

Determining the Countries Prepared for a Future Pandemic Based on Analysis and Visualization of Novel COVID-19 Virus

Prafulla Kumar Dubey

Dept. of Computer Science and
Engineering
SRM Institute of Science and Technology
New Delhi, India
praffullakrdubey@gmail.com

Udbhav Naryani

Dept. of Computer Science and
Engineering
SRM Institute of Science and Technology
New Delhi, India
udbhav.naryani@gmail.com

Medhavi Malik

Dept. of Computer Science and
Engineering
SRM Institute of Science and Technology
New Delhi, India
medhavimalik28@gmail.com

Abstract - The COVID-19 has been announced as a pandemic by WHO. At the time of writing this paper more than three lakh people have already got infected from this disease. This virus has shown different statistics around the globe. As the virus spread, some countries were able to handle the adverse conditions and were able to control the spread of virus whereas many other countries were not able to handle the spread. Due to this there is a need to determine which countries need to invest more in health care so that they are prepared for similar adverse conditions in the future. We propose to satisfy this need by analysing the number of total patients, recovered patients and total number of deaths in different countries. Using the above three attributes we perform the analysis.

Keywords- COVID-19, Corona Virus, Data Analysis, Health Care Investments, Spatial Analysis.

I. INTRODUCTION

The Coronavirus disease, termed as COVID-19 is a widespread infectious disease. This ongoing spread has been declared as a pandemic by WHO on 30th January 2020 [1]. This infection spreads rapidly in places that are confined and restricted. Usual symptoms for this virus includes shortened breath, fever and cough. Sputum production, sore throat and muscle pain are rare symptoms [5]. While most of the cases can result in only few symptoms [6] some of the symptoms lead to serious pneumonia as well as failure of different types of organs [7].

The rate of deaths per number of diagnosed cases is 4.1%; however, it ranges from 0.2% to 15% depending on age and other health problems. In order to be prepared for a pandemic like this, the hospitals in the country need to have a proper strategy to be able to manage properly their staff, supplies and space in order to provide best care to the patients [8]. This outbreak has become a large scale clinical threat for the general public and also for the healthcare workers around the world whereas the knowledge about the virus is very limited [9].

As of 22 March 2020 there are more than three lakh cases. Out of these more than thirteen thousand people have lost their lives and more than Ninety-five thousand patients have successfully recovered from this disease, percentage of people who died per the number of people who were diagnosed comes out to be 4.1%. Once a person tests positive for the disease he/she should be isolated to

prevent the spread of disease. If the person is isolated, then the spread of disease can be prevented completely. Now to prevent the spread of this virus in a country requires the following-

1. Social Distancing

Social Distancing is a practice that should be adopted by all the citizens of the country. If citizens avoid closely and physically interacting with each other, then the virus spread can be prevented by large extent. It can take up to 14 days for the symptoms of virus to show up. No one can be sure is the other person is having the virus or carrying the virus in their hand or not. Hence social distancing is the key to stop virus spread [2].

2. Hygiene

Even if a person comes in contact with an infected person, if the person keeps good hygiene then the chances of that person getting infected are very less. Countries whose citizens have good Hygiene will have a limited spread of this virus [3].

3. Testing/Screening of People

In order to determine whether a person is suffering from COVID-19, proper testing for this infection should be made available for the people. If people know that they are infected, then they can be isolated. Countries should also test the people coming from other countries at the airport itself. Countries which will make proper and accurate testing and screening facilities available for the general public would be able to identify majority of the cases and can prevent the spread.

4. Isolation

Isolation is the most important role of a country in preventing the virus to spread. This infection can only be

spread through physical touch. If a person tests positive for the COVID-19 virus, then that person should be isolated or quarantined from everyone else. If

II. METHODOLOGY

Data Science is considered as one of the most trending topics around the globe. Data Analysis is a part of Data Science which is used to analyze the data and find out information from large bundle of data. Big Data Analysis can also be one of the approaches for analyzing the data but in this paper we have used Data Science technique as this technique provides more benefits and gives a better performance. The steps involved in this technique for analysis task are:

1. Data Import: The relevant dataset of COVID-19 is “covid_19_data.csv”. It is a comma separated values file which contains five thousand six hundred and thirty two rows and eight columns. This dataset has been provided by Johns Hopkins University and this dataset is used for analysis purpose.

2. Data Cleaning: It can be considered as the most tedious and time taking process but this step cannot be ignored as it is one of the most important steps. The dataset that we use is very big and has been obtained from various sources therefore it contains various null values and duplicate values. This is the main reason for data cleaning, and after cleaning the data we have the dataset containing only useful data.

