

COVID-19 Data Analysis

I. Introduction

COVID-19 has been a terrible disease for many countries. This disease has kill at least nine hundred thousands people around the world with no ending in sight. Day by day, the case are increasing. Some countries have so many cases more than any other country for example nowadays United States, India, and Brazil are the top three countries with the most cases. There are countries though that are not really affected or the cases is pretty small compared to the top three, such countries are South Korea and New Zealand. The number of cases and the severity of how COVID-19 affects countries will give an idea about how a healthcare should prepare. Therefore identifying factors that may contribute to the severity of COVID-19 cases are important.

1.2 Problem

As severity varied form each country, data such as GDP per capita, happiness index are indetified and correlated to the number of cases and the number of deaths in the country. This parameter will hopefully gives an idea why some country have more cases than any other country.

1.3 Interest

Public health academics and stakeholder may be interested in knowing these factor to further prepare about the increasing number each day and prepare for pandemic that may be happening in the near future.

2. Data

Data is obtained from kaggle for the world happiness report and covid-19 data is obtained from John Hopkins. The world happiness report include country data, GDP per capita and et cetera. In the COVID-19 data, there are list of countries and case per day.

Firstly country and region is set as index and I interested to analyze all the parameter included such as happiness score, happiness ranking, gdp per capita, perception of corruption, social support, freedom to make life choices, and generosity. For the second data set, I will take the total case in the last day available which is April 30th 2020. Both for the case and death caused by COVID-19

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	4/21/20	4/22/20	4/23/20	4/24/20	4/25/20	4/26/20	4/27/20	4/28/20	4/29/20	4/30/20
0	NaN	Afghanistan	33.0000	65.0000	0	0	0	0	0	0	...	1092	1176	1279	1351	1463	1531	1703	1828	1939	2171
1	NaN	Albania	41.1533	20.1683	0	0	0	0	0	0	...	609	634	663	678	712	726	736	750	766	773
2	NaN	Algeria	28.0339	1.6596	0	0	0	0	0	0	...	2811	2910	3007	3127	3256	3382	3517	3649	3848	4006
3	NaN	Andorra	42.5063	1.5218	0	0	0	0	0	0	...	717	723	723	731	738	738	743	743	743	745
4	NaN	Angola	-11.2027	17.8739	0	0	0	0	0	0	...	24	25	25	25	25	26	27	27	27	27
5	NaN	Antigua and Barbuda	17.0608	-61.7964	0	0	0	0	0	0	...	23	24	24	24	24	24	24	24	24	24
6	NaN	Argentina	-38.4161	-63.6167	0	0	0	0	0	0	...	3031	3144	3435	3607	3780	3892	4003	4127	4285	4428
7	NaN	Armenia	40.0691	45.0382	0	0	0	0	0	0	...	1401	1473	1523	1596	1677	1746	1808	1867	1932	2066
8	Australian Capital Territory	Australia	-35.4735	149.0124	0	0	0	0	0	0	...	104	104	104	105	106	106	106	106	106	106
9	New South Wales	Australia	-33.8688	151.2093	0	0	0	0	3	4	...	2969	2971	2976	2982	2994	3002	3004	3016	3016	3025
10	Northern Territory	Australia	-12.4634	130.8456	0	0	0	0	0	0	...	28	28	28	28	28	28	28	28	28	28
11	Queensland	Australia	-28.0167	153.4000	0	0	0	0	0	0	...	1024	1024	1026	1026	1026	1030	1033	1034	1033	1033
12	South Australia	Australia	-34.9285	138.6007	0	0	0	0	0	0	...	437	438	438	438	438	438	438	438	438	438
13	Tasmania	Australia	-41.4545	145.9707	0	0	0	0	0	0	...	201	205	207	207	207	212	214	218	219	221
14	Victoria	Australia	-37.8136	144.9631	0	0	0	0	1	1	...	1336	1336	1337	1343	1346	1349	1349	1354	1361	1364
15	Western Australia	Australia	-31.9505	115.8605	0	0	0	0	0	0	...	546	546	546	548	549	549	549	550	551	551
16	NaN	Austria	47.5162	14.5501	0	0	0	0	0	0	...	14873	14925	15002	15071	15148	15225	15274	15357	15402	15452
17	NaN	Azerbaijan	40.1431	47.5769	0	0	0	0	0	0	...	1480	1518	1548	1592	1617	1645	1678	1717	1766	1804
18	NaN	Bahamas	25.0343	-77.3963	0	0	0	0	0	0	...	65	65	72	73	78	80	80	80	80	81
19	NaN	Bahrain	26.0275	50.5500	0	0	0	0	0	0	...	1973	2027	2217	2518	2588	2647	2723	2811	2921	3040

20 rows × 104 columns

Total cases	
Country/Region	
Afghanistan	2171
Albania	773
Algeria	4006
Andorra	745
Angola	27
Antigua and Barbuda	24
Argentina	4428
Armenia	2066
Australia	6766
Austria	15452
Azerbaijan	1804
Bahamas	81
Bahrain	3040
Bangladesh	7667
Barbados	81
Belarus	14027
Belgium	48519
Belize	18
Benin	64
Bhutan	7

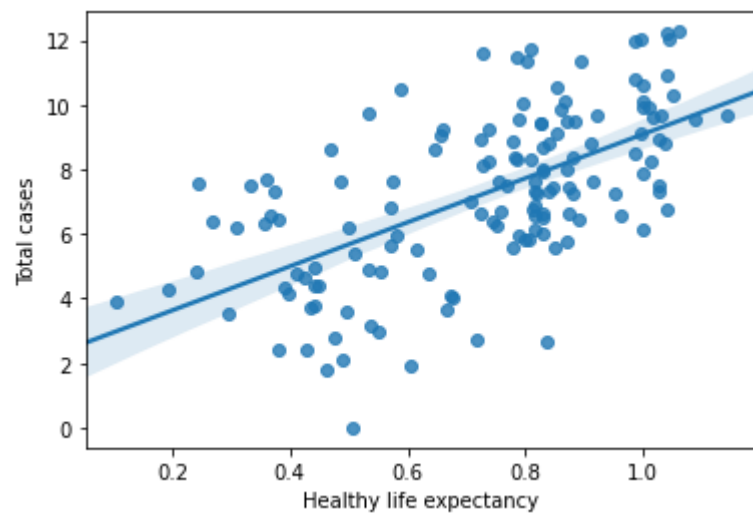
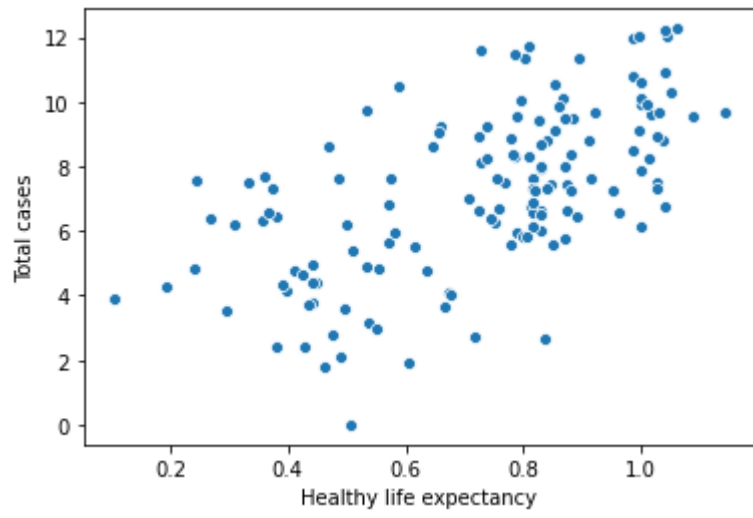
This 3 dataset will be joined and will be analyzed.

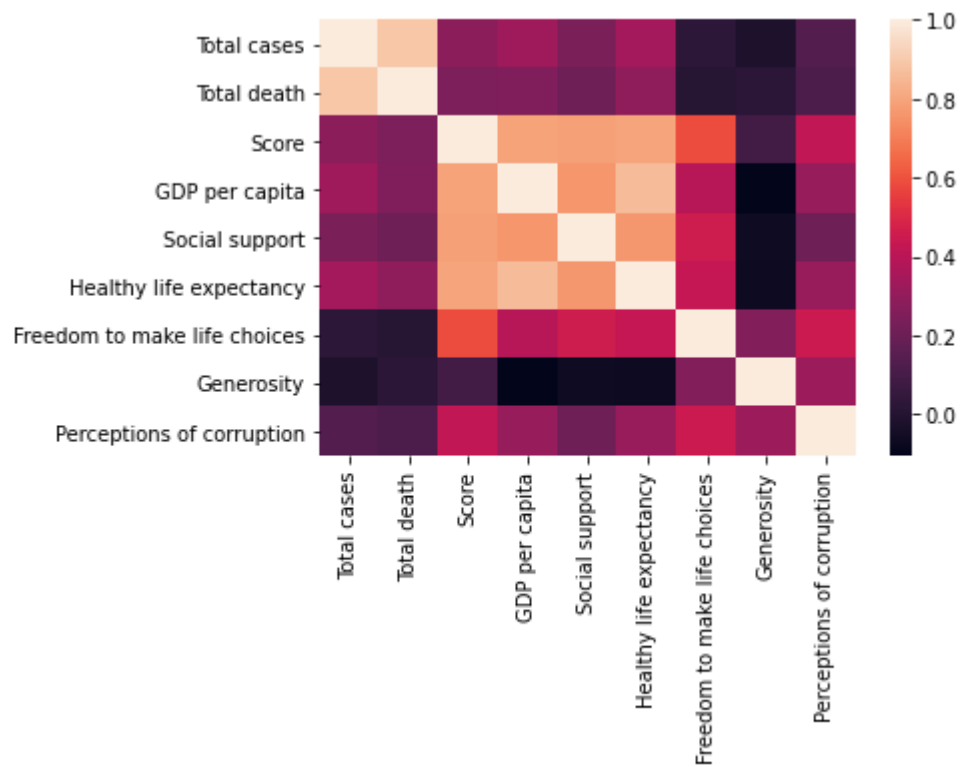
3. Result

Pearson correlation indicates that some are not an important factor for COVID-19 total case and total death. Total case correlates strongly with total death but this is a no question as with more cases, more deaths are expected.

Variable that are pretty significant that may be contributes to total cases and total deaths are GDP per capita, healthy life expectancy, and happiness score.

This may be because the country is rich and so they can afford more test and have better data collection than developing country.





Recommendation

I think that there will be a need of better data collection as pretty controversial when the richer the country is the more of the total cases.