Written Assignment 2: Instructions

Help

When you're ready to submit your solution, go to the assignments list.

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

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 people who rated that movie *and Star Wars: Episode IV - A New Hope*. Order with the highest percentage first, and submit the top 5.

Output Format

For each part, your output should be a plain text file with the top 5 movie IDs, one per line. Upload this file using the submit button for each part (1 file per part). You can easily create text file for submission by using Notepad (Windows) or a similar editor and then cutting/pasting the top 5 movie IDs from your analysis.

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*: 72.7%
