**Week 1:**

**Scraping libraries:**

* Requests,
* Beautiful Soup (bs4)

Requests - for Fetching html from the website

Beautifulsoup- for finding the elements/tags in a webpage

**Dataset**

* Cooking website : <https://www.jamieoliver.com/>
* 858 different recipes

**Task Breakdown**

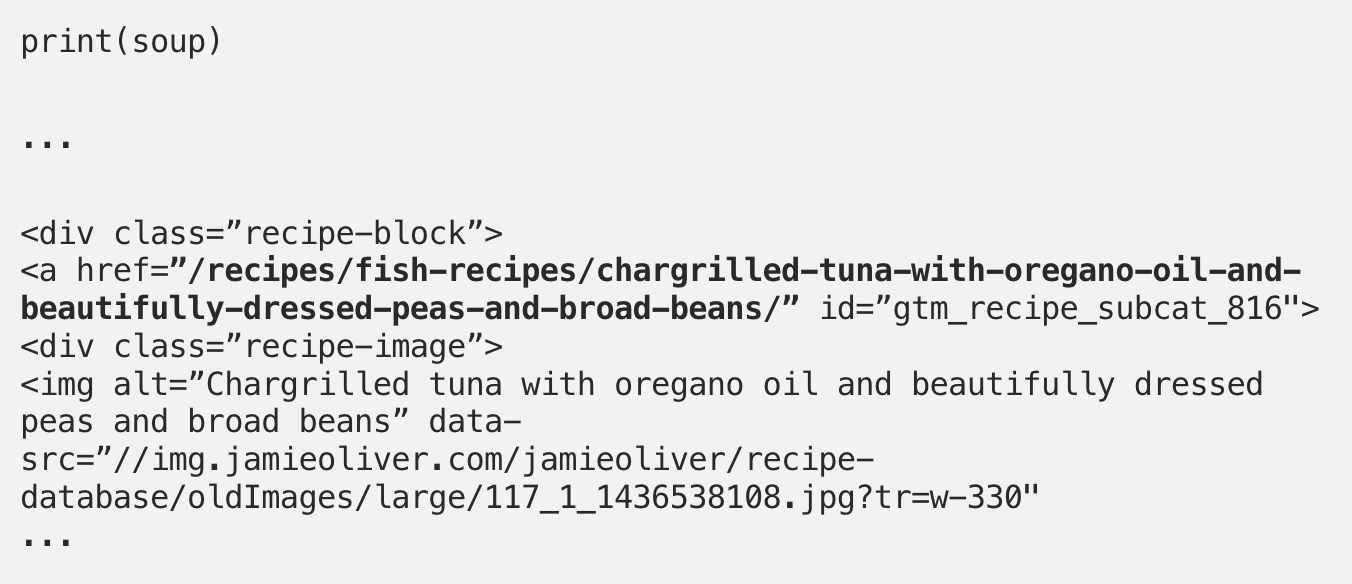
* Step 1: obtain URLs to each recipe page
  + Example : [**https://www.jamieoliver.com/recipes/duck-recipes/roast-duck-with-marsala-gravy/**](https://www.jamieoliver.com/recipes/duck-recipes/roast-duck-with-marsala-gravy/)
* Step 2: Within each of these URLs, find recipe attributes: recipe name, ingredients, serving size, cooking time, and difficulty.

**Approach**

Used requests to retrieve content from the URL and BeautifulSoup to extract the required information from our requested content. Our target website is written in HTML so we had to use the ‘html.parser’.

Requests - for Fetching html from the website

Beautifulsoup- for finding the elements/tags in a webpage



Finding of the href’s in our ‘soup’ - using BeautifulSoups find\_all method.

We then made use of the get method to extract the href from inside each ‘a’ tag.

For every url, scraped:

* 'recipe\_name',
* 'serves',
* 'cooking\_time',
* 'difficulty',
* 'ingredients'

