

# Exploratory Data Analysis of Anime Dataset

**Aniket Zambare(210260060)**  
Second year, Department of Physics  
Indian Institute of Technology, Bombay

**Nahush Kolhe(210260034)**  
Second year, Department of Physics  
Indian Institute of Technology, Bombay

**Abstract**—Anime is a hand-drawn and computer-generated animation originating from Japan. After the works of cartoonist Osamu Tezuka in the 1960s, anime developed a large domestic audience. In the twenty first century, anime has changed significantly. Anime has become commercially profitable in western countries. Many anime movies have started to get streamed in movie theaters outside Japan and are available on international streaming platforms like Netflix and Prime Video as well. In this report we try to give a detailed exploratory data analysis of the anime data set obtained from MyAnimeList and find the factors affecting popularity and ratings of an anime.

■ **THE INTRODUCTION** Anime was kept for many years at the fringe of culture, as a genre either too childish or too violent. In Asia, Japanese comics and animation have been influential since 1980s to present. They have entered the cultural boundaries of foreign countries, and even gives positive impact to businessmen to promote their products or services. Gadgets of Japanese cartoon character, such as *Doraemon*, *Sailormoon*, *Dragonball*, and *Pokemon* are extremely popular among children and young people. This is shows that perception of people on anime and youth culture has been changing staggeringly.

Anime can entangle fantasy elements from Japanese culture, or even other cultures, and concerns of related to today's society. Saito (2007) refers to anime as Japan's 'greatest cultural export'. Whereas Napier (2005) says that anime can be seen as a part of "fantasy cape".

Some fans may have great fondness towards the characters from an anime that they even dress-up (cosplay) or act like them. As anime have a lot of different types of characters, it's easy for the viewers to relate to one (or more) of the characters. Some young characters in anime show strong and adult like behaviour while facing the problems. Many of the viewers can relate to this. These factors affect the popularity of anime and

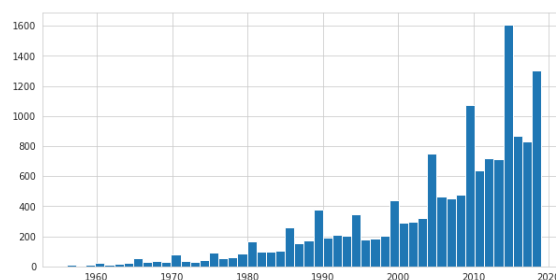
their genres in various age groups.

## Basic Data Analysis

In this section we look at the plots of some parameters related to anime vs number of anime. These parameters can be number of episodes, members or score given to the anime. These plots will give us insights about overall anime community and people's very general preferences.

First we look at the plot of number of anime released per year. This can be seen in the plot below.

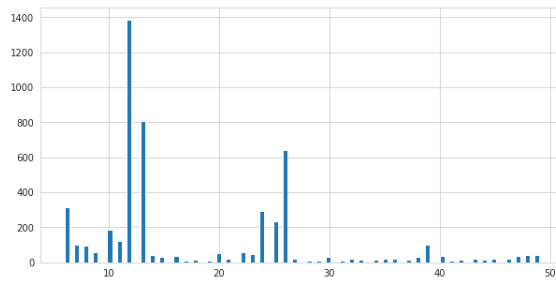
We can see that roughly number of anime produced,



**Figure 1.** Year vs number of anime released

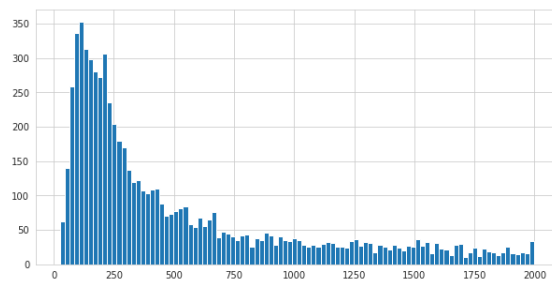
and hence released, has increased as the time progresses. However there is a decrease in production of anime in an year followed by 2016, which corresponds to the global peak.

