Exploratory Data Analysis of MyAnimeList 2023 Dataset

Mukund Valluru

January 31, 2025

Contents

1 Introduction

This EDA is intended to understand the **MyAnimeList** dataset from various perspectives, taking into consideration both the anime-level information (e.g., Score, Episodes, Type, Genres, etc.) and user-level details such as Gender, Mean Score, Total Entries, etc. Our objectives are as follows:

- To examine important univariate distributions, such as how anime are scored or how many episodes they have.

- Investigate multivariate relationships—whether some variables (e.g., popularity, user ratings) correlate with official scores or show trends among different user demographics.

These findings can help understand what types of shows are most common, which are rated highest, and how audience tastes evolve.

2 Univariate Analysis

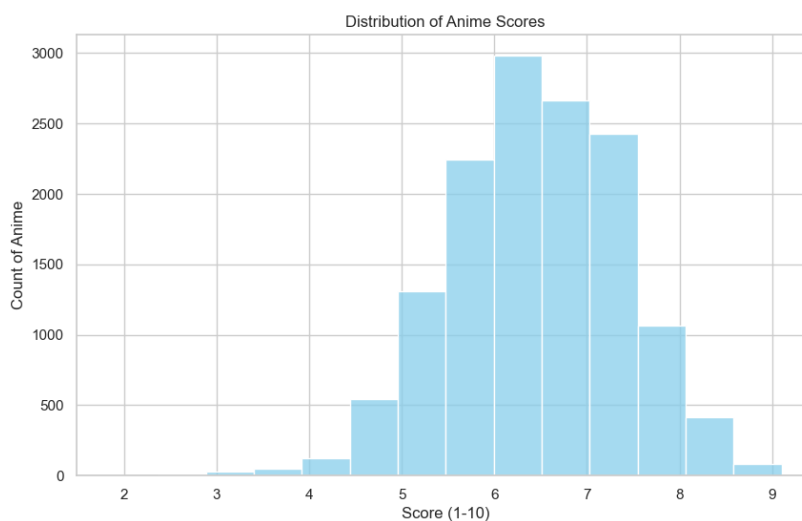## 2.1 Anime Attributes

2.1.1 Distribution of Anime Scores

Figure 1: Distribution of Anime Scores (1–10)

Most anime cluster around **6–8** on MyAnimeList's 1–10 scale. Few shows drop below 4 or rise above 9, suggesting that ultra-low or ultra-high scores are comparatively rare.

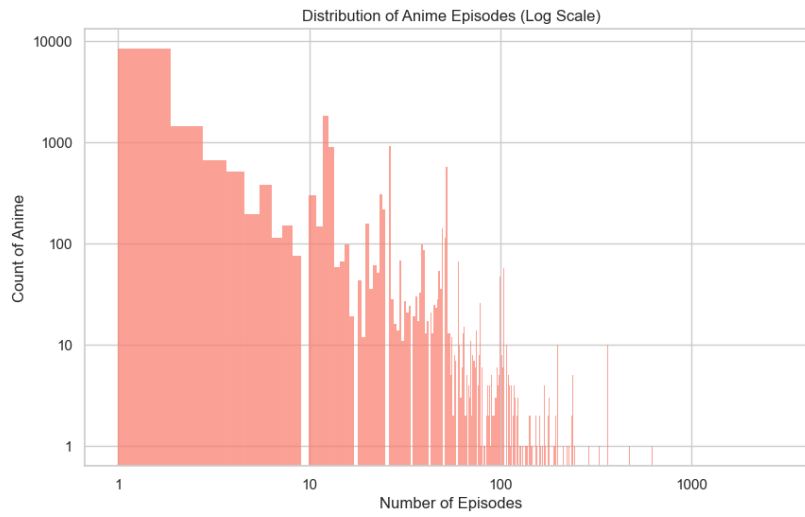### 2.1.2 Distribution of Anime Episodes (Log Scale)

Figure 2: Distribution of Anime Episodes (Log Scale)

