# 상권영역 별 분석 바탕 기반 공원 조성 제안

18101997 이소현 19102023 정태경 21101434 박선영

github repo 주소: https://github.com/JTK01/DataMining

1 주제 선정 배경

2 분석 목적

3 데이터 획득 및 이해

4 분석

5 결과

6 기대효과

7 한계점 및 추후 개선 방안

# 1. 주제 선정 배경

#### 1. 주제 선정 배경

- 유현준 교수, 도심 속 공원 부족 문제점 지적

사람들이 모여 살면 갈등 생기기 마련 이를 해결하는 방법 중 하나 '공간의 구조를 바꾸는 것'



현재 도심 속 공간의 대부분은 이동하는 공간



도심 속 '공짜로 머무를 수 있는 공간' 필요

#### 1. 주제 선정 배경

#### - 기존 공원들의 문제점





공원과 상권의 공간이 분리되어 공원이 주는 '공간의 통합' 없음 공원은 접근성이 매우 중요, but 현재 공원들은 너무 멀리 떨어져 있음

# 2. 분석 목적

상권 데이터(인구 수, 연령, 점포 수, 아파트 단지 수)를 feature로 Clustering

Cluster 별 특징에 맞는 공원 구성안 제시

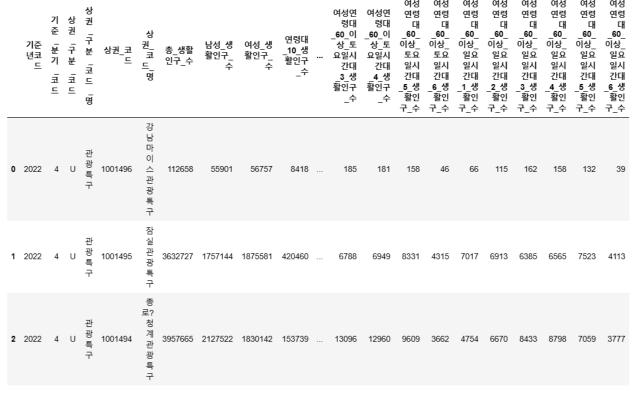
- 데이터 획득 방법: 서울 열린 데이터 광장 > 골목상권정보 > 상권 데이터
- 사용할 데이터 파일 1) 상주 인구 데이터.csv
  - 2) 생활 인구 데이터 .csv
  - 3) 직장 인구 데이터.csv
  - 4) 점포 데이터.csv
  - 5) 아파트 데이터.csv
  - 6) 상권영역 데이터.csv
- 2017년 1분기 ~ 2022 4분기 데이터
   ⇒ 이 중 2022 4분기 데이터만 사용

## 1) 상주인구 데이터

#	Column	Non-Null Count Dtype
		00470
0	기운_년_코드	39178 non-null int64 39178 non-null int64
ı	기운_문기_피드	391/8 non-null int64
2	상권_구문_되느	39178 non-null object 39178 non-null object
3	상권_구운_코드_병	39178 non-null object
4	상권 코드	39178 non-null int64
5	총 상주인구 수 남성 상주인구 수	39178 non-null int64
6	남성 상수인구 수	39178 non-null int64
7	여성 상수인구 수	391/8 non-null int64
8	연령대 10 상주인구	
9	연령대 20 상주인구	
10	연령대 30 상주인구	
11	연령대 40 상주인구	
12	연령대 50 상주인구	
13	연령대 60 이상 상취	
14	남성연령대 10 상주	인구 수 - 39178 non-null int64
15	남성연령대 20 상주	인구 수 39178 non-null int64
16	남성연령대 30 상주	인구 수 39178 non-null int64
17		인구 수 39178 non-null int64
18		인구 수 39178 non-null int64
19		상주인구 수 39178 non-null int64
20	여성연령대 10 상주	인구 수 39178 non-null int64
21	여성연령대 20 상주	인구 수 39178 non-null int64
22	여성연령대 30 상주	인구 수 39178 non-null int64 인구 수 39178 non-null int64
23	여성연령대 40 상주	인구 수 39178 non-null int64
24		인구 수 39178 non-null int64
25		상주인구 수 39178 non-null int64
26	총 가구 수	39178 non-null int64
27	아파트 가구 수	39178 non-null int64
28	비 아파트 가구 수	39178 non-null int64
29	상권 코드 명	39178 non-null object
		• •

