

두개의 데이터 셋을 합치고 처리하기

데이터 참고 사이트

- our world in data
 - <https://ourworldindata.org/covid-vaccinations> (<https://ourworldindata.org/covid-vaccinations>)
- bloomberg.com
 - <https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/> (<https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/>)
- 코로나 19(COVID-19) 실시간 상황판
 - <https://coronaboard.kr/> (<https://coronaboard.kr/>)

In [81]:



```
from IPython.display import display, Image
import os, warnings
import re
warnings.filterwarnings(action='ignore')
```

01 파일 확인 및 불러오기

In [82]:



```
os.listdir()
```

Out[82]:

```
[ '.git',
  '.idea',
  '.ipynb_checkpoints',
  '01_dataScraping_corona01_today.ipynb',
  '01_dataScraping_corona02.py',
  '01_dataScraping_corona02_all.ipynb',
  '02_dataScraping_bloomberg.ipynb',
  '02_dataScraping_bloomberg.py',
  '03_corona_vaccine_merge.ipynb',
  '03_corona_vaccine_merge.py',
  '04_analysis_corr.ipynb',
  '05_people_request.ipynb',
  '06_corona_analysis.ipynb',
  '20210921_00_datamerge.csv',
  '20210921_00_datamerge.xlsx',
  '20210929_chrome_driver_01.ipynb',
  '2021_0911_13',
  '2021_0911_13.zip',
  '2021_0920_00',
  'chromedriver_91.exe',
  'data',
  'html_pdf',
  'project01_01',
  'README.md',
  'WEB_sel_project11_naverNews_corona_0731.ipynb']
```

In [83]:

```
path = os.getcwd() + 'WWdata'
```

In [84]:

```
for dirname, _, filenames in os.walk(path):  
    for filename in filenames:  
        print(os.path.join(dirname, filename))
```

```
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-08-07_corona.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-08-07_corona.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-09-19_corona.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-09-19_corona.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-09-20_corona.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-09-20_corona.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-09-28_corona.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-09-28_corona.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-09-29_corona.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\2021-09-29_corona.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210808_16datamerge.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210808_16datamerge.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210808_16_today_corona.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210808_16_today_corona.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210808_16_vaccine_bloomberg.csv  
v  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210808_16_vaccine_bloomberg.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210920_00_vaccine_bloomberg.csv  
v  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210920_00_vaccine_bloomberg.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210921_00_datamerge.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210921_00_datamerge.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210929_00_vaccine_bloomberg.csv  
v  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210929_00_vaccine_bloomberg.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210929_01_datamerge.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210929_01_datamerge.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210929_14_vaccine_bloomberg.csv  
v  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210929_14_vaccine_bloomberg.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210930_00_datamerge.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210930_00_datamerge.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210930_00_vaccine_bloomberg.csv  
v  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210930_00_vaccine_bloomberg.xlsx  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\country.csv  
C:\Users\Wtoto\Documents\Github\corona_analysis\data\country.xlsx
```

In [85]:



```
import pandas as pd

# excel 데이터 셋도 pd.read_excel로 가져올 수 있음.
corona = pd.read_csv("./data/2021-09-29_corona.csv")
vaccine = pd.read_csv("./data/20210930_00_vaccine_bloomberg.csv")
country_code = pd.read_excel("./data/country.xlsx")

corona.shape, vaccine.shape, country_code.shape
```

Out[85]:

((222, 15), (207, 6), (227, 4))

In [86]:



```
corona.head()
```

Out[86]:

	국가	치명률 (%)	완치률 (%)	발생률	인구수	치료중합계	치료중1일	위중중합계	위중중1일	확진자합계	확진1일
0	미국	1.6	76.1	133098	331002651	9823512.0	-14871.0	21296.0	113357.0	44055692	113357.0
1	인도	1.3	97.8	24434	1380004385	284976.0	-4806.0	8944.0	4743.0	33718984	4743.0
2	브라질	2.8	95.3	100592	212559417	403027.0	-7475.0	8318.0	15395.0	21381790	15395.0
3	영국	1.8	80.9	113959	67886011	1344510.0	-1886.0	831.0	34520.0	7736235	34520.0
4	러시아	2.8	88.9	51305	145934462	626809.0	3117.0	2300.0	22430.0	7487138	22430.0

In [87]:



```
vaccine.head()
```

Out[87]:

	국가	백신접종수	백신보유율(인구)	1차접종	2차접종	일별접종수
0	Global Total	6.199444e+09	-999.0	-999.0	-999.0	29152868.0
1	Mainland China	2.203185e+09	78.7	78.6	73.0	3171286.0
2	India	8.765386e+08	32.0	46.7	16.8	7178901.0
3	EU	5.640300e+08	63.5	67.3	64.7	1052903.0
4	U.S.	3.911526e+08	61.1	64.4	55.8	701422.0

In [88]:



```
country_code.head()
```

Out[88]:

	국문(코로나보드)	영문(블룸버그)	나라	비고_뜻
0	가나	Ghana	NaN	NaN
1	가봉	Gabon	NaN	NaN
2	가이아나	Guyana	NaN	NaN
3	감비아	Gambia	NaN	NaN
4	과들루프	NaN	프랑스	Guadeloupe

코로나에 별칭을 제외한 나라명 가져오기

In [89]:



```
corona['국가'].str.extract('([ㄱ-ㅣ가-힣]+)')
```

Out[89]:

	0
0	미국
1	인도
2	브라질
3	영국
4	러시아
...	...
217	마셜
218	바누아투
219	사모아
220	세인트헬레나
221	미크로네시아

222 rows × 1 columns

In [90]:



```
corona['한글(국가명)'] = corona['국가'].str.extract('([ㄱ-ㅣ가-힣]+)', expand=False)  
corona.columns
```

Out[90]:

```
Index(['국가', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료중1일',  
      '위중증합계',  
      '위중증1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합  
계', '완치1일', '한글(국가명)'],  
      dtype='object')
```

In [91]:



```
col = ['국가', '한글(국가명)', '치명(%)', '완치(%)', '발생률',
       '인구수', '치료중합계', '치료중1일', '위중증합계', '위중증1일',
       '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일']

new_corona = corona[col].copy()
new_corona
```

Out[91]:

	국 가	한 글 (국 가 명)	치 명 (%)	완 치 (%)	발 생 률	인 구 수	치 료 중 합 계	치 료 중 1 일	위 중 증 합 계	위 중 증 1 일	확 진 자 합 계
0	미 국 US	미 국	1.6	76.1	133098	331002651	9823512.0	-14871.0	21296.0	113357.0	44055692
1	인 도 IN	인 도	1.3	97.8	24434	1380004385	284976.0	-4806.0	8944.0	4743.0	33718984
2	브 라 질 BR	브 라 질	2.8	95.3	100592	212559417	403027.0	-7475.0	8318.0	15395.0	21381790
3	영 국 GB	영 국	1.8	80.9	113959	67886011	1344510.0	-1886.0	831.0	34520.0	7736234
4	러 시 아 RU	러 시 아	2.8	88.9	51305	145934462	626809.0	3117.0	2300.0	22430.0	7487138
...
217	마 셜 제 도 MH	마 셜	-	100.0	68	59190	0.0	NaN	NaN	NaN	4
218	바 누 아 투 VU	바 누 아 투	25	75.0	13	307145	0.0	NaN	NaN	NaN	4
219	사 모 아 WS	사 모 아	-	100.0	15	198414	0.0	NaN	NaN	NaN	4
220	세 인 트 헬 레 나 SH	세 인 트 헬 레 나	-	100.0	329	6077	0.0	NaN	NaN	NaN	4

	국가	한글 (국가명)	치명 (%)	완치 (%)	발생률	인구수	치료중합 계	치료중1 일	위중증 합계	위중증1 일	확진자합 계
221	미 크 로 네 시아 FM	미 크 로 네 시아	-	100.0	9	115023	0.0	NaN	NaN	NaN	.

222 rows × 16 columns

두 데이터 셋 연결

In [92]:



```
new_corona.head()
```

Out [92]:

국가		한글 (국가명)	치명 (%)	완치 (%)	발생률	인구수	치료중합 계	치료중1 일	위중증 합계	위중증1 일	확진자합 계	
0	미국 US	미국	1.6	76.1	133098	331002651	9823512.0	-14871.0	21296.0	113357.0	44055692	1
1	인도 IN	인도	1.3	97.8	24434	1380004385	284976.0	-4806.0	8944.0	4743.0	33718984	
2	브라질 BR	브라질	2.8	95.3	100592	212559417	403027.0	-7475.0	8318.0	15395.0	21381790	
3	영국 GB	영국	1.8	80.9	113959	67886011	1344510.0	-1886.0	831.0	34520.0	7736235	:
4	러시아 RU	러시아	2.8	88.9	51305	145934462	626809.0	3117.0	2300.0	22430.0	7487138	:

In [93]:



```
country_code.head()
```

Out[93]:

	국문(코로나보드)	영문(블룸버그)	나라	비고_뜻
0	가나	Ghana	NaN	NaN
1	가봉	Gabon	NaN	NaN
2	가이아나	Guyana	NaN	NaN
3	감비아	Gambia	NaN	NaN
4	과들루프	NaN	프랑스	Guadeloupe

데이터 병합

- 두개의 데이터 셋의 공통되는 열을 지정하여 이를 기준으로 데이터를 병합한다.

