DOCUMENTATION:

IT’S A WEB SCRAPING PROJECT USING PYTHON EXTRACTING USEFUL DATA FROM THE E-COMMERCE SITE [www.newegg.com](http://www.newegg.com)

and in that extraction of information about Graphics Cards like product name ,shipping price, brand name, product price etc. in the form of tables in the excel spreadsheet using anaconda prompt.

Using javascript beautifier, we correct the Html code in Sublime text.

Web scraping :using beautifulsoup library(bs4)

**Web scraping**, **web harvesting**, or **web data extraction** is [data scraping](https://en.wikipedia.org/wiki/Data_scraping) used for [extracting data](https://en.wikipedia.org/wiki/Data_extraction) from [websites](https://en.wikipedia.org/wiki/Website).[[1]](https://en.wikipedia.org/wiki/Web_scraping#cite_note-Boeing2016JPER-1)Web scraping software may access the World Wide Web directly using the [Hypertext Transfer Protocol](https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol), or through a web browser. While web scraping can be done manually by a software user, the term typically refers to automate processes implemented using a [bot](https://en.wikipedia.org/wiki/Internet_bot) or [web crawler](https://en.wikipedia.org/wiki/Web_crawler). It is a form of copying, in which specific data is gathered and copied from the web, typically into a central local [database](https://en.wikipedia.org/wiki/Database) or spreadsheet, for later [retrieval](https://en.wikipedia.org/wiki/Data_retrieval) or [analysis](https://en.wikipedia.org/wiki/Data_analysis).

## Libraries required for web scraping

As we know, python is a open source programming language. You may find many libraries to perform one function. Hence, it is necessary to find the best to use library. I prefer **BeautifulSoup** (python library), since it is easy and intuitive to work on. Precisely, I’ll use two Python modules for scraping data:

* **Urllib2**: It is a Python module which can be used for fetching URLs. It defines functions and classes to help with URL actions (basic and digest authentication, redirections, cookies, etc). For more detail refer to the [documentation page](https://docs.python.org/2/library/urllib2.html).
* **BeautifulSoup:** It is an incredible tool for pulling out information from a webpage. You can use it to extract tables, lists, paragraph and you can also put filters to extract information from web pages. In this article, we will use latest version BeautifulSoup 4. You can look at the installation instruction in its [documentation page](http://www.crummy.com/software/BeautifulSoup/bs4/doc/).

BeautifulSoup does not fetch the web page for us. That’s why, I use urllib2 in combination with the BeautifulSoup library.