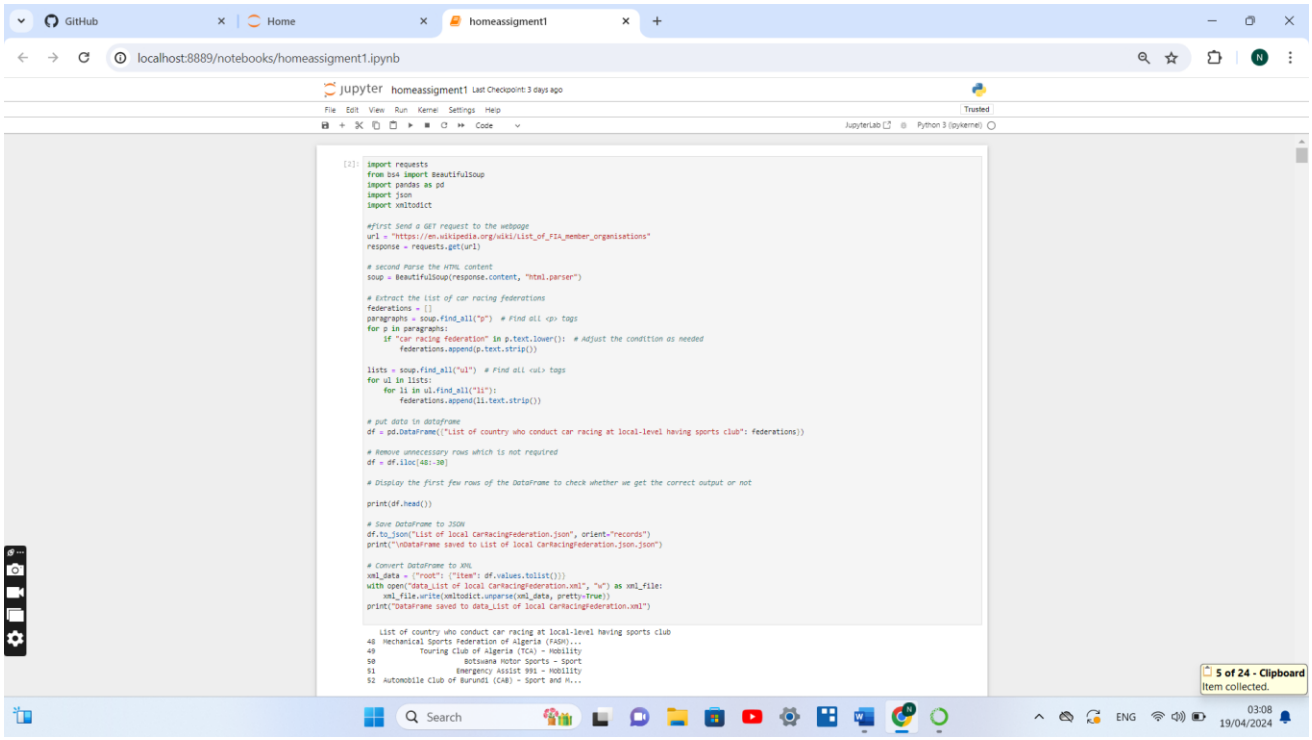
Report for HomeAssigment1

The code is written in Jupiter's notebook. Here is a brief of what work has been performed in the code

1. In the initial stage as per the requirement of the task data has been collected from the " **list of FIA member organizations"**. "The FIA is **the governing body for world motor sport** and the federation of the world's leading motoring organizations" The list contains all the local federation and clubs of a different country that comes under this or are members of this all listed down so the dataset taken from the source:-
https://en.wikipedia.org/wiki/List_of_FIA_member_organisation

And results have been stored based on the requirement in List of local CarRacingFederation.json and data_List of local CarRacingFederation.xml few rows just for checking are also printed on the screen here in this both are provided.