In [94]:



```
df_corona = new_corona.merge(country_code, left_on='한글(국가명)',  
                             right_on='국문(코로나보드)')  
  
df_corona.head()
```

Out[94]:

	국가	한글(국가명)	치명(%)	완치(%)	발생률	인구수	치료중합계	치료중1일	위중증합계	위중증1일	확진자합계	
0	미국	미국	1.6	76.1	133098	331002651	9823512.0	-14871.0	21296.0	113357.0	44055692	1
1	인도	인도	1.3	97.8	24434	1380004385	284976.0	-4806.0	8944.0	4743.0	33718984	
2	브라질	브라질	2.8	95.3	100592	212559417	403027.0	-7475.0	8318.0	15395.0	21381790	
3	영국	영국	1.8	80.9	113959	67886011	1344510.0	-1886.0	831.0	34520.0	7736235	:
4	러시아	러시아	2.8	88.9	51305	145934462	626809.0	3117.0	2300.0	22430.0	7487138	:

In [95]:



```
df_corona.drop(['국가', '국문(코로나보드)', '나라', '비고_뜻'], axis=1, inplace=True)
df_corona.head()
```

Out [95]:

	한글 (국가명)	치명 (%)	완치 (%)	발생률	인구수	치료중합 계	치료중1 일	위중중 합계	위중중1 일	확진자합 계	확진1 일
0	미국	1.6	76.1	133098	331002651	9823512.0	-14871.0	21296.0	113357.0	44055692	11335
1	인도	1.3	97.8	24434	1380004385	284976.0	-4806.0	8944.0	4743.0	33718984	474
2	브라질	2.8	95.3	100592	212559417	403027.0	-7475.0	8318.0	15395.0	21381790	1539
3	영국	1.8	80.9	113959	67886011	1344510.0	-1886.0	831.0	34520.0	7736235	3452
4	러시아	2.8	88.9	51305	145934462	626809.0	3117.0	2300.0	22430.0	7487138	2243

In [96]:



```
vaccine.head()
```

Out [96]:

	국가	백신접종수	백신보유율(인구)	1차접종	2차접종	일별접종수
0	Global Total	6.199444e+09	-999.0	-999.0	-999.0	29152868.0
1	Mainland China	2.203185e+09	78.7	78.6	73.0	3171286.0
2	India	8.765386e+08	32.0	46.7	16.8	7178901.0
3	EU	5.640300e+08	63.5	67.3	64.7	1052903.0
4	U.S.	3.911526e+08	61.1	64.4	55.8	701422.0

두 개의 데이터 셋을 이용하여 병합

- 공통 열 이름 : '영문(블룸버그)', '국가'

In [97]:



```
df_corona_all = df_corona.merge(vaccine, left_on='영문(블룸버그)', right_on='국가')
df_corona_all.head()
```

Out[97]:

	한글 (국가명)	치명 (%)	완치 (%)	발생률	인구수	치료중합 계	치료중1 일	위중중 합계	위중중1 일	확진자합 계	...	시
0	미국	1.6	76.1	133098	331002651	9823512.0	-14871.0	21296.0	113357.0	44055692	...	2
1	인도	1.3	97.8	24434	1380004385	284976.0	-4806.0	8944.0	4743.0	33718984	...	1
2	브라질	2.8	95.3	100592	212559417	403027.0	-7475.0	8318.0	15395.0	21381790	...	8
3	영국	1.8	80.9	113959	67886011	1344510.0	-1886.0	831.0	34520.0	7736235	...	1
4	러시아	2.8	88.9	51305	145934462	626809.0	3117.0	2300.0	22430.0	7487138	...	8

5 rows × 22 columns

In [98]:



```
df_corona_all.drop(['국가'], axis=1, inplace=True)
df_corona_all
```

Out[98]:

	한글 (국가명)	치명 (%)	완치 (%)	발생률	인구수	치료중합 계	치료중1 일	위중중 합계	위중중1 일	확진자합 계	...
0	미국	1.6	76.1	133098	331002651	9823512.0	-14871.0	21296.0	113357.0	44055692	...
1	인도	1.3	97.8	24434	1380004385	284976.0	-4806.0	8944.0	4743.0	33718984	...
2	브라질	2.8	95.3	100592	212559417	403027.0	-7475.0	8318.0	15395.0	21381790	...
3	영국	1.8	80.9	113959	67886011	1344510.0	-1886.0	831.0	34520.0	7736235	...
4	러시아	2.8	88.9	51305	145934462	626809.0	3117.0	2300.0	22430.0	7487138	...
...
173	마카오	-	88.7	109	649335	8.0	1.0	NaN	1.0	71	...
174	몬트세랫	3.0	90.9	6611	4992	2.0	NaN	NaN	NaN	33	...
175	사모아	-	100.0	15	198414	0.0	NaN	NaN	NaN	3	...
176	세인트헬레나	-	100.0	329	6077	0.0	NaN	NaN	NaN	2	...
177	미크로네시아	-	100.0	9	115023	0.0	NaN	NaN	NaN	1	...

