Practical CS5052
Movie Recommender System

I have implemented all three parts and this report will summarize my effort.

Summary Part 1:

In part 1, I have implemented search by user id, search by movie ID or title and search by genre.

For user Id search I have given 7 options,

1. Show titles: shows the titles of the user ID entered.
2. Movies Watched: shows the number of distinct movies watched by a user.
3. Genres Watched: shows the number of genres watched by the user.
4. Genre ratings: shows the average ratings of genres, given by the user.
5. Genre watch count shows the number of movies from each genre watched by the user.
6. Favourite genre based on rating: This gives the favourite genre of the user based on his rating.
7. Favourite genre based on watch count: This gives the favourite genre of the user based on his watch count.

For movie ID and movie name search, I have given 2 options:

1. Watch count: gives the number of times the movie was watched.
2. Rating: gives the average rating of the movie, given by the users.

To give the program user the right movie, I have added the data that matches the letters entered e.g if someone enters “men” or “Men” or any such combination, he will get a list of movies that start with the letters “MEN”.

For Genre search, I have given the option to see all the movies of that genre. Again, I have given the added ability to search if one enters “rom”, he will get romance movies.

I have added the Year search, which gives all the movies from that year.

Lastly for part 1, I have added Top N rated and Top N watched movies, which gives out ‘n’, number entered by user, highest rated or most watched movies.

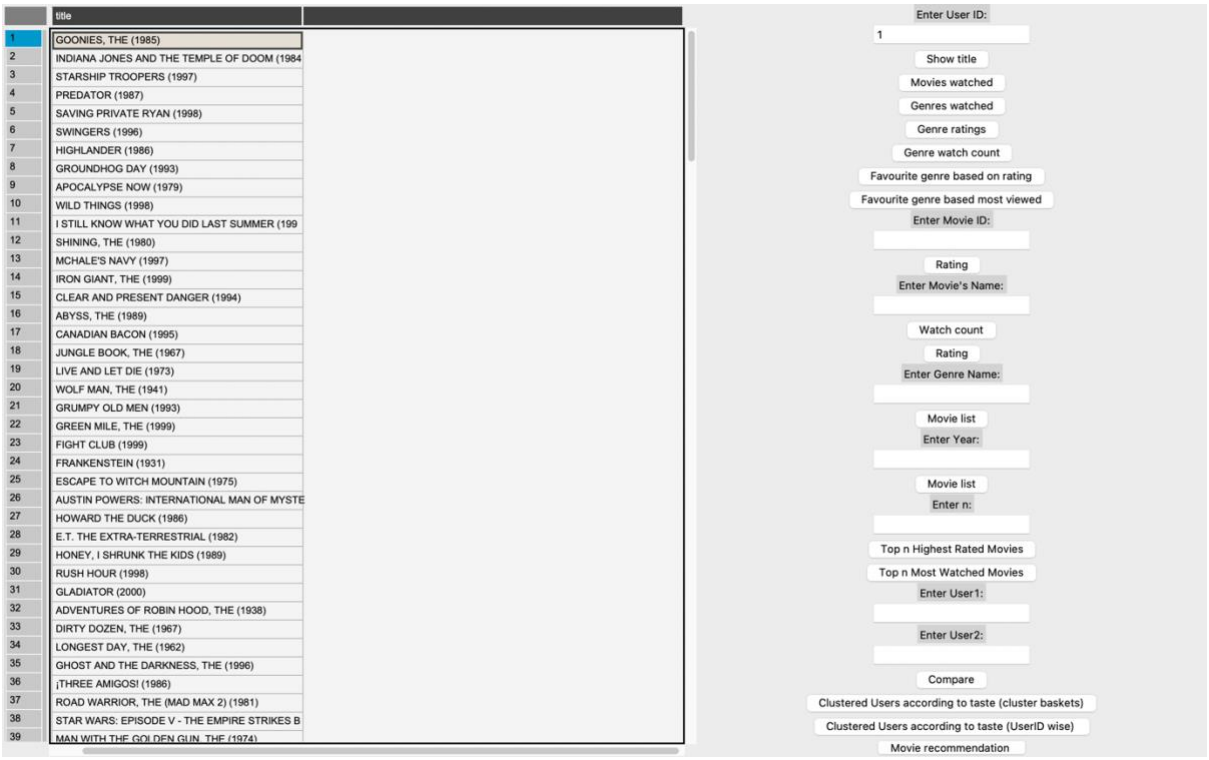


Figure 1 user id title list search

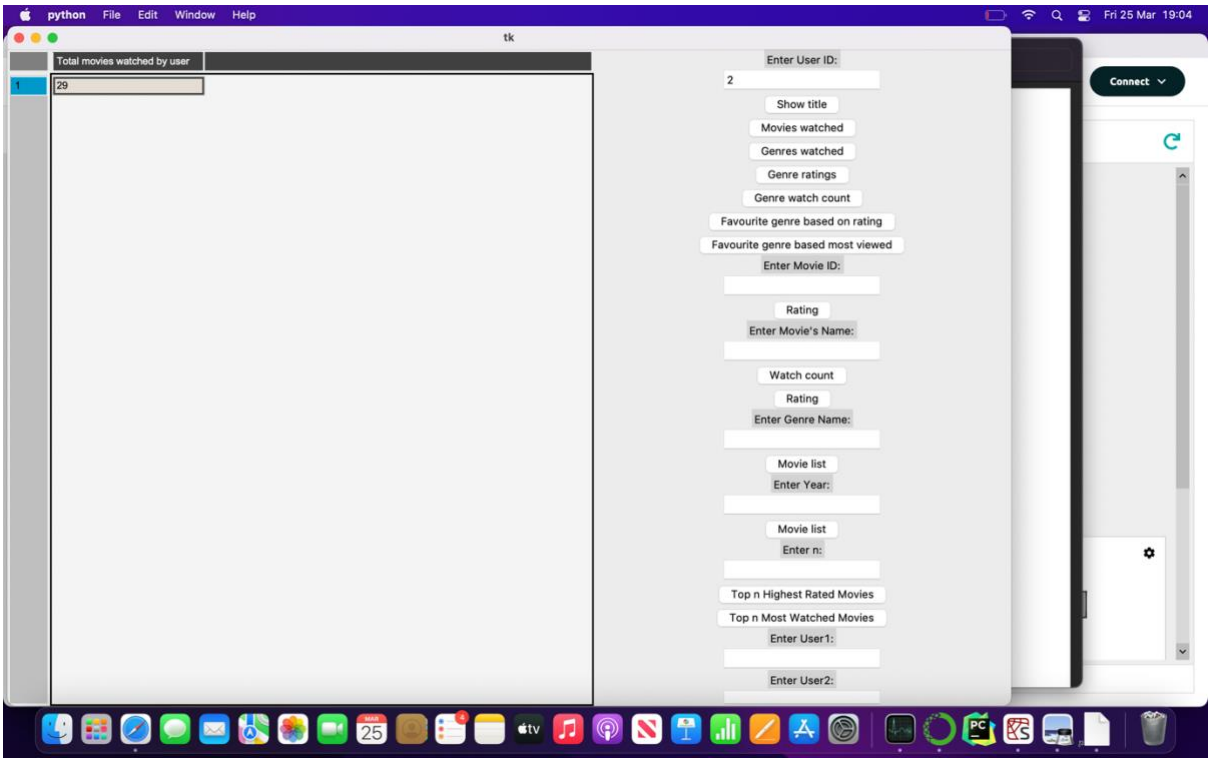
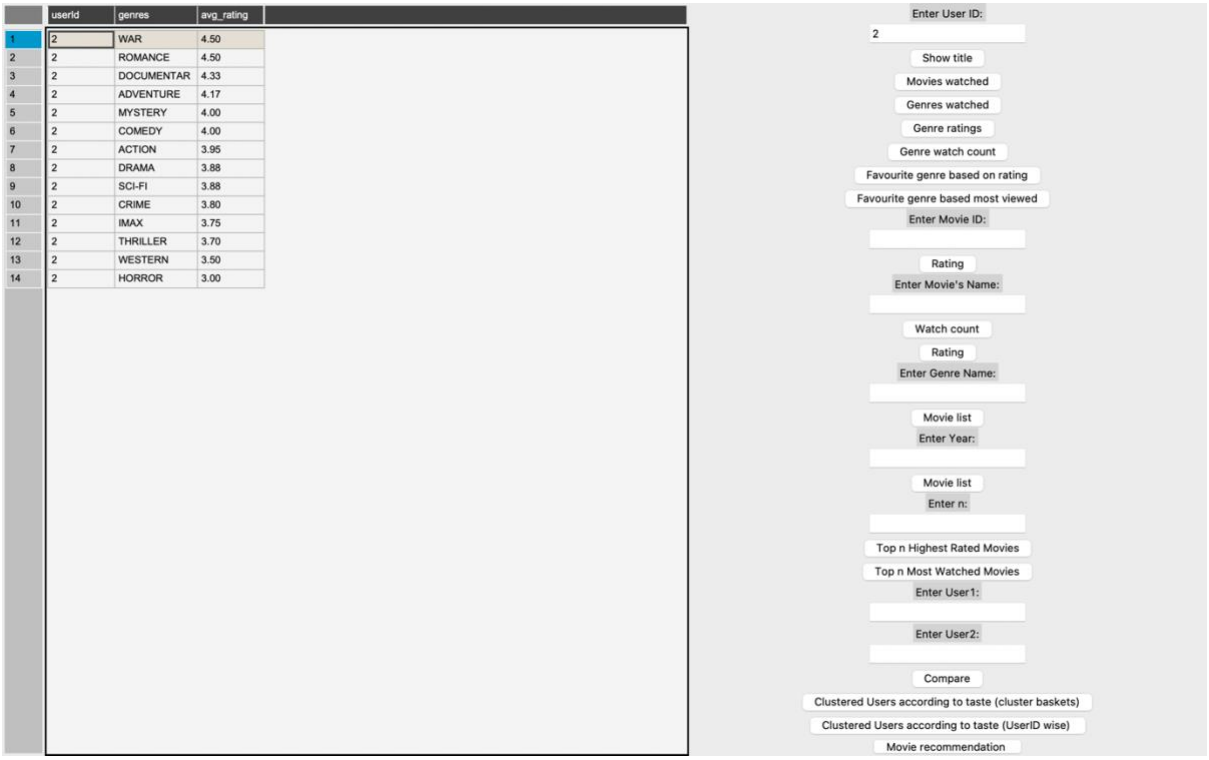


