

빅데이터 분석 및 서비스 기획 부트캠프 자유 팀프로젝트

서울시 상권 공공데이터 EDA

: 코로나 전 후 서울시 상권에 미친 영향과 현황 분석

Date : 22.12.30

Team. 골목식당

강서현, 맹지주, 이우성

1

주제 선정

1. 주제 선정 배경
2. 수집데이터 소개

2

분석 프로세스 설계

1. 데이터 수집
2. 전처리
3. 분석
4. 시각화

3

데이터 분석 및 결과

1. 가설과 인사이트
2. 결과

주제 선정

리서치와 분석을 통한 주제 선정

: 코로나가 서울시 상권에 끼친 영향 분석


곧 마스크 해제가 진행되며 코로나 전후로 골목 상권에 미친 영향과 현황을 파악해 보기 위해 상권 분석을 주제로 선정


The JoongAng


구독


거리두기 강화 직격탄, 골목상권 매출 3분의 1 줄었다

입력 2020.12.17. 오전 12:04 · 수정 2020.12.17. 오전 12:06 · 기사원문


 손해용 기자

 34

 25









지난주 전국 소상공인 매출 분석

서울 38%나 줄어 타격 가장 커

노래방 -91%, 실내체육시설 -66%

연말 대목은커녕 장사 접어야할 판

구축/개발 환경	Python 3.7, PyCharm, Jupyter Notebook
기간	22. 12. 21(수) ~ 22. 12. 30(금)
참여인원	강서현, 맹지주, 이우성 (3명)
패키지 활용	크롤링 : pandas, BeautifulSoup, sklearn 전처리 : pandas (numpy), konlpy(kkma) word2vec 시각화 : WordCloud, plotly, cufflinks, matplotlib, seaborn, folium, Pyproj(위치좌표변환)

키워드 분석

상권 분석에 앞서 어떤 키워드를 가지고 있는 지 뉴스 크롤링 진행
(새해 기간을 제외하고 '골목상권'으로 2021년 네이버 뉴스 데이터 크롤링)

<WordCloud 사용 관련 키워드 확인>
(골목, 상권 키워드 제외)

데이터 출처

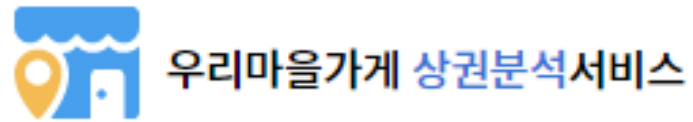
주제 선정

분석 프로세스 설계

1. 데이터 수집
2. 전처리
3. 분석
4. 시각화
5. 활용

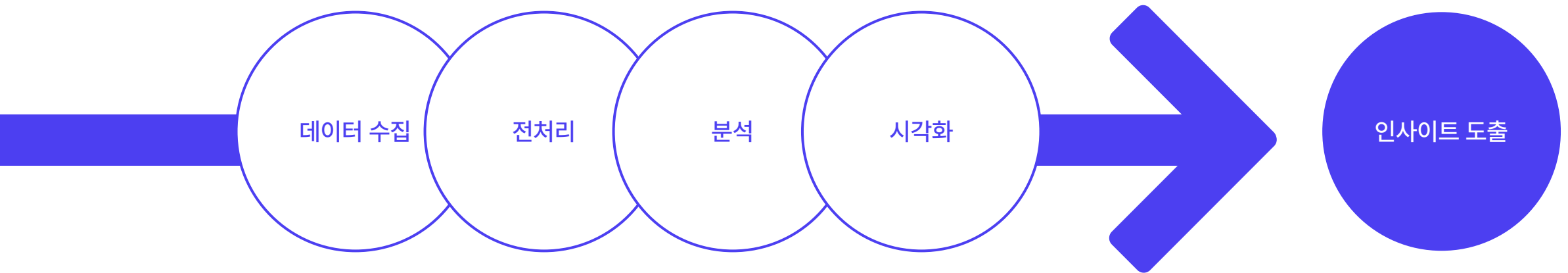
결과

참고자료



Analyze process

작업과정 : 데이터 수집 ~ 인사이트 도출



가설 설정

가설을 세우고 검증하는 방식

가설1

거리두기가 시행된 시점 모든 구의 매출이 떨어졌을 것이다

가설1-1

거리두기로 인해 특정 상권에 영향을 줬을 것이다

가설1-2

각 구 마다의 매출의 차이는 유동인구로 인해 발생할 것이다

가설2

주거인구와 폐업률은 연관성이 있을 것이다

거리두기가 시행된 시점 모든 구의 매출이 떨어졌을 것이다

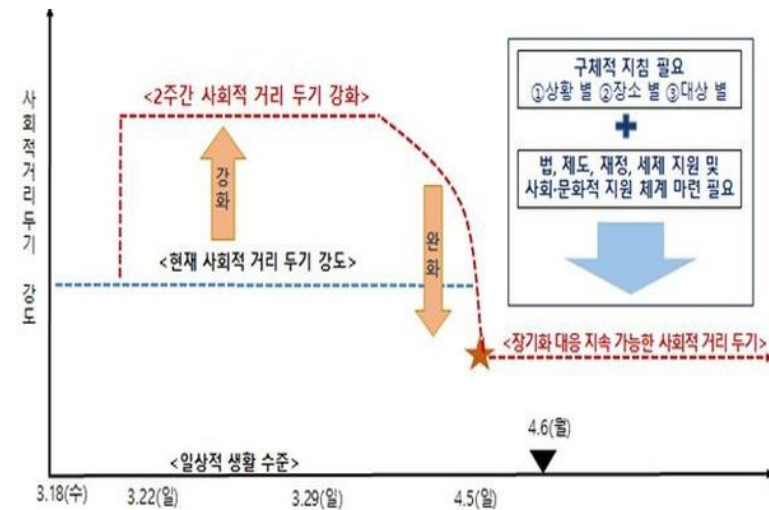
이유 : 거리두기가 실행되며 유동인구가 적어지고 소상공인들의 매출감소로 연결되었을 것이기 때문

