



4(회적거리두기)조_권성구/김용호/김진주/이대경/황수빈



multicampus



- 1.프로젝트 배경
- 2.팀 구성 및 역할
- 3.수행 절차 및 방법
- 4.결론 및 향후 과제
- 5.느낀점



01. 프로젝트 배경

1.프로젝트 주제

코로나 백신이 영화 매출액에 미치는 영향 분석 및 시각화

2.프로젝트 개요

지역에 따라 백신접종이 영화산업에 미치는 영향을 파악하기 위해 제1도시(서울)와 제2도시(부산)의 데이터 분석 및 비교

[포토뉴스] 백신 접종 땐 영화관람료 할인

기사입력

"백신 맞으면 영화표 반값"...지자체들도 혜택 경쟁













01. 프로젝트 배경

3.프로젝트 구조



DATA RABINITED ATA

공공데이터 OPEN API



크롤링 데이터





데이터 수집

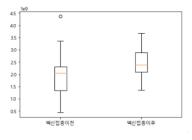


- Numpy 및 Pandas 패키지
- Selenium 및 matplotlib 패키지

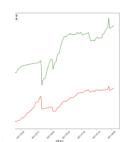
데이터 처리 및 분석







시각화





02. 팀 구성 및 역할

1.팀 구성 및 역할

김용호(팀장)

- 백신통계 및 영화데이터 전처리
- 2019년 영화데이터 전처리
- 서울 및 부산 백신과 영화 시각화

권성구

- 주별 영화데이터 수집 자동화
- 서울의 거리두기와 영화 시각화

김진주

- 백신통계API 데이터수집
- 2020년 영화데이터 전처리
- 부산의 거리두기와 영화 시각화

이대경

- 백신통계 데이터 전처리
- 2021년 영화데이터 전처리
- 서울 상관분석 및 시각화

황수빈

- 주별 영화데이터 수집 자동화
- 부산 상관분석 및 시각화



1. Work-Flow

구분	기간	활동	비고
	08.02 ~ 08.06	주제 선정 및 데이터 검색	
기획	08.10	프로젝트 아이디어발표	
	08.11 ~ 08.12	주제 구체화 및 데이터 수집	1차 피드백
분석	08.13 ~ 08.17	데이터 전처리	
正省	08.18 ~ 08.20	데이터 분석 및 시각화	2차 피드백
바이 미 아서	08.21 ~ 08.23	최종 결과물 작성	
보완 및 완성	08.24	프로젝트 발표	



2. 데이터 명세

1. 영화관입장권 통합전산망 (KOBIS)

- 데이터명 : 지역별 영화점유율

- 제공형태: 기간별 Excel 파일

- 요약: 지역별, 매출액, 관객수



- 특이사항: Selenium 크롤링을 통해 데이터수집 자동화



```
sdate_range=pd.date_range('2021-01-01','2021-06-23',freq = '7D') #rea/ : 2020-12-28
  sdate_range=pd.Series(sdate_range.format())
  list(sdate_range)
  edate_range=pd.date_range('2021-01-07','2021-07-01',freq = '7D') #real : 2020-12-31
  edate_range=pd.Series(edate_range.format())
  list(edate_range)
 # webdriver 레체 셈섬
 driver =webdriver.Chrome('./driver/chromedriver')
 url = https://www.kobis.or.kr/kobis/business/stat/them/findAreaShareList.do
 driver.get(url)
for start,end in date_range:
     # 시작날짜 변경
     first_elem = driver.find_element_by_name("sSearchFrom")
     first_elem.clear()
     first_elem.send_keys(start)
     # 끝날짜 변경
     second_elem = driver.find_element_by_name("sSearchTo")
     second_elem.clear()
     second_elem.send_keys(end)
     search_button = driver.find_element_bv_xpath('//*[@id="searchForm"]/div/div[2]')
     search_button.click()
     # 엑셀 다운로드
     excel_button = driver.find_element_by_xpath('//*[@id="content"]/div[3]/div/a')
     excel_button.click()
     # 짤업참에서 확인 누르기
     pop = Alert(driver)
     pop.accept()
```



2. 데이터 명세

2. 공공 데이터포털 (data.go.kr)

