DATA STRUCTURES CS-102

VIRTUAL CLASS HACKATHON

CORONA VIRUS DATA ANALYSIS

Team Details:

1) Pranav K M – 19BCS088 – First problem statement

2) Sanchit Goel – 19BCS095 – Second and Third problem statement

3) B M Kalpajeet – 19BEC007 – Third problem and Research Article

4) Gourab Chakraborty – 19BEC014 – First problem statement

5) Jaynit Patel – 19BEC019 – Third problem statement

References and Data Sets:

- https://www.mohfw.gov.in/
- https://covindia.com/#latest-update
- https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov
- https://www.worldometers.info/coronavirus/country/
- https://www.statista.com/page/covid-19-coronavirus
- https://www.coronavirus.gov/
- https://covid19api.com/
- https://api.covid19api.com/total/dayone/country/
- https://api.covid19api.com/countries
- https://pomber.github.io/covid19/timeseries.json
- John Hopkins University

Contents

1) Purpose of the Research Topic Study

28,97,645 confirmed cases, 2,02,880 deaths and no medicine or vaccine till date. This has been the gripping tale of Corona virus, named Covid-19, till date. And all this in a span of just 4 months. All it takes is miniscule virus to enter our body. What began from a supposed bat-meal in Wuhan has turned out as a worldwide pandemic, threatening to wipe off a fourth of the population and much of the economy, if not contained. The impact is massive. The inability to find any particular trend or a biological vaccine leaves the globe with an uncertain future. Even if this pandemic end, countries will remain as hotspots and high-risk travel destinations. Even though the mortality rate might not be high, the Covid-19 affects all sorts of people, leaving no disposition. But there might just be some trends.

As budding data scientists, this immense pandemic leaves us a lot of data, dynamically changing every moment, at our disposal. In this short research, we try to play with this incredible amount of data and try to find specific trends that may help us to decode the containment policies and spread of this virus. We first try to analyze the latest data to point out high-risk travel destinations for students. Secondly, we also focus on the age group of 15 to 25 years, try to understand how this sect is getting affected and how fast. Our third motive is to analyze how many days it takes for a country to reach a death-state.

As students of data structures and data visualization, we believe that our research can be accomplished using various data structures and graphical visualizations. The research focuses on a global scale and uses highly verified and legitimate data sets. The research intends to pin-point locations that will have a lasting effect of this pandemic spread, the resistance to mortality of a particular age group and the average amount of time it takes for this spread to attain death state and how other cases can culminate into recoveries. The research expects to study these trends, among others, and perhaps even achieve a new quarantine policy that may help the world and governments, to stabilize economy and curtail people's fears.

2) What are the important observations found in the data analysis done in this research?

The spread of the Corona virus on a world-wide scale has many facets. Much of these depend on the testing of samples, transportation of people – immigration, emigration and general movement, logistical factors, biological capacity and immunity of the people affected, economical and even geographical factors.

- The number of Corona virus cases is growing at an exponential rate daily. The factor of the exponential differs from country to country, but it has maintained its exponential nature. The value of the exponent is greater for countries like USA, Spain, Italy – countries which had greater initial exposure to travel. While countries like Japan, South Korea, Australia had lesser value of the exponent, as they put early travel restrictions.

- The confirmed cases in India was doubling at a rate of 4 and a half days a little before lockdown, at a rate of 6.2 days post 1st lockdown, and is now doubling at a rate of 8 days. This is a significant improvement compared to other countries- like the USA where cases double every 3 days, and Spain with 4 days- though India has had a major economic setback due to this. Thus, social distancing is the key.

- While graphing results as per the legitimate data, we found out that the number of deaths did not follow any significant trend in general, as it depended on which country handled the situation better. But when looked individually, the death rate too follows a mild exponential growth. It seems more like a prey-predator model, where the general death rate due to Corona virus has valleys due to the reduction in number of deaths and an increase in recovered cases.

- On an average, after the first confirmed case, it takes about 16 to 18 days for the first recovery to take place. Thereafter, the recovery cases thoroughly outnumbers the daily death cases. This gives an indication that if the confirmed cases curve can be flattened and decreased to zero over a period of two/three weeks, then with the current mortality rates, there is the possibility of ending the pandemic within 6 weeks, with the hope that there are no remaining cases, and no resurfacing.

- As per the World Health Organization and Ministry of Health and Family Welfare's (India) data, the research found out that the fatality rate of the age group 65 and above has a whopping mortality rate of 47.7%. The 15 to 25 years age group has a mortality rate of barely 2%, mostly due to better immunity. This makes this age group very important for the coming future.

- The most common countries that students prefer for internships or higher studies like USA, Germany, UK are at a higher risk as per the current situation. Another major country is Australia and as per the current scenario, it is at a better situation and is a viable option.

- In what could be considered prime, about 68% of the recoveries were in the age group of 15 to 30 years. This points at a possibility of resistance in this age groups and perhaps a chance to study their recovery biology.

- In what seems a tough pill to digest, the vaccine for Corona virus is still ages away in development and vaccinating at a massive scale will too take time. In this scenario, social distancing is the only possible way forward.

3) Data files and the sites used for study

- World Health Organization
- Ministry of Health and Family Welfare, India
- The US Department of Health and Health Services
- Covindia.com
- Worldometers.info
- Statista.com

4) Data Types and Data Structures used

- Integers and floats
- Strings

- Time datatype
- json (Lists)

- Json (Dictionary)
- Web scraping technique

5) Final Results and Conclusion

On the face of this enormous and seemingly unstoppable pandemic, the crux of the world's strategy to recover relies on how to get the spread of the virus plateaued and move over to a state of normalcy. Things are worse than they are already, with the World Health Organization declaring that there's a huge possibility of recurrence of Corona virus, even later this year. Amidst all this uncertainty, our research and data analysis concludes on some very fixed trends that could be of salient essence in drawing out better plans or policies. **The mortality rate of 60+ years is higher, compared to the age group 15 to 30 years. This implies that if we can draw out a statistic of the aged, and secure them without travel and isolation, the death rate could be controlled.** Also, the age group of 15 to 25, which consists of students majorly, need to be kept safe, yet not let their academics affected. **Since the recovery rate of this age group is also higher, their plasma and blood traits could be studied (with consent) to find a vaccine. Countries with longer time to achieve death state, and least time to attain recovery state must try to completely flatten the fatality and spread curve within 6 weeks.**

As per our study, countries with the present fatality rate of 7% and higher must be kept under strict lockdown, with the countrymen being ethical and follow social distancing. Considering the situation of India, **our data analysis points that case growth in India has been modest so far, courtesy the lockdown**, and it has saved some lives. If lives are priceless, they're worth saving at any cost — that's the current mantra. The media frenzy around corona, our own panic regarding this disease, a desire to be seen as a responsible country in the eyes of the international community are making us look at corona differently than almost anything else we have in the past. There is definitely a genuine, underlying intention to protect Indians from this disease. However, we cannot ignore the economic reality of our country and our existing problems versus the rest of the world. **In view of our research, maybe a plan with somewhat less stringent measures may be more sustainable – using the working age group to maintain sustained economy, securing the aged as per census records and focusing of a vaccine using the recovered patients' help.** Frankly, it might be all we can afford as a nation anyway. As we continue with the lockdown, it is time we unlock our minds, analyze all available data and see this problem from every perspective before we decide what to do next, which is the essence of this research.