Data Mining – Assignment 01

README FILE

Methodology for each question:

Ques1.

1.) Read the file and select the data between ‘start_date’ and ‘end_date’. To count the total number of Confirmed, Recovered and Deceased between the given dates, we have grouped the data according to the ‘status’ and add the total column for each group.

2.) Read the file and select the data between ‘start_date’ and ‘end_date’. To count the total number of Confirmed, Recovered and Deceased between the given dates for ‘Delhi’, we have grouped the data according to the ‘status’ and add the ‘Delhi’ column for each group.

3.) Read the file and select the data between ‘start_date’ and ‘end_date’. To count the total number of Confirmed, Recovered and Deceased between the given dates for ‘Delhi’ and ‘Maharashtra’ (Sum of both states count), we have grouped the data according to the ‘status’ and add the data of column ‘Delhi’ and ‘Maharashtra’ respectively and then sum these two columns to produce the desired output.

4.) Read the file and select the data between ‘start_date’ and ‘end_date’. To find the highest affected states in terms of Confirmed, Recovered and Deceased, we have grouped the data according to the ‘status’ and we have find the maximum for each group and then we have searched the state to which this maximum value belongs using idxmax.

5.) Read the file and select the data between ‘start_date’ and ‘end_date’. To find the lowest affected states in terms of Confirmed, Recovered and Deceased, we have grouped the data according to the ‘status’ and we have find the minimum for each group and then we have searched the state to which this minimum value belongs using idxmin.

6.) Read the file and select the data between ‘start_date’ and ‘end_date’. To find the day and count with highest spike in a day for ‘Delhi’ in terms of Confirmed, Recovered and Deceased, we have first sorted our data on the basis of the data in column ‘Delhi’ in descending order, then we grouped the data according to the status and printed the first value for each status.

7.) Read the file and select the data between ‘start_date’ and ‘end_date’. To find the active cases between the given dates, we have grouped our data on the basis of ‘status’ and then we have calculated the sum for each group according to the states. Then to access the status we have transposed our data to calculate Active cases using

Active = Confirmed – (Recovered + Deceased)

Ques 2.

1.) To plot the area trend line for the total Confirm, Recovered and Deceased between the given dates, we have grouped our data according to the status and date then we find the cumulative sum for each category in status. We used matplotlib to plot the area trend line. We dropped the column for total.

2.) To plot the area trend line for the total Confirm, Recovered and Deceased for the state ‘Delhi’ between the given dates, we have grouped our data according to the status and date then we find the cumulative sum for each category in status for the column Delhi. We used matplotlib to plot the area trend line.

3.) To plot the area trend line for the active cases between the given dates, we have grouped our data according to the status and date then we find the cumulative sum for each category in status. We used matplotlib to plot the area trend line. To calculate the active case, we have used the formula

Active = Confirmed – (Recovered + Deceased)

Ques3.

To find the slope and intercept for Confirmed, Recovered and Deceased, We have first converted dates into numeric data then using the mean, variance and covariance, we have calculated the slopes and intercepts for the status. Then we plotted the graph using

Y = (Slope * x) + Intercept

confirmed_intercept = -11.684415584415774

confirmed_slope = 12.214269205370902

recovered_intercept = -158.442662337662

recovered_slope = 12.305528285274045

deceased_intercept = 8.948441558441562

deceased_slope = 0.19023332599603782

Assumptions:

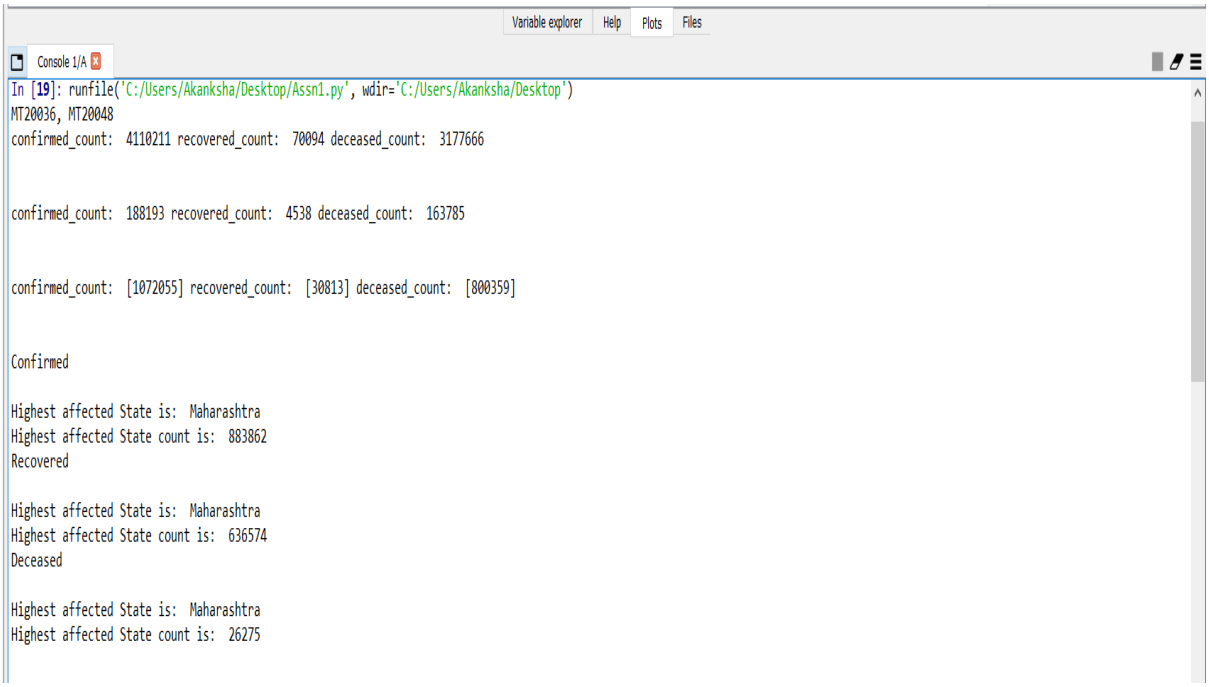
1. We have directly read the json file from the link.

2. For question 1.4, 1.5 and 1.7, we have assumed that states does not contain total and Union Territories.

3. For question 2, we have considered Cumulative Sum.

4. For Question 1.1 We have included Union Territories.

Results:

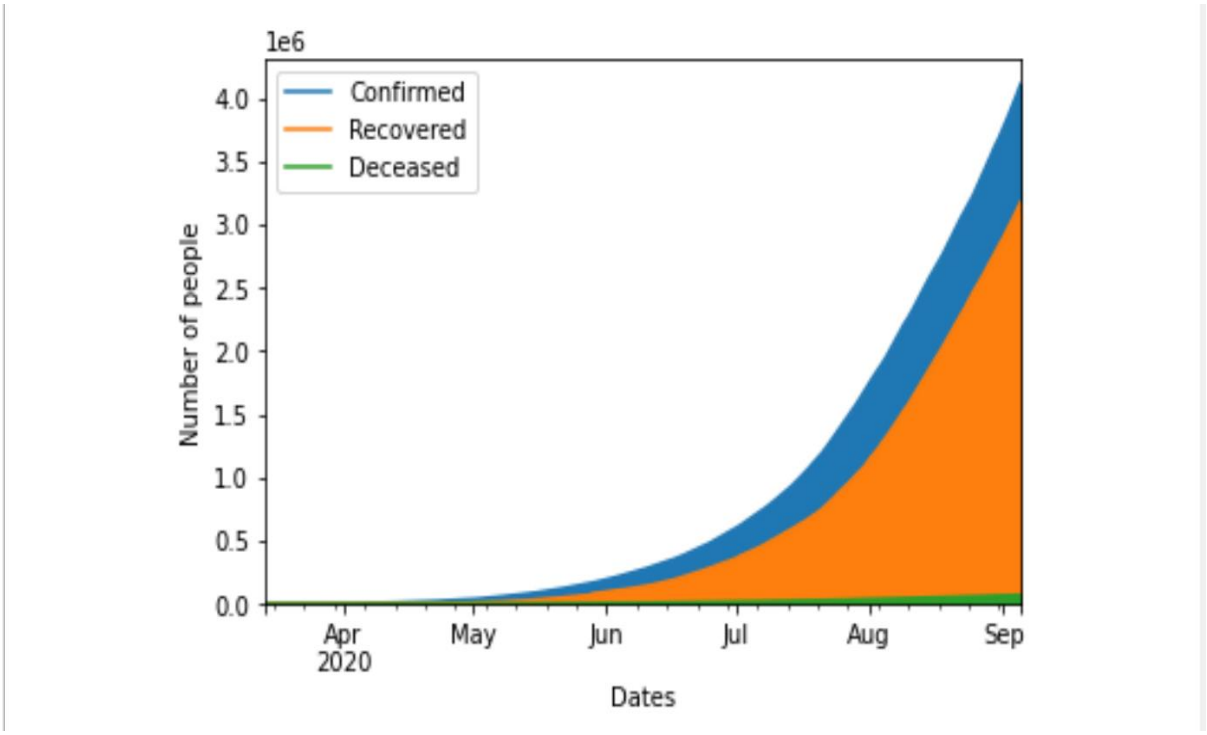

Variable explorer
Help
Plots
Files
Console 1/A
In [19]: runfile('C:/Users/Akanksha/Desktop/Assn1.py', wdir='C:/Users/Akanksha/Desktop')
MT20036, MT20048
confirmed_count: 4110211 recovered_count: 70094 deceased_count: 3177666
confirmed_count: 188193 recovered_count: 4538 deceased_count: 163785
confirmed_count: [1072055] recovered_count: [30813] deceased_count: [800359]
Confirmed
Highest affected State is: Maharashtra
Highest affected State count is: 883862
Recovered
Highest affected State is: Maharashtra
Highest affected State count is: 636574
Deceased
Highest affected State is: Maharashtra
Highest affected State count is: 26275

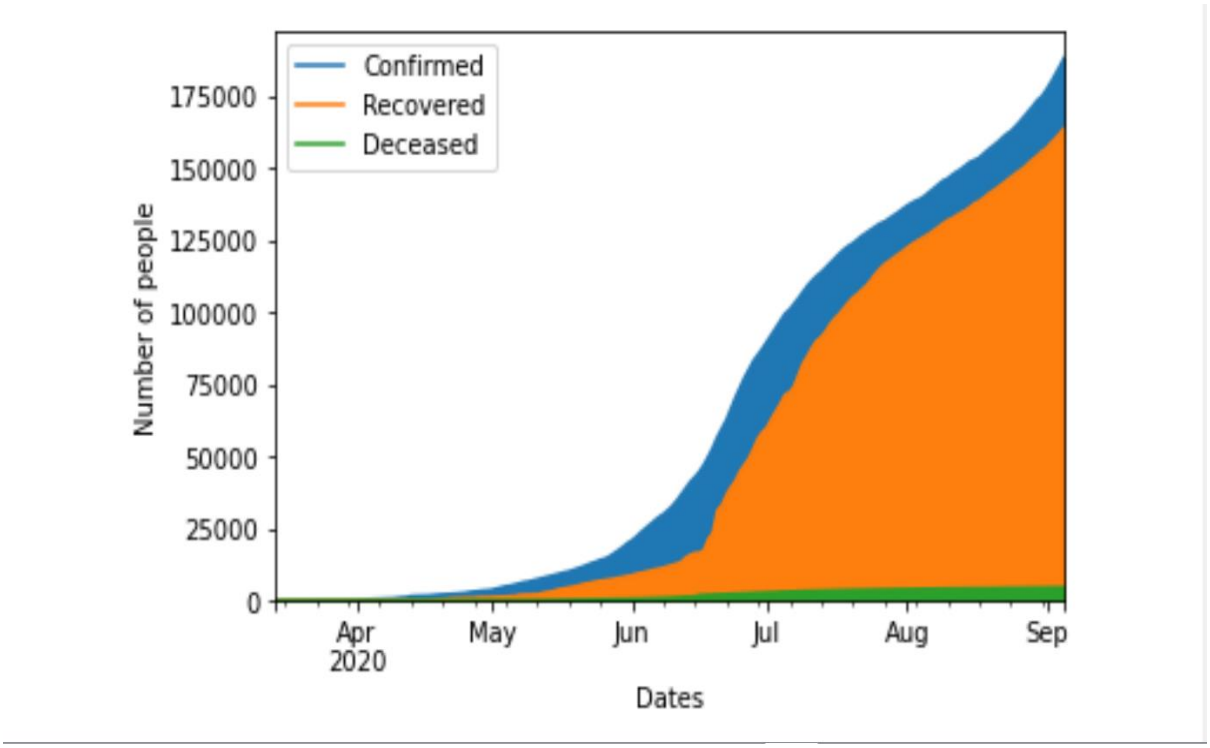
```
Confirmed

Lowest affected State is:  Mizoram
Lowest affected State count is:  1062
Recovered

Lowest affected State is:  713
Lowest affected State count is:  Mizoram
Deceased

Lowest affected State is:  0
Lowest affected State count is:  Mizoram


Confirmed

Day:  2020-06-23 00:00:00
Count:  3947
Recovered

Day:  2020-06-20 00:00:00
Count:  7725
Deceased

Day:  2020-06-16 00:00:00
Count:  437
```

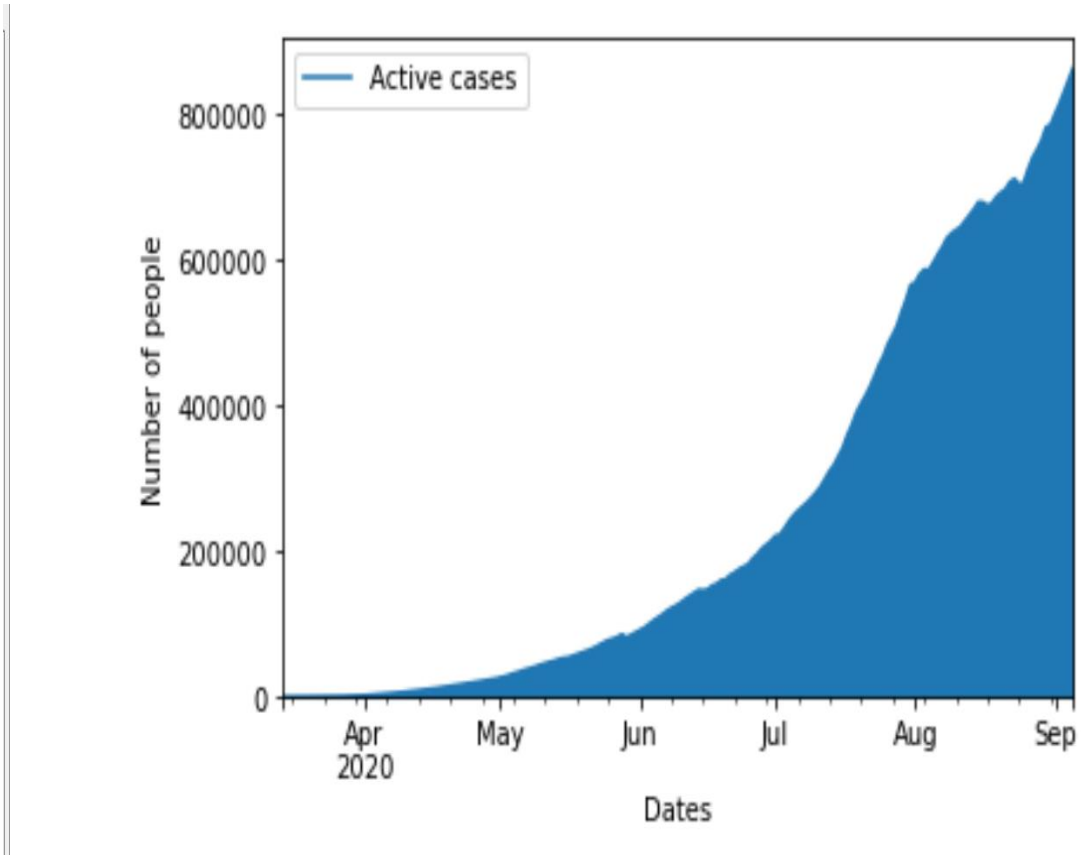
Console 1/A

status	Confirmed	Deceased	Recovered	Active
Andhra Pradesh	487331	4347	382104	100880
Arunachal Pradesh	4914	8	3381	1525
Assam	123821	356	95061	28404
Bihar	145861	750	128376	16735
Chhattisgarh	43163	356	20487	22320
Goa	20455	229	15281	4945
Gujarat	103006	3093	83647	16266
Himachal Pradesh	7018	53	4942	2023
Haryana	74272	781	58579	14912
Jharkhand	49772	462	34330	14980
Karnataka	389232	5709	283299	100224
Kerala	84759	338	62554	21867
Maharashtra	883862	26275	636574	221013
Meghalaya	2916	15	1527	1374
Manipur	6883	36	4975	1872
Madhya Pradesh	71880	1544	54649	15687
Mizoram	1062	0	713	349
Nagaland	4128	8	3419	701
Odisha	120221	591	93774	25856
Punjab	61527	1808	43849	15870
Rajasthan	89363	1122	73245	14996
Sikkim	1914	5	1348	561
Telangana	137885	877	104603	32405
Tamil Nadu	457697	7751	398366	51580
Tripura	14524	136	8483	5905
Uttar Pradesh	259765	3843	195959	59963
Uttarakhand	23961	330	15982	7649
West Bengal	177701	3510	150801	23390

```
confirmed_intercept =  -11.684415584415774
confirmed_slope =  12.214269205370902
recovered_intercept =  -158.442662337662
recovered_slope =  12.305528285274045
deceased_intercept =  8.948441558441562
deceased_slope =  0.19023332599603782
```




Confirmed
Recovered
Deceased
175000
150000
125000
100000
75000
50000
25000
0
Number of people
Apr
2020
May
Jun
Jul
Aug
Sep
Dates


Active cases
800000
600000
400000
200000
0
Number of people
Apr
2020
May
Jun
Jul
Aug
Sep
Dates