- 데이터명: 코로나 예방접종 통계 데이터 조회 서비스

- 제공형태 : OPEN API 형태

- 요약: 지역명칭, 1차 및 2차 당일통계, 누적통계

	accumulatedFirstCnt	accumulated SecondCnt	baseDate	firstCnt	secondCnt	sido	totalFirstCnt	totalSecondCnt
0	61232	0	2021-03-11 00:00:00	12595	0	서울특별시	73827	0
1	74892	0	2021-03-12 00:00:00	10972	0	서울특별시	85864	0
2	86342	0	2021-03-13 00:00:00	9654	0	서울특별시	95996	0
3	96073	0	2021-03-14 00:00:00	1704	0	서울특별시	97777	0
4	97777	0	2021-03-15 00:00:00	900	0	서울특별시	98677	0
155	4042352	1551572	2021-08-13 00:00:00	28270	135399	서울특별시	4070622	1686971
156	4071166	1687149	2021-08-14 00:00:00	43931	121688	서울특별시	4115097	1808837
157	3452328	1160526	2021-08-15 00:00:00	29747	39226	서울특별시	3482075	1199752
158	4145167	1848323	2021-08-16 00:00:00	3875	1162	서울특별시	4149042	1849485
159	4149520	1849577	2021-08-17 00:00:00	123845	51259	서울특별시	4273365	1900836





```
#코로나 10 예방집중 통계데이터(서울특별시)
servicekey='
base_url = "https://api.odcloud.kr/api"
vaccine_endpoint = "/15077756/v1/vaccine-stat"

vaccine_url = base_url+vaccine_endpoint+"?serviceKey="+servicekey+"&cond[sido::EQ]="+quote("서울특별시")

response = urllib.request.urlopen(vaccine_url)
json_str = response.read().decode("utf-8")

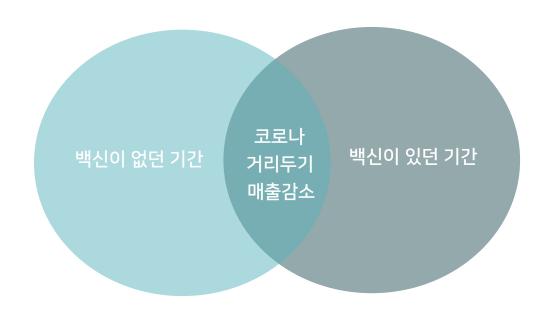
rawdata=json.loads(json_str)
seoul=json_normalize(rawdata['data'])
seoul

#busan.to_csv("./result/vc_seoul.csv",encoding="utf-8-sig")
```



3. 데이터 샘플링 기준 선정 (표본추출)

어떤 데이터들을 어떤 기준으로 비교할 것인가? '동일한 조건 ' 을 가진 표본 추출



코로나 백신 접종 전

2020년 8월 16일 ~ 2021년 3월 10일

코로나 백신 접종 후

2021년 3월 11일 ~ 2021년 6월 30일

- 동월기준으로 선택하지 않은 이유 : 매출액에 큰 영향을 미칠 것으로 생각되는 요소인 거리두기를 고려할 수 없음
- 백신 접종후의 표본에서 7월 이후를 제외한 이유 : 거리두기 개편안(5단계->4단계) 적용에 따라 동일한 표본을 구할 수 없음



4. 데이터 탐색 및 전처리

1. 코로나 백신 접종 전/후의 매출액 결측치 처리

```
np.set_printoptions(precision=6, suppress=True)
seoul_bv_sell = np.array(seoul_bv['매출액'])
seoul_av_sell = seoul_av['매출액']

len(seoul_bv_sell)
len(seoul_av_sell)
29
16
```



```
지역 매출액 관객수 주 지역 매출액 관객수 주 이 서울시 2370034110 245237 2021-03-12 이 서울시 3361656490 368804 2020-08-16 1 서울시 2089378170 221908 2021-03-19 1 서울시 2189550260 247399 2020-08-23 2 서울시 2156717370 223590 2021-03-26 2 서울시 1842116650 200100 2020-08-30 3 서울시 2394668990 261118 2021-04-02 3 서울시 1733144880 188395 2020-09-06 4 서울시 1826626560 193633 2021-04-09 4 서울시 1571785690 172993 2020-09-13 5 서울시 1481982720 158687 2021-04-16 5 서울시 1686750342 194747 2020-09-20
```

```
tmp = []
  for i in range(13):
     tmp.append(np.random.randint(1722277400,3097454600,dtype='int64'))
  #접종이후의 데이터가 접종 전보다 부족해서, 결측치를 평균값으로 대체
  seoul_av_sell = np.append(seoul_av_sell,tmp )
  len(seoul by sell)
  len(seoul av sell)
  29
   29
  array([3361656490, 2189550260, 1842116650, 1733144880, 1571785690,
         1686750342, 4365204790, 2310220710, 1635498260, 2051049940,
         2345716060, 2686707390, 2595964430, 2087858250, 1340672930,
          978608420, 752040660, 606209800, 1249933430, 754663270,
          446257140, 761371720, 2050589220, 2287327000, 2358611200,
         2299673230, 1955377880, 2226500250, 2370034110], dtype=int64)
i array([2370034110, 2089378170, 2156717370, 2394668990, 1826626560,
         1481982720, 1709790330, 1796128240, 2713673630, 1365881880,
         2551193710, 2886481360, 3308331430, 3216698320, 3014873260,
         3675390440, 2829551649, 2276646354, 2500665097, 2478983920,
         2149687962, 2959886167, 2039959888, 3026195954, 1788117360,
         1755376030, 1963498570, 1803241784, 2272225233], dtype=int64)
```



