

--

**title:** "Dataset: Lista de películas en los cines de Reino Unido"

**author:** Alberto Lucas Navío

**date:** 24/10/19

## ## Título del dataset

Películas en cartelera en Reino Unido

## ## Subtítulo del dataset

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

## ## Imagen

Imagen mostrando el póster de varias de las películas que están en cartelera.

## ## Descripción

Este conjunto de datos se ha extraído de la página web [www.filmaffinity.com](http://www.filmaffinity.com). Este sitio web es sobre cine y tiene una sección dedicada a las películas que se encuentran en cartelera en este momento. Así mismo, esta página permite que sus usuarios valoren cada película, de forma que podemos ver una nota media para cada película.

En este caso, se han extraído los datos de esta sección para obtener un listado de las películas que se estén emitiendo en los cines del Reino Unido además de varias características de interés de estas películas. En concreto, este conjunto de datos está formado por una lista de las películas que están actualmente en cartelera en el Reino Unido junto con varias características de las mismas, como su director, su reparto, el género y una nota.

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.

Por otra parte, se ha creado un segundo web scraper, por si un usuario desea realizar una búsqueda en la base de datos de Film Affinity. En este caso, el scraper, utilizará un término de búsqueda introducido por el usuario, llevará a

cabo una búsqueda en Film Affinity y producirá un fichero csv con información acerca de cada uno de los resultados de la búsqueda.

Como se ha mencionado anteriormente, los datos se han obtenido del sitio web [www.filmaffinity.com](http://www.filmaffinity.com). Para obtener este conjunto de datos, se ha hecho uso del lenguaje de programación Python y de técnicas de Web Scraping.

## ## Contenido

Para cada película, se recogen las siguientes características:

- \* **Director**: director de la película.
- \* **Cast**: reparto de la película.
- \* **Genre**: género de la película.
- \* **Rating**: media de la nota valoración de la película por parte de los usuarios que la han votado en la página [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.

Las películas aparecen ordenadas según las valoraciones de los usuarios de Film Affinity.

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. En el segundo fichero, el contenido será muy similar. Es un archivo que contiene los resultados de la búsqueda realizada por el usuario. Para cada resultado se incluye:

- **Name**: nombre 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)
- **Director**: director de la película
- **Cast**: reparto de la película

## ## Inspiración

Este conjunto de datos puede ser utilizado por cualquier persona que tenga interés en el cine de actualidad en diversos ámbitos. Ya sea por parte de un usuario que quiere decidir qué película ver en el cine, bien por parte de un

crítico que quiere saber cuáles son las películas mejor valoradas en el panorama actual.

## ## 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