

A stylized illustration of a city skyline. In the foreground, there is a traditional Korean building with a red base and green tiled roof. Behind it are several modern skyscrapers in various colors (white, orange, yellow, blue). A large orange sun is on the left, and a white cloud is on the right. The background is a light yellow color.

서울시 요식업 창업을 위한 업종&지역 추천 알고리즘 학습용 데이터 구축

I • MULCAM • U

TABLE OF CONTENTS

- 
- 01 프로젝트 배경
- 02 팀 구성 및 역할
- 03 수행절차 및 방법
- 04 결론 및 향후 과제

01

프로젝트 배경



1. 프로젝트 배경

청년 창업 열기 지속... '젊은 사장님' 증가율 1위

| 고용 부진 영향...부동산업 제외 30세 미만 창업기업 상반기 8만개로 15% ↑

어디에?

무슨 업종을?

02

팀 구성 및 역할



OUR TEAM



이새봄

서울 생활인구, 주민등록인구
(내국인, 외국인),
가구유형 데이터,
수집 및 전처리, 시각화



이주남

서울특별시 일반음식점 인
허가 정보
서울특별시 학교 데이터
수집 및 전처리, 시각화



★황수빈

서울특별시 일반음식점
인허가 정보
서울특별시 병원 데이터
수집 및 전처리, 시각화

OUR TEAM



여정문

서울특별시 일반음식점
인허가 정보
수집 및 전처리, 시각화
매출액, 개업률·폐업률
서울특별시 학교 기본 정보
시각화



왕기영


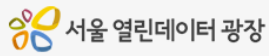





매출액, 개업률·폐업률,
임대료 데이터
수집 및 전처리, 시각화

03

수행절차 및 방법



WORK-FLOW

구분	기간	활동	도구
사전 기획	9/8(수) ~ 9/10(금)	● 주제 선정 및 일정 수립	
분석	9/13(월) ~ 9/23(목)	● AWS 서버 구축 및 하둡, 스파크 설치/ 데이터 전처리	 
수정/보완	9/24(금)	● 데이터 DB 저장	 
	9/25(토)~9/26(일)	● 데이터 시각화	
	9/26(일)~9/29(수)	● 발표준비 및 포트폴리오 제작	 
프로젝트 발표	9/30(목)	● 프로젝트 발표	

DATA SCHEMA

출처

데이터이름

제공 형태

요약

서울
열린데이터 광장

서울특별시 일반음식점 인허가 정보

CSV

구별로 분할하여 업태로 구분

서울특별시 학교 기본 정보

CSV

구별로 학교정보 구분

자치구단위 서울생활인구 일별 집계표

CSV

구별로 야간·주간 서울생활
인구 구분
(2018.04~2019.12)

서울시 가구원수별 가구수(구별) 통계

CSV

구별로 1인 ~ 7인
이상 가구수 분리

서울시 우리마을가게 상권분석서비스(상권-추정매출)

CSV

구별로 남·녀 매출비율,
00-24시 매출비율 및
연령별 매출비율

서울시 주민등록인구 (구별) 통계

CSV

구별로 2017-2019년도별
남자·여자 및 총인구

DATA SCHEMA

출처

데이터이름

제공 형태

요약

서울
열린데이터 광장

서울특별시 우리마을가게 상권분석서비스 지
역/상권별 현황 중 개/폐업수(률)

json

웹크롤링 이용

서울특별시 병원 인허가 정보

csv

구별 병원 정보

서울시 등록외국인 현황 (연령별/구별) 통계

csv

구별로 구분하여 연령별
통계

서울시 주민등록인구 정보 (연령대)

csv

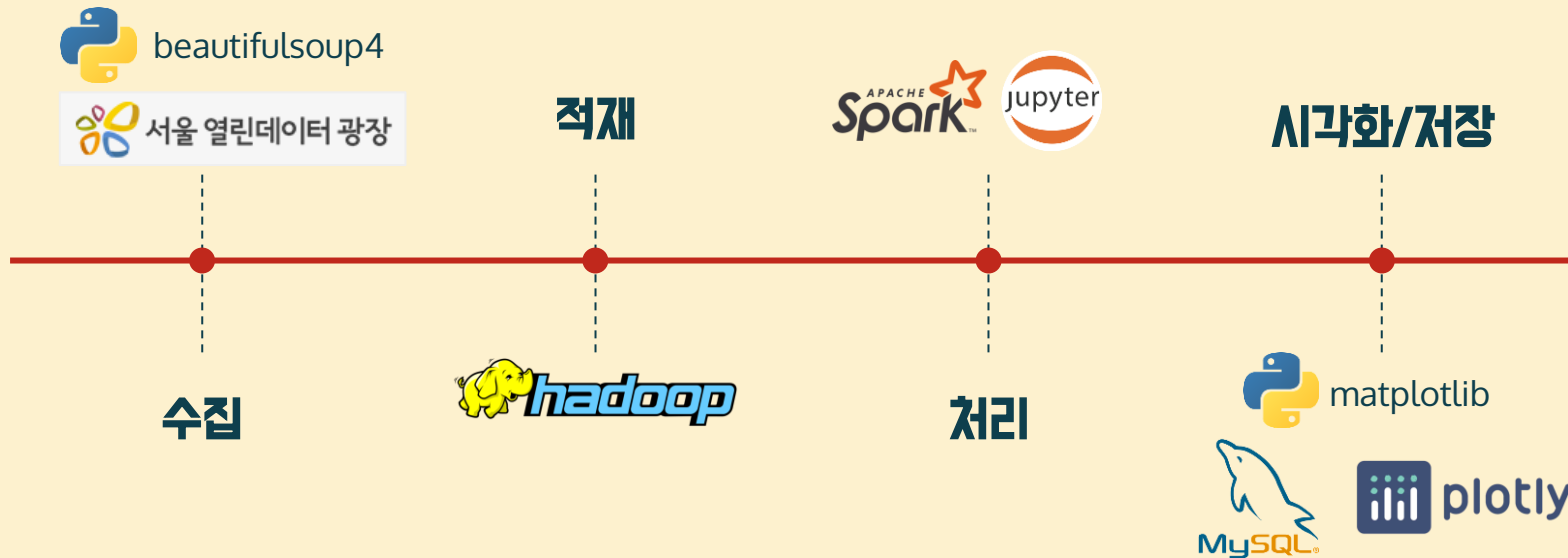
읍면동별 5살 구간으로
나누어진 연령별 통계

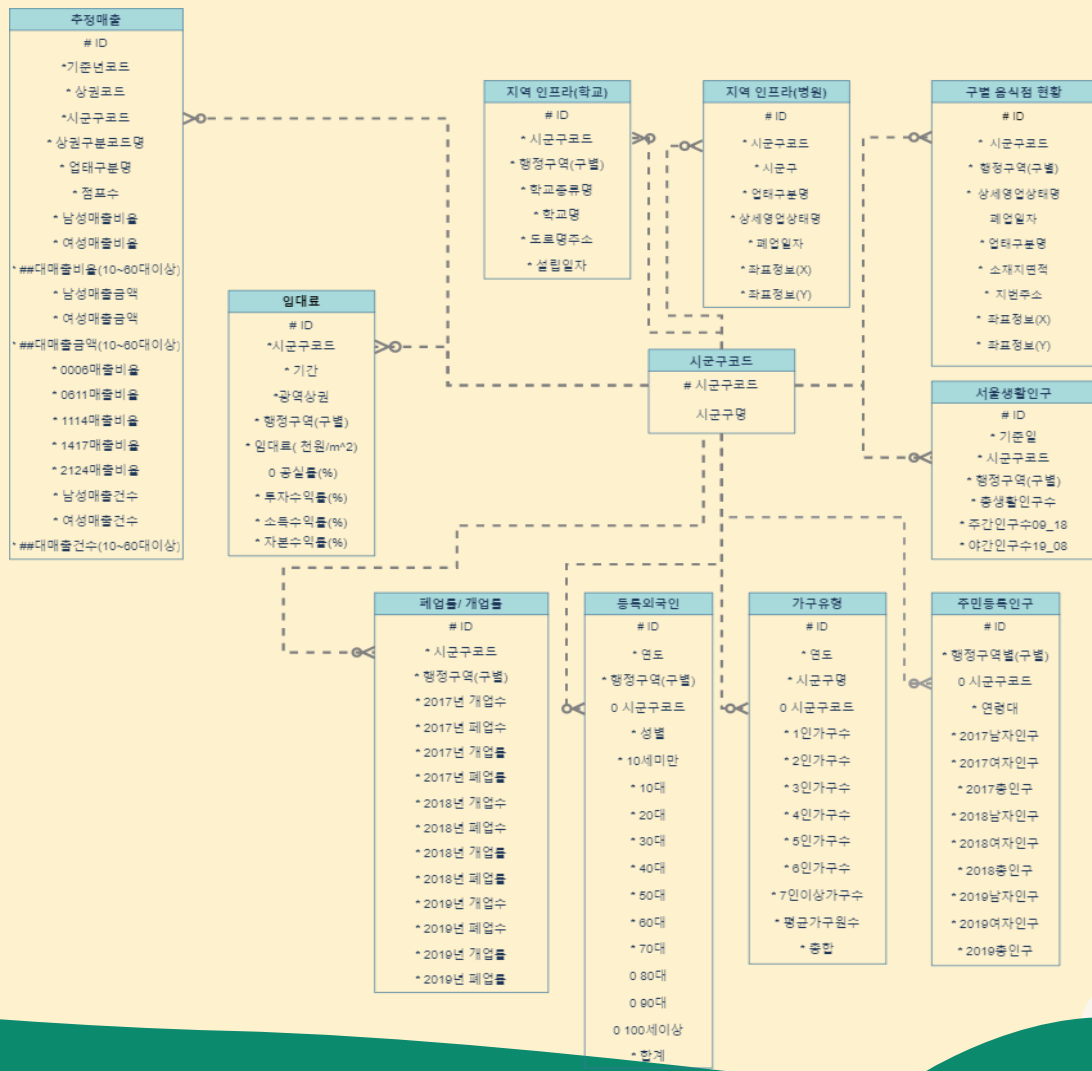
서울시 매장용빌딩 임대료·공실을 및 수익률 통
계

xls

상권별로 임대료, 공실을
· 투자수익률, 소득수익
률, 자본수익률 구분
(원본 데이터에서
5개구가 생략됨)

