

# KUIAI 해커톤

---

아이티에스

통계학과 구병모  
산업경영공학부 김가영  
산업경영공학부 안정수



1 개발  
배경

2 분석  
개요

3 모델  
구현

4 활용  
방안

1

# 개발배경

# 개발배경

## • 상권

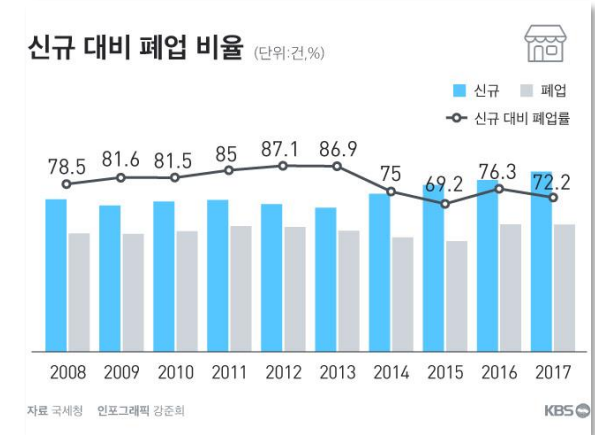
: 지정기업이나 다수기업에서 판매하는 상품이나 서비스를 판매할 확률이 '0' 이상인 잠재적 고객을 포함하고 있는 지리상으로 묘사된 지역

## • 입지

: 경제활동을 잘하기 위해 적당한 사업장의 장소를 찾아서 선택하는 것으로, 사업장(점포)이 소재(所在)하는 위치조건

## • 분석 필요성

: 점포 입지는 한번 정해지고 나면 쉽게 바꿀 수 없는 장기적인 성격을 가지는 고정 투자이므로 무엇보다 신중하게 결정



# 2

## 분석개요

---

2.1 분석목표

2.2 분석내용

# 분석 목표

## • 서울 인구 밀집 현상

:우리나라에서 토지 대비 인구 밀도가 가장 높은 서울의 입지 분석은  
유의미한 경제분석 이라는 의의를 가짐

## • 건축물 입지 점수 예측 모델 구축

: 서울시 건축물 생애 이력 데이터의 도로명 주소로 입지 점수를 예측

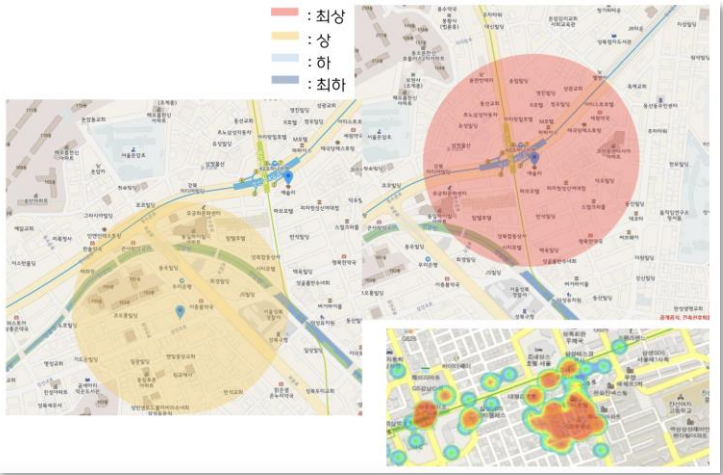
## • HEATMAP 시각화

: 도로명 주소 검색 시 해당 건물의 반경 500m을 heatmap으로 표현

## • 건축물의 등급화

: 입지 점수를 기반으로 4개의 등급으로 나눠 영역으로 등급을 표현

행정구역	2021년12월_총인구수	2021년12월_세대수
전국 (10000)	51,638,809	23,472,895
서울특별시 (1)	9,509,458	4,426,007
부산광역시 (2)	3,350,380	1,544,663
대구광역시 (3)	2,385,412	1,063,893
인천광역시 (4)	2,948,375	1,298,647
광주광역시 (5)	1,441,611	645,712
대전광역시 (6)	1,452,251	664,417
울산광역시 (7)	1,121,592	482,650
세종특별자치	371,895	153,649
경기도 (4100)	13,565,450	5,841,995
강원도 (4200)	1,538,492	746,220
충청북도 (4300)	1,597,427	760,672
충청남도 (4400)	2,119,257	1,001,915
전라북도 (4500)	1,786,855	849,001
전라남도 (4600)	1,832,803	903,108
경상북도 (4700)	2,626,609	1,276,846
경상남도 (4800)		
제주특별자치		



