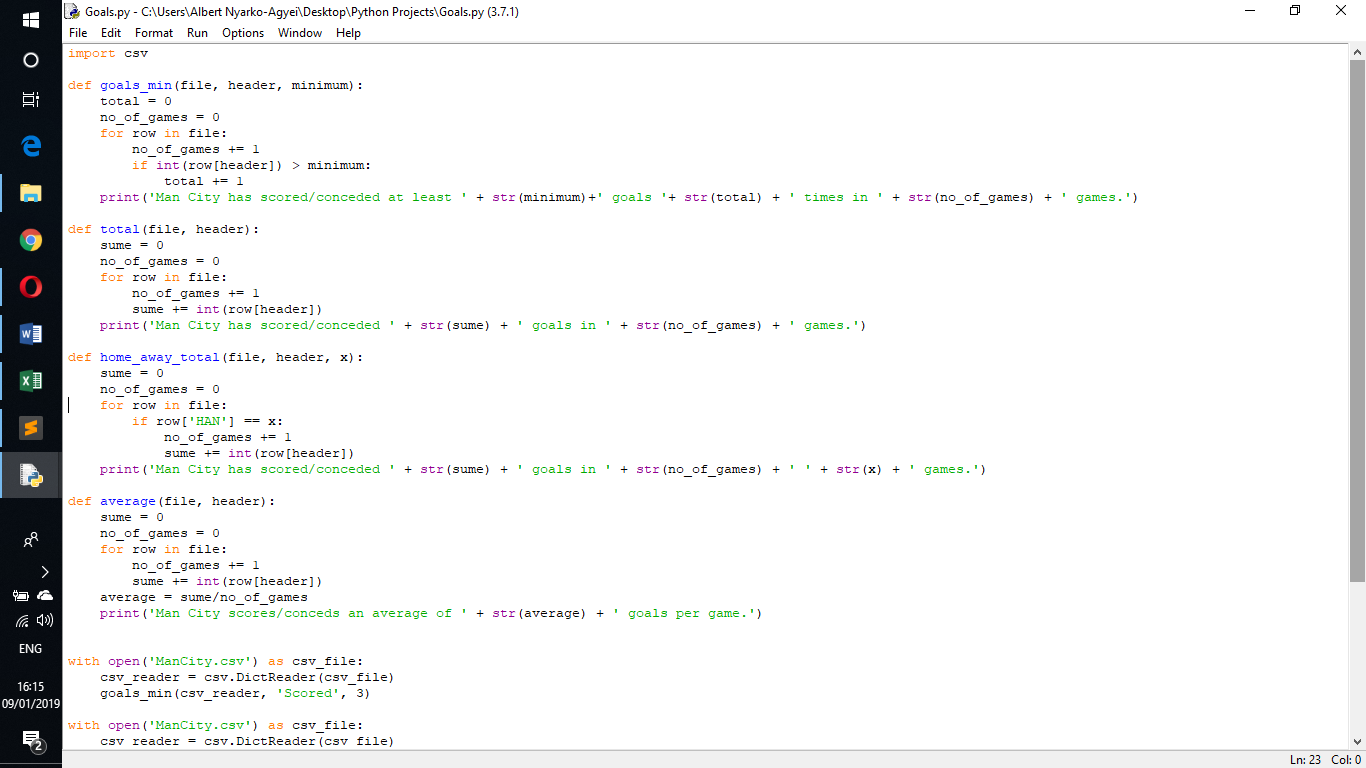
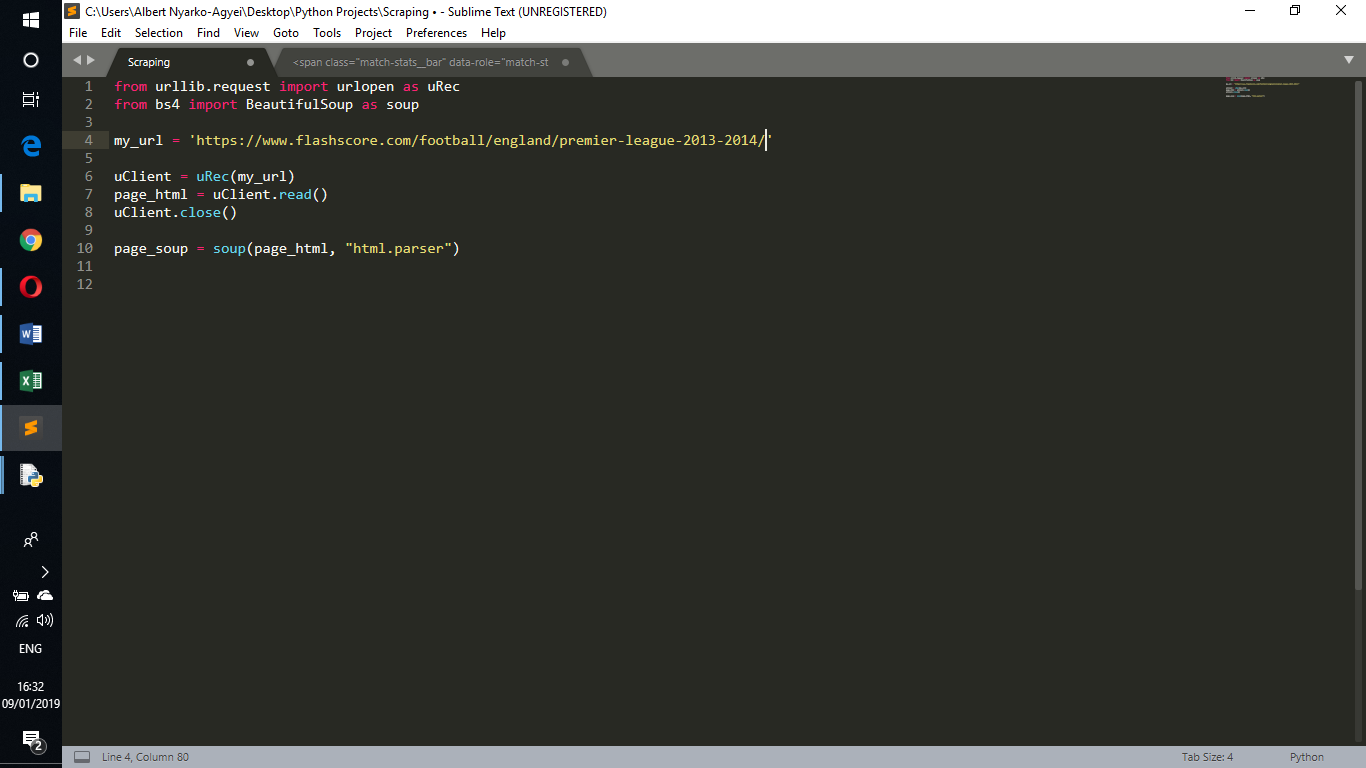


