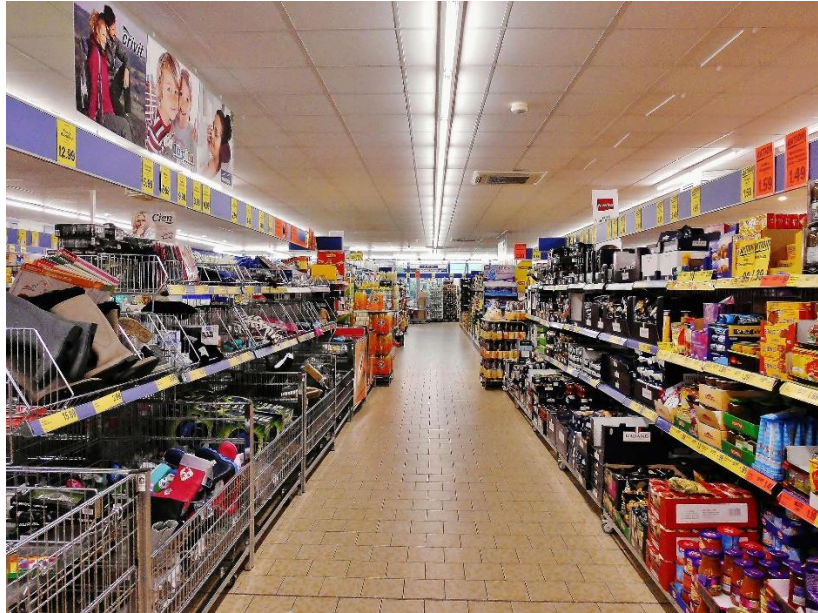


## Tipología y ciclo de vida de los datos

### Práctica 1: scrapper precios supermercado

*Mikel Álvarez Rua*

*Mikel Tobar del Barrio*



*Figura 1: Pasillo de un supermercado*

#### ***Introducción***

En esta práctica, queremos realizar un proyecto que obtenga todos los precios de productos del supermercado DIA. Para ello, se desarrollará un scrapper en Python y después se obtendrá un archivo .csv con la lista de todos los precios.

#### ***1. Contexto***

En esta práctica, se pide crear un dataset a partir de los datos contenidos en una web. Se ha escogido este proyecto para poder, en primer lugar, hacer una recopilación de los precios de los elementos vendidos en un gran supermercado muy popular entre la ciudadanía.

Posteriormente, y a partir de este dataset, sería posible hacer un estudio de los precios de los bienes vendidos en los supermercados DIA, analizar los productos más baratos y caros por sección, estimar cuál sería el precio que una persona o familia paga por la cesta de la compra, etc.

#### ***2. Descripción del dataset***

En este dataset se mostrará, en formato.csv, una columna con la sección a la que pertenece el producto, otra con su nombre, y finalmente su precio.

#### ***3. Contenido***

El dataset consta de los siguientes campos:

- **Sección:** Sección a la que pertenece el producto.
- **Producto:** Nombre del producto para su identificación.
- **Precio:** Precio del producto (en Euros).

Los datos se han recogido el día 9 de abril de 2021, por lo que reflejarán la situación de los precios a dicha fecha.

#### ***4. Agradecimientos***

Los datos extraídos proceden del sitio web del supermercado DIA. Para poder realizar el web scraping, que posibilita la obtención de los datos, se ha empleado el lenguaje de programación Python y la biblioteca BeautifulSoup.

Hemos podido observar cómo existen ejemplos de repositorios que recopilan precios de supermercados<sup>1</sup> o artículos académicos que analizan los mismos<sup>2</sup>. No obstante, si bien es cierto que dichas ideas han dado un impulso inicial a nuestra inspiración, nuestro proyecto quiere desmarcarse de estos enfoques y dar una visión particular de la información recogida.

#### ***5. Inspiración***

A nivel de grupo, existían varias ideas a la hora de realizar esta práctica. Finalmente, los primeros intentos fueron en vano por dificultades técnicas a la hora de realizar el web scraping en las páginas escogidas.

Fue entonces cuando se reparó en el interés de poder realizar una recopilación de los precios de un supermercado, siendo la compra de productos de necesidad básica algo vital en el día a día de casi todas las personas.

Particularmente, las personas que realizamos estudios universitarios, como estudiantes o jóvenes profesionales, podemos encontrar dificultades como consecuencia del elevado precio de muchos bienes básicos (comida, vivienda, etc.), por lo que una recopilación de precios de productos de supermercado es de gran interés de cara a un análisis futuro.

En el primer ejemplo que nos ha podido servir de inspiración, se realiza un scraping de los principales supermercados del Reino Unido. Sin embargo, en nuestro proyecto, con el objetivo de simplificar el análisis, se realizará un estudio de un solo supermercado.

El segundo ejemplo que ha ayudado a encontrar el camino para la realización de este proyecto es el análisis detallado de los precios de consumo que se realiza en el artículo citado. No obstante, este eventual análisis a realizar quedaría como un posible fin al que dirigir el dataset, totalmente fuera del alcance de esta práctica.

#### ***6. Licencia***

Se escoge como licencia la **CC BY-SA 4.0 License**. Los términos de uso son atribución a los autores en caso de utilización, no utilización para fines comerciales y uso de la misma licencia.

#### ***7. Código y dataset***

Se pueden ver como elementos incluidos en este repositorio. El archivo .csv obtenido se publica también en Zenodo con DOI 10.5281/zenodo.4675690.

## 8. Autores

<i>Contribuciones</i>	<i>Firma</i>
Investigación previa	 
Redacción de las respuestas	 
Desarrollo código	 

## 9. Referencias

1. *Github: Uk Supermarket scraper* [en línea]. Actualizada: 5 de mayo de 2019. [Fecha de consulta: 6 de abril de 2021]. Disponible en: <https://github.com/ciarans/Boozio-UK-Supermarket-Scraper>
2. Méndez, José & Paredes, García & Oubiña, Javier & Rubio, Natalia. (2009). La información de precios unitarios al consumidor: un análisis intramarca. *International Marketing Trends Conference*.