Figure 2 user id total movies watched count

	Total genres
1	14

Enter User ID:
2
Show title
Movies watched
Genres watched
Genre ratings
Genre watch count
Favourite genre based on rating
Favourite genre based most viewed
Enter Movie ID:
Rating
Enter Movie's Name:
Watch count
Rating
Enter Genre Name:
Movie list
Enter Year:
Movie list
Enter n:
Top n Highest Rated Movies
Top n Most Watched Movies
Enter User1:
Enter User2:
Compare
Clustered Users according to taste (cluster baskets)
Clustered Users according to taste (UserID wise)
Movie recommendation

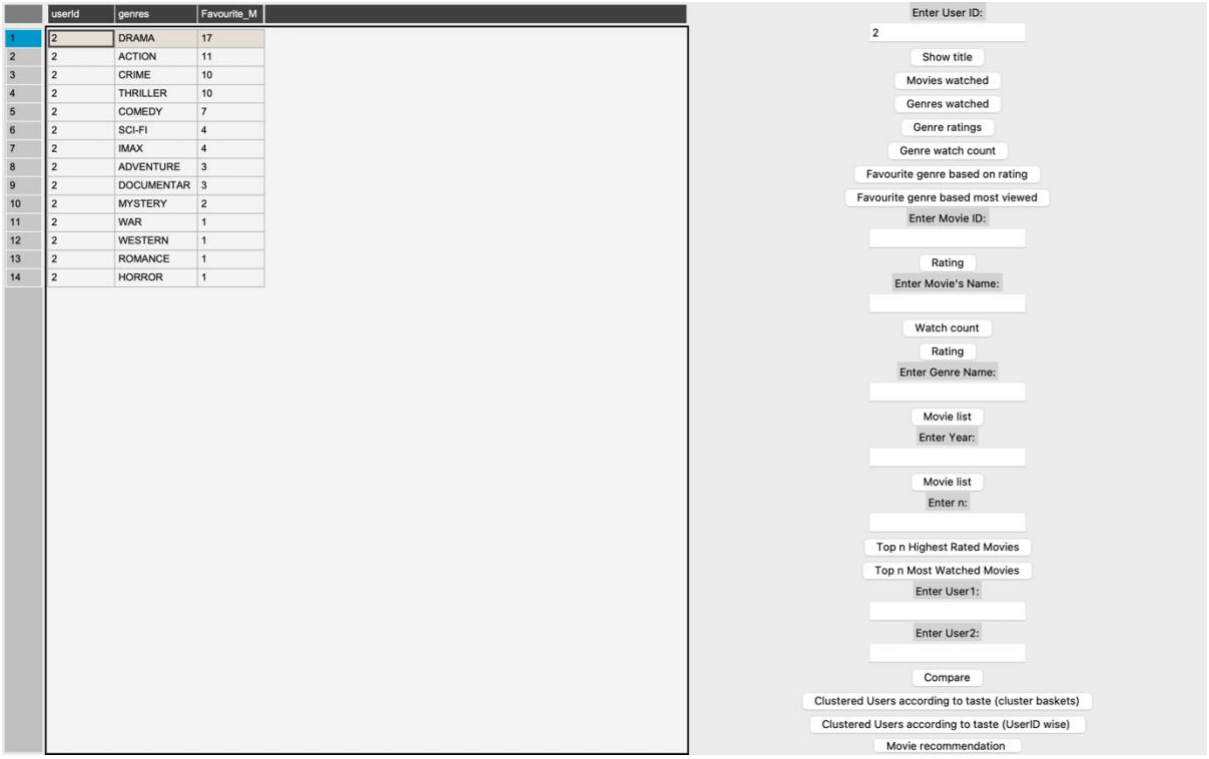
Figure 3 user id total genres watched



	userId	genres	avg_rating
1	2	WAR	4.50
2	2	ROMANCE	4.50
3	2	DOCUMENTAR	4.33
4	2	ADVENTURE	4.17
5	2	MYSTERY	4.00
6	2	COMEDY	4.00
7	2	ACTION	3.95
8	2	DRAMA	3.88
9	2	SCI-FI	3.88
10	2	CRIME	3.80
11	2	IMAX	3.75
12	2	THRILLER	3.70
13	2	WESTERN	3.50
14	2	HORROR	3.00

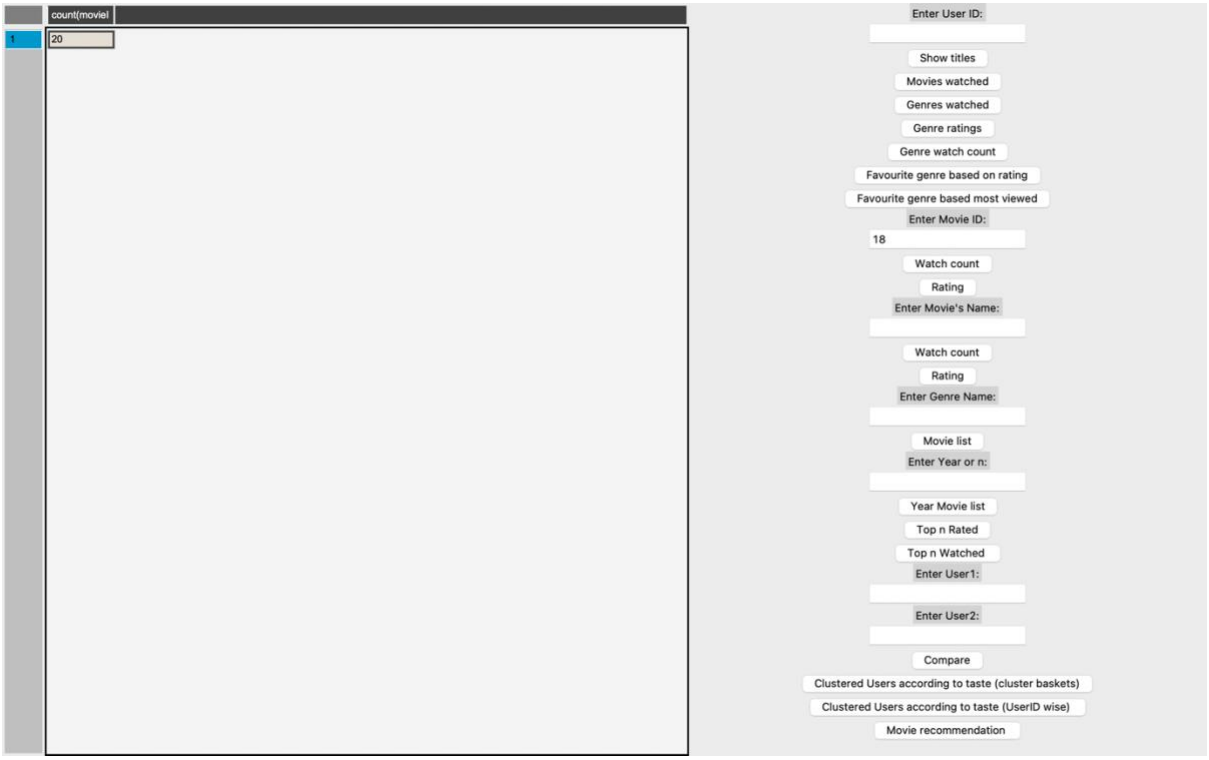
Enter User ID:
2
Show title
Movies watched
Genres watched
Genre ratings
Genre watch count
Favourite genre based on rating
Favourite genre based most viewed
Enter Movie ID:
Rating
Enter Movie's Name:
Watch count
Rating
Enter Genre Name:
Movie list
Enter Year:
Movie list
Enter n:
Top n Highest Rated Movies
Top n Most Watched Movies
Enter User1:
Enter User2:
Compare
Clustered Users according to taste (cluster baskets)
Clustered Users according to taste (UserID wise)
Movie recommendation

