SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

Data Visualization on COVID-19 day to day analysis and prediction

Project Supervisor: **Dr. Joshila Grace**

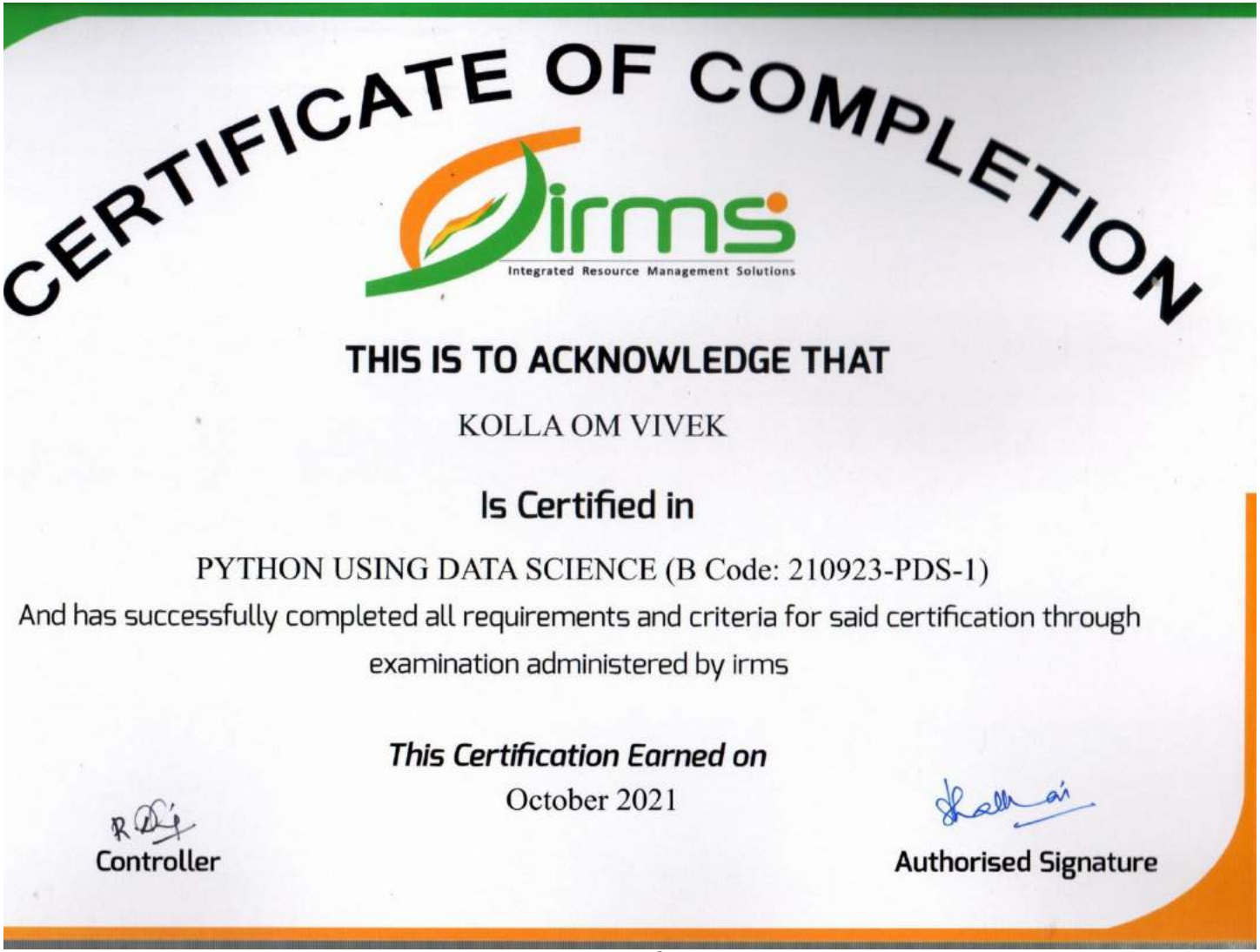
Name of the Student: **Mr. Kolla Om Vivek**

Register Number: **39110511**

Presentation Outline

- Course Certificate
- Introduction
- Objectives
- System Architecture
- Module Implementation
- Project Implementation
- Results and Discussions
- Conclusion
- References

Course Certificate



CERTIFICATE OF COMPLETION

irms
Integrated Resource Management Solutions

THIS IS TO ACKNOWLEDGE THAT

KOLLA OM VIVEK

Is Certified in

PYTHON USING DATA SCIENCE (B Code: 210923-PDS-1)

And has successfully completed all requirements and criteria for said certification through examination administered by irms

This Certification Earned on

October 2021

Controller

Authorised Signature

Introduction

- Corona virus identified in **2019** ,SARS-CoV-2. This virus was firstly identified in **People's Pepublic Of China** in **Wuhan.**

- **Data Science** for COVID-19 presents leading-edge research on data science techniques for the **detection, treatment and elimination of COVID-19.**

- During the current coronavirus pandemic, monitoring the **evolution of COVID-19** cases is of utmost importance for the authorities to make informed **policy decisions** (e.g., lock-downs), and to raise awareness in the general public for taking appropriate public health measures.

Introduction

- **Data Visualization** gives us a clear idea of what is the information, means by giving it **visual context through graphs.**

- This makes the data more natural for the **human mind** to comprehend and therefore makes it easier to identify **trends, patterns and outliers within large data sets.**

- At the end, we can analyze the data by using the graphs , bar plots, etc., and get some useful insights from the data.

Objectives

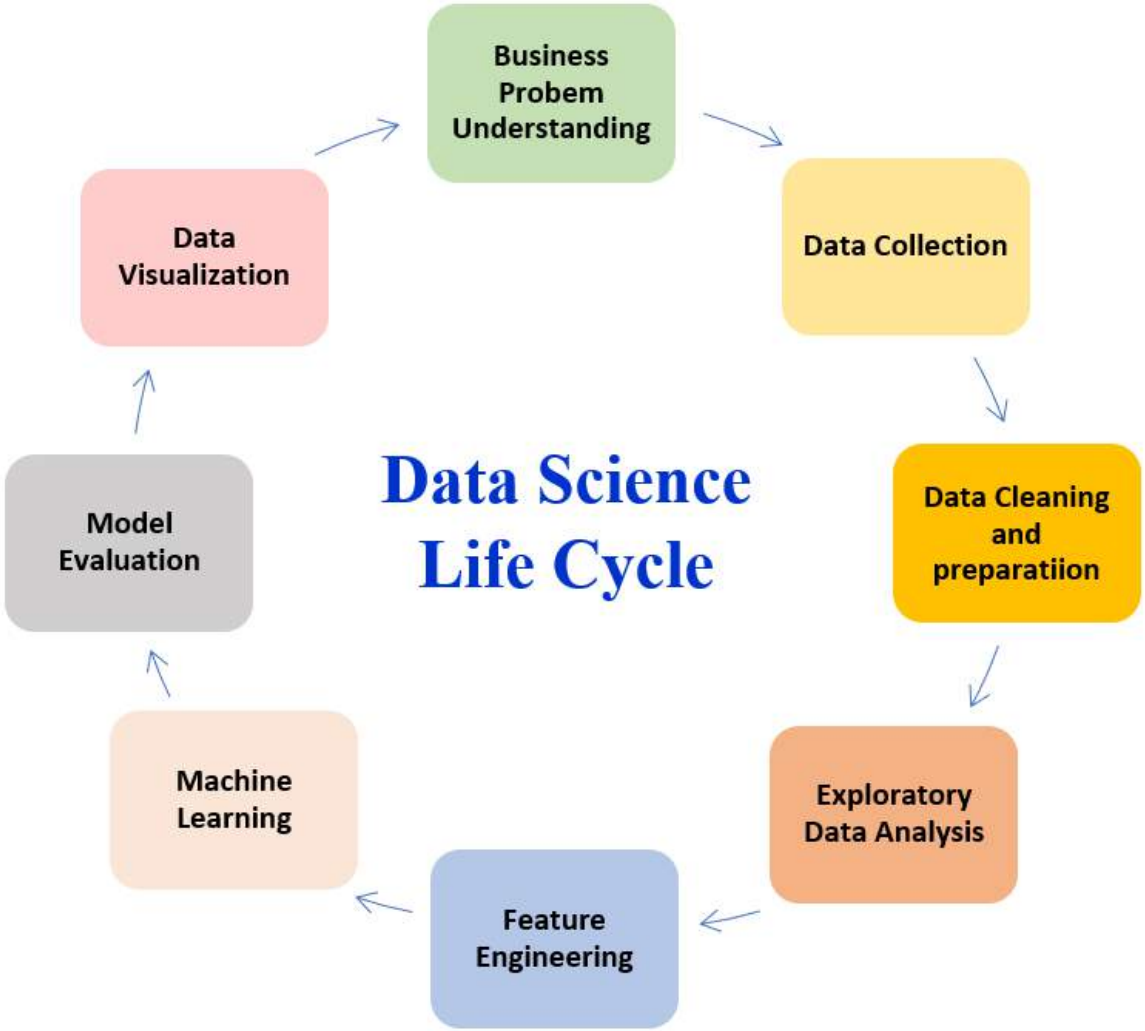
Objective of the project :

- Main goal is to express **data in easier way** so that everyone can understand
- Visualizing the data of covid-19 of all the country's around the world and all the state's of India

Scope of the project :

- To get some useful data about Covid-19 like **Case's, Deaths and Conformed case's**
- We are going to achieve the objective by performing various processes like **cleaning and analyzing** the data by using various plotting tools like **matplotlib, plotly , seaborn .**

System Architecture



Module Implementation

This Module contains mainly 3 parts:

1. Datasets(Covid-19) are csv file's which contains the raw data.
2. We used **plotly** library to visualize the data, which is an open-source and we use **pywebio library** for web-application and it is mostly used by data scientists for **data visualization.**
3. The plots are plotted after the data is processed and the data is ready for the analysis.

Project Implementation

- Modules.
- Hardware and Software Requirements.
- Construction.
- Measurement and Analysis.
- Methodology.

Modules

Pandas :

- pandas is a Python library for **data analysis** .
- Pandas is built on top of **two core** Python libraries
  - **matplotlib** for data visualization
  - **NumPy** for mathematical operations.
  - Pandas acts as a wrapper over these libraries, allowing you to access many of matplotlib's and NumPy's methods with less code.

Modules

Numpy :

- **NumPy** is a Python library used for **working with arrays.**
- In Python we have lists that serve the purpose of arrays, but they are **slow** to process.
- NumPy aims to provide an array object that is up to **50x faster** than traditional Python lists.
- The array object in NumPy is called **ndarray** , it provides a lot of supporting functions that make working with **ndarray very easy.**
- Arrays are very frequently used in data science, where **speed and resources** are very important.

Modules

Regular Expressions :

- A **Regular Expressions (RegEx)** is a special sequence of characters that uses a **search pattern** to find a string or set of strings
- It can detect the **presence or absence** of a text by matching with a particular pattern, and also can split a pattern into one or more sub-patterns.
- Python provides a **re** module that supports the use of **regex** in Python.
- Its primary function is to **offer a search**, where it takes a regular expression and a string.

Modules

Time :

- The Python time module provides many **ways of representing time in code**, such as objects, numbers, and strings
- It also provides functionality other than representing time
  - like **waiting during code execution** and measuring the efficiency of your code.

Modules

Matplotlib :

- Matplotlib is an amazing **visualization library** in Python for 2D plots of arrays
- Matplotlib is a multi-platform data visualization library built **on NumPy arrays**
- One of the greatest benefits of visualization is that it allows us **visual access to huge amounts of data** in easily digestible visuals.
- Matplotlib consists of several plots like **line, bar, scatter, histogram etc.**

Modules

Pywebio :

- **PyWebIO** provides a diverse set of imperative functions to obtain user **input and output** content on the **browser**, turning the browser into a **“rich text terminal”**, and can be used **to build simple web applications or browser-based GUI applications**.

Modules

Pygame :

- The pygame library is an open-source module for the Python programming language specifically intended to help you make **games and other multimedia applications.**
- Built on top of the **highly portable SDL** (Simple DirectMedia Layer) development library, pygame can run across many platforms and operating systems.

Modules

Plotly :

- The plotly Python library is an **interactive, open-source plotting library** that supports **over 40** unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.

Hardware and Software Requirements

Hardware:

- Desktop , Keyboard , Mouse , Screen , Hard Drive , GPU , CPU and RAM

Software:

- System requirements for Python Installation:
  1. Operating system: **Linux- Ubuntu 16.04 to 17.10**, or Windows 7 to 10, with 2GB RAM (4GB preferable)
  2. You have to install Python 3.6 and related packages
  3. Jupyter Notebook , IDLE , Atom

Construction

- I’m taken **two dataset’s.** one set is to represent **all country’s data** and another one is to represent the data of **states of India .**
- After collecting the data set’s from the website, By using **numpy and pandas** I converted data in the way that we can able to see the resultant graph’s .
- By using **matplotlib,plotly and seaborn** libraries we represent the data into graph’s
- At the end by using **pywebio** we’re going to create a **front-end** for a better experience .

Measurement and Analysis

- From the graph's they can able to analyze the **new case's, deaths and conformed cases .**

- For all the countries and all the states of India .

Methodology

- **Methodology of Data Visualization on Covid19 Dataset :**
  - **Step 1:** Import the needed modules.
  - **Step 2:** Get your dataset and read it.
  - **Step 3:** using numpy and pandas clean the data
  - **Step 4:**Once the data is clean now we are good to go with the visualization part.
  - **Step 5:** Start visualizing the data by considering different columns.

Methodology

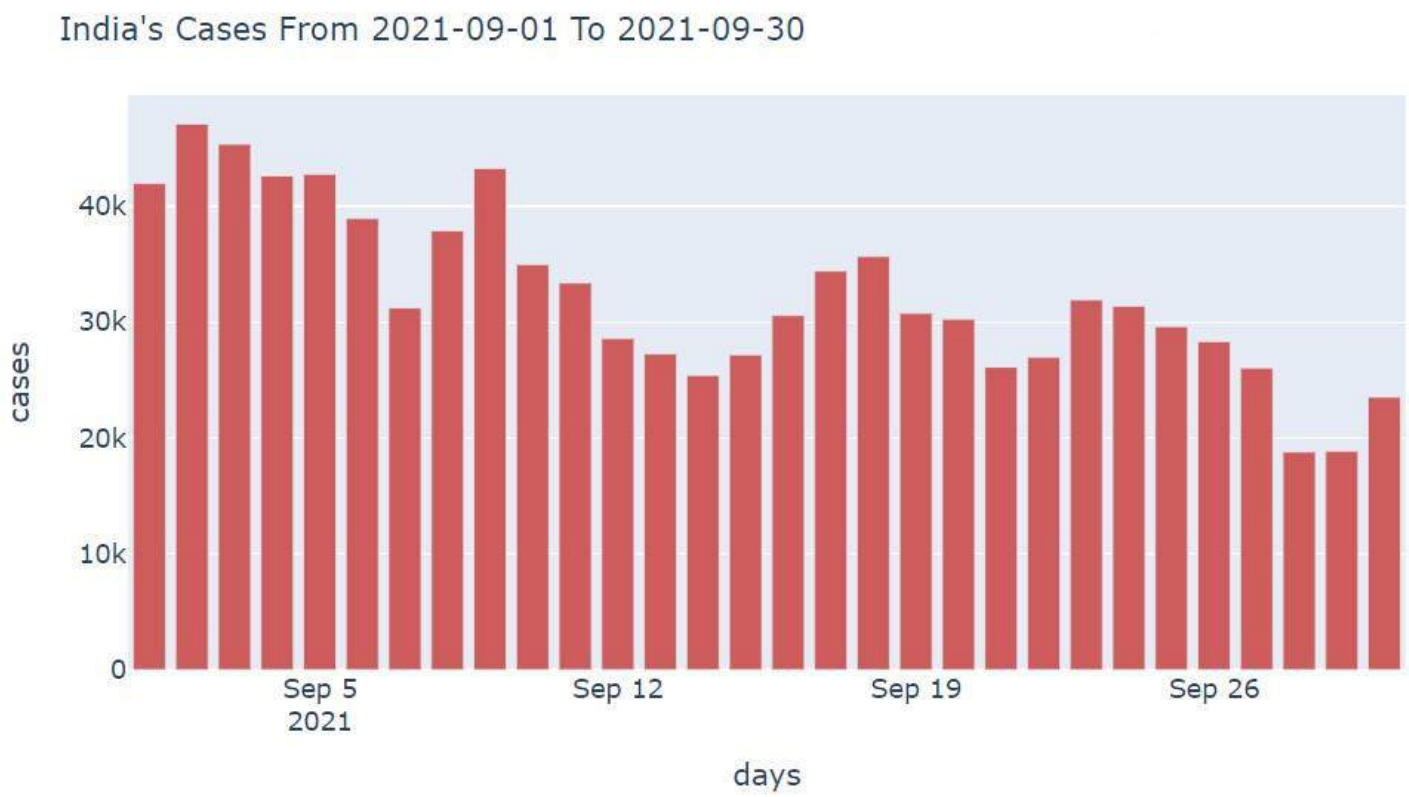
- **Step 6:** Plot the graphs with the help of data and modules imported
- **Step 7:** Generating the results in the pywebio application
- **Step 8:** Once visualizing the data is completed start analyzing the data.
- **Step 9:** predicting tomorrow's Covid-19 cases.

Results and Discussion

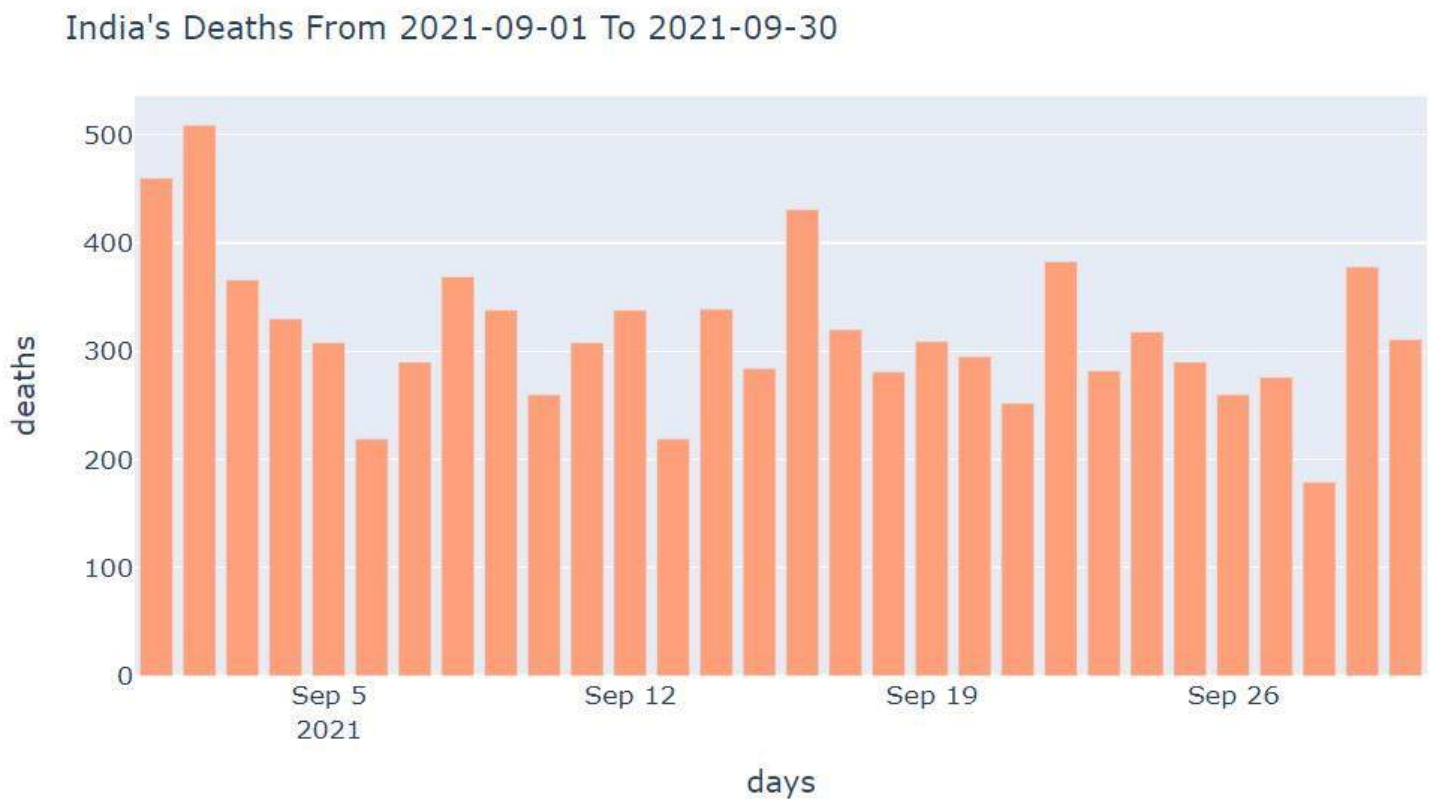
Bar Graph

day's VS new_cases's and death's

day's VS new_case's



day's VS new_death's



Line Graph

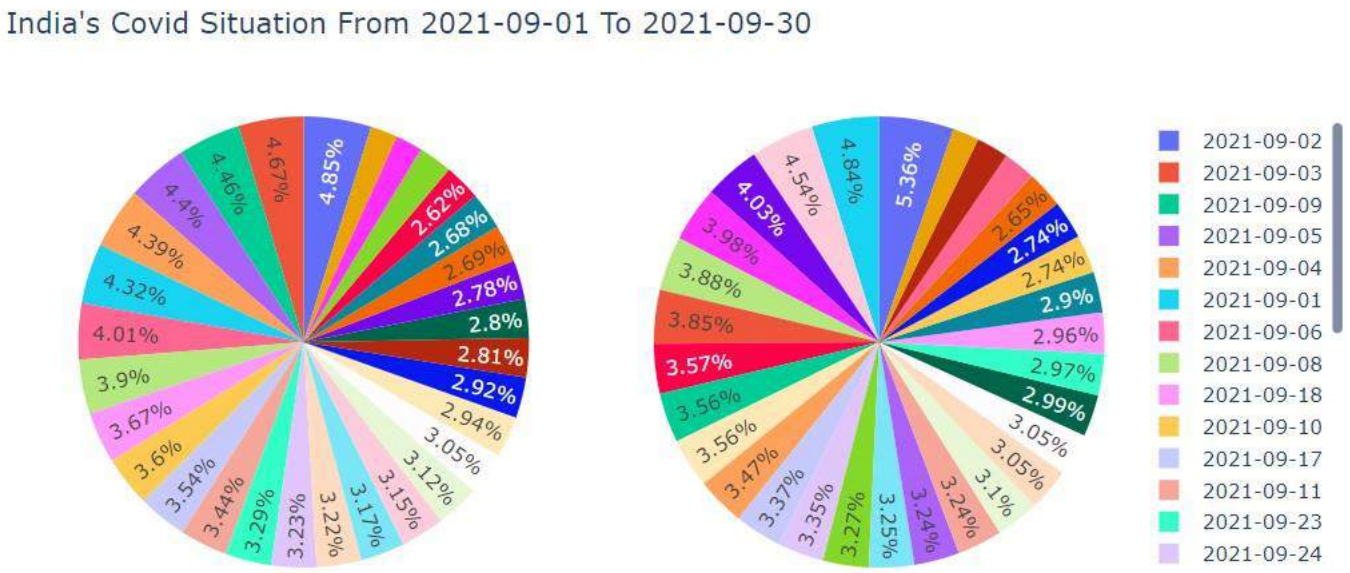
day's VS new_cases's and death's

day's VS new_case's

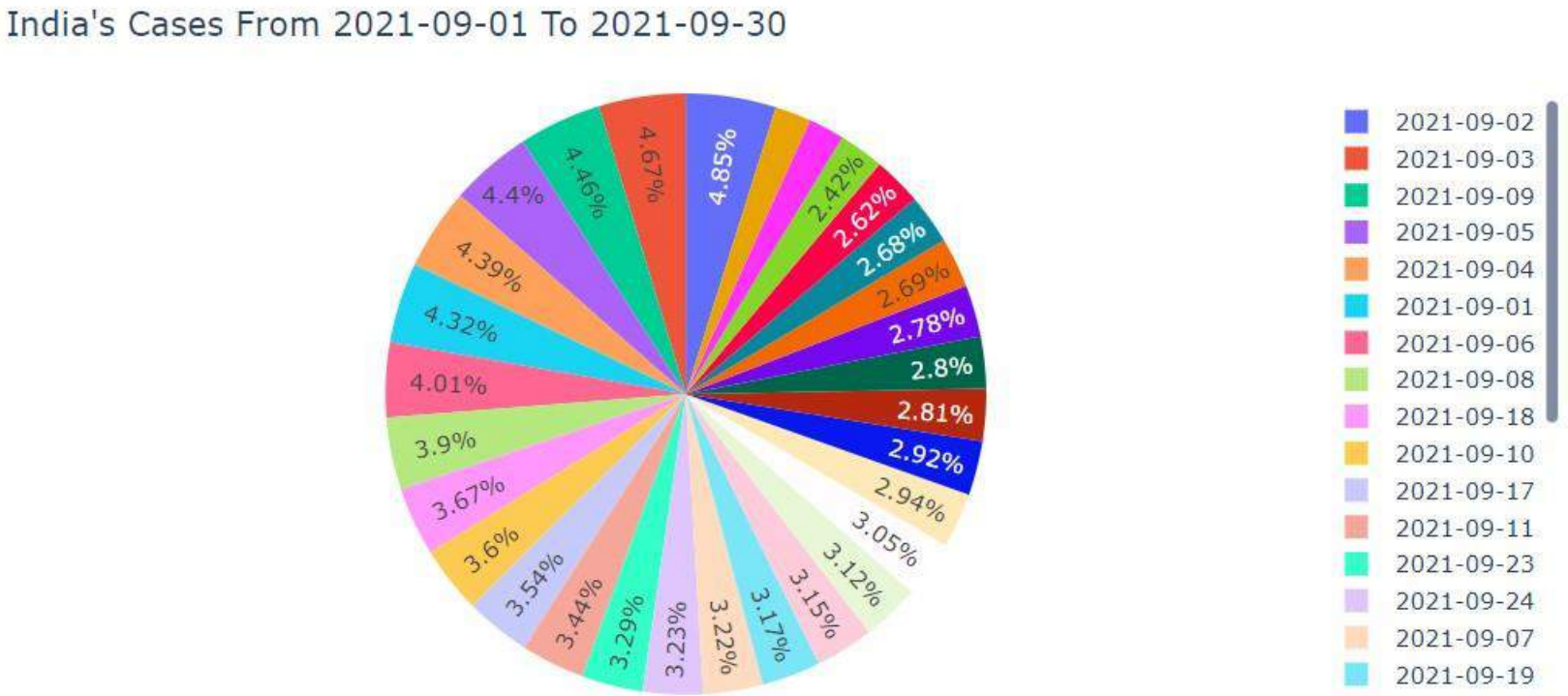
day's VS new_death's

Pie Graph

day's VS new_cases's and death's

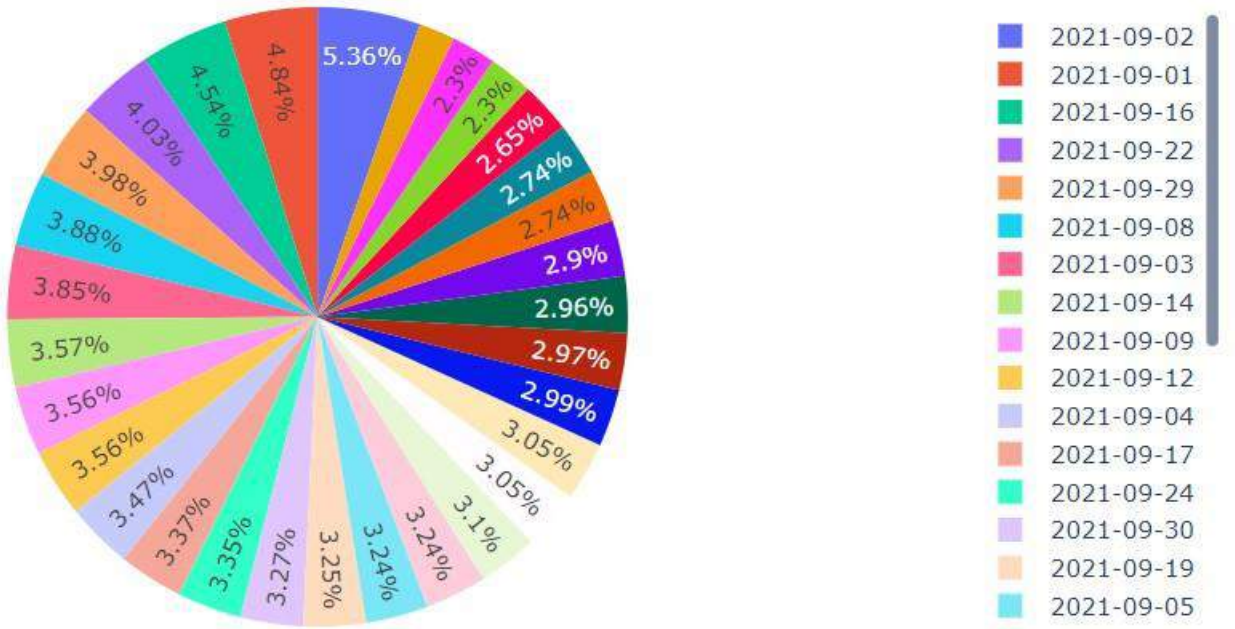


day's VS new_case's



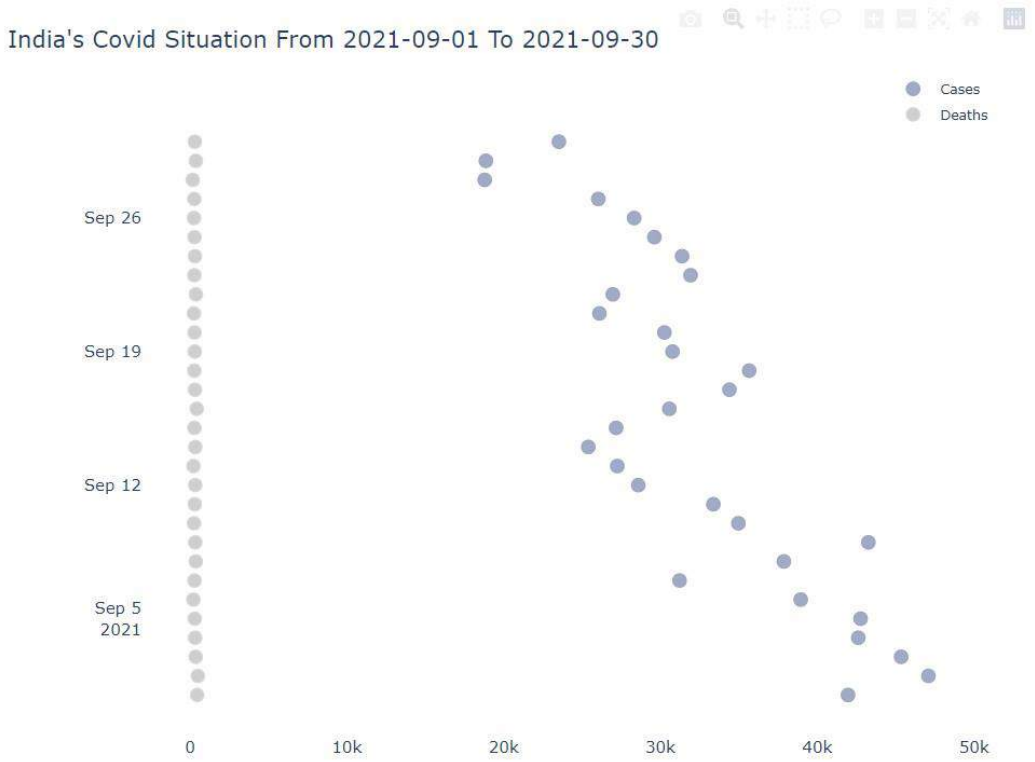
day's VS new_death's

India's Deaths From 2021-09-01 To 2021-09-30



Scatter Graph

day's VS new_cases's and death's

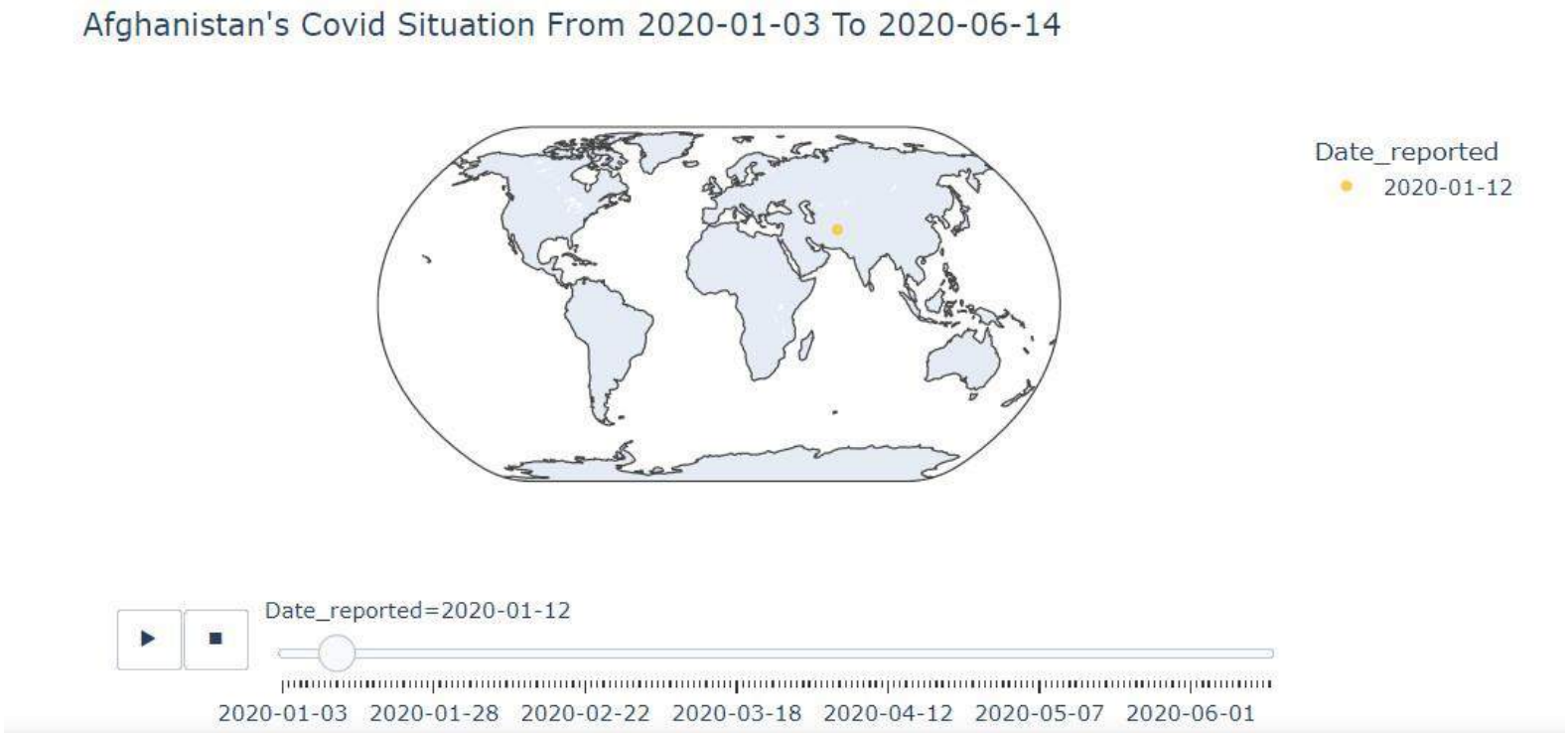


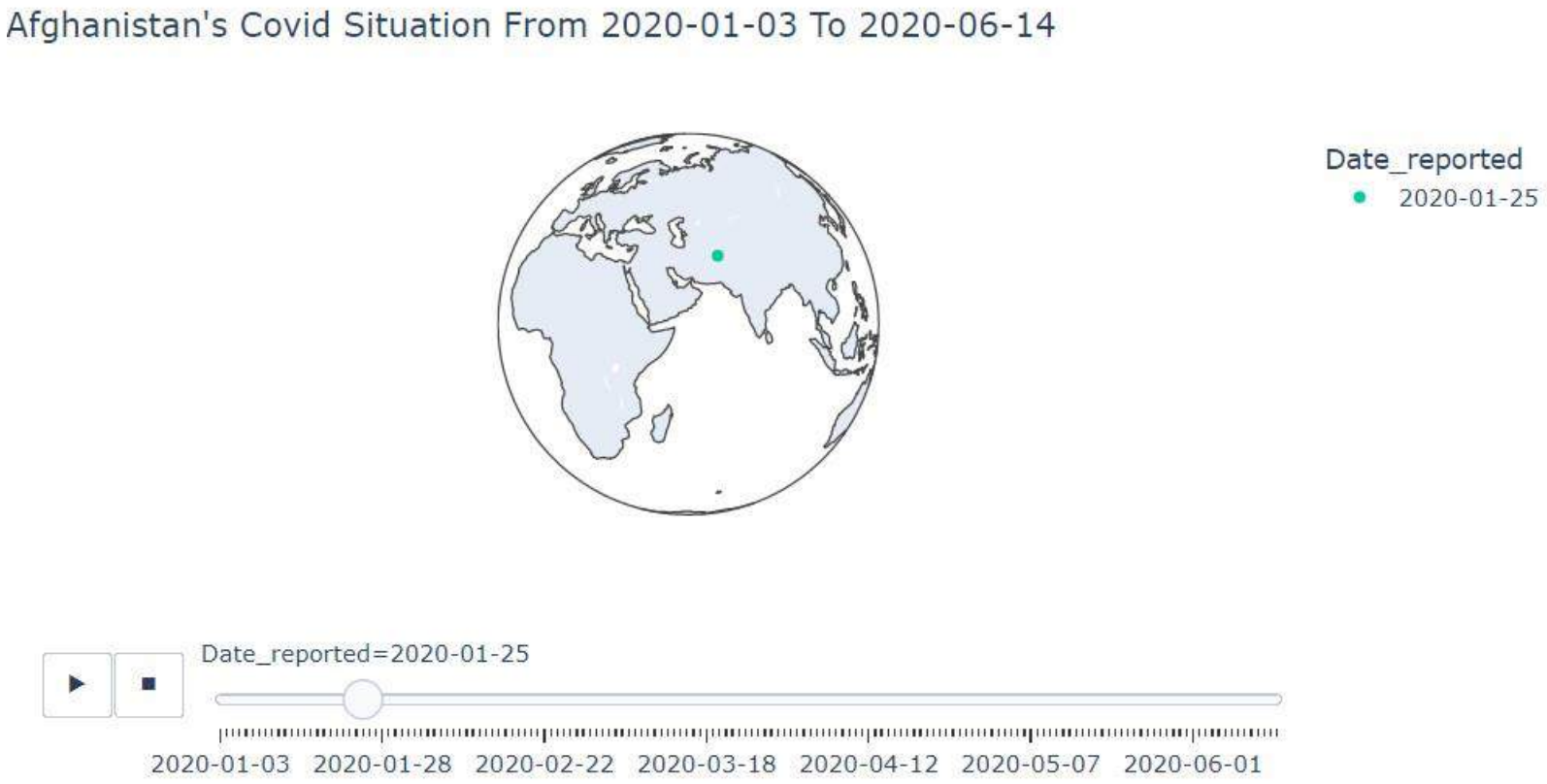
day's VS new_case's

day's VS new_death's

On Map

Please Wait While Loading The Data .....




Afghanistan's Covid Situation From 2020-01-03 To 2020-06-14
Date_reported
2020-01-25
Date_reported=2020-01-25
2020-01-03
2020-01-28
2020-02-22
2020-03-18
2020-04-12
2020-05-07
2020-06-01

Conclusion

- They can able to know the statistics on **Covid-19 pandemic** of all the country's and all the states of India.
- Result is going to be always in **graph's.** so, they can understand the **situation** very easily and able to take **precautions.**
- my graphical estimations require large amounts of data from active participants, but provide insightful depictions of the progress of the pandemic in different regions, offering an estimation of active number of cases and deaths in **different geographical areas.**

References

[1] **Dataset**

- Country's around the world - https://covid19.who.int
- All the states of India - https://www.kaggle.com/sudalairajkumar/covid19-in-india

[2] **Data processing**

- Numpy - https://www.w3schools.com/python/numpy_intro.asp
- Pandas - https://www.w3schools.com/python/pandas/default.asp

Data validation

- https://www.w3schools.com/python/python_regex.asp

[3] **Data visualization**

- Matplotlib - https://www.w3schools.com/python/matplotlib_pyplot.asp

[4] **Web application**

- Pywebio - https://www.pyweb.io/tutorial.html