```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

covid= pd.read_csv(r"C:\Users\Ashish\OneDrive\Pyhton Project file ALL\
covid_19_data.csv")

covid.head()
```

| | 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 |

```
# It's return the number of no null values for each coloumn, NULL
value menas missing Values

covid.count()
```

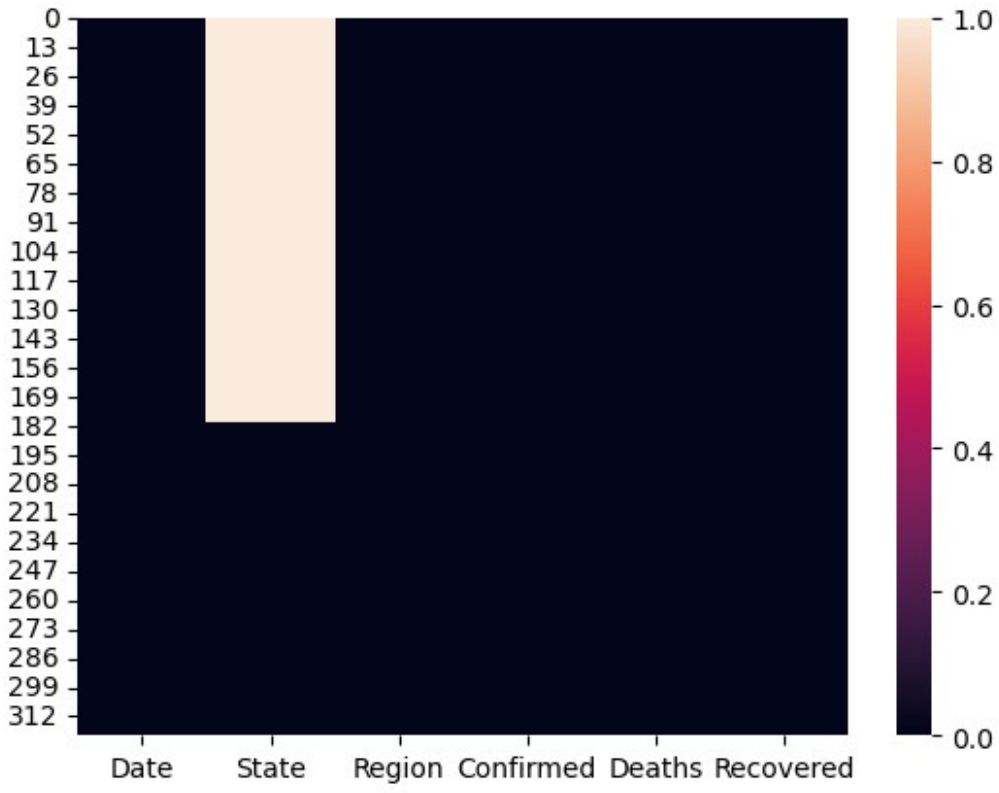
```
Date          321
State         140
Region        321
Confirmed     321
Deaths        321
Recovered     321
dtype: int64
```

```
covid.isnull().sum()
```

```
Date            0
State         181
Region          0
Confirmed       0
Deaths          0
Recovered       0
dtype: int64
```

```
# The white line show that their are null value present in that
particular coloumn

sns.heatmap(covid.isnull())
plt.show()
```




## *Show the number of "confirmed", "Death" and "Recovered" cases in each region?*

```
covid.head()
```

| | 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 |

```
covid.groupby("Region").sum().head(7) # To find the data for whole dataframe!
```

| Region | Date | State | Confirmed | Deaths | Recovered |
|---|---|---|---|---|---|
| Afghanistan | 4/29/2020 | 0 | 1939 | 60 | 252 |
| Albania | 4/29/2020 | 0 | 766 | 30 | 455 |
| Algeria | 4/29/2020 | 0 | 3848 | 444 | 1702 |
| Andorra | 4/29/2020 | 0 | 743 | 42 | 423 |
| Angola | 4/29/2020 | 0 | 27 | 2 | 7 |
| Antigua and Barbuda | 4/29/2020 | 0 | 24 | 3 | 11 |
| Argentina | 4/29/2020 | 0 | 4285 | 214 | 1192 |

```
covid.groupby("Region")["Confirmed"].sum().sort_values(ascending =
False).head(7) # To find the data for particular coloumn!
```

```
Region
US         1039909
Spain       236899
Italy       203591
France      166543
UK          166441
Germany     161539
Turkey      117589
Name: Confirmed, dtype: int64
```

```
# To find the data for multiple coloumn!
# You need double square bracket to make it a list or it will give
tupel subset error
# You also need to specify the "By" parameter to tell which coloumn
you want sorting

covid.groupby("Region")
[["Confirmed","Recovered"]].sum().sort_values(by = "Confirmed" ,
ascending = False).head(7)
```

| Region | Confirmed | Recovered |
|---|---|---|
| US | 1039909 | 120720 |
| Spain | 236899 | 132929 |
| Italy | 203591 | 71252 |
| France | 166543 | 49118 |
| UK | 166441 | 857 |
| Germany | 161539 | 120400 |
| Turkey | 117589 | 44040 |

# *Remove all record where confirmed cases is less then 10?*

```
covid.head()
```

| | 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 |

```
# To find all the value in confirmed coloumn which is less then 10!

covid[covid.Confirmed < 10].head(2)
```

```
         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
```

```
# To remove the value which is less then 10, we use "~" sign

covid = covid[~(covid.Confirmed < 10)]

covid.head(50).sort_values(by = "Confirmed" , ascending =
True).head(2)
```

```
         Date State    Region  Confirmed  Deaths  Recovered
27  4/29/2020   NaN   Burundi         11       1          4
45  4/29/2020   NaN  Dominica         16       0         13
```

# *In which region, maximum number of confirmed cases were recored?*

```
covid.head(2)
```

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

```
# To find the max confirmed cases with index for covied with region!
covid.groupby("Region")["Confirmed"].sum().idxmax()
```

```
'US'
```

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

```
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
```

# *In which region, minimum number of death cases were recored?*

```
covid.head(2)
```

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

```
covid.groupby("Region")["Deaths"].sum().sort_values(ascending =
True).head(20)
```

```
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
Name: Deaths, dtype: int64
```

## *How many confirmed, death & recovered cases were reported from india till 29th april 2022?*

These record are for date - 29th April 2022 that why we didnt use date filtring!

covid.head(2)

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

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

```
covid[covid.Region == "Yemen"]
```

```
Empty DataFrame
Columns: [Date, State, Region, Confirmed, Deaths, Recovered]
Index: []
```

```
covid[covid.Region == "US"].head(2)
```

```
          Date    State Region  Confirmed  Deaths  Recovered
181  4/29/2020  Alabama     US       6912     256          0
182  4/29/2020   Alaska     US        355       9          0
```

## *Sort the entire data with number of confirmed cases in ascending order?*

```
covid.head(2)
```

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

```
covid.sort_values(by = ["Confirmed"], ascending = True).head(2)
```

```
          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
```

## *Sort the entire data with number of Recovred cases in ascending order?*

```
covid.sort_values(by = ["Recovered"], ascending = True).head(2)
```

| | Date | State | Region | Confirmed | Deaths | Recovered |
|---|---|---|---|---|---|---|
| 199 | 4/29/2020 | Colorado | US | 14758 | 766 | 0 |
| 276 | 4/29/2020 | Ontario | Canada | 16978 | 1153 | 0 |