International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

Exploratory Data Analysis on outbreak of CORONAVIRUS

S SHASHANK RAJ

Student, Department of Computer Science, ICFAI University, Telangana, India.

Abstract - *A prevailing virus which was first reported in the Wuhan, a city in China, now spread across the world, which ignited a prodigious health and economic crisis. The World Health Organization (WHO) has announced the eruption of the novel coronavirus "a public health emergency of international concern". In this paper, we will take a momentary look at the current crisis and then drive into Kaggle's "Novel Corona Virus 2019 Dataset".*

Key Words: CoV- Coronavirus, WHO – World Health Organization, MERS-CoC – Middile East Respiratory Syndrome, SARS-CoV – Severe Acute Respiratory Syndrome, EDA – Exploratory Data Analysis

1. INTRODUCTION

The WHO, says that the coronaviruses (CoV) are a substantial family of viruses that cause ailment ranging from the normal cold to a degenerative or a serious disease such as Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV). This virus is a new strain which is not identified in humans formerly, it is recognized as the cause of the former outbreak referred to as 2019-nCoV or Wuhan coronavirus. It can be transmitted between people and animals, after a detailed investigation it was found that the SARS-CoV was transmitted from civet cats to humans and MERS-CoV from dromedary camels to humans, and many more coronaviruses are transmitting in between animals which are not yet infected humans.

1.1 Signs and Prevent of CoV

Regular signs of infection of CoV include respiratory symptoms, cough, fever and heavy breathing difficulties. If it becomes for severe, the infection can also cause pneumonia, SARS, kidney failure and even death. The typical recommendations to prevent CoV infections include regular hand washing, covering mouth and nose when coughing and sneezing, thoroughly cooking meat and eggs. Avoid close contact with anyone showing symptoms of respiratory illness such as coughing and sneezing.

1.2 How CoV effected

Sixteen cities in China, with a amalgamate population of more than 50 million people, are on lockdown. Airlines over the globe have cancelled flights to and from China, various countries moving out their citizens on special flights and further placing them under firm quarantine. To make this more substandard, stock markets have plunged in China and markets across the world are affected by this scenario, few analysts predict that this outbreak of CoV is a threat to the global economy and it is capable of many more geopolitical consequences.

2. DESCRIPTION OF DATASET

The "Novel Corona Virus Dataset", published on Kaggle, has been used for analysis in this paper, which is collected by the John Hopkins University. A team has collected the data from various sources like WHO, local CDC and media outlets. They have also created a real-time dashboard which monitors the spread of the virus over the globe. This dataset contains daily information of the number of affected cases, deaths of people and recovery. Please note that the number of cases on any given day is the cumulative number, and this data has been made available from 22 January, 2020.

A total of 1719 observations and 8 columns are present in this dataset.

Table -1: Description of Data

Column	Description	Data
S no	Serial Number	Numerical
Date	Date Time-Observation	Date&Time
Province	Province-Observations	Categorical
Country	Country of Observation	Categorical
Last Update	Time in UTC	Date&Time
Confirmed	No of Confirmed Cases	Numerical
Deaths	No of deaths	Numerical
Recovered	No of recovered cases	Numerical

3. DATA PREPARATION

The first column 'Sno' is just a row number hence it doesn't add any value to the analysis. The 'Last Update' column shows the same labels as the 'Date' Column for most of the cases, so we removed these two attributes. Except for the 'Province/State' Column, none of the columns have null values and further analysis shows that names of provinces are missing for some countries. In this case, we cannot assume or fill missing values from any master list. By using duplicated() method we can conclude that all the observations in the dataset are unique. The Data shows that the virus has spread to 34 countries across Asia, Europe and America. To make the data uniform we will extract the dates from the timestamp and we will use them for analysis. Since the data is cumulative, we need to use the max function with group by in order to get the maximum number of reported cases for each country and the data confirms that China has

the most number of reported cases and nearly all of the 1789 deaths. On a positive note, China also has 7862 recoveries, followed by Singapore which has 24.

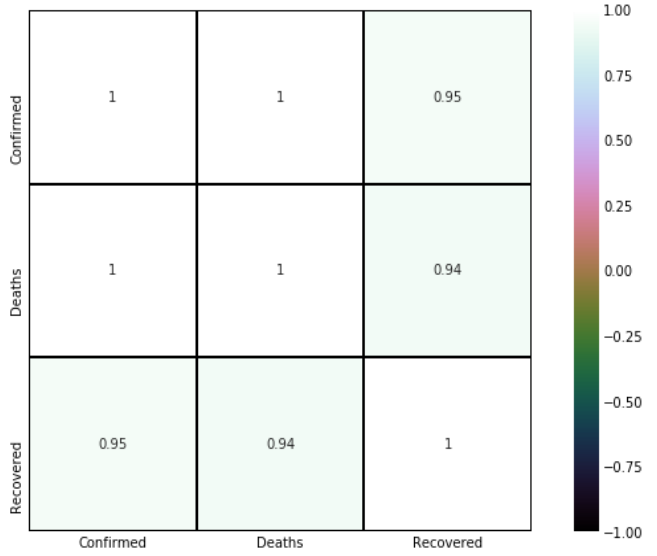

Chart -1: Correlation Matrix

A Correlation matrix is used to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses

Table -2: Country wise Data of the Confirmed cases, Deaths and recovered people

Country	Confirmed	Deaths	Recovered
Australia	5	0	4
Belgium	1	0	1
Brazil	0	0	0
Cambodia	1	0	1
Canada	5	0	1
China	59989	1789	7862
Egypt	1	0	0
Finland	1	0	1
France	12	1	4
Germany	16	0	1
Hong Kong	60	1	2
India	3	0	3
Italy	3	0	0
Ivory Coast	0	0	0
Japan	66	1	12
Macau	10	0	5
Malaysia	22	0	7
Mexico	0	0	0
Nepal	1	0	1
Philippines	3	1	1
Russia	2	0	2
Singapore	77	0	24
South Korea	30	0	10
Spain	2	0	2
Sri Lanka	1	0	1
Sweden	1	0	0
Taiwan	22	1	2
Thailand	35	0	15
UK	9	0	8
US	3	0	2
UAE	9	0	4
Vietnam	16	0	7
Others	454	0	0

4. DATA VISUALIZATION

For Visualization of data, we will use Matplotlib and Seaborn which are the two most powerful python libraries. Matplotlib is the popular 2D visualization library in python. Seaborn, which is built on top of matplotlib, helps to build better looking and more complex visualizations like heatmaps.

4.1 Most affected provinces in China

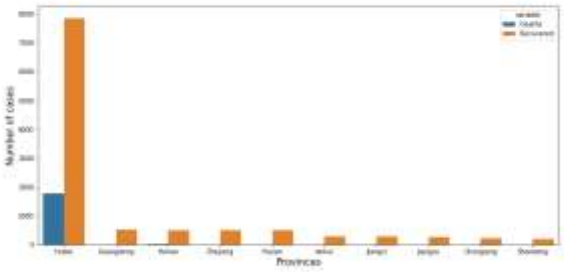

Fig -1: Most affected provinces in China

The Chinese province of Hubei is the epicenter of the outbreak. It has significantly more reported cases than all the other provinces combined. There are some provinces where there have been no deaths and all affected patients have recovered.

4.2 Mortality

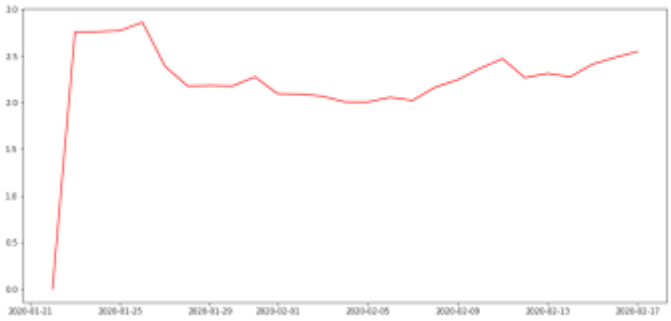

Fig -1: Mortality over time

The mortality rate has never crossed 3% and is gradually reducing to 2.5%. More recoveries in the coming weeks might reduce this further.

4.3 Most affected countries

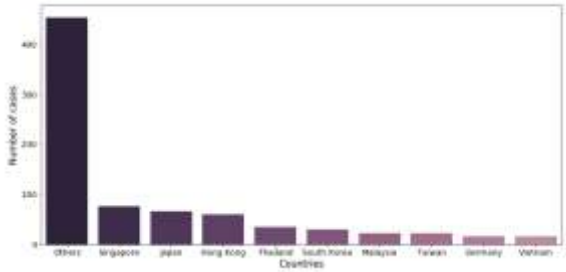

Fig -3: Most affected countries, besides China

Countries which are geographically close to China, like Singapore, Japan, Hong Kong and Thailand, have reported more cases than other Asian and European countries. Germany is an exception and has the highest number of cases in Europe.

4.5 Deaths vs. Recoveries

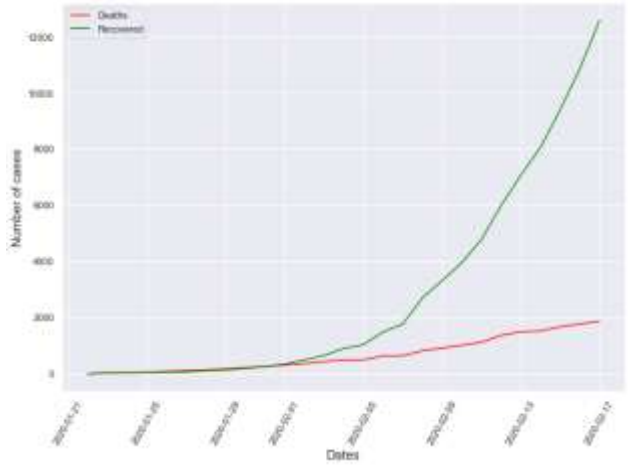

Fig -4: Rate of deaths vs rate of recovery

During the first few weeks, the rate of deaths was the same as that of recoveries. Since the 1st of February, the rate of recovery has shot up and is showing a positive trend. There was 12583 recoveries on the 17th of February compared to 1868 deaths. The recovery rate will continue to increase as more people get to know the symptoms and are prompt in seeking medication.