Figure 4 user id genre list sorted by highest ratings



	userId	genres	Favourite_M
1	2	DRAMA	17
2	2	ACTION	11
3	2	CRIME	10
4	2	THRILLER	10
5	2	COMEDY	7
6	2	SCI-FI	4
7	2	IMAX	4
8	2	ADVENTURE	3
9	2	DOCUMENTAR	3
10	2	MYSTERY	2
11	2	WAR	1
12	2	WESTERN	1
13	2	ROMANCE	1
14	2	HORROR	1

Figure 5 user id genre list sorted by most watched



	count(movieI
1	20

Figure 6 movie ID watch count

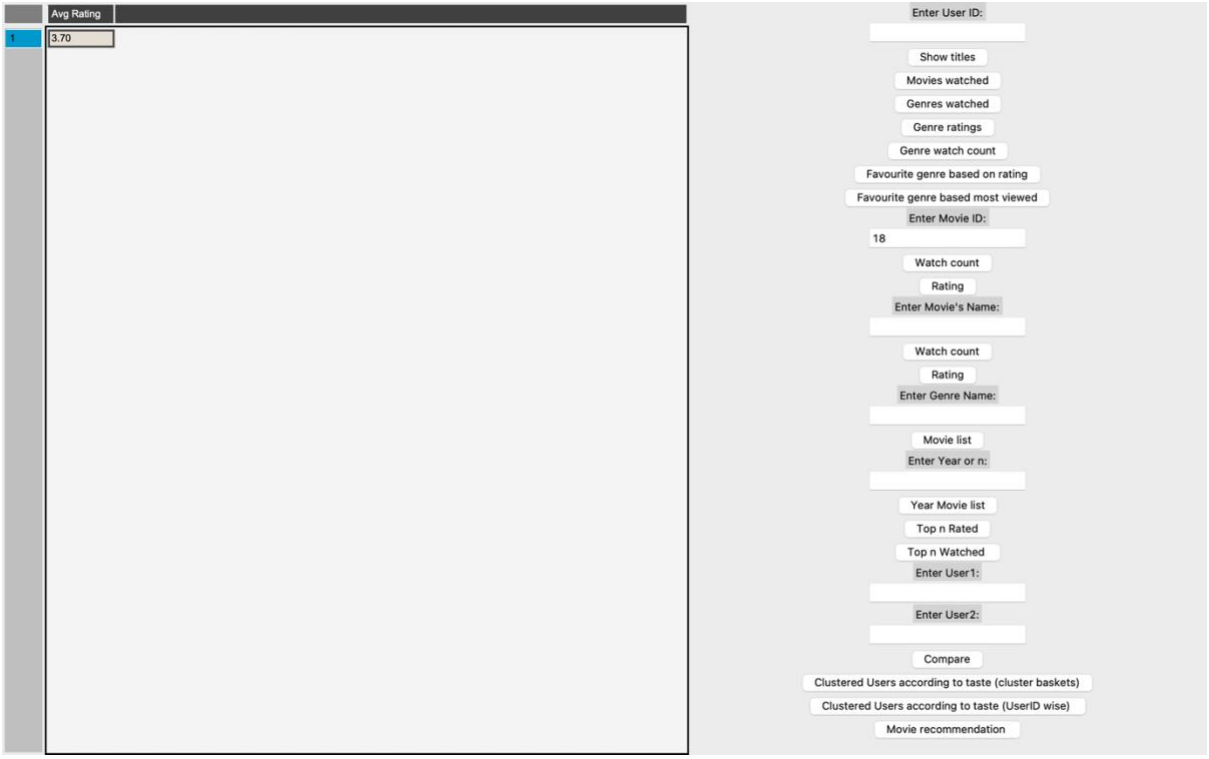


Figure 7 movie Id rating

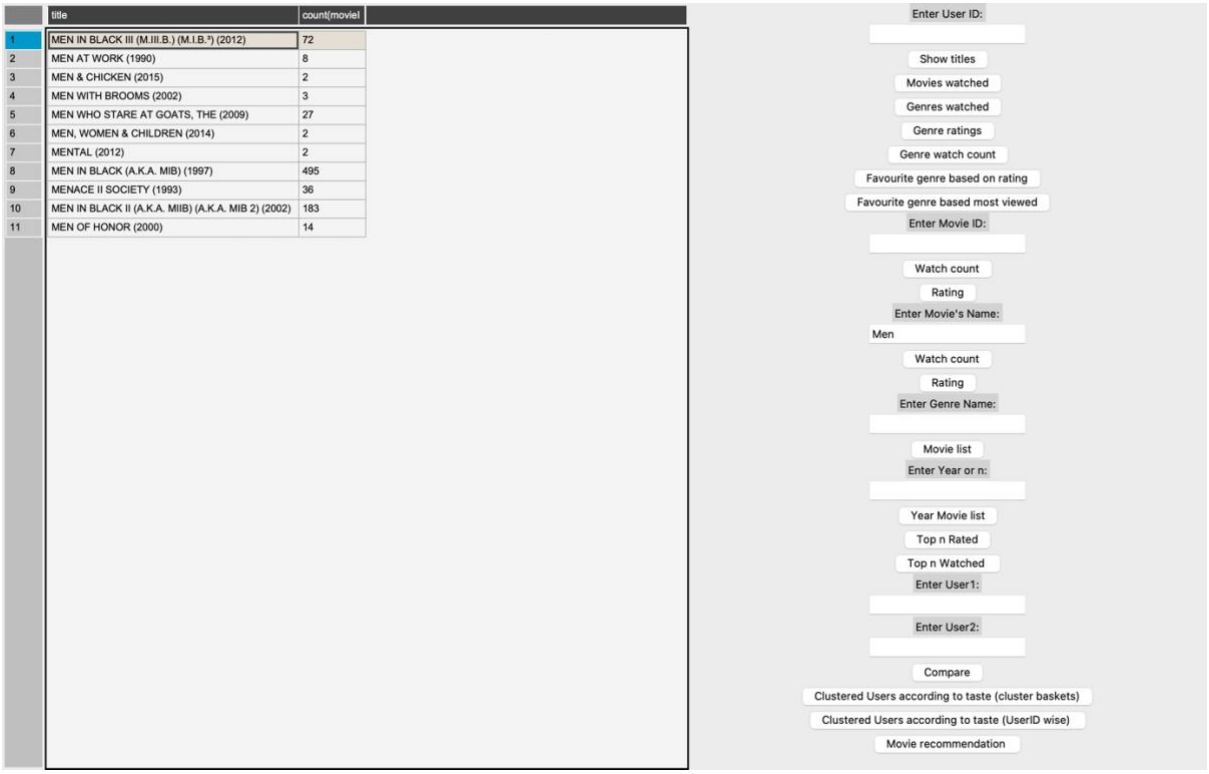
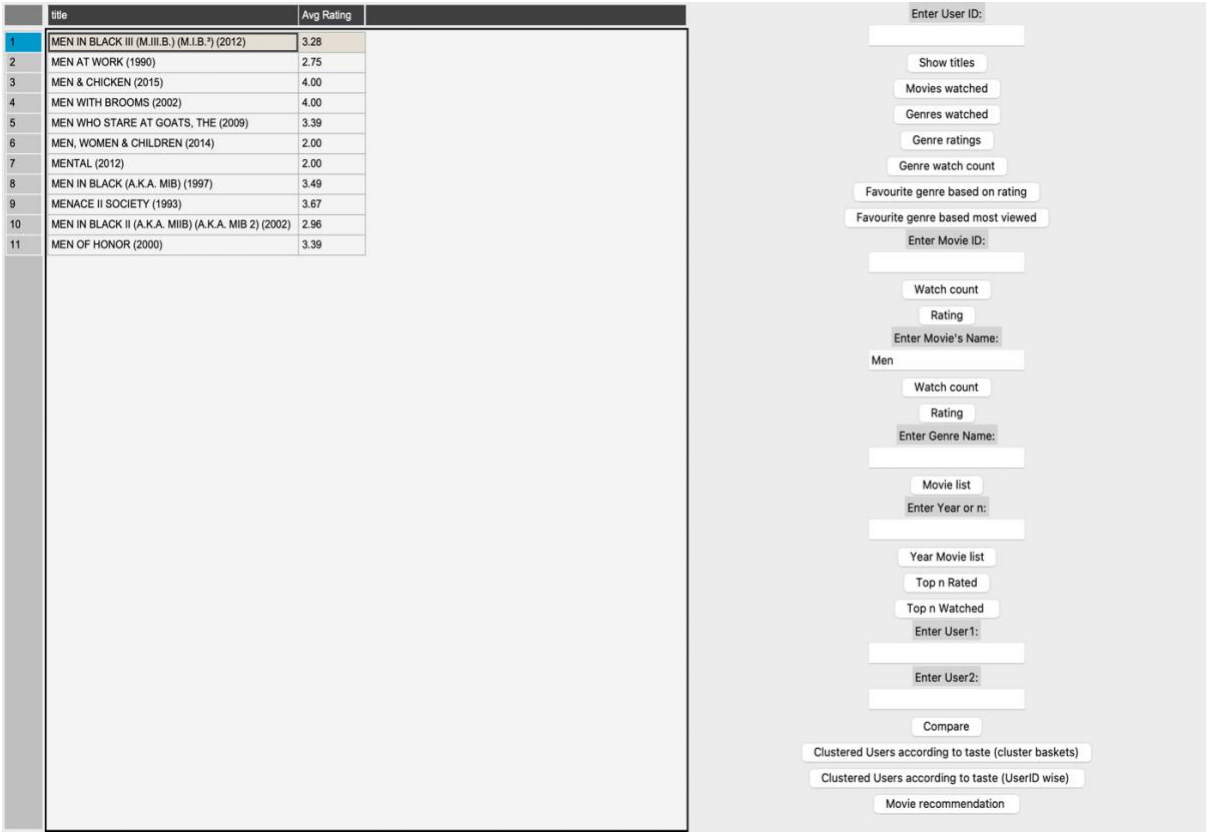