# 분석 내용

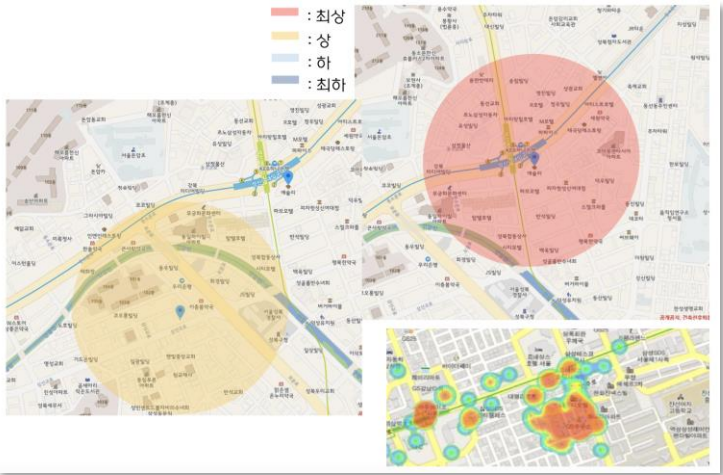
## • 서울시 건축시설의 상권 데이터

- 건축물 생애이력 (판매시설)
- 건축물 생애이력(제 1종 근린생활시설)
- 건축물 생애이력(제 2종 근린생활시설)
- 상권 코드, 매출, 생활인구
- 점포 정보
- 지하철역 정보
- 학교 정보

## • 변수 설정

- 설명 변수: 서울시 건축물 생애이력 및 기타 거리 데이터, 평균 생활 인구
- 반응 변수: 상권의 평균 매출액

행정구역	2021년12월_총인구수	2021년12월_세대수
전국 (10000)	51,638,809	23,472,895
서울특별시 (1)	9,509,458	4,426,007
부산광역시 (2)	3,350,380	1,544,663
대구광역시 (3)	2,385,412	1,063,893
인천광역시 (4)	2,948,375	1,298,647
광주광역시 (5)	1,441,611	645,712
대전광역시 (6)	1,452,251	664,417
울산광역시 (7)	1,121,592	482,650
세종특별자치	371,895	153,649
경기도 (4100)	13,565,450	5,841,995
강원도 (4200)	1,538,492	746,220
충청북도 (4300)	1,597,427	760,672
충청남도 (4400)	2,119,257	1,001,915
전라북도 (4500)	1,786,855	849,001
전라남도 (4600)	1,832,803	903,108
경상북도 (4700)	2,626,609	1,276,846
경상남도 (4800)		
제주특별자치		



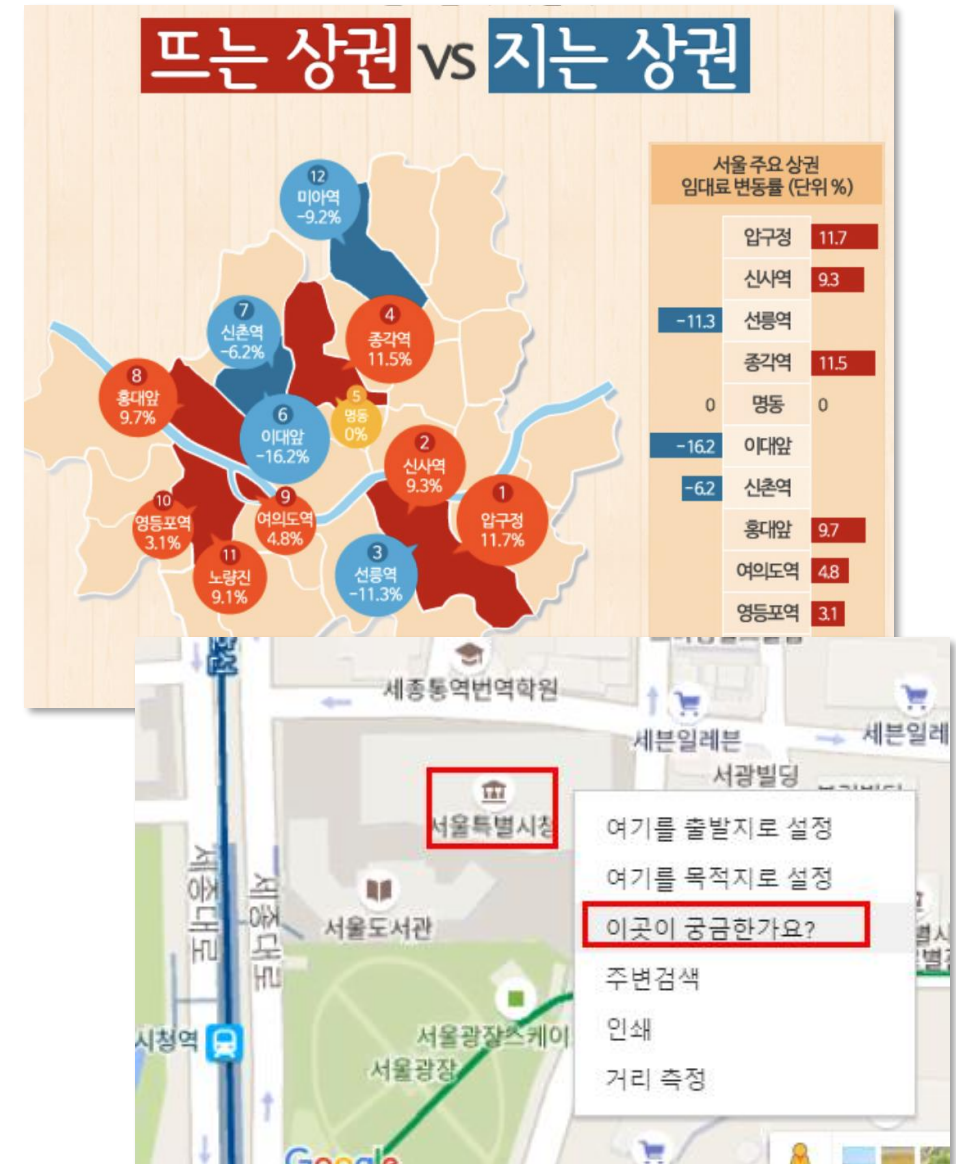
## 분석 내용

### • 주소 좌표화

: Kakao API 및 Google API의 Geocoding을 이용하여 주소를 위도 경도로 변환 후, **하버사인 거리**로 표현

### • 데이터들의 관계 규정

: 변수들간의 상관 분석 등의 검정 실시





# 3

## 모델구현

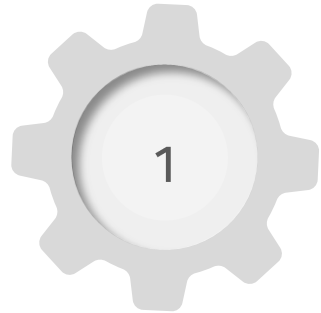
---

3.1 데이터셋 구축

3.2 데이터 학습 및 입지점수 모델 개발

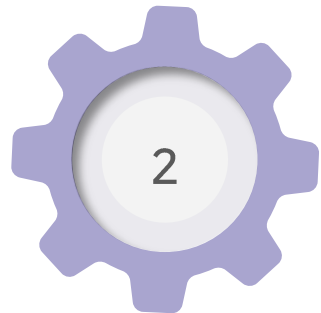
3.3 최종 함수 구현

### 전체 모델 구현 흐름도



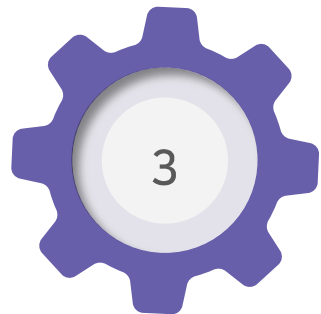
#### 데이터셋 구축

- 1.1 생애이력 데이터 결합
- 1.2 좌표 정보 추출
- 1.3 거리 계산, 평균 생활인구, 평균 매출액(y) 추출
- 1.4 건물별 생애 이력 데이터 결합



#### 데이터 학습 및 입지점수 예측모델 개발

- 2.1 데이터 결측치 처리
- 2.2 중복 도로명 주소 처리
- 2.3 입지점수 예측모델 개발



#### 최종 함수 구현

- Input: 도로명 주소
- Output: 입지점수, 점수별 히트맵, 건물별 등급 지도

1.1 생애이력 데이터 결합

2650	2651	연령 구분	서울특별시영등포구영등포동8가85-12	서울특별시영등포구영등포동157-11	NaN	NaN	0	0	1	0.0	60.000	300.00	240.00	0.00	0.00	제2종근로자 철근콘크리트구조	(철근)콘크리트	1	4	0.00	0	0.00	0	0
2651	2652	연령 구분	서울특별시영등포구영등포동8가91	서울특별시영등포구영등포동154	당산푸르지오	상가동	0	24	0	0.0	329.780	1848.70	989.34	0.00	0.00	제2종근로자 철근콘크리트구조	(철근)콘크리트	1	3	11.45	0	0.00	0	0

120009 rows × 42 columns

- 총 120009건의 생애이력 데이터가 결합되었습니다.

용도별, 지역구별 엑셀 데이터를 결합하여 120009건의 데이터셋 구축

## 1.2 좌표 정보 추출

건물 생애이력 도로명주소 좌표 ← 구글 api 활용

지하철역, 학교, 대규모 점포, 상권 좌표 ← 카카오 api 활용

## 1.3 거리 계산, 평균 생활인구, 평균 매출액(y) 추출

건물별로 모든 지하철역, 학교, 대규모 점포들과  
의 거리를 비교한 결과 **소요 시간**이 컸음

각각의 건물별로 일정 크기의 버퍼(ex. 5km, 10km)를 그려서 해당  
버퍼에 속한 지하철역, 학교, 대규모 점포들과 건물 간의 거리 중 **최소값**  
을 선택

하버사인 공식으로 거리 산출

	도로명주소	위도	경도	xy_tuple	geometry	haver_dis_subway
0	서울특별시 강동구 양재대로 1540	37.542920	127.142450	(37.5429196, 127.1424496)	POINT (968414.138 1949348.008)	0.316026
1	서울특별시 강동구 양재대로 1449	37.535248	127.138685	(37.5352477, 127.13868470000001)	POINT (968078.277 1948498.129)	0.307749
2	서울특별시 강동구 양재대로128길	37.544508	127.144746	(37.5445084, 127.14474569999999)	POINT (968617.642 1949523.507)	0.207319
3	서울특별시 강동구 천호대로 1156	37.533434	127.139603	(37.533434299999996, 127.13960279999999)	POINT (968158.619 1948296.632)	0.487983

1.3 거리 계산, 평균 생활인구, 평균 매출액(y) 추출

- 평균 생활인구와 평균 매출액(y)을 구하기 위해서 상권코드별로 2020년 총 매출액과 총 생활인구를 합산한 결과를 활용
- 도로명주소별 500m 반경 내에 있는 상권코드를 구하여 해당 상권들의 총 매출액과 총 생활인구의 평균을 내서 평균 생활인구, 평균 매출액(y)으로 사용
- 도로명주소별 가장 가까운 지하철역, 학교, 대규모 점포와의 거리 & 평균 매출액, 평균 생활인구 도출

building\_total

	도로명주소	위도	경도	haver_dis_subway	haver_dis_school	haver_dis_mart	mean_sales	mean_pop
0	서울특별시 강동구 양재대로 1540	37.542920	127.142450	0.316026	0.459123	0.110586	2.551361e+10	18225491.8
1	서울특별시 강동구 양재대로 1449	37.535248	127.138685	0.307749	0.416656	0.139283	4.715353e+10	20419730.5
2	서울특별시 강동구 양재대로128길	37.544508	127.144746	0.207319	0.498036	0.232098	2.896793e+10	22305878.2
3	서울특별시 강동구 천호대로 1156	37.533434	127.139603	0.487983	0.432536	0.346491	3.770563e+10	15761681.4
4	서울특별시 강동구 천호대로 1238	37.535540	127.148192	0.765237	0.442360	0.746099	3.966936e+10	12552664.0

가장 가까운 지하철  
역과의 거리

가장 가까운 학교와의 거리

가장 가까운 마트와의거리

평균 매출액

평균 생활인구

1.4 건물별 생애 이력 데이터 결합

- 건물별 생애 이력 데이터 중 건물 용도, 연면적, 층수, 주차장 개수, 승강기 개수만을 추출
- 앞서 만든 데이터셋과 건물별 생애 이력 데이터를 결합하여 분석에 활용할 최종 데이터셋을 구축

	도로명주소	위도	경도	haver_dis_subway	haver_dis_school	haver_dis_mart	mean_sales	mean_pop	용도	연면적 (㎡)	자주식주차장	기계식주차장	승강기 개수	층수
0	서울특별시 강동구 양재대로 1540	37.542920	127.142450	0.316026	0.459123	0.110586	2.551361e+10	18225491.8	판매 시설	1543.47	21	0	0	2
1	서울특별시 강동구 양재대로 1449	37.535248	127.138685	0.307749	0.416656	0.139283	4.715353e+10	20419730.5	판매 시설	2772.93	13	0	1	3
2	서울특별시 강동구 양재대로 128길	37.544508	127.144746	0.207319	0.498036	0.232098	2.896793e+10	22305878.2	판매 시설	3640.85	0	0	0	4
3	서울특별시 강동구 천호대로 1156	37.533434	127.139603	0.487983	0.432536	0.346491	3.770563e+10	15761681.4	판매 시설	1625.67	16	0	1	3

## 2.1 데이터 결측치 처리

- 도로명주소의 500m 반경에 상권이 하나도 없을 경우 평균매출액(y)과 평균 생활인구 값은 NaN
- 해당 경우 상권이 하나도 없어서 입지 점수를 산출할 수 없기에 빼는 것이 적합하지만 input으로 해당 도로명주소가 들어오면 오류가 날 수 있기에 0으로 처리

