

TIPOLOGÍA Y CICLO DE VIDA DE LOS DATOS

PRÁCTICA 1: WEB SCRAPING

Autor: **JAVIER CAÑÓN ÁLVAREZ**

1. Contexto

En esta época de incertidumbre sobre la compra de coches nuevos, el mercado de coches de segunda mano ha adquirido una gran relevancia. Por ello, se ha decidido hacer una exploración de los diferentes coches que están a la venta en una de las páginas web más conocidas de venta de segunda mano: <https://www.coches.com/coches-segunda-mano>. Esta web nos proporciona información sobre todos los coches de segunda mano que se encuentran a la venta en España. Nos indica los siguientes datos: coches, combustible, año, cv, km, ubicación y precio.

2. Definir un título para el dataset

El título elegido para el dataset será: **Coches Segunda Mano – Web Scraping**

3. Descripción del dataset

Este dataset nos proporciona los datos sobre casi 1000 coches que se encuentran a la venta en la web <https://www.coches.com/coches-segunda-mano>. Con ellos podremos hacer una inspección y análisis de los coches que están a la venta, sus precios y sus características principales: combustible, año, cv, km, ubicación y precio.

4. Representación gráfica

La siguiente imagen puede representar el dataset correctamente:



5. Contenido

Mediante Web-Scraping se han extraído los datos de las 50 primeras páginas de la web de venta. Por cada página se disponen una lista los coches que están a la venta de manera cronológica. Analizando la estructura de la página se ha conseguido localizar el contenido HTML sobre los coches que nos interesaba, y mediante Python (librerías Request y BeautifulSoup) se han extraído los datos a un archivo CSV.

Los datos almacenados son:

- Coche: el coche puesto a la venta.
- Ubicación: lugar donde se vende el coche.
- Combustible: combustible/tipo de propulsión que tiene.
- CV: potencia en CVs
- Anno: año desde su fabricación.
- Kilometros: número de kms hasta la actualidad.
- Precio: precio de venta

6. Agradecimientos

Agradecer al dominio “coches.com” de donde se han obtenido los datos para realizar este ejercicio de Web Scraping.

7. Inspiración

Como ya se ha dicho en la introducción, el mercado de coches de segunda mano ha subido en los últimos tiempos, debido la intrusión de las nuevas tecnologías de propulsión (coches híbridos y eléctricos) y la cada vez más restrictiva normativa de emisiones contaminantes. Las personas no quieren gastarse el dinero en un coche nuevo, y aquí es donde entra en juego los coches de segunda mano.

Este conjunto de datos nos permitirá analizar el mercado de coches de segunda mano actual. Podremos ver clasificar los coches según sus km, potencia, combustible, ubicación... y poder decidir si un coche que nos interesa está en el precio de mercado o por el contrario es más caro o más barato. De una manera sencilla y rápida se podrá hacer un sondeo del mercado.

8. Licencia

La licencia elegida para el dataset será “**Released Under CC0: Public Domain License**”, porque serán datos libremente compartidos sin restricción ninguna.

9. Código

Ver archivo de Python: *jcanona_practica1_web_scraping.ipynb*

10. Dataset

Ver archivo CSV obtenido: *coches_segunda_mano.csv*

Contribuciones	Firma
Investigación previa	Javier Cañón Álvarez
Redacción de las respuestas	Javier Cañón Álvarez
Desarrollo código	Javier Cañón Álvarez