

Exploratory Data Analysis of COVID-19-Global-Data:

Dataset:

The datasets consist of 209271 rows and 8 columns i.e
'Date_reported', 'Country_code', 'Country', 'WHO_region', 'New_cases', 'Cumulative_cases',
'New_deaths', 'Cumulative_deaths'.

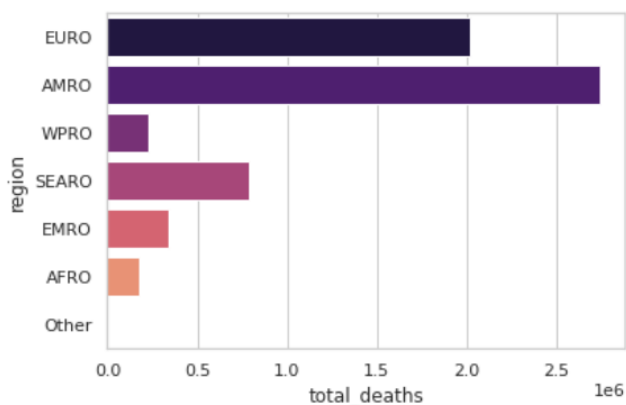
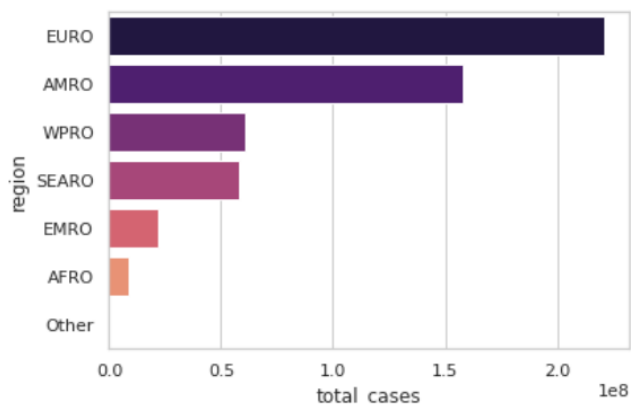
Column	Non-Null Count	Dtype
0 Date_reported	209271 non-null	object
1 Country_code	208388 non-null	object
2 Country	209271 non-null	object
3 WHO_region	209271 non-null	object
4 New_cases	209271 non-null	int64
5 Cumulative_cases	209271 non-null	int64
6 New_deaths	209271 non-null	int64
7 Cumulative_deaths	209271 non-null	int64

- Country_code have 236 unique values and 7 different region in the WHO_region. The time interval is of 883 days for each one of 236 country.
- New_cases and New_deaths are the no. of new cases and deaths recorded on each day where the Cumulative cases and death keep the sum of total cases and deaths.
- Country_code is missing for one country and the country name is also not there (Other is written) so we will replace the country code " " with Ot as the country code.

Q1. Which region is mostly affected by Covid-19?

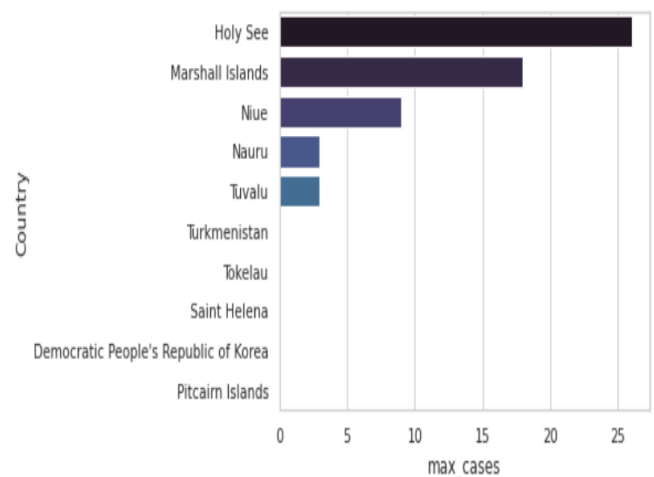
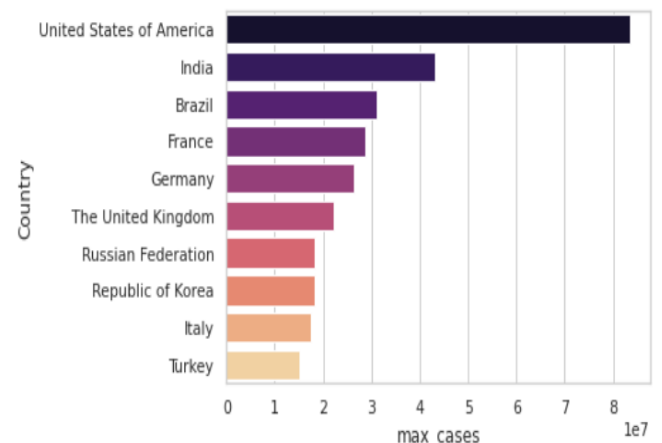
- The EURO region and the AMRO were affected very much as compared to the remaining region as it recorded nearly 221373828 total cases and 2014515 deaths in EURO and 157725452 total cases and 2744395 total deaths in AMRO
- And other recorded cases and deaths are as follows:
- WPRO 60771271 231635
- SEARO 58157405 788853
- EMRO 21780549 342872
- AFRO 9007048 172686
- Other 764 13

The more number of cases doesn't imply that more deaths also. One of the reasons of less death rate might be good medical facilities available in that region.



Q2. Which countries have the most and least number of total cases?

Country	max_cases
United States of America	83551386
India	43168585
Brazil	31060017
France	28687573
Germany	26452148
The United Kingdom	22305897
Russian Federation	18343964
Republic of Korea	18141835
Italy	17457950
Turkey	15073722

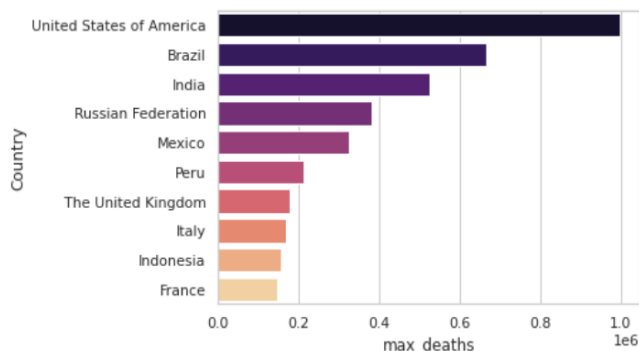


Country	max_cases
Holy See	26
Marshall Islands	18
Niue	9
Nauru	3
Tuvalu	3
Turkmenistan	0
Tokelau	0
Saint Helena	0
Democratic People's Republic of Korea	0
Pitcairn Islands	0

It's totally clear from the graphs that the most affected countries are the United States of America, India, Brazil, France, Germany, The United Kingdom, the Russian Federation, the Republic of Korea, Italy, and Turkey. Out of these 10 Countries, 8 are from EURO and AMRO regions which are mostly affected by the covid.

Similarly, the least affected counties are Holy See, Marshall Islands, Niue, Nauru, Tuvalu, Turkmenistan, Tokelau, Saint Helena, the Democratic People's Republic of Korea, and the Pitcairn Islands. Most of the countries are from the WPRO region.

Q3. Which countries have the most and least number of total deaths?



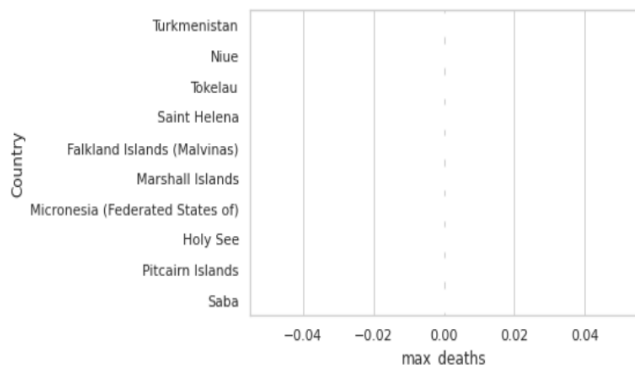
Country	max_deaths
United States of America	998070
Brazil	666801
India	524651
Russian Federation	379363
Mexico	324951
Peru	213205
The United Kingdom	178749
Italy	166835
Indonesia	156604
France	145123

Most dead are recorded United States of America, Brazil, India, Russian Federation, Mexico, Peru, The United Kingdom, Italy, Indonesia, and France. Out of these 10 Countries, 7 are from the top ten countries with most no. of cases.

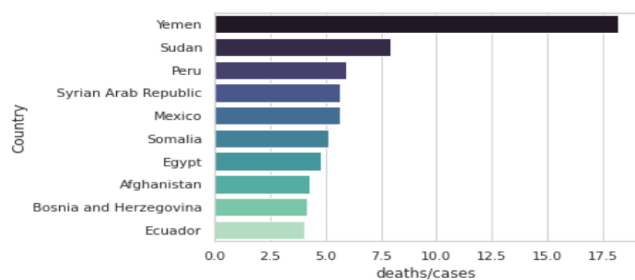
Similarly, the least affected counties are Turkmenistan, Niue, Tokelau, Saint Helena, Falkland Islands (Malvinas), Marshall Islands, Micronesia (Federated States of), Holy See, Pitcairn Islands, and Saba.

Most of the countries are from the WPRO region.

Country	max_deaths
Turkmenistan	0
Niue	0
Tokelau	0
Saint Helena	0
Falkland Islands (Malvinas)	0
Marshall Islands	0
Micronesia (Federated States of)	0
Holy See	0
Pitcairn Islands	0
Saba	0

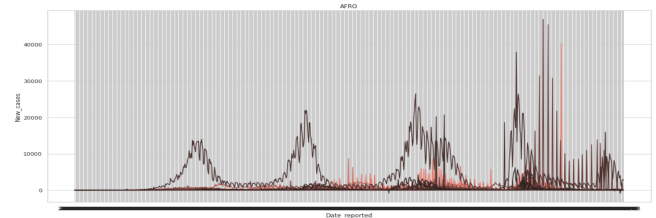
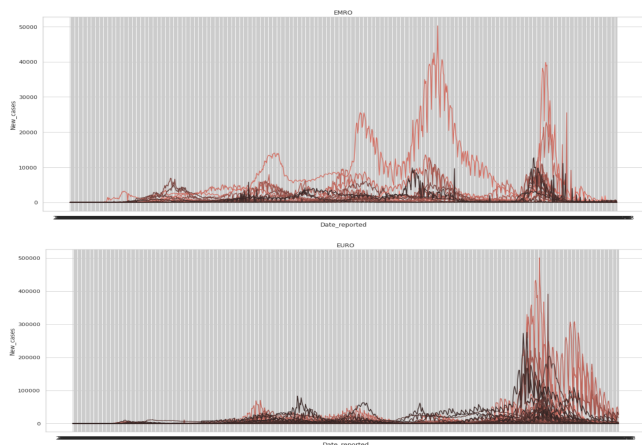


Q4. Which country has highest deaths/cases ratio ?

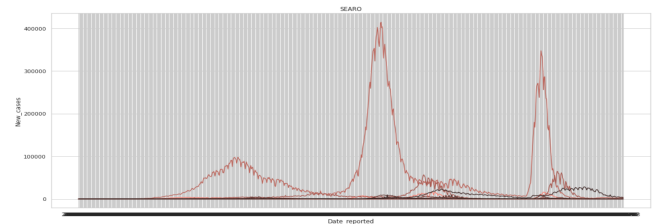
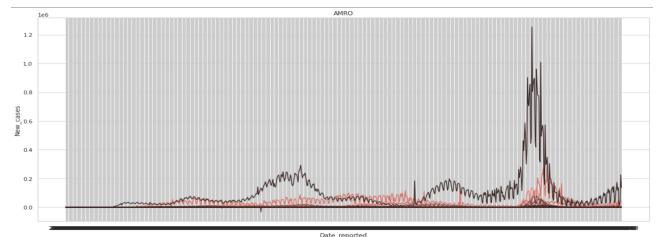
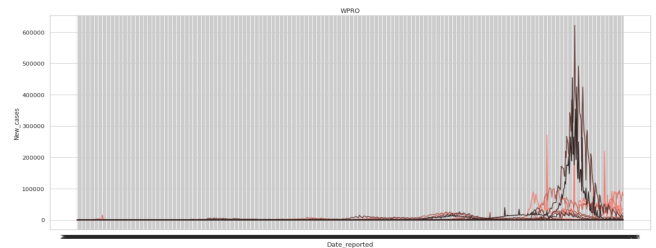


The Death/Cases may be Good parameter than the total of the death. Suprising the death/cases ratio is nearly 18 for Yemen compare to other countries . All this country high death /cases ratio this might be due poor medical facility. All this countries includes Yemen, Sudan, Peru, Syrian Arab Republic, Mexico, Somalia, Egypt, Afghanistan, Bosnia and Herzegovina, and Ecuador.

Q5. Does each regions follow same pattern in cases increase ?



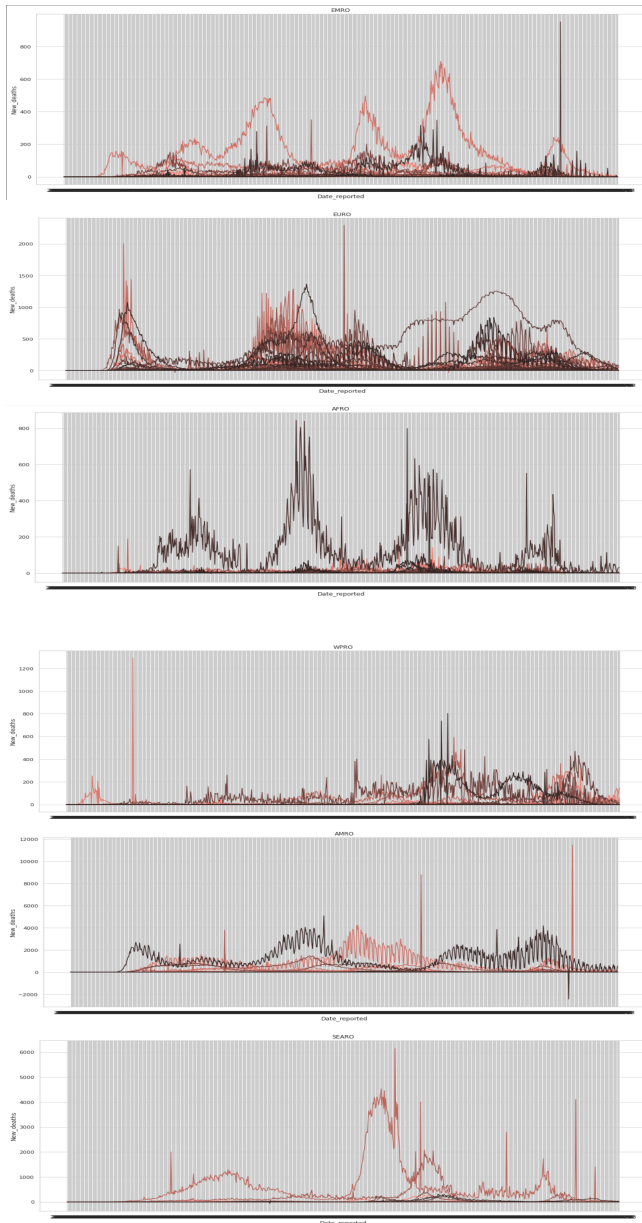
- Most of the country have some what similar pattern in EMRO but one country has affected very much as compare to that of other.
- Euro country has shown the similar pattern increases in cases for all the countries but the number of the increase in cases are varing for each country.
- AFRO region dont have do much countries in it, most of the counties are not affected that much one country has affect very as compare to other and one country has some very sharp peaks.



- WPRO region countries have similar pattern it seems that in the starting phase the covid was not that active in this region but in the last days the covid was detected in significant number.
- AMRO region dont have plenty of countries that have somewhat similar pattern.
- In SEARO region the one country has shown very hugh number od cases as

compare to the other countries in that region.

Q6. Does each region follow same pattern in deaths increase ?



have recorded negative deaths which is not possible we must replace the negative values by the zero.

- One country in AFRO region has very fluctuating graph with very sharp peaks .

CODE:

https://colab.research.google.com/drive/1w1wXJ8m_SGaSZRtW7NHYAFZfNLpOeN9B#scrollTo=CawZ8OROxvPd

- Mostly each region countries follow somewhat similar pattern but one-two countries have more no of deaths as compare to the other remaining countries in the regions.
- As compare to cases graphs the the death graphs are very fluctuating they have a sudden large peaks in the graph as compare to reaming whole graphs.
- In AMRO region some countries also