

```
import pandas as pd
```

```
data =pd.read_csv(r"C:\Users\ankus\Downloads\Covid 19.csv")
```

```
data
```

	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 x 6 columns]
```

```
data.count()          #this shows values in column  
#null values means missing values
```

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

```
data.isnull().sum()   #this show number of null values present
```

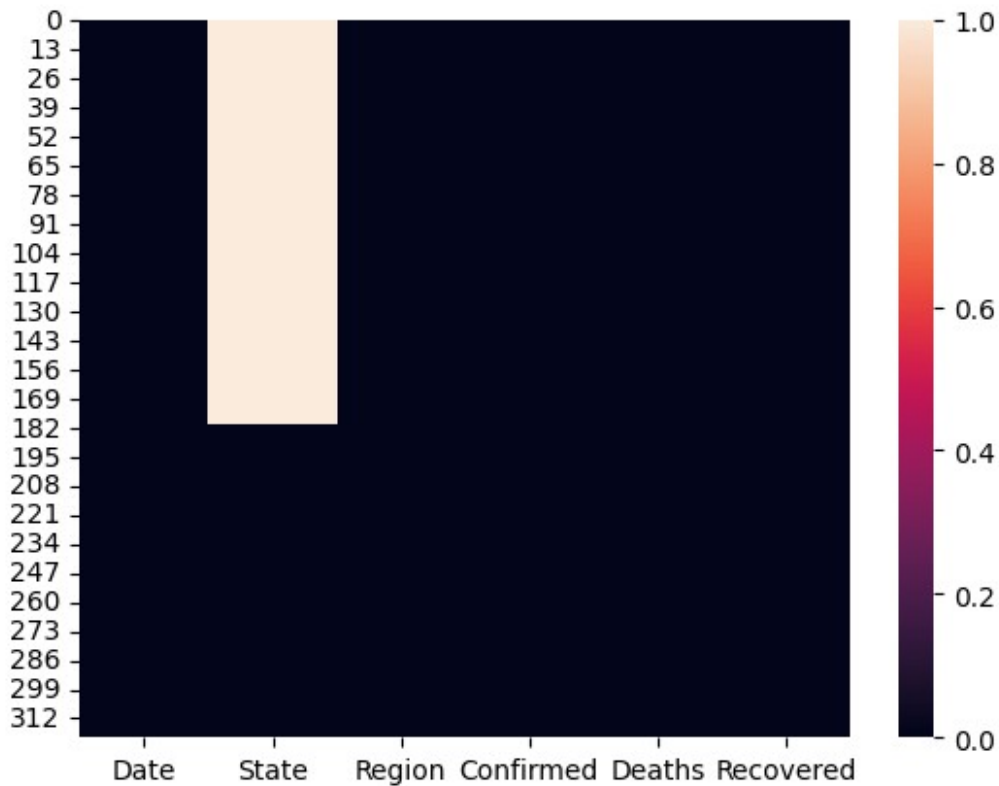
```
Date      0  
State     181  
Region    0  
Confirmed 0  
Deaths    0  
Recovered 0  
dtype: int64
```

```
#import seaborn as sns  
#import matplotlib.pyplot as plt  
#sns.heatmap(df.isnull())  
#plt.show()
```

```
import seaborn as sns          #Seaborn & Matplotlib are for  
Visualization
```

```
import matplotlib.pyplot as plt
```

```
sns.heatmap(data.isnull()) #show heatmap for Null Values in state
column
plt.show()
```



```
#Q1- Show the number of confirmed, Death and Recovered Cases in each
Region.
#df.groupby('Region').Sum().head(50)
#df.groupby('Region')
['Confirmed'].sum().sort_values(ascending=False).head(20)
#df.groupby('Region')['confirmed',Recovered].sum()
```

```
data.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

```
data.groupby('Region').sum().head(20)
```

	Date
\	
Region	
Afghanistan	4/29/2020

Albania	4/29/2020
Algeria	4/29/2020
Andorra	4/29/2020
Angola	4/29/2020
Antigua and Barbuda	4/29/2020
Argentina	4/29/2020
Armenia	4/29/2020
Australia	4/29/20204/29/20204/29/20204/29/20204/29/20204...
Austria	4/29/2020
Azerbaijan	4/29/2020
Bahamas	4/29/2020
Bahrain	4/29/2020
Bangladesh	4/29/2020
Barbados	4/29/2020
Belarus	4/29/2020
Belgium	4/29/2020
Belize	4/29/2020
Benin	4/29/2020
Bhutan	4/29/2020
State	
\	
Region	
Afghanistan	0
Albania	0
Algeria	0
Andorra	0
Angola	0

Antigua and Barbuda	0
Argentina	0
Armenia	0
Australia	Australian Capital TerritoryNew South WalesNor...
Austria	0
Azerbaijan	0
Bahamas	0
Bahrain	0
Bangladesh	0
Barbados	0
Belarus	0
Belgium	0
Belize	0
Benin	0
Bhutan	0
	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

```
#maximum number of confirmed cases is in US Region. (us confirmed cases = 1039909(MAX))
data.groupby('Region')['Confirmed'].sum().sort_values(ascending = False).head(20)
```

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

```
#least confirmed cases are in Yemen Region. (Yemen confirmed cases = 6(MINIMUM))
data.groupby('Region')['Confirmed'].sum().sort_values(ascending = True).head(20)
```

Region	
Yemen	6
Western Sahara	6
Bhutan	7
Papua New Guinea	8
Sao Tome and Principe	8
Mauritania	8
MS Zaandam	9
Gambia	10
Holy See	10
Suriname	10
Burundi	11
Seychelles	11

Nicaragua	13
Saint Kitts and Nevis	15
Dominica	16
Namibia	16
Saint Vincent and the Grenadines	16
Saint Lucia	17
Belize	18
Fiji	18

Name: Confirmed, dtype: int64

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

```
-----
-----
ValueError                                Traceback (most recent call
last)
```

```
Cell In[16], line 1
```

```
----> 1 data.groupby('Region')['Confirmed', 'Recovered'].sum()
```

```
File C:\Python312\Lib\site-packages\pandas\core\groupby\
generic.py:1960, in DataFrameGroupBy.__getitem__(self, key)
    1956 # per GH 23566
    1957 if isinstance(key, tuple) and len(key) > 1:
    1958     # if len == 1, then it becomes a SeriesGroupBy and this is
actually
    1959     # valid syntax, so don't raise
-> 1960     raise ValueError(
    1961         "Cannot subset columns with a tuple with more than one
element. "
    1962         "Use a list instead."
    1963     )
    1964 return super().__getitem__(key)
```

```
ValueError: Cannot subset columns with a tuple with more than one
element. Use a list instead.
```

```
#Q2 -Remove all the records where Confirmed Cases is less Than 10.
data.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

```
data[data.Confirmed<10] #this show confirmed cases less than 10
```

	Date	State
Region \		
18	4/29/2020	NaN
Bhutan		
98	4/29/2020	NaN
Zaandam		MS

105	4/29/2020	NaN	
Mauritania			
126	4/29/2020	NaN	Papua New Guinea
140	4/29/2020	NaN	Sao Tome and Principe
177	4/29/2020	NaN	Western Sahara
178	4/29/2020	NaN	
Yemen			
184	4/29/2020	Anguilla	
UK			
192	4/29/2020	Bonaire, Sint Eustatius and Saba	
Netherlands			
194	4/29/2020	British Virgin Islands	
UK			
203	4/29/2020	Diamond Princess cruise ship	
Canada			
272	4/29/2020	Northwest Territories	
Canada			
284	4/29/2020	Recovered	
Canada			
285	4/29/2020	Recovered	
US			
288	4/29/2020	Saint Barthelemy	
France			
289	4/29/2020	Saint Pierre and Miquelon	
France			
305	4/29/2020	Tibet	Mainland
China			

	Confirmed	Deaths	Recovered
18	7	0	5
98	9	2	0
105	8	1	6
126	8	0	0
140	8	0	4
177	6	0	5
178	6	0	1
184	3	0	3
192	5	0	0
194	6	1	3
203	0	1	0
272	5	0	0
284	0	0	20327
285	0	0	120720
288	6	0	6
289	1	0	0
305	1	0	1

```
#now removing the less than 10 confirmed cases from dataframe.
data= data[~(data.Confirmed<10)] #To remove the records satisfying a
particular condition
```

```
data #confirmed cases less than 10 is removed here
```

	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 x 6 columns]
```

```
#Q3 - In which Region , MINIMUM number of Deaths cases were recorded ?
```

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

```
#Below country has Minimum Number of death cases
```

```
#Gambia 1
#Equatorial Guinea 1
#Eswatini 1
#Cabo Verde 1
#Maldives 1
#Guinea-Bissau 1
#Liechtenstein 1
#Brunei 1
#Burundi 1
#Botswana 1
#Suriname
```

```
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
Eswatini	1
Cabo Verde	1
Maldives	1
Guinea-Bissau	1
Liechtenstein	1
Brunei	1
Burundi	1
Botswana	1
Suriname	1
Benin	1
Djibouti	2
Angola	2
Libya	2
Chad	2
West Bank and Gaza	2
Belize	2
Zambia	3
Malawi	3
Nicaragua	3
Syria	3
Ethiopia	3
Antigua and Barbuda	3
Gabon	3
Hong Kong	4
Zimbabwe	4

Name: Deaths, dtype: int64

#Q.4.-How many Confirmed ,Death & Recovered cases were reported from India till 29 April 2020?

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

#India stats are Below

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

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

#Q.5.-sort the entire data wrt No. of recovered cases in descending order.

```
data.sort_values(by=['Recovered'],ascending= False).head(20)
```

#Max Recovered from Spain, the United States & Germany country.

#India Recoverd 8437.

#Region Confirmed Deaths Recovered

Spain 236899 24275 132929
US 0 0 120720
Germany 161539 6467 120400

	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 8233	4/29/2020	NaN	Israel	15834	215
------------	-----------	-----	--------	-------	-----