



살기 좋은 동네 찾기 서비스

가구원수에 따른 살기 좋은 동네의 기준은 어떤 것인가?

요기 오때



Content

S

I 분석 배경

1) 분석 배경

2) 분석 환경

II 데이터 전처리

III 데이터 분석

IV 분석 결과

1) 시각화

2) 플랫폼

V 기대효과

VI 한계점

분석 배경



서울시 구별 집값 순위

순위	지역	평당가	25평대	32평대	순위	지역	평당가	25평대	32평대
1위	서울 강남구	3648만	9.0억	12.4억	13위	서울 양천구	1858만	4.9억	5.8억
2위	서울 서초구	3013만	8.8억	10.6억	14위	서울 서대문구	1746만	4.2억	5.6억
3위	서울 용산구	2657만	6.5억	8.5억	15위	서울 강서구	1732만	4.1억	5.7억
4위	서울 송파구	2534만	7.3억	8.2억	16위	서울 동대문구	1636만	4.5억	5.0억
5위	서울 성동구	2342만	6.0억	7.7억	17위	서울 관악구	1597만	3.7억	4.6억
6위	서울 마포구	2192만	6.3억	7.0억	18위	서울 은평구	1551만	4.0억	5.0억
7위	서울 광진구	2143만	5.8억	7.1억	19위	서울 성북구	1529만	3.8억	4.8억
8위	서울 중구	2099만	5.0억	6.7억	20위	서울 구로구	1459만	3.7억	4.7억
9위	서울 종로구	2051만	5.0억	6.6억	21위	서울 노원구	1456만	3.5억	4.3억
10위	서울 동작구	2048만	5.2억	6.8억	22위	서울 중랑구	1385만	3.1억	4.1억
11위	서울 영등포구	2023만	4.7억	6.0억	23위	서울 금천구	1352만	2.8억	4.5억
12위	서울 강동구	2014만	5.2억	6.0억	24위	서울 강북구	1273만	2.9억	4.1억
					25위	서울 도봉구	1237만	3.0억	3.9억

분석 환경



사이킷 런
(sklearn)



R Studio



Python 3.7



Pandas



Numpy



맷플롯립
(matplotlib)

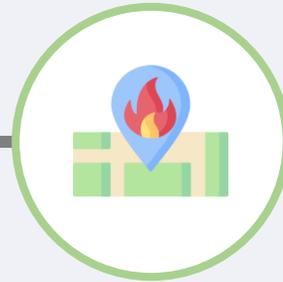
데이터 전처리



데이터 수집



데이터 정제



API 위도, 경도



가중치

[2-1] 운영여부 확인

1) 결측치 처리

- 필요 없는 컬럼 삭제
- 폐업일자 'NaN'값 확인
 - 영업중 : '0'으로 치환
 - 폐업 : '최종데이터갱신일자'으로 치환
- 폐업일자 '20170101' 이전 삭제

코드 예시(1)

```
1 hosp['폐업일자'] = hosp['폐업일자'].fillna(0)
2
3 hosp
```

코드 예시(2)

```
df2 = df1.drop(df1[(df1['폐업일자'] != 0) & (df1['폐업일자'] <= 20170101)].index)
df2
```

■ 통합 데이터 예시

년	상세영업 상태명	폐업 일자	휴업시작 일자	휴업종료 일자	인허가일 자	사업장명	지번주소	도로명주소
2017	영업중	0	0	0	19601124	서울특별시서북병원	서울특별시 은평구 역촌 동 산 31-1	서울특별시 은평구 갈현로7길 49-1
2017	영업중	0	0	0	19610515	가톨릭대학교 은평 성모병원	서울특별시 은평구 진관 동 55	서울특별시 은평구 진관3로 15- 35
2017	영업중	0	0	0	19620201	국립정신건강센터	서울특별시 광진구 중곡 동 30-1	서울특별시 광진구 용마산로 127
2017	영업중	0	0	0	19660107	삼육서울병원	서울특별시 동대문구 휘 경동 283-1	서울특별시 동대문구 망우로 82
2017	영업중	0	0	0	19660107	청량리정신병원	서울특별시 동대문구 청 량리동 46-3	서울특별시 동대문구 제기로 144
...
2019	영업중	0	0	0	20191030	엘치과병원	서울특별시 동대문구 답 십리동 94-1	서울특별시 동대문구 전농로 110
2019	영업중	0	0	0	20191031	미림한방병원	서울특별시 강남구 논현 동 2	서울특별시 강남구 도산대로 110
2019	영업중	0	0	0	20191213	서울스마트요양병 원	서울특별시 양천구 신정 동 1268	서울특별시 양천구 중앙로 181
2019	영업중	0	0	0	20191213	연세바로척병원	서울특별시 성동구 성수 동2가 279	서울특별시 성동구 성수동2가 아 차산로13길 38
2019	영업중	0	0	0	20191226	치유본한방병원	서울특별시 관악구 신림 동 1638-1	서울특별시 관악구 신원로 35

[2-2] 운영여부 확인

2) 연도별 구분

- 2017 ~ 2019 연도별
- 연도별 시설 운영 여부 및 개수 확인

코드 예시(2017년)

2017년

```
1 # 인허가 일자 20180101 이전까지 불러오기
2 df_2017 = df.drop(df[df['인허가일자'] >= 20180101].index)
3
4 df_2017
```

코드 예시(2018년)

