

**ENGINEERING CLINICS**

Report submitted by,

**Jayanthi 18BCS075**

**Deepthi 18BCS111**

**Shilpa Shri 18BCS122**

**Esaiarasi 18BCS105**

**Reena Dharshini 18BCS099**

**Department of Computer Science and Engineering**

**Kumaraguru College of Technology**

**Chinnavedampatti**

**Coimbatore**

**DECEMBER 2020**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S. NO** | **PAGE NO.** | **CONTENTS** |
| **1** | **2** | **Introduction** |
| **2** | **3** | **Dataset details** |
| **3** | **4** | **How to predict using Data Visualization** |
| **4.1** | **5** | **No. of cases over time** |
| **4.2** | **7** | **No. of deaths over time** |
| **4.3** | **8** | **No. of recoveries over time** |
| **4.4** | **9** | **No. of active cases over time** |
| **5** | **10** | **Country Wise Comparison** |
| **5.1** | **12** | **Worldwide Comparison** |
| **5.2** | **13** | **Worldwide Confirmed cases over time** |
| **5.3** | **14** | **Worldwide Confirmed recoveries over time** |
| **5.4** | **15** | **Worldwide mortality rate over time** |
| **5.5** | **16** | **Worldwide recovery rate over time** |
| **6** | **17** | **Inference from the project** |
| **7** | **18** | **Literature Study** |
| **8** | **28** | **Reference** |
| **9** | **29** | **Source code** |
| **10** | **38** | **Dataset** |

**Introduction:**

In this period, Coronavirus cases increase day to day.So we chose this project to compare and visualize the covid cases.And to know how Data science and data analytics will help to analyze and predict the spread of Covid 19 cases.

In this project we visualize Data visualization which means it is the graphical representation of information and data and by using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data to evaluate the graph related to covid cases. Moreover Charts and graphs can feel like a comforting way of making sense of complex and overwhelming data sets. When dealing with quantitative data that is so large, data visualization helps people to feel like they can understand the world around them.

Analysis the data of the coronavirus to know the no.of cases in each country and comparing the status with each country.

The **Dataset** contains:

· Country name

· Last updated

· No: of confirmed cases

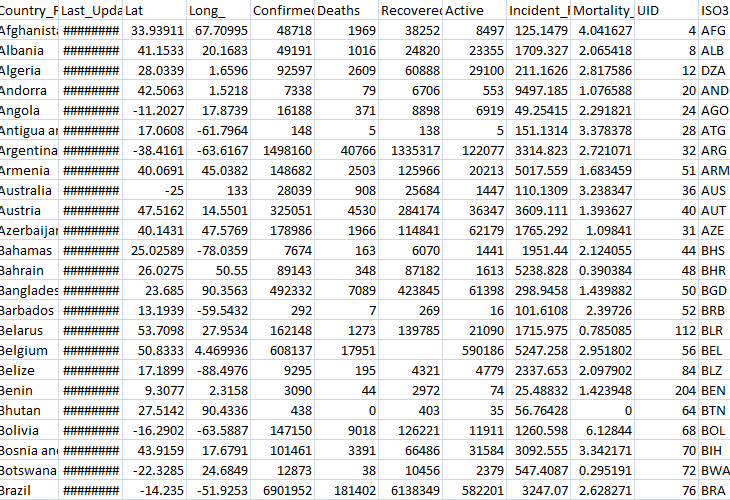
· No.of deaths

· No.of Recoveries

· No.of Active case

· Incidence rate

· Mortality rate.



Here, we compare these datasets with each country to deliver the status and show how to predict the Covid cases. And analyze the confirmed cases to understand how the recovery rate is changing over time.

Using data visualization to track the coronavirus outbreak is:

Being able to visualize the spread of the virus can help raise awareness, reveal its impact, and ultimately assist in prevention efforts.

**How to predict using Data Visualization:**

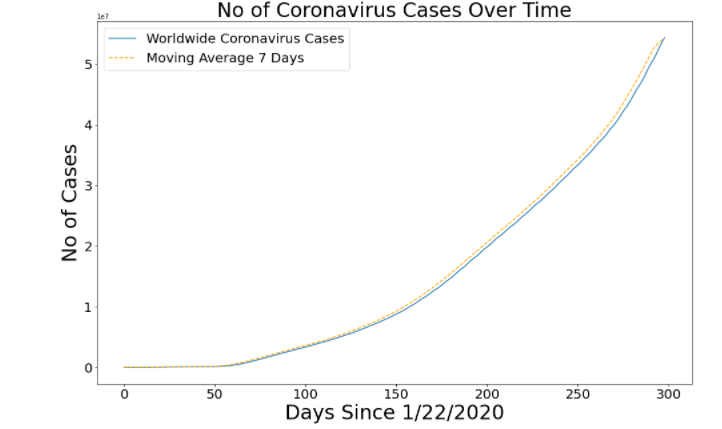
Some of these measures to predict the Covid Cases are

· Compare the Cases with the country

· Predict the mortality rate

· Predict the possible recovery rate.

With the help of graphs and charts,we analyze to predict the cases and visualize the data through data analytics and science.



2019-2020 Novel coronavirus outbreak -In early December 2019 ,a new Coronavirus designated SARS-COV.2 was identified in wuhan,china on March 11,2020 had now identified as global pandemic by world health organisation

The breadth and depth of data visualisation offered by the team at Our World in Data has continued to expand and now includes a data explorer.

The team has also incorporated the data collected by the Blavatnik School of Government tracking of the international responses to the pandemic.

#### 

#### 

#### **RESPONSES**

1. The reported data is subject to delays in aggregation and completeness.

2. There are varying levels of detail from different authorities and this variability may compound as the outbreak proceeds.