DATA PIPELINE





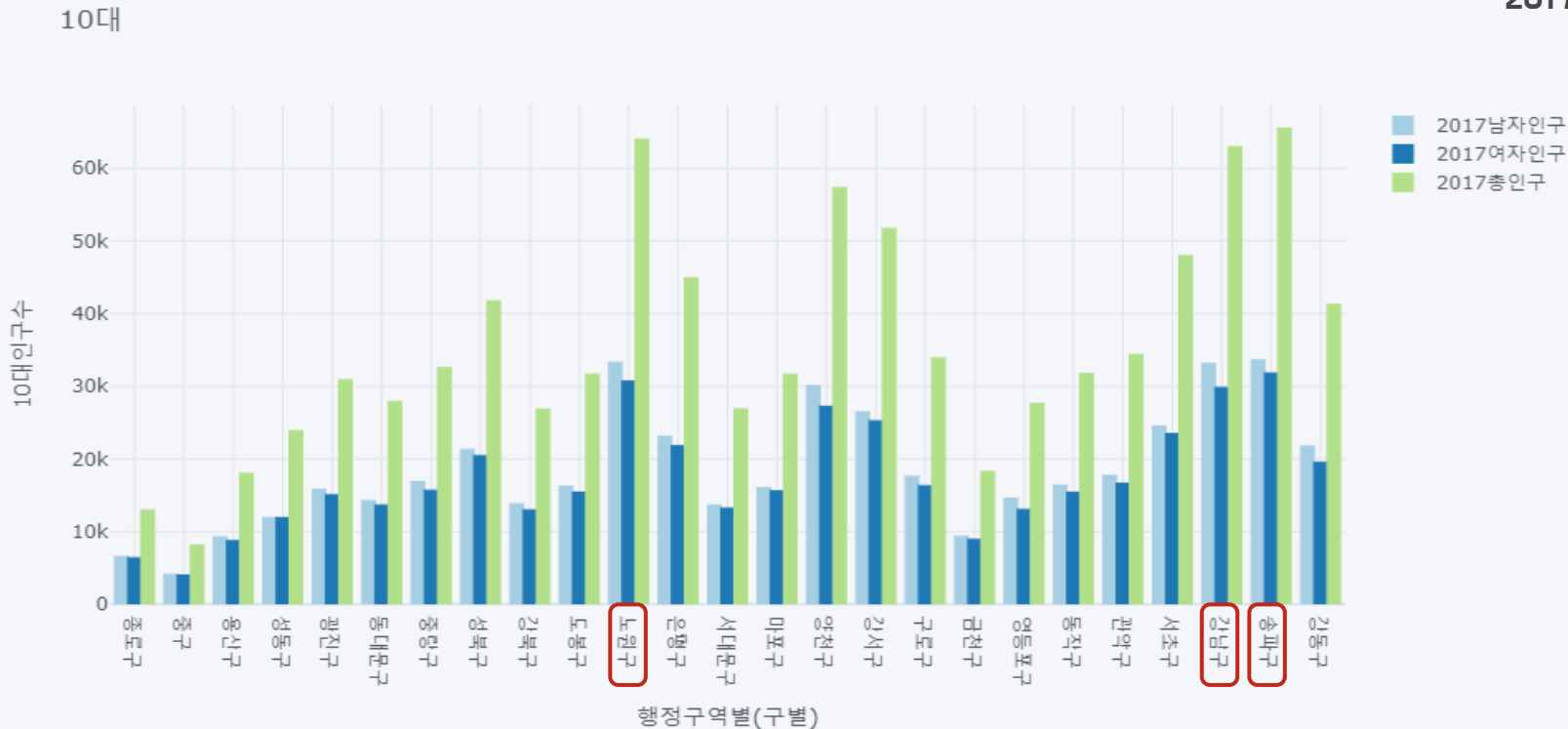
04

테이터 시각화

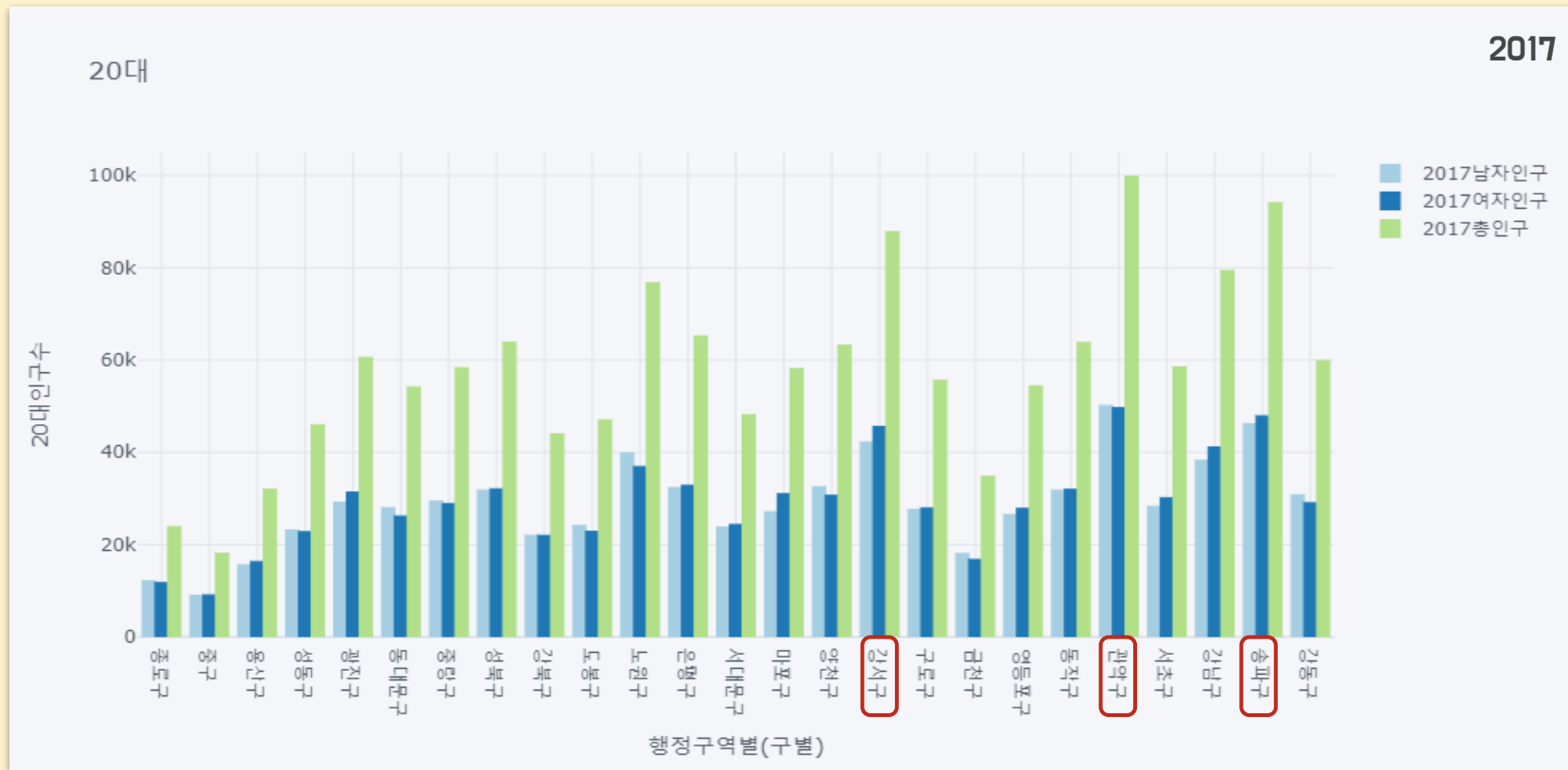


1. 주민등록인구 (내국인)

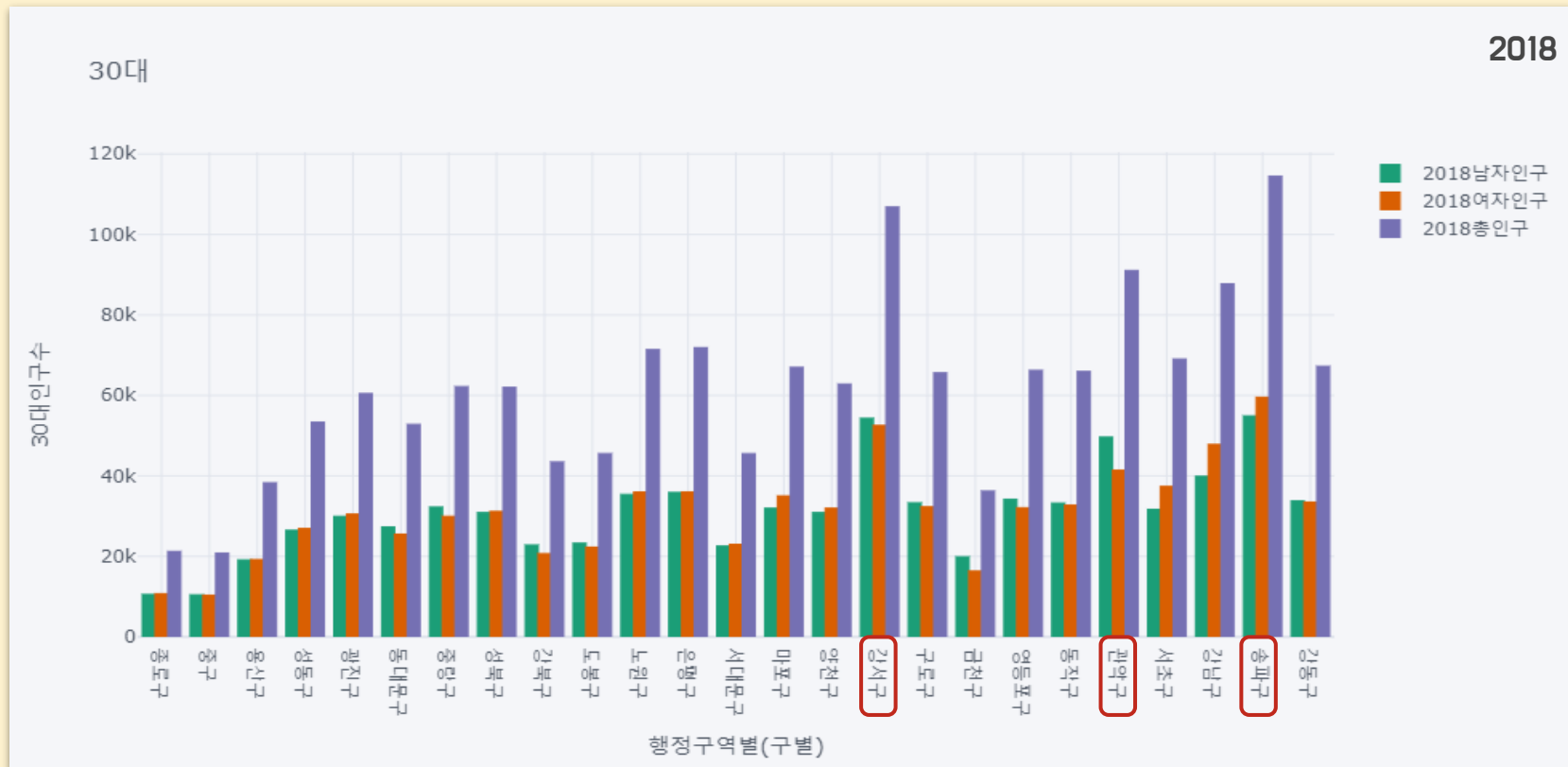
2017



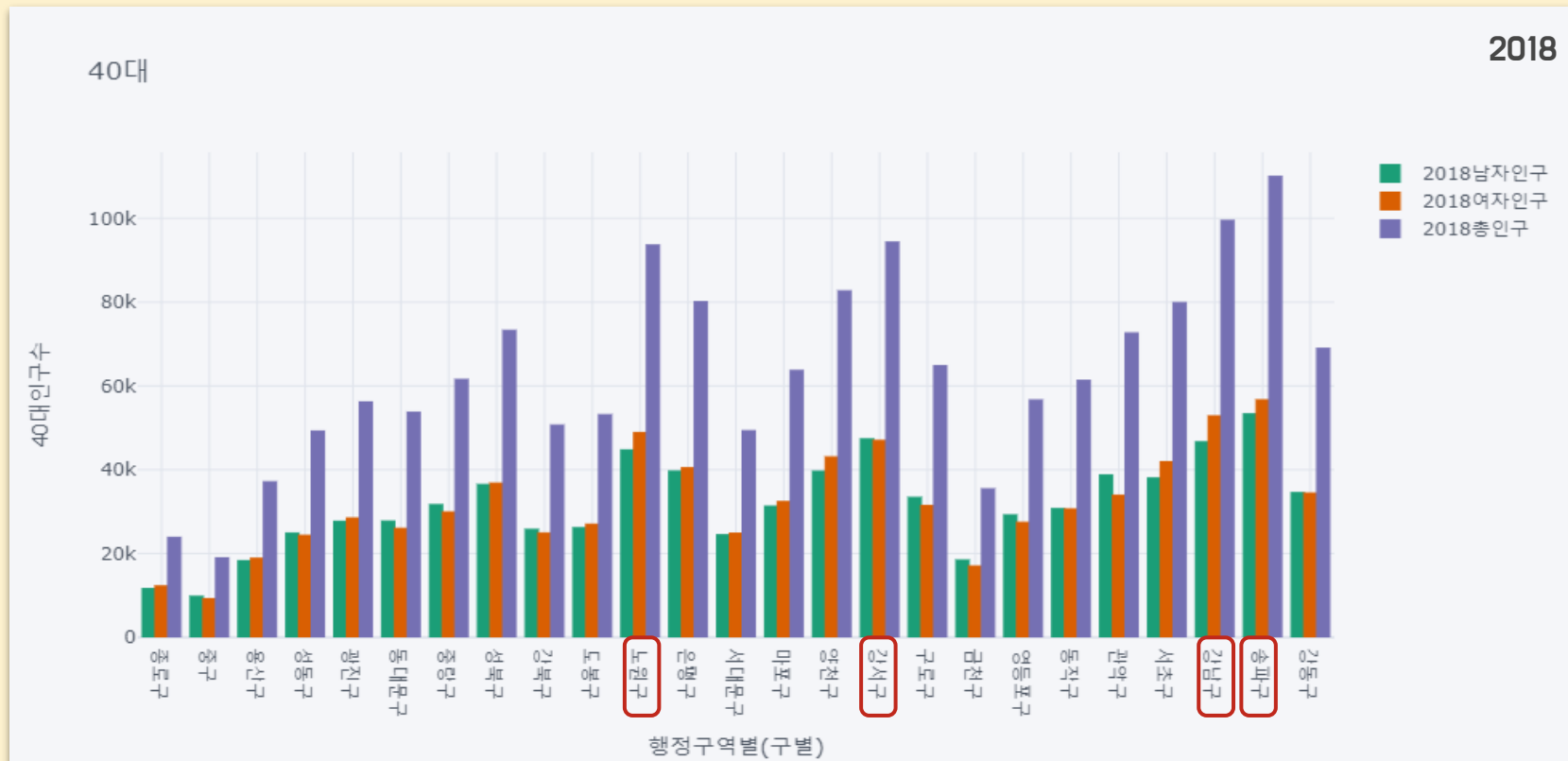
1. 주민등록인구 (내국인)



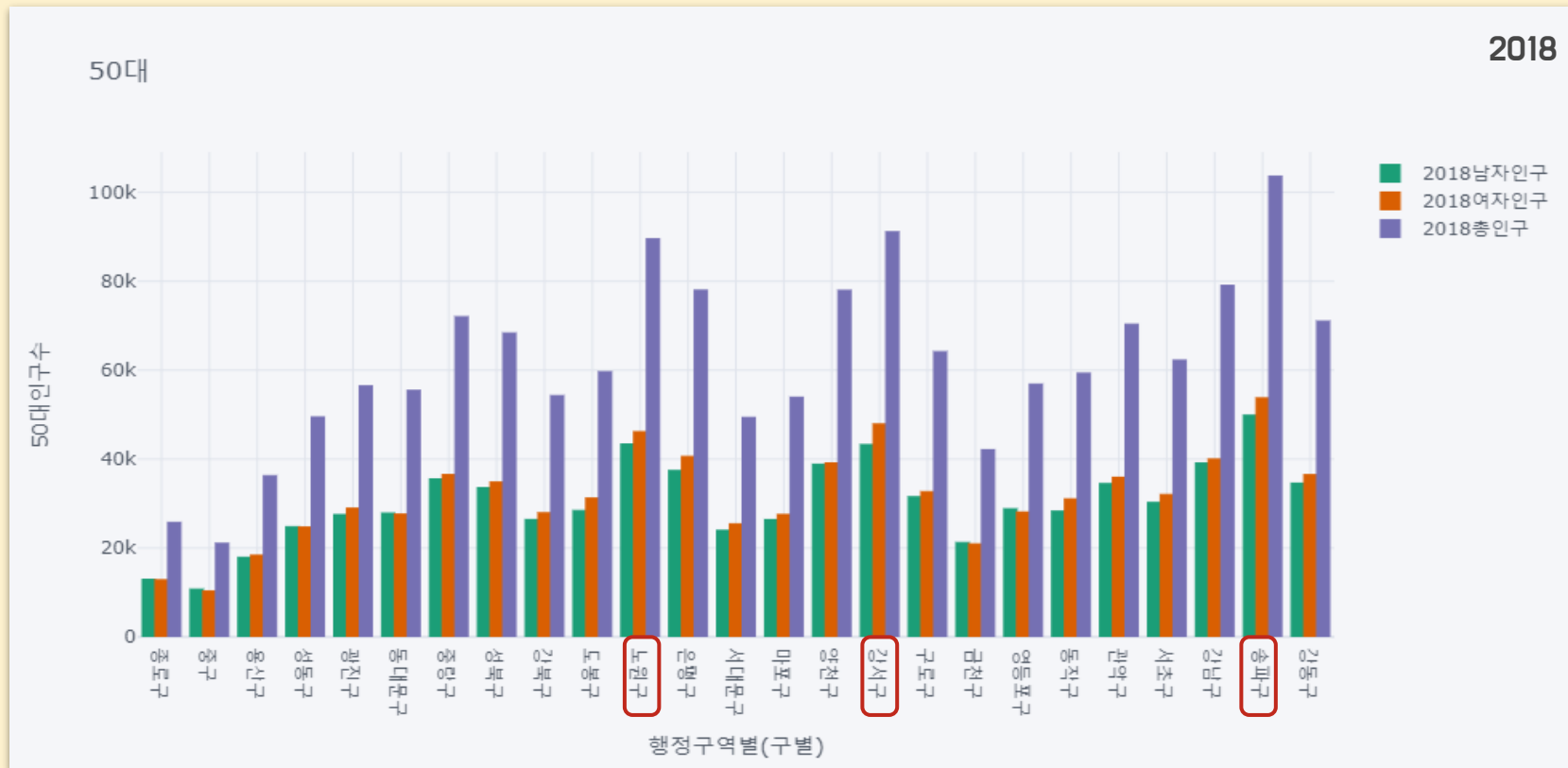
1. 주민등록인구 (내국인)



