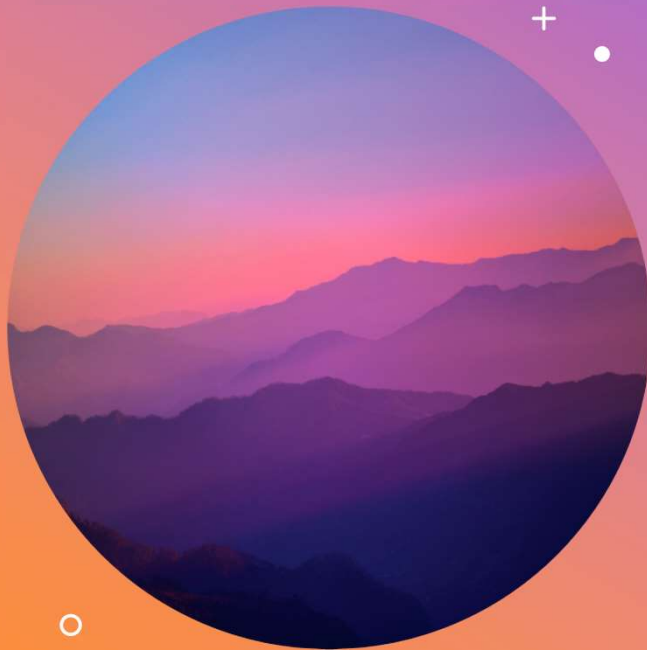


8/28/2022

1

MOVIELENS DATA ANALYSIS



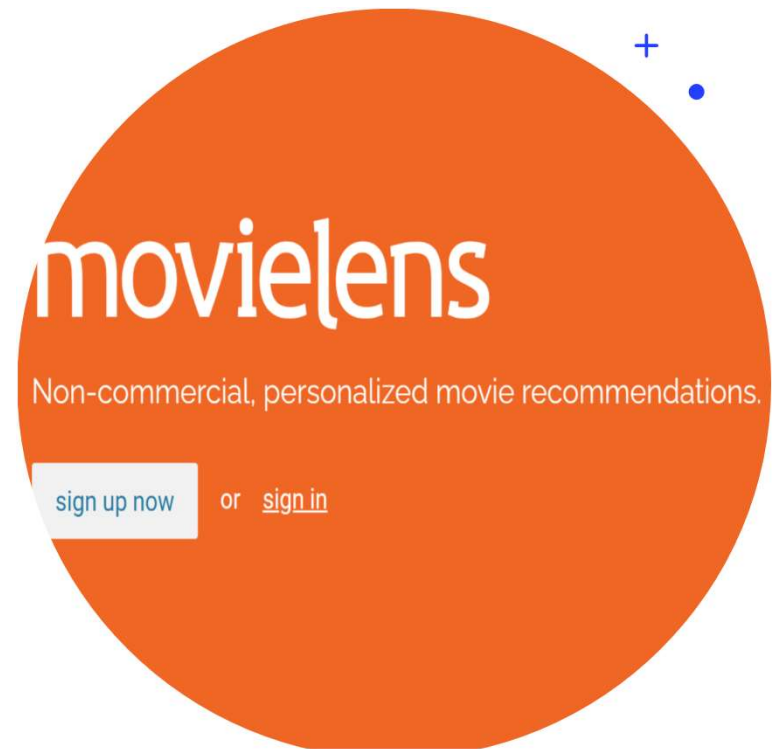
# MOVIELENS RECOMMENDATION SYSTEM

Gamze Turan

# Overview

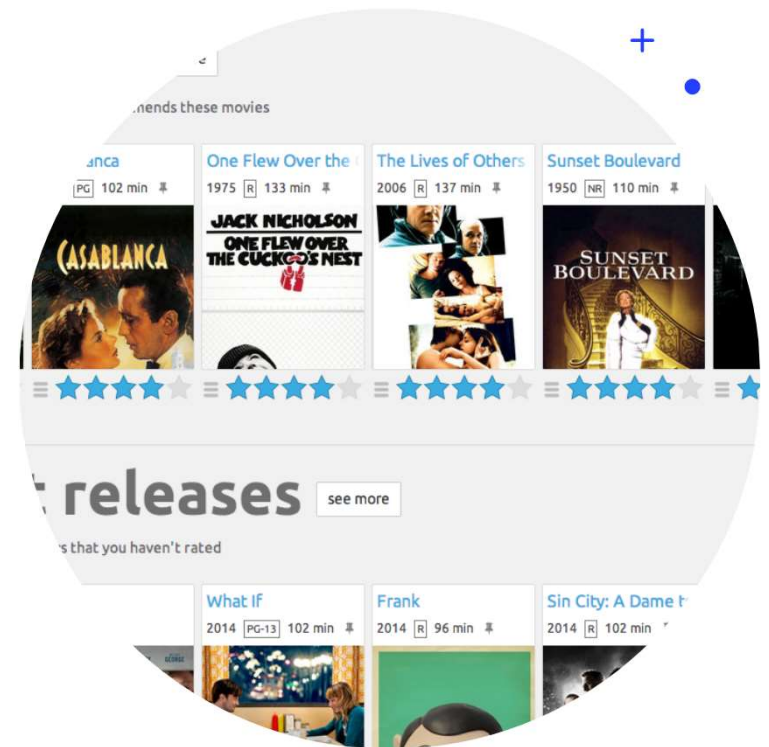
MovieLens is a website that helps users find movies that they will like.

It uses ratings given by the user to build a custom taste profile of that particular user and then utilizes that information to recommend other movies for the user to watch.



# Business Understanding

Our goal is to build a variety of recommendation engines and improve upon predictions iteratively so that the end user can be provided with better movie suggestions.

