

2023-1 데이터사이언스캡스톤프로젝트

Data Analysis Presentation

포동포동

김보현 김민

CONTENTS



- | | |
|----|----------|
| 01 | 주제 리마인드 |
| 02 | 전체 데이터셋 |
| 03 | 데이터별 EDA |
| 04 | 상관관계 분석 |
| 05 | 결론 |

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

서울시 공동육아나눔터 입지 선정

가족 돌봄 기능을 보완해 부모의 육아부담을 줄이고,
양육 친화적인 사회 환경을 조성하고자 운영되는 공간



- 주제 리마인드
- 전체 데이터셋
- 데이터별 EDA
- 상관관계 분석
- 결론

1. 데이터셋 변경

서울시민 KCB 생활금융 통계	서울 빅데이터 캠퍼스
▼	
서울시 자치구별 1인당 지역내총생산 및 수준지수(2015년 기준) 통계	서울 열린데이터 광장

2. 데이터셋 추가



행정구별 예산, 한부모 가정수, 키즈카페 등을 고려하는 건 어떨까요?

서울시 저소득 한부모가족 통계	서울시 열린데이터 광장
서울시 예산결산총괄 통계	서울시 열린데이터 광장
서울시 유원시설업(기타) 인허가 정보	서울시 열린데이터 광장

키즈카페 등 안전성 검사 대상이 아닌 유기사설 또는 유기기구를 설치하여 운영하는 업소

- 주제 리마인드
- 전체 데이터셋
- 데이터별 EDA
- 상관관계 분석
- 결론

1. 데이터셋 변경

서울시민 KCB 생활 만족도 통계	서울 빅데이터 캠퍼스
서울시 자치구별 지역내총생산 및 수준지수(2019년 기준) 통계	서울 열린데이터 광장

최종 데이터셋 선정 완료 !



행정구별 예산, 한부모 가정수, 키즈카페 등을 고려하는 건 어떨까요?

서울시 저소득 한부모가족 통계	서울시 열린데이터 광장
서울시 예산결산총괄 통계	서울시 열린데이터 광장
서울시 유원시설업(기타) 인허가 정보	서울시 열린데이터 광장

키즈카페 등 안전성 검사 대상이 아닌 유기사설 또는 유기기구를 설치하여 운영하는 업소

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

데이터셋이 10개나?

10개의 데이터셋을 모두 합쳐 하나의 데이터셋 만들기

데이터셋을 어떤 변수를 기준으로 매핑 해야 할까?

위치를 기준으로 합쳐 보기

자치구는 너무 포괄적어서 적합하지 않은 것 같으므로,
행정동을 기준으로 해보자!



행정동 Column 유무 확인

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

데이터셋	행정동 열 유무
서울시 어린이집 정보	X (주소)
서울시 유원시설 정보	X (주소)
서울시 공동육아나눔터 정보	X (주소)
생활인구	0
직장인구	0
행정동별 혼인/이혼	0
어린이 보호구역 지정 통계	0
저소득/한부모 가정 통계	X (자치구)
자치구별 예산 총괄	X (자치구)
자치구별 지역 내 총생산	X (자치구)

행정동 Column 유무 확인

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

데이터셋	행정동 열 유무
서울시 어린이집 정보	X (주소)
서울시 유원시설 정보	X (주소)
서울시 공동육아나눔터 정보	X (주소)
생활인구	0
직장인구	0
유형 1) 주소 정보만 있는 데이터셋	
서울시 공동육아나눔터 정보 데이터셋의 경우, 행이 30개 뿐이라 직접 행정동 찾아서 입력	
어린이집과 유원시설 데이터셋에 대해서는 카카오 API를 통한 행정동 추출 시도	
행정동별 주민등록 인구는	0
어린이 보호구역 지정 통계	0
저소득/한부모 가정 통계	X (자치구)
자치구별 예산 총괄	X (자치구)
자치구별 지역 내 총생산	X (자치구)

행정동 Column 유무 확인

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

행정동 열 있는 데이터는
No problem



데이터셋	행정동 열 유무
서울시 어린이집 정보	X (주소)
서울시 유원시설 정보	X (주소)
서울시 공동육아나눔터 정보	X (주소)
유형 2) 자치구 정보만 있는 데이터셋 (자치구별 통계)	
자치구에 속한 모든 행정동에 같은 값 할당	
행정동별 혼인/이혼	0
어린이 보호구역 지정 통계	0
저소득/한부모 가정 통계	X (자치구)
자치구별 예산 총괄	X (자치구)
자치구별 지역 내 총생산	X (자치구)

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

행정동 추출하기

1. 카카오 API

- 서버 에러로 API 연결이 끊김 → `time.sleep()`을 통해 해결
- 입력 데이터셋 개수와 출력 데이터셋 개수가 다른, API 상의 문제 → 9366개 중 679개의 행정동 추출 실패

2. 크롤러

- 도로명주소 안내 시스템 크롤링을 통한 행정동 추출 시도 > 정확한 장소를 찾지 못하면 검색 결과 X
- 네이버 지도 크롤링을 통한 행정동 추출 시도 + 전체 주소에서 핵심 주소만 추출하여 검색 및 크롤링 시도
- > 여전히 230개 행정동 추출 실패 & '고척1동'이 '고척동'으로만 나와서 mapping 불가능



1) 주소 정제

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

행정동 추출하기

1. 주소 정제

정확한 결과 도출을 위해 핵심 부분만 추출하는
주소 정제 작업

2. 주소 -> 좌표 변환

지오서비스를 활용해, 주소를 위도 및 경도로 변환
전체 9359개 데이터 중 변환이 안된 4개는 직접 처리

3. 카카오 API 활용해 행정동 추출

위도와 경도를 주소로 변환하고,
주소에서 행정동을 최종 추출 & 데이터 병합

```
def get_addrs(x):  
    addr = x.split()  
    return " ".join(addr[2:4])
```

주소 정제 함수

field1	field2	_GC_TYPE	_X	_Y
5007	약령사로 154	정	127.0458	37.58156
5008	관악로 285	정	126.9554	37.48939
5009	신평로11다길 18	유	126.907	37.49991
5010	도당로 13	정	127.0343	37.65941
5011	문정로 83	정	127.1299	37.48954
5012	갈현로7길 9	정	126.9093	37.60361
5013	송이로31길 32-10	정	127.1343	37.48984
5014	양천로14길 118	정	126.8082	37.56882
5015	길음로 119	정	127.0165	37.61082
5016	역촌1동 35-65	유	126.9175	37.61085
5017	도신로4길 10	정	126.8957	37.50249

좌표 변환 결과



주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

통계청행정	행정자치부	시도명	시군구명	행정동명	공동육아L	2021_한부	2021_저소	2020_한부	2020_저소	2019_한부	2019_저소	2018_한부	2018_저소	2021_예산	2021_세입	2021_세출
1101053	11110530	서울	종로구	사직동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101054	11110540	서울	종로구	삼청동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101055	11110550	서울	종로구	부암동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101056	11110560	서울	종로구	평창동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101057	11110570	서울	종로구	무악동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101058	11110580	서울	종로구	교남동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101060	11110600	서울	종로구	가회동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101061	11110615	서울	종로구	종로1.2.3.4가동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101063	11110630	서울	종로구	종로5.6가동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101064	11110640	서울	종로구	이화동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101067	11110670	서울	종로구	창신1동	1	96	214	114	220	140	227	154	237	662770	695108	540729
1101068	11110680	서울	종로구	창신2동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101069	11110690	서울	종로구	창신3동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101070	11110700	서울	종로구	송인1동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101071	11110710	서울	종로구	송인2동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101072	11110515	서울	종로구	청운효자동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1101073	11110650	서울	종로구	혜화동	0	96	214	114	220	140	227	154	237	662770	695108	540729
1102052	11140520	서울	중구	소공동	0	72	263	94	266	93	272	103	268	653442	647475	536751
1102054	11140540	서울	중구	회현동	0	72	263	94	266	93	272	103	268	653442	647475	536751
1102055	11140550	서울	중구	명동	0	72	263	94	266	93	272	103	268	653442	647475	536751
1102057	11140570	서울	중구	필동	0	72	263	94	266	93	272	103	268	653442	647475	536751

424 rows × 63 columns

행정동명, 공동육아나눔터유무, 어린이보호구역현황, 어린이집수, 연도별 저소득 가구수, 연도별 한부모
가구수, 연도별 예산현액 & 세입 & 세출, 총생활인구수, 성별 생활인구수, 나이대별 생활인구수, 직장인구수,
연도별 혼인건수, 연도별 GDP ...



주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

생활 인구 데이터

형태

총생활인구수	남자0세부터 9세생활인구 수	남자10세부 터14세생활 인구수	남자30세부 터34세생활 인구수	남자35세부 터39세생활 인구수	남자40세부 터44세생활 인구수	남자45세부 터49세생활 인구수	여자0세부터 9세생활인구 수	여자10세부 터14세생활 인구수	여자25세부 터29세생활 인구수	여자30세부 터34세생활 인구수	여자35세부 터39세생활 인구수
23050.525414	376.461406	167.316184	1036.331230	1290.318078	1199.036465	1341.700725	413.434509	221.482378	1289.548565	1222.913410	1368.620025
5056.038923	62.542217	47.922133	185.886491	225.446885	233.844939	236.765876	88.232968	54.200151	208.088268	198.518384	266.113977
15235.972430	428.735398	332.435485	363.460493	487.424951	513.138530	584.285075	436.183110	310.662880	429.958440	384.796150	606.040526
18042.284852	583.821039	422.296752	376.874308	519.554757	553.489355	734.932078	606.922678	354.606187	443.184777	456.192115	671.908807
6575.949430	231.906654	148.312047	145.121536	200.780400	195.878479	222.667330	271.936491	171.897375	174.067480	180.246811	258.628149

424 rows × 12 columns

총 생활인구수, 성별 & 나이대별 생활인구수
- 아이와 직장인(203040) 타겟

결측치 > 없음!

life.isna().sum() # 결측치 없음

```

통계청행정동코드      0
행자부행정동코드      0
시도명                  0
시군구명                0
행정동명                0
총생활인구수            0
남자0세부터9세생활인구수  0
남자10세부터14세생활인구수  0
남자30세부터34세생활인구수  0
남자35세부터39세생활인구수  0
남자40세부터44세생활인구수  0
남자45세부터49세생활인구수  0
여자0세부터9세생활인구수  0
여자10세부터14세생활인구수  0
여자25세부터29세생활인구수  0
여자30세부터34세생활인구수  0
여자35세부터39세생활인구수  0
여자40세부터44세생활인구수  0
dtype: int64

```



생활 인구 데이터

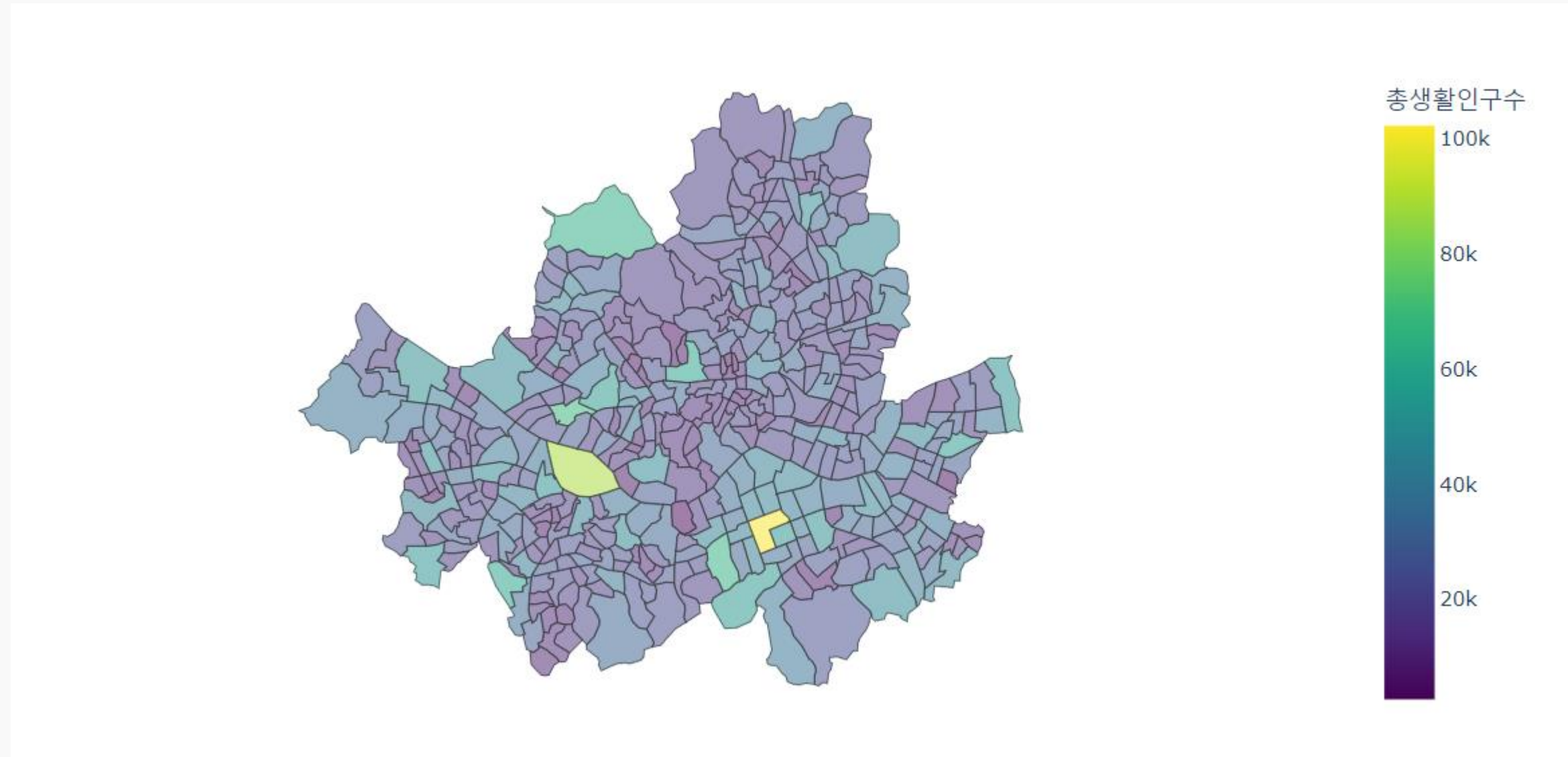
주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론



