

## EJERCICIO 1

Para la realización de nuestra práctica, el contexto que hemos planteado es la apertura de una tienda online de venta de calzado exclusivo. A la hora de establecer una ventaja competitiva respecto a la competencia, una de las claves va a ser estar, mínimamente, dentro de la media de un precio de mercado en el que se mueva nuestra competencia.

Por ello hemos decidido hacer un scrapping sobre la sección de calzado de un líder en el sector, para tener una referencia en cuanto a precios y con el conocimiento sobre los precios que nos ofrecen nuestros proveedores ver en que margen de beneficio se mueven, además vamos a sacar los descuentos para poder entender que estrategias siguen a la hora de ofertar sus productos.

Planteamos también mediante este scrapping poder hacer price matching para poder ser altamente competitivos en el mercado al que iría esta hipotética empresa.

## EJERCICIO 2.

El título del dataset obtenido es SVD\_sneakers.csv

## EJERCICIO 3.

El dataset cuenta con 1521 rows sin contar el header y 6 atributos por cada row:

```
(1521, 6)
```

Los atributos son:

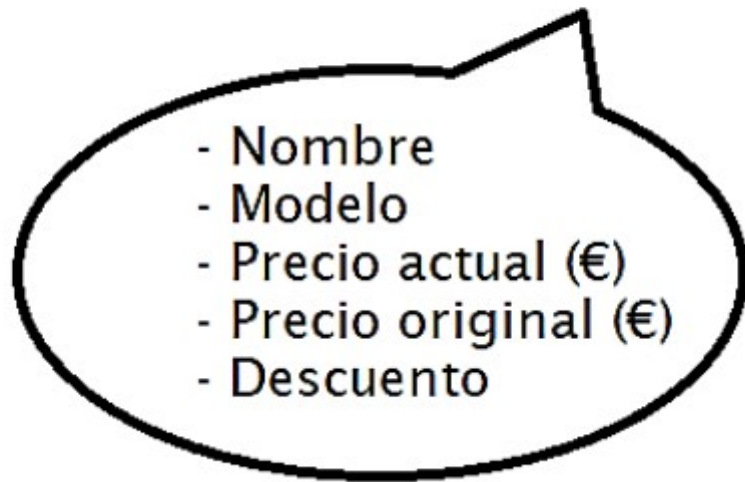
- Id: identificador único, type integer.
- Name: nombre del producto, type string.
- Model: modelo del producto, type string.
- Actual price: precio actual del producto, type string.
- Old price: precio anterior del producto, type string.
- Discount: porcentaje aplicado de descuento, type string.

No hemos realizado transformaciones sobre los atributos dado que no era el objeto del ejercicio.

#### EJERCICIO 4.

	name	model	actual price	old price	discount
0	Converse	One Star OX	95,00	95,00	0%
1	Puma	Slipstream Low Retro	100,00	100,00	0%
2	Puma	Slipstream Low Retro	100,00	100,00	0%
3	Converse	Women's Run Star Motion	120,00	120,00	0%
4	Converse	Women's Chuck 70 Hi	95,00	95,00	0%

**S | V | D** ➡ CALZADOS



#### EJERCICIO 5.

Los campos obtenidos son los anteriormente mencionados en el ejercicio 3, el periodo temporal no es un ítem al que podamos tener acceso vía scrapping de la web por lo que son todos los modelos de calzado a la venta del día 7 de abril a las 12:51.

Para la obtención del dataset hemos usado las siguientes tecnologías:

- Entorno de desarrollo: JetBrains Pycharm
- Dev. tools: git
- Lenguaje: Python
- Librerías:
  - BeautifulSoup: extrae datos de archivos HTML y XML, proporciona formas ideomáticas de navegar, buscar y en general parsear esos archivos.
  - Pandas: herramienta de análisis y manipulación de datos, open-source rápida, potente, flexible y fácil de usar.

Hemos creado una clase con 5 métodos, los más relevantes para la extracción de la información son:

- `scrape_product_info()`: en éste método realizamos las requests a las páginas de la web para obtener el archivo el cual parsear con BeautifulSoup, una vez obtenido hacemos una búsqueda por el tag y la clase donde se almacena la información de los productos y a continuación aplicamos el método propio de nuestra clase, `set_product_attributes`
- `set_product_attributes()`: en este método recibimos como argumento la el objeto creado en BeautifulSoup respecto al conjunto de información del request a la página web y extraemos los atributos que queremos, guardándolos en una lista en Python para su más fácil manejo.
- Los demás métodos, tenemos un getter para la lista anterior, uno que nos crea una carpeta donde guardar la información y por último un método que nos lo transforma a objeto DataFrame de pandas y a .csv como se solicita en el ejercicio.

## EJERCICIO 6

Agradecer al propietario de la página web, Proudplaces, S.L. (SVD) por poner a disposición publica los datos extraídos de este proyecto.

Los pasos seguidos a nivel ético y legal:

- Hemos comprobado que la obtención del dataset no contraviene el legal notice de la página web:
  - <https://www.sivasdescalzo.com/es/legal-notice>
- Comprobamos que en ningún caso se está extrayendo información que vulnere derechos de autor o de marca registrada.
- Hemos comprobado el robots.txt de la web viendo que el autor no dice nada al contrario de nuestras acciones.
- Sólo hemos rastreado información pública.
- No hemos causado daño sobrecargando el servidor con muchas peticiones, ítem que hemos tenido en cuenta en el código con el que se ha realizado el scrapping.

## EJERCICIO 7

El conjunto de datos nos permite obtener información detallada acerca del calzado a la venta de la compañía SiVasDescalzo. Esta empresa, fundada en 2011 con una inversión inicial de 90.000 € ha crecido exponencialmente desde entonces, siendo ahora una de las referencias en la venta de deportivas exclusivas y obteniendo la mayor parte de sus beneficios en la venta online de sus productos. Por esta razón, analizar la información de su página web puede ayudarnos a entender sus estrategias de negocio y utilizar este conocimiento para aumentar los beneficios de nuestra hipotética empresa.

## EJERCICIO 8

Licencia CC BY-NC-SA 4.0, la elegimos porque es de las más comunes en cuanto a publicación de datasets, la licencia permite compartir y adaptar el conjunto de datos pero sin propósitos comerciales.

## EJERCICIO 9

<https://github.com/martinleal5/Pra1-Web-scraping>

## EJERCICIO 10

FALTA LINK DEL VIDEO

Contribuciones	Firma
Investigación previa	EPM, MLM
Redacción de las respuestas	EPM, MLM
Desarrollo del código	EPM, MLM