

# Anime Recommendations Database

Group 4 :

Choy Hei Kiu Max  
Fong Sze Pui Oscar  
Wan Ching Yin  
Lo Tsz Kin



# Introduction



This database is contains information from 73516 user preference data from >12000 animes , each user can add rating after they finish watching animate , therefor , we



# Dataset Introduction

anime.csv

- anime\_id: anime's unique id identifying
- name: Full name of the anime
- genre: comma separated list of genre for the anime
- type: the type of the anime e.g. movie , TV , etc
- episodes : How many episodes in the anime ( if 1 mean it is movie)
- rating : The average rating for the anime
- members : amount of locals who belong to the "group" in this anime.

anime\_top100\_csvformat.xlsx

- anime\_id: anime's unique id identifying
- name: Full name of the anime
- genre: comma separated list of genre for the anime
- type: the type of the anime e.g. movie , TV , etc
- episodes : How many episodes in the anime ( if 1 mean it is movie)



# Problem description-Dataset problem

Since there are more than 10000 rows of data, there are more than

A	B	C	D	E	F	G	H	I	J	K
	name	genre	type	episodes	rating	members				
18099	Recorder to Randomer Mi??Comedy	School	Screen	Slice of Life*	TV	12	6.8	12240		
11491	Recorder to Randomer Do??Comedy	School	Screen	Slice of Life*	TV	13	6.81	29599		
13377	Recorder to Randomer Re??Comedy	School	Screen	Slice of Life*	TV	13	7	17281		
10821	Suite Precure?Movie: Tottomodo! Kokoro ga Tsumaku	Kids	Movie	Shojo*	Movie	1	7.09	2874		
9893	Suite Precure??Action	Fantasy	Movie	Shojo*	TV	48	7.36	7314		
8211	Jewelpet Twinkle??Fantasy	Magic	School	Shojo*	TV	52	7.67	4110		
19993	Kamitani Boy Pikkanbee??Adventure	Comedy	Kids*		TV	54	6.2	64		
4688	Minta Aegchan??Comedy	Ecchi	Romance*	OVA	1	6.01	690			
5938	Oreiga! My Melody Kira??Comedy	Fantasy	Magic*	TV	52	6.56	741			
5520	Oreiga! My Melody Sakkin??Comedy	Fantasy	Magic*	TV	52	7.1	966			
681	Tenbatsu Angel Rubbe??Comedy	Magic	Sci-Fi*	OVA	1	5.3	1192			
30919	Sen no Method: An Shoujo no Kyuujitsu??Comedy	Fantasy	Slice of Life*	Special	1	6.49	4215			
9759	Yanagihime: Eiyuu Retouden??Comedy	Joshi	Slice of Life*	Special	4	6.68	9748			
7062	Hilamari Sketch x ? ??Comedy	School	Slice of Life*	TV	12	8.0	15748			
31080	Arise Happy??Comedy	School	Slice of Life*	TV	12	7.1	33040			
10296	Kachou wa Maid-sama!: Goshujin-sama to Aetaji??Com	School	Shojo*	Special	1	7.1	41241			
28511	Precure All Stars Movie: Hira no Carnival??Action	Comedy	Fantasy	Movie	Music	Shojo*	Movie	1	6.58	980
25681	Kamisama Hajimemashita??Comedy	Demons	Fantasy	Romance	Shojo*	Supernatural	TV	12	8.28	91796
30727	Santa Hirono no Sotokata ??Comedy	Ecchi	Harem	Romance	School*	TV	Unknown	72890		
32261	Kimi no Na wa,	Drama, Romance, School, Supernatural	Movie		1	9.37	2006			
5114	Fullmetal Alchemist: Brotherhood	Action, Adventure, Drama, Fantasy, Magic, Military, Shounen	TV		64	9.26	79365			



# Methodology dataset And problem solving

Because the number of ratings given to each set of anime is very different. As we can see, if we only focus on the scale of the number of ratings, then it must be a large-scale anime that has a better rating. It's not a fair rubric for a small anime. Therefore, it is more fair to select only the top 100 raters for each anime set for comparison.

Since the data files we got are excel file, .csv and raw data file (.dat) and files might include some problematic data (e.g. missing data/unknown value), we need to read 3 different file formats and combine the dataset into one and fix the problematic data.

# Anaylsis

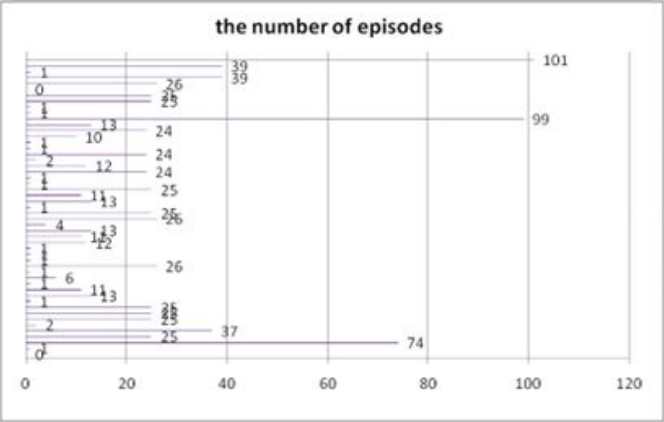




# What are the Top Animes based on user rating?

title	rating
<a href="#">Fullmetal Alchemist (TV)</a>	8.72
<a href="#">Death Note (TV)</a>	8.86

According to the user rating, We can conclude that Fullmetal Alchemist and Death Note are the top anime of anime1 and anime2 respectively. Their user ratings are 8.72 and 8.86 respectively.





# What are the 4 largest(smallest) anime in terms of user rating?

Extreme Anime Rating Values

変数: rating

極端観測値					
最低			最高		
値	name	観測値	値	name	観測値
7.35	Sword Art Online II	24	9.06	Clannad: After Story	40
7.44	Shouwa Genroku Rakugo Shinjuu	89	9.13	Hunter x Hunter (2011)	45
7.46	Highschool of the Dead	25	9.17	Steins;Gate	8
7.48	Deadman Wonderland	42	9.26	Fullmetal Alchemist: Brotherhood	4

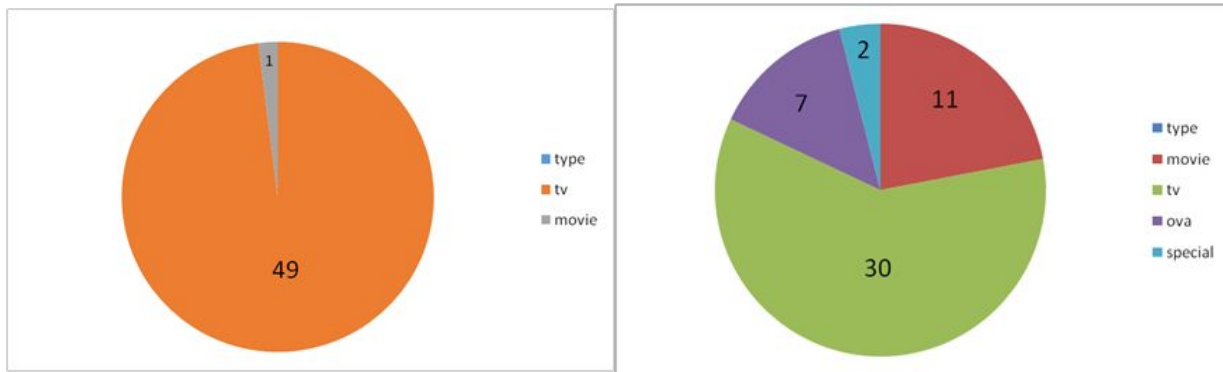
According to the figure 1, it shows 4 of the lowest / 4 of the highest rated anime from the dataset. 4 of the lowest are Sword Art Online 2(7.35), Shouwa Genroku Rakugo Shinjuu (7.44), Highschool of the Dead (7.46) and Deadman Wonderland (7.48). 4 of the highest are Clannad: After story (9.06), Hunter x Hunter (2011) (9.13), Steins;Gate (9.17), Fullmetal Alchemist: Brotherhood (9.26).