178 rows × 21 columns

In [99]:



```
df_corona_all.columns
```

Out[99]:

```
Index(['한글(국가명)', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료  
중1일', '위중증합계',  
      '위중증1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합  
계', '완치1일', '영문(블룸버그)',  
      '백신접종수', '백신보유율(인구)', '1차접종', '2차접종', '일별접종수'],  
      dtype='object')
```

In [100]:



```
sel = ['한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '일별접종수', '치명(%)', '완치(%)',
        '치료중합계', '치료중1일', '위중증합계', '위중증1일', '확진자합계', '확진자1일',
        '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종수',
        '백신보유율(인구)']
df_corona_all_n = df_corona_all[sel].copy()
df_corona_all_n
```

Out[100]:

	한글(국가명)	영문(블룸버그)	1차접종	2차접종	일별접종수	치명(%)	완치(%)	발생률	인구수	치료중합계
0	미국	U.S.	64.4	55.8	701422.0	1.6	76.1	133098	331002651	9823512.0	...	2
1	인도	India	46.7	16.8	7178901.0	1.3	97.8	24434	1380004385	284976.0	...	{
2	브라질	Brazil	71.3	41.9	1317163.0	2.8	95.3	100592	212559417	403027.0	...	{
3	영국	U.K.	73.0	67.1	63126.0	1.8	80.9	113959	67886011	1344510.0	...	
4	러시아	Russia	32.6	28.5	214966.0	2.8	88.9	51305	145934462	626809.0	...	4
...	
173	마카오	Macau	51.2	44.6	4123.0	-	88.7	109	649335	8.0	...	
174	몬트세랫	Montserrat	29.7	27.8	2.0	3.0	90.9	6611	4992	2.0	...	
175	사모아	Samoa	49.3	25.0	981.0	-	100.0	15	198414	0.0	...	
176	세인트헬레나	St. Helena	72.7	58.9	8.0	-	100.0	329	6077	0.0	...	
177	미크로네시아	Micronesia	41.6	-999.0	266.0	-	100.0	9	115023	0.0	...	

178 rows × 21 columns

In [101]:



```
df_corona_all_n.columns
```

Out[101]:

```
Index(['한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '일별접종수', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료중1일', '위중증합계', '위중증1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종수', '백신보유율(인구)'], dtype='object')
```

In [102]:



```
df_corona_all_n.sort_values(['2차접종'], ascending=False)
```

Out[102]:

	한글 (국가명)	영문(블룸버그)	1차접종	2차접종	일별접종수	치명(%)	완치(%)	발생률	인구수	치료중합계	...	위중증합계
161	지브롤터	Gibraltar	117.4	116.40	30.0	1.8	97.1	164198	33691	62.0	...	2.0
110	몰디브	Maldives	105.3	90.30	759.0	0.3	97.9	156470	540544	1579.0	...	23.0
33	포르투갈	Portugal	86.9	84.00	23657.0	1.7	95.5	104792	10196709	30495.0	...	68.0
81	카타르	Qatar	84.4	79.40	4415.0	0.3	99.2	82108	2881053	1282.0	...	15.0
108	싱가포르	Singapore	82.1	79.20	23733.0	0.1	83.1	15687	5850342	15469.0	...	30.0
...
149	남수단	South Sudan	0.6	0.20	996.0	1.1	96.6	1070	11193725	283.0	...	4.0
130	베냉	Benin	1.3	0.20	1696.0	0.7	75.3	1894	12123200	5510.0	...	5.0
153	예멘	Yemen	1.0	0.05	244.0	19.0	62.1	303	29825964	1707.0	...	23.0
116	콩고민주공화국	DR Congo	0.1	0.05	432.0	1.9	54.3	635	89561403	24920.0	...	NaN
177	미크로네시아	Micronesia	41.6	-999.00	266.0	-	100.0	9	115023	0.0	...	NaN

178 rows × 21 columns

In [103]:



```
df_corona_all_n.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 178 entries, 0 to 177
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   한글(국가명)          178 non-null    object
1   영문(블룸버그)        178 non-null    object
2   1차접종                178 non-null    float64
3   2차접종                178 non-null    float64
4   일별접종수            178 non-null    float64
5   치명(%)               178 non-null    object
6   완치(%)               175 non-null    float64
7   발생률                178 non-null    object
8   인구수                178 non-null    object
9   치료종합계            175 non-null    float64
10  치료중1일             145 non-null    float64
11  위중증합계            131 non-null    float64
12  위중증1일             146 non-null    float64
13  확진자합계           178 non-null    int64
14  확진자1일            146 non-null    float64
15  사망자합계            178 non-null    int64
16  사망자1일            114 non-null    float64
17  완치합계              175 non-null    float64
18  완치1일              130 non-null    float64
19  백신접종수          178 non-null    float64
20  백신보유율(인구)    178 non-null    float64
dtypes: float64(14), int64(2), object(5)
memory usage: 30.6+ KB
```