We can see that most anime have **1–13** episodes (common for shorter series or OVAs). A large group clusters around **12–24** episodes (a typical one-season TV anime), while some run for more than **100** or even over **1000** episodes.
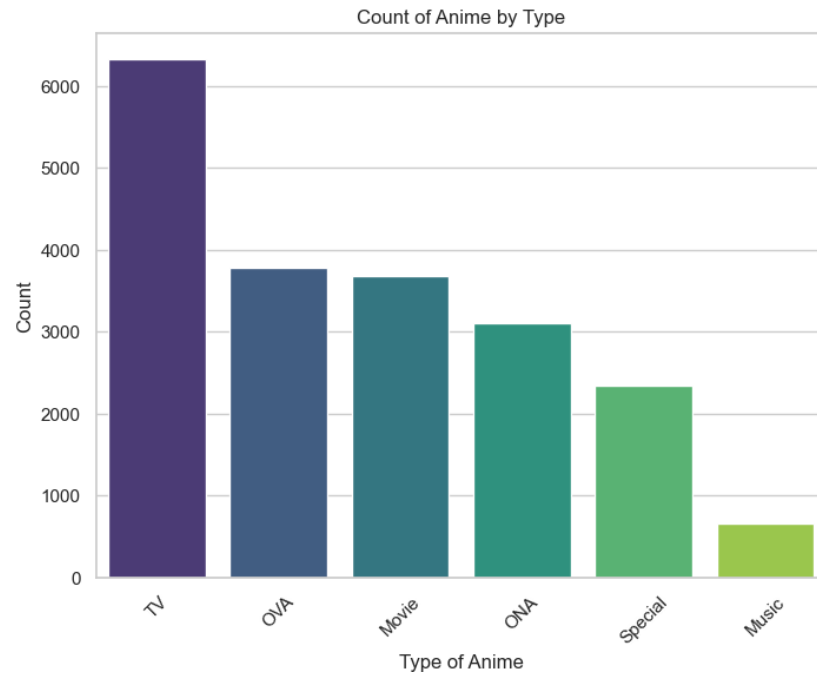
2.1.3 Count of Anime by Type

Figure 3: Count of Anime by Type

TV series dominate the dataset, followed by **OVA** and **Movie**. **ONA**, **Special**, and **Music** entries appear less frequently, indicating more niche or special releases.
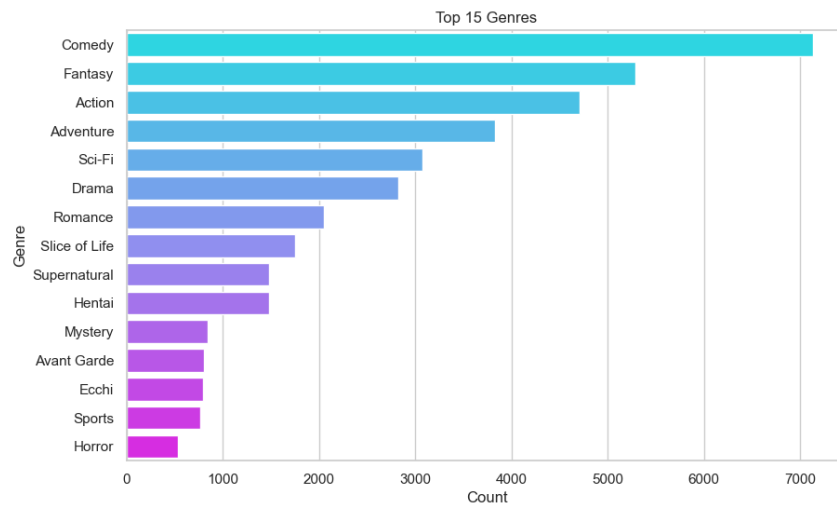
2.1.4 Top 15 Genres

Figure 4: Top 15 Genres

The genres with the highest frequencies assigned are **Comedy**, **Fantasy**, and **Action**. Adventure and Sci-Fi are not far behind, indicating that the audience enjoys highly escapist and speculative themes common in anime.