## 2.2 중복 도로명 주소

- 하나의 도로명주소에서 여러 개의 건물이 할당될 수 있음
  - 도로명주소를 그룹으로 하여 중복된 도로명주소의 건물들의 컬럼 정보를 평균
- 최종 78835개의 도로명 주소를 모델링에 활용

data1																
	도로명주소	위도	경도	haver_dis_subway	haver_dis_school	haver_dis_mart	mean_sales	mean_pop	연면적(m²)	자주식주차장	기계식주차장	승강기개수	층수	제1종근린생활시설	제2종근린생활시설	판매시설
0	서울특별시 강남구	37.517236	127.047325	0.531267	0.292754	0.348237	3.518745e+10	9939498.4	823.103670	3	1	0	4	0.545455	0.446970	0.007576
1	서울특별시 강남구 0	37.519082	127.019666	0.300327	0.609858	0.094908	7.587799e+10	16127592.5	457.606667	3	0	0	4	0.333333	0.666667	0.000000
2	서울특별시 강남구 14-3	37.517236	127.047325	0.531267	0.292754	0.348237	3.518745e+10	9939498.4	606.960000	4	0	0	4	0.000000	1.000000	0.000000
3	서울특별시 강남구 30	37.498875	127.040052	0.365576	0.820344	0.277399	9.419313e+10	25950123.2	1176.520000	5	0	1	6	1.000000	0.000000	0.000000
4	서울특별시 강남구 31-3	37.517236	127.047325	0.531267	0.292754	0.348237	3.518745e+10	9939498.4	642.930000	4	0	0	5	0.000000	1.000000	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
78830	서울특별시 중랑구 중랑천로5길 13	37.590036	127.072308	1.357060	0.284616	0.958115	0.000000e+00	0.0	347.440000	0	0	0	4	1.000000	0.000000	0.000000
78831	서울특별시 중랑구 중랑천로5길 5	37.590131	127.072864	1.310062	0.238341	0.976836	0.000000e+00	0.0	319.500000	0	0	0	5	1.000000	0.000000	0.000000
78832	서울특별시 중랑구 중랑천로5길 7	37.590059	127.072713	1.322061	0.249798	0.966523	0.000000e+00	0.0	466.660000	0	0	0	6	1.000000	0.000000	0.000000
78833	서울특별시 중랑구 중랑천로5길 9	37.590051	127.072615	1.330507	0.258169	0.964003	0.000000e+00	0.0	455.260000	0	0	0	6	1.000000	0.000000	0.000000
78834	서울특별시 중랑구 중랑천로8길 4	37.590653	127.073555	1.245052	0.203945	1.020726	3.303573e+10	13054920.5	187.090000	0	0	0	4	1.000000	0.000000	0.000000
78835 rows × 16 columns																

2.3 입지점수 예측 모델 개발

단순선형회귀, 라쏘회귀, 릿지회귀, XGBoost, LightGBM, K-nearest Neighbors 회귀모델을 고려하여 비교

성능	단순선형회귀	라쏘	릿지	XGBoost	LightGBM	Knn
r2-score	0.2654	0.2654	0.2654	0.79550	0.6878	0.87594
mae	36520919362	36520919362	36520745612	21500392828	25816916270	6887593272
mse	5.753138795414 207e+21	5.753138795 442346e+21	5.753178453 315655e+21	1.601665333 9250584e+21	2.444480046 997519e+21	9.716139353 286528e+20
rmse	75849448221	75849448221	75849709645	40020811260	49441683294	31170722406

가장 높은 성능을 보인 K-nearest Neighbors 회귀모델을 선택해서 학습을 진행하고 예측값을 산출

data_score																	
	도로명주소	위도	경도	haver_dis_subway	haver_dis_school	haver_dis_mart	mean_sales	mean_pop	연면적(m <sup>2</sup> )	자주식주차장	기계식주차장	승강기개수	층수	제1종근린생활시설	제2종근린생활시설	판매시설	y_score
0	서울특별시 강남구	37.517236	127.047325	0.531267	0.292754	0.348237	3.518745e+10	9939498.4	823.103670	3	1	0	4	0.545455	0.446970	0.007576	3.243109e+10
1	서울특별시 강남구 0	37.519082	127.019666	0.300327	0.609858	0.094908	7.587799e+10	16127592.5	457.606667	3	0	0	4	0.333333	0.666667	0.000000	5.668244e+10
2	서울특별시 강남구 14-3	37.517236	127.047325	0.531267	0.292754	0.348237	3.518745e+10	9939498.4	606.960000	4	0	0	4	0.000000	1.000000	0.000000	3.243109e+10
3	서울특별시 강남구 30	37.498875	127.040052	0.365576	0.820344	0.277399	9.419313e+10	25950123.2	1176.520000	5	0	1	6	1.000000	0.000000	0.000000	9.419313e+10
4	서울특별시 강남구 31-3	37.517236	127.047325	0.531267	0.292754	0.348237	3.518745e+10	9939498.4	642.930000	4	0	0	5	0.000000	1.000000	0.000000	3.243109e+10