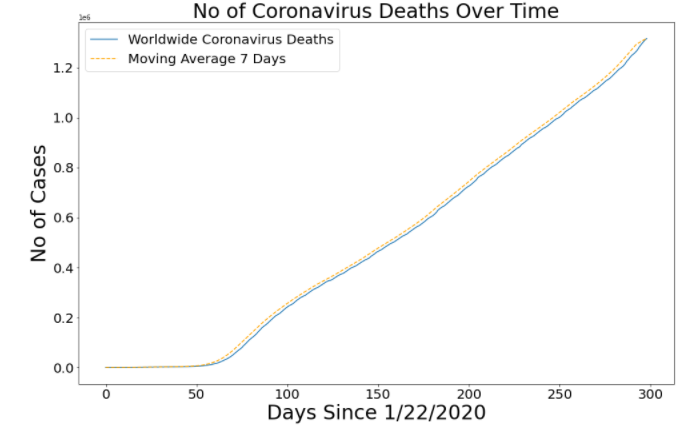
3. Countries which experience surges in suspected cases of COVID-19 may be unable to test all suspected cases.

Some countries have prioritised the testing of only the severely ill and hospitalised patients.

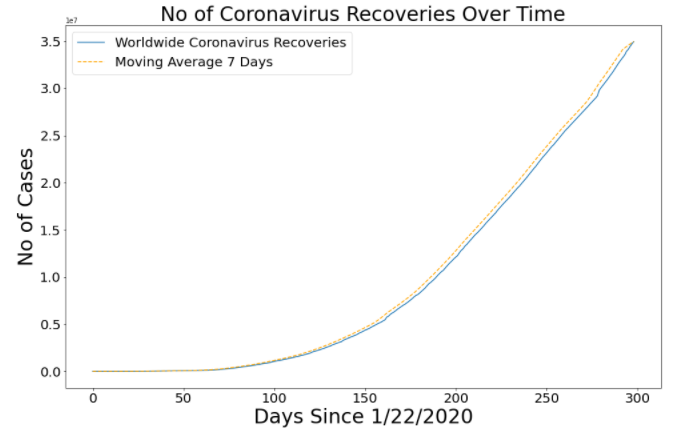
The number of infections in the community in these countries may be higher than officially reported.

4. Most people do recover from COVID-19 but most countries are no longer reporting the number of people who have recovered from illness.

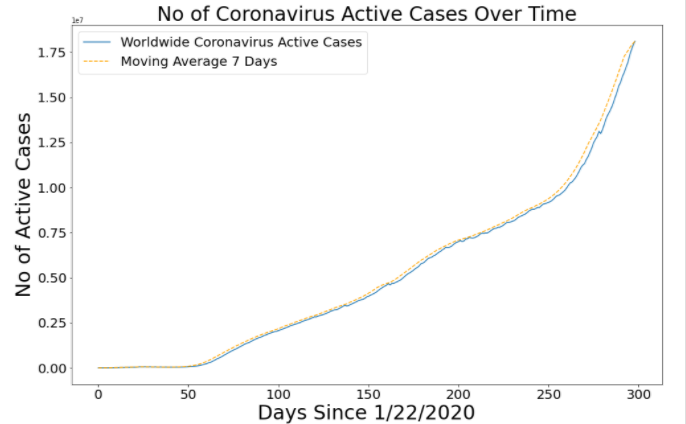
5. Direct comparison between countries at different levels of activity may give rise to erroneous conclusions.



The above graph was generated to visualise the number of deaths over time.



The above graph depicts and shows the recovery over time.



This graph shows the number of active cases over a certain period of time.

The world has seen in 2020 an unprecedented global outbreak , a new strain of coronavirus, causing the COVID-19 pandemic, and radically changing our lives and work conditions.

Here we have implemented data visualization methods.The results reveal intriguing information including increased efforts in topics such as active case,recovery, deaths etc.,

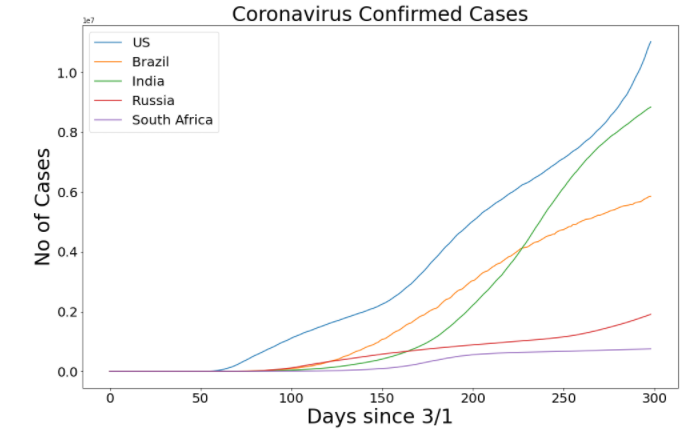
It was observed from all the graph that the number of COVID 19 active cases also increased,recovery from the pandemic was also at a high level and the deaths were also seen to be high in all the months.

The methodology will improve future large volume visualisation and discovery systems but also hope our visualisation interfaces will currently aid researchers and the general public to tackle the numerous issues in the fight against the COVID-19 pandemic.COVID-19 trend analysis to provide a top-down and bottom-up approach and provided an open interface for discovering COVID-19 research with the aim to aid in solving various issues , we present our methodology for visualising research information of a large volume.

**COUNTRY WISE COMPARISON:**

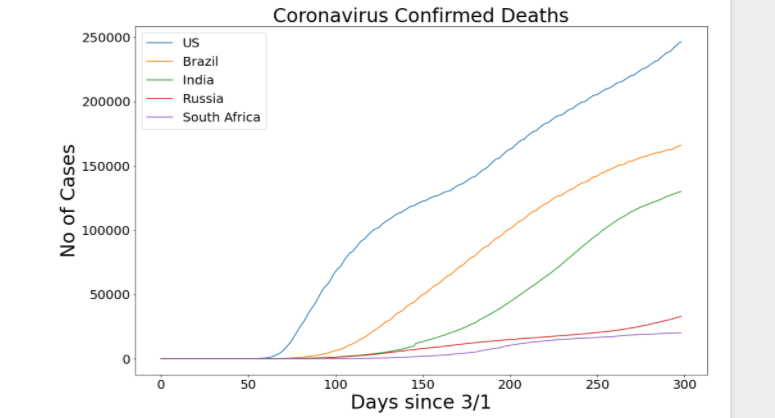
Data comparison was done with real time data sets of five chosen countries:

* India
* Brazil
* USA
* Russia
* South Africa



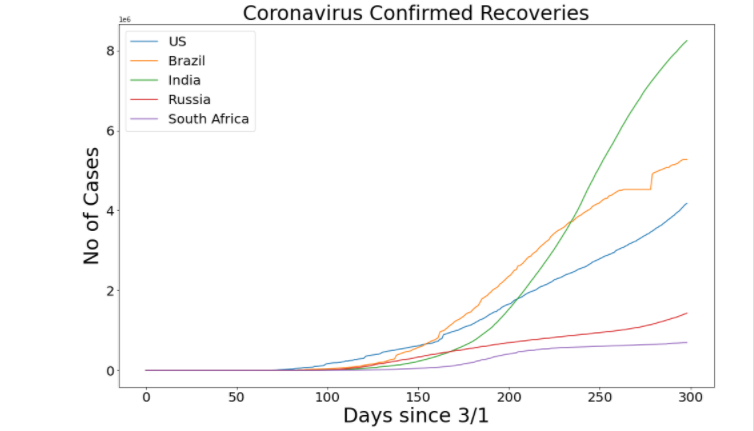
**What can we infer?**

* The population and the number of CoronaVirus cases are not directly proportional. Though India’s population is more compared to the United States, the number of cases of the United States is more compared to India



**What can we infer?**

* Though the US has more advanced hospital and medical facilities compared to India, the number of deaths has been more in the US and Brazil compared to that of India. The increase in confirmed cases and confirmed deaths are directly proportional and in this case and are on an increase.



**What can we infer:**

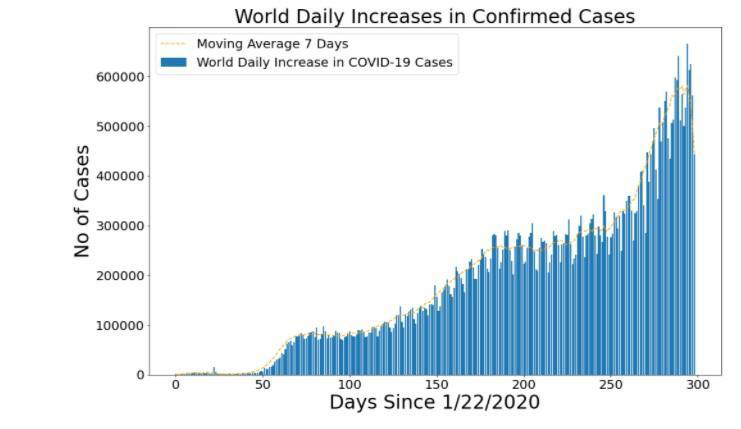
* India has the highest recovery rate compared to the other four countries. The second highest recovery rate is of Brazil.

**WORLDWIDE COMPARISON**

Worldwide daily Increase graphs represent the deaths, recoveries, and confirmed cases across the world for a better understanding. We have used the Line chart to know the increase and decrease of the daily dataset. The x-axis represents the Days since 1/22/2020 and the y-axis represents the No.of cases. We have used matplotlib.pyplot to plot the graph.

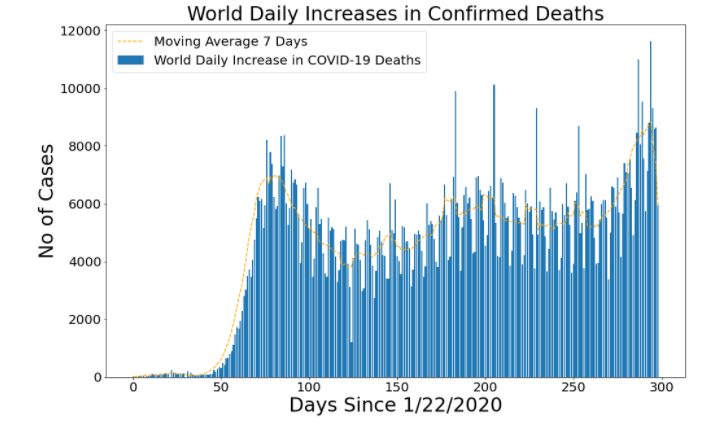
Worldwide daily Increase in confirmed cases

As it is noticed the cases worldwide are increasing slowly from 1/22/2020 till Dec 2020. X-axis value is taken for 50 days of interval for the understanding of the third party person. The increased line is compared with the Moving Average of 7 days i.e a week to understand the increase of confirmed cases and analyze it better.



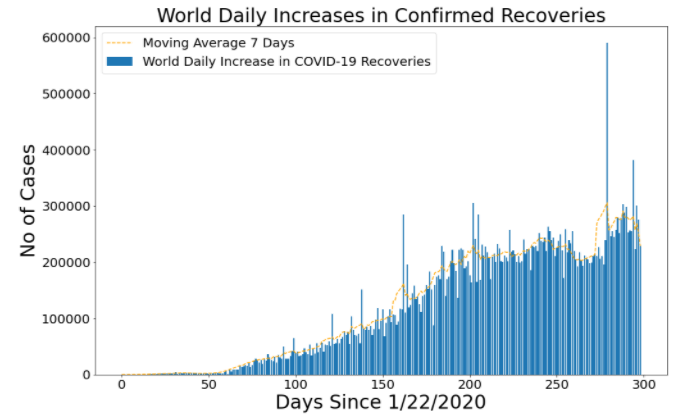
**Worldwide daily Increase in Confirmed Deaths:**

As it is noticed the death worldwide was very less at the starting i.e from 1/22/2020 but gradually the deaths have increased in the range of 50 -100 but in between 100 -150 there has been a decrease and after 150 there has been a fluctuation of confirmed Deaths from high to low till Dec 2020. The confirmed deaths is compared with the moving average of 7 days to analyze the deaths worldwide better.



**Worldwide daily Increase in Confirmed recoveries:**

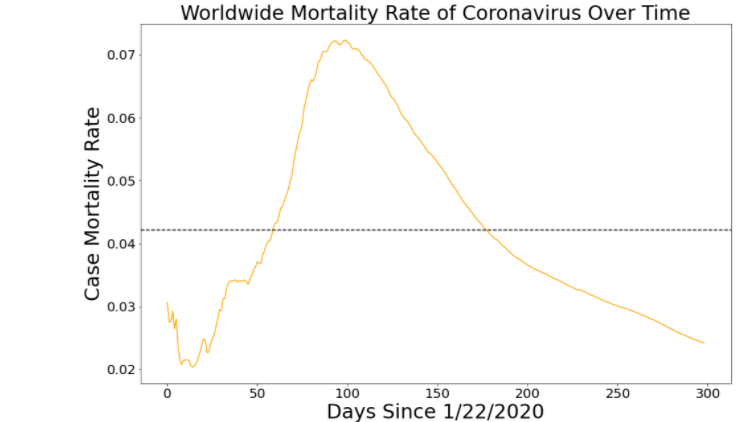
Confirmed recoveries have increased gradually from 1/22/2020 till Dec 2020. In the range between 250-300, there was a high recovery worldwide but after that, it decreased a bit. As the graph shows there has been a little bit of fluctuation between the high and low. But overall the recoveries have been increased compared to the starting stage of the COVID-19.The confirmed recoveries is compared with the moving average of 7 days to analyze the recoveries worldwide better.



Worldwide Mortality rate and recovery rate over time graph represent the mortality rate and recovery rate worldwide. The rate graph is plotted by finding the mean mortality rate and recovery rate np.mean(). For the graph matplotlib.pyplot is used.

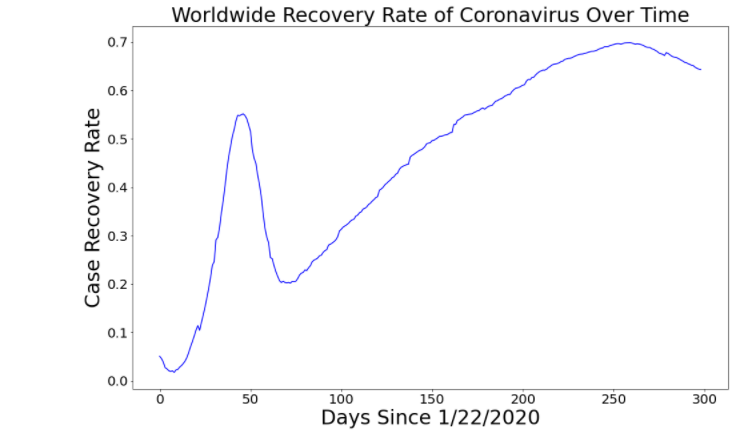
**Worldwide Mortality rate over time:**

The graph tells that in the starting the rate increases a bit in the range of 50 -150 the rate reaches to the peaks and after 150 the death rate has decreased compared to the previous days.



**Worldwide Recovery rate over time:**

The graph represents the recovery rate the range between 0-50 the recovery rate has reached the high but in the between of 50-100 the recovery rate is decreased a lot after the drop the rate has been constantly increasing and till the Dec2020 the recovery has increased a lot compared to the starting stage



**FINAL INFERENCES FROM THE PROJECT:**

* The population and the number of CoronaVirus cases are not directly proportional. Though India’s population is more compared to the United States, the number of cases of the United States is more compared to India
* Though the US has more advanced hospital and medical facilities compared to India, the number of deaths has been more in the US and Brazil compared to that of India. The increase in confirmed cases and confirmed deaths are directly proportional and in this case and are on an increase.
* India has the highest recovery rate compared to the other four countries. The second highest recovery rate is of Brazil.
* There has been a worldwide increase in the number of cases. During the start the number of cases were low, then during the peak it spiked high and eventually after the lockdown measures, it reduced to a small extent.
* The mortality rate was high during the peak spread and it dropped and reduced to very low compared to the start
* The recovery rate has comparatively increased compared to the starting months of the pandemic,

**LITERATURE STUDY**

6 papers were chosen on Covid 19 Analysis and were analysed for Literature Study

**Paper 1:**

<https://www.computer.org/csdl/magazine/cs/2020/06/09222822/1nTpVintIu4>

**AIM:**

Visualization techniques have been front-and-center in the efforts to communicate the science around COVID-19 to the very broad audience of policy makers, scientists, healthcare providers, and the general public. The deadly impact of COVID-19 is driving a massive amount of research that aims at understanding the various characteristics of the pandemic. While there is no vaccine, considerable effort has been devoted to understanding the spread of the disease in different places in the world. The speed with which the disease has spread throughout the world demands agile solutions to understand and estimate the disease progression. In this article, summarize and illustrate with examples how visualization can help understand different aspects of the pandemic.

**SUMMARY OF THE PAPER:**

Dashboards and interactive tools to analyze COVID-19 data:

Johns Hopkins University Dashboard, The New York Times charts (used with permission), and INF-UFRGS Brazil Dashboard.

Johns Hopkins University (JHU) was the first to track and display information on cases and death totals for different countries and states in the United States. Along with lists of total counts and histograms, a *bubble map* composed of circles of different radii allows a visual inspection of how serious the pandemic is around the world. This representation is useful to communicate trends, such as the average daily counts for the past week.He showshow *multidimensional projections* and *network visualizations* canhelpthe *literature exploration* ofpapers that describe novel coronavirus research.

(Top) Heatmaps matrices are useful for comparing time series such as the total deaths for different countries. Columns can be aligned by the first date after reaching a certain threshold, which allows us to compare when countries passed through specific checkpoints. (Bottom) Searching for places with similar timelines of deaths to Italy.He shows the results of searching for regions similar to Italy concerning the number of deaths. The results are listed in a ranking, with pairwise comparisons that detail attributes of the different locations and evolution charts, aligned by the day of the first death.

The applications used to study the author for Covid19 visualization are: Literature exploration, contact tracing, spread of claims and fact checking, and simulation of the transport and spread of the novel coronavirus.

*Contact tracing* is another application that relies on graph visualization to trace the network of people who may have been in contact with a COVID-19 patient, an activity essential to control the dissemination of the disease and essential for directing social distancing regulations.

**CONCLUSION:**

The above paragraph author explains the importance of data visualization in the analysis of scientific simulations and shows the visualization of simulating the transport and spread of novel coronavirus in closed spaces, which shows how an infected person can disseminate the virus indoors.

**PAPER 2:**

Analysis of COVID-19 Impact using Data Visualization. (10 pages)

**AIM:**

The world has suffered from many crises and pandemics in the past but it’s the creativity and inventiveness of its people and their rigorous efforts with the capacity to think out of the box which has made them combat and overcome those situations. The world is facing a similar situation with the inception of COVID-19. The deadly new coronavirus first detected in Wuhan; the capital of China’s Hubei province has sat its foot across the globe by infecting millions of people worldwide. Nevertheless, tough times ask for draconian measures and smart solutions, so the objective of our research work is to analyze the data with the help of visualization using Python to delineate and bring out a result by comparing the COVID-19 outbreak in different continents and countries like the USA, China, India, Italy, and Taiwan. For differentiation, we have used the total number of confirmed cases, the total number of casualties, and the total number of recovered cases of the COVID-19.Additionally, we have also compared the total number of tests conducted by the countries mentioned above. Our research will also concentrate on what makes COVID-19 pandemic different from other epidemics like SARS (2003), MERS (2012), Ebola (2014) by comparing their mortality rate, contagiousness, and symptoms. For all the comparison we have used data visualization.This research provides a comprehensive understanding of COVID-19 and compares its impact on different regions of the world with the help of Data Visualization and it will also help to derive a better solution for future emergencies.

**SUMMARY OF THE PAPER:**

The COVID-19 pandemic has become an insurmountable problem for the world's people, so this prompts us to find out the different impacts of COVID-19 in different regions due to the factors stated above. We will be differentiating the total number of cases, casualties, recoveries, lockdown dates, and tests conducted in different continents and countries to gain better insight into the effects of the COVID-19 across the globe. This analysis is based on the information under balanced periodic panel from the advent of the COVID-19 virus aka coronavirus to April 15, 2020, this study has incorporated the official information disclosed by government websites of different factors like lockdown period, mortality rate, contagiousness, due to which there is the different impact of this virus on different regions of the world. Two key aspects are the data source and representation of data.The datasets taken is the total number of cases, the total number of recoveries, the total number of sufferers died, the total number of tests taken, and the lockdown dates. Additionally, we have taken a dataset to compare various epidemics which include features like contagiousness, the total number of cases, and the mortality rate.

Data Visualization: In this paper, we’ve created a descriptive model by analyzing the collected data and applying statistical analysis to it, which encompasses the collection, analysis, interpretation, presentation, and modeling of data. This model classifies the COVID-19 pandemic in 2 different categories i.e. at a country level and at the continent level. Both of these categories have some common features and some unique features and we have drawn different conclusions from the collected data using inferential analysis which is the part of statistical analysis. The future application of the information presented is promising with many possible ideas in the eradication of the COVID-19 and control of the situation. In future, we can use many different kinds of technology to change the idea of dealing with these problems and situations like we can achieve the classification of the disease, scanning the symptoms of patients, rectify the manufacturing ramifications, achieve the complexity of chemical mixing easily and accurately through artificial intelligence, machine learning, etc.

All of the ideas are just the beginning, the applications of technology go to the boundaries of imagination which are practically limitless.

**CONCLUSION:**

In this paper, firstly we demonstrated the total number of cases of the COVID-19 along with the total recovered and casualties across the continents with the help of plots and charts. The information presented and cases evaluated above has directed us to the inference that among various continents, Europe has been severely affected by COVID-19 pandemic with the highest number of cases, and recoveries (till April 15th, 2020). Talking about countries, The USA has been severely affected by the COVID-19 as compared to the other countries.Different factors have led some countries to suffer more severely than others. After studying we have come to the conclusion that lockdown and travel restrictions have played an important role in containing COVID-19.We have also concluded that COVID-19 has the highest number of cases that means it has the highest contagiousness as compared to other epidemics and has the lowest mortality rate.

**Paper 3:**

Interactive Visualization and Simplified Pattern Discovery in the COVID-19 Open Research Dataset(CORD-19)

**AIM:**

This work explores Interactive Visualization for the COVID-19 Open Research Dataset (CORD-19) Dataset. Developed a series of easy to use online interactive text visualization based on different percentages of diseases and chemical entities from the CORD-19 Dataset.This is to enable the study of patterns based on the frequency of entities in a very large dataset of about 2.6 million disease and chemical entities extracted from 31,376 papers. This will be useful to medical professionals, especially those who are not familiar with data mining techniques to interact with diseases, symptoms, drugs and chemicals texts entities to study patterns, relationships and trends to derive insights about the COVID-19 disease from publications about the disease and similar diseases. These extracted entities will also be made publicly available so that more work can be done with the dataset.

**SUMMARY**

The COVID-19 Open Research Dataset (CORD-19) is a growing resource of scientific papers on COVID-19 and related historical coronavirus research. CORD-19 is designed to facilitate the development of text mining and information retrieval.This dataset is intended to mobilize researchers to apply recent advances in natural language processing to generate new insights in support of the fight against this infectious disease. The corpus is updated regularly as new research is published in peer-reviewed publications and archival services like bioRxiv, medRxiv, and other.This work aims to provide a simple interface for medical professionals via an interactive web-based visualization tool using Scattertext, a Python text visualization library.Each of the data point in the text visualization, when clicked shows the text data in different contexts from the CORD-19 dataset.There is also a search box where users of this tool can type in words related to diseases and chemicals they suspect may be significant to their study.E.g name of a particular medication or symptom.Data points with deeper color tone implies that the particular word has many occurrences in the mined dataset.

**CONCLUSION**

This presented a text visualization method to interact with extracted diseases and chemical entities data that was extracted using Named Entity Recognition from the COVID-19 Open Research.

**Paper 4:**

Visualising COVID-19 Research Pierre Le Bras, Azimeh Gharavi, David A. Robb, Ana F. Vidal, Stefano Padilla, and Mike J. Chantler Strategic Futures Laboratory, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK.

**AIM**

The world has seen in 2020 an unprecedented global outbreak of SARS-CoV-2, a new strain of coronavirus, causing the COVID-19 pandemic, and radically changing our lives and work conditions. Many scientists are working tirelessly to find a possible vaccine. In this paper, we develop a visualisation method.Information mapping and trend analysis, to provide a top-down and bottom-up browsing and search interface for quick discovery of topics and research resources.We apply this method on two recently released publications datasets (Dimensions’ COVID-19 dataset and the Allen Institute for AI’s CORD-19). The results revealed information including increased efforts in topics such as social distancing; cross-domain initiatives.We believe our methodology will improve future large volume visualisation and discovery systems but also hope our visualisation interfaces will currently aid scientists,researchers, and the general public to tackle the numerous issues in the fight against the COVID-19 pandemic.

