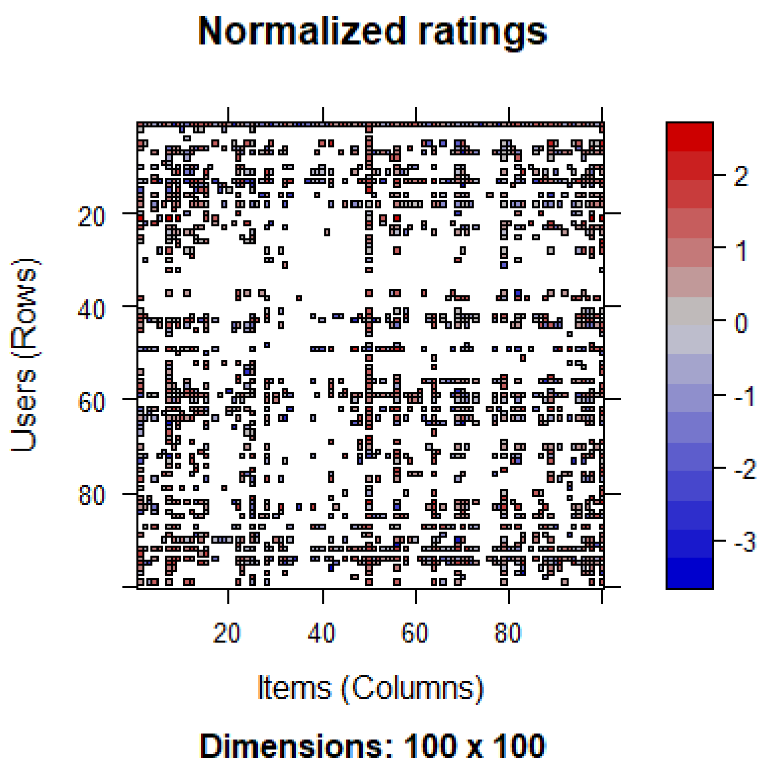
Personalized Movie Recommendation System using Collaborative Filtering

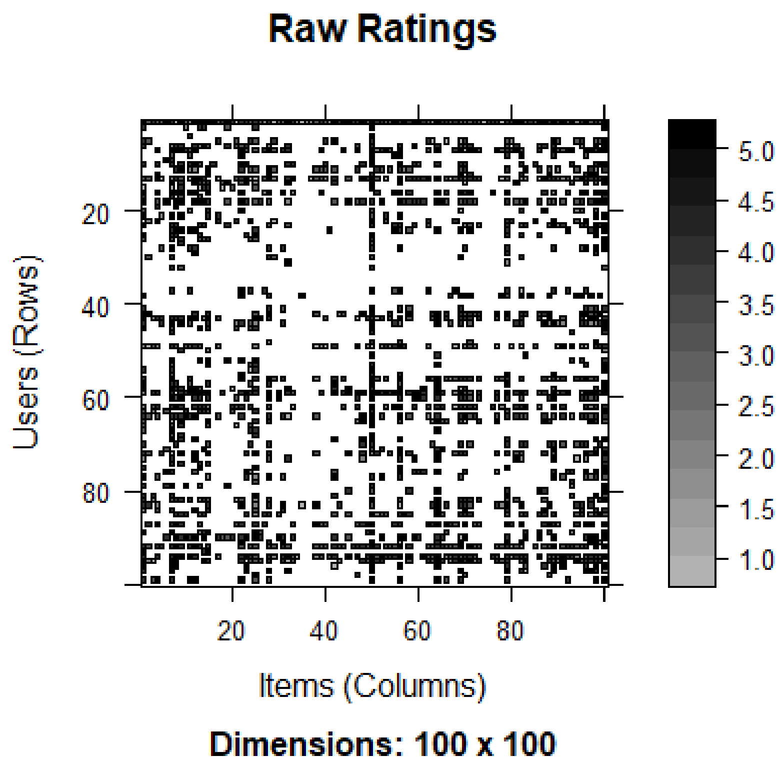
Author: Nandini Ethirajulu

```
##This R script demonstrates the implementation of movie recommendation
systems using the MovieLens dataset, facilitated by the recommenderlab
package. The dataset comprises approximately 100,000 ratings from 943 users
on 1664 movies, collected over seven months via the MovieLens website

#install.packages("recommenderlab")
library("recommenderlab")



data(MovieLense)



########### Exploratory Data Analysis and Preprocessing ###########




#?MovieLense

Moviedata =MovieLense[]


dim(getRatingMatrix(Moviedata)) ## 943 users 1664 movies are present in this
dataset

## [1]  943 1664

##analysing the matrix with limited dataset
getRatingMatrix(Moviedata)[1:10, 1:10]

## 10 x 10 sparse Matrix of class "dgCMatrix"

##   [[ suppressing 10 column names 'Toy Story (1995)', 'GoldenEye (1995)',
'Four Rooms (1995)' ... ]]