2018년

```
1 # 인허가 일자 20190101 이전 / 폐업일자 20180101이후 까지 불러오기
2 df_2018 = df.drop(df[(df['인허가일자'] >= 20190101) | ((df['폐업일자'] != 0) & (df['폐업일자'] < 20180101))].index)
3
4 df_2018
```

전처리 데이터 예시

년	상세영업 상태명	폐업 일자	휴업시작 일자	휴업종료 일자	인허가일 자	사업장명	지번주소	도로명주소
2017	영업중	0	0	0	19601124	서울특별시서북병원	서울특별시 은평구 역촌 동 산 31-1	서울특별시 은평구 갈현로7길 49-1
2017	영업중	0	0	0	19610515	가톨릭대학교 은평 성모병원	서울특별시 은평구 진관 동 55	서울특별시 은평구 진관3로 15- 35
2017	영업중	0	0	0	19620201	국립정신건강센터	서울특별시 광진구 중곡 동 30-1	서울특별시 광진구 용마산로 127
2017	영업중	0	0	0	19660107	삼육서울병원	서울특별시 동대문구 휘 경동 283-1	서울특별시 동대문구 망우로 82
2017	영업중	0	0	0	19660107	청량리정신병원	서울특별시 동대문구 청 량리동 46-3	서울특별시 동대문구 제기로 144
...
2019	영업중	0	0	0	20191030	엘치과병원	서울특별시 동대문구 답 십리동 94-1	서울특별시 동대문구 전농로 110
2019	영업중	0	0	0	20191031	미림한방병원	서울특별시 강남구 논현 동 2	서울특별시 강남구 도산대로 110
2019	영업중	0	0	0	20191213	서울스마트요양병 원	서울특별시 양천구 신정 동 1268	서울특별시 양천구 중앙로 181
2019	영업중	0	0	0	20191213	연세바로척병원	서울특별시 성동구 성수 동2가 279	서울특별시 성동구 성수동2가 아 차산로13길 38
2019	영업중	0	0	0	20191226	치유본한방병원	서울특별시 관악구 신림 동 1638-1	서울특별시 관악구 신원로 35

[3] 업종 분리

1) 휴게 음식점

- 업태구분 명에 따라 분리
 - 편의점, 카페, 제과제빵, 백화점
- 기타 휴게음식점에서 분리

2) 집단 급식소

- 업태구분 명에 따라 분리
 - 어린이집, 유치원

■ 전처리 데이터 예시

사업장명	업태구분명
동선유치원	학교
오송유치원	어린이집
대학교구유지재단 천호유치원	어린이집
라온유치원	어린이집
서울청림유치원	어린이집
...	...
영산유치원	어린이집
신지유치원	어린이집
우촌유치원	어린이집
상원유치원	어린이집
미파유치원	어린이집



사업장명	업태구분명
동선유치원	유치원
오송유치원	유치원
대학교구유지재단 천호유치원	유치원
라온유치원	유치원
서울청림유치원	유치원
...	...
영산유치원	유치원
신지유치원	유치원
우촌유치원	유치원
상원유치원	유치원
미파유치원	유치원

[4] 같은 주소 처리

1) 백화점

- 주소 같은 백화점 1개로 처리

53461	20200909	폐업	20200917	서울특별시 구도구 신노림중 692 니큐르 시티	서울특별시 구도구 성진로 662 시아2층 (신노림 등 디큐브시티)	공략	백화점
53580	20200923	영업	0	서울특별시 강남구 압구정동 429 현대백화점본점	서울특별시 강남구 압구정로 165 현대백화점본점 지하1층 (압구정동)	영양사의 반찬가게(한시적)	백화점
53581	20200923	영업	0	서울특별시 강남구 삼성동 159-7 현대백화점	서울특별시 강남구 테헤란로 517 현대백화점 지하1층 (삼성동)	영양사의 반찬가게(한시적)	백화점
53593	20200923	영업	0	서울특별시 양천구 목동 916 현대하이퍼 리온 지하2층	서울특별시 양천구 목동동로 257 지하2층 (목동 현대하이퍼리온)	영양사의 반찬가게	백화점

1335 rows x 7 columns



48479	20200114	폐업	20200220	서울특별시 송파구 신천동 29번지 롯데캐슬드림타워엔드롯데캐슬드림	서울특별시 송파구 올림픽로 300 롯데캐슬드림 지하1층 (신천동)	교토마블	백화점	서울특별시 송파구 올림픽로	300
32074	20200416	영업	0	서울특별시 영등포구 당산동2가 161-11번지	서울특별시 영등포구 양산로 90-5 1층 (당산동2가)	이십사캐이커피 (24k coffee)	백화점	서울특별시 영등포구 양산로	90-5
51610	20200702	폐업	20200801	서울특별시 중구 황학동 2545 롯데캐슬베네치아	서울특별시 중구 청계천로 400 지하2층 (황학동 롯데캐슬베네치아)	(주)에그탐탐	백화점	서울특별시 중구 청계천로	400

86 rows x 11 columns

2) 영화관

- 관만 다른 같은 영화관 1개로 처리

126	영업중	0	서울특별시 성동구 성수동1가 656-335번지	메가박스중앙(주) 성수지점 더부티크 102호	20191213163233	영화관
130	영업중	0	서울특별시 성동구 성수동1가 656-335번지	메가박스중앙(주) 성수지점 MX관	20191213162857	영화관
131	영업중	0	서울특별시 성동구 성수동1가 656-335번지	메가박스중앙(주) 성수지점 2관	20191213162807	영화관
127	영업중	0	서울특별시 성동구 성수동1가 656-335번지	메가박스중앙(주) 성수지점 더부티크 101호	20191213163035	영화관
128	영업중	0	서울특별시 성동구 성수동1가 656-335번지	메가박스중앙(주) 성수지점 5관	20191213162959	영화관