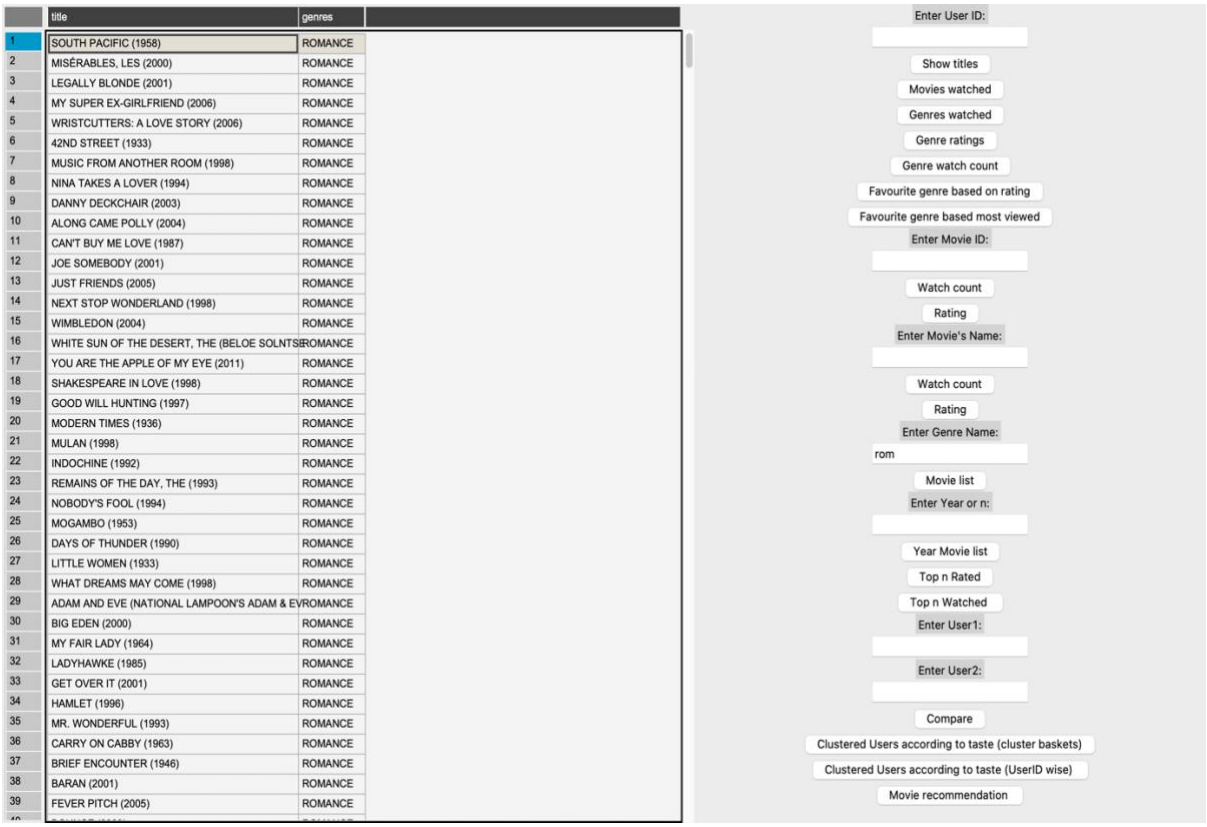


Figure 8 movie name watch count



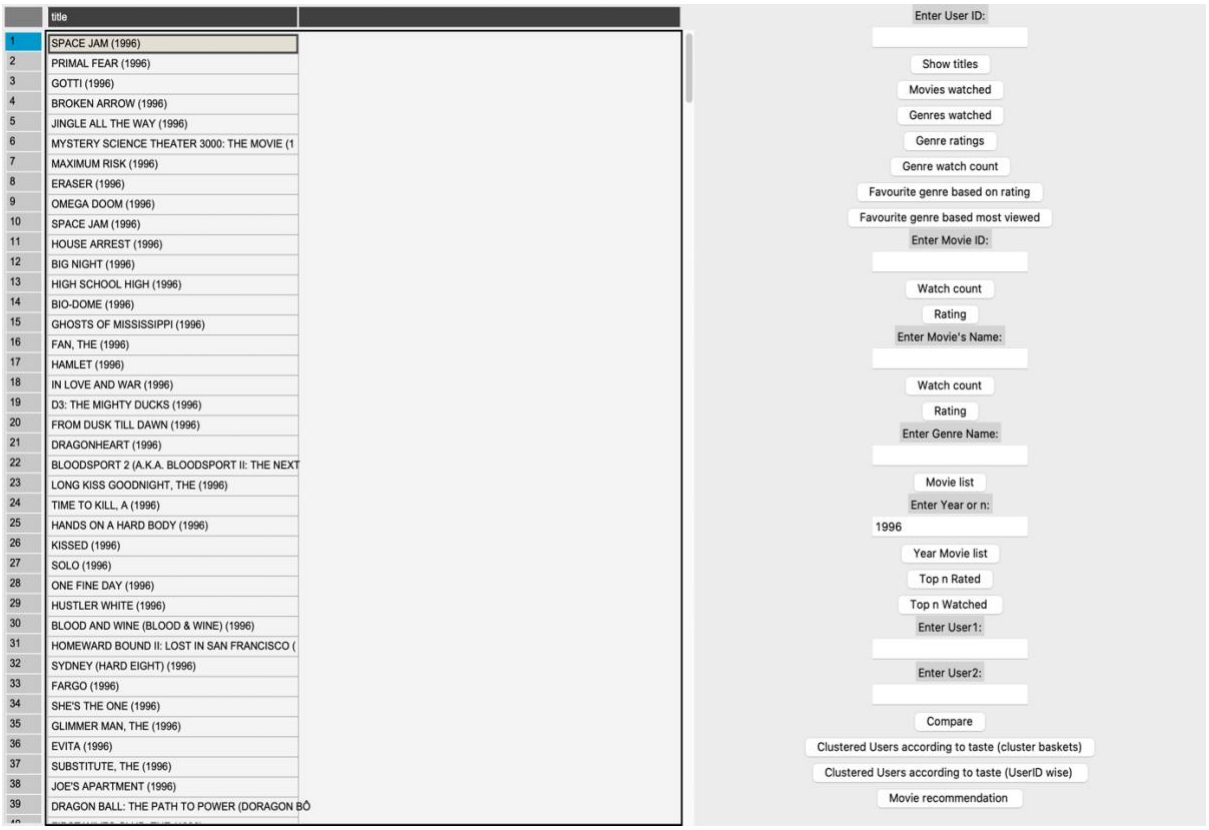
	title	Avg Rating
1	MEN IN BLACK III (M.III.B.) (M.I.B.³) (2012)	3.28
2	MEN AT WORK (1990)	2.75
3	MEN & CHICKEN (2015)	4.00
4	MEN WITH BROOMS (2002)	4.00
5	MEN WHO STARE AT GOATS, THE (2009)	3.39
6	MEN, WOMEN & CHILDREN (2014)	2.00
7	MENTAL (2012)	2.00
8	MEN IN BLACK (A.K.A. MIB) (1997)	3.49
9	MENACE II SOCIETY (1993)	3.67
10	MEN IN BLACK II (A.K.A. MIIB) (A.K.A. MIB 2) (2002)	2.96
11	MEN OF HONOR (2000)	3.39