_	기준_ 년_코 드	기 준_ 분 기_ 코 드	상 권_구 분_코 드	상 권_구 분_코 드_명	5 E	총 상 주인 구 수	남성 상주 인구 수	여성 상주 인구 수	연령 대 10 상주 인구 수	연령대 20 상 주인구 <sup></sup> 수	여성 연령 대 10 상주 인구 수	여성연 령대 20 상주인 구 수	여성연 령대 30 상주인 구 수	여성 연령 대 40 상주 인구 수	여성 연령 대 <b>50</b> 상주 인구 수	여성연 령대 60 이 상 상 주인구 수	총 가 구 수	아 파 트 가 구 수	비 아 파트 가구 수	상권 코드 명
0	2022	4	U	관 광 투 구	1001496	29	16	13	2	4	2	2	1	2	4	2	18	0	18	강남 마이 스 관 광특 구
1	2022	4	U	관 광 투 구	1001495	9639	4250	5389	819	2254	395	1435	1454	728	548	829	6122	0	6122	잠실 관광 특구
2	2022	4	U	관 광 투 구	1001494	1977	1303	674	45	218	20	80	67	67	125	315	1591	0	1591	종로? 청계 관광 특구

#### 2) 생활인구 데이터



print(df\_2.columns[:20])

Index(['기준 년코드', '기준\_분기\_코드', '상권\_구분\_코드', '상권\_구분\_코드\_명', '상권\_코드', '상권\_코드\_명', '총\_생활인구\_수', '남성\_생활인구\_수', '여성\_생활인구\_수', '연령대\_10\_생활인구\_수', '연령대\_20\_생활인구\_수', '연령대\_30\_생활인구\_수', '연령대\_30\_생활인구\_수', '연령대\_40\_생활인구\_수', '연령대\_50\_생활인구\_수', '연령대\_60\_이상\_생활인구\_수', '시간대\_1\_생활인구\_수', '시간대\_1\_생활인구\_수', '시간대\_5\_생활인구\_수'], '시간대\_5\_생활인구\_수'], '서간대\_5\_생활인구\_수'], '서간대\_5\_생활인구\_수')

# 3) 직장인구 데이터

25 여성연령대\_50\_직장\_인구\_수 37744 non-null int64 26 여성연령대\_60\_이상\_직장\_인구\_수 37744 non-null int64

	#	Column Non-Nu	II Count Dtype					상																
	1 2 3 4	기준_분기_코드 상권_구분_코드 상권_구분_코드_명 <u>장권_코드</u> 37	37744 non-null int64 37744 non-null int64 37744 non-null object 37744 non-null object 744 non-null int64		기준 _년 월_ 코드	기 준 - 분기 - 코 드	상권 '구분 '코드	권 '구분 '코드 -	상권_코 드	상권 _코 드_ 명	총_직 장_인 구_수	남성_ 직장_ 인구_ 수	여성_ 직장_ 인구_ 수	연령 대 _10_ 직장_ 인구_ 수	3( 직정 인구	령 연령 대 대 )40 <sub>.</sub> 날 직장	! 연령   대 <b>50</b> _ 직장	령대 _ <b>60</b> _ 이상_ 직장_	여성 연령 대 _10_ 직장_ 인구_ 수	여성 연령 대 _ <b>20</b> _ 직장_ 인구_ 수	여성 연령 대 _ <b>30</b> _ 직장_ 인구_ 수	여성 연령 대 _ <b>40</b> _ 직장_ 인구_ 수	여성 연령 대 _ <b>50</b> _ 직장_ 인구_ 수	여성연 령대 _ <b>60</b> _ 이상_ 직장_ 인구_ 수
_			7744 non-null object					명																
Г	7 8 9	총_직장_인구_수 남성_직장_인구_수 여성_직장_인구_수 연령대_10_직장_인구_수 연령대_20_직장_인구_수	37744 non-null int64 37744 non-null int64 37744 non-null int64 37744 non-null int64 37744 non-null int64	0	2022	4	U	관 광 투 구	1001496	강남 마이 스 관광 특구	15904	8904	7000	59 .	336	51 288	1246	225	34	2353	2824	1361	350	78
	11 12 13	연령대_30_직장_인구_수 연령대_40_직장_인구_수 연령대_50_직장_인구_수	37744 non-null int64 37744 non-null int64 37744 non-null int64	1	2022	4	U	관 광 투 구	1001495	잠실 관광 특구	24375	14199	10176	120 .	418	36 3490	2297	1685	69	2238	2379	2121	2217	1152
_	15 16 17	연령대_60_이상_직장_인구_수 남성연령대_10_직장_인구_수 남성연령대_20_직장_인구_수 남성연령대_30_직장_인구_수 남성연령대_40_직장_인구_수	37744 non-null int64 37744 non-null int64 37744 non-null int64 37744 non-null int64 37744 non-null int64	2	2022	4	U	관 광 투 구	1001494	종 로? 청계 관광 특구	32935	17150	15785	68 .	49	11 459 <sup>.</sup>	4158	1190	53	3072	5054	4892	2243	471
	19 20 21 22 23	남성연령대_50_직장_인구_수 남성연령대_60_이상_직장_인구 여성연령대_10_직장_인구_수 여성연령대_20_직장_인구_수 여성연령대_30_직장_인구_수 여성연령대_40_직장_인구_수	37744 non-null int64 子_수 37744 non-null int64 37744 non-null int64 37744 non-null int64 37744 non-null int64 37744 non-null int64																					

