

# 부산 청년 인구 감소의 지역별 차이 탐색

Exploring Regional Variations in the Factors Influencing Youth Population Decline in Busan

## 연구 배경

- 부산의 청년 인구의 감소 - 부산의 청년 인구 감소가 장기화되면서, 도시 쇠퇴 징후로 나타나고 있음
- 기존 정책의 한계 - 청년 인구 감소의 심각성과 복잡성에 비해 부산의 청년 정책은 효과가 적고 부족함
- 청년 인구 감소 영향 요인의 공간적 이질성 - 청년 인구 변화의 양상과 영향 요인은 지역별로 상이함

## 연구 목적

지역별 대응 정책 제안을 위한  
부산 청년 인구 감소의 지역별 차이 탐색

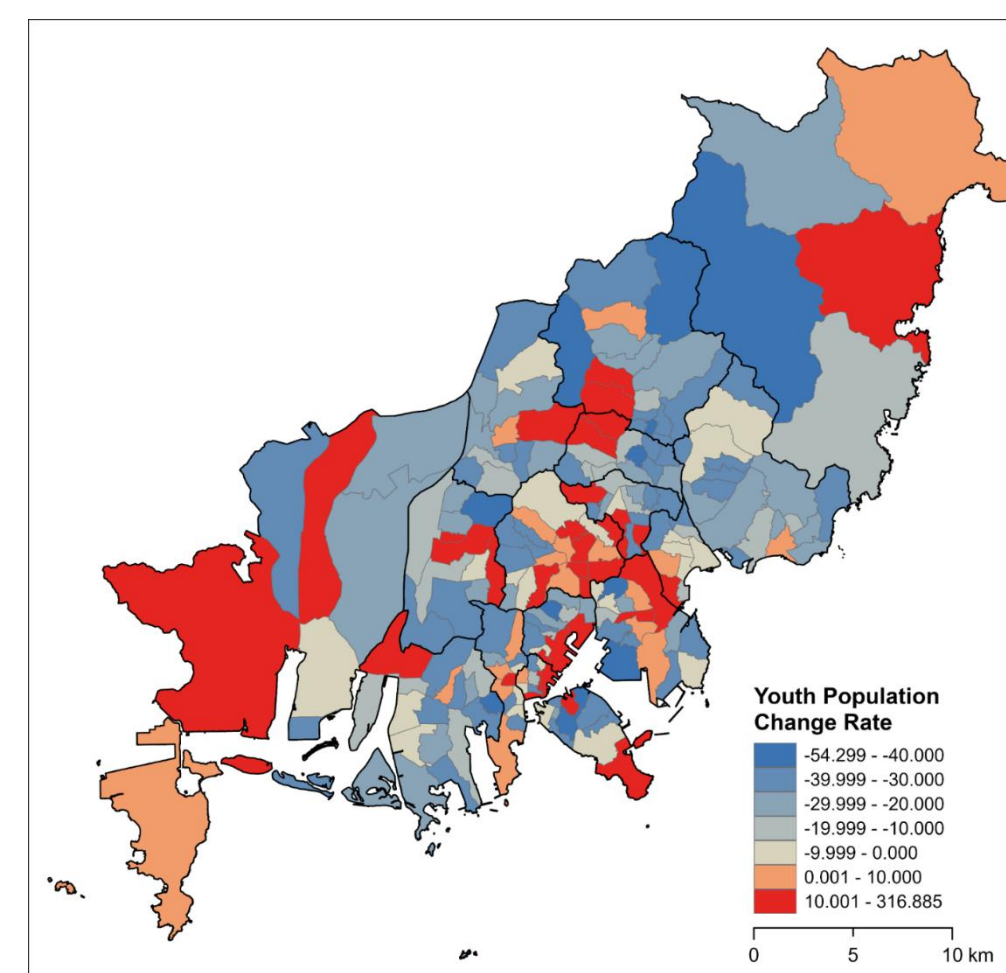
- 기계학습 모형 구축
- 전역적 해석
- 지역별 차이 해석

## 연구 방법 및 자료

### 1. 연구 범위 및 자료

- 연구 범위: 2024년 부산광역시 읍면동
- 종속변수: 행정동 단위 청년 인구 변화율 (2019-2024년)
- 독립변수: 주거, 경제, 인프라 관련 총 14개 변수