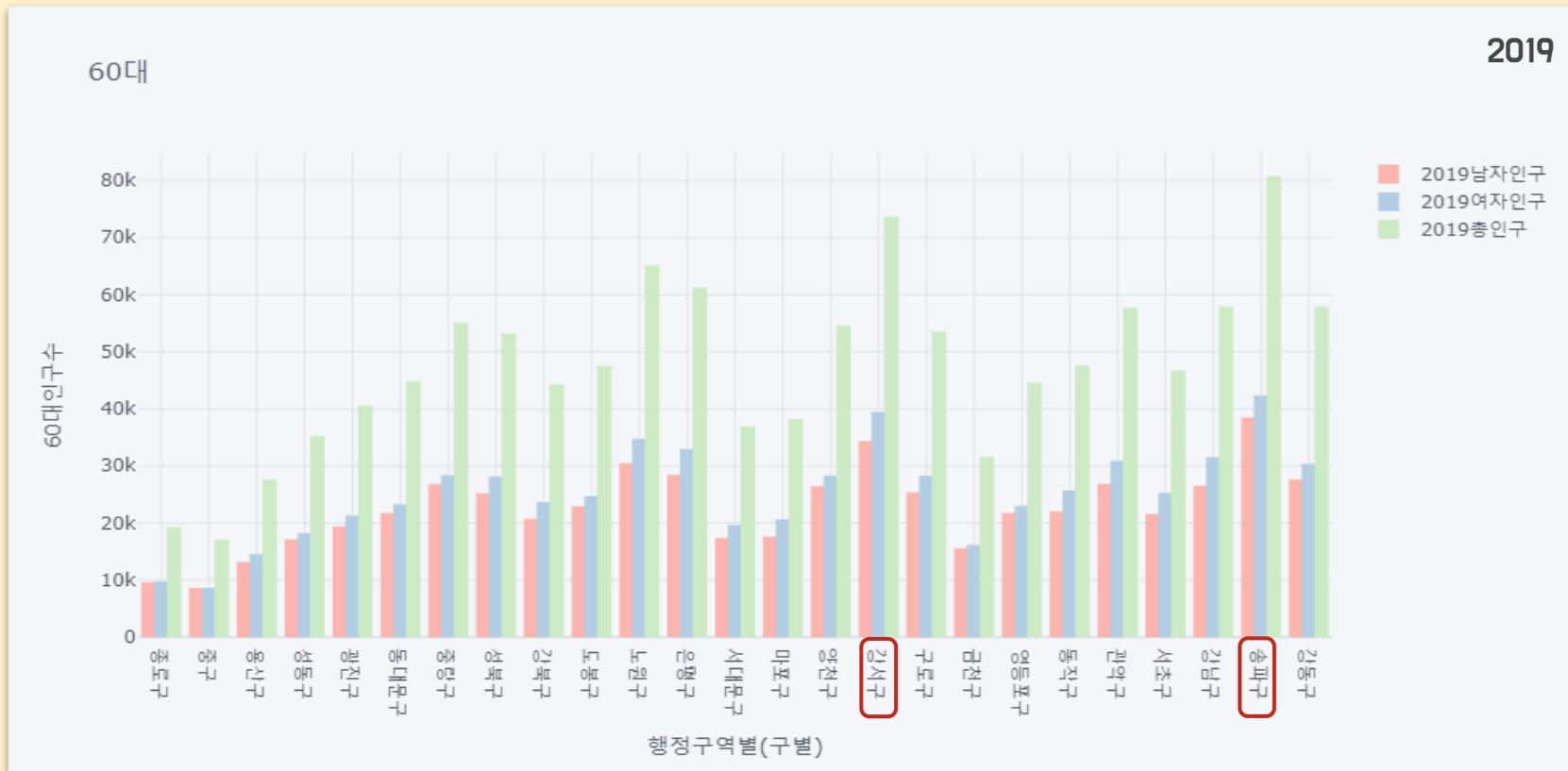
1. 주민등록인구 (내국인)



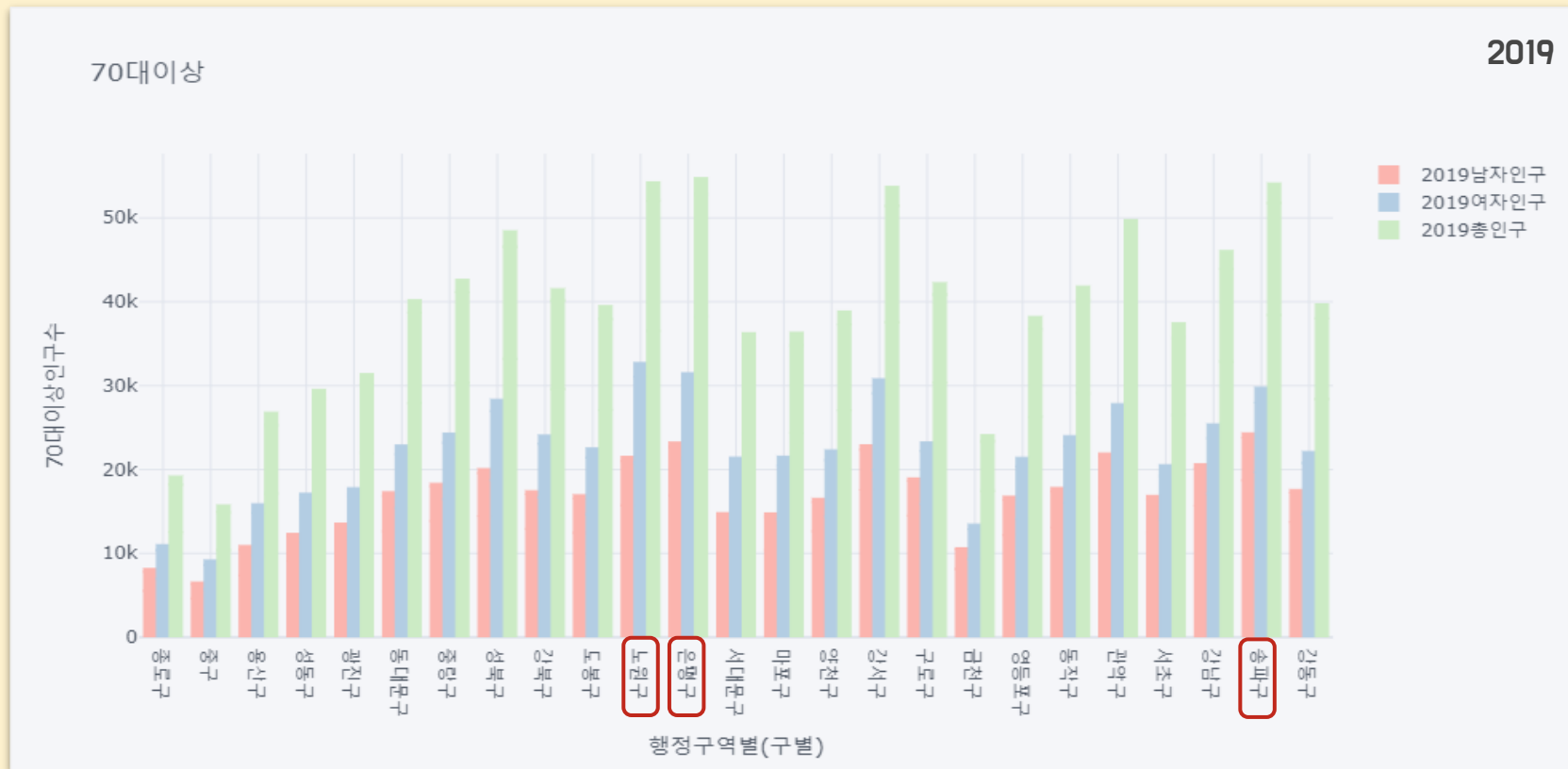
1. 주민등록인구 (내국인)



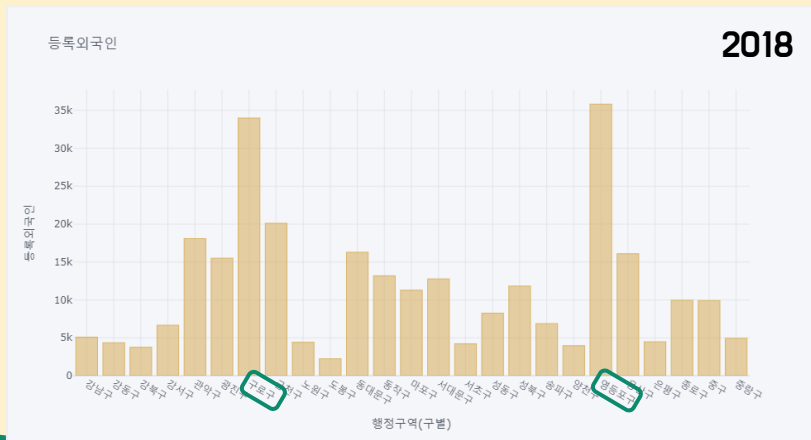
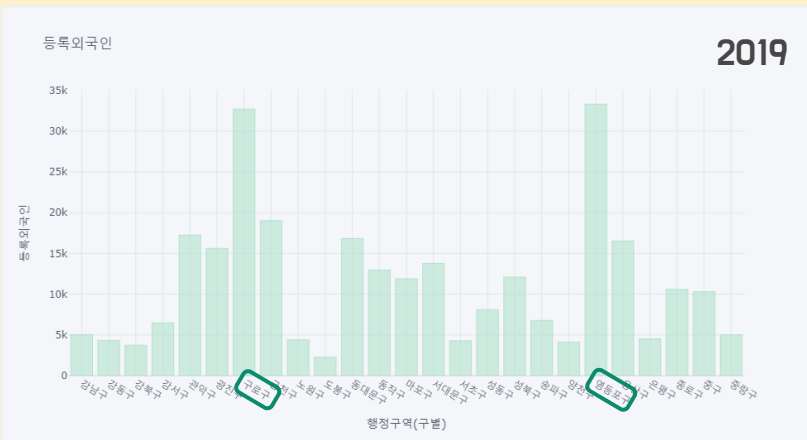
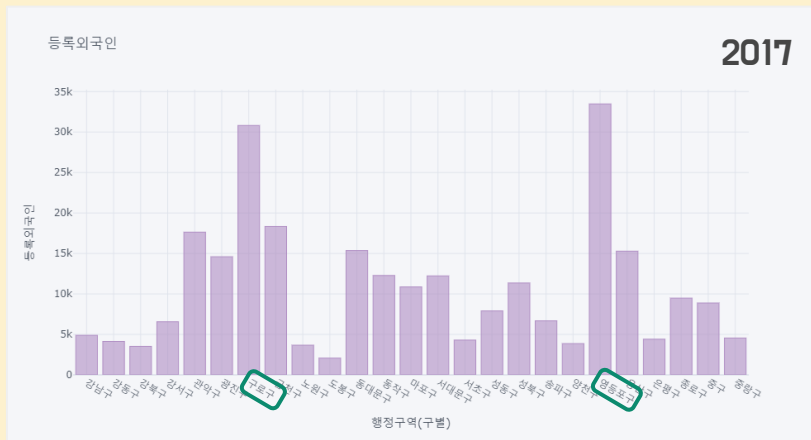
1. 주민등록인구 (내국인)



1. 주민등록인구 (내국인)



1. 주민등록인구 (외국인)