# 4) 점포 데이터

#	Column	Non-Null Count Dtype	
0 1 2 3 4 5 6	서비스_업종_	드 898068 non-null inte 도 898068 non-null obje 도명 898068 non-null obj 898068 non-null int64 898068 non-null objec 코드 898068 non-null obj 코드_명 898068 non-null obj	64 ect iect ct iect
9 10 11 12 13 14	개업_율 개업_점포_수 폐업_률 폐업_점포_수		1 1

기	[준_년_ 코드	기준_분 기_코드	상권_구 분_코드	상권_구분 _코드_명	상권_코 드	상권_코드_ 명	서비스_업종_ 코드	서비스_업 종_코드_명	점포 _수	유사_업종 _점포_수	개업 _ <sup>율</sup>	개업_ 점포_ 수	폐업 _률	폐업_ 점포_ 수	프랜차이 즈_점포_ 수
0	2022	4	U	관광특구	1001496	강남 마이 스 관광특 구	CS300043	전자상거래 업	6	6	0	0	0	0	0
1	2022	4	U	관광특구	1001496	강남 마이 스 관광특 구	CS300042	주유소	1	1	0	0	0	0	0
2	2022	4	U	관광특구	1001496	강남 마이 스 관광특 구	CS300041	예술품	10	10	0	0	0	0	0

## 5) 아파트 데이터

#_	Column	Non-Null Count Dtype
0 1 2 3	기준_년_코드 기준_분기_코드 강권_구분_코드 <u>상권_구분_코드_</u> 명 상권_코드	35121 non-null int64 35121 non-null int64 35121 non-null object 35121 non-null object 35121 non-null int64
5	장권_코드_명	35121 non-null object
6	아파트_단지_수	35121 non-null int64
7		_미만_세대_수 34362 non-null fl
8		_세대_수 32584 non-null floa
9	아파트_면적_99_제곱미터	
10		d_세대_수 13083 non-null floa
11		d_세대_수 8275 non-null floa
12	––– . – .	∥대_수 30052 non-null float
13	아파트_가격_1_억_세대_ન	
14	아파트_가격_2_억_세대_ન	
15	아파트_가격_3_억_세대_ન	
16	아파트_가격_4_억_세대_ન	
17	아파트_가격_5_억_세대_쉬	
18		∥대_수 7601 non-null float
19	아파트_평균_면적	35121 non-null int64
20	아파트_평균_시가	35121 non-null int64

	기준 _년_ 코드	기준 분기 코드	상권 구분 코드	상 권 구 분 코 드 명	상권_코 드	상 권 「코 드 「명	아파트 단지 수	아파트_ 면적 _66_제 곱미터_ 미만_세 대_수	아파트 _면적 _66_제 곱미터 _세대_ 수	아파트 _면적 _99_제 곱미터 _세대_ 수	 아파트_ 면적 _165_제 곱미터_ 세대_수	아파 트_가 격_1_ 억_미 만_세 대_수	아파 트_ 가격 _1_ 억_ 세대 _수	아파 트_ 가격 _2_ 억_ 세대 _수	아파 트_가 격_3_ 억_세 대_수	아파트 _가격 _ <b>4</b> _억_ 세대_ 수	아파 트_가 격_ <b>5</b> _ 억_세 대_수	아파트 _가격 _6_억_ 이상_ 세대_ 수	아파트'평균'면적	아파트_평 균_시가
0	2022	4	R	전 통 시 장	2130326	고 덕골목형상점가	2	657.0	5.0	NaN	 NaN	NaN	2.0	4.0	NaN	129.0	21.0	506.0	53	633301800
1	2022	3	R	전 통 시 장	2130326	고 덕 골 목 형 상점 가	2	657.0	5.0	NaN	 NaN	NaN	2.0	4.0	NaN	129.0	21.0	506.0	53	633301800
2	2022	2	R	전 통 시 장	2130326	고 덕 골 목 형 상 점 가	2	657.0	5.0	NaN	 NaN	NaN	2.0	4.0	NaN	129.0	21.0	506.0	53	633301800