독립변수 분류	변수명
주거 환경	신규주택 비율, 노후주택 비율, 아파트 비율, 연립·다세대 주택 비율, 정비사업 시행 유무, 평균 공시지가, 총 주택 수
경제	사업체 수, 종사자 수
인프라 접근성	교통시설 접근성, 문화시설 접근성, 공원 접근성
토지이용	토지이용 압축도, 토지이용 혼합도



종속 변수

- 행정동 단위 20-39세 청년 인구 변화율
- 2019년에서 2024년 사이의 변화율을 계산
- 청년 인구 증감의 상대적 강도를 비교하기 위하여 단순 인구 수가 아닌 청년 인구 변화율을 사용함

$$\frac{(2024\text{년 청년인구} - 2019\text{년 청년인구})}{(2024\text{년 청년인구})}$$

### 2. 기계학습 모형 설계

- 다음처럼 네 가지 머신러닝 모형을 설계하여 정확도가 가장 높은 모형 선정
- 훈련 및 검증 데이터는 8:2로 분할해 검증 데이터에 대해 RMSE와 R<sup>2</sup> 산출
- Tree-Structured Parzen Estimator(TPE) 기반 베이지안 알고리즘을 활용해 각 기계학습 모형의 최적 하이퍼파라미터를 추정, 학습에 사용함

Random Forest (RF)	Support Vector Regression (SVR)
5-Fold 교차 검증 나무 수: 200~1500 최대 깊이: 3~20	5-Fold 교차 검증 학습률: 0.01, 상호작용 깊이: 2~5 나무 수: 100~1000
XGBoost (XGB)	Feed-forward Neural Network (FNN)
5-Fold 교차 검증, RBF 커널 사용 C(규제 강도): 0.01~100 ε: 0.0001~1, γ: 0.0001~10	dropout, weight decay 적용 ReLU 활성화 함수, 4개의 은닉층 사용 (512, 256, 128, 32)

### 3. GeoShapley 적용

- 선정된 모형에 GeoShapley를 적용하여 주요 영향 요인 해석

#### SHAP

- 설명가능한 AI 기법 중 하나. 기계학습 모형은 그 예측 과정과 변수의 영향을 알 수 없기 때문에, Shapley 값을 기반으로 예측값에 대한 설명 가능성을 확보하기 위한 방법으로 활용됨
- 예측에 대한 변수별 기여도는 알 수 있지만, 공간 변수의 효과와 지역별 차이는 파악할 수 없음

#### GeoShapley

- 본 연구는 부산 청년 인구 감소 영향 요인을 지역별로 파악하기 위해 GeoShapley를 활용함
- 예측값을 전역적 기준값, 위치 효과, 변수 효과, 위치와 변수의 상호작용 효과로 분해해 단순히 변수의 전역적 기여도뿐 아니라 지역별 주요 기여도 차이까지 파악할 수 있음

$$\hat{y} = \varphi_0 + \varphi_{GEO} + \sum_{j=1}^p \varphi_j + \sum_{j=1}^p \varphi_{(GEO, j)}$$

( $\varphi_0$ : 전역적 기준값,  $\varphi_{GEO}$ : 위치 효과,  $\sum_{j=1}^p \varphi_j$ : 변수 효과,  $\sum_{j=1}^p \varphi_{(GEO, j)}$ : 위치×변수 상호작용 효과)

