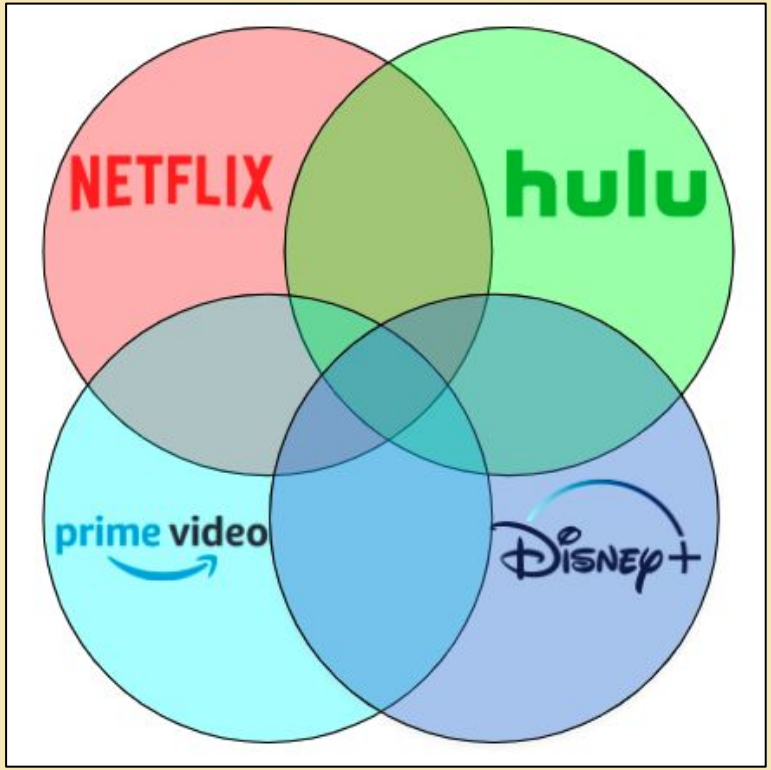
Exploratory Analysis of Streaming Platform Content

Katrina Agate
Shivani Athavale
Laura Becerra
Jake Daly
Jinyu Zhao



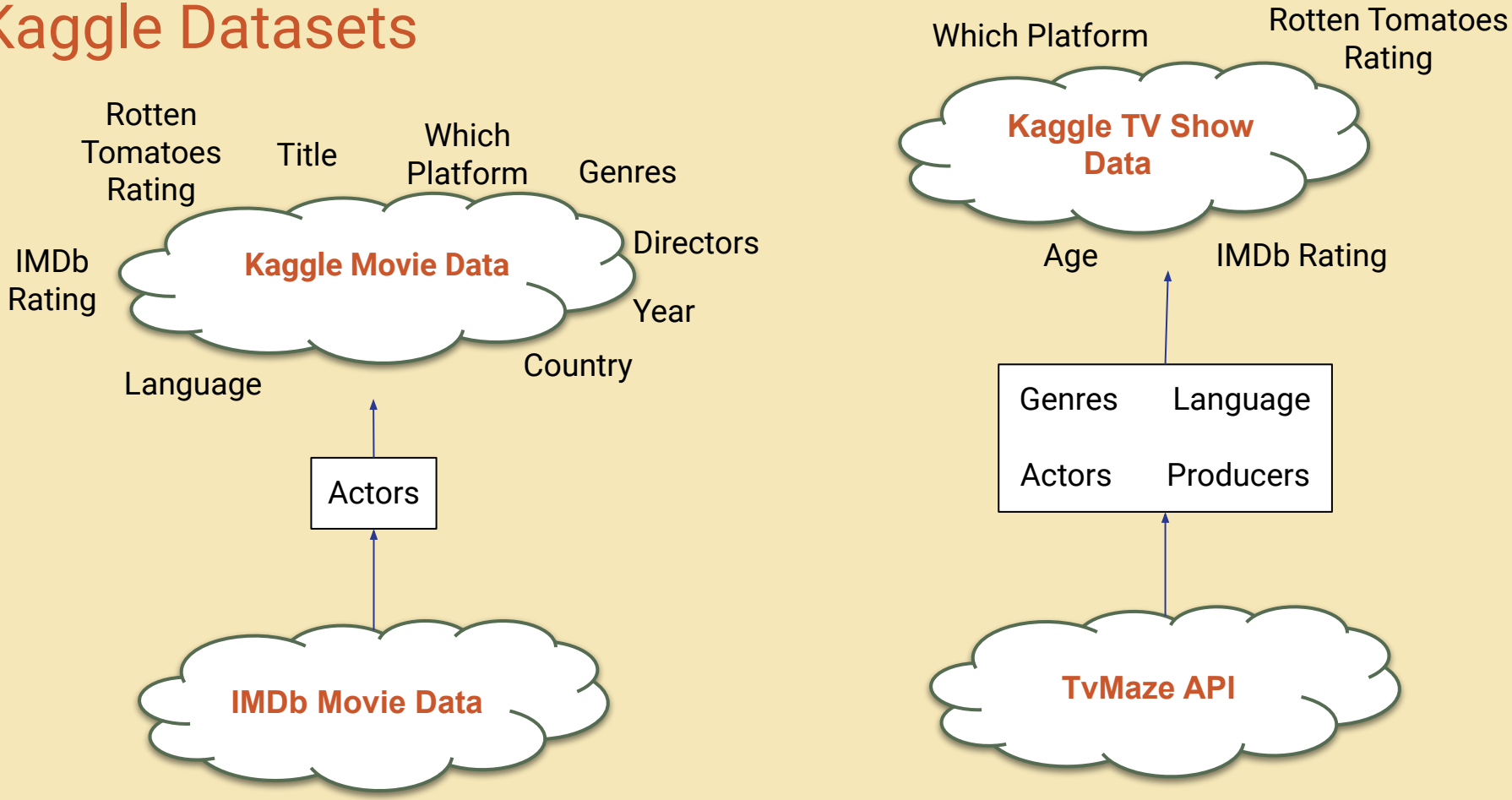
Project Motivation

How do different movie and TV show streaming platforms differ from each other in their content?

Project Motivation

Which platform would we recommend to someone?

Kaggle Datasets



Dataset Description: Movies

Kaggle dataset:

Title	Year	Age	IMDB rating	Rotten Tomatoes rating	Netflix	Hulu	Prime Video	Disney+
Inception	2010	13+	8.8	87%	1	0	0	0

Directors	Genres	Country	Language	Runtime
Christopher Nolan	Action,Adventure,Sci-Fi,Thriller	United States	English	148

Extracted from IMDb Movie dataset:

Actors
Leonardo DiCaprio, Joseph Gordon-Levitt, Ellen Page, Ken Watanabe

Dataset Description: TV Shows

Kaggle dataset:

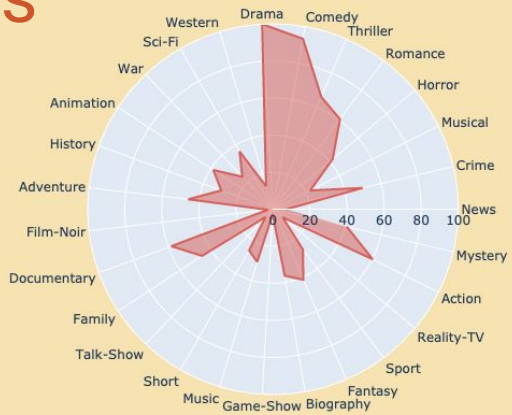
Title	Year	Age	IMDB rating	Rotten Tomatoes rating	Netflix	Hulu	Prime Video	Disney+
Breaking Bad	2008	18+	9.5	96%	1	0	0	0

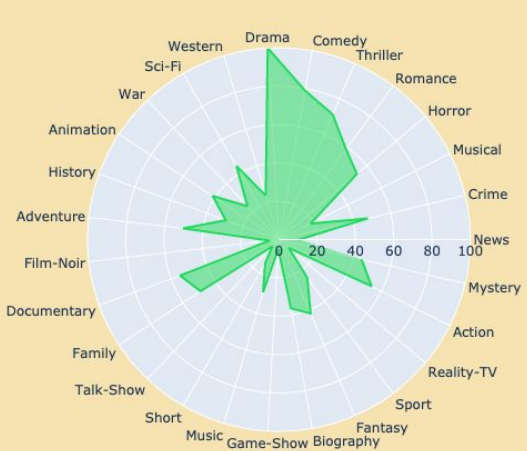
Extracted from TvMaze API:

Actors	Producers	Genres	Language
Bryan Cranston,Aaron Paul,...Laura Fraser	Thomas Schnauz,Michelle MacLaren,...Melissa Bernstein	Drama, Crime, Thriller	English

