1- web scraping a una web de espectaculos ([Web\_Scraping\_PR1\_UOC/espectaculos.csv at master · jonortizabalia/Web\_Scraping\_PR1\_UOC · GitHub](https://github.com/jonortizabalia/Web_Scraping_PR1_UOC/blob/master/espectaculos.csv))

2- web scrapping a filmin ([Web\_Scraping\_PR1\_UOC/espectaculos.csv at master · jonortizabalia/Web\_Scraping\_PR1\_UOC · GitHub](https://github.com/jonortizabalia/Web_Scraping_PR1_UOC/blob/master/espectaculos.csv))

3- web scrapping a filmaffinity ([Practica1-Web-scraping/best\_films.csv at main · ldetorreUOC/Practica1-Web-scraping · GitHub](https://github.com/ldetorreUOC/Practica1-Web-scraping/blob/main/best_films.csv))

4-web scrapping hidrografía [web-scraping/csv at master · rsalamanquesb/web-scraping · GitHub](https://github.com/rsalamanquesb/web-scraping/tree/master/csv)

5- web scraping web libros [PRA1-WebScraping/Informe at master · iboda001/PRA1-WebScraping · GitHub](https://github.com/iboda001/PRA1-WebScraping/tree/master/Informe)

6- productos de ferretería [GitHub - dcanete/PRA1WebScraping: Práctica 1 web scraping. Tipología y ciclo de vida de los datos. UOC](https://github.com/dcanete/PRA1WebScraping)

7- subastas pisos [GitHub - anderestebanez/Practica-1-Web-scraping: Práctica 1: Web scraping](https://github.com/anderestebanez/Practica-1-Web-scraping/tree/master)

8- blog [GitHub - obachiller/WebScraping: Proyecto Web Scraping UOC](https://github.com/obachiller/WebScraping?tab=readme-ov-file)

9- covid [GitHub - SevillaFe/UOC\_Web-Scraping: UOC Repository for Web Scraping](https://github.com/SevillaFe/UOC_Web-Scraping/tree/master)

10-nanosatelites ([GitHub - beejeke/nanosat-scraping: Repositorio para la práctica 1 (Web scraping) de la asignatura Tipología y ciclo de vida de los datos del máster en Ciencia de Datos (UOC)](https://github.com/beejeke/nanosat-scraping))

RECURSOS PARA WEB SCRAPING:

[GitHub - lkuffo/web-scraping: Más de 50 ejemplos de web scraping utilizando: Requests | Scrapy | Selenium | LXML | BeautifulSoup](https://github.com/lkuffo/web-scraping)

**(no necesario Selenium)**

Idealista: [Scrapy-Idealista/README.md at master · David-Carrasco/Scrapy-Idealista · GitHub](https://github.com/David-Carrasco/Scrapy-Idealista/blob/master/README.md)

Indeed: [Indeed\_Job\_Scraper/Indeed\_Job\_Scraper.py at master · Ram-95/Indeed\_Job\_Scraper · GitHub](https://github.com/Ram-95/Indeed_Job_Scraper/blob/master/Indeed_Job_Scraper.py)

Carrefour (usar Selenium): [Carrefour\_WebScraping/Carrefour\_WebScraping.py at main · carlosvertiz/Carrefour\_WebScraping · GitHub](https://github.com/carlosvertiz/Carrefour_WebScraping/blob/main/Carrefour_WebScraping.py)

[CarrefourTracker/pdf/PRAC1\_WebScraping.pdf at main · equipoUOC/CarrefourTracker · GitHub](https://github.com/equipoUOC/CarrefourTracker/blob/main/pdf/PRAC1_WebScraping.pdf)

PC Componentes:

[scrapper-pccomponentes/PRAC1 Web Scrapping.pdf at master · mtldevai/scrapper-pccomponentes · GitHub](https://github.com/mtldevai/scrapper-pccomponentes/blob/master/PRAC1%20Web%20Scrapping.pdf)

All recipes

[Recipes](https://www.allrecipes.com/recipes/)

