

MyAnimeList

ANIME RECOMMENDATION ENGINE

YILU CHEN

ANDREW GATCHALIAN

SARA HART

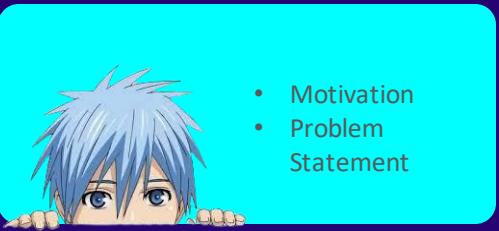
HSUAN-YI LIN

RAKESH VENKATA SUBRAMANIYAN



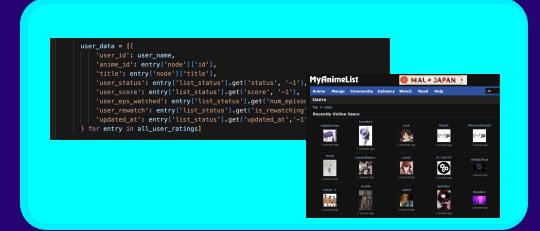
TABLE OF CONTENTS

INTRODUCTION

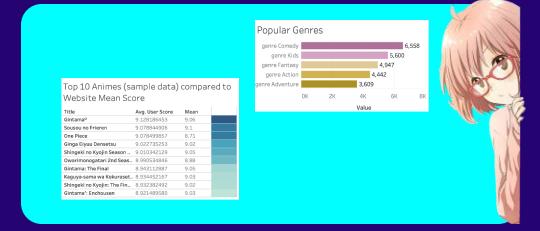


- Motivation
- Problem Statement

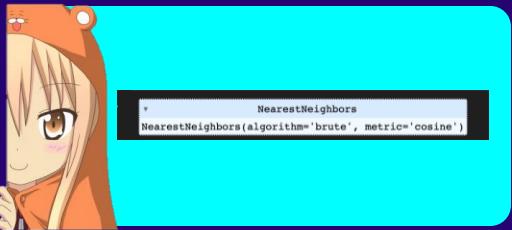
DATA PRE-PROCESSING



EXPLORATORY ANALYSIS



KNN MODEL



DEMO



CONCLUSION



INTRODUCTION



WHY ANIME?

- *More than one-third of the world population* watch anime
- It's a **global phenomenon**. The top 5 countries with the most anime watchers are:

1. Japan (76%)
2. United States (72%)
3. Philippines (71%)
4. France (63%)
5. Brazil (56%)





PROBLEM STATEMENT

Can accurate recommendations be given to viewers based on historical viewing data?

DATA PRE- PROCESSING



DATA COLLECTION

1.1 Pull all Animes

```
CLIENT_ID = '018269e286b49a7ca078f74376d45377'  
  
# Initialize an empty list to store all the data  
all_anime_list = []  
df_AAL = pd.DataFrame(columns=['anime_id', 'title',
```

```
#update this (ex. 10400 - 20800)  
anime_id_start = 0  
anime_id_limit = 10000
```

A	B	C	D	E	F	G	H	I	J
anime_id	title	mean	genres	studios	synopsis	media_type	num_episodes		
1	Cowboy Bebop	8.75	Action, Adult	Sunrise	Crime is	tv	26		
5	Cowboy Bebop	8.38	Action, Adult	Bones	Another	movie	1		
6	Trigun	8.21	Action, Adult	Madhouse	Vash the	tv	26		
7	Witch Hunter	7.25	Action, Dete	Sunrise	Robin Sena	tv	26		
8	Bouken Ou B	6.94	Adventure, F	Toei Animati	It is the dark	tv	52		
15	Eyeshield 21	7.92	Shounen, Sp	Gallop	Shy,	tv	145		
16	Hachimitsu t	8	Adult Cast, C	J.C.Staff	Yuuta	tv	24		
17	Hungry Heart	7.55	Comedy, Sho	Nippon Anim	As the	tv	52		
18	Initial D Four	8.16	Action, Dram	A.C.G.T.	Takumi	tv	24		
19	Monster	8.87	Adult Cast, D	Madhouse	Dr. Kenzou	tv	74		
20	Naruto	7.99	Action, Adve	Pierrot	Moments	tv	220		
21	One Piece	8.71	Action, Adve	Toei Animati	Barely	tv	0		
22	Tennis no Ou	7.86	School, Shou	Trans Arts	At the	tv	178		
23	Ring ni Kakeru	6.4	Action, Com	Toei Animati	In order to	tv	12		
24	School Rumble	7.89	Comedy, Gag	Studio Come	Just the	tv	26		
25	Sunabouzu	7.38	Action, Adve	Gonzo	The Great	tv	24		
26	Technolyze	7.76	Action, Dram	Madhouse	In the dark	tv	22		
27	Trinity Blood	7.29	Action, Super	Gonzo	Following	tv	24		
28	Yakitate!! Ja	7.91	Comedy, Gol	Sunrise	While	tv	69		
29	Zipang	7.48	Action, Adult	Studio Deen	A Ripple in	tv	26		
30	Neon Genesis	8.35	Action, Avan	Gainax, Tats	Fifteen	tv	26		
31	Neon Genesis	7.47	Drama, Mecl	Gainax, Prod	In the year	movie	1		
32	Neon Genesis	8.55	Avant Garde	Gainax, Prod	Shinji Ikari	movie	1		
33	Kenpuu Denk	8.57	Action, Adve	OLM	Guts, a man	tv	25		
43	Koukaku Kidou	8.27	Action, Adult	Production I.	In the year	movie	1		
44	Rurouni Kenshi	8.7	Action, Adult	Studio Deen	When	ova	4		
45	Rurouni Kenshi	8.29	Action, Adve	Gallop, Studi	In the final	tv	94		

DATA COLLECTION

1.2 Pull Usernames of Viewers

MyAnimeList



Anime Manga Community Industry Watch Read Help All

Users

Top > Users

Recently Online Users

reiddDemise	kumata1	mxk	Rejain	Marciovinicius9
2 seconds ago	7 seconds ago	2 seconds ago	1 second ago	1 second ago
larasl	Constellation-	LaviS	el_capone	eXtInCtPool
1 second ago	1 second ago	1 second ago	1 second ago	1 second ago
rehan-1	Gedrik	noisuf	BePotter	BogdanJ
1 second ago	1 second ago	1 second ago	1 second ago	1 second ago

```
# Add a 5-second delay before making the next request  
time.sleep(5)
```

```
# Now you have collected at least 100 unique usernames  
for username in unique_usernames:  
    print(username)
```

[7] ⏪ 13.1s

```
... Collected 21 unique usernames  
Collected 41 unique usernames  
Collected 61 unique usernames
```

```
from bs4 import BeautifulSoup
```

```
# The URL to fetch  
url = "https://myanimelist.net/users.php?lucky=1"
```

```
# Minimum number of unique usernames you want to collect  
minimum_user_count = 10000
```

A	B
1	username
2	#NAME?
3	-----9-----
4	#NAME?
5	--Ollie---
6	--OoFF---
7	--DarkSoul--
8	--Dreamy--
9	--Kuro--
10	--Lunium--
11	--Mayhem--
12	--Sleepy--
13	--bee--
14	#NAME?
15	--I--
16	#NAME?
17	-ARZ-
18	-Agito-
19	-Akkey-
20	-AlexL-
21	#NAME?
22	-Amnesty-
23	-Anii-
24	-Aozora-



DATA COLLECTION

1.3 Get Anime Ratings from Users

```
# THIS IS FOR CHUNKING DATA
if df_user_ratings.shape[0] > 1000000: #if number of rows exceed 1,000,000 (for excel) export to excel
    df_user_ratings.to_csv(f'all_user_ratings_2_{list_counter}.csv', index=False)
    list_counter += 1
# Initialize an empty list
all_user_ratings = []
```



```
user_data = [
    'user_id': user_name,
    'anime_id': entry['node']['id'],
    'title': entry['node']['title'],
    'user_status': entry['list_status'].get('status', '-1'),
    'user_score': entry['list_status'].get('score', '-1'),
    'user_eps_watched': entry['list_status'].get('num_episodes_watched', '-1'),
    'user_rewatch': entry['list_status'].get('is_rewatching', '-1'),
    'updated_at': entry['list_status'].get('updated_at', '-1')
} for entry in all_user_ratings]
```



The screenshot shows a Jupyter Notebook cell containing Python code for collecting anime ratings. The code uses a list comprehension to iterate through `all_user_ratings` and extract user data for each entry. The resulting data is stored in a variable named `user_data`, which is a list of dictionaries. Each dictionary contains the following keys: `user_id` (set to `user_name`), `anime_id` (extracted from the `node` key of the entry), `title` (extracted from the `title` key of the entry), `user_status` (extracted from the `list_status` key of the entry), `user_score` (extracted from the `list_status` key of the entry), `user_eps_watched` (extracted from the `list_status` key of the entry), `user_rewatch` (extracted from the `list_status` key of the entry), and `updated_at` (extracted from the `list_status` key of the entry). The code is intended to handle large datasets by chunking them into smaller CSV files if they exceed 1,000,000 rows.

	user_id	anime_id	title	user_status	user_score	user_eps_watched	user_rewatch	updated_at
0	NV_	52034	"Oshi no Ko"	completed	8	11	False	2023-06-29T00:34:07+00:00
1	NV_	55791	"Oshi no Ko" Season 2	plan_to_watch	0	0	False	2023-06-29T18:14:53+00:00
2	NV_	31646	3-gatsu no Lion	completed	6	22	False	2017-03-18T19:53:53+00:00
3	NV_	35180	3-gatsu no Lion 2nd Season	completed	7	22	False	2018-03-31T23:53:30+00:00
4	NV_	38101	5-toubun no Hanayome	completed	7	12	False	2019-03-31T12:44:07+00:00
...
4799	Rhythmdavis4	40784	Wu Liuqi: Zui Qiang Fa Xing Shi	watching	9	6	False	2021-06-04T02:14:08+00:00
4800	Rhythmdavis4	28	Yakitate!! Japan	watching	0	0	False	2021-03-16T14:48:35+00:00
4801	Rhythmdavis4	37779	Yakusoku no Neverland	completed	8	12	False	2020-09-07T22:36:37+00:00
4802	Rhythmdavis4	39617	Yakusoku no Neverland 2nd Season	dropped	0	0	False	2021-02-06T23:00:01+00:00
4803	Rhythmdavis4	7785	Yojouhan Shinwa Taikei	plan_to_watch	0	0	False	2020-10-13T14:11:43+00:00



DATA CLEANING



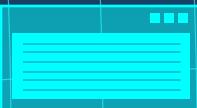
```
combined_data = combined_data.dropna()
```

```
combined_data = combined_data.drop_duplicates()
```

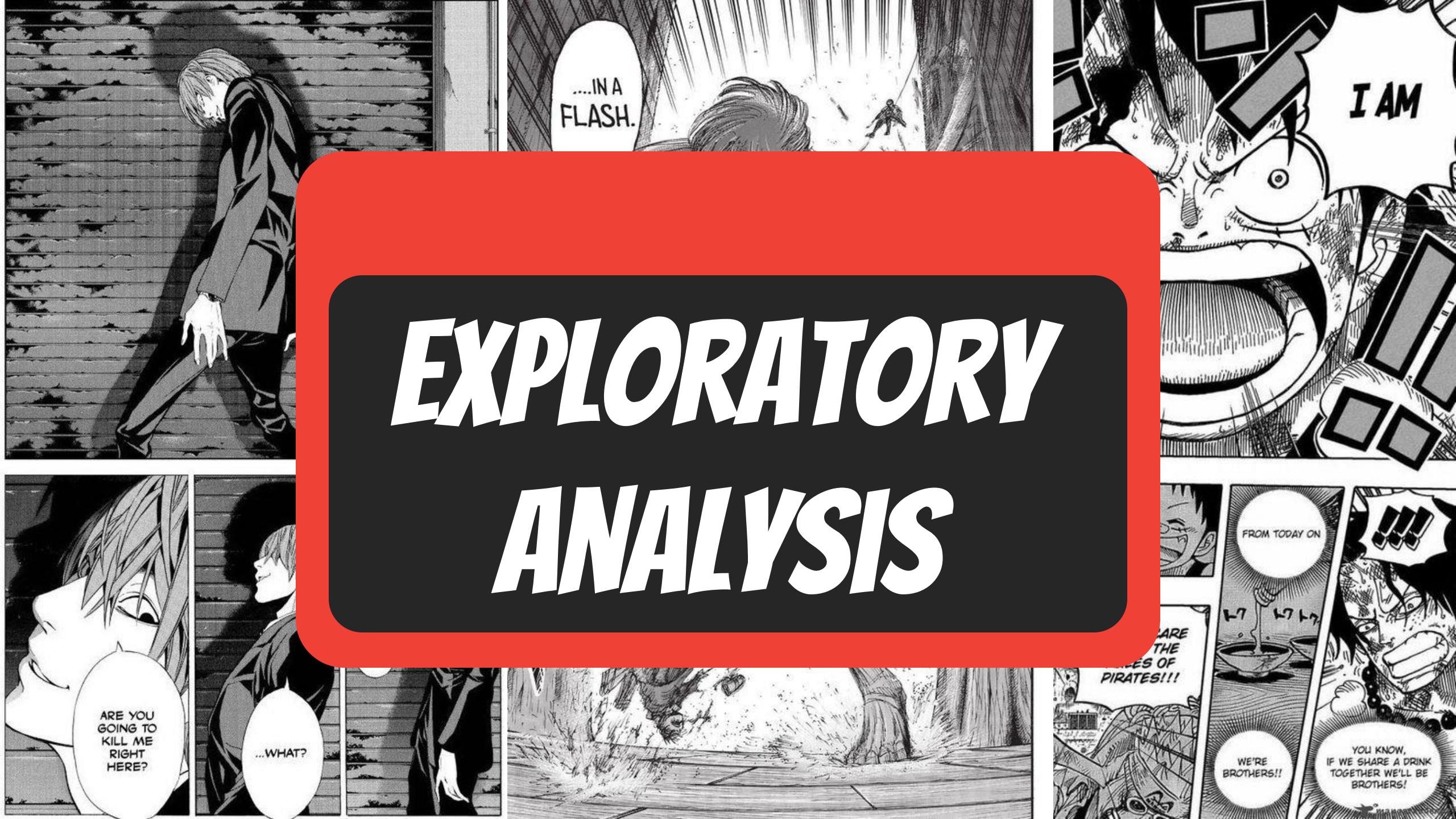
```
#remove all rows that are animes that users havent watched  
filtered_df = filtered_combined_data[(filtered_combined_data['user_status'] != 'plan_to_watch')]  
filtered_df.reset_index(drop=True, inplace=True)
```

```
combined_data['updated_at'] = pd.to_datetime(combined_data['updated_at'])  
combined_data = combined_data.sort_values(by='updated_at', ascending=False)  
combined_data
```

```
dummy_variables_genres = test_df['genres'].str.get_dummies(' ', '').add_prefix('genre_')
```



```
#remove all rows where user rating is 0 - assumes this means they did not rate it  
filtered_df = filtered_df[~(filtered_df['user_score'] == 0)]  
filtered_df.reset_index(drop=True, inplace=True)
```



EXPLORATORY ANALYSIS

ARE YOU
GOING TO
KILL ME
RIGHT
HERE?

...WHAT?

....IN A
FLASH.

I AM

FROM TODAY ON

ARE
THE
LIES OF
PIRATES!!!

WE'RE
BROTHERS!!

YOU KNOW,
IF WE SHARE A DRINK
TOGETHER WE'LL BE
BROTHERS!

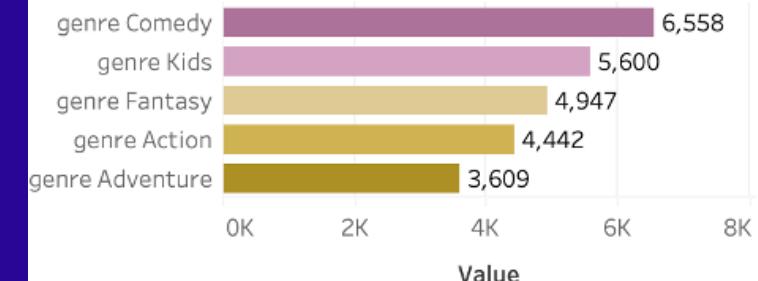
FINDINGS

Top 10 Animes (sample data) compared to Website Mean Score

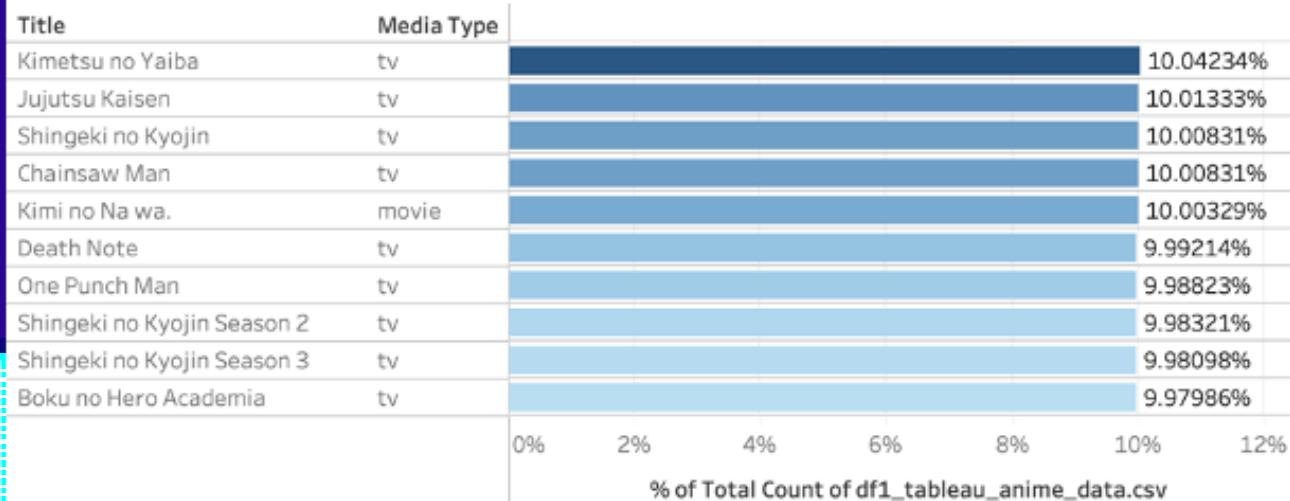
Title	Avg. User Score	Mean
Gintama°	9.128186453	9.06
Sousou no Frieren	9.078844906	9.1
One Piece	9.078499857	8.71
Ginga Eiyuu Densetsu	9.022735253	9.02
Shingeki no Kyojin Season ..	9.010342129	9.05
Owarimonogatari 2nd Seas..	8.990534846	8.88
Gintama: The Final	8.943112887	9.05
Kaguya-sama wa Kokuraset..	8.934452167	9.03
Shingeki no Kyojin: The Fin..	8.932382492	9.02
Gintama': Enchousen	8.921489580	9.03



Popular Genres



Most Popular Animes (% of total)



FINDINGS

Top 10 most Completed Animes

Title	Media Type	Mean	Avg. User Score	
Kimi no Na wa.	movie	8.84	8.573302814	17,912
Koe no Katachi	movie	8.93	8.796756209	17,757
Shingeki no Kyojin Season 3	tv	8.62	8.519124450	17,726
Shingeki no Kyojin Season 2	tv	8.51	8.594115642	17,606
Boku no Hero Academia 2nd Season	tv	8.1	7.527523718	17,603
Shingeki no Kyojin Season 3 Part 2	tv	9.05	9.019778346	17,595
Kimetsu no Yaiba Movie: Mugen Ressha-hen	movie	8.61	8.204267425	17,575
Jujutsu Kaisen 0 Movie	movie	8.47	8.323507654	17,573
Boku dake ga Inai Machi	tv	8.3	7.930553969	17,510
Kimetsu no Yaiba: Yuukaku-hen	tv	8.78	8.397312750	17,490

0K 5K 10K 15K 20K
Count of df1_tableau_anime_data.csv



Top 5 most Dropped Animes

Title	Media Type	Mean	Avg. User Score	
Boruto: Naruto Next Generations	tv	6.04	3.764241088	5,442
Ex-Arm	tv	2.9	1.431438792	5,032
Platinum End	tv	6.05	3.816373967	3,872
Gibiate	tv	3.81	2.520192887	3,318
Super Dragon Ball Heroes	ona	5.32	3.628657617	3,315

0K 1K 2K 3K 4K 5K 6K
Count of df1_tableau_anime_data.csv



KNN MODEL



KNN MODELING

1.

```
import pandas as pd
from sklearn.neighbors import NearestNeighbors

# read the user ratings
user_ratings = pd.read_csv('cleaned/anime_user_ratings_cleaned_final.csv')
```

2.

```
# Create a user-item matrix with anime_id as rows, user_id as columns, and ratings as values.
user_item_matrix = user_ratings.pivot_table(index='anime_id', columns='user_id', values='user_score', fill_value=0)
```

4.

user_id	jR4dh3y	Arkhalis-	Artoria-	-Black-Reaper-	CandyConfetti-	Cheesecake-	Dutchy-	EyeOfHorus-	FMA	Froggu-	...	zuo_bai	z
anime_id	1	8	9	9	7	6	7	8	8	8	5	...	0
5	9	9	10	7	8	7	8	9	9	9	9	...	0
6	9	7	9	8	8	7	7	9	8	9	9	...	0
7	0	7	0	0	0	0	0	0	0	0	9	...	0
8	0	0	0	0	0	0	0	0	0	0	0	...	0
...
56979	0	0	0	0	0	0	0	0	0	0	0	...	0
57003	0	0	0	0	0	0	0	0	0	0	0	...	0
57034	0	0	0	0	0	0	0	0	0	0	0	...	0
57036	0	0	0	0	0	0	0	0	0	0	0	...	0
57037	0	0	0	0	0	0	0	0	0	0	0	...	0

3.

```
# Initialize and fit the KNN model
knn_model = NearestNeighbors(n_neighbors=5, metric='cosine', algorithm='brute')
knn_model.fit(user_item_matrix)
```

NearestNeighbors
NearestNeighbors(algorithm='brute', metric='cosine')

5.

standardized_item_matrix													
user_id	jR4dh3y	Arkhalis-	Artoria-	-Black-Reaper-	CandyConfetti-	Cheesecake-	Dutchy-	EyeOfHorus-	FMA	Froggu-	...	zuo_bai	z
anime_id	1	2.285187	3.432290	2.910231	2.058921	1.559091	2.208980	2.922255	2.294446	2.217382	1.160487
5	2.634866	3.432290	3.281726	2.058921	2.251882	2.208980	2.922255	2.645087	2.555592	2.474626
6	2.634866	2.588581	2.910231	2.420564	2.251882	2.208980	2.512148	2.645087	2.217382	2.474626
7	-0.512244	2.588581	-0.433231	-0.472586	-0.519282	-0.446172	-0.358599	-0.510676	-0.488297	2.474626
8	-0.512244	-0.364401	-0.433231	-0.472586	-0.519282	-0.446172	-0.358599	-0.510676	-0.488297	-0.482186
...
56979	-0.512244	-0.364401	-0.433231	-0.472586	-0.519282	-0.446172	-0.358599	-0.510676	-0.488297	-0.482186
57003	-0.512244	-0.364401	-0.433231	-0.472586	-0.519282	-0.446172	-0.358599	-0.510676	-0.488297	-0.482186
57034	-0.512244	-0.364401	-0.433231	-0.472586	-0.519282	-0.446172	-0.358599	-0.510676	-0.488297	-0.482186
57036	-0.512244	-0.364401	-0.433231	-0.472586	-0.519282	-0.446172	-0.358599	-0.510676	-0.488297	-0.482186
57037	-0.512244	-0.364401	-0.433231	-0.472586	-0.519282	-0.446172	-0.358599	-0.510676	-0.488297	-0.482186

16139 rows × 18145 columns

KNN FINDINGS

- Z-score normalization improved model
- Popular titles exhibit more stable and consistent neighbors (recommendations less sensitive to changes)
- Titles with mid-low popularity demonstrated greater sensitivity
- Model achieved reasonable accuracy in recommendations.
- Note: the model recommends titles that are similarly enjoyed (not necessarily content wise)
 - Any similarity in content suggests that users who enjoyed the targeted anime tended to enjoy other similar shows
 - Ex. if a user enjoys romance anime, nearby neighbors in the model may be in the romance genre

DEMO 1

ARE YOU
GOING TO
KILL ME
RIGHT
HERE?

...WHAT?

....IN A
FLASH.

I AM

FROM TODAY ON

ARE
THE
LIES OF
PIRATES!!!

WE'RE
BROTHERS!!

YOU KNOW,
IF WE SHARE A DRINK
TOGETHER WE'LL BE
BROTHERS!

"AKIRA" (1988)

Type: Movie

Episodes: 1

Status: Finished Airing

Aired: Jul 16, 1988

Producers: Mainichi Broadcasting System, Kodansha, TOHO, Streamline Pictures

Licensors: Funimation, Bandai Entertainment, Geneon Entertainment USA

Studios: Tokyo Movie Shinsha

Source: Manga

Genres: Action, Adventure, Horror, Sci-Fi, Supernatural

Themes: Gore, Military

Demographic: Seinen

Duration: 2 hr. 4 min.

Rating: R+ - Mild Nudity

myanimelist.net/anime/47/Akira?q=akira&cat=anime

MyAnimeList

Anime Manga Community Industry Watch Read Help

Akira

Score: 8.16 (500,949 users) Ranked #409 Popularity #208 Members 830,294

Movie | Tokyo Movie Shinsha

Add to List Select Episodes: 0/1

* Your list is public by default.

Synopsis

Japan, 1988. An explosion caused by a young boy with psychic powers tears through the city of Tokyo and ignites the fuse War III. In order to prevent any further destruction, he is captured and taken into custody, never to be heard from again. In 2019, a restored version of the city known as Neo-Tokyo—an area rife with gang violence and terrorism against the current stands in its place. Here, Shoutarou Kaneda leads "the Capsules," a group of misfits known for riding large, custom motorc constant conflict with their rivals "the Clowns."

During one of these battles, Shoutarou's best friend Tetsuo Shima is caught up in an accident with an esper who finds hims Tokyo after escaping confinement from a government institution. Through this encounter, Tetsuo begins to develop his own as the government seeks to quarantine this latest psychic in a desperate attempt to prevent him from unleashing the could once again bring the city to its knees.

[Written by MAL Rewrite]

Background

Akira won the Silver Scream Award, the audience award of the Amsterdam-based Imagine Film Festival, in 1988. It also received the American Anime Awards' Best Anime Feature award in 2007.

myanimelist.net/anime/47/Akira?q=akira&cat=anime



"AKIRA" (1988)



```
# Choose the target anime  
target_anime_id = 47 # Rep
```

```
# Locate the target anime'  
target_anime_row = standar
```

...

	title	genres	studios	synopsis	media_type
0	Cowboy Bebop	Action, Adult Cast, Award Winning, Sci-Fi, Space	Sunrise	Crime is timeless. By the year 2071, humanity ...	tv
24	Koukaku Kidoutai	Action, Adult Cast, Award Winning, Detective, ...	Production I.G	In the year 2029, Niihama City has become a te...	movie
142	Mononoke Hime	Action, Adventure, Award Winning, Fantasy	Studio Ghibli	When an Emishi village is attacked by a fierce...	movie
409	Perfect Blue	Adult Cast, Avant Garde, Drama, Horror, Psycho...	Madhouse	J-pop idol group CHAM! has spent the last two ...	movie

title
Cowboy Bebop
Koukaku Kidoutai
Mononoke Hime
Perfect Blue



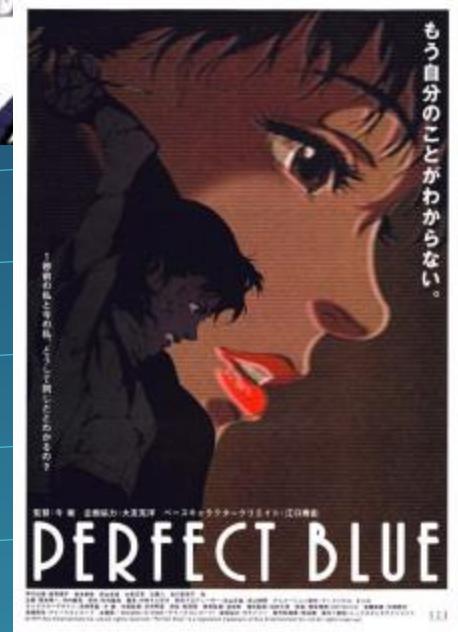
"AKIRA" (1988)



Synopsis:

A secret military project endangers Neo-Tokyo when it turns a biker gang member into a rampaging psychic psychopath who can only be stopped by a teenager, his gang of biker friends and a group of psychics.

RECOMMENDATIONS:



DEMO 2



"GIRLFRIEND, GIRLFRIEND" (2021)

Information

Type: TV

Episodes: 12

Status: Finished Airing

Aired: Jul 3, 2021 to Sep 18, 2021

Premiered: Summer 2021

Broadcast: Saturdays at 02:25 (JST)

Producers: Mainichi Broadcasting

System, Kodansha, Half H.P

Studio, Nikkatsu, NBCUniversal Entertainment

Japan, Forecast Communications, Myrica Music

Licensors: None found, add some

Studios: Tezuka Productions

Source: Manga

Genres: Comedy, Romance

Themes: Harem, School

Demographic: Shounen

Duration: 24 min. per ep.

Rating: PG-13 - Teens 13 or older

myanimelist.net/anime/43969/Kanojo_mo_Kanojo?q=kanojo&cat=anime

MyAnimeList

Hide Ads

Anime Manga Community Industry Watch Read Help

Kanojo mo Kanojo
Girlfriend, Girlfriend

Details Characters & Staff Episodes Videos Stats Reviews Recommendations Interest Stacks News Forum Clubs Pictures

Top > Anime > Kanojo mo Kanojo

Score **6.54**
172,133 users

Ranked #6381 Popularity #642 Members 353,965
Summer 2021 | TV | Tezuka Productions

Add to List Select Episodes: 0/12

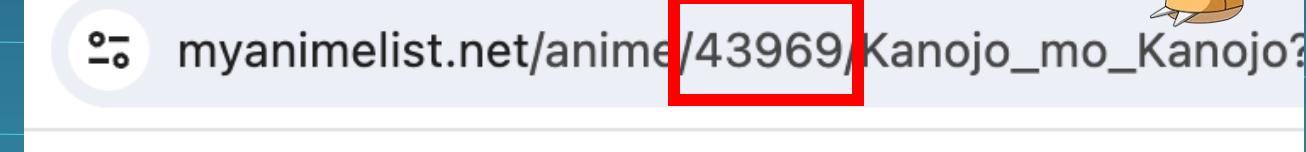
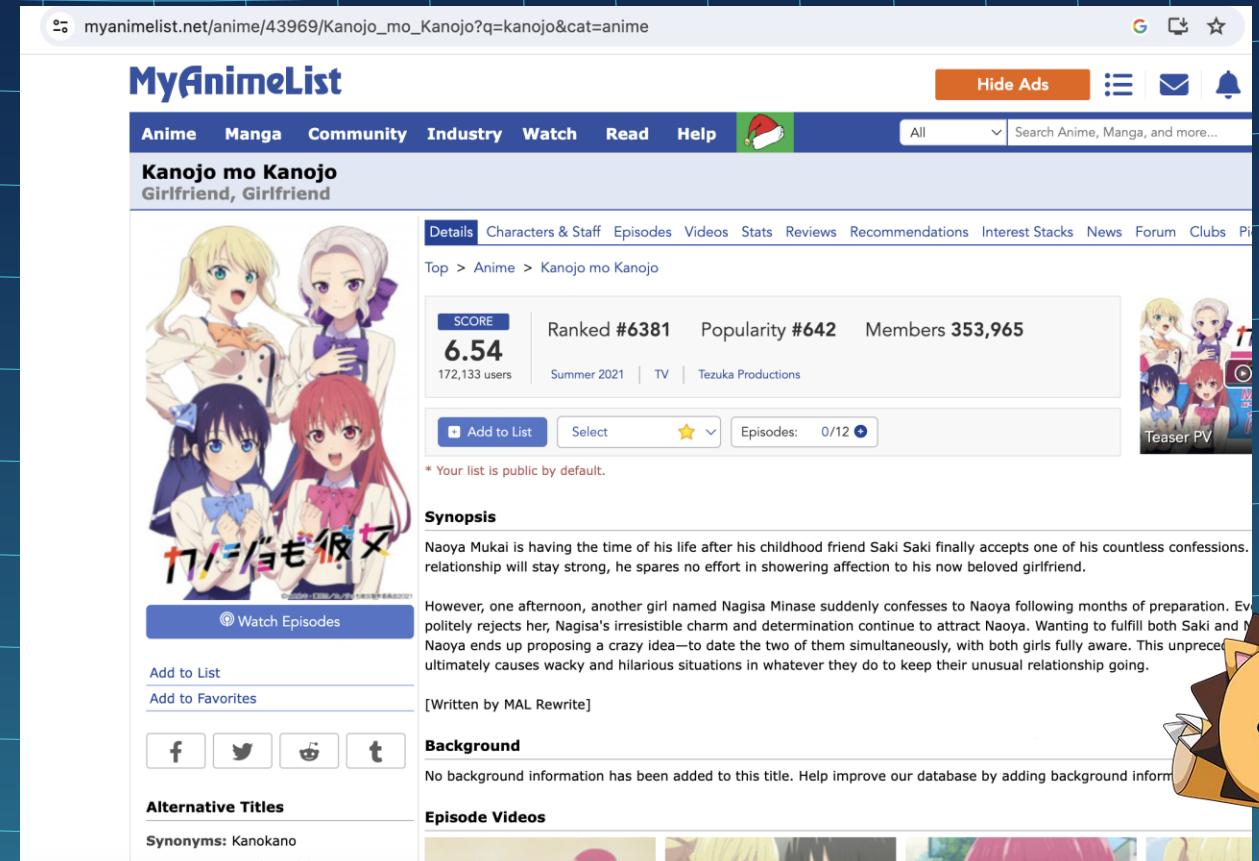
* Your list is public by default.

