

ASSIGNMENT – 1

(Total Marks - 150)

(This is your first assignment in this course. This will help you to build your thinking ability in the field of data. Try to explore more and report more analysis in your report file.)

Dataset: https://api.covid19india.org/states_daily.json

Download and save it as a JSON file.

The above link contains covid19 data arranged in date wise and state wise fashion. Variable Status dictates whether the JSON array has data of “Confirmed”, “Recovered” or “Deceased”.

Instructions:

- Use python language and submit working code.
- You need to submit the README.pdf, and working Code file
- You are allowed to use libraries such as pandas, matplotlib, etc
- Mention methodology, assumptions, and results you may have in README.pdf.
- This assignment can be done in pairs, only one student has to submit, other one can simply check in at the time of submission. An excel sheet will be shared later to put your group details.
- Submission format:
If working in pairs: A1_ML_<Register_Number1>_<Name1>_<Register_Number2>_<Name2>
If working individually: A1_ML_<Register_Number>_<Name>
- For Q1 and Q2:
 - o Your code should be flexible to take any start and end dates
 - o Submit all answers and plots in README file
 - o Date format: format='%Y%m%d' e.g: 2020-03-14

Question 1: Data Manipulation

Total Marks - 70

1. Count the total number of “Confirmed”, “Recovered” and “Deceased” from 14-Mar-2020 to 30-Sept-2020 and report the numbers.
2. Count the total number of “Confirmed”, “Recovered” and “Deceased” from 14-Mar-2020 to 05-Sept-2020 for the state of Delhi (dl)
3. Report total count of “Confirmed”, “Recovered” and “Deceased” count from states Delhi + Karnataka (ka) (Sum of both states count) from 14-Mar-2020 to 05-Sept-2020.
4. Report the highest affected state in terms of “Confirmed”, “Recovered” and “Deceased” with the count till 05-Sept-2020 from 14-Mar-2020.
5. Report the lowest affected state in terms of “Confirmed”, “Recovered” and “Deceased” with the count till 05-Sept-2020 from 14-Mar-2020.

6. Find the day and count with the highest spike in a day in the number of cases for the state Delhi for “Confirmed”, “Recovered” and “Deceased” between dates 14-Mar-2020 and 05-Sept-2020.
7. Report active cases (Assume active = Confirmed - (Recovered + Deceased)) state wise for all states separately on date 05-Sept-2020 (This date only) starting from 14-March-2020.

Question 2: Plotting

Total Marks - 30

1. Plot the area trend line for total “Confirmed”, “Recovered” and “Deceased” cases from 14-Mar-2020 to 05-Sept-2020.
2. Plot the area trend line for total “Confirmed”, “Recovered” and “Deceased” cases for the state Delhi (dl) from 14-Mar-2020 to 05-Sept-2020.
3. Plot the area trend line for active cases. Assume active = Confirmed - (Recovered + Deceased) from 14-Mar-2020 to 05-Sept-2020.

Question 3: Linear Regression

Total Marks - 50

Implement a linear regression on the state Delhi data over dates, separately for “Confirmed”, “Recovered” or “Deceased” and report intercept and slope coefficients for all 3 cases from 14-Mar-2020 to 05-Sept-2020.