

# 코로나 데이터 분석

## 학습 목표

- 코로나 데이터와 백신 접종률 데이터를 수집하여 지금까지 내용을 나라별로 분석해 본다.
- 데이터 분석을 통해 pandas에 대해 좀 더 깊이 알아본다.

## 데이터가 수집된 사이트

- 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 [2]:



```
os.listdir(os.getcwd())
```

Out[2]:

```
['.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 [20]:



```
import pandas as pd  
  
corona = pd.read_csv("./data/20210929_01_datamerge.csv")  
corona.shape
```

Out[20]:

```
(178, 21)
```

## Learn 01. 보이지 않는 전체 내용을 표시해 보기

In [21]:



```
print( pd.options.display.max_rows )  
print( pd.options.display.max_columns )
```

```
500  
500
```

In [22]:



```
pd.options.display.max_columns = 500  
pd.options.display.max_rows = 500
```

In [23]:



```
corona.head()
```

Out[23]:

	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)	1차 접종
0	미 국	U.S.	132772	331002651	43942335	390114340.0	60.9	13.0	47.5	0.132755	64.3
1	인 도	India	24431	1380004385	33714241	870566900.0	31.8	103.0	95.0	0.024431	46.2
2	브 라 질	Brazil	100520	212559417	21366395	232250880.0	56.3	29.0	61.5	0.100520	71.1
3	영 국	U.K.	113959	67886011	7701715	93500856.0	70.0	20.0	26.0	0.113451	73.0
4	러 시 아	Russia	51151	145934462	7464708	89682024.0	30.6	77.0	99.5	0.051151	32.4

In [24]:



```
corona.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype  
---  --
 0   국가명                178 non-null   object 
 1   eng_code              178 non-null   object 
 2   발생률                178 non-null   int64  
 3   인구수                178 non-null   int64  
 4   확진자_합계          178 non-null   int64  
 5   백신접종            178 non-null   float64 
 6   접종가능률(인구)     178 non-null   float64 
 7   발생률_순위           178 non-null   float64 
 8   접종률_순위           178 non-null   float64 
 9   확진자비율(인구)     178 non-null   float64 
10   1차접종              178 non-null   float64 
11   2차접종              178 non-null   float64 
12   위중증               131 non-null   object 
13   치명(%)              178 non-null   object 
14   완치(%)              175 non-null   float64 
15   확진자1일            151 non-null   float64 
16   사망자합계           178 non-null   int64  
17   사망자1일            121 non-null   float64 
18   완치합계             175 non-null   float64 
19   완치1일              133 non-null   float64 
20   접종비율(일간)       178 non-null   float64 
dtypes: float64(13), int64(4), object(4)
memory usage: 29.3+ KB
```

**가설 : 백신 접종은 코로나 감염에 효과가 있다.**

## 접종률(인구)와 발생률의 관계

In [25]:



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

**한글 표시하기**

In [26]:

```
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 [27]:

```
corona.columns
```

Out[27]:

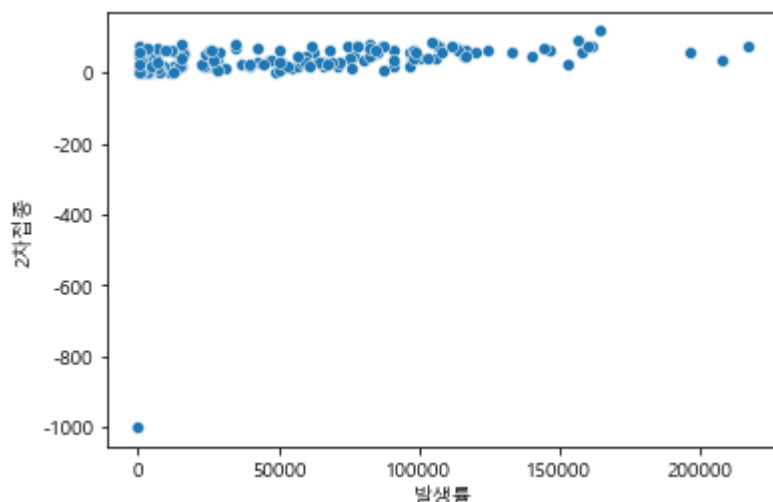
```
Index(['국가명', 'eng_code', '발생률', '인구수', '확진자_합계', '백신접종', '접종가  
능률(인구)',  
      '발생률_순위', '접종률_순위', '확진자비율(인구)', '1차접종', '2차접종', '위중  
증', '치명(%)',  
      '완치(%)', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '접  
종비율(일간)'],  
      dtype='object')
```

In [28]:

```
sns.scatterplot(x='발생률', y='2차접종', data=corona)
```

Out[28]:

<AxesSubplot: xlabel='발생률', ylabel='2차접종'>



## 이상치 제거

In [29]:

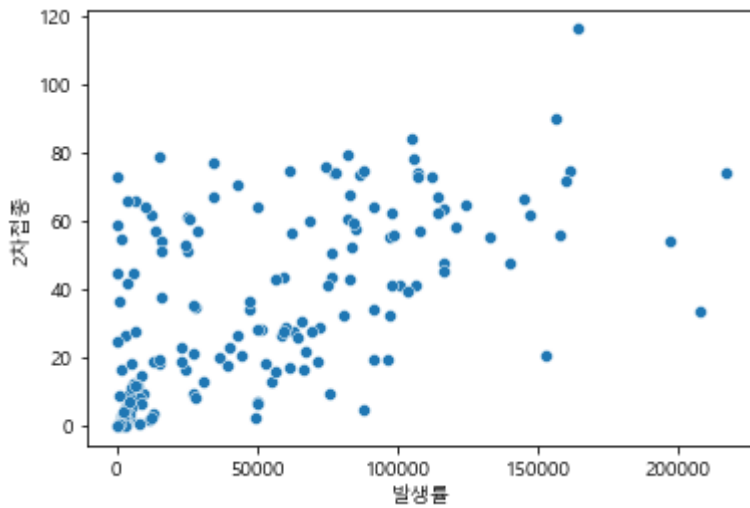
```
corona = corona[corona['2차접종'] > 0]
```

In [30]:

```
sns.scatterplot(x='발생률', y='2차접종', data=corona)
```

Out[30]:

<AxesSubplot:xlabel='발생률', ylabel='2차접종'>

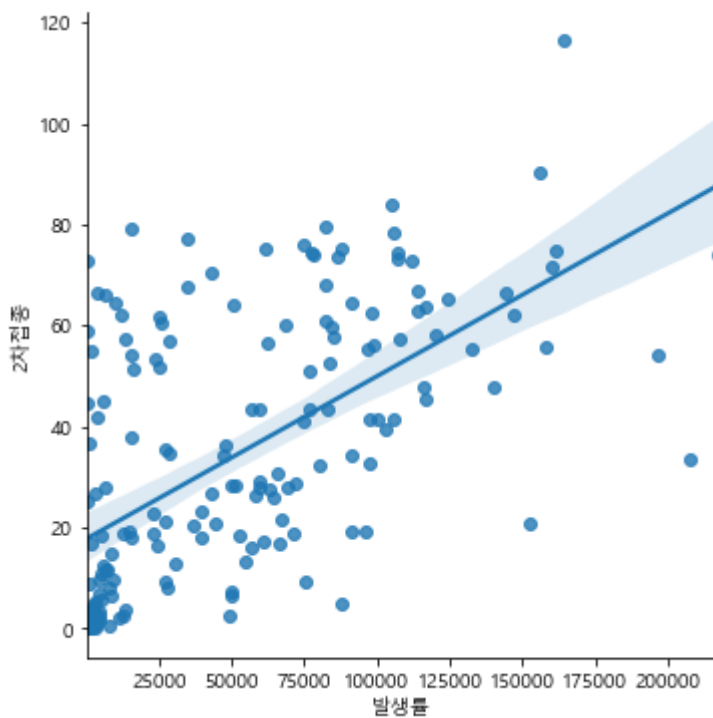


In [31]:

```
sns.lmplot(x='발생률', y='2차접종', data=corona)
```

Out[31]:

<seaborn.axisgrid.FacetGrid at 0x2518c8b5be0>



In [33]:



```
corona[['발생률', '2차접종']].corr()
```

Out[33]:

	발생률	2차접종
발생률	1.000000	0.596321
2차접종	0.596321	1.000000

확인 결과 나라별 백신 접종률과 인구당 확진자 비율은 0.582의 양의 상관관계를 갖는다.

- 백신을 접종하면 할수록 인구당 확진자 비율은 양의 관계로 늘어난다.

## 백신 발생률과 사망률(인구당)은 어떠한가?

In [34]:



```
corona.head()
```

Out[34]:

	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)	1차 접종
0	미 국	U.S.	132772	331002651	43942335	390114340.0	60.9	13.0	47.5	0.132755	64.3
1	인 도	India	24431	1380004385	33714241	870566900.0	31.8	103.0	95.0	0.024431	46.2
2	브 라 질	Brazil	100520	212559417	21366395	232250880.0	56.3	29.0	61.5	0.100520	71.1
3	영 국	U.K.	113959	67886011	7701715	93500856.0	70.0	20.0	26.0	0.113451	73.0
4	러 시 아	Russia	51151	145934462	7464708	89682024.0	30.6	77.0	99.5	0.051151	32.4

In [35]:



```
### 사망률 = 사망자수/인구수
corona['사망률'] = corona['사망자합계'] / corona['인구수']
corona.head(15)
```

Out[35]:

	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)	접
0	미 국	U.S.	132772	331002651	43942335	390114340.0	60.9	13.0	47.5	0.132755	6
1	인 도	India	24431	1380004385	33714241	870566900.0	31.8	103.0	95.0	0.024431	4
2	브 라 질	Brazil	100520	212559417	21366395	232250880.0	56.3	29.0	61.5	0.100520	7
3	영 국	U.K.	113959	67886011	7701715	93500856.0	70.0	20.0	26.0	0.113451	7
4	러 시 아	Russia	51151	145934462	7464708	89682024.0	30.6	77.0	99.5	0.051151	3
5	터 키	Turkey	83789	84339067	7066688	108344730.0	65.1	44.0	38.0	0.083789	6
6	프 랑 스	France	107174	65273511	6995628	93817816.0	72.4	23.0	22.0	0.107174	7
7	이 란	Iran	66192	83992949	5559691	43372270.0	26.0	62.0	107.5	0.066192	3
8	아 르 헨 티 나	Argentina	116204	45195774	5251940	51153324.0	56.9	18.0	58.5	0.116204	6
9	콜 롬 비아	Colombia	97335	50882891	4952690	39610550.0	39.3	33.0	80.0	0.097335	5
10	스 페 인	Spain	105907	46754778	4951640	69867540.0	75.1	26.0	18.0	0.105907	8
11	이 탈 리 아	Italy	77157	60461826	4662087	84158584.0	69.7	51.0	27.0	0.077108	7
12	독 일	Germany	50275	83783942	4212201	107030470.0	64.4	78.0	40.0	0.050275	6



국가명	eng_code	발생률	인구수	확진자_합계	백신접종	접종 가능 률 (인구)	발생 률_순 위	접종 률_순 위	확진자비 율(인구)	접
13 인도네시아	Indonesia	15397	273523615	4211460	136941020.0	25.7	110.0	110.0	0.015397	3
14 멕시코	Mexico	28199	128932753	3635807	99366400.0	38.9	95.0	81.0	0.028199	4

가설 : 백신은 중증과 사망자 감소에 효과가 있다.

## 2-1 백신 접종률과 사망률의 관계는 어떠한가?

In [38]:

```
corona[['사망률', '2차접종']].corr()
```

Out [38]:

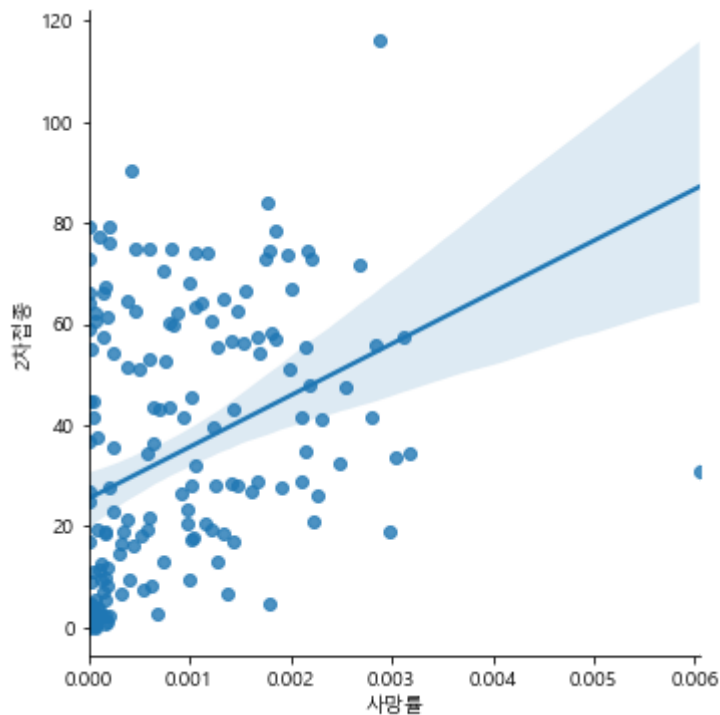
	사망률	2차접종
사망률	1.000000	0.355511
2차접종	0.355511	1.000000

In [39]:

```
sns.lmplot(x='사망률', y='2차접종', data=corona)
```

Out[39]:

<seaborn.axisgrid.FacetGrid at 0x2518c9eb370>



**백신 접종률과 사망률은 0.324의 양의 상관관계를 갖는다.**

- 데이터로 봤을 때, 이 부분에 대한 확인이 필요해 보인다.

**2-2 백신 접종률과 증증 감소의 관계는 어떠한가?**

In [40]:

```
corona.head(3)
```

Out[40]:

	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)	1차 접종
0	미 국	U.S.	132772	331002651	43942335	390114340.0	60.9	13.0	47.5	0.132755	64.3
1	인 도	India	24431	1380004385	33714241	870566900.0	31.8	103.0	95.0	0.024431	46.2
2	브 라 질	Brazil	100520	212559417	21366395	232250880.0	56.3	29.0	61.5	0.100520	71.1

In [41]:

```
corona['위중증'].unique()
```

Out[41]:

```
array(['21819Wn-275', '8944', '8318', '878Wn-38', '2300', '633', '1609',  
      '6124Wn-85', '1200Wn-37', '542', '837', '488Wn+5', '1435', nan,  
      '4798', '168Wn+15', '3170', '177', '978Wn-72', '1025Wn+2',  
      '160Wn-17', '516Wn-5', '1066Wn-44', '36Wn+6', '389Wn+9', '638',  
      '3324', '1408Wn-5', '641', '4015Wn-18', '206Wn-10', '1267Wn+49',  
      '35Wn-7', '74Wn-5', '552Wn-46', '251Wn+26', '221', '406Wn+20',  
      '209', '74Wn+3', '446Wn-30', '240Wn+6', '316', '323Wn-14', '200',  
      '5', '227Wn-17', '443Wn-16', '759', '220', '432Wn-108', '61', '23',  
      '10Wn-2', '119Wn+20', '102Wn+5', '13Wn-1', '66Wn+1', '681',  
      '418Wn-33', '15Wn-3', '791Wn-7', '203Wn+11', '324Wn+5', '20Wn-2',  
      '90', '192', '74Wn+6', '160Wn-1', '4Wn+1', '92', '16', '25Wn-3',  
      '11', '18', '21Wn+2', '131', '1', '3', '61Wn+9', '1124', '19Wn+2',  
      '32', '25', '12', '16Wn-1', '42', '327', '33Wn-6', '127', '301',  
      '9Wn-2', '27Wn-3', '152', '46Wn+15', '8', '9Wn-1', '67', '10Wn-29',  
      '10', '19', '22', '30Wn-3', '24', '26Wn+3', '15', '7', '4',  
      '34Wn-4', '2', '2Wn-4'], dtype=object)
```

## Learn02. Pandas 문자열 처리

- [].str.strip(): 앞뒤 공백을 제거
- [].str.lstrip(): 앞 공백을 제거
- [].str.rstrip(): 뒤 공백을 제거

- `df.str.split('구분자', n=1, expand=True)` : 구분자를 기준으로 n개로 나눈다. `expand=True`(여러컬럼), `False`(하나의 컬럼)

In [42]:



```
corona[['위중증_수', '위중증_증감']] = corona['위중증'].str.split('Wn', n=2, expand=True)
corona.head(15)
```

Out[42]:

	국가명	eng_code	발생률	인구수	확진자_합계	백신접종	접종 가능 률 (인구)	발생 률_순 위	접종 률_순 위	확진자비율(인구)	접종률
0	미국	U.S.	132772	331002651	43942335	390114340.0	60.9	13.0	47.5	0.132755	6
1	인도	India	24431	1380004385	33714241	870566900.0	31.8	103.0	95.0	0.024431	4
2	브라질	Brazil	100520	212559417	21366395	232250880.0	56.3	29.0	61.5	0.100520	7
3	영국	U.K.	113959	67886011	7701715	93500856.0	70.0	20.0	26.0	0.113451	7
4	러시아	Russia	51151	145934462	7464708	89682024.0	30.6	77.0	99.5	0.051151	3
5	터키	Turkey	83789	84339067	7066688	108344730.0	65.1	44.0	38.0	0.083789	6
6	프랑스	France	107174	65273511	6995628	93817816.0	72.4	23.0	22.0	0.107174	7
7	이란	Iran	66192	83992949	5559691	43372270.0	26.0	62.0	107.5	0.066192	3
8	아르헨티나	Argentina	116204	45195774	5251940	51153324.0	56.9	18.0	58.5	0.116204	6
9	콜롬비아	Colombia	97335	50882891	4952690	39610550.0	39.3	33.0	80.0	0.097335	5
10	스페인	Spain	105907	46754778	4951640	69867540.0	75.1	26.0	18.0	0.105907	8
11	이탈리아	Italy	77157	60461826	4662087	84158584.0	69.7	51.0	27.0	0.077108	7
12	독일	Germany	50275	83783942	4212201	107030470.0	64.4	78.0	40.0	0.050275	6

국가명	eng_code	발생률	인구수	확진자_합계	백신접종	접종 가능 률 (인구)	발생 률_순 위	접종 률_순 위	확진자비 율(인구)	집
13 인도네시아	Indonesia	15397	273523615	4211460	136941020.0	25.7	110.0	110.0	0.015397	3
14 멕시코	Mexico	28199	128932753	3635807	99366400.0	38.9	95.0	81.0	0.028199	4

In [43]:



```
### 사망률 = 사망자수/인구수
corona['위중증률'] = corona['위중증_수'] / corona['인구수']
corona.head(15)
```

```
-----
TypeError                                Traceback (most recent call last)
~Wanaconda3\lib\site-packages\pandas\core\ops\array_ops.py in _na_arithmetic
_op(left, right, op, is_cmp)
    141     try:
--> 142         result = expressions.evaluate(op, left, right)
    143     except TypeError:

~Wanaconda3\lib\site-packages\pandas\core\computation\expressions.py in ev
aluate(op, a, b, use_numexpr)
    234         # error: "None" not callable
--> 235         return _evaluate(op, op_str, a, b) # type: ignore[misc]
    236     return _evaluate_standard(op, op_str, a, b)

~Wanaconda3\lib\site-packages\pandas\core\computation\expressions.py in _e
valuate_numexpr(op, op_str, a, b)
    119     if result is None:
--> 120         result = _evaluate_standard(op, op_str, a, b)
    121

~Wanaconda3\lib\site-packages\pandas\core\computation\expressions.py in _e
valuate_standard(op, op_str, a, b)
     68     with np.errstate(all="ignore"):
--> 69         return op(a, b)
     70
```

**TypeError:** unsupported operand type(s) for /: 'str' and 'int'

During handling of the above exception, another exception occurred:

```
TypeError                                Traceback (most recent call last)
<ipython-input-43-0a97e14a9e3f> in <module>
      1 ### 사망률 = 사망자수/인구수
----> 2 corona['위중증률'] = corona['위중증_수'] / corona['인구수']
      3 corona.head(15)

~Wanaconda3\lib\site-packages\pandas\core\ops\common.py in new_method(self,
other)
     63         other = item_from_zerodim(other)
     64
--> 65         return method(self, other)
     66
     67     return new_method

~Wanaconda3\lib\site-packages\pandas\core\arraylike.py in __truediv__(self,
other)
    111     @unpack_zerodim_and_defer("__truediv__")
    112     def __truediv__(self, other):
--> 113         return self._arith_method(other, operator.truediv)
    114
    115     @unpack_zerodim_and_defer("__rtruediv__")

~Wanaconda3\lib\site-packages\pandas\core\series.py in _arith_method(self, o
ther, op)
```

```

4996         lvalues = extract_array(self, extract_numpy=True)
4997         rvalues = extract_array(other, extract_numpy=True)
-> 4998         result = ops.arithmetic_op(lvalues, rvalues, op)
4999
5000         return self._construct_result(result, name=res_name)

~Wanaconda3WlibWsite-packagesWpandasWcoreWopsWarray_ops.py in arithmetic_op
(left, right, op)
    187     else:
    188         with np.errstate(all="ignore"):
--> 189             res_values = _na_arithmetic_op(lvalues, rvalues, op)
    190
    191     return res_values

~Wanaconda3WlibWsite-packagesWpandasWcoreWopsWarray_ops.py in _na_arithmetic_
_op(left, right, op, is_cmp)
    147         # will handle complex numbers incorrectly, see GH#32047
    148         raise
--> 149     result = _masked_arith_op(left, right, op)
    150
    151     if is_cmp and (is_scalar(result) or result is NotImplemented):

~Wanaconda3WlibWsite-packagesWpandasWcoreWopsWarray_ops.py in _masked_arith_
_op(x, y, op)
    89         if mask.any():
    90             with np.errstate(all="ignore"):
---> 91                 result[mask] = op(xrav[mask], yrav[mask])
    92
    93     else:

```

`TypeError: unsupported operand type(s) for /: 'str' and 'int'`

## 에러 발생

- `TypeError: unsupported operand type(s) for /: 'str' and 'int'`

## Learn 03. 데이터 타입 변경하기

- `df.astype(dtype)`: 모든 열의 데이터 타입을 변경
- `df.astype({'컬럼명': dtype})`: 특정 컬럼의 데이터 타입을 변경
  - str, int, float, category 등
  - int8, int16, int32, int64와 같이 변경 가능



In [44]:



```
corona.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 177 entries, 0 to 176
Data columns (total 24 columns):
#   Column                Non-Null Count  Dtype
---  -
0   국가명                177 non-null    object
1   eng_code              177 non-null    object
2   발생률                177 non-null    int64
3   인구수                177 non-null    int64
4   확진자_합계          177 non-null    int64
5   백신접종            177 non-null    float64
6   접종가능률(인구)     177 non-null    float64
7   발생률_순위          177 non-null    float64
8   접종률_순위          177 non-null    float64
9   확진자비율(인구)     177 non-null    float64
10  1차접종               177 non-null    float64
11  2차접종               177 non-null    float64
12  위중증                131 non-null    object
13  치명(%)              177 non-null    object
14  완치(%)              174 non-null    float64
15  확진자1일            151 non-null    float64
16  사망자합계           177 non-null    int64
17  사망자1일            121 non-null    float64
18  완치합계             174 non-null    float64
19  완치1일              133 non-null    float64
20  접종비율(일간)       177 non-null    float64
21  사망률               177 non-null    float64
22  위중증_수            131 non-null    object
23  위중증_증감           58 non-null     object
dtypes: float64(14), int64(4), object(6)
memory usage: 34.6+ KB
```

In [45]:



```
corona['위중증_수'] = corona['위중증_수'].astype('int64')
```

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-45-5f8714d9e82b> in <module>
----> 1 corona['위중증_수'] = corona['위중증_수'].astype('int64')

~Wanaconda3\lib\site-packages\pandas\core\generic.py in astype(self, dtype,
copy, errors)
    5875         else:
    5876             # else, only a single dtype is given
-> 5877             new_data = self._mgr.astype(dtype=dtype, copy=copy, errors=e
rrors)
    5878             return self._constructor(new_data).__finalize__(self, method=
"astype")
    5879

~Wanaconda3\lib\site-packages\pandas\core\internals\managers.py in astype(s
elf, dtype, copy, errors)
    629         self, dtype, copy: bool = False, errors: str = "raise"
    630     ) -> "BlockManager":
--> 631         return self.apply("astype", dtype=dtype, copy=copy, errors=erro
rs)
    632
    633     def convert(

~Wanaconda3\lib\site-packages\pandas\core\internals\managers.py in apply(se
lf, f, align_keys, ignore_failures, **kwargs)
    425         applied = b.apply(f, **kwargs)
    426         else:
--> 427             applied = getattr(b, f)(**kwargs)
    428         except (TypeError, NotImplementedError):
    429             if not ignore_failures:

~Wanaconda3\lib\site-packages\pandas\core\internals\blocks.py in astype(sel
f, dtype, copy, errors)
    671         vals1d = values.ravel()
    672         try:
--> 673             values = astype_nansafe(vals1d, dtype, copy=True)
    674         except (ValueError, TypeError):
    675             # e.g. astype_nansafe can fail on object-dtype of str
ings

~Wanaconda3\lib\site-packages\pandas\core\dtypes\cast.py in astype_nansafe(a
rr, dtype, copy, skipna)
    1072         # work around NumPy brokenness, #1987
    1073         if np.issubdtype(dtype.type, np.integer):
-> 1074             return lib.astype_intsafe(arr.ravel(), dtype).reshape(arr.s
hape)
    1075
    1076         # if we have a datetime/timedelta array of objects
pandas\_libs\lib.pyx in pandas._libs.lib.astype_intsafe()

ValueError: cannot convert float NaN to integer
```

## 에러 발생

- ValueError: cannot convert float NaN to integer

In [46]:



```
corona['위중증_수'].unique()
```

Out[46]:

```
array(['21819', '8944', '8318', '878', '2300', '633', '1609', '6124',  
      '1200', '542', '837', '488', '1435', nan, '4798', '168', '3170',  
      '177', '978', '1025', '160', '516', '1066', '36', '389', '638',  
      '3324', '1408', '641', '4015', '206', '1267', '35', '74', '552',  
      '251', '221', '406', '209', '446', '240', '316', '323', '200', '5',  
      '227', '443', '759', '220', '432', '61', '23', '10', '119', '102',  
      '13', '66', '681', '418', '15', '791', '203', '324', '20', '90',  
      '192', '4', '92', '16', '25', '11', '18', '21', '131', '1', '3',  
      '1124', '19', '32', '12', '42', '327', '33', '127', '301', '9',  
      '27', '152', '46', '8', '67', '22', '30', '24', '26', '7', '34',  
      '2'], dtype=object)
```

## Learn 04. 어떻게 NaN을 체크할 것인가?

In [47]:



```
corona.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 177 entries, 0 to 176
Data columns (total 24 columns):
#   Column                Non-Null Count  Dtype
---  -
0   국가명                177 non-null    object
1   eng_code              177 non-null    object
2   발생률                177 non-null    int64
3   인구수                177 non-null    int64
4   확진자_합계          177 non-null    int64
5   백신접종            177 non-null    float64
6   접종가능률(인구)     177 non-null    float64
7   발생률_순위           177 non-null    float64
8   접종률_순위           177 non-null    float64
9   확진자비율(인구)     177 non-null    float64
10  1차접종              177 non-null    float64
11  2차접종              177 non-null    float64
12  위중증              131 non-null    object
13  치명(%)              177 non-null    object
14  완치(%)              174 non-null    float64
15  확진자1일            151 non-null    float64
16  사망자합계           177 non-null    int64
17  사망자1일            121 non-null    float64
18  완치합계             174 non-null    float64
19  완치1일              133 non-null    float64
20  접종비율(일간)       177 non-null    float64
21  사망률               177 non-null    float64
22  위중증_수            131 non-null    object
23  위중증_증감          58 non-null     object
dtypes: float64(14), int64(4), object(6)
memory usage: 34.6+ KB
```

## 위중증\_수가 isnull()인 것을 살펴보기

In [48]:

```
corona[ corona['위증증_수'].isnull() ]
```

Out[48]:

	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)	1차 접종	2차접 종	위 증 증 수
13	인 도 네 시 아	Indonesia	15397	273523615	4211460	1.369410e+08	25.7	110.0	110.0	0.015397	32.2	18.10	Na
41	네 팔	Nepal	27226	29136808	793271	1.264563e+07	22.2	98.0	116.0	0.027226	23.2	21.20	Na
42	베 트 남	Vietnam	7917	97338579	770640	3.836725e+07	19.9	123.0	122.0	0.007917	31.5	8.20	Na

In [49]:

```
corona['위증증_수'].isnull().sum()
```

Out[49]:

46

In [50]:

```
corona['위증증_수'].isna().sum()
```

Out[50]:

46

(의문) 왜 47개국은 위증증 통계가 없을까?

위증증\_수2를 만들고, 위증증수가 없는 것은 -99999로 만들자.

In [51]:

```
corona['위증증_수2'] = corona['위증증_수'].copy()
```

In [52]:



```
corona.loc[ corona['위중증_수2'].isnull(), "위중증_수2"] = -99999
corona[ corona['위중증_수2'].isnull() ]
```

Out[52]:

국가명	eng_code	발생률	인구수	확진자_합계	백신접종	접종가능률(인구)	발생률_순위	접종률_순위	확진자비율(인구)	1차접종	2차접종	위중증	치명(%)	완치(%)	확진자_1일	사망자합계	사망자_1일	완치합계	완치_1일	접종비율(일간)	eng_name
-----	----------	-----	-----	--------	------	-----------	--------	--------	-----------	------	------	-----	-------	-------	--------	-------	--------	------	-------	----------	----------

In [53]:



```
corona['위중증_수2'] = corona['위중증_수2'].astype('int64')
```

In [54]:



```
### 사망률 = 사망자수/인구수
corona['위중증률'] = corona['위중증_수2'] / corona['인구수']
corona.head(15)
```

Out[54]:

	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)	접종비율
0	미국	U.S.	132772	331002651	43942335	390114340.0	60.9	13.0	47.5	0.132755	6
1	인도	India	24431	1380004385	33714241	870566900.0	31.8	103.0	95.0	0.024431	4
2	브라질	Brazil	100520	212559417	21366395	232250880.0	56.3	29.0	61.5	0.100520	7
3	영국	U.K.	113959	67886011	7701715	93500856.0	70.0	20.0	26.0	0.113451	7
4	러시아	Russia	51151	145934462	7464708	89682024.0	30.6	77.0	99.5	0.051151	3
5	터키	Turkey	83789	84339067	7066688	108344730.0	65.1	44.0	38.0	0.083789	6
6	프랑스	France	107174	65273511	6995628	93817816.0	72.4	23.0	22.0	0.107174	7
7	이란	Iran	66192	83992949	5559691	43372270.0	26.0	62.0	107.5	0.066192	3
8	아르헨티나	Argentina	116204	45195774	5251940	51153324.0	56.9	18.0	58.5	0.116204	6
9	콜롬비아	Colombia	97335	50882891	4952690	39610550.0	39.3	33.0	80.0	0.097335	5
10	스페인	Spain	105907	46754778	4951640	69867540.0	75.1	26.0	18.0	0.105907	8
11	이탈리아	Italy	77157	60461826	4662087	84158584.0	69.7	51.0	27.0	0.077108	7
12	독일	Germany	50275	83783942	4212201	107030470.0	64.4	78.0	40.0	0.050275	6

국가명	eng_code	발생률	인구수	확진자_합계	백신접종	접종가능률(인구)	발생률_순위	접종률_순위	확진자비율(인구)	접
13 인도네시아	Indonesia	15397	273523615	4211460	136941020.0	25.7	110.0	110.0	0.015397	3
14 멕시코	Mexico	28199	128932753	3635807	99366400.0	38.9	95.0	81.0	0.028199	4

In [55]:



```
corona.columns
```

Out [55]:

```
Index(['국가명', 'eng_code', '발생률', '인구수', '확진자_합계', '백신접종', '접종가능률(인구)',
      '발생률_순위', '접종률_순위', '확진자비율(인구)', '1차접종', '2차접종', '위중증', '치명(%)',
      '완치(%)', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '접종비율(일간)', '사망률',
      '위중증_수', '위중증_증감', '위중증_수2', '위중증률'],
      dtype='object')
```

In [57]:



```
corona_sel = ['국가명', '발생률', '인구수', '확진자_합계', '백신접종', '2차접종',
              '사망률', '위중증', '위중증_증감', '위중증_수', '위중증_수2', '위중증률']
corona_part = corona[corona_sel]
corona_part
```

7	이란	66192	83992949	5559691	4.337227e+07	16.80	0.001427	6124	-85	-85	6124	6124	7.29
8	아르헨티나	116204	45195774	5251940	5.115332e+07	47.70	0.002543	1200	-37	-37	1200	1200	2.65
9	콜롬비아	97335	50882891	4952690	3.961055e+07	32.60	0.002480	542	None	None	542	542	1.06
10	스페인	105907	46754778	4951640	6.986754e+07	78.40	0.001846	837	None	None	837	837	1.79
11	이탈리아	77157	60461826	4662087	8.415858e+07	74.40	0.002162	488	+5	+5	488	488	8.07

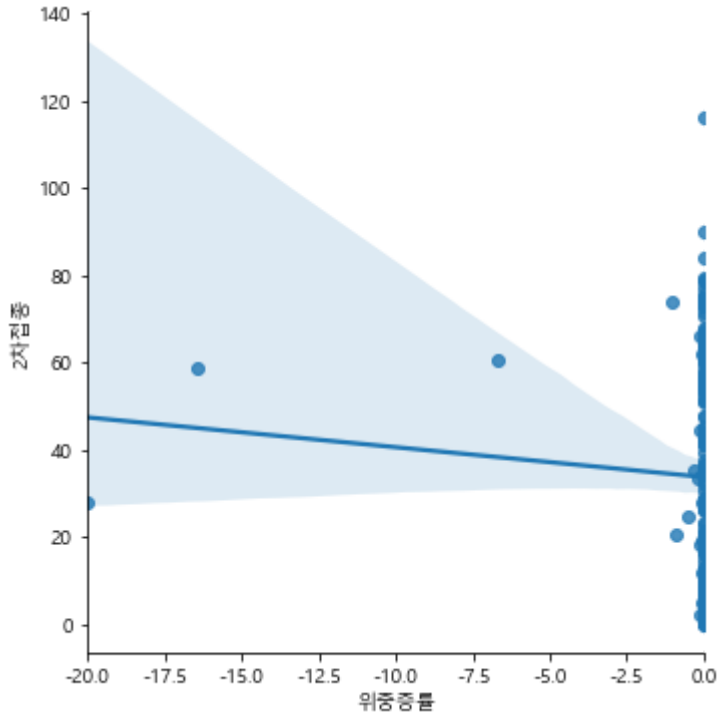


In [59]:

```
sns.lmplot(x='위중증률', y='2차접종', data=corona_part)
```

Out[59]:

<seaborn.axisgrid.FacetGrid at 0x2518cce2f70>



In [60]:

```
corona_part[['위중증률', '2차접종']].corr()
```

Out[60]:

	위중증률	2차접종
위중증률	1.00000	-0.05124
2차접종	-0.05124	1.00000

**결과 확인 : 인구당 접종률과 위중증률은 거의 관계가 없다.**

**오류가 있을까?**

- 오류: -99999 데이터가 있었기에 정확하지 않을 수 있다.

In [61]:

```
corona_part['위중증'].notnull().sum()
```

Out[61]:

131

In [62]:

```
corona_part = corona_part[ corona_part['위중증'].notnull() ]  
corona_part['위중증'].isnull().sum()
```

Out[62]:

0

In [63]:

```
corona_part[['위중증률', '2차접종']].corr()
```

Out[63]:

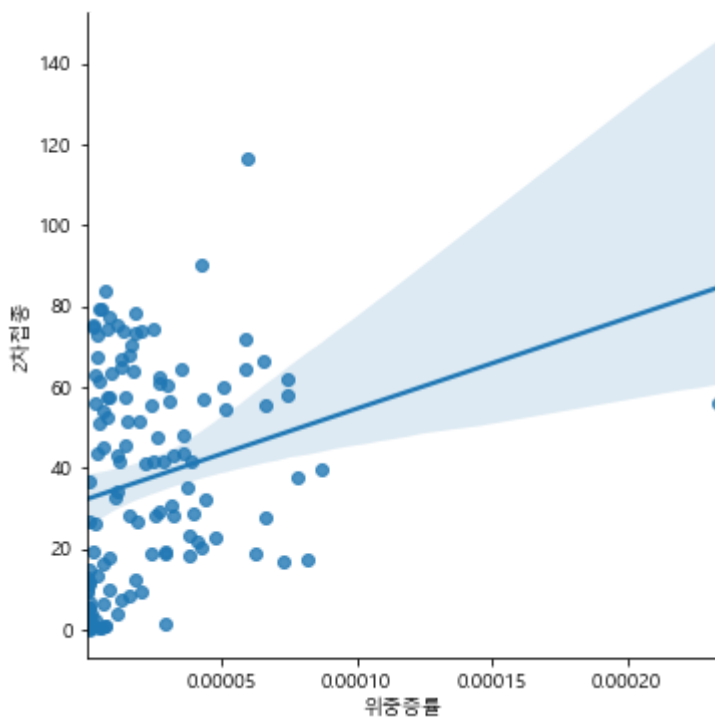
	위중증률	2차접종
위중증률	1.000000	0.234562
2차접종	0.234562	1.000000

In [64]:

```
sns.lmplot(x='위중증률', y='2차접종', data=corona_part)
```

Out[64]:

<seaborn.axisgrid.FacetGrid at 0x2518c9ebc70>



확인 : 데이터가 있는 131개국의 인구당 접종률과 위증증률은 0.2755로 양의 상관관계를 갖는다.

In [65]:



corona											
1	미국	USA	100020	212000111	21000000	2.022000e+08	66.0	20.0	0.0	0.100020	71.1
2	중국	China	1400000000	800000000	800000000	1.000000e+09	66.0	20.0	0.0	0.140000	71.1
3	영국	U.K.	113959	67886011	7701715	9.350086e+07	70.0	20.0	26.0	0.113451	73.0
4	러시아	Russia	51151	145934462	7464708	8.968202e+07	30.6	77.0	99.5	0.051151	32.4
5	터키	Turkey	83789	84339067	7066688	1.083447e+08	65.1	44.0	38.0	0.083789	64.4
6	프랑스	France	107174	65273511	6995628	9.381782e+07	72.4	23.0	22.0	0.107174	77.4
7	이란	Iran	66192	83992949	5559691	4.337227e+07	26.0	62.0	107.5	0.066192	35.5
8	아르헨티나	Argentina	116204	45195774	5251940	5.115332e+07	56.9	18.0	58.5	0.116204	65.6

실습과제 : 인구 100만 이상의 국가에 대해서 다시 수행해 보자.

- 접종률과 인구당 발생률 관계 확인
- 접종률과 사망률과 중증률 관계 확인

의문 : 그렇다면 전체 국가가 아닌 30개의 국가의 데이터를 확인해 보면 어떨까?

## 02. 접종률이 높은 30개의 국가의 데이터를 확인해 보기

In [66]:



```
corona_top30 = corona[corona['접종률_순위']<=30]
corona_top30.shape
```

Out[66]:

(30, 26)

In [67]:



```
corona_top30
```

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)	1차접 종	2차 접종	
3	영 국	U.K.	113959	67886011	7701715	9.350086e+07	70.0	20.0	26.0	0.113451	73.0	67.0	87
6	프 랑 스	France	107174	65273511	6995628	9.381782e+07	72.4	23.0	22.0	0.107174	77.4	74.4	
10	스 페 인	Spain	105907	46754778	4951640	6.986754e+07	75.1	26.0	18.0	0.105907	80.9	78.4	
11	이 탈	Italy	77157	60461826	1662087	8.415858e+07	69.7	51.0	27.0	0.077108	71.3	71.1	41

## Learn 05. 데이터의 인덱스를 다시 재설정하기

- [].reset\_index(inplace=True)

In [68]:



```
corona_top30.reset_index(inplace=True)  
corona_top30
```

Out [68]:

	index	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)	1차접 종	2차 접종
0	3	영 국	U.K.	113959	67886011	7701715	9.350086e+07	70.0	20.0	26.0	0.113451	73.0	67
1	6	프 랑 스	France	107174	65273511	6995628	9.381782e+07	72.4	23.0	22.0	0.107174	77.4	74
2	10	스 페 인	Spain	105907	46754778	4951640	6.986754e+07	75.1	26.0	18.0	0.105907	80.9	78
3	11	이 탈	Italy	77157	60461826	1662087	8.415858e+07	69.7	51.0	27.0	0.077108	71.3	71

## Learn 06. 데이터를 정렬시키기 2가지

- sort\_index(): 인덱스를 기준으로 정렬
- sort\_values(by=[], ascending=False, axis=0,1, inplace=True or False): 컬럼의 값을 기준으로 데이터 정렬

- by : 정렬할 열 선택
- axis : 0 : index, 1: columns

In [69]:

```
corona_top30.columns
```

Out [69]:

```
Index(['index', '국가명', 'eng_code', '발생률', '인구수', '확진자_합계', '백신접종',
      '접종가능률(인구)',
      '발생률_순위', '접종률_순위', '확진자비율(인구)', '1차접종', '2차접종', '위중증',
      '치명(%)',
      '완치(%)', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '접종비율(일간)',
      '사망률',
      '위중증_수', '위중증_증감', '위중증_수2', '위중증률'],
      dtype='object')
```

In [76]:

```
corona_sel = ['국가명', '발생률', '인구수', '확진자_합계', '백신접종', '확진자비율(인구)',
              '발생률_순위', '접종률_순위', '사망률', '2차접종',
              '위중증_수2', '위중증_증감', '위중증률']
corona_top30_part = corona_top30[corona_sel]
corona_top30_part
```

Out [76]:

	국 가 명	발생률	인구수	확진자_합계	백신접종	확진자비율(인구)	발생률_순위	접종률_순위	사망률	2차접종	위중증_수2	위중증_증감	위중증률
0	영국	113959	67886011	7701715	9.350086e+07	0.113451	20.0	26.0	0.002006	67.0	878	-38	0.0013
1	프랑스	107174	65273511	6995628	9.381782e+07	0.107174	23.0	22.0	0.001785	74.4	1609	None	0.0026
2	스페인	105907	46754778	4951640	6.986754e+07	0.105907	26.0	18.0	0.001846	78.4	837	None	0.0039
3	이탈리아	77157	60461826	4662087	8.415858e+07	0.077108	51.0	27.0	0.002162	74.4	488	+5	0.0080

In [77]:

```
# 접종률_순위를 기준으로 정렬, 내림차순
corona_top30_part_sorted = corona_top30_part.sort_values(by=['접종률_순위'], ascending=False)
```

In [78]:

```
corona_top30_part_sorted
```

Out[78]:

	국 가 명	발생률	인구수	확진자_합계	백신접종	확진자비율(인구)	발생률_순위	접종률_순위	사망률	2차접종	위중증_수2	위중증_증감	위중증_위험
26	지브롤터	164198	33691	5532	7.950200e+04	0.164198	4.0	1.0	0.002879	116.4	2	None	0.00
21	몰디브	156294	540544	84484	7.276640e+05	0.156294	8.0	2.0	0.000427	90.3	23	None	0.00
10	아랍에미레	74361	9890402	735457	1.987280e+07	0.074361	56.0	3.0	0.000212	76.0	-99999	NaN	-0.0

In [79]:

```
corona_top30_part_sorted.reset_index(inplace=True)
```

In [80]:

```
corona_top30_part_sorted
```

Out[80]:

	index	국 가 명	발생률	인구수	확진자_합계	백신접종	확진자비율(인구)	발생률_순위	접종률_순위	사망률	2차접종	위중증_수2	위중증_증감	위중증_위험
0	26	지브롤터	164198	33691	5532	7.950200e+04	0.164198	4.0	1.0	0.002879	116.4	2	Non	
1	21	몰디브	156294	540544	84484	7.276640e+05	0.156294	8.0	2.0	0.000427	90.3	23	Non	
2	10	아랍에미레	74361	9890402	735457	1.987280e+07	0.074361	56.0	3.0	0.000212	76.0	-99999	NaN	

상관관계 확인하기

In [81]:



```
corona_top30_part_sorted[['2차접종', '확진자비율(인구)']].corr()
```

Out[81]:

	2차접종	확진자비율(인구)
2차접종	1.000000	0.299941
확진자비율(인구)	0.299941	1.000000

In [82]:



```
corona_top30_part_sorted[['2차접종', '사망률']].corr()
```

Out[82]:

	2차접종	사망률
2차접종	1.000000	0.341103
사망률	0.341103	1.000000

In [83]:



```
corona_top30_part_sorted[['2차접종', '위중증률']].corr()
```

Out[83]:

	2차접종	위중증률
2차접종	1.000000	0.00598
위중증률	0.00598	1.000000

접종률 상위 30개국의 데이터 확인 결과 큰 연관관계가 보이지 않는다.

실습 : 위중증 데이터는 처리가 안되었다. 이를 처리해 보자.

### 03. 200만 이상의 국가에 대해서 확인

In [84]:



```
corona_200m = corona[corona['인구수'] >= 2000000]  
corona_200m.shape
```

Out[84]:

(136, 26)

In [85]:



```
corona_200m.reset_index(inplace=True)
corona_200m.head(10)
```

Out[85]:

index	국가명	eng_code	발생률	인구수	확진자_합계	백신접종	접종 가능 률 (인구)	발생 률_순 위	접종 률_순 위	확진자 율(인구)
0	미국	U.S.	132772	331002651	43942335	390114340.0	60.9	13.0	47.5	0.132
1	인도	India	24431	1380004385	33714241	870566900.0	31.8	103.0	95.0	0.024
2	브라질	Brazil	100520	212559417	21366395	232250880.0	56.3	29.0	61.5	0.100
3	영국	U.K.	113959	67886011	7701715	93500856.0	70.0	20.0	26.0	0.113
4	러시아	Russia	51151	145934462	7464708	89682024.0	30.6	77.0	99.5	0.051
5	터키	Turkey	83789	84339067	7066688	108344730.0	65.1	44.0	38.0	0.083
6	프랑스	France	107174	65273511	6995628	93817816.0	72.4	23.0	22.0	0.107
7	이란	Iran	66192	83992949	5559691	43372270.0	26.0	62.0	107.5	0.066
8	아르헨티나	Argentina	116204	45195774	5251940	51153324.0	56.9	18.0	58.5	0.116
9	콜롬비아	Colombia	97335	50882891	4952690	39610550.0	39.3	33.0	80.0	0.097



In [87]:



```
corona_200m.columns
```

Out[87]:

```
Index(['index', '국가명', 'eng_code', '발생률', '인구수', '확진자_합계', '백신접종',  
      '접종가능률(인구)',  
      '발생률_순위', '접종률_순위', '확진자비율(인구)', '1차접종', '2차접종', '위중  
증', '치명(%)',  
      '완치(%)', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '접  
종비율(일간)', '사망률',  
      '위중증_수', '위중증_증감', '위중증_수2', '위중증률'],  
      dtype='object')
```

In [88]:



```
corona_sel = ['국가명', '발생률', '인구수', '확진자_합계', '백신접종', '2차접종',  
             '확진자비율(인구)', '발생률_순위', '접종률_순위', '사망률',  
             '위중증_수2', '위중증_증감', '위중증률']  
corona_200m_part = corona_200m[corona_sel]  
corona_200m_part.head(10)
```

Out [88]:

	국 가 명	발생률	인구수	확진자_ 합계	백신접종	2차 접종	확진자비 율(인구)	발생 률_순 위	접종 률_순 위	사망률	위 중_ 증
0	미 국	132772	331002651	43942335	390114340.0	55.3	0.132755	13.0	47.5	0.002142	218
1	인 도	24431	1380004385	33714241	870566900.0	16.4	0.024431	103.0	95.0	0.000324	89
2	브 라 질	100520	212559417	21366395	232250880.0	41.4	0.100520	29.0	61.5	0.002798	83
3	영 국	113959	67886011	7701715	93500856.0	67.0	0.113451	20.0	26.0	0.002006	8
4	러 시 아	51151	145934462	7464708	89682024.0	28.4	0.051151	77.0	99.5	0.001408	23
5	터 키	83789	84339067	7066688	108344730.0	52.7	0.083789	44.0	38.0	0.000751	6
6	프 랑 스	107174	65273511	6995628	93817816.0	74.4	0.107174	23.0	22.0	0.001785	16
7	이 란	66192	83992949	5559691	43372270.0	16.8	0.066192	62.0	107.5	0.001427	61
8	아 르 헨 티 나	116204	45195774	5251940	51153324.0	47.7	0.116204	18.0	58.5	0.002543	12
9	콜 롬 비 아	97335	50882891	4952690	39610550.0	32.6	0.097335	33.0	80.0	0.002480	5

In [90]:



```
corona_200m_part_sorted = corona_200m_part.sort_values(by=['2차접종'], ascending=False)
corona_200m_part_sorted.head(10)
```

Out[90]:

	국 가 명	발생률	인구수	확진자_ 합계	백신접종	2차 접종	확진자비 율(인구)	발생 률_ 순위	접종 률_ 순위	사망률	위중증 _수2
33	포르투갈	104718	10196709	1067775	15944324.0	84.0	0.104718	27.0	13.0	0.001762	74
80	카타르	82082	2881053	236482	4696753.0	79.4	0.082082	47.0	8.0	0.000210	16
102	싱가포르	15305	5850342	89539	9286999.0	79.2	0.015305	111.0	11.0	0.000014	27
10	스페인	105907	46754778	4951640	69867540.0	78.4	0.105907	26.0	18.0	0.001846	837
44	아랍에미리트	74361	9890402	735457	19872800.0	76.0	0.074361	56.0	3.0	0.000212	-99999
68	덴마크	61777	5792202	357827	8811697.0	75.1	0.061777	67.0	17.0	0.000458	15
6	프랑스	107174	65273511	6995628	93817816.0	74.4	0.107174	23.0	22.0	0.001785	1609
11	이탈리아	77157	60461826	4662087	84158584.0	74.4	0.077108	51.0	27.0	0.002162	488
65	아일랜드	78116	4937786	385721	7195154.0	74.0	0.078116	50.0	20.0	0.001055	66
24	칠레	86460	19116201	1652795	31552752.0	73.8	0.086460	41.0	9.5	0.001959	389

### 상관관계 확인하기

In [94]:



```
corona_200m_part_sorted[['2차접종', '확진자비율(인구)']].corr()
```

Out[94]:

	2차접종	확진자비율(인구)
2차접종	1.000000	0.610009
확진자비율(인구)	0.610009	1.000000

In [95]:



```
corona_200m_part_sorted[['2차접종', '사망률']].corr()
```

Out[95]:

	2차접종	사망률
2차접종	1.000000	0.401819
사망률	0.401819	1.000000

In [96]:



```
corona_200m_part_sorted[['2차접종', '위증증률']].corr()
```

Out[96]:

	2차접종	위증증률
2차접종	1.000000	0.15664
위증증률	0.15664	1.000000

## 결과 확인

01. 인구 200만 이상의 국가의 백신 접종률과 인구당 확진자 비율은 양의 상관관계(0.610009)의 관계를 갖는다.

02. 인구 200만 이상의 국가의 백신 접종률과 사망률은 양의 상관관계(0.401819)의 관계를 갖는다.

03. 인구 200만 이상의 국가의 백신 접종률과 위증증률은 양의 상관관계(0.15664)의 관계를 갖는다.(관계가 크지 않음)

## 실습해 보기

- 위증증 nan의 값에 대해 처리가 되지 않았다. 이를 처리 후, 확인해 보자.

