



What can we learn from Movie Ratings?

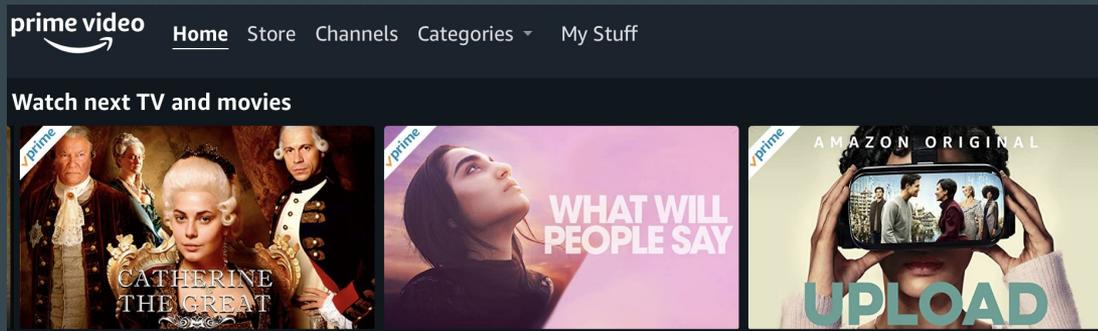


CPSC 547 Info Viz

Niloofer Zarif, Deepansha Chhabra, Lucie Polakova

Introduction

- What are recommendation systems?
- How are they built and trained?
- What is the design trend?
 - Collaborative filtering and content-based filtering
 - Deep learning models
 - Pipelined systems: frontend & backend
- What are the design challenges now?
- How can we help?



Our Project :

- Use data to extract knowledge
- Use visualization as a means for both discovery and presentation
- Use cases:
 - ML Engineer
 - Movie Producer
 - Theatre Manager

Data Collection & Sources:

- Kaggle
 - Netflix
 - Crawled
 - The Movie Database
- Ratings
 - User privacy
- Movies
 - Meaningful properties
- Merging

Data abstraction and data aggregation

- Datasets of 1,100 items (movies) and 26 million items (ratings)

| Categorical | Ordered |
|------------------------|--------------------------|
| Original language: 9 | Budget: 0-400,000,000 |
| Genre: 19 | Revenue: 0-1,900,00,000 |
| View rating: 9 | Year: 1930-2020 |
| Release month: 12 | Rating: 1-5 |
| Languages spoken: 26 | Runtime: 70-210 |
| Original languages: 10 | Number of votes: 1-5,639 |



- Splitting a list of genres into multiple strings
- Aggregating budget/revenue into 10-15 groups
- Aggregating languages spoken into 6 groups
- Filtering outliers and items with missing attributes



Live demo:

https://public.tableau.com/views/MovieRatings_16395453139120/Whatisthebestgenre?:language=en-GB&publish=yes&:display_count=n&:origin=viz_share_link

Rating vs. time

Ratings over time

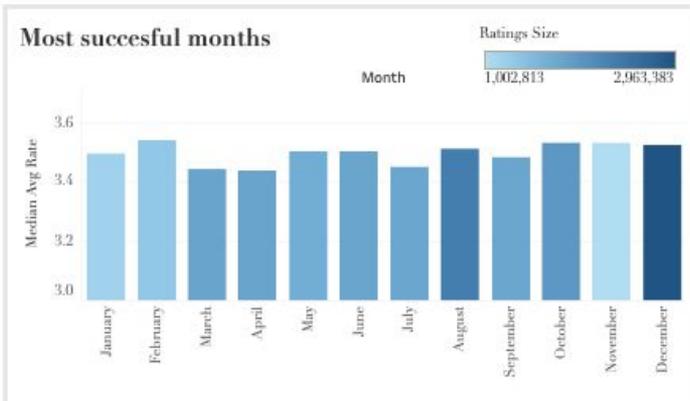
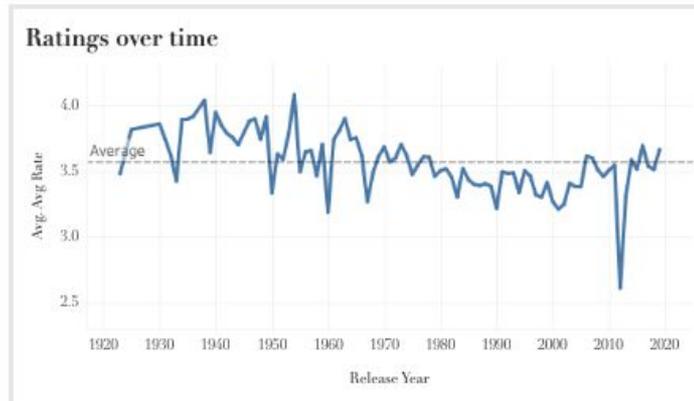
- see the development of average rating over time (e.g. movies were generally better before 1975, then it became less consistent)

