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

```
In [2]:  df=pd.read_csv(r"E:\new download\4. covid_19_data.csv")
```

```
In [15]:  df
```

Out[15]:

	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

Count not null values

```
In [5]:  df.count()
```

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

```
In [7]:  df.notnull().sum()
```

```
Out[7]:  Date         321
         State        140
         Region       321
         Confirmed    321
         Deaths       321
         Recovered    321
         dtype: int64
```

```
In [9]:  df.isnull().sum()
```
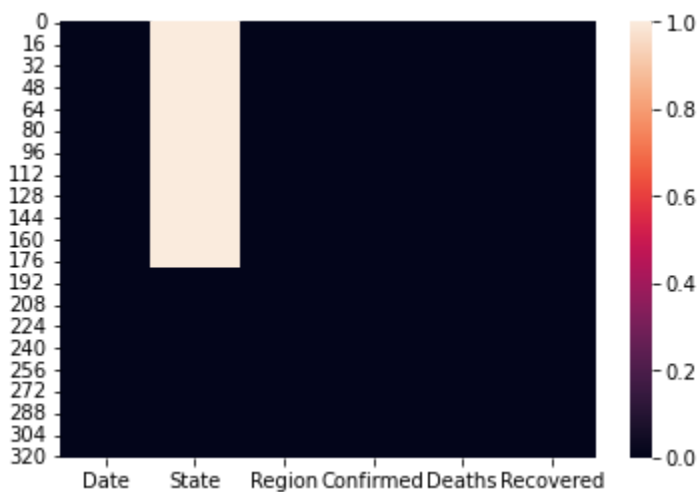
```
Out[9]:  Date           0
         State        181
         Region         0
         Confirmed      0
         Deaths         0
         Recovered      0
         dtype: int64
```

Loading [MathJax]/extensions/Safe.js

Import seaborn and show null by heat map

```
In [11]:  import seaborn as sns
          import matplotlib.pyplot as plt
```

```
In [14]:  sns.heatmap(df.isnull())
          plt.show()
```


Q.1) Show the number of confirmed, deaths and recoverd cases in each region.

```
In [21]:  df.head()
```

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
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 [23]:  df.groupby('Region').sum()
```

Loading [MathJax]/extensions/Safe.js

`Out[23]:`

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 [27]:`
```python
df.groupby('Region')['Confirmed', 'Deaths'].sum()
```

C:\Users\hp\AppData\Local\Temp\ipykernel_9128\1006431545.py:1: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.
  df.groupby('Region')['Confirmed', 'Deaths'].sum()

`Out[27]:`

Region	Confirmed	Deaths
Afghanistan	1939	60
Albania	766	30
Algeria	3848	444
Andorra	743	42
Angola	27	2
...	...	...
West Bank and Gaza	344	2
Western Sahara	6	0
Yemen	6	0
Zambia	97	3
Zimbabwe	32	4

187 rows × 2 columns

Q.2) Remove all the records where Confirmed case is Less than 10.

`In [29]:`
```python
df=df[~(df['Confirmed']<10)]
```

Loading [MathJax]/extensions/Safe.js

```
In [30]:   df
```

Out[30]:

	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

Q.3) In which region, maximum number of confirmed cases were recorded?

```
In [36]:   df.groupby('Region').Confirmed.sum().sort_values(ascending=False).head(7)
```

Out[36]:
```
Region
US         1039909
Spain       236899
Italy       203591
France      166536
UK          166432
Germany     161539
Turkey      117589
Name: Confirmed, dtype: int64
```

Q.4) In which region, minimum number of death cases were recorded?

```
In [41]:   df.groupby('Region').Deaths.sum().sort_values(ascending= True).head(25)
```

```
Out[41]:  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
          Name: Deaths, dtype: int64
```

Q.5) How many confirmed,death and recovered cases were reported from India till 29 april 2020?

In [52]: ```python
df[(df['Region']=='India')].groupby('Region').sum()
```

Out[52]:

|        | Confirmed | Deaths | Recovered |
|--------|-----------|--------|-----------|
| **Region** |       |        |           |
| **India** | 33062  | 1079   | 8437      |

## Q.6.a) Sort the entire data wrt No. of Confirmed cases in descending order.

In [57]: ```python
df.sort_values(by=['Confirmed'], ascending=False).head(20)
```

Loading [MathJax]/extensions/Safe.js

Out[57]:

	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
61	4/29/2020	NaN	Germany	161539	6467	120400
167	4/29/2020	NaN	Turkey	117589	3081	44040
262	4/29/2020	New Jersey	US	116365	6771	0
134	4/29/2020	NaN	Russia	99399	972	10286
76	4/29/2020	NaN	Iran	93657	5957	73791
22	4/29/2020	NaN	Brazil	79685	5513	34132
229	4/29/2020	Hubei	Mainland China	68128	4512	63616
249	4/29/2020	Massachusetts	US	60265	3405	0
232	4/29/2020	Illinois	US	50358	2215	0
195	4/29/2020	California	US	48747	1946	0
15	4/29/2020	NaN	Belgium	47859	7501	11283
278	4/29/2020	Pennsylvania	US	46327	2373	0
251	4/29/2020	Michigan	US	40399	3670	0
116	4/29/2020	NaN	Netherlands	38802	4711	0
128	4/29/2020	NaN	Peru	33931	943	10037

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

In [60]:
```python
df.sort_values(by=['Recovered'], ascending=False).head(20)
```

```
Out[60]:
```

	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

```
In [ ]:
```