

MOVIE RECOMMENDATIONS

for online streaming service **TUBI**



TOO MANY CHOICES





free ad-supported streaming service

- Increased User Engagement
- Enhanced User Experience
- Higher Ad Revenue



- ML model
- Top 5 movie recommended to a user

DATA USED

Dated March 29, 1996 - September 24, 2018

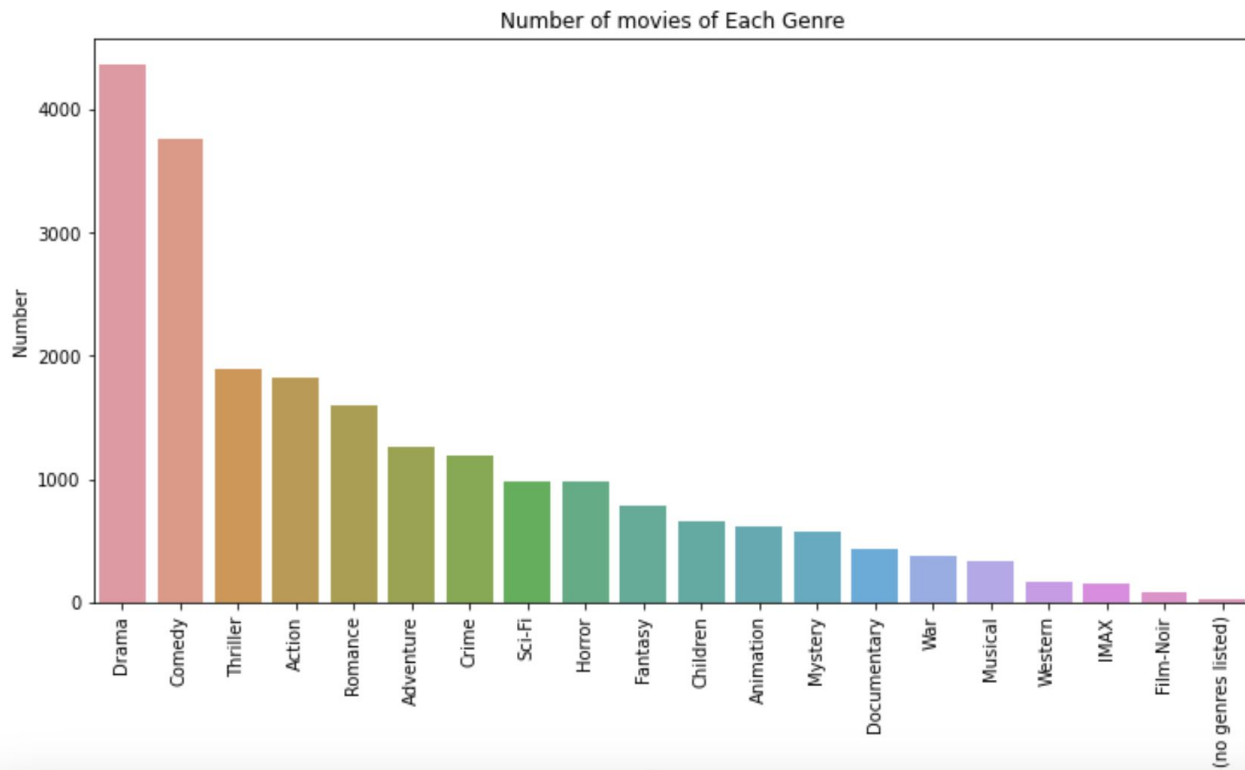
MovieLens Dataset Details

Dataset	Shape	Columns
Ratings	100,836 rows × 4 columns	userid, movieid, rating, timestamp
Movies	9,742 rows × 3 columns	movieid, title, genre
Tags	3,683 rows × 4 columns	userId, movieid, tag, timestamp
Links	9,742 rows × 3 columns	movieid, imdbid, tmdbid

