FORECASTING NOVEL CORONAVIRUS PHASE III USING TIME-SERIES ANALYSIS

By

Yerraballi Suresh Kumar Reddy

19MCAR0080

Under the guidance of

Dr. Gagesh D

Associate Professor

School of CS & IT

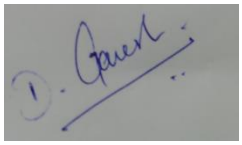
Jain (Deemed-to-be-University)

Department of MCA School of Computer Science & IT, Jain Knowledge Campus, Jayanagar 9th Block, Bangalore

NOV - 2021

CERTIFICATE

This is to certify that the report titled ***Forecasting Novel Coronavirus Phase III using Time-Series Analysis*** *is a bonafide record of work done by* ***Yerraballi Suresh Kumar Reddy (19MCAR0080)*** *of Jain (Deemed to be University), Bengaluru, in partial fulfillment of the requirements of V Semester MCA during the year 2021.*



Project Guide

Head of Department (MCA)

Head School of CS &IT

Examiner-1

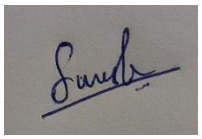
Examiner-2

Department of MCA School of Computer Applications,
School of CS & IT,
Jain Knowledge Campus, Jayanagar 9th Block, Bangalore
Nov-2021

DECLARATION

I affirm that project work titled "***Forecasting Novel Coronavirus Phase III using Time-Series Analysis***" admitted in partial fulfillment for the award of **MASTER OF COMPUTER APPLICATIONS (General)** is the original work carried out by me. It has not formed from the part of any other project work submitted for award of any degree or diploma, either in this or any other University.

Yerraballi Suresh Kumar Reddy



19MCAR0080

ACKNOWLEDGEMENT

I do acknowledge the support and encouragement of all people who helped me throughout the completion of this project.

I conjointly wish to give thanks **Dr. M. N Nachappa**, Head of School of Computer Science & IT, Jain deemed to be university, Knowledge campus, Bangalore, Karnataka for his support and cordial cooperation.

I give thanks the project Committee Members, **Prof. Bhuvana J,** Mentor and Program coordinator, Department of Master of Computer Application for providing for providing the support and steerage to try to analysis work. Her timely direction and motivation helped me to stay my patience throughout this journey.

Moving further, I would like to give my sincere gratitude to Project Centric Learning Coordinator, **Dr. Manju Bargavi S.K.**, Professor, Department of Master of Computer Application for helping me in completing my project work as per the university requirements.

I am highly indebted to my guide **Dr. Ganesh D**, Associate Professor, Department of MCA - School of Computer Science & IT for providing a valuable guidance throughout the course of the project to submit successfully.

I am indebted to my parents and Friends; whose constant support and blessings gave a driving force to go on and to all my friends for their continued moral and material support throughout the course of the project. I am grateful to the Almighty for giving me strength and ability to complete the project successfully.

Yerraballi Suresh Kumar Reddy

19MCAR0080

ABSTRACT

Early 2020, the World Health Organization declared the novel coronavirus (COVID-19) outbreak as a global pandemic spread over 170 countries. The unavailability of any COVID-19 drugs at Phase one and unexpected outbreak of Phase two leads to significant economic loss globally. Expect from World Health Organization (WHO) said, while many people now have immunity against coronavirus due to previous infections or vaccination, an experimental approach is needed to implement conditional measures to avoid a Phase three.

However, it is unavoidable. Further, it is unclear when Phase three will occur. So, forecasting techniques play a significant role in yielding accurate predictions by designing better strategies and making effective decisions. Forecasting models in Time Series Analysis require comprehensive historical data, which assists in identifying the trend and seasonality using the study of confirmed, recovered, and death cases. This helpsto measure the occurrence of phase three and allows the government to alter its extensive decision against possible threats and consequences.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 PROJECT DESCRIPTION

Early 2020, the World Health Organization declared the novel coronavirus (COVID-19) outbreak as a global pandemic spread over 170 countries. The unavailability of any COVID-19 drugs at Phase one and unexpected outbreak of Phase two leads to significant economic loss globally. Expect from World Health Organization (WHO) said, while many people now have immunity againstcoronavirus due to previous infections or vaccination, an experimental approach is needed to implement conditional measures to avoid a Phase three.

However, it is unavoidable. Further, it is unclear when Phase three will occur. So, forecasting techniques play a significant role in yielding accurate predictions by designing better strategies and making effective decisions. Forecasting models in Time Series Analysis require comprehensive historical data, which assists in identifying the trend and seasonality using the study of confirmed, recovered, and death cases. This helps to measure the occurrence of phase three and allows the government to alter its extensive decision against possible threats and consequences.

1.2 EXISTING SYSTEM

There are many websites that gives information on Covid19 such as ‘my gov’ and ‘Covid19India’.

1.3 OBJECTIVES

The main objective is to analyze and predict trends of COVID-19 with the help of Time-series models. To extend a new interactive web application that visually represents the spread of the COVID-19 pandemic and it’s forecast across the different regions in India. And to support the government in making an extensive decision against possible threats and consequences with the help of accurate prediction of phase III.

1.4 PURPOSE, SCOPE, AND APPLICABILITY

1.4.1 PURPOSE

The COVID-19 pandemic has led to dramatic loss of human life worldwide and presents an unprecedented challenge to public health, food systems and the world of work. It is one of the challenges mankind has ever faced. Since there is high chance of Covid-19 Phase III emerging, it is required for experimental approach to implement conditional measures to avoid Covid-19 Phase III in India.

1.4.2 SCOPE

The major problem considered in this pandemic is that people have refused to understand that following Covid guidelines will prevent future waves with different variants. There are different factors behind the each Phase of Covid-19. The several researches made has found that factors like complacency, superspreader events, more infectious variants and insufficient vaccine coverage causedPhase II. And they have predicted the factors that may cause Phase III listing Individual choices, vaccination coverage and variants.

1.4.3 APPLICABILITY

Forecasting the novel coronavirus using time series analysis helps to measure the occurrence of Phase III and allows the government to alter its extensive decision against possible threats and consequences. This application creates awareness and gives updates news on this pandemic. Researchers are benefitted with the research model details. On the whole the applicants who all get benefitted by this project are researchers, government and public people.

1.5 OVERVIEW OF THE REPORT

This report contains the detailed flow and description of Forecasting Novel coronavirus Phase III using Time-series Analysis. There is introduction to the project on objectives, existing systems, scope and applicability. Following that system analysis and requirements are detailed. Furthermore System design, implementation, testing and conclusion on the next 90 days forecast details on Covid19 Phase III using the best time series forecast model.

2. SYSTEM ANALYSIS & REQUIREMENT

2.1 PROJECT DEFINITION

The Covid-19 pandemic has led to a dramatic loss of human life worldwide and presents an unpredented challenge to public health, food systems and the world of work. It is one of the prevalent challenges mankind has ever faced and there is a lot of uncertainty prevailing over the future with respect to Covid-19. However, an experimental approach is needed to implement conditional measures to avoid a Phase III India.

The major problem is that people have refused to understand that followingCOVID guidelines will prevent forthcoming waves with different variants. The main objective of this project is to analyse and predict trends of COVID-19 with the help ofTime-series models and factors that caused Phase II. To extend a new interactive web app that visually represents the spread of the COVID-19 pandemic and its forecast COVID-19 Phase III across the different regions in India. Furthermore, to support the government in making an extensive decision against possible threats and consequences with the help of accurate prediction of Phase III.

2.2 REQUIREMENTS SPECIFICATIONS

The existing system has so many limitations. These limitations are taken into consideration and are considered as necessary requirements to develop ourapplication. Forecasting the cases by state-wise and district-wise. Since people are taking vaccination, this also should be considered to forecast the Phase III.

2.3 BLOCK DIAGRAM

2.4 SYSTEM REQUIREMENTS

2.4.1 USER CHARACTERISTIC

Our Application facilitates three different types of users, namely Admin, Researchers, and Developers. The task of the admin is to control and manage the entire flow of the website. The developer creates and maintains the whole datasets and source code of the application. Researchers are benefited with the help of supplemented source code and research models implemented in our project.

2.4.2 SOFTWARE AND HARDWARE REQUIREMENTS

Software Requirements	Configuration
OS	Windows 10
Testing Tools	
Linker	flask
IDE	Jupyter Notebook, VS Code
Packages	Pandas, numpy, keras, tensorflow, plotly, seaborn, matplotlib
Virtualization S/W	Tableau

Table 1, Software Requirements

Hardware Requirements	Configuration
Processor	i5, i7 or above
RAM	4GB or above
Graphics card	
Disk capacity	

Table 2, Hardware Requirements

2.4.3 CONSTRAINTS

2.5 CONCEPTUAL MODEL

3. SYSTEM DESIGN

3.1 SYSTEM ARCHITECTURE

The 3-tier architecture is a well-established software application architecture that builds applications into three physical and logical computing tiers as mention:

- The presentation tier-0, or client user interface.
- The application tier-1, where data is processed.
- The data tier-2, where the data associated with the application is stored and managed.

The client tier consists of an end-user interface and web module, and here users can utilize COVID19 completed statistical details, forecasted details, user-centric interactive graphs, and other filters.

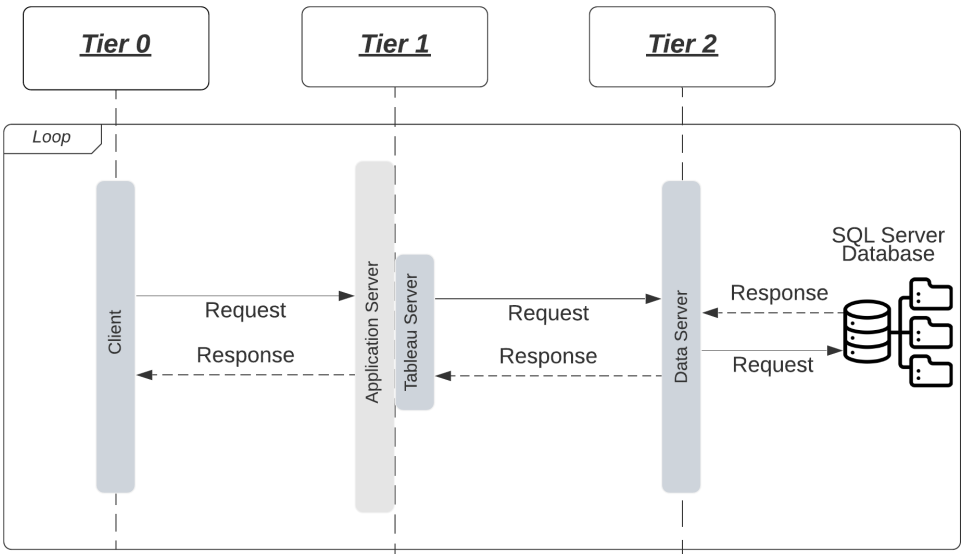


Figure-1, System Architecture

The Application-tier consists of an embedded Tableau public server for interactive dashboards and handles third-party sign-in. This tier will be hosted with the help of a feasible cloud server (Amazon S3) and Tableau Public server. The Data-tier consists of all the necessary data associated with the application in the form of the relational

database system that can be stored and managed. Here Microsoft SQL Server is addressed for implementation.

3.2 MODULE DESIGN

Web Module explains the experience of a user when interacting with Daily Pulse website. Entering the website user has access to interactive, knowledge gain and contact pages. The user is provided with interactive Covid19 dashboards in Home, Forecast and Vaccination pages. To gain more knowledge about Covid19 in India user are provided with campaigns, posters, videos, etc. on Awareness and Blog pages.To know more about the website user can visit About Us page. If the end user is a researcher and wanted to know more details on research models should register and login the website. This ensures user authentication.

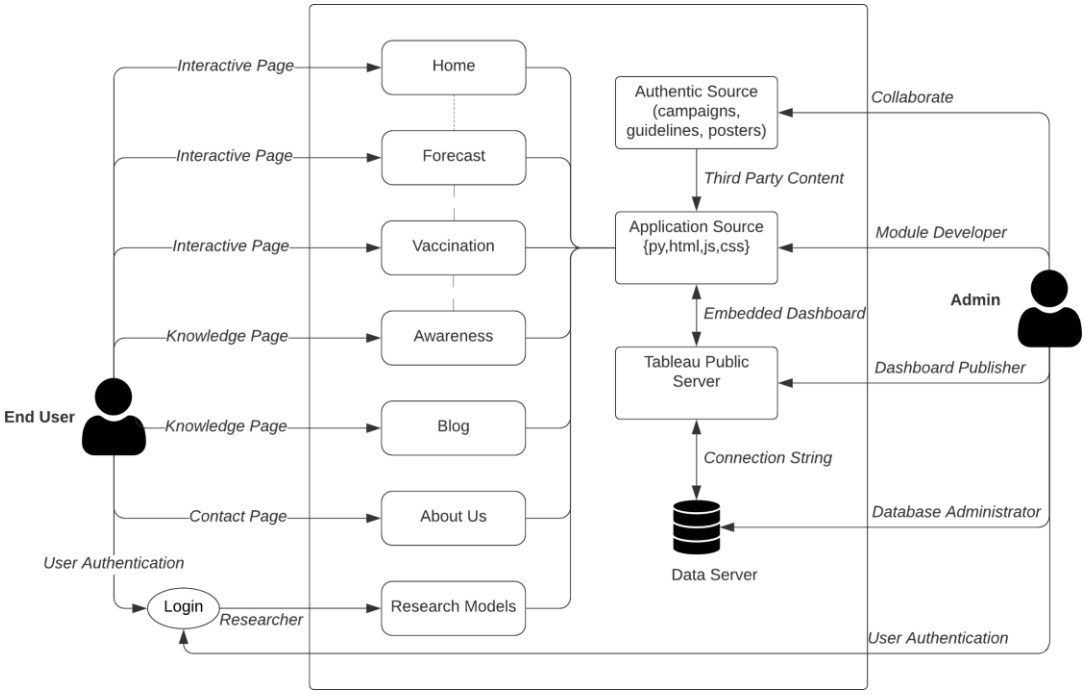


Figure-2, Web Module

By default, Admin has permission on the data server as a database administrator. Further, Admin has a Tableau Public Server account to publish Covid19 dashboards. Admin acts as a module developer from testing till the deployment of the application. Also, Admin collaborates with popular sites to pull relevant content and incorporate in web application.

The **Forecast module** explains how time-series models are used to forecast theimpact of Covid19 Phase III in India. Forecasting is broken down into four key phases. In Phase one, the data is collected from the Covid19 and the vaccination database. Phase two involves further data processing to produce meaningful information by organizing and manipulating the data collected. In phase three, the time-series models are trained with data after checking for the stationarity condition. Further different validation metrics are used to know the performance of the model. Finally, phase four ensures forecasting and selecting the most suitable model out ofall time-series models that best suits the dataset.

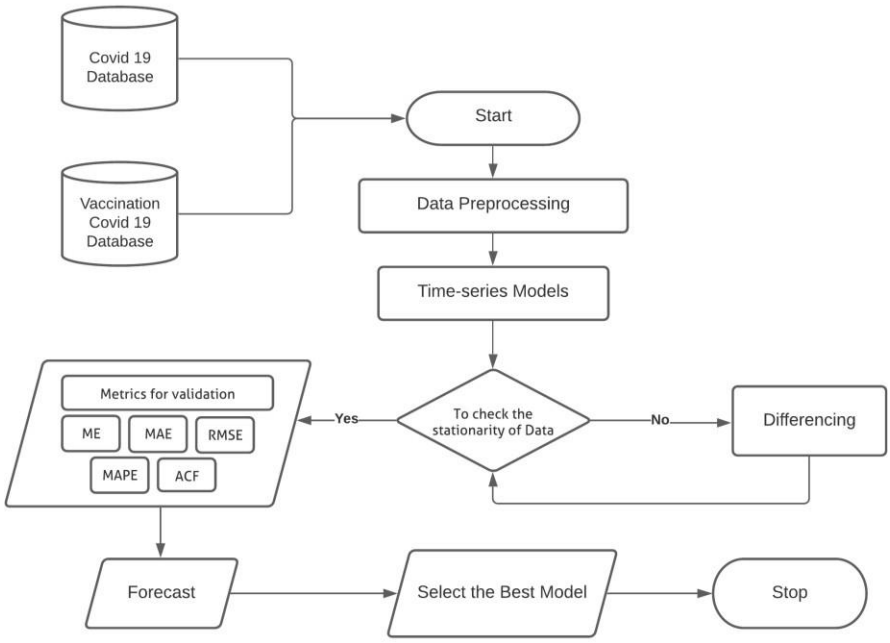


Figure-3, Forecast Module

Database module explains the backend of the web application. The datasets are collected from Covid19 India.org with API requests. The data collected is stored in CSV format using Python IDE. Following the file is processed with the SSIS package. Data flow in SSIS defines the flow of your entire data from a source to a destination. In data source configuration, it checks for condition. If the data exists, itis dropped; else, the data is transformed to desired form and incrementally inserted in SQL server destination. The upcoming forecasting analytics process is done with the help report query.

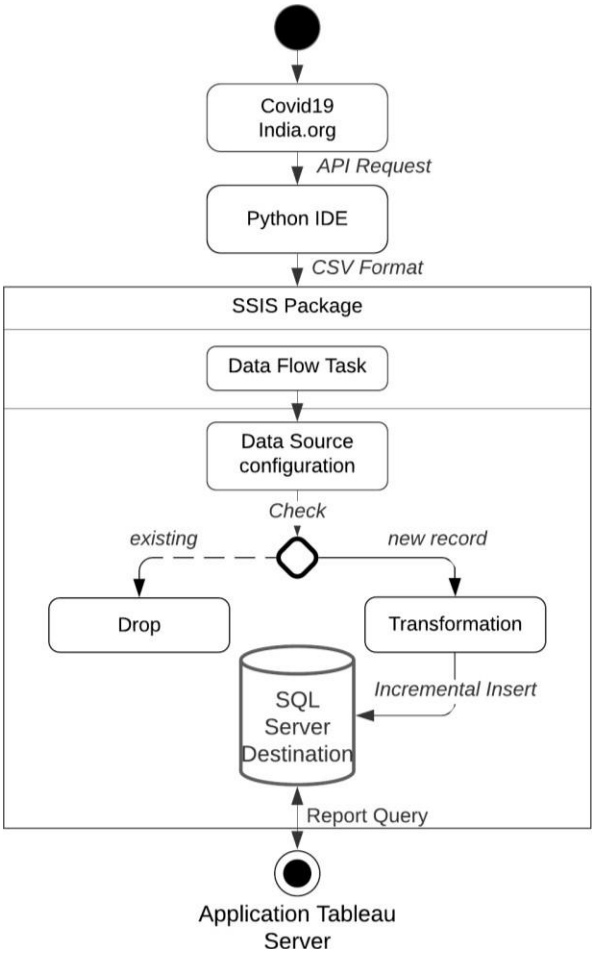


Figure-4, Database Module

3.3 DATABASE DESIGN

3.3.1 TABLES AND RELATIONSHIPS

Entity-Relationship Data Model (ERM) is a logical representation of the data for an organization or a business area, mainly expressed in terms of Entities, Attributes, and Relationships. An ERD (Entity-Relationship Diagram) is a graphical representation of an Entity-Relationship Model. Identified entities are period, state details, state-wise cases, and vaccination. As shown in figure-5, attributes for the identified entities were state code, state name, district name, confirmed, recovered, deceased, reported date, vaccination details, and other related fields discussed in the data dictionary. Attributes are appropriately categorized into different types, namely unique, composite, single-valued, multi- valued, derived, and key attribute. Relationships between entities are state details

having state-wise cases details. Period entity disclosed state-wise cases and vaccination recorded on a reported date. The cardinality of relationships is shown in the figure-5.

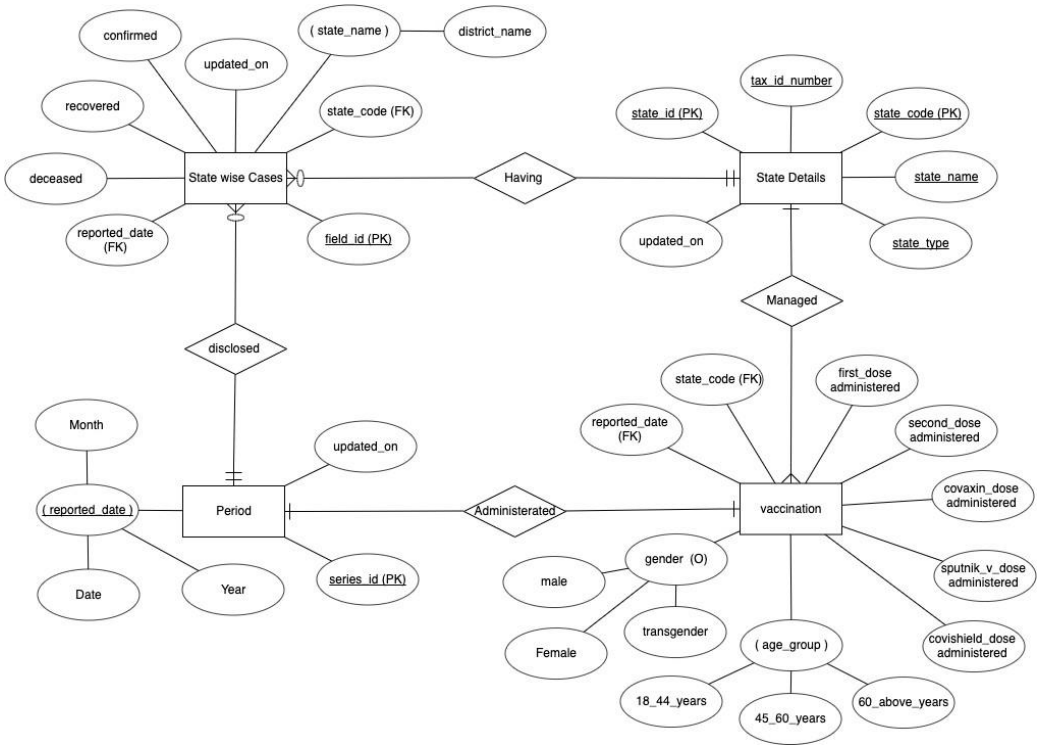


Figure-5, Entity Relationship Diagram

The primary data source is in the tabular form, so required database normalization was considered while designing the table relation. Once the fine-grained data are stored, it concerns performance, storage optimization, cloud migration, and cost-effectiveness. With the assistant of ERD, database scheme design originated efficiently, as shown in figure-6. The period and State Details table act as a superset for the table, implemented with the primary and foreign key references. reported_date, state_code are the widely used primary key related to all other tables. Additional cumulative records can be easily fetched using a simple report query with the help tableau tool and other filters provided for end-user. As part of future research, the work shown in figure-6 is subject to slight changes. Updated date columns were included to do a sanity check for the data accuracy while scheduled incremental insert.

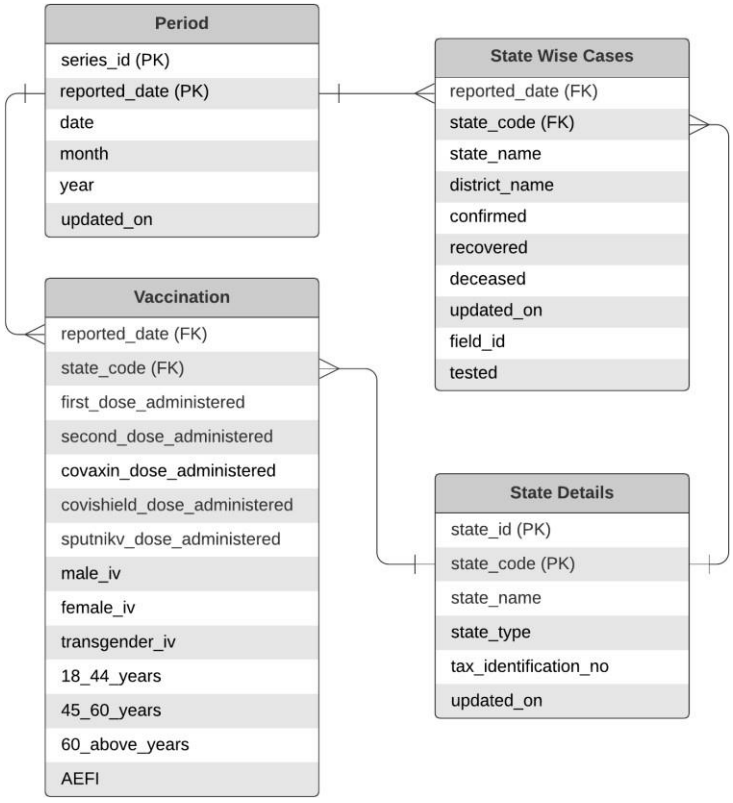


Figure-6, Entity Relationship Diagram

3.3.2 DATA INTEGRITY AND CONSTAINTS

Integrity constraints are a set of rules. Restrictions on the contents of the database or on database operations.

Domain Integrity Constraints can be defined as the definition of a valid set of values for an attribute. NOT NULL, CHECK, UNIQUE, DEFAULT. In our database design we've considered state_code, state_name, district_name, and reported_date as NOT NULL integrity constraint. In CHECK integrity constraint we're checking whether confirmed, recovered, and deceased cases are greater than or equals to zero. Following we're checking the reported_date is not greater than current date. By default, we're setting confirmed, recovered, and deceased as '0' using DEFAULT integrity constraint. state_code and reported_date columns follow UNIQUE integrity constraint.

Entity Integrity Constraint states that primary key value can't be null. As it is used to identify individual rows in relation. Here we considered state_code and reported_date as primary key in state_details and period tables respectively.

Referential Integrity Constraint is specified between two tables which establishes parent child relationship between tables. All the primary key enabled fields are acts as foreign key in vaccination and state_wise_cases tables.

3.4 INTERFACE DESIGN AND PROCEDURAL DESIGN

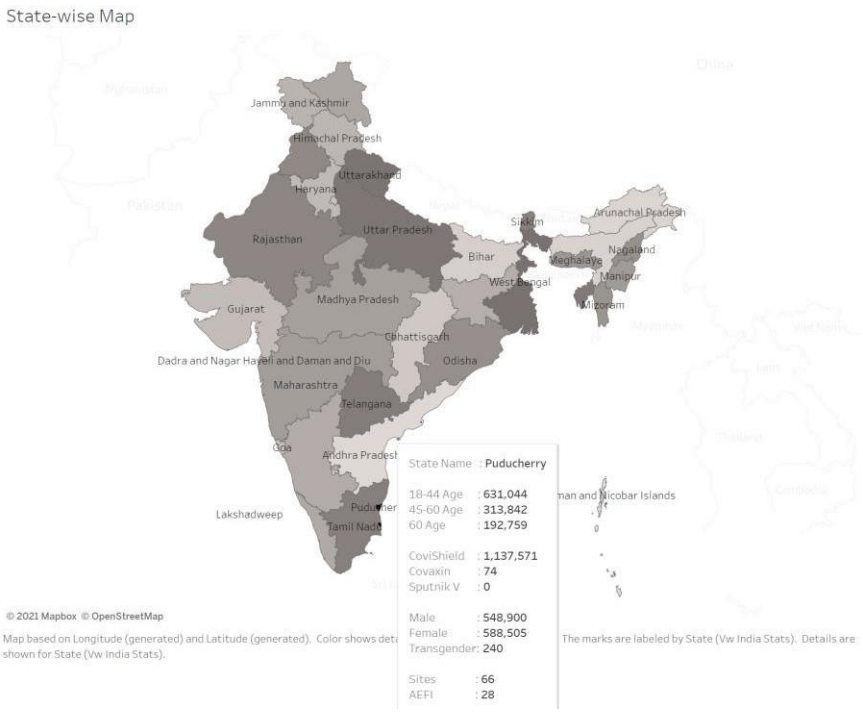
3.4.1 USER-INTERFACE DESIGN

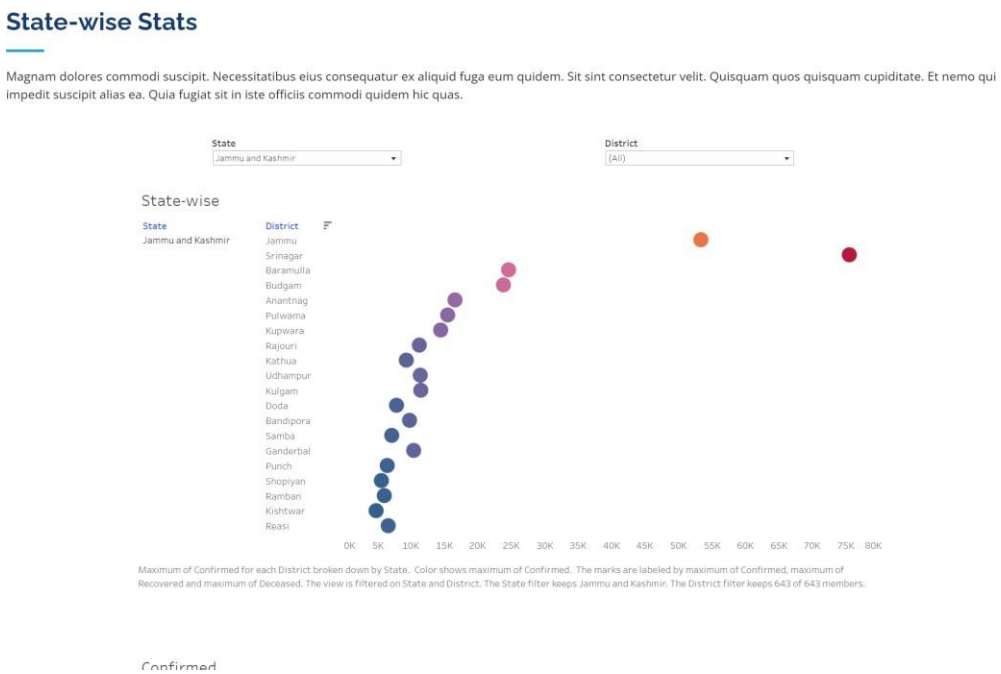
Our User-Interface design facilitates three different types of users, namely Admin, Researchers, and Developers. The task of the admin is to control and manage the entire flow of the website. The developer creates and maintains the whole datasets and source code of the application. Researchers are benefited with the help of supplemented source code and research models implemented in our project. All the tasks are carried over the website using the Tableau server and accessible in standard web browsers. Since users are more comfortable grasping context from pictorial representations, we make this advantage of a project to show the results in graphical which is interactive and easy to use.

FORECAST: - It is a collection of statistical graphs and a dashboard. The interactive dashboard covers different charts such as regional geographic map, pie

chart, line chart, scatter plot, heat map, and bar chart for representing various statistical inferences. The dashboard also shows the number of views and downloads.

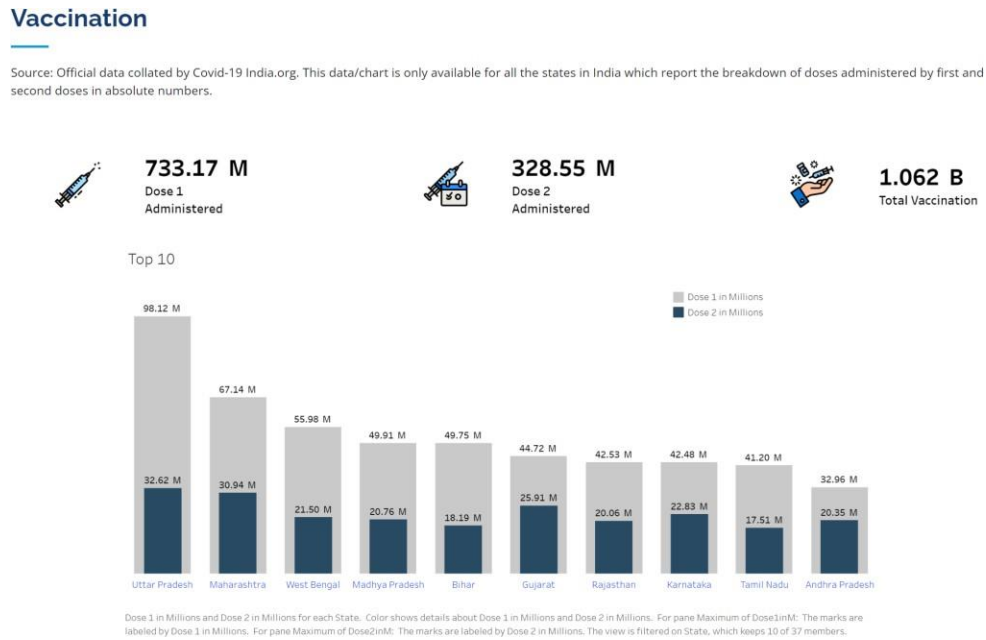
- Statistically, we are projecting the information on daily & total confirmed cases, daily & total recovery cases, and daily & total deceased.
- We are projecting the forecasted confirmed, recovery, deceased cases for the upcoming 90 days.
- Filters are enabled for an interactive experience for individual viewers by providing state-wise, district-wise, and date-wise customization.
- Further, we were delivering a rate of growth in confirmed, recovery, and deceased.





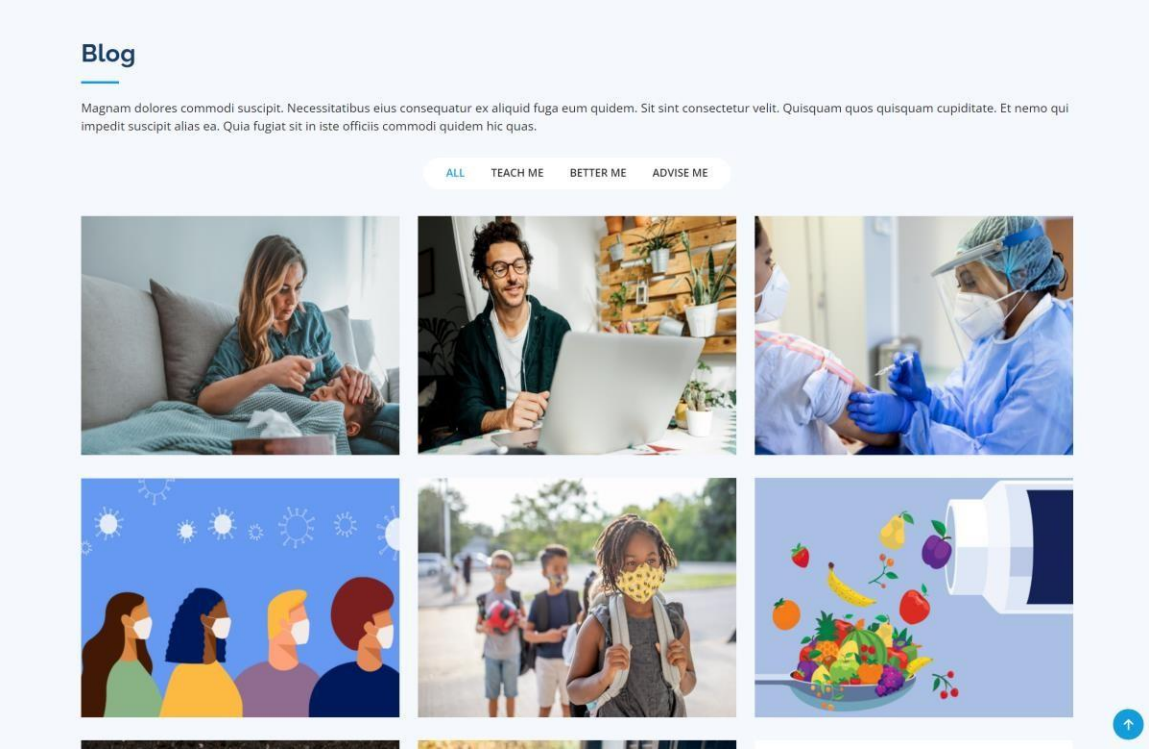
VACCINATION: - This page shows all aspects of vaccination details in India. Particularly,

- The distribution of the different vaccines taken by individuals all over India is statistically compared against the first dose, the second dose, age category, and gender-wise.
- The availability of vaccines information is given state-wise.
- The vaccination registration applications and websites details are provided
- along with their guidelines.



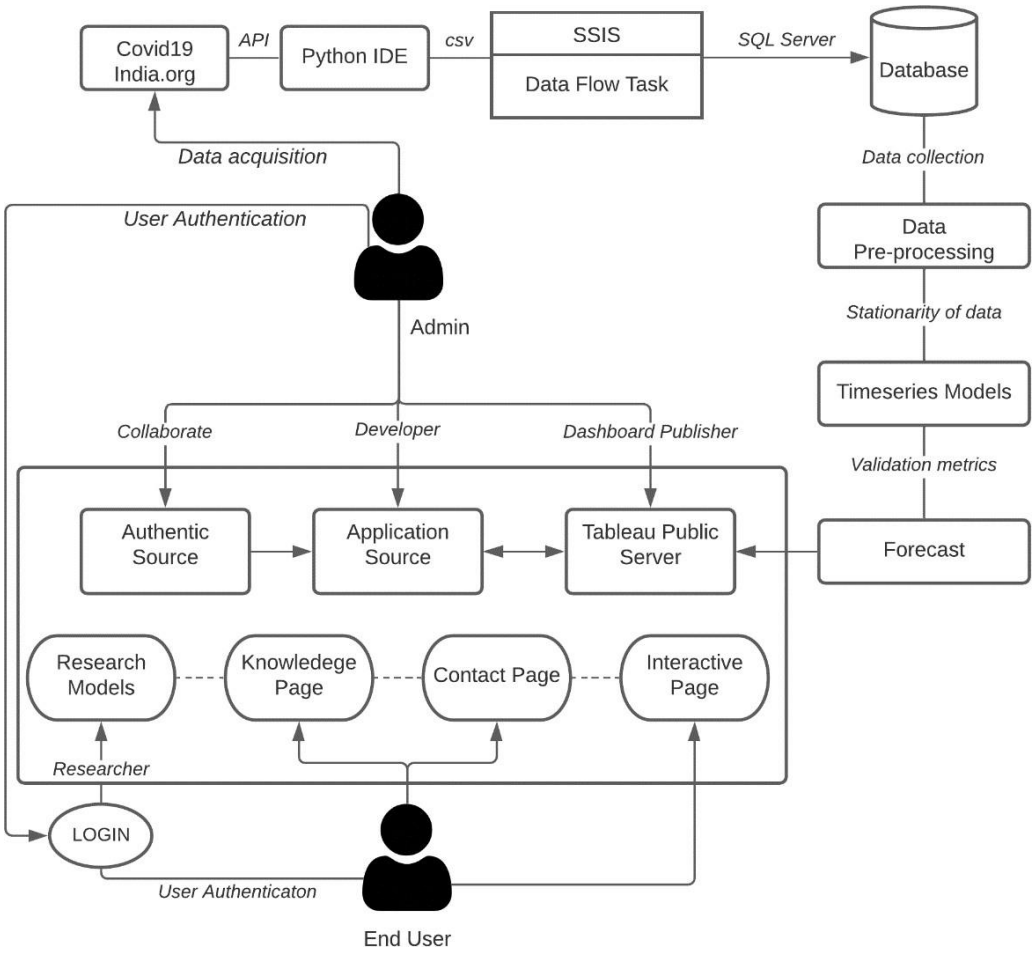
BLOG:- Posters and videos included concerns on self-awareness among the society such as symptoms, social distancing, post covid precautions, etc.

- General guidelines to be followed on quarantine, public gathering, and post covid recovery by the Ministry of Health Services.
- NGO campaigns, social hubs, and their contact details are provided.
- Government helpline to reach out at the earliest stage of the virus.



AWAERNESS:- This page is a collection of articles and blogs related to Covid19. Each blog contains a title, posted date, and description.

3.4.2 APPLICATION FLOW/CLASS DIAGRAM



4. IMPLEMENTATION

4.1 IMPLEMENTATION APPROACHES

Google Colab

4.2 CODING STANDARD

Python

4.3 CODING DETAILS

```
from google.colab import drive
drive.mount('/content/drive')
# importing required libraries
import pandas as pd
import numpy as np
import datetime as dt
import keras
import tensorflow as tf

import matplotlib.pyplot as plt
import plotly.express as px
from plotly.subplots import make_subplots
import plotly.graph_objects as go
import seaborn as sns
# Approach 1
# This dataset is directly connected via API
case_time_series = pd.read_csv('https://api.covid19india.org/csv/latest/case_time_series.csv')
# saving the dataframe
case_time_series.to_csv('case_time_series.csv')
# Approach 2
# Accessing the dataset .csv from drive
case_time_series = pd.read_csv('/content/drive/MyDrive/Project/case_time_series.csv')
case_time_series["Date_YMD"]=pd.to_datetime(case_time_series.Date_YMD,format="%Y-%m-
    %d")
```

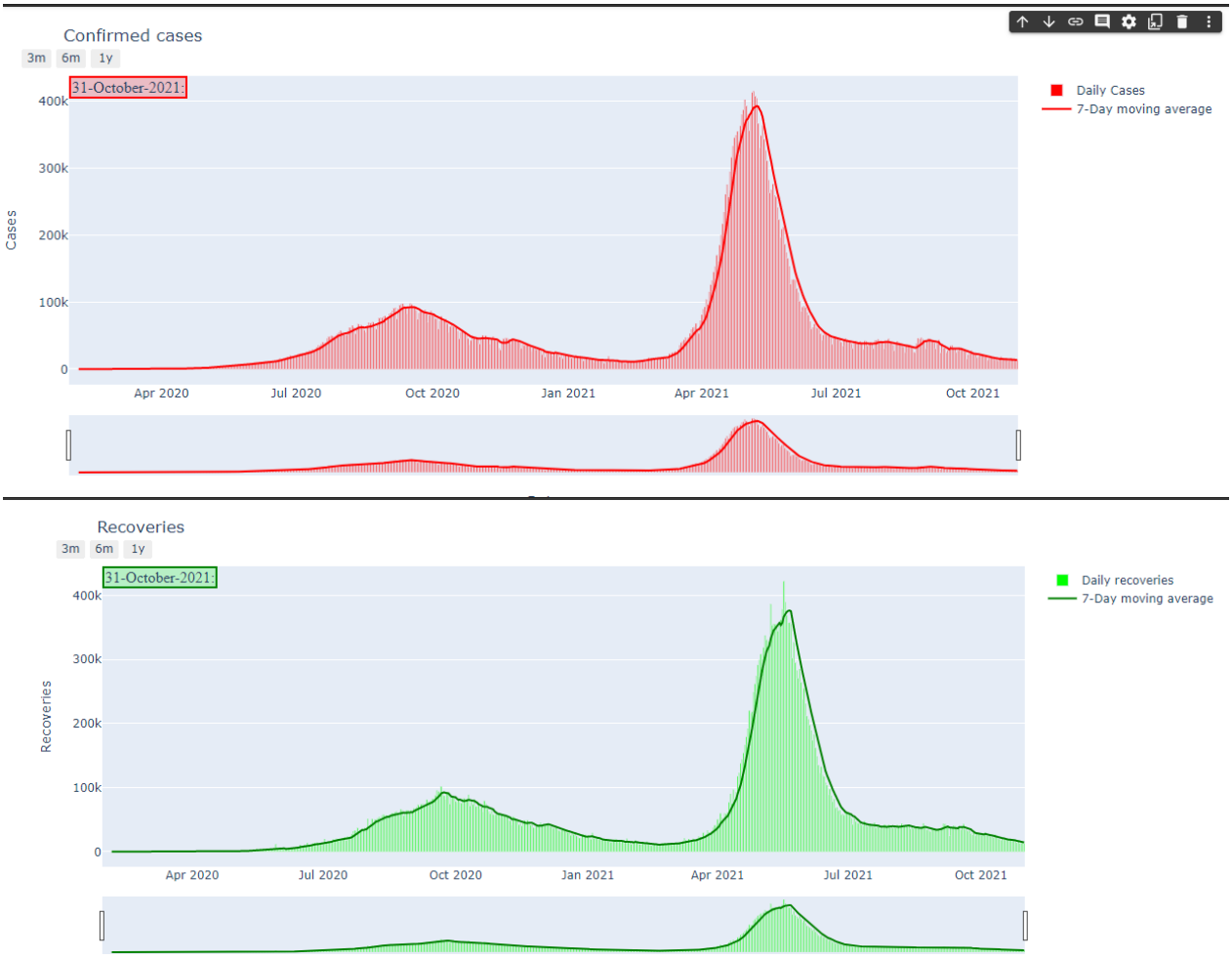
```
case_time_series["Date"]=pd.to_datetime(case_time_series.Date,format="%d %B %Y")
case_time_series["Date"]=case_time_series["Date"].dt.strftime('%d-%B-%Y')
#Cases
TC=case_time_series["Total Confirmed"]
TR=case_time_series["Total Recovered"]
TD=case_time_series["Total Deceased"]
X=case_time_series["Daily Confirmed"]
Y=case_time_series["Daily Recovered"]
Z=case_time_series["Daily Deceased"]
D=case_time_series["Date_YMD"]
DF=case_time_series["Date"]
TOTC=case_time_series["Total Confirmed"]
TOTR=case_time_series["Total Recovered"]
TOTD=case_time_series["Total Deceased"]
DAC=(X-Y-Z)  #Daily Active cases
TAC=(TC-TR-TD) #Total Active cases
#Moving Averages
case_time_series['CMA'] = X.rolling(window=7).mean()
case_time_series['RMA'] = Y.rolling(window=7).mean()
case_time_series['DMA'] = Z.rolling(window=7).mean()
case_time_series['AMA'] = DAC.rolling(window=7).mean()

CMA=case_time_series["CMA"]
RMA=case_time_series["RMA"]
AMA=case_time_series["AMA"]
DMA=case_time_series["DMA"]
#Daily change in stats
delC=X.diff()
delR=Y.diff()
delD=Z.diff()
delA=DAC.diff()
```

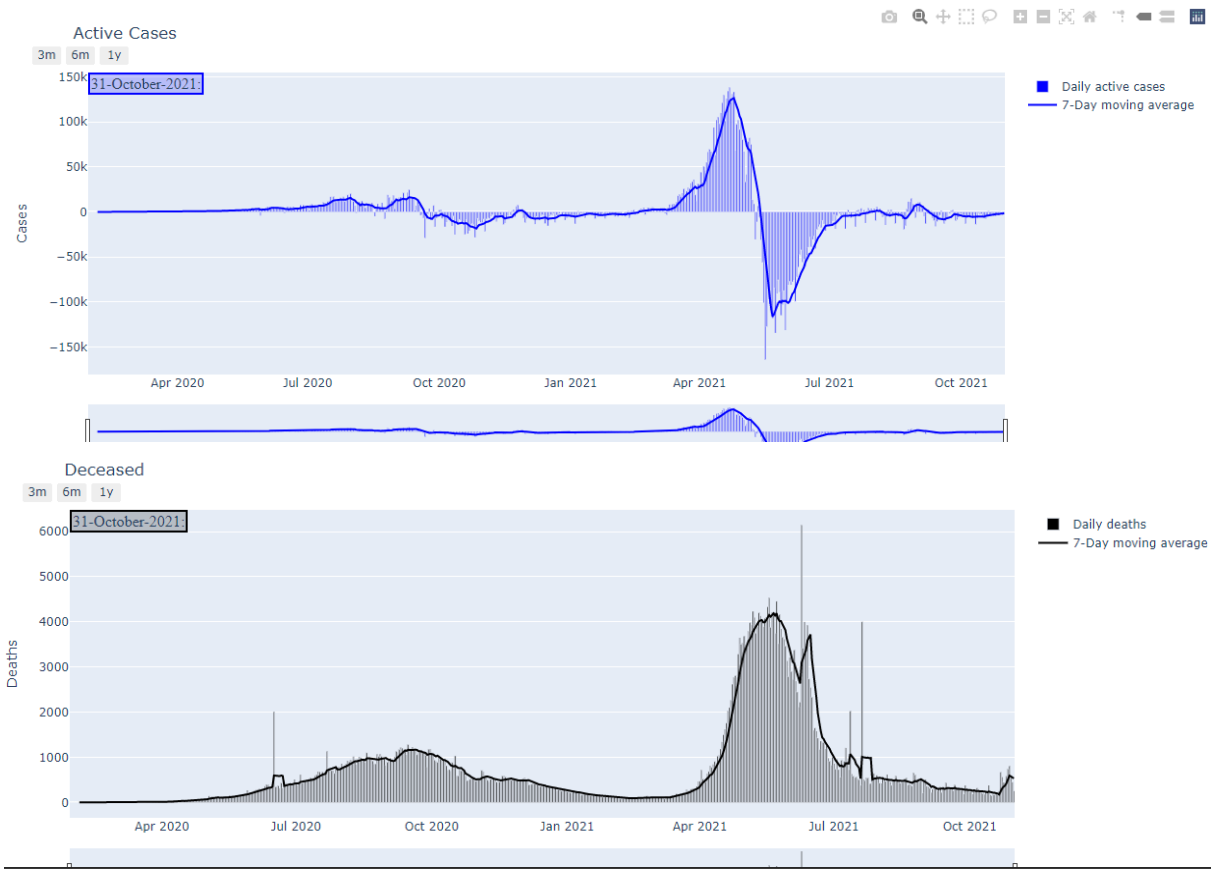
```
Atoday=X[len(X)-1]-Y[len(Y)-1]-Z[len(Z)-1]
TOTA=TOTC-TOTR-TOTD
#Confirmed Cases
confirmed = go.Figure(data=[go.Bar(x=D,y=X,
                            name="Daily Cases",
                            marker = dict(color = 'rgba(255, 0, 0,1)'),
                            text="Peak: {}, Average: {}".format(X.max(),round(X.mean()),2)),
                        go.Scatter(x=D, y=CMA,
                            mode='lines',
                            name="7-Day moving average",
                            line_color="red")]
                    )
confirmed.add_annotation(text="{}:<br>"
                        .format(DF[len(DF)-1]),
                        bgcolor='rgba(255, 0, 0,0.2)',
                        font=dict(size=16,family="orbitron"),
                        borderwidth=2,
                        bordercolor="red",
                        x=0,
                        xanchor="left",
                        xref="paper",
                        y=1,
                        yanchor="top",
                        yref="paper",
                        showarrow=False,
                        align="left"
                    )
confirmed.update_layout(
    margin=dict(l=20,r=20,b=50,t=70),
    title="Confirmed cases",
    xaxis_title="Date",
    yaxis_title="Cases",
```

```
    xaxis=dict(
        rangeselector=dict(
            buttons=list([dict(count=3,step="month",stepmode="backward"),
                          dict(count=6,step="month",stepmode="backward"),
                          dict(count=1,step="year",stepmode="backward")])
        ),
        rangeslider=dict(visible=True),
    )
)
```

Forecasting_Covid19_Phase_3_using_Time_Series_Models - Colaboratory (google.com)

4.4 SCREENSHOTS




Active Cases
3m 6m 1y
31-October-2021
150k
100k
50k
0
−50k
−100k
−150k
Cases
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Apr 2021
Jul 2021
Oct 2021
Daily active cases
7-Day moving average
Deceased
3m 6m 1y
31-October-2021
6000
5000
4000
3000
2000
1000
0
Deaths
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Apr 2021
Jul 2021
Oct 2021
Daily deaths
7-Day moving average

6. CONCLUSION

This report contains the detailed flow and description of Forecasting Novel coronavirus Phase III using Time-series Analysis. There is introduction to the project on objectives, existing systems, scope and applicability. Following that system analysis and requirements are detailed. Furthermore System design, implementation, testing and conclusion on the next 90 days forecast details on Covid19 Phase III using the best time series forecast model.

REFERENCES

www.github.com (for datasets Covid19 india.api)

www.youtube.com (tutorials)

Novel Corona Virus 2019 Dataset | Kaggle

Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET - ScienceDirect

Time Series Analysis of COVID-19 Data to Study the Effect of Lockdown and Unlock in India | SpringerLink

Prediction of confirmed cases of and deaths caused by COVID-19 in Chile through time series techniques: A comparative study (plos.org)

Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming (nih.gov)

Correlation Determination between COVID-19 and Weather Parameters Using Time Series Forecasting: A Case Study in Pakistan (hindawi.com)