

## 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 [1]:



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

### 01 파일 불러오기

In [6]:



```
os.listdir()
```

Out[6]:

```
['.git',
 '.idea',
 '.ipynb_checkpoints',
 '01_dataScraping_corona01_today.ipynb',
 '01_dataScraping_corona02.ipynb',
 '01_dataScraping_corona02.py',
 '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',
 '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 [9]:

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

In [10]:

```
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\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
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
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
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210929_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 [11]:

```
import pandas as pd

# excel 데이터 셋도 pd.read_excel로 가져올 수 있음.
corona = pd.read_csv("./data/2021-09-28_corona.csv")
vaccine = pd.read_csv("./data/20210929_00_vaccine_bloomberg.csv")
country_code = pd.read_excel("./data/country.xlsx")

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

Out[11]:

```
((222, 12), (207, 6), (227, 4))
```

In [12]:



```
corona.head()
```

Out [12]:

	국 가	위중증	치 명 (%)	완치 (%)	발생률	인구수	확진자_ 합계	확진자1 일	사망자 합계	사망자 1일	완치합계
0	미 국 us	21819\n-275	1.6	76.0	132772	331002651	43942335	187382.0	709119	2781.0	33394833.
1	인 도 IN	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	337.0	32976760.
2	브 라 질 BR	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	218.0	20361191.
3	영 국 GB	878\n-38	1.8	80.7	113959	67886011	7701715	37485.0	136208	40.0	6219111.
4	러 시 아 RU	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	852.0	6635485.

In [13]:



```
vaccine.head()
```

Out [13]:

	국가	백신접종수	Enough_for_percent_of_people:	1차접종	2차접종	일별접종수
0	Global Total	6.172363e+09	-999.0	-999.0	-999.0	31740868.0
1	Mainland China	2.200202e+09	78.6	78.6	73.0	3223429.0
2	India	8.705669e+08	31.8	46.2	16.4	7488509.0
3	EU	5.629421e+08	63.4	67.2	64.6	999304.0
4	U.S.	3.901143e+08	60.9	64.3	55.3	716762.0

In [14]:



```
country_code.head()
```

Out [14]:

	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 나라명이 있으면 해당되는 열을 붙여라

In [15]:



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

Out [15]:

	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 [19]:



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

Out[19]:

```
Index(['국가', '위중증', '치명(%)', '완치(%)', '발생률', '인구수', '확진자_합계',
      '확진자1일', '사망자합계',
      '사망자1일', '완치합계', '완치1일', 'hanguk_code', '한글표기'],
      dtype='object')
```

In [20]:



```
col = ['국가', '한글표기', '위중증', '치명(%)', '완치(%)', '발생률',
       '인구수', '확진자_합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일']

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

Out[20]:

	국 가	한 글 표 기	위중증	치 명 (%)	완치 (%)	발생률	인구수	확진자_ 합계	확진자1 일	사망자 합계	사망자 1일	
0	미 국 US	미 국	21819\n-275	1.6	76.0	132772	331002651	43942335	187382.0	709119	2781.0	3%
1	인 도 IN	인 도	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	337.0	3%
2	브 라 질 BR	브 라 질	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	218.0	2%
3	영 국 GB	영 국	878\n-38	1.8	80.7	113959	67886011	7701715	37485.0	136208	40.0	
4	러 시 아 RU	러 시 아	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	852.0	(
...	...	...	...	...	...	...	...	...	...	...	...	
217	마 셜 제 도 MH	마 셜	NaN	-	100.0	68	59190	4	NaN	0	NaN	
218	바 누 아 투 vu	바 누 아 투	NaN	25	75.0	13	307145	4	NaN	1	NaN	
219	사 모 아 ws	사 모 아	NaN	-	100.0	15	198414	3	NaN	0	NaN	
220	세 인 트 헬 레 나 SH	세 인 트 헬 레 나	NaN	-	100.0	329	6077	2	NaN	0	NaN	

	국 가	한 글 표 기	위중증	치 명 (%)	완치 (%)	발생률	인구수	확진자_ 합계	확진자1 일	사망자 합계	사망자 1일
221	미 크 로 네 시 아 FM	미 크 로 네 시 아	NaN	-	100.0	9	115023	1	NaN	0	NaN

222 rows × 13 columns

## 두 데이터 셋 연결

In [21]:

```
new_corona.head()
```

Out[21]:

	국 가	한 글 표 기	위중증	치 명 (%)	완치 (%)	발생률	인구수	확진자_ 합계	확진자1 일	사망자 합계	사망자 1일	완치
0	미 국 us	미 국	21819\n-275	1.6	76.0	132772	331002651	43942335	187382.0	709119	2781.0	33394
1	인 도 IN	인 도	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	337.0	32976
2	브 라 질 BR	브 라 질	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	218.0	20361
3	영 국 GB	영 국	878\n-38	1.8	80.7	113959	67886011	7701715	37485.0	136208	40.0	6219
4	러 시 아 RU	러 시 아	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	852.0	6635

In [22]:



```
country_code.head()
```

Out[22]:

	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

In [23]:



```
country_code.columns = ['han_code', 'eng_code', 'country', 'etc']  
country_code.head()
```

Out[23]:

	han_code	eng_code	country	etc
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



In [24]:



```
df_corona = new_corona.merge(country_code, left_on='한글표기',
                              right_on='han_code')

df_corona.head()
```

Out[24]:

	국가	한글표기	위중증	치명(%)	완치(%)	발생률	인구수	확진자_합계	확진자1일	사망자_합계	사망자1일	완치
0	미국	미국	21819	1.6	76.0	132772	331002651	43942335	187382.0	709119	2781.0	33394
1	인도	인도	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	337.0	32976
2	브라질	브라질	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	218.0	20361
3	영국	영국	878	1.8	80.7	113959	67886011	7701715	37485.0	136208	40.0	6219
4	러시아	러시아	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	852.0	6635

In [25]:



```
df_corona.drop(['국가', 'han_code', 'country', 'etc'], axis=1, inplace=True)
df_corona.head()
```

Out [25]:

	한글 표기	위중증	치명 (%)	완치 (%)	발생률	인구수	확진자_합계	확진자1일	사망자_합계	사망자1일	완치합계
0	미국	21819\n-275	1.6	76.0	132772	331002651	43942335	187382.0	709119	2781.0	33394833.0
1	인도	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	337.0	32976760.0
2	브라질	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	218.0	20361191.0
3	영국	878\n-38	1.8	80.7	113959	67886011	7701715	37485.0	136208	40.0	6219111.0
4	러시아	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	852.0	6635485.0

In [26]:



```
vaccine.head()
```

Out [26]:

	국가	백신접종수	Enough_for_percent_of_people:	1차접종	2차접종	일별접종수
0	Global Total	6.172363e+09	-999.0	-999.0	-999.0	31740868.0
1	Mainland China	2.200202e+09	78.6	78.6	73.0	3223429.0
2	India	8.705669e+08	31.8	46.2	16.4	7488509.0
3	EU	5.629421e+08	63.4	67.2	64.6	999304.0
4	U.S.	3.901143e+08	60.9	64.3	55.3	716762.0

In [27]:



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

Out [27]:

	한글 표기	위중증	치명 (%)	완치 (%)	발생률	인구수	확진자_합계	확진자1일	사망자_합계	사망자1일	완치합계
0	미국	21819\n-275	1.6	76.0	132772	331002651	43942335	187382.0	709119	2781.0	33394833.0
1	인도	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	337.0	32976760.0
2	브라질	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	218.0	20361191.0
3	영국	878\n-38	1.8	80.7	113959	67886011	7701715	37485.0	136208	40.0	6219111.0
4	러시아	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	852.0	6635485.0

In [28]:



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

Out[28]:

	한글 표기	위중증	치명 (%)	완치 (%)	발생률	인구수	확진자_합계	확진자1일	사망자_합계	사망자1일	완치
0	미국	21819\n-275	1.6	76.0	132772	331002651	43942335	187382.0	709119	2781.0	333948
1	인도	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	337.0	329767
2	브라질	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	218.0	203611
3	영국	878\n-38	1.8	80.7	113959	67886011	7701715	37485.0	136208	40.0	6219
4	러시아	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	852.0	66354
...	...	...	...	...	...	...	...	...	...	...	...
173	마카오	NaN	-	90.0	108	649335	70	3.0	0	NaN	
174	몬트세랫	NaN	3.0	90.9	6611	4992	33	NaN	1	NaN	
175	사모아	NaN	-	100.0	15	198414	3	NaN	0	NaN	
176	세인트헬레나	NaN	-	100.0	329	6077	2	NaN	0	NaN	
177	미크로네시아	NaN	-	100.0	9	115023	1	NaN	0	NaN	

178 rows × 18 columns

In [29]:



```
df_corona_all.columns
```

Out[29]:

```
Index(['한글표기', '위중증', '치명(%)', '완치(%)', '발생률', '인구수', '확진자_합계', '확진자1일',  
      '사망자합계', '사망자1일', '완치합계', '완치1일', 'eng_code', '백신접종수',  
      'Enough_for_percent_of_people:', '1차접종', '2차접종', '일별접종수'],  
      dtype='object')
```

In [31]:



```
sel = [ '한글표기', 'eng_code', '위중증', '치명(%)', '완치(%)', '발생률', '인구수', '확진자_합계', '확진자_1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종수', 'Enough_for_percent_of_people:', '1차접종', '2차접종', '일별접종수' ]
df_corona_all_n = df_corona_all[sel].copy()
df_corona_all_n
```

Out[31]:

	한글표기	eng_code	위중증	치명(%)	완치(%)	발생률	인구수	확진자_합계	확진자1일	사망자합계	사망자1일
0	미국	U.S.	21819\n-275	1.6	76.0	132772	331002651	43942335	187382.0	709119	271
1	인도	India	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	31
2	브라질	Brazil	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	21
3	영국	U.K.	878\n-38	1.8	80.7	113959	67886011	7701715	37485.0	136208	4
4	러시아	Russia	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	81
...	...	...	...	...	...	...	...	...	...	...	...
173	마카오	Macau	NaN	-	90.0	108	649335	70	3.0	0	1
174	몬트세랫	Montserrat	NaN	3.0	90.9	6611	4992	33	NaN	1	1
175	사모아	Samoa	NaN	-	100.0	15	198414	3	NaN	0	1
176	세인트헬레나	St. Helena	NaN	-	100.0	329	6077	2	NaN	0	1
177	미크로네시아	Micronesia	NaN	-	100.0	9	115023	1	NaN	0	1

178 rows × 18 columns

In [32]:



```
df_corona_all_n.columns
```

Out [32]:

```
Index(['한글표기', 'eng_code', '위중증', '치명(%)', '완치(%)', '발생률', '인구수',  
      '확진자_합계',  
      '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종수',  
      'Enough_for_percent_of_people:', '1차접종', '2차접종', '일별접종수'],  
      dtype='object')
```

In [33]:



```
df_corona_all_n.columns = ['국가명', 'eng_code', '위중증', '치명(%)', '완치(%)', '발생률', '인구수',
                           '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종', '접종가능률(인구)',
                           '접종비율(일간)']
df_corona_all_n
```

Out [33]:

	국가명	eng_code	위중증	치명(%)	완치(%)	발생률	인구수	확진자_합계	확진자1일	사망자_합계	사망자1일
0	미국	U.S.	21819\n-275	1.6	76.0	132772	331002651	43942335	187382.0	709119	271
1	인도	India	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	31
2	브라질	Brazil	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	21
3	영국	U.K.	878\n-38	1.8	80.7	113959	67886011	7701715	37485.0	136208	4
4	러시아	Russia	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	81
...	...	...	...	...	...	...	...	...	...	...	...
173	마카오	Macau	NaN	-	90.0	108	649335	70	3.0	0	1
174	몬트세랫	Montserrat	NaN	3.0	90.9	6611	4992	33	NaN	1	1
175	사모아	Samoa	NaN	-	100.0	15	198414	3	NaN	0	1
176	세인트헬레나	St. Helena	NaN	-	100.0	329	6077	2	NaN	0	1
177	미크로네시아	Micronesia	NaN	-	100.0	9	115023	1	NaN	0	1

