Video Games Success Predictor

Florencia Doval, Serena West, Brian Delgado

BACKGROUND

Video games, integral to modern culture, have evolved beyond mere entertainment, shaping our interactions and leisure activities. With digital technology deeply ingrained in our lives, it's crucial to grasp what drives video game success. Analyzing and predicting this success unveils insights into not only consumer behavior and market trends, but the socio-historical and economic impact of the gaming industry.

HYPOTHESIS

1. Video games released on multiple platforms have higher average review scores than those released on a single platform.
2. The average review score of games from the "Action" genre is significantly different from the overall average score of video games.
3. The distribution of review scores is independent of the game's platform (e.g., PC, console).

DATA COLLECTION

We collected our data from the top 3 online game databases: igdb_games, MobyGames, and GiantBomb. Our final joined database consisted of more than 1000 entries with game's average score,release_date,genres,involved_companies,rating, etc. From an igdb API we extracted thousands of game names, platforms, and user_ratings. And, from GiantBomb we extracted all the remaining features to support our analysis.

HYPOTHESIS TESTING

Methodology:

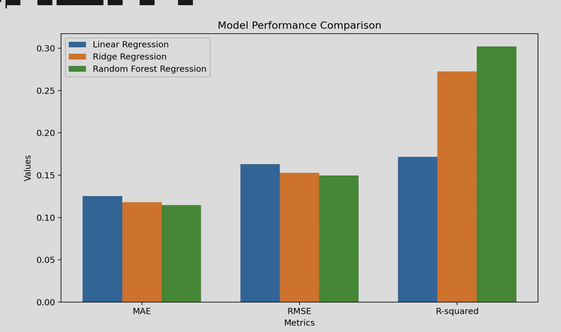
1. : Two-sample t-test was used. Our p-value caused us to accept the null hypothesis, which states that multiplayer and single player games have no significant difference in ratings.
2. A two-sample t-test compared the average scores of games identified as Shooters with those that are not. The p-value came to about 0.764, even higher than the first tested hypothesis. Meaning that shooter and non-shooter games have no significant difference in ratings

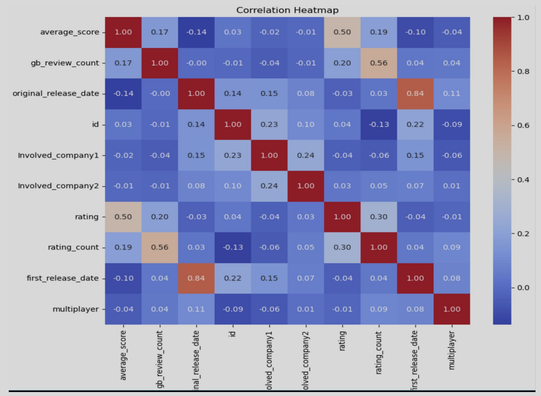
3. The dataset was divided into two groups: games with reviews greater than the median are categorized as 'high review' games, and those with reviews less than or equal to the median as 'low review' games. A two-sample t-test was used to compare the average scores between these two groups. With the low p-value, we were able to reject the null hypothesis which suggests that in most cases, people are more likely to leave reviews if they have positive comments for the game.

	Hypothesis 1	Hypothesis 2	Hypothesis 3
T-statistic	-0.740	-0.30	6.521
P-value	0.460	0.764	1.122

MACHINE LEARNING COMPONENT

Three regression models (Linear, Ridge, and Random Forest) were evaluated using MAE, MSE, RMSE, and R2 score metrics via cross-validation. Random Forest emerged as the best performer, adept at capturing non-linear relationships in our feature-rich dataset. While we addressed data inconsistencies, further feature engineering is needed for optimization. Lower MAE, MSE, and RMSE values denote better performance, with a higher R2 score indicating a superior fit. Overall, Random Forest Regression excelled, demonstrating its ability to handle complex data and yield accurate predictions across different splits. Further improving the Random Forest predictor is relevant in the context of the game industry where player behavior and game success are influenced by complex, non-linear relationships such as gameplay mechanics, graphics, storyline, genre, and platform.





ML RESULTS

- Random forest regression has a similar performance to linear and ridge regression in our dataset, but it typically yields better predictive performance, especially for our complex non-linear dataset.
- The higher R-squared (R2) score of 0.302 suggests a better fit of the model to the data, implying that approximately 30.2% of the variance in video game success can be explained by the features included in the model.

CONCLUSION

These findings underscore the complexity of predicting video game success and suggest that while certain factors like review volume play a significant role, others such as game genre or multiplayer capability might not have as straightforward an impact. The insights from this study can assist developers and marketers in understanding the nuanced influences on video game success, possibly guiding decisions in game development and promotional strategies.

LIMITATIONS NEXT STEPS

Our Machine Learning models have a relatively low accuracy for different metrics, further hyperparameters tuning, feature engineering and cross-validation can help improve our results. Because of this, our model will not be as strong for other unfamiliar datasets.