covid-19-dataset

January 30, 2024

0.1 Working on Real Project with Python on 'Covid-19 Dataset'

0.1.1 COVID-19 SMALL DATSET

we have taken a small dataset of covid-19, for easy understanding purpose. The data used here is till 29-Apri-2020 and has records as on 29-April-2020. This data is available as aCSV file, download from kaggle. We will analyze this data using the pandas dataframe.

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

[2]:
```
data = pd.read_csv('D:/data analystics/Python for data analytics/Python␣
↪Projects/Covid-19 Dataset/file.csv')
```

[3]:
```
data.count()
```

[3]:
```
Date         321
State        140
Region       321
Confirmed    321
Deaths       321
Recovered    321
dtype: int64
```

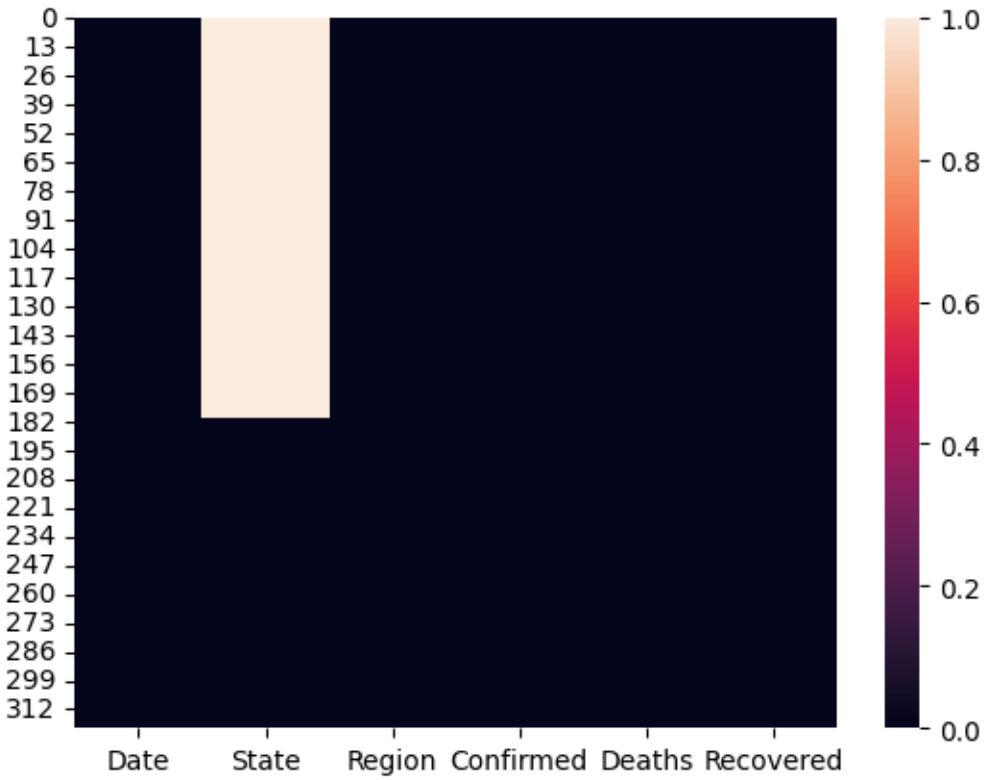
[5]:
```
data.isnull().sum()
```

[5]:
```
Date           0
State        181
Region         0
Confirmed      0
Deaths         0
Recovered      0
dtype: int64
```

[6]:
```
import seaborn as sns
```

[7]:
```
import matplotlib.pyplot as plt
```

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



0.1.2 Question.1. Show the number of confirmed, Deaths and Recovered cases in each Region.

[ ]:
```
# data.groupby('Region').sum().head(20)
```

[ ]:
```
#data.groupby('Region')['Confirmed'].sum().sort_values(ascending= False).
↪head(10)
```

[13]:
```
data.groupby('Region')[['Confirmed', 'Recovered']].sum()
```

[13]:

Region	Confirmed	Recovered
Afghanistan	1939	252
Albania	766	455
Algeria	3848	1702
Andorra	743	423
Angola	27	7
…	…	…
West Bank and Gaza	344	71
Western Sahara	6	5
Yemen	6	1

```
Zambia                    97          54
Zimbabwe                  32           5

[187 rows x 2 columns]
```

0.1.3 Question.2. Remove all the reocrds where confirmed Cases is less than 10.

[14]:
```
data = data[~(data.Confirmed < 10)]
```

[ ]:
```
#data.head(20)
```

0.1.4 Question.3. In which Region, maximum number of Confirmed cases were recorded?

[15]:
```
data.groupby('Region').Confirmed.sum().sort_values( ascending= False).head(20)
```

[15]:
```
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
```

0.1.5 Question.4. In which Region, minimum number of Deaths Cases were recorded?

[16]:
```
data.groupby('Region').Deaths.sum().sort_values( ascending= True).head(50)
```

[16]:
```
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
```

0.1.6 Question.5. How many Confirmed, Deaths & Recovered cases were Reported from india till 29 april 2020?

[18]:
```
data[data.Region == 'India']
```

[18]:

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

0.1.7 Question.6A. Sort the entire data wrt No. of Confirmed cases in ascending order.

[19]:
```
data.sort_values( by = ['Confirmed'], ascending= True).head(20)
```

[19]:

	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
59	4/29/2020	NaN	Gambia	10	1	8
318	4/29/2020	Yukon	Canada	11	0	0
217	4/29/2020	Greenland	Denmark	11	0	11
256	4/29/2020	Montserrat	UK	11	1	2
144	4/29/2020	NaN	Seychelles	11	0	6
27	4/29/2020	NaN	Burundi	11	1	4
306	4/29/2020	Turks and Caicos Islands	UK	11	1	2
118	4/29/2020	NaN	Nicaragua	12	0	1
215	4/29/2020	Grand Princess	Canada	12	0	2
206	4/29/2020	Falkland Islands (Malvinas)	UK	13	0	2
270	4/29/2020	Northern Mariana Islands	US	14	2	0
136	4/29/2020	NaN	Saint Kitts and Nevis	15	0	2
138	4/29/2020	NaN	Saint Vincent and the Grenadines	15	0	2
45	4/29/2020	NaN	Dominica	16	0	2
201	4/29/2020	Curacao	Netherlands	16	0	2
114	4/29/2020	NaN	Namibia	16	0	2
137	4/29/2020	NaN	Saint Lucia	16	0	2
260	4/29/2020	New Caledonia	France	18	0	2

306	12	1	5
118	13	3	7
215	13	0	0
206	13	0	11
270	14	2	0
136	15	0	4
138	16	0	8
45	16	0	13
201	16	1	13
114	16	0	8
137	17	0	15
260	18	0	17

0.1.8 Question.6B. Sort the entire data wrt No. of Recovered cases in descending order.

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

[20]:

	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
..	…	…	…	…	…	…
258	4/29/2020	Nevada	US	4934	230	0
257	4/29/2020	Nebraska	US	3851	56	0
255	4/29/2020	Montana	US	451	16	0
254	4/29/2020	Missouri	US	7660	338	0
274	4/29/2020	Ohio	US	17303	937	0

[304 rows x 6 columns]