

# Pandemic COVID-19 Analysis and Visualization

Md. Abrar Jahin

**2<sup>nd</sup> year, Industrial Engineering and  
Management**

# Abstract

The 2019-nCoV is a contagious coronavirus that hailed from Wuhan, China. This new strain of virus has struck fear in many countries as cities are quarantined and hospitals are overcrowded. This dataset will help us understand how 2019-nCoV is spread around the world. In the following dataset global death rate, new cases, new recovered, confirmed, active, change in 1 week and %increment, total population, timestamps etc. have been included and considered. Useful Python data analysis & visualization libraries numpy, pandas, matplotlib, seaborn etc. have been used to answer

# Motivation

This analysis and visualization is the key factor of survival for 2020 while the USA is facing continuous death increase, economic fall and losing equilibrium of the system in spite of having such great power and developed technology. I tried to visualize the chain multiplication of the virus overtime and compare the death, active, recovered cases within a certain country and globally as well. The map showing the Red Spots indicates Corona affected zones and epidemic span, comparison with SARS, Ebola, H1N1 etc. also have been shown to relate the insights to reduce the spread, find a pattern. Less affected and fast recovered countries and their reasons have been analyzed.

# Dataset(s)

The datasets have been collected from:

1. <https://cgdv.github.io/challenges/COVID-19/datasource/>
2. <https://github.com/CSSEGISandData/COVID-19>
3. [https://github.com/imdevskp/covid\\_19\\_jhu\\_data\\_web\\_scrap\\_and\\_cleaning](https://github.com/imdevskp/covid_19_jhu_data_web_scrap_and_cleaning)

There are confirmed positive, death, recovered, active cases mentioned for each countries and the data regarding spread and death rate in other epidemics have been enlisted. The datasets are authentic, reliable and clean enough compared to other sites.

# Data Preparation and Cleaning

At first the pandemic cases are totally out of control and pattern-less which caused outliers for a few times. Data cleaning including finding null values, filtering a certain number of cases as well as merging data resulted in our final outputs. The inner join, group-by etc. have been done based on the countries most of the time. It is the lengthiest process among all steps and takes a lot of patience and trials. Fortunately there were no null values which didn't cause panic while visualizing graphs. Still the death, recover, confirmed cases are changing continuously which compels the data sets to be upgraded from time to time.

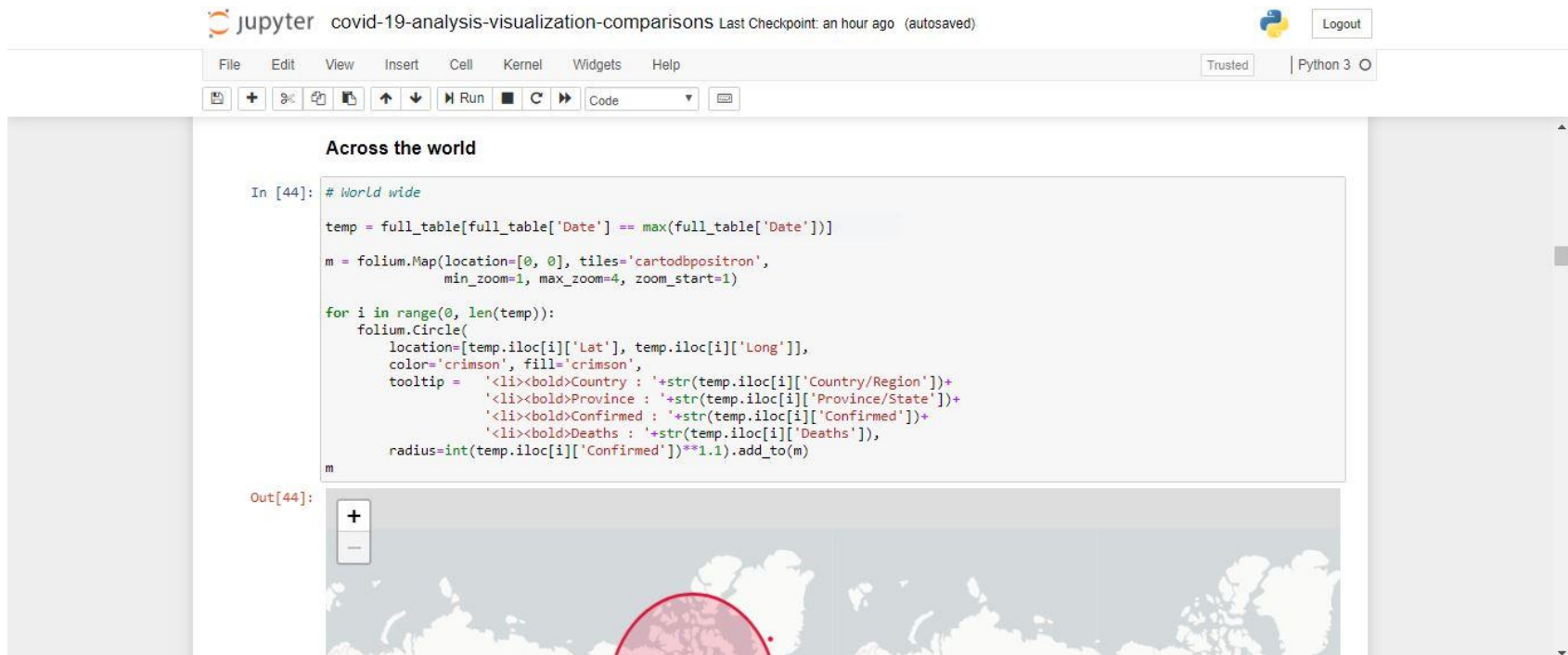
# Research Question(s)