* 사회적 거리두기(Social Distancing)란

전염병이 지역사회에 확산되는 것을 막기 위해 실시된 감염 예방 캠페인



<사회적 거리두기 단계 강화와 완화 이미지>



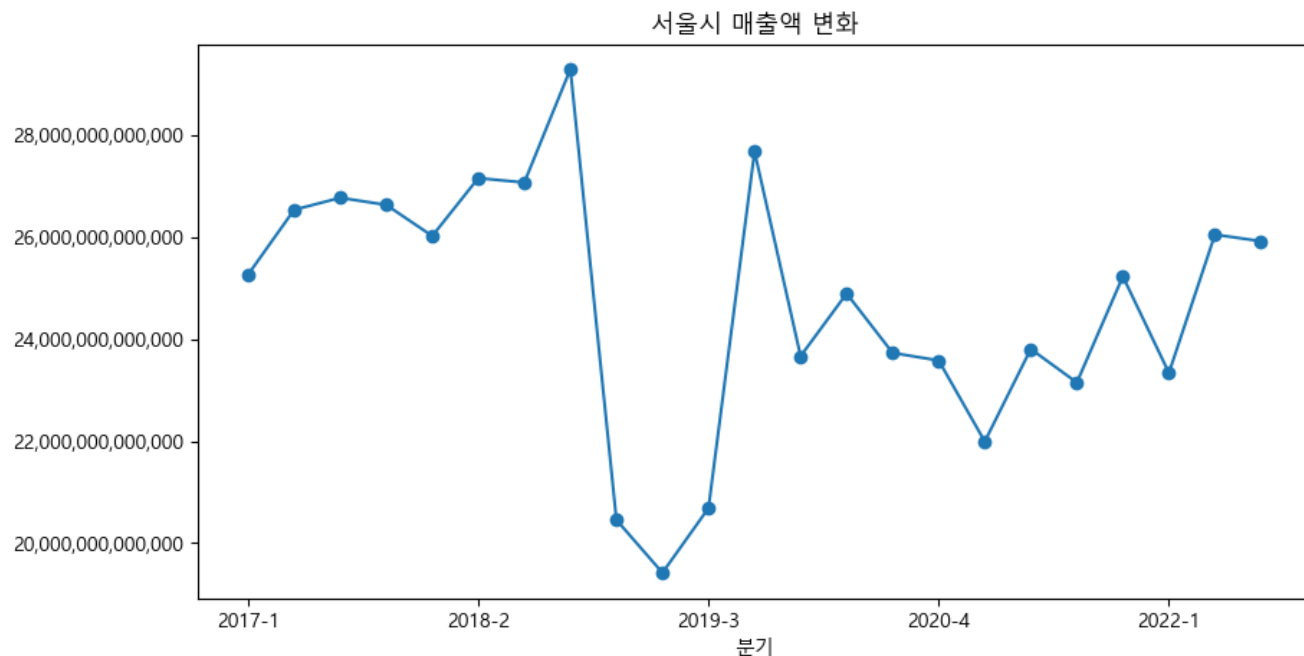
가설1 거리두기가 시행된 시점 모든 구의 매출이 떨어졌을 것이다

거리두기(2020.02.29~2021.10.31) 이후 매출은 어땠을까

분석결과

▶ 거리두기가 전체적인 매출에 영향을 끼치는 지 상관관계를 확인하기 어려웠다

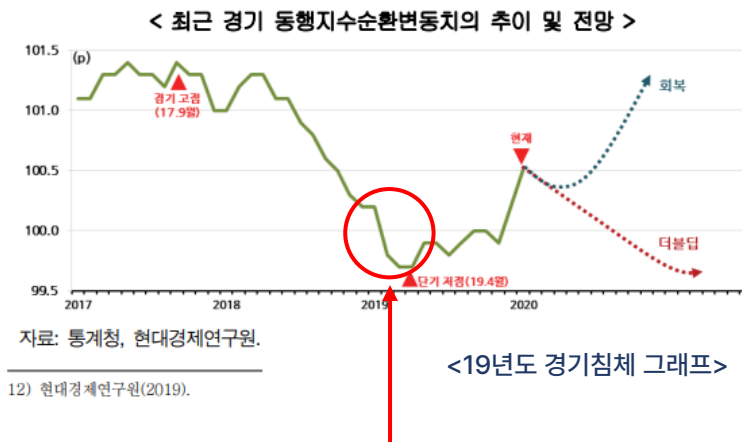
코로나가 시작되며 20년도 1분기가 급격하게 떨어지지만, 거리두기 단계가 가장 심한 시점 전체적인 매출에 영향을 준다는 직접적인 상관관계를 확인하기 어려웠다



왜 그랬을까

제로섬게임

코로나로 인해 지는 산업이 있다면 뜨는 산업이 있기 때문에 결국 전체적인 매출이 떨어지는 영향을 주고 있다고 보기 어렵다. 오히려 경기침체와 같은 상황이 상권의 매출액에 변화 큰 변화를 더 주고 있었다