178 rows × 18 columns





In [34]:



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

Out[34]:

	국가명	eng_code	위중증	치명(%)	완치(%)	발생률	인구수	확진자_합계	확진자_1일	사망_자합계	사망_자1일	완치률
161	지브롤터	Gibraltar	2	1.8	97.1	164198	33691	5532	15.0	97	NaN	53%
110	몰디브	Maldives	23	0.3	97.9	156294	540544	84484	122.0	231	NaN	82%
33	포르투갈	Portugal	74\5	1.7	95.5	104718	10196709	1067775	600.0	17962	7.0	10192%
81	카타르	Qatar	16	0.3	99.2	82082	2881053	236482	94.0	605	NaN	2345%
108	싱가포르	Singapore	27\3	0.1	84.0	15305	5850342	89539	1647.0	80	2.0	751%
...	...	...	...	...	...	...	...	...	...	...	...	...
129	베냉	Benin	5	0.7	75.3	1894	12123200	22958	NaN	154	NaN	172%
149	남수단	South Sudan	4	1.1	96.8	1068	11193725	11954	6.0	128	NaN	115%
153	예멘	Yemen	23	18.9	62.0	301	29825964	8988	NaN	1703	NaN	55%
116	콩고민주공화국	DR Congo	NaN	1.9	54.3	635	89561403	56862	245.0	1084	1.0	308%
177	미크로네시아	Micronesia	NaN	-	100.0	9	115023	1	NaN	0	NaN	

178 rows × 18 columns

In [35]:



```
df_corona_all_n.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 178 entries, 0 to 177
Data columns (total 18 columns):
#   Column          Non-Null Count  Dtype
---  -
0   국가명          178 non-null   object
1   eng_code        178 non-null   object
2   위중증          131 non-null   object
3   치명(%)         178 non-null   object
4   완치(%)         175 non-null   float64
5   발생률          178 non-null   object
6   인구수          178 non-null   object
7   확진자_합계     178 non-null   int64
8   확진자1일       151 non-null   float64
9   사망자합계      178 non-null   int64
10  사망자1일       121 non-null   float64
11  완치합계        175 non-null   float64
12  완치1일         133 non-null   float64
13  백신접종       178 non-null   float64
14  접종가능률(인구) 178 non-null   float64
15  1차접종         178 non-null   float64
16  2차접종         178 non-null   float64
17  접종비율(일간)  178 non-null   float64
dtypes: float64(10), int64(2), object(6)
memory usage: 26.4+ KB
```

In [36]:



```
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 18 columns):
 #   Column          Non-Null Count  Dtype  
---  --
 0   국가명          178 non-null   object  
 1   eng_code        178 non-null   object  
 2   위중증          131 non-null   object  
 3   치명(%)         178 non-null   object  
 4   완치(%)         175 non-null   float64 
 5   발생률          178 non-null   int32   
 6   인구수          178 non-null   object  
 7   확진자_합계     178 non-null   int64   
 8   확진자1일       151 non-null   float64 
 9   사망자합계      178 non-null   int64   
10   사망자1일       121 non-null   float64 
11   완치합계        175 non-null   float64 
12   완치1일         133 non-null   float64 
13   백신접종       178 non-null   float64 
14   접종가능률(인구) 178 non-null   float64 
15   1차접종         178 non-null   float64 
16   2차접종         178 non-null   float64 
17   접종비율(일간)  178 non-null   float64 
dtypes: float64(10), int32(1), int64(2), object(5)
memory usage: 25.7+ KB
```

In [37]:



```
df_corona_all_n['발생률_순위'] = df_corona_all_n['발생률'].rank(ascending=False) # ascending=False
df_corona_all_n['접종률_순위'] = df_corona_all_n['접종가능률(인구)'].rank(ascending=False) # ascending=False
df_corona_all_n
```

Out[37]:

	국가명	eng_code	위중증	치명률(%)	완치률(%)	발생률	인구수	확진자_합계	확진자1일	사망자_합계	사망자1일
0	미국	U.S.	21819\n-275	1.6	76.0	132772	331002651	43942335	187382.0	709119	276
1	인도	India	8944	1.3	97.8	24431	1380004385	33714241	21093.0	447699	36
2	브라질	Brazil	8318	2.8	95.3	100520	212559417	21366395	14423.0	594702	26
3	영국	U.K.	878\n-38	1.8	80.7	113959	67886011	7701715	37485.0	136208	4
4	러시아	Russia	2300	2.8	88.9	51151	145934462	7464708	21559.0	205531	86
...	...	...	...	...	...	...	...	...	...	...	...
173	마카오	Macau	NaN	-	90.0	108	649335	70	3.0	0	1
174	몬트세랫	Montserrat	NaN	3.0	90.9	6611	4992	33	NaN	1	1
175	사모아	Samoa	NaN	-	100.0	15	198414	3	NaN	0	1
176	세인트헬레나	St. Helena	NaN	-	100.0	329	6077	2	NaN	0	1
177	미크로네시아	Micronesia	NaN	-	100.0	9	115023	1	NaN	0	1

178 rows × 20 columns



In [38]:



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

Out[38]:

	국 가 명	eng_code	위 중 증	치 명 (%)	완 치 (%)	발 생 률	인 구 수	확 진 자 _ 합 계	확 진 자 1 일	사 망 자 합 계	사 망 자 1 일	완 치 율
161	지 브 롤 터	Gibraltar	2	1.8	97.1	164198	33691	5532	15.0	97	NaN	537
110	몰 디 브	Maldives	23	0.3	97.9	156294	540544	84484	122.0	231	NaN	8267
44	아 랍 에 미 레 이트	UAE	NaN	0.3	99.0	74361	9890402	735457	277.0	2094	NaN	72784
171	그 린 란 드	Greenland	2	-	70.4	9952	56770	565	NaN	0	NaN	39
37	쿠 바	Cuba	406\n+20	0.8	95.4	76528	11326616	866808	6009.0	7330	51.0	82672
...	...	...	...	...	...	...	...	...	...	...	...	...
145	부 르 키 나 파 소	Burkina Faso	NaN	1.3	97.6	679	20903273	14199	27.0	181	1.0	1385
153	예 멘	Yemen	23	18.9	62.0	301	29825964	8988	NaN	1703	NaN	557
149	남 수 단	South Sudan	4	1.1	96.8	1068	11193725	11954	6.0	128	NaN	1157
164	차 드	Chad	NaN	3.5	96.4	306	16425864	5034	2.0	174	NaN	485
116	콩 고 민 주 공 화 국	DR Congo	NaN	1.9	54.3	635	89561403	56862	245.0	1084	1.0	3085

178 rows × 20 columns

In [39]:



```
df_corona_all_n.columns
```

Out[39]:

```
Index(['국가명', 'eng_code', '위중증', '치명(%)', '완치(%)', '발생률', '인구수', '확  
진자_합계',  
      '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종',  
'접종가능률(인구)', '1차접종',  
      '2차접종', '접종비율(일간)', '발생률_순위', '접종률_순위'],  
      dtype='object')
```



In [46]:



```
sel = [ '국가명', 'eng_code', '발생률', '인구수', '백신접종',
        '접종가능률(인구)', '발생률_순위', '접종률_순위', '1차접종', '2차접종',
        '위중증', '치명(%)', '완치(%)', '확진자_합계',
        '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '접종비율(일간)']

df_corona_all_n = df_corona_all_n[sel]
df_corona_all_n
```

Out[46]:

	국 가 명	eng_code	발생률	인구수	백신접종	접종 가능 률 (인 구)	발생 률_순 위	접종 률_순 위	1차 접종	2차접 종	위중증
0	미 국	U.S.	132772	331002651	390114340.0	60.9	13.0	47.5	64.3	55.3	21819\n275
1	인 도	India	24431	1380004385	870566900.0	31.8	103.0	95.0	46.2	16.4	8944
2	브 라 질	Brazil	100520	212559417	232250880.0	56.3	29.0	61.5	71.1	41.4	8318
3	영 국	U.K.	113959	67886011	93500856.0	70.0	20.0	26.0	73.0	67.0	878\n-38
4	러 시 아	Russia	51151	145934462	89682024.0	30.6	77.0	99.5	32.4	28.4	2300
...	...	...	...	...	...	...	...	...	...	...	..
173	마 카 오	Macau	108	649335	649564.0	47.8	175.0	73.0	51.2	44.6	NaN
174	몬 트 세 랫	Montserrat	6611	4992	2871.0	28.7	128.0	103.0	29.7	27.8	NaN
175	사 모 아	Samoa	15	198414	149431.0	37.2	177.0	86.0	49.3	25.0	NaN
176	세 인 트 헬 레 나	St. Helena	329	6077	7892.0	65.8	171.0	35.0	72.7	58.9	NaN
177	미 크 로 네 시 아	Micronesia	9	115023	72085.0	34.7	178.0	90.0	41.6	-999.0	NaN

178 rows × 20 columns

In [47]:



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

Out[47]:

	국가명	eng_code	발생률	인구수	백신접종	접종 가능 률(인 구)	발생 률_순 위	접종 률_순 위	1차 접종	2차접 종	위
131	세이셸	Seychelles	217058	98347	1.484020e+05	77.3	1.0	14.0	80.5	74.1	Na
96	몬테네그로	Montenegro	207781	628066	4.504510e+05	36.2	2.0	87.0	38.3	33.6	Na
141	안도라	Andorra	196583	77265	9.343000e+04	60.7	3.0	49.0	67.0	54.3	
161	지브롤터	Gibraltar	164198	33691	7.950200e+04	116.9	4.0	1.0	117.4	116.4	
78	바레인	Bahrain	161543	1701575	2.561756e+06	86.3	5.0	7.0	78.4	74.9	4\h
...	...	...	...	...	...	...	...	...	...	...	
159	니제르	Niger	247	24206644	6.218530e+05	1.3	174.0	168.0	1.7	0.4	Na
173	마카오	Macau	108	649335	6.495640e+05	47.8	175.0	73.0	51.2	44.6	Na
107	중국	Mainland China	67	1439323776	2.200202e+09	78.6	176.0	12.0	78.6	73.0	Na
175	사모아	Samoa	15	198414	1.494310e+05	37.2	177.0	86.0	49.3	25.0	Na
177	미크로네시아	Micronesia	9	115023	7.208500e+04	34.7	178.0	90.0	41.6	-999.0	Na

178 rows × 20 columns

## 인구당 확진자 비율

In [48]:



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

Out[48]:

	국 가 명	eng_code	발생률	인구수	백신접종	접종 가능 률 (인 구)	발생 률_순 위	접종 률_순 위	1차 접종	2차접 종	...	치 명 (%)
0	미 국	U.S.	132772	331002651	390114340.0	60.9	13.0	47.5	64.3	55.3	...	1.6
1	인 도	India	24431	1380004385	870566900.0	31.8	103.0	95.0	46.2	16.4	...	1.3
2	브 라 질	Brazil	100520	212559417	232250880.0	56.3	29.0	61.5	71.1	41.4	...	2.8
3	영 국	U.K.	113959	67886011	93500856.0	70.0	20.0	26.0	73.0	67.0	...	1.8
4	러 시 아	Russia	51151	145934462	89682024.0	30.6	77.0	99.5	32.4	28.4	...	2.8
...	...	...	...	...	...	...	...	...	...	...	...	...
173	마 카 오	Macau	108	649335	649564.0	47.8	175.0	73.0	51.2	44.6	...	-
174	몬 트 세 랫	Montserrat	6611	4992	2871.0	28.7	128.0	103.0	29.7	27.8	...	3.0
175	사 모 아	Samoa	15	198414	149431.0	37.2	177.0	86.0	49.3	25.0	...	-
176	세 인 트 헬 레 나	St. Helena	329	6077	7892.0	65.8	171.0	35.0	72.7	58.9	...	-
177	미 크 로 네 시 아	Micronesia	9	115023	72085.0	34.7	178.0	90.0	41.6	-999.0	...	-

178 rows × 21 columns

In [49]:



```
df_corona_all_n.columns
```

Out[49]:

```
Index(['국가명', 'eng_code', '발생률', '인구수', '백신접종', '접종가능률(인구)', '발  
생률_순위',  
      '접종률_순위', '1차접종', '2차접종', '위중증', '치명(%)', '완치(%)', '확진자_  
합계', '확진자1일',  
      '사망자합계', '사망자1일', '완치합계', '완치1일', '접종비율(일간)', '확진자비  
율(인구)'],  
      dtype='object')
```

In [51]:



```
sel = ['국가명', 'eng_code', '발생률', '인구수', '확진자_합계', '백신접종', '접종가능률(인구)', '발생률_순위', '확진자비율(인구)', '1차접종', '2차접종', '위중증', '치명(%)', '완치(%)', '확진자1일', '사망자1일', '완치합계', '완치1일', '접종비율(일간)']
df_corona_all_n = df_corona_all_n[sel]
df_corona_all_n
```

Out[51]:

	국 가 명	eng_code	발생률	인구수	확진자_합계	백신접종	접종 가능 률 (인구)	발생 률_순 위	접종 률_순 위	확진자비 율(인구)
0	미 국	U.S.	132772	331002651	43942335	390114340.0	60.9	13.0	47.5	0.132755
1	인 도	India	24431	1380004385	33714241	870566900.0	31.8	103.0	95.0	0.024431
2	브 라 질	Brazil	100520	212559417	21366395	232250880.0	56.3	29.0	61.5	0.100520
3	영 국	U.K.	113959	67886011	7701715	93500856.0	70.0	20.0	26.0	0.113451
4	러 시 아	Russia	51151	145934462	7464708	89682024.0	30.6	77.0	99.5	0.051151
...	...	...	...	...	...	...	...	...	...	...
173	마 카 오	Macau	108	649335	70	649564.0	47.8	175.0	73.0	0.000108
174	몬 트 세 랫	Montserrat	6611	4992	33	2871.0	28.7	128.0	103.0	0.006611
175	사 모 아	Samoa	15	198414	3	149431.0	37.2	177.0	86.0	0.000015
176	세 인 트 헬 레 나	St. Helena	329	6077	2	7892.0	65.8	171.0	35.0	0.000329
177	미 크 로 네 시아	Micronesia	9	115023	1	72085.0	34.7	178.0	90.0	0.000009

178 rows × 21 columns

In [52]:



```
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" )
```

```
20210929_01
C:\Users\Wtoto\Documents\Github\corona_analysis\data\
C:\Users\Wtoto\Documents\Github\corona_analysis\data\20210929_01
```

In [53]:



```
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 [54]:



```
import seaborn as sns
```

## 백신 접종률과 인구당 확진자 비율의 상관관계

In [64]:



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

In [65]:



```
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 [66]:

```
df_corona_all_n.columns
```

Out [66]:

```
Index(['국가명', 'eng_code', '발생률', '인구수', '확진자_합계', '백신접종', '접종가  
능률(인구)',  
      '발생률_순위', '접종률_순위', '확진자비율(인구)', '1차접종', '2차접종', '위중  
증', '치명(%)',  
      '완치(%)', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '접  
종비율(일간)'],  
      dtype='object')
```

In [67]:

```
df_corona_all_n.head()
```

Out [67]:

	국 가 명	eng_code	발생률	인구수	확진자_ 합계	백신접종	접종 가능 률 (인 구)	발생 률_순 위	접종 률_순 위	확진자비 율(인구)	...
0	미 국	U.S.	132772	331002651	43942335	390114340.0	60.9	13.0	47.5	0.132755	...
1	인 도	India	24431	1380004385	33714241	870566900.0	31.8	103.0	95.0	0.024431	...
2	브 라 질	Brazil	100520	212559417	21366395	232250880.0	56.3	29.0	61.5	0.100520	...
3	영 국	U.K.	113959	67886011	7701715	93500856.0	70.0	20.0	26.0	0.113451	...
4	러 시 아	Russia	51151	145934462	7464708	89682024.0	30.6	77.0	99.5	0.051151	...

