

# Instrucciones Miniproyecto 2

## El Set de Datos

El set de datos con que trabajaremos es un subconjunto del dataset público de IMDB (la base de datos mundial de información de películas). Se ha recortado solo para hacerlo mas manejable (por ejemplo, solo hay actores que comienzan con la letra A).

Hay 5 archivos csv que deberemos cargar en nuestra base de datos:

- movies.csv (4652 filas)
- actors.csv (3655 filas)
- directors.csv (3655 filas)
- movies\_actors.csv (154 filas)
- movies\_directors.csv (4255 filas)

## El trabajo a realizar

1. Usar el MySQL Workbench para crear un base de datos que llamaremos "Cine" y las 5 tablas donde cargaremos los archivos de datos
  - movies(id, name, year, rank)
  - actors(id, first\_name, last\_name)
  - directors(id, first\_name, last\_name)
  - movies\_actors(actor\_id, movie\_id, role)
  - movies\_directors(director\_id, movie\_id)
2. Escribir un programa **Python CargaBD.py** que cargue la información de los archivos csv en las tablas respectivas. Corra el programa para revisar que efectivamente se cargó la información.
3. Escribir un programa Python Queries.py que efectúe 3 consultas interesantes a la base de datos:

- 3.1 La primera query nos va a entregar una lista de los directores que tienen más de 3 películas ordenadas en orden decreciente:

```
SELECT d.last_name, d.first_name, COUNT(movie_id) AS 'How Many'
FROM movies_directors AS md JOIN directors AS d ON d.id =
md.director_id
GROUP BY d.last_name, d.first_name
HAVING COUNT(movie_id) > 3
ORDER BY COUNT(movie_id) DESC;
```

- 3.2 La segunda query es muy similar a la anterior pero el ranking es ahora de actores y se despliega la cuenta de películas para todos ordenados por apellido:

```
SELECT a.last_name, a.first_name, COUNT(movie_id)
FROM actors AS a JOIN movies_actors as ma on ma.actor_id = a.id
GROUP BY a.last_name, a.first_name
ORDER BY a.last_name, a.first_name;
```

- 3.3 La tercera query despliega una lista de películas, el año, su director y el puntaje (rank) solo para las películas con rank mayor a 8 ordenadas en forma decreciente:

```
SELECT m.name as 'Movie', m.year AS 'Year', d.last_name AS 'Director',
m.rank as 'Rank'
FROM (movies_directors AS md JOIN movies as m on m.id = md.movie_id)
JOIN directors AS d ON d.id = director_id
WHERE m.rank > 8
ORDER BY m.rank DESC;
```

***\*En cada caso use un loop de print para imprimir las filas resultantes***

4. Escriba un programa ***Python Dataframes.py*** que efectúe la consulta 3.3, pero el resultado lo cargue en un dataframe Pandas df1 con las siguientes columnas: Pelicula, Año, Director, Puntaje
  - Recorte ese dataframe usando loc de modo de tomar solo las primeras 10 filas y solo las columnas Película y Puntaje. Imprima el nuevo Dataframe resultante.
  - Recorte nuevamente df1 ahora usando iloc de modo de tomar las filas 20 a la 50 y todas las columnas. Imprima el nuevo Dataframe resultante.