총 생활인구수가 가장 많았던 행정동: 역삼1동, 여의동, 서교동, 서초3동, 진관동
총 생활인구수가 가장 적었던 행정동: 삼청동, 창신3동, 신월6동, 둔촌1동, 반포본동

생활 인구 데이터

주제 리마인드

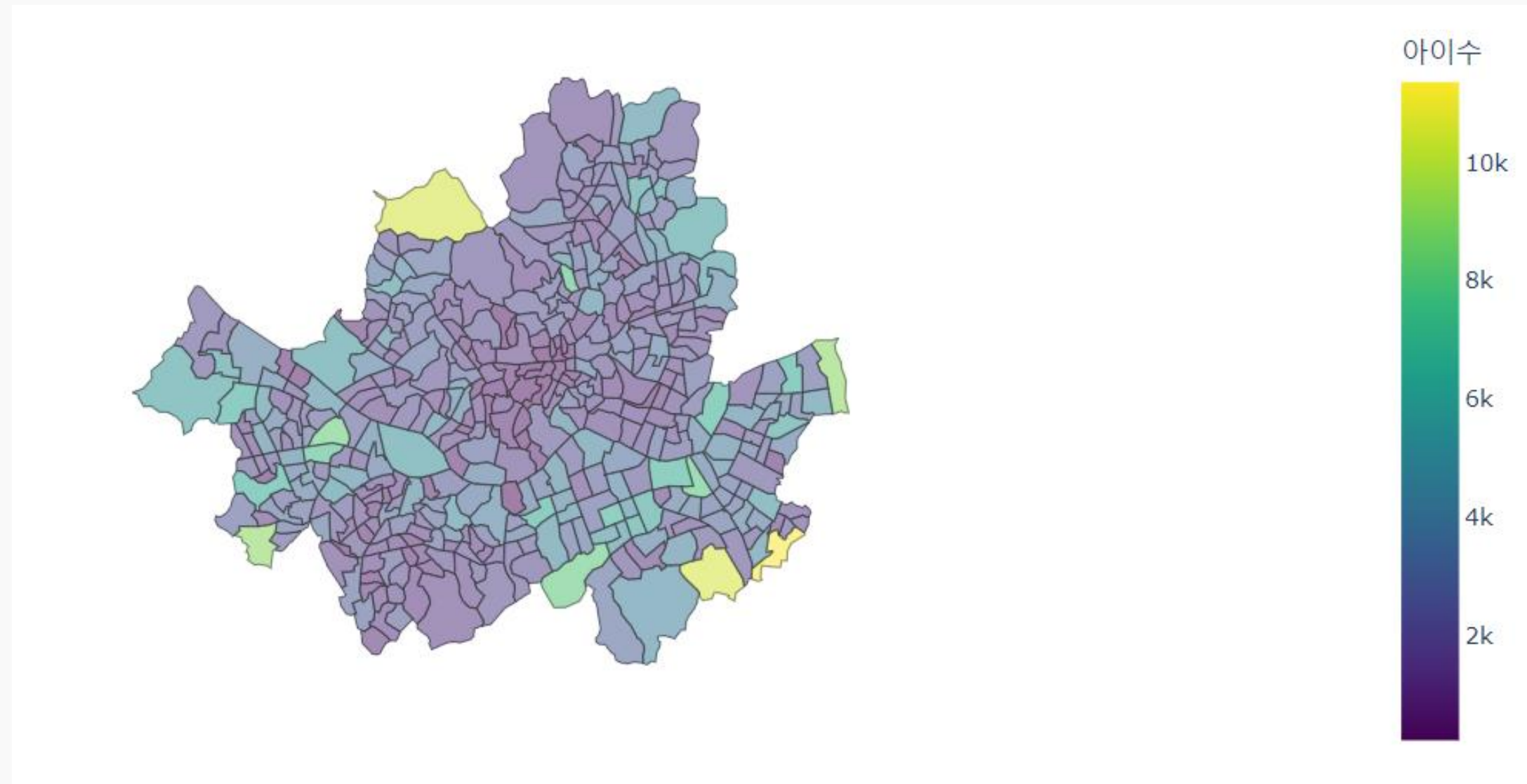
전체 데이터셋

데이터별 EDA

상관관계 분석

결론

성별 0세~9세, 10세~14세 타겟으로 '아이 수' 변수 생성 > 시각화



아이수가 가장 많았던 행정동: 위례동, 세곡동, 진관동, 강일동, 오류2동
아이수가 가장 적었던 행정동: 가리봉동, 삼청동, 을지로동, 창신1동, 반포본동

생활 인구 데이터

주제 리마인드

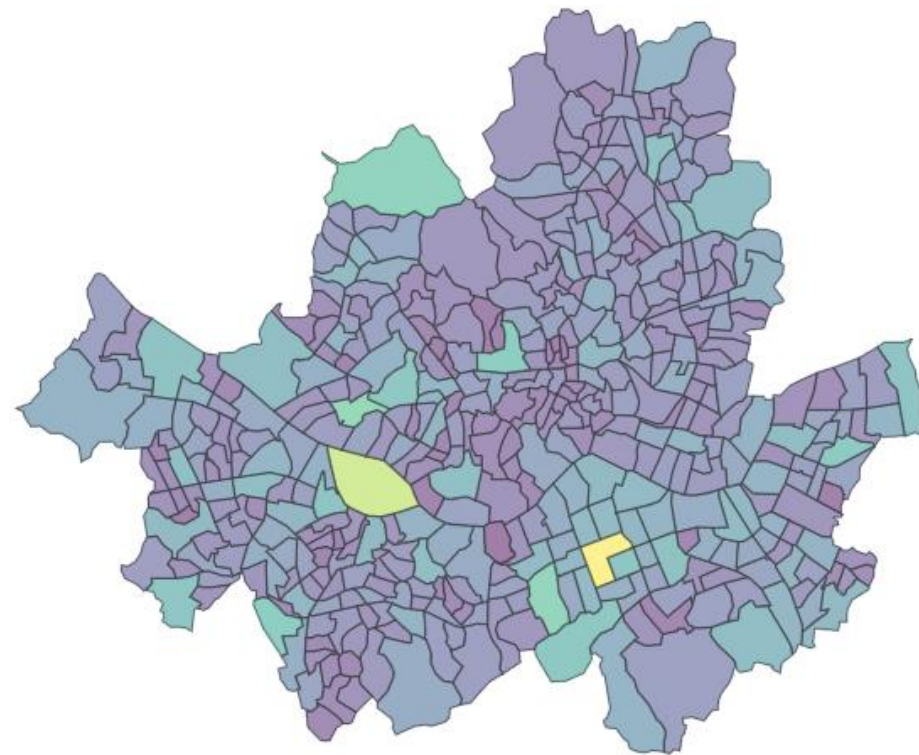
전체 데이터셋

데이터별 EDA

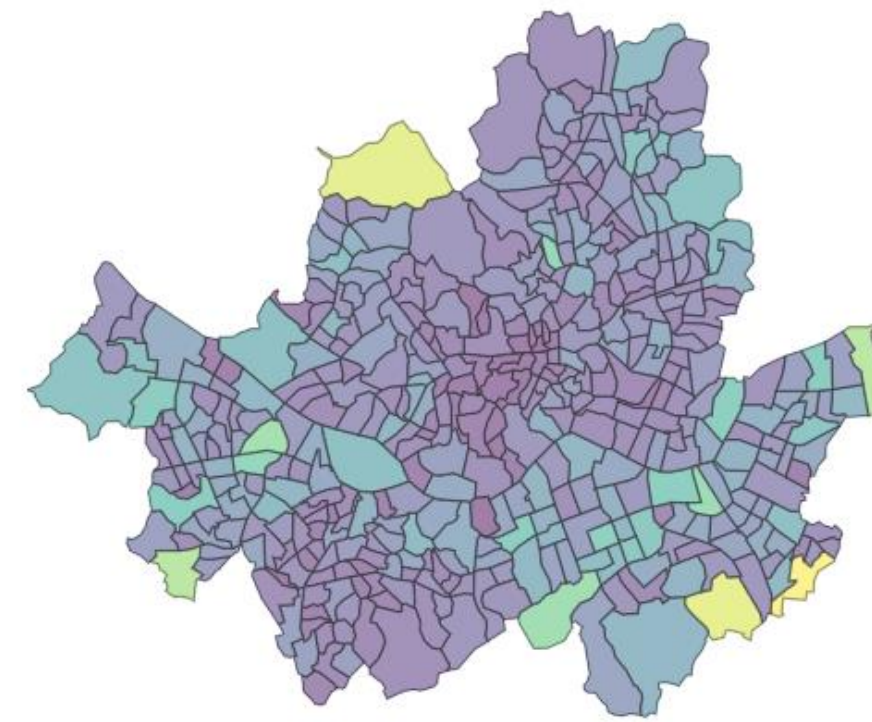
상관관계 분석

결론

총 생활인구 수



아이 수



- 총 생활인구가 가장 많았던 강남구 역삼동은 상대적으로 아이가 많이 살고 있지 않았음
- 서초구 반포본동은 총 생활인구수와 아이 수 모두 가장 적었음
- 진관동과 삼성동은 아이수와 전체인구 비례함
- 송파구 위례동, 강남구 세곡동은 전체 인구에 비해 아이가 많이 살고 있음

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

혼인 데이터

형태

2018혼인	2019혼인	2020혼인	2021혼인	2022혼인
40.0	35.0	40.0	42.0	43.0
11.0	18.0	22.0	11.0	11.0
38.0	34.0	34.0	35.0	32.0
69.0	63.0	58.0	46.0	43.0
24.0	18.0	18.0	17.0	12.0
...
60.0	50.0	44.0	19.0	19.0
72.0	77.0	80.0	82.0	60.0
257.0	232.0	195.0	173.0	143.0
186.0	155.0	155.0	234.0	189.0
285.0	217.0	275.0	216.0	208.0

424 rows × 5 columns

5개년 혼인 건수

결측치

```
marriage.isna().sum()
```

```

통계청행정동코드    0
행자부행정동코드    0
시도명                0
시군구명              0
행정동명              0
2018혼인              0
2019혼인              0
2020혼인              0
2021혼인              1
2022혼인              2
dtype: int64

```

> 결측치 확인

```

marriage['2021혼인'] = marriage['2021혼인'].fillna(marriage['2021혼인'].mean())
marriage['2022혼인'] = marriage['2022혼인'].fillna(marriage['2022혼인'].mean())

```

평균으로 결측치 보간



혼인 데이터

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

'5개년 혼인 평균' 변수 생성 > 지도 시각화



5개년 혼인 평균이 가장 높았던 행정동: 사직동, 삼청동, 부암동, 평창동, 무악동 -> 모두 종로구
5개년 혼인 평균이 가장 낮았던 행정동: 둔촌1동, 둔촌2동, 암사1동, 천호2동, 길동 -> 모두 강동구

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

어린이 보호 구역 데이터

형태

어린이보호구역현황
3.0
NaN
5.0
1.0
1.0
...
3.0
5.0
5.0
5.0
10.0

424 rows × 1 columns

행정동별 어린이 보호구역 개수

결측치

```
safezone.isna().sum()
```

```
통계청행정동코드      0
행자부행정동코드      0
시도명                0
시군구명              0
행정동명              0
어린이보호구역현황    145
dtype: int64
```

> 결측치가 많다!

```
safezone['어린이보호구역현황'] = safezone['어린이보호구역현황'].fillna(0)
```

```
safezone.isna().sum() # 결측치 처리 확인
```

```
통계청행정동코드      0
행자부행정동코드      0
시도명                0
시군구명              0
행정동명              0
어린이보호구역현황      0
dtype: int64
```