655 rows x 7 columns



372	영업중	0	서울특별시 강서구 동진동 639-23번지 3~9층	CGV 동진 4관	20191205111621	영화관	20190125	서울특별시 강서구 동진동	639-23번지
321	영업중	0	서울특별시 마포구 동교동 160-4번지 아일렉스	메가박스중앙(주) 홍대지점 2관	20191218104848	영화관	20190923	서울특별시 마포구 동교동	160-4번지
129	영업중	0	서울특별시 성동구 성수동1가 656-335번지	메가박스중앙(주) 성수지점 4관	20191213162927	영화관	20191106	서울특별시 성동구 성수동1가	656-335번지

122 rows x 15 columns

[5] API이용 좌표, 주소 구하기

1) 카카오 API

- 주소 > 좌표 변환 사용
- 플랫폼 구현 시 시설 위치 나타내기 위해 필요

```
def get_xy(addr):
    url = 'https://dapi.kakao.com/v2/local/search/address.json?query=' + addr
    headers = {"Authorization": API KEY}
    results = requests.get(urlparse(url).geturl(), headers=headers)
    json_obj = results.json()

    global val

    if len(json_obj['documents']) == 0:
        val = ['0', '0']

    else:
        for document in json_obj['documents']:
            val = [document['y'], document['x']]

    return val
```

2) vworld API

- 좌표 > 주소 변환 사용
- 카카오 API로 변환 안되는 데이터에 사용

데이터 예시

사업장명	지번주소	도로명주소	위도	경도
서울특별시서북병원	서울특별시 은평구 역촌동 산 31-1	서울특별시 은평구 갈현로7길 49-1	37.604297	126.903688
가톨릭대학교 은평성모병원	서울특별시 은평구 진관동 55	서울특별시 은평구 진관3로 15-35	37.642175	126.918375
국립정신건강센터	서울특별시 광진구 중곡동 30-1	서울특별시 광진구 용마산로 127	37.565004	127.085767
삼육서울병원	서울특별시 동대문구 휘경동 283-1	서울특별시 동대문구 망우로 82	37.589726	127.062542
청량리정신병원	서울특별시 동대문구 청량리동 46-3	서울특별시 동대문구 제기로 144	37.586070	127.049088
...
엘치과병원	서울특별시 동대문구 답십리동 94-1	서울특별시 동대문구 전농로 110	37.574573	127.057846
미림한방병원	서울특별시 강남구 논현동 2	서울특별시 강남구 도산대로 110	37.516176	127.021058
서울스마트요양병원	서울특별시 양천구 신정동 1268	서울특별시 양천구 증양로 181	37.512742	126.855774
연세바로척병원	서울특별시 성동구 성수동2가 279	서울특별시 성동구 성수동2가 아차산로13길 38	37.546292	127.062173
치유본한방병원	서울특별시 관악구 신림동 1638-1	서울특별시 관악구 신원로 35	37.482227	126.928894

[6] 2020년 예측

1) 과거 데이터 통합

- 2020년 데이터가 없는 요소에 대해 2015년 ~ 2019년 데이터를 바탕으로 증감률에 따라 예측
- 매년 증감률에 대한 평균값을 2019년 데이터에 적용

```
# 증감률 구하기
office = list(file['관서'].unique())
crime = list(file['죄종'].unique())

per_list=[]
office_list = []
crime_list = []

for i in range(len(office)) :
    for j in range(len(crime)) :
        df = file[(file['관서']==office[i]) & (file['죄종']==crime[j])].reset_index(drop=True)
        crime_list.append(((df['건수'][len(df['건수'])-1]-df['건수'][0])/sum(df['건수'])))
```

2017 ~ 2020 데이터 예시

	구분	2015	2016	2017	2018	2019	2016-2015	2017-2016	2018-2017	2019-2018	비율 평균	2020
0	종로구	8,988	8,988	8,986	9,220	9,265	0.000000	-0.022252	2.604051	0.488069	0.767467	9336
1	중구	8,651	8,621	8,641	8,694	8,665	-0.346781	0.231992	0.613355	-0.333563	0.041251	8668
2	용산구	6,977	7,206	7,291	7,248	7,637	3.282213	1.179573	-0.589768	5.366998	2.309754	7813
3	성동구	5,299	5,712	5,662	5,780	6,050	7.793923	-0.875350	2.084069	4.671280	3.418481	6256
4	광진구	4,981	4,925	4,852	4,774	4,616	-1.124272	-1.482234	-1.607585	-3.309594	-1.880921	4529
5	동대문구	7,890	8,019	8,096	8,149	8,211	1.634981	0.960219	0.654644	0.760830	1.002669	8293
6	종랑구	5,392	5,682	5,698	5,729	5,731	5.378338	0.281591	0.544051	0.034910	1.559722	5820
7	성북구	5,301	5,288	5,373	5,301	5,584	-0.245237	1.607413	-1.340034	5.338615	1.340190	5658
8	강북구	4,360	4,440	4,445	4,525	4,339	1.834862	0.112613	1.799775	-4.110497	-0.090812	4335
9	도봉구	4,400	4,405	4,464	4,471	4,556	0.113636	1.339387	0.156810	1.901141	0.877744	4595
10	노원구	8,745	9,267	9,308	9,270	9,510	5.969125	0.442430	-0.408251	2.588997	2.148075	9714
11	은평구	6,835	6,455	6,554	6,428	6,369	-5.559620	1.533695	-1.922490	-0.917859	-1.716569	6259

