

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 [121]:



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

01 파일 불러오기

In [122]:



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

Out [122]:

```
['.ipynb_checkpoints',  
'01_02_Pandas알아보기_v10.pdf',  
'01_03_pandas_01.html',  
'01_03_pandas_01.ipynb',  
'01_03_pandas_01.pdf',  
'01_03_pandas_02_california.html',  
'01_03_pandas_02_california.ipynb',  
'01_03_pandas_02_california.pdf',  
'01_04_titanic_dataset.html',  
'01_04_titanic_dataset.ipynb',  
'01_04_titanic_dataset.pdf',  
'01_04_화장품관련키워드분석_ing.ipynb',  
'01_05_titanic_dataset_pandas_etc.html',  
'01_05_titanic_dataset_pandas_etc.ipynb',  
'01_05_titanic_dataset_pandas_etc.pdf',  
'01_06_corona_analysis.ipynb',  
'01_unit01_03_pandas.zip',  
'20210921_00_datamerge.csv',  
'20210921_00_datamerge.xlsx',  
'california_housing_test.csv',  
'california_housing_train.csv',  
'debug.log',  
'decision_first_model.csv',  
'pandas_신재용_0115.pdf',  
'team_12.csv',  
'team_12.xlsx',  
'yelp_pandas_허수경.pdf',  
'[REF]PyClass02_Bike_Model_Pandas.ipynb',  
'상위키워드리스트_20210108.xlsx']
```

In [123]:



```
import pandas as pd  
  
corona = pd.read_csv("20210921_00_datamerge.csv")  
corona.shape
```

Out [123]:

(178, 21)

Learn 01. 보이지 않는 전체 내용을 표시해 보기

In [124]:



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

500

500

In [125]:



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

In [126]:



```
corona.head()
```

Out[126]:

	국 가 명	eng_code	발생률	인구수	확진자_ 합계	백신접종	접종 률 (인 구)	발생 률_순 위	접종 률_순 위	확진자비 율_전체 인구	1차 접종
0	미 국	U.S.	129632	331002651	42908619	385586020.0	60.2	13.0	46.0	0.129632	63.8
1	인 도	India	24277	1380004385	33483170	808389300.0	29.6	101.0	99.0	0.024263	43.7
2	브 라 질	Brazil	99924	212559417	21239783	222322660.0	53.9	27.0	62.0	0.099924	69.4
3	영 국	U.K.	109444	67886011	7429746	93002090.0	69.6	21.0	25.0	0.109444	72.7
4	러 시 아	Russia	49986	145934462	7294672	87810616.0	29.9	77.0	97.0	0.049986	31.9

In [7]:



```
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   접종률(인구)         178 non-null    float64  
 7   발생률_순위           178 non-null    float64  
 8   접종률_순위           178 non-null    float64  
 9   확진자비율_전체인구  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            130 non-null    float64  
16   사망자합계           178 non-null    int64   
17   사망자1일            104 non-null    float64  
18   완치합계             175 non-null    float64  
19   완치1일              115 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 [127]:



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

한글 표시하기

In [128]:

```
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")

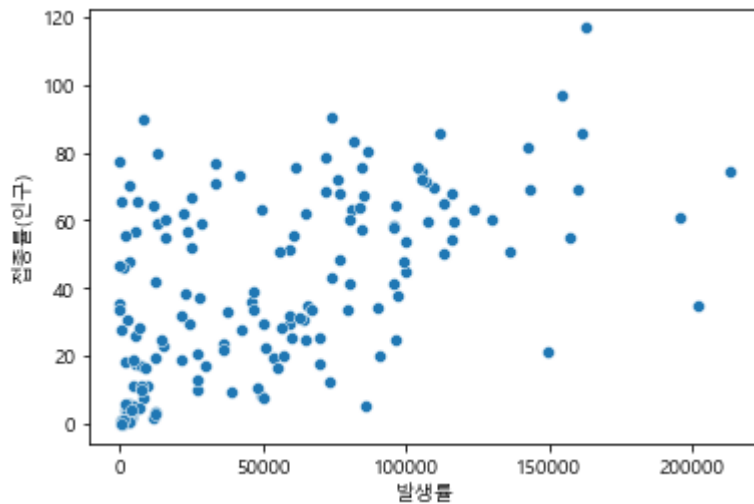
matplotlib.rcParams['axes.unicode_minus'] = False
```

In [129]:

```
sns.scatterplot(x='발생률', y='접종률(인구)', data=corona)
```

Out[129]:

<AxesSubplot:xlabel='발생률', ylabel='접종률(인구)'>

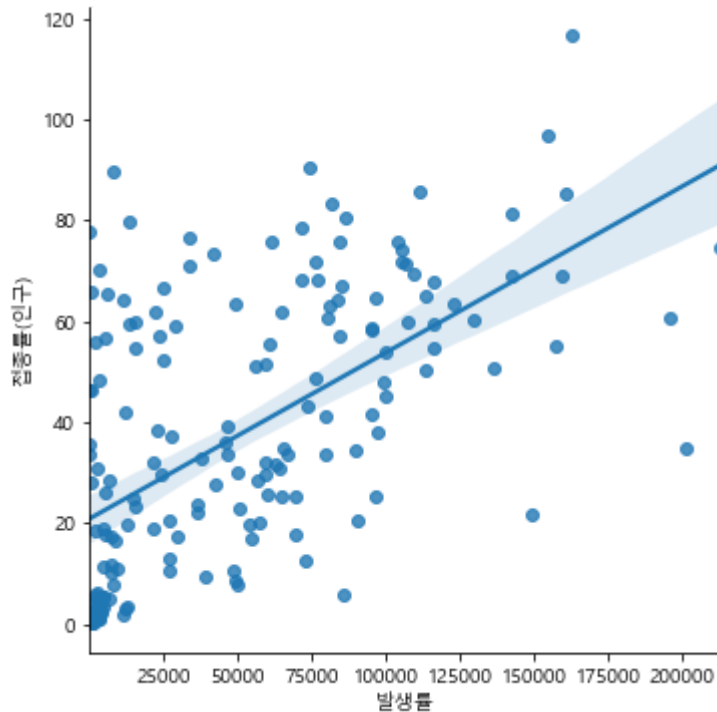


In [130]:

```
sns.lmplot(x='발생률', y='접종률(인구)', data=corona)
```

Out[130]:

<seaborn.axisgrid.FacetGrid at 0x1b63d38fd30>



In [131]:

```
corona[['발생률', '접종률(인구)']].corr()
```

Out[131]:

	발생률	접종률(인구)
발생률	1.000000	0.582878
접종률(인구)	0.582878	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 [132]:



```
corona.head()
```

Out[132]:

	국 가 명	eng_code	발생률	인구수	확진자_ 합계	백신접종	접종 률 (인 구)	발생 률_순 위	접종 률_순 위	확진자비 율_전체 인구	1차 접종
0	미 국	U.S.	129632	331002651	42908619	385586020.0	60.2	13.0	46.0	0.129632	63.8
1	인 도	India	24277	1380004385	33483170	808389300.0	29.6	101.0	99.0	0.024263	43.7
2	브 라 질	Brazil	99924	212559417	21239783	222322660.0	53.9	27.0	62.0	0.099924	69.4
3	영 국	U.K.	109444	67886011	7429746	93002090.0	69.6	21.0	25.0	0.109444	72.7
4	러 시 아	Russia	49986	145934462	7294672	87810616.0	29.9	77.0	97.0	0.049986	31.9

In [133]:



```
### 사망률 = 사망자수/인구수
corona['사망률'] = corona['사망자합계'] / corona['인구수']
corona
```

ia	97107	50882891	4941064	3.833164e+07	38.0	30.0	80.0	0.097107	49.3	31.70	542	2.5
ain	105434	46754778	4929546	6.909466e+07	74.3	24.0	18.0	0.105434	80.4	77.00	1028	1.7
aly	76678	60461826	4636111	8.237767e+07	68.2	49.0	29.0	0.076678	73.3	67.90	530	2.8
sia	15328	273523615	4192695	1.234548e+08	23.1	109.0	111.0	0.015328	29.4	16.80	NaN	3.4
ny	10560	83783012	4153067	1.051074e+08	63.4	79.0	100.0	0.010560	67.1	63.00	1217	2.3

가설 : 백신은 중증과 사망자 감소에 효과가 있다.

2-1 백신 접종률과 사망률의 관계는 어떠한가?

In [134]:

```
corona[['사망률', '접종률(인구)']].corr()
```

Out [134]:

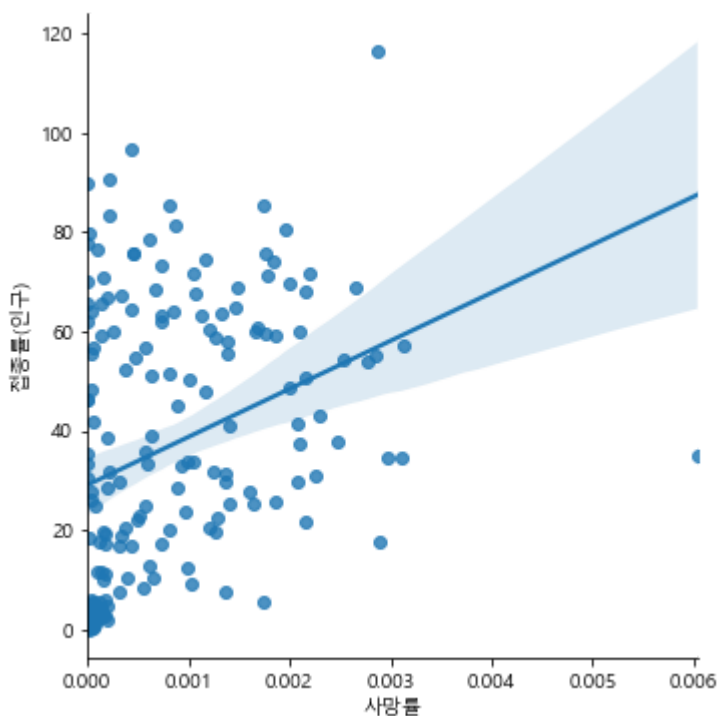
	사망률	접종률(인구)
사망률	1.000000	0.324012
접종률(인구)	0.324012	1.000000

In [135]:

```
sns.lmplot(x='사망률', y='접종률(인구)', data=corona)
```

Out [135]:

<seaborn.axisgrid.FacetGrid at 0x1b63d24de80>



백신 접종률과 사망률은 0.324의 양의 상관관계를 갖는다.

- 데이터로 봤을 때, 이 부분에 대한 확인이 필요해 보인다.

2-2 백신 접종률과 중증 감소의 관계는 어떠한가?

In [136]:



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

Out [136]:

	국 가 명	eng_code	발생률	인구수	확진자_ 합계	백신접종	접종 률 (인 구)	발생 률_순 위	접종 률_순 위	확진자비 율_전체 인구	1차 접종
0	미 국	U.S.	129632	331002651	42908619	385586020.0	60.2	13.0	46.0	0.129632	63.8
1	인 도	India	24277	1380004385	33483170	808389300.0	29.6	101.0	99.0	0.024263	43.7
2	브 라 질	Brazil	99924	212559417	21239783	222322660.0	53.9	27.0	62.0	0.099924	69.4

In [137]:



```
corona['위중증'].unique()
```

Out [137]:

```
array(['24327Wn-523', '8944', '8318', '1020', '2300', '1832Wn-168', '633',  
      '6707Wn-58', '1496Wn-20', '542', '1028', '530Wn+11', nan, '1217',  
      '4798', '113Wn+2', '3170', '177', '1043Wn+9', '1191Wn+26',  
      '208Wn+11', '618Wn+2', '28Wn-1', '1458Wn-42', '434Wn-4', '638',  
      '1403', '4387', '714Wn-12', '4840Wn-124', '218', '952Wn+103', '43',  
      '86Wn-4', '776Wn+3', '181Wn-2', '221', '260Wn+3', '45Wn+4',  
      '509Wn+2', '503Wn+42', '219Wn+7', '380Wn-5', '348', '200',  
      '339Wn-4', '5', '759', '489', '220', '391Wn-5', '108', '37Wn+1',  
      '16Wn-1', '60Wn+1', '85Wn+4', '11', '63Wn-3', '515', '681',  
      '22Wn-1', '784Wn-16', '166Wn+4', '19Wn-8', '90', '332Wn-1',  
      '63Wn+3', '100Wn+6', '192', '2', '109', '19Wn+1', '21Wn-3',  
      '20Wn+1', '131', '1', '23', '3', '1124', '36Wn-2', '18', '32',  
      '20', '12', '19', '42', '381', '45', '127Wn-31', '19Wn-4', '292',  
      '152', '55Wn-2', '21Wn+3', '9', '7Wn-1', '67', '8', '39', '10',  
      '22', '24', '15', '7', '4', '6', '2Wn+1', '22Wn-2'], dtype=object)
```

Learn02. Pandas 문자열 처리

- `[],str.strip()` : 앞뒤 공백을 제거
- `[],str.lstrip()` : 앞 공백을 제거
- `[],str.rstrip()` : 뒤 공백을 제거
- `[],str.split('구분자', n=1, expand=True)` : 구분자를 기준으로 n개로 나눈다. `expand=True`(여러컬럼), `False`(하나의 컬럼)

In [138]:



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

17	크 라 이 나	Ukraine	53749	43733762	2350646	1.655874e+07	19.8	75.0	119.0	0.053749	27.4	12.3
18	페 루	Peru	65723	32971854	2167008	2.315884e+07	34.9	60.0	84.0	0.065723	42.0	27.9
19	말 레 이 시 아	Malaysia	64816	32365999	2097830	4.037506e+07	62.0	61.0	42.0	0.064816	67.3	56.2
20	네 덜 란 드	Netherlands	115918	17134872	1986241	2.332039e+07	67.7	17.0	30.0	0.115918	70.0	63.4
21	이 라 크	Iraq	49187	40222493	1978412	6.672182e+06	8.5	80.0	138.0	0.049187	10.7	6.3

In [139]:



```
### 사망률 = 사망자수/인구수
corona['위중증률'] = corona['위중증_수'] / corona['인구수']
corona
```

```
-----
-
TypeError                                Traceback (most recent call last)
~Wanaconda3\lib\site-packages\pandas\core\ops\warray_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-139-94a0e7fdd58a> in <module>
      1 ### 사망률 = 사망자수/인구수
----> 2 corona['위중증률'] = corona['위중증_수'] / corona['인구수']
      3 corona

~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\warraylike.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, other, 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)

~Wanaconda3\lib\site-packages\pandas\core\ops\warray_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

~Wanaconda3\lib\site-packages\pandas\core\ops\warray_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):

~Wanaconda3\lib\site-packages\pandas\core\ops\warray_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 [140]:



```
corona.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 24 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   접종률(인구)          178 non-null    float64  
 7   발생률_순위           178 non-null    float64  
 8   접종률_순위           178 non-null    float64  
 9   확진자비율_전체인구  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            130 non-null    float64  
16   사망자합계           178 non-null    int64   
17   사망자1일            104 non-null    float64  
18   완치합계             175 non-null    float64  
19   완치1일              115 non-null    float64  
20   접종비율(일간)       178 non-null    float64  
21   사망률               178 non-null    float64  
22   위중증_수            131 non-null    object  
23   위중증_증감           50 non-null     object  
dtypes: float64(14), int64(4), object(6)
memory usage: 33.5+ KB
```

In [141]:



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

```
-----  
ValueError                                Traceback (most recent call last)  
<ipython-input-141-5f8714d9e82b> 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=errors)  
    5878             return self._constructor(new_data).__finalize__(self, method=  
"astype")  
    5879  
  
~Wanaconda3\lib\site-packages\pandas\core\internals\managers.py in astype(self, dtype, copy, errors)  
    629         self, dtype, copy: bool = False, errors: str = "raise"  
    630     ) -> "BlockManager":  
--> 631         return self.apply("astype", dtype=dtype, copy=copy, errors=errors)  
    632  
    633     def convert(  
  
~Wanaconda3\lib\site-packages\pandas\core\internals\managers.py in apply(self, 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(self, 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 strings  
ings  
  
~Wanaconda3\lib\site-packages\pandas\core\dtypes\cast.py in astype_nansafe(arr, 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.shape)  
    1075  
    1076         # if we have a datetime/timedelta array of objects
```

ValueError: cannot convert float NaN to integer

에러 발생

- ValueError: cannot convert float NaN to integer

In [142]:



```
corona['위중증_수'].unique()
```

Out[142]:

```
array(['24327', '8944', '8318', '1020', '2300', '1832', '633', '6707',  
      '1496', '542', '1028', '530', nan, '1217', '4798', '113', '3170',  
      '177', '1043', '1191', '208', '618', '28', '1458', '434', '638',  
      '1403', '4387', '714', '4840', '218', '952', '43', '86', '776',  
      '181', '221', '260', '45', '509', '503', '219', '380', '348',  
      '200', '339', '5', '759', '489', '220', '391', '108', '37', '16',  
      '60', '85', '11', '63', '515', '681', '22', '784', '166', '19',  
      '90', '332', '100', '192', '2', '109', '21', '20', '131', '1',  
      '23', '3', '1124', '36', '18', '32', '12', '42', '381', '127',  
      '292', '152', '55', '9', '7', '67', '8', '39', '10', '24', '15',  
      '4', '6'], dtype=object)
```

Learn 04. 어떻게 NaN을 체크할 것인가?

