Date: April 23, 2019
To: Levi Ackerman, Senior Data Analyst, Janzen Consulting Group
From: Andrew Garcia, Assistant to the Senior Data Analyst
RE: How different factors affect MyAnimeList anime score

In 2017, Makoto Shinkai's *Your Name.* became the highest grossing anime film worldwide, with a total of US$281 million.[1] The American anime localization company Funimation is looking for the next popular anime film to bring to the US. To help with this decision, they are analyzing the website MyAnimeList, which ranks anime by popularity according to users.

I have been tasked to see how various factors affect the anime score (out of ten) on the website. My findings will help Funimation decide if they should localize popular anime movies. I hypothesize that anime popularity rank, overall rank, the number of users who add a title to their list, and the number of users who add a title to their favorites affect the average score of an anime. This memo will confirm that these factors do affect average anime scores.

To complete this task, I analyzed MyAnimeList data as of June 2018. The data was compiled by Matěj Račinský, a data scientist from Prague, Czechia.[2] The MyAnimeList data represents individual anime and the fans of anime who use the website. In total, the data consists of 14,487 anime and 302,675 unique users. I narrowed down total anime to 2,530 anime movies for the purposes of my analysis.

Popularity rank on MyAnimeList is determined by the number of users who have added the title to their list. The more users who add an anime to their list, the more popular the anime is. A lower rank number indicates a higher popularity rank, so the most popular anime on the site would have a rank of one.

Table 1: Summary Statistics

	Freq	Min	Max	Med.	Mean	SD
Popularity	2,530.00	33.00	14,487.00	9,116.50	8,196.55	4,018.72
Rank	2,530.00	2.00	12,860.00	7,318.00	6,759.98	3,868.18
Favorites	2,530.00	0.00	34,912.00	1.00	113.71	1,016.91
Members	2,530.00	9.00	730,076.00	634.50	14,015.33	45,657.01
Score	2,530.00	2.52	9.20	6.30	6.23	1.17

Overall rank is the weighted rank on based on the top anime page of MyAnimeList. It consists of roughly 13,000 anime, with each anime having its own individual rank. Since it excludes "not yet aired" and rated "R18+" titles, there are less ranks than total anime on the site. Additionally, some anime are

[1] Pineda, Rafael A. (2017, January 17). Shinkai's 'your name.' Tops Spirited Away as Highest Grossing Anime Film Worldwide. *Anime News Network.* Retrieved from https://www.animenewsnetwork.com/news/2017-01-17/shinkai-your-name-tops-spirited-away-as-highest-grossing-anime-film-worldwide/.111141
[2] Račinský, Matěj. Github profile. *Github.* Retrieved from https://github.com/racinmat

unranked since not enough users have given it a score. I have removed these unranked anime from my analysis. Like popularity rank, lower rank numbers indicate a higher rank.

Each anime also has a certain number of users who add the title to their list and/or favorites list. This is represented by the favorites and members variables in the data. Since less users use the favorites list compared to the main list, it may not be as representative of the population compared to members. It is also possible for an anime to have no users who have added it to their favorites.

Additionally, each anime has a score out of ten, which is the average of a weighted score rated by users. The lowest score the user can rate an anime is one, and the highest score is ten. Anime which have not aired yet are not given a score (N/A). In the data, these unreleased anime have a score of zero. Since it is not possible for a user to give a score of zero, I removed the anime with no score from the data. Thus, anime movies that have not aired as of June 2018 will not be included in this analysis.

Table 2: Regression Results

	Dependent variable:
	Average Anime Movie Score
Popularity	-0.00001
	(0.00001)
Rank	-0.0002***
	(0.00001)
Members	0.00000***
	(0.00000)
Favorites	-0.00002
	(0.00002)
Constant	7.744***
	(0.043)
Observations	2,530
R^2	0.646
Adjusted R^2	0.646
Residual Std. Error	0.698 (df = 2525)
F Statistic	1,153.684*** (df = 4; 2525)
Note:	*p<0.1; **p<0.05; ***p<0.01

Upon checking for regression assumptions, I found that my model violated normality, independent errors, and homoscedasticity. This negatively effects hypothesis test, confidence interval, and standard error reliability. There were also multicollinearity issues due to the popularity variable being too correlated. However, the assumption would not be violated if popularity was removed from the model. Despite the assumption violations, I still performed a multiple OLS regression.

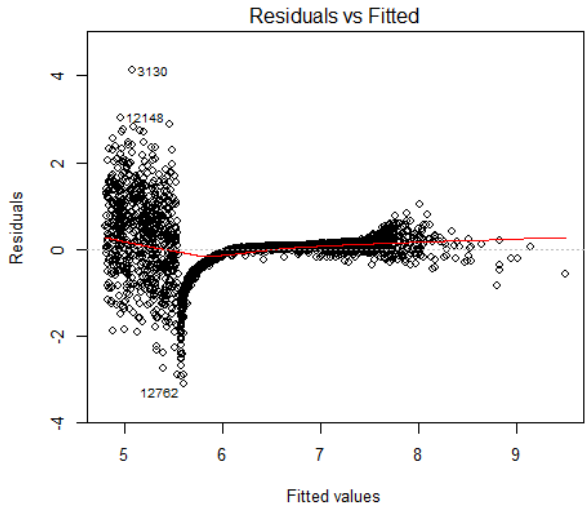
I found that my model of score = 7.744 - 0.00001(popularity) - 0.0002(rank) + 0(members) - 0.00002(favorites) + ε to be statistically significant at an alpha of 0.05. For each coefficient, I will be holding all other variables equal. When all other predictors are 0, the average anime score is 7.74, $p < 0.05$. When popularity rank increases by one, the average anime score decreases by 0.00001, $p > 0.05$. Each increase in rank corresponds to an average anime score decrease of 0.0002, $p < 0.05$. For each additional member that adds an anime to their list, there is approximately no change to the average anime score, $p < 0.05$. For each additional member that adds an anime to their favorites, the average anime score decreases by

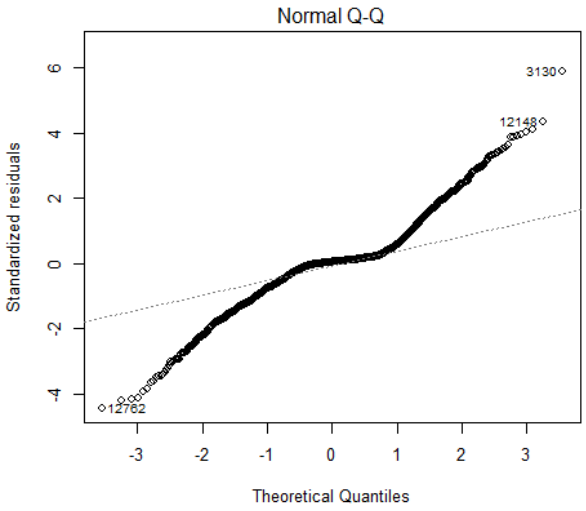
0.00002, $p > 0.05$. An adjusted R^2 value of 0.65 shows that 65% of variance in anime score is accounted for by the predictor variables. A Cohen's f^2 value of 1.8 is a huge effect size.

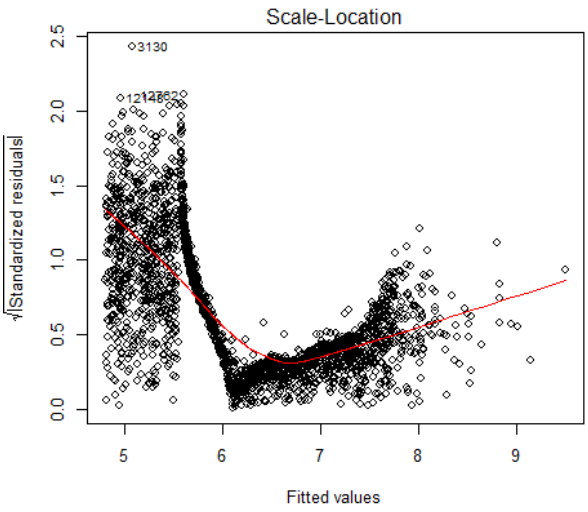
The results support my hypothesis that that anime popularity rank, overall rank, the number of users who add a title to their list, and the number of users who add a title to their favorites affect the average score of an anime. From this result, I recommend that Funimation brings popular anime movies to the United States. These findings could be generalized to the population of released anime that are ranked and scored on MyAnimeList, since movies are only a subset of anime on the site.

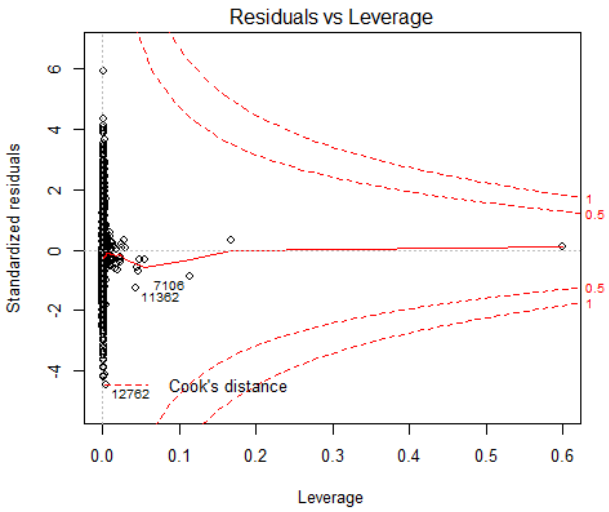
Although my model is statistically significant, it is not substantively significant. Since there are approximately 15,000 anime on MyAnimeList, each with an individual rank and popularity rank, it will take an increase or decrease of thousands of ranks to make a one point change in anime score. It will also take thousands of users adding an anime to their list or favorites to be able to affect the anime score by one point. Because it will take thousands of rank changes or users to make a meaningful change, none of the model coefficients are substantially significant.

One limitation of this study is that the data is almost a year old, so it cannot account for anime movies that have aired since June 2018. Another limitation is that the study does not account for users that might have more than one account. Notionally, someone may try to make multiple accounts to attempt to review bomb a specific or popular anime title by giving it negative reviews with each account. Alternatively, they could add a title to their list and favorites to try and inflate its score. Although this is unlikely to skew the score, popularity rank, or rank of an anime, further analysis into MyAnimeList users may be able to answer this.


Residuals vs Fitted
Residuals
Fitted values
3130
12148
12762


Normal Q-Q
Standardized residuals
Theoretical Quantiles
3130
12148
12762


Scale-Location
√|Standardized residuals|
Fitted values
3130
12148
12762


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
7106
11362
12762