# We Know What You Want:


# Movies & Models

UC DAVIS 2023 MSBA BAX 401

AJAIY PRAVEEN, SRIPRIYA SRINIVASAN, FRED XUE, JINNY ZHONG

# Table of Contents

## Executive Summary

This report summarizes the prediction of movie ratings for users based on the recommendation system our team is building. This tests out our recommendation system for implementation on our platform. An accurate recommendation system can effectively improve user experience, prolong user lifespan, broaden our marketing flexibility, and ultimately sustain revenue for our business.

Our recommendation system is fairly mature at this stage. We accurately predicted team members' rating on three movies in 6 out of 8 applicable cases. We achieved this even under data constraints.

Then, we took it to the next level. We made predictions on new users and movies where we have no prior data. These "cold start" cases are particularly tricky in the industry, but our system made it work. We successfully produced rating predictions for new movies based on existing user data, and for brand new users in cases with and without prior data.

Behind the scenes, the core of our success is the method of collaborative filtering. It is a machine learning technique that uses similarities between users and movies simultaneously to make predictions. Think about it this way. We tend to ask for recommendations from friends who have similar taste in movies to our own. Collaborative filtering works the same way. We can get into the details in our next meeting. For now, rest assured that our predictions are as robust as they can be given the data at hand. In our implementation, we took steps to eliminate user bias and item bias that could cloud our predictions.

We're not stopping here though. With more data and testing, we can build a robust and accurate recommendation system that could pinpoint each user's interest and offer a collection of tailor-made films to watch. This will allow us to stay in the game and even edge out competitors in an industry that is ever more reliant on recommendation systems.

## Introduction

How often do we waste time trying to pick the best flick to watch over dinner? Ever ended up regretting watching a suggested movie? As a part of MovieRecommender.Inc, our key objective is to provide the best recommendations to our clients. We analyze historic movie ratings from numerous users along with personalized user parameters to deliver the perfect recommendation system.

## Problem Formulation

Our goal is to create a recommendation system primarily using collaborative filtering. Both historical data from DMBI and internal employee data are available. For starters we develop a basic model to predict movie rating for team members with both user and item-based filters. We then enhance the model and predict ratings for newly joined customers whom we have no data on. On obtaining initial data on new customers, we feed this data to our system to see if it affects our predictions.

## Data Characteristics

We have two datasets – 98 rows of ratings (1 to 5, hate to love)  in employee dataset and 280 rows in DBMI dataset on 50 movies (Details in *Appendix 1-4*).

In terms of popularity (i.e., the number of users that rated the movie). The most popular movies are typically Hollywood box office champions such as *Avatar* (84) while the least known titles are about niche cultures such as *Ida* (4). Besides, in terms of public preferences, we compared the average ratings among movies. *Inception* (4.88) received the highest user average adjusted rating while *A Separation* (2.44) was rated the lowest. Also, we found that about 72% of ratings are 5 and 4 and only 2% of ratings are 1. This shows that either the users are lenient overall, or the set of movies is generally critically acclaimed. We also discovered that most users rated about 7 to 20 movies out of 50 with a few movie "fanatics" who have rated more than 40 movies.

## Model Development, Predictions and Results

**For task one**, we developed several models to make predictions and do comparisons.

The picked 3 movies are: *La La Land*, *Gravity* and *The Imitation Game*.

Model 1 - 3: Item-Item Based Collaborative Filters with Original Ratings, Mean Centered, and Z-scores

Starting off, we used Item-Item Collaborative Filtering (IICF) with original movie ratings. To predict ratings for our members. We applied this model to both the class data and the combined data of class and DBMI ratings. Then, we used IICF on normalized movie ratings by centering each rating by the overall average for each team member. To mitigate the impacts of different variances among individuals in Model 1&2, we standardized and replaced the ratings with Z-scores. After that, we calculated the cosine similarity matrix of all movies and got the top 15 neighbors of each movie with the highest similarity rankings. The results are summarized in the table below. (Details in *Appendix 5-7*)

| | Actual | | | Item Item with zscores | | | Item Item with original scores | | | Item Item with mean centered | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | La la land | Gravity | The Intimation Game | La la land | Gravity | The Intimation Game | La la land | Gravity | The Intimation Game | La la land | Gravity | The Intimation Game |
| Fred | 4.00 | 3.00 | - | 4.04 | 4.01 | 4.30 | 2.27 | 2.49 | 2.42 | 4.00 | 3.28 | 2.76 |
| Ajaiy | 3.00 | 1.00 | 1.00 | 3.05 | 2.24 | 3.44 | 0.93 | 1.24 | 1.03 | 3.89 | 2.55 | 2.08 |
| Sripriya | - | - | - | 3.84 | 3.87 | 4.27 | 0.94 | 0.91 | 0.89 | 4.11 | 3.00 | 2.50 |
| Jinny | 5.00 | 5.00 | 3.00 | 4.01 | 4.02 | 4.37 | 3.54 | 1.67 | 1.65 | 4.00 | 3.17 | 2.66 |

We used the z-score method as one of the potential paths in **task 2.** We still standardized ratings as the first step, but this time we calculated average, standard deviation and cosine based on the DMBI dataset.

| Name | Winter's bone | Serious Man | Son of Soul |
|---|---|---|---|
| Fred | 2.83 | 2.70 | 3.29 |
| Ajaiy | 2.02 | 2.07 | 2.46 |
| Sripriya | 2.75 | 2.67 | 3.25 |
| Jinny | 2.86 | 2.77 | 3.35 |

Results are in the table above. Besides, we could add new data or more content features as another path, like movie genre, target audience, launch time, box office takings etc. and make classification among movie attributes. Also, user-based methods are useful paths to find similar users and make predictions.

Model 4: User-User Based Collaborative Filters (UUCF) *(Appendix 8)*

We developed a UUCF model that clusters similar users together. Since user's ratings are subjective, it is essential to normalize the ratings first. Then based on their Pearson correlation method "n" similar users to "user of interest" (say, user A) are identified. With the help of the ratings given by similar users we then predict ratings by user. If a particular movie hasn't been rated by any of the 'n' users, the model is unable to generate a predicted rating for user A, this is observed usually when the value of 'n' is small. Below are the predicted ratings for the members of our team for 3 movies of our choice. (Question 1) :

| Movie Title | La La Land | | | | Gravity | | | | The Imitation Game | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| n - value | 5 | 15 | 30 | Actual Rating | 5 | 15 | 30 | Actual Rating | 5 | 15 | 30 | Actual Rating |
| User 332 - Shaolong | 3.85 | 3.41 | 3.51 | 4 | 3.49 | 3.53 | 3.67 | 3 | N/A | 4.28 | 4.26 | - |
| User 342 - Ajaiy | 2.4 | 2.15 | 1.99 | 3 | 0.02 | 1.03 | 1.29 | 1 | 1.13 | 1.12 | 1.14 | 1 |
| User 362 - Jinny | 5 | 5 | 5 | 5 | N/A | 5 | 4.92 | 5 | 4.33 | 4.25 | 4.37 | 4 |

As "n" increases the prediction accuracy decreases, as our pool of similar users becomes larger, i.e lowering correlation value threshold. We now make predictions for specific movies of interest (as per Question 2). Here compared to the above case, the values of n are much higher, this is because these movies are much rarely watched as compared to the ones in the previous question. Hence more number of similar users (higher 'n' value) are required to make a prediction. This issue can be avoided in the IICF.

| Movie Title | Winter's Bone | | | A Serious Man | | | Son of Saul | | |
|---|---|---|---|---|---|---|---|---|---|
| n - value | 100 | 150 | 200 | 100 | 150 | 200 | 100 | 150 | 200 |
| User 332 - Shaolong | N/A | 4.18 | 4.18 | 3.437 | 3.721 | 3.721 | N/A | N/A | 4.005 |
| User 342 - Ajaiy | 1.023 | 1.023 | 1.023 | 1.023 | 1.023 | 1.023 | 1.023 | 1.023 | 1.023 |
| User 362 - Jinny | 4.226 | 4.226 | 4.225 | 4.342 | 4.432 | 4.317 | 4.432 | 4.432 | 4.317 |

In UUCF using Pearson correlation we also observe NaN values of correlation due to variance being 0, hence for a fairly new user or a user with very few number of ratings like Sripriya (4th team member) the ratings cannot be predicted. This issue is also resolved in the IIFC method.

Different users have subjective opinions of ratings, some are harsh while some generous. Using a boxplot of average ratings of every user we observe among our team members that Ajaiy has harshest while Jinny has most lenient ratings.(*Appendix 9*)

**In Task 3**, we are predicting the rating of 3 new users (Amy, Camille, and Shachi) on *Avatar*, *Inception*, and *The Wolf of Wall Street*. Having no prior data on new users poses a challenge. We define 3 paths to approach this challenge. The predictions from all three paths are in the table below.

| Path | Data | *Inception* | *Avatar* | *The Wolf of Wall Street* |
|---|---|---|---|---|
| Simple Average (1) | Employee | 4.40 | 3.99 | 4.17 |
| | Employee + DBMI | 4.49 | 4.02 | 4.12 |
| IICS of Actual Ratings (2) | Employee | 3.99 | 4.00 | 3.99 |
| | Employee + DBMI | 3.99 | 3.99 | 4.00 |
| IICS of Normalized Ratings (3) | Employee | 4.05 | 3.90 | 3.96 |
| | Employee + DBMI | 3.90 | 3.70 | 3.91 |

First, we make predictions based on simple averages of actual ratings for the 3 movies with historical data. We assume that pooling the ratings across the entire dataset, the averages should at least have some generalizability towards typical new users. In our second path, we predict ratings for new users based on cosine similarity between movies with IICF of prior ratings for 10 films most similar with the 3 targeted movies (*Appendix 10*). We eliminate some bias from the first approach by filtering and predicting based on similar content, but are still exposed to user biases. To eliminate that, in our third path we normalize historical user ratings. Then, we filter for the most similar movies with mean-centered ratings (*Appendix 11*). Eliminating biases, our predictions are getting closer to true ratings.