Input: 도로명 주소

Output: 입지점수, 점수별 히트맵, 건물별 등급 지도

```
def analyze_score():
    location = input('도로명 주소 입력: ')

    score = data_score[data_score['도로명주소'] == location]['y_score']
    score = float(score)

    col = data_score.loc[data_score[data_score['도로명주소'] == location]['color'].index[0], 'color']

    latitude, longitude = data_score[data_score['도로명주소'] == location]['위도'], data_score[data_score['도로명주소'] == location]['경도']

    global seoul

    seoul = folium.Map(location=[latitude, longitude], zoom_start=16) #확대정도 지정

    folium.Marker([latitude, longitude],
                  popup=location,
                  tooltip=location,
                  icon=folium.Icon('red', icon='star'),
                  ).add_to(seoul)

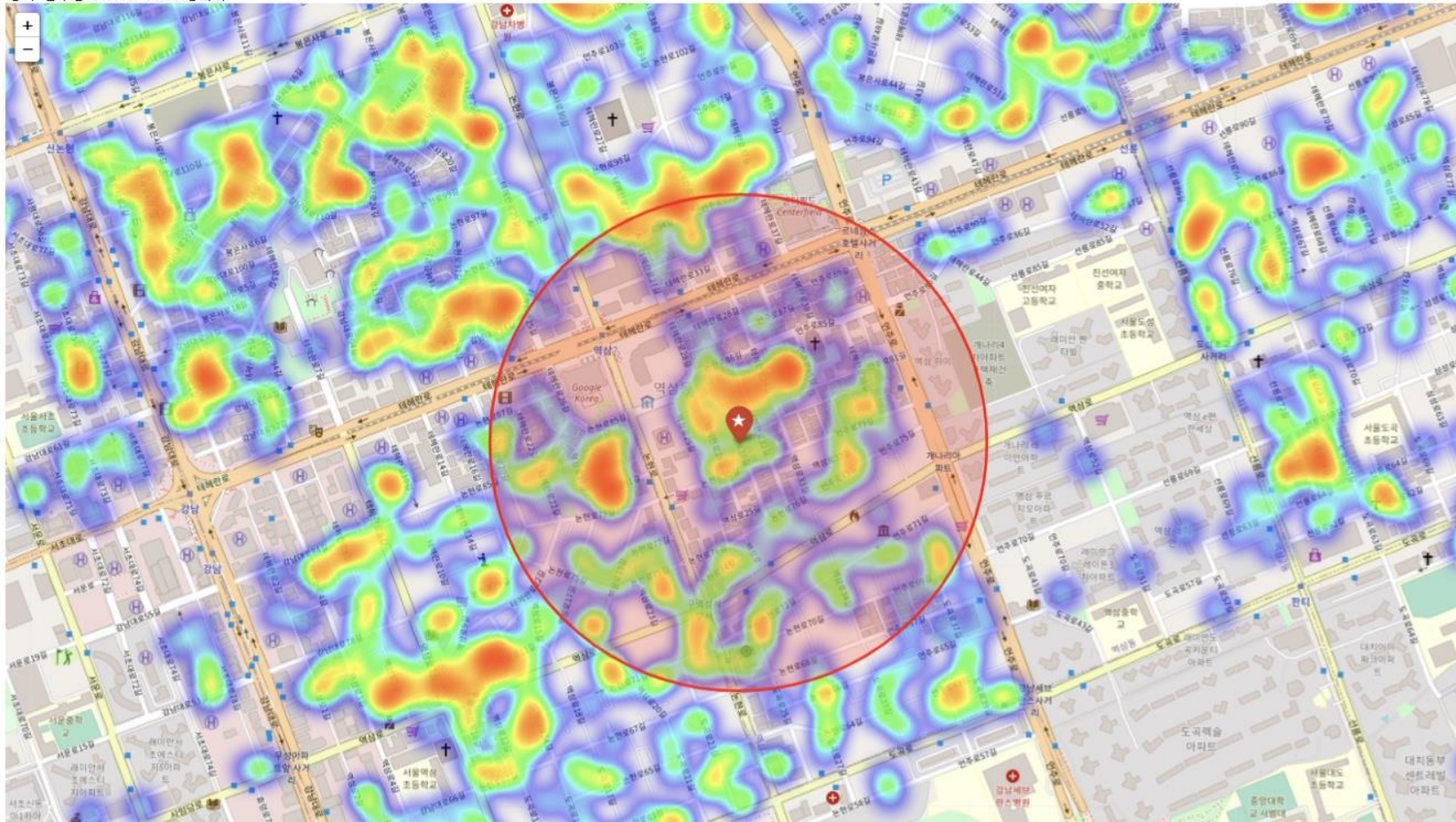
    seoul.add_child(plugins.HeatMap(zip(data_score['위도'],
                                       data_score['경도'],
                                       data_score['y_score']), radius=20))

    folium.Circle([latitude, longitude],
                  color=col,
                  fill_color=col,
                  radius = 500,
                  tooltip=location).add_to(seoul)

    print('입지 점수는 {} 입니다'.format(score))

    return seoul
```