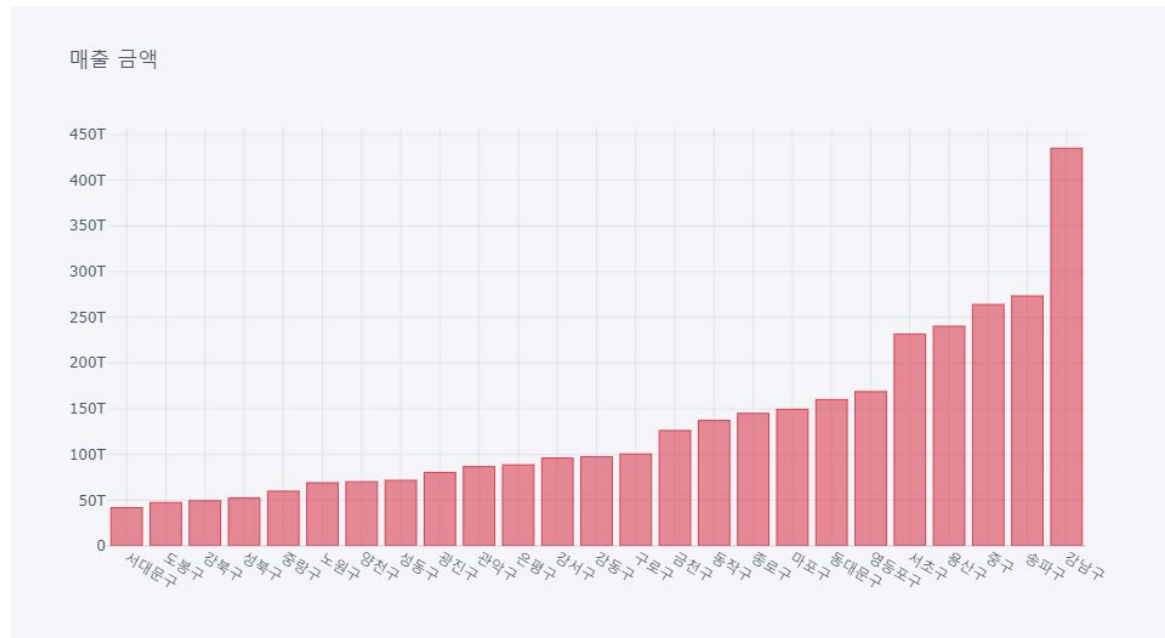
2019년 1분기 성장률 급락은 민간 부문이 부진한 가운데 정부의 성장기여도가 축소되면서 발생한 현상이다

거리두기로 인해 특정 상권에 영향을 줬을 것이다

이유 : 전체 매출 내 특정 상권에 영향이 있을 것이라 판단하였다.

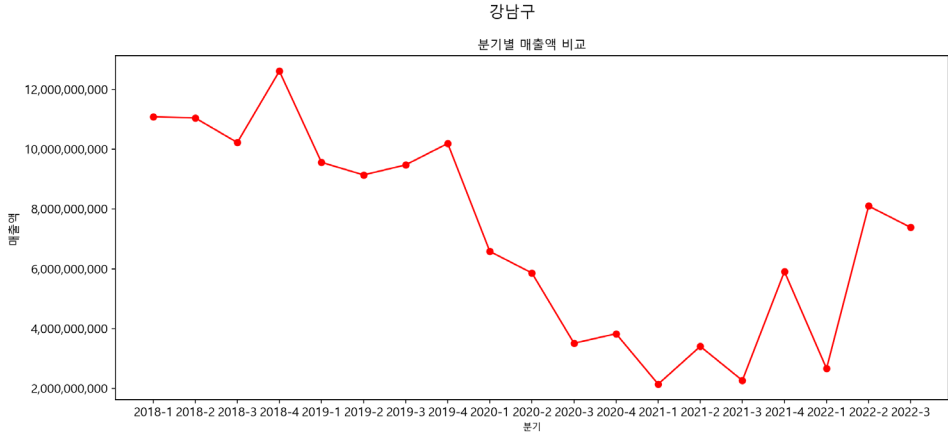
매출이 가장 높은 강남구의 '노래방', 'PC방', '의약품', '편의점'의 매출을 살펴보았다

