

# DATASET: Películas en cartelera en Reino Unido

Alberto Lucas Navío

24 de Octubre de 2019

## Lista de películas que están actualmente en cartelera en los cines del Reino Unido

### Descripción

Este conjunto de datos se ha extraído del sitio web Film Affinity. Éste, es un sitio web sobre cine que almacena información sobre cualquier obra audiovisual realizada a lo largo de la historia del cine. Una sección del sitio web está dedicada a las películas que están en cartelera en la actualidad. Es la información de esta sección la que queremos extraer mediante nuestro web scraper. Así mismo, este sitio permite que sus usuarios valoren las películas con una nota del 1 al 10. De esta forma, cada película tiene una nota, que es la media de las valoraciones de los usuarios.

Teniendo esto en cuenta, mediante web scraping vamos a obtener un listado de las películas que se está emitiendo en el momento en que se ejecute en los cines del Reino Unido.

Deseamos obtener el nombre de las mismas, así como información de utilidad para alguien que intente decir qué película va a ver en el cine, como son: director, reparto, género de la película o la valoración de los usuarios. Además, se extraerá un enlace por si el usuario desea consultar información adicional de la película.

Junto a esta información, vamos a descargar el póster de cada una de las películas que están en cartelera. Estas imágenes se almacenarán en la misma carpeta desde la que se ejecute el scraper.

Como se ha mencionado anteriormente, el sitio web del que se van a extraer los datos es FilmAffinity.com, en concreto su versión del Reino Unido. Para extraer este conjunto de datos, se va a hacer uso del lenguaje de programación Python y de técnicas de Web Scraping.

## Imagen identificativa



## Contexto

Como se ha comentado, la temática de este conjunto de datos es el cine de actualidad. Recoge películas que se están emitiendo en el momento en que se lance el scraper en cines del Reino Unido. Film Affinity dispone de una base de datos con infinidad de películas, así como una enorme cantidad de información acerca de prácticamente todas las películas de la historia del cine.

Además, las películas están organizadas clasificadas en función a distintas variables, como son su género, el año en que fueron estrenadas, o su valoración. En este caso, nos interesa la sección en la que aparecen los estrenos más recientes.

## Contenido

Para cada película, se recogen una serie de características. Éstas serán exportadas a un fichero csv. En el fichero, se incluirán las siguientes características:

- **Name:** nombre de la película
- **Director:** director de la película
- **Cast:** reparto de la película
- **Genre:** género de la película
- **Rating:** Nota de la película. Es la media de la nota otorgada por todos los usuarios que hayan votado esta obra en [www.filmaffinity.com](http://www.filmaffinity.com)
- **Link:** enlace que conduce hasta la página de Film Affinity acerca de esa película en concreto, donde se puede consultar toda la información sobre la misma, además de reviews y estadísticas sobre la película.

Cabe destacar que en el fichero, las películas aparecen ordenadas según la valoración de los usuarios.

Además de esto, el scraper incluye una función que el usuario puede ejecutar si lo desea para descargar una imagen de cada una de las películas que están en cartelera. Así pues, si sólo desea leer la información acerca de la película puede acceder al fichero csv y leerla, y si además desea ver el póster de cada película puede descargar las imágenes correspondientes del sitio web.

## Inspiración

Este conjunto de datos puede ser utilizado en diversos ámbitos. Cualquier persona que estén intentando decidir qué película ver en el cine puede acceder de forma rápida al listado de opciones, junto con información que puede ayudarle a tomar una decisión. De la misma manera, una distribuidora puede utilizar este conjunto de datos para tener una idea de qué películas han tenido una mejor acogida, y empezar a planificar su lanzamiento en formato físico.

Así mismo, si se realiza el proceso de forma periódica, durante un periodo de tiempo largo, es posible realizar análisis interesantes para las productoras de cine. Por ejemplo, si la presencia de ciertos actores condiciona como es la película valorada por parte de los usuarios, si las películas de ciertos directores están mejor valoradas que las dirigidas por otros, etc.

## Agradecimientos

Está claro que la información sobre qué películas están en cartelera en los cines es pública y de hecho son los propios cines los que se benefician de que esta información se reproduzca.

Sin embargo, Film Affinity recoge esta información, la organiza y la estructura, además de ofrecer una valoración mediante las votaciones de sus usuarios.

Hay que agradecer a FilmAffinity y a sus usuarios la labor que hacen en torno a la recopilación y organización de información relativa al cine, no solo a las películas que están en cartelera, sino de toda la historia del cine.

## Licencia

La licencia escogida para la publicación de este conjunto de datos ha sido **\*\*CC BY-SA 4.0 License\*\***. Los motivos que han llevado a la elección de esta licencia tienen que ver con la idoneidad de las cláusulas que esta presenta en relación con el trabajo realizado:

- \* Se debe proveer el nombre del creador del conjunto de datos generado, indicando los cambios que se han realizado. De esta manera, se reconoce el trabajo ajeno y en qué medida se han realizado aportaciones en relación con el trabajo original.
- \* Se permite un uso comercial. Esto haría que incrementen las probabilidades de que una empresa utilice los datos generados y realicen trabajos de calidad que reporten cierto reconocimiento al autor original.
- \* Las contribuciones realizadas a posteriori sobre el trabajo publicado bajo esta licencia deberán distribuirse bajo la misma. Esto hace que el trabajo del autor original continúe distribuyéndose bajo los términos que él mismo planteó.

## Recursos

1. Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data
2. Mitchel, R. (2015). Web Scraping with Python: Collecting Data from the Modern Web. O'Reilly Media, Inc. Chapter 1. Your First Web Scraper.
3. Subirats, L., Calvo, M. (2019). Web Scraping. Editorial UOC

