**Documento de descripción del dataset**

|  |  |  |  |
| --- | --- | --- | --- |
| DATOS DEL ALUMNO | | | |
| Nombre y apellidos | FERNANDO JOSE MATEOS GOMEZ | | |
| Titulación | ⬜ IC ⬛ IS ⬜ TI ⬜ DG | | |
| Grupo | ⬜ 1 ⬜ 2 ⬜ 3 ⬛ 4 ⬜ 5 | Subgrupo de laboratorio | ⬜ 1 ⬛ 2 ⬜ 3 ⬜ 4 |

|  |  |
| --- | --- |
| INFORMACIÓN DEL DATASET | |
| Nombre | IMDB 5000 Movie Dataset |
| URL | https://www.kaggle.com/carolzhangdc/imdb-5000-movie-dataset |
| ¿El dataset contiene varios ficheros csv? ⬜ Sí ⬛ No  En caso de responder sí, indique el que ha elegido: | |
| Listado con datos de películas en los últimos años, indicando el formato de color, el director, el actor principal, su idioma y el dinero gastado, entre otros.  El formato usado es una versión simplificada con 15 columnas de datos en vez de 28 como el original | |

|  |  |
| --- | --- |
| DESCRIPCIÓN DE LAS COLUMNAS | |
| Nº de colummas | 15 |
| color | Descripción: Formato de color de la película  Tipo de dato: str |
| director\_name | Descripción: Nombre director  Tipo de dato: str |
| num\_critic\_for\_reviews | Descripción: Número de crítica de la audiencia  Tipo de dato: int |
| duration | Descripción: Duración en minutos de cada película  Tipo de dato: int |
| genres | Descripción: Conjunto con los géneros que abarca la película  Tipo de dato: list |
| actor\_1\_name | Descripción: Nombre del actor protagonista  Tipo de dato: str |
| movie\_title | Descripción: Título de la película  Tipo de dato: str |
| num\_user\_for\_reviews | Descripción: Numero de reseñas por los videntes  Tipo de dato: int |
| language | Descripción: Idioma de la película  Tipo de dato: str |
| country | Descripción: País donde se rodó la película  Tipo de dato: str |
| content\_rating | Descripción: Calificación para la audiencia  Tipo de dato: str |
| budget | Descripción: Presupuesto, en dólares  Tipo de dato: int |
| title\_year | Descripción: Año de lanzamiento  Tipo de dato: int |
| imdb\_score | Descripción: Calificación por IMDB  Tipo de dato: float |
| aspect\_ratio | Descripción: Relación de aspecto  Tipo de dato: float |

|  |  |
| --- | --- |
| DESCRIPCIÓN DE LOS METODOS | |
| read\_file | *Input: (Rute) Ruta relativa con el archivo csv deseado*    *Ouput: Movies(str,str,int,int,list,str,str,,int,str,str,str,int,int,float,float)* |
| Filter\_by\_genres | *Input: (movies,genres) Dataset con las peliculas de tipo Movies, y una lista con generos*  *Ouput: Listado con todas las peliculas que son de algún genero de la lista genres* |
| get\_genres\_by\_year | *Input: (movies,year) Dataset de tipo Movies, y un año de busqueda, por defecto vale 2009*  *Output: Listado con todos los generos de todas las peliculas, se repiten, que se encuentran en el año seleccionado* |
| count\_genres\_by\_year | *Input: (movies,year) Dataset de tipo Movies, y un año de búsqueda*  *Output: Devuelve una lista de tuplas con todas las veces que se repite un género determinado en un año*Tipo de dato: int |
| average\_users\_by\_rating | *Input: (movies,rate) Dataset de tipo Movies, y marca de la película, con edad recomendada para ver la película, por defecto 'PG'*  *Output: Devuelve una media correspondiente al número de críticas por los usuarios para las películas con edad recomendada* |
| get\_movies\_ordered\_by\_title | *Input: (movies,order,min\_of\_ocurrences) Dataset de tipo Movies, orden en que se ordena el dataset (Por defecto de forma ascendente), minimo de elementos que se retornarán, por defecto 10*    *Outut: Devuelve un listado de películas* ordenadas *a antojo del usuario, en función del titulo de la película. Solo se mostrarán tantas películas como*  *películas haya o las enviadas por parámetro* |
| get\_movies\_max\_budget | *Input: (movies) Dataset de tipo Movies*  *Outut: Devuelve un listado con las películas con máximo presupuesto* |
| get\_number\_films\_lang\_color | *Input: (movies,lang,color) Dataset de tipo Movies, lenguaje str, color de la película str*  *Outut: Devuelve un número con la cantidad de películas que hablan en el idioma seleccionado y que coincidan con el campo color* |
| list\_of\_countries\_speaklang | *Input: (movies,lang) Dataset de tipo Movies, y un lenguaje str*  *Outut: Lista con los paises que hablan el idioma seleccionado* |
| max\_min\_of\_film\_by\_duration | Input: (movies) Dataset de tipo Movies  Outut: Una tupla con la duracion mayor y menor del dataset |
| counter\_films\_of\_color | Input: (movies) Dataset de tipo Movies  Outut: Un diccionario que devuelve el numero de peliculas en función del color en que esté |
| get\_longest\_shorter\_by\_country | Input: (movies) Dataset de tipo Movies  Outut: Un diccionario con el que se devuelve la pelicula más larga y más corta en función del pais |
| ar\_med\_num\_critic\_by\_director | Input: (movies) Dataset de tipo Movies  Outut: Un diccionario de tuplas con la media por la critica en función del director y el valor total de la critica |
| get\_films\_by\_content\_rating | Input: (movies) Dataset de tipo Movies  Outut: Un diccionario que devuelve las peliculas por rating del vidente |
| get\_user\_rate\_films | *Input: (movies,n,select) Dataset de tipo Movies, un numero entero que indica cuantas peliculas quiero ver, filtro que indica si quiero a las más vistas o las menos o ambas*  *Outut: Un diccionario que devuelve las n peliculas mejores y peores valoradas*  *Error: Devuelve None en caso de que n sea un numero menor o igual que 0* |