개수이므로, 0으로 결측치 보간



어린이 보호 구역 데이터

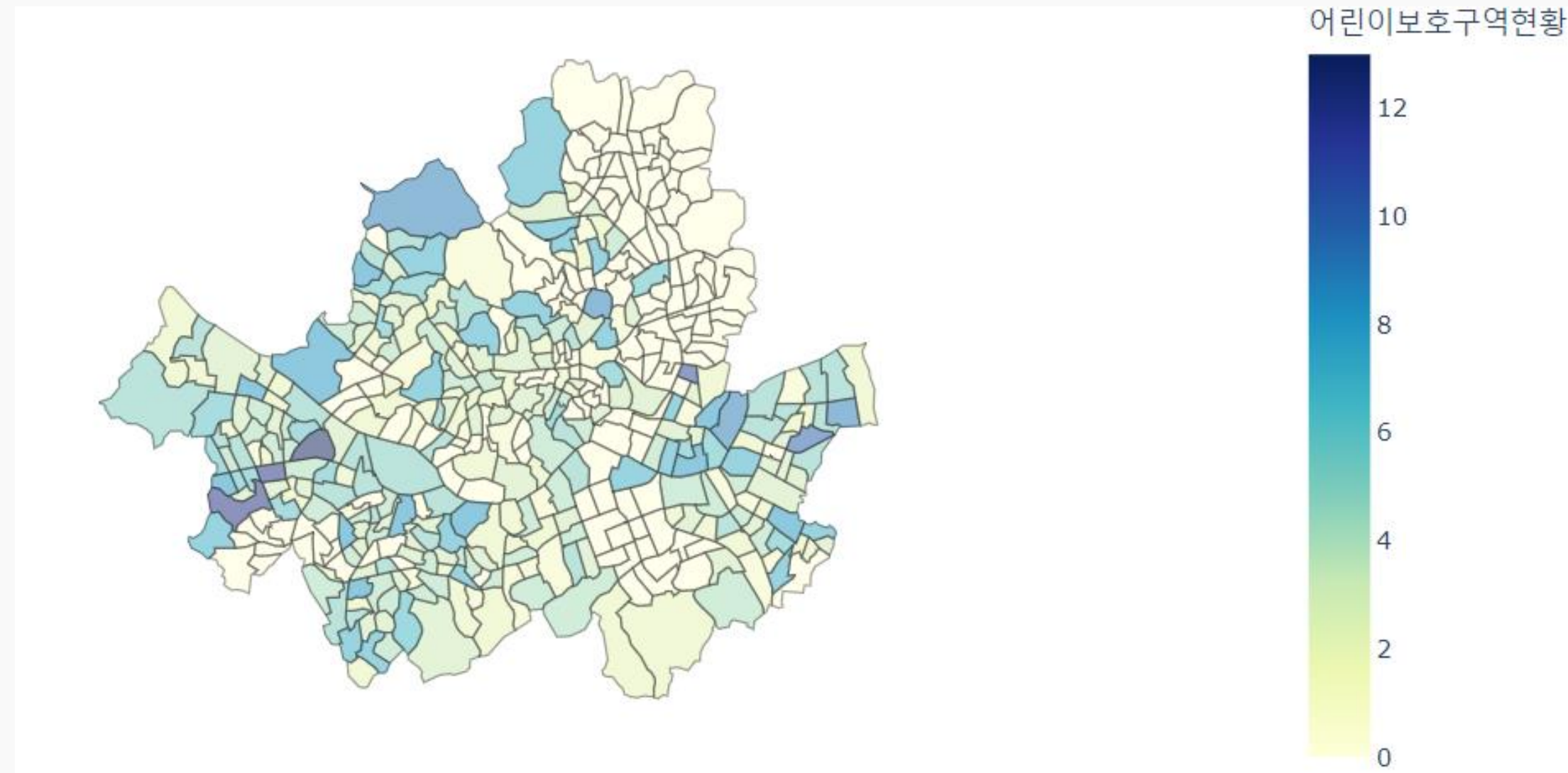
주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론



어린이 보호 구역이 가장 많았던 행정동: 목5동, 신정3동, 신정4동, 중곡3동, 길동
-> 결측치를 모두 0으로 처리했으므로, 가장 적었던 행정동은 무의미함

어린이 보호 구역 데이터

주제 리마인드

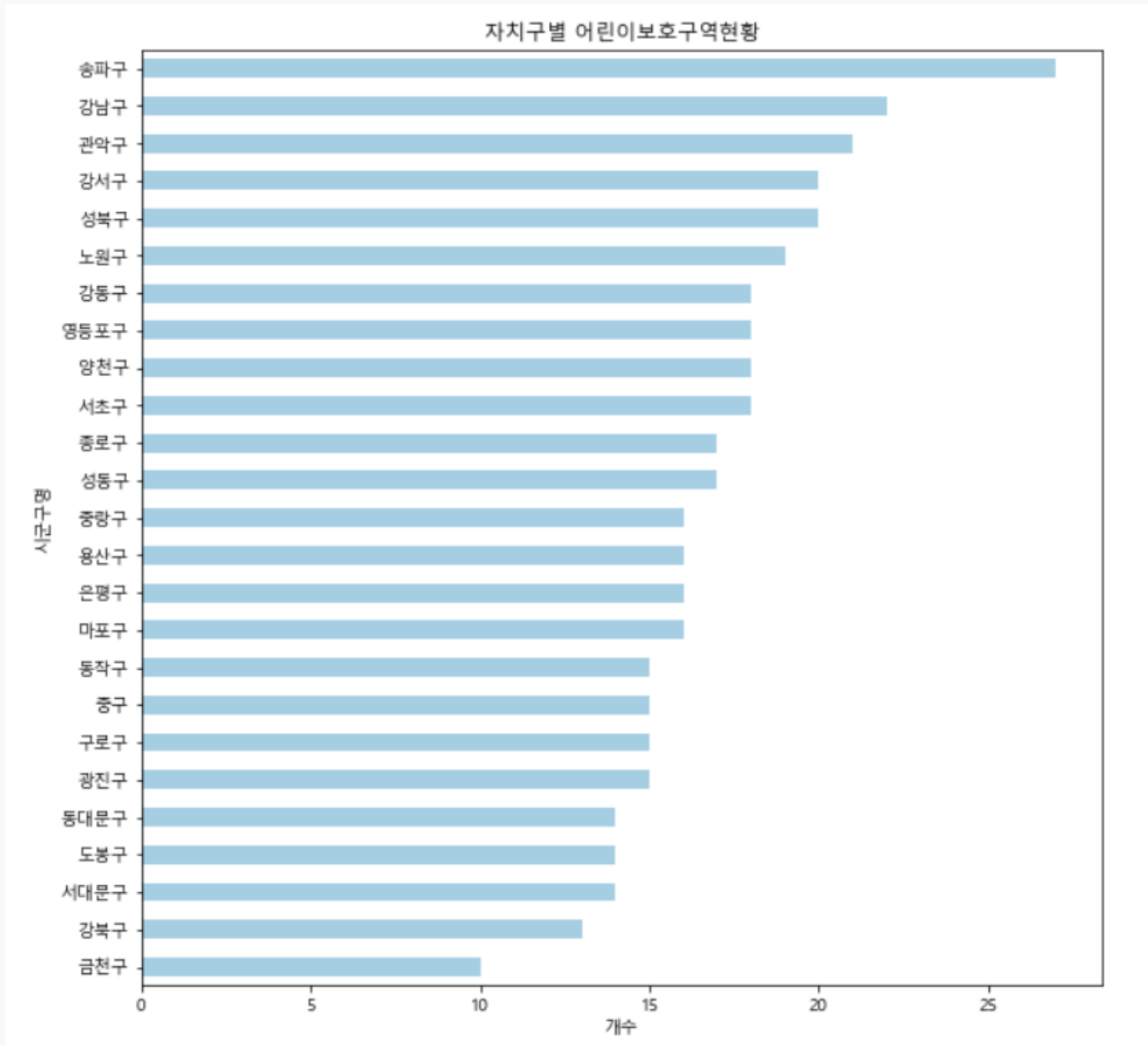
전체 데이터셋

데이터별 EDA

상관관계 분석

결론

'자치구'로 그룹화 > 자치구별 현황 집계 (행정동은 결측치가 많으므로)



어린이 보호 구역이 **많은** 자치구 : 송파구, 강남구, 관악구
어린이 보호 구역이 **적은** 자치구 : 서대문구, 강북구, 금천구

GDP 데이터

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

형태

2018GDP	2018GDP 구성비	2018GDP1 인당	2018GDP 수준지수	2019GDP	2019GDP 구성비	2019GDP1 인당	2019GDP 수준지수	2020GDP	2020GDP 구성비	2020GDP1 인당	2020GDP 수준지수
31868698	7.50	204380	468.07	33780610	7.70	218704	484.74	33893980	7.60	223603	482.96
31868698	7.50	204380	468.07	33780610	7.70	218704	484.74	33893980	7.60	223603	482.96
31868698	7.50	204380	468.07	33780610	7.70	218704	484.74	33893980	7.60	223603	482.96
31868698	7.50	204380	468.07	33780610	7.70	218704	484.74	33893980	7.60	223603	482.96
31868698	7.50	204380	468.07	33780610	7.70	218704	484.74	33893980	7.60	223603	482.96

424 rows × 12 columns

3개년(2018-2020) 인당 GDP, 구성비, 수준지수

결측치

```
gdp.isna().sum() # 결측치 없음

통계청행정동코드    0
행자부행정동코드    0
시도명              0
시군구명            0
행정동명            0
2018GDP              0
2018GDP구성비        0
2018GDP1인당         0
2018GDP수준지수      0
2019GDP              0
2019GDP구성비        0
2019GDP1인당         0
2019GDP수준지수      0
2020GDP              0
2020GDP구성비        0
2020GDP1인당         0
2020GDP수준지수      0
dtype: int64
```

결측치 없음



GDP 데이터

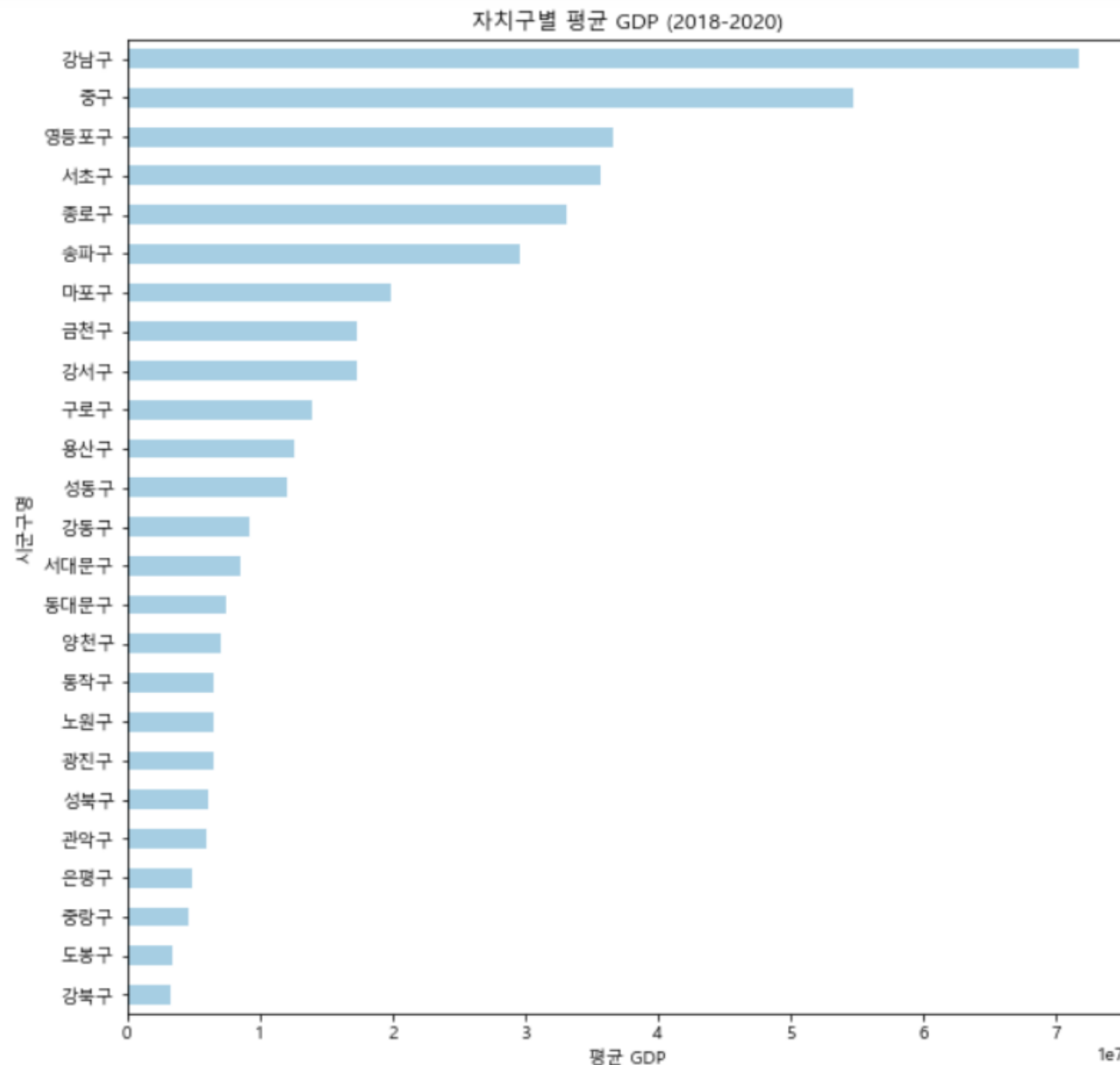
주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론



'3개년 GDP 평균' 변수 생성 > bar plot
- 자치구별 통계이므로, 행정동 무의미

평균 GDP가 **높은** 자치구 : 강남구, 중구, 영등포구
평균 GDP가 **낮은** 자치구 : 중랑구, 도봉구, 강북구

GDP 데이터

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

	2018GDP	2019GDP	2020GDP	3개년평균
count	25.00	25.00	25.00	25.00
mean	16949684.48	17437088.60	17781796.36	17389523.15
std	16799554.53	17234094.94	17948822.10	17323121.95
min	3207289.00	3283572.00	3356956.00	3282605.67
25%	6342279.00	6527100.00	6438328.00	6460191.67
50%	8825479.00	9348668.00	9317128.00	9163758.33
75%	19474677.00	19943128.00	20199596.00	19872467.00
max	69186319.00	71852681.00	74260359.00	71766453.00

GDP는 점차 높아지고 있음을 확인

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

직장인구 데이터

형태

총_직장_인구_수	연령대_20_직장_인구_수	연령대_30_직장_인구_수	연령대_40_직장_인구_수
10012.710692	1445.641509	3188.132075	2985.075472
837.927419	115.225806	236.217742	243.895161
733.290323	141.612903	230.258065	198.677419
2227.870968	205.290323	644.612903	847.806452
NaN	NaN	NaN	NaN
...
NaN	NaN	NaN	NaN
545.500000	45.822581	90.935484	114.967742
315.258065	23.516129	51.661290	54.951613
561.370656	91.339768	135.332046	133.366795
680.511521	73.917051	137.013825	169.672811

424 rows × 4 columns

결측치

총_직장_인구_수	49
연령대_20_직장_인구_수	49
연령대_30_직장_인구_수	49
연령대_40_직장_인구_수	49

49개의 행정동의 직장인구 데이터 결측



생활인구 데이터를 기준으로 클러스터링 후
해당 군집의 평균으로 보간하기로 결정 !

주제 리마인드

전체 데이터셋

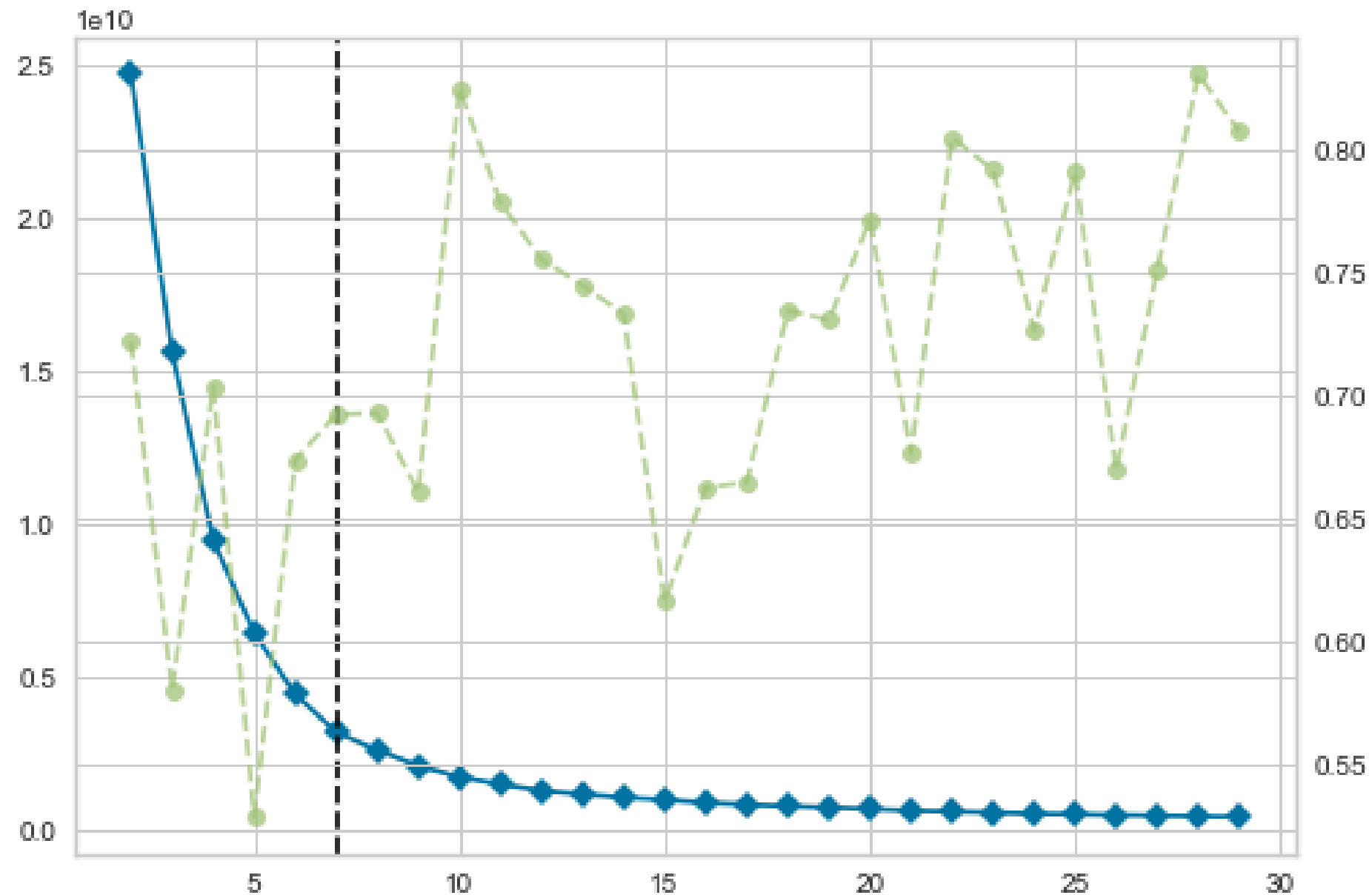
데이터별 EDA

상관관계 분석

결론

직장인구 데이터

결측치



K-Means 클러스터링 적용
최적의 군집 개수 K는 ???



[Elbow Method]

지정한 K의 범위에 대하여
각 K일 때의 군집 내 분산을 계산
: 그래프가 빠른 속도로 감소하는
elbow 지점을 적절한 K로 고려

주제 리마인드

전체 데이터셋

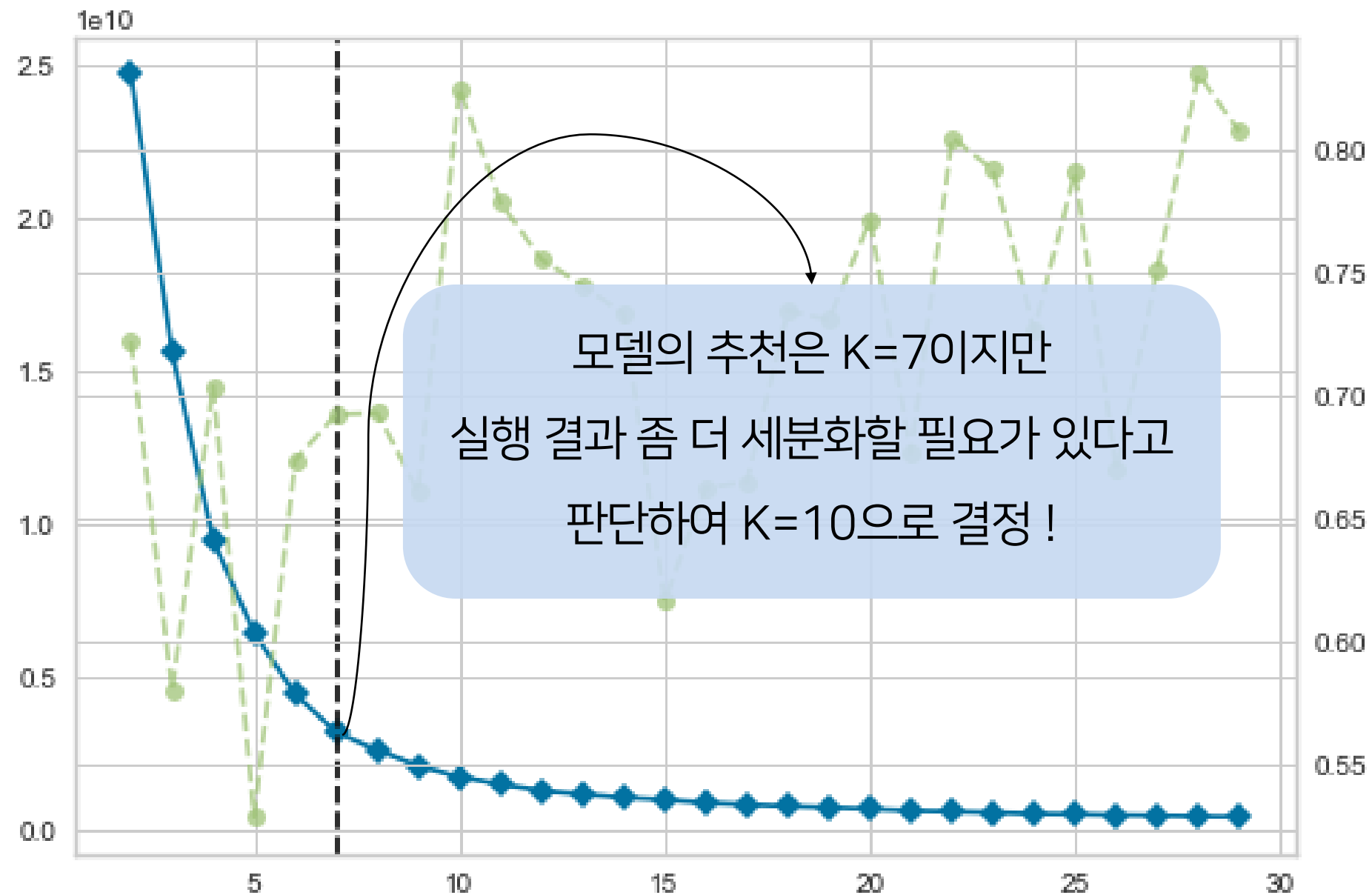
데이터별 EDA

상관관계 분석

결론

직장인구 데이터

결측치



K-Means 클러스터링 적용

최적의 군집 개수 K는 ???

군집화를 적용한 다음,

각 군집의 직장인구 수의 평균으로

결측치 보간 완료!

지정한 K의 범위에 대하여

각 K일 때의 군집 내 분산을 계산

: 그래프가 빠른 속도로 감소하는

elbow 지점을 적절한 K로 고려

직장인구 데이터

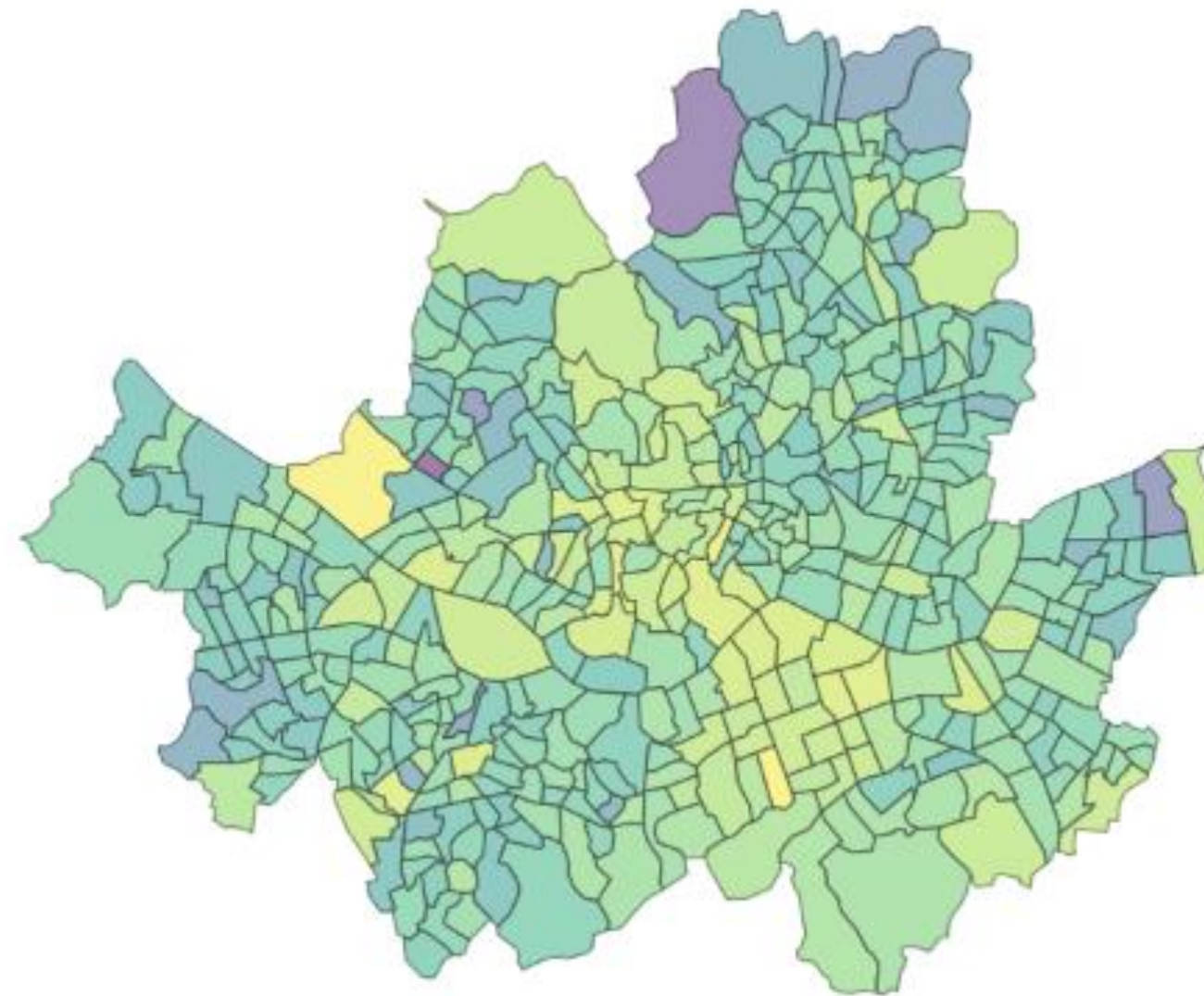
주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론



20-30대 직장인구비

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

전체 직장인구 대비 20-40대 직장인구 비율이 **높았던** 행정동 : 마포구 상암동 및 용산, 서초, 강남구 일대
전체 직장인구 대비 20-40대 직장인구 비율이 **낮았던** 행정동 : 강북, 도봉, 노원구 일대

