

Práctica 1 - Tipología y ciclo de vida de los datos

Francisco Javier Morales Hidalgo y Antonio
Martín Espinosa

- 1. Contexto. Explicar en qué contexto se ha recolectado la información. Explicar por qué el sitio web elegido proporciona dicha información.**
Se ha recogido la información de los videojuegos actualmente disponibles en el catálogo de la cadena *Mediamarkt* para las diferentes plataformas existentes. La web de la que hemos obtenido los datos '<https://www.mediamarkt.es>' sirve tanto para informar del contenido de su catálogo como para realizar compras online.
- 2. Título. Definir un título que sea descriptivo para el dataset.**
Mediamarkt Web Scraper.
- 3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído. Es necesario que esta descripción tenga sentido con el título elegido.**
Colección del fabricante, género, precio, tipo de dispositivo, título, clasificación por edades y plataforma de todos los títulos de las diferentes plataformas disponibles en la web de la cadena *Mediamarkt*.
- 4. Representación gráfica. Dibujar un esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.**
- 5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se han recogido.**
El dataset contiene los campos del fabricante, género, precio, tipo de dispositivo, título, clasificación por edades y plataforma de todos los videojuegos disponibles en el momento de la ejecución del script (.../03/2022) y se han recogido mediante web scraping de la url
'<https://www.mediamarkt.es/es/category/juegos-ps5-8533.html?page=>'.
- 6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares. Justificar qué pasos se han seguido para actuar de acuerdo a los principios éticos y legales en el contexto del proyecto.**
- 7. Inspiración. Explicar por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.**
- 8. Licencia. Seleccionar una de estas licencias para el dataset resultante y justificar el motivo de su selección:**

Elegimos la licencia Open Database, cubriendo los contenidos con la licencia Database Contents License. Esta decisión se basa en la flexibilidad que ofrece la licencia para modificar los derechos de uso entre unos componentes de nuestro proyecto y otros.

- 9. Código. Adjuntar en el repositorio Git el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.**
<https://github.com/uoctcvd-martin-morales/practica1>
- 10. Dataset. Publicar el dataset obtenido(*) en formato CSV en Zenodo con una breve descripción. Obtener y adjuntar el enlace del DOI.**
- 11. Vídeo. Se debe hacer entrega de un vídeo explicativo de la práctica en donde cada uno de los integrantes del grupo explique con sus propias palabras tanto las respuestas del proyecto como el código utilizado para llevar a cabo la extracción. El vídeo debe ser enviado a través de un enlace a Google Drive que deben proporcionar, junto con el enlace al repositorio Git, al momento de entregar la práctica.**