

DMG

Assignment 1

Methodology and Assumptions

Following are libraries:

- 1-numpy
- 2-pandas
- 3-matplotlib
- 4-json
- 5-datetime

Methodology :

Preprocessing of Data : In the Preprocessing of data we opened the json file and analyse it .After analysing the data we prepare the dataframe with the help of pandas library and get the data as:

| | an | ap | ar | as | br | ch | ct | date | dd | dl | ... | sk | status | tg | tn | tr | tt | un | up | ut | wb |
|---|----|----|----|----|----|----|----|-----------|----|----|-----|----|-----------|----|----|----|----|----|----|----|----|
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 14-Mar-20 | 0 | 7 | ... | 0 | Confirmed | 1 | 1 | 0 | 81 | 0 | 12 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14-Mar-20 | 0 | 1 | ... | 0 | Recovered | 0 | 0 | 0 | 9 | 0 | 4 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14-Mar-20 | 0 | 1 | ... | 0 | Deceased | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15-Mar-20 | 0 | 0 | ... | 0 | Confirmed | 2 | 0 | 0 | 27 | 0 | 1 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15-Mar-20 | 0 | 1 | ... | 0 | Recovered | 1 | 0 | 0 | 4 | 0 | 0 | 0 | 0 |

5 rows x 41 columns

After that we did some preprocessing with data obtained.Following are the preprocessing we did on data for our problem statements

- 1-we convert date column to datetime format as it was string.
- 2-we convert all the other values to type int for performing mathematical operations.
- 3-we separate and get three different dataframe which is "Confirmed" , "Deceased" ,"Recovered".
- 4-We find the state and Union Territory as a list from the data and use it when needed.

Answer to Question 1

For the date range we include the both starting and ending date. And for total cases we added the 'tt' column otherwise we sum the column value for the state.

For Answer 1_7 we consider only states not the Union Territory.

- 1.) Total number of "Confirmed" cases from 14-March-2020 to 05-Sept-2020 = **4110211**

Total number of Recovered cases from 14-March-2020 to 05-Sept-2020 = **3177666**

Total number of Deceased cases from 14-March-2020 to 05-Sept-2020 = **70094**

2.) For Delhi state only :-

Total number of Confirmed cases from 14-March-2020 to 05-Sept-2020 = **188193**

Total number of Recovered cases from 14-March-2020 to 05-Sept-2020 = **163785**

Total number of Deceased cases from 14-March-2020 to 05-Sept-2020 = **4538**

3.) For Delhi and Maharashtra combinedly (Delhi + Maharashtra) :-

Total number of Confirmed cases from 14-March-2020 to 05-Sept-2020 = **1072055**

Total number of Recovered cases from 14-March-2020 to 05-Sept-2020 = **800359**

Total number of Deceased cases from 14-March-2020 to 05-Sept-2020 = **30813**

4.)

Highest affected state is 'mh' in terms of "Confirmed" cases from 14-March-2020 to 05-Sept-2020 = **883862**

Highest affected state is 'mh' in terms of "Recovered" cases from 14-March-2020 to 05-Sept-2020 = **636574**

Highest affected state is 'mh' in terms of "Deceased" cases from 14-March-2020 to 05-Sept-2020 = **26275**

5.)

Lowest affected state is 'mz 'in terms of "Confirmed" cases from 14-March-2020 to 05-Sept-2020 = **1062**

Lowest affected state is 'mz 'in terms of "Recovered" cases from 14-March-2020 to 05-Sept-2020 = **713**

Lowest affected state is 'mz 'in terms of "Deceased" cases from 14-March-2020 to 05-Sept-2020 = **0**

6.)

Highest Spike in Number of "Confirmed" cases in a single day in Delhi is **3947** on 2020-06-23.

Highest Number of Recovered cases in a Single day in Delhi is **7725** on 2020-06-20

Highest Number of Deceased cases in a Single day in Delhi is **437** on 2020-06-16

7.) Active cases on 05-sept-2020 from 14-March-2020 state wise for all states Separately.

'ap' :: 100880

'ar' :: 1525

'as' :: 28404

'br' :: 16735

'ch' :: 2143

'ct' :: 22320

'ga' :: 4945

'gj' :: 16266

'hp' :: 2023

'hr' :: 14912

'jh' :: 14980

'ka' :: 100224

'kl' :: 21867

'mh' :: 221013

'ml' :: 1374

'mn' :: 1872

'mp' :: 15687

'mz' :: 349

'nl' :: 701

'or' :: 25856

'pb' :: 15870

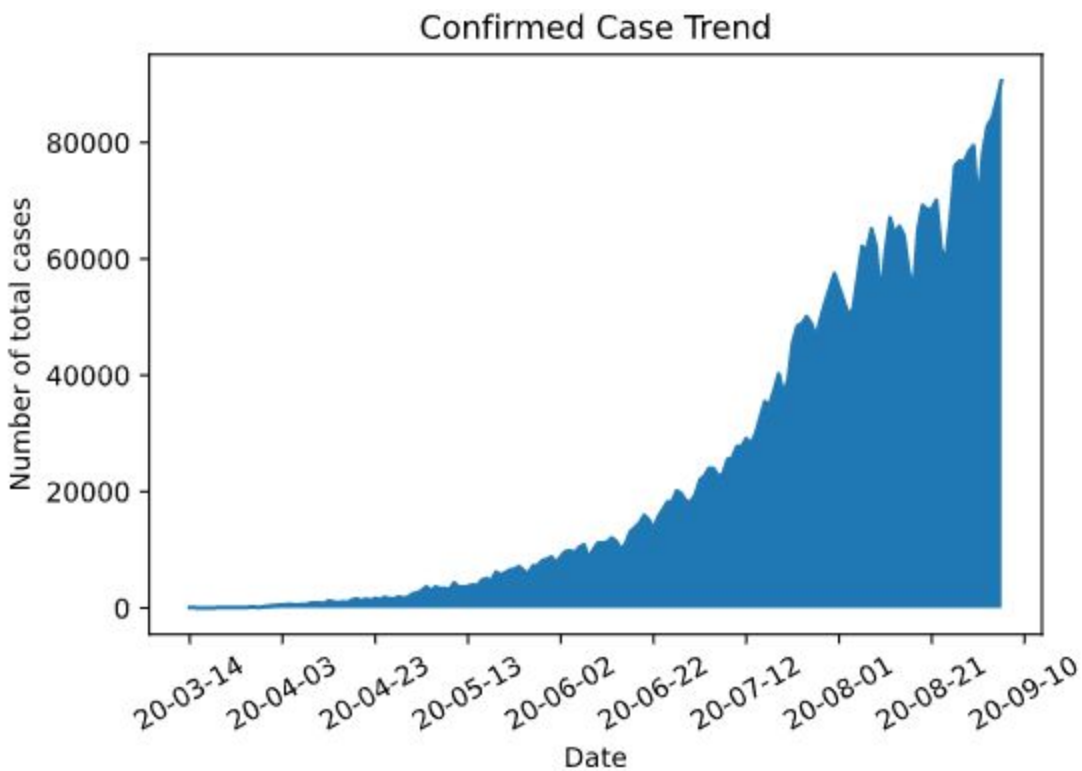
'rj' :: 14996
'sk' :: 561
'tg' :: 32405
'tn' :: 51580
'tr' :: 5905
'up' :: 59963
'ut' :: 7649
'wb' :: 23390

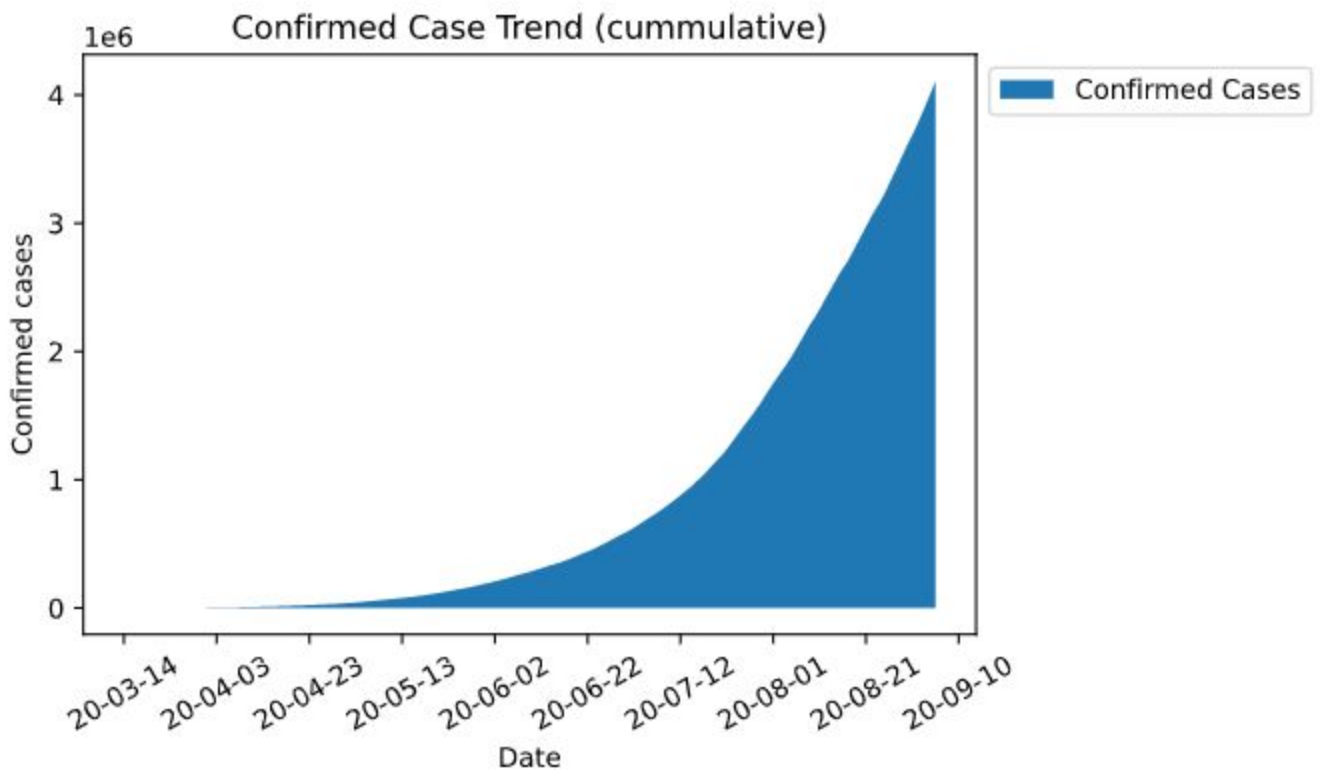
Answer to Question 2

We plot the graph using matplotlib for Question 2_1 and 2_2 we plot the graph for total cases per day and for Answer 2_3 we plot the graph with cumulative cases with respect to days.

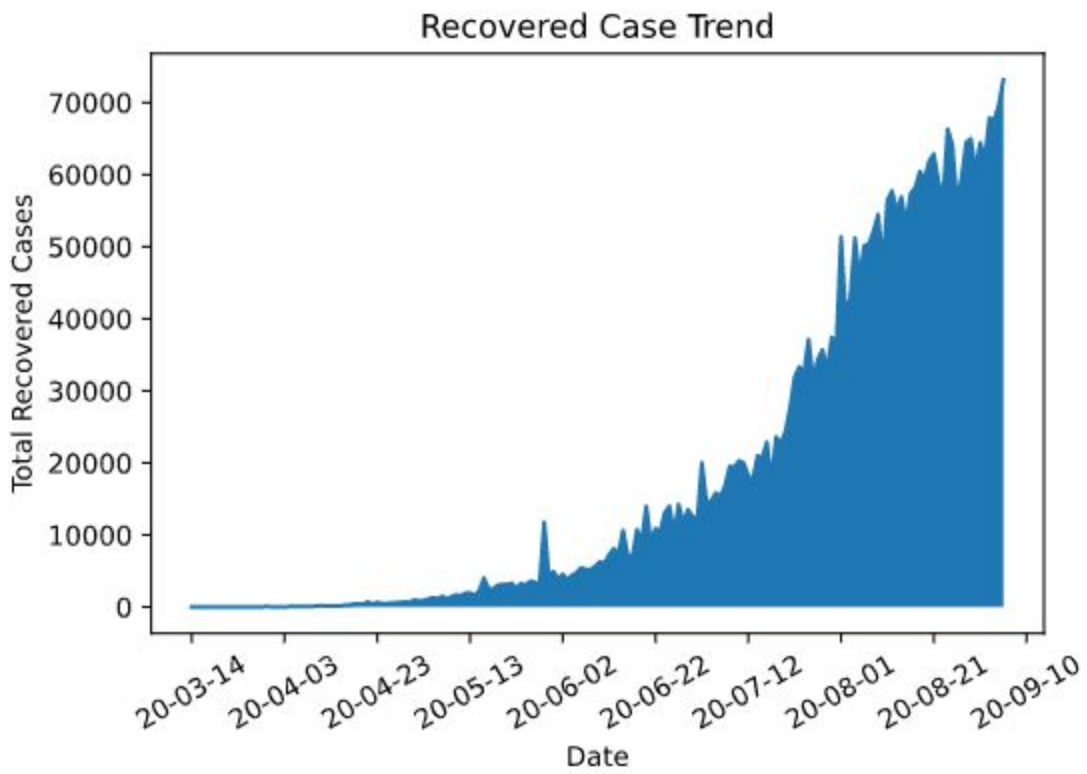
a.)

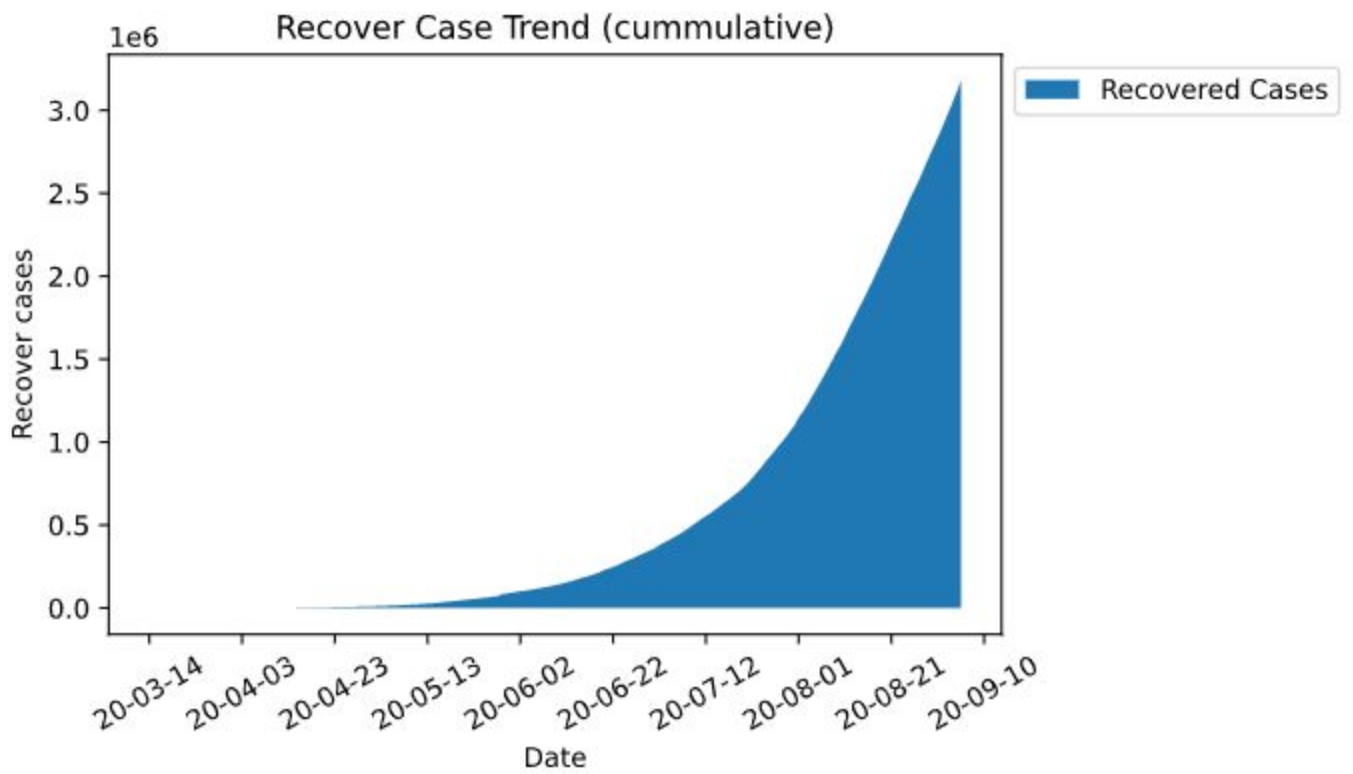
Area trend line for total “Confirmed” cases from 14-Mar-2020 to 05-Sept-2010.



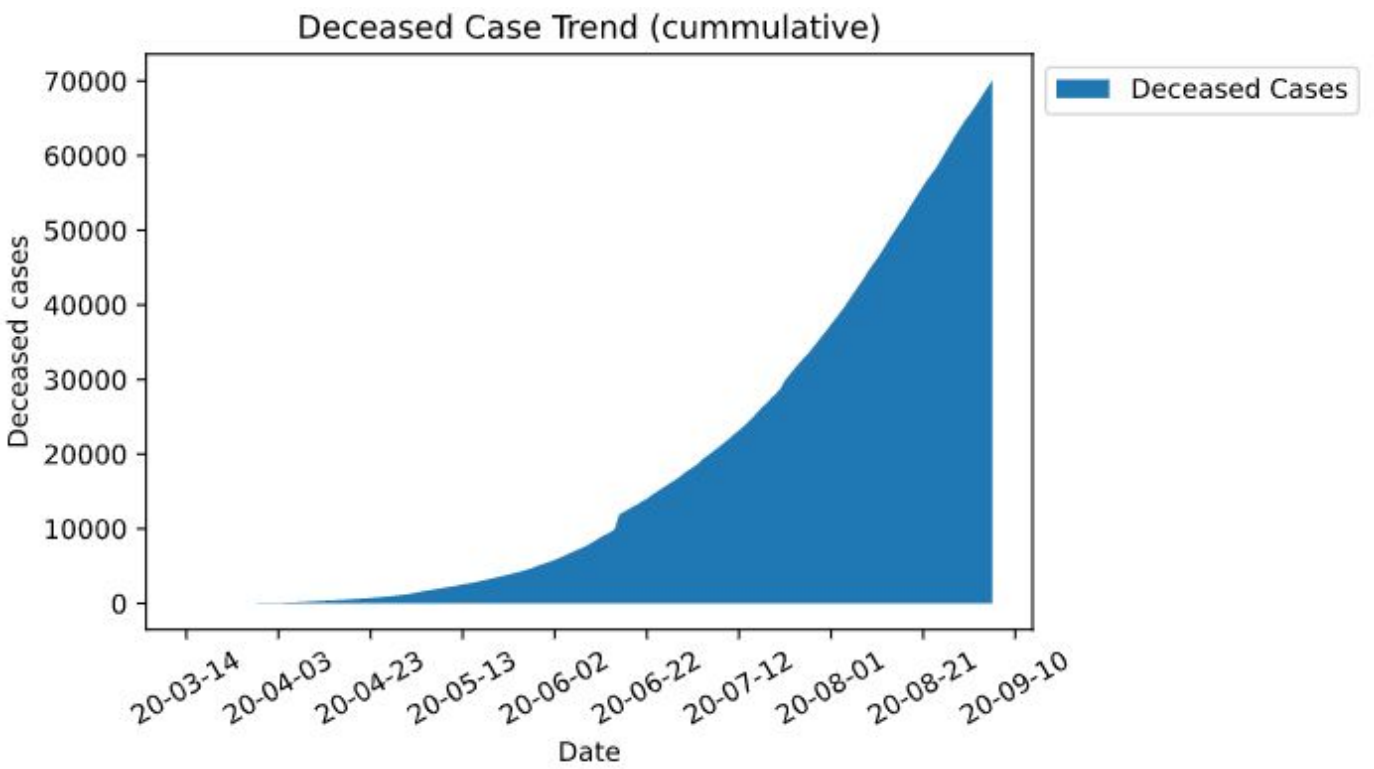
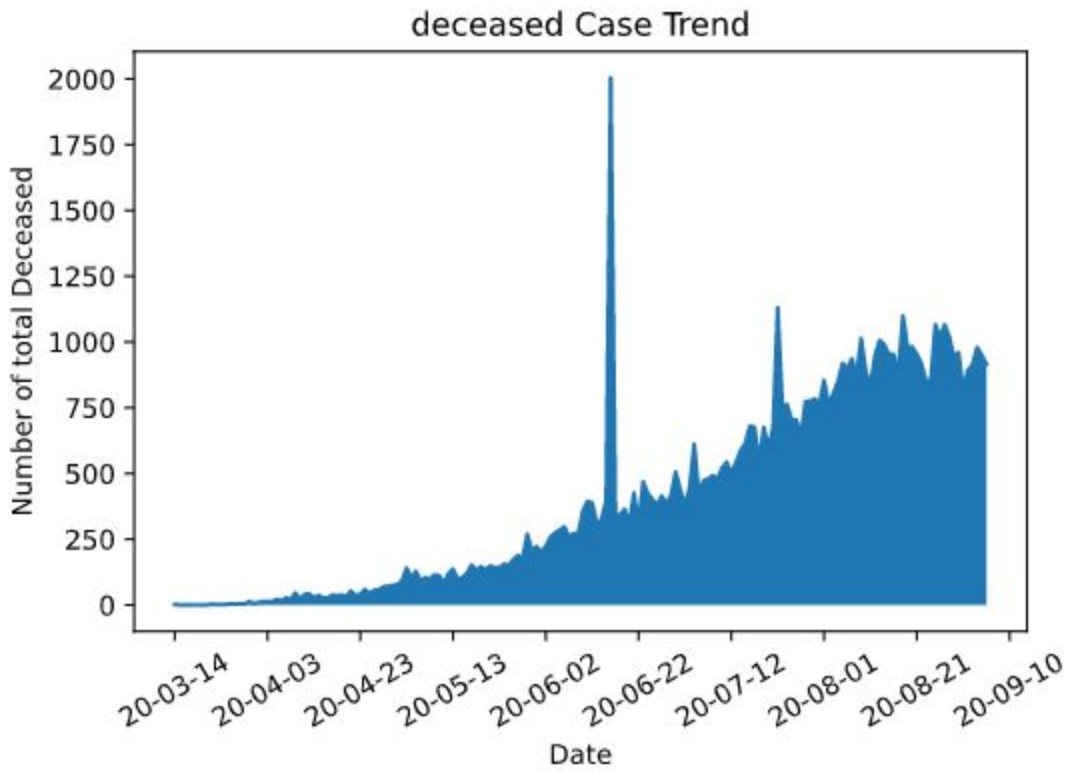


Area trend line for total “Recovered” cases from 14-Mar-2020 to 05-Sept-2010.



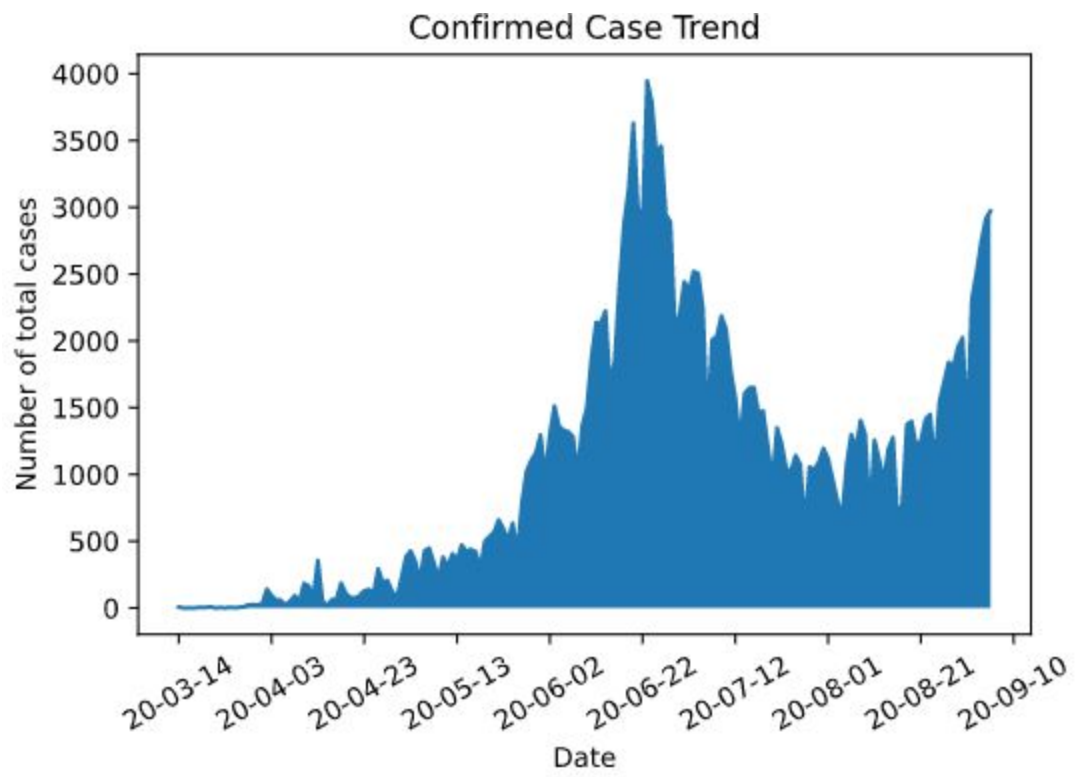


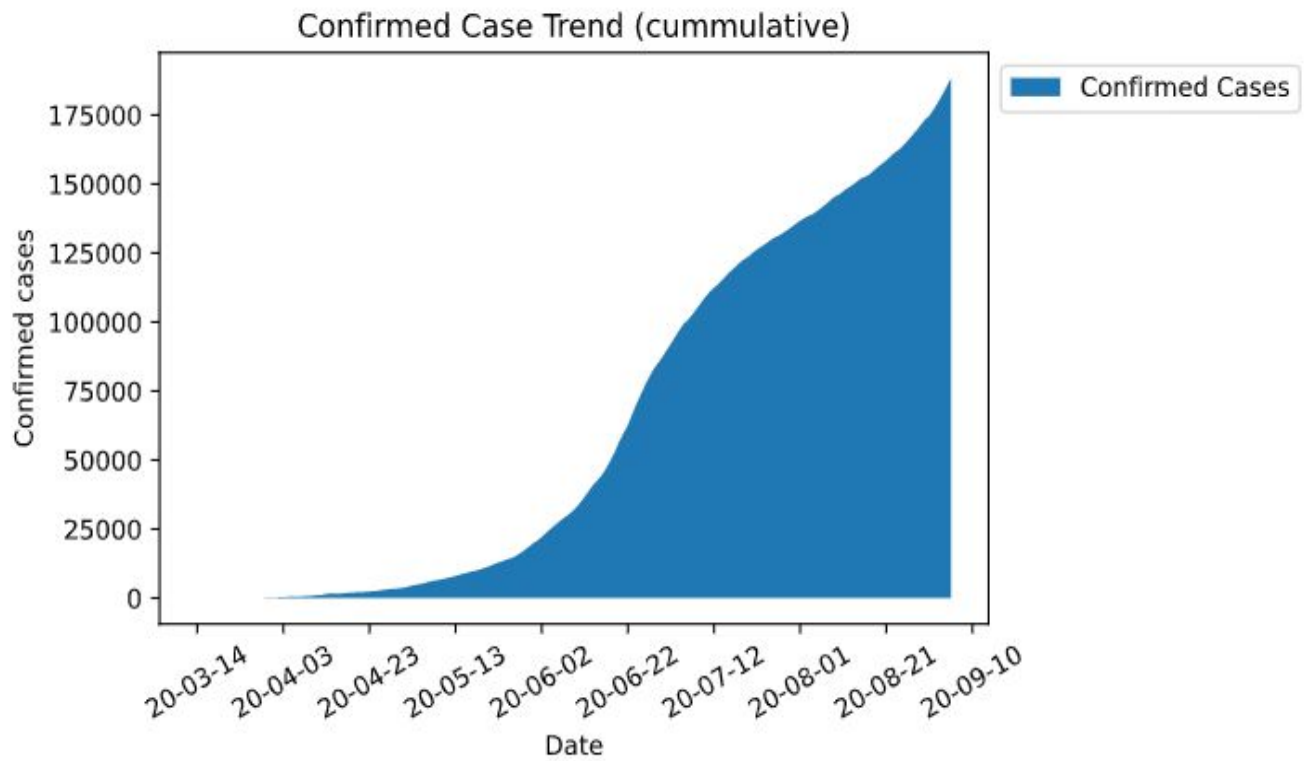
Area trend line for total “Decreased” cases from 14-Mar-2020 to 05-Sept-2010.



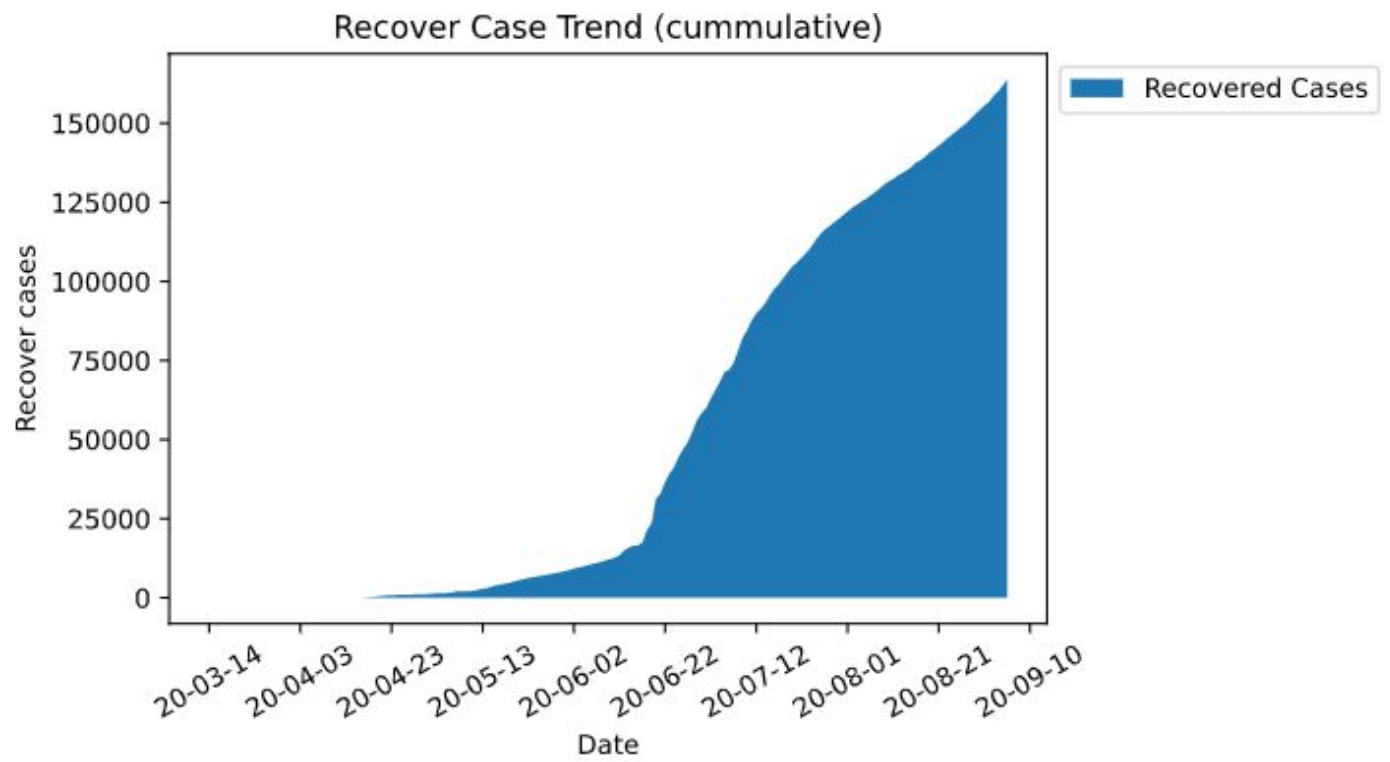
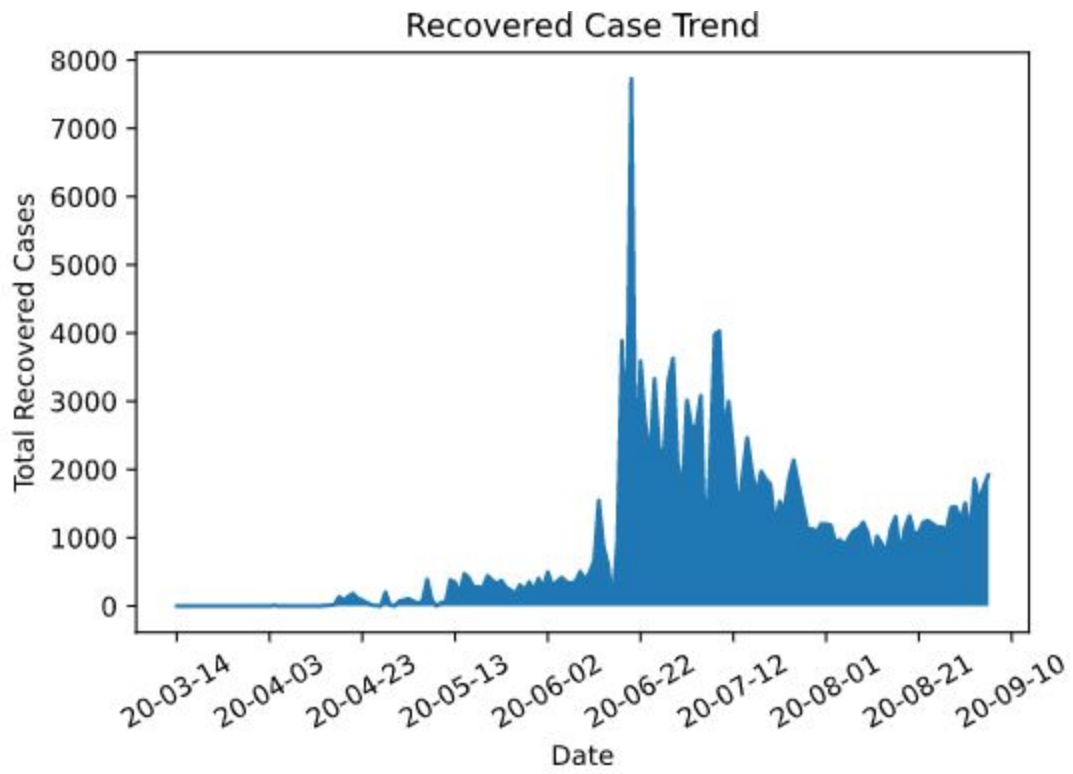
b.)

Area trend line for total “Confirmed” cases from 14-Mar-2020 to 05-Sept-2010
For state Delhi(dl) .

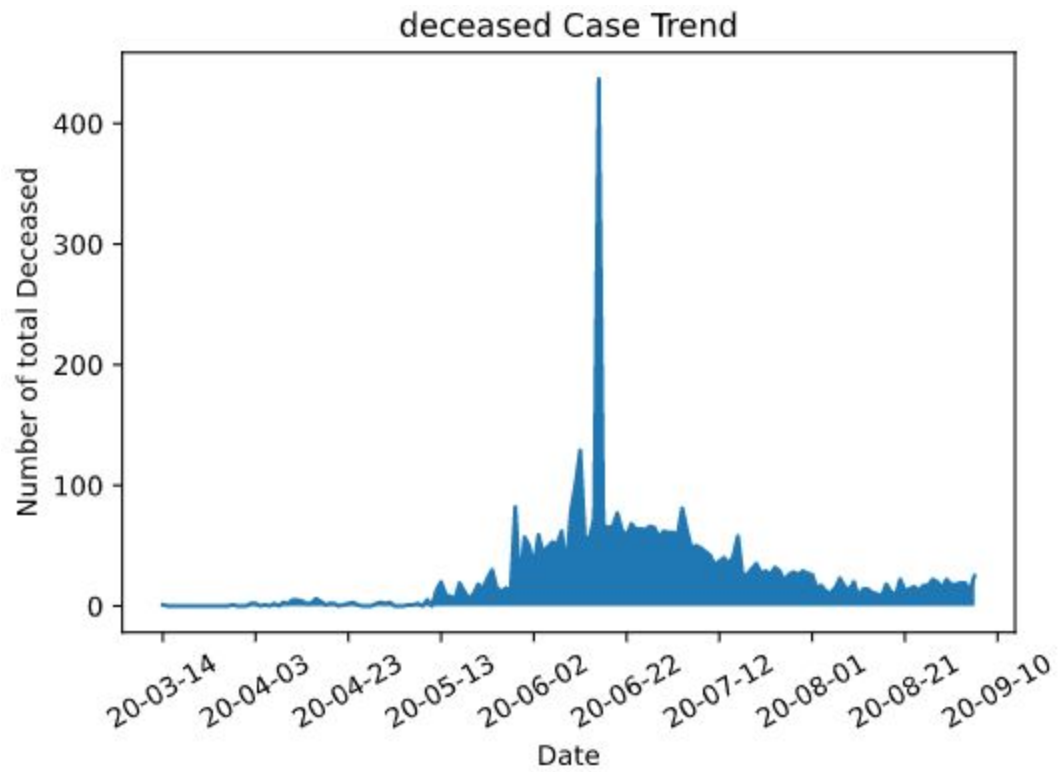


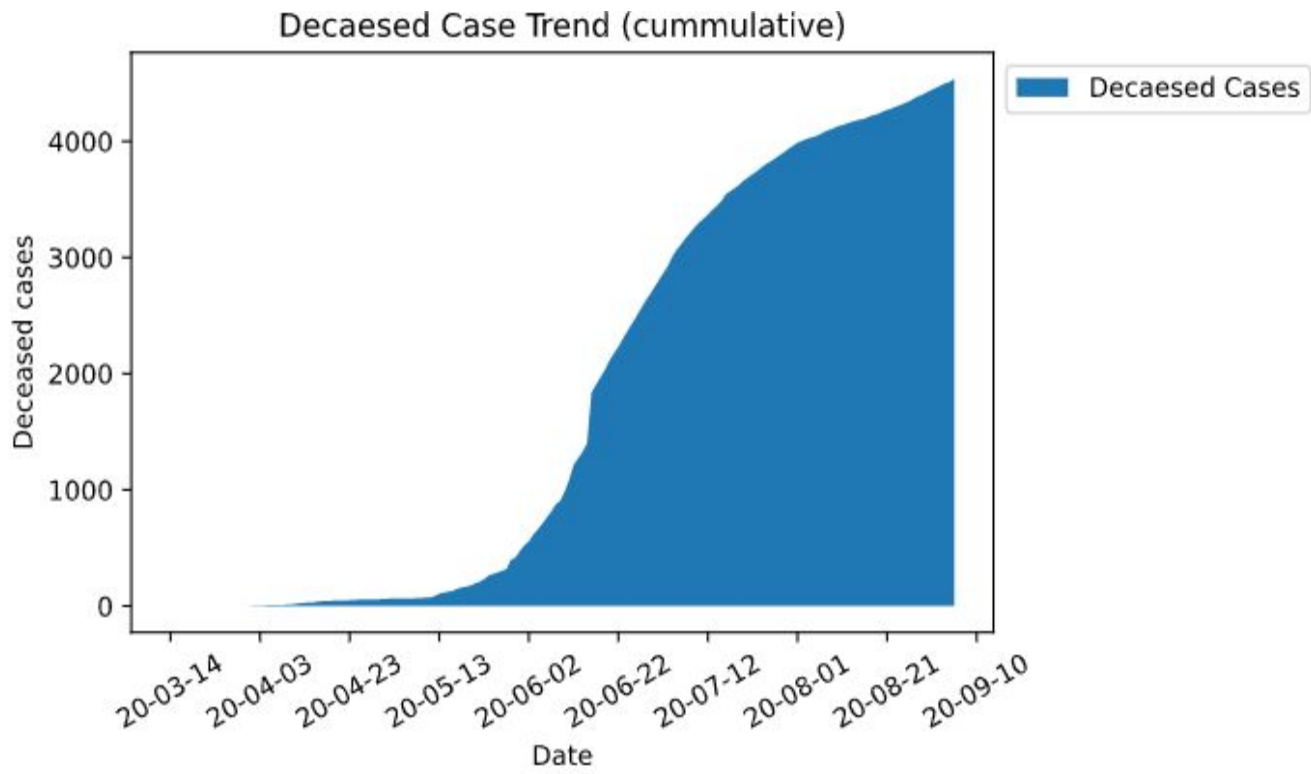


Area trend line for total “Recovered” cases from 14-Mar-2020 to 05-Sept-2010
For state Delhi(dl) .



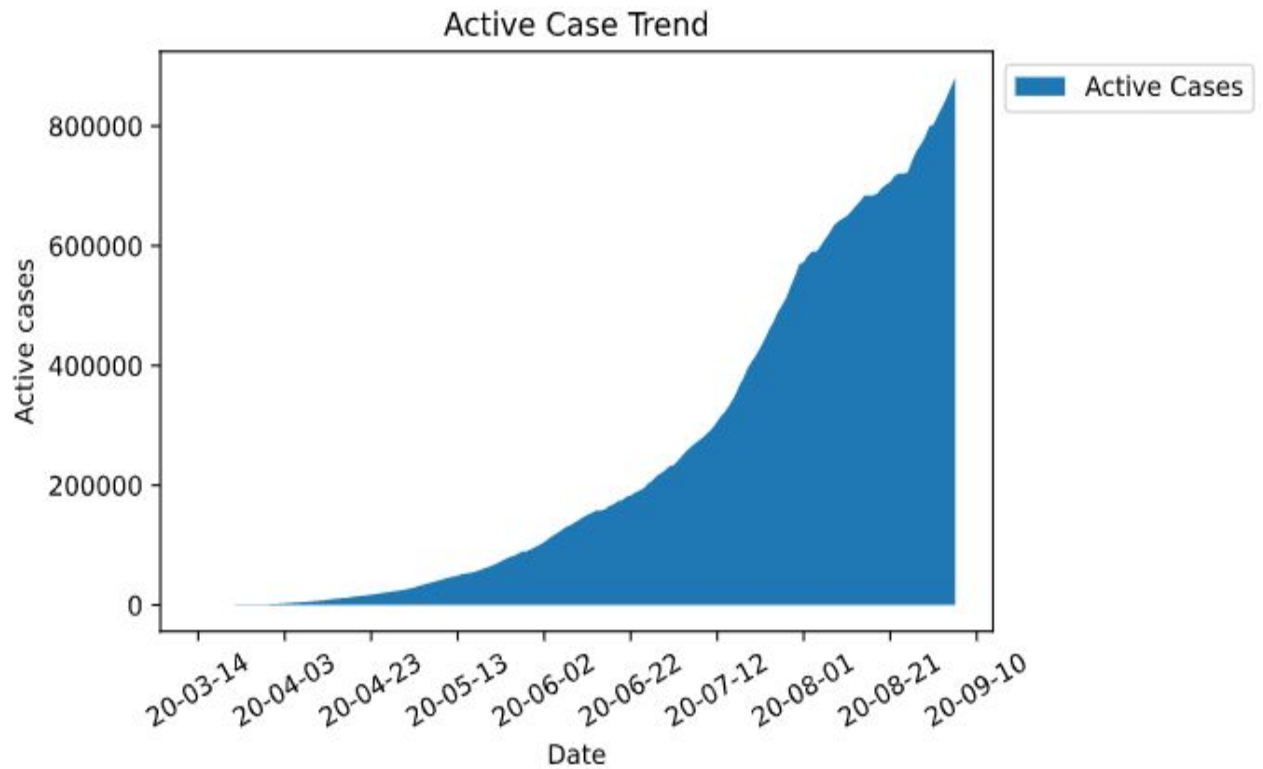
Area trend line for total “Deceased” cases from 14-Mar-2020 to 05-Sept-2010
For state Delhi(dl) .





c.)

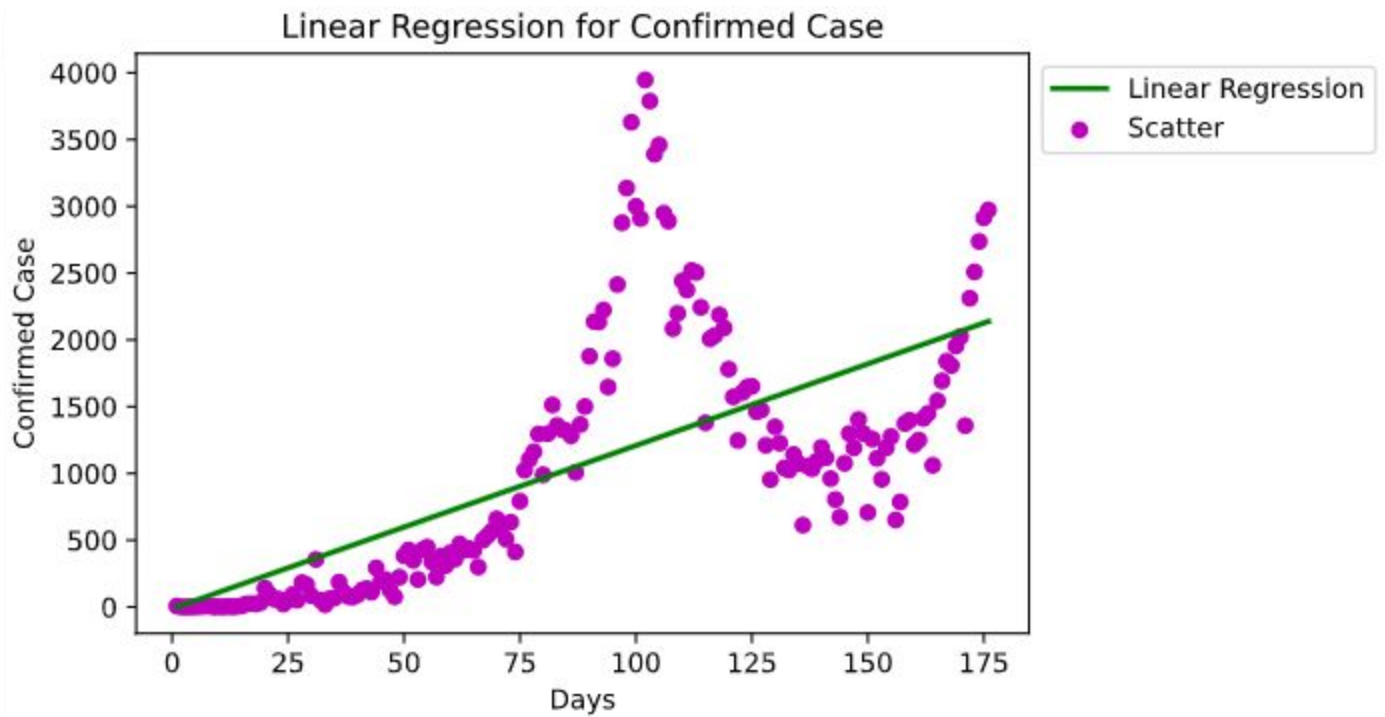
Area trend line for active cases from 14-Mar-2020 to 05-Sept-2020.



Q3.)

For our problem statement it is two variable linear regression so we try a linear function that Predict the responses as a function of independent variable.

A linear regression on the state Delhi data for "Confirmed" cases from 14-Mar-2020 to 05-Sept-2020.



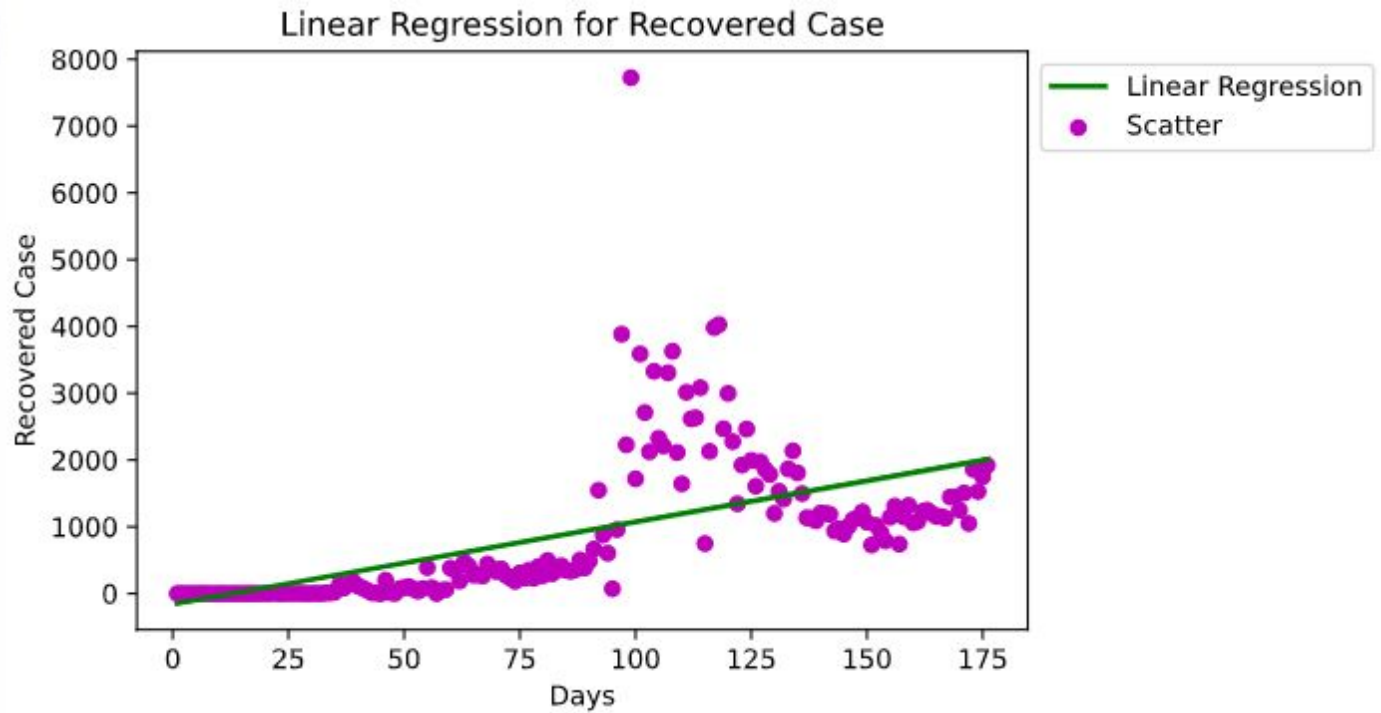
Intercept = -11.684415584415774

Slope coefficient = 12.2142692053709

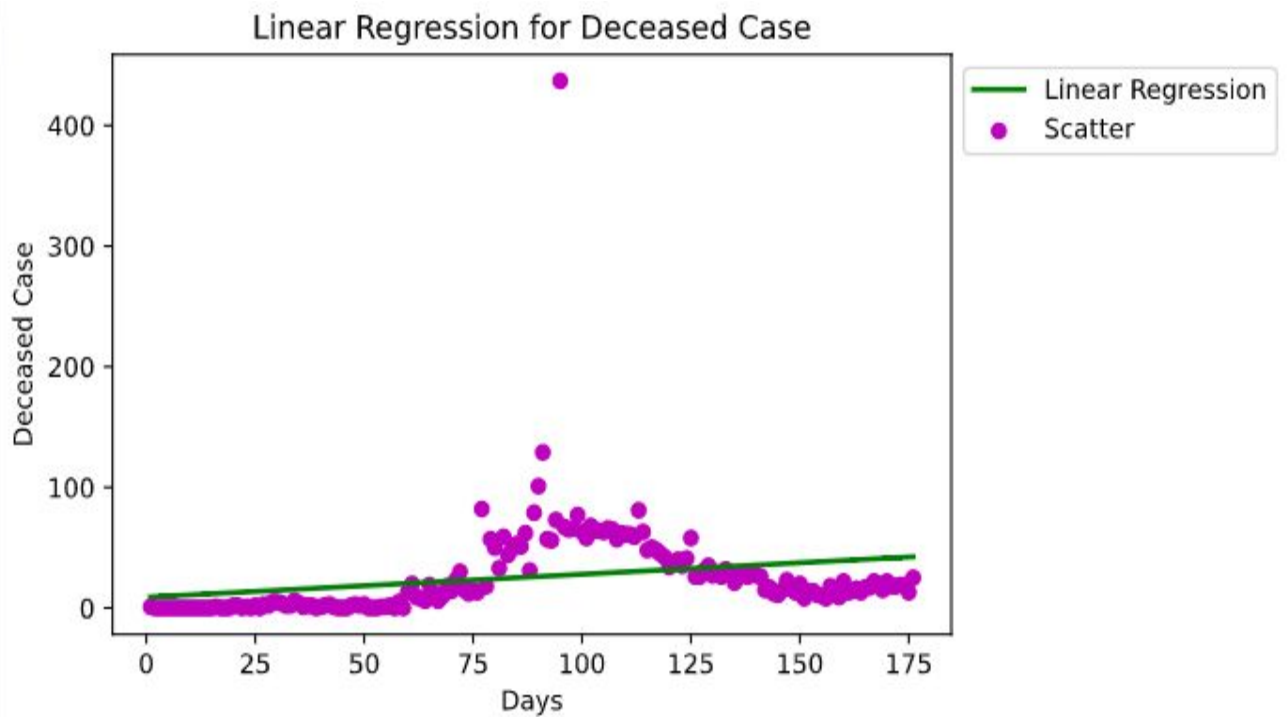
A linear regression on the state Delhi data for "Recovered" cases from 14-Mar-2020 to 05-Sept-2020.

Intercept = -158.44266233766245

Slope coefficient = 12.305528285274049



A linear regression on the state Delhi data for “Deceased” cases from 14-Mar-2020 to 05-Sept-2020.



Intercept = 8.948441558441559
Slope coefficient = 0.19023332599603787

References :

<https://www.geeksforgeeks.org/linear-regression-python-implementation/>
<https://thispointer.com/data-analysis-in-python-using-pandas/>
<https://www.tutorialspoint.com/matplotlib/>
<https://matplotlib.org/3.3.1/contents.html>