In [104]:



```
df_corona_all_n['발생률'] = df_corona_all_n['발생률'].astype(int)
df_corona_all_n.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 178 entries, 0 to 177
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  --
 0   한글(국가명)          178 non-null    object
 1   영문(블룸버그)        178 non-null    object
 2   1차접종                178 non-null    float64
 3   2차접종                178 non-null    float64
 4   일별접종수            178 non-null    float64
 5   치명(%)               178 non-null    object
 6   완치(%)               175 non-null    float64
 7   발생률                178 non-null    int32
 8   인구수                178 non-null    object
 9   치료중합계            175 non-null    float64
10   치료중1일             145 non-null    float64
11   위중증합계            131 non-null    float64
12   위중증1일             146 non-null    float64
13   확진자합계            178 non-null    int64
14   확진자1일             146 non-null    float64
15   사망자합계            178 non-null    int64
16   사망자1일             114 non-null    float64
17   완치합계              175 non-null    float64
18   완치1일               130 non-null    float64
19   백신접종수           178 non-null    float64
20   백신보유율(인구)     178 non-null    float64
dtypes: float64(14), int32(1), int64(2), object(4)
memory usage: 29.9+ KB
```

In [105]:



```
df_corona_all_n['발생률순위'] = df_corona_all_n['발생률'].rank(ascending=False) # ascending=False :
df_corona_all_n['2차접종률순위'] = df_corona_all_n['2차접종'].rank(ascending=False) # ascending=False :
df_corona_all_n
```

Out[105]:

	한글 (국가명)	영문(블룸버그)	1차 접종	2차접 종	일별접종 수	치 명 (%)	완치 (%)	발생률	인구수	치료종합 계
0	미국	U.S.	64.4	55.8	701422.0	1.6	76.1	133098	331002651	9823512.0	...	44
1	인도	India	46.7	16.8	7178901.0	1.3	97.8	24434	1380004385	284976.0	...	36
2	브라질	Brazil	71.3	41.9	1317163.0	2.8	95.3	100592	212559417	403027.0	...	27
3	영국	U.K.	73.0	67.1	63126.0	1.8	80.9	113959	67886011	1344510.0	...	17
4	러시아	Russia	32.6	28.5	214966.0	2.8	88.9	51305	145934462	626809.0	...	11
...
173	마카오	Macau	51.2	44.6	4123.0	-	88.7	109	649335	8.0
174	몬트세랫	Montserrat	29.7	27.8	2.0	3.0	90.9	6611	4992	2.0
175	사모아	Samoa	49.3	25.0	981.0	-	100.0	15	198414	0.0
176	세인트헬레나	St. Helena	72.7	58.9	8.0	-	100.0	329	6077	0.0
177	미크로네시아	Micronesia	41.6	-999.0	266.0	-	100.0	9	115023	0.0

178 rows × 23 columns

In [106]:



```
df_corona_all_n.sort_values(['2차접종률순위'], ascending=True)
```

Out[106]:

	한글 (국가명)	영문(블룸버그)	1차접종	2차접종	일별접종수	치명(%)	완치(%)	발생률	인구수	치료증합계	...	확진률
161	지브롤터	Gibraltar	117.4	116.40	30.0	1.8	97.1	164198	33691	62.0	...	5%
110	몰디브	Maldives	105.3	90.30	759.0	0.3	97.9	156470	540544	1579.0	...	84%
33	포르투갈	Portugal	86.9	84.00	23657.0	1.7	95.5	104792	10196709	30495.0	...	1068%
81	카타르	Qatar	84.4	79.40	4415.0	0.3	99.2	82108	2881053	1282.0	...	236%
108	싱가포르	Singapore	82.1	79.20	23733.0	0.1	83.1	15687	5850342	15469.0	...	91%
...
149	남수단	South Sudan	0.6	0.20	996.0	1.1	96.6	1070	11193725	283.0	...	11%
130	베냉	Benin	1.3	0.20	1696.0	0.7	75.3	1894	12123200	5510.0	...	22%
153	예멘	Yemen	1.0	0.05	244.0	19.0	62.1	303	29825964	1707.0	...	9%
116	콩고민주공화국	DR Congo	0.1	0.05	432.0	1.9	54.3	635	89561403	24920.0	...	56%
177	미크로네시아	Micronesia	41.6	-999.00	266.0	-	100.0	9	115023	0.0

178 rows × 23 columns

In [107]:



```
df_corona_all_n.columns
```

Out [107]:

```
Index(['한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '일별접종수', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료중1일', '위중증합계', '위중증1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종수', '백신보유율(인구)', '발생률순위', '2차접종률순위'], dtype='object')
```

In [108]:



```
df_corona_all_n.columns
```

Out [108]:

```
Index(['한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '일별접종수', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료중1일', '위중증합계', '위중증1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종수', '백신보유율(인구)', '발생률순위', '2차접종률순위'], dtype='object')
```

In [109]:



```
sel = [ '한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '발생률순위', '2차접종률순위',  
        '일별접종수', '치명(%)', '완치(%)', '발생률',  
        '인구수', '치료중합계', '치료중1일', '위중증합계', '위중증1일', '확진자합계', '확진자1일',  
        '사망자1일', '완치합계', '완치1일', '백신접종수', '백신보유율(인구)']
```

```
df_corona_all_n = df_corona_all_n[sel]
```

```
df_corona_all_n
```

Out [109]:

	한글 (국가명)	영문(블룸 버그)	1차 접종	2차접 종	발생 률순 위	2차접 종률 순위	일별접종 수	치 명 (%)	완치 (%)	발생률	...	위중증 합계	우
0	미국	U.S.	64.4	55.8	13.0	52.5	701422.0	1.6	76.1	133098	...	21296.0	115
1	인도	India	46.7	16.8	103.0	119.0	7178901.0	1.3	97.8	24434	...	8944.0	4
2	브라질	Brazil	71.3	41.9	29.0	70.5	1317163.0	2.8	95.3	100592	...	8318.0	15
3	영국	U.K.	73.0	67.1	20.0	24.0	63126.0	1.8	80.9	113959	...	831.0	34
4	러시아	Russia	32.6	28.5	77.0	89.0	214966.0	2.8	88.9	51305	...	2300.0	25
...
173	마카오	Macau	51.2	44.6	175.0	66.0	4123.0	-	88.7	109	...	NaN	
174	몬트세랫	Montserrat	29.7	27.8	128.0	94.0	2.0	3.0	90.9	6611	...	NaN	
175	사모아	Samoa	49.3	25.0	177.0	99.0	981.0	-	100.0	15	...	NaN	
176	세인트헬레나	St. Helena	72.7	58.9	171.0	42.0	8.0	-	100.0	329	...	NaN	
177	미크로네시아	Micronesia	41.6	-999.0	178.0	178.0	266.0	-	100.0	9	...	NaN	

178 rows × 23 columns

In [110]:



```
df_corona_all_n.sort_values(['발생률순위'])
```

Out[110]:

	한글 (국가명)	영문(블룸버그)	1차 접종	2차접 종	발생 률순 위	2차접 종순 위	일별접 종수	치 명 (%)	완치 (%)	발생률	...	위중 합계	위중 증1 일
131	세이셸	Seychelles	80.5	74.1	1.0	14.5	154.0	0.5	97.9	217058	...	NaN	NaN
96	몬테네그로	Montenegro	38.6	33.7	2.0	83.0	2132.0	1.5	93.0	207781	...	NaN	NaN
141	안도라	Andorra	68.7	55.9	3.0	51.0	113.0	0.9	98.6	196622	...	4.0	3.0
161	지브롤터	Gibraltar	117.4	116.4	4.0	1.0	30.0	1.8	97.1	164198	...	2.0	NaN
78	바레인	Bahrain	78.4	75.0	5.0	11.0	3999.0	0.5	99.2	161571	...	5.0	48.0
...
159	니제르	Niger	1.7	0.4	174.0	170.5	9294.0	3.4	95.9	248	...	NaN	6.0
173	마카오	Macau	51.2	44.6	175.0	66.0	4123.0	-	88.7	109	...	NaN	1.0
107	중국	Mainland China	78.6	73.0	176.0	18.0	3171286.0	4.8	94.2	67	...	NaN	25.0
175	사모아	Samoa	49.3	25.0	177.0	99.0	981.0	-	100.0	15	...	NaN	NaN
177	미크로네시아	Micronesia	41.6	-999.0	178.0	178.0	266.0	-	100.0	9	...	NaN	NaN

178 rows × 23 columns

인구당 확진자 비율

In [111]:



```
df_corona_all_n['확진자비율(인구)'] = df_corona_all_n['확진자합계'].astype(int)/df_corona_all_n['인구']
df_corona_all_n
```

Out[111]:

	한글 (국가명)	영문(블룸버그)	1차 접종	2차접 종	발생 률순 위	2차접 종률 순위	일별접종 수	치 명 (%)	완치 (%)	발생률	...	위중증1 일	3
0	미국	U.S.	64.4	55.8	13.0	52.5	701422.0	1.6	76.1	133098	...	113357.0	4
1	인도	India	46.7	16.8	103.0	119.0	7178901.0	1.3	97.8	24434	...	4743.0	3
2	브라질	Brazil	71.3	41.9	29.0	70.5	1317163.0	2.8	95.3	100592	...	15395.0	2
3	영국	U.K.	73.0	67.1	20.0	24.0	63126.0	1.8	80.9	113959	...	34520.0	1
4	러시아	Russia	32.6	28.5	77.0	89.0	214966.0	2.8	88.9	51305	...	22430.0	1
...
173	마카오	Macau	51.2	44.6	175.0	66.0	4123.0	-	88.7	109	...	1.0	
174	몬트세랫	Montserrat	29.7	27.8	128.0	94.0	2.0	3.0	90.9	6611	...	NaN	
175	사모아	Samoa	49.3	25.0	177.0	99.0	981.0	-	100.0	15	...	NaN	
176	세인트헬레나	St. Helena	72.7	58.9	171.0	42.0	8.0	-	100.0	329	...	NaN	
177	미크로네시아	Micronesia	41.6	-999.0	178.0	178.0	266.0	-	100.0	9	...	NaN	