5 rows × 21 columns

In [69]:

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

Out [69]:

```
count    178.000000  
mean      28.207865  
std       81.895828  
min      -999.000000  
25%        8.425000  
50%       28.300000  
75%       57.475000  
max      116.400000  
Name: 2차접종, dtype: float64
```



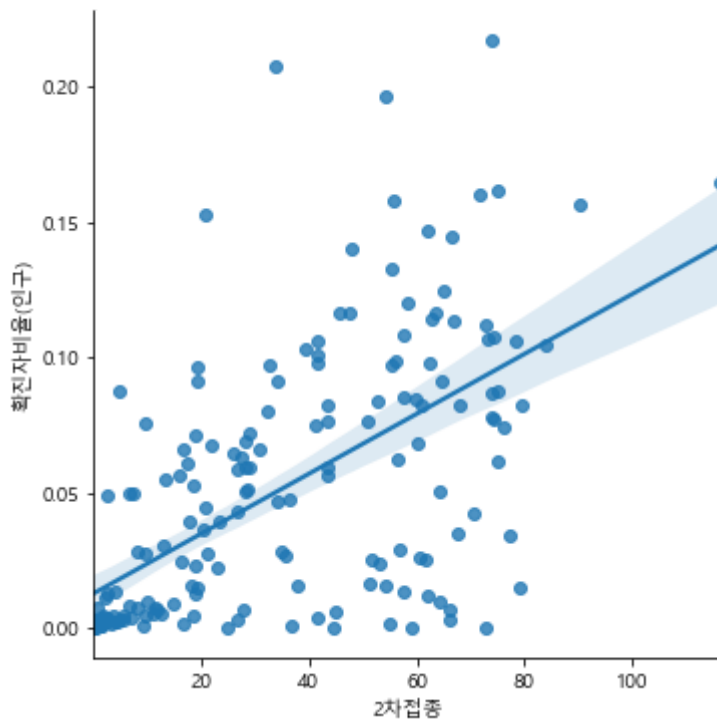
## 확진자 비율과 백신 접종률 상관관계 확인해 보기

In [71]:

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

In [73]:

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

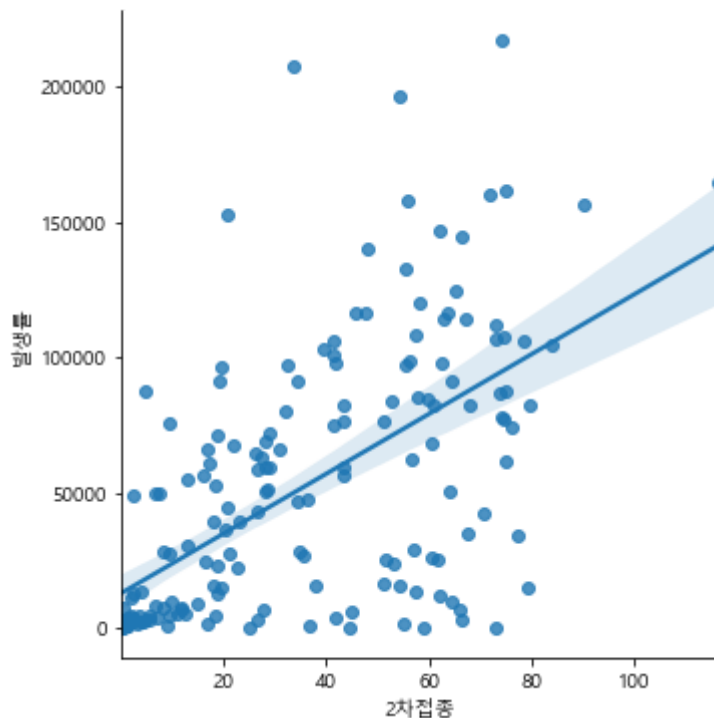


## 인구당 확진자 발생 비율과 백신 접종률 상관관계 그래프 그려보기

In [74]:



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



In [75]:



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

Out[75]:

0.5963209756874164

In [76]:



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

Out[76]:

0.5962873390099418

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

**02 현재까지의 누적된 데이터로 확인 결과 백신 2차 접종률과 확진자 발생 비율은 양의 상관관계를 갖는다.**

In [ ]:



