

# 빅데이터를 이용한 서울시 젠트리피케이션 예측

---

---

2017.08

---

Gentri Hackers

---

박민제 손원희 송예설 심규호

---

# 목차

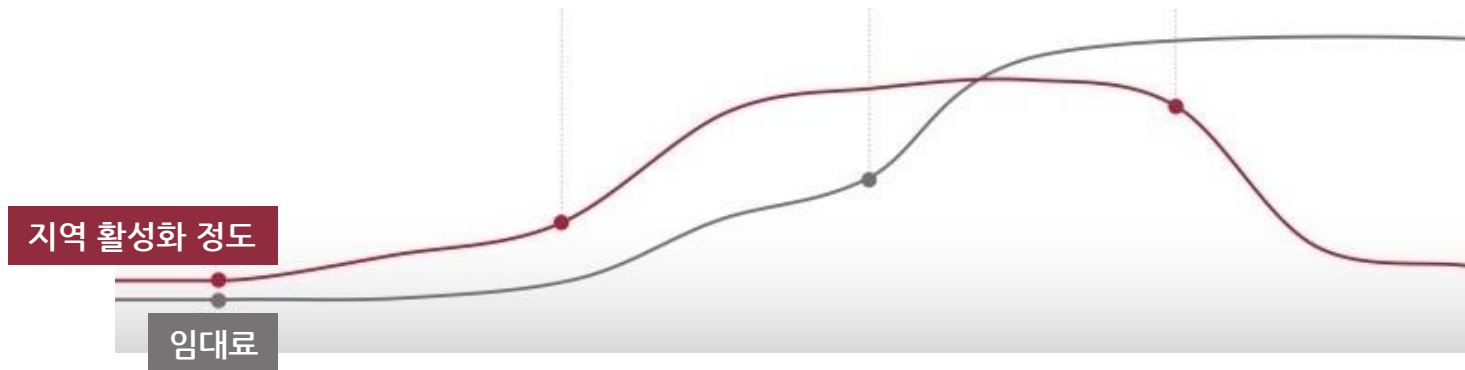
- 
1. 서론
  2. 연구 목표
  3. 데이터 수집
  4. 데이터 분석
  5. 정책 제언
  6. 한계점

# 젠트리피케이션 정의 및 연구의 필요성

## 지역별 매매, 전월세 시장

- 투기, 권리금 장사 등으로 인한 세입자 부담 증가
- 부동산 시장 차별화 심화
- 지역별 빈부 격차

## 젠트리피케이션 발생 과정



젠트리피케이션은 시민들이 대응하기 어려울 정도로 급격히 일어나며  
주거환경에 직접적으로 영향을 미치므로 큰 피해가 발생함  
**BUT 시민 개인의 힘으로 대응하기에 한계가 있음!!**

# 지역별 젠트리피케이션 발생 확률과 유형 예측을 통한 맞춤 정책 시행

## 목표 #1 젠트리피케이션 발생 확률을 의미

### 하는 젠트리피케이션 종합 지수 예측

- 젠트리피케이션 종합 지수는 젠트리피케이션 발생 확률을 0부터 1까지의 숫자로 나타낸 지표임
- 절대적 수치가 아닌, 행정구역 간 비교가 가능하도록 하는 상대적 수치임

## 목표 #2 젠트리피케이션 유형 분류 및 예측

- 발생 여부 예측에서 그치지 않고, 그 원인을 파악하기 위해 젠트리피케이션이 발생할 것으로 기대되는 지역을 대상으로 젠트리피케이션 유형까지 예측함
- 젠트리피케이션 유형 정의 및 판별을 위해 정량적, 정성적 요인 모두 고려함

젠트리피케이션이 발생할 것으로 예상되는 지역과 그 심각성, 원인에 따라  
세분화되고 지역 특화된 정책을 시행할 수 있음

### 제약조건

- Prototype에서 데이터 수집 지역은 서울로 제한, 동 단위 데이터 수집
- 2012년부터 2016년까지 최근 5개년의 데이터로 분석을 진행함

## 필요한 데이터를 수집 후 전처리 과정에서 연평균 변화율로 환산

### 사전 정보 탐색

젠트리피케이션의 발생과 상관관계가 높으면서 지역의 특성을 반영하는 데이터를 수집하기 위해 조사한 결과, 젠트리피케이션과 관련이 높은 지표를 크게 **경제, 문화, 관광** 분야로 분류할 수 있으며, 특히 경제 분야의 영향력이 가장 큼