4. 데이터 탐색 및 전처리

2. 영화 및 접종데이터 데이터 통합

지역	매출액	관객수	주
서울시	625176800	71052	2021-01-01
서울시	446257140	50249	2021-01-08
서울시	761371720	85940	2021-01-15
서울시	2050589220	230553	2021-01-22
서울시	2287327000	246303	2021-01-29
서울시	2358611200	247998	2021-02-05
서울시	2299673230	241331	2021-02-12
서울시	1955377880	215607	2021-02-19
서울시	2226500250	242385	2021-02-26
서울시	2370034110	245237	2021-03-05



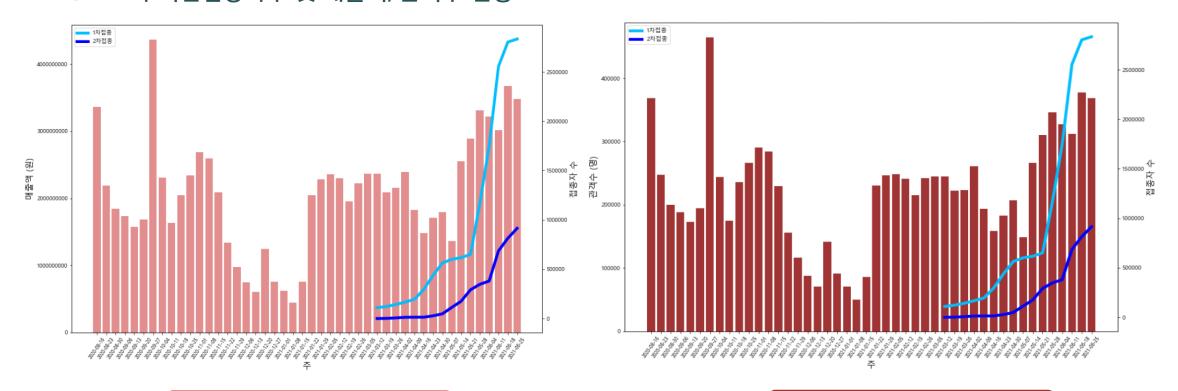
	일자	1차접종누계	2차접종누계	주간1차접종자수	주간2차접종자수
0	20210318	109536	0	0	0
1	20210325	122249	1725	12713	1725
2	20210401	141584	6227	19335	4502
3	20210408	166563	12397	24979	6170
4	20210415	195071	13502	28508	1105

	지역	매출액	관객수	주	1차접종누계	2차접종누계
0	서울시	3381656490	368804	2020-08-16	NaN	NaN
1	서울시	2189550260	247399	2020-08-23	NaN	NaN
2	서울시	1842116650	200100	2020-08-30	NaN	NaN
3	서울시	1733144880	188395	2020-09-06	NaN	NaN
4	서울시	1571785690	172993	2020-09-13	NaN	NaN
			•	• •		
30	서울시	2370034110	245237	2021-03-12	109536.0	0.0
31	서울시	2089378170	221908	2021-03-19	122249.0	1725.0
32	서울시	2156717370	223590	2021-03-26	141584.0	6227.0
33	서울시	2394668990	261118	2021-04-02	166563.0	12397.0
34	서울시	1826626560	193633	2021-04-09	195071.0	13502.0



