

COVID19 데이터 분석

김용주

장은준

INDEX

1. Project Summary
 - Project background and purpose
 - Member and Role

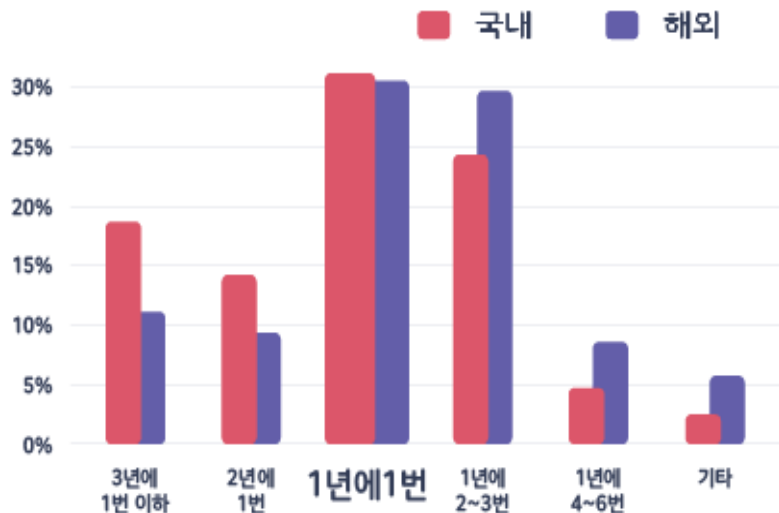
2. Process
 - Data used
 - Preprocessing
 - Data Analysis

3. Comprehensive Insight

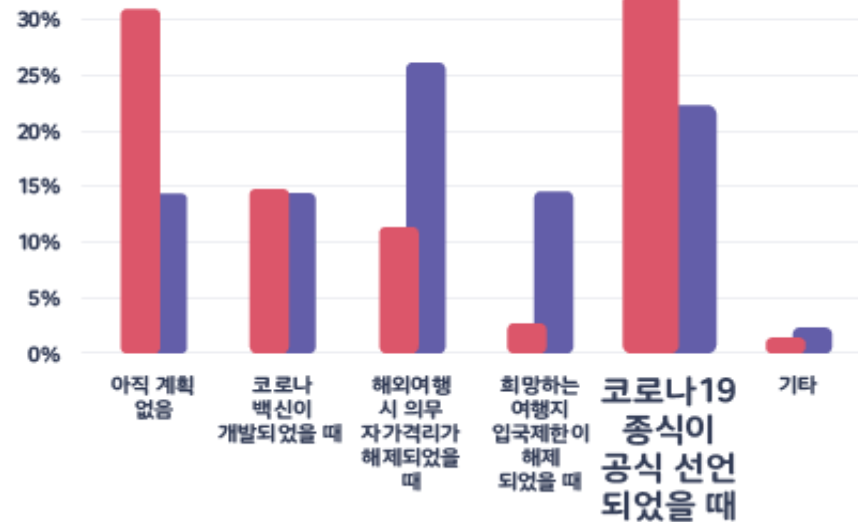
Project Summary

Project background

1. 코로나19 발생 이전 평균 해외여행 빈도



2. 다음 해외여행 계획시기



Project background

 아이뉴스24 **PiCK** | 2021.08.08. | 네이버뉴스

포스트 코로나 준비하는 여행업계 "향후 폭발수요 노린다"

 뉴스시스 | 2021.08.09. | 네이버뉴스

야놀자 조사, 응답자 80% "코로나 이후 여행 빈도 줄어"

 동아일보 |  B1면 1단 | 2021.08.10. | 네이버뉴스

코로나로 여행 횟수 줄어도 고급호텔 이용 185% 늘어

Project purpose

1. 코로나 바이러스 데이터 시각화를 통해서 현 코로나 상황을 진단한다.
2. 코로나 바이러스 데이터 시각화를 통해서 현 코로나 백신 현황을 진단한다.
3. 현 코로나 상황과 백신 현황에 따라서 여행사업 추이를 진단한다.
4. 이러한 세 과정을 통해서 현 코로나 상황에 따른 여행사업 현황을 짐작한다.

Member and Role



김용주

- 데이터 전처리 및 시각화 분석
- ppt 제작



장은준

- 데이터 전처리 및 시각화 분석
- ppt 제작

Process

Data used(1) - 크롤링 데이터

서울특별시 코로나 현황 사이트 데이터 크롤링 :

<https://www.seoul.go.kr/coronaV/coronaStatus.do>

확진자 현황 (30001번~현재)

확진자 현황 (1번~30000번)

자치구 전체



100



검색

연번	환자	확진일	거주지	여행력	접촉력	퇴원현황
75136	236412	2021-08-22	은평구	-	감염경로 조사중	-
75135	237181	2021-08-22	영등포구	-	감염경로 조사중	-
75134	237252	2021-08-22	서초구	-	기타 확진자 접촉	-
75133	236423	2021-08-22	구로구	-	기타 확진자 접촉	-
75132	237182	2021-08-22	강남구	-	기타 확진자 접촉	-

Data used(1) - 크롤링 데이터

Crawling code

```
import pandas as pd
import numpy as np
import requests

def get_seoul_covid19_30001_current(page_num):
    start_no = (page_num - 1) * 100

    url = 'https://news.seoul.go.kr/api/27/getCorona19Status/get_status_ajax.php?draw={}'.format(page_num)
    url = '{}&order%5B0%5D%5Bcolumn%5D=0&order%5B0%5D%5Bdir%5D=desc&start={}&length=100&search%5Bvalue%5D=&search%5B'

    response = requests.get(url)
    json_data = response.json()

    return json_data

page_list = []
all_page = 404

for page_num in range(1, all_page + 1):
    one_page_json_data = get_seoul_covid19_30001_current(page_num)
    one_page_df = pd.DataFrame(one_page_json_data['data'])
    page_list.append(one_page_df)

    time.sleep(0.5)

corona_df = pd.concat(page_list)
for_columns = pd.read_html('https://www.seoul.go.kr/coronaV/coronaStatus.do')
corona_df.columns = for_columns[4].columns.tolist()
```

Crawling result

	연번	환자	확진일	거주지	여행력	접촉력	퇴원현황
0	<p class='corona19_no'>70870</p>	222111	2021-08-13	기타	-	감염경로 조사중	<b class="">>-
1	<p class='corona19_no'>70869</p>	221232	2021-08-13	기타	-	감염경로 조사중	<b class="">>-
2	<p class='corona19_no'>70868</p>	221356	2021-08-13	성북구	-	감염경로 조사중	<b class="">>-
3	<p class='corona19_no'>70867</p>	221710	2021-08-13	동대문구	-	감염경로 조사중	<b class="">>-
4	<p class='corona19_no'>70866</p>	220502	2021-08-13	타시도	-	기타 확진자 접촉	<b class="">>-

