

Dataset: Extracto de la lista de vehículos en venta publicados por el portal **tucarro.com.co**

Por: Gilberto J. Martínez

Noviembre 10 de 2.018

Descripción

El dataset generado en el presente repositorio, consta de un extracto del listado de vehículos en venta publicados en el portal web **tucarro.com.co**, el cual se fundamenta en el segmento de automóviles marca **Chevrolet** modelos desde el **2008 hasta el 2018** y ordenado **ascendentemente por precio**.

Así mismo solo se incluyen las primeras tres páginas, equivalentes a 144 registros, o vehículos coincidentes con el filtro establecido.

Imagen identificativa

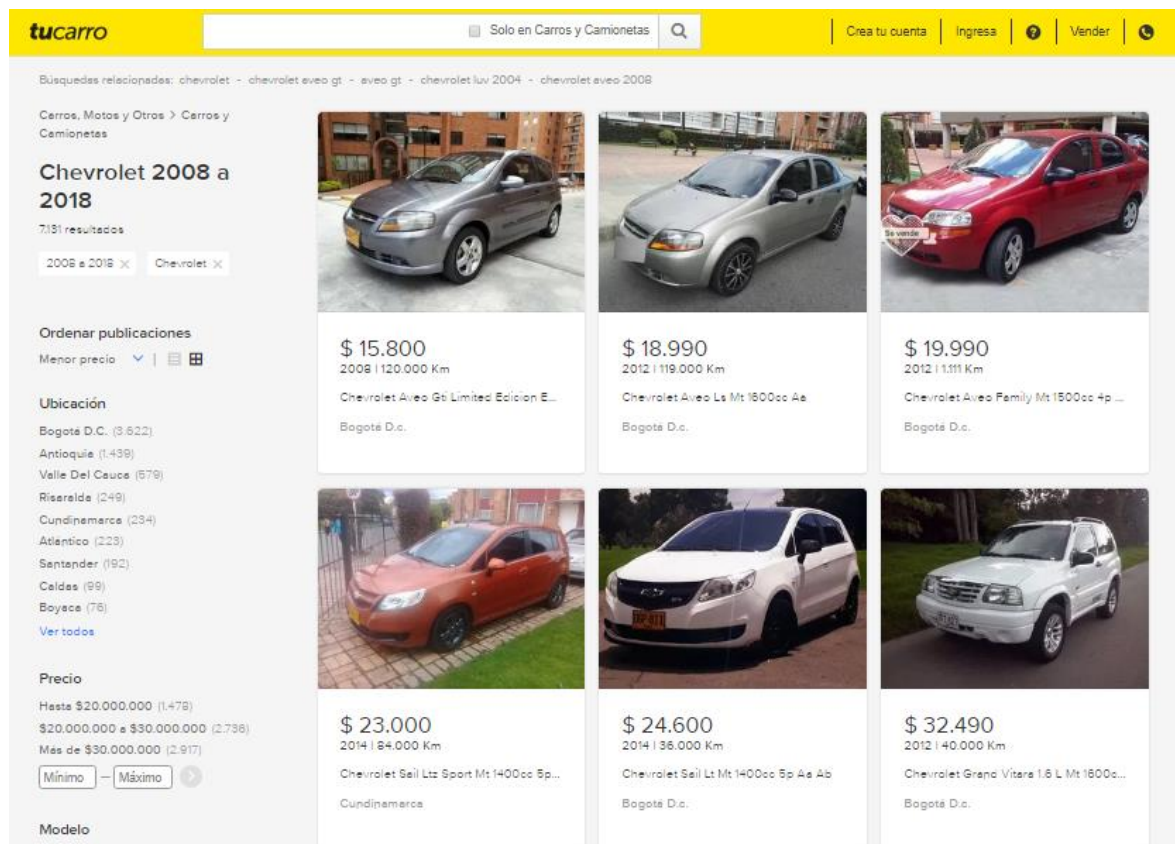


Ilustración 1: Imagen de muestra del portal web

Contexto

Este interesante portal web permite a sus visitantes acceder a un listado de vehículos que se encuentran en venta en Colombia, vehículos de distintas marcas, precios, modelos y ubicación dentro del territorio nacional.

En este ejercicio didáctico de Web Scraping, se ha conseguido transformar esta información en un dataset para un posible análisis futuro. El script adjunto al presente repositorio permite filtrar cierta información que podría ser relevante para el usuario, como lo es: el rango de modelo de los vehículos, y la marca.

Contenido

El dataset consta de los siguientes atributos:

- Descripción del vehículo
- Modelo y Kilometraje
- Precio de venta
- Ubicación.

Los datos recogidos están actualizados hasta el día 10 de noviembre de 2018, se pasó por parámetros al Script como modelo inicial 2008, modelo final 2018, y marca Chevrolet.

Ej:

Python sctucarro.py 2008 2018 Chevrolet

Agradecimientos

Los datos contenidos en el dataset del presente ejercicio son propiedad del portal **tucarro.com**.

Inspiración

Los motivos que me inspiraron a realizar el presente ejercicio sobre este reconocido portal, son mi gusto por los automóviles, mi interés en adquirir uno a futuro y la complejidad que presenta este sitio web en su código.

Licencia

La licencia seleccionada para la presente obra es **Released Under CC BY-NC-SA 4.0 License**, toda vez que se trata de una práctica de contexto netamente académico, con la cual no se pretende obtener ningún beneficio económico o comercial, y podrá ser usada como ejemplo para realizar scripts derivados del mismo haciendo uso del mismo tipo de licencia.

Codigo Fuente

El código fuente de este ejercicio fue desarrollado en el lenguaje de programación Python v 3.7, no se ha probado en otras versiones, la sintaxis de uso es la siguiente:

Python sctucarro.py Year1 Year2 [brand]

Recursos

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