

Statistical Machine Learning, Forecasting and Inferences from Pandemic Data

Devaashish Sharma

July 2, 2020

Content

		Page No
Work Done !		1
List of Figures		3-4
Chapter 1	INTRODUCTION	5-6
1.1	Introduction to work done and motivation	5
1.2	Project Statement and objective	5
1.3	Organization of Report	6
Chapter 2	BACKGROUND MATERIAL	7-16
2.1	Conceptual Overview	7-8
2.2	Technologies Involved	8-16
Chapter 3	METHODOLOGY	17-18
3.1	Methodology adopted and Block Diagram	17-18
Chapter 4	IMPLEMENTATION	19-29
4.1	Modules and Evaluation Report	19-29
Chapter 5	RESULTS AND ANALYSIS	30-33
5.1	Machine Learning and Prediction	30-33
Chapter 6	CONCLUSIONS & FUTURE SCOPE	34
REFERENCES		35

Work Done !

Pandemic(pan-de-mic) year 2020, the year of Rat, the first in the twelve-year cycle according to the Chinese zodiac calendar. But, no one knew that it would force everyone to live like one for months because of the spread of an airborne virus called the nCoV-19 (Novel Coronavirus) which would claim thousands of lives and infect millions of people to struggle with severe respiratory disorders. It's a large family of viruses that causes severe diseases such as Middle East Respiratory Syndrome (MERS-CoV), Severe Acute Respiratory Syndrome (SARS-CoV), since the behaviour of the virus reported in humans was different than previously encountered cases it was called 'novel' meaning new. It's believed to be zoonotic in nature means it can transmit between animals and humans. And its symptoms in human include sore throat, fever, shortness of breath, dry cough and pneumonia severe case, death. The source of coronavirus is believed to be a wet market in Wuhan city of Hubei Province, China. After which there has been a chain reaction contributory to global infected population. Walloping developed countries like United States, India, France, Italy, Spain, Brazil, Russia and many more.

Taking into account the seriousness of the situation WHO (World Health organisation) an independent and integral functioning body of UN (United Nations) responsible for monitoring global public health declared it a pandemic on 11th March 2020. Which is considered to be a delayed decisive action. Which makes the authorities of WHO questionable. Considering the outspread of the disease, and no availability of any kind prescribed drug or vaccine for prevention, countries globally imposed nationwide lockdown with the intension to contain and break the chain reaction of transmission from one person to another. The study analyses the current global scenario and generate a prediction on the next ten days from the last collected data on 16th June.

The analytics and data modelling is done on MacOS Cataline version 10.15.5 operating system using Python version 3.8 scripting on Jupiter Lab, utilising widely used standard scientific computational third party libraries, like matplotlib, sklearn for modelling and prediction bases on the dataset, and all the graphs used, have been derived from the analytics, unless mentioned. The second paragraph should briefly discuss the methodology that was adopted.

LIST OF FIGURES

Figure No	Figure Title	Page No
1	Depicting the Respiratory System and organ morphology inside the body	7
2	JupyterLab	15
3	QGIS 3.10 LTS	15
4	Showing graph post importing map.shp file in Python	16
5	Showing Data Life Cycle	17
6	Showing top ten countries with confirmed COVID-19 cases on linear scale	19
7	Showing top ten countries with confirmed COVID-19 cases on exponential scale	20
8	Showing global situation as of 22 nd January 2020	21
9	Showing global situation as of 16 th February 2020	21
10	Showing global situation as of 16 th March 2020	21
11	Showing Global Situation as of 16 th April 2020	22
12	Showing Global Situation as of 16 th May 2020	22
13	Showing Global Situation as of 16 th June 2020	22
14	Pie Chart depicting a total percentage of Confirmed, Dead and Recovered cases as on 16 th June 2020	23
15	Country wise situation report sorted based on confirmed cases	24
16	Graph of Global Confirmed (in Blue), Global Deaths (in Red) and Global Recovered (in Green) on linear scale	25
17	Graph of Global Confirmed (in Blue), Global Deaths (in Red) and Global Recovered (in Green) on exponential scale	26
18	Comparing the situation report of US being on top with the most confirmed cases and cumulative sum of other countries as of 16 th June 2020	27
19	Comparison of situation reports inclusive of top ten countries and cumulative sum of other countries entitled 'Others' as their cumulative confirmed cases	27
20	Pie Chart showing the percentage of total confirmed cases	28
21	Showing the stagewise ground report in India as of 16 th June 2020 Most to Least affected, from dark to light	29
22	Shows the predicted values from SVM Regression model in green dashed lines and the confirmed values in blue on test data	30
23	Actual confirmed cases in blue line and prediction values via purple dashed line from the SVM Regression model	30
24	Shows the forecast values for the next ten days based on SVM regression model	31
25	Shows the actual confirmed cases via blue line and the prediction values using green dashed line from linear regression.	31
26	Depicts the linear regression model in yellow dashed line and the confirmed cases in blue line on the graph (on linear scale)	32
27	Depicts the linear regression model in yellow dashed line vs confirmed cases in the blue on exponential scale.	32
28	Shows the prediction values for next 10 days from which we can observe that it also indicates an increment in future new positive cases	33

29	Shows the global mortality rate (Total Deaths/ Total Confirmed) over time in yellow while the black dashed line represents the mean mortality rate over the timeline.	33
----	---	----

Chapter 1 Introduction

1.1 *Introduction to the work done and motivation*

(Pan-de-mic) year 2020, the year of Rat, the first in the twelve-year cycle according the Chinese zodiac calendar. But, no one knew it would force everyone to live like one because of a microorganism that one can't see with bare eyes. We have often heard on our childhood stories than even an ant in the trunk of an elephant can cause serious trouble to it. And it seems to be analogous to the spread of the coronavirus outbreak. Coronavirus is a family of viruses which have been known in the past for causing serious respiratory disorders in the past such as Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV). Coronaviruses are zoonotic in nature, i.e. can be transmitted between animals and human being.

Although, the virus responsible for the outbreak of global pandemic is not analogues to the previous cases and resembles more closely found in bats and so it was named nCoV-19. Novel meaning new coronavirus. Its first case was reported in December 2019 as an unknown cause of pneumonia. It's source of origin is believed to be a wet market in Wuhan city of Hubei province, China. Later it was realized that the virus can be airborne and can transmit from one person to another primarily through droplets of saliva while speaking or discharge from the nose when an infected person coughs or sneezes. So, it becomes important that one also learns respiratory etiquettes like coughing into flexed elbow rather than covering with hands or handkerchief.

At the current time we have no potential prescribed drug or vaccine for treatment or prevention from the infection and therefor it's important that we maintain proper health and hygiene in our routine life. The symptoms of coronavirus in humans include fever, cough, shortness of breath, body aches, sore throat, loss of taste or smell, runny nose, nausea, diarrhea, pneumonia and death in severe cases.

Data collection and data wrangling is performed to formulate a prediction model for future forecast on possible new infection cases, deaths and recoveries based on statistical and machine learning approach. And analysis on top ten countries who are currently walloped by lack of medical and health care facilities. The demand doesn't meet the requirement of the market. And a theoretical analysis on its impact on economy and job market.

1.2 *Project Statement and Objective*

In the past also there have been cases of code blue, medical term implying immediate medical attention and critical care. After World War I (1914-1918) there was an outbreak of Spanish Flu, named Spanish flu since Spain was the first to accept the outbreak. Which was caused by H1N1 influenza virus. It would cause agony to the extent as if all the bones in the body are on the verge to break or fracture. And the skin tone would turn blue, then purple and black in the end. It's often factualized to have claimed more life's than both world wars combined. It eradicated almost a third

of the then population on the globe. And approximately six percent of the then population in India. On the worst side, India also faced its worst drought of all time during that period. And the death rate of women was comparatively more compared to men. Although, its live virus vaccine came into existence during 2009 influenza, commonly referred as Swine Flu. A doze of which would help the body to develop immunity just like a firewall to develop resistance against it. There are many more examples that can be examplated from the 20th century and early 21st century like Swine Flu, Bird Flu, Bubonic Plague often referred as Black Death.

But, the most common situation in the past was the lack of involvement of technology in monitoring and recording of the data for further speculation and study. So, in order to monitor the current situation, it's important to evaluate the patterns that one can find during their pandemic study and the most important of all is data collection and prediction on what might be coming up next, so we can possibly meet the demands and on time.

The web scraping is done based on vibrant global reports repositories but the main concern still lies underneath it is those are the cases that have been affirmatively tested positive on the coronavirus disease but there are many more people out there who might be asymptomatic and do poses as a threat of acting as a potential vector to the spread of the disease. As, mentioned the virus is airborne so. While coughing and sneezing the tiny water droplets may land on tables, furniture, doorknobs from where it can spread from one person to another if they come directly in physical contact with them. So, it's important to maintain physical distancing as well so the chain reaction to the spread of infection can be controlled, monitored and contained easily.

1.3 *Organization of Report*

In this report Chapter 2 discussed the bases theory and the ideas perceived behind the spread of coronavirus and the technological knowledge and utilised libraries are shared. Chapter 3 Describes briefly about the methodology adopted and the data flow diagram from problem understanding to concluding. Chapter 4 the most interesting part implementation and graphs withdrawn from the data itself making it easy to derive conclusions. Chapter 5 Discusses briefly about the conclusions withdrawn from the graphs and machine learning based prediction. Chapter 6 Introspects into the futuristic plans on coronavirus and discussed about general health and proper sanity one should maintain to avoid any incidents of infection.

Chapter 2 Background Knowledge

2.1 Conceptual Overview

So, it was clear in China by the end of December 2019 that there was an underlying cause for causing pneumonia and this can't be coincidental that the number of new infected individuals rose exponentially showing up similar symptoms. So, it's tedious and somewhat difficult to realize the situation at the beginning. But now we can summarize the whole process if someone gets infected with the virus.

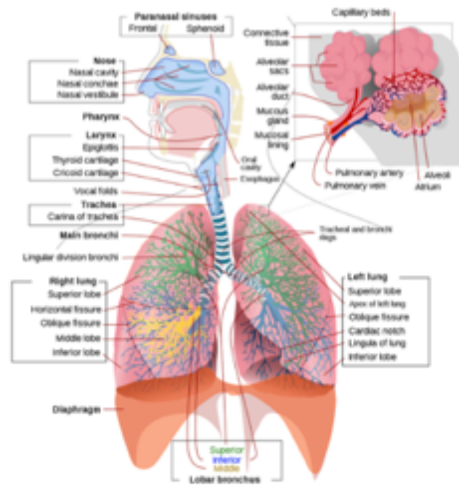


Fig Depicting Human Respiratory System,

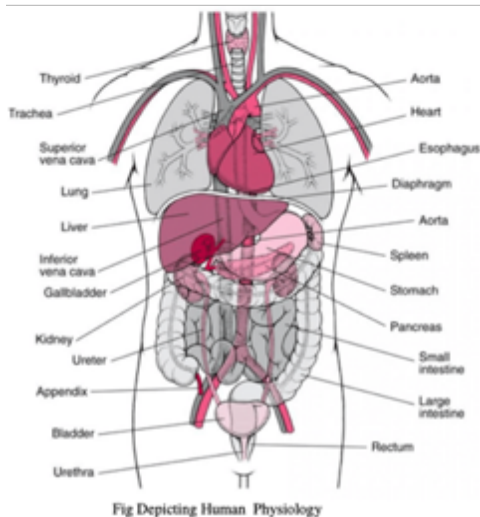
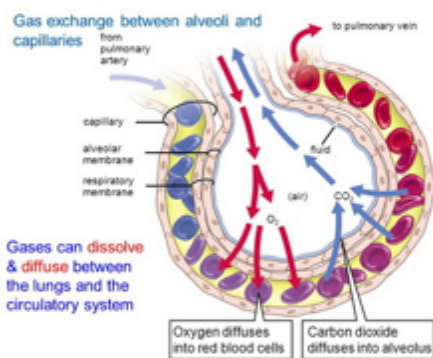


Fig Depicting Human Physiology

Considering the possibility of someone getting infected by corona virus. It can take upto fourteen days to show up symptoms like fever, cough, shortness of breath, rashes on skin, or discoloration of fingers or toes, body aches, sore throat, loss of taste or smell, runny nose, nausea, diarrhea, pneumonia and death in severe cases. As the virus is airborne and can be transmitted via small droplets from nose, or saliva while speaking or coughing. The single stranded RNA genome carrier, virus hijacks the oral cavity cells lying below the mucus membrane of the throat from where it replicated itself and generated multiple identical copies of itself. The replicated copies thus start affecting other nearby cells and starts to spread down the throat, from where it can enter the lungs and down the trachea. Which also triggers the immune system of our body against foreign infection. As the virus resides inside the host cells, the immune system starts to destroy its own cells can be classified as an autoimmune disease or scientifically called as cytokine storm. At this point the patient starts to feel difficulty breathing, shortness of breath, chest pains or pressure, loss of speech or movement. As the infection becomes more aggressive the immune system targets the alveoli which is the focus point for gas exchange between the respiratory system and the circulatory system. As the lungs gets filled with fluid due to autoimmune disorder the person starts to feel symptoms of pneumonia.