한부모/저소득 가정 데이터

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

형태

	2021_한부모_가구 수	2021_저소득_가구 수	2020_한부모_가구 수
0	96	214	114
1	96	214	114
2	96	214	114
3	96	214	114
4	96	214	114
...
419	189	1073	268
420	189	1073	268
421	189	1073	268
422	189	1073	268
423	189	1073	268

424 rows × 8 columns

결측치

2021_한부모_가구수	0
2021_저소득_가구수	0
2020_한부모_가구수	0
2020_저소득_가구수	0
2019_한부모_가구수	0
2019_저소득_가구수	0
2018_한부모_가구수	0
2018_저소득_가구수	0

결측치 없음

한부모/저소득 가정 데이터

주제 리마인드

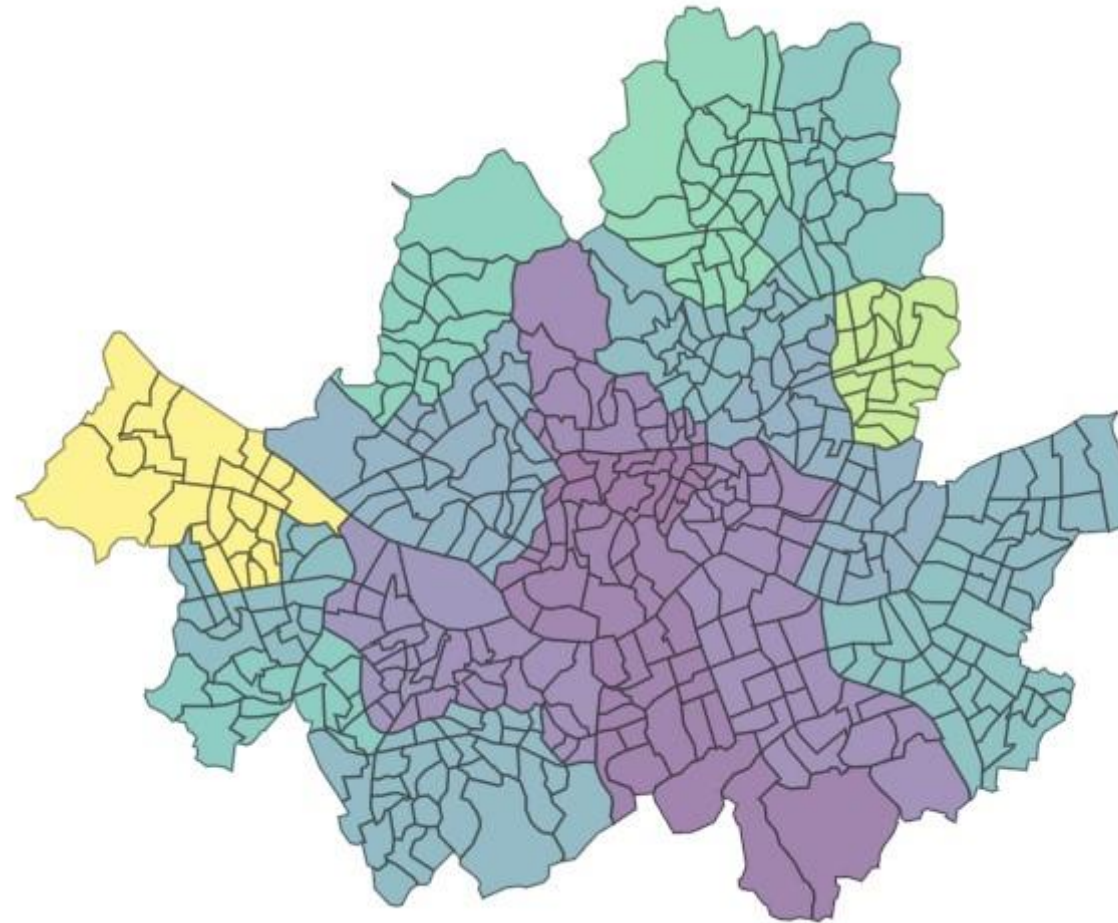
전체 데이터셋

데이터별 EDA

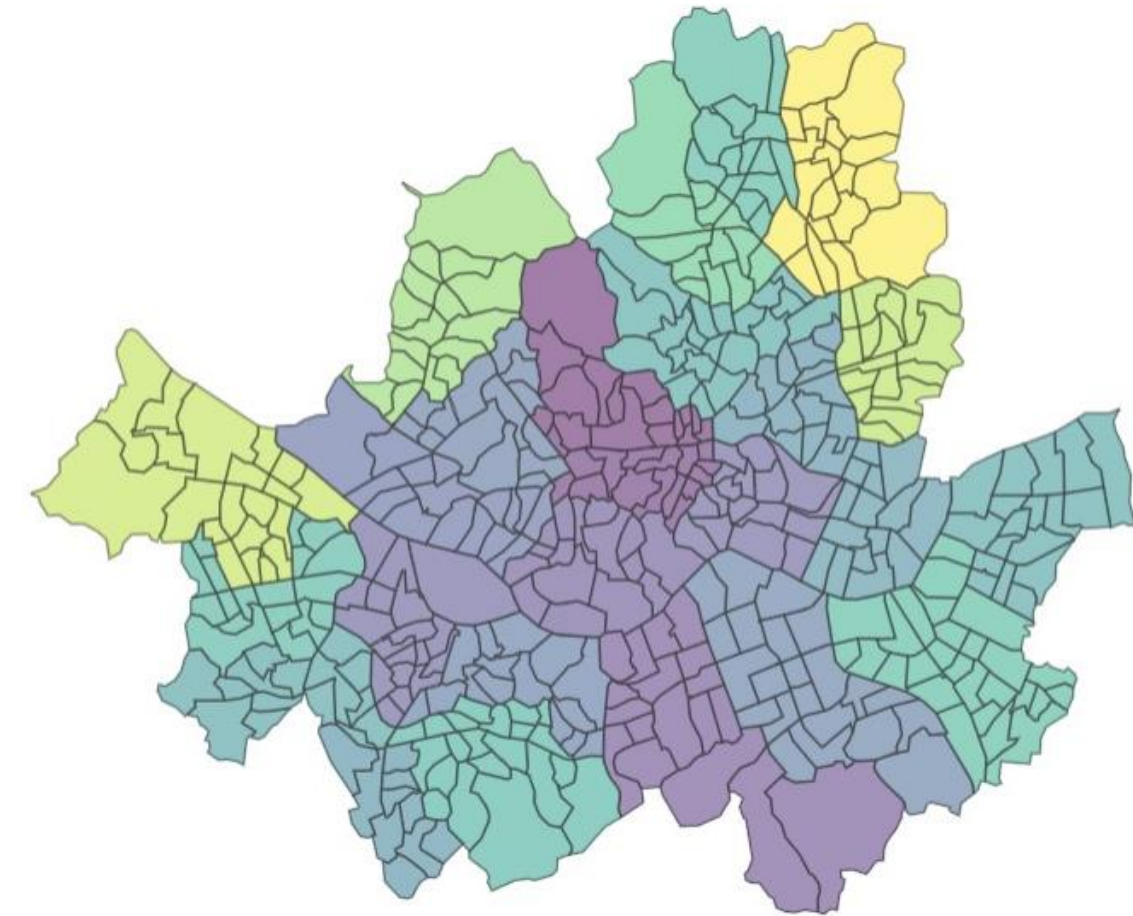
상관관계 분석

결론

한부모 가정



저소득 가정



한부모/저소득 가정이 많았던 자치구 : 강서구, 은평구, 강북구, 도봉구, 노원구, 중랑구

자치구 예산 데이터

- 주제 리마인드
- 전체 데이터셋
- 데이터별 EDA
- 상관관계 분석
- 결론

형태

2021_예산현액	2020_예산현액	2019_예산현액	2018_예산현액
662770	682470	562397	497363
662770	682470	562397	497363
662770	682470	562397	497363
662770	682470	562397	497363
662770	682470	562397	497363
...
1170691	1154710	896075	749439
1170691	1154710	896075	749439
1170691	1154710	896075	749439
1170691	1154710	896075	749439
1170691	1154710	896075	749439

424 rows × 12 columns

결측치

2021_예산현액	0
2021_세입	0
2021_세출	0
2020_예산현액	0
2020_세입	0
2020_세출	0
2019_예산현액	0
2019_세입	0
2019_세출	0
2018_예산현액	0
2018_세입	0
2018_세출	0

결측치 없음

자치구 예산 데이터

주제 리마인드

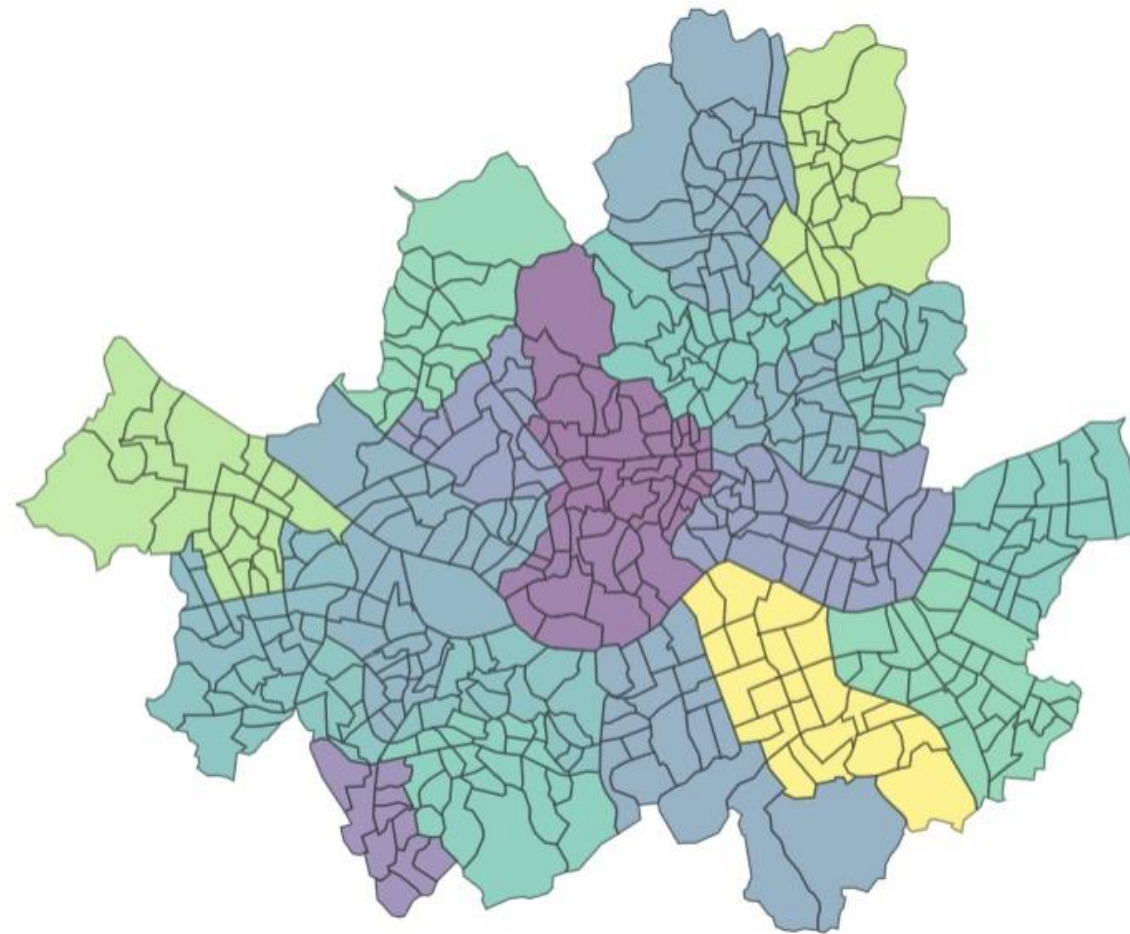
전체 데이터셋

데이터별 EDA

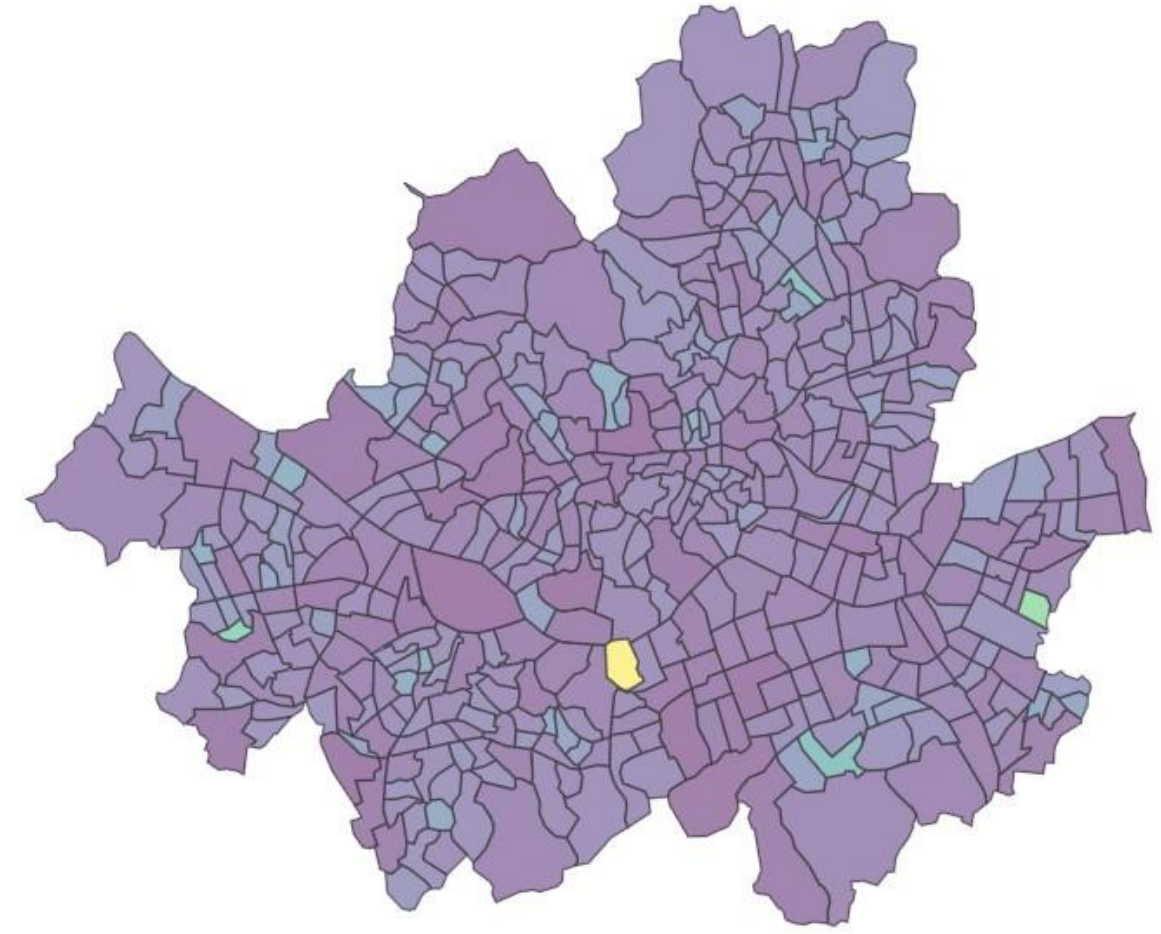
상관관계 분석

결론

자치구별 예산 현액



생활인구 대비 예산



예산 현액 자체는 강남구가 타 자치구에 비해 높게 나타났지만,
생활인구 대비 예산은 대부분의 자치구 및 행정동에서 비슷하게 나타났음



- 주제 리마인드
- 전체 데이터셋
- 데이터별 EDA
- 상관관계 분석
- 결론

어린이집 데이터



형태

어린이집_수
9.0
3.0
3.0
3.0
4.0
...
4.0
13.0
14.0
8.0
16.0

결측치

어린이집_수5

면 적 3.8 평	1.0	311	1776	434	1794	...	4741739	1.1	12174	26.98	4750297	1.1	12340	26.65	NaN	NaN
인 원 2 평	1.0	102	745	129	766	...	71852681	16.5	141362	313.32	74260359	16.7	147023	317.55	NaN	NaN
잠 실 7 평	0.0	206	1280	300	1348	...	29962528	6.9	46033	102.03	30336592	6.8	46082	99.53	NaN	NaN
상 일 동	0.0	189	1073	268	1013	...	9348668	2.1	22807	50.55	9317128	2.1	23271	50.26	NaN	NaN
문 춘 1 동	0.0	189	1073	268	1013	...	9348668	2.1	22807	50.55	9317128	2.1	23271	50.26	NaN	NaN

5개 행정동 데이터에 어린이집 수 데이터 결측

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

어린이집 데이터

결측치

```
data.loc[108, '어린이집_수']=12  
data.loc[354, '어린이집_수']=7  
data.loc[401, '어린이집_수']=1  
data.loc[407, '어린이집_수']=25  
data.loc[419, '어린이집_수']=4
```

```
df = data[['총_직장_인구_수', '연령대_20_직장_인구_수', '연령대_30_직장_인구_수', '연령대_40_직장_인구_수']]  
df.isnull().sum()
```

```
총_직장_인구_수      0  
연령대_20_직장_인구_수  0  
연령대_30_직장_인구_수  0  
연령대_40_직장_인구_수  0  
dtype: int64
```

결측 개수가 많지 않아, 개별 행정동에 대해 직접 어린이집 정보 검색!
검색 결과를 바탕으로 어린이집 수 결측치 보간

어린이집 데이터

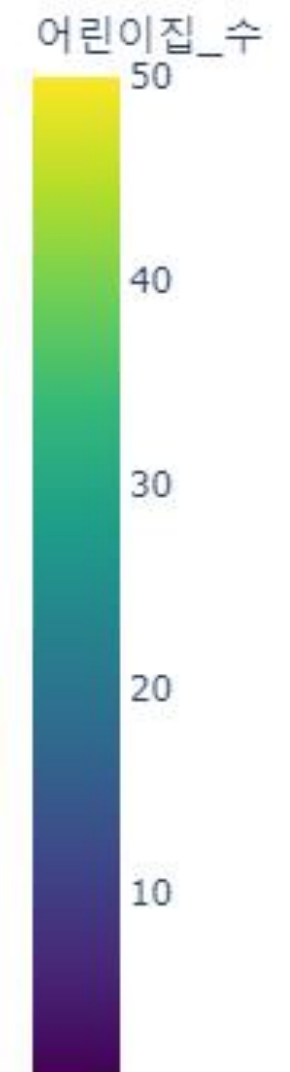
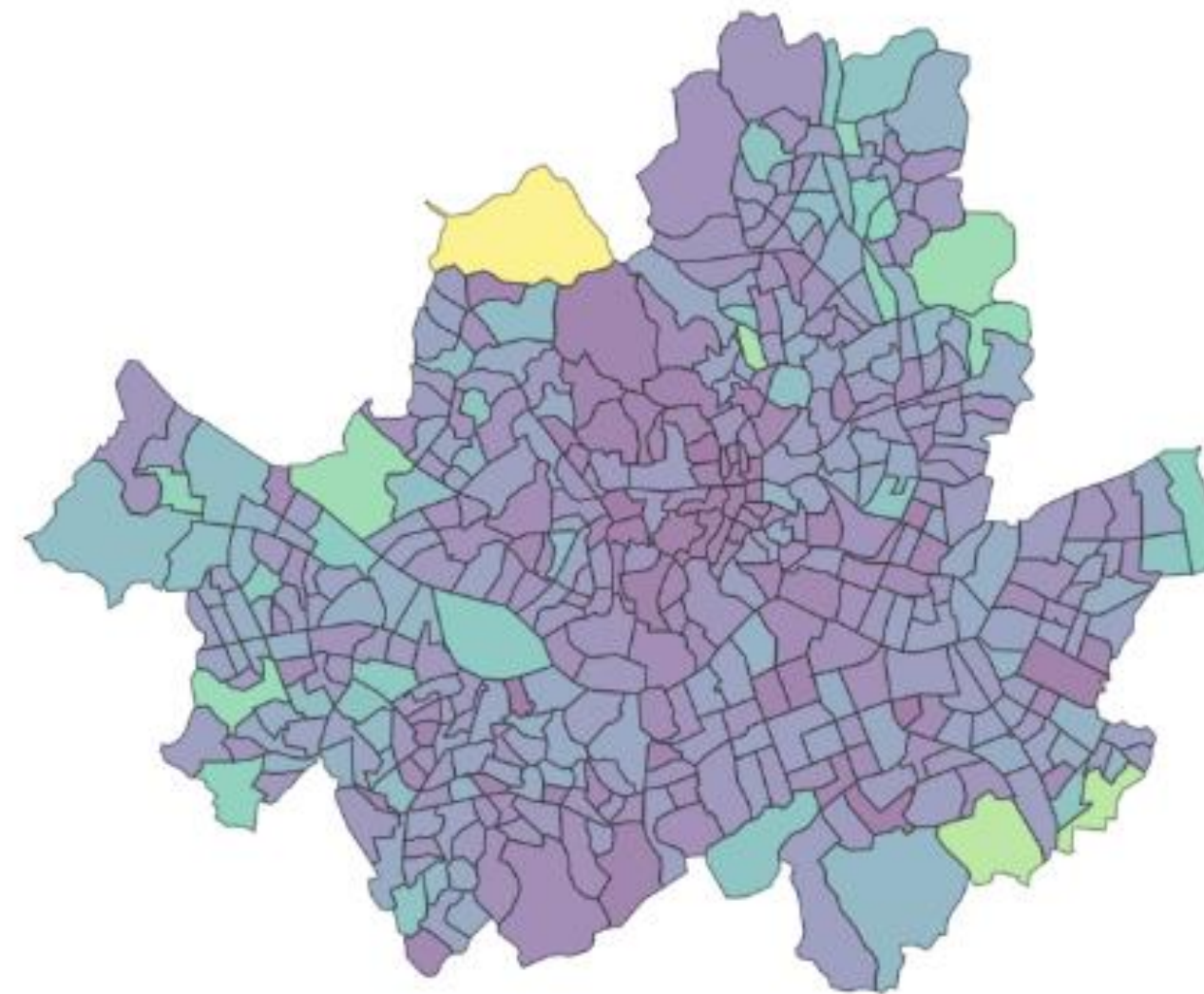
주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론



