

Covid-19 Data Analysis Using Pandas

Project Done By - SYED ABDUR RAHMAN UDDIN, " Copyright - rahmansyed862@gmail.com "

Dataset used in this project is till April 2020. Downloaded from Kaggle. [Small Dataset]

```
In [3]: #importing libraries
```

```
In [13]: import pandas as pd
```

```
In [14]: import seaborn as sns
```

```
In [15]: import matplotlib.pyplot as plt
```

```
In [16]: #Assigning the Covid-19 dataset csv file to Data Object ( Any name for Data Object can be used )
```

```
In [17]: data = pd.read_csv(r"C:\Users\ROB\Downloads\4. covid_19_data.csv")
```

```
In [103]: data
```

Out[103]:

	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 [19]: data.count()
```

```
Out[19]: Date          321
State          140
Region         321
Confirmed      321
Deaths         321
Recovered      321
dtype: int64
```

```
In [10]: data.isnull()
```

```
Out[10]:
```

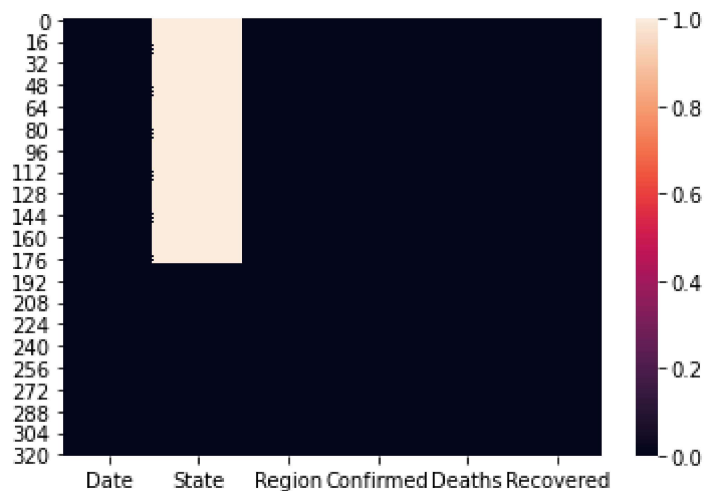
	Date	State	Region	Confirmed	Deaths	Recovered
0	False	True	False	False	False	False
1	False	True	False	False	False	False
2	False	True	False	False	False	False
3	False	True	False	False	False	False
4	False	True	False	False	False	False
...
316	False	False	False	False	False	False
317	False	False	False	False	False	False
318	False	False	False	False	False	False
319	False	False	False	False	False	False
320	False	False	False	False	False	False

321 rows × 6 columns

```
In [88]: sns.set(color_codes=True)
```

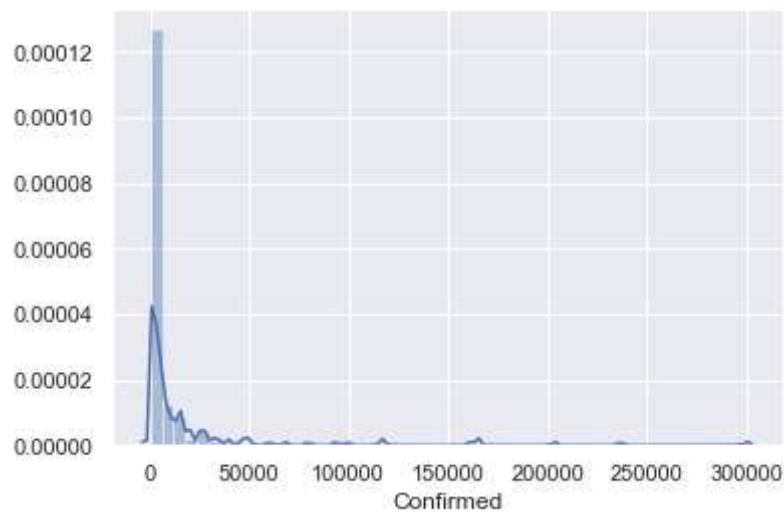
```
In [20]: sns.heatmap(data.isnull())
```

