

Dataset: Precios diarios de Mediamarkt

Descripción

El conjunto de datos que se han generado son el listado de diferentes productos de una categoría (en este caso televisores) que posee la tienda Mediamarkt durante un día en concreto. Las variables que contiene este listado son el nombre del producto, el precio y la fecha de obtención del dato.

Imagen identificativa



Esta imagen refleja una gran variedad de tipos de productos que vende Mediamarkt.

Contexto

El conjunto de datos corresponde al listado de un tipo de productos que posee la tienda Mediamarkt. Los tipos de productos pueden ser por ejemplo monitores, webcams o impresoras (mirar el sitemap-productlist.xml de Mediamarkt para poder obtener otros ejemplos).

Contenido

Para cada producto se recogen las siguientes características:

- Producto: indica el nombre del producto.
- Precio(EUR): indica el precio del producto en Euros.
- Fecha dato: Indica la fecha de obtención de los datos sobre el producto.

Estos datos han sido recogidos a partir de un script en Python que hace uso de técnicas de Web Scraping. Para hacer funcionar el Script hace falta indicarle la url que hace referencia a una categoría y el nombre de la categoría(Sera el nombre del fichero csv más la fecha en la que se ha ejecutado el script). Ejemplo:

```
$ python MediamarktScraper.py Televisores https://www.mediamarkt.es/es/category/_televisores-701210.html
```

Para ver el listado de categorías junto a sus url se puede acceder al sitemap que tiene publico Mediamarkt: <https://www.mediamarkt.es/sitemap/sitemap-productlist.xml>

Agradecimientos

Los datos han sido recogidos de la página web de Medimarkt (tienda online) para ello se ha hecho uso de técnicas de Web Scraping aprendidas en la asignatura de “Tipología i cicle de vida de les dades” de la UOC.

Inspiración

Este conjunto de datos podría utilizarse en situaciones muy diversas. Por ejemplo, cuando un posible cliente quiere adquirir un producto pero espera a que este baje sus precio. El posible cliente pueda hacer uso de la información recogida para ver si la rebaja realizada vale la pena o no. Ya que por ejemplo puede ser que haya habido otra rebaja mayor en la historia de ese producto. De esa manera podrá valorar si vale la pena esperar a que ese producto pueda volver a conseguir esa rebaja o no.

Otro uso puede ser por ejemplo para ayudar a distinguir al cliente si un producto tiene un precio erróneo o no (comprobado si la diferencia de precio respecto otras ocasiones es muy exagerada). De esta manera se ahorrará los tramites de cancelación de pedido si la empresa desea cancelarlo.

Licencia

La licencia que se ha utilizado ha sido la CC BY-NC-SA 4.0. Los motivos por los que se ha escogido esta licencia han sido los siguientes:

- Permite copiar y distribuir la información en cualquier formato. De esta forma puede ser vista por personas que tengan la intención de comprar en Mediamarkt.
- Atribuir el nombre del creador de los datos e indicar si se han realizado cambios sobre esta. Siempre es bueno reconocer el trabajo de otros. Además, si se realizan cambios sobre este siempre es bueno saber cuáles han sido.
- No permite el uso comercial. Esta información ha sido generada a partir de la página web de Mediamarkt, por lo que no se debe comercializar con esa información.

Código fuente y dataset

El código fuente y dataset están ubicados en mi repositorio de github:
<https://github.com/aprietomo/Web-scraping>