Steps used in this are

#first Send a GET request to the webpage

second Parse the HTML content

Extract the list of car racing federations

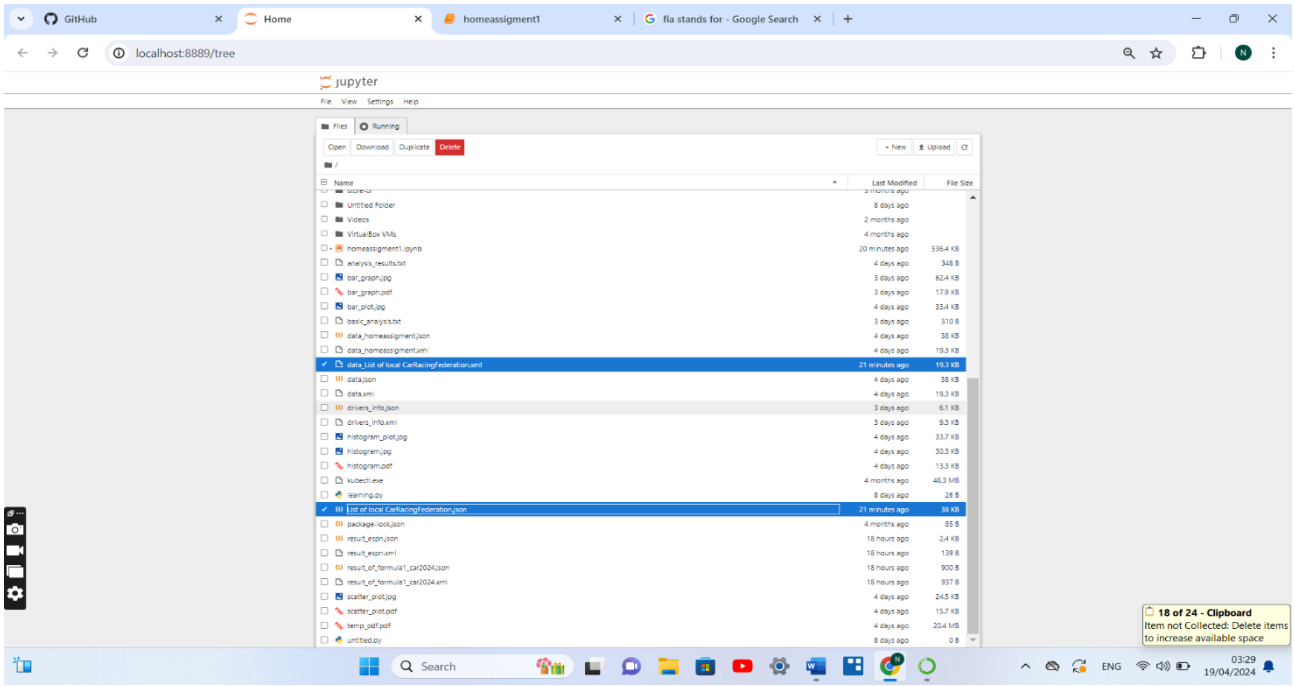
put data in dataframe

Remove unnecessary rows which is not required

Save DataFrame to JSON

Convert DataFrame to XML

These are also commented in the main program



2. next to full fill the other requirement. The extracted from next source which is https://www.fia.com/sites/default/files/brochure_sport_grant_programme_a4_v8_web.pdf which was pdf it contains information of grant and funds by FIA in 2020 and 2021 based on region and based on pillar division to local car racing federation and then this data first extracted and processed for further analysis here mean, median , mode correlation function have been performed data of 2020 and 2021 first merged based on region and also based on pillar divison then comparison made this analysis for visual reprentation stored in graph such as Bar Graph, histogram, scatter plot is done this visual analysis is stored both in jpg and pdf format with the following name “bar_plot_funding based on region 2020 and 2021.jpg”, “histogram_plot_funding based on region 2020 and 2021.jpg”,

“scatter_plot_funding based on region 2020 and 2021”.jpg, “bar_plot_funding based on region 2020 and 2021”.pdf, “histogram_plot_funding based on region 2020 and 2021”.pdf, “scatter_plot_funding based on region 2020 and 2021.jpg” have been stored.

Basic analysis such as mean, median, mode, correlation, and statistical analysis are stored in a text file with the “basic analysis.txt” For these all data extracted from pdf this pdf is also stored which was extracted from the source with the name “temp.pdf”.

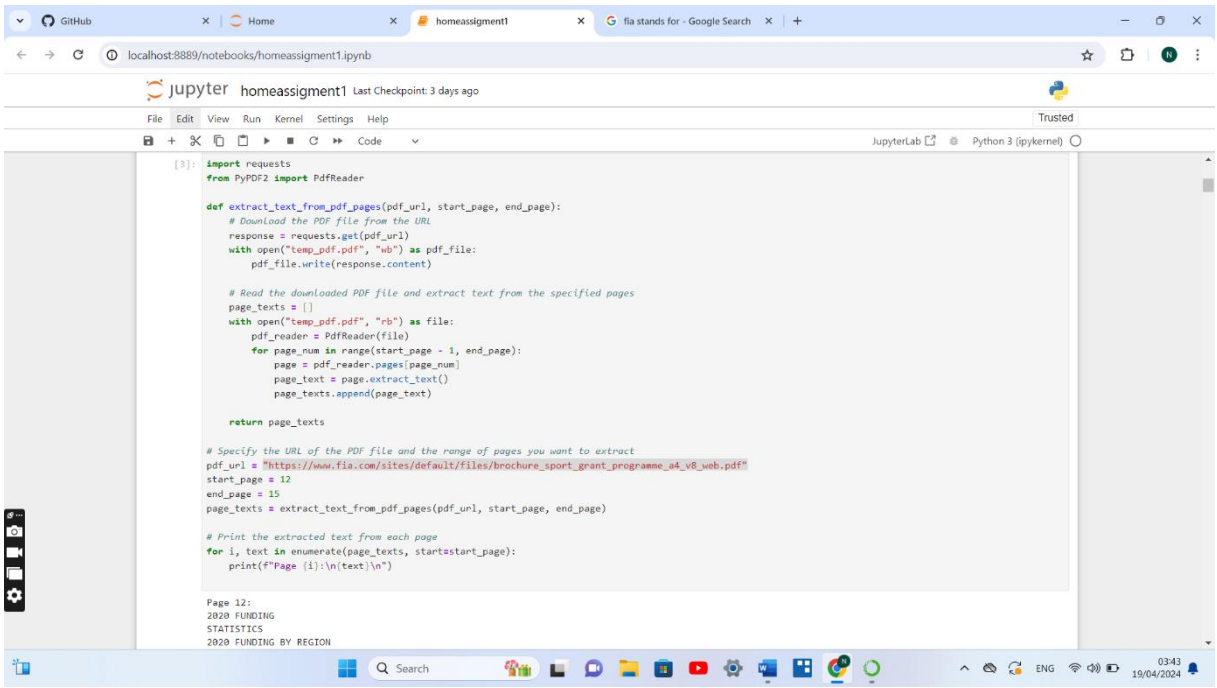
The steps used for this are

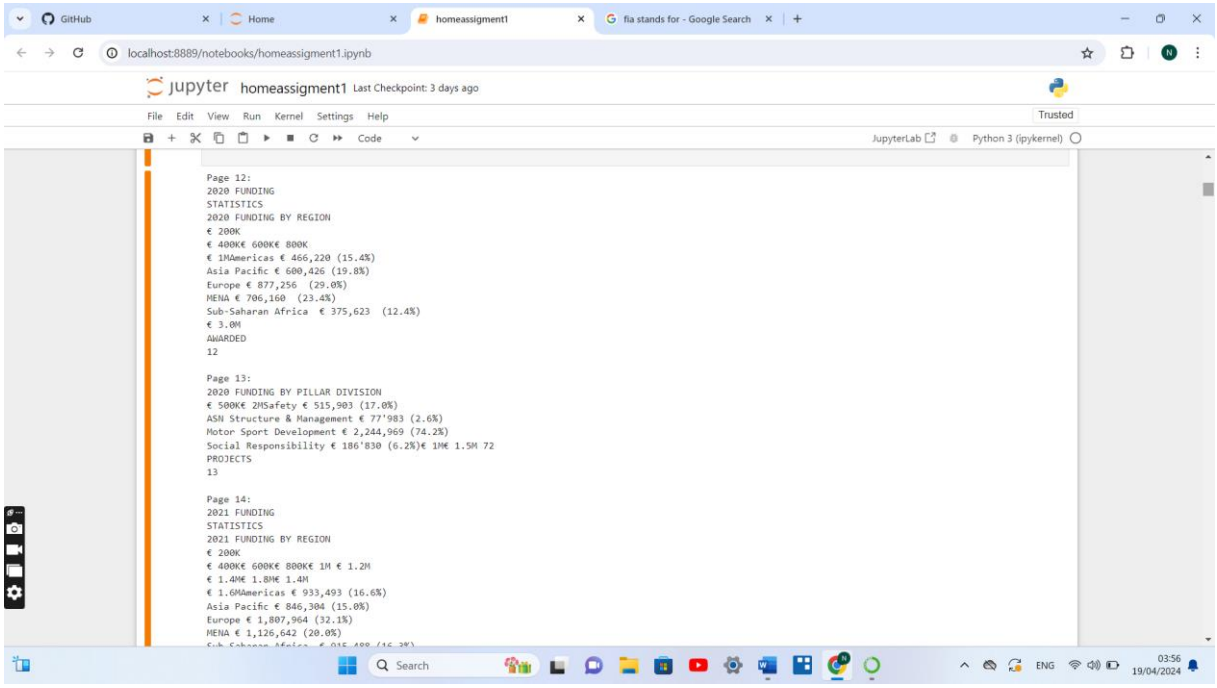
data extracted from the source which was pdf

Then download the PDF file from the URL

Read the downloaded PDF file and extract text from the specified pages

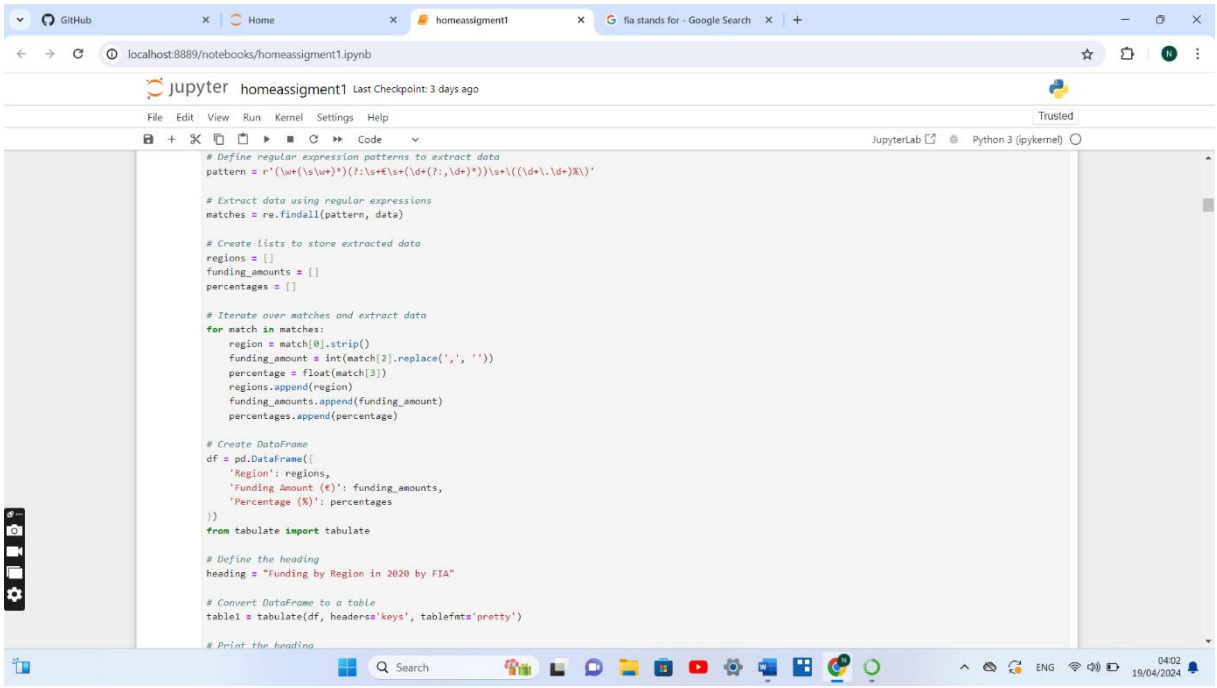
Print the extracted text from each page

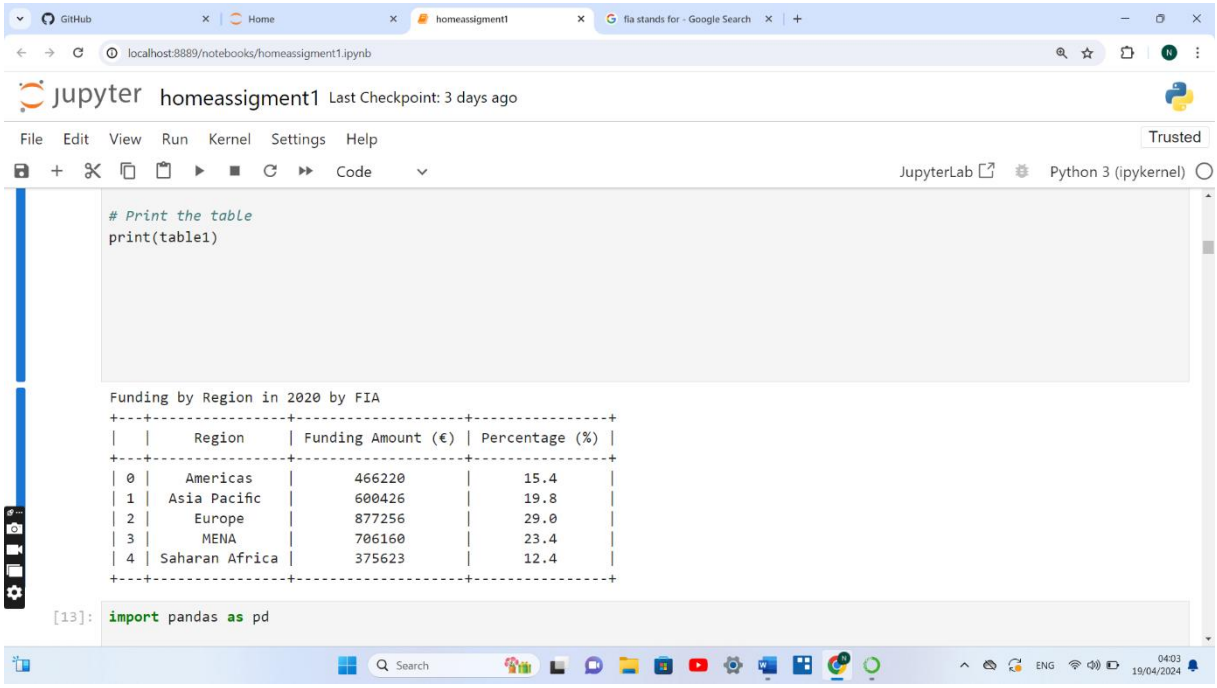




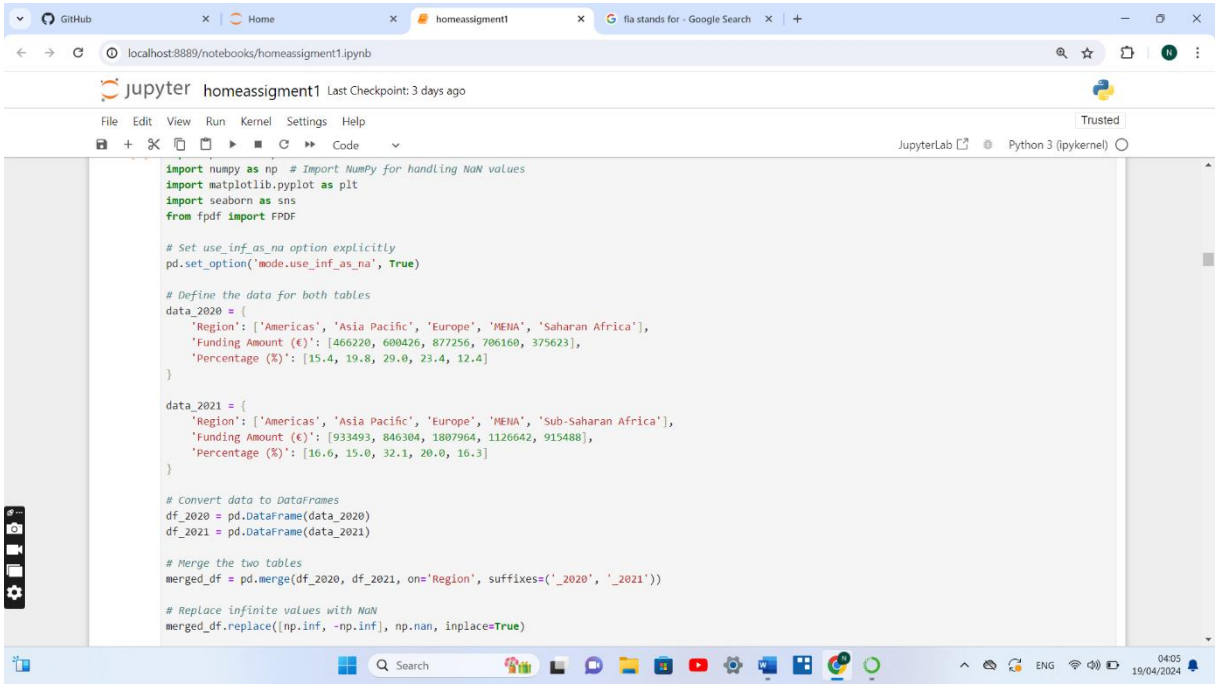
Then after extracting data from the pdf and having output