Data used(2) – 크롤링 데이터

서울특별시 코로나19 백신 예방접종 현황

<http://data.seoul.go.kr/dataList/OA-20914/S/1/datasetView.do>

접종일	접종대상자	당일 1차접종자 수	1차접종 누계	1차접종률(%)	당일 2차접종자 수	2차접종 누계
2021.08.23	9,575,355	74,785	4,856,923	50.7	130,725	2,342,494
2021.08.22	9,795,426	7,449	4,780,478	48.8	561	2,210,777
2021.08.21	9,575,355	65,430	4,771,762	49.8	26,500	2,210,176
2021.08.20	9,575,355	116,425	4,704,843	49.1	66,759	2,183,532
2021.08.19	9,575,355	85,052	4,587,049	47.9	57,683	2,116,503
2021.08.18	5,762,857	93,303	4,500,520	78.1	62,833	2,058,630
2021.08.17	5,503,804	93,303	4,405,330	80.0	62,833	1,995,627
2021.08.16	4,915,812	128,626	4,273,365	86.9	94,563	1,900,836
2021.08.15	4,802,523	123,845	4,149,042	86.4	51,259	1,849,485
2021.08.14	4,802,523	3,875	4,145,096	86.3	1,162	1,848,314
2021.08.13	4,788,775	29,755	4,115,097	85.9	39,280	1,808,837
2021.08.12	4,737,032	43,931	4,070,622	85.9	121,688	1,686,971
2021.08.11	4,774,960	28,270	4,041,944	84.6	135,399	1,551,492
2021.08.10	4,419,973	31,041	4,010,500	90.7	35,492	1,515,914
2021.08.09	4,419,959	48,625	3,961,449	89.6	34,387	1,481,476
2021.08.08	4,364,248	70,831	3,890,343	89.1	35,253	1,446,195
2021.08.07	4,356,536	6,105	3,884,200	89.2	106	1,446,093
2021.08.06	4,341,603	22,222	3,860,574	88.0	8,011	1,427,142

Data used(2) – 크롤링 데이터

Crawling result

1	vaccin_data							
	접종일	접종대상 자	당일 1차접 종자 수	1차접종 누 계	1차접종 률(%)	당일 2차접종 자 수	2차접종 누 계	2차접종 률(%)
0	2021.08.15	4802523	3875	4149042	86.4	1162	1849485	38.5
1	2021.08.14	4802523	29755	4145096	86.3	39280	1848314	38.5
2	2021.08.13	4788775	43931	4115097	85.9	121688	1808837	37.8
3	2021.08.12	4737032	28270	4070622	85.9	135399	1686971	35.6
4	2021.08.11	4774960	31041	4041944	84.6	35492	1551492	32.5
...
112	2021.04.25	1023346	23323	350631	34.2	3411	18513	1.8
113	2021.04.24	1023346	200	350427	34.2	0	18513	1.8
114	2021.04.23	1027740	10195	345764	33.6	119	18376	1.8
115	2021.04.22	1023538	27074	317386	31.0	2479	15901	1.6
116	2021.04.21	1019322	22501	291857	28.6	2358	13543	1.3

117 rows × 8 columns

Data used(3) - 여행사업 데이터

코로나19 여행관련 xlsx

구분		세부항목	2021년 06월	2021년 05월	2021년 04월	2021년 03월	2021년 02월	2021년 01월	2020년 12월	2020년 11월	...	2010년 10월	2010년 09월	2010년 08월	2010년 07월	2010년 06월	2010년 05월	2010년 04월	2010년 03월	2010년 02월	2010년 01월
0	관광/여행	호텔업	89.7	85.8	77.1	73.9	70.7	57.6	67.5	92.7	...	114.6	105.0	113.5	101.8	100.0	110.6	102.3	103.9	94.5	98.1
1	관광/여행	여관업	67.7	67.6	64.4	61.7	53.0	58.0	65.0	66.6	...	106.7	100.8	114.2	105.6	99.5	99.8	103.1	104.3	98.5	104.3
2	관광/여행	여행사업	22.5	17.4	17.7	16.7	13.3	13.4	18.0	18.3	...	98.5	78.9	85.6	89.7	87.7	95.0	82.6	63.6	67.0	75.3

Preprocessing(1) - 크롤링 데이터

#연번 전처리

```
def extract_number(num_string):  
    if type(num_string) == str:  
        num_string = num_string.replace("corona19", "")  
        num = re.sub("[^0-9]", "", num_string)  
  
        #위의 정규화 식은 밑의 정규화 식을 생략하고 바로 함수화한 것.  
        #r = re.compile("[^0-9]")  
        #r.sub("", num_string)  
  
        num = int(num)  
        return num  
  
    else:  
        return num_string
```

```
corona_df['연번'] = corona_df['연번'].map(extract_number)
```

#퇴원현황 전처리

```
def extract_korean(text):  
    extracted_text = re.sub("[^가-힣]", "", text)  
  
    return extracted_text
```

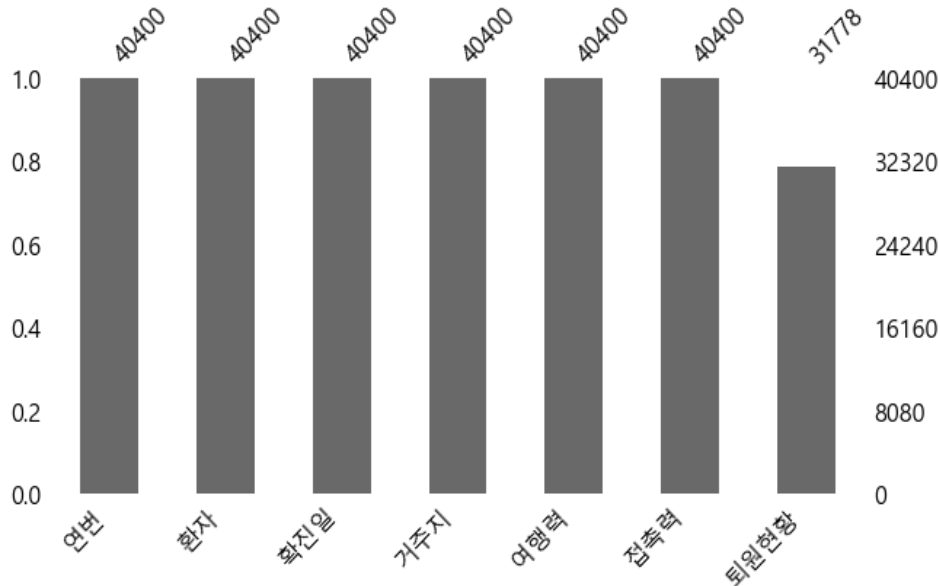
```
corona_df['퇴원현황'] = corona_df['퇴원현황'].map(extract_korean)  
corona_df.loc[corona_df['퇴원현황'].isin([""]), '퇴원현황'] = np.nan
```

	연번	환자	확진일	거주지	여행력	접촉력	퇴원현황
0	70870	222111	2021-08-13	기타	-	감염경로 조사중	NaN
1	70869	221232	2021-08-13	기타	-	감염경로 조사중	NaN
2	70868	221356	2021-08-13	성북구	-	감염경로 조사중	NaN
3	70867	221710	2021-08-13	동대문구	-	감염경로 조사중	NaN

Preprocessing(1) - 크롤링 데이터

Null value handling

1. '퇴원현황' value handling
2. `corona_data['퇴원현황'].fillna('퇴원전')`



2. '여행력' value handling

"-" 를 여행력 없음으로 간주, 보기 쉽게 바꿀 예정

```
counted_null2 = len(corona_data[corona_data['여행력'] == '-'])  
print('여행력 column null 값 개수 : ', counted_null2)  
corona_data[corona_data['여행력'] == '-'].sample(5)
```

Preprocessing(2) – 크롤링 데이터

접종일 월, 일로 나눔

```
1 #접종일을 년도, 월과 일로 분리
2 vaccin_data['년도'] = vaccin_data['접종일'].str.split(".", expand = True)[0]
3 vaccin_data['월'] = vaccin_data['접종일'].str.split(".", expand = True)[1]
4 vaccin_data['일'] = vaccin_data['접종일'].str.split(".", expand = True)[2]
```

```
1 #자료형 변경
2 vaccin_data['년도'] = vaccin_data['년도'].astype(int)
3 vaccin_data['월'] = vaccin_data['월'].astype(int)
4 vaccin_data['일'] = vaccin_data['일'].astype(int)
```

```
import datetime
t = ['월', '화', '수', '목', '금', '토', '일']
```

#요일 컬럼 생성

```
day_list = []
for i in range(0, len(vaccin_data)) :
    s = datetime.datetime(vaccin_data['년도'][i], vaccin_data['월'][i], vaccin_data['일'][i])
    day_list.append(t[s.weekday()])
sr1 = pd.Series(day_list, name='요일')
vaccin_data = pd.concat([vaccin_data, sr1], axis = 1)

vaccin_data
```


Preprocessing(3) - 여행사업 데이터

Data array code

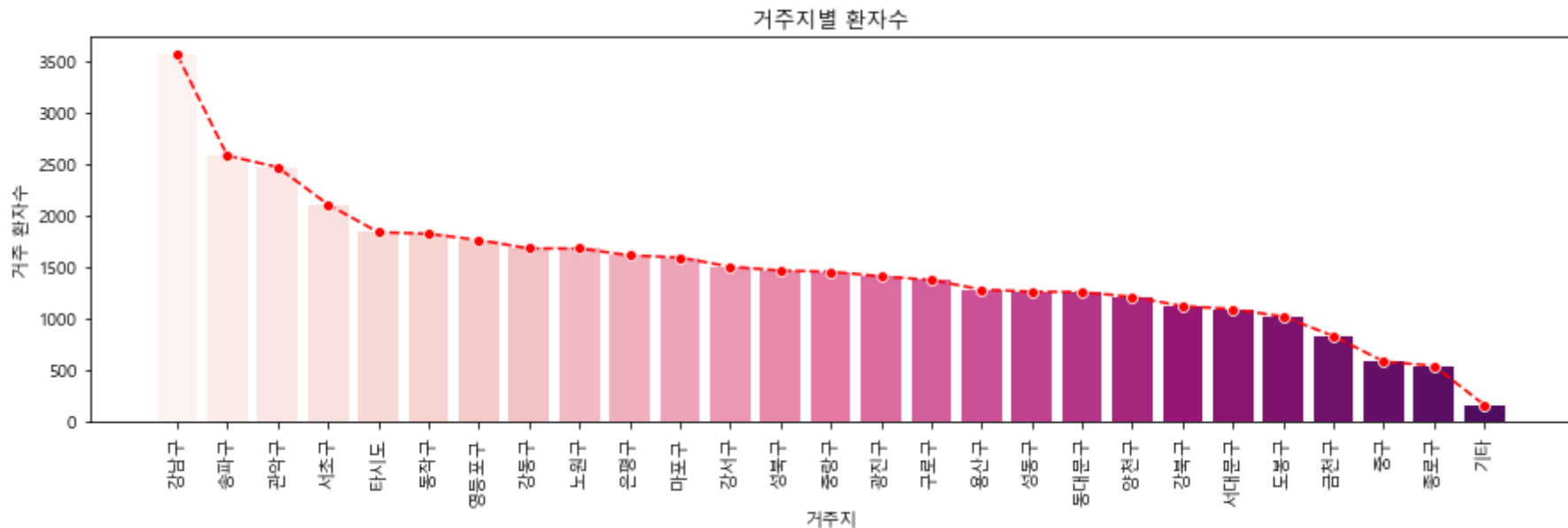
```
: trip_industry_df.T  
  
wanted_data = trip_industry_df.T.iloc[2:20,2]  
visualization_df = wanted_data.reset_index()  
visualization_df['results'] = visualization_df['results'].astype(int)
```

