Recommendation System using Collaborative Filtering

Student Names:

Anisha Samant

Sheetal Rajgure

Issam (Sam) Tamer

Siwei (David) Ran

School: *UC Davis*

Course Number and Name*: BAX 401 Information, Insight and Impact*

Instructor*: Ashwin Aravindakshan*

Date: *2022-11-27*

Table of Contents

Executive Summary: Plentiful and easily accessible content, in today's Information Age, has brought about a wide range of choices at consumers' disposal. Recommender Systems help save time, decision-making & search costs for consumers by providing relevant recommendations, and as a result, enhance customer experience leading to long-term benefits for firms. The objective of this report is to explore and build a Recommendation System using Collaborative Filtering (CF) and predict movie ratings of users in different scenarios. We explored Cosine Similarity, Manhattan Distance and Euclidean Distance, accounted for user bias and item bias using Mean Centered methods, and finally predicted the ratings using weighted sum of similarities and ratings. Of these, User-User Mean Centered Cosine Similarity proved to be giving comparatively better results when we have some information about the users as well as items (here movies). To approach the cold start problem, we examined different approaches -- a. Mean ratings of closest cohort (UC Davis) b. Mean ratings of user clusters created using demographic data from external sources, IMDB and Rotten Tomatoes c. Mean ratings of movie clusters created using genre, etc. We noticed that predictions improve when we have base information about the new users, thus the performance of recommender systems, in general, boosts with more information at hand. We also probed further scope of making more accurate predictions with sophisticated ML algorithms and granular user and movie information.

Introduction: Companies ranging from Youtube to Amazon all rely on various recommendation algorithms to provide a personalized platform for their customers. Recommender systems are extremely popular among movie streaming platforms. Relevant movie recommendations based on past viewing history, genre preferences and also general popularity boosts viewership and ultimately customer retention and revenue. In this report, we will be predicting movie ratings for different Users using CF techniques that take into account past interaction and behavior.

Problem Formulation: First problem statement is to predict our group members' ratings for any three given movies and compare the results of different techniques used with the actual ratings. Since we have a fairly large number of users and a variety of movies, CF metrics based on similarity would help find patterns amongst users to give good predictions. Next, we will predict ratings in different scenarios: a. New movie - predict group members' ratings for 3 new movies, b.New User - predict ratings of new users with no prior information for 3 given movies c.New User with base information - predict ratings of new users with some information for 3 given movies. Lastly, we will examine the value of recommendation engines in practice and approach to refine the process.

Data Description: The Analytics Team has been provided with Movie Ratings ranging from 1 to 5 for 50 different movies. The movies have been rated by two sets of users- 98 MSBA students from UC Davis (UC Davis Dataset) and 280 Random Users (DBMI Dataset). The dataset with the students contain their full name and group number. The data lacks any demographic information for any of the users, also, considering not all users have seen all included movies, there are multiple data points with "Did not Watch". Before proceeding with our analysis, the "Did Not Watch's" were replaced with Null.

Model Development and Results: First, we selected Life of Pi, Dunkirk and Imitation Game to predict on. The movies were selected such that there was enough data provided by all group members to test different CF approaches and determine the most effective one. We started by first combining the UC Davis and DBMI data with the intent of getting more data points to calibrate our filter. We then filtered movies and users with low exposure for simplicity and to remove outliers. Movies that were rated by less than 15 users, and users who rated less than 5 movies were dropped, leaving 45 movies and 350 users. The team members' reviews for the chosen 3 movies were deleted from the data.

Starting off with a User-User collaborative filtering approach to make predictions, we calculated the predictions for the movies using three methods: Manhattan Distance, Euclidean Distance and Cosine Similarity. Comparatively, the Cosine Similarity measure was more accurate than the other two across the 3 member ratings. However, the predictions were still quite off, so we proceeded with our next two approaches- user-user mean centered and item-item mean centered, to reduce the possibility of any user bias or item bias. After getting the averages per user and per movie, we proceeded to calculate the cosine similarity using both the filtering types. The predictions came out to be much more accurate, with the predictions from item-item-mean squared performing the best in this case overall (1.).

To give an example, Sheetal's reviews for the 3 movies were (5,4,4) respectively. User-User Cosine Similarity predicted (2.9, 2.18, 2.82). U-U mean centered predicted much better with (4.24, 4.29, 4.35), and I-I mean centered brought the predictions even closer (4, 4.04, 4.11). This trend was seen across all team member's predictions. Item bias would be suspected to be very high in this scenario as different users prefer different genres, and many users would also have favorites that would lead them to rate it higher. Also, from the ratings we see that Sheetal and David are much kinder with their reviews with 5s-4s, and Anisha and Sam fall on the more moderate side with 3s-4s. Again, reinforcing that different users have different approaches to ratings, with some being more generous. Normalizing the ratings helped reduce the biases as seen from better predictions in both of the mean centered approaches.

For parts 2, 3 and 4, we examined three paths to approach these situations:

1. Averages and similarities from the available data of movie ratings by UC Davis class as well as the additional DBMI dataset. Predicting ratings based on the weighted sum-products of ratings and similarities using CF approaches mentioned previously.
2. Averages from external data sources e.g. IMDB, Rotten Tomatoes Audience ratings, etc. which also provides demographics (gender and age) data of movie ratings*(for instance, 3.2)*
3. Clustering users and items on the basis of certain parameters and assigning the average rating of the cluster to new user/new item. For e.g., i. Cluster movies on the basis of genre -- if we do not have any user rating for a new movie, we predict ratings based on the average rating of the cluster it belongs to.

   ii. Cluster users based on demographics, behaviors and interests -- if we do not have any information for a new user, we predict ratings based on the average rating of the cluster it belongs to.

