

Movie Recommender

Understanding User Preferences with Data

Kai Cansler

Outline

01

**Business
Problem**

02

**Data
Understanding**

03

Visualizations

04

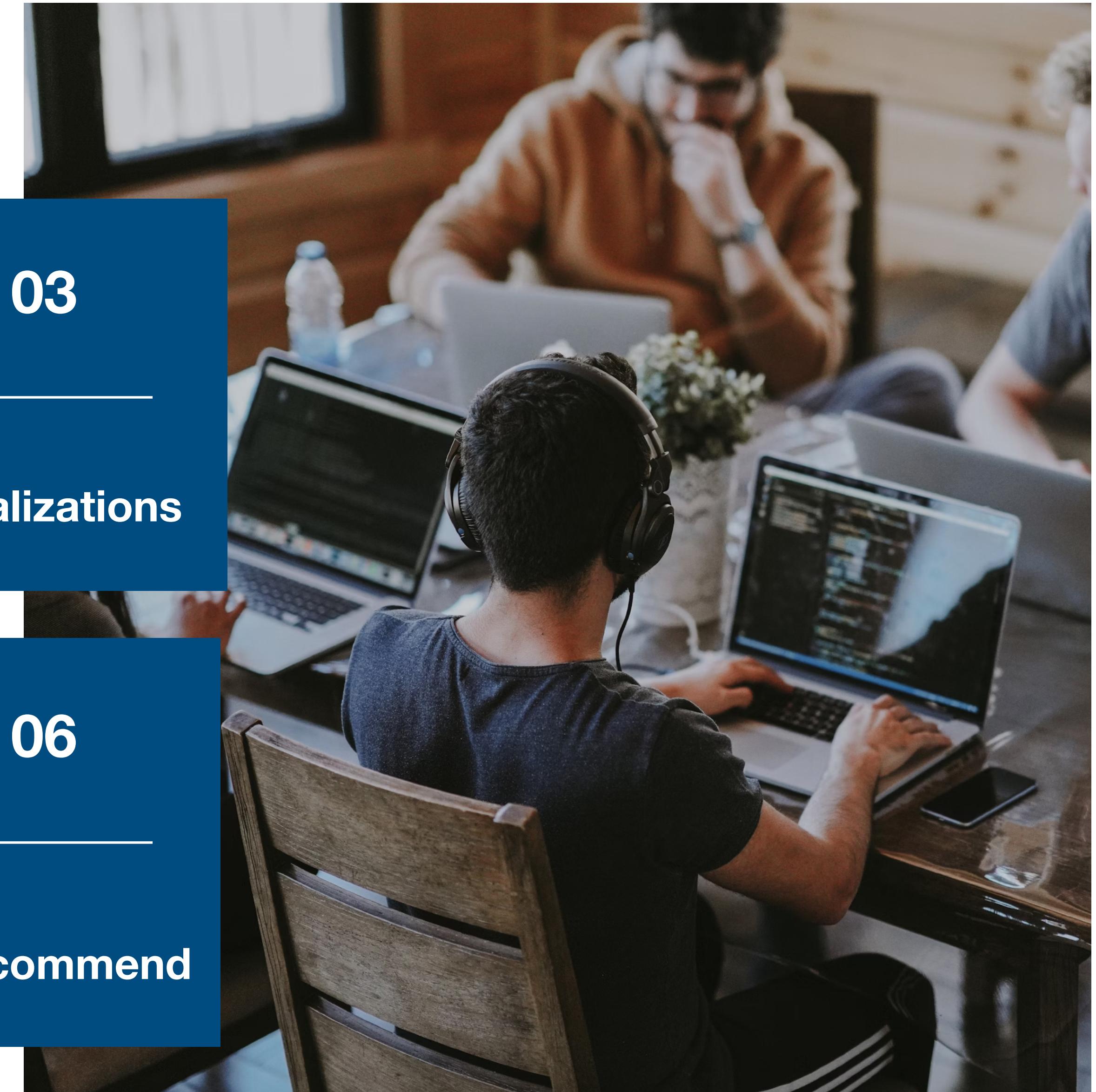
Modeling

05

Conclusion

06

Recommend



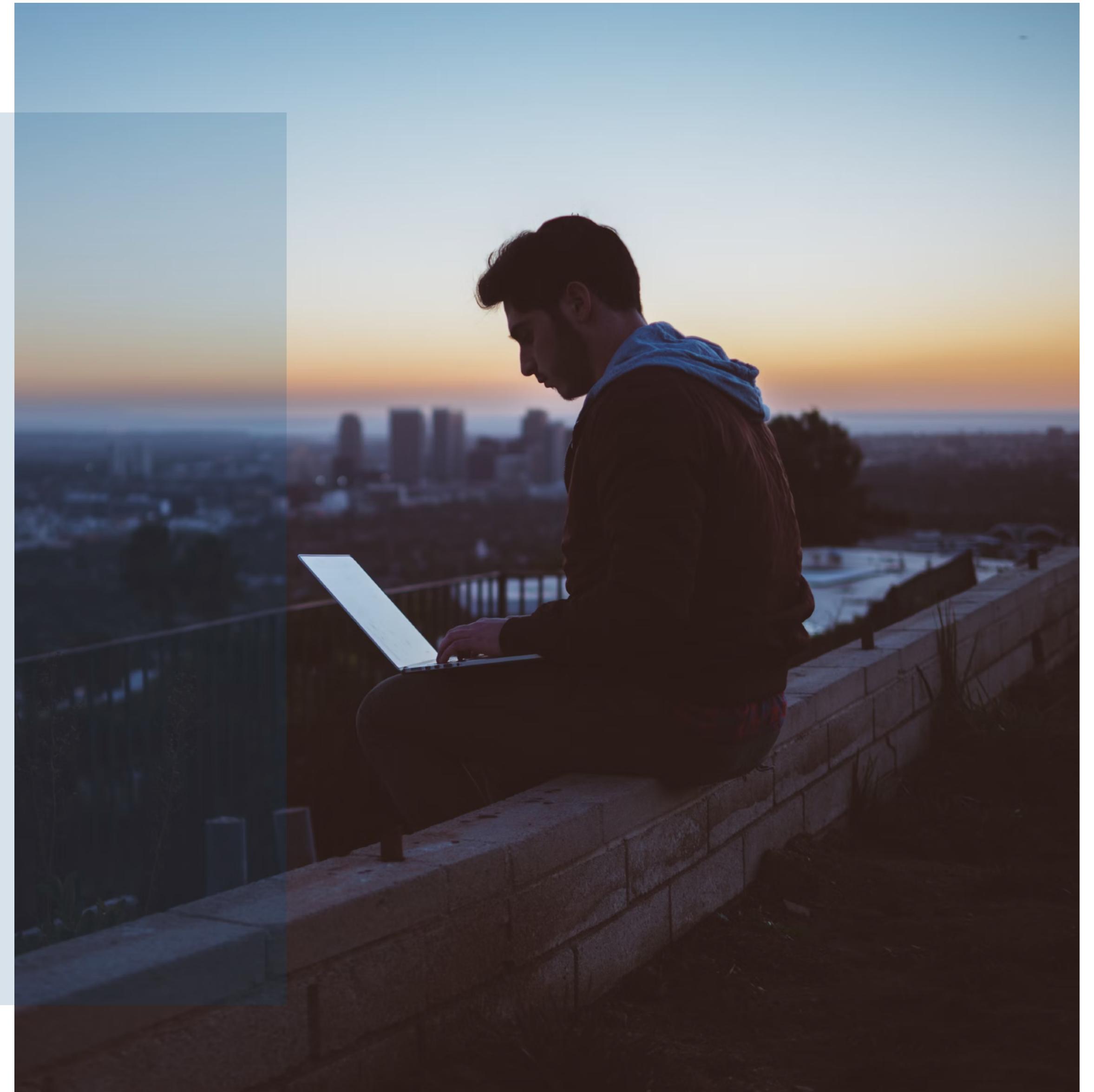
Business Problem

- Customer Satisfaction
-

- Historical Ratings
-

