Working on Real Project with Python on 'COVID-19 Dataset'

(A part of Big Data Analysis)

COVID-19 SMALL DATASET

We have taken a small dataset of Covid-19, just for your understaning purpose. You have to work on the original dataset which contains about 19000 rows.

The data used here is till 29-April-2020 and has records as on 29-April-2020.

This data is available as a CSV file, downloaded from Kaggle.

We will analyze this data using the Pandas DataFrame.

```
import pandas as pd
```

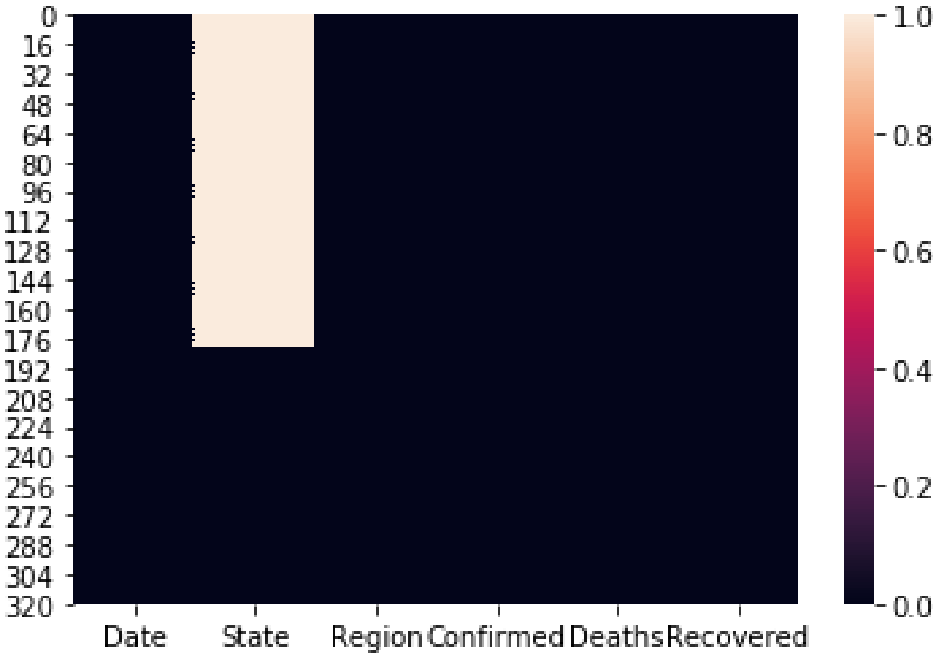
```
----------------------------------------------------------------------
-----
ModuleNotFoundError                       Traceback (most recent call
last)
Cell In[1], line 1
----> 1 import pandas as pd

ModuleNotFoundError: No module named 'pandas'
```

```
data = pd.read_csv(r"C:\Users\91999\Desktop\Youtube Video\DSL\Videos\
12. Covid 19 Project 4\covid_19_data.csv")
```

```
data
```

```
          Date     State          Region  Confirmed  Deaths  Recovered
0    4/29/2020       NaN     Afghanistan       1939      60        252
1    4/29/2020       NaN         Albania        766      30        455
2    4/29/2020       NaN         Algeria       3848     444       1702
3    4/29/2020       NaN         Andorra        743      42        423
4    4/29/2020       NaN          Angola         27       2          7
..         ...       ...             ...        ...     ...        ...
316  4/29/2020   Wyoming              US        545       7          0
317  4/29/2020  Xinjiang  Mainland China         76       3         73
318  4/29/2020     Yukon          Canada         11       0          0
319  4/29/2020    Yunnan  Mainland China        185       2        181
```

```
320  4/29/2020  Zhejiang  Mainland China       1268         1       1263

[321 rows x 6 columns]

# 1.
# df.count()
# df.isnull().sum()

data.count()

# Null Values means Missing Values

Date         321
State        140
Region       321
Confirmed    321
Deaths       321
Recovered    321
dtype: int64

data.isnull().sum()

Date           0
State        181
Region         0
Confirmed      0
Deaths         0
Recovered      0
dtype: int64

# 2.
# import seaborn as sns
# import matplotlib.pyplot as plt
# sns.heatmap(df.isnull())
# plt.show()

import seaborn as sns

import matplotlib.pyplot as plt

sns.heatmap(data.isnull())
plt.show()
```