어린이집 수가 많았던 행정구 : 진관동, 공릉2동, 세곡동, 여의동, 상암동

어린이집 데이터

주제 리마인드

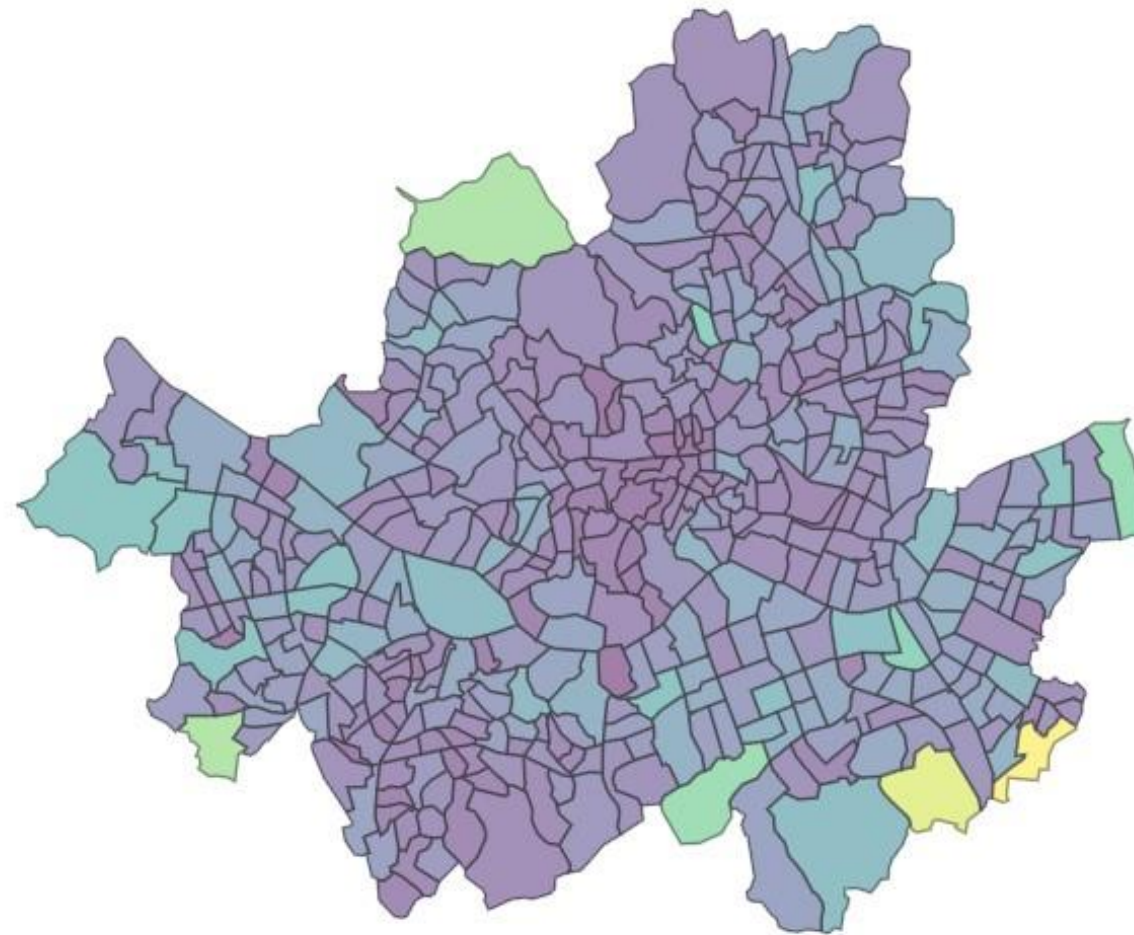
전체 데이터셋

데이터별 EDA

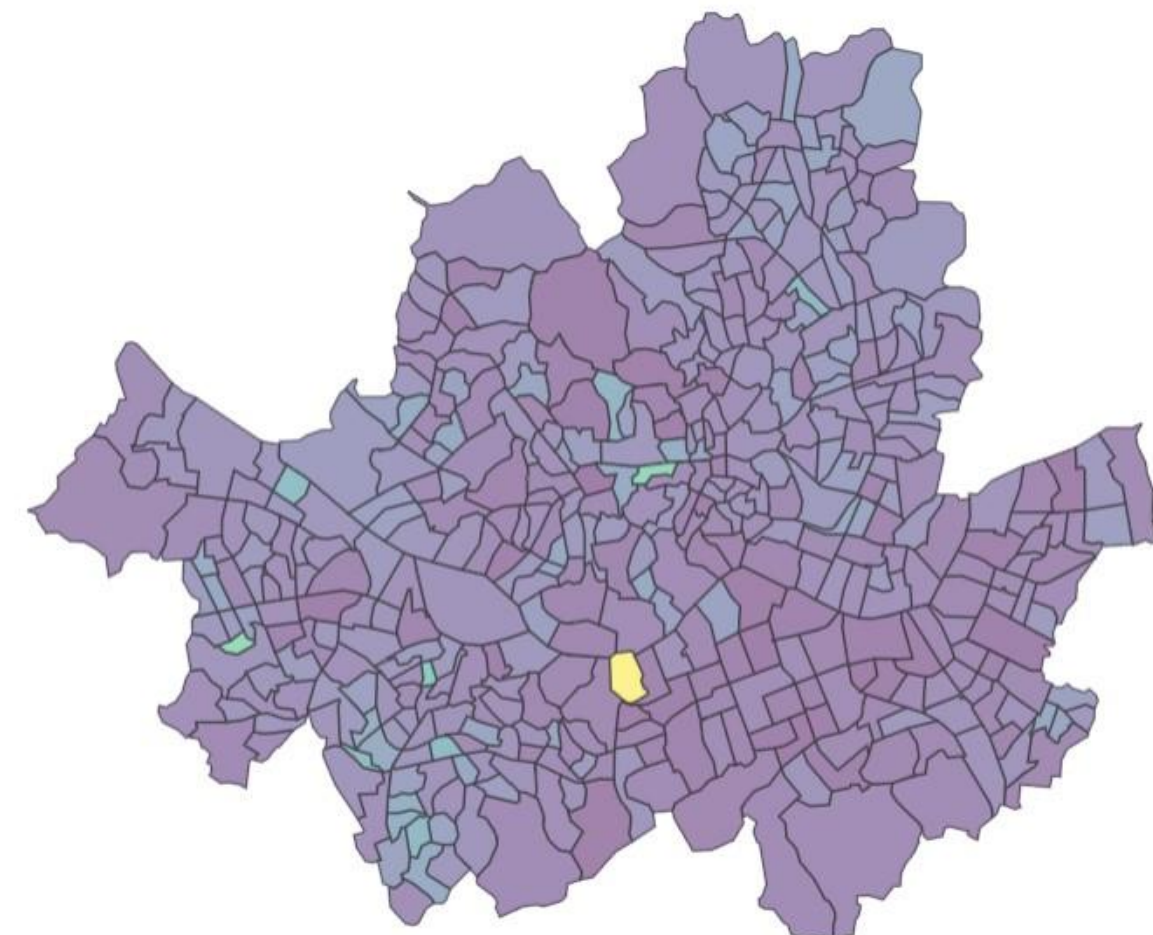
상관관계 분석

결론

유아인구



유아인구 대비 어린이집 수



유아인구가 많은 행정동과 어린이집 수가 많은 행정동이 대부분 겹침

= 유아인구 대비 어린이집 수는 모든 행정동에서 비슷하게 나타남

(반포본동은 유아인구 대비 어린이집 수가 상대적으로 많았음)



- 주제 리마인드
- 전체 데이터셋
- 데이터별 EDA
- 상관관계 분석
- 결론

공동육아나눔터 데이터



형태

	시군구명	행정동명	공동육아나눔터유무
0	종로구	사직동	0.0
1	종로구	삼청동	0.0
2	종로구	부암동	0.0
3	종로구	평창동	0.0
4	종로구	무악동	0.0
...
419	강동구	둔촌1동	0.0
420	강동구	둔촌2동	0.0
421	강동구	암사1동	0.0
422	강동구	천호2동	0.0
423	강동구	길동	0.0

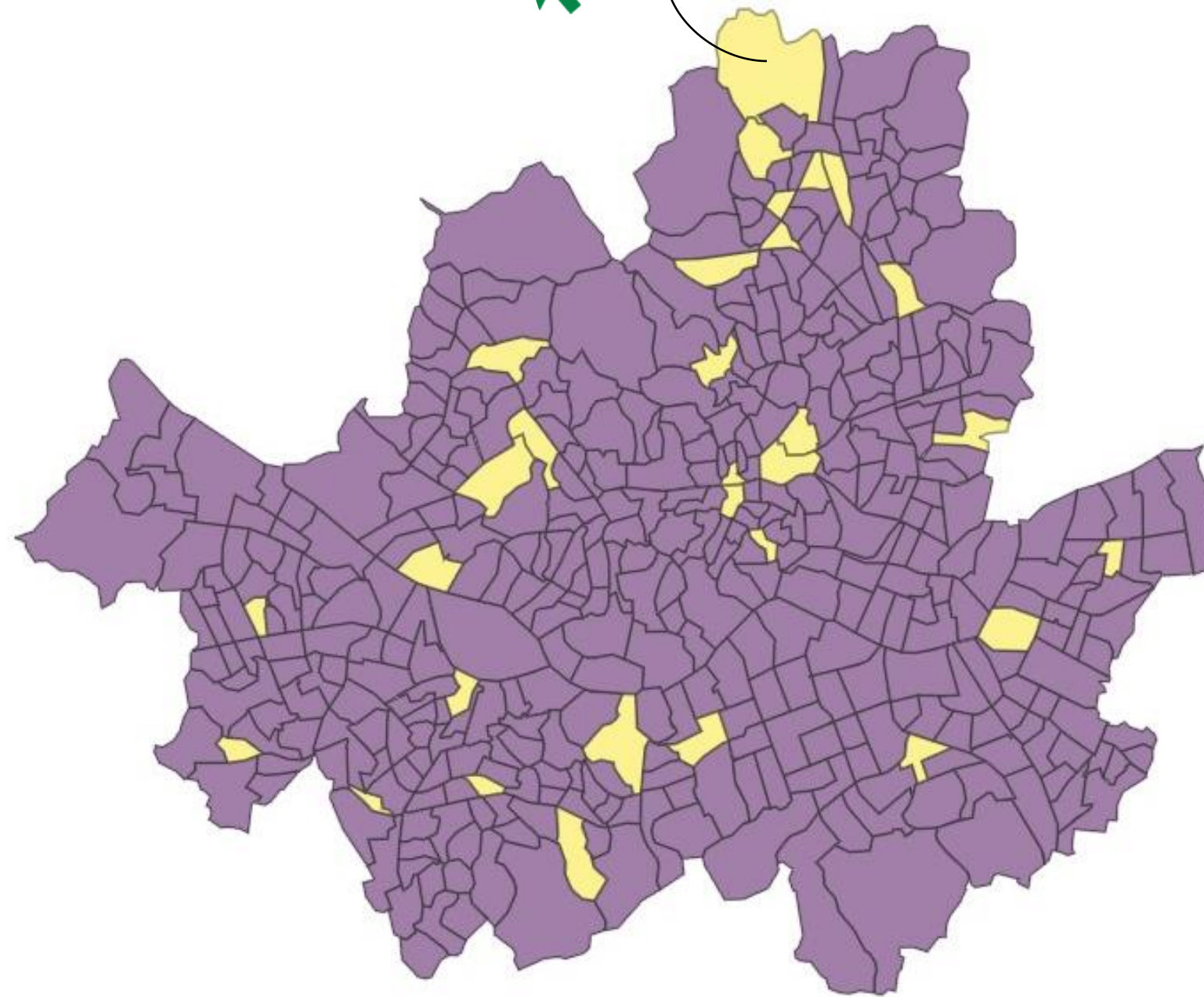
424 rows × 1 columns

결측치

공동육아나눔터유무 0
결측치 없음

공동육아나눔터 데이터

공동육아나눔터 有



용산구, 서초구, 강남구 일대 **상업지역**과 **외곽 주거지역**에 공동육아나눔터가 부족해 보임

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

공동육아나눔터 데이터

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

	2021_한부모_가구수	2021_저소득_가구수	총생활인구수	2021_예산현액	총_직장_인구_수	2022혼인	어린이보호_구역현황	키즈카페_수	어린이집_수	아동인구	아동인구비
count	424.000000	424.000000	424.000000	4.240000e+02	424.000000	422.000000	279.000000	150.00000	424.000000	424.000000	424.000000
mean	178.299528	1018.080189	24498.104517	1.073777e+06	1619.203171	84.099526	3.908602	1.56000	11.379717	2530.989266	0.103867
std	72.482996	488.041920	12256.917916	2.354672e+05	4132.004323	47.096702	2.239218	0.95172	7.030475	1662.480376	0.040175
min	72.000000	214.000000	2710.765852	6.534420e+05	8.181818	5.000000	1.000000	1.00000	1.000000	221.735909	0.014165

	2021_한부모_가구수	2021_저소득_가구수	총생활인구수	2021_예산현액	총_직장_인구_수	2022혼인	어린이보호_구역현황	키즈카페_수	어린이집_수	아동인구	아동인구비	20-30대직장인구
count	30.000000	30.000000	30.000000	3.000000e+01	30.000000	29.000000	17.000000	10.000000	30.000000	30.000000	30.000000	30.000000
mean	194.733333	1064.066667	24948.295846	1.061853e+06	759.558143	81.689655	3.235294	1.600000	10.766667	2275.559945	0.088421	504.953011
std	70.297856	454.762836	9559.198653	1.992704e+05	954.696790	32.601603	2.194579	1.074968	5.130593	1353.240675	0.035179	742.406549
min	72.000000	214.000000	8542.106721	6.534420e+05	109.310811	22.000000	1.000000	1.000000	2.000000	225.324190	0.022910	72.229730

공동육아나눔터가 없는 행정동과 비교한 공동육아나눔터가 있는 행정동의 특징

- 한부모/저소득 가구 수가 더 많음 = 복지제도적 특징이 드러남
 - 예상과는 달리 아동 및 신혼부부 인구가 평균 이하
- 직장인구 수가 평균 이하 = 외벌이 가정이 많은 것으로 예상됨

행정동별 인구 관련

총생활인구수, 직장인구수, 혼인평균 상관관계

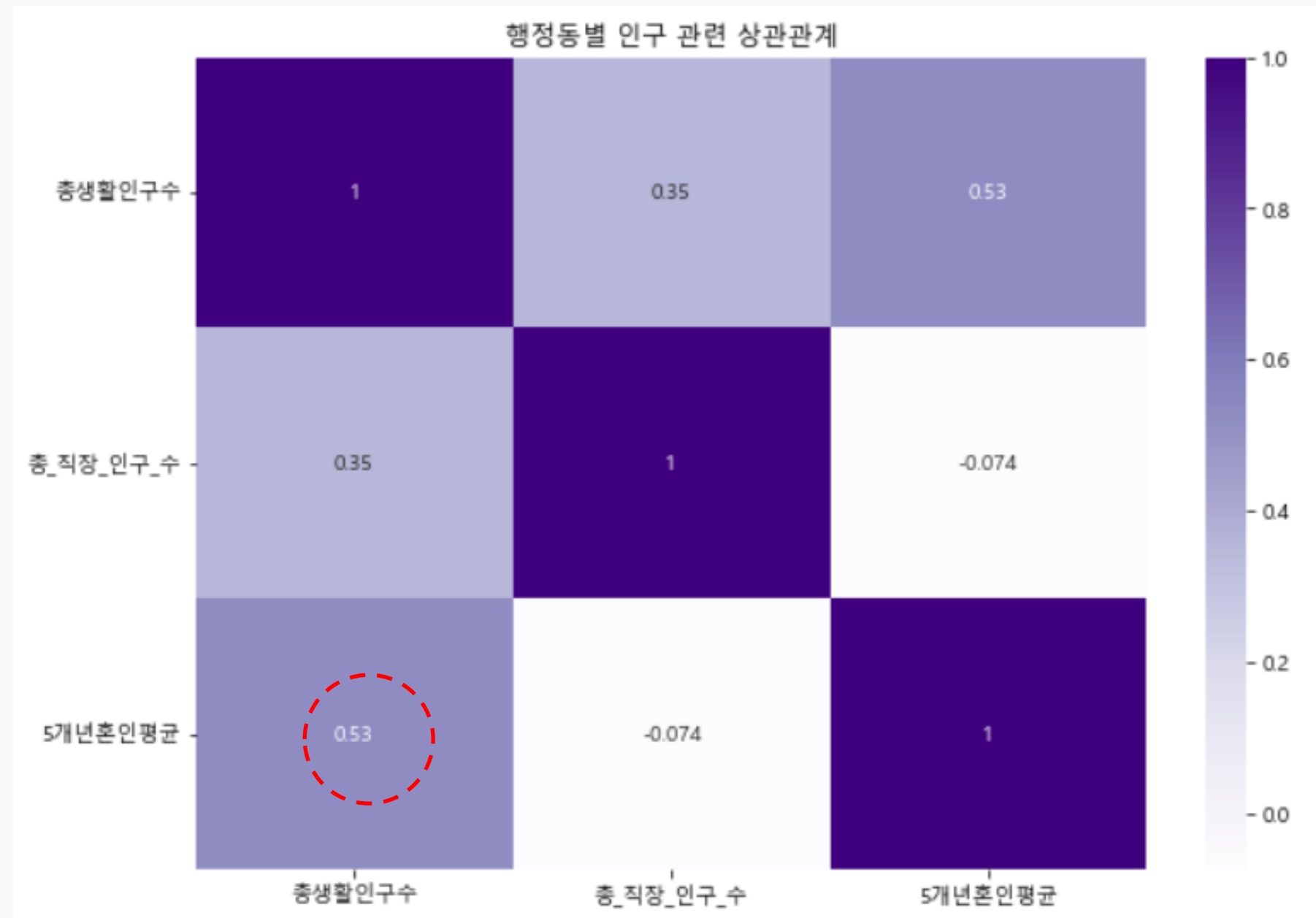
주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론



