Exploring Indie Games on the Steam Games Dataset

1st David Demitri Africa
Department of Information Systems & Computer Science
Ateneo de Manila University
Quezon City, Philippines
david.africa@obf.ateneo.edu

2nd Mark Kevin A. Ong Yiu
Department of Information Systems & Computer Science
Ateneo de Manila University
Quezon City, Philippines
mark.ongyiu@obf.ateneo.edu

Abstract—Our paper discusses the significance and growth of indie games in the video game industry, with a focus on Steam as the primary platform. We highlight that indie games have become a major driving force in the industry, despite their small development teams and limited marketing. This paper presents insights derived from a dataset obtained from Steam, exploring the landscape of indie games, their growth, player engagement, and characteristics. We conclude that indie games are popular due to their accessibility, diverse genres, and lower prices. Furthermore, we identify the prevalence of casual and multiplayer games among top indie titles.

Index Terms—indie games, video games, Steam, players, data visualization

I. INTRODUCTION

Minecraft, with over 200 million sales [1], is one of the best-selling video games of all time. It also happens to be an "indie" game (defined as having been built and maintained by a small team of developers) at least until its acquisition by Microsoft in 2014 [2]. Despite being characterized by non-existent institutional capital and little marketing, indie games (like Minecraft, and indeed others like Undertale, Hades, and Stardew Valley) have become the backbone of the modern video game industry. From Fig. 1, it immediately becomes clear that indie games make up a significant portion of all games in Steam, the largest personal computer (PC) gaming distribution platform [3].
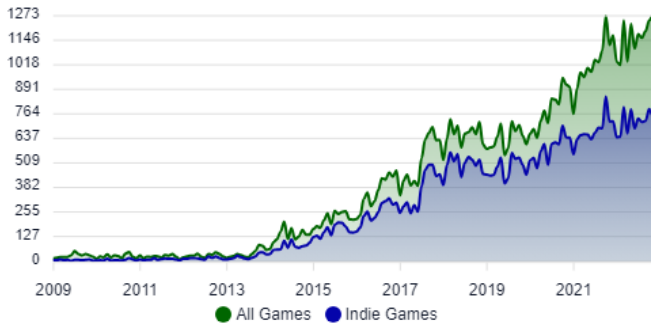

Fig. 1. The number of games (in green) and the number of indie games (in blue) released in Steam from Jan 2009 to Dec 2022

In this paper, we explore the different games in Steam to uncover valuable insights regarding the neglected aspects of indie games.

II. MATERIALS AND METHODS

Our dataset was taken from the "Steam Games Dataset" on Kaggle, which was scraped by Martin Bustos (or Fronkon Games) using the Application Programming Interface (API) provided by Steam [4]. The dataset contains extensive information on $74,689$ different games available in the Steam platform. These include the title, description, and price (in USD) of each game, the date that each game was released, the developers and genres of each game, the Metacritic score (aggregated score from various reviews) of each game, and many other relevant information.

Thankfully, the scraping process was already quite thorough, so the dataset required minimal cleaning. Cleaning steps mostly involved the removal of nonsensical years, i.e., games that were released before the launch of Steam as well as games that have not yet been released. For this purpose, we have arbitrarily chosen to focus on games released between 2009 and 2022 ($n = 67,670$). Moreover, a significant number of games had a Metacritic score of 0. We have assumed that this indicates that a game has not yet been rated in Metacritic. Thus, for any aggregation or analysis involving the Metacritic score of various games, we have chosen to ignore games that have a Metacritic score of 0.

In this paper, games are classified as "indie" based on whether or not the games includes "indie" as one of their genres, resulting in $46,547$ indies games from 2009 to 2022.

To further explore this dataset, we have created an interactive visualization and data story, which can be found on this website. The website was built using NextJS [5], with the data stored in MongoDB Atlas [6]. Visualizations for genres other than indie games can also be found in the website.

An important assumption being made in this analysis is that the Steam Games dataset is a representative subset of the gaming industry. This is a reasonable assumption to make because of Steam's market share and significance in video game distribution. However, it is possible that some types of games are deliberately excluded by the platform (i.e., games using AI art [7]) which reduces the representativeness of the dataset. Future studies may consider exploring other video game distributors.

III. LANDSCAPE OF THE GAME INDUSTRY

A first glance at the dataset shows that a significant portion of the games are classified as indie games (see Fig. 1). We

visualize this using a big angry number (BAN). Indie games make up a big portion of existing Steam games, comprising 68.79% out of the 67,670 games on the platform released from 2009 to 2022.

In addition, we also found that the number of indie games enjoyed steady growth through the years. This trend was visualized in two ways. First, a heatmap over time to show the number of indie games released yearly going from tens in 2009 to thousands in 2022. This heatmap can be found on the website. Second, we also show indie games as a proportion of total games, representing this graph as a stacked area chart (see Fig. 1). We find that monthly growth in indie games closely follows the same spikes and trends in the broader dataset, with a correlation between the two time series of $r = 0.99$.

Finally, we sense-check these findings by comparing it to the growth in other genres. We note that percentage metrics (like relative change) are not so sensible considering that some genres are small enough to make it sensitive to outliers (like the "360 Video" genre). In terms of absolute size, indie games ($n = 46,547$) outstrip the largest game genres like action ($n = 27,354$) and adventure ($n = 25,253$) in terms of number of games released significantly. This is visualized as a heatmap in the home page of the website.

IV. Growth and Player Engagement

After establishing the prominence of indie games, the natural next question to ask is: Why are indie games so popular? While a serious statistical inquiry into the causal variables (i.e., using Granger causality [8] and transfer entropy [9]) is outside the scope of this paper, we have three conjectures on the possible causes of indie game popularity, which we will assess based on visualizations and correlations.

Our first conjectured cause is that indie games are accessible to consumers due to low prices. We visualize this using a line graph of the average monthly price of indie games released in that month (see Fig. 2). Interestingly, we find no clear trend in the graph. This is unexpected because we should expect prices to increase over time due to inflation. Using least squares regression to find the line of best fit gives us a coefficient of $m = -0.0105$, confirming this as it is very close to zero. Hence, this may be a good reason why indie games are popular. That is, as more and more developers publish games at an affordable price, the more accessible they get to a wider audience.

Our second conjectured cause is that indie games enjoy a lot of diversity in genre. To visualize this, we use a heatmap of yearly game releases divided by genre (intersection with indie genre). The margins of this paper is too narrow to contain the heatmap, but this can be viewed in our website, under the indie games data story. From the heatmap, we find that not only do indie games currently have releases in all genres, they also experienced an expansion in genre coverage over the years. The number of different genres of indie games grows from 9 in 2009 to 20 in 2022. This is evidence that the number of different audiences captured by indie games increased as the coverage of different styles and interests increased.
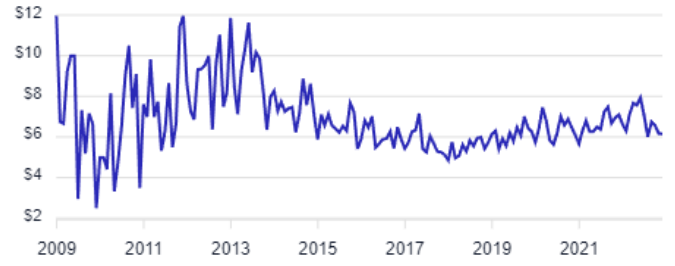

Fig. 2. The average monthly price (USD) of indie game releases from Jan 2009 to Dec 2022

Our third conjectured cause is that the number of indie games pushed competition, increasing the quality of games over time. To assess this we can take the average Metacritic rating of indie games over time to see if games really have been improving (see Fig. 3). Looking at the line graph of the average Metacritic rating over time for indie games, we find no particularly strong trend. Thus, this is not great evidence for indie game popularity either way.