Figure 9 movie name rating



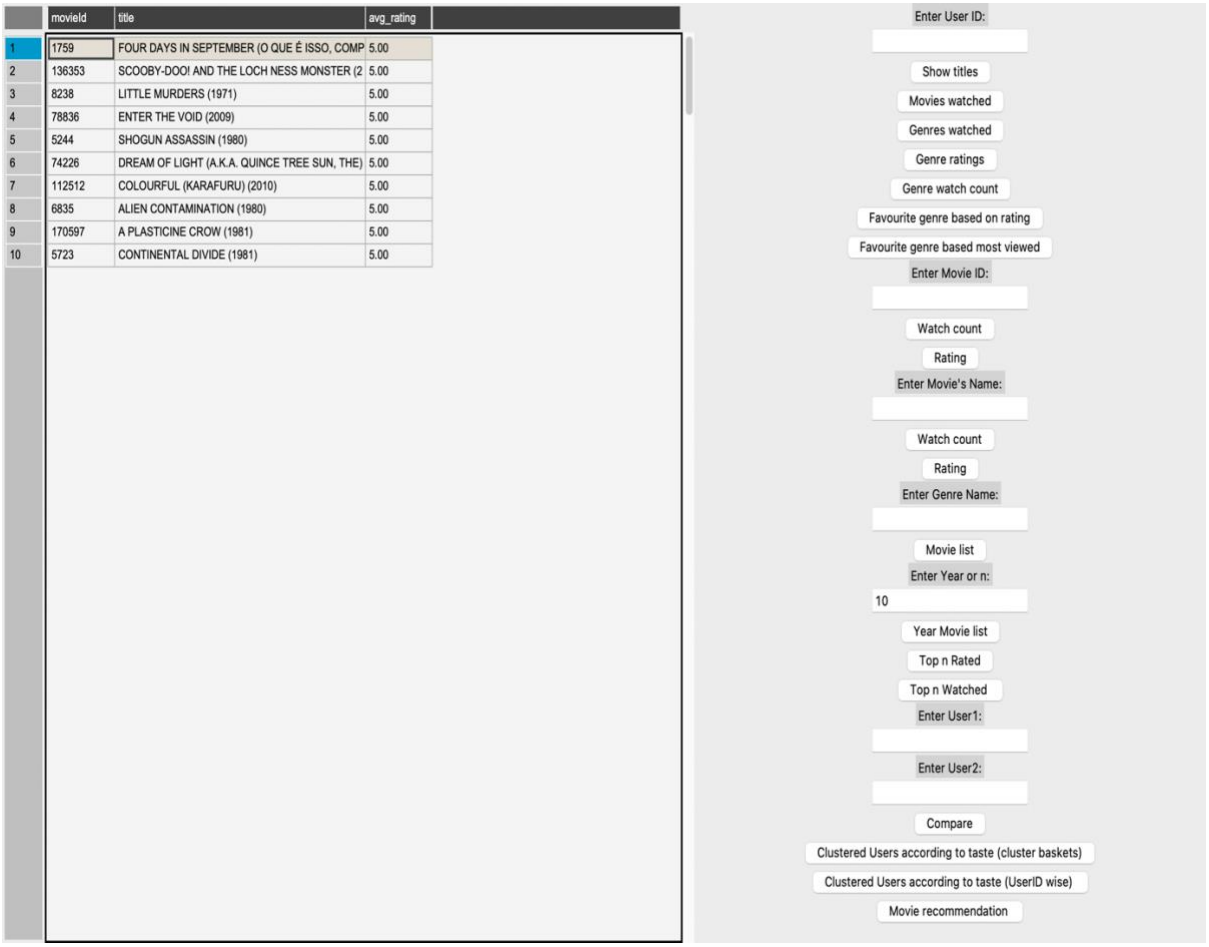
	title	genres
1	SOUTH PACIFIC (1958)	ROMANCE
2	MISÉRABLES, LES (2000)	ROMANCE
3	LEGALLY BLONDE (2001)	ROMANCE
4	MY SUPER EX-GIRLFRIEND (2006)	ROMANCE
5	WRISTCUTTERS: A LOVE STORY (2006)	ROMANCE
6	42ND STREET (1933)	ROMANCE
7	MUSIC FROM ANOTHER ROOM (1998)	ROMANCE
8	NINA TAKES A LOVER (1994)	ROMANCE
9	DANNY DECKCHAIR (2003)	ROMANCE
10	ALONG CAME POLLY (2004)	ROMANCE
11	CAN'T BUY ME LOVE (1987)	ROMANCE
12	JOE SOMEBODY (2001)	ROMANCE
13	JUST FRIENDS (2005)	ROMANCE
14	NEXT STOP WONDERLAND (1998)	ROMANCE
15	WIMBLEDON (2004)	ROMANCE
16	WHITE SUN OF THE DESERT, THE (BELOE SOLNTSE	ROMANCE
17	YOU ARE THE APPLE OF MY EYE (2011)	ROMANCE
18	SHAKESPEARE IN LOVE (1998)	ROMANCE
19	GOOD WILL HUNTING (1997)	ROMANCE
20	MODERN TIMES (1936)	ROMANCE
21	MULAN (1998)	ROMANCE
22	INDOCHINE (1992)	ROMANCE
23	REMAINS OF THE DAY, THE (1993)	ROMANCE
24	NOBODY'S FOOL (1994)	ROMANCE
25	MOGAMBO (1953)	ROMANCE
26	DAYS OF THUNDER (1990)	ROMANCE
27	LITTLE WOMEN (1933)	ROMANCE
28	WHAT DREAMS MAY COME (1998)	ROMANCE
29	ADAM AND EVE (NATIONAL LAMPOON'S ADAM & EV	ROMANCE
30	BIG EDEN (2000)	ROMANCE
31	MY FAIR LADY (1964)	ROMANCE
32	LADYHAWKE (1985)	ROMANCE
33	GET OVER IT (2001)	ROMANCE
34	HAMLET (1996)	ROMANCE
35	MR. WONDERFUL (1993)	ROMANCE
36	CARRY ON CABBY (1963)	ROMANCE
37	BRIEF ENCOUNTER (1946)	ROMANCE
38	BARAN (2001)	ROMANCE
39	FEVER PITCH (2005)	ROMANCE

