

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

```
In [2]: dataset1=pd.read_csv('worldometer_data.csv')
```

```
In [3]: dataset1.head(5)
```

Out[3]:

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	Total
0	USA	North America	3.311981e+08	5032179	NaN	162804.0	NaN	
1	Brazil	South America	2.127107e+08	2917562	NaN	98644.0	NaN	
2	India	Asia	1.381345e+09	2025409	NaN	41638.0	NaN	
3	Russia	Europe	1.459409e+08	871894	NaN	14606.0	NaN	
4	South Africa	Africa	5.938157e+07	538184	NaN	9604.0	NaN	

## Step 1

## Data Cleaning/Missing Value 'NaN'/Unknown replaced with 'Zero'

```
In [4]: dataset1.isnull().sum() # Total count of Unknow values in each column
```

```
Out[4]: Country/Region      0
Continent                  1
Population                  1
TotalCases                  0
NewCases                   205
TotalDeaths                 21
NewDeaths                   206
TotalRecovered              4
NewRecovered                206
ActiveCases                  4
Serious,Critical            87
Tot Cases/1M pop            1
Deaths/1M pop               22
TotalTests                  18
Tests/1M pop                18
WHO Region                  25
dtype: int64
```

```
In [5]: dataset2=dataset1.fillna(0)
dataset2
```

Out[5]:

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	To
0	USA	North America	3.311981e+08	5032179	0.0	162804.0	0.0	
1	Brazil	South America	2.127107e+08	2917562	0.0	98644.0	0.0	
2	India	Asia	1.381345e+09	2025409	0.0	41638.0	0.0	
3	Russia	Europe	1.459409e+08	871894	0.0	14606.0	0.0	
4	South Africa	Africa	5.938157e+07	538184	0.0	9604.0	0.0	
...	...	...	...	...	...	...	...	...
204	Montserrat	North America	4.992000e+03	13	0.0	1.0	0.0	
205	Caribbean Netherlands	North America	2.624700e+04	13	0.0	0.0	0.0	
206	Falkland Islands	South America	3.489000e+03	13	0.0	0.0	0.0	
207	Vatican City	Europe	8.010000e+02	12	0.0	0.0	0.0	
208	Western Sahara	Africa	5.986820e+05	10	0.0	1.0	0.0	

209 rows × 16 columns

```
In [6]: #dataset2.to_excel("C:\\Users\\pragati\\Desktop\\DATAMiningFinalProject\\Final Pr
```

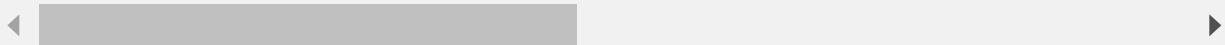
## DATA Cleaning is Done : The new excel we have is "df\_Cleandata"

In [7]: dataset2

Out[7]:

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	To
0	USA	North America	3.311981e+08	5032179	0.0	162804.0	0.0	
1	Brazil	South America	2.127107e+08	2917562	0.0	98644.0	0.0	
2	India	Asia	1.381345e+09	2025409	0.0	41638.0	0.0	
3	Russia	Europe	1.459409e+08	871894	0.0	14606.0	0.0	
4	South Africa	Africa	5.938157e+07	538184	0.0	9604.0	0.0	
...	...	...	...	...	...	...	...	...
204	Montserrat	North America	4.992000e+03	13	0.0	1.0	0.0	
205	Caribbean Netherlands	North America	2.624700e+04	13	0.0	0.0	0.0	
206	Falkland Islands	South America	3.489000e+03	13	0.0	0.0	0.0	
207	Vatican City	Europe	8.010000e+02	12	0.0	0.0	0.0	
208	Western Sahara	Africa	5.986820e+05	10	0.0	1.0	0.0	

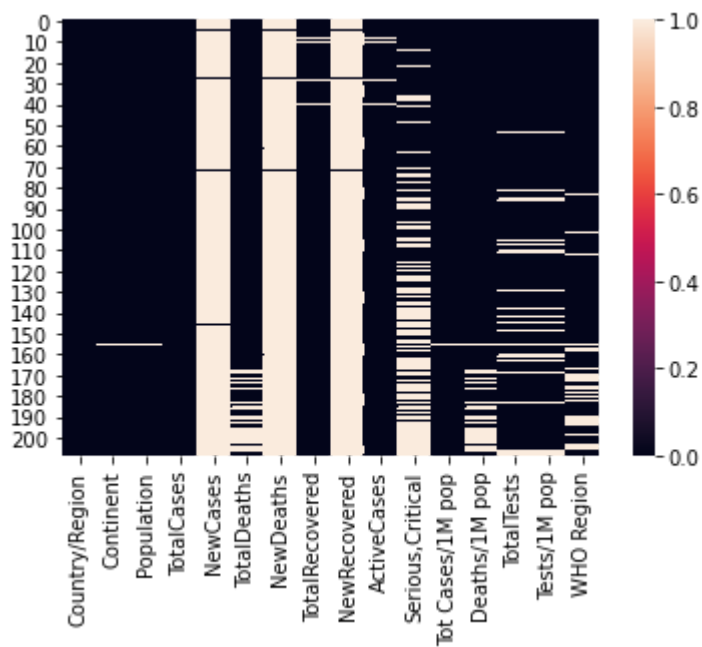
209 rows × 16 columns



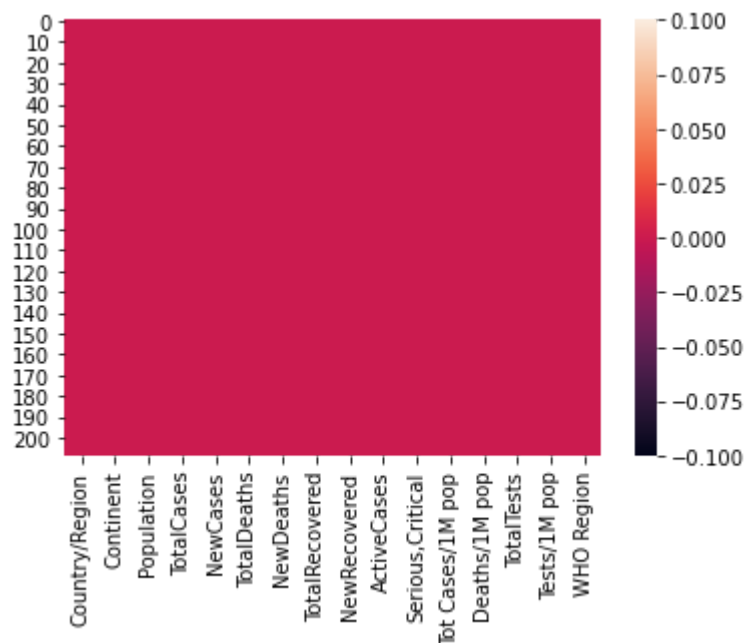
In [8]: `import seaborn as sns`  
`import matplotlib.pyplot as plt`

In [ ]:

```
In [9]: sns.heatmap(dataset1.isnull())  
plt.show()
```



```
In [10]: sns.heatmap(dataset2.isnull())  
plt.show()
```



## Step 2

**Q1. Show the number of Confirmed, Deaths and Recovered cases in each Region.**

In [11]: `dataset2.groupby('Country/Region').sum()`

Out[11]:

	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewI
Country/Region							
<b>Afghanistan</b>	39009447.0	36896	0.0	1298.0	0.0	25840.0	
<b>Albania</b>	2877470.0	6016	0.0	188.0	0.0	3155.0	
<b>Algeria</b>	43926079.0	33626	0.0	1273.0	0.0	23238.0	
<b>Andorra</b>	77278.0	944	0.0	52.0	0.0	828.0	
<b>Angola</b>	32956300.0	1483	0.0	64.0	0.0	520.0	
...	...	...	...	...	...	...	...
<b>Vietnam</b>	97425470.0	747	0.0	10.0	0.0	392.0	
<b>Western Sahara</b>	598682.0	10	0.0	1.0	0.0	8.0	
<b>Yemen</b>	29886897.0	1768	0.0	508.0	0.0	898.0	
<b>Zambia</b>	18430129.0	7164	0.0	199.0	0.0	5786.0	
<b>Zimbabwe</b>	14883803.0	4339	0.0	84.0	0.0	1264.0	

209 rows × 13 columns



In [12]: `dataset2.groupby('Country/Region').sum().head()`

Out[12]:

	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered	NewI
Country/Region							
<b>Afghanistan</b>	39009447.0	36896	0.0	1298.0	0.0	25840.0	
<b>Albania</b>	2877470.0	6016	0.0	188.0	0.0	3155.0	
<b>Algeria</b>	43926079.0	33626	0.0	1273.0	0.0	23238.0	
<b>Andorra</b>	77278.0	944	0.0	52.0	0.0	828.0	
<b>Angola</b>	32956300.0	1483	0.0	64.0	0.0	520.0	



```
In [13]: dataset2.groupby('Country/Region')['TotalCases'].sum().sort_values(ascending=False)
```

```
Out[13]: Country/Region
USA          5032179
Brazil       2917562
India        2025409
Russia       871894
South Africa 538184
Mexico       462690
Peru         455409
Chile        366671
Colombia     357710
Spain        354530
Name: TotalCases, dtype: int64
```

```
In [14]: dataset2.groupby('Country/Region')['TotalCases', 'TotalRecovered'].sum()
```

<ipython-input-14-2c2293e3bbcf>:1: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead.

```
dataset2.groupby('Country/Region')['TotalCases', 'TotalRecovered'].sum()
```

```
Out[14]:
```

	TotalCases	TotalRecovered
Country/Region		
<b>Afghanistan</b>	36896	25840.0
<b>Albania</b>	6016	3155.0
<b>Algeria</b>	33626	23238.0
<b>Andorra</b>	944	828.0
<b>Angola</b>	1483	520.0
...	...	...
<b>Vietnam</b>	747	392.0
<b>Western Sahara</b>	10	8.0
<b>Yemen</b>	1768	898.0
<b>Zambia</b>	7164	5786.0
<b>Zimbabwe</b>	4339	1264.0

209 rows × 2 columns

## Q2) Remove all the records where Confirmed cases is Less Than 10,000.

In [15]: dataset2

Out[15]:

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	Total
0	USA	North America	3.311981e+08	5032179	0.0	162804.0	0.0	
1	Brazil	South America	2.127107e+08	2917562	0.0	98644.0	0.0	
2	India	Asia	1.381345e+09	2025409	0.0	41638.0	0.0	
3	Russia	Europe	1.459409e+08	871894	0.0	14606.0	0.0	
4	South Africa	Africa	5.938157e+07	538184	0.0	9604.0	0.0	
...	...	...	...	...	...	...	...	...
204	Montserrat	North America	4.992000e+03	13	0.0	1.0	0.0	

In [16]: dataset2=dataset2[~(dataset2.TotalCases<1000)]

In [17]: dataset2

Out[17]:

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	To
0	USA	North America	3.311981e+08	5032179	0.0	162804.0	0.0	
1	Brazil	South America	2.127107e+08	2917562	0.0	98644.0	0.0	
2	India	Asia	1.381345e+09	2025409	0.0	41638.0	0.0	
3	Russia	Europe	1.459409e+08	871894	0.0	14606.0	0.0	
4	South Africa	Africa	5.938157e+07	538184	0.0	9604.0	0.0	
...	...	...	...	...	...	...	...	...
140	Cyprus	Asia	1.208238e+06	1208	0.0	19.0	0.0	
141	Georgia	Asia	3.988368e+06	1206	0.0	17.0	0.0	
142	Burkina Faso	Africa	2.095485e+07	1158	0.0	54.0	0.0	
143	Niger	Africa	2.428143e+07	1153	0.0	69.0	0.0	
144	Togo	Africa	8.296582e+06	1012	0.0	22.0	0.0	

145 rows × 16 columns



### Q3) In which , Maximum number of Confirmed cases were recorded?

```
In [18]: dataset2.groupby('Country/Region').TotalCases.sum().sort_values(ascending=False).
```

```
Out[18]: Country/Region
USA      5032179
Brazil   2917562
India    2025409
Russia   871894
South Africa  538184
Mexico   462690
Peru     455409
Chile    366671
Colombia 357710
Spain    354530
Iran     320117
UK       308134
Saudi Arabia 284226
Pakistan 281863
Bangladesh 249651
Italy    249204
Turkey   237265
Argentina 228195
Germany  215210
France   195633
Name: TotalCases, dtype: int64
```

### Q4. In which Country/Region, minimum number of Deaths cases were recorded?

```
In [19]: dataset2.groupby('Country/Region').TotalDeaths.sum().sort_values(ascending=True).
```

```
Out[19]: Country/Region
Uganda      5.0
Rwanda      5.0
Iceland     10.0
Jordan      11.0
Sri Lanka   11.0
Namibia     15.0
Mozambique  15.0
Georgia     17.0
Maldives    19.0
Cyprus       19.0
New Zealand 22.0
Togo        22.0
Singapore   27.0
Guinea-Bissau 27.0
Cabo Verde  27.0
Slovakia    29.0
Suriname    29.0
Latvia      32.0
Uruguay     37.0
Benin       38.0
Name: TotalDeaths, dtype: float64
```

**Q5. Sort the entire data wrt No. of Recovered cases in descending order.**

```
In [20]: dataset2.sort_values(by=['TotalRecovered'],ascending =False).head(10)
```

```
Out[20]:
```

	Country/Region	Continent	Population	TotalCases	NewCases	TotalDeaths	NewDeaths	TotalRecovered
0	USA	North America	3.311981e+08	5032179	0.0	162804.0	0.0	331198100
1	Brazil	South America	2.127107e+08	2917562	0.0	98644.0	0.0	212710700
2	India	Asia	1.381345e+09	2025409	0.0	41638.0	0.0	1381345000
3	Russia	Europe	1.459409e+08	871894	0.0	14606.0	0.0	1459409000
4	South Africa	Africa	5.938157e+07	538184	0.0	9604.0	0.0	593815700
7	Chile	South America	1.913251e+07	366671	0.0	9889.0	0.0	191325100
6	Peru	South America	3.301632e+07	455409	0.0	20424.0	0.0	330163200
5	Mexico	North America	1.290662e+08	462690	6590.0	50517.0	819.0	129066200
10	Iran	Asia	8.409762e+07	320117	0.0	17976.0	0.0	840976200
13	Pakistan	Asia	2.212959e+08	281863	0.0	6035.0	0.0	221295900

## Step 3

```
In [21]: #Visualizing: Worldwide #Covid-19 cases
```

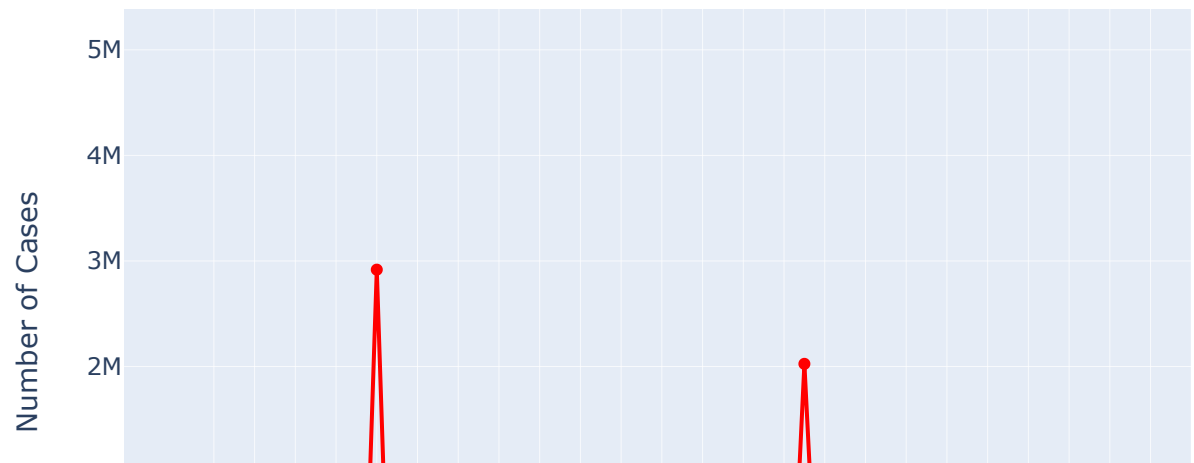
```
In [22]: import plotly
import plotly.graph_objs as go
import numpy as np
```

```
In [23]: TotalCases = dataset2.groupby('Country/Region').sum()['TotalCases'].reset_index()
TotalRecovered = dataset2.groupby('Country/Region').sum()['TotalRecovered'].reset_index()
Population = dataset2.groupby('Country/Region').sum()['Population'].reset_index()
```

```
In [24]: #conda install -c plotly plotly=5.4.0
```

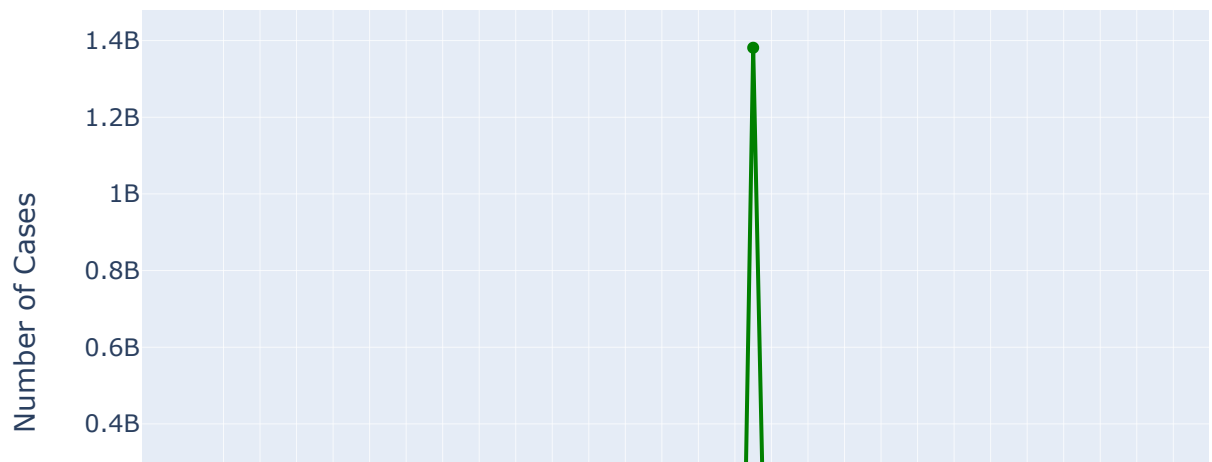
```
In [25]: fig=go.Figure()  
#Plotting datewise confirmed cases  
fig.add_trace(go.Scatter(x=TotalCases['Country/Region'], y=TotalCases['TotalCases'],  
fig.update_layout(title='Worldwide NCOVID-19 Cases', xaxis_tickfont_size=14,yaxis=  
fig.show()
```

## Worldwide NCOVID-19 Cases



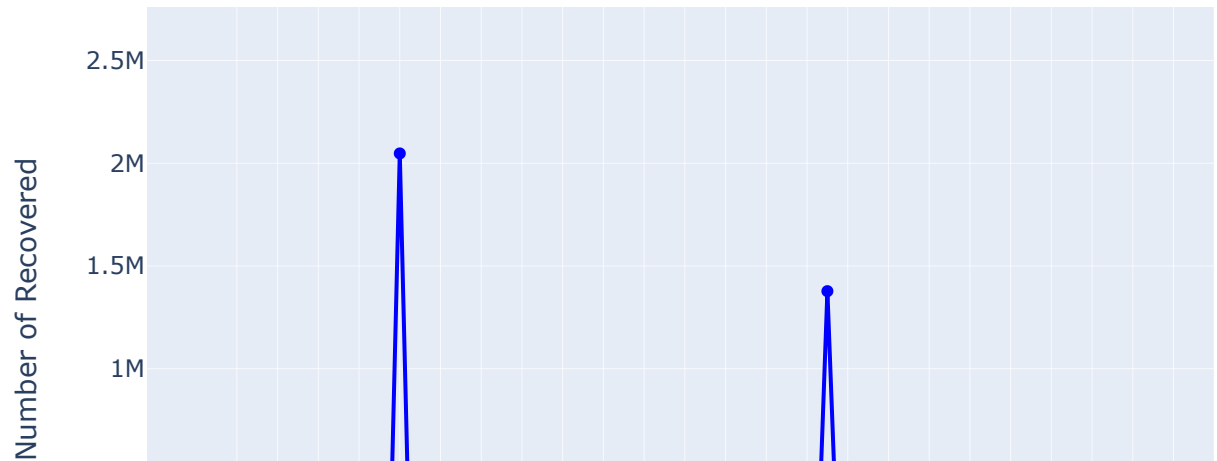
```
In [26]: fig=go.Figure()  
#Plotting datewise confirmed cases  
fig.add_trace(go.Scatter(x=TotalCases['Country/Region'], y=TotalCases['TotalCases'],  
fig.add_trace(go.Scatter(x=TotalRecovered['Country/Region'], y=TotalRecovered['TotalRecovered'],  
fig.add_trace(go.Scatter(x=Population['Country/Region'], y=Population['Population'],  
fig.update_layout(title='Worldwide NCOVID-19 Cases', xaxis_tickfont_size=14,yaxis_tickfont_size=14,  
fig.show()
```

## Worldwide NCOVID-19 Cases



```
In [27]: fig=go.Figure()
#Plotting datewise confirmed cases
fig.add_trace(go.Scatter(x=TotalRecovered['Country/Region'], y=TotalRecovered['Total Recovered'],
fig.update_layout(title='Worldwide NCOVID-19 Cases', xaxis_tickfont_size=14,yaxis=
fig.show()
```

## Worldwide NCOVID-19 Cases



In [ ]:

In [ ]:

## Forecasting Total Number of Cases Worldwide

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: