

Ejercicios Spark

Dataset

Se utilizará el dataset llamado `movie_metadata.csv` donde se listan un grupo de películas así como sus respectivas características.

Preguntas

Aspecto económico

No nos interesa saber qué director tiene más películas en la lista (dado que el dataset puede ser limitado), pero lo que si nos interesa es notar otras características que nos permitan realizar mejor inversiones en la producción de una película.

- ¿ Qué director es el que hace más dinero ? La columna donde se muestra el ingreso de la película es la que se llama **gross**.
- Hay veces que la película ha recaudado bastante dinero, pero esto también debido a una gran inversión hecha. ¿Cuáles son las 5 películas que más margen (ingresos - presupuesto) han tenido? El presupuesto es la columna llamada **budget**.

Calidad

- Una película no solamente debe dar ganancias si no también ser del agrado de los críticos. Para esto la página web IMDB cuenta con un ranking del 1 al 10 que califica el gusto de ellos por la película. ¿Qué director hace *mejores* películas en cuanto a crítica?
- Pero no solo la opinión de los críticos es importante, si no también la opinión de los espectadores. Para esto tenemos la columna `movie_facebook_likes`. Encontrar los 5 mejores directores cuyas películas sean de la predilección de los espectadores.
- ¿Existe alguna relación entre la calificación de los críticos y el gusto de los espectadores?

Elección de actores

Se dice que hay actores que actúan mejor en ciertos géneros de películas. Teniendo en cuenta el número de likes en facebook que cada actor tiene, ¿qué actores son lo más apreciados por género?