The second part required predicting movie ratings of three new movies 'Winter's Bone', 'A Serious Man' and 'Son of Saul', for which, the ratings of UC Davis students were not available. Although none of the UC Davis students rated these three movies, we noticed there were users in the DBMI dataset who had rated them. This additional information helped us use the first path mentioned above to tackle the cold-start problem. Out of the 280 users, we filtered only those who rated the three movies of interest. Taking into account biases, we used U-U Mean Centered and I-I Mean Centered approaches respectively, explained earlier.*(2.1)*. Though both methods are very reliable, we believe predictions from the U-U Mean Centered method are closer based on our literature review in (*2*).

For the third part, we predicted 3 movie ratings (Inception, Avatar, and The Wolf of Wall Street) for new users - Camille, Shachi and Amy. Since we do not have any information about them we cannot apply CF approach. We decided to use the average rating from internal and external resources to predict their rating. The internal source including the data of MSBA students and DBMI users has the rating range from 4.4 to 4.52, 3.99 to 4.02, and 4.1 to 4.17 for each respective movie (In Appendix). The average rating of external resources from IMDB are 4.35, 3.9, and 3.9(scale of 5) respectively. We also used Rotten Tomatoes' ratings *(3.3)* and considered the demographic data available at IMDB, female and 30-44 years group to benchmark Amy and Shachi's ratings while female and 20-29 years group for Camille. From our perspective, the UC Davis MSBA cohort is closer to Amy, Shachi and Camille due to similar backgrounds, interests and behaviors, as compared to the random and very diverse cohort of external sources. Hence to start with, we give more importance to the MSBA dataset's average ratings.

Our approach to solving the cold-start problem changes when we receive move preference data on these users. Now, we can predict the ratings of the same three movies (Avatar, The Wolf of Wall Street, and Inception) for Camille, Shachi, and Amy on a more individualized basis. This changes into a problem that can be solved by a U-U Mean Centered approach. We found the Cosine similarity of each of them with the complete dataset of UC Davis+DBMI subjects. We filtered for students who have rated all 5 movies to make the subset more representative. The results, summarized in the appendix, represent a more distinguished spread between each admin than a simple average, with values changing by up to 1.26 when compared to our previous recommendations [4.1]. We believe these new scores are more representative of the admin’s preferences because they’re based on scores of people who share similar tastes, rather than a general average of the population.

Recommendations and Managerial Implications: Firms are always looking for new ways to get an edge on their competition, and extending the methods we’ve applied in this report to other businesses enables them to do this. By determining the best recommendation (across any product or service) at an individual customer or product level, advertisements can be more targeted: less money is wasted on irrelevant ads and conversion rates are higher because customers see the ads that are most relevant to them. Customers, in turn, get a better experience with the company: less time is spent sifting through the meaningless noise that comes with other blanket advertising strategies, increasing customer engagement. These both help increase advertising’s return on investment and make a firm more profitable than competitors who do not use recommendation-model-based advertising strategies.

Additionally, more ML models and methods like content based filtering could be applied to predict ratings from new customers and on new movies. Instead of predicting them by average public ratings, using detailed, specific, and quantifiable data, these models could help determine the factors affecting ratings and form a systematic algorithm for future use.

Conclusion: While effective, the methods discussed in this report can be improved on by increasing the granularity of the data collected, such as adding movie genre, and collecting some demographic information from the subjects, such as gender, age, and nationality. These would allow recommendations to be targeted further, and help us better understand the appropriate scope of making predictions. Currently, we are limited in generalizing the data beyond MSBA students because they exclusively make up our data. We’ve concluded that using user-user or item-item mean centered approaches perform better for providing accurate recommendations than cosine similarity or distance-based rankings for our sparse, non-granular dataset. The cold start problem was addressed by applying CF techniques, which reduced to simple averages of our internal data because user similarities couldn’t be defined. After some base preferences were added, we could then improve our naive recommendations using the same CF methods.

References & Notes

1. #Chapter 9: Recommendation Systems, *Mining of Massive Datasets*, by Jure Leskovic, Anand Rajaram and Jeffrey Ullman
2. BOSTRÖM, P., & FILIPSSON, M. (2017). *Comparison of User Based and Item Based Collaborative Filtering Recommendation Services*. http://www.diva-portal.org/smash/get/diva2:1111865/FULLTEXT01.pdf.

