

--

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

author: Alberto Lucas Navío

date: 24/10/19

Descripción

Este conjunto de datos se ha extraído de la página web 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.

Como se ha mencionado anteriormente, los datos se han obtenido del sitio web 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

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.

Si extraemos los datos y producimos un archivo de forma periódica durante un periodo de tiempo largo, se pueden realizar análisis interesantes para productores de cine. Por ejemplo, si las películas de ciertos directores están

mejor valoradas que, si influye la aparición de ciertos actores en el reparto, etc.

Sigüientes pasos

1. Extraer imágenes de la página web
2. Incluir un enlace al trailer de la película en el csv generado.
3. Gestionar licencias.

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.