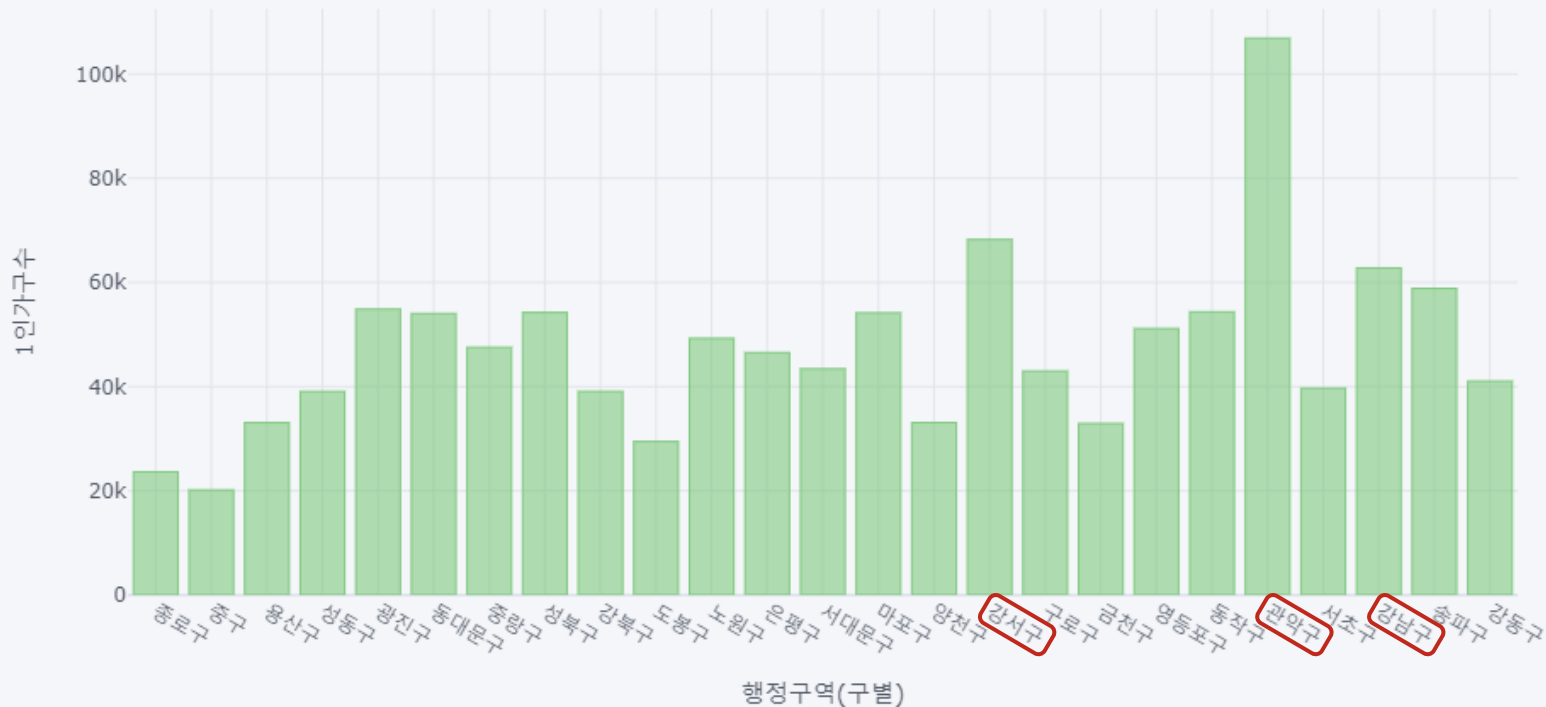
👉 외국인 수는
큰 변화없이
유지

2. 가구유형

1인가구수

👉 2017/2018/2019 구별 3개년 데이터
 👉 1인 가구수 ~ 7인 이상 가구 수

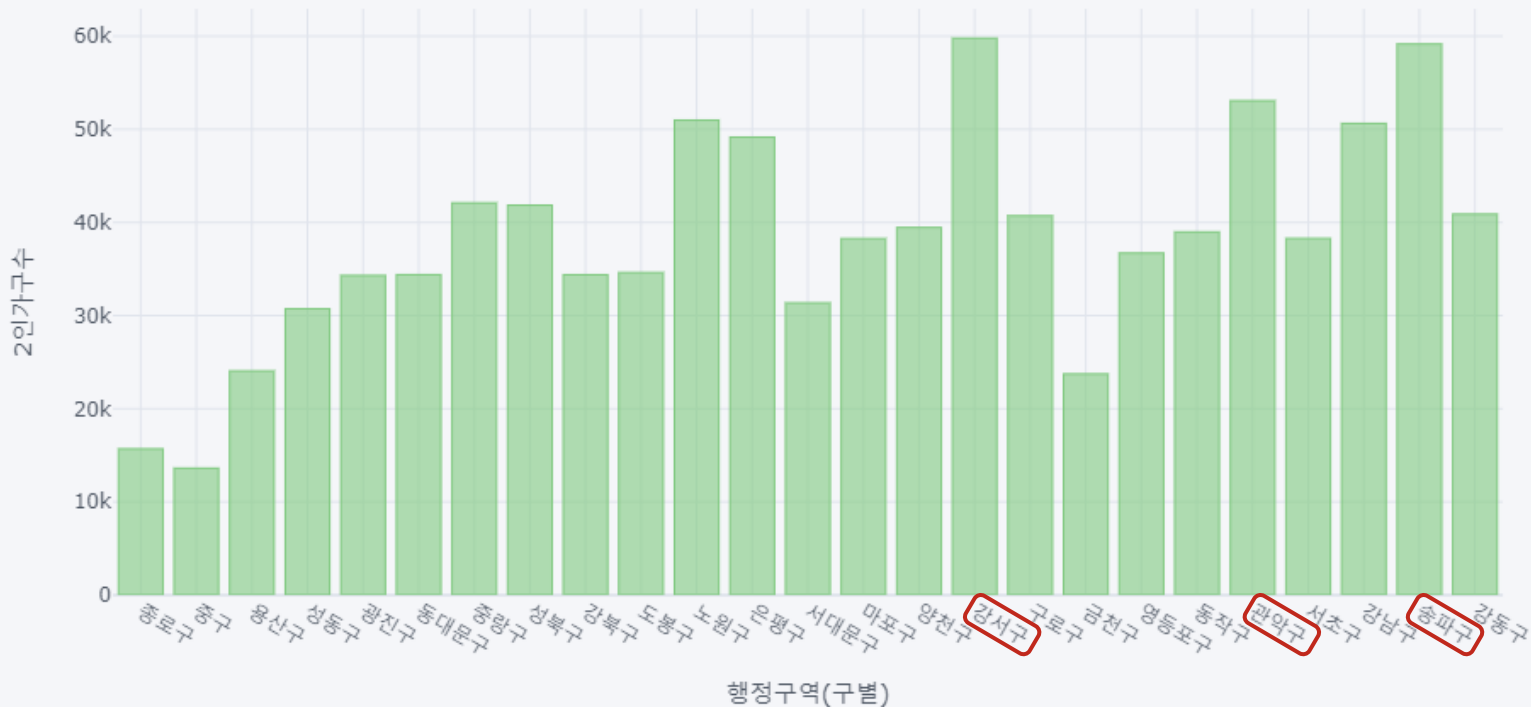
2017



2. 가구유형

2인가구수

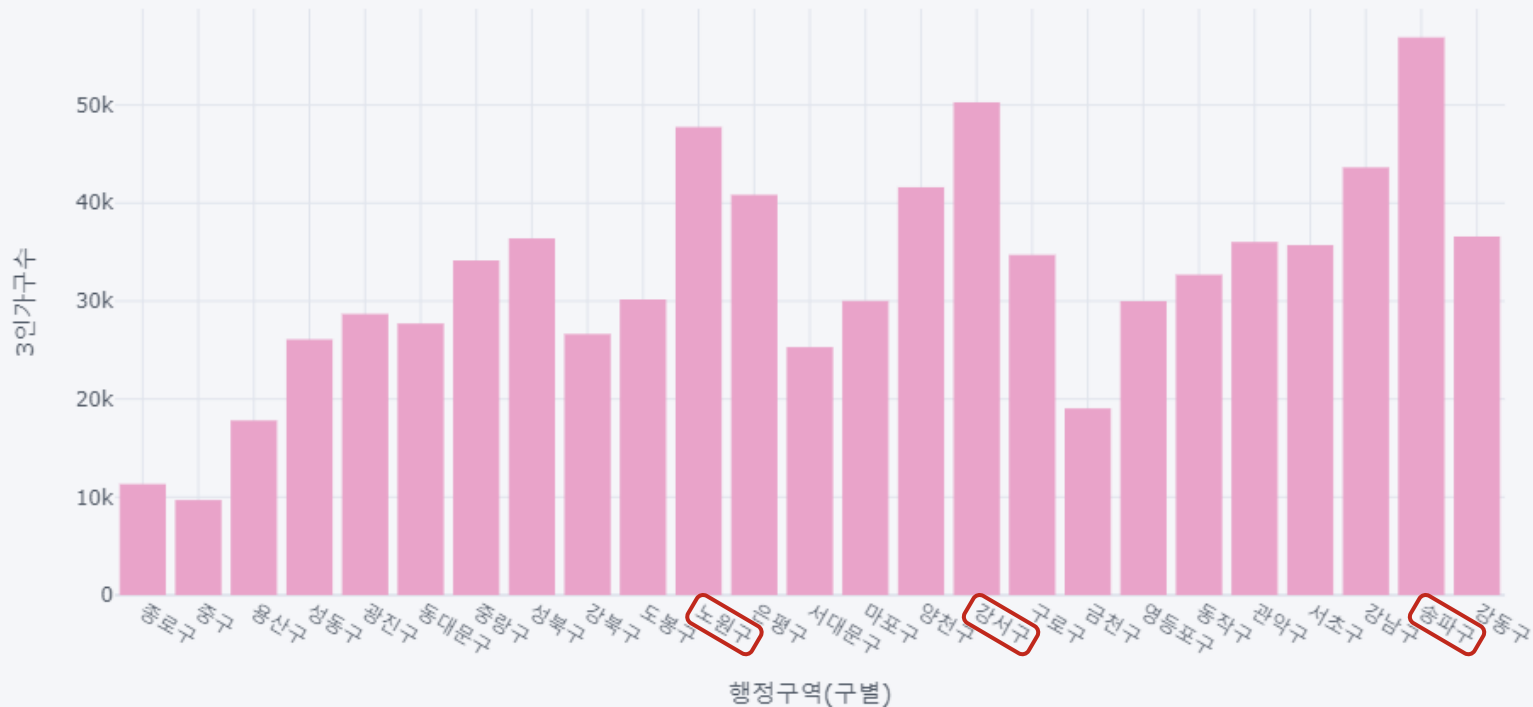
2017



2. 가구유형

3인가구수

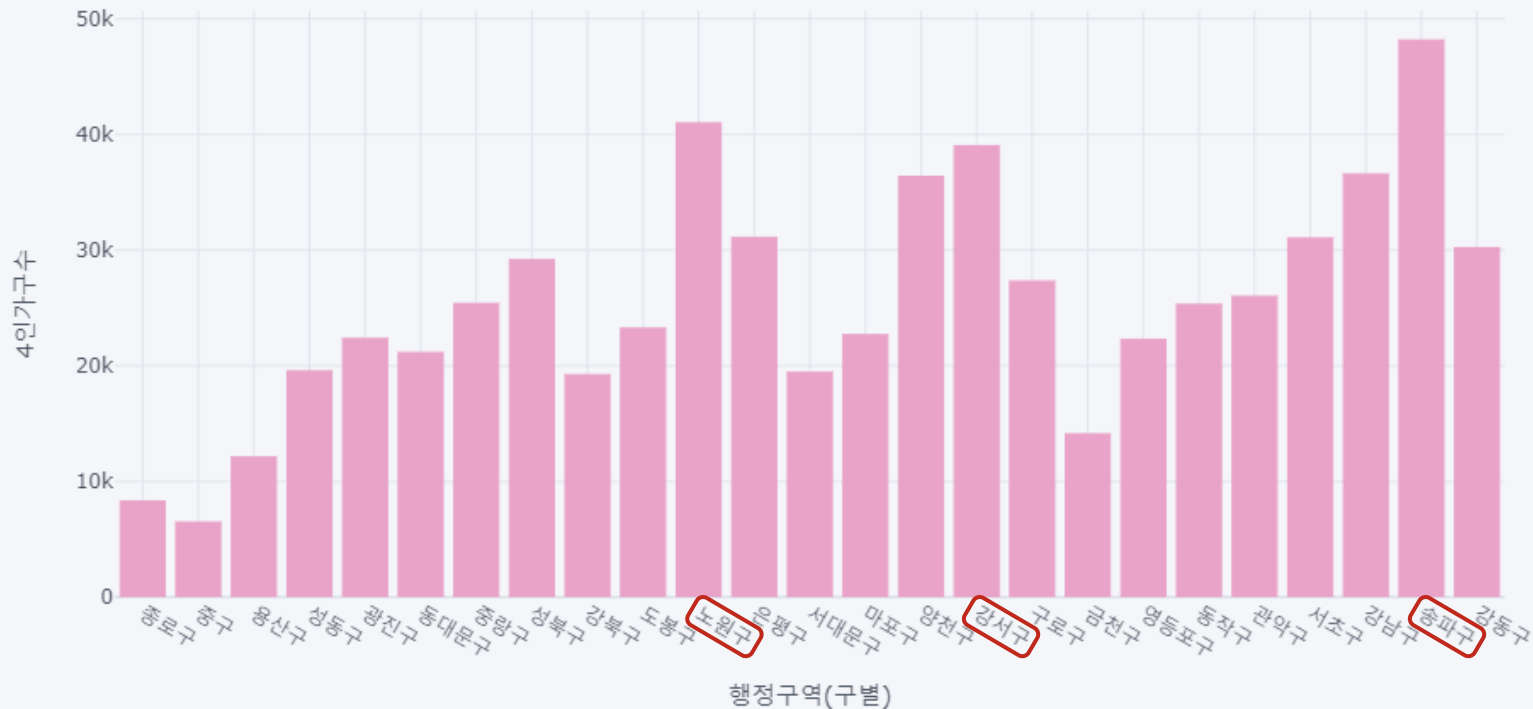
2018



2. 가구유형

2018

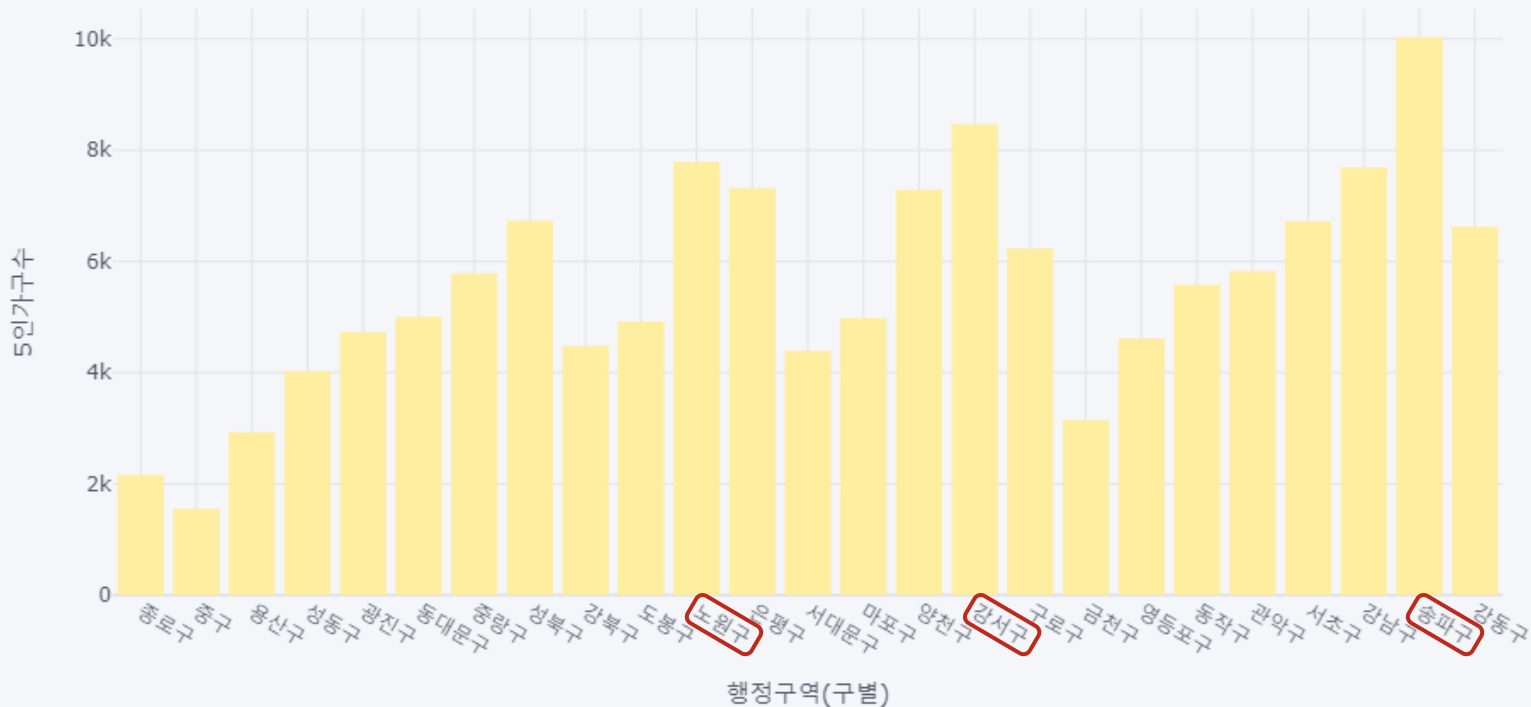
4인가구수