The plot below shows the distribution of number of episodes of anime released. The peak corresponds



**Figure 2.** Number of episodes vs number of anime

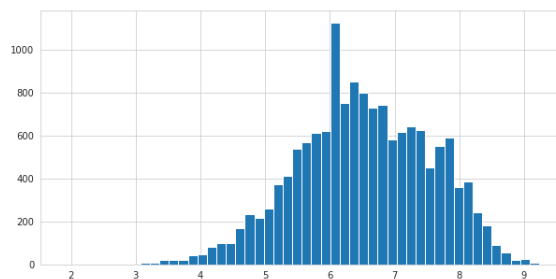
to 12-13 episodes. So this means that anime studios prefers producing anime with 12 to 13 episodes. This is mostly because production of anime with too many episodes can be very expensive, and you cover decent amount of story-line in 12 to 13 episodes. Popularity and ratings for the same are discussed next section. We get the information about average number of anime people have watched from the plot below: Total



**Figure 3.** Members vs number of anime

number of members who have watched at least one anime is around 650 millions.

Lastly, we look at plot showing number of anime with a particular score. The average score given to anime



**Figure 4.** Score vs number of anime

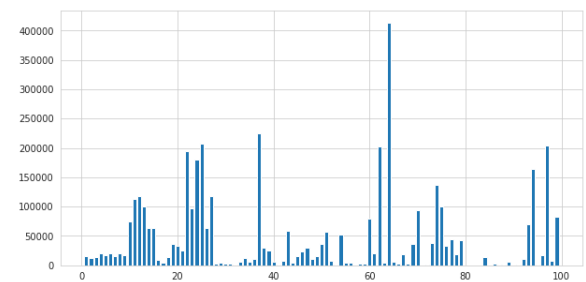
is 6.52. Most of the anime have a score between 6-7,

meaning apart from few anime with more that 8 rating, most of the anime are mediocre.

## Relations between Parameters of anime

In this section we look if there are any relations between two different parameters of anime. These parameters can be genre, number of episodes, rating etc.

Number of members, who have watched an anime, depends on the number of episodes it has. If there are too many episodes with a mediocre story-line, some may consider this as simply time consuming. The peaks are around 13 and 26 episodes. The nearby

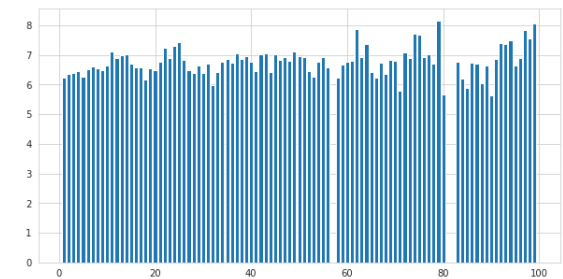


**Figure 5.** Number of episodes vs members

regions of these peaks also have similar values. Unlike these, the peak at around 65 episodes arises due to a particular anime.

The graph below tells us that number of episodes and score are not tightly related.

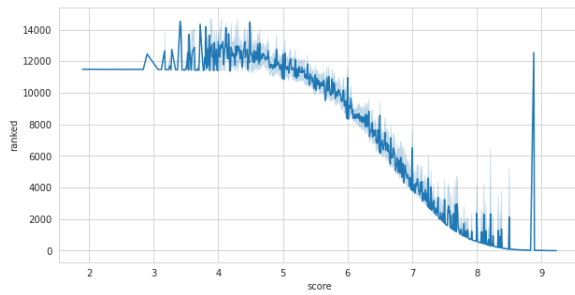
So the rating given to anime is mostly based on other



**Figure 6.** Number of episodes vs Score

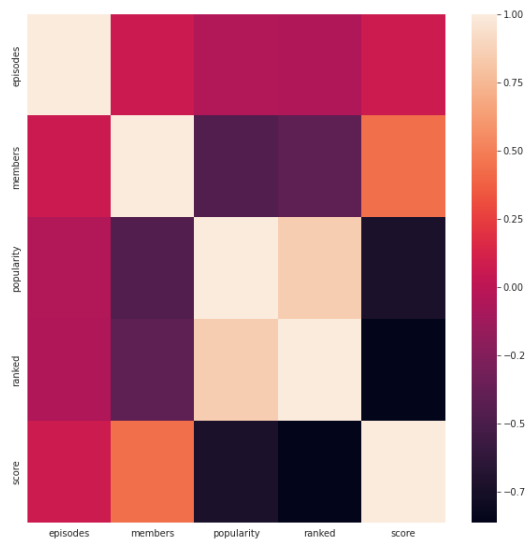
factors like story-line, animation and OSTs.

The plot between score and ranking tells us that ranking is proportional to the score of an anime, which is quite obvious. The appearance of an outlier point is not obvious. The anime which corresponds to this point has very few members in its community. So its score is very biased causing its ranking to be low.