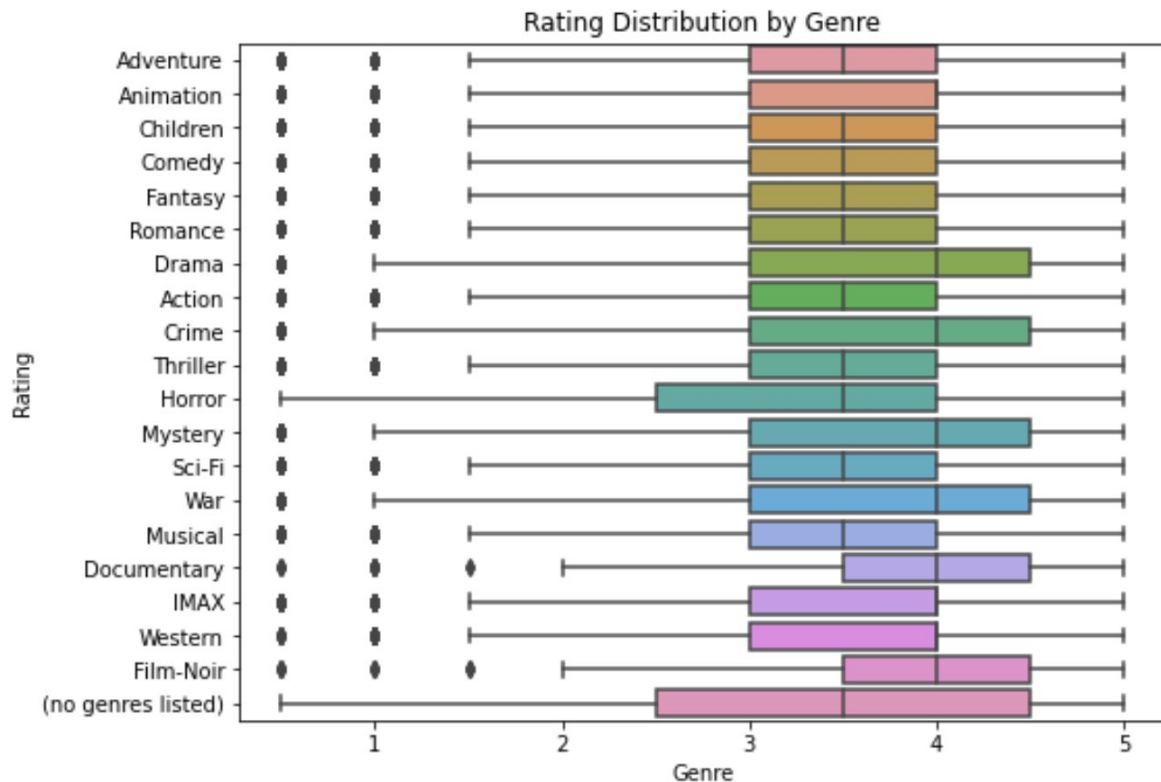
EDA Tools: Python Pandas, Numpy

Visualization Tools: Seaborn and Matplotlib

MOVIES DATA SET: 20 genres, 9742 movies



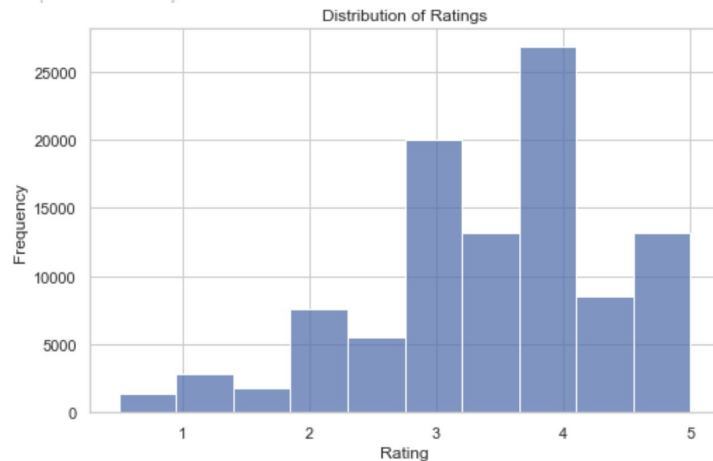
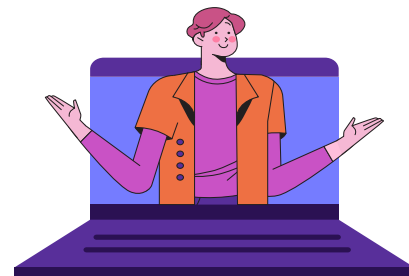
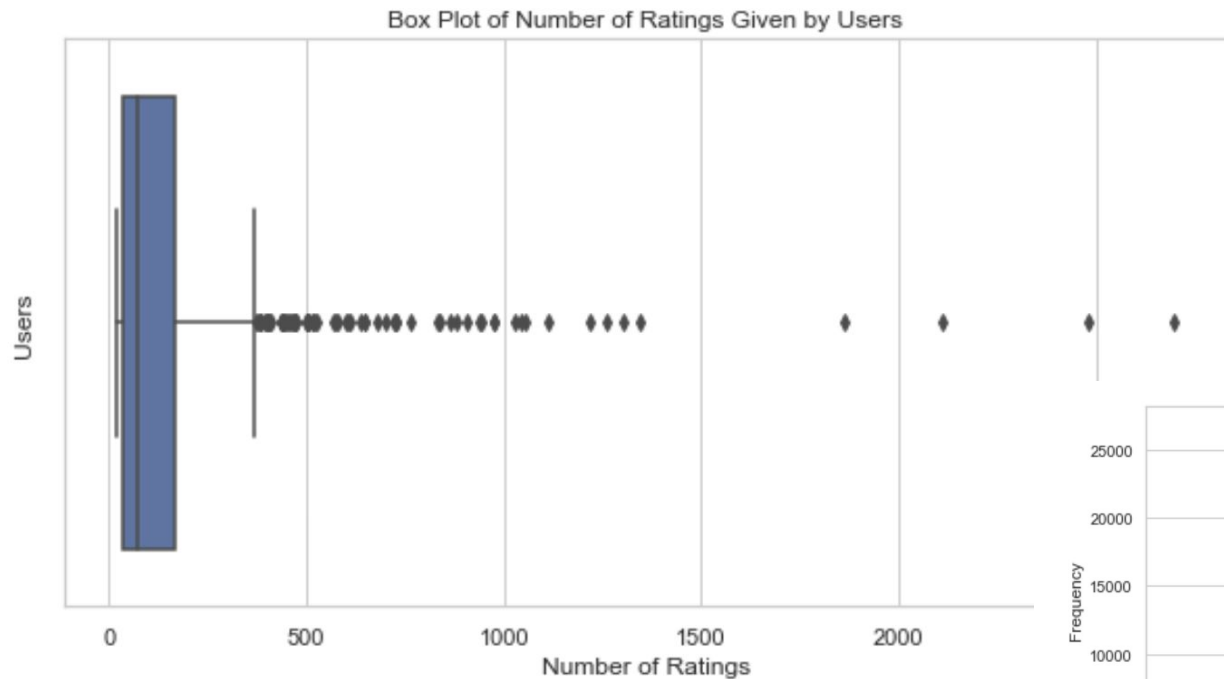
RATINGS DATA SET:



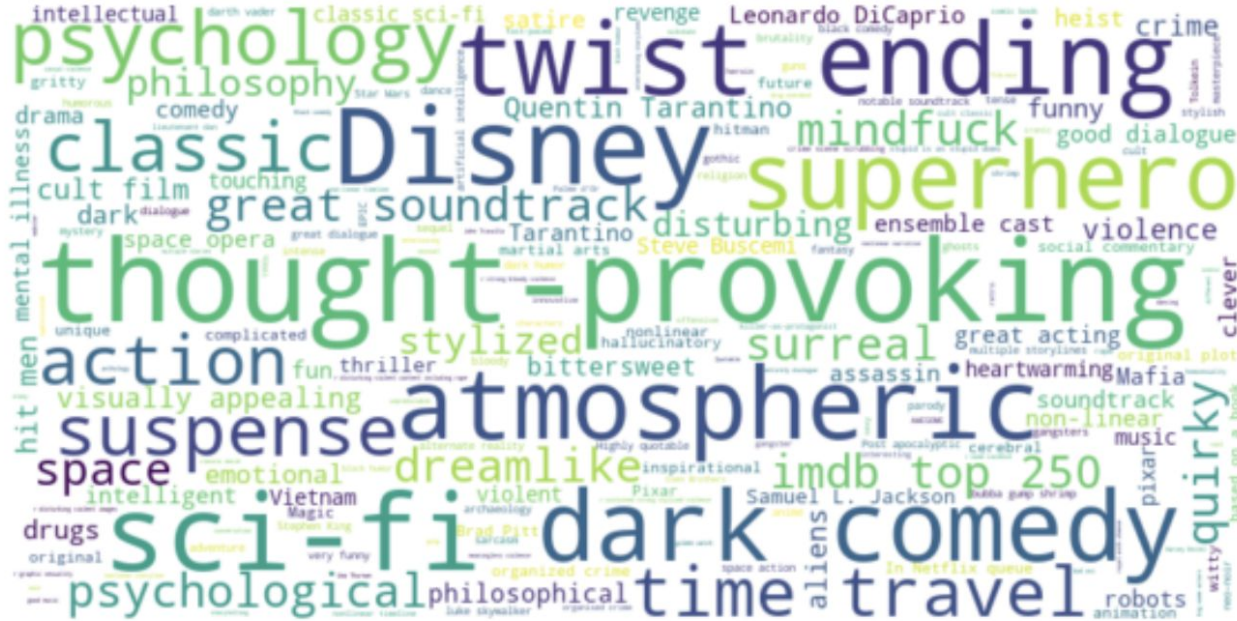
**FOR ALL GENRES
COMBINED:**

mean	3.560346
std	0.161471
min	3.258195
25%	3.453882
50%	3.536144
75%	3.638387
max	3.920115

Distribution of 100836 ratings by 610 users



TAGS DATA SET: 3683 ROWS



**1584 UNIQUE
TAGS
FOR 1553
MOVIES
BY 58 USERS**

**USER #474 LEFT
1507 TAGS**

MODELING

COLLABORATIVE FILTERING

Baseline - KNN Basic
KNN Baseline
Matrix Factorization

CONTENT FILTERING

NearestNeighbours

EVALUATION METRICS

Root Mean Square Error, Mean Absolute Error

EVALUATIONS RESULTS

	name	RMSE	MAE
0	KNN Basic	0.85	0.65
1	KNN Baseline	0.75	0.56
2	KNN Baseline Tuned	0.60	0.45
3	SVD	0.37	0.22
4	ALS	0.68	0.51

01

BEST MODEL: SINGULAR VALUE DECOMPOSITION

RMSE: 0.37

MAE: 0.22

Params:

'n_factors': 100

'n_epochs': 40

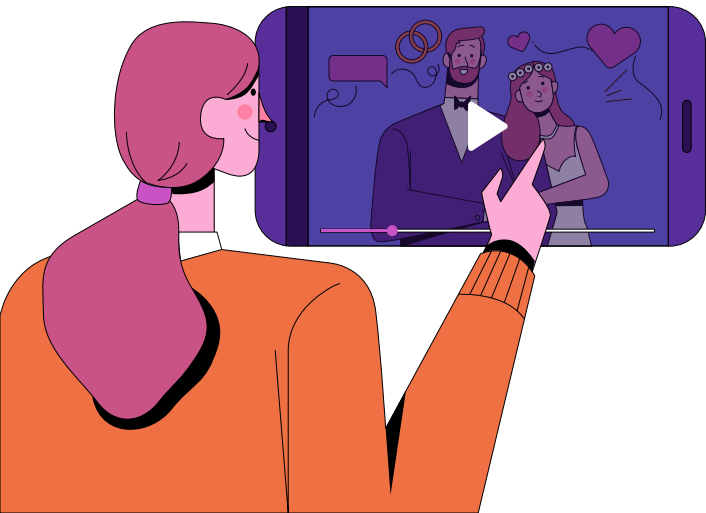
'lr_all': 0.01

'reg_all': 0.02



CONTENT BASED FILTERING

- **content = title + genres + tag**
- **TF-IDF matrix**
- **NearestNeighbors**
- **Cosine similarity**



Titanic

	title	genres	year	
984	Ghosts of the Abyss	Documentary	2003	1206
901	Z	Drama	1969	1191
1341	Illusionist, The	Drama	2006	1484
1317	North Country	Drama	2005	1400
949	May	Drama	2002	1463
667	Magnolia	Drama	1999	831
71	My Family	Drama	1995	1451
715	My Life	Drama	1993	1417
1051	My Life Without Me	Drama	2003	988
1366	There Will Be Blood	Drama	2007	142

```
titanic_rows['tag'].value_counts()
```

```
romance      280
shipwreck    280
```

Shrek

	title	genres	year
	Into the Woods	Adventure	1991
	Shrek 2	Adventure	2004
	Wild Tales	Comedy	2014
	Up	Adventure	2009
	Hobbit: The Desolation of Smaug, The	Adventure	2013
	Bill & Ted's Excellent Adventure	Adventure	1989
	Croods, The	Adventure	2013
	Sintel	Animation	2010
	Holes	Adventure	2003
	Aladdin	Adventure	1992

```
shrek_rows['tag'].value_counts()
```

```
fairy tales    1020
```



RECOMMENDATIONS

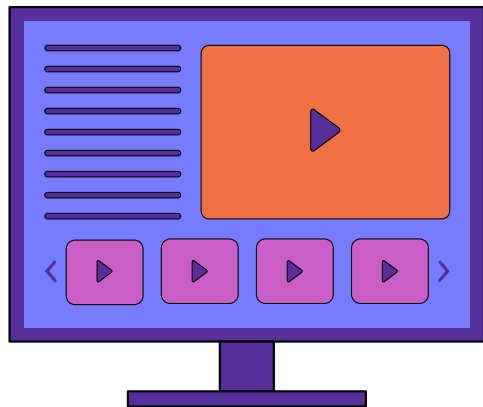


**Introduce recommendation system based on
SVD Matrix Factorisation with tuned
hyperparameters**

Introduce content based filtering



NEXT STEPS



Collect data and evaluate changes in:

User engagement, viewing time, user retention, ads revenue

Introduce users demographic data

Collect user feedback to provided recommendations

Develop hybrid model

THANKS!

github.com/dolgorp
linkedin.com/purbuyeva