[6] 가중치 값 부여

1) 피어슨상관계수

- 가구원수 별 크게 영향을 미치는 요소를 분석하기 위해 피어슨 상관계수를 이용

범위	상관관계
$-1.0 < R < -0.7$	강한 음적 선형 관계
$-0.7 < R < -0.3$	뚜렷한 음적 선형 관계
$-0.3 < R < -0.1$	약한 음적 선형 관계
$-0.1 < R < 1.0$	거의 무시될 수 있는 선형 관계
$+1.0 < R < +0.3$	약한 양적 선형 관계
$+0.3 < R < +0.7$	뚜렷한 양적 선형 관계
$+0.7 < R < +1.0$	강한 양적 선형 관계

* 피어슨 상관계수 지표

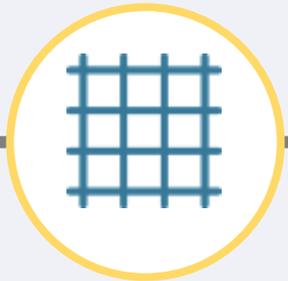
2) 증감률에 따른 상관계수

	요소	1인 가구 상관계수	2인 가구 상관계수	4인 가구 상관계수
0	cctv	0.461408	0.327351	0.206458
1	어린이집/유치원	-0.200600	0.092261	0.333847
2	초중고등학교	0.281919	0.350073	0.321225
3	대학교	0.000000	0.000000	0.000000
4	백화점	0.076570	-0.134271	0.110704
5	버스정류장	-0.114068	-0.240309	-0.005293
6	병원	0.065630	0.058194	0.153373
7	영화관	0.075359	-0.038948	-0.211472
8	음식점	0.018174	0.061029	-0.250236
9	제과제빵	-0.373063	-0.135929	-0.130313
10	지하철역	0.022836	0.109487	0.037033
11	카페	-0.022973	-0.019614	-0.238499
12	편의점	0.296920	0.267358	-0.104014

3) 가중치 값 부여

	요소	1인 가구	2인 가구	4인 가구
0	cctv	3.5	3.5	3.0
1	어린이집/유치원	1.0	2.0	3.5
2	초중고등학교	3.0	3.5	3.5
3	대학교	2.5	2.5	2.5
4	백화점	2.0	1.0	3.0
5	버스정류장	1.0	1.0	2.0
6	병원	2.0	2.0	3.0
7	영화관	2.0	2.0	1.0
8	음식점	2.0	2.0	1.0
9	제과제빵	1.0	1.0	1.0
10	지하철역	2.0	3.0	2.0
11	카페	2.0	2.0	1.0
12	편의점	3.0	3.0	1.0

데이터 분석



그리드화



군집분석



구별 그리드 개수



점수 부여

[1] 그리드화

1) 그리드 나누기

- 행정동을 그리드화 하여 분할
 행정동 전체를 하나의 유형으로 판단하기에 행정동 내에서도 물리적 환경요인들이 차이가 있을 수 있기 때문에 400m*400m 구역으로 행정동을 분할

또한 공간적 분석단위의 생태적 오류를 방지하기 위하여 분석단위를 400m×400m 격자로 나누었다. 범죄분석 관련 연구에서 300m~400m 격자의 크기가 범죄발생의 변이를 적정하게 잡아냄을 관련 연구들(Stucky, et al., 2012; Spader, et al., 2015; Robinson et al., 2016)은 보여주고 있다.