Our predictions remain high across the board as the 3 targeted movies are blockbusters. Their popularity can mask the effectiveness of our predictions. Predicting ratings of a diverse set of films could yield more apparent results. In addition, we theorized more ways to predict ratings when we face "cold start" user cases. First, most streaming platforms collect information on preference when a new user lands for the first time. Second, we can use UUCF based on demographic information. Lastly, we can gather data that could indirectly inform user taste: views and clicks when browsing movies could indicate interest.

Earlier we did not have any information on these new customers, but now we have their ratings for 5 movies. In **task 4**, we first calculate the cosine similarity between the movies to be predicted vs the movies whose data we now have for each new customer. We then calculate (*Appendix 12*) the weighted average of the rating by providing a higher weight if the movies are more similar to each other. Customers are more likely to give the same rating to similar movies, hence this can be used for new customers. The employee and full dataset resulted in some decimal differences, but on rounding we got the same results.

| User | Inception | Avatar | Wolf of Wall Street |
|---|---|---|---|
| Amy | 3 (2.95) | 3 (2.9) | 3 (2.92) |
| Shachi | 4 (4.17) | 4 (4.15) | 4 (4.16) |
| Camille | 3 (3.17) | 3 (3.16) | 3 (3.14) |

$$Predicted\ Rating = \frac{\sum_1^j CosineSimilarity_{(i,j)} * Rating_j}{\sum_1^j CosineSimilarity_{(i,j)}}$$

Amy's and Camille's prediction as compared to that in task 3, has dropped by 1 for all movies , Shachi's remains the same for all. This change in rating is because earlier we made predictions purely based on the item and no user data, whereas now we have taken into consideration the data of the new users which helps fine tune the model to make a more personalized and accurate prediction.

## Value Creation, Recommendations and Implications

Using such recommendation engines in eCommerce and mobile commerce, the search and transaction costs per user could be reduced. As users spend long hours online purchasing, the basket size, retention rates and conversion rates are all improved. For retailers like Walmart, accurate predictions could help the company to adjust the assortment strategy and balance inventory level. With UUCFs, it could analyze preference differences between consumer groups and help targeted promotions. With IICFs, stores could adjust assortment among similar products to grow profits.

Potentially we could improve the methods by adding more dimensions of movie content and customer profiles. Second, benchmarking with public data sources would expose potential opportunities. Besides, objective value functions could be considered to refine the recommendation process. Instead of making predictions simply based on similarity and scores, a weighted objective function considering the relative importance of each movie could be helpful. The importance index could be based on movie popularity (e.g. number of reviews, box office earnings) and sentiments (e.g. Rotten Tomatoes reviews). Thus, popular movies with positive reviews could get higher weights and be recommended to more users.

To improve areas in customer analytics, benchmarking with open industrial data would help understand market trends and changes in customer behaviors before deep-diving in a specific segment. Richer dimensions provide more angles in analyzing customer journeys and where to make interventions in social media. It is suggested to consider relative priorities when conducting customer analytics.

## Conclusion

We have made predictions for both existing as well as new users by employing different methods. In case of new users or existing users with few ratings, we are inclined to use the IICF method. But for existing users, UUCF produces better predictions at about 75% accuracy and we would recommend that. Had we collected preliminary data for new users about demographics, or choice of language, genre, actors and so on, we could have attempted a UUCF model as well. On the other hand, for existing users with a significant number of ratings, we can use either UUCF or IICF models.

## Appendix

***Appendix 1: Count of Non-Zero Rating by Movie***

As a note, the employee dataset doesn't include ratings on 3 movies: *Winter's Bone, A Serious Man,* and *Son of Saul* as they were not included in the initial survey. "Did Not Watch" or NA is counted as no input.

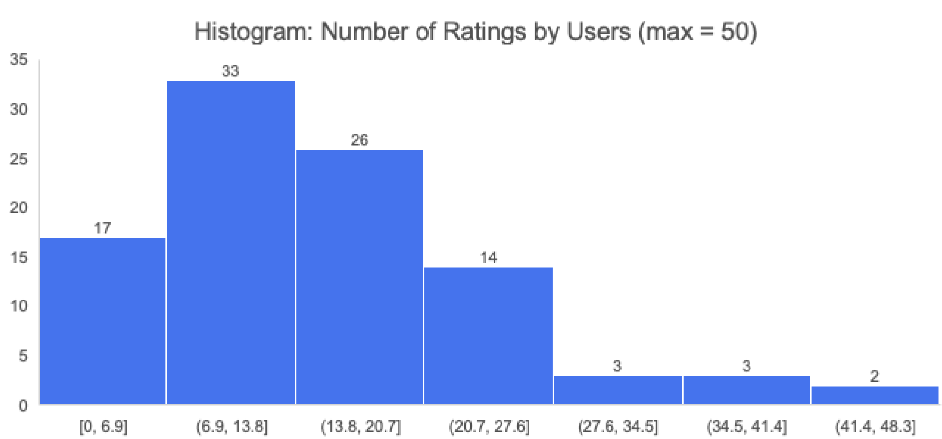
| Movies | Count of Non-0 Rating |
|---|---|
| [Avatar] | 84 |
| [Inception] | 82 |
| [The Wolf of Wall Street] | 81 |
| [Life of Pi] | 69 |
| [La La Land] | 67 |
| [Toy Story 3] | 65 |
| [Gravity] | 58 |
| [The Imitation Game] | 56 |
| [Dunkirk] | 55 |
| [Black Swan] | 51 |
| [The Social Network] | 49 |
| [Mad Max: Fury Road] | 49 |
| [The Grand Budapest Hotel] | 46 |
| [The King's Speech] | 39 |
| [Inglourious Basterds] | 39 |
| [Zero Dark Thirty] | 38 |
| [Arrival] | 38 |
| [12 Years a Slave] | 35 |
| [Birdman] | 33 |
| [Call Me by Your Name] | 32 |
| [The Shape of Water] | 30 |
| [Hugo] | 27 |
| [Argo] | 25 |
| [The Descendants] | 23 |
| [Spotlight] | 22 |
| [Lincoln] | 22 |
| [Moonlight] | 21 |
| [Boyhood] | 19 |
| [Blue is the Warmest Colour] | 17 |
| [Manchester by the Sea] | 17 |
| [Three Billboards Outside Ebbing, Missouri] | 17 |
| [The Fighter] | 17 |
| [The Tree of Life] | 15 |
| [Up in the Air] | 14 |
| [True Grit] | 12 |
| [Amour] | 10 |
| [The Artist] | 10 |
| [Precious] | 10 |
| [The White Ribbon] | 9 |
| [A Prophet] | 8 |
| [A Separation] | 7 |
| [Toni Erdmann] | 7 |
| [Inside Llewyn Davis] | 7 |
| [Leviathan] | 6 |
| [Beasts of the Southern Wild] | 6 |
| [The Secret in Their Eyes] | 5 |
| [Ida] | 4 |

### *Appendix 2: Average Movie Ratings by Users*

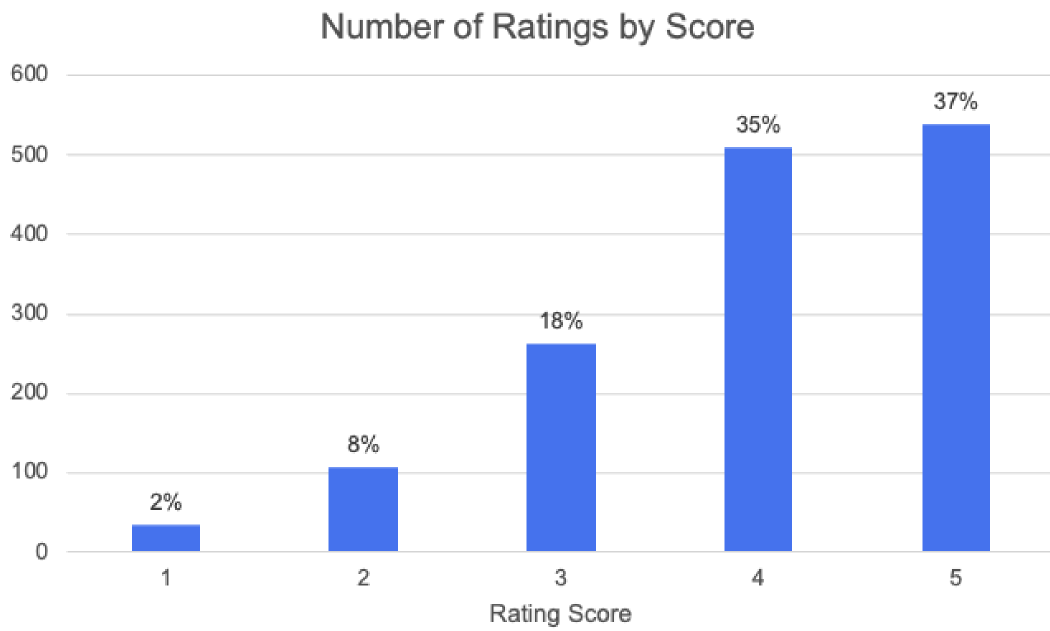
| Movie | Original Ratings |
|---|---|
| Son of Saul | n/a |
| Winter's Bone | n/a |
| A Serious Man | n/a |
| Inception | **4.40** |
| Black Swan | **4.29** |
| **The Imitation Game** | **4.27** |
| The Grand Budapest Hotel | **4.26** |
| Inglourious Basterds | **4.26** |
| **Ida** | **4.25** |
| Call Me by Your Name | **4.25** |
| Three Billboards Outside Ebbing, Missouri | **4.18** |
| The Wolf of Wall Street | **4.17** |
| The Social Network | **4.12** |
| Life of Pi | **4.12** |
| Dunkirk | **4.05** |
| 12 Years a Slave | **4.00** |
| Argo | **4.00** |
| True Grit | **4.00** |
| Avatar | **3.99** |
| Arrival | **3.97** |
| The Tree of Life | **3.93** |
| Up in the Air | **3.93** |
| Moonlight | **3.90** |
| The Shape of Water | **3.90** |
| Zero Dark Thirty | **3.89** |
| Gravity | **3.88** |
| **The King's Speech** | **3.87** |
| Hugo | **3.85** |
| La La Land | **3.85** |
| The Descendants | **3.83** |
| The Secret in Their Eyes | **3.80** |
| Toy Story 3 | **3.78** |
| **The White Ribbon** | **3.78** |
| Spotlight | **3.77** |
| Blue is the Warmest Colour | **3.76** |
| The Fighter | **3.76** |
| Leviathan | **3.67** |
| Manchester by the Sea | **3.65** |
| Mad Max: Fury Road | **3.63** |
| A Prophet | **3.63** |
| Amour | **3.60** |
| Lincoln | **3.59** |
| Birdman | **3.58** |
| The Artist | **3.50** |
| Precious | **3.50** |
| Beasts of the Southern Wild | **3.50** |
| Toni Erdmann | **3.43** |
| Boyhood | **3.42** |
| Inside Llewyn Davis | **3.29** |
| A Separation | **3.14** |

| Movie | Ratings Adjusted With User Average |
|---|---|
| Son of Saul | n/a |
| Winter's Bone | n/a |
| A Serious Man | n/a |
| Inception | **4.88** |
| **Ida** | **4.83** |
| Black Swan | **4.61** |
| Inglourious Basterds | **4.52** |
| The Grand Budapest Hotel | **4.51** |
| **The Imitation Game** | **4.48** |
| Call Me by Your Name | **4.43** |
| Three Billboards Outside Ebbing, Missouri | **4.41** |
| The Wolf of Wall Street | **4.39** |
| The Social Network | **4.28** |
| Life of Pi | **4.22** |
| Dunkirk | **4.21** |
| The Tree of Life | **4.14** |
| Moonlight | **4.12** |
| True Grit | **4.10** |
| 12 Years a Slave | **4.06** |
| Avatar | **3.99** |
| Argo | **3.96** |
| **The White Ribbon** | **3.94** |
| The Secret in Their Eyes | **3.92** |
| Arrival | **3.92** |
| Up in the Air | **3.91** |
| Zero Dark Thirty | **3.88** |
| Gravity | **3.85** |
| Blue is the Warmest Colour | **3.83** |
| The Shape of Water | **3.78** |
| Hugo | **3.76** |
| The Descendants | **3.73** |
| A Prophet | **3.69** |
| La La Land | **3.68** |
| Toy Story 3 | **3.62** |
| Leviathan | **3.61** |
| The Fighter | **3.61** |
| **The King's Speech** | **3.60** |
| Spotlight | **3.57** |
| Amour | **3.47** |
| Precious | **3.34** |
| Mad Max: Fury Road | **3.29** |
| Manchester by the Sea | **3.26** |
| Beasts of the Southern Wild | **3.26** |
| Lincoln | **3.24** |
| The Artist | **3.17** |
| Birdman | **3.13** |
| Toni Erdmann | **3.11** |
| Boyhood | **2.89** |
| Inside Llewyn Davis | **2.88** |
| A Separation | **2.44** |

### *Appendix 3: Count of Total Ratings by User*



*Appendix 4: Count of Rating Score Across all User Ratings*




*Appendix 5: Model 1, Item-Item Collaborative Filters with Original Scores*

| user | La La Land | Avatar | Life of Pi | Inception | The Wolf of W | Toy Story 3 | The King's S | The Grand B | Black Swan | The Social Network |
|---|---|---|---|---|---|---|---|---|---|---|
| Fred | 4 | 5 | | 5 | 5 | | 5 | | | |
| Ajaiy | 3 | 2 | 2 | 1 | 2 | 1 | | | | |
| Sripriya | | 4 | 4 | | | | | | | |
| Jinny | 5 | 5 | | 5 | | 4 | 5 | 5 | 3 | |

| | Cosine Similarity | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sim(1,2) | Sim(1,3) | Sim(1,4) | Sim(1,5) | Sim(1,6) | Sim(1,7) | Sim(1,8) | Sim(1,9) | Sim(1,10) |
| Deno1 | 7.07 | 7.07 | 7.07 | 7.07 | 7.07 | 7.07 | 7.07 | 7.07 | 7.07 |
| Deno2 | 8.37 | 4.47 | 7.14 | 5.39 | 4.12 | 7.07 | 5.00 | 3.00 | 0.00 |
| **CosineM1** | 0.7736797 | 0.749548 | 0.7193332 | 0.6840894 | 0.6710463 | 0.6621177 | 0.6547967 | 0.6451648 | 0.6082978 |
| **CosineM2** | 0.7367459 | 0.6869723 | 0.6864672 | 0.6538814 | 0.6451429 | 0.6154741 | 0.6005972 | 0.59323 | 0.5857671 |
| **CosineM3** | 0.7533215 | 0.7474057 | 0.727957 | 0.6962558 | 0.6930729 | 0.6718697 | 0.6708153 | 0.6460261 | 0.6166931 |

| user | Movie 1 | Movie 2 | Movie 3 |
|---|---|---|---|
| Fred | 2.30 | 2.45 | 2.45 |
| Ajaiy | 0.94 | 1.24 | 1.13 |
| Sripriya | 0.99 | 0.92 | 0.90 |
| Jinny | 3.03 | 1.63 | 1.58 |

### *Appendix 6: Model 2, Item-Item Collaborative Filters with Mean Centered*

| user | Movie 1 | Precious | The White Ri | Hugo | Inception | Black Swan | Avatar | Toy Story 3 | Ida | The Shape of Water |
|---|---|---|---|---|---|---|---|---|---|---|
| Fred | 0 | 0 | 0 | 0 | 1.333333333 | 0 | 1 | 0 | 0 | 0 |
| Ajaiy | -1 | 0 | 0 | 0 | -2.66666667 | 0 | -2 | -1.5 | 0 | 0 |
| Sripriya | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Jinny | 1 | 0 | 0 | 0 | 1.333333333 | 0 | 1 | 1.5 | 0 | 0 |

| user | LalaLand | Precious | The White Ri | Hugo | Inception | Black Swan | Avatar | Toy Story 3 | Ida | The Shape of Water |
|---|---|---|---|---|---|---|---|---|---|---|
| Fred | 4 | | | | 5 | | 5 | | | 4 |
| Ajaiy | 3 | | | | 1 | | 2 | 1 | | |
| Sripriya | | | | | | | 4 | | | |
| Jinny | 5 | | | | 5 | 3 | 5 | 4 | | |
| Average | 4 | #DIV/0! | #DIV/0! | #DIV/0! | 3.666666667 | 3 | 4 | 2.5 | #DIV/0! | 4 |

| | Cosine Similarity | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sim(1,2) | Sim(1,3) | Sim(1,4) | Sim(1,5) | Sim(1,6) | Sim(1,7) | Sim(1,8) | Sim(1,9) | Sim(1,10) |
| Cosine1 | 0.075388 | 0.065263 | 0.057097 | 0.046491 | 0.031059 | 0.030489 | 0.030136 | 0.028905 | 0.025063 |
| AbsCosine1 | 0.075388 | 0.065263 | 0.057097 | 0.046491 | 0.031059 | 0.030489 | 0.030136 | 0.028905 | 0.025063 |
| Cosine2 | 0.15 | 0.14 | 0.08 | 0.07 | 0.06 | 0.02 | 0.02 | 0.02 | 0.02 |
| AbsCosine2 | 0.149427466 | 0.138117955 | 0.076679892 | 0.071770284 | 0.064097584 | 0.022519338 | 0.019969038 | 0.018565304 | 0.015466986 |
| Cosine3 | 0.157533979 | 0.133486352 | 0.102798245 | 0.088306186 | 0.086229544 | 0.060527063 | 0.053826336 | 0.04507807 | 0.040924758 |
| AbsCosine3 | 0.157533979 | 0.133486352 | 0.102798245 | 0.088306186 | 0.086229544 | 0.060527063 | 0.053826336 | 0.04507807 | 0.040924758 |

| user | Movie 1 | Movie 2 | Movie 3 |
|---|---|---|---|
| Fred | 4.24 | 3.03 | 2.79 |
| Ajaiy | 3.41 | 2.94 | 2.03 |
| Sripriya | 4.00 | 3.00 | 2.50 |
| Jinny | 4.35 | 3.03 | 2.68 |

### *Appendix 7: Model 3, Item-Item Based Collaborative Filters with Z-Scored*

(a) Step 1: Standardized Z-score of ratings

| | X.The.Social.Network. | X.A.Prophet. | X.Amour. | X.The.King.s.Speech. | X.La.La.Land. | X.Boyhood. | X.Inception. | X.A.Separation. | X.The.Artist. | X.The.White.Ribbon. |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.6440915 | 0.00000000 | 0.0000000 | 0.0000000 |
| 2 | -0.1428571 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.6440915 | 0.00000000 | 0.0000000 | 0.0000000 |
| 3 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | -0.4337759 | 0.00000000 | 0.4629101 | 0.0000000 |
| 4 | 1.0238095 | 0.0000000 | 0.0000000 | 0.0000000 | -1.6552829 | 0.0000000 | -1.5116433 | 0.00000000 | 0.0000000 | 0.0000000 |
| 5 | -0.1428571 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.6022941 | 0.6440915 | 0.00000000 | 0.0000000 | 0.0000000 |
| 6 | 1.0238095 | 0.0000000 | 0.0000000 | 1.0685803 | -1.6552829 | 0.0000000 | -0.4337759 | 0.00000000 | 0.0000000 | 0.0000000 |
| 7 | 1.0238095 | 0.0000000 | 0.0000000 | 1.0685803 | 1.0278773 | -0.4380321 | -0.4337759 | 0.00000000 | 0.0000000 | 0.0000000 |
| 8 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.00000000 | 0.0000000 | 0.0000000 |
| 9 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 1.0278773 | 0.0000000 | 0.6440915 | 0.00000000 | 0.0000000 | 0.0000000 |
| 10 | -0.1428571 | 0.0000000 | 0.0000000 | 0.0000000 | 0.1334906 | 0.0000000 | -0.4337759 | 0.00000000 | 0.0000000 | 0.0000000 |
| 11 | -0.1428571 | 0.0000000 | -0.5581563 | 0.0000000 | 0.0000000 | 0.0000000 | -0.4337759 | 0.00000000 | 0.0000000 | 0.0000000 |
| 12 | -0.1428571 | 0.0000000 | 0.0000000 | -0.8257211 | 0.1334906 | 0.0000000 | -0.4337759 | 0.00000000 | 0.4629101 | 0.0000000 |
| 13 | -0.1428571 | 0.0000000 | 0.0000000 | -0.8257211 | -0.7608962 | 0.0000000 | 0.6440915 | 0.00000000 | 0.0000000 | 0.0000000 |
| 14 | 0.0000000 | 0.7245688 | 0.0000000 | 0.0000000 | -0.7608962 | 0.0000000 | -0.4337759 | 0.00000000 | 0.0000000 | 0.0000000 |
| 15 | 1.0238095 | 0.0000000 | 0.0000000 | -1.7728718 | -0.7608962 | 0.6022941 | 0.6440915 | 1.18019369 | 0.0000000 | 0.0000000 |
| 16 | 1.0238095 | 0.0000000 | 0.0000000 | 0.0000000 | 0.1334906 | 0.0000000 | 0.6440915 | 0.00000000 | 0.0000000 | 0.0000000 |
| 17 | -0.1428571 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | -0.4337759 | 0.00000000 | 0.0000000 | 0.0000000 |
| 18 | -2.4761905 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 1.6426202 | 0.6440915 | 0.00000000 | 0.0000000 | 0.0000000 |
| 19 | 1.0238095 | -1.2076147 | -0.5581563 | -1.7728718 | -0.7608962 | 0.0000000 | 0.6440915 | 0.00000000 | 0.0000000 | 0.2286648 |
| 20 | -0.1428571 | -1.2076147 | 0.0000000 | -1.7728718 | 1.0278773 | 0.0000000 | 0.6440915 | 0.00000000 | -0.4629101 | 0.2286648 |
| 21 | 0.0000000 | 0.0000000 | 1.3023647 | 1.0685803 | -2.5496697 | 0.0000000 | 0.0000000 | 0.00000000 | 0.0000000 | 0.0000000 |
| 22 | -1.3095238 | 0.0000000 | 0.0000000 | 0.0000000 | 1.0278773 | 0.0000000 | -1.5116433 | 0.00000000 | 0.0000000 | 0.0000000 |
| 23 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | -3.6673781 | 0.00000000 | 0.0000000 | 0.0000000 |
| 24 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | -0.7608962 | 0.0000000 | -2.5895107 | 0.00000000 | 0.0000000 | 0.0000000 |
| 25 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 | -0.7608962 | 0.0000000 | -0.4337759 | 0.00000000 | 0.0000000 | 0.0000000 |

(b) Step 2: Similarity matrix (i.e., Cosine matrix) - Sample data

| | X.The.Social.Network. | X.A.Prophet. | X.Amour. | X.The.King.s.Speech. | X.La.La.Land. | X.Boyhood. | X.Inception. | X.A.Separation. | X.The.Artist. |
|---|---|---|---|---|---|---|---|---|---|
| X.The.Social.Network. | 1.000000000 | 0.026666032 | 0.055839910 | −0.025873397 | −0.191756801 | −0.3258071967 | 0.165821054 | −0.009425320 | 0.0943898830 |
| X.A.Prophet. | 0.026666032 | 1.000000000 | 0.464233584 | 0.432650968 | 0.002999977 | 0.0294529666 | 0.010932820 | 0.043983222 | 0.1126872339 |
| X.Amour. | 0.055839910 | 0.464233584 | 1.000000000 | 0.318117078 | −0.009273278 | −0.0176082284 | 0.130794504 | −0.066656980 | 0.1818202630 |
| X.The.King.s.Speech. | −0.025873397 | 0.432650968 | 0.318117078 | 1.000000000 | 0.022514942 | 0.1063154462 | 0.008895721 | 0.074759596 | −0.0413377603 |
| X.La.La.Land. | −0.191756801 | 0.002999977 | −0.009273278 | 0.022514942 | 1.000000000 | 0.1296745355 | 0.158465758 | −0.100605781 | 0.0238331688 |
| X.Boyhood. | −0.325807197 | 0.029452967 | −0.017608228 | 0.106315446 | 0.129674536 | 1.0000000000 | 0.052551051 | 0.514666902 | −0.2210436238 |
| X.Inception. | 0.165821054 | 0.010932820 | 0.130794504 | 0.008895721 | 0.158465758 | 0.0525510509 | 1.000000000 | −0.013316141 | 0.0367343142 |
| X.A.Separation. | −0.009425320 | 0.043983222 | −0.066656980 | 0.074759596 | −0.100605781 | 0.5146669021 | −0.013316141 | 1.000000000 | −0.2973808571 |
| X.The.Artist. | 0.094389883 | 0.112687234 | 0.181820263 | −0.041337760 | 0.023833169 | −0.2210436238 | 0.036734314 | −0.297380857 | 0.9999999999 |
| X.The.White.Ribbon. | 0.072653421 | 0.169742853 | 0.320883647 | −0.106858601 | 0.069874483 | −0.1001625046 | 0.013283600 | −0.067417262 | 0.6362090102 |
| X.Zero.Dark.Thirty. | 0.028079601 | 0.205399513 | 0.142824820 | 0.160638077 | 0.009076173 | 0.1425718677 | 0.188436752 | 0.002943247 | 0.1783596050 |
| X.Avatar. | 0.127667101 | 0.096270785 | 0.189165945 | 0.208058492 | 0.113009194 | −0.0111780889 | 0.366982959 | −0.066284034 | 0.1292189781 |
| X.Spotlight. | −0.086817831 | 0.057760015 | −0.030584276 | 0.162432689 | 0.161347978 | 0.2305631489 | 0.033648726 | 0.179672265 | 0.1995717947 |
| X.Precious. | 0.037068593 | 0.196931590 | 0.510829468 | 0.120531511 | 0.083140305 | 0.0759301468 | 0.061010124 | −0.072685476 | 0.2420910129 |
| X.The.Tree.of.Life. | 0.006443851 | 0.296324595 | 0.645201657 | 0.125979043 | −0.071161535 | 0.1571625415 | 0.087428217 | −0.024659297 | 0.3702156406 |
| X.12.Years.a.Slave. | 0.080191628 | 0.232379001 | 0.289429146 | 0.236791769 | 0.066924644 | −0.1387101522 | −0.006816131 | −0.193932695 | 0.4364357804 |
| X.Blue.is.the.Warmest.Colour. | 0.294082851 | 0.057298985 | 0.019463498 | −0.017722192 | −0.054447947 | −0.1387227833 | 0.054622469 | −0.069789988 | 0.3107365886 |
| X.Up.in.the.Air. | 0.001570199 | −0.059843917 | 0.045583928 | 0.047662483 | −0.036493631 | 0.0002563754 | −0.073434107 | 0.064054590 | 0.5088387036 |
| X.Inglourious.Basterds. | 0.098212510 | 0.285603890 | 0.162375254 | 0.177533422 | 0.057201451 | −0.1535791065 | −0.002132869 | −0.025748290 | 0.4550474624 |
| X.Mad.Max..Fury.Road. | 0.270839873 | −0.166033763 | −0.091975838 | 0.043345410 | −0.175350166 | 0.0759985693 | 0.045211981 | 0.271697138 | 0.0596331010 |
| X.Moonlight. | 0.062654926 | 0.145208350 | 0.383848308 | 0.052745452 | −0.009231830 | −0.0351356937 | 0.107419351 | 0.051774092 | 0.4169450693 |
| X.Birdman. | 0.067858515 | 0.173079205 | 0.240956152 | 0.156474878 | 0.058119197 | 0.0614893077 | 0.028265288 | 0.049698716 | 0.0894346006 |
| X.Manchester.by.the.Sea. | −0.061572204 | −0.083809270 | −0.006901477 | 0.047459080 | 0.063816465 | 0.1066362465 | 0.004002340 | 0.044337531 | −0.0622460935 |

(c) Step 3: Predicted top 15 neighbors of each movie - Sample data

| | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 |
|---|---|---|---|---|---|---|---|---|
| X.The.Social.Network. | X.The.Social.Network. | X.Blue.is.the.Warmes… | X.Dunkirk. | X.Mad.Max..Fury.Ro… | X.The.Wolf.of.Wall.Stre… | X.Inception. | X.Black.Swan. | X.Avatar. |
| X.A.Prophet. | X.A.Prophet. | X.Amour. | X.The.King.s.Speech. | X.The.Fighter. | X.The.Tree.of.Life. | X.Inglourious.Basterds. | X.Three.Billboards.Out… | X.Toni.Erdmann. |
| X.Amour. | X.Amour. | X.The.Tree.of.Life. | X.Precious. | X.A.Prophet. | X.Moonlight. | X.The.Fighter. | X.Ida. | X.The.White.Ribb |
| X.The.King.s.Speech. | X.The.King.s.Speech. | X.A.Prophet. | X.Amour. | X.12.Years.a.Slave. | X.Avatar. | X.Black.Swan. | X.Toni.Erdmann. | X.True.Grit. |
| X.La.La.Land. | X.La.La.Land. | X.Arrival. | X.Dunkirk. | X.Life.of.Pi. | X.Spotlight. | X.Inception. | X.Boyhood. | X.Ida. |
| X.Boyhood. | X.Boyhood. | X.A.Separation. | X.The.Descendants. | X.Hugo. | X.True.Grit. | X.Spotlight. | X.Call.Me.by.Your.Name. | X.The.Tree.of.Life |
| X.Inception. | X.Inception. | X.Toy.Story.3. | X.Gravity. | X.Dunkirk. | X.Avatar. | X.The.Wolf.of.Wall.Str… | X.The.Imitation.Game. | X.Arrival. |
| X.A.Separation. | X.A.Separation. | X.Boyhood. | X.Call.Me.by.Your.N… | X.Inside.Llewyn.Dav… | X.The.Descendants. | X.Mad.Max..Fury.Road. | X.True.Grit. | X.Toni.Erdmann. |
| X.The.Artist. | X.The.Artist. | X.Ida. | X.The.White.Ribbon. | X.Beasts.of.the.Sou… | X.Leviathan. | X.Up.in.the.Air. | X.Inglourious.Basterds. | X.Lincoln. |
| X.The.White.Ribbon. | X.The.White.Ribbon. | X.Beasts.of.the.South… | X.Ida. | X.The.Secret.in.The… | X.The.Artist. | X.Leviathan. | X.The.Fighter. | X.Moonlight. |
| X.Zero.Dark.Thirty. | X.Zero.Dark.Thirty. | X.Argo. | X.Toni.Erdmann. | X.The.Fighter. | X.Gravity. | X.Arrival. | X.Ida. | X.Dunkirk. |
| X.Avatar. | X.Avatar. | X.Gravity. | X.Inception. | X.The.Wolf.of.Wall.… | X.Dunkirk. | X.Toy.Story.3. | X.Life.of.Pi. | X.The.King.s.Spee |
| X.Spotlight. | X.Spotlight. | X.Inglourious.Basterds. | X.Birdman. | X.True.Grit. | X.The.Grand.Budapest.… | X.12.Years.a.Slave. | X.Toni.Erdmann. | X.Dunkirk. |
| X.Precious. | X.Precious. | X.The.Tree.of.Life. | X.Toni.Erdmann. | X.Amour. | X.Ida. | X.The.Fighter. | X.True.Grit. | X.Birdman. |
| X.The.Tree.of.Life. | X.The.Tree.of.Life. | X.Amour. | X.Precious. | X.The.White.Ribbon. | X.Ida. | X.Beasts.of.the.South… | X.The.Secret.in.Their.E… | X.Moonlight. |
| X.12.Years.a.Slave. | X.12.Years.a.Slave. | X.Inglourious.Basterds. | X.Toni.Erdmann. | X.Ida. | X.The.Artist. | X.The.Fighter. | X.Birdman. | X.True.Grit. |
| X.Blue.is.the.Warmest.Col… | X.Blue.is.the.Warme… | X.The.Artist. | X.The.Social.Network. | X.Dunkirk. | X.Beasts.of.the.Southe… | X.Ida. | X.Leviathan. | X.The.White.Ribb |
| X.Up.in.the.Air. | X.Up.in.the.Air. | X.The.Secret.in.Their… | X.The.Artist. | X.The.White.Ribbon. | X.Beasts.of.the.Southe… | X.Moonlight. | X.Leviathan. | X.Mad.Max..Fury. |
| X.Inglourious.Basterds. | X.Inglourious.Baster… | X.12.Years.a.Slave. | X.Toni.Erdmann. | X.The.Artist. | X.True.Grit. | X.The.Fighter. | X.Spotlight. | X.Three.Billboard |
| X.Mad.Max..Fury.Road. | X.Mad.Max..Fury.Ro… | X.Up.in.the.Air. | X.Dunkirk. | X.Inside.Llewyn.Dav… | X.A.Separation. | X.The.Social.Network. | X.Gravity. | X.Moonlight. |
| X.Moonlight. | X.Moonlight. | X.The.Secret.in.Their… | X.The.White.Ribbon. | X.Beasts.of.the.Sou… | X.Ida. | X.Up.in.the.Air. | X.The.Artist. | X.The.Tree.of.Life |
| X.Birdman. | X.Birdman. | X.12.Years.a.Slave. | X.Precious. | X.Toni.Erdmann. | X.Black.Swan. | X.Spotlight. | X.The.Fighter. | X.Up.in.the.Air. |
| X.Manchester.by.the.Sea. | X.Manchester.by.the… | X.True.Grit. | X.Inside.Llewyn.Davis. | X.The.Descendants. | X.Call.Me.by.Your.Name. | X.Toni.Erdmann. | X.Three.Billboards.Out… | X.Moonlight. |
| X.Lincoln. | X.Lincoln. | X.Ida. | X.The.Artist. | X.Beasts.of.the.Sou… | X.The.White.Ribbon. | X.Leviathan. | X.The.Tree.of.Life. | X.Toni.Erdmann. |
| X.Hugo. | X.Hugo. | X.Up.in.the.Air. | X.Boyhood. | X.12.Years.a.Slave. | X.The.Descendants. | X.Ida. | X.The.Artist. | X.Lincoln. |

(d) R Codes of model 1 - Key highlights

## Cosine calculation

```
# create cosine function
getCosine <- function(x,y)
{
  this.cosine <- sum(x*y) / (sqrt(sum(x*x)) * sqrt(sum(y*y)))
  return(this.cosine)
}
```

## Similarity matrix

```
for(i in 1:ncol(hw4_new)) {
  # Loop through the columns for each column
  for(j in 1:ncol(hw4_new)) {
    # Fill in placeholder with cosine similarities
    similarity[i,j] <- getCosine(as.matrix(hw4_new[i]),as.matrix(hw4_new[j]))
    absCosine[i,j] <- abs(similarity[i,j])
  }
}
View(similarity)
View(absCosine)
```

## Top neighbors selection

```
# get the top 15 neighbors
neighbours <- matrix(NA, nrow=ncol(similarity),ncol=15,dimnames=list(colnames(similarity)))
for(i in 1:ncol(hw4_new))
{
  neighbours[i,] <- t(rownames(head(n=15,as.matrix(similarity[order(similarity[,i],decreasing=TRUE),][,i]))))
}
View(neighbours)
```

*Appendix 8: User-User Collaborative Filtering*

**Structure of Input Data :** We have combined both the class data as well as DBMI Data. Totally we have 378 users, starting from user id 0 to 377.

| userid | title | rating |
|---|---|---|
| 0 | The Social Network | 5 |
| 0 | The King's Speech | 3 |
| 0 | La La Land | 4 |
| 0 | Inception | 5 |
| 0 | Avatar | 5 |
| 0 | 12 Years a Slave | 2 |
| 0 | Up in the Air | 5 |
| 0 | Birdman | 4 |
| 0 | Lincoln | 3 |
| 0 | Hugo | 4 |
| 0 | Gravity | 5 |
| 0 | The Wolf of Wall Street | 5 |
| 0 | Life of Pi | 5 |
| 1 | The Social Network | 3 |
| 1 | The King's Speech | 4 |
| 1 | La La Land | 5 |
| 1 | Boyhood | 4 |
| 1 | Inception | 5 |
| 1 | Avatar | 3 |
| 1 | Up in the Air | 3 |

**Step 0 : Inputting and Structuring data**

```
df = pd.read_excel("data.xlsx")
df.head()
matrix_data = df.pivot_table(index='userid',columns='title',values='rating')
matrix_data.head()
```

**Step 1 : Normalizing the ratings to remove User Bias**

We do so by subtracting the average rating from each individual value, this gives a negative rating value for those below average and positive for ones above.

```
mat = matrix_data.subtract(matrix_data.mean(axis=1), axis = 'rows')
mat.head()
```

**Step 2 : Defining similarity between users**

This is done in two methods, one using Pearson Correlation and the other using Cosine Similarity.

**Method 1 : Cosine Similarity :** Calculates the cosine of the angle between the two vectors that represent a pair of users.

```
In [5]:  # User similarity matrix using cosine similarity
         user_similarity_cosine = cosine_similarity(mat.fillna(0))
         user_similarity_cosine
```

```
Out[5]: array([[ 1.        , -0.0994042 ,  0.04651303, ...,  0.23534666,
                 0.04518771,  0.10963225],
               [-0.0994042 ,  1.        , -0.20246457, ..., -0.07663063,
                -0.23659147, -0.15111763],
               [ 0.04651303, -0.20246457,  1.        , ...,  0.53785287,
                -0.15122507,  0.        ],
               ...,
               [ 0.23534666, -0.07663063,  0.53785287, ...,  1.        ,
                -0.05422456, -0.12677314],
               [ 0.04518771, -0.23659147, -0.15122507, ..., -0.05422456,
                 1.        , -0.10693227],
               [ 0.10963225, -0.15111763,  0.        , ..., -0.12677314,
                -0.10693227,  1.        ]])
```

The downside of this method is that it cannot accept N/A values, and hence NA values need to be encoded as 0, this would misrepresent the ratings and hence we did not use this method in our predictions, it is only for illustration purposes.

**Method 2 : Pearson Correlation :** Calculated correlation values between pairs of users. Correlation values range from -1 to +1. Higher the value, the more similar the users are.

```
# User similarity matrix using Pearson correlation
user_similarity = mat.T.corr()
user_similarity.head()
```

]:

| userid | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|---|
| userid | | | | | | | | | |
| 0 | 1.000000 | -0.198680 | -1.0 | -0.316228 | 0.218218 | 0.566947 | 0.577350 | -5.222330e-01 | 0.03 |
| 1 | -0.198680 | 1.000000 | NaN | -0.174078 | 0.192308 | 0.011606 | -0.816497 | -6.500800e-17 | 0.20 |
| 2 | -1.000000 | NaN | 1.0 | NaN | NaN | NaN | NaN | NaN | |
| 3 | -0.316228 | -0.174078 | NaN | 1.000000 | -0.292770 | -0.727607 | NaN | 8.807048e-01 | -0.47 |
| 4 | 0.218218 | 0.192308 | NaN | -0.292770 | 1.000000 | 0.785714 | NaN | -4.000000e-01 | -0.32 |

5 rows × 377 columns

**Step 3 :** Here we identify 'n' users similar to the user of interest

In [7]:
```
# Number of similar users
n = 30

# User similarity threshold
user_similarity_threshold = 0.3

# Get top n similar users
similar_users = user_similarity[
    user_similarity[picked_userid]>user_similarity_threshold][picked_userid].sort_values(ascending=False)[:n]

# Print out top n similar users
print(f'The similar users for user {picked_userid} are', similar_users)
```

```
The similar users for user 362 are userid
377    1.000000
281    1.000000
329    1.000000
192    1.000000
351    1.000000
69     1.000000
49     1.000000
165    1.000000
19     1.000000
288    0.918559
110    0.880705
361    0.880705
352    0.871315
366    0.836660
84     0.790569
218    0.763763
282    0.763763
256    0.725866
```

**Step 4 :** Using the similar users list, a movie score is predicted for all movies watched by the similar users based on the similarity index, this score is then used to compute a predicted rating for our user of interest by adding the average rating of the user of interest to movie score.

In [11]:
```
# A dictionary to store item scores
item_score = {}

# Loop through items
for i in similar_user_movies.columns:
  # Get the ratings for movie i
  movie_rating = similar_user_movies[i]
  # Create a variable to store the score
  total = 0
  # Create a variable to store the number of scores
  count = 0
  # Loop through similar users
  for u in similar_users.index:
    # If the movie has rating
    if pd.isna(movie_rating[u]) == False:
      # Score is the sum of user similarity score multiply by the movie rating
      score = similar_users[u] * movie_rating[u]
      # Add the score to the total score for the movie so far
      total += score
      # Add 1 to the count
      count +=1
  # Get the average score for the item
  item_score[i] = total / count

# Convert dictionary to pandas dataframe
item_score = pd.DataFrame(item_score.items(), columns=['movie', 'movie_score'])

# Sort the movies by score
ranked_item_score = item_score.sort_values(by='movie_score', ascending=False)

# Select top m movies
m = 50
ranked_item_score.head(m)
```

Out[11]:

| | movie | movie_score |
|---|---|---|
| 2 | A Separation | 0.798630 |
| 19 | La La Land | 0.433773 |
| 30 | The Grand Budapest Hotel | 0.406546 |
| 17 | Inception | 0.396069 |
| 3 | Amour | 0.395810 |
| 6 | Avatar | 0.387676 |
| 16 | Ida | 0.363048 |

```
# Average rating for the picked user
avg_rating = matrix_data[matrix_data.index == picked_userid].T.mean()[picked_userid]

# Print the average movie rating for user 1
print(f'The average movie rating for user {picked_userid} is {avg_rating:.2f}')
```

```
The average movie rating for user 332 is 4.20
```

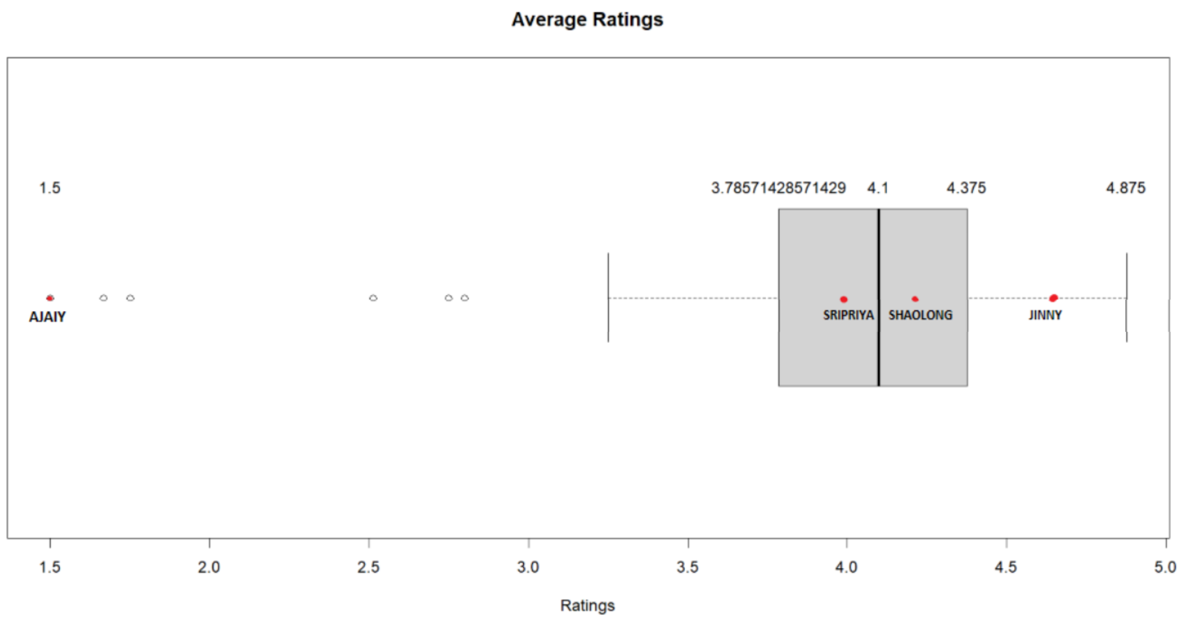
```
# Calcuate the predicted rating
ranked_item_score['predicted_rating'] = ranked_item_score['movie_score'] + avg_rating

ranked_item_score.head(m)
```

]:

| | movie | movie_score | predicted_rating |
|---|---|---|---|
| 6 | The Grand Budapest Hotel | 1.285714 | 5.485714 |
| 3 | Inception | 0.542857 | 4.742857 |
| 0 | Avatar | 0.309524 | 4.509524 |
| 8 | The Wolf of Wall Street | 0.285714 | 4.485714 |
| 9 | Toy Story 3 | 0.066667 | 4.266667 |
| 5 | Life of Pi | 0.042857 | 4.242857 |
| 4 | La La Land | -0.200000 | 4.000000 |
| 7 | The Shape of Water | -0.433333 | 3.766667 |
| 2 | Gravity | -0.714286 | 3.485714 |
| 1 | Birdman | -1.714286 | 2.485714 |

*Appendix 9 : Boxplot for all users average ratings with our team members highlighted to visualize rating leniency.*




*Appendix 10a: Item-Item Cosine Similarity of Actual Movies Ratings, Class Data (Partial of 50x50)*

| | The Social N | A Prophet | Amour | The King's S | La La Land | Boyhood | Inception | A Separatio | The Artist | The White R | Zero Dark Tl | Avatar | Spotlight | Precious | The Tree of | 12 Years a | Blue is the V | Son of Saul | Up in the Air |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| The Social N | 1 | 0.311725 | 0.286864 | 0.456811 | 0.562902 | 0.395197 | 0.70833 | 0.199984 | 0.250169 | 0.270678 | 0.707969 | 0.623574 | 0.47518 | 0.236087 | 0.339673 | 0.548338 | 0.372221 | NA | 0.38579 |
| A Prophet | 0.311725 | 1 | 0.416691 | 0.166013 | 0.259349 | 0.306412 | 0.301432 | 0.274247 | 0.444281 | 0.530541 | 0.226652 | 0.250797 | 0.221178 | 0.348977 | 0.378299 | 0.152654 | 0.197238 | NA | 0.281089 |
| Amour | 0.286864 | 0.416691 | 1 | 0.394901 | 0.280838 | 0.20774 | 0.261447 | 0.230535 | 0.271151 | 0.623254 | 0.310886 | 0.333358 | 0.303854 | 0.565529 | 0.726503 | 0.393522 | 0.32331 | NA | 0.223398 |
| The King's S | 0.456811 | 0.166013 | 0.394901 | 1 | 0.653784 | 0.294491 | 0.556918 | 0.257086 | 0.155831 | 0.311629 | 0.392281 | 0.638545 | 0.476426 | 0.351618 | 0.479076 | 0.578874 | 0.264809 | NA | 0.329882 |
| La La Land | 0.562902 | 0.259349 | 0.280838 | 0.653784 | 1 | 0.404249 | 0.751457 | 0.192922 | 0.222048 | 0.308465 | 0.47567 | 0.769086 | 0.516463 | 0.338197 | 0.371573 | 0.54804 | 0.34792 | NA | 0.304196 |
| Boyhood | 0.395197 | 0.306412 | 0.20774 | 0.294491 | 0.404249 | 1 | 0.470078 | 0.550498 | 0.258008 | 0.23296 | 0.386979 | 0.39555 | 0.278364 | 0.273368 | 0.26142 | 0.460945 | 0.329931 | NA | 0.41035 |
| Inception | 0.70833 | 0.301432 | 0.261447 | 0.556918 | 0.751457 | 0.470078 | 1 | 0.259842 | 0.338492 | 0.294738 | 0.651036 | 0.816309 | 0.441343 | 0.31776 | 0.330745 | 0.536792 | 0.34281 | NA | 0.392622 |
| A Separatio | 0.199984 | 0.274247 | 0.230535 | 0.257086 | 0.192922 | 0.550498 | 0.259842 | 1 | 0.274367 | 0.402309 | 0.251397 | 0.193519 | 0.291234 | 0.307408 | 0.300972 | 0.357833 | 0.278261 | NA | 0.403767 |
| The Artist | 0.250169 | 0.444281 | 0.271151 | 0.155831 | 0.222048 | 0.258008 | 0.338492 | 0.274367 | 1 | 0.520479 | 0.311953 | 0.309882 | 0.236172 | 0.389358 | 0.439439 | 0.273844 | 0.281953 | NA | 0.452672 |
| The White R | 0.270678 | 0.530541 | 0.623254 | 0.311629 | 0.308465 | 0.23296 | 0.294738 | 0.402309 | 0.520479 | 1 | 0.332756 | 0.322335 | 0.425067 | 0.709682 | 0.660098 | 0.399267 | 0.470177 | NA | 0.368321 |
| Zero Dark Tl | 0.707969 | 0.226652 | 0.310886 | 0.392281 | 0.47567 | 0.386979 | 0.651036 | 0.251397 | 0.311953 | 0.332756 | 1 | 0.597099 | 0.442594 | 0.352342 | 0.334511 | 0.490999 | 0.396897 | NA | 0.456775 |
| Avatar | 0.623574 | 0.250797 | 0.333358 | 0.638545 | 0.769086 | 0.39555 | 0.816309 | 0.193519 | 0.309882 | 0.322335 | 0.597099 | 1 | 0.447028 | 0.349611 | 0.404168 | 0.594862 | 0.340935 | NA | 0.426835 |
| Spotlight | 0.47518 | 0.221178 | 0.303854 | 0.476426 | 0.516463 | 0.278364 | 0.441343 | 0.291234 | 0.236172 | 0.425067 | 0.442594 | 0.447028 | 1 | 0.330932 | 0.321492 | 0.445524 | 0.385514 | NA | 0.259179 |
| Precious | 0.236087 | 0.348977 | 0.565529 | 0.351618 | 0.338197 | 0.273368 | 0.31776 | 0.307408 | 0.389358 | 0.709682 | 0.352342 | 0.349611 | 0.330932 | 1 | 0.779437 | 0.395697 | 0.454698 | NA | 0.424727 |
| The Tree of | 0.339673 | 0.378299 | 0.726503 | 0.479076 | 0.371573 | 0.26142 | 0.330745 | 0.300972 | 0.439439 | 0.660098 | 0.334511 | 0.404168 | 0.321492 | 0.779437 | 1 | 0.475317 | 0.429456 | NA | 0.453589 |
| 12 Years a | 0.548338 | 0.152654 | 0.393522 | 0.578874 | 0.54804 | 0.460945 | 0.536792 | 0.357833 | 0.273844 | 0.399267 | 0.490999 | 0.594862 | 0.445524 | 0.395697 | 0.475317 | 1 | 0.47241 | NA | 0.363878 |
| Blue is the V | 0.372221 | 0.197238 | 0.32331 | 0.264809 | 0.34792 | 0.329931 | 0.34281 | 0.278261 | 0.281953 | 0.470177 | 0.396897 | 0.340935 | 0.385514 | 0.454698 | 0.429456 | 0.47241 | 1 | NA | 0.396043 |
| Son of Saul | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | 1 | NA |
| Up in the Air | 0.38579 | 0.281088 | 0.223398 | 0.329882 | 0.304196 | 0.41035 | 0.392622 | 0.403767 | 0.452672 | 0.368321 | 0.456775 | 0.426835 | 0.259179 | 0.424727 | 0.453589 | 0.363878 | 0.396043 | NA | 1 |

*Appendix 10b: Item-Item Cosine Similarity of Normalized Movies Ratings, Class Data (Partial of 50x50)*

| | The Social N | A Prophet | Amour | The King's S | La La Land | Boyhood | Inception | A Separatio | The Artist | The White R | Zero Dark Tl | Avatar | Spotlight | Precious | The Tree of | 12 Years a | Blue is the V | Son of Saul | Up in the Air |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| The Social N | 1 | 0.097834 | 0.045761 | −0.06299 | −0.1568 | −0.33901 | 0.113245 | −0.19216 | 0.180167 | 0.218485 | −0.05071 | 0.063951 | −0.25115 | −0.01139 | 0.1432 | 0.013651 | 0.195703 | NaN | −0.01497 |
| A Prophet | 0.097834 | 1 | 0.273392 | 0.151256 | 0.015029 | −0.1502 | 0.070455 | −0.33862 | 0.472991 | 0.18426 | 0.194351 | 0.258515 | 0.022496 | 0.162155 | 0.161518 | 0.192979 | 0.127401 | NaN | −0.05526 |
| Amour | 0.045761 | 0.273392 | 1 | 0.048932 | 0.015657 | −0.24587 | 0.051496 | −0.41219 | 0.394309 | 0.374774 | 0.081798 | 0.157896 | −0.10853 | 0.284212 | 0.348642 | 0.265267 | 0.104925 | NaN | 0.128771 |
| The King's S | −0.06299 | 0.151256 | 0.048932 | 1 | −0.03096 | 0.056625 | −0.29384 | 0.008958 | 0.09502 | −0.20809 | 0.132981 | 0.162425 | 0.037461 | −0.00742 | −0.18775 | 0.148147 | 0.008049 | NaN | 0.004471 |
| La La Land | −0.1568 | 0.015029 | 0.015657 | −0.03096 | 1 | 0.055999 | −0.02746 | −0.12566 | 0.02276 | 0.080707 | −0.11174 | 0.041973 | −0.03998 | 0.045124 | −0.04551 | −0.0373 | 0.037558 | NaN | −0.058 |
| Boyhood | −0.33901 | −0.1502 | −0.24587 | 0.056625 | 0.055999 | 1 | −0.2405 | 0.377385 | −0.27493 | −0.36282 | 0.079882 | −0.15804 | 0.273557 | −0.05172 | −0.15188 | −0.16828 | −0.28025 | NaN | −0.16514 |
| Inception | 0.113245 | 0.070455 | 0.051496 | −0.29384 | −0.02746 | −0.2405 | 1 | −0.31723 | 0.133582 | 0.188747 | −0.04275 | 0.011161 | −0.18567 | 0.060548 | 0.196676 | 0.007639 | 0.104624 | NaN | 0.000307 |
| A Separatio | −0.19216 | −0.33862 | −0.41219 | 0.008958 | −0.12566 | 0.377385 | −0.31723 | 1 | −0.44205 | −0.40803 | −0.15104 | −0.30065 | 0.327629 | −0.26996 | −0.49314 | −0.43097 | −0.36114 | NaN | −0.30403 |
| The Artist | 0.180167 | 0.472991 | 0.394309 | 0.09502 | 0.02276 | −0.27493 | 0.133582 | −0.44205 | 1 | 0.435349 | 0.046381 | 0.093816 | 0.005839 | 0.185571 | 0.166814 | 0.339735 | 0.174047 | NaN | 0.305072 |
| The White R | 0.218485 | 0.18426 | 0.374774 | −0.20809 | 0.080707 | −0.36282 | 0.188747 | −0.40803 | 0.435349 | 1 | −0.11127 | 0.116821 | −0.28413 | 0.144008 | 0.482241 | 0.116499 | 0.084538 | NaN | 0.37404 |
| Zero Dark Tl | −0.05071 | 0.194351 | 0.081798 | 0.132981 | −0.11174 | 0.079882 | −0.04275 | −0.15104 | 0.046381 | −0.11127 | 1 | −0.04537 | 0.019502 | 0.035142 | −0.0081 | 0.197545 | 0.025813 | NaN | −0.11521 |
| Avatar | 0.063951 | 0.258515 | 0.157896 | 0.162425 | 0.041973 | −0.15804 | 0.011161 | −0.30065 | 0.093816 | 0.116821 | −0.04537 | 1 | −0.23624 | 0.082609 | 0.24656 | 0.037775 | 0.195354 | NaN | 0.086365 |
| Spotlight | −0.25115 | 0.022496 | −0.10853 | 0.037461 | −0.03998 | 0.273557 | −0.18567 | 0.327629 | 0.005839 | −0.28413 | 0.019502 | −0.23624 | 1 | −0.12313 | −0.45091 | −0.06985 | −0.1627 | NaN | −0.21838 |
| Precious | −0.01139 | 0.162155 | 0.284212 | −0.00742 | 0.045124 | −0.05172 | 0.060548 | −0.26996 | 0.185571 | 0.144008 | 0.035142 | 0.082609 | −0.12313 | 1 | 0.632711 | 0.116229 | 0.064966 | NaN | −0.00915 |
| The Tree of | 0.1432 | 0.161518 | 0.348642 | −0.18775 | −0.04551 | −0.15188 | 0.196676 | −0.49314 | 0.166814 | 0.482241 | −0.0081 | 0.24656 | −0.45091 | 0.632711 | 1 | 0.107042 | 0.203744 | NaN | 0.175144 |
| 12 Years a | 0.013651 | 0.192979 | 0.265267 | 0.148147 | −0.0373 | −0.16828 | 0.007639 | −0.43097 | 0.339735 | 0.116499 | 0.197545 | 0.037775 | −0.06985 | 0.116229 | 0.107042 | 1 | 0.191684 | NaN | 0.331063 |
| Blue is the V | 0.195703 | 0.127401 | 0.104925 | 0.008049 | 0.037558 | −0.28025 | 0.104624 | −0.36114 | 0.174047 | 0.084538 | 0.025813 | 0.195354 | −0.1627 | 0.064966 | 0.203744 | 0.191684 | 1 | NaN | 0.152799 |
| Son of Saul | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 1 | NaN |
| Up in the Air | −0.01497 | −0.05526 | 0.128771 | 0.004471 | −0.058 | −0.16514 | 0.000307 | −0.30403 | 0.305072 | 0.37404 | −0.11521 | 0.086365 | −0.21838 | −0.00915 | 0.175144 | 0.331063 | 0.152799 | NaN | 1 |