Result Df

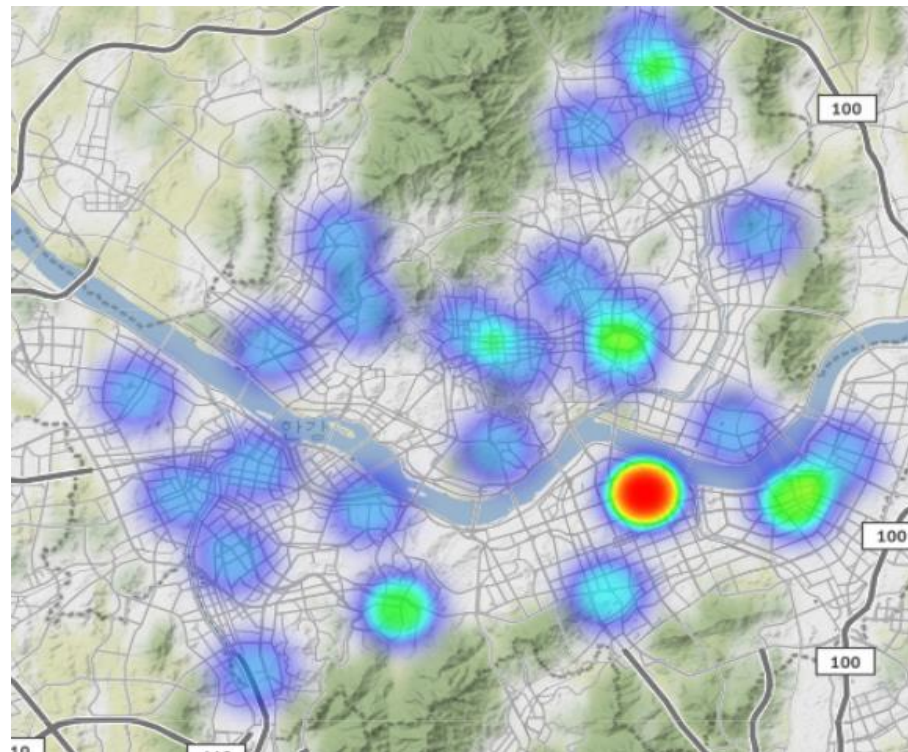
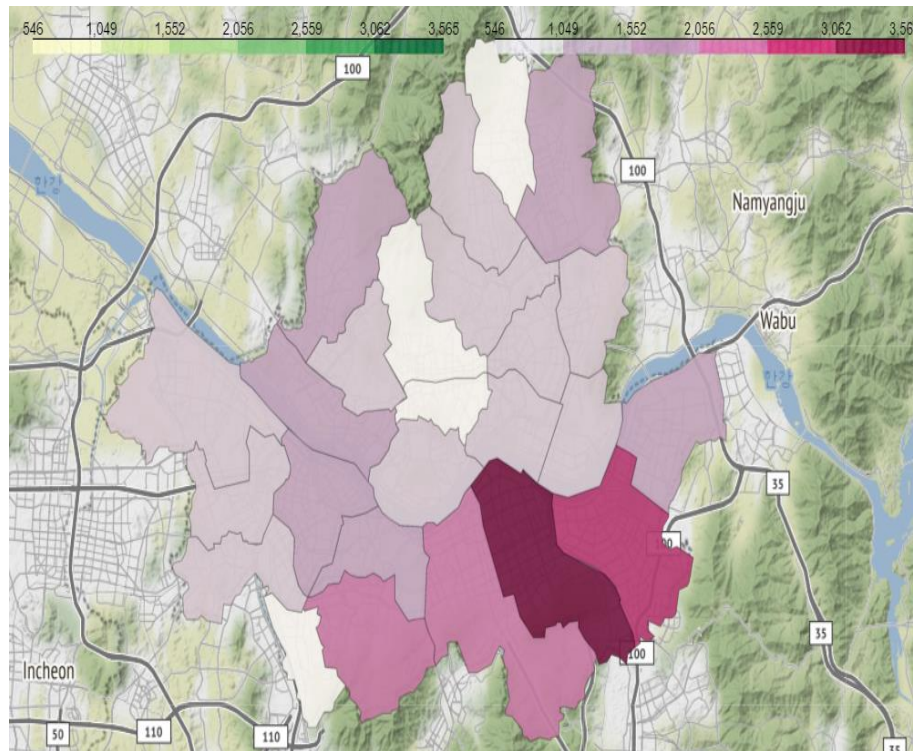
	날짜	results
17	2020년 01월	110.8
16	2020년 02월	53.6
15	2020년 03월	25.6
14	2020년 04월	12.7
13	2020년 05월	12.1
12	2020년 06월	14.1
11	2020년 07월	12.4

