

# Práctica 1 Web Scraping

Alejandra Cristina Marrero Suarez  
Luis Miguel Moreno López

# Contexto

Nuestro proyecto surge con el objetivo de **aventurar qué géneros y autores de libros están teniendo mayor impacto**, tendencia o éxito en España en un período reciente de tiempo.

Se decidió seleccionar los libros más vendidos recientemente usando la técnica de *web scraping* sobre la **página web [www.todos.tuslibros.com](http://www.todos.tuslibros.com)**, en particular sobre el subdominio *mas\_vendido* donde aparecen los 100 libros más vendidos.

No es casualidad elegir esta página como fuente de datos ya que dentro de esta asociación se encuentran más de 400 librerías de todo el país, siendo así una muestra realmente representativa.

## Título del dataset

Titulamos nuestro dataset: ***libros\_más\_vendidos***.

## Descripción del dataset

Ofrecemos la información habitual asociada a un libro (autor, número de páginas, ...) de los 100 libros más vendidos en este momento en España.

## Representación gráfica

ISBN:	Idioma de publicación :	Idioma original :	Ilustrador:	Nº páginas:	País de publicación :	Peso:	Traductor:	autor	materias	título	
978-84-9066-929-7	Español	Español		400	España	559 gramos		Terra Alta	Ficción moderna y contemporánea	Independencia	
978-84-204-5459-7	Español	Español		668	España	100 gramos			Ficción moderna y contemporánea	Tomás Nevinson	
978-84-17860-79-0	Español	Español		452	España	744 gramos		La invenci	Ensayos literarios	El infinito en un junco	
978-84-9066-848-1	Español	Español		240	España	352 gramos			Ficción moderna y contemporánea	El huerto de Emerson	
978-84-670-6113-0	Español	Español		312	España	526 gramos			Ficción moderna y contemporánea	Delparaiso	
978-84-322-3773-7	Español	Español		320	España	426 gramos			Ficción moderna y contemporánea	Uévame a casa	
978-84-16327-81-2	Español	Inglés		320	España	416 gramos		Terés Lorient, Mireia	Narrativa romántica histórica	El duque y yo (Bridgerton 1)	
978-84-339-8087-8	Español	Inglés		384	España	445 gramos		Bach, Mauricio	Ficción moderna y contemporánea	Klara y el sol	
978-84-233-5868-7	Español	Español		304	España	218 gramos			Ficción moderna y contemporánea	Los privilegios del ángel	
978-84-08-23551-4	Español	Español		416	España	612 gramos		Premio PI	Obra de misterio y suspense	Aquitania	
978-84-9066-929-7	Español	Español		400	España	559 gramos		Terra Alta	Ficción moderna y contemporánea	Independencia	
978-84-204-5459-7	Español	Español		668	España	100 gramos			Ficción moderna y contemporánea	Tomás Nevinson	
978-84-17860-79-0	Español	Español		452	España	744 gramos		La invenci	Ensayos literarios	El infinito en un junco	
978-84-9066-848-1	Español	Español		240	España	352 gramos			Ficción moderna y contemporánea	El huerto de Emerson	
978-84-670-6113-0	Español	Español		312	España	526 gramos			Ficción moderna y contemporánea	Delparaiso	
978-84-322-3773-7	Español	Español		320	España	426 gramos			Ficción moderna y contemporánea	Uévame a casa	
978-84-16327-81-2	Español	Inglés		320	España	416 gramos		Terés Lorient, Mireia	Narrativa romántica histórica	El duque y yo (Bridgerton 1)	
978-84-339-8087-8	Español	Inglés		384	España	445 gramos		Bach, Mauricio	Ficción moderna y contemporánea	Klara y el sol	
978-84-233-5868-7	Español	Español		304	España	218 gramos			Ficción moderna y contemporánea	Los privilegios del ángel	
978-84-08-23551-4	Español	Español		416	España	612 gramos		Premio PI	Obra de misterio y suspense	Aquitania	
978-84-18524-00-4	Español	Japonés		96	España	220 gramos		Morinaga, Makoto	Forma física y alimentación	Adelgaza con Kaoru	
978-84-17977-58-0	Español	Inglés		352	España	460 gramos		Cardeloso, Concha	Ficción histórica	Hamnet	
978-84-08-23796-9	Español	Español		384	España				Ficción moderna y contemporánea	Terra Alta	
978-84-17921-37-8	Español	Español	FORÉS, David G.	192	España				Thriller (infantil/juvenil)	Una herencia peligrosa (Aman	
978-84-18107-59-7	Español	Sueco		512	España	750 gramos			Ficción histórica	1794	
978-84-339-9903-0	Español	Español		192	España	268 gramos			Obra de misterio y suspense	Un amor	
978-84-9838-149-8	Español	Francés	Saint-Exupéry, Antoine de	96	España	146 gramos		CARRIL, BONIFACIO DEL	Cuestiones personales y sociales: au	El Principito	
978-84-666-6854-5	Español	Español		528	España	900 gramos			Obra de misterio y suspense	Rey blanco	
978-84-666-6441-7	Español	Español		568	España	820 gramos		Francisco Ferriz Blanquer;	Género policiaco y misterio	Reina roja	
978-84-233-5877-9	Español	Español		304	España	536 gramos			Premio NI	Ficción moderna y contemporánea	El lunes nos querrán
978-84-204-5466-5	Español	Español		688	España	1127 gramos			Ficción histórica	Línea de fuego	
978-84-18107-57-3	Español	Italiano		272	España			Mayor, Carlos	Sicilia	La red de protección (Comisar	
978-84-204-5432-0	Español	Español		208	España	320 gramos			Ficción moderna y contemporánea	Miss Marte	
978-84-16327-82-9	Español	Inglés		352	España	454 gramos		Arruti Illaramendi, Rosa	Narrativa romántica histórica	El vizconde que me amó (Brid)	
978-84-18526-13-8	Español	Español		576	España	590 gramos			Una lectur	Ficción moderna y contemporánea	Biografía de la luz

# Contenido

Los datos obtenidos son los siguientes:

- a. **Título:** El título del libro.
- b. **Autor:** Persona que ha realizado el libro.
- c. **Editorial:** Casa editora que publica el libro.
- d. **Url:** Enlace a la web del libro en el dominio de Todostuslibros
- e. **Colección:** Conjunto ordenado de libros al que pertenece.
- f. **Dimensiones:** Tamaño del libro
- g. **EAN:** número del artículo europeo, numeración que identifica el artículo.
- h. **Encuadernación:** Formato del libro siendo esta de bolsillo o tapa dura.
- i. **Fecha de publicación:**
- j. **ISBN:** código normalizado internacional para libros
- k. **Idioma de publicación:** Idioma del país donde se esté vendiendo.
- l. **Idioma original:** Idioma original en que se escribió.
- m. **Ilustrador:** Artista gráfico que se especializa en la mejora de la comunicación escrita, a través de representaciones visuales que se corresponden con el contenido.
- n. **Nº de páginas:** nº de páginas que contiene cada libro.
- o. **País de publicación:** País original donde se publicó el libro.
- p. **Peso:** peso del libro
- q. **Traductor:** Persona encargada de traducir el libro.
- r. **Materias:** Temas o géneros que toca el libro.

El proceso de obtención de datos se realiza recorriendo las 10 páginas de libros donde está contenida información básica de 10 libros del ranking. Además en cada una de estas páginas también disponemos de 10 urls, cada una de ellas nos lleva a información más característica de un determinado libro.

# Agradecimientos

Tal como expone la web de todostuslibros en su apartado de “quienes somos”, los datos de los libros más vendidos son proporcionados directamente por todas las librerías españolas asociadas a su proyecto. Por tanto, nuestro dataset es directamente obtenido de estas librerías.

## Inspiración

En la misma línea de trabajo encontramos datasets de años anteriores. **La empresa Statista ya ha realizado algún dataset relacionado con los libros más vendidos.** De forma particular cuenta con el dataset de libros más vendidos en España de 2020.

<https://es.statista.com/estadisticas/808007/ranking-de-los-libros-mas-vendidos-espana/>

También podemos valorar en este espacio a **otras empresas con sus propias listas de “best sellers”** basadas en sus propias ventas como son Fnac, El corte Inglés, Amazon, Grupo Planeta, Carrefour ...

Volviéndonos a centrar en el dataset de Statista, que es el más representativo encontrado. El breve análisis que incorpora centra su atención sobre el autor que más vendió aunque simplemente a título de “primero de la lista”, sin contar posibles agregaciones o un análisis más amplio. Como algo distintivo añade el número de ejemplares vendidos de cada libro lo cual otorga mayor información respecto a la diferencia de ventas entre puestos. Si algo se le puede echar en cara a este dataset es que solo tiene datos de un ranking de diez libros, en comparación con nuestro dataset de cien posiciones.

Respondiendo a la pregunta de qué nos pueden aportar estos datos. Podemos ver su importancia en entender un mercado como el del libro cuya facturación interior en España ronda los 2400 millones de euros. Gracias a estos datos y estudios similares como los mencionados podemos comprender al consumidor de este mercado, el lector. Estos datos nos permiten ver sus inquietudes, sus temas favoritos, sus autores consagrados... y así adaptarnos a él. Un claro ejemplo práctico es poder establecer mayor stock de libros de determinados autores, organizar encuentros con estos autores e incluso descubrir potenciales nuevos intereses de los lectores tanto en temas como en escritores.

## Licencia

Para nuestro dataset optariamos por una licencia de base de datos abiertas (Database released under Open Database License) con el fin de que otros puedan obtener nuestros datos o contribuir en ellos. Para ello mantendremos nuestra base de datos abierta y sin ningún medio de pago con el fin de que nuestros datos se puedan usar como un componente más de un programa diseñado para la distribución comercial.

Contribuciones	Firma
Investigación previa	Luis Miguel Moreno Lopez, Alejandra Cristina Marrero Suárez.
Redacción de las respuestas	Luis Miguel Moreno Lopez, Alejandra Cristina Marrero Suárez.
Desarrollo del código	Luis Miguel Moreno Lopez, Alejandra Cristina Marrero Suárez.