**Basic Steps**

1. dir to see directory

2. cd folder name to change the folder

5. scrapy to see the available commands

6. scrapy startproject gdp\_debt to start that new project

7. scrapy genspider gdp “website link” to create the spider file.

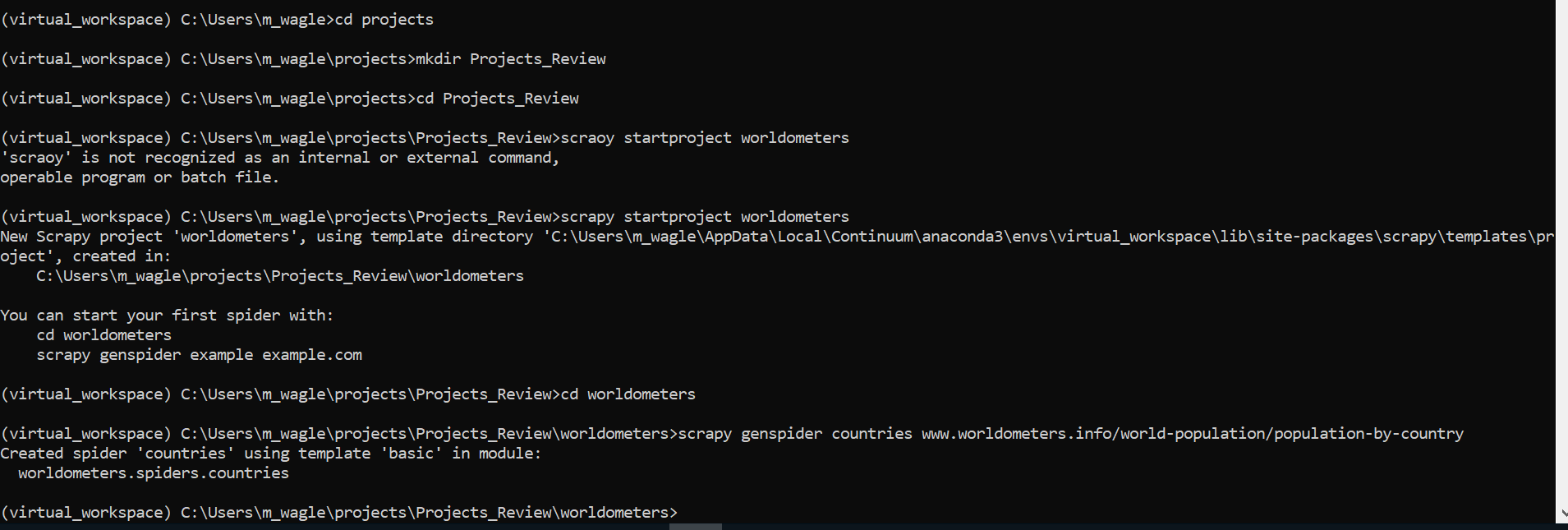
Here gdp is the name of spider. Remembet to remove / from end of website and https:// from beginning it should be like [www.manilwagle.com](http://www.manilwagle.com)

8. do all the work in spider file that was just generated, i.e gdp spider

9. scrapy crawl nameofspider(projectname) to run that spider file

10. scrapy crawl countries -o population\_dataset.json to export the file

**(From beginning to up to generating spider)**



**Before Scraping Any website**

1. Check to make sure the website is allowed to be scrapped

2. Do the following

* type webpage name tinydeal.com/robots.txt
* Ctrl+F and type in webpage name, if you can’t find the webpage name, it means you can scrap. If you find the webpage name, you aren’t allowed to scrap