Movie Genres

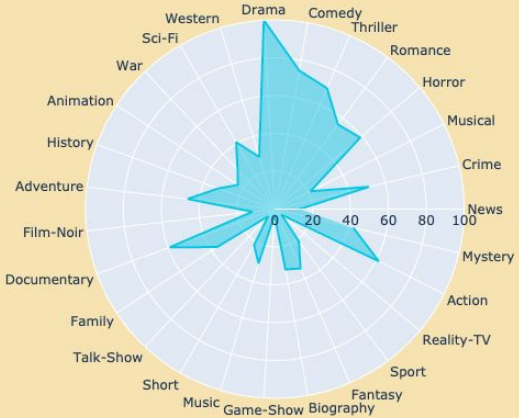
NETFLIX

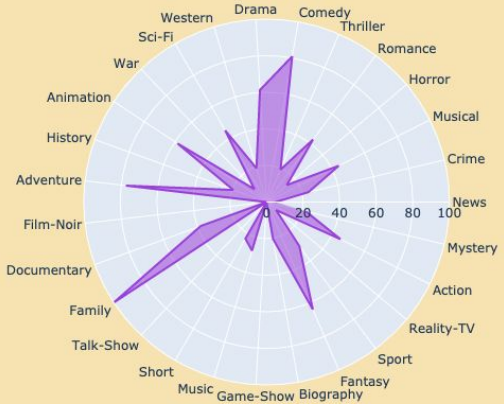




hulu

prime video





Disney+

Movie Description

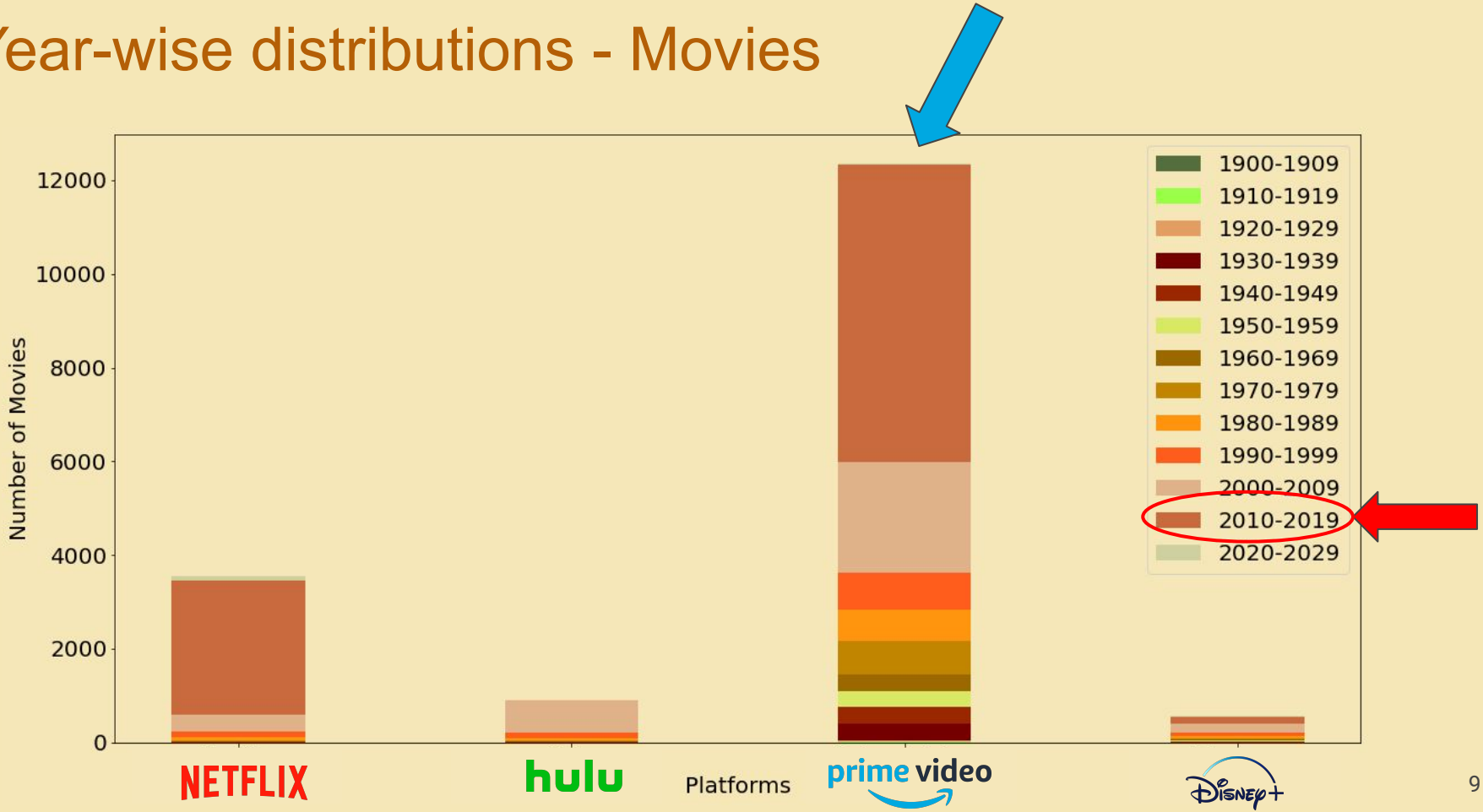
