SEMESTER PROJECT REPORT

- <u>TOPIC:</u> Identification of convection-soil moisture regimes from reanalysis data
- <u>NAME:</u> Barish Sarkhel
- <u>REG NO.:</u> 20201242
- <u>SEMESTER:</u> 6th
- <u>COURSE:</u> Data Science Semester Project, DS-3613
- <u>SUPERVISOR:</u> Joy Merwin Monteiro

➢ Introduction:

In this project I have tried to see the distribution of soil and rainfall across different cities in India, across different seasons and the final goal is to make a prediction model based on their relations. Here I have discussed all those plotting and relation that I have tried to form. All the work I had done using Python programming language. I have used data from two Sources IMD data and ERA-5 data.

➢ Those files are containing in netCDF format. The first part of the project about understood those files, reading them and turning them in plotting or data set in interpretable format to understand about them properly.

• Overall rainfall plotting in India:

At first after downloading the data I have tried to plot those to see how they are distributed overall India. I plotted them for each of the day of a year. Here I am giving one of them.
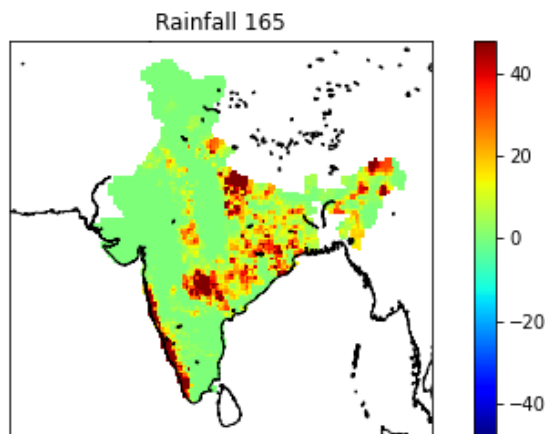

• Plotting rainfall of some specific locations of India:

Now I took 5 year netCDF file of rainfall and merge them and then I exclude a particular longitude and latitude data from that. Then plot them for major cities of India to see are there any variation among those cities. Here are some of them. What it turns out that there is a significant difference between them in term of amount, frequency etc.
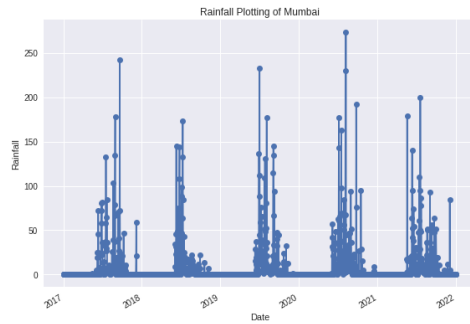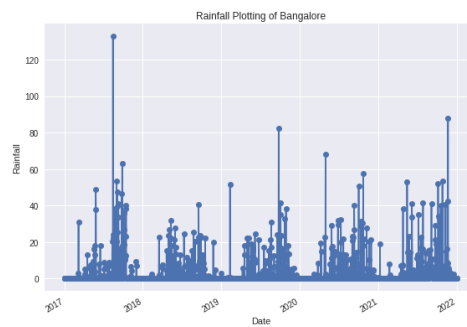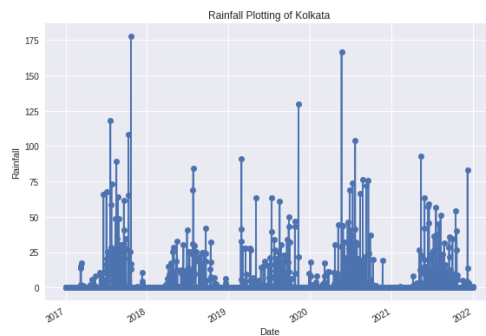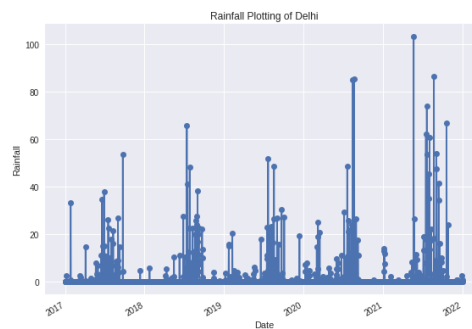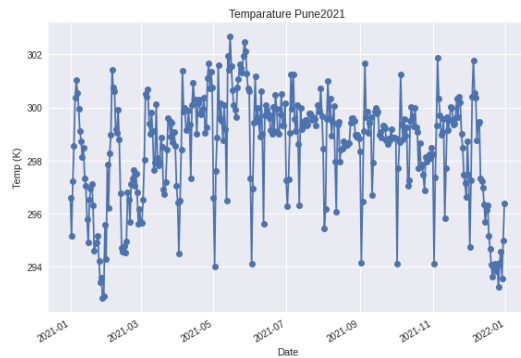
Mumbai
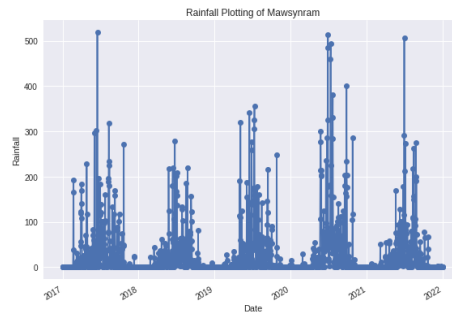
Mawsynram
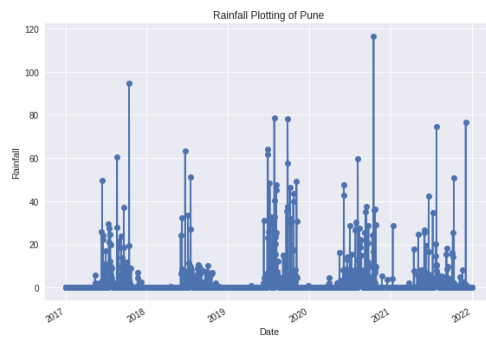
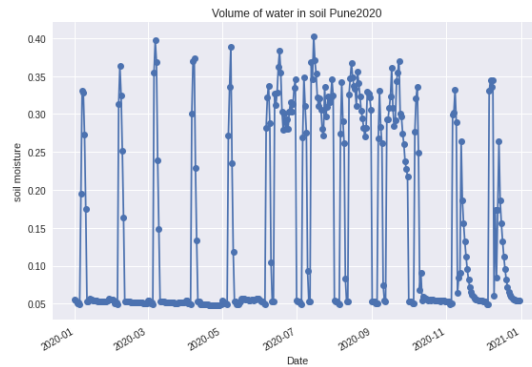

Pune

Bengalore

Kolkata

Delhi

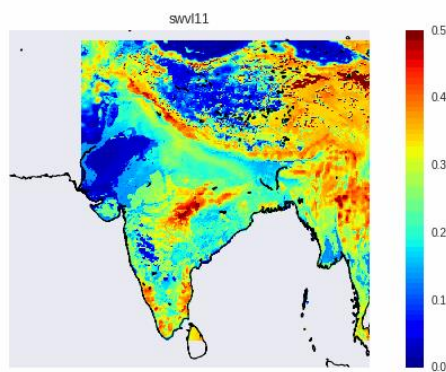- Plotting for soil water volume and temperature:

Then I took data from the ERA-5 and tried to do some spatial plotting as I did for rainfall and tried to see how they look. Also tried to see their distribution among different cities as before. Here I attached some of them.
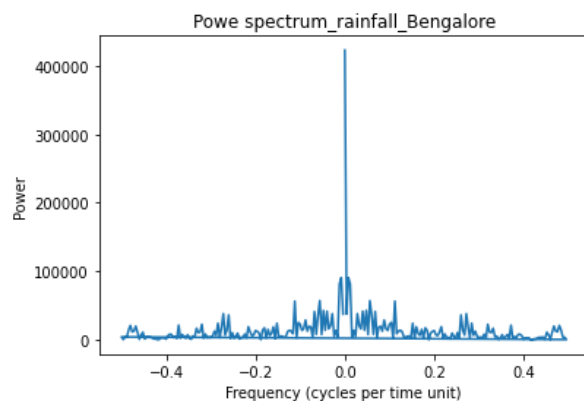
Pune (Temperature, 2021)
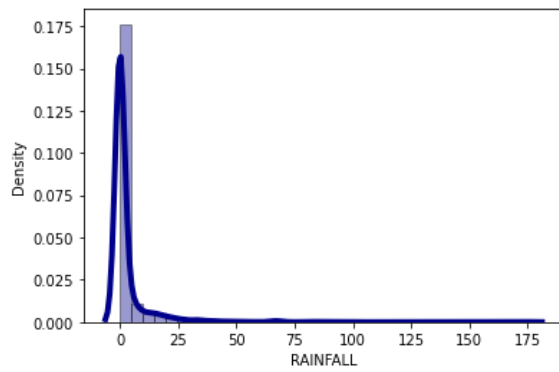

Pune (Soil water Volume, 2020)

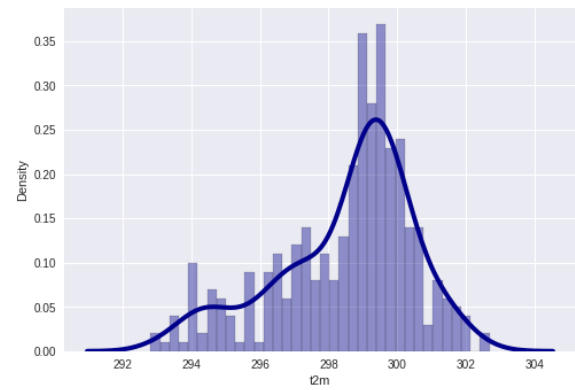- Power spectrum and Probability distribution function plotting:

Then I tried to see the power spectrum plotting of the rainfall and tried to see the frequencies where the plot getting peaks. Here I am giving the example of the city Bengalore.

Also I had plotted the probability distribution of them. Hare I attached them also.
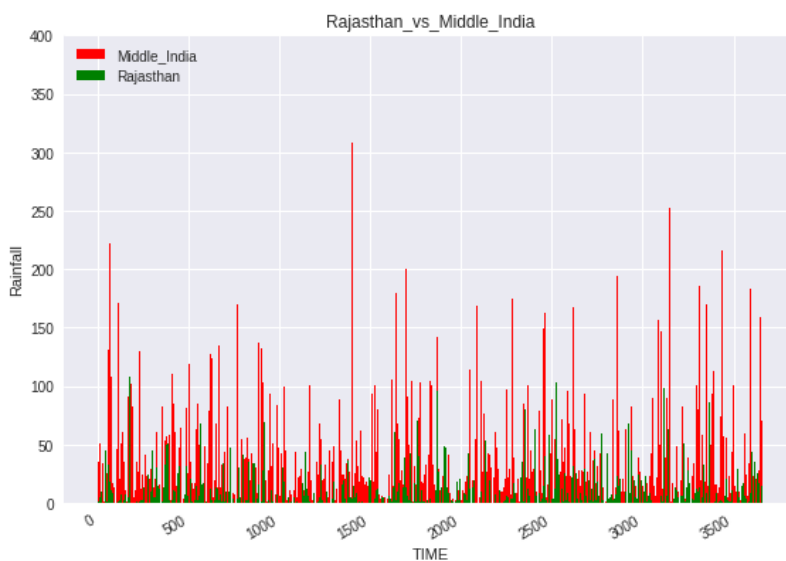
PDF of rainfall in Pune

PDF of temperature in Bangalore

- Bar plots:

Here I took 30 years June to September data and took some data point of Middle India and Rajasthan to see their variation in rainfall over the years.

These are all type of plotting I did to see their distribution and to get some primary idea about the them across different time and different places in India.

➢ Now the next part started where I tried to find some relation between rainfall and soil water volume. Here I am giving all the method I had tried to see if there is some relation between them.

- Linear regression:

At first I tried to see if there is some linear relation between them. I plotted rainfall along the X-axis and Soil water volume in Y-axis.

Regarding the data set, here both the rainfall data and soil water volume taken from the ERA-5. In case of soil water volume data it was given in four different steps there. I averaged them by scaling the data based on their level.
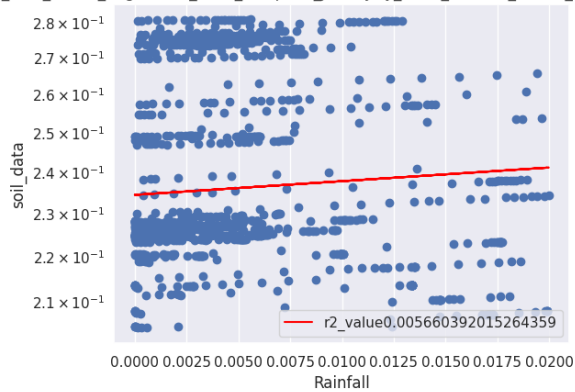
So after doing this, I got this plotting
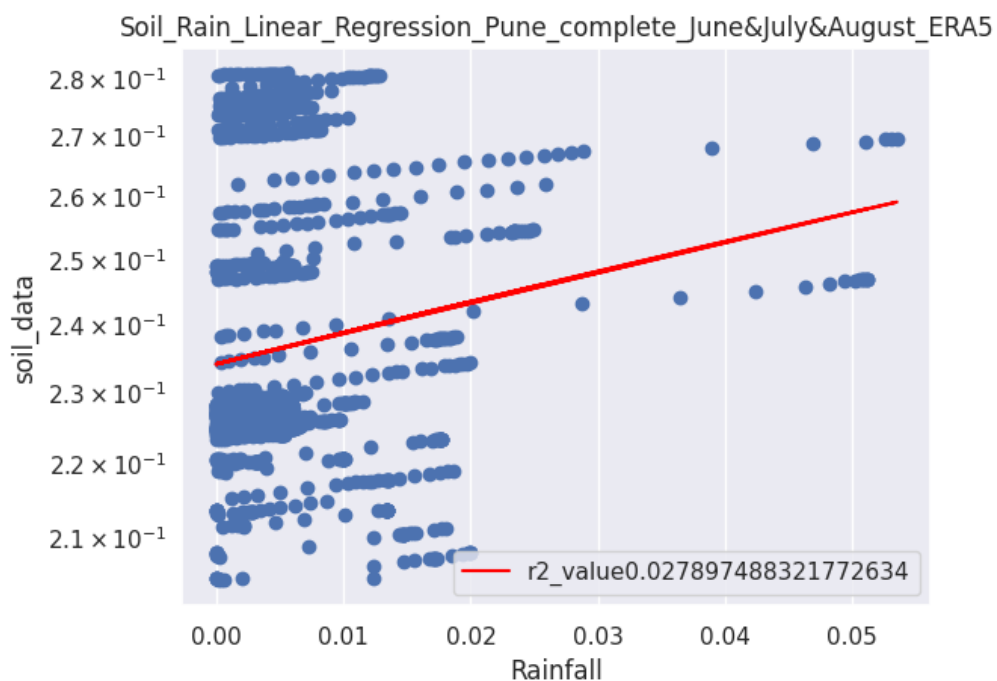
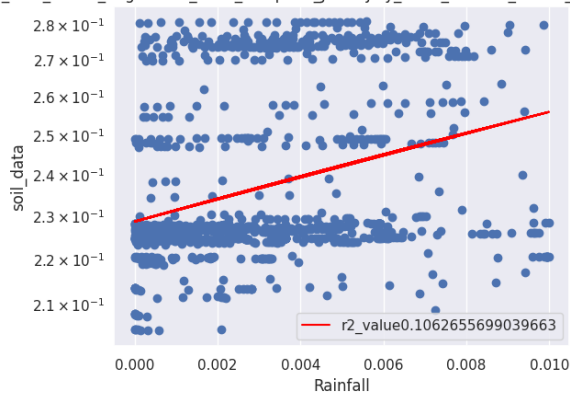
But we can see the R^2 value is too low here.

- Linear regression excluding outliers:

I tried some plotting by excluding but it did not developed so much.

Soil_Rain_Linear_Regression_Pune_complete_June&July_ERA5_exclude_outlier_rain_greater_0.02


Soil_Rain_Linear_Regression_Pune_complete_June&July_ERA5_exclude_outlier_rain_greater_0.01


- Linear regression by averaging 4 days rainfall and soil water volume:

To find some relation I averaged 4 days rainfall value and plot the same plot again. Also I did the same thing for soil data also. Plots are here. There is not much development here.
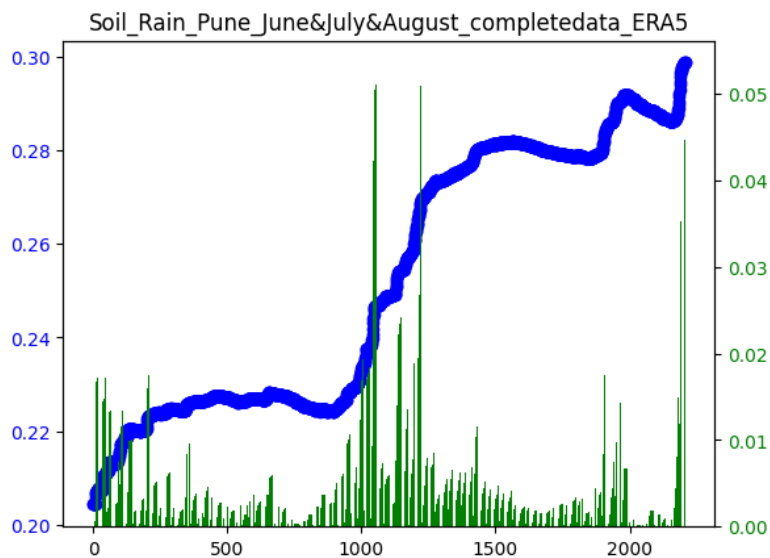
Average_on_4days_soil_Rain_Linear_Regression_Pune_complete_June&July&August_ERA5

Average_on_4days_rain_Rain_Linear_Regression_Pune_complete_June&July&August_ERA5
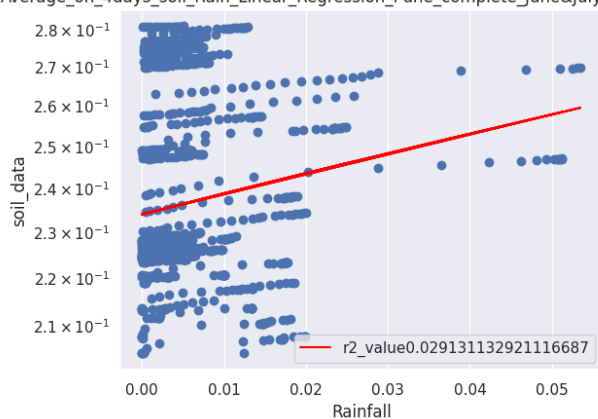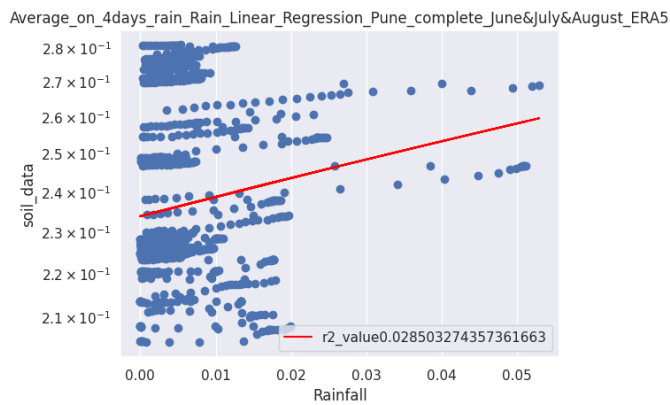
- Plotting rain and soil water volume together to see relation:

Now I put rain fall as bar plot and soil water volume as scatter plot.

Soil_Rain_Pune_June&July&August_completedata_ERA5

From this it seems there is some relation among rainfall and soil water volume. As rainfall happen soil water volume increases and after some time it decreases. The curve follows this concept. To see this deeply I divided the plot into some parts like this
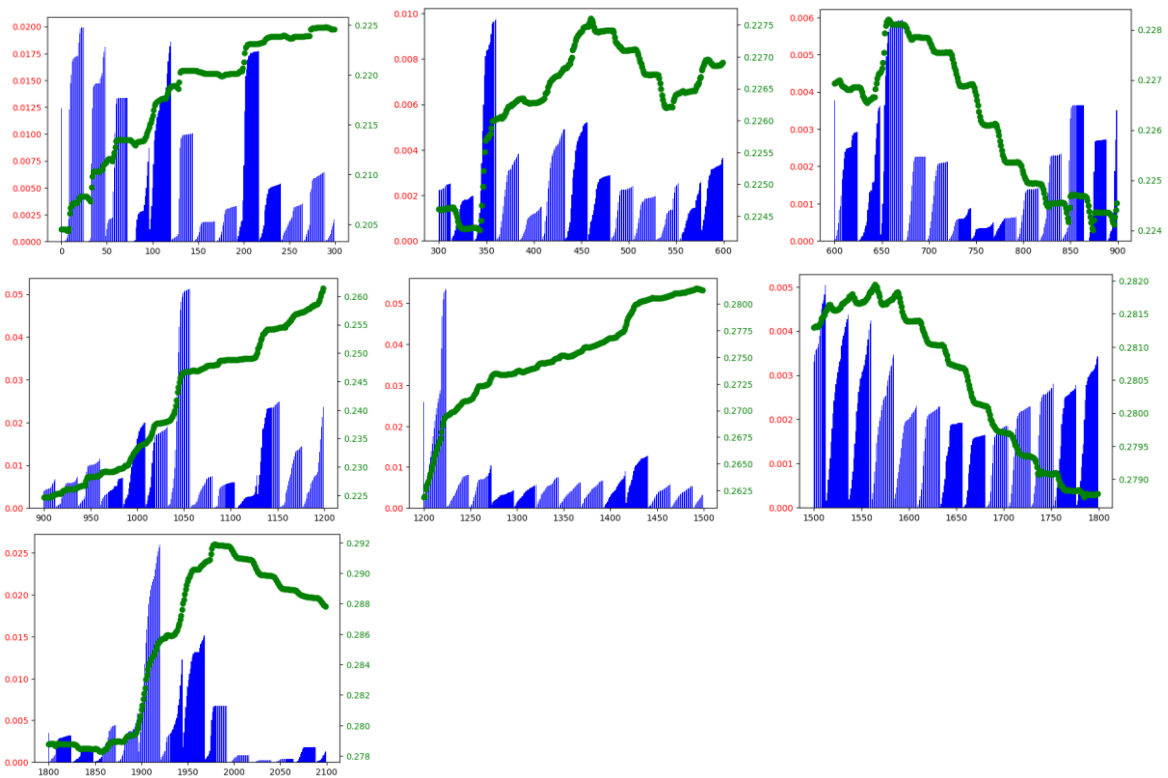
These plots also support the analogy which resembles RC circuit.

- Plotting these plots on daily data:

Now I averaged these data on daily basis for these 3 months and put them together and I got these plots
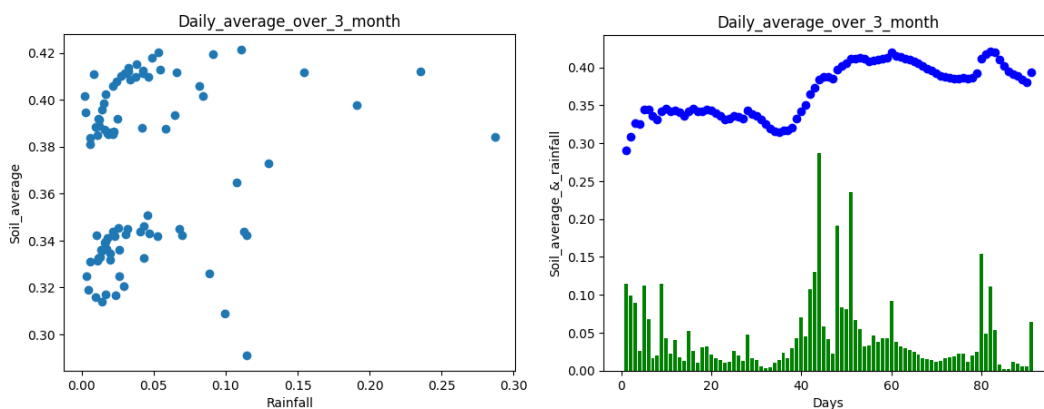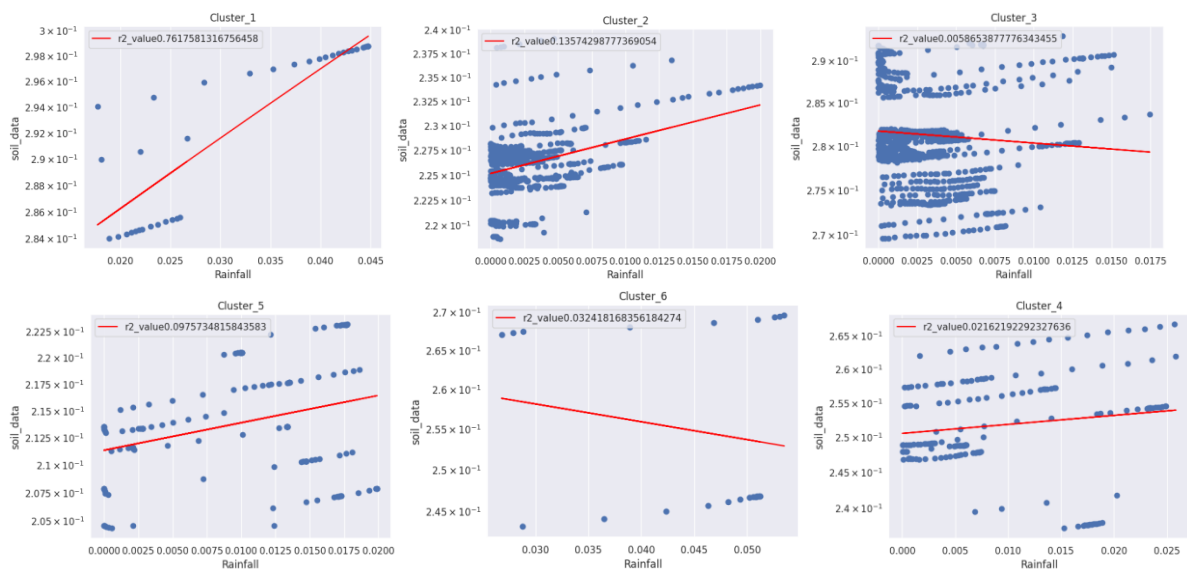
- K-mean clustering:

Then I tried to do K-mean clustering to find some good relation at least in some clusters. I did 6-mean cluster sampling and it looks something like this
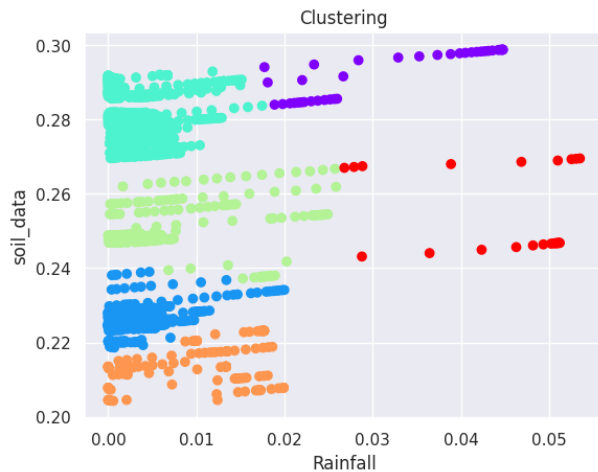
Now I plot linear regression in each of the cluster

It seems it is slightly better in some cluster like in cluster 1 and cluster 2 but overall not so significant development in this process also.

➢ Codes:

Here is the link where I combined all the codes that I had made to do the project:

https://colab.research.google.com/drive/1zebdxHm9g8pwd8USzbUmjPFclQz2jaP1?usp=share_link

➢ Conclusion:

So throughout the project I did different kind of plotting to understand the behaviour of rainfall and soil moisture and try to get some idea about their correlation. As I mention before through the plotting of the soil water volume and rainfall some correlation is found which tells about the memory of the soil. Now from the analogy of RC circuit we can try to develop some model about which will help in predict this.