Most successful months:

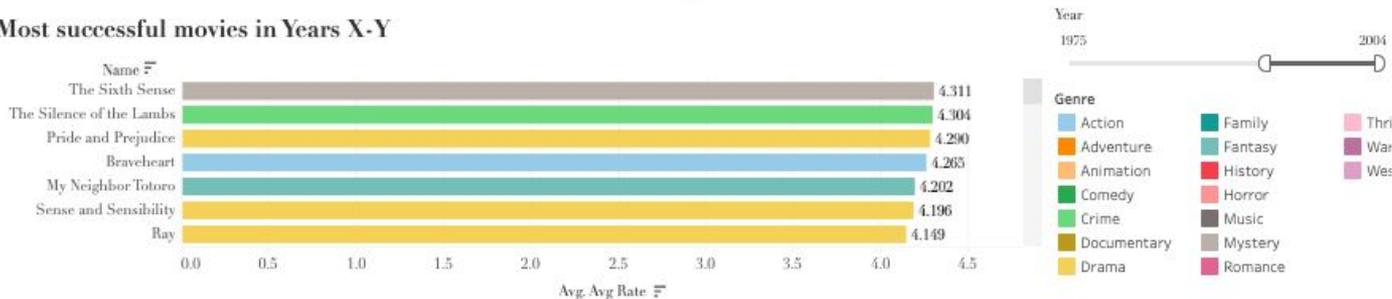
- see how which months the highest rated movies came out and what's the proportion to other months (e.g. in December the most famous movies come out, however, the highest rated movies usually came out in October/November)

Most successful movies in years X-Y:

- explore the highest rated movies throughout history and their main genres using the filter on the right



Most successful movies in Years X-Y



Movies Across the World

Average rating vs Country availability:

- explore and compare different countries throughout the years by filters on the right [e.g. movies available in Brazil tend to get low rating across all years]

Spoken Languages:

- see the rating based on how many languages are spoken in the movie [e.g. movies with 3 languages spoken are statistically less popular]

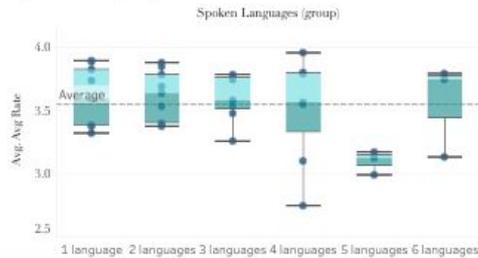
Original Language:

- see the rating based on original languages of the movies [e.g. original korean movies are on average the least popular]

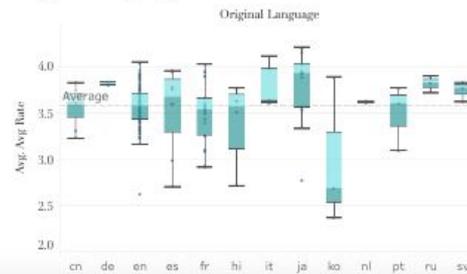
Average rating vs Country availability



Spoken Languages



Original Language



Do money matter?

Budget:

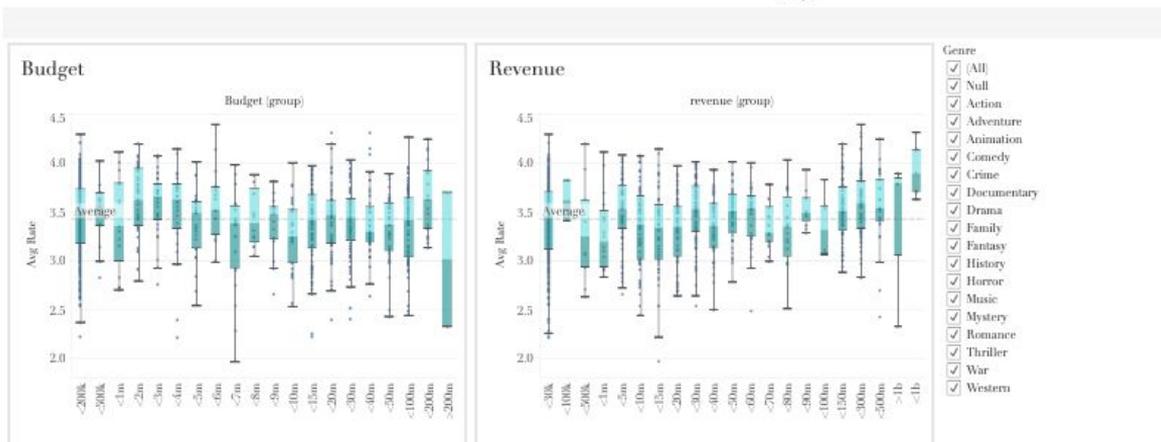
explore the user rating based on the budget for the movie and different genres (e.g. low budget movies span across the whole scale, however movies for \$1-2m typically have higher than average rating)

