**Trabajo fin de Máster:  
SISTEMA DE RECOMENDACIONES DE PELICULAS**

# INTRODUCCIÓN:

Los sistemas de recomendación son algoritmos destinados a sugerir items relevantes a los usuarios, principalmente a partir de elementos de búsquedas o su histórico. Existen diferentes modelos, como “popularity”, y “colaborative”; el primero hace recomendaciones globales pero sin personalizar la misma; y el segundo utiliza datos de otros usuarios para generar recomendaciones a perfiles similares. Estas herramientas son fundamentales para servicios web (Amazon, Youtube, Neflix… ).

# OBJETIVOS

**PRIMARIO:** Construir dos modelos de recomendación de película, uno básico y otro colaborativo.

# METODOLOGÍA

Para el análisis descriptivo y construcción del sistema de recomendación se extraen los datos del portal web MovieLens

Los CSV utilizados son metadatos de 45,000 películas enumeradas en el conjunto de datos completo de MovieLens. El conjunto es anterior a julio de 2017. Estos incluyen reparto, equipo, palabras clave de la trama, presupuesto, ingresos, carteles, fechas de lanzamiento, idiomas, compañías de producción, países, recuentos de votos TMDB y promedios de votos.

Contienen 26 millones de calificaciones de 270,000 usuarios para las 45,000 películas. Las clasificaciones están en una escala de 1-5 y se han obtenido del sitio web oficial de GroupLens.

Para el sistema de recomendación se han utilizado las cvs:

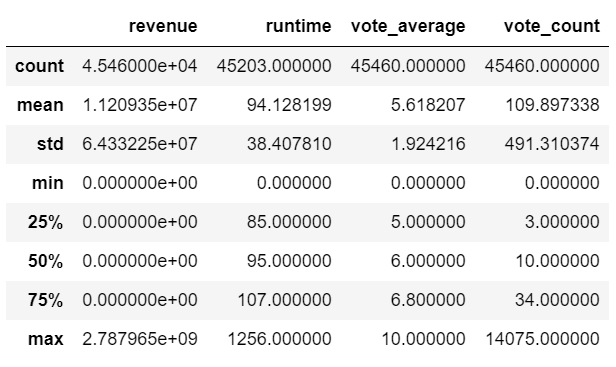
* movies\_metadata.csv: El archivo principal de metadatos de películas. Contiene información sobre 45,000 películas presentadas en el conjunto de datos Full MovieLens. Las características incluyen carteles, fondos, presupuesto, ingresos, fechas de lanzamiento, idiomas, países de producción y empresas.
* Créditos.csv: Consiste en información sobre el elenco y el equipo de todas nuestras películas.
* ratings\_small.csv: El subconjunto de 100,000 calificaciones de 700 usuarios en 9,000 películas.

Análisis: Se realizó un análisis descriptivo de los datos y un sistema de recomendación. Para las variables continuas, se calcularon medias, desviación standard, medianas y percentiles; Para las categóricas, se calcularon los porcentajes. Para el análisis de los datos se utilizó Jupyter notebook Python. Para el sistema de recomendación se utilizó SurPRISE (Simple Python RecommendatIon System Engine).

# REsumen de resultados

Se parte de una base de datos de 45466 películas

**Características basales**



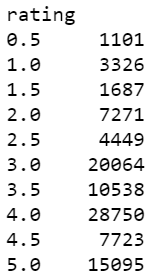
**Visualización del rating**



671 usuarios y 9066 películas valoradas.

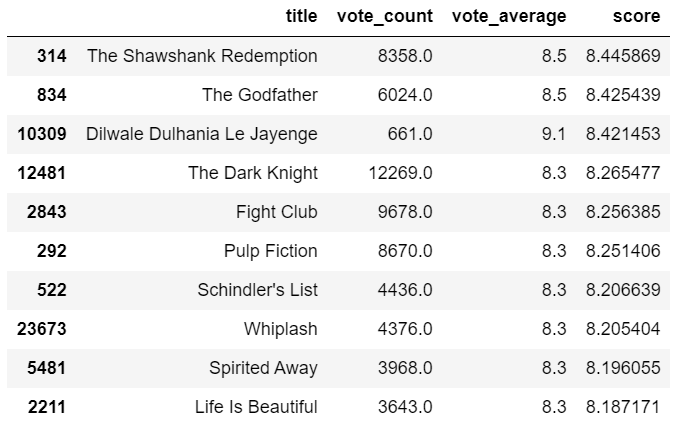
28000 valoraciones con una puntuación de 4 y unos 20000 con puntuación en 3.

**Valoración real rating**

****

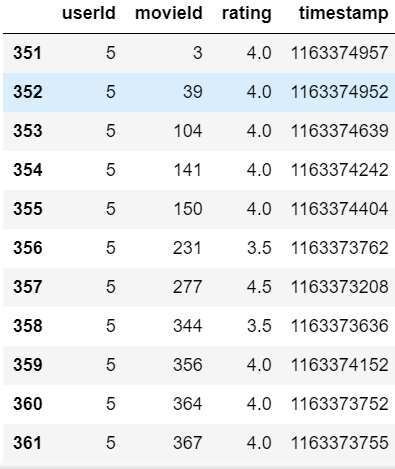
**Filtrado demográfico - Sistema de recomendación básico.**

Tras el análisis, obtenemos las 10 mejores películas según las calificaciones ponderadas de los usuarios. Se trata de una recomendación genérica, no personalizada para el usuario, ya que no dispone de un histórico de un usuario.

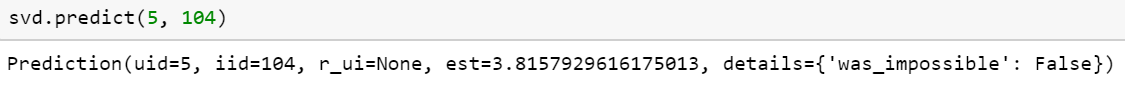


**Sistema de recomendación colaborativo**

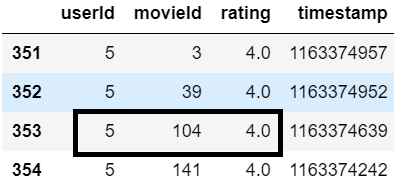
Creo un objeto SVD para llamar al método de ajuste que hace que RMSE sea menor y luego se transforma en dos matrices.

****

Realizamos una simulación de recomendación. Seleccionamos un usuario ('5'), y un película ('104').



Predicción rating 3,8; real 4,0

** Buena aproximación de recomendación**

# CONCLUSIONES

1. El análisis de bases de metadatos de usuarios mediante filtros permite crear sistemas de recomendaciones tanto genéricos como individualizados.
2. Las recomendaciones colaborativos permiten mejorar la experiencia del usuario, frente a los genéricos que son más elementales.