2. 가구유형

5인가구수

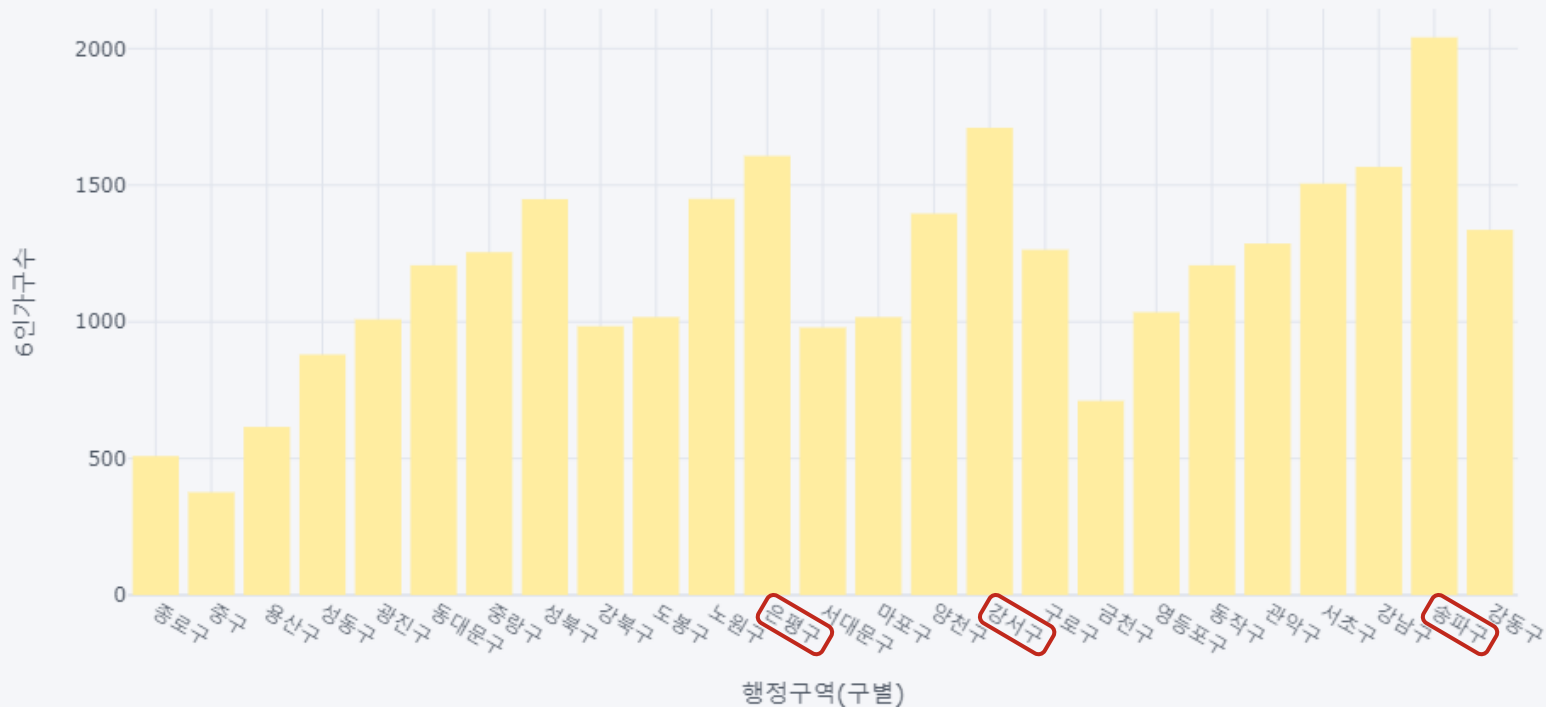
2019



2. 가구유형

6인가구수

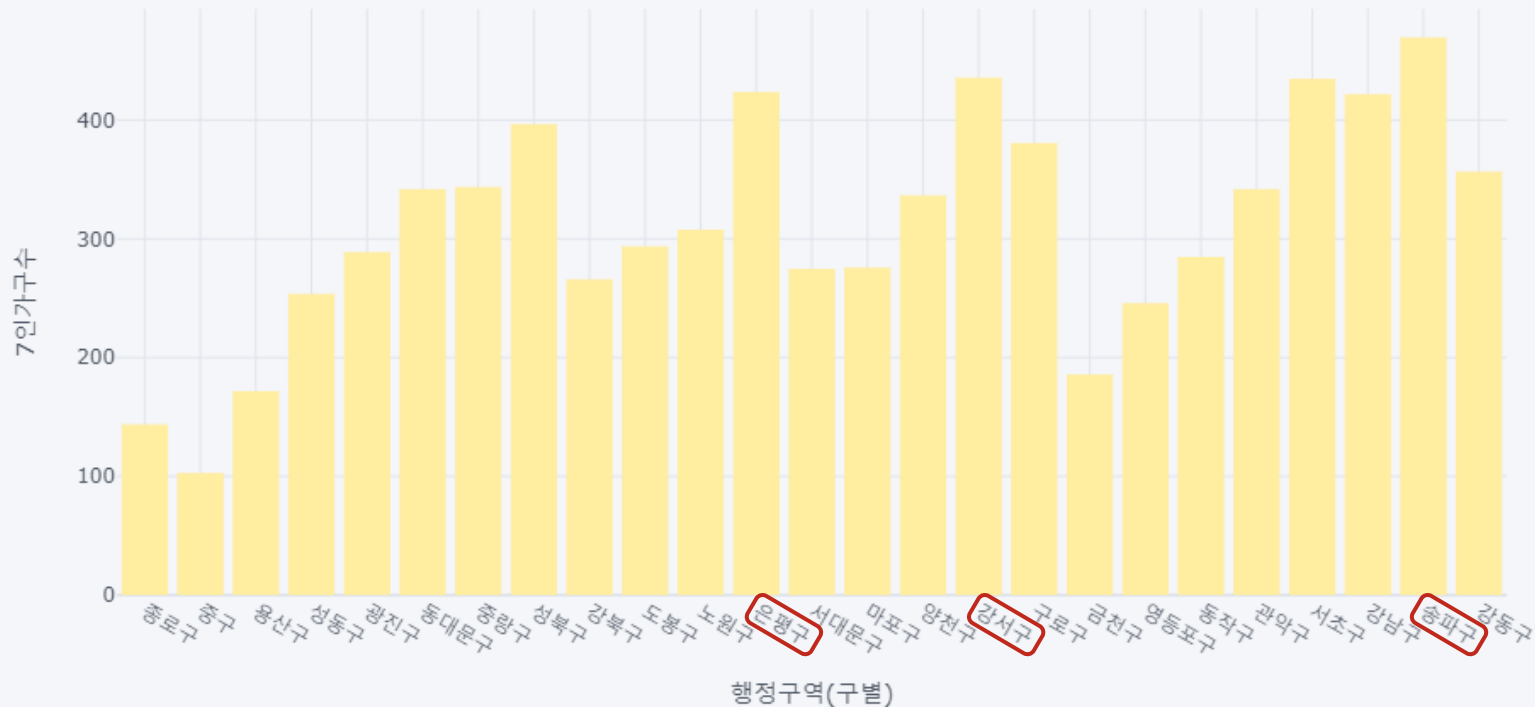
2019



2. 가구유형

7인가구수

2019



3. 생활인구

주간인구수09_18



유동인구가 생활인구로 대체되며 2018.04 이후부터 데이터 존재

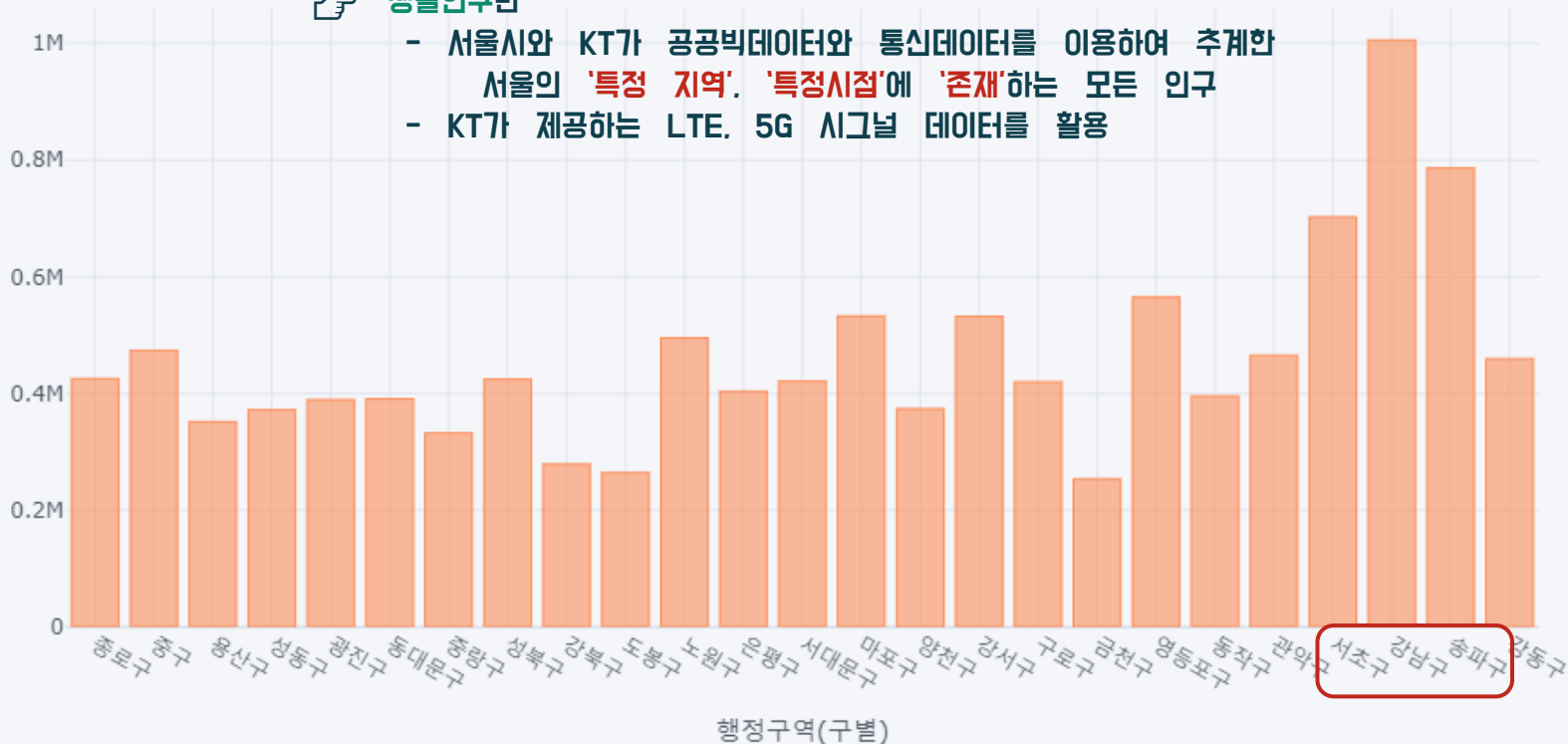
2018



생활인구란

- 서울시와 KT가 공공빅데이터와 통신데이터를 이용하여 추계한 서울의 '특정 지역', '특정시점'에 '존재'하는 모든 인구
- KT가 제공하는 LTE, 5G 시그널 데이터를 활용

