

COVID-19 DATASET

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
%matplotlib inline

In [2]: data=pd.read_csv(r'C:\Users\mujjj\Downloads\4. 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.shape

Out[4]: (321, 6)

In [5]: data.count()

Out[5]: Date      321
State    140
Region   321
Confirmed 321
Deaths   321
Recovered 321
dtype: int64

In [6]: data.columns

Out[6]: Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object')

In [7]: data.isnull().sum()

Out[7]: Date      0
State    181
Region    0
Confirmed 0
Deaths    0
Recovered 0
dtype: int64

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

Out[8]: <function matplotlib.pyplot.show(close=None, block=None)>
```

Q:1) Show the number of confirmed deaths and recoverd ceases in each region

```
In [9]: data.columns

Out[9]: Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object')

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()

Out[11]: Region
Afghanistan      1939
Albania           766
Algeria          3848
Andorra           743
Angola            27
...
West Bank and Gaza      344
Western Sahara          6
Yemen                   6
Zambia                  97
Zimbabwe                32
Name: Confirmed, Length: 187, dtype: int64

In [12]: data.columns

Out[12]: Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object')

In [13]: data.groupby('Region')['Deaths'].sum().sort_values(ascending=False).head(10)

Out[13]: Region
US              60967
Italy           27682
UK              26166
Spain           24275
France          24121
Belgium         7501
Germany         6467
Iran            5957
Brazil          5513
Netherlands     4727
Name: Deaths, dtype: int64

In [15]: data.groupby('Region')['Confirmed','Recovered'].sum()

Out[15]:
```

	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

Remove all records where confirmed cases is less then 10

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

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
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 [17]: data[data.Confirmed<10]

Out[17]:
```

	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.Confirmed<10]

Out[18]:
```

	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 [19]: data=data[~(data.Confirmed<10)]

In [20]: data

Out[20]:
```

	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

304 rows × 6 columns

in which Region maximum number of confirmed cases were recorded

```
In [22]: data.groupby('Region').Confirmed.sum().sort_values(ascending=False)

Out[22]: Region
US              1039909
Spain           236899
Italy           203591
France          166536
UK              166432
...
Seychelles      11
Burundi         11
Suriname        10
Gambia          10
Holy See        10
Name: Confirmed, Length: 180, dtype: int64
```

in which Region minimum number of confirmed death cases were recorded

```
In [23]: data.columns

Out[23]: Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object')

In [26]: data.groupby('Region').Deaths.sum().sort_values(ascending=True).head(12)

Out[26]: 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
Name: Deaths, dtype: int64
```

how many confirmed, deaths and Recovered cases were reported from nepal till 29 April 2020

```
In [27]: data.head()

Out[27]:
```

	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 [28]: data[data.Region=='Nepal']

Out[28]:
```

	Date	State	Region	Confirmed	Deaths	Recovered
115	4/29/2020	NaN	Nepal	57	0	16

```
In [33]: data[data.Region=='US'].head(5)

Out[33]:
```

	Date	State	Region	Confirmed	Deaths	Recovered
181	4/29/2020	Alabama	US	6912	256	0
182	4/29/2020	Alabama	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

Sort the entire data wrt no of confirmed cases in ascending order

```
In [35]: data.sort_values(by = ['Confirmed'], ascending=True)

Out[35]:
```

	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
...
57	4/29/2020	NaN	France	165093	24087	48228
168	4/29/2020	NaN	UK	165221	26097	0
80	4/29/2020	NaN	Italy	203591	27682	71252
153	4/29/2020	NaN	Spain	236899	24275	132929
265	4/29/2020	New York	US	299691	23477	0

304 rows × 6 columns

Sort the entire data wrt no of Recovered cases in descending order

```
In [38]: data.columns

Out[38]: Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object')

In [40]: data.sort_values(by=['Recovered'],ascending=False)

Out[40]:
```

	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
...
258	4/29/2020	Nevada	US	4934	230	0
257	4/29/2020	Nebraska	US	3851	56	0
255	4/29/2020	Montana	US	451	16	0
254	4/29/2020	Missouri	US	7660	338	0
274	4/29/2020	Ohio	US	17303	937	0

304 rows × 6 columns

```
In [ ]:
```