Characterizing videos and users in Netflix

RQ0:
How many unique users and videos are there in this dataset?
Answer:
Users = 480189
Movies = 17770

RQ1:
What are the top most rated videos (movies, series, documentary etc.) on Netflix? Show the first 50 with their names and average rating (for your convenience we show all the videos so you can compare).

RQ2:
What are the rating distrubition of Netflix videos?
Draw a cumulitive distribution function (CDF) for all the ratings combining all the videos. This research question, naively, will help understand the quality of Netflix videos!

RQ3: Actually the result of RQ2 can be very much dominated by a few movies (they have so many users and many high ratings because they are extraordinary movies/series, such as the lord of the rings, and the breaking bad). Better, we should take one rating from each movies (median) and then draw the CDF.

***Don’t consider movies with less than 50 users.

Hint: For drawing the CDF of medians, we need to first convert the floating point medians to rounded integer (use the round function in python), and then typecast to integer if necessary.

RQ4:

Does phychology play a role in movie rating? How true it is that sometimes rating depend more on a user’s attitude than a actual movie’s quality?
Draw 4 different CDFs (in one graph) of rating for these 4 users (1488844,1248029,880166,543865). Are they very different in behavior?

RQ5:

Hey, let’s build a movie recommender. Yes, I am serious. Say, I watched one movie, what other movies should I watch based on my interest on that movie?

What are the 50 most similar movies (considering users’ common interests) for a given movie ID?

***Consider movies with at least 50 users.

Formula for calculating similarity:
|A ^ B| / |A U B| (Watch the video to understand more).