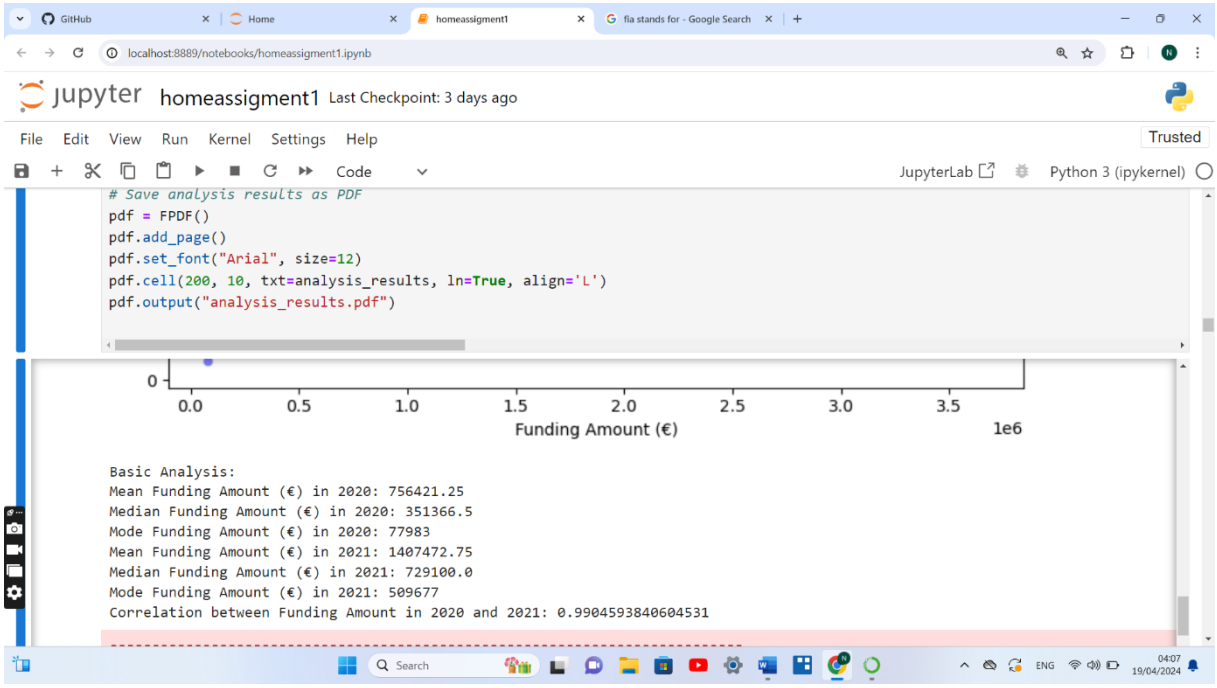
The next step which was performed was forming data frame and after that from data frame form table to make the things more representable. Here to perform the things regular expressions are used.



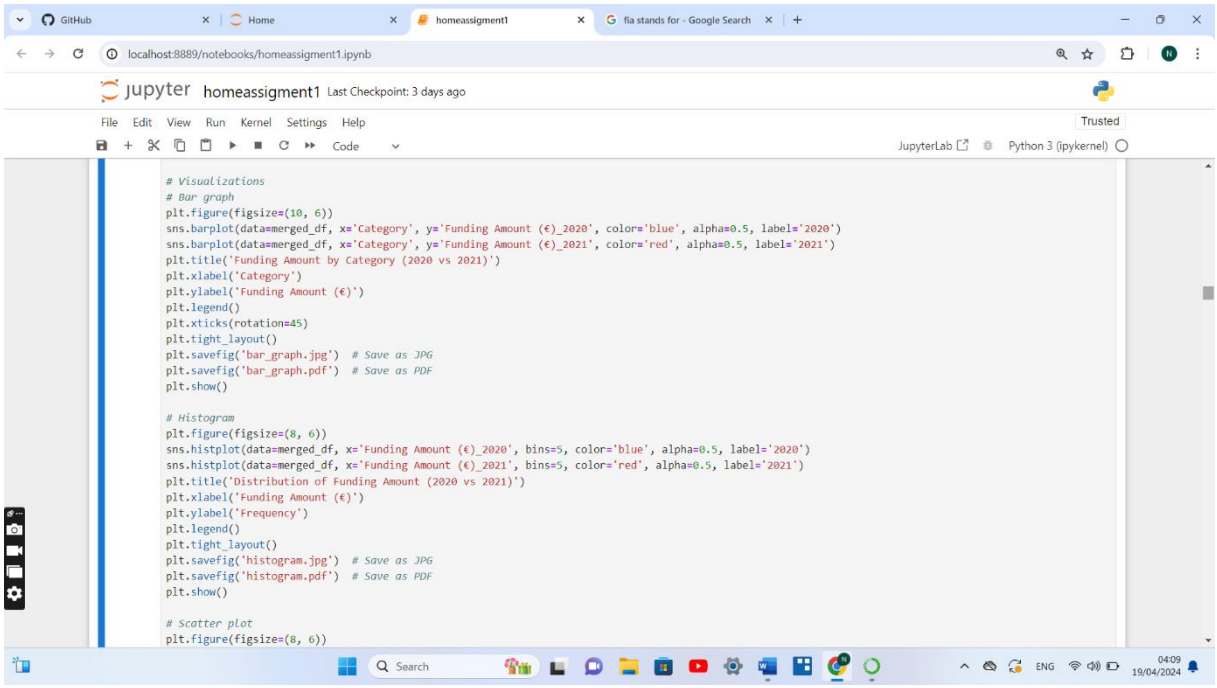


After this 2020 and 2021 data was merged for analysis with this preprocessing is also done.