Appendix:

1. Movies Chosen: Life of Pi, Dunkirk, Imitation Game

Collaborative Filtering metric: Cosine Similarity

Team Mem: Anisha Samant			
Method	**Life of Pi**	**Dunkirk**	**Imitation Game**
User-User	3	2.12	2.69
User-User Mean Centered	3.68	3.71	3.74
Item-Item Mean Centered	3.87	3.87	3.87
Actual Rating	4	3	4

Team Mem: Sheetal Rajgure			
Method	**Life of Pi**	**Dunkirk**	**Imitation Game**
User-User	2.9	2.18	2.82
User-User Mean Centered	4.24	4.29	4.35
Item-Item Mean Centered	4.02	4.04	4.11
Actual Rating	5	4	4

Team Mem: Issam 'Sam' Tamer			
Method	**Life of Pi**	**Dunkirk**	**Imitation Game**
User-User	2.95	2	2.62
User-User Mean Centered	3.7	3.73	3.77
Item-Item Mean Centered	3.96		4.004
Actual Rating			4

Team Mem: Siwei 'David' Ran			
Method	**Life of Pi**	**Dunkirk**	**Imitation Game**
User-User	2.99	2.09	2.69
User-User Mean Centered	3.99	4	4.13
Item-Item Mean Centered	3.99	4.02	4.05
Actual Rating	5	4	4

Excel Calculation Snapshots for User-User Mean Centered Method (Showing calculations only for one member- Issam “Sam”. Please note that for calculations, all users (UC DAVIS cohort + BDMI data) were used. Due to space restrictions, only few are shown in the screenshot)

	Issam "Sam"	User 1	User 2	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17
The Social Network	0	0.76923077	-1.0526316	0	-0.3	0.25	0	0	0.4375	0	0	-0.5581395	-0.9534884	0	0	-0.25	0.36363636
A Prophet	0	0	0	0	0	0	0	0	0	0	0	1.44186047	-0.9534884	0	0	0	0
Amour	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	0.04651163	0	0	0	0
The King's Speech	0	-1.2307692	-0.0526316	0	-0.3	0	0.5	0	0.4375	0	0	1.44186047	1.04651163	0	-0.2222222	0.75	0
La La Land	1.25	-0.2307692	0.94736842	0.3	0.7	0	0	0.375	-0.5625	0.64705882	0	1.44186047	1.04651163	-1.1	0.77777778	-0.25	0.36363636
Boyhood	0	0	-0.0526316	0	0	0.25	0	0	0	-0.3529412	0	-0.5581395	0.04651163	0	0	0	0
Inception	1.25	0.76923077	0.94736842	-0.7	0.7	1.25	0	0	0.4375	-0.3529412	0	-0.5581395	1.04651163	0.9	0.77777778	0	-0.6363636
A Separation	0	0	0	0	0	0	0.5	0	0	0	0	-0.5581395	0.04651163	0	0	0	0
The Artist	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	1.04651163	0	0	0	0
Zero Dark Thirty	0	0	0	0	0	0.25	0	0	0	0	0	-0.5581395	1.04651163	0	0	0	0
Avatar	1.25	0.76923077	-1.0526316	0.3	0	-0.75	0.5	0	0.4375	0.64705882	0	1.44186047	0.04651163	-1.1	0.77777778	0	0.36363636
Spotlight	0	0	0	-0.7	0	0.25	0	0	0	0	0	-0.5581395	0.04651163	0	0	0	0
Precious	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	0.04651163	0	0	0	0
The Tree of Life	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	-0.9534884	0	0	0	0
12 Years a Slave	0	-2.2307692	0	0	0	-0.75	0	0	0.4375	0	0	1.44186047	0.04651163	0	0	0	0
Blue is the Warmest Colour	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	0.04651163	0	0	0	0
Up in the Air	0	0.76923077	-1.0526316	0	0	0.25	0	0	0	0.64705882	0	-0.5581395	0.04651163	0	0.77777778	0	0.36363636
Inglourious Basterds	0.25	0	-0.0526316	0	0	0.25	0	0	0.4375	0.64705882	0	-0.5581395	-0.9534884	0	0	0	0
Mad Max: Fury Road	-1.75	0	-1.0526316	0	0	1.25	0	0	-0.5625	0	0	-0.5581395	0.04651163	0	0	0	0
Moonlight	-1.75	0	-0.0526316	0	0	0	0	0	0	-0.3529412	0	1.44186047	0.04651163	0	0	0	0
Birdman	0	-0.2307692	-0.0526316	0	0	-0.75	0	0	-1.5625	0	0	1.44186047	-0.9534884	0	0	0	0
Manchester by the Sea	0	0	-0.0526316	0	0	0	0	0	0	-1.3529412	0	-0.5581395	1.04651163	0	0	-0.25	0
Lincoln	0	-1.2307692	0	0	0	0	-1.5	0	0	0	0	-0.5581395	-1.9534884	0	0	0	0
Hugo	0	-0.2307692	0	0	0	0	0	0	0	0	0	-0.5581395	-0.9534884	0	0	-0.25	0
The Shape of Water	0	0	0	0	0	0	0	0	0	-0.3529412	0	1.44186047	0.04651163	-2.1	0	-0.25	0
Three Billboards Outside Ebbing, Missouri	0	0	-1.0526316	0.3	0	0	0	0.375	-0.5625	-1.3529412	0	-0.5581395	1.04651163	0	0	-0.25	0
Argo	0	0	0	0	0	0.25	0	0	0	0	0	-0.5581395	1.04651163	0	0	0	0
Gravity	0	0.76923077	0	0.3	-0.3	0	0	0.375	-0.5625	0	0	-0.5581395	1.04651163	0.9	-1.2222222	0	0.36363636
Black Swan	0	0	0	0	0	-1.75	0	0	0	0.64705882	0	-0.5581395	-0.9534884	0.9	0	-0.25	0
Leviathan	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	-1.9534884	0	0	0	0
The Wolf of Wall Street	0.25	0.76923077	-0.0526316	-0.7	0	0.25	0.5	-1.625	0.4375	0	0	1.44186047	1.04651163	0.9	-1.2222222	0	-0.6363636
True Grit	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	-1.9534884	0	0	0	0
The Descendants	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	-1.9534884	0	0	0	0
The Secret in Their Eyes	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	0.04651163	0	0	0	0
Life of Pi	0	0.76923077	0.94736842	0.3	0.7	0.25	0	-0.625	0.4375	0	0	1.44186047	1.04651163	0.9	0	-0.25	-1.6363636
Arrival	0	0	0.94736842	0	0	0.25	-0.5	0	0	0.64705882	0	-0.5581395	1.04651163	0	0	0	1.36363636
Call Me by Your Name	0	0	0	0	-0.3	0	0	0.375	0	-0.3529412	0	1.44186047	1.04651163	-0.1	0	-0.25	0
The Grand Budapest Hotel	0	0	-0.0526316	0.3	0.7	0	0	0.375	-0.5625	0.64705882	0	-0.5581395	1.04651163	0	-1.2222222	0.75	0.36363636
Dunkirk	0	0	0.94736842	0.3	-1.3	-0.75	0	0.375	0.4375	0.64705882	0	1.44186047	-0.9534884	-0.1	0	0.75	0
Inside Llewyn Davis	0	0	0	0	0	0	0	0	0	0	0	-0.5581395	0.04651163	0	0	0	0
Toy Story 3	-0.75	0	0	0	-0.3	-0.75	0	0	0	-0.3529412	0	-0.5581395	1.04651163	0	0.77777778	0	-0.6363636
The Imitation Game	0	0	0.94736842	0	0	0.25	0	0	0.4375	-0.3529412	0	-0.5581395	1.04651163	0	0	0	0
The Fighter	0	0	0	0	0	0.25	0	0	0	0	0	-0.5581395	-1.9534884	0	0	0	0

Cosine Similarity	User 1	User 2	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12	User 13	User 14	User 15	User 16	User 17
The Social Network	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A Prophet	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amour	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
The King's Speech	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
La La Land	-0.2884615	1.18421053	0.375	0.875	0	0	0.46875	-0.703125	0.80882353	0	1.80232558	1.30813953	-1.375	0.97222222	-0.3125	0.45454545
Boyhood	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Inception	0.96153846	1.18421053	-0.875	0.875	1.5625	0	0	0.546875	-0.4411765	0	-0.6976744	1.30813953	1.125	0.97222222	0	-0.7954545
A Separation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
The Artist	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Zero Dark Thirty	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Avatar	0.96153846	-1.3157895	0.375	0	-0.9375	0.625	0	0.546875	0.80882353	0	1.80232558	0.05813953	-1.375	0.97222222	0	0.45454545
Spotlight	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Precious	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
The Tree of Life	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12 Years a Slave	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Blue is the Warmest Colour	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Up in the Air	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Inglourious Basterds	0	-0.0131579	0	0	0.0625	0	0	0.109375	0.16176471	0	-0.1395349	-0.2383721	0	0	0	0
Mad Max: Fury Road	0	1.84210526	0	0	-2.1875	0	0	0.984375	0	0	0.97674419	-0.0813953	0	0	0	0
Moonlight	0	0.09210526	0	0	0	0	0	0	0.61764706	0	-2.5232558	-0.0813953	0	0	0	0
Birdman	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Manchester by the Sea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lincoln	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hugo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
The Shape of Water	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Three Billboards Outside Ebbing, Missouri	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Argo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gravity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Black Swan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Leviathan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
The Wolf of Wall Street	0.19230769	-0.0131579	-0.175	0	0.0625	0.125	-0.40625	0.109375	0	0	0.36046512	0.26162791	0.225	-0.3055556	0	-0.1590909
True Grit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
The Descendants	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
The Secret in Their Eyes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Life of Pi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Arrival	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Call Me by Your Name	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
The Grand Budapest Hotel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dunkirk	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Inside Llewyn Davis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Toy Story 3	0	0	0	0.225	0.5625	0	0	0	0.26470588	0	0.41860465	-0.7848837	0	-0.5833333	0	0.47727273
The Imitation Game	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
The Fighter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Num	1.82692308	2.96052632	-0.3	1.975	-0.875	0.75	0.0625	1.59375	2.22058824	0	2	1.75	-1.4	2.02777778	-0.3125	0.43181818
Deno1	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499	3.39116499
Deno2	3.50823208	3.30868077	1.44913767	2.02484567	3.122499	1.87082869	1.96850197	2.43669859	2.80755284	0	5.88257182	6.62623398	3.3015148	2.74873708	1.5	2.5584086
Cosine	0.15356174	0.26385488	-0.0610468	0.28762475	-0.0826336	0.11821656	0.00936257	0.19287213	0.23323363	0	0.1002568	0.07787936	-0.1250449	0.21753956	-0.0614341	0.04977166
Abs	0.15356174	0.26385488	0.06104677	0.28762475	0.08263363	0.11821656	0.00936257	0.19287213	0.23323363	0	0.1002568	0.07787936	0.12504487	0.21753956	0.06143415	0.04977166