Figure1 Depicting the Respiratory System and organ morphology inside the body

Which is fatal and can cause multiple organ failure primarily heart, liver, kidney etc., due or insufficient oxygen supply. At this point the patient is required to be attached to a ventilator for oxygen support system. It can spread to nearby organs too like the heart, can reach to other organs via their respective oxygen supply. Additionally, as the immune cells accumulate near the lungs it can cause vascular trauma or blood clots in the blood stream, scientifically defined as thrombosis.

Even after someone recovers from the corona virus the recovery is mostly unfeasible. And periodic checkup for damage assessment is required. It's being observed that men are more susceptible to the virus than women, probably because they are more involved in earning for the family and therefore, are more prone to get in contact with an infected individual.[1] Whereas children are less susceptible to the disease and can be asymptomatic to it as their immune system is not so well developed compared to adults. In the past due to lack of technology and scientific foresight it was hard to determine the disease and difficult to diagnose. So, in order to contain the spread of the pervasive disease the whole village were burnt down so that it would perish with the fire and turn into ashes. But, with the advancement of technology and accessibility of mobile phone in every hand we can generate an early warning system near to the probable red, orange or containment zone.

2.2 *Technologies Involved*

Python 3.8.2 is used to develop and derive all the scripts and conclusions. Python package manager, Preferred Installer Program (pip) is used to install all the third-party scientific development libraries. It doesn't withdraw memory over RAM like anaconda which is more of a software for developing Python, R, Scala, Julia and other programming language projects. Whereas pip provides the functionality of install package dependencies and creating virtual environments utilizing which we can segregate the package dependencies of different projects can host a master child package management system using virtualenv python package.

List of project dependencies:

Z shell terminal command execution output: `pip list >> project_pkg_dependency.txt`

This will save all the utilized packages inside a file named `project_pkg_dependency.txt`

Package	Version
-----	-----
alabaster	0.7.12
applaunchservices	0.2.1
appnope	0.1.0
astroid	2.4.2
atomicwrites	1.4.0
attrs	19.3.0
autopep8	1.5.3
Babel	2.8.0
backcall	0.2.0
bleach	3.1.5
bokeh	2.1.0

branca	0.4.1
Brotli	1.0.7
Certify	2020.4.5.2
chardet	3.0.4
click	7.1.2
click-plugins	1.1.1
cligj	0.5.0
cloudpickle	1.4.1
cmdstanpy	0.4.0
colorlover	0.3.0
convertdate	2.2.1
cufflinks	0.17.3
cycler	0.10.0
Cython	0.29.20
dash	1.12.0
dash-core-components	1.10.0
dash-html-components	1.0.3
dash-renderer	1.4.1
dash-table	4.7.0
decorator	4.4.2
defusedxml	0.6.0
Deprecated	1.2.10
9ehaviour	1.1.0
diff-match-patch	20181111
docutils	0.16
entrypoints	0.3
ephem	3.7.7.1
fbprophet	0.6
Fiona	1.8.13.post1
flake8	3.8.3
Flask	1.1.2
Flask-Compress	1.5.0
folium	0.11.0
future	0.18.2
geopandas	0.7.0
helpdev	0.7.1
holidays	0.10.2
idna	2.9
imagesize	1.2.0
intervaltree	3.0.2
ipykernel	5.3.0
ipython	7.15.0
ipython-genutils	0.2.0
ipywidgets	7.5.1
isort	4.3.21
itsdangerous	1.1.0
jedi	0.15.2

Jinja2	2.11.2
joblib	0.15.1
json5	0.9.5
jsonschem	3.2.0
jupyter-client	6.1.3
jupyter-core	4.6.3
jupyterlab	2.1.4
jupyterlab-server	1.1.5
keyring	21.2.1
kiwisolver	1.2.0
10ehavi-lunar-calendar	0.2.1
lazy-object-proxy	1.4.3
LunarCalendar	0.0.9
mapclassify	2.2.0
MarkupSafe	1.1.1
matplotlib	3.2.1
mccabe	0.6.1
mistune	0.8.4
mplcursors	0.3
mpmath	1.1.0
munch	2.5.0
nbconvert	5.6.1
nbformat	5.0.7
notebook	6.0.3
numpy	1.18.5
numpydoc	1.0.0
packaging	20.4
pandas	1.0.4
pandocfilters	1.4.2
parso	0.5.2
pathtools	0.1.2
pexpect	4.8.0
pickleshare	0.7.5
Pillow	7.1.2
pip	20.1.1
plotly	4.8.1
pluggy	0.13.1
10ehaviour10-client	0.8.0
prompt-toolkit	3.0.5
psutil	5.7.0
ptyprocess	0.6.0
pycodestyle	2.6.0
pydocstyle	5.0.2
pyflakes	2.2.0
Pygments	2.6.1
pylint	2.5.3
pyls	0.1.6

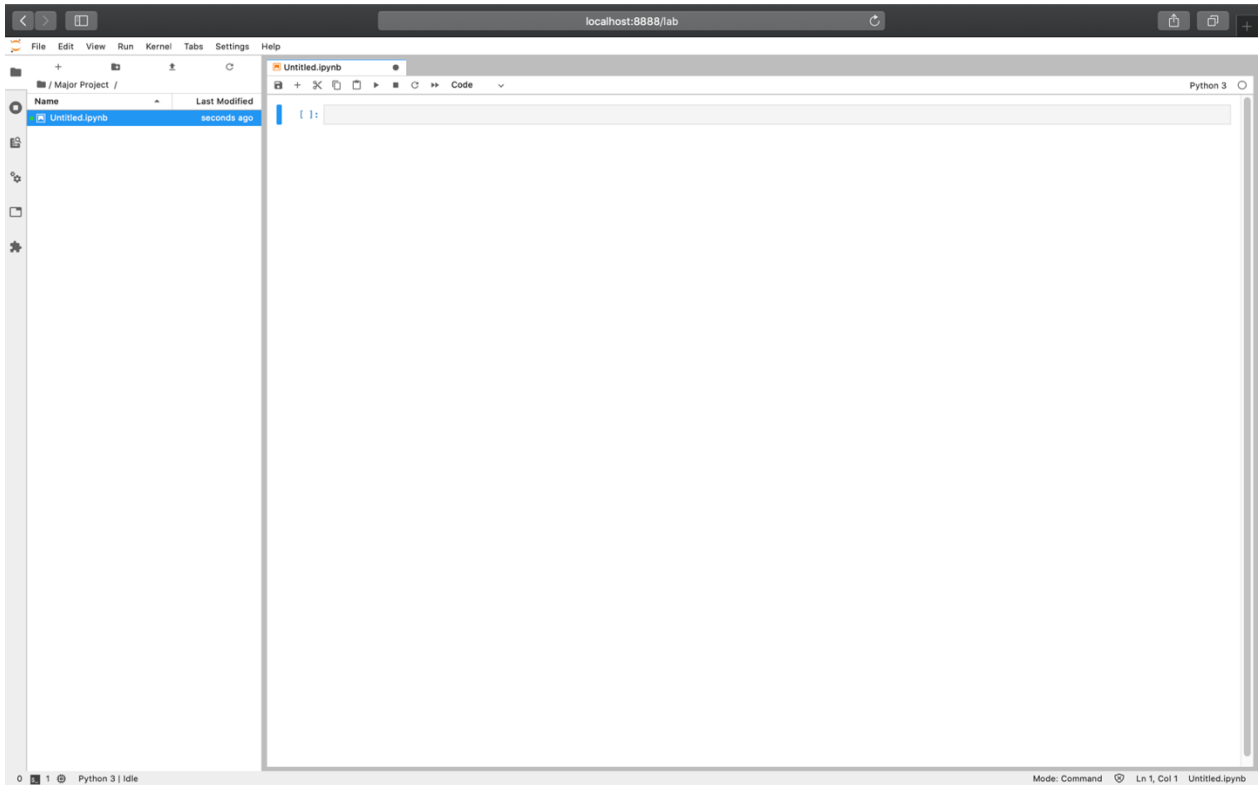
PyMeeus	0.3.7
pyobjc	6.2
pyobjc-core	6.2
pyobjc-framework-Accounts	6.2
pyobjc-framework-AddressBook	6.2
pyobjc-framework-AdSupport	6.2
pyobjc-framework-AppleScriptKit	6.2
pyobjc-framework-AppleScriptObjC	6.2
pyobjc-framework-ApplicationServices	6.2
pyobjc-framework-AuthenticationServices	6.2
pyobjc-framework-AutomaticAssessmentConfiguration	6.2
pyobjc-framework-Automator	6.2
pyobjc-framework-AVFoundation	6.2
pyobjc-framework-AVKit	6.2
pyobjc-framework-BusinessChat	6.2
pyobjc-framework-CalendarStore	6.2
pyobjc-framework-CFNetwork	6.2
pyobjc-framework-CloudKit	6.2
pyobjc-framework-Cocoa	6.2
pyobjc-framework-Collaboration	6.2
pyobjc-framework-ColorSync	6.2
pyobjc-framework-Contacts	6.2
pyobjc-framework-ContactsUI	6.2
pyobjc-framework-CoreAudio	6.2
pyobjc-framework-CoreAudioKit	6.2
pyobjc-framework-CoreBluetooth	6.2
pyobjc-framework-CoreData	6.2
pyobjc-framework-CoreHaptics	6.2
pyobjc-framework-CoreLocation	6.2
pyobjc-framework-CoreMedia	6.2
pyobjc-framework-CoreMediaIO	6.2
pyobjc-framework-CoreML	6.2
pyobjc-framework-CoreMotion	6.2
pyobjc-framework-CoreServices	6.2
pyobjc-framework-CoreSpotlight	6.2
pyobjc-framework-CoreText	6.2
pyobjc-framework-CoreWLAN	6.2
pyobjc-framework-CryptoTokenKit	6.2
pyobjc-framework-DeviceCheck	6.2
pyobjc-framework-DictionaryServices	6.2
pyobjc-framework-DiscRecording	6.2
pyobjc-framework-DiscRecordingUI	6.2
pyobjc-framework-DiskArbitration	6.2
pyobjc-framework-DVDPlayback	6.2
pyobjc-framework-EventKit	6.2
pyobjc-framework-ExceptionHandler	6.2
pyobjc-framework-ExecutionPolicy	6.2

pyobjc-framework-ExternalAccessory	6.2
pyobjc-framework-FileProvider	6.2
pyobjc-framework-FileProviderUI	6.2
pyobjc-framework-FinderSync	6.2
pyobjc-framework-FSEvents	6.2
pyobjc-framework-GameCenter	6.2
pyobjc-framework-GameController	6.2
pyobjc-framework-GameKit	6.2
pyobjc-framework-GameplayKit	6.2
pyobjc-framework-ImageCaptureCore	6.2
pyobjc-framework-IMServicePlugIn	6.2
pyobjc-framework-InputMethodKit	6.2
pyobjc-framework-InstallerPlugins	6.2
pyobjc-framework-InstantMessage	6.2
pyobjc-framework-Intents	6.2
pyobjc-framework-IOSurface	6.2
pyobjc-framework-iTunesLibrary	6.2
pyobjc-framework-LatentSemanticMapping	6.2
pyobjc-framework-LaunchServices	6.2
pyobjc-framework-libdispatch	6.2
pyobjc-framework-LinkPresentation	6.2
pyobjc-framework-LocalAuthentication	6.2
pyobjc-framework-MapKit	6.2
pyobjc-framework-MediaAccessibility	6.2
pyobjc-framework-MediaLibrary	6.2
pyobjc-framework-MediaPlayer	6.2
pyobjc-framework-MediaToolbox	6.2
pyobjc-framework-Metal	6.2
pyobjc-framework-MetalKit	6.2
pyobjc-framework-ModelIO	6.2
pyobjc-framework-MultipeerConnectivity	6.2
pyobjc-framework-NaturalLanguage	6.2
pyobjc-framework-NetFS	6.2
pyobjc-framework-Network	6.2
pyobjc-framework-NetworkExtension	6.2
pyobjc-framework-NotificationCenter	6.2
pyobjc-framework-OpenDirectory	6.2
pyobjc-framework-OSAKit	6.2
pyobjc-framework-OSLog	6.2
pyobjc-framework-PencilKit	6.2
pyobjc-framework-Photos	6.2
pyobjc-framework-PhotosUI	6.2
pyobjc-framework-PreferencePanels	6.2
pyobjc-framework-PubSub	6.2
pyobjc-framework-PushKit	6.2
pyobjc-framework-Quartz	6.2
pyobjc-framework-QuickLookThumbnails	6.2