<서울시 각 구의 매출> 가장 높은 매출을 내고 있는 강남구

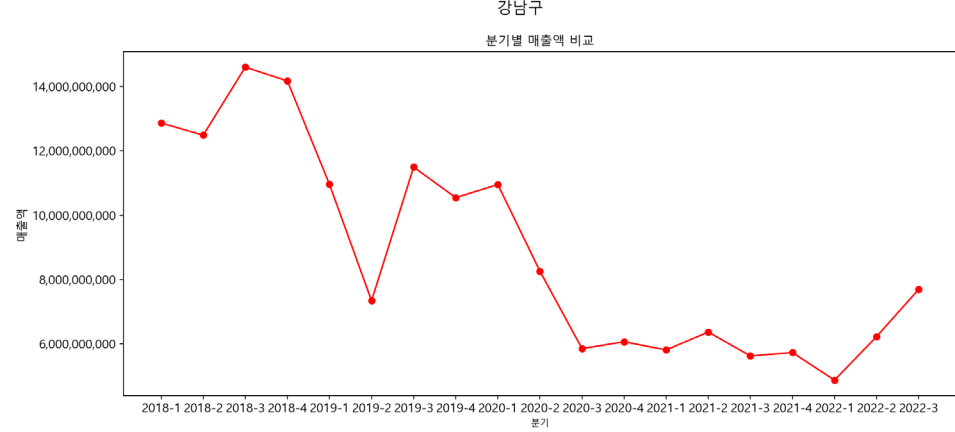


가설1-1 거리두기로 인해 특정 상권에 영향을 줬을 것이다 => △

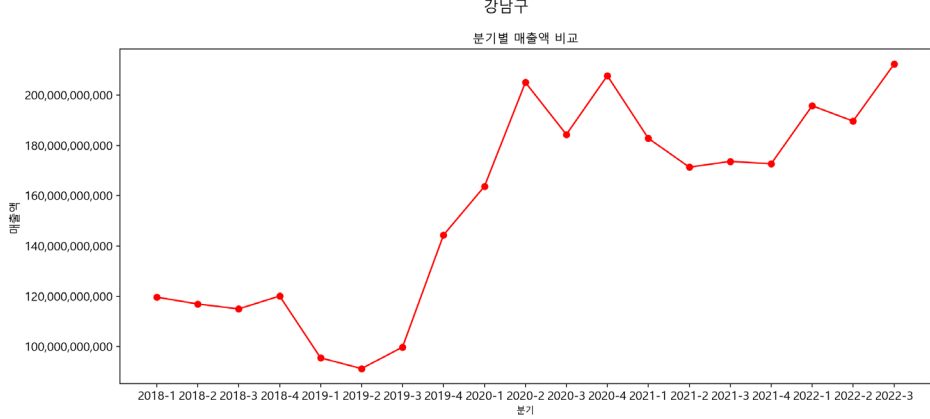
<강남구_노래방 매출 (2018-2022)>



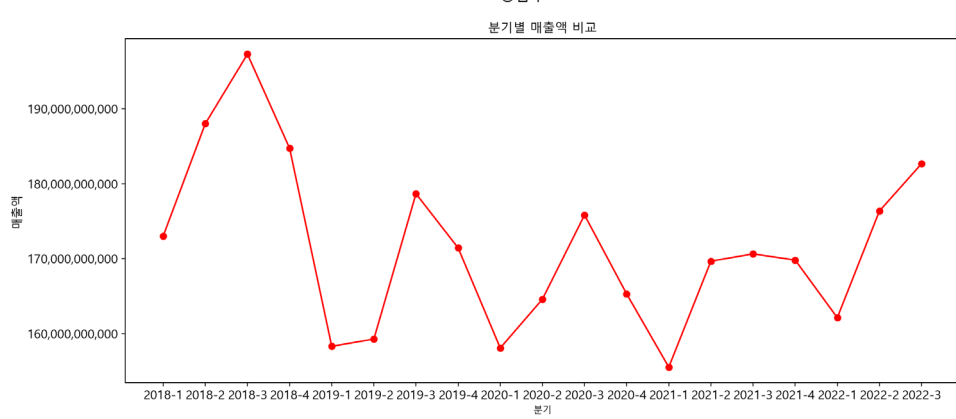
<강남구_PC방 매출 (2018-2022)>



<강남구_의약품 매출 (2018-2022)>



<강남구_편의점 매출 (2018-2022)>



분석결과

▶

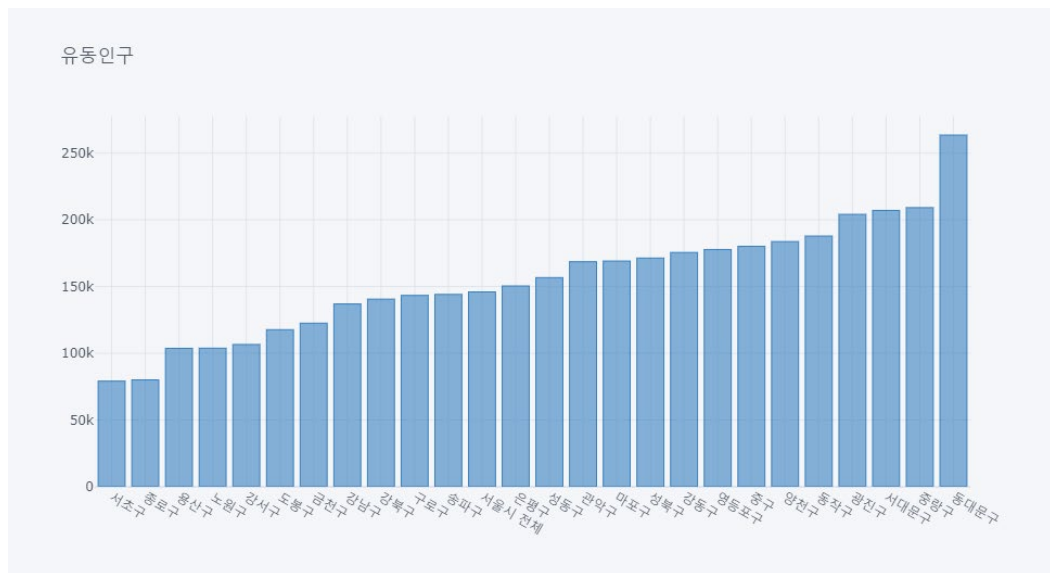
특정 업종은 매출에 영향을 주었다

노래방과 PC방 업종에 경우 타격을 입고, 의약품은 매출 상승 추이를 확인할 수 있었고, 편의점의 경우 코로나 거리두기로 인한 연관관계를 확인하기 어려웠다