178 rows × 24 columns

In [112]:



```
df_corona_all_n.columns
```

Out[112]:

```
Index(['한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '발생률순위', '2차접  
종률순위', '일별접종수',  
      '치명(%)', '완치(%)', '발생률', '인구수', '치료종합계', '치료중1일', '위중증  
합계', '위중증1일',  
      '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일',  
      '백신접종수',  
      '백신보유율(인구)', '확진자비율(인구)'],  
      dtype='object')
```

In [113]:



```
sel = ['한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '발생률순위', '2차접종률순위', '확진자  
일별접종수', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료중1일', '위중증합계',  
'확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종수',  
'백신보유율(인구)']  
df_corona_all_n = df_corona_all_n[sel]  
df_corona_all_n
```

Out[113]:

	한글(국가명)	영문(블룸버그)	1차접종	2차접종	발생률순위	2차접종률순위	확진자비율(인구)	일별접종수	치명(%)	완치(%)	...	위중증합계
0	미국	U.S.	64.4	55.8	13.0	52.5	0.133098	701422.0	1.6	76.1	...	21296.0
1	인도	India	46.7	16.8	103.0	119.0	0.024434	7178901.0	1.3	97.8	...	8944.0
2	브라질	Brazil	71.3	41.9	29.0	70.5	0.100592	1317163.0	2.8	95.3	...	8318.0
3	영국	U.K.	73.0	67.1	20.0	24.0	0.113959	63126.0	1.8	80.9	...	831.0
4	러시아	Russia	32.6	28.5	77.0	89.0	0.051305	214966.0	2.8	88.9	...	2300.0
...
173	마카오	Macau	51.2	44.6	175.0	66.0	0.000109	4123.0	-	88.7	...	NaN
174	몬트세랫	Montserrat	29.7	27.8	128.0	94.0	0.006611	2.0	3.0	90.9	...	NaN
175	사모아	Samoa	49.3	25.0	177.0	99.0	0.000015	981.0	-	100.0	...	NaN
176	세인트헬레나	St. Helena	72.7	58.9	171.0	42.0	0.000329	8.0	-	100.0	...	NaN
177	미크로네시아	Micronesia	41.6	-999.0	178.0	178.0	0.000009	266.0	-	100.0	...	NaN

178 rows × 24 columns

파일 생성

In [114]:



```
from datetime import datetime
import os

now = datetime.now()
file_make_time = "%04d%02d%02d_%02d" % (now.year, now.month, now.day, now.hour)
print( file_make_time )

path_dir = os.getcwd() + "WWdataWW"
path_file = path_dir + file_make_time
print( path_dir, path_file, sep="\n" )
```

20210930_00

C:\Users\wtoto\Documents\Github\corona_analysis\data\

C:\Users\wtoto\Documents\Github\corona_analysis\data\20210930_00

In [115]:



```
df_corona_all_n.to_csv(path_file + "_datamerge.csv", index=False)
df_corona_all_n.to_excel(path_file + "_datamerge.xlsx", index=False)
```

In [116]:



```
import seaborn as sns
```

백신 접종률과 인구당 확진자 비율의 상관관계

In [117]:



```
from matplotlib import font_manager, rc
import matplotlib.pyplot as plt
import platform
import matplotlib
```

In [118]:

```
path = "C:/Windows/Fonts/malgun.ttf"
if platform.system() == "Windows":
    font_name = font_manager.FontProperties(fname=path).get_name()
    rc('font', family=font_name)
elif platform.system()=="Darwin":
    rc('font', family='AppleGothic')
else:
    print("Unknown System")

matplotlib.rcParams['axes.unicode_minus'] = False
```

In [119]:

```
df_corona_all_n.columns
```

Out[119]:

```
Index(['한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '발생률순위', '2차접종률순위', '확진자비율(인구)', '일별접종수', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료중1일', '위중중합계', '위중중1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종수', '백신보유율(인구)'], dtype='object')
```

In [120]:

```
df_corona_all_n.head()
```

Out[120]:

	한글(국가명)	영문(블룸버그)	1차접종	2차접종	발생률순위	2차접종률순위	확진자비율(인구)	일별접종수	치명(%)	완치(%)	...	위중중합계	위중중1일
0	미국	U.S.	64.4	55.8	13.0	52.5	0.133098	701422.0	1.6	76.1	...	21296.0	113357.0
1	인도	India	46.7	16.8	103.0	119.0	0.024434	7178901.0	1.3	97.8	...	8944.0	4743.0
2	브라질	Brazil	71.3	41.9	29.0	70.5	0.100592	1317163.0	2.8	95.3	...	8318.0	15395.0
3	영국	U.K.	73.0	67.1	20.0	24.0	0.113959	63126.0	1.8	80.9	...	831.0	34520.0
4	러시아	Russia	32.6	28.5	77.0	89.0	0.051305	214966.0	2.8	88.9	...	2300.0	22430.0