5. 분석 및 시각화

1. 코로나 백신접종자수 및 매출액/관객수 현황



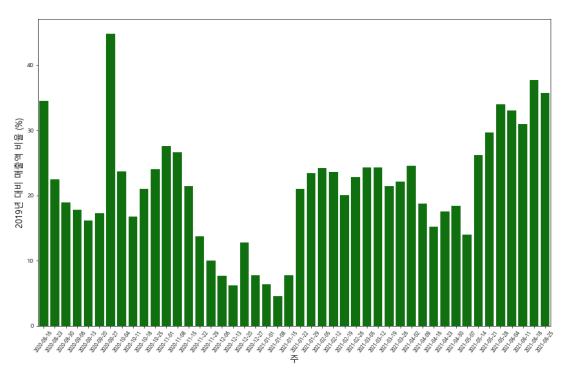
서울 백신접종자 수 및 매출액

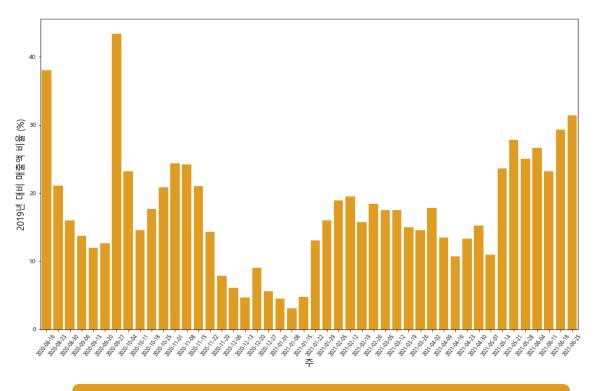
서울 백신접종자 수 및 관객수



5. 분석 및 시각화

2. 서울 및 부산 2019년 대비 코로나이후 매출액 비율 [매출액 비율 (%) = (주별 매출액 / 2019년 주 평균) *100]





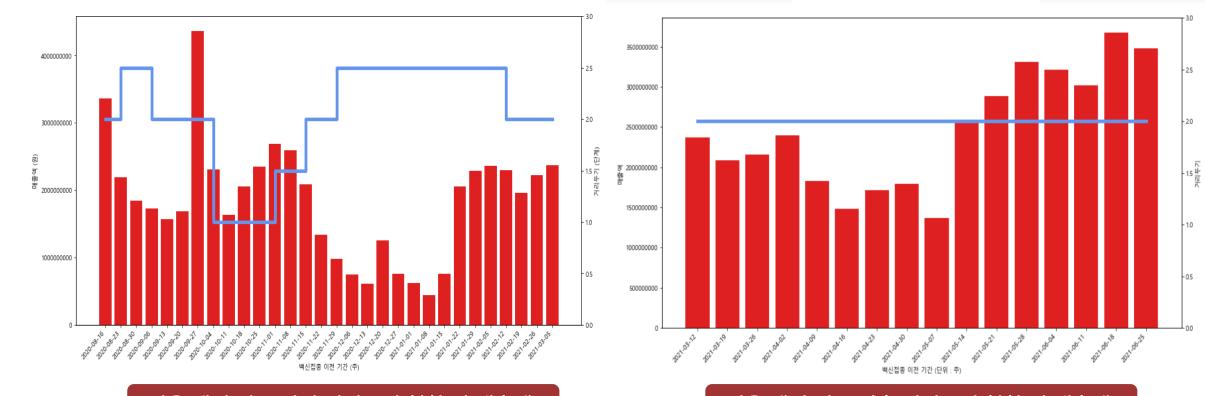
서울 2019년 매출액 평균대비 코로나이후 매출액

부산 2019년 매출액 평균대비 코로나이후 매출액



5. 분석 및 시각화

2. 백신접종 전/후 거리두기에 따른 매출액 비교 (서울)



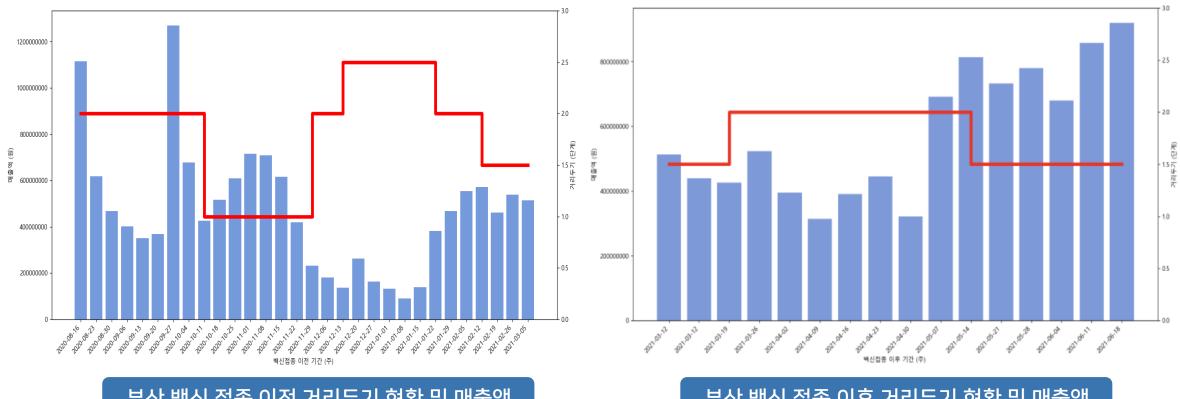
서울 백신 접종 이전 거리두기 현황 및 매출액

서울 백신 접종 이후 거리두기 현황 및 매출액



5. 분석 및 시각화

2. 백신접종 전/후 거리두기에 따른 매출액 비교 (부산)