	Issam "Sam"	
The Social Network		3.68938029
A Prophet		3.74497202
Amour		3.75048053
The King's Speech		3.7845439
La La Land	5	5
Boyhood		3.73860405
Inception	5	5
A Separation		3.72748488
The Artist		3.73430244
Zero Dark Thirty		3.72677074
Avatar	5	5
Spotlight		3.74700156
Precious		3.76120215
The Tree of Life		3.75132503
12 Years a Slave		3.72038523
Blue is the Warmest Colour		3.69962421
Up in the Air		3.71189871
Inglourious Basterds	4	4
Mad Max: Fury Road	2	3
Moonlight	2	2
Birdman		3.63281871
Manchester by the Sea		3.69262832
Lincoln		3.70604803
Hugo		3.72434486
The Shape of Water		3.68773107
Three Billboards Outside Ebbing, Missouri		3.70991161
Argo		3.76013579
Gravity		3.75050309
Black Swan		3.8397822
Leviathan		3.74316539
The Wolf of Wall Street		4
True Grit		3.74600309
The Descendants		3.72645975
The Secret in Their Eyes		3.74623704
Life of Pi		3.70209602
Arrival		3.72968084
Call Me by Your Name		3.72089441
The Grand Budapest Hotel		3.76706041
Dunkirk		3.73424166
Inside Llewyn Davis		3.7286006
Toy Story 3	3	3
The Imitation Game	4	3.7732725
The Fighter		3.73507837

2.1 Question 2: Predict Ratings for the following movies for your group members

- Movies: Winter's Bone, A Serious Man, Son of Saul

Results:

Using User-User Mean Centered Method:

	Winter's Bone	Son of Saul	A Serious Man
Issam 'Sam' Tamer	3.68	3.73	3.55
Sheetal Rajgure	4.17	4.26	4.36
Anisha Samant	3.38	3.69	3.60
Siwei 'David' Ran	3.76	4.03	4.12

Using Item-Item Mean Centered Method:

	Winter's Bone	Son of Saul	A Serious Man
Issam 'Sam' Tamer	2.74	3.22	2.61
Sheetal Rajgure	2.89	3.44	2.77
Anisha Samant	2.79	3.30	2.72
Siwei 'David' Ran	2.84	3.41	2.75

Excel Calculation Snapshots for User-User Mean Centered Method (Showing calculations only for one member):

User	User 12	User 13	User 139	User 144	User 146	User 155	User 162	User 168	User 173	Sam	Sheetal	Anisha	Siwei
The Social Network	1	3	5	4	4	4	5				4	2	3
The King's Speech	3	5	4	3	3	5	3	3					3
La La Land	3	5	4	4	2	5		3		5		5	
Boyhood	1	4		4			3					4	
Inception	1	5	5	5	5	5	3	5	5	5	4	4	5
Zero Dark Thirty	1	5			5		2		4		5		
Avatar	3	4	3	3	2	5	2	5	4	5	3	3	3
12 Years a Slave	3	4			5	2	3	4				4	
Blue is the Warmest Colour	1	4		3		3						3	
Inglourious Basterds	1	3	4		5		4	4	5	4	4	4	
Mad Max: Fury Road	1	4	4				3	5	4	2		2	
Moonlight	3	4	5							2			
Manchester by the Sea	1	5		5		3	4					4	
Lincoln	1	2			4	4	2	4	5			4	
Three Billboards Outside Ebbing, Misso	1	5	3		4	3			4				4
Argo	1	5	2		4		1	5	5		5		
Gravity	1	5	5	5	4	4	2	4	3			4	5
Black Swan	1	3			4	5	4					5	
The Wolf of Wall Street	3	5	2		5	5	4	5	3	4	5	4	5
Life of Pi	3	5	3	5		3	3	5			5	4	5
Arrival	1	5	2				4		5				4
Call Me by Your Name	3	5		3								3	
The Grand Budapest Hotel	1	5	5	5			4	4				5	4
Dunkirk	3	3	4		5				4		4	3	4
Toy Story 3	1	5	5	3		2	5	3	5	3		3	
The Imitation Game	1	5	3		4	5		5	3	4	4	4	4
Winter's Bone	1			2			3		5				
Son of Saul	1	4			4			4					
A Serious Man	1		4			3							
Average	1.620689655	4.333333333	3.78947368	3.857143	4.058824	3.882353	3.2	4.25	4.266667	3.777778	4.3	3.7	4.083333