Q.1) Show the number of Confirmed , Deaths and Recovered cases in each Region.

```
#df.groupby('Region').sum().head(50)
#df.groupby('Region')
['Confirmed'].sum().sort_values(ascending=False).head(20)
#df.groupby('Region')['Confirmed', 'Recovered'].sum()

data.head(2)
```

```
        Date State       Region  Confirmed  Deaths  Recovered
0  4/29/2020   NaN  Afghanistan       1939      60        252
1  4/29/2020   NaN      Albania        766      30        455
```

```
#data.groupby('Region').sum().head(20)

#data.groupby('Region')['Confirmed'].sum().sort_values(ascending =
False).head(10)

data.groupby('Region')['Confirmed', 'Recovered'].sum()
```

```
                    Confirmed  Recovered
Region
Afghanistan              1939        252
Albania                   766        455
Algeria                  3848       1702
```

```
Andorra                    743        423
Angola                      27          7
...                        ...        ...
West Bank and Gaza         344         71
Western Sahara               6          5
Yemen                        6          1
Zambia                      97         54
Zimbabwe                    32          5

[187 rows x 2 columns]
```

Q2) Remove all the records where Confirmed Cases is Less Than 10.

```
#df.Confirmed < 10
#df[df.Confirmed < 10]
#df[~(df.Confirmed < 10)]
#df = df[~(df.Confirmed < 10)]

data.head(2)

        Date State       Region  Confirmed  Deaths  Recovered
0  4/29/2020   NaN  Afghanistan       1939      60        252
1  4/29/2020   NaN      Albania        766      30        455

data = data[~(data.Confirmed < 10)] # To remove the records satisfying
a particular condition

#data.head(20)
```

Q.3) In which Region, maximum number of Confirmed cases were recorded ?

```
#df.groupby('Region').Confirmed.sum().sort_values(ascending =
False).head(20)

data.head(2)

        Date State       Region  Confirmed  Deaths  Recovered
0  4/29/2020   NaN  Afghanistan       1939      60        252
1  4/29/2020   NaN      Albania        766      30        455

data.groupby('Region').Confirmed.sum().sort_values(ascending =
False).head(20)

Region
US                     1039909
Spain                   236899
```

```
Italy              203591
France             166543
UK                 166441
Germany            161539
Turkey             117589
Russia              99399
Iran                93657
Mainland China      82862
Brazil              79685
Canada              52865
Belgium             47859
Netherlands         38998
Peru                33931
India               33062
Switzerland         29407
Ecuador             24675
Portugal            24505
Saudi Arabia        21402
Name: Confirmed, dtype: int64
```

Q.4) In which Region, minimum number of Deaths cases were recorded ?

```
#df.groupby('Region').Deaths.sum().sort_values(ascending =
True).head(50)

data.head(2)

        Date State        Region  Confirmed  Deaths  Recovered
0  4/29/2020   NaN  Afghanistan       1939      60        252
1  4/29/2020   NaN      Albania        766      30        455

data.groupby('Region').Deaths.sum().sort_values(ascending =
True).head(50)

Region
Laos                                0
Mongolia                            0
Mozambique                          0
Cambodia                            0
Fiji                                0
Namibia                             0
Nepal                               0
Madagascar                          0
Macau                               0
Papua New Guinea                    0
Rwanda                              0
Saint Kitts and Nevis               0
Bhutan                              0
Dominica                            0
```

```
Central African Republic            0
Saint Lucia                         0
Holy See                            0
Sao Tome and Principe               0
Yemen                               0
Western Sahara                      0
Eritrea                             0
Vietnam                             0
Saint Vincent and the Grenadines    0
Timor-Leste                         0
Uganda                              0
Grenada                             0
South Sudan                         0
Seychelles                          0
Liechtenstein                       1
Maldives                            1
Gambia                              1
Eswatini                            1
Guinea-Bissau                       1
Equatorial Guinea                   1
Mauritania                          1
Cabo Verde                          1
Benin                               1
Burundi                             1
Suriname                            1
Brunei                              1
Botswana                            1
West Bank and Gaza                  2
Angola                              2
Belize                              2
Djibouti                            2
Chad                                2
Libya                               2
MS Zaandam                          2
Nicaragua                           3
Syria                               3
Name: Deaths, dtype: int64
```

Q.5) How many Confirmed , Deaths & Recovered cases were reported from India till 29 April 2020 ?

```
#df[df.Region == 'Country_name']

data.head(2)
```

```
        Date State       Region  Confirmed  Deaths  Recovered
0  4/29/2020   NaN  Afghanistan       1939      60        252
1  4/29/2020   NaN      Albania        766      30        455
```

```
data[data.Region == 'India']
```

```
         Date State Region  Confirmed  Deaths  Recovered
74  4/29/2020   NaN  India      33062    1079       8437
```

```
data[data.Region == 'Yemen']
```

```
          Date State Region  Confirmed  Deaths  Recovered
178  4/29/2020   NaN  Yemen          6       0          1
```

```
data[data.Region == 'US']
```

```
          Date                         State Region  Confirmed  Deaths
\
181  4/29/2020                       Alabama     US       6912     256

182  4/29/2020                        Alaska     US        355       9

186  4/29/2020                       Arizona     US       7209     308

187  4/29/2020                      Arkansas     US       3193      57

195  4/29/2020                    California     US      48747    1946

199  4/29/2020                      Colorado     US      14758     766

200  4/29/2020                   Connecticut     US      26767    2169

202  4/29/2020                      Delaware     US       4655     144

204  4/29/2020  Diamond Princess cruise ship     US         49       0

205  4/29/2020          District of Columbia     US       4106     205

208  4/29/2020                       Florida     US      33193    1218

213  4/29/2020                       Georgia     US      25775    1101

216  4/29/2020                Grand Princess     US        103       3

219  4/29/2020                          Guam     US        141       5

224  4/29/2020                        Hawaii     US        613      16

231  4/29/2020                         Idaho     US       1952      60

232  4/29/2020                      Illinois     US      50358    2215

233  4/29/2020                       Indiana     US      17182     964

235  4/29/2020                          Iowa     US       6843     148
```

```
240  4/29/2020                        Kansas     US       3839     134
241  4/29/2020                      Kentucky     US       4537     234
243  4/29/2020                     Louisiana     US      27660    1845
245  4/29/2020                         Maine     US       1056      52
248  4/29/2020                      Maryland     US      20849    1078
249  4/29/2020                 Massachusetts     US      60265    3405
251  4/29/2020                      Michigan     US      40399    3670
252  4/29/2020                     Minnesota     US       4644     319
253  4/29/2020                   Mississippi     US       6569     250
254  4/29/2020                      Missouri     US       7660     338
255  4/29/2020                       Montana     US        451      16
257  4/29/2020                      Nebraska     US       3851      56
258  4/29/2020                        Nevada     US       4934     230
261  4/29/2020                 New Hampshire     US       2058      60
262  4/29/2020                    New Jersey     US     116365    6771
263  4/29/2020                    New Mexico     US       3213     112
265  4/29/2020                      New York     US     299691   23477
268  4/29/2020                North Carolina     US      10180     382
269  4/29/2020                  North Dakota     US       1033      19
270  4/29/2020      Northern Mariana Islands     US         14       2
274  4/29/2020                          Ohio     US      17303     937
275  4/29/2020                      Oklahoma     US       3473     214
277  4/29/2020                        Oregon     US       2446     101
278  4/29/2020                  Pennsylvania     US      46327    2373
280  4/29/2020                   Puerto Rico     US       1433      86
285  4/29/2020                     Recovered     US          0       0
```

```
287  4/29/2020                  Rhode Island      US      8247      251

298  4/29/2020                South Carolina      US      5882      231

299  4/29/2020                  South Dakota      US      2373       13

302  4/29/2020                     Tennessee      US     10366      195

303  4/29/2020                         Texas      US     27257      754

307  4/29/2020                          Utah      US      4497       45

308  4/29/2020                       Vermont      US       862       47

310  4/29/2020                Virgin Islands      US        57        4

311  4/29/2020                      Virginia      US     14962      522

312  4/29/2020                    Washington      US     14070      801

313  4/29/2020                 West Virginia      US      1110       38

315  4/29/2020                     Wisconsin      US      6520      308

316  4/29/2020                       Wyoming      US       545        7


     Recovered
181          0
182          0
186          0
187          0
195          0
199          0
200          0
202          0
204          0
205          0
208          0
213          0
216          0
219          0
224          0
231          0
232          0
233          0
235          0
240          0
241          0
243          0
245          0
```

```
248          0
249          0
251          0
252          0
253          0
254          0
255          0
257          0
258          0
261          0
262          0
263          0
265          0
268          0
269          0
270          0
274          0
275          0
277          0
278          0
280          0
285     120720
287          0
298          0
299          0
302          0
303          0
307          0
308          0
310          0
311          0
312          0
313          0
315          0
316          0
```

Q. 6-A) Sort the entire data wrt No. of Confirmed cases in ascending order.

```
#df.sort_values(by= ['Confirmed'] , ascending = True)

data.head(2)
```

```
        Date State        Region  Confirmed  Deaths  Recovered
0  4/29/2020   NaN   Afghanistan       1939      60        252
1  4/29/2020   NaN       Albania        766      30        455
```

```
data.sort_values( by = ['Confirmed'] , ascending = True).head(50)
```

```
          Date                             State  \
285  4/29/2020                         Recovered   
284  4/29/2020                         Recovered   
203  4/29/2020      Diamond Princess cruise ship   
305  4/29/2020                             Tibet   
289  4/29/2020         Saint Pierre and Miquelon   
184  4/29/2020                          Anguilla   
192  4/29/2020  Bonaire, Sint Eustatius and Saba   
272  4/29/2020             Northwest Territories   
288  4/29/2020                  Saint Barthelemy   
178  4/29/2020                               NaN   
194  4/29/2020            British Virgin Islands   
177  4/29/2020                               NaN   
18   4/29/2020                               NaN   
126  4/29/2020                               NaN   
140  4/29/2020                               NaN   
105  4/29/2020                               NaN   
98   4/29/2020                               NaN   
156  4/29/2020                               NaN   
70   4/29/2020                               NaN   
59   4/29/2020                               NaN   
144  4/29/2020                               NaN   
27   4/29/2020                               NaN   
256  4/29/2020                        Montserrat   
318  4/29/2020                             Yukon   
217  4/29/2020                         Greenland   
306  4/29/2020          Turks and Caicos Islands   
206  4/29/2020       Falkland Islands (Malvinas)   
118  4/29/2020                               NaN   
215  4/29/2020                    Grand Princess   
270  4/29/2020          Northern Mariana Islands   
136  4/29/2020                               NaN   
45   4/29/2020                               NaN   
138  4/29/2020                               NaN   
114  4/29/2020                               NaN   
201  4/29/2020                           Curacao   
137  4/29/2020                               NaN   
260  4/29/2020                     New Caledonia   
281  4/29/2020                           Qinghai   
55   4/29/2020                               NaN   
16   4/29/2020                               NaN   
90   4/29/2020                               NaN   
64   4/29/2020                               NaN   
21   4/29/2020                               NaN   
5    4/29/2020                               NaN   
163  4/29/2020                               NaN   
279  4/29/2020              Prince Edward Island   
4    4/29/2020                               NaN   
271  4/29/2020                Northern Territory   
180  4/29/2020                               NaN   
```

```
152  4/29/2020                                 NaN

                                   Region  Confirmed  Deaths  Recovered
285                                    US          0       0     120720
284                                Canada          0       0      20327
203                                Canada          0       1          0
305                         Mainland China          1       0          1
289                                France          1       0          0
184                                    UK          3       0          3
192                           Netherlands          5       0          0
272                                Canada          5       0          0
288                                France          6       0          6
178                                 Yemen          6       0          1
194                                    UK          6       1          3
177                        Western Sahara          6       0          5
18                                 Bhutan          7       0          5
126                      Papua New Guinea          8       0          0
140                 Sao Tome and Principe          8       0          4
105                            Mauritania          8       1          6
98                             MS Zaandam          9       2          0
156                              Suriname         10       1          8
70                               Holy See         10       0          2
59                                 Gambia         10       1          8
144                            Seychelles         11       0          6
27                                Burundi         11       1          4
256                                    UK         11       1          2
318                                Canada         11       0          0
217                               Denmark         11       0         11
306                                    UK         12       1          5
206                                    UK         13       0         11
118                             Nicaragua         13       3          7
215                                Canada         13       0          0
270                                    US         14       2          0
136                 Saint Kitts and Nevis         15       0          4
45                               Dominica         16       0         13
138  Saint Vincent and the Grenadines         16       0          8
114                               Namibia         16       0          8
201                           Netherlands         16       1         13
137                           Saint Lucia         17       0         15
260                                France         18       0         17
281                        Mainland China         18       0         18
55                                   Fiji         18       0         12
16                                 Belize         18       2          9
90                                   Laos         19       0          7
64                                Grenada         20       0         13
21                               Botswana         23       1          5
5                     Antigua and Barbuda         24       3         11
163                           Timor-Leste         24       0          6
279                                Canada         27       0          0
4                                  Angola         27       2          7
```

```
271                          Australia         28       0         25
180                           Zimbabwe         32       4          5
152                        South Sudan         34       0          0
```

Q. 6-B) Sort the entire data wrt No. of Recovered cases in descending order.

```
#df.sort_values(by= ['Recovered'] , ascending = False)

data.sort_values( by = ['Recovered'] , ascending = False).head(50)
```

```
          Date            State                Region  Confirmed
Deaths  \
153  4/29/2020              NaN                 Spain     236899
24275
285  4/29/2020        Recovered                    US          0
0
61   4/29/2020              NaN               Germany     161539
6467
76   4/29/2020              NaN                  Iran      93657
5957
80   4/29/2020              NaN                 Italy     203591
27682
229  4/29/2020            Hubei        Mainland China      68128
4512
57   4/29/2020              NaN                France     165093
24087
167  4/29/2020              NaN                Turkey     117589
3081
22   4/29/2020              NaN                Brazil      79685
5513
158  4/29/2020              NaN           Switzerland      29407
1716
284  4/29/2020        Recovered                Canada          0
0
78   4/29/2020              NaN               Ireland      20253
1190
8    4/29/2020              NaN               Austria      15402
580
107  4/29/2020              NaN                Mexico      17799
1732
15   4/29/2020              NaN               Belgium      47859
7501
134  4/29/2020              NaN                Russia      99399
972
128  4/29/2020              NaN                  Peru      33931
943
151  4/29/2020              NaN           South Korea      10765
247
```

```
74   4/29/2020              NaN                   India      33062
1079
79   4/29/2020              NaN                  Israel      15834
215
33   4/29/2020              NaN                   Chile      14885
216
42   4/29/2020              NaN                 Denmark       9008
443
101  4/29/2020              NaN                Malaysia       5945
100
133  4/29/2020              NaN                 Romania      11978
693
124  4/29/2020              NaN                Pakistan      15525
343
97   4/29/2020              NaN              Luxembourg       3769
89
41   4/29/2020              NaN          Czech Republic       7579
227
130  4/29/2020              NaN                  Poland      12640
624
141  4/29/2020              NaN            Saudi Arabia      21402
157
56   4/29/2020              NaN                 Finland       4906
206
162  4/29/2020              NaN                Thailand       2947
54
83   4/29/2020              NaN                   Japan      13895
413
171  4/29/2020              NaN    United Arab Emirates      11929
98
264  4/29/2020  New South Wales               Australia       3016
40
150  4/29/2020              NaN            South Africa       5350
103
14   4/29/2020              NaN                 Belarus      13181
84
2    4/29/2020              NaN                 Algeria       3848
444
73   4/29/2020              NaN                 Iceland       1797
10
220  4/29/2020        Guangdong          Mainland China       1588
8
47   4/29/2020              NaN                 Ecuador      24675
883
131  4/29/2020              NaN                Portugal      24505
973
11   4/29/2020              NaN                 Bahrain       2921
8
34   4/29/2020              NaN                Colombia       6207
```

```
278
75   4/29/2020              NaN             Indonesia       9771
784
88   4/29/2020              NaN                Kuwait       3740
24
77   4/29/2020              NaN                  Iraq       2003
92
48   4/29/2020              NaN                 Egypt       5268
380
309  4/29/2020         Victoria             Australia       1361
18
38   4/29/2020              NaN               Croatia       2062
67
9    4/29/2020              NaN            Azerbaijan       1766
23

     Recovered
153     132929
285     120720
61      120400
76       73791
80       71252
229      63616
57       48228
167      44040
22       34132
158      22600
284      20327
78       13386
8        12779
107      11423
15       11283
134      10286
128      10037
151       9059
74        8437
79        8233
33        8057
42        6366
101       4087
133       3569
124       3425
97        3134
41        3108
130       3025
141       2953
56        2800
162       2665
83        2368
```

```
171       2329
264       2284
150       2073
14        2072
2         1702
73        1656
220       1557
47        1557
131       1470
11        1455
34        1411
75        1391
88        1389
77        1346
48        1335
309       1291
38        1288
9         1267
```