“COVID-19 STREAMLINE DISEASE STUDY APPROACH METHOD”

PROJECT REPORT

Submitted to:

JAIN – Deemed to be University, Bangalore, Karnataka

In partial fulfillment of the University Regulations for the Degree in

BACHELOR OF SCIENCES (HONOURS)

IN

DATA SCIENCE AND ANALYTICS

Submitted by:

Yatin Kande

20BSR18040

Department of Data Science and Analytics School of Sciences

Batch: 2020-2023

DEPARTMENT OF
DATA SCIENCE AND ANALYTICS

This is to certify that the project entitled

'COVID-19 STREAMLINE DISEASE STUDY APPROACH METHOD'

Is the approved record of the project done by:

YATIN KANDE (20BSR18040)

Under the guidance of Dr. UTHAMAKUMAR A

In the partial fulfillment of the requirement for the award of degree in

BACHELOR OF SCIENCE (HONOURS)

DATA SCIENCE AND ANALYTICS

from

Jain University, Bangalore.

STUDENT	GUIDE	EXAMINER
YATIN KANDE	DR. UTHAMAKUMAR A	
(20BSR18040)		

DECLARATION

I, **YATIN KANDE** hereby declare that this dissertation titled **" COVID-19 STREAMLINE DISEASE STUDY APPROACH METHOD"** has been the outcome of an original study carried out under the guidance of **Dr. Uthamakumar A** towards the partial fulfilment of the B.Sc. Data Science and Analytics degree of the JAIN (Deemed-to-be University) during the year 2023-2024. This study has not been submitted for any degree, diploma or certificate.

Place: Bangalore

Date: 20/04/2023

GUIDE

DR. UTHAMAKUMAR A

STUDENT

YATIN KANDE
20BSR18040

DEPARTMENT OF

DATA SCIENCE AND ANALYTICS

BONAFIDE

This is to certify that the internship report titled **"COVID-19 STREAMLINE DISEASE STUDY APPROACH METHOD"** is the bonafide work of **Yatin Kande (20BSR18040)** for the degree of Bachelor of Science (Honors) in Data Science and Analytics.

Certified further, that to the best of my knowledge the work reported here in does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this way or any other candidate.

HEAD OF THE DEPARTMENT

PROGRAM HEAD

ACKNOWLEDGEMENT

I would like to take this opportunity to express my gratitude and appreciation to Jain University for having deputed me for this assignment. I want to express my heartfelt gratitude to Dr. Asha Rajiv, Director, Jain University for giving me this opportunity to work on my mini project.

My sincere thanks are expressed to Dr. Arathi Sudarshan, Head of the Department, of Data Science and Analytics, for her constant guidance and encouragement.

I would also like to thank Dr. Uthama Kumar A, as my project mentor, for guiding me and helping in me for completion of the project assigned to me successfully during 6 months of tenure of my project work.

Yatin Kande
20BSR18040

TABLE OF CONTENTS

S.no	Title	Page no
1	Table of contents	6
2	List of figures	7
3	Abstract	9
4	History	9
5	Symptoms	10
6	Variants	11
7	Transmission	17
8	Test	17
9	Transmission Vaccination	19
10	Literature Review	25
11	Methodology	26
12	Data Collection	27
13	Chapter 1 Dataset explanation	28
14	Chapter 2 Data Analysis and visualizations	30
15	Chapter 3 Predictions	46
16	Conclusion	48
17	Future Scope	49
18	References	50

LIST OF FIGURES

Figure no.	Name of the figure	Page no.
1	Currently circulating variants of concern (VOCs)	14
2	Previously circulating VOCs	14
3	Omicron subvariants under monitoring	16
4	Viral vector vaccines	20
5	Genetic vaccines	21
6	Inactivated vaccines	23
7	Attenuated vaccines	24
8	Dataset	28
9	Distribution of number of Active Cases	30
10	Distribution of Number of Closes Cases	30
11	Weekly Growth of different types of Cases in India	31
12	Growth of different types of cases	32
13	Recovery Rate and Mortality Rate	33
14	Date wise Growth Factor of different Types of cases	34
15	Datewise Growth Factor of Active and Closed Cases	35
16	Top 20 Countries as per Confirmed and Death cases	35

17	Rate of Doubling for confirmed around the world	36
18	Top 15 countries with most number of active cases	37
19	Top 15 countries with most number of closed cases	37
20	Top 15 countries as per the high number confirmed cases per day	38
21	Top 15 countries as per lowest confirmed cases per day having more than 1000 confirmed cases	38
22	Top 25 countries with maximum survival probability having more than 1000 confirmed cases	39
23	Confirmed cases – Logarithmic scale	39
24	Weekly Progress of Different Types of Cases in India	40
25	Weekly Progress of Different Types of Cases in China	41
26	Weekly Progress of Different Types of Cases in US	41
27	Elbow Method	42
28	Dendogram	43
29	Gradient colour map	44
30	Cluster plot	45
31	Linear Regression	46
32	Support Vector Machine Regressor	47
33	Global CO2 emissions and intensity	48

ABSTRACT

To contain the spread of the COVID-19 pandemic, effective prediction, and forecasting are required. Data analysis and data visualization are discussed as solutions to these problems. Machine learning (ML) algorithms have shown promise in accurately predicting and forecasting COVID-19 spread. The presentation provides an overview of COVID-19, its transmission, and the available data sources. COVID-19 prediction and forecasting make use of a variety of ML techniques such as time series forecasting and Linear regression. Data quality, data scarcity, and model interoperable are among the challenges that ML faces in COVID-19 prediction. The potential impact of machine learning on COVID-19 prediction and forecasting is discussed, including improving prediction accuracy and insights on assisting or decision-making processes.

HISTORY

The COVID-19 (Coronavirus Disease 2019) virus causes an infectious respiratory illness. In December 2019, the first cases of COVID-19 were reported in Wuhan, China, and the virus quickly spread to other parts of China and then to other countries.

The World Health Organization (WHO) declared COVID-19 a global health emergency in January 2020, and by March 2020, the outbreak had become a pandemic, with cases reported in nearly every country on the planet.

The virus is thought to have originated in a wet market in Wuhan, where live animals were sold for food. The virus is thought to have begun in bats and spread to humans via an intermediate animal host, possibly a pangolin.

The first COVID-19 cases in China were reported to health officials in December 2019, but it wasn't until January 7, 2020, that Chinese officials confirmed the virus was a new strain of coronavirus. The Chinese government tried to contain the outbreak by placing Wuhan and other cities under lockdown, but the virus spread anyway.

The virus has spread to other countries by the end of January, including India, Japan, South Korea, Thailand, and the United States. On January 30, 2020, the WHO declared a global health emergency and urged countries to take action to contain the virus.

As the virus spread, countries around the world began to implement measures to slow its spread, such as closing borders, prohibiting large gatherings, and mandating masks and physical separation. These policies had a significant impact on daily life, forcing many businesses to close and forcing people to work from home.

The pandemic had a significant impact on the global economy, with many countries experiencing significant economic downturns. To try to mitigate the economic impact of the pandemic, governments around the world implemented stimulus measures.

As the pandemic progressed, scientists around the world raced to develop COVID-19 vaccines and treatments. The first COVID-19 vaccine was approved for use in December 2020, and vaccination campaigns began in many countries around the world. Multiple vaccines have been developed and approved for use as of early 2022, providing hope for an end to the pandemic.

However, the pandemic continues to have a global impact, with many countries still experiencing high levels of infection and death. New strains of the virus have emerged, some of which are more contagious and potentially lethal than previous strains. The pandemic's ongoing nature has highlighted the importance of global cooperation and the need for robust public health systems to respond to it.

SYMPTOMS

The COVID-19 (Coronavirus Disease 2019) virus causes an infectious respiratory illness. COVID-19 symptoms can range from mild to severe and appear anywhere between 2 and 14 days after virus exposure.

The following are the most common COVID-19 symptoms:

1. Fever: A temperature above 100.4°F (38°C) is considered a fever.
2. Cough: The most common type of cough associated with COVID-19 is a dry cough.
3. Breathing difficulties or shortness of breath: This could be a sign of severe COVID-19, especially if it worsens over time.
4. Fatigue: Even after resting, feeling tired or exhausted is a common symptom of COVID-19.
5. Muscle or body aches: Body aches and pains are also common symptoms, especially in the early stages of the illness.
6. Taste or smell loss: This is a less common symptom, but it has been reported by some COVID-19 patients.
7. Sore throat: COVID-19 can cause a sore throat, especially in the early stages of the illness.
8. Congestion or a runny nose and throat: While less common than other symptoms, COVID-19 can cause congestion or a runny nose.
9. Headache: Some COVID-19 patients report headaches as a symptom.

10. Nausea or vomiting: These symptoms, especially in severe cases, can be associated with COVID-19.
11. Diarrhoea: Although this is a less common symptom, it has been reported by some COVID-19 patients.

It is important to note that not everyone infected with COVID-19 will show all of these symptoms, and some people may not show any symptoms at all whiles still transmitting the virus to others.

VARIANTS

Alpha (B.1.1.7 lineage)

- In late December 2020, a new SARS-CoV-2 variant of concern, B.1.1.7 lineage, also known as Alpha variant or GRY (formerly GR/501Y.V1), was reported in the United Kingdom based on whole-genome sequencing of samples from SARS-CoV-2 positive patients.

- The B.1.1.7 variant was identified in a commonly used commercial assay characterized by the absence of the S gene (S-gene target failure, SGTF) PCR samples, in addition to being detected by genomic sequencing. The viral genome of the B.1.1.7 variant contains 17 mutations. The spike (S) protein has eight mutations (69-70 deletions, 144 deletions, N501Y, A570D, P681H, T716I, S982A, D1118H). N501Y increases the spike protein's affinity for ACE 2 receptors, enhancing viral attachment and subsequent entry into host cells.

- This type of worry was circulating in the United Kingdom as early as September 2020, and it was based on various model projections. It was reported to be 43% to 82% more transmissible, surpassing previous SARS-CoV-2 variants to become the dominant SARS-CoV-2 variant in the UK. At the end of December 2020, the B.1.1.7 variant was reported in the United States (US).

- An initial matched case-control study found no significant difference between the B.1.1.7 lineage variant and other existing variants in the risk of hospitalization or associated mortality. However, subsequent studies have found that people infected with the B.1.1.7 lineage variant had a higher severity of disease than people infected with other circulating virus variants.

- A large, matched cohort study conducted in the United Kingdom found that patients infected with the B.1.1.7 lineage variant had a mortality hazard ratio of 1.64 (95% confidence interval 1.32 to 2.04, P0.0001) patients with previously circulating strains. Another study found that the B 1.1.7 variant was linked to an increased risk of death when compared to other SARS-CoV-2 variants (HR= 1.61, 95% CI 1.42-1.82). Individuals with confirmed B.1.1.7 variant of concern had a higher risk of death (adjusted hazard ratio 1.67, 95% CI 1.34-2.09) than those with non-1.1.7 SARS-CoV-2.

- During the early stages of the pandemic, the B.1.1.7 variant emerged as one of the most prevalent SARS-CoV-2 strains circulating in the United States.

Beta (B.1.351 lineage)

- Another SARS-CoV-2 variant is B.1.351, also known as the Beta variant or GH501Y. The second wave of COVID-19 infections was caused by V2 with multiple spike mutations, which was first detected in South Africa in October 2020.
- The B.1.351 variant contains nine mutations in the spike protein (L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, and A701V), three of which are located in the RBD and increase the binding affinity for the ACE receptors. At the end of January 2021, SARS-CoV-2 501Y.V2 (B.1.351 lineage) was reported in the United States.
- This variant has been linked to an increased risk of transmission and a decreased ability to be neutralized by monoclonal antibody therapy, convalescent sera, and post-vaccination sera.

Gamma (P.1 lineage)

- A third of concern, the P.1 variant, also known as the s Gamma variant or GR/501Y.V3, was discovered in Brazil in December 2020 and was first detected in the United States in January 2021.
- The spike protein in the B.1.1.28 variant has ten mutations (L18F, T20N, P26S, D138Y, R190S, H655Y, T1027I V1176, K417T, E484K, and N501Y). The RBD contains three mutations (L18F, K417N, and E484K), which are like the B.1.351 variant.
- This variant may have reduced neutralization by monoclonal antibody therapies, convalescent sera, and post-vaccination sera.

Delta (B.1.617.2 lineage)

- B.1.617.2, also known as the Delta variant, was identified in December 2020 in India and was responsible for the deadly second wave of COVID-19 infections in April 2021 in India. This variant was first discovered in the United States in March 2021.
- The Delta variant was initially thought to be an interesting variant. This variant, on the other hand, spread rapidly around the world, prompting the WHO to classify it as a VOC in May 2021.
- The B.1.617.2 variant harbours ten mutations (T19R, (G142D*), 156del, 157del, R158G, L452R, T478K, D614G, P681R, D950N) in the spike protein.

Omicron (B.1.1.529 lineage)

- The fifth variant of concern B.1.1.529, also known as the Omicron variant by the WHO, was discovered in South Africa on November 23, 2021, following an increase in the number of COVID-19 cases.
- Omicron was snappily honoured as a VOC due to further than 30 changes to the shaft protein of the contagion along with the sharp rise in the number of cases observed in South Africa. The reported mutations include T91 in the envelope, P13L, E31del, R32del, S33del, R203K, G204R in the nucleocapsid protein, D3G, Q19E, A63T in the matrix, N211del/ L212I, Y145del, Y144del, Y143del, G142D, T95I, V70del, H69del, A67V in the N-terminal sphere of the shaft, Y505H, N501Y, Q498R, G496S, Q493R, E484A, T478K, S477N, G446S, N440K, K417N, S375F, S373P, S371L, G339D in the receptor- binding sphere of the shaft, D796Y in the emulsion peptide of the shaft, L981F, N969K, Q954H in the heptad reprise 1 of the shaft as well as multiple other mutations in the non-structural proteins and shaft protein.
- According to preliminary modeling, Omicron has a 13-fold increase in viral infectivity and is 2.8 times more infectious than the Delta variant. Early reports also indicate that monoclonal antibodies such as Bamlanivimab and the Rockefeller University antibody C144 are likely to have reduced efficacy against the Omicron variant; however, REGN-COV2 (Casirivimab and Imdevimab) and the Rockefeller University antibody C135 are predicted to be effective against Omicron based on early modeling studies.
- The Spike mutation K417N (also seen in the Beta variant) combined with E484A is predicted to have a massively disruptive effect, increasing the likelihood of vaccine breakthroughs for Omicron.
- The Omicron (B.1.1.529) became the dominant VOC in many countries, and many subvariants were identified, including BA.1, BA.2, BA.3, BA.4, and BA.5.
- According to the CDC, the Omicron VOC is the most common SARS-CoV-2 variant in the United States.

