

# Práctica 1

Integrantes:

Victor Morant, Aitor Jara

1. En el contexto de la venta de aceite y lubricante para motor, buscamos recabar información sobre las principales marcas comercializadas actualmente, así como información sobre sus respectivos productos. Por ello nos parece una buena idea centrarnos en la construcción de un conjunto de datos a partir del portal web de una empresa pequeña/mediana dentro del sector, ya que dichos datos proporcionan información clave al respecto.
2. El título seleccionado para el *dataset* es el de **Ánàlisis de Aceites de Motor para Vehículos de Combustión**
3. Esta práctica se ha realizado bajo el contexto de la asignatura Tipología y ciclo de vida de los datos, perteneciente al Máster en Ciencia de Datos de la Universitat Oberta de Catalunya. En ella, se aplican técnicas de *web scraping* mediante el lenguaje de programación Python para extraer así datos de la web <https://www.tuaceitedemotor.com> y generar un *dataset*.

4.

	Name	Brand	Quantity (Liters)	Price (€)	Datetime	Price by Liter (€)	Found in link (url)
1							
2	Motul Gear 300 75W90 1L	Motul	1.0	15.25	2020-11-07 18:49:49.524540	15.25	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
3	Motul Gear Competition 75w140 1L	Motul	1.0	17.7	2020-11-07 18:49:49.524540	17.7	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
4	Castrol Syntans Multivehicla 75w90 1L	Castrol	1.0	12.86	2020-11-07 18:49:49.524540	12.86	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
5	Castrol Syntans V FE 75w80 1L	Castrol	1.0	12.34	2020-11-07 18:49:49.524540	12.34	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
6	Ceroll Racing Coolant 1L	Ceroll	1.0	14.8	2020-11-07 18:49:49.524540	14.8	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
7	Repsol Elite Long Life 50700 50400 5w30 5L	Repsol	5.0	27.8	2020-11-07 18:49:49.524540	5.5600000000000005	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
8	Total Quartz Ineo First 0W30 5L	Total	5.0	34.57	2020-11-07 18:49:49.524540	6.914	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
9	Total Quartz Ineo ECS 5w30 5L	Total	5.0	29.35	2020-11-07 18:49:49.524540	5.87	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
10	Motul Specific VW505 01 502 00 5W40 5L	Motul	5.0	27.2	2020-11-07 18:49:49.524540	5.4399999999999995	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
11	Motul 8100 Eco-clean+ 5W30 C1 5L	Motul	5.0	39.0	2020-11-07 18:49:49.524540	7.8	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
12	Total Quartz 7000 10w40 5L	Total	5.0	18.71	2020-11-07 18:49:49.524540	3.742	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
13	Euroil Syntence FS 0w30 5L	Euroil	5.0	44.07	2020-11-07 18:49:49.524540	8.814	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
14	Euroil Evolence 0w20 5L	Euroil	5.0	26.84	2020-11-07 18:49:49.524540	5.368	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
15	Motul Specific VW 508.00 509.00 0w20 5L	Motul	5.0	50.82	2020-11-07 18:49:49.524540	10.164	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
16	Motul 7100 4T 10W40 1L	Motul	1.0	11.4	2020-11-07 18:49:49.524540	11.4	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
17	Motul 7100 4T 10W60 4L	Motul	4.0	37.87	2020-11-07 18:49:49.524540	9.4675	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
18	Motul 300V Competition 15W50 5L	Motul	5.0	63.15	2020-11-07 18:49:49.524540	12.629999999999999	<a href="https://www.tuaceitedemotor.com//">https://www.tuaceitedemotor.com//</a>
19	Castrol ATF Dexron II Multivehicla 1L	Castrol	1.0	6.99	2020-11-07 18:49:49.524540	6.99	<a href="https://www.tuaceitedemotor.com/aceite-cambio-automat">https://www.tuaceitedemotor.com/aceite-cambio-automat</a>
20	Castrol Transmax ATF Dex III Multivehicla 1L	Castrol	1.0	9.21	2020-11-07 18:49:49.524540	9.21	<a href="https://www.tuaceitedemotor.com/aceite-cambio-automat">https://www.tuaceitedemotor.com/aceite-cambio-automat</a>
21	Castrol Transmax ATF Dex III Multivehicla 20L	Castrol	20.0	137.76	2020-11-07 18:49:49.524540	6.888	<a href="https://www.tuaceitedemotor.com/aceite-cambio-automat">https://www.tuaceitedemotor.com/aceite-cambio-automat</a>
22	Castrol Transmax CVT 1L	Castrol	1.0	15.42	2020-11-07 18:49:49.524540	15.42	<a href="https://www.tuaceitedemotor.com/aceite-cambio-automat">https://www.tuaceitedemotor.com/aceite-cambio-automat</a>

5. El *dataset* contiene diferentes atributos que pretenden responder a varias aspectos del producto. Estos atributos y su significado son los siguientes:

- Name:** Título que describe el producto
- Brand:** Marca fabricante del producto
- Quantity (Liters):** Cantidad contenida en el producto
- Price (€):** Precio del producto
- Datetime:** Fecha y hora de la extracción
- Price by Liter (€):** Precio por litro
- Found in link (url):** url de los datos extraídos

6. El propietario de este conjunto de datos es Sergi Dols Icardo, fundador de la empresa, y hasta la fecha no hay citas de investigación o análisis anteriores (que sean públicos), sobre la empresa o sus productos.
7. La idea de este proyecto viene motivada por la necesidad de recabar información sobre uno de los productos más recurridos en el sector del transporte, que es el lubricante para motor. Con ello lo que se pretende es, primero, tener una idea de las principales marcas y productos que se comercializan hoy en día, y dos, conocer los precios unitarios (por litro) de los mismos. Con todo ello podría resultar interesante para, por ejemplo, un particular, intentar descubrir alguna tendencia que pueda ayudar en la toma de decisiones, por ejemplo, la de realizar, o no, una compra de un producto concreto. Así mismo, para una empresa que busque irrumpir en el sector podría ser interesante conocer el catálogo de productos que se está comercializando actualmente para poder hacer una primera aproximación de los productos que debería tener en cuenta para realizar una primera compra de stock.
8. La licencia escogida para el *dataset* ha sido la de *Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International* (CC BY-NC-SA 4.0 License), ya que es una licencia muy permisiva en términos de reutilización del conjunto de los datos y la única restricción es la de evitar su uso para fines comerciales, citando, además, la fuente original.
9. El repositorio se encuentra en el siguiente enlace:  
<https://github.com/UOC-DataTypCyc-Team/PEC1>
10. El dataset, además de en el enlace de GitHub, se puede encontrar también en el portal web de Zenodo, en el siguiente enlace:

<https://zenodo.org/record/4263160#.X6gnL1NKjOQ>

El DOI: 10.5281/zenodo.4263160

11. Las principales contribuciones según apartado, por los contribuyentes:

**Contribuciones Firma**

Investigación previa	V.M	A.J
Redacción de las respuestas	V.M	A.J
Desarrollo código	V.M	A.J