



FUNDAMENTOS DE PROGRAMACIÓN

Proyecto de cuatrimestre: Películas

Autor: Lidia García Barragán

En este proyecto trabajaremos con datos extraídos de Kaggle (www.kaggle.com) correspondientes a las películas estrenadas entre los años de 2007 y 2011. Los registros con los que obraremos contienen diversas informaciones de las películas que nos permitirán elaborar una serie de funciones útiles para explorar los datos de los que disponemos.

Trabajaremos con un fichero en formato CSV llamado *Dataset Proyecto FP.csv* ubicado en la carpeta “data” del proyecto. Cada registro del fichero de entrada ocupa una línea y contiene siete informaciones separadas por puntos y comas sobre las películas: el título, el género, las valoraciones de Rotten Tomatoes, las valoraciones de la audiencia, las ganancias en millones de dólares, el año de estreno y la fecha de estreno. Abajo podemos ver en la Ilustración 1 una foto de algunas líneas del fichero CSV.

	A	B	C	D	E	F
11	50/50;	Comedy;	93;93;8;	2011;11/12/2011		
12	88 Minutes;	Drama;	5;51;30;	2007;22/01/2007		
13	A Dangerous Method;	Drama;	79;89;20;	2011;29/03/2011		
14	A Nightmare on Elm Street;	Horror;	13;40;35;	2010;02/01/2010		
15	A Serious Man ;	Drama;	89;64;7;	2009;11/02/2009		
16	A Very Harold and Kumar Christmas;	Comedy;	72;71;19;	2011;16/11/2011		
17	Abduction;	Action;	4;46;35;	2011;28/01/2011		
18	Across the Universe;	Romance;	54;84;45;	2007;23/05/2007		
19	Adventureland ;	Comedy;	89;56;10;	2009;19/06/2009		
20	Albert Nobbs;	Drama;	53;43;8;	2011;01/06/2011		
21	Alice in Wonderland;	Adventure;	52;72;200;	2010;04/08/2010		
22	Alien vs. Predator -- Requiem;	Horror;	14;37;40;	2007;17/12/2007		
23	Aliens in the Attic ;	Adventure;	30;46;45;	2009;03/01/2009		
24	All About Steve ;	Comedy;	6;35;15;	2009;18/04/2009		
25	All Good Things;	Drama;	33;64;20;	2010;29/08/2010		

Ilustración 1: Foto del fichero CSV

En el proyecto tendremos que averiguar, por una parte, una lista con las películas que coincidan con un género, las películas cuyas ganancias no coincidan, la media de la valoraciones en un año, la suma de las ganancias producidas por un género y las películas mejores y peores valoradas, y por otra parte tendremos que generar dos diccionarios con unas claves concretas.

Habrà un fichero llamado **peliculas.py** en la carpeta “src” el cuál contendrá las siguientes funciones definidas en su interior:

1. **recorre_fichero**: Recibe la ruta de un fichero CSV codificado en UTF-8, y devuelve una lista de tuplas de tipo películas (str,str,int,int,int,int,int) conteniendo todos los datos almacenados en el fichero.
2. **filtrar_por_genero**: Recibe una lista de tuplas de tipo películas y un género de tipo str, y devuelve otra lista de tuplas (str, int, int, int) con los nombre de las películas que coincidan con el género cinematográfico seleccionado anteriormente, además del año en el que salió, las ganancias y la fecha de estreno. El género puede ser ‘Comedy’, ‘Drama’, ‘Action’, ‘Adventure’, ‘Horror’, ‘Romance’ o ‘Thriller’



3. **ganancias_sin_repetir**: Recibe una lista de tuplas de tipo películas y devuelve otra lista de tuplas que contenga todos los títulos de las películas (str) que no hayan registrado el mismo número de ganancias con su respectivo género(str), las valoraciones de Rotten Tomatoes (int), las valoraciones de la audiencia(int) y el año (int) y fecha de estreno (int)
4. **calcula_media_año**: Recibe una lista de tuplas de tipo películas y un año de tipo int y devuelve una lista de tuplas (año, media) de tipo (int,int) con la media numérica de las valoraciones que hizo la audiencia a todas las películas que salieron el año seleccionado.
5. **calcula_ganancias_genero**: Recibe una lista de tuplas de tipo películas y un género de tipo str y devuelve una lista de tuplas (género, ganancias) de tipo (str,int) que muestra las ganancias totales que ha tenido ese género concreto.
6. **peliculas_peores_valoradas**: Recibe una lista de tuplas de tipo películas y un año de tipo int y devuelve una lista de tuplas (peores_peliculas) de tipo (str) que nos muestra las películas que han obtenido una calificación menor al 20% en Rotten Tomatoes en el año seleccionado.
7. **peliculas_mejores_valoradas**: Recibe una lista de tuplas de tipo películas y un parámetro *ranking* (de tipo int que vaya del 1 al 100) y devuelve una lista de tuplas (mejores_peliculas) de tipo (str) con las películas mejor valoradas por la audiencia en un ranking cuyo número está marcado por el parámetro dado. El parámetro *ranking* tendrá por defecto el número cero.
8. **diccionario_de_genero**: Recibe una lista de tuplas de tipo películas y un género de tipo str y devuelve un diccionario {str,[str,int,int]} cuyas claves son el título de la película, el año en el que salió y el dinero ganado con la película. A cada clave se le asigna una lista con los registros que coincidan con el género seleccionado
9. **diccionario_de_año**: Recibe una lista de tuplas de tipo películas y un año de estreno y devuelve un diccionario {int,[int,int]} cuyas claves son las valoraciones de la audiencia y las valoraciones de Rotten Tomatoes. A cada clave se le asigna una lista con los registros que cumplan las condiciones marcadas por las claves: en el que caso de la valoración de la audiencia que supere el 60% y el de Rotten Tomatoes el 40%.

Al final del proyecto, en la carpeta "src", habrá un fichero **peliculas_TEST.py** donde estarán importadas todas las funciones del módulo películas y todos los datos del fichero CSV cargados, en adición mostrará en consola los datos leídos. Incluye además llamadas a todas las funciones implementadas mostrando los resultados en la consola.