**SUMMARY**

We have presented one of the first works combining analysis and visualisation of large-volume literature datasets to highlight the impact of COVID-19 on many research communities. We show that it is possible to integrate advanced statistical topic modelling techniques into a visualisation pipeline which quickly: (a) abstracts thousands of publication entries into smaller themes; (b) extracts trend information; and (c) produces at-a-glance semantic visualoverviews of rapidly changing corpora. This method, its techniques and interfaces, can help scientists browse, search,and access knowledge faster, and stay abreast of evolving themes.We have presented analysis, using topic and visual information, through different themes to summarise interesting aspects of the information inside the large volume of research literature.

**CONCLUSION**

This analysis highlights: (a) the development of research regarding social distancing for the first time in 70 years; (b) insights into cross-domain initiatives to understand the consequences of this unprecedented situation; (c) the evolution in medical topics; and (d) the unfolding of the pandemic through publications. We hope the methods and findings may be useful as a reference guide for similar systems,to stimulate new ideas and directions of research, and to help in the fight against this pandemic.

**PAPER 5:**

<https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3632420>

**AIM:**

The COVID-19 deadly virus was first reported in Wuhan, China and so far gone on to spread to more than 50 countries across the globe. WHO declared COVID-19 as a pandemic on 11th March 2020 and there has been rising cases of infection of the virus. This study presents data analysis and visualization of the confirmed, recovered, active and death cases in Nigeria. Cases were also compared to global active, confirmed, recovered and death cases. Different line graph analysis, dot graph analysis and the use of SIR model to further provide diagrammatic view and analysis of the cases in Nigeria. The model and analysis will help to interpret patterns of the virus and assist in further research and provide relevant information to the public concerning the virus.

**SUMMARY OF THE PAPER:**

WHO declared COVID-19 as a pandemic on 11th March 2020 and there has been rising cases of infection of the virus.

This study presents data analysis and visualization of the confirmed, recovered, active and death cases in Nigeria.

Cases were also compared to global active, confirmed, recovered and death cases.

Different line graph analysis, dot graph analysis and the use of SIR model to further provide diagrammatic view and analysis of the cases in Nigeria.

The model and analysis will help to interpret patterns of the virus and assist in further research and provide relevant information to the public concerning the virus.

**CONCLUSION:**

The above paragraph author explains the importance of data visualization in the analysis of scientific simulations and shows the visualization of simulating the transport and spread of novel coronavirus in closed spaces, which shows how an infected person can disseminate the virus indoors.

**PAPER 6:**

<https://journals.sagepub.com/doi/full/10.1177/2053951720939236>

**AIM:**

In response to the ubiquitous graphs and maps of COVID-19, artists, designers, data scientists, and public health officials are teaming up to create counter-plots and subaltern maps of the pandemic. In this intervention, we describe the various functions served by these projects. First, they offer tutorials and tools for both dataviz practitioners and their publics to encourage critical thinking about how COVID-19 data is sourced and modeled—and to consider which subjects are not interpellated in those data sets, and why not. Second, they demonstrate how the pandemic’s spatial logics inscribe themselves in our immediate material landscapes. And third, they remind us of our capacity to personalize and participate in the creation of meaningful COVID visualizations—many of which represent other scales and dimensions of the pandemic, especially the quarantine quotidian.

**SUMMARY**

In this intervention, they describe the various functions served by these projects.First, they offer tutorials and tools for both dataviz practitioners and their publics to encourage critical thinking about how COVID-19 data is sourced and modeled to consider which subjects are not interpellated in those data sets, and why not.Second, they demonstrate how the pandemic’s spatial logics inscribe themselves in our immediate material landscapes. And third, they remind us of our capacity to personalize and participate in the creation of meaningful COVID visualizations—many of which represent other scales and dimensions of the pandemic, especially the quarantine quotidian. And third, they remind us of our capacity to personalize and participate in the creation of meaningful COVID visualizations—many of which represent other scales and dimensions of the pandemic, especially the quarantine quotidian.

Together, the official maps and contour-plots acknowledge that the pandemic plays out differently across different scales: COVID-19 is about global supply chains and infection counts and TV ratings for presidential press conferences, but it is also about local dynamics and neighborhood mutual aid networks and personal geographies of mitigation and care.

## Indexical landscapes

Designers and artists and media-makers have also helped us recognize how our everyday environments are themselves indexing COVID’s presence. The pandemic has rendered itself visible, audible, and tangible in our material landscapes, transforming those spaces themselves into environmental data

## Participatory reflections

More personal, intimate, and participatory data projects have focused on COVID-19 data as tools for reflection. These engagements with the data of crisis draw out human (and sometimes non-human) stories and interactions and ask users to situate themselves within the overwhelming global narrative of emergency.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CONCLUSION:**

With COVID-19 data and the veritable flood of graphics being produced, we have a chance to strengthen our collective data literacy. By understanding the origins of epidemiological data and reinforcing the importance of context and non-quantitative forms of data, we can push for a richer, more diverse discourse through data visualization.

**REFERENCE:**

* <https://www.computer.org/csdl/magazine/cs/2020/06/09222822/1nTpVintIu4>
* <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3632420>
* <https://journals.sagepub.com/doi/full/10.1177/2053951720939236>
* Visualising COVID-19 Research Pierre Le Bras, Azimeh Gharavi, David A. Robb, Ana F. Vidal, Stefano Padilla, and Mike J. Chantler Strategic Futures Laboratory, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK.
* Interactive Visualization and Simplified Pattern Discovery in the COVID-19 Open Research Dataset(CORD-19) - Anonymous ACL submission
* Analysis of COVID-19 Impact using DATA Visualization Ritik Dixit,Rishika Kushwah,Samay Pashine, Computer Science and Engineering Acropolis Institute of Technology and Research indore,India

