

Assignment 7

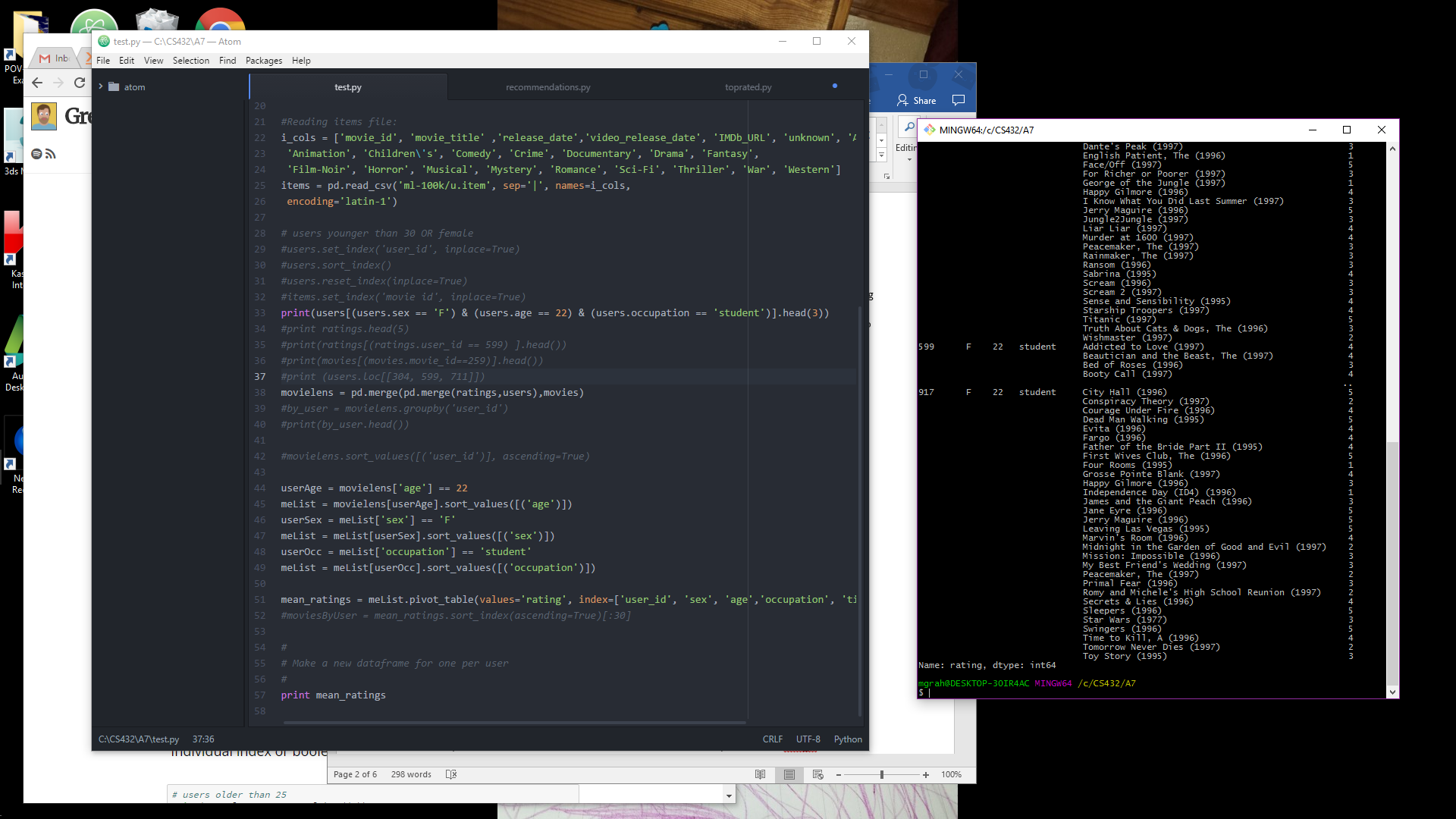
CS 432 Spring 2017



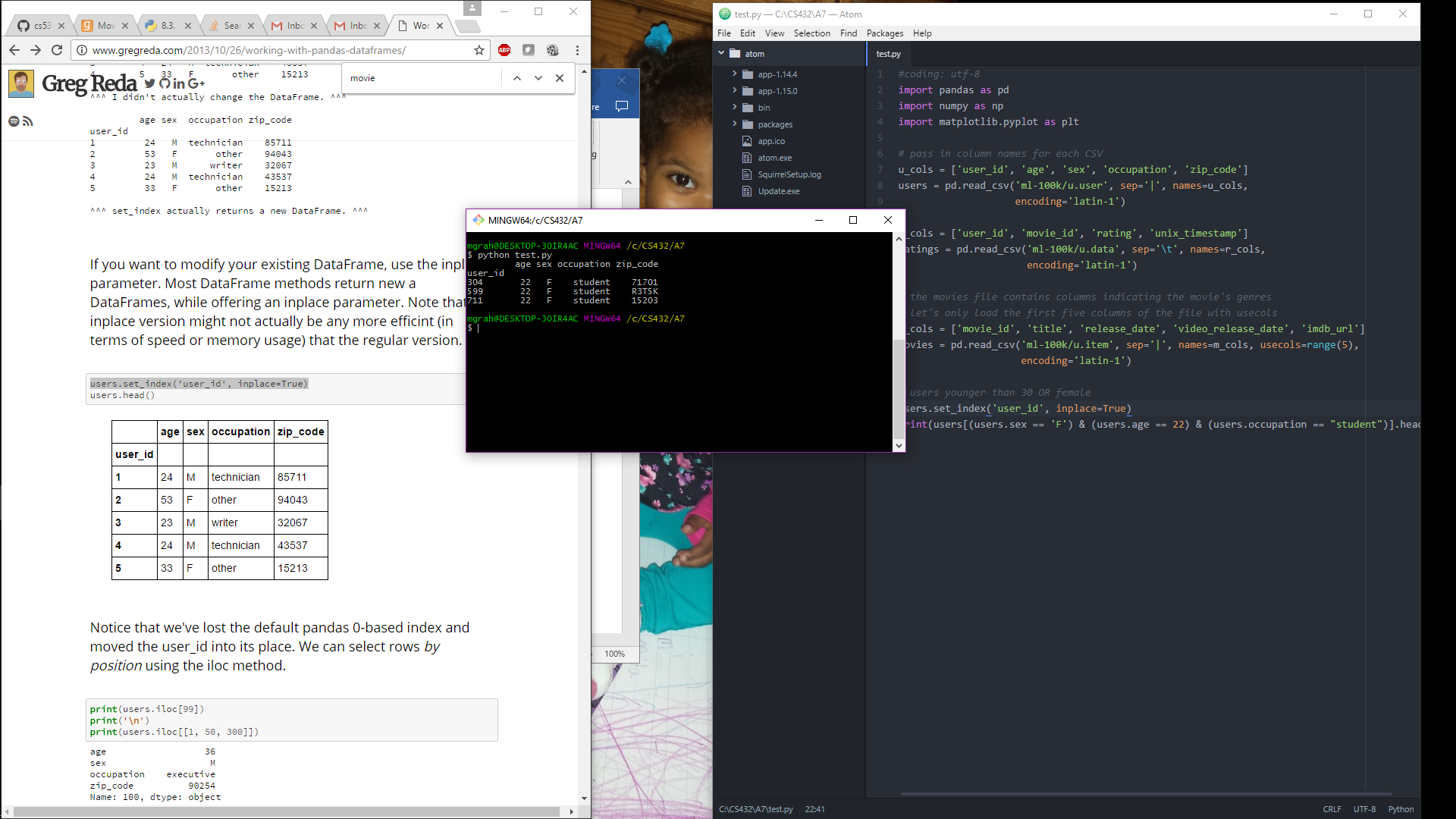
March 29, 2017

Michelle Graham

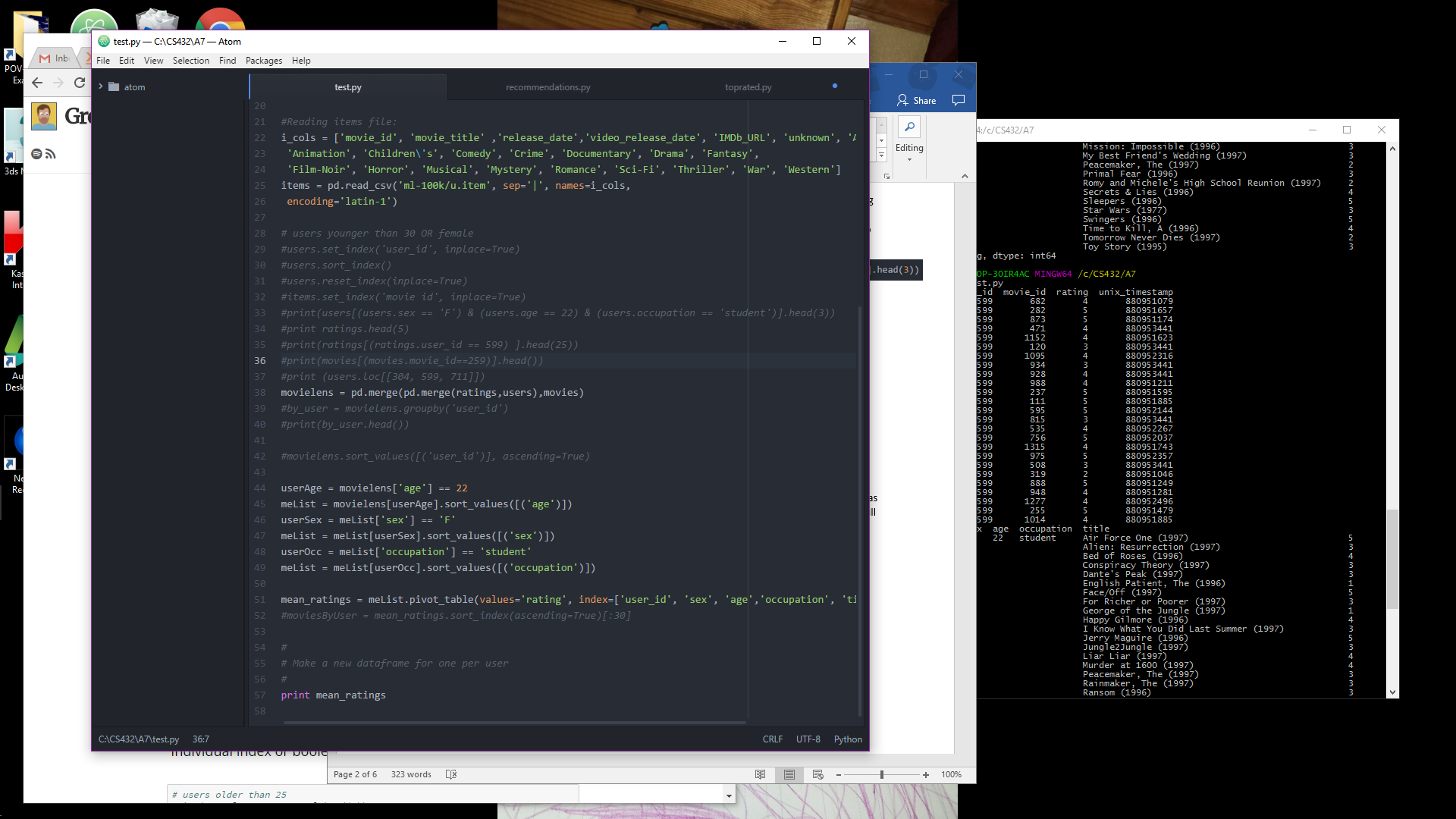
In order to analyze the MovieLens data, I found a very interesting tutorial on data frames using Python’s Pandas. However, isolating dating per user has proven to be quite challenging for me. I was unable to figure out how to successfully create any loops to properly handle the data. Instead, I had to select the desired data by isolating information by column.



