

Lizzy Brunn

DSCI 510 Final Report

Monday, December 15, 2025

Indie Game Success Analysis

Project Description: As game development becomes more accessible, competition for indie game developers rises. Despite advances, game development remains a lengthy, time consuming, and oftentimes expensive process. Considering its hefty monetary and time-related costs, insights into what factors predict a game's success are very valuable to developers. While large AAA development studios have entire teams dedicated to market research and predictors of success, indie studios that lack such teams must compete in the same diverse and competitive landscape. As a result, researching the attributes that correspond with success for indie titles, specifically, can aid indie developers in deciding the direction of their game.

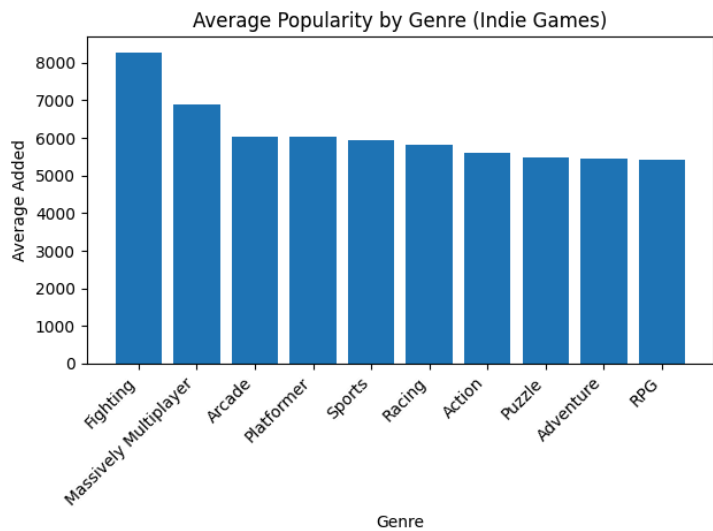
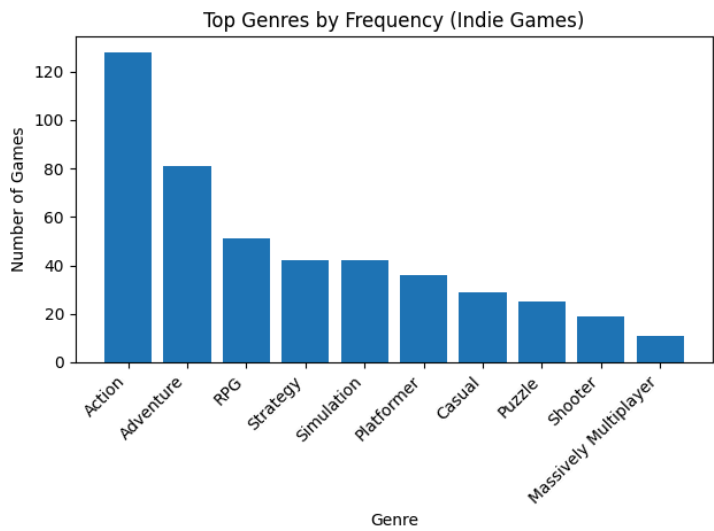
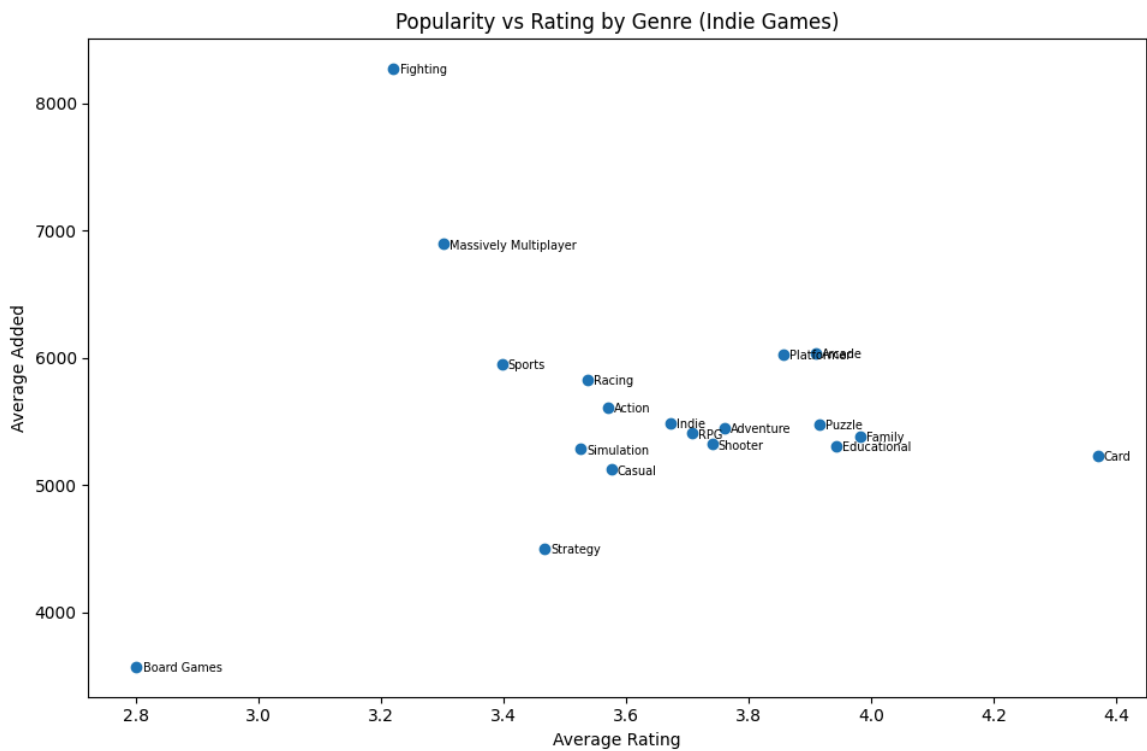
Research Question: What genres and tags are most strongly associated with success in Indie games across platforms?

Data: Using the RAWG Video Game Database API, which stores comprehensive data on 500,000 games across 50 platforms, I retrieved information on 750 of the most popular games. Since this project is primarily concerned with Indie game development, this sample was filtered for games that included the "Indie" genre tag. The final dataset for analysis consisted of 207 Indie games.

Data Cleaning: After collecting information on 750 games, the data was cleaned to only include categories useful for this analysis. These included: title, release date, rating (the rating score), ratings_count (the number of ratings), added (how many people purchased/added this game to their library), genres (all genres the game belongs to), and tags (specific game attributes). Rows containing missing or invalid values for key numeric variables were excluded to ensure consistency in analysis. Then, in order to remove large studios and focus the sample on indie games, the data was filtered to only include games that had “Indie” as one of their genres.

Analysis/Results: To analyze the attributes of successful games, a number of analyses were conducted on the cleaned dataset. I performed descriptive statistics to determine the average added and rating across all titles, as well as the top genres/tags by frequency, average added, and average rating. A Pearson correlation was conducted to examine the relationship between popularity (library additions) and perceived quality (rating score). The results found that the average added across all the indie games was 5487.39, and average rating was 3.57 (out of 5). The most frequent genres, excluding Indie, were Action, Adventure, and RPG. The three most popular genres, those added to the most libraries, were Fighting, Massively Multiplayer, and Arcade. The most acclaimed genres, with the highest ratings, were Card, Family, and Educational. The most frequent tags were Singleplayer, Steam Achievements, and Full Controller Support. The top tags by average added were Dark Fantasy, Dark, and Action-Adventure. The top tags by average rating were Surreal, Hand-Drawn, and Action-Adventure. Finally, the Pearson correlation between rating and added returned a score of 0.305.

Visualization: Using Matplotlib, three primary visualizations served to better illustrate these findings and to uncover further insights. Bar charts were used to display and contrast frequency and average popularity by genre. A scatter plot was used to visualize the relationship between popularity and rating at the genre level.



Premise and Conclusion:

These results uncovered several insights that answer this paper's research question: What genres and tags correlate most strongly with success in Indie games across platforms? When examining popularity and perceived quality measured across genres and tags, notable differences emerged across both genre and tag. Findings suggest that a variety of genre-related factors and tag-specific attributes contribute to indie game success. Since "success" is a broad term, this project examines it across two dimensions, popularity and acclaim, and discovered variation in results between the two. Genres with the highest popularity varied from those with the highest acclaim, and the same pattern appeared for tags. The result of the Pearson correlation test between rating and added suggest a weak positive relationship between a game's popularity and its acclaim. Additionally, this relationship varies by genre. Some genres, like Fighting, see a lower rating score relative to very high popularity. In contrast, the highest rated genre, Card, has a popularity similar to many other genres. These results mean that popular genres do not necessarily have equally high ratings, and the critical acclaim of highly rated genres does not mean that genre is the most popular. Depending on developer intentions and personal goals, these results can be utilized in a number of ways to inform design decisions. Developers may be more concerned with acclaim rather than popularity, or vice versa, and may shift their game's attributes to reflect those that excelled in their target category. They may also seek to garner a balance of popularity and acclaim, and choose attributes like Arcade and Action-Adventure that showed top averages in both metrics of success.

Changes from Original Proposal: While this project did not change much conceptually, the API used for data collection changed between the project proposal and the final product. Initially

seeking to use [Itch.io](https://itch.io), API restrictions made this functionally impossible for the amount of games I sought to collect. After running into similar issues using the Steam API, the RAWG Video Game Database API presented a robust solution that not only executed easily and without error, but provided cross-platform information that strengthened the overall findings. This API change meant some of my target variables also changed. RAWG does not have reliable price information or release date, so the attributes examined shifted solely to genre and tag, which RAWG held a more robust list of. Additionally, RAWG has slightly different information surrounding performance, so I decided to use library add count and rating score as the metrics for success. Since this database included many large developers, I increased the sample size of data collected to 750 so that more indie games could be captured. After filtering this dataset for only indie games, the final data used for analysis consisted of 207 indie titles, smaller than the originally proposed 300-500 games.

Future Work: Future work could expand this project in many ways. If an API provided it, examining pricing and release date would add a useful layer of insight into indie game success. Additionally, collecting a larger sample of games would increase the strength of these findings. Since this project was looking at factors that determine popularity, these games were collected from the most popular games in the API database. However the average indie developer may desire insight into game performance beyond that of wildly successful titles. Future research could explore methods for randomly sampling titles and extend the analysis to include predictors of underperformance or limited reach. Examining both success and failure would provide a more complete picture of the indie game market.