**Source Code:**

import numpy as np

import matplotlib.pyplot as plt

import matplotlib.colors as mcolors

import pandas as pd

import random

import math

import time

from sklearn.model\_selection import RandomizedSearchCV, train\_test\_split

from sklearn.preprocessing import MinMaxScaler

from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error

import datetime

import operator

%matplotlib inline

confirmed\_df = pd.read\_csv(r'C:\Users\Welcome\Desktop\New folder\New folder\time\_series\_covid\_19\_confirmed.csv')

deaths\_df = pd.read\_csv(r'C:\Users\Welcome\Desktop\New folder\New folder\time\_series\_covid\_19\_deaths.csv')

recoveries\_df = pd.read\_csv(r'C:\Users\Welcome\Desktop\New folder\New folder\time\_series\_covid\_19\_recovered.csv')

case\_country\_df= pd.read\_csv(r'C:\Users\Welcome\Desktop\New folder\New folder\cases\_country.csv')

confirmed\_df=confirmed\_df.replace(np.nan,"",regex=True)

deaths\_df=deaths\_df.replace(np.nan,"",regex=True)

recoveries\_df=recoveries\_df.replace(np.nan,"",regex=True)

case\_country\_df=case\_country\_df.replace(np.nan,"",regex=True)

cols = confirmed\_df.keys()

confirmed = confirmed\_df.loc[:, cols[4]:cols[-1]]

deaths = deaths\_df.loc[:, cols[4]:cols[-1]]

recoveries = recoveries\_df.loc[:, cols[4]:cols[-1]]

dates = confirmed.keys()

world\_cases = []

total\_deaths = []

mortality\_rate = []

recovery\_rate = []

total\_recovered = []

total\_active = []

for i in dates:

confirmed\_sum = confirmed[i].sum()

death\_sum = deaths[i].sum()

recovered\_sum = recoveries[i].sum()

# confirmed, deaths, recovered, and active

world\_cases.append(confirmed\_sum)

total\_deaths.append(death\_sum)

total\_recovered.append(recovered\_sum)

total\_active.append(confirmed\_sum-death\_sum-recovered\_sum)

# calculate rates

mortality\_rate.append(death\_sum/confirmed\_sum)

recovery\_rate.append(recovered\_sum/confirmed\_sum)

def daily\_increase(data):

d = []

for i in range(len(data)):

if i == 0:

d.append(data[0])

else:

d.append(data[i]-data[i-1])

return d

def moving\_average(data, window\_size):

moving\_average = []

for i in range(len(data)):

if i + window\_size < len(data):

moving\_average.append(np.mean(data[i:i+window\_size]))

else:

moving\_average.append(np.mean(data[i:len(data)]))

return moving\_average

# window size

window = 7

# confirmed cases

world\_daily\_increase = daily\_increase(world\_cases)

world\_confirmed\_avg= moving\_average(world\_cases, window)

world\_daily\_increase\_avg = moving\_average(world\_daily\_increase, window)

# deaths

world\_daily\_death = daily\_increase(total\_deaths)

world\_death\_avg = moving\_average(total\_deaths, window)

world\_daily\_death\_avg = moving\_average(world\_daily\_death, window)

# recoveries

world\_daily\_recovery = daily\_increase(total\_recovered)

world\_recovery\_avg = moving\_average(total\_recovered, window)

world\_daily\_recovery\_avg = moving\_average(world\_daily\_recovery, window)

# active

world\_active\_avg = moving\_average(total\_active, window)

days\_since\_1\_22 = np.array([i for i in range(len(dates))]).reshape(-1, 1)

world\_cases = np.array(world\_cases).reshape(-1, 1)

total\_deaths = np.array(total\_deaths).reshape(-1, 1)

total\_recovered = np.array(total\_recovered).reshape(-1, 1)

days\_in\_future = 10

future\_forcast = np.array([i for i in range(len(dates)+days\_in\_future)]).reshape(-1, 1)

adjusted\_dates = future\_forcast[:-10]

start = '1/22/2020'

start\_date = datetime.datetime.strptime(start, '%m/%d/%Y')

future\_forcast\_dates = []

for i in range(len(future\_forcast)):

future\_forcast\_dates.append((start\_date + datetime.timedelta(days=i)).strftime('%m/%d/%Y'))

X\_train\_confirmed, X\_test\_confirmed, y\_train\_confirmed, y\_test\_confirmed = train\_test\_split(days\_since\_1\_22[50:], world\_cases[50:], test\_size=0.05, shuffle=False)

def flatten(arr):

a = []

arr = arr.tolist()

for i in arr:

a.append(i[0])

return a

adjusted\_dates = adjusted\_dates.reshape(1, -1)[0]

plt.figure(figsize=(16, 10))

plt.plot(adjusted\_dates, world\_cases)

plt.plot(adjusted\_dates, world\_confirmed\_avg, linestyle='dashed', color='orange')

plt.title('No of Coronavirus Cases Over Time', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('No of Cases', size=30)

plt.legend(['Worldwide Coronavirus Cases', 'Moving Average {} Days'.format(window)], prop={'size': 20})

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.plot(adjusted\_dates, total\_deaths)

plt.plot(adjusted\_dates, world\_death\_avg, linestyle='dashed', color='orange')

plt.title('No of Coronavirus Deaths Over Time', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('No of Cases', size=30)

plt.legend(['Worldwide Coronavirus Deaths', 'Moving Average {} Days'.format(window)], prop={'size': 20})

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.plot(adjusted\_dates, total\_recovered)

plt.plot(adjusted\_dates, world\_recovery\_avg, linestyle='dashed', color='orange')

plt.title('No of Coronavirus Recoveries Over Time', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('No of Cases', size=30)

plt.legend(['Worldwide Coronavirus Recoveries', 'Moving Average {} Days'.format(window)], prop={'size': 20})

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.plot(adjusted\_dates, total\_active)

plt.plot(adjusted\_dates, world\_active\_avg, linestyle='dashed', color='orange')

plt.title('No of Coronavirus Active Cases Over Time', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('No of Active Cases', size=30)

plt.legend(['Worldwide Coronavirus Active Cases', 'Moving Average {} Days'.format(window)], prop={'size': 20})

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.bar(adjusted\_dates, world\_daily\_increase)

plt.plot(adjusted\_dates, world\_daily\_increase\_avg, color='orange', linestyle='dashed')

plt.title('World Daily Increases in Confirmed Cases', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('No of Cases', size=30)

plt.legend(['Moving Average {} Days'.format(window), 'World Daily Increase in COVID-19 Cases'], prop={'size': 20})