Data Analysis(1) - 크롤링 데이터

1. 거주지별 환자수

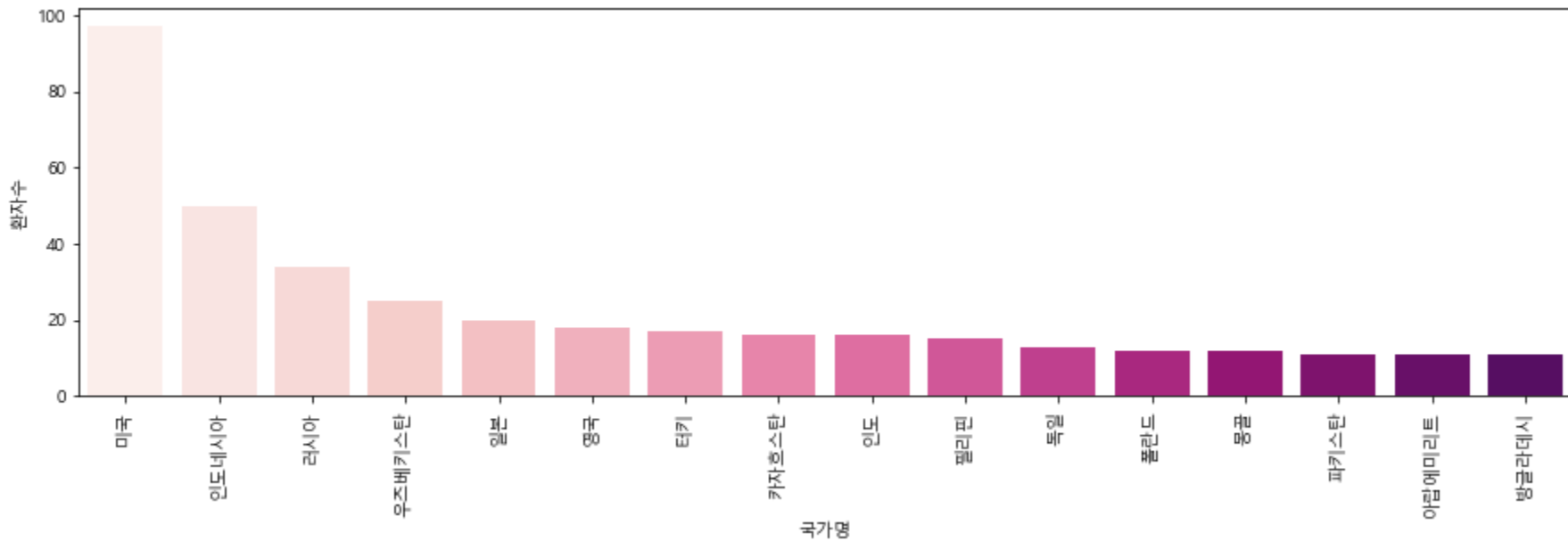


Data Analysis(1) - 크롤링 데이터



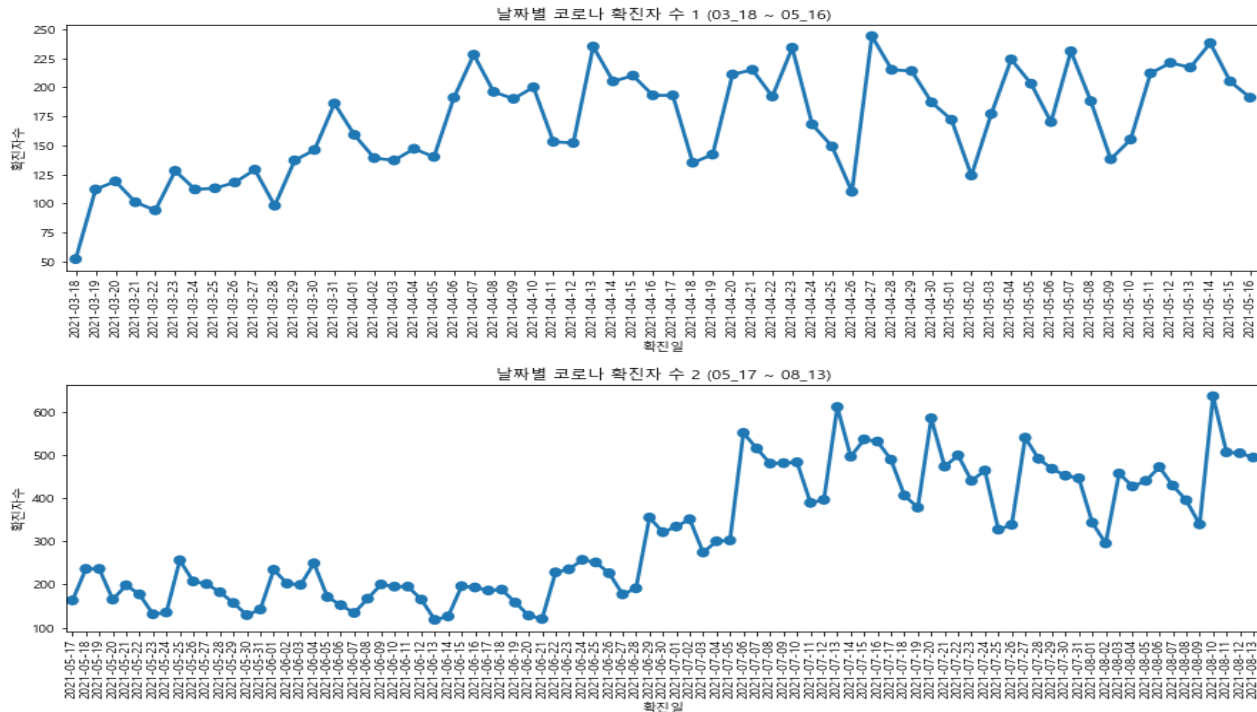
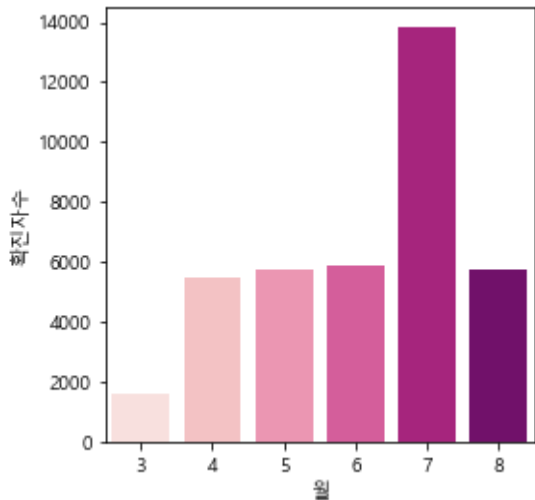
Data Analysis(1) - 크롤링 데이터