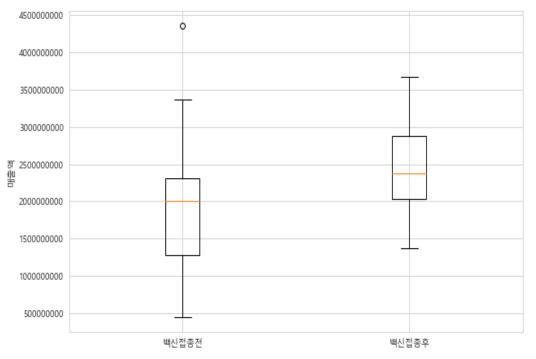
부산 백신 접종 이전 거리두기 현황 및 매출액

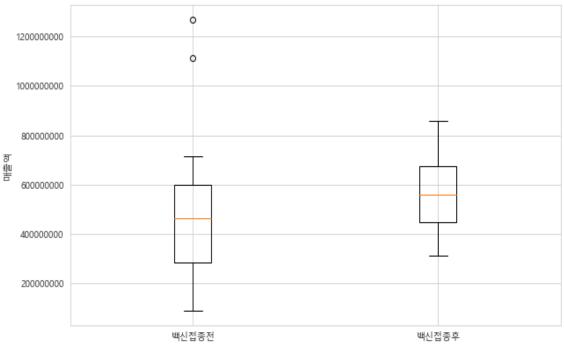
부산 백신 접종 이후 거리두기 현황 및 매출액



5. 분석 및 시각화

3. 서울 및 부산 백신접종 전/후 매출액 비교 BoxPlot





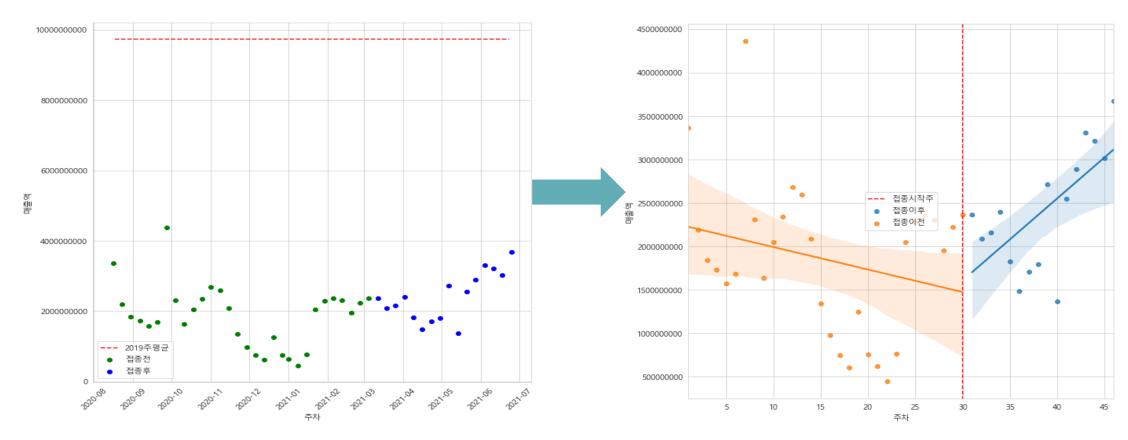
서울 백신 접종 전/후 매출액 비교

부산 백신 접종 전/후 매출액 비교



5. 분석 및 시각화

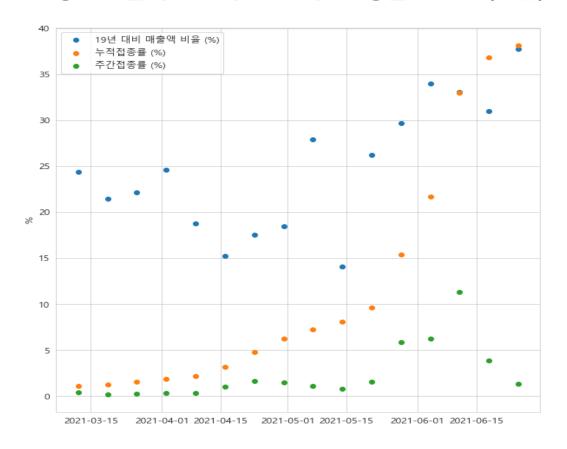
4. 백신 접종 전/후의 산점도와 회귀직선 (서울)

