This is an old code of mine that I had used for an Internship Fair at my college. It is 100% written by me and I’ll explain the execution logic behind it.

The main libraries used are BeautifulSoup and requests. Requests is used to make GET requests on the URL and BeautifulSoup is used to parse and extract data from it. There is an alternative module that I have also used called scrapy which is very fast and requires minimal coding.

So I found the base URL of Internshala which I passed through request and BeautifulSoup modules. I also found out the total number of pages and the specific change in the URL so that I could mimic that inside the loop. I then find out the div tags I need from the Inspector menu on Google Chrome and pass it into BeautifulSoup to find the company details and stipend. Since the company part was already ready with me, I just modified the code to add the stipend part and therefore they are looped separately. I then save all the details inside a list called all\_companies.  
  
Finally to save the data in csv format, I use the python csv library and iterate over the all\_companies list after adding the header.

That concludes my code. Also I have added a sleep function after every GET request to avoid being blocked by the firewall of Internshala.

This is all. It originally took me 2-3 hours to write this code about a year ago and took me 30 mins now to brush up and add the stipend part. I have also added comments to explain the code better