Synopsis
Naoya Mukai is having the time of his life after his childhood friend Saki Saki finally accepts one of his countless confessions. He's so smitten that his relationship will stay strong, he spares no effort in showering affection to his now beloved girlfriend. However, one afternoon, another girl named Nagisa Minase suddenly confesses to Naoya following months of preparation. Even though Naoya politely rejects her, Nagisa's irresistible charm and determination continue to attract Naoya. Wanting to fulfill both Saki and Nagisa's wishes, Naoya ends up proposing a crazy idea—to date the two of them simultaneously, with both girls fully aware. This unprecedented arrangement ultimately causes wacky and hilarious situations in whatever they do to keep their unusual relationship going.

[Written by MAL Rewrite]

Background
No background information has been added to this title. Help improve our database by adding background information.

Episode Videos



"GIRLFRIEND, GIRLFRIEND" (2021)

```
# Choose the target anime for w  
target_anime_id = 43969 # Repla
```

...	title	genres	studios	synopsis	media_type
14649	5-toubun no Hanayome	Comedy, Harem, Romance, School, Shounen	Tezuka Productions	Fuutarou Uesugi is an ace high school student,...	tv
15376	Kobayashi-san Chi no Maid Dragon S	Fantasy, Slice of Life	Kyoto Animation	As Tooru continues on her quest to become the ...	tv
15707	5-toubun no Hanayome jj	Comedy, Harem, Romance, School, Shounen	Bibury Animation Studios	Through their tutor Fuutarou Uesugi's diligent...	tv
16350	Hige wo Soru. Soshite Joshikousei wo Hirou.	Drama, Romance	Project No.9	When regular salaryman Yoshida wakes up one Su...	tv

title
5-toubun no Hanayome
Kobayashi-san Chi no Maid Dragon S
5-toubun no Hanayome jj
Hige wo Soru. Soshite Joshikousei wo Hirou.



"GIRLFRIEND, GIRLFRIEND" (2021)

RECOMMENDATIONS:



Synopsis:

A honest young man is compelled to accept every girl who asks to be romantically involved with him. The protagonist, Naoya, is a first-year in high school. He asks out Saki, a girl he has been in love with, and she agrees to be his girlfriend. He's currently on cloud nine. But then, a beautiful girl named Nagisa comes up to Naoya one day and tells him that she wants to go out with him. Naoya feels himself wavering because Nagisa is such a good girl, and he comes to a decision. A new kind of romcom from a rather unexpected decision is about to begin.

CONCLUSION



SUMMARY

- *Personalized anime recommendations based on MAL historical user data*
- *Recommendations accuracy is subjective*
- ***Limitations***
 - *Limited dataset*
 - *Model sensitivity which may lead to inaccurate recommendations*
- ***Future Improvements***
 - *Increase the sample size*
 - *Expand training data with additional features*
 - *Invert model to recommend shows based on users with similar tastes*



THANK YOU!

