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

La información que queremos obtener corresponde a las 250 mejores películas de todos los tiempos según la crítica de la web IMDB. Esta información puede sernos útil para añadirlas y mantenerlas actualizadas en la base de datos. Así como poder mantenerla actualizada cada cierto tiempo lanzando nuestro procedimiento con determinada frecuencia.
IMDB es una de las webs con más información de películas de todo el mundo, y con una amplia crítica y valoraciones que podemos aprovecharla para nuestro propósito.

2. Definir un título para el dataset. Elegir un título que sea descriptivo.

Scraping IMBD: Información y valoración de las mejores películas

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).

Recabaremos la información detallada de cada película, tanto información relacionada con detalles de la propia película como la valoración de los críticos de cine. Y de esta manera poder aprovecharla en lo que nos pueda ser necesario, manteniéndola actualizada frecuentemente.

4. Representación gráfica. Presentar esquema o diagrama que identifique el dataset visualmente y el proyecto elegido

Imagen que podemos identificar con la informadición que vamos a obtener:

5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

El dataset contiene los siguientes campos:

- Nombre: nombre de la película en español
- Ranking: posición en el ranking de las 250 mejores películas de todos los tiempos.
- Puntuación IMDB: media de toda las puntuaciones de los críticos .
- Año: año de la película.

- URL Imagen: dirección web de la imagen miniatura de la película.
- Título Original: título original de la película.
- Duración: duración de la película en horas y minutos.
- Fecha Estreno (Pais): fecha de estreno de la película y país de estreno.
- Géneros: 1 o más géneros en los que se puede catalogar la película.
- URL Poster: dirección web del póster de la película.
- Nombre imagen: nombre y extensión de la imagen almacenada en disco.
- Resumen: breve resumen de la película.
- Director: director o directores de la película.
- Escritores: escritor o escritores de la película.
- Actores: actor o actores de la película.

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.

El propietario del conjunto de datos obtenido es IMDb and STARMETER, en la web no aparecen citas a análisis de datos anteriores o realizados por ellos.
Pero a través de artículos de internet como:
https://www.datacentric.es/blog/insight/exito-netflix-datos/
se puede observar que es algo que se realiza y se aprovecha de dicho análisis de datos para, junto con otras técnicas de big data, obtener distintas ventajas en un negocio que está cada vez más en auge.

7. Inspiración. Explique 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 dataset nos puede ser útil en caso de querer maneja información de películas determinadas, podrías aprovecharla para que plataformas de streaming como netflix, hbo, … la utilicen para generar valor en sus aplicaciones y mejorar tanto la información de las películas como tener un mejor y más interesante repertorio.

8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

La licencia que mejor se adaptaría al trabajo obtenido sería la de Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0). Se ha elegido esta porque creo que habría que limitar el aprovechamiento comercial de la información obtenida, aunque si se permita aprovecharla para otras aplicaciones que no sean comerciales.

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

En la dirección https://github.com/davidarbona/imdbscraping/tree/main/src se puede obtener el código fuente en Python con el que he obtenido el dataset.

10. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

El dataset obtenido con la ejecución del código fuente referenciado anteriormente se puede encontrar en el DOI siguiente:

https://doi.org/10.5281/zenodo.4671026