

Descripción

En el presente proyecto, hemos creado un web crawler capaz de navegar por la web del supermercado *Carrefour*, extrayendo los datos de los diferentes productos (precios, ofertas y promociones). Este ejercicio tiene como finalidad entender y poner en práctica los conceptos aprendidos en la asignatura de *Tipología y Ciclo de Vida de los Datos* para realizar un Web Scraping.

Miembros del equipo

Actividad realizada en conjunto por los alumnos:

- Ismael Rosende Rey
- Chantal López Procopio

Título Dataset

El título escogido para este dataset es **CarrefourDailyPricing**, ya que guarda relación directa con el propósito de la araña que hemos creado para este proyecto, que se explica detalladamente a continuación.

Descripción del Dataset

El dataset **CarrefourDailyPricing** está formado por una lista con los productos del supermercado Carrefour, guardando información como la categoría del producto, su precio actual, precio por kilogramo, ofertas y promociones.

Representación gráfica

(en proceso)

Contenido del dataset

El dataset **CarrefourDailyPricing** contiene los siguientes atributos:

- **Categoría:** Sección del supermercado a la que pertenece un producto.
- **Descripción:** Nombre que tiene asociado el producto en el supermercado.
- **Precio:** Precio de venta actual.
- **Precio/Kg:** Precio por cada kilogramo
- **Precio previo:** Precio anterior a la oferta.
- **Precio oferta:** Precio de venta durante la oferta.
- **Promociones:** Promociones activas para cada producto.
- **Enlace:** Enlace donde se puede encontrar el producto en cuestión.

La araña está programada para actualizar el dataset una vez al día, a las 8:00hrs, generando un csv nuevo cogiendo los atributos ya mencionados.

Agradecimientos

Para elaborar este dataset se recurre a la web de Carrefour para extraer los datos de los productos que tienen dentro de su oferta comercial. Agradecemos por su aporte a este proyecto como nuestra fuente principal no solo de datos, sino también durante nuestra formación y aprendizaje.

Inspiración

Este conjunto de datos busca recoger diariamente información de interés sobre cada producto en el supermercado *Carrefour*. Durante la elaboración de este proyecto, no solo buscamos la extracción de datos, sino también poder llevar un registro donde comparar las subidas y bajadas de precios en un período de tiempo determinado. Por esa razón, hemos programado la araña para que recorra la web diariamente, y así poder remarcar los productos en oferta, cada cuánto tiempo ocurren y su evolución en el tiempo. También, serviría para estudiar el impacto que tienen sobre la economía las crisis como la que vivimos actualmente del covid.

Licencia

(seleccionar una)

Código y dataset

El código fuente y el dataset generado con nuestro script “carrefourScraping”, se pueden encontrar en el siguiente enlace: <https://github.com/equipoUOC/CarrefourTracker>

Contribuciones al proyecto

| Contribuciones | Firma |
|-----------------------------|----------|
| Investigación previa | IRR, CLP |
| Redacción de las respuestas | IRR, CLP |
| Desarrollo código | IRR, CLP |

Recursos

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