

# **Tipología y ciclo de vida de los datos**

## **Práctica 1: Web Scraping**

**Iosu Rodríguez Alfaro**

# Dataset: Precios de Smartphone de MediaMarkt

## Descripción

El dataset obtenido en esta práctica contiene los precios de los Smartphone de un día determinado. Se recogen el identificador del producto, el nombre, el precio y la fecha en la que se ha obtenido.

## Imagen



## Contexto

La materia del conjunto de datos trata sobre los precios de los Smartphone de la tienda MediaMarkt en un día determinado. Cada vez que apliquemos el algoritmo de web Scraping, se generará un fichero csv con la información de precios de ese día.

## Contenido

Los campos que incluye son:

- fecha en la que se han recogido los datos con formato dd-mm-aa.
- id: identificador único del producto.
- nombre: nombre descriptivo del Smartphone.
- precio: precio del Smartphone en ese día.

El periodo de tiempo de los datos es el del día en que se ejecuta el algoritmo.

## Agradecimientos

Los datos han sido obtenidos desde la página oficial de MediaMarkt, desde la sección Smartphone: [https://www.mediemarkt.es/es/category/\\_smartphones-701189.html](https://www.mediemarkt.es/es/category/_smartphones-701189.html)

Se ha empleado un script en Python que aplica técnicas de Web Scraping para extraer la información.

## **Inspiración**

El conjunto de datos podría obtenerse para comparar los precios del MediaMarkt con los nuestros, en caso de dedicarnos a la venta online de Smartphone.

Además cada día obtendríamos un csv con los precios de ese día. Podríamos realizar un análisis de la evolución de los precios de cada producto o hacer estudios de cuándo puede resultar más económico comprar un Smartphone en MediaMarkt.

## **Licencia**

La licencia escogida es: Released Under CC BY-SA 4.0 License.

He escogido esta licencia ya que se permite el uso comercial. Debido a la naturaleza del dataset, el uso más lógico que se le puede dar, es un uso comercial. Además, se permite copiar y redistribuir los datos en cualquier medio o formato, además de mezclar, transformar y construir sobre el material. Aunque los datos sean obtenidos de la empresa MediaMarkt, son datos sobre precios, que cualquier usuario puede ver en su web. Dar esta licencia permite que otras webs puedan hacer publicidad del MediaMarkt y sus precios, o permite a las empresas comparar sus precios con ellos. También da la posibilidad de ampliar los datos de los precios con precios de otras empresas.

## **Código fuente y dataset**

Se encuentra en el repositorio de GitHub siguiente:

<https://github.com/iosurodriguez/P1-WebScraping>