**About Scrapy**

Scrapy is a framework we use to do web scraping. It is a python class. Scrapy has 5 components. They are 1. Spiders component, 2. Pipelines, 3. Middleware, 4. Engine & 5. Scheduler. The Spider component is where we define what we will have to extract from the web page. There are 5 kinds of spiders in Scrapy which are defined as spider classes. They are scrapySpider class, crawlSpider class, XMLFeedSpider for scraping XML files, CSVFeedSpider ­­­­class to scrape comma separated files and Sitemap Spiderclass to scrape sitemaps. For this project, we have used **scrapySpider class**.

**Setting up the development environment in Windows 10**

1. Install python (3.7.1) -> <https://www.python.org/downloads/release/python-371/>

Use the link *Windows x86-64 executable installer* for 64-bit OS and use *Windows x86 executable installer* link for 32-bit OS.

One the installation is completed, go to the command prompt. To confirm that installation is successful, type python -V in the command prompt. It shows "Python 3.7.1"

2. Install python virtual environment package.

From command prompt type -> pip install virtualenv

3. Create a virtual environment

From command prompt type -> virtualenv virtual\_workspace

Once virtual environment is created cd to that directory ->cd virtual\_workspace

To activate it, from command prompt type -> Scripts\activate

Now, each package we install will only belong to that virtual environment.

4. Install scrapy

from the command prompt, type -> pip install scrapy==1.5.1

5. Once the installation is complete, we can go ahead and create a new project by typing the command ->scrapy startproject capstone\_project

6. type ->cd capstone\_project.

The code editor that is used is Visual Studio Code (VS code editor). Download the same.

To launch vs code editor from the command prompt, type -> code .

As we are using python as a programming language, we need to install an extension called Python in the VS code editor. Click on the extensions icon and search for python and click install. Upon completion, click Ctrl+shift+p and type open workspace settings.

Inside the workspace settings, we need to tell the VS Code that we would be using the virtual environment that we have created.

Type "python.pythonPath":"path...." (eg"python.pythonPath":"c:\\Users\\Hiran\\Desktop\\virtual\_workspace\\Scripts\\python.exe".

press ctrl+s to save the file.

press Ctrl+shift+p and type python: create terminal, we will get an integrated terminal with the virtual environment already activated.