Assignment 1 Instructions: Non-Personalized Recommenders

Overview

This assignment will explore non-personalized recommendations. You will be given a 20x20 matrix where columns represent movies, rows represent users, and each cell represents a user-movie rating.

Download matrix data (https://eventing.coursera.org/api/redirectStrict/A-VreFPS_Sn91xga1z-BGCMnewP2HFWsC39JTwAHyF6Okhu8munszaUd7hBZZPw9fVlj3s3lbeKFoPvs5CmLwA.CoKd9qVR89XgIRN1VgKdbw.vgrr2sb1jEYxkVzX8pVFc0yNrSR8AlT3HOvDzibr t3s6cZsY6Kf0CH2Z0ESZ76gTyUjnHdbpmNmqFRyEP-4I55NsdxZTD-GdDx29Mgbp4NNvp9LrL4TFYqcnFii5S0U7WtQOBid61U2YE0tVtPWWDVpleIiXdHs2rTzRAf5veSfY4JlmdsPeLa3WL9gYmb9iQLSFII-L3Kqocb16kNoJ-UvcEXwbfnhhB8eKfBhvffFm3BWse4-59zMq-2ZF2N3iaZJ4-pglCNjJPdkwPB76322e1-p7ekLeizuQarNCd7prFTxeWcRAaVPqFz2ellZyUyL9PDhxz7tbr0Z2OZOi2aSKBifDP6LbaJAmGfNzFo2_OP21iT_Kwz3wbiBhx3YL-n6GphlFkD6Vh9bEyFlvivrgBjywr3GxfLxUT4RBFR5T4dPHeZrLtPF0mczc)

Deliverables

There are 4 deliverables for this assignment. Each deliverable represents a different analysis of the data provided to you. For each deliverable, you will submit a list of the top 5 movies as ranked by a particular metric. The 4 metrics are:

1. Mean Rating: Calculate the mean rating for each movie, order with the highest rating listed first, and submit the top 5.

2. % of ratings 4+: Calculate the percentage of ratings for each movie that are 4 or higher. Order with the highest percentage first, and submit the top 5.

3. Rating Count: Count the number of ratings for each movie, order with the most number of ratings first, and submit the top 5.

4. Top 5 Star Wars: Calculate movies that most often occur with *Star Wars: Episode IV - A New Hope (1977)* using the $(x + y)/x$ method described in class. In other words, for each movie, calculate the percentage of Star Wars raters who also rated that movie. Order with the highest percentage first, and submit the top 5.

Output Format

For each part, you will need to enter your movie ID and scores into the appropriate quiz questions below.

Notes

- You can either do this by hand or using a spreadsheet program.

- Blank cells mean there is no rating. Do not factor them into your calculations.

Examples

In order to check your calculations, the values for *Raiders of the Lost Ark* are:

- Mean: 2.91

- Rating Count: 11

- % of 4+: 27.2%

- Association with *Star Wars Episode IV*: 46.7%