Isfahan University of Technology
Information Technology Fundamentals
Dr. Manshaei
Homework 2: Module 03 Web Systems
Deadline: Tuesday 13^{th} **Khordad, at** 23 : 59

We only accept the homework **delivered via lms(lms.iut.ac.ir), before the deadline**.

In this assignment you should first go over the attached spreadsheet. The spreadsheet includes headers that provide movie ID (and name, in case you're curious) and user ID (but not name). The second column represents gender, with 1=female and 0=male. The rest of the cells are a 20x20 rating matrix (user-movie ratings) where blank cells represent the lack of a rating.

There are 5 deliverables for this assignment. Each deliverable represents a different analysis of the data provided to you. The first three deliverables represent non-personalized summary statistics; The last two represent a demographic analysis to explore whether using gender would be wise given this dataset.

1. **Mean Rating:** Calculate the mean rating for each movie, order with the highest rating listed first, and submit the top three (along with the mean scores for the top two).
2. **Rating Count (popularity):** Count the number of ratings for each movie, order with the most number of ratings first, and submit the top three (along with the counts for the top two).
3. **% of ratings 4+ (liking):** Calculate the percentage of ratings for each movie that are 4 or higher. Order with the highest percentage first, and submit the top three (along with the percentage for the top two). Notice that the three different measures of "best" reflect different priorities and give different results; this should help you see why you need to be thoughtful about what metrics you use.
4. **Mean rating difference by gender:** First, recompute the mean rating for each movie separately for males and for females. And calculate the overall mean rating (across all ratings) for males and females. Submit the two movies that have the greatest differences (one where men are most above women, and one where women are most above men) along with the differences in average. Also submit the difference in overall rating averages (female average - male average).
5. Next, compute the % of ratings 4+ separately for males and females. You'll be asked to submit two movies as above (largest difference in each direction). And again you'll indicate whether men or women are more likely to rate movies 4 stars or above.