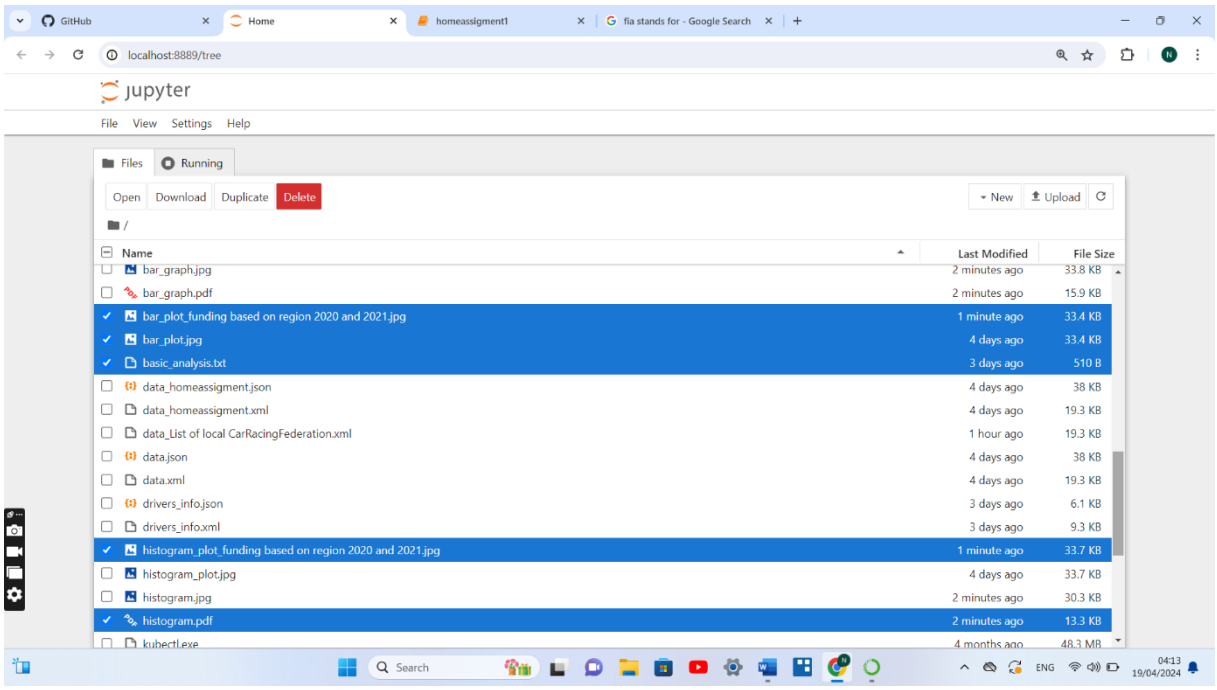




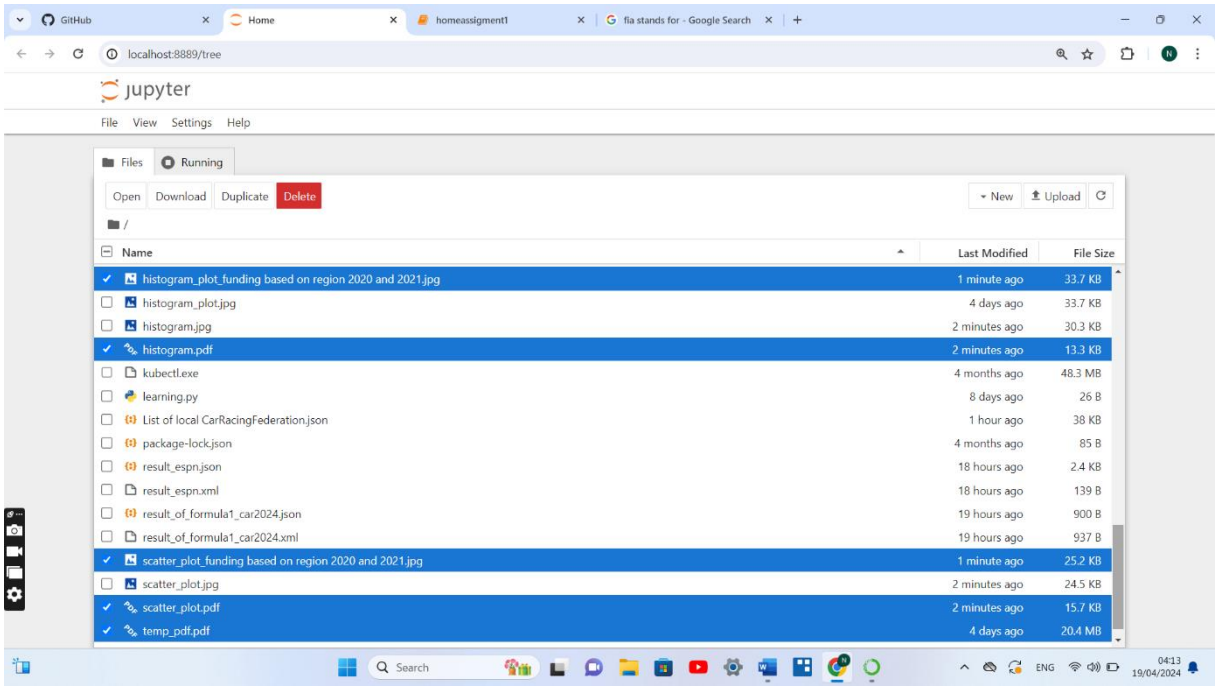
After this data visulaisation task was performed



Results are stored in above mentioned file in both format



.

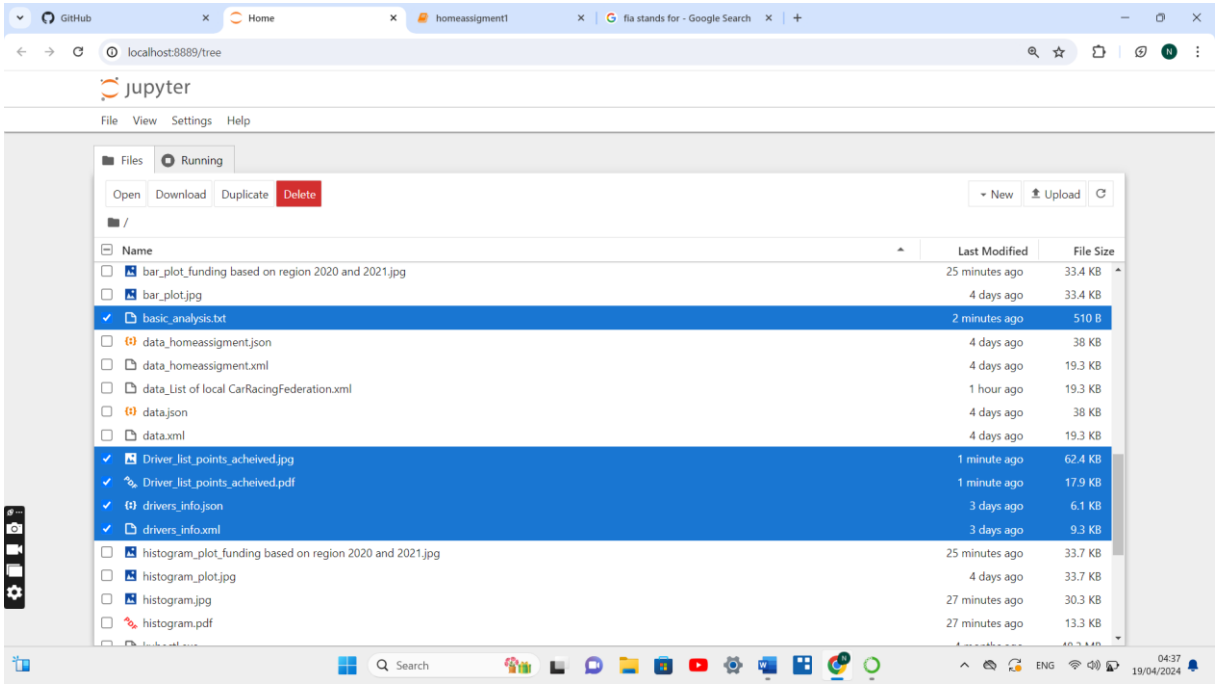


3. then data extracted from 3rd data source list of drivers for participate in local or internation car racing race in different races their performance are first stored in data frame to make it more representable data stored in table then for analysis graph have been formed this data stored in both xml and json format.

Source:- https://www.motorsport.com/f1/results/2024/japanese-gp-639931/

Graph stored with name:-Driver_list_points_acheived.jpg, Driver_list_points_acheived.pdf

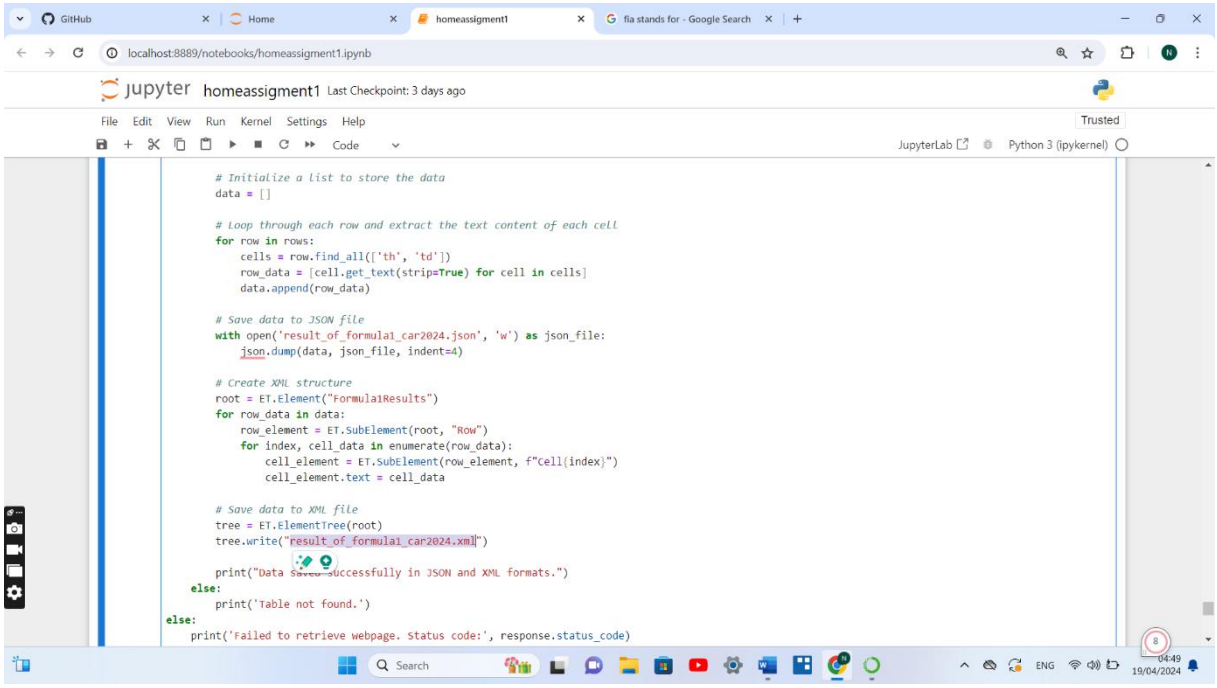
Data extracted stored in json and xml with the name :- drivers_info.xml", drivers_info.json



4.Then at the end of the assignment two data set is extracted from two sources for results of international car racing “Formula 1 “ 2024 and other car racing results such as “Espn”which is supported and govern at local level the data was extracted and stored both as xml and json format to understand better data extracted converted into data frame for better analysis.

Source:- https://www.formula1.com/en/results.html

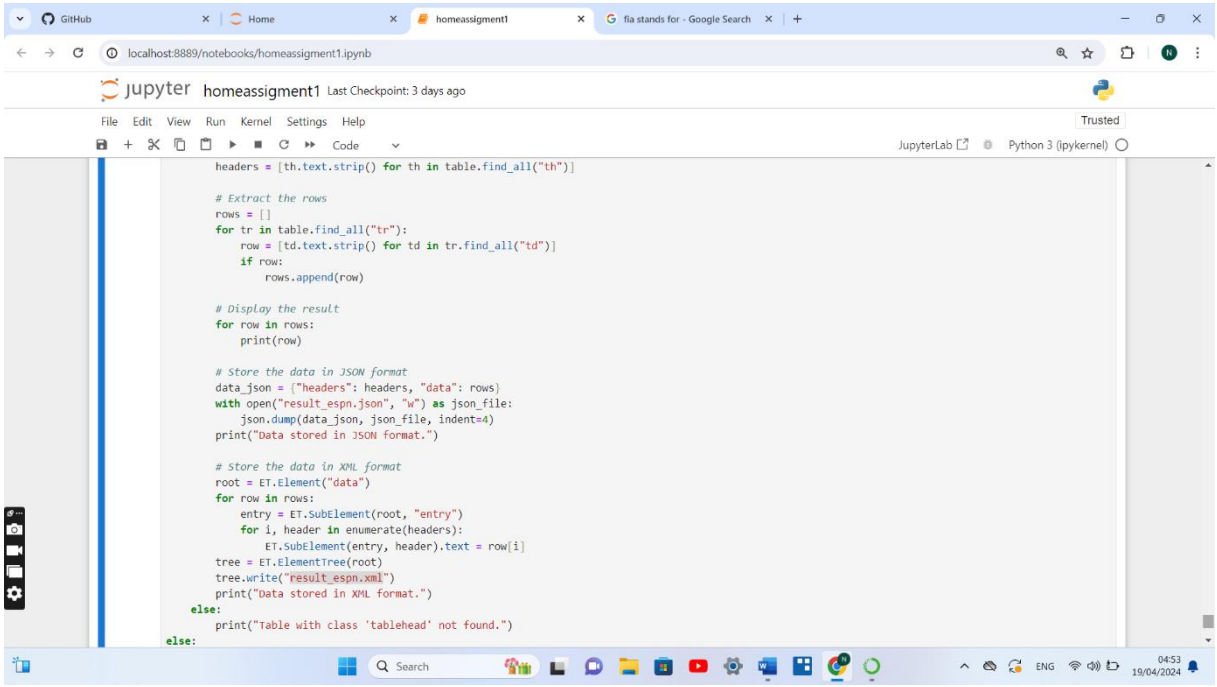
Data stored with the file name:- result_of_formula1_car2024.json, result_of_formula1_car2024.xml.

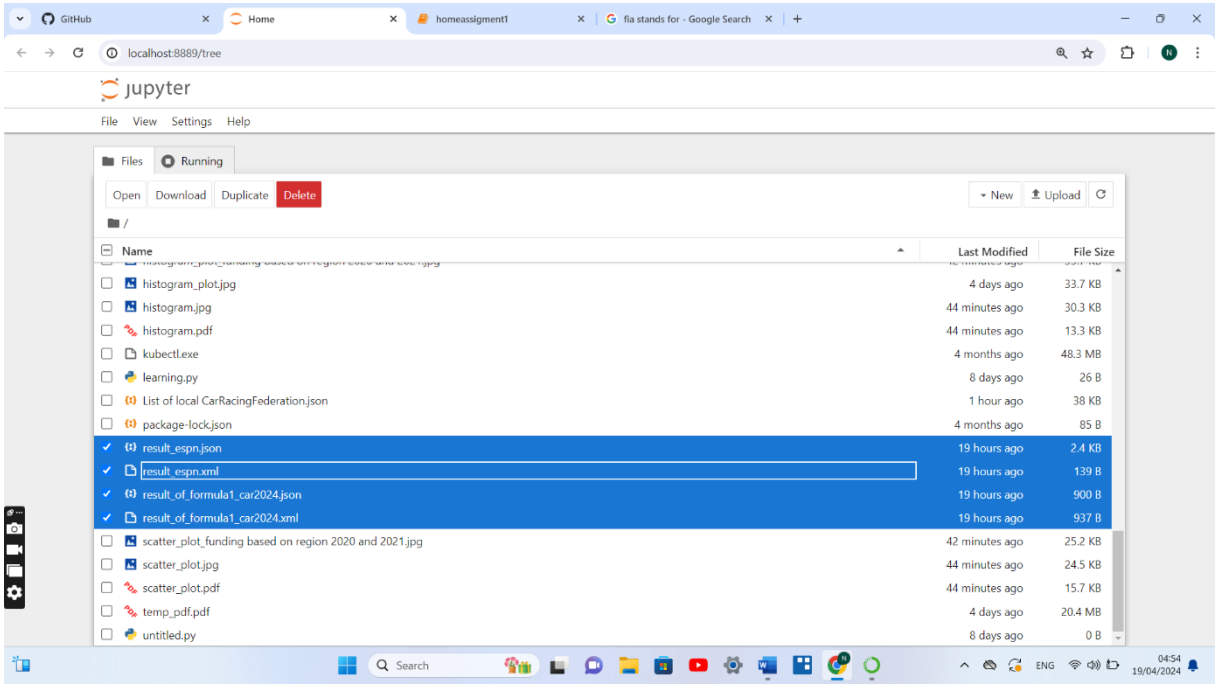


Other data source as mentioned above

Source:-“ https://www.espn.com/racing/results

Data stored :- result_espn.json, result_espn.xml





Overall there were 5 data sources were collected and data were analyzed from different data sources as per the requirement and stored in different file formats as required to represent different features of analysis, such as funds driver information local and international car racing results or a list of international and local car racing federation and clubs are represented in the task which was performed.