IJCRR
Section: Healthcare
ISI Impact Factor
(2019-20): 1.628
IC Value (2019): 90.81
SJIF (2020) = 7.893

Copyright@IJCRR

Exploratory Data Analysis on the Epidemiology of Coronavirus (COVID-19) Pandemic Outbreak

Ashima Tyagi¹, Deepak Chahal², Latika Kharb²

¹Research Scholar, MNNIT, Allahabad, India; ²Professor, Jagan Institute of Management Studies (JIMS), Rohini, New Delhi, India.

ABSTRACT

Severe Acute Respiratory Syndrome (SARS), first identified in the Guangdong province of China in 2002, affected the world. The transmission of the virus is from animals to animals and also to humans. In late 2019, another novel coronavirus SARS-CoV2 was identified, which became the cause of the coronavirus disease 2019 (COVID-19) that originated from Wuhan, China and quickly spread worldwide, leading to millions of cases and thousands of deaths. The data is extracted from the google sheets associated with a dashboard created by Johns Hopkins University that depicts the countries in the world that have coronavirus infection. The dataset, available from January 22 2020, consists of daily updated information on the infected countries along with the deaths and recoveries from the COVID-19 virus. We showed the effect of COVID-19 worldwide, representing the countries that are majorly affected by the infection and scrutinizing coronavirus trends in India. The techniques and measures that can be adapted to control the pandemic are also discussed in this study. The coronavirus has affected 213 countries around the world till now. The analysis is done using Python Programming Language, dated January 22, 2020 to August 1, 2020, while the virus continues to spread. The paper scrutinizes trends of covid-19 in India by discussing the confirmed, recovered and death cases. Results suggest that coronavirus infection is increasing greatly in Maharashtra, Tamil Nadu, Andhra Pradesh and Delhi. It can be concluded that even though the infection originated from China, the country is successful in its attempt to decrease the transmission of the virus.

Key Words: Coronavirus, COVID-19, Pandemic, Python language, Severe acute respiratory syndrome (SARS)

INTRODUCTION

Severe Acute Respiratory Syndrome (SARS), first identified in the Guangdong province of China in 2002.¹ In the past two decades, Severe Acute Respiratory Syndrome (SARS)-2003 and Middle East Respiratory Syndrome (MERS)-2012, have caused large-scale pandemics and it was thought that SARS-CoV which is majorly found in bats could lead to another epidemic in future. In late 2019, the outbreak of Coronavirus disease or COVID-19 first reported in Wuhan, China, became a pandemic in just a few months, infecting millions of people worldwide.² In late 2019, another novel coronavirus SARS-CoV2 was identified, which became the cause of the coronavirus disease 2019 (COVID-19) that originated from Wuhan, China and quickly spread worldwide, leading to millions of cases and thousands of deaths.³ The plague, which started on December 12, 2019, caused millions of infections, including thousands of deaths. It is unclear about how the

virus spread in humans, but some reports suggest that it came from the animal market in Wuhan, China.

Coronaviruses are a broad family of viruses that cause an ailment in animals or humans. In humans, it causes respiratory infections. People infected with the virus show symptoms like dry cough, high fever, tiredness, and difficulty breathing. Older people and people with previous medical problems like respiratory diseases, diabetes, cardiovascular diseases, and cancer are more vulnerable to the virus. One can catch the virus from another infected person through small droplets from the nose or mouth while coughing or exhaling and also by touching the surface where the droplets have been landed.

At this time, no vaccine or specific treatment is available for COVID-19, but clinical trials are going on. This paper analyzes the trends of the coronavirus cases along with the deaths and the recoveries all over the world. The study also

Corresponding Author:

Latika Kharb, Professor, Jagan Institute of Management Studies (JIMS), Rohini, New Delhi, India.

Email: latika.kharb@jimsindia.org

ISSN: 2231-2196 (Print)

ISSN: 0975-5241 (Online)

Received: 11.01.2020

Revised: 02.02.2021

Accepted: 24.02.2021

Published: 30.03.2021

discusses the timeline of coronavirus pandemic in India. Furthermore, the paper presents some techniques that the affected countries can adopt to curb the transmission of the virus, followed by the measures to control the disease.