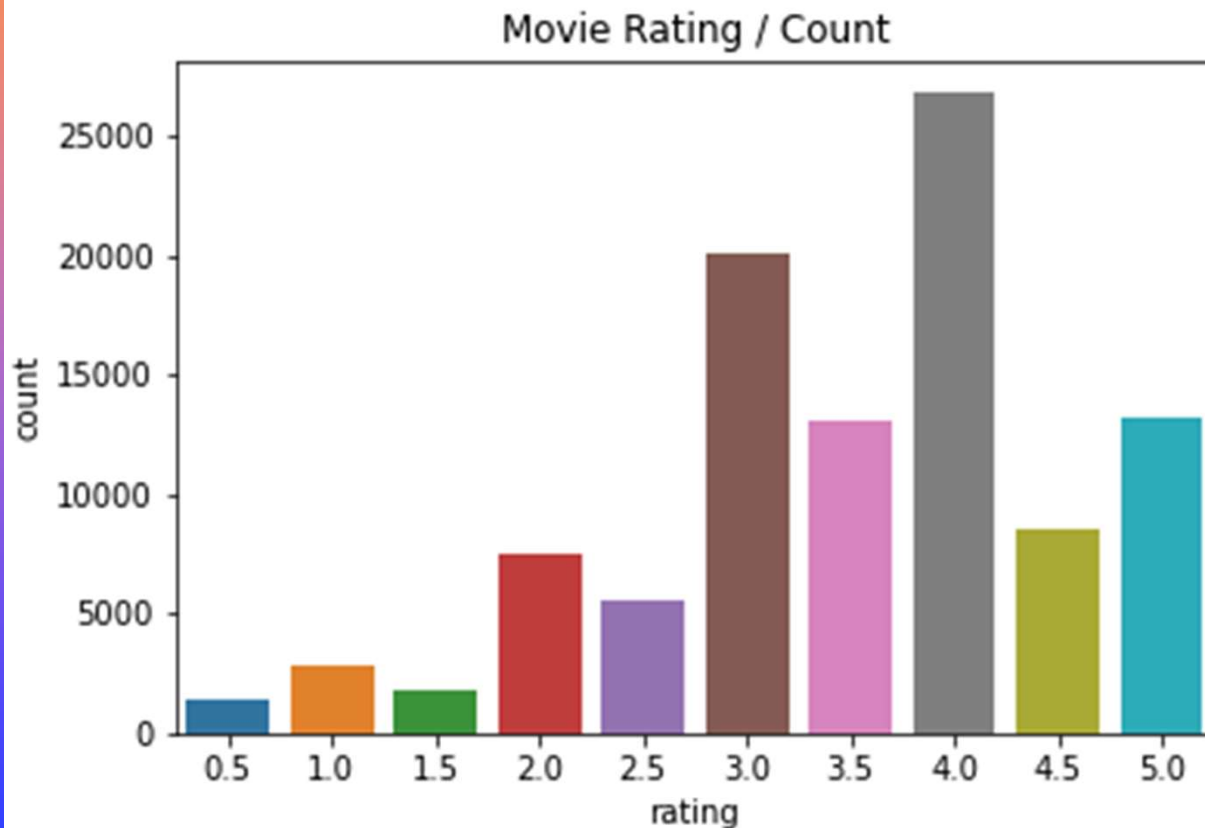


# Data Understanding

- The datasets describe ratings and free-text tagging activities from MovieLens(<https://movielens.org/>), a movie recommendation service.
- It contains 100836 ratings and 3683 tag applications across 9742 movies. These data were created by 610 users.