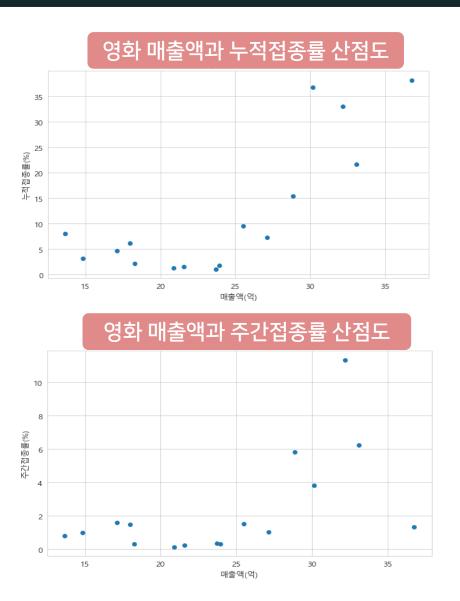




5. 분석 및 시각화

5. 영화 매출액과 누적/주간 백신접종률 산점도 (서울)

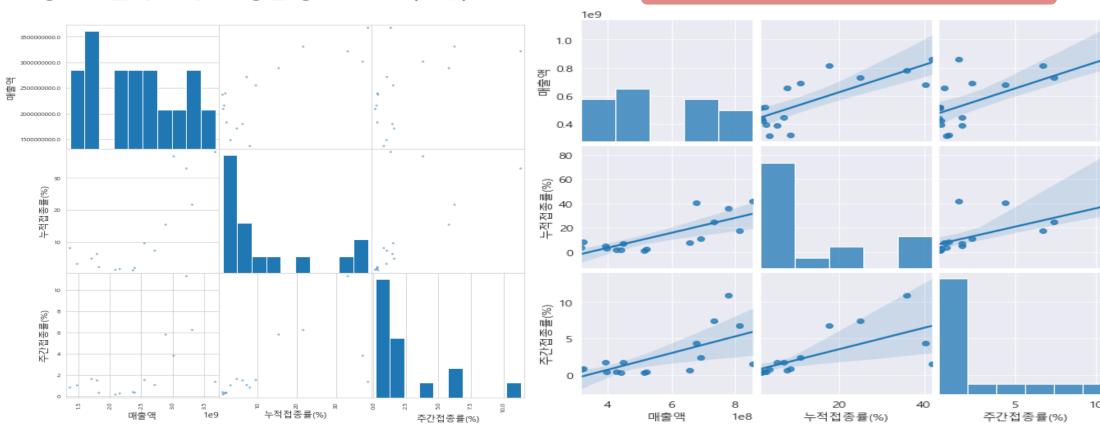






5. 분석 및 시각화

6. 영화 매출액과 백신접종률 종합산점도 (서울)



누적/주간 접종률과 영화매출 산점도



5. 분석 및 시각화

7. 상관계수 분석 및 시각화 (서울)

pearsonr(final_df_modi['매출액'],final_df_modi['주간접종률(%)'])

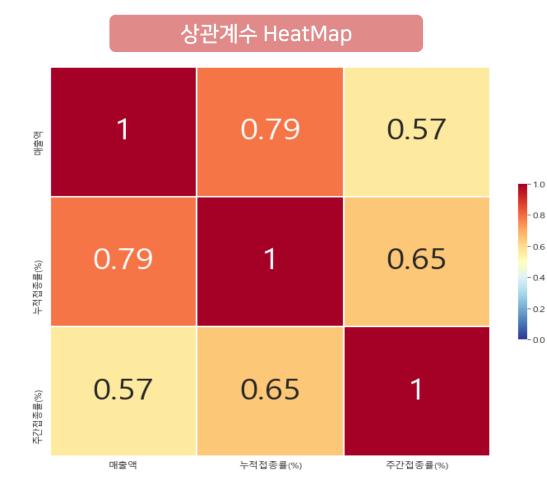
#p value가 0.05 미만이므로 유의미한 상관성이 있으며 # 상관계수가 0.57이므로, 보통의 상관관계가 있다고 볼 수 있다.

(0.5688300866681443, 0.021480252505826934)

pearsonr(final_df_modi['매출액'],final_df_modi['누적접종률(%)'])

#p value가 0.05 미만이므로 유의미한 상관성이 있으며 # 상관계수가 0.79이므로, 아주 강한 상관관계가 있다고 볼 수 있다.