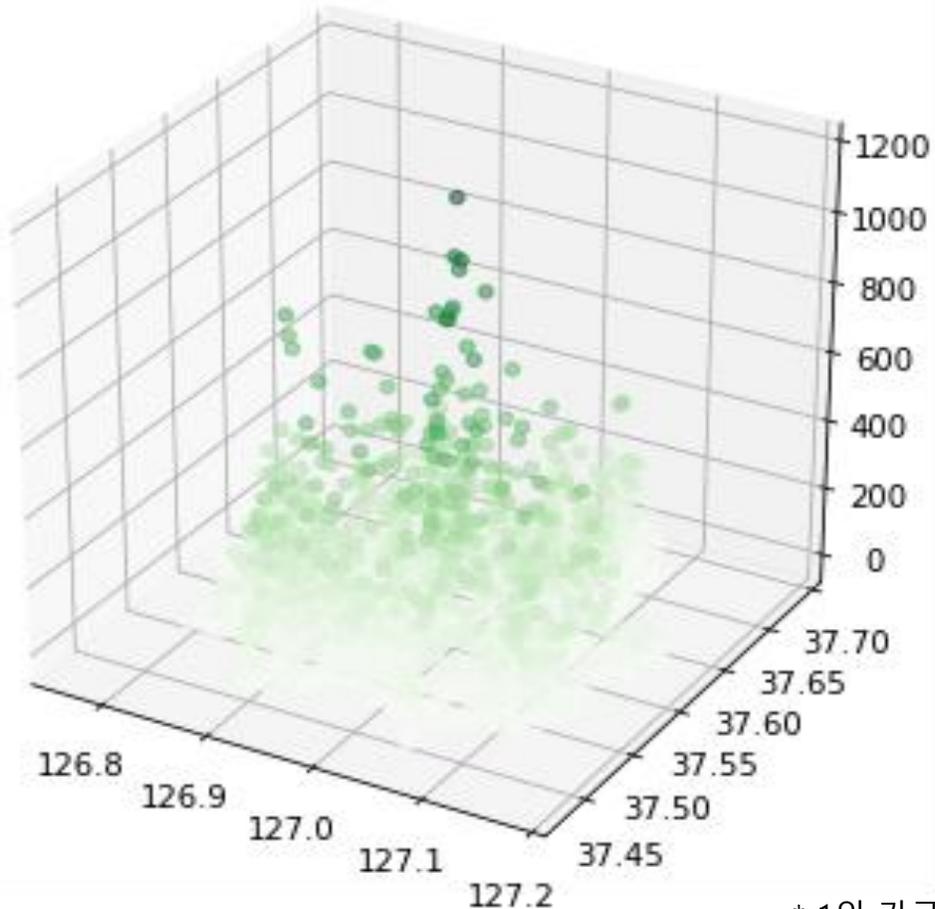
* 학회지 논문

2) 구역별 요

- 12^슈 요소 : 백화점, 버스정류장, 병원, 영화관, 제과제빵, 지하철역, 카페, 편의점, 음식점, 어린이집/유치원, 초중고등학교, 대학교



[2] 군집분석



* 1인 가구

1) 유클리디안 거리 사용

- 12개 요소에 대한 위도 경도 값 정확도 떨어짐
- 12개의 요소의 개수를 좌표로 두고 원점으로 부터의 유클리디안의 거리 공식을 사용하여 좌표 별 거리를 구하여 군집으로 나눈다

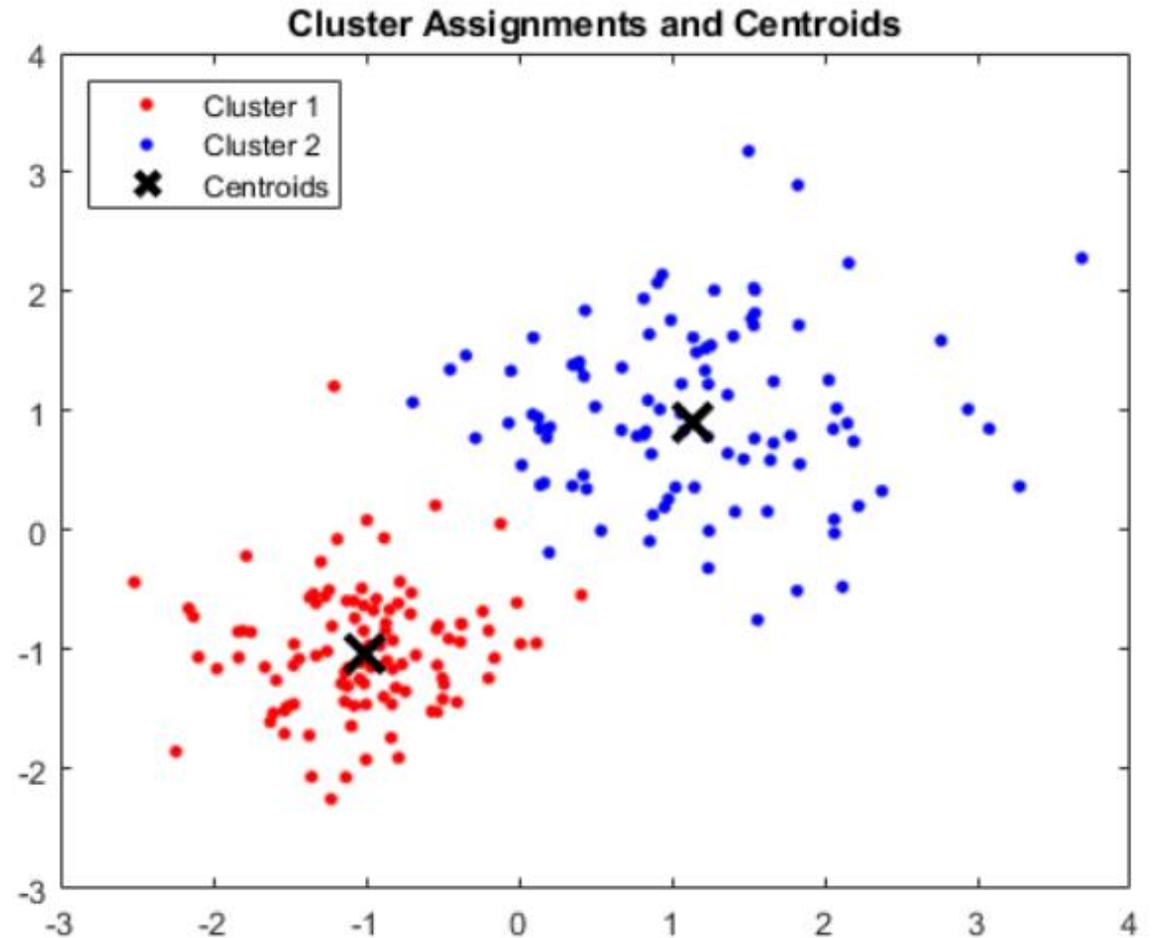
	X좌표(경도)	Y좌표(위도)	거리
0	126.7595 ~ 126.7654	37.5535 ~ 37.557599999999994	1.000000
1	126.7654 ~ 126.771300000000001	37.5535 ~ 37.557599999999994	8.246211
2	126.771300000000001 ~ 126.777200000000001	37.5494 ~ 37.5535	1.000000
3	126.777200000000001 ~ 126.7831	37.5453 ~ 37.5494	1.000000
4	126.777200000000001 ~ 126.7831	37.5535 ~ 37.557599999999994	1.000000

