

## Descripción de la Práctica

Componentes del grupo: Alberto Cuevas Gonzáles y Carlos Salas Kufliovski

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

Se ha recolectado información sobre las películas mejor valoradas de diferentes géneros. El análisis de las películas se ha realizado a nivel mundial, contando las producciones de todos los países, y de todos los años.

Para ello, se ha recurrido a la página de FilmAffinity donde obtenemos datos sobre la valoración de las películas e información sobre las mismas, como el año de estreno, director(es), el reparto y el género.

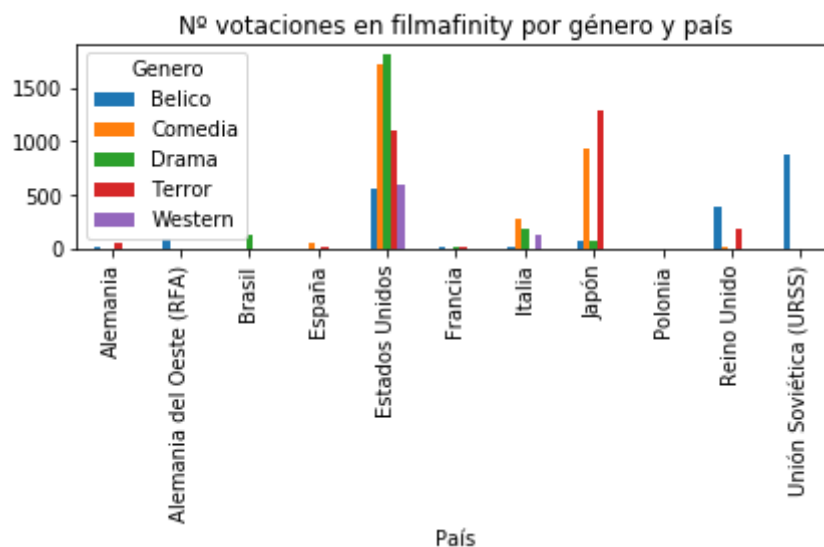
### 2. Definir un título para el dataset. Elegir un título que sea descriptivo.

El título del csv es Peliculas.csv porque la descarga de datos en el archivo está relacionada a las películas en su conjunto, sin especial tratamiento de datos posterior.

### 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).

El dataset contiene las 30 películas más valoradas en la página web FilmAffinity en los géneros de bélico, comedia, drama, horror y western.

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



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

El dataset incluye los siguientes campos:

- Genero
- Nombre de la película
- Año estreno
- País
- Valoración FA
- Director(es)
- Cantidad de votos
- Reparto

Los datos fueron extraídos el 10 de abril de 2021.

El método de recolección fue a través de web scraping, utilizando el lenguaje de programación Python y el método de BeautifulSoup.

**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.**

Los propietarios del conjunto de datos son Alberto Cuevas González y Carlos Salas Kuflijevski.

Se han realizado análisis similares con otras páginas web, por ejemplo, con la web [IMDB](#)

**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.**

El conjunto de datos es interesante tanto para aquellas personas entusiastas del cine, como para obtener información útil a partir de los datos recogidos.

Por ejemplo, con el conjunto de datos se pueden responder en qué géneros tienen mejores votaciones en estrenos más recientes, y en qué géneros las películas clásicas permanecen imbatibles por los nuevos estrenos.

También se pueden realizar cruces de qué actores obtienen mejores puntuaciones en qué géneros, de manera que en el futuro podrían especializarse en el tipo de genero en el que sobresalen más.

En comparación con el análisis de IMDB presentado en la respuesta 6., podemos ver que nuestros datos obtenidos tienen mucha mayor profundidad, obteniendo los géneros de las películas, directores, reparto y país.

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

Released Under CC0: Public Domain License

Released Under CC BY-NC-SA 4.0 License

Released Under CC BY-SA 4.0 License

Database released under Open Database License, individual contents under Database Contents License

Other (specified above)

Unknown License

La licencia bajo la que distribuimos nuestro código es la “Released Under CC BY-NC-SA 4.0 License”. El motivo de elegir esta licencia es porque queremos permitir que el código se comparta, esto es, copiar y redistribuir el material en cualquier medio o formato; y que se pueda adaptar, remezclar, transformar y crear a partir del material.

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

El código se puede obtener en este [enlace](#) a GitHub.

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

El dataset se puede obtener en este [enlace](#) a GitHub.

Contribuciones	Firma
Investigación previa	A.C., C.S.
Redacción de las respuestas	A.C., C.S.
Desarrollo código	A.C., C.S.