**Data Analysis and Visualization OF PYTHON on COVID-19**

Major Project Report

Submitted to the Centurion University

In partial fulfillment of requirements for the award of degree

**Bachelor of Technology in**

**Electronics and Communication Engineering**



by

Ch.Praneeth

211801131002

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING CENTURION UNIVERSITY OF TECHNOLOGY AND MANAGEMENT (CUTM-AP)VIZIANAGARAM)**

**535003**

**CENTURION UNIVERSITY OF TECHNOLOGY AND MANAGEMENT**

**(CUTM-AP) VIZIANAGARAM**

### DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING



**CERTIFICATE**

This is to certify that their report entitled "**Data Analysis and Visualization on covid-19**" submitted by CH.Praneeth**(211801131002)** in **Department of Electronics and Communication Engineering** in partial fulfilment of the B.Tech. Degree in **Electronics and Communication Engineering** is a bonafide record of the seminar work carried out by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Aswin kumar

Department of Computer Science

And EngineeringCenturion University (CUTM)

**External Examiner**

**ABSTRACT**

The main objective of this project is to give information you will learn how to preprocess and merge datasets to calculate needed measures and prepare them for an Analysis. In this project, we are going to work with the COVID19 dataset, published by John Hopkins University, which consists of the data related to the cumulative number of confirmed cases, per day, in each Country. Also, we have another dataset consist of various life factors, scored by the people living in each country around the globe. We are going to merge these two datasets to see if there is any relationship between the spread of the virus in a country and how happy people are, living in that country.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **S. NO.** | **TITLE** | **PAGE** |

**Abstract**  3

**List of Figures** 6

**1**. **Introduction** 7

### 2. Existing and Proposed System 8

### 3. System Requirements 9

3.1 Hardware Requirements 9

3.2 Software Requirements 9

1. **Methodology** 10  4.1 Modular Design 10

4.1.1 Data Acquisition 11

4.1.2 Data Processing 11

4.1.3 Implementation 11

4.1.4 Analysis 11

4.1.5 Visualization 11

4.2 What is Python

4.3.Modules 12

4.3.1 Pandas 13

4.3.2 Plotly. Express

4.3.3 folium 14

1. **Data set** 15

**6.**  **Implementation** 17

1. **Results** 20

1. **Conclusion** 25

1. **References** 26

**LIST OF FIGURES**

# FIG. NO. TITLE PAGE NO

**7.1 pair plot on covid-19 daily cases 20**

**7.2 bar graph of number of cases daily 20**

### 

**7.3 bar graph in total cases in india 21**

**7.4 BAR graph in number of cases in states 21**

**7.5 sunburst graph pf different cases 22**

**7.6 heat map of different states**  22

7.**7 map of Indian states 23**

**DASH BOARD 23**

**1.INTRODUCTION**

Data science is the study of data to extract knowledge and insights from the data and apply knowledge and actionable insights. In this tutorial, we will work on IPL Data Analysis and Visualization Project using Python where we Indian government has announced the lockdown and closure of educational institutions as a logical solution to enforce social distancing within communities. The nationwide lockdown has had a tremendous impact on the education system of the country, especially for students from rural areas. Since the Indian education system is dominated by classroom study, the present scenario has made the functioning of the educational institutions go very difficult. All educational activities like examinations, school admissions, entrance tests of various universities and competitive examinations, others, are being held during this period. As the days are passing by with no immediate solution to stop this outbreak, the closure of schools and universities is hugely affecting the learning across the country. The structure of the Indian education system i.e. learning methodology, teaching techniques & assessment methodologies, is quite affected, resulting in a shift to online education with 35 Educational Resurgence Journal Volum2,Issue 5,Jan.2021 ISSN 2581-9100 most focus on virtual education to accomplish the set aims and objectives. But only a handful of schools and universities could adopt such methods and the low-income private and government schools are quite inefficient to adopt the same, thus resulting in a shutdown.

**2.Existing and Proposed System**

There are many other visualizations available on the web for having a proper analysis for the requirement to be full filled. But I believe that they are outdated and the techniques and technology used while creating those visualizations are far outdated from what we have today.

Therefore, given the state of technology now and the changes made to programming languages. In my visualization I’ve used and added different types of visualization modules and graphs which weren’t in use previously. The graphs I’ve produces are interactive so that the users can also operate them. I’ve linked the graphs with an html file so that the user may view the 3D

1. **SYSTEM REQUIREMENTS**

**3.1. Hardware** **Requirements:**

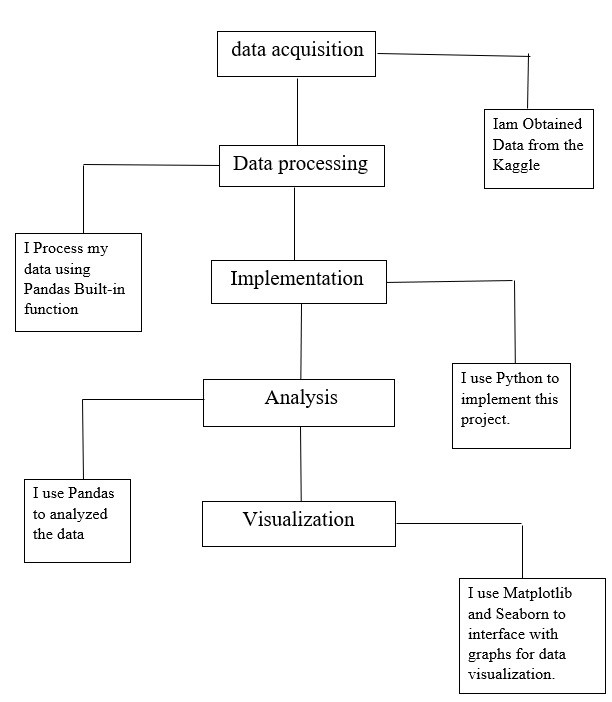
1. Processor-Pentium IV Processor
2. Speed-1.1GHZ
3. RAM-256MB
4. Storage-20GB

**3.2. Software Requirements:**

1. Operating system -XP/7/8/8.1/10
2. IDLE used-Python

**4. METHODOLOGY**

### 4.1. Modular Design



#### 4.1.1. Data Acquisition

For collecting the raw data I used online sources like Kaggle, GitHub, Google Scholar, etc. Data acquisition is theprocess of measuring physical world conditions and phenomena such as electricity, sound,temperature and pressure. This is done through the use of various sensors which sample the environment’s analog signals and transform them to digital signals using an analog-to-digital converter.

#### 4.1.2. Data Processing

Now the acquisited data is processed in excel for filtering the garbage values. I also computed the data using mathematical calculations in excel for a better result. Data preprocessing is the process oftransforming raw data into a useful, understandable format**.** Real-world or raw data usually has inconsistent formatting, human errors, and can also be incomplete. Data preprocessing resolves such issues and makes datasets more complete and efficient to perform data analysis.

#### 4.1.3. Implementation

After processing the data, I imported the data in python environment for analysis. Implementation is the realization of an application, or execution of a plan, idea, model, design, specification, standard, algorithm, or policy.

#### 4.1.4. Analysis

After implementation the data has to analyze for visualization. In analysis stage the user can clearly analyze the data in the form of rows and columns for a clear understanding.Data analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains.

#### 4.1.5. Visualization

After analysis we come to the visualization stage where we used some python built-in-modules such as plotly where I produced interactive graphs for a better representation of the data.Data Visualization is used to communicate information clearly and efficiently to users by the usage of information graphics such as tables and charts. It helps users in analyzing a large amount of data in a simpler way. It makes complex data more accessible, understandable, and usable.

**4.2**. **What is Python?**

* Python is a high-level, interpreted, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation.
* Python uses garbage collection and has dynamic typing. It supports a variety of programming paradigms, including procedural, objectoriented, and functional programming as well as structured programming (especially this). Due to its extensive standard library, it is frequently referred to as a "batteries included" language.
* Python was created by Guido van Rossum in the late 1980s to replace the ABC programming language, and it was originally made available as Python 0.9.0 in 1991.
* Python frequently causes programmers to fall in love due to the enhanced productivity it offers. The edit-test-debug cycle is extraordinarily quick because there is no compilation step.
* Python programs are simple to debug since a segmentation failure is never caused by a bug or incorrect input. Instead, the interpreter raises an exception when it finds a mistake. The interpreter prints a stack trace if the application doesn't catch the exception.
* Setting breakpoints, evaluating arbitrary expressions, inspecting local and global variables, stepping through the code one line at a time, and other features are all possible with a source level debugger. Python's ability to perform introspection is demonstrated by the debugger, which is developed in Python.

### 4.3. Modules

For this project I’ve used the following modules:-

1.Pandas

2.Plotly. Express

3.Plotly. Graph

4.folium

#### 4.3.1. Pandas

For the purpose of manipulating and analyzing data, the Python programming language has a software package called pandas. It includes specific data structures and procedures for working with time series and mathematical tables. It is free software distributed under the BSD license's three clauses. The word is derived from "panel data," a phrase used in econometrics to refer to data sets that contain observations for the same persons throughout a range of time periods. Python data analysis is a play on words in the name of the thing. When Wes McKinney worked as a researcher at AQR Capital from 2007 to 2010, he began creating the pandas that would eventually become famous.

4.3.2. Plotly Express

An interactive, open-source plotting toolkit for Python, plotly provides over 40 different chart types for a variety of statistical, financial, geographic, scientific, and three-dimensional use-cases.

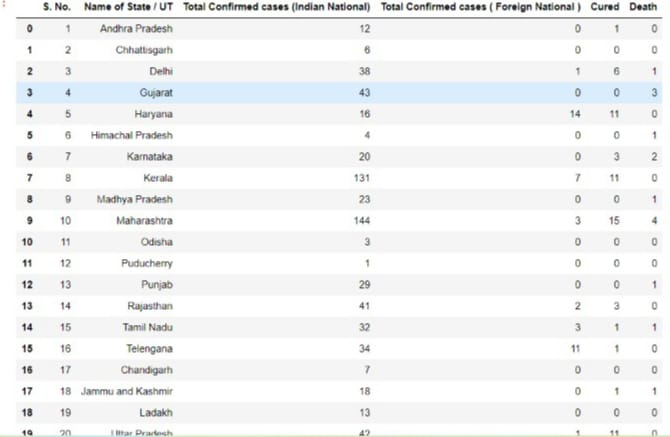
Plotly a python tool that is built on top of the Plotly JavaScript library (plotly.js), allows users to create stunning interactive web-based visualizations that can be viewed in Jupyter notebooks, saved to standalone HTML files, or used as a component of web applications that are entirely written in Python and served using Dash. To distinguish it from the JavaScript library, the plotly Python library is sometimes called "plotly.py."

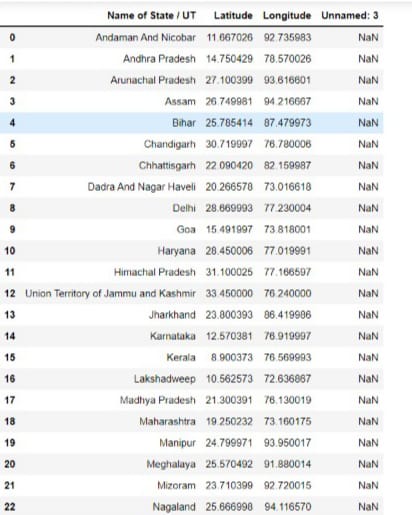
4.3.3Folium:

* folium builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the leaflet.js library. Manipulate your data in Python, then visualize it in on a Leaflet map via folium. folium makes it easy to visualize data that’s been manipulated in Python on an interactive leaflet map. Folium is a powerful data visualization library in Python that was built primarily to help people visualize geospatial data. With Folium, one can create a map of any location in the world. Folium is actually a python wrapper for leaflet.js which is a javascript library for plotting interactive maps.

**5.DATASET**

So for performing any visualizations the main thing is the data. These data can be of any type i.e. raw data, filtered data, selective data etc. A data set (sometimes spelled dataset) is a group of data. In the case of tabular data, a data set relates to one or more database tables, where each row refers to a specific record in the corresponding data set and each column to a specific variable.





**IMPLEMENTATION**

import numpy as np

import pandas as pd

import seaborn as sns

import plotly.graph\_objects as go

import matplotlib.pyplot as plt

import folium

df = pd.read\_csv('case\_time\_series (2).csv')

sns.pairplot(df)

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

data = pd.read\_csv('case\_time\_series (2).csv')

Y = data.iloc[61:,1].values

R = data.iloc[61:,3].values

D = data.iloc[61:,5].values

X = data.iloc[61:,0]

plt.figure(figsize=(25,8))

ax = plt.axes()

ax.grid(linewidth=0.4, color='#8f8f8f')

ax.set\_facecolor("black")

ax.set\_xlabel('\nDate',size=25,color='#4bb4f2')

ax.set\_ylabel('Number of Confirmed Cases\n',

size=25,color='#4bb4f2')

plt.xticks(rotation='vertical',size='20',color='red')

plt.yticks(size=20,color='red')

plt.tick\_params(size=20,color='red')

for i,j in zip(X,Y):

ax.annotate(str(j),xy=(i,j+100),color='white',size='13')

ax.annotate('Second Lockdown 15th April',

xy=(15.2, 860),

xytext=(19.9,500),

color='white',

size='25',

arrowprops=dict(color='red',

linewidth=0.025))

plt.title("COVID-19 IN : Daily Confirmed\n",

size=50,color='#28a9ff') #28a9ff

ax.plot(X,Y,

color='#1F77B4',

marker='\*',

linewidth=4,

markersize=15,

markeredgecolor='#28a9ff')

data = pd.read\_csv('district (1).csv')

data.head()

re=data.iloc[:15,5].values

de=data.iloc[:15,4].values

co=data.iloc[:15,3].values

x=list(data.iloc[:15,0])

plt.figure(figsize=(13,10))

ax=plt.axes()

ax.set\_facecolor('black')

ax.grid(linewidth=0.4, color='#8f8f8f')

plt.xticks(rotation='vertical',size='20',color='black')#ticks of X

plt.yticks(size='20',color='black')

ax.set\_xlabel('\nDistrict',size=25,color='#4bb4f2') #17th may 2020

ax.invert\_xaxis()

ax.set\_ylabel('No. of cases\n',size=25,color='#4bb4f2')

plt.tick\_params(size=20,color='white')

ax.set\_title('andhrapradesh District wise breakdown\n',size=50,color='#28a9ff')

plt.bar(x,co,label='re')

plt.bar(x,re,label='re',color='green')

plt.bar(x,de,label='re',color='red')

for i,j in zip(x,co):

ax.annotate(str(int(j)),xy=(i,j+3),color='white',size='15')

plt.legend(['Confirmed','Recovered','Deceased'],

fontsize=20)

import plotly.express as px

import pandas as pd

DISTRICTS = ["westgodavari", "vizianagaram", "visakhapatnam", "srikakulam", "prakasam", "nellore",

"kurnool", "krishna", "kadapa", "guntur","east godavari","chittor","ananthapur"]

ACTIVE= ["47", "100", "13", "134", "34","99",

"187", "53", "3", "10", "41","7","19"]

CONFIRMED = ["122", "177", "52", "417", "102","367" ,"611" ,

"150", "66", "14", "75", "7","70"]

RECOVERED = ["4", "0", "0", "8", "0","15",

"19", "3", "0", "0", "1","1","1"]

df = pd.DataFrame(dict(DISTRICTS=DISTRICTS, ACTIVE=ACTIVE, CONFIRMED=CONFIRMED, RECOVERED=RECOVERED))

fig = px.sunburst(df, path=['DISTRICTS','ACTIVE','CONFIRMED','RECOVERED'] )

fig.show()

import pandas as pd

df= pd.read\_csv('Covid cases in India (3).csv') #states of covid cases

df.style.background\_gradient(cmap='Reds')

data= pd.read\_csv('Covid cases in India (3).csv')

Indian\_Cord=pd.read\_csv('Indian Coordinates.csv')

dd=pd.merge(Indian\_Cord,df,on='Name of State / UT')

dd

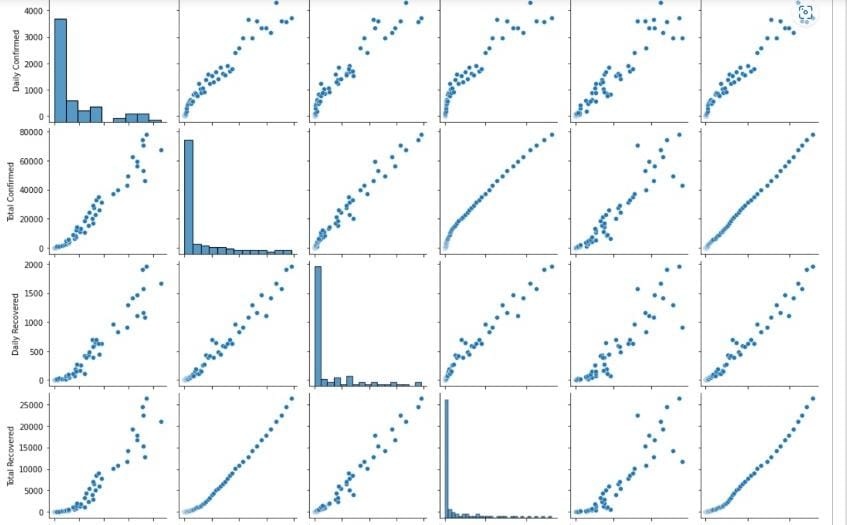
map=folium.Map(location=[20,70],zoom\_start=4,tiles='Stamenterrain')

for lat,long,value, name in zip(dd['Latitude'],dd['Longitude'],dd['Total Cases'],dd['Name of State / UT']):folium.CircleMarker([lat,long],radius=value\*0.8,popup=('<strong>State</strong>: '+str(name).capitalize()+'<br>''<strong>Total Cases</strong>: ' + str(value)+ '<br>'),color='red',fill\_color='red',fill\_opacity=0.2).add\_to(map

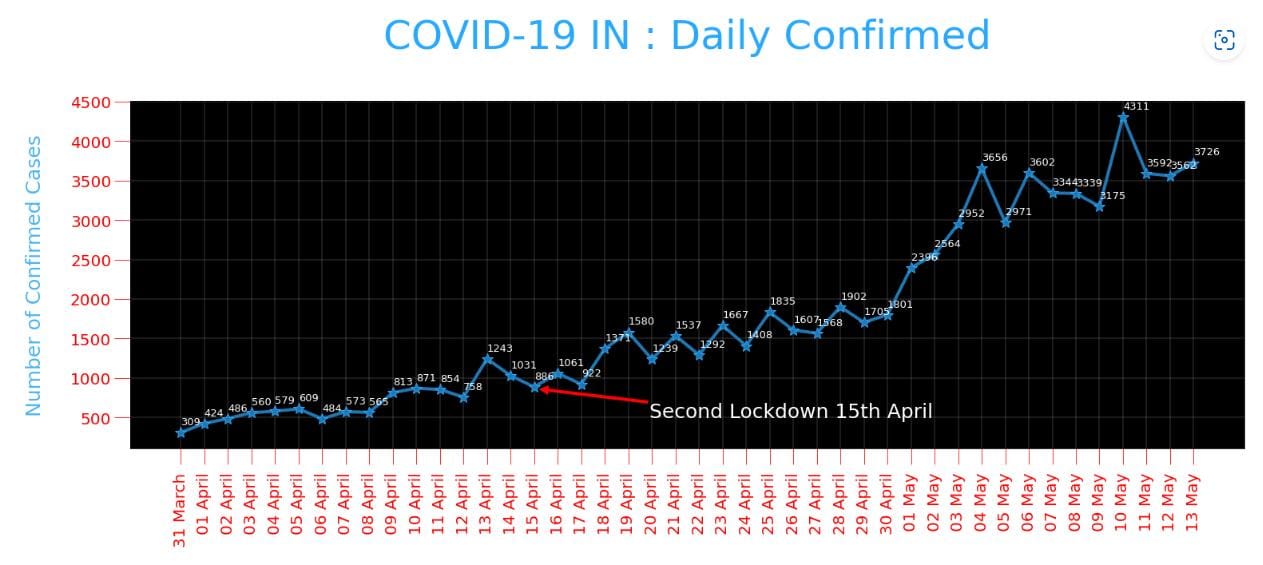
map

**RESULT**

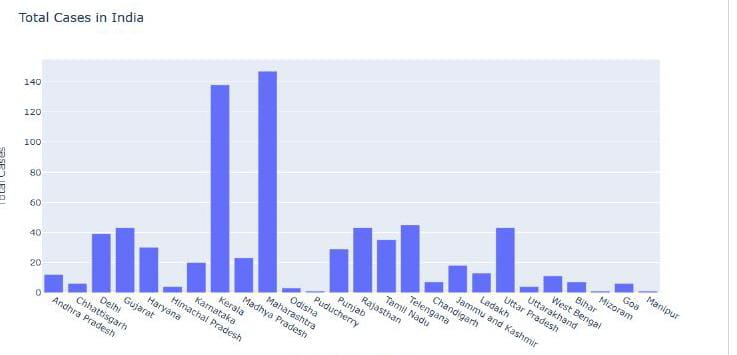
**PAIR PLOT ON DAILY CASES**



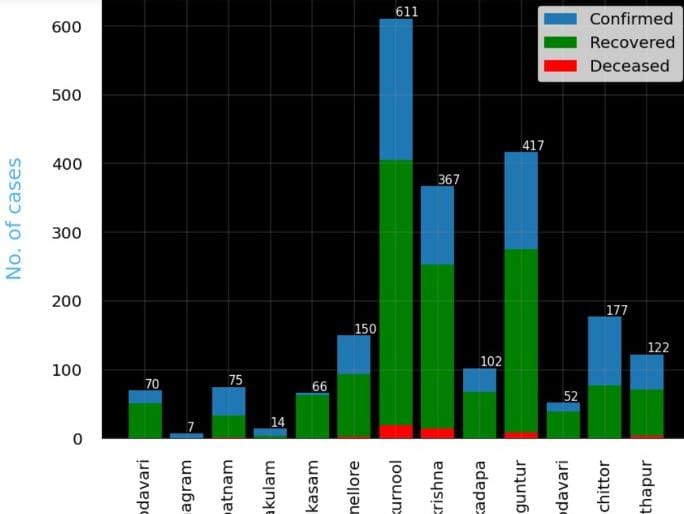
**DAILY CONFIRMED CASES**



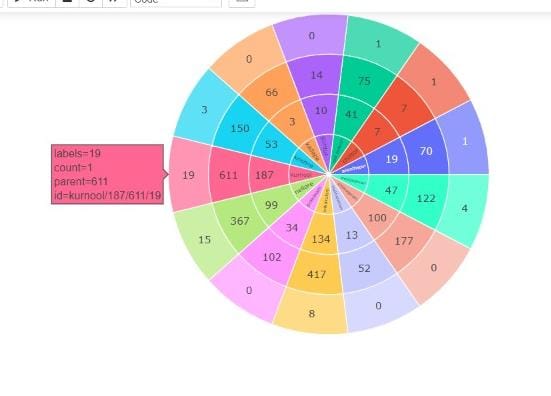
Total cases in different states



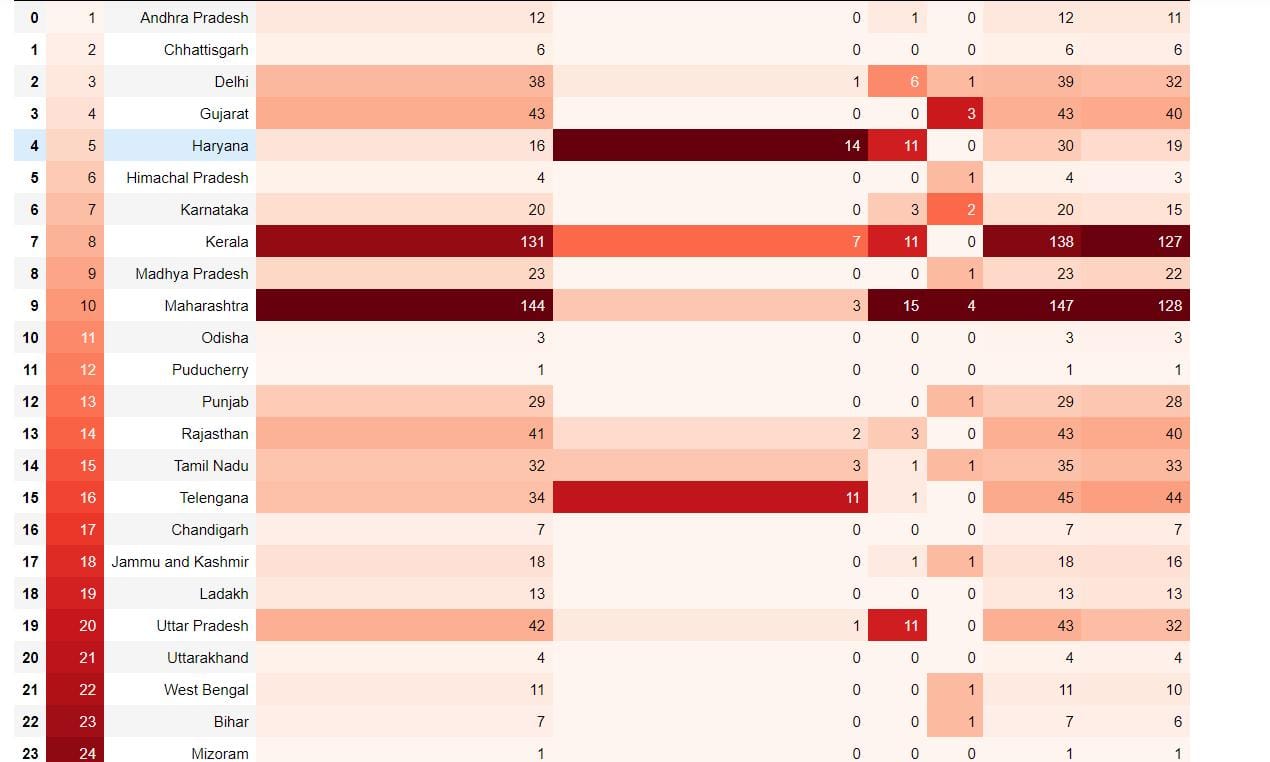
NUMBER OF CASES IN STATES



SUNBURST GRAPH OF DIFFERENT STATES



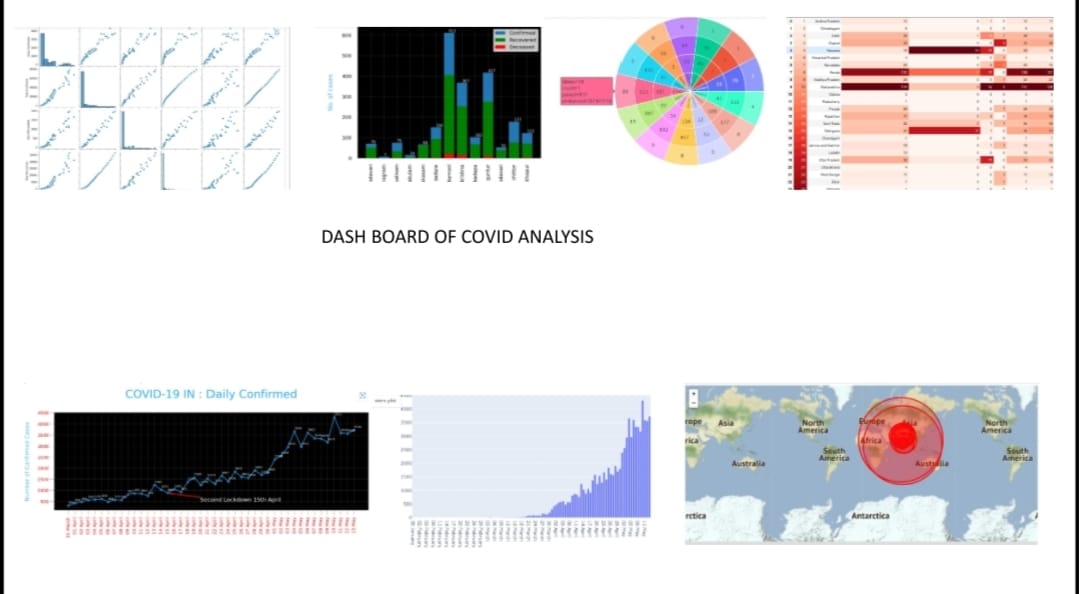
**HEAT MAP OF DIFFERENT STATES**



**MAP**



**DASH BOARD**



**CONCLUSION**

Trust you preferred our undertaking on COVID-19 Data examination and Visualization utilizing Python. We just recorded a nuts and bolts to medium high level investigation here, to provide you with a thought of how to utilize the informational collection. You can connect your own information investigation COVID-19 with Python libraries and even truly do some Data Science projects. COVID-19 has impacted immensely the education sector of India. Though it has created many challenges, various opportunities are also evolved. The Indian Govt. and different stakeholders of education have explored the possibility of Open and Distance learning by adopting different digital technologies to cope up with the present crisis of COVID-19. India is not fully equipped to make education reach all corners of the nation via digital platforms. The students who aren’t privileged like the others will suffer due to the present choice of digital platforms. The priority should be to utilize digital technology to create an advantageous position for millions of young students in India. It is need of the hour for the educational institutions to strengthen their knowledge and Information Technology infrastructure to be ready for facing COVID-19 like situations. Even if the COVID-19 crisis stretches longer, there is an urgent need to take efforts on maximum utilization of online platforms. India should develop creative strategies to ensure that all children must have sustainable access to learning during pandemic COVID19. As online practice is benefitting the students immensely, it should be continued after the lockdown. Further detailed statistical study may be undertaken to explore the impact of COVID-19 on education system of India.

**8.REFERENCES**

1. International Journal of Recent Technology and Engineering

(IJRTE) ISSN: 2277-3878, Volume-7 Issue-4S, November 2018

1. International Journal of Engineering Science InventionISSN (Online):

2319 – 6734, ISSN (Print): 2319 – 6726 [www.ijesi.org](https://jovian.ai/outlink?url=http%3A%2F%2Fwww.ijesi.org) Volume 3 Issue 6ǁ June 2014 ǁ PP.07-11

1. <https://coed.dypvp.edu.in/educational-resurgence-journal/documents/jan-2021/35-40.pdf>
2. <https://medium.com/codex/covid-19-analysis-with-python-b898181ea627>

5.DATA SET FROM KAGGLE

6. W3School (<https://www.w3schools.com/python>)

7.Plotly (<https://plotly.com/python/>)

8.Pandas (<https://pandas.pydata.org/>