Project Python Internship Report on

**“Corona Based Analysis on .csv File”**

Submitted to



**Visvesvaraya Technological University**

Jnana Sangama, Belagavi

In partial fulfillment of the requirements for the award of Certificate in

Python Internship in Computer Science and Engineering

By

**Name: Swapnadeep Kapuri**

**USN: 2BU20CS098**

**Name: Omkar Bhandari**

**USN: 2GI20CS080**

Under the guidance of

**Dr. Sunil F Rodd**

**Dr. Shrinivas Mangalwede**

Computer Science and Engg. Department

S.G.B.I.T. Belagavi



**NBA ACCREADTED, ISO CERTIFIED**

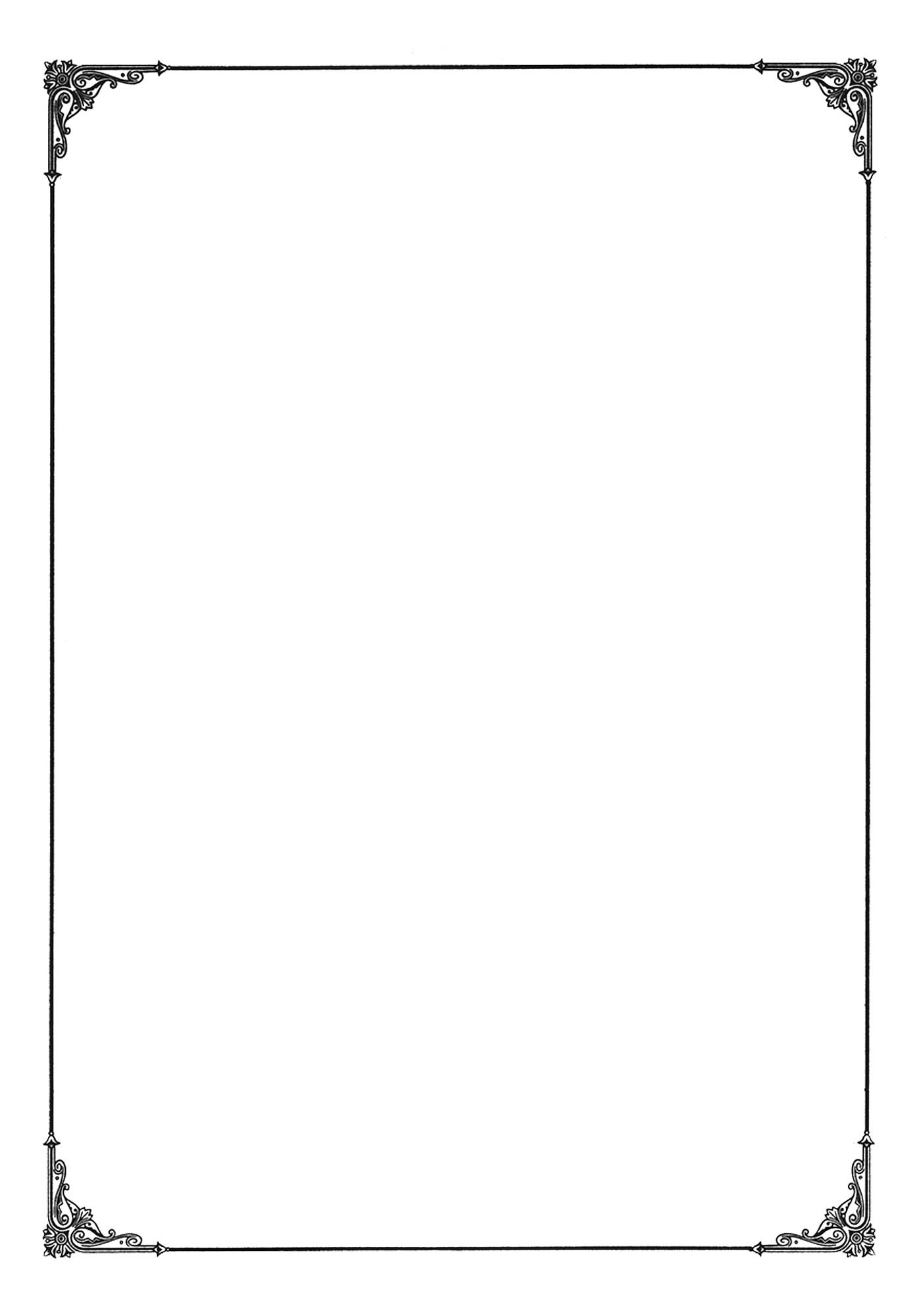
Department of Computer Science and Engineering

S.S EDUCATION TRUST

**S. G. BALEKUNDRI INSTITUTE OF TECHNOLOGY**

Shivabasavanagar, Belagavi-10, Karnataka.

**2022 - 2023**

****

**S. G. BALEKUNDRI INSTITUTE OF TECHNOLOGY**

Shivabasavanagar, Belagavi-10, Karnataka.



NBA Accredited, ISO Certified

Department of Computer Science and Engineering

**CERTIFICATE**

Certified that the Project Python Internship Report on entitled **“Corona Based Analysis on .csv File”** is a Bonafede work carried out by **Mr Swapandeep Kapuri (2BU20CS098) and Mr Omkar Bhandari (2GI20CS098)** in fulfillment for the award of Certificate in Python Machine Learning, from JThorn Technology Services Private Limited Belagavi, during the year 2022-2023. The Internship Project report has been approved as it satisfies the academic requirements in respect of Python Machine Learning Project work prescribed for the said Python Machine Learning Certification Program.

Guide

Dr. Sunil F Rodd HOD

Dr. Shrinivas Mangalwede Dr. B. S. Halakarnimath

Principal

Dr. B. R. Patagundi

**ACKNOWLEDGEMENT**

It is my privilege that I was able to work in this python machine learning project and I like to express my gratitude to these people, without them I may not be able to create this project. Their precious time and their guidance have led me to competition of this project without many obstacles.

First and the foremost, I wish to record our sincere gratitude to Management of this college and to our beloved Professor, **Dr. B.R. Patagundi**, Principal of S. G. Balekundri Institute of Technology, Belagavi for his constant support and encouragement in preparation of this report and for making available library and laboratory facilities needed to prepare this report.

My sincere thanks to **Dr. B.S. Halakarnimath**, Head, Department of Computer Science and Engineering in S.G.B.I.T. for the valuable suggestions and guidance through the period of preparation of this report.

I express my sincere gratitude to our beloved guide **Guide Dr. Sunil F Rodd**, **Professor** and **Dr. Shrinivas Mangalwede, Professor** in the Dept. of Computer Science and Engineering of S.G.B.I.T., Belagavi and JThorn Technology Private Limited for guiding me in investigations for this project. Our numerous discussions with him were extremely helpful. I told them that in their esteem guidance, encouragement and inspiration received from them helped me a lot.

The course on Python Machine Learning Internship Program was very helpful to us in giving the necessary background information and inspiration in choosing this topic for the Internship project. The class room lectures very much knowledgeable, informative and interesting.

I wish to thank all the source materials available from our college library and the digital library for providing me in the source materials, books for the required project and source materials.

Finally, I would like to thank my fellow classmates, friends and my parents for the constant support and encouragement for the completion of this project.

**By Swapnadeep Kapuri**

**(2BU20CS098)**

**Omkar Bhandari**

**(2GI20CS080)**

**Belagavi:**

**Date:**

**ABSTRACT**

This document is meant for describing all the features and procedures that were followed while developing the system. This document specially mentions the details of the project and how it was developed, the primary requirement, as well as various features and functionalities of the project and the procedures followed in achieving these objectives.

The COVID-19 pandemic in India has disrupted our country in an unprecedented way, and data analytics have played a crucial role in understanding and mitigating its impact. In this project, we explore COVID-19 data using Python and machine learning techniques. The data is in a .csv file, which we load into a pandas data frame for analysis. We perform exploratory data analysis to understand the trends and patterns in the data.

Overall, this project demonstrates how Python and machine learning techniques can be used to gain insights into the COVID-19 pandemic in India and make predictions about its future impact. The dashboard provides a user-friendly interface to explore the data and stay up to date with the latest COVID-19 statistics in India.

**Keywords:** Reading the .csv file contents related to corona pandemic in India, Analytics based on that using Python machine learning library called Pandas, getting results related to the problem statement, solving real life problems based on analytics.

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter** | **Contents** | **Page No** |
| **1** | **Introduction**   * **Introduction to the project** * **Problem Definition** * **Motivations/ Objectives** | **7** |
| **2** | **System Requirements**   * **Software Requirements**   + **Backend + Frontend** * **Hardware Requirements**   + **Functional/Non-functional** | **13** |
| **3** | **Implementations**   * **Checking the .csv file** * **Analysis of the contents** * **Code of the project** | **14** |
| **4** | **Results of the Project**   * **Screenshots** | **29** |
| **5** | **Conclusions**   * **Project Conclusion** * **References** | **33** |

Figures Used

|  |  |
| --- | --- |
| Figure 1.1 | Main Menu of our program |
| Figure 1.2 | First Six rows of the column are: |
| Figure 1.3 | Use Case Scenario |
| Figure 1.4 | Columns contained in the data set |
| Figure 1.5 | Does the dataframe contain any null value |
| Figure 1.6 | Last six columns in the row |
| Figure 1.7 | Check the values for the columns /"Confirmed Indian National" and /"Confirmed Foreign National" |
| Figure 1.8 | Dropping the columns Sno and Time |
| Figure 1.9 | Maximum cases happening in each state is |
| Figure 1.10 | Total Cases Happening in each states are as follows |
| Figure 1.11 | The percentage of fatal, active, and cured covid pandemic figures: |
| Figure 1.12 | States with most affected cases are: |
| Figure 1.13 | Total Deaths till August 11th 2021 |

Chapter 1 Introduction

1.1: Introduction to the Project

1. **Data Collection:** The first step in any data analysis project is data collection. For this project, you can download the COVID-19 dataset from a reliable source, such as the Kaggle website or the official Indian government website. This dataset contains daily COVID-19 cases and deaths reported in India since the beginning of the pandemic.
2. **Data Pre-processing:** Once you have the data, the next step is to clean and pre-process it. You can use Python Pandas library to read the .csv file and convert it into a dataframe. Then, you can perform some basic data cleaning operations such as removing duplicates, handling missing values, and changing the data types of some columns.
3. **Exploratory Data Analysis:** After pre-processing the data, you can start exploring it to gain insights and identify patterns. You can use Python's matplotlib or seaborn library to create different types of visualizations such as line plots, bar plots, and heatmaps to show how COVID-19 cases and deaths have changed over time in India.
4. **Feature Engineering:** Feature engineering is the process of creating new features from existing data that can help improve the performance of machine learning models. In this project, you can create new features such as daily cases, daily deaths, and rolling averages to capture the trends and patterns in the data.

Through this project we can automate the data processing required in minutes of time using python’s pandas framework.

**1.2: Problem Definition**

1. Get the first 6 rows.

2. Check the columns contained in the dataset.

3. Check the shape of the dataframe.

4. Check if the dataframe has any null values.

5. List the last 6 rows.

6. check the values for the columns ‘’ConfirmedIndianNational’ and ‘ConfirmedForeignNational’.

7. Drop the columns 'Sno' and 'Time'.

8. Find out the maximum cases until 11th Aug 2021 for each state.

9. Find out the total confirmed cases till 11th Aug 2021.

10. Find out the percentage for Active, fatal, and cured cases.

11. Check the data for the 10 most affected states with covid-19 in India.

12. Find out the total death till Aug 2021.

The problem addressed in this project is to analyse the COVID-19 pandemic situation in India by exploring the daily reported cases and deaths data. The goal is to gain insights into the trends and patterns in the data and predict the future trends using machine learning algorithms. The project aims to help understand the severity of the pandemic in India and provide valuable information for decision-making and resource allocation.

**1.3: Motivations/ Objectives**

**Motivations:**

The motivation for this project is to help understand the severity of the COVID-19 pandemic in India by analysing the daily reported cases and deaths data. The COVID-19 pandemic has had a significant impact on India, with many cases and deaths reported daily. The motivation behind this project is to help understand the severity of the pandemic in India by analysing the daily reported cases and deaths data.

This project aims to provide valuable information for decision-making and resource allocation by government agencies, healthcare organizations, and individuals. The project can also help in identifying potential areas of improvement in the COVID-19 response in India and contribute to the global fight against the pandemic.

**Objectives:**

The specific objectives of this project are:

1. To pre-process the COVID-19 dataset and clean the data to ensure it is accurate and usable.
2. To explore the data and gain insights into the trends and patterns in the COVID-19 cases and deaths in India.
3. To create new features using feature engineering to capture the trends and patterns in the data.
4. To provide insights and information to decision-makers and the public about the severity of the COVID-19 pandemic in India and its future trends.

Chapter 2 System Requirements

2.1: Software Requirements

* Backend

1. You will need .csv configured in the same folder/directory as the python project.
2. You will need Python 3.10.9 and pandas 1.5.2 or upwards as a data framework, to work as a mediator between backend and frontend.
3. Excel program or text program to access the .csv file directly.

* Frontend

1. You will need latest version of graphical based processing unit or GPU since it utilizes multiple data related models in it.
2. Spyder IDE for python code visualization and implementations.

2.2: Hardware Requirements

* Functional Requirements

1. It should be running on Windows 10/11 as an operating system.
2. The processor must be of intel core i3 10th generation and upwards.
3. The processor must be of architecture x86-64, with a 64-bit architecture with dual core processor.
4. It should have free space of up to 100 Mb for the program to run properly.

* Non-Functional Requirements

1. It should have good hardware compatibility with the framework as well as the software’s which are going to be used.
2. The processor must be streamlined to perform high level of data manipulation.
3. Robust system software if any unknown damage occurs to the computational device.

Chapter 3 Implementations

**3.1: Checking the .csv file**

The .csv file name is CaseStudy-11.csv which must be in the same directory as the python file, since it does not uses OS python module through which it can directed to that particular file location by the user.

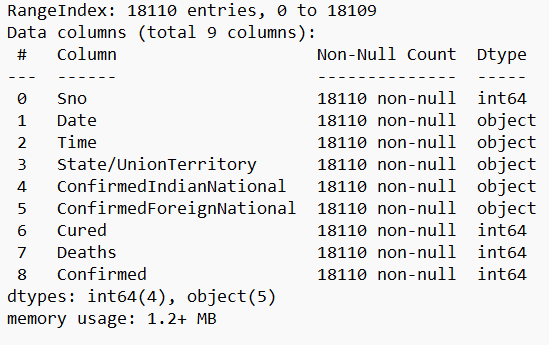
As soon as the python project detects the file, the operations related to it can be started.

**3.2: Analysis of the content**

It has 18,110 rows from starting date 1/30/2020 till 8/11/2021 i.e from 30th of January 2020 till 11th of August 2021.

It has 9 columns namely Sno, Date, Time, State/Union Territory, Confirmed Indian National, Confirmed Foreign National, Cured, Deaths and Confirmed.

These are its analytical information:



**3.3: Code of the project**

import numpy as np

import pandas as pd

from matplotlib import pyplot as plt

stdata=pd.read\_csv("CaseStudy-11.csv")

t=True

while(t==True):

n=int(input(""" Welcome to Python internship project done by Swapnadeep Kapuri and Omkar Bhandari

Menu (select any one of it)

1.Get the 1st six rows

2.Check the columns contained in the data set

3.Check the shape of the dataframe

4.Check if the data frame has any null values

5.List the last six rows

6.Check the values for the columns /"Confirmed Indian National" and /"Confirmed Foreign National"

7.Drop the columns /'Sno' and /'Time'

8.Find the maximum cases until 11th August 2021 for each state

9.Find out total confirmed cases till 11th August 2021

10.Find the percentage of the active, fatal and cured cases

11.Check out the data for the 10 most affected states with covid-19 in India

12.Find out total deaths till August 2021

13.EXIT

DEFAULT:Error

Enter your option:"""))

if(n==1):

print("The 1st six rows are:")

print(stdata.head(6))

if(n==2):

print("The columns contained in the data set are:")

print(stdata.columns)

if(n==3):

print("Shape of the dataframe is (i.e rows and columns):")

print(stdata.shape)

if(n==4):

print("Does the Dataframe contains null values?")

print(stdata.isnull().sum())

if(n==5):

print("The last six rows are:")

print(stdata.tail(6))

if(n==6):

print("The values of the columns of Confirmed Indian and Foreign Nationals are:")

print(stdata[["ConfirmedIndianNational","ConfirmedForeignNational"]])

if(n==7):

print("As dropping columns might disturb .csv files instead columns except of Sno and Time will be displayed")

print(stdata[["Date","State/UnionTerritory","ConfirmedIndianNational","ConfirmedForeignNational","Cured","Deaths","Confirmed"]])

print("Note: By using stdata.drop(['Sno','Time']) method we can drop the required columns if required")

if(n==8):

print("The maximum cases until 11th August 2021 for each state is:")

result=stdata[["Date","State/UnionTerritory",'Confirmed']].loc['1/30/2020':'8/11/2021'].groupby(["State/UnionTerritory"]).max()["Confirmed"]

print(result)

if(n==9):

print("Total Confirmed cases until 11th August 2021 is:")

result2=stdata[["Date","State/UnionTerritory",'Confirmed']].loc['1/30/2020':'8/11/2021'].groupby(["State/UnionTerritory"]).sum()["Confirmed"]

print(result2)

if(n==10):

print("The percentage of active cases are:")

result3=((stdata.sum()["Confirmed"]-stdata.sum()["Cured"])/stdata.sum()["Confirmed"])\*100

print(result3)

print("The percentage of fatal cases i.e deaths are:")

result9=(stdata.sum()["Deaths"]/stdata.sum()["Confirmed"])\*100

print(result9)

print("The percentage of cured cases are:")

result10=(stdata.sum()["Cured"]/stdata.sum()["Confirmed"])\*100

print(result10)

print("The percentage in the pie chart is as following:")

x=np.array([result3,result9,result10])

labels = ['Active', 'Fatal', 'Cured']

colors = ['red', 'blue', 'green']

plt.pie(x,labels=labels, colors=colors)

plt.title("The percentage of covid cases")

plt.show()

if(n==11):

print("The data of 10 most affected states in India are:")

result4=stdata[["Date","State/UnionTerritory",'Confirmed']].loc['1/30/2020':'8/11/2021'].groupby(["State/UnionTerritory"]).sum()["Confirmed"].sort\_values().tail(10)

print(result4)

if(n==12):

print("Total Deaths till 11th August 2021 is:")

result5=stdata.loc['1/30/2020':'8/11/2021'].sum()["Deaths"]

print(result5)

if(n==13):

t=False

if(type(n)!=int or n>13 or n<1):

print("Wrong Option typed!! Please type a correct option!!")

Chapter 4 Results of the Project

4.1: Screenshots and their functionality

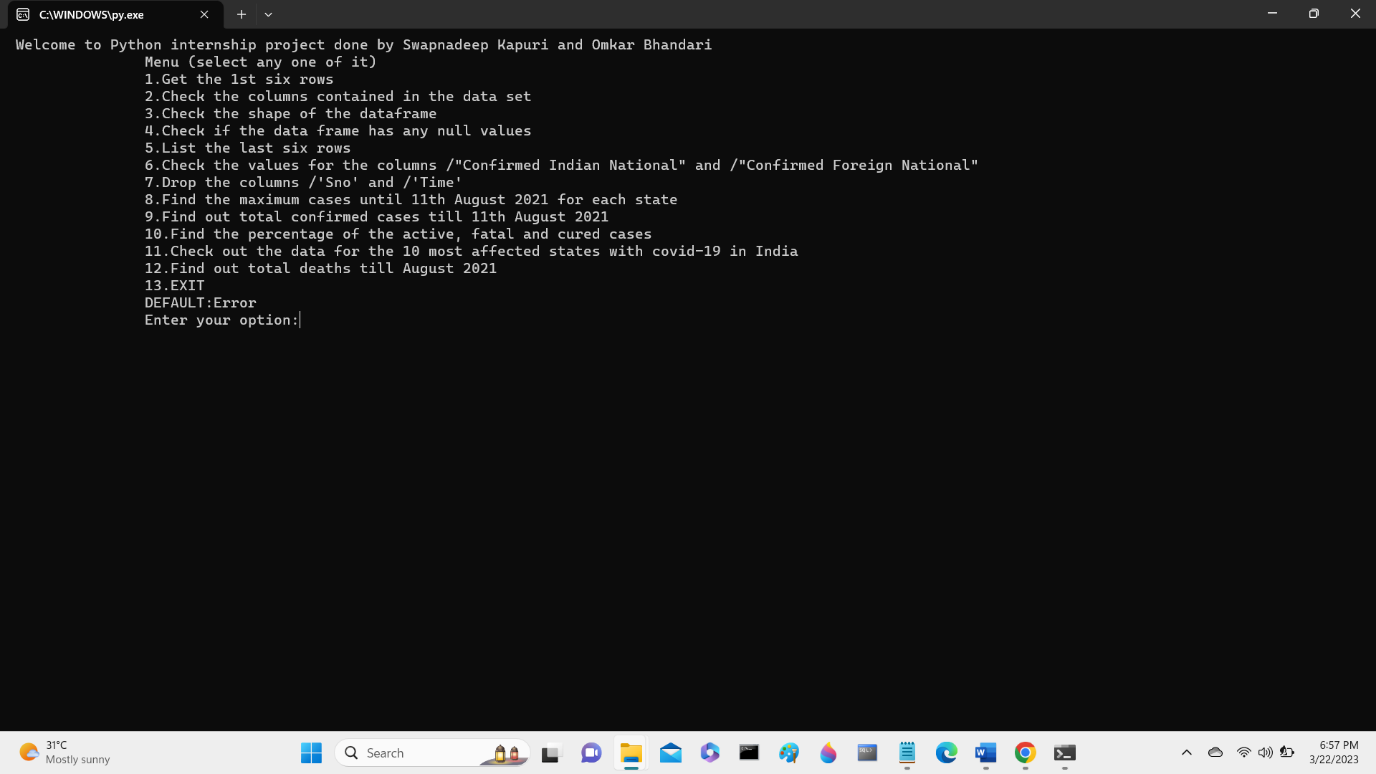
Figure1.1: Main Menu of our program.

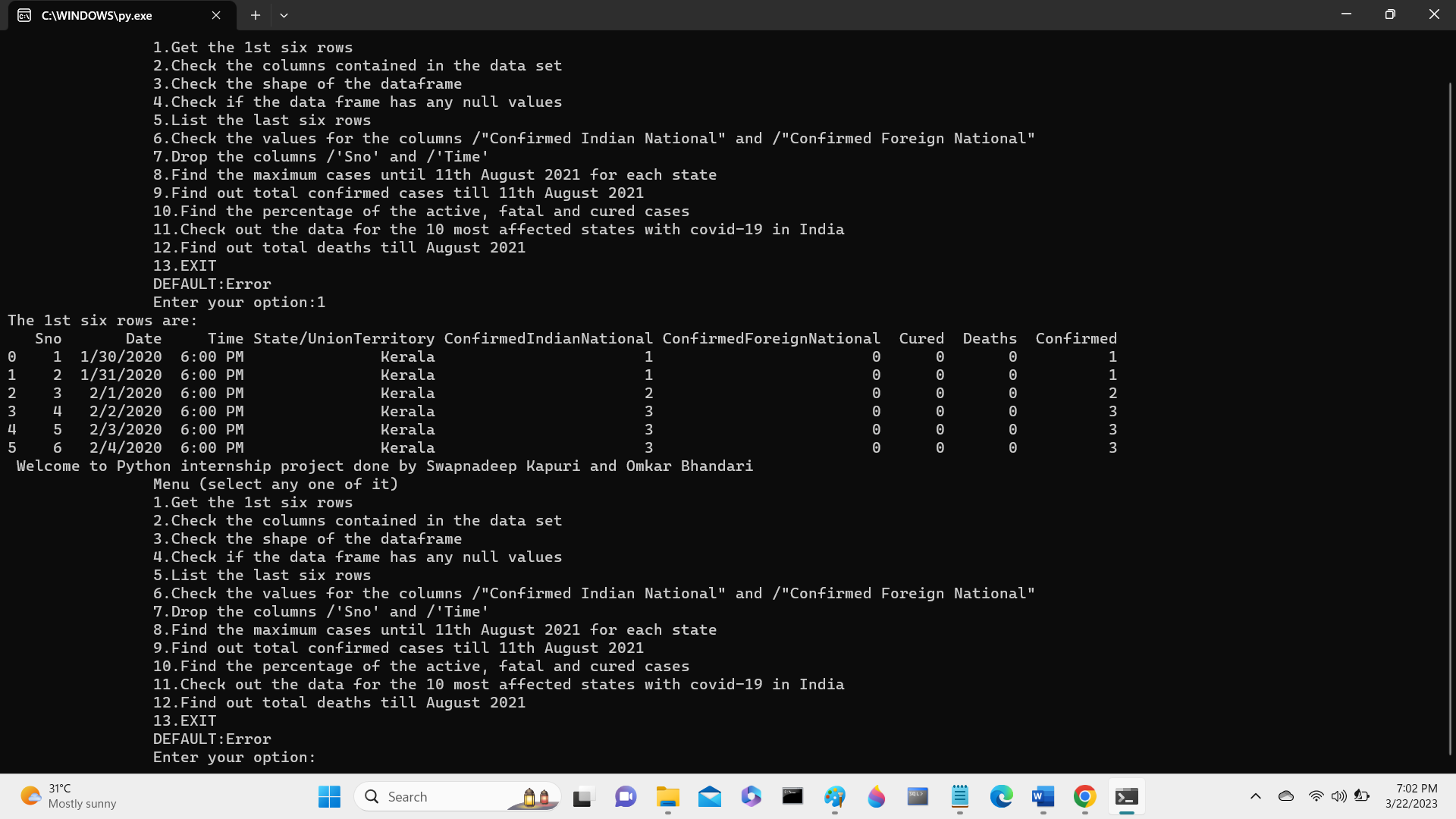
Figure 1.2: First Six rows of the column are: 

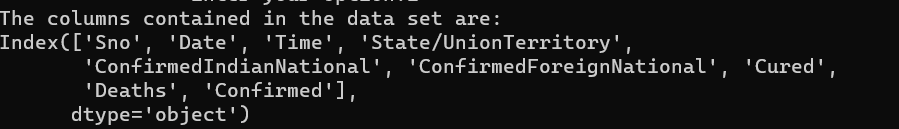
Figure 1.3: Columns contained in the data set

Figure 1.4: Shape of the data frame

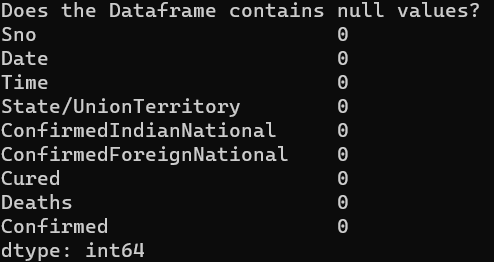
Figure 1.5: Does the dataframe contain any null value

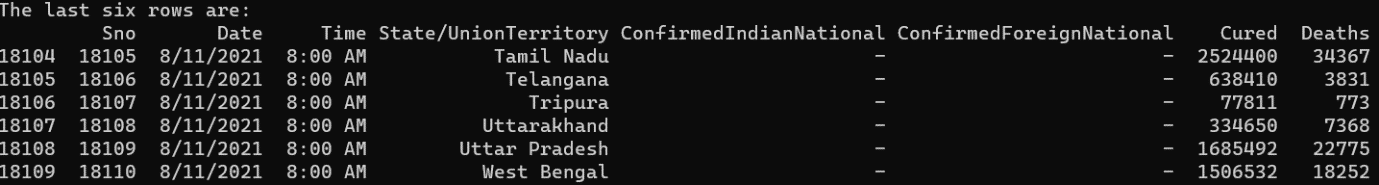
Figure 1.6: Last six columns in the row

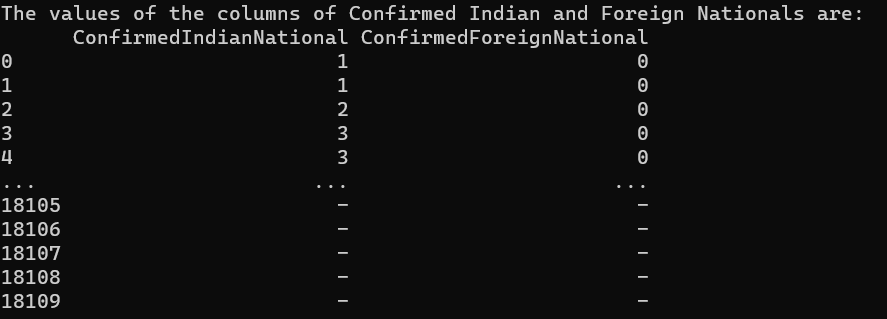
Figure 1.7: Check the values for the columns /"Confirmed Indian National" and /"Confirmed Foreign National"

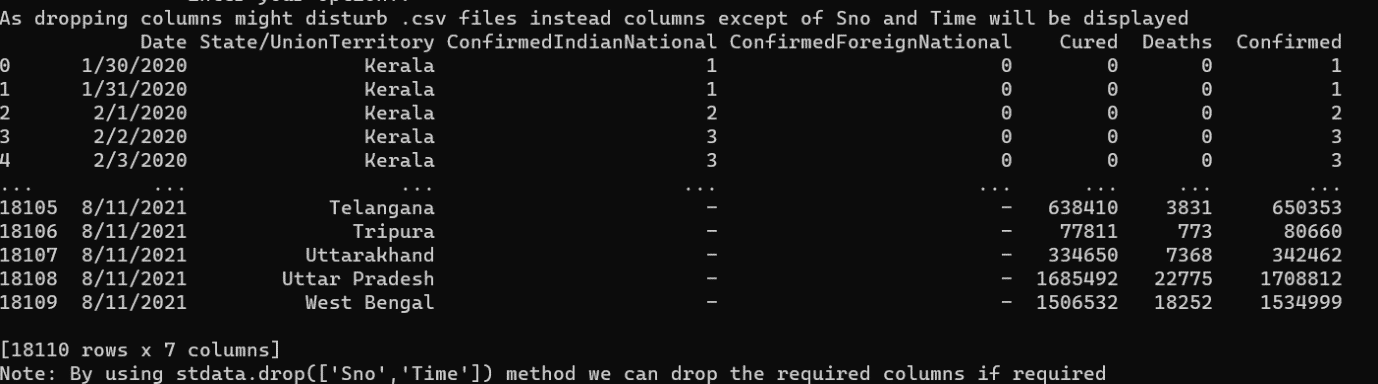
Figure 1.8: Dropping the columns Sno and Time

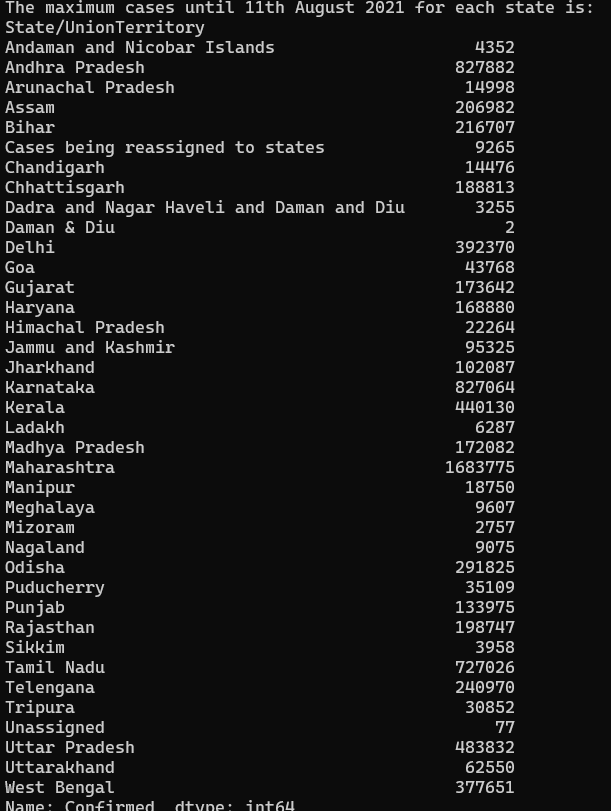
Figure 1.9:Maximum cases happening in each state is

Figure 1.10: Total Cases Happening in each states are as follows

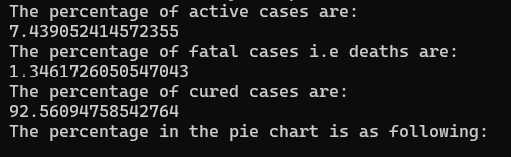
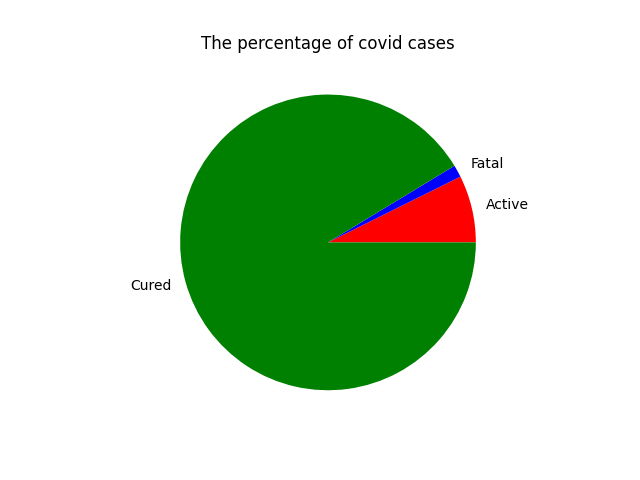
Figure 1.11: The percentage of fatal, active and cured covid pandemic figures: 

Figure 1.12: States with most affected cases are:

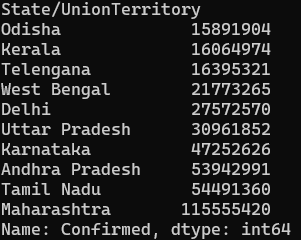


Figure 1.13: 

Chapter 5 Conclusion

5.1: Project Conclusion

* India has been significantly impacted by the COVID-19 pandemic, with a large number of cases and deaths reported across the country.
* The pandemic has had a disproportionate impact on certain states and regions within India, with some areas experiencing much higher infection and mortality rates than others.
* The implementation of various government policies, such as lockdowns and vaccination drives, has had a significant impact on the trajectory of the pandemic in India.
* The pandemic has highlighted significant inequalities and weaknesses in India's healthcare system, and there is a need for investment and reform to strengthen the system and better prepare for future pandemics.
* The use of data analysis and visualization tools such as Pandas, Matplotlib, and NumPy can provide valuable insights into the spread and impact of the pandemic, as well as inform policy decisions and interventions.

6.2: References

1. Under the Guidance of Professor Srinivas Mangalwede sir, the project has been completed under Python Machine Learning Internship Programe.
2. **For multiple issues:** [**https://stackoverflow.com/**](https://stackoverflow.com/)