Predicting Game of the Year (GOTY) 2024: A Bayesian Analysis of Video Game Metrics from Metacritic*

A Logistic Regression Approach Analysing Data from 2014–2023 to Forecast GOTY Success at The Game Awards

Ziqi Zhu

November 29, 2024

This study explores the factors influencing a video game's likelihood of winning the prestigious Game of the Year (GOTY) award by analyzing historical data from 2014 to 2023, sourced from Metacritic. A Bayesian logistic regression model was used to predict the 2024 GOTY winner, incorporating metrics such as critic scores, user reviews, genre, and engagement levels. The model demonstrated strong performance, with predicted rankings closely aligning with actual 2024 nominees, showing games with high critic scores and user engagement are more likely to win the award. These findings provide insights into the evolving dynamics of the gaming industry, detecting potential biases in the award process.

Table of contents

1

1 Introduction

Over the past decade, the video game industry has experienced significant growth, evolving from a niche market into a global entertainment powerhouse valued at approximately \$244.22 billion in 2024 (Intelligence 2024). Annual accolades such as the Game of the Year (GOTY) award at The Game Awards (TGA) have become critical markers of success for developers and publishers, influencing sales, visibility, and industry trends. Since its inception in 2014, TGA has been produced and hosted by Canadian games journalist Geoff Keighley. The nominees for

the 2024 GOTY were announced on November 18, with the final winner set to be revealed on December 12. This paper aims to predict the winner of the 2024 GOTY by analyzing historical data from 2014 to 2023 collected from Metacritic, through factors such as number of reviews from users and critics, positivity rates, release dates, and genres. Using a Bayesian logistic regression model, this study provides a data-driven framework to forecast which nominated game is most likely to claim this prestigious title.

The estimand for this study is the relationship between various factors associated with games on Metacritic—such as scores, positivity ratios from players and critics, and genre—and the probability of a video game winning the GOTY award at The Game Awards in 2024.

The result shows that higher critic scores, user engagement (measured by the number of user reviews on Metacritic), and specific genres are significant predictors of a video game's likelihood of winning the GOTY award. Critic scores, in particular, emerged as the strongest driver, with each additional point on Metacritic increasing the odds of winning. While user scores also positively influenced outcomes, their effect was less certain due to variability. Genres like Action RPG and Open-World Adventure demonstrated a natural advantage, reflecting industry and audience preferences.

The GOTY award holds significant cultural and financial importance within the video game industry. A win—or even a nomination—can dramatically increase a game's visibility, boost sales, and enhance the reputation of its developers and publishers. Understanding the factors that contribute to GOTY success provides valuable insights for stakeholders, including game developers, publishers, and marketers, enabling them to strategize and optimize future projects.

A case in point is Concord, a multiplayer first-person hero shooter game released on August 23, 2024, which was a commercial failure. Despite being backed by Sony Entertainment, Concord failed to exceed 700 simultaneous players on the Steam platform (Peters 2024), and discontinued in just 2 weeks on September 6th resulting in millions of dollars in losses for the publisher (Tassi 2024). On Metacritic, Concord received a mediocre Metascore of 62/100 based on 46 critic reviews and a dismal user score of 1.7/10 from 1,170 user reviews. This stark disparity highlights the importance of understanding player preferences and delivering games that resonate with the market. As the case of Concord demonstrates, failing to align with audience expectations can have dire consequences, reinforcing the importance of using data-driven approaches to predict and optimize success in a highly competitive industry.

2 Data

2.1 Overview

I use the statistical programming **language R** (R Core Team 2023), and packages **lubridate** (Grolemund and Wickham 2011), **dplyr** (Wickham et al. 2023), **tidyverse** (Wickham et

al. 2019) to process the data. Figures and tables are generated using package **patchwork** (Pedersen 2024), **knitr** (**kabble?**), packages **broom.mixed** (Bolker and Robinson 2024) and **modelsummary** (Arel-Bundock 2022) are used to view model summaries, **bayesplot** (Gabry et al. 2019) is used to diagnose for model. The data were scraped from Metacritic best games from 2014 to 2024 (Metacritic 2024), with a modified Python scrape scripts originally from Bruno Vieira Ribeiro's repo (Ribeiro 2021). The raw dataset originally have 5968 observations, and after cleaning process Section A, there are 4207 left for analyse, 3969 of them are games released between 2014-2023, 239 were released in 2024. I consider use a Bayesian logistic regression model to predict the winner of The Game Award for Game of the Year 2024.

2.2 Measurement

The raw dataset was scraped from Metacritic, a website that aggregates reviews for films, television shows, music albums, and video games. This study focuses on video games released between 2014 and 2024. For each game listed on Metacritic, the dataset includes various features such as release date, genre, the number of reviews from both critics and users, the positive ratio of reviews from critics and users, as well as the aggregated scores from critics and users. The data also includes downloadable content (DLC) alongside full games. For example, the action RPG "The Witcher 3: Wild Hunt," released in 2015, and its DLCs "The Witcher 3: Blood and Wine" and "The Witcher 3: Hearts of Stone" are all included as seperate entries. Since The Game Awards (TGA) revised their rules in 2024 to make DLCs eligible for GOTY nominations, these entries are retained in the analysis. In addition to the cleaned version of raw data, I add a column goty_status to mark the games that won GOTY in past years. Preview of first several rows of cleaned dataset ordered by Metascore can be found in Table 1.

Table 1: Preview of Best Games from 2014 to 2023 provided by Metacritic

Name	Release date	Score	Genre	User Score	User Positivity	Critic Positivity	Critics	Users	GOTY Status
The Legend of Zelda: Breath of the Wild	2017-03-03	97	Open-WorldAction	8.9	87	91	117	23624	1
Red Dead Redemption 2	2018-10-26	97	Open-WorldAction	8.9	87	90	109	27705	0
Grand Theft Auto V	2014-11-18	97	Open-WorldAction	8.5	84	100	66	11944	0

2.3 Outcome variables

The primary outcome variable is GOTY status, coded as 1 if a game wins the GOTY award and 0 otherwise. This variable serves as the response variable in the Bayesian logistic regression model. However, due to the highly imbalanced nature of the dataset—only 10 games out of 3,993 listed on Metacritic between 2014 and 2023 won GOTY—the model focuses on predicting the mean probability of winning GOTY for games released in 2024. This approach allows for comparison of the likelihood of winning among potential contenders.

2.4 Predictor variables

2.4.1 Score and User Score

Score also called Metascore is a score to measure quality of a game in scale of 100 the average score of the platform that has most critics reviewed, for example a game recieve 92 on PC end by 50 critics and 94 on Playstation 5 by 70 critics, the 94 will take acount. User score is similar to it but is a combination of user reviews for all platforms on Metacritic, in sclae of 10.
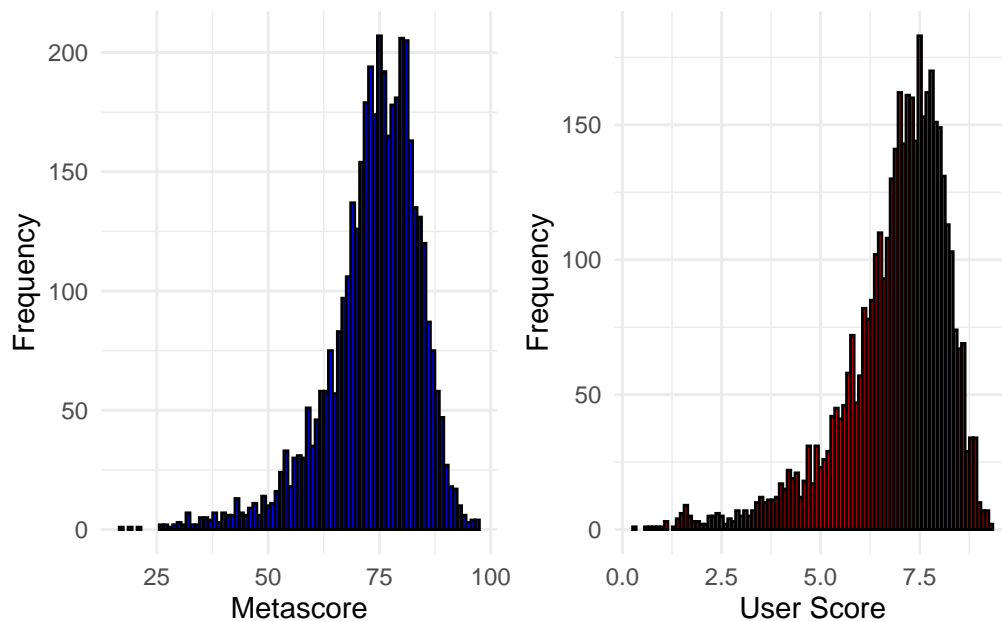

Figure 1: Distribution of scores rated by critics and users on Metacritic

The Metascore and user score in Figure 1 presents the distribution of scores among all of the games in the dataset, showing a similar unimodal and slightly skewed left with peak around

75. Suggesting that most games are considered "good" by critics. While there are more games with low User Scores (below 5) compared to the Metascore distribution, showing useres might be more strict when reviewing a game.

2.4.2 Critics Positivity and User Positivity

The positivity of user and critics is the positivy rate of a critic or user. Note that the sentiment of critics can be positive, mixed, or negative, but they don't necessarily to have a score for review, so the sum of positive, mixed, and negative ratio is not 100% for critics, while usesr have to give a score for reviewing so User Positivity will be exact ratio.

The figure Figure 2 illustrates that critics exhibit a wider spread in their positive review ratios, while user positive ratios tend to cluster around 65%. This indicates that games are generally less likely to receive extreme reviews from users, where overwhelmingly positive or negative feedback dominates. In contrast, critics show more variability, suggesting that some games elicit highly polarized opinions from professional reviewers. This divergence highlights the differing perspectives between critics and the broader gaming audience.
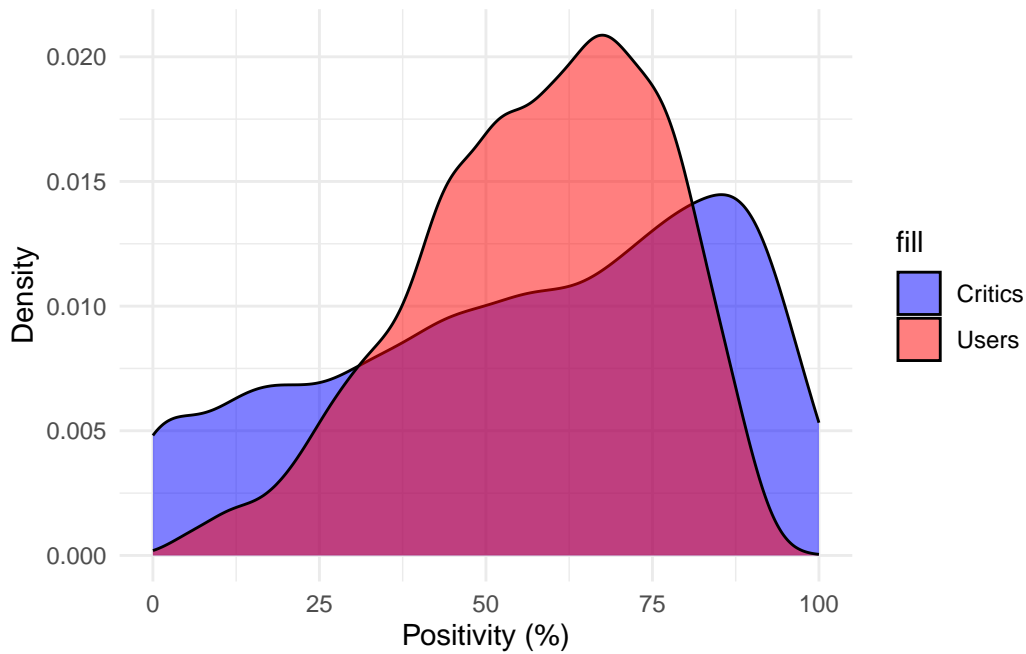

Figure 2: Comparison of Critics Positivity vs. User Positivity Distribution

2.4.3 Users and Critics

The variables Users and Critics represent the number of user and critic reviews for a game, respectively. Generally, a higher number of reviews indicates greater attention and a larger player base. To address the influence of extreme values and improve model stability, the users predictor is log-transformed. This helps reduce the skewness caused by outliers, as illustrated in Table **?@fig-reviews-sum**, where the maximum number of user reviews is significantly larger than the median.

The scatterplot Figure 3 highlights the positive correlation between the number of user reviews and critic reviews, indicating that games attracting more critics also tend to engage a larger player base. This relationship underscores the importance of these variables as predictors in the model, reflecting both professional and community attention toward the games.
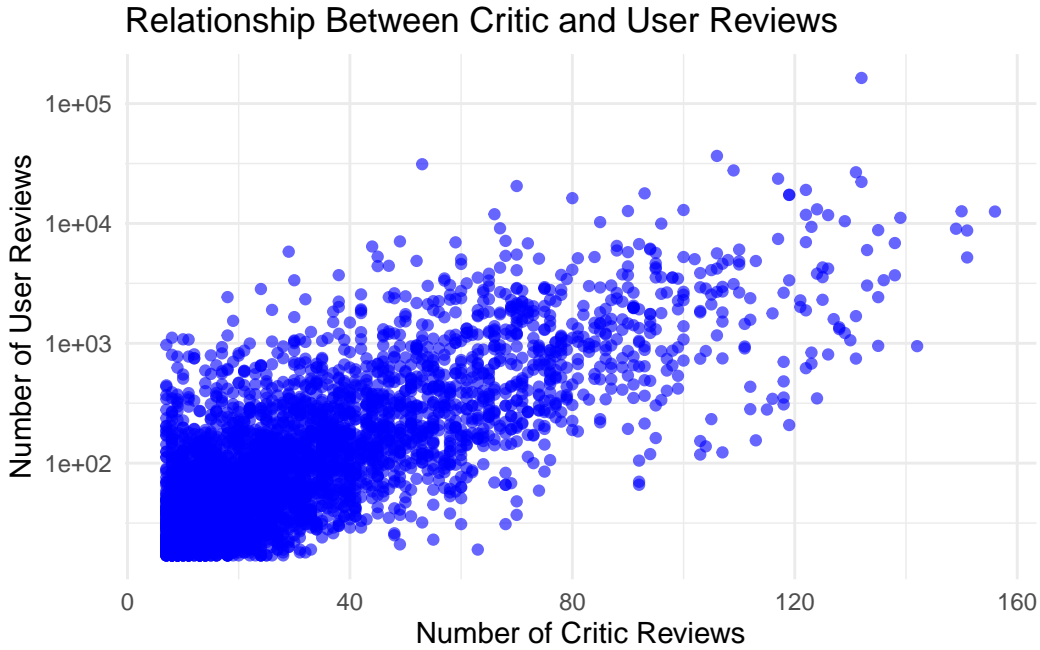

Figure 3

Table 2: Minimum, quartiles, median, and maximum of review numbers for games on Metacritic 2014-2024

Number of user reviews	Number of critics reviews
Min. : 17.0	Min. : 7.00
1st Qu.: 34.0	1st Qu.: 14.00
Median : 76.0	Median : 24.00

7

Number of user reviews	Number of critics reviews
Mean : 467.2	Mean : 32.26
3rd Qu.: 228.5	3rd Qu.: 43.00
Max. :163733.0	Max. :156.00

2.4.4 Genre

Genre is the category of a game, this table Table 3 shows the count of top 10 genre of all games from 2014 to 2024. A game is classified to the most related genre on Metacritic.

Table 3: Top 10 Genres by Frequency in Metacritic Best games 2014-2024

Genre	Count
ActionRPG	292
ActionAdventure	255
2DPlatformer	216
Open-WorldAction	201
FPS	161
Survival	158
Adventure	154
JRPG	146
Point-and-Click	129
Turn-BasedTactics	124

3 Model

The Bayesian logistic regression model is designed to predict whether a game achieves Game of the Year (GOTY) winner based on various predictors I introduced in data section. Posterior checks and diagnostics are included in Appendix D.

3.1 Model set-up

The logistic regression model I will be using is:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \beta_0 + \beta_1 \times \text{score} + \beta_2 \times \text{user\_score} + \beta_3 \times \text{critic\_positivity} \tag{1}$$

$$+ \beta_4 \times \text{critics} + \beta_5 \times \text{user\_positivity} + \beta_6 \times \log(\text{users}) + b_{\text{genre}} \tag{2}$$

8

$$\beta_0 \sim \text{Normal}(0, 2.5)$$
$$\beta_1 \sim \text{Normal}(0, 2.5)$$
$$\beta_2 \sim \text{Normal}(0, 2.5)$$
$$\beta_3 \sim \text{Normal}(0, 2.5)$$
$$\beta_4 \sim \text{Normal}(0, 2.5)$$
$$\beta_5 \sim \text{Normal}(0, 2.5)$$
$$\beta_6 \sim \text{Normal}(0, 2.5)$$
$$b_{\text{genre}} \sim \text{Normal}(0, \tau_{\text{genre}})$$
$$\tau_{\text{genre}} \sim \text{Exponential}(1)$$

where,

- $\hat{p}$ represents the probability that a game is a GOTY winner.
- β_0 represents the intercept term of the logistic regression.
- β_1 is the coefficient corresponding to the **Metascore** (critics' aggregated score).
- β_2 is the coefficient corresponding to the **User Score** (average user rating).
- β_3 is the coefficient corresponding to the **Critic Positivity** percentage (positive critic reviews).
- β_4 is the coefficient corresponding to the **Number of Critics** (total number of critic reviews).
- β_5 is the coefficient corresponding to the **User Positivity** percentage (positive user reviews).
- β_6 is the coefficient corresponding to the **Log-transformed Number of User Reviews**.
- b_{genre} represents the random effect for each game genre.
- τ_{genre} is the standard deviation of the random effects for genres, drawn from an exponential prior.

I run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). I use the default priors from `rstanarm`, reflecting no strong assumptions about the weight of each predictors' effects.

4 Results

4.1 Model result

Table 4 shows a 95% credible interval of each coefficient in the logistic regression model. The large negative intercept -29.92 reflects the rarity of GOTY winners in the dataset, which not give much information. Estimate 0.18 for score(Metascore) and exclusion of 0 on 95%

confidence shows it is associated with increased odds of winning GOTY. Number of user that leave a review also significantly increases the odds of winning GOTY with an estimate of 1.5. While other predictors like positive ratio of users and critics and user score shows slightly positive or negative impact, but the confidence interval that include 0 showing no strong evidence for these predictors' importance. An estimate value of 0.49 suggests moderate genre-specific effects, implying that some genres perform systematically better than others in the model. The Metacritic score shows positive correlation as it reflect the general quality of a game and games with higher score has higher chance of winning GOTY. The coefficient of number of review from user shows games with more user engagement are significantly more likely to win.

Table 4: Summary of the model

term	estimate	std.error	conf.low	conf.high	group
(Intercept)	-29.92	5.81	-39.81	-21.09	NA
score	0.18	0.10	0.02	0.35	NA
user_score	0.80	0.85	-0.56	2.19	NA
critic_positivity	-0.02	0.04	-0.09	0.04	NA
critics	-0.02	0.01	-0.04	0.00	NA
user_positivity	-0.06	0.07	-0.17	0.05	NA
log(users)	1.50	0.37	0.93	2.13	NA
sd_(Intercept).genre	0.49	NA	NA	NA	genre

4.2 Prediction result

To get a prediction by the Bayesian model, I use games that are released in 2024 and generate posterior predictions by the new data frame. The predict probability (mean_prob) was used as the primary metric for comparison because the data contains significant class imbalance (10 winner out of 3993 games for past decade). Given that only a small proportion of games achieve GOTY winner, the model is more likely to assign low probabilities for the majority of games, resulting in most predictions being 0 if using a hard classification threshold like 0.5. By using the mean_prob, it allows us to rank games based on their likelihood of being the winners rather than forcing a binary classification, which are more informative for identifying potential GOTY contenders, especially when making comparisons within the 2024 dataset. In addition, I selected the top 10th percentile of games based on their mean_prob and set their goty_status to 1, marking them as potential winners of GOTY. The results are displayed in appendix Section B. For example, the first row shows that "Elden Ring: Shadow of the Erdtree" has a mean_prob of 0.06675, indicating that it is predicted to win the GOTY award in 267 out of 4000 iterations.

5 Discussion

5.1 Games with Higher Metascores and User Engagement Have a Greater Chance of Winning

From the results section of the model, Metascore and the number of user reviews emerge as key predictors for winning GOTY. A comparison of the distribution of each variable between games that won and those that did not, as shown in Figure 4, reveals that GOTY winners consistently have higher Metascores (a), with minimal overlap. The median Metascore for winners is significantly higher than that for non-winners. This highlights the importance of critic evaluations, which provide a standardized and relatively objective assessment of games, minimizing biases compared to user reviews. This aligns with the Game Awards voting system, where 90% of the voting power is held by a jury of 100+ gaming media outlets, and only 10% is derived from public votes.

Regarding the number of reviews, winners generally exhibit a higher log-transformed number of user reviews compared to non-winners. Both the distributions of critic (e) and user reviews (f) for non-winners skew heavily toward fewer reviews, while winners display tighter clustering at higher values, again with minimal overlap. This indicates that top-tier games garner the most attention. A higher number of user reviews suggests a game has a large, engaged player base, reflecting its widespread appeal and relevance within the gaming community, which gives it a distinct advantage in competition.

Other variables, such as user scores (c) and positivity ratios (d), exhibit significant overlap between the two groups, making them less critical for prediction. While these metrics reflect player sentiment to some extent, they are not precise measures of a game's quality. This is partly because users do not necessarily need to own a game to leave a review, introducing potential biases and inaccuracies. Consequently, these variables lack the differentiation required to effectively distinguish winners from non-winners in the GOTY selection process.

5.2 Genre-Level Analysis

There is ongoing speculation that the GOTY selection process favors certain types of games. Through this study, I compared the genres of past GOTY nominees (2014–2023) shown in the left plot of Figure 5 with the top 10 genres of potential 2024 winners—games predicted to win at least once across all iterations—shown in the right plot of Figure 5.

In both plots, genres such as Open-World Action, Survival, Action RPG, and Action Adventure consistently dominate, suggesting a potential genre bias in the GOTY selection process. This observation is further supported by the genre counts for past TGA GOTY winners in Table 5. For instance, in the past nominees, "Open-World Action" has been a particularly prominent category, whereas in the predicted top genres for 2024, "Action RPG" leads the way. This
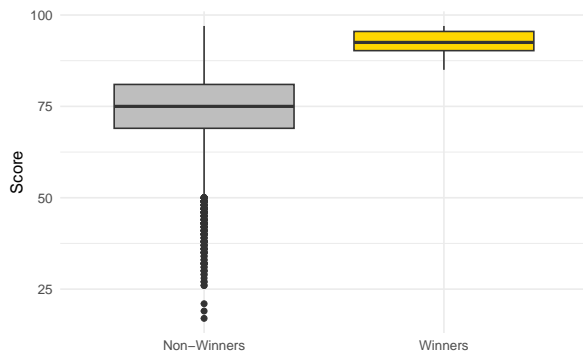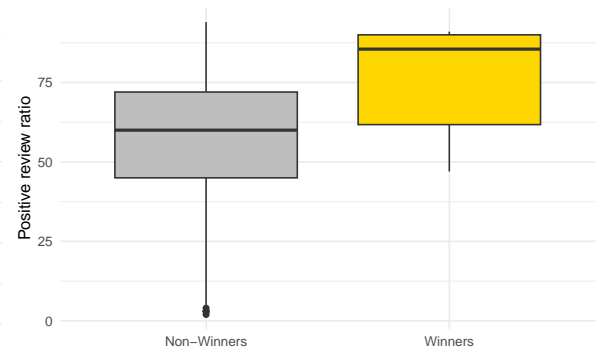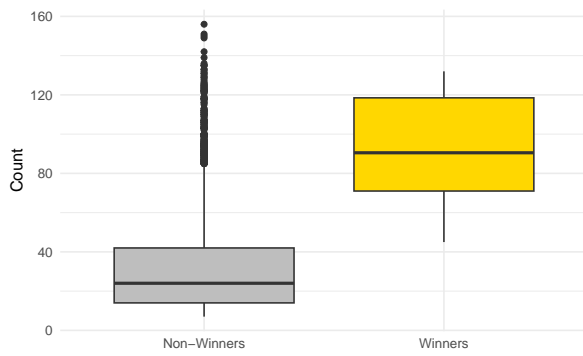
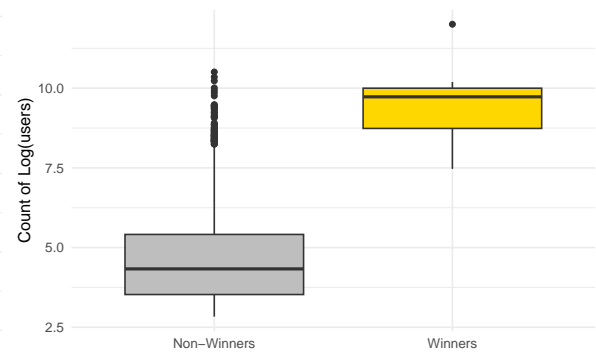(a) Metascore
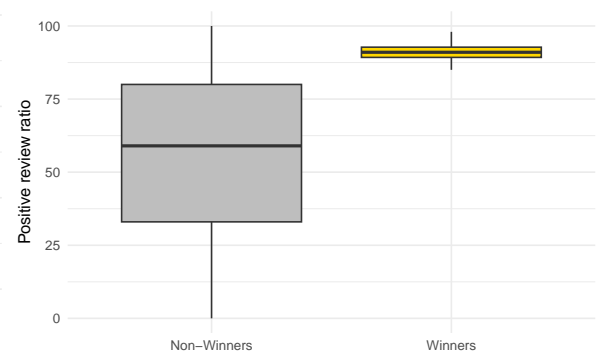
(b) Critic Positivity
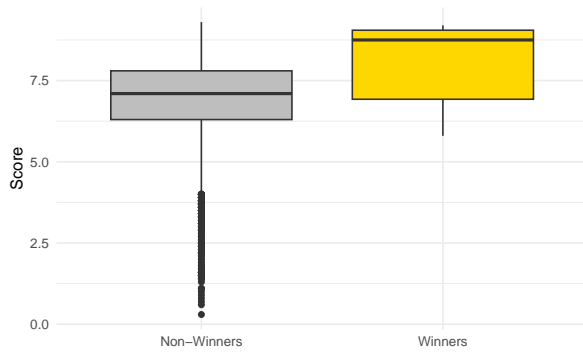
(c) User score

(d) User Positivity

(e) Number of Critic Reviews

(f) Log(Number of User Reviews)

Figure 4: Compare between games that win GOTY and Not

pattern highlights the potential influence of genre preferences on the GOTY outcomes and reinforces the idea that certain genres have a competitive edge.

Table 5: Genre of games won GOTY from 2014 to 2023

Genre	Count
ActionRPG	2
LinearActionAdventure	2
WesternRPG	2
ActionAdventure	1
Open-WorldAction	1
Survival	1
TacticalFPS	1

These genres often deliver expansive narratives, engaging gameplay, and immersive worlds, which align well with GOTY selection criteria that favor innovation, depth, and technical achievements. They typically feature AAA titles that receive significant attention due to their large marketing budgets and development scale. "Open-World" mechanics have been particularly popular among critics and players, likely due to their ability to provide significant replay value and exploration opportunities. Conversly, genres like "Metroidvania" and "Card Battle" have smaller fan bases and limited innovation compared to mainstream genres could also contribute to their reduced likelihood of nomination.

Another thing worth notice is the growth of JRPG(Japanese Role-Playing Game) games, plot Figure 6 shows the trend of average score of JRPG and number of JRPG games released by time. The average JRPG scores show a clear upward trajectory, particularly from 2017 onwards, reaching a peak in 2020 with 81.3 average score of 6 games, and reaching a new peak of this year with an average of 83.0. This indicates a consistent improvement in the quality of JRPG titles, which could be attributed to advancements in game development technologies, storytelling, and gameplay mechanics that resonate with critics and audiences alike. The strong emphasis on narrative-driven gameplay and unique artistic styles may have contributed to this positive reception.

On the other hand, the number of JRPG releases shows a more variable pattern. Between 2016 and 2019, there was a steady increase in the number of JRPG titles, reflecting a surge in popularity during that period. However, the count dropped significantly around 2020, possibly due to the global pandemic affecting production timelines and game releases. Despite this dip, the number of JRPG games rebounded in the subsequent years, signaling a recovery in the genre's production and showcasing a notable presence of JRPGs among potential GOTY contenders.
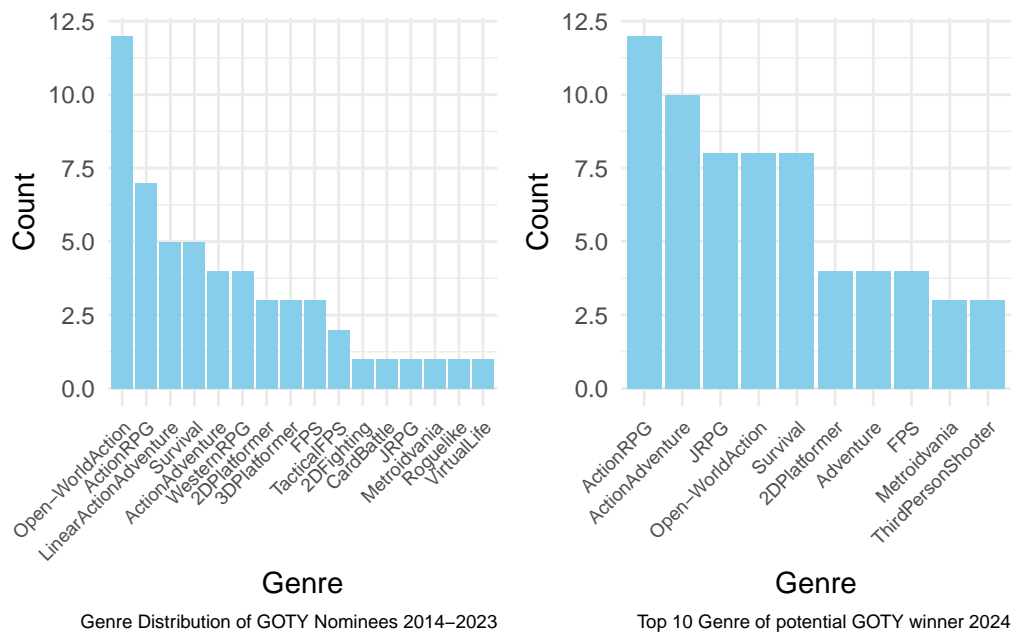
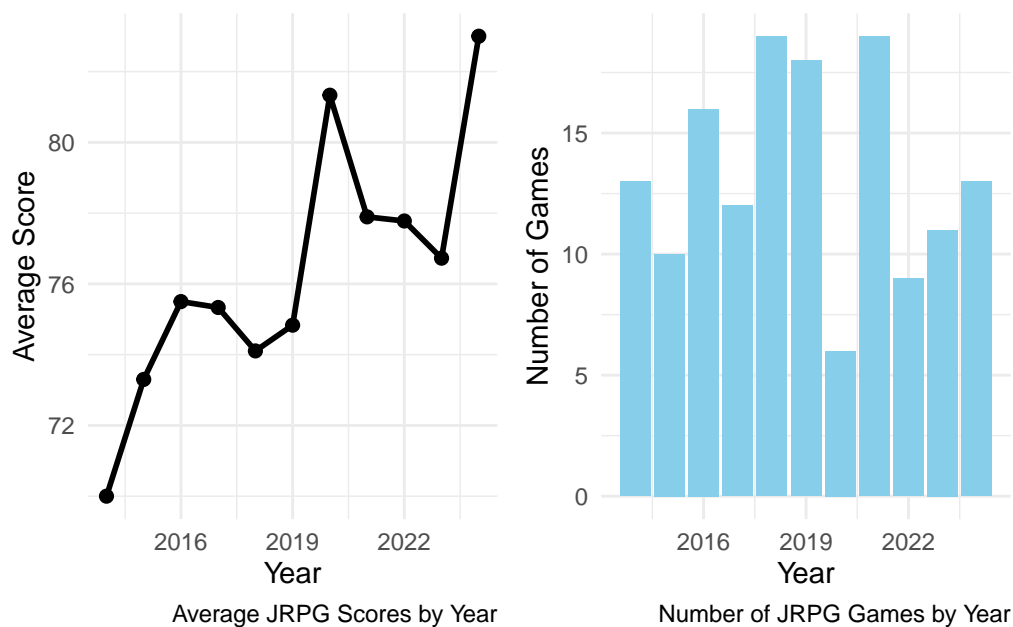Figure 5: Genre of GOTY nominees 2014-2023 and genre of potential winner for 2024

Figure 6: Trend of average score and number of JRPG games by time

5.3 Predict result and actual nominees for 2024

From the predicted results, we observe a significant overlap between the high-ranked games in the model's predictions and the actual nominees for the 2024 GOTY. The six nominees—*Astro Bot*, *Balatro*, *Black Myth: Wukong*, *Elden Ring: Shadow of the Erdtree*, *Final Fantasy VII Rebirth*, and *Metaphor: ReFantazio*—are ranked in positions 6, 22, 5, 1, 4, and 2, respectively, demonstrating the model's strong overall performance.

However, there are notable differences between the predicted results and the actual nominees. For instance, *Dragon Age: The Veilguard* was ranked 3rd in the predictions but did not make the official nominee list, while *Balatro* did, despite ranking 22nd. Two potential reasons for this discrepancy are evident. First, the influence of user scores introduces uncertainty, as previously discussed. While *Dragon Age: The Veilguard* has a significantly lower user score (3.8) compared to *Balatro* (8.3), its exceptionally high number of user reviews (6,825) likely elevated its predicted probability of winning compared to *Balatro*'s modest 488 reviews. This highlights the substantial positive impact of user engagement as measured by the number of reviews.

Second, the genre difference between the two games may explain the divergence. *Dragon Age: The Veilguard* is an RPG with a Western theme, boasting a larger player base and higher popularity of discussion. In contrast, *Balatro* is a card battle game built around the mechanics of Texas Hold'em Poker, a niche genre with a smaller audience. Historically, GOTY nominees have shown a preference for action and RPG titles, with two out of ten past winners coming from Western RPGs. This genre advantage likely strengthened *Dragon Age: The Veilguard*'s standing in the model's predictions while working against *Balatro*'s less mainstream appeal.

When it comes to the actual winner, I believe that DLCs should not compete directly with full games for the GOTY award, as they require ownership of the base game to play. A more appropriate approach would be to establish a separate category for DLCs/expansions or include them in the "Best Ongoing Game" category to recognize their contributions without diminishing the competition among standalone titles. The nomination of *Elden Ring: Shadow of the Erdtree* has sparked significant controversy online, with many questioning whether a DLC should compete for the award.

If we also exclude the unqualified game *Dragon Age: The Veilguard* from the prediction list and focus on nominees, *Metaphor: ReFantazio* and *Final Fantasy VII Rebirth* emerge as strong contenders for GOTY. While *Metaphor: ReFantazio* has a higher predicted probability of winning (0.04075) compared to *Final Fantasy VII Rebirth* (0.02875), the latter benefits from significantly higher user engagement and the enduring popularity of the Final Fantasy franchise. Additionally, *Final Fantasy VII Rebirth* aligns closely with historically favored genres, making it a formidable competitor despite the model's lower probability estimate.

5.4 Weaknesses and next steps

The most significant challenge in this study is the severe imbalance in the dataset, with only 10 GOTY winners out of approximately 4,000 games. This imbalance may limit the model's ability to generalize and predict accurately. Future studies could address this by assigning heavier weights to the minority class (GOTY winners) in the model or by expanding the "winner" class to include all past nominees, which would increase the representation of high-quality games.

Another limitation is the restricted feature representation. Potentially influential factors such as game price, marketing budgets, and player sentiment on social media or other platforms are not included in the dataset. The current model heavily relies on data from Metacritic, which may not fully capture qualitative aspects of games. Future analyses could incorporate data from additional sources like SteamDB (Pavel Djundik 2024), which provides valuable metrics such as 24-hour peak player counts and owner estimates for games released on Steam. Integrating these data sources would offer a more comprehensive understanding of factors driving GOTY success.

Appendix

A Data cleaning

To prepare the dataset for analysis, I followed these steps:

1. Column Selection and Formatting: Removed irrelevant columns such as publisher and developer that were not essential to the analysis. Reformatted the release_date column into a standardized "Month-Day-Year" (MDY) format for filtering.

2. Handling Missing Values, Redundant rows: Removed rows with missing values in key columns like user_score and user_positivity. These missing values primarily occurred for games with too few user reviews, as Metacritic does not provide aggregated scores for such entries. Use "distinct()" to filter out redundant rows, there are 24 redundant rows possibly because of a repeat page in scraping process.

3. Filtering and Grouping Categorical Variables: Filtered out games with fewer user ratings than the 25th percentile (less than 17 user reviews) to improve reliability, as low user engagement can result in unreliable user scores. Combined less common genres (those with fewer than 5 entries) into an "Others" category to simplify the analysis and ensure meaningful comparisons between genres.

4. Adding New Columns: Created a goty_status column to identify Game of the Year (GOTY) winners. Manually compiled a list of past GOTY winners from 2014–2023 (goty_winners) and assigned a value of 1 to games in this list, with 0 for all other entries.

B Predict Result

Table 6: Potential winner of 2024 GOTY

Rank	name	Score	Genre	User Score	critics	users	Predictit prob
1	Elden Ring: Shadow of the Erdtree	94	ActionRPG	8.1	70	2858	0.06675
2	Metaphor: ReFantazio	94	JRPG	8.9	50	1524	0.04075
3	Dragon Age: The Veilguard	82	WesternRPG	3.8	72	6825	0.04000
4	Final Fantasy VII Rebirth	92	ActionRPG	9.0	151	5204	0.02875
5	Black Myth: Wukong	81	ActionRPG	8.3	92	6727	0.02650
6	Astro Bot	94	3DPlatformer	9.2	138	3695	0.02550
7	Stellar Blade	81	ActionAdventure	9.2	138	6864	0.01725
8	Anomaly Agent	84	2DPlatformer	8.8	11	1074	0.01375
9	The Last of Us Part II Remastered	90	Survival	7.7	91	1964	0.01175
10	Animal Well	91	Metroidvania	8.0	33	477	0.01000
11	Dragon's Dogma 2	86	ActionRPG	6.4	91	2059	0.00850
12	Silent Hill 2	86	Survival	9.2	105	2550	0.00700
13	Satisfactory	91	Sandbox	8.9	22	262	0.00700
14	Senua's Saga: Hellblade II	81	ActionAdventure	7.4	100	2544	0.00650
15	Sons of the Forest	86	Survival	7.3	17	175	0.00625

16	Horizon Forbidden West: Complete Edition	89	Open-WorldAction	7.3	38	518	0.00600
17	Persona 3 Reload	87	JRPG	8.6	70	1205	0.00500
18	Call of Duty: Black Ops 6	83	TacticalFPS	6.8	58	1157	0.00425
19	Final Fantasy XIV: Dawntrail	79	MMORPG	5.3	35	1047	0.00425
20	Helldivers 2	82	ThirdPersonShooter	7.4	69	2300	0.00375
21	Destiny 2: The Final Shape	89	FPS	7.6	29	343	0.00375
22	Balatro	90	CardBattle	8.3	34	488	0.00350
23	Tekken 8	90	3DFighting	7.6	98	868	0.00300
24	Star Wars Outlaws	75	Open-WorldAction	5.3	91	2645	0.00250

C Ideal Survey

An ideal survey for selecting the Game of the Year needs to be carefully conduct to show diversity of the stakeholders in game industry and also games itself,

C.1 Sampling Frame

The sampling frame need to include diverse stakeholders in the gaming industry to ensure fairness and inclusivity. The target population can be divided into the following categories:

- Industry Experts: Include critics, developers, journalists, and influencers with expertise in gaming.

- Players: The player samples should cover all gaming market globally and all platforms include PC, console (e.g., PlayStation, Xbox, Switch), or mobile devices.

The ideal sampling size will be approximately 10000, with ~1000 sample from experts and ~9000 from players distributed across demographics and platforms.

C.2 Survey Distributing

First we need to compile a list of industry experts using publicly available sources. We can find developers and publishers from online databases like Moby (Mobygames 2024), and critics or journalists from professional website LinkedIn, Metacritic, and TGA juro list. Once the list is done, an email will be sent to each person.

The players will be recruit through gaming forums, social media platforms (e.g., Reddit, Discord, Twitter),there will be a post with introduction and a link of the survey. Small incentives like discount codes, and in-game rewards can be reward to player who finish the survey to boost participation.

The survey should be conducted annually, 1–2 months before the GOTY announcement.

C.3 Sampling Methodology

Stratified Sampling will be used to divide the population into strata based on expertise (expert vs. player), region, and platform. For example, if we aim to collect 1000 samples, and the sampling proportions is set up with 30% expert and 70% players, with North America (30%), Europe (30%), Asia-Pacific (30%), Other (10%), and from different platforms in PC (40%), Console (50%), Mobile (10%). 210 player respondents will from North America and 84 of them use PC as primary gaming platform.

C.4 Survey Develiery

The survey would be implemented using Google form for general accessibility, with an introduction of what the survey aims for and statement of anonymity for respondents to encourage honest feedback. It also need to obtain ethics review board clearances from data protection regulations.

The questions should be combination of most closed-ended (e.g., multiple-choice, rate by scale) and some open-ended questions to collect information in scale of number and also allow certain amount of freedom from respondents. More closed-ended question can save both respondents time and time to analyze the answers.

Once the survey is designed, there would be a pilot study with a small group of respondents to refine questions and address ambiguities.

C.5 Survey Questions (use nominees for 2024 as example)

This survey is designed to gather insights from players and industry workers like you to identify the best video game of the year. Your input will play a crucial role in shaping a comprehensive, fair, and representative selection process for the GOTY award. The survey includes questions about your gaming preferences, experiences with the nominated games, and your thoughts on what defines a Game of the Year. Your responses will remain completely anonymous, and no personal data will be shared or used beyond this research. This survey should take less than 5 minutes to complete. Thank you for taking the time to share your insights.

1. **What is your primary region?**

- North America
- Europe
- Asia
- South America
- Africa
- Oceania
- Other

2. **What is your primary gaming platform?**

- PC
- PlayStation
- Xbox
- Nintendo Switch
- Mobile
- Other

3. **How would you classify your gaming expertise?**

- Casual Player
- Hardcore Player
- Industry Worker

4. **Which game(s) from the following list have you played?(Check all that apply)**

- Astro Bot
- Balatro
- Black Myth: Wukong
- Elden Ring: Shadow of the Erdtree
- Final Fantasy VII Rebirth
- Metaphor: ReFantazio

5. **On a scale of 1 to 10, how would you rate each of the following aspects of these games?** (will be seperate for games checked previous in question)

- Story/Narrative
- Gameplay/Mechanics
- Visuals/Graphics
- Audio/Soundtrack
- Innovation/Originality
- Replayability

6. **Which game do you believe should win GOTY?**

- Astro Bot
- Balatro
- Black Myth: Wukong
- Elden Ring: Shadow of the Erdtree
- Final Fantasy VII Rebirth
- Metaphor: ReFantazio
- Another game, please specify

7. **Why do you think this game deserves GOTY?** (Open-ended)

8. **Do you believe DLCs/expansions should be eligible for GOTY?**

- Yes
- No
- Unsure

9. **Rank the following genres based on your preference**

- RPG
- Action
- FPS
- Survival
- Adventure

10. **How important are the following factors when deciding the best game?** (Rate each on a scale of 1 to 5)

- Critical acclaim (e.g., Metascore, IGN)
- User engagement (e.g., review from players, number of players)
- Impact/Progress for industry
- Cultural significance

D Model details

D.1 Posterior predictive check

The posterior predictive check graph Figure 7 shows a concentration toward 0, reflects the model's strong tendency to predict a low probability of y=1 (GOTY winner), consistent with the class imbalance in the dataset, where only 10 games out of ~4000 are winners. The closely overlap between y and y_rep This indicates that the model fits the data well.

D.2 Diagnostics

Trace plot Figure 9 shows generally overlapping chains and Figure 8 also shows values close to 1.0 for all parameters indicate good convergence. I didn't apply residual vs. fitted and AUC checks because of the class imbalance of dataset.
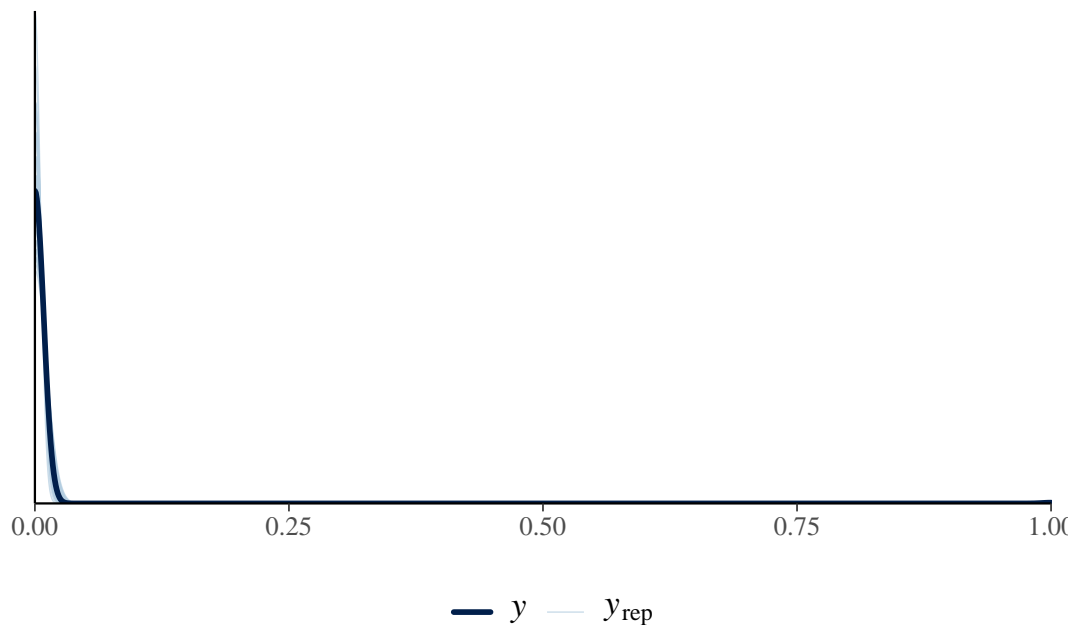
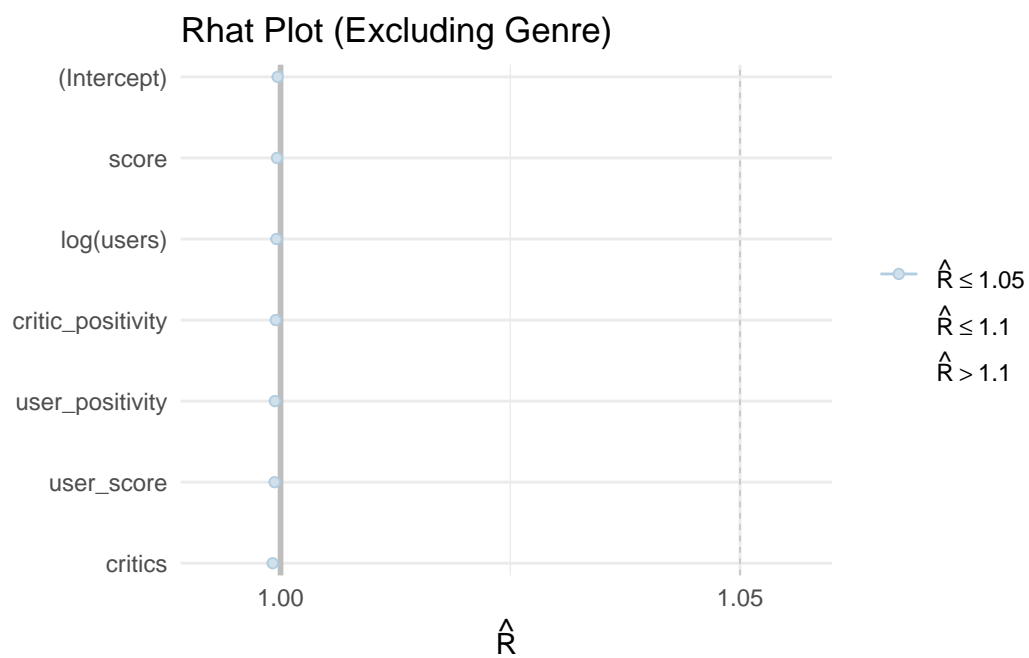Figure 7: Posterior distribution for logistic regression model
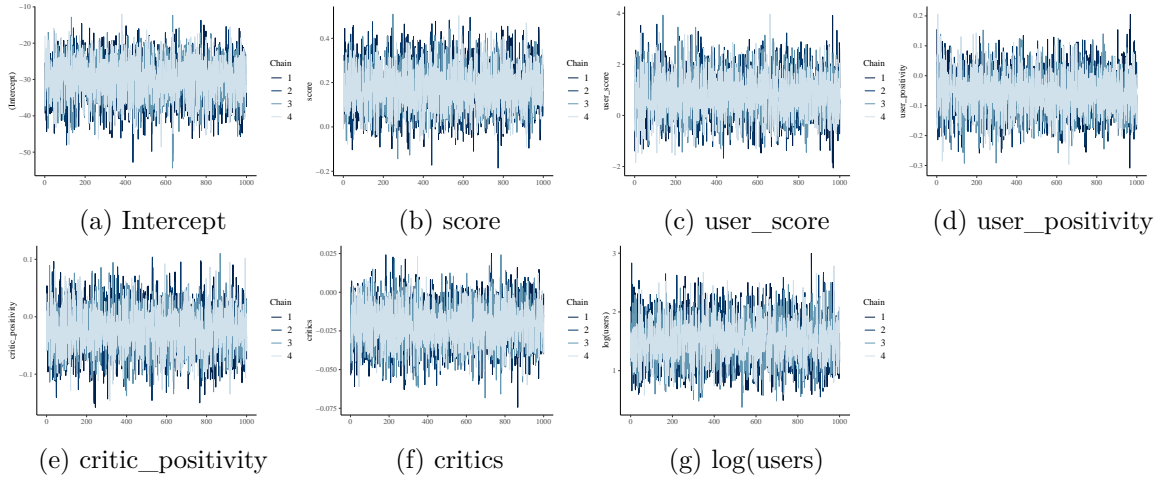

Figure 8: Rhat plot for model

(a) Intercept (b) score (c) user_score (d) user_positivity

(e) critic_positivity (f) critics (g) log(users)

Figure 9: Trace plot

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bolker, Ben, and David Robinson. 2024. *Broom.mixed: Tidying Methods for Mixed Models*. https://CRAN.R-project.org/package=broom.mixed.

Gabry, Jonah, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. 2019. "Visualization in Bayesian Workflow." *J. R. Stat. Soc. A* 182: 389–402. https://doi.org/10.1111/rssa.12378.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Intelligence, Mordor. 2024. "Global Gaming Market - Growth, Trends, and Forecasts (2024-2029)." 2024. https://www.mordorintelligence.com/industry-reports/global-gaming-market.

Metacritic. 2024. "Metacritic: Browse Games by Score." https://www.metacritic.com/browse/game/.

Mobygames. 2024. "List of All Video Game Companies." https://www.mobygames.com/company/.

Pavel Djundik. 2024. "Database of Everything on Steam." https://steamdb.info/.

Pedersen, Thomas Lin. 2024. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

Peters, Jay. 2024. "Sony Is Shutting down Concord Developer Firewalk Studios." The Verge. 2024. https://www.theverge.com/2024/10/29/24282900/sony-shutting-down-concord-firewalk-studios.

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

Ribeiro, Bruno Vieira. 2021. "projectGames." *GitHub Repository.* https://github.com/BrunoBVR/projectGames; GitHub.

Tassi, Paul. 2024. "New Report Says Sony's Concord Cost $400 Million to Make." Forbes. 2024. https://www.forbes.com/sites/paultassi/2024/09/21/new-report-says-sonys-concord-cost-400-million-to-make/.

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.* https://CRAN.R-project.org/package=dplyr.