An Probing Covid-19 Data Analysis Across World

Priya Joshi, Vipra Bohara, Ram Kishan Bairwa

Abstract—The aim of the report is to provide the data analysis of **COVID-19**, a pandemic started in December 2019 and still continued from which lot of lives affected worse.
In this report, we will learn that how to pre-process and merge datasets and we will calculate required measures and produce them for an analysis. By studying data from various countries we have a datasheet contains three datasets. One includes the number of confirmed covid-19 cases and another includes confirmed covid-19 deaths in each country taken cumulatively on daily basis. Another one consists of a number of life factor score of people of each such country so that we can analyse if there is any specific pattern or trend between the spread of covid-19 and contentment score of citizens of that country.
Index Terms- Covid-19, Data Analysis.

I. INTRODUCTION

COVID-19 DATA ANALYSIS

As a data analyst, one always faced with a problem to solve or a question to answer. Our job is to work with different data sources. We should know what are the steps we should take to prepare the data sets and how to look at our data set to find a good measure to calculate for establishing our analysis and to solve the problem in this report, we are going to answer to this question.

Is there any relationship between the spread and related deaths due to corona virus in a country and how happy people are living in that country?

In order to answer the question, we're going to work with three data sets.

Covid19 confimed cases Data set and covid19 deaths data set published by John Hopkins University, which consists of the data related to a cumulative number of confirmed cases indifferent countries per day. The another data, is a data set related to World Happiness Report. The world happiness report is on annual publication of the United Nations .This data set consists of scores given by people those people who are living in different countries to various life factors, such as freedom to make life choices, healthy life expectancy, social support etc.

We're going to merge these data sets and calculate a measure which can help us to find the answer to the question that we have. And at the end, we will visualize the results.

We will work with python* programming language and with jupyter notebook along with Microsoft excel also some modules in python, such as pandas NumPy, matplotlib and Seaborn.

Data Description

Data contains multiple files which includes -

Priya Joshi, Computer Science and Engineering, Yagyavalkya Institute of Technology, Sitapura Jaipur, India.
Vipra Bohara, Assistant Professor ECE, JECRC Sitapura, Jaipur, India.
Ram Kishan Bairwa, Assistant Professor CSE, Yagyavalkya Institute of Technology Sitapura, Jaipur, India.

Covid19_confirmed_dataset:

Having Columns ['province/state', 'country/region', 'Latitute', 'Longitude' and cumulative cases between
22nd January 2020 to 30th April 2020]

covid_confirmed_cases

Covid19_deaths_dataset:
Having Columns ['province/state', 'country/region', 'Latitute', Longitude' and death cases between
22 January 2020 to 30 April 2020]

covid_death_cases

Worldwide_happiness_report : The World Happiness Report is a publication of the United Nations Sustainable Development Solutions Network. It **contains articles and rankings of national happiness**, based on respondent ratings of their own lives, which the report also correlates with various (quality of) life factors.
Having Columns ['overall rank', 'country or re score', 'GDP per capita', 'Social support', 'Healthy life expectancy', 'Freedom to make life choices', 'Generosity', 'Perceptions of corruption']

worldwide_happiness_report.csv

A. Per Capita GDP?

Per capita gross domestic product (GDP) is a financial metric that breaks down a country's economic output per person and is calculated by dividing the GDP of a nation by its population.

1) KEY TAKEAWAYS

Per capita gross domestic product (GDP) calculates a country's economic yield per person and is measured by dividing the GDP of a country by its population.Per capita GDP is a global measure for gauging the wealth of nations and is used by economists, along with GDP, to examine the wealth of a country based on its economic growth.Rich, small countries and more developed industrial countries prone to have the highest per capita GDP.

B. Social support in a country.

Social support was based **on self-reports of access to support from relatives and friends**. Our outcome indicate that the linkage between social capital and health is not confined to high-income countries but increased across many geographical regions in any case of their national-income level.

C. Freedom to make life choices

Freedom of choice describes **an individual's chance and autonomy to perform an action selected from at least** two obtainable options, unconstrained by external parties.

D. Life expectancy of health?

Healthy life expectancy   is **the expected number of remaining years of life spent in good health from a particular age** also we can say it is typically birth or age sixty-five, assuming current rates of mortality and morbidity.

II. RESULT AND ANALYSIS

Visualizing the results using seaborn

As the base paper of sarvam mittal conclude that [1]The main aim of the paper is to study and analyze the COVID-19 spread in India since the day of outbreak and pattern of spreading of virus in India and to understand whyNational and local authorities are having a difficult time in dealing with the COVID-19. Moreover to study about the common symptoms of COVID-19 that are observed till now, age wise spread of COVID -19 to observe which age group is affected most, the spread of disease in India compared to other countries, the state wise trend of the epidemic to get detail understanding of how this is spreading and also to analyze the Healthcare sector of India and lastly to predict the future of epidemic in India.

So,I priya joshi from the previous work of sarvam mittal appreciate its efforts as he compared various fields and shows the great results from his work I get inspired and decided to see thatBy studying data from various countries i have a datasheet contains three datasets. One includes the number of confirmed covid-19 cases and another includes confirmed covid-19 deaths in each country taken cumulatively on daily basis. Another one consists of a number of life factor score of people of each such country so that we can analyse if there is any specific pattern or trend between the spread of covid-19 and contentment score of citizens of that country.

In this task we are going to introduce seaborn module, which is a very handed tool for visualization. We have different columns related to different life factors.Here we have maximum infection rate and death rate of each country and we will see the relationship between the related columns that are GDP per capita, social support, health life expectancy and freedom to make life choices.

Table-1 maximum infection rate comparison with other attributes of world happiness report of various countries.

	max_infection_rate	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
Afghanistan	232.0	0.350	0.517	0.361	0.000
Albania	34.0	0.947	0.848	0.874	0.383
Algeria	199.0	1.002	1.160	0.785	0.086
Argentina	291.0	1.092	1.432	0.881	0.471
Armenia	134.0	0.850	1.055	0.815	0.283

And also we have the maximum infection rate here by plotting every each of these column.So, we have GDP per capita on the x axis and we have the maximum infection rate on y axis.

1. Analysis 1

Plotting GDP v/s Maximum Infection Rate

Out[168]: <AxesSubplot:xlabel='GDP per capita', ylabel='max_infection_rate'>

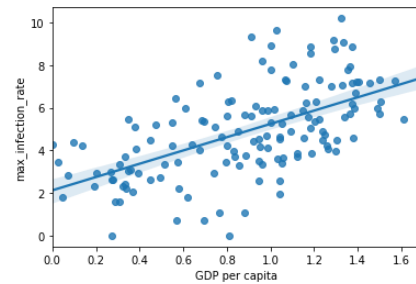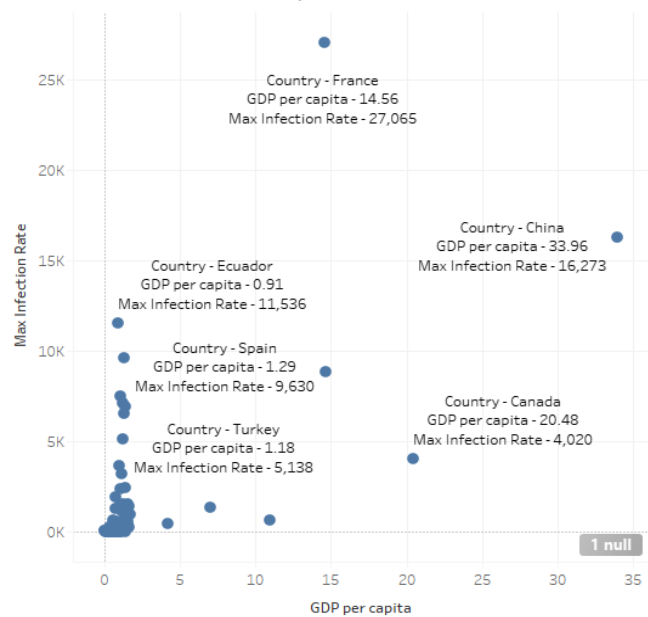Figure- 1Plotting GDP v/s Maximum Infection Rate

So, this graph shows us that as GDP per capita grows, the maximum number of infection

Maximum Infection Rate v/s GDP per Capita

rate is also growing

So in this slope, as we can see there is increase. There is a positive correlation between maximum Infection and GDP per capita. Actually, there is a positive slope between these two variables.

2. Analysis 2

Plotting Social Support v/s Maximum Infection Rate

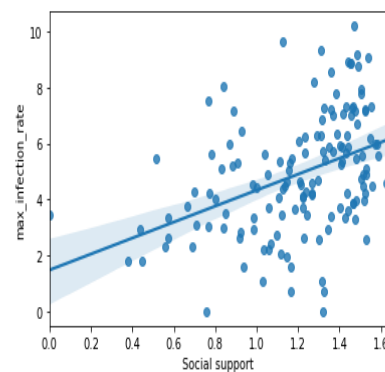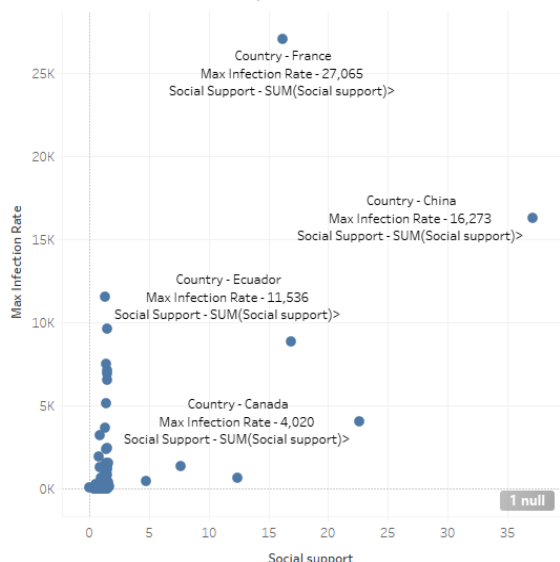Out[169]: <AxesSubplot:xlabel='Social support', ylabel='max_infection_rate'>

Figure – 2

Plotting Social Support v/s Maximum Infection Rate

Maximum Infection Rate v/s Social Support

4. Analysis 4

Plotting Freedom to Make Life Choices v/s Maximum Infection Rate

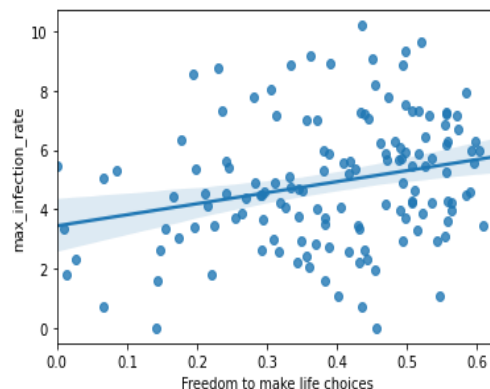Out[171]: <AxesSubplot:xlabel='Freedom to make life choices', ylabel='max_

Figure –4 Plotting Freedom to Make Life Choices v/s Maximum Infection Rate

Maximum Infection Rate v/s Freedom to make Life Choices

3. Analysis 3

Plotting Health Life Expectancy v/s Maximum Infection Rate

<AxesSubplot:xlabel='Healthy life expectancy', ylabel='max_infection_rate'>

Figure – 3 Plotting Health Life Expectancy v/s Maximum Infection Rate

Maximum Infection Rate v/s Healthy Life Expectancy
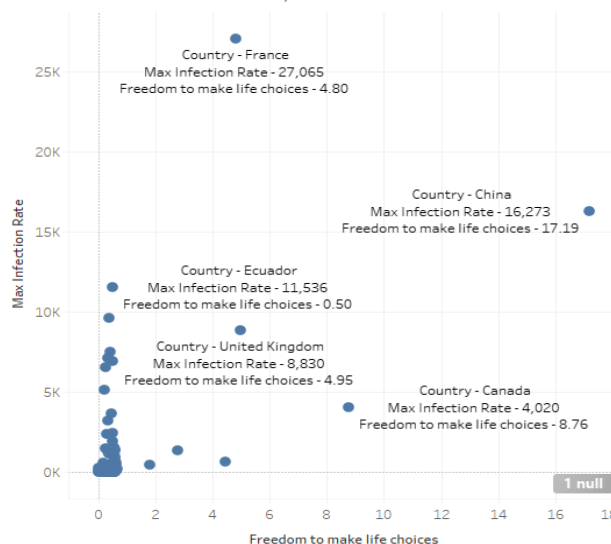
Now, next we have different columns related to different life factors. And also we have the maximum death rate here by plotting every each of these column.

So, we have GDP per capita on the x axis and we have the maximum death rate on y axis.

Table-2 maximum death rate comparison with other attributes of world happiness report of various countries.

In [365]: data= corona_data.join(happiness_report_csv,how="inner")
data.head()

Out[365]:

	max_death_rate	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
Afghanistan	7.0	0.350	0.517	0.361	0.000
Albania	4.0	0.947	0.848	0.874	0.383
Algeria	30.0	1.002	1.160	0.785	0.086
Argentina	13.0	1.092	1.432	0.881	0.471
Armenia	3.0	0.850	1.055	0.815	0.283

5. Analysis 5

Plotting GDP v/s Maximum Death Rate
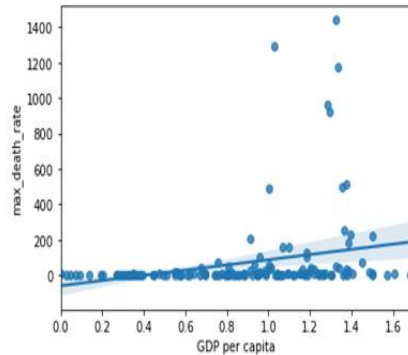
Figure-5 Plotting GDP v/s Maximum Death Rate
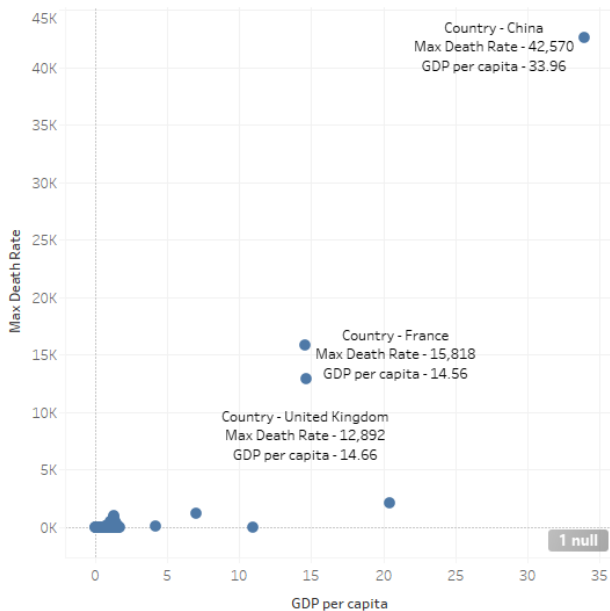
6. Analysis 6

Plotting Social Support v/s Maximum Death Rate
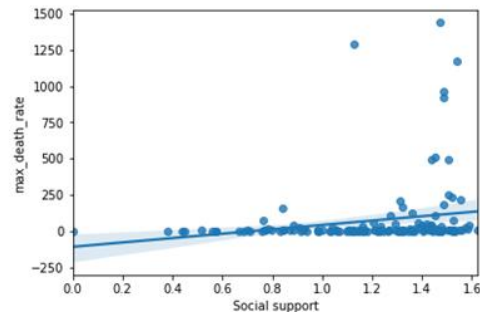
Figure-6 Plotting Social Support v/s Maximum Death Rate
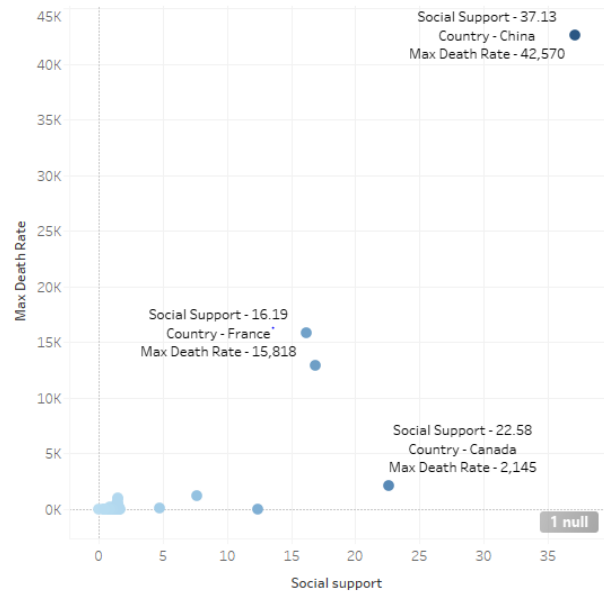
7. Analysis 7

Plotting Health life Expectancy v/s Maximum Death Rate
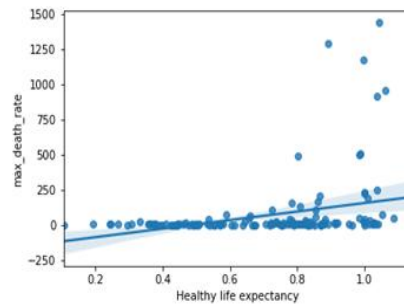
Figure-7 Plotting Health life Expectancy v/s Maximum Death Rate
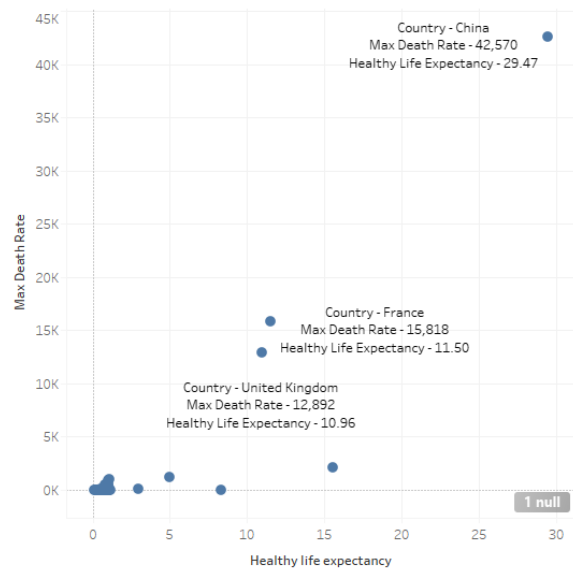
8. Analysis 8

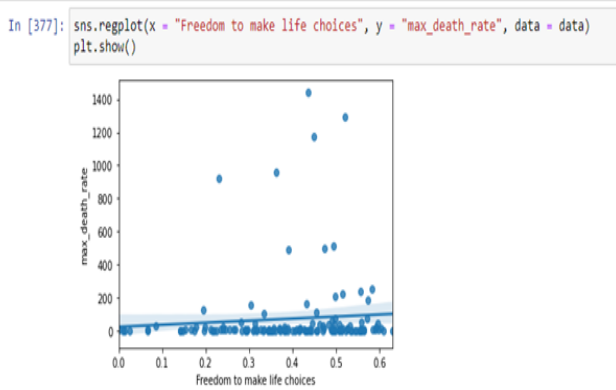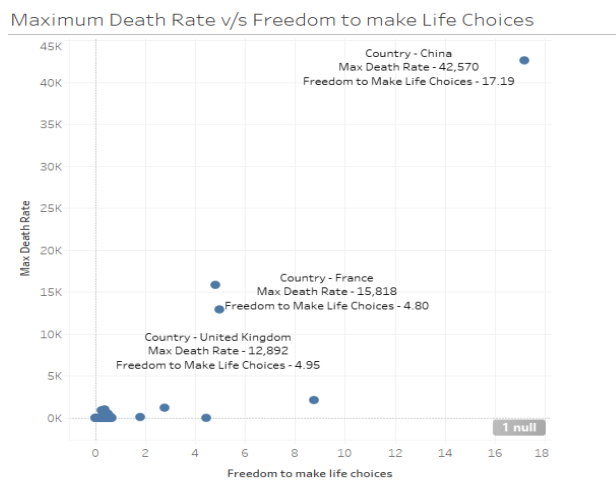### Plotting Freedom to Make Life Choices v/s Maximum Death Rate

Figure – 8 Plotting Freedom to Make Life Choices v/s Maximum Death Rate

After analysing all the graphs we can easily compare our data with the different values which are increasing rapidly, we had calculated the values or countries where maximum infection and maximum death rates are rising.

Table-3 comparing values of different countries.

Country/Region	Max Infection Rat	Max Death Rate	GDP per capi	Social suppo	Healthy life expectan	Freedom to make life choice
Belgium	2454	496	1.356	1.504	0.986	0.473
Brazil	7502	493	1.004	1.439	0.802	0.39
China	14840	1290	1.029	1.125	0.893	0.521
France	26843	1438	1.324	1.472	1.045	0.436
Germany	6933	510	1.373	1.454	0.987	0.495
Italy	6557	919	1.294	1.488	1.039	0.231
Spain	9630	961	1.286	1.484	1.062	0.362
United Kingdom	8681	1172	1.333	1.538	0.996	0.45

From this observation we can easily make out the conclusion that developed countries like China, France, United kingdom, Italy, Spain, Germany are more prone to the infection as compare to less developed nations.

III. CONCLUSION

An effective research consists of rigorous and in depth analysis of data sets and also observing pattern in between them so as to suffice the basic purpose or major objective of the research. After analysing the datasets from verified sources as mentioned in the report, we have observed very great insights from this pandemic data. Covid-19 has made humans to again stop at this scenario to look around and then think carefully about their progress in medical sciences. Although this pandemic has taken many lives but it has also alerted the humans about the progress we all lack in our medical science domain and hygiene we need to maintain in our daily habits.

As we moved through our analysis we observed some beautiful intriguing patterns between datasets. We saw the number of covid-19 cases were more in the countries we thought and remember as so called "developed nations" but there may be certain reasons behind them as developed nations were going through more testing per person as compared to other developing countries. We also observed that countries having maximum infection rates at some point of time were not directly proportional in the maximum death rates at similar time references. There are numbers of possibilities for reasons of these insights that is better health facilities in some countries could have reduced the number of deaths even if infection rate was very high comparatively.

REFERENCES

[1] **Sarvam Mittal**International Journal of Engineering and Technical Research 9(4):580 DOI:10.17577/IJERTV9IS040550
[2] https://www.mohfw.gov.in/
[3] https://www.covid19india.org/
[4] https://github.com/CSSEGISandData/COVID-19
[5] https://worldhappiness.report/
[6] https://pib.gov.in/PressReleasePage.aspx?PRID1539877
[7] https://www.worldometers.info/coronavirus/country/india/
[8] D. Gondauri, E.Mikautadze and M. Batiashvili, Research on covid- 19 virus spreading statistics based on the examples of the cases from different countries, Electron J Gen Med.2020;17(4),em(209)