Decoding Anime Ratings: Insights from MyAnimeList Data

1. Motivation

Why do certain anime receive higher user scores on MyAnimeList (MAL), and what characteristics of anime, as well as user demographic and behavioral aspects, contribute to these ratings? This analysis is crucial for elevating the anime-watching experience and optimizing content offerings for industry stakeholders. Anime, with its diverse genres and a passionate global fanbase, provides a unique opportunity to explore consumer behavior. Investigating the reasons behind higher user ratings offers valuable insights into content characteristics that resonate with viewers, shaping preferences and influencing engagement. For producers and streaming services hosting anime content, understanding these dynamics is critical for strategic content creation and marketing, enhancing audience satisfaction and overall success in a competitive landscape.

2. Dataset

The primary dataset for our analysis is the "MyAnimeList Datasets - 2023" from Kaggle (DSFelix, 2024), collected between August 11th and October 6th, 2023. Comprising information on 20,000+ anime entries, 700,000+ registered users, and 24,325,191 user-anime interactions, this dataset from MAL is ideal for addressing our question of interest. Leveraging MAL's extensive user base, actively involved in tracking, discussions, and recommendations, it provides substantial data on user demographics and behavior. With features like average user score, user gender, genre, type of animation, episode count, and duration, we can thoroughly investigate the characteristics, popularity, and viewership of various animes. Access to scores from 270,033 users enables us to delve into user interactions and analyze rating patterns. The dataset's depth makes it suitable for discerning patterns influencing user ratings. Figure 1 in Section 5.1: *Suppl. Exploratory Data Analysis* provides an initial visual overview of some relevant features in the dataset, offering insights into attribute distributions. For a comprehensive feature list with corresponding descriptions, refer to Section 5.11: *Suppl. Dataset Features.*

3. Milestone 1

This milestone explores factors influencing higher user scores on MyAnimeList (MAL), examining anime characteristics, user demographics, and behaviors. It analyzes the impact of anime genre, type, source, episode count, and user gender on average user ratings. Methodologically, it utilizes Kruskal-Wallis and Mann-Whitney U tests to identify significant differences, alongside correlation and regression analyses for predictive insights. In addition to generated plots for these analyses, Section 5: *Supplementary Materials* offers a closer look at the average score distributions across various subgroups (Section 5.2: *Suppl. Score Distributions*).

3.1 Methodology

Prior to analysis, the anime dataset underwent pre-processing steps, including one-hot encoding of genres and types, as well as the removal of rows with missing or invalid data. In examining the user dataset, which includes gender and features related to watching behavior, data filtering was performed to ensure robustness, removing null entries and users with zero watch history. Exclusion criteria also considered users with scores falling below 1, and required that these users had a *total entries* value greater than 1 in an attempt to preserve raw behavioral data and maintain integrity. The selection of the tests to be used for the analyses was guided by several assumptions tailored to the dataset's characteristics. To cut down on false positives, the alpha level of significance was defined as 0.005 (Benjamin et al., 2017). Given the ordinal nature of user scores, reducing groupings to means was deemed inappropriate. Instead, the rank order of user scores was interpreted, and significance tests suitable for ordinal data, including the Kruskal-Wallis test (H) and Mann-Whitney

U test (U), were utilized. To address the issue of multiple pairwise comparisons in the U tests, the Bonferroni correction was applied. This involved adjusting the level of significance by dividing it by the number of unique combinations being tested. Additionally, p values, effect sizes using Cliff's δ (the ordinal data equivalent of Cohen's d) (Macbeth et al., 2011), and power for each pairwise comparison conducted in the U tests were generated. Power was computed via simulation: synthetic data was generated under the null hypothesis, and the proportion of simulations yielding significant results at the specified alpha level was calculated.

Hypothesis testing was conducted to explore differences in anime scores across various types (e.g., TV, movie, etc.) and sources (e.g., manga, original, etc.) using H and U tests. Furthermore, a gender-specific analysis was performed to examine the influence of user gender on anime scores. Correlations between episode count and average user scores across different anime types were assessed. Outliers were identified and removed using the interquartile range (IQR) method to ensure the robustness of estimates. Regression analyses were conducted to understand how genre influences average user scores, employing Ordinary Least Squares (OLS) regression with regression coefficients tested for significance. Additionally, the predictive capability of genre and anime type on average scores was explored using OLS regression. An 80/20 train/test data split was employed for model evaluation, generating Mean Squared Error (MSE) and R^2 scores. It's important to acknowledge potential biases arising from assumptions about the data distribution and filtering methods employed.

3.2 Results

Hypothesis testing aimed to identify significant differences in user scores among anime types and sources. The H test revealed significant evidence ($H = 1247, p = 1.214 \times 10^{-268}$) of differences among anime types, with subsequent pairwise U tests confirming significant differences in several comparisons. Notable comparisons such as 'TV' vs 'Original Net Animation' (ONA) ($p = 2.414 \times 10^{-222}$, $\delta = 0.508$, power $= 1$), 'TV' vs 'Original Video Animation' (OVA) ($p = 2.670 \times 10^{-97}$, $\delta = 0.333$, power $= 1$), and 'TV' vs 'Special' ($p = 1.644 \times 10^{-101}$, $\delta = 0.33$, power $= 1$) exhibited moderate to large effect sizes and high power. In Cliff's δ, a moderate effect size ($0.33 \leq |\delta| < 0.474$) suggests a noticeable but not statistically significant divergence in score distributions, while a large effect size ($|\delta| \geq 0.474$) indicates a more substantial distinction (Wan et al., 2019). High power ($1 - \beta \geq 0.8$) indicates a greater likelihood of detecting true effects, reducing the risk of false negatives and bolstering confidence in study results. Significant differences in user scores among different anime sources were indicated by the H test ($H = 2144, p = 0.0$), with pairwise U tests further confirming these differences. Notable pairs, excluding 'Other' or 'Unknown' sources due to their ambiguous categorization, revealed significant score differences and substantial effect sizes and power. For example, 'Manga' vs 'Original' ($p = 5.831 \times 10^{-224}$, $\delta = 0.436$, power $= 1$) and 'Visual Novel' vs 'Light Novel' ($p = 2.26 \times 10^{-21}$, $\delta = 0.34$, power $= 0.894$) demonstrated these findings. The analysis also explored the influence of user gender on anime scores. Results from the H test revealed significant differences ($H = 2025, p = 0.0$) in user scores based on gender. Subsequent pairwise U tests indicated that while all gender pairs had p values below the adjusted significance level, the effect sizes were generally small or negligible. Heatmaps depicting the magnitudes of the p values, power values, and effect sizes for all the pairwise U tests generated can be found in Figures 4, 5, and 6 in Section 5.3: *Suppl. Hypothesis Testing.*

Correlation analyses examined the relationship between episode count and average scores across different anime types. Correlation coefficients varied across types, with notable correlations observed for 'ONA' ($\rho = 0.38$) and 'OVA' ($\rho = 0.17$), indicating moderate positive correlations between episode count and user scores for these types. Figure 7 in Section 5.4: *Suppl. Correlations* illustrates all generated correlations between anime type and episode counts.

Regression analysis was conducted to understand the influence of genre on average scores. All genres exhibited significant coefficients ($p < 0.005$) except for the 'Boys Love' genre. The OLS regression analysis revealed varying predictive capabilities for average user scores based on genre and anime type. Assessments were made using genre alone and in combination with anime types, with mean squared error and R^2 scores calculated for each genre. Results indicated that only a few predictions based solely on genre achieved R^2 scores above 0.3 ('Horror' and 'Avant-Garde'), while a limited number of type-genre combinations surpassed an R^2 of 0.5 ('ONA-Horror', 'ONA-Suspense', 'ONA-Ecchi', 'TV-Avant Garde'). Detailed summaries of the outputs for the OLS regression models are presented in Figures 8, 9, and 10 in Section 5.5: *Suppl. Regression.*

3.3 Analysis

In the examination of user scores across different types and sources of anime, significant variations in scoring patterns were discovered, particularly when comparing mainstream 'TV' anime with 'Specials', 'ONA's, and 'OVA's. These differences suggest that content aimed at dedicated or niche audiences tends to elicit distinct scoring behaviors. Similarly, it was found that the source material of an anime—whether it's from a novel, manga, original, or lesser-known sources—can influence scoring patterns, with more pronounced differences observed for mainstream sources. However, the impact of user gender on anime scores, while statistically significant, showed relatively small practical differences, indicating subtle variations in scoring preferences that would require further investigation. Overall, the findings emphasize the importance of considering content types, sources, and user demographic factors in analyzing average user scores within the anime community.

In the correlation analysis between the number of episodes in an anime and average scores, examination was focused on how episode counts relate to scores within specific medium types. Positive correlation coefficients indicate that as episode counts increase, scores tend to rise, while negative correlations suggest the opposite. Movies, with only one episode, exhibited no correlation. TV anime displayed a weak positive correlation of 0.07, likely due to varied episode counts. 'Specials' showed minimal association, possibly due to their limited episodes. 'OVA's displayed a moderate positive correlation of 0.17, indicating a relationship with scores, albeit with variability. 'ONA's exhibited the most notable correlation at 0.38, suggesting a pronounced relationship between episode count and average scores. Further investigation is warranted to understand the nuances driving this relationship, potentially offering insights into viewer preferences within the 'ONA' category.

In the initial regression analysis, hypothesis testing on the coefficients of each genre was conducted without dropping a column from the feature set. This approach enabled assessment of the impact of each genre on the score independently, providing insights into their individual effects. "Boys Love" was the only genre with a non-significant p value, suggesting it has minimal influence compared to others. Most genres had positive coefficients, indicating a positive effect on scores, except for 'Horror," 'Avant-Garde', and 'Ecchi', which tended to receive lower scores. This suggests certain genres are more favorably received, potentially impacting overall success. Predictability within genre groups was then examined, revealing that only 'Horror' and 'Avant-Garde' had R^2 scores above 0.3, showing some predictability. Expanding to genre and anime type combinations, only four combinations had R^2 scores above 0.5, indicating higher predictability. Notably, 'ONA's, particularly those labeled as 'Horror', 'Suspense', and 'Ecchi', showed high predictability, likely due to their ability to evoke consistent viewer responses through intense plotlines and thematic elements, suggests a strong relationship between genre, type, and scores. However, limitations include small sample sizes and the influence of other factors beyond genre and type. This underscores the need for more advanced modeling and consideration of additional variables for improved accuracy. The prevalence of 'ONA's in the highly predictable group suggests unique viewer behaviors specific to this category, requiring further investigation.

To summarize, the analysis uncovered significant variations in scoring patterns across different anime types and sources. Niche content tends to show distinct scoring behaviors, while mainstream source material influences scores. Though gender differences were statistically significant, they were practically minor. The correlation and regression analyses revealed nuanced preferences and predictability within certain genres and anime types. These findings highlight the complexity of factors influencing user scores in the anime community, emphasizing the need for comprehensive consideration of content types, sources, and demographic factors in future analyses.

4. Milestone 2

This milestone sets out to explore the intricate dynamics of anime scoring. With access to a mixture of categorical and numerical anime features, this study aims to forecast average user scores. Additionally, it delves into uncovering latent structures and inherent clusters among users, derived from their viewing habits. Methodologically, it utilizes Ordinary Least Squares (OLS) Regression, Gradient Boosted (GB) and Random Forest (RF) Regressor models to project an anime's average score. Furthermore, it leverages Principal Component Analysis (PCA) and K-Means clustering to gain insights into the underlying structures and patterns within the user dataset, providing a comprehensive understanding of anime scoring dynamics and user behavior on MyAnimeList (MAL).

4.1 Methodology

The anime dataset underwent filtering of anime entries with missing or invalid values (given as -1) for *score*, *episodes*, and *rank*. To streamline the analysis and address complexities associated with *genre* features, which could have multiple labels per entry, these features were excluded. The remaining features were *type, episodes, status, producers, licensors, studios, source, rating, favorites, scored by,* and *member*. Definitions for these features can be found in Section 5.11: *Suppl. Dataset Features.* Categorical features were pre-processed to handle unknown and infrequent values by assigning them to a new variable labeled 'Other/Unknown'. Subsequently, one-hot encoding was applied, with the 'Other/Unknown' category within each feature dropped to serve as the reference feature for regression analyses. Additionally, 'currently airing' was dropped from the *status* feature to establish it as the reference variable due to its minority representation. These steps were taken to prevent multicollinearity issues in the regression analyses; it ensures that the encoded features are linearly independent and avoids redundancy, making the data suitable for OLS regression and compatible with ensemble regression methods. Numerical features *favorites, scored by,* and *members* were used to create the following relative user engagement metrics for each anime: *favorites per member, scored by per member, favorites per entry,* and *scored by per entry.* Prior to regression analyses, correlation among independent features was assessed to identify multicollinearity. An 80/20 train/test data split was employed for regression model evaluation, generating Mean Squared Error (MSE) and R^2 scores. In the OLS regression model, coefficients were tested for significance using an alpha level set at 0.005 (Benjamin et al., 2017) to reduce false positives. For tuning the RF Regressor model to forecast average scores, GridSearchCV was utilized to find optimal hyperparameter values for number of estimators, max depth, minimum samples per split, and minimum samples per leaf. Feature importances were computed for all ensemble regression models to identify the top 10 features.

In examining the user dataset, which includes user gender and features related to watch behavior, data filtering was performed to ensure robustness, removing null entries and users with zero watch history. Exclusion criteria also considered users with scores falling below 1, and required that these users had a *total entries* value greater than 1 in an attempt to preserve raw behavioral data and maintain integrity. The user features of interest were: *days watched, episodes watched, watching, completed, on hold, dropped, plan to watch,* and *rewatched.* Outliers in each feature were identified as values beyond 3 times the IQR, above Q3 or below Q1, and subsequently removed from the dataset. User *mean score* was reformatted into four score quantiles, with the 1st quantile representing the lowest group of mean scores and so forth. Before conducting PCA, correlation among independent features was examined to detect multicollinearity. The Kaiser criterion guided the selection of principal components, retaining only those with eigenvalues greater than 1.0. Data points were visualized after dimensionality reduction with PCA and color coded by either score quantiles or user gender. Silhouette analysis determined the optimal number of clusters for K-Means clustering and corresponding clusters were visualized. It's important to acknowledge potential biases arising from assumptions about the data distribution and filtering methods employed.

4.2 Results

Various regression analyses were conducted to predict average anime scores using both numerical and categorical features from the anime dataset. Initial correlation analysis revealed strong correlations ($\rho > 0.7$) among features *favorites, scored by,* and *members*, as well as among engineered metrics like *favorites per entry* and *scored by per entry* (see Section 5.6: *Suppl. Anime Dataset Correlations*). The OLS Regression model yielded an R^2 value of 0.515 and MSE of 0.431, with approximately 65% of coefficients deemed statistically significant. Noteworthy relationships between independent variables and anime scores are detailed in Table 1. Additionally, the GB Regressor model achieved an R^2 value of 0.707 and MSE of 0.259, while the RF Regressor model performed slightly better with an R^2 value of 0.719 and MSE of 0.248. Hyperparameter tuning for the RF Regressor model via GridSearchCV identified optimal parameters, leading to similar model performance with an R^2 value of 0.72 and MSE of 0.247. The top features for the tuned RF Regressor model, as determined by descending order of feature importance, include *favorites per entry*, *favorites*, *scored by per member*, *favorites by per member*, *members*, *episodes*, *scored by per entry*, *scored by*, *rating: R+ - Mild Nudity*, and *type: TV*. Feature importance plots are shown in Section 5.8: *Suppl. Feature Importances.*

Table 1: Key Associations between Independent Variables and Anime Score

Variable	β	p-value
Favorites	-3.97×10^{-5}	8.99×10^{-79}
Members	1.05×10^{-6}	3.16×10^{-5}
Source: Manga	0.42	4.24×10^{-74}
Producer: Aniplex	0.27	0.0003
Rating: PG-13	0.35	5.32×10^{-6}
Rating: R-17+	0.33	4.18×10^{-5}
Favorites per Member	90.78	0.0
Scored by per Member	0.58	4.22×10^{-20}
Favorites per Entry	-3.14×10^{-9}	9.25×10^{-79}

PCA and K-Means aimed to uncover latent structures and patterns within users, derived from their activity on MAL. Preliminary correlation analysis revealed strong correlations ($\rho > 0.8$) among user features *days watched, episodes watched,* and *completed* (see Section 5.9: *Suppl. User Dataset Correlations*). Kaiser criterion determined that the top 2 principal components should be considered for further analysis. These components explain about 61% of the variance in the data. Features *days watched* and *on hold* were identified as having the highest magnitude of correlations with principal components 1 and 2, respectively. Silhouette analysis determined the optimal number of clusters for K-Means to be 2. Visualizations of data points after dimensionality reduction did not reveal notable clusters explained by pre-determined labels (score quantiles and user gender). See Section 5.10: *Suppl. PCA & K-Means* for all PCA and K-Means related visualizations.

4.3 Analysis

In the regression analyses, particularly with Ordinary Least Squares (OLS), notable relationships between independent variables and average anime scores were observed. Positive coefficients for categorical variables suggest that selecting a specific level is associated with higher average scores compared to the reference category ('Other/Unknown'), holding other variables constant. Similarly, positive coefficients for numerical features indicate that increasing the feature is linked to higher average scores, while negative coefficients suggest the opposite trend. In particular, *favorites* and *favorites per entry* displayed very small negative coefficients, suggesting a minor dampening effect on anime scores. This implies that while user favoritism may indicate popularity, it doesn't guarantee higher scores. Conversely, *members* showed a very small positive coefficient, indicating a slight positive influence on scores in relation to the amount of users belonging to the anime's MAL fanbase. Certain categorical features like *Source: Manga, Producer: Aniplex,* and ratings such as *PG-13* and *R-17+* also had positive coefficients, suggesting a positive relationship with anime scores. The feature *scored by per member* also showed a positive association. However, features like *favorites, members,* and *scored by* were highly correlated, which might complicate interpreting coefficients for *favorites* and *members.* The notably high positive coefficient for *favorites per member* suggests a strong positive relationship with anime scores, indicating that a higher ratio correlates strongly with higher scores. It's important to remember that while these coefficients offer insights into relationships, they do not imply causation.

The GB and RF Regressors are ensemble methods commonly used for regression tasks. GB Regressor iteratively combines weak learners to improve predictive accuracy, while RF constructs multiple decision trees on random subsets of the data and averages their predictions. In this analysis, both methods generally outperformed OLS, indicating potential non-linear relationships in anime data that OLS may not effectively capture. Comparing the two, RF offers the advantage of reduced overfitting compared to GB. However, it may struggle to capture nuanced data patterns. Despite this distinction, both methods exhibited similar predictive accuracy, suggesting their suitability for modeling anime score data. Model tuning involves adjusting hyperparameters to optimize performance. While tuning the RF model aimed to enhance its predictive power, the default configuration already demonstrated satisfactory performance. This highlights the robustness of the default RF model in capturing essential patterns in anime score data without extensive tuning. Feature importances across all ensemble models highlight the significance of user engagement metrics in predicting average anime scores. Notably, metrics such as *favorites*, *members*, and engineered user engagement ratios consistently emerged as influential features. In addition, the RF model identified episode count, R-rated anime, and TV-type anime among its top 10 features. This suggests that factors related to content characteristics, such as episode count and rating, are key contributors to the model's predictive accuracy.

When reducing the dimensionality of user watch behavior using PCA and visualizing the data points, a correlation was observed between specific aspects of user activity and the principal components. Specifically, the number of days spent watching anime on MAL showed a correlation with the first principal component, while the count of anime entries on hold correlated with the second principal component. However, when categorizing entries by pre-defined criteria like user gender and score quantiles, distinct separations were not evident. This suggests that user activity is not solely influenced by gender or scoring behavior. Further exploration of user demographics or additional user-anime interaction features may reveal more natural groupings.

5. Supplementary Materials

5.1 Exploratory Data Analysis

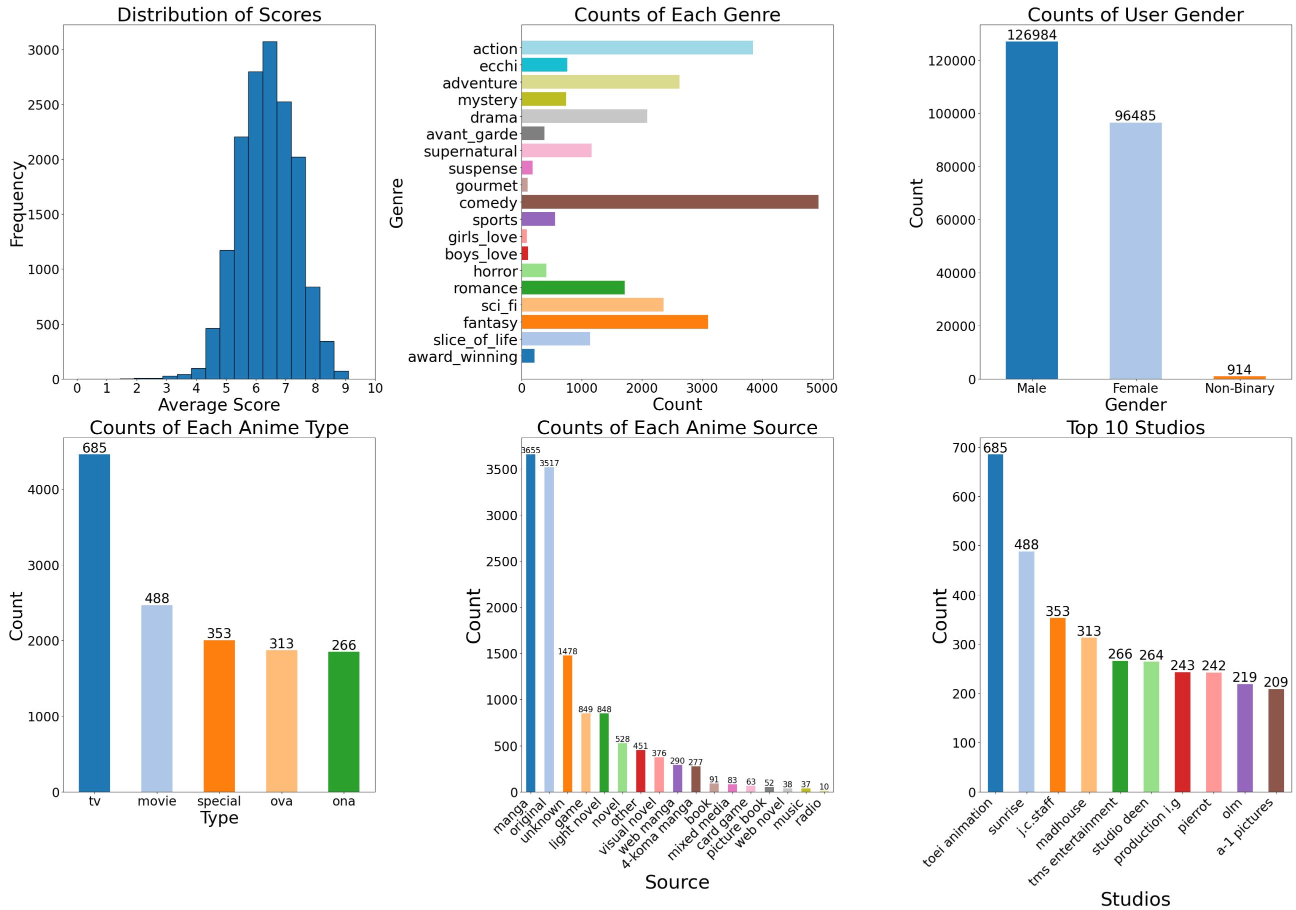

Figure 1: Exploratory analysis of anime-related features and user demographics. Subplots include the distribution of average scores, genre composition, user gender distribution, counts of anime types and sources, and the top 10 studios based on produced anime count.

5.2 Score Distributions

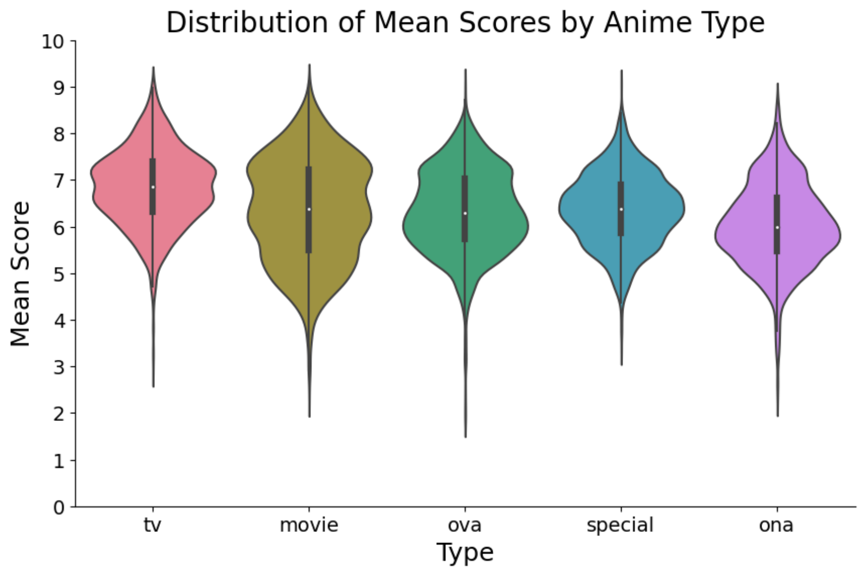

Figure 2: Violin plots representing the distribution of average user scores across various anime types, providing insight into their shape, central tendency, and variability.

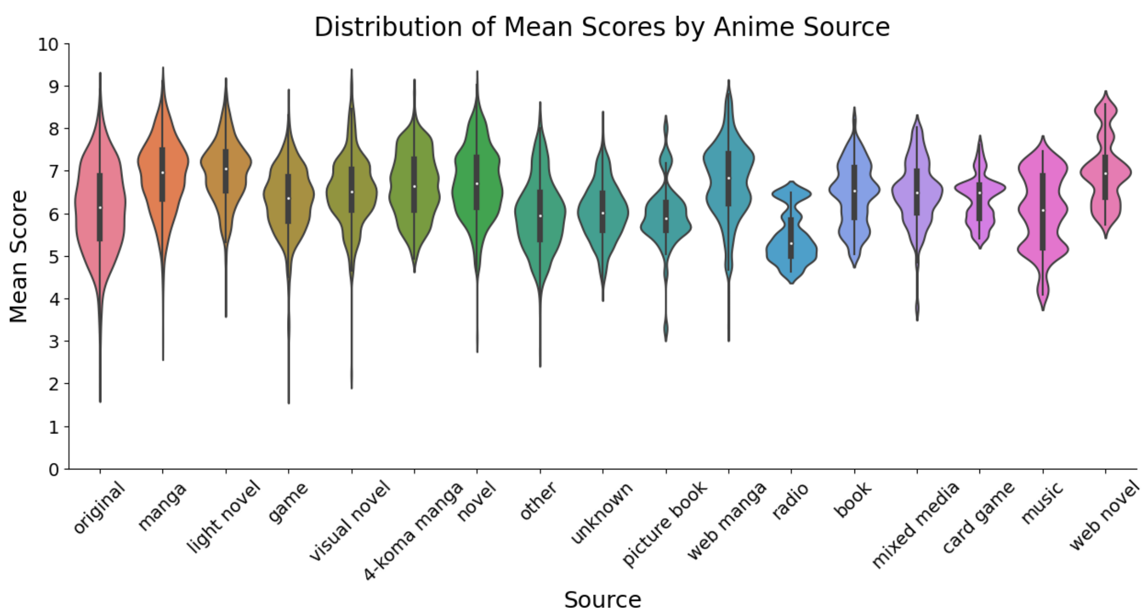

Figure 3: Violin plots representing the distribution of average user scores across various anime sources, providing insight into their shape, central tendency, and variability.

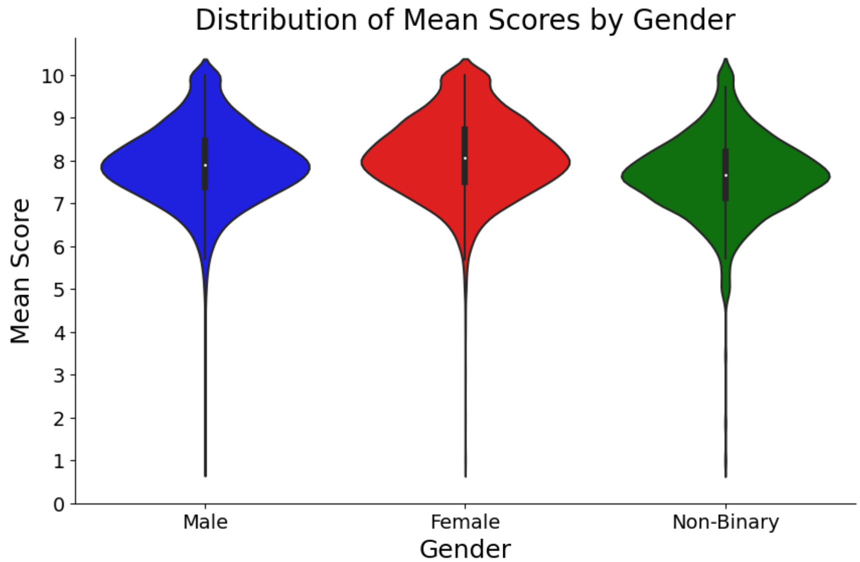

Figure 4: Violin plots representing the distribution of average user scores across various genders, providing insight into their shape, central tendency, and variability.

5.3 Hypothesis Testing

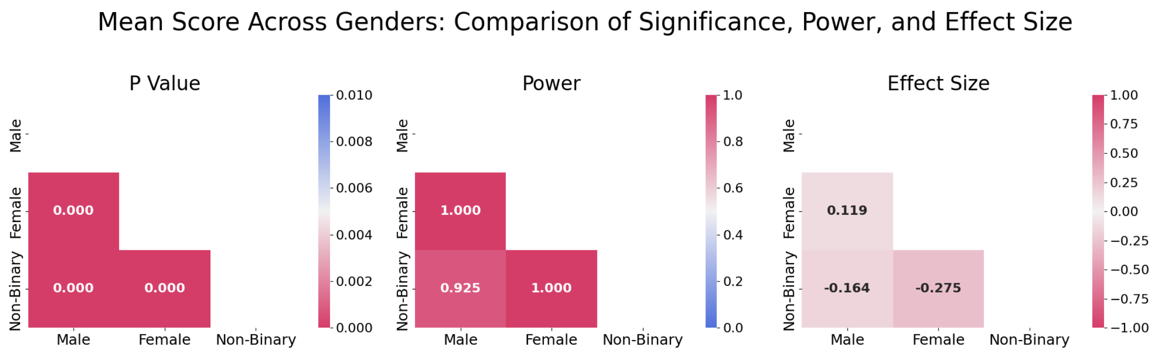

Figure 5: Heatmaps depicting the significance, power, and effect size of mean user scores across genders. For p values, lower values (pink) relate to values getting increasing lower than the adjusted alpha level. For power values, higher values (pink) relate to higher power. For effect size, larger magnitudes of effect are shown in pink.

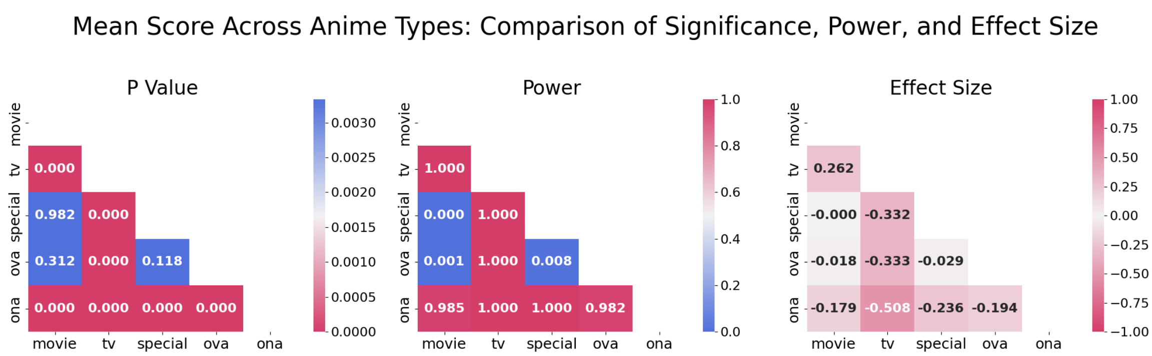

Figure 6: Heatmaps depicting the significance, power, and effect size of mean user scores across anime types. For p values, lower values (pink) relate to values getting increasing lower than the adjusted alpha level. For power values, higher values (pink) relate to higher power. For effect size, larger magnitudes of effect are shown in pink.

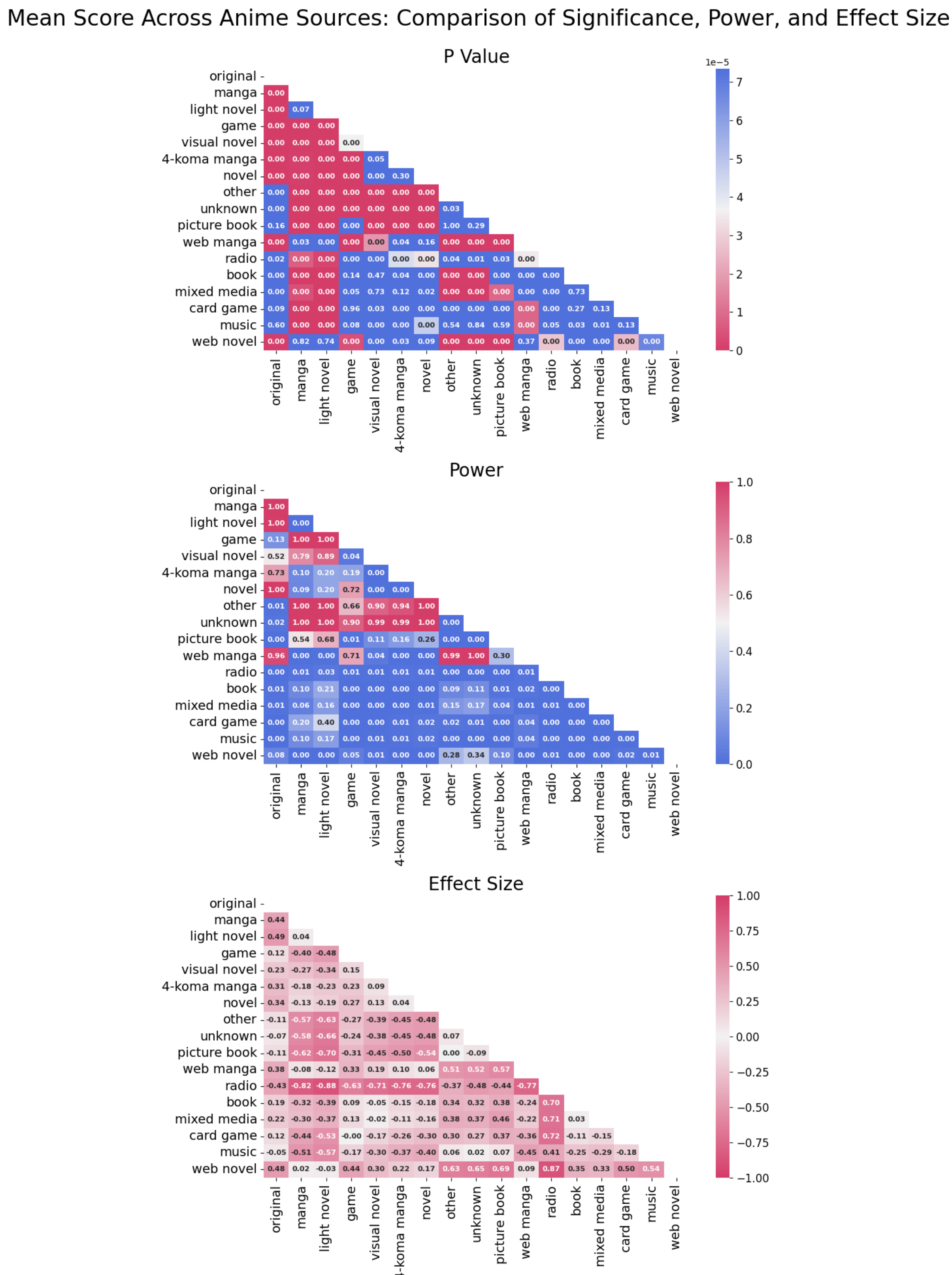

Figure 7: Heatmaps depicting the significance, power, and effect size of mean user scores across anime sources. For p values, lower values (pink) relate to values getting increasing lower than the adjusted alpha level. For power values, higher values (pink) relate to higher power. For effect size, larger magnitudes of effect are shown in pink.

5.4 Correlations

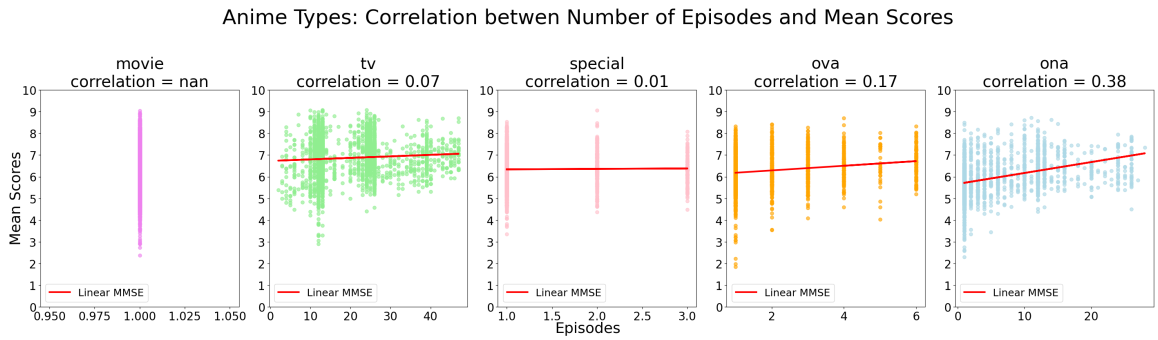

Figure 8: Scatter plots depicting the correlation between episode counts and mean user scores, within subgroups of anime types. Linear MMSE line was generated to visually depict the linear relationship, if any was found.

5.5 Regression

OLS Regression Results

Dep. Variable:	score	R-squared:	0.205
Model:	OLS	Adj. R-squared:	0.204
Method:	Least Squares	F-statistic:	171.1
Date:	Sun, 14 Apr 2024	Prob (F-statistic):	0.00
Time:	07:53:35	Log-Likelihood:	-15731.
No. Observations:	12643	AIC:	3.150e+04
Df Residuals:	12623	BIC:	3.165e+04
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	5.9957	0.017	361.118	0.000	5.963	6.028
genre_award_winning	0.7168	0.059	12.217	0.000	0.602	0.832
genre_slice_of_life	0.2235	0.027	8.169	0.000	0.170	0.277
genre_fantasy	0.2156	0.019	11.307	0.000	0.178	0.253
genre_sci_fi	0.1031	0.021	4.969	0.000	0.062	0.144
genre_romance	0.3859	0.023	16.845	0.000	0.341	0.431
genre_horror	-0.3548	0.044	-8.066	0.000	-0.441	-0.269
genre_boys_love	0.0651	0.083	0.782	0.434	-0.098	0.228
genre_girls_love	0.2577	0.090	2.856	0.004	0.081	0.434
genre_sports	0.5467	0.038	14.530	0.000	0.473	0.620
genre_comedy	0.2237	0.017	13.500	0.000	0.191	0.256
genre_gourmet	0.3244	0.084	3.867	0.000	0.160	0.489
genre_suspense	0.3956	0.065	6.131	0.000	0.269	0.522
genre_supernatural	0.3427	0.027	12.563	0.000	0.289	0.396
genre_avant_garde	-1.0254	0.045	-22.573	0.000	-1.114	-0.936
genre_drama	0.4283	0.021	20.102	0.000	0.387	0.470
genre_mystery	0.5043	0.033	15.236	0.000	0.439	0.569
genre_adventure	0.1544	0.020	7.804	0.000	0.116	0.193
genre_ecchi	-0.1267	0.033	-3.890	0.000	-0.191	-0.063
genre_action	0.2938	0.018	16.298	0.000	0.258	0.329

Omnibus:	72.774	Durbin-Watson:	1.647
Prob(Omnibus):	0.000	Jarque-Bera (JB):	84.258
Skew:	-0.130	Prob(JB):	5.05e-19
Kurtosis:	3.303	Cond. No.	14.9

Figure 9: Summary report for OLS Regression model addressing how the genre of an anime influence its average user score.

Genre	Train Length	Test Length	R2 Score	MSE
genre_award_winning	170	43	-0.024512015032588375	0.9268000542902687
genre_slice_of_life	912	228	0.09000349348791403	0.724102278718819
genre_fantasy	2482	621	0.11894728103964047	0.5807626042261316
genre_sci_fi	1888	473	0.13067189205548813	0.6132955157361287
genre_romance	1374	344	0.07536942847678763	0.5876295937155595
genre_horror	328	82	0.4614993787963517	0.5538464153191018
genre_boys_love	83	21	-0.28829755747115215	0.7734383486222514
genre_girls_love	70	18	-0.19671596621415177	0.6027832466846103
genre_sports	444	112	0.004649407557719698	0.8184178519213405
genre_comedy	3953	989	0.12572169864431382	0.6392053645172637
genre_gourmet	81	21	-0.20746634277148512	0.8307784617143331
genre_suspense	145	37	0.07960288288302209	0.9390341838840899
genre_supernatural	929	233	0.12527584124366486	0.6601106726066186
genre_avant_garde	303	76	0.3079616890182142	0.5871891392578771
genre_drama	1672	418	0.07005703125162088	0.6796583893665968
genre_mystery	590	148	0.15136376146342723	0.6834249568113993
genre_adventure	2104	526	0.13478171325340127	0.5966301725485893
genre_ecchi	607	152	0.06101520280149042	0.5647219047129635
genre_action	3081	771	0.10141681207735209	0.6858882179541234

Figure 10: Results of OLS Regression model attempting to predict the average user score of an anime based on the average user scores of other animes in the same genre. Shows the length of the train and test sets used, along with computed R^2 and MSE.

Type	Genre	Train Length	Test Length	R2 Score	MSE
tv	genre_award_winning	43	11	-0.47780291958110865	0.4980415677439179
movie	genre_award_winning	115	29	-0.22077526350947907	1.1945717071458397
ova	genre_award_winning	8	3	-22.539535478988896	0.43521985596708435
tv	genre_slice_of_life	375	94	0.06879184185823528	0.5417189535587391
movie	genre_slice_of_life	109	28	-0.27725951932974424	1.192987109237804
ova	genre_slice_of_life	86	22	-0.23896644666358524	0.7516717277378884
special	genre_slice_of_life	164	42	0.010935603892796353	0.6502811889718689
ona	genre_slice_of_life	176	44	0.027152235724424933	0.4669253195202088
tv	genre_fantasy	886	222	-0.03348820224662541	0.5560863992123991
movie	genre_fantasy	522	131	0.19392658860386502	0.6360707956940163
ova	genre_fantasy	360	90	0.004908831482522857	0.8230187874348959
special	genre_fantasy	292	73	-0.2797554939536644	0.5265298821266173
ona	genre_fantasy	420	106	0.03768696326637255	0.4383748351134855
tv	genre_sci_fi	722	181	0.06211830051332656	0.6484719619245161
movie	genre_sci_fi	382	96	0.08678380320990298	0.9443822263081868
ova	genre_sci_fi	389	98	-0.1339947699343531	0.7083851021188124
special	genre_sci_fi	232	59	-0.09087534641973272	0.622889697453717
ona	genre_sci_fi	161	41	0.08748028648408279	0.7514246449766399
tv	genre_romance	628	157	0.0564592081732358	0.3926285262558862
movie	genre_romance	169	43	0.07235915971168727	0.7962777663876991
ova	genre_romance	283	71	-0.0015253308271088173	0.6414924115140264
special	genre_romance	159	40	-0.004022765592724076	0.36835681351200134
ona	genre_romance	134	34	-0.5009234510881166	0.5420359126061057
tv	genre_horror	110	28	-0.028458405325135105	0.6498818168776005
movie	genre_horror	64	17	0.4279025754244804	0.5540425049541233
ova	genre_horror	80	21	-0.03534655245332585	0.931358316813585
special	genre_horror	26	7	0.36430507776610666	0.898470445700169
ona	genre_horror	45	12	0.5227962755513795	0.8511551441514733
tv	genre_boys_love	24	6	-14.376383590152617	0.9444516498484855
movie	genre_boys_love	12	3	-0.033097963291204735	0.30926361474435143
ova	genre_boys_love	25	7	-3.2011358953697586	1.3945199048610013
ona	genre_boys_love	14	4	-9.935330046288893	1.8516247600878681
tv	genre_girls_love	34	9	-1.4730021138160705	1.4088479326178096
ova	genre_girls_love	13	4	-5.569868741525355	6.195714716695488
special	genre_girls_love	10	3	-31.01351740005731	0.824525925925922
tv	genre_sports	205	52	0.16594714121987808	0.382501175273629
movie	genre_sports	79	20	-0.24936634071655095	1.4314052633906558
ova	genre_sports	52	14	-0.1528959097476461	0.9826249481218958
special	genre_sports	75	19	-0.1271155330682816	0.571944630093498
ona	genre_sports	32	8	-0.018566549187537085	0.3666314378698207
tv	genre_comedy	1534	384	0.11766567075808321	0.5612949921676827
movie	genre_comedy	516	130	0.3061095997160982	0.6453104558930669
ova	genre_comedy	584	146	0.073933170877582	0.6820522382816921
special	genre_comedy	796	199	0.07957167846687907	0.4546405780831358

Type	Genre	Train Length	Test Length	R2 Score	MSE
ona	genre_comedy	522	131	0.21210249536999637	0.5712220638010527
tv	genre_gourmet	48	12	0.038774471070451066	0.7159120962162584
special	genre_gourmet	11	3	-2.5861975397973676	0.6608166666666621
ona	genre_gourmet	14	4	-0.5302088833765308	0.3604407024793418
tv	genre_suspense	60	15	-0.09775962880263256	0.7467468404010595
movie	genre_suspense	28	8	-5.761316559888018	4.554665819554442
ova	genre_suspense	15	4	-5.2142768838440965	0.9341223333333408
special	genre_suspense	16	5	-0.9526224808090764	2.2119619882202137
ona	genre_suspense	24	7	0.770234918381175	0.2665603182609254
tv	genre_supernatural	374	94	0.18809519177892087	0.5264315429975063
movie	genre_supernatural	138	35	0.28500655384103024	0.6082277064379156
ova	genre_supernatural	144	37	-0.12757758161827604	0.8368877516168707
special	genre_supernatural	141	36	0.3358747563289258	0.4317766714130489
ona	genre_supernatural	130	33	0.2612176254526182	0.719356807725284
tv	genre_avant_garde	14	4	0.9547376566344093	0.04590167396562966
movie	genre_avant_garde	208	52	-0.24632714270555334	1.1450001928998303
ova	genre_avant_garde	14	4	-2.72254325933448	5.3214453869047516
special	genre_avant_garde	24	6	-1.5406729739064473	0.5006325520833376
ona	genre_avant_garde	42	11	-1.9188183576246352	0.7910769668067258
tv	genre_drama	671	168	0.09807764357597903	0.46175563640244527
movie	genre_drama	426	107	0.10665382906762211	0.7727612919724052
ova	genre_drama	274	69	-0.18873272649025252	0.8216037222480569
special	genre_drama	178	45	-0.04946373067512866	0.49022311390597295
ona	genre_drama	121	31	0.08127770705582404	0.7450445833092455
tv	genre_mystery	256	64	-0.036394037560639525	1.0297909727750445
movie	genre_mystery	101	26	-0.9756318118640865	1.2685075948945792
ova	genre_mystery	85	22	-0.11490087216191025	0.7387803095841349
special	genre_mystery	84	21	0.01097765381463156	0.32823745390993836
ona	genre_mystery	63	16	0.07788204482769023	0.8195838415920422
tv	genre_adventure	828	208	0.07232463341670936	0.5466592444824415
movie	genre_adventure	519	130	0.17213649157926425	0.5938507377647686
ova	genre_adventure	298	75	-0.09849813607436464	0.683732658406161
special	genre_adventure	256	65	-0.008943763288782769	0.39045106337398594
ona	genre_adventure	200	51	0.09895852131480942	0.4532650186143555
tv	genre_ecchi	248	62	-0.28972375704315434	0.2825832059446833
movie	genre_ecchi	12	4	-13.861042183622688	2.1335812499999793
ova	genre_ecchi	188	48	0.022471502868033677	0.7553647385542512
special	genre_ecchi	129	33	-0.12086228290416079	0.35086900636420015
ona	genre_ecchi	28	7	0.5779025053014093	0.2071516676227312
tv	genre_action	1192	298	0.000954825301193063	0.5538237469698178
movie	genre_action	560	140	0.15885933101859606	0.8708572203874974
ova	genre_action	528	132	-0.07599064059134131	0.7795168085150816
special	genre_action	385	97	-0.0778823382611975	0.6050339722466422
ona	genre_action	416	104	0.1459634970648882	0.6137927499036843

Figure 11: Results of OLS Regression model attempting to predict the average user score of an anime based on the average user scores of other animes in the same genre and anime type combination. Shows the length of the train and test sets used, along with computed R^2 and MSE.

5.6 Anime Dataset Correlations

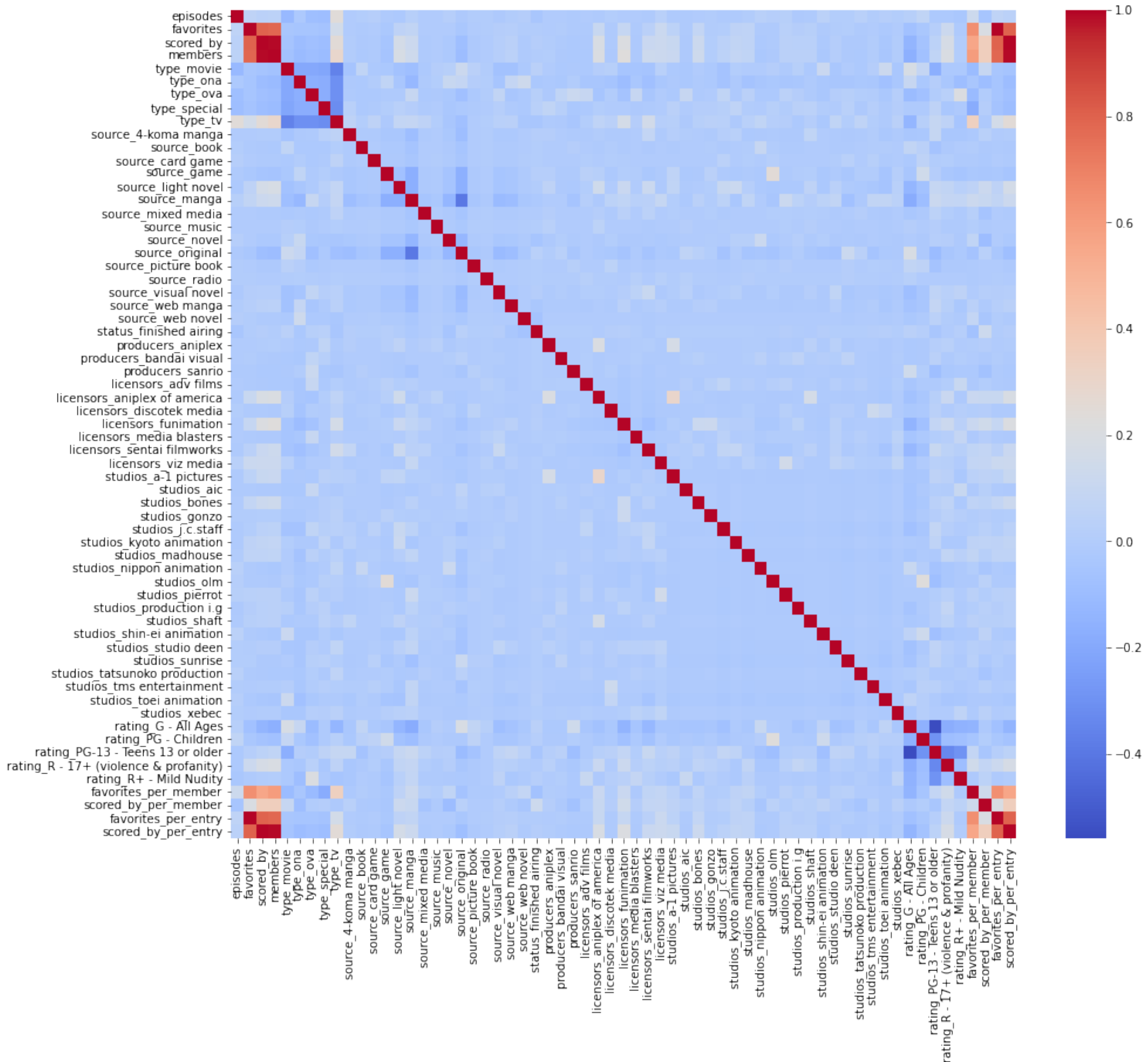

Figure 12: Correlation matrix for numerical and one hot encoded categorical features in the anime dataset

5.7 OLS Regressor

Feature Name	Coefficient	Standard Error	P Value	Statistical Significance
episodes	0.00045982960309688887	0.0002068874385506	0.026264430709657217	Fail to reject
favorites	-3.970622643828114e-05	2.0944945493174e-06	8.99485906296757e-79	Reject
scored_by	-5.946015831273613e-07	4.0895145217662187e-07	0.1459869017965672	Fail to reject
members	1.0542346084092082e-06	2.53205515792742e-07	3.1594775292822986e-05	Reject
type_movie	-0.2592961882989817	0.6587751036984112	0.6938821429877091	Fail to reject
type_ona	-0.5442383868661573	0.6588137994358507	0.4087735972254096	Fail to reject
type_ova	-0.3856208284486886	0.6589022958054948	0.5583945421432088	Fail to reject
type_special	-0.30219271593370767	0.658851221863391	0.646483882848546	Fail to reject
type_tv	-0.30559201977551087	0.658745759511591	0.642729588303206	Fail to reject
source_4-koma manga	0.36698562770385723	0.0485738549929484	4.546234683212732e-14	Reject
source_book	0.37416918733731236	0.08109445754277361	3.998943407325266e-06	Reject
source_card game	0.07525787049294066	0.09358107555251303	0.42130068500425577	Fail to reject
source_game	0.10561881321183414	0.031905344976536326	0.0009349752421536968	Reject
source_light novel	0.34884339020021105	0.03359992291412052	3.999456440587419e-25	Reject
source_manga	0.42429135937803814	0.022945441250684036	4.2363319377634712e-75	Reject
source_mixed media	0.02341689951958484	0.087693896283999	0.789451554783287	Fail to reject
source_music	-0.1810874014387696	0.12394407332530614	0.14403538298702864	Fail to reject
source_novel	0.366287478536894	0.03715840479559256	8.074844628192928e-23	Reject
source_original	-0.043047950584834946	0.021886734780213992	0.0492278326367198	Fail to reject
source_picture book	-0.01986155773797249	0.10222283600636141	0.8459475107486022	Fail to reject
source_radio	-0.44165629387898653	0.20941784286740653	0.03497169647434499	Fail to reject
source_visual novel	0.22268244972838458	0.04378520785619153	3.726988268291468e-07	Reject
source_web manga	0.3841461568114869	0.04885687274413189	4.1458084750562715e-15	Reject
source_web novel	0.38155191599646293	0.12434178987081564	0.0021565610547804644	Reject
status_finished airing	-0.17162115898548086	0.09440481155640924	0.06910487670757585	Fail to reject
producers_aniplex	0.26714875451319064	0.07425750937497028	0.00032271042925936897	Reject
producers_bandai visual	0.1758409455703362	0.07125837605910638	0.013616741006012508	Fail to reject
producers_sanrio	0.138107553620482	0.07551348995016766	0.06744235330914311	Fail to reject
licensors_adv films	-0.0067651241852755145	0.06126349007183361	0.9120731946365062	Fail to reject
licensors_aniplex of america	0.3549775365896659	0.05530864019130293	1.4418860280214483e-10	Reject
licensors_discotek media	0.24596124362148392	0.04616273398891603	1.013728386149629e-07	Reject

Feature Name	Coefficient	Standard Error	P Value	Statistical Significance
licensors_funimation	0.1938574686591922	0.027398496589738284	1.58870999260362e-12	Reject
licensors_media blasters	-0.17570080379321973	0.06574571065543325	0.007542560243982669	Fail to reject
licensors_sentai filmworks	0.2503422588540404	0.029160007816495883	1.0417458468721508e-17	Reject
licensors_viz media	0.21537454345910828	0.05978222721308261	0.0003165081221346493	Reject
studios_a-1 pictures	0.2254774840093084	0.0560847619315718	5.854896381241369e-05	Reject
studios_aic	0.3289955580062849	0.06574248690434156	5.701314735495869e-07	Reject
studios_bones	0.4481349836534471	0.0639403554223617	2.5609573967531313e-12	Reject
studios_gonzo	0.19577630292571044	0.06964851953380635	0.004949590314508863	Reject
studios_j.c.staff	0.1679337881550806	0.040865815512612945	3.9984577771305844e-05	Reject
studios_kyoto animation	0.5349945684998494	0.070190375687496	2.722307819412672e-14	Reject
studios_madhouse	0.2952729191836339	0.043964205089173994	1.9659636568366568e-11	Reject
studios_nippon animation	0.5159384228428359	0.058143040418203555	8.287626455071137e-19	Reject
studios_olm	0.44430440875765526	0.05358971775840464	1.2682232976920438e-16	Reject
studios_pierrot	0.1669640473861982	0.049113296157243615	0.0006775445286195525	Reject
studios_production i.g	0.5363177103714089	0.048130084662337015	1.1389650000578816e-28	Reject
studios_shaft	0.2912225923736405	0.06875106695308213	2.296674752319847e-05	Reject
studios_shin-ei animation	0.45278160399911305	0.057662813535050886	4.504293785770169e-15	Reject
studios_studio deen	0.2097384006334274	0.04736538224652182	9.607325410519272e-06	Reject
studios_sunrise	0.4728024492266465	0.03587366465623744	2.4338578208303044e-39	Reject
studios_tatsunoko production	0.3313804588756901	0.0619737684309398	9.133043178808835e-08	Reject
studios_tms entertainment	0.43754356035798375	0.04528236204144724	5.419004187555197e-22	Reject
studios_toei animation	0.31799109387861196	0.030412116169579607	1.856277973826734e-25	Reject
studios_xebec	0.3342372348876132	0.06916351251446373	1.3677839198337522e-06	Reject
rating_G - All Ages	-0.028972886674683673	0.07809099231904977	0.7106345866732264	Fail to reject
rating_PG - Children	0.1744312725436899	0.08076696193962646	0.03082056065784809	Fail to reject
rating_PG-13 - Teens 13 or older	0.3531394086510826	0.07753905087656728	5.316798568178027e-06	Reject
ing_R - 17+ (violence & profanity)	0.32902075969726347	0.08027110669854715	4.184375563767347e-05	Reject
rating_R+ - Mild Nudity	-0.02957949028102616	0.08086743442684284	0.7145387864276258	Fail to reject
favorites_per_member	90.7839357616658	2.165010568159503	0.0	Reject
scored_by_per_member	0.5841996814961047	0.06348790476118886	4.222552935009221e-20	Reject
favorites_per_entry	-3.1402534064770864e-09	1.6566087653966177e-10	9.247962138864471e-79	Reject
scored_by_per_entry	-4.7086304462595356e-11	3.234555422893752e-11	0.14549926551714049	Fail to reject

Figure 13: Results of OLS Regression model attempting to predict the average user score of an anime based on numerical and categorical anime characteristics (sans genre). Shows variable coefficient, standard error, p value, and statistical significance

5.8 Feature Importances

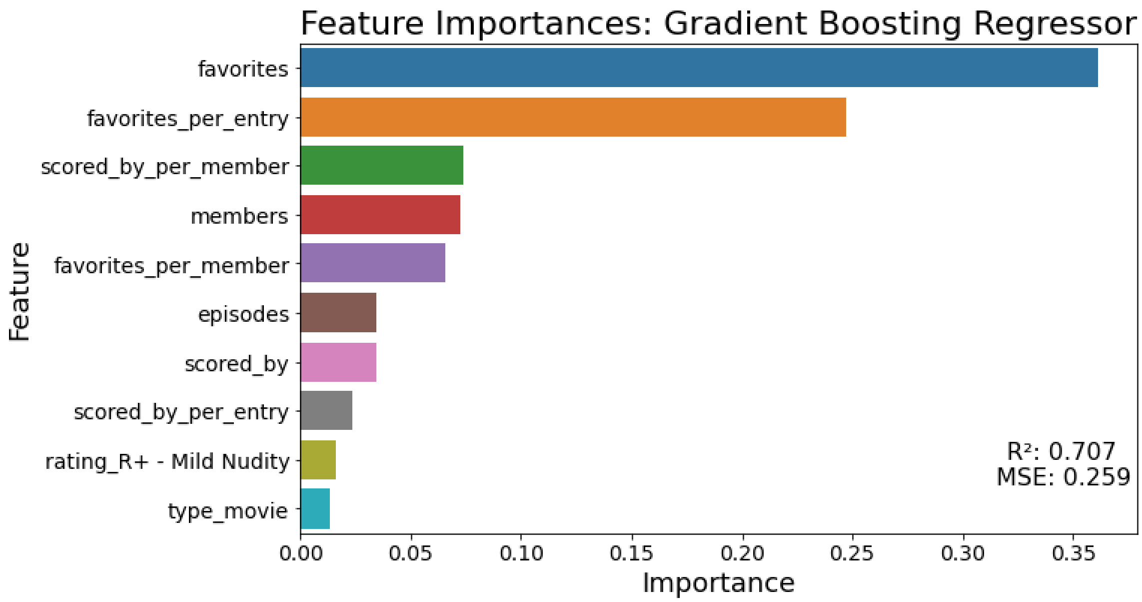

Figure 14: Top 10 features for Gradient Boosted Regressor model

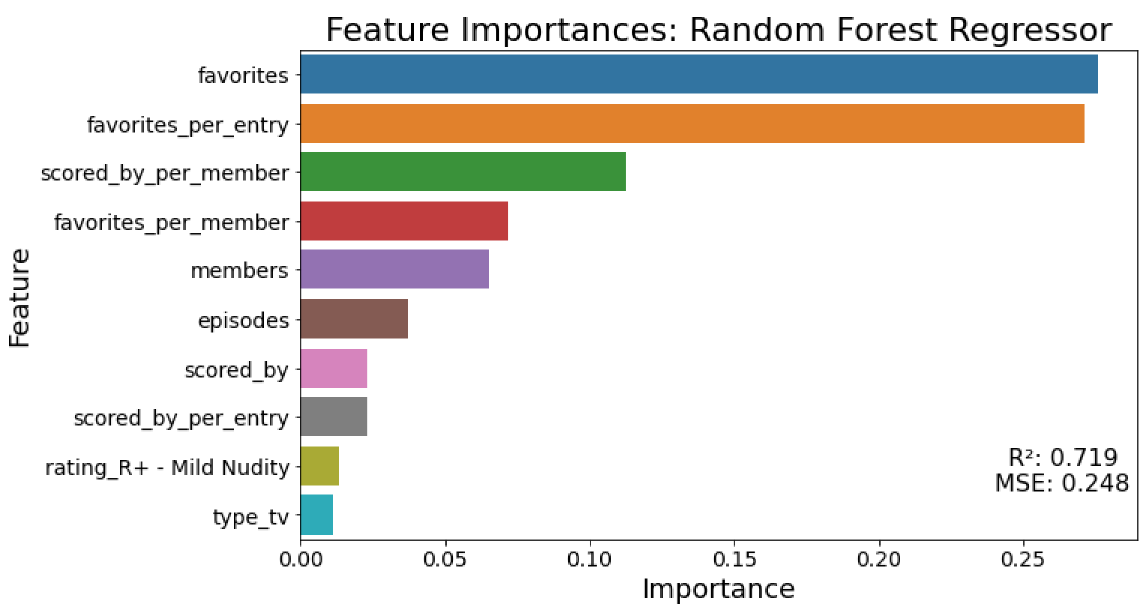

Figure 15: Top 10 features for Random Forest Regressor model

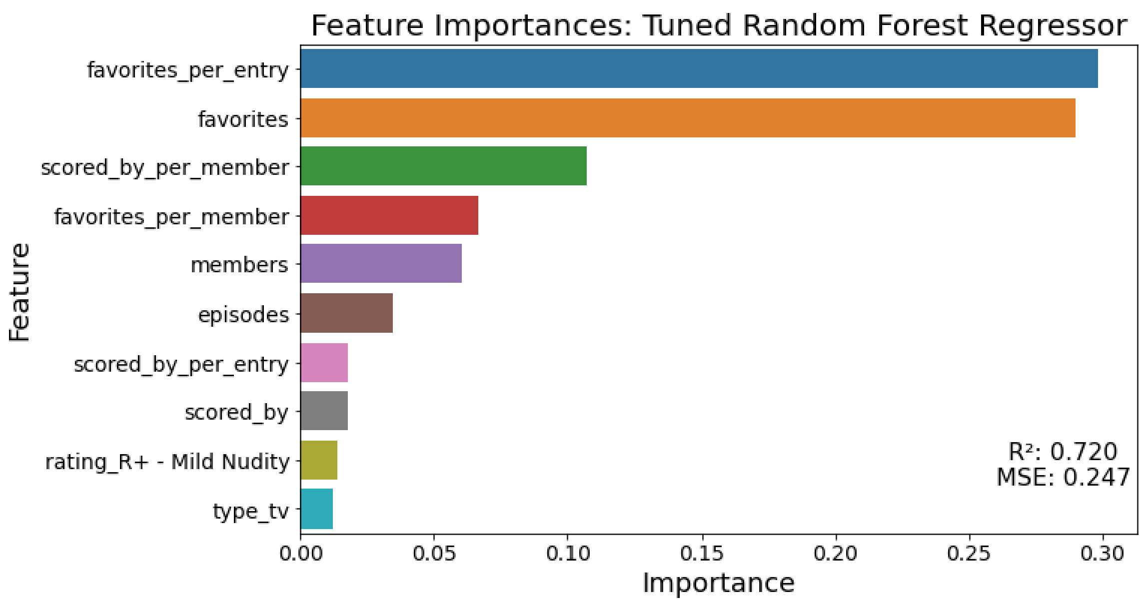

Figure 16: Top 10 features for tuned Random Forest Regressor model

5.9 User Dataset Correlations

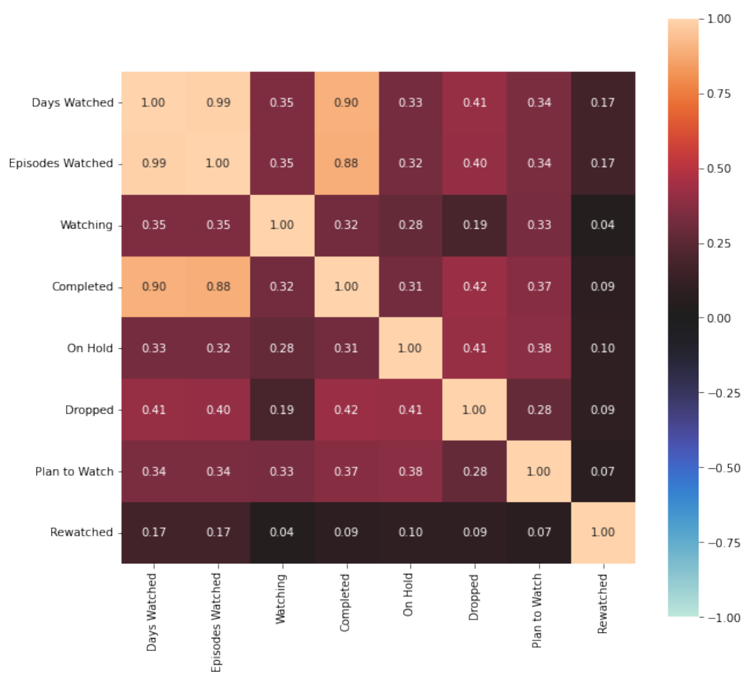

Figure 17: Correlation matrix for user activity features in the user dataset

5.10 PCA & K-Means

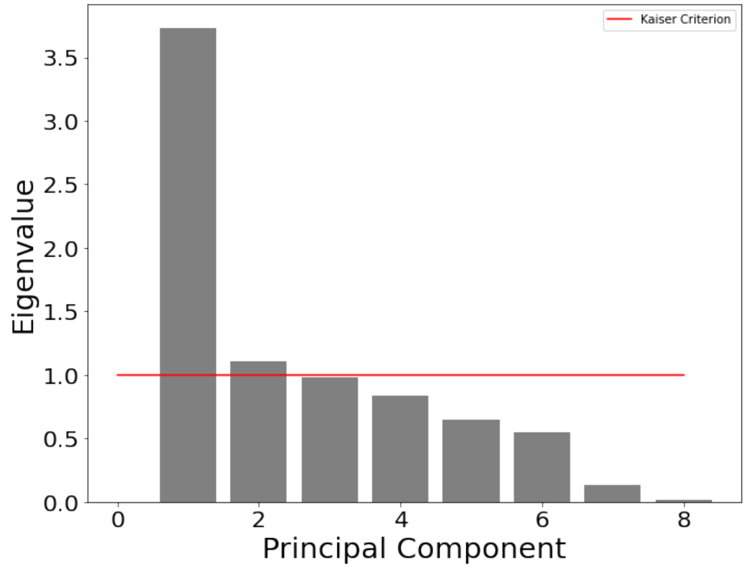

Figure 18: Kaiser criterion plot showing that 2 principal components should be considered for dimensionality reduction

	Days Watched	Episodes Watched	Watching	Completed	On Hold	Dropped	Plan to Watch	Rewatched
0	0.477891	0.475112	0.260956	0.461270	0.281640	0.305775	0.284575	0.10592
1	-0.323771	-0.327472	0.321964	-0.284764	0.552152	0.251957	0.475106	-0.09541

Figure 19: Correlation between each user activity feature and each principal component

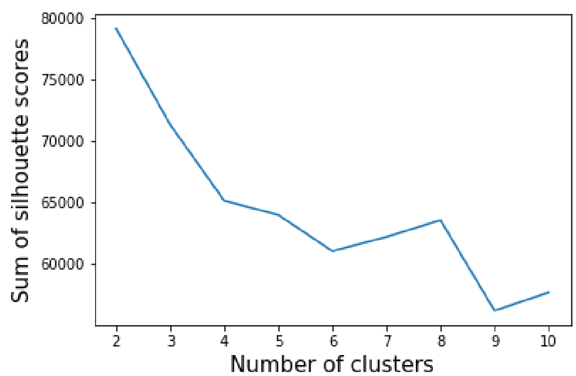

Figure 20: Silhoutte analysis plot showing that 2 clusters is optimal for K-Means clustering. Also suggests that there might be additional structure in the data that could be captured by increasing the number of clusters to 8.

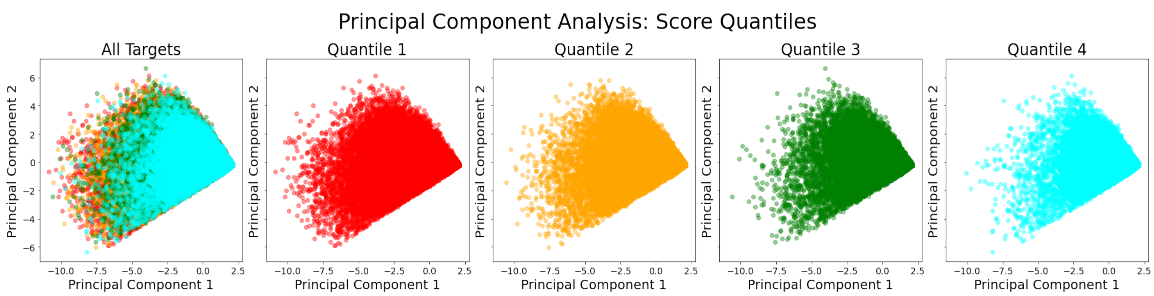

Figure 21: Visualization of data points in the principal component subspace, color coded by score quantiles. 1 is lowest quantile (lower average user scores)and 4 is highest quantile (higher average user scores).

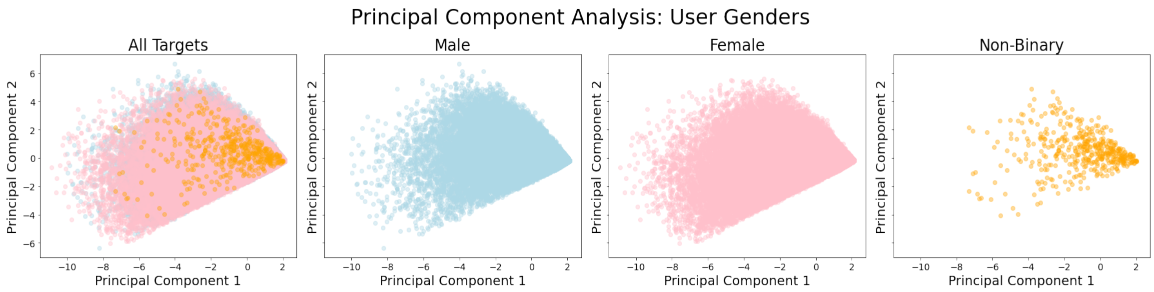

Figure 22: Visualization of data points in the principal component subspace, color coded by user genders

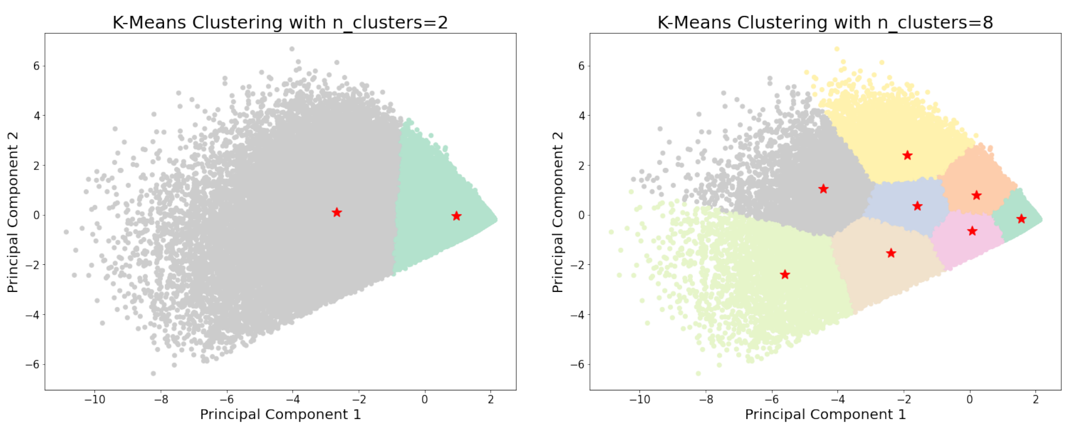

Figure 23: K-Means clustering results for 2 clusters and for 8 clusters. A small peak in the silhouette analysis implied that increasing cluster count to 8 could capture more underlying data structure.

5.11 Dataset Features

Full dataset containing all listed Animes in MyAnimeList.

Variable	Description
Anime ID	anime id on MyAnimeList
Name	anime original name
English Name	English version name
Other Name	Japanese version name
Score	average score
Genres	related genres
Synopsis	brief description
Type	type of animation (movie, anime, OVA. . .)
Episodes	number of episodes. Movies are considered having 1 episode
Aired	period when anime was aired
Premiered	season when the anime was released
Status	current status (airing, hiatus, finished. . .)
Producers	related production companies
Licensors	related streaming platforms and licensors
Studios	related animation studios
Source	source material of the anime (originated from manga, light novel, movie or tv)
Duration	duration of the movie or each episode
Rating	age restriction
Rank	rank position on MyAnimeList website (based on Score criteria)
Popularity	popularity position on MyAnimeList
Favorites	number of users that marked the anime as favorite
Scored By	number of users that rated the anime
Members	number of users that added the anime to the watch list
Image Url	banner url

Full dataset containing all users in MyAnimeList.

Variable	Description
Mal ID	user id
Username	nickname
Gender	user gender
Birthday	birthday
Location	user's location or country
Joined	the joined date on MyAnimeList Platform (ISO format)
Days Watched	total number of days the user spent on MyAnimeList
Mean Score	the average score the user gives to the watched animes
Watching	number of animes currently being watched by the user
Completed	number of animes finished by the user
On Hold	number of animes that the user stopped watching but kept it into its list
Dropped	number of animes that the user stopped watching and removed from its list
Plan to Watch	number of animes that the user has added into the list but did not started watching
Total Entries	total number of animes into the user's list
Rewatched	number of animes rewatched
Episodes Watched	number of episodes watched from all anime

Full dataset containing all ratings from all users on MyAnimeList.

Variable	Description
User ID	user id on MyAnimeList Platform
Username	nickname
Anime ID	anime id on MyAnimeList Platform
Anime Title	anime original name
Rating	score that the user rated the anime

References

Daniel J. Benjamin, Jesse Berger, Magnus Johannesson, Brian A. Nosek, Eric Wagenmakers, Richard Berk, and Valen ... Johnson. Redefine statistical significance. https://doi.org/10.31234/osf.io/mky9j, 2017.

DSFelix. MyAnimesList datasets - 2023. https://www.kaggle.com/datasets/dsfelix/animes-dataset-2023?select=users-details-transformed-2023.csv, January 2024.

Guillermo Macbeth, Eugenia Razumiejczyk, and Rubén Ledesma. Cliff's delta calculator: A non-parametric effect size program for two groups of observations. *Universitas Psychologica*, 10:545–555, 05 2011. doi: 10.11144/Javeriana.upsy10-2.cdcp.

Zhiyuan Wan, Xin Xia, David Lo, and Gail Murphy. How does machine learning change software development practices? *IEEE Transactions on Software Engineering*, PP:1–1, 08 2019. doi: 10.1109/TSE.2019.2937083.