In [1]:
```
import pandas as pd
```

In [2]:
```
data=pd.read_csv(r"C:\Users\Bhavya's Surface\Documents\pandas\covid19data.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

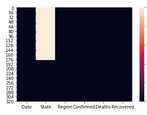
In [4]:
```
#to find out the null values and total
data.isnull().sum()
```

Out[4]:
```
Date           0
State        181
Region         0
Confirmed      0
Deaths         0
Recovered      0
dtype: int64
```

In [5]:
```
#to plot the null values in seaborn heatmap
import seaborn as sns
import matplotlib.pyplot as plt
```

In [6]:
```
sns.heatmap(data.isnull())
```

Out[6]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x122dc543190>
```



In [7]:
```
#no. of confirmed deaths and recovered cases in each region
data.groupby('Region').sum()
```

Out[7]:

Region	Confirmed	Deaths	Recovered
Afghanistan	1939	60	252
Albania	766	30	455
Algeria	3848	444	1702
Andorra	743	42	423
Angola	27	2	7
...	...	...	...
West Bank and Gaza	344	2	71
Western Sahara	6	0	5
Yemen	6	0	1
Zambia	97	3	54
Zimbabwe	32	4	5

187 rows × 3 columns

In [8]:
```
data.groupby('Region')['Recovered'].sum().sort_values(ascending=False).head()
```

Out[8]:
```
Region
Spain             132929
US                129720
Germany           120400
Mainland China     77610
Iran               73791
Name: Recovered, dtype: int64
```

In [9]:
```
data.groupby('Region')['Deaths'].sum().sort_values(ascending=False).head()
```

Out[9]:
```
Region
US         60967
Italy      27682
UK         26166
Spain      24275
France     24121
Name: Deaths, dtype: int64
```

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

```
<ipython-input-10-20fd7b835859>: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]:

Region	Confirmed	Recovered
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

In [11]:
```
#removing the records where the confirmed cases are less than 10
data=data[~(data.Confirmed<10)]
data
```

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
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 [12]:
```
#which region the max, no. of confirmed records are present
data.groupby('Region').Confirmed.sum().sort_values(ascending=False).head()
```

Out[12]:
```
Region
US         1039909
Spain       236899
Italy       203591
France      166536
UK          166432
Name: Confirmed, dtype: int64
```

In [14]:
```
#deaths confirmed
data.groupby('Region').Deaths.sum().sort_values(ascending=False).head()
```

Out[14]:
```
Region
US         60967
Italy      27682
UK         26165
Spain      24275
France     24121
Name: Deaths, dtype: int64
```

In [15]:
```
#no. of confirmed deaths & recovered cases reported in 2020 of april
data[data.Region=='India']
```

Out[15]:

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

In [17]:
```
data[data.Region=='Yemen']
```

Out[17]:

Date	State	Region	Confirmed	Deaths	Recovered

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

Out[19]:

	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
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
287	4/29/2020	Rhode Island	US	8247	251	0
288	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

In [24]:
```
#sorting wrt to no. of cases
data.sort_values(by=['Confirmed'], ascending=True).head()
```

Out[24]:

	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

In [25]:
```
data.sort_values(by=['Confirmed'], ascending=False).head()
```

Out[25]:

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

In [ ]:
```

```