## 연구 결과

### 1. 기계학습 모형 구축 결과

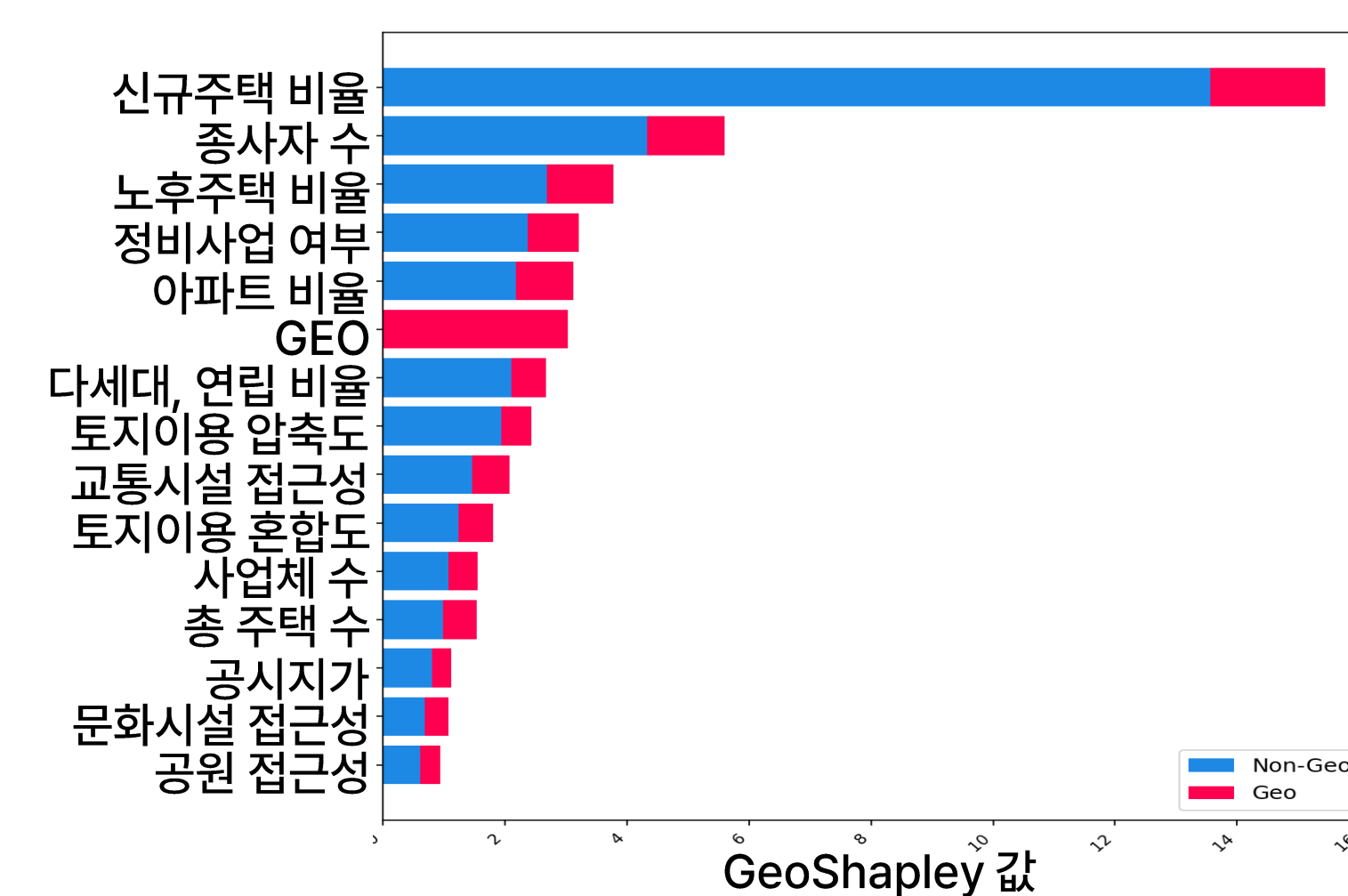
정확도가 가장 높은 FNN으로 최종 선정

Model	RMSE	R <sup>2</sup>
XGB	18.12	0.68
SVR	18.96	0.67
RF	18.04	0.70
FNN	16.51	0.70