가설
1-2

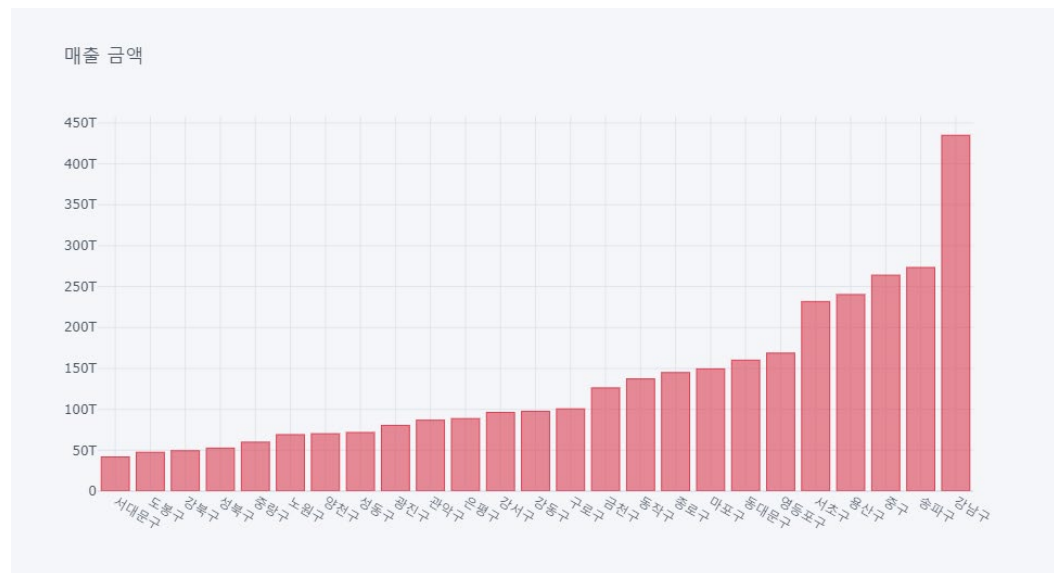
각 구 마다의 매출 차이는 유동인구로 인해 발생할 것이다

이유 : 각 구마다 유동인구가 다르고 소상공인들의 매출감소로 연결되었을 것이기 때문
서울시의 지역구별로 유동인구와 매출을 확인해보았다



<서울시 각 구별 유동인구 (21년도)>

- 상위 3개 구 : 동대문구, 중랑구, 서대문구
- 하위 3개 구 : 서초구, 종로구, 노원구



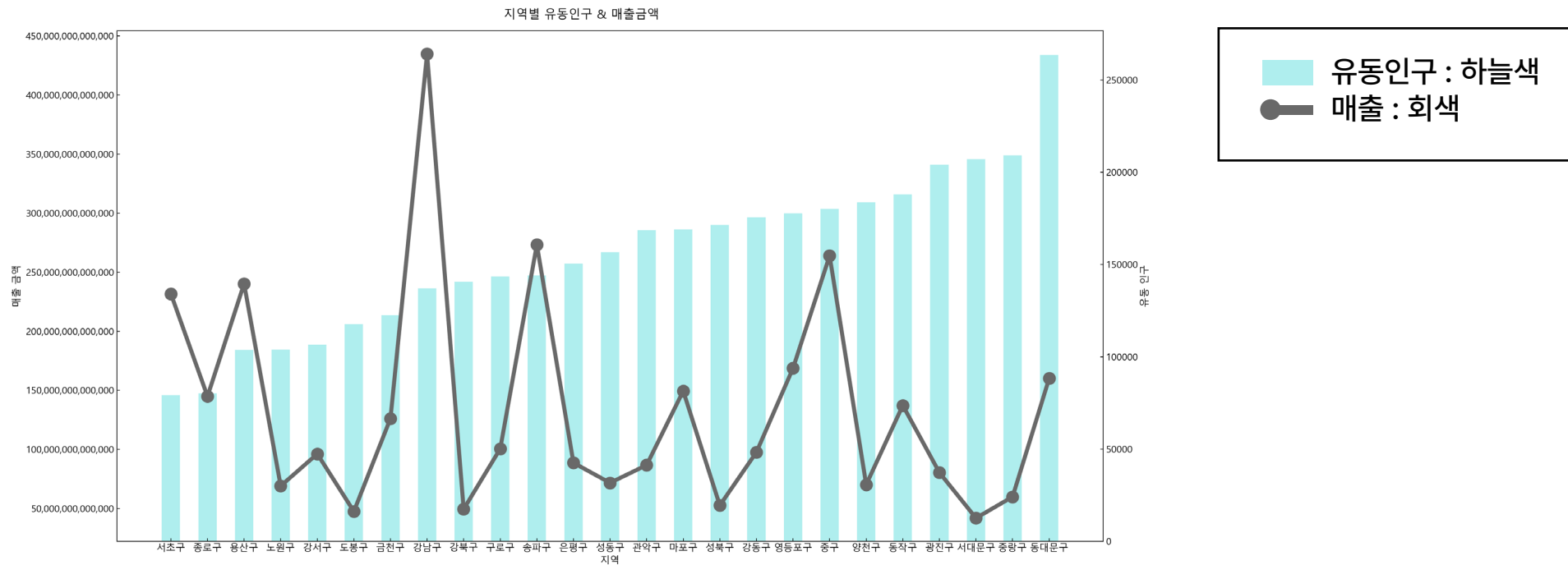
<서울시 각 구별 매출 (21년도)>

- 상위 3개 구 : 강남구, 송파구, 중구
- 하위 3개 구 : 서대문구, 도봉구, 강북구

가설1-2 각 구마다의 매출 차이는 유동인구로 인해 발생할 것이다 => X

유동인구가 많은 곳은 매출이 높고, 유동인구가 적은 곳은 매출이 낮을까
유동인구가 많은 3개의 구와 하위 3개구의 매출을 보았다 (2021년도)