* Euclidean Distance 적용 데이터 예시

[2] 군집분석

K-Means

- k-평균 알고리즘(K-means algorithm)
- 주어진 데이터를 k개의 클러스터로 묶는 알고리즘
- 각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작



* K-Means 예시

[2] 군집분석

2) Label Encoding

- 범수형 데이터인 X좌표(경도), Y좌표(위도)를 분석에 용이하도록 수치형 데이터로 변경

3)

MinMaxScaler

- 거리, 점수 데이터의 경우 최대, 최소값 범위 차이가 매우 큼
 - 거리 : 1.0 ~ 472.65
 - 점수 : 1.0 ~ 570.0
- 모든 feature 값이 0~1사이에 있도록 데이터를 재조정

4) K-Means

```
kmeans = KMeans(n_clusters=9, init='k-means++',
                 max_iter=300, random_state=0).fit(data)
```

```
data['target'] = data.점수
data['cluster'] = kmeans.labels_
data['idx'] = data.index
```

* K-Means 코드 예시

- cluster 개수 : 9
- init='k-means++' : 초기 중심 설정 방식
- max_iter=300 : 최대 반복 횟수

[2] 군집분석

	X좌표(경도)	Y좌표(위도)	거리	점수	target	cluster	idx
0	0	31	0.000000	0.000000	0.000000	4	0
1	1	31	0.015364	0.019332	0.019332	4	1
2	2	30	0.000000	0.000000	0.000000	4	2
3	3	29	0.000000	0.000000	0.000000	4	3
4	3	31	0.000000	0.000000	0.000000	4	4
...
2048	70	33	0.055673	0.115993	0.115993	3	2048
2049	70	34	0.024949	0.059754	0.059754	3	2049
2050	70	35	0.000878	0.001757	0.001757	3	2050
2051	71	32	0.007596	0.017575	0.017575	3	2051
2052	71	33	0.004912	0.019332	0.019332	3	2052

2053 rows × 7 columns

* K-Means 데이터 예시

[2] 군집분석

5) 각 그리드별 cluster 번호 부여

- 각 위경도에 포함되는 cluster 번호 부여

6) 각 클러스터 당 12개 요소 평균치 계산

- 각 클러스터의 군집들이 가진 12개의 요소에 대한 평균치를 계산
- 군집들의 평균치를 알 수 있다

7) 클러스터 별 점수

- 총점 = sum(12개 요소 평균치 * 가구원수 별 가중치)

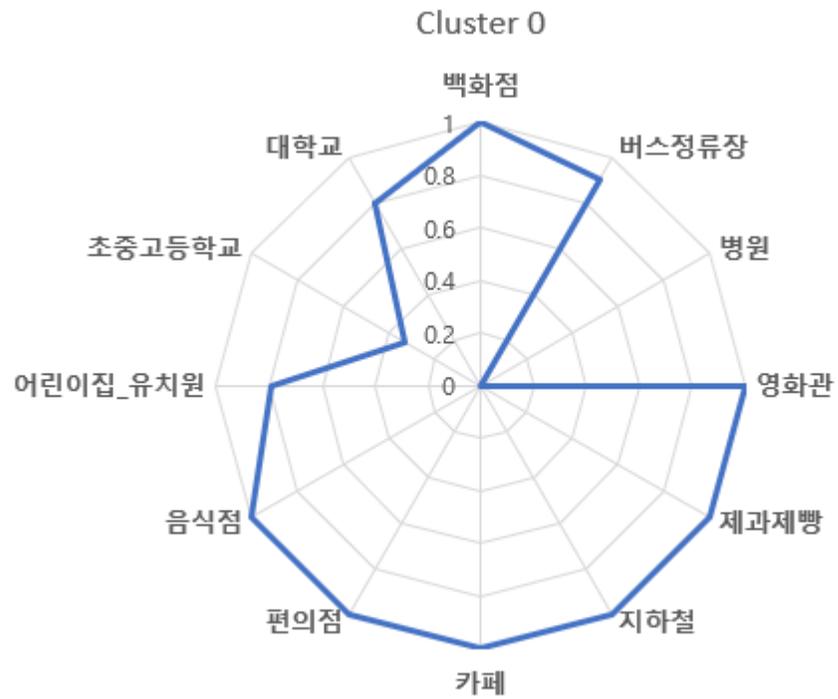
	백화점 개수	버스정류장 개수	병원 개 수	영화관 개수	제과제빵 개수	지하철 개수	카페 개수	편의점 개수	음식점 개 수	어린이집_유치 원 개수	초중고등학 교 개수	대학교 개수	cluster
0	0.030172	6.262931	0.146552	0.129310	0.685345	0.318966	12.232759	4.443966	74.594828	0.051724	0.573276	0.051724	0.0
1	0.018957	5.909953	0.284360	0.047393	0.251185	0.175355	6.985782	3.075829	39.630332	0.014218	0.649289	0.014218	1.0
2	0.015748	3.822835	0.299213	0.043307	0.149606	0.165354	5.708661	2.712598	36.830709	0.011811	0.673228	0.011811	2.0
3	0.012448	6.518672	0.244813	0.041494	0.103734	0.157676	4.556017	2.016598	29.228216	0.037344	0.788382	0.037344	3.0
4	0.000000	4.510204	0.251701	0.034014	0.061224	0.129252	6.489796	2.346939	32.231293	0.013605	0.646259	0.013605	4.0
5	0.005025	5.160804	0.160804	0.025126	0.140704	0.105528	4.211055	2.130653	23.924623	0.040201	0.527638	0.040201	5.0
6	0.029963	4.295880	0.359551	0.063670	0.438202	0.172285	9.029963	3.565543	53.790262	0.014981	0.468165	0.014981	6.0
7	0.000000	6.394737	0.258772	0.021930	0.140351	0.118421	7.039474	3.478070	40.201754	0.017544	0.640351	0.017544	7.0
8	0.010949	6.277372	0.328467	0.047445	0.485401	0.248175	7.894161	3.872263	51.609489	0.062044	0.755474	0.062044	8.0

