https://docs.scrapy.org/en/latest/intro/tutorial.html

* a website is made out of 3 core things
  + HTML
    - the bones/structure of the house
  + CSS
    - design/style of the house
  + a programming language (e.g. JavaScript)
    - functionality of the house (e.g. you open the door and there is a greeting, etc.)
* Document Object Model (DOM)
  + object oriented representation of a webpage
* url.com/robots.txt
  + check to see if you can scrape a specific page on a site and obtain other information
* scrapy startproject *projectName*
* scrapy crawl *spiderName*
* scrapy shell *url* 
  + **In [1]:** response.xpath('*element xpath*').get()
  + **Out[1]:** '<h1>Books</h1>'
  + **In [2]:** response.css('img').xpath('@src').getall()
  + **Out[2]:**

['../../../media/cache/2c/da/2cdad67c44b002e7ead0cc35693c0e8b.jpg',

'../../../media/cache/26/0c/260c6ae16bce31c8f8c95daddd9f4a1c.jpg',

'../../../media/cache/3e/ef/3eef99c9d9adef34639f510662022830.jpg',

'../../../media/cache/32/51/3251cf3a3412f53f339e42cac2134093.jpg',

'../../../media/cache/be/a5/bea5697f2534a2f86a3ef27b5a8c12a6.jpg',

'../../../media/cache/68/33/68339b4c9bc034267e1da611ab3b34f8.jpg',

'../../../media/cache/92/27/92274a95b7c251fea59a2b8a78275ab4.jpg',

'../../../media/cache/3d/54/3d54940e57e662c4dd1f3ff00c78cc64.jpg',

'../../../media/cache/66/88/66883b91f6804b2323c8369331cb7dd1.jpg',

'../../../media/cache/58/46/5846057e28022268153beff6d352b06c.jpg',

'../../../media/cache/be/f4/bef44da28c98f905a3ebec0b87be8530.jpg',

'../../../media/cache/10/48/1048f63d3b5061cd2f424d20b3f9b666.jpg',

'../../../media/cache/5b/88/5b88c52633f53cacf162c15f4f823153.jpg',

'../../../media/cache/94/b1/94b1b8b244bce9677c2f29ccc890d4d2.jpg',

'../../../media/cache/81/c4/81c4a973364e17d01f217e1188253d5e.jpg',

'../../../media/cache/54/60/54607fe8945897cdcced0044103b10b6.jpg',

'../../../media/cache/55/33/553310a7162dfbc2c6d19a84da0df9e1.jpg',

'../../../media/cache/09/a3/09a3aef48557576e1a85ba7efea8ecb7.jpg',

'../../../media/cache/0b/bc/0bbcd0a6f4bcd81ccb1049a52736406e.jpg',

'../../../media/cache/27/a5/27a53d0bb95bdd88288eaf66c9230d7e.jpg']

* scrapy items
  + starting URL's variable must be named *start\_urls = \_\_\_\_\_\_\_\_\_\_\_*

BeautifulSoup4

* to walk through HTML tree
  + soup.prettify()