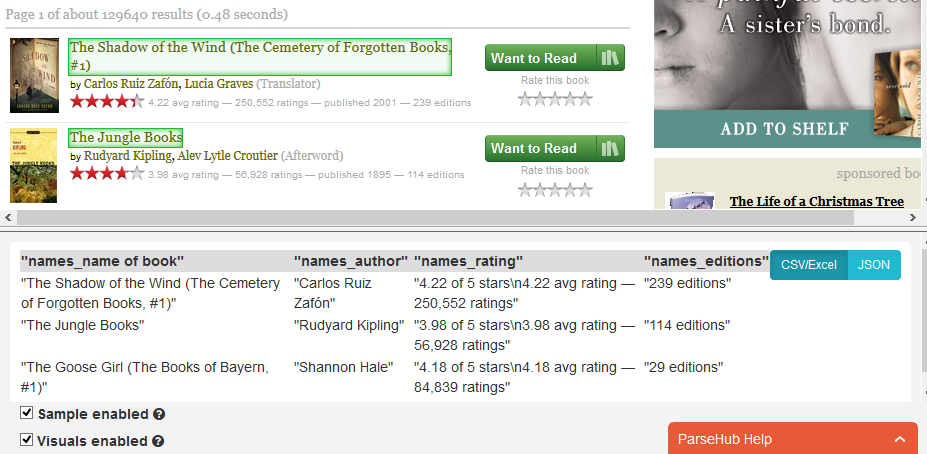
Task 1 <- Pick at least 3 web scraping toolkits and use them to extract the data from a web site of your choice

I tried many tools but few were easier to use then others. The three toolkits I chose to write the report on, are:

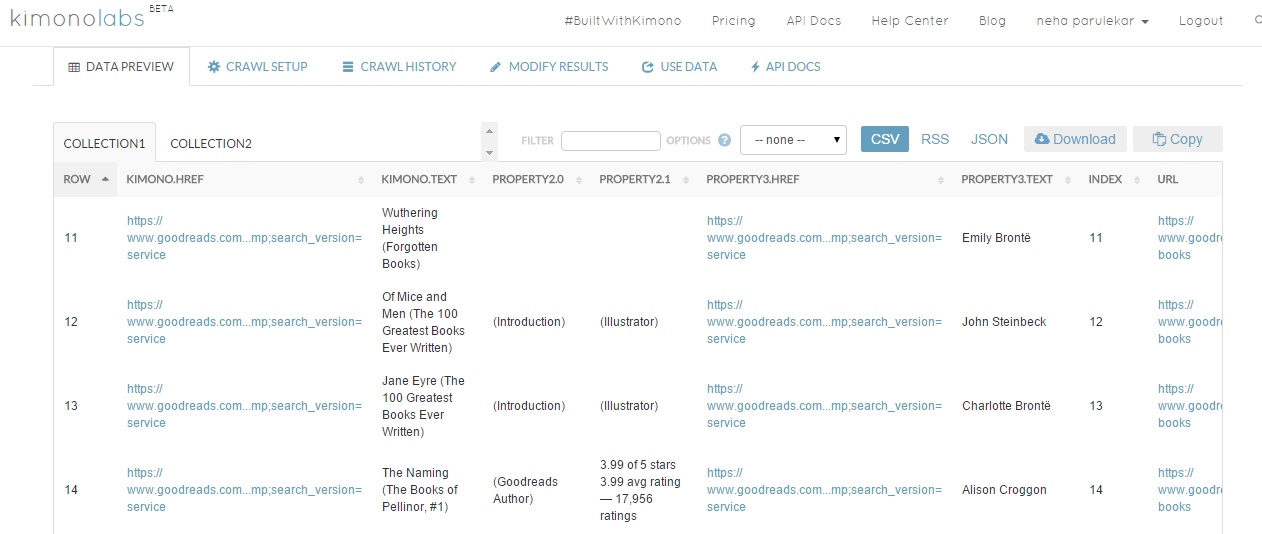
* Kimono
* Parse Hub
* Import.Io
* I chose Goodreads.com to Scrape. My aim was to get the name of the book, the author name, rating and No of editions for the books.
* The screen shot below is of Parse hub:

The toolkit is available for free use as well as for Premium and Professional use as well. The tool was kind of difficult to use as one has to adhere to the steps and it takes getting use to. One thing which I would like to bring to attention is it only works with Firefox. Which kind of is a negative as you have to specially download Firefox if you don’t use it. One of the positive things it tutorials which help learing a lot easier



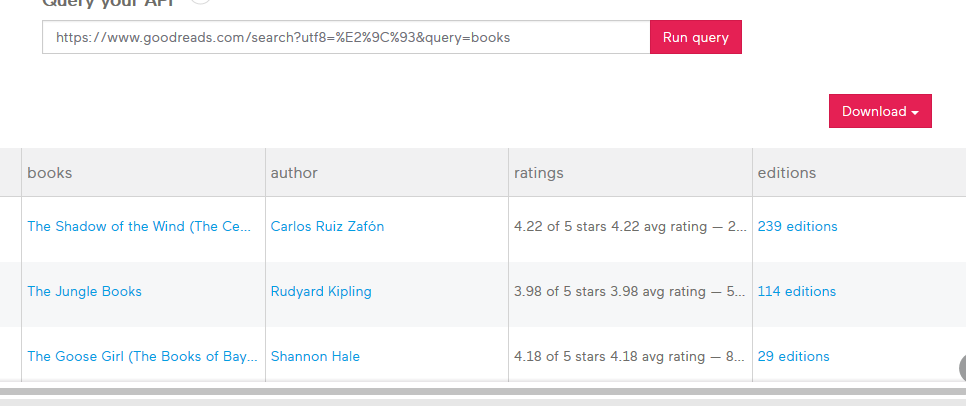
* The screen shot below is of Kimono:

Free API is available for Kimono. The toolkit is pretty easy to use. The purpose of Kimono is to convert unorganized data to an organized data. Kimono Labs allows you to extract this data either on demand or as a scheduled job. Once you've extracted the data, it then allows you to either download it via a file or extract it via their own API. This is where Kimono really shines—it basically allows you to take any website or data source and turn it into an API or automated export.



* The screen shot below is of Import.io:

The tool kit is pretty user friendly. The setup takes a little time. But the tutorials available help you get the desired results in a very short time.  The data that users collect is stored on import.io’s cloud servers and can be downloaded as CSV, Excel, Google Sheets or JSON and shared. Users can also generate an API from the data allowing them to easily integrate live web data into their own applications or third party analytics and visualization software.



Recommendation:

Though It is not very easy to and takes getting used to I prefer Pasre hub. Once you get a hang of it, the tool is very efficient in extracting data required. Based on the data I could scrape,With the other two I would still need to filter the data. I got the exact results I was aiming for in Parse Hub.

Screenshot of the dataframe of the Pasrehub extracted file in R