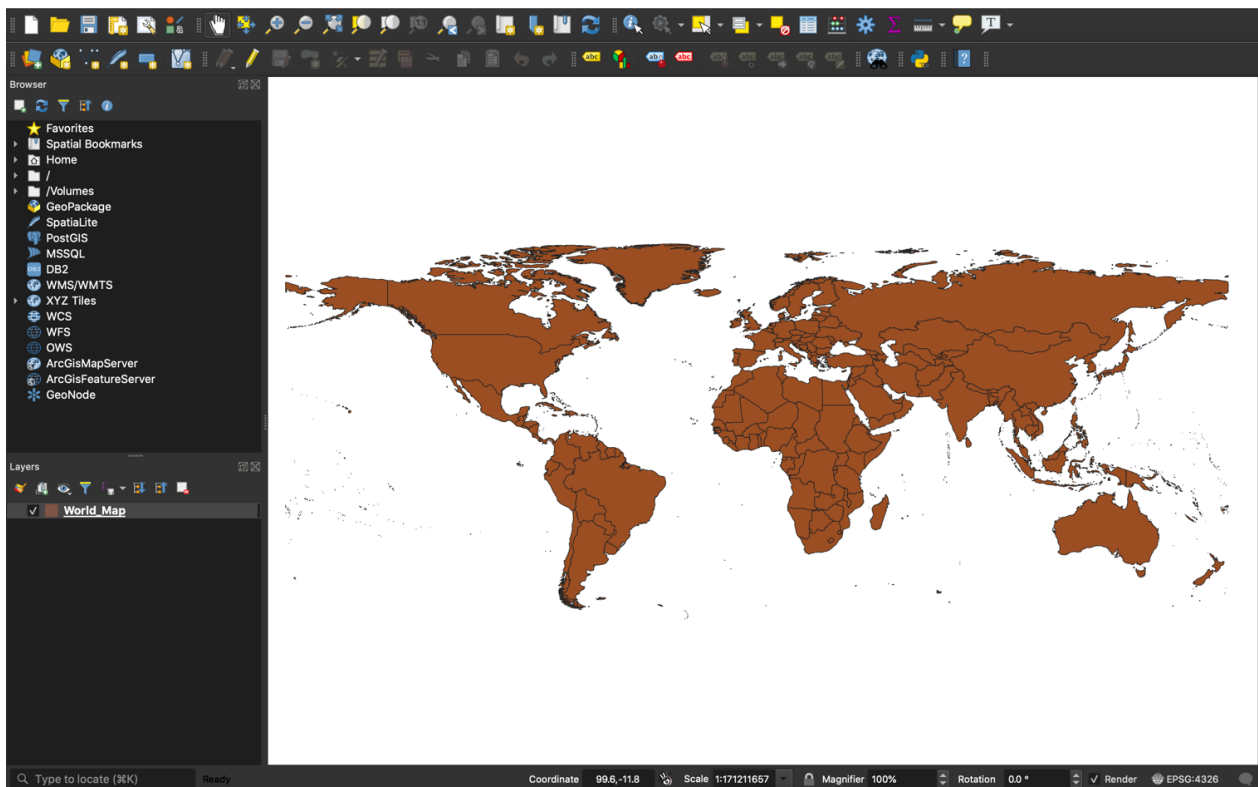
pyobjc-framework-SafariServices	6.2
pyobjc-framework-SceneKit	6.2
pyobjc-framework-ScreenSaver	6.2
pyobjc-framework-ScriptingBridge	6.2
pyobjc-framework-SearchKit	6.2
pyobjc-framework-Security	6.2
pyobjc-framework-SecurityFoundation	6.2
pyobjc-framework-SecurityInterface	6.2
pyobjc-framework-ServiceManagement	6.2
pyobjc-framework-Social	6.2
pyobjc-framework-SoundAnalysis	6.2
pyobjc-framework-Speech	6.2
pyobjc-framework-SpriteKit	6.2
pyobjc-framework-StoreKit	6.2
pyobjc-framework-SyncServices	6.2
pyobjc-framework-SystemConfiguration	6.2
pyobjc-framework-SystemExtensions	6.2
pyobjc-framework-UserNotifications	6.2
pyobjc-framework-VideoSubscriberAccount	6.2
pyobjc-framework-VideoToolbox	6.2
pyobjc-framework-Vision	6.2
pyobjc-framework-WebKit	6.2
pyparsing	2.4.7
pyproj	2.6.1.post1
PyQt5	5.12.3
PyQt5-sip	12.8.0
PyQtWebEngine	5.12.1
pyrsistent	0.16.0
pystan	2.19.1.1
python-dateutil	2.8.1
python-jsonrpc-server	0.3.4
python-language-server	0.31.10
pytz	2020.1
PyYAML	5.3.1
pyzmq	19.0.1
QdarkStyle	2.8.1
QtAwesome	0.7.2
qtconsole	4.7.4
QtPy	1.9.0
requests	2.23.0
retrying	1.3.3
rope	0.17.0
scikit-learn	0.23.1
scipy	1.4.1
seaborn	0.10.1
Send2Trash	1.5.0
setuptools	47.1.1

setuptools-git	1.2
Shapely	1.7.0
six	1.15.0
snowballstemmer	2.0.0
sortedcontainers	2.2.2
Sphinx	3.1.0
sphinxcontrib-applehelp	1.0.2
sphinxcontrib-devhelp	1.0.2
sphinxcontrib-htmlhelp	1.0.3
sphinxcontrib-jsmath	1.0.1
sphinxcontrib-qthelp	1.0.3
sphinxcontrib-serializinghtml	1.1.4
spyder	4.1.3
spyder-kernels	1.9.1
sympy	1.6
terminado	0.8.3
testpath	0.4.4
threadpoolctl	2.1.0
toml	0.10.1
tornado	6.0.4
traitlets	4.3.3
typing-extensions	3.7.4.2
ujson	1.35
urllib3	1.25.9
watchdog	0.10.2
wcwidth	0.2.4
webencodings	0.5.1
Werkzeug	1.0.1
widgetsnbextension	3.5.1
wrapt	1.12.1
wurlitzer	2.0.0
yapf	0.30.0

And for complete interactive python project development and machine Jupiter lab environment **Figure2** is used as picturized below:



And an open source, cross platform software named QGIS 3.10 LTS is used for viewing desktop geographical information system application for viewing and analysis of geospatial data. **Figure 3**



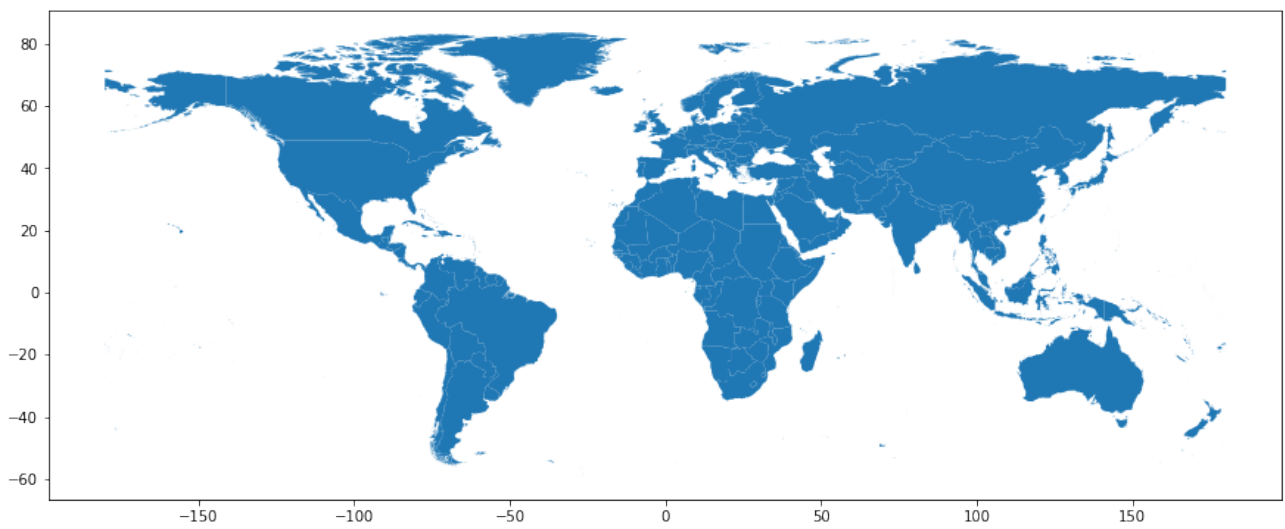


Figure 4 Showing graph post importing map.shp file in Python

Chapter 3 Methodology

3.1 Methodology Adopted and Block diagram

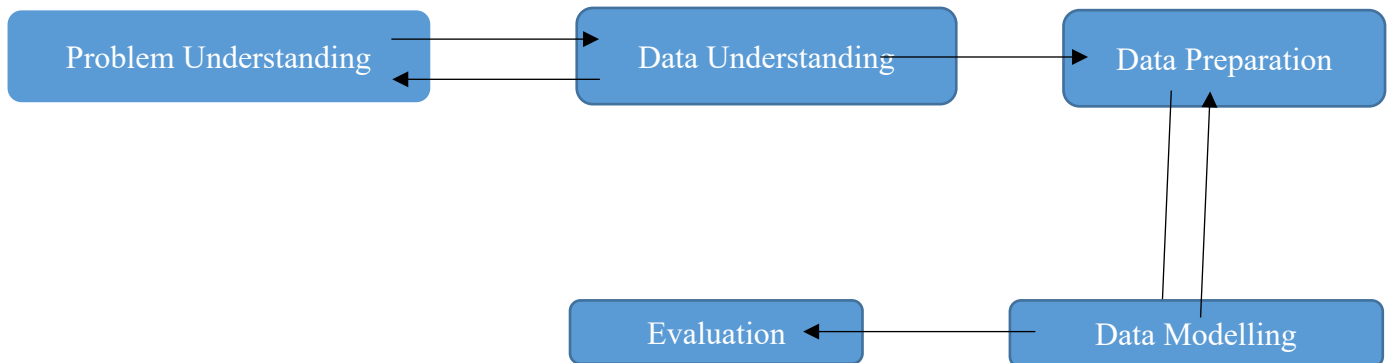


Figure 5 Showing Data Life Cycle

As depicted in the above figure any data science project is completed in five major steps

1. Problem Understanding:

To find out the right solution to the problem one needs to break the problem into small chunks to make sense out of it, in. such a way that they make complete sense when interlinked together and seek out what kind of data and what kind of attributes are required for making the best of the problem. This is often referred to as the bottom-up approach where first the problem statement is defined, and later data modelling or data scrapping is done based on the sub-parts that one wishes to analyze. On the contrary there is a top to bottom approach on which the second step data understanding is done first according the problem statement is defined making sense from the provided data.

We have taken top to bottom approach where we defined the problem statement to define the COVID-19 situation around top ten most impacted countries and analyze other factors like recover, death rate, mortality and make a choropleth map from the data collected for every single day to the data.

2. Data Understanding:

For making the best out of the data we have times_confirmed.csv file with attributes like country, province, latitude, longitude, date wise tabulation of confirmed cases from coronavirus. Which is really helpful in mapping out detailed description from everyday progress and best suited for machine learning model building. Similarly, we have times_death.csv and times_recovery.csv with same attributes as times_confirmed.csv but with cumulative deceased and recovered cases. Then we have Map.shp file which have geometrical coordinated for every country,

which help in developing a choropleth map of the world and monitor daily cases via sharp shift in colors on the map.

3. Data Preparation:

It includes pre-processing of raw data prior modelling and analysis. Or it can be defined as the pre-manipulation of raw data and making it suitable to modelling and analysis. It includes editing and conditioning of data making it best suitable for modelling. Like in Map.shp and times_confirms.csv file the name of country may vary like 'UK' and 'United Kingdom', 'Viet' and 'Vietnam' etc. which actually represents same country on the globe but computer being case sensitive to analysis one needs to edit out all such minor rectifications of significance otherwise the modelling or mapping might not work.

4. Data Modelling:

It primarily includes processing of the conditions data making it workable. If there is any sort of error encountered while modelling then we might need to reassessment of data preparation, step 3. In the project we have used two unique techniques for data modelling i.e. linear regression and SVM.

Linear Logistic Regression:

It's a supervised machine learning algorithm. It's more of like, finding a relationship between the scalar dependent entity present in the data and the explanatory variable, independent entity. Sklearn is the best suited library for such modelling where we first, find out which model is fitting our data best i.e. linear, quadratic and polynomial based on which the system computes and returns the best way for modelling post assessment.

SVM (Support Vector Machine):

Again, a supervised learning approach as it working on the predefined data, we collected post rectification from errors. It's often used for regression, classification, and outliers detection. Based on which it generates a prediction model. It's based on algorithm that depend on behaviour of data on classification and regression approach. Which makes it helpful in prediction and forecasting futuristic values.

5. Evaluation

After model building and being in coherence with the predicted values we can initiate to ponder on our values and observations that we make with our data set.

Chapter 4 Implementation

4.1 Modules and Evaluation Report

Every linear graph withdrawn from the data is redrawn in Y-axis to logarithmic scale as logarithmic scale can show sharp changes and stability with vertical graph line and horizontal graph line respectively. The top ten countries have been picked up by sorting out from the list of countries based on current confirmed cases.

