Assignment 6

Ryan Dill
CS 432
April 01, 2018

1. Find 3 users who are closest to you in terms of age, gender, and occupation. For each of those 3 users:

- what are their top 3 favorite films?
- bottom 3 least favorite films?

Based on the movie values in those 6 tables (3 users X (favorite + least)), choose a user that you feel is most like you. Feel free to note any outliers (e.g., "I mostly identify with user 123, except I did not like ``Ghost'' at all").

This user is the "substitute you".

For this question I created a python script name assignment6_1.py. This script took euclidean distance from each user to me. Here is a snippet from that code.

```
def euclideanDistance(P, Q):
    if( len(P) != len(Q) ):
        return -1
    sumOfSquares = 0
    for i in range(0, len(P)):
        sumOfSquares += (P[i] - Q[i]) * (P[i] - Q[i])
    return math.sqrt(sumOfSquares)
```

Reading the sorted results of that code, I am able to see the three users with the closest distance to myself. They are as follows:

124
183
272

I used assignment6_2.py to find the top and bottom movies for each user.

User 124:

Top Movies:
Monty Python & The Quest for the Holy Grail
Seven
Monty Python & The Life of Brian

Bottom Movies:
Its a Wonderful Life

User 183:

Top Movies:
The Good, The Bad and the Ugly
Star Trek IV: The Voyage Home
Sabrina

Bottom Movies:
101 Dalmations
Barb Wire

User 272:

Top Movies:
The Usual Suspects
Braveheart
Pulp Fiction

Bottom Movies:
Gone with the Wind

Overall, I have to choose user 272 as my closest match. He had a large list of movies that I saw or that I would like to see.

2. Which 5 users are most correlated to the substitute you? Which 5 users are least correlated (i.e., negative correlation)?

Most of this code is taken from https://github.com/arthur-e/Programming-Collective-Intelligence/blob/master/chapter2/recommendations.py. The results are as follows.

Correlate	User ID
0.6321265341207492	199
0.5072342224829962	238
0.899228803025897	782
0.6585061837190669	658
0.8660254037844402	634

Negative Correlate	User ID
-0.04828045495852676	205
-0.010689310486389712	286
-0.03292431988832048	709
-0.018486846666159188	663
-0.03464794643376185	825

3. Here I used assignment6_4.py. Most of this code is taken from https://github.com/arthur-e/Programming-Collective-Intelligence/blob/master/chapter2/recommendations.py. The results are as follows.

Recommend
'Boys, Les'
Sliding Doors
Basic Instinct
My Fair Lady
Brassed Off

Not Recommended
Anne Frank Remembered
Bushwhacked
Shooting Fish
Excess Baggage
Killer(Bulletproof Heart)

4. Here I used assignment6_5.py. This code is lifted from http://www.gregreda.com/2013/10/26/using\-pandas\-on\-the\-movielens\-dataset/. Here is the results from this code.

```
Recommend
They Made Me a Criminal
Marlene Dietrich: Shadow and Light
The Saint of Fort Washington
Someone Else's America
Star Kid

Not Recommended
The Eye of Vichy
Butterfly Kiss
Daens
JLG/JLG – autoportrait de decembre
Touki Bouki
```

I have never seen nor heard of these movies, so I can't say if this recommendation gives good data.