- The dataset is distributed among four csv files: `links.csv`, `movies.csv`, `ratings.csv`, `tags.csv`.

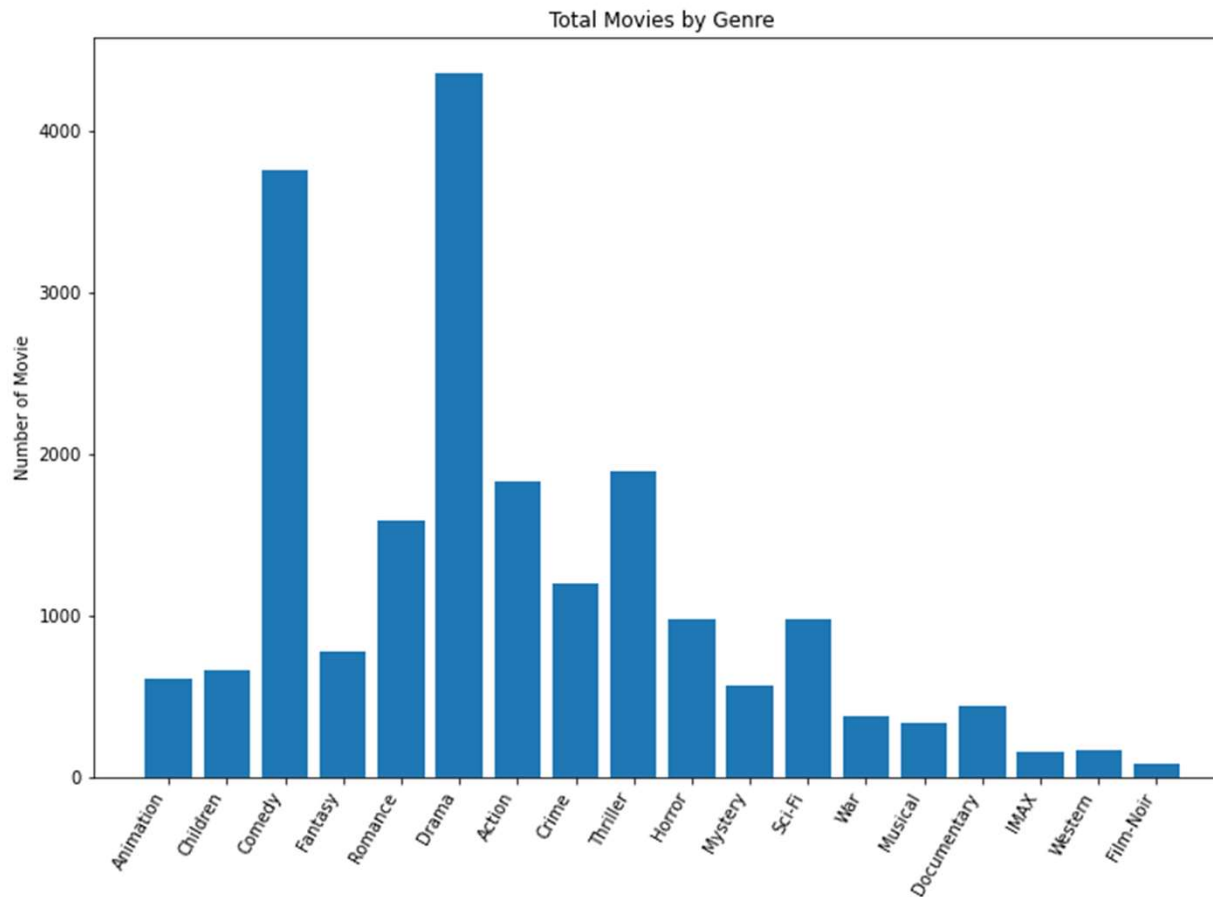
\*\*Source: F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19.  
<https://doi.org/10.1145/2827872>

# Data Analyzing



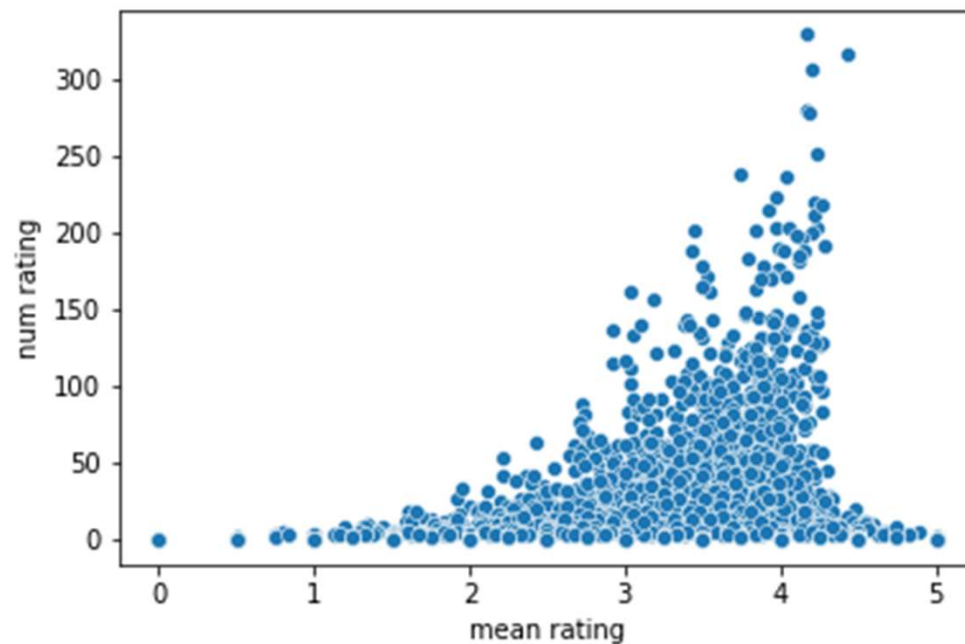
- The mean rating given by users is approximately 3.5.
- The most common rating in the dataset is 4.
- Most of the ratings in the dataset are above 3.

# Data Analyzing



**Drama** is the most popular genre in our dataset with **4357** movies, followed by **Comedy** with **3755** movies.

# Data analyzing



Movies that are more popular (higher number of ratings) in our dataset have higher mean ratings as well.

threshold  100Genre 

These are the recommendations for the users with the following filters  
 Minimum number of ratings: 100  
 User's choice of genre: All

	title	mean rating	num rating
277	Shawshank Redemption, The (1994)	4.429022	317.0
659	Godfather, The (1972)	4.289062	192.0
2226	Fight Club (1999)	4.272936	218.0
922	Godfather: Part II, The (1974)	4.259690	129.0
6313	Departed, The (2006)	4.252336	107.0
914	Goodfellas (1990)	4.250000	126.0
6708	Dark Knight, The (2008)	4.238255	149.0
46	Usual Suspects, The (1995)	4.237745	204.0
899	Princess Bride, The (1987)	4.232394	142.0
224	Star Wars: Episode IV - A New Hope (1977)	4.231076	251.0

threshold  105Genre 

These are the recommendations for the users with the following filters  
 Minimum number of ratings: 105  
 User's choice of genre: Animation

	title	mean rating	num rating
4360	Finding Nemo (2003)	3.960993	141.0
322	Lion King, The (1994)	3.941860	172.0
0	Toy Story (1995)	3.920930	215.0
3568	Monsters, Inc. (2001)	3.871212	132.0
3194	Shrek (2001)	3.867647	170.0
5374	Incredibles, The (2004)	3.836000	125.0
506	Aladdin (1992)	3.792350	183.0
512	Beauty and the Beast (1991)	3.770548	146.0

# New User – Naive Recommendation Engine

- Model recommends the most popular movies in our dataset. Movies are sorted by both the number of ratings and their average rating in our dataset.
- I allow users to specify their preferred genre as well.



User 1 Movie Ratings	
Movie Title	Rating
Iron Man (2008)	5.0
The Fault in our Stars (2014)	2.0
Avengers: Age of Ultron (2015)	4.5

User 2 Movie Rating	
Movie Title	Rating
Iron Man (2008)	5.0
The Notebook (2004)	2.5
Captain America: The First Avenger (2011)	4.5

Recommend Captain America: Age of Ultron (2015) to User 1.