**Figure 7.** Score vs Ranking

Above heat map summarises all the correlations



**Figure 8.** Correlation matrix of all parameters between the parameters.

## Analysis of Members

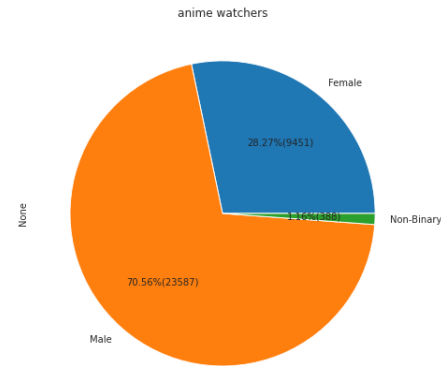
In this section, we will look at the characteristics, age and gender, of the members. From the figure 9, we can conclude that anime is especially popular in males. However, interests of females in anime is growing as well. So in few years, anime members won't be dominated by males only.

The age wise distribution of profiles is not very surprising. Anime are especially popular in college students and post graduate students.

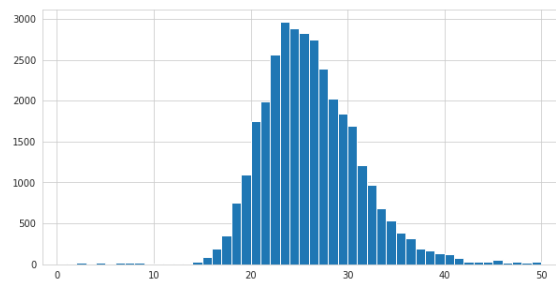
Figure 11 is similar to figure 10. In females and males, most of the members have age around 23.

## Genre-wise Analysis

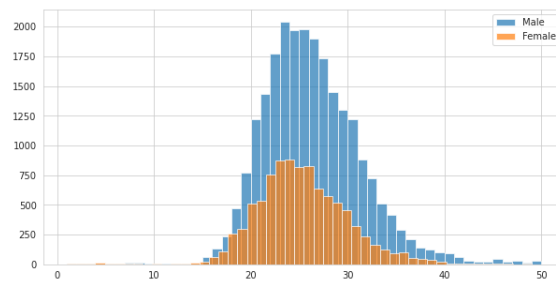
Genre is a very important parameter of anime. One of the reason anime having a huge number of fans



**Figure 9.** Gender-wise distribution of members



**Figure 10.** Age vs Number of profiles

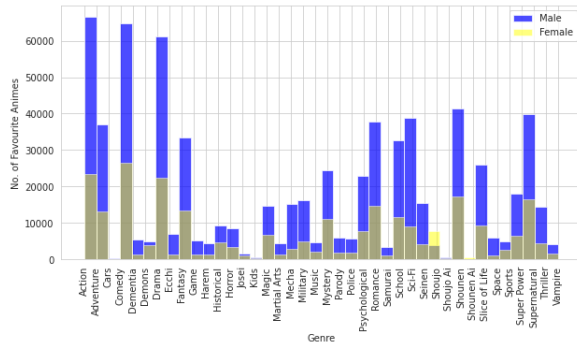


**Figure 11.** Comparison of age-wise distribution of Male and Female members

is that there is a lot of variety in the genres. Other parameters like popularity, number of anime with that genre, or even characteristics of the members like age and gender. These genres can be action, shounen, romance, comedy, thriller, horror and many more.

First we look at the genres preferred by males and females. Figure below shows us the plot between number of males (blue) and females (yellow) having a particular genre (plotted on x axis) as their favourite genre. It can be seen that females have more interest towards genres like 'shoujo' and 'shounen-ai' than boys.

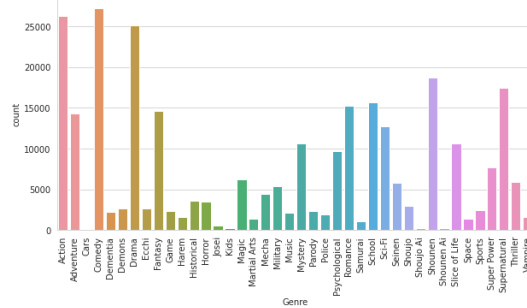
Although overall distribution for males and females



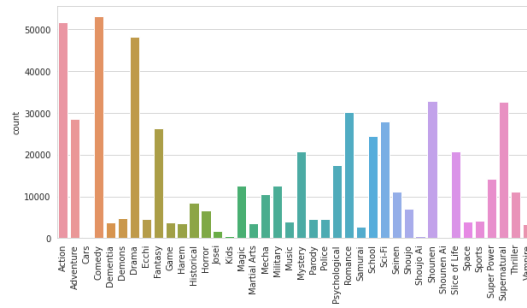
**Figure 12.** Number of males and females vs favourite genre

may look similar, males have very high interests in some particular genres like 'action' and 'comedy' compared to other genres.

