

Part 1

1. When I created my test data, I would like to have enough common users for each of the movies to guarantee the similarity function would work properly when using my test file.
2. Yes. I designed my test data in the exact same format with the real data, and I tried to raise both positive and negative similarity in the test data. Also, I must make sure that there are no duplicates in my test file, i.e., one user cannot rate the same movie twice.
3. Yes. I first created a reference similarity matrix as the “reference” solution, then used the similarity matrix mathematically to create the test data, i.e., the movie ratings have the same means with my reference, etc.

Part 2

The reasons I split my code of the assignment into several functions are:

- 1) Each function has a determined functionality and distinct parameter names, so it is easier to write a function and passing the variable name into functions is good to help avoid variable name confusion.
- 2) It is simply easier to debug.