Anaconda shell

* Create scrapy project –
  + Scrapy startproject whiskeyscraper
* Go to project directory –
  + Cd whiskeyscraper
* Open scrapy shell –
  + Scrapy shell
* Fetch all items from a page –
  + Fetch(‘https://www.whiskyshop.com/scotch-whisky’)
  + This returns a response variable
* Access different elements within response –
  + Response.css(‘div.product-item-info’)
* Get the name of each product –
  + products = response.css(‘div.product-item-info’)
  + products.css(‘a.product-item-link::text’).get()
* Get the price of each product –
  + products.css(‘span.price::text’).get().replace(‘£’,’’)
* Get the link for each product -
  + products.css('a.product-item-link').attrib['href']

Python script in spider folder of the project

* Create a class for the spider (code in whiskeyspider.py file)
  + Basic (without item loader)
* import scrapy  
    
  class WhiskeySpider(scrapy.Spider):  
   name = 'whiskey'  
   start\_urls = ['https://www.whiskyshop.com/scotch-whisky']  
    
   def parse(self,response):  
   for products in response.css('div.product-item-info'):  
   yield {  
   'name': products.css(‘a.product-item-link::text’).get(), 'price' : products.css(‘span.price::text’).get().replace(‘£’,’’),  
   'link': products.css('a.product-item-link').attrib['href']  
   }
* With item loader –
  + Edit the items.py file to include the fields that we are capturing through our spider
* import scrapy  
    
  class WhiskeyscraperItem(scrapy.Item):  
   # define the fields for your item here like:  
   name = scrapy.Field()  
   price = scrapy.Field()  
   link = scrapy.Field()

Anaconda shell

* Crawl and save the output in a .json file
* -O saves the output to a new file or replaces it if the filename already exists
* -o appends it to the existing filename