀 데이터는 지역 민심 고려로 국가적 비공개 처리 상태였습니다.

개선 방안: 빅데이터로 동 단위 잠재 위험 지표를 자체 생성할 예정입니다.



모델 및 로직의 아쉬운 점



가중치 부재 및 선택 편향

한계점: 단순 합산으로 변수들이 동일 영향력을 가정함. 방어 시설이 이미 치안 수요 높은 지역에 설치됨.

개선: AHP 분석으로 차별적 가중치 부여. 지표를 인구 밀도나 범죄 건수 대비 재정 규화할 예정입니다.

가중치 모델

CCTV(0.4) + 안심벨(0.3) + 경찰서(0.3)

시설 재정규화

CCTV 수 ÷ (인구 밀도 × 범죄율)

**더 안전한 도시를 향한 한 걸음
모두가 안심하는 변화를 만들니다**

더 나은 모습으로 발전하겠습니다.

감사합니다.