## 6) 상권영역 데이터

	기준_년윌_코드	상권_구분_코드	상권_구분_코드_명	상권_코드	상권_코드_명	엑스좌표_값	와이좌표_값	시군구_코드	행정동_코드	형태정보
0	202112	А	골목상권	2110753	시흥동 은행나무사거리	191934	438813	11545	11545710	NaN
1	202112	Α	골목상권	2110427	노원 을지대학교병원	206265	459431	11350	11350611	NaN
2	202112	А	골목상권	2110755	동일여자고등학교	192300	439070	11545	11545680	NaN

#	Column	Non-Null Count Dtyp	ре
0	기준_년월_코	⊆ 1671 non-null	int64
1	상권_구분_코	⊆ 1671 non-null	object
2	상권_구분_코	<u>드</u> _명 1671 non-null	l object
3	상권_코드	1671 non-null	int64
4	상권_코드_명	1671 non-null	object
5	엑스좌표_값	1671 non-null	int64
6	와이좌표_값	1671 non-null	int64
7	시군구_코드	1671 non-null	int64
8	행정동_코드	1671 non-null	int64
9	형태정보	O non-null	float64

# I. 데이터 전처리 II. Clustering

#### I. 데이터 전처리

- 인구 데이터 ⇨ 총 인구 수, 상권 별 평균 나이, 나이 합 추출 및 계산

```
import pandas as pd
df1 = pd.read csv("data₩서울시 우리마을가게 상권분석서비스(상권-상주인구).csv", encoding="cp949")
pop1_22 = df1.loc[(df1.기준 년 코드 == 2022) & (df1.기준 분기_코드 == 4)].iloc[:.[4.5.8.9.10.11.12.13]]
pop1_22['ave_age1'] = (15*pop1_22.iloc[:,2]+25*pop1_22.iloc[:,3]+35*pop1_22.iloc[:,4]
                     +45*pop1_22.iloc[:,5]+55*pop1_22.iloc[:,6]+65*pop1_22.iloc[:,7])/pop1_22.iloc[:,1]
pop1 22['sum age1'] = 15*pop1 22.iloc[:.2]+25*pop1 22.iloc[:.3]+35*pop1 22.iloc[:.4]
                      +45*pop1 22.iloc[:.5]+55*pop1 22.iloc[:.6]+65*pop1 22.iloc[:.7]
pop1_iloc = pop1_22.iloc[:,[0.1,-2,-1]]
pop1_iloc.rename(columns={'상권 코드': '상권_코드', '총 상주인구 수': '총_상주인구_수'}. inplace=True)
population1 = pop1 iloc
population1
```

# - 계산된 인구 데이터 셋 3개

	상권_코드	총_상주인구_수	ave_age1	sum_age1
0	1001496	29	48.103448	1395
1	1001495	9639	39.563751	381355
2	1001494	1977	51.605969	102025
3	1001493	2703	41.115427	111135
4	1001492	1089	54.081726	58895
1626	2110005	1796	42.978842	77190
1627	2110004	471	46.210191	21765
1628	2110003	3300	41.963636	138480
1629	2110002	629	48.354531	30415
1630	2110001	1170	43.427350	50810

	상권_코드	총_생활인구_수	ave_age2	sum_age2
0	1001496	112658	38.979788	4391385
1	1001495	3632727	38.870040	141204245
2	1001494	3957665	44.877106	177608550
3	1001493	3249964	42.836948	139218540
4	1001492	4667755	41.852843	195358815
1664	2110005	149283	43.133746	6439135
1665	2110004	379929	41.787308	15876210
1666	2110003	331722	42.350161	14048480
1667	2110002	323459	42.412037	13718555
1668	2110001	179890	43.854077	7888910

	상권_코드	총_직장_인구_수	ave_age3	sum_age3
0	1001496	15904	37.966549	603820
1	1001495	24375	42.464205	1035065
2	1001494	32935	41.611508	1370475
3	1001493	6054	41.243806	249690
4	1001492	104830	38.751216	4062290
1625	2110005	1006	43.349901	43610
1626	2110004	32	48.125000	1540
1627	2110003	485	46.030928	22325
1628	2110002	475	37.863158	17985
1629	2110001	1066	44.643527	47590

