Covid 19 Project

January 26, 2024

1 Rakesh Kumar

2 Working on Real Project with python on COVID -19 Dataset

(A part of big data Analysis) covid -19 SMALL DATASET

```python
[22]: import pandas as pd
      co=pd.read_csv(r"C:\Users\Dell\Desktop\Covid 19 Data Files\Covid 19.csv")
```

```python
[2]: co
```

```
[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
     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]
```

```python
[3]: co.head(5)
```

```
[3]:       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
```

```python
[5]: co.tail(5)
```

```
[5]:          Date      State          Region  Confirmed  Deaths  Recovered
     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
```

```
[6]: co.shape
```

```
[6]: (321, 6)
```

```
[7]: co.index
```

```
[7]: RangeIndex(start=0, stop=321, step=1)
```

```
[8]: co.columns
```

```
[8]: Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'],
     dtype='object')
```

```
[10]: co.count()
```

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

```
[12]: co.isnull().sum()
```

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

```
[13]: co.notnull().sum()
```

```
[13]: Date         321
      State        140
      Region       321
      Confirmed    321
      Deaths       321
      Recovered    321
```

```
dtype: int64
```

[15]: `co.notna().sum()`

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

[16]: `co.nunique()`

```
[16]: Date           1
      State        137
      Region       187
      Confirmed    282
      Deaths       142
      Recovered    195
      dtype: int64
```

[17]: `co.value_counts()`

```
[17]: Date        State                  Region          Confirmed  Deaths  Recovered
      4/29/2020   Alabama                US              6912       256     0
      1
                  Oregon                 US              2446       101     0
      1
                  Northern Territory     Australia       28         0       25
      1
                  Northwest Territories  Canada          5          0       0
      1
                  Nova Scotia            Canada          935        28      0
      1
       ..
                  Guangxi                Mainland China  254        2       252
      1
                  Guangdong              Mainland China  1588       8       1557
      1
                  Guam                   US              141        5       0
      1
                  Guadeloupe             France          149        12      82
      1
                  Zhejiang               Mainland China  1268       1       1263
      1
      Name: count, Length: 140, dtype: int64
```

```
[19]: co.dtypes
```

```
[19]: Date        object
      State       object
      Region      object
      Confirmed    int64
      Deaths       int64
      Recovered    int64
      dtype: object
```

```
[20]: co
```

```
[20]:           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]
```

```
[21]: co
```

```
[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
      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]
```

```
[23]: #  2.
      #  import seaborn as sns
      #  import matplotlib.pyplot as plt
      #  sns.heatmap (df.null())
```
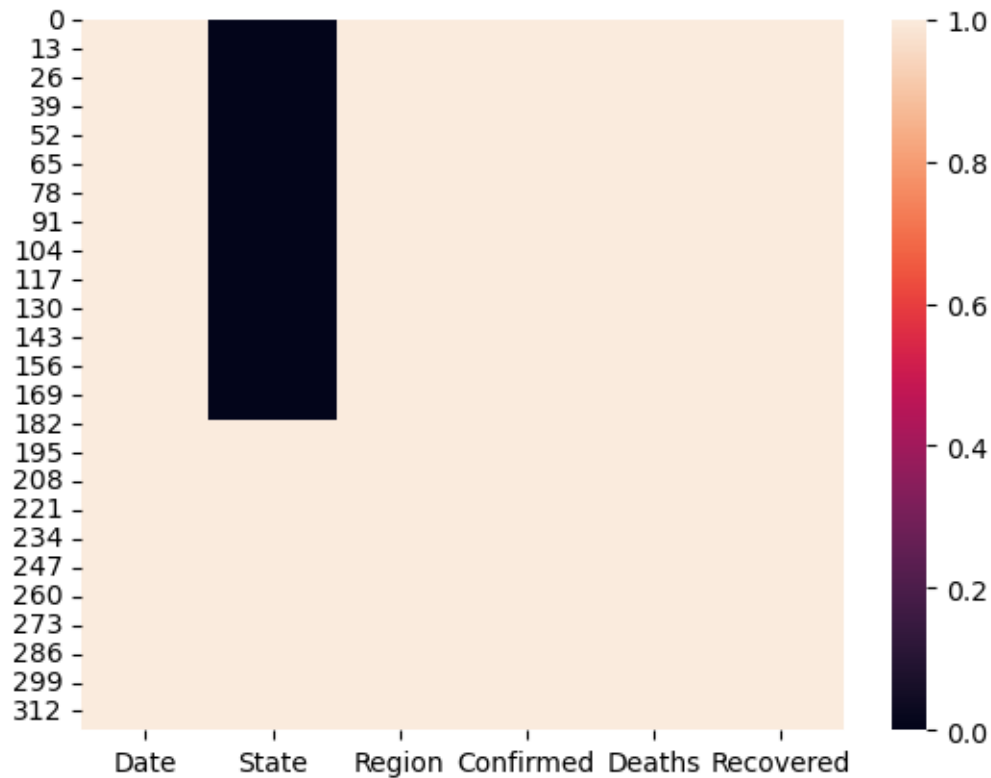
```
#  plt.show()
```

[25]:
```
import seaborn as sns
import matplotlib.pyplot as plt
sns.heatmap(co.isnull())
```

[25]: <Axes: >

[26]:
```
sns.heatmap(co.notnull())
```

[26]: <Axes: >

3 Q. 1) Show the number of Confirmed , Deaths and Recoverd cases in eah Region.

```
[59]: co.groupby("Region").sum()
```

```
[59]:                     Date State  Confirmed  Deaths  Recovered
      Region
      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
      ...                    ...    ...        ...     ...        ...
      West Bank and Gaza 4/29/2020   0        344       2         71
      Western Sahara   4/29/2020     0          6       0          5
      Yemen            4/29/2020     0          6       0          1
      Zambia           4/29/2020     0         97       3         54
      Zimbabwe         4/29/2020     0         32       4          5

      [187 rows x 5 columns]
```

```
[60]: co.groupby("Region").Confirmed.sum().sort_values(ascending=False)
```

```
[60]: 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
```

```
[61]: co.groupby("Region").Deaths.sum().sort_values(ascending=False)
```

```
[61]: Region
      US              60967
      Italy           27682
      UK              26166
      Spain           24275
      France          24121
                        …
      Holy See            0
      Grenada             0
      South Sudan         0
      Nepal               0
      Laos                0
      Name: Deaths, Length: 187, dtype: int64
```

```
[62]: co.groupby("Region").Recovered.sum().sort_values(ascending=False)
```

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

4 Q. 2) Remove all the records where confirmed cases is less than 10.

```
[80]: co=co[~(co.Confirmed <10)]    #Answer    To Remove the records Satisfying a␣
      ↪particular condition.
```

```
[79]: co
```

```
[79]:          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]
```

5 Q. 3) In which Region, Maximum number of Confirmed cases were recorded ?

```
[103]: import pandas as pd
       aa=pd.read_csv(r"C:\Users\Dell\Desktop\Covid 19 Data Files\Covid 19.csv")
```

```
[104]: aa.groupby("Region").Confirmed.sum().sort_values(ascending=False).head(1)   #␣
       ↪Answer   US  1039909
```

```
[104]: Region
       US      1039909
       Name: Confirmed, dtype: int64
```

```
[105]: aa.groupby("Region").Confirmed.sum().sort_values(ascending=False).head(20)   #␣
       ↪Answer   US  1039909
```

```
[105]: 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
```

6 Q. 4) In Which Region ,Minimum number of Deaths cases were recorded ?

[93]: `co`

[93]:
```
          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]
```

[94]:
```python
import pandas as pd
cd=pd.read_csv(r"C:\Users\Dell\Desktop\Covid 19 Data Files\Covid 19.csv")
```

[101]: `cd.groupby("Region").Deaths.sum().sort_values().head(20)        # Answer`

[101]:
```
Region
Laos              0
Mongolia          0
Mozambique        0
Cambodia          0
```

```
Fiji                        0
Namibia                     0
Nepal                       0
Madagascar                  0
Macau                       0
Papua New Guinea            0
Rwanda                      0
Saint Kitts and Nevis       0
Bhutan                      0
Dominica                    0
Central African Republic    0
Saint Lucia                 0
Holy See                    0
Sao Tome and Principe       0
Yemen                       0
Western Sahara              0
Name: Deaths, dtype: int64
```

[ ]:

7 Q. 5) How Many Confirmed ,Deaths & Recovered Cases were reported from india till 29 th april 2020?

[119]: `cd[cd.Region=="India"]`   *#Answer*

[119]:
```
        Date  State Region  Confirmed  Deaths  Recovered
74  4/29/2020   NaN  India      33062    1079       8437
```

[120]: `cd[cd.Region=="India"].sum()`   *#Answer*

[120]:
```
Date         4/29/2020
State                0
Region           India
Confirmed        33062
Deaths            1079
Recovered         8437
dtype: object
```

[ ]:

[122]: `cd.groupby("Region").get_group("India").sum()`   *#Answer*

[122]:
```
Date         4/29/2020
State                0
Region           India
Confirmed        33062
```

```
Deaths          1079
Recovered       8437
dtype: object
```

[123]: `cd.groupby("Region").get_group("India")`   *#Answer*

[123]:
	Date	State	Region	Confirmed	Deaths	Recovered
74	4/29/2020	NaN	India	33062	1079	8437

[124]: `cd`

[124]:
	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]

8 Q. 6-A) Sort the entire data write No. of Confirmed Cases in Ascending Order.

[127]: `cd.sort_values(by="Confirmed",ascending=True)`   *# answer*

[127]:
	Date	State	Region	Confirmed	Deaths	Recovered
285	4/29/2020	Recovered	US	0	0	120720
284	4/29/2020	Recovered	Canada	0	0	20327
203	4/29/2020	Diamond Princess cruise ship	Canada	0	1	0
305	4/29/2020	Tibet	Mainland China	1	0	1
289	4/29/2020	Saint Pierre and Miquelon	France	1	0	0
..	…	…	…	…	…	…
57	4/29/2020	NaN	France	165093	24087	

```
                 48228
168  4/29/2020                              NaN          UK     165221   26097
0
80   4/29/2020                              NaN       Italy     203591   27682
71252
153  4/29/2020                              NaN       Spain     236899   24275
132929
265  4/29/2020                         New York          US     299691   23477
0

[321 rows x 6 columns]
```

```python
cd.sort_values(by="Confirmed",ascending=True).head(50)     # answer
```

```
                 Date                                State
Region  Confirmed  Deaths  Recovered
285  4/29/2020                          Recovered
US           0        0     120720
284  4/29/2020                          Recovered
Canada         0        0      20327
203  4/29/2020       Diamond Princess cruise ship
Canada         0        1        0
305  4/29/2020                              Tibet          Mainland
China          1        0        1
289  4/29/2020          Saint Pierre and Miquelon
France         1        0        0
184  4/29/2020                           Anguilla
UK         3        0        3
192  4/29/2020  Bonaire, Sint Eustatius and Saba
Netherlands      5        0        0
272  4/29/2020             Northwest Territories
Canada         5        0        0
288  4/29/2020                 Saint Barthelemy
France         6        0        6
178  4/29/2020                              NaN
Yemen          6        0        1
194  4/29/2020           British Virgin Islands
UK         6        1        3
177  4/29/2020                              NaN          Western
Sahara         6        0        5
18   4/29/2020                              NaN
Bhutan         7        0        5
126  4/29/2020                              NaN          Papua New
Guinea         8        0        0
140  4/29/2020                              NaN      Sao Tome and
Principe         8        0        4
105  4/29/2020                              NaN
```

Mauritania   8   1   6
98   4/29/2020   NaN   MS
Zaandam   9   2   0
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
144   4/29/2020   NaN
Seychelles   11   0   6
27   4/29/2020   NaN
Burundi   11   1   4
256   4/29/2020   Montserrat
UK   11   1   2
318   4/29/2020   Yukon
Canada   11   0   0
217   4/29/2020   Greenland
Denmark   11   0   11
306   4/29/2020   Turks and Caicos Islands
UK   12   1   5
206   4/29/2020   Falkland Islands (Malvinas)
UK   13   0   11
118   4/29/2020   NaN
Nicaragua   13   3   7
215   4/29/2020   Grand Princess
Canada   13   0   0
270   4/29/2020   Northern Mariana Islands
US   14   2   0
136   4/29/2020   NaN   Saint Kitts and
Nevis   15   0   4
45   4/29/2020   NaN
Dominica   16   0   13
138   4/29/2020   NaN   Saint Vincent and the
Grenadines   16   0   8
114   4/29/2020   NaN
Namibia   16   0   8
201   4/29/2020   Curacao
Netherlands   16   1   13
137   4/29/2020   NaN   Saint
Lucia   17   0   15
260   4/29/2020   New Caledonia
France   18   0   17
281   4/29/2020   Qinghai   Mainland
China   18   0   18
55   4/29/2020   NaN
Fiji   18   0   12

13

```
16   4/29/2020                                        NaN
Belize          18      2          9
90   4/29/2020                                        NaN
Laos            19      0          7
64   4/29/2020                                        NaN
Grenada         20      0          13
21   4/29/2020                                        NaN
Botswana        23      1          5
5    4/29/2020                                        NaN          Antigua and
Barbuda         24      3          11
163  4/29/2020                                        NaN                Timor-
Leste           24      0          6
279  4/29/2020              Prince Edward Island
Canada          27      0          0
4    4/29/2020                                        NaN
Angola          27      2          7
271  4/29/2020              Northern Territory
Australia       28      0          25
180  4/29/2020                                        NaN
Zimbabwe        32      4          5
152  4/29/2020                                        NaN                South
Sudan           34      0          0
```

9 Q. 6-B) Sort the entire data write no. of recovered cases in descending order.

```
[129]:  cd.sort_values(by="Recovered",ascending=False)      #answer
```

```
[129]:           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
        ..         ...             ...       ...        ...     ...        ...
        248  4/29/2020        Maryland        US      20849    1078          0
        246  4/29/2020        Manitoba    Canada        275       6          0
        243  4/29/2020       Louisiana        US      27660    1845          0
        241  4/29/2020        Kentucky        US       4537     234          0
        215  4/29/2020  Grand Princess    Canada         13       0          0

        [321 rows x 6 columns]
```

```
[133]:  cd.sort_values(by="Recovered",ascending=False).head(50)      #answer
```

```
[133]:        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   4/29/2020        NaN          Israel     15834     215
       8233
       33   4/29/2020        NaN           Chile     14885     216
       8057
       42   4/29/2020        NaN         Denmark      9008     443
       6366
       101  4/29/2020        NaN        Malaysia      5945     100
```

	Date	Province/State	Country	Confirmed	Deaths	Recovered
						4087
133	4/29/2020	NaN	Romania	11978	693	3569
124	4/29/2020	NaN	Pakistan	15525	343	3425
97	4/29/2020	NaN	Luxembourg	3769	89	3134
41	4/29/2020	NaN	Czech Republic	7579	227	3108
130	4/29/2020	NaN	Poland	12640	624	3025
141	4/29/2020	NaN	Saudi Arabia	21402	157	2953
56	4/29/2020	NaN	Finland	4906	206	2800
162	4/29/2020	NaN	Thailand	2947	54	2665
83	4/29/2020	NaN	Japan	13895	413	2368
171	4/29/2020	NaN	United Arab Emirates	11929	98	2329
264	4/29/2020	New South Wales	Australia	3016	40	2284
150	4/29/2020	NaN	South Africa	5350	103	2073
14	4/29/2020	NaN	Belarus	13181	84	2072
2	4/29/2020	NaN	Algeria	3848	444	1702
73	4/29/2020	NaN	Iceland	1797	10	1656
220	4/29/2020	Guangdong	Mainland China	1588	8	1557
47	4/29/2020	NaN	Ecuador	24675	883	1557
131	4/29/2020	NaN	Portugal	24505	973	1470
11	4/29/2020	NaN	Bahrain	2921	8	1455
34	4/29/2020	NaN	Colombia	6207	278	1411
75	4/29/2020	NaN	Indonesia	9771	784	1391
88	4/29/2020	NaN	Kuwait	3740	24	1389
77	4/29/2020	NaN	Iraq	2003	92	1346

```
48    4/29/2020              NaN            Egypt       5268     380
1335
309   4/29/2020         Victoria        Australia       1361      18
1291
38    4/29/2020              NaN          Croatia       2062      67
1288
9     4/29/2020              NaN       Azerbaijan       1766      23
1267
```

10 Rakesh Kumar

```
[ ]:
```