
CS 6220 Data Mining — Assignment 1

Exploring the MovieLens Dataset

On this assignment, you'll describe how a simple recommendation system can be crafted to leverage user ratings and generate individual recommendations for movies, products, and anything else that can be rated.

Below are two links to material that will serve as a solid starting point for what you will submit:

1. <https://goo.gl/qZUEm8> - A notebook with some initial data exploration of the MovieLens dataset.
2. <http://www.gregreda.com/2013/10/26/using-pandas-on-the-movielens-dataset/> - A blog post containing some further analysis of that dataset.

Objective:

Download the data (link 1 above contains instructions on how you can do that) and extend the presented data exploration to include:

1. [10 pts] An aggregate on the number of rating done for each particular genre, e.g., Action, Adventure, Drama, Science Fiction, ...
2. [5 pts] The top 5 ranked genres by women on most number of rating.
3. [5 pts] The top 5 ranked genres by men on most number of rating.
4. [30 pts] Pick a genre of your choice and provide average movie's ratings by the following four time intervals during which the movies were released (a) 1970 to 1979 (b) 1980 to 1989 (c) 1990 to 1999 (d) 2000 to 2009. Also, if you observed any issues with data in any of these ranges, please mention it.
5. [30 pts] A function that given a **genre** and a **rating_range** (i.e. [3.5, 4]), returns all the movies of that genre and within that rating range sorted by average rating. Using an example, demonstrate that your function works.

-
6. [20 pts] Present one other statistic, figure, aggregate, or plot that you created using this dataset, along with a short description of what interesting observations you derived from it. This question is meant to give you a freehand to explore and present aspects of the dataset that interests you.

You may use the code available on both of the provided notebook files as a starting point.
Submission: Submit your ipynb file through the assignment submission portal on Canvas.