2. 국가별 해외유입 환자



Data Analysis(1) - 크롤링 데이터

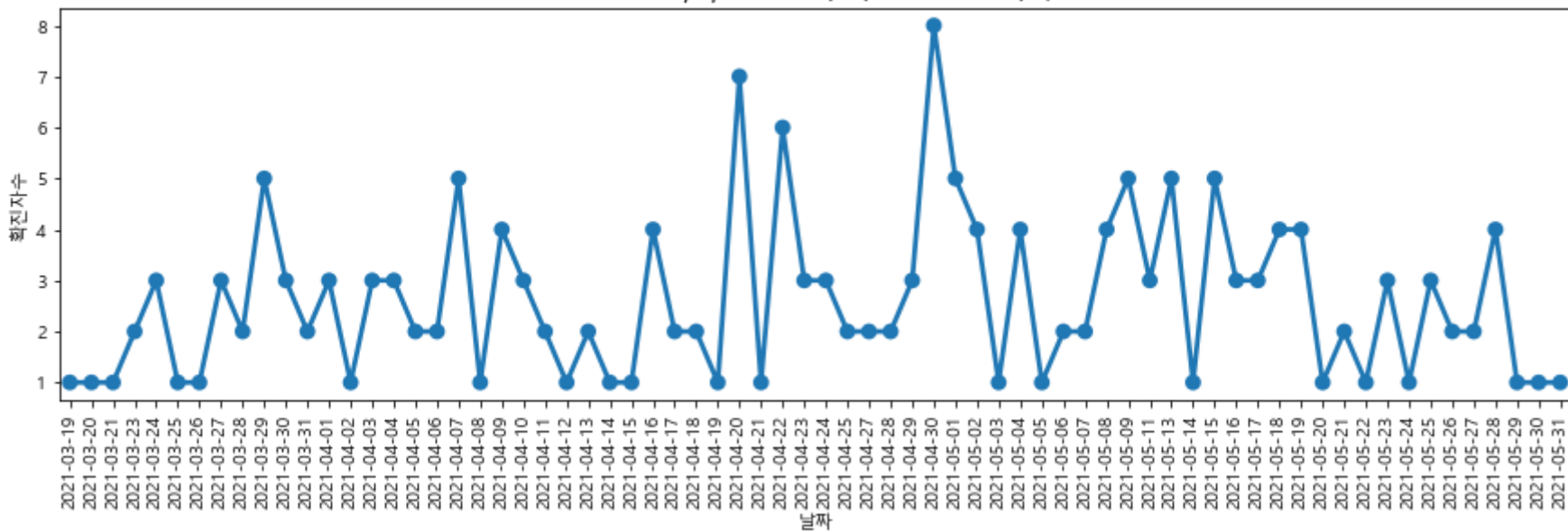
3. 날짜별 감염자 수



Data Analysis(1) - 크롤링 데이터

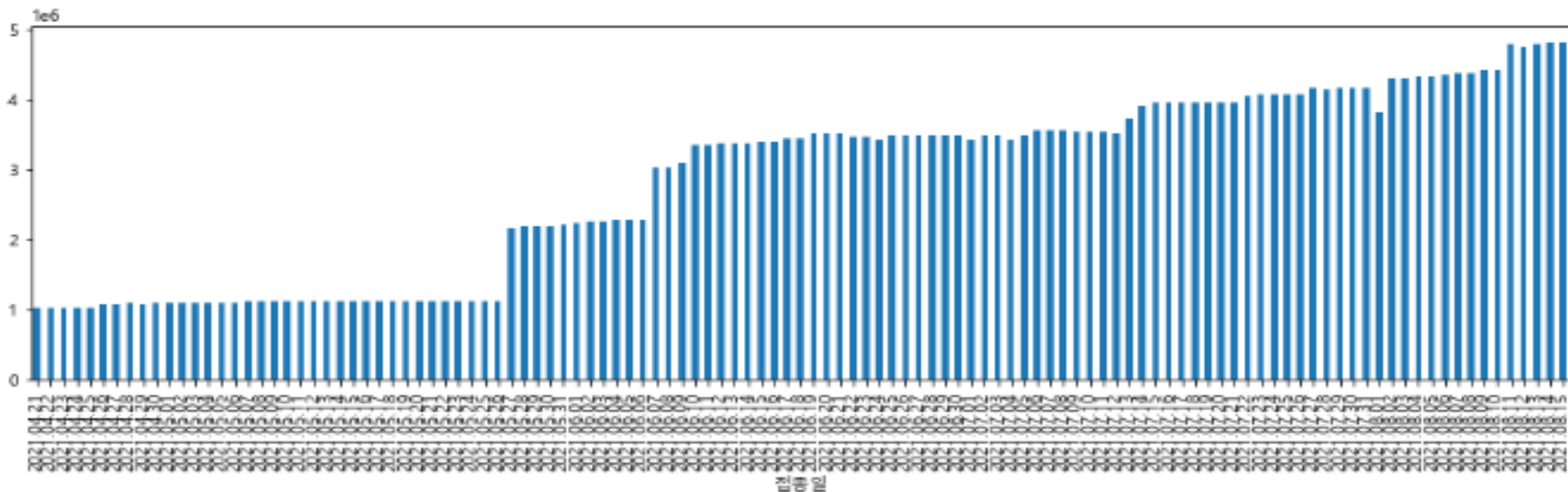
4. 날짜별 해외유입 환자수

2021. 3,4,5 월 해외유입 환자수



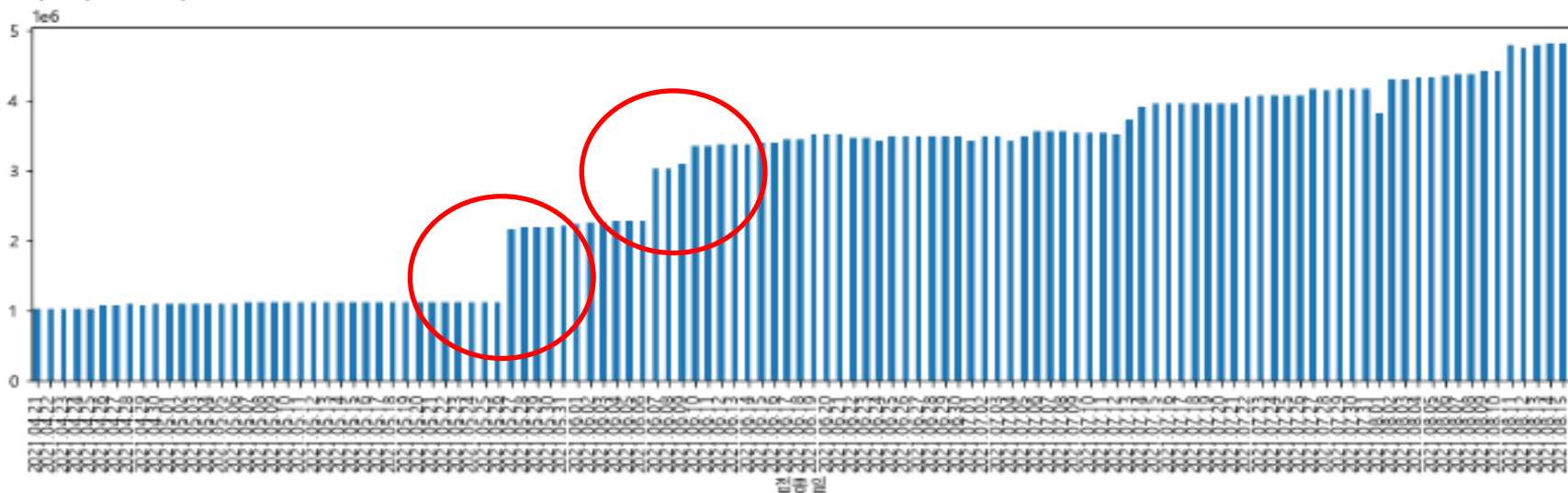
Data Analysis(2) – 예방접종 데이터

접종일별 예방접종 대상자 확인



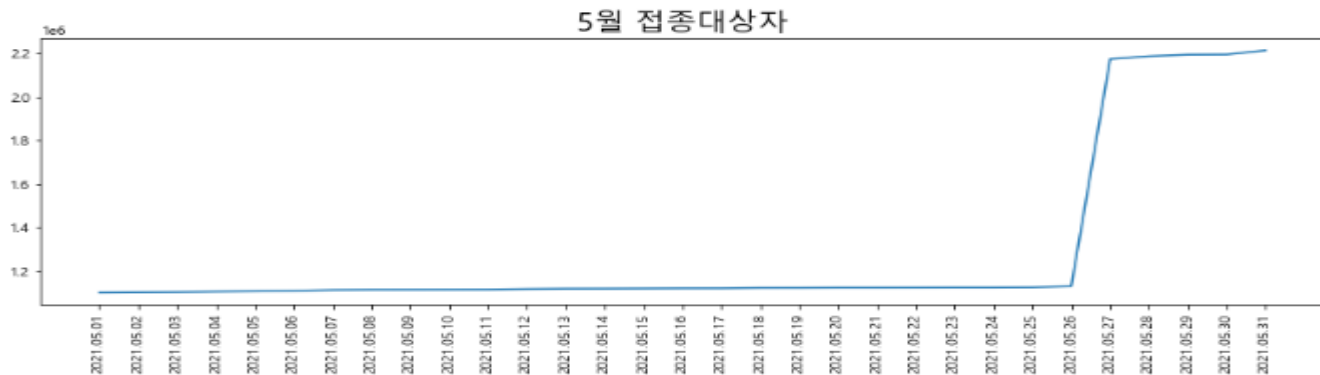
Data Analysis(2) – 예방접종 데이터

특이점 확인



Data Analysis(2) – 예방접종 데이터

특이점 1

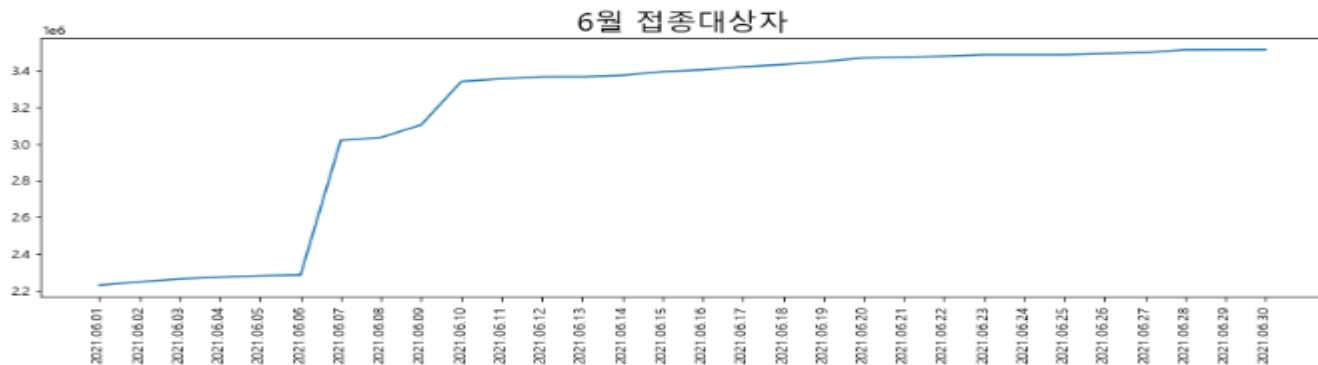


