Investing in Success: How Resource Allocation Decisions Affect User Satisfaction in Digital Games*

An Truong

December 10, 2025

This study examines whether resource-intensive game design features predict user satisfaction on Steam beyond game quality alone. From around 4,000 games(1997-2025), I was able to show that design features signficantly improve predictions compared to Metacritic score alone($F = 18.22$, $p < 0.01$). When including an interaction with game quality and price, these models were not significant($p = 0.05$) compared to the design model, suggesting the design features apply universally. The key significant features acquired include cross-platform compatibility(Mac/Linux Support) and game longevity(days since release). These findings suggest that indie developers should prioritize cross-platform support over other design features such as DLC, Achievements, and extensive language localization when resources are limited. The model explains 36% of the variance in user satisfaction(Adj. R-squared = 0.361).

Introduction

The advancement of technology has made making games more accessible since it was commercially available in 1972 with the release of Pong on the Atari console. Since then, not only have games been accessible but also the resources used to create them. Today, popular video games aren't just made from big AAA title developers such as Electronic Arts(EA), Activation Blizzard, and Nintendo, who have made timeless classics through an expansive talent pool and budget, they are many small developers who are more limited in their resources and budget. Despite their limited resources, smaller or indie developers have made titles that rival if not out-competing AAA titles with positive reviews and customer satisfaction. This raises the

*Project repository available at: https://github.com/antruong1101/MATH261A-Paper2#.

question of whether developers need the high budget resources and game features like Downloadable Content(DLC), language support, or platform compatibility to make games that have high-user satisfaction.

On the game distribution service Steam from Valve, thousands of games are released annually where most developers are working with constrained budgets, thus are limited to what features they can allocate for the game. Because they are not well-sourced like their AAA counterparts, resource allocation decisions weigh more on these smaller developers. Industry guidance on a game's market performance tend to be anecdotal rather than being more rigorous in its quantitative analysis. In this paper, I aim to address these issues using a multiple regression model to understand the association between various game features and user satisfaction to see whether what features sell better than others. Since price and rating scores from accredited rating sources such as Metacritic are popular talking points in a game's value and user satisfaction, their effects with game features will also be looked into in this paper.

In this paper, I will be using multiple regression to see whether these features have a strong association with Steam user satisfaction. Comparisons will be made on models with the main predictor that I believe are important in determining user satisfaction such as price which which is a common talking point in how well a game is and its accessibility as well as Metacritic score . The specific comparisons will be a model with just the main predictor, the second is the main model with all the game features, and last is the full model that adds the interactions between the main predictor and the game features. The metrics of comparison between the models will consist of using the Root Mean Square Error(RMSE), Adjusted-R^2, Akaike Information Criterion(AIC), and F-statistic to measure model significance and fit with the data. Nested ANOVA comparisons are also made to compare significant differences in the zero, main, and full models to better understand what factors contribute most to variability.

The structure of this paper consist of the **Data** sections that covers the dataset in detail, the **Methods** section that covers the models and statistical methods used to understand and solve the problem of this paper, the **Results** section that describe the fitted models and implications, and the **Discussion** section that points out the implications of the key findings towards the research question.

Data

The dataset used in this study comes from the open-source data science platform HuggingFace from the user FronkonGames that consists of over 110,000 published games on Steam(Bustos 2025). Collecting the necessary data included shifting column names, column factoring, filtering for games that had any number of reviews, had more than 0 owners(not publicly released games), and those who had a Metacritic score. After cleaning the data, I was able to work with around 4000 titles from 1997 to 2025 and 23 predictor variables. An alternative dataset that have similar predictors but less titles is SteamDB extraction from Jon Garrastatxu which has around 400 titles(Garrastatxu 2025).

Predictor variables that are most relevant in my analysis: Metacritic score which is a weight average of ratings from professional critics ranging from 0-100. Price which is the market price of the game title. The count of DLC, achievement goals, supported languages for a given title. Indicator variables of whether the game is support on operating systems Mac or Linux. The days since its release date from the date of November 19th, 2025 when I started on this paper. For the response variable, I will be using the percentage of positive reviews which is calculated from dividing the number of positive reviews with its sum with negative reviews.

The bivariate relationships of the positive review percentages with the predictor values are shown in Figure 1. From comparing the relationships of the Metacritic score and the counts for platform, DLC, language support, achievement counts along with price with positive review percentage, all relationships are positive. Although most of the relationships make sense, it is interesting that for the plot with price has a positive relationship since that indicates that as a game gets more expensive the more likely that it will receive greater user satisfaction compared to the cheaper games.

Figure 1: Bivariate relationships between Steam user satisfaction and video game design features.The design features are shown to have positive linear relationships with percenntage of positive reviews.

Methods

The statistical methods analyses done in this paper using the R programming language(R Core Team 2025). The libraries used in this analysis include: tidyverse(Wickham et al. 2019), knitr(Xie 2025), cowplot(Wilke 2025), and kableExtra(Zhu 2024).

To better understand whether or not the listed game features are need in order to have positive user satisfaction on Steam, I compared three regression models which consist of the zero model with Metacritic score by itself, the full model with the listed features, and the model with features and interactions with Metacritic score. From these models, I calculated the RMSE, Adjusted-R^2 value, AIC, and F-statistic values to compare model significance and complexity that is shown in Figure 2. The RMSE for the full model is lower compared to the zero model indicating a better fit but when the interactions are added the RMSE doesn't decrease as much. Similar patterns are shared between the Adjusted-R^2 value where the full model with interactions is slightly better than the full model with no interactions. AIC values shows that including the interactions with the full model actually increases it suggesting that it has too many parameters. Comparing the full model with the zero by AIC values show that the full model has a better balance of fit and complexity than it. As for the F-statistic, it is to note that the model significance is tested between an intercept only model, thus the zero model has the greatest model. Although a higher F-statistic hints better overall model significance, it doesn't mean the interaction model is worse. Thus it is better to compare the models with an ANOVA table instead of the F-statistic as seen in hierachial regression("Hierarchical Linear Regression | UVA Library" n.d.).

Evaluation Metrics	Metacritic only	Design Features	Interactions
RMSE	0.1183578	0.1164996	0.1163526
Adjusted R-squared	0.3417721	0.3611511	0.3617955
AIC	-5673.3603356	-5785.0366682	-5783.0622541
F-statistic	2062.3455267	281.5377689	161.7554700

Figure 2: Evaluation metrics for hierachial models with Metacritic score as primary predictor.Models compared are Metacritic score only, added design features, added interactions between design features and Metacritic score. The main effects model (middle column) optimizes fit.

From using a nested anova table to conduct a hierarchical F-test shown in Figure 3, the comparison between the three models from the evaluation metrics is more interpret-able. Including the game design features is better than the Metacritic score only model with a F-value of `18.22` and when including the interactions, it isn't much better fitting than the game design model based on the F-value of `1.67`. It is also to note that when the anova tables to compare the models, multiple hypotheses test are carried out thus increasing the probability of Type I error.

These results also rely on the assumption of having normal errors or a sufficient sample size which is true in our case with approximately 4000 samples.

Adding design features significantly improved mode fit compared to the Metacritic score only model. This validates that resource allocation towards design decisions including pricing, DLC count, platform support, achievements, multi-language support explains the Steam user satisfaction better than just the video game quality alone. But for the effect of design features on the Steam user satisfaction does not meaningfully vary by video game quality. Although this supports that the design features are optimal compared to the baseline and interactions model, an alternated method that I used was using price as the primary predictor variable. The results had significant results compared to the analysis with the Metacritic score as the primary predictor variable.

Model	Description	k	df	RSS	F	p
1	Metacritic only	1	3969	55.63	—	—
2	+ Design features	8	3962	53.90	18.22	$< .001$
3	+ Interactions	14	3956	53.76	1.67	.050

Note.

k = number of predictors

Figure 3: Hierachial ANOVA table with Metacritic score as primary predictor.Hierachial F-test suggest that the model with design features is optimal compared to Metacritic score only and interactions.

In using the price as our primary predictor variable, the evaluation metrics differ with the price only and the interactions of the design features on price in Figure 4. This evaluation's model with the addition of interactions with price is better compared to the last model with the interactions on Metacritic score, but not by much. It is to note that the design feature models in both evaluations are the same. The RMSE between the design feature model and the model with interactions is a `0.004` difference indicating that model error in predicting on the original data is not that much different and not significant. This pattern continues for Adjusted-R^2 value with an increase of `0.003` which suggest a slightly better fit relative to the number of predictors. What is different from these evaluation metrics compared to the evaluation on Metacritic score is that the AIC value for the interaction model is lower by approximately 17, indicating that the interaction model has a better balance between fit and complexity. These change in evaluation metrics is slightly better fitting, but not worth including the interactions on price.

Evaluation Metrics	Price only	Design Features	Interactions
RMSE	0.1446163	0.1164996	0.1160636
Adjusted R-squared	0.0173092	0.3611511	0.3649626
AIC	-4082.0087435	-5785.0366682	-5802.8175611
F-statistic	70.9278696	281.5377689	163.9714571

Figure 4: Evaluation metrics for hierachial models with price as primary predictor.Models compared are price only, added design features, added interactions between design features and price.

Conducting the same hierarchical anova table like before on the Metacritic score-focused model, including the interactions on price with the design features is better in terms of F-value of `4.96` compared to the design feature model but it is still not significant with a p-value of `0.05` which is right at the threshold in Figure 5. It is also shown that the design features model is significantly better fitting than the price only model which supports that resource allocation towards design decisions explains the Steam user satisfaction better than video game price alone. Although the interaction model is slightly better than the interaction model on Metacritic score, it is still not signifcantly better enough further analyze in this study. After looking into the evaluation metrics of using video game quality and price as primary predictors, the model with just design features is optimal in both in fit and complexity, thus I will be using this in my final analysis.

Model	Description	k	df	RSS	F	p
1	Price only	1	3969	83.05	—	—
2	+ Design features	8	3962	53.90	308.01	$< .001$
3	+ Interactions	14	3956	53.49	4.96	.050

Note.

k = number of predictors

Figure 5: Hierachial ANOVA table with price as primary predictor.Hierachial F-test suggest that the model with design features is optimal compared to price only and interactions.

Since there are many other design features that weren't available in the dataset , I believe that it is still important to name some that are notable and relevant in modern games such as having multiplayer, anti-cheat software, cross-platform compatibility, and micro-transactions, some of which are present in the alternate data source from Zenodo(Garrastatxu 2025). The analyses I have done is on a subset of the many game design features that could effect Steam user satisfaction.

Results

The multiple regression model with percentage of positive Steam user reviews as the response and game design features listed before as predictor variables is modeled as:

$$Y_i = \beta_0 + \beta_1 \text{Metacritic}_i + \beta_2 \text{Price}_i + \beta_3 \text{DLC}_i + \beta_4 \text{Mac}_i + \beta_5 \text{Linux}_i + \beta_6 \text{Achievements}_i + \beta_7 \text{Languages}_i + \beta_8 \text{DaysSince}_i$$

Where:

- Y_i = percentage of positive reviews for game i
- Metacritic score = critic rating (0-100)
- Price = market price (USD)
- Mac, Linux = binary platform indicators
- DLC, Achievements, Languages = feature counts
- Days Since Release = days since release (as of Nov 19, 2025)
- _i ~ N(0, ^{2}) assumed by large sample size `(n = 3,971)`

Predictor	Coefficient (SE)
Intercept	0.2440 (0.0141)***
Metacritic Score	0.0079 (1.811e-04)***
Price	-1.900e-04 (1.787e-04)
DLC Count	-1.379e-04 (1.949e-04)
Mac Support	0.0176 (0.0052)***
Linux Support	0.0187 (0.0056)***
Achievement Count	-2.146e-05 (3.517e-05)
Language Count	5.826e-04 (4.108e-04)
Days Since Release	-8.839e-06 (1.371e-06)***

Note.
Standard errors in parentheses. *** p < .001, ** p < .01, * p < .05. N = 3971, Residual SE = 0.1166, Adj. R^2 = 0.3612.

Figure 6: Coefficient summary of model with video game design features.Predictor coefficient, standard error, and p-value categorization shown for intercept and design features. Cross-platform support(Mac/Linux) show strongest effects beyond game quality.

The summary table of the model with design features shows the coefficients in Figure 6 including standard errors and p-value categorization for the each predictor. To interpret Metacritic score for example, for an increase in score, the expected positive review percentage increases for `0.008` assuming the other predictors are constant. The game design features in this model that have a statistically significant relationship with positive review percentage based on p-value less than 0.05 are Metacritic score, Mac and Linux support, and not necessarily a design feature but a post-launch feature of a video game's age since release. This indicates that overall video game quality score, cross-platform compatibility, and game longevity are meaningfully contribute to explaining the Steam user satisfaction for a given video game in their catelogue. But because the Adjusted-R^2 value is 0.361 which is meaningful since user satisfaction is influenced by a plethora of uncontrolled factors such as personal interest, community dynamics, and genre conventions.

To validate our assumption, the histogram of the residuals and the QQ-plot of the model with design features is shown in Figure 7. The residual plot shows a lack of random scatter and a cluster from the positive review percentages of 0.7 - 1 which reveals a majority of games on Steam tend to have positive feedback. From the histogram, the distribution of errors is approximately normal and the QQ-plot follows the line of best fit.

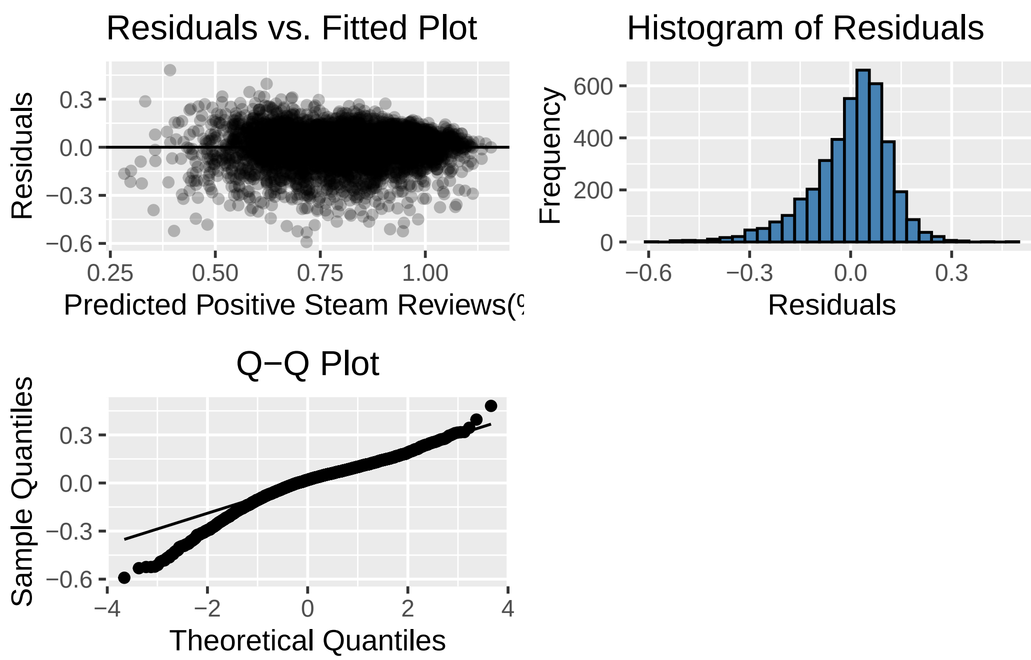

Figure 7: Residual plot between residuals and predicted percentage of positive reviews shows clusters in mostly positve reviews. Histogram of residuals with an approximately normal distribution and QQ-plot that mostly follows the line of best fit suggesting the errors are normally distributed.

Discussion

In this paper, I examined whether resource allocation decisions in game design predict Steam user satisfaction beyond just game quality. From using hierachial ANOVA test and evaluation metrics on nested regression models, there is strong evidence that design features matter ($F = 18.22$, $p < .001$) compared to the models with interaction on game quality or price as the primary predictor (both $p = 0.05$). This suggests that game design features are optimal without additional effect of critic ratings and pricing strategies.

Key Findings

From comparing two hierarchical regression structures with primary predictors on video game Metacritic score and price. The model that was optimal in fit and complexity is the model with the design features with no interactions between the primary predictor. In looking into this model, the most statistically significant predictors in explaining the percentage of positive reviews is Metacritic score, Cross-compatibility(e.g. Mac, Linux), and days since release. This addresses the research question regarding indie developers not needing to have resource-intensive feature decisions such as DLC, achievements, and language support. This allows indie developers to allocate their resources and budget into other features that can be more captivating to users and thus increase user satisfaction.

Future Direction

There is are a plethora of design decisions that greatly influence Steam user satisfaction that are not used in this study's dataset, this introduces potential weakness in this analyses. For future improvements including more features through other video game related datasets or doing a more thorough scrape of a Steam database for this study. Since the Adjusted-R^2 for the model with design features is 0.361, this suggests that there are game design features that are unaccounted for in the dataset that explains the Steam user-satisfaction that can be considered for further analysis.

References

Bustos, Martin. 2025. "FronkonGames/Steam-Games-Dataset · Datasets at Hugging Face." https://huggingface.co/datasets/FronkonGames/steam-games-dataset.

Garrastatxu, Jon. 2025. "Steam Game Metadata with Review Scores Extracted from Steamdb." *Zenodo*. Zenodo. https://zenodo.org/records/17478381.

"Hierarchical Linear Regression | UVA Library." n.d. Accessed December 4, 2025. https://library.virginia.edu/data/articles/hierarchical-linear-regression.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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.

Wilke, Claus O. 2025. *Cowplot: Streamlined Plot Theme and Plot Annotations for 'Ggplot2'*. https://doi.org/10.32614/CRAN.package.cowplot.

Xie, Yihui. 2025. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. https://yihui.org/knitr/.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://doi.org/10.32614/CRAN.package.kableExtra.