The multifaceted nature that determine an anime show's popularity involves (findings of the paper, WIP).*

Justin Abando

19 April 2023

There are online lists and forums where users can vote on what they believe to be the most popular anime such as MyAnimeList (MAL). However, there are no metrics to determine why specific anime stand out, with some shows being rated lower in terms of quality, yet have more popularity based on MAL user responses. Data was retrieved to determine what factors could be responsible for a show's popularity. It was found that (results of analysis, WIP). The findings of this paper suggests reasons why a particular anime becomes popular, which can be relevant to those that are new to watching anime or those that want something new to watch as they would gravitate to shows that receive traction by a broader demographic.

1 Introduction

Japanese animation (referred hereafter as anime) is a medium that has been exploding in popularity in recent decades. With an already large catalogue of shows to choose from, many newcomers to the medium seek guidance in what shows they should start watching as their entry to anime. One such website that helps with this is MyAnimeList (MAL), which serves as both a social networking platform for anime fans, but also as a social cataloguing website to keep track, rank, and review the shows that you can watch. MAL is one of the most frequently used anime catalogues online, with millions of active users and website visits every month.

With a large user base, MAL members can personally rate and review the shows that they watch and catalogue them in their anime lists. While this is useful at the individual level to keep track of their own journey and see the depth and breadth of the shows they watch, it is also useful at the community level. The rating and reviews that are given from individuals

---

*Code and data are available at: https://github.com/justabando/malanalysis.

are aggregated and averaged using all of the data from other MAL users to give an overall subjective ranking of each show in terms of popularity and rating. It should be noted that a show's ranking is not defined by objective means, but is the majority voice of all the subjective reviews left by the users of the website. With MAL, popularity and rating are both metrics used to sort and determine the top anime based on the reviews and ratings that were given to them by users. While the rating of the show matters for popularity to an extent, a large part of it's score is dependent on the content of the show itself. The rating is dependent on a scale of 1-10, and takes the average score of all the users that have rated the show. Meanwhile, popularity as a ranking metric demonstrates how much exposure the show has gotten and its outreach across as many demographics as possible, and is a ranking that is determined by the amount of users that rated the show regardless if it was positive or not.

This presents two ways for newcomers to judge a show before watching. They can be inclined to start a show that is subjectively rated as one of the best as ranked by the MAL community, or choose one that has gained the most traction despite any apparent flaws or drawbacks that may deter a positive watching experience. As such, it should be noted that popularity is not a metric that is inherently positive, but can also imply the notorious status of a show, especially if it is ranked low in terms of rating but manages to be much more popular than its rating score suggests. This creates an interesting method to identify shows that have managed to be popular despite its poor ranking performance. It could be that the show has notoriously bad traits that often gets shared to others to ridicule or prank unsuspecting friends to watch and experience a show that can be "so bad that its good". The anime may have been able to maintain a good quality throughout its run time, but eventually hit a point in the plot that polarizes the show's watchers. Regardless, there are many ways to interpret the collective opinion of a show on MAL with the presence of ranking both its ratings and popularity. While the rating of a show is indicative of how well-received the show is strictly by the subject and content of the show itself, the popularity of a show can be shrouded in mystery and depend on many logistical factors the put an anime in the limelight.

With all of this in mind, the objective of this analysis is to determine the factors that influence an anime's popularity. How an anime becomes well-known and circulated through the community and even seeps through the cracks and becomes mainstream is an interesting topic to discuss due to the many facets of the show in terms of its production, marketing, and teams that work on them. To determine these exact causes, a statistical analysis is required to contextualize the exact factors that are relevant in determining an anime's popularity. These factors will be expanded upon in Section 2.

It was found that (the results of paper, WIP).

The findings of this paper suggest that the popularity of a show is influenced by a variety of factors that may not be obvious by just watching the show itself. These findings can help someone who is new to watching anime, or those that want something new to watch to find something that gained traction within the community. Popularity can be a factor that is more important than the ranking of the show's quality in many cases, and this is due to the context that the show presents. An anime may be very well-received, but it may be less known within

the broader community due to subject matter that may be difficult to process, whether it would be highly philosophical, graphic, or leans in any other kind of extreme direction that may lose the interest of a casual viewer. With less overall viewership of an anime, the show may still be ranked relatively high in terms of rating due to its niche community that watches the show, which can inflate its overall ranking in its ratings. This is why ratings are not necessarily the best metric to depend on for newcomers to find a show to watch, as a niche show may not be widely accessible to a broad audience. However, popularity ranking can be much more indicative of a potential show someone can watch due to its exposure that is not dependent on the taste or bias of the viewer. As such, the (factors that influence popularity that will be discussed on, WIP).

This statistical analysis is divided into several sections, with Section 2 giving context and describing the variables that will be used to determine the underlying reason behind an anime's popularity. Section 3 describes the model used to create the correlation between a show's popularity and the factors that possibly influence it. Section 4 demonstrate the overall findings of the statistical analysis and puts into perspective how these factors affect a show's popularity, and Section 5 covers a discussion of the observations that were made and inferring the causes based on the results of the data.

2 Data

2.1 Data Collection

This report focuses on the data available from MyAnimeList (MAL) since July 2022. MAL is a social networking platform for anime and manga that also doubles as a database for said mediums. The data itself was obtained from Kaggle from Andreu Vall Hernandez, which gathered data from the official MAL API, as well as a third-party Jikan API. A limitation to the way the data was gathered was that it was aggregated from an outside source, however there is currently no way to publicly access the official MAL API without requesting it through an application process as it is currently in active development.

Despite the data from MAL being used in this analysis, there are other anime databases that could have been used, such as Anilist and AniDB, MAL has its prominence within the anime community as exemplified by its high user and visitor counts, making it a more viable option to compare the primary measurement of interest: popularity.

Regardless, data from MAL was imported and analyzed in R (R Core Team 2023). To aid in this analysis, the `dplyr` (Wickham et al. 2023) and `ggplot2` (Wickham 2016) packages from the `tidyverse` package (Wickham et al. 2019) were used to clean the data set and generate graphs to visualize it respectively.

2.2 Data Cleaning and Description {sec-data-cleaning}

Cleaning involved filtering specific variables within the data set to include that were essential to the analysis. The final data set contains the following variables:

- `title`: title of the anime.
- `type`: the medium of how the anime is delivered (e.g. tv show, movie, etc.).
- `score`: aggregated score of the anime out of 10, contributed by users of MAL that both have the anime in their list as well as rated it numerically out of 10.
- `scored_by`: the number of MAL users that have the anime in their list and rated the anime out of 10.
- `members`: the number of MAL users that have the anime in their list.
- `episodes`: the number of episodes that a given anime has.
- `source`: the type of source material that the anime was adapted from (e.g. manga, light novel, anime original, etc.).
- `start_year`: the year in which the anime first started to broadcast.
- `start_season`: the season in which the anime broadcasted in.
- `genres`: the genres of the anime.
- `studios`: the names of the studios that produced the anime.
- `licensors`: the names of the licensors that release the anime to different audiences.

The important variable of note to consider from this list is `members`. This is a quantitative variable that popularity ranking on MAL is completely dependent on. Other than `title`, the rest of the variables will be tested to see if they have any kind of influence on the amount of users that have specific shows on their lists.

3 Model

4 Results

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Limitations and next steps

Appendix

A Additional details

References

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation.*