Caso práctico: Identificar y extraer datos relevantes para un proyecto analítico, empleando herramientas específicas de web scraping

Solución PRA1 – Webscraping

Estudiante:

- **Daniel Cañete Román (dcanete@uoc.edu)**
- **Balpreet Kaur Singh (bkaur@uoc.edu)**

INDICE

1 Contexto

Explicar en qué contexto se ha recolectado la información. Explicar por qué el sitio web elegido proporciona dicha información. Indicar la dirección del sitio web.

La recolección de datos solamente tiene sentido cuando estos datos recabados se transforman en una información útil para que se pueda convertir en el conocimiento objetivo de nuestro estudio/análisis final según un contexto dado. Por lo tanto, primero de todo comentemos el entorno en el cuál hemos decidido llevar a cabo la captura de los datos para su posterior análisis.

Empezaremos mencionando los datos que estaremos recolectando: se trata de una tienda virtual de productos de ferretería llamada “Ferreterias Industriales”. Como podemos ver en la propia página web (www.ferreteriaindustriales.com), hay una gran variedad de productos disponibles.

Para el desarrollo del ejercicio se ha partido de un caso de negocio inventado, pero que podría ser similar a un caso real:
“Somos una empresa de reciente creación que pretende abrir una ferretería en un barrio residencial. El plan de negocio es muy ambicioso y pretende ampliar en los próximos años con nuevos locales, estableciendo una marca que se comercializará en un futuro como franquicia.
Una tarea muy importante para el futuro negocio es establecer un catálogo que estará compuesto de la suma de los productos en venta por varios proveedores que hemos seleccionado. Algunos de estos proveedores nos han enviado su catálogo en versión digital, pero no es el caso de "Ferreterías Industriales", que nos remite a su página web.
Tras hablar con este proveedor en particular, hemos acordado que podemos obtener la información de su página web a condición de que no saturamos sus servidores y obtengamos los datos fuera del horario no laboral, que es cuando tienen comprobado que sus clientes acceden más a la web.
En una segunda fase, unificaremos la oferta de todos los proveedores, seleccionando los productos más interesantes para nuestro catálogo. Para ello tendremos que buscar una única clasificación y estructura de los datos común.”

Por tanto, en el ámbito de esta práctica, el principal objetivo es recopilar los datos de los productos pero es importante que estos datos dispongan todas las características posibles, ya que tenemos que poder comparar con los productos de otros proveedores para seleccionar los más competitivos de cada categoría.

Así, entre los objetivos principales de estudio, podemos analizar:

- Los productos disponibles online que tiene.
- A qué precios está ofertando sus productos.
- Los puntos fuertes y débiles del proveedor frente a otros proveedores.
- A partir de los datos recolectados, seleccionar los productos que se quedarán en nuestro catálogo y márgenes de negocio con el que se incrementarán las tarifa.
- Estudiar nuestra catalogación efectiva y eficiente de productos y variantes a partir de la categorización y oferta de productos de nuestro proveedor.

A partir de estos datos podremos estudiar, como hemos dicho, a nuestro proveedor principal conociendo sus productos y precios a partir de su portafolio disponible online y gracias a la recopilación de datos y su posterior análisis y convertirlo así en conocimiento útil que podremos aplicar en nuestra estrategia de compras, necesario para poder tener mejores márgenes y mejorar la oferta de nuestros productos y, por consiguiente, nuestra cuota de mercado de forma eficaz y eficiente basándonos en la estrategia de datos.

2 Título

Definir un título conciso y que sea descriptivo para el dataset.

Se propone en título **"Productos de Ferretería"** ya que cumple con las siguientes características:

- Sencillo
- Fácil de recordar
- Pone fácilmente en contexto de a qué se refiere

Se han estudiado otras opciones:

- Ponerlo en inglés, pero todo está en español, incluido campos y nombres de los productos
- “Productos Ferreterías Industriales”, pero particulariza en la fuente, cuando lo interesante del dataset son los productos, precio y clasificación.

3 Descripción del dataset

Desarrollar una breve descripción del conjunto de datos que se ha extraído. Es necesario que esta descripción sea coherente con el título elegido.

Un dataset es una forma de salvar un tipo de datos o los contenidos de una única tabla de datos donde cada columna de la tabla representa una variable en particular y cada fila representa a una muestra determinada del conjunto de datos.