Web de interés para hacer la práctica: [Python Selenium: Scraping A Recipe Website | by Galina Blokh | Analytics Vidhya | Medium](https://medium.com/analytics-vidhya/python-selenium-web-scraping-in-eight-steps-7d33b263f399)

* [GitHub - shaansubbaiah/allrecipes-scraper: 🥗 Scrapy spider to scrape recipe and nutritional data from](https://github.com/shaansubbaiah/allrecipes-scraper/tree/main) [allrecipes.com](http://allrecipes.com)
* [GitHub - hhursev/recipe-scrapers: Python package for scraping recipes data](https://github.com/hhursev/recipe-scrapers)
* [GitHub - phuongdtrn/Python-Web-Scraping-Allrecipes: This project details the steps for web scraping over 13,000+ recipes to analyze factors that could contribute to a higher star ratings from 1 to 5.](https://github.com/phuongdtrn/Python-Web-Scraping-Allrecipes/tree/main)
* [GitHub - anand-07-a/recipe\_scraper: Web scrapers for extracting recipes from popular cooking websites.](https://github.com/anand-07-a/recipe_scraper)

Es una web que banea a los webscrapers después de 10-15 descargas rápidas.

*No acabo de entender lo del uso de user-agent, si ya indicamos en selenium cuál estamos usando?*

1. **Contexto:** El objetivo principal es construir un conjunto de datos estructurado que permita estudiar qué recetas tienen más éxito de cada país. Es una fuente fiable debido a que “utilizar el about us para explicarlo”.
2. **Título:** Recetas de diferentes países
3. **Descripción del dataset:**
4. **Representación gráfica:** esto lo haría con Tableau

**Contenido:** Esto sin más, no es problema. En cuanto al periodo de tiempo, se puede añadir la fecha cuando se subió la receta.   
Propuesta de campos:   
**url\_receta:** El enlace permanente que servirá como identificador único de cada receta. Es de tipo texto.

**titulo:** El nombre descriptivo de la receta, en formato texto.

**fecha:**

**autor:**

**descripcion\_corta:** Un breve resumen de la receta que se encuentra típicamente al inicio de la página. Es de tipo texto.

**calificacion\_media:** La puntuación promedio de la receta. Es un campo numérico (flotante), esencial para medir la calidad percibida.

**num\_reviews:** La cantidad de valoraciones que ha recibido la receta, un número entero que ayuda a medir la popularidad y el nivel de participación.

**tiempo\_preparacion:** El tiempo estimado necesario para la fase de preparación de la receta. Se extraerá como texto (ej. "30 minutes") para posterior conversión.

**tiempo\_coccion:** El tiempo necesario para la cocción o el reposo. Se extraerá como texto.

**tiempo\_total:**

**ingredientes:** La lista completa y estandarizada de los ingredientes requeridos. Se puede almacenar como una lista de texto o una cadena de texto para su posterior análisis.

**porciones:** La cantidad de personas para las que está pensada la receta. Es un campo numérico (entero).

**categoria\_principal:** La clasificación general de la receta (ej. "Desserts" o "Main Dish"). Es de tipo texto.

**calorias:**

**grasas:**

**carbohidratos:**

**proteínas:**

1. **Propietario:** Utilizar el repositorio de github del enlace o otros trabajos de análisis de recetas.
2. **Inspiración:** Podríamos enfocarlo en que queremos montar un restaurante y queremos saber qué recetas o platos son los más populares.
3. **Licencia:** La que sea
4. **Código:** en python
5. **Dataset:** Definir qué datos queremos obtener.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nombre** | **Autor** | **Valoración** | **n\_val** | **Reviews** | **Publicado** | **Prep time** | **Total time** | **Ingredientes** | **Nutrition facts** | **Comentarios** | **n\_home\_cooks** |
| **Chimichurri** | **John Mitze** | **4.5** | **230** | **174** | **21/03/2024** | **15** | **15** |  |  | **So yummy** | **525** |
|  |  |  |  |  |  |  |  |  |  |  |  |

1. **Video:**