

CODING PROBLEM

The problem mentioned below revolves around movies dataset. The dataset contains 4 files which are follows,

File Name	Description / Schema
movies.dat	MovieID – Title – Genres
ratings.dat	UserID – MovieID – Rating – Timestamp
users.dat	UserID – Gender – Age – Occupation – ZipCode
README	Additonal information / explanation about the above three files

The dataset can be downloaded from the link : <http://grouplens.org/datasets/movielens/1m/>

You are required to write a code (Preferably in Java / Mapreduce / Python) to get results for the following questions,

1. Top ten most viewed movies with their movies Name (Ascending or Descending order)
2. Top twenty rated movies (Condition : The movie should be rated/viewed by at least 40 users)
3. Top twenty rated movies (which is calculated in the previous step) with no of views in the following age group
(Age group :
 1. Young (<20 years),
 2. Young Adult(20-40 years),
 - 3.adult (> 40years))
4. Top ten critics (Users who have given very low ratings; Condition : The users should have at least rated 40 movies)

Instructions:

- 1) Please make your program readable and well structured. Showcase your object oriented skills and/or functional programming skills.
- 2) Although the input size of the JSON/CSV is limited to few MBs, please make your code so that it can be extended to adapt to file sizes greater than heap size of the JVM or your primary memory.
- 3) We are not particular about formatting the output. It should just be readable.
- 4) You would get 3 Days to revert with the solution.
- 5) You can send us the ZIP file with the source code.