

After reading the assignment, I had no idea of what could be done, the first thing I did was to do a google search.

I was just searching on how to extract information from a website using python, as well as sort the information to be able to achieve the goal.

From my research, I realised that the process of retrieving information from a website is called web-scraping.

I then narrowed down my research to how web scraping is done using python.

The process was tedious because as I dig deeper into the subject matter, I realised that there are many ways and many methods to use for this process.

- ❖ First I used QBSC to break down the task, so as to clearly map out what I have to do, what results I need and which tools will help me achieve them.
- ❖ Secondly, I took time to come up with an algorithm, which is drawing out a sketch of how the program is going to run.
- ❖ Next, I sourced for and downloaded the libraries that I identified that I will need for the program.
- ❖ Following the libraries, I identified processes that will run in the program, then I wrote them in as functions.
- ❖ After putting the above together, I then started writing the code.
- ❖ During the process of coding, I had a lot of errors but at each time, I copied the error, posted it on google and then saw how other developers on platforms like StackOverflow who have encountered errors before were able to solve it.
- ❖ After completing the code, and running it for final testing, I noticed some attribute errors so I used the try and except functions to effectively handle the error.
- ❖ When writing the program, I made a couple of assumptions based on the understanding I had of the task we had to do. I made the assumption that most popular books are books that

have many ratings. This is because taking the place of a user, I find it easier to just give a 5-star rating than to type a paragraph of text to give a written review about a book. Again, I had the understanding that many book buyers and readers simply look at the ratings of books before buying them.

- ❖ One major roadblock that I faced was when I had to extract the information about the books from the website. I tried several web drivers, and many other methods, but the google chrome on my pc was not up to date so it could not work. Nevertheless, I used the terminal to update the software, download new dependencies.

Now when it comes to the libraries I used and why I choose to use them;

- ☐ The very first library I installed was the BeautifulSoup library. BeautifulSoup is a Python library that is used to pull the data out of HTML and XML files. It enables us to create a parse tree from page source code that can be used to extract data in a hierarchical and more readable manner. I choose to use it because it is fast, lenient, and conductively parses the code as a web browser would do.
- ☐ The next library I used was the request library. The requests library is the common standard for making HTTP requests in Python. It permits us to take away the complexities of making requests by putting it into a simple API so that we can simply use the information that it returns.