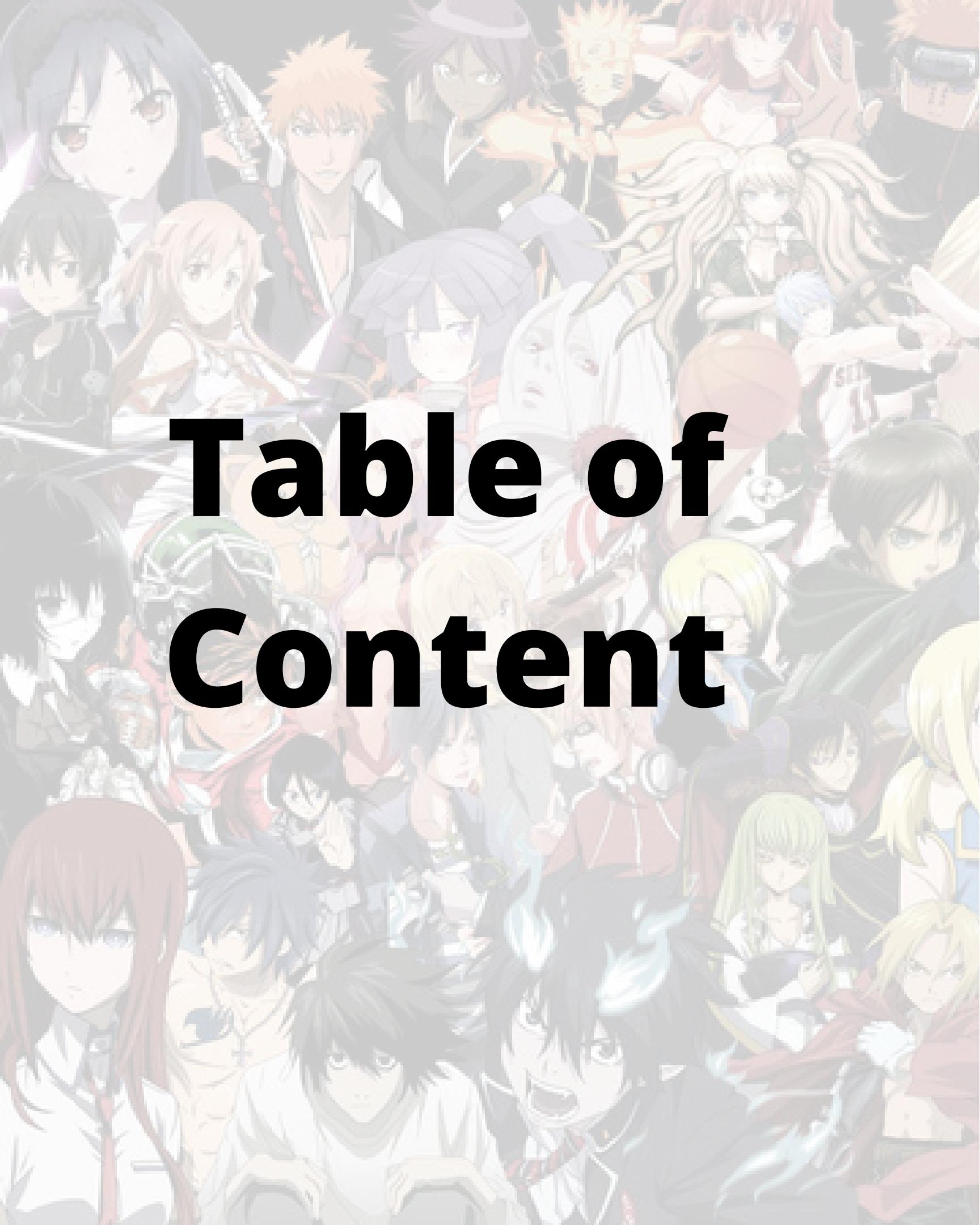


# Anime Popularity Prediction System

SC1015 Mini Project

Group 11

Sim Oi Liang Oi Yeek Sheng



# Table of Content

- 01 Introduction**
- 02 Data Preparation**
- 03 Data Visualisation**
- 04 Machine Learning**
- 05 Conclusion**



# 01 Introduction

# 01 Introduction

HERNAN VALDIVIESO · UPDATED 9 MONTHS AGO

320

New Notebook

Download (693 MB)

## Anime Recommendation Database 2020

Recommendation data from 320.0000 users and 16.000 animes at myanimelist.net

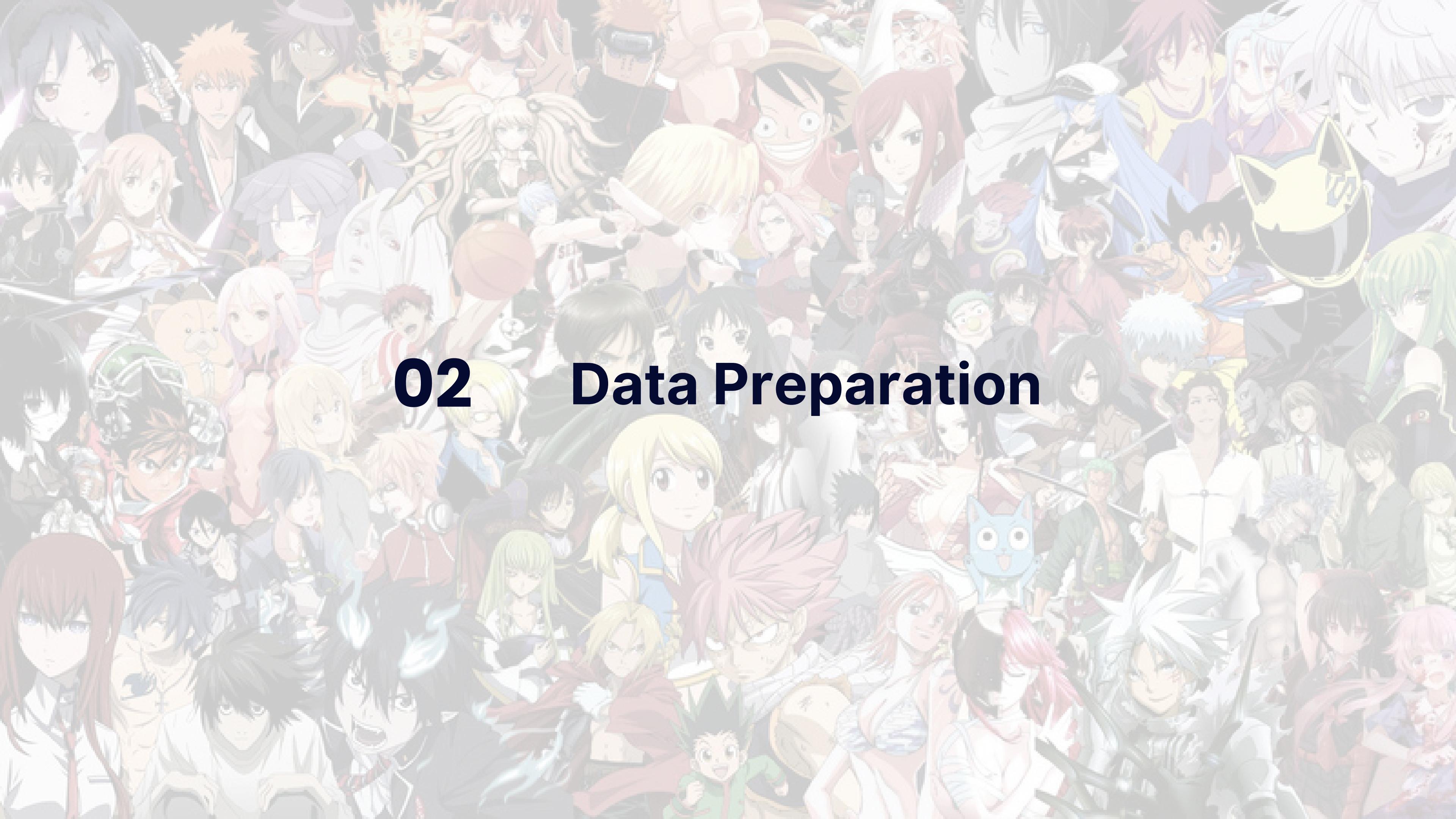
Source : Kaggle Datasets  
Anime Recommendation Database 2020 by  
Hernan Valdivieso



## 01 Introduction

### Motivation

- Predict the popularity of the Anime produced based on synopsis, genres and other more
- Give suggestions on the premiered date based on other Anime
- Provide Anime Studios with the genres that will be most likely be popular



## 02 Data Preparation

## 02 Data Preparation

### Merge csv file

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17562 entries, 0 to 17561
Data columns (total 35 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   MAL_ID      17562 non-null   int64  
 1   Name         17562 non-null   object  
 2   Score        17562 non-null   object  
 3   Genres       17562 non-null   object  
 4   English name 17562 non-null   object  
 5   Japanese name 17562 non-null   object  
 6   Type         17562 non-null   object  
 7   Episodes     17562 non-null   object  
 8   Aired        17562 non-null   object  
 9   Premiered    17562 non-null   object  
 10  Producers    17562 non-null   object  
 11  Licensors    17562 non-null   object  
 12  Studios      17562 non-null   object  
 13  Source        17562 non-null   object  
 14  Duration     17562 non-null   object  
 15  Rating        17562 non-null   object  
 16  Ranked        17562 non-null   object  
 17  Popularity    17562 non-null   int64  
 18  Members       17562 non-null   int64  
 19  Favorites     17562 non-null   int64  
 20  Watching      17562 non-null   int64  
 21  Completed     17562 non-null   int64  
 22  On-Hold       17562 non-null   int64  
 23  Dropped        17562 non-null   int64  
 24  Plan to Watch 17562 non-null   int64  
 25  Score-10      17562 non-null   object  
 26  Score-9       17562 non-null   object  
 27  Score-8       17562 non-null   object  
 28  Score-7       17562 non-null   object  
 29  Score-6       17562 non-null   object  
 30  Score-5       17562 non-null   object  
 31  Score-4       17562 non-null   object  
 32  Score-3       17562 non-null   object  
 33  Score-2       17562 non-null   object  
 34  Score-1       17562 non-null   object  
dtypes: int64(9), object(26)
memory usage: 4.7+ MB
```

anime.csv



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16214 entries, 0 to 16213
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   MAL_ID      16214 non-null   int64  
 1   Name         16214 non-null   object  
 2   Score        16214 non-null   object  
 3   Genres       16214 non-null   object  
 4   synopsis     16206 non-null   object  
dtypes: int64(1), object(4)
memory usage: 633.5+ KB
```

animeSynopsis.csv



```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16214 entries, 0 to 16213
Data columns (total 39 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Anime_id    16214 non-null   int64  
 1   Name         16214 non-null   object  
 2   Score        16214 non-null   object  
 3   Genres       16214 non-null   object  
 4   English name 16214 non-null   object  
 5   Japanese name 16214 non-null   object  
 6   Type         16214 non-null   object  
 7   Episodes     16214 non-null   object  
 8   Aired        16214 non-null   object  
 9   Premiered    16214 non-null   object  
 10  Producers    16214 non-null   object  
 11  Licensors    16214 non-null   object  
 12  Studios      16214 non-null   object  
 13  Source        16214 non-null   object  
 14  Duration     16214 non-null   object  
 15  Rating        16214 non-null   object  
 16  Ranked        16214 non-null   object  
 17  Popularity    16214 non-null   int64  
 18  Members       16214 non-null   int64  
 19  Favorites     16214 non-null   int64  
 20  Watching      16214 non-null   int64  
 21  Completed     16214 non-null   int64  
 22  On-Hold       16214 non-null   int64  
 23  Dropped        16214 non-null   int64  
 24  Plan to Watch 16214 non-null   int64  
 25  Score-10      16214 non-null   object  
 26  Score-9       16214 non-null   object  
 27  Score-8       16214 non-null   object  
 28  Score-7       16214 non-null   object  
 29  Score-6       16214 non-null   object  
 30  Score-5       16214 non-null   object  
 31  Score-4       16214 non-null   object  
 32  Score-3       16214 non-null   object  
 33  Score-2       16214 non-null   object  
 34  Score-1       16214 non-null   object  
 35  Name_y        16214 non-null   object  
 36  Score_y       16214 non-null   object  
 37  Genres_y      16214 non-null   object  
 38  synopsis      16206 non-null   object  
dtypes: int64(9), object(30)
memory usage: 4.9+ MB
```

## 02 Data Preparation

Change the Dtype of data  
object → float

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16214 entries, 0 to 16213
Data columns (total 39 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   Anime_id         16214 non-null   int64  
 1   Name              16214 non-null   object  
 2   Score             16214 non-null   object  
 3   Genres            16214 non-null   object  
 4   English name     16214 non-null   object  
 5   Japanese name    16214 non-null   object  
 6   Type              16214 non-null   object  
 7   Episodes          16214 non-null   object  
 8   Aired              16214 non-null   object  
 9   Premiered         16214 non-null   object  
 10  Producers         16214 non-null   object  
 11  Licensors         16214 non-null   object  
 12  Studios            16214 non-null   object  
 13  Source             16214 non-null   object  
 14  Duration           16214 non-null   object  
 15  Rating             16214 non-null   object  
 16  Ranked             16214 non-null   object  
 17  Popularity         16214 non-null   int64  
 18  Members            16214 non-null   int64  
 19  Favorites          16214 non-null   int64  
 20  Watching            16214 non-null   int64  
 21  Completed          16214 non-null   int64  
 22  On-Hold             16214 non-null   int64  
 23  Dropped             16214 non-null   int64  
 24  Plan to Watch      16214 non-null   int64  
 25  Score-10            16214 non-null   object  
 26  Score-9              16214 non-null   object  
 27  Score-8              16214 non-null   object  
 28  Score-7              16214 non-null   object  
 29  Score-6              16214 non-null   object  
 30  Score-5              16214 non-null   object  
 31  Score-4              16214 non-null   object  
 32  Score-3              16214 non-null   object  
 33  Score-2              16214 non-null   object  
 34  Score-1              16214 non-null   object  
 35  Name_y              16214 non-null   object  
 36  Score_y              16214 non-null   object  
 37  Genres_y            16214 non-null   object  
 38  synopsis            16206 non-null   object  
dtypes: int64(9), object(30)
memory usage: 4.9+ MB
```

## 02 Data Preparation

### Delete all unnecessary data

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16214 entries, 0 to 16213
Data columns (total 39 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Anime_id    16214 non-null   int64  
 1   Name        16214 non-null   object  
 2   Score       16214 non-null   object  
 3   Genres      16214 non-null   object  
 4   English name 16214 non-null   object  
 5   Japanese name 16214 non-null   object  
 6   Type        16214 non-null   object  
 7   Episodes     16214 non-null   object  
 8   Aired        16214 non-null   object  
 9   Premiered    16214 non-null   object  
 10  Producers    16214 non-null   object  
 11  Licensors    16214 non-null   object  
 12  Studios      16214 non-null   object  
 13  Source       16214 non-null   object  
 14  Duration     16214 non-null   object  
 15  Rating       16214 non-null   object  
 16  Ranked       16214 non-null   object  
 17  Popularity   16214 non-null   int64  
 18  Members      16214 non-null   int64  
 19  Favorites    16214 non-null   int64  
 20  Watching     16214 non-null   int64  
 21  Completed    16214 non-null   int64  
 22  On-Hold      16214 non-null   int64  
 23  Dropped      16214 non-null   int64  
 24  Plan to Watch 16214 non-null   int64  
 25  Score-10     16214 non-null   object  
 26  Score-9      16214 non-null   object  
 27  Score-8      16214 non-null   object  
 28  Score-7      16214 non-null   object  
 29  Score-6      16214 non-null   object  
 30  Score-5      16214 non-null   object  
 31  Score-4      16214 non-null   object  
 32  Score-3      16214 non-null   object  
 33  Score-2      16214 non-null   object  
 34  Score-1      16214 non-null   object  
 35  Name_y       16214 non-null   object  
 36  Score_y      16214 non-null   object  
 37  Genres_y     16214 non-null   object  
 38  synopsis      16206 non-null   object  
dtypes: int64(9), object(30)
memory usage: 4.9+ MB
```



```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16214 entries, 0 to 16213
Data columns (total 22 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Anime_id    16214 non-null   int64  
 1   Name        16214 non-null   object  
 2   Score       16214 non-null   object  
 3   Genres      16214 non-null   object  
 4   Premiered    16214 non-null   object  
 5   Studios      16214 non-null   object  
 6   Source       16214 non-null   object  
 7   Watching     16214 non-null   int64  
 8   Completed    16214 non-null   int64  
 9   On-Hold      16214 non-null   int64  
 10  Plan to Watch 16214 non-null   int64  
 11  Score-10     16214 non-null   object  
 12  Score-9      16214 non-null   object  
 13  Score-8      16214 non-null   object  
 14  Score-7      16214 non-null   object  
 15  Score-6      16214 non-null   object  
 16  Score-5      16214 non-null   object  
 17  Score-4      16214 non-null   object  
 18  Score-3      16214 non-null   object  
 19  Score-2      16214 non-null   object  
 20  Score-1      16214 non-null   object  
 21  synopsis      16206 non-null   object  
dtypes: int64(5), object(17)
memory usage: 2.8+ MB
```

## 02 Data Preparation

### Replace "Unknown" with appropriate value

MAL_ID	Name	Score	Genres	Premiered	Studios	Source	Watched	Completed	On-Hold	Plan to Watch	Score-1	Score-2	Score-3	Score-4	Score-5	Score-6	Score-7	Score-8	Score-9	Score-10	Score-11
29830	Tamayura: Sotsug	7.8	Comedy, Drama	Unknown	TYO Animations	Original	239	3059	143	3752	298.0	441.0	743.0	517.0	164.0	87.0	20.0	12.0	2.0	9.0	
29831	Tamayura: Sotsug	8.1	Comedy, Drama	Unknown	TYO Animations	Original	333	3041	141	4455	492.0	539.0	637.0	384.0	142.0	71.0	16.0	16.0	6.0	9.0	
29832	Panpaka Panta M	Unkn	Adventure, Kids	Unknown	DLE	Original	11	67	6	74	7.0	5.0	1.0	5.0	5.0	11.0	3.0	1.0	5.0	4.0	
29833	Mochi Mochi no K	Unkn	Kids	Unknown	Unknown	Other	9	118	4	73	4.0	1.0	7.0	7.0	11.0	15.0	11.0	6.0	6.0	7.0	
29834	Moshidora Recap	5.94	Drama, Sports	Unknown	Production I.G	Unknown	30	197	8	237	4.0	6.0	7.0	15.0	24.0	42.0	13.0	2.0	3.0	4.0	
29835	47 Todoufukuen Sp	Unkn	Kids, Comedy	Unknown	Pollyanna Graphics	Unknown	9	108	5	85	8.0	Unkn	3.0	6.0	19.0	19.0	4.0	1.0	2.0	15.0	
29836	Akuei to Gacching	Unkn	Kids, Comedy	Fall 2004	Unknown	Unknown	9	93	5	76	6.0	Unkn	6.0	3.0	6.0	9.0	11.0	6.0	1.0	12.0	
29837	Yakushiji Ryousuke	Unkn	Police, Super	Unknown	Doga Kobo	Light novel	35	174	13	261	5.0	3.0	10.0	15.0	27.0	19.0	5.0	4.0	5.0	9.0	

Replace with the average score given by users

Delete row with unknown Premiered, Studios and Source

Replace the unknown in Score-n with 0

## 02 Data Preparation

### Calculate the Total Voters

	Anime_id	Name	Score-10	Score-9	Score-8	Score-7	Score-6	Score-5	Score-4	Score-3	Score-2	Score-1
0	1	Cowboy Bebop	229170.0	182126.0	131625.0	62330.0	20688.0	8904.0	3184.0	1357.0	741.0	1580.0
1	5	Cowboy Bebop: Tengoku no Tobira	30043.0	49201.0	49505.0	22632.0	5805.0	1877.0	577.0	221.0	109.0	379.0
2	6	Trigun	50229.0	75651.0	86142.0	49432.0	15376.0	5838.0	1965.0	664.0	316.0	533.0
3	7	Witch Hunter Robin	2182.0	4806.0	10128.0	11618.0	5709.0	2920.0	1083.0	353.0	164.0	131.0
4	8	Bouken Ou Beet	312.0	529.0	1242.0	1713.0	1068.0	634.0	265.0	83.0	50.0	27.0

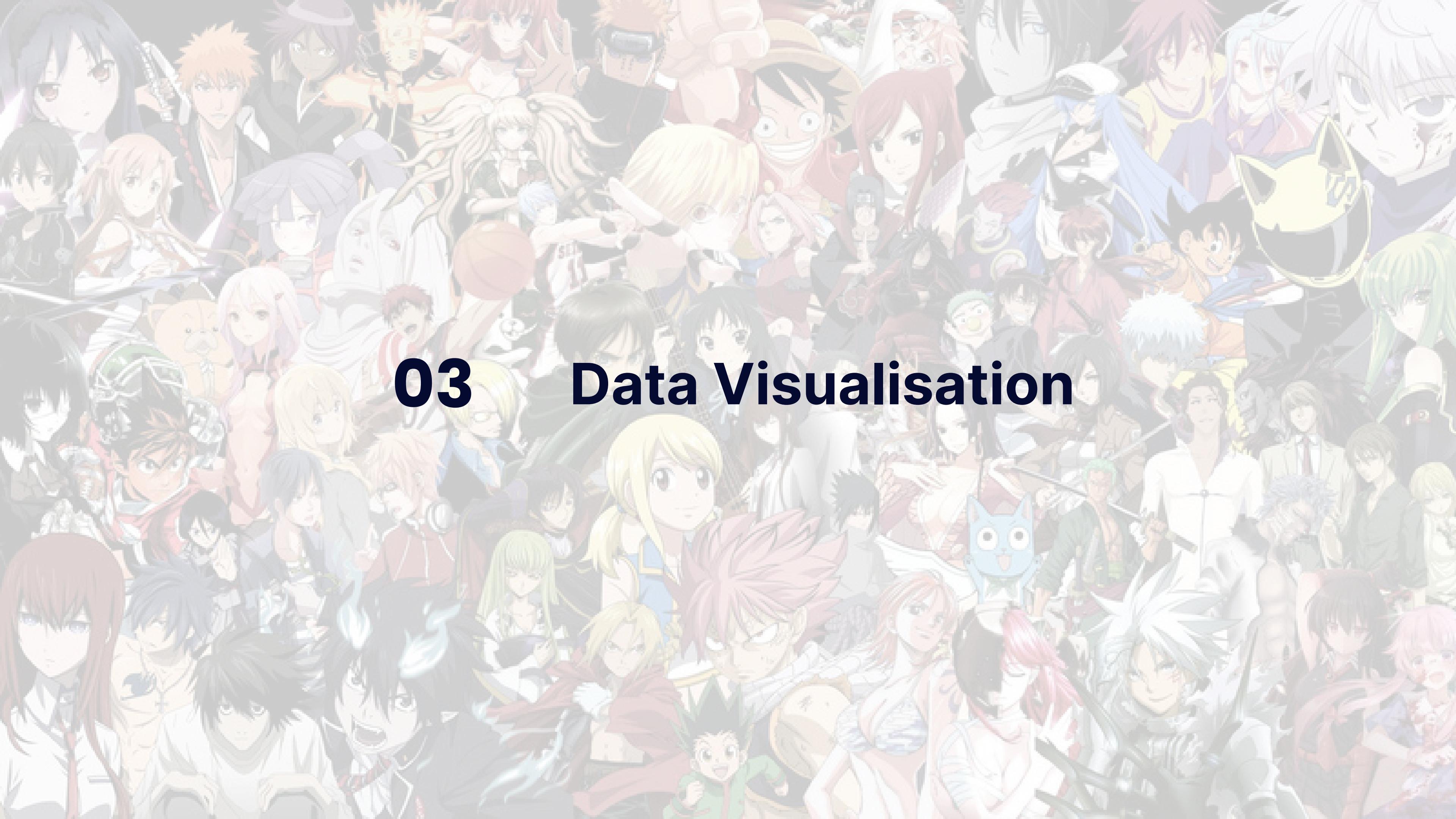
Total Voters = Sum of Score-10 to Score-1

## 02 Data Preparation

### Redefine Popularity

	Anime_id	Name	Watching	Completed	On-Hold	Plan to Watch
0	1	Cowboy Bebop	105808	718161	71513	329800
1	5	Cowboy Bebop: Tengoku no Tobira	4143	208333	1935	57964
2	6	Trigun	29113	343492	25465	146918
3	7	Witch Hunter Robin	4300	46165	5121	33719
4	8	Bouken Ou Beet	642	7314	766	3394

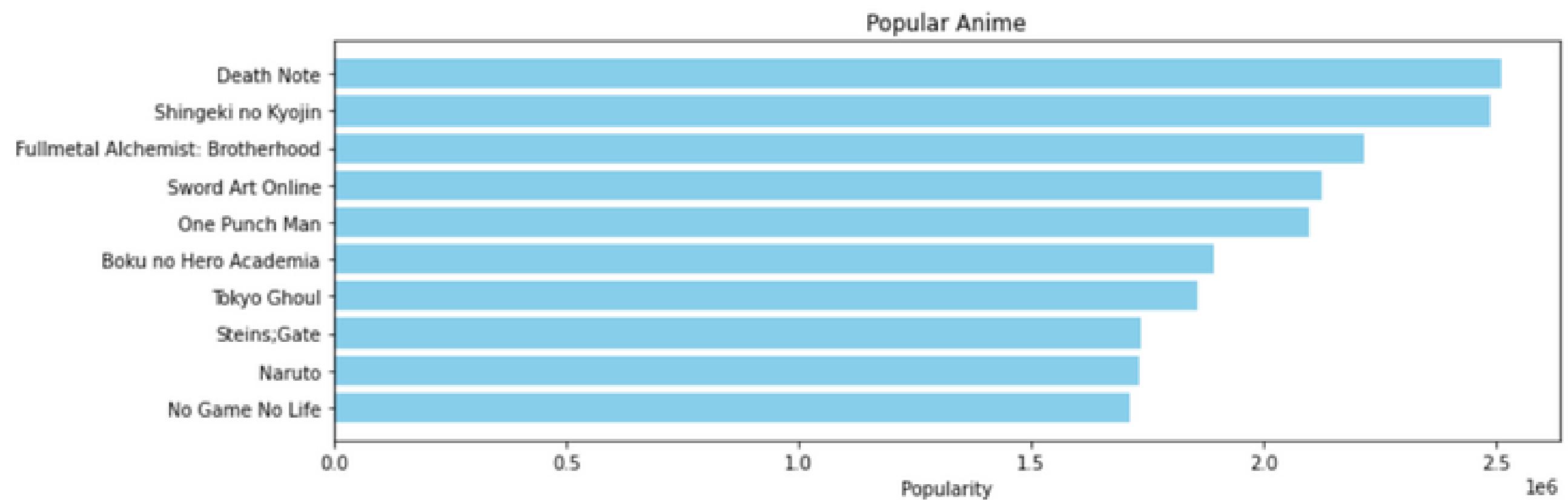
$\text{Popularity} = \text{Watching} + \text{Completed} + \text{On-Hold} + \text{Plan to Watch}$



# 03 Data Visualisation

# 03 Data Visualisation

## Top 10 Popular Animes



# 03 Data Visualisation

## Top 10 Popular Animes



**Death Note**



**Shingeki No  
Kyojin**



**Fullmetal  
Alchemist:  
Brotherhood**



**Sword Art Online**



**One Punch Man**



**Boku No Hero  
Academia**



**Tokyo Ghoul**



**Steins;Gate**



NO GAME NO LIFE

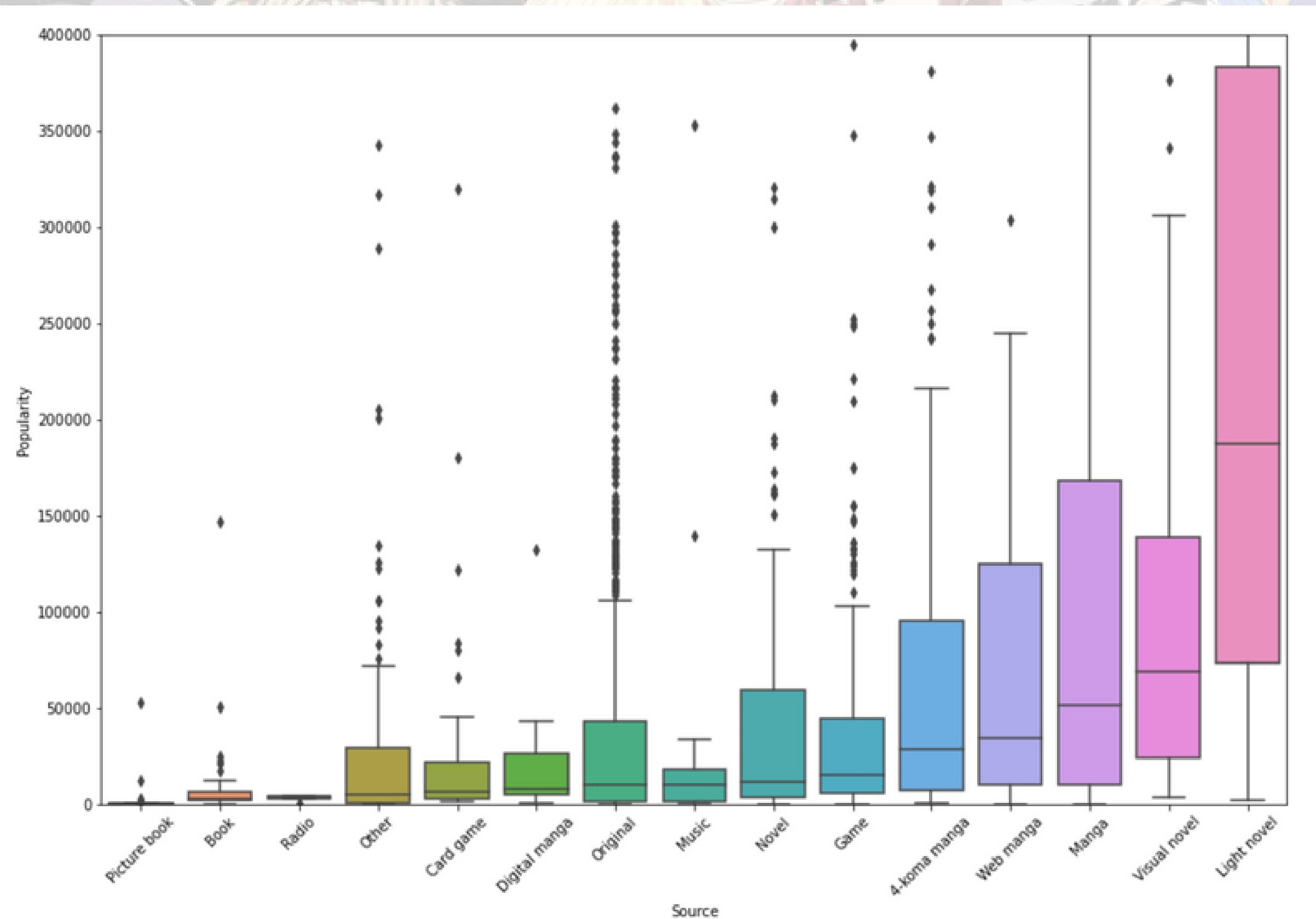
**No Game No Life**



**Naruto**

# 03 Data Visualisation

## Relationship between Source and Popularity



# 03 Data Visualisation

## Light Novel Animes



**Sword Art Online**



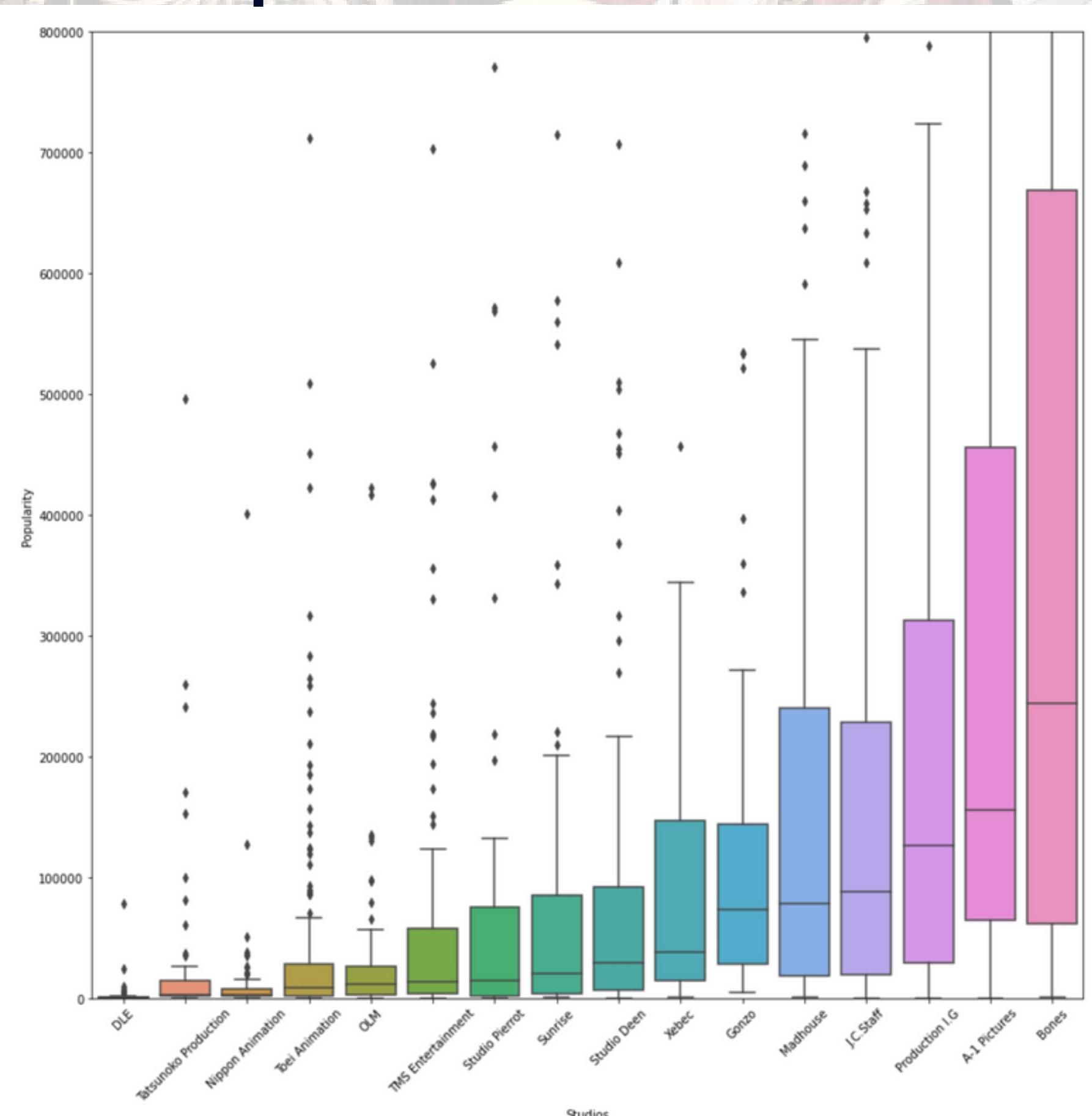
**Toradora!**



**No Game No Life**

# 03 Data Visualisation

## Relationship between Studios and Popularity



## 03 Data Visualisation

Most Popular Studio



bones



**Fullmetal  
Alchemist:  
Brotherhood**



**Boku No Hero  
Academia**



**NORAGAMI**

**Noragami**

# 03 Data Visualisation

## Underrated and Overrated Animes

	Name	Underrated	Popularity	Score		Name	Overrated	Popularity	Score
	Pui Pui Molcar	2783	3310	8.33		School Days	2765	498446	5.68
	Tentai Senshi Sunred 2nd Season	2263	6642	7.90		Boruto: Naruto Next Generations	2649	416055	5.81
	Hello! Lady Lynn	2216	3055	7.57		Diabolik Lovers	2528	224406	5.30
Mobile Suit Gundam: The Origin - Advent of the...		2147	10840	8.09		Tokyo Ghoul:re 2nd Season	2394	446952	6.33
	Igano Kabamaru	2122	8385	7.87		Yosuga no Sora: In Solitude, Where We Are Leas...	2351	340982	6.26
	Hanada Shounen-shi	2082	11242	8.01		Isekai wa Smartphone to Tomo ni.	2350	322466	6.24
	BanG Dream! Garupa☆Pico: Oomori	2065	5612	7.63		Isekai Cheat Magician	2334	165264	5.38
	Trapp Ikka Monogatari	2056	5763	7.63		Ousama Game The Animation	2331	150411	5.00
Konnichiwa Anne: Before Green Gables		2054	5137	7.59		Big Order (TV)	2308	157337	5.40
	Perrine Monogatari	2047	3947	7.52		Mayoiga	2306	166871	5.53

# 03 Data Visualisation

## Most Underrated Anime



**Pui Pui Molcar**

Popularity : 3310

Score : 8.33

## Most Overrated Anime

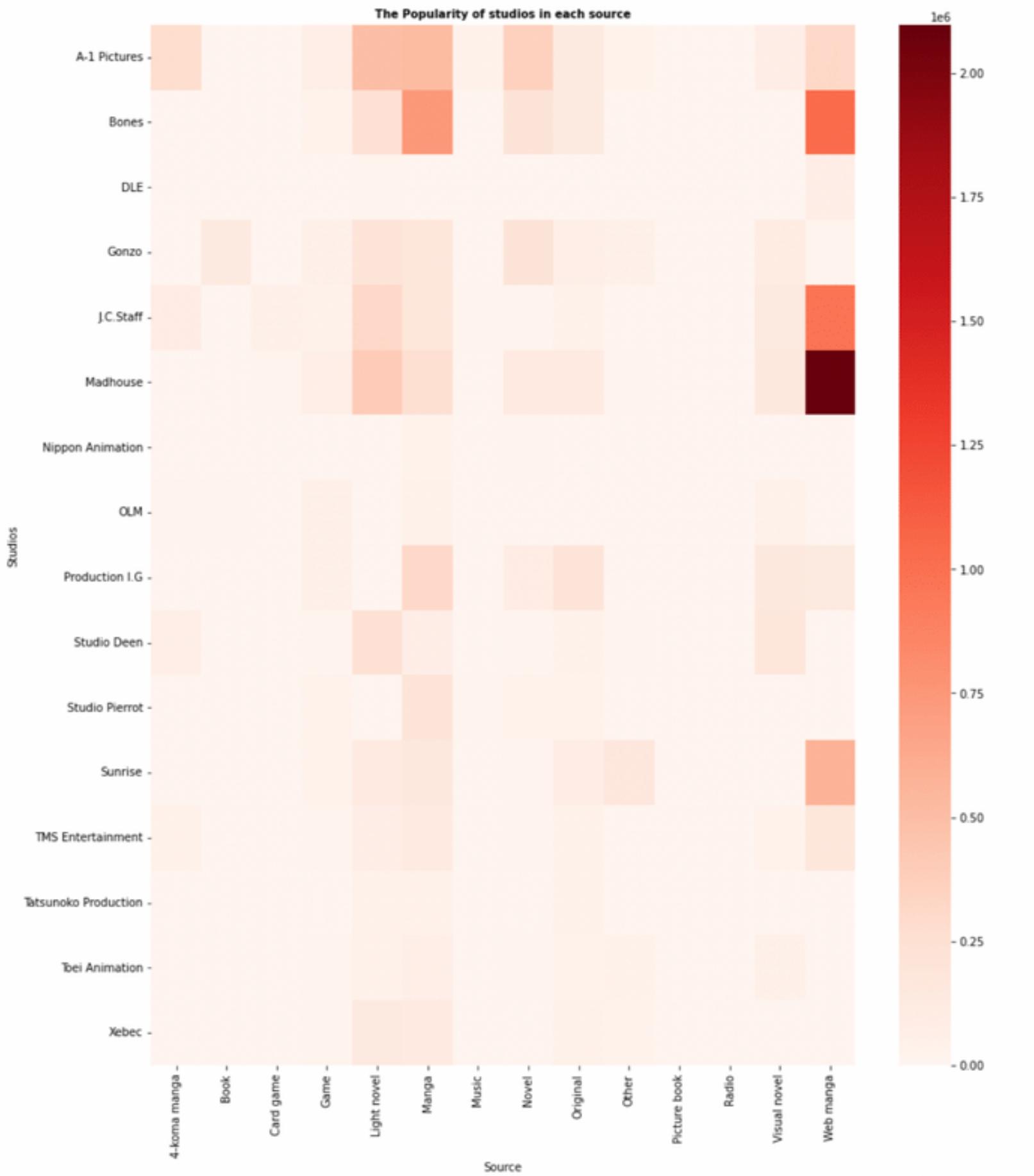


**School Days**

Popularity : 498446

Score : 5.68

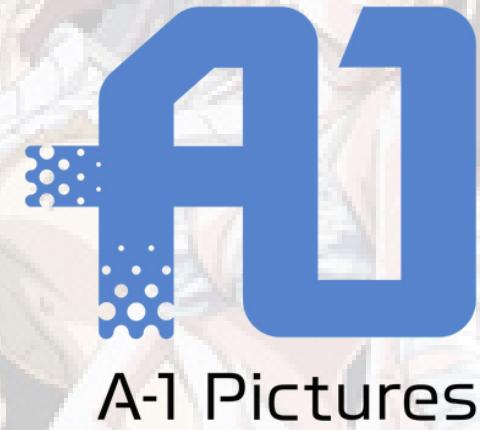
# 03 Data Visualisation



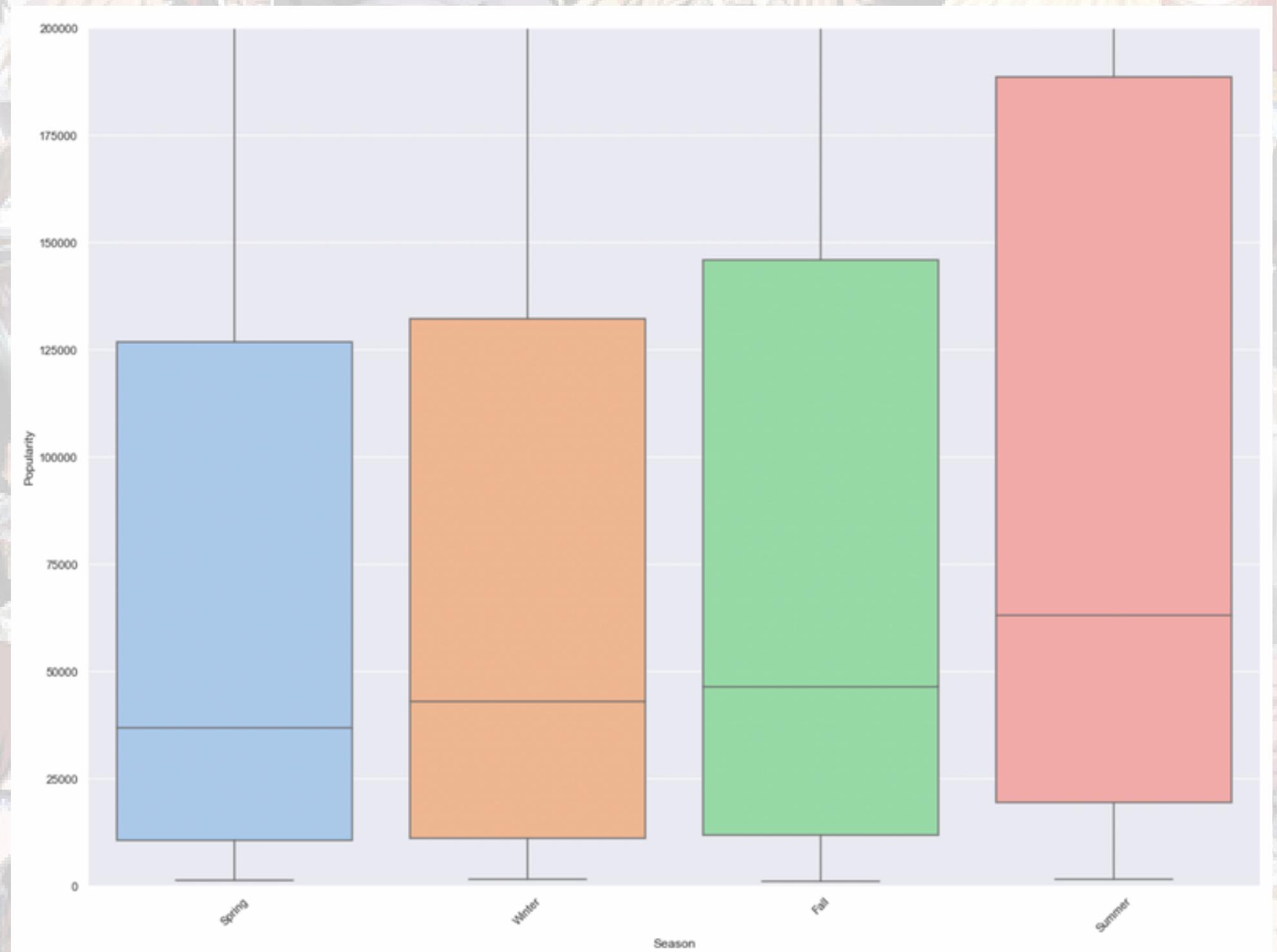
## Top Studios for Web Manga



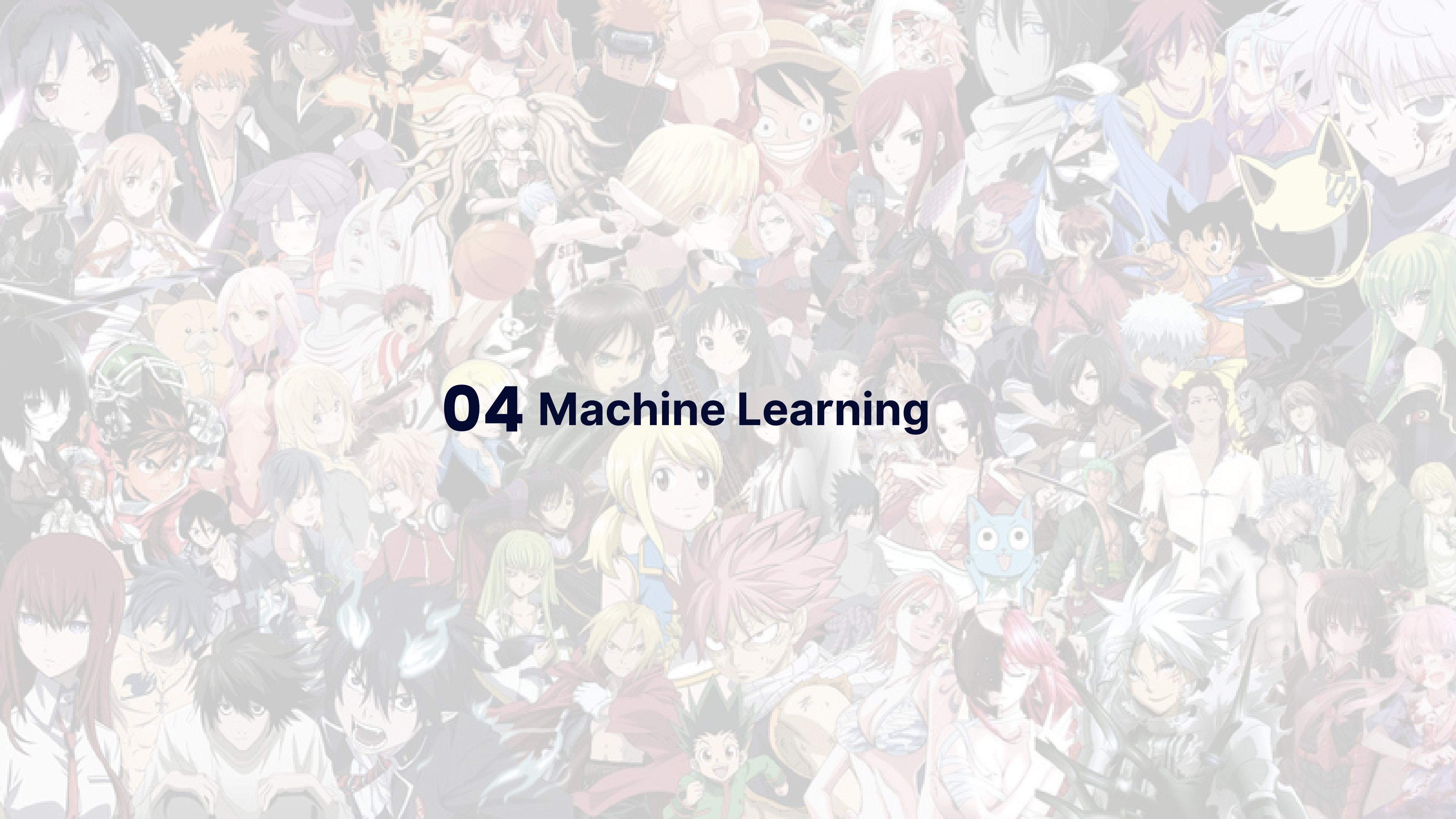
## Top Studios for Manga



# 03 Data Visualisation



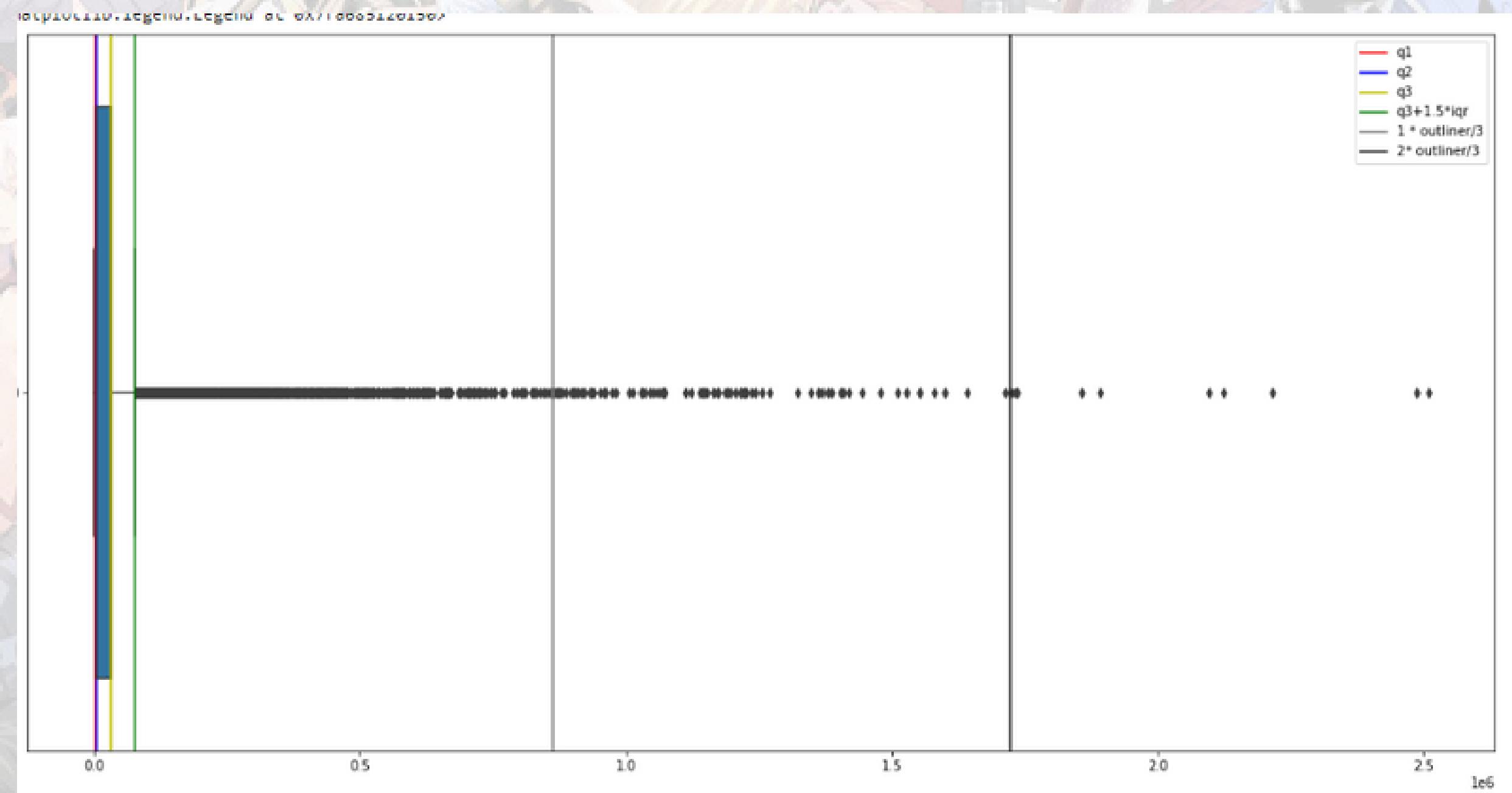
**Anime which premiered in Summer have a higher popularity**

A dense, overlapping collage of numerous anime and manga characters. The characters are drawn in a variety of styles and colors, creating a busy and diverse visual background.

## 04 Machine Learning

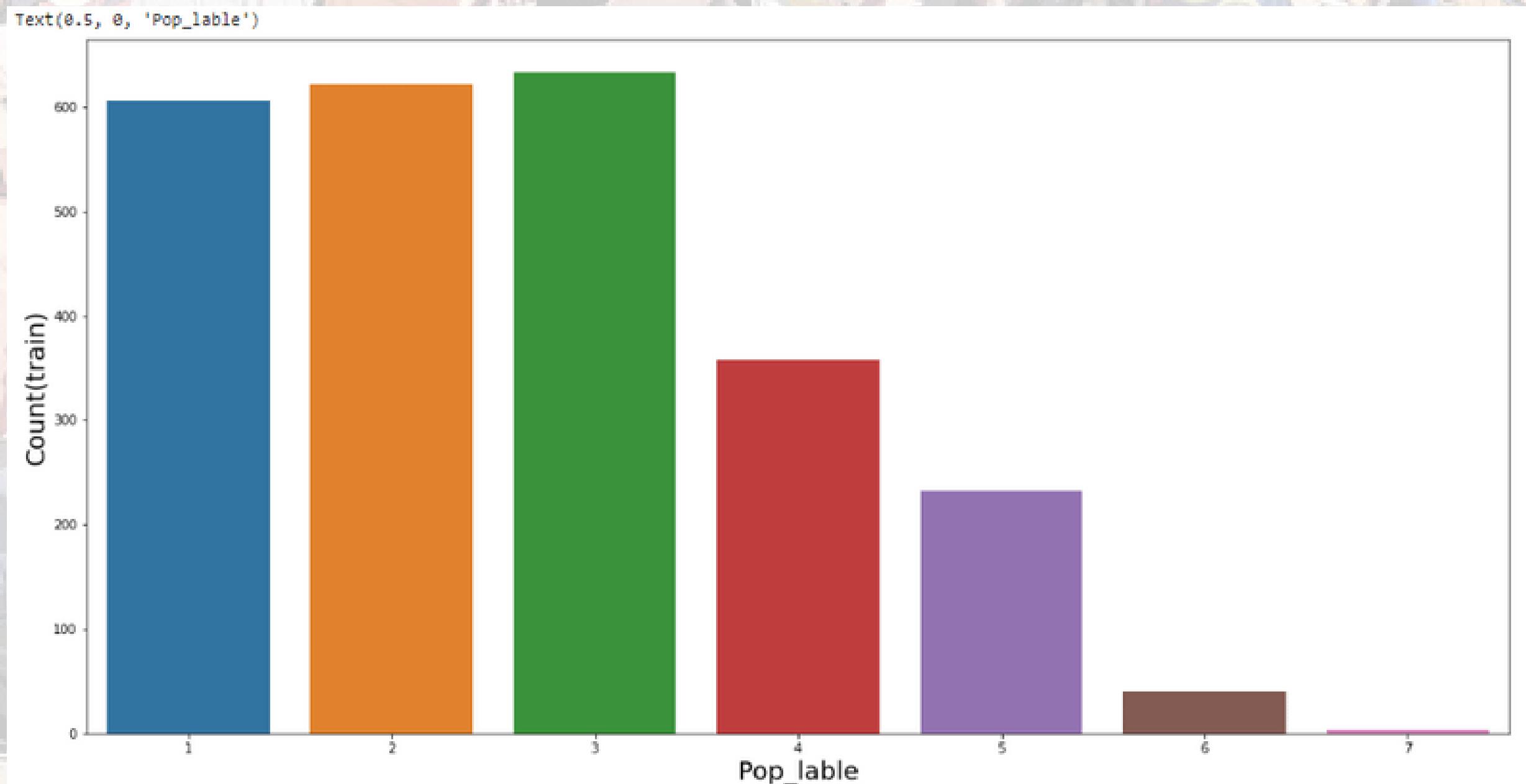
## 04 Machine Learning

### Separate Data into Segments based on Popularity

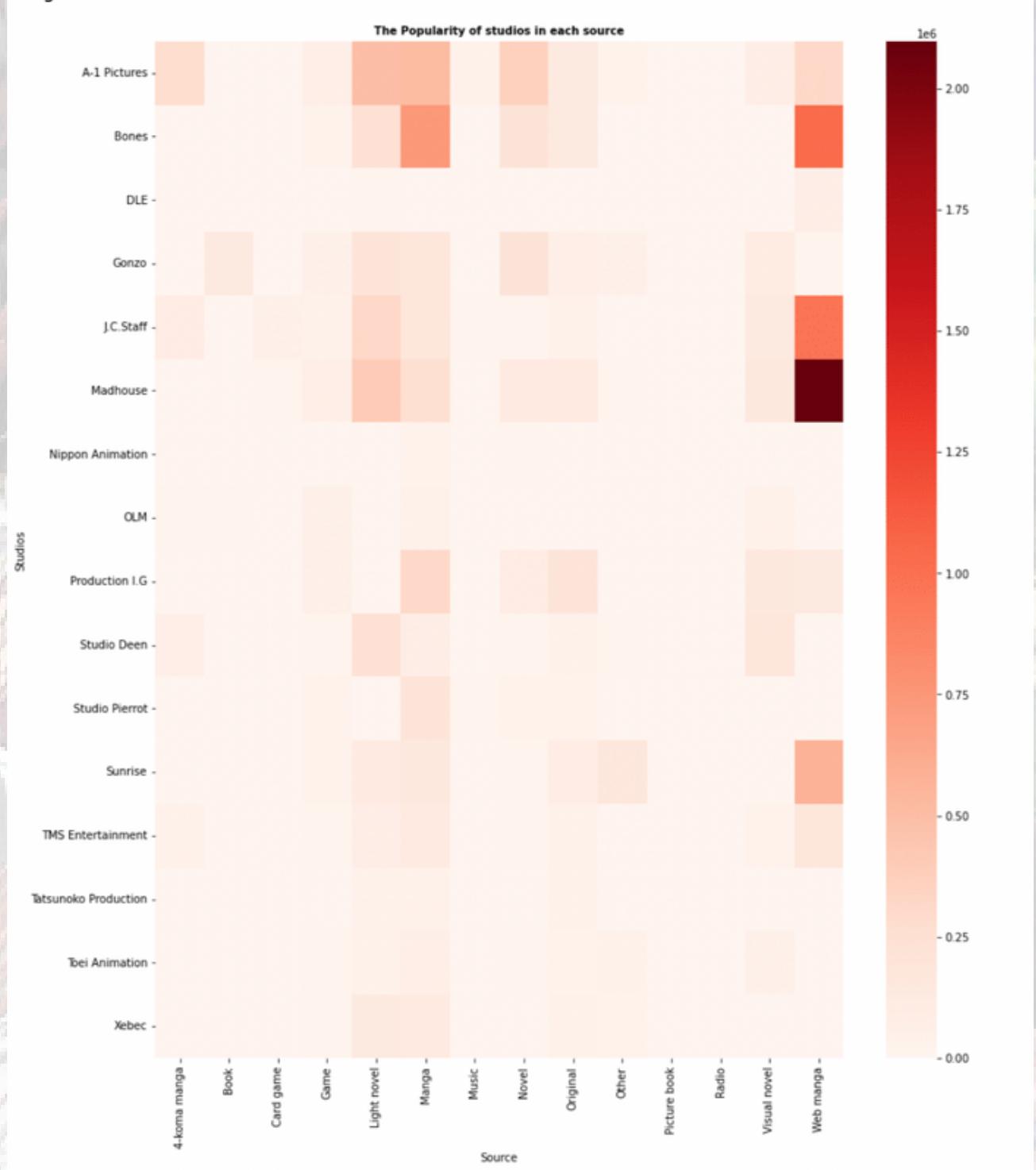
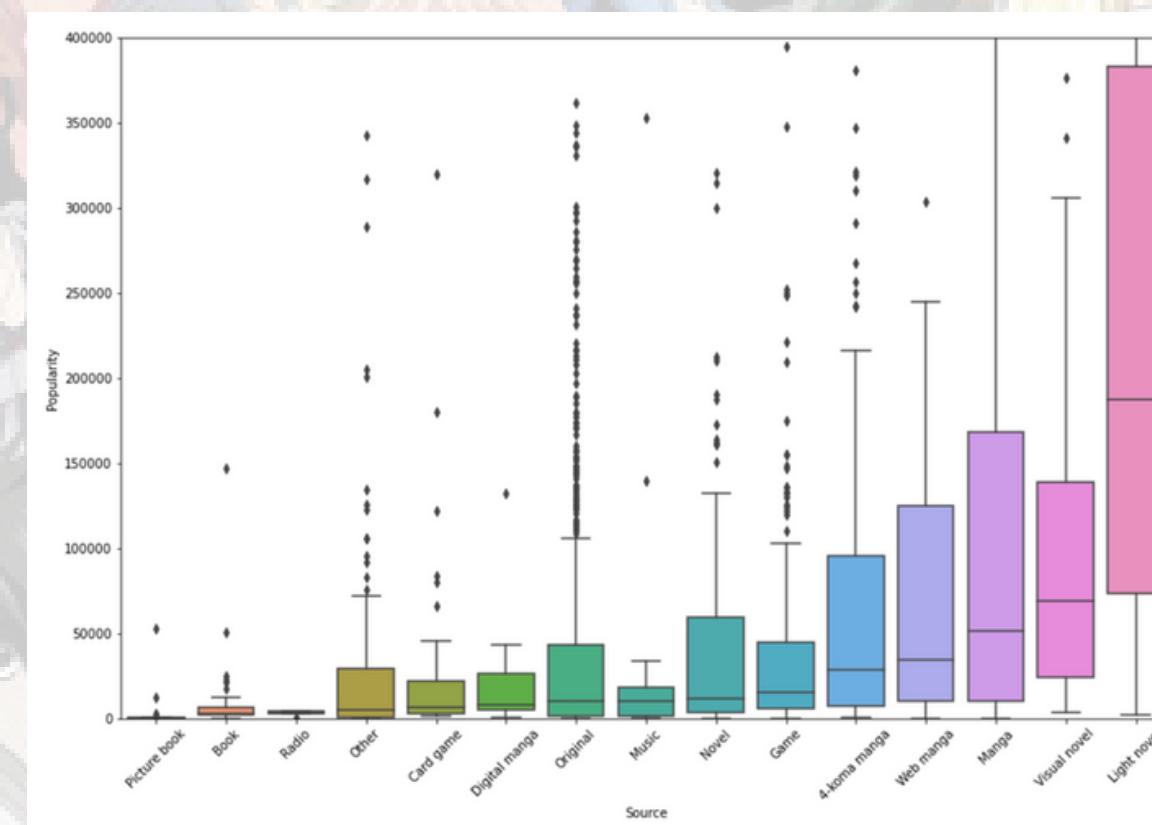
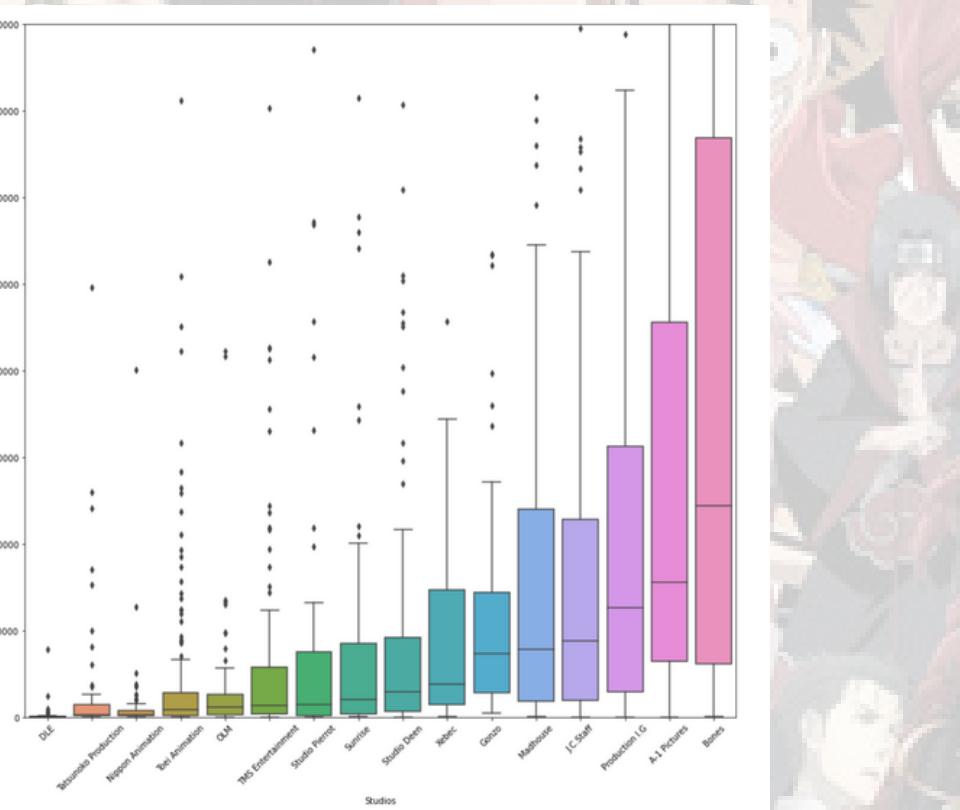
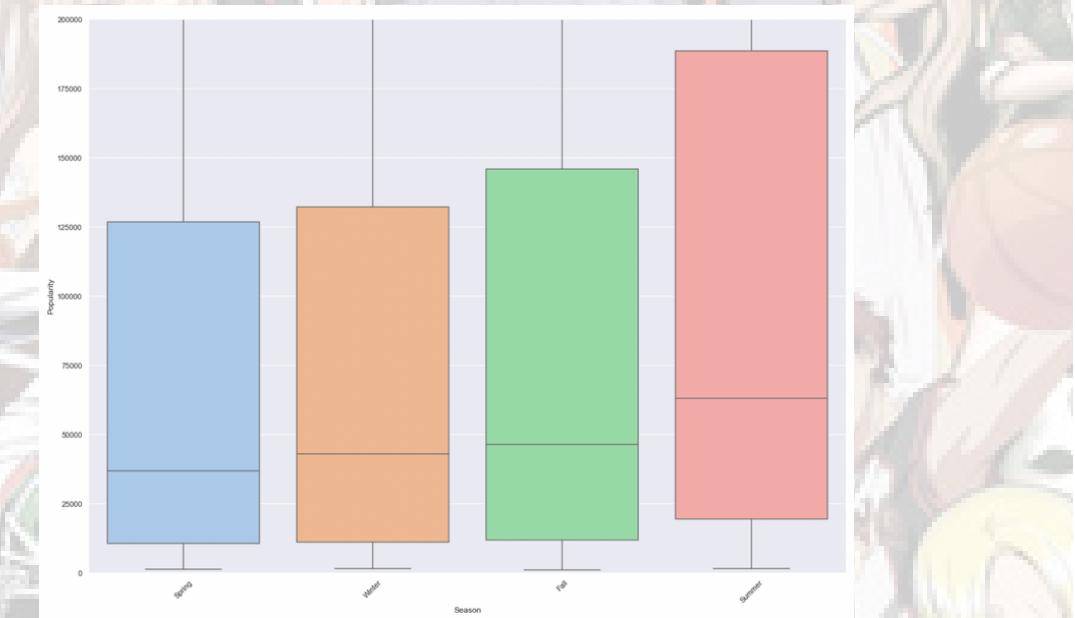


## 04 Machine Learning

### Number of Anime in Each Segment



# 04 Machine Learning



## 04 Machine Learning

### One Hot Encoding

<b>id</b>	<b>color</b>
1	red
2	blue
3	green
4	blue

One Hot Encoding

<b>id</b>	<b>color_red</b>	<b>color_blue</b>	<b>color_green</b>
1	1	0	0
2	0	1	0
3	0	0	1
4	0	1	0

## 04 Machine Learning

### Removing 'Stop' Words and tokenize synopsis

#### Synopsis

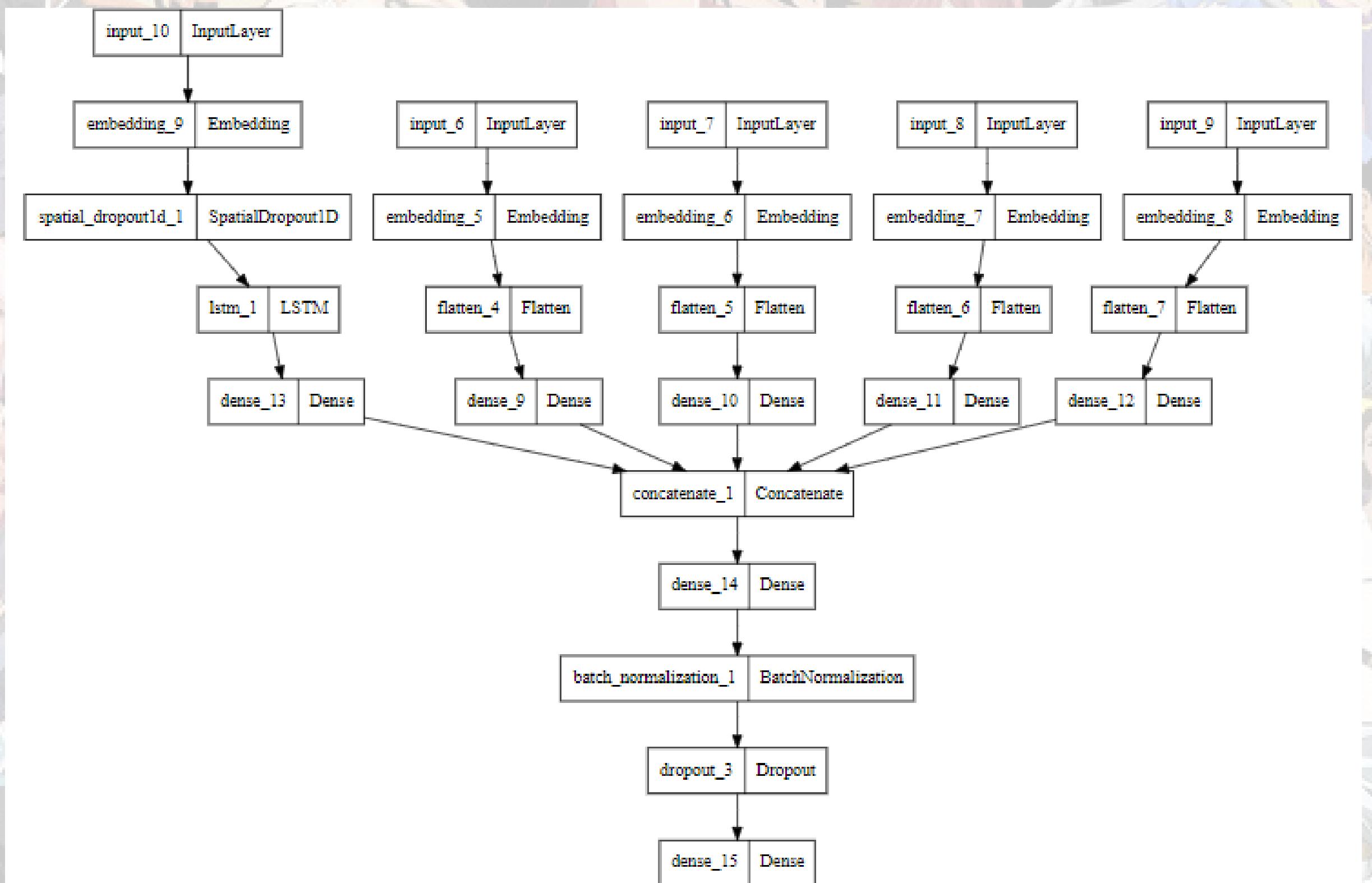
[Edit](#)

Brutal murders, petty thefts, and senseless violence pollute the human world. In contrast, the realm of death gods is a humdrum, unchanging gambling den. The ingenious 17-year-old Japanese student Light Yagami and sadistic god of death Ryuk share one belief: their worlds are rotten.

For his own amusement, Ryuk drops his "Death Note" into the human world. Light stumbles upon it, deeming the first of its rules ridiculous: the human whose name is written in this note shall die. However, the temptation is too great, and Light experiments by writing a felon's name, which disturbingly enacts his first murder.

Aware of the terrifying godlike power that has fallen into his hands, Light—under the alias "Kira"—follows his wicked sense of justice with the ultimate goal of cleansing the world of all evil-doers. The meticulous mastermind detective L is already on his trail, but as Light's brilliance rivals L's, the grand chase for Kira turns into an intense battle of wits that can only end when one of them is dead.

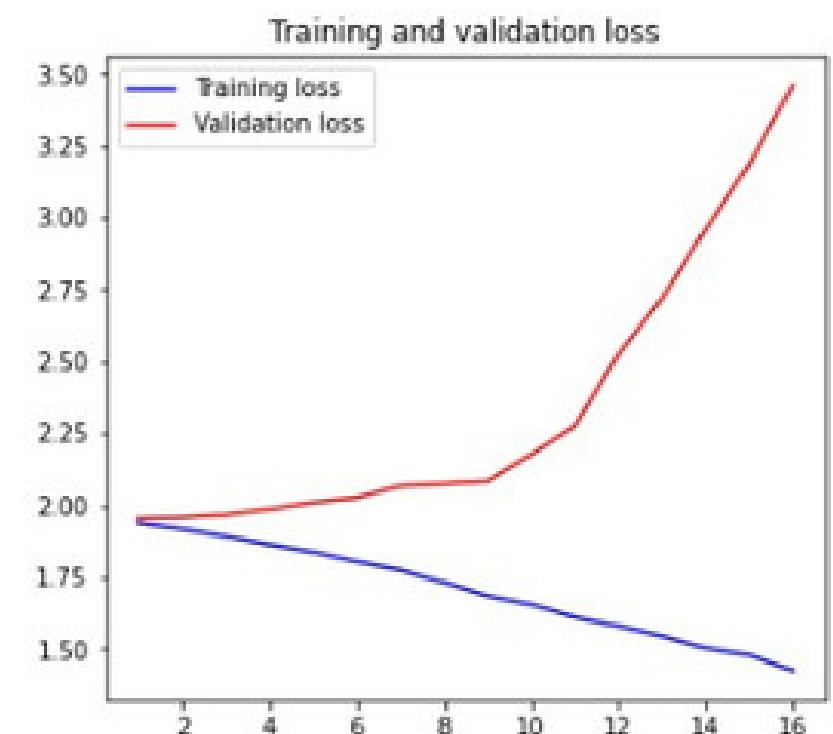
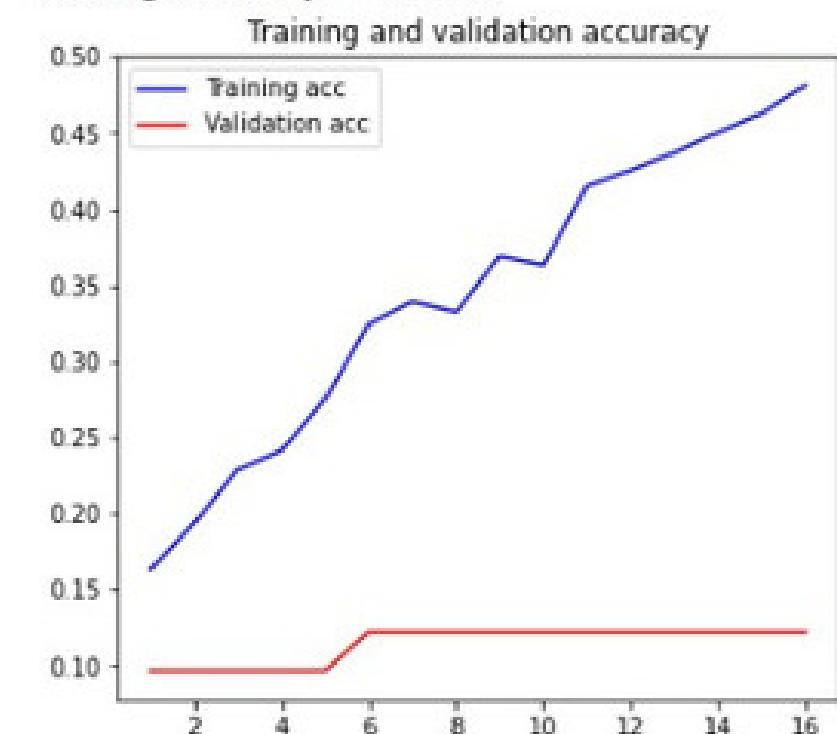
# 04 Machine Learning



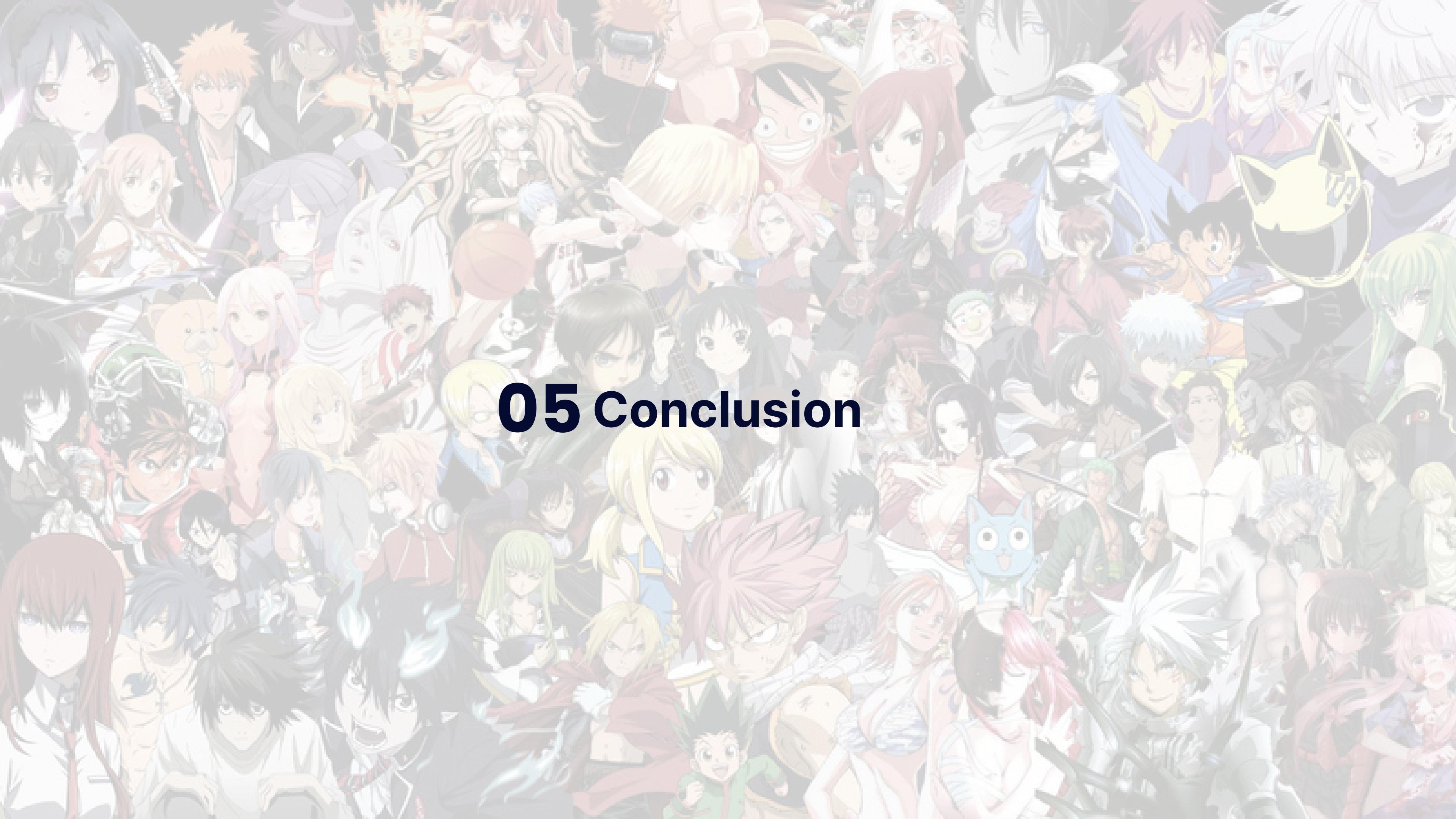
## 05 Conclusion

# Accuracy of Neural Network

```
10/10 [=====] - 3s 263ms/step - loss: 1.4250 - accuracy: 0.4812 - val_loss: 3.4585 - val_accuracy: 0.121
Epoch 16: early stopping
Training Accuracy: 0.1435
Testing Accuracy: 0.1221
```



```
test = now we want to predict
```



# 05 Conclusion

# 05 Conclusion

**Spy x Family** Edit

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

Top > Anime > Spy x Family

**SCORE** **9.09** Ranked #3 Popularity #356 Members 498,055

109,772 users Spring 2022 | TV | Wit Studio, CloverWorks

Add to List Select Episodes: 0/12

\* Your list is public by default.

**PV 3** Play More videos

**Synopsis** Edit

For the agent known as "Twilight," no order is too tall if it is for the sake of peace. Operating as Westalis' master spy, Twilight works tirelessly to prevent extremists from sparking a war with neighboring country Ostania. For his latest mission, he must investigate Ostanian politician Donovan Desmond by infiltrating his son's school: the prestigious Eden Academy. Thus, the agent faces the most difficult task of his career: get married, have a child, and play family.

Twilight, or "Loid Forger," quickly adopts the unassuming orphan Anya to play the role of a six-year-old daughter and prospective Eden Academy student. For a wife, he comes across Yor Briar, an absent-minded office worker who needs a pretend partner of her own to impress her friends. However, Loid is not the only one with a hidden nature. Yor moonlights as the lethal assassin "Thorn Princess." For her, marrying Loid creates the perfect cover. Meanwhile, Anya is not the ordinary girl she appears to be; she is an esper, the product of secret experiments that allow her to read minds. Although she uncovers their true identities, Anya is thrilled that her new parents are cool secret agents! She would never tell them, of course. That would ruin the fun.

Under the guise of "The Forgers," the spy, the assassin, and the esper must act as a family while carrying out their own agendas. Although

**Alternative Titles**

Japanese: SPYxFAMILY

**Code Cell** run this cell repeatedly to get prediction

```
name,synopsis,genres,season,studio,source = get_input()
```

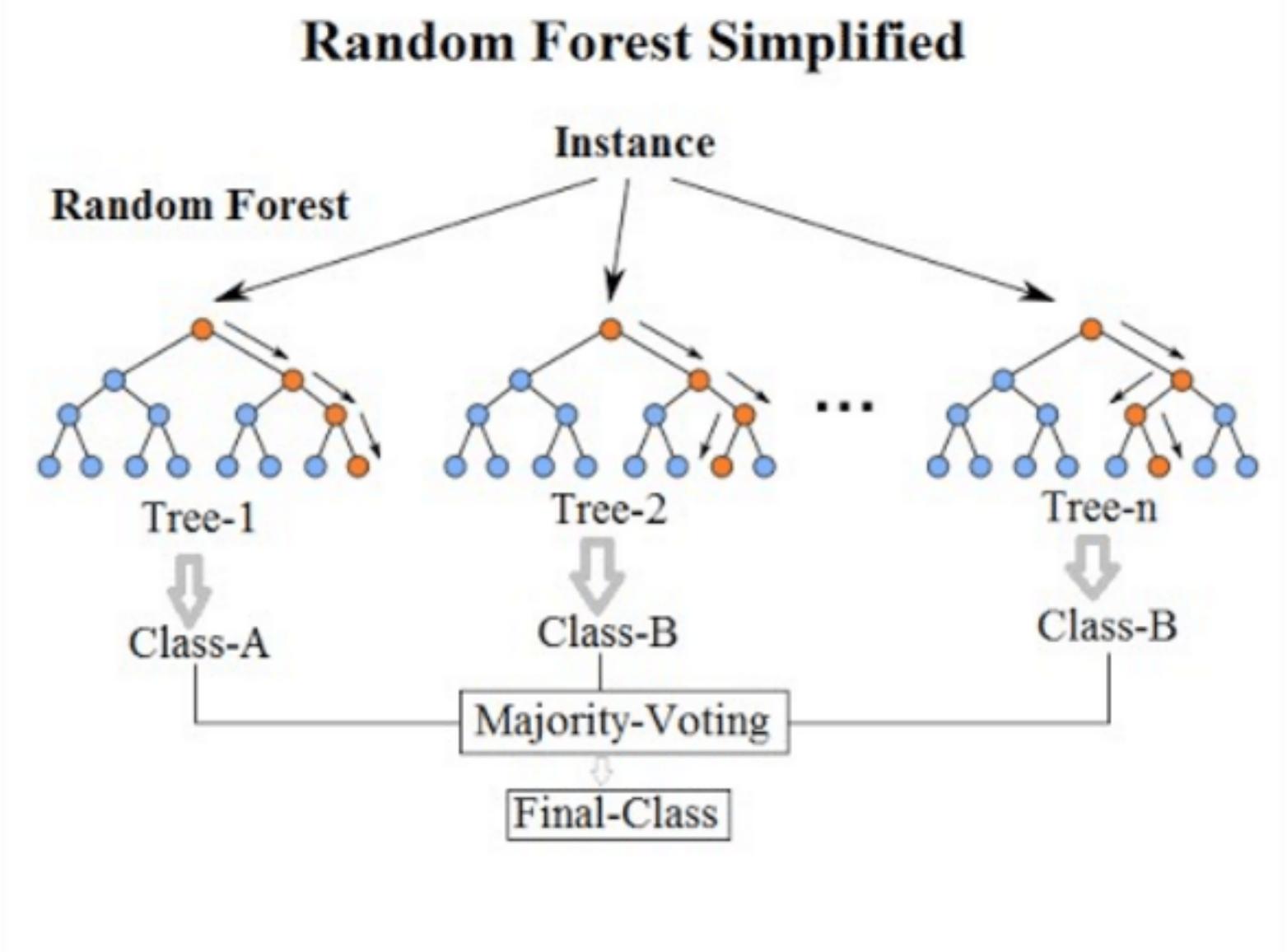
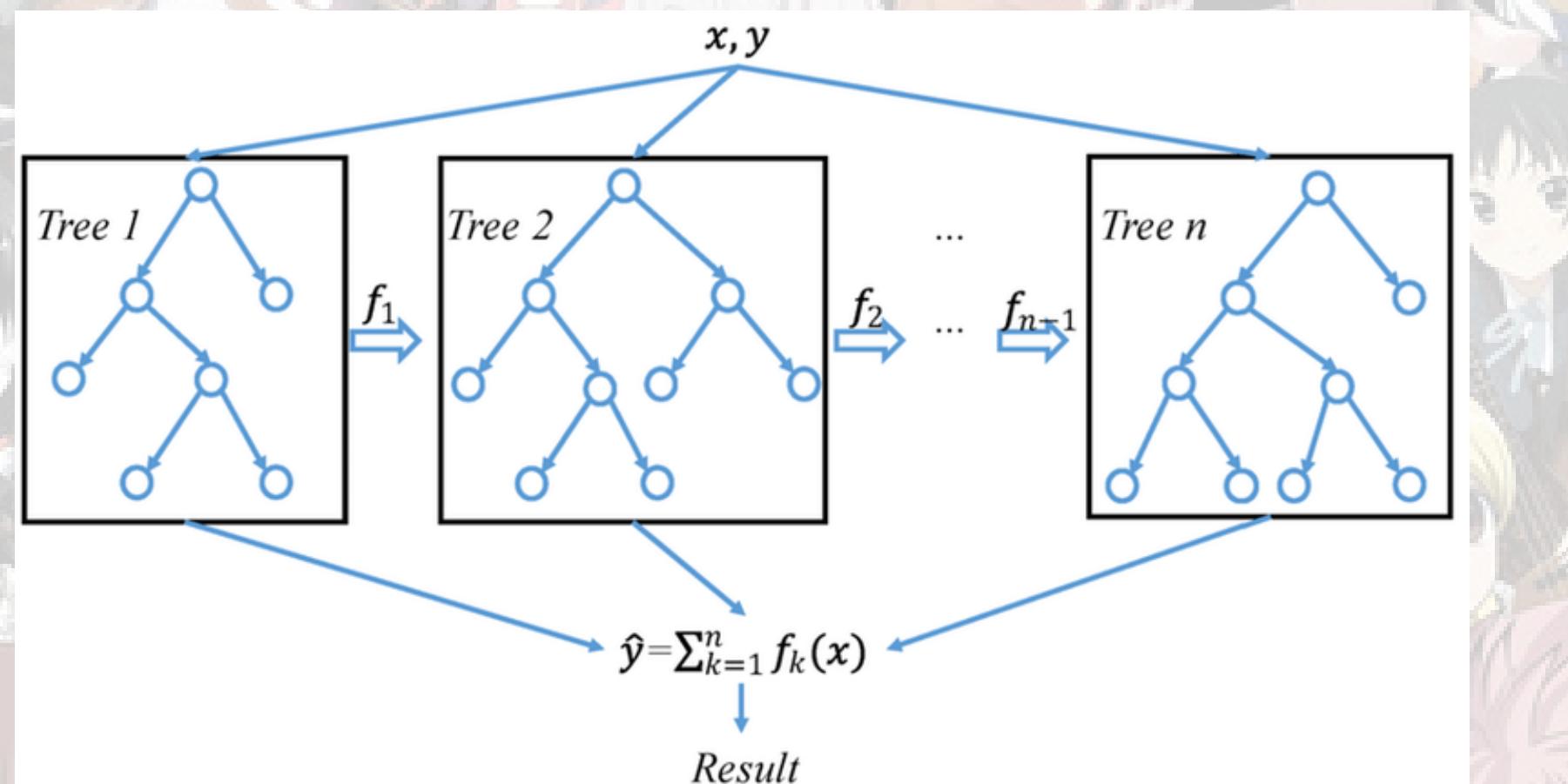
Name of anime: Spy X Family  
Synopsis of anime: For the agent known as "Twilight," no order is too tall if it is for the sake of peace. Operating as Westalis' master spy, Twilight works tirelessly to prevent ext  
Genres of anime (XX,YY)): Action, Comedy  
season anime premiered: Spring  
source of original story: Manga  
Studios produce (XX,YY):Wit Studio, CloverWorks

[32]: pred= get\_pre(name,synopsis,genres,season,studio,source,gen\_mapper,se\_mapper,st\_mapper,so\_mapper,Y\_mapper,model,maxlen)  
print(pred)

4

# 05 Conclusion

## Suggestions



## 05 Conclusion

### Suggestions