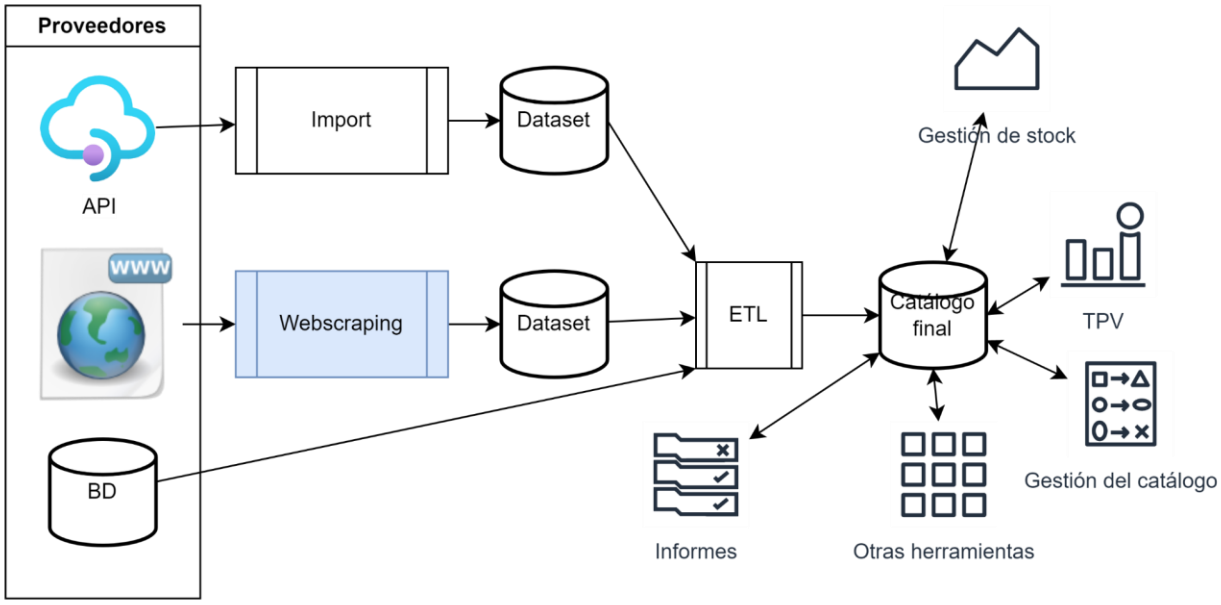
En el caso particular de la práctica, podemos definirlo como un dataset que incluye datos de varios productos ofertados en la plataforma online de “Ferreterías Industriales”, con sus correspondientes características que incluye el título del producto, precio, fabricante, identificador o sku, tipo, marca comerciante, forma de presentación, número de referencia, diámetro, capacidad, categoría y subcategoría.

4 Representación gráfica.

Dibujar un esquema o diagrama que refleje visualmente el dataset y el proyecto elegido.

Como se comentaba en el capítulo de contexto, en el escenario completo, la parte de webscraping es sólo una parte de todo el proceso de generación de catálogo. Se hace este proceso porque en el caso de uno de los proveedores no se dispone de otras alternativas como el acceso directo al catálogo en BD, o el acceso a través de una API.

Tras la obtencion de los dataset es necesario generar procesos de carga, transformación y grabación de datos (ETL) para generar el catálogo final, del que se podrán nutrir las herramientas que sean necesarias para el ejercicio normal de la empresa de ferretería.



5 Contenido

Explicar los campos que se incluyen en el dataset y el período de tiempo al que pertenecen los datos.

Se captura la información más relevante disponible de cada uno de los productos. Se compone de los siguientes campos:

- **url**: Enlace. Dirección web donde está la información del producto
- **title**: Cadena. Nombre del producto
- **price**: Importe. Precio del proveedor para el producto. Es importante tener en cuenta que es en euros y con impuestos incluidos. Al ser un proveedor, el IVA se podrá repercutir.
- **fabricante**: Lista de valores. Identifica al fabricante del producto.
- **sku**: Numérico. Referencia del producto. Es importante guardarlo porque será necesario utilizarlo cuando realicemos pedidos a este proveedor.
- **Articulo**: Cadena. Nombre alternativo del producto que se facilita en la ficha técnica
- **Presentación**: Cadena. Unidad de medida (opcional)
- Referencia Proveedor: Numérico. Otra referencia del producto. No se tiene claro en este punto cuál de las dos referencias se utilizará al hacer el pedido. Ante la duda, se almacenan los dos datos (opcional)
- **Diámetro**: Numérico. Diámetro del producto (opcional)
- **Capacidad**: Numérico. Volumen de almacenamiento del producto (opcional)
- **Categoría**: Lista de valores. Categoría en la que se ha clasificado el producto
- **Subcategoría**: Lista de valores. Subcategoría en la que se ha clasificado el producto

Si se considera interesante en futuras fases, **se podrían normalizar fácilmente los datos** sacando a tablas separadas valores posibles de Categoría, Subcategoría y fabricante.