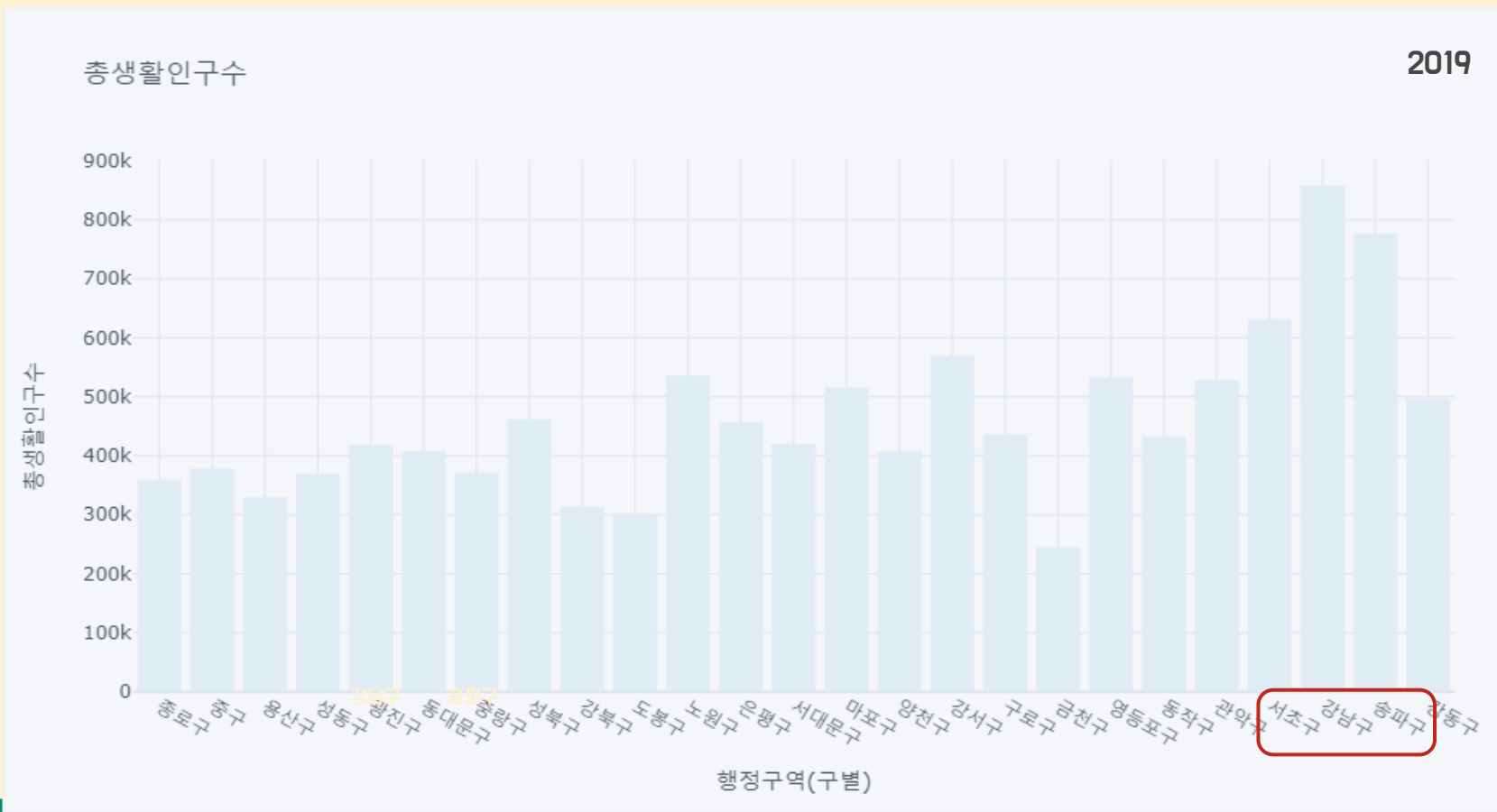
주간인구수09_18



DATA VISUALIZATION

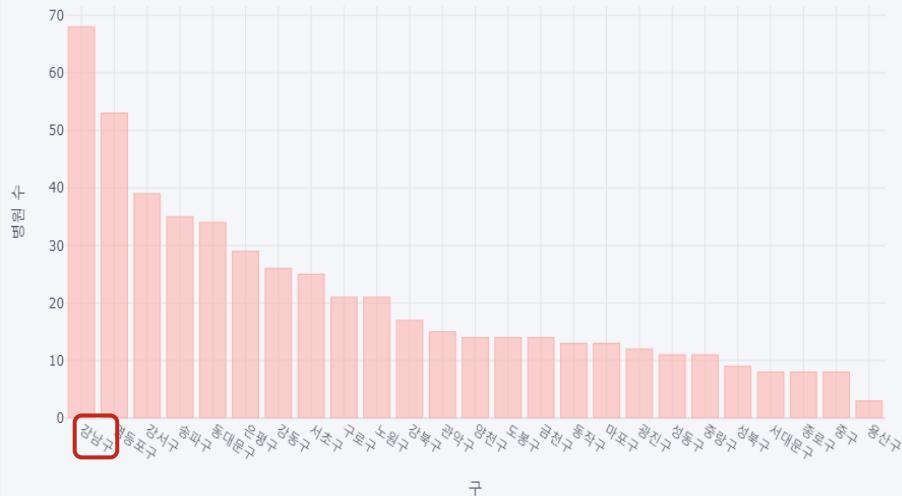


3. 생활인구



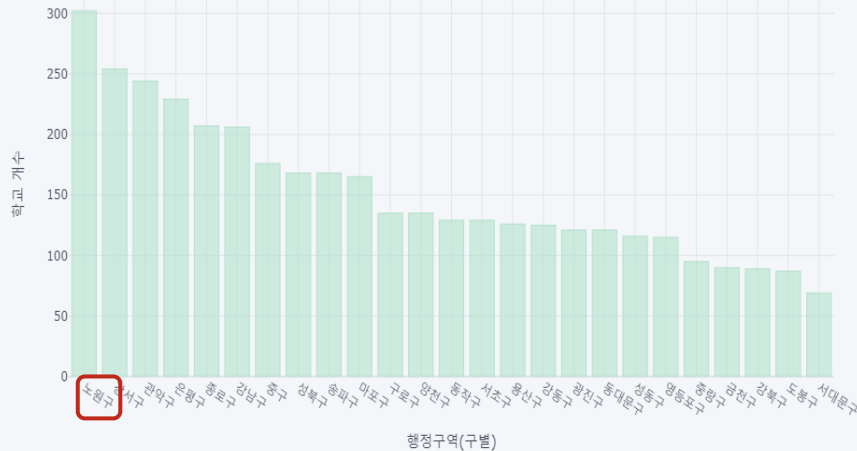
4. 병원 / 학교 인프라

서울시 구별 병원 수 현황



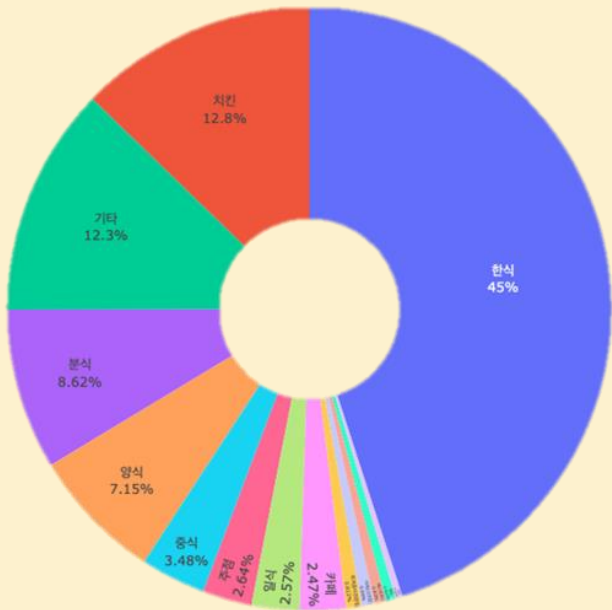
- 👉 강남구 병원 多
- 👉 용산구가 가장 적음

서울시 구별 학교 수 현황



- 👉 노원구 학교 多
- 👉 2019년 이후 개교한 학교 데이터는 제외함

5. 서울시 업태별 음식점 현황



- 한식
- 치킨
- 기타
- 분식
- 양식
- 중식
- 주점
- 일식
- 카페
- 외국음식전문점
- 식육(고기집)
- 패스트푸드
- 횃집
- 뷔페



서울시 전체

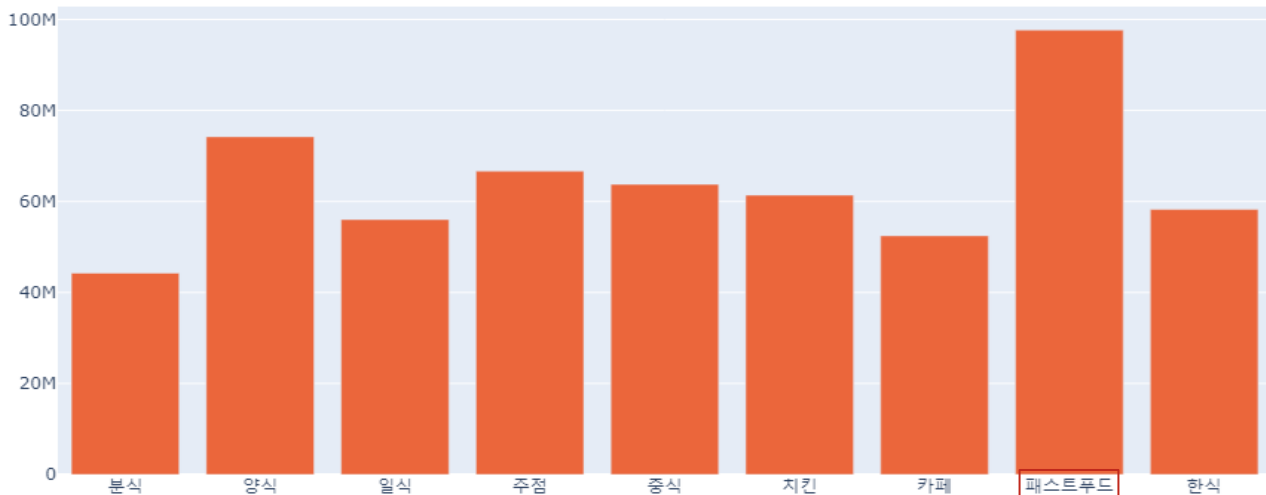
업태별 음식점 개수를
백분위로 표현

6. 구별/업태별 추정매출액

👉 모든 업태의 총 평균매출 순위 : 1) 서초구 2) 강남구 3) 용산구

서초구에서는 패스트푸드 업종이 강세

서초구 업종별 평균매출액

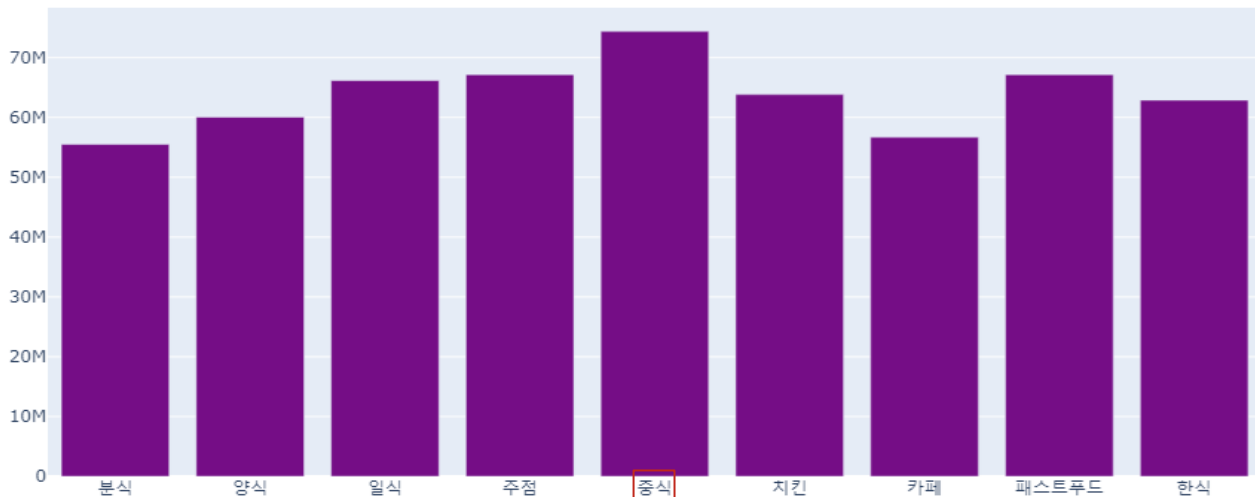


6. 구별/업태별 추정매출액



강남구에서는 중식이 제일 강세인 업종

강남구 업종별 평균매출액

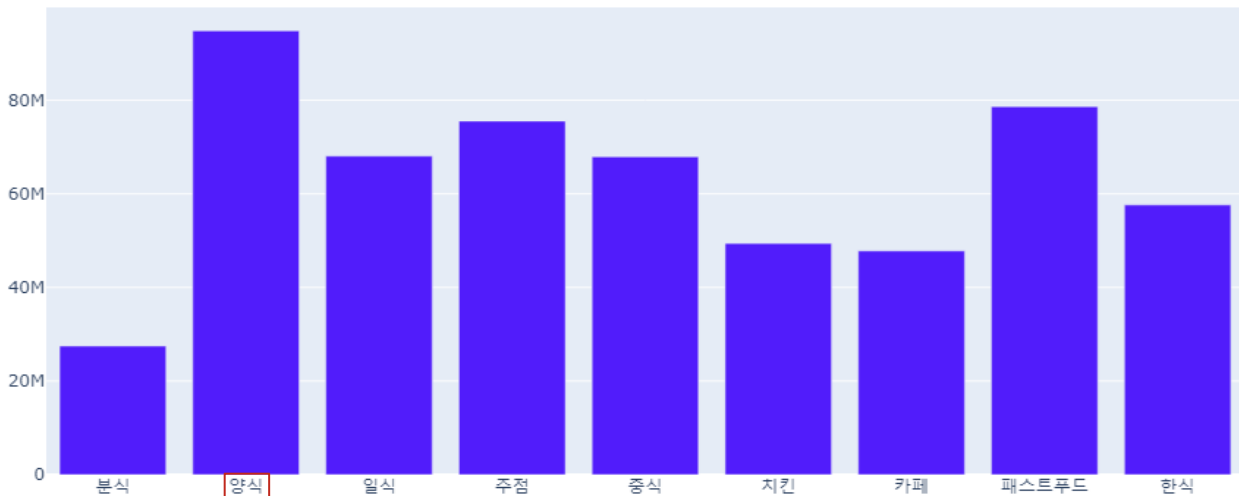


6. 구별/업태별 추정매출액



이태원 등 이국적인 식당들이 많은 용산구는 양식이 강세

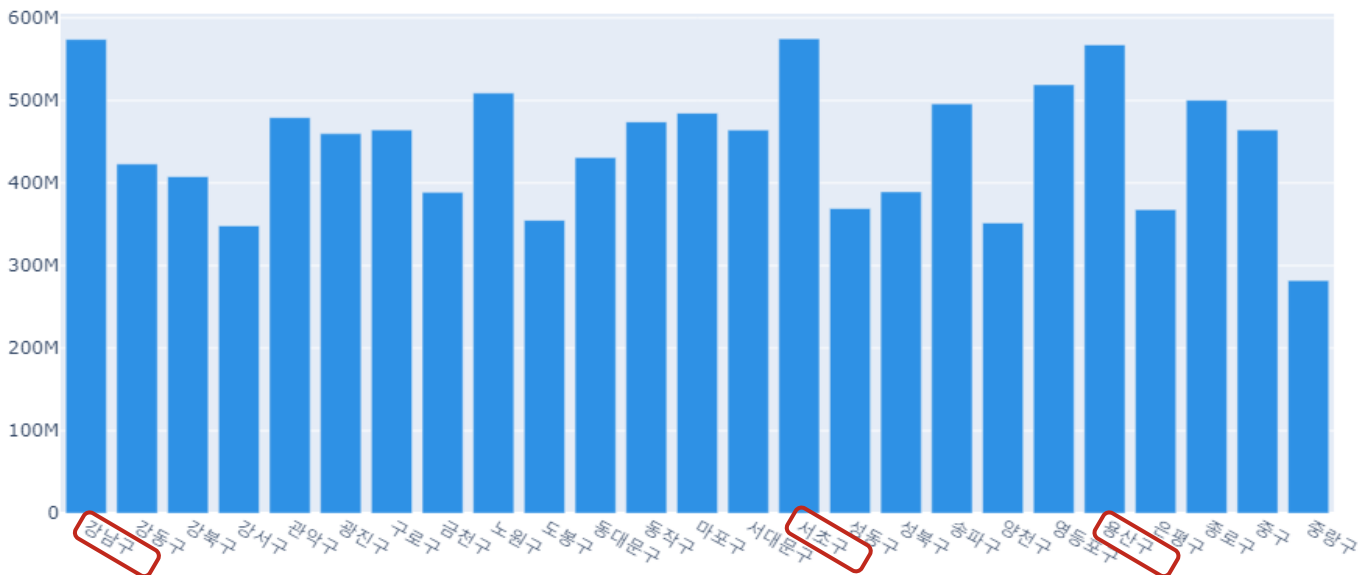
용산구 업종별 평균매출액

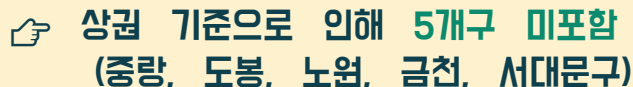
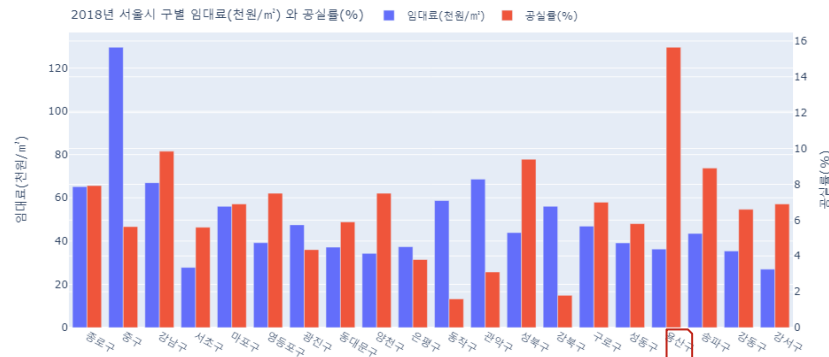