<서울시 유동인구 + 매출>



분석결과

각 구마다의 유동인구와 매출의 연관관계는 크게 발견할 수 없었다

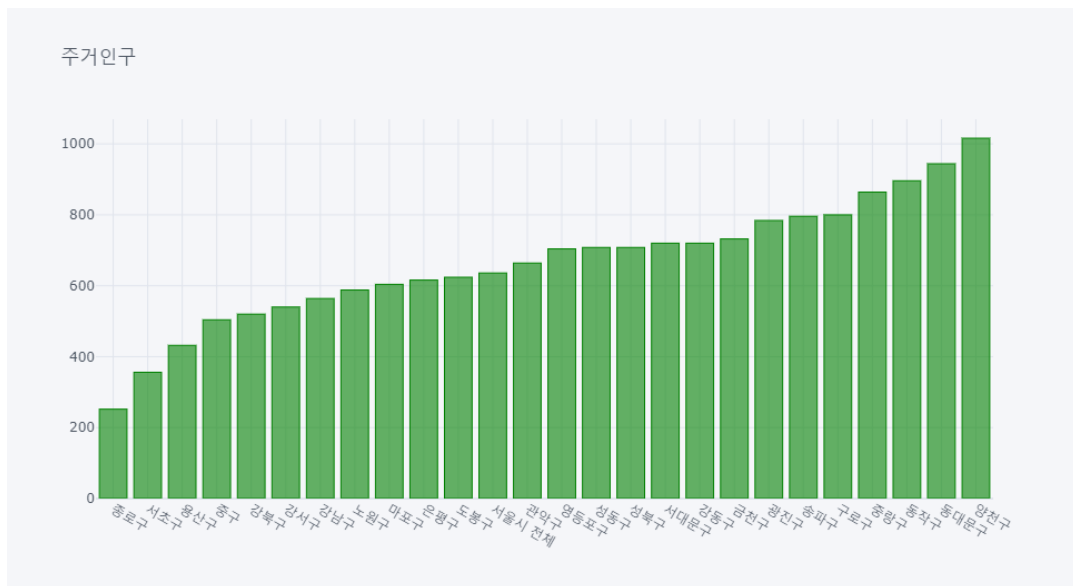
유동인구가 많다고 매출이 높은 것은 아니었다

가설1-2 각 구마다의 매출 차이는 유동인구로 인해 발생할 것이다

그렇다면 어떠한 이유로 차이점 발생할까?

생활인구 유형별로 확인해 보았다

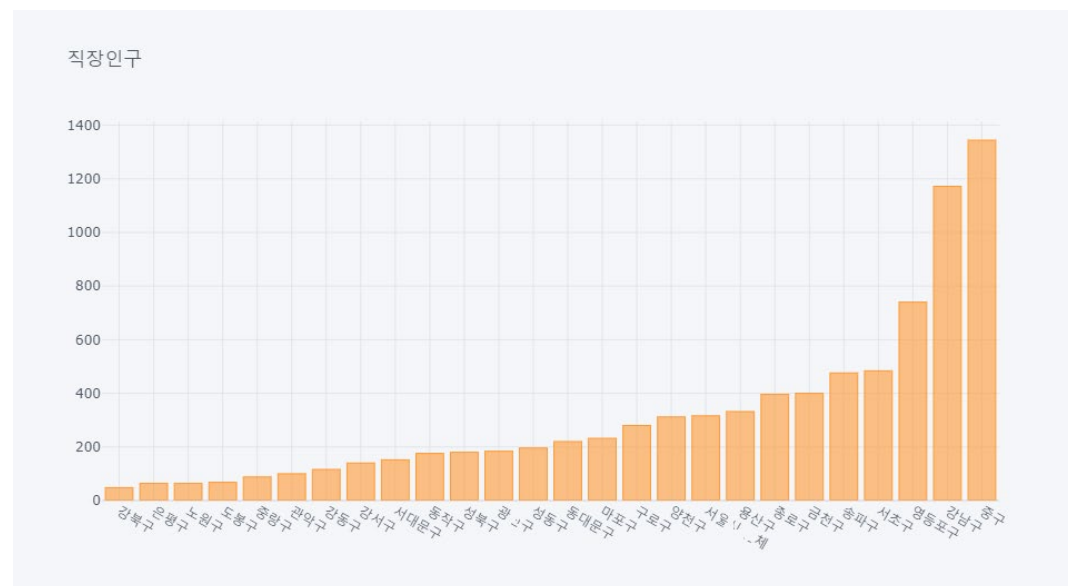
<서울시 주거인구>



주거인구 수

- 상위 3개 구 : 양천구, 동대문구, 동작구
- 하위 3개 구 : 종로구, 서초구, 용산구

<서울시 직장인구>



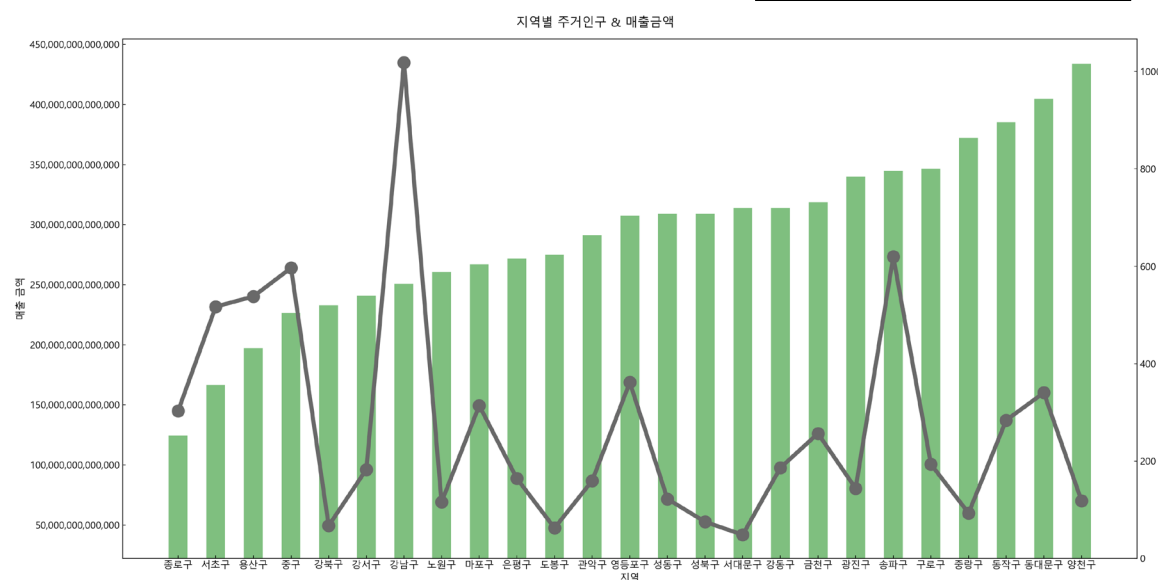
직장인구 수

- 상위 3개 구 : 강남구, 중구, 영등포구
- 하위 3개 구 : 강북구, 은평구, 노원구

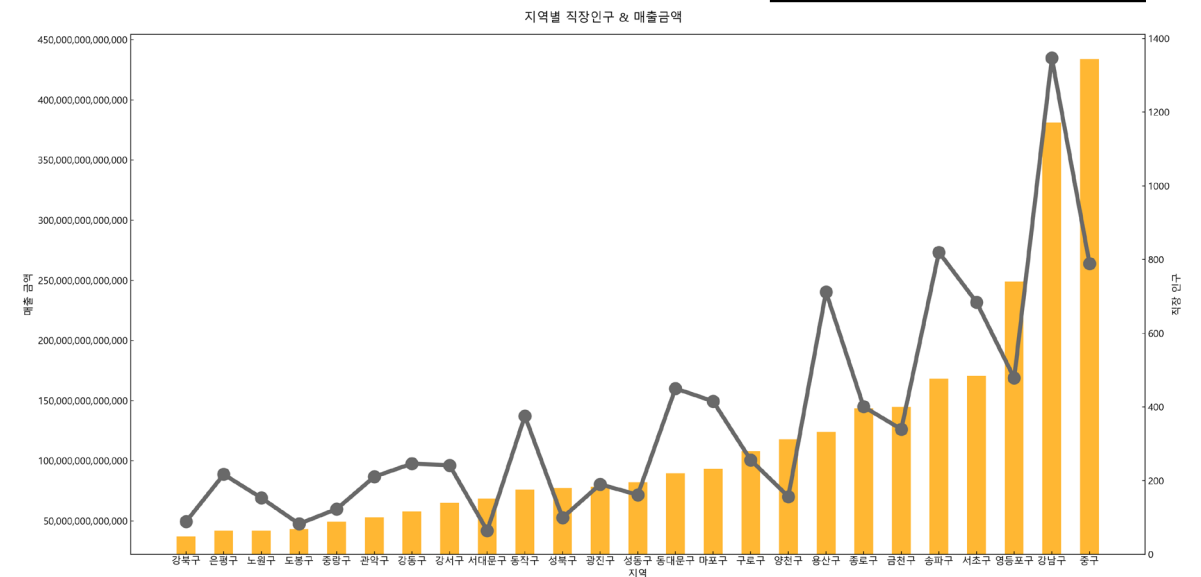
가설1-2 각 구마다의 매출 차이는 유동인구로 인해 발생할 것이다 => X

주거인구와 직장인구가 매출에 어떤 영향을 끼치는 지 확인

<서울시 주거인구 + 매출>



<서울시 직장인구 + 매출>



분석결과

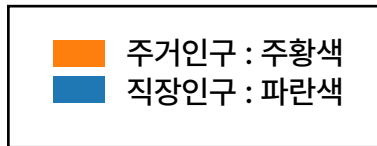
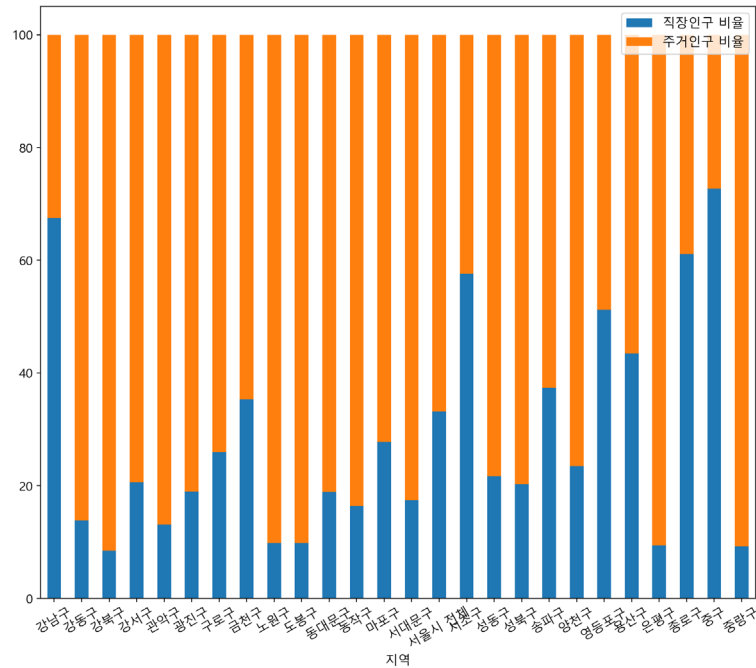
직장 인구 수와 매출은 대체로 비례하는 추이를 볼 수 있다.

코로나 시기 (2021) 주거인구에 따른 매출은 상관 관계를 가지지 않았고
직장인구와 매출은 연관성 있는 그래프를 그리며 직장인이 많은 곳에 매출액이 많은 경향을 보였다

가설 2

주거인구와 폐업률은 연관성이 있을 것이다

이유 : 직장인구가 코로나 시기 상권 매출에 많은 영향을 주었다면 (가설 1-2 결과),
주거인구가 많은 곳에는 비교적 타격이 덜할 것이다 = 폐업률 낮을 것이다



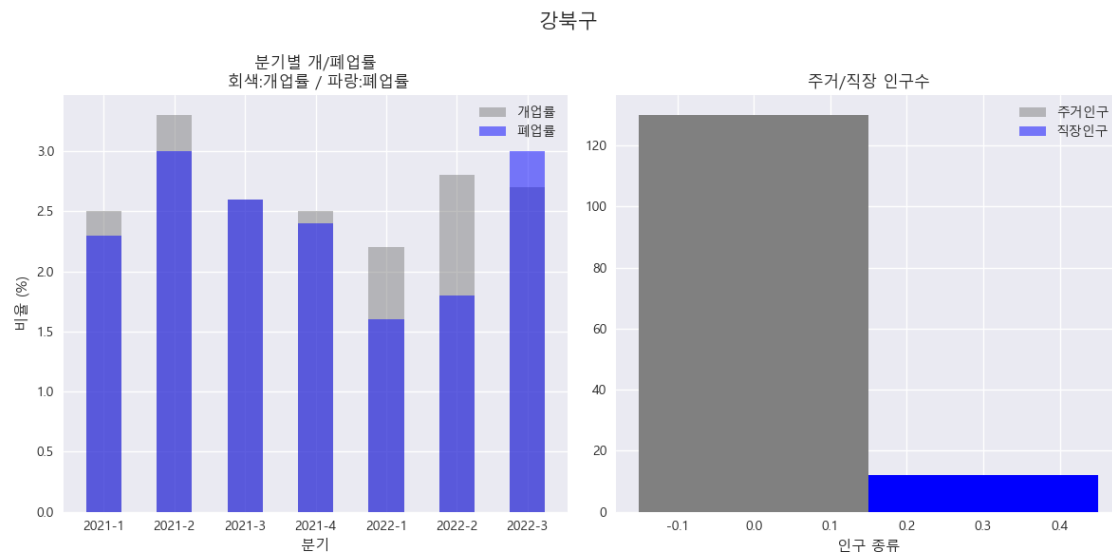
- 주거인구 비율
- 상위 3개 구 : 강북구, 은평구, 중랑구
- 직장인구 비율
- 상위 3개 구 : 중구, 종로구, 강남구

