Old Dominion Univeristy

CS 495: Introduction to Web Science
Instructor: Michael L. Nelson, Ph.D
Fall 2014 4:20pm - 7:10pm R, ECSB 2120

Assignment # 8

George C. Micros UIN: 00757376

Honor Pledge

I pledge to support the Honor System of Old Dominion University. I will refrain from any form of academic dishonesty or deception, such as cheating or plagIiarism. I am aware that as a member of the academic community it is my responsibility to turn in all suspected violations of the Honor Code. I will report to a hearing if summoned.

Signed _____
November 13, 2014

George C. Micros

Written Assignment 8

Fall 2014
CS 495: Introduction to Web Science
Dr. Michael Nelson

November 13, 2014

Contents

Chapter 1
Written Assignment 8

The goal of this project is to use the basic recommendation principles we have learned for user-collected data. You will modify the code given to you which performs movie recommendations from the MovieLense data sets.The MovieLense data sets were collected by the GroupLens Research Project at the University of Minnesota during the seven-month period from September 19th, 1997 through April 22nd, 1998. It is available for download from `http://www.grouplens.org/node/73`

There are three files which we will use:

- u.data: 100,000 ratings by 943 users on 1,682 movies. Each user has rated at least 20 movies. Users and items are numbered consecutively from 1. The data is randomly ordered. This is a tab separated list of
  user id — item id — rating — timestamp
  The time stamps are unix seconds since 1/1/1970 UTC.
  Example:
  196 242 3 881250949
  186 302 3 891717742
  22 377 1 878887116
  244 51 2 880606923
  166 346 1 886397596
  298 474 4 884182806
  115 265 2 881171488
- u.item: Information about the 1,682 movies. This is a tab separated list of
  movie id — movie title — release date — video release date — IMDb URL — unknown — Action — Adventure — Animation —Children's — Comedy — Crime — Documentary — Drama — Fantasy — Film-Noir — Horror — Musical — Mystery — Romance — Sci-Fi — Thriller — War — Western —
  The last 19 fields are the genres, a 1 indicates the movie is of that genre, a 0 indicates it is not; movies can be in several genres at once. The movie ids are the ones used in the u.data data set.
  Example:
  161—Top Gun (1986)—01-Jan-1986——http://us.imdb.com/M/title-exact?Top162—On Golden Pond (1981)—01-Jan-1981——http://us.imdb.com/M/title-exact?On163—Return of the Pink Panther, The (1974)—01-Jan-1974——http://us.imdb.com/M/title-exact?Return
- u.user: Demographic information about the users. This is a tab separated list of:
  user id — age — gender — occupation — zip code
  The user ids are the ones used in the u.data data set.
  Example:
  1—24—M—technician—85711   2—53—F—other—94043   3—23—M—writer—32067
  4—24—M—technician—43537 5—33—F—other—15213

The code for reading from the u.data and u.item files and creating recommendations is described in the book Programming Collective Intelligence (check email for more details). You are to modify recommendations.py to answer the following questions. Each question your program answers correctly will award you 1 point.

1.1 Question 1

1.1.1 The Question

What 5 movies have the highest average ratings? Show the movies and their ratings sorted by their average ratings.

1.1.2 The Answer

5.0000	They Made Me a Criminal (1939)
5.0000	Star Kid (1997)
5.0000	Someone Else's America (1995)
5.0000	Santa with Muscles (1996)
5.0000	Saint of Fort Washington The (1993)
5.0000	Prefontaine (1997)
5.0000	Marlene Dietrich: Shadow and Light (1996)
5.0000	Great Day in Harlem A (1994)
5.0000	Entertaining Angels: The Dorothy Day Story (1996)
5.0000	Aiqing wansui (1994)
4.6250	Pather Panchali (1955)
4.5000	Some Mother's Son (1996)
4.5000	Maya Lin: A Strong Clear Vision (1994)
4.5000	Everest (1998)
4.5000	Anna (1996)
4.4911	Close Shave A (1995)
4.4664	Schindler's List (1993)
4.4661	Wrong Trousers The (1993)
4.4568	Casablanca (1942)

Table 1.1: Highest Average Rating

1.2 Question 2

1.2.1 The Question

What 5 movies received the most ratings? Show the movies and the number of ratings sorted by number of ratings.

1.2.2 The Answer

583	Star Wars (1977)
509	Contact (1997)
508	Fargo (1996)
507	Return of the Jedi (1983)
485	Liar Liar (1997)

Table 1.2: Most Ratings

1.3 Question 3

1.3.1 The Question

What 5 movies were rated the highest on average by women? Show the movies and their ratings sorted by ratings.

1.3.2 The Answer

5.0000 Year of the Horse (1997)
5.0000 Visitors The (Visiteurs Les) (1993)
5.0000 Telling Lies in America (1997)
5.0000 Stripes (1981)
5.0000 Someone Else's America (1995)
5.0000 Prefontaine (1997)
5.0000 Mina Tannenbaum (1994)
5.0000 Maya Lin: A Strong Clear Vision (1994)
5.0000 Foreign Correspondent (1940)
5.0000 Faster Pussycat! Kill! Kill! (1965)
5.0000 Everest (1998)

4.6329 Schindler's List (1993)

4.6316 Close Shave A (1995)

4.5625 Shawshank Redemption The (1994)

4.5333 Wallace & Gromit: The Best of Aardman Animation (1996)

Table 1.3: Highest Average Rating by Women

1.4 Question 4

1.4.1 The Question

What 5 movies were rated the highest on average by men? Show the movies and their ratings sorted by ratings.

1.4.2 The Answer

5.0000 Great Day in Harlem A (1994)
5.0000 They Made Me a Criminal (1939)
5.0000 Quiet Room The (1996)
5.0000 Hugo Pool (1997)
5.0000 Prefontaine (1997)
5.0000 Letter From Death Row A (1998)
5.0000 Marlene Dietrich: Shadow and Light (1996)
5.0000 Star Kid (1997)
5.0000 Delta of Venus (1994)
5.0000 Saint of Fort Washington The (1993)
5.0000 Santa with Muscles (1996)
5.0000 Aiqing wansui (1994)
5.0000 Love Serenade (1996)
5.0000 Leading Man The (1996)
5.0000 Entertaining Angels: The Dorothy Day Story (1996)
5.0000 Little City (1998)

4.6667 Two or Three Things I Know About Her (1966)
4.6667 Little Princess The (1939)

4.6250 Pather Panchali (1955)

4.5000 A Chef in Love (1996)
4.5000 Anna (1996)
4.5000 Sliding Doors (1998)
4.5000 Grosse Fatigue (1994)
4.5000 Bitter Sugar (Azucar Amargo) (1996)

4.4734 Casablanca (1942)

Table 1.4: Highest Average Rating by Men

1.5 Question 5

1.5.1 The Question

What movie received ratings most like Top Gun? Which movie received ratings that were least like Top Gun (negative correlation)?

1.5.2 The Answer

1.000000000 Ponette (1996)
1.000000000 The Innocent (1994)
1.000000000 Blood For Dracula (Andy Warhol's Dracula) (1974)
1.000000000 Man of the Year (1995)
1.000000000 Perfect Candidate A (1996)
1.000000000 Two or Three Things I Know About Her (1966)
1.000000000 Two Bits (1995)
1.000000000 Flower of My Secret The (Flor de mi secreto La) (1995)
1.000000000 Hear My Song (1991)
1.000000000 Dangerous Ground (1997)
1.000000000 Shiloh (1997)
1.000000000 Shooter The (1995)
1.000000000 Four Days in September (1997)
1.000000000 Albino Alligator (1996)
1.000000000 Last Time I Saw Paris The (1954)
1.000000000 That Old Feeling (1997)
1.000000000 Maybe Maybe Not (Bewegte Mann Der) (1994)
1.000000000 King of the Hill (1993)
1.000000000 Gone Fishin' (1997)

Table 1.5: Rating Most like Top Gun

-1.0000 Bad Moon (1996)
-1.0000 Frisk (1995)
-1.0000 Loch Ness (1995)
-1.0000 Telling Lies in America (1997)
-1.0000 Year of the Horse (1997)
-1.0000 Nil By Mouth (1997)
-1.0000 Love and Death on Long Island (1997)
-1.0000 Switchback (1997)
-1.0000 Carried Away (1996)
-1.0000 Two Much (1996)
-1.0000 Heidi Fleiss: Hollywood Madam (1995)
-1.0000 Broken English (1996)
-1.0000 Tetsuo II: Body Hammer (1992)
-1.0000 Caro Diario (Dear Diary) (1994)
-1.0000 Joy Luck Club The (1993)
-1.0000 Beat the Devil (1954)
-1.0000 Carpool (1996)
-1.0000 World of Apu The (Apur Sansar) (1959)

Table 1.6: Rated Least like Top Gun

1.6 Question 6

1.6.1 The Question

Which 5 raters rated the most films? Show the raters' IDs and the number of films each rated.

1.6.2 The Answer

737	405
685	655
636	13
540	450
518	276

Table 1.7: Raters with the Most Reviews

1.7 Question 9

1.7.1 The Question

What movie was rated highest on average by men over 40? By men under 40?

1.7.2 The Answer

5.0000 World of Apu The (Apur Sansar) (1959)
5.0000 Unstrung Heroes (1995)
5.0000 Two or Three Things I Know About Her (1966)
5.0000 They Made Me a Criminal (1939)
5.0000 Strawberry and Chocolate (Fresa y chocolate) (1993)
5.0000 Star Kid (1997)
5.0000 Spice World (1997)
5.0000 Solo (1996)
5.0000 Rendezvous in Paris (Rendez-vous de Paris Les) (1995)
5.0000 Prefontaine (1997)
5.0000 Poison Ivy II (1995)
5.0000 Marlene Dietrich: Shadow and Light (1996)
5.0000 Little Princess The (1939)
5.0000 Little City (1998)
5.0000 Leading Man The (1996)
5.0000 Late Bloomers (1996)
5.0000 Indian Summer (1996)
5.0000 Hearts and Minds (1996)
5.0000 Great Day in Harlem A (1994)
5.0000 Grateful Dead (1995)
5.0000 Faithful (1996)
5.0000 Double Happiness (1994)
5.0000 Boxing Helena (1993)
5.0000 Aparajito (1956)
5.0000 Ace Ventura: When Nature Calls (1995)
4.8000 Pather Panchali (1955)
4.6667 Whole Wide World The (1996)
4.6667 A Chef in Love (1996)
4.6471 Close Shave A (1995)
4.6000 Shanghai Triad (Yao a yao yao dao waipo qiao) (1995)

Table 1.8: Highest Average Rating by Men over 40

5.0000 Star Kid (1997)
5.0000 Santa with Muscles (1996)
5.0000 Saint of Fort Washington The (1993)
5.0000 Quiet Room The (1996)
5.0000 Prefontaine (1997)
5.0000 Perfect Candidate A (1996)
5.0000 Maya Lin: A Strong Clear Vision (1994)
5.0000 Magic Hour The (1998)
5.0000 Love in the Afternoon (1957)
5.0000 Love Serenade (1996)
5.0000 Letter From Death Row A (1998)
5.0000 Leading Man The (1996)
5.0000 Hugo Pool (1997)
5.0000 Entertaining Angels: The Dorothy Day Story (1996)
5.0000 Delta of Venus (1994)
5.0000 Crossfire (1947)
5.0000 Angel Baby (1995)
5.0000 Aiqing wansui (1994)

4.5000 Winter Guest The (1997)
4.5000 Two or Three Things I Know About Her (1966)
4.5000 Sum of Us The (1994)
4.5000 Sliding Doors (1998)
4.5000 Man of No Importance A (1994)
4.5000 Little Princess The (1939)
4.5000 Innocents The (1961)
4.5000 Grosse Fatigue (1994)
4.5000 Fille seule La (A Single Girl) (1995)
4.5000 Boy's Life 2 (1997)
4.5000 Anna (1996)

4.4762 Wallace & Gromit: The Best of Aardman Animation (1996)

4.4754 Casablanca (1942)

4.4706 Paths of Glory (1957)

Table 1.9: Highest Average Rating by Men under 40

1.8 Question 10

1.8.1 The Question

What movie was rated highest on average by women over 40? By women under 40?

1.8.2 The Answer

```
5.0000 Wrong Trousers The (1993)
5.0000 Visitors The (Visiteurs Les) (1993)
5.0000 Top Hat (1935)
5.0000 Tombstone (1993)
5.0000 Swept from the Sea (1997)
5.0000 Shallow Grave (1994)
5.0000 Shall We Dance? (1937)
5.0000 Safe (1995)
5.0000 Quest The (1996)
5.0000 Pocahontas (1995)
5.0000 Nightmare Before Christmas The (1993)
5.0000 Mina Tannenbaum (1994)
5.0000 Mary Shelley's Frankenstein (1994)
5.0000 Ma vie en rose (My Life in Pink) (1997)
5.0000 Letter From Death Row A (1998)
5.0000 In the Bleak Midwinter (1995)
5.0000 Great Dictator The (1940)
5.0000 Grand Day Out A (1992)
5.0000 Gold Diggers: The Secret of Bear Mountain (1995)
5.0000 Funny Face (1957)
5.0000 Foreign Correspondent (1940)
5.0000 Bride of Frankenstein (1935)
5.0000 Best Men (1997)
5.0000 Band Wagon The (1953)
5.0000 Balto (1995)
5.0000 Angel Baby (1995)
4.8000 Once Were Warriors (1994)
4.7000 Fantasia (1940)
4.6667 Last of the Mohicans The (1992)
4.5714 Christmas Carol A (1938)
```

Table 1.10: Highest Average Rating by Women over 40

5.0000 Year of the Horse (1997)
5.0000 Wedding Gift The (1994)
5.0000 Umbrellas of Cherbourg The (Parapluies de Cherbourg Les) (1964)
5.0000 Telling Lies in America (1997)
5.0000 Stripes (1981)
5.0000 Someone Else's America (1995)
5.0000 Prefontaine (1997)
5.0000 Nico Icon (1995)
5.0000 Mina Tannenbaum (1994)
5.0000 Maya Lin: A Strong Clear Vision (1994)
5.0000 Horseman on the Roof The (Hussard sur le toit Le) (1995)
5.0000 Heaven's Prisoners (1996)
5.0000 Grace of My Heart (1996)
5.0000 Faster Pussycat! Kill! Kill! (1965)
5.0000 Everest (1998)
5.0000 Don't Be a Menace to South Central While Drinking Your Juice in the Hood (1996)
5.0000 Backbeat (1993)

4.8182 Wallace & Gromit: The Best of Aardman Animation (1996)

4.8000 Paradise Lost: The Child Murders at Robin Hood Hills (1996)

4.8000 Anne Frank Remembered (1995)

4.7021 Shawshank Redemption The (1994)

4.7000 Shall We Dance? (1996)

Table 1.11: Highest Average Rating by Women under 40