

Practica 1 – Web Scrapping



Contexto

Los datos han sido recogidos de la página web: www.carrefour.es.

Se ha recogido la información de todos los portátiles vendidos por Carrefour España en el contexto de que si eres un competidor directo, como puede ser una empresa de venta de ordenadores y periféricos por internet, necesitaras saber que modelos vende la competencia y al precio que los vende. Por eso se han recogido los más de 1200 modelos que vende Carrefour, con su precio actual y su precio anterior si ha sido rebajado.

Descripción del dataset

Describir el dataset es una tarea sencilla ya que este se compone básicamente de tres campos:

Nombre del producto: marca del portátil, modelo y descripción.

Precio anterior: precio al que se encontraba anteriormente.

Precio actual: precio actual del portátil.

Estos productos y precios corresponden al mes de abril de 2019.

Agradecimientos

Los datos han sido recolectados desde la página web de Carrefour España. Para ellos hemos utilizado el lenguaje de programación Python, usando librerías de Web Scraping para extraer la información deseada.

Inspiración

Creo que es un buen punto de partida para una empresa de venta de productos informáticos por internet. Obteniendo los precios de venta y los productos de tus principales competidores puedes variar el precio de tus productos para que sean más competitivos.

Licencia

Ya que estos datos van a ser usados de forma interna podemos basarnos en la licencia CCO: Public Domain Licence.

La persona que ha asociado una obra a este documento ha dedicado la misma al dominio público, liberándola de forma mundial y en la medida que lo permita la ley, de todos sus derechos de propiedad intelectual, incluyendo todos los derechos conexos.

Puede copiar, modificar, distribuir la obra y hacer comunicación pública, incluso para fines comerciales, sin pedir permiso.

Recursos

- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- Subirats, L., Calvo, M. (2018). Web Scraping. Editorial UOC.