* 1인 가구 각 cluster의 요소별 평균치 예시

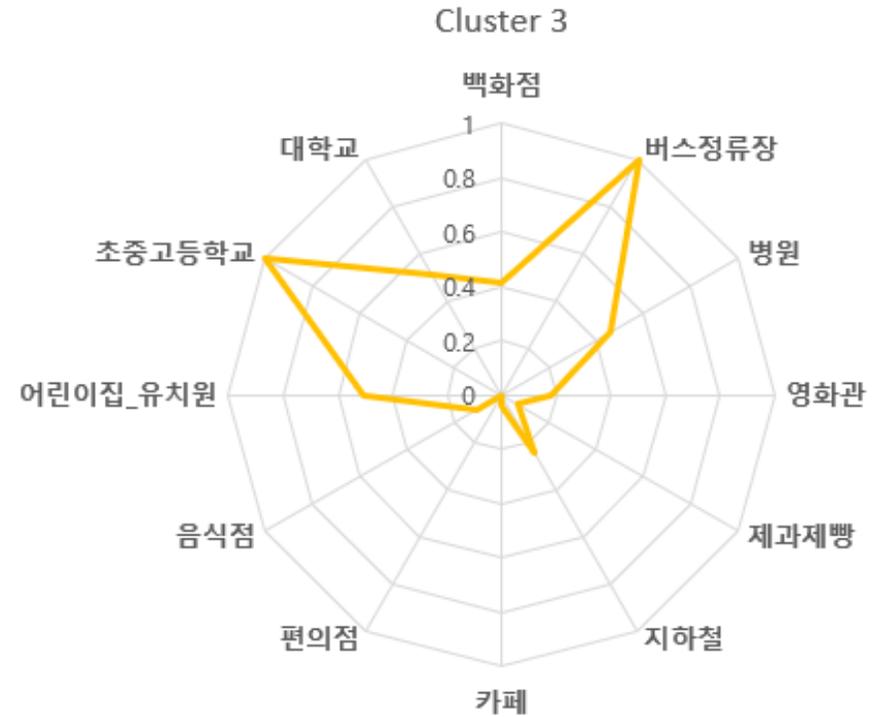
	X좌표(경도)	Y좌표(위도)	cluster	1인가구_총점	2인가구_총점	4인가구_총점
0	126.9011 ~ 126.90700000000001	37.5453 ~ 37.5494	0	197.086207	108.096983	84.393382
1	126.9011 ~ 126.90700000000001	37.5494 ~ 37.5535	0	197.086207	108.096983	84.393382
2	126.9011 ~ 126.90700000000001	37.5535 ~ 37.557599999999994	0	197.086207	108.096983	84.393382
3	126.9011 ~ 126.90700000000001	37.557599999999994 ~ 37.561699999999995	0	197.086207	108.096983	84.393382
4	126.90700000000001 ~ 126.91290000000001	37.5453 ~ 37.5494	0	197.086207	108.096983	84.393382

* 각 cluster별 점수 예시

[2] 군집분석



* 1인 가구 cluster_0의 요소별 평균치



* 1인 가구 cluster_3의 요소별 평균치

[3] 구별 그리드 개수

1) 구별 그리드 개수 계산

- 자치구 경계 그리드 위치 기반
- 동일한 X좌표를 가진 그리드 '번째'의 max, min값을 찾아 count

2) 가구원수별 점수 컬럼 추가

- 그리드 '번째'별로 가구원수별 점수 컬럼 추가

예) 그리드 3179번째

- 1인 가구일 때 그리드 점수 : 55
- 2인 가구일 때 그리드 점수 : 56
- 4인 가구일 때 그리드 점수 : 33

	구	그리드_개수
0	강남구	796
1	강동구	534
2	강북구	738
3	강서구	1079
4	관악구	646
5	광진구	374
6	구로구	528
7	금천구	407
8	노원구	1198
9	도봉구	797
10	동대문구	433

* 구별 그리드 개수 예시

그리드 번째	1인 가구 점수	2인 가구 점수	4인 가구 점수	
0	3177	0.0	0.0	0
1	3178	0.0	0.0	0
2	3179	55.0	56.0	33
3	3251	114.0	114.0	68
4	3252	15.0	15.0	8
...
71	4516	27.0	27.0	18
72	4517	6.0	6.0	12
73	4588	0.0	0.0	0
74	4589	3.0	3.0	6
75	4590	8.0	8.0	3

