#README

Goal of program

MovieRecommender.py is a Python program that generates personalized movie recommendations using user–user collaborative filtering.

The program takes as input a CSV file containing a user's past movie ratings, such as one exported from IMDb, MovieLens, or another rating platform. After loading this file, the program compares the user's ratings to thousands of other users from the MovieLens dataset.

It computes similarity scores between users based on the movies they have both rated, identifies the users whose tastes most closely match the input user, and then recommends movies that these similar users enjoyed but the input user has not yet rated.

Finally, the program outputs a ranked list of the top 10 movie recommendations, each weighted by similarity and rating strength.

You can increase the amount of users sampled from data set by editing the SAMPLE_N parameter potentially increasing the accuracy of recommendation, but it will increase the loading time of program. From personal experience a SAMPLE_N of 2 million users seemed to be accurate enough without increasing the loading time by too much. If you want to use full dataset of 32M users input None.

To see if the code is successful I highly recommend watching your top recommended movie!!!, it only gets better the more data you give it on your preferences.

How to run code

To run the program:

1. Download the folder/file provided that has both the program (MovieRecommender.py) and dataset (ml-32m)
2. ensure that folder ml-32m is in the same folder as the code, and contains the files:
    a. movies.csv
    b. ratings.csv.

For input files, you can either use the file of test data I provided (my own personal MovieLens ratings from my account, included in folder "Yash MovieLens Ratings.csv") or alternatively use

your own IMDB or MovieLens past ratings by exporting your own personal csv file.
To do so:

- On MovieLens: go to your profile → “Export Ratings”
- On IMDb: go to your ratings → “Export CSV”

When prompted please copy paste the path to your file (it doesn't need to be in the program folder), the program should output the top 10 recommendations in both terminal and in a file generated in the same folder as the program.

Notes

- The program accepts many different input CSV formats and automatically detects column names like movieId, movie_id, rating, etc.
- Invalid rows or missing ratings are skipped safely.
- If some movie IDs in your file do not exist in MovieLens, the program warns you but still runs correctly.