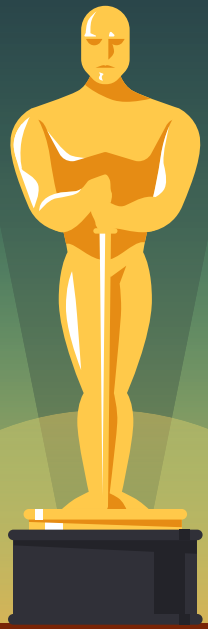
RECOMMENDER ENGINE OPTIMIZATION



STEVEN L TRUONG

Friday, 04/30/2021

CELEBRATING 10 YEARS OF

FANDOR

STREAMING INDEPENDENT CINEMA

CELEBRATING 10 YEARS OF

FANDOR

STREAMING INDEPENDENT CINEMA

Become an independently original program

01

02

- Become an independently original program
- Optimize the recommender engine

CELEBRATING 10 YEARS OF

FANDOR

STREAMING INDEPENDENT CINEMA

03

- Become an independently original program
- Optimize the recommender engine
- Reduce consumers churning rate

WHY?

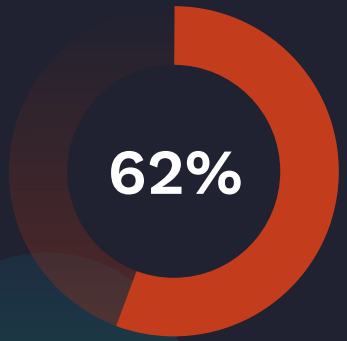
WHY DO WE NEED TO IMPROVE THE RECOMMENDER ENGINE?

Motivation

According to Deloitte, among customers who cancel the services

Motivation

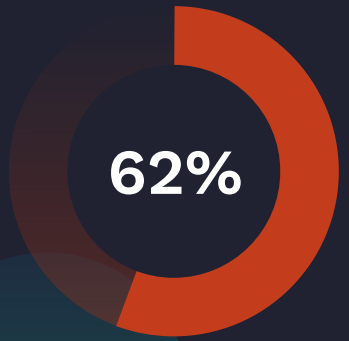
According to Deloitte, among customers who cancel the services



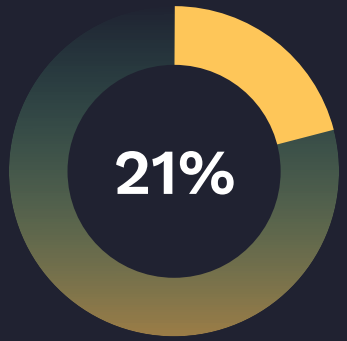
ONE SPECIFIC SHOW
THEN CANCEL

Motivation

According to Deloitte, among customers who cancel the services



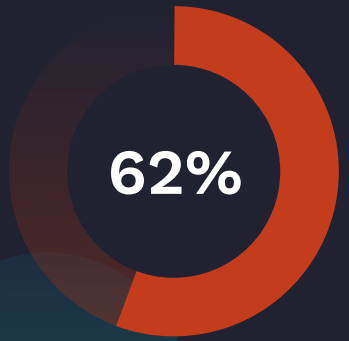
ONE SPECIFIC SHOW THEN CANCEL



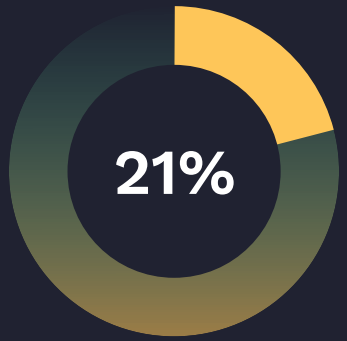
LACKS OF NEW CONTENT

Motivation

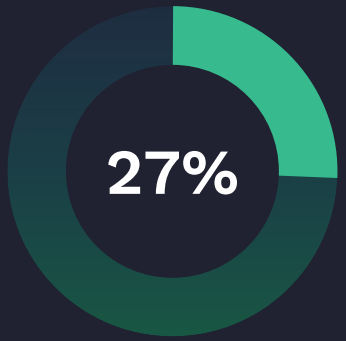
According to Deloitte, among customers who cancel the services



ONE SPECIFIC SHOW THEN CANCEL



LACKS OF NEW CONTENT



STAY IF MORE EXCLUSIVE CONTENT

Solution?