* 강남구 그리드 점수 예시

[3] 구별 그리드 개수

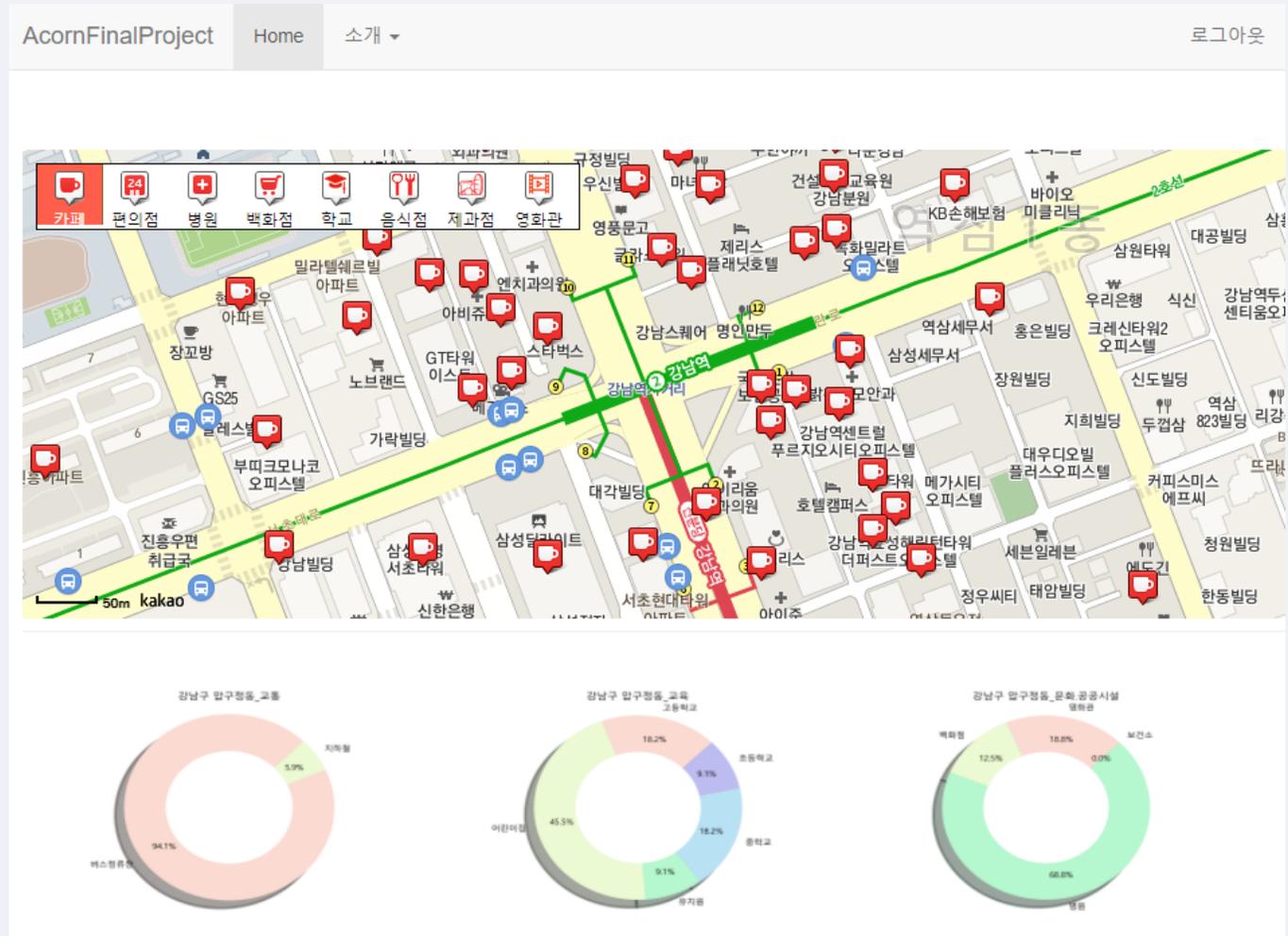
3) 자치구 가구별 점수 컬럼 추가

- 구별 그리드 점수 데이터 기반
- 1인 가구 점수 : 자치구별 1인 가구 점수 총점 / 자치구 그리드 개수
- 2인 가구 점수 : 자치구별 2인 가구 점수 총점 / 자치구 그리드 개수
- 4인 가구 점수 : 자치구별 4인 가구 점수 총점 / 자치구 그리드 개수

	자치구	1인 가구 점수	2인 가구 점수	4인 가구 점수
0	강남구	34.948052	35.006133	19.660173
1	강동구	33.902836	34.120798	20.079832
2	강북구	69.118037	69.514920	46.918435
3	강서구	9.635245	9.678642	5.510218
4	관악구	70.561056	70.916832	43.308911
5	광진구	27.700544	27.708711	15.927405
6	구로구	82.016633	82.280527	49.173631
7	금천구	71.477218	71.674261	41.972822
8	노원구	47.194227	47.670349	33.364932
9	도봉구	50.778993	51.154632	34.391685
10	동대문구	94.945413	95.493119	55.804587

* 자치구 가구별 점수 예시

플랫폼 http://192.168.0.12:5000



기대효과



- 가구원수에 따른 가구 위치 선정의 요소를 알 수 있다.



- 예산 및 요소에 따라 거주지 위치 선택이 가능하다.



- 소상공인, 자영업자, 기업들에게 입지 선정에 도움을 줄수있다.

한계점

- 데이터 부족
- 정확한 위도, 경도, 주소를 불러오는데 한계
- 모든 업체를 구하는데 한계
- 데이터 양이 많아 웹페이지 로딩 시간 지연



THANK YOU