#### - 점포 데이터 ⇒ 음식 관련 점포 수 계산

```
df = pd.read_csv("data₩서울시 우리마을가게 상권분석서비스(상권-점포).csv", encoding="cp949")
|store_2022 = df.loc[(df['기준_년_코드'] == 2022) & (df['기준_분기_코드'] == 4)]
store_iloc = store_2022.iloc[:.[4. 7. 8]]
target_names = ['편의점', '슈퍼마켓', '커피-음료', '분식전문점', '패스트푸드점',
             '제과점', '호프-간이주점', '치킨전문점']
target_rows = store_iloc[store_iloc['서비스_업종_코드_명'].isin(target_names)]
data = {
   '상권_코드': [].
   '점포_수': []
for code in target rows['상권 코드'].unique():
   temp = target_rows[target_rows['상권_코드'] == code]
   sum of temp = temp['점포 수'].sum(skipna=True)
   data['상권_코드'].append(code)
   data['점포 수'].append(sum of temp)
store = pd.DataFrame(data)
store
```

	상권_코드	점포_수
0	1001496	72
1	1001495	299
2	1001494	524
3	1001493	445
4	1001492	1007
1649	2110005	12
1650	2110004	4
1651	2110003	27
1652	2110002	8
1653	2110001	7

#### - 아파트 데이터 ⇒ 아파트 단지 수 column 추출

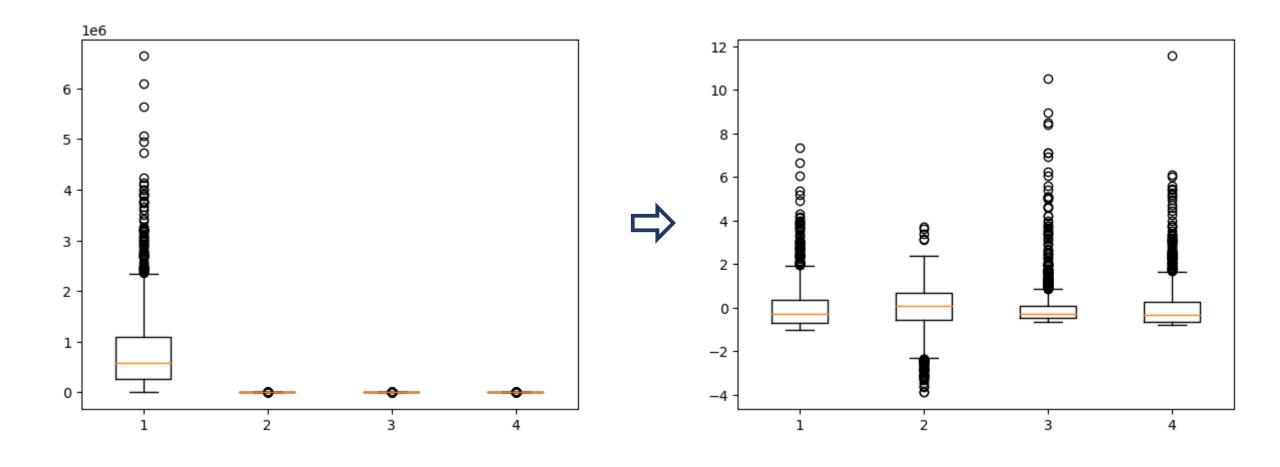
```
apt_2022 = df.loc[(df['기준_년_코드'] == 2022) & (df['기준_분기_코드'] == 4)]
apt_iloc = pd.DataFrame(apt_2022.iloc[:,[4, 6]])
apt_iloc.drop_duplicates(subset=['상권_코드'], keep='first', inplace=True)
apt = apt_iloc
apt
```

	상권_코드	아파트_단지_수
0	2130326	2
4	2130325	5
8	2130324	17
12	2130323	30
16	2130322	12
5912	2110001	21
5916	1001495	30
5920	1001494	9
5924	1001493	11
5928	1001491	72

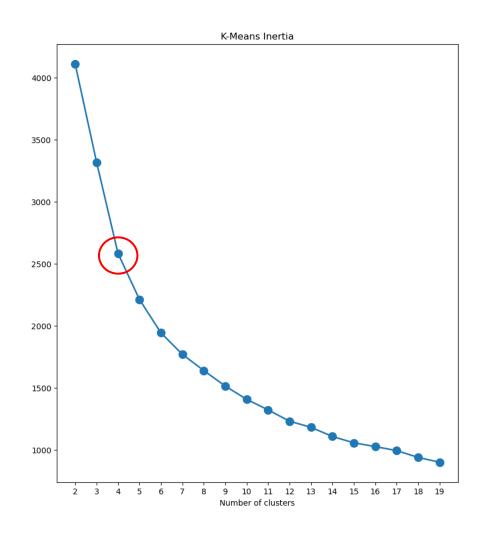
# - 최종 데이터 셋

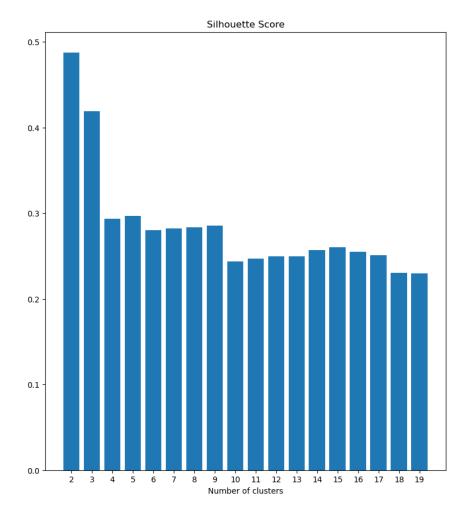
30
-
9
11
72
2
40
6
70
7
21

## - StandardScaler로 표준화



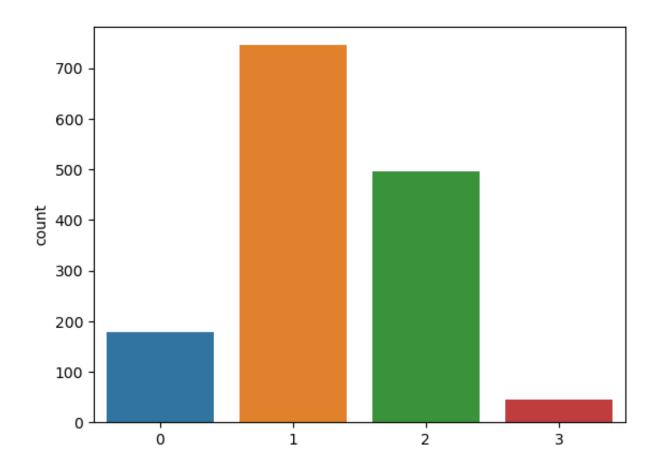
# II. Clustering



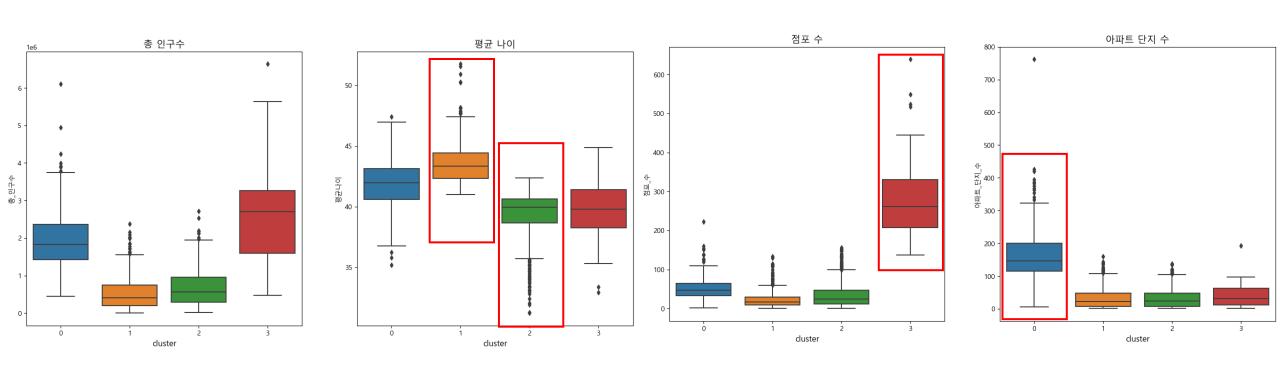


## **II. Clustering**

kmeans = KMeans(n\_clusters= 4, random\_state = 10)