4.5 Confirmed Cases

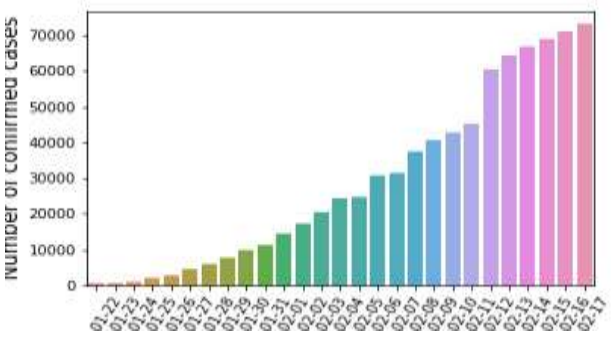

Fig -5: Number of confirmed cases over time

The number of cases reporting daily has been increasing drastically by nearly 700% since the 28th January. The number of cases reported on the 17th of February was 2034. This shows that the virus is highly contagious and is spreading rapidly.

4.6 Recoveries vs. Deaths vs. Confirmed cases

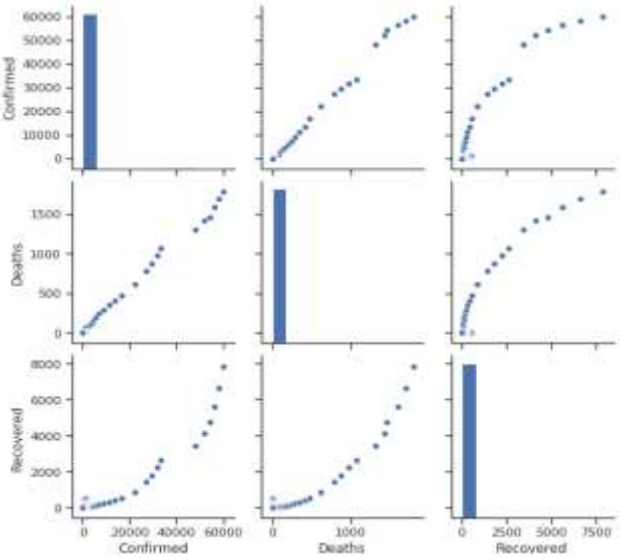

Fig -6: Recoveries vs. Deaths vs. Confirmed

Maximum number of Confirmed cases has been recovered, while few deaths taken place even when treated with medication

6. CONCLUSION

The analysis shows the alarming rate which the Wuhan coronavirus is spreading. At least 2200 people have now died during the epidemic, and continuing, exceeding the 774 fatalities reported during the SARS outbreak seven years ago.

ACKNOWLEDGEMENT

I wish to express my sincere appreciation to my mentor,

Dr S Vairachilai and my friend C Vamshidhar Reddy, who convincingly guided and encouraged me to be professional and do the right thing. Without their persistent help and support, the goal of this paper would not have been realized.

REFERENCES

[1] Epidemiological and Clinical Features of 197 Patients Infected with 2019 Novel Coronavirus in Chongqing, China: A Single Center Descriptive Study, *Preprints with The Lancet* (Feb 19, 2020)

[2] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[3] Utilize State Transition Matrix Model to Predict the Novel Corona Virus Infection Peak and Patient Distribution, *Preprints with The Lancet* (Feb 19, 2020)

[4] Understanding Unreported Cases in the 2019-Ncov Epidemic Outbreak in Wuhan, China, and the Importance of Major Public Health Interventions, *Preprints with The Lancet* (Feb 4, 2020)

[5] Liu, Z., Magal, P., Seydi, O. and Webb, G., 2020. Understanding Unreported Cases in the 2019-Ncov Epidemic Outbreak in Wuhan, China, and the Importance of Major Public Health Interventions. China, and the Importance of Major Public Health Interventions (February 3, 2020).

[6] Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, Wong BH, Poon RW, Cai JJ, Luk WK, Poon LL. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. Journal of virology. 2005 Jan 15;79(2):884-95.

[7] Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, Yuen KY. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerging Microbes & Infections. 2020 Jan 1;9(1):221- 36.

[8] Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DG. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance. 2020 Jan 23;25(3):2000045.

[9] World Health Organization. Novel Coronavirus—Japan (ex-China). Available online: https://www.who.int/csr/don/16-january-2020-novel-coronavirus-japan-exchina/en/ (accessed on 9 Feb. 2020).

[10] Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, Mchugh TD, Memish ZA, Drosten C, Zumla A. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. International Journal of Infectious Diseases. 2020;91:264-6.

[11] Gralinski LE, Menachery VD. Return of the Coronavirus: 2019-nCoV. Viruses. 2020 Feb;12(2):135.

[12] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P. A novel coronavirus from patients with pneumonia in China, 2019. New England Journal of Medicine. 2020 Jan 24..

BIOGRAPHIES

S SHASHANK RAJ
Final Year Student at ICFAI University, Hyderabad
Majoring in Computer Science