

# Ejercicios sobre Procesamiento de información en Internet (Web Scraping)

## Ejercicio resuelto. Nombres de bebés: de Internet a un DataFrame

En la siguiente url,

<https://www.enterat.com/servicios/nombres-nino-nina.php>

hemos encontrado una tabla con los nombres de bebés más populares en 2019, en el sentido de ser los más elegidos por los padres para bautizar a sus hijos:

### Los 100 nombres de niño y niña más puestos en España (sean o no en español)

Posición	Nombres de niño	Nombres de niña
1	HUGO	LUCIA
2	LUCAS	SOFIA
3	MARTIN	MARTINA
4	DANIEL	MARIA
5	PABLO	PAULA
6	MATEO	JULIA
7	ALFONSO	EMMA

Se pide acceder a esa página desde Python, localizar la tabla y convertirla en un DataFrame para poder procesarla posteriormente.

## Apartado 1

En primer lugar, debes hacer lo siguiente mediante un programa en Python: (a) acceder a la url, (b) cargar el texto de la página, (c) buscar la tabla, (d) cargar las filas (etiqueta tr) y (e) las celdas (etiqueta td).

Comprobamos el resultado de estos pasos imprimiendo alguna fila y alguna celda.

## Apartado 2

Ahora iniciamos en procesamiento de los datos que tenemos, y vamos comprobando el resultado paso a paso. (a) Selecciona las celdas de una fila cualquiera, y (b) extrae el contenido (el texto) de las celdas de esa fila. (c) El dato entero que indica el número de orden debería ser un entero, y no una cadena de caracteres...

## Apartado 3

Es frecuente necesitar cada columna en una lista (porque los datos de cada columna son homogéneos, y así se facilita su procesamiento en algunos programas. Prepara los datos en tres listas, una por cada columna, y almacena el resultado en sendas listas.

## Apartado 4

Pasa el contenido de estas columnas a un DataFrame de la librería pandas.

