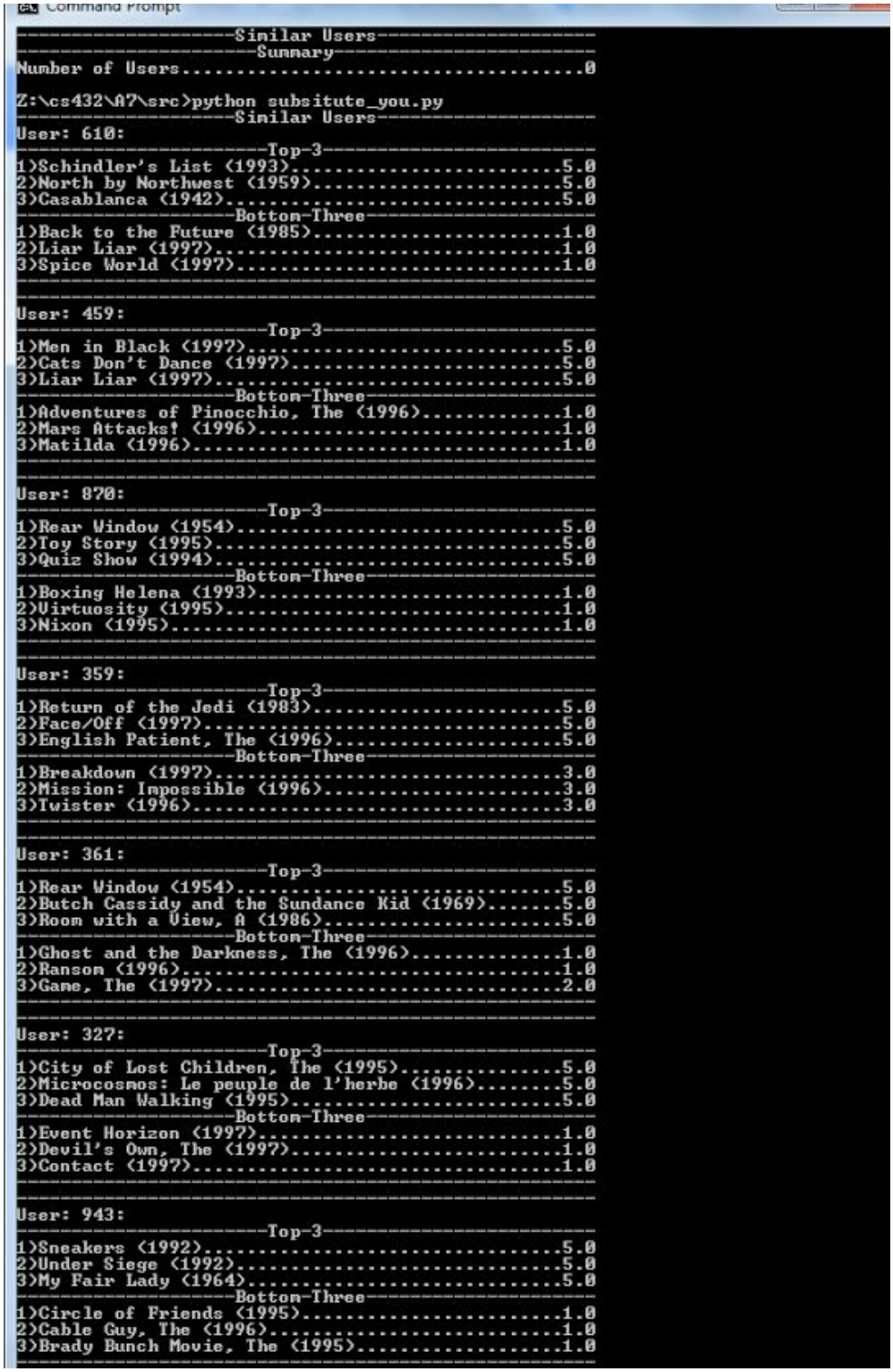
Alexander Wilson
A7 CS 432
3-31-16
1)



Top Matches

User 610:

Top Movies

1) Schindler's List
2) North by Northwest
3) Casablanca

Bottom Movies

1) Back to the Future
2) Liar Liar
3) Spice World

User 459:

Top Movies

1) Men in Black
2) Cats Don't Dance
3) Liar Liar

Bottom Movies

1) Adventures of Pinocchio, The
2) Mars Attacks!
3) Matilda

User 870:

Top Movies

1) Rear Window
2) Toy Story
3) Quiz Show

Bottom Movies

1) Boxing Helena
2) Virtuosity
3) Nixon

To find three users who are close to me based off age, gender, and occupation required a search for a 22 year old, student that is male. I was able to find those students who share these commonalities with q1.py. The program worked by creating a dictionary of people, movies, and ratings. It would then read to find my credentials on line 59 as to start formulating matches. The program ran by creating blank lists that the data will be uploaded to which will be sorted. After all the information was gathered and sorted the program would find the highest match cases. I ended up with 19 people who matched my credentials. I found to mostly identify with user 1870 because he is 22, a student, and his favorite movie is Rear Window. I also like Quiz Show a lot. However, I am not a fan of Toy Story. User 870 is substitute you.

2)

User 870 Correlation: Conditions- Name=User Gender=M Occupation=Student

Top Correlations

1) User 598… 1.0
2) User 355… 1.0
3) User 875… 0.946729262406
4) User 400… 0.877346477798
5) User 250… 0.863731292725

Least Correlated

1) User 242… -0.707106781187
2) User 35… -0.785714285714
3) User 427… -0.84162541153
4) User 310… -0.866025403784
5) User 866... -0.867527617236

For part two, I was asked to find the five users who most correlated to the substitute me also to find the five users who least correlate to me. I used the same method as in part one to gather the ratings and users. I used a function found on page 11 of Programming Collective Intelligence. Above are the values that correlate to my imaginary user, 870. There was only two people who were perfect matches to my credentials.

3) To compute the ratings for all the films that the substitute me has not seen I used q6.py. It allowed me to find the top 5 recommendations for films and the bottom 5 recommendations. As it appears I would absolutely hate Eat Drink Man Woman, I think it might be true. It was able to calculate the top five movies by using the genres to find out what type of movie would be in my top and bottom categories. Rear window is a thriller / action and since their is 19 different genres it calculates based off the combination of genres. The program then accounts for the user's least favorite movie genre and title to find similar movies that the user has not seen yet. My least popular movies for the most part fell into the drama category which is not my cup of tea. After the program would compile a list of the movies with proper genera it would then take the movie rating system to further develop the list.

```
Top 5 recommendations for films
Entertaining Angels: The Dorothy Day Story (1996)      Most likely rating:  5.0
Great Day in Harlem, A (1994)      Most likely rating:  5.0
They Made Me a Criminal (1939)      Most likely rating:  5.0
Aiqing wansui (1994)      Most likely rating:  5.0
Someone Else's America (1995)      Most likely rating:  5.0

Bottom 5 recommendations for films
Eat Drink Man Woman (1994)      Most likely rating:  0
Clerks (1994)      Most likely rating:  0
Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)      Most likely rating:  0
Trainspotting (1996)      Most likely rating:  0
Vertigo (1958)      Most likely rating:  0
```

4)

My favorite movie is Usual suspects, mainly because of the character Verbal. My least favorite movie was Sliding Doors For this part of the assignment I used the a modified version of q3.py. I do not believe my results were very accurate because recommendations are subjective. I also really like Good Will Hunting which I am recommended not to see.

```
sirius:~/cs432/A7/src> python temporary3.py
Most favorite movie: 12|Usual Suspects, The (1995)|14-Aug-1995|



Top 5 most correlated movies:
Nico Icon (1995)  ( correlation: 1.0 )
Savage Nights (Nuits fauves, Les) (1992)  ( correlation: 1.0 )
Ed (1996)  ( correlation: 1.0 )
Gang Related (1997)  ( correlation: 1.0 )
Cement Garden, The (1993)  ( correlation: 1.0 )
Bottom 5 least correlated movies:
New Jersey Drive (1995) ( correlation: -1.0 )
8 Seconds (1994) ( correlation: -1.0 )
Slingshot, The (1993) ( correlation: -1.0 )
Inkwell, The (1994) ( correlation: -1.0 )
Phat Beach (1996) ( correlation: -1.0 )

Least favorite movie: 1680|Sliding Doors (1998)|01-Jan-1998|
Top 5 most correlated movies:
Hoodlum (1997)  ( correlation: 1.0 )
Ulee's Gold (1997)  ( correlation: 1.0 )
Rosewood (1997)  ( correlation: 1.0 )
Good Will Hunting (1997)  ( correlation: 1.0 )
Restoration (1995)  ( correlation: 1.0 )
Bottom 5 least correlated movies:
I Know What You Did Last Summer (1997) ( correlation: -1.0 )
Smilla's Sense of Snow (1997) ( correlation: -1.0 )
Halloween: The Curse of Michael Myers (1995) ( correlation: -1.0 )
Amistad (1997) ( correlation: -1.0 )
She's So Lovely (1997) ( correlation: -1.0 )
```