<주거인구와 직장인구의 비율>

가설 2 주거인구와 폐업률은 연관성이 있을 것이다

주거인구 비율이 높은 강북구와 직장인구 비율이 높은 중구 개폐업률 비교

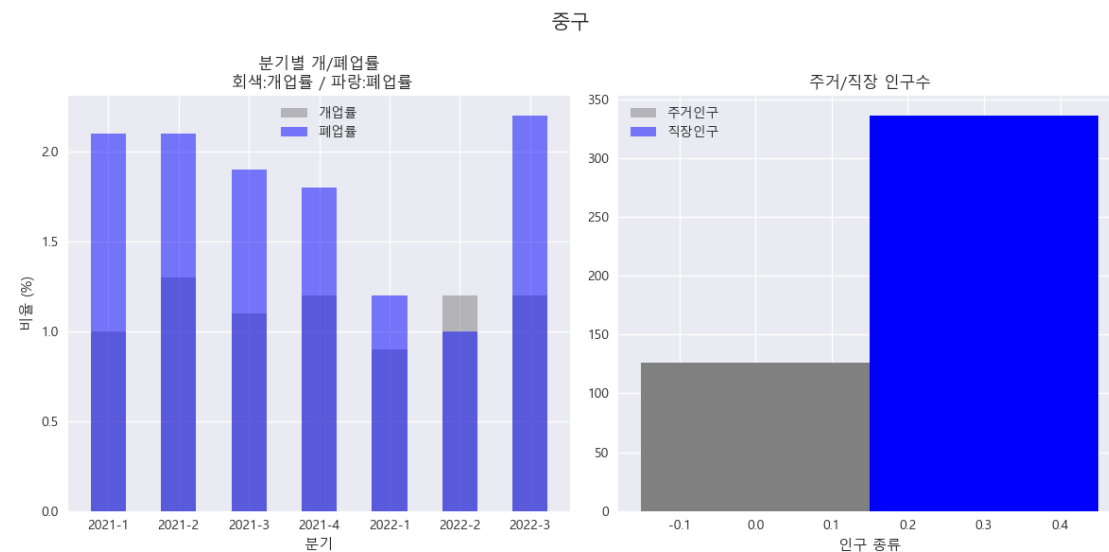
개업률 : 회색
폐업률 : 파란색



<강북구의 개폐업률>

* 회색이 보이면 폐업률보다 개업률이 높은 것

<주거인구 비율이 높은 강북구>



<중구의 개폐업률>

* 회색을 덮으면 폐업률이 더 높은 것

<직장인구의 비율이 높은 중구>

▶ 주거인구가 많은 곳은 비교적 폐업률이 낮았고, 직장인구 비율이 높은 중구는 폐업률이 높았다

분석결과 종합

가설 1

거리두기가 시행된 시점 모든 구의 매출이 떨어졌을 것이다

- ▶ 거리두기가 전체적인 매출에 영향을 끼치는 지 상관관계를 확인하기 어려웠다

가설1-1

거리두기로 인해 특정 상권에 영향을 줬을 것이다

- ▶ 특정 업종의 매출액에 영향을 주었다
Ex) 매출 : 노래방 하락, 의약품 상승

가설1-2

각 구 마다의 매출의 차이는 유동인구로 인해 발생할 것이다

- ▶ 유동인구와 매출의 연관관계는 크게 발견할 수 없다
유동인구가 많다고 매출이 높은 것은 아니었다
- ▶ 직장 인구 수와 매출은 대체로 비례하는 추이.

가설 2

주거인구가 많은 곳은 폐업률이 낮을 것이다

- ▶ 주거인구가 많은 곳은 비교적 폐업률이 낮았고, 직장인구 비율이 높은 중구는 폐업률이 높았다

한계점 및 어려웠던 점

배경지식 부족으로 인한 분석 한계

- 해당 분야에 대한 기초 지식이 부족하여 얘기치 못한 현상에 대한 해석이 어려웠다
(ex) 19년도의 매출 급락)
- 인과관계를 확인하기 위해 다른 요인으로 인한 변수가 많았다

타율이 낮은 데이터 분석 방법

- 지도 짚어보기/요일별 분석 등 적합하지 않은 접근도 겪었다
- 프로젝트의 주제와 목적 변경 야기

확증편향의 위험

- 뚜렷한 인과관계를 확인하고자 하는 마음이 있었고, 보고 싶은 것만 보게 될 위험에 처했었다.
- 어려워 현상에 대해 명확히 밝혀내지 못한 것이 아쉬웠다

빅데이터 분석 및 서비스 기획 부트캠프 자유 팀프로젝트

서울시 상권 공공데이터 EDA : 코로나가 상권에 미친 영향 과 현황 분석

E.O.D

Date : 22.12.30

Team. 골목식당

강서현, 맹지주, 이우성