Currently circulating variants of concern (VOCs):

WHO label	Pango lineage•	GISAID clade	Nextstrain clade	Additional amino acid changes monitored°	Earliest documented samples	Date of designation
Omicron*	B.1.1.529	GR/484A	21K, 21L, 21M, 22A, 22B, 22C, 22D	+S:R346K +S:L452X +S:F486V	Multiple countries, Nov-2021	VUM: 24-Nov-2021 VOC: 26-Nov-2021

(fig 1: Currently circulating variants of concern (VOCs))

Previously circulating VOCs:

WHO label	Pango lineage•	GISAID clade	Nextstrain clade	Earliest documented samples	Date of designation
Alpha	B.1.1.7	GRY	20I (V1)	United Kingdom, Sep-2020	VOC: 18-Dec-2020 Previous VOC: 09-Mar-2022
Beta	B.1.351	GH/501Y.V2	20H (V2)	South Africa, May-2020	VOC: 18-Dec-2020 Previous VOC: 09-Mar-2022
Gamma	P.1	GR/501Y.V3	20J (V3)	Brazil, Nov-2020	VOC: 11-Jan-2021 Previous VOC: 09-Mar-2022
Delta	B.1.617.2	G/478K.V1	21A, 21I, 21J	India, Oct-2020	VOI: 4-Apr-2021 VOC: 11-May-2021 Previous VOC: 7-Jun-2022

(fig 2: previously circulating VOCs)

<u>Omicron subvariants under monitoring (as of 9 February 2023):</u>

Pango lineage[#] (+ mutation)	GISAID clade	Nextstrain clade	Relationship to circulating VOC lineages	Spike genetic features	Earliest documented samples
BF.7*	GRA	22B	BA.5 sublineage	BA.5 + S:R346T	24-01-2022
BQ.1[$]	GRA	22E	BA.5 sublineage	BQ.1 and BQ.1.1: BA.5 + S:R346T, S:K444T, S:N460K	07-02-2022
BA.2.75[§]	GRA	22D	BA.2 sublineage	BA.2.75: BA.2 + S:K147E, S:W152R, S:F157L, S:I210V, S:G257S, S:D339H, S:G446S, S:N460K, S:Q493R reversion	31-12-2021
CH.1.1[§]	GRA	22D	BA.2 sublineage	BA.2.75 + S:L452R, S:F486S	27-07-2022

XBB[μ]	GRA	22F	Recombinant of BA.2.10.1 and BA.2.75 sublineages, i.e. BJ1 and BM.1.1.1, with a breakpoint in S1	BA.2+ S:V83A, S:Y144-, S:H146Q, S:Q183E, S:V213E, S:G252V, S:G339H, S:R346T, S:L368I, S:V445P, S:G446S, S:N460K, S:F486S, S:F490S	13-08-2022
XBB.1.5	GRA	23A	Recombinant of BA.2.10.1 and BA.2.75 sublineages, i.e. BJ1 and BM.1.1.1, with a breakpoint in S1	XBB + S:F486P (see rapid risk assessment)	05-01-2022
XBF	GRA		Recombinant of BA.5.2.3 and CJ.1 (BA.2.75.3 sublineage)	BA.5 + S:K147E, S:W152R, S:F157L, S:I210V, S:G257S, S:G339H, S:R346T, S:G446S, S:N460K, S:F486P, S:F490S	27-07-2022

(fig 3: Omicron subvariants under monitoring)

SARS-CoV-2 Variants of Interest (VOIs)

VOIs are defined as variants with specific genetic markers that have been linked to changes that may result in increased transmissibility or virulence, decreased neutralization by antibodies obtained through natural infection or vaccination, the ability to evade detection, or a decrease in the effectiveness of therapeutics or vaccination. Since the start of the pandemic, WHO has described eight variants of interest (VOIs), namely Epsilon (B.1.427 and B.1.429); Zeta (P.2); Eta (B.1.525); Theta (P.3); Iota (B.1.526); Kappa (B.1.617.1); Lambda (C.37); and Mu (B.1.621); and the Epsilon (B.1.427 and B.1.429), Eta (B.1.525), Iota (B.1.526), Kappa (B.1.617.1), Zeta (P.2), Mu (B.1.621, B.1.621.1), and B.1.617.3 have been designated as VOIs by the CDC.

TRANSMISSION

COVID-19 is primarily transmitted through respiratory droplets, which are produced when an infected person speaks, coughs, sneezes, or breathes. People nearby can inhale the droplets, or they can land on surfaces and objects, where they can remain infectious for hours or days.

COVID-19 can be transmitted in a variety of ways, including:

- COVID-19 is most commonly transmitted through close contact with an infected person, which is defined as being within 6 feet of the person for 15 minutes or more in 24 hours. People come into close contact when they live together, work together, or have intimate contact.
- COVID-19 spreads via respiratory droplets, which are produced when an infected person speaks, coughs, sneezes, or breathes. People nearby can inhale these droplets, or they can land on surfaces and objects, where they can remain infectious for hours to days.
- COVID-19 can also be spread through airborne transmission in some cases, which occurs when small respiratory droplets or particles containing the virus remain in the air for an extended period and are inhaled by others in the same space.
- COVID-19 can be spread by touching a contaminated surface or object and then touching their mouth, nose, or eyes.
- People who are infected with COVID-19 but do not exhibit any symptoms can still spread the virus.

To prevent the spread of COVID-19, it is critical to practice good hygiene, such as frequent hand washing, mask use, and physical separation. Vaccination is also an important tool in limiting the spread of COVID-19.

TESTS

In India, several types of COVID-19 tests are available for the diagnosis and surveillance of the SARS-CoV-2 virus. Here are some of the most commonly used tests in the country:

- RT-PCR test: The Reverse Transcription Polymerase Chain Reaction (RT-PCR) test is the gold standard for COVID-19 diagnosis. This test detects SARS-CoV-2 virus genetic material in respiratory samples such as nasopharyngeal swabs, oropharyngeal swabs, and sputum samples. RT-PCR tests are available throughout the country in designated hospitals and labs.

- Rapid antigen test (RAT): The Rapid Antigen Test (RAT) is a point-of-care test that detects the presence of viral proteins in respiratory samples. This test produces results in 15-30 minutes and is useful for mass screening, particularly in areas with a high prevalence of COVID-19. RATs are available in primary care clinics, community health centers, and district hospitals across the country.
- TrueNat test: The TrueNat test detects the genetic material of the SARS-CoV-2 virus in respiratory samples using reverse transcriptase loop-mediated amplification (RT-LAMP) technology. TrueNat machines can be found in a variety of hospitals and labs across the country.
- CBNAAT test: The Cartridge-Based Nucleic Acid Amplification Test (CBNAAT) detects the genetic material of the SARS-CoV-2 virus in respiratory samples using the GeneXpert platform. CBNAAT machines are available throughout the country in designated hospitals and labs.
- Antibody test: Antibodies to the SARS-CoV-2 virus are detected in blood samples using the Antibody test, also known as a serology test. This test is used to identify previous infections with the virus and is not recommended for current infection diagnosis. Antibody tests are available across the country in designated hospitals and labs.
- Saliva test: The diagnostic procedure known as a saliva test for COVID-19 is taking a sample of a person's saliva and testing it for the presence of the SARS-CoV-2 virus. The non-invasive nature of this test makes it potentially more practical and comfortable than other diagnostic procedures involving nasal swabs.

It is important to note that COVID-19 tests are free of charge at government-run health facilities in India, and people with COVID-19 symptoms are encouraged to get tested as soon as possible to prevent the virus from spreading.

VACCINATIONS

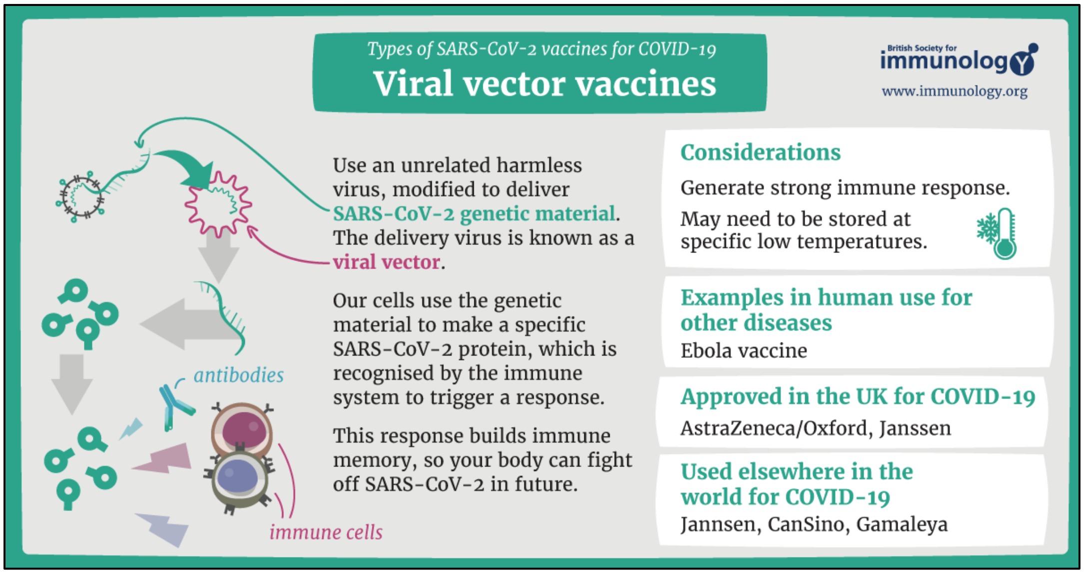
1. Viral vector vaccines for COVID-19

COVID-19 viral vector vaccines are a type of vaccine that employs a weakened or modified virus to deliver genetic material into cells. This genetic material directs the production of a specific protein, such as the spike protein found on the surface of the SARS-CoV-2 virus, which causes COVID-19. This protein then causes an immune response, allowing the body to recognize and fight the virus in the future.

There are several steps involved in developing and manufacturing COVID-19 viral vector vaccines:

1. Designing the genetic material that will be delivered into cells is the first step in developing a viral vector vaccine. This is typically a small piece of DNA or RNA that codes for the spike protein found on the surface of the SARS-CoV-2 virus in COVID-19 vaccines.
2. The next step is to choose a viral vector capable of efficiently delivering genetic material into cells. Adenoviruses, which cause mild respiratory infections in humans, and vesicular stomatitis viruses, which cause mild flu-like symptoms in humans and animals, are two of the most commonly used viral vectors.
3. After a viral vector has been chosen, it can be modified to make it safer and more effective. Scientists may, for example, remove specific genes from the viral vector to prevent it from replicating in the body or causing disease.
4. Cell culture techniques are then used to manufacture the viral vector vaccine. This entails mass production of cells capable of producing the viral vector and genetic material. Following that, the viral vector is combined with the genetic material and purified to remove any impurities.
5. The viral vector vaccine is typically administered by injection into the upper arm muscle tissue. The viral vector enters cells and delivers the genetic material, instructing cells to produce the spike protein.
6. The spike protein produced by cells causes the immune system to produce antibodies and activate immune cells capable of recognizing and combating the SARS-CoV-2 virus.
7. To provide long-term protection, many viral vector vaccines require multiple doses. To ensure that the immune response remains strong, booster shots may be administered several weeks or months after the initial vaccination.

While viral vector vaccines for COVID-19 have been shown in clinical trials to be highly effective, there are some potential risks and side effects associated with these vaccines. Some people may experience mild to moderate side effects such as fever, fatigue, or soreness at the injection site, for example. More serious side effects, such as blood clots or allergic reactions, may occur in rare cases. However, the benefits of vaccination, which include reducing the spread of COVID-19 and preventing severe illness, hospitalization, and death, outweigh the risks.



(fig 4: Viral vector vaccines)

2. Genetic vaccines for COVID-19

Messenger RNA (mRNA), a brief genetic fragment, is used in genetic COVID-19 vaccines, also known as mRNA vaccines, to start an immune response against the SARS-CoV-2 virus. Unlike conventional vaccinations, these ones rely on genetic instructions to tell the body's cells to generate a particular protein found on the surface of the virus rather than a virus that has been weakened or changed. The body can then recognize and combat the virus if it is encountered in the future thanks to the immunological response that is sparked by this protein.

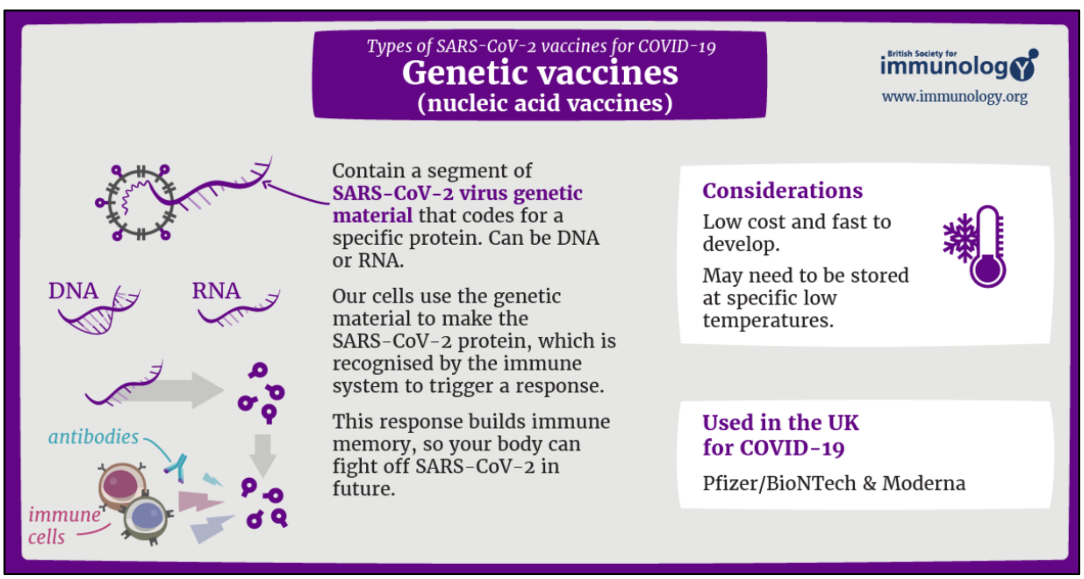
The following steps are Involved In creating genetic COVID-19 vaccines:

1. Designing the mRNA sequence that codes for the viral protein is the first step in creating a genetic vaccination. This is normally a segment of mRNA for the COVID-19 vaccine that codes for the spike protein present on the surface of the SARS-CoV-2 virus.
2. Synthesis: After the mRNA sequence has been created, it is produced in a lab using enzymes that join the nucleotides, the building blocks of RNA molecules, in a predetermined order. After purification, the mRNA is checked for purity.
3. Lipid nanoparticle encapsulation: The mRNA is next wrapped into lipid nanoparticles, which assist it to penetrate cells and shield them from cellular destruction. The lipid

nanoparticles are intended to target cells in the body, such as immune cells, and are composed of a lipid bilayer that surrounds the mRNA.

4. Manufacturing: Cell culture procedures are then used to create the mRNA vaccine. In order to do this, a lot of cells that can create lipid nanoparticles and mRNA must be grown. The lipid nanoparticles and the mRNA are then combined before being processed to remove any contaminants.
5. Administration: The mRNA vaccination is commonly given by injection into the upper arm's muscle tissue. The lipid nanoparticles are delivered into the body, where they transfer the mRNA and direct cells to create the spike protein.
6. Immune response: The body's immune system produces antibodies and activates immune cells that can recognize and combat the SARS-CoV-2 virus as a result of the spike protein produced by cells.
7. Booster shots: To offer durable protection, many genetic vaccinations necessitate several doses. To make sure that the immune response is still strong a few weeks or months after the original immunization, booster doses may be administered.

While genetic COVID-19 vaccines have demonstrated excellent efficacy in clinical studies, there are certain possible hazards and adverse reactions attached to these vaccinations. For instance, some individuals may have mild to moderate adverse effects like fever, exhaustion, or discomfort at the injection site. More severe side effects, like allergic responses, can occasionally happen. The advantages of immunization, such as lowering the spread of COVID-19 and avoiding serious disease, hospitalization, and death, often outweigh these dangers.



(fig 5: Genetic vaccines)

3. Inactivated vaccines for COVID-19

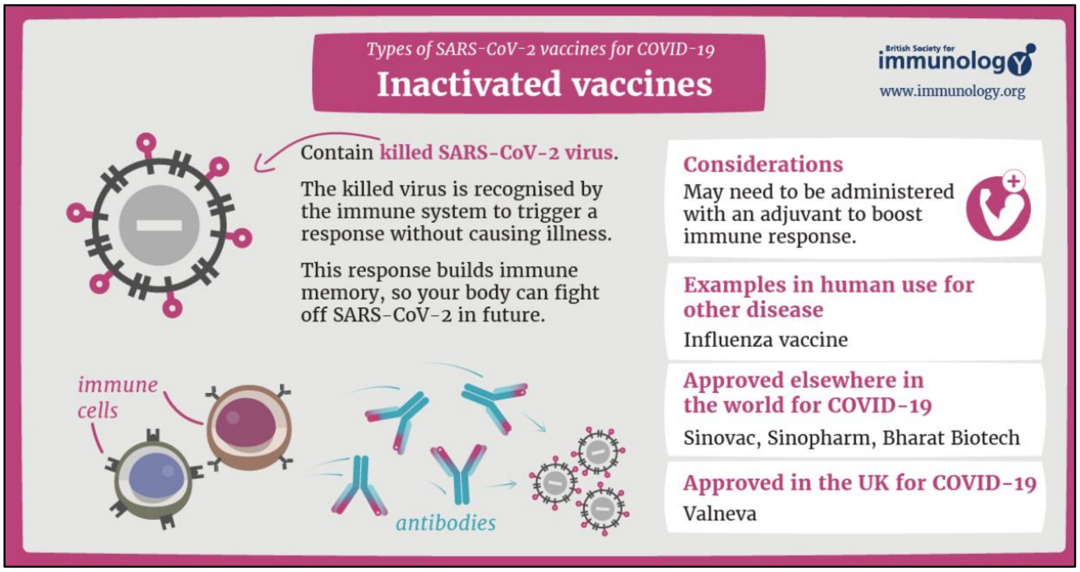
An inactivated or killed variant of the SARS-CoV-2 virus is used in the COVID-19 inactivated vaccine to elicit an immunological response. These vaccines are created by inactivating the virus with chemicals or heat after it has been grown in high quantities in the lab. Numerous additional vaccinations, such as those for polio, hepatitis A, and influenza, have been developed using this method.

The following steps are involved in creating and manufacturing COVID-19 inactivated vaccines.

1. Isolation of the virus: Isolating the virus from a patient sample is the first step in creating an inactivated vaccine for COVID-19. This often entails taking a sample of respiratory secretions from an infected person, such as saliva or nose swabs, and cultivating the virus in a lab.
2. The virus is treated with a chemical or physical substance that inactivates the virus while maintaining its structural integrity after it has been cultivated. Formaldehyde or beta-propiolactone therapy are two common ways to inactivate viruses.
3. Following inactivation, the virus is purified to remove contaminants and concentrate the viral antigens that will be used to elicit an immune response. Specific chemicals that can be discovered on the virus's surface which are recognized by the immune system are called antigens.
4. Purified inactivated virus is combined with adjuvants, which are compounds that improve the immune response to the vaccination, to create the vaccine. Because the virus has been rendered inactive and may not be able to trigger a robust enough immune response on its own, adjuvants are frequently added to inactivated vaccinations.
5. Quality Control: After that, the finished vaccine formulation is examined for purity, effectiveness, and safety. To guarantee that the vaccination is both safe and effective, a number of exacting laboratory testing, animal studies, and clinical trials are required.
6. Administration: The inactivated vaccine is commonly injected into the upper arm's muscular tissue. As soon as the vaccination is administered, the body responds by recognizing and attacking the SARS-CoV-2 virus.
7. Immune Reaction: The immune system responds to the viral antigens in the vaccination by producing antibodies and immune cells that are active and ready to combat a live virus in the future.

Clinical trials have demonstrated the safety and efficacy of inactivated COVID-19 vaccines, with some vaccines having efficacy rates of up to 90%. However, because inactivated vaccines use a fully inactivated virus, there is a chance that the vaccination will interact with the immune system or infect people who already have compromised immune systems. Therefore, inactivated vaccines are not advised for those with immunosuppression or who have specific medical conditions that raise their risk of vaccine-related adverse events.

In conclusion, COVID-19 inactivated vaccines have proven to be a useful weapon in the fight against the pandemic, and ongoing research and development work to further enhance their safety and effectiveness.



(fig 6: Inactivated vaccines)

4. Attenuated vaccines for COVID-19

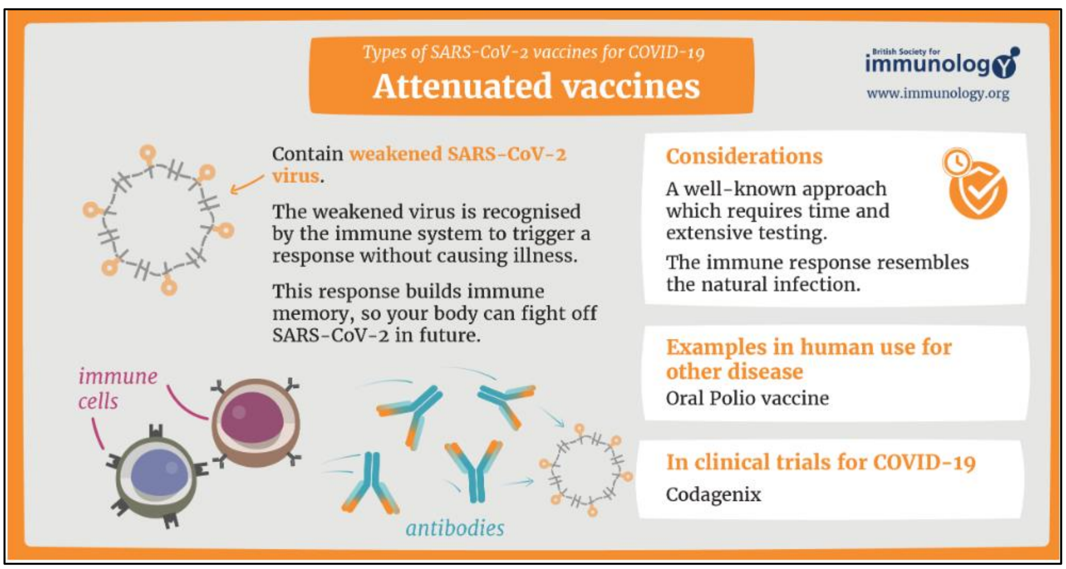
In order to create COVID-19 attenuated vaccines, live viruses are modified to lessen their virulence—or capacity to spread disease—while maintaining their viability. Next, a vaccine made from this altered virus is used to elicit an immune response and protect against subsequent infections. The procedures for creating COVID-19 attenuated vaccines are as follows:

1. Isolation of the virus: Isolating the virus from a patient sample is the first step in creating an attenuated vaccine for COVID-19. This entails taking a sample of respiratory secretions from an infected person, such as saliva or nose swabs, and cultivating the virus in a lab.
2. After being cultivated, the virus is altered in a way that lessens its virulence while still enabling it to multiply in the body. Numerous techniques, such as genetic modification or modifying the virus to grow in cells from a different host species, can be used to accomplish this.
3. Formulation of the Vaccine: A vaccine that can elicit an immunological response is then made using the modified virus. In order to achieve this, it may be necessary to cultivate

huge quantities of the virus and purify it in order to produce a concentrated vaccine, or it may be necessary to make a viral vector vaccine that employs the altered virus as a vehicle for genetic information that induces an immune response.

4. Quality Control: After that, the finished vaccine formulation is examined for purity, effectiveness, and safety. To guarantee that the vaccination is both safe and effective, a number of exacting laboratory testing, animal studies, and clinical trials are required.
5. Administration: The attenuated vaccination is normally injected into the upper arm's muscular tissue. The vaccine, when administered, sets up an immune response that identifies and combats the SARS-CoV-2 virus.
6. Immune Reaction: The immune system responds to the viral antigens in the vaccination by producing antibodies and immune cells that are active and ready to combat a live virus in the future.

COVID-19 attenuated vaccinations have proven to be quite successful at preventing the virus. However, there is a chance that the vaccination could have negative consequences, especially in people with compromised immune systems, because attenuated vaccines use a live virus that has been altered to lessen its virulence. As a result, attenuated vaccines are typically not advised for individuals who are immunocompromised or have specific medical conditions that enhance their risk of vaccine-related side effects.



(fig 7: Attenuated vaccines)

LITERATURE REVIEW

"A Comprehensive Review of COVID-19 Pandemic Forecasting and Simulation Models: Past, Present, and Future Directions" by Singh et al. (2021); This review paper takes an in-depth look at the many COVID-19 forecasting and simulation models utilized throughout the epidemic. The authors analyze the merits and limits of each model, emphasizing the importance of continuing to develop more accurate and adaptive models.

"Visualization of COVID-19 Pandemic Data: A Review" by Boukhelifa et al. (2020); The various visualization strategies used to depict COVID-19 pandemic data are summarised in this review article. The authors stress the significance of clear and accurate visualizations in shaping policy decisions and communicating danger to the general public.

"Forecasting COVID-19 Daily Cases in the United States with a Bayesian Model and Implications for Social Distancing" by Roosa et al. (2020): The daily COVID-19 cases in the United States were predicted by this study using a Bayesian model. The researchers discovered that social isolation practices can dramatically lower the number of cases, and that swift adoption of these practices is essential to lessening the pandemic's overall impact.

"COVID-19 Data Analysis and Forecasting Using a Hybrid Intelligent System: A Comparative Study" by Arora et al. (2020): A Comparative Study" (2020): In this work, the effectiveness of various machine learning algorithms used to analyze and forecast COVID-19 data was examined. The scientists discovered that a hybrid intelligent system that integrates several algorithms may anticipate COVID-19 situations more precisely than any single method.

"COVID-19 Data Analytics and Visualization: A Review" by Das et al. (2021): The many data analytics and visualization methods used to analyze COVID-19 data are summarized in this review article. The authors underline the necessity for standardized techniques for data collecting and analysis while discussing the benefits and drawbacks of each strategy.

"A Comprehensive Review of COVID-19 Pandemic and the Role of Forecasting and Predictive Analytics" by Mishra et al. (2021): In handling the COVID-19 pandemic, this review paper highlights the use of forecasting and predictive analytics. The authors highlight the difficulties involved in modeling infectious diseases and give an outline of the many models and methods used to forecast the spread of the virus.

"Visualization of COVID-19 Data: A Review" by Wu et al. (2021): The many COVID-19 data visualisation methods are discussed in this review article. The authors stress the significance of accurate and clear visualisations for presenting complicated information to the general audience and decision-makers.

"Machine Learning and AI Techniques in COVID-19 Forecasting: A Comprehensive Review" by Sarwar et al. (2021): An overview of the various machine learning and artificial intelligence strategies utilised in COVID-19 forecasting is given in this review article. The authors stress the need of incorporating data from several sources for precise forecasting and analyse the advantages and disadvantages of each technique.

"COVID-19 Forecasting: Model Selection, Uncertainty, and Value of Information" by Reich et al. (2021): The significance of model choice and uncertainty in COVID-19 forecasting are covered in this article. The authors address the potential effects of various sources of uncertainty on forecasting accuracy and emphasise the need for transparent and repeatable modelling practises.

"Visualization of COVID-19 Vaccine Distribution Data: A Review" by Choi et al. (2021): The various visualisation methods used to depict COVID-19 vaccination distribution data are discussed in this review paper. The authors stress the need of concise and understandable visualisations in conveying the status of vaccination distribution and pinpointing areas for improvement.

METHODOLOGY

- **Data collection:** For any research or modeling, it is essential to get accurate and trustworthy data on COVID-19 cases, deaths, and other pertinent variables. Data can be retrieved through academic research repositories, government websites, or public health organizations
- **Data cleaning and pre-processing:** Data preparation and cleaning are necessary after data collection in order to guarantee correctness and consistency. Duplicate data should be eliminated, errors should be fixed, and data formats should be standardized.
- **Data analysis and modeling:** The COVID-19 data can be analyzed and modeled using a variety of statistical and machine learning algorithms. Time series analysis, regression analysis, grouping, and classification are a few typical methods.
- **Visualization**: Insights from COVID-19 data can be shared with a larger audience using visualization approaches. Popular visualization tools include graphs, charts, heat maps, and geographic maps.
- **Forecasting and predictions:** Modelling methods can be used to create forecasts and predictions about the spread of COVID-19, such as the number of cases or fatalities over a specific time. These forecasts can be used to guide resource allocation and public health policy.

- **Evaluation and validation**: Finally, it's critical to assess and confirm the precision of any study or modeling. To do this, one can compare forecasts with actual results or use statistical accuracy measures like mean absolute error or root mean squared error.

DATA COLLECTION

- Official government data: A lot of nations and regions have official government websites or public health organisations that offer the most recent COVID-19 data. Information on the quantity of instances, fatalities, recoveries, hospitalisations, and other pertinent factors can be included in this.
- Academic research repositories: Access to COVID-19 data is made possible by a number of academic research repositories, including the European Centre for Disease Prevention and Control (ECDC), the COVID-19 Open Research Dataset (CORD-19), and the Centre for Systems Science and Engineering at Johns Hopkins University.
- Social media data: Real-time data on COVID-19 conversations, trends, and public attitude can be gathered via social media platforms like Twitter and Facebook.
- Surveys and questionnaires: Surveys and questionnaires can be used to gather information on issues like COVID-19 vaccine reluctance, societal perceptions of pandemic-related regulations, and the pandemic's effects on mental health.
- Medical records: These records can offer thorough information on COVID-19 patients' demographics, comorbidities, and clinical results.
- Mobility data: Data acquired by Google and Apple, for example, can be used to analyse changes in human movement patterns and predict the spread of COVID-19.

CHAPTER 1

DATASET EXPLANATION

The virus (more precisely, a coronavirus) known as the 2019 Novel Coronavirus (2019-nCoV) has been determined to be the origin of an outbreak of respiratory illnesses that was originally discovered in Wuhan, China. Early reports suggest that a huge seafood and animal market was connected to a significant number of the cases in the Wuhan, China outbreak, indicating possible human-to-animal transmission. The fact that an increasing percentage of patients reportedly have not been exposed to animal markets suggests that person-to-person transmission is taking place. It's unclear at this time how quickly or sustainably this virus is spreading among humans.

SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
1	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1	0	0
2	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14	0	0
3	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6	0	0
4	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1	0	0
5	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0	0	0
6	01/22/2020	Guangdong	Mainland China	1/22/2020 17:00	26	0	0
7	01/22/2020	Guangxi	Mainland China	1/22/2020 17:00	2	0	0
8	01/22/2020	Guizhou	Mainland China	1/22/2020 17:00	1	0	0
9	01/22/2020	Hainan	Mainland China	1/22/2020 17:00	4	0	0
10	01/22/2020	Hebei	Mainland China	1/22/2020 17:00	1	0	0
11	01/22/2020	Heilongjiang	Mainland China	1/22/2020 17:00	0	0	0
12	01/22/2020	Henan	Mainland China	1/22/2020 17:00	5	0	0
13	01/22/2020	Hong Kong	Hong Kong	1/22/2020 17:00	0	0	0
14	01/22/2020	Hubei	Mainland China	1/22/2020 17:00	444	17	28
15	01/22/2020	Hunan	Mainland China	1/22/2020 17:00	4	0	0
16	01/22/2020	Inner Mongolia	Mainland China	1/22/2020 17:00	0	0	0
17	01/22/2020	Jiangsu	Mainland China	1/22/2020 17:00	1	0	0
18	01/22/2020	Jiangxi	Mainland China	1/22/2020 17:00	2	0	0
19	01/22/2020	Jilin	Mainland China	1/22/2020 17:00	0	0	0
20	01/22/2020	Liaoning	Mainland China	1/22/2020 17:00	2	0	0
21	01/22/2020	Macau	Macau	1/22/2020 17:00	1	0	0
22	01/22/2020	Ningxia	Mainland China	1/22/2020 17:00	1	0	0
23	01/22/2020	Qinghai	Mainland China	1/22/2020 17:00	0	0	0
24	01/22/2020	Shaanxi	Mainland China	1/22/2020 17:00	0	0	0
25	01/22/2020	Shandong	Mainland China	1/22/2020 17:00	2	0	0

(fig 8: Dataset)

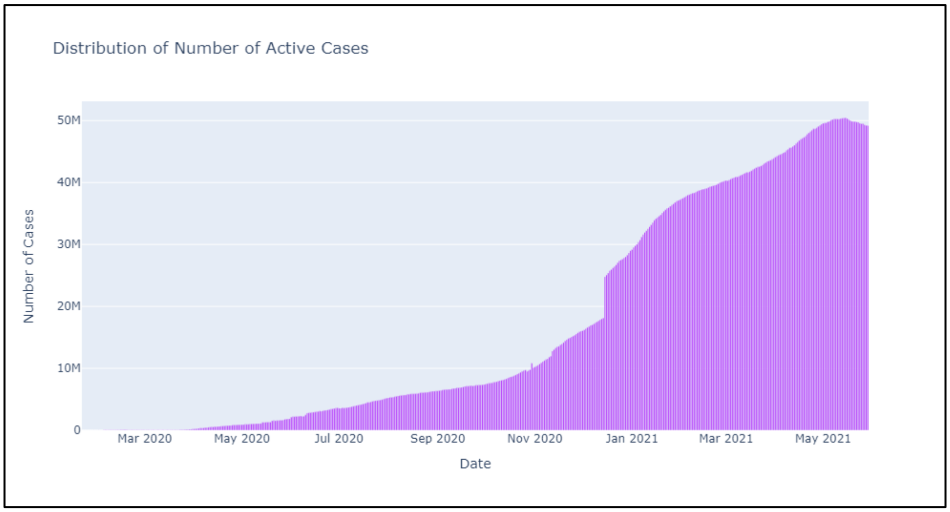
The Dataset can be used for various purposes such as analyzing trends and patterns in COVID-19 cases and deaths across different geographic regions, identifying areas with high rates of cases and deaths, and assessing the effectiveness of interventions aimed at reducing the spread of the virus. To use the data, it may be necessary to clean and preprocess the data, remove duplicates, and potentially fill in any missing values before analysis.

The following is the description of the attributes of the dataset:

- Sno - Serial number
- ObservationDate - The date range from January 22, 2020, to May 29, 2021, for COVID-19 observation
- Province/State - Province or state where COVID-19 observations were made (could be empty)
- Country/Region - Different countries of the world where COVID-19 observations were made
- Last Update - The last time (in UTC) when the data was updated for a given province or country
- Confirmed - Cumulative number of confirmed COVID-19 cases till the date of observation
- Deaths - Cumulative number of deaths due to COVID-19 till the date of observation
- Recovered - Cumulative number of COVID-19 patients who have recovered till the date of observation

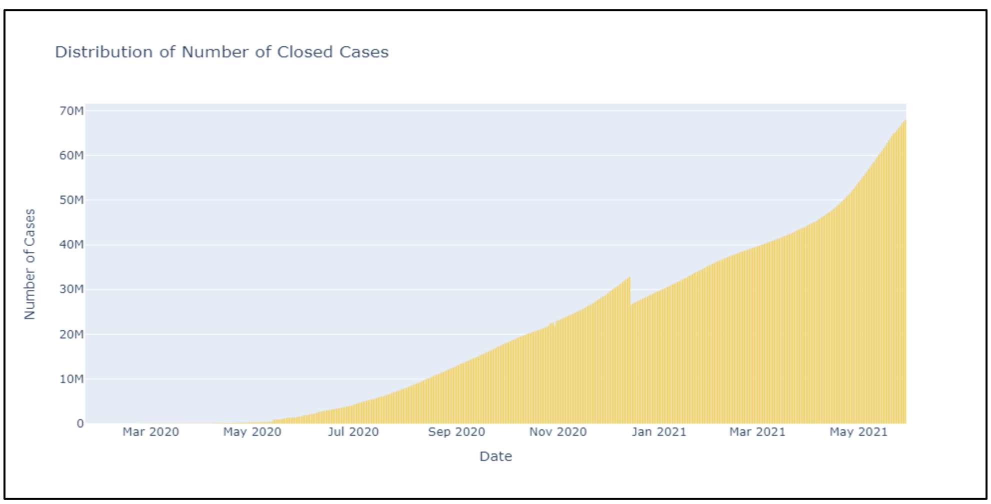
CHAPTER 2

DATA ANALYSIS



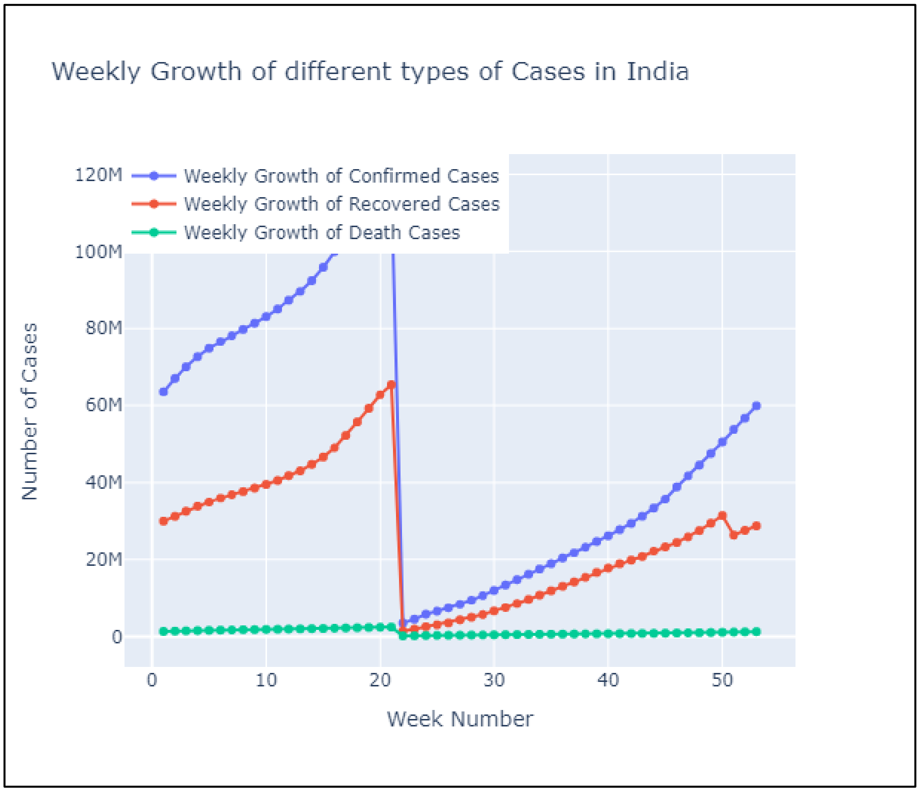
(fig 9: Distribution of Number of Active Cases)

The distribution of the number of active COVID-19 cases over time is depicted in the graph. The x-axis of the chart represents the date, while the y-axis shows the number of active cases. Active cases are calculated by subtracting the total number of recovered cases and deaths from the total number of confirmed cases till that date. The graph shows that the number of active cases started out at 149.512K and has since grown to 49.19M.



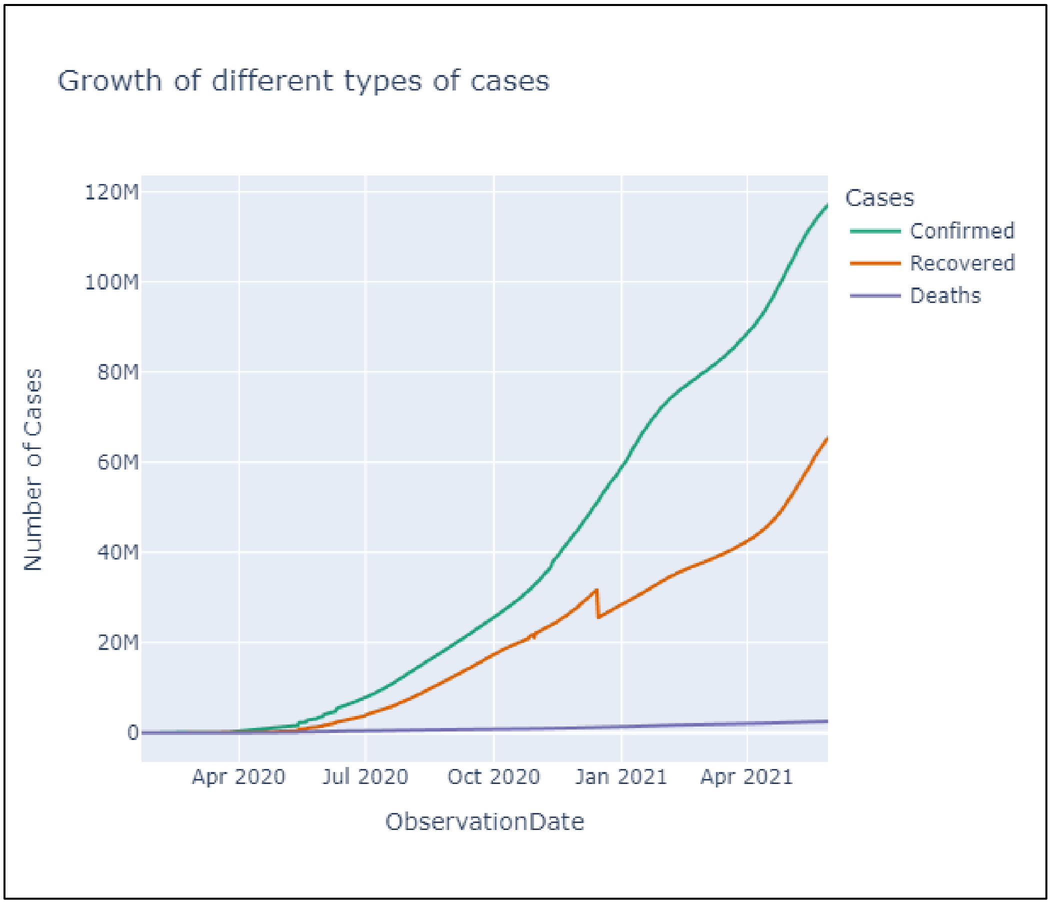
(fig 10: Distribution of Number of Closes Cases)

The distribution of the total number of resolved COVID-19 cases through time is depicted in the graph. The x-axis of the chart represents the date, while the y-axis shows the number of closed cases. Closed cases are calculated by adding the total number of recovered cases and deaths till that date. The graph demonstrates that the total number of closed cases increased from 230K to 67.6M over time.



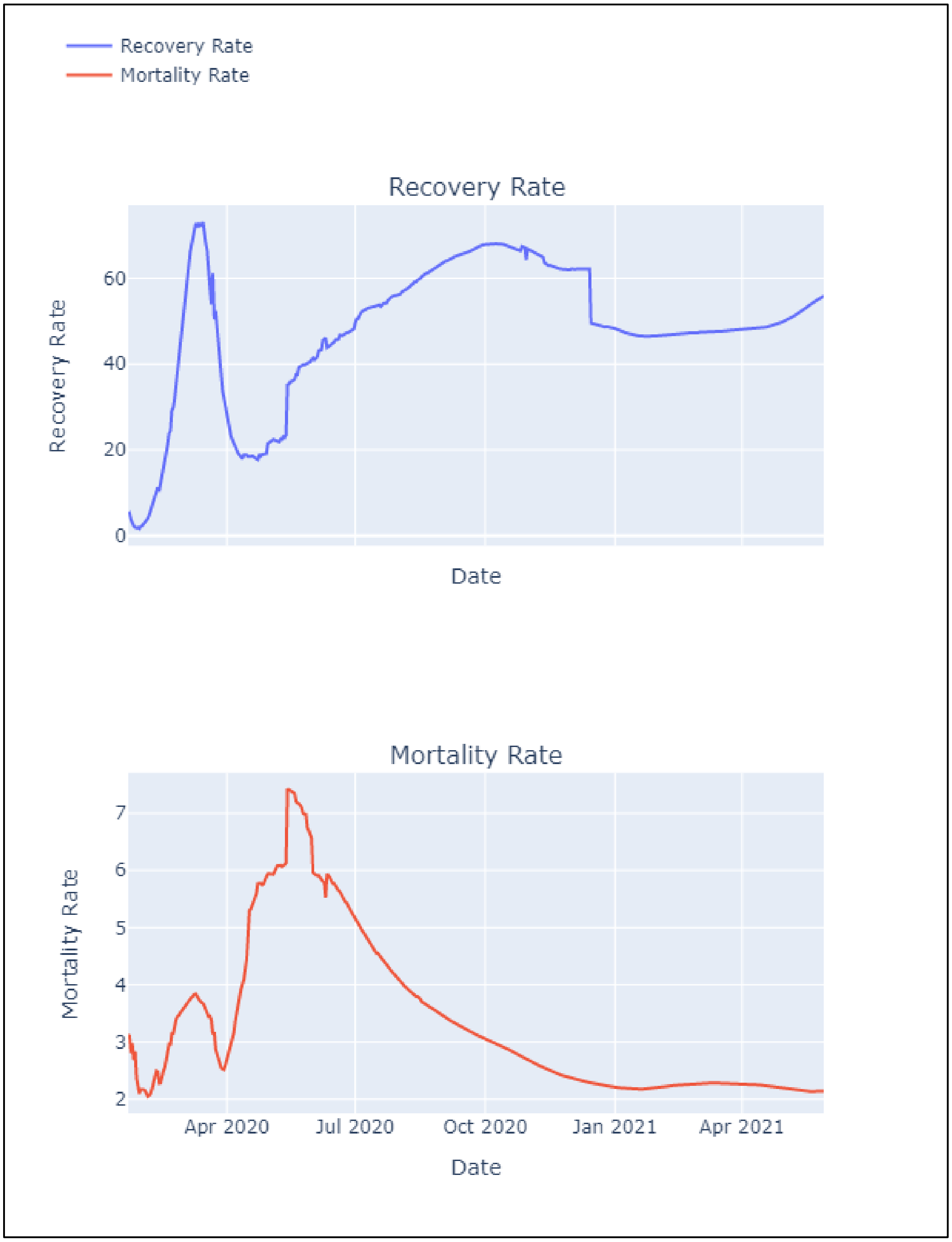
(fig 11: Weekly Growth of different types of Cases in India)

Based on the above graph, we can predict that the COVID-19 situation in the country mentioned is still evolving and the number of confirmed cases is increasing every week, but the recovery rate is also increasing. However, the death cases are consistently low during a certain period of time. The situation is subject to change in the future depending on various factors such as government policies, vaccination drives, and public behavior towards the pandemic.We can see that the number of weekly growth of confirmed cases is increasing every 1 week. Even recovered cases are also increases but not as the speed of confirmed cases. The death cases is as low as possible and consistent for a period of time.



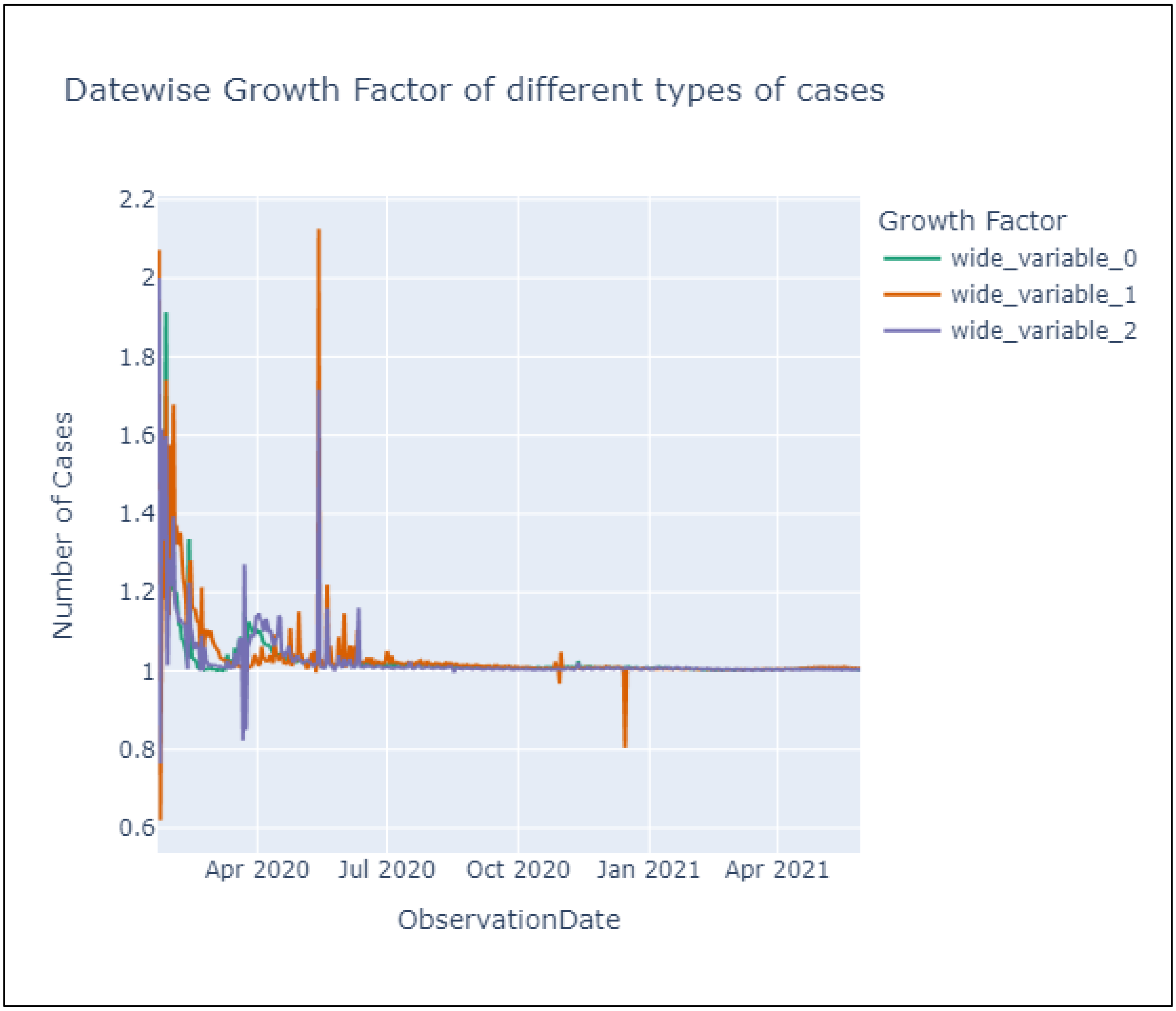
(fig 12: Growth of different types of cases)

The graph displays the overall rise in confirmed, recovered, and death case counts.



(fig 13: Recovery Rate and Mortality Rate)

The first graph demonstrates that the recovery rate has grown, having initially decreased and then slowly and steadily rising. The mortality rate in the second graph also rose in the first few months before decreasing smoothly over time.



(fig 14: Date wise Growth Factor of different Types of cases)

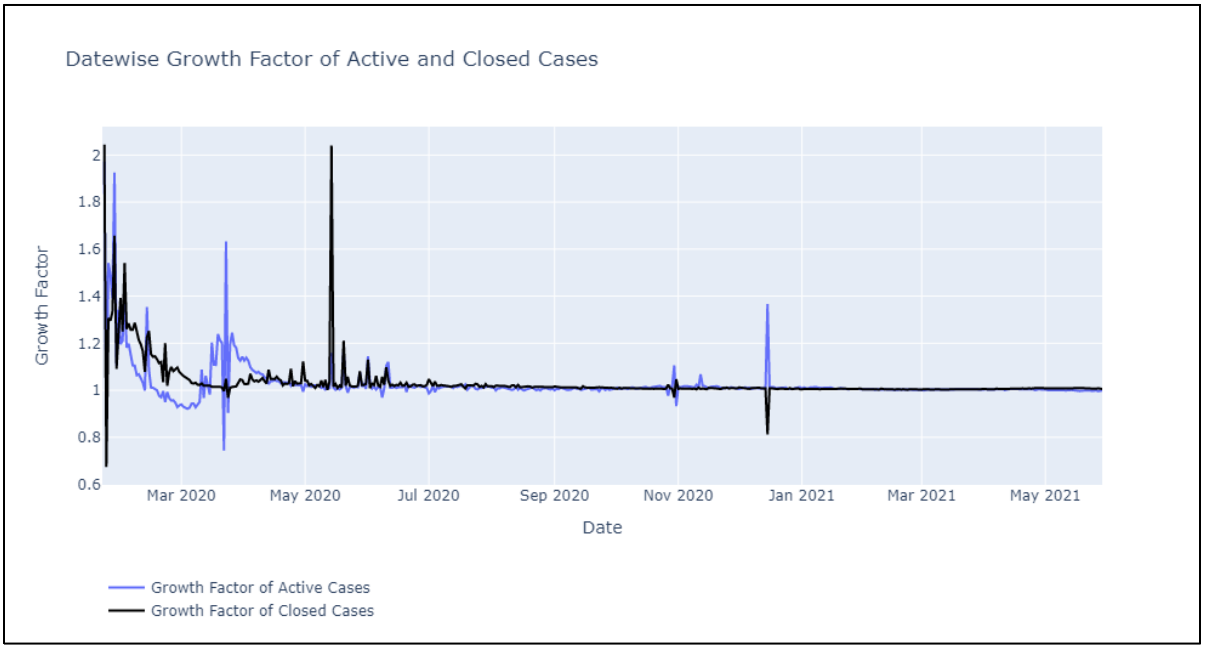
Growth Factor

The rate at which a quantity multiplies itself over time is known as its growth factor. It goes like this:

Formula: New (Confirmed, Recovered, Deaths) for each day divided by new (Confirmed, Recovered, Deaths) for the day before.

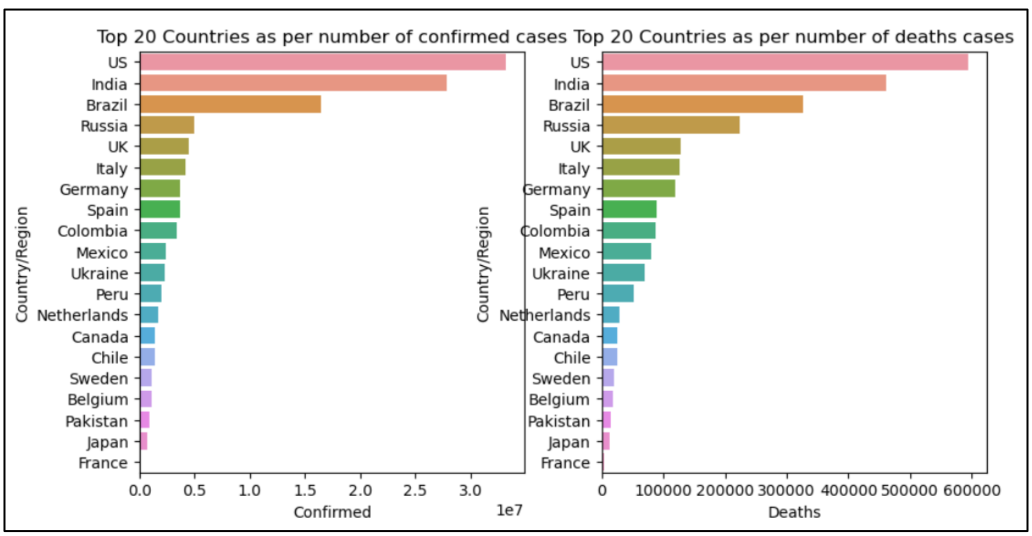
A growth factor of less than 1 indicates that the number of cases is decreasing, while a growth factor greater than 1 indicates that the number of cases is increasing. If the growth factor remains constant at 1, then there has been no change in any kind of situation.

In the above graph the growth rate is most of the time above 1for different types of cases.



(fig 15: Datewise Growth Factor of Active and Closed Cases)

The initial growth rate for active cases was given, and after years of fluctuation, it has stabilised to a steady rate that is almost equal to 1. However, compared to active cases, closed cases showed a higher growth rate in the initial stage.



(fig 16: Top 20 Countries as per Confirmed and Death cases)

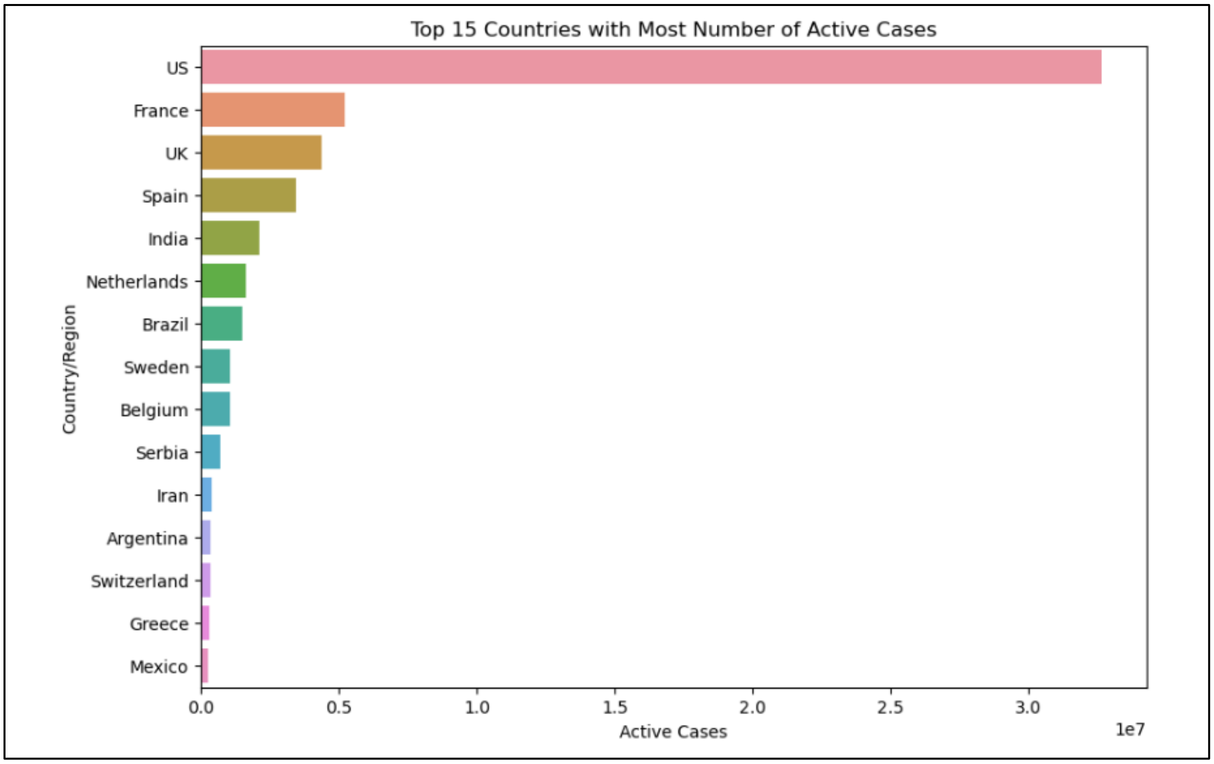
The above graphs show the top 20 countries for number of confirmed cases and number of death cases.

	No. of cases	Days since first Case	Number of days for doubling
0	560	0 days	0 days
1	1120	2 days	2 days
2	2240	4 days	2 days
3	4480	5 days	1 days
4	8960	8 days	3 days
5	17920	11 days	3 days
6	35840	16 days	5 days
7	71680	25 days	9 days
8	143360	50 days	25 days
9	286720	58 days	8 days
10	573440	64 days	6 days
11	1146880	72 days	8 days
12	2293760	86 days	14 days
13	4587520	114 days	28 days
14	9175040	152 days	38 days
15	18350080	194 days	42 days
16	36700160	260 days	66 days
17	73400320	327 days	67 days
18	146800640	458 days	131 days

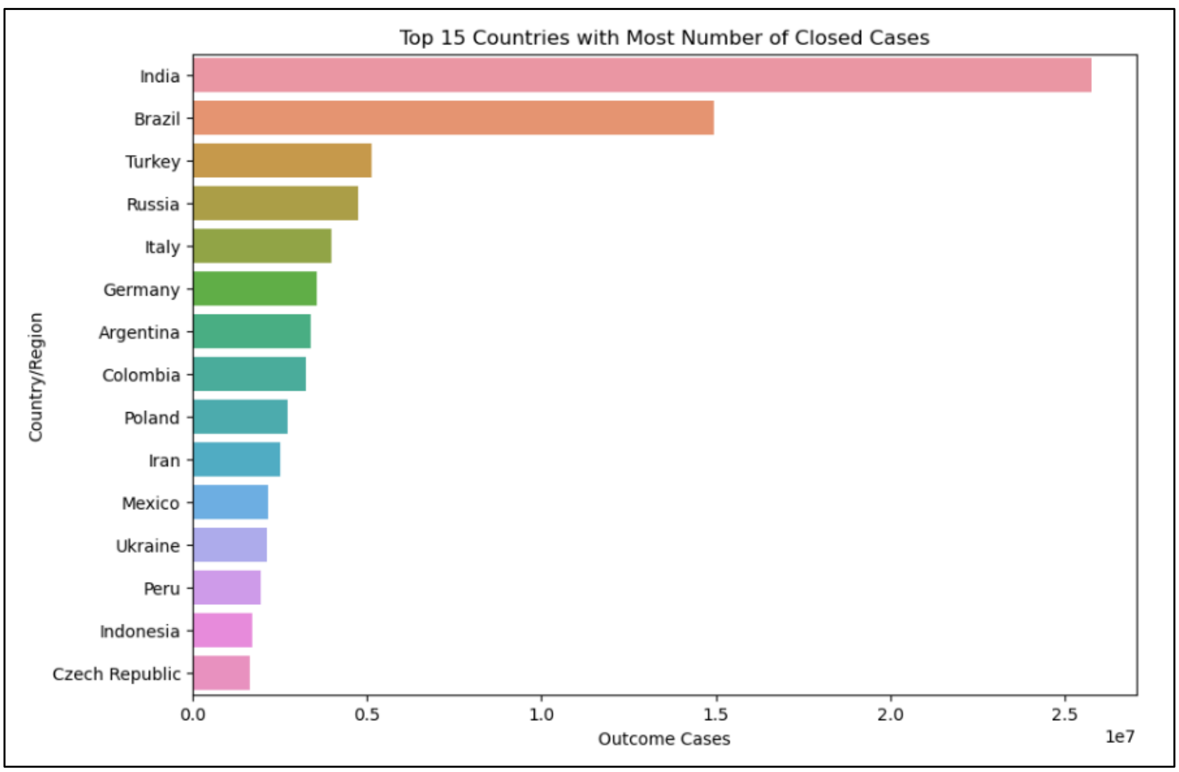
(fig 17: Rate of Doubling for confirmed around the world)

The global rate of doubling for confirmed cases refers to how long it takes for the total number of confirmed COVID-19 cases to double. It is a key indicator for tracking the propagation of the virus and determining the efficiency of anti-viral efforts.

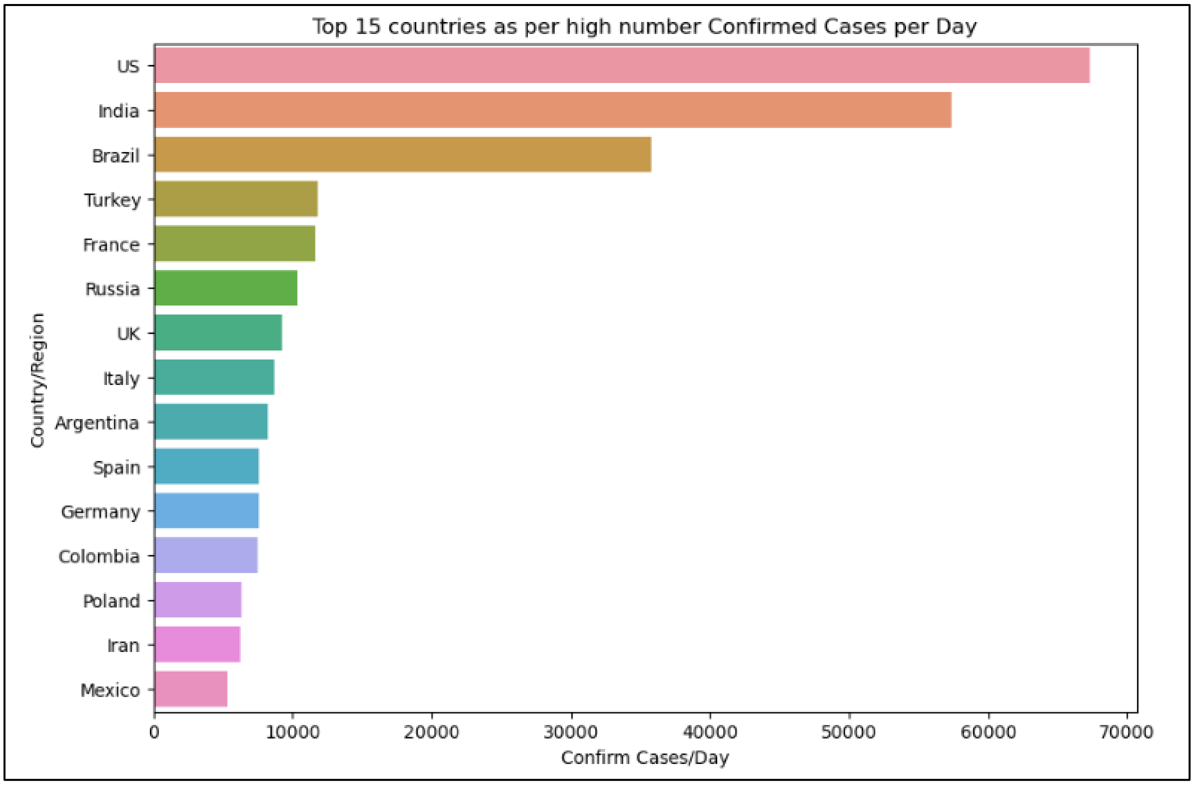
The number of confirmed cases is shown on a logarithmic scale over time to compute the rate of doubling. This is due to the virus's exponential growth rate, and showing the data on a logarithmic scale will assist see this growth more vividly. After plotting the data, the rate of doubling can be computed by calculating how long it takes for the number of confirmed cases to double.



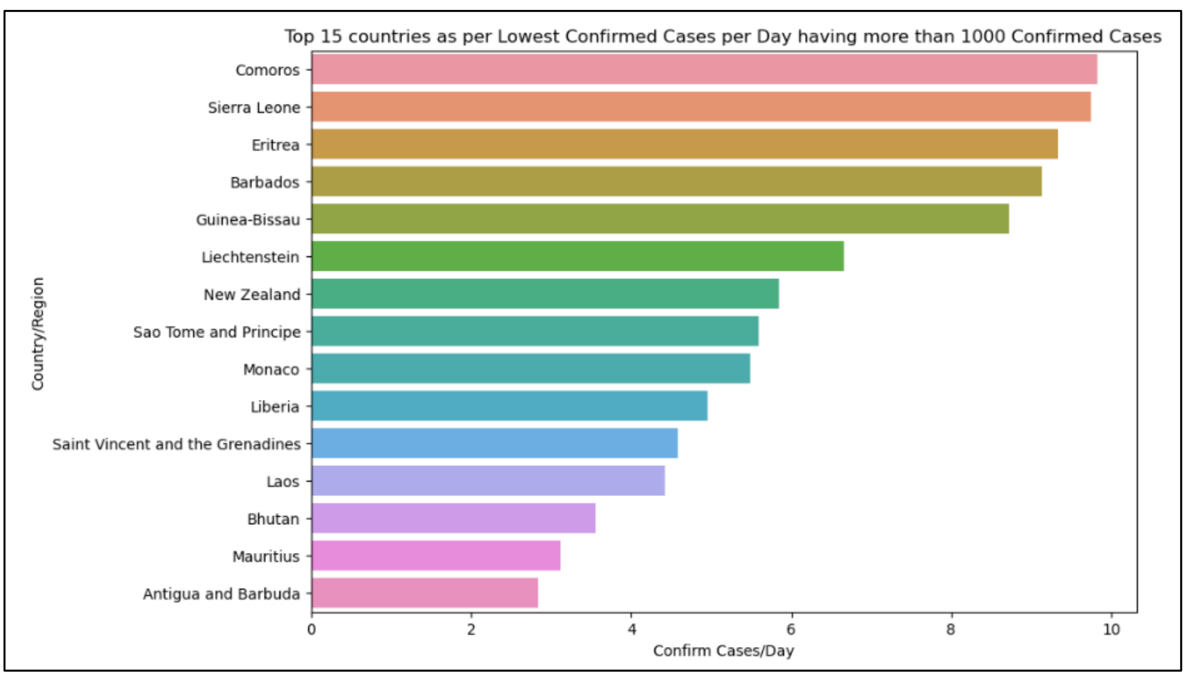
(fig 18: Top 15 countries with most number of active cases)



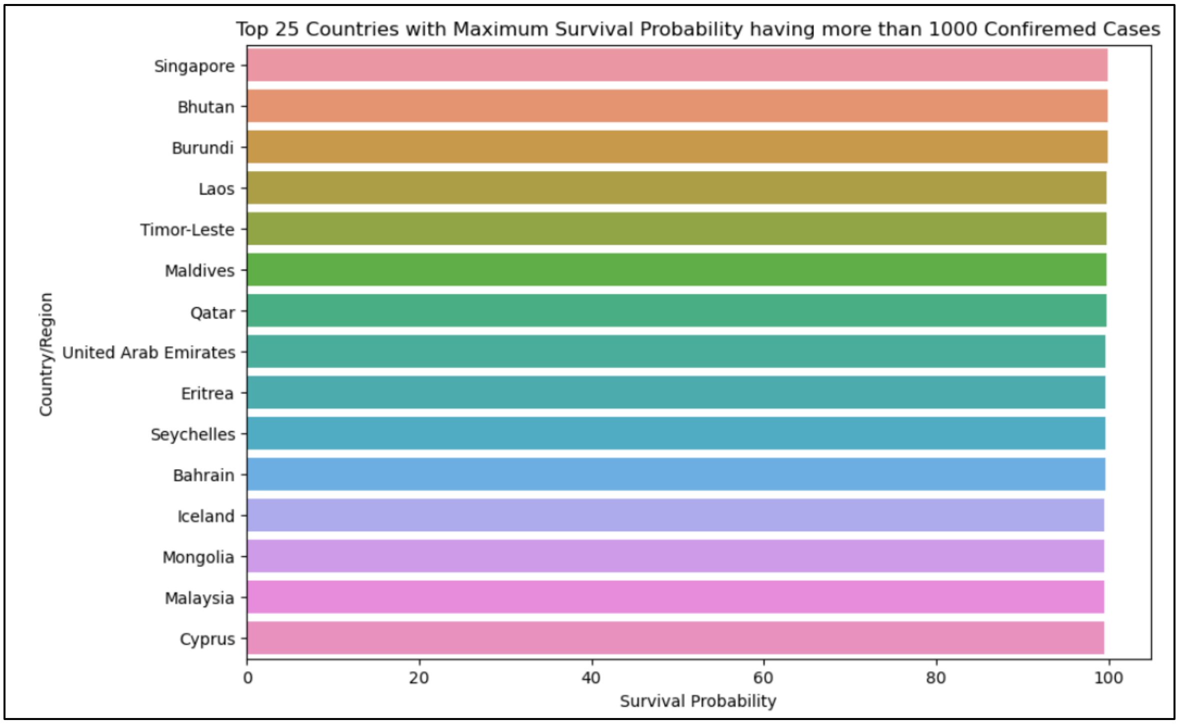
(fig 19: Top 15 countries with most number of closed cases)



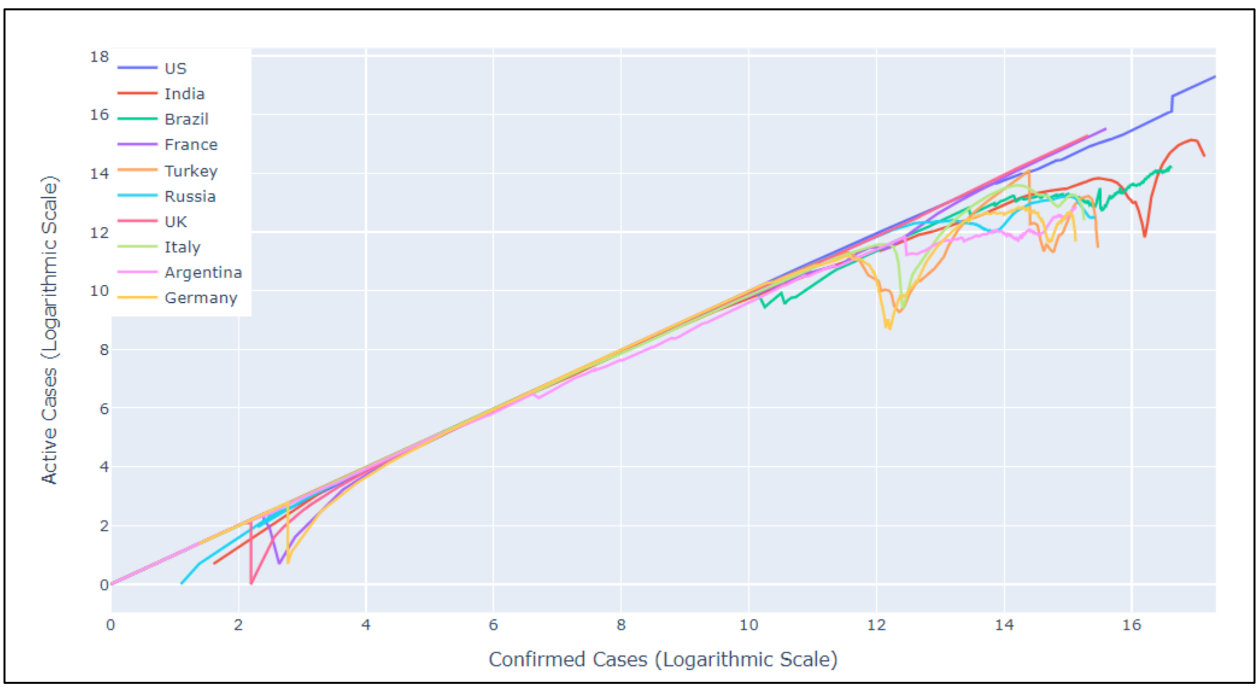
(fig 20: Top 15 countries as per the high number confirmed cases per day)



(fig 21: Top 15 countries as per lowest confirmed cases per day having more than 1000 confirmed cases)

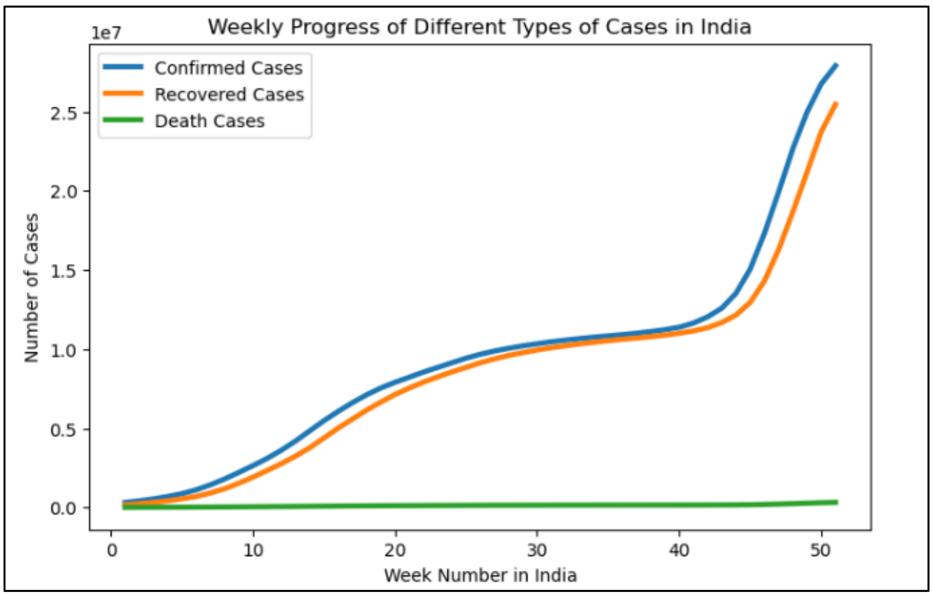


(fig 22: Top 25 countries with maximum survival probability having more than 1000 confirmed cases)



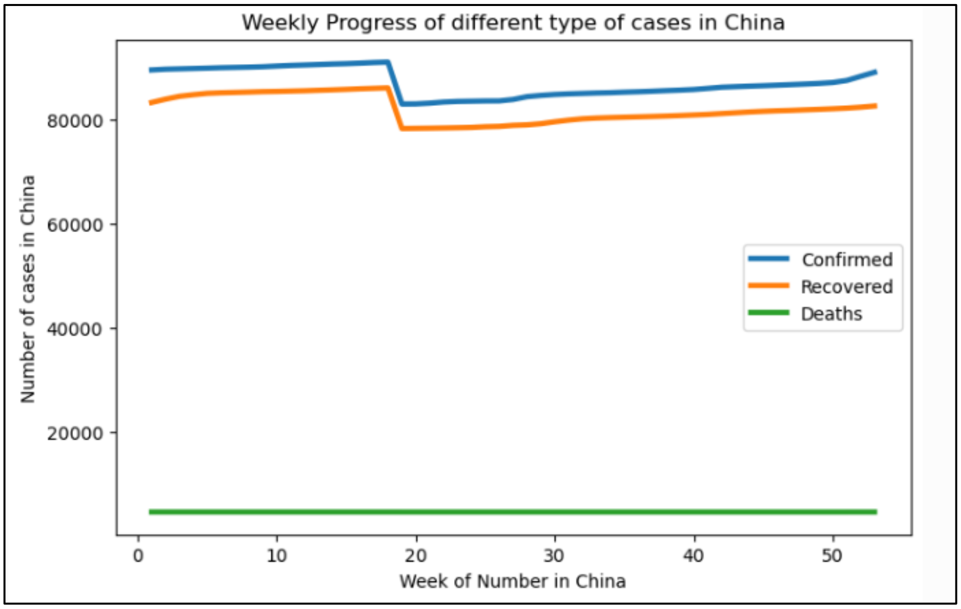
(fig 23: Confirmed cases – Logarithmic scale)

It's pretty evident that the disease is spreading in the same manner everywhere, but a if particular country is following pandemic-controlling practices rigorously the results are evident in the graph. Most of the countries will follow the same trajectory as the USA, which is **"Uncontrolled Exponential Growth"**. There are a few countries where the pandemic controlling practices seems to be working accurately, few classic examples are China, Germany, Italy, Spain, and Turkey has started showing that dip indicating there are somehow got control over COVID-19. Countries like the United Kingdom, and Russia are following similar lines as that of the United States, indicating the growth is still exponential among those countries. Iran is showing some occasional drops.



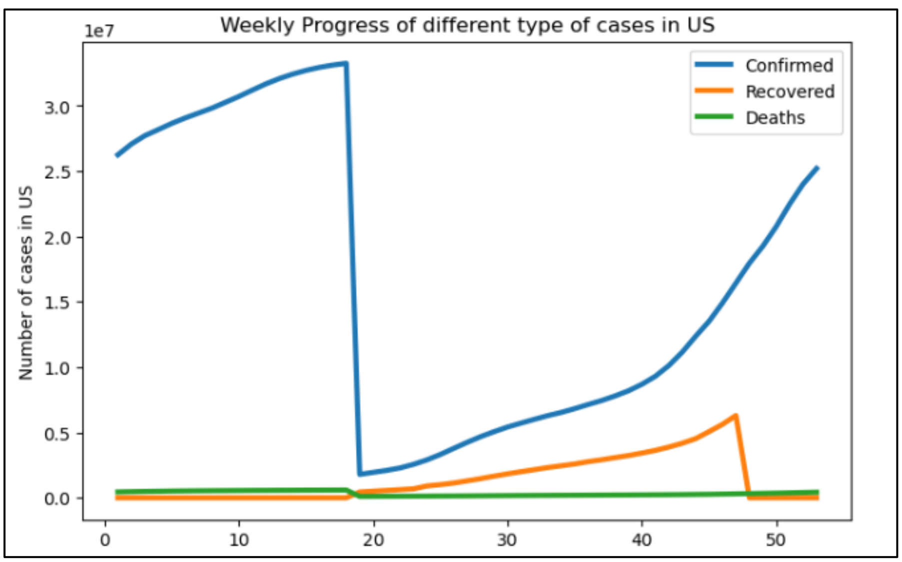
(fig 24: Weekly Progress of Different Types of Cases in India)

The number of confirmed and recovered cases have incresed over time while the death cases remained at 0.



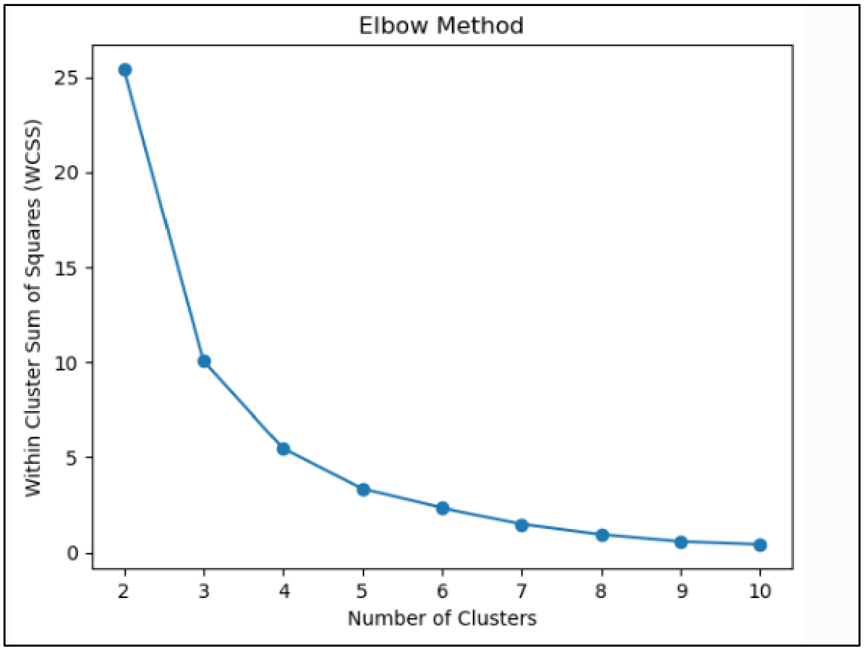
(fig 25: Weekly Progress of Different Types of Cases in China)

The number of confirmed and recovered cases was very high from the begenning while the death cases remained the same.



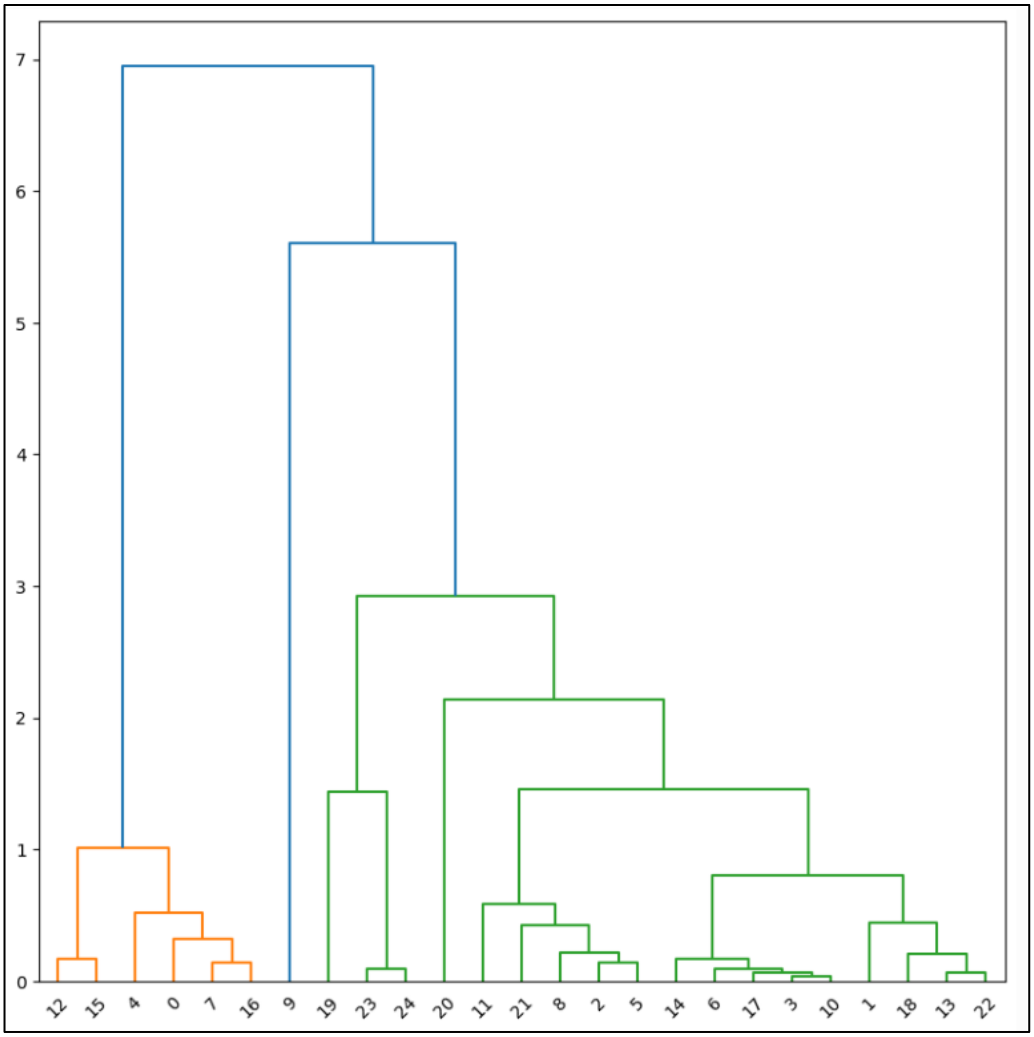
(fig 26: Weekly Progress of Different Types of Cases in US)

The number of confirmed cases was high in the beginning later fell down to 0 and again rose up. Bothe recovered and death were constant at 0 for some period later recovered rose till 0.5 and again fell down to 0.



(fig 27 : Elbow Method)

Here I am plotting the K-Means clustering on the Mortality and Recovery rates of each country. The Elbow Method is then used to determine the optimal number of clusters to use for the K-Means clustering. We can see that the optimal cluisters are 10.



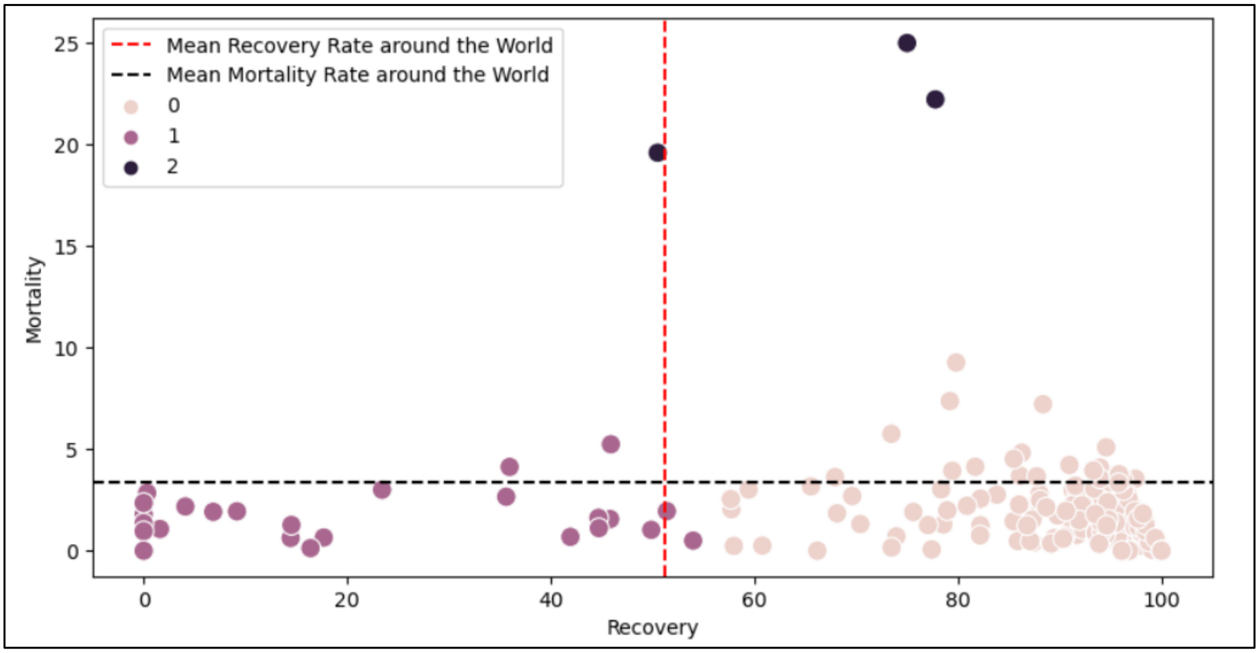
(fig 28: Dendogram)

In the above graph we can see that the dendrograms is used to identify groups of countries with similar COVID-19 patterns based on their mortality and recovery rates. We can determine the optimal number of clusters to divide the countries and identify the countries that are more similar to each other in terms of their COVID-19 patterns are 25.

Country/Region	Confirmed	Recovered	Deaths	Mortality	Recovery	Clusters
US	33251939.00	0.00	594306.00	1.79	0.00	1.00
UK	4496823.00	15481.00	128037.00	2.85	0.34	1.00
Spain	3668658.00	150376.00	79905.00	2.18	4.10	1.00
Netherlands	1671967.00	26810.00	17889.00	1.07	1.60	1.00
Sweden	1068473.00	0.00	14451.00	1.35	0.00	1.00
Belgium	1059763.00	0.00	24921.00	2.35	0.00	1.00
France	119524.00	59192.00	986.00	0.82	49.52	1.00
Mexico	2411503.00	1924865.00	223455.00	9.27	79.82	2.00
India	27894800.00	25454320.00	325972.00	1.17	91.25	0.00
Brazil	16471600.00	14496224.00	461057.00	2.80	88.01	0.00
Russia	4995613.00	4616422.00	118781.00	2.38	92.41	0.00
Italy	4213055.00	3845087.00	126002.00	2.99	91.27	0.00
Germany	3684672.00	3479700.00	88413.00	2.40	94.44	0.00
Colombia	3363061.00	3141549.00	87747.00	2.61	93.41	0.00
Ukraine	2257904.00	2084477.00	52414.00	2.32	92.32	0.00
Peru	1947555.00	1897522.00	68978.00	3.54	97.43	0.00
Canada	1384373.00	1322602.00	25451.00	1.84	95.54	0.00
Chile	1369597.00	1295499.00	29047.00	2.12	94.59	0.00
Pakistan	918936.00	839322.00	20736.00	2.26	91.34	0.00
Japan	742539.00	666786.00	12877.00	1.73	89.80	0.00
Mainland China	91072.00	86117.00	4636.00	5.09	94.56	0.00
Australia	30096.00	23579.00	910.00	3.02	78.35	0.00
Hong Kong	11837.00	11571.00	210.00	1.77	97.75	0.00

(fig 29: Gradient colour map)

The following is a gradient color map of the top 15 countries in each cluster, where the clusters are identified using the K-Means clustering algorithm applied on the mortality rate and recovery rate of countries.



(fig 30: Cluster plot)

Few Countries belonging to Cluster 0: ['India', 'Brazil', 'Turkey', 'Russia', 'Italy', 'Argentina', 'Germany', 'Colombia', 'Iran', 'Poland']

Few Countries belonging to Cluster 1: ['US', 'France', 'UK', 'Spain', 'Netherlands', 'Sweden', 'Belgium', 'Serbia', 'Switzerland', 'Greece']

Few Countries belonging to Cluster 2: ['Yemen', 'MS Zaandam', 'Vanuatu']

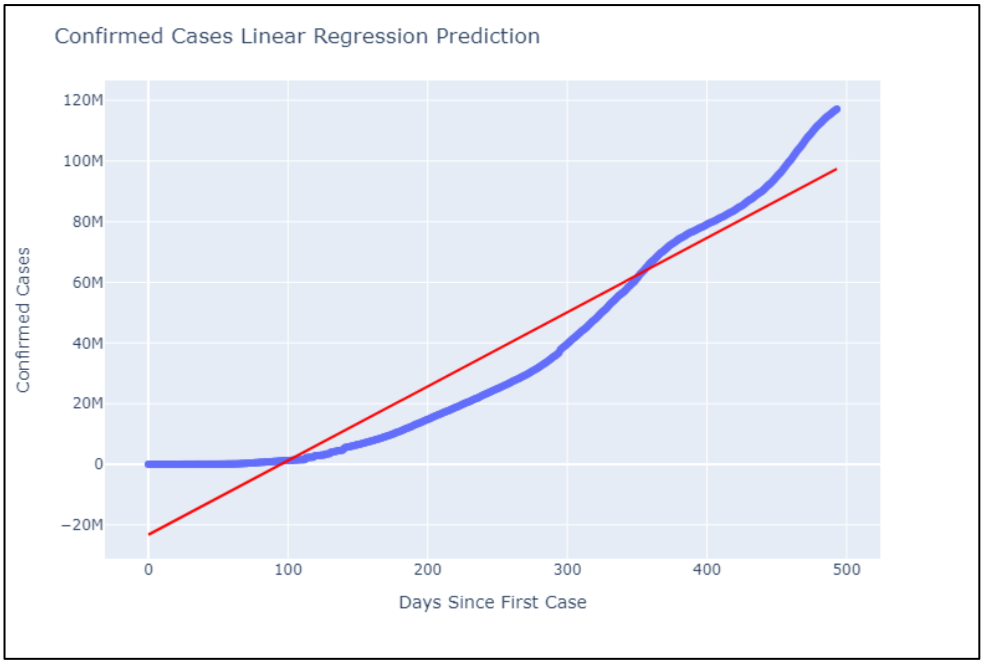
Cluster 2 is a set of countries which have really High Mortality Rate and considerably Good Recovery Rate. Basically, few countries among these clusters have seen already the worst of this pandemic but are now recovering with healthy Recovery Rate.

Cluster 0 is set of countries which have Low Mortality Rate and really High Recovery Rate. These are the set of countries who has been able to control the COVID-19 by following pandemic controlling practices rigorously.

Cluster 1 is set of countries which have Low Mortality Rate and really Low Recovery Rate. These countries need to pace up their Recovery Rate to get out it, some of these countries have really high number of Infected Cases but Low Mortality is positive sign out of it.

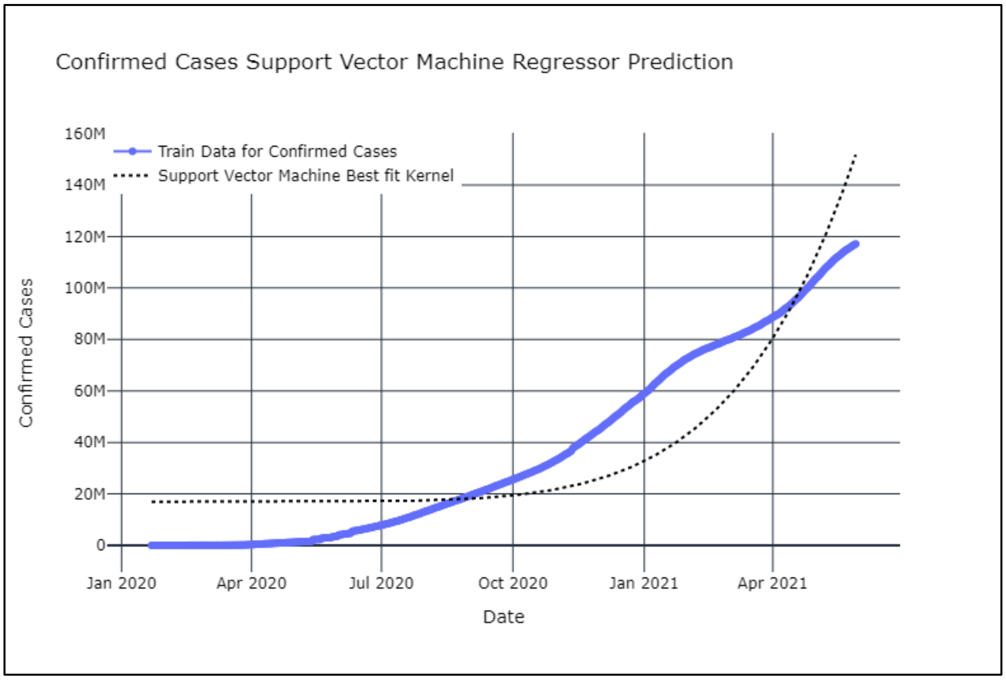
CHAPTER 3

PREDICTIONS



(fig 31: Linear Regression)

- In this case, I chose a training set that contains 95% of the data and a validation set that contains the remaining 5%.
- After splitting the data, the Scikit-Learn LinearRegression function is used to train a linear regression model. The model is trained using the column "Days Since" as the independent variable and the column "Confirmed" as the dependent variable.
- The model is then applied to the validation set to produce predictions. Predictions are saved in the field prediction_valid_linreg.
- The method then computes the model's root mean squared error (RMSE) on the validation set. The RMSE indicates how well the model fits the data. The better the fit, the lower the RMSE.
- The code generates a scatter plot of the confirmed cases over time, with a trendline showing the predictions of the linear regression model. The RMSE value is also displayed on the plot and is printed to the console. With an error of about 4,536 confirmed examples, the RMSE number shows that the model's predictions are relatively reliable.



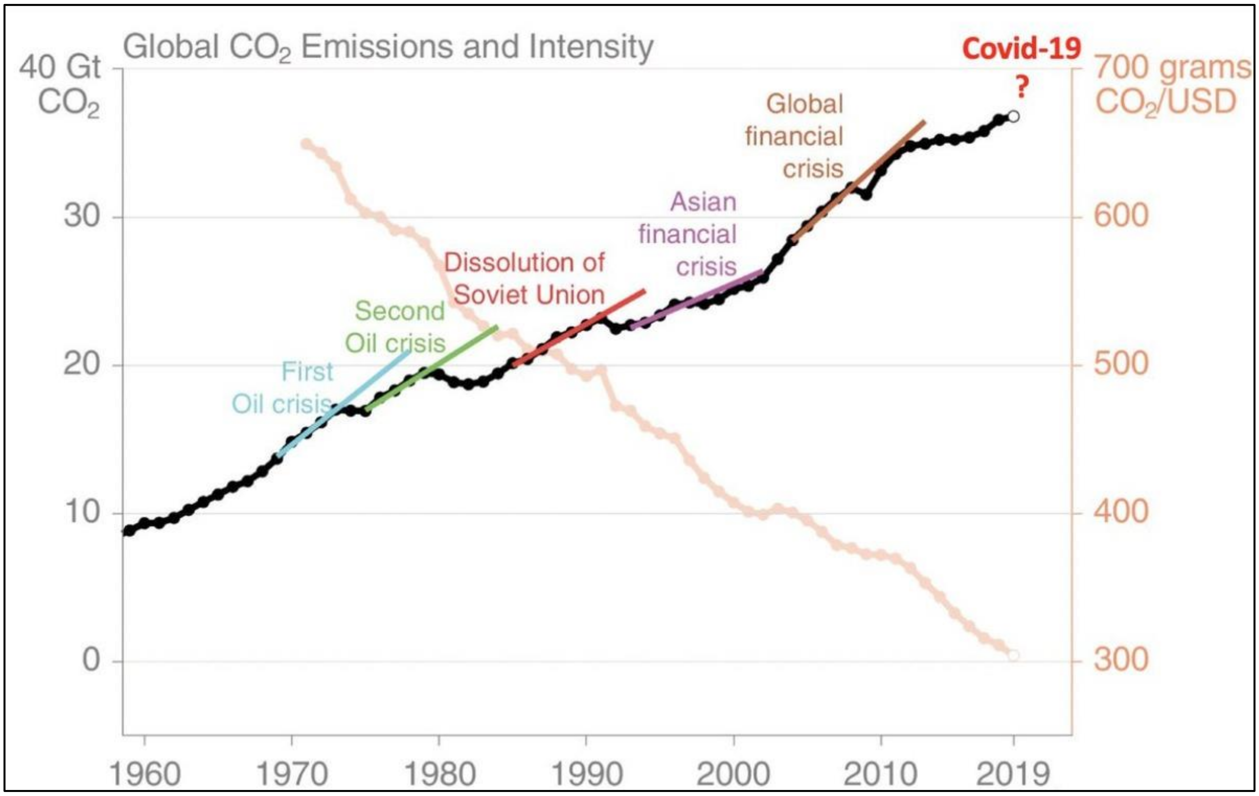
(fig 32: Support Vector Machine Regressor)

- For anticipating COVID-19 confirmed instances, this method use Support Vector Machines (SVM). It separates the dataset into two parts: training and validation, using the iloc approach with a 95:5 split. The Scikit-Learn SVM function is then used with the parameters C=1, degree=6, kernel=poly, and epsilon=0.01. It applies the SVM model to the training data and uses the predict technique to produce predictions for the validation data.
- The code then computes and appends the root mean square error (RMSE) between the predicted and actual values on the validation set to a list named model_scores. It outputs the SVM model's RMSE value.
- The code then generates a scatter plot with the Plotly Express package, displaying the confirmed cases from the training set as well as the forecasted confirmed cases from the validation set, both utilising the best-fit kernel from SVM. It also displays the projected confirmed cases for the full dataset that were forecasted using the same kernel.
- The code produces a graph displaying the verified instances for the training set as well as the anticipated confirmed cases for the validation set and the overall dataset. The graph also displays the best-fit SVM kernel and the RMSE value for the SVM model.

CONCLUSION

As we can see, COVID-19 does not have a high mortality rate, which is the most encouraging takeaway. Furthermore, a healthy Recovery Rate shows that the condition is curable. The only cause for concern is the infection's exponential growth rate.

Countries such as the United States, Spain, the United Kingdom, and Italy are having difficulty containing the pandemic, demonstrating how deadly negligence can be. The necessity of the hour is to undertake COVID-19 pandemic control practises such as testing, contact tracing, and quarantine at a rate faster than disease spread at the country level.



(fig 33: Global CO2 emissions and intensity)

The rationale for including this graph in the conclusion is that there is an intriguing pattern to observe here: every time the world's carbon emissions have decreased, the world economy has fallen. The 2008 recession is a prime example.

The growth of Confirmed and Death Cases seems to has calmed down a lot over the years, which is a really good sign. Hope this goes like that for a brief period. There should not be any new country emerging as the new epicentre of COVID-19 just like USA happened to be that epicentre for brief period. In case if any new country emerges as new epicentre, the Growth of Confirmed Cases will shoot up again.

FUTURE SCOPE

- Forecasting the pandemic's future trajectory using machine learning and statistical modelling, as well as identifying viable treatments to slow the virus's spread.
- Creating new data visualisation tools to help the public, politicians, and healthcare professionals understand COVID-19 trends and patterns.
- Integrating social and behavioural data into prediction models to better understand how human behaviour influences virus propagation and the efficiency of public health treatments.
- Analysing the pandemic's long-term economic implications, especially the effects on certain businesses and vulnerable populations.
- Examining the consequences of COVID-19 on mental health and well-being, as well as developing solutions to assist individuals and communities in dealing with the pandemic's psychological toll.
- Investigating and analysing the possible influence of novel COVID-19 vaccines, treatments, and diagnostic techniques on the path of the pandemic.
- Investigating the ethical and legal implications of COVID-19 data collecting, analysis, and sharing, including privacy, security, and transparency concerns.
- Investigating the pandemic's influence on global health security and preparedness, as well as developing methods to increase international collaboration and coordination in pandemic response.
- Investigating the role of artificial intelligence and other emerging technologies in COVID-19 response, such as contact tracing, vaccine development, and public health messaging.
- Creating effective communication methods to promote vaccination uptake and combat vaccine hesitancy, as well as determining the most effective routes for reaching certain demographics.

REFERENCES

[1] Potestio, L., Martora, F., Fabbrocini, G., Battista, T., & Megna, M. (2023). Safety and Efficacy of Covid-19 Vaccination in Patients Undergoing Biological Treatments for Psoriasis. *Psoriasis: Targets and Therapy*, 11-18.

[2] Qin, B., Hu, Z., Liu, W., Sun, Y., Wang, Y., Zhang, H., ... & He, Y. (2023). Anxiety and Depression Among College Students During the COVID-19 Lockdown: A Cross-Sectional Survey in Jiangsu, China. *Risk Management and Healthcare Policy*, 711-723.

[3] Kajjimu, J., Dreifuss, H., Tagg, A., Dreifuss, B., & Bongomin, F. (2023). Undergraduate Learning in the COVID-19 Pandemic: Lessons Learned and Ways Forward. *Advances in Medical Education and Practice*, 355-361.

[4] Tan, Cedric C. S.; Lam, Su Datt; Richard, Damien; et al. (27 May 2022). "Transmission of SARS-CoV-2 from humans to animals and potential host adaptation". Nature Communications. 13 (1): 2988. Bibcode:2022NatCo..13.2988T. doi:10.1038/s41467-022-30698-6. ISSN 2041-1723. PMC 9142586. PMID 35624123. Retrieved 28 February 2023.

[5] Patil, S. B. (2020). Indian Publications on SARS-CoV-2: A bibliometric study of WHO COVID-19 database. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*(5), 1171-1178.

[6] Sohn, K. H., Lee, M. N., Sim, D. W., Kim, S., Cho, Y. S., Kwon, H. S., & Kim, S. H. (2022). Impact of Wearing Face Masks on Patients with Severe Asthma During the COVID-19 Pandemic. *Journal of Asthma and Allergy*, 1065-1068.

[7] Commissioners of the Lancet Covid-19 Commission, Task Force CamotLC-C, Commission SaSotLC-C, Commissioners of the Lancet Covid-19 Commission, Colloborators:, Reddy KS. Priorities for the COVID-19 pandemic at the start of 2021: statement of the Lancet COVID-19 Commission. Lancet. 2021;397(10278):947-50

[8] Cascella, M., Rajnik, M., Aleem, A., Dulebohn, S., & Di Napoli, R. (2023). Features, evaluation, and treatment of coronavirus (COVID-19). *StatPearls*.