Assignment 7

1. For this part of the assignment I used a slightly modified version on a code developed by Kevin Clemmons called data_extractor.py and substitute_you.py to create the resulting information below:

```
---------------------------Similar Users---------------------------
User: 33:
---------------------------Top-3---------------------------
1)Titanic (1997)......................................5.0
2)Game, The (1997)....................................4.0
3)Air Force One (1997)................................4.0
---------------------------Bottom-Three---------------------------
1)Liar Liar (1997)....................................3.0
2)Soul Food (1997)....................................3.0
3)Devil's Own, The (1997)............................3.0
---------------------------------------------------------------

User: 66:
---------------------------Top-3---------------------------
1)Return of the Jedi (1983)..........................5.0
2)Ransom (1996)......................................5.0
3)Air Force One (1997)...............................5.0
---------------------------Bottom-Three---------------------------
1)English Patient, The (1996)........................1.0
2)Muppet Treasure Island (1996).....................1.0
3)Excess Baggage (1997).............................1.0
---------------------------------------------------------------

User: 37:
---------------------------Top-3---------------------------
1)Pulp Fiction (1994)................................5.0
2)Raiders of the Lost Ark (1981)....................5.0
3)Terminator, The (1984)............................5.0
---------------------------Bottom-Three---------------------------
1)Jurassic Park (1993)..............................1.0
2)Twister (1996)....................................2.0
3)Arrival, The (1996)...............................2.0
---------------------------------------------------------------
```

The data I used was my name "Matt", my age "23", my gender "M", and my occupation "student" to get these results. I got a few more results, but chose these three. I chose User 37 as my substitute because I like all 3 of his top 3, though I did kind of like Jurassic Park even though they did not.

2. Using another slightly modified code developed by Kevin Clemmons called correlation.py I found users correlated most and least with User 37, and these were the results.

```
---------------------Results for: 37---------------------
-----------------Top Five Most Correlated-----------------
1)93.................................................+1.0
2)37.................................................+1.0
3)156................................................+1.0
4)163................................................+1.0
5)310................................................+1.0
-----------------Top Five Least Correlated-----------------
1)615................................................-1.0
2)641................................................-1.0
3)672................................................-1.0
4)736................................................-1.0
5)491................................................-1.0
---------------------------------------------------------------
```

3. Using a slightly modified code developed by Chevelle Taylor-Sakyi called rate.py I found the movies that would be best and worst recommended for User 37 based on correlation.

Assignment 7

```
Top 5 recommendations for films
Santa with Muscles (1996)  Movie Rate: 5.0
Great Day in Harlem, A (1994)  Movie Rate: 5.0
They Made Me a Criminal (1939)  Movie Rate: 5.0
Aiqing wansui (1994)  Movie Rate: 5.0
Someone Else's America (1995)  Movie Rate: 5.0

Bottom 5 recommendations for films
Rock, The (1996)    Movie Rate:  0
Twister (1996)    Movie Rate:  0
Speed (1994)    Movie Rate:  0
Clear and Present Danger (1994)    Movie Rate:  0
Dragonheart (1996)    Movie Rate:  0
```

4. Using another slightly modified code developed by Chevelle Taylor-Sakyi called like.py and the data given, I chose "Star Wars" (now known as Star Wars Episode IV) as my favorite from the list, and "Batman and Robin" as my least favorite. The resulting list based on correlation was a bit confusing admittedly.

```
Favorite movie: 50|Star Wars (1977)
/usr/lib/python2.7/dist-packages/scipy/stats/stats.py:2514: RuntimeWarning: invalid value encountered in double_sca
lars
  r = r_num / r_den
Top 5 most correlated movies:
Letter From Death Row, A (1998)  ( correlation: 1.0 )
Old Lady Who Walked in the Sea, The (Vieille qui marchait dans la mer, La) (1991)  ( correlation: 1.0 )
Scarlet Letter, The (1926)  ( correlation: 1.0 )
He Walked by Night (1948)  ( correlation: 1.0 )
Hurricane Streets (1998)  ( correlation: 1.0 )
Bottom 5 least correlated movies:
Bewegte Mann, Der (1994) ( correlation: -1.0 )
Sudden Manhattan (1996) ( correlation: -1.0 )
I Like It Like That (1994) ( correlation: -1.0 )
Sliding Doors (1998) ( correlation: -1.0 )
Slingshot, The (1993) ( correlation: -1.0 )

Least favorite movie: 254|Batman & Robin (1997)
Top 5 most correlated movies:
Apostle, The (1997)  ( correlation: 1.0 )
Once Upon a Time... When We Were Colored (1995)  ( correlation: 1.0 )
Farinelli: il castrato (1994)  ( correlation: 1.0 )
Once Were Warriors (1994)  ( correlation: 1.0 )
Welcome To Sarajevo (1997)  ( correlation: 1.0 )
Bottom 5 least correlated movies:
Contempt (Mépris, Le) (1963) ( correlation: -1.0 )
Mixed Nuts (1994) ( correlation: -1.0 )
Fluke (1995) ( correlation: -1.0 )
Portrait of a Lady, The (1996) ( correlation: -1.0 )
Search for One-eye Jimmy, The (1996) ( correlation: -1.0 )
```

I have never heard of any of the movies listed as most/least correlated for either film, except the Scarlet Letter and I've still never seen it. I can't really say whether or not I like most of them because I've never seen any of them, but I'm pretty sure a good number of them are not like Star Wars.