6. 구별/업태별 추정매출액

👉 모든 업태(총 9종)의 평균매출액을 더한 서울시 구별 추정매출 순위

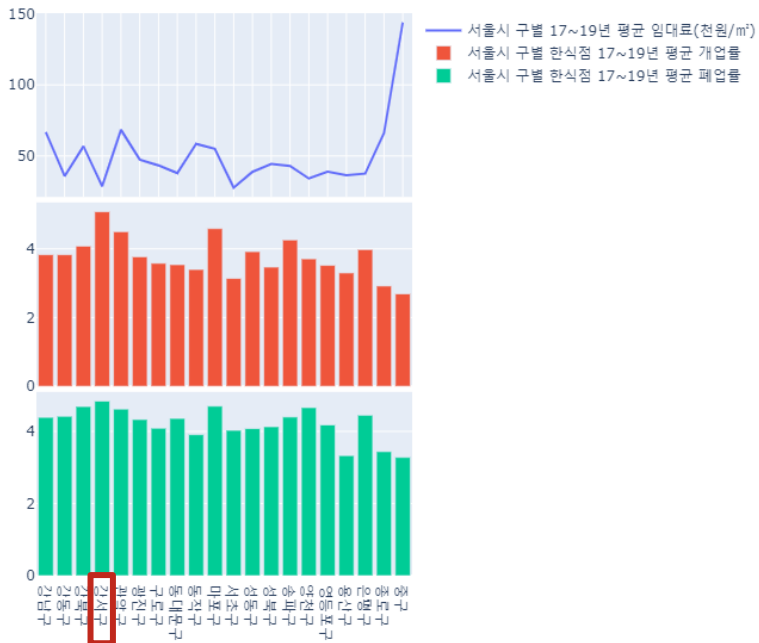
서울시 구별 총평균매출액(9개 업종 평균매출 총합)



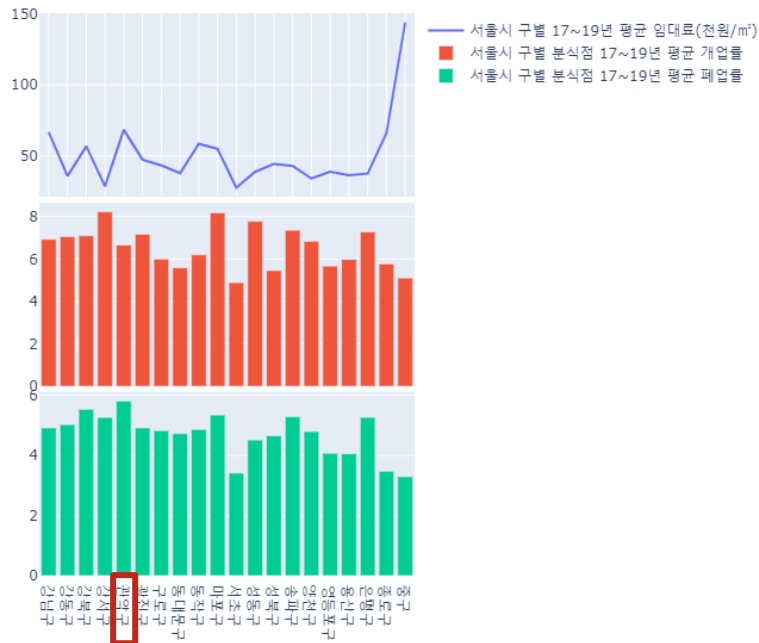


a) 한식점 • 분식점

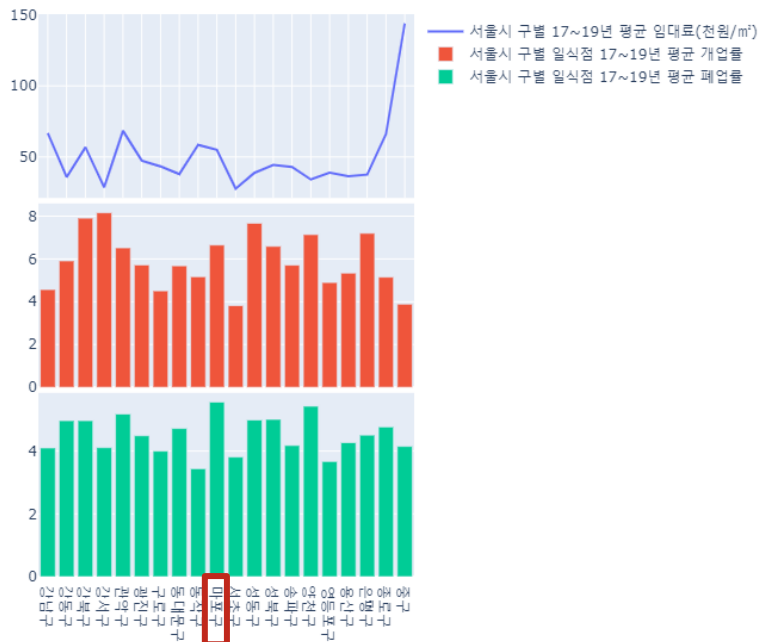
17~19년 서울시 한식점 평균 임대료(천원/㎡)와 평균 개·폐업을 비교



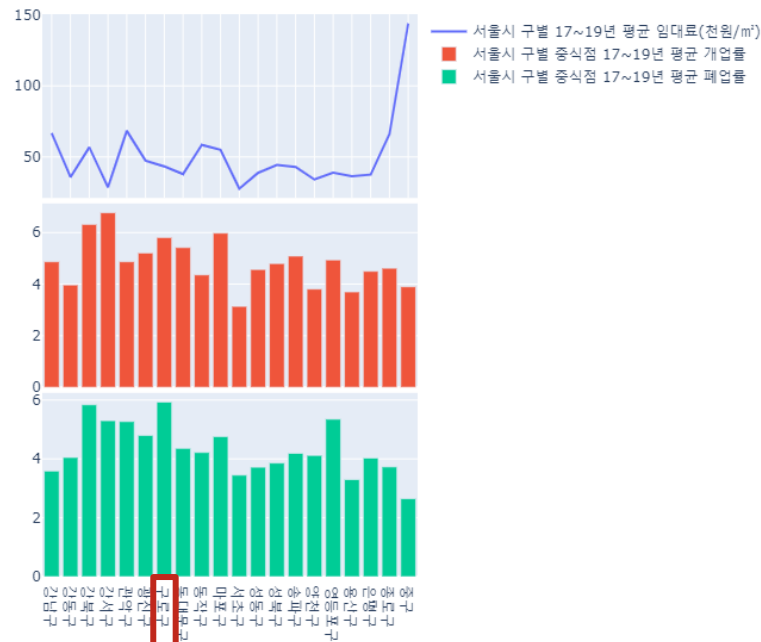
17~19년 서울시 분식점 평균 임대료(천원/㎡)와 평균 개·폐업을 비교



b) 일식점 • 중식점

17~19년 서울시 일식점 평균 임대료(천원/m²)와 평균 개·폐업을 비교

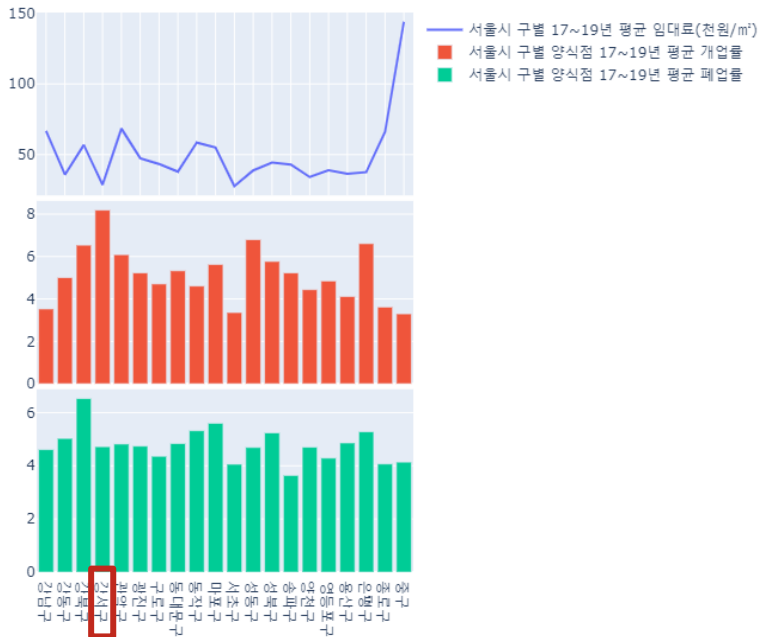
17~19년 서울시 중식점 평균 임대료(천원/㎡)와 평균 개·폐업을 비교



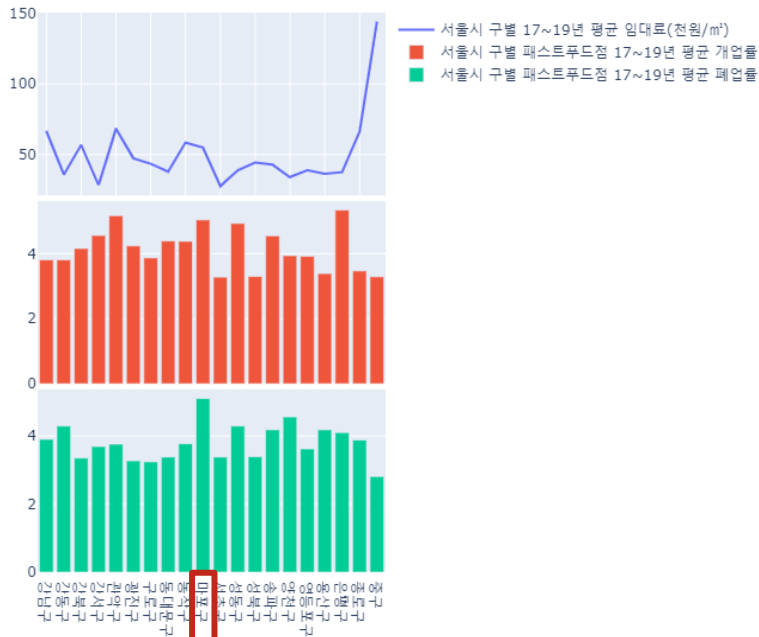
8. 서울시 업태별 평균 임대료와 평균 개폐업률 비교

c) 양식점 • 패스트푸드점

17~19년 서울시 양식점 평균 임대료(천원/m²)와 평균 개·폐업률 비교



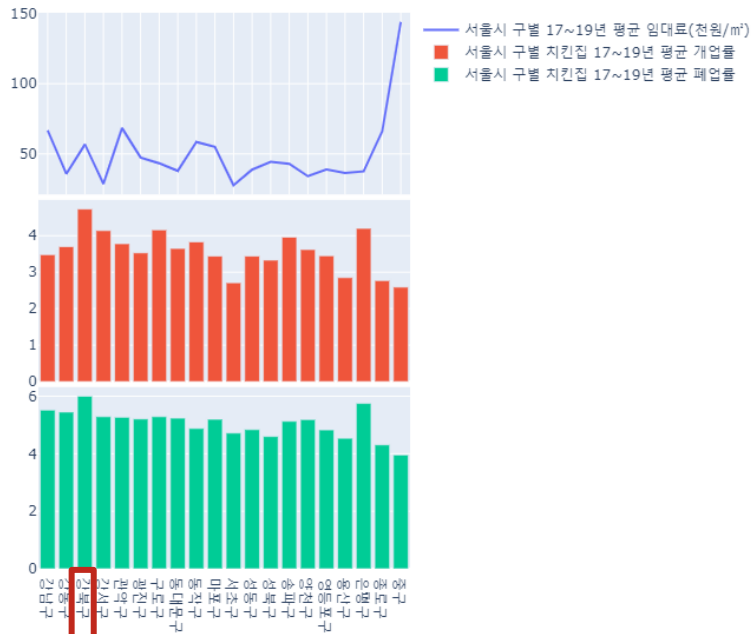
17~19년 서울시 패스트푸드점 평균 임대료(천원/m²)와 평균 개·폐업률 비교



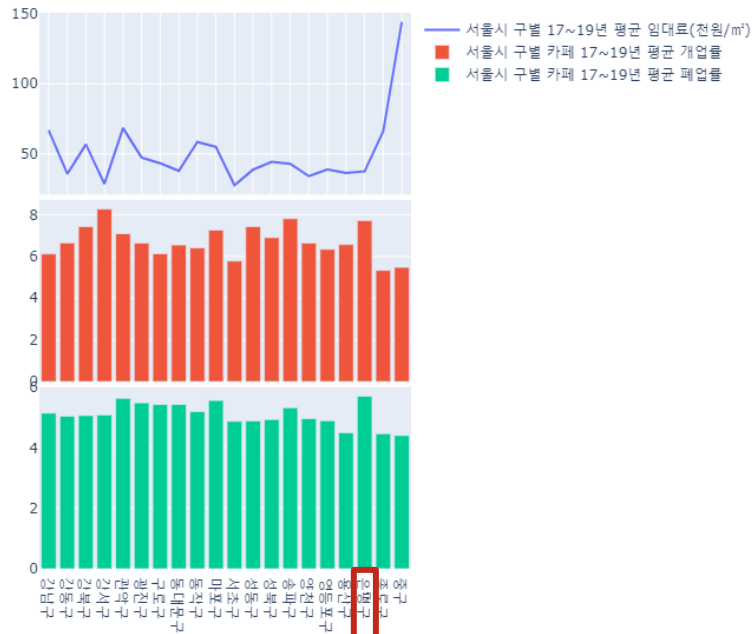
8. 서울시 업태별 평균 임대료와 평균 개폐업률 비교

d) 치킨집 • 카페

17~19년 서울시 치킨집 평균 임대료(천원/㎡)와 평균 개•폐업률 비교



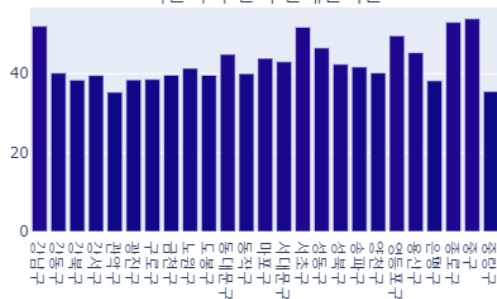
17~19년 서울시 카페 평균 임대료(천원/㎡)와 평균 개•폐업률 비교



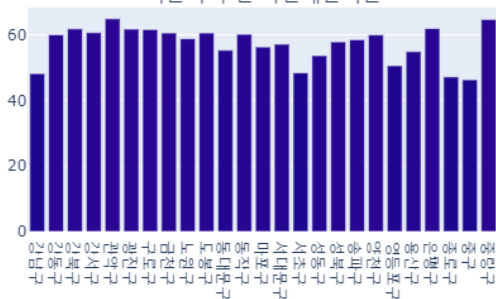
10. 서울생활인구와 매출 비교(주/야간)