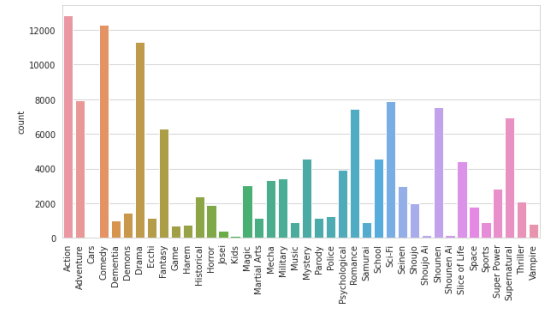
Similarly we can analyse how the preferences of members changes with their age. The division of age in three groups is based on the fact that most members younger than 22 years are in college or high school. Whereas members in the age group (22-30) years have graduated from college, started the job and are settling in their lives. Final group of members older than 30 years have more or less settled in their lives



**Figure 12(a).** Favorite genre for the age group of (0-22) years



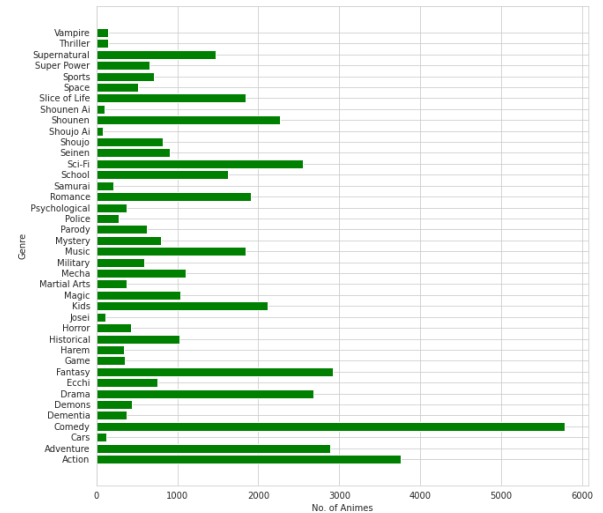
**Figure 12(b).** Favorite genre for the age group of (22-30) years



**Figure 12(c).** Favorite genre for the age group of (>30) years

Graphs of all three age groups are more or less similar. However differences between the number of people preferring two different genres is much higher in group 2 compared to other two groups.

It was also observed that anime studios prefer making anime of some genres more. This can be seen in the figure 13. It can be easily seen that there are a

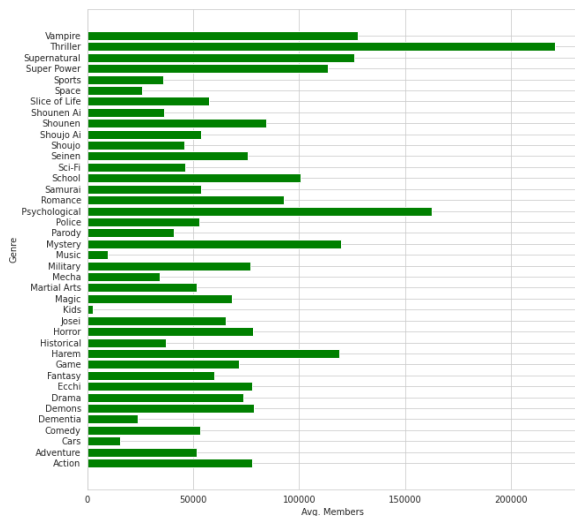


**Figure 13.** Genre vs number of anime

lot of number of 'comedy' anime, whereas genres like 'cars' and 'vampire' have very few anime. It is also very interesting to see that 'thriller' genre has very few anime. This tells us that although it's a popular genre, for some reasons like cost of production and difficulty in writing the story-line are causing this genre to have very few, but surely very popular, anime.

This popularity can be seen from the plot in figure 14.

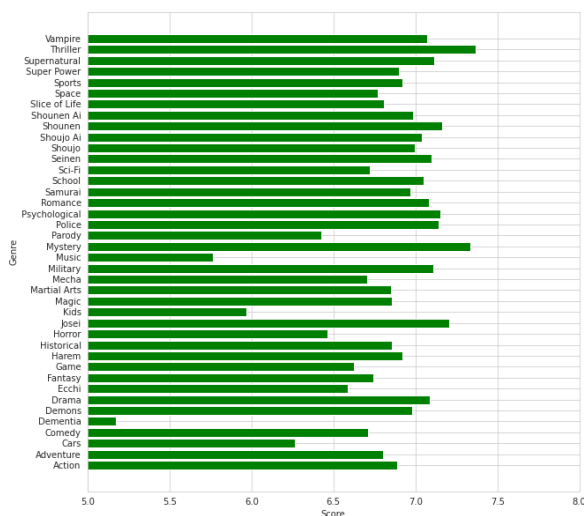
As discussed at the beginning of this section, some



**Figure 14.** Genre vs average number of members

genre like 'thriller' and 'psychological' are much more popular than genres like 'kids' and 'music'. This plot indicates the interests for members. Lastly we see if there's any relation between score and genre. As shown in figure 15, there is not very strong relation between these two. Except for few genres like 'music' and 'dementia', most of the genres have similar ratings.

This shows that there are good anime in all genres.



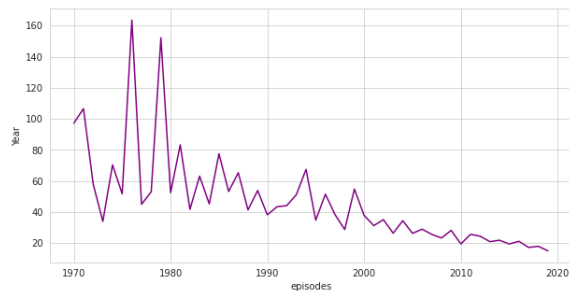
**Figure 15.** Genre vs score

## Trends of Parameters

Now we come to the last part of this report. In this section we will see some of past year trends of some parameters of anime like score and number of episodes. We will try to loosely predict some of the future aspects of these parameters.

First we begin with trend of number of episodes. This is trend is shown in figure 16.

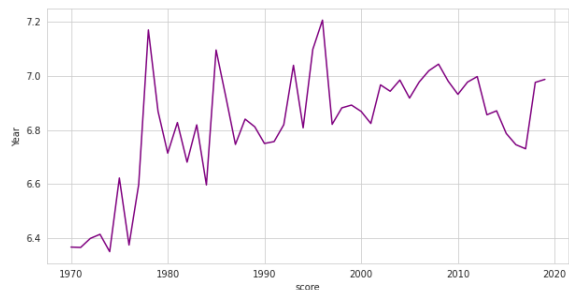
It can be seen that average number of episodes have



**Figure 16.** trend of number of episodes

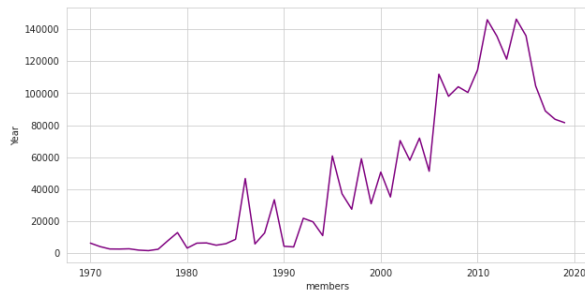
decreased with time. This means that anime studios are preferring to produce anime with lesser number of anime. One of the reasons for this is members prefer watching anime with less number of episodes as it's less time consuming. Another reason is producing an anime with too many episodes can be very expensive. Now we will look at the trend of the scores of anime. This trend has a positive slope on an average.

This basically indicates that the quality of anime has



**Figure 17.** trend of score

been increasing with time. One of the reasons for this can be advancement in technology, which is allowing the anime studios to produce anime with much better animation, sound tracks and cost-effectiveness. This fact is causing anime to seek attention of more people. This can be seen in figure 18:



**Figure 18.** trend of number members

## Conclusion

In this report, we first analysed the parameters of anime and their interrelations. Then we explored the genres and their relations with other parameters of anime. Finally we inspected the trends of some parameters of anime. From these trends, we can conclude that anime industry will grow bigger as the time progresses.

## REFERENCES

1. Datasets were obtained from  
<https://www.kaggle.com/datasets/marlesson/myanimelist-dataset-animes-profiles-reviews>
2. <https://towardsdatascience.com/exploratory-data-analysis-on-anime-data-468cc15e13b8>
3. [www.wikipedia.org](http://www.wikipedia.org)
4. stackexchange
5. GeeksforGeeks