Recommend Avengers: Age of Ultron (2015) to User 2.

## Existing User – Collaborative-Filtering based Recommendation Engine

Model recommends those movies that haven't been watched by this user and were loved by other users with similar taste.

# Recommended movies

These are the recommendations for the users with the following filters

User id: new

Minimum number of ratings: 100

User's choice of genre: Children

	title	mean rating	num rating
6770	WALL·E (2008)	4.057692	104.0
7036	Up (2009)	4.004762	105.0
4360	Finding Nemo (2003)	3.960993	141.0
322	Lion King, The (1994)	3.941860	172.0
0	Toy Story (1995)	3.920930	215.0
815	Willy Wonka & the Chocolate Factory (1971)	3.873950	119.0
3568	Monsters, Inc. (2001)	3.871212	132.0
3194	Shrek (2001)	3.867647	170.0
5374	Incredibles, The (2004)	3.836000	125.0
506	Aladdin (1992)	3.792350	183.0

When user is new, use naive recommendation engine and only look at the most popular movies.

These are the recommendations for the users with the following filters

User id: 5

Minimum number of ratings: 100

User's choice of genre: Children

	movieid	Expected Rating	title	num rating
8	4306	3.9	Shrek (2001)	170.0
6	6377	3.9	Finding Nemo (2003)	141.0
1	8961	3.9	Incredibles, The (2004)	125.0
4	1097	3.9	E.T. the Extra-Terrestrial (1982)	122.0
10	68954	3.9	Up (2009)	105.0

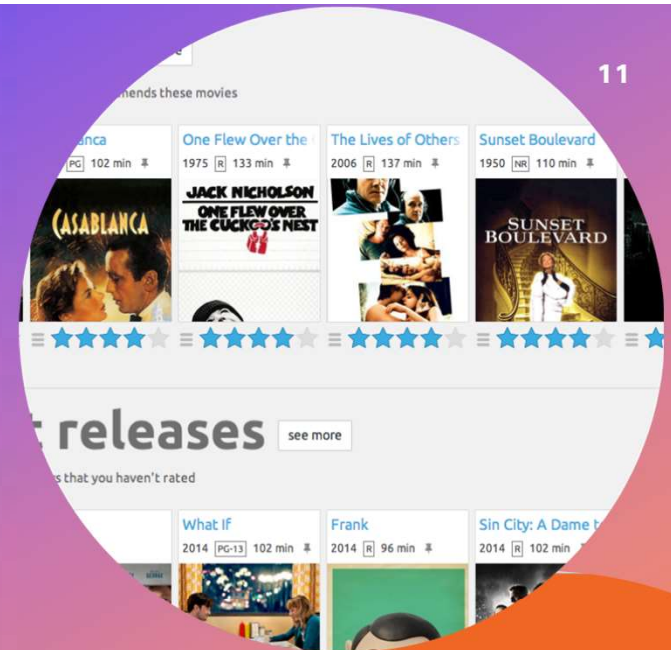
When user has provided some ratings in the database, consider that information and try to find similar movies based on his/her taste.

# CONCLUSION

I analyzed a variety of movie recommendation systems on the famous MovieLens database and provided general recommendations based upon movie popularity or the average ratings given by other users in the database.

Then progressed to some collaborative filtering based engines which try to find similar movies or users to make their predictions.

Finally, hybrid system of our naive recommendation engine to filter options for genre and minimum number of ratings to give users some control over these recommendations.



## movielens

Non-commercial, personalized movie re

[sign up now](#)

or [sign in](#)



# Future Work



## **Allow new user ratings**

I would like to add functionality in our final dashboard to allow new users to rate some movies and then to utilize that information to improve our recommendation system.

## **Utilize tag information**

I also didn't make use of tag information in this part of the analysis. We would like to make word embeddings from tags and other meta information about the movie and use it in our model.

## **Scrape movie metadata from the internet**

I can also make use of the links dataset and scrape more information about each movie from the internet. This could involve significant features like cast, director, plot, etc.

8/28/2022

13

+



o



# THANK YOU

Email: [ginaturan15@gmail.com](mailto:ginaturan15@gmail.com)

Github: @gt1719

Linkedin: <https://www.linkedin.com/in/gamze-turan-958a8012b/>

•