

COVID-19 SMALL DATASET

I have taken a small dataset of Covid-19, just for understanding purpose. I have to work on the original dataset which contains 19000 rows.

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

The data is available as a CSV file, downloaded from kaggle.

I will analyze this data using the Pandas DataFrame.

In [2]:

```
import pandas as pd
```

In [3]:

```
data = pd.read_csv(r'C:\Users\harsh\Desktop\Projects\Python\Covid - 19 Dataset Analysis\
```

In [4]:

```
data
```

Out[4]:

	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()
```



Q1) 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').sum().head(20)
```

Out[10]:

	Confirmed	Deaths	Recovered
Region			
Afghanistan	1939	60	252
Albania	766	30	455
Algeria	3848	444	1702
Andorra	743	42	423
Angola	27	2	7
Antigua and Barbuda	24	3	11
Argentina	4285	214	1192
Armenia	1932	30	900
Australia	6752	91	5715
Austria	15402	580	12779
Azerbaijan	1766	23	1267
Bahamas	80	11	23
Bahrain	2921	8	1455
Bangladesh	7103	163	150
Barbados	80	7	39
Belarus	13181	84	2072
Belgium	47859	7501	11283
Belize	18	2	9
Benin	64	1	33
Bhutan	7	0	5

In [11]:

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

Out[11]:

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

In [12]:

```
data.groupby('Region')['Deaths'].sum().sort_values(ascending = False).head(20)
```

Out[12]:

Region	
US	60967
Italy	27682
UK	26166
Spain	24275
France	24121
Belgium	7501
Germany	6467
Iran	5957
Brazil	5513
Netherlands	4727
Mainland China	4633
Canada	3155
Turkey	3081
Sweden	2462
Mexico	1732
Switzerland	1716
Ireland	1190
India	1079
Portugal	973
Russia	972

Name: Deaths, dtype: int64

In [13]:

```
data.groupby('Region')['Recovered'].sum().sort_values(ascending = False).head(20)
```

Out[13]:

```
Region
Spain          132929
US              120720
Germany        120400
Mainland China  77610
Iran            73791
Italy           71252
France          49118
Turkey          44040
Brazil          34132
Switzerland     22600
Canada          20327
Ireland         13386
Austria         12779
Mexico          11423
Belgium         11283
Russia          10286
Peru            10037
South Korea      9059
India            8437
Israel           8233
Name: Recovered, dtype: int64
```

Q2) Remove all the records where Confirmed is less than 10.

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[data.Confirmed < 10]
```

Out[15]:

	Date	State	Region	Confirmed	Deaths	Recovered
18	4/29/2020	NaN	Bhutan	7	0	5
98	4/29/2020	NaN	MS Zaandam	9	2	0
105	4/29/2020	NaN	Mauritania	8	1	6
126	4/29/2020	NaN	Papua New Guinea	8	0	0
140	4/29/2020	NaN	Sao Tome and Principe	8	0	4
177	4/29/2020	NaN	Western Sahara	6	0	5
178	4/29/2020	NaN	Yemen	6	0	1
184	4/29/2020	Anguilla	UK	3	0	3
192	4/29/2020	Bonaire, Sint Eustatius and Saba	Netherlands	5	0	0
194	4/29/2020	British Virgin Islands	UK	6	1	3
203	4/29/2020	Diamond Princess cruise ship	Canada	0	1	0
272	4/29/2020	Northwest Territories	Canada	5	0	0
284	4/29/2020	Recovered	Canada	0	0	20327
285	4/29/2020	Recovered	US	0	0	120720
288	4/29/2020	Saint Barthelemy	France	6	0	6
289	4/29/2020	Saint Pierre and Miquelon	France	1	0	0
305	4/29/2020	Tibet	Mainland China	1	0	1

In [18]:

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

In [20]:

```
data[data.Confirmed < 10]
```

Out[20]:

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

Q3) In which region, maximum number of Confirmed cases were recorded?

In [7]:

```
data.head()
```

Out[7]:

	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

In [10]:

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

Out[10]:

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

Q4) In which region, minimum number of Deaths cases were recorded?

In [13]:

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

Out[13]:

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

Q5) How many Confirmed, Deaths and Recovered cases were reported from India till 29 April 2020?

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 [16]:

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

Out[16]:

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

Q6-A) Sort the entire data wrt No. of Confirmed cases in ascending order.

In [19]:

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

Out[19]:

	Date	State	Region	Confirmed	Deaths	Recovered
285	4/29/2020	Recovered	US	0	0	120720
284	4/29/2020	Recovered	Canada	0	0	20327
203	4/29/2020	Diamond Princess cruise ship	Canada	0	1	0
305	4/29/2020	Tibet	Mainland China	1	0	1
289	4/29/2020	Saint Pierre and Miquelon	France	1	0	0
184	4/29/2020	Anguilla	UK	3	0	3
192	4/29/2020	Bonaire, Sint Eustatius and Saba	Netherlands	5	0	0
272	4/29/2020	Northwest Territories	Canada	5	0	0
288	4/29/2020	Saint Barthelemy	France	6	0	6
178	4/29/2020	NaN	Yemen	6	0	1
194	4/29/2020	British Virgin Islands	UK	6	1	3
177	4/29/2020	NaN	Western Sahara	6	0	5
18	4/29/2020	NaN	Bhutan	7	0	5
126	4/29/2020	NaN	Papua New Guinea	8	0	0
140	4/29/2020	NaN	Sao Tome and Principe	8	0	4
105	4/29/2020	NaN	Mauritania	8	1	6
98	4/29/2020	NaN	MS Zaandam	9	2	0
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
144	4/29/2020	NaN	Seychelles	11	0	6
27	4/29/2020	NaN	Burundi	11	1	4
256	4/29/2020	Montserrat	UK	11	1	2
318	4/29/2020	Yukon	Canada	11	0	0
217	4/29/2020	Greenland	Denmark	11	0	11
306	4/29/2020	Turks and Caicos Islands	UK	12	1	5
206	4/29/2020	Falkland Islands (Malvinas)	UK	13	0	11
118	4/29/2020	NaN	Nicaragua	13	3	7
215	4/29/2020	Grand Princess	Canada	13	0	0
270	4/29/2020	Northern Mariana Islands	US	14	2	0
136	4/29/2020	NaN	Saint Kitts and Nevis	15	0	4
45	4/29/2020	NaN	Dominica	16	0	13
138	4/29/2020	NaN	Saint Vincent and the Grenadines	16	0	8

	Date	State	Region	Confirmed	Deaths	Recovered
114	4/29/2020	NaN	Namibia	16	0	8
201	4/29/2020	Curacao	Netherlands	16	1	13
137	4/29/2020	NaN	Saint Lucia	17	0	15
260	4/29/2020	New Caledonia	France	18	0	17
281	4/29/2020	Qinghai	Mainland China	18	0	18
55	4/29/2020	NaN	Fiji	18	0	12
16	4/29/2020	NaN	Belize	18	2	9
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

Q6-B) Sort the entire data wrt No. of Recovered cases in descending order.

In [22]:

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

Out[22]:

	Date	State	Region	Confirmed	Deaths	Recovered
153	4/29/2020	NaN	Spain	236899	24275	132929
285	4/29/2020	Recovered	US	0	0	120720
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
284	4/29/2020	Recovered	Canada	0	0	20327
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

	Date	State	Region	Confirmed	Deaths	Recovered
73	4/29/2020	NaN	Iceland	1797	10	1656
220	4/29/2020	Guangdong	Mainland China	1588	8	1557
47	4/29/2020	NaN	Ecuador	24675	883	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