Member: **Issam 'Sam' Tamer**

	1	2	3	4	5	6	7	8	9	10	11	12	13	
User	User 12	User 13	User 139	User 144	User 146	User 155	User 162	User 168	User 173	Sam	Sheetal	Anisha	Siwei	IMDB
The Social Network	-0.620689655	-1.333333333	1.210526316	0.1428571	-0.058824	0.1176471	1.8	0	0	0	-0.3	-1.7	-1.083333	0.0844828
The King's Speech	1.379310345	0.666666667	0.210526316	-0.857143	-1.058824	1.1176471	-0.2	-1.25	0	0	0	0	-1.083333	0.1844828
La La Land	1.379310345	0.666666667	0.210526316	0.1428571	-2.058824	1.1176471	0	-1.25	0	1.2222222	0	1.3	0	0.1844828
Boyhood	-0.620689655	-0.333333333	0	0.1428571	0	0	-0.2	0	0	0	0	0.3	0	0.1344828
Inception	-0.620689655	0.666666667	1.210526316	1.1428571	0.9411765	1.1176471	-0.2	0.75	0.7333333	1.2222222	-0.3	0.3	0.9166667	-0.065517
Zero Dark Thirty	-0.620689655	0.666666667	0	0	0.9411765	0	-1.2	0	-0.266667	0	0.7	0	0	-0.115517
Avatar	1.379310345	-0.333333333	-0.78947368	-0.857143	-2.058824	1.1176471	-1.2	0.75	-0.266667	1.2222222	-1.3	-0.7	-1.083333	0.0844828
12 Years a Slave	1.379310345	-0.333333333	0	0	0.9411765	-1.882353	-0.2	-0.25	0	0	0	0.3	0	0.2344828
Blue is the Warmest Colour	-0.620689655	-0.333333333	0	-0.857143	0	-0.882353	0	0	0	0	0	-0.7	0	0.0344828
Inglourious Basterds	-0.620689655	-1.333333333	0.210526316	0	0.9411765	0	0.8	-0.25	0.7333333	0.2222222	-0.3	0.3	0	-0.565517
Mad Max: Fury Road	-0.620689655	-0.333333333	0.210526316	0	0	0	-0.2	0.75	-0.266667	-1.777778	0	-1.7	0	0.2844828
Moonlight	1.379310345	-0.333333333	1.210526316	0	0	0	0	0	0	-1.777778	0	0	0	-0.115517
Manchester by the Sea	-0.620689655	0.666666667	0	1.1428571	0	-0.882353	0.8	0	0	0	0	0.3	0	0.0844828
Lincoln	-0.620689655	-2.333333333	0	0	-0.058824	0.1176471	-1.2	-0.25	0.7333333	0	0	0.3	0	-0.165517
Three Billboards Outside Ebbing, Misso	-0.620689655	0.666666667	-0.78947368	0	-0.058824	-0.882353	0	0	-0.266667	0	0	0	-0.083333	0.2344828
Argo	-0.620689655	0.666666667	-1.78947368	0	-0.058824	0	-2.2	0.75	0.7333333	0	0.7	0	0	0.0344828
Gravity	-0.620689655	0.666666667	1.210526316	1.1428571	-0.058824	0.1176471	-1.2	-0.25	-1.266667	0	0	0.3	0.9166667	0.0344828
Black Swan	-0.620689655	-1.333333333	0	0	-0.058824	1.1176471	0.8	0	0	0	0	1.3	0	0.1844828
The Wolf of Wall Street	1.379310345	0.666666667	-1.78947368	0	0.9411765	1.1176471	0.8	0.75	-1.266667	0.2222222	0.7	0.3	0.9166667	0.2844828
Life of Pi	1.379310345	0.666666667	-0.78947368	1.1428571	0	-0.882353	-0.2	0.75	0	0	0.7	0.3	0.9166667	0.1344828
Arrival	-0.620689655	0.666666667	-1.78947368	0	0	0	0.8	0	0.7333333	0	0	0	-0.083333	0.1344828
Call Me by Your Name	1.379310345	0.666666667	0	-0.857143	0	0	0	0	0	0	0	-0.7	0	-0.365517
The Grand Budapest Hotel	-0.620689655	0.666666667	1.210526316	1.1428571	0	0	0.8	-0.25	0	0	0	1.3	-0.083333	0.3344828
Dunkirk	1.379310345	-1.333333333	0.210526316	0	0.9411765	0	0	0	-0.266667	0	-0.3	-0.7	-0.083333	-0.265517
Toy Story 3	-0.620689655	0.666666667	1.210526316	-0.857143	0	-1.882353	1.8	-1.25	0.7333333	-0.777778	0	-0.7	0	-0.515517
The Imitation Game	-0.620689655	0.666666667	-0.78947368	0	-0.058824	1.1176471	0	0.75	-1.266667	0.2222222	-0.3	0.3	-0.083333	0.1844828
Winter's Bone	-0.620689655	0	0	-1.857143	0	0	-0.2	0	0.7333333	0				-0.265517
Son of Saul	-0.620689655	-0.333333333	0	0	-0.058824	0	0	-0.25	0	0				-0.115517
A Serious Man	-0.620689655	0	0.210526316	0	0	-0.882353	0	0	0	0				-0.315517