65세~74세 고령층을 대상으로 한 신종 코로나바이러스 감염증(코로나19) 백신 대규모 접종이 27일 시작됐다. 여기에 30세 이상이라면 네이버·카카오 등을 활용한 잔여백신(노쇼백신) 당일 예약 백신 접종도 가능해지면서 사실상 일반인 접종에 돌입했다는 평가다.

이날 오후 6시 기준 신규 1차 접종자는 64만명을 기록하면서 일일 접종 최다 기록을 경신했다. 이에 따라 28일 0시 기준 1차 누적 접종자는 전국민 대비 9.1%를 넘길 전망이다. 첫날 접종이 이같이 물리면서 이날 함께 실시된 '잔여백신 당일예약'은 하늘의 별 따기가 됐다.

Data Analysis(2) – 예방접종 데이터

특이점 2



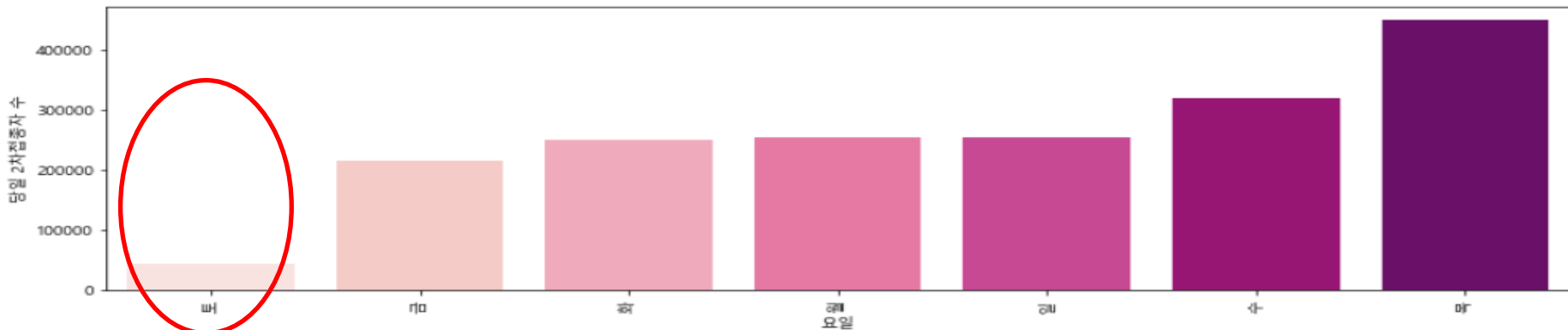
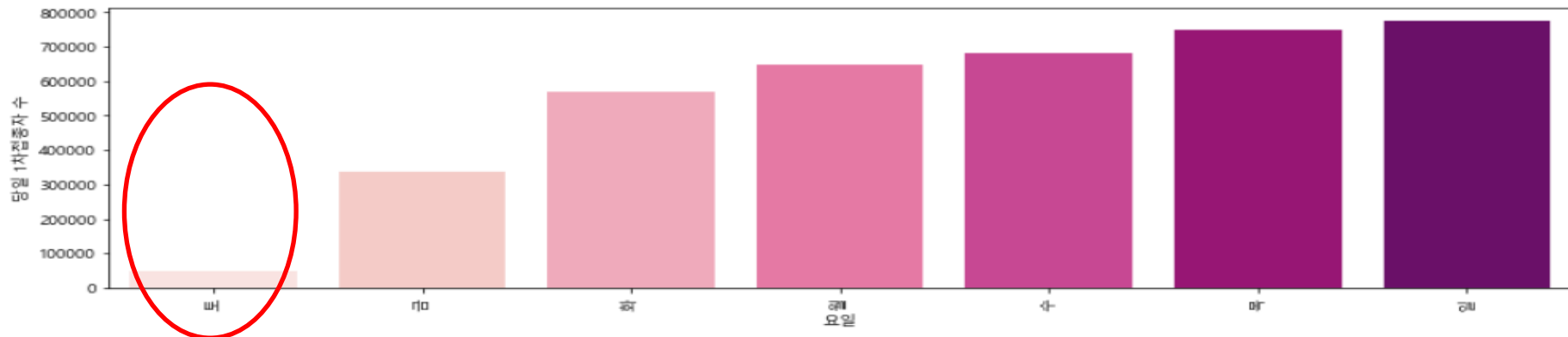
오늘(7일)부터 군 30세 미만 장병들을 대상으로 코로나19 백신 예방접종이 시작됐다. 미국 화이자 백신으로 1·2차 접종이 이뤄진다.