5 rows × 24 columns

In [121]:



```
df_corona_all_n['2차접종'].describe()
```

Out [121]:

```
count      178.000000
mean        28.392135
std         81.917729
min        -999.000000
25%          9.000000
50%        28.350000
75%        57.700000
max        116.400000
Name: 2차접종, dtype: float64
```

확진자 비율과 백신 접종률 상관관계 확인해 보기

In [122]:



```
df_corona_all_n = df_corona_all_n[df_corona_all_n['2차접종'] > 0 ]
```

In [123]:



```
df_corona_all_n[ ['2차접종', '확진자비율(인구)' ]].corr()
```

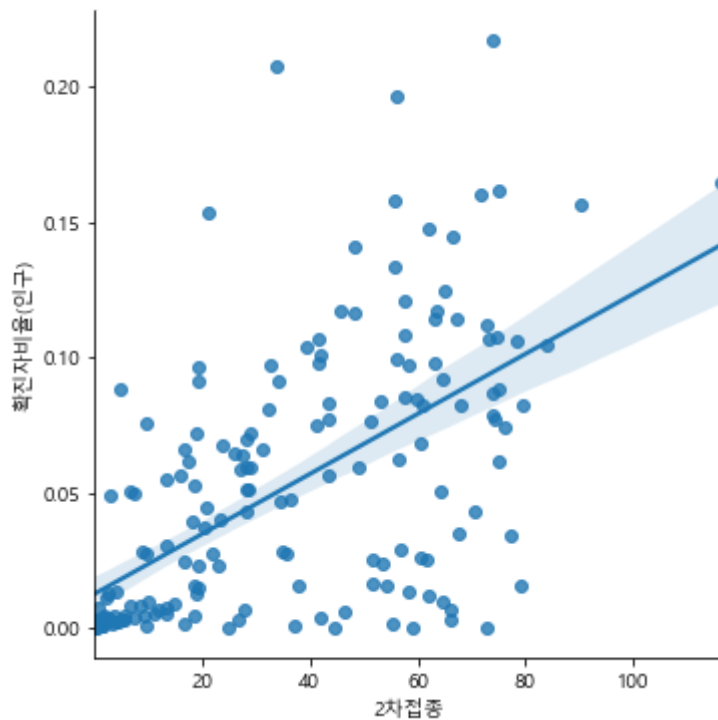
Out [123]:

	2차접종	확진자비율(인구)
2차접종	1.00000	0.59704
확진자비율(인구)	0.59704	1.00000

In [124]:



```
sns.lmplot(x='2차접종', y='확진자비율(인구)', data=df_corona_all_n)
plt.show()
```

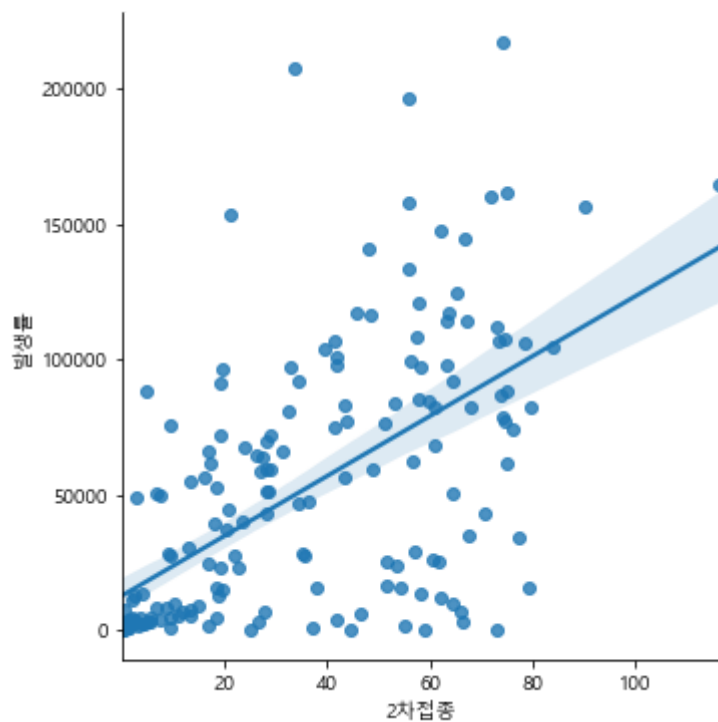


인구당 확진자 발생 비율과 백신 접종률 상관관계 그래프 그려보기

In [125]:



```
sns.lmplot(x='2차접종', y='발생률', data=df_corona_all_n)
plt.show()
```



In [126]:



```
df_corona_all_n.corr()['2차접종']['발생률']
```

Out[126]:

0.5970275728730123

In [127]:



```
df_corona_all_n.corr()['2차접종']['확진자비율(인구)']
```

Out[127]:

0.5970396083176327

01 현재까지의 누적된 데이터로 확인 결과 백신 2차 접종률과 인구당 확진자 비율은 양의 상관관계를 갖는다.

02 현재까지의 누적된 데이터로 확인 결과 백신 2차 접종률과 100만명당 발생률은 양의 상관관계를 갖는다.

In []:

