Covid 19 data analysis

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

In [4]:
```
df=pd.read_csv('C:\\Users\\shriy\\OneDrive\\Documents\\covid19_data.csv')
```

In [5]:
```
df.head(10)
```

Out[5]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	female_smoke
0	AFG	Asia	Afghanistan	24-02-2020	5.0	5.0	NaN	NaN	NaN	NaN	...	Na
1	AFG	Asia	Afghanistan	25-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	...	Na
2	AFG	Asia	Afghanistan	26-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	...	Na
3	AFG	Asia	Afghanistan	27-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	...	Na
4	AFG	Asia	Afghanistan	28-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	...	Na
5	AFG	Asia	Afghanistan	29-02-2020	5.0	0.0	0.714	NaN	NaN	NaN	...	Na
6	AFG	Asia	Afghanistan	01-03-2020	5.0	0.0	0.714	NaN	NaN	NaN	...	Na

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	female_smoke
7	AFG	Asia	Afghanistan	02-03-2020	5.0	0.0	0.000	NaN	NaN	NaN	...	Na
8	AFG	Asia	Afghanistan	03-03-2020	5.0	0.0	0.000	NaN	NaN	NaN	...	Na
9	AFG	Asia	Afghanistan	04-03-2020	5.0	0.0	0.000	NaN	NaN	NaN	...	Na

10 rows × 67 columns

In [6]:
```
df.drop(['excess_mortality_cumulative_absolute','excess_mortality','excess_mortality_cumulative_per_million','excess_mortality_cum
```

In [7]:
```
df.head()
```

Out[7]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	gdp_per_capita
0	AFG	Asia	Afghanistan	24-02-2020	5.0	5.0	NaN	NaN	NaN	NaN	...	1803.987
1	AFG	Asia	Afghanistan	25-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	...	1803.987
2	AFG	Asia	Afghanistan	26-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	...	1803.987
3	AFG	Asia	Afghanistan	27-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	...	1803.987
4	AFG	Asia	Afghanistan	28-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	...	1803.987

5 rows × 63 columns

In [8]:
```
df.drop(['iso_code'],axis=1,inplace=True)
```

In [9]:
```
df.head()
```

Out[9]:

	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per_million	...
0	Asia	Afghanistan	24-02-2020	5.0	5.0	NaN	NaN	NaN	NaN	0.126	...
1	Asia	Afghanistan	25-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	0.126	...
2	Asia	Afghanistan	26-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	0.126	...
3	Asia	Afghanistan	27-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	0.126	...
4	Asia	Afghanistan	28-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	0.126	...

5 rows × 62 columns

In [10]:
```
df.rename(columns={'location':'country','date':'observational_date'},inplace=True)
```

In [11]:
```
df.head()
```

Out[11]:

	continent	country	observational_date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per
0	Asia	Afghanistan	24-02-2020	5.0	5.0	NaN	NaN	NaN	NaN	
1	Asia	Afghanistan	25-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	
2	Asia	Afghanistan	26-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	
3	Asia	Afghanistan	27-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	
4	Asia	Afghanistan	28-02-2020	5.0	0.0	NaN	NaN	NaN	NaN	

5 rows × 62 columns

In [12]:
```
df['observational_date']=pd.to_datetime(df['observational_date'])
```

In [13]:
```
df.head()
```

Out[13]:

	continent	country	observational_date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per
0	Asia	Afghanistan	2020-02-24	5.0	5.0	NaN	NaN	NaN	NaN	
1	Asia	Afghanistan	2020-02-25	5.0	0.0	NaN	NaN	NaN	NaN	
2	Asia	Afghanistan	2020-02-26	5.0	0.0	NaN	NaN	NaN	NaN	
3	Asia	Afghanistan	2020-02-27	5.0	0.0	NaN	NaN	NaN	NaN	
4	Asia	Afghanistan	2020-02-28	5.0	0.0	NaN	NaN	NaN	NaN	

5 rows × 62 columns

In [14]:
```
df.describe()
```

Out[14]:

	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per_million	new_cases_per_millio
count	1.580460e+05	1.580160e+05	1.568650e+05	1.404130e+05	140587.000000	140457.000000	157311.000000	157281.00000