도로명 주소 입력 :: 서울특별시 강남구 30  
입지 점수 94193134920.0 입니다

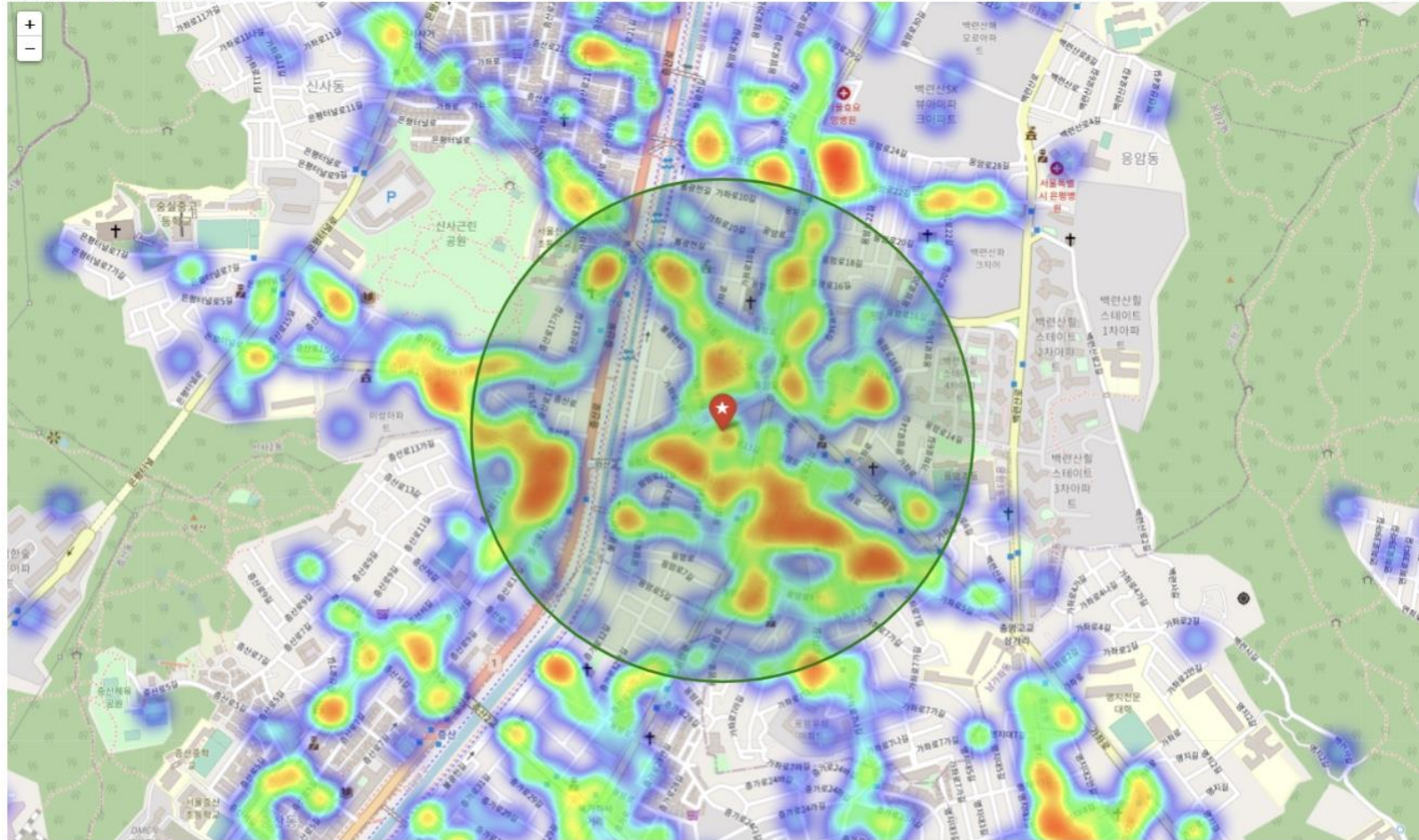


최상 - 빨강 ●  
상 - 분홍 ●  
하 - 녹색 ●  
최하 - 남색 ●

- 도로명주소: 서울특별시 강남구 30
- 등급: 최상 ●
- 입지점수: 94193134920



도로명 주소 입력 :: 서울특별시 은평구 응암로13길 9-1  
입지 점수는 17356960526.2 입니다



최상 - 빨강 ●  
상 - 분홍 ●  
하 - 녹색 ●  
최하 - 남색 ●

