

Anime Analysis

Flatiron Capstone Project

By: Rachael Bryant

Presented March 8th 10am

Business Problem



- [My Anime List](#) is a website that ranks, discusses, and categorizes anime with the help of their vast community.
 - *Anime is an entertainment medium usually of moving drawings or animations originating from Japan.
- MyAnimeList wants to know how to write their synopses in order for anime to have higher scores from viewers. Specifically **whether any aspect of the description meant for marketing influences the scoring of an anime**. They would also like to know whether their recommendation system is working well with their opinion based recommendations.



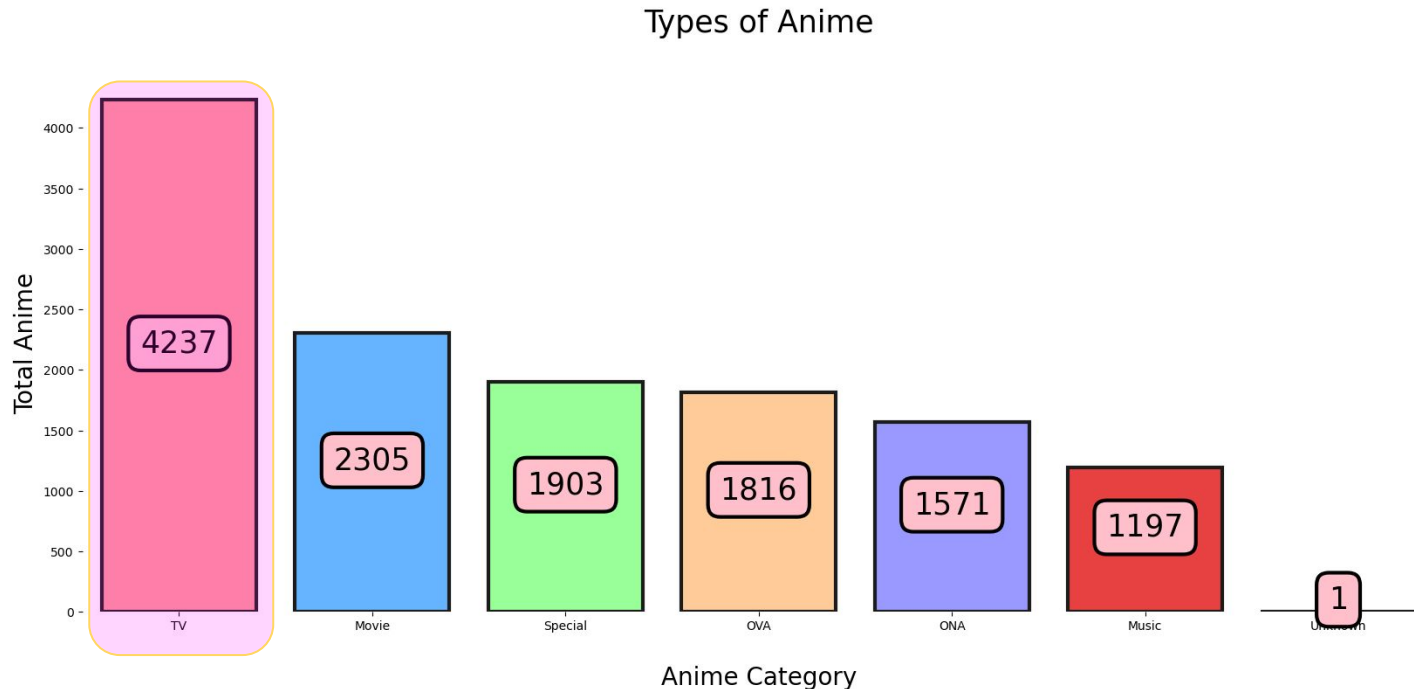
Data



- The primary dataset was sourced from Kaggle. It was scrapped from the website My Anime List.
- About 13,000 records
 - Including synopsis of the anime, rank, rating, popularity, and genres.