```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x220402203d0>
```



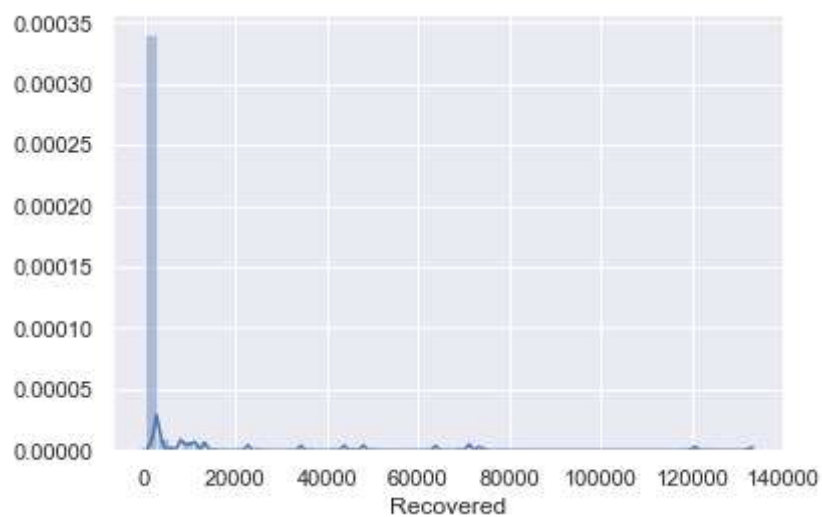
```
In [89]: sns.distplot(data['Confirmed'])
```

```
Out[89]: <matplotlib.axes._subplots.AxesSubplot at 0x22011bd4820>
```



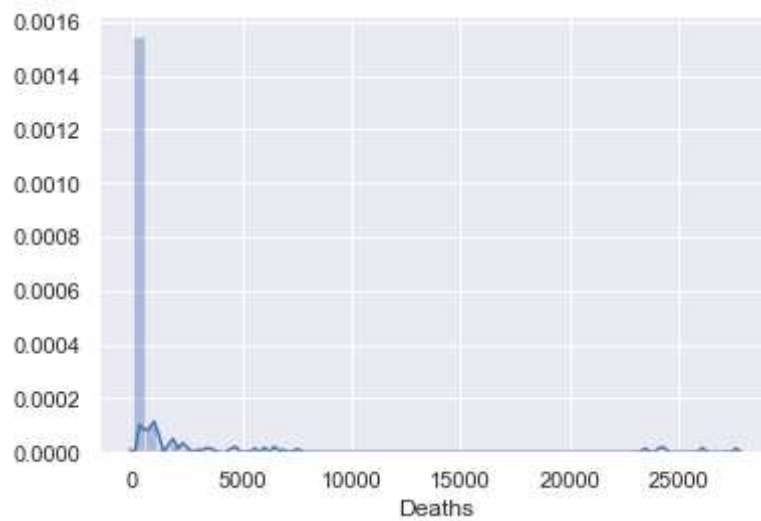
```
In [90]: sns.distplot(data['Recovered'])
```

```
Out[90]: <matplotlib.axes._subplots.AxesSubplot at 0x22011c8bd60>
```



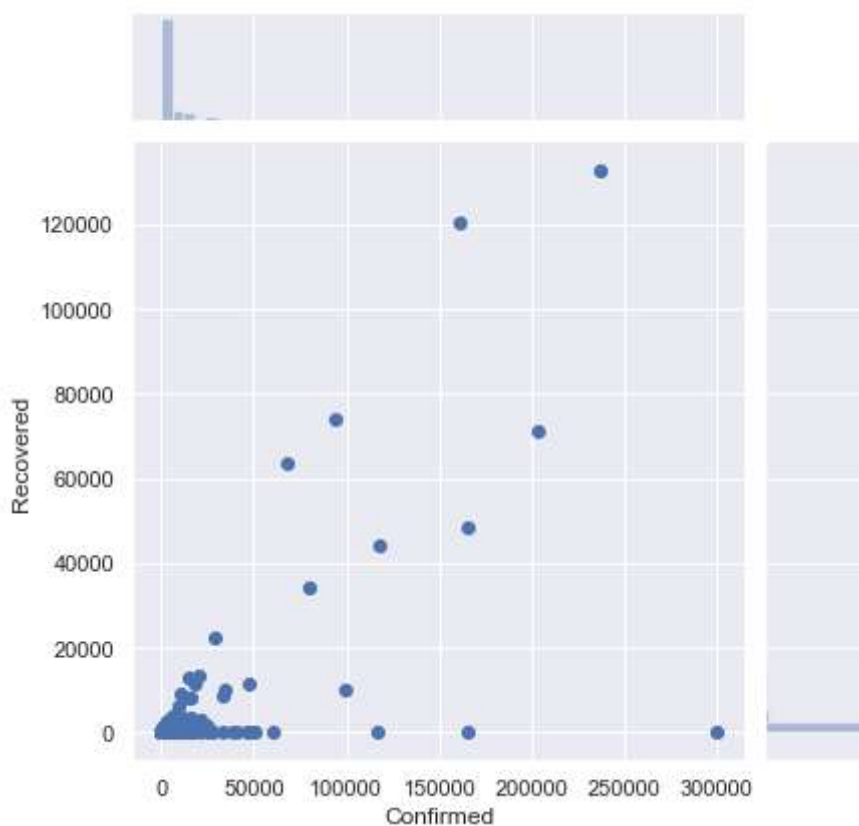
```
In [91]: sns.distplot(data['Deaths'])
```

```
Out[91]: <matplotlib.axes._subplots.AxesSubplot at 0x22011b6e0a0>
```



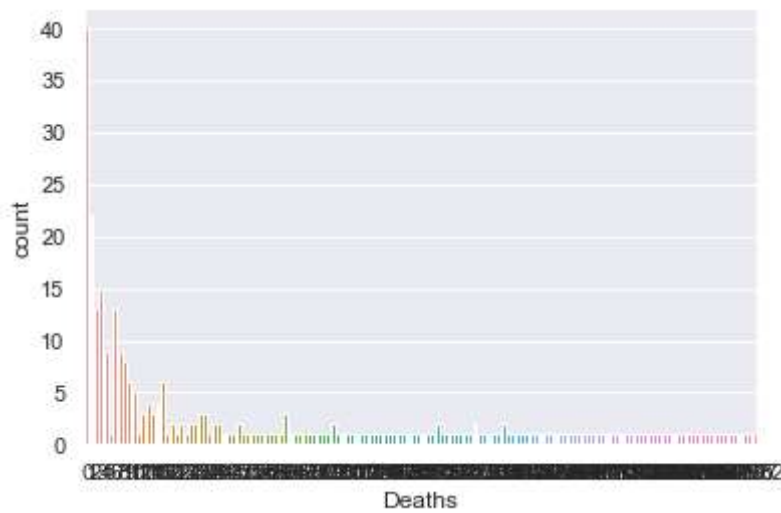
```
In [92]: sns.jointplot(data['Confirmed'], data['Recovered'])
```

```
Out[92]: <seaborn.axisgrid.JointGrid at 0x22011d93580>
```



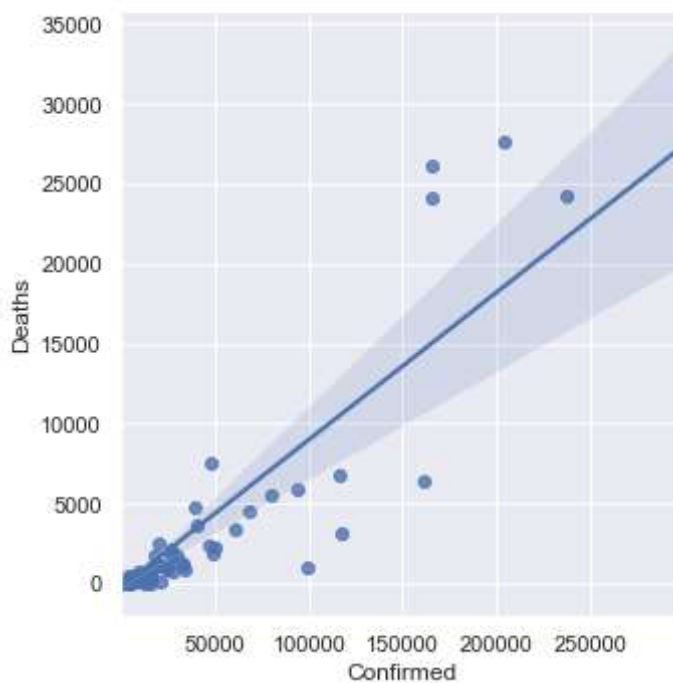
```
In [93]: sns.countplot(data['Deaths'])
```

```
Out[93]: <matplotlib.axes._subplots.AxesSubplot at 0x220120413a0>
```



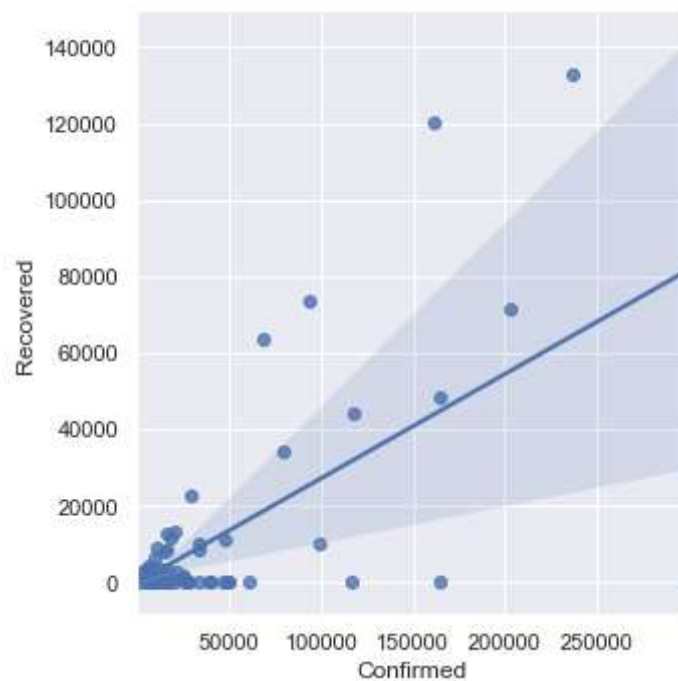
```
In [94]: sns.lmplot(x='Confirmed', y='Deaths', data=data)
```

```
Out[94]: <seaborn.axisgrid.FacetGrid at 0x22011e057f0>
```



```
In [95]: sns.lmplot(x='Confirmed', y='Recovered', data=data)
```

```
Out[95]: <seaborn.axisgrid.FacetGrid at 0x220122cbc70>
```



```
In [21]: data.head(3)
```

```
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

Q) Show Number of Confirmed, Deaths and Recovered Records.

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

```
Out[23]:
```

	Confirmed	Deaths	Recovered
Region			
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

Q) Show total Number of Confirmed Cases in all Regions.

```
In [25]: data.groupby('Region')['Confirmed'].sum()
```

```
Out[25]: Region
Afghanistan      1939
Albania           766
Algeria          3848
Andorra           743
Angola            27
...
West Bank and Gaza  344
Western Sahara      6
Yemen              6
Zambia            97
Zimbabwe          32
Name: Confirmed, Length: 187, dtype: int64
```

Q) Show total Number of Death Cases in all Regions.

```
In [27]: data.groupby('Region')['Deaths'].sum()
```

```
Out[27]: Region
Afghanistan      60
Albania          30
Algeria          444
Andorra          42
Angola           2
...
West Bank and Gaza  2
Western Sahara    0
Yemen            0
Zambia           3
Zimbabwe         4
Name: Deaths, Length: 187, dtype: int64
```

Q) Show total Number of Recovered Cases in all Regions.

```
In [28]: data.groupby('Region')['Recovered'].sum()
```

```
Out[28]: Region
Afghanistan      252
Albania          455
Algeria          1702
Andorra          423
Angola           7
...
West Bank and Gaza  71
Western Sahara     5
Yemen             1
Zambia            54
Zimbabwe          5
Name: Recovered, Length: 187, dtype: int64
```

Confirmed Cases in Descending Order.

```
In [31]: data.groupby('Region')['Confirmed'].sum().sort_values(ascending = False)
```

```
Out[31]: Region
US              1039909
Spain           236899
Italy           203591
France          166543
UK              166441
...
Sao Tome and Principe    8
Papua New Guinea        8
Bhutan                  7
Western Sahara          6
Yemen                   6
Name: Confirmed, Length: 187, dtype: int64
```


Death Cases in Descending Order.

```
In [32]: data.groupby('Region')['Deaths'].sum().sort_values(ascending = False)
```

```
Out[32]: Region
US          60967
Italy       27682
UK          26166
Spain       24275
France      24121
...
Fiji         0
Cambodia     0
Mozambique   0
Mongolia     0
Laos         0
Name: Deaths, Length: 187, dtype: int64
```

Recovered Cases in Descending Order.

```
In [33]: data.groupby('Region')['Recovered'].sum().sort_values(ascending = False)
```

```
Out[33]: Region
Spain      132929
US         120720
Germany    120400
Mainland China  77610
Iran       73791
...
Cabo Verde    2
Yemen         1
MS Zaandam    0
South Sudan   0
Papua New Guinea  0
Name: Recovered, Length: 187, dtype: int64
```

Q) Remove all the records where Confirmed Cases is Less than 10

In [35]: data.Confirmed<10

Out[35]:

0	False
1	False
2	False
3	False
4	False
...	
316	False
317	False
318	False
319	False
320	False

Name: Confirmed, Length: 321, dtype: bool

In [36]: data[data.Confirmed<10]

Out[36]:

	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
105	4/29/2020	NaN	Mauritania	8	1	6
126	4/29/2020	NaN	Papua New Guinea	8	0	0
140	4/29/2020	NaN	Sao Tome and Principe	8	0	4
177	4/29/2020	NaN	Western Sahara	6	0	5
178	4/29/2020	NaN	Yemen	6	0	1
184	4/29/2020	Anguilla	UK	3	0	3
192	4/29/2020	Bonaire, Sint Eustatius and Saba	Netherlands	5	0	0
194	4/29/2020	British Virgin Islands	UK	6	1	3
203	4/29/2020	Diamond Princess cruise ship	Canada	0	1	0
272	4/29/2020	Northwest Territories	Canada	5	0	0
284	4/29/2020	Recovered	Canada	0	0	20327
285	4/29/2020	Recovered	US	0	0	120720
288	4/29/2020	Saint Barthelemy	France	6	0	6
289	4/29/2020	Saint Pierre and Miquelon	France	1	0	0
305	4/29/2020	Tibet	Mainland China	1	0	1

In [37]: `data[~(data.Confirmed<10)]`

Out[37]:

	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 [38]: `data=data[~(data.Confirmed<10)]`

In [39]: `data`

Out[39]:

	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) In which Region, Maximum number of Death Cases were recorded ?

```
In [40]: data.groupby('Region').Deaths.sum().sort_values(ascending=False)
```

```
Out[40]: Region
US                60967
Italy             27682
UK                26165
Spain            24275
France           24121
...
Saint Kitts and Nevis    0
Central African Republic 0
Saint Lucia             0
Seychelles              0
Cambodia                0
Name: Deaths, Length: 180, dtype: int64
```

```
In [ ]:
```