- 도로명주소: 서울특별시 은평구 응암로13길 9-1
- 등급: 하 ●
- 입지점수: 17356960526

4

활용방안

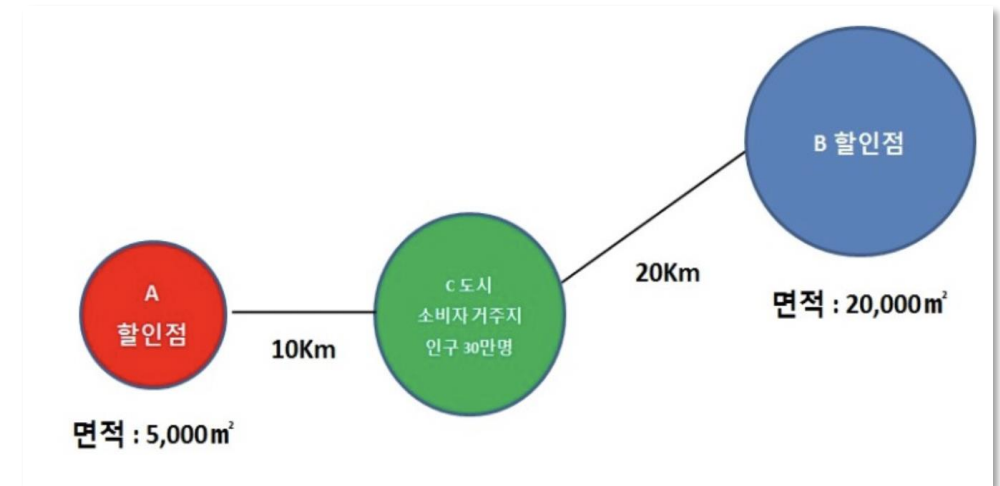
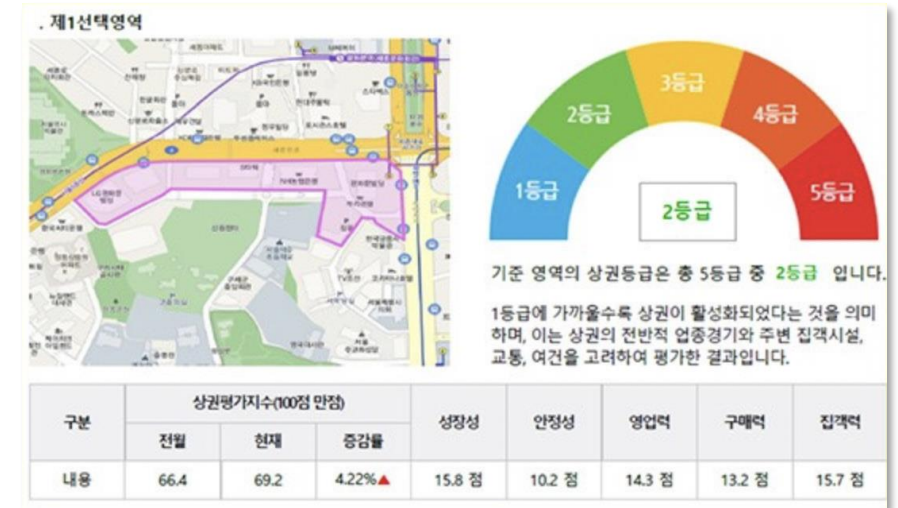
# 활용방안

## • 상권 평가 지수

: 상권 내 음식, 서비스 소매 등 전반적인 업종 등을 기타 설비 서비스를 종합하여 산출한 등급

## • 허프의 확률 모델의 확장

: 면적에 비례하고 이동거리의 공간 마찰 계수의 승에 반비례하는 흡인력을 전제



**감사합니다**

---