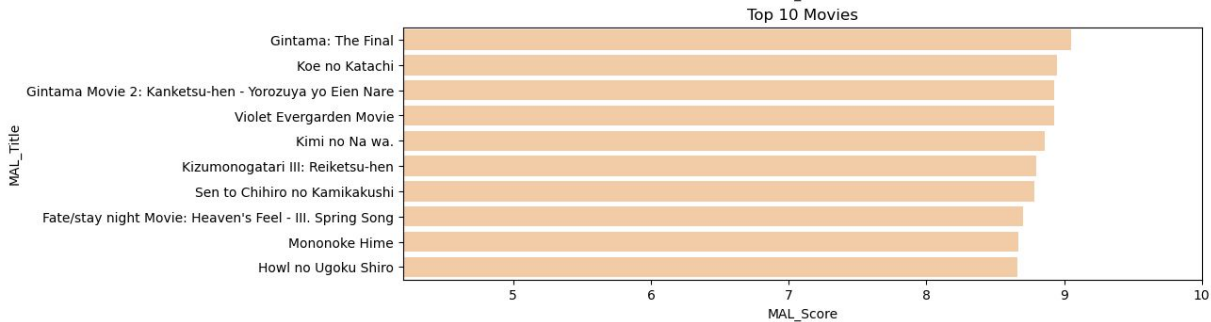
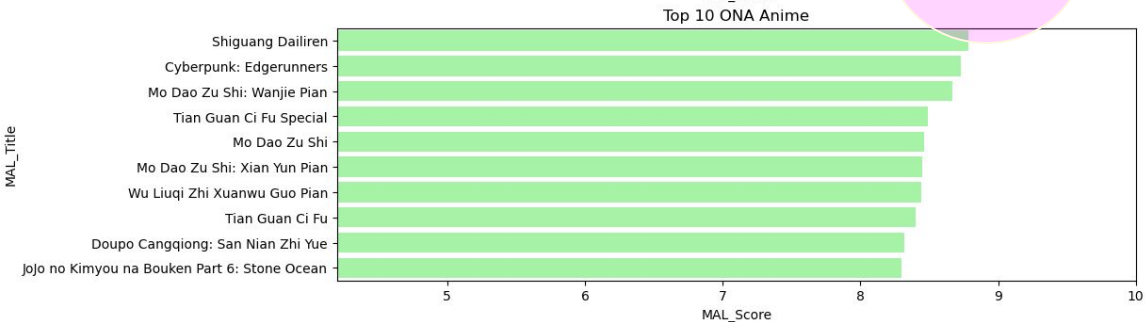
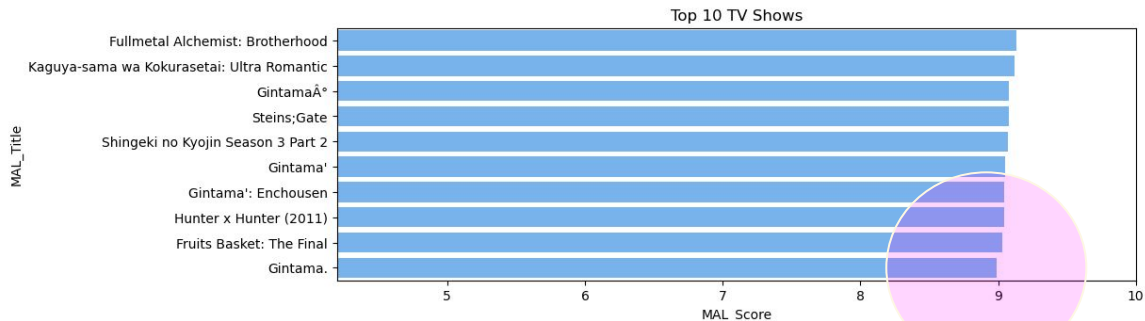
Types of Anime

- TV is the biggest category



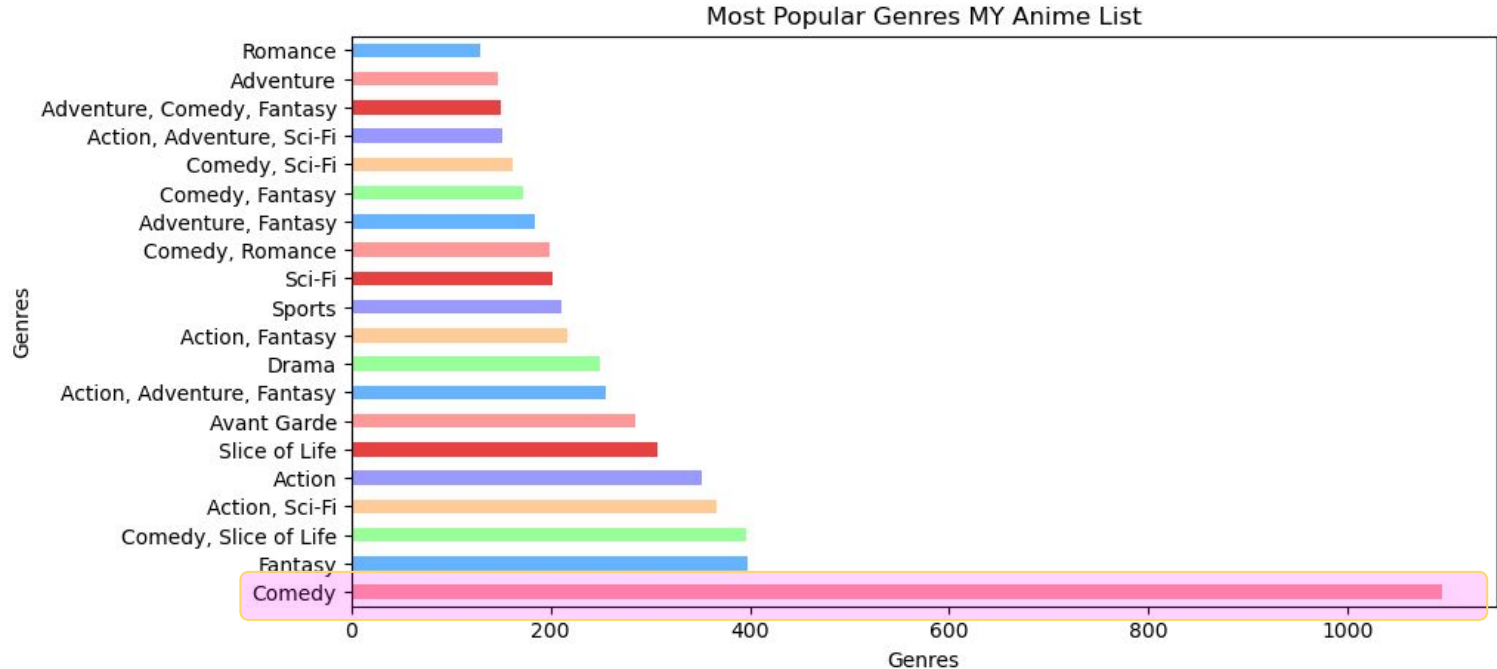
Top anime per Type

- The top anime of TV, ONA and Movies are listed.
- TV anime are more highly rated in general.



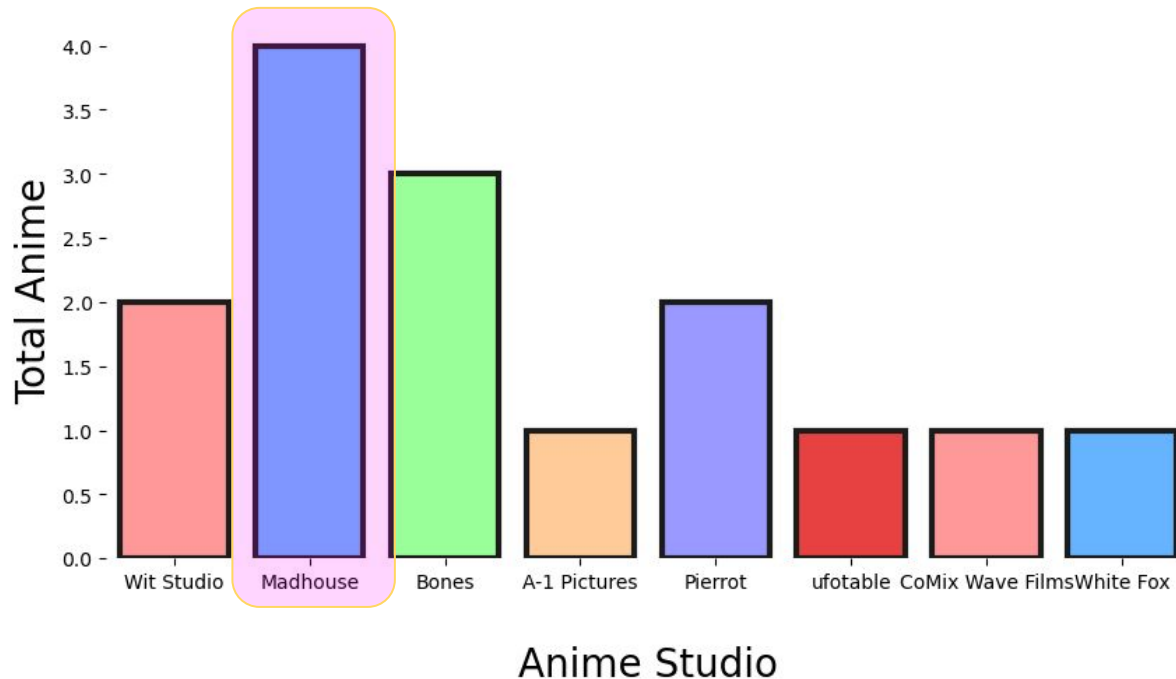
Focusing Genre. How many anime per genre

- Comedy has the most anime



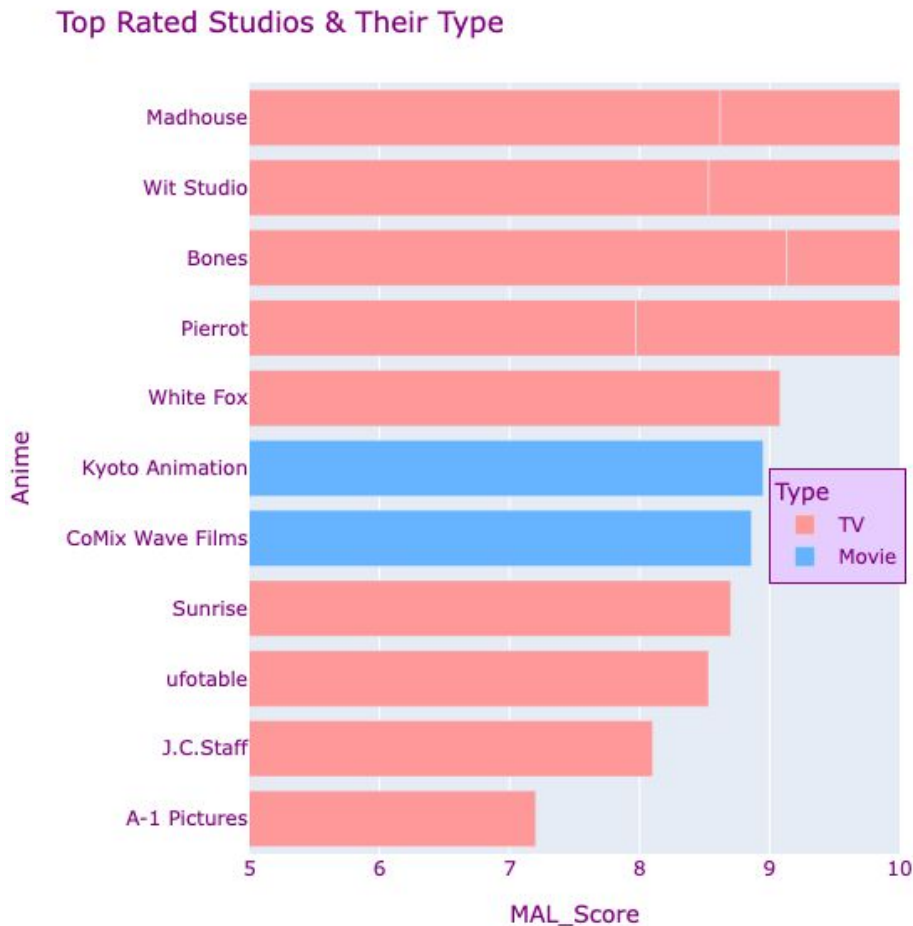
Studios that produced the Anime

- Madhouse Studio has the most anime recorded
-



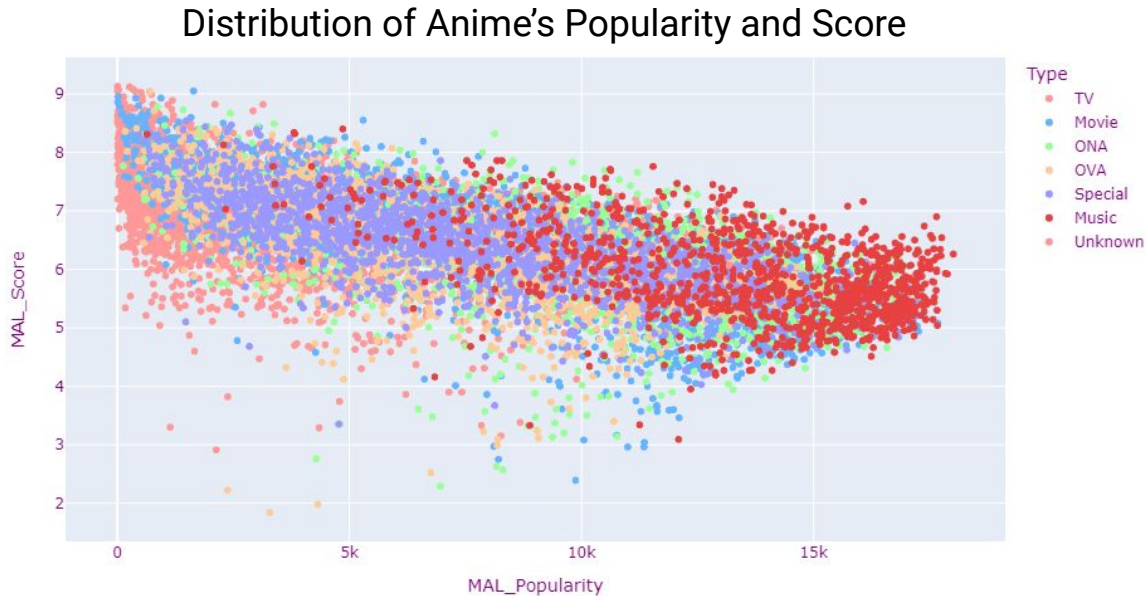
Type of anime that Studios produce

- The most highly scored studios are **Madhouse**, **Wit Studio** and **Bones**. They reflect the highest rated anime in general.



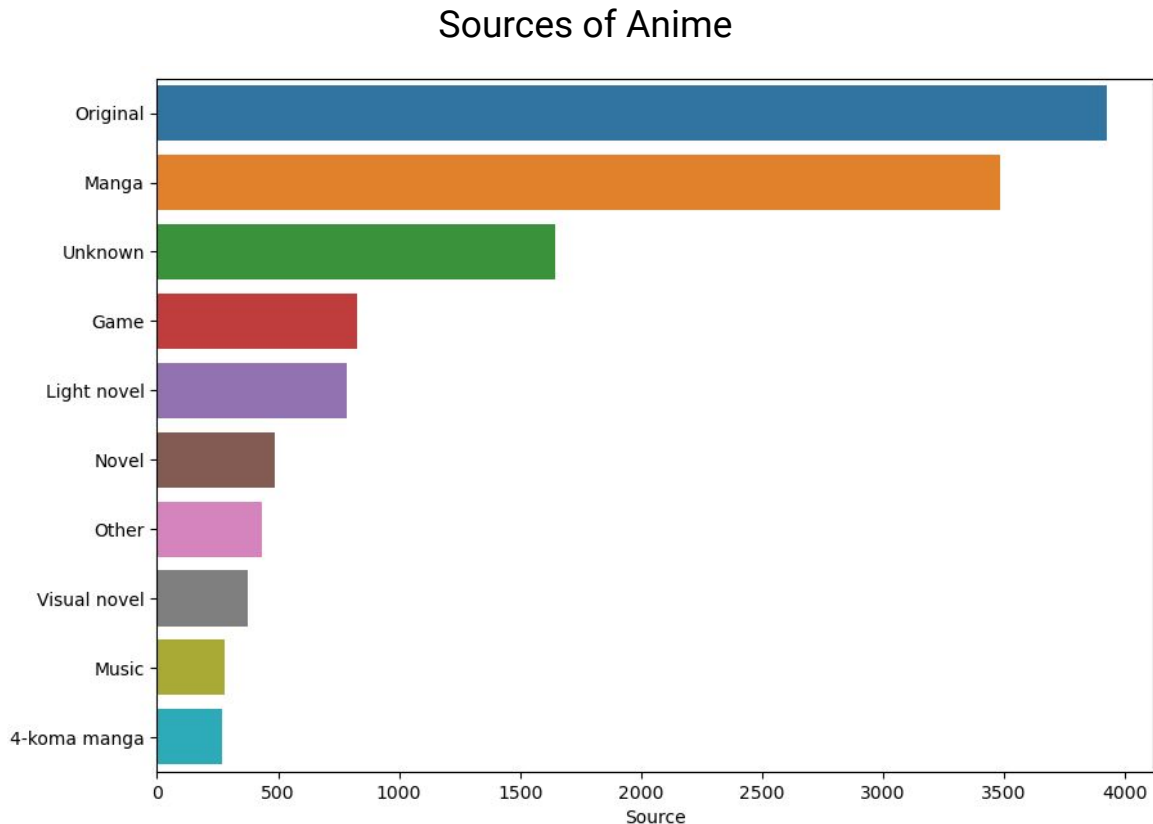
Popularity in relation to Scores of Anime

- The lower ranked anime and more highly scored overall are TV and movie.

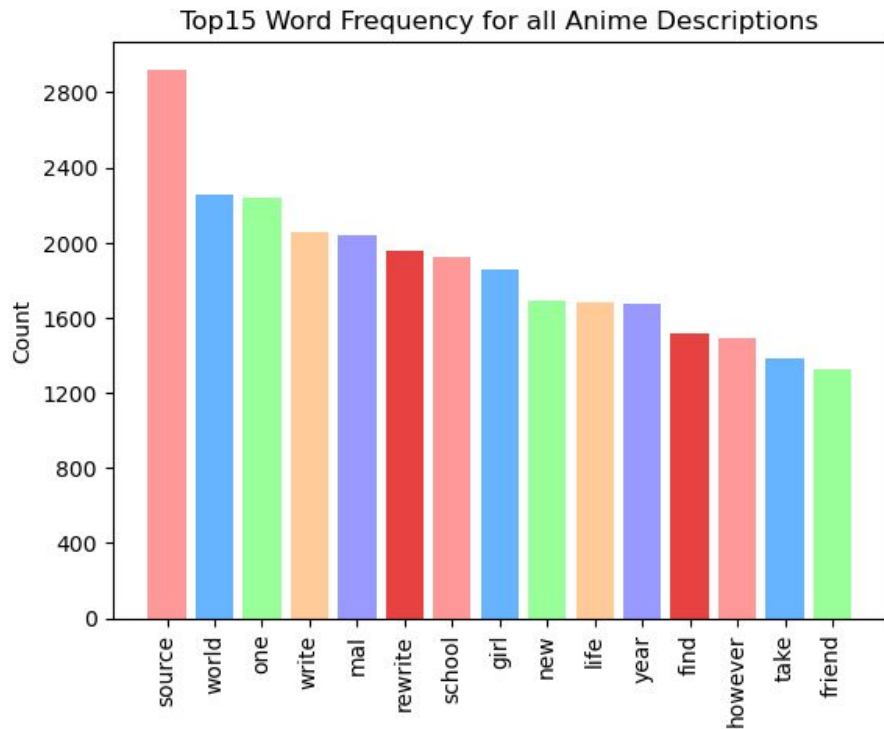


Sources of Anime

- More anime are created based on original ideas than manga.

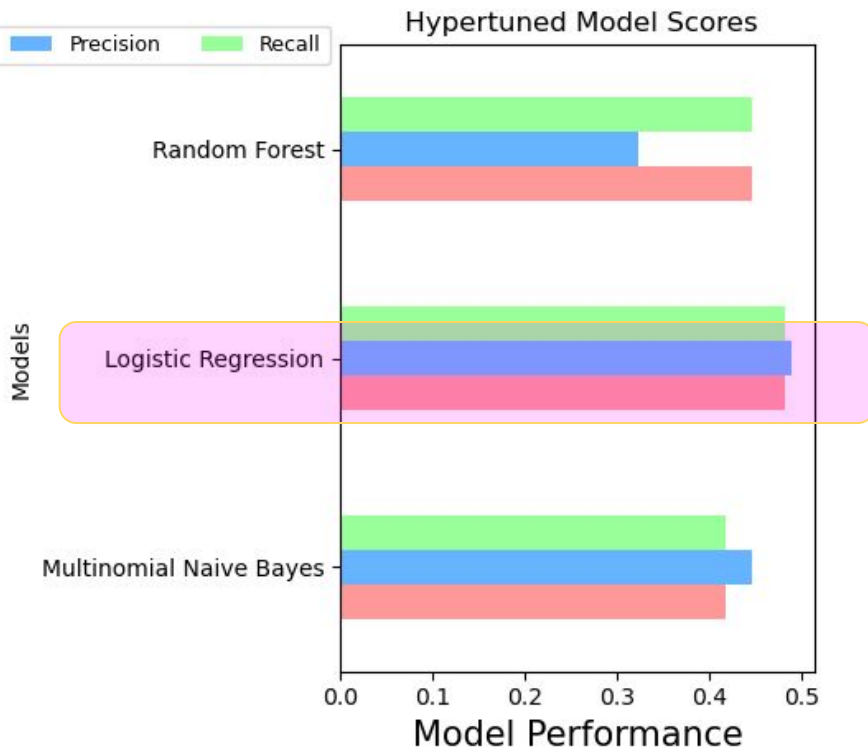


Most Frequent Words Across the descriptions



- Source, world, one, write, and mal (stands for My Anime List) were the most commonly used words

Best models



- Many models were performed upon the data to find the ones that were best at predicting the score based on the synopses. However, all performed poorly.

Recommendation System

- The recommendation system takes in the **genre** and provides similar anime and lists the **rating** for reference.

```
give_recommendation("Death Note")
```

Recommendations for Death Note viewers :


No	Anime Name	Rating
1	Munou na Nana	7.191000
2	Death Note: Rewrite	7.701000
3	Warau Salesman Special Program	6.161000
4	Zhongguo Jingqi Xiansheng	6.071000
5	Mahou Shoujo Ikusei Keikaku	6.961000
6	Higurashi no Naku Koro ni Kai	8.191000
7	Summertime Render	8.491000
8	Higurashi no Naku Koro ni Rei	7.411000
9	Mouryou no Hako	7.151000
10	Mouryou no Hako: Chuuzenji Atsuko no Jikenbo - Hako no Yurei no Koto	6.601000

In conclusion

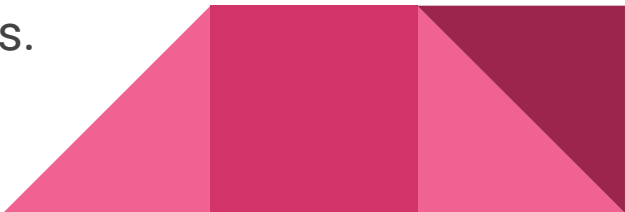
- Although these models could be used to predict the score of an anime due to the synopsis, **they are not deemed to be optimal models with low reliability** in accuracy, precision, and recall. The hypertuning performed and the unsupervised K-means modeling did not improve the models enough to perform them upon new data.
- The exploratory **analysis exposed interesting trends** in the most commonly listed Studio, Source of the anime, Type of anime, as well as genres and themes. These can be used in marketing and improving the site for visitors.
- A **recommendation model** was created in order to display the best next anime for someone who enjoyed the anime they inputted. It also helpfully has the score of the anime there for those who find importance in scores.



Recommendations

- The most frequent words, especially the top 10 of 'source', 'one', 'world', 'write', 'mal', 'rewrite', 'school', 'girl', 'new', 'life' can be utilized to better understand that many anime are based in schools with a high number of girl characters. This can be used for **marketing certain anime** or to ensure that a synopsis will meet certain criteria by including these words.
 - It is recommended to **fill in many Nan and Null data and 'unknown' items** those in with meaningful information in order to perform a better analysis of the entire dataset.
 - The **recommendation model** created could be used to better ensure that the recommendations on MyAnimeList site are up to date and make sense according to the dataset and not just someone's opinion.
 - **Utilizing other aspects of anime**, such as box office gross winnings, could be implemented into MyAnimeList for more variety of information.
- 

Next Steps

- I would like to **combine this dataset with more data** from other anime list and recommendation sites, such as Anime_Planet or Anilist. Since they do not allow web scraping, I would reach out for the information.
 - I would like to look at the **other information on MyAnimeList such as Manga**, Manhwa, and Manhwa (Japanese, Chinese, and Korean comics) to understand their patterns, modeling and make a recommendation system for them as well.
 - I would **utilize other hypertuned models after adding more data** to model genres upon the synopses to create useful models.
- 

Thank you for your time

Rachael Bryant

Email: rachaelbryant94@gmail.com

Github: @rabrya0072



***Made with word cloud in Python**