

Dataset: Comparación de ordenadores portátiles

Rubén Gómez Márquez

09 de abril de 2020

Descripción

El conjunto de datos generados dentro de esta práctica conforma un conjunto de datos asociados a ordenadores portátiles actuales. Dentro del conjunto de ordenadores existentes podemos encontrar de la mayor parte, por no decir de todas, las marcas más relevantes en lo que en este ámbito se refiere. El dataset extraído contiene datos referentes al nombre del producto, marca, valoración por parte de los compradores, precio actual y descuento realizado.

Imagen identificativa

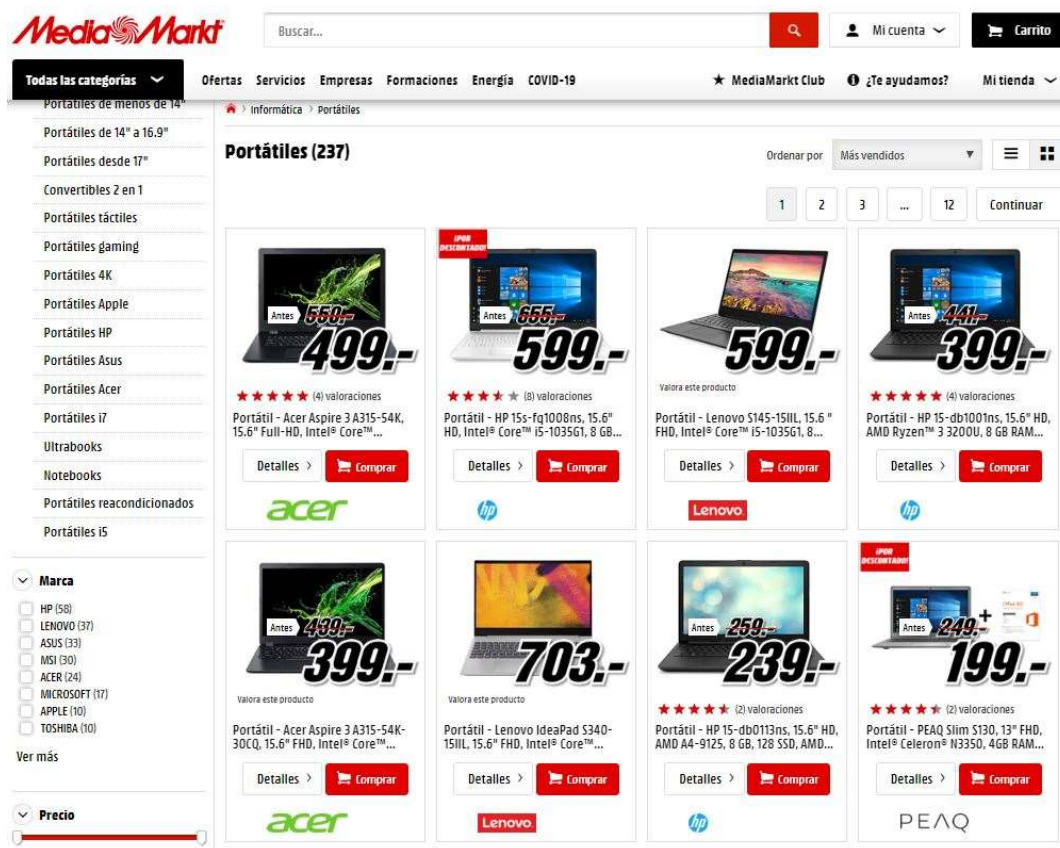


Figura 1: Fragmento de la página web de la cual se extrae la información (<https://www.mediamarkt.es/es/category/port%C3%A1tiles-701175.html?view=PRODUCTGRID&page=1&sort=topseller>)

Contexto

El conjunto de datos corresponde a las características de los diferentes ordenadores portátiles de las diferentes marcas, haciendo hincapié en el precio del producto. Si bien los datos extraídos se encuentran en un estado brutos, es decir, requieren una limpieza, son un buen ejemplo de cómo se pueden obtener datos de forma sencilla de cara a realizar un estudio de mercado que posteriormente ofrecer al consumidor. Estos estudios podrían decantar la balanza entre comprar el producto de una marca u otra y, por tanto, mejorar los índices de beneficios.

Contenido

Para cada uno de los ordenadores portátiles que se registran en la página web se registran los siguientes atributos:

- **Marca:** hace referencia a la empresa manufacturadora.
- **Nombre:** nombre del producto, del cual también pueden extraerse características del mismo.
- **Precio actual:** precio que nos costaría el producto.
- **Precio antiguo:** precio anterior a la realización de un descuento.
- **Descuento:** diferencia entre el precio actual y el precio antiguo, es decir, descuento directo realizado al producto.
- **Valoración:** opinión aportada por el consumidor tras la compra del producto. Está ponderada entre 1 y 5, expresando 1 que el producto no satisface las necesidades del consumidor y 5 que el consumidor está contento con la compra y el producto es de su agrado.

Agradecimientos

Mediante el uso de técnicas de Web Scraping bajo el lenguaje de programación Python, ha sido extraída la información alojada en las páginas HTML de la empresa MediaMarkt.

Inspiración

Como exponía brevemente en los conceptos de este mismo documento, la idea de este Dataset es la realización de un análisis de mercado de un producto, en este caso ordenadores portátiles, que podría ser de cualquier producto y en cualquier página que resultase como intermediario.

Este análisis de mercado podría ser de gran utilidad para el consumidor, con el fin de decantarse entre un producto con la ayuda de un análisis que le facilitaría ver que producto se adecúa mejor a sus necesidades, e incluso para diferentes empresas que buscan información acerca de las empresas de la competencia.

Lo que está claro es que estudios de este tipo pueden ser de gran beneficio para las empresas ya que pueden cambiar la opinión del consumidor.

Licencia

Para la publicación del Dataset se escoge la licencia Released Under **CC BY-SA 4.0 License**, bajo la premisa de que se permite su uso comercial, que es el fin mismo del Dataset, y se le aporta cierto reconocimiento al autor a la par que mantiene los términos del proyecto realizado.

Código fuente y dataset

El código fuente se puede obtener en el siguiente enlace:

<https://github.com/rgommar/Web-Scraping/blob/PRACTICA-1.-WEB-SCRAPING/Scraping%20portatiles.ipynb>

Recursos

1. Subirats, L., Calvo, M. (2019). Web Scraping. Editorial UOC.
2. Masip, D. (2010). El lenguaje Python. Editorial UOC.
3. Tutorial de Github <https://guides.github.com/activities/hello-world>.
4. <https://www.youtube.com/watch?v=rhnMvvmfBFI> (Pequeño tutorial introductorio a web scraping y BeautifulSoup, por Errodringer, creador de contenido en la plataforma youtube)
5. <https://www.youtube.com/watch?v=XQgXKtPSzUI> (Pequeño tutorial introductorio a web scraping y BeautifulSoup, por Data Science Dojo, creador de contenido en la plataforma youtube)