Project: Investigating Covid 19 data.

Introduction

Covid 19 pandemic has changed the world dramatically and there are a lot of unanswered questions about it. In this notebook,we will try to discover more about it using data informed strategy. Any observation - even small ones - if it is accurately interpreted,It will give us a great weapon in our war with the virus.

Data and Data Sources

1- Timeseres data:

Four Datasets contains time series data of global confirmed, recovered, deaths and vaccinations cases for all the countries in the world till 17-6-2021.

Source: Johns Hopkins Coronavirus Resource Center here. (https://coronavirus.jhu.edu/about/how-to-use-our-data)

2- Demographic data: consists of merging two datasets including:

1- Country
2- Population
3- Population density (people/Km2)
4- Total tests coducted by each country till 17-6-2021
5- Continent
6- Income level
7- Median age of each country.

Note: Combination of data needed work with spreadsheet, manual entry and modification of data so it will be provided after pereperation.

Sources:

A) The United Nations Statistics Division here. (https://unstats.un.org/unsd/demographic-social/products/dyb/index.cshtml#censusdatasets)

B) Worldometers here. (https://www.worldometers.info/coronavirus/)

What is the behavior of Covid19?

Using data and answering these two questions:

- Are Covid waves till now repeating with a pattern,ie are they coming in specific months?
- Does the waves' strength change or not?

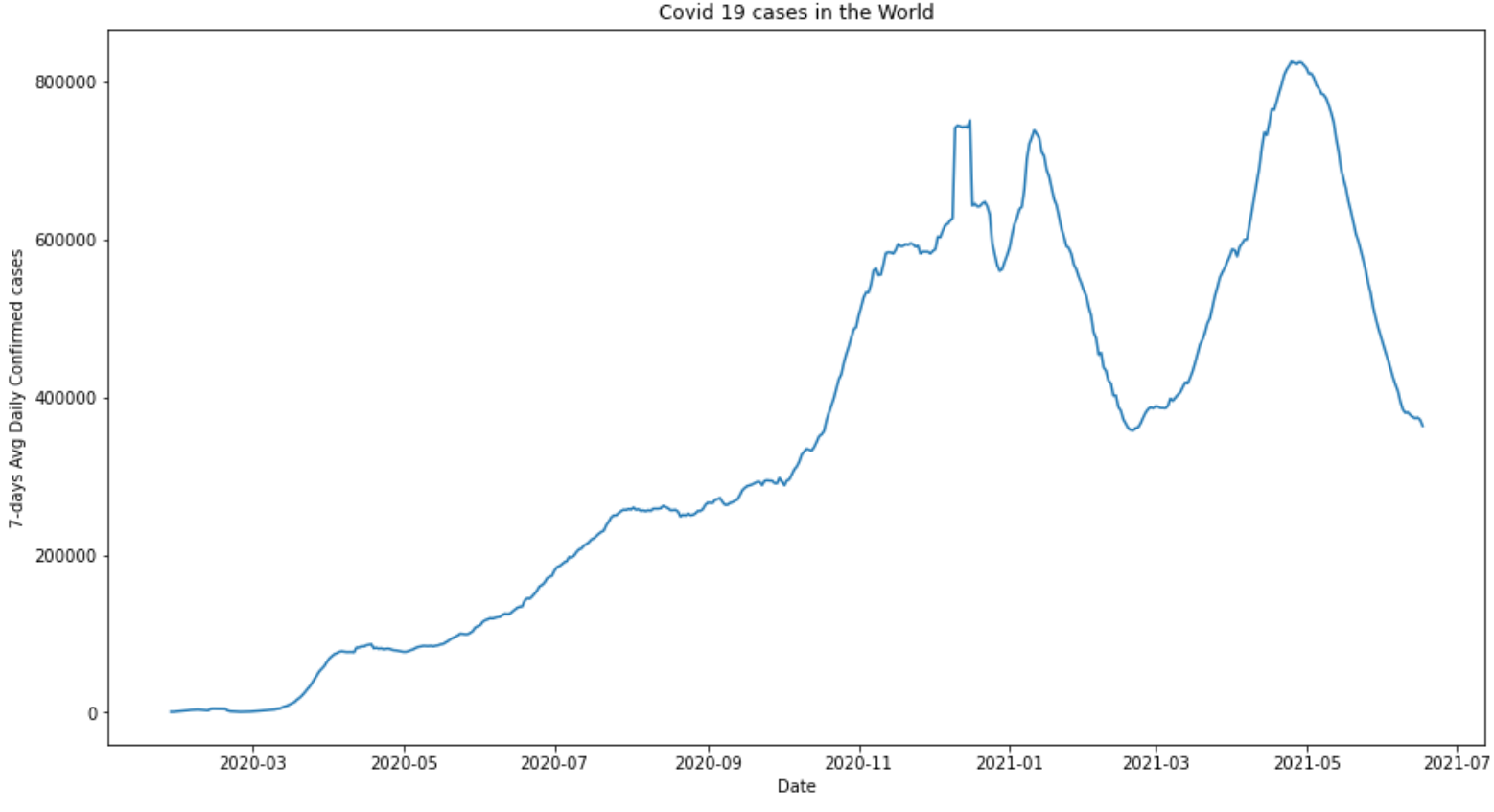
We found that:

1- The occurrence time of waves and their strength are greatly different from one country to another. We definitly can't identify a pattern that all countries follow either in wave strength or time of occurrence.

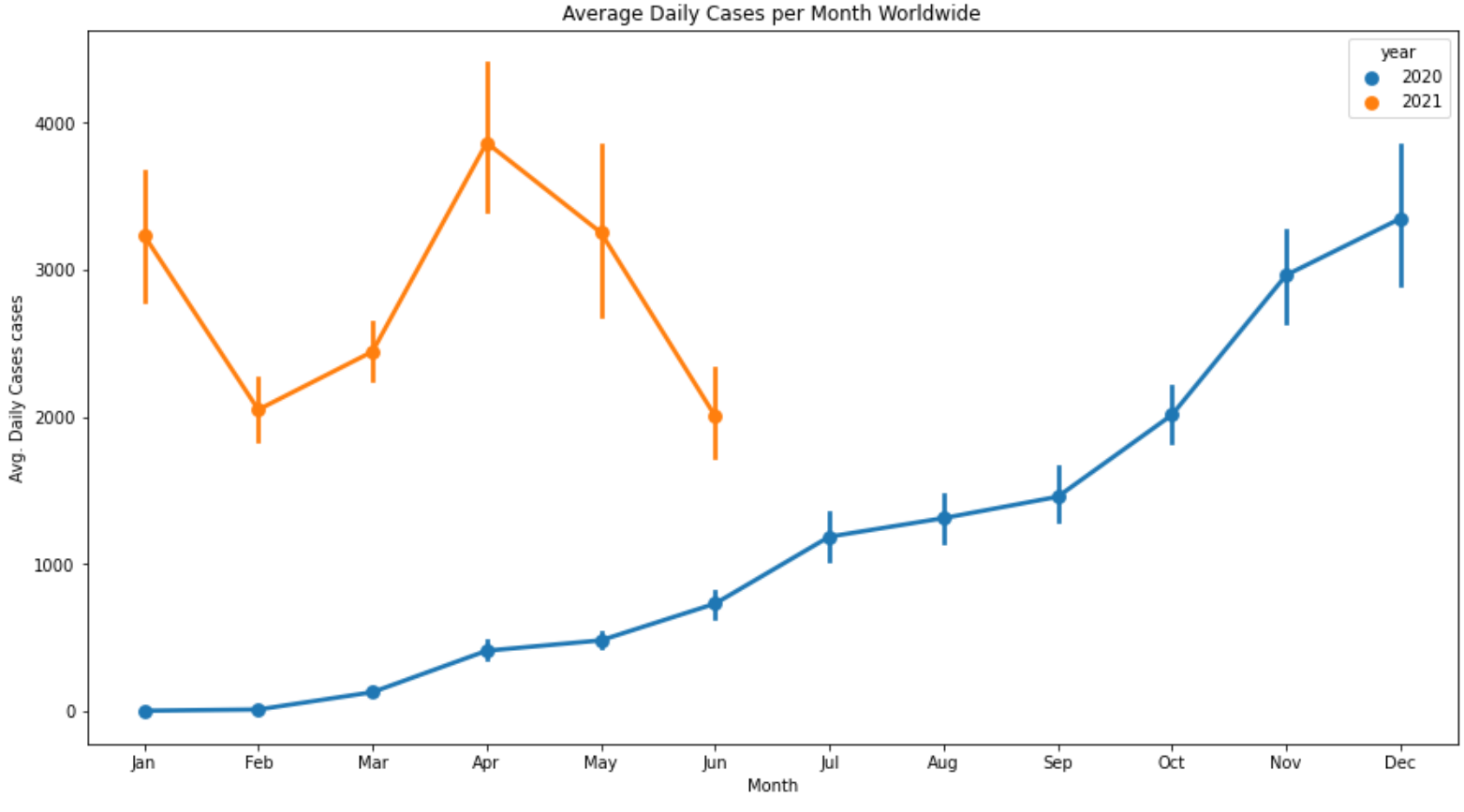
2- Worldwide data may be misleading if it is taken alone as different countries has very different conditions.So, we can't generalize any insights from it.

Here is the Worldwide cases distribution

Out[88]:



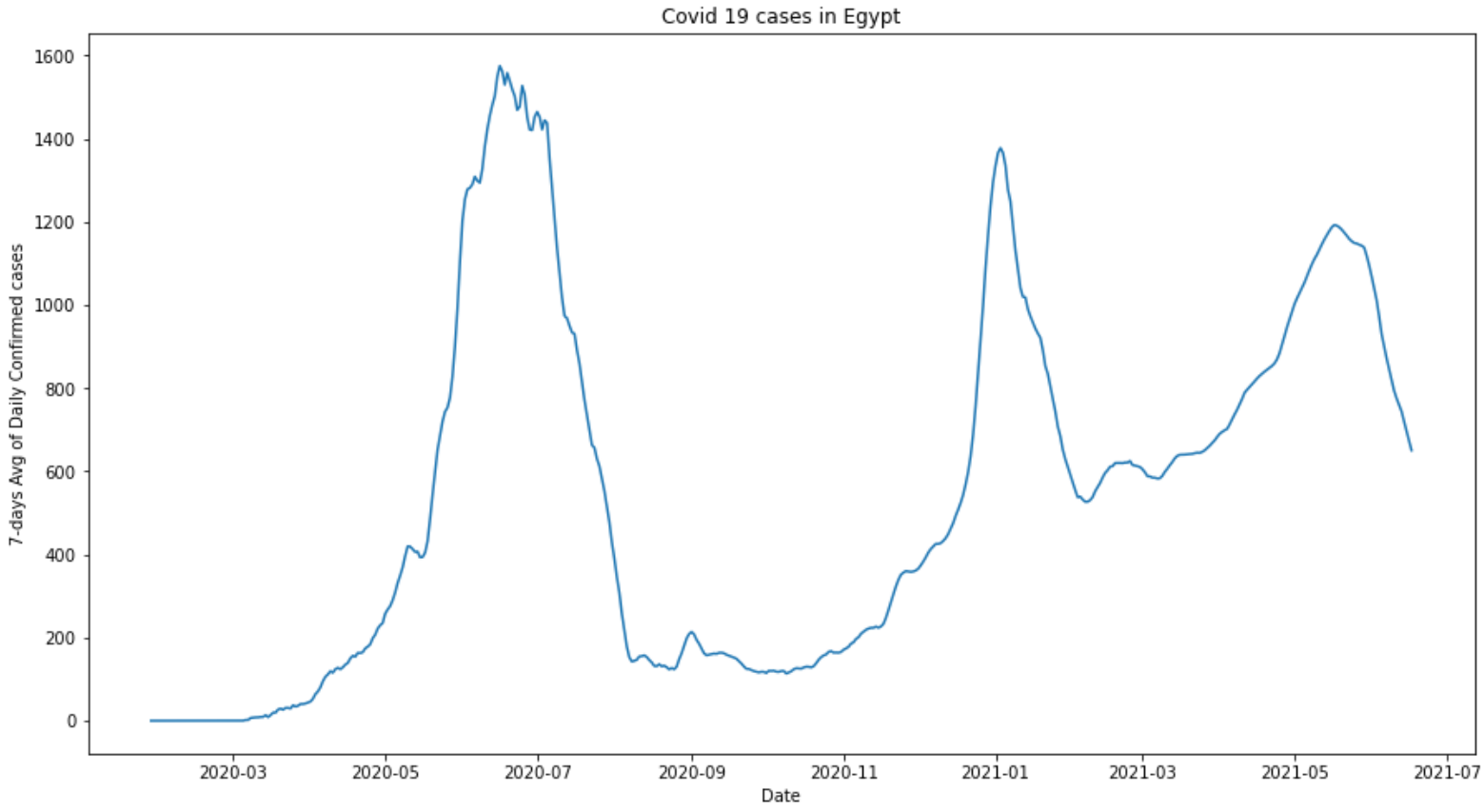
We can see that the wave strength is increasing over time.



Here, the strongest wave came in june 2020.

But for Egypt Daily the case is different

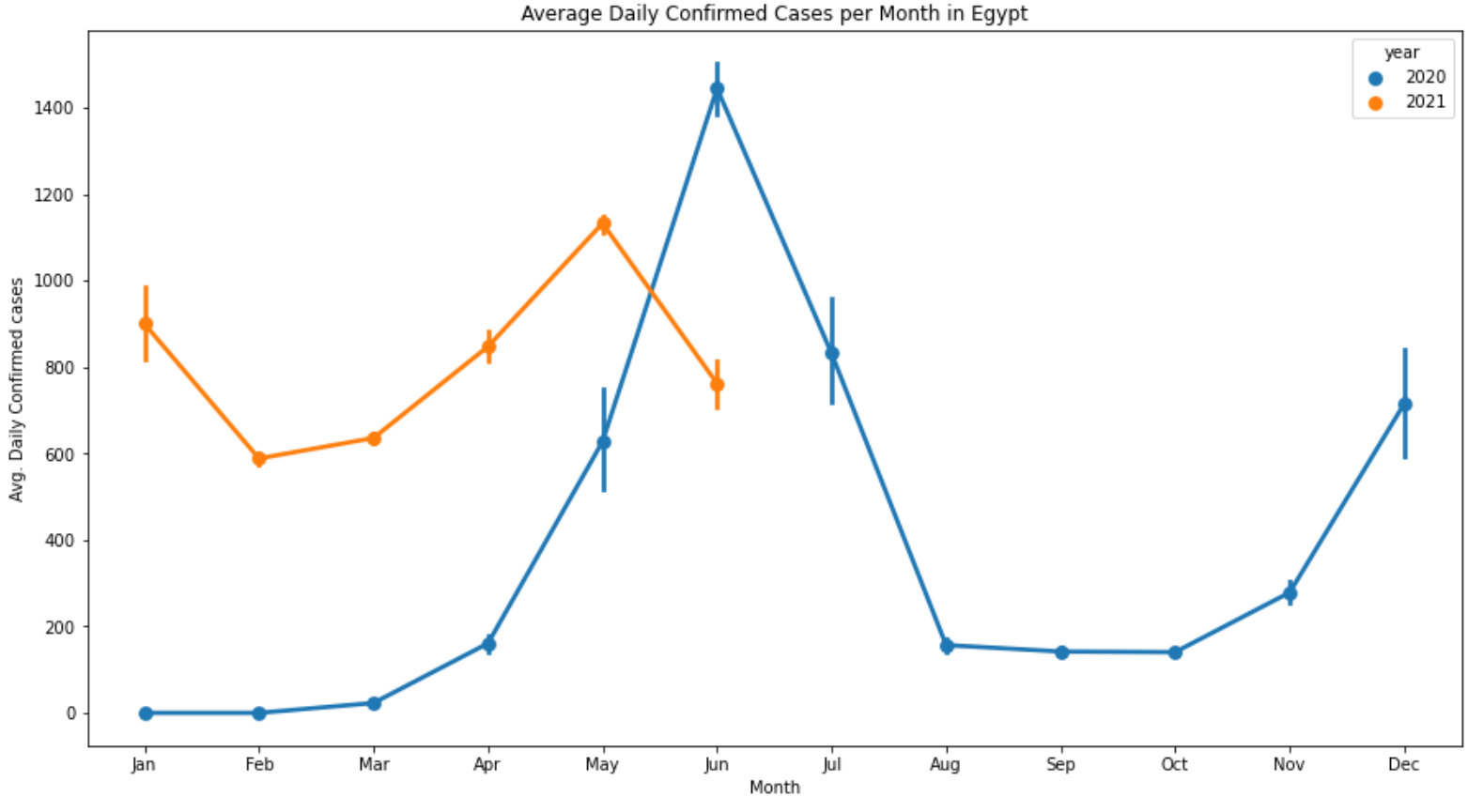
Out[138]:



1- Here, It is clear that wave strength is decreasing over time.

2- There is a reatitive wave started in about Feb 2020 , and also 2021.

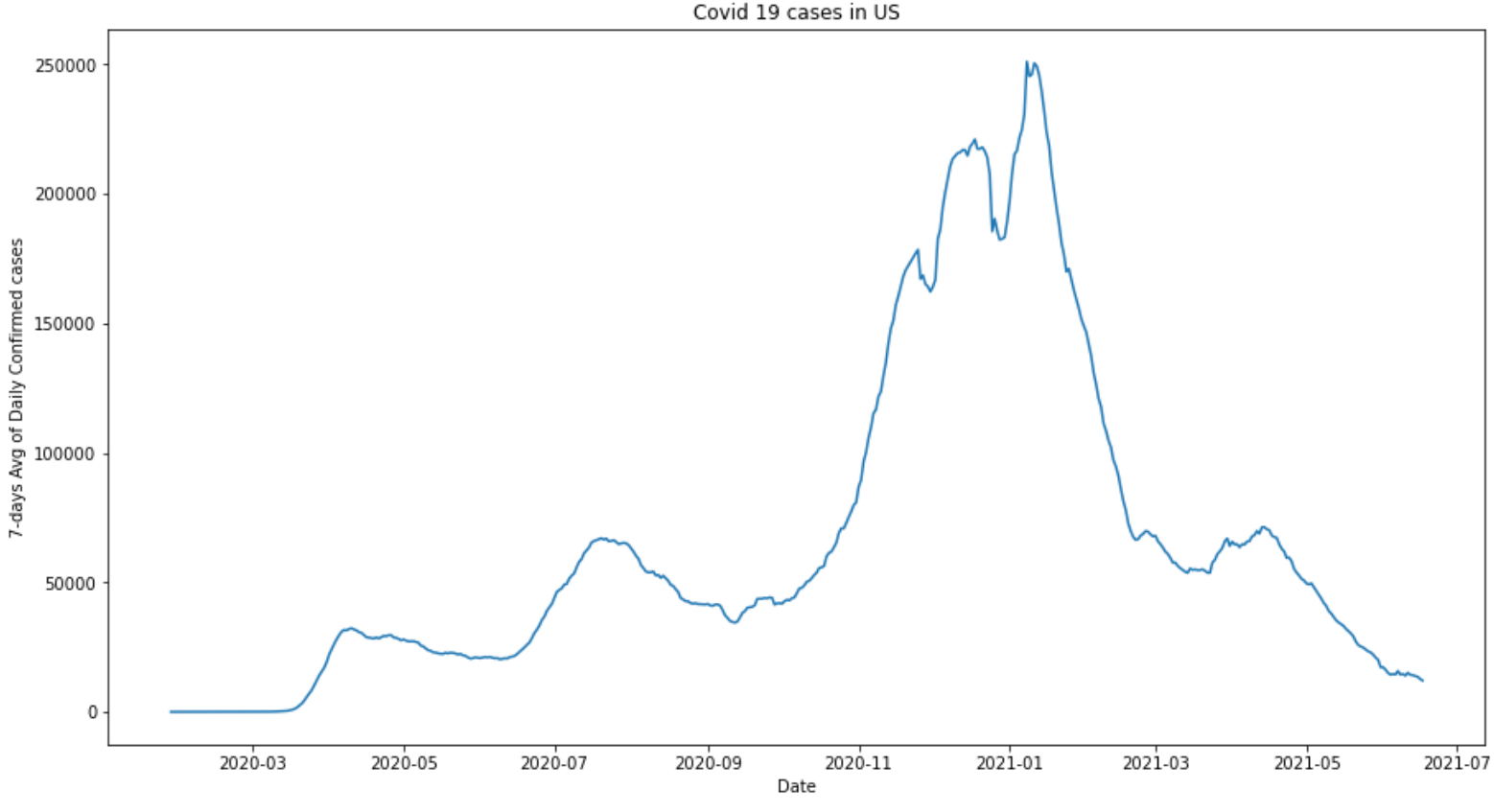
2. There is a relative wave started in about Feb 2020, and also 2021.



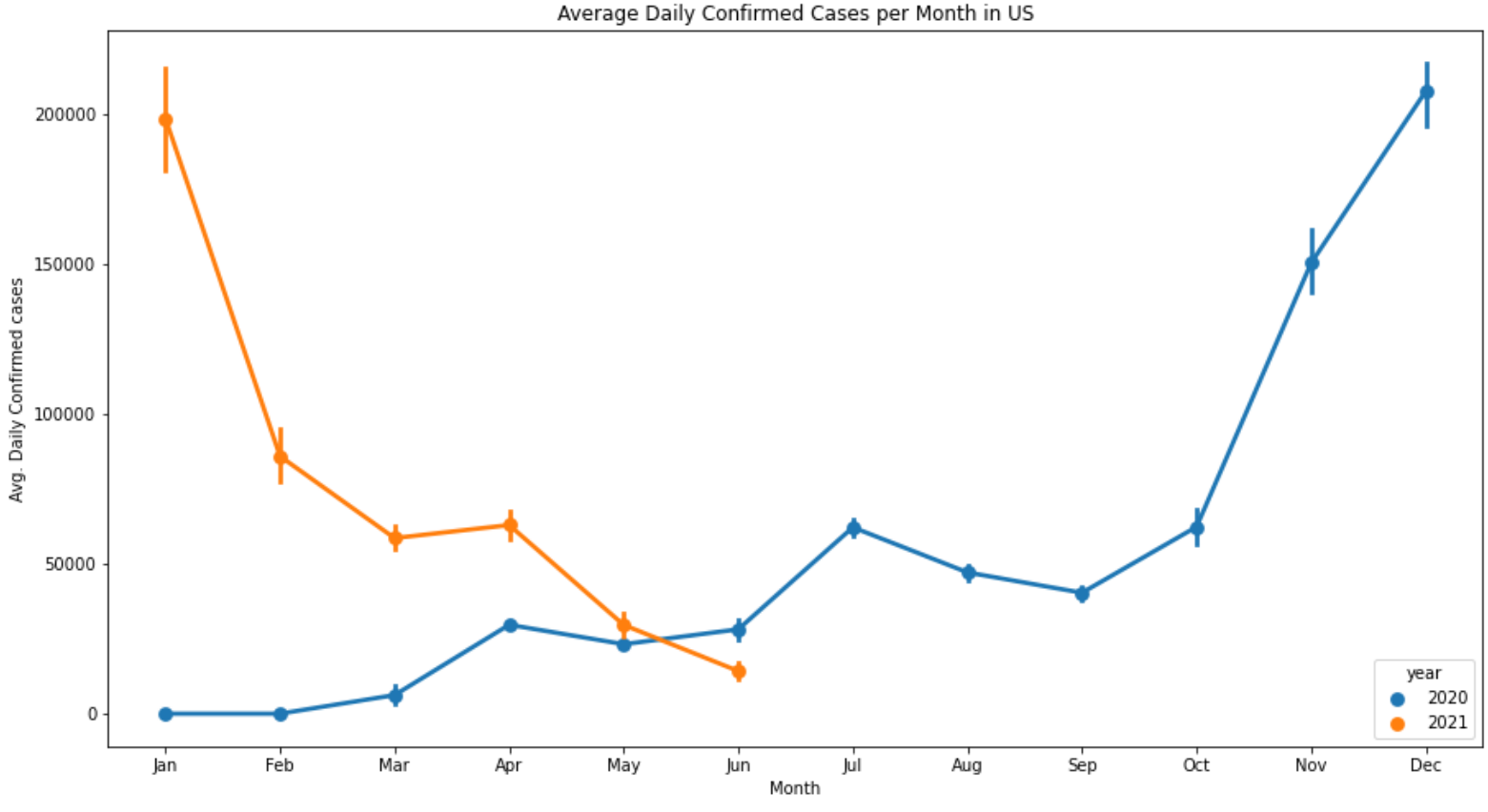
Strongest wave came in june 2020

And for United States

Out[139]:



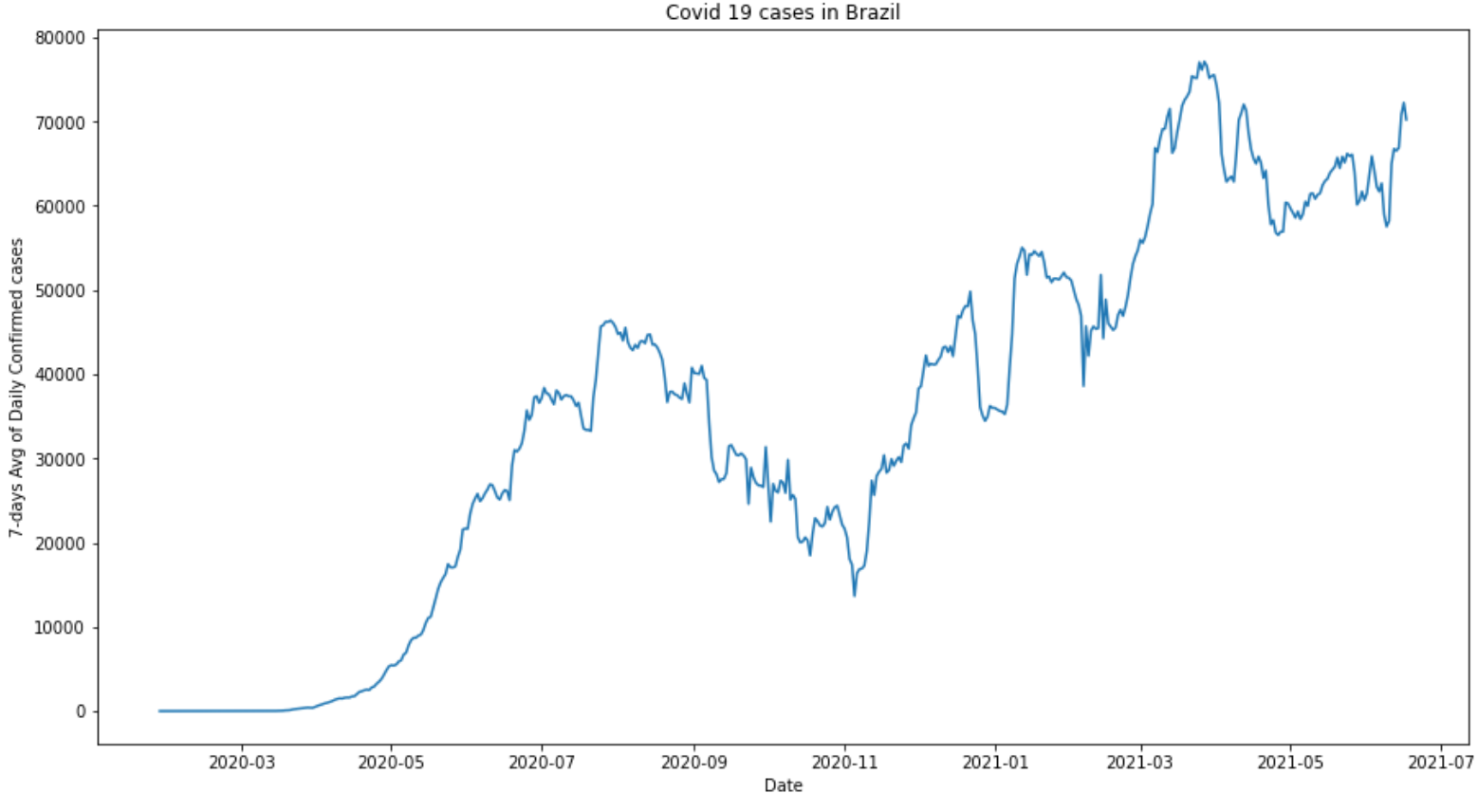
Here, It is clear that wave strength is decreasing unlike worldwide curve.



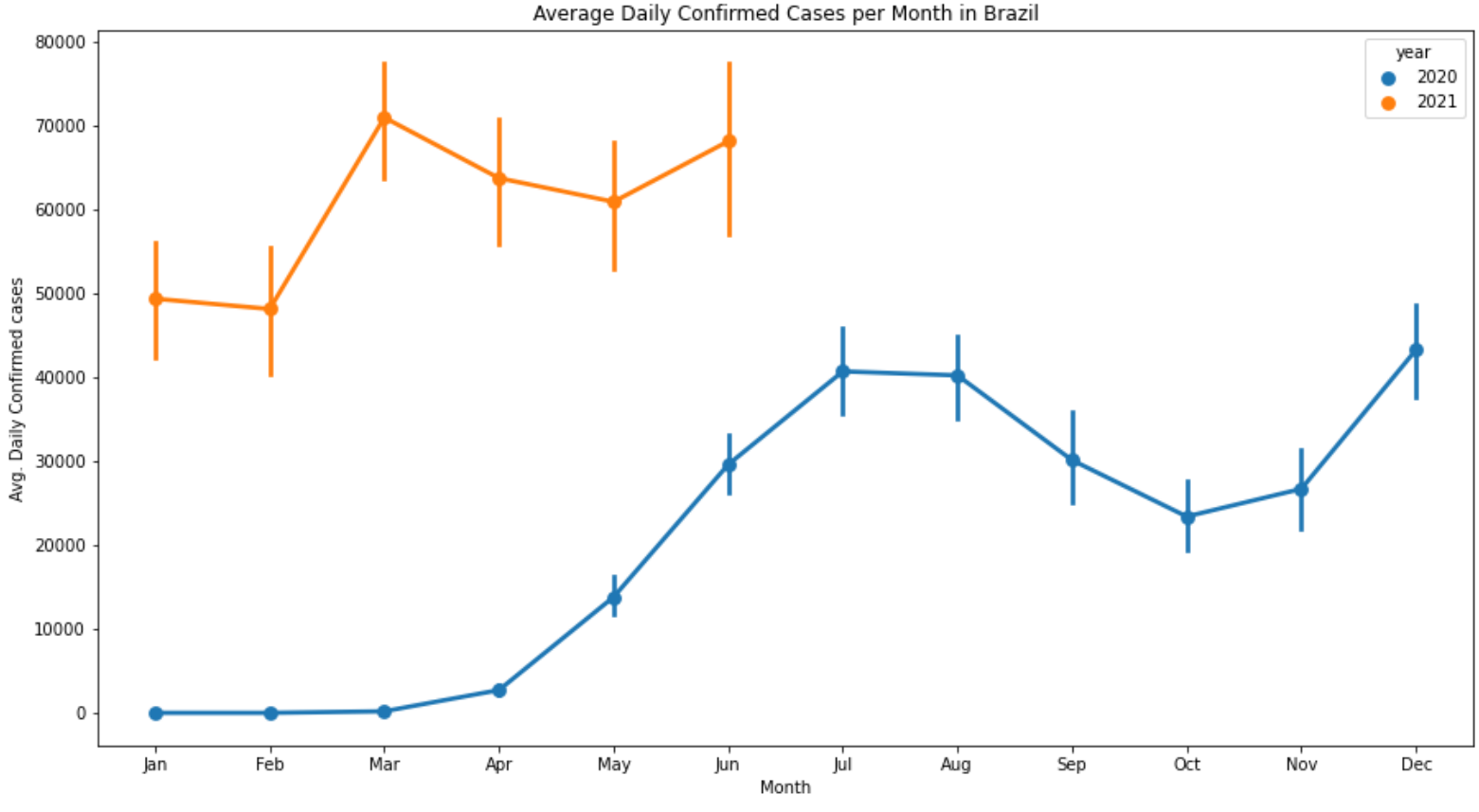
Strongest wave came in Dec 2020 an Jan 2021

And for Brazil

Out[94]:



Here, the wave strength is getting stronger with time.



Strongest wave came in march 2021.

Conclusions

1- It is clear that the occurrence time of waves and their strength are greatly different from one country to another. We definitly can't identify a pattern that all countries follow either in wave strength or time of occurrence. we can only find few details that some countries share.

2- Worldwide data may be misleading if it took alone as different countries has very different conditions.So, we can't generalize any insights from it.

3- Egypt shows a repetitive pattern of covid waves that starts between Feb and March. It happened in 2020 and also in 2021.

Recommendation

Countries with repeatitive wave patterns should make great use of this. They should take proactive moves designing their future strategies about food imports,medcine, etc.. taking into consideration the reasonable likelihood of being hit by a wave on the next pattern.

How Can we rely on these data to make comparisons and extract conclusions?

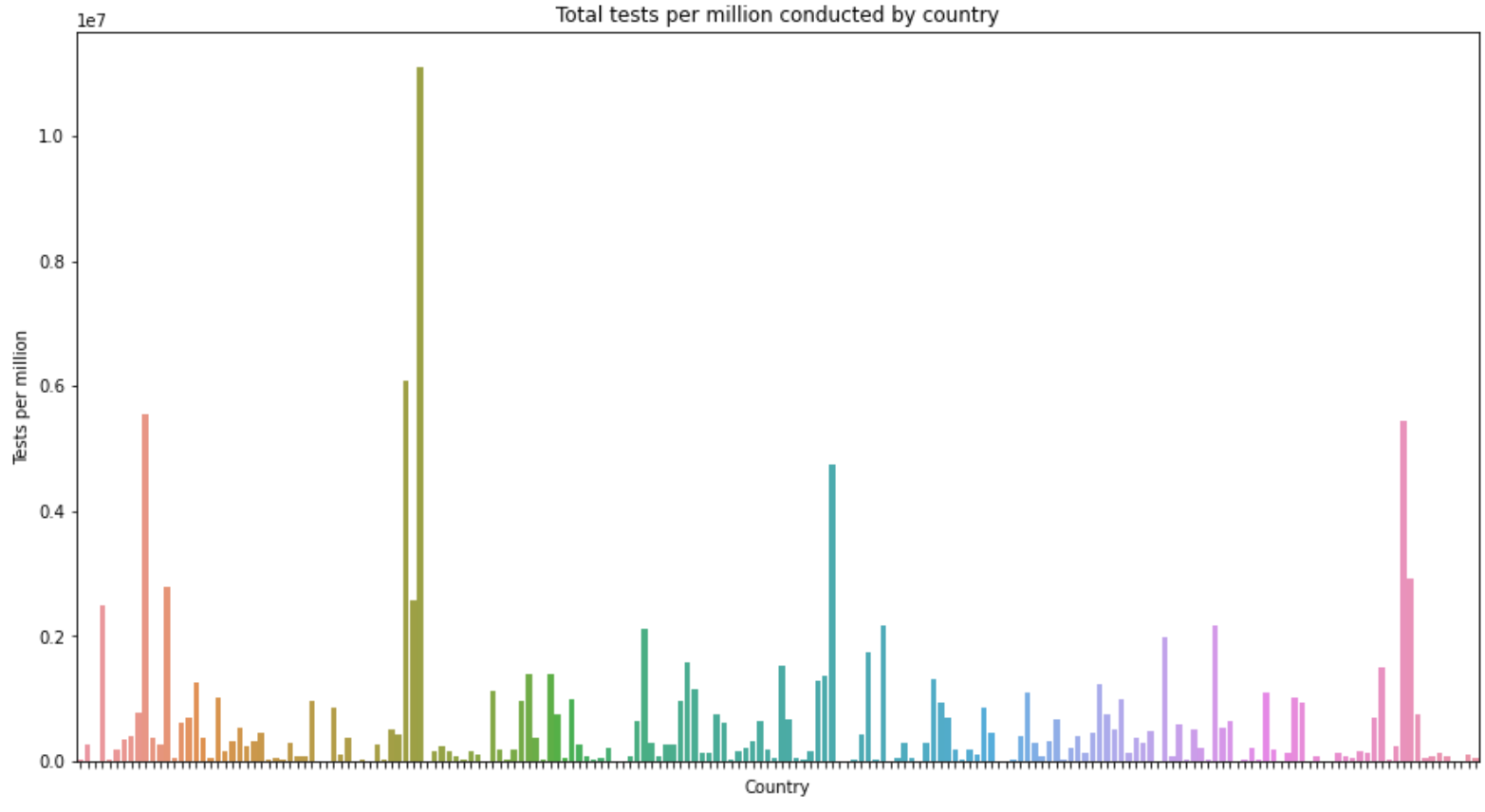
By dividing this question to the following questions

- Do all countries have similar or close values for covid tests per million ratio?
- Does tests per million ratio affect number of cases per million discoverd or cases discoverd per million ratio is a constant ratio that isn't affected by test rates?

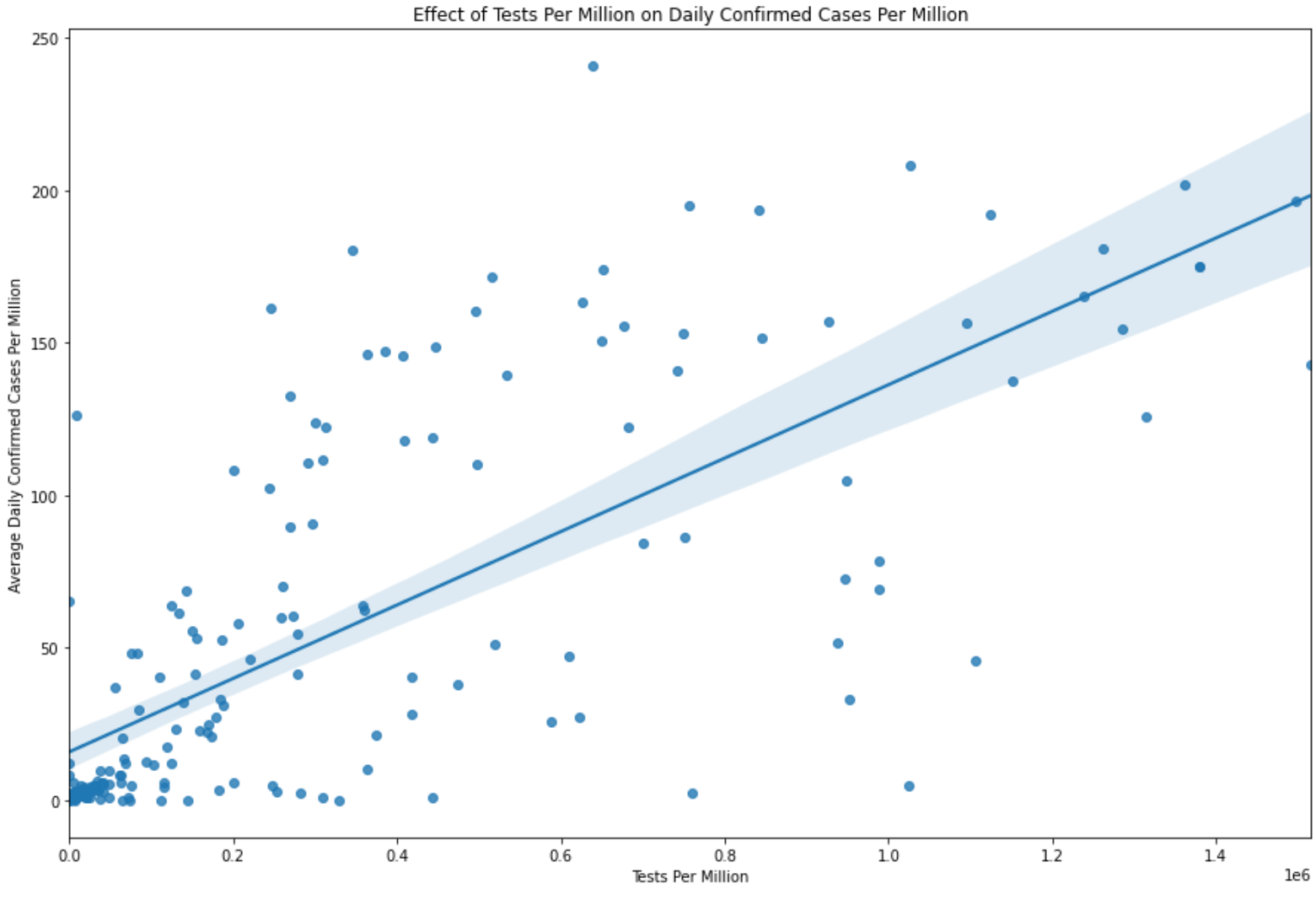
we found that:

1- There is a wide gap of test rates between countries.

2- High test rates result in increase in daily cases per million ratio.



Here, We can see obviously that there is a wide gap of test rates between countries



And here we can see that Tests Per Million and Average Daily Confirmed Cases Per Million are directly proportional. So, high test rates result in more new cases discovery.

Finally, we can see the effect clearly here

Conclusion

It is so clear that high test rates result in increase in daily cases per million ratio.

Recommendation

we need to split our investigation of any parameter effect on daily cases into 2 levels.

1- Effect of the parameter on daily cases for high test rates

2- Effect of the parameter on daily cases for low test rates

So as not to get inaccurate observation due to interference between the parameter effect and high test rate effect. Finally, we can rely on the Effect of the parameter on daily cases for high test rates as the higher test rates, the more accurate data will be.

What is the effect of Population density on daily cases?

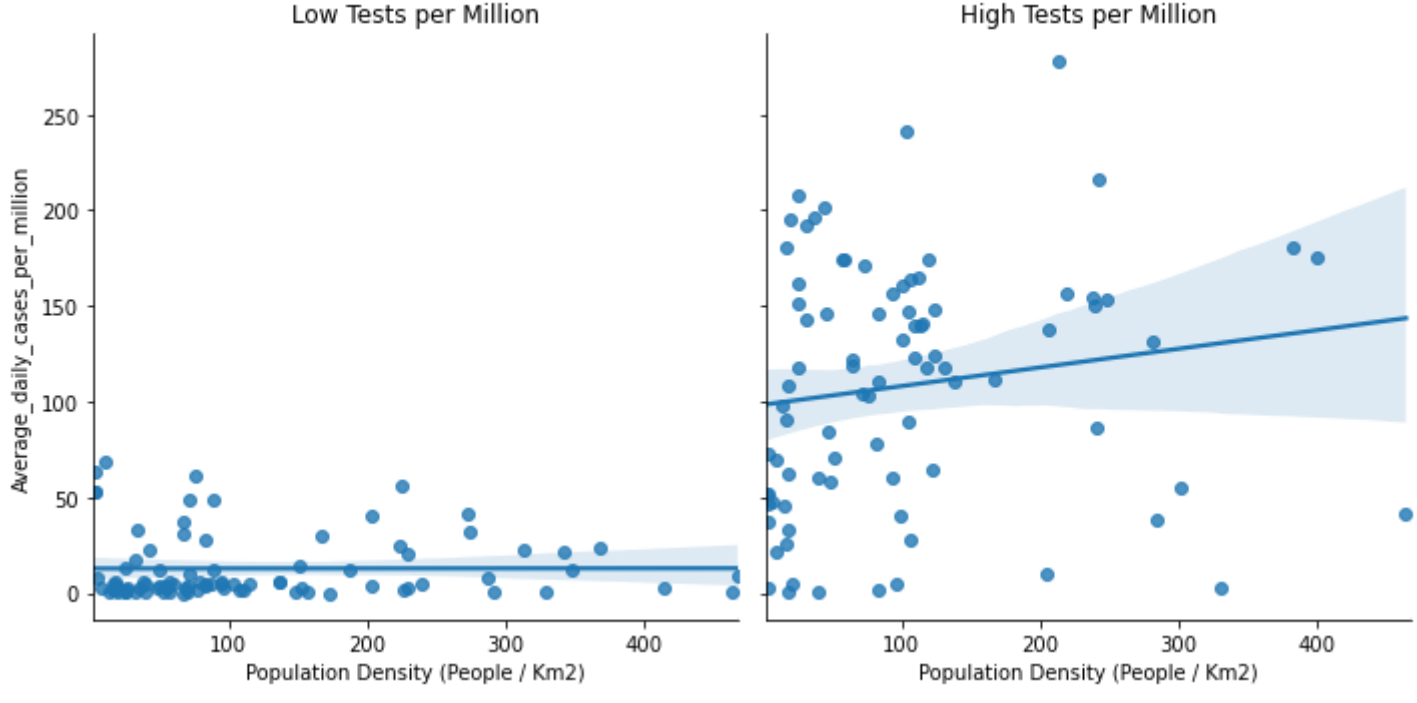
we found that:

1- For low test rates, it seems that Population density has no effect on daily cases.

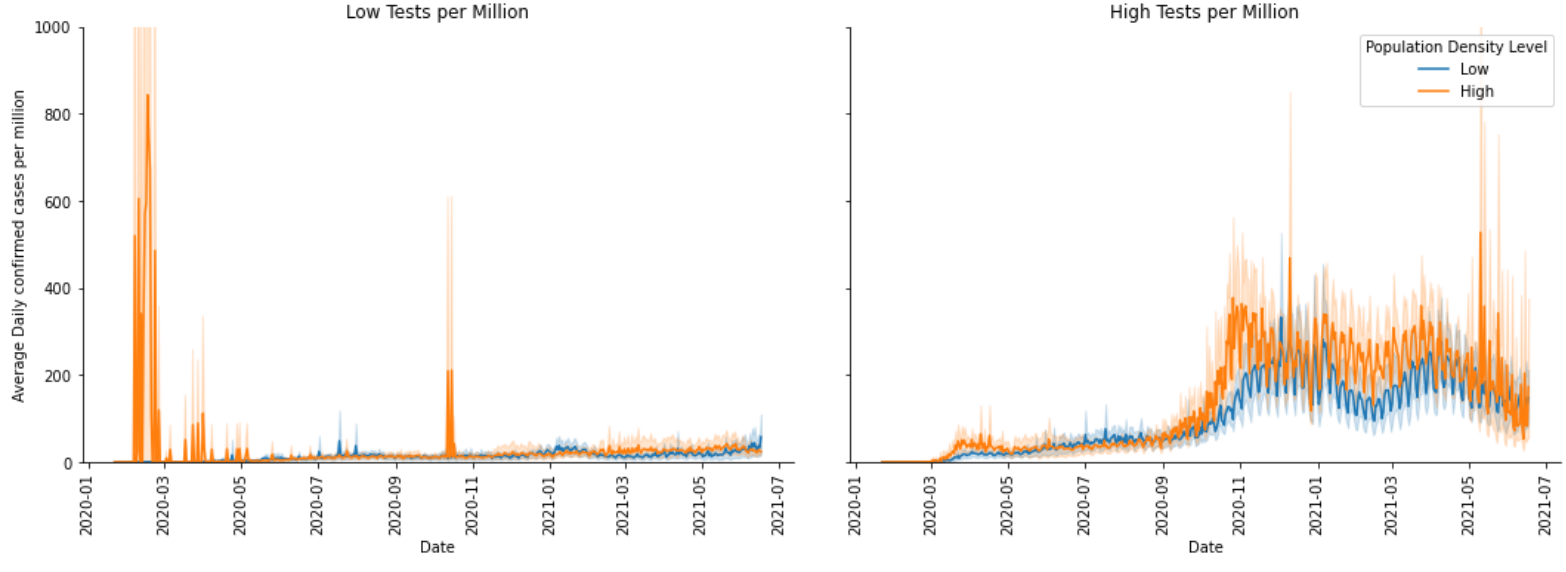
2- For high test rates we can see an increase in average daily cases per million for high population density.

We can see this clearly here

```
<ipython-input-112-45a50bbecf09>:9: UserWarning: FixedFormatter should only be used t
ogether with FixedLocator
  ax.set_xticklabels(ax.get_xticklabels());
```



```
<ipython-input-113-a810d1b0c5db>:7: UserWarning: FixedFormatter should only be used t
ogether with FixedLocator
  ax.set_xticklabels(ax.get_xticklabels(), rotation=90);
```



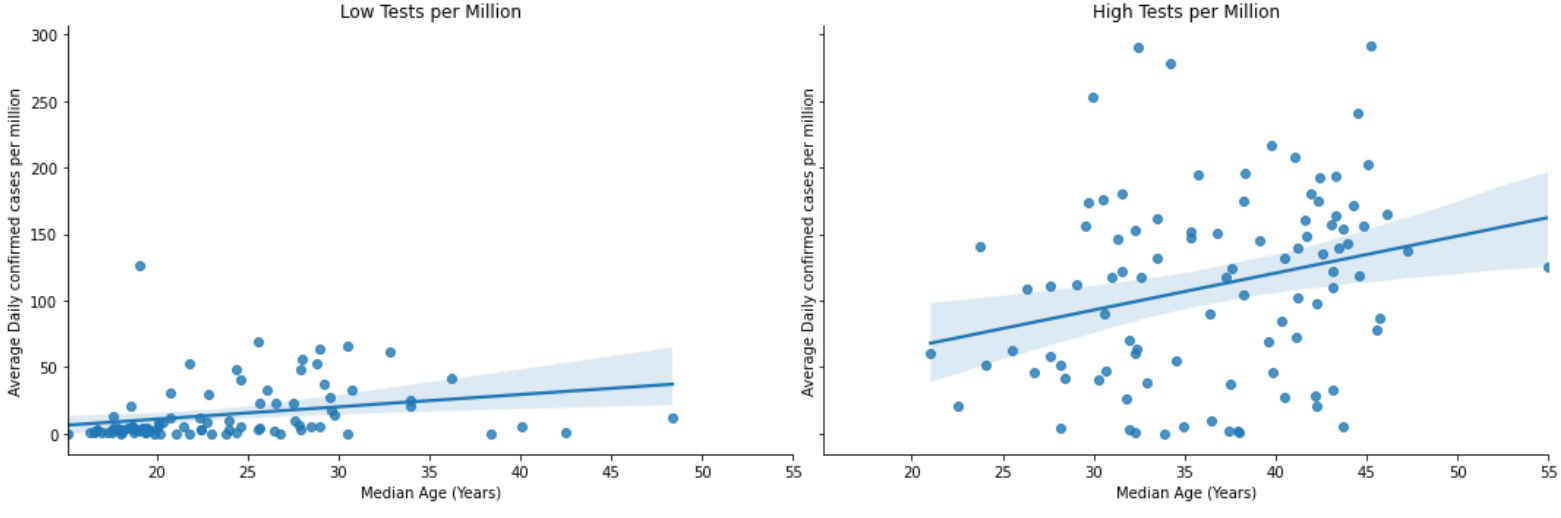
What is the effect of median age on daily cases?

we found that:

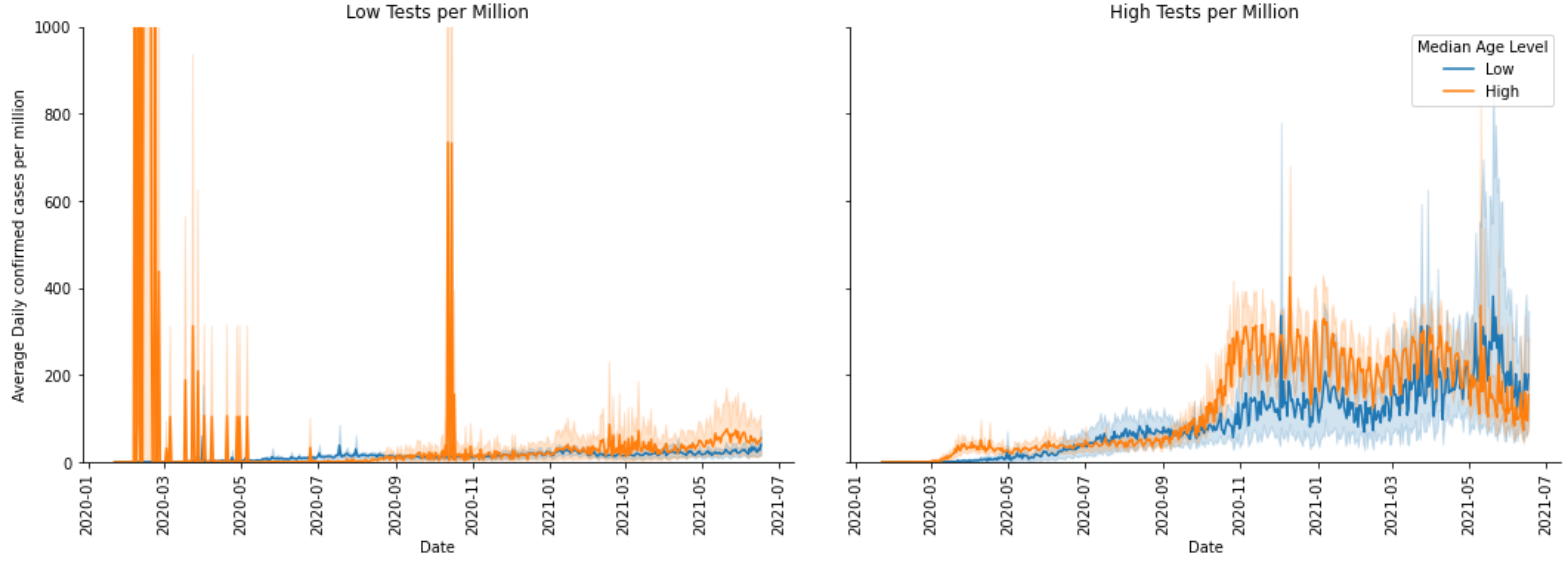
For both low test rates and high test rates we can see an increase in average daily cases per million with increasing median age but it is more clear for high test rates.

We can see this clearly here

```
<ipython-input-122-c846ecd63822>:10: UserWarning: FixedFormatter should only be used
together with FixedLocator
  ax.set_xticklabels(ax.get_xticklabels());
```



```
<ipython-input-123-36914bb646c8>:6: UserWarning: FixedFormatter should only be used t
ogether with FixedLocator
  ax.set_xticklabels(ax.get_xticklabels(), rotation=90)
```



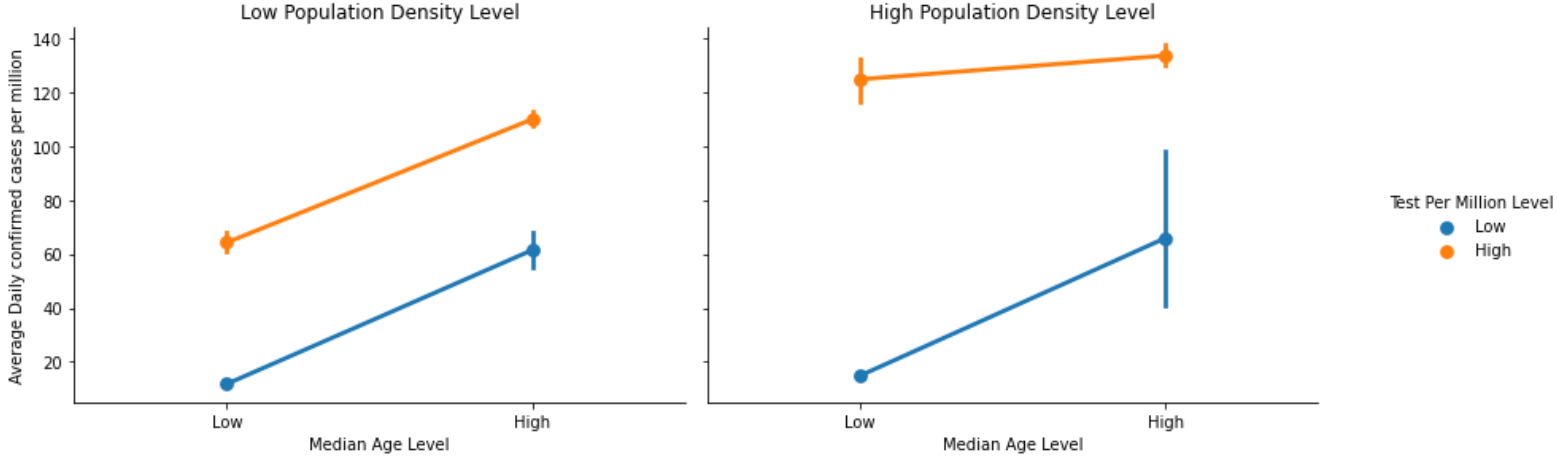
Now, Lets Combine the effect of Median age with population density

What is the effect of Median age on daily cases per million for different population density levels?

we found that:

1- As concluded before high test rates results in more cases discovery than low test rates.

2- Countries with higher median age are experiencing higher number of cases. This can be shown for either high and low test rates but the effect is not so big for high population density countries with high test rates.


Low Population Density Level
High Population Density Level
Average Daily confirmed cases per million
140
120
100
80
60
40
20
Low
High
Median Age Level
Low
High
Median Age Level
Test Per Million Level
Low
High

What is the effect of population density on daily cases per million for different Median age levels?

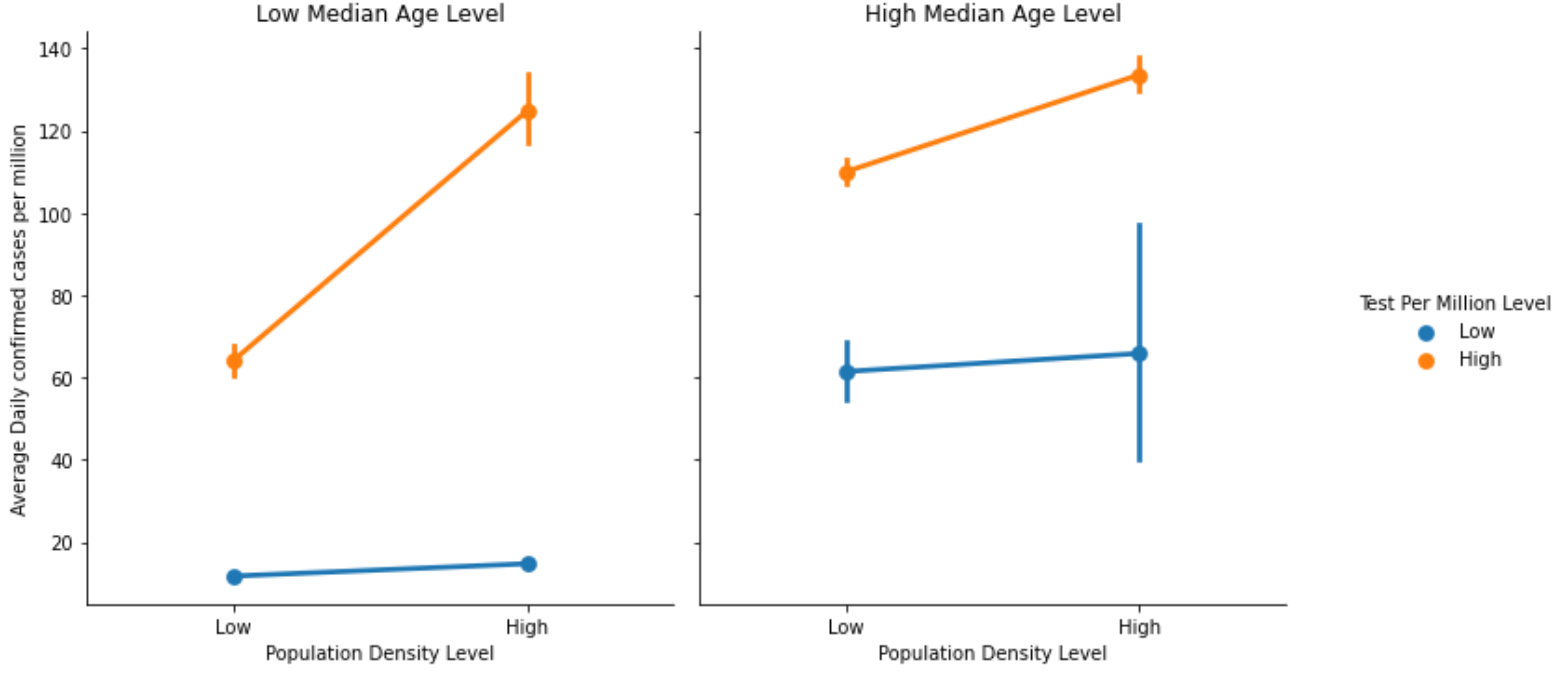
we found that:

1- Again and as concluded before high test rates results in more cases discovery than low test rates.

2- Countries with high test rates are experiencing high number of cases with increasing population density.

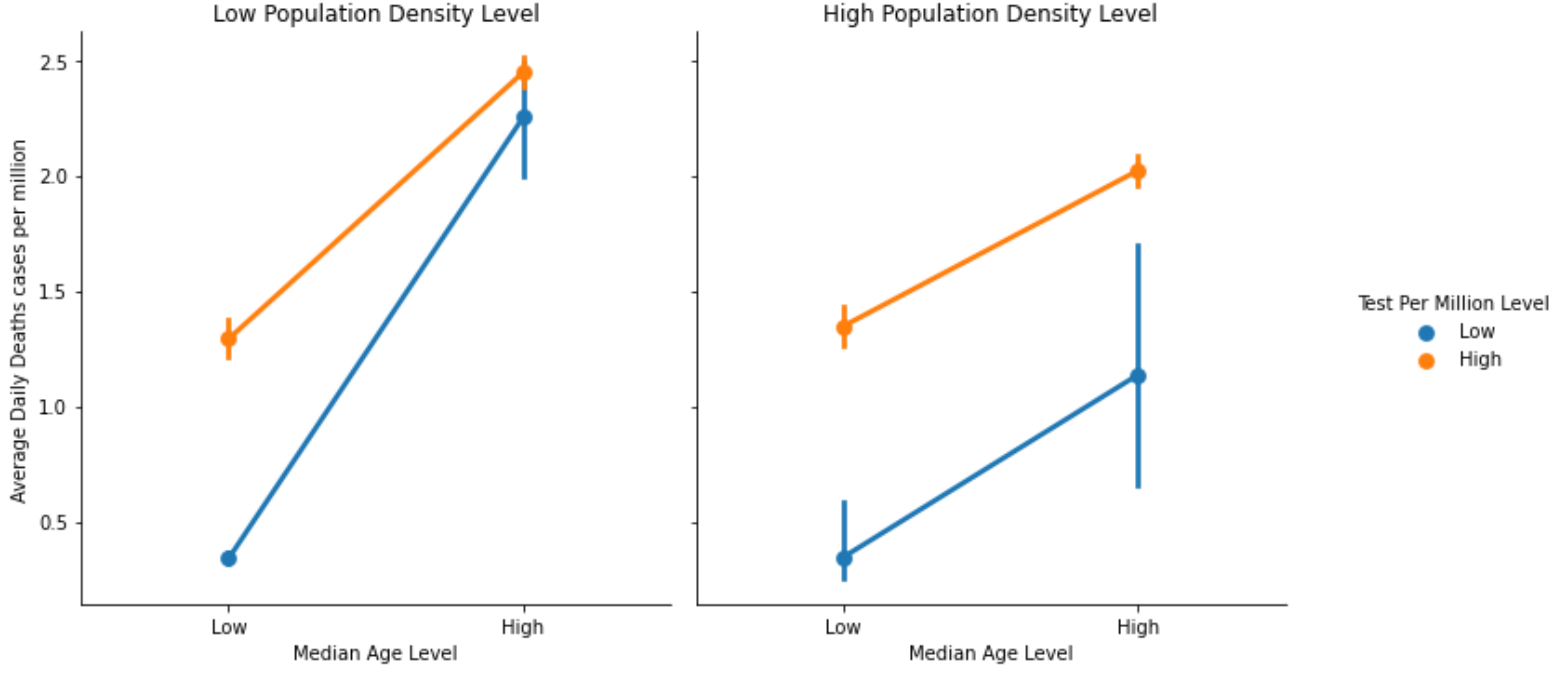
3 - For Countries with low test rates:

a) With low median age level: there is no difference shown. b) With high median age level: there is nearly no difference in average cases but the variance of number of cases for high population density cases is grater.


Low Median Age Level
High Median Age Level
Average Daily confirmed cases per million
140
120
100
80
60
40
20
Low
High
Population Density Level
Low
High
Population Density Level
Test Per Million Level
Low
High

What is the effect of Poulation density and Median age on daily deaths per million?

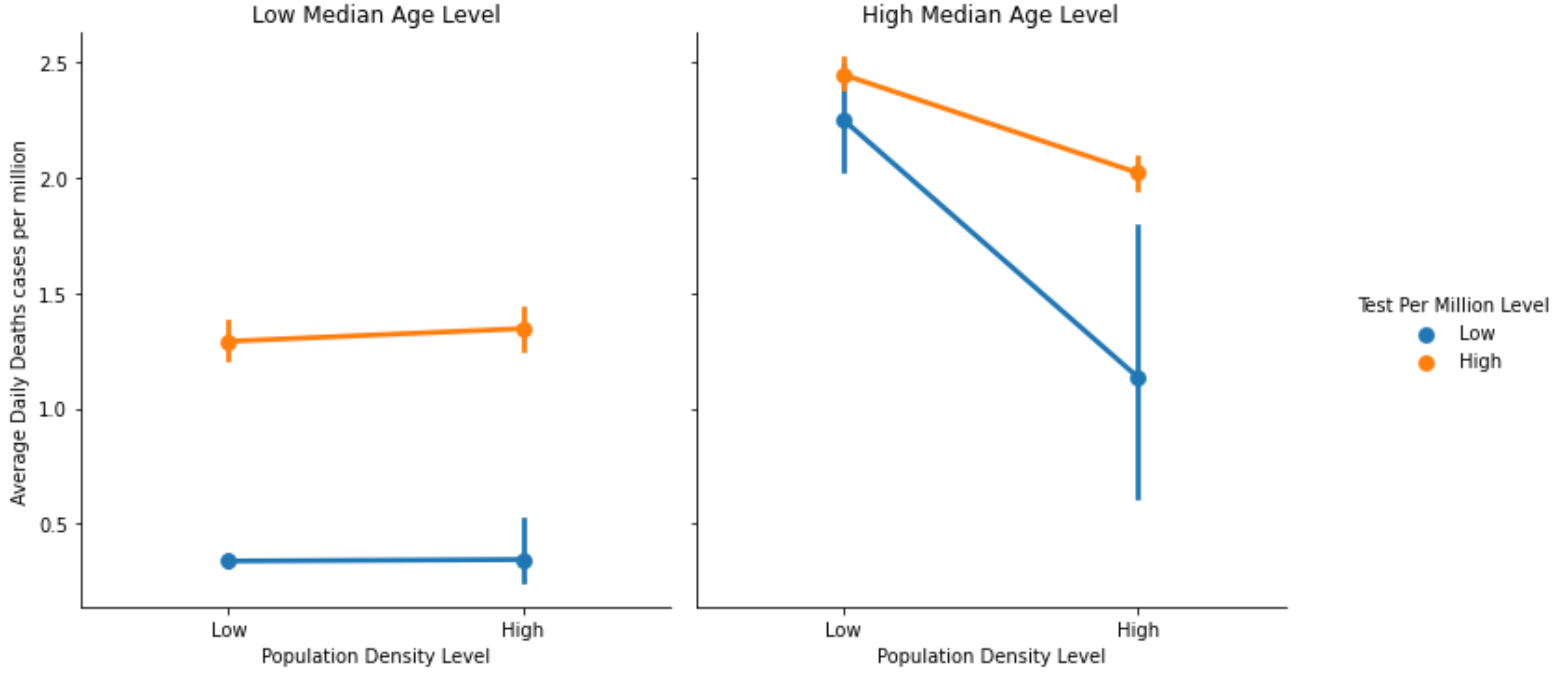
1 - Median age



Observation

1- Countries with high Median age experience higher deaths rates under all conditions.

2- Population Density

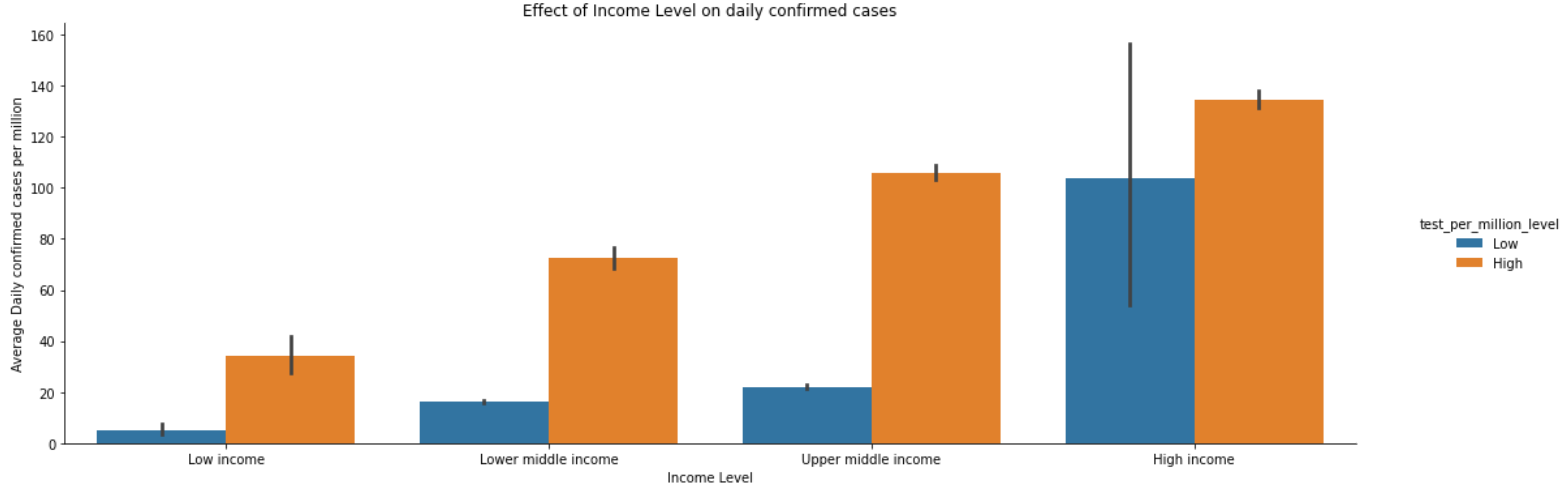


Observation

1 - Countries with low Median age have about the same deaths per million ratio for both population density level.

2 - For Countries with high Median we can see that deaths per million are inversly proportional to population density for both test rates.

What is the effect of Income Level on daily cases?



Observation

1 - Under all conditions, we can see the higher the income the higher the cases per million.

What is the effect of Continent on daily cases?

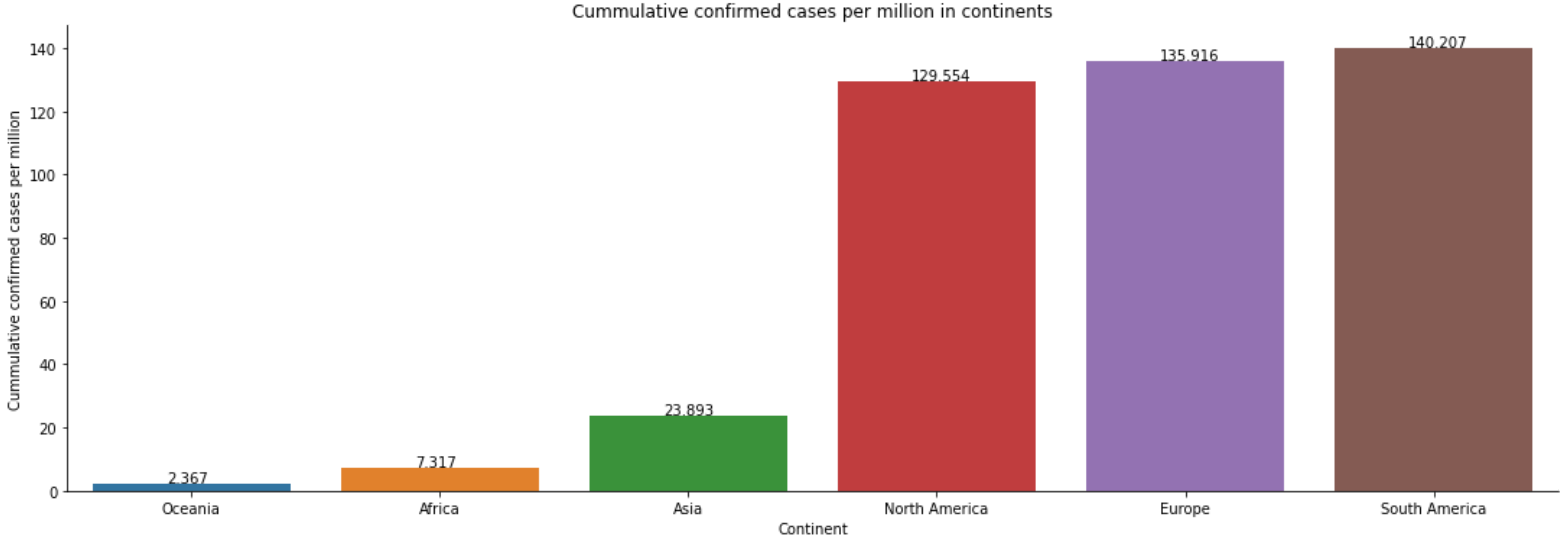
We found that

1 - Even Africa has lower cases per million and deaths per million but has the second highest deaths to confirmed ratio. This can reflect the heath care system of Africa.

2- South America is the most suffering continent with highest cummulative cases, deaths and deaths to confirmed ratio.

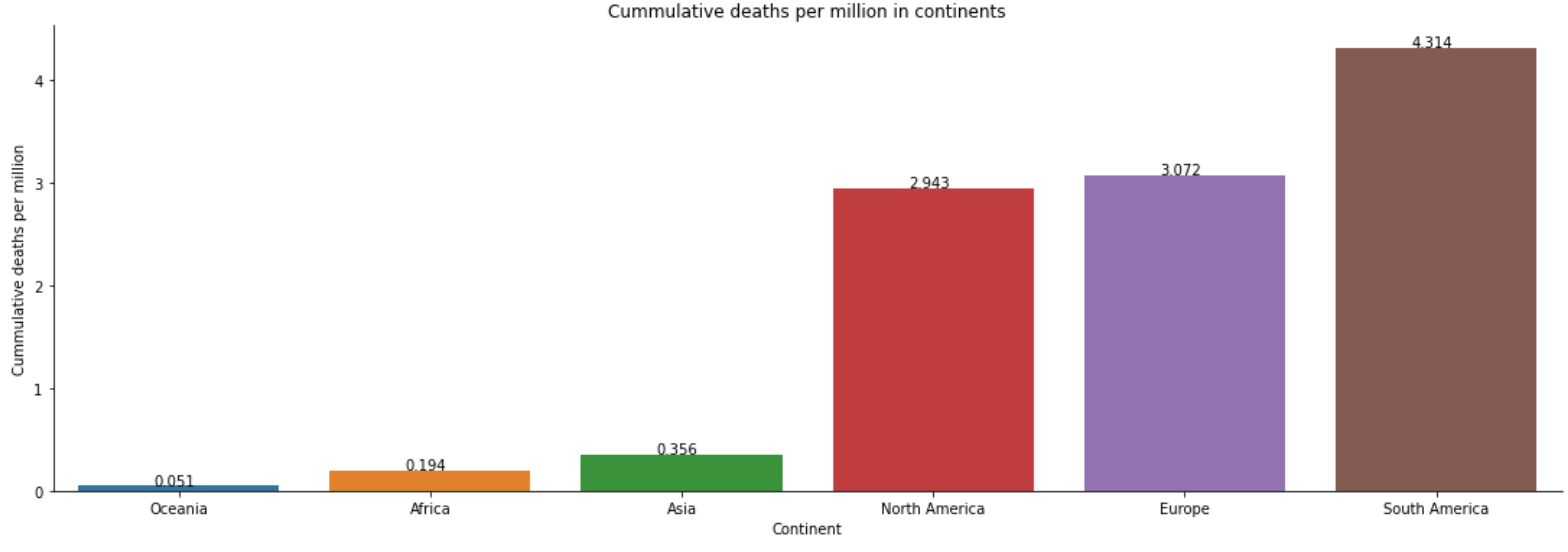
3- Europe and North America reflect better heath care systems comparing their death to confirmed ratios with their very high cases per million.

1. Cummulative confirmed cases per million



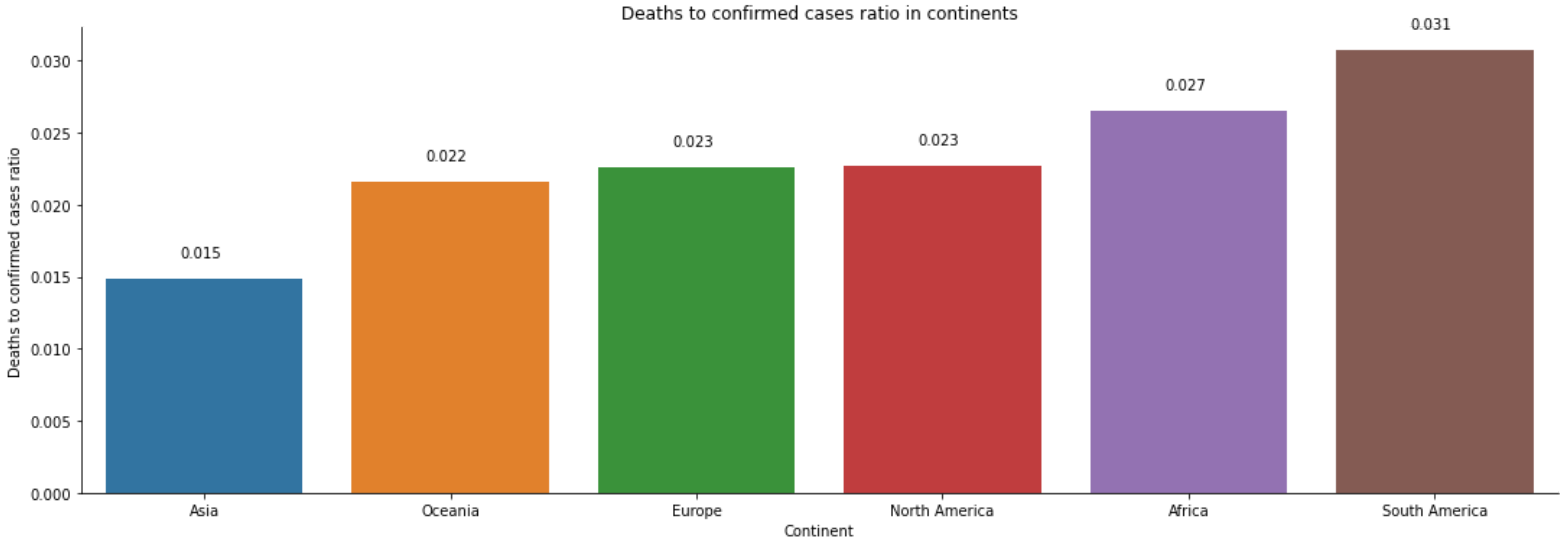
South America, Eroupe and North America has highest cummulative cases per million.

2. Cummulative deaths per million



South America, Eroupe and North America has highest Cummulative deaths per million.

3. Deaths to confirmed cases ratio



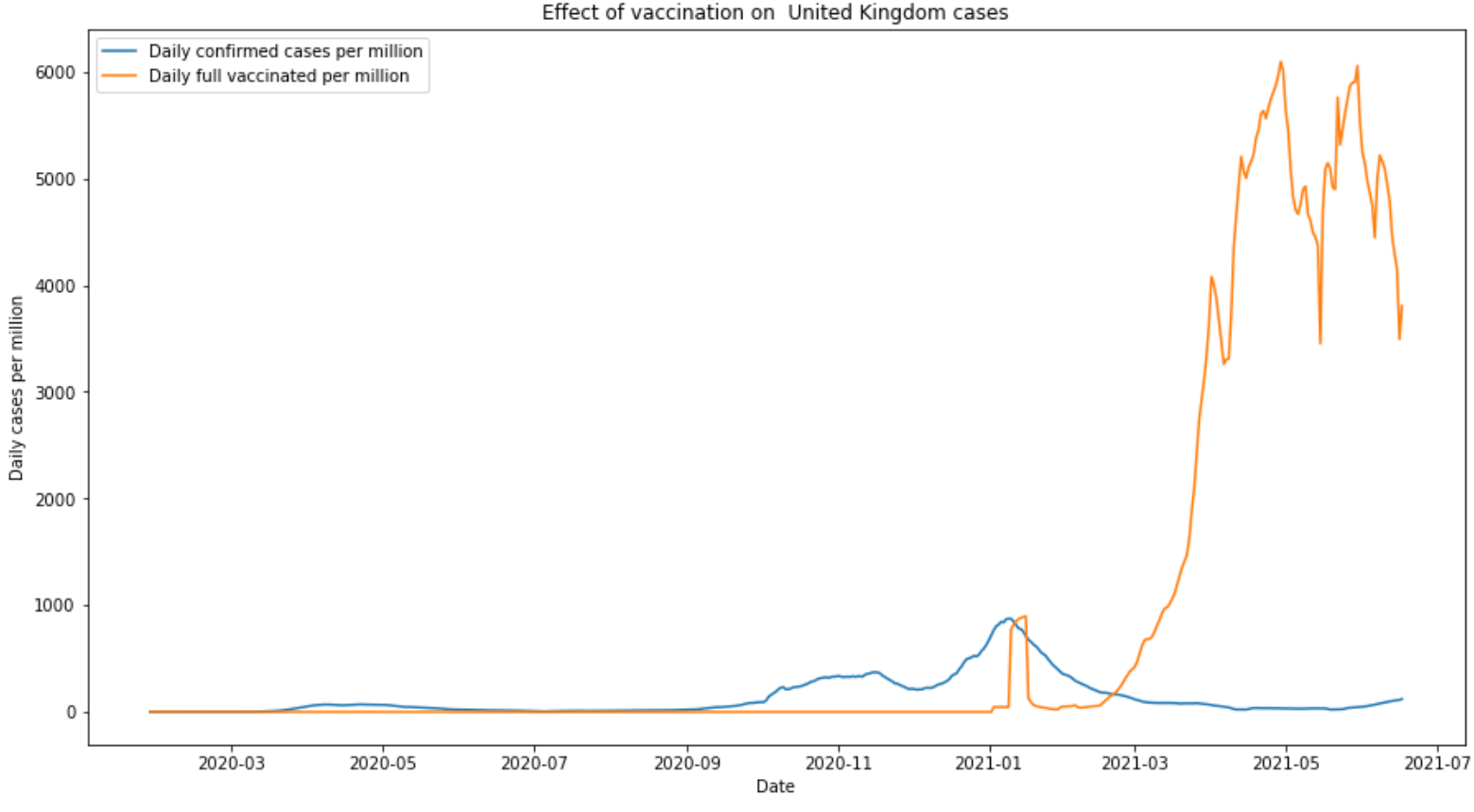
South America, Africa has highest Cummulative deaths per million.

What is the effect of Vaccinations on daily cases?

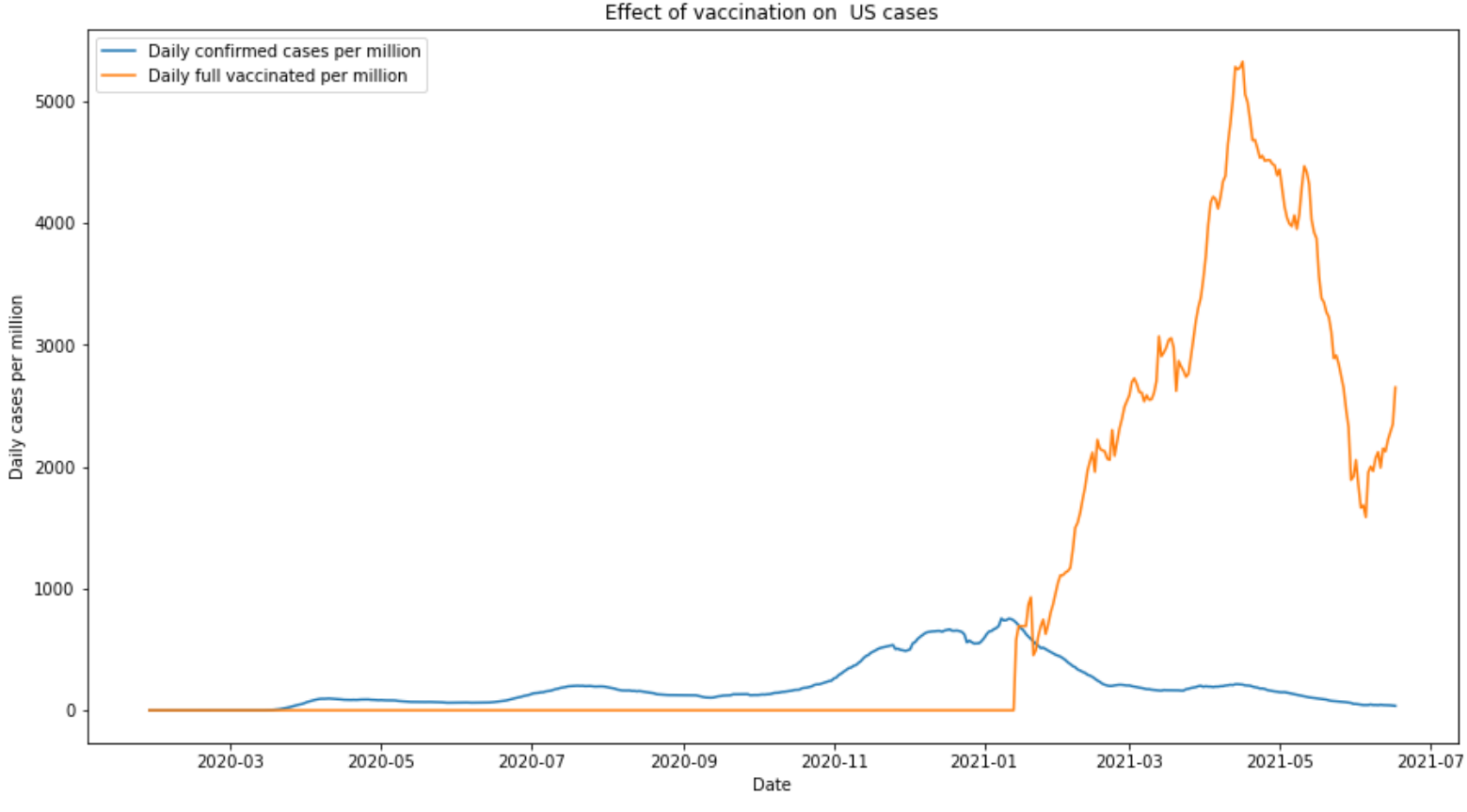
We found that

FOr US and Uk we can notice a high decrease in daily cases from the start of vaccination.

United Kingdom



United States



Conclusions

Timeseries Analysis

1- It is clear that the occurrence time of waves and their strength are greatly different from one country to another. We definitly can't identify a pattern that all countries follow either in wave strength or time of occurrence. we can only find few details that some countries share.

2- Worldwide data may be misleading if it took alone as different countries has very different conditions.So, we can't generalize any insights from it.

3- The strength of waves worldwide is continously incrasing but in Egypt,US and Italy the third wave is lower than the second one. This can explain point 2.

4- Egypt, United Kingdom and US shares a wave in months [Dec,Jan,Feb].

5- Egypt and Italy share a repetitive pattern of covid waves that starts between Feb and March. It happened in 2020 and also in 2021.

6- United Kingdom and India don't have repeatitive patterns in waves till now.

Effect of Test rates on daily cases

7- There is a wide gap of test rates between countries.

8- High test rates result in increase in daily cases per million ratio. This High test rates will be the more likeliy to represent the actual case.

Effect of Population density on daily cases

9- For low test rates it seems to have no effect.

10 - For high test rates we can see an increase in average daily cases per million for high population density.

Effect of median age on daily cases

11- For both low test rates and high test rates we can see an increase in average daily cases per million for high median age counties but it is more clear for high test rates.

12- Countries with high Median age has higher averager deaths under all conditions.

13- Countries with low Median age have about the same deaths per million ratio for both population density level.

14- For Countries with high Median we can see that deaths per million are inversly proportional to population density for both test rates.

Effect of Income level on daily cases

15- Under all conditions, we can see the higher the income the higher the cases per million.

Effect of continent on daily cases

16- we can see that even Africa has lower cases per million and deaths per million but has the second highest deaths to confirmed ratio. This can reflect quality of the heath care system of Africa.

17- South America is the most suffering continent.

18- Europe and North America reflect better heath care systems with their low death to confirmed ratios compared to their very high cases per million.

Effect of vaccination on daily cases

19 - for US and Uk we can notice a high decrease in daily cases from the start of vaccination.

Recommendations

1- Countries with repeatitive wave patterns should make great use of this. They should take proactive moves designing their future strategies about food imports,medcine, etc.. taking into consideration the reasonable likelihood of being hit by a wave on the next pattern.

2- Study should be conducted to relate why high income countries has the highest cases per million, is that have any relation with the food happits, daily routine?