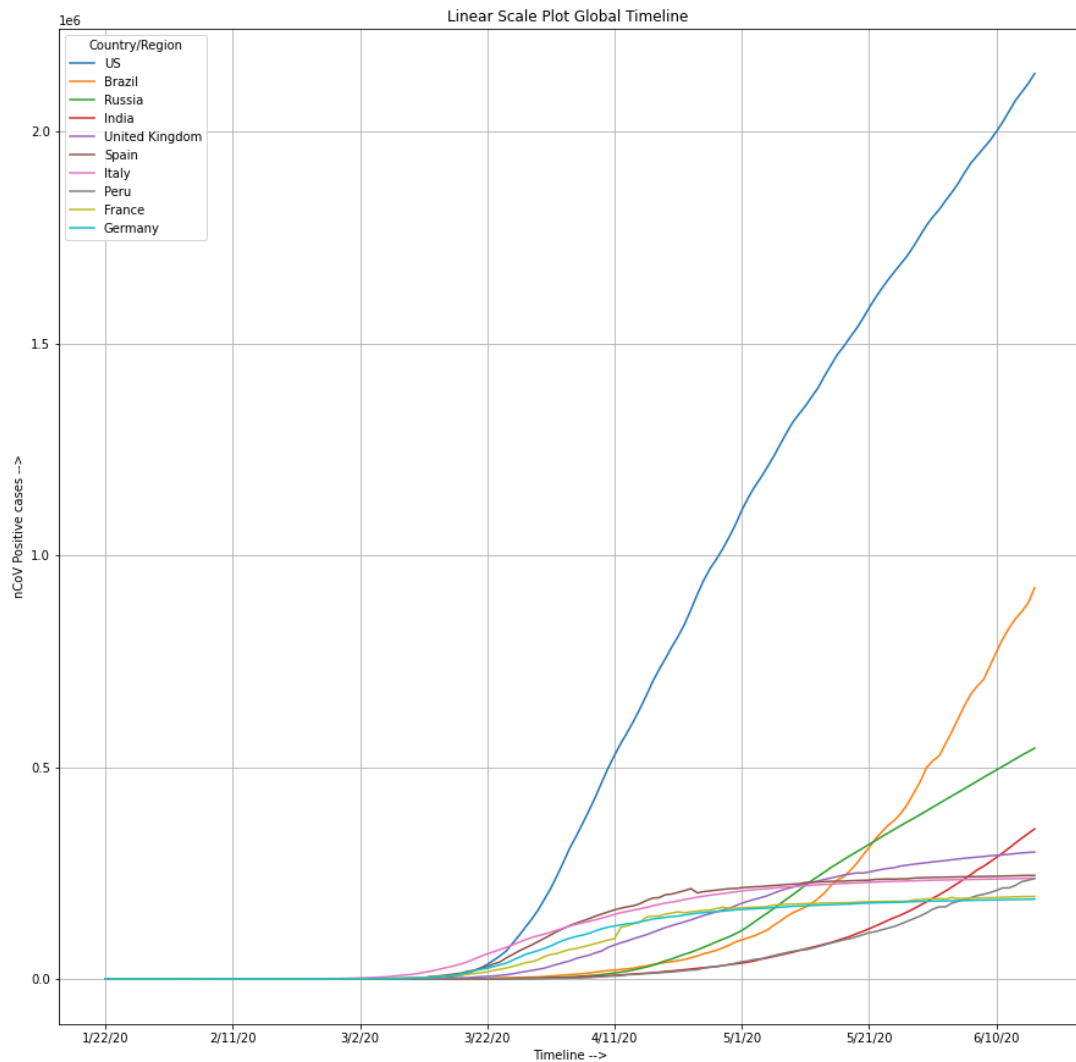


Figure 6 Showing top ten countries with confirmed COVID-19 cases on linear scale

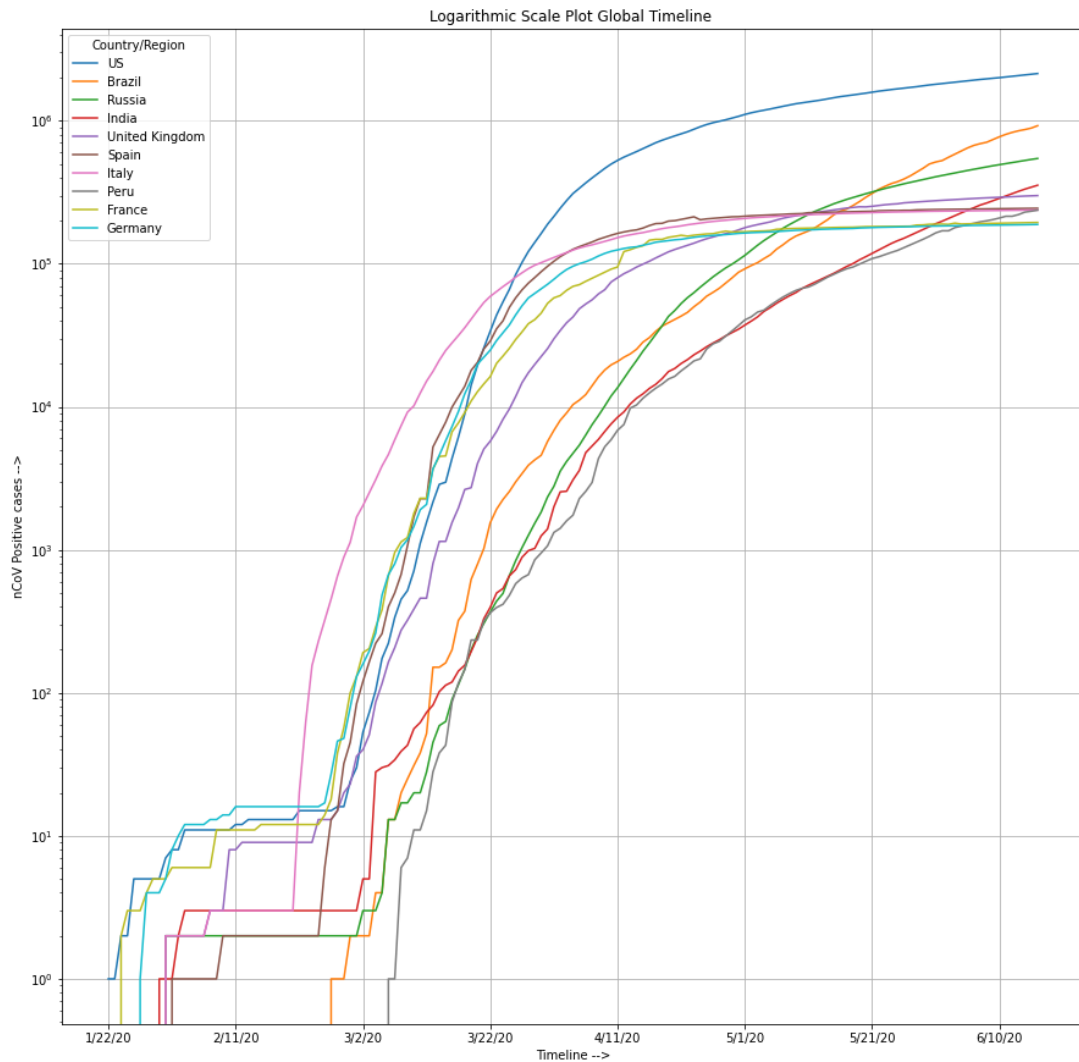


Figure 7 Showing top ten countries with confirmed COVID-19 cases on exponential scale

From the graph depicted in Figure 7 we can see when the first case was reported in the past via a straight vertical line rising from x-axis to the scale of 10^0 which is, sort of difficult to make out from the linear scale plots because of overlapping. From the graph it's clear that the first case in India was first reported on 30th January 2020 after which there has been less than 10 cases till the 3rd of March, as we can see from the red horizontal line on the graph between 10^0 and 10^1 . Similarly, we can deduce conclusions for other countries too.

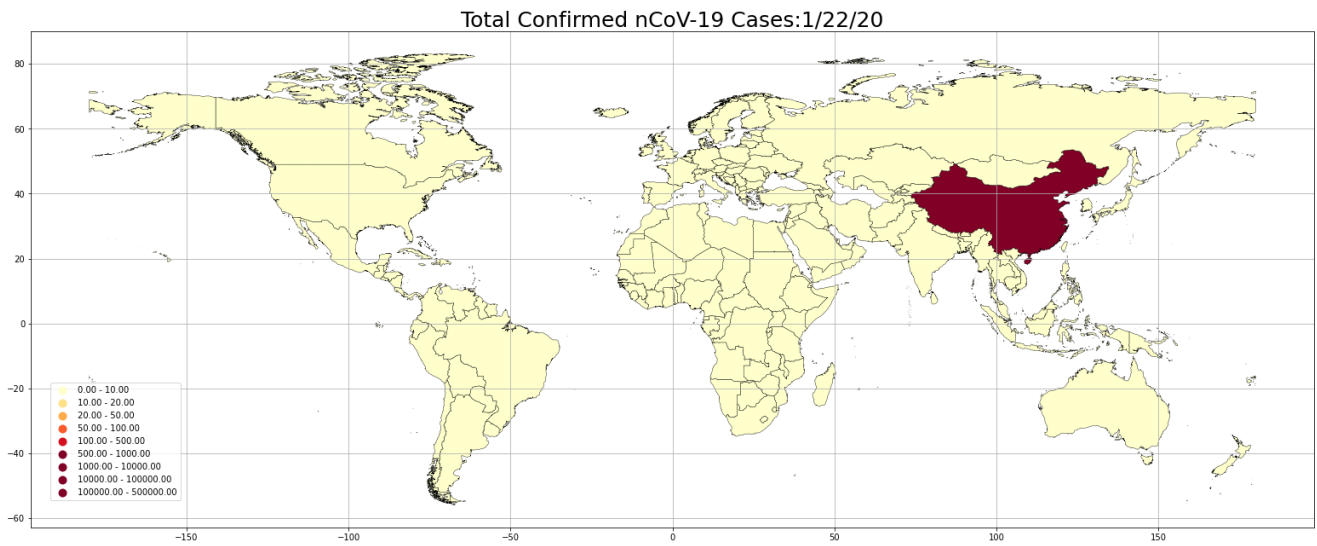


Figure 8 Showing global situation as of 22nd January 2020

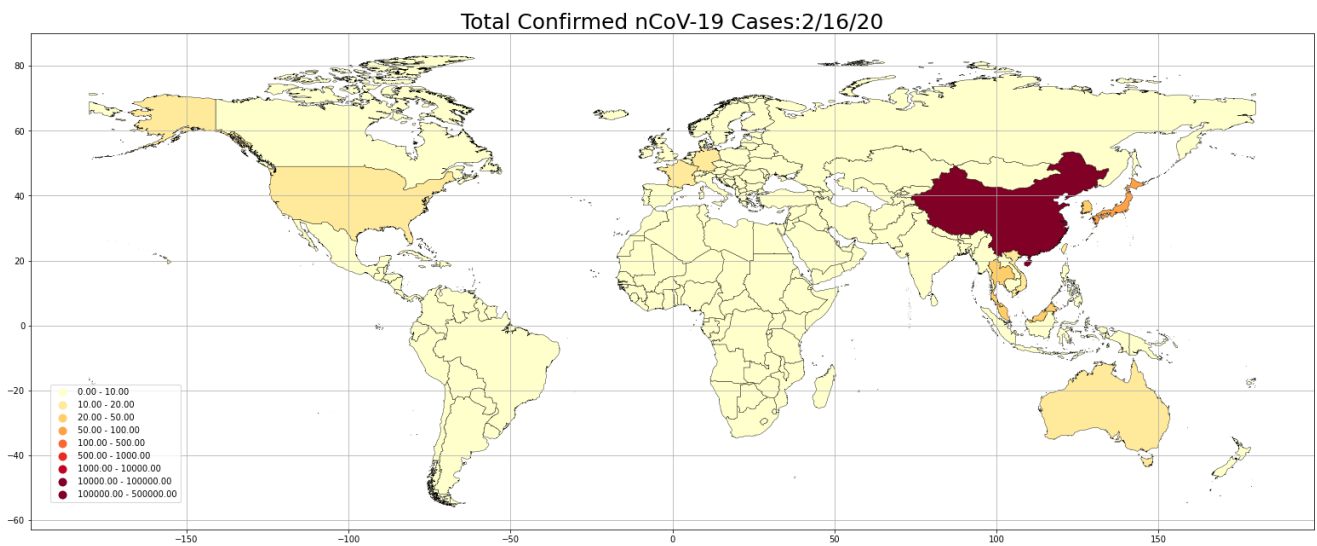


Figure 9 Sowing Global Situation as of 16th February 2020

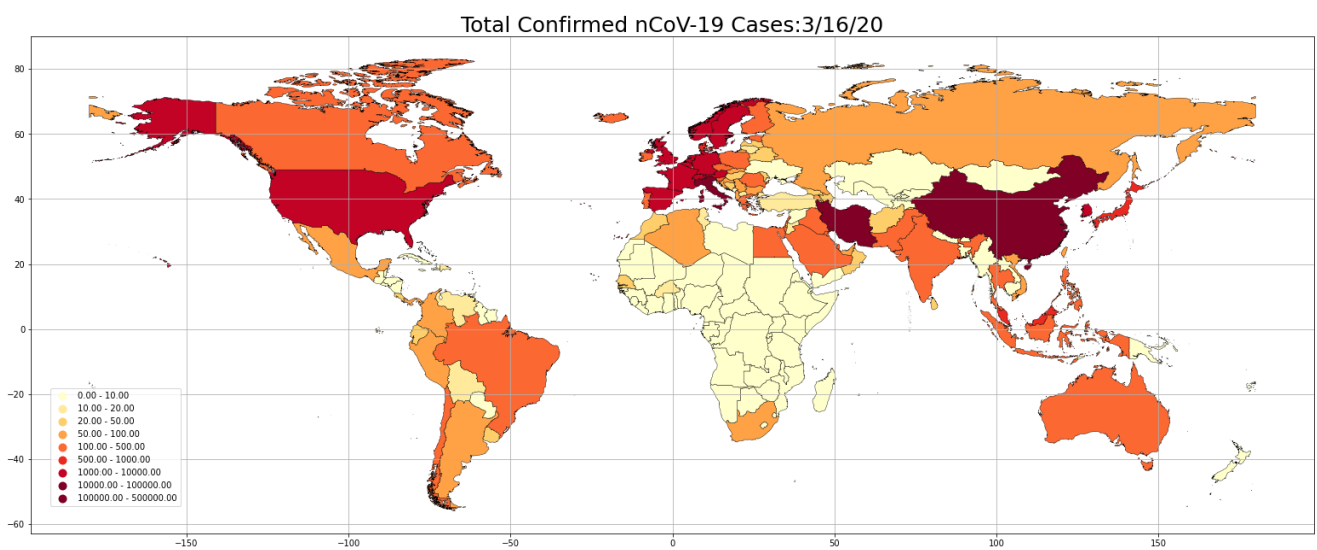


Figure 10 Showing Global Situation as of 16th March 2020

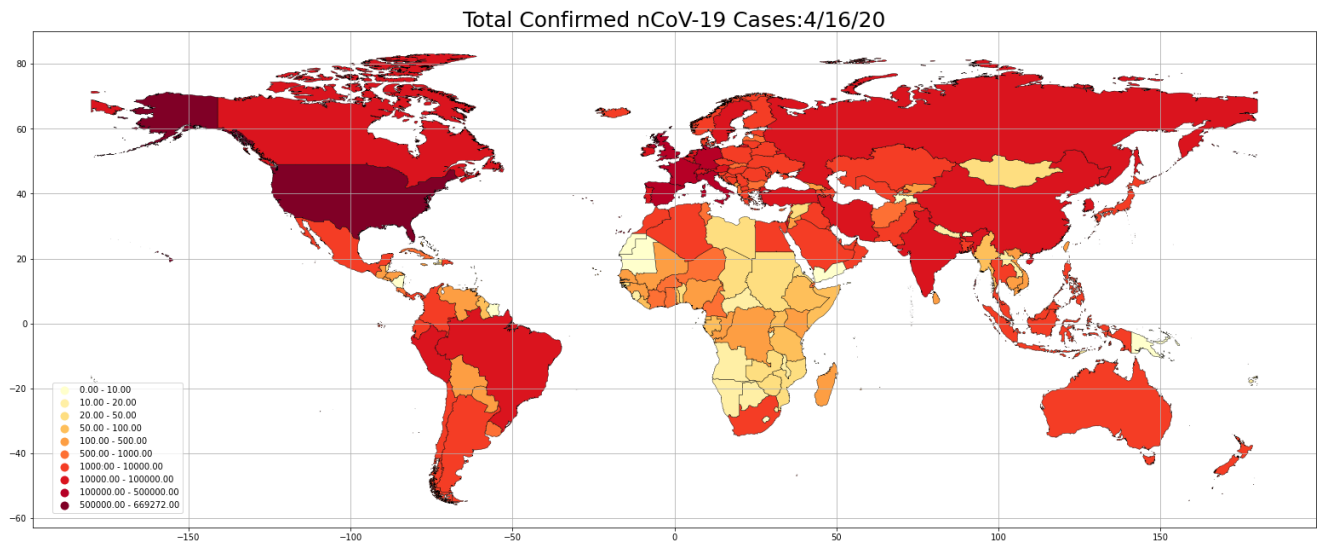


Figure 11 Showing Global Situation as of 16th April 2020

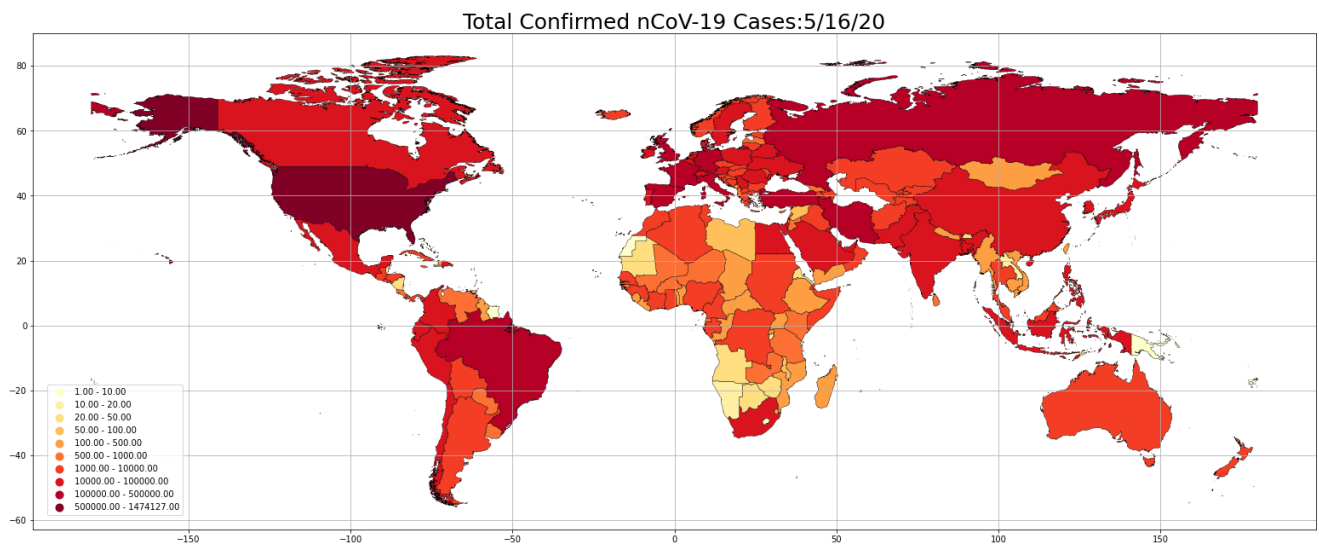


Figure 12 Showing Global Situation as of 16th May 2020

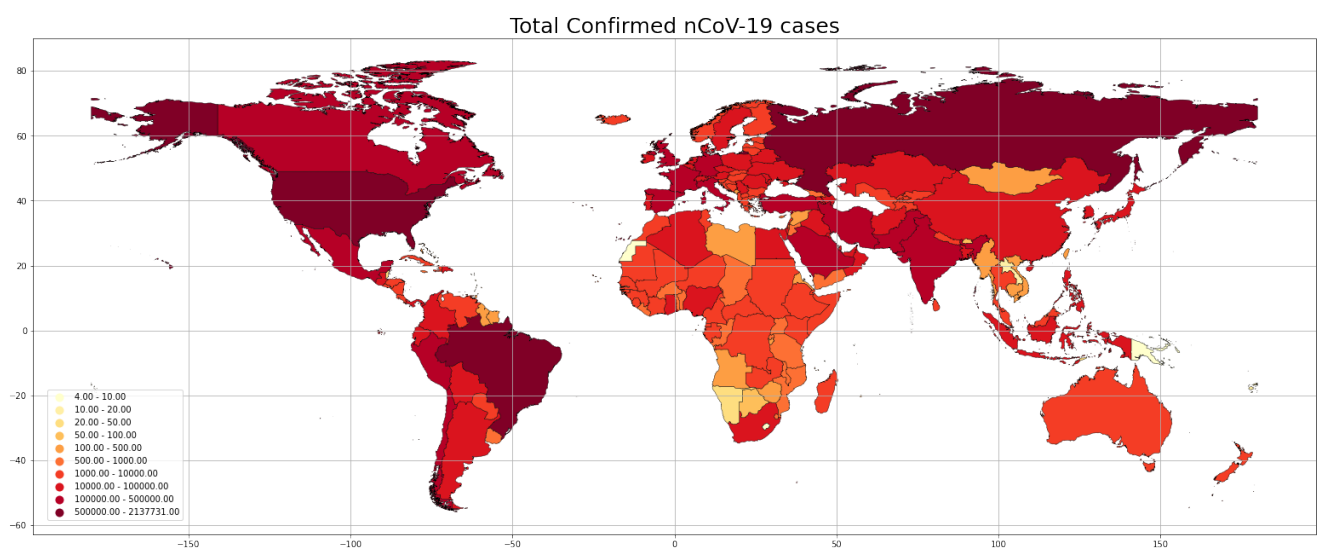


Figure 13 Showing global situation as of 16th June 2020
(last when the data was collected)

Global nCoV-19 Cases 16_June_20

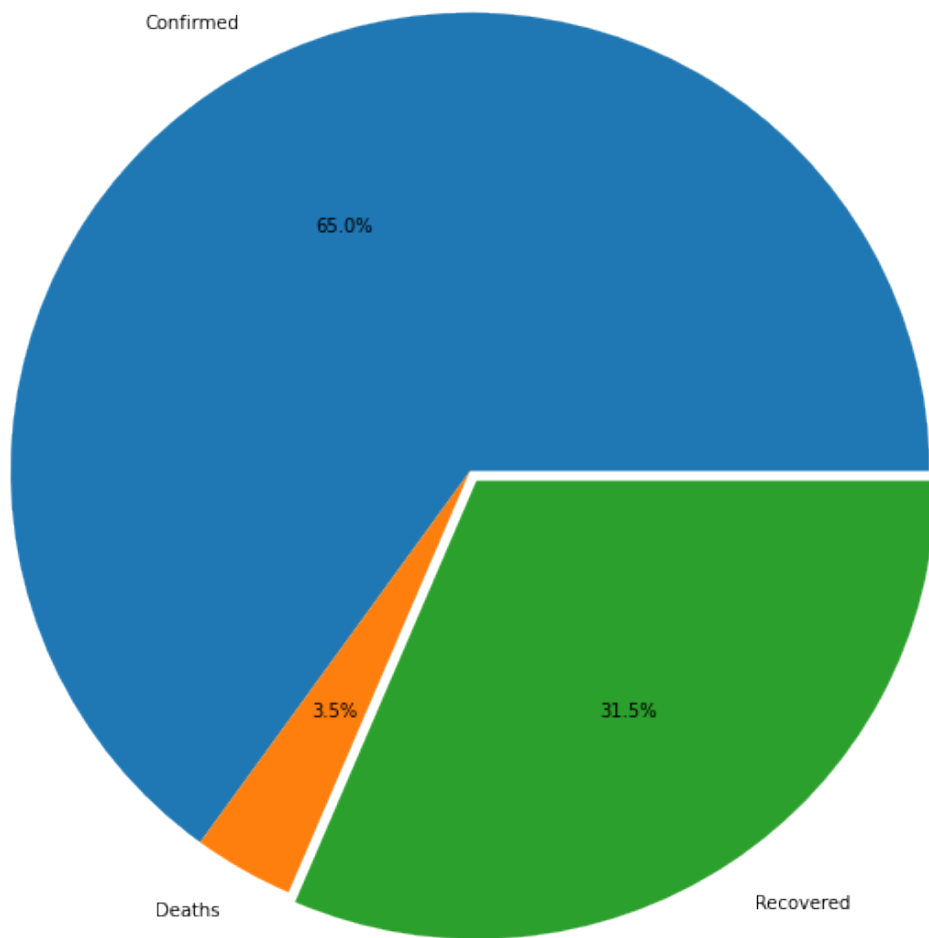


Figure 14 Depicting a total percentage of Confirmed, Dead and Recovered cases as on 16th June 2020

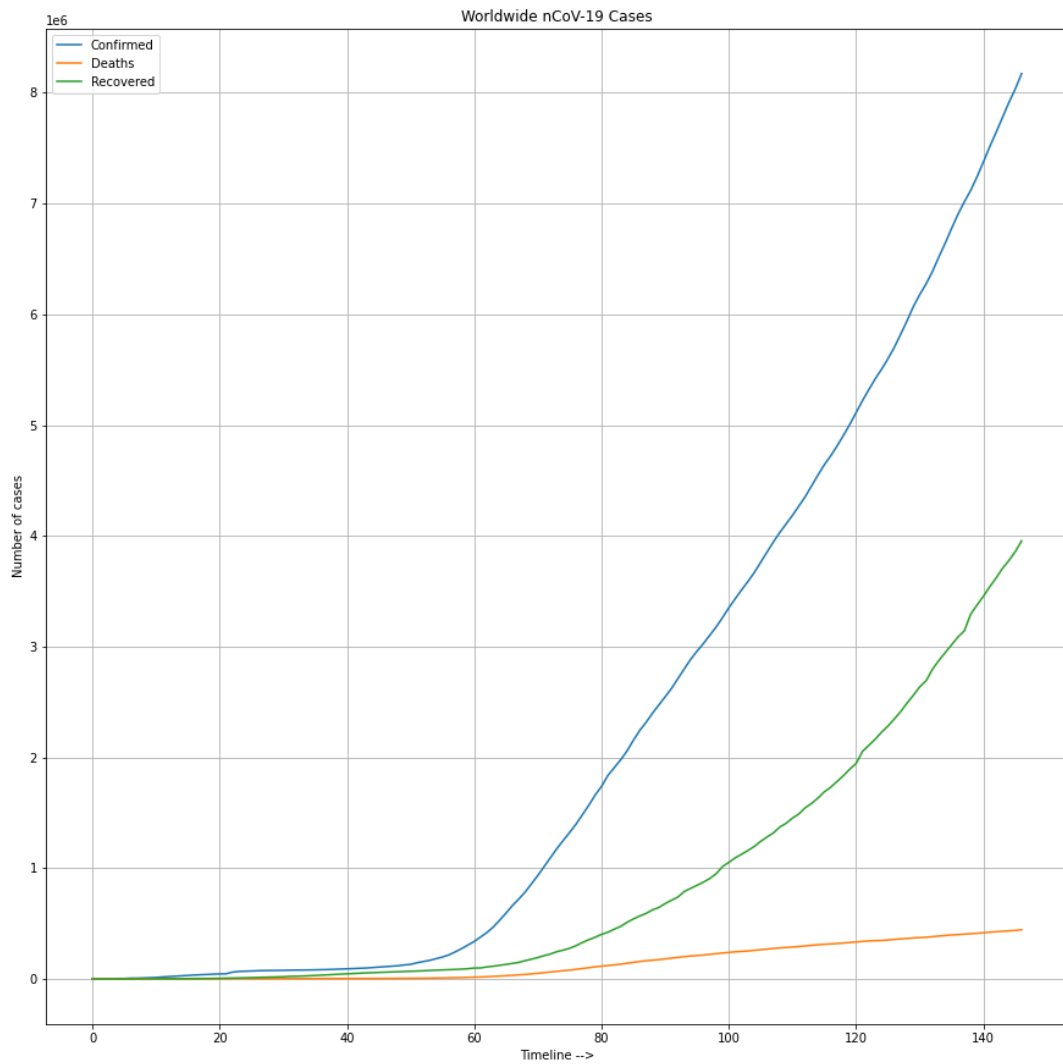


Figure 16 Graph of Global Confirmed (in Blue), Global Deaths (in Red) and Global Recovered (in Green) on linear scale

