Allan Chesarone

The url of the public github repo: https://github.com/ACHEZ9/movies-2

The url of the CodeClimate report: https://codeclimate.com/github/ACHEZ9/movies-2

The Algorithm:

```
predict(user, movie)
        get most_similar users to user
        for each of the most similar users
                if the user has rated the movie
                        add thier rating to the total
                        increase the count of movies

        return the average rating

most_similar(user)
        sort the users in the training set by similiarity to the user. Most similar appear first

similarity(user, user2)
        get all the movies both users have seen

        for each movie that they have in common
                add 1 minus their difference in rating/3
```

The advantages are that it will be using the most similar users to the user for each prediction it is making. This can be changes but just changing the similarity method.The disadvantage is that the most_similar method takes a long time to run for each user, making the total time very long if there are many different users for predictions.

The Analysis and Benchmarking:
u1 file

The test was run on the first 1000 samples from u1
The average prediction error is 0.8724738800831067
The standard deviation of the errors is 0.6992409255975227
The RMSE of the prediction is 1.1181004174302727
The time it took to run 1000 predictions, was 0.746627 seconds

The test was run on the first 10000 samples from u1
The average prediction error is 0.8472281768972766
The standard deviation of the errors is 0.6726137296541018
The RMSE of the prediction is 1.081760053361133
The time it took to run 10000 predictions, was 7.873147 seconds

The test was run on the first 20000 samples from u1

The average prediction error is 0.8348751225123806
The standard deviation of the errors is 0.6728151496180255
The RMSE of the prediction is 1.0722391038129468
The time it took to run 20000 predictions, was 23.289482 seconds

u2 file

The test was run on the first 1000 samples from u2
The average prediction error is 0.8675091112617573
The standard deviation of the errors is 0.6906031380280226
The RMSE of the prediction is 1.108830353289589
The time it took to run 1000 predictions, was 1.777585 seconds

The test was run on the first 10000 samples from u2
The average prediction error is 0.8201692885508228
The standard deviation of the errors is 0.6559127913857875
The RMSE of the prediction is 1.0501901026887808
The time it took to run 10000 predictions, was 18.379282 seconds

The test was run on the first 20000 samples from u2
The average prediction error is 0.8284006003107688
The standard deviation of the errors is 0.6637057996925291
The RMSE of the prediction is 1.0614861954546282
The time it took to run 20000 predictions, was 35.231675 seconds

The running time is increasing exponentially, so if the user set was to increase by 1000 or 10000, it would take a very long time to run through a set of predictions for every users, as the most_similar method goes though ever single user in the set.