DATASET

The dataset used for analysis has been taken from the repository of Kaggle.⁴ The data is extracted from the google sheets associated with a dashboard created by Johns Hopkins

University that depicts the countries in the world that have coronavirus infection. The dataset, available from 22 January 2020, consists of daily updated information on the infected countries along with the deaths and recoveries from the COVID-19 virus. The analysis is done using Python Programming Language, dated 22 January 2020 to 1st August 2020, while the virus continues to spread.³ Table 1 depicts the total coronavirus cases in different countries from January 22nd to 1st August. The data is processed for the exploratory analysis of coronavirus in different countries.

Table 1: Dataset of COVID-19

	Date	State/Province	Country	Confirmed	Deaths	Recovered
0	01/22/2020	Anhui	Mainland China	1	0	0
1	01/22/2020	Beijing	Mainland China	14	0	0
2	01/22/2020	Chongqing	Mainland China	6	0	0
3	01/22/2020	Fujian	Mainland China	1	0	0
4	01/22/2020	Gansu	Mainland China	0	0	0
...	...	...	...	...	...	...
77462	08/01/2020	Zaporizhia Oblast	Ukraine	800	21	597
77463	08/01/2020	Zeeland	Netherlands	867	69	0
77464	08/01/2020	Zhejiang	Mainland China	1270	1	1268
77465	08/01/2020	Zhytomyr Oblast	Ukraine	1896	37	1485
77466	08/01/2020	Zuid-Holland	Netherlands	13262	1311	0

COVID-19 IN WORLD

More than 200,000 cases of the infection are recorded daily since the pandemic began. Coronavirus continues to affect the lives of people by multiplying itself rapidly in humans. The virus has forced many countries to implement lockdown to save the lives of the people. COVID-19 cases are almost 18 million in the world, causing more than 650,000 deaths. All the 213 countries and territories, along with the two international conveyances have been analyzed for the novel coronavirus.

Total coronavirus cases in the world

The unknown virus originated from China, gradually affected almost all the countries of the world as cited in Figure 1 and 2. Within a few days, there were thousands of cases worldwide, leading to hundreds of deaths. The python code shown below depicts the total number of coronavirus cases worldwide, along with the total deaths and full recoveries from the virus.⁵

According to figure 2, the total number of confirmed cases of coronavirus in the world is 17,860,311, causing the total death toll to 682,773. The recovered instances around the globe are 10,553,550, while the active cases are more than 6 million, which is predicted to grow in the coming months.

```
df = data.groupby(["Date", "Country"])[["Date", "Country", "Confirmed",
'Deaths', 'Recovered']].sum().reset_index()
sorted_By_Confirmed=df.sort_values('Confirmed', ascending=False)
sorted_By_Confirmed=sorted_By_Confirmed.drop_duplicates('Country')

world_Confirmed_Total=sorted_By_Confirmed['Confirmed'].sum()
world_Deaths_Total=sorted_By_Confirmed['Deaths'].sum()
world_Recovered_Total=sorted_By_Confirmed['Recovered'].sum()

Active=world_Confirmed_Total-world_Deaths_Total-world_Recovered_Total

veri={'Total Confirmed cases in the world':world_Confirmed_Total,
'Total Deaths cases in the world':world_Deaths_Total,
'Total Recovered cases in the world':world_Recovered_Total,
'Total Active Cases':Active}
veri=pd.DataFrame.from_dict(veri, orient='index', columns=['Total'])
print("27/03/2020")
veri.style.background_gradient(cmap='Greens')
```

Figure 1: Python code for total cases worldwide.

	Total
Total Confirmed cases in the world	17860311
Total Deaths cases in the world	682773
Total Recovered cases in the world	10553550
Total Active Cases	6623988

Figure 2: Worldwide cases of coronavirus.

Country-Wise Confirmed COVID-19 cases

Since the first case was discovered in China, the infection spread like wildfire and affected almost all the countries of the world within a few months. The year 2020 experienced the deadliest outbreak in history, causing millions of deaths

worldwide. The code shown in Figure 3 visualizes the spread of coronavirus in the world by illustrating those countries that are severely affected by the virus.⁶

```
import seaborn as sns
import matplotlib.pyplot as plt
sorted_By_Confirmed1=sorted_By_Confirmed.head(13)
x=sorted_By_Confirmed1.Country
y=sorted_By_Confirmed1.Confirmed
plt.rcParams['figure.figsize'] = (12, 8)
sns.barplot(x,y,order=x,palette="Paired").set_title
('Total Cases Worldwide')
```

Figure 3: Python code for visualizing confirmed cases in the world.

The histogram shown below illustrates the country-wise spread of the coronavirus infection. United States of America, Brazil, India, and Russia are the top 4 countries that are severely hit by the COVID-19 pandemic. The cases are proliferating in these countries, causing complete or partial lockdown in most of the places. Figure 4 depicts the highest number of confirmed cases, i.e., 4,620,444 are in the United States (US). Total Confirmed cases in Brazil are 2,707,877, while India has 1,750,723 followed by Russia with 843,890 cases. Coronavirus cases in China are comparatively less, with a value of 84,385. China being the origin of the virus, does not stand even in the top 10 countries that are affected by the virus.

Figure 4: Country-wise confirmed cases.

Recovery rate and Death rate

A patient suffering from coronavirus shows the symptoms after 14 days. The other important finding has been the reported cases of infectivity despite lack of symptoms.⁵ After getting diagnosed with the infection, the medications start and recovery is intended. The recovery from the virus depends on various factors like previous medical problems, immune systems, and healthy intake. In the cases where the strength of the virus is low, there is a high chance of the patient getting recovered quickly. But, in situations where the virus has affected the lungs and other body parts, the recovery

takes time and even may lead to death. Python code shown below in Figure 5 calculates the death rate and the recovery rate for all the countries that are infected with the virus. Table 2 below depicts the recovery rate, death rate and total cases rate for all the 213 countries.^{6,7}

```
Recovered_rate=(sorted_By_Confirmed['Recovered']*100)/sorted_By_Confirmed['Confirmed']
Deaths_rate=(sorted_By_Confirmed['Deaths']*100)/sorted_By_Confirmed['Confirmed']
cases_rate=(sorted_By_Confirmed.Confirmed*100)/world_Confirmed_Total

sorted_By_Confirmed['Active']=sorted_By_Confirmed['Confirmed']-
sorted_By_Confirmed['Deaths']-sorted_By_Confirmed['Recovered']
sorted_By_Confirmed['Recovered_Cases_Rate']=pd.DataFrame(Recovered_rate)
sorted_By_Confirmed['Deaths_Cases_Rate_%']=pd.DataFrame(Deaths_rate)
sorted_By_Confirmed['Total_Cases_Rate_%']=pd.DataFrame(cases_rate)
print("Sorted By Confirmed Cases")
sorted_By_Confirmed.style.background_gradient(cmap="Reds")]
```

Figure 5: Calculating the recovery rate and death rate.

Table 2: Recovery rate and Death rate

Date	Country	Confirmed	Deaths	Recovered	Active	Recovered_Cases_Rate	Deaths_Cases_Rate_	Total_Cases_Rate_
28748	US	4620444	154447	1461885	3004112	31.6395	3.34289	25.8699
28594	Brazil	2707877	93563	2637962	575332	75.2612	3.45522	15.1814
28649	India	1750723	37364	1145829	567730	65.4375	2.1342	9.80231
28712	Russia	843890	14034	645316	184540	76.4682	1.66301	4.72495
28726	South Africa	503290	8153	342461	152676	68.0445	1.61994	2.81792
28685	Mexico	434193	47472	334067	51854	77.124	10.9334	2.43105
28706	Peru	407492	19021	287127	101344	70.462	4.66782	2.28155
28606	Chile	357656	9533	330507	17618	92.4067	2.6654	2.00253
22857	UK	314992	43991	1372	269629	0.435567	13.9655	1.76364
28651	Iran	306752	16982	265830	23940	86.6596	5.53807	1.71751
28607	Colombia	306101	10330	160708	135143	52.4879	3.37382	1.71431

The histogram presented below illustrates the recovery rate and the death rate in 10 countries that are majorly affected. It is observed that the recovery rate for Brazil, India, Russia and South Africa above 65% while that of the United States is very less. Despite having the largest number of confirmed cases, the recovery rate of the United States is as low as 31%. The cases in the US are increasing at a high rate, but the recovery rate is flat. It can be concluded Chile has the highest recovery rate, with 92.4%.