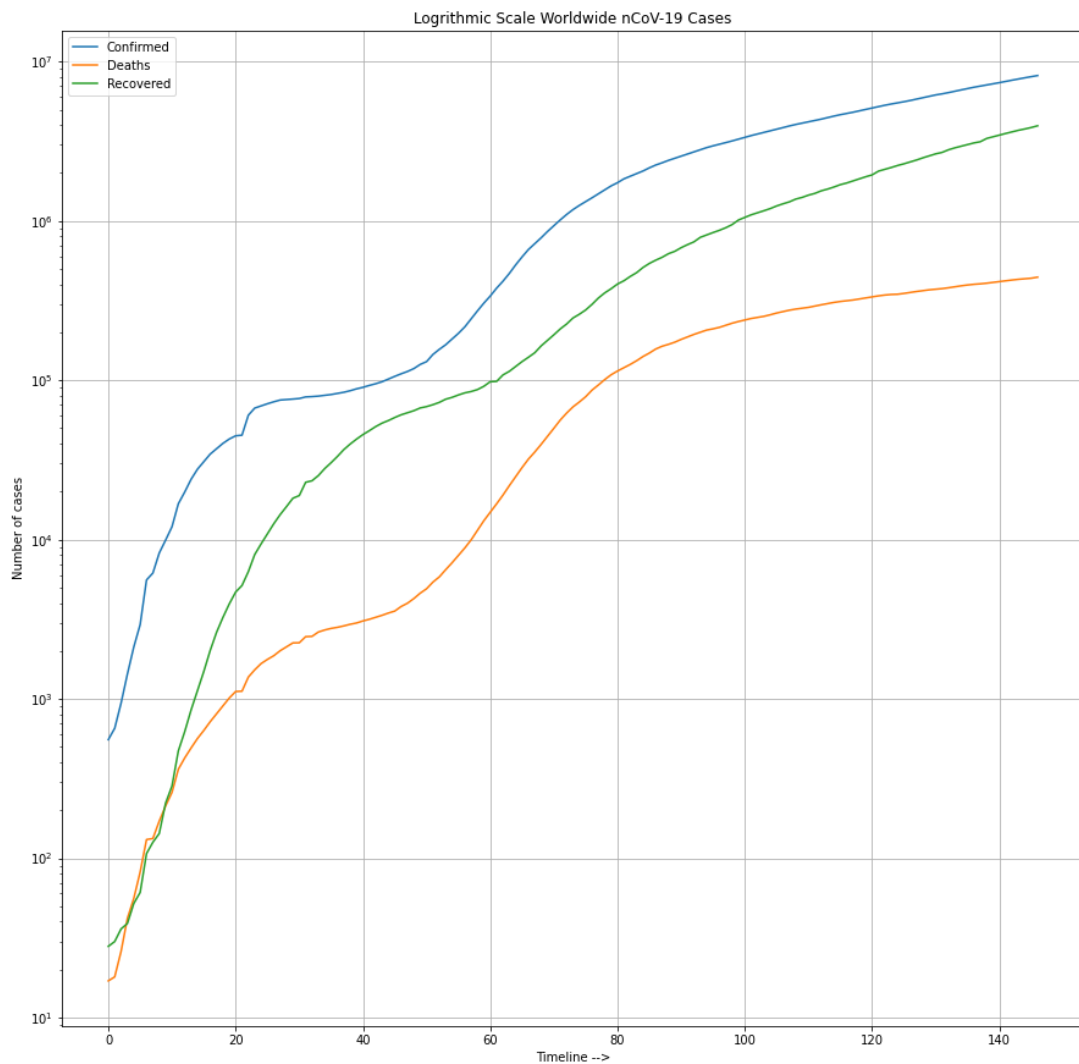


Figure 17 Graph of Global Confirmed (in Blue), Global Deaths (in Red) and Global Recovered (in Green) on exponential scale

From figure 14 nothing much is clarified from for first 30 days, but from figure 15 it's evident that we did have faced days when the death rate was more as compared to recovery rate as we were not able to fully understand the behaviour of the virus and the lockdown hadn't been imposed then we might be facing more deaths than recovery.

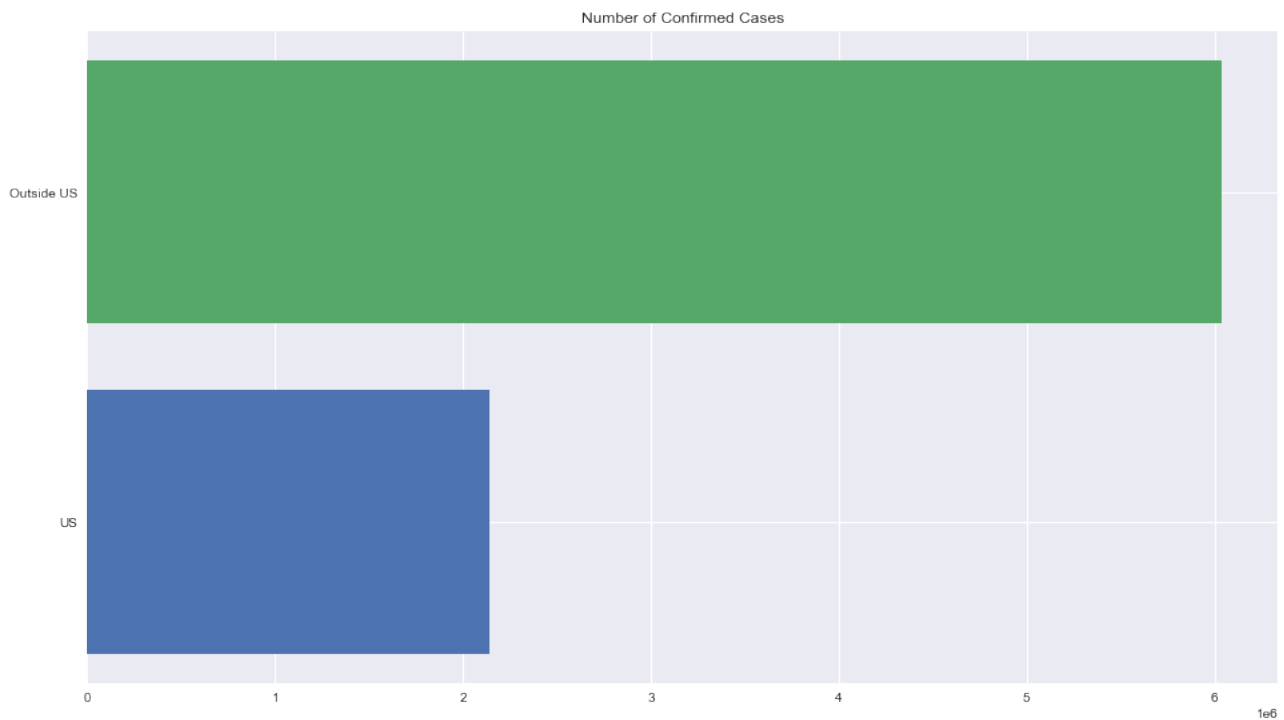


Figure 18 Comparing the situation report of US being on top with the most confirmed cases and cumulative sum of other countries as of 16th June 2020

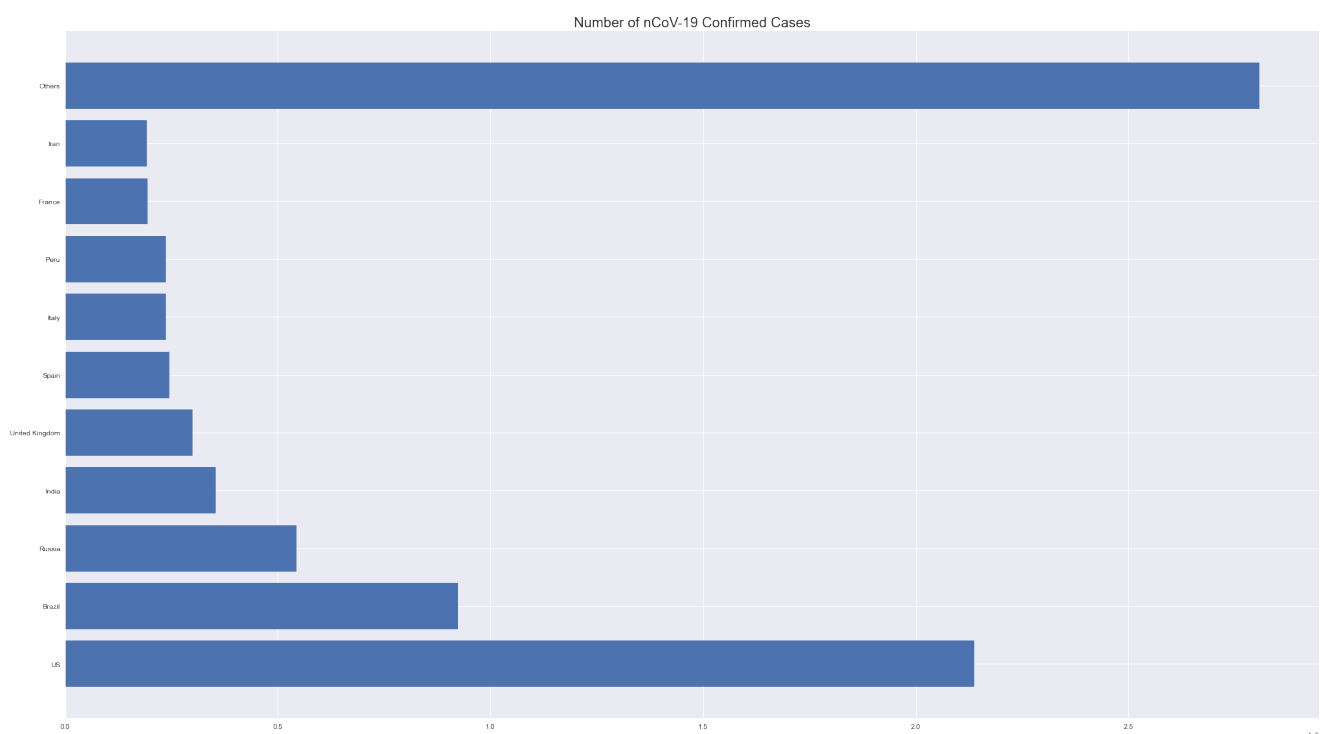


Figure 19 Comparison of situation reports inclusive of top ten countries and cumulative sum of other countries entitled 'Others' as their cumulative confirmed cases