2018년도

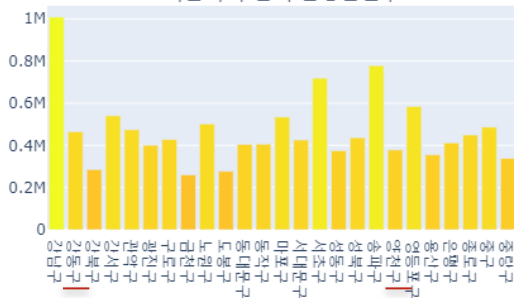
서울시 구별 주간매출비율



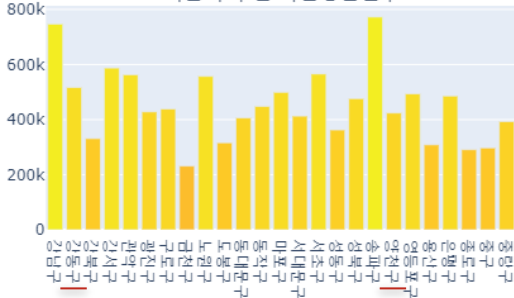
서울시 구별 야간매출비율



서울시 구별 주간생활인구



서울시 구별 야간생활인구



👉 생활인구가 2018년 데이터부터 존재하기 때문에 2개년만 비교

👉 한계

→ 데이터간 주/야간 시간대가 완전히 일치하지 않음

생활인구 주간 : 09시 ~ 18시

생활인구 야간 : 19시 ~ 08시

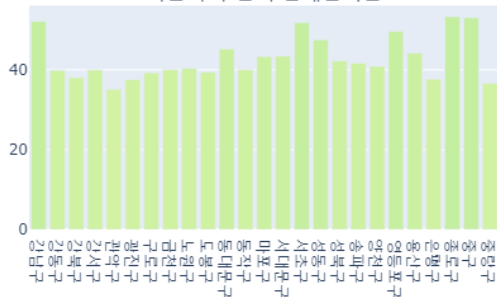
매출비율 주간 : 06시 ~ 17시

매출비율 야간 : 17시 ~ 06시

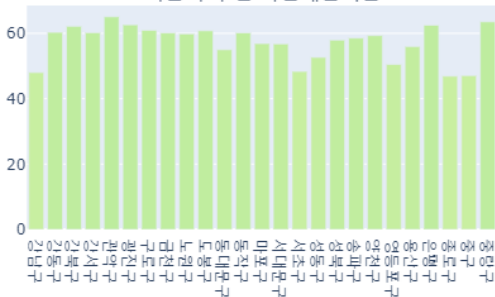
10. 서울생활인구와 매출 비교(주/야간)

2019년도

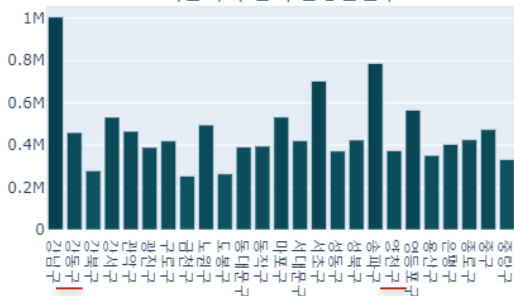
서울시 구별 주간매출비율



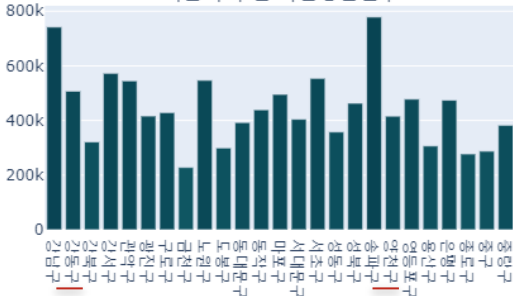
서울시 구별 야간매출비율



서울시 구별 주간생활인구



서울시 구별 야간생활인구



👉 생활인구가 2018년 데이터부터 존재하기 때문에 2개년만 비교

👉 한계

→ 데이터간 주/야간 시간대가 완전히 일치하지 않음

생활인구 주간 : 09시 ~ 18시

생활인구 야간 : 19시 ~ 08시

매출비율 주간 : 06시 ~ 17시

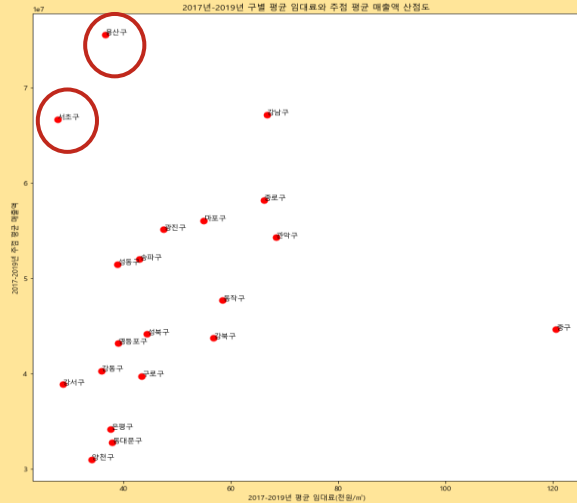
매출비율 야간 : 17시 ~ 06시



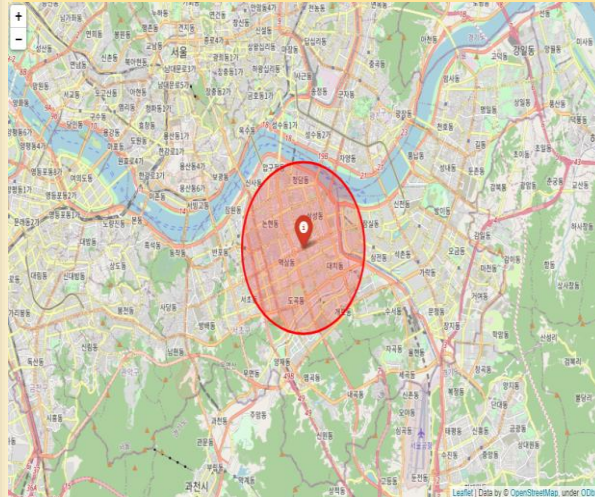
초대규모 자본으로 주점을 창업하고 싶은 루피에게 최적 위치를 추천한다면?



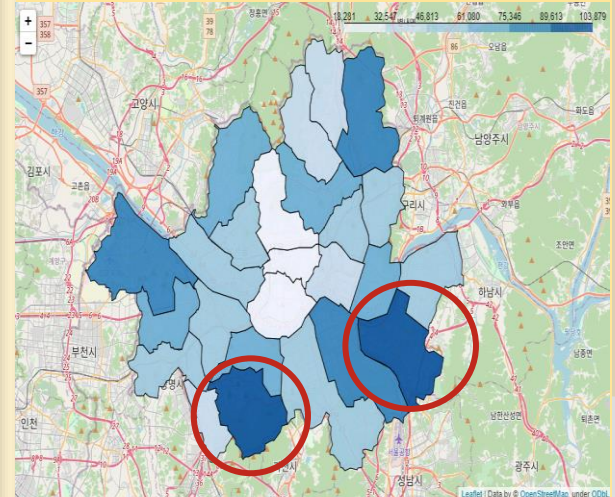
임대로 대비 매출액이 타지역에 비해 높은 '용산구', '서초구' 선정



주점의 수가 적은 '강남구' 👉 경쟁력 UP

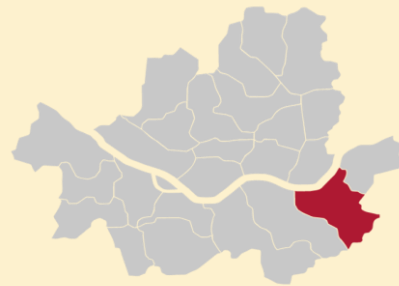


야간 인구, 20대가 많은 '관악구' 와 '송파구' 👉 최대 주점 업종 소비자 확보





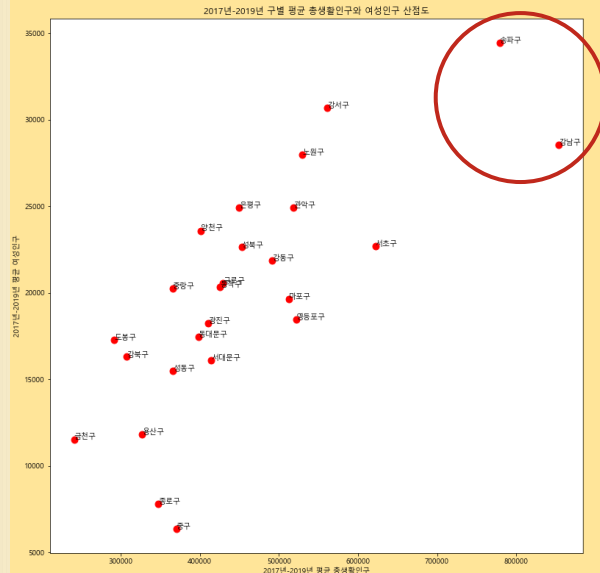
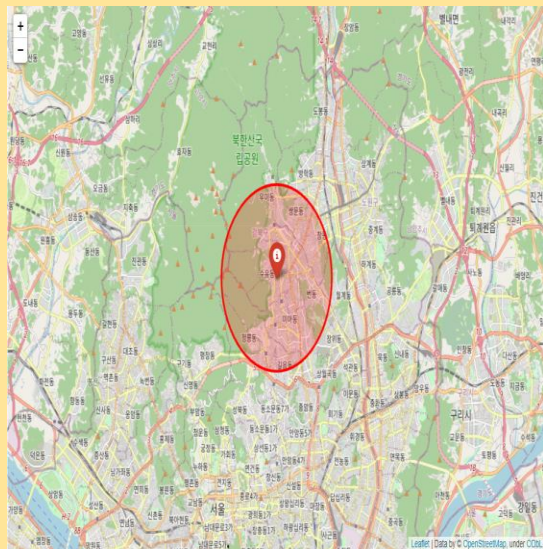
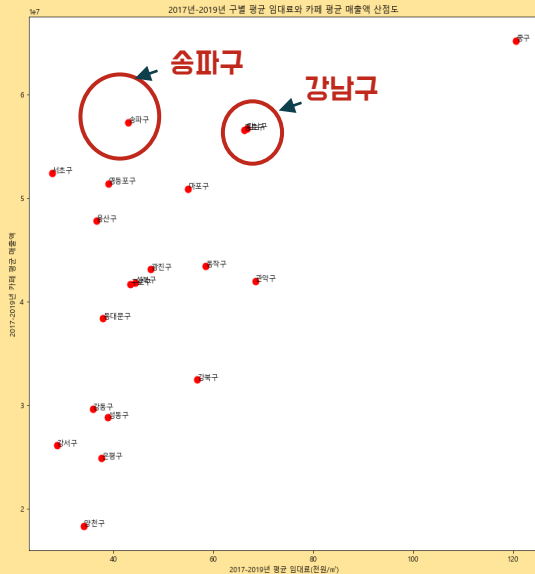
대규모 자본으로 카페를 창업하고 싶은
보로로에게 최적 위치를 추천한다면?



매출액이 높고 임대료가 중간인
'송파구' 와 둘 다 높은 '강남구' 선정

카페 밀집도가 가장 낮은 '강북구'

유동인구, 여성인구가 많은
'강남구' 와 '송파구'

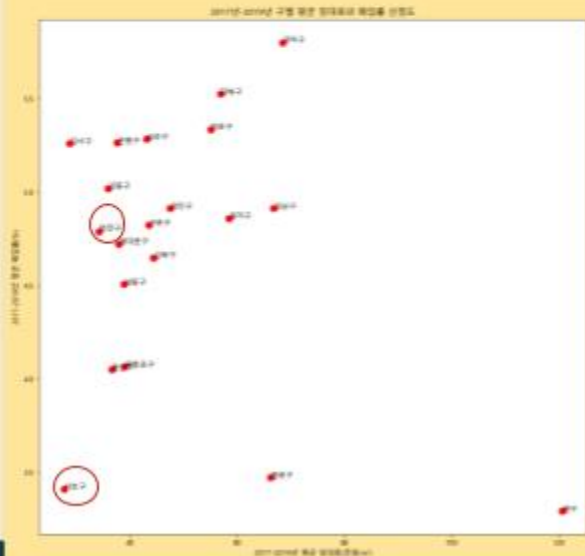




소규모 자본으로 분식집을 창업하고 싶은
패티에게 최적 위치를 추천한다면?



임대료와 폐업률이 가장 낮은 구와
'서초구' 와 상대적으로 낮은 '용산구' 선정



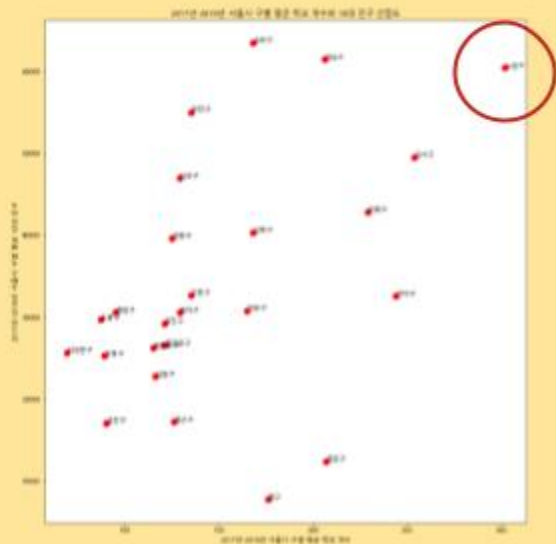
분식집의 수가 적은 '광진구'

경쟁력 UP



학교와 10대가 가장 많은

'노원구'



04

결론 및 향후과제



IMPRESSIONS AFTER PROJECT

👉 비슷한 데이터베이스를 가진 서울시의 '우리마을가게 상권분석 서비스'와의 차별성 모색 필요

👉 SNS(인스타그램, 트위터, 네이버 블로그 등)의 동적 크롤링을 통해 실시간 맛집 트렌드 데이터 추가

👉 Django(장고)를 통한 웹서비스 구현

👉 임대료와 추정매출 데이터가 상권 기준 → 데이터간 표본의 통일성 부족

👉 동별 DB가 아닌 구별 DB 구축으로 창업 지역 추천시 세부적인 장소 선정의 한계

05

느낌점



IMPRESSIONS AFTER PROJECT

👉 데이터 수집 및 가공

리눅스 OS, 우분투, 하둡 및 스파크에 대한 이해가 완전하지 않은 상태여서, 해당 프로그램을 데이터 수집과 전처리에 활용하기 어렵다고 느꼈습니다.
그러나 우분투에서 스파크와 하둡을 여러 번 연결 해 보고, 다양한 종류의 오류를 해결하다 보니 점차 친숙해지고 익숙하게 다룰 수 있게 되었습니다.

👉 팀 내 협업

워크플로우에 대한 이해가 서로 달라서 정해진 기간까지 각자에게 할당된 산출물이 나오지 않아서 힘들었습니다. 기본적인 약속과 계획을 지키지 않는 경우가 잦아 사전에 협의한 워크플로우대로 진행할 수 없었습니다.
팀원으로서 전체를 통솔하거나 지휘할 수 없기에 산출물이 당장 필요한 경우 다른 팀원의 데이터 전처리 및 가공, 시각화를 맡아 처리하였습니다.

감사합니다:-)





QUESTION

