

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

(A part of Big Data Analysis)

COVID-19 SMALL DATASET

```
In [1]: ▶ import pandas as pd
```

```
In [2]: ▶ data = pd.read_csv(r"C:\Users\KIIT\Desktop\Projects\Covid 19\Covid_19_data.csv")
```

```
In [3]: ▶ data
```

Out[3]:

	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 × 6 columns

```
In [4]: data.count()
```

```
Out[4]: Date      321  
State    140  
Region   321  
Confirmed 321  
Deaths   321  
Recovered 321  
dtype: int64
```

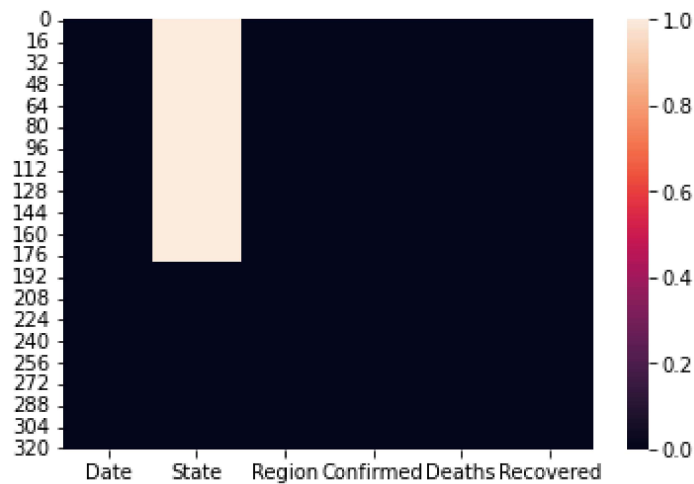
```
In [5]: data.isnull().sum()
```

```
Out[5]: Date      0  
State    181  
Region    0  
Confirmed  0  
Deaths     0  
Recovered  0  
dtype: int64
```

```
In [6]: import seaborn as sns
```

```
In [7]: import matplotlib.pyplot as plt
```

```
In [8]: sns.heatmap(data.isnull())  
plt.show()
```



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

In [9]: `data.head(2)`

Out[9]:

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

In [10]: `data.groupby('Region')['Confirmed', 'Recovered'].sum()`

C:\Users\KIIT\AppData\Local\Temp\ipykernel_5748\581960954.py:1: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.

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

Out[10]:

	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 × 2 columns

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

In [11]: `data.head(2)`

Out[11]:

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

```
In [12]: data = data[~(data.Confirmed < 10)]
```

```
In [13]: data.head(20)
```

Out[13]:

	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
5	4/29/2020	NaN	Antigua and Barbuda	24	3	11
6	4/29/2020	NaN	Argentina	4285	214	1192
7	4/29/2020	NaN	Armenia	1932	30	900
8	4/29/2020	NaN	Austria	15402	580	12779
9	4/29/2020	NaN	Azerbaijan	1766	23	1267
10	4/29/2020	NaN	Bahamas	80	11	23
11	4/29/2020	NaN	Bahrain	2921	8	1455
12	4/29/2020	NaN	Bangladesh	7103	163	150
13	4/29/2020	NaN	Barbados	80	7	39
14	4/29/2020	NaN	Belarus	13181	84	2072
15	4/29/2020	NaN	Belgium	47859	7501	11283
16	4/29/2020	NaN	Belize	18	2	9
17	4/29/2020	NaN	Benin	64	1	33
19	4/29/2020	NaN	Bolivia	1110	59	117
20	4/29/2020	NaN	Bosnia and Herzegovina	1677	65	710

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

```
In [14]: data.head(2)
```

Out[14]:

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

```
In [15]: data.groupby('Region').Confirmed.sum().sort_values(ascending = False).head(20)
```

```
Out[15]: Region
US          1039909
Spain       236899
Italy       203591
France      166536
UK          166432
Germany     161539
Turkey     117589
Russia      99399
Iran        93657
Mainland China 82861
Brazil      79685
Canada      52860
Belgium     47859
Netherlands 38993
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 ?

```
In [16]: data.head(2)
```

```
Out[16]:
```

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

```
In [17]: data.groupby('Region').Deaths.sum().sort_values(ascending = True).head(50)
```

```
Out[17]: Region
Cambodia                0
Seychelles              0
Saint Lucia             0
Central African Republic 0
Saint Kitts and Nevis   0
South Sudan            0
Rwanda                 0
Grenada                0
Macau                 0
Madagascar            0
Nepal                 0
Namibia               0
Saint Vincent and the Grenadines 0
Mozambique            0
Holy See              0
Timor-Leste           0
Mongolia              0
Uganda               0
Laos                 0
Eritrea              0
Vietnam              0
Fiji                0
Dominica            0
Gambia             1
Equatorial Guinea   1
Eswatini            1
Cabo Verde         1
Maldives           1
Guinea-Bissau      1
Liechtenstein      1
Brunei             1
Burundi           1
Botswana           1
Suriname           1
Benin              1
Djibouti           2
Angola             2
Libya              2
Chad               2
West Bank and Gaza 2
Belize             2
Zambia             3
Malawi             3
Nicaragua          3
Syria              3
Ethiopia           3
Antigua and Barbuda 3
Gabon              3
Hong Kong          4
Zimbabwe           4
Name: Deaths, dtype: int64
```

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

In [18]: `data.head(2)`

Out[18]:

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

In [19]: `data[data.Region == 'India']`


Out[19]:

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

In [20]: `data[data.Region == 'Yemen']`

Out[20]:

	Date	State	Region	Confirmed	Deaths	Recovered
--	------	-------	--------	-----------	--------	-----------

In [377]:  data[data.Region == 'US']

Out[377]:

	Date	State	Region	Confirmed	Deaths	Recovered
181	4/29/2020	Alabama	US	6912	256	0
182	4/29/2020	Alaska	US	355	9	0
186	4/29/2020	Arizona	US	7209	308	0
187	4/29/2020	Arkansas	US	3193	57	0
195	4/29/2020	California	US	48747	1946	0
199	4/29/2020	Colorado	US	14758	766	0
200	4/29/2020	Connecticut	US	26767	2169	0
202	4/29/2020	Delaware	US	4655	144	0
204	4/29/2020	Diamond Princess cruise ship	US	49	0	0
205	4/29/2020	District of Columbia	US	4106	205	0
208	4/29/2020	Florida	US	33193	1218	0
213	4/29/2020	Georgia	US	25775	1101	0
216	4/29/2020	Grand Princess	US	103	3	0
219	4/29/2020	Guam	US	141	5	0
224	4/29/2020	Hawaii	US	613	16	0
231	4/29/2020	Idaho	US	1952	60	0
232	4/29/2020	Illinois	US	50358	2215	0
233	4/29/2020	Indiana	US	17182	964	0
235	4/29/2020	Iowa	US	6843	148	0
240	4/29/2020	Kansas	US	3839	134	0
241	4/29/2020	Kentucky	US	4537	234	0
243	4/29/2020	Louisiana	US	27660	1845	0
245	4/29/2020	Maine	US	1056	52	0
248	4/29/2020	Maryland	US	20849	1078	0
249	4/29/2020	Massachusetts	US	60265	3405	0
251	4/29/2020	Michigan	US	40399	3670	0
252	4/29/2020	Minnesota	US	4644	319	0
253	4/29/2020	Mississippi	US	6569	250	0
254	4/29/2020	Missouri	US	7660	338	0
255	4/29/2020	Montana	US	451	16	0
257	4/29/2020	Nebraska	US	3851	56	0
258	4/29/2020	Nevada	US	4934	230	0
261	4/29/2020	New Hampshire	US	2058	60	0
262	4/29/2020	New Jersey	US	116365	6771	0
263	4/29/2020	New Mexico	US	3213	112	0
265	4/29/2020	New York	US	299691	23477	0
268	4/29/2020	North Carolina	US	10180	382	0

	Date	State	Region	Confirmed	Deaths	Recovered
269	4/29/2020	North Dakota	US	1033	19	0
270	4/29/2020	Northern Mariana Islands	US	14	2	0
274	4/29/2020	Ohio	US	17303	937	0
275	4/29/2020	Oklahoma	US	3473	214	0
277	4/29/2020	Oregon	US	2446	101	0
278	4/29/2020	Pennsylvania	US	46327	2373	0
280	4/29/2020	Puerto Rico	US	1433	86	0
285	4/29/2020	Recovered	US	0	0	120720
287	4/29/2020	Rhode Island	US	8247	251	0
298	4/29/2020	South Carolina	US	5882	231	0
299	4/29/2020	South Dakota	US	2373	13	0
302	4/29/2020	Tennessee	US	10366	195	0
303	4/29/2020	Texas	US	27257	754	0
307	4/29/2020	Utah	US	4497	45	0
308	4/29/2020	Vermont	US	862	47	0
310	4/29/2020	Virgin Islands	US	57	4	0
311	4/29/2020	Virginia	US	14962	522	0
312	4/29/2020	Washington	US	14070	801	0
313	4/29/2020	West Virginia	US	1110	38	0
315	4/29/2020	Wisconsin	US	6520	308	0
316	4/29/2020	Wyoming	US	545	7	0

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

In [21]: `data.head(2)`

Out[21]:

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

```
In [22]: data.sort_values( by = ['Confirmed'] , ascending = True).head(50)
```

Out[22]:

	Date	State	Region	Confirmed	Deaths	Recovered
156	4/29/2020	NaN	Suriname	10	1	8
70	4/29/2020	NaN	Holy See	10	0	2
59	4/29/2020	NaN	Gambia	10	1	8
318	4/29/2020	Yukon	Canada	11	0	0
217	4/29/2020	Greenland	Denmark	11	0	11
256	4/29/2020	Montserrat	UK	11	1	2
144	4/29/2020	NaN	Seychelles	11	0	6
27	4/29/2020	NaN	Burundi	11	1	4
306	4/29/2020	Turks and Caicos Islands	UK	12	1	5
118	4/29/2020	NaN	Nicaragua	13	3	7
215	4/29/2020	Grand Princess	Canada	13	0	0
206	4/29/2020	Falkland Islands (Malvinas)	UK	13	0	11
270	4/29/2020	Northern Mariana Islands	US	14	2	0
136	4/29/2020	NaN	Saint Kitts and Nevis	15	0	4
138	4/29/2020	NaN	Saint Vincent and the Grenadines	16	0	8
45	4/29/2020	NaN	Dominica	16	0	13
201	4/29/2020	Curacao	Netherlands	16	1	13
114	4/29/2020	NaN	Namibia	16	0	8
137	4/29/2020	NaN	Saint Lucia	17	0	15
260	4/29/2020	New Caledonia	France	18	0	17
16	4/29/2020	NaN	Belize	18	2	9
55	4/29/2020	NaN	Fiji	18	0	12
281	4/29/2020	Qinghai	Mainland China	18	0	18
90	4/29/2020	NaN	Laos	19	0	7
64	4/29/2020	NaN	Grenada	20	0	13
21	4/29/2020	NaN	Botswana	23	1	5
5	4/29/2020	NaN	Antigua and Barbuda	24	3	11
163	4/29/2020	NaN	Timor-Leste	24	0	6
279	4/29/2020	Prince Edward Island	Canada	27	0	0
4	4/29/2020	NaN	Angola	27	2	7
271	4/29/2020	Northern Territory	Australia	28	0	25
180	4/29/2020	NaN	Zimbabwe	32	4	5
152	4/29/2020	NaN	South Sudan	34	0	0
100	4/29/2020	NaN	Malawi	36	3	7

	Date	State	Region	Confirmed	Deaths	Recovered
300	4/29/2020	St Martin	France	38	3	24
110	4/29/2020	NaN	Mongolia	38	0	10
51	4/29/2020	NaN	Eritrea	39	0	19
159	4/29/2020	NaN	Syria	43	3	21
244	4/29/2020	Macau	Macau	45	0	34
204	4/29/2020	Diamond Princess cruise ship	US	49	0	0
31	4/29/2020	NaN	Central African Republic	50	0	10
32	4/29/2020	NaN	Chad	52	2	19
310	4/29/2020	Virgin Islands	US	57	4	0
115	4/29/2020	NaN	Nepal	57	0	16
210	4/29/2020	French Polynesia	France	58	0	50
94	4/29/2020	NaN	Libya	61	2	18
17	4/29/2020	NaN	Benin	64	1	33
196	4/29/2020	Cayman Islands	UK	73	1	10
296	4/29/2020	Sint Maarten	Netherlands	75	13	33
267	4/29/2020	Ningxia	Mainland China	75	0	75

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

```
In [23]: data.sort_values( by = ['Recovered'] , ascending = False).head(50)
```

Out[23]:

	Date	State	Region	Confirmed	Deaths	Recovered
153	4/29/2020	NaN	Spain	236899	24275	132929
61	4/29/2020	NaN	Germany	161539	6467	120400
76	4/29/2020	NaN	Iran	93657	5957	73791
80	4/29/2020	NaN	Italy	203591	27682	71252
229	4/29/2020	Hubei	Mainland China	68128	4512	63616
57	4/29/2020	NaN	France	165093	24087	48228
167	4/29/2020	NaN	Turkey	117589	3081	44040
22	4/29/2020	NaN	Brazil	79685	5513	34132
158	4/29/2020	NaN	Switzerland	29407	1716	22600
78	4/29/2020	NaN	Ireland	20253	1190	13386
8	4/29/2020	NaN	Austria	15402	580	12779
107	4/29/2020	NaN	Mexico	17799	1732	11423
15	4/29/2020	NaN	Belgium	47859	7501	11283
134	4/29/2020	NaN	Russia	99399	972	10286
128	4/29/2020	NaN	Peru	33931	943	10037
151	4/29/2020	NaN	South Korea	10765	247	9059
74	4/29/2020	NaN	India	33062	1079	8437
79	4/29/2020	NaN	Israel	15834	215	8233
33	4/29/2020	NaN	Chile	14885	216	8057
42	4/29/2020	NaN	Denmark	9008	443	6366
101	4/29/2020	NaN	Malaysia	5945	100	4087
133	4/29/2020	NaN	Romania	11978	693	3569
124	4/29/2020	NaN	Pakistan	15525	343	3425
97	4/29/2020	NaN	Luxembourg	3769	89	3134
41	4/29/2020	NaN	Czech Republic	7579	227	3108
130	4/29/2020	NaN	Poland	12640	624	3025
141	4/29/2020	NaN	Saudi Arabia	21402	157	2953
56	4/29/2020	NaN	Finland	4906	206	2800
162	4/29/2020	NaN	Thailand	2947	54	2665
83	4/29/2020	NaN	Japan	13895	413	2368
171	4/29/2020	NaN	United Arab Emirates	11929	98	2329
264	4/29/2020	New South Wales	Australia	3016	40	2284
150	4/29/2020	NaN	South Africa	5350	103	2073
14	4/29/2020	NaN	Belarus	13181	84	2072
2	4/29/2020	NaN	Algeria	3848	444	1702
73	4/29/2020	NaN	Iceland	1797	10	1656

	Date	State	Region	Confirmed	Deaths	Recovered
47	4/29/2020	NaN	Ecuador	24675	883	1557
220	4/29/2020	Guangdong	Mainland China	1588	8	1557
131	4/29/2020	NaN	Portugal	24505	973	1470
11	4/29/2020	NaN	Bahrain	2921	8	1455
34	4/29/2020	NaN	Colombia	6207	278	1411
75	4/29/2020	NaN	Indonesia	9771	784	1391
88	4/29/2020	NaN	Kuwait	3740	24	1389
77	4/29/2020	NaN	Iraq	2003	92	1346
48	4/29/2020	NaN	Egypt	5268	380	1335
309	4/29/2020	Victoria	Australia	1361	18	1291
38	4/29/2020	NaN	Croatia	2062	67	1288
9	4/29/2020	NaN	Azerbaijan	1766	23	1267
320	4/29/2020	Zhejiang	Mainland China	1268	1	1263
227	4/29/2020	Henan	Mainland China	1276	22	1254

By - Dipisha Shivangi

In []: ▶