(0.7879655314227366, 0.0002888065130348548)



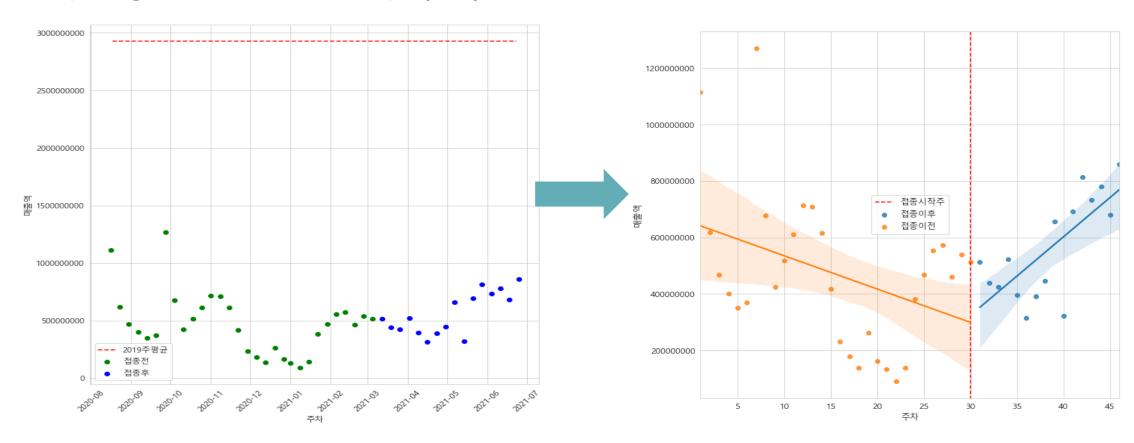
- 0.8

- 0.4



5. 분석 및 시각화

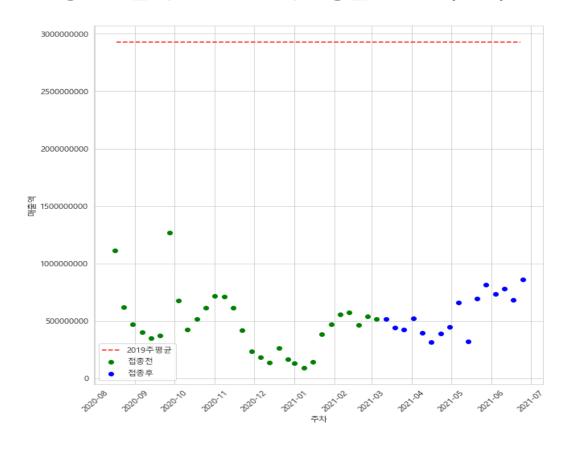
4. 백신 접종 전/후의 산점도와 회귀직선 (부산)

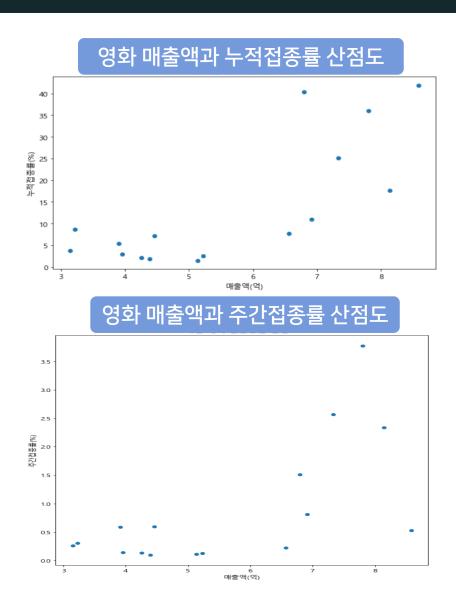




5. 분석 및 시각화

5. 영화 매출액과 주간/누적 접종률 산점도 (부산)