NETFLIX

hulu

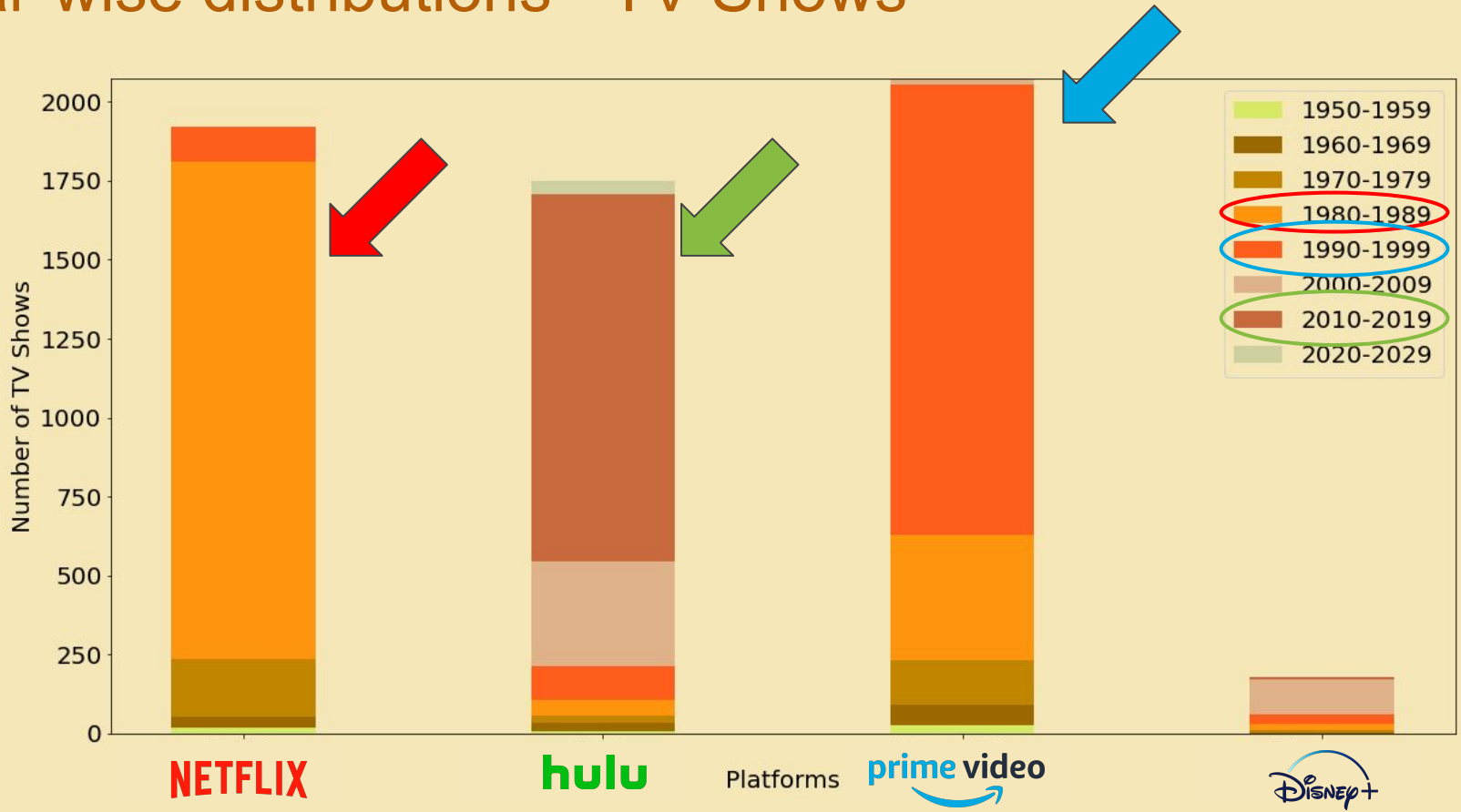
prime video

Disney+

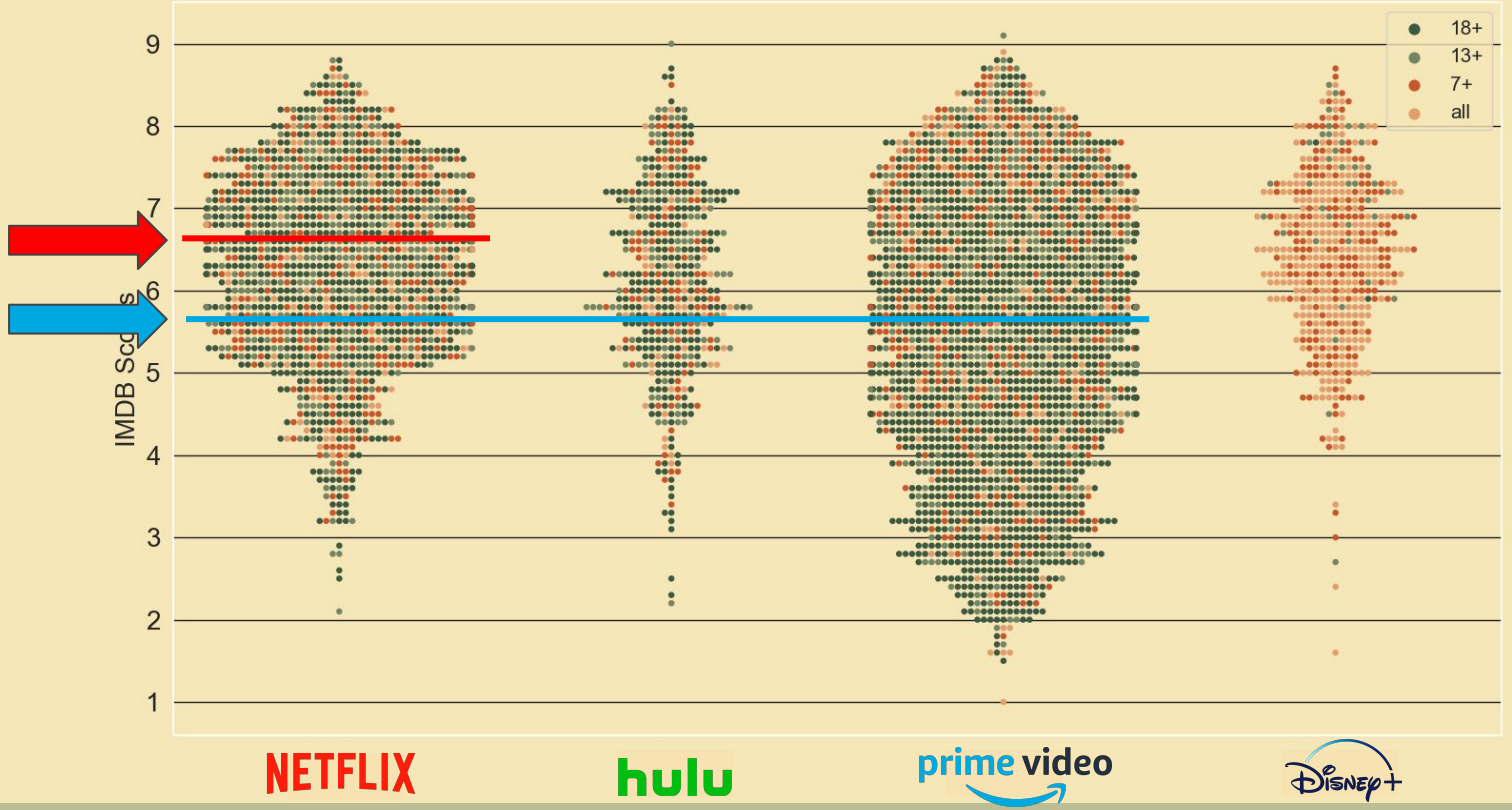
Year-wise distributions - Movies



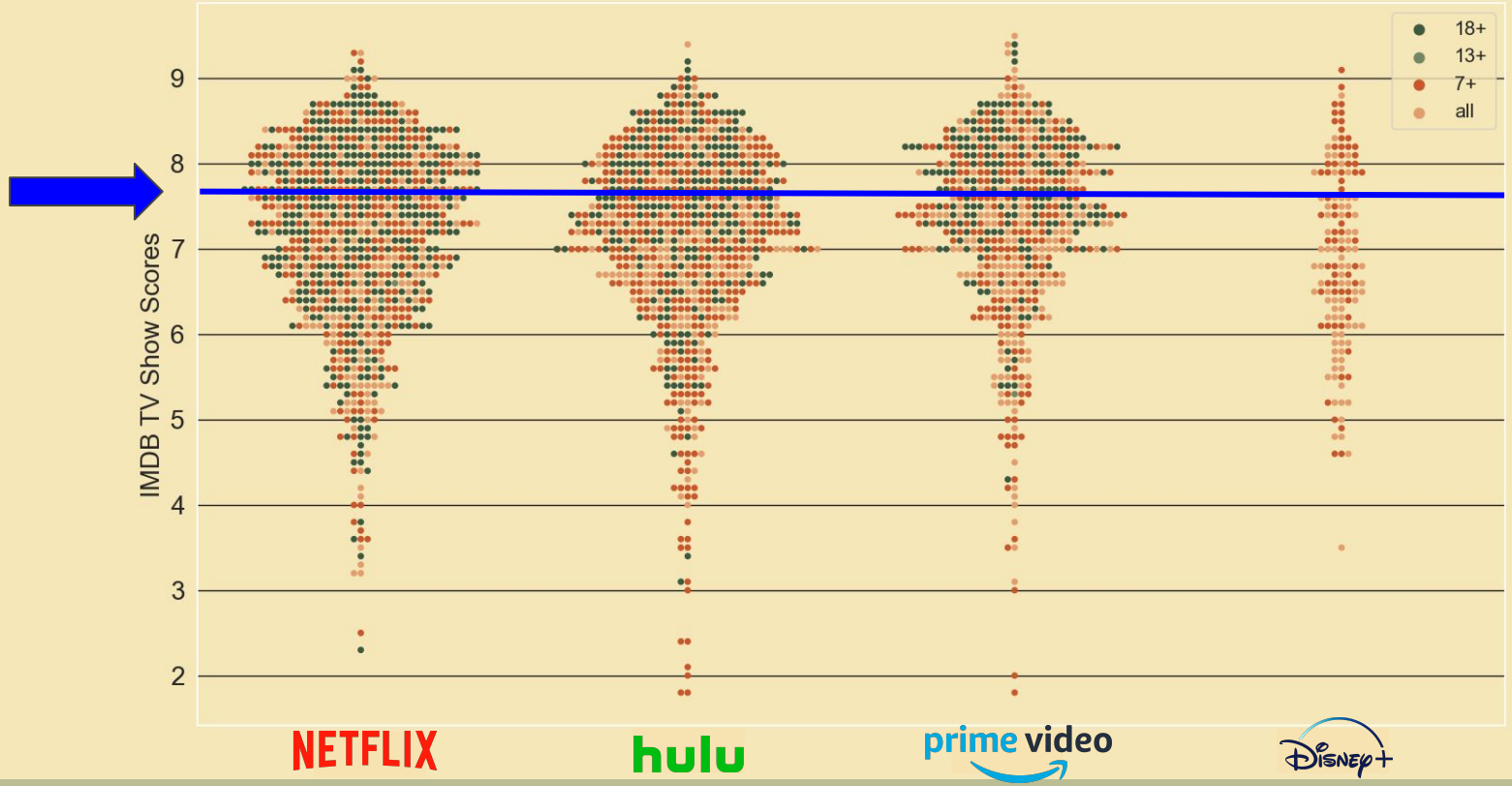
