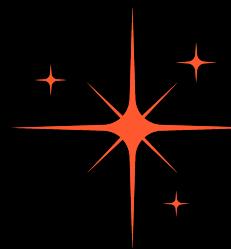


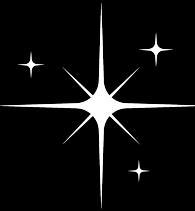


# Anime Recommender System



Alberto Cotumaccio - 1852040  
Giovanni Montobbio - 1845035





# Are they making too much anime?

Half of all TV Anime aired after 2010

Three comments on a YouTube video, where people is complaining about the huge amount of anime

<https://www.youtube.com/watch?v=GCBUZP9MA-w>



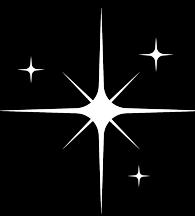
*"Instead of focusing on quality over quantity, the anime industry does the complete opposite and makes dozens of awful shows every year"*



*"Being an anime fan is like having a super nice grandma. Sure, you love their cooking, but sometimes they insist on giving you more than you asked for"*

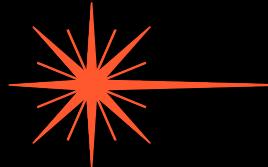
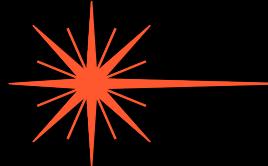
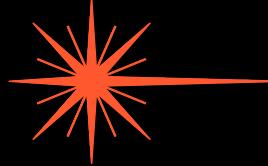


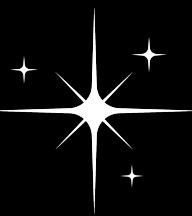
*"Years ago, having as much anime as we do now would've been a dream come true. Now, it's somewhat of a nightmare"*



# Solution: Recommender System



-  Improve the user experience
-  Increase satisfaction with the anime content
-  Assist new users in discovering the anime world
-  Help existing users to explore new anime options



# Types of Recommender Systems



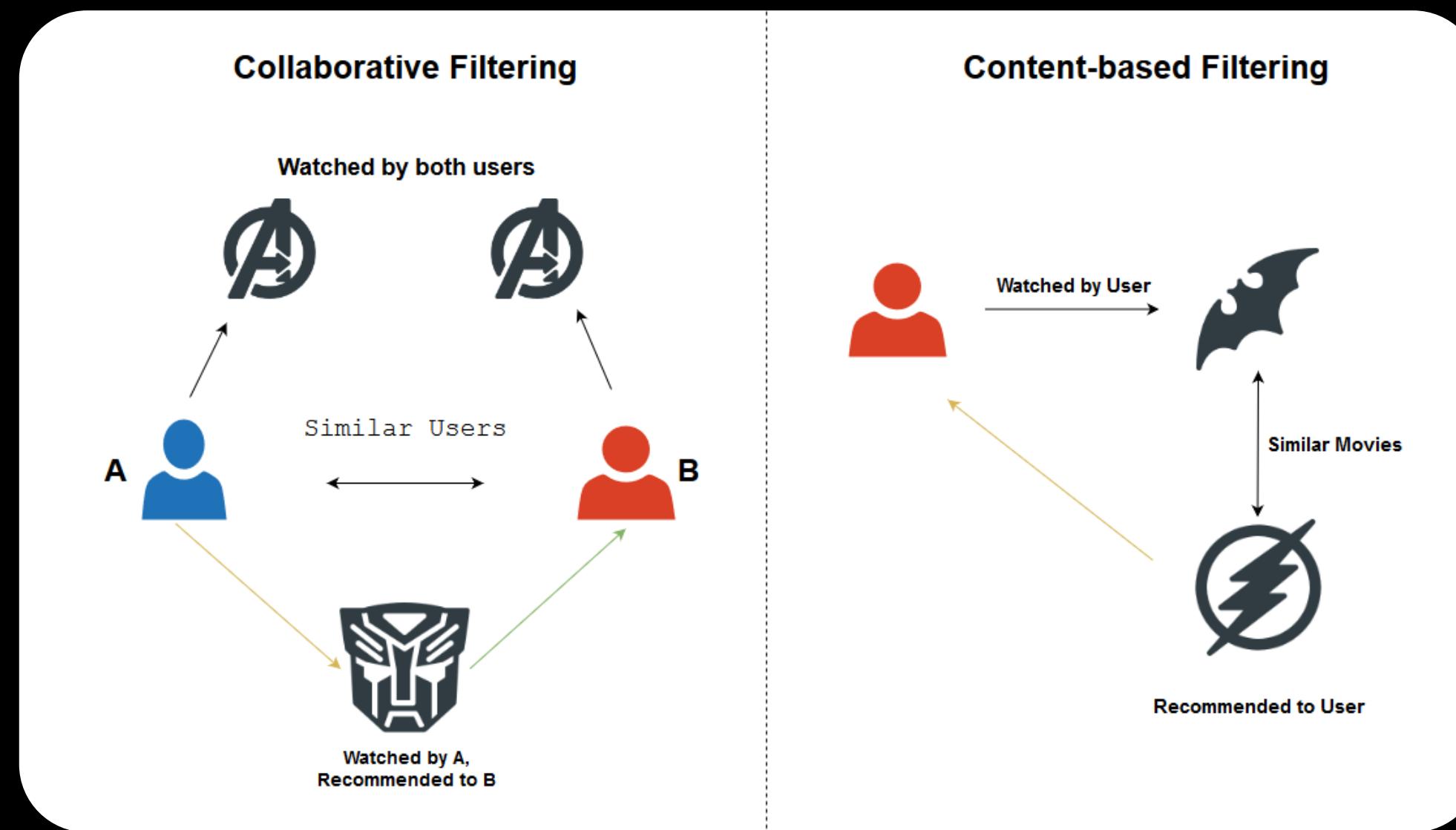
## Collaborative Filtering

- tries to find similarities between users

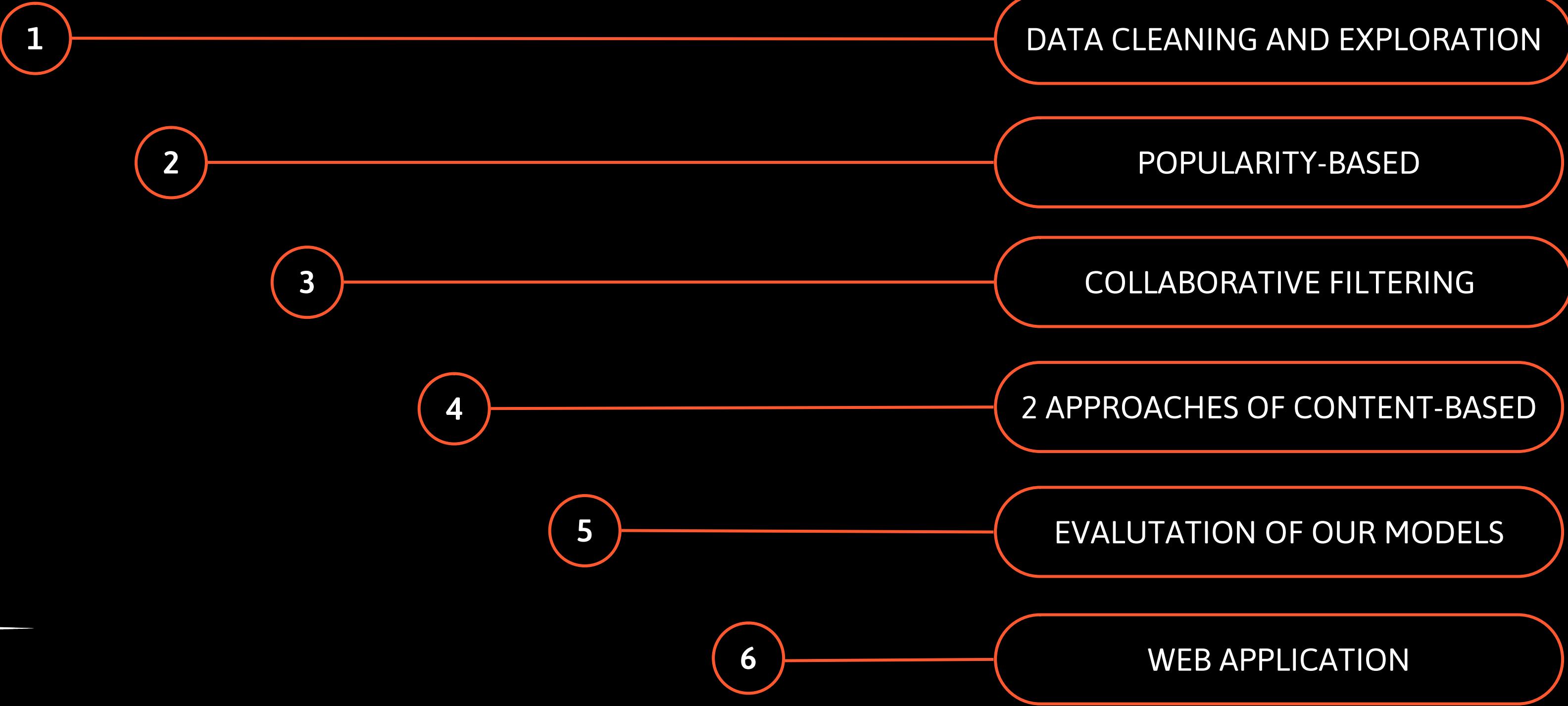


## Content-Based Filtering

- Uses the attributes of an item
- Similar features (e.g. the genre)



# Overview of our project



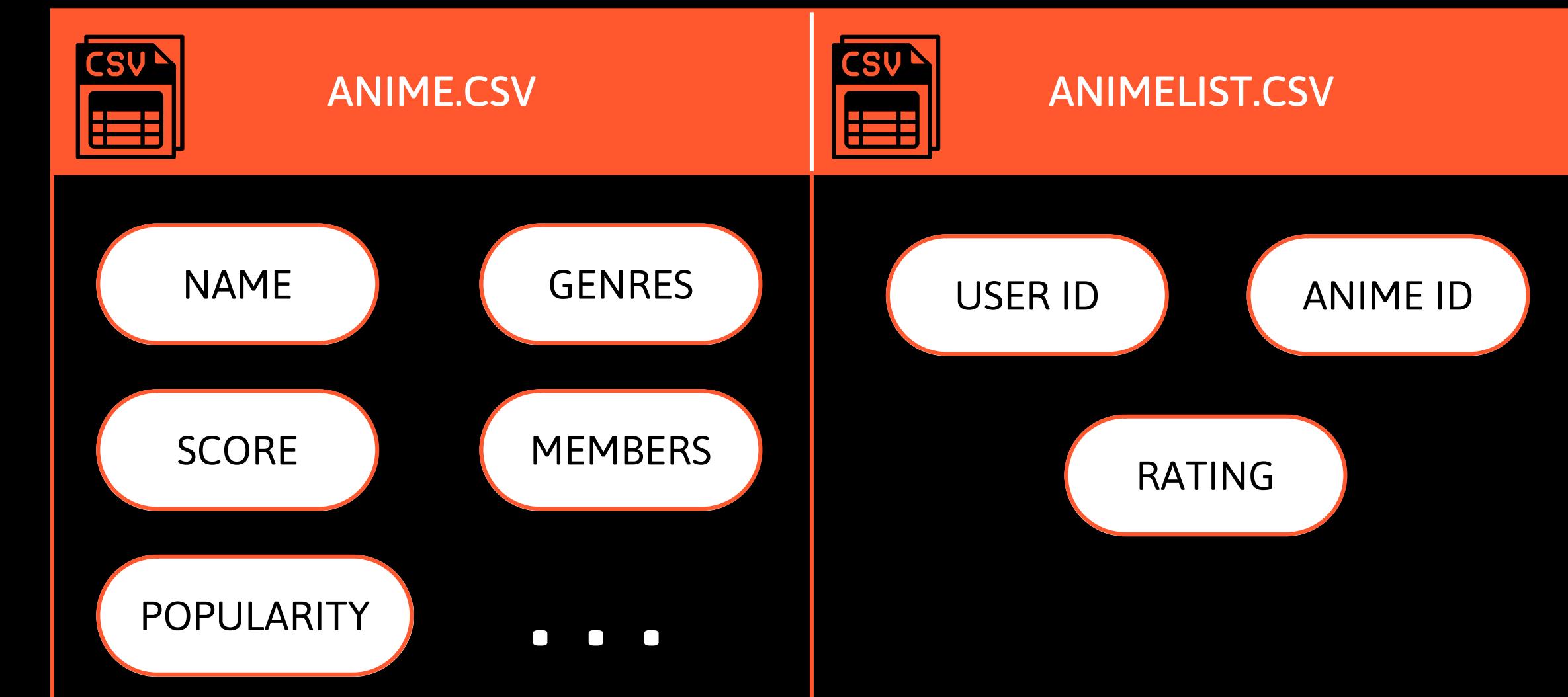
# The Dataset

The dataset we used is available on Kaggle here :  
<https://www.kaggle.com/datasets/hernan444/anime-recommendation-database-2020>

MyAnimeList Database 2020

Large Dataset  
- 109 million reviews  
- 17.562 anime  
- 325.772 unique users

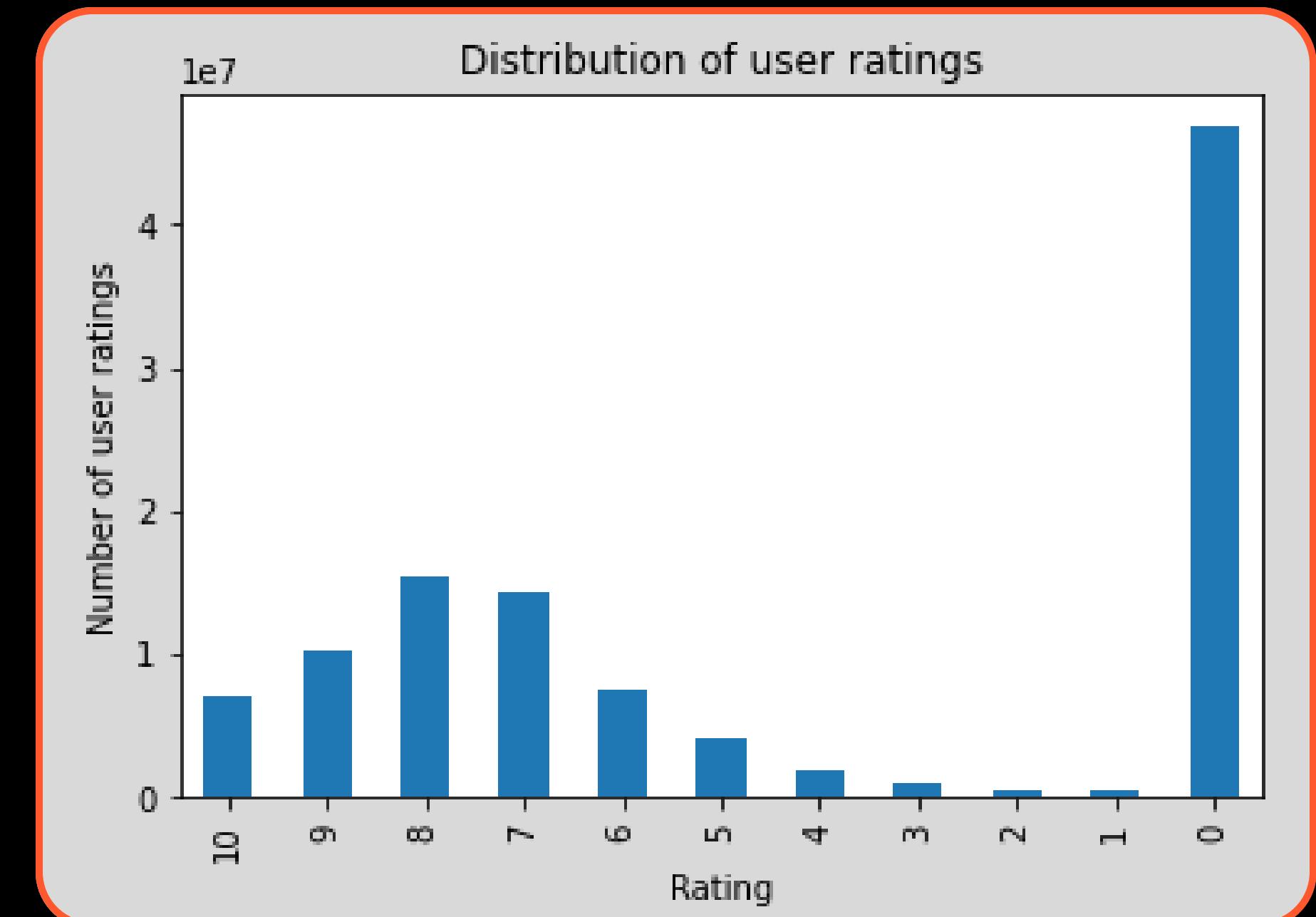
clean dataset



# Data exploration

## 1) Distribution of user ratings

- Large number of **unrated reviews**
  - uncomfortable assign a score
  - just a written review



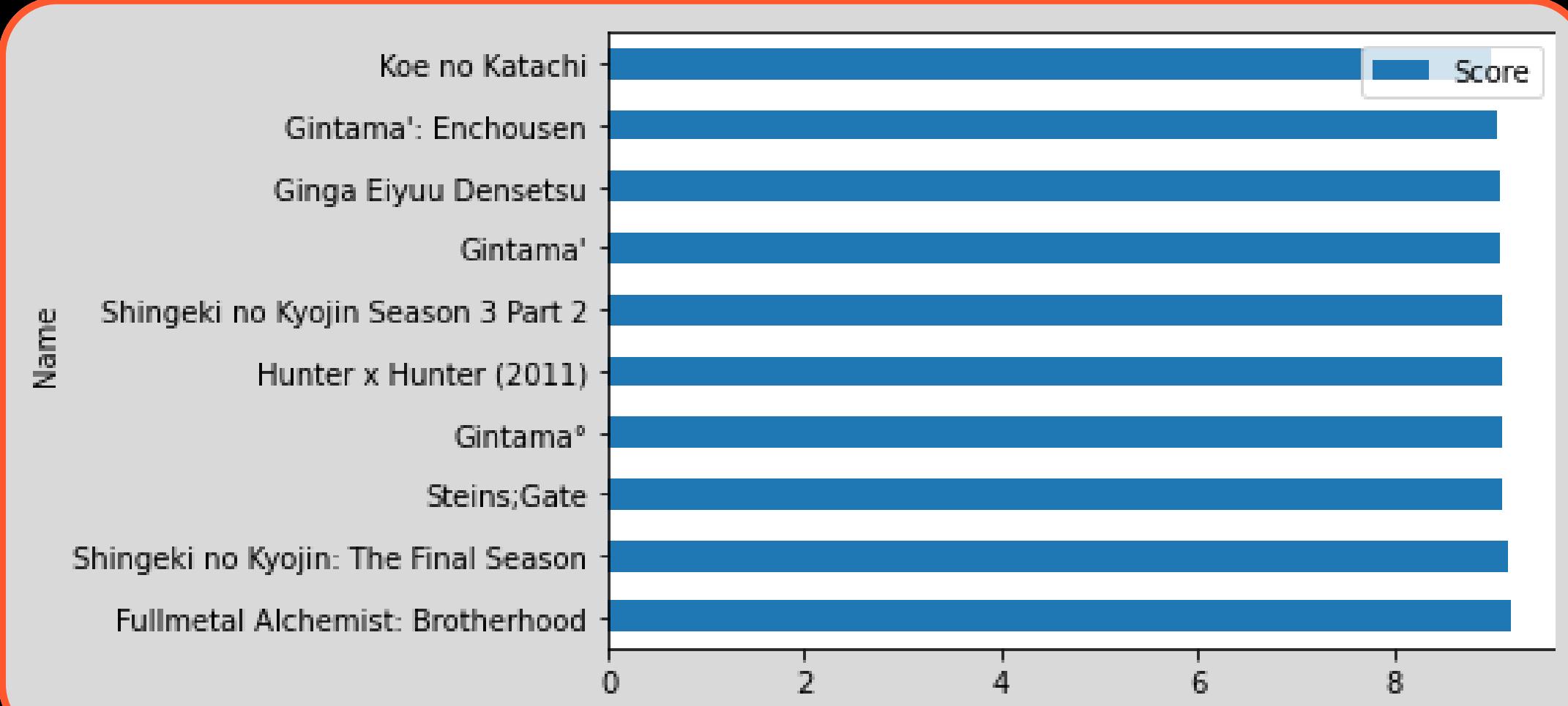
# Data exploration

## 2) Top 10 Highest Average Rating

Score column analysis

Ranging between 9.0/10 and 9.19/10

Some of those anime may not be widely known

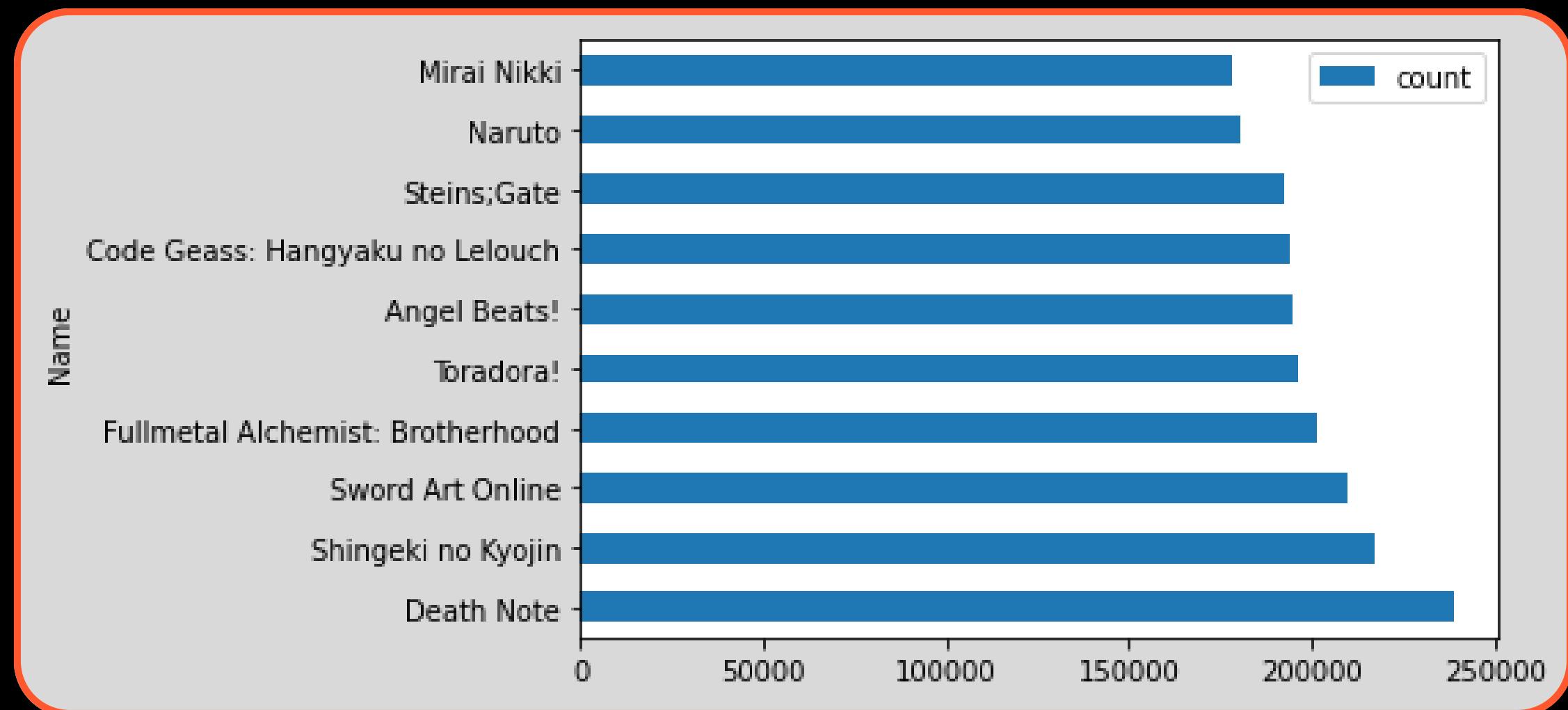


# Data exploration

## 3) Top 10 Most Reviewed Anime

very large and dedicated  
fan base

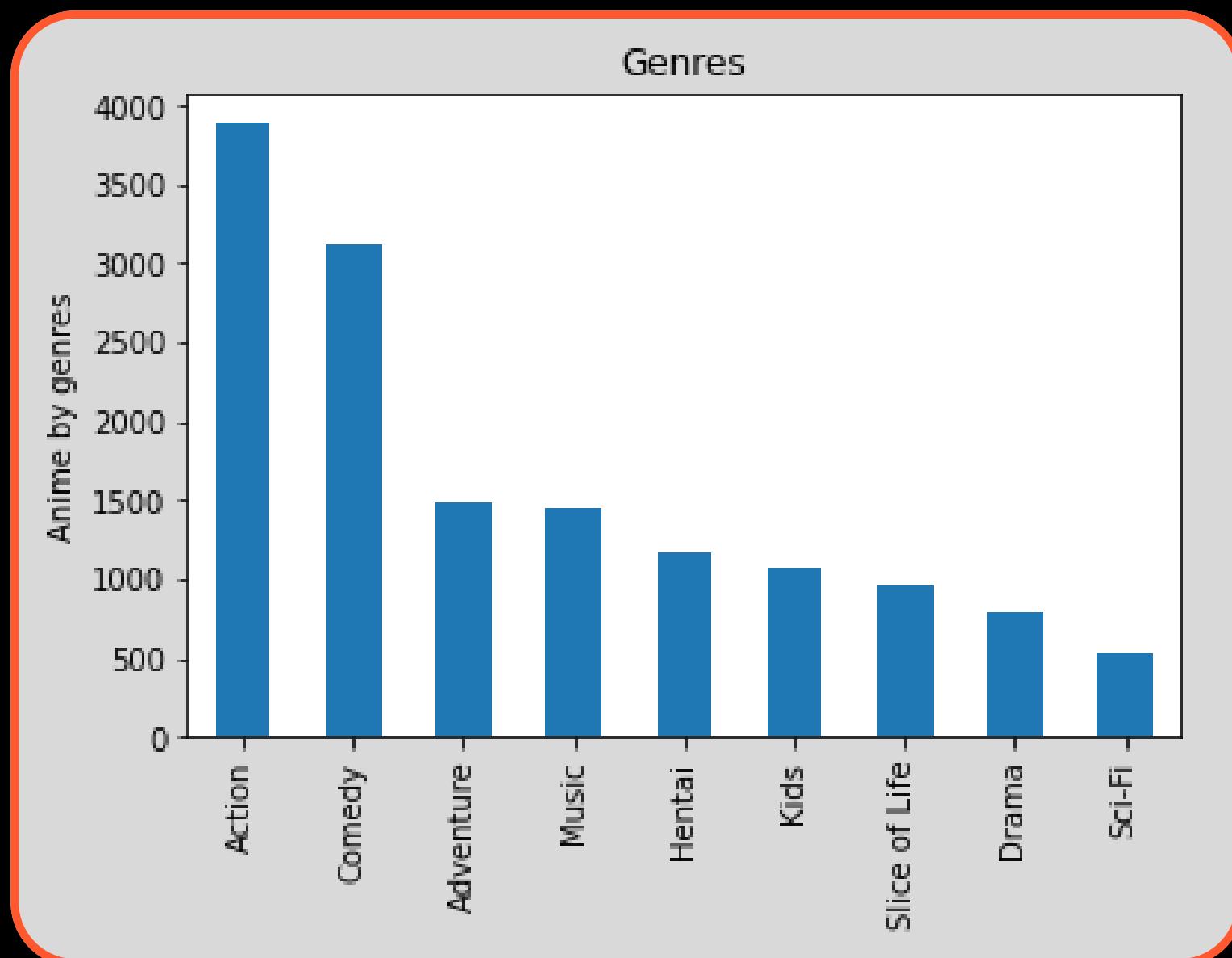
Death Note has 238.790 reviews

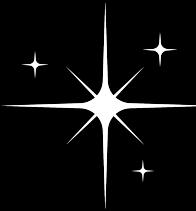


# Data exploration

## 4) Classification by genres

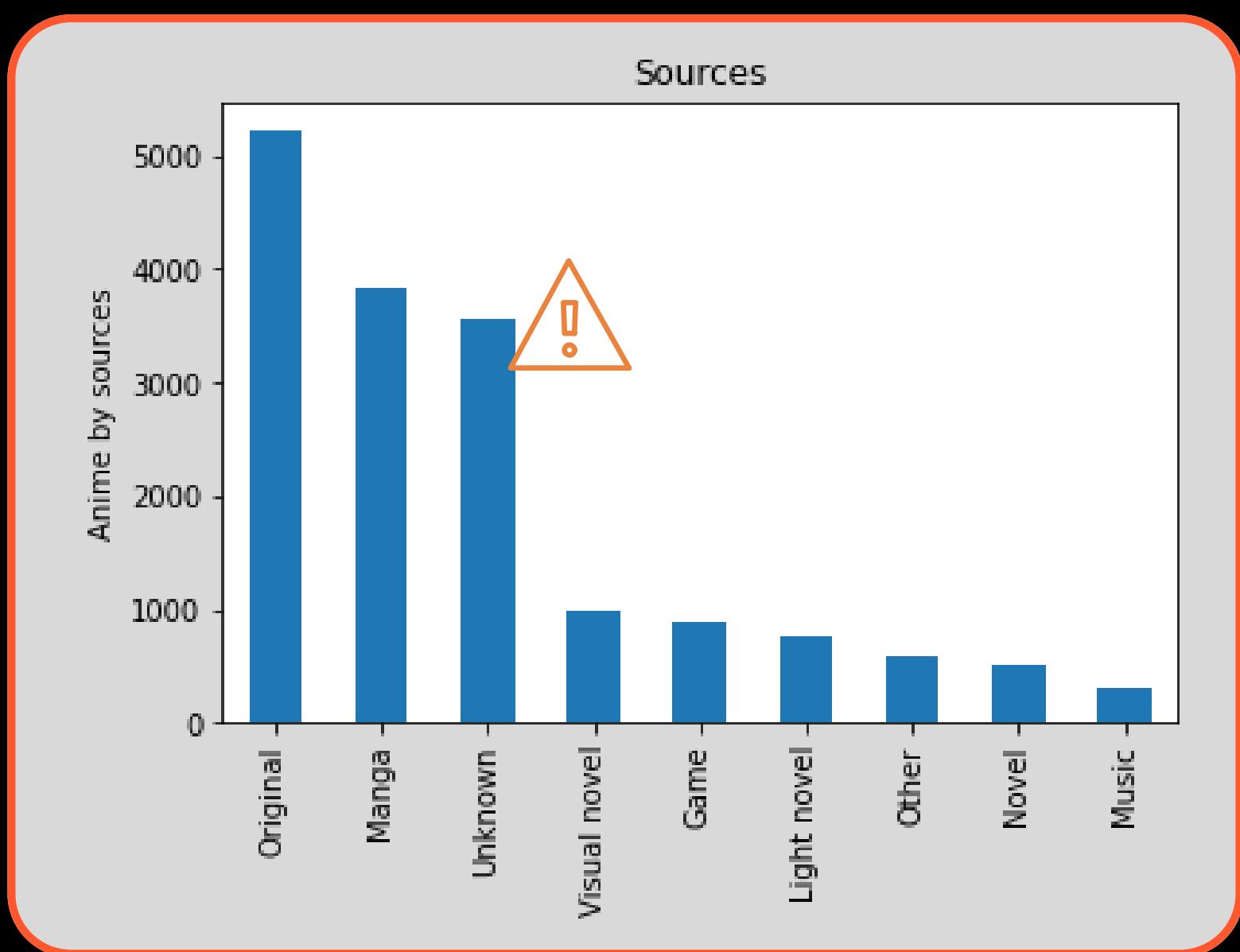
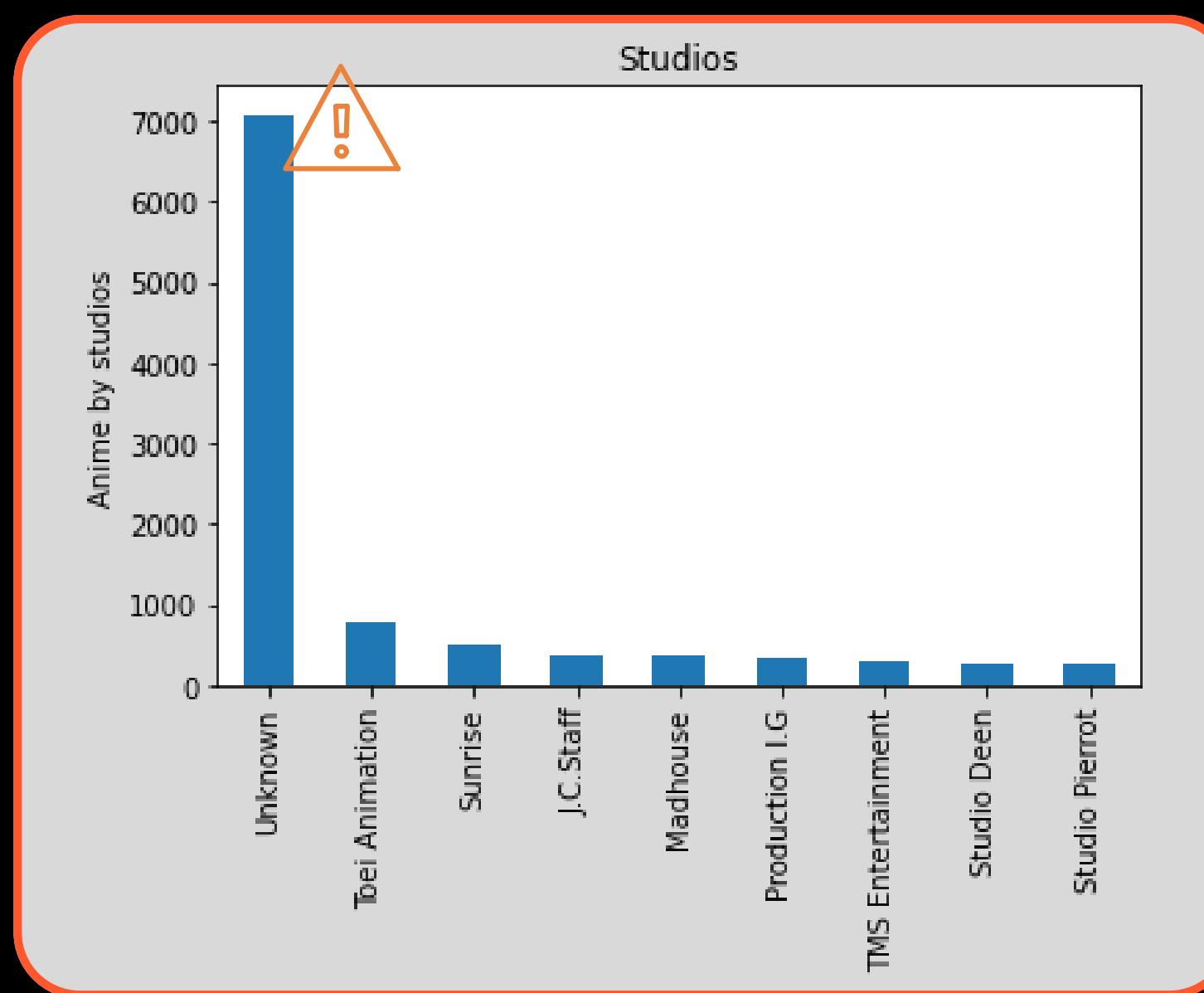
The best feature that we can use to help users find content that matches with their interests





# Data exploration

## 5) Exploration other features



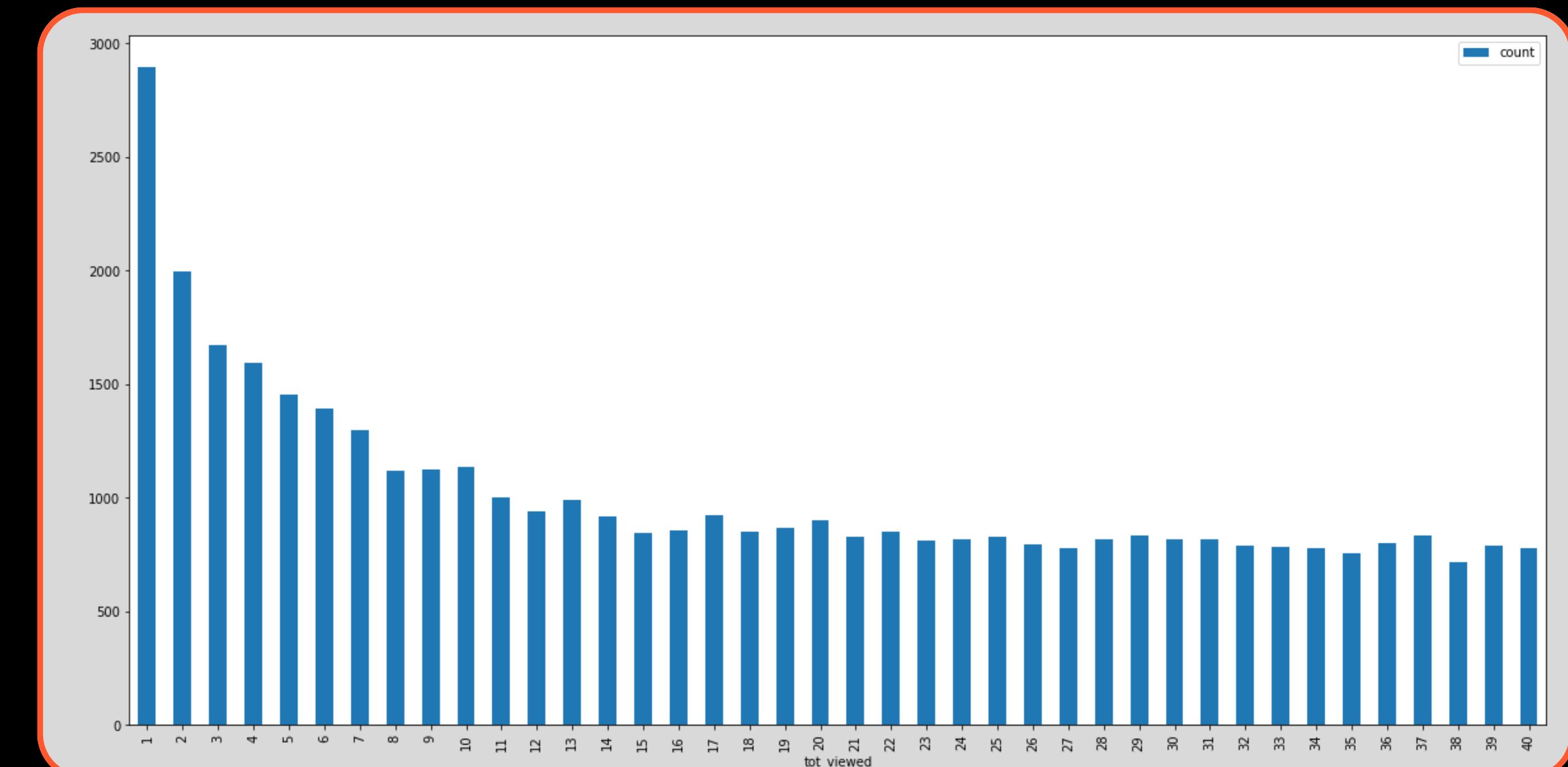
# Data exploration

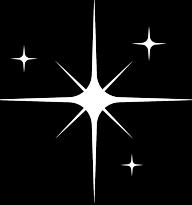
## 6) Watched Anime

most of the users reviewed  
**only one** anime

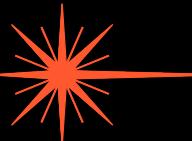
There are some users  
who have reviewed more  
than **10.000** anime!

- experencied users





# Popularity-Based



20 most popular anime



Avoid **cold start problem** and it can help new users

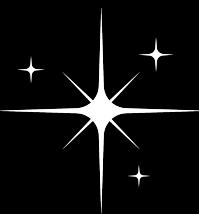


Scale most relevant factors

- **Score:** 50%
- **Members:** 25%
- **Favorites:** 25%



MAL_ID		Name	Score	Members	Favorites	Weighted_score
3971	5114	Fullmetal Alchemist: Brotherhood	1.000000	0.868280	1.000000	0.967070
1393	1535	Death Note	0.923706	1.000000	0.789505	0.909229
7448	16498	Shingeki no Kyojin	0.903270	0.977542	0.706004	0.872521
5683	9253	Steins;Gate	0.989101	0.683965	0.807182	0.867337
6474	11061	Hunter x Hunter (2011)	0.987738	0.646414	0.800776	0.855667
11	21	One Piece	0.908719	0.522377	0.688610	0.757107
11281	32281	Kimi no Na wa.	0.968665	0.666779	0.386344	0.747613
1431	1575	Code Geass: Hangyaku no Lelouch	0.935967	0.611643	0.492007	0.743896
10438	30276	One Punch Man	0.915531	0.820167	0.295981	0.736803
9881	28851	Koe no Katachi	0.974114	0.535848	0.339572	0.705912
8882	23273	Shigatsu wa Kimi no Uso	0.938692	0.556980	0.378204	0.703142
1574	1735	Naruto: Shippuuden	0.859673	0.596151	0.460275	0.693943



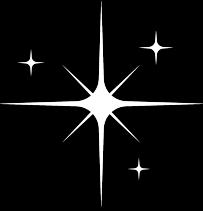
# Collaborative Filtering (ALS)

- making recommendations based on the preferences of similar users
- matrix factorization method that decomposes a matrix R into two latent factors U and V
- Minimize the loss function by applying gradient descent
- Scalability**
- Sparsity**
- Customization**
- Cold start**
- Sensitivity to initialization**



$$\arg \min_{U,V} \sum_{\{i,j|r_{i,j} \neq 0\}} (r_{i,j} - u_i^T v_j)^2 + \lambda \left( \sum_i n_{u_i} \|u_i\|^2 + \sum_j n_{v_j} \|v_j\|^2 \right)$$





---

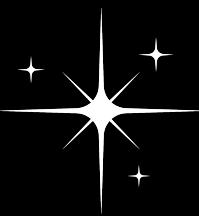
# Implementation

-  Training of the model
-  **K-fold cross-validation** was used to find best combination of parameters
  - Best model was used to make predictions on the dataset
-  Model performance evaluated with:
  - **MSE** average squared distance between the observed and predicted values
  - **MAE** treats the positive and negative errors observed as absolute
  - **RMSE** deviations are squared to prevent positive and negative values from cancelling each other out

$$RMSE = \sqrt{\frac{\sum_{i=1}(\hat{y}_i - y_i)^2}{n}}$$

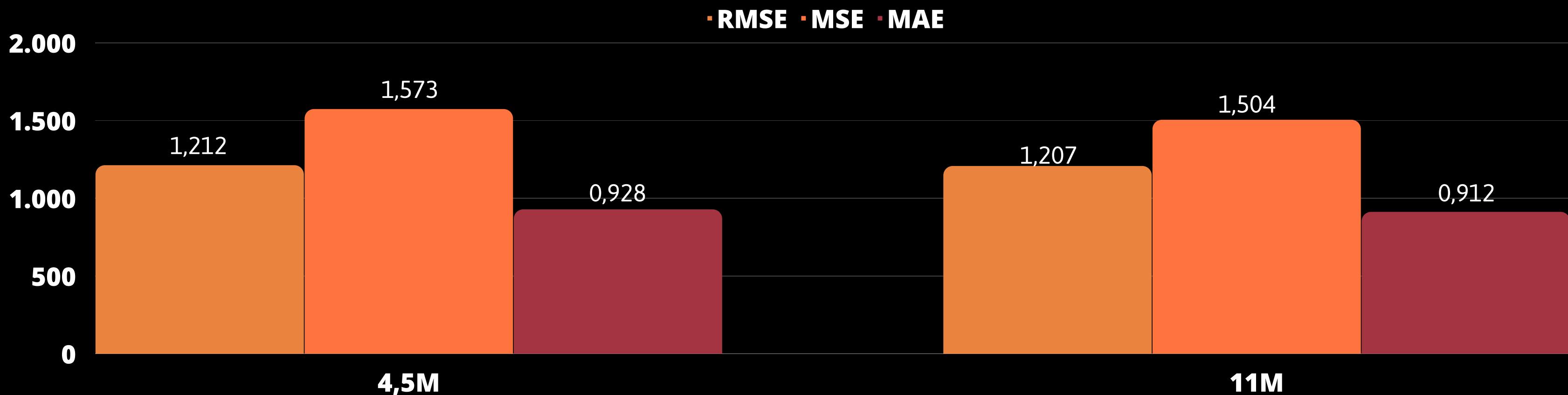
$$MAE = \frac{1}{N} * \sum_{i=1} |y_i - \hat{y}_i|$$

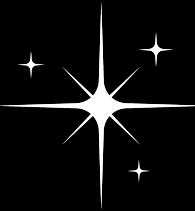
$$MSE = \frac{1}{N} * \sum_{i=1} (y_i - \hat{y}_i)^2$$



# Results

We evaluated the model with **RMSE**, **MSE** and **MAE** with different number of rows





# Content-Based with User Profile

# User Profile

MAL_ID	Name	rating
0	Captain Tsubasa J	10.0
1	Sazae-san	10.0
2	Lupin tai Holmes	10.0
3	Doraemon: Nobita to Mirai Note	10.0
4	Ansatsu Kyoushitsu	3.3
5	Hatena☆Illusion	8.8
6	Imouto Paradise! 3 The Animation	9.5
7	Kyoei Tankou-sho	4.4
8	Tate no Yuusha no Nariagari Season 3	10.0

- Personalized recommendations ✓
- Limited by user ratings ✗

# 2 One Hot Encoding technique for genres

# Weights

Action	20.0
Adventure	17.0
Comedy	17.0
Drama	10.0
Sci-Fi	11.0
Space	0.0
Mystery	6.0
Shounen	17.0
Police	0.0
Supernatural	8.0
Magic	0.0
Fantasy	17.0
Sports	12.0
Josei	0.0
Romance	8.0
Slice of Life	6.0
Cars	0.0
Seinen	0.0
Horror	0.0
Psychological	0.0
Thriller	0.0
Super Power	0.0
Martial Arts	0.0
School	3.0
Ecchi	8.0
Vampire	0.0
Military	0.0
Historical	0.0
Dementia	0.0
Mecha	4.0
Demons	0.0

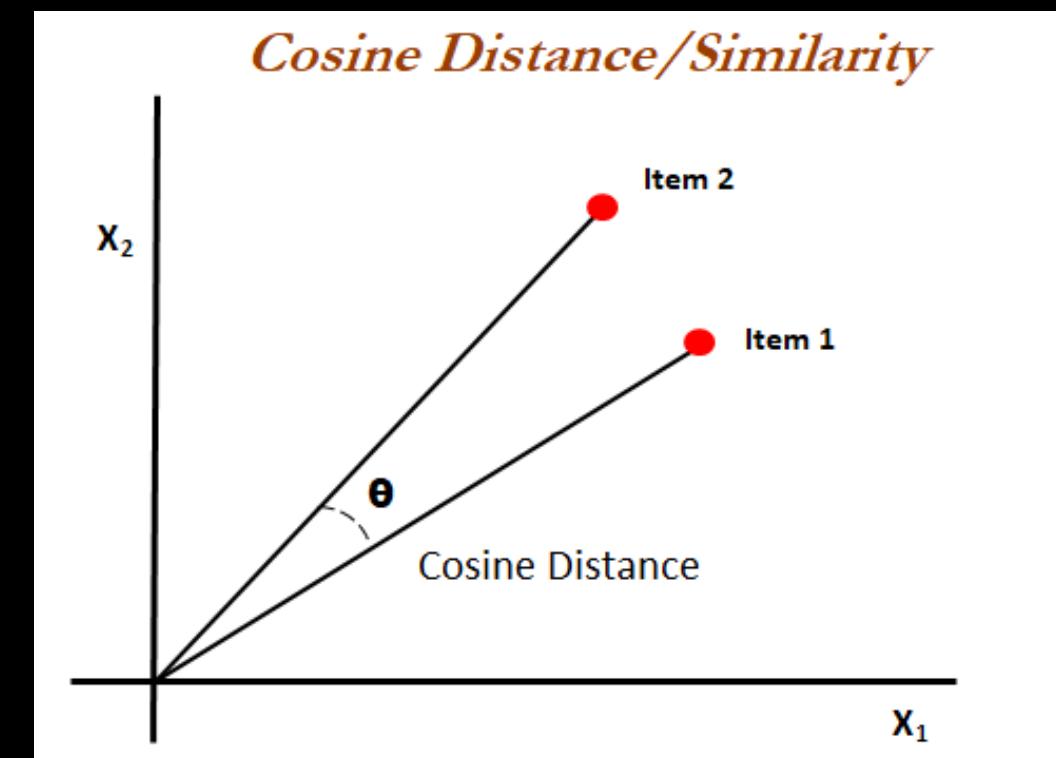
# Content-Based with Cosine Similarity

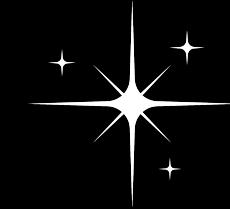
We use **cosine similarity** which measures the similarity between two vectors by taking the cosine of the angle between them. [0,1]

TfidfVectorizer library. Matrix representing the anime data in numerical form useful to calculate the similarity between every pair of anime, based on their **genre**

efficient to compute and easy to interpret ✓  
does not take into account the preferences of individual users ✗

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





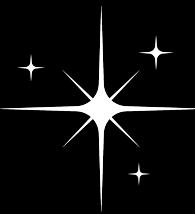
# Evaluation Content-Based



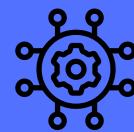
Precision and Recall to evaluate the performance of both our content-based recommendation algorithms

- 1) Select a subset of users who have made a *large number of ratings*
- 2) For each user we apply 2 evaluation functions to calculate the precision and recall for both recommendation algorithms
- 3) Calculate the average precision and recall for both algorithm

		PREDICTED	
		POSITIVE	NEGATIVE
ACTUAL	POSITIVES	TRUE POSITIVES	FALSE NEGATIVES
	NEGATIVE	FALSE POSITIVES	TRUE NEGATIVES

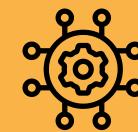


# Evaluation Content-Based



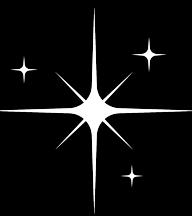
## USER-BASED EVALUATION FUNCTION

1. It first creates the User Profile of the user, taking only reviews with rating greater than 5.
2. With that User Profile we invoke the user-based recommendation to get a list of recommended titles
3. Calculates the precision and recall comparing the recommended anime titles with the remaining anime rated by the user (which were not used for recommendation).



## ITEM-BASED EVALUATION FUNCTION

1. It first selects the ratings made by the user
2. We take only the top rated anime, invoking on it the item-based recommendation to get a list of recommended anime titles
3. Calculates the precision and recall comparing the recommended anime titles with the remaining ratings made by the user (which were not used for recommendation).



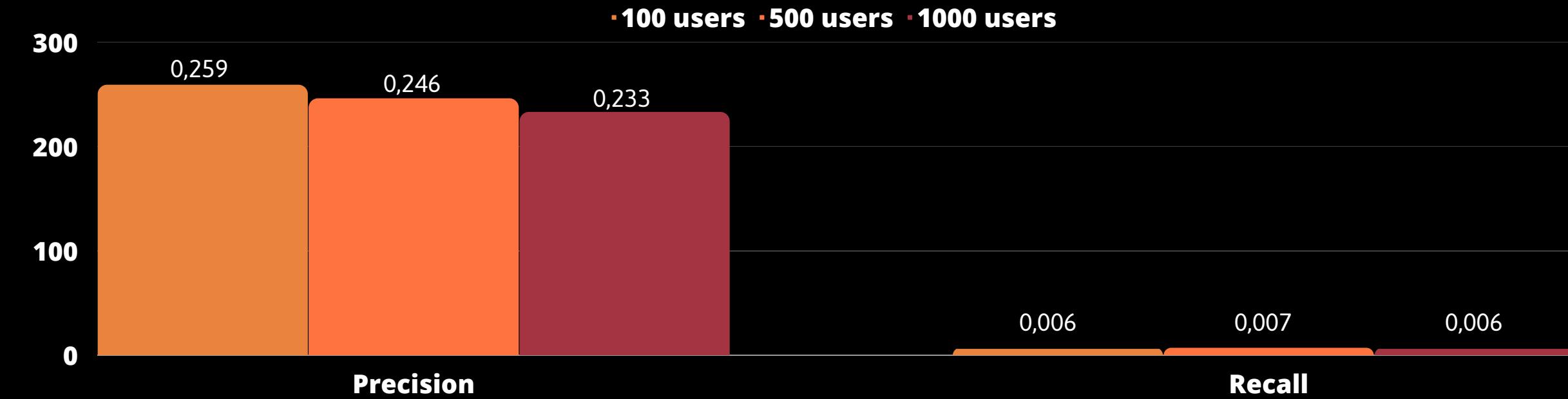
# Results



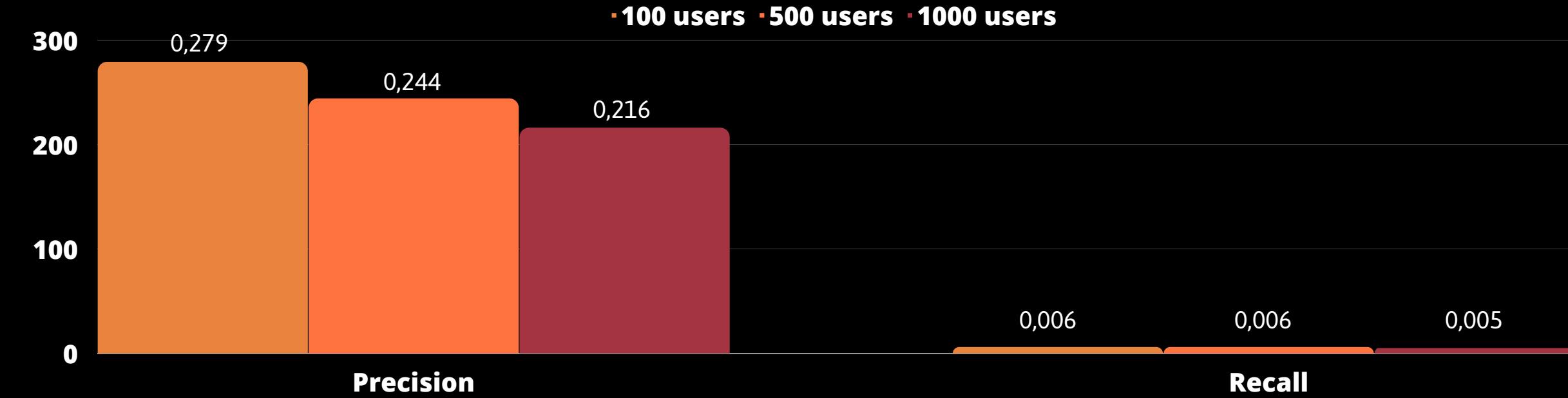
We ran evaluation for both methods with different number of users



User-Based:



Item-Based:

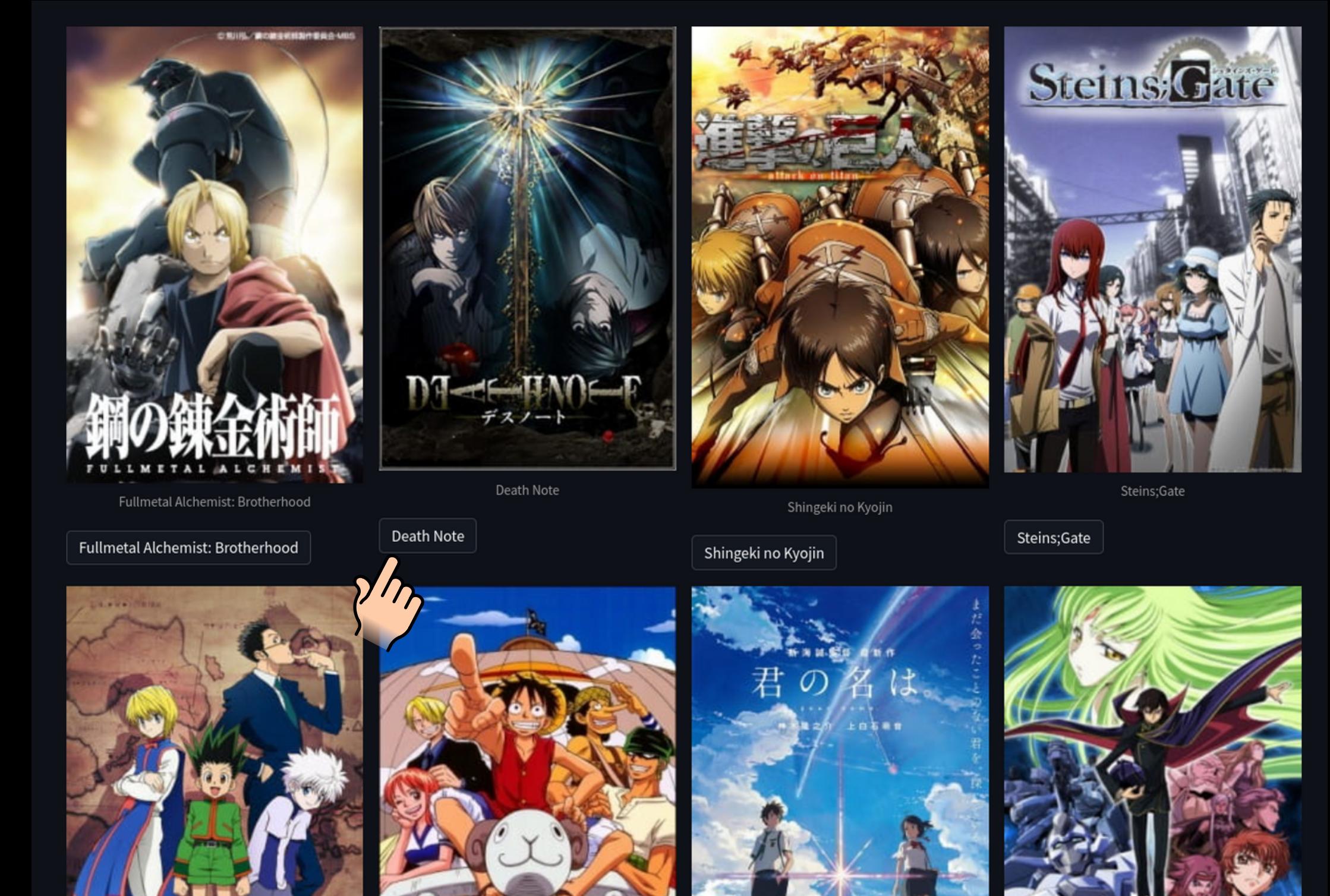


# Web Application

Streamlit library (python)

Landing page: Top 20 Animes

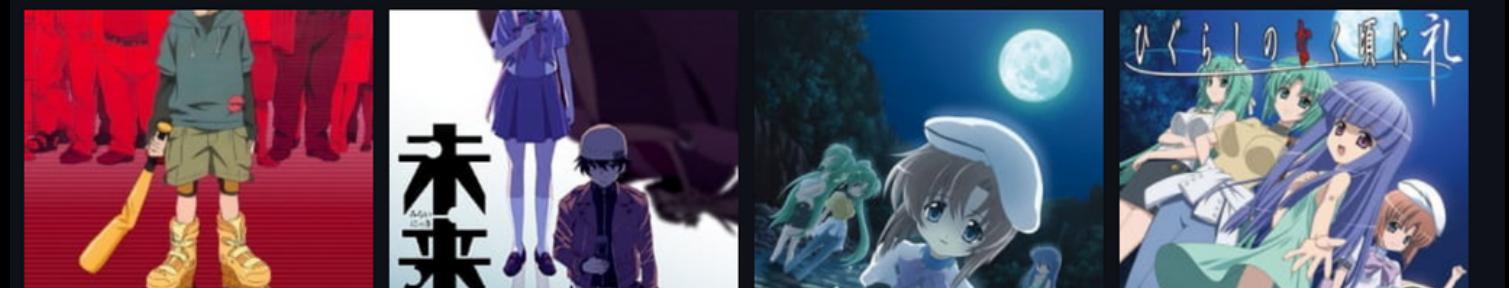
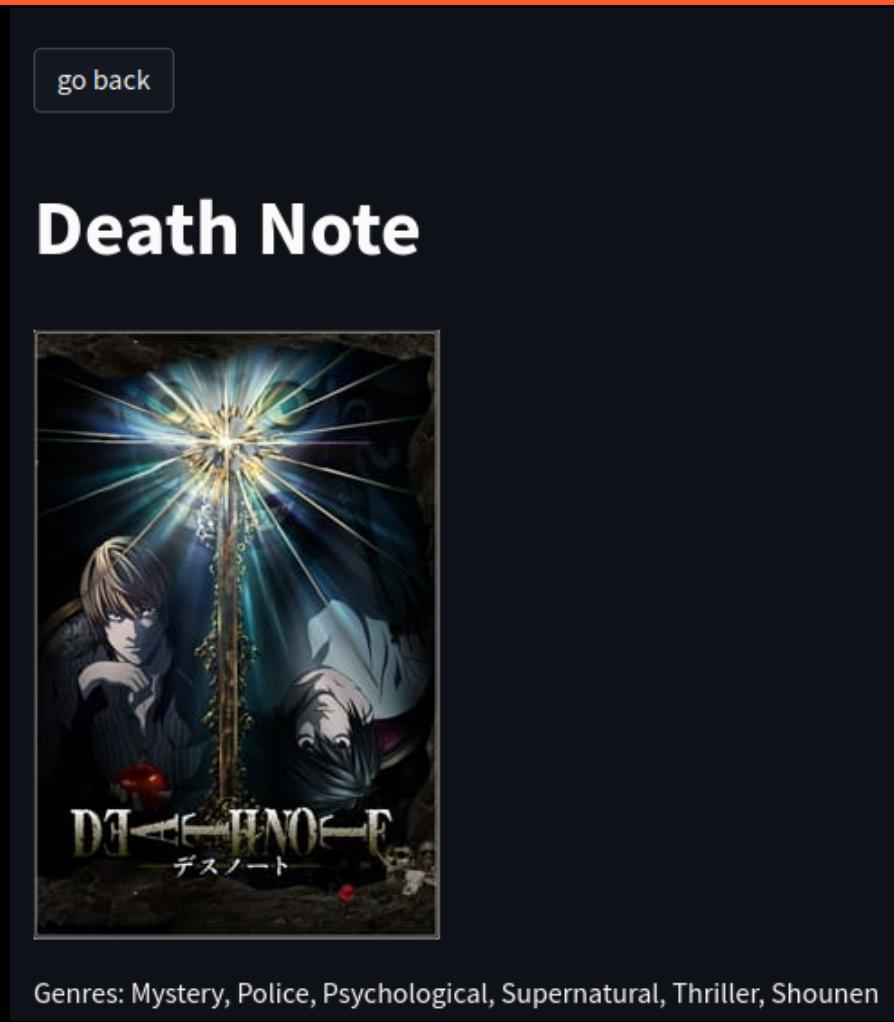
clickable buttons

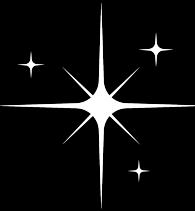


# Anime pages

The metadata of the selected anime are shown

Is invoked the content-based (item-based) algorithm





# Recommendation page

Give ratings to your favourite animes

Choose animes you have watched

One Piece X Hunter x Hunter X Dragon Ball X

One Piece

Select rating for One Piece

1

8

10

Hunter x Hunter

Select rating for Hunter x Hunter

1

10

Dragon Ball

Select rating for Dragon Ball

1

5

10



Submit



Your 20 recommended anime

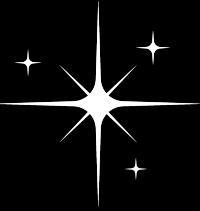
	MAL_ID	Name	Score	Genres	English name	Japanese name	Type	Episodes	Aired
200	5478	Bikkuriman 2000	<NA>	Adventure Fantasy Kids	Unknown	ピックリマン2000	TV	68	Nov 1,
664	40076	Hataraku Saibou: Necchuushou - Moshimo Pocari Sweat	6.6600	Comedy	Unknown	ボカリスエットweb movie   はたらく細胞第11.5	ONA	1	Jul 8, 2
1256	40655	Melon no Kirime	<NA>	Kids Music Slice of Life	Unknown	メロンの切り目	Music	1	Aug 1, :
1818	40957	Shin Chuuka Ichiban! 2nd Season	6.6300	Comedy Shounen	Unknown	真・中華一番!	TV	Unknown	Jan 11,
2009	37284	Tanuki to Kitsune Specials	<NA>	Comedy	Unknown	タヌキとキツネ	Special	2	May 15
2080	11099	Tiger & Bunny Pilot	6.1900	Action Comedy Super Power	Tiger & Bunny Pilot	TIGER & BUNNY (タイガー・アンド・バニー)	Special	1	May 27
5386	26023	Haiyore! Nyaruko-san F	7.5000	Sci-Fi Comedy Parody Romance	Unknown	這いよれ! ニャル子さんF	OVA	1	Jun 19
5423	38389	He Wei Dao x Re:Cong Ling Kaishi de Yi Shijie Shenghuo	5.9000	Fantasy	Unknown	合味道X《Re: 从零开始的异世界生活》	ONA	2	Sep 21
5748	8925	Kobe to Watashi	4.5700	Comedy Slice of Life	Unknown	神戸と私	ONA	1	May 9,
6520	32175	Shounen Maid	7.3000	Comedy Shoujo Slice of Life	Shounen Maid	少年メイド	TV	12	Apr 8, 2



give ratings to anime



Is invoked the content-based  
(user-based) algorithm



---

**Thanks for the  
attention**

---