

# Práctica 1 - Films to Watch Scraper

Autor: Hugo Mourisco Quirós

noviembre 2021

---

## Contexto

---

En esta práctica se desea tener un primer acercamiento a la construcción de datasets sobre películas que permitan realizar consultas rápidas y aplicar diferentes modelos predictivos o de análisis. El dataset de esta práctica en particular se construye a partir de la extracción de datos de la plataforma The Movie Database. The Movie Database es una plataforma que tiene como objetivo ofrecer a los amantes del cine un lugar en el que consultar información relacionada con las películas.

---

## Título

---

El título que se asigna a esta práctica es Films to Watch Scraper. El objetivo es poder construir un dataset que recoga información de películas seleccionadas a partir de dos parámetros: el género e idioma original.

---

## Descripción del dataset

---

El dataset recoge registros de películas según el género e idioma seleccionado. TODO;

---

## Representación gráfica

---

TODO;

---

## Contenido

---

El dataset se compone de registros que almacenan información de cada una de las películas que se extraen de The Movie Database. Se recogen las siguientes características:

- **Título:** nombre de la película en inglés.
  - **Género:** el tema general de una película que sirve para su clasificación.
  - **Idioma original:** el idioma original en el que fue filmada la película.
  - **Año de estreno:** fecha de estreno de la película.
  - **Duración:** duración en segundos de la película.
  - **Director:** director principal de la película.
  - **Actores principales:** listado de los actores principales que participan en la película.
  - **Palabras clave:** palabras clave que sirven para realizar clasificaciones de la película.
  - **Ranking:** puntuación de la película en base a la valoración de los usuarios.
- 

## Agradecimientos

---

Los datos han sido recolectados de la plataforma The Movie Database. El acceso y la extracción de los datos de las películas de dicha plataforma se ha realizado gracias al lenguaje de programación Python, junto a la librería BeautifulSoup, la cual facilita realizar técnicas de web scrapping sobre las páginas HTML de la plataforma.

---

## Inspiración

---

TODO;

---

## Licencia

---

Para esta práctica se utiliza una licencia Creative Commons (CC) debido a la extracción de información de una plataforma protegida por derechos de autor en la que al hacer web scrapping en lugar de utilizar su propia API no solicitamos permiso directamente a la plataforma. La versión asignada es la 4.0, que incluye la cobertura de bases de datos. Y la condición principal es que se pueda utilizar el dataset para fines no comerciales, por lo tanto la licencia seleccionada sería **Released Under CC BY-NC-SA 4.0 License**.

---

## Código

---

El código, así como los diferentes recursos necesarios para esta práctica se encuentran accesibles en el siguiente repositorio:

<https://github.com/hoguku/filmsToWatchScraper>

---

## Dataset

---

El dataset obtenido se encuentra en formato CSV en Zenodo:

<https://zenodo.org/api/files/213ebb8b-b830-4484-bf7f-12ce7a996016/filmsToWatch.csv>