Hackerrank: Challenge Recommendation

Data quality and any errors:

1. “Challeges.csv”: many challenge problems with missing domain and subdomain
2. “submissions.csv”: some users had very small sample data to study

Data Processing Steps:

1. Due to limited time and sampling size, I decided not to use any fancy machine learning algorithm; instead, I created two dictionaries to see which problem the user chooses to solve sequentially regardless of the result.
2. Create two average difficulties list for each user (one with solved=True and another one with solved=False); this will be another factor for program to choose next problem set
3. Create a dictionary with a list of challenge problems along with difficulty level

Model:

1. As mentioned above, no fancy math/model was used
2. For each user, calculate average difficulties for solved=True and solved=False
3. Based on history, find candidate problems that each user would choose (depending on whether user got current question correctly or not)
4. If the number of suggested problem is less than 10, look at the domain the user was looking and pull N number of popular problems based on the user’s scores