Anime Rating Analysis

Github: <https://github.com/Uthkarshh/Empirical-Analysis-Project>

Video: <https://drive.google.com/drive/u/0/folders/1a-DAzl8bzZ4RXs-yWky1tmxcHDQgV6HJ>

# Introduction

## Motivation

For a long time, anime has been a genre that has been mostly followed by a very small and specific group of fans. However, there has been a massive surge in the popularity of anime all over the world. The demand for anime has increased significantly in recent years, growing by 118% in just the last two years alone. In December 2021, anime’s overall global demand share was 7.11%, up from 4.2% in

January 2020 according to Data from consulting firm Parrot Analytics. This means that anime now ranks as one of the most popular content genres worldwide. This is largely thanks to the release of hit anime series such as My Hero Academia, Demon Slayer, Jujutsu Kaisen, Spy × Family and anime movies such as Kimi No Na Wa (Your Name), Demon Slayer the Movie: Mugen Train, A Silent Voice.

Now there is an insatiable demand for anime all over the world, and it has become one of the most profitable genres out there.

## Significance:

MAL is the world’s most active online anime and manga social network where fans discuss anime/manga and rate them. Users can rate anime on a scale from 1-10, with the ratings often dependent on different factors such as genre, anime studios, number of episodes etc. The main motive of this project is to analyze the factors that play a major role in the rating of an anime and to find a correlation between the rating and the different factors. As a result, this model will be helpful for studio house to look upon different factors before investing in projects and for users to determine if they would want to watch the anime.

## Objectives:

With the rising popularity of anime, many well-known streaming services such as Netflix, Amazon Prime Video, and Hulu have started carrying a larger selection of anime titles. Anime streaming services such as Crunchyroll have also begun to experiment with lesser-known anime in an

effort to attract more viewers. This has led to a drastic increase in the number of ratings for anime shows and movies on MAL, as more and more viewers are tuning in. The purpose of this project is to explore how this influx of anime viewership has affected the number of ratings for anime shows and movies on MAL.

## Related Work (Background):

Although there have not been many extensive studies done on different anime’s ratings and popularity, in this paper by (Cho, Schmalz, Keating, & Lee, 2018) an analysis of 396 recommendation request threads from the online forum at Anime News Network was conducted in order to identify and understand relevant information features for anime recommendations. These features include work, theme, genre, audience, mood, and artwork/visual style. The findings of the analysis can be used to help recommend appropriate anime titles to users based on their specific preferences and this paper by (Simon) provides us with a general introduction to how anime has entered the global market, as well as how it is seen as an aspect of culture. The growth and acceptance of anime globally, challenges of anime and empirical study conducted to increase understanding of cultural perceptions of anime and to identify the factors bearing on its popularity are all touched on in this article.

## Dataset:

The two datasets “anime.csv” & “anime\_rating.csv” are taken from Kaggle and they contain details about 12277 anime and ratings provided by 7813737 users. Each user is able to rate the anime they have watched from 1-10, and this dataset is a compilation of those ratings.

### anime.csv

**anime\_id:** unique id of the anime from MAL

**name:** title of the anime.

**genre:** a list of genres the anime is tagged on

**type:** the type of anime such as movie, Series, OVA, etc.

**episodes:** total number of episodes in the anime

**rating:** average rating of the anime on a scale of 1 to 10

**total\_members:** number of fans in the community of the anime

studio: Production studio of the Anime

### aime\_rating.csv

**user\_id:** An ID unique to each user

**anime\_id:** Id of the anime rated by the user

rating: Rating assigned by the user to anime on a scale of 1 to 10. (-1 is given when the user has seen the anime but hasn’t provided any rating to it)

## Detail design of Features and Analysis:

**Data loading, cleaning, and pre-processing:**

In this stage, I utilized Pandas, NumPy and Seaborn libraries for tasks such as identifying duplicate and null values, identifying and analyzing outliers and summarizing the datasets. This was very helpful in further understanding the data I was working with.

**Data Visualization:**

I used Matplotlib and seaborn to visualize the dataset and get a better understanding of it. I was able to answer questions such as:

What are the top 15 Anime based on number of members in the community and average ratings?

What are the number of anime in each category?

What does the distribution of ratings look like? Etc.

## Implementation:

**Model Training:**

Once I've completed the encoding process, I will split the dataset into training and testing datasets. I will use the training dataset to fit different models and choose the model with the best accuracy for predicting ratings based on different features in the dataset.

## Preliminary Results:

Graphical user interface, text, application

Description automatically generated

Graphical user interface, application

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Chart, box and whisker chart

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence

Graphical user interface, application

Description automatically generated with medium confidence

## Project Management:

### Implementation status report:

So far, I've loaded and cleaned the data from the datasets. Next, I’ve performed a descriptive analysis on the data which helped me understand it better. Then I Visualized various attributes of the data to get deeper insights about the dataset and relation between different features. And lastly, I encoded the categorical variables in the data so that they can be used to train models.

### Work to be completed:

The dataset needs to be split into Training and testing sets respectively and fit different models on the training set. Post which all the models will be evaluated based on different metrics and the best model among them will be chosen

# Bibliography

Cho, H., Schmalz, M. L., Keating, S. A., & Lee, J. H. (2018). Analyzing anime users’ online forum queries for recommendation using content analysis.

Simon, J. (n.d.). Japanese Anime: Factors Leading to Acceptance or Rejection.