In [143]:



```
corona.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 24 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   접종률(인구)          178 non-null    float64
7   발생률_순위           178 non-null    float64
8   접종률_순위           178 non-null    float64
9   확진자비율_전체인구  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            130 non-null    float64
16  사망자합계            178 non-null    int64
17  사망자1일            104 non-null    float64
18  완치합계              175 non-null    float64
19  완치1일               115 non-null    float64
20  접종비율(일간)        178 non-null    float64
21  사망률                178 non-null    float64
22  위중증_수             131 non-null    object
23  위중증_증감           50 non-null     object
dtypes: float64(14), int64(4), object(6)
memory usage: 33.5+ KB
```

위중증_수가 isnull()인 것을 살펴보기

In [144]:



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

_code	발생률	인구수	확진자_합계	백신접종	접종률_순위 (인구)	발생률_순위	접종률_순위	확진자비율_전체인구	1차접종	2차접종	위증증	치명(%)	완치(%)
onesia	15328	273523615	4192695	1.234548e+08	23.1	109.0	111.0	0.015328	29.4	16.80	NaN	3.4	95.3
Nepal	26960	29136808	785541	1.178218e+07	20.7	97.0	115.0	0.026960	21.8	19.50	NaN	1.4	95.8
UAE	74113	9890402	733003	1.944587e+07	90.5	52.0	3.0	0.074113	84.5	74.50	NaN	0.3	98.9

In [145]:



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

Out[145]:

47

In [146]:



```
corona['위증증_수'].isna().sum()
```

Out[146]:

47

(의문) 왜 47개국은 위증증 통계가 없을까?

위증증_수2를 만들고, 위증증수가 없는 것은 -99999로 만들자.

In [147]:



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

In [148]:



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

Out[148]:

국가명	eng_code	발생률	인구수	확진자_합계	백신접종	접종률(인구)	발생률_순위	접종률_순위	확진자비율_전체인구	1차접종	2차접종	위중증	치명(%)	완치(%)	확진자1일	사망자합계	사망자1일	완치합계	완치1일	접종비율(일간)	사망률
-----	----------	-----	-----	--------	------	---------	--------	--------	------------	------	------	-----	-------	-------	-------	-------	-------	------	------	----------	-----

In [149]:



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

In [150]:



```
### 사망률 = 사망자수/인구수
corona['위중증률'] = corona['위중증_수2'] / corona['인구수']
corona
```

5	프랑스	France	106557	65273511	6955333	9.226241e+07	71.2	23.0	22.0	0.106557	76.7	72.7
6	터키	Turkey	81187	84339067	6847259	1.051146e+08	63.2	45.0	41.0	0.081187	63.3	50.5
7	이란	Iran	64794	83992949	5442232	4.200386e+07	25.2	62.0	107.0	0.064794	34.1	16.6
8	아르헨티나	Argentina	115923	45195774	5239232	4.901757e+07	54.5	16.0	61.0	0.115923	64.5	44.3

In [56]:



```
corona.columns
```

Out [56]:

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

In [151]:



```
corona_sel = ['국가명', '발생률', '인구수', '확진자_합계', '백신접종', '접종률(인구)',  
              '사망률', '위중증', '위중증_증감', '위중증_수', '위중증_수2', '위중증률']  
corona_part = corona[corona_sel]  
corona_part
```

8	헨 티 나	115923	45195774	5239232	4.901757e+07	54.5	0.002532	1496	n-20	-20	1496	1496	3.310
9	콜 롬 비 아	97107	50882891	4941064	3.833164e+07	38.0	0.002474	542	None	542	542	542	1.065
10	스 페 인	105434	46754778	4929546	6.909466e+07	74.3	0.001835	1028	None	1028	1028	1028	2.198
11	이 탈 ...	76678	60461826	4636111	8.237767e+07	68.2	0.002155	530	n+11	+11	530	530	8.765

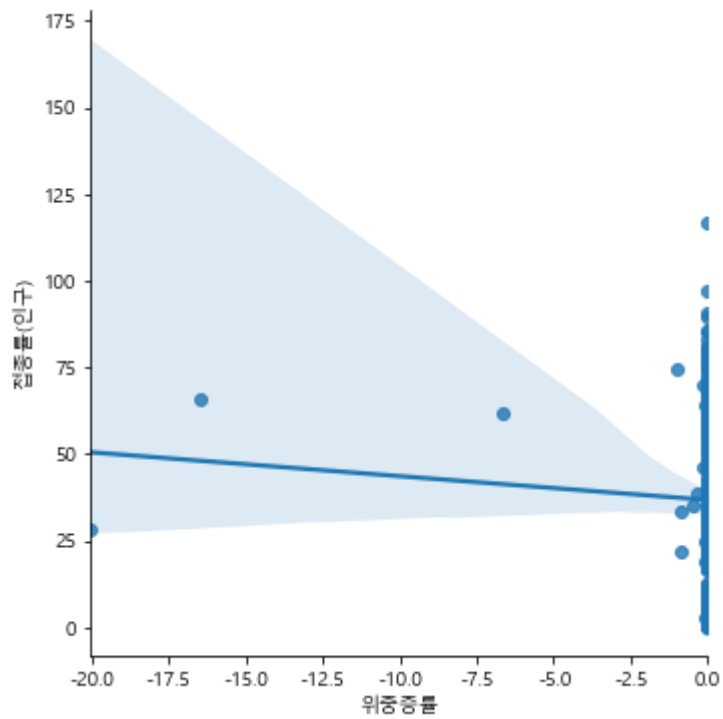
In [152]:



```
sns.lmplot(x='위중증률', y='접종률(인구)', data=corona_part)
```

Out [152]:

<seaborn.axisgrid.FacetGrid at 0x1b63cbbc490>



In [153]:



```
corona_part[['위중증률', '접종률(인구)']].corr()
```

Out [153]:

	위중증률	접종률(인구)
위중증률	1.00000	-0.04982
접종률(인구)	-0.04982	1.00000

결과 확인 : 인구당 접종률과 위중증률은 거의 관계가 없다.