Spoken Languages:

explore the user rating based on the revenue for the movie and different genres (e.g. documentaries tend to have above-average rating no matter the revenue)

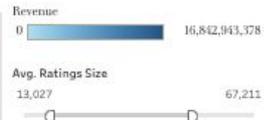
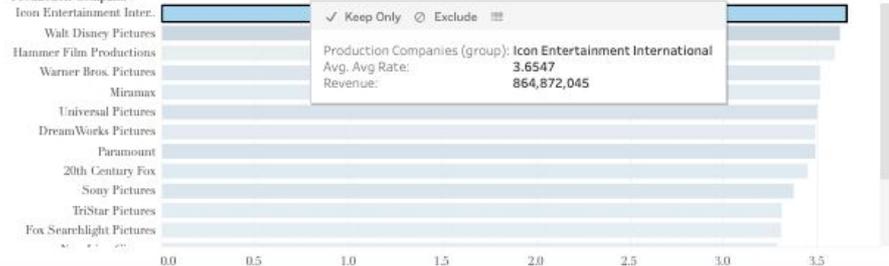
Production companies rating:

see the rating based on the production companies and explore the revenue based on the number of ratings on the movies from the respective companies (e.g. Icon Entertainment International has a higher average rating than Paramount, who has significantly higher number of ratings)



Production companies rating

Production Compan., ▾



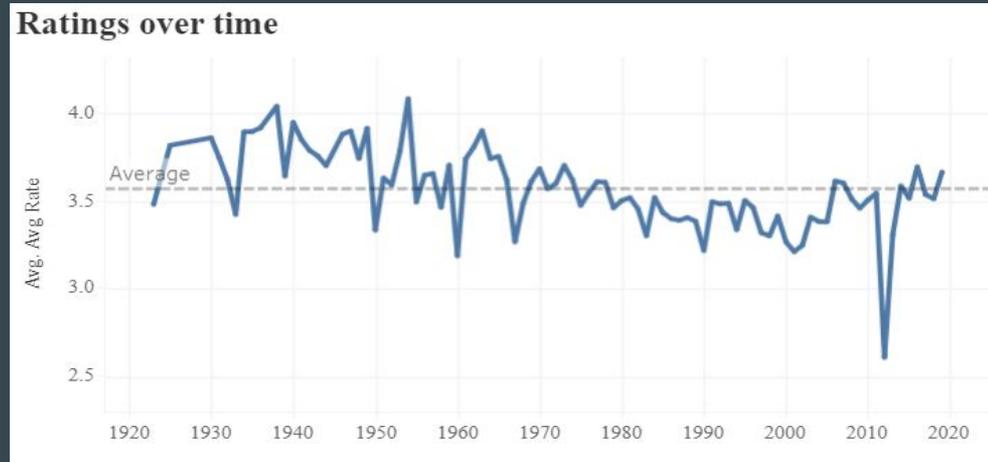
Avg. Avg. Rate ▾

Main Findings/Results:

- Movies with the Genre Mystery have maintained a consistent rating of 3 and above .
- Movies released in February, October received the highest ratings.
- December has the highest number of ratings.
- The genres drama and comedy are more popular.
- Movies with lower budget experienced fluctuations in the average rating.
- Documentaries maintained a higher rating overall.
- Movies more popular in Russia and australia.



Main Findings/Results:



- Production companies like Icon Entertainment International had the highest average rating.
- Movies with less number of languages spoken are more popular.
- Original Korean movies are least popular.
- Crime-Comedy movies are not as popular as the crime-mystery movies.
- On the overall, movies received higher ratings before 1975 .

Limitations/Future Work

- Predict trends for the future.
- Using larger dataset for better analyze the trends.
- Include more Interactivity in the visualizations by providing options of different viz techniques for the user to choose from.
- Could explore other complex visualizations which could help to analyze data in a more detailed manner.

Thank you

Q & A

