

# Projet 5 : Notation de films

## Dataset :

La société GroupLens Research récolte les données de notations de films par les utilisateurs du site web MovieLens. Ces données sont accessibles ici :

<http://grouplens.org/datasets/movielens/>

Elles se présentent sous la forme d'une matrice (utilisateur, film) avec pour entrée la note du film.

Vous pouvez commencer par utiliser le jeu de données possédant 100 000 notations, avant d'essayer d'analyser les plus gros dossiers (1M, 10M ou 20M).

## Objectif :

Pour un premier pas sur l'utilisation de ces données, vous pouvez vous appuyer sur ce qui est présenté dans ce lien :

<http://www.gregreda.com/2013/10/26/using-pandas-on-the-movielens-dataset/>

Il est possible d'utiliser ce dataset pour construire un moteur de recommandation de film, en identifiant les utilisateurs aux profils similaires et en leur proposant des films communs.

Ce problème peut-être approché de différentes façons. Ce projet sera l'occasion d'introduire les méthodes de factorisation matricielle, et notamment la SVD (décomposition en valeurs singulières), qui trouve aussi son usage dans le calcul des composantes principales que nous avons vu en cours (PCA), et est une alternative à la diagonalisation de la matrice de covariance.

Ces méthodes ont des applications diverses en analyse d'image, de sons, de textes et en chimiométrie.