COVID-19 Dashboard

A Summer Internship Report

Submitted By

Parth Prajapati (180770107583)

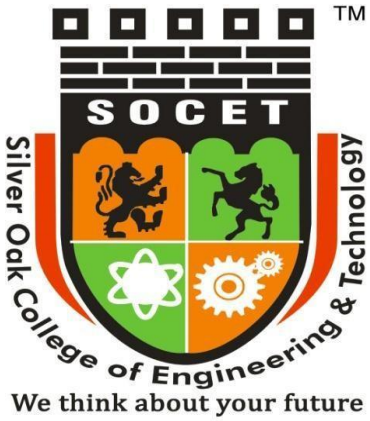
Under Subject of

Summer Internship

(3170001)

B.E. - IV, Semester - VII

(Department of Computer Engineering)



SILVER OAK COLLEGE OF ENGINEERING AND TECHNOLOGY

Opp. Bhagwat Vidyapith, Near Gota Cross Road, Ahmedabad- 382481

Silver Oak College of Engineering & Technology

Department of Computer Engineering

2021 - 2022

CERTIFICATE

This is to certify that the project entitled "**COVID-19 Dashboard"** has been carried out by **"Parth Prajapati (180770107583)"** under my guidance in fulfillment of the Summer Internship (3170001) Subject of Bachelor of Engineering in **Computer Engineering**– 7th Semester of Gujarat Technological University, Ahmedabad during the academic year 2021- 2022.

Name of Guide	**Head of Department**
Prof. Hemal Patel	Dr. Satvik Khara
(Computer Engineering)	(Computer Engineering)

Candidate's Declaration

I hereby declare that the Summer Internship report titled "COVID-19 Data Analysis with Dashboard" submitted towards the completion of Summer Internship in 7th semester of Bachelor of Computer Engineering in Silver Oak College of Engineering & Technology, Ahmedabad is an authenticate record of mine work carried out.

I further declare that to the best of my knowledge the report of C.E. 7th semester.

Candidate's Name **: Prajapati Parth Ramanbhai**

Branch **: Computer Engineering**

Enrollment Number **: 180770107583**

Candidate's Signature **: Parth**

Submitted to:

Silver Oak College of Engineering & Technology, Ahmedabad

Affiliated to:

Gujarat Technological University

Abstract

COVID-19 has been recognized as a global threat, and several studies are being conducted in order to contribute to the fight and prevention of this pandemic. Understanding the spread of COVID-19 is critical to informing public health decisions and lessening its impact on communities. We're supporting the development of data platforms to help model disease and projects that explore the use of diverse public datasets to more accurately predict the spread of the virus. due to the vast amount of data available on COVID-19 from various sources, there is a need to review the roles of data analysis in controlling the spread of COVID-19, presenting the main challenges and directions of COVID-19 data analysis, as well as providing a framework on the related existing applications and studies to facilitate future research on COVID-19 analysis

Acknowledgement

We would like to enlarge our sincerely thanks with a deep sense of thankfulness and respect to all those who has given us vast help and guidance during this project. We would like to convey our sincere thanks to our faculty guide **Prof. Hemal Patel** for providing a vision about the system and for giving us an opportunity to undertake such a great challenging and innovative work. We are grateful for the guidance, encouragement, understanding and insightful support given in the development process.

We would like to extend my gratitude to **Prof. Satvik Khara** Head of Computer Engineering Department, Silver Oak College of Engineering and Technology, Ahmedabad, for his continuous encouragement and motivation.

Last but not the least we would like to mention here that we are greatly indebted to each and everybody who has been associated with our project at any stage but whose name does not find a put in this acceptance.

Your Sincerely,

Parth Prajapati (180770107583)

TABLE OF CONTENT

1. INTRODUCTION

1.1 Project Summary

COVID-19 has infected millions of people around the world. Countries try to take effective measures to combat the pandemic, prevent the spread, decrease the risk of infection. In this Data Analysis and Data Visualization, I use Covid-19 Data which not easy understand and not easy to analysis so with Python create data to understand. also make dashboard which we can use in future for better understanding and prediction.

1.2 Project Scope

Using COVID-19 data to fight & contain the pandemic with advanced analytics is critical to protect public health & save lives. Using spatial data through Data analysis and Dashboard will allow us to beat Coronavirus faster because of better understanding of Data.

1.3 Objectives

The pandemic has already taken grip over peoples' life. Since the start of the pandemic, some countries are facing problem of ever-increasing cases. Through the data analysis of cases one can analyses how countries all over the world are doing in terms of controlling the pandemic. Analyzing data leads to adapt the prevention model of the countries that are doing great in terms of lowering the graph. Predictions are made with the dataset available to the individual/country/organizations, thus helping them to decide how far they are able to control the pandemic or up to how much extent they should guide preventive measures. Through this project, a step towards helping people to understand the spread and predict the cases in their country is done. This project also gives an insight of how a country is doing in terms of limiting the spread.

1.4 Literature Review

There are a lot of research papers published that are related to covid-19. Some of them to name can be research work related to vaccine or other medical drugs that can help to recover. Deep analysis is done on the people who recovered which can shed some light on how to deal with the active cases. Data scientists all over the world are busy in making sense out of the available data and predict the near future. Finding trend pattern, feature selection, forecasting techniques are being applied in and out to come to a conclusion. Rajan Gupta and Saibal Kumar Pal, in their research paper 'Trend analysis and forecasting of covid-19 break in India' used exploratory data analysis to report the situation in the time period of January to March in India. They use time series forecasting methods to predict the future trends. A very famous machine learning model- Arima model prediction was used and the inferred a result that predicted huge range of the number of likely covid-19 positive cases in April and May. The average that was forecasted was a detection of approximately 7000 patients in a span of 30 days in April. However in reality the figures were pretty higher. Another research paper by the department of CSE and IT of Northcap University, India in collaboration with Defence Research and Development Organization (DRDO) India also covered data from January 30 to March 30, 2020. They used regression models for forecasting. According to them, expected cases may rise to about 5000 in a two-week time period. This was far more accurate however actual scenario showed a bigger upsurge. This research paper can also be of help to several other sectors or other branches of healthcare as immunity power is very related to fighting with Covid-19.

2. System Requirement Study

2.1 User Characteristics

- Students: for educational purpose
- Researchers: to predict and planning according to data
- Government: with advanced analytics is critical to protect public health & save lives

2.2 Hardware and Software Characteristics

2.2.1 Hardware Characteristics

- **Memory (RAM):** At least 2 GB available, 4 GB or more recommended.
- **Hard Disk:** requires a minimum of 1 GB of available hard-disk space.
- **Display:** At least 1440x900 or 1600x900 (16:9) required. Lower resolutions such as 1024x768 or 1280x800 aren't supported, as certain controls (such as closing the start-up screen) display beyond those resolutions.
- **CPU:** 1 gigahertz (GHz) 64-bit (x64) processor or better recommended.
- **Processor**: i5 Processor or more

2.2.2 Software Characteristics

- **Operating System:** Windows 8.1 or later
- **.NET Framework:** 4.8
- **Browser:** Google Chrome 54 or later.
- **Language:** python
- **For Development:** Jupyter Notebook, PowerBI, MS Excel

3. System Analysis

3.1 Study of Current System

In this we 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, which consists of the data related to the cumulative number of confirmed cases, per day, in each Country. with dashboard try to address the large amounts of information in number terms and in visualization. journalists, non-governmental organizations and members of the communities and Others too may find this useful.

3.2 Requirement of this System

The User will require dataset which is a covid19 dataset on GitHub.
time_series_covid19_confirmed_global.csv
time_series_covid19_deaths_global.csv
time_series_covid19_recovered_global.csv

columns:

Province/State,
Country/Region,
Lat,
Long,
Dates

Pandas: pandas is a software library written for the Python programming language for data manipulation and analysis.

Data frame: Pandas Data frame is 2D, mutable and heterogeneous tabular data structure with labelled axes. Data frame can be made of more than one series (series can only contain single list with index.

4. Implementation and Planning

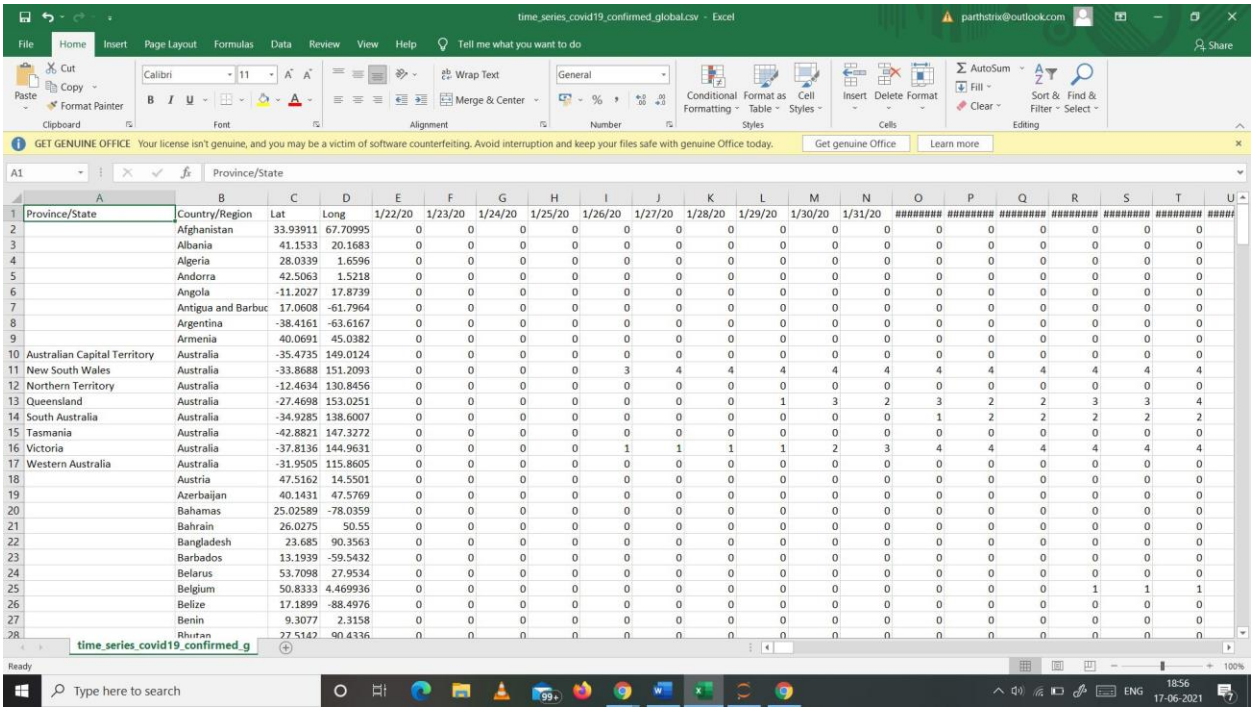
4.1 Technologies & Implementation Environment

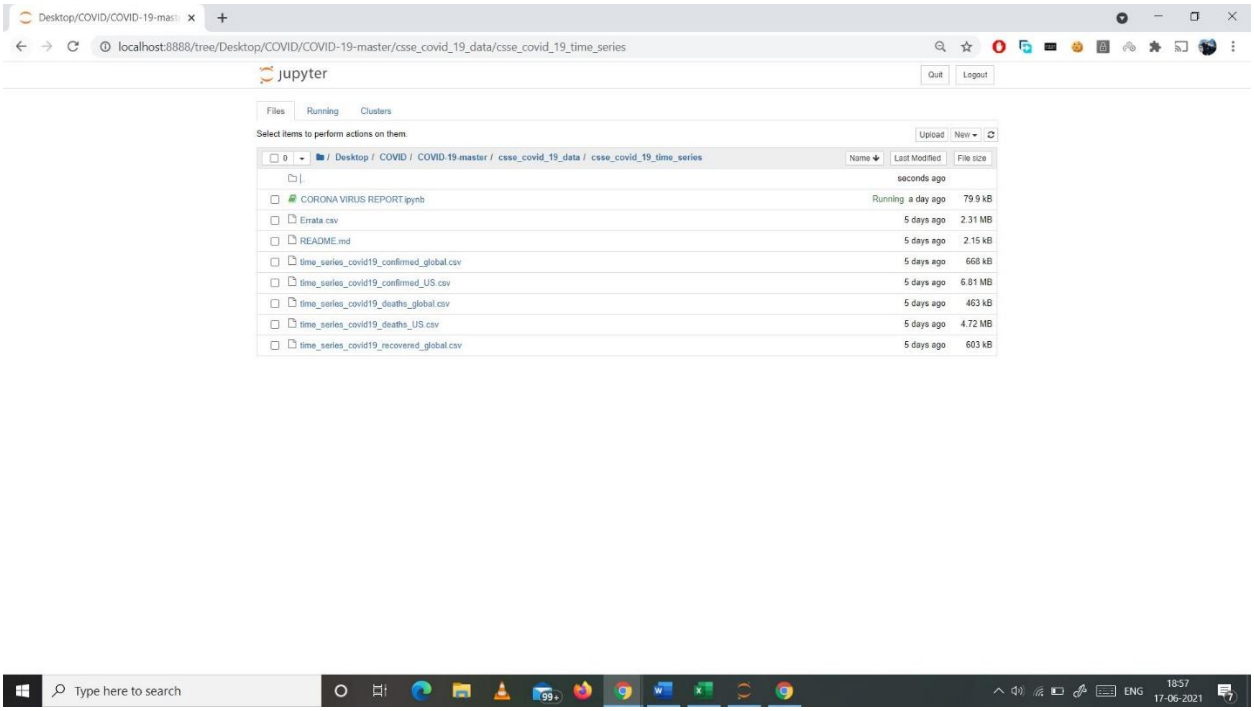
Technologies	Windows/Linux/Mac OS
Implementation Environment	Microsoft Excel Anaconda Navigator Jupyter Notebook Python 3 PowerBI

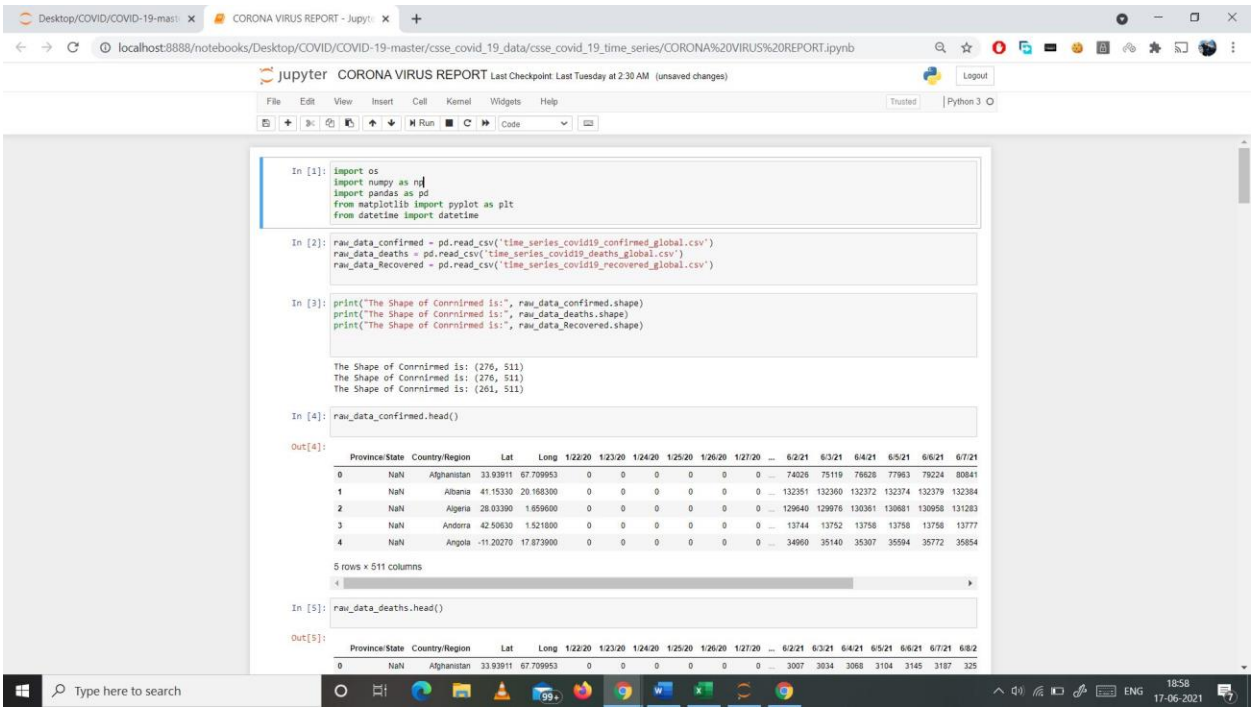
4.2 Program/Modules Specification:

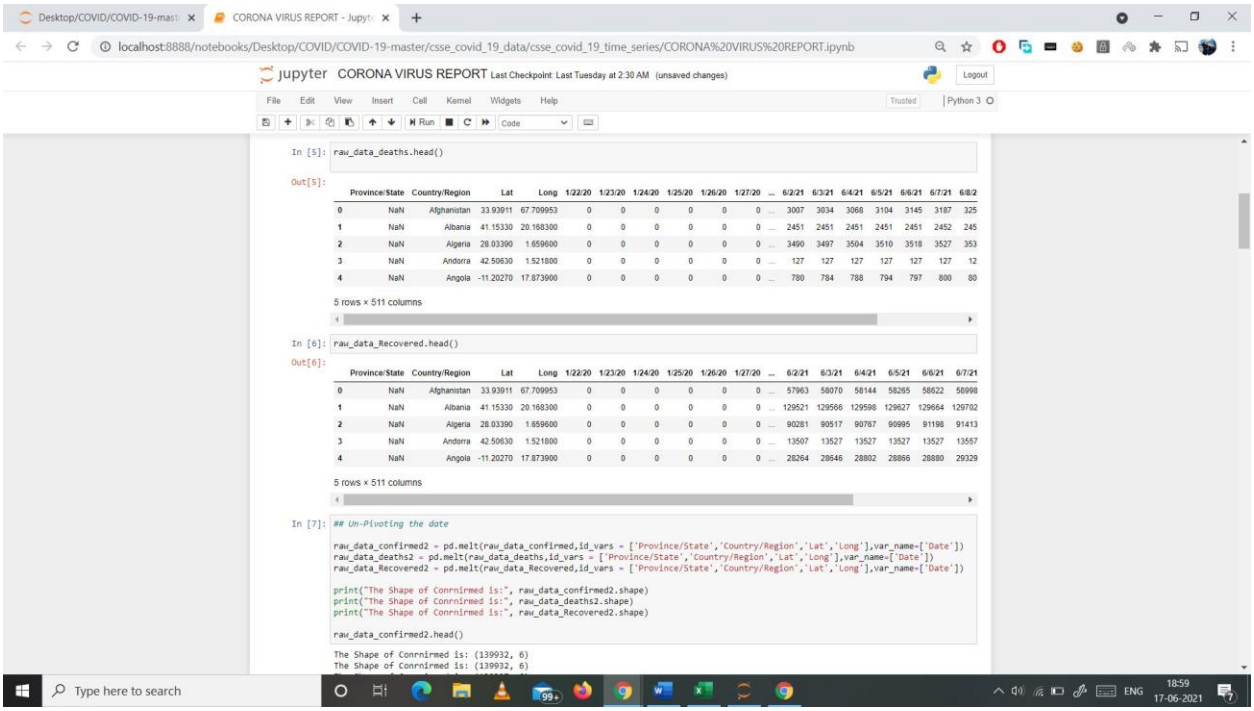
Data comes in many forms, all of it messy. Whether we're talking about missing data, unstructured data, or data that lacks regular structure, you need methods to cleanse data before you can process it to improve its quality. To cleaning data , In module un-pivot table and make column of date. After this count number of null values in every columns and try to remove this null values. Use Full Join to joint tables so we can work easily. And last create final table on daily base numbers of deaths , confirm case and recovered case. This final table extract to one file and use in PowerBI to create dashboard.

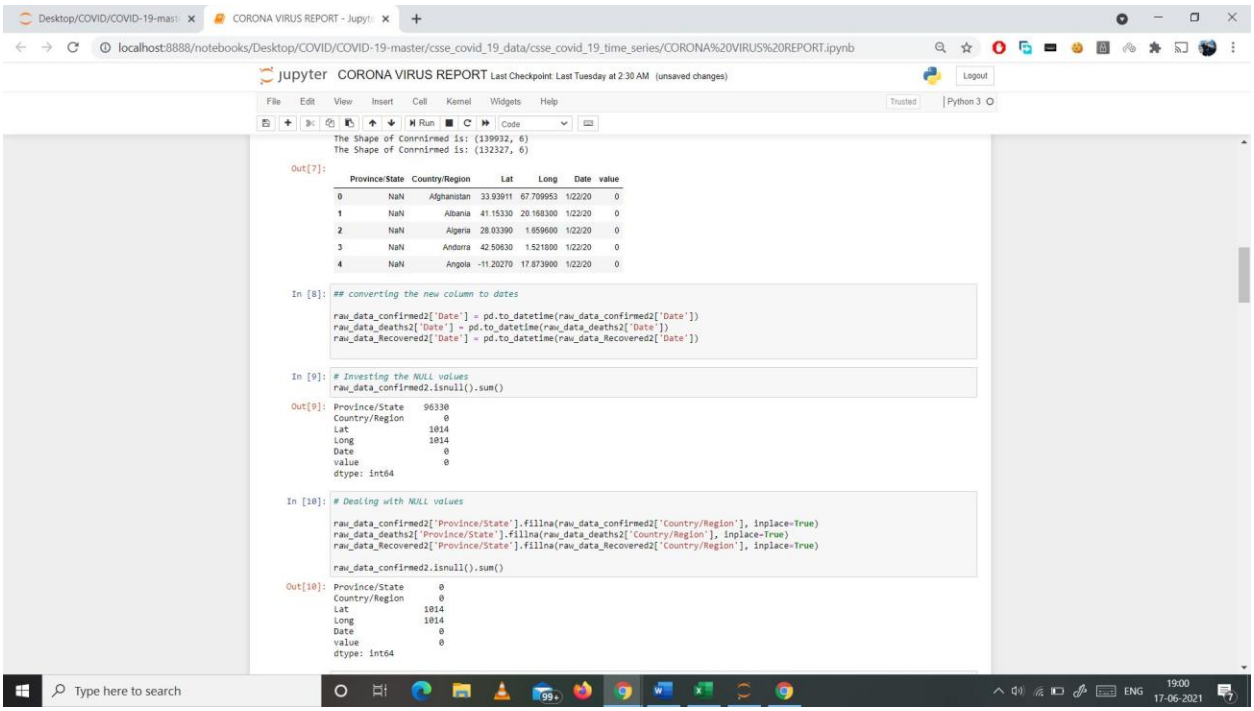
SCREENSHOTS:

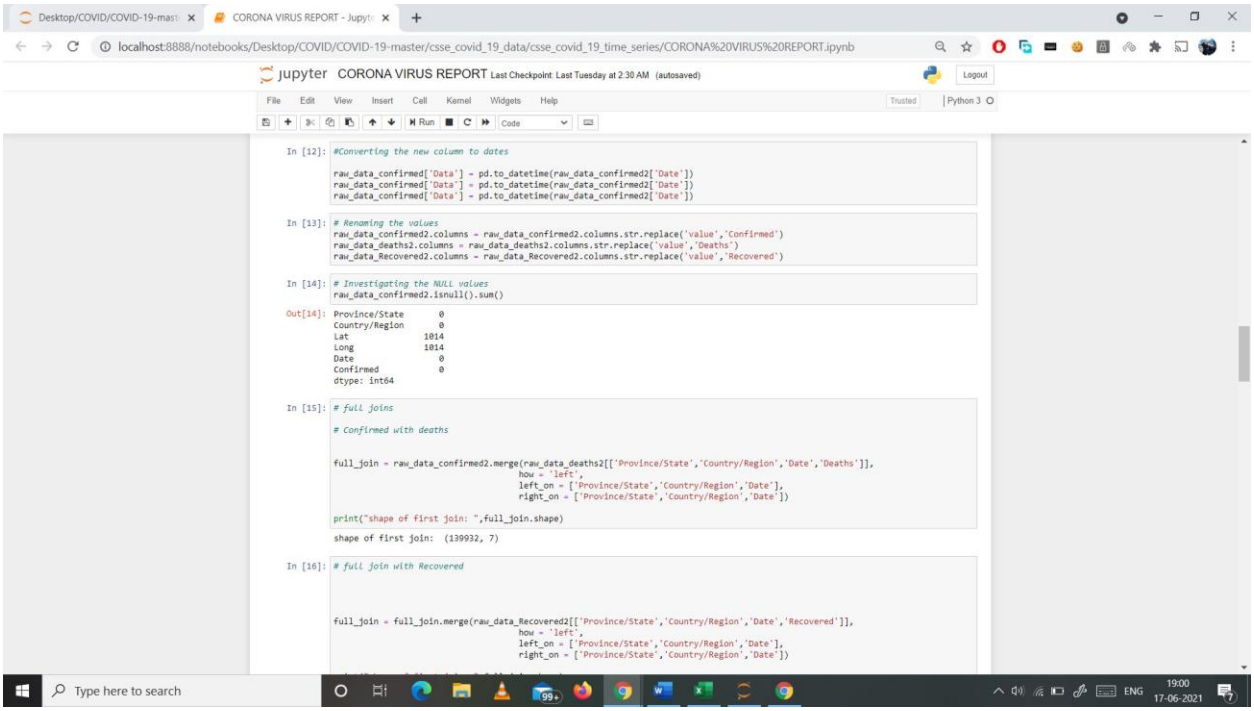


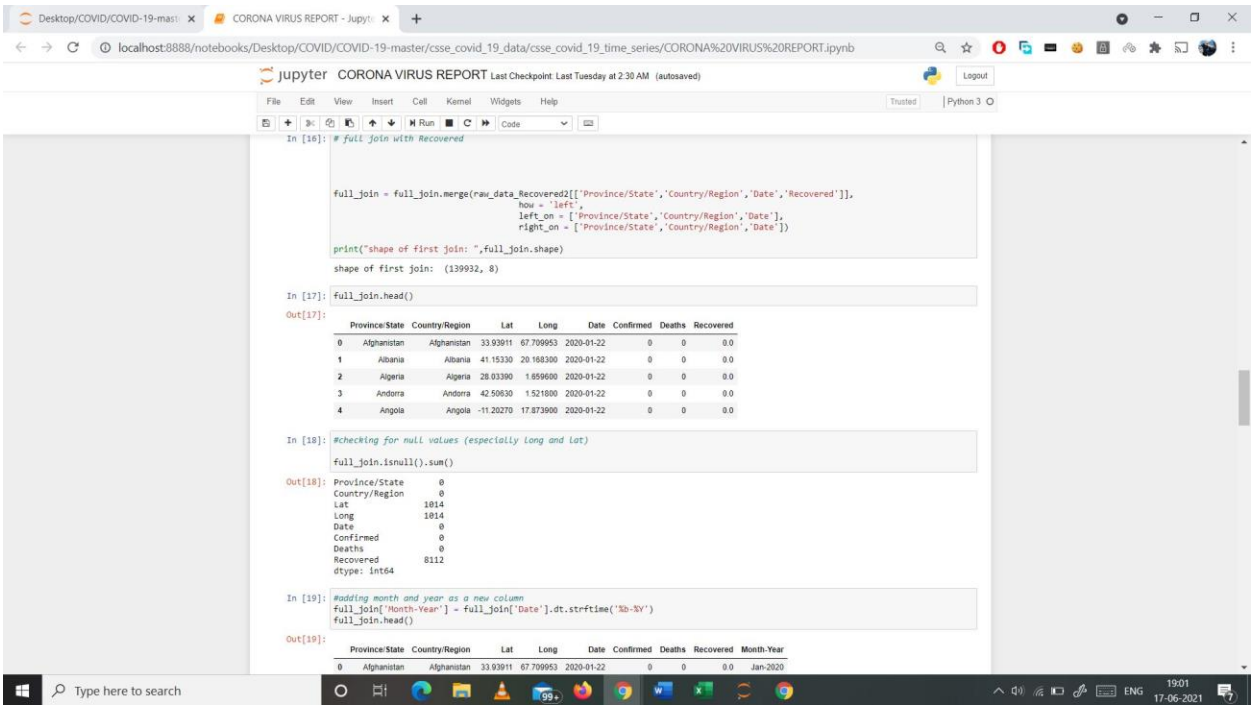


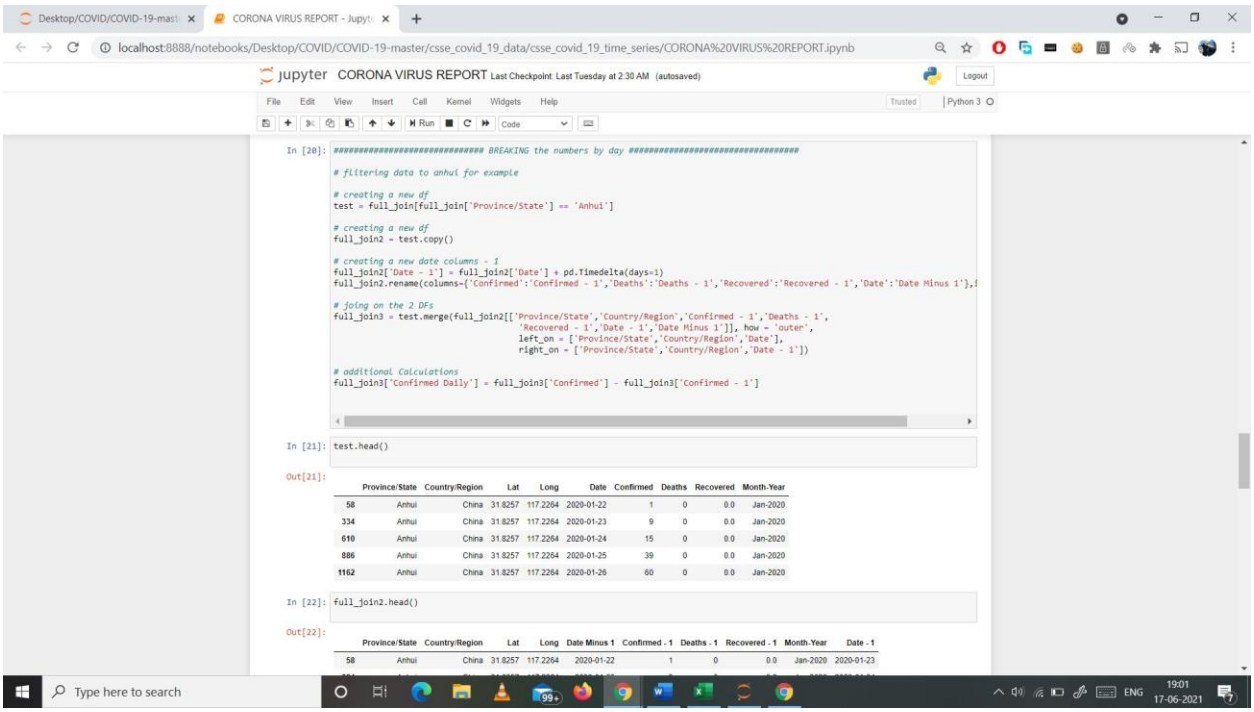

Desktop/COVID/COVID-19-mast
CORONA VIRUS REPORT - Jupyt
localhost:8888/notebooks/Desktop/COVID/COVID-19-master/csse_covid_19_data/csse_covid_19_time_series/CORONA%20VIRUS%20REPORT.ipynb
jupyter CORONA VIRUS REPORT Last Checkpoint: Last Tuesday at 2:30 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [1]:
import os
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from datetime import datetime
In [2]:
raw_data_confirmed = pd.read_csv('time_series_covid19_confirmed_global.csv')
raw_data_deaths = pd.read_csv('time_series_covid19_deaths_global.csv')
raw_data_Recovered = pd.read_csv('time_series_covid19_recovered_global.csv')
In [3]:
print("The Shape of Connrimed is:", raw_data_confirmed.shape)
print("The Shape of Connrimed is:", raw_data_deaths.shape)
print("The Shape of Connrimed is:", raw_data_Recovered.shape)
The Shape of Connrimed is: (276, 511)
The Shape of Connrimed is: (276, 511)
The Shape of Connrimed is: (261, 511)
In [4]:
raw_data_confirmed.head()
Out[4]:
Province/State Country/Region Lat Long 1/22/20 1/23/20 1/24/20 1/25/20 1/26/20 1/27/20 ... 6/2/21 6/3/21 6/4/21 6/5/21 6/6/21 6/7/21
0 NaN Afghanistan 33.93911 67.709953 0 0 0 0 0 0 ... 74026 75119 76628 77963 79224 80841
1 NaN Albania 41.15330 20.168300 0 0 0 0 0 0 ... 132351 132360 132372 132374 132379 132384
2 NaN Algeria 28.03390 1.659600 0 0 0 0 0 0 ... 129640 129976 130361 130681 130958 131283
3 NaN Andorra 42.50630 1.521800 0 0 0 0 0 0 ... 13744 13752 13758 13758 13758 13777
4 NaN Angola -11.20270 17.873900 0 0 0 0 0 0 ... 34960 35140 35307 35594 35772 35854
5 rows × 511 columns
In [5]:
raw_data_deaths.head()
Out[5]:
Province/State Country/Region Lat Long 1/22/20 1/23/20 1/24/20 1/25/20 1/26/20 1/27/20 ... 6/2/21 6/3/21 6/4/21 6/5/21 6/6/21 6/7/21 6/8/2
0 NaN Afghanistan 33.93911 67.709953 0 0 0 0 0 0 ... 3007 3034 3068 3104 3145 3187 325
Type here to search
ENG
18:58
17-06-2021

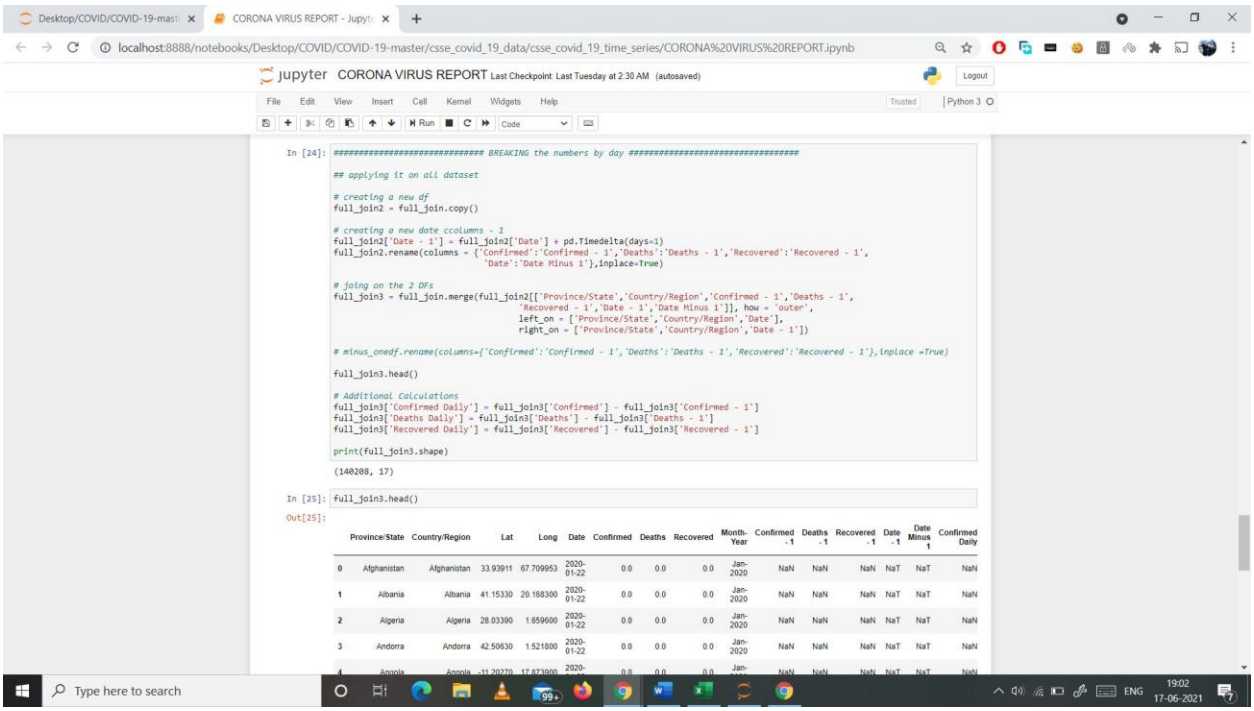

Desktop/COVID/COVID-19-mast
CORONA VIRUS REPORT - Jupyt
localhost:8888/notebooks/Desktop/COVID/COVID-19-master/csse_covid_19_data/csse_covid_19_time_series/CORONA%20VIRUS%20REPORT.ipynb
jupyter CORONA VIRUS REPORT Last Checkpoint: Last Tuesday at 2:30 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [5]:
raw_data_deaths.head()
Out[5]:
Province/State Country/Region Lat Long 1/22/20 1/23/20 1/24/20 1/25/20 1/26/20 1/27/20 ... 6/2/21 6/3/21 6/4/21 6/5/21 6/6/21 6/7/21 6/8/2
0 NaN Afghanistan 33.93911 67.709953 0 0 0 0 0 0 ... 3007 3034 3068 3104 3145 3187 325
1 NaN Albania 41.15330 20.168300 0 0 0 0 0 0 ... 2451 2451 2451 2451 2451 2452 245
2 NaN Algeria 28.03390 1.659600 0 0 0 0 0 0 ... 3490 3497 3504 3510 3518 3527 353
3 NaN Andorra 42.50630 1.521800 0 0 0 0 0 0 ... 127 127 127 127 127 127 12
4 NaN Angola -11.20270 17.873900 0 0 0 0 0 0 ... 780 784 788 794 797 800 80
5 rows × 511 columns
In [6]:
raw_data_Recovered.head()
Out[6]:
Province/State Country/Region Lat Long 1/22/20 1/23/20 1/24/20 1/25/20 1/26/20 1/27/20 ... 6/2/21 6/3/21 6/4/21 6/5/21 6/6/21 6/7/21
0 NaN Afghanistan 33.93911 67.709953 0 0 0 0 0 0 ... 57963 58070 58144 58265 58622 58998
1 NaN Albania 41.15330 20.168300 0 0 0 0 0 0 ... 129521 129566 129598 129627 129664 129702
2 NaN Algeria 28.03390 1.659600 0 0 0 0 0 0 ... 90281 90517 90767 90995 91198 91413
3 NaN Andorra 42.50630 1.521800 0 0 0 0 0 0 ... 13507 13527 13527 13527 13527 13557
4 NaN Angola -11.20270 17.873900 0 0 0 0 0 0 ... 28264 28646 28802 28866 28880 29329
5 rows × 511 columns
In [7]:
Un-Pivoting the date
raw_data_confirmed2 = pd.melt(raw_data_confirmed,id_vars = ['Province/State','Country/Region','Lat','Long'],var_name=['Date'])
raw_data_deaths2 = pd.melt(raw_data_deaths,id_vars = ['Province/State','Country/Region','Lat','Long'],var_name=['Date'])
raw_data_Recovered2 = pd.melt(raw_data_Recovered,id_vars = ['Province/State','Country/Region','Lat','Long'],var_name=['Date'])
print("The Shape of Connrimed is:", raw_data_confirmed2.shape)
print("The Shape of Connrimed is:", raw_data_deaths2.shape)
print("The Shape of Connrimed is:", raw_data_Recovered2.shape)
raw_data_confirmed2.head()
The Shape of Connrimed is: (139932, 6)
The Shape of Connrimed is: (139932, 6)
Type here to search
ENG
18:59
17-06-2021


Desktop/COVID/COVID-19-mast
CORONA VIRUS REPORT - Jupyt
localhost:8888/notebooks/Desktop/COVID/COVID-19-master/csse_covid_19_data/csse_covid_19_time_series/CORONA%20VIRUS%20REPORT.ipynb
jupyter CORONA VIRUS REPORT Last Checkpoint: Last Tuesday at 2:30 AM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
The Shape of Conrnirmed is: (139932, 6)
The Shape of Conrnirmed is: (132327, 6)
Out[7]:
Province/State Country/Region Lat Long Date value
0 NaN Afghanistan 33.93911 67.709953 1/22/20 0
1 NaN Albania 41.15330 20.168300 1/22/20 0
2 NaN Algeria 28.03390 1.659600 1/22/20 0
3 NaN Andorra 42.50630 1.521800 1/22/20 0
4 NaN Angola -11.20270 17.873900 1/22/20 0
In [8]:
converting the new column to dates
raw_data_confirmed2['Date'] = pd.to_datetime(raw_data_confirmed2['Date'])
raw_data_deaths2['Date'] = pd.to_datetime(raw_data_deaths2['Date'])
raw_data_Recovered2['Date'] = pd.to_datetime(raw_data_Recovered2['Date'])
In [9]:
# Investing the NULL values
raw_data_confirmed2.isnull().sum()
Out[9]:
Province/State 96330
Country/Region 0
Lat 1014
Long 1014
Date 0
value 0
dtype: int64
In [10]:
Dealing with NULL values
raw_data_confirmed2['Province/State'].fillna(raw_data_confirmed2['Country/Region'], inplace=True)
raw_data_deaths2['Province/State'].fillna(raw_data_deaths2['Country/Region'], inplace=True)
raw_data_Recovered2['Province/State'].fillna(raw_data_Recovered2['Country/Region'], inplace=True)
raw_data_confirmed2.isnull().sum()
Out[10]:
Province/State 0
Country/Region 0
Lat 1014
Long 1014
Date 0
value 0
dtype: int64
Type here to search
ENG
19:00
17-06-2021


Desktop/COVID/COVID-19-mast
CORONA VIRUS REPORT - Jupyt
localhost:8888/notebooks/Desktop/COVID/COVID-19-master/csse_covid_19_data/csse_covid_19_time_series/CORONA%20VIRUS%20REPORT.ipynb
jupyter CORONA VIRUS REPORT Last Checkpoint: Last Tuesday at 2:30 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [12]:
#Converting the new column to dates
raw_data_confirmed['Data'] = pd.to_datetime(raw_data_confirmed2['Date'])
raw_data_confirmed['Data'] = pd.to_datetime(raw_data_confirmed2['Date'])
raw_data_confirmed['Data'] = pd.to_datetime(raw_data_confirmed2['Date'])
In [13]:
# Renaming the values
raw_data_confirmed2.columns = raw_data_confirmed2.columns.str.replace('value','Confirmed')
raw_data_deaths2.columns = raw_data_deaths2.columns.str.replace('value','Deaths')
raw_data_Recovered2.columns = raw_data_Recovered2.columns.str.replace('value','Recovered')
In [14]:
# Investigating the NULL values
raw_data_confirmed2.isnull().sum()
Out[14]:
Province/State 0
Country/Region 0
Lat 1014
Long 1014
Date 0
Confirmed 0
dtype: int64
In [15]:
# full joins
# Confirmed with deaths
full_join = raw_data_confirmed2.merge(raw_data_deaths2[['Province/State','Country/Region','Date','Deaths']],
how = 'left',
left_on = ['Province/State','Country/Region','Date'],
right_on = ['Province/State','Country/Region','Date'])
print("shape of first join: ",full_join.shape)
shape of first join: (139932, 7)
In [16]:
# full join with Recovered
full_join = full_join.merge(raw_data_Recovered2[['Province/State','Country/Region','Date','Recovered']],
how = 'left',
left_on = ['Province/State','Country/Region','Date'],
right_on = ['Province/State','Country/Region','Date'])
Type here to search
ENG
19:00
17-06-2021


Desktop/COVID/COVID-19-mast
CORONA VIRUS REPORT - Jupyt
localhost:8888/notebooks/Desktop/COVID/COVID-19-master/csse_covid_19_data/csse_covid_19_time_series/CORONA%20VIRUS%20REPORT.ipynb
jupyter CORONA VIRUS REPORT Last Checkpoint: Last Tuesday at 2:30 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [16]:
# full join with Recovered
full_join = full_join.merge(raw_data_Recovered2[['Province/State','Country/Region','Date','Recovered']],
how = 'left',
left_on = ['Province/State','Country/Region','Date'],
right_on = ['Province/State','Country/Region','Date'])
print("shape of first join: ",full_join.shape)
shape of first join: (139932, 8)
In [17]:
full_join.head()
Out[17]:
Province/State Country/Region Lat Long Date Confirmed Deaths Recovered
0 Afghanistan Afghanistan 33.93911 67.709953 2020-01-22 0 0 0.0
1 Albania Albania 41.15330 20.168300 2020-01-22 0 0 0.0
2 Algeria Algeria 28.03390 1.659600 2020-01-22 0 0 0.0
3 Andorra Andorra 42.50630 1.521800 2020-01-22 0 0 0.0
4 Angola Angola -11.20270 17.873900 2020-01-22 0 0 0.0
In [18]:
#checking for null values (especially Long and Lat)
full_join.isnull().sum()
Out[18]:
Province/State 0
Country/Region 0
Lat 1014
Long 1014
Date 0
Confirmed 0
Deaths 0
Recovered 8112
dtype: int64
In [19]:
#adding month and year as a new column
full_join['Month-Year'] = full_join['Date'].dt.strftime('%b-%Y')
full_join.head()
Out[19]:
Province/State Country/Region Lat Long Date Confirmed Deaths Recovered Month-Year
0 Afghanistan Afghanistan 33.93911 67.709953 2020-01-22 0 0 0.0 Jan-2020
Type here to search
ENG
19:01
17-06-2021


Desktop/COVID/COVID-19-mast
CORONA VIRUS REPORT - Jupyt
localhost:8888/notebooks/Desktop/COVID/COVID-19-master/csse_covid_19_data/csse_covid_19_time_series/CORONA%20VIRUS%20REPORT.ipynb
jupyter CORONA VIRUS REPORT Last Checkpoint: Last Tuesday at 2:30 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [20]:
################################ BREAKING the numbers by day ##################################
# filtering data to anhui for example
# creating a new df
test = full_join[full_join['Province/State'] == 'Anhui']
# creating a new df
full_join2 = test.copy()
# creating a new date columns - 1
full_join2['Date - 1'] = full_join2['Date'] + pd.Timedelta(days=1)
full_join2.rename(columns={'Confirmed':'Confirmed - 1','Deaths':'Deaths - 1','Recovered':'Recovered - 1','Date':'Date Minus 1'},i
# joing on the 2 DFs
full_join3 = test.merge(full_join2[['Province/State','Country/Region','Confirmed - 1','Deaths - 1',
'Recovered - 1','Date - 1','Date Minus 1']], how = 'outer',
left_on = ['Province/State','Country/Region','Date'],
right_on = ['Province/State','Country/Region','Date - 1'])
# additional Calculations
full_join3['Confirmed Daily'] = full_join3['Confirmed'] - full_join3['Confirmed - 1']
In [21]:
test.head()
Out[21]:
Province/State Country/Region Lat Long Date Confirmed Deaths Recovered Month-Year
58 Anhui China 31.8257 117.2264 2020-01-22 1 0 0.0 Jan-2020
334 Anhui China 31.8257 117.2264 2020-01-23 9 0 0.0 Jan-2020
610 Anhui China 31.8257 117.2264 2020-01-24 15 0 0.0 Jan-2020
886 Anhui China 31.8257 117.2264 2020-01-25 39 0 0.0 Jan-2020
1162 Anhui China 31.8257 117.2264 2020-01-26 60 0 0.0 Jan-2020
In [22]:
full_join2.head()
Out[22]:
Province/State Country/Region Lat Long Date Minus 1 Confirmed - 1 Deaths - 1 Recovered - 1 Month-Year Date - 1
58 Anhui China 31.8257 117.2264 2020-01-22 1 0 0.0 Jan-2020 2020-01-23
Type here to search
ENG
19:01
17-06-2021



jupyter CORONA VIRUS REPORT Last Checkpoint: Last Tuesday at 2:30 AM (autosaved)

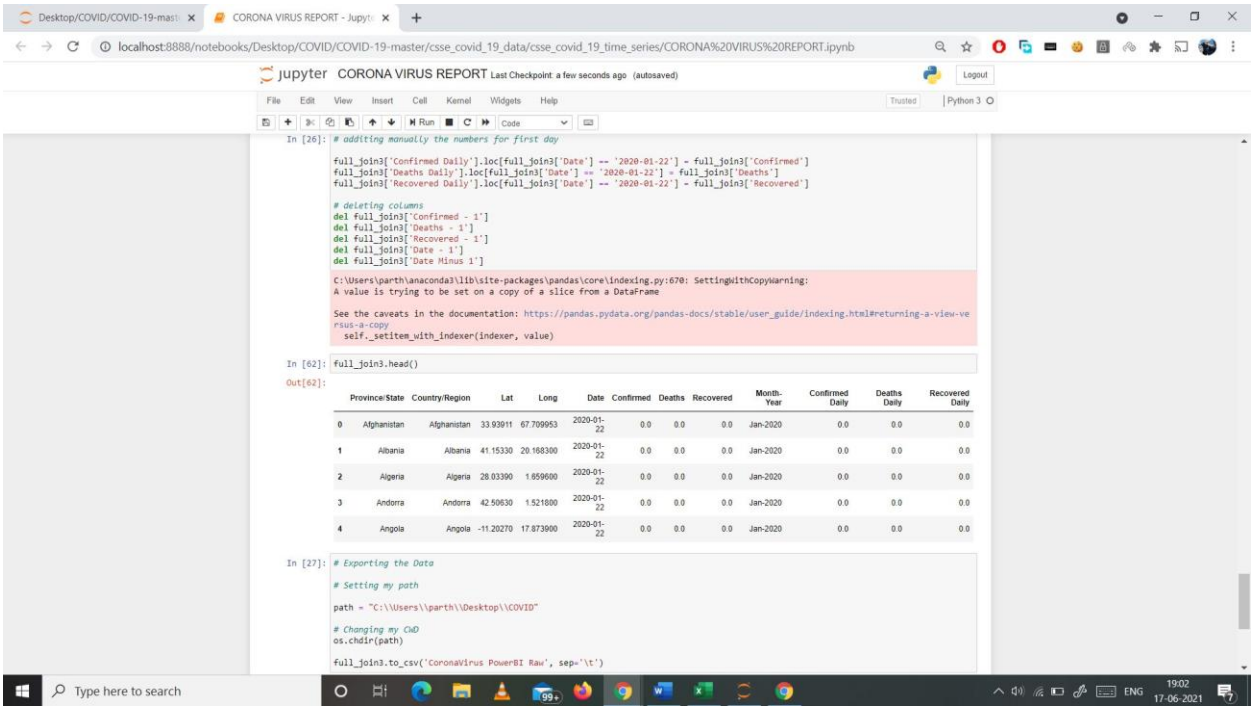
```
In [24]:
############################## BREAKING the numbers by day ################################

## applying it on all dataset

# creating a new df
full_join2 = full_join.copy()

# creating a new date ccolumns - 1
full_join2['Date - 1'] = full_join2['Date'] + pd.Timedelta(days=1)
full_join2.rename(columns = {'Confirmed':'Confirmed - 1','Deaths':'Deaths - 1','Recovered':'Recovered - 1',
                             'Date':'Date Minus 1'},inplace=True)


# joing on the 2 DFs
full_join3 = full_join.merge(full_join2[['Province/State','Country/Region','Confirmed - 1','Deaths - 1',
                                         'Recovered - 1','Date - 1','Date Minus 1']], how = 'outer',
                                        left_on = ['Province/State','Country/Region','Date'],
                                        right_on = ['Province/State','Country/Region','Date - 1'])

# minus_onedf.rename(columns={'Confirmed':'Confirmed - 1','Deaths':'Deaths - 1','Recovered':'Recovered - 1'},inplace =True)

full_join3.head()

# Additional Calculations
full_join3['Confirmed Daily'] = full_join3['Confirmed'] - full_join3['Confirmed - 1']
full_join3['Deaths Daily'] = full_join3['Deaths'] - full_join3['Deaths - 1']
full_join3['Recovered Daily'] = full_join3['Recovered'] - full_join3['Recovered - 1']

print(full_join3.shape)
```

```
(140208, 17)
```

```
In [25]: full_join3.head()
```

Out[25]:

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered	Month-Year	Confirmed - 1	Deaths - 1	Recovered - 1	Date - 1	Date Minus 1	Confirmed Daily
0	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-22	0.0	0.0	0.0	Jan-2020	NaN	NaN	NaN	NaT	NaT	NaN
1	Albania	Albania	41.15330	20.168300	2020-01-22	0.0	0.0	0.0	Jan-2020	NaN	NaN	NaN	NaT	NaT	NaN
2	Algeria	Algeria	28.03390	1.659600	2020-01-22	0.0	0.0	0.0	Jan-2020	NaN	NaN	NaN	NaT	NaT	NaN
3	Andorra	Andorra	42.50630	1.521800	2020-01-22	0.0	0.0	0.0	Jan-2020	NaN	NaN	NaN	NaT	NaT	NaN
4	Angola	Angola	-11.20270	17.873900	2020-01-22	0.0	0.0	0.0	Jan-2020	NaN	NaN	NaN	NaT	NaT	NaN



jupyter CORONA VIRUS REPORT Last Checkpoint: a few seconds ago (autosaved)

```
In [26]: # additing manually the numbers for first day

full_join3['Confirmed Daily'].loc[full_join3['Date'] == '2020-01-22'] = full_join3['Confirmed']
full_join3['Deaths Daily'].loc[full_join3['Date'] == '2020-01-22'] = full_join3['Deaths']
full_join3['Recovered Daily'].loc[full_join3['Date'] == '2020-01-22'] = full_join3['Recovered']

# deleting columns
del full_join3['Confirmed - 1']
del full_join3['Deaths - 1']
del full_join3['Recovered - 1']
del full_join3['Date - 1']
del full_join3['Date Minus 1']
```

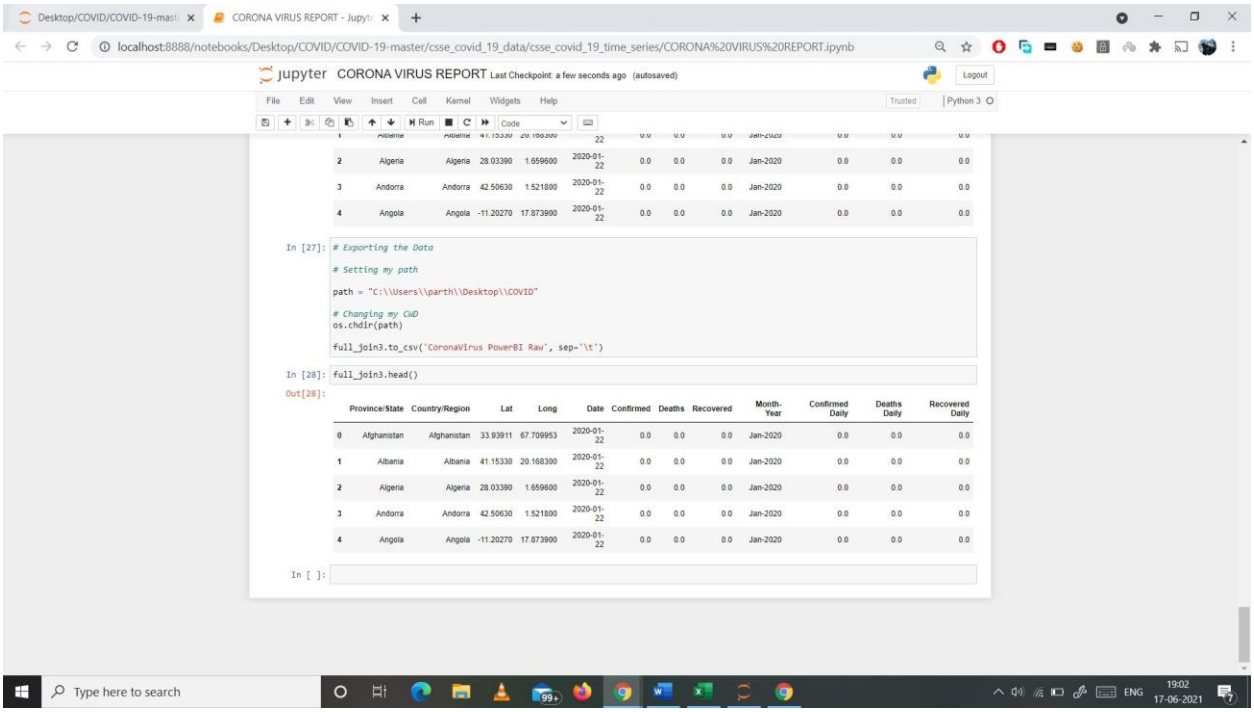
```
C:\Users\parth\anaconda3\lib\site-packages\pandas\core\indexing.py:670: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  self._setitem_with_indexer(indexer, value)
```

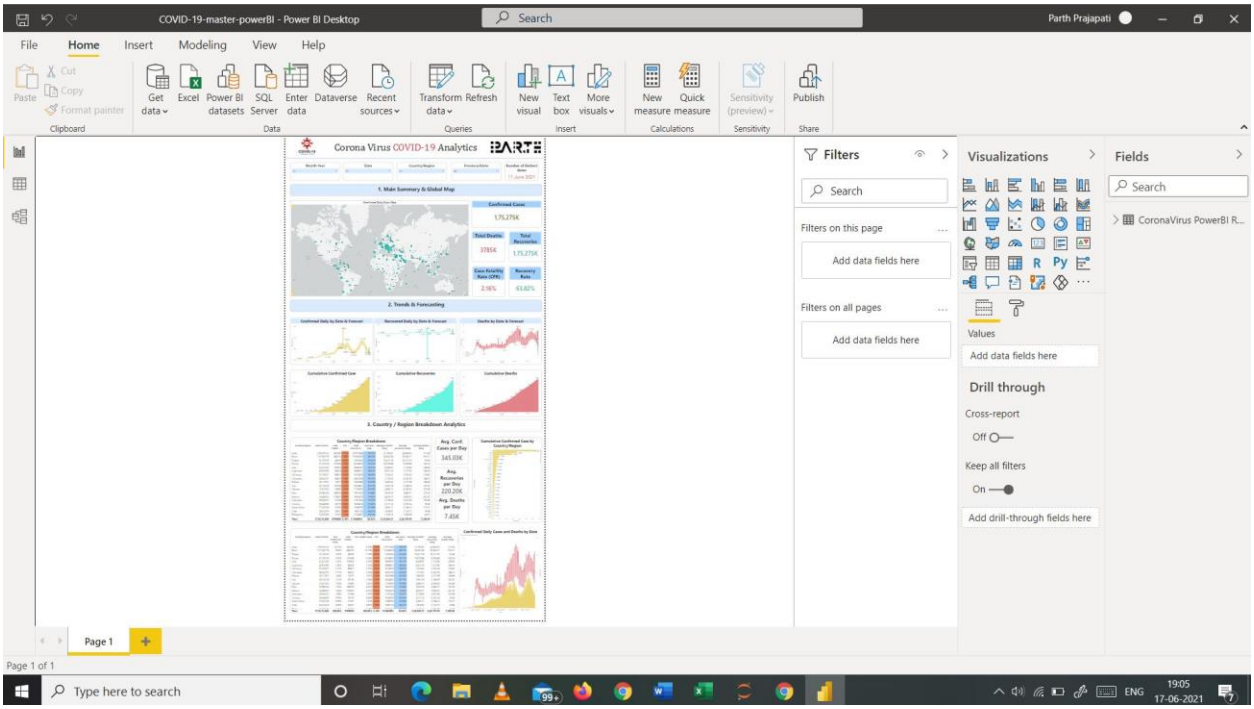
```
In [62]: full_join3.head()
```

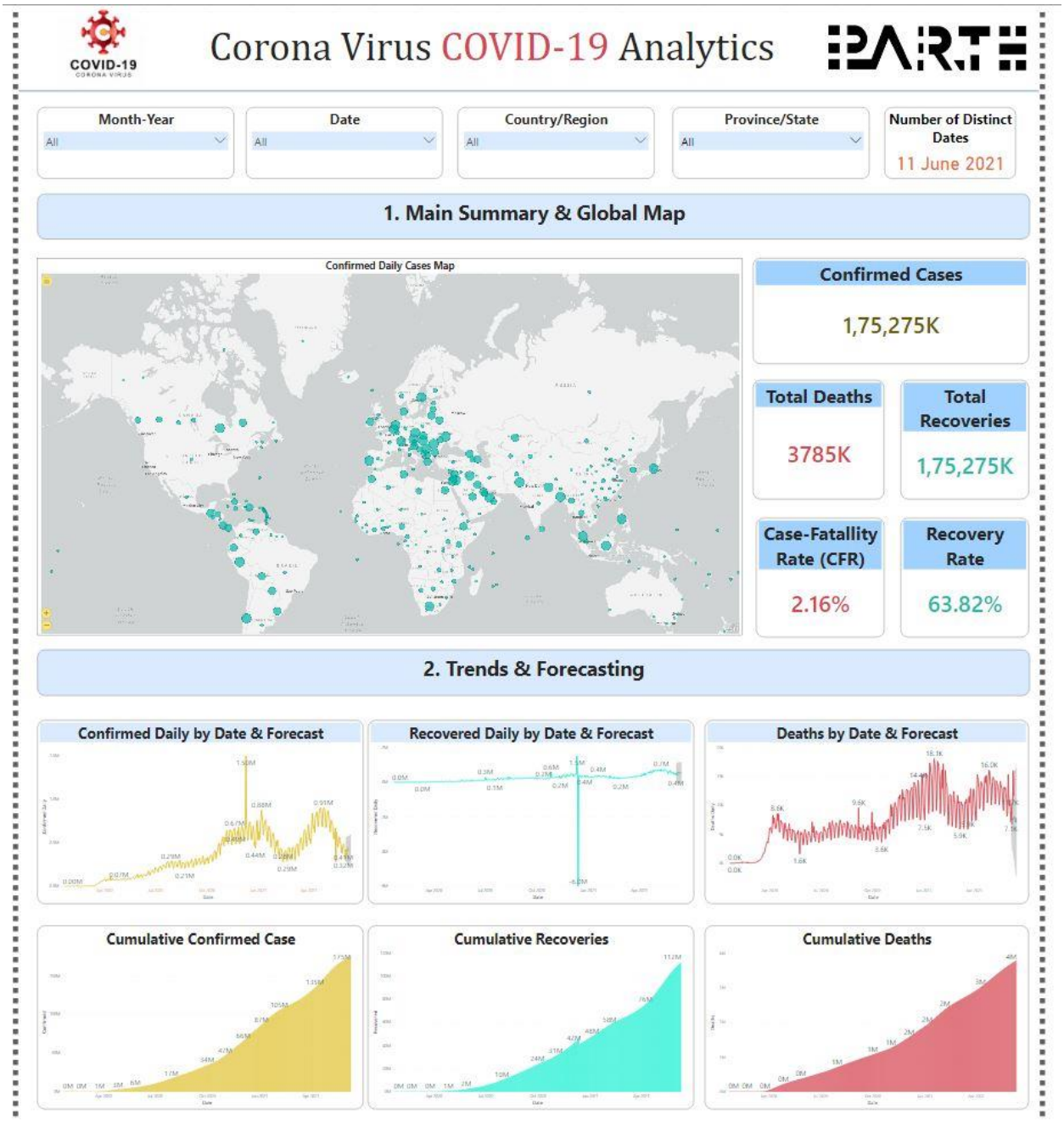
Out[62]:

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered	Month-Year	Confirmed Daily	Deaths Daily	Recovered Daily
0	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-22	0.0	0.0	0.0	Jan-2020	0.0	0.0	0.0
1	Albania	Albania	41.15330	20.168300	2020-01-22	0.0	0.0	0.0	Jan-2020	0.0	0.0	0.0
2	Algeria	Algeria	28.03390	1.659600	2020-01-22	0.0	0.0	0.0	Jan-2020	0.0	0.0	0.0
3	Andorra	Andorra	42.50630	1.521800	2020-01-22	0.0	0.0	0.0	Jan-2020	0.0	0.0	0.0
4	Angola	Angola	-11.20270	17.873900	2020-01-22	0.0	0.0	0.0	Jan-2020	0.0	0.0	0.0

```
In [27]: # Exporting the Data

# Setting my path

path = "C:\\Users\\parth\\Desktop\\COVID"

# Changing my CWD
os.chdir(path)

full_join3.to_csv('CoronaVirus PowerBI Raw', sep='\t')
```


CORONA VIRUS REPORT - Jupyter
localhost:8888/notebooks/Desktop/COVID/COVID-19-master/csse_covid_19_data/csse_covid_19_time_series/CORONA%20VIRUS%20REPORT.ipynb
jupyter CORONA VIRUS REPORT Last Checkpoint: a few seconds ago (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Logout
Trusted
Python 3
In [27]:
# Exporting the Data
# Setting my path
path = "C:\\Users\\parth\\Desktop\\COVID"
# Changing my CWD
os.chdir(path)
full_join3.to_csv('CoronaVirus PowerBI Raw', sep='\t')
In [28]:
full_join3.head()
Out[28]:
Province/State Country/Region Lat Long Date Confirmed Deaths Recovered Month-Year Confirmed Daily Deaths Daily Recovered Daily
0 Afghanistan Afghanistan 33.93911 67.709953 2020-01-22 0.0 0.0 0.0 Jan-2020 0.0 0.0 0.0
1 Albania Albania 41.15330 20.168300 2020-01-22 0.0 0.0 0.0 Jan-2020 0.0 0.0 0.0
2 Algeria Algeria 28.03390 1.659600 2020-01-22 0.0 0.0 0.0 Jan-2020 0.0 0.0 0.0
3 Andorra Andorra 42.50630 1.521800 2020-01-22 0.0 0.0 0.0 Jan-2020 0.0 0.0 0.0
4 Angola Angola -11.20270 17.873900 2020-01-22 0.0 0.0 0.0 Jan-2020 0.0 0.0 0.0
In [ ]:
Type here to search
ENG
19:02
17-06-2021


COVID-19-master-powerBI - Power BI Desktop
Search
Parth Prajapati
File Home Insert Modeling View Help
Paste Cut Copy Format painter
Clipboard
Get data Excel Power BI datasets SQL Server Enter data Dataverse Recent sources
Data
Transform data Refresh
Queries
New visual Text box More visuals
Insert
New measure Quick measure
Calculations
Sensitivity (preview)
Sensitivity
Publish
Share
Corona Virus COVID-19 Analytics
Filters
Search
Filters on this page
Add data fields here
Filters on all pages
Add data fields here
Visualizations
Values
Add data fields here
Drill through
Cross-report
Off
Keep all filters
On
Add drill-through fields here
Fields
Search
CoronaVirus PowerBI R...
Page 1
Page 1 of 1
Type here to search
ENG
19:05
17-06-2021


COVID-19
Corona Virus COVID-19 Analytics
Month-Year
All
Date
All
Country/Region
All
Province/State
All
Number of Distinct Dates
11 June 2021
1. Main Summary & Global Map
Confirmed Daily Cases Map
Confirmed Cases
1,75,275K
Total Deaths
3785K
Total Recoveries
1,75,275K
Case-Fatallity Rate (CFR)
2.16%
Recovery Rate
63.82%
2. Trends & Forecasting
Confirmed Daily by Date & Forecast
Recovered Daily by Date & Forecast
Deaths by Date & Forecast
Cumulative Confirmed Case
Cumulative Recoveries
Cumulative Deaths

3. Country / Region Breakdown Analytics

Country/Region Breakdown

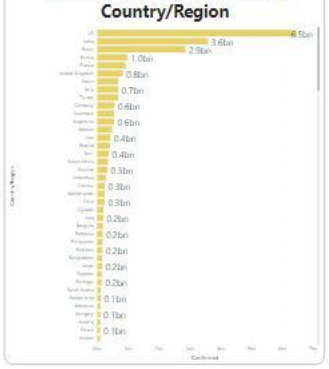
Country/Region	Total Confirm	Total Deaths	CFR	Total Recoveries	Recovery Rate	Average Confirm Daily	Average Recoveries Daily	Average Deaths Daily
India	2,93,59,155	367081	1.25%	27911384	95.07%	57,793.61	54,943.67	722.60
Brazil	1,72,96,118	484235	2.80%	15260459	88.23%	34,047.48	30,040.27	953.22
Turkey	53,19,359	48593	0.91%	5192945	97.62%	10,471.18	10,222.33	95.66
Russia	51,20,578	123568	2.41%	4726870	92.31%	10,079.88	9,304.86	243.24
Italy	42,41,760	126924	2.99%	3949597	93.11%	8,349.92	7,774.80	249.85
Argentina	40,93,090	84628	2.07%	3668672	89.63%	8,057.26	7,221.80	166.59
Germany	37,20,811	89821	2.41%	3574000	96.05%	7,324.43	7,035.43	176.81
Colombia	36,94,707	94615	2.56%	3435109	92.97%	7,273.05	6,762.03	186.25
Poland	28,77,007	74515	2.59%	2647688	92.03%	5,663.40	5,211.98	146.68
Iran	30,13,078	81796	2.71%	2624802	87.11%	5,931.26	5,166.93	161.02
Ukraine	22,81,303	53683	2.35%	2176093	95.39%	4,490.75	4,283.65	105.68
Peru	19,98,056	188100	9.41%	1955350	97.86%	3,933.18	3,849.11	370.28
Mexico	24,48,820	229823	9.39%	1950419	79.65%	4,820.51	3,839.41	452.41
Indonesia	18,94,025	52566	2.78%	1735144	91.61%	3,728.40	3,415.64	103.48
Czechia	16,64,839	30219	1.82%	1628416	97.81%	3,277.24	3,205.54	59.49
South Africa	17,30,106	57592	3.33%	1598293	92.38%	3,405.72	3,146.25	113.37
Chile	14,61,419	30472	2.09%	1382124	94.57%	2,876.81	2,720.72	59.98
Philippines	13,00,349	22507	1.73%	1216497	93.55%	2,559.74	2,394.68	44.31
Total	**17,52,74,868**	**3785050**	**2.16%**	**111860581**	**63.82%**	**3,45,029.27**	**2,20,197.99**	**7,450.89**

Avg. Conf. Cases per Day
345.03K

Avg. Recoveries per Day
220.20K

Avg. Deaths per Day
7.45K

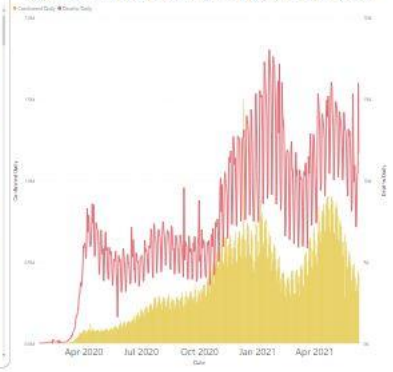
Cumulative Confirmed Case by Country/Region



Country/Region Breakdown

Country/Region	Total Confirm	%GT Confirmed Daily	Total Deaths	%GT Deaths Daily	CFR	Total Recoveries	Recovery Rate	Average Confirm Daily	Average Recoveries Daily	Average Deaths Daily
India	2,93,59,155	16.75%	367081	9.70%	1.25%	27911384	95.07%	57,793.61	54,943.67	722.60
Brazil	1,72,96,118	9.87%	484235	12.79%	2.80%	15260459	88.23%	34,047.48	30,040.27	953.22
Turkey	53,19,359	3.03%	48593	1.28%	0.91%	5192945	97.62%	10,471.18	10,222.33	95.66
Russia	51,20,578	2.92%	123568	3.26%	2.41%	4726870	92.31%	10,079.88	9,304.86	243.24
Italy	42,41,760	2.42%	126924	3.35%	2.99%	3949597	93.11%	8,349.92	7,774.80	249.85
Argentina	40,93,090	2.34%	84628	2.24%	2.07%	3668672	89.63%	8,057.26	7,221.80	166.59
Germany	37,20,811	2.12%	89821	2.37%	2.41%	3574000	96.05%	7,324.43	7,035.43	176.81
Colombia	36,94,707	2.11%	94615	2.50%	2.56%	3435109	92.97%	7,273.05	6,762.03	186.25
Poland	28,77,007	1.64%	74515	1.97%	2.59%	2647688	92.03%	5,663.40	5,211.98	146.68
Iran	30,13,078	1.72%	81796	2.16%	2.71%	2624802	87.11%	5,931.26	5,166.93	161.02
Ukraine	22,81,303	1.30%	53683	1.42%	2.35%	2176093	95.39%	4,490.75	4,283.65	105.68
Peru	19,98,056	1.14%	188100	4.97%	9.41%	1955350	97.86%	3,933.18	3,849.11	370.28
Mexico	24,48,820	1.40%	229823	6.07%	9.39%	1950419	79.65%	4,820.51	3,839.41	452.41
Indonesia	18,94,025	1.08%	52566	1.39%	2.78%	1735144	91.61%	3,728.40	3,415.64	103.48
Czechia	16,64,839	0.95%	30219	0.80%	1.82%	1628416	97.81%	3,277.24	3,205.54	59.49
South Africa	17,30,106	0.99%	57592	1.52%	3.33%	1598293	92.38%	3,405.72	3,146.25	113.37
Chile	14,61,419	0.83%	30472	0.81%	2.09%	1382124	94.57%	2,876.81	2,720.72	59.98
Philippines	13,00,349	0.74%	22507	0.59%	1.73%	1216497	93.55%	2,559.74	2,394.68	44.31
Total	**17,52,74,868**	**100.00%**	**3785050**	**100.00%**	**2.16%**	**111860581**	**63.82%**	**3,45,029.27**	**2,20,197.99**	**7,450.89**

Confirmed Daily Cases and Deaths by Date



Conclusion & Future Work

Through this project, the analysis on COVID-19 data has been performed successfully. The analysis on this pandemic spread has been done and compared between different countries. The analysis of confirmed cases, active cases, recovered cases and deaths are done separately to give a clear look on how the virus is spreading, which countries are getting affected mostly and how different countries are recovering. A separate analysis on cases of any separate country can be done and predictions of different cases both around the world. Also, with Dashboard we can study data easily it can be helpful for future prediction on covid19.

References

https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started

https://www.kaggle.com/