##
## 1   5 3 4 3 3 5 4 1 5 3
## 2   4 . . . . . . . . 2
## 3   . . . . . . . . . .
## 4   . . . . . . . . . .
## 5   4 3 . . . . . . . .
## 6   4 . . . . . 2 4 4 .
```

```
## 7  . . . 5 . . 5 5 5 4
## 8  . . . . . . 3 . . .
## 9  . . . . . 5 4 . . .
## 10 4 . . 4 . . 4 . 4 .

########### Normalizing the data ###########


Moviedata_Normalize <- normalize(Moviedata)
Moviedata_Normalize

## 943 x 1664 rating matrix of class 'realRatingMatrix' with 99392 ratings.
## Normalized using center on rows.

image(Moviedata_Normalize[1:100,1:100], main = "Normalized ratings")
```

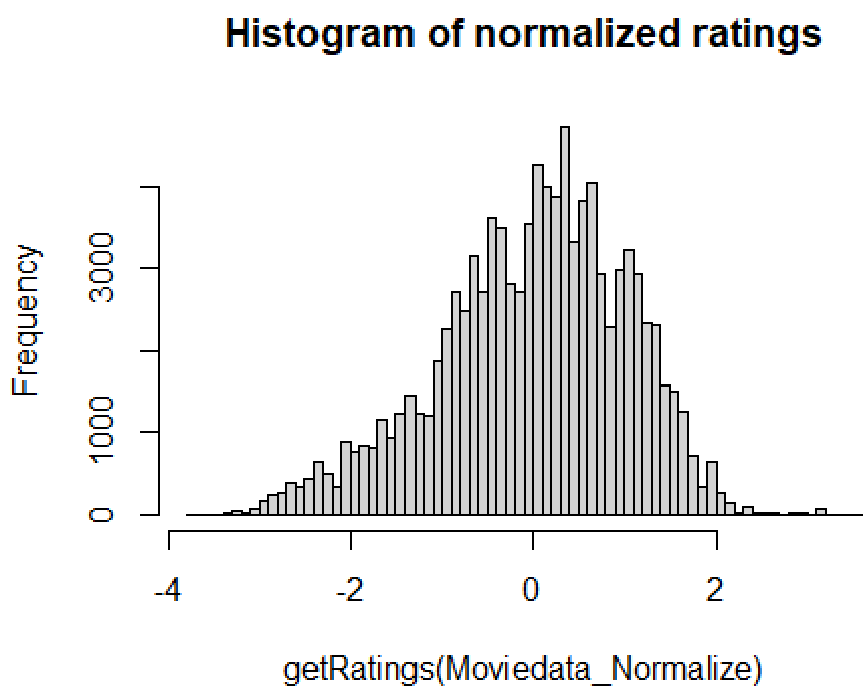



```
image(Moviedata[1:100, 1:100], main = "Raw Ratings")
```

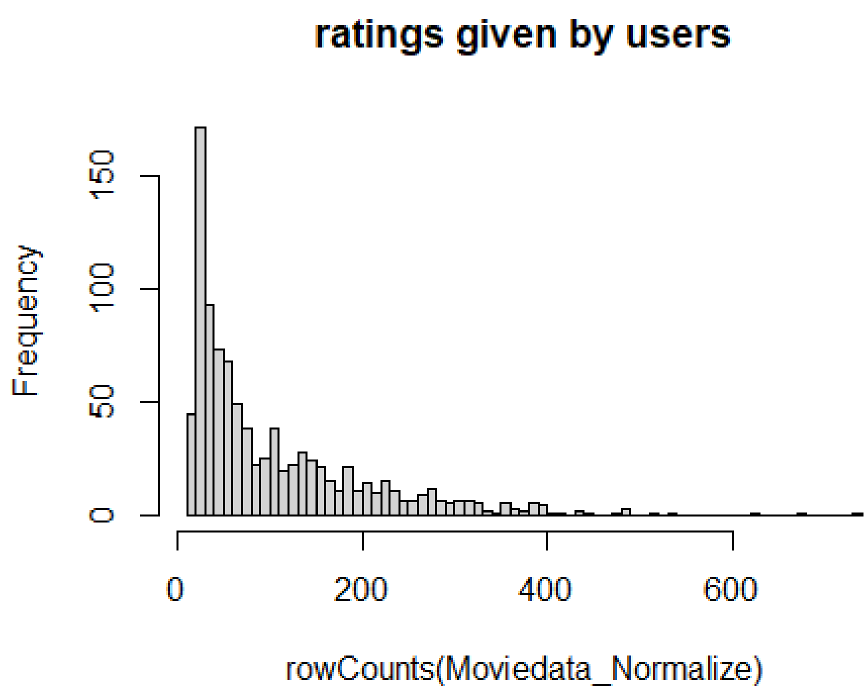



```
getRatingMatrix(Moviedata_Normalize)[1:10, 1:10]

## 10 x 10 sparse Matrix of class "dgCMatrix"

##   [[ suppressing 10 column names 'Toy Story (1995)', 'GoldenEye (1995)',
'Four Rooms (1995)' ... ]]

##
## 1   1.3948339 -0.6051661 0.3948339 -0.6051661 -0.6051661 1.3948339
0.3948339
## 2   0.2950820  .          .          .          .          .          .
## 3   .          .          .          .          .          .          .
## 4   .          .          .          .          .          .          .
## 5   1.1257143  0.1257143 .          .          .          .          .
## 6   0.3605769  .          .          .          .          .         -
1.6394231