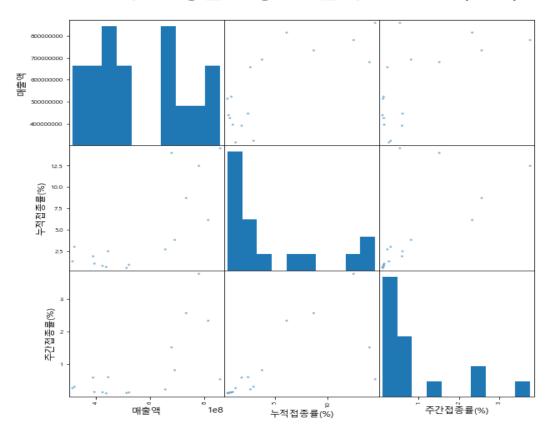




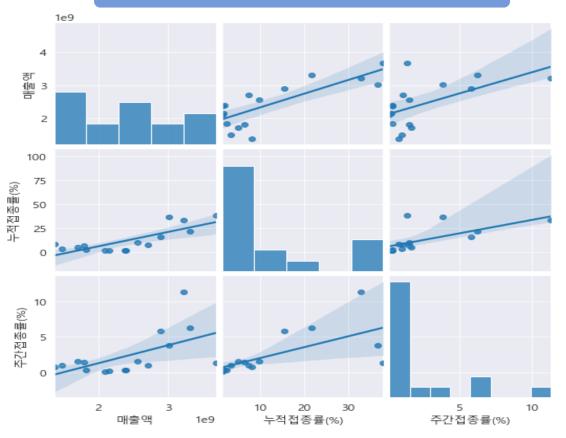


5. 분석 및 시각화

6. 코로나 백신접종률과 영화매출액의 산점도 (부산)



누적/주간 접종률과 영화매출 산점도





5. 분석 및 시각화

7. 상관계수 분석 및 시각화 (부산)

```
pearsonr(fin_df_modi['매출액'],fin_df_modi['주간접종률(%)'])
# pvalue가 0.05미만이므로 유의미한 상관성이 있으며
# 상관계수가 0.65이므로 보통의 상관관계 있음
```

(0.652243442896062, 0.006175325270177162)

```
pearsonr(fin_df_modi['매출액'],fin_df_modi['누적접종률(%)'])
# pvalue가 0.05 미반이므로 유의미한 상관성이 있으며
# 상관계수가 0.77이므로, 아주 강한 상관관계가 있다고 볼 수 있다.
```

(0.7704029130695069, 0.00047926863271377876)

상관계수 HeatMap



- 0.8

- 0.6

- 0.4

- 0.2



04. 결론 및 향후과제

1.결론

- 1. 코로나 백신과 영화 매출액에는 상관관계가 있다.
- 2. 누적접종률이 주간접종률보다 강한 상관관계를 갖고 있다.
- 3. 코로나 백신과 영화매출액의 상관관계는 지역별 차이가 있다.
- 4. 영화산업에서 코로나 백신의 경제적효과가 있을 것 같다.



2. 한계점 및 해결방안

- 1. 거리두기를 고려하다보니 계절성을 고려하고, 적절한 크기의 표본을 선정하는데 어려움이 있었다→ 접종 이후의 데이터 확보 필요
- 2. 지역별 차이에 대한 표본이 부산 한 곳이므로 , 이를 명확하게 확인할 수 없었다 → 타 지역 데이터분석 필요
- 3. 코로나에 둔감하게 된 사람들의 심리 고려하지 못했다 → 설문자료 및 여론조사 관련 추가 데이터수집 필요





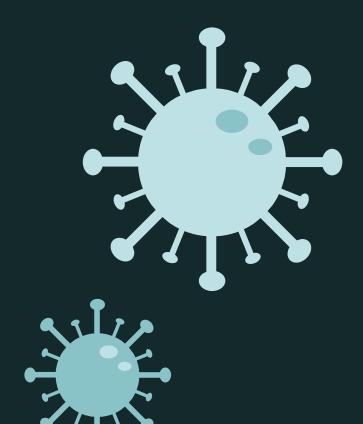
05. 느낀점

1.느낀점

- 프로젝트 진행에 유효한 데이터를 수집하기 어려웠다.
- 실제 데이터는 데이터 수집 및 전처리에 많은 노력이 필요했다.
- GIT으로 형상관리를 진행하면서 폴더구조, 파일명 등 사전에 협의가 필요한 부분이 있었다

2. 아쉬운점

- 비대면으로 프로젝트를 진행하면서 서로가 원하는 의도한 내용에 대해 전달이 어려웠다.
- 배달앱,OTT데이터 등 최근 소비패턴관련 데이터를 구하지 못해서 처음 기획과 달리 범위가 좁혀진게 아쉬웠다



Thank U



비고. 데이터 출처

- 프로젝트 GIT 저장소 : https://github.com/vvspearlvvs/MultiCampus_Project
- 영화관입장권 통합전산망 (KOBIS) : https://www.kobis.or.kr/kobis/business/stat/them/findAreaShareList.do
- 코로나19 예방접종 통계 데이터 조회 서비스(data.go.kr): https://www.data.go.kr/data/15077756/openapi.do