

MINI PROJECT REPORT ON COVID-19 DATA ANALYSIS

UNDER THE GUIDENCE OF

Mr. Amir Khan

Technical Trainer

T&D dept.

GLA University, Mathura.

Under taken by

NIMISHA PACHAURI	181500430
PARAKH TIWARI	181500450

CONTENT

Abstract

1. Objective

2. Introduction

3. Problem

4. Implementation

5. Progress

6. Work Remaining

7. Screenshots

8. References

Abstract

This project is entitled to “Covid-19 Data Analysis”.

This project is an outcome of summer training done on “Introduction of Data Analysis for Business”. The training was based on how to play with data to extract useful information. A large amount of data has been short summarised and thoroughly analysed using data analysis.

The main objective of the project is to analyse the COVID-19 data of India and the world. In the process, using Jupyter Notebook, the data was accessed which was stored in a CSV file and then different queries were generated for analysing the data in the form of graphs, charts and different other diagrammatic format which made it easy to understand the spread of the disease.

The Covid data has been recorded in a CSV file which was quite difficult to be read and analyse for which this project has been done. Various tools have been used for analysing the data such as python pandas and Jupyter Notebook, after which a conclusion has been drawn about the data set.

With the help of this project, a huge amount of data has been easily analysed for making a brief report on the spread and contain of covid and the process of this has been explained in this report and the objective has been fulfilled.

Objective

To analyse the data of covid19 contain and spread across the country and the world among the population using various diagrammatic formats to understand the spread of the pandemic covid-19 and make a brief report on the same in order help the authorities to do the possible to rein the flow of the disease and save lives. This project is nothing less than a contribution for our government and concerned authorities to understand the disease containment and bring back the normal routine life as it was before corona.

Introduction

The project aims at deeply analysing the contain and spread of the world pandemic Corona virus, also known as Covid19 in Indian subcontinent and other parts of the world, which has made a foot fall in December 2019 from the city of Wuhan, China. The project uses various technologies for analysing the spread of the virus through charts, graphs, tables, maps and various representations methods.

Today, as we all know how much the virus is spreading and making its roots across the country and the world, the data that we will represent in this project, will give us a thorough study of the numbers upto which the virus is spreading among the locals, how steeply the number of patients are rising, the number of recoveries, active patients and the number of deceased. It also gives a thorough study of the data of specified places where the cases are more, places where the cases are less, states which are able to control the disease, the population effected, groups of ages of the people who are effected, who among males and females are effected, who are at higher risk of infection, who has lower risk of infection and many other deep studies related to the pandemic.

The project also gives a time series analysis of the pandemic which helps us to predict the future projection of the disease, which will help us to plan the future actions that should be taken for controlling and stopping its contain.

This could help us to understand the extent of corona virus spread in the country and the world and, in turn, it could help us to study what effects

the measures have done, taken in the past, and now what shall be done in future to control the spread.

The project will reduce the searching and analysing of covid19 data for the users and analysts who has to search at different places for such data, maintain the record of the data and has to generate reports of the contain and spread of the data. The project will also help the government, doctors, police, media, journalists and other concerned authorities who require the data for analysing purpose and report making and to place the future actions that should be taken for controlling the same, knowing the past mistakes.

Problem

The problem is a large set of given data of patient details of covid-19 effected people of whom some are confirmed cases, some are active, some recovered and some were unfortunate deaths. To analyse the amount of people effected and those who are non-effected was difficult to understand. Those who are active and those who recovered from the disease were difficult to compare. Also to analyse the age groups effected the most, the people who succumbed to the disease was difficult to be known. To make the task easy, this project has been implemented.

With the implementation of this project, the problem of analysation has been resolved as the projection of every available data or information has been done separately in form of easily understandable formats to reduce the confusion generated by the data.

Implementation

The project uses python as a working methodology. Various libraries of python is been used in this project to make this project working more effective, efficient and apprehensive. The project works on Jupyter Notebook along with other technologies where each tech used in this project has a different purpose. Some techs give us data analysed in graphical form, whereas some techs represent data tabular form and all techs used in this project gives a better cumulative result and the purpose of the project is been fulfilled.

The main platform on which the project works if Jupyter Notebook. The other techs include are matplotlib, plotly, plotly.express , plotly.graph_objects , cufflinks, plotly.offline , folium.

What is Pandas?

It is a high-level data manipulation tool developed by Wes McKinney. It is built on the Numpy package and its key data structure is called the DataFrame. DataFrames allow you to store and manipulate tabular data in rows of observations and columns of variables. There are several ways to create a DataFrame.

What is NumPy?

NumPy is a python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

What is DataFrame?

Pandas DataFrame is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the data, rows, and columns.

Progress

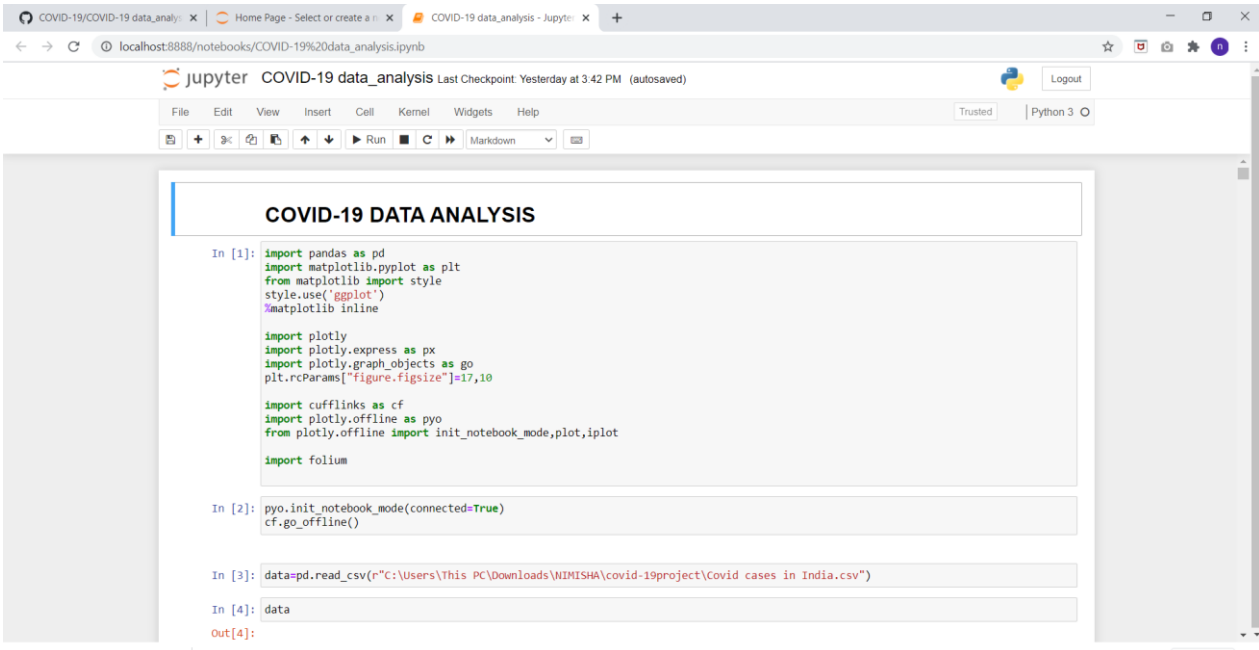
The project is going as follows:

- All python libraries have been imported in the jupyter notebook.
- The table containing States/UT, confirmed cases, cured and deaths has been imported.
- Number of active cases have been calculated using the formula 'Total-(Deaths + Confirmed)' and the table has been displayed.
- Coloured tables has been displayed.
- The data has been sorted.
- 'Total Confirmed Cases' has been plotted on bar graph according to states.
- The dot graph has been plotted.
- Latitude and Longitude coordinates of Indian States and UTs has been plotted so that the Indian map can be displayed.
- Indian map has been displayed which includes the number of cases found in India. The map show red area where the number of cases are larger.
- Table of number of cases found in Italy, Korea, Wuhan and India has been displayed. It also show the dates and time it took to surpass 100 cases in their countries.
- Various graphs have been displayed show number of cases found in several countries.

Work Remaining

- Time Series Analysis is remaining.
- How data can increase or decrease in India and various countries in coming future left for being analysed.
- When a country can observe its peak of cases and when will the downfall of cases can be seen is left for being analysed.
- The overall report of data which includes when the pandemic has started, when it was on its peak, how much population got affected, number of people got cured and people who succumbed to the disease is left for being analysed.

Screenshots



COVID-19 DATA ANALYSIS

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import style
style.use('ggplot')
%matplotlib inline

import plotly
import plotly.express as px
import plotly.graph_objects as go
plt.rcParams["figure.figsize"]=17,10

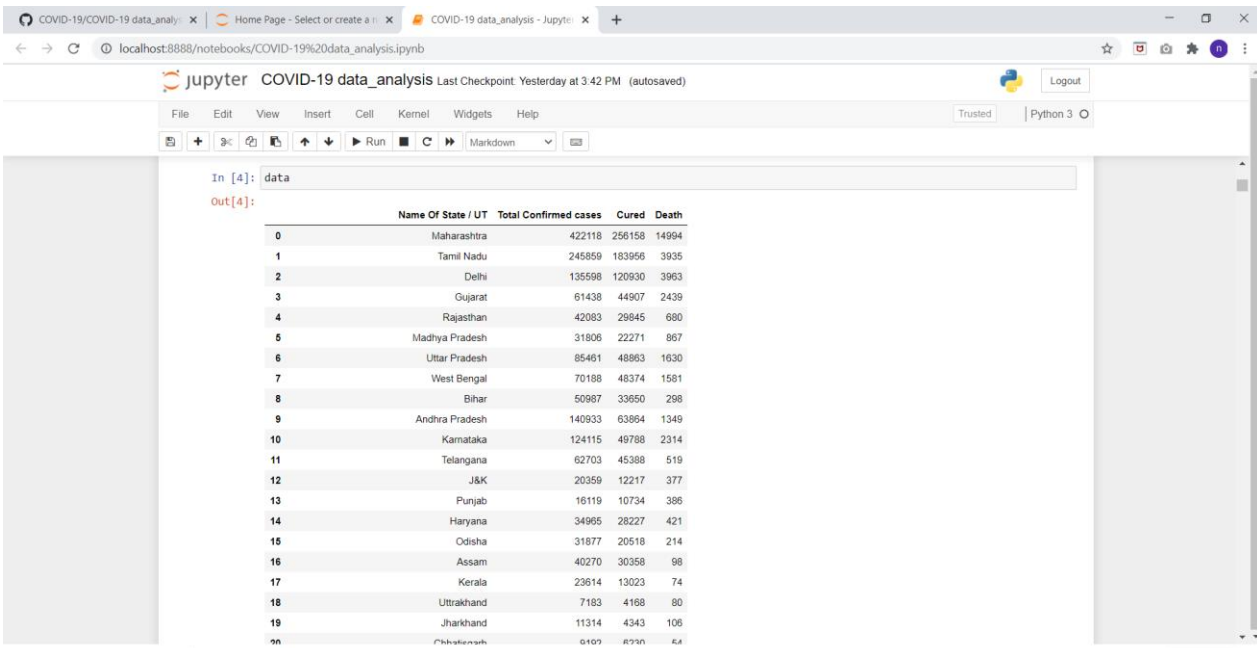
import cufflinks as cf
import plotly.offline as pyo
from plotly.offline import init_notebook_mode, plot, iplot

import folium

In [2]: pyo.init_notebook_mode(connected=True)
cf.go_offline()

In [3]: data=pd.read_csv(r"C:\Users\This PC\Downloads\WIMISHA\covid-19project\Covid cases in India.csv")

In [4]: data
Out[4]:
```



COVID-19 DATA ANALYSIS

```
In [4]: data
Out[4]:
```

	Name Of State / UT	Total Confirmed cases	Cured	Death
0	Maharashtra	422118	256158	14994
1	Tamil Nadu	245859	183956	3935
2	Delhi	135598	120930	3963
3	Gujarat	61438	44907	2439
4	Rajasthan	42083	29845	680
5	Madhya Pradesh	31806	22271	867
6	Uttar Pradesh	85461	48863	1630
7	West Bengal	70188	48374	1581
8	Bihar	50987	33650	298
9	Andhra Pradesh	140933	63864	1349
10	Karnataka	124115	49788	2314
11	Telangana	62703	45388	519
12	J&K	20359	12217	377
13	Punjab	16119	10734	386
14	Haryana	34965	28227	421
15	Odisha	31877	20518	214
16	Assam	40270	30358	98
17	Kerala	23614	13023	74
18	Uttarakhand	7183	4168	80
19	Jharkhand	11314	4343	106
20	Chhattisgarh	6107	4930	64

COVID-19/COVID-19 data_analysis x Home Page - Select or create a n x COVID-19 data_analysis - Jupyter x +

localhost:8888/notebooks/COVID-19%20data_analysis.ipynb

jupyter COVID-19 data_analysis Last Checkpoint: Yesterday at 3:42 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

19 Jharkhand 11314 4343 106

20 Chhattisgarh 9192 6230 54

21 Himachal Pradesh 2564 1459 13

22 Tripura 4996 3327 21

23 Chandigarh 1051 667 15

24 Ladakh 1404 1095 7

25 Goa 5913 4211 45

26 Manipur 2621 1689 5

27 Puducherry 3472 2100 49

28 Nagaland 1693 635 4

29 Andaman and Nicobar Islands 548 214 5

30 Meghalaya 823 215 5

31 Arunachal Pradesh 1591 918 3

32 Dadra and Nagar Haveli and Daman and Diu 1149 725 2

33 Mizoram 408 247 0

34 Sikkim 639 231 1

```
In [5]: Total_cases_INDIA=data['Total Confirmed cases'].sum()
print('Total Confirmed COVID-19 Cases in India is',Total_cases_INDIA)

Total Confirmed COVID-19 Cases in India is 1697054

In [6]: data['Active cases']=data['Total Confirmed cases']-(data['Death']+data['Cured'])

In [7]: data

Out[7]:
```

COVID-19/COVID-19 data_analysis x Home Page - Select or create a n x COVID-19 data_analysis - Jupyter x +

localhost:8888/notebooks/COVID-19%20data_analysis.ipynb

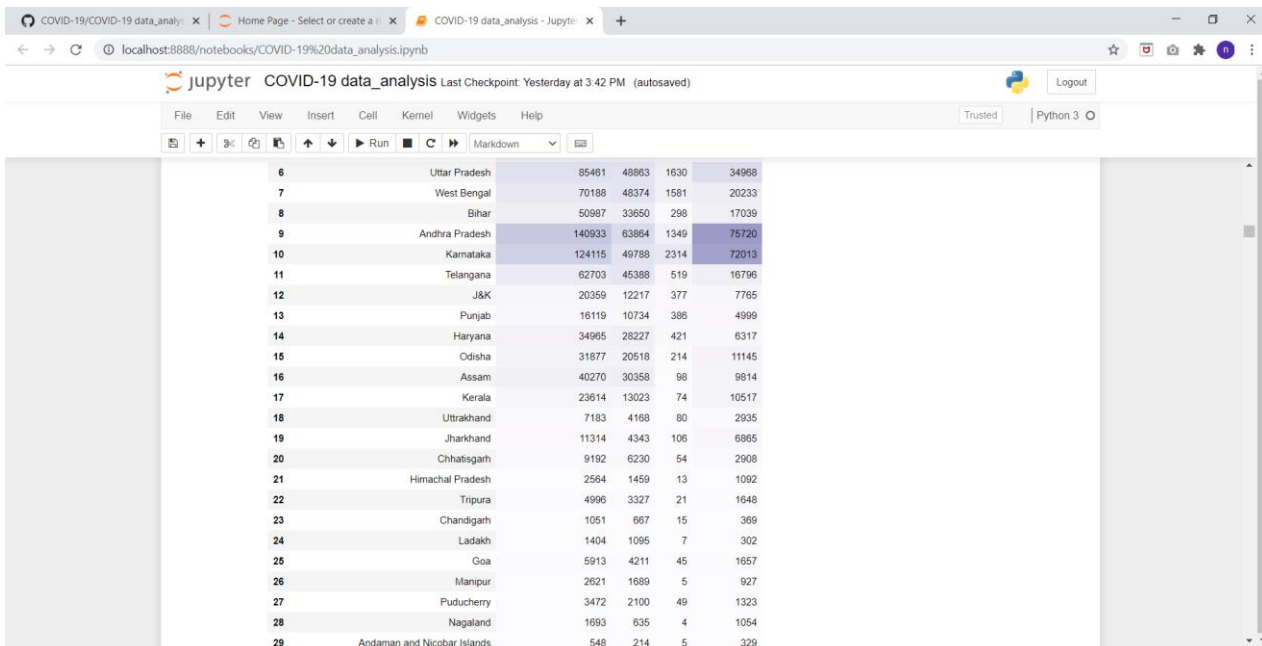
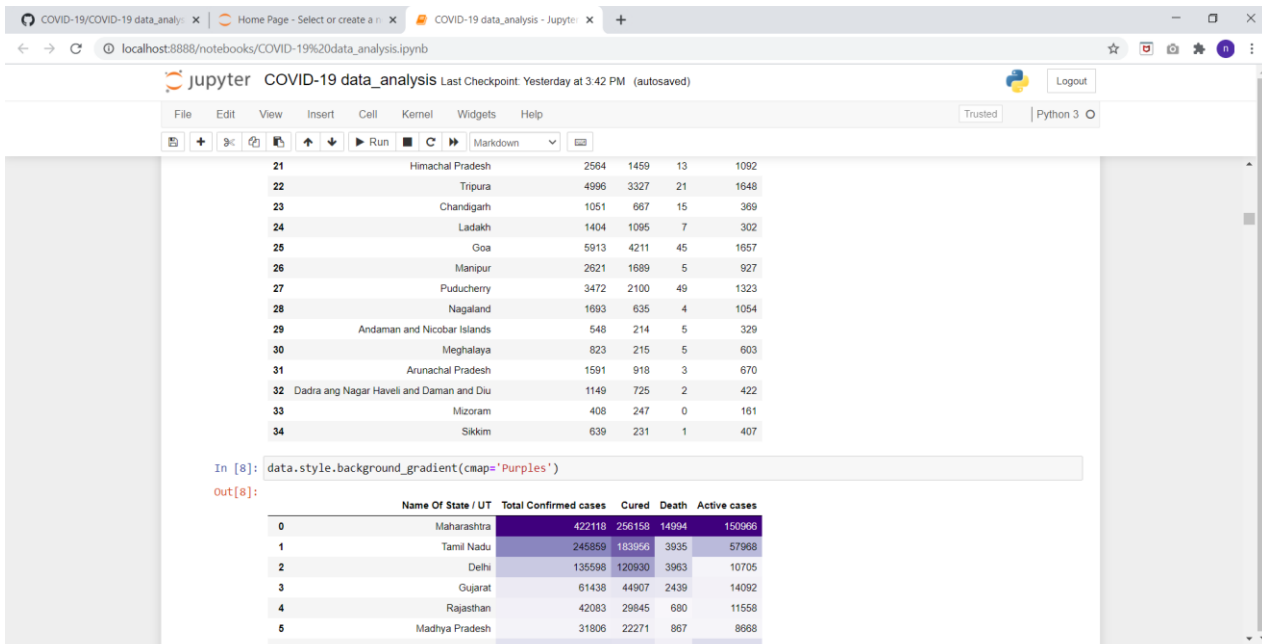
jupyter COVID-19 data_analysis Last Checkpoint: Yesterday at 3:42 PM (autosaved)

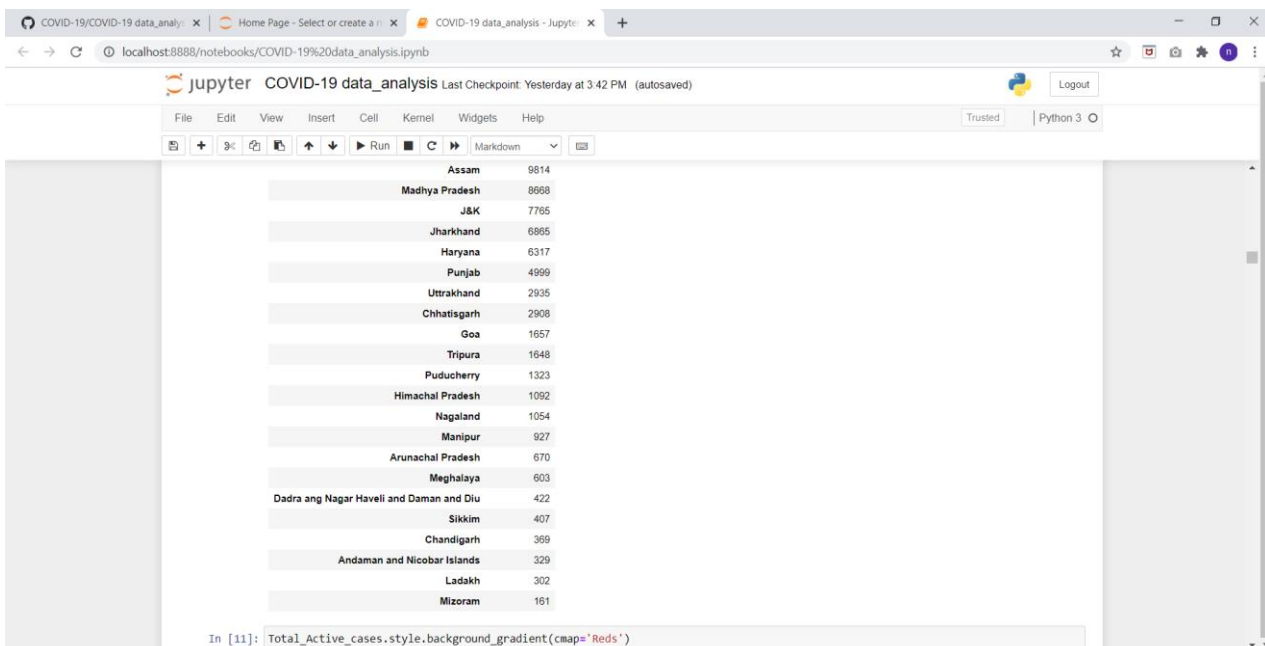
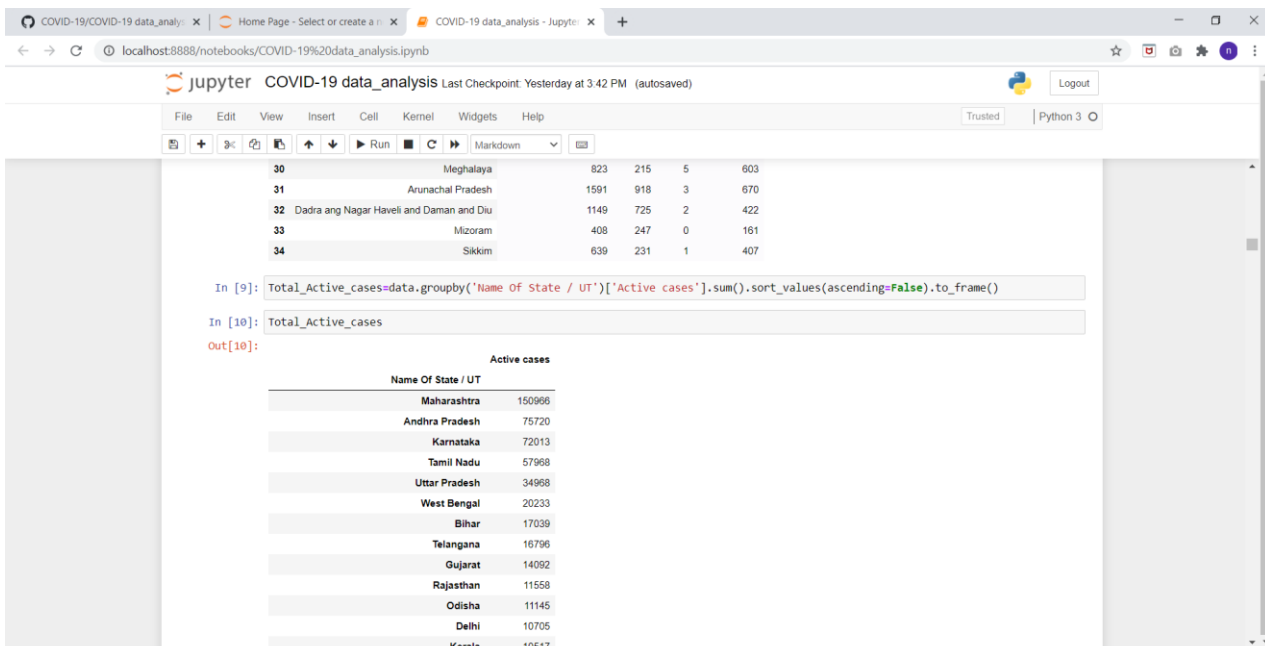
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

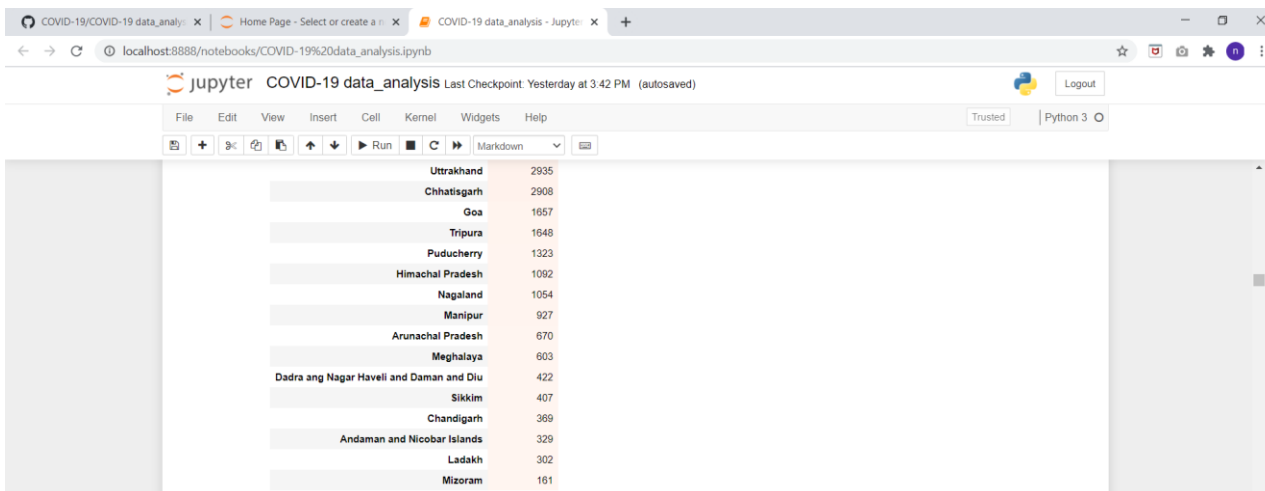
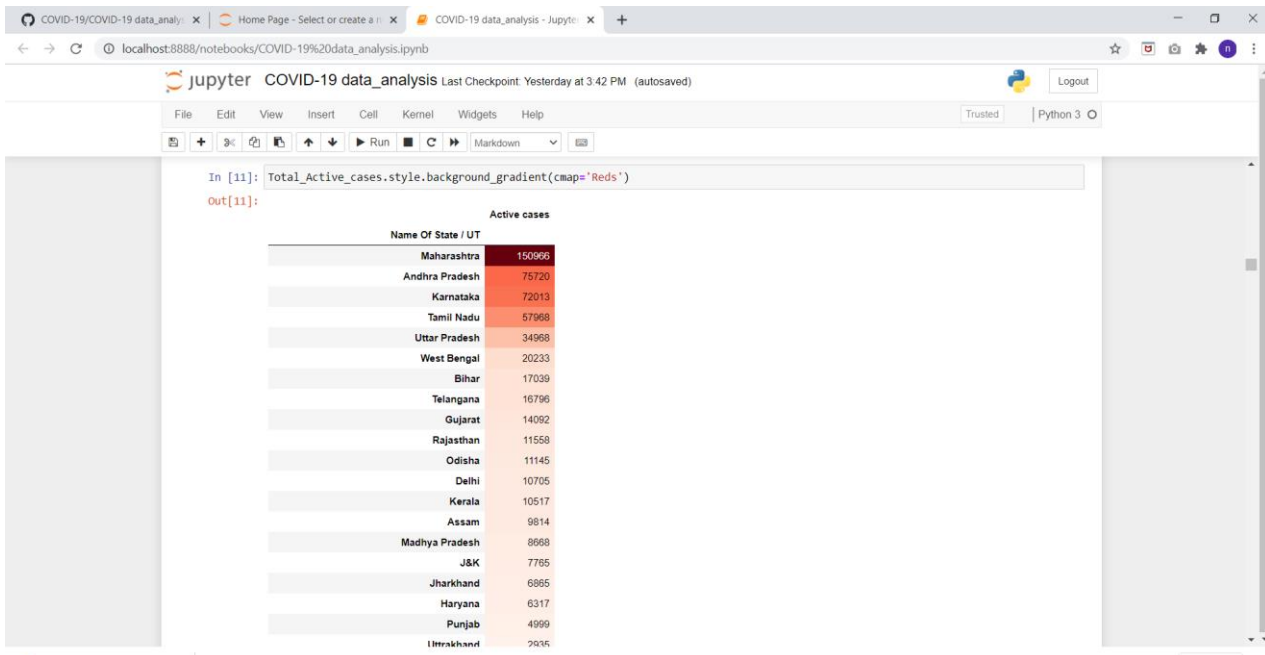
```
In [7]: data

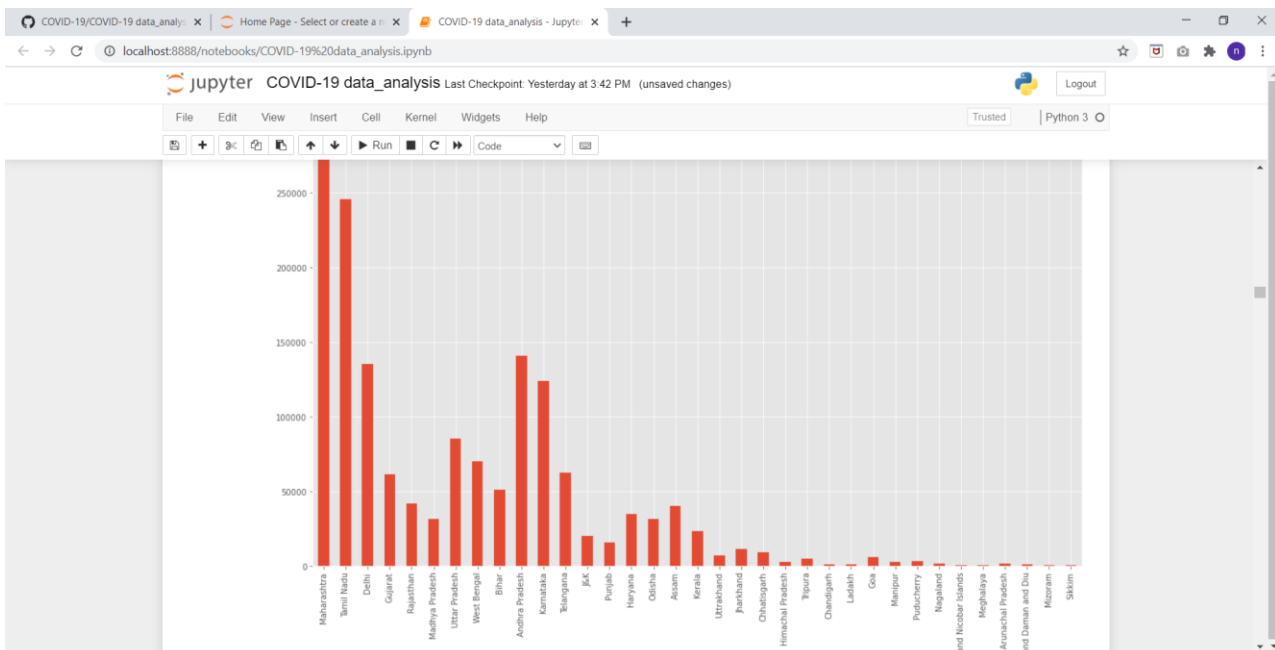
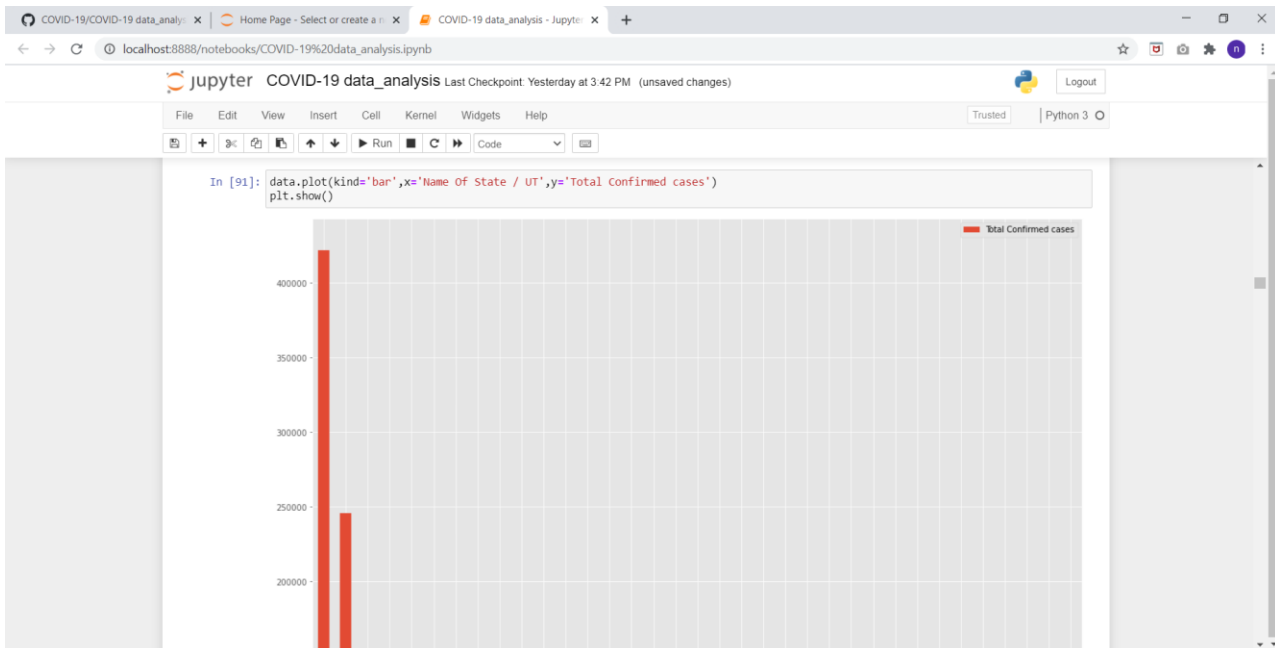
Out[7]:
```

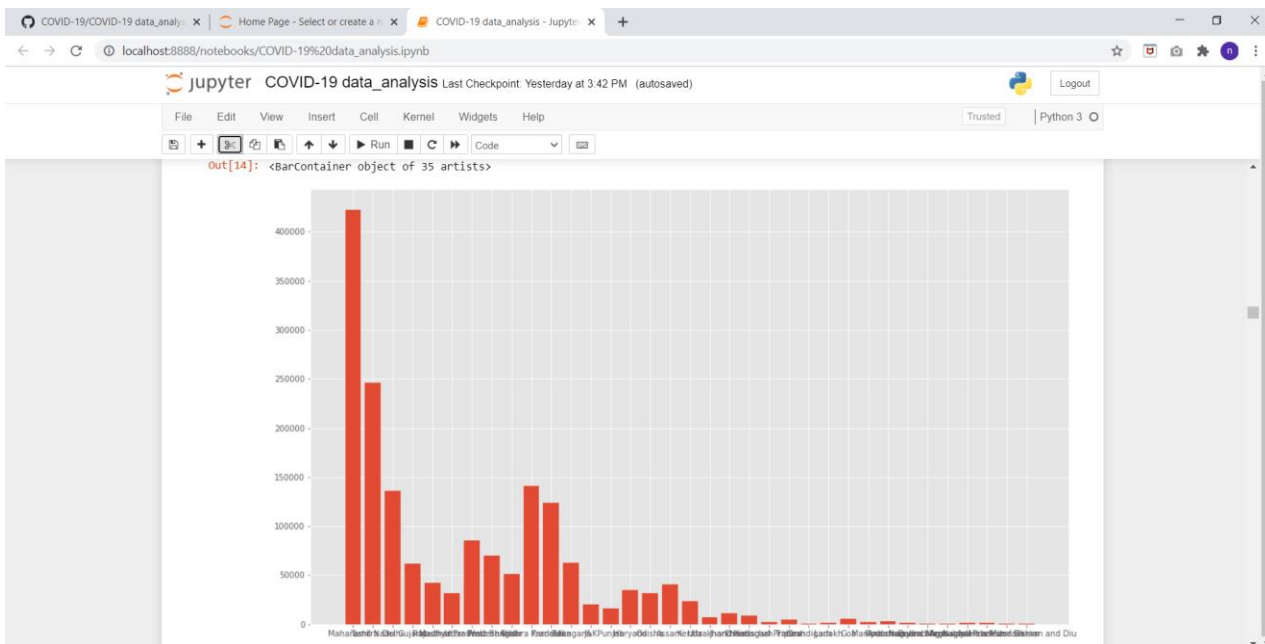
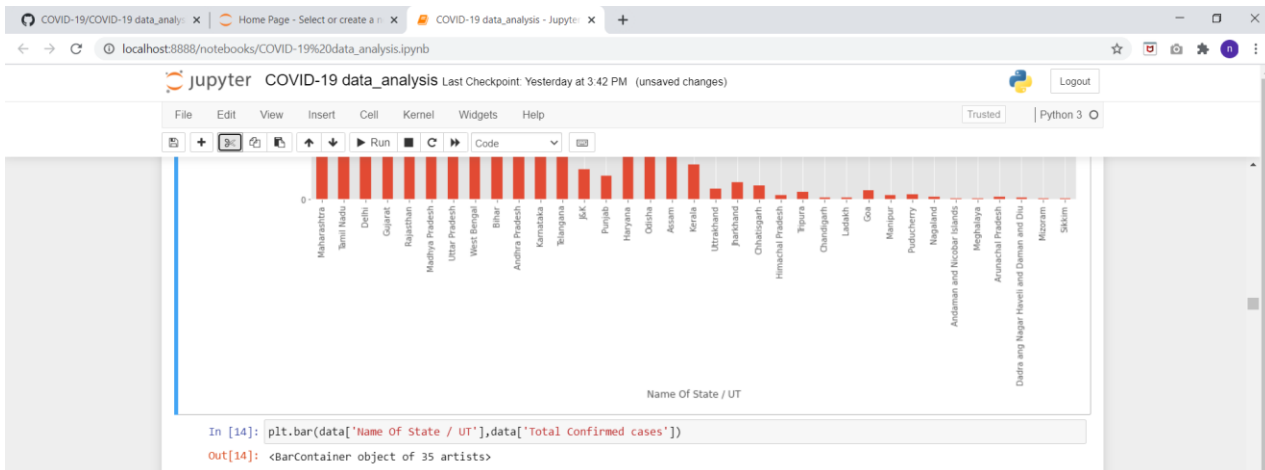
	Name Of State / UT	Total Confirmed cases	Cured	Death	Active cases
0	Maharashtra	422118	256158	14994	150966
1	Tamil Nadu	245859	183956	3935	57968
2	Delhi	135598	120930	3963	10705
3	Gujarat	61438	44907	2439	14092
4	Rajasthan	42083	29845	680	11558
5	Madhya Pradesh	31806	22271	867	8668
6	Uttar Pradesh	85461	48863	1630	34968
7	West Bengal	70188	48374	1581	20233
8	Bihar	50987	33650	298	17039
9	Andhra Pradesh	140933	63864	1349	75720
10	Karnataka	124115	49788	2314	72013
11	Telangana	62703	45388	519	16796
12	J&K	20359	12217	377	7765
13	Punjab	16119	10734	386	4999
14	Haryana	34965	28227	421	6317
15	Odisha	31877	20518	214	11145
16	Assam	40270	30358	98	9814
17	Kerala	23614	13023	74	10517
18	Uttarakhand	7183	4168	80	2935
19	Jharkhand	11314	4343	106	6865
20	Chhattisgarh	9192	6230	54	2908

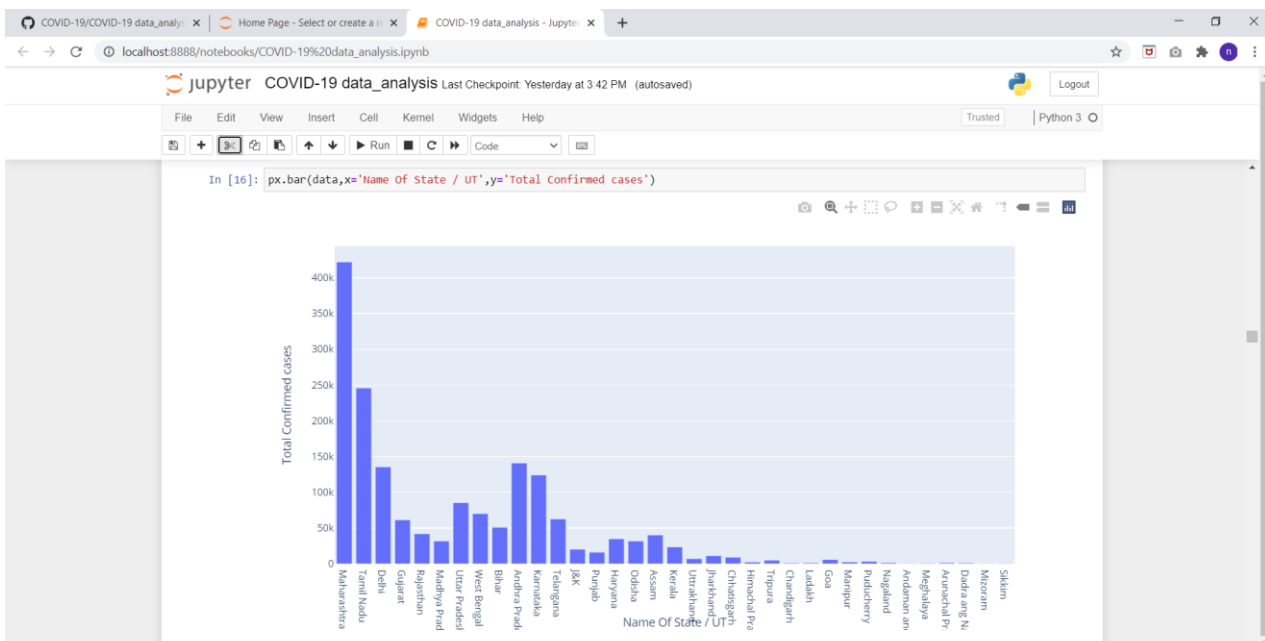
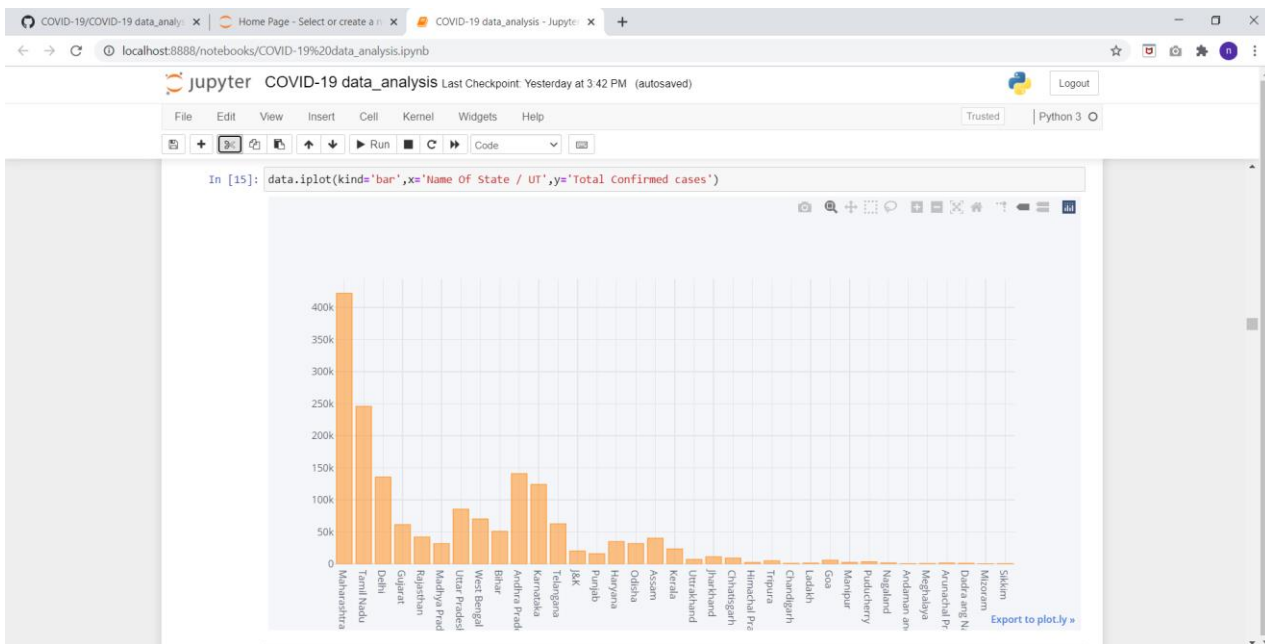


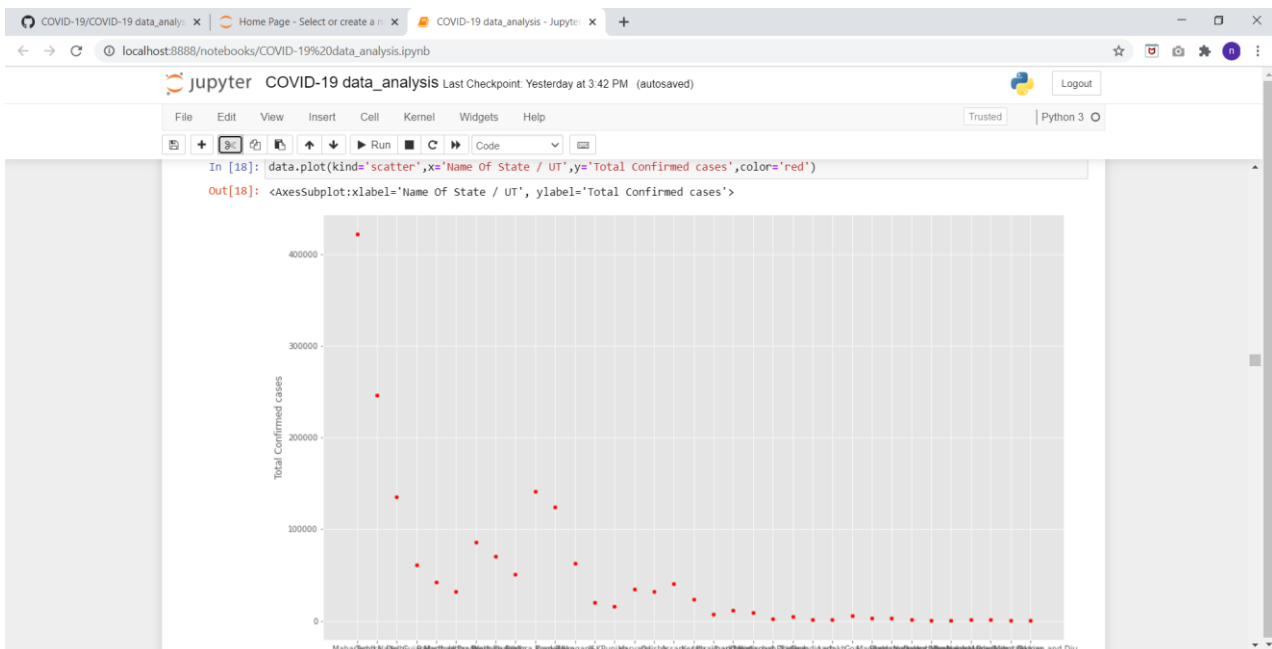
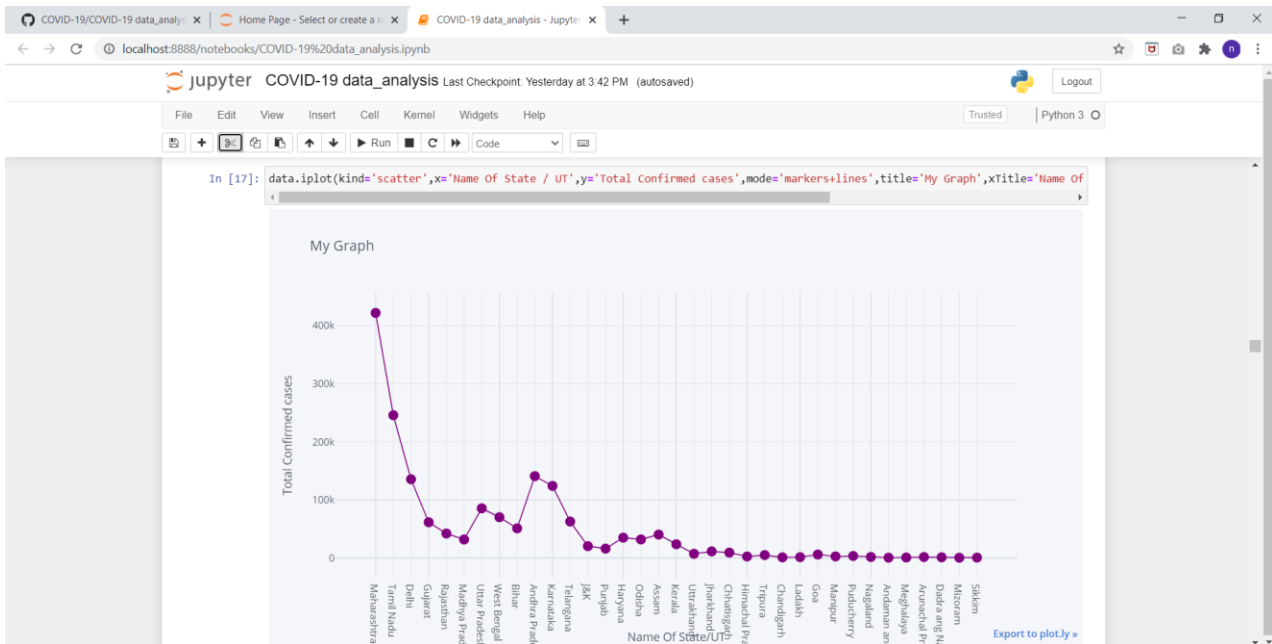


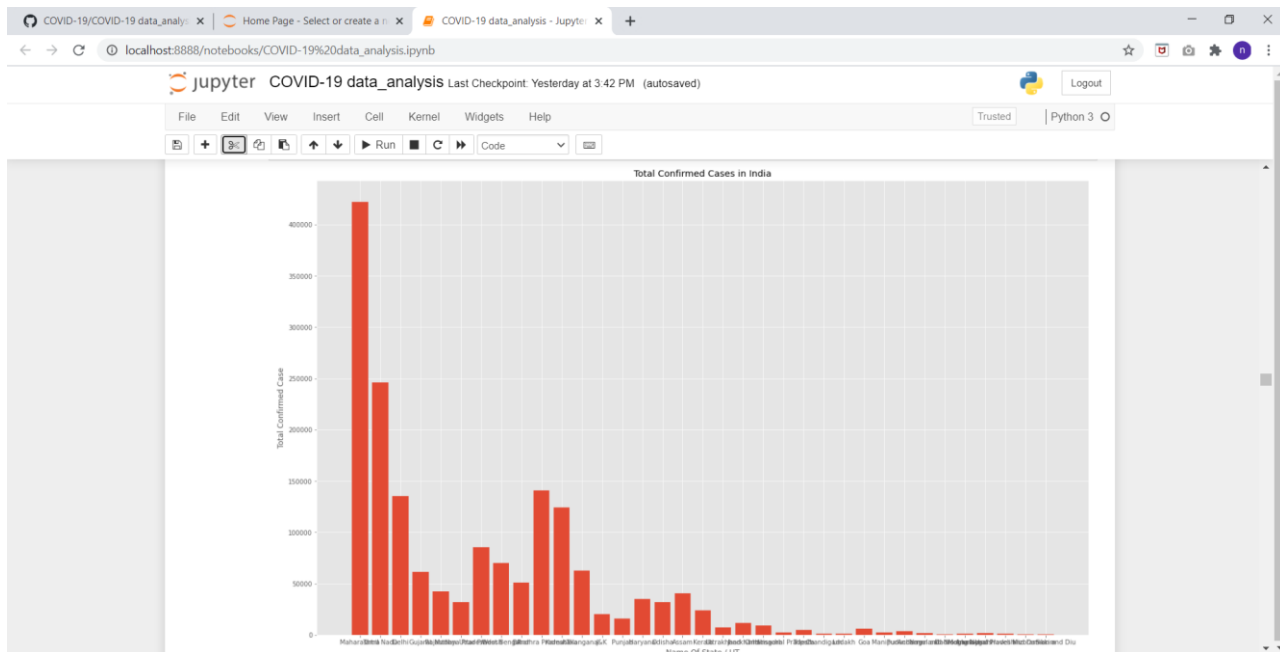


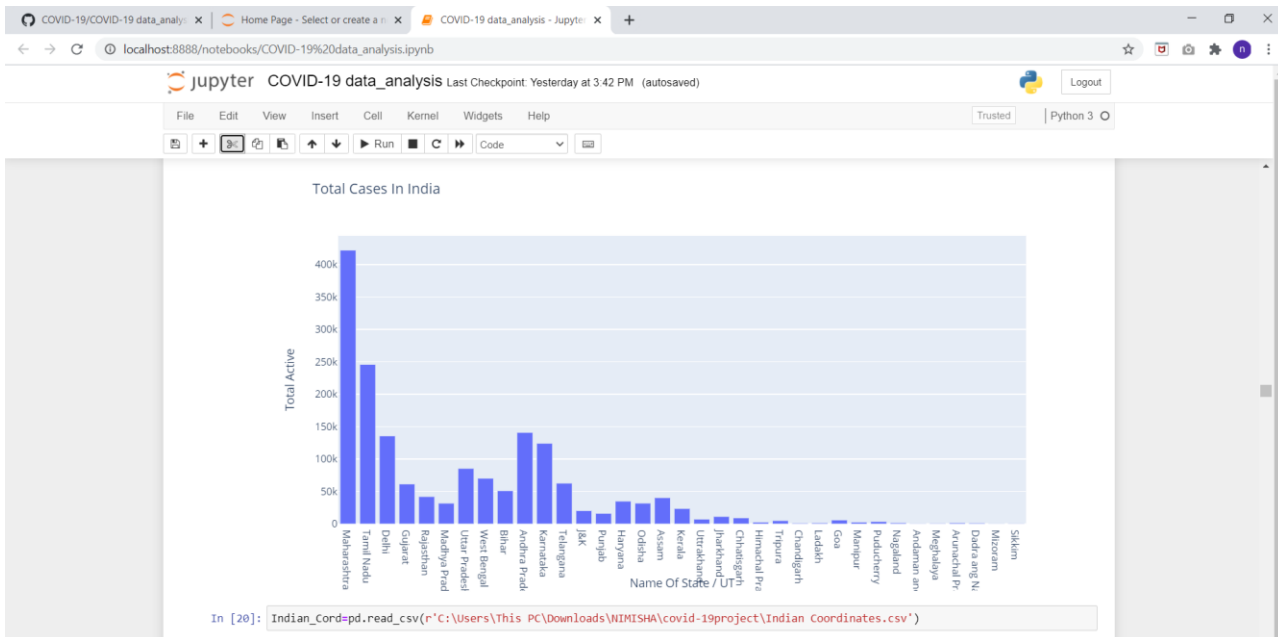












COVID-19/COVID-19 data_analy: x Home Page - Select or create a n: x COVID-19 data_analysis - Jupyter: x +

localhost:8888/notebooks/COVID-19%20data_analysis.ipynb

jupyter COVID-19 data_analysis Last Checkpoint: Yesterday at 3:42 PM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

In [20]: `Indian_Cord=pd.read_csv(r'C:\Users\This PC\Downloads\NIMISHA\covid-19project\Indian Coordinates.csv')`

In [21]: `Indian_Cord`

Out[21]:

	Name Of State / UT	Latitude	Longitude
0	Andaman And Nicobar	11.667026	92.735983
1	Andhra Pradesh	14.750429	78.570026
2	Arunachal Pradesh	27.100399	93.616601
3	Assam	26.749981	94.216667
4	Bihar	25.785414	87.479973
5	Chandigarh	30.719997	76.780006
6	Chhattisgarh	22.090420	82.159987
7	Dadra And Nagar Haveli	20.266578	73.016618
8	Delhi	28.669993	77.230004
9	Goa	15.491997	73.818001
10	Haryana	28.450006	77.019991
11	Himachal Pradesh	31.100025	77.166597
12	Union Territory of Jammu and Kashmir	33.450000	76.240000
13	Jharkhand	23.800393	86.419986
14	Karnataka	12.570381	76.919997
15	Kerala	8.900373	76.569993
16	Lakshadweep	10.562573	72.636867
17	Madhya Pradesh	21.300391	76.130019
18	Maharashtra	19.250232	73.160175

```
COVID-19/COVID-19 data_analysis | Home Page - Select or create a notebook | COVID-19 data_analysis - Jupyter | +
localhost:8888/notebooks/COVID-19%20data_analysis.ipynb
jupyter COVID-19 data_analysis Last Checkpoint: Yesterday at 3:42 PM (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3
+ - Run Code
18 Maharashtra 19.250232 73.160175
19 Manipur 24.799971 93.950017
20 Meghalaya 25.570492 91.880014
21 Mizoram 23.710399 92.720015
22 Nagaland 25.666998 94.116570
23 Orissa 19.820430 85.900017
24 Puducherry 11.934994 79.830000
25 Punjab 31.519974 75.980003
26 Rajasthan 26.449999 74.639981
27 Sikkim 27.333330 88.616647
28 Telangana 18.112400 79.019300
29 Tamil Nadu 12.920386 79.150042
30 Tripura 23.835404 91.279999
31 Uttar Pradesh 27.599981 78.050006
32 Uttarakhand 30.320409 78.050006
33 West Bengal 22.580390 88.329947
34 Union Territory of Ladakh 34.100000 77.340000

In [22]: data_full=pd.merge(Indian_cord,data,on = 'Name Of State / UT')

In [23]: data_full
```

```
COVID-19/COVID-19 data_analysis | Home Page - Select or create a notebook | COVID-19 data_analysis - Jupyter | +
localhost:8888/notebooks/COVID-19%20data_analysis.ipynb
jupyter COVID-19 data_analysis Last Checkpoint: Yesterday at 3:42 PM (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3
+ - Run Code

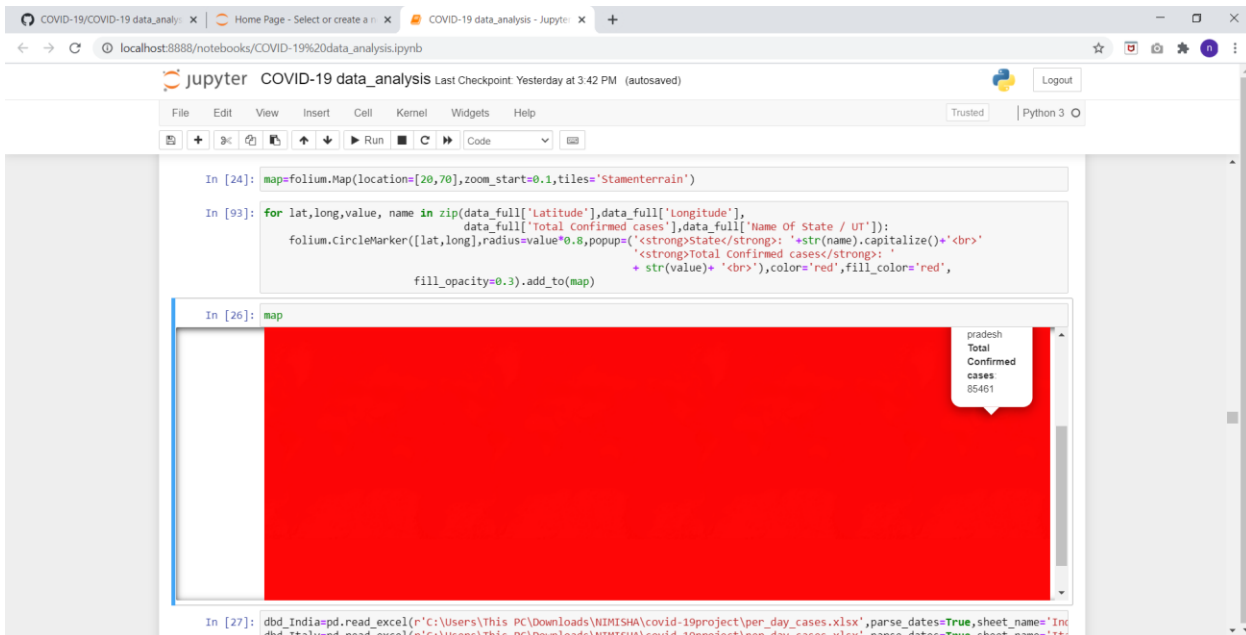
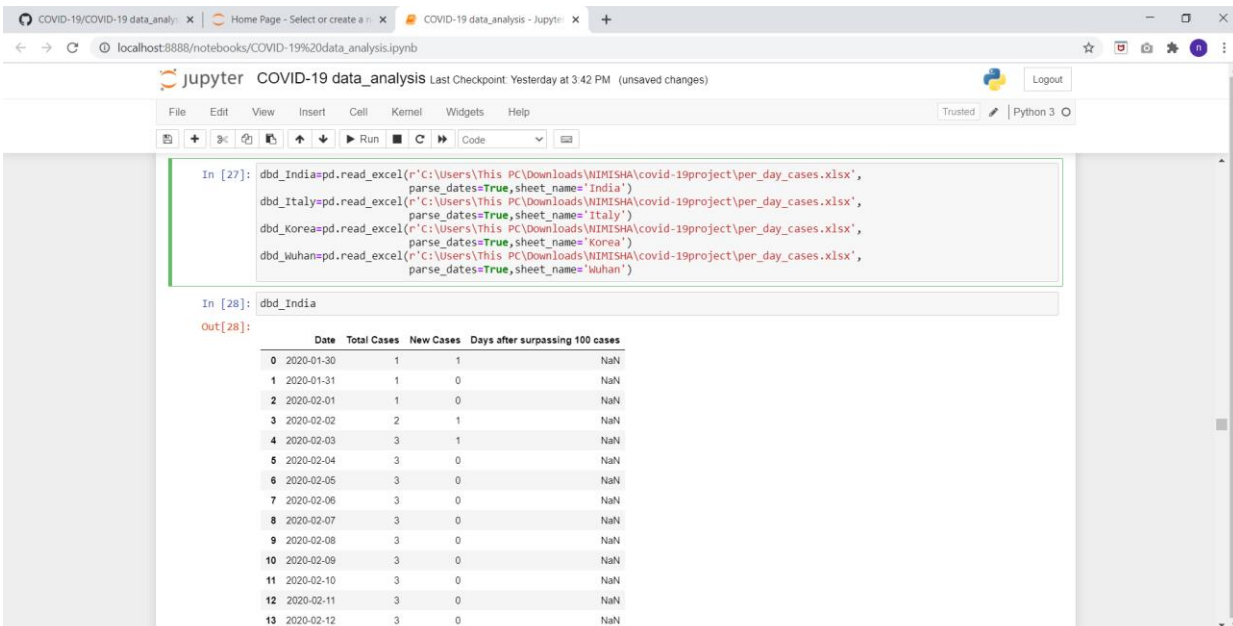
In [22]: data_full=pd.merge(Indian_cord,data,on = 'Name Of State / UT')

In [23]: data_full
Out[23]:
   Name Of State / UT  Latitude  Longitude  Total Confirmed cases  Cured  Death  Active cases
0  Andhra Pradesh    14.750429   78.570026          140933      63864    1349      75720
1      Delhi        28.669993   77.230004          135598    120930    3963    10705
2  Haryana          28.450006   77.019991          34965    28227     421     6317
3  Karnataka        12.570381   76.919997          124115    49788    2314    72013
4  Kerala            8.900373   76.569993           23614    13023     74    10517
5  Maharashtra       19.250232   73.160175          422118    256158   14994   150966
6  Punjab           31.519974   75.980003           16119    10734     386    4999
7  Rajasthan         26.449999   74.639981           42083    29845     680    11558
8  Tamil Nadu        12.920386   79.150042          245859    183956    3935    57968
9  Uttar Pradesh     27.599981   78.050006           85481    48863    1630    34968

In [24]: map=folium.Map(location=[20,70],zoom_start=0.1,tiles='Stamenterrain')

In [25]: for lat,long,value, name in zip(data_full['Latitude'],data_full['Longitude'],data_full['Total Confirmed cases'],data_full['Name']):
folium.CircleMarker([lat,long],radius=value*0.8,popup=folium.Popup(f'<strong>State</strong>: {str(name).capitalize()}<br><strong>Total Confirmed cases</strong>: {value}'))

In [26]: map
```

The screenshot shows a Jupyter Notebook interface with the following code and output:

```
In [27]: dbd_India=pd.read_excel(r'C:\Users\This PC\Downloads\NIMISHA\covid-19project\per_day_cases.xlsx',
      parse_dates=True,sheet_name='India')
      dbd_Italy=pd.read_excel(r'C:\Users\This PC\Downloads\NIMISHA\covid-19project\per_day_cases.xlsx',
      parse_dates=True,sheet_name='Italy')
      dbd_Korea=pd.read_excel(r'C:\Users\This PC\Downloads\NIMISHA\covid-19project\per_day_cases.xlsx',
      parse_dates=True,sheet_name='Korea')
      dbd_Wuhan=pd.read_excel(r'C:\Users\This PC\Downloads\NIMISHA\covid-19project\per_day_cases.xlsx',
      parse_dates=True,sheet_name='Wuhan')

In [28]: dbd_India
```

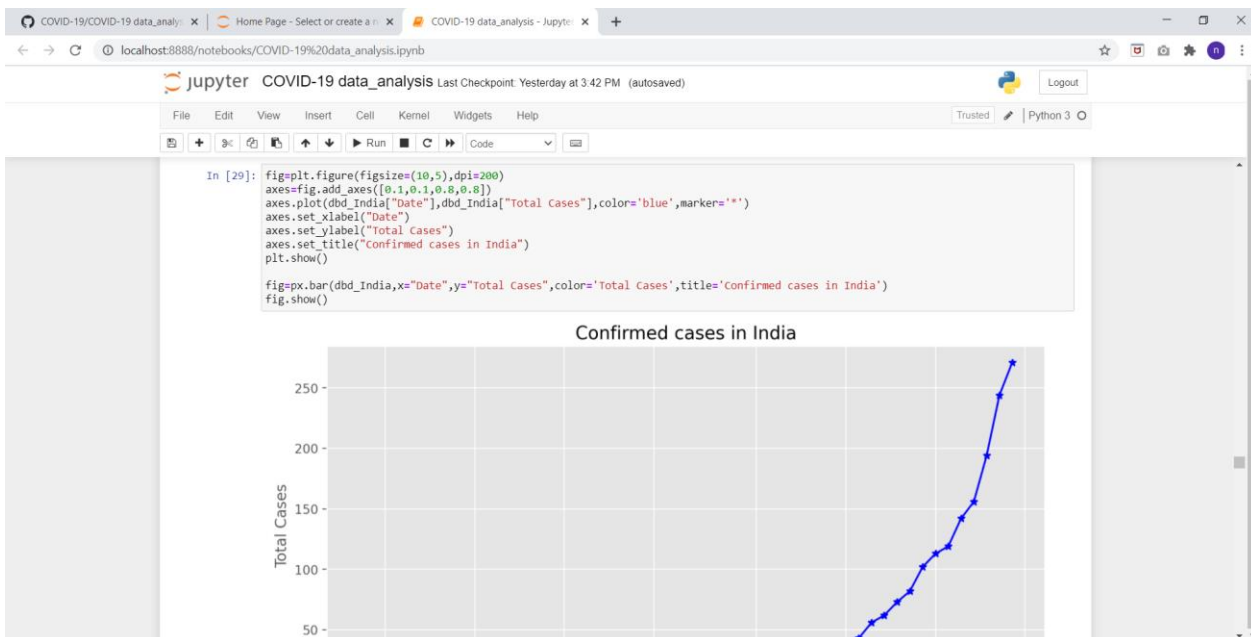
The output is a table of COVID-19 data for India:

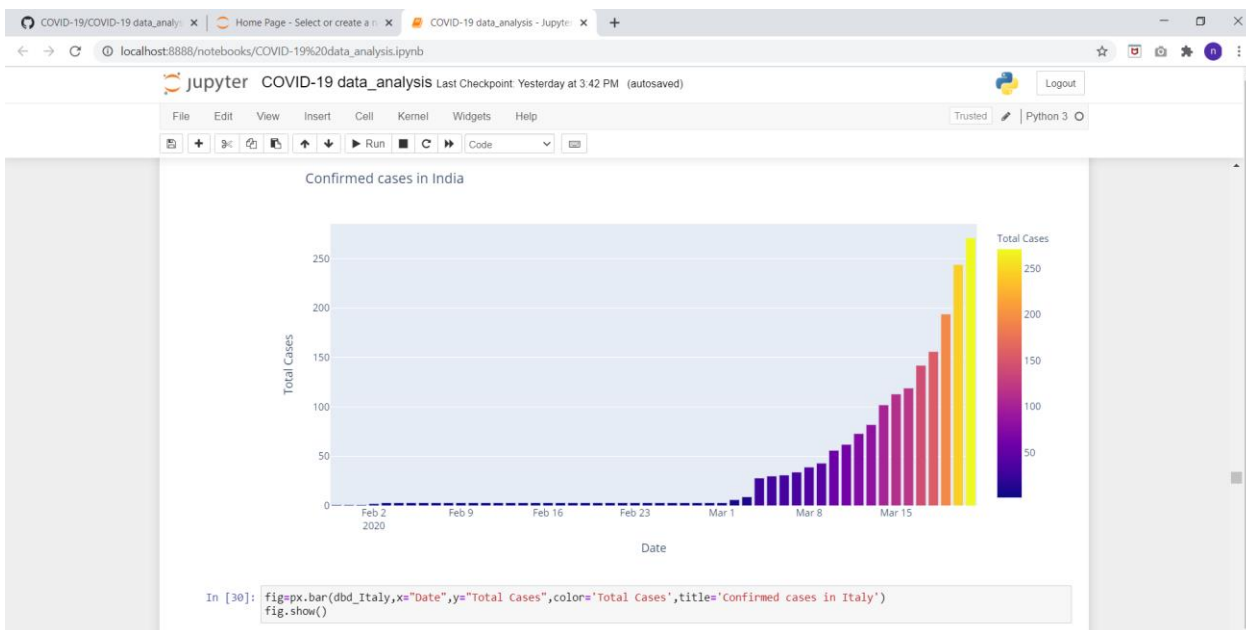
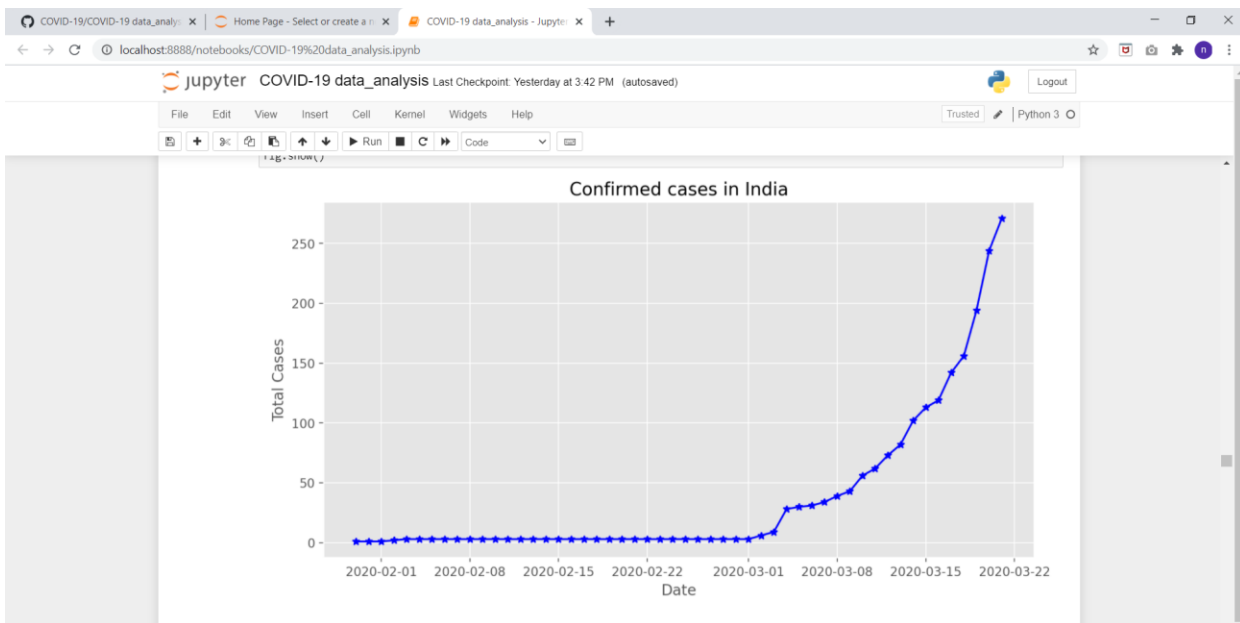
	Date	Total Cases	New Cases	Days after surpassing 100 cases
0	2020-01-30	1	1	NaN
1	2020-01-31	1	0	NaN
2	2020-02-01	1	0	NaN
3	2020-02-02	2	1	NaN
4	2020-02-03	3	1	NaN
5	2020-02-04	3	0	NaN
6	2020-02-05	3	0	NaN
7	2020-02-06	3	0	NaN
8	2020-02-07	3	0	NaN
9	2020-02-08	3	0	NaN
10	2020-02-09	3	0	NaN
11	2020-02-10	3	0	NaN
12	2020-02-11	3	0	NaN
13	2020-02-12	3	0	NaN

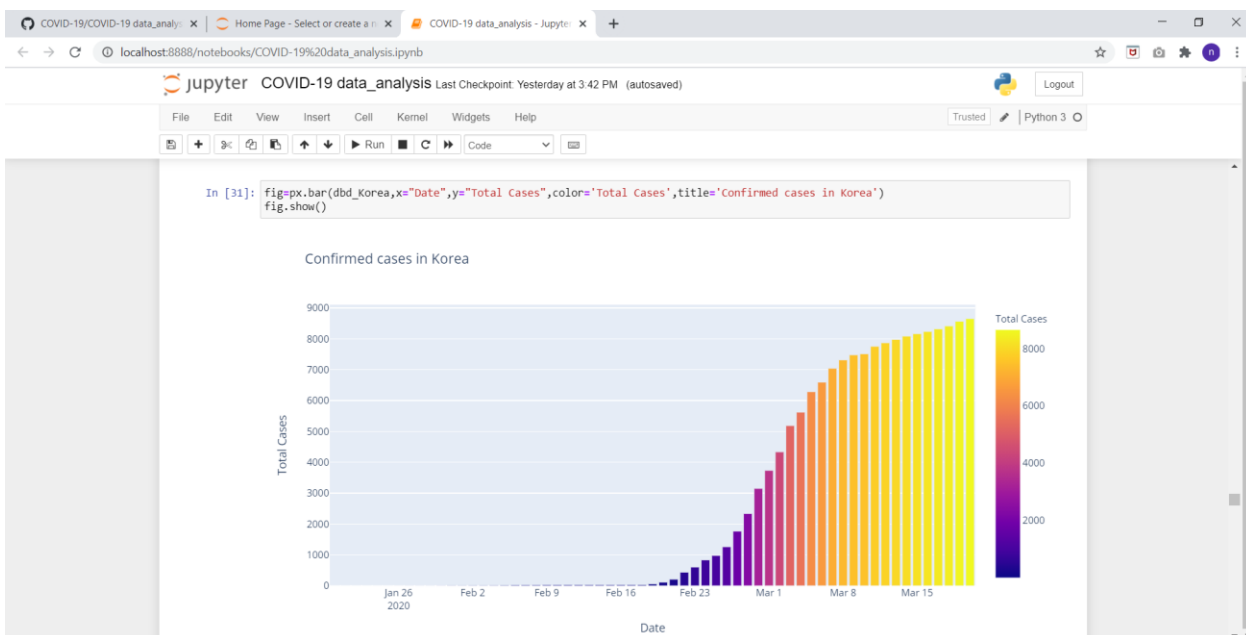
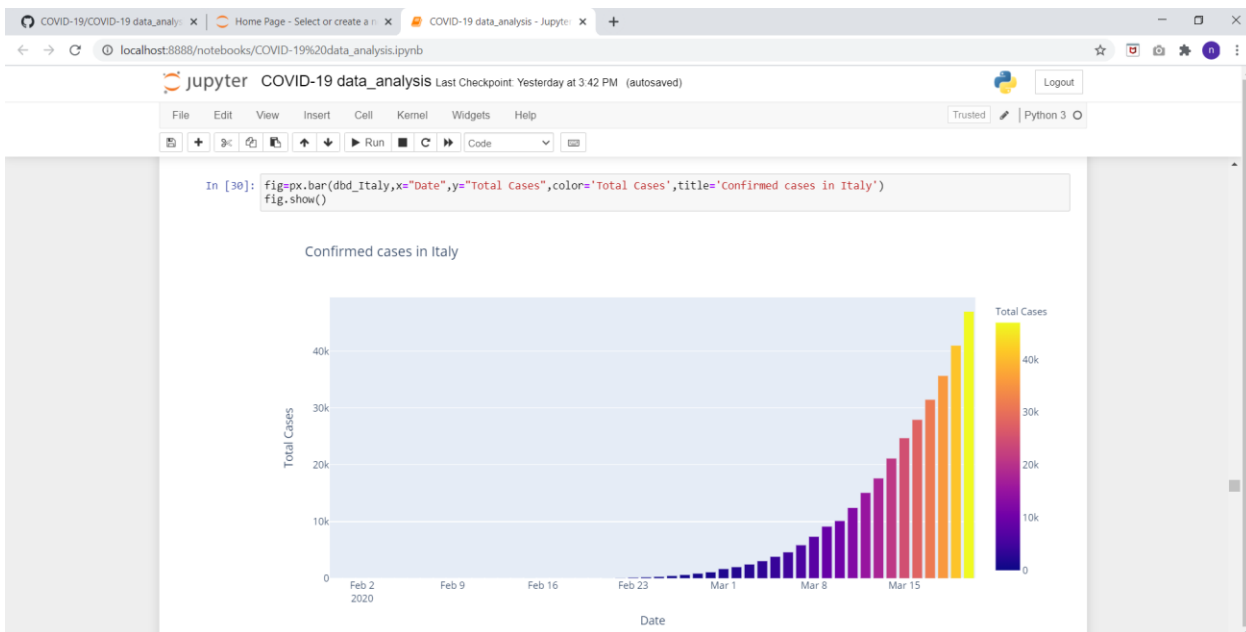
COVID-19/COVID-19 data_analysis | Home Page - Select or create a notebook | COVID-19 data_analysis - Jupyter | localhost:8888/notebooks/COVID-19%20data_analysis.ipynb

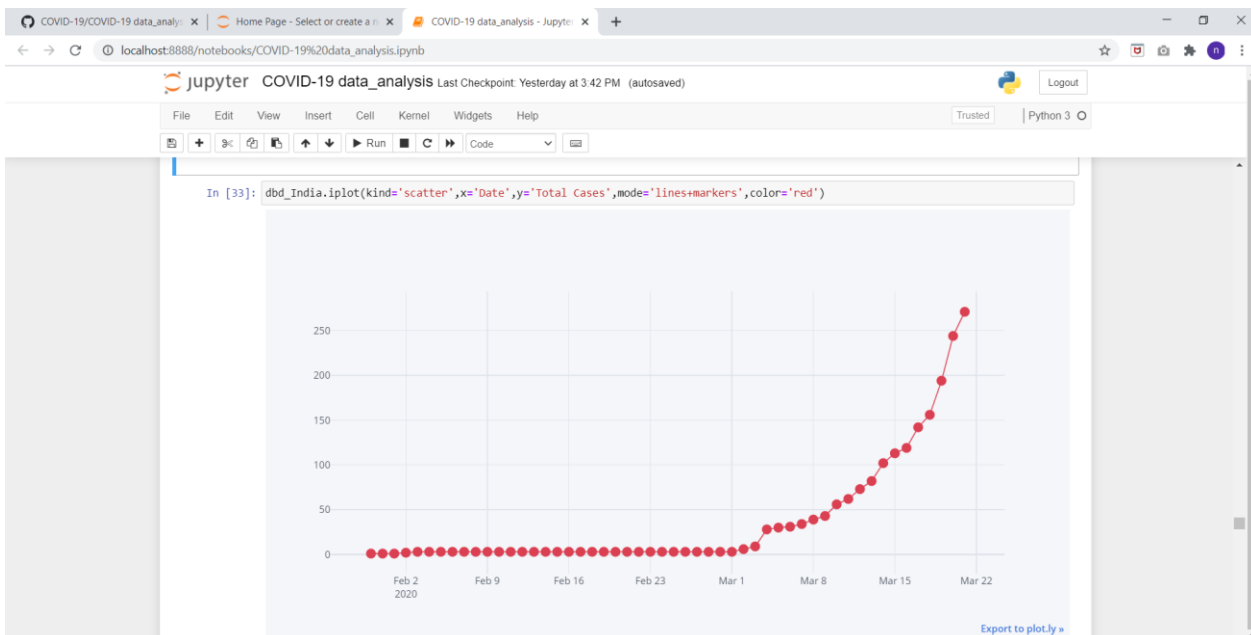
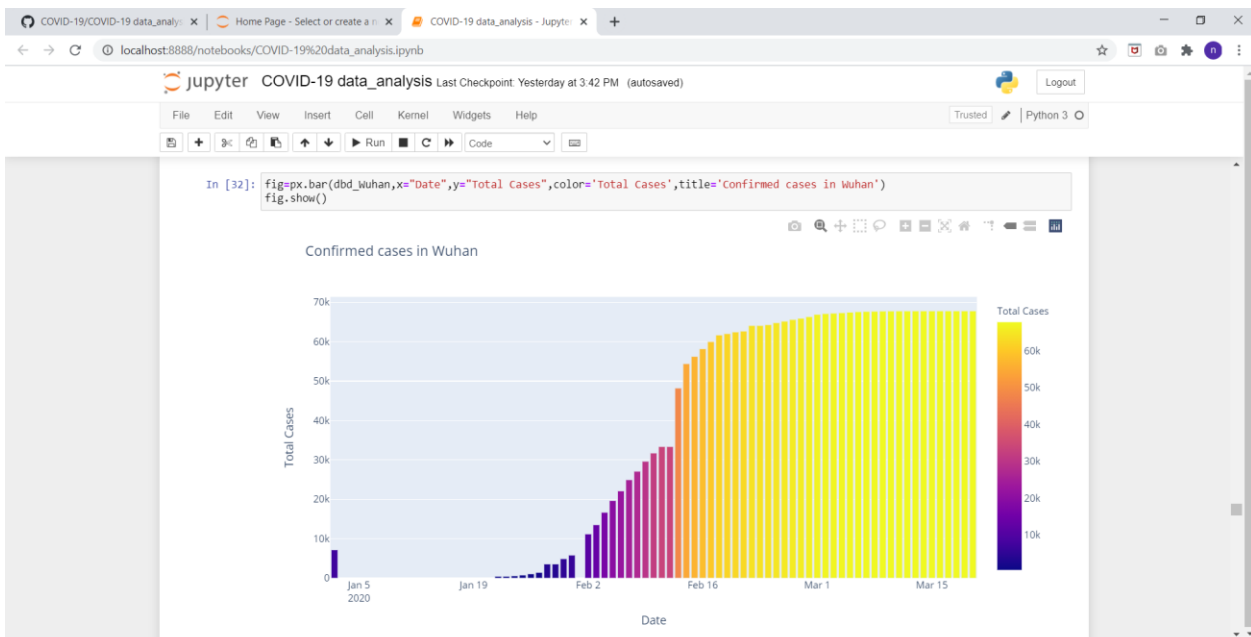
jupyter COVID-19 data_analysis Last Checkpoint: Yesterday at 3:42 PM (autosaved) Logout

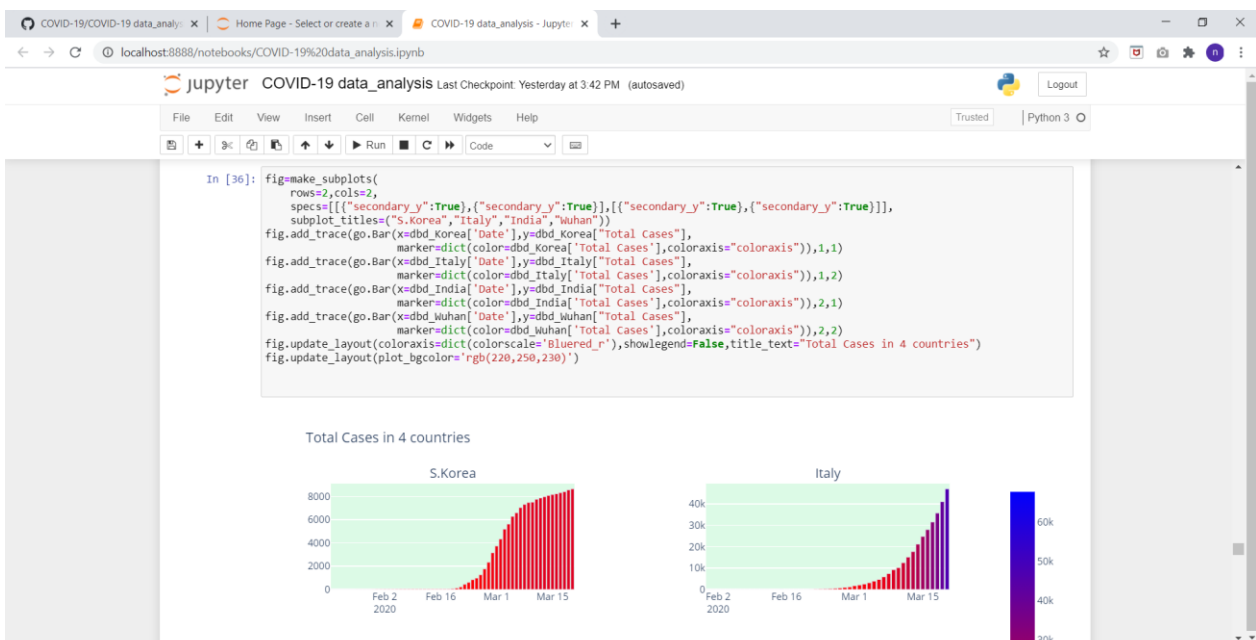
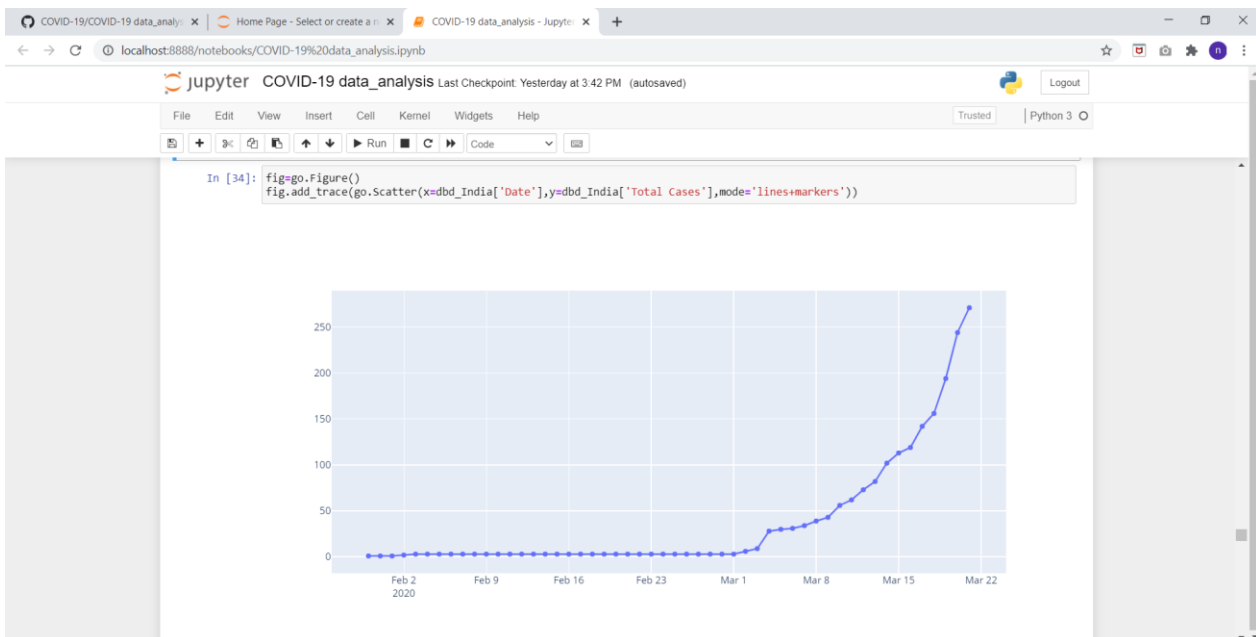
	Date	Confirmed	Deaths	Recovered
29	2020-02-28	3	0	NaN
30	2020-02-29	3	0	NaN
31	2020-03-01	3	0	NaN
32	2020-03-02	6	3	NaN
33	2020-03-03	9	3	NaN
34	2020-03-04	28	19	NaN
35	2020-03-05	30	2	NaN
36	2020-03-06	31	1	NaN
37	2020-03-07	34	3	NaN
38	2020-03-08	39	5	NaN
39	2020-03-09	43	4	NaN
40	2020-03-10	56	13	NaN
41	2020-03-11	62	6	NaN
42	2020-03-12	73	11	NaN
43	2020-03-13	82	9	NaN
44	2020-03-14	102	20	0.0
45	2020-03-15	113	11	1.0
46	2020-03-16	119	6	2.0
47	2020-03-17	142	23	3.0
48	2020-03-18	156	14	4.0
49	2020-03-19	194	38	5.0
50	2020-03-20	244	50	6.0
51	2020-03-21	271	27	7.0

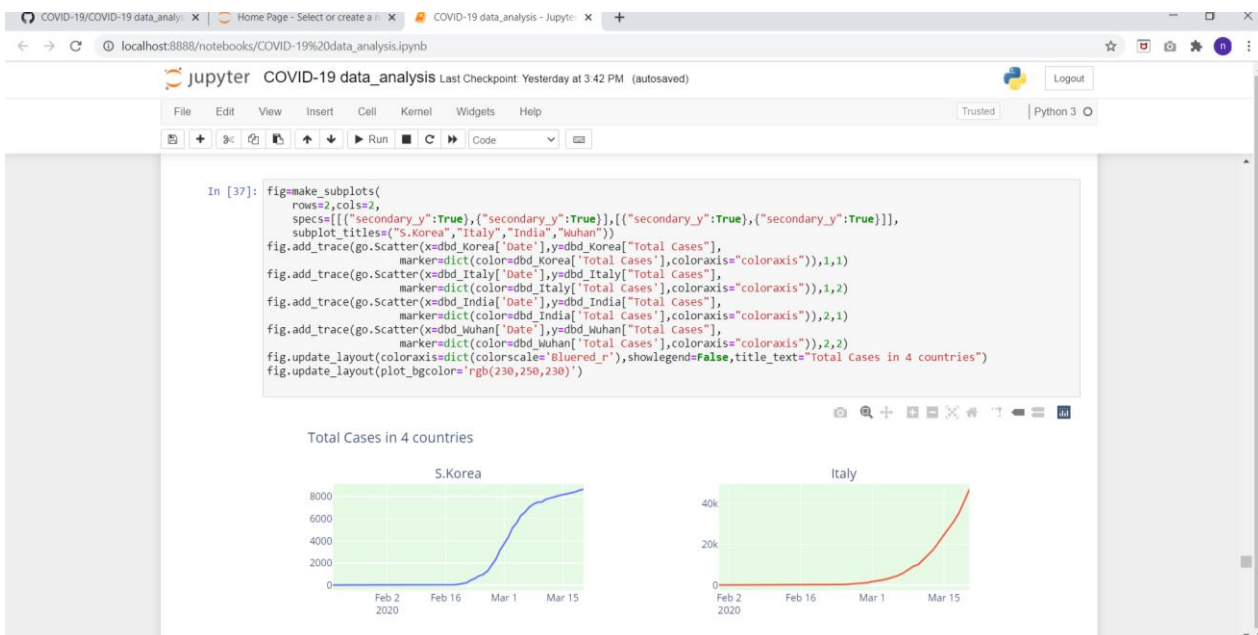
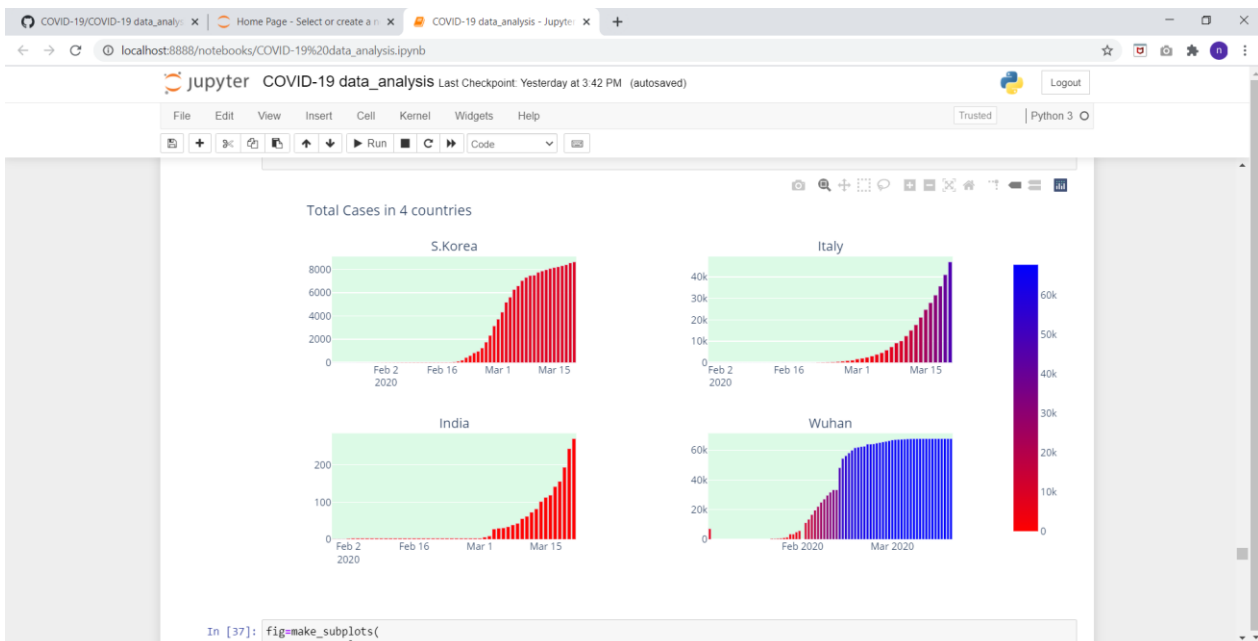


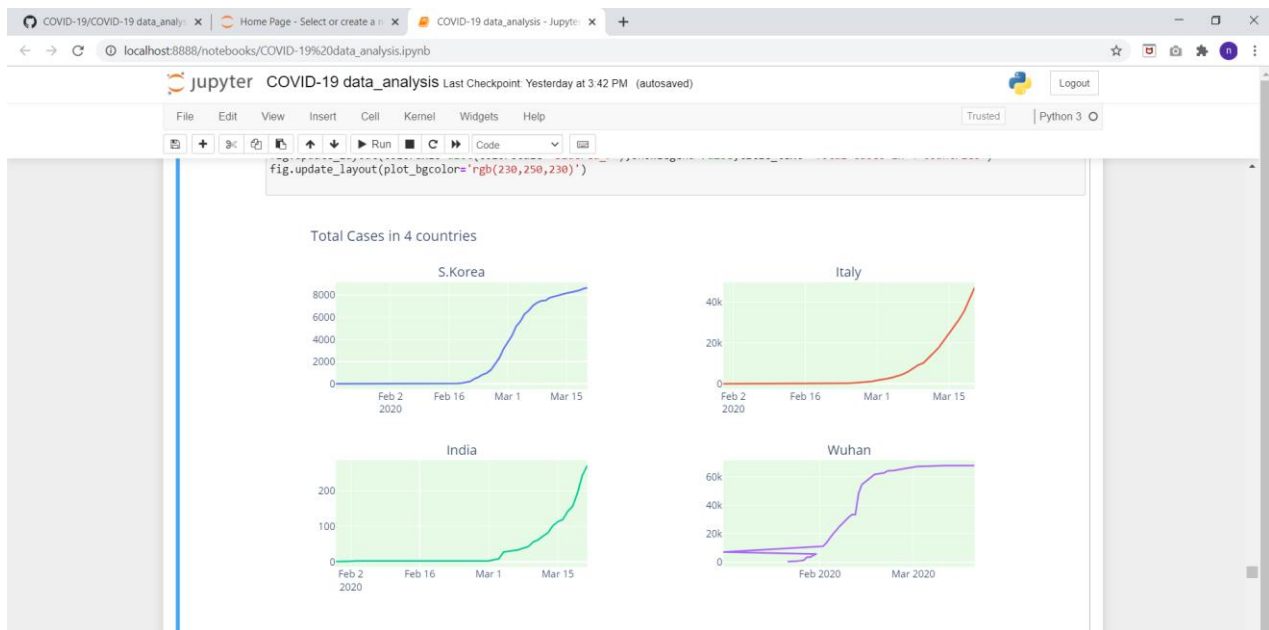












References

The source of the data used in the project is:

- ❖ google.com
- ❖ mohfw.gov.in (The official site of Ministry Of Health And Family Welfare of India)
- ❖ covid19india.org
- ❖ worldometers.info/coronavirus
- ❖ Some other sites as well for collecting data.
- ❖ *(Data upto 31st July, 2020)*