Hypothesis and solution path

Hypothesis and solution path

Pre-launch

Implement SOTA algorithms

01

Hypothesis and solution path

Pre-launch

Implement SOTA algorithms

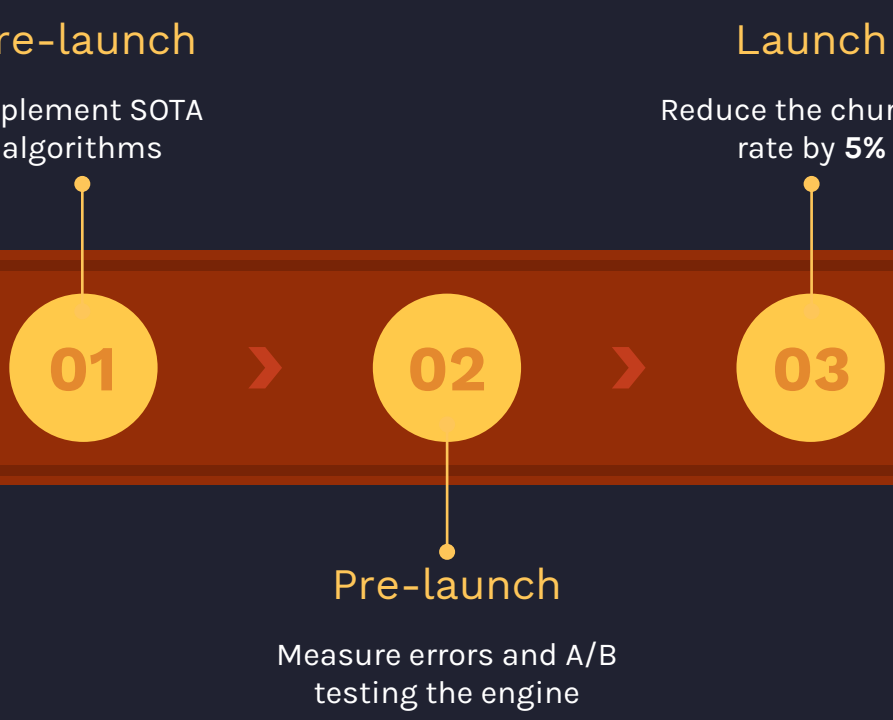
01

02

Pre-launch

Measure errors and A/B testing the engine

Hypothesis and solution path



Hypothesis and solution path

01

Pre-launch

Implement SOTA algorithms

02

Pre-launch

Measure errors and A/B testing the engine

03

Launch

Reduce the churning rate by **5%**

04

3-month

Reduce churning rate by **15%**

Data and tools

Data

Acquire from **MovieLens** website

EDA/Visualization

Google Sheet and Tableau

Language

Python, SQL, HTML, CSS

Packages

surprise, numpy, pandas, requests

API

The Movie Database API

Deployment

Flask, SQLAlchemy

HOW DO WE APPROACH THE PROBLEM?