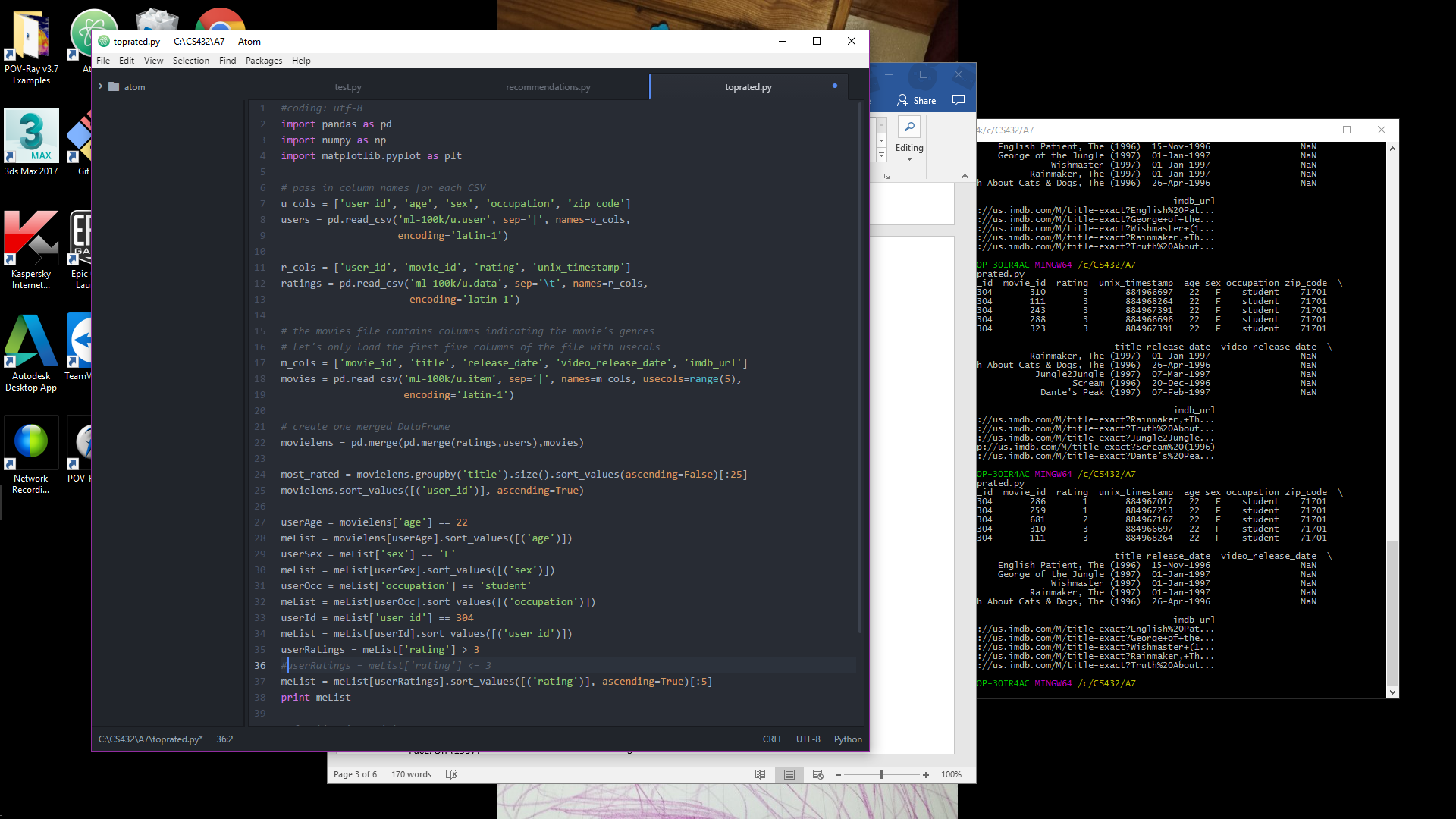
There were only five users in total that are identical to me in age, gender, and occupation! I isolated the first three out of that five to proceed with further analysis.

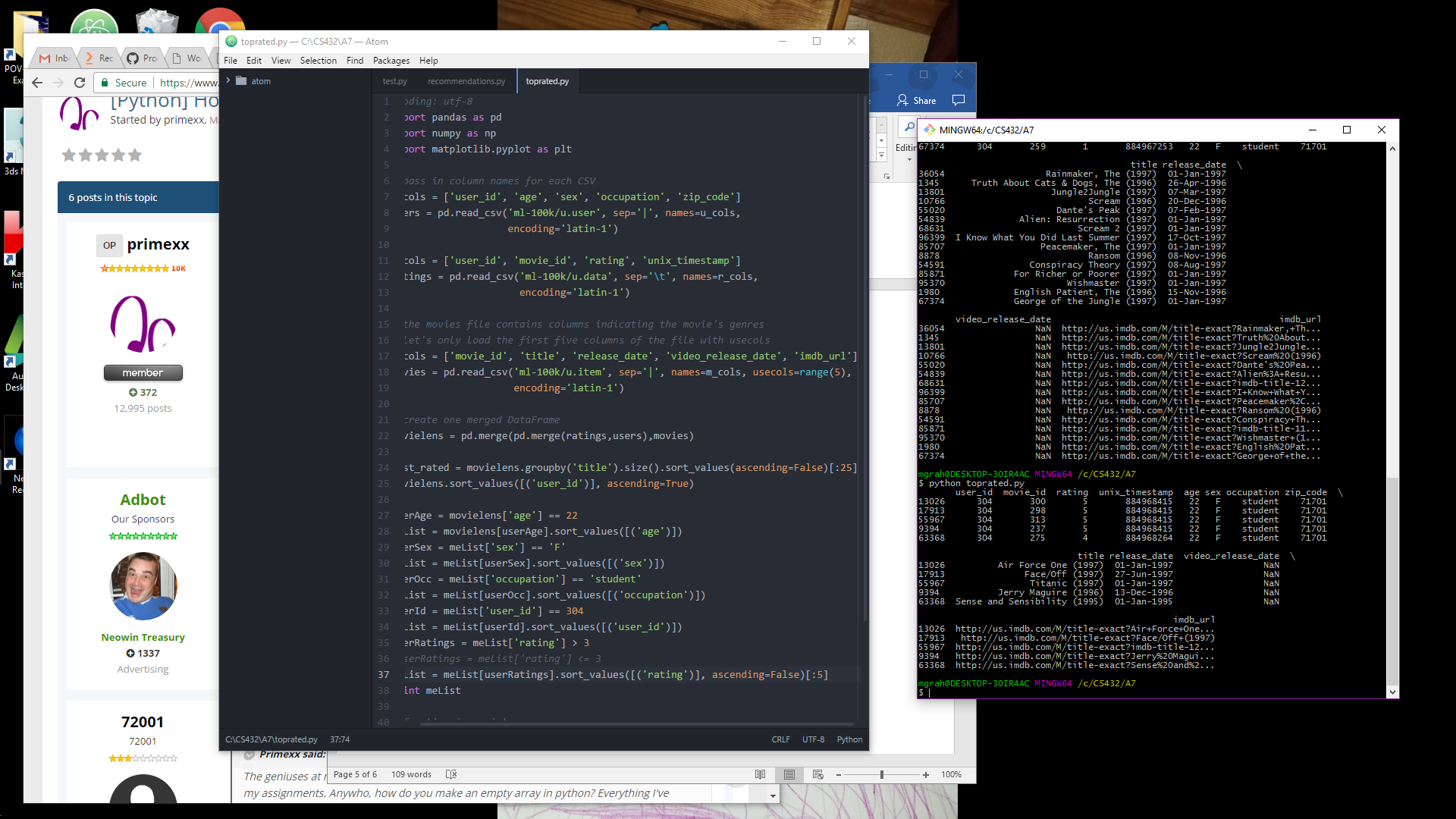


Before I could access how the user rating, I had to merge the data from each of the three files used: u.item, u.data. and u.user. This is possible because the user and data files share the column “user\_id” and the data and time files share the column “movie\_id.”



To access user-specific data, I had to learn how to select the desired information by column. Because I could not figure out how to loop through the data due to various errors, I simply specified each component of the data frame that I wanted to access. For example, I set the “user\_id” to 304 for the first user, because it was one of the three users I previously found that is most like me in age, occupation, and gender. Additionally, I sorted the data so that everything would be easier to read. I was even able to split the data for readability purposes. After showing how this was done for one user, I will reuse the same data with the other two user’s id numbers.





**User 304’s Top 5 Movies:**

Movie Rating

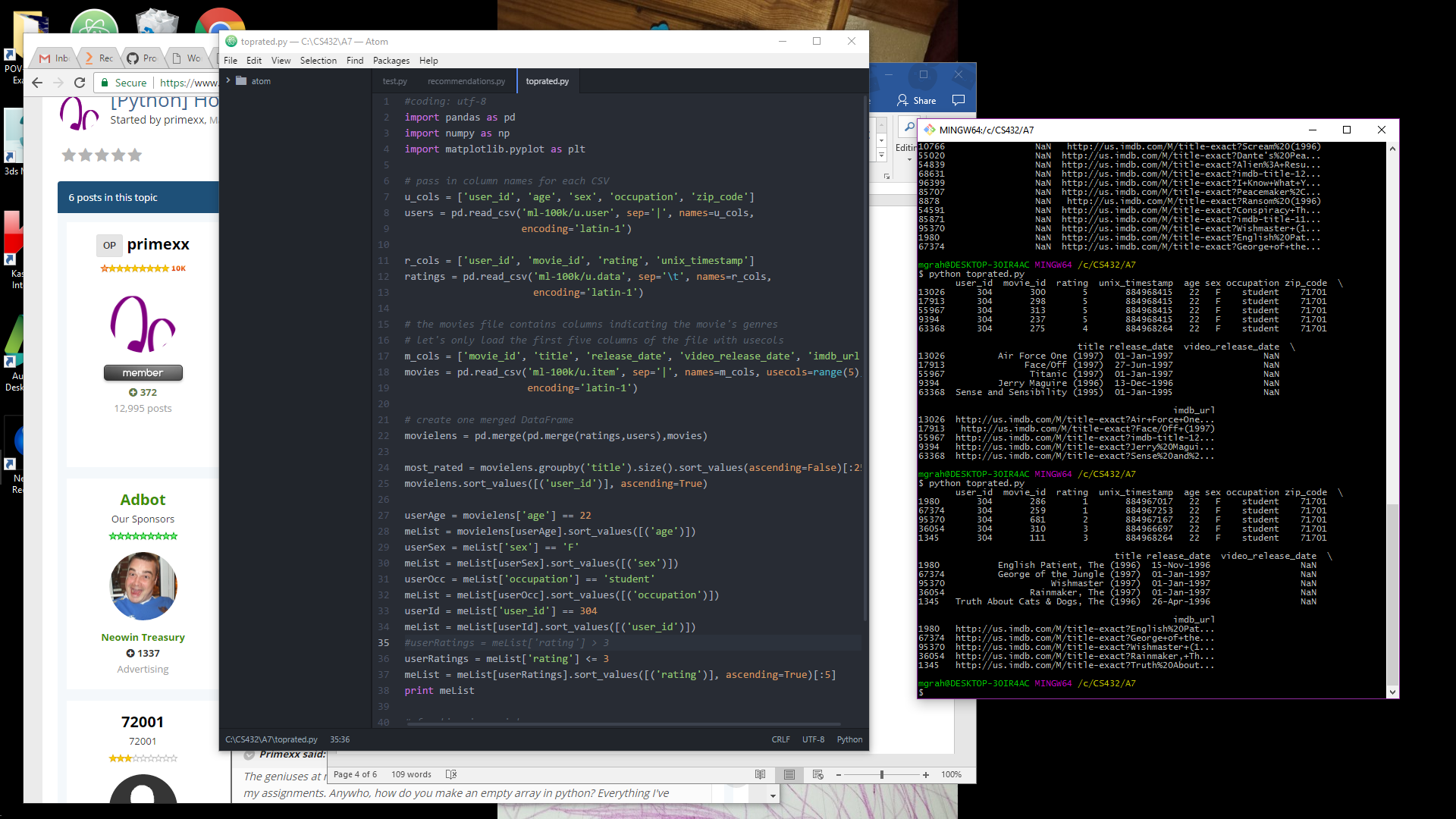
Air Force One (1997) 5

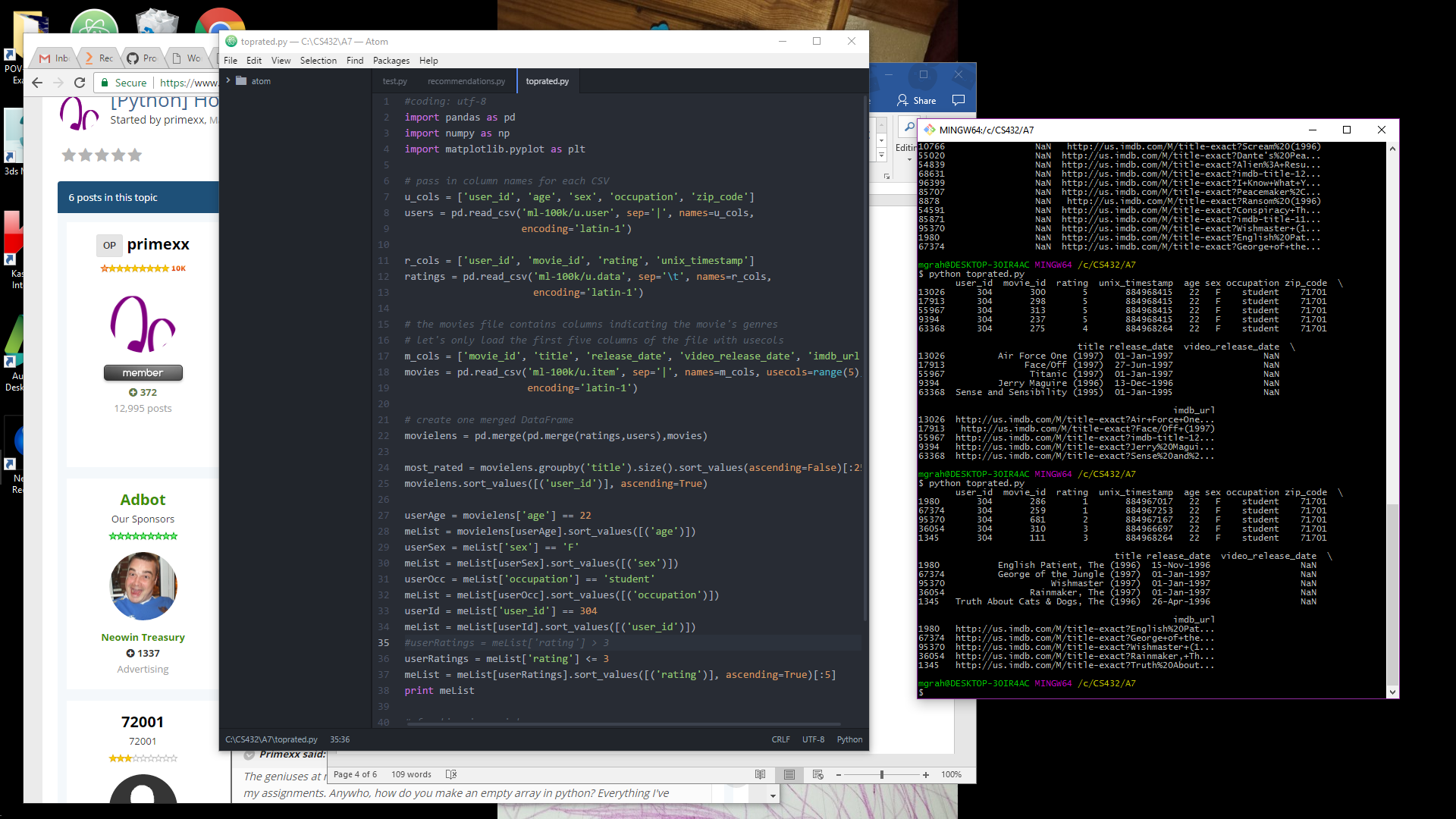
Face/Off (1997) 5

Titanic (1997) 5

Jerry Maguire (1996) 5

Sense and Sensibility (1995) 4





**User 304’s Bottom 5 Movies:**

Movie Rating

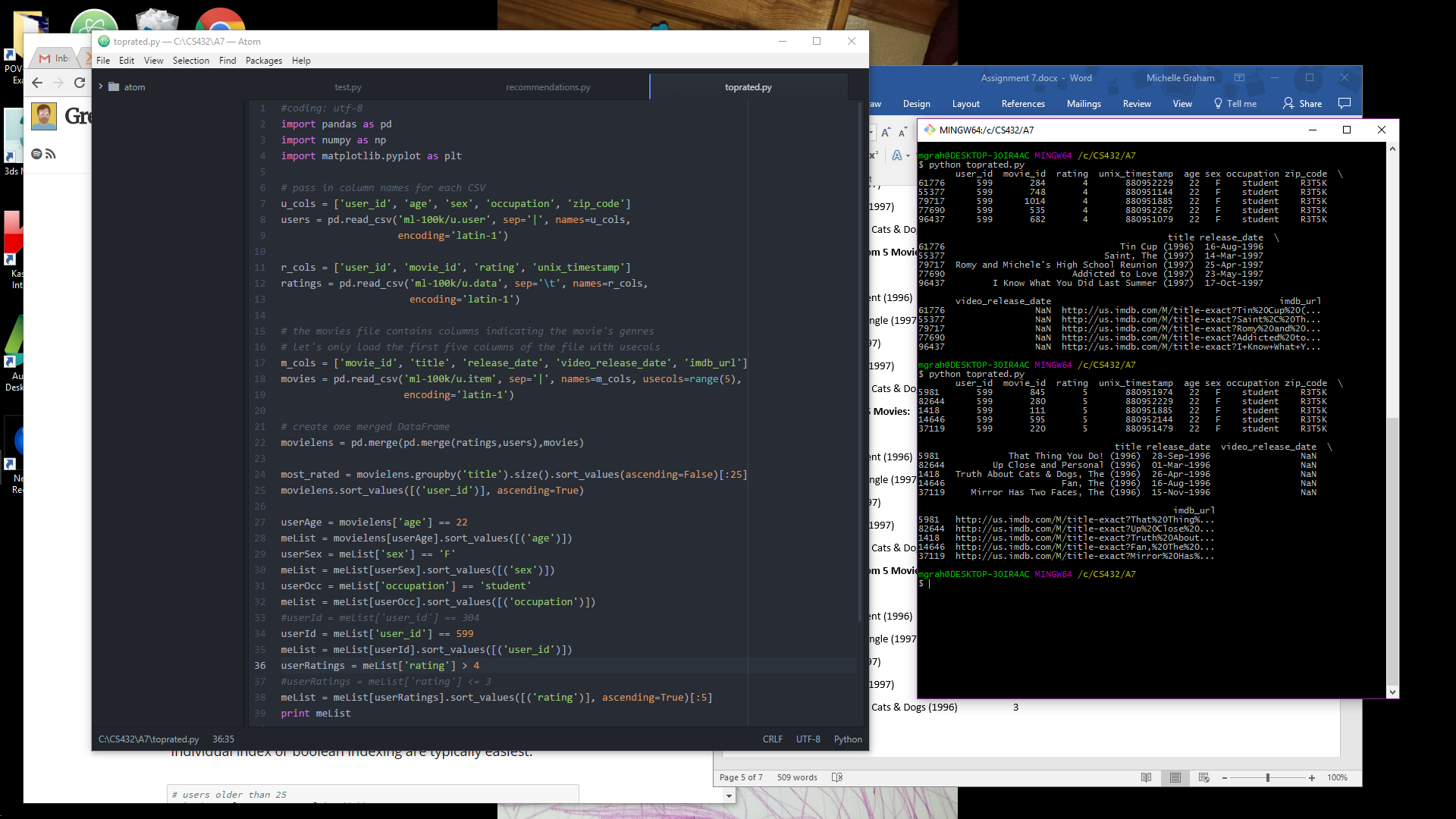
The English Patient (1996) 1

George of the Jungle (1997) 1

Wishmaster (1997) 2

The Rainmaker (1997) 3

The Truth About Cats & Dogs (1996) 3



**User 599’s Top 5 Movies:**

Movie Rating

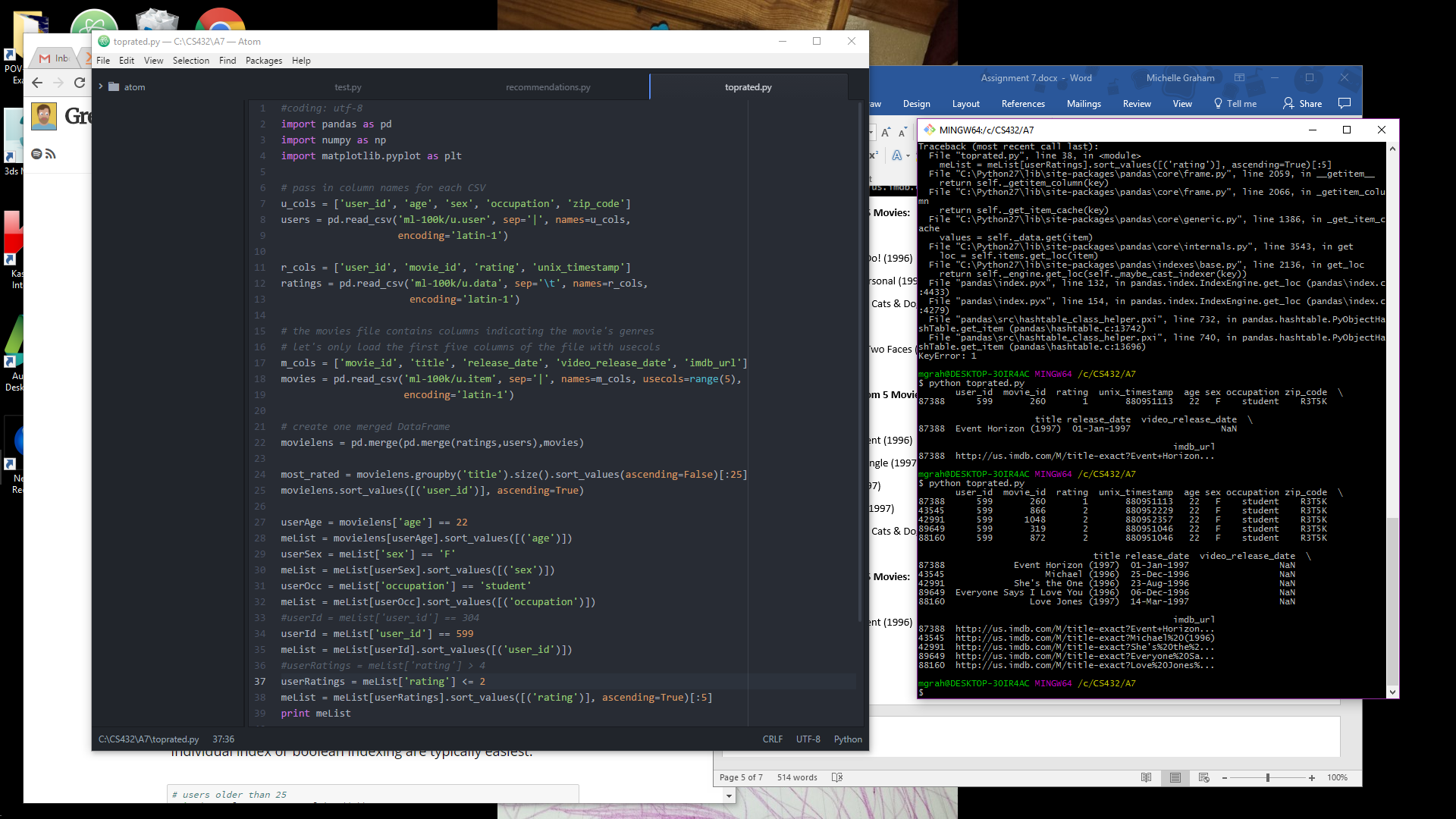
That Thing You Do! (1996) 5

Up Close and Personal (1996) 5

The Truth About Cats & Dogs (1996) 5

The Fan (1996) 5

The Mirror Has Two Faces (1996) 5



**User 599’s Bottom 5 Movies:**

Movie Rating

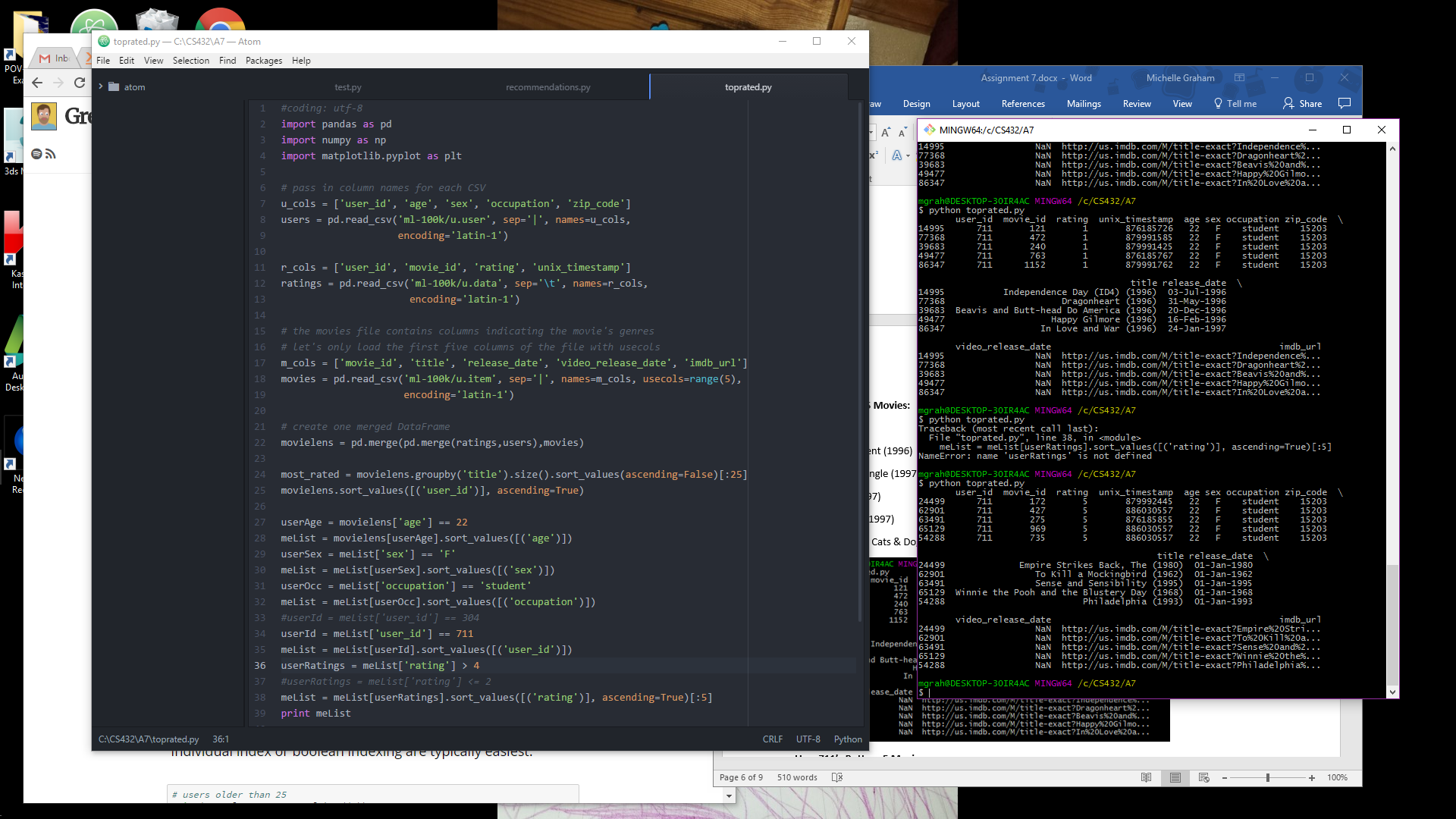
Event Horizon (1997) 1

Michael (1996) 2

She’s the One (1996) 2

Everyone Says I Love You (1996) 2

Love Jones (1997) 2



**User 711’s Top 5 Movies:**

Movie Rating

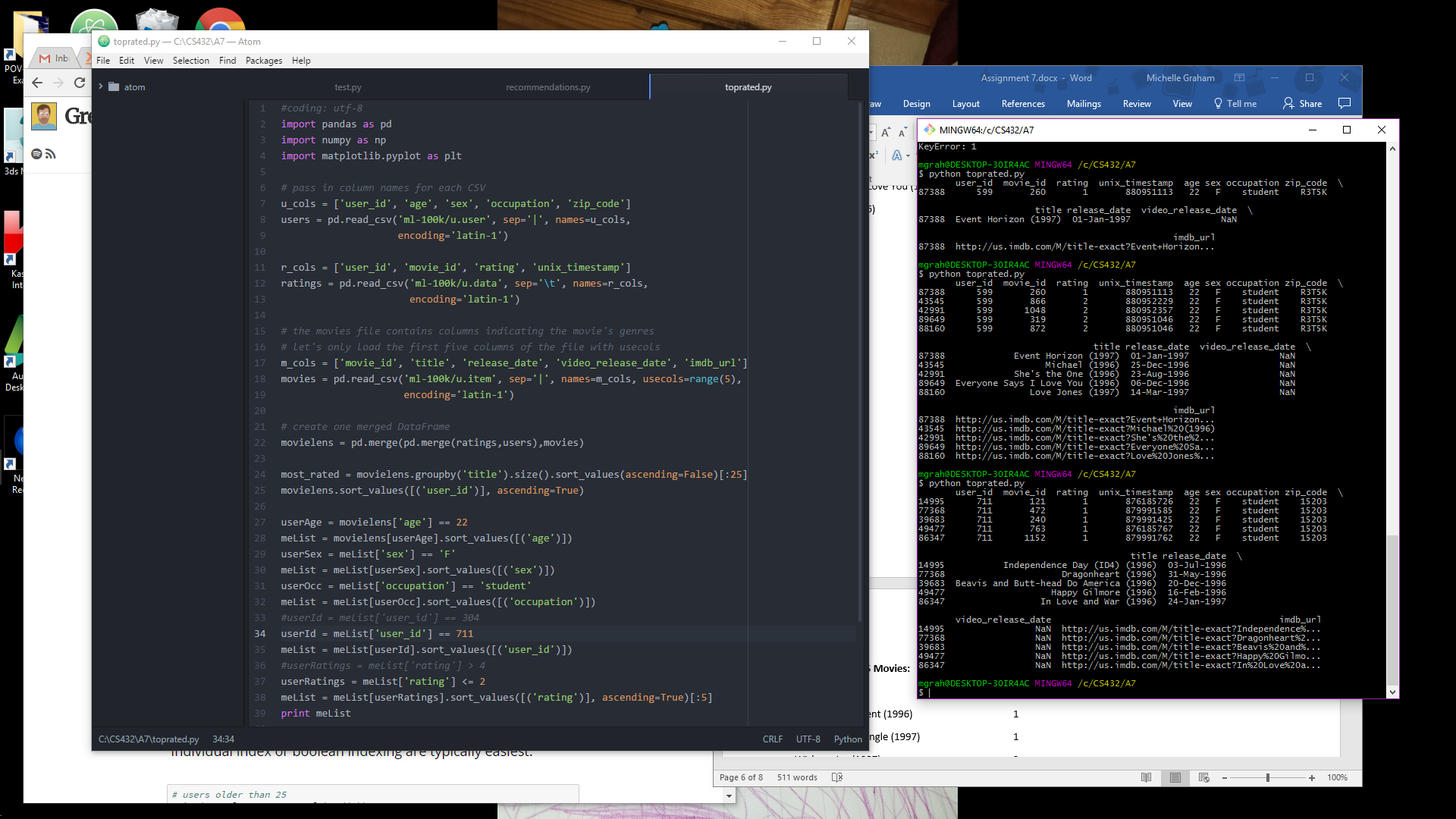
The Empire Strikes Back (1980) 5

To Kill a Mockingbird (1962) 5

Sense and Sensibility (1995) 5

Winnie the Pooh and the Blustery Day (1968) 5

Philadelphia (1993) 5



**User 711’s Bottom 5 Movies:**

Movie Rating

Independence Day (ID4) (1996) 1

Dragonheart (1996) 1

Beavis and Butt-head Do America (1996) 1

Happy Gilmore (1996) 1

In Love and War (1996) 1

I scrolled through the “u.item” data file in order to pick out movies I have seen. Below, I have listed my own most and least favorite movies in order to compare my own interests to that of my substitute user. Based on my preferences, I am most closely related to user number 711.

**My Top 5 Movies:**

Movie Rating

Pulp Fiction 5

3 Ninjas: High Noon At Mega Mountain (1998) 5

The Empire Strikes Back (1980) 5

Powder (1995) 5

The Nightmare Before Christmas (1993) 5

**My Bottom 5 Movies:**

Movie Rating

Beavis and Butt-head Do America (1996) 1

Gone with the Wind (1939) 1

Old Yeller (1957) 1

Coneheads (1993) 2

William Shakespeare's Romeo and Juliet(1996) 2

Resources:

<https://github.com/arthur-e/Programming-Collective-Intelligence/blob/master/chapter2/recommendations.py>

<http://www.gregreda.com/2013/10/26/working-with-pandas-dataframes/>

<http://www.michaeltsmith.org.uk/ipython/Lesson%201,%20(detour)%20Python,%20pandas%20and%20Movie%20Ratings%20(MovieLens).html>

<https://www.packtpub.com/books/content/recommending-movies-scale-python>