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.bar(adjusted\_dates, world\_daily\_death)

plt.plot(adjusted\_dates, world\_daily\_death\_avg, color='orange', linestyle='dashed')

plt.title('World Daily Increases in Confirmed Deaths', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('No of Cases', size=30)

plt.legend(['Moving Average {} Days'.format(window), 'World Daily Increase in COVID-19 Deaths'], prop={'size': 20})

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.bar(adjusted\_dates, world\_daily\_recovery)

plt.plot(adjusted\_dates, world\_daily\_recovery\_avg, color='orange', linestyle='dashed')

plt.title('World Daily Increases in Confirmed Recoveries', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('No of Cases', size=30)

plt.legend(['Moving Average {} Days'.format(window), 'World Daily Increase in COVID-19 Recoveries'], prop={'size': 20})

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

def country\_plot(x, y1, y2, y3, y4, country):

# window is set as 14 in in the beginning of the notebook

confirmed\_avg = moving\_average(y1, window)

confirmed\_increase\_avg = moving\_average(y2, window)

death\_increase\_avg = moving\_average(y3, window)

recovery\_increase\_avg = moving\_average(y4, window)

plt.figure(figsize=(16, 10))

plt.plot(x, y1)

plt.plot(x, confirmed\_avg, color='red', linestyle='dashed')

plt.legend(['{} Confirmed Cases'.format(country), 'Moving Average {} Days'.format(window)], prop={'size': 20})

plt.title('{} Confirmed Cases'.format(country), size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('# of Cases', size=30)

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.bar(x, y2)

plt.plot(x, confirmed\_increase\_avg, color='red', linestyle='dashed')

plt.legend(['Moving Average {} Days'.format(window), '{} Daily Increase in Confirmed Cases'.format(country)], prop={'size': 20})

plt.title('{} Daily Increases in Confirmed Cases'.format(country), size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('# of Cases', size=30)

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.bar(x, y3)

plt.plot(x, death\_increase\_avg, color='red', linestyle='dashed')

plt.legend(['Moving Average {} Days'.format(window), '{} Daily Increase in Confirmed Deaths'.format(country)], prop={'size': 20})

plt.title('{} Daily Increases in Deaths'.format(country), size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('# of Cases', size=30)

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.bar(x, y4)

plt.plot(x, recovery\_increase\_avg, color='red', linestyle='dashed')

plt.legend(['Moving Average {} Days'.format(window), '{} Daily Increase in Confirmed Recoveries'.format(country)], prop={'size': 20})

plt.title('{} Daily Increases in Recoveries'.format(country), size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('# of Cases', size=30)

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

# helper function for getting country's cases, deaths, and recoveries

def get\_country\_info(country\_name):

country\_cases = []

country\_deaths = []

country\_recoveries = []

for i in dates:

country\_cases.append(confirmed\_df[confirmed\_df['Country/Region']==country\_name][i].sum())

country\_deaths.append(deaths\_df[deaths\_df['Country/Region']==country\_name][i].sum())

country\_recoveries.append(recoveries\_df[recoveries\_df['Country/Region']==country\_name][i].sum())

return (country\_cases, country\_deaths, country\_recoveries)

def country\_visualizations(country\_name):

country\_info = get\_country\_info(country\_name)

country\_cases = country\_info[0]

country\_deaths = country\_info[1]

country\_recoveries = country\_info[2]

country\_daily\_increase = daily\_increase(country\_cases)

country\_daily\_death = daily\_increase(country\_deaths)

country\_daily\_recovery = daily\_increase(country\_recoveries)

country\_plot(adjusted\_dates, country\_cases, country\_daily\_increase, country\_daily\_death, country\_daily\_recovery, country\_name)

compare\_countries = ['US', 'Brazil', 'India', 'Russia', 'South Africa']

graph\_name = ['Coronavirus Confirmed Cases', 'Coronavirus Confirmed Deaths', 'Coronavirus Confirmed Recoveries']

for num in range(3):

plt.figure(figsize=(16, 10))

for country in compare\_countries:

plt.plot(get\_country\_info(country)[num])

plt.legend(compare\_countries, prop={'size': 20})

plt.xlabel('Days since 3/1', size=30)

plt.ylabel('No of Cases', size=30)

plt.title(graph\_name[num], size=30)

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

mean\_mortality\_rate = np.mean(mortality\_rate)

plt.figure(figsize=(16, 10))

plt.plot(adjusted\_dates, mortality\_rate, color='orange')

plt.axhline(y = mean\_mortality\_rate,linestyle='--', color='black')

plt.title('Worldwide Mortality Rate of Coronavirus Over Time', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('Case Mortality Rate', size=30)

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

mean\_recovery\_rate = np.mean(recovery\_rate)

plt.figure(figsize=(16, 10))

plt.plot(adjusted\_dates, recovery\_rate, color='blue')

plt.title('Worldwide Recovery Rate of Coronavirus Over Time', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('Case Recovery Rate', size=30)

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

plt.figure(figsize=(16, 10))

plt.plot(adjusted\_dates, total\_deaths, color='r')

plt.plot(adjusted\_dates, total\_recovered, color='green')

plt.legend(['death', 'recoveries'], loc='best', fontsize=25)

plt.title('Worldwide Coronavirus Cases', size=30)

plt.xlabel('Days Since 1/22/2020', size=30)

plt.ylabel('No of Cases', size=30)

plt.xticks(size=20)

plt.yticks(size=20)

plt.show()

confirmed\_df.columns

X\_data=confirmed\_df[['1/22/20', '1/23/20','1/24/20', '1/25/20', '1/26/20', '1/27/20','11/6/20', '11/7/20', '11/8/20', '11/9/20', '11/10/20', '11/11/20','11/12/20', '11/13/20', '11/14/20', '11/15/20']]

Y\_data=confirmed\_df['Lat']

X\_data.head()

scalar=MinMaxScaler()

scalar.fit(X\_data)

new=scalar.transform(X\_data)

new

**Dataset:**

<https://drive.google.com/drive/folders/1NTkcU8Dw1XbPbsjN8AjMPSaGrCKLhztH?usp=sharing>