*Appendix 11a: Item-Item Cosine Similarity of Actual Movies Ratings, Class + DBMI Data (Partial of 50x50)*

| | The Social N | A Prophet | Amour | The King's S | La La Land | Boyhood | Inception | A Separatio | The Artist | The White R | Zero Dark Tl | Avatar | Spotlight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **The Social N** | 1 | 0.269625 | 0.216214 | 0.571513 | 0.608298 | 0.383359 | 0.717668 | 0.255281 | 0.320411 | 0.192774 | 0.595409 | 0.67338 | 0.490684 |
| **A Prophet** | 0.269625 | 1 | 0.302771 | 0.245621 | 0.250761 | 0.290441 | 0.240663 | 0.389553 | 0.353563 | 0.288218 | 0.208368 | 0.233489 | 0.240942 |
| **Amour** | 0.216214 | 0.302771 | 1 | 0.319894 | 0.237586 | 0.198899 | 0.223006 | 0.231291 | 0.357973 | 0.403379 | 0.232023 | 0.237028 | 0.235381 |
| **The King's S** | 0.571513 | 0.245621 | 0.319894 | 1 | 0.662118 | 0.402351 | 0.628466 | 0.332923 | 0.362033 | 0.21659 | 0.381307 | 0.644882 | 0.495078 |
| **La La Land** | 0.608298 | 0.250761 | 0.237586 | 0.662118 | 1 | 0.431774 | 0.719333 | 0.28155 | 0.321761 | 0.215607 | 0.4122 | 0.77368 | 0.446938 |
| **Boyhood** | 0.383359 | 0.290441 | 0.198899 | 0.402351 | 0.431774 | 1 | 0.43786 | 0.421998 | 0.314186 | 0.207583 | 0.343878 | 0.411769 | 0.370961 |
| **Inception** | 0.717668 | 0.240663 | 0.223006 | 0.628466 | 0.719333 | 0.43786 | 1 | 0.279072 | 0.321286 | 0.201837 | 0.557575 | 0.851802 | 0.482852 |
| **A Separatio** | 0.255281 | 0.389553 | 0.231291 | 0.332923 | 0.28155 | 0.421998 | 0.279072 | 1 | 0.314457 | 0.274884 | 0.238312 | 0.259016 | 0.318819 |
| **The Artist** | 0.320411 | 0.353563 | 0.357973 | 0.362033 | 0.321761 | 0.314186 | 0.321286 | 0.314457 | 1 | 0.303997 | 0.34742 | 0.291441 | 0.33704 |
| **The White R** | 0.192774 | 0.288218 | 0.403379 | 0.21659 | 0.215607 | 0.207583 | 0.201837 | 0.274884 | 0.303997 | 1 | 0.270703 | 0.213979 | 0.211352 |
| **Zero Dark Tl** | 0.595409 | 0.208368 | 0.232023 | 0.381307 | 0.4122 | 0.343878 | 0.557575 | 0.238312 | 0.34742 | 0.270703 | 1 | 0.495387 | 0.463023 |
| **Avatar** | 0.67338 | 0.233489 | 0.237028 | 0.644882 | 0.77368 | 0.411769 | 0.851802 | 0.259016 | 0.291441 | 0.213979 | 0.495387 | 1 | 0.473384 |
| **Spotlight** | 0.490684 | 0.240942 | 0.235381 | 0.495078 | 0.446938 | 0.370961 | 0.482852 | 0.318819 | 0.33704 | 0.211352 | 0.463023 | 0.473384 | 1 |

*Appendix 11b: Item-Item Cosine Similarity of Normalized Movies Ratings, Class + DBMI Data (Partial of 50x50)*

| | The Social N | A Prophet | Amour | The King's S | La La Land | Boyhood | Inception | A Separatio | The Artist | The White R | Zero Dark Tl | Avatar | Spotlight |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **The Social Net** | 1 | 0.015768 | 0.004335 | −0.12165 | −0.16894 | −0.18326 | −0.00291 | −0.08181 | 0.026013 | 0.03248 | −0.01152 | −0.04514 | −0.0226 |
| **A Prophet** | 0.015768 | 1 | 0.11994 | 0.077483 | 0.016389 | −0.00022 | −0.0814 | −0.03353 | −0.12647 | 0.165372 | −0.03554 | 0.076997 | −0.031 |
| **Amour** | 0.004335 | 0.11994 | 1 | 0.103091 | −0.08541 | 0.014042 | −0.03955 | 0.070074 | −0.02242 | 0.085667 | −0.05407 | −0.01584 | −0.028 |
| **The King's Spe** | −0.12165 | 0.077483 | 0.103091 | 1 | 0.000363 | −0.02504 | −0.08092 | 0.06866 | 0.10043 | 0.011491 | 0.013103 | −0.01653 | 0.028386 |
| **La La Land** | −0.16894 | 0.016389 | −0.08541 | 0.000363 | 1 | −0.01571 | 0.046491 | −0.02125 | −0.06338 | 0.065263 | −0.1128 | 0.030489 | −0.09272 |
| **Boyhood** | −0.18326 | −0.00022 | 0.014042 | −0.02504 | −0.01571 | 1 | −0.01992 | 0.243613 | −0.01152 | −0.12917 | 0.013317 | −0.10581 | 0.091607 |
| **Inception** | −0.00291 | −0.0814 | −0.03955 | −0.08092 | 0.046491 | −0.01992 | 1 | −0.02647 | −0.0535 | −0.0207 | −0.02465 | −0.03605 | −0.04992 |
| **A Separation** | −0.08181 | −0.03353 | 0.070074 | 0.06866 | −0.02125 | 0.243613 | −0.02647 | 1 | −0.13727 | −0.1659 | −0.07081 | −0.17916 | 0.117388 |
| **The Artist** | 0.026013 | −0.12647 | −0.02242 | 0.10043 | −0.06338 | −0.01152 | −0.0535 | −0.13727 | 1 | −0.01994 | −0.07293 | −0.0311 | 0.012266 |
| **The White Ribb** | 0.03248 | 0.165372 | 0.085667 | 0.011491 | 0.065263 | −0.12917 | −0.0207 | −0.1659 | −0.01994 | 1 | 0.002839 | 0.06902 | −0.14789 |
| **Zero Dark Thirt** | −0.01152 | −0.03554 | −0.05407 | 0.013103 | −0.1128 | 0.013317 | −0.02465 | −0.07081 | −0.07293 | 0.002839 | 1 | −0.02482 | 0.068215 |
| **Avatar** | −0.04514 | 0.076997 | −0.01584 | −0.01653 | 0.030489 | −0.10581 | −0.03605 | −0.17916 | −0.0311 | 0.06902 | −0.02482 | 1 | −0.07358 |
| **Spotlight** | −0.0226 | −0.031 | −0.028 | 0.028386 | −0.09272 | 0.091607 | −0.04992 | 0.117388 | 0.012266 | −0.14789 | 0.068215 | −0.07358 | 1 |

*Appendix 12: Part 4 Calculations (New User Ratings with Some Data)*

| | [Inception] | [Avatar] | [The Wolf of Wall Street] |
|---|---|---|---|
| [The Social Network] | 0.719116 | 0.687974 | 0.75407 |
| [A Prophet] | 0.21835 | 0.228056 | 0.239612 |
| [Amour] | 0.209932 | 0.195524 | 0.189689 |
| [The King's Speech] | 0.649641 | 0.646697 | 0.620413 |
| [La La Land] | 0.707379 | 0.774084 | 0.670436 |
| [Boyhood] | 0.431219 | 0.418917 | 0.388716 |
| [Inception] | 1 | 0.863107 | 0.814106 |
| [A Separation] | 0.287574 | 0.27848 | 0.300717 |
| [The Artist] | 0.317543 | 0.287278 | 0.308578 |
| [The White Ribbon] | 0.158477 | 0.162401 | 0.154672 |
| [Zero Dark Thirty] | 0.523176 | 0.457468 | 0.530656 |
| [Avatar] | 0.863107 | 1 | 0.78761 |
| [Spotlight] | 0.497648 | 0.48374 | 0.494049 |
| [Precious] | 0.288566 | 0.271756 | 0.311837 |
| [The Tree of Life] | 0.275999 | 0.2395 | 0.227474 |
| [12 Years a Slave] | 0.53044 | 0.511255 | 0.516062 |
| [Blue is the Warmest ( | 0.369966 | 0.337936 | 0.333582 |
| [Son of Saul] | 0.125376 | 0.103493 | 0.1483 |
| [Up in the Air] | 0.533608 | 0.471962 | 0.497787 |
| [Inglourious Basterds] | 0.586449 | 0.56571 | 0.622727 |
| [Mad Max: Fury Road] | 0.664714 | 0.657573 | 0.671928 |
| [Moonlight] | 0.425366 | 0.394588 | 0.385699 |

| | | | |
|---|---|---|---|
| [Gravity] | 0.718996 | 0.669188 | 0.675186 |
| [Black Swan] | 0.634356 | 0.639194 | 0.619674 |
| [Ida] | 0.202036 | 0.175653 | 0.212008 |
| [Leviathan] | 0.219192 | 0.192337 | 0.236694 |
| [The Wolf of Wall Stre | 0.814106 | 0.78761 | 1 |
| [True Grit] | 0.317685 | 0.272438 | 0.323322 |
| [The Descendants] | 0.359629 | 0.314313 | 0.349015 |
| [The Secret in Their Ey | 0.215148 | 0.215984 | 0.26466 |
| [Life of Pi] | 0.770414 | 0.786958 | 0.737607 |
| [Arrival] | 0.6002 | 0.562826 | 0.589985 |
| [Call Me by Your Nam | 0.455193 | 0.474729 | 0.469457 |
| [Winter's Bone] | 0.103672 | 0.086387 | 0.079157 |
| [The Grand Budapest | 0.622723 | 0.620311 | 0.598169 |
| [Dunkirk] | 0.634139 | 0.623451 | 0.649252 |
| [Inside Llewyn Davis] | 0.246369 | 0.209305 | 0.213852 |
| [A Serious Man] | 0.101577 | 0.090687 | 0.084021 |
| [Toy Story 3] | 0.688166 | 0.713104 | 0.660939 |
| [Beasts of the Souther | 0.169663 | 0.213227 | 0.138262 |

| | Inception | Avatar | Wolf | Amy | Sachi | Camille |
|---|---|---|---|---|---|---|
| | 0.28856603 | 0.27175595 | 0.311836563 | 2 | 4 | 4 |
| | 0.53044034 | 0.511255 | 0.51606242 | 2 | 5 | 3 |
| | 0.66471372 | 0.6575727 | 0.671927793 | 4 | 5 | 4 |
| | 0.63982085 | 0.64593489 | 0.667147017 | 3 | 4 | 1 |
| | 0.68816567 | 0.71310449 | 0.660939182 | 3 | 3 | 4 |
| Sum | 2.81170661 | 2.79962303 | 2.827912976 | | | |

Avatar Rating for Amy 2.96 = (0.271*2 + 0.511*2 + 0.657*4 + 0.645*3 +0.713*3) / 2.811

$$Predicted\ Rating = \frac{\sum_1^j CosineSimilarity_{(i,j)} * Rating_j}{\sum_1^j CosineSimilarity_{(i,j)}}$$

***i*** is the movie, whose rating is to be predicted (Inception, Avatar, Wolf of Wall Street)

***j*** is the ratings of movies in the new data (Precious, 12 Years a Slave, Mad Max Fury Road ,Black Swan & Toy Story 3)