## 7   .          .          .          1.0350000  .          .
1.0350000
## 8   .          .          .          .          .          .         -
0.7966102
## 9   .          .          .          .          .          0.7272727 -
0.2727273
## 10 -0.2065217  .          .         -0.2065217  .          .         -
0.2065217
##
## 1  -2.6051661  1.3948339 -0.6051661
## 2   .          .         -1.7049180
```

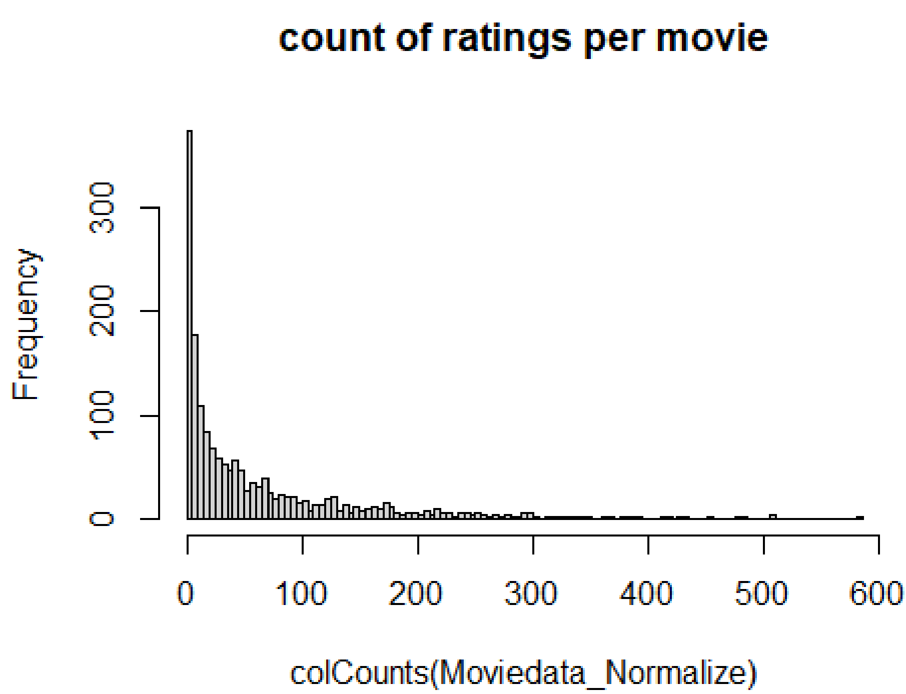
```
## 3    .           .           .
## 4    .           .           .
## 5    .           .           .
## 6    0.3605769   0.3605769   .
## 7    1.0350000   1.0350000   0.0350000
## 8    .           .           .
## 9    .           .           .
## 10   .          -0.2065217   .

########### Dataset Visualization ###########

hist(getRatings(Moviedata_Normalize), breaks = 100, main = "Histogram of
normalized ratings")
```




```
hist(rowCounts(Moviedata_Normalize), breaks = 100, main = "ratings given by
users")
```




```
hist(colCounts(Moviedata_Normalize), breaks = 100, main = "count of ratings
per movie")
```



```r
#a) Developing a user-based recommender system that outputs a user’s top ten recommendations.





