

Dataset: Productos electrónicos de la página web Tecnomari

Manel Muñiz Perálvarez

9 de abril de 2022

Título

“Productos electrónicos recogido en la página web Tecnomari a día 11 de abril de 2022.”

Representación gráfica - Imagen Identificativa

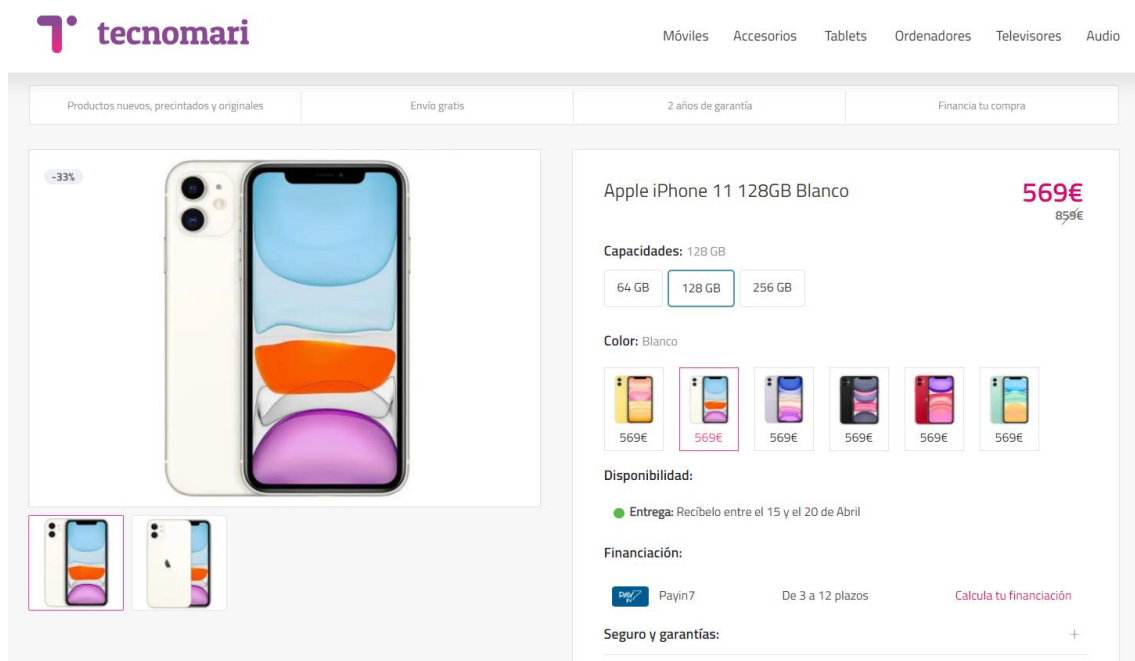


Figura 1: Link de producto “Apple iPhone 11 126GB Blanco”

Contexto

Nuestro conjunto de datos generado como parte de esta práctica de Web Scrapping reúne los diferentes productos que podemos encontrar dentro de una página de tecnología, así como sus distintos precios.

Los datos recogidos en esta página web se centran en recolectar todos los productos de los cuáles dispone la empresa y algunas de sus características de venta (precio, descuento...). Dentro de nuestro conjunto de datos generado podemos encontrar smartphones, portátiles, fundas, pantallas, accesorios, televisores, cargadores, baterías, auriculares y pendrives.

Contenido

Para cada producto que encontramos en nuestro conjunto de datos, se recogen las siguientes características:

- **Name:** Nombre del producto.
- **Price:** Precio.
- **Discount Price:** Precio con el descuento aplicado.
- **Product link:** Link de la imagen del producto.
- **Capacity:** Capacidad de almacenamiento del producto.
- **Discount Percentage:** Descuento aplicado en %.
- **Categories:** Categorías a las cuáles pertenece un producto.

Todos los datos recogidos en <https://tecnomari.es/> se encuentran actualizados a día 11 de abril de 2022.

Agradecimientos

Los datos han sido recolectados desde la propia página web de Tecnomari (<https://tecnomari.es/>). Tecnomari es una empresa de informática nacida en 2012 que apostó desde el primer momento por el envío online y su focalización en la venta de productos móviles. Actualmente se encuentra en tiendas en diversas partes de España en las cuáles también se pueden recoger los pedidos y asistir a una postventa física de los productos comprados. Dentro de su crecimiento, Tecnomari se encuentra ahora mismo vendiendo a nivel europeo, tanto en España como Francia, Alemania, Italia y UK haciéndola así una gran competidora del sector.

Para la extracción de los datos, hemos utilizado el lenguaje de programación Python y técnicas de Web Scrapping para extraer las informaciones alojadas en las páginas HTML.

Todos los datos recogidos en Tecnomari se encuentran abiertos de forma pública a todos los usuarios que accedan a la página web y su recolección se ha hecho de una forma ética que sólo busca recoger sus productos y gestionar su información.

Inspiración

El presente conjunto de datos tiene en mente principalmente el ser utilizado para estudiar y recolectar los precios, datos y nuevos productos que se añadan a la empresa Tecnomari ya sea por la propia empresa o por algún competidor que desee extraer todo este tipo de información para ver cómo evoluciona en el tiempo.

Con esta información se podrían llegar a predecir nuevos precios en productos nuevos, nuevos descuentos que se aplicarán en un futuro gracias al uso de la minería de datos elaborando modelos predictivos.

Licencia

La licencia seleccionada para la presentación y uso de este Dataset es "Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)"

Hemos seleccionado esta licencia debido a que nos permite y nos da la libertad de compartir nuestro conjunto de datos generado en cualquier tipo de medio o formato de forma libre y que se adapte y trabaje sobre el material generado. Creo que es una buena licencia que en el ámbito académico encaja ya que estamos trabajando con unos datos públicos disponibles para todo el mundo con el único propósito de practicar algunas de las técnicas aprendidas en la asignatura.

A esto, sumado que no buscamos ningún tipo de atribución personal y todo el crédito ha sido dado a la página que gestiona esta información (<https://tecnomari.es/>) y no pretendemos comercializar de ninguna de las maneras el contenido generado, hacen de "Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)" una buena elección para los datos generados.

Código fuente y Dataset

Tanto el código fuente escrito para la extracción de datos como el Dataset generados pueden encontrarse en los siguientes links:

- Zenodo: <https://zenodo.org/record/6450277>
- Github: https://github.com/Manelm5/Web_Scrapping_PR1_UOC
- Video: <https://drive.google.com/drive/folders/14NdBOm-m8Olq2V6mKWoSIOAZllrgYhD?usp=sharing>
- DOI Zenodo: 10.5281/zenodo.6450277

Recursos

1. Subirats, L., Calvo, M. (2018). Web Scraping. Editorial UOC.
2. Masip, D. (2019) El lenguaje Python. Editorial UOC.
3. Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
4. BeautifulSoup documentation, <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
5. BS general documentation: <https://pypi.org/project/beautifulsoup4/>
6. Web scraping con Python. Extraer datos de una web. Guía de inicio de BeautifulSoup: <https://j2logo.com/python/web-scraping-con-python-guia-inicio-beautifulsoup/>
7. Simon Munzert, Christian Rubba, Peter Meißner, Dominic Nyhuis. (2015). Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons.

8. Tutorial de Github <https://guides.github.com/activities/hello-world>.