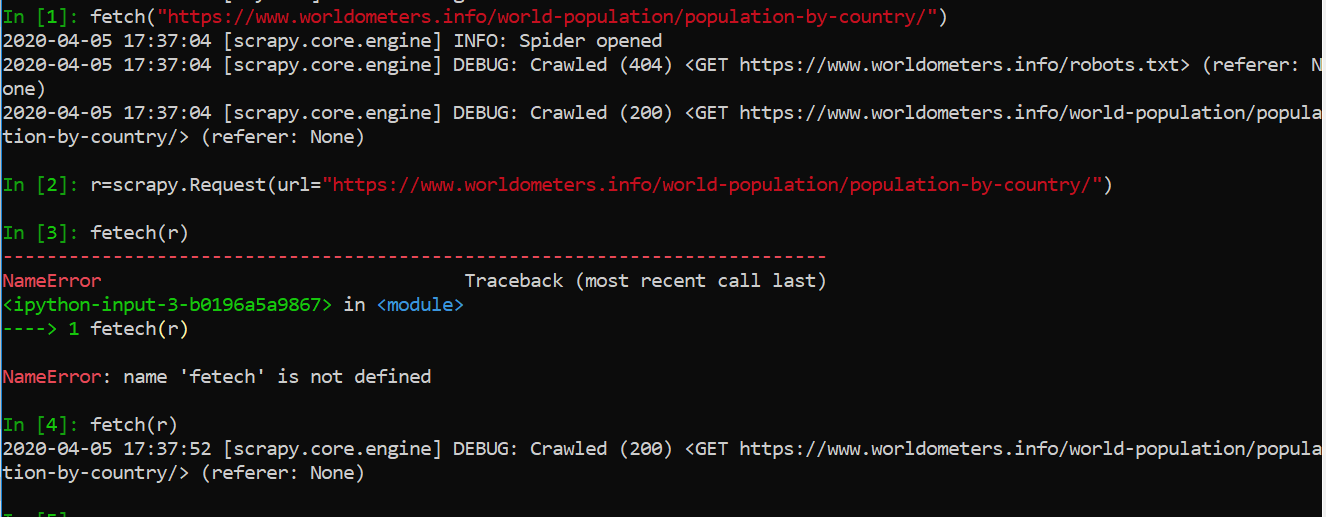
**3. Disable java script**

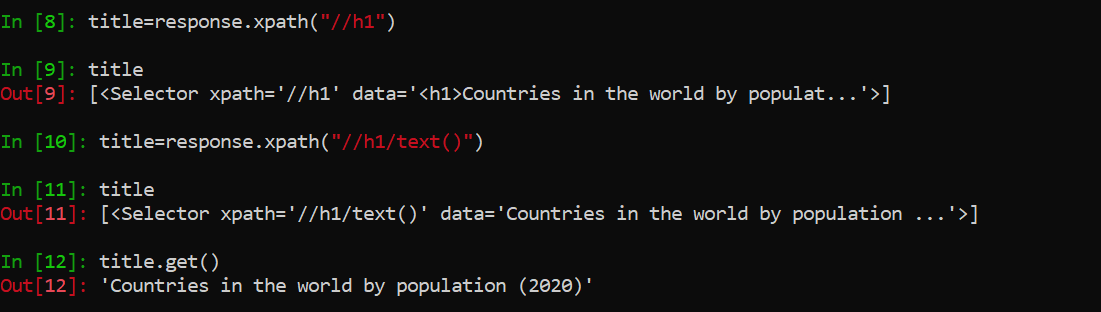
* CTRL+SHIFT + I to open developer tool
* CTRL+SHIFT+ P to open command palate clicking on element
* In the search box type in JavaScript and choose the option disable JavaScript
* Refresh the page again using CTRL+R
* Disabling JavaScript will hide all the java script driven images
* Copy the website link from the top
* Go back to Anaconda prompt to start creating projects

**To get title from the website**

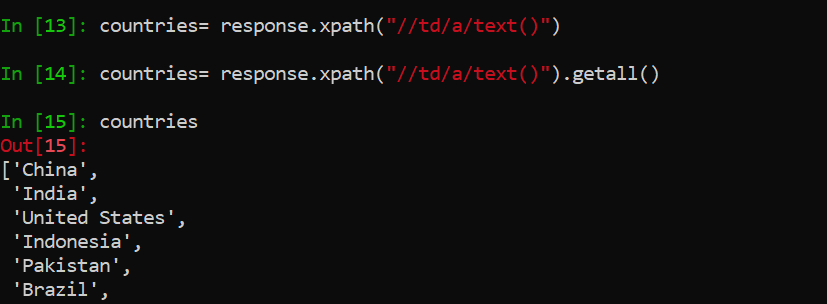
**Type scrapy shell to get started and load the website**

**Use ctrl+D to exot scrappy shell**





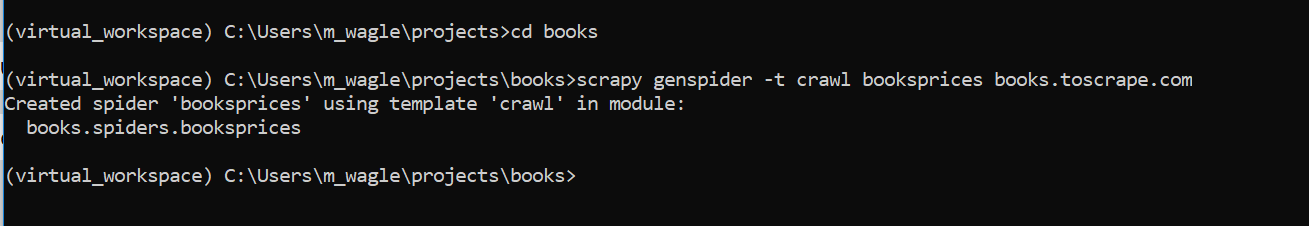
To get list of countries



Make sure you are in same level as scrapy.cfg file before executing the spider

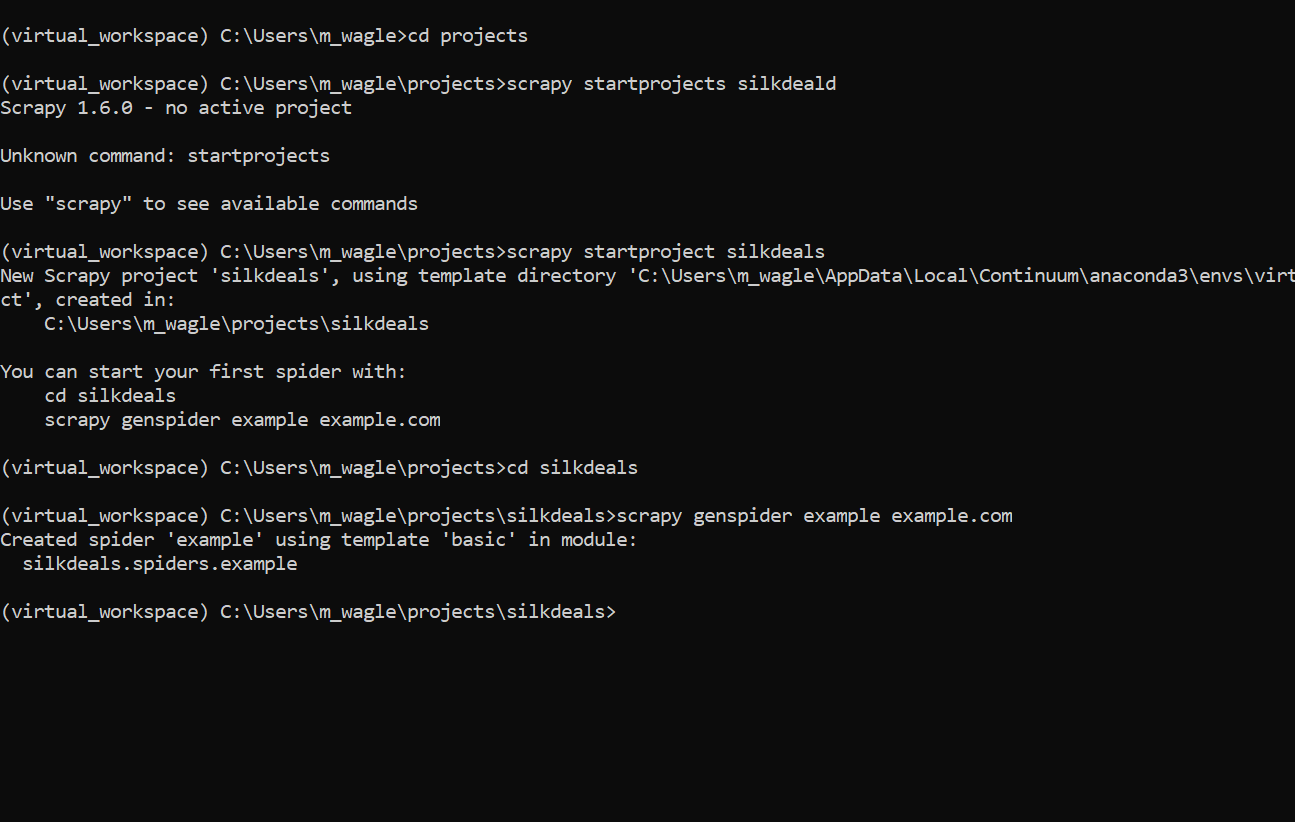
In scarpy, all the data should always be returned ad dict, so we use yield function for that.

**Starting Crawl Spider Project**



Selenium

To execute file: python .\basics.py



**Comment Code Block Ctrl+K+C/Ctrl+K+U**