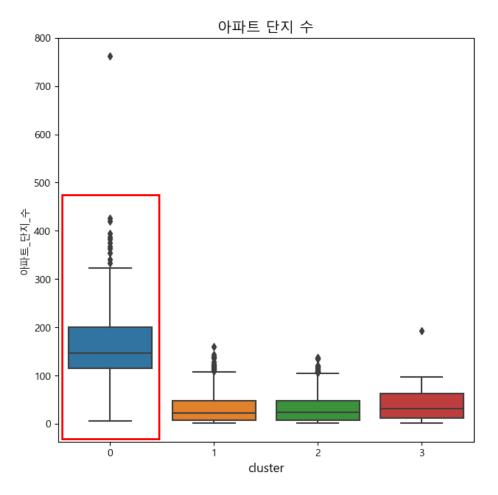


# II. Clustering



# 5. 결과

# 5. 결과



상권_코드	상권_코드_명	cluster
2110659	화곡역 4번	0
2110662	까치산역 1번	0
2110712	남구로역 5번	0
2111003	삼전역 4번	0
2111010	석촌역 7번	0
2110717	남구로역 2번	0
2110654	까치산역 4번	0
2111015	석촌역 3번	0
2110468	응암역 3번	0
2110544	망리단길	0

Cluster 0 = 아파트 단지 수가 많은 클러스터

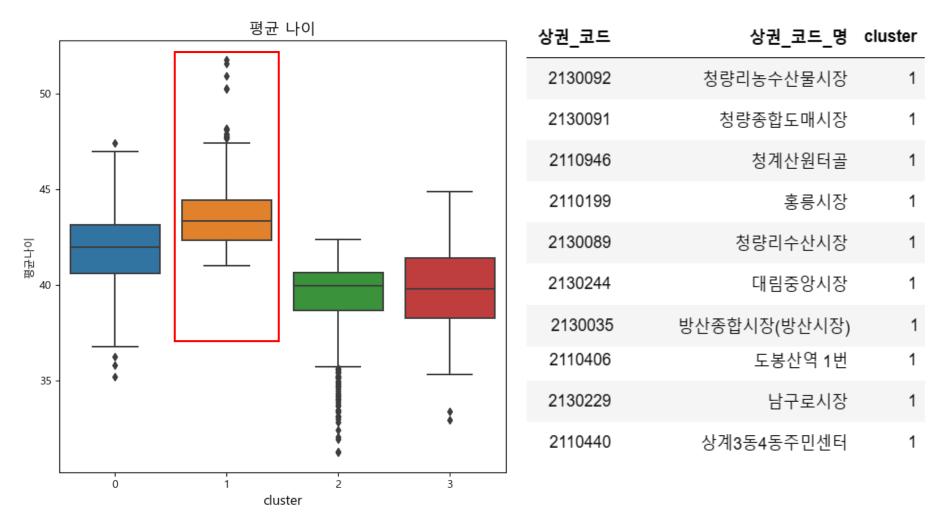
#### ✓ 조깅/ 산책할 수 있는 산책로



## ✓ 가족들과 함께 할 수 있는 공간



#### 5. 결과



Cluster 1 = 평균 연령이 높은 클러스터

- ✓ 계단이나 가파른 경사는 없도록
- ✓ 편하게 쉴 수 있는 공간

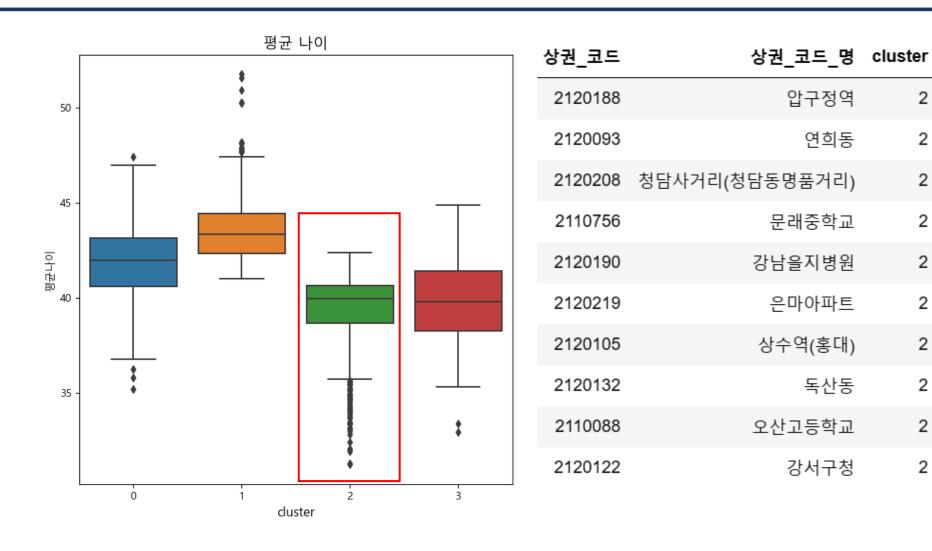


#### ✓ 어르신들이 사용할 수 있는 시설물



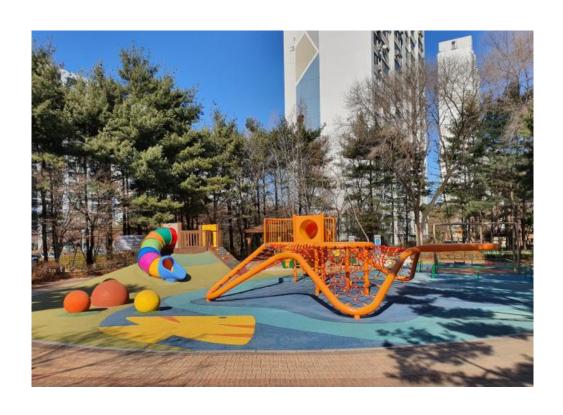
핀란드 헬싱키에 마련된 어르신 놀이터. (사진=이건웅 교수 제공)

#### 5. 결과



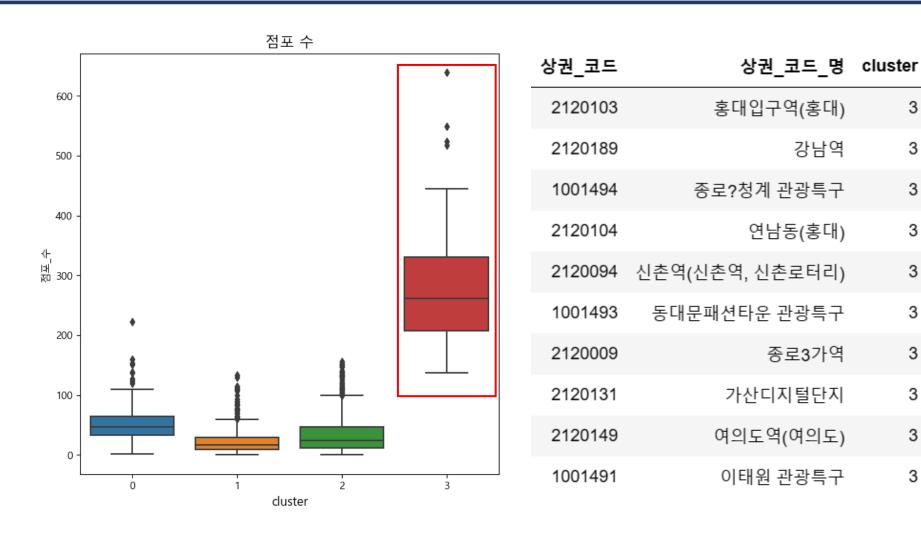
Cluster 2 = 평균 연령이 낮은 클러스터

# ✓ 아이들이 놀 수 있는 놀이기구 설치





#### 5. 결과



Cluster 3 = (음식 관련) 점포 수가 많은 클러스터

- ✓ 음식을 먹을 수 있는 테이블 설치
- ✓ 음식물 포함 기타 쓰레기를 버릴 수 있는 분리수거 쓰레기 통 마련





# 6. 기대 효과

#### 6. 기대효과

아주 작은 공간이더라도 주변 특성을 파악하여 효율적인 공원을 만든다면 모두가 공유하고 공통된 추억을 함께 할 수 있는 공간 탄생



서로가 서로를 마주보는 시간이, 공통된 공간이 많이 생긴다면 어렴풋한 타인이 실체화되어 서로에 대한 오해, 갈등 감소 기대

# 7. 한계점 및 추후 개선 방안

#### 7. 한계점 및 추후 개선 방안

- 상권 근처의 공원 여부 파악 X

- 실제 사용가능한 공간은 고려 X

⇒ 관련 데이터 확보 or 지자체에서 관련 정보 구함

#### 활용 자료 출처

조선일보: [유현준의 도시 이야기] 공원과 스타벅스의 차이

https://www.chosun.com/site/data/html\_dir/2019/12/19/2019121903847.html

유튜브 셜록현준 : 강남문제, 공원과 벤치부터? 건축가가 본 공원으로 세상을 바꾸는 방법

https://www.youtube.com/watch?v=m1WUJETal2U&t=612s

<u>서울 열린데이터 광장 - 인기그룹데이터 - 골목상권분석정보</u>

https://data.seoul.go.kr/dataList/3/literacyView.do

서울특별시 빅데이터 캠퍼스 - 데이터 분석 사례 - 지하철 공실 문제 해결을 위한 공유오피스 도입 및 활성화 방안 https://bigdata.seoul.go.kr/noti/selectNoti.do?r\_id=P260&bbs\_seq=605&ac\_type=A1&sch\_type=&sch\_text=&currentPage=1