### 데이터 수집

#### 경제

- 부동산 거래 현황: 임대료, 매매가, 지방세 징수액, 지가변동률
- 인구 이동: 전출입 빈도, 유동인구 수
- 지역 경제 상황: 사업체 수 및 연령, 창업률, 프랜차이즈 수

#### 문화

- 문화예술 인프라: 업종 별 분포 또는 밀도, 업종 다양성 증감률
- 문화예술 행사: 문화행사 개최 수
- 화제성: 소셜 데이터

#### 관광

- 관광객: 외국인/내국인 관광객 수, 관광객 소비액
- 관광지: 문화재 등 유명 관광지 수
- 화제성: 소셜 데이터

### 데이터 전처리

#### 연평균 변화율

: 데이터의 첫 연도부터 마지막 연도까지 일정한 비율로 성장/감소했다고 가정했을 때의 그 성장/감소율

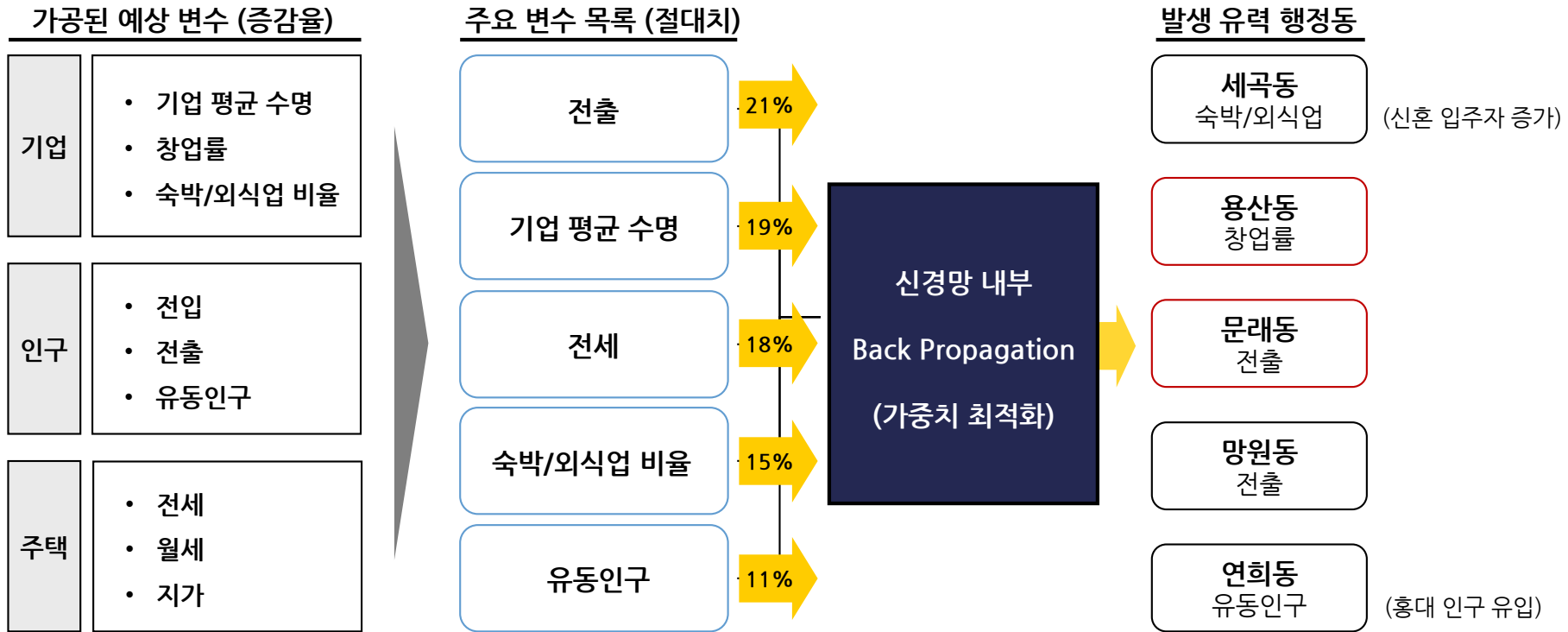
- List-up 한 데이터 중 주어진 기간 내 확보가 가능하며 유의미한 영향력을 가진 데이터를 대상으로 전처리함
- 각 데이터마다 절대량의 크기가 매우 다르다는 점, 젠트리피케이션은 어떤 지역이 과거에 비해 얼마나 변화되었는지가 중요하다는 점에서 모든 데이터를 연평균 변화율로 환산하였음