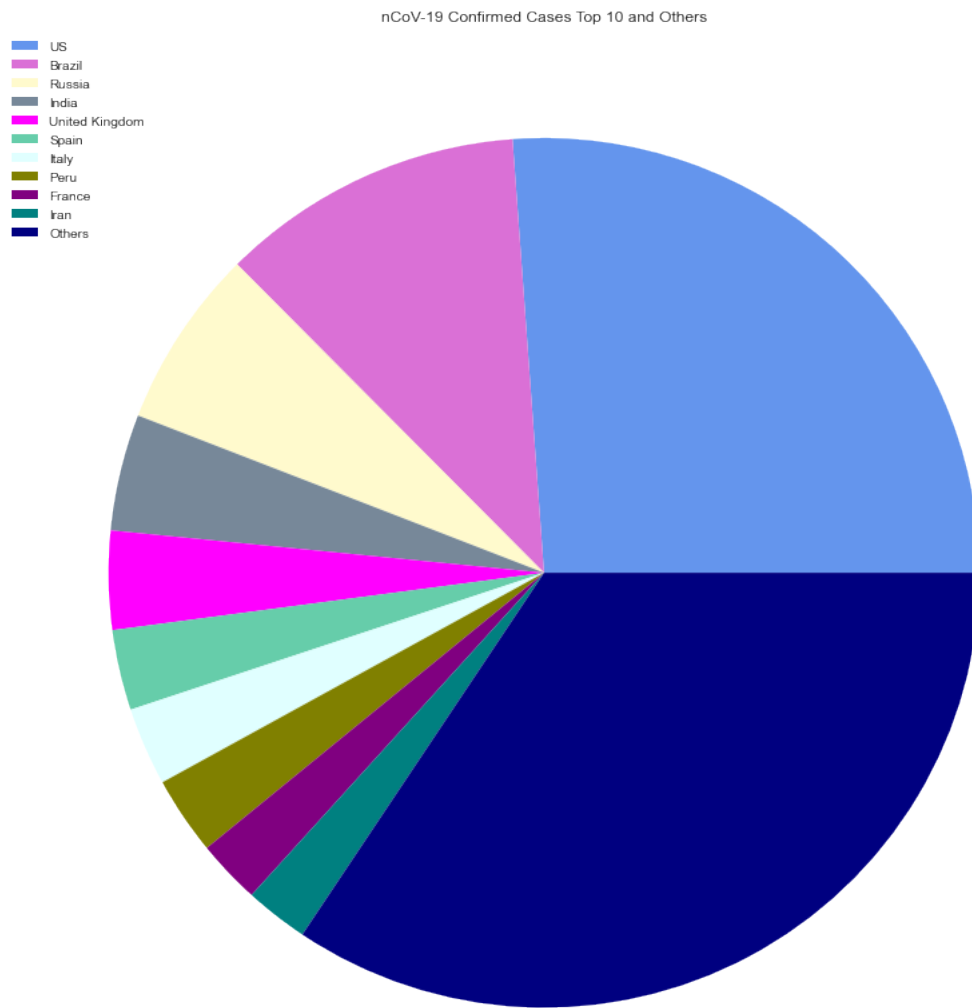


Figure 20 Pie Chart showing the percentage of total confirmed cases
Mentioning top ten most impacted countries while other countries are totalled as ‘Others’

	Name of State / UT	Active Cases	Cured/Discharged/Migrated	Deaths	Total Confirmed cases
0	Andaman and Nicobar Islands	11	33	0	44
1	Andhra Pradesh	3244	3509	88	6841
2	Arunachal Pradesh	88	7	0	95
3	Assam	2145	2166	8	4319
4	Bihar	2093	4644	41	6778
5	Chandigarh	50	302	6	358
6	Chhattisgarh	736	1036	9	1781
7	Dadra and Nagar Haveli and Daman and Diu	36	9	0	45
8	Delhi	26351	16500	1837	44688
9	Goa	544	85	0	629
10	Gujarat	5962	17082	1533	24577
11	Haryana	4406	3748	118	8272
12	Himachal Pradesh	180	372	8	560
13	Jammu and Kashmir	2454	2781	63	5298
14	Jharkhand	709	1121	9	1839
15	Karnataka	2980	4456	94	7530
16	Kerala	1366	1236	20	2622
17	Ladakh	563	85	1	649
18	Madhya Pradesh	2455	8152	476	11083
19	Maharashtra	50057	57851	5537	113445
20	Manipur	341	159	0	500
21	Meghalaya	18	25	1	44
22	Mizoram	120	1	0	121
23	Nagaland	87	92	0	179
24	Odisha	1178	2974	11	4163
25	Puducherry	111	99	6	216
26	Punjab	838	2461	72	3371
27	Rajasthan	3059	9849	308	13216
28	Sikkim	66	4	0	70
29	Tamil Nadu	20709	26782	528	48019
30	Telangana	2188	3027	191	5406
31	Tripura	658	433	1	1092
32	Uttarakhand	701	1216	25	1942
33	Uttar Pradesh	5064	8610	417	14091
34	West Bengal	5386	6028	495	11909

Figure 21 Showing the stagewise ground report in India as of 16th June 2020 most to Least affected, from dark to light

Chapter 5 Results and Analysis

5.1 Machine Learning and Prediction

As generally taught we split the complete data into two sets training and testing set. Keeping the 15% of the data for testing and other 85% data for training. Using the SVM model we pre-set the kernel to poly (as the data points are not equally separated), sigmoid (when we expect a situation of saturation on the curve) and Radial Basic Function or rbf (default parameter of SVM kernel for regression). Using randomised search for iterative search method over kernel poly was found to give best results amongst all of the predefined parameters. Creating a regression model over the training data it calculates the prediction values it develops the prediction model values for the corresponding testing values which are depicted using graph in figure 18

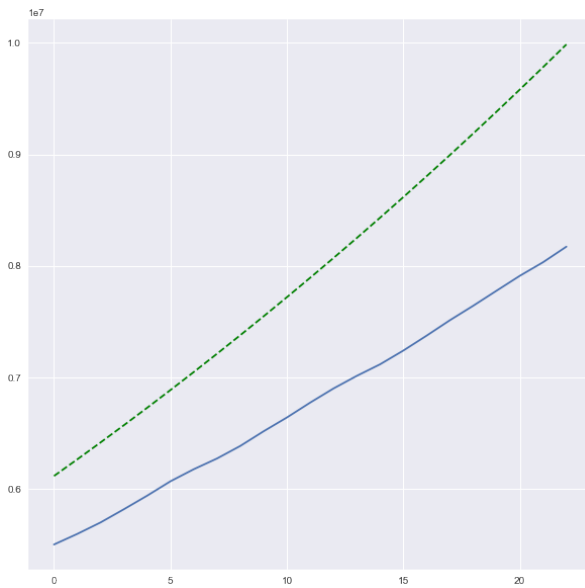
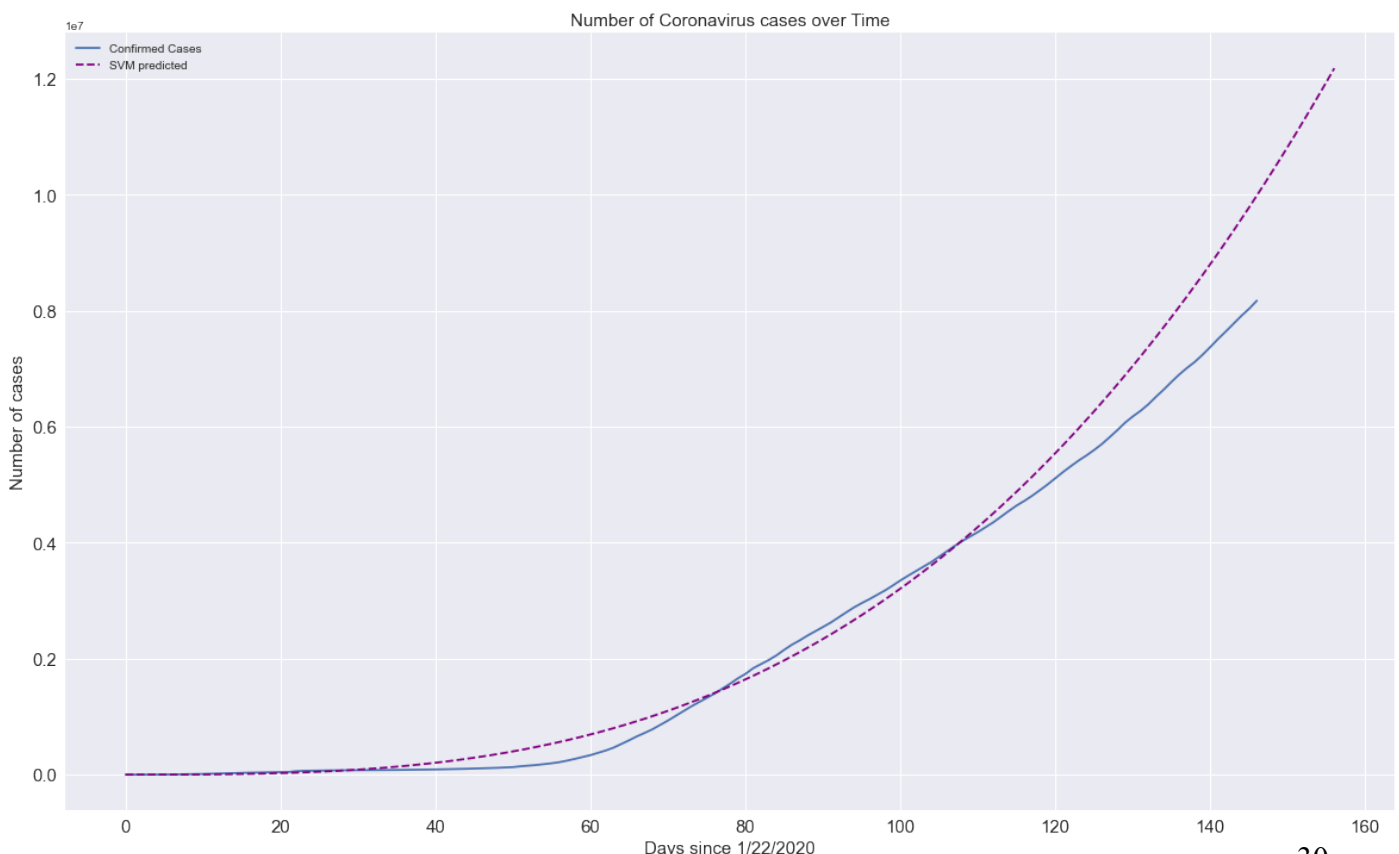


Figure 22 Shows the predicted values from SVM Regression model in green dashed lines and the confirmed values in blue on test data

Clearly, there is an increment in the mean squared error with time as the model is clearly overfitting bases on the data it was fed.

Figure 23 below shows the actual confirmed cases in blue line and prediction values via purple dashed line from the SVM Regression model.



```
# Forecasting for next 10 days via SVM model
print('SVM model forecasting ')
set(zip(future_forecast_dates[-10:], svm_pred[-10:]))
```

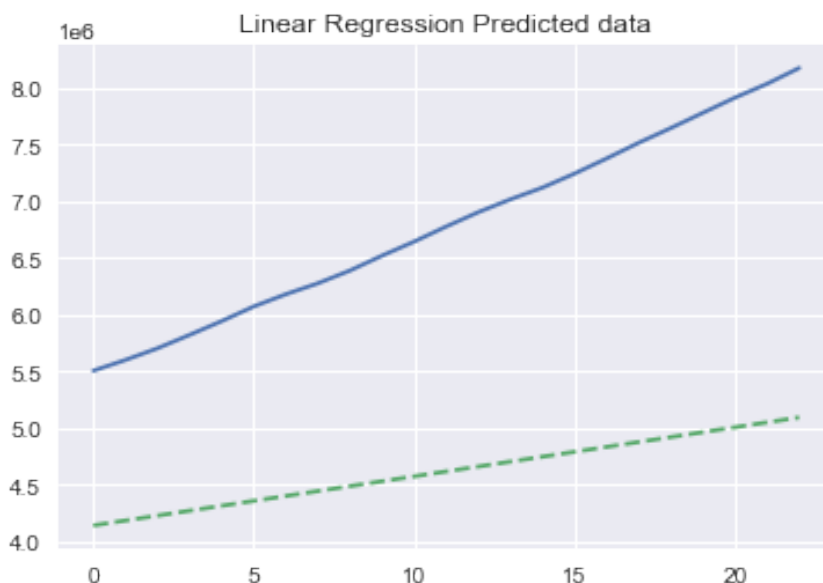
SVM model forecasting

```
{('06/17/2020', 10193168.4122999),
 ('06/18/2020', 10402599.18964365),
 ('06/19/2020', 10614879.3185499),
 ('06/20/2020', 10830028.0607374),
 ('06/21/2020', 11048064.6544874),
 ('06/22/2020', 11269008.3654249),
 ('06/23/2020', 11492878.4435499),
 ('06/24/2020', 11719694.1310499),
 ('06/25/2020', 11949474.6935499),
 ('06/26/2020', 12182239.3732374)}
```

Figure 24 Shows the forecast values for the next ten days based on SVM regression model

The increment indicated in the predicted values on new confirmed cases is a statistical evidence for no possible demise in the rate of increment in the spread of the virus. Which is also not a good indicator for medical and health

facilities as we have already started to face demise in the availability of medical facilities in the upcoming near future.



Second, approach we opted is linear regression, as we have already covered the possibility of logistic regression with the support vector machine regression model. The model predicted demise in the rate of new upcoming positive cases on coronavirus on test data set with time which indicates an underfitting model based on training dataset.

Figure 25 Shows the actual confirmed cases via blue line and the prediction values using green dashed line from linear regression.

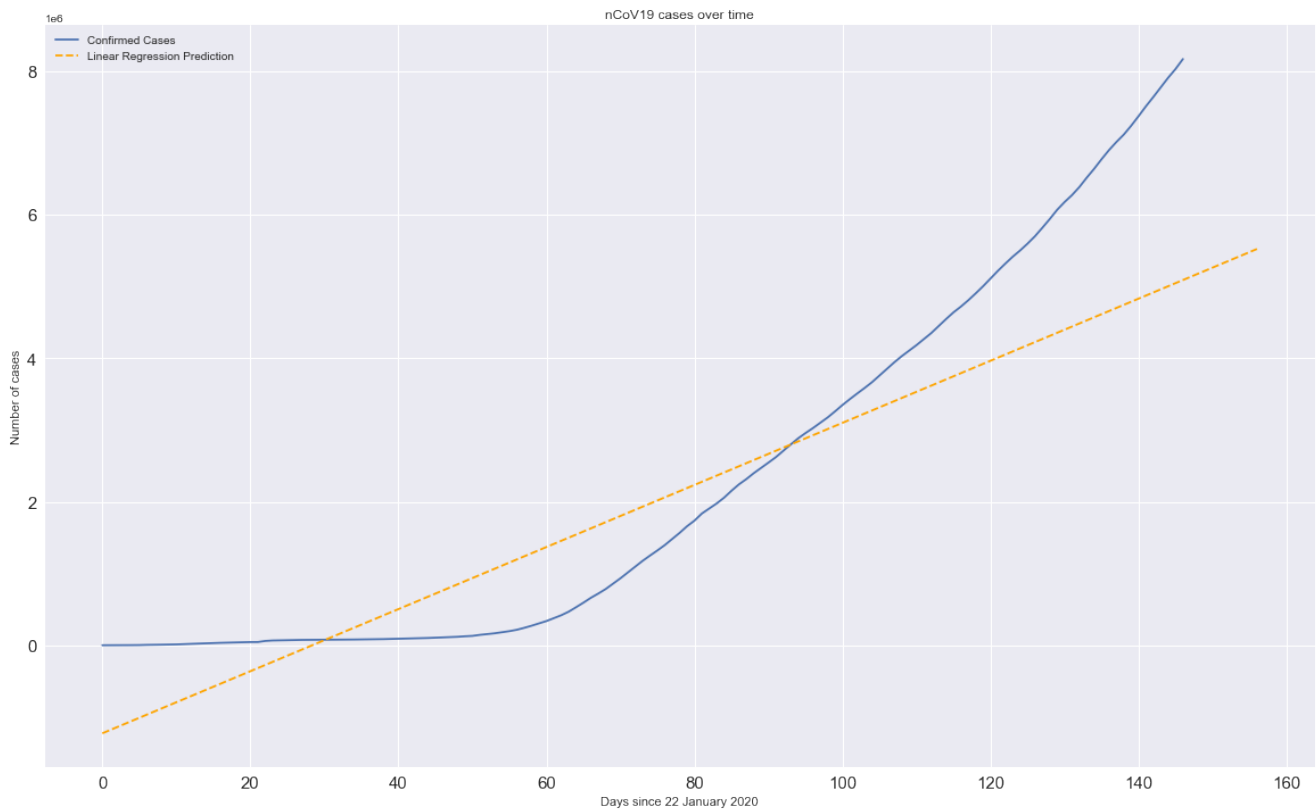


Figure 26 Depicts the linear regression model in yellow dashed line and the confirmed cases in blue line on the graph (on linear scale)

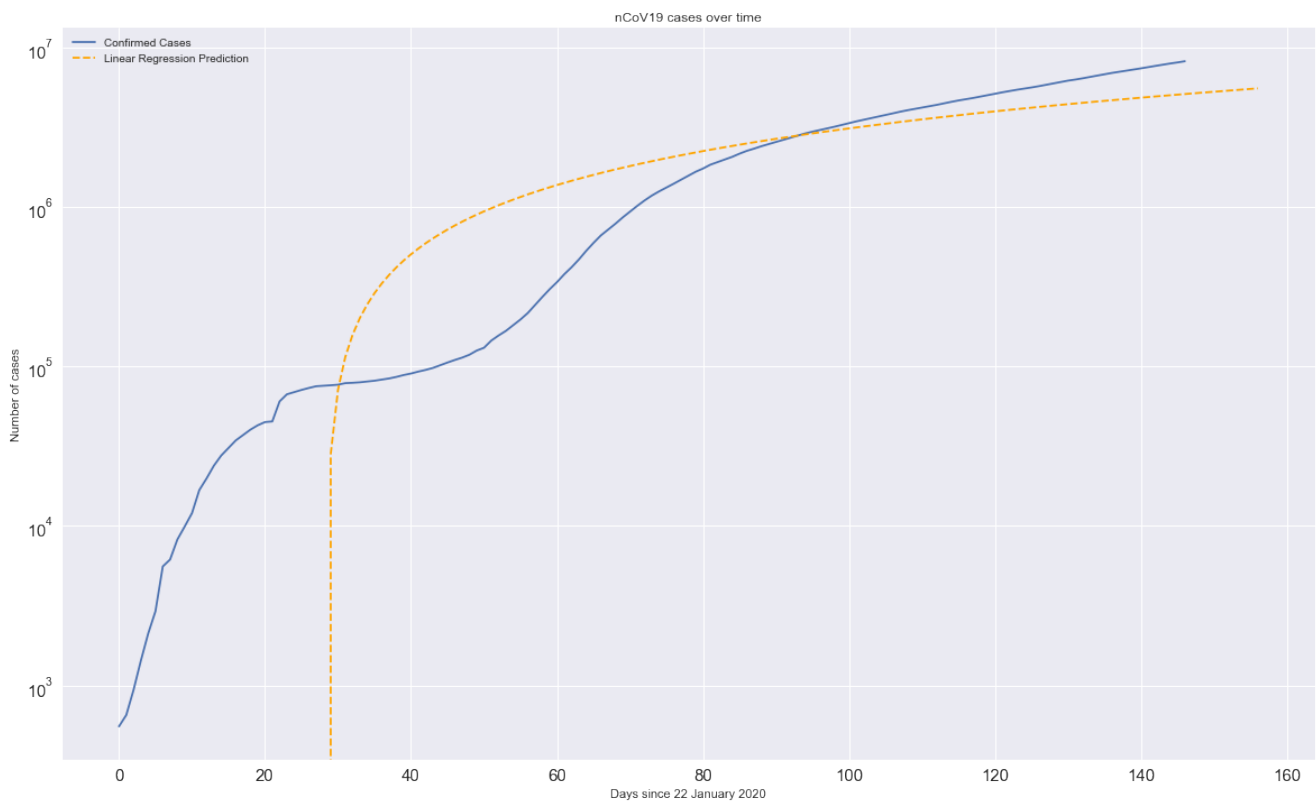


Figure 27 Depicts the linear regression model in yellow dashed line vs confirmed cases in the blue on exponential scale.


```
# Linear Regression prediction for next 10 days
print('Linear Regression future predicitions: ')
print(linear_pred[-10:])
```

```
Linear Regression future predicitions:
[[5135541.6300236 ]
 [5178825.10966168]
 [5222108.58929976]
 [5265392.06893785]
 [5308675.54857593]
 [5351959.02821401]
 [5395242.50785209]
 [5438525.98749017]
 [5481809.46712825]
 [5525092.94676633]]
```

Figure 28 Shows the prediction values for next 10 days from which we can observe that it also indicates an increment in future new positive cases.

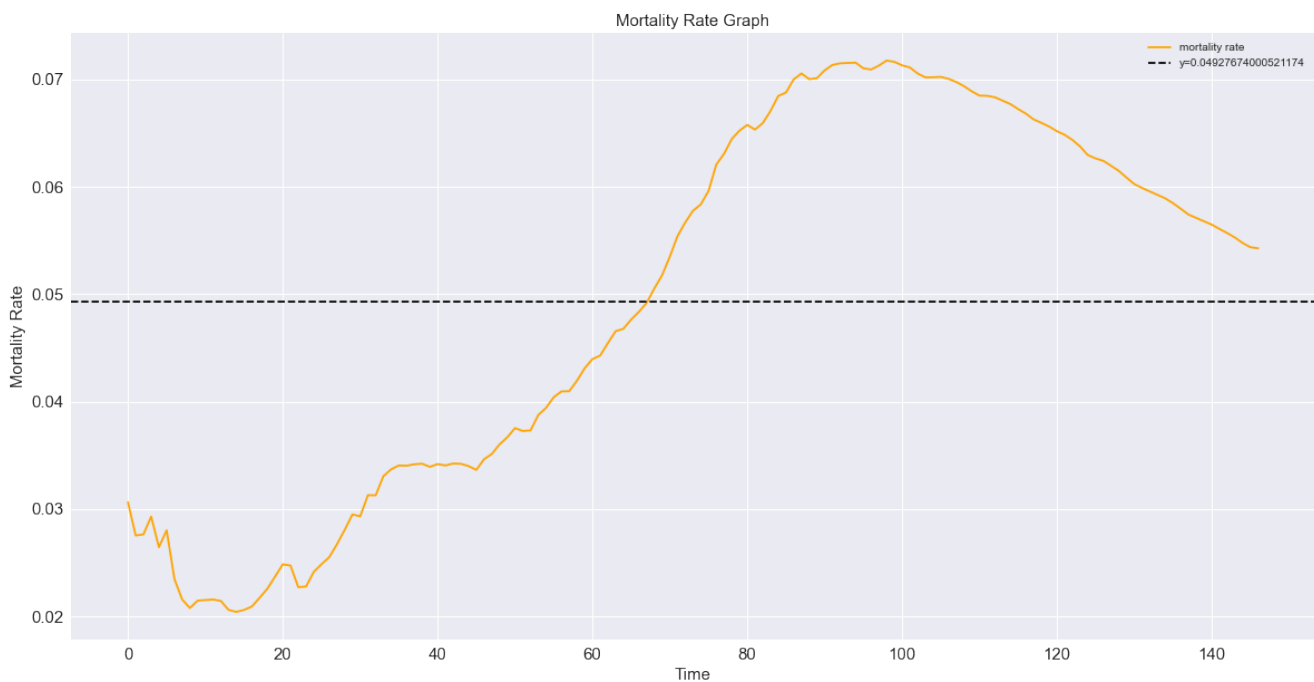


Figure 29 Shows the global mortality rate (Total Deaths/ Total Confirmed) over time in yellow while the black dashed line represents the mean mortality rate over the timeline.

Chapter 6 Conclusion and Future Scope

Coronavirus infection is going to spread in upcoming days and the number of new positive cases, deaths will rise in the near future. Precaution is better than cure, it's advised that we maintain proper physical distancing and health rather than being on a ventilator and struggling to breath. The WHO in this regards have directed various guidelines on how should one go out only when necessary like for buying groceries and daily households. It's being advised to wear mask at all time when in open or in public because the primary entry point for the virus is via the oral cavity. Which also indicates the importance of proper brushing and gargling with hot salty water at night which helps to remove the old mucus membrane in the back lining of the throat and develop a new one which acts as a protective covering inside the oral cavity.

One must avoid touching his eyes and face while in public as it's important that we our self might not turn out to be the carriers for the virus. Wash hands for at least twenty seconds which will help to kill all the bacteria and micro-organisms via increment in osmotic pressure inside them. Avoid physical contacting and maintain social distancing by self-isolating in homes whenever one feels minor symptoms like high fever, cough, difficulty breathing, nausea, fatigue. This will help your dear one's not to catch infection or being a vector for the virus. Small children can be asymptotic to the virus as they don't have a well-developed immune system compared to adults. So, they tend to be asymptomatic. Serious conditions like pneumonia can be fatal to small children and elderly as this would affect their breathing and might interrupt proper oxygen supply to other organs resulting in organ failure.

Use of alcohol based sanitizers is important and must for all. But, might be flammable so avoid their usage prior handling fire or one can suffer sever burns. Maintaining healthy diet, mental health and maintaining fitness via proper exercise makes one less susceptible to corona virus. And it's advised for people who have recovered from coronavirus to donate their plasma, as antibodies of a recovered individual can help an infected individual survive through corona.

We need to keep a check on the current situation from corona cases, if there is a need to re-lockdown we should do it, without a second thought, as loss of life is an irreparable loss whereas economy can be revived again. Throughout the world MSME, hospitality, civil aviation, agriculture and allied sectors are badly hurt which includes hotels, restaurants and even popular education destinations like Australia, Canada, Germany will be ignored till this pandemic fly round. Scholars are expecting the worst recession since the Great Depression in the 1930s. Due to Pandemic domestic consumption and consumer sentiment are weak which ultimately results in delayed investment that will further create pressure on the growth which we calculate in terms of GDP.

REFERENCES

Journal / Conference Papers

- [1] Zhong Zheng, Zhixian Yao, Ke Wo, Junhua Zheng “The diagnosis of pandemic coronavirus pneumonia: A review of laboratory examination and laboratory test”, Journal of Clinical Virology, 2020, 128
- [2] Manisha Putra, Malavika Kesavan, Kerri Breckney, David N. Hackney, and Kimberlyn M. Roosa “Forecasting the impact of coronavirus disease during delivery hospitalization: an aid for resource utilisation”, COVID-19 Pragnency Research, 2020
- [3] Mūcahid Barstugan, Umut Ozkaya, and Saban Ozturk “Coronavirus (COVID-19) Classification using CT Images by Machine Learning Methods”, 2020
- [4] Reza Kamali Kakhi, Mohammad Kamali Kakhi and Alireza Neshani “COVID-19 target A specific target for coronavirus detection”, Gene Reports 2020, 20

Web

- [1] Gender and COVID-19, who.int
- [2] scikit-learn.org
- [3] matplotlib.org
- [4] plotly.com

Data acquisition and data repository

- [1] WHO (<https://covid19.who.int/WHO-COVID-19-global-data.csv>)
- [2] Johns Hopkins Whitenig School of Engineering , Center for System Science and Engineering (<https://github.com/CSSEGISandData/COVID-19>)
- [3] Ministry of Health and Family Welfare, Government of India (<https://www.mohfw.gov.in>)

###