### 2. 전역적 영향요인

#### 주요 해석

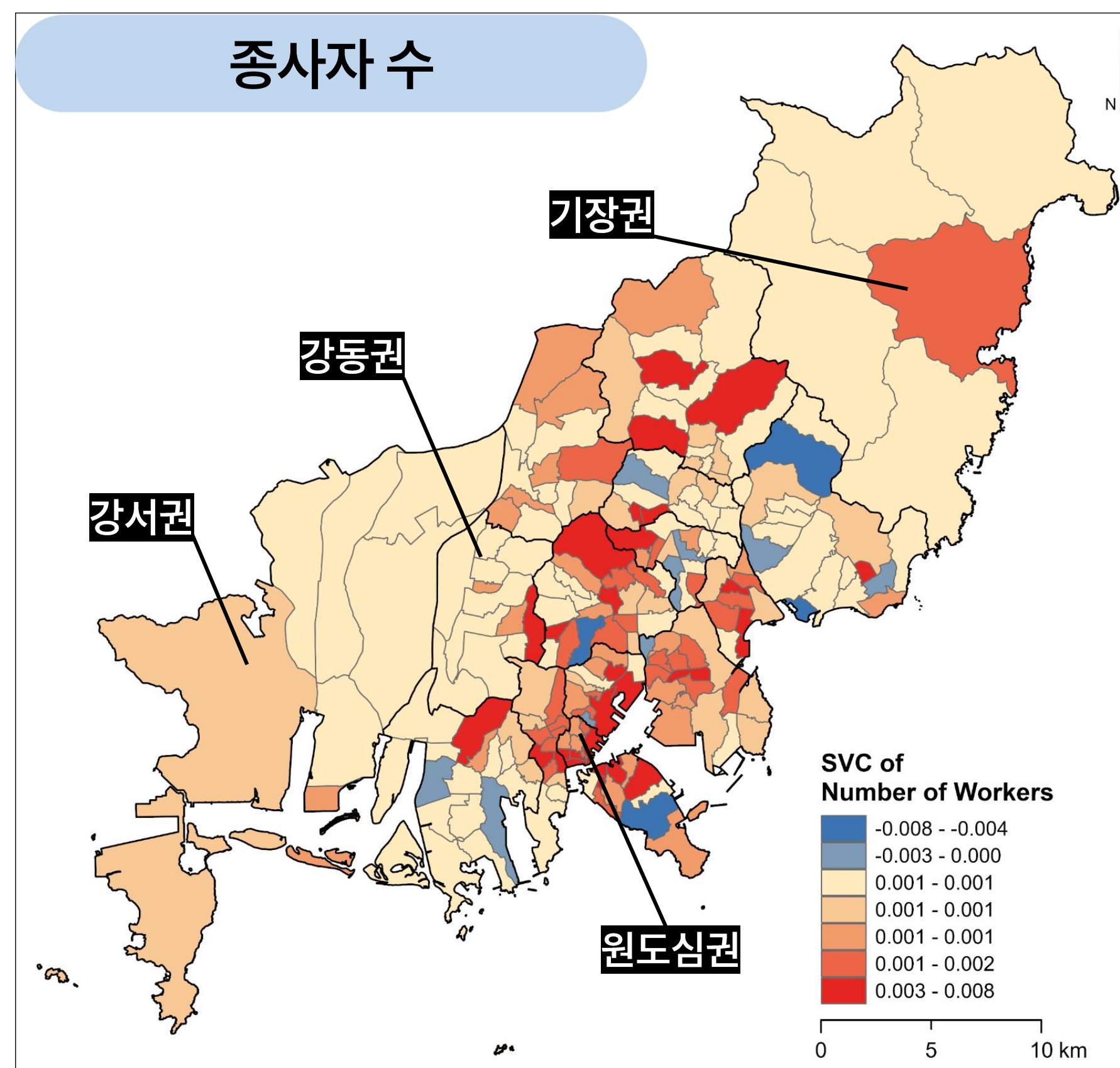
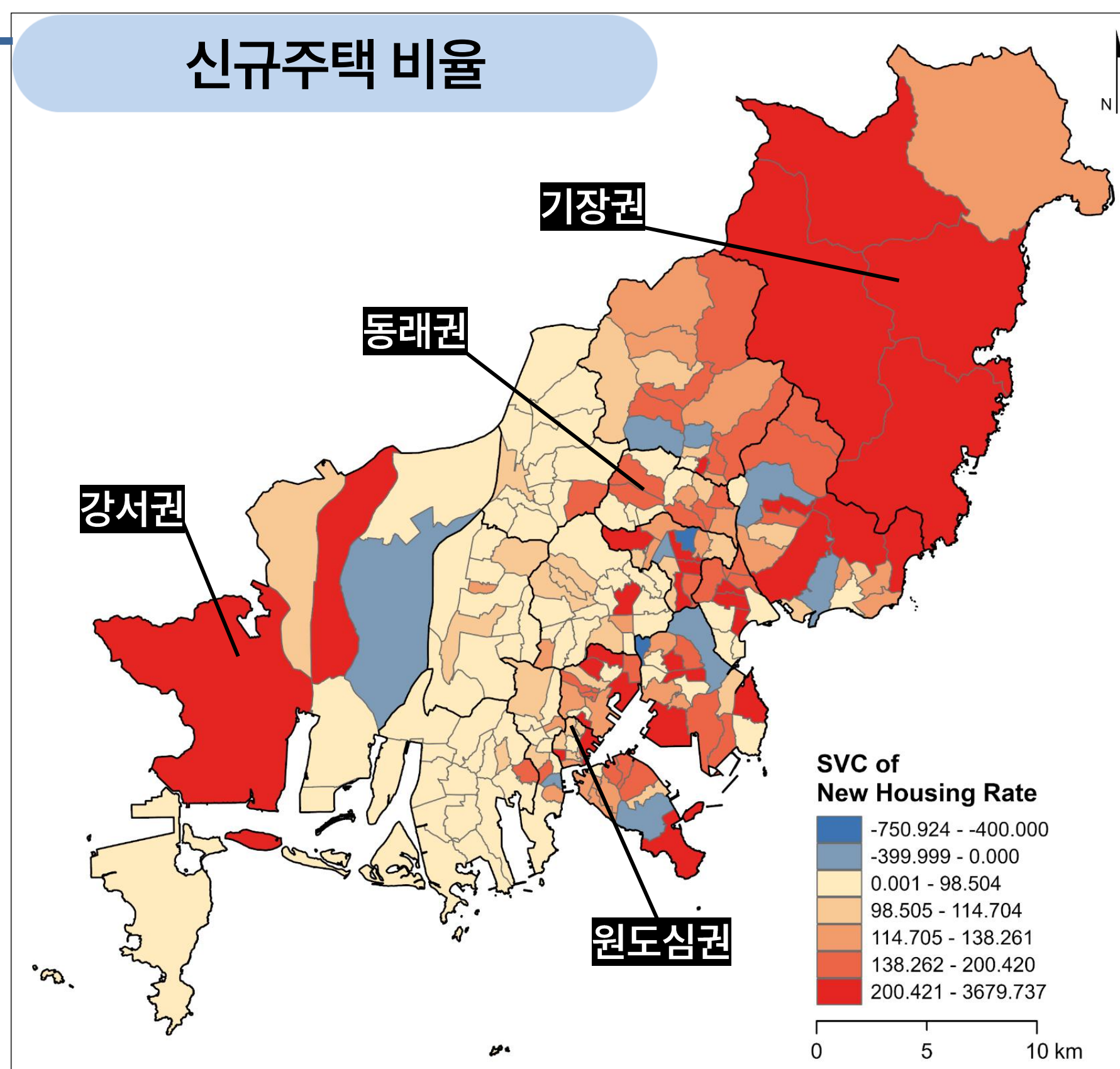
- GeoShapley 결과, 총 14개 독립변수 중 신규주택 비율과 종사자 수가 청년 인구 변화율 예측값의 정확도 향상에 상대적으로 큰 기여도를 보임
- 주거환경과 고용구조가 모형 예측 과정에서 주요하게 반영되었음을 의미함
- 위치 효과와 위치×변수 상호작용 효과가 확인되며, 이는 변수의 영향이 지역적 맥락에 따라 달라질 수 있음을 시사함