국방부는 군내 30세 미만 장병·군무원 등 41만4000명 중 접종에 동의한 35만8000명(3일 기준, 동의률 86.5%)을 대상으로 다음달 중순까지 6주 간 접종을 진행한다고 7일 발표했다. 군 병원과 사단급의무대 등 91개 군 접종기관에서 이뤄진다.

또 이날 하루 접종건수는 총 85만7000건(1차 83만8000건, 2차 2만건)으로 지난 5월 27일 접종 71만6000건(1차 66만2000건, 2차 5만5000건)보다 많은 일일 최다 규모를 기록했다. 1차 누적 접종자도 800만명을 초과해 인구 대비 접종률 16.4%를 달성했다.

Data Analysis(2) – 예방접종 데이터

요일별 1차, 2차 접종자 수



Data Analysis(2) – 예방접종 데이터

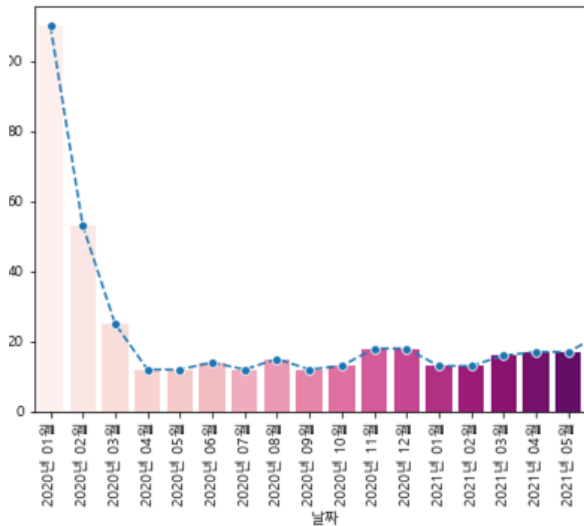
요일별 1차, 2차 접종자 수



Data Analysis(3) - 여행사업 데이터

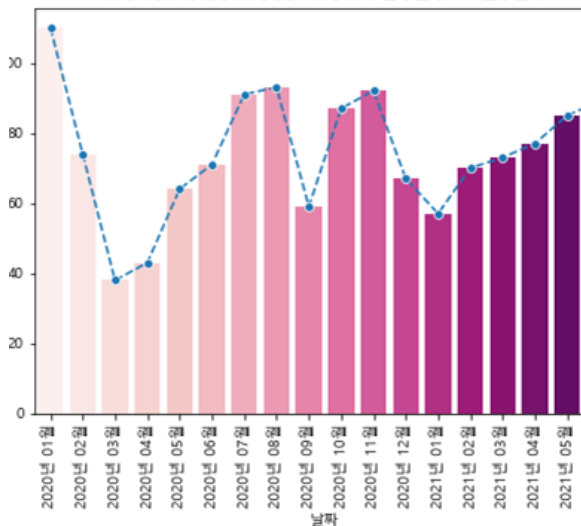
1. 여행사업 추이

travel business from 2020 - 2021



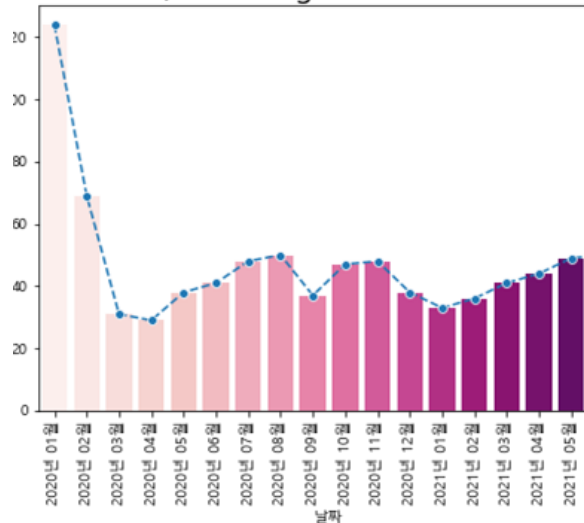
2. 호텔사업 추이

hotel business from 2020 - 2021



3. 관광/여행 지수 추이

tourism/travel degree from 2020 - 2021



Comprehensive Insight

Insight(1)

1. 날짜별 해외유입 환자수 선그래프를 살펴보면, 4.27일을 기점으로 해외유입 환자수가 증가함.
2. 여행사업 데이터의 여행사업과 관광/여행 지수를 살펴보면 4월부터 상승함. 이 사실을 통해서 사람들이 4월부터 여행을 많이 다님을 알 수 있으며 여행을 많이 다니기 때문에 해외유입 환자수가 증가했음을 유추 가능.
3. 4.27일을 기점으로 하여 국내 확진자 수를 보니, 급격한 상승세를 보인다. 이를 통해서 사람들이 확진자가 많이 나오더라도 상관없이 여행을 많이 다님을 알 수 있다.

코로나 이후 여행 사업은 침체가 되었지만, 시간이 장기화 됨에 따라서 확진자수가 많이 나옴에도 불구하고 여행에 대한 사람들의 수요가 높아짐.

따라서 현재 여행사업이 우려와는 달리 심각하게 침체될 것으로 보이지 않음.

Insight(2)

1. 코로나19 예방접종 관련 데이터를 살펴보면, 예방접종을 받으려는 대상자가 시간이 지남에 따라 지속적으로 증가하고 있음을 확인할 수 있다.
2. 또한, 예방접종 데이터를 통해서 토요일에 가장 적은 접종자 수가 발생한다는 것을 확인할 수 있으며 이를 통해서 주말에 여가시간을 보내는 인원이 많다는 것을 유추할 수 있다.
3. 결론적으로, 코로나 19 예방접종 대상자가 지속적으로 증가하고 있으며, 토요일에 접종자가 가장 적은 것으로 보아, 주말을 이용하여 여행을 가는 인원이 많아질 것이다.

예방접종 대상자가 증가함에 따라 사람들이 갖는 감염병에 관한 두려움이 감소할 것으로 예상됨

또한 토요일에 접종자가 가장 적은 것으로 보아 코로나19와 상관없이 주말을 이용해 여가시간을 보내는 인원이 많은 것으로 예상됨

따라서 시간이 지남에 따라 코로나19 창궐 이전처럼 주말에 국내여행을 즐기는 사람들이 많아질 것으로 예상된다.

감사합니다
Thank You :)