- Personal and Creative
-

- Relevant



Data Understanding

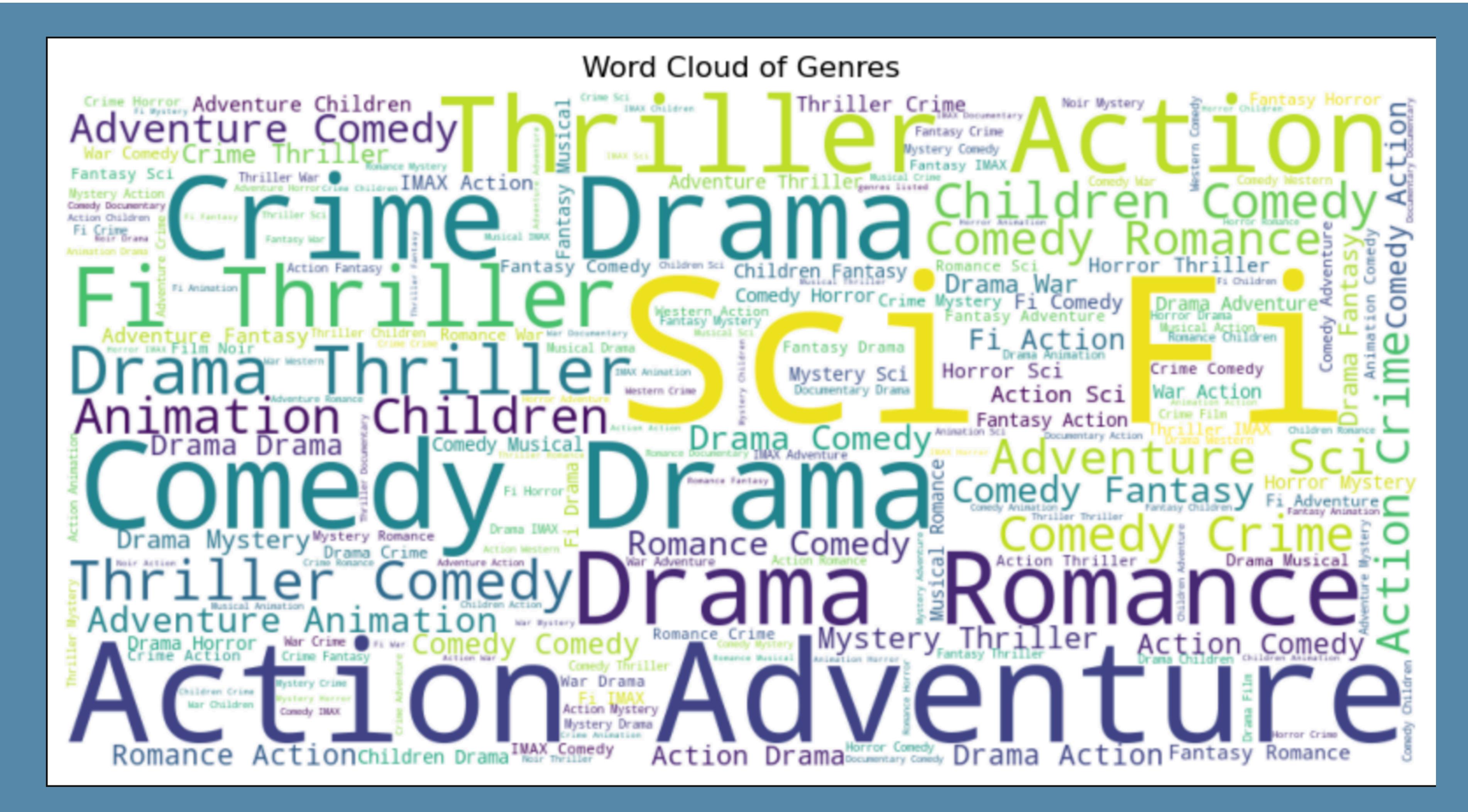
- ml-latest-small dataset
- 610 Users from 1996 - 2018
- 100,836 Ratings
- 9,742 Movies



Word Cloud

Genre Volume

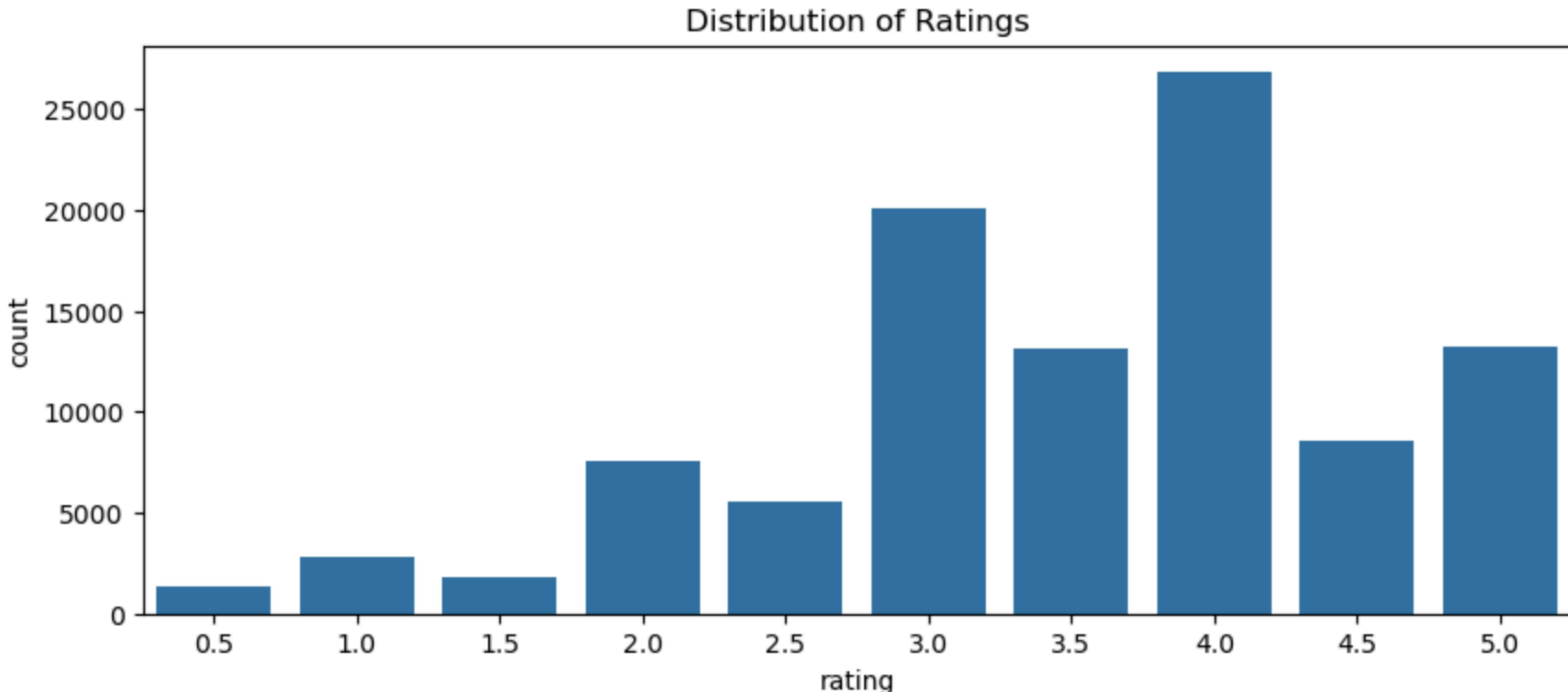
Sci-Fi and Action Adventure are most Popular



Distribution of Ratings

How do Users Rate Movies?

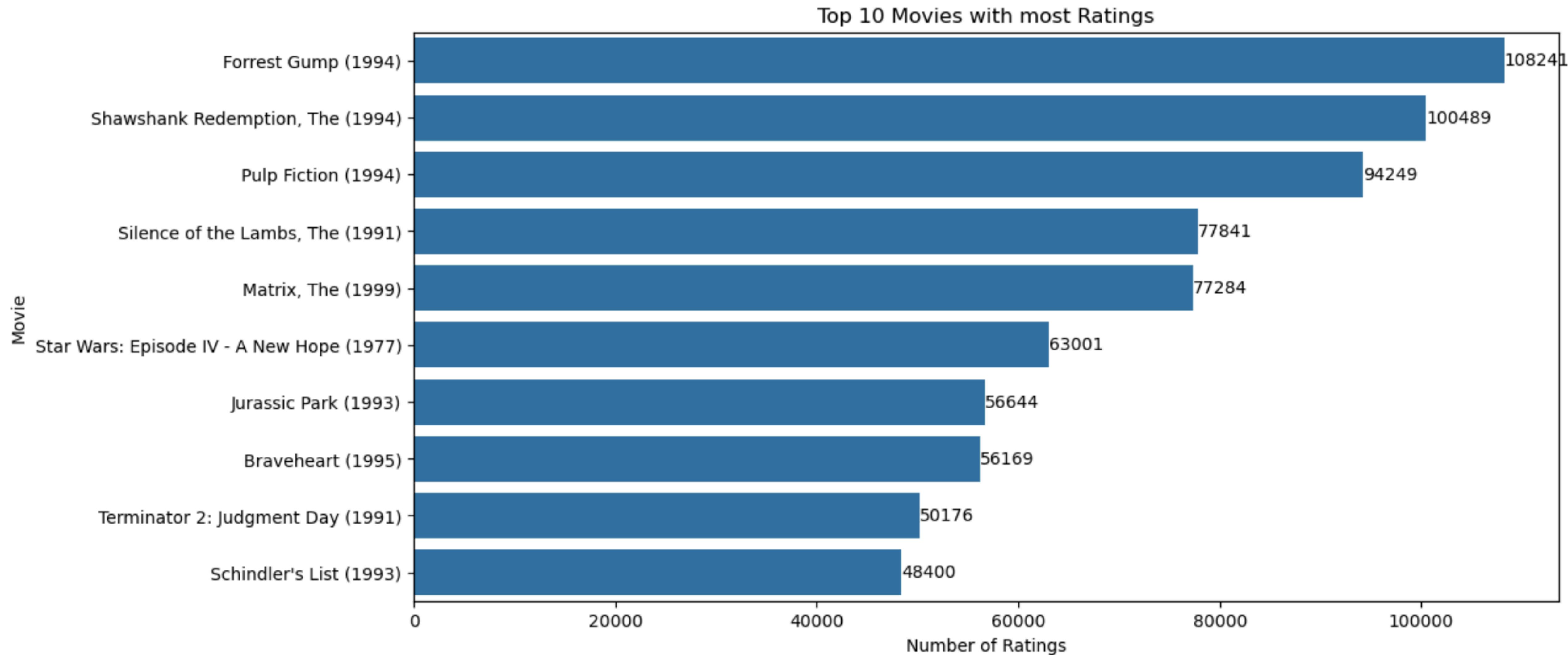
- Greater than 3-stars on Average
- Low volume of 0.5-stars



Most Ratings

Top 10 Movies

- Forrest Gump with Most Ratings
- 4 Sci-Fi Movies out of 10



Modeling

Collaborative Filtering

- Ratings of Similar Users

Root Mean Square Error

- Average Distance of Predictions



Recommendations

K-Nearest Neighbors

- RMSE is ~0.87
- Strength:
Item-Item Similarity,
Fast Processing
- Weakness:
Scalability and
Cold Start Problem

	title	genres	knn_distance
	Spider-Man 2 (2004)	Action Adventure Sci-Fi IMAX	0.267842
	Star Wars: Episode II - Attack of the Clones (...)	Action Adventure Sci-Fi IMAX	0.303643
	Minority Report (2002)	Action Crime Mystery Sci-Fi Thriller	0.338738
	X2: X-Men United (2003)	Action Adventure Sci-Fi Thriller	0.361211
	X-Men (2000)	Action Adventure Sci-Fi	0.370047
	Pirates of the Caribbean: The Curse of the Bla...	Action Adventure Comedy Fantasy	0.375362
	Lord of the Rings: The Fellowship of the Ring,...	Adventure Fantasy	0.379254
	Matrix Reloaded, The (2003)	Action Adventure Sci-Fi Thriller IMAX	0.387891
	Lord of the Rings: The Two Towers, The (2002)	Adventure Fantasy	0.388087
	Bourne Identity, The (2002)	Action Mystery Thriller	0.389831

Recommendations

Singular Value Decomposition

- RMSE is ~0.88
- Strength:
Scalability and
Dimensionality
- Weakness:
Cold Start Problem

	title	rating
0	Streetcar Named Desire, A (1951)	4.768913
1	Touch of Evil (1958)	4.729551
2	Grand Day Out with Wallace and Gromit, A (1989)	4.725915
3	Hoop Dreams (1994)	4.706459
4	His Girl Friday (1940)	4.696811
5	Rear Window (1954)	4.675742
6	Shawshank Redemption, The (1994)	4.673842
7	Louis C.K.: Live at the Beacon Theater (2011)	4.669133
8	Seventh Seal, The (Sjunde inseglet, Det) (1957)	4.661953
9	Guess Who's Coming to Dinner (1967)	4.659831

Next Steps

1. Collect More Data - User Meta Data
2. Use Single Value Decomposition for Scalability
3. Hybrid Recommendation System

Thank you.

Kai Cansler



Email:
cansler.kai@gmail.com



GitHub:
<https://github.com/kai-cansler>



Work Cited:

<https://unsplash.com/photos/group-of-people-using-laptop-computer-QckxruozjRg>

<https://unsplash.com/photos/man-in-black-jacket-standing-on-blue-lighted-room-KJ241ZAOYwU>

<https://unsplash.com/photos/man-sitting-on-concrete-brick-with-opened-laptop-on-his-lap-Z3ownETsdNQ>

<https://unsplash.com/photos/woman-sitting-in-front-of-laptop-LVooQvKjLjw>