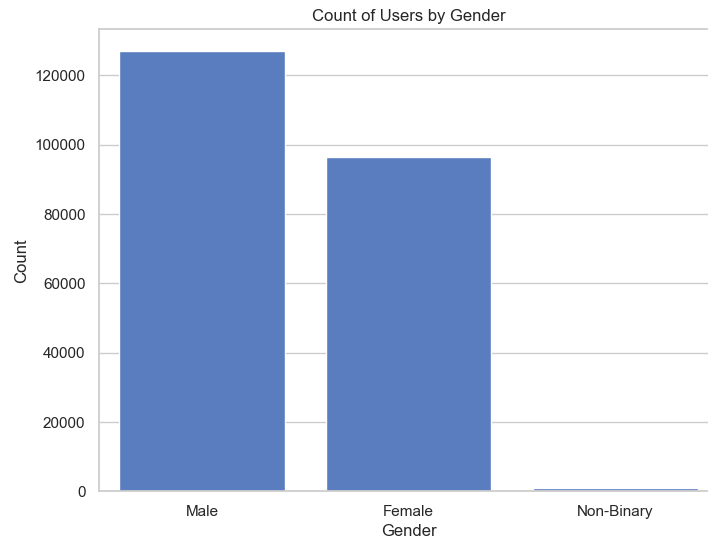## 2.2   User Attributes

2.2.1   Count of Users by Gender

Figure 5: Count of Users by Gender

As illustrated by the data, most of the users identify their gender as **Male**, followed by **Female**, with a small fraction identifying as **Non-Binary**.
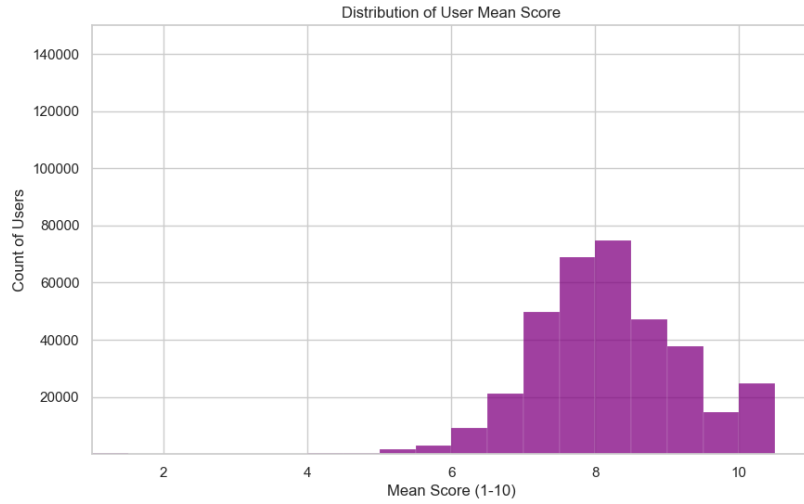
2.2.2 Distribution of User Mean Score

Figure 6: Distribution of User Mean Score (1–10)

Most users' average rating lies around **6–8**, mirroring the typical anime score distribution. This suggests that overall, users tend to rate shows favorably, and extremely low or high averages are less common.
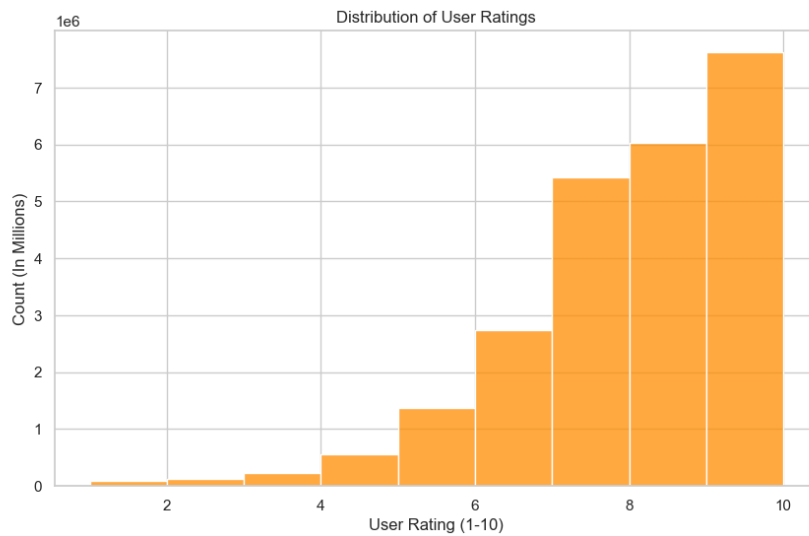
2.2.3 Distribution of User Ratings

Figure 7: Distribution of User Ratings (1–10)

Individual user ratings are weighted toward the upper end (**8–10**). This pattern indicates an enthusiastic user base, as lower ratings (1–3) are relatively rare.

3 Multivariate Analysis

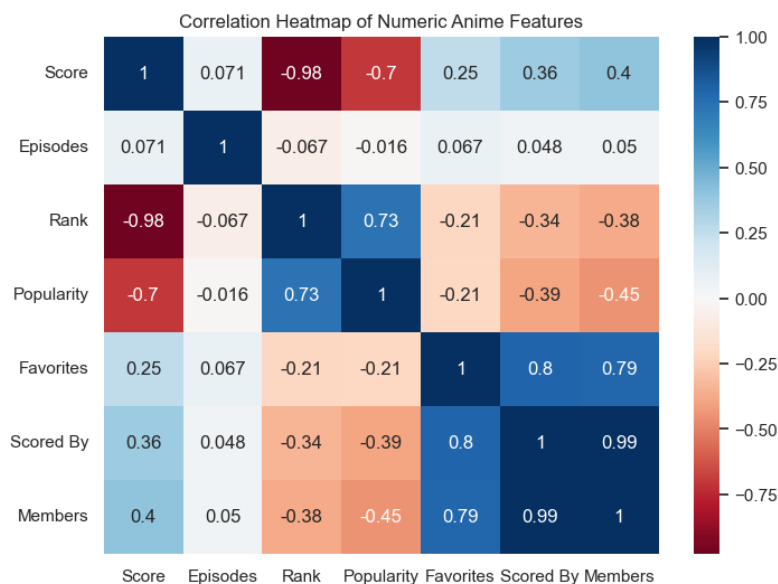## 3.1 Anime-Level Relationships

3.1.1 Correlation Heatmap

Figure 8: Correlation Heatmap of Numeric Anime Features

- `Rank` shows a strong *negative* correlation with `Score` and `Popularity`.

- `Favorites`, `Scored By`, and `Members` are all strongly interrelated, indicating that popular shows gather more total interactions.
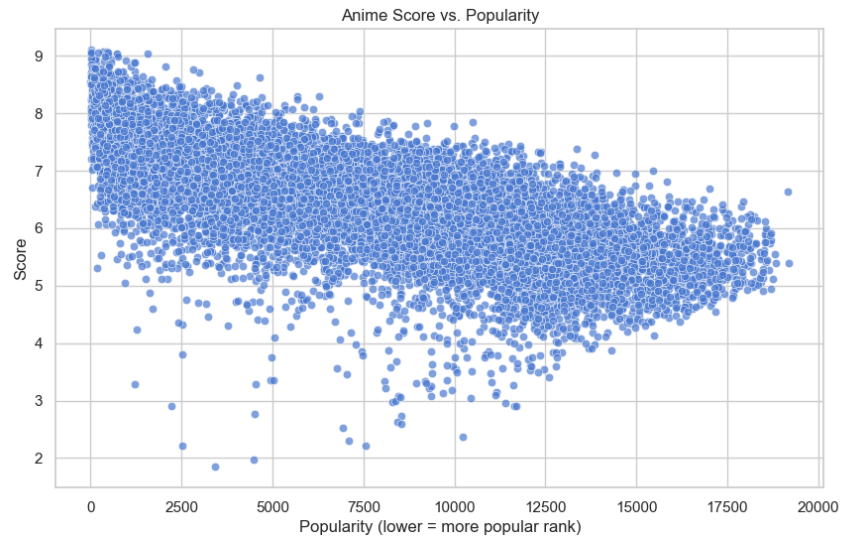
3.1.2 Score vs. Popularity

Figure 9: Anime Score vs. Popularity

A downward trend would indicate that programs with lower popularity rank (i.e., more popular) tend to have higher average scores. As the rank value increases (less popular), average scores tend to drop.
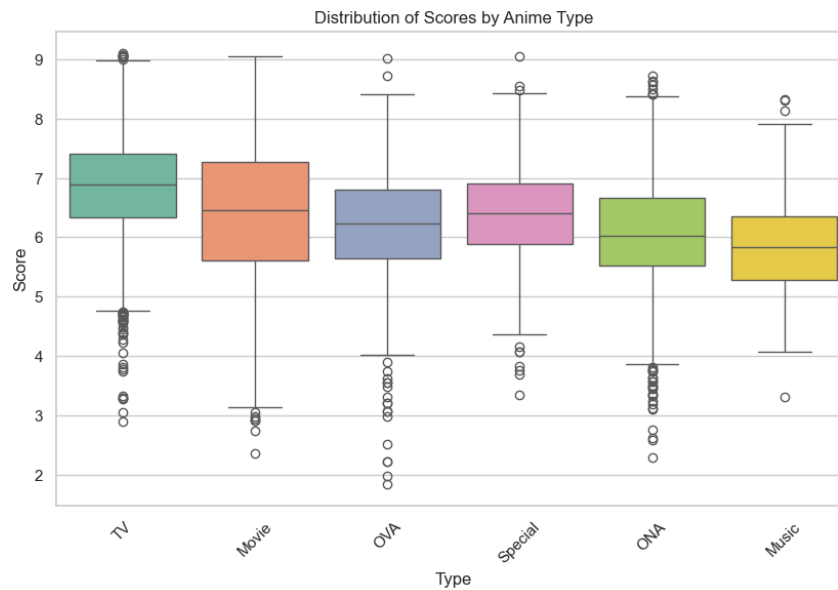
3.1.3 Box Plot: Score by Type

Figure 10: Distribution of Scores by Anime Type

Movies and **TV** often show higher median scores than categories like **Music** or **ONA**. **Specials** and **OVAs** have broader distributions, reflecting their often niche or experimental nature.
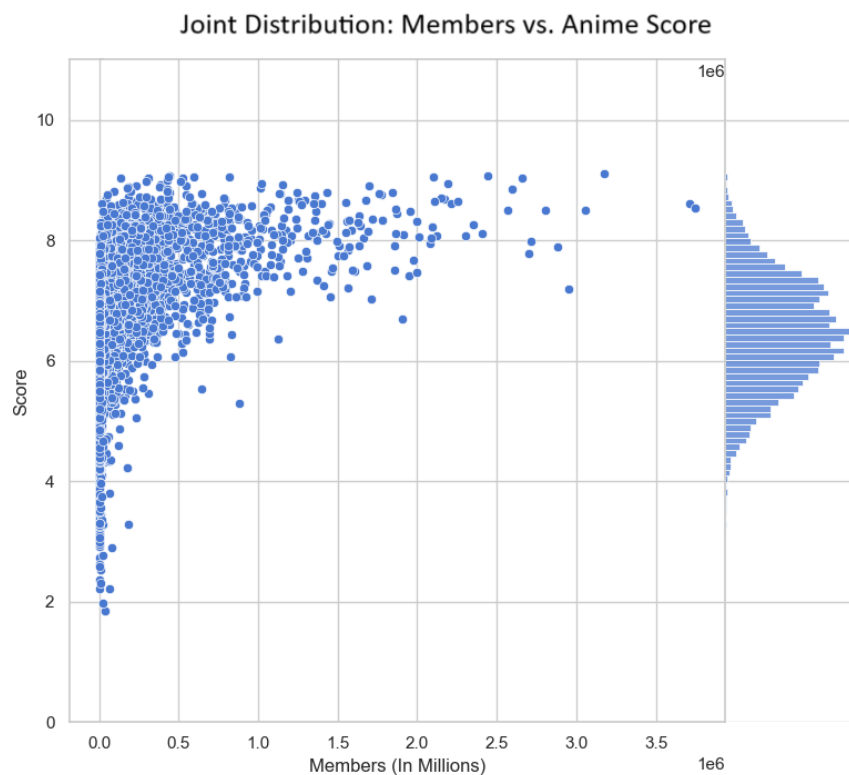
3.1.4   Joint Distribution: Members vs. Score

Figure 11: Joint Distribution: Members vs. Anime Score

Shows with fewer members are distributed more widely regarding `Score`, while highly followed titles tend to occupy the range of **6–9**. The marginal histogram indicates a skewed membership distribution: a small fraction of anime attract huge followings.
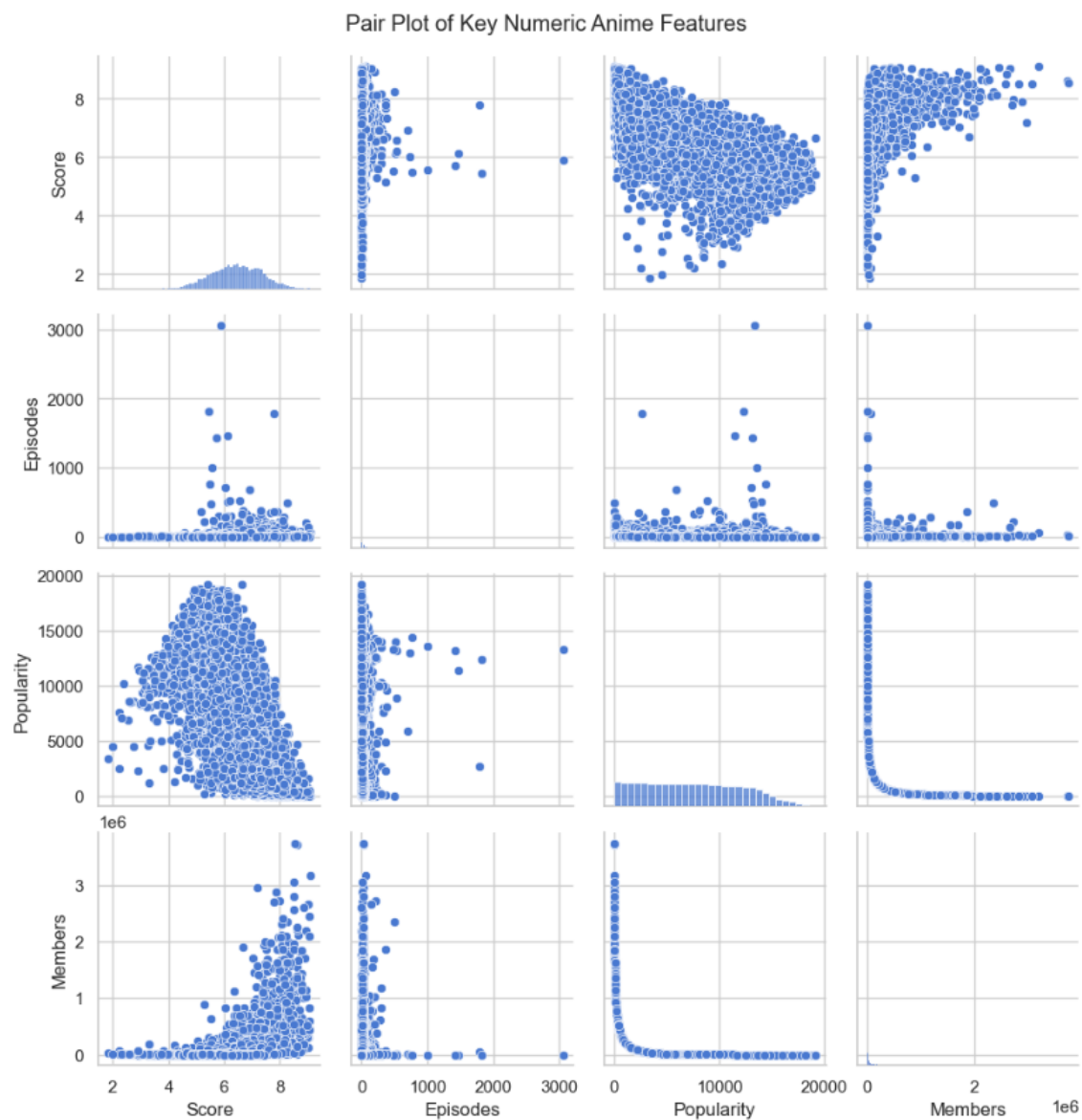
3.1.5 Pair Plot of Key Numeric Features

Figure 12: Pair Plot of Key Numeric Anime Features

This matrix shows relationships among `Episodes`, `Score`, `Popularity`, and `Members`. Note the strong negative correlation between `Popularity` and `Score`, as well as the limited correlation between `Episodes` and `Score`.

3.2 User-Level Relationships

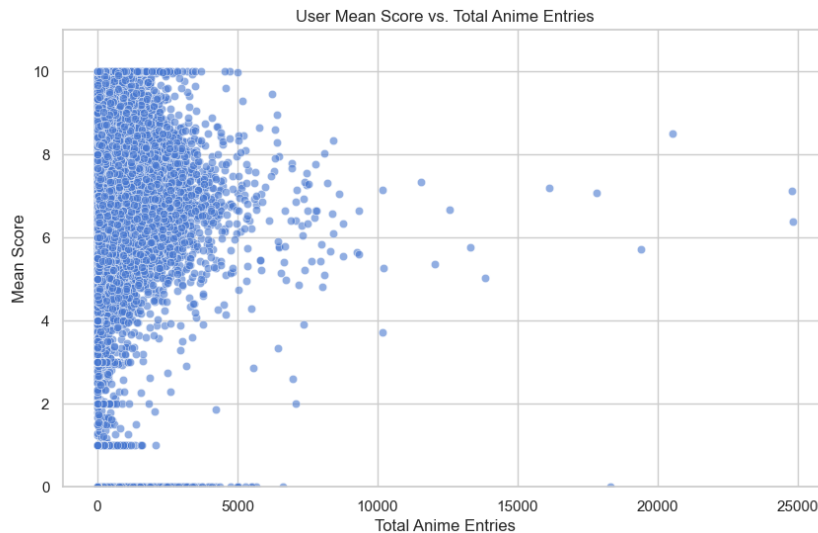### 3.2.1 User Mean Score vs. Total Anime Entries

Figure 13: User Mean Score vs. Total Anime Entries

Highly engaged users can vary widely in mean score—some have completed thousands of entries. But many cluster around **6–8**. Only a small fraction show extremely high or low mean scores and have 10k+ entries.
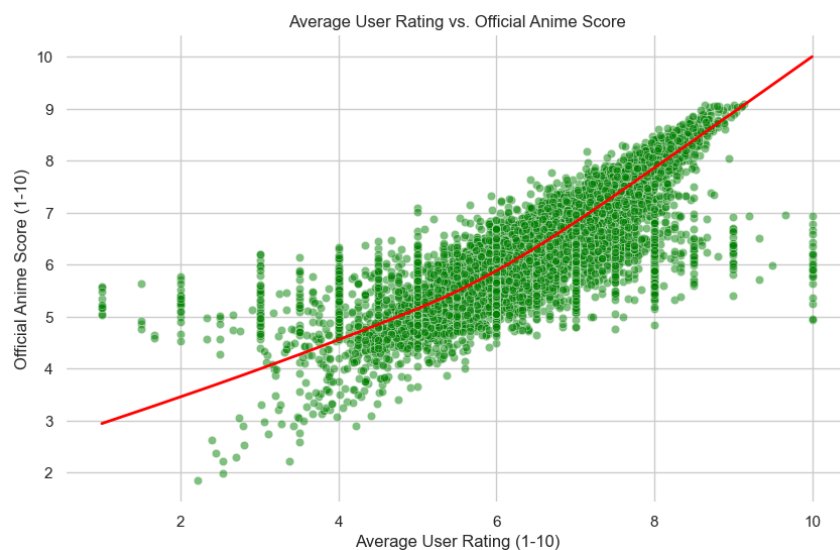
3.2.2 Average User Rating vs. Official Anime Score

Figure 14: Average User Rating vs. Official Anime Score

There is a clear positive relationship between the *average user rating* and the `Score` assigned on MyAnimeList. The locally weighted regression line highlights the generally smooth, *increasing* trend.

4 Top 10 Anime
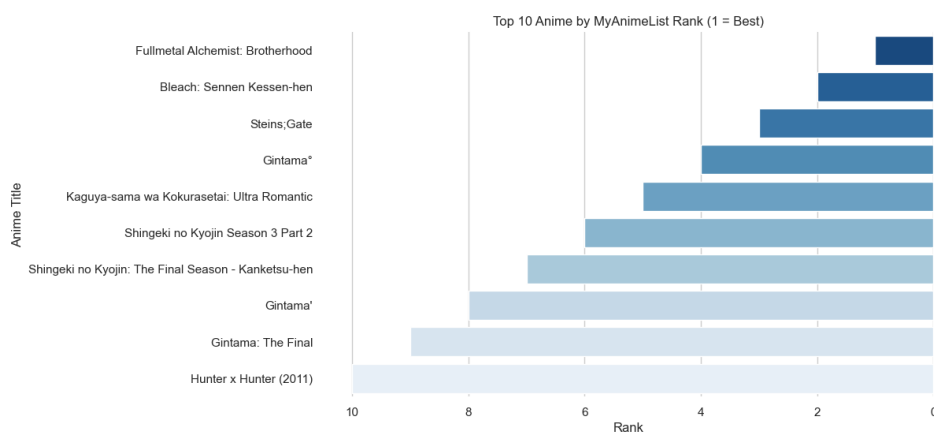
4.1 Top 10 Anime by MyAnimeList Rank

Figure 15: Top 10 Anime by MAL Rank (1 = Best)

Fullmetal Alchemist: Brotherhood, **Bleach: Sennen Kessen-hen**, and **Steins;Gate** are among the most highly ranked shows, reflecting their critical acclaim and strong fan reception.
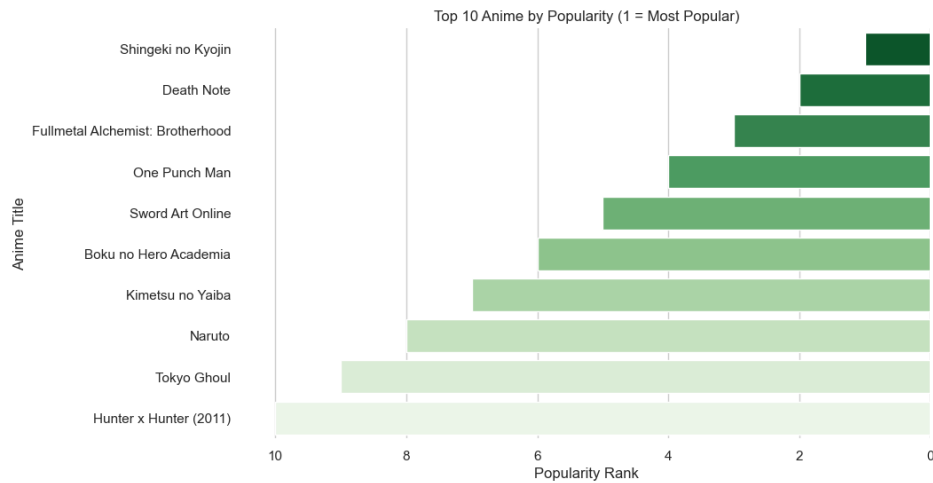
4.2 Top 10 Anime by Popularity

Figure 16: Top 10 Anime by Popularity (1 = Most Popular)

Titles such as **Hunter x Hunter (2011)**, **Tokyo Ghoul**, and **Naruto** rank among the most followed, reflecting their widespread appeal.
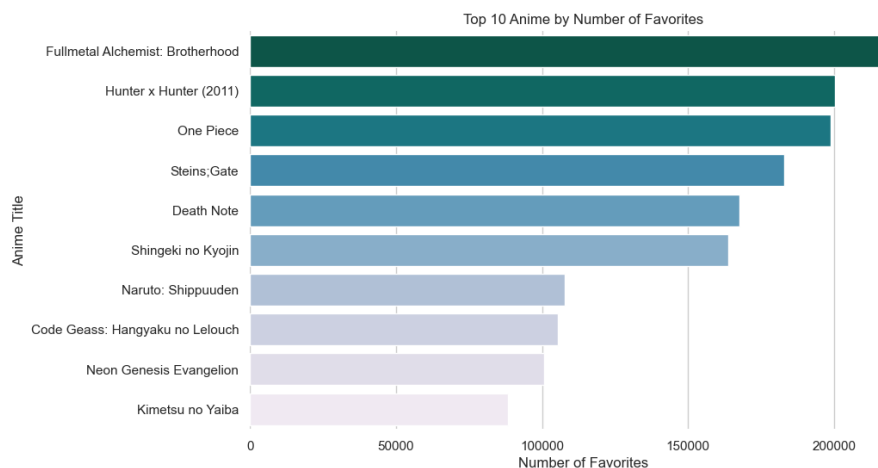
4.3 Top 10 Anime by Number of Favorites

Figure 17: Top 10 Anime by Number of Favorites

Fullmetal Alchemist: Brotherhood, **Hunter x Hunter (2011)**, and **One Piece** top the list, indicating a large, passionate fan base to these series.

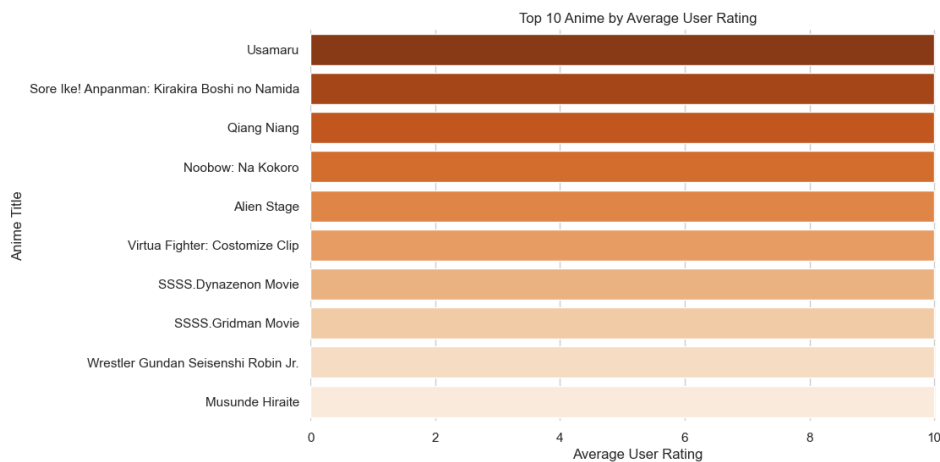## 4.4   Top 10 Anime by Average User Rating

Figure 18: Top 10 Anime by Average User Rating

Some of these entries are relatively unknown short-form or niche productions but have extremely high average ratings among the small group of fans who have seen them.

5   Conclusion

- **Dominant Formats & Genres**: The most common format for anime is TV; the most popular genres are Comedy, Fantasy, and Action.

- **Score Patterns**: Both the official MAL scores and user ratings tend to cluster in the 6–8 range, reflecting an overall positive bias in the fanbase.

- **Popularity & Engagement**: The most popular shows—with the most favorites or members—often score in the 7–9 range.

- **Top Titles**: Well-known series (*Fullmetal Alchemist: Brotherhood, Hunter x Hunter,* etc.) consistently rank among the top-rated, most popular, or most favorited.

In general, this EDA points out the major traits of anime production and their popularity on MAL, offering an avenue for subsequent research, recommendations, and strategic actions within the context of the anime industry.