- What is the impact of Covid 19 on marginalized populations?
- Based on US/Western populations across LMICs, where are these populations most at risk based on the current rate of spread of Covid-19?
- Based on current situations and physical distancing norms where are the next 'hot zones' going to develop around the world?
- Given levels of physical distancing – will this be sufficient to flatten the curve?
- What will be the impact of no-go zones - marginalized areas in cities - on the ability to stop the spread of the virus
- Are countries that have active citizen engagement mitigating the spread of Covid 19 more effectively?
- What is the nature of the changing relationship between China and Taiwan, Hong Kong, the US, Germany, Italy, France, the UK and others?
- What would be the impact of religious distress due to COVID 19 on particular community in India?
- What has been the economic impact of Covid 19 on the US, China, Russia, and Germany?

# Methods

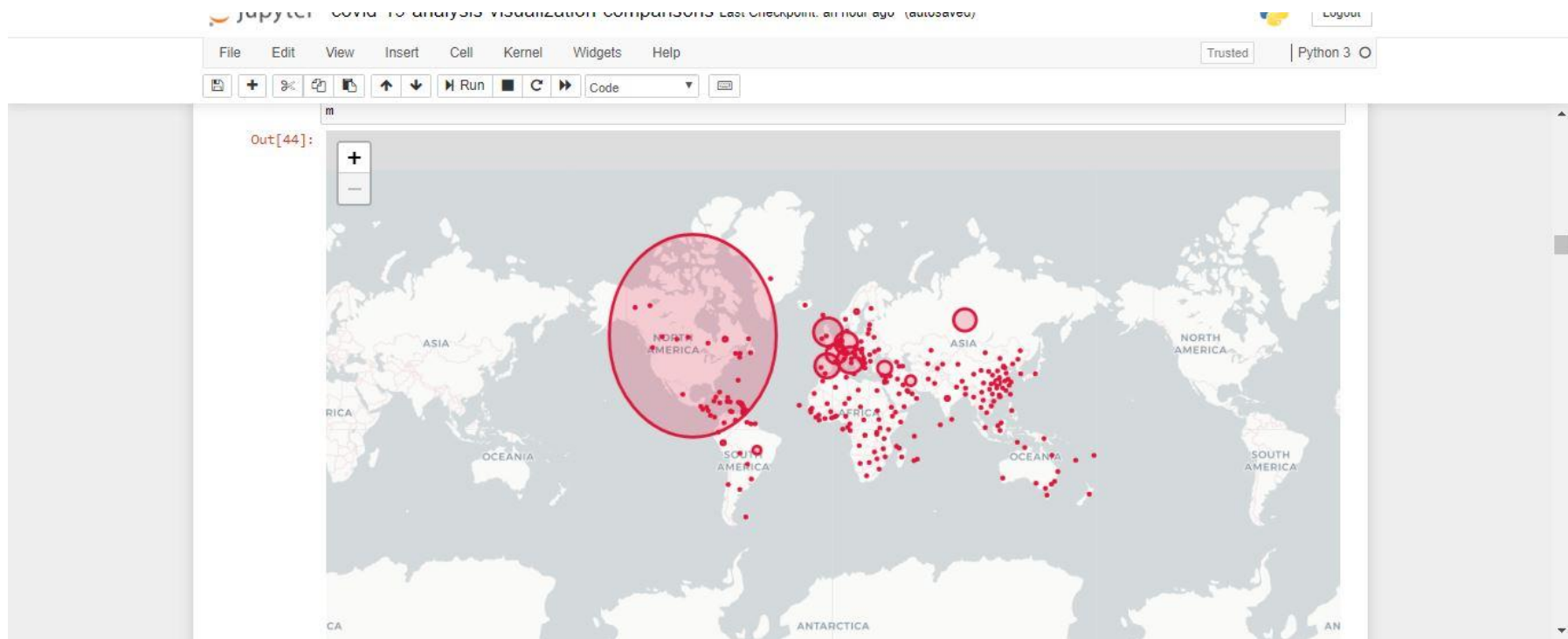
Using Python numpy, pandas, matplotlib, seaborn, datetime, math libraries the whole analysis has been done including graphs, dataframes, charts, maps etc. The whole methods have been attached with the existing pdf file.

