

# Tipología y ciclo de vida de los datos - Práctica 1

---

La práctica ha sido realizada individualmente por Ángel Baltasar Sánchez

1. Título del dataset. Poned un título que sea descriptivo.

PRODUCTOS\_BARRABES\_ONLINE\_10\_2017

2. Subtítulo del dataset. Agregad una descripción ágil de vuestro conjunto de datos por vuestro subtítulo.

Lista de productos ofertados en la tienda online [www.barrabes.com](http://www.barrabes.com), disponibles a día 31/10/2017.

3. Imagen. Agregad una imagen que identifique vuestro dataset visualmente



La imagen sería descriptiva de un humano yendo a la montaña, y nuestro dataset es un conjunto de elementos necesarios para que el ser humano pueda explorar la montaña

4. Contexto. ¿Cuál es la materia del conjunto de datos?

El conjunto de datos es un listado de productos que se venden online en la web [www.barrabes.com](http://www.barrabes.com), estos productos incluyen diferentes categorías: ropa de montaña de hombre, mujer y niños, calzado de montaña, productos de esquí, alpinismo y trekking así como material especializado en montaña.

5. Contenido. ¿Qué campos incluye? ¿Cuál es el periodo de tiempo de los datos y cómo se ha recogido?

Los datos han sido capturados mediante la técnica de Web Scrapping utilizando un script en Python basado en el script de búsqueda que se puede encontrar en la siguiente referencia de la bibliografía de la asignatura:

Richard, Lawson. Web Scrapping with Python. Packt publishing Ltd, 2015

Los datos capturados se corresponden al listado de productos a día 31/10/2017

Cada linea del dataset contiene información relacionada con un producto en concreto, los datos han sido extraídos mediante la interpretación de las etiquetas html <meta ...> que están contenidas en cada página de producto del website. Los diferentes campos capturados son:

SKU : Identificador único del producto dentro de la tienda,  
categoria: Categoría en la que se ubica al producto dentro de la tienda online  
url: URL de acceso al producto que hemos accedido mediante Web Scrapping  
marca: Marca de l producto  
nombre: nombre del producto  
precio: precio del producto en €  
descripcion: descripción extendida del producto  
imagen\_url: url a la imagen del producto  
url: url al producto que se incluye en los datos encontrados mediante web scrapping  
caract1: Características del producto  
caract2: Características del producto  
caract3: Características del producto  
caract4: Características del producto  
caract5: Características del producto  
caract6: Características del producto  
caract7: Características del producto  
caract8: Características del producto  
caract9: Características del producto  
caract10: Características del producto

6. Agradecimientos. ¿Quién es propietario del conjunto de datos? Incluid citas de investigación o análisis anteriores.

Los datos son propiedad de BARRABES SKI MONTAÑA S.L.U., propietaria de la web [www.barrabes.com](http://www.barrabes.com)

7. Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaría responder la comunidad?

El motivo de extraer este conjunto de datos es poner en prueba las técnicas de Web Scrapping con Phyton aprendidas en la asignatura “tipología y ciclo de vida de los datos” del máster en ciencia de datos de la UOC.

Con esta base de datos, podemos ver los productos, categorías y precios que ofrece [barrabes.com](http://barrabes.com) en su tienda online.

Podría ser posible estimar que marcas son sus best-sellers teniendo en cuenta la diversidad de productos que tienen de cada marca, entendiendo que cuando tienes un best-seller ofrecerías tantos productos de dicha marca como pudieras.

#### 8. Licencia. Seleccionad una de estas licencias y decid porqué la habéis seleccionado:

La licencia que selecciono sería CC BY-SA 4.0 License, ya que desearía que:

- Pueda ser distribuido en cualquier otro medio o formato
- Pueda ser tratado y transformado
- Cualquier tratamiento se le haga referencia la fuente original
- Sin fines comerciales, ya que se trata de un trabajo educativo
- Y sea compartida de igual a igual, es decir utilizando esta misma licencia

Añadiremos en GitHub, en la sección de Wiki la siguiente linea.

This work is licensed under a Creative Commons Attribution Share-Alike 4.0 License.

#### 9. Código: Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python, que os ha ayudado a generar el dataset

El código está contenido en el fichero “webScrappingBarrabes.py”

#### 10. Dataset: Dataset en formato CSV

Los datos están contenidos en el fichero “productos.csv”