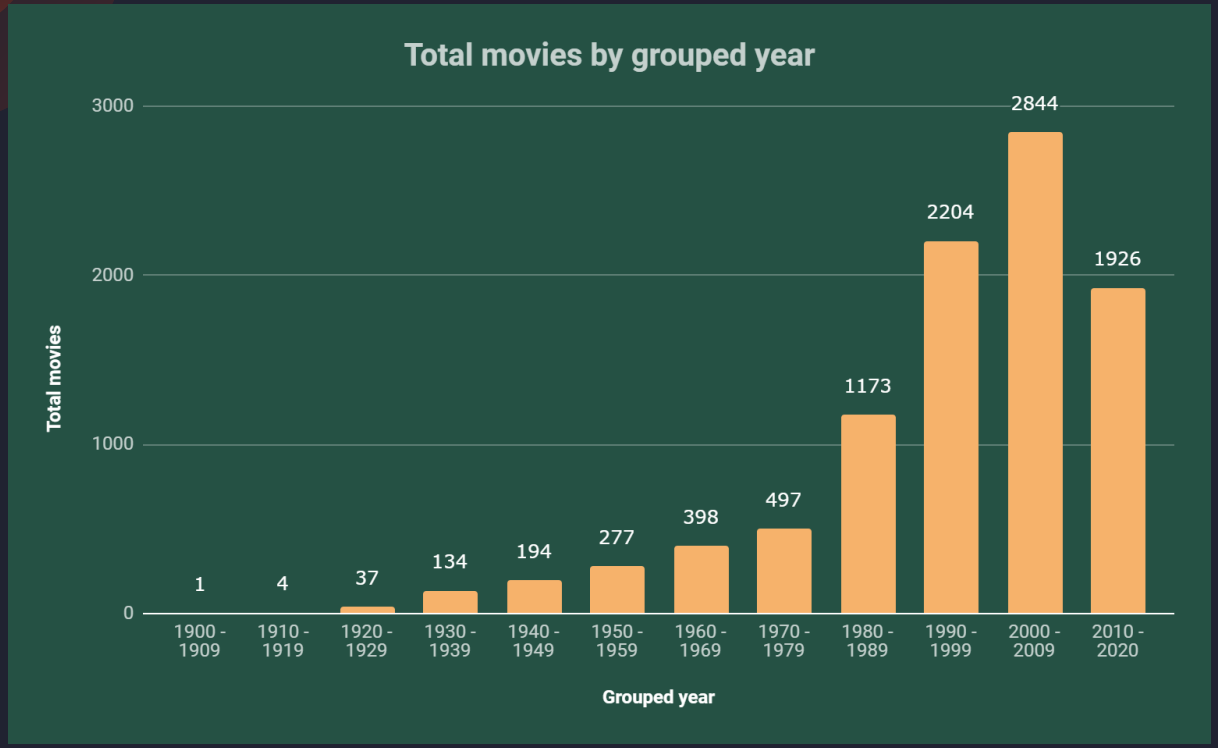
01

User-based collaborative filtering. Recommend similar movies based on user’s preference.