- 총생활인구수와 5개년혼인평균은 0.53의 높은 상관관계
- 직장인구수와 생활인구수에는 차이 존재, 0.35의 상관관계
 - 직장인구수와 혼인은 관계 없음

아이 관련

공동육아나눔터유무, 어린이보호구역, 어린이집수, 아이 수 상관관계

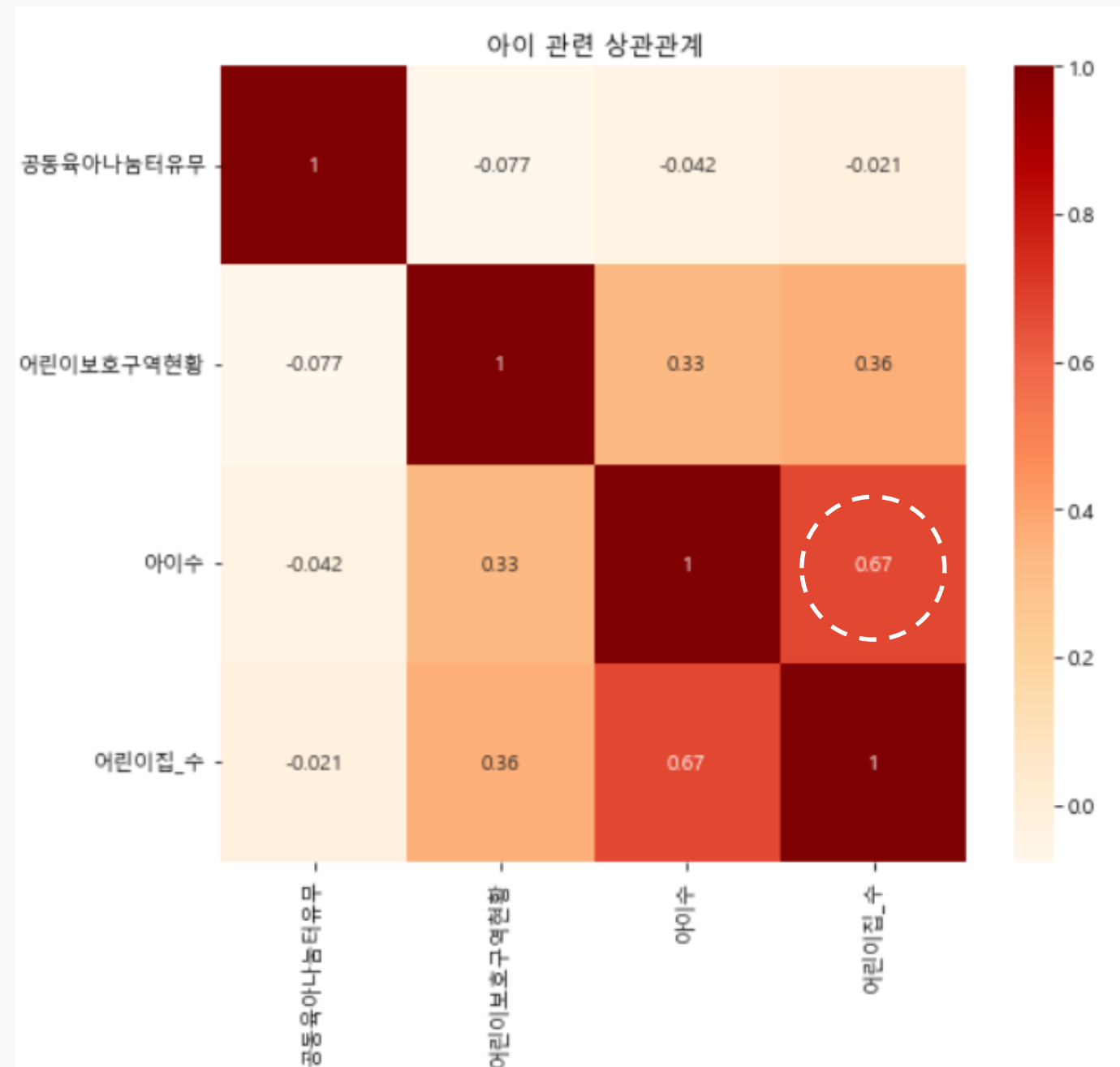
주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론



- 아이수와 어린이집 수는 0.67의 높은 상관관계
- 어린이보호구역현황은 아이수 / 어린이집 수와 약 0.3의 상관관계
- 공동육아나눔터유무는 어린이보호구역현황과 아이수, 어린이집수와 관련이 없음

어린이보호구역 데이터에는 결측치가 많았고,
공동육아나눔터는 개수가 많지 않기 때문에
유의미한 결과가 나오기 어려움

전체 변수 상관관계

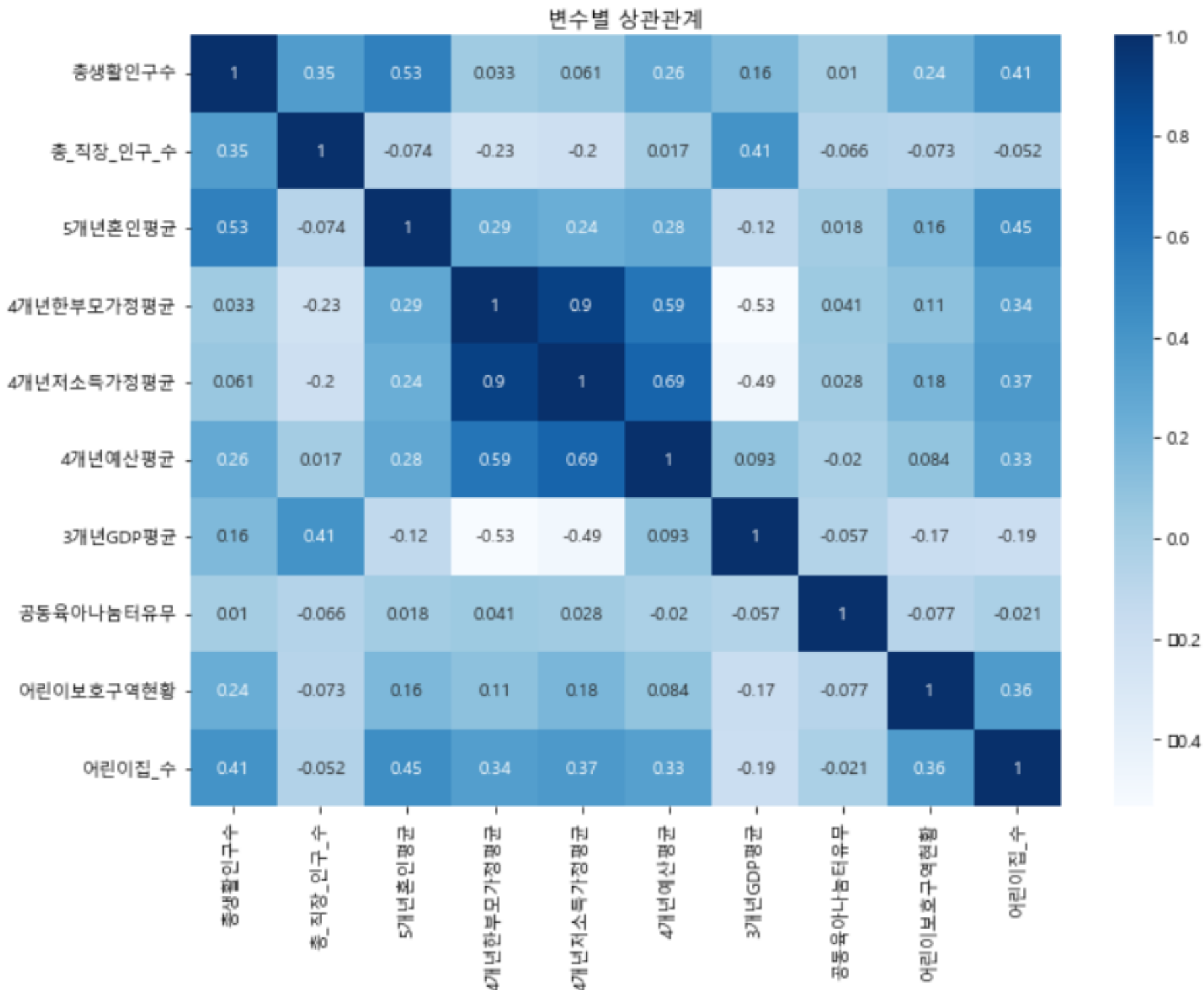
주제 리마인드

전체 데이터셋

데이터별 EDA

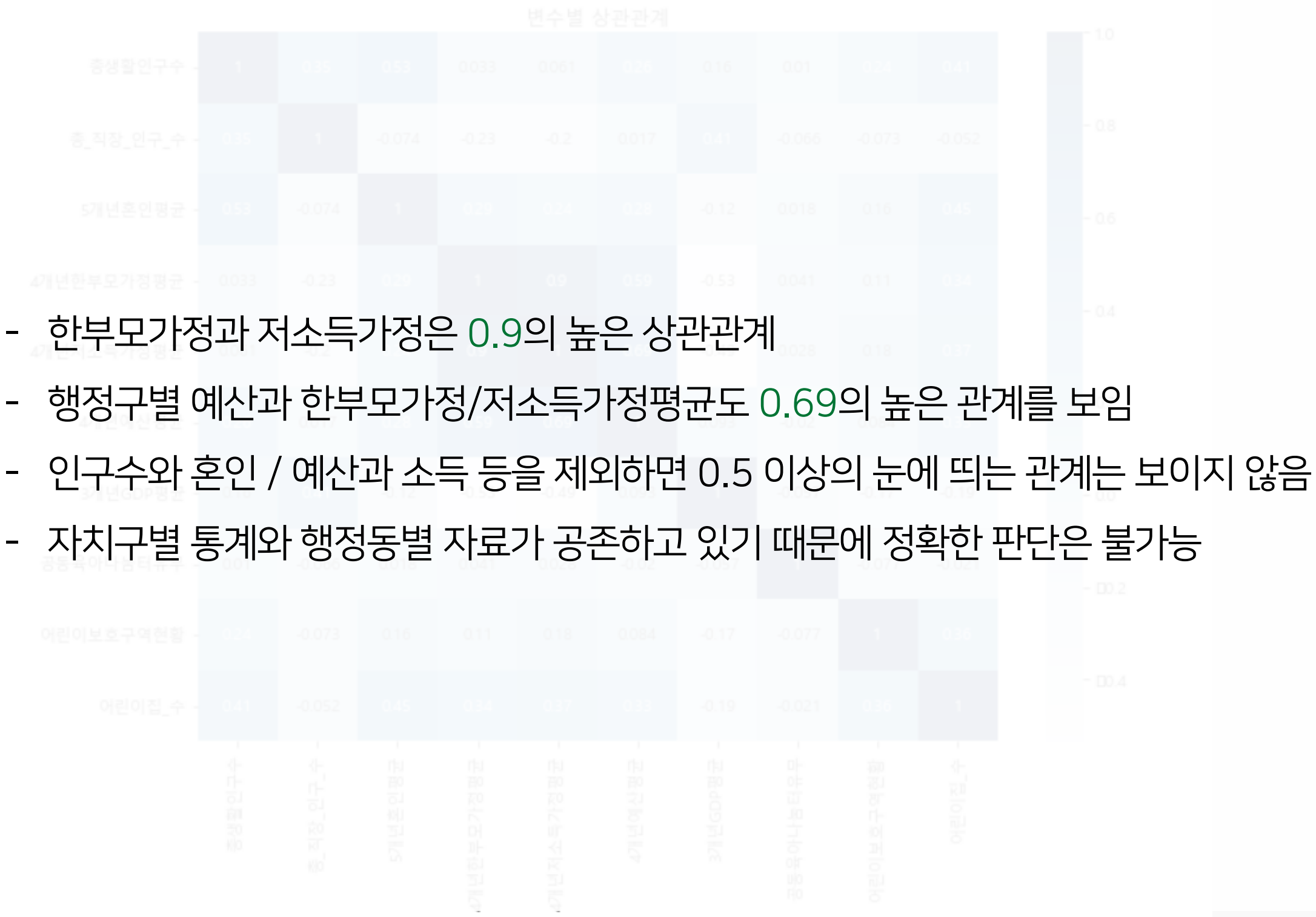
상관관계 분석

결론



전체 변수 상관관계

- 주제 리마인드
- 전체 데이터셋
- 데이터별 EDA
- 상관관계 분석
- 결론



EDA 결과 요약

주제 리마인드

전체 데이터셋

데이터별 EDA

상관관계 분석

결론

데이터 수집 & 전처리

- 공동육아나눔터 입지선정을 위한 아동, 인구, 경제 관련 9종류의 데이터 수집
- 행정동을 추출하여 전체 데이터를 행정동 기준으로 병합
- 결측치 보간

공동육아나눔터 설치 행정동 특징

- 서울 서부 외곽 주거지역 및 20-40대 직장인구가 많은 상업지역에 공동육아나눔터 부족
- 상대적으로 저소득/한부모 가정의 수가 높음 → 복지제도적 성격을 가짐
- 상대적으로 직장인구 수가 적음 → 외벌이 가정이 많다 = 가구 내 수입이 상대적으로 적다
- 유아 인구 및 신혼부부 연령대 인구가 전체 평균보다 낮음 → 인구 분포에 따른 배치 필요

상관분석

- 생활인구와 혼인 수, 유아 인구와 어린이집 수가 유의미한 양의 상관관계를 가짐
- 한부모 가정 수와 저소득 가정 수는 높은 양의 상관관계를 가짐 & 자치구 예산과도 높은 상관관계를 가짐

감사합니다 😊