# Logistic Regression을 적용해 젠트리피케이션 발생 확률 계산

## Training Set

True : 서울시 젠트리피케이션 종합대책 기반 발생 행정동

False : 동별 전출 인구 / 전세 가격 동향 분석을 바탕으로 한 하위 10개 행정동



지역 정보 자동 업데이트로 즉각적 동향 파악 및 빠른 대응 가능

# 각 행정동의 연관검색어와 유형별 키워드 비교를 통해 유형 분류

## 유형 분류 과정

각 행정동 연관검색어  
웹 크롤링

유형별 키워드  
리스트 업

연관검색어와  
키워드 비교

유형별 순위

## 유형별 키워드

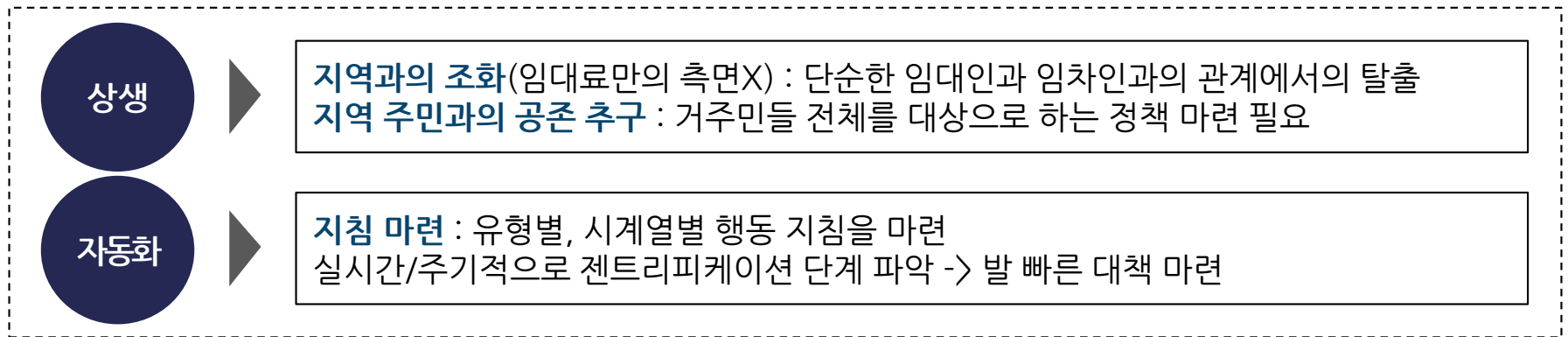
관광	상업경제	문화예술
쇼핑	시장	시장
방송	부동산	전통
전통	데이트	맛집
문학	방송	문학
인물	영화관	도서관
축제	쇼핑	벽화
거리	맛집	거리
공원	놀거리	문학
...	...	...

## 유형 분류 결과

관광 관련 키워드			문화예술 관련 키워드			상업 관련 키워드		
순위	행정동	키워드 수	순위	행정동	키워드 수	순위	행정동	키워드 수
1	명동	7	1	명동	7	1	명동	7
2	서원동	6	2	서원동	6	2	서원동	6
3	삼선동	5	3	삼선동	5	3	삼선동	5
4	삼청동	5	4	삼청동	5	4	삼청동	5
5	중곡동	5	5	중곡동	5	5	중곡동	5
6	혜화동	4	6	혜화동	4	6	혜화동	4
7	성수동	3	7	성수동	3	7	성수동	3
8	종로동	3	8	종로동	3	8	종로동	3
9	방배동	2	9	방배동	2	9	방배동	2
10	방이동	2	10	방이동	2	10	방이동	2

- 시계열 데이터가 아니기 때문에 변화추이는 알기 어려움
- 사람들에게 알려진 지역명과 행정동명이 다를 수 있음 (e.g. '홍대'와 '서교동')
- 유형별 키워드는 분석자의 자의적인 판단에 의함

## 정책 방향성 제안



주요 변수 목록(절대치)	제안
<div>전출 21%</div>	거주민들의 전출을 최소화할 수 있는 방안 모색
<div>기업 평균 수명 19%</div>	창업 초기 뿐만이 아니라 창업을 지속할 수 있는 여건 마련 필요
<div>전세 18%</div>	기존의 임대료 및 전·월세 대책의 유형별(경제, 문화 등) 접근 필요



## 분야별 정책 제안 및 기대 효과

### 경제적 관점

- 프랜차이즈 입점 규제 방안 : 프랜차이즈의 무자비한 난입은 동네의 고유 문화 상실에 근원이 됨
  - ‘상생’이라는 이념적 가치를 실질적 이익과의 연관
- 국토연구원 대응 방안(2017.08.28)
- 부동산의 금융 상품화 주의 요망

### 문화적 관점

- **전출의 가중치↑** : 예술가들에의 주거 보급 일환으로, 독거 노인과의 셰어링 산업(성동구 한지붕 룸쉐어링 정책 참고)
- **문화행사를 지원**하거나 주민들에게 **거주의 혜택을 부여**하여 **전출을 최소화**할 수 있도록 함

### 관광적 관점

- 국소적으로 관광 동선이 한정되지 않도록 다각화 & 공공편의시설 확충을 통한 관광에서의 편의 제공
- 가게 등록 시, 주변 문화와의 소통을 고려하여 정부 지원 : 지역의 고유한 문화를 극대화할 수 있는 방향으로의 입점 유도

**종합적 관점 필요** : 저렴한 임대료 및 지가만이 젠트리피케이션의 원인인 것은 아님을 이해해야 하며, 경제, 문화, 관광의 다양한 측면에서 지역 주민과의 공존과 상생을 추구하는 방향이 필요함

#### 신규 창업자

데이터를 기반으로 시장 진입 시점 및 지역을 전략적으로 계획 가능

#### 기존 거주자

확실한 정부 정책 수립으로 인해 낮은 퇴출 가능성

#### 기존 영세사업자

젠트리피케이션 단계에 따라 사업 철수 시점 계획 가능

#### 일반인

유형별, 단계별 젠트리피케이션 지역 탐색 가능

유형별 지역에 따른 새로운 사업 기회 탐색 가능

# 한계점 및 발전 방안

Current		Next Step
영역	세부 영역	Key Questions
Data	데이터 종류	<p>? 다방면적 관점에 대한 데이터</p> <p>! 문화, 관광적 관점에 대해 분석해 볼 수 있는 데이터 수집</p>
	연속성	<p>? 시계열 데이터의 중요성 (과거 젠트리피케이션이 발생한 지역의 현재 경향은 관련성이 적음)</p> <p>! 이미 발생한 지역의 발생 시기 주변의 경향성 조사</p>
Logic	기준 및 논리	<p>? Train Data의 Output Value 설정 기준</p> <p>! 해외 예시 등 활용해 Sample 갯수 증가</p>
Variance	복잡성	<p>? 각 데이터의 개별적 특성이 고려되지 않음</p> <p>! 데이터 수집 방법 / 수집 항목 간결화, 명확화</p>

감사합니다

---

# APPENDIX - Matlab 코드 실행 화면

The screenshot displays the MATLAB environment with the 'GentrHackers.m' script open in the editor. The script contains the following code:

```
1 %% GentrHackers Neural Network Learning
2 %% Initialization
3 clear ; close all; clc
4
5 %% Setup the parameters
6 input_layer_size = 9;
7 hidden_layer_size = 10;
```

The Command Window shows the execution progress, including iterations and costs:

```
Iteration 1274 | Cost: 7.821128e-01
Iteration 1275 | Cost: 7.821128e-01
Iteration 1276 | Cost: 7.821128e-01
Iteration 1277 | Cost: 7.821128e-01
Iteration 1278 | Cost: 7.821128e-01
Iteration 1279 | Cost: 7.821128e-01
Iteration 1281 | Cost: 7.821128e-01
```

The workspace (작업 공간) lists the variables and their sizes:

- cost: 1276x1 double
- costFunction: @(p)nnCostFun
- h1: 206x10 double
- h2: 206x2 double
- hidden\_layer\_size: 10
- INIT\_EPSILON: 3
- input\_layer\_size: 9
- J: 27.4216
- lambda: 3
- m: 206
- nn\_params: 122x1 double
- num\_labels: 2
- options: 1x1 struct
- P: 206x9 double
- prediction: 206x1 double
- Theta1: 10x10 double