### 3. 지역별 영향요인 차이

- GeoShapley 결과, 각 변수의 기여 방향과 강도는 지역별로 상이하게 나타남, 신규주택 비율과 종사자 수, 둘의 위치×변수 상호작용 효과가 큰 중요도를 보임

- 주로 신규주택 비율 상승이 청년 인구 변화율을 높이는 방향으로 작용
- 강서 및 기장권: 신규주택 비율이 예측값을 증가시키는 경향이 강함
- 원도심 및 동래 일부: 예측값을 감소시키는 방향으로 기여함
- 전반적으로 종사자 수가 청년 인구 변화율을 높이는 방향으로 작용함
- 원도심권: 종사자 수가 청년 인구 변화율을 높이는 경향이 강함
- 기장·강동·강서: 변화율을 증가시키는 방향으로 작용하지만, 비교적 약함



- 강서·기장권**  
신도시가 있어 신규주택 공급, 일자리가 청년 정착 기반으로 작용하는 지역
- 원도심권**  
기존 도심의 인프라를 기반으로 고용 기회가 청년을 강하게 유입시키는 지역
- 동래·해운대 일부**  
주거·고용과 함께 생활권 인프라 등 복합적 요소를 고려하는 것이 중요한 지역

## 연구 결론

### 연구 결론

- 부산 청년 인구 변화와 영향 요인의 지역적 맥락을 발견하고 해석함
- 신규주택 비율의 기여도는 강서권과 기장권 등 외곽 지역에서 높게 나타나, 원도심과 동래권 일부 지역에서는 낮게 나타남
- 종사자 수의 기여도는 원도심권에서 상대적으로 높게 나타남

### 연구 의의

- 청년 인구 감소를 지역별 차이가 있는 공간적 문제로 접근해 다른 시각을 제시함
- 지역별 맞춤형 정책 수립 및 시행의 기초 자료로 활용될 수 있을 것으로 기대됨

### 향후 연구

- 집계구·격자 등 세밀한 공간 단위를 활용하여 지역별 영향 요인의 차이를 정교하게 분석할 수 있음
- 해석 결과를 직접적인 지역 맞춤형 정책으로 연결