

Descripción del dataset

El objetivo de esta actividad será la creación de un dataset a partir de los datos contenidos al web. Tenéis que indicar las siguientes características del dataset general:

Título del dataset

Libros ofertados en la página web de la librería *Elkar Liburu Denda*.

Subtítulo del dataset

El conjunto de datos generado en esta práctica recoge las características principales de los libros disponibles en [Elkar Liburu Denda](#). Las variables que se recogen en el conjunto de datos son el título, el autor / la autora, precio de venta, precio de venta para los clientes que disponen de la Tarjeta Cliente y el link que contiene información aún más detallada.

Imagen

La página web de donde se han extraído los datos es de la siguiente manera:

Quiénes somos | Donde estamos | Recibir boletín | Trabaja con nosotros

euskara | **español** | english | français

ekar

email
 Contraseña

LIBROS MÚSICA E-BOOK DVD JUEGOS PAPELERÍA

Mi cesta 0 0.00 €

Ordenar por: Pag. 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14

Filtros

LIBROS

LITERATURA

LIT.INFANTIL Y JUVENIL

OCIO

TEMA VASCO

BELLAS ARTES

ECONOMÍA Y DERECHO

TEXTO E IDIOMAS

CIENCIAS

CIENCIAS HUMANAS Y SOCIALES

REVISTAS

PRECIO

< 5 €

> 5 € < 10 €

> 10 € < 20 €

> 20 € < 30 €

> 30 € < 40 €

IDIOMA

EUSKARA

CASTELLANO

INGLÉS

FRANCÉS

CATALÁN

SIT.

FN STOCK

El color de las mareas
 MIKEL ALVIRA PALACIO S
 23,00 € 21,85 €

Bioklik
 ABEL LUZARRAGA ZARRABETIA
 25,00 € 23,75 €

Tretatxu, Lapurren Gobernadores
 HECTOR ORTEGA LAHERA
 20,00 € 19,00 €

La bruja
 CAMILLA LACKBERG
 24,00 € 22,80 €

Zenbakiak Asmatzen
 GIANNI RODARI / ALESSANDRO BANNA (IL.)
 15,00 € 14,25 €

Donde Mis Pies Me Lleven
 ERIKA RAMOS
 17,90 € 17,01 €

Zaldi Mamarroa (xix. Igartza Saria)
 EKAITZ GOIENETXEA CERECEDA

Toda La Verdad
 KAREN CLEVELAND

Columna De Fuego, Una - Los Pilares De
 KEN FOLLETT

La jaula de sal
 IBON MARTIN

Arbasoen Ituna
 TOTI MARTINEZ DE LEZEA

Ez Duzu Abusatuko - Pederastia Kasuak
 ALBERTO BARANDIARAN / IÑIGO

Contexto

Los datos han sido recolectados desde página web de la librería más importate de la CAV: [Elkar Liburu Denda](#). Estos establecimientos venden libros, música, e-books, papelería, juguetería educacional... Para poder extraer solo los datos que nos interesaban, se han hecho uso de técnicas de *Web Scraping* aplicadas mediante el lenguaje de programación Python.

Contenido

Las variables que se recogen en el conjunto de datos son el **título** (cadena de caracteres), **el autor / la autora** (cadena de caracteres), **precio de venta** (float), **precio de venta para los clientes que disponen de la Tarjeta Cliente** (float) y **el link** que contiene información aún más detallada (cadena de caracteres).

Agradecimientos

Como se ha mencionado anteriormente, el conjunto de datos se ha generado desde la página web de [Elkar Liburu Denda](#), el objetivo no ha sido en ningún momento obtener beneficio económico y se han tomado las medidas necesarias para evitar la vulneración de los derechos de propiedad intelectual de los titulares de la página web.

Inspiración

La idea de la construcción del conjunto de datos nace de mi interés hacia la literatura. Quizás por querer hacer frente a la extendida falsa idea de la guerra entre números y letras, me pareció buena idea aplicar mis primeras nociones en *Web Scraping* en una tienda de libros. En principio, el objetivo del dataset generado era poder aplicarlo en el ámbito bibliotecario, introduciendo la evolución de ventas, los/las bibliotecarios/as tendrían a su disposición de una herramienta muy útil para poder satisfacer de los/las usuarios. Con la evolución de la práctica, también he visto que podría ser muy útil para los colegios, centros culturales o incluso para editoriales. Cabe destacar la gran fuente de interés que fue [este blog](#).

Licencia

Seleccionad una de estas licencias y decid por qué la habéis seleccionado:

- Released Under CC0: Public Domain License
- Released Under CC BY-NC-SA 4.0 License
- Released Under CC BY-SA 4.0 License
- Database released under Open Database License, individual contents under Database Contents License
- Other (specified above)
- Unknown License

La licencia escogida para la publicación de este conjunto de datos ha sido CC BY-NA 4.0 License. Por los siguientes motivos: se debe otorgar el crédito apropiado, proporcionar un enlace a la licencia e indicar si se realizan cambios, no puede usarse el material con fines comerciales y en el caso de que se remixe, transforme o construya sobre el material, debe distribuir sus contribuciones bajo la misma licencia que el original.

Código

*El código con el que se ha generado el dataset, generado en Python, está en la carpeta **Código**.*

Dataset: Dataset en formato CSV

*El Dataset generado está en la carpeta **Fichero-csv**.*