The Command Window also displays the final Theta1 matrix:

```
>> Theta1
Theta1 =
-2.1074 -0.1656 -0.0423 0.0275 0.2086 0.1459 0.0123 -0.1064 0.1745 -0.0718
-4.6098 -0.1676 -0.0017 -0.0440 0.1589 0.1974 0.0152 0.0482 0.2333 -0.0107
2.5160 0.0434 -0.1591 0.0939 -0.1114 -0.0210 -0.0188 0.0341 -0.1693 0.1664
4.6098 0.1676 0.0017 0.0440 -0.1589 -0.1974 -0.0152 -0.0482 -0.2333 0.0107
2.1074 0.1656 0.0423 -0.0275 -0.2086 -0.1459 -0.0123 0.1064 -0.1745 0.0718
-2.1074 -0.1656 -0.0423 0.0275 0.2086 0.1459 0.0123 -0.1064 0.1745 -0.0718
2.1074 0.1656 0.0423 -0.0275 -0.2086 -0.1459 -0.0123 0.1064 -0.1745 0.0718
2.1074 0.1656 0.0423 -0.0275 -0.2086 -0.1459 -0.0123 0.1064 -0.1745 0.0718
4.6098 0.1676 0.0017 0.0440 -0.1589 -0.1974 -0.0152 -0.0482 -0.2333 0.0107
4.6098 0.1676 0.0017 0.0440 -0.1589 -0.1974 -0.0152 -0.0482 -0.2333 0.0107
```

# 참고문헌

## 활용 데이터 출처

서울특별시 서울부동산정보광장(<http://land.seoul.go.kr/land/index.jsp?sigunguCd=11000&admin=>)  
KB부동산 (<http://nland.kbstar.com/quics?page=kbland>)  
크레딧하우스 (<https://kredithouse.com/>)  
KOSIS 국가통계포털  
([http://kosis.kr/statisticsList/statisticsList\\_01List.jsp?vwcd=MT\\_ZTITLE&parmTabId=M\\_01\\_01#SubCont](http://kosis.kr/statisticsList/statisticsList_01List.jsp?vwcd=MT_ZTITLE&parmTabId=M_01_01#SubCont))  
국토교통부 실거래가공개시스템 (<http://rtdown.molit.go.kr/rtns/rqs/initRtTrList.do#>)  
통계청(MDIS) (<https://mdis.kostat.go.kr/index.do>)  
온나라부동산정보 (<http://www.onnara.go.kr>)  
국토교통 통계누리 (<http://stat.molit.go.kr>)  
국토연구원 (<http://www.krihs.re.kr>)  
KDI 한국개발연구원 (<http://www.kdi.re.kr>)  
한국감정원 부동산통계정보 (<http://www.r-one.co.kr/rone/>)  
공공데이터포털 (<http://www.data.go.kr/dataset/3050988/openapi.do>)  
서울열린데이터광장 (<http://data.seoul.go.kr/index.jsp>)  
우리마을가게 상권분석서비스 (<http://golmok.seoul.go.kr/>)  
통계지리정보서비스 (<http://sgis.kostat.go.kr/> )  
서울통계 (<http://stat.seoul.go.kr/>)

# 참고문헌

## 참고 논문

서울특별시 기획조정실(2015), 서울시 젠트리피케이션 종합대책.

황준기(2015), 문화주도적 젠트리피케이션 현상에 의한 장소성 변화 연구, 서울시립대학교 대학원 행정학 석사논문.

허자연(2015), 상업공간의 젠트리피케이션 과정 및 사업자 변화에 관한 연구 : 경리단길 사례, 서울도시연구, 16(2), 19p-33p.

김상현(2016), 성수동 지역의 젠트리피케이션 과정 및 특성 연구, 문화콘텐츠연구, (7), 81-105.

남기범(2016), 국내 젠트리피케이션 논의의 쟁점과 현안 진단, 부동산 포커스.

양승호(2016), 젠트리피케이션 체감지표 선정에 관한 연구 - 서울특별시 상업지역을 중심으로, 광운대학교 박사학위 논문.

윤윤채(2016), 상업용도 변화 측면에서 본 서울시의 상업 젠트리피케이션 속도 연구, 서울도시연구, 17(4), 17p-32p.

김영호(2017), 공간 회귀와 공간 필터링을 이용한 서울시 젠트리피케이션의 발생 원인 및 특징 분석, 한국도시지리학회지, 19(3), 77p-86p.