

PRÁCTICA 1

1. **Contexto.** Explicar en qué contexto se ha recolectado la información. Explicar por qué el sitio web elegido proporciona dicha información.

Se ha buscado una página web de juegos de mesa <https://zacatrus.es/juegos-de-mesa.html> para poder recopilar información de esto y saber los precios de mercado.

2. **Título.** Definir un título que sea descriptivo para el dataset.

El título del dataset va a ser juegos.csv, pues va a contener información de juegos de mesa.

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.

El dataset va a contener información de: nombre del juego, precio, números de comentarios y rating de este. Además, se guardan las imágenes de cada juego en una carpeta.

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 ha recogido.

Los datos son recogidos de la página web, obteniendo el contenido html y guardando, evitando realizar más peticiones y por tanto evitando bloqueos. Se procede a crear una cabecera http para falsear el user-agent.

Una vez obtenida y guardada la información de la web, se vuelve a leer el fichero html y se obtiene información de cada juego de mesa.

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.

Se ha procedido a leer el fichero robots.txt, verificando que se podía extraer el contenido y se ha comprobado los términos legales de la página web para comprobar que se podía extraer esta información siempre que no sea con fines de comercialización o distribución a terceros.

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.

Este conjunto de datos puede ser interesante para tener una orientación de los precios y tipos de juegos de mesa que hay en el mercado.

8. **Licencia.** Seleccionar una de estas licencias para el dataset resultante y justificar el motivo de su selección:

La licencia es CC BY-NC-SA 4.0 License. Esto se debe a:

- Reconocimiento: Debe reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios. Puede hacerlo de cualquier manera razonable, pero no de una manera que sugiera que tiene el apoyo del licenciador o lo recibe por el uso que hace.
- No Comercial: No puede utilizar el material para una finalidad comercial.
- Compartir Igual: Si remezcla, transforma o crea a partir del material, deberá difundir sus contribuciones bajo la misma licencia que el original.

9. **Código.** Adjuntar en el repositorio Git el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

Tanto el código fuente como extraer datos como el dataset generado se encuentran en el este repositorio.

10. **Dataset.** Publicar el dataset obtenido(*) en formato CSV en Zenodo con una breve descripción. Obtener y adjuntar el enlace del DOI.

(*) Si existe algún impedimento para publicar el dataset real, se deberá justificar esta situación y realizar y publicar en Zenodo un dataset simulado. En este caso, el dataset real se comunicará al profesor de forma privada (p.ej., enlace de Google Drive).