	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per_million	new_cases_per_millio
mean	2.355755e+06	1.086580e+04	1.073002e+04	5.569075e+04	170.920732	170.023846	26673.381473	151.92418
std	1.413462e+07	8.066765e+04	7.760773e+04	2.920949e+05	832.430023	810.618130	45110.679176	634.74365
min	1.000000e+00	-7.493700e+04	-6.223000e+03	1.000000e+00	-1918.000000	-232.143000	0.001000	-13876.28200
25%	1.819000e+03	1.000000e+00	6.571000e+00	7.500000e+01	0.000000	0.143000	575.183000	0.04000
50%	2.397050e+04	7.700000e+01	1.028570e+02	7.400000e+02	2.000000	2.429000	4400.740000	11.06400
75%	2.785592e+05	1.020000e+03	1.080857e+03	6.986000e+03	19.000000	20.429000	34336.762500	96.31700
max	4.059612e+08	4.235318e+06	3.438947e+06	5.789567e+06	18057.000000	14705.714000	535910.138000	51427.49100

8 rows × 58 columns

In [15]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 160934 entries, 0 to 160933
Data columns (total 62 columns):
 #   Column                                Non-Null Count   Dtype
---  ------                                --------------   -----
 0   continent                             151277 non-null  object
 1   country                               160934 non-null  object
 2   observational_date                    160934 non-null  datetime64[ns]
 3   total_cases                           158046 non-null  float64
 4   new_cases                             158016 non-null  float64
 5   new_cases_smoothed                    156865 non-null  float64
 6   total_deaths                          140413 non-null  float64
 7   new_deaths                            140587 non-null  float64
 8   new_deaths_smoothed                   140457 non-null  float64
 9   total_cases_per_million               157311 non-null  float64
 10  new_cases_per_million                 157281 non-null  float64
 11  new_cases_smoothed_per_million        156135 non-null  float64
 12  total_deaths_per_million              139691 non-null  float64
 13  new_deaths_per_million                139865 non-null  float64
 14  new_deaths_smoothed_per_million       139735 non-null  float64
 15  reproduction_rate                     121653 non-null  float64
 16  icu_patients                          22669 non-null   float64
 17  icu_patients_per_million              22669 non-null   float64
```

```
18  hosp_patients                                 23440 non-null   float64
19  hosp_patients_per_million                     23440 non-null   float64
20  weekly_icu_admissions                         5164 non-null    float64
21  weekly_icu_admissions_per_million             5164 non-null    float64
22  weekly_hosp_admissions                        10476 non-null   float64
23  weekly_hosp_admissions_per_million            10476 non-null   float64
24  new_tests                                     66141 non-null   float64
25  total_tests                                   67328 non-null   float64
26  total_tests_per_thousand                      67328 non-null   float64
27  new_tests_per_thousand                        66141 non-null   float64
28  new_tests_smoothed                            81560 non-null   float64
29  new_tests_smoothed_per_thousand               81560 non-null   float64
30  positive_rate                                 76322 non-null   float64
31  tests_per_case                                75762 non-null   float64
32  tests_units                                   83793 non-null   object
33  total_vaccinations                            42775 non-null   float64
34  people_vaccinated                             40731 non-null   float64
35  people_fully_vaccinated                       38038 non-null   float64
36  total_boosters                                15154 non-null   float64
37  new_vaccinations                              35423 non-null   float64
38  new_vaccinations_smoothed                     79276 non-null   float64
39  total_vaccinations_per_hundred                42775 non-null   float64
40  people_vaccinated_per_hundred                 40731 non-null   float64
41  people_fully_vaccinated_per_hundred           38038 non-null   float64
42  total_boosters_per_hundred                    15154 non-null   float64
43  new_vaccinations_smoothed_per_million         79276 non-null   float64
44  new_people_vaccinated_smoothed                78083 non-null   float64
45  new_people_vaccinated_smoothed_per_hundred    78083 non-null   float64
46  stringency_index                              126087 non-null  float64
47  population                                    159882 non-null  float64
48  population_density                            143243 non-null  float64
49  median_age                                    133496 non-null  float64
50  aged_65_older                                 132048 non-null  float64
51  aged_70_older                                 132780 non-null  float64
52  gdp_per_capita                                134118 non-null  float64
53  extreme_poverty                               88375 non-null   float64
54  cardiovasc_death_rate                         132466 non-null  float64
55  diabetes_prevalence                           139407 non-null  float64
56  female_smokers                                102727 non-null  float64
57  male_smokers                                  101253 non-null  float64
58  handwashing_facilities                        66357 non-null   float64
59  hospital_beds_per_thousand                    119789 non-null  float64
60  life_expectancy                               150216 non-null  float64
61  human_development_index                       131941 non-null  float64
```

```
dtypes: datetime64[ns](1), float64(58), object(3)
memory usage: 76.1+ MB
```

In [16]:
```
df=df.fillna("NA")
```

In [17]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 160934 entries, 0 to 160933
Data columns (total 62 columns):
 #   Column                                 Non-Null Count   Dtype
---  ------                                 --------------   -----
 0   continent                              160934 non-null  object
 1   country                                160934 non-null  object
 2   observational_date                     160934 non-null  datetime64[ns]
 3   total_cases                            160934 non-null  object
 4   new_cases                              160934 non-null  object
 5   new_cases_smoothed                     160934 non-null  object
 6   total_deaths                           160934 non-null  object
 7   new_deaths                             160934 non-null  object
 8   new_deaths_smoothed                    160934 non-null  object
 9   total_cases_per_million                160934 non-null  object
 10  new_cases_per_million                  160934 non-null  object
 11  new_cases_smoothed_per_million         160934 non-null  object
 12  total_deaths_per_million               160934 non-null  object
 13  new_deaths_per_million                 160934 non-null  object
 14  new_deaths_smoothed_per_million        160934 non-null  object
 15  reproduction_rate                      160934 non-null  object
 16  icu_patients                           160934 non-null  object
 17  icu_patients_per_million               160934 non-null  object
 18  hosp_patients                          160934 non-null  object
 19  hosp_patients_per_million              160934 non-null  object
 20  weekly_icu_admissions                  160934 non-null  object
 21  weekly_icu_admissions_per_million      160934 non-null  object
 22  weekly_hosp_admissions                 160934 non-null  object
 23  weekly_hosp_admissions_per_million     160934 non-null  object
 24  new_tests                              160934 non-null  object
 25  total_tests                            160934 non-null  object
 26  total_tests_per_thousand               160934 non-null  object
 27  new_tests_per_thousand                 160934 non-null  object
 28  new_tests_smoothed                     160934 non-null  object
 29  new_tests_smoothed_per_thousand        160934 non-null  object
```

```
 30  positive_rate                                160934 non-null  object
 31  tests_per_case                               160934 non-null  object
 32  tests_units                                  160934 non-null  object
 33  total_vaccinations                           160934 non-null  object
 34  people_vaccinated                            160934 non-null  object
 35  people_fully_vaccinated                      160934 non-null  object
 36  total_boosters                               160934 non-null  object
 37  new_vaccinations                             160934 non-null  object
 38  new_vaccinations_smoothed                    160934 non-null  object
 39  total_vaccinations_per_hundred               160934 non-null  object
 40  people_vaccinated_per_hundred                160934 non-null  object
 41  people_fully_vaccinated_per_hundred          160934 non-null  object
 42  total_boosters_per_hundred                   160934 non-null  object
 43  new_vaccinations_smoothed_per_million        160934 non-null  object
 44  new_people_vaccinated_smoothed               160934 non-null  object
 45  new_people_vaccinated_smoothed_per_hundred   160934 non-null  object
 46  stringency_index                             160934 non-null  object
 47  population                                   160934 non-null  object
 48  population_density                           160934 non-null  object
 49  median_age                                   160934 non-null  object
 50  aged_65_older                                160934 non-null  object
 51  aged_70_older                                160934 non-null  object
 52  gdp_per_capita                               160934 non-null  object
 53  extreme_poverty                              160934 non-null  object
 54  cardiovasc_death_rate                        160934 non-null  object
 55  diabetes_prevalence                          160934 non-null  object
 56  female_smokers                               160934 non-null  object
 57  male_smokers                                 160934 non-null  object
 58  handwashing_facilities                       160934 non-null  object
 59  hospital_beds_per_thousand                   160934 non-null  object
 60  life_expectancy                              160934 non-null  object
 61  human_development_index                      160934 non-null  object
dtypes: datetime64[ns](1), object(61)
memory usage: 76.1+ MB
```

In [18]:
```
df2=df.groupby(['country','observational_date'])[['total_cases','new_cases']].sum().reset_index()
```

In [19]:
```
df2.head()
```

Out[19]:

	country	observational_date	total_cases	new_cases
0	Afghanistan	2020-01-03	5.0	0.0
1	Afghanistan	2020-01-04	192.0	26.0
2	Afghanistan	2020-01-05	2171.0	344.0
3	Afghanistan	2020-01-06	15836.0	656.0
4	Afghanistan	2020-01-07	31848.0	403.0

In [ ]: