

Escape the bubble!

A trade-off between personalization
and diversity in movie recommendations

About us



**Alexandra
Zimmermann-Rösner**
Dipl.- Biologist

📍 SH, HH, remote



Hassan Mohamed, MSc.
Physicist

📍 Leipzig, Berlin,
remote



Konstanze Braun, MSc.
Psychologist

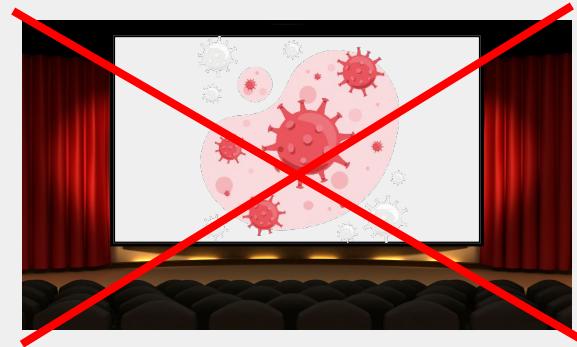
📍 throughout Germany,
remote



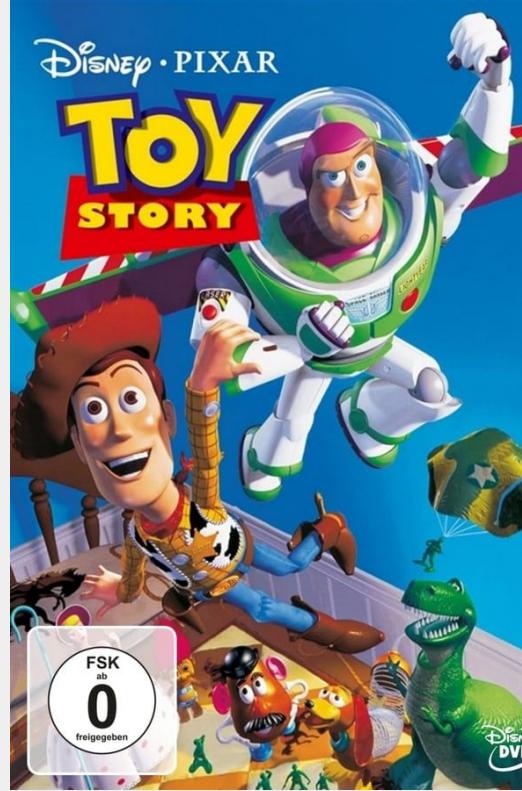
Jana Conradi, MSc.
Bio Scientist

📍 remote, hybrid in
NI, HB, HH

Meet Mia



The seed item

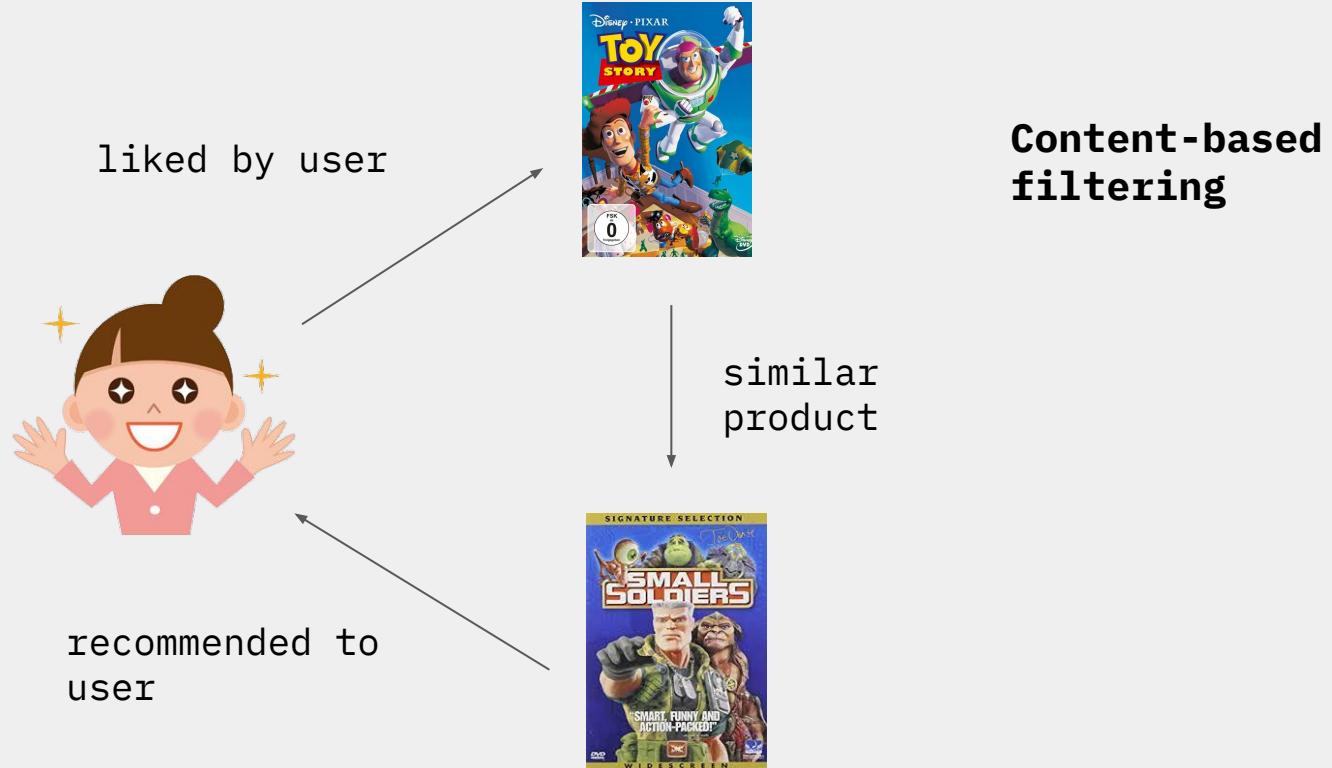


Recommen- dations



What are recommender systems?

Information Aggregation on user's preferences

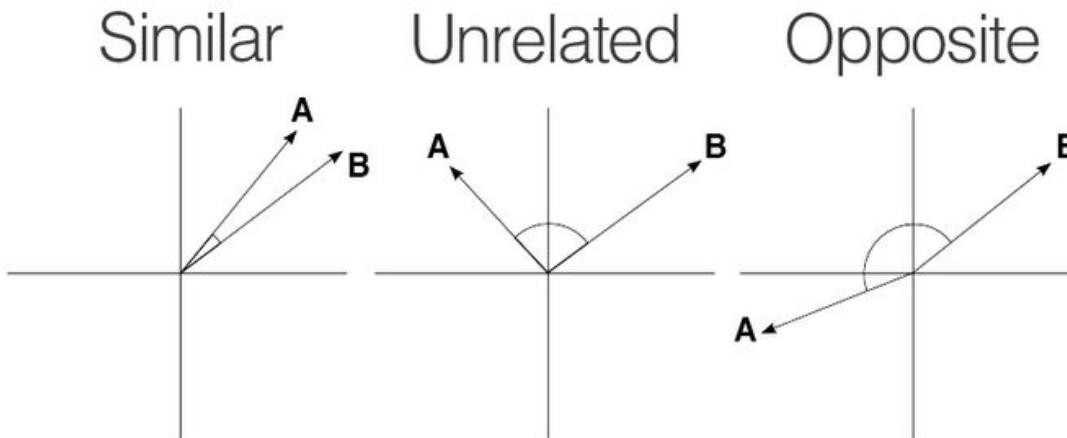


How to compute similarity

The Movie Database (TMDB)

- 45,000 movies
- movie features: description, keywords, genres

cosine similarity of the movies



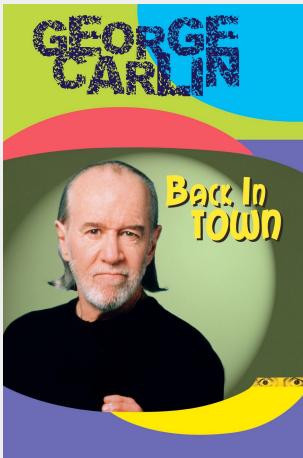
Source: [pyimagesearch](#)

Limitations

...works only with well described movies

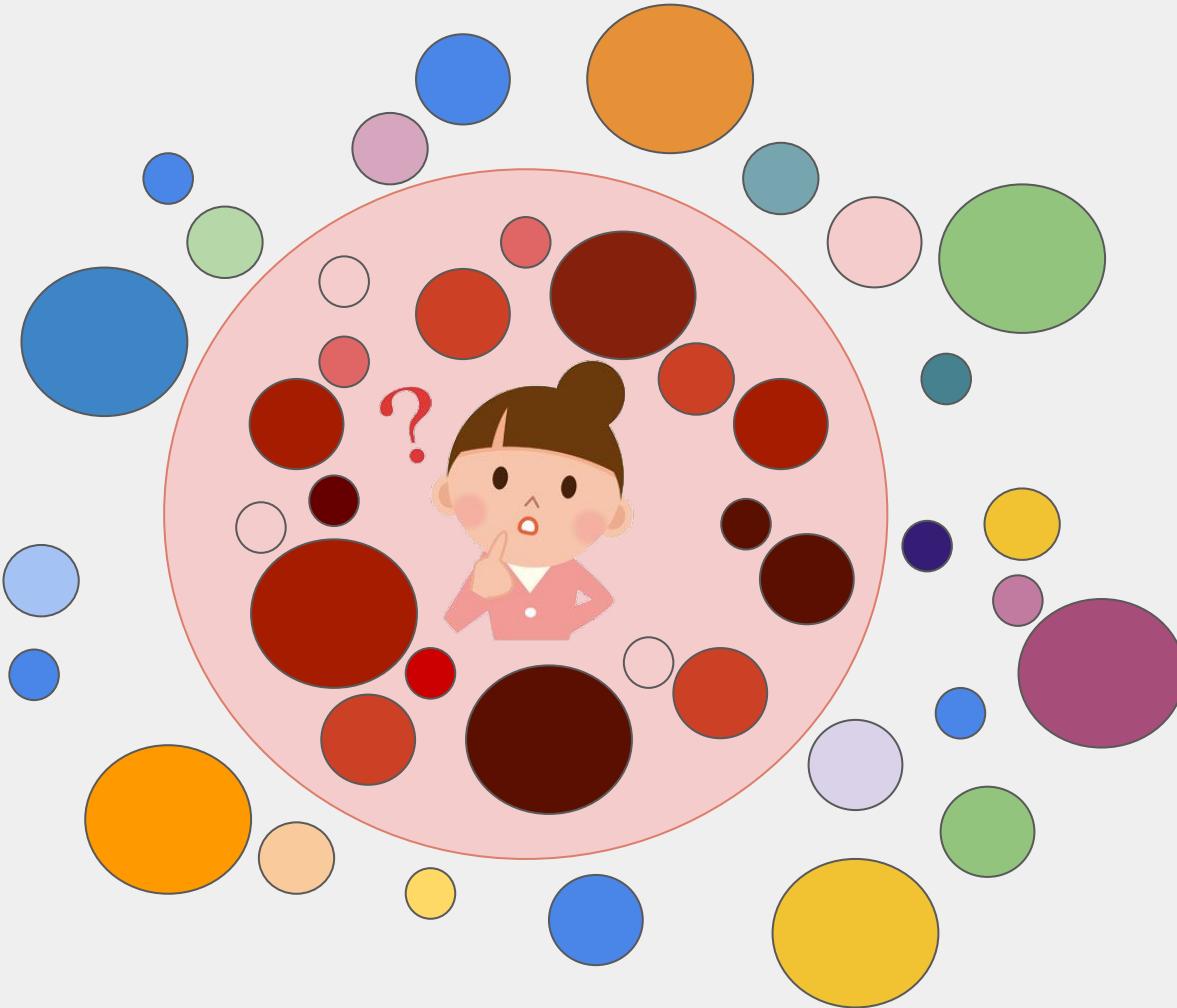
sparse informations problematic:

- old movies
- niche movies



only genre -> high similarity

The filter bubble



Collaborative Filtering:

The Idea

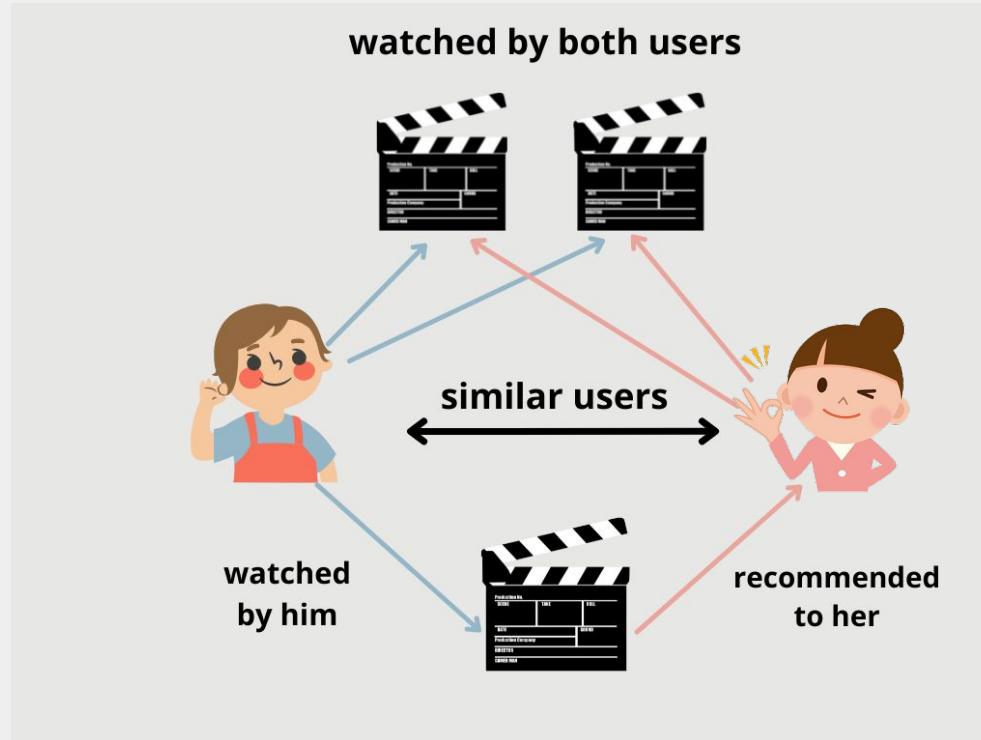


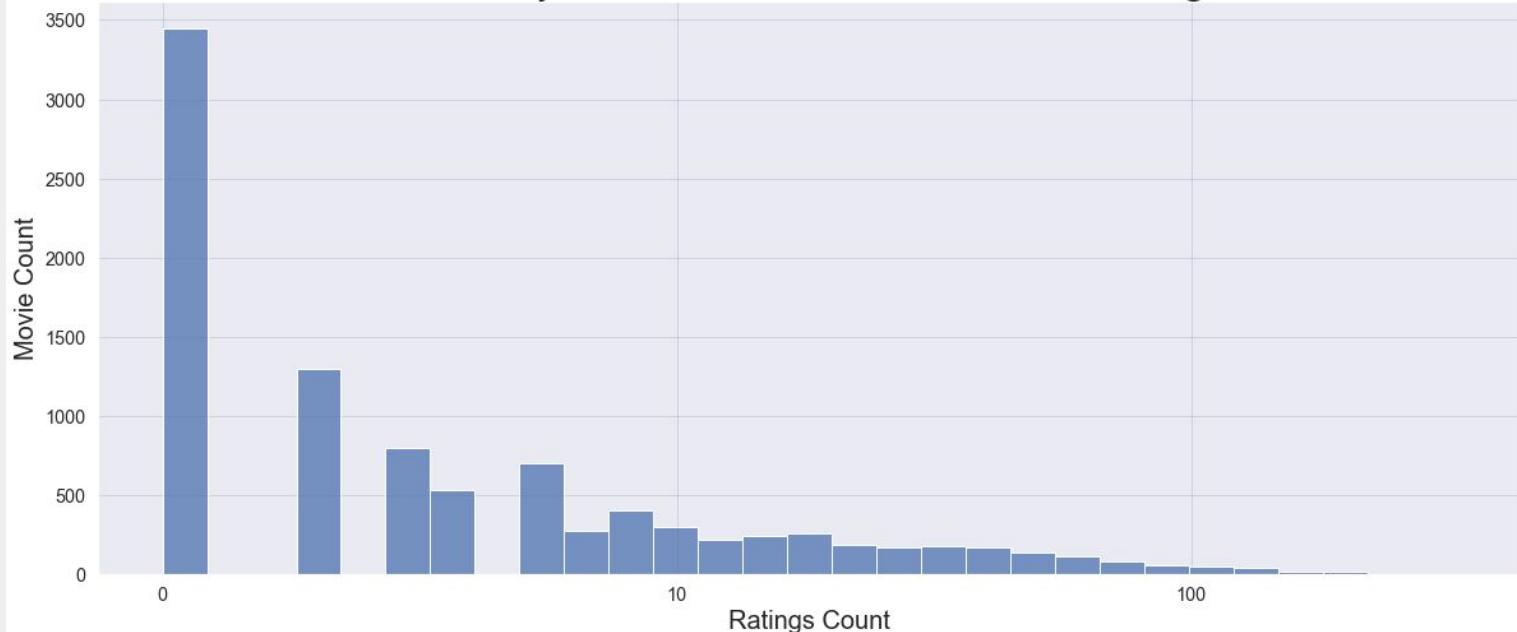
image created via Canva

How many ratings are there per movie?

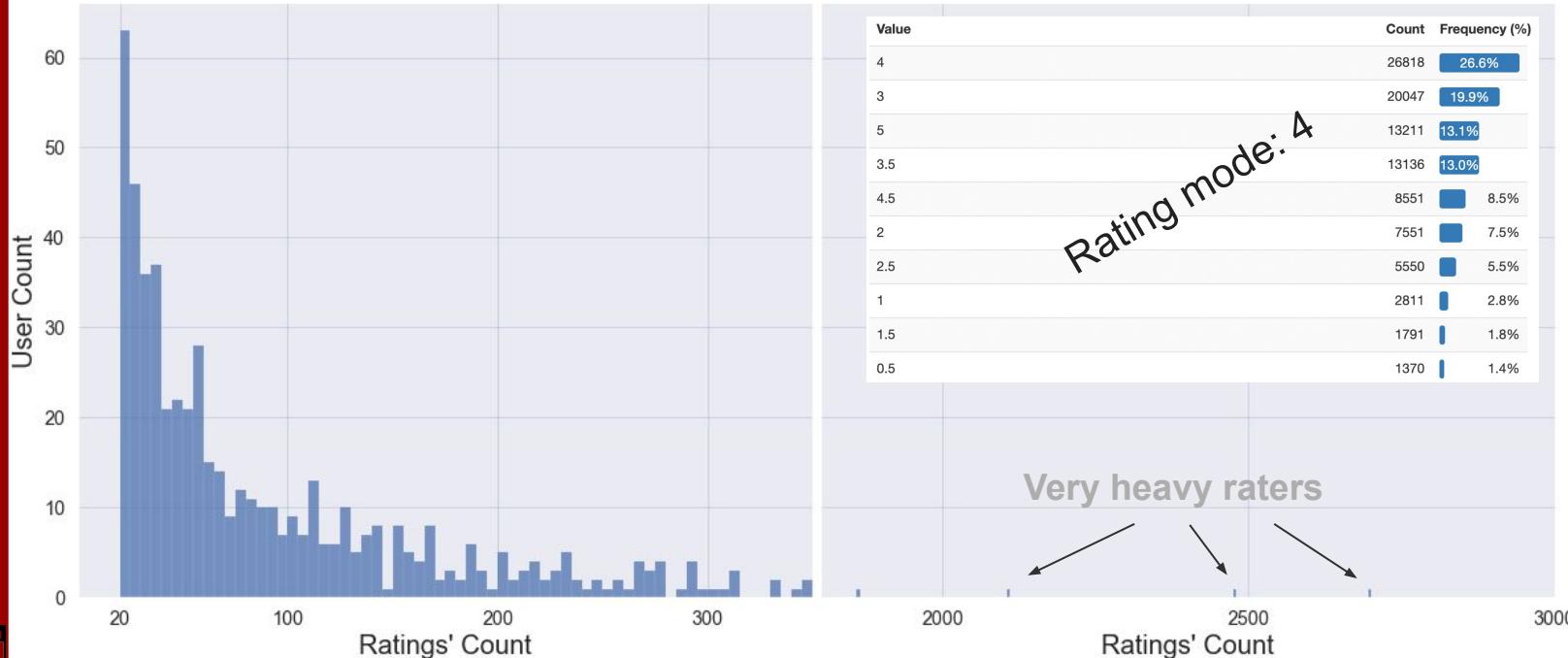
MovieLens Dataset (*small*)

- 100,000 ratings
- 9,500 movies
- 610 users

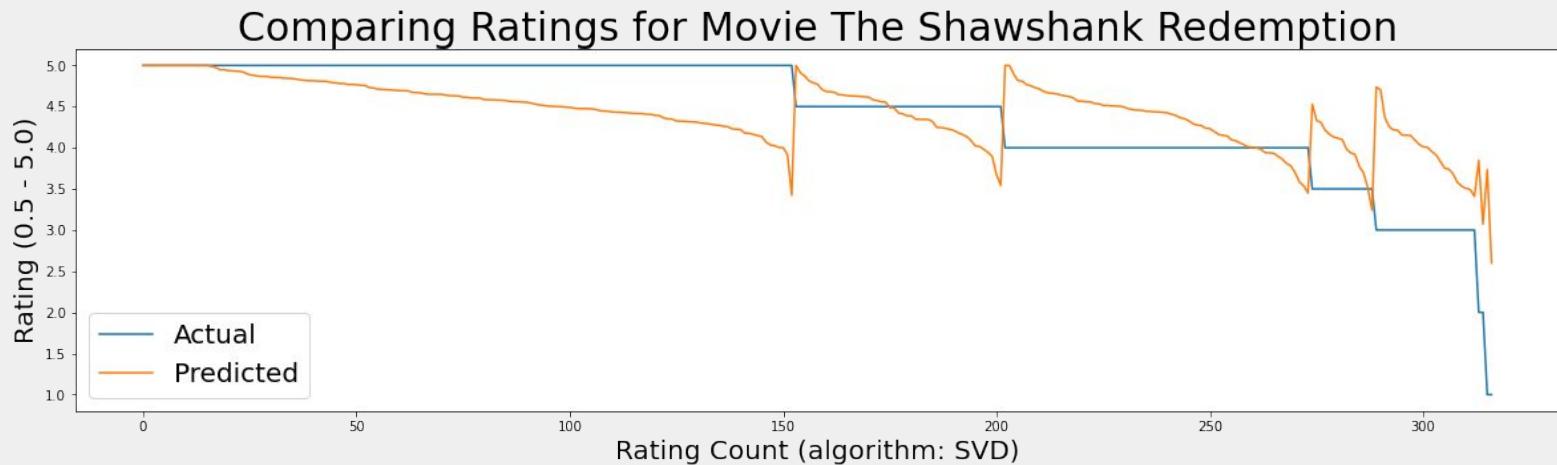
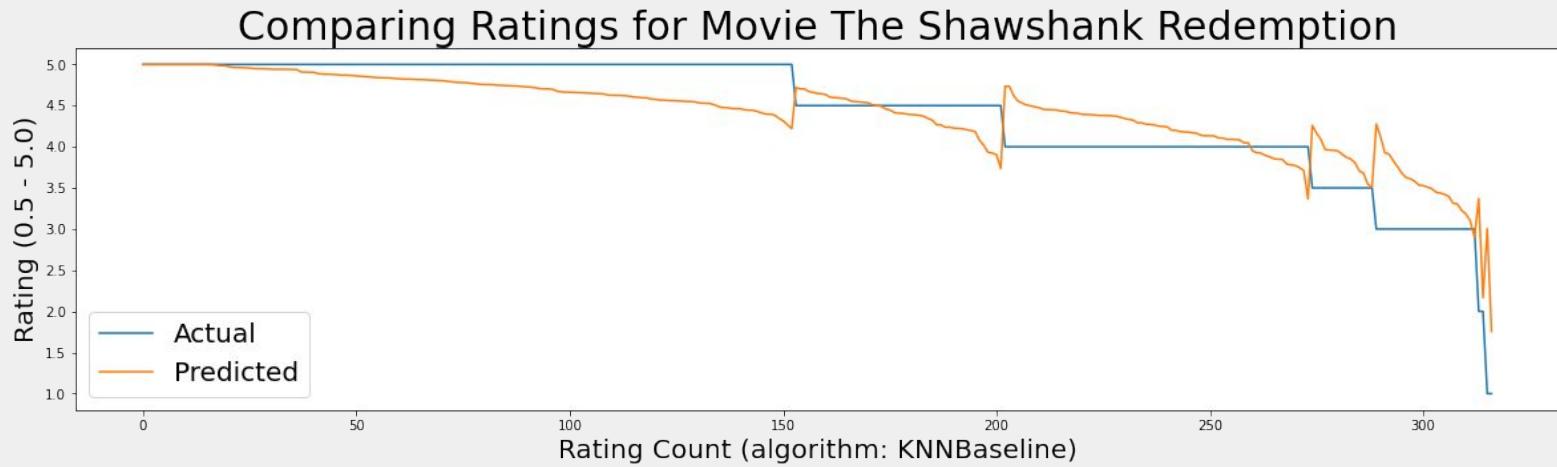
How many movies received this number of ratings?



How many movies did the users rate?

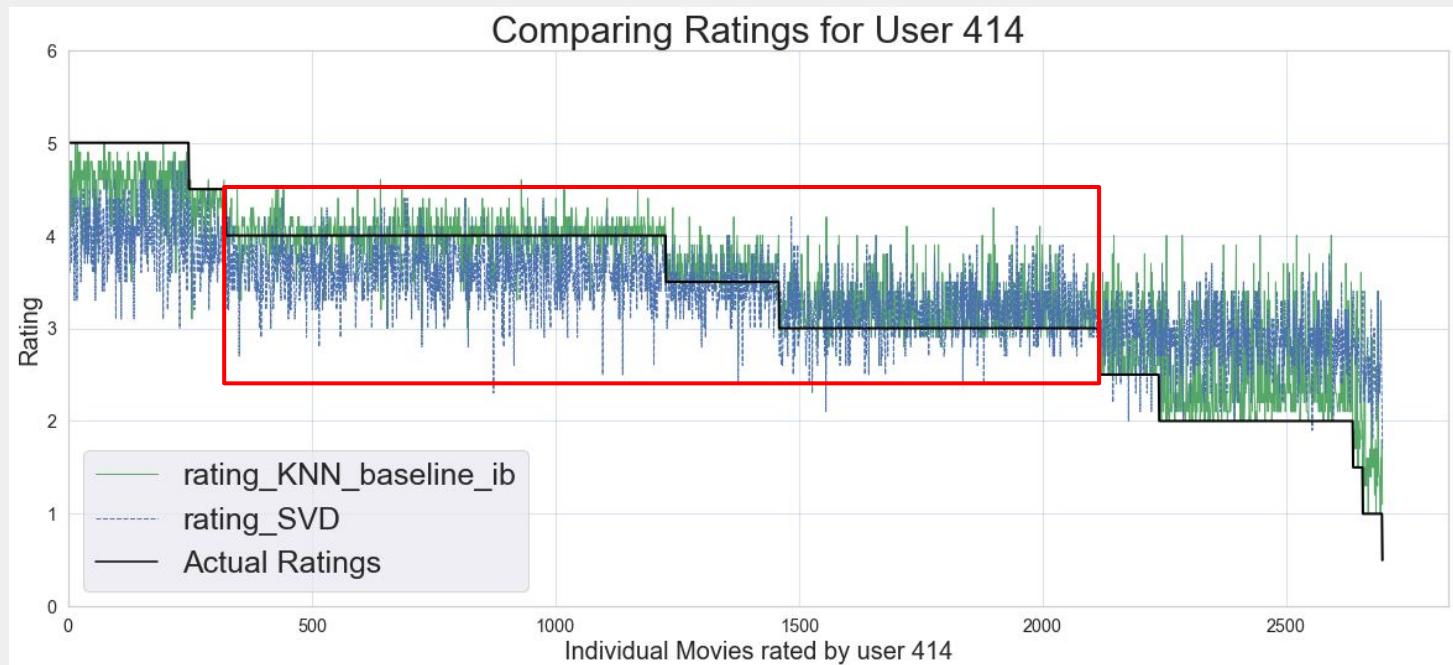


Comparison of two algorithms regarding one popular movie



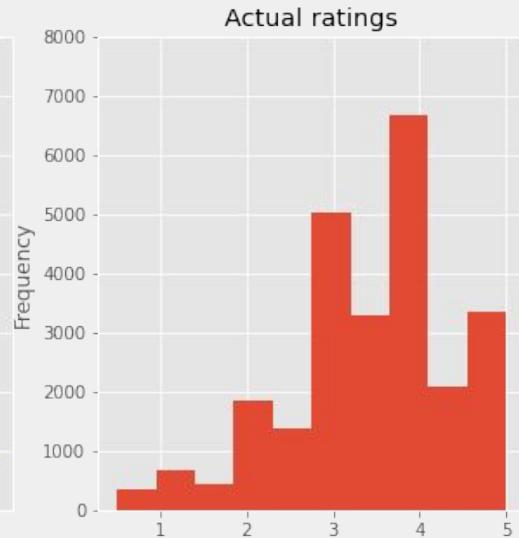
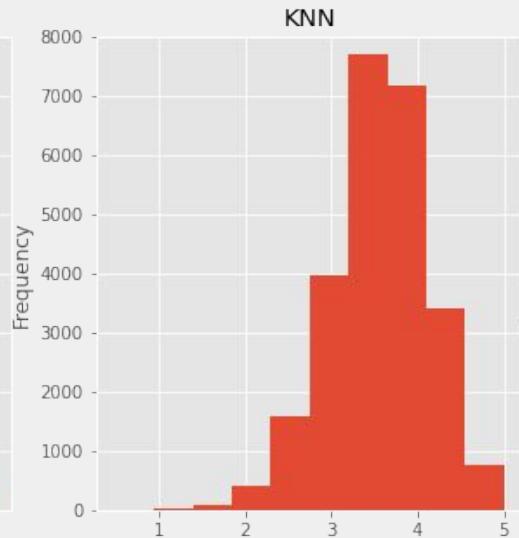
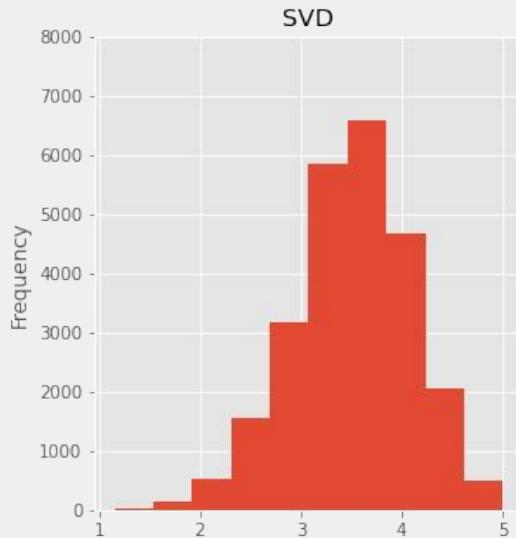
Comparison of two algorithms regarding a heavy rater

The most frequent rater



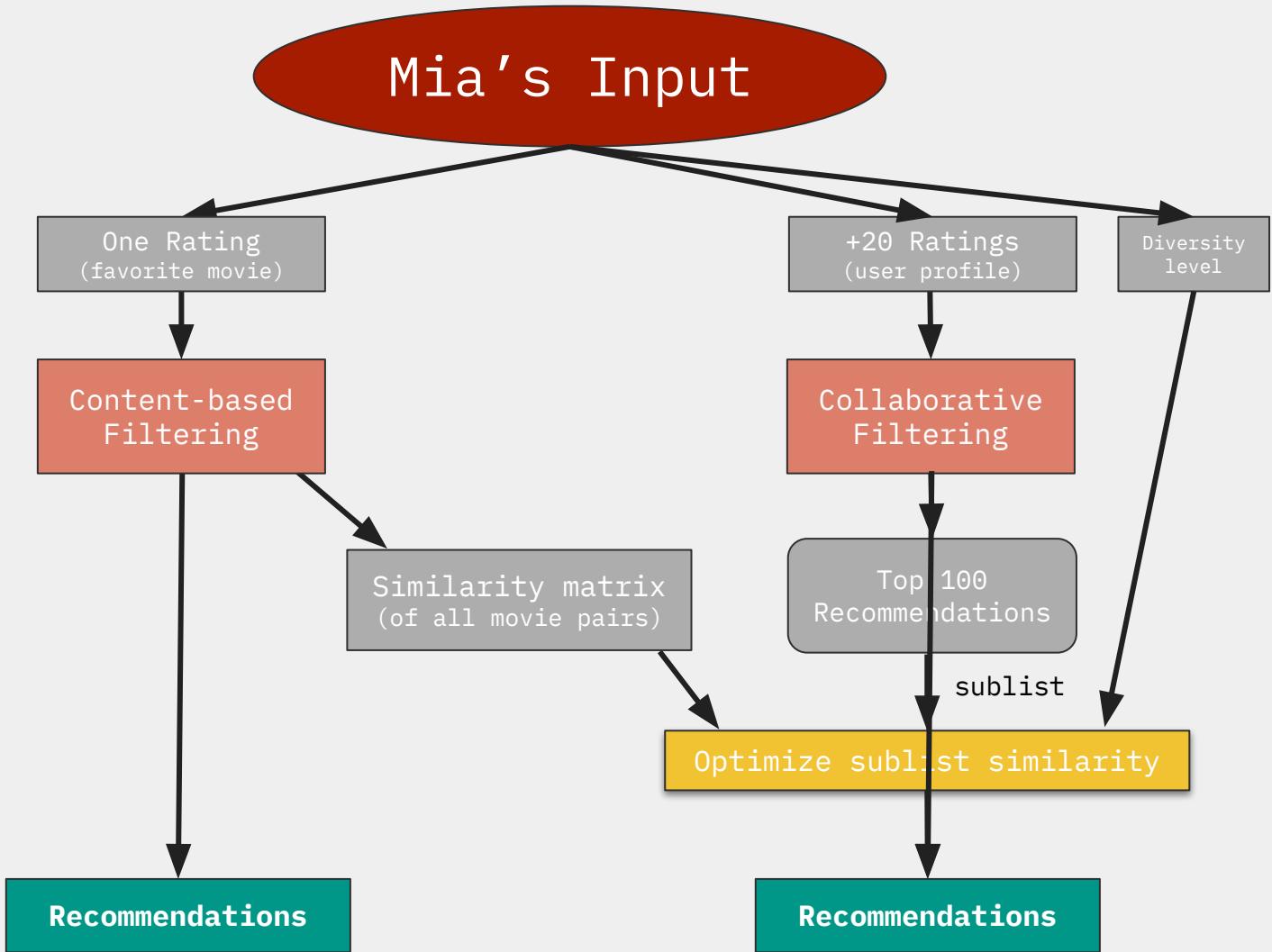
Comparison of the distribution of the predicted and the actual ratings

The algorithms tend to “normalize” the ratings

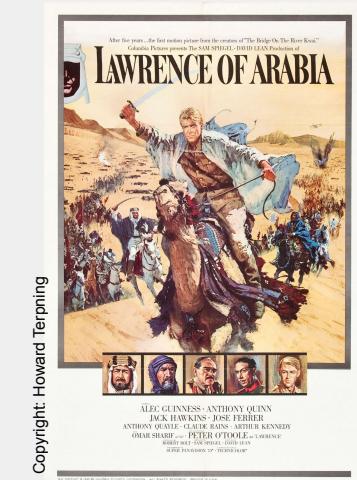
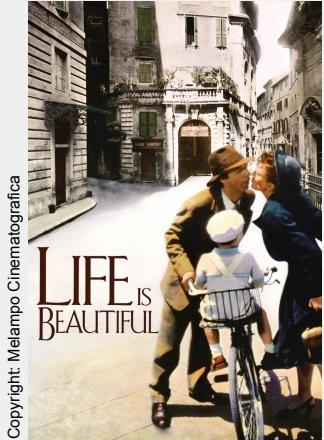


SVD (Singular Value Decomposition) algorithm was chosen for generating the Top 100 Recommendation list

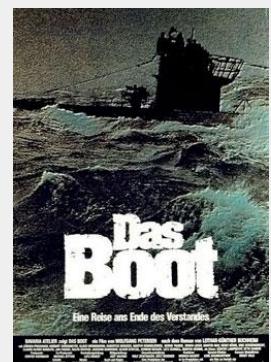
Hybrid Recommender Engine



Hybrid Recommender Engine: Results

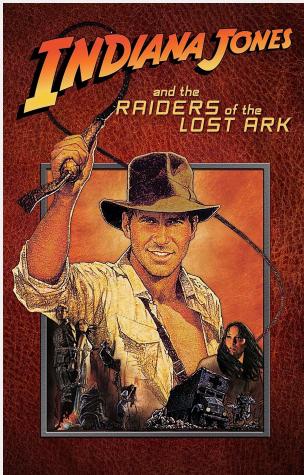
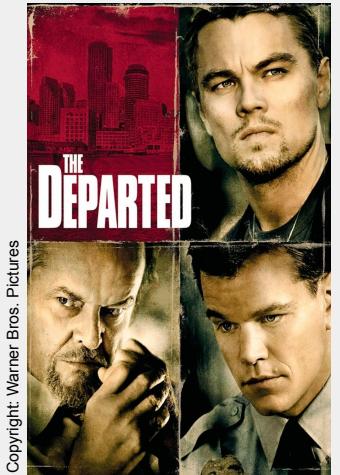
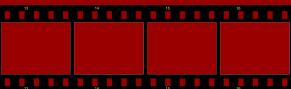


A recommendation list with
high intra-list similarity

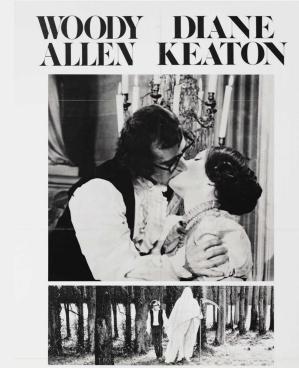


The six movies deal with wars / armies -> **low diversity**

Hybrid Recommender Engine: Results

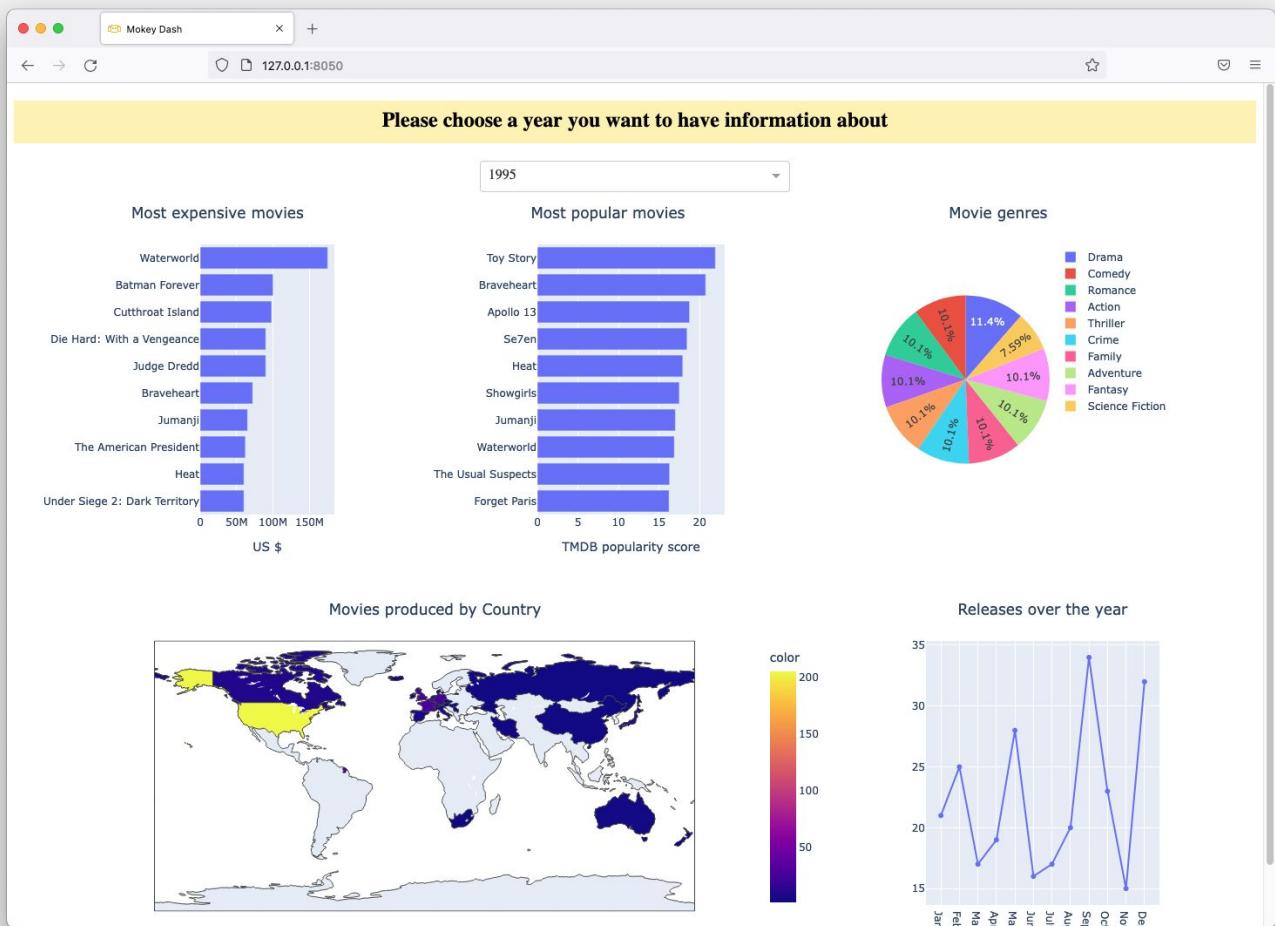


A recommendation list with
low intra-list similarity

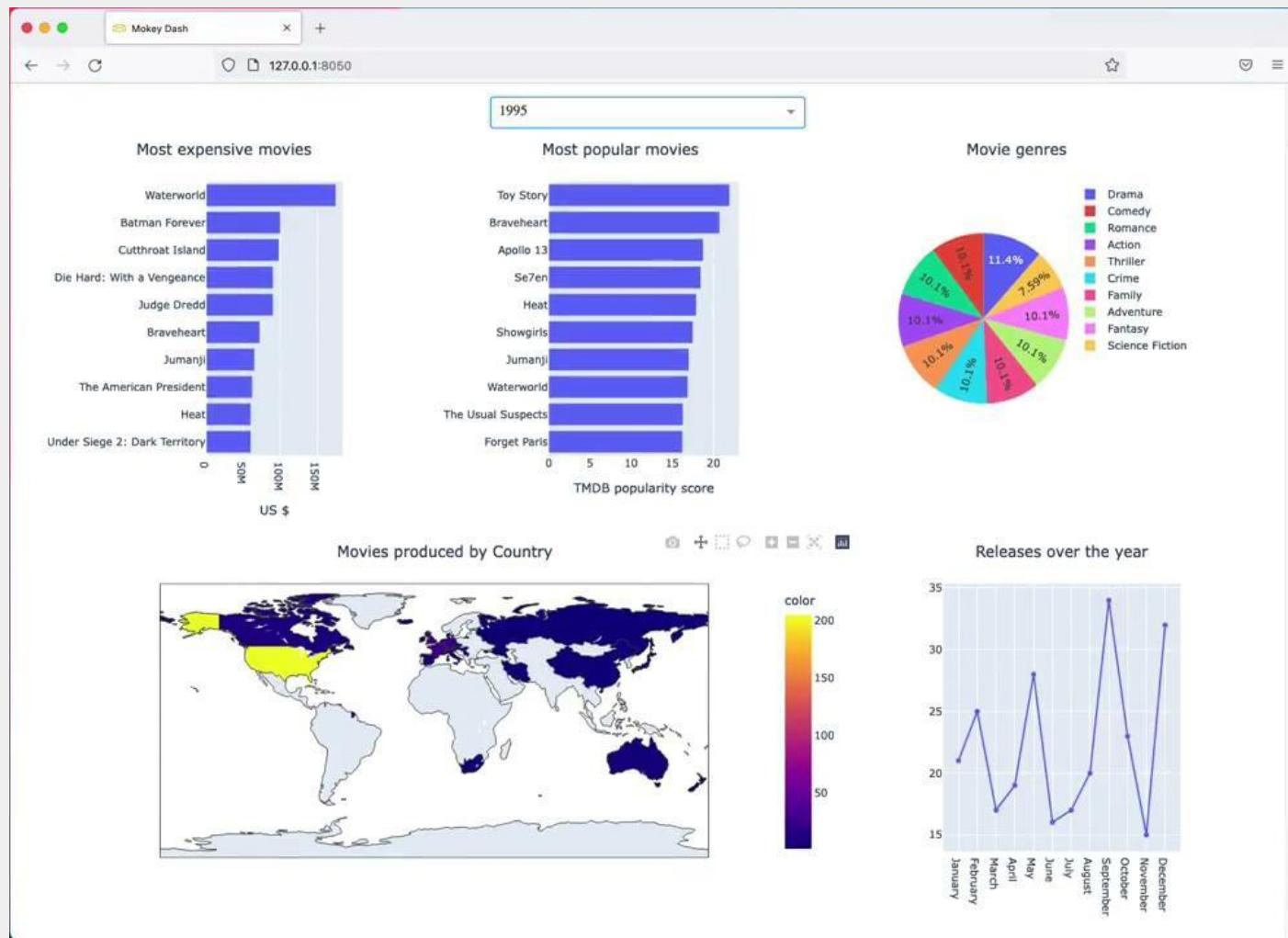


Comedy/history/action movies from the 1970s, 1980s, 1990s, 2000s -> high diversity

Dashboard



Dashboard



Outlook

- Online A/B testing
- Specify user groups
- Time series analysis
- More EDA



Balmy Bouffle

*The
End*