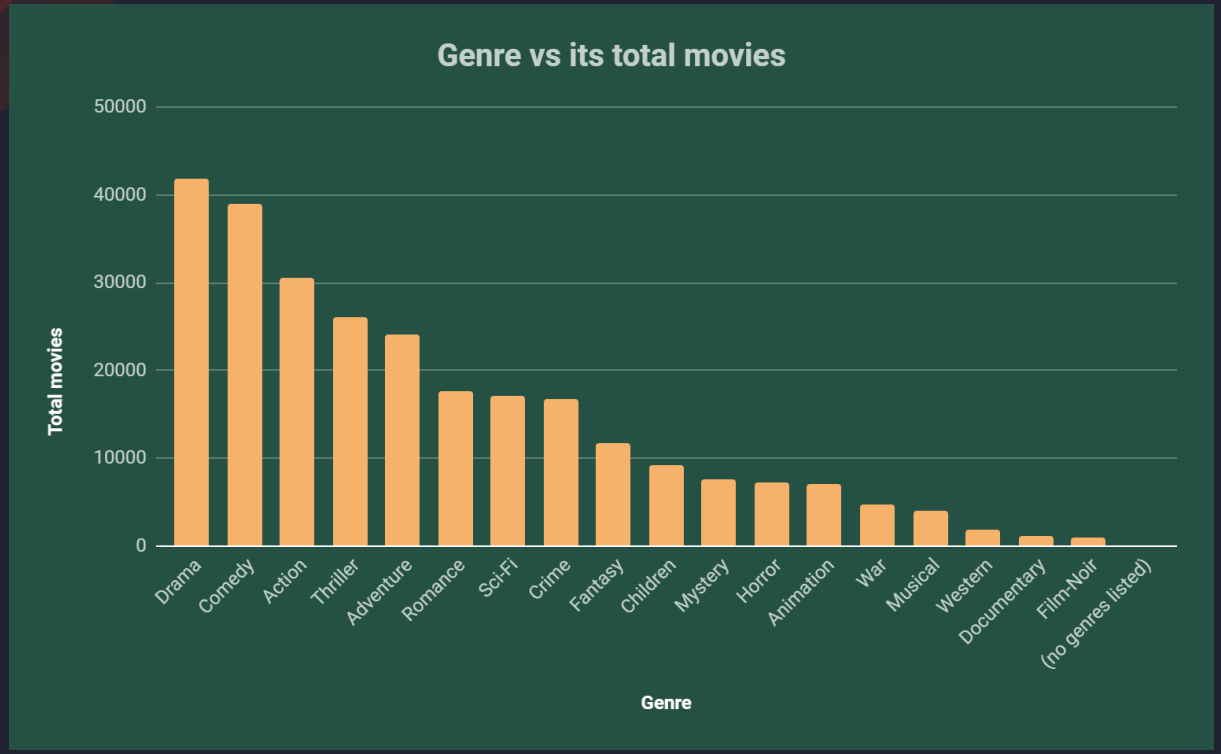
02

Item-based collaborative filtering. Recommend similar movies based on its characteristics.

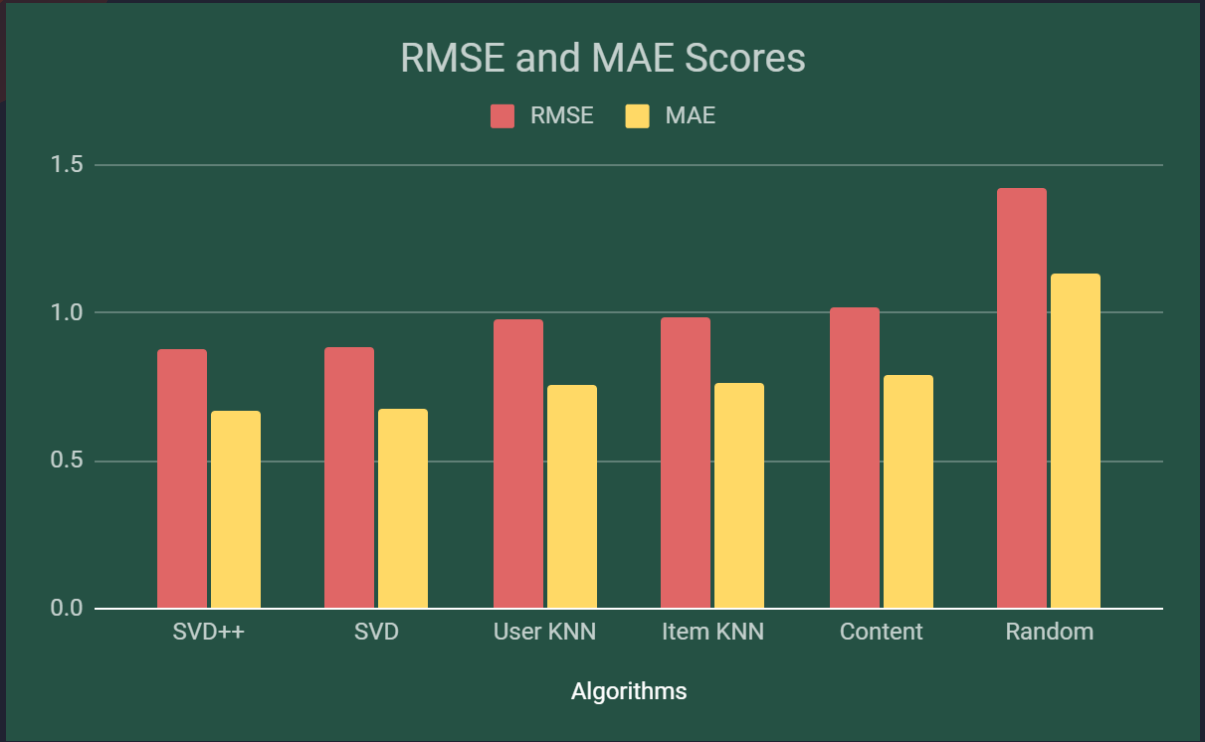
Results


Total movies by grouped year
3000
2000
1000
0
Total movies
1
4
37
134
194
277
398
497
1173
2204
2844
1926
1900 - 1909
1910 - 1919
1920 - 1929
1930 - 1939
1940 - 1949
1950 - 1959
1960 - 1969
1970 - 1979
1980 - 1989
1990 - 1999
2000 - 2009
2010 - 2020
Grouped year

Total movies by decade

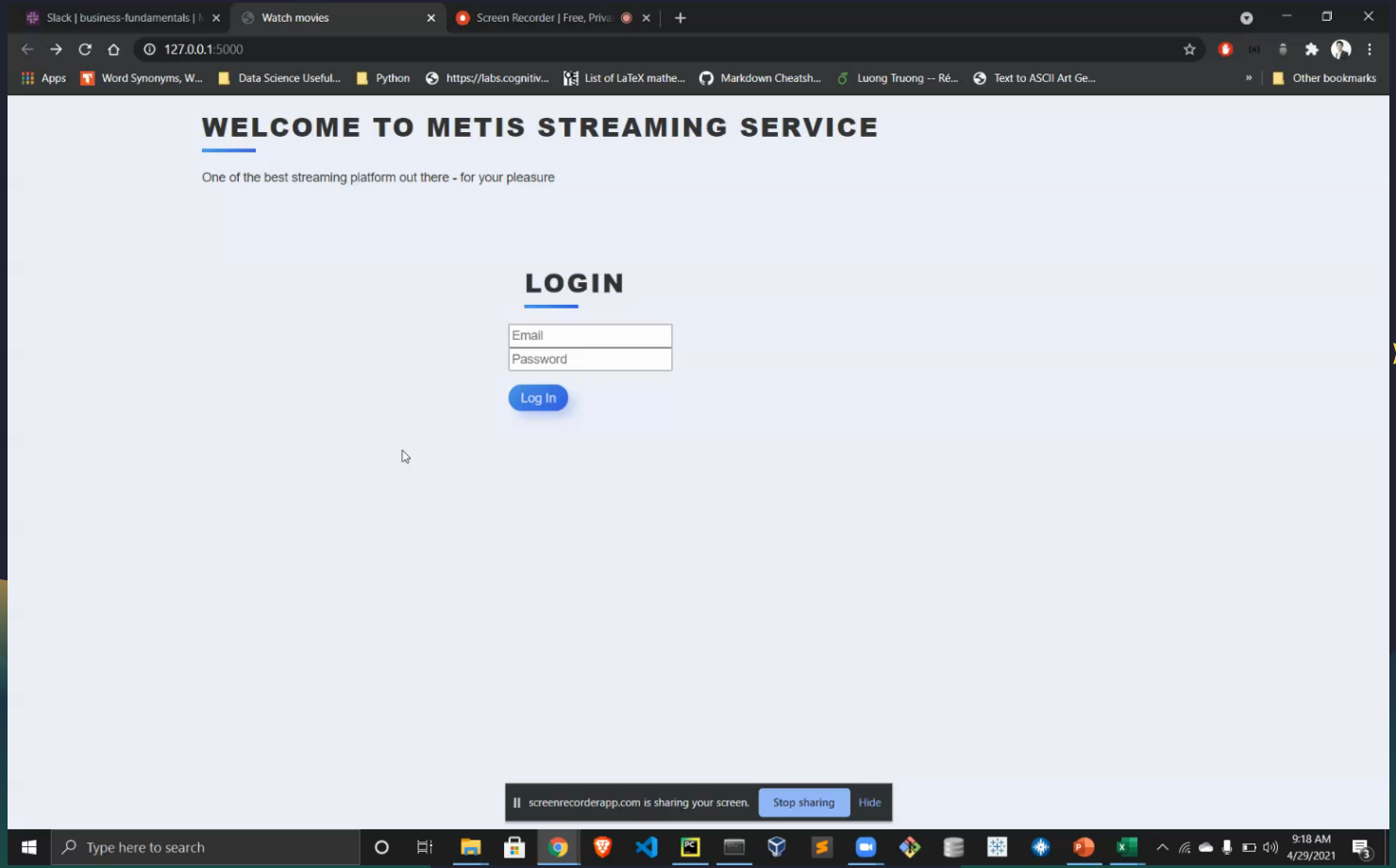

Genre vs its total movies
50000
40000
30000
20000
10000
0
Total movies
Drama
Comedy
Action
Thriller
Adventure
Romance
Sci-Fi
Crime
Fantasy
Children
Mystery
Horror
Animation
War
Musical
Western
Documentary
Film-Noir
(no genres listed)
Genre

Total movies
by genre


RMSE and MAE Scores
RMSE
MAE
1.5
1.0
0.5
0.0
SVD++
SVD
User KNN
Item KNN
Content
Random
Algorithms

The lower the score, the better the algorithm

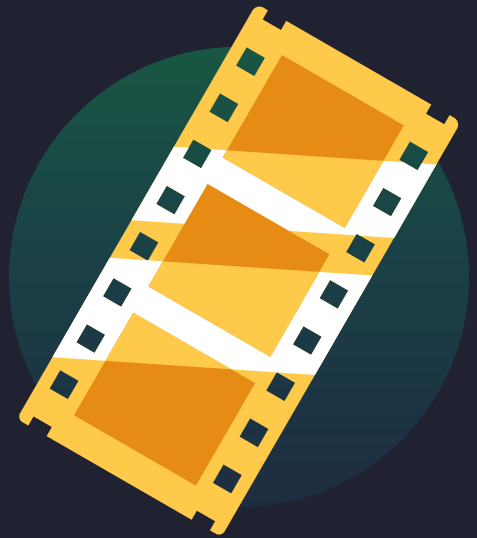
Demo


Slack | business-fundamentals |
Watch movies
Screen Recorder | Free, Priva
127.0.0.1:5000
Apps
Word Synonyms, W...
Data Science Useful...
Python
https://labs.cognitiv...
List of LaTeX mathe...
Markdown Cheatsh...
Luong Truong -- Ré...
Text to ASCII Art Ge...
Other bookmarks
WELCOME TO METIS STREAMING SERVICE
One of the best streaming platform out there - for your pleasure
LOGIN
Email
Password
Log In
screenrecorderapp.com is sharing your screen.
Stop sharing
Hide
Type here to search
9:18 AM
4/29/2021

Future Work

Algorithm

Restricted Boltzmann Machines (RBM's)
Deep Neural Networks (RNN)
Tensorflow Recommenders (TFRS)



Data

Scale it up using Apache Spark
Amazon Deep Scalable Sparse Tensor Network Engine (DSSTNE)

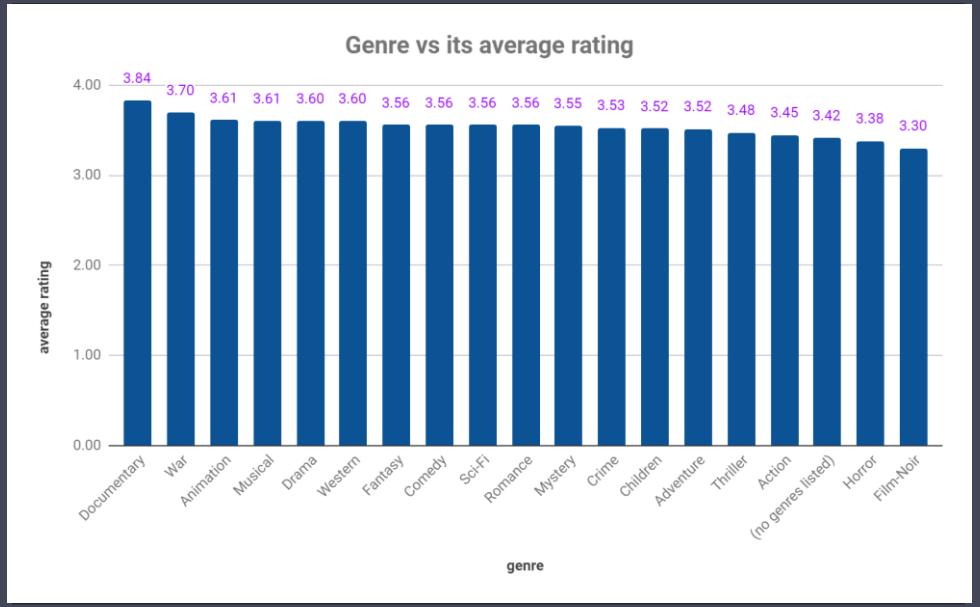
Thank you!

STEVEN L TRUONG

https://github.com/luongtruong77

https://www.linkedin.com/in/luongtruong77/

Appendix



Average rating by genre

The Documentary genre has the highest average rating, followed by War and Animation. Horror and Film-Noir have the lowest rating of all.

Appendix

```
1 df.sample(20)
```

executed in 31ms, finished 12:18:42 2021-04-29

	movieId	title	year	userId	rating	sub_genre_1	sub_genre_2	sub_genre_3	sub_genre_4	sub_genre_5
68958	5943	Maid in Manhattan (2002)	2002	606	3.0	Comedy	Romance	None	None	None
90223	71902	Spread (2009)	2009	432	3.5	Drama	Romance	None	None	None
8924	318	Shawshank Redemption, The (1994)	1994	517	4.0	Crime	Drama	None	None	None
8093	296	Pulp Fiction (1994)	1994	437	5.0	Comedy	Crime	Drama	Thriller	None
23018	1080	Monty Python's Life of Brian (1979)	1979	608	2.0	Comedy	None	None	None	None
55863	3618	Small Time Crooks (2000)	2000	448	4.0	Comedy	Crime	None	None	None
15567	588	Aladdin (1992)	1992	330	3.0	Adventure	Animation	Children	Comedy	Musical
32496	1527	Fifth Element, The (1997)	1997	477	4.0	Action	Adventure	Comedy	Sci-Fi	None
27579	1246	Dead Poets Society (1989)	1989	318	3.5	Drama	None	None	None	None
23457	1094	Crying Game, The (1992)	1992	95	4.0	Drama	Romance	Thriller	None	None
71247	6523	Mr. Baseball (1992)	1992	599	2.0	Comedy	Romance	None	None	None
12747	457	Fugitive, The (1993)	1993	470	4.0	Thriller	None	None	None	None
6973	260	Star Wars: Episode IV - A New Hope (1977)	1977	292	4.0	Action	Adventure	Sci-Fi	None	None
85098	53322	Ocean's Thirteen (2007)	2007	28	1.5	Crime	Thriller	None	None	None
19276	788	Nutty Professor, The (1996)	1996	151	5.0	Comedy	Fantasy	Romance	Sci-Fi	None
45896	2628	Star Wars: Episode I - The Phantom Menace (1999)	1999	382	3.5	Action	Adventure	Sci-Fi	None	None
59948	4036	Shadow of the Vampire (2000)	2000	414	3.0	Drama	Horror	None	None	None
99862	158238	The Nice Guys (2016)	2016	249	4.5	Crime	Mystery	Thriller	None	None
43136	2396	Shakespeare in Love (1998)	1998	201	5.0	Comedy	Drama	Romance	None	None
95227	99007	Warm Bodies (2013)	2013	125	4.5	Comedy	Horror	Romance	None	None

20 random rows from over 100,000 rows

Each movie has one or more genres associated with it. For example, Aladdin (1992) is considered “Adventure | Animation | Children | Comedy | Musical” movie whereas Dead Poets Society (1989) is just a Drama.

Appendix

$$\text{Similarity}(p,q) = \cos\theta = \frac{p \cdot q}{\|p\|\|q\|} = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^{\,2}}\sqrt{\sum_{i=1}^{n} q_i^{\,2}}}$$

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_i)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_j)^2}}$$

Cosine similarity (top) and adjusted cosine similarity (bottom) formulas

We use cosine similarity to measure the similarity between movies (they are identical if `alpha` is 0 degree and totally different if `alpha` is 90 degree.