오류가 있을까?

- 오류: -99999 데이터가 있었기에 정확하지 않을 수 있다.

In [154]:



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

Out [154]:

131

In [155]:



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

Out [155]:

0

In [156]:



```
corona_part[['위중증률', '접종률(인구)']].corr()
```

Out [156]:

	위중증률	접종률(인구)
위중증률	1.000000	0.275532
접종률(인구)	0.275532	1.000000

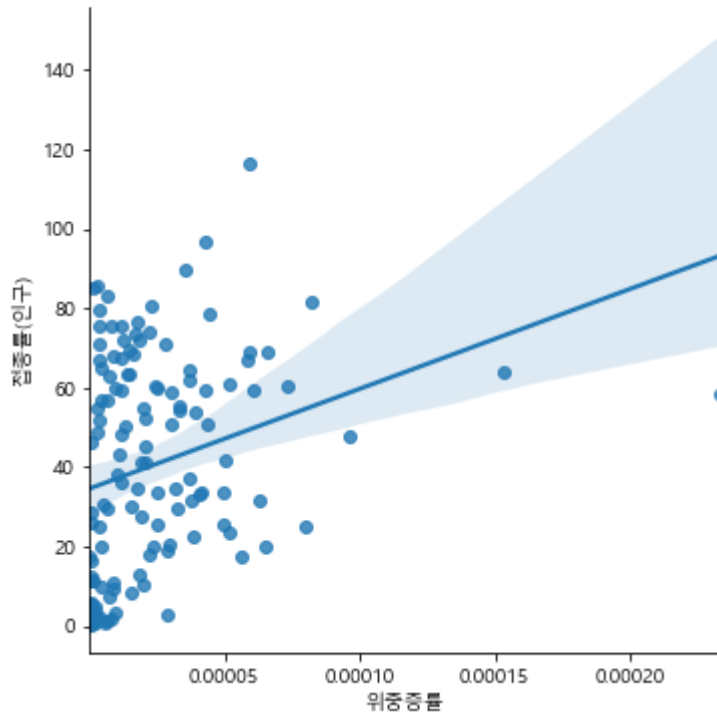
In [157]:



```
sns.lmplot(x='위중증률', y='접종률(인구)', data=corona_part)
```

Out [157]:

<seaborn.axisgrid.FacetGrid at 0x1b63e44dc40>



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

In [158]:



```
corona
```

eng_code	발생률	인구수	확진자_합계	백신접종	접종률(인구)	발생률_순위	접종률_순위	확진자비율_전체인구	1차접종	2차접종	위중
U.S.	129632	331002651	42908619	3.855860e+08	60.2	13.0	46.0	0.129632	63.8	54.60	2432715
India	24277	1380004385	33483170	8.083893e+08	29.6	101.0	99.0	0.024263	43.7	14.50	89
Brazil	99924	212559417	21239783	2.223227e+08	53.9	27.0	62.0	0.099924	69.4	37.80	83
U.K.	109444	67886011	7429746	9.300209e+07	69.6	21.0	25.0	0.109444	72.7	66.50	10
...

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 [159]:



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

Out [159]:

(30, 26)

In [160]:



```
corona_top30
```

Out [160]:

	국 가 명	eng_code	발생률	인구수	확진자_ 합계	백신접종	접종 률(인 구)	발생 률_순 위	접종 률_순 위	확진자비 율_전체 인구	1차접 종	2차 접종
3	영 국	U.K.	109444	67886011	7429746	9.300209e+07	69.6	21.0	25.0	0.109444	72.7	66.5
5	프 랑 스	France	106557	65273511	6955333	9.226241e+07	71.2	23.0	22.0	0.106557	76.7	72.7
10	스 페 인	Spain	105434	46754778	4929546	6.909466e+07	74.3	24.0	18.0	0.105434	80.4	77.0
11	이 탈 리	Italy	76678	60461826	4636111	8.237767e+07	68.2	49.0	29.0	0.076678	73.3	67.9

Learn 05. 데이터의 인덱스를 다시 재설정하기

- `{}.reset_index(inplace=True)`

In [166]:



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

3	3	11	이탈리아	Italy	76678	60461826	4636111	8.237767e+07	68.2	49.0	29.0	0.076678
4	4	20	네덜란드	Netherlands	115918	17134872	1986241	2.332039e+07	67.7	17.0	30.0	0.115918
5	5	24	칠레	Chile	86196	19116201	1647746	3.078569e+07	80.6	38.0	9.0	0.086196
6	6	25	캐나다	Canada	41687	37742154	1573359	5.506036e+07	73.4	86.0	19.0	0.041687

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 [167]:



```
corona_top30.columns
```

Out[167]:

```
Index(['level_0', 'index', '국가명', 'eng_code', '발생률', '인구수', '확진자_합계',  
      '백신접종',  
      '접종률(인구)', '발생률_순위', '접종률_순위', '확진자비율_전체인구', '1차접  
종', '2차접종', '위중증',  
      '치명(%)', '완치(%)', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완  
치1일', '접종비율(일간)',  
      '사망률', '위중증_수', '위중증_증감', '위중증_수2', '위중증률'],  
      dtype='object')
```

In [168]:



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

Out [168]:

	국 가 명	발생률	인구수	확진자_합계	백신접종	접종률(인구)	확진자비율_전체인구	발생률_순위	접종률_순위	사망률	위
0	영국	109444	67886011	7429746	9.300209e+07	69.6	0.109444	21.0	25.0	0.001992	
1	프랑스	106557	65273511	6955333	9.226241e+07	71.2	0.106557	23.0	22.0	0.001778	
2	스페인	105434	46754778	4929546	6.909466e+07	74.3	0.105434	24.0	18.0	0.001835	
3	이탈리아	76678	60461826	4636111	8.237767e+07	68.2	0.076678	49.0	29.0	0.002155	
4	네덜란드	115918	17134872	1986241	2.332039e+07	67.7	0.115918	17.0	30.0	0.001058	
5	칠레	86196	19116201	1647746	3.078569e+07	80.6	0.086196	38.0	9.0	0.001955	
6	캐나다	41687	37742154	1573359	5.506036e+07	73.4	0.041687	86.0	19.0	0.000726	
7	이스라엘	142297	8655535	1231660	1.473209e+07	81.4	0.142297	11.0	8.0	0.000871	
8	벨기에	105251	11589623	1219814	1.644224e+07	71.8	0.105251	25.0	21.0	0.002200	
9	포르투갈	104156	10196709	1062048	1.555099e+07	75.7	0.104156	26.0	15.0	0.001756	
10	쿠바	71505	11326616	809911	1.776627e+07	78.4	0.071505	56.0	11.0	0.000605	
11	아랍에미리트	74113	9890402	733003	1.944587e+07	90.5	0.074113	52.0	3.0	0.000210	-95

국 가 명	발생률	인구수	확진자_ 합계	백신접종	접종 률(인 구)	확진자비 율_전체 인구	발생 률_순 위	접종 률_순 위	사망률	위 .
12 우 루 과 이	111588	3473730	387627	6.017599e+06	85.5	0.111588	20.0	5.0	0.001741	
13 아 일 랜 드	76019	4937786	375367	7.119653e+06	71.9	0.076019	51.0	20.0	0.001049	
14 덴 마 크	61274	5792202	354913	8.773849e+06	75.6	0.061274	65.0	16.0	0.000454	
15 바 레 인	161182	1701575	274264	2.535900e+06	85.4	0.161182	5.0	6.0	0.000816	
16 카 타 르	81785	2881053	235626	4.658268e+06	83.3	0.081785	44.0	7.0	0.000210	
17 노 르 웨 이	33664	5421241	182501	7.586733e+06	70.8	0.033664	91.0	23.0	0.000155	
18 중 국	67	1439323776	95738	2.174043e+09	77.6	0.000067	176.0	12.0	0.000003	-95
19 몰 디 브	154615	540544	83576	7.206660e+05	96.9	0.154615	8.0	2.0	0.000424	
20 싱 가 포 르	13299	5850342	77804	9.104751e+06	79.8	0.013299	111.0	10.0	0.000010	
21 세 이 셸	212950	98347	20943	1.434900e+05	74.7	0.212950	1.0	17.0	0.001159	-95
22 아 루 바	142826	106766	15249	1.544470e+05	68.9	0.142826	10.0	27.0	0.001480	
23 아 이 슬 란 드	33419	341243	11404	5.477880e+05	76.7	0.033419	92.0	13.0	0.000097	
24 맨 섬	84273	85033	7166	1.289150e+05	75.8	0.084273	42.0	14.0	0.000447	
25 지 브 롤 터	162625	33691	5479	7.926900e+04	116.6	0.162625	4.0	1.0	0.002879	

국 가 명	발생률	인구수	확진자_합계	백신접종	접종률(인구)	확진자비율_전체인구	발생률_순위	접종률_순위	사망률	위
26 산마리노	159706	33931	5419	4.691700e+04	69.0	0.159706	6.0	26.0	0.002652	
27 버뮤다	71887	62278	4477	8.619400e+04	68.4	0.071887	55.0	28.0	0.000674	
28 부탄	3366	771608	2597	1.046117e+06	70.2	0.003366	141.0	24.0	0.000004	-95
29 그린란드	8173	56770	464	1.023570e+05	89.8	0.008173	122.0	4.0	0.000000	

In [169]:



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

In [170]:



```
corona_top30_part_sorted
```

Out[170]:

	국 가 명	발생률	인구수	확진자_ 합계	백신접종	접종 률(인 구)	확진자비 율_전체 인구	발생 률_순 위	접종 률_순 위	사망률	위 .
25	지 브 롤 터	162625	33691	5479	7.926900e+04	116.6	0.162625	4.0	1.0	0.002879	
19	몰 디 브	154615	540544	83576	7.206660e+05	96.9	0.154615	8.0	2.0	0.000424	
11	아 랍 에 미 레 이 트	74113	9890402	733003	1.944587e+07	90.5	0.074113	52.0	3.0	0.000210	-95
29	그 린 란 드	8173	56770	464	1.023570e+05	89.8	0.008173	122.0	4.0	0.000000	
12	우 루 과 이	111588	3473730	387627	6.017599e+06	85.5	0.111588	20.0	5.0	0.001741	
15	바 레 인	161182	1701575	274264	2.535900e+06	85.4	0.161182	5.0	6.0	0.000816	
16	카 타 르	81785	2881053	235626	4.658268e+06	83.3	0.081785	44.0	7.0	0.000210	
7	이 스 라 엘	142297	8655535	1231660	1.473209e+07	81.4	0.142297	11.0	8.0	0.000871	
5	칠 레	86196	19116201	1647746	3.078569e+07	80.6	0.086196	38.0	9.0	0.001955	
20	싱 가 포 르	13299	5850342	77804	9.104751e+06	79.8	0.013299	111.0	10.0	0.000010	
10	쿠 바	71505	11326616	809911	1.776627e+07	78.4	0.071505	56.0	11.0	0.000605	
18	중 국	67	1439323776	95738	2.174043e+09	77.6	0.000067	176.0	12.0	0.000003	-95

국 가 명	발생률	인구수	확진자_합계	백신접종	접종률(인구)	확진자비율_전체인구	발생률_순위	접종률_순위	사망률	위
23 아이슬란드	33419	341243	11404	5.477880e+05	76.7	0.033419	92.0	13.0	0.000097	
24 맨섬	84273	85033	7166	1.289150e+05	75.8	0.084273	42.0	14.0	0.000447	
9 포르투갈	104156	10196709	1062048	1.555099e+07	75.7	0.104156	26.0	15.0	0.001756	
14 덴마크	61274	5792202	354913	8.773849e+06	75.6	0.061274	65.0	16.0	0.000454	
21 세이셸	212950	98347	20943	1.434900e+05	74.7	0.212950	1.0	17.0	0.001159	-95
2 스페인	105434	46754778	4929546	6.909466e+07	74.3	0.105434	24.0	18.0	0.001835	
6 캐나다	41687	37742154	1573359	5.506036e+07	73.4	0.041687	86.0	19.0	0.000726	
13 아일랜드	76019	4937786	375367	7.119653e+06	71.9	0.076019	51.0	20.0	0.001049	
8 벨기에	105251	11589623	1219814	1.644224e+07	71.8	0.105251	25.0	21.0	0.002200	
1 프랑스	106557	65273511	6955333	9.226241e+07	71.2	0.106557	23.0	22.0	0.001778	
17 노르웨이	33664	5421241	182501	7.586733e+06	70.8	0.033664	91.0	23.0	0.000155	
28 부탄	3366	771608	2597	1.046117e+06	70.2	0.003366	141.0	24.0	0.000004	-95
0 영국	109444	67886011	7429746	9.300209e+07	69.6	0.109444	21.0	25.0	0.001992	
26 산마리노	159706	33931	5419	4.691700e+04	69.0	0.159706	6.0	26.0	0.002652	
22 아루바	142826	106766	15249	1.544470e+05	68.9	0.142826	10.0	27.0	0.001480	

국가명	발생률	인구수	확진자_합계	백신접종	접종률(인구)	확진자비율_전체인구	발생률_순위	접종률_순위	사망률	위
27 버뮤다	71887	62278	4477	8.619400e+04	68.4	0.071887	55.0	28.0	0.000674	
3 이탈리아	76678	60461826	4636111	8.237767e+07	68.2	0.076678	49.0	29.0	0.002155	
4 네덜란드	115918	17134872	1986241	2.332039e+07	67.7	0.115918	17.0	30.0	0.001058	

In [171]:



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

In [172]:



```
corona_top30_part_sorted
```

Out[172]:

	index	국가명	발생률	인구수	확진자_합계	백신접종	접종률(인구)	확진자비율_전체인구	발생률_순위	접종률_순위	사망률
0	25	지브롤터	162625	33691	5479	7.926900e+04	116.6	0.162625	4.0	1.0	0.0021
1	19	몰디브	154615	540544	83576	7.206660e+05	96.9	0.154615	8.0	2.0	0.0004
2	11	아랍에미리트	74113	9890402	733003	1.944587e+07	90.5	0.074113	52.0	3.0	0.0001
3	29	그린란드	8173	56770	464	1.023570e+05	89.8	0.008173	122.0	4.0	0.0001
4	12	우루과이	111588	3473730	387627	6.017599e+06	85.5	0.111588	20.0	5.0	0.0011
5	15	바레인	161182	1701575	274264	2.535900e+06	85.4	0.161182	5.0	6.0	0.0001
6	16	카타르	81785	2881053	235626	4.658268e+06	83.3	0.081785	44.0	7.0	0.0001
7	7	이스라엘	142297	8655535	1231660	1.473209e+07	81.4	0.142297	11.0	8.0	0.0001
8	5	칠레	86196	19116201	1647746	3.078569e+07	80.6	0.086196	38.0	9.0	0.0011
9	20	싱가포르	13299	5850342	77804	9.104751e+06	79.8	0.013299	111.0	10.0	0.0001
10	10	쿠바	71505	11326616	809911	1.776627e+07	78.4	0.071505	56.0	11.0	0.0001
11	18	중국	67	1439323776	95738	2.174043e+09	77.6	0.000067	176.0	12.0	0.0001

index	국가명	발생률	인구수	확진자_합계	백신접종	접종률(인구)	확진자비율_전체인구	발생률_순위	접종률_순위	사망률	
12	23	아이슬란드	33419	341243	11404	5.477880e+05	76.7	0.033419	92.0	13.0	0.0000
13	24	맨섬	84273	85033	7166	1.289150e+05	75.8	0.084273	42.0	14.0	0.0000
14	9	포르투갈	104156	10196709	1062048	1.555099e+07	75.7	0.104156	26.0	15.0	0.0011
15	14	덴마크	61274	5792202	354913	8.773849e+06	75.6	0.061274	65.0	16.0	0.0000
16	21	세이셸	212950	98347	20943	1.434900e+05	74.7	0.212950	1.0	17.0	0.0011
17	2	스페인	105434	46754778	4929546	6.909466e+07	74.3	0.105434	24.0	18.0	0.0011
18	6	캐나다	41687	37742154	1573359	5.506036e+07	73.4	0.041687	86.0	19.0	0.0000
19	13	아일랜드	76019	4937786	375367	7.119653e+06	71.9	0.076019	51.0	20.0	0.0011
20	8	벨기에	105251	11589623	1219814	1.644224e+07	71.8	0.105251	25.0	21.0	0.0021
21	1	프랑스	106557	65273511	6955333	9.226241e+07	71.2	0.106557	23.0	22.0	0.0011
22	17	노르웨이	33664	5421241	182501	7.586733e+06	70.8	0.033664	91.0	23.0	0.0000
23	28	부탄	3366	771608	2597	1.046117e+06	70.2	0.003366	141.0	24.0	0.0000
24	0	영국	109444	67886011	7429746	9.300209e+07	69.6	0.109444	21.0	25.0	0.0011
25	26	산마리노	159706	33931	5419	4.691700e+04	69.0	0.159706	6.0	26.0	0.0021
26	22	아루바	142826	106766	15249	1.544470e+05	68.9	0.142826	10.0	27.0	0.0011

