Evaluation of the relationship between COVID-19 spread and various Global Parameters

Akshat Gupta
MNNIT Allahabad

Abstract **- COVID-19 is a disease caused by the novel coronavirus and was first reported in the Wuhan province of China. It is a highly contagious disease and has caused over 1.6 million deaths globally. In this paper, the infections and deaths caused by this disease (up to the 7th of December 2020) have been evaluated and compared to various parameters present in the Human freedom index and world happiness report using Python and its various libraries. This paper attempts to analyze and visualize the relationships between the spread and effect of corona virus in different countries of the world.**

Index Terms **- COVID-19, Human Freedom Index, Python, World Happiness Report.**

I.INTRODUCTION

The novel coronavirus is a type of RNA virus that is a part of the sub-family Orthocoronavirinae, family Coronaviridae. It causes a respiratory tract infection in humans that has been termed as COVID-19. The symptoms of COVID-19 resemble the common cold and its severity can vary from mild to lethal depending on the person and their immunity. COVID-19 is closely like SARS and MERS, the latter being more lethal. The first case of due to this virus was first reported in the Wuhan province of China.

[1] The Human Freedom Index is worldwide ranking of civil, economic, and personal freedom. The Human Freedom Index measures various aspects of freedoms such as the social, economic, political freedom and it evaluates the degree to which people are free to enjoy the major freedoms often referred to as civil liberties—freedom of speech, religion, movement, etc. in the countries in the survey. In addition, it includes indicators on rule of law, crime and violence, freedom of movement, and legal discrimination against same-sex relationships. We also include nine variables pertaining to women-specific freedoms that are found in various categories of the index. [2] The World Happiness Report is also a worldwide ranking of national happiness based on various factors that determine the quality of life. The Happiness Score is explained by the following factors: GDP per capita, Healthy Life Expectancy, Freedom to make life choices etc.

This paper utilizes four datasets for analysis. Two out of four datasets viz. infections and deaths [3] (hereby addressed as dataset 1 and dataset 2 respectively) due to coronavirus have been utilized for analysis. These are time-series datasets published and updated by the WHO daily. The 3rd and 4th datasets are [1] the Human freedom Index and [4] the World happiness report (hereby addressed as dataset 3 and dataset 4).

II.METHODOLOGY

To derive relationships between the distinct variables in datasets 1 through 4, correlation matrices have been used. A correlation matrix is as the name suggests is a matrix that presents the correlation between all the variables. Based on this the fit of data is described using scatter plots which have been drawn using the matplotlib.pyplot library in python.

Pandas (library) has been used to create data frames of the various datasets. All datasets have first been filtered and cleaned based on relevance and requirement for analysis. Each dataset has been aggregated, grouped, and indexed based on country. Then the maximum death rate and infection rate was calculated for each country and presented as a data frame using the line of code:

```
countries = list(corona_dataset_aggregated.index)
max_infection_rates = []

for country in countries:
max_infection_rates.append(corona_dataset_aggregated.loc[country].diff().max())
corona_dataset_aggregated['max infection rate'] = max_infection_rates
```

This code is specific for the [3] dataset 1 and calculates the maximum number of infections a country reports in a single day. It produces a list which is converted to a data frame that appears as follows:

```
In [35]: corona_data.head()
```

Out[35]:

Country/Region	max infection rate
Afghanistan	915.0
Albania	846.0
Algeria	1133.0
Andorra	299.0
Angola	355.0

```
In [37]: death_data.head()
```

Out[37]:

Country/Region	max death rate
Afghanistan	42.0
Albania	19.0
Algeria	30.0
Andorra	6.0
Angola	12.0

Table 1

[1][4] Datasets 3 and 4 have also been filtered and converted to data frames. Now that all 4 datasets are filtered, to compare the maximum infection rates and maximum death rates variation based on various parameters present in dataset 3 and dataset 4, they have been combined using inner joins.

After successful joining of the data, we now have 4 final data frames:

1. 1.Human freedom index – Max Infection Rate (represented by data.corr())
2. Human Freedom index – Max Death Rate (represented by data_deaths.corr())
3. World Happiness report- Max Infection Rate (representedby happiness_infection_data.corr())
4. World Happiness report -Max Death rate (represented by happiness_death_data.corr())
5. Using the. corr() function in python we create correlation matrices.

All 4 correlation matrices are shown below.

Now we create plots to graphically represent the correlations that are given in the correlation matrix using the matplotlib.pyplot library. The following variables were chosen for comparison in the world happiness report against maximum infection and death rates:

1. GDP per capita
2. Social Support
3. Healthy life expectancy
4. Freedom to make life choices
5. Generosity

The following variables were chosen for comparison in the Human Freedom Index against maximum infection and death rates:

1. Human Freedom Score(hf_score)
2. Political Freedom Score(pf_score)
3. Economic Freedom Score(ef_score)

```
In [27]: happiness_infection_data.corr()
```

Out[27]:

	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	max infection rate
GDP per capita	1.000000	0.754906	0.835462	0.379079	-0.079662	0.253117
Social support	0.754906	1.000000	0.719009	0.447333	-0.048126	0.182747
Healthy life expectancy	0.835462	0.719009	1.000000	0.390395	-0.029511	0.261623
Freedom to make life choices	0.379079	0.447333	0.390395	1.000000	0.269742	0.060731
Generosity	-0.079662	-0.048126	-0.029511	0.269742	1.000000	-0.125275
max infection rate	0.253117	0.182747	0.261623	0.060731	-0.125275	1.000000

```
In [38]: happiness_death_data.corr()
```

Out[38]:

	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	max death rate
GDP per capita	1.000000	0.754906	0.835462	0.379079	-0.079662	0.147376
Social support	0.754906	1.000000	0.719009	0.447333	-0.048126	0.145110
Healthy life expectancy	0.835462	0.719009	1.000000	0.390395	-0.029511	0.205851
Freedom to make life choices	0.379079	0.447333	0.390395	1.000000	0.269742	0.116149
Generosity	-0.079662	-0.048126	-0.029511	0.269742	1.000000	-0.191924
max death rate	0.147376	0.145110	0.205851	0.116149	-0.191924	1.000000

Table 2

```
In [22]: data_deaths.corr()
Out[22]:
```

	hf_score	hf_rank	hf_quartile	pf_rol	pf_ss_homicide	pf_ss_disappearances_violent	pf_ss_disappearances_fatalities
hf_score	1.000000	-0.976128	-0.928518	0.837371	0.233121	0.454195	0.466837
hf_rank	-0.976128	1.000000	0.968113	-0.818322	-0.197345	-0.374685	-0.383587
hf_quartile	-0.928518	0.968113	1.000000	-0.763528	-0.194366	-0.318923	-0.337971
pf_rol	0.837371	-0.818322	-0.763528	1.000000	0.340018	0.384180	0.381328
pf_ss_homicide	0.233121	-0.197345	-0.194366	0.340018	1.000000	0.023412	0.070285
pf_ss_disappearances_violent	0.454195	-0.374685	-0.318923	0.384180	0.023412	1.000000	0.825459
pf_ss_disappearances_fatalities	0.466837	-0.383587	-0.337971	0.381328	0.070285	0.825459	1.000000
pf_ss_disappearances_injuries	0.433746	-0.351852	-0.310665	0.343555	0.057714	0.696925	0.924895
pf_ss_disappearances	0.656795	-0.586487	-0.529214	0.575340	0.099713	0.830986	0.854934
pf_ss_women_fgm	0.357891	-0.362984	-0.363659	0.270723	0.029952	0.096301	0.173496
pf_ss_women	0.619791	-0.631579	-0.630729	0.415056	-0.088723	0.246250	0.275435
pf_ss	0.740191	-0.695259	-0.669068	0.667242	0.621444	0.482509	0.538900
pf_movement	0.692775	-0.690178	-0.669875	0.429536	-0.066645	0.310861	0.344309
pf_expression_killed	0.127004	-0.051044	-0.036546	0.132419	0.119740	0.364793	0.380135
pf_expression_jailed	0.192127	-0.197155	-0.165880	0.120048	-0.094622	0.184909	0.104310
pf_expression_influence	0.767041	-0.782625	-0.745565	0.666099	-0.029103	0.223547	0.223149
pf_expression_control	0.792982	-0.791356	-0.736344	0.729275	0.112144	0.310356	0.308520
pf_expression	0.778418	-0.772944	-0.737748	0.613440	-0.004203	0.352130	0.321514

Table 3

```
In [46]: data.corr()
Out[46]:
```

	hf_score	pf_rol	pf_ss_women	pf_ss	pf_movement	pf_expression	pf_score	ef_government	ef_legal	ef_money
hf_score	1.000000	0.837371	0.619791	0.740191	0.692775	0.778418	0.952638	0.158256	0.793798	0.698807
pf_rol	0.837371	1.000000	0.415056	0.667242	0.429536	0.613440	0.784903	-0.137639	0.910607	0.566157
pf_ss_women	0.619791	0.415056	1.000000	0.630683	0.447890	0.449449	0.665892	0.050387	0.420669	0.296296
pf_ss	0.740191	0.667242	0.630683	1.000000	0.384237	0.445185	0.744773	-0.105112	0.670146	0.441035
pf_movement	0.692775	0.429536	0.447890	0.384237	1.000000	0.629902	0.756099	0.094006	0.371288	0.368795
pf_expression	0.778418	0.613440	0.449449	0.445185	0.629902	1.000000	0.829715	0.154067	0.503724	0.419677
pf_score	0.952638	0.784903	0.665892	0.744773	0.756099	0.829715	1.000000	0.036989	0.696650	0.514170
ef_government	0.158256	-0.137639	0.050387	-0.105112	0.094006	0.154067	0.036989	1.000000	-0.138258	0.167824
ef_legal	0.793798	0.910607	0.420669	0.670146	0.371288	0.503724	0.696650	-0.138258	1.000000	0.584887
ef_money	0.698807	0.566157	0.296296	0.441035	0.368795	0.419677	0.514170	0.167824	0.584887	1.000000
ef_trade	0.816533	0.670223	0.436969	0.569770	0.506096	0.516803	0.683934	0.147352	0.639004	0.725318
ef_regulation	0.726628	0.694254	0.343454	0.530445	0.326540	0.398274	0.576746	0.114051	0.729571	0.572193
ef_score	0.881008	0.757789	0.428999	0.590581	0.460724	0.548646	0.695424	0.316431	0.792159	0.851056

Table 4

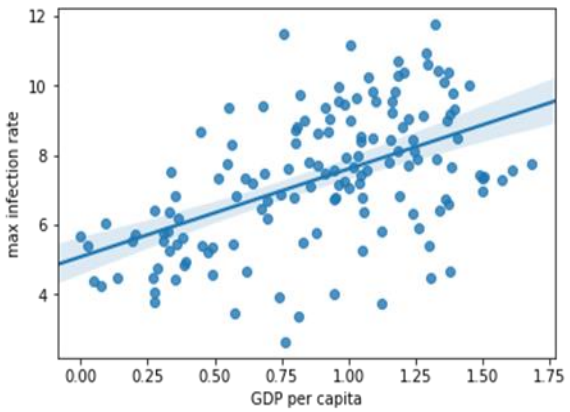

Fig 1

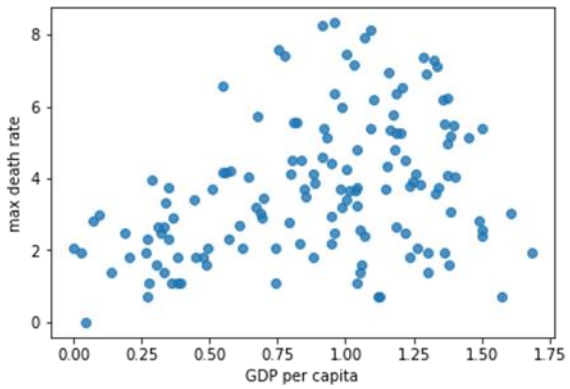

Fig 2

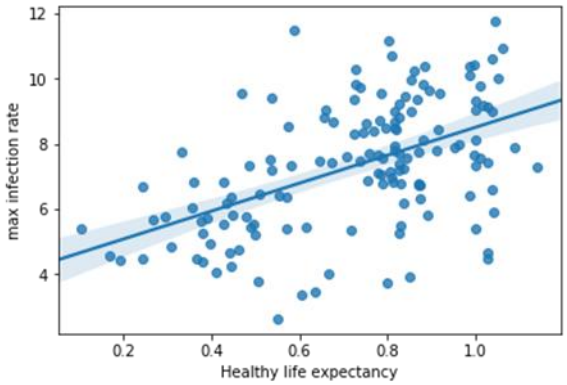

Fig 3

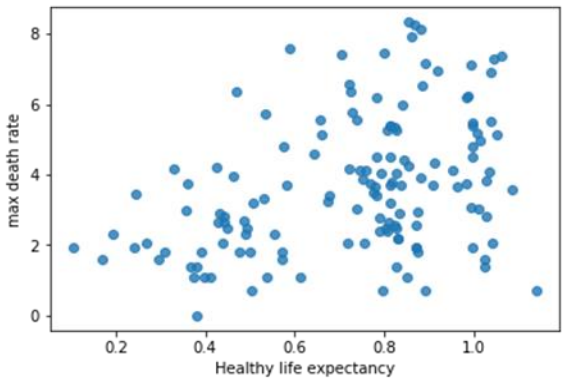

Fig 4

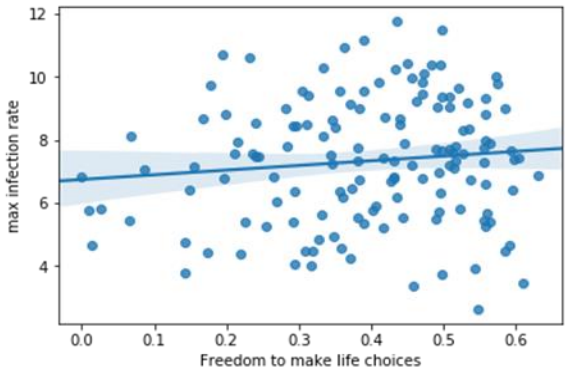

Fig 5

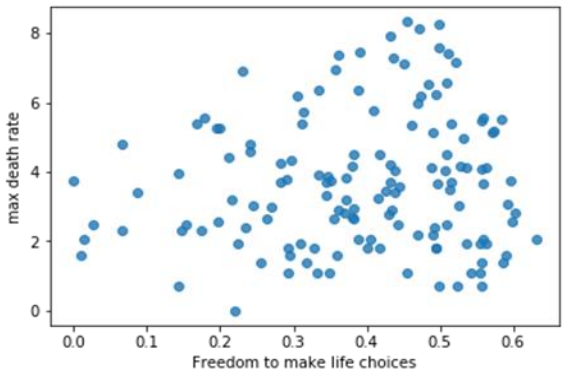

Fig 6

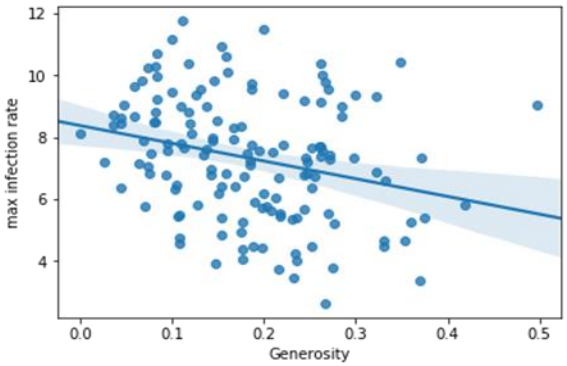

Fig 7

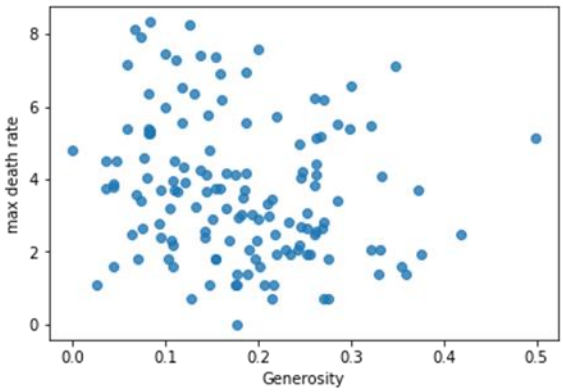

Fig 8

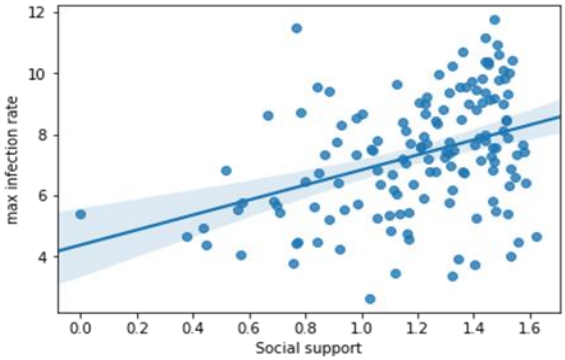

Fig 9

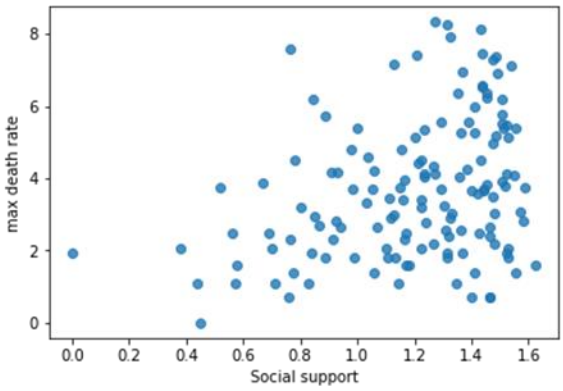

Fig 10

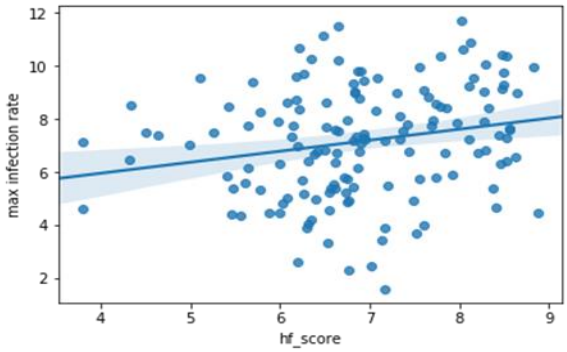

Fig 11

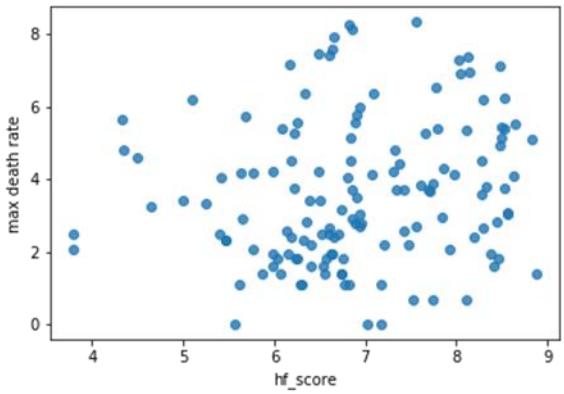

Fig 12

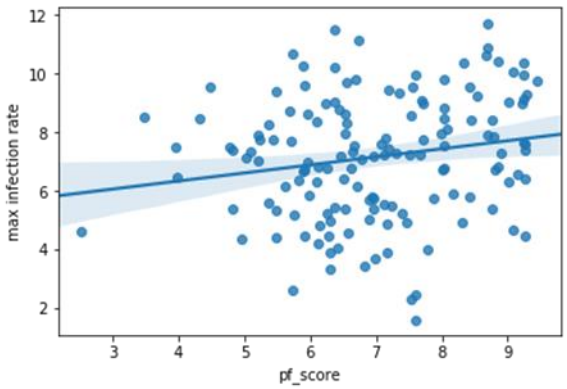

Fig 13

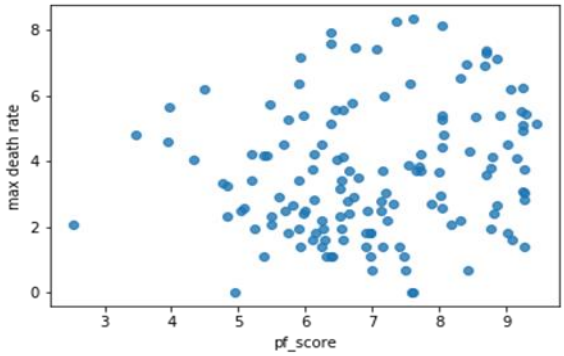

Fig 14

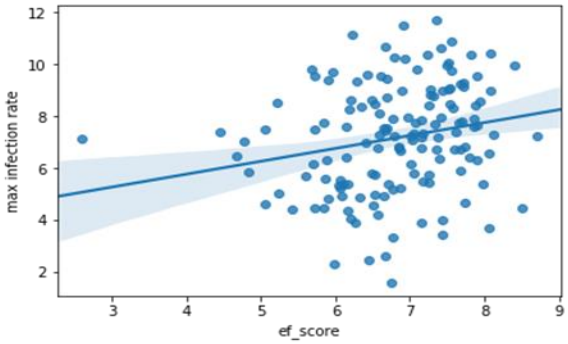

Fig 15

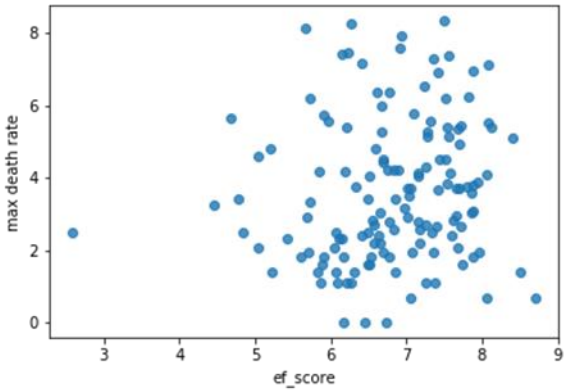

Fig 16

III. ANALYSIS

1. Table 1- Maximum Infection and Maximum Death Rate
2. Table 2-Correlation matrix between World Happiness Report and Table 1.
3. Table 3- Correlation Matrix Between Human Freedom Index and Maximum Infection Rate
4. Table 4-Correlation Matrix Between Human Freedom Index and Maximum Death Rate
5. Figure 1- GDP per Capita vs Maximum Infection Rate. Correlation – Positive (0.253117)
6. Figure 2- GDP per Capita vs Maximum Death Rate. Correlation – Positive (0.147376)
7. Figure 3- Healthy Life Expectancy vs Maximum Infection Rate. Correlation – Positive (0.261623)
8. Figure 4- Healthy Life Expectancy vs Maximum Death Rate. Correlation – Positive (0.205851)
9. Figure 5- Freedom to make life choices vs Maximum Infection Rate. Correlation – Positive (0.060731)
10. Figure 6- Freedom To make life choices vs Maximum Death Rate. Correlation – Positive (0.116149)

11. Figure 7- Generosity vs Maximum Infection Rate. Correlation – Negative (-0.125275)
12. Figure 8- Generosity vs Maximum Death Rate. Correlation – Negative (-0.191924)
13. Figure 9- Social Support vs Maximum Infection Rate. Correlation – Positive (0.182747)
14. Figure 10- Social Support vs Maximum Death Rate. Correlation – Positive (0.145110)
15. Figure 11- Human Freedom Score vs Maximum Infection Rate. Correlation – Positive (0.162267)
16. Figure 12- Human Freedom Score vs Maximum Death Rate. Correlation – Positive (0.080459)
17. Figure 13- Political Freedom Score vs Maximum Infection Rate. Correlation – Positive (0.163547)
18. Figure 14- Political Freedom Score vs Maximum Death Rate. Correlation – Positive (0.112207)
19. Figure 15- Economic Freedom Score vs Maximum Infection Rate. Correlation – Positive (0.129850)
20. Figure 16- Economic Freedom Score vs Maximum Death Rate. Correlation – Positive (0.015723)

IV.CONCLUSIONS

It is observed that a positive correlation exists between all the chosen variables and the maximum infection rate and the maximum death rate barring one exception. There is negative correlation between Generosity and maximum infection and death rates. This means that since there is a positive correlation between two variables such as GDP per capita or Healthy life expectancy and Maximum infection and death rates it indicates that there exists a direct proportionality between the two variables. This suggests that the more the GDP per capita of a particular country the greater the number of cases reported. Similarly, the greater the healthy life expectancy the greater the infection rate. Hence as the prosperity of the country increases so does the maximum number of cases reported. One reason for this trend could be the fact that there is more comprehensive testing in the developed countries compared to the developing and underdeveloped countries. However, on evaluating the same relationship to the maximum death rate of a country, a positive correlation or a direct proportionality is found. This indicates that the number of deaths is also larger in the countries with a greater GDP per capita, Social, political freedom score etc. This means that in general the more developed countries have been worst hit despite the presence of greater medical facilities and research potential. A similar trend has been observed in the comparison with the Human freedom score, Political freedom score and Economic freedom score. All Display a positive correlation, and this shows that more the freedom given to the citizens the greater is the infection rate and death rates. A more comprehensive analysis can be conducted comparing various other parameters to attain a better idea about the trend of spread of Covid-19 in the world.

REFERENCE

[1] https://www.kaggle.com/gsutters/the-human-freedom-index
[2] https://www.fraserinstitute.org/research/human-freedom-index
[3] https://www.who.int
[4] https://www.kaggle.com/mathurinache/world-happiness-report