# Findings





# Findings



# Findings

The USA has been suffering from COVID-19 mostly in spite of having developed medical equipments and technologies which is a remarkable threat to the other developing as well as developed countries too. Italy has almost been damaged by it though they possess the best medical facilities. The economic condition is breaking in these developed European countries as well as South East Asian countries including India, Bangladesh, Pakistan etc. just because of having an innumerable and unorganized population it became so tough to keep them quarantined for a few months. The people are blind to religious cultures and taboos like basking in the sun will keep them safe from the virus. The virus didn't spread within a day rather it took a few months which was ignored by the administrators and resulted in uncontrolled situation. The only fear is that there is

# Findings

No vaccine for the virus still invented which can control that chain multiplication. Moreover, the government of developed countries are providing services and foods to the nation while the developing and under developed countries are deprived of that which compels the govt. to keep some of the industries open just to keep the country mobile.

Social distancing can definitely flatten the curve and lengthen the contamination and the medical team will get time to treat the patients otherwise a huge portion of the nation will be dead if quarantine rule is not being followed. The active citizens i.e. military, police, doctors, food delivery men are contributing the highest in this situation and if the proper regulations are not being followed, it can

# Findings

Even cause more danger to the general people. The jobless people are in tension regarding managing food for themselves where simple sympathy and help to each other can make the society more worthy for living and pandemic is temporary but that love towards each other is permanent.

# Limitations

There are a lot of limitations:

- The pandemic situation is still prevailing and data is changing and growing fast.
- Continuous data growth causes limitations of analysis.
- Many countries are recovering while other countries are still in the red zone which can't be visualize continuously.
- The contamination based on gender could have been shown, but lack of data caused limitations.
- Generally the old people are highly affected due to this pandemic but there are exceptions too which have been ignored.

# Conclusions

The pandemic situation intensifies the value of this research which can give certain insights about the COVID-19. The nature of its spread, multiplication rate, death rate, recoveries, economic conditions, predictions of next decades can easily be expressed by visualizing these simple graphs. The helpless military and nuclear power proves that simple virus can be the reason of destruction of the whole mankind indicating another prediction called 'Biological War'. The whole situation is dependent upon the researchers, biologists, chemists, data scientists, doctors who are putting restless efforts to formulate a vaccine and save the human being. The research is not totally correct, limitations lie within data and more sophisticated codes could be performed which hasn't been run. So this overall procedures have been performed to leave a mark and contribute to the pandemic history.

# Acknowledgements

I would like to acknowledge Johns Hopkins University for open-sourcing their dataset. Their dataset is transformed into a format that is easier for Jupyter Notebook to handle.

I believe that epidemic data should be openly available and easily accessible for health professionals and data scientists. This dataset would serve as a starting point for people to gather more data about epidemics, not just statistics, but also new stories, government responses etc.

# References

1. <https://cgdv.github.io/challenges/COVID-19/datasource/>
2. <https://github.com/CSSEGISandData/COVID-19>
3. [https://github.com/imdevskp/covid\\_19\\_jhu\\_data\\_web\\_scrap\\_and\\_cleaning](https://github.com/imdevskp/covid_19_jhu_data_web_scrap_and_cleaning)

The datasets have been collected from the above sites and rest of the data processing, cleaning, transformation, analysis and visualization have been done by me.