Year-wise distributions - TV Shows



IMDB Movie Scores and Ages



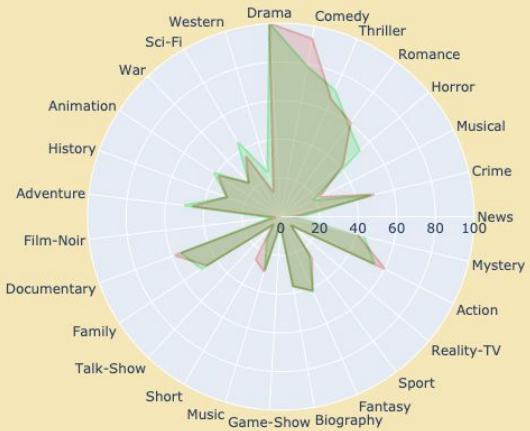
IMDB TV Show Scores and Ages



Recommendation Settings

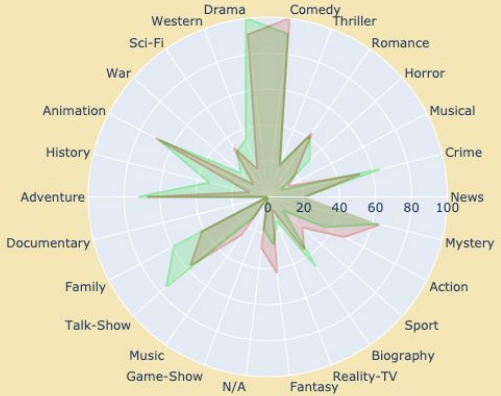
- How do we model the platforms?
- How do we model the users?
- How do we measure similarity?

Movie Genres



NETFLIX

hulu



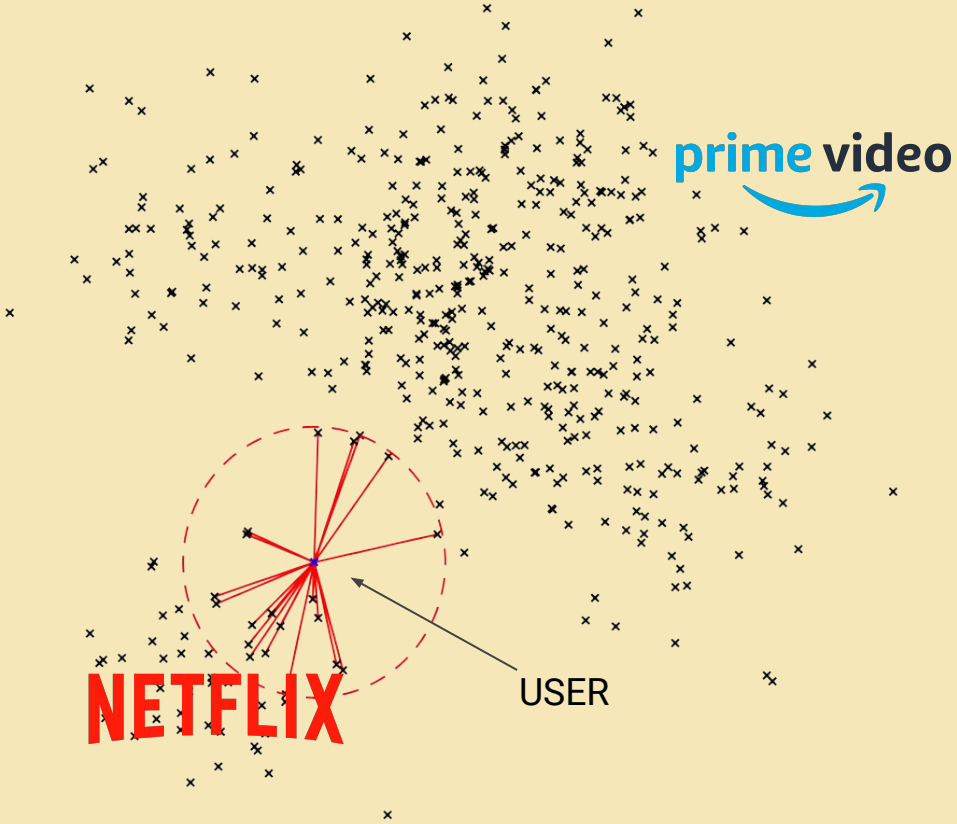
TV Show Genres

Our Approach

- Content-based show recommendation
  - TF-IDF statistics of their plot summary
  - Concatenate with one-hot vector encoding other features

- K nearest neighbor
  - Find closest cluster(platform) using cosine similarity

K Nearest Neighbor

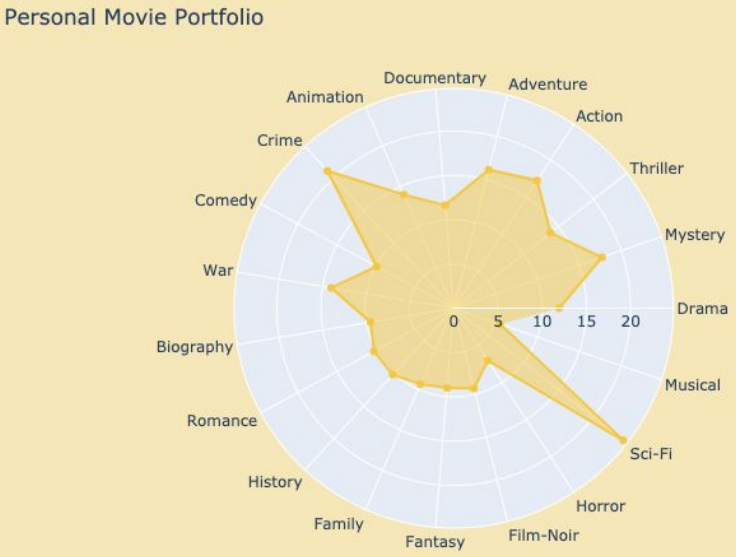


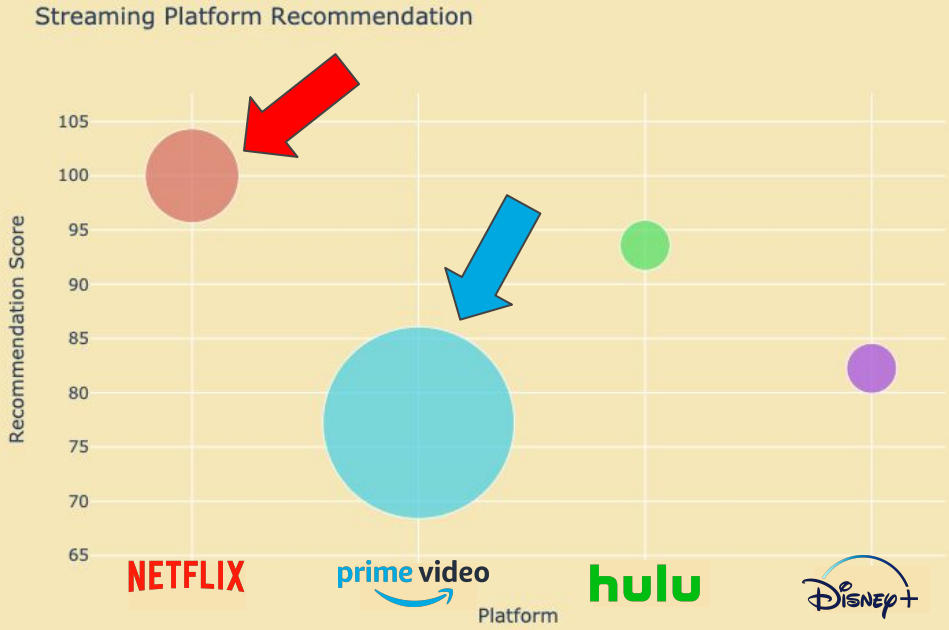
Demo

A sample user who has watched

Inception, Avatar, Back to the Future and The Dark Knight

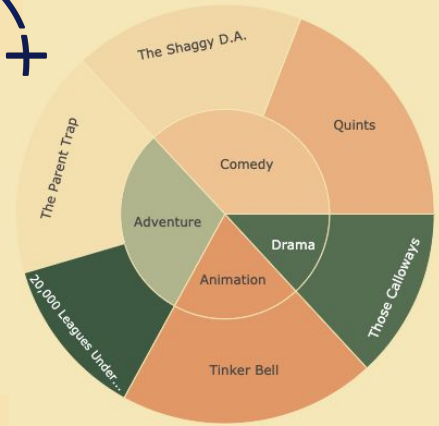
Personal Movie Portfolio



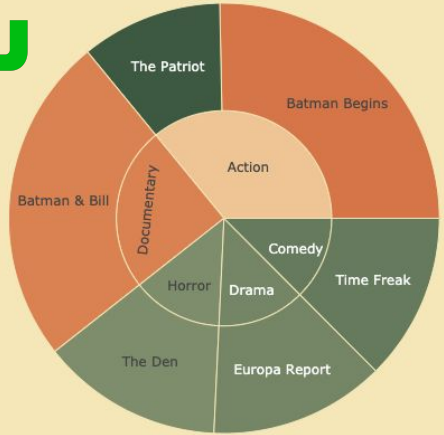


NETFLIX

Disney+

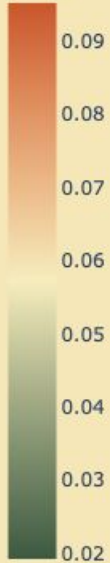


hulu



prime video

Score



A closer look ...

NETFLIX

Conclusions

➡ Quantity over quality

➡ Widest selection/largest balance of genres

➡ Most content from recent years (21st century)

➡ Least content, mostly for children

- **Recommender System:**
  - Clustering algorithm and recommender system combined
  - Potential direction for recommender system research

Thank you!