	1	2	3	4	5	6	7	8	9	11	12	13	
	10	10	10	10	10	10	10	10	10	10	10	10	
The Social Network	0	0	0	0	0	0	0	0	0	0	0	0	
The King's Speech	0	0	0	0	0	0	0	0	0	0	0	0	
La La Land	1.685823755	0.81481481	0.25731	0.1746	-2.5163	1.36601	0	-1.5278	0	0	1.58889	0	
Boyhood	0	0	0	0	0	0	0	0	0	0	0	0	
Inception	-0.75862069	0.81481481	1.479532	1.39683	1.15033	1.36601	-0.2444	0.91667	0.8963	-0.3667	0.36667	1.12037	
Zero Dark Thirty	0	0	0	0	0	0	0	0	0	0	0	0	
Avatar	1.685823755	-0.4074074	-0.96491	-1.0476	-2.5163	1.36601	-1.4667	0.91667	-0.3259	-1.5889	-0.8556	-1.3241	
12 Years a Slave	0	0	0	0	0	0	0	0	0	0	0	0	
Blue is the Warmest Colour	0	0	0	0	0	0	0	0	0	0	0	0	
Inglourious Basterds	-0.13793103	-0.2962963	0.046784	0	0.20915	0	0.17778	-0.0556	0.16296	-0.0667	0.06667	0	
Mad Max: Fury Road	1.103448276	0.59259259	-0.37427	0	0	0	0.35556	-1.3333	0.47407	0	3.02222	0	
Moonlight	-2.45210728	0.59259259	-2.15205	0	0	0	0	0	0	0	0	0	
Manchester by the Sea	0	0	0	0	0	0	0	0	0	0	0	0	
Lincoln	0	0	0	0	0	0	0	0	0	0	0	0	
Three Billboards Outside Ebbing, M	0	0	0	0	0	0	0	0	0	0	0	0	
Argo	0	0	0	0	0	0	0	0	0	0	0	0	
Gravity	0	0	0	0	0	0	0	0	0	0	0	0	
Black Swan	0	0	0	0	0	0	0	0	0	0	0	0	
The Wolf of Wall Street	0.30651341	0.14814815	-0.39766	0	0.20915	0.24837	0.17778	0.16667	-0.2815	0.15556	0.06667	0.2037	
Life of Pi	0	0	0	0	0	0	0	0	0	0	0	0	
Arrival	0	0	0	0	0	0	0	0	0	0	0	0	
Call Me by Your Name	0	0	0	0	0	0	0	0	0	0	0	0	
The Grand Budapest Hotel	0	0	0	0	0	0	0	0	0	0	0	0	
Dunkirk	0	0	0	0	0	0	0	0	0	0	0	0	
Toy Story 3	0.482758621	-0.5185185	-0.94152	0.66667	0	1.46405	-1.4	0.97222	-0.5704	0	0.54444	0	
The Imitation Game	-0.13793103	0.14814815	-0.17544	0	-0.0131	0.24837	0	0.16667	-0.2815	-0.0667	0.06667	-0.0185	
Winter's Bone	0	0	0	0	0	0	0	0	0	0	0	0	
Son of Saul	0	0	0	0	0	0	0	0	0	0	0	0	
A Serious Man	0	0	0	0	0	0	0	0	0	0	0	0	
Numerator	1.777777778	1.88888889	-3.22222	1.19048	-3.4771	6.05882	-2.4	0.22222	0.07407	-1.9333	4.86667	-0.0185	
Deno1	3.399346342	3.39934634	3.399346	3.39935	3.39935	3.39935	3.39935	3.39935	3.39935	3.39935	3.39935	3.39935	
Deno2	4.982728791	4.47213595	4.599771	3.70328	3.86538	4.44575	4.60435	3	2.98887	2.02485	3.76829	2.62996	
Cosine Similarity	0.104957822	0.12424989	-0.20607	0.09457	-0.2646	0.40091	-0.1533	0.02179	0.00729	-0.2809	0.37992	-0.0021	0.2267
Abs Cosine	0.104957822	0.12424989	0.206074	0.09457	0.26463	0.40091	0.15334	0.02179	0.00729	0.28088	0.37992	0.00207	2.04067
Winter's Bone	-0.62068966	0	0	-1.8571	0	0	-0.2	0	0.73333				
Son of Saul	-0.62068966	-0.3333333	0	0	-0.0588	0	0	-0.25	0	0			
A Serious Man	-0.62068966	0	0.210526	0	0	-0.8824	0	0	0	0			
	Sam												
Winter's Bone	3.677440178												
Son of Saul	3.730516785												
A Serious Man	3.55124724												

3. Question 3:

Predict Movie ratings for new customers: Camille, Shachi and Amy

For the Movies: Avatar, The Wolf of Wall Street, Inception

Results:

Using Averages from the available data of movie ratings by UC Davis MSBA class as well as the additional DBMI dataset to predict the rating of 3 movies for Camille, Shachi and Amy.

	[Inception]	[Avatar]	[The Wolf of Wall Street]
MSBA Average	4.4	3.99	4.17
DBMI Average	4.52	4.02	4.1
MSBA&DBMI Average	4.49	4.02	4.12

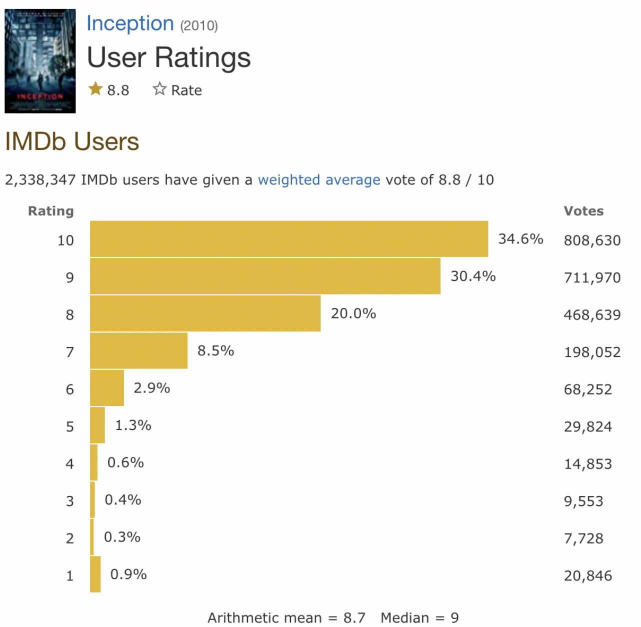
Excel Calculation Snapshots for Average (Showing calculations only for MSBA class dataset):

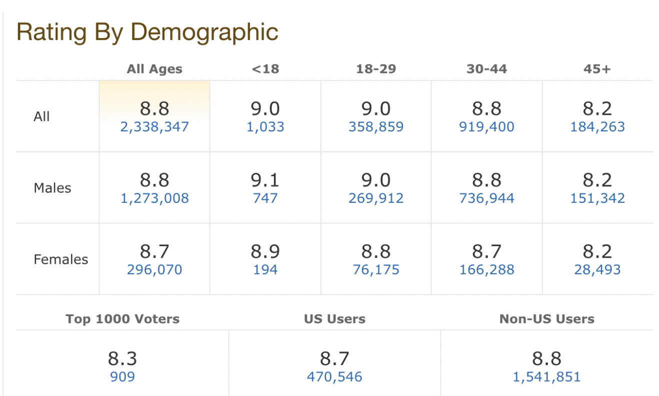
	[Inception]		[Avatar]		[The Wolf of Wall Street]
	5		5		5
	5		5		4
	4		5		4
	3		5		4
	5		3		3
	4		5		5
	4		5		4
	5		5		4
	4		3		5
	4		5		5
	4		4		4
	5		4		5
	4		4		5
	5		4		5
	5		3		5
	4		4		4
	5		4		5
	5		2		5
	5		5		2
	3		5		4
	1		1		5
	2		3		4
	4		3		2
	4		5		5
	5		3		2
	5		3		2
	5		3		4
	5		5		5
	4		4		3
	5		4		4
	5		4		3
	5		5		5
	4		4		1
	5		5		5
	5		5		5
	1		4		4
	5		3		4
	5		1		5
	5		4		3
	5		3		4
	4		5		5
	5		5		5
	4		5		5
	5		4		5
	5		5		4
	4		5		4
	4		3		4
	5		3		4
	5		5		5
	5		4		4
	4		4		5
	4		2		4
	5		4		2
	5		3		5
	1		2		5
	4		3		5
	4		4		3
	5		5		4
	5		4		5
	4		5		5
	5		3		5
	5		4		5
	5		5		3
	5		5		2
	5		5		4
	5		3		5
	4		2		4
	5		3		5
	5		3		5
	5		5		5
	4		3		4
	5		4		5
	4		5		4
	5		5		3
	4		5		4
	4		3		5
	3		3		3
	5		5		5
	3		5		4
	5		4		4
	5		3		4
	5		5	Average	4.172839506
Average	4.402439024		5		
			5		
		Average	3.988095238		

Using Averages from the external resources such as IMDB and Rotten Tomatoes of movie ratings to predict the rating of 3 movies for Camille, Shachi and Amy.

IMDB	[Inception]	[Avatar]	[The Wolf of Wall Street]
All Age Female	8.7	7.8	7.8
Age <18 Female	8.9	7.6	7.5
Age 18-29 Female	8.8	7.7	7.9
Age 30-44 Female	8.7	7.8	7.8
Age 45+ Female	8.2	8	7.5

3.2 Snapshots for Average (Showing only for Inception on IMDB):





Rating By Demographic

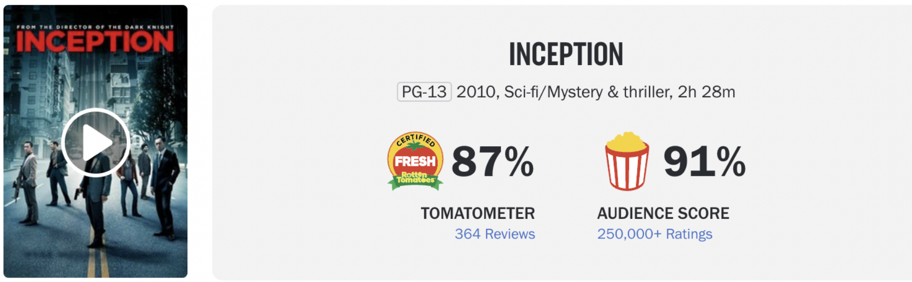
	All Ages	<18	18-29	30-44	45+
All	8.8 2,338,347	9.0 1,033	9.0 358,859	8.8 919,400	8.2 184,263
Males	8.8 1,273,008	9.1 747	9.0 269,912	8.8 736,944	8.2 151,342
Females	8.7 296,070	8.9 194	8.8 76,175	8.7 166,288	8.2 28,493

Top 1000 Voters	US Users	Non-US Users
8.3 909	8.7 470,546	8.8 1,541,851

3.3 Snapshots for Average (Showing only for Rotten Tomatoes):

Rotten Tomatoes	[Inception]	[Avatar]	[The Wolf of Wall Street]
Audience Score	91%	82%	83%
Tomatometer	87%	82%	80%

Snapshots for Average (Showing only for Inception on Rotten Tomatoes):



4.1

Movie	MSBA & DBMI Average	User-User Mean Centered			Differential		
		Shachi	Camille	Amy	Shachi	Camille	Amy
Inception	4.49	4.30	3.23	3.50	0.19	1.26	0.99
Avatar	4.02	4.56	3.20	3.33	-0.54	0.82	0.69
The Wolf of Wall Street	4.12	4.09	3.04	3.46	0.03	1.08	0.66