In figure 2, the ODS SELECT statement restricts the output to the "ExtremeObs" table for producing extreme observations for variable rating. And we use the UNIVARIATE procedure for producing descriptive statistics results and use NEXTROBS= option to request a specific number of extreme observations. (here we use 4). The ID statement requests that the extreme observations are to be identified using the value of the anime name as well as the observation number.

```
37  
38 ods select ExtremeObs;  
39 proc univariate data=final nextrobs=4;  
40 title 'Extreme Anime Rating Values';  
41 id Name;  
42 var rating;  
43 run;  
44
```



## The proportion of anime types ( TV, Movie, Special, OVA)



Combining anime 1 and anime 2, we can see that the most popular anime types are TV.



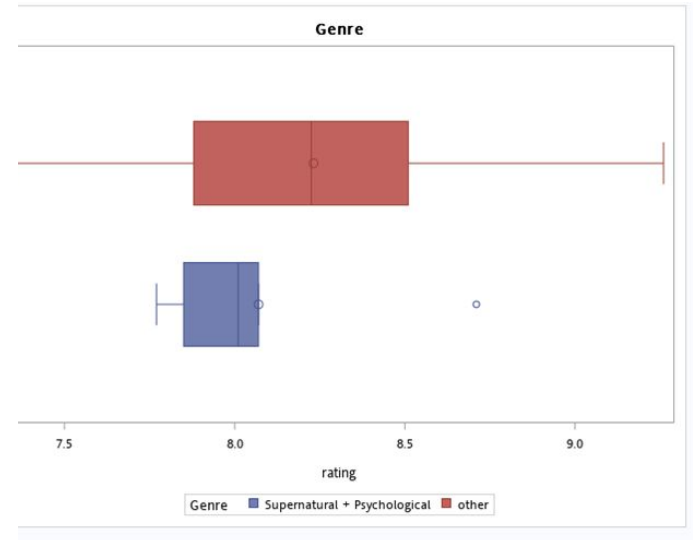
## List the number of anime in a specific genre

According to the result above, the genre Comedy has the highest number of anime followed by Supernatural. It seems that these two genres are more common among the dataset compared to the others genre Thriller and Psychological.

Genre			
Supernatural	Thriller	Comedy	Psychological
35	12	37	15

# Compare rating of the specified anime genre(Psychological+Supernatural) to other combined genre using boxplot

As we can see in the figure, the distribution of ratings for other genres has both the lowest point of rating in the minimum value and highest point of rating in maximum value compared to the 'Supernatural + Psychological' group. The mean and median of both groups is in the range of 8.0 to 8.5. However, there is an outlier for the 'Supernatural+Psychological' group which has a value larger than 8.5.





# Conclusion-Limitation

- Limitation
  - The datasheet is outdated as it was updated before 6 years ago (2016)
  - The statistical results (graph , figure) only describe the ranking from 2016 and cannot represent the year of 2017-2022
  - Some data is inputting with the wrong format



## Conclusion-Recommendation for improvement

- Start with the way you store your data . Keep data fresh.
- A well thought out data model will ensure that there is no redundancy in data
- Using the correct tables and fields can actually store the data.
- Make sure that when the data inputting is the right format



# Reference

1. Anila, Q. (2020). Building A Recommendation System for Anime. Published in Towards Data Science. Retrieved from <https://towardsdatascience.com/building-a-recommendation-system-for-anime-566f86>
2. Anime Recommendations Database. (2016). Published in Kaggle. Retrieved from <https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database>
3. Krishna, A. (2019) Analysis of Anime Recommendations Dataset. Published in ACM VIT. Retrieved from <https://medium.com/@krishnax16/analysis-of-anime-ratings-dataset-54cbde90f23d4acea8>
4. Vinayak, N. (2020). Exploratory Data Analysis on Anime Data. Published in Towards Data Science. Retrieved from <https://towardsdatascience.com/exploratory-data-analysis-on-anime-data-468cc15e13b8>



~~~~Thank you~~~~