The aim of this project is to extract patterns from football match stats. I started off by entering the data manually but this was taking too long so I turned to web scrapping.



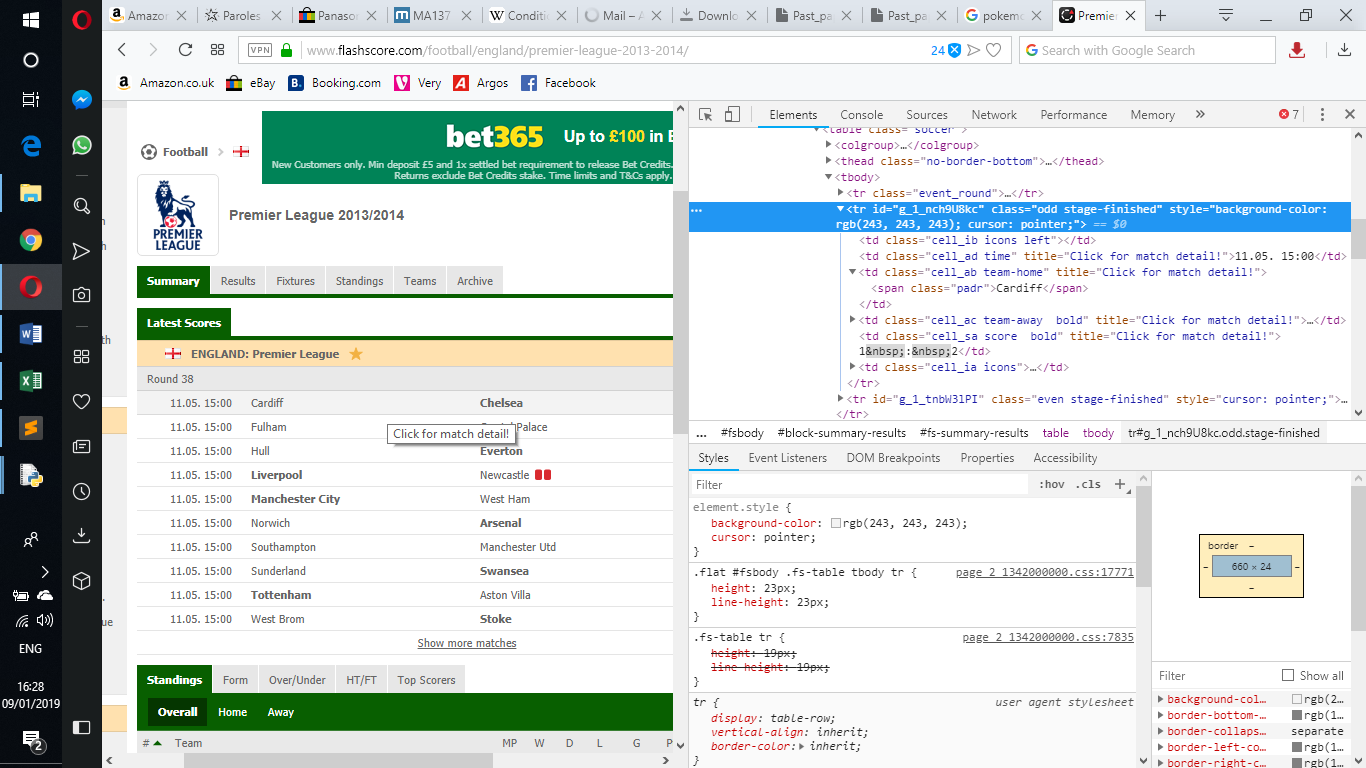
This is some of the code I have written to analyse the csv’s once they are complete with the data.

One problem here is that I always have to open the csv for each function I run otherwise there is an error. Is there a way to leave the csv open and run numerous functions?



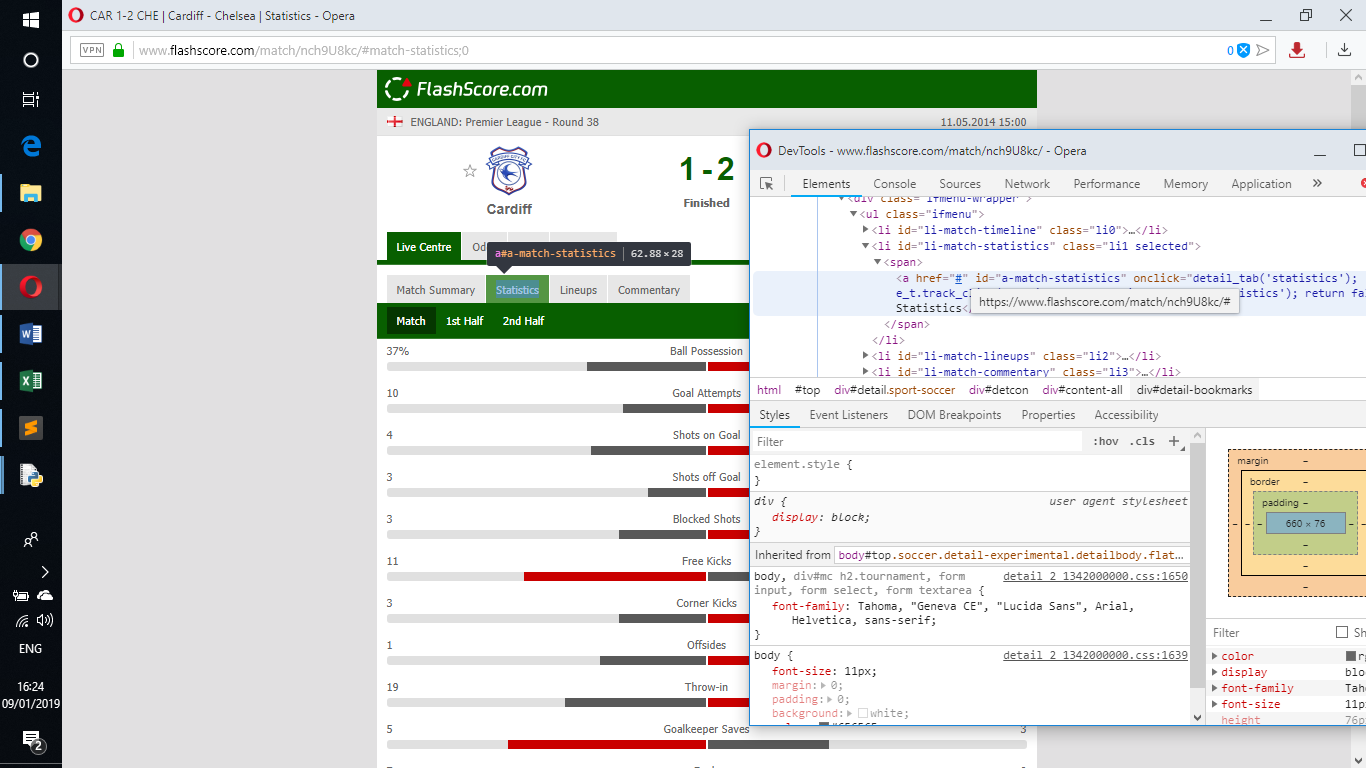
Currently this is the code I used to examine the webpages where I get my data from. It doesn’t work in the normal python IDLE so I have to run it in the windows cmd.

This is the way to locate the data within each webpage so once you have run this code you then type page\_soup.body.div + whatever else that will lead you to the information you need. This is what I am having problems with because I do not understand which tags my data is under.

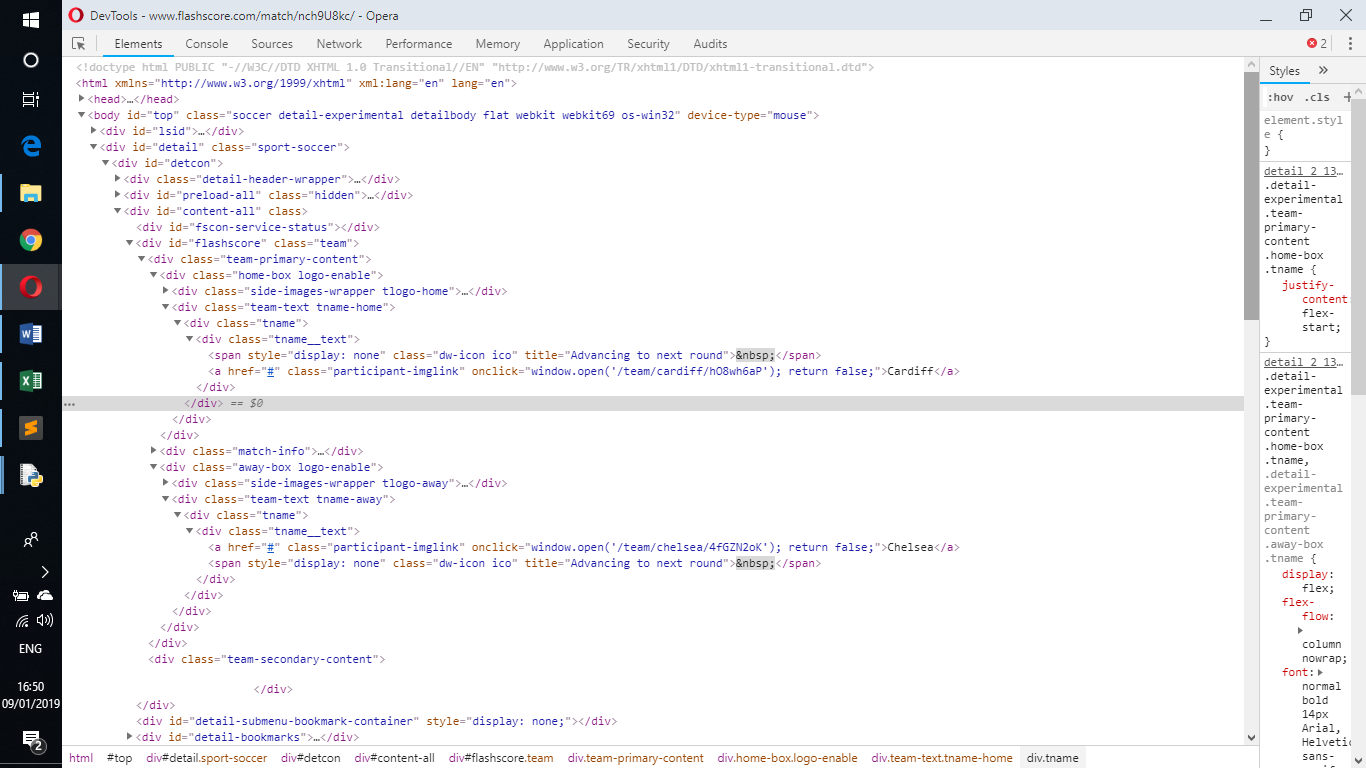


This is the webpage I am sourcing my data from. Whenever I click a row in the table another page opens up so there must be a “href” for each row.

One step is to get the program to loop through all the individual hrefs and open the match statistics before scrapping can begin. But my main problem is finding the tag refers to the right part of the webpage.



This is an example of one of the pages opened when a row in the table is clicked and it contains the actual information I need.



One of the first things I will need to locate in the webpage are the team names. But I don’t know which tags will allow me to extract that specific piece of data.