Figure 10 genre movie list search



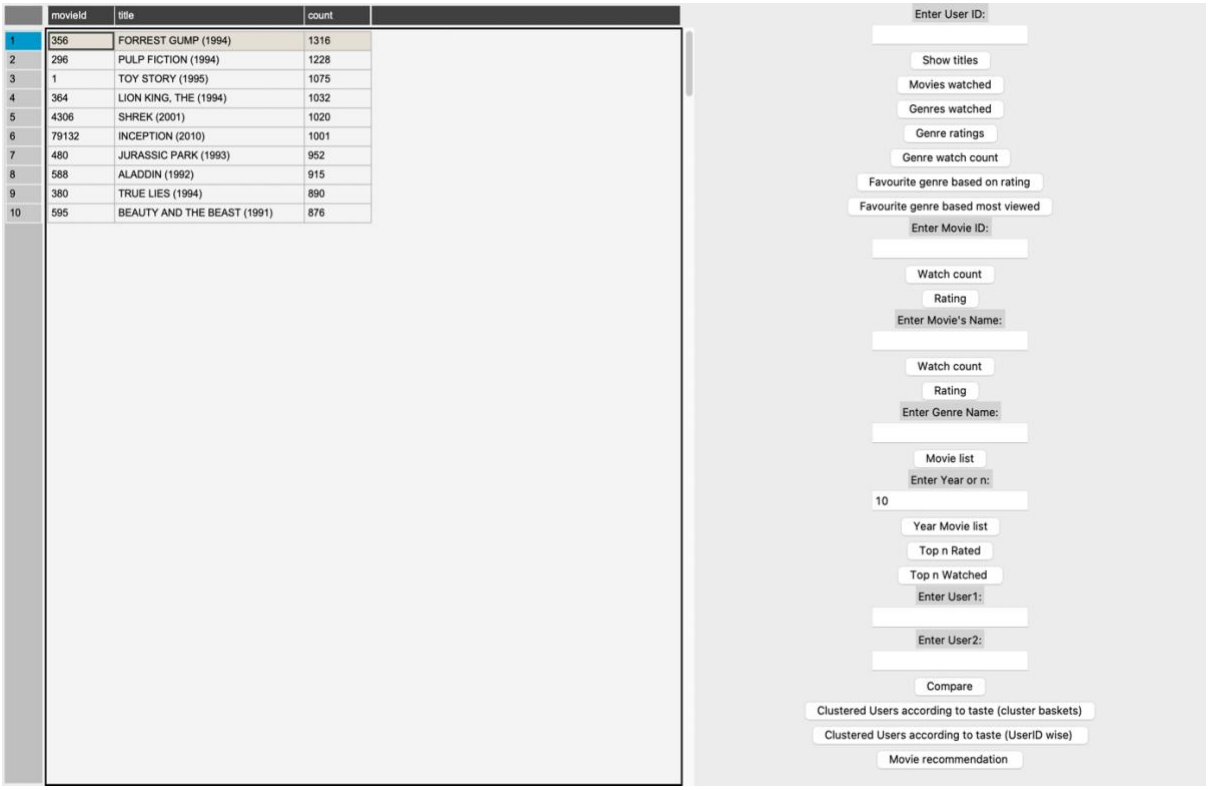
	title
1	SPACE JAM (1996)
2	PRIMAL FEAR (1996)
3	GOTTI (1996)
4	BROKEN ARROW (1996)
5	JINGLE ALL THE WAY (1996)
6	MYSTERY SCIENCE THEATER 3000: THE MOVIE (1
7	MAXIMUM RISK (1996)
8	ERASER (1996)
9	OMEGA DOOM (1996)
10	SPACE JAM (1996)
11	HOUSE ARREST (1996)
12	BIG NIGHT (1996)
13	HIGH SCHOOL HIGH (1996)
14	BIO-DOME (1996)
15	GHOSTS OF MISSISSIPPI (1996)
16	FAN, THE (1996)
17	HAMLET (1996)
18	IN LOVE AND WAR (1996)
19	D3: THE MIGHTY DUCKS (1996)
20	FROM DUSK TILL DAWN (1996)
21	DRAGONHEART (1996)
22	BLOODSPORT 2 (A.K.A. BLOODSPORT II: THE NEXT
23	LONG KISS GOODNIGHT, THE (1996)
24	TIME TO KILL, A (1996)
25	HANDS ON A HARD BODY (1996)
26	KISSED (1996)
27	SOLO (1996)
28	ONE FINE DAY (1996)
29	HUSTLER WHITE (1996)
30	BLOOD AND WINE (BLOOD & WINE) (1996)
31	HOMEWARD BOUND II: LOST IN SAN FRANCISCO (
32	SYDNEY (HARD EIGHT) (1996)
33	FARGO (1996)
34	SHE'S THE ONE (1996)
35	GLIMMER MAN, THE (1996)
36	EVITA (1996)
37	SUBSTITUTE, THE (1996)
38	JOE'S APARTMENT (1996)
39	DRAGON BALL: THE PATH TO POWER (DORAGON BÔ

Figure 11 year wise movie search



	movieId	title	avg_rating
1	1759	FOUR DAYS IN SEPTEMBER (O QUE É ISSO, COMP	5.00
2	136353	SCOOBY-DOO! AND THE LOCH NESS MONSTER (2	5.00
3	8238	LITTLE MURDERS (1971)	5.00
4	78836	ENTER THE VOID (2009)	5.00
5	5244	SHOGUN ASSASSIN (1980)	5.00
6	74226	DREAM OF LIGHT (A.K.A. QUINCE TREE SUN, THE)	5.00
7	112512	COLOURFUL (KARAFURU) (2010)	5.00
8	6835	ALIEN CONTAMINATION (1980)	5.00
9	170597	A PLASTICINE CROW (1981)	5.00
10	5723	CONTINENTAL DIVIDE (1981)	5.00

Figure 12 Top n rated



	movieId	title	count
1	356	FORREST GUMP (1994)	1316
2	296	PULP FICTION (1994)	1228
3	1	TOY STORY (1995)	1075
4	364	LION KING, THE (1994)	1032
5	4306	SHREK (2001)	1020
6	79132	INCEPTION (2010)	1001
7	480	JURASSIC PARK (1993)	952
8	588	ALADDIN (1992)	915
9	380	TRUE LIES (1994)	890
10	595	BEAUTY AND THE BEAST (1991)	876

Figure 13 top n watched

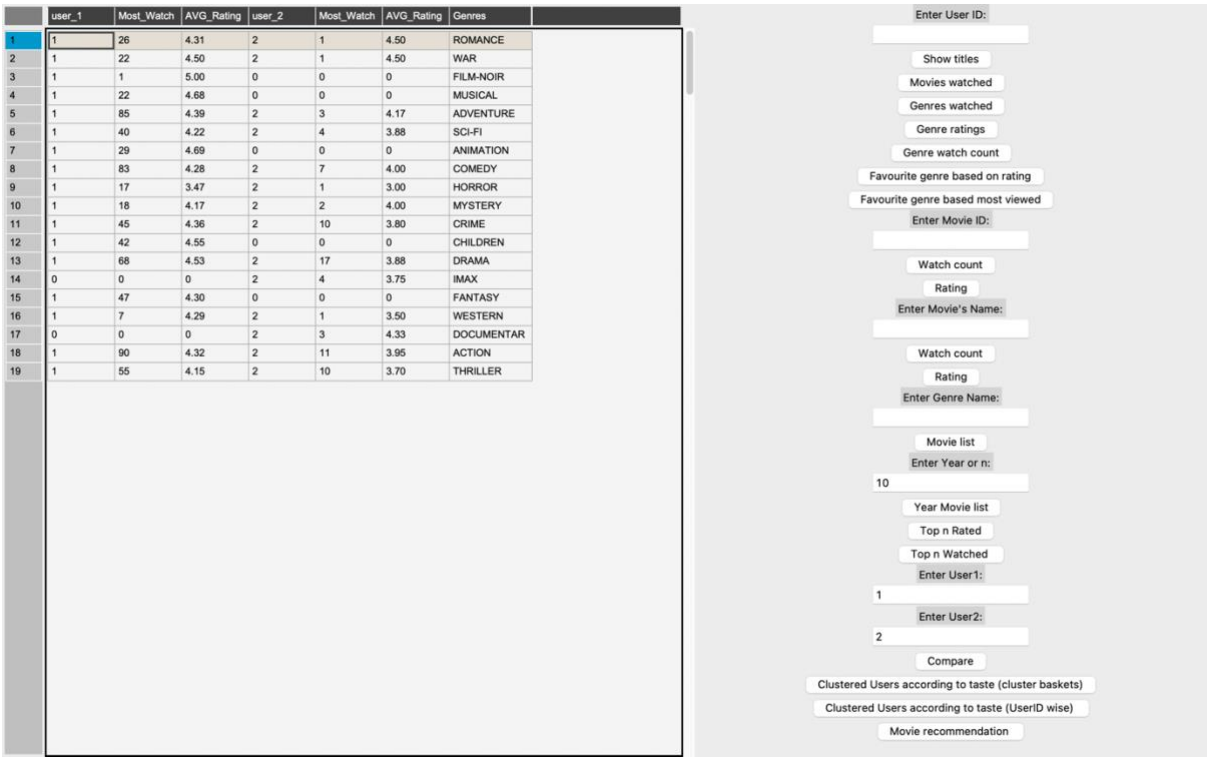
Summary Part 2:

For this part, I implemented the Favourite genre based on ratings and Favourite genre based on most watched, by user Id (6 and 7 buttons in part 1).
I added the functionality to compare the taste of 2 users based on the genre ratings and number of movies watched from each genre.
Finally for part 2, I implemented K means clustering algorithm for clustering users according to their taste. To this end, taste is defined by the number of movies watched by the user from each genre. I decided on using the genre most watched by the user because of the abnormality of different genres available. For Example, Film Noir genre isn't a genre that has many movies, if a user has watched only one movie and liked it with a high rating, it would not reflect his real preference. I have added 2 buttons for viewing this:

1. Cluster basket view: which puts the user in cluster baskets or columns of their assigned cluster.
2. User Id view: that shows the user Id wise view of cluster.

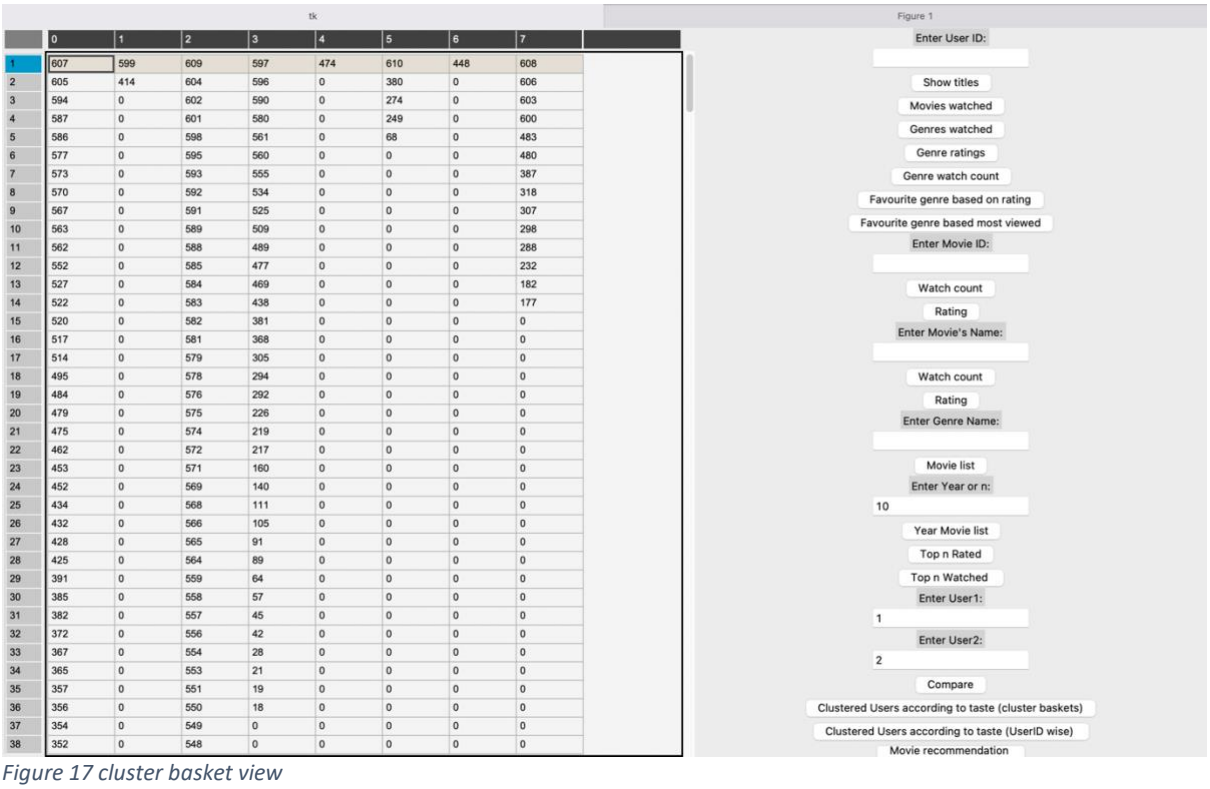
Figure 14 favourite genre (rating based)

Figure 15 favourite genre (most viewed)



	user_1	Most_Watch	AVG_Rating	user_2	Most_Watch	AVG_Rating	Genres
1	1	26	4.31	2	1	4.50	ROMANCE
2	1	22	4.50	2	1	4.50	WAR
3	1	1	5.00	0	0	0	FILM-NOIR
4	1	22	4.68	0	0	0	MUSICAL
5	1	85	4.39	2	3	4.17	ADVENTURE
6	1	40	4.22	2	4	3.88	SCI-FI
7	1	29	4.69	0	0	0	ANIMATION
8	1	83	4.28	2	7	4.00	COMEDY
9	1	17	3.47	2	1	3.00	HORROR
10	1	18	4.17	2	2	4.00	MYSTERY
11	1	45	4.36	2	10	3.80	CRIME
12	1	42	4.55	0	0	0	CHILDREN
13	1	68	4.53	2	17	3.88	DRAMA
14	0	0	0	2	4	3.75	IMAX
15	1	47	4.30	0	0	0	FANTASY
16	1	7	4.29	2	1	3.50	WESTERN
17	0	0	0	2	3	4.33	DOCUMENTAR
18	1	90	4.32	2	11	3.95	ACTION
19	1	55	4.15	2	10	3.70	THRILLER

Figure 16 user comparison



	0	1	2	3	4	5	6	7
1	607	599	609	597	474	610	448	608
2	605	414	604	596	0	380	0	606
3	594	0	602	590	0	274	0	603
4	587	0	601	580	0	249	0	600
5	586	0	598	561	0	68	0	483
6	577	0	595	560	0	0	0	480
7	573	0	593	555	0	0	0	387
8	570	0	592	534	0	0	0	318
9	567	0	591	525	0	0	0	307
10	563	0	589	509	0	0	0	298
11	562	0	588	489	0	0	0	288
12	552	0	585	477	0	0	0	232
13	527	0	584	469	0	0	0	182
14	522	0	583	438	0	0	0	177
15	520	0	582	381	0	0	0	0
16	517	0	581	368	0	0	0	0
17	514	0	579	305	0	0	0	0
18	495	0	578	294	0	0	0	0
19	484	0	576	292	0	0	0	0
20	479	0	575	226	0	0	0	0
21	475	0	574	219	0	0	0	0
22	462	0	572	217	0	0	0	0
23	453	0	571	160	0	0	0	0
24	452	0	569	140	0	0	0	0
25	434	0	568	111	0	0	0	0
26	432	0	566	105	0	0	0	0
27	428	0	565	91	0	0	0	0
28	425	0	564	89	0	0	0	0
29	391	0	559	64	0	0	0	0
30	385	0	558	57	0	0	0	0
31	382	0	557	45	0	0	0	0
32	372	0	556	42	0	0	0	0
33	367	0	554	28	0	0	0	0
34	365	0	553	21	0	0	0	0
35	357	0	551	19	0	0	0	0
36	356	0	550	18	0	0	0	0
37	354	0	549	0	0	0	0	0
38	352	0	548	0	0	0	0	0

Figure 17 cluster basket view

Summary part 3:

For this part I have implemented the whole application with Pandas table, this gives a nice view of all the data. The other option was printing, which was untidy and difficult to interpret.

For the User cluster, I have provided the cluster assignment graph, this was done using ml.feature was used, function ‘PCA’.

Lastly, I have added a movie recommendation system, which uses the rating given by the user to predict his preferences. To this end, pyspark’s ml.recommendation was used with its function ‘ALS’.

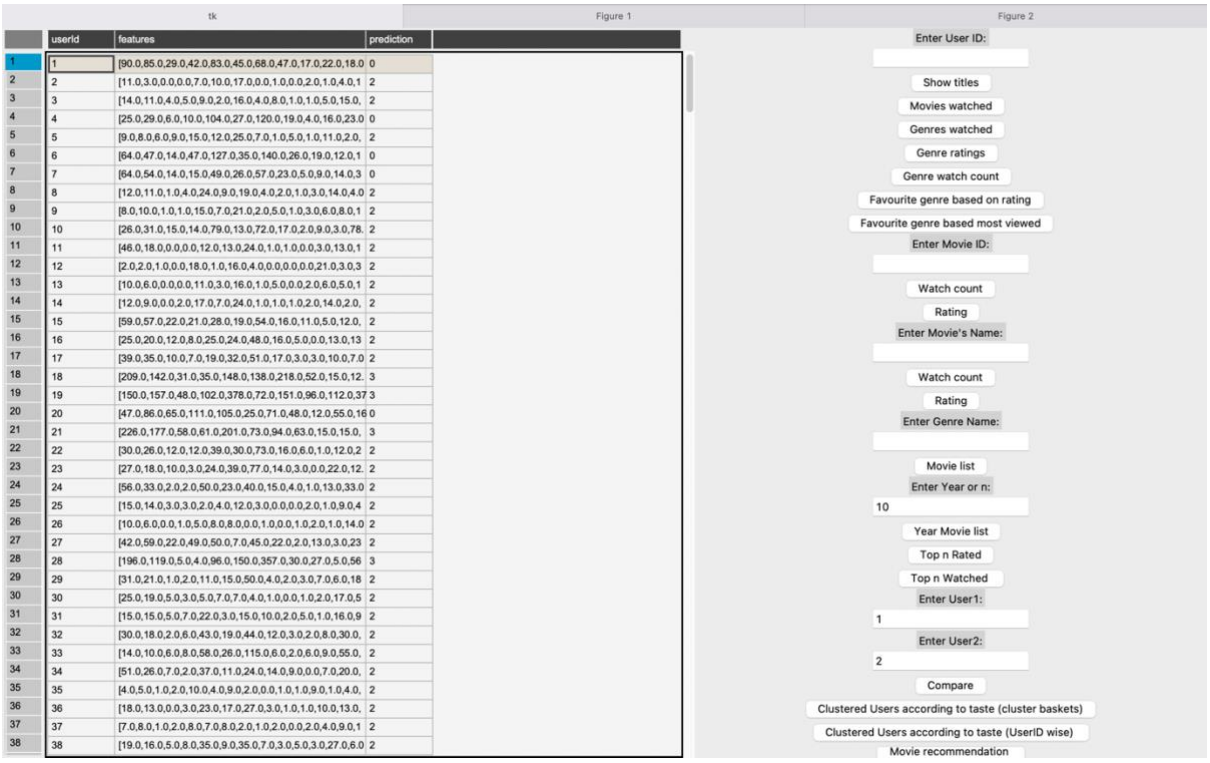


Figure 18 movie recommendation system

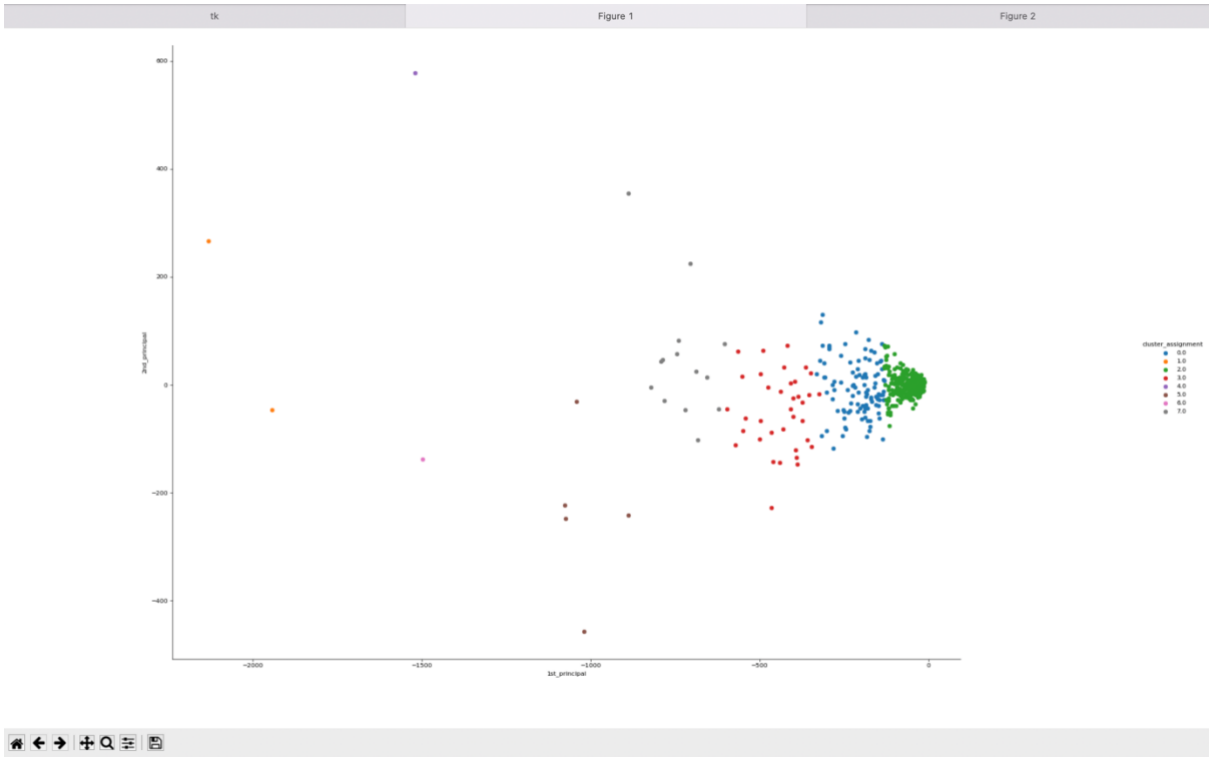


Figure 19 cluster graph visualization

Other effort:

I also implemented this program on a console, using spark submit. It was a fine application having a limitation that you can't have visualizations loaded onto it. I used a command that takes input in runtime, not possible with python's input command. But as mentioned before, the lack of visualizations forced me to switch to an application.

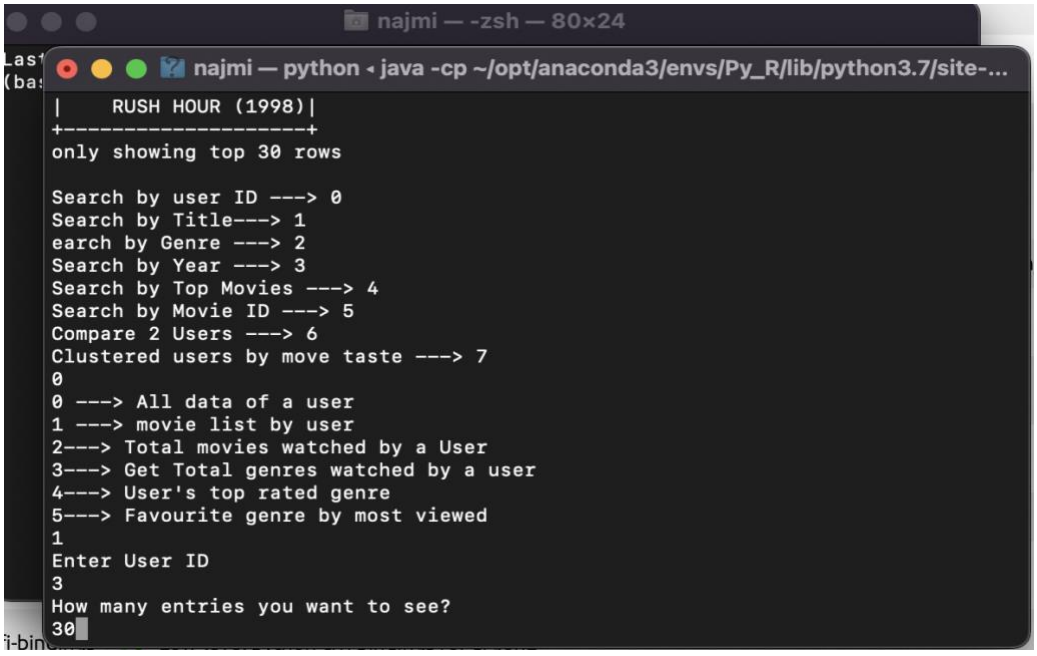


Figure 20 console app interface

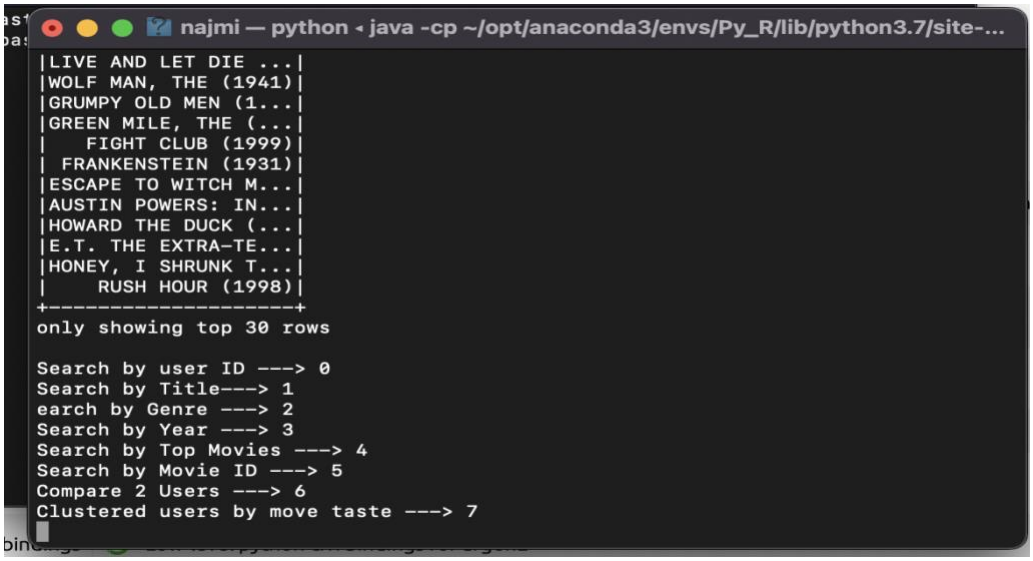


Figure 21 some result