	index	국 가 명	발생률	인구수	확진자_ 합계	백신접종	접종 률(인 구)	확진자비 율_전체 인구	발생 률_순 위	접종 률_순 위	사망 률
27	27	버 뮤 다	71887	62278	4477	8.619400e+04	68.4	0.071887	55.0	28.0	0.0001
28	3	이 탈 리 아	76678	60461826	4636111	8.237767e+07	68.2	0.076678	49.0	29.0	0.0021
29	4	네 덜 란 드	115918	17134872	1986241	2.332039e+07	67.7	0.115918	17.0	30.0	0.0011

상관관계 확인하기

In [173]:



```
corona_top30_part_sorted[['접종률(인구)', '확진자비율_전체인구']].corr()
```

Out[173]:

	접종률(인구)	확진자비율_전체인구
접종률(인구)	1.000000	0.198458
확진자비율_전체인구	0.198458	1.000000

In [174]:



```
corona_top30_part_sorted[['접종률(인구)', '사망률']].corr()
```

Out[174]:

	접종률(인구)	사망률
접종률(인구)	1.000000	0.017558
사망률	0.017558	1.000000

In [175]:



```
corona_top30_part_sorted[['접종률(인구)', '위중증률']].corr()
```

Out[175]:

	접종률(인구)	위중증률
접종률(인구)	1.000000	0.075372
위중증률	0.075372	1.000000

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

실습 : 위증증 데이터는 처리가 안되었다. 이를 처리해 보자.

In []:



03. 200만 이상의 국가에 대해서 확인

In [176]:



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

Out[176]:

(136, 26)

In [177]:



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

Out[177]:

	index	국가명	eng_code	발생률	인구수	확진자_합계	백신접종	접종률(인구)	발생률_순위	접종률_순위	확진자비율_전체인구	1차접종
0	0	미국	U.S.	129632	331002651	42908619	3.855860e+08	60.2	13.0	46.0	0.129632	63.8
1	1	인도	India	24277	1380004385	33483170	8.083893e+08	29.6	101.0	99.0	0.024263	43.7
2	2	브라질	Brazil	99924	212559417	21239783	2.223227e+08	53.9	27.0	62.0	0.099924	69.4
3	3	영국	U.K.	109444	67886011	7429746	9.300209e+07	69.6	21.0	25.0	0.109444	72.7

In [178]:



```
corona_sel = ['국가명', '발생률', '인구수', '확진자_합계', '백신접종', '접종률(인구)', '확진자비율_전체인구', '발생률_순위', '접종률_순위', '사망률', '위중증_수2', '위중증_증감', '위중증률']
corona_200m_part = corona_200m[corona_sel]
corona_200m_part
```

Out [178]:

	국 가 명	발생률	인구수	확진자_합계	백신접종	접종률 (인구)	확진자비율_전체인구	발생률_순위	접종률_순위	사망률	위중증_수2	위중증_증감	
0	미국	129632	331002651	42908619	3.855860e+08	60.2	0.129632	13.0	46.0	0.002091	24327	-523	7
1	인도	24277	1380004385	33483170	8.083893e+08	29.6	0.024263	101.0	99.0	0.000323	8944	None	6
2	브라질	99924	212559417	21239783	2.223227e+08	53.9	0.099924	27.0	62.0	0.002779	8318	None	3
3	영국	109444	67886011	7429746	9.300209e+07	69.6	0.109444	21.0	25.0	0.001992	1020	None	1
..	..												

In [179]:



```
corona_200m_part_sorted = corona_200m_part.sort_values(by=['접종률(인구)'], ascending=False)
corona_200m_part_sorted
```

Out [179]:

	국 가 명	발생률	인구수	확진자_합계	백신접종	접종률 (인구)	확진자비율_전체인구	발생률_순위	접종률_순위	사망률	위중증_수2	위중증_증감	
42	아랍에미리트	74113	9890402	733003	1.944587e+07	90.5	0.074113	52.0	3.0	0.000210	-99999	NaN	-1
64	우루과이	111588	3473730	387627	6.017599e+06	85.5	0.111588	20.0	5.0	0.001741	11	None	3

상관관계 확인하기

In [180]:



```
corona_200m_part_sorted[['접종률(인구)', '확진자비율_전체인구']].corr()
```

Out [180]:

	접종률(인구)	확진자비율_전체인구
접종률(인구)	1.000000	0.608814
확진자비율_전체인구	0.608814	1.000000

In [181]:



```
corona_200m_part_sorted[['접종률(인구)', '사망률']].corr()
```

Out [181]:

	접종률(인구)	사망률
접종률(인구)	1.000000	0.37719
사망률	0.37719	1.000000

In [182]:



```
corona_200m_part_sorted[['접종률(인구)', '위중증률']].corr()
```

Out [182]:

	접종률(인구)	위중증률
접종률(인구)	1.000000	0.187706
위중증률	0.187706	1.000000

결과 확인

01. 인구 200만 이상의 국가의 백신 접종률과 인구당 확진자 비율은 양의 상관관계(0.6088)의 관계를 갖는다.

02. 인구 200만 이상의 국가의 백신 접종률과 사망률은 양의 상관관계(0.37719)의 관계를 갖는다.

03. 인구 200만 이상의 국가의 백신 접종률과 위중증률은 양의 상관관계(0.187)의 관계를 갖는다.(관계가 크지 않음)

실습해 보기

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

In []:

