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 any- thing else that can be rated.

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

1. A notebook with some initial data exploration of the MovieLens dataset. LINK 1

2. A blog post containing some further analysis of that dataset. LINK 2

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

- [10 pts] An aggregate on the number of rating done for each particular genre, e.g., Action, Adventure, Drama, Science Fiction, ...

- [5 pts] The top 5 ranked genres by women on most number of rating.

- [5 pts] The top 5 ranked genres by men on most number of rating.

- [30 pts] Pick a genre of your choice and provide a breakdown of movie's ratings by four year ranges (a) 1919 to 1939 (b) 1939 to 1959 (c) 1959 to 1979 (d) 1979 to 2000

- [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 rating.

- [20 pts] Some 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.

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 BB.