dim(Moviedata)#943 1664

## [1]  943 1664

Moviedatasample <- sample(Moviedata, 600)
dim(Moviedatasample) #600 1664

## [1]  600 1664

Moviedata_evaluationScheme<- evaluationScheme(Moviedatasample,method = "split", given = 15, train=0.5, goodRating=4)
Moviedata_evaluationScheme

## Evaluation scheme with 15 items given
## Method: 'split' with 1 run(s).
## Training set proportion: 0.500
## Good ratings: >=4.000000
## Data set: 600 x 1664 rating matrix of class 'realRatingMatrix' with 60543 ratings.

##Developing a user-based recommender system
Moviedata_userbased_model<-
Recommender(getData(Moviedata_evaluationScheme,"train"), "UBCF")
Moviedata_userbased_model

## Recommender of type 'UBCF' for 'realRatingMatrix'
## learned using 300 users.

predictions_userbased <- predict(Moviedata_userbased_model,
Moviedata[603:607], n = 10)


########### Illustrating top 10 recommendations for 5 users on five users.
###########

print("top 10 recommendations for 5 users with a user-based recommender system are as follows: ")

## [1] "top 10 recommendations for 5 users with a user-based recommender system are as follows: "

as(predictions_userbased, "list")

## $`0`
##  [1] "Pillow Book, The (1995)"
```

```
##  [2] "Emma (1996)"
##  [3] "Shiloh (1997)"
##  [4] "Close Shave, A (1995)"
##  [5] "Antonia's Line (1995)"
##  [6] "Marvin's Room (1996)"
##  [7] "Lone Star (1996)"
##  [8] "Shanghai Triad (Yao a yao yao dao waipo qiao) (1995)"
##  [9] "Secrets & Lies (1996)"
## [10] "Con Air (1997)"
##
## $`1`
##  [1] "Marvin's Room (1996)"
##  [2] "Close Shave, A (1995)"
##  [3] "Paradise Lost: The Child Murders at Robin Hood Hills (1996)"
##  [4] "Secrets & Lies (1996)"
##  [5] "Nightmare Before Christmas, The (1993)"
##  [6] "Homeward Bound: The Incredible Journey (1993)"
##  [7] "Bedknobs and Broomsticks (1971)"
##  [8] "Hunt for Red October, The (1990)"
##  [9] "Cinderella (1950)"
## [10] "Strictly Ballroom (1992)"
##
## $`2`
##  [1] "Glengarry Glen Ross (1992)"
##  [2] "Good, The Bad and The Ugly, The (1966)"
##  [3] "Raging Bull (1980)"
##  [4] "Unforgiven (1992)"
##  [5] "Sunset Blvd. (1950)"
##  [6] "Night of the Living Dead (1968)"
##  [7] "Once Upon a Time in the West (1969)"
##  [8] "Die xue shuang xiong (Killer, The) (1989)"
##  [9] "Living in Oblivion (1995)"
## [10] "Switchblade Sisters (1975)"
##
## $`3`
##  [1] "Oscar & Lucinda (1997)"     "Wild Things (1998)"
##  [3] "Lost in Space (1998)"       "Brassed Off (1996)"
##  [5] "Ed Wood (1994)"             "Citizen Kane (1941)"
##  [7] "Sound of Music, The (1965)" "Close Shave, A (1995)"
##  [9] "As Good As It Gets (1997)"  "Apostle, The (1997)"
##
## $`4`
##  [1] "Until the End of the World (Bis ans Ende der Welt) (1991)"
##  [2] "Mr. Holland's Opus (1995)"
##  [3] "Rumble in the Bronx (1995)"
##  [4] "Birdcage, The (1996)"
##  [5] "Spitfire Grill, The (1996)"
##  [6] "Mirror Has Two Faces, The (1996)"
##  [7] "Mission: Impossible (1996)"
##  [8] "First Wives Club, The (1996)"
```

```
##  [9] "Fan, The (1996)"
## [10] "Eraser (1996)"

 # b) Developing a item-based recommender system that outputs same user’s top
ten recommendations.


## item-based recommender system
Moviedata_itembased_model<-
Recommender(getData(Moviedata_evaluationScheme,"train"), "IBCF")
Moviedata_itembased_model

## Recommender of type 'IBCF' for 'realRatingMatrix'
## learned using 300 users.

predictions_itembased <- predict(Moviedata_itembased_model,
Moviedata[603:607], n = 10)



print("top 10 recommendations for 5 users with a item-based recommender
system are as follows: ")

## [1] "top 10 recommendations for 5 users with a item-based recommender
system are as follows: "

as(predictions_itembased, "list")

## $`0`
##  [1] "Mighty Aphrodite (1995)"           "Postino, Il (1994)"
##  [3] "Mr. Holland's Opus (1995)"         "Birdcage, The (1996)"
##  [5] "Brothers McMullen, The (1995)"     "Apollo 13 (1995)"
##  [7] "Outbreak (1995)"                   "Shawshank Redemption, The
(1994)"
##  [9] "Ace Ventura: Pet Detective (1994)" "Forrest Gump (1994)"
##
## $`1`
##  [1] "Joy Luck Club, The (1993)"
##  [2] "Boys (1996)"
##  [3] "Ballad of Narayama, The (Narayama Bushiko) (1958)"
##  [4] "Promesse, La (1996)"
##  [5] "Hearts and Minds (1996)"
##  [6] "Color of Night (1994)"
##  [7] "Johnny 100 Pesos (1993)"
##  [8] "Faust (1994)"
##  [9] "Stranger, The (1994)"
## [10] "Gabbeh (1996)"
##
## $`2`
##  [1] "Horseman on the Roof, The (Hussard sur le toit, Le) (1995)"
##  [2] "Unhook the Stars (1996)"
```

```
##  [3] "Paradise Lost: The Child Murders at Robin Hood Hills (1996)"
##  [4] "Crossfire (1947)"
##  [5] "Man of No Importance, A (1994)"
##  [6] "Night Flier (1997)"
##  [7] "Gang Related (1997)"
##  [8] "Bewegte Mann, Der (1994)"
##  [9] "Paris, France (1993)"
## [10] "Bitter Sugar (Azucar Amargo) (1996)"
##
## $`3`
##  [1] "Kansas City (1996)"
##  [2] "Alaska (1996)"
##  [3] "Soul Food (1997)"
##  [4] "Heaven's Prisoners (1996)"
##  [5] "Carried Away (1996)"
##  [6] "Twelfth Night (1996)"
##  [7] "Chungking Express (1994)"
##  [8] "Prefontaine (1997)"
##  [9] "Old Lady Who Walked in the Sea, The (Vieille qui marchait dans la
mer, La) (1991)"
## [10] "Paradise Lost: The Child Murders at Robin Hood Hills (1996)"
##
## $`4`
##  [1] "Women, The (1939)"
##  [2] "Madame Butterfly (1995)"
##  [3] "White Balloon, The (1995)"
##  [4] "Maya Lin: A Strong Clear Vision (1994)"
##  [5] "Promesse, La (1996)"
##  [6] "Daniel Defoe's Robinson Crusoe (1996)"
##  [7] "Blue Angel, The (Blaue Engel, Der) (1930)"
##  [8] "Wonderful, Horrible Life of Leni Riefenstahl, The (1993)"
##  [9] "Whole Wide World, The (1996)"
## [10] "Double Team (1997)"
```

```
###general comparison and understanding between part A (user-based
recommender system) and B (item-based recommender system)


# Part  A provides movie recommendations based on similarities between the
user's . Whereas, Part B provides movie recommendations based on the
similarities observed between the movies of the selected 5 users.

#c)

######### Understanding and comparing the first users ratings given in the
dataset to the  recommendations given to the user from user based recommender
system model -


print("First Users ratings: ")
```

```
## [1] "First Users ratings: "
```

```r
(as(MovieLense[603,], "list"))
```

```
## $`0`
##                                                            Twelve Monkeys
(1995)
##
5
##                                                             Seven (Se7en)
(1995)
##
5
##                                                       Usual Suspects, The
(1995)
##
5
##                                                    Muppet Treasure Island
(1996)
##
3
##                                                                Braveheart
(1995)
##
4
##                                                                 Star Wars
(1977)
##
5
##                                                              Pulp Fiction
(1994)
##
4
##                                                                  Stargate
(1994)
##
2
##                                                              Blade Runner
(1982)
##
5
##                                                                     Fargo
(1996)
##
4
##                                                                   Platoon
(1986)
##
1
##                                                  Empire Strikes Back, The
```

```
(1980)
##
5
##                                                     Princess Bride, The
(1987)
##
4
##                                                 Raiders of the Lost Ark
(1981)
##
3
##                                                                  Aliens
(1986)
##
2
##                                                          Apocalypse Now
(1979)
##
4
##                                                      Return of the Jedi
(1983)
##
5
##                                                                   Alien
(1979)
##
4
##                                     Indiana Jones and the Last Crusade
(1989)
##
4
##                                                 When Harry Met Sally...
(1989)
##
4
##                                                Star Trek: First Contact
(1996)
##
4
##                                 Star Trek VI: The Undiscovered Country
(1991)
##
3
##                                             Star Trek: The Wrath of Khan
(1982)
##
3
##                                     Star Trek III: The Search for Spock
(1984)
##
```

```
4
##                                          Star Trek IV: The Voyage Home
(1986)
##
4
##                                                    Fifth Element, The
(1997)
##
5
##                                                     Starship Troopers
(1997)
##
2
##                                                                  Heat
(1995)
##
1
##                                                                Scream
(1996)
##
3
##                                                            Liar Liar
(1997)
##
4
##                                                               Titanic
(1997)
##
5
##                                                            G.I. Jane
(1997)
##
4
##                                               Star Trek: Generations
(1994)
##
4
##                                                            True Lies
(1994)
##
4
##                                                         Mary Poppins
(1964)
##
2
##                                   Day the Earth Stood Still, The
(1951)
##
5
##                                         Star Trek: The Motion Picture
```

```
(1979)
##
4
##                                       Star Trek V: The Final Frontier
(1989)
##
3
## Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb
(1963)
##
4
##                                                          Benny & Joon
(1993)
##
3
##                                                            Saint, The
(1997)
##
5
##                                                   Tomorrow Never Dies
(1997)
##
4
##                                                     Raise the Red Lantern
(1991)
##
4
##                                               Island of Dr. Moreau, The
(1996)
##
2
##                                             Beautician and the Beast, The
(1997)
##
4
##                                       Ghost in the Shell (Kokaku kidotai)
(1995)
##
5
##                                                Man in the Iron Mask, The
(1998)
##
5
```

```
print("First Users recommendations given based on the user based recommender system model: ")
```

```
## [1] "First Users recommendations given based on the user based recommender system model: "
```

```r
as(predictions_userbased, "list")[1]

## $`0`
##  [1] "Pillow Book, The (1995)"
##  [2] "Emma (1996)"
##  [3] "Shiloh (1997)"
##  [4] "Close Shave, A (1995)"
##  [5] "Antonia's Line (1995)"
##  [6] "Marvin's Room (1996)"
##  [7] "Lone Star (1996)"
##  [8] "Shanghai Triad (Yao a yao yao dao waipo qiao) (1995)"
##  [9] "Secrets & Lies (1996)"
## [10] "Con Air (1997)"

##goals of a recommender system – relevance, novelty, serendipity, diversity:
##Analysis each of them -

##Relevance -- can be understood as the overlap of the user's
preferences(based on their original ratings) and the recommendations obtained
with the model. From overall analysis, we could see that there is no
significant overlap of movies between these sets of movies -the genres do not
align with users interests . This in turn means that these recommendations
are not relevant to a great extent for this user.



##Novelty -- this metric indicates whether the user is provided with new or
more wider variety of movie genre recommendations apart from those the user
has already watched and rated. As we analysed from relevance metric that
there is no much overlap between the movie recommendations and user ratings,
we can conclude by saying that user based recommender system has achieved its
goal of providing high novelty to the user.


##serendipity - Good serendipity indicates if the user has been recommended
with an suprising or an unexpected movie which is newer to him to watch and
rate, however it is more relevant and highly matching with the genre of his
interests. With the given information, we can assess that user has a wider
variety interest in watching different genres (total of 47 movies rated), and
even the recommendations suggest wide variety of genres. So there are
possibilities that serendipity is achieved for this user.


##diversity -- illustrates if the user was recommended with movies of
different genres, themes, or even from different time periods. From the
original data - user's ratings we can understand that user has watched and
rated movies from 1950 to 1998. However, movie recommendations were not
provided from wide range of years. This indicates there is no much diversity
in this recommendation for this user.
```