To complete Objective one I used GCP cloud.

First you create vm machine on GCP.

I follow below documents to create vm and install jupyter in vm

https://sinanartun.medium.com/how-to-setup-jupyter-notebook-on-google-cloud-platform-42751e585fc7

Then I create a folder

mkdir scraping (make folder)

Nano file.py (create python file)

Copy code from file.py (which I give to you) and paste here

Press ctrl+s (for save)

Ctrl+x (exit)

The run command

Python file.py

I used tmux to run script to keep script in running even my system is off.

I follow below video to do above :

https://www.youtube.com/watch?v=lEKp2O7MTfY&t=1s

tmux new -s demo

(to create session, here session name is demo)

demotmux ls (to check how many session is running)

demotmux a -t demo (to check the script is running or not)

After scraping all data,

```
96
97
98
99
100
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Counter exceeded 100. Stopping the loop.
Data saved to 'github_repositories1.csv'
```

The reason of these extra lines : counter exceed 100.stop....

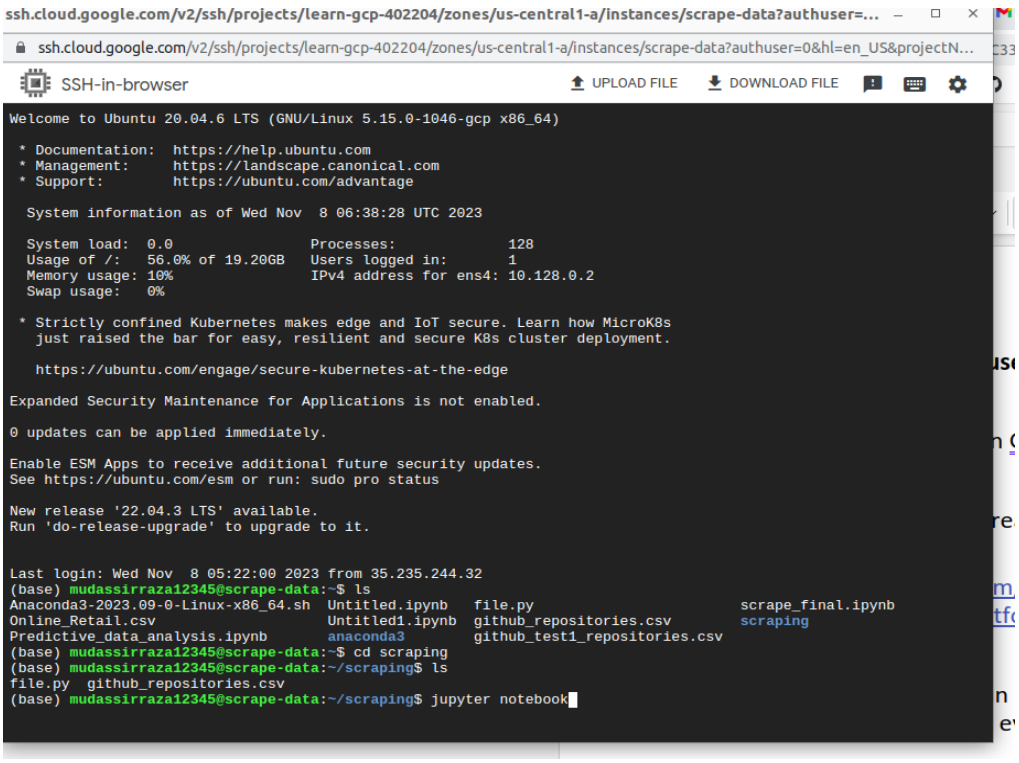
The thread is waiting till all data fatching then it will shutdown

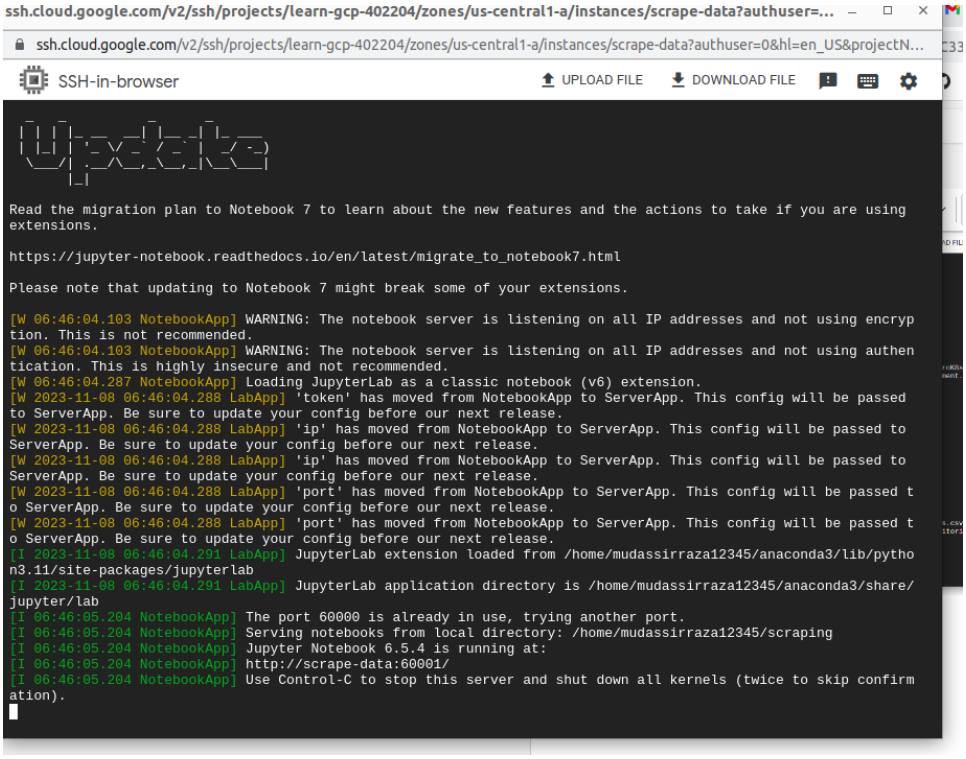
That why come but it will stop after all fetching by threads

I will start exploratory data analysis.

run command `jupyter-notebook`

In the scrapping folder

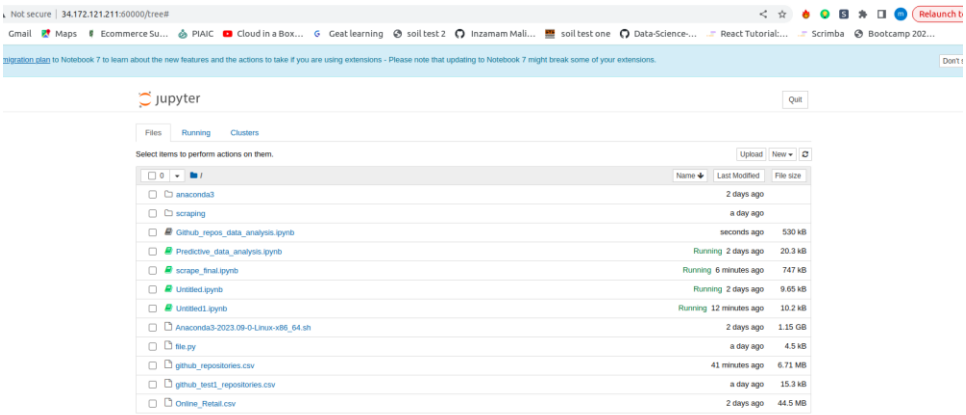




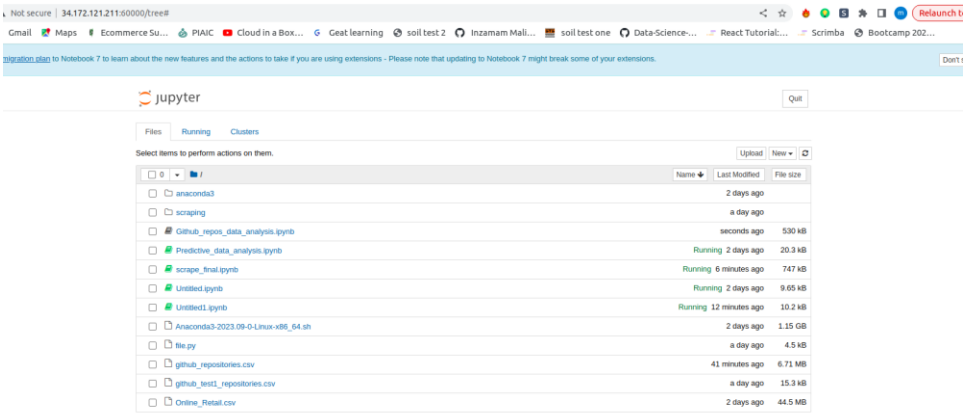
Then go to google

http://<your vm ip>:60000/

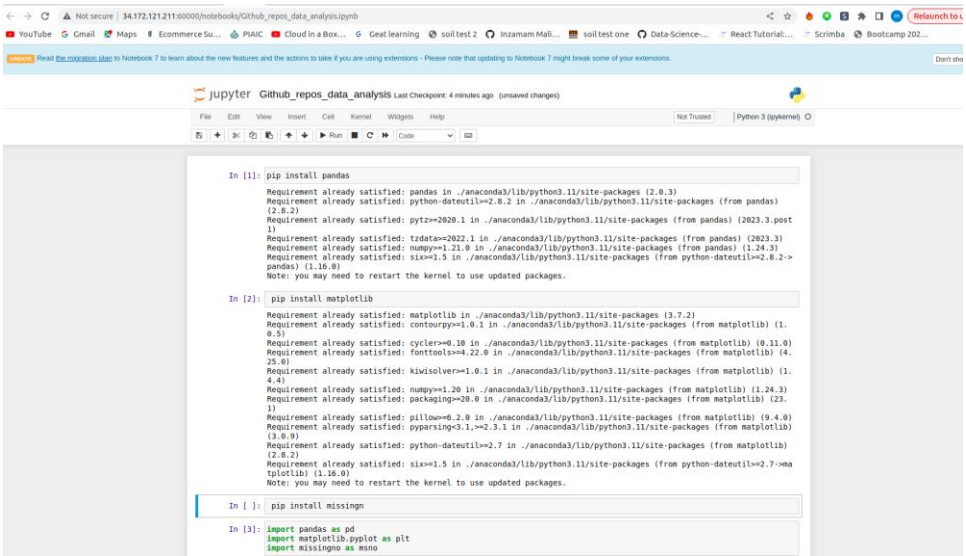
This juptyer terminal is open



click to upload and select the scrape data analysis notebook from system which I send you and upload.



Then click to Github_repos_data_analysis.ipynb



This screen come and then click to cell and run each cell by shift+Enter