3. Wrangling Data: It is the process in which data in its raw form is used to map and transform into numerous forms that are different from previous form. This newly formed form of data is more precise for the task of analysis.

4. Data Visualization: Visualization can be termed as one of the key feature of data analysis task. The relationship among various attributes present in the dataset can be visualized in the form various graphs or patterns. The result of these graphs are used for analysis purpose.

5. Modeling Data: Modeling of data is done to determine the various relationships among the various information from dataset and diagrams. There can be various types of diagrams which include histogram, replot, catplot, barplot, scatter-plot, histogram and etc.

The dataset had two hundred and sixty one “State/Province” categorical data and one hundred and fifty six “Country/Region” categorical data. As the names of these countries were too big so for ease of visualization these are categorical data are encoded to numeric data i.e. 0 to 260 for “State/Province” and 0 to 155 for “Country/Region”. The library used for it is the “sklearn.preprocessing.OneHotEncoder.”

This library is used to transform the categorical data into one-hot numeric array. The inputs are transformed into array like of strings or integers denoting the values that were taken from discrete features [10].

Table I: Attributes of the data set.

Attribute Value	Data Type	Representation	Summary
Serial Number	Integer	Sno	Stores the serial number
Observation Date	Date	ObservationDate	Date on which COVID-19 has been observed in the patient
State or Province	Integer	State/Province	An unique integer to represent the name of the State or Province
Country or Region	Integer	Country/Region	An unique integer to represent the name of the Country or Region
Update	Integer	LastUpdate	Stores the updated time and date of the cases
Confirmed	Integer	Confirmed	Total confirmed cases
Deaths	Integer	Deaths	Total deaths occurred
Recovered	Integer	Recovered	Total patients recovered



Fig. 1. Methodology.

III. PROPOSED WORK

In this paper we propose to analyze the dataset of COVID-19 and find out the major findings by analyzing the various graphs.

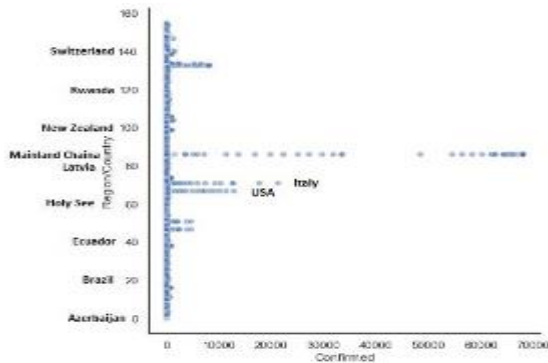


Fig. 2. Confirmed VS. Region/Country.

We can analyze from figure 2 that, there are approximate seventy thousand cases that are confirmed. Mainland China out of various countries present in the dataset has the maximum number of confirmed cases of COVID-19. After Mainland China, Italy and USA are the countries which have more number of confirmed cases than other countries.

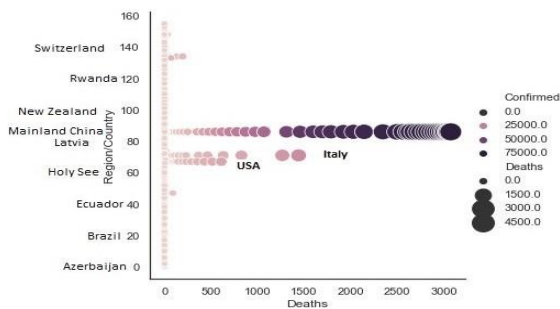


Fig. 3. Deaths VS. Region/Country.

From figure 3 it can be analysed that as Mainland China had the maximum number of cases of COVID-19, therefore we can analyse that maximum death has occurred in Mainland China then followed by Italy and then followed by USA.

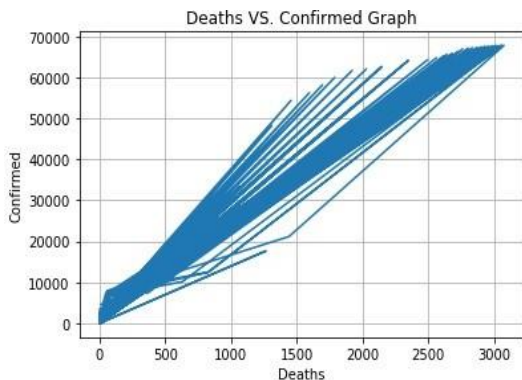


Fig. 4. Deaths VS. Confirmed.

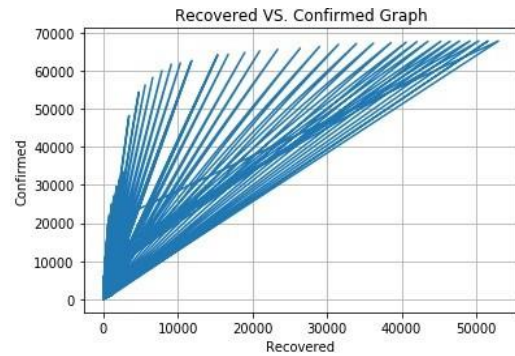


Fig. 5. Recovered VS. Confirmed.

From the figures 4 and 5 we can see that, there is a trend followed by number of deaths and number of recovered patients. These figures show that as time passed, the trend of people recovering started increasing and the number of people dying started decreasing in major countries.

IV. RESULT

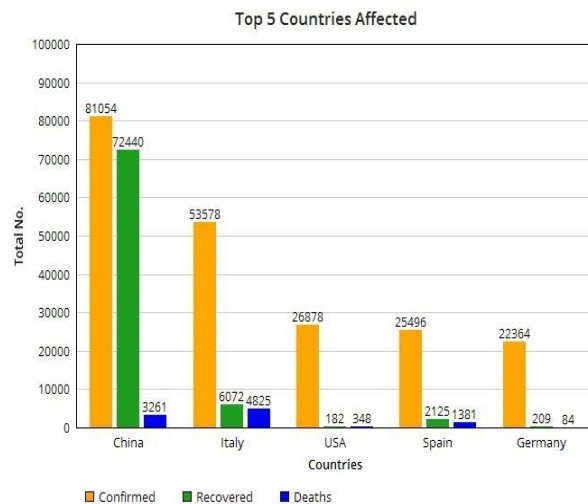


Fig. 6. Graph for Top 5 Countries Affected.

The final analysis from the dataset can be concluded as the result. We analyse that the top 5 countries till date that are affected with the Novel COVID-19 virus are China, Italy, USA, Spain and Germany. China records the maximum number of confirmed cases and is considered as the epicentre for the birth of this virus. China also records maximum number of recovered cases which show that China has a good and well developed medical facility and it can be ready for future pandemic.

Whereas, Italy stands in the second position but it fails to fight against the virus as it has more number of deaths due to COVID-19 virus.

On the other hand, countries like USA, Spain, and Germany have many affected patients but there are not as many deaths as compared to Italy.

V.CONCLUSION

From the present scenario we can suggest that China will be prepared for future pandemic and various other countries that take preventive measures before spread of any such virus beforehand would be ready for any future pandemic.

REFERENCES

- [1]. N Zhu, D Zhang, W Wang, et al. A novel coronavirus from patients with pneumonia in China, 2019, published online Jan 24. DOI:10.1056/NEJMoa2001017.
- [2]. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Science. 16 March 2020. [Preprint ahead of publication]
- [3]. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395:497–506. doi:10.1016/S0140-6736(20)30183-5.
- [4]. European Centre for Disease Prevention and Control Public health management of persons having had contact with novel coronavirus cases in the European Union. European Centre for Disease Prevention and Control, 2020
<https://www.ecdc.europa.eu/en/publications-data/public-health-management-persons-having-had-contact-novel-coronavirus-cases>
- [5]. "Coronavirus Disease 2019 (COVID-19) Symptoms". Centers for Disease Control and Prevention. United States. 10 February 2020
- [6]. L. Li et al., "Characterizing the Propagation of Situational Information in Social Media During COVID-19 Epidemic: A Case Study on Weibo," in IEEE Transactions on Computational Social Systems.
- [7]. Salzberger, B., Glück, T. & Ehrenstein, B. Successful containment of COVID-19: the WHO-Report on the COVID-19 outbreak in China. Infection (2020).
<https://doi.org/10.1007/s15010-020-01409-4>
- [8]. Wong, J., Goh, Q.Y., Tan, Z. et al. Preparing for a COVID-19 pandemic: a review of operating room outbreak response measures in a large tertiary hospital in Singapore. Can J Anesth/J Can Anesth (2020). <https://doi.org/10.1007/s12630-020-01620-9>
- [9]. C. Wang, P.W. Horby, F.G. Hayden, G.F. Gao. "A novel coronavirus outbreak of global health concern Lancet, 395 (2020), pp. 470-473, 10.1016/S0140-6736(20)30185-9.
- [10]. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>