6 Propietario

Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en su defecto, justificar esta búsqueda con análisis similares. Indicar qué pasos se han seguido para actuar de acuerdo con los principios éticos y legales en el contexto del proyecto elegido.

El propietario se denomina **JERONIMO MONTERRUBIO E HIJOS SL** y el domicilio social es Avd. Castilla y Leon nº25 - 05450 Casavieja (Avila), España.

Otros datos del propietario:

- Teléfono de contacto: 918. 678.722.
- Email: info@ferreteriaindustriales.es
- Datos registrales: Registro Mercantil de Avila
- Tomo 75
- Libre 33
- Folio 183
- Hoja nº AV-1813
- Inscripción 1ª
- Número de identificación fiscal: B05142633
- Contacto de Administración: José Alfredo Monterrubio Goméz.

Sobre los **pasos seguidos** se indican a continuación:

- No se han hecho pruebas en horario laboral, intentando hacerlo en **fines de semana y última hora del día**
- No es necesario tener todos los datos para conseguir los objetivos académicos, por lo que **se ha puesto un límite de productos a descargar.**
- Para evitar sobrecarga, **se ha puesto un delay de 1 segundo** entre acceso y acceso. Destacamos que este tiempo se ha aumentado tras comprobar que existían mecanismo de protección ante ataques cuando recibimos el siguiente error: "ConnectionError: HTTPSConnectionPool(host='www.ferreteriasindustriales.es', port=443): Max retries exceeded with url: /decoracion/111050-adhesivo-decoracion-45cmx15m-pvc-bl-adh-brillo-blanco-45-geckofix-8718483100316.html (Caused by NewConnectionError('<urllib3.connection.HTTPSConnection object at 0x000002339F91A580>: Failed to establish a new connection: [WinError 10013] Intento de acceso a un socket no permitido por sus permisos de acceso'"

En un caso real el volumen de descarga sería mayor y lo apropiado es contactar con el proveedor y solicitarle la BD del catálogo en formato digital antes de realizar este proceso como se ha comentado en el capítulo de "*Contexto*". **Un proceso de webscraping se haría tras descartar otras opciones como acceso a una API o incluso que se provea en formato digital los datos**.

7 Inspiración

Explicar por qué puede ser interesante este conjunto de datos y qué preguntas se pretenden responder con ellos. Es necesario comparar con los análisis anteriores o análisis similares presentados en el apartado 6.

Este conjunto de datos puede ser interesante dado que tiene detalles e información útil sobre los productos de nuestro proveedor principal de productos de ferretería, dado que es la empresa proveedora número uno en España.

Respecto al contenido de dataset, es interesante desde el punto de vista didáctico por ser muy completo, con un volumen de datos no muy grande y con características de distinto tipo: clasificación, numéricas y descriptivas.
También es interesante desde el punto de vista comercial porque tiene una gran cantidad de productos de ferretería, pero en ese caso sería necesario una extracción completa de los datos.

Es posible que los científicos de datos necesiten limpiar, reestructurar y procesar dichos conjuntos de datos utilizando herramientas de big data para llegar a conclusiones valiosas tal como:

- La información clave del producto
- Lógica de estructura del catálogo
- Materiales relacionados
- Conectores para tienda online y offline.

Además, el estudio del dataset nos va a permitir categorizar todo tipo de producto según el portafolio de productos ofertados por nuestro proveedor y destacar los productos estrellas en nuestras ofertas.

Finalmente, también debemos indicar que teniendo en cuenta los datos de nuestro proveedor principal, podemos ayudarnos a crear una base de datos o solamente como dataset identificando y consolidando los principales productos y tipo de características de productos más buscadas en las plataformas online para la organización y fácil búsqueda.

Podremos ayudarnos a consolidar una base de datos, por ejemplo, con una estructura más consolidada y relacionada teniendo en cuenta los productos estrellas y con buenas ofertas de nuestro proveedor. A estos productos podremos añadir los productos de otros proveedores con los que hemos decidido trabajar.

Una vez se hayan recopilado y cotejado todos los datos relevantes para nuestra base de datos de catálogos de productos, se puede definir una estructura que tenga sentido para nuestras operaciones diarias, semanales o incluso para cierres mensuales y/o anuales.

Así se puede llegar a organizar nuestro catálogo en función de:

- Familias y subfamilias
- Categorías y subcategorías
- Atributos de producto: Es lo que distingue a cada producto en tu base de datos y permite localizarlo en tiendas online. Aparte de su nombre y precio, se puede incluir los atributos específicos (unidad de medida, fabricante,, referencia,, etc).
- Variantes: Cuando un mismo producto está disponible en distintas variantes de tamaño, color, material, composición, etc.
- Stock: si queremos mantener el número de unidades que tenemos en almacén para así controlar cuando tenemos que hacer nuevos pedidos.

8 Licencia

Seleccionar una licencia adecuada para el dataset resultante y justificar el motivo de su elección. Ejemplos de licencias que pueden considerarse:

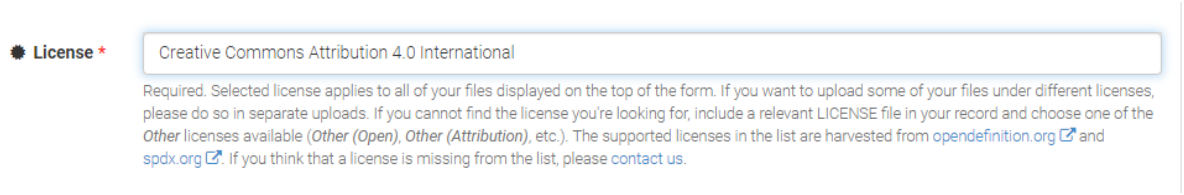
Nuestra primera elección fue usar la Licencia abierta de BBDD llamado **Open DB License**, en él es un acuerdo de licencia copyleft **diseñado para permitir a los usuarios compartir, modificar y usar libremente una base de datos**, manteniendo la misma libertad para los demás. La Licencia Abierta de Bases de Datos fue creada con el fin de permitir a los usuarios compartir sus datos con libertad y sin temor a los derechos de autor o cuestiones de propiedad. Este sistema permite a los usuarios hacer uso libre de los datos contenidos en el repositorio sin temor a la infracción de derechos de autor, y a partir de los datos que han recogido añadirlas a las bases de datos, calculadas o elaboradas por ellos mismos. La licencia establece los derechos de los usuarios que poseen cuando hacen uso de los datos contenidos dentro de la base de datos, así como el procedimiento correcto para la atribución de crédito para la base de datos, y para la presentación de modificación o alteración de datos. Con esto, se consigue comparar y compartir información y datos de forma más fácil. Los usuarios ya no necesitan verificar las

repercusiones de la infracción de derechos de autor o información robada cuando se trabaja bajo la Licencia Abierta de Bases de Datos.

Pero una vez dentro de **Zenodo**, plataforma para publicar proyectos, trabajos y datasets, al subir y compartir el archivo csv de dataset al público, nos hemos encontrado con un pequeño problema ya que **no hemos tenido una opción para seleccionar *Open DB License*** tal cual. En cambio, si hemos tenido una opción muy similar en **Zenodo: *License Creative Commons Attribution 4.0 International***, la cual se define como una licencia en la cual se otorga el crédito adecuado, proporcionando un enlace al detalle legal e indica si se realizaron cambios. **Puede hacer la publicación y compartir el dataset de cualquier manera razonable, pero no de ninguna manera que sugiera que el licenciante lo respalda**.

No hay restricciones adicionales a las descritas: no puede aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros de hacer cualquier cosa que permita la licencia.

Por esta razón, publicaremos nuestro data para que sea un dataset abierto y útil para cualquier análisis sin compromiso alguno ni requerimientos de autor pero si respetando las reglas básicas (no tan estrictas) de la licencia cuando un usuario analista vaya a trabajar sobre estos datos u obtener información útil y así adquirir conocimiento diferenciable.



> **Share** — copy and redistribute the material in any medium or format
> **Adapt** — remix, transform, and build upon the material
> for any purpose, even commercially.
>
> The licensor cannot revoke these freedoms as long as you follow the license terms.

9 Código

Código implementado para la obtención del dataset, preferiblemente en Python o, alternativamente, en R.

El código está disponible en https://github.com/dcanete/PRA1WebScraping. Se ha dado permiso al profesor, no obstante, los fuentes se entregan también con este documento.

Lo más característico del código:

- Se incluye un fichero **Read.me** con introducción al contenido.
- Se han configurado una serie de **constantes para permitir cierta configuración**.
- Parte de esa configuración está enfocada a no saturar la web origen de los datos:
    - **Delay** entre acceso y acceso. Por defecto un segundo.
    - Número **máximo de páginas** por categoría que se van a mirar.
    - Número **máximo de productos** que se van a descargar su detalle.
- Se ha implementado un **control de errores** para algunas instrucciones, ya que hemos detectado problemas con algunos productos que se han controlado.
- Se incluye un log de la aplicación con niveles de debug, info y errores (activado nivel debug)

Destacar las siguientes secciones de los fuentes:

- **Constantes**: donde se puede configurar el algoritmo
- Función **getProduct**: recibe una url de un producto y descarga todo sus propiedades
- Función **save**: recibe una lista de producto y lo guarda en el fichero csv
- **Bucle de categorías**: para cada categoría configurada, recorre todas sus páginas (o hasta llegar a un límite) capturando las urls de los productos encontrados.
- **Bucle de productos**: para cada producto encontrado, se llama el método *getProduct*. Cada 25 productos, se almacenan datos en el fichero csv.

A continuación, se presenta el código completo del algoritmo:

```
from bs4 import BeautifulSoup
from datetime import datetime
import requests
import sys
import time
import csv
import logging
logging.basicConfig(filename='scraper.log',         encoding='utf-8',         level=logging.DEBUG,
format='%(asctime)s %(message)s', datefmt='%m/%d/%Y %I:%M:%S %p')
#logging.basicConfig(stream=sys.stdout, encoding='utf-8', level=logging.DEBUG, format='%(asctime)s
%(message)s', datefmt='%m/%d/%Y %I:%M:%S %p')

############################
# Constantes
############################
base = "https://www.ferreteriasindustriales.es/"
# Número máximo de páginas por categoría (para no saturar la página origen)
max_pages = 15
# Número máximo de productos que se mira el detalle (para no saturar la página origen)
max_products = 10000
# Nombre del fichero de datos
filename = "../dataset/data-" + datetime.now().strftime("%Y%m%d%H%M%S") + ".csv"
# Categorías de productos (tienen páginas distintas)
categorias = ["56-soportes-de-cerrajeria", "174-burletes",
              "14-para-el-hogar", "22-jardin-y-piscinas", "16-ferreteria", "19-fontaneria",
              "20-material-electrico", "25-maquinaria", "15-cerrajeria", "17-climatizacion",
              "176-cantoneras"]
# Segundos de retraso que se pone entre llamada y llamada a pagina de detalle (para no saturar la
página origen)
```

```
delay = 1
products_links = []


#####################################################################
# Método que almacena en un diccionario los datos de un producto
#####################################################################
def getProduct(str):

    #inicializamos
    productData={}

    productData['url'] = str
    productData['title'] = ""
    productData['price'] = ""
    productData['fabricante'] = ""
    productData['sku'] = ""
    productData['Articulo'] = ""
    productData['Marca Comercial'] = ""
    productData['Presentacion'] = ""
    productData['Referencia Proveedor'] = ""
    productData['Diametro'] = ""
    productData['Capacidad'] = ""
    productData['Categoria'] = ""
    productData['Subcategoria'] = ""

    try:
        page = requests.get(str)
        soup = BeautifulSoup(page.content, features="lxml")

    except Exception:
        logging.error ("---- Error al obtener el detalle del producto")
        return productData


    # Título
    # productData['title'] = soup.title.string
    productData['title'] = soup.find("h1", {"itemprop": "name"}).get_text()

    # Precio
    productData['price'] = soup.find("span", {"itemprop": "price"}).get_text()

    # Referencia
    # Hay algunos casos donde no viene el fabricante
    try:
        productData['fabricante']   =   soup.find("div",   {"class":   "product-reference"}).find
("a").get_text()
    except Exception:
        logging.debug ("---- Sin fabricante")

    productData['sku'] = soup.find("span", {"itemprop": "sku"}).get_text()

    # Categorización
    productData["Categoria"]            =            soup.find_all("li",            {"itemprop":
"itemListElement"})[1].find("span").get_text()
    productData["Subcategoria"]           =           soup.find_all("li",           {"itemprop":
"itemListElement"})[2].find("span").get_text()

    # Ficha técnica. Hay que recorrer el HTML para encontrar todas las propiedades de la ficha
técnica
```

```
    # Hay algunos casos donde no viene ficha técnica
    try:
        props = soup.find("dl", {"class": "data-sheet"}).contents
        tam=int((len(props)-1)/4)
        for i in range(0,tam):
            productData[props[i*4+1].string] = props[i*4+3].string
    except Exception:
        logging.debug ("---- Sin ficha técnica")

    # Devuelve el diccionario relleno
    return productData

################################################################
# Método que almacena en fichero los productos que se han capturado
################################################################
def save (products):
    logging.info('-----Almacenamiento-----')
    with open(filename, 'w', newline='') as csvFile:
        writer = csv.writer(csvFile)
        for product in products:
            try:
                writer.writerow(product)
            except Exception:
                logging.error ("---- Problema al guardar: " & str(products))

    products = []
    return products


########################################
# Inicio del proceso
########################################
logging.info('-----Inicio del proceso-----')

# Bucle de categorias
for cat in categorias:
    logging.info('--Buscando productos en ' + cat + '...')

    # Bucle de páginas. Se para cuando se alcance el máximo o no se encuentren más productos
    for i in range(1,max_pages+1):
        url = base + cat +  "?page=" + str(i)

        page = requests.get(url)
        soup = BeautifulSoup(page.content, features="lxml")
        tam = len (soup.find_all('h5', {"class": "product-title"}))
        logging.debug ("----" + str(tam) + " productos encontrados")

        if (tam)==0:
            logging.info('----No se encuentran más productos. Páginas revisadas: ' + str(i))
            break

        for link in soup.find_all('h5', {"class": "product-title"}):
              aLink = link.find("a",href=True).get('href')
              products_links.append(aLink)

        # Para no saturar la web origen
        time.sleep (delay)

# Quita enlaces duplicados
products_links = list(dict.fromkeys(products_links))
```

```
# Reduce la lista para no saturar web origen
products_links = products_links[0:max_products]

# Inicializa la lista de productos
item=["url","title","price","fabricante","sku","Articulo", "Marca Comercial", "Presentacion" ,
      "Referencia Proveedor", "Diametro", "Capacidad", "Categoria", "Subcategoria"]
products = []
products.append(item)


# Recorremos todos los productos viendo el detalle
i=1
size = len (products_links)
logging.info("-----Procesando " + str(size) + " productos-----")
for url in products_links:
    productData = getProduct(url)
    logging.debug ("--Procesado " + productData["title"] + "(" + str(i) + "/" + str(size) + ")")
    i=i+1
    item = [productData ["url"],productData ["title"], productData ["price"],
            productData ["fabricante"], productData ["sku"], productData ["Articulo"],
            productData ["Marca Comercial"], productData ["Presentacion"] ,
            productData ["Referencia Proveedor"], productData ["Diametro"],
            productData ["Capacidad"], productData ["Categoria"], productData ["Subcategoria"]]
    products.append(item)

    if (i%25==0):
        save (products)


    # Para no saturar el servidor origen
    time.sleep (delay)

# Por último, almacenamos el contenido de la lista en CSV
logging.info('-----Almacenamiento-----')
with open(filename, 'w', newline='') as csvFile:
    writer = csv.writer(csvFile)
    for product in products:
        writer.writerow(product)

logging.info('-----Fin del proceso-----\n\n')
```

10 Dataset.

*Publicar el dataset obtenido en formato CSV en Zenodo, incluyendo una breve descripción de la misma. Obtener y adjuntar el enlace del DOI del dataset (https://doi.org/...). El dataset también deberá incluirse en la carpeta **/dataset** del repositorio.*

Se ha publicado en zenodo incluyendo detalles prácticos para los usuarios como la descripción de los campos, tipo de datos, etc.

Dataset "Productos de Ferreteria" de la empresa Ferreterias Industriales

Kaur, Balpreet; Cañete Roman, Dani

Presentamos el dataset **"Productos de Ferreteria"** de la empresa Ferreterias Industriales, en el cual podemos encontar los siguientes attributos y detalles de los productos ofrecidos por la empresa online Ferreteria Industriales, la empresa número uno como comercializadora y especializada en productos de ferreteria, fontaneria, mobiliario de jardin, riego, cerrajeria y en maquinas agricola.

En el dataset podemos encontrar los diferentes tipos de productos junto con sus caracteristicas principales incluyendo precio, tamaño, fabricante, categoria, entre otros.

Se capturan la información más relevante que se compone de los siguientes campos:

- url: Enlace. Dirección web donde está la información del producto
- title: Cadena. Nombre del producto
- price: Importe. Precio del proveedor para el producto. Es importante tener en cuenta que es en euros y con impuestos incluidos. Al ser un proveedor, el IVA se podrá repercutir.
- fabricante: Lista de valores. Identifica al fabricante del producto.
- sku: Numérico. Referencia del producto. Es importante guardarlo porque será necesario utilizarlo cuando realicemos pedidos a este proveedor.
- Articulo: Cadena. Nombre alternativo del producto que se facilita en la ficha técnica
- Presentacion: Cadena. Unidad de medida (opcional)
- Referencia Proveedor: Numérico. Otra referencia del producto. No se tiene claro en este punto cuál de las dos referencias se utilizará al hacer el pedido. Ante la duda, se almacenan los dos datos (opcional)
- Diametro: Numérico. Diámetro del producto (opcional)
- Capacidad: Numérico. Volumen de almacenamiento del producto (opcional)
- Categoria: Lista de valores. Categoría en la que se ha clasificado el producto
- Subcategoria: Lista de valores. Subcategoría en la que se ha clasificado el producto

Si se considera interesante en futuras fases, se podrían normalizar fácilmente los datos sacando a tablas separadas valores posibles de Categoría, Subcategoría y fabricante.

Este dataset está disponible para el público sin restricciones alguno, bajo la licencia **License Creative Commons Attribution 4.0 International.**

Este dataset es el resultado de un caso práctico de analisis de datos llevado a cabo y trabajado durante la duración de la asignatura de "Tipologia y el ciclo de vida de datos" del master universitario Data Science de la universidad UOC.

7 views
6 downloads
See more details...

Indexed in OpenAIRE

Publication date:
April 19, 2023

DOI:
DOI 10.5281/zenodo.7845032

Keyword(s):
Dataset "Productos de Ferreteria" | Ferreterias Industriales | Ferreteria | Herramientas | Fontaneria | Mobiliario de jardin | Riego | Cerrajeria | Máquinas Agricola

License (for files):
Creative Commons Attribution 4.0 International

Versions

Version 1 — Apr 19, 2023
10.5281/zenodo.7845032

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.7845031. This DOI represents all versions, and will always resolve to the latest one. Read more.

Share

Está publicado en https://doi.org/10.5281/zenodo.7845032 con el siguiente Badge:

11 Vídeo

*Realizar un breve vídeo explicativo de la práctica **(máximo 10 minutos)**, que deberá contar con la participación de los dos integrantes del grupo. En el vídeo se deberá realizar una presentación del proyecto, destacando los puntos más relevantes, tanto de las respuestas a los apartados como del código utilizado para extraer los datos. Indicar el enlace del vídeo (https://drive.google.com/…), que deberá ubicarse en el Google Drive de la UOC.*

Se incluye en el drive de la entrega (https://drive.google.com/drive/folders/1ZuaHJcx7jHUO-vj2e9O_04gtjCoAKPur?usp=sharing), en la carpeta “Video”

12 Referencias

A continuación, se incluye un resumen de urls que componen la entrega para facilitar la localización de los ficheros:

- Fuentes: https://github.com/dcanete/PRA1WebScraping
- Dataset: https://zenodo.org/record/7845032#.ZEAXxnZBzD5
- Documento: https://docs.google.com/document/d/1Ncy_6mQzNkXpTf-ictA9TCWfF0EHF_PTpHo_CETJyHc
- Presentación: https://docs.google.com/presentation/d/1UiUI2LxnSkkTcxdFiM9X0URV7yG3hdo40sRMTQC0Nio
- Carpeta drive: aunque solo se pedía el video en Drive, se ha incluido además documentos, dataset y código en una carpeta: https://drive.google.com/drive/folders/1ZuaHJcx7jHUO-vj2e9O_04gtjCoAKPur?usp=sharing

13 Tabla de contribuciones

Contribución	Firma
Investigación previa	DCR, BKS
Redacción de las respuestas	DCR, BKS
Desarrollo del código	DCR, BKS
Participación en el vídeo	DCR, BKS