Figure 6 also depicts that the death rate for the US and Brazil is around 3% that is less than that death rate in the UK and Mexico, which is more than 10%. India's death rate is relatively low, with a value of 2.1%, but it is expected to grow in the upcoming days due to the significant increase of the cases in India.

Figure 6: Coronavirus Recovery and Death Rate (%) for 11 countries.

COVID-19 IN INDIA

India, the second most populated country in the world, reported its first coronavirus case on 30th January 2020 in Kerala. The affected person had a travel history from Wuhan, China. On 2nd March 2020, a 45-year old man was then reported positive in Delhi, who had traveled from Italy in the past days, and a 24-year old engineer who moved from the United Arab Emirates was also tested positive. Since then, coronavirus cases never stopped increasing, and within a few months, there were thousands of COVID-19 cases. The infection affected almost all the states of India, causing thousands of deaths in the country.⁷⁻⁹

The bar chart shown in Figure 7 tells about the timeline of coronavirus cases in India. The total number of confirmed cases in India by 1st April was 1,998, which went to 37,257 by 1st May. The infection spread itself so fast that by 1st June, there were more than two lakhs cases. On 16th July 2020, India experienced a gradual increase in the virus with the highest growth of cases per day, i.e., 25,602 cases in one day, leading to a total of 1 million infections. Till 1st August, there were more than 17 million confirmed cases in India.

Figure 7: Coronavirus confirmed cases in India from 22/3/2020 to 01/08/2020.

Timeline of confirmed, recovered and death cases in India

The deadly virus multiplied itself expeditiously, causing millions of cases. Gradually, a more significant number of deaths were encountered, but the recovery rate was increased. The following python code in Figure 8 visualizes the confirmed, recovered and death cases in India.^{9,10}

Line graph shown in figure 9, clearly signifies that the recovered cases in India increase with the increase in the confirmed cases. On 1st June, the recovered cases were 95754, and by 1st August, more than 11 million cases have been recovered from the infection. The death rate also increased but at a slower pace. On 1st May, the total reported deaths in

India were 1,223, which grew to 17,834 by 1st July and by 1st August, the total death toll reached 37,364.

```
x=datetime
y=confirmed
y1=deaths
y2=recovered
# Initialize the matplotlib figure
fig, ax = plt.subplots(figsize=(10,10))

fig.autofmt_xdate()
plt.plot(x,y,color='magenta',label='Confirmed Cases')
plt.plot(x,y1,color='red',label='Deaths Cases')
plt.plot(x,y2,color='green',label='Recovered Cases')
plt.title('Increase in Coronavirus cases in India.')
plt.xlabel('Date')
plt.ylabel('Total Cases')
ax.legend(ncol=2, loc='upper left', frameon=True)

plt.show()
```

Figure 8: Python code for total cases in India.

Figure 9: Confirmed, recovered and death cases in India.

Coronavirus Pandemic in the Indian States

The epidemic of coronavirus disease-2019 (COVID-19) establishes a medical emergency of worldwide concern with an exceptionally high danger of spread and affect the entire worldwide.⁶ The virus that began from Kerala slowly spread itself to other states of India. Before the beginning of lockdown in India, the infection already affected most of the states. After Kerala and Delhi reported their first case, the virus-infected people from Jaipur, Hyderabad and slowly it multiplied itself to almost all the states of India. In India, too, the pandemic's spread and impact have been very diverse across different states.⁷ The code below in Figure 10 is used to analyze Indian states for the confirmed cases and recovered cases from coronavirus pandemic.

```
import matplotlib.pyplot as plt
import seaborn as sns
f, ax = plt.subplots(figsize=(10, 5))
sns.set_color_codes("pastel")
sns.barplot(x="Confirmed", y="State/UnionTerritory", data=states_10,
            label="Confirmed", color="b")

# Plot the cured cases
sns.set_color_codes("muted")
sns.barplot(x="Cured", y="State/UnionTerritory", data=states_10,
            label="Recovered", color="b")

ax.set(xlabel='Total Confirmed and Recovered Cases in India')
# Add a legend and informative axis label
ax.legend(ncol=2, loc="lower right", frameon=True)

sns.despine(left=True, bottom=True)
```

Figure 10: Python code for determining cases in different states.

From the bar graph shown in Figure 11, it can be observed that Maharashtra has the highest number of coronavirus cases, i.e., 422,118, followed by Tamil Nadu, Andhra Pradesh and Delhi with confirmed cases of 245,859, 140,933, 135,598 respectively. It can also be noted that Delhi shows a higher recovery rate with total recovered cases of 120,930.^{6,9}

Figure 11: Confirmed and recovered cases for Indian states.

TECHNIQUES ADOPTED BY CHINA TO FIGHT THE PANDEMIC

The recovery rate for China is the highest in the world. China is highly successful in controlling the crisis. The three techniques that the Chinese Government has adopted to contain the virus can be used in other countries to curb the pandemic and save the lives of the people.

Mass Surveillance and Big Data

In December 2019, a novel virus named COVID-19 emerged in the city of Wuhan, China.⁸ The Chinese government collects an immense amount of data about its citizens, and they are using it in novel ways to control the coronavirus. Every time a person leaves the apartment complex, she has to scan her face. Those images, as well as the images captured in the CCTV cameras (around 200 million), installed everywhere, are fed into the central database and then analyzed using machine learning as shown in Figure 12. If you are inside a camera, it is likely that the Chinese government has analyzed your face and knows where you are.

Figure 12: Face being Scanned by the machines and Cameras installed everywhere to capture people.

Another significant source of data is the WeChat application that shares accurate information on viruses on the internet.

The app is linked to the debit cards of each person and is bound to use it while buying items. This data is used to try the highest tech epidemic control ever attempted.

Whenever a person visits the hospital, the hospital will register her with the authorities; the authorities will pull the name of the person from the big database. They have an algorithm that will be able to tell all the different places that the person has been in the last 14 days and all the people with whom she had close contact with. So, for example, a person bought bread with the WeChat app, the baker will be notified that the person has to self-quarantine for 14 days and the people that were next to her on the flight have to self-isolate for 14 days. Every single citizen has a code red, amber or green, and automatic color coding determines where you can go. In some cities, you cannot drive on the road if you aren't green (Figure 13).

Figure 13: App showing the person is Green and safe.

Complete Lockdown

Cities in China were in lockdown for months. Your temperature is checked, entering most buildings. People are not allowed to eat in restaurants. Every single person that comes to China has to self-quarantine for two weeks. Any flight coming into the country is screened. Only one person from each household is allowed to go out every three days to buy essential items (Figure 14). The city Hubei in China also banned all vehicles from the road, and one can only drive with a special permit.⁷⁻¹⁰

Figure 14: Checking the temperature before entering to a Complex.

Promoting Solidarity

COVID-19 is more infectious compared to other known viruses (such as SARS and MERS) albeit with a lower mortality rate.⁹ Complete lockdown and tracing those with the virus with big data and the glue that holds it all together is the intense promotion of solidarity. The government has framed it as a people's war against the virus. The doctors had cut their hair because there isn't time to have a shower. These messages of solidarity can be seen on social media and on the screens that are installed in every public place.

NEW TECHNOLOGY TO TEST THE VIRUS

Abbott Laboratories, American medical devices and health care company headquartered in Illinois, United States, unveiled a COVID-19 test that can identify if a person is infected with the coronavirus in just 5 minutes. Being very small and portable, it can be used to diagnose loads of infections. The test is done with the ID Now equipment (Figure 15) that weighs 3.18 kg.

Figure 15: Abbott's ID Now System.

All the suspected patients were tested by RT-PCR. The technology works by mixing a nasal or throat swab with a chemical solution, which is then inserted into the ID Now equipment. The ID Now identifies and amplifies sequences of the coronavirus genome by ignoring other virus contaminations. It takes 5 minutes for a positive test, while the technology can take up to 13 minutes for checking a negative result to rule out an infection.^{9,10}

SIMULATING AN EPIDEMIC

To eradicate the virus, it is essential to break the chain of the coronavirus, meaning that human to human transfer of the virus has to be stopped. However, several technical issues remain unsolved and further research is needed⁹ from time to time. Several steps can be taken by all the affected countries to tackle the epidemic. Some of which are:

- **Identify and Isolate:** It states that the person infected with the virus must be identified and should be isolated immediately.
- **Social Distancing:** Maintaining a social distance of up to 1 meter can reduce the chances of getting infected with the virus.
- **Travel restrictions:** Strict constraints on traveling to places should help break the chain of the coronavirus. Rigorous checking on the airports and railway stations can do much better in curbing the transmission of the virus.
- **Proper Hygiene:** Good hygiene is the best way to reduce the chances of getting infected with the virus. Activities like washing hands frequently, not touching face often, wearing a mask, etc. can save the person and his family too.^{8,9,10}

CONCLUSION

The paper discussed the coronavirus pandemic that has taken thousands of lives of the people around the world. The disease turned into an epidemic affecting many countries and taking millions of lives. The countries that are greatly affected by the coronavirus are United States, Brazil, India, Russia and South Africa. The study discusses the recovery rates for the countries affected by the virus. Results show that the recovery rate of the US is very low while that of India is relatively higher. The paper scrutinizes trends of covid-19 in India by discussing the confirmed, recovered and death cases. Results suggest that coronavirus infection is increasing greatly in Maharashtra, Tamil Nadu, Andhra Pradesh and Delhi. It can be concluded that even though the infection originated from China, the country is successful in its attempt to decrease the transmission of the virus. The techniques and the methods adopted by the Chinese Government are explained briefly that helped control the effect of the virus in the country. Few measures like social distancing, identifying the infected and isolating them, and so on can be taken into account to eradicate the infection.

ACKNOWLEDGEMENT

Authors acknowledge the immense help received from the scholars whose articles are cited and included in references of this manuscript. The authors are also grateful to authors / editors / publishers of all those articles, journals and books from where the literature for this article has been reviewed and discussed.

Conflict of interest: NIL

Financial support: NIL

REFERENCES

1. Hussain A, Kaler J, Tabrez E, Mishra AS, Jain NM, et al. Novel COVID-19: A Comprehensive Review of Transmission, Manifestation, and Pathogenesis. *Cureus* 2020;12(5): e8184.
2. Nagendra Kumar Rai, Anushruti Ashok & Butchi Raju Akondi. Consequences of chemical impact of disinfectants: safe preventive measures against COVID-19. *Crit Rev Toxicol* 2020;50(6):513-520.
3. Gagliardi I, Patella G, Michael A, Serra R, Provenzano M, Andreucci M. COVID-19 and the kidney: From Epidemiology to Clinical Practice. *J Clin Med* 2020;9(8):2506.
4. Novel Coronavirus Dataset: <https://www.kaggle.com/sudalairaj-kumar/novel-corona-virus-2019-dataset>
5. Ng BH, Nik Abeed NN, Abdul Hamid MF, Soo CI, Low HJ, Ban Y-LA. What happens when we treat the “Typhoid Mary” of COVID-19. *Resp Case Rep* 2020;8(6):e00604.
6. Bhati A, Jagetiya A. Prediction of COVID-19 Outbreak in India adopting Bhilwara Model of Containment 5th International Conference on Communication and Electronics Systems (ICES), Coimbatore, India. 2020;951-956.
7. Jalan J, Sen A. Containing a pandemic with public actions and public trust: the Kerala story. *Ind Econ Res* 2020;55:105-124.
8. Shuja J, Alanazi E, Alasmay W, Alashaikh A. COVID-19 open source data sets: a comprehensive survey. *Appl Intell* 2020. Preprint at medRxiv. doi: 10.1007/s10489-020-01862-6.
9. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, Seneviratne A, Hu W, Janicke H, Jha SK. A survey of COVID-19 contact tracing apps. *IEEE Access*. 2020;8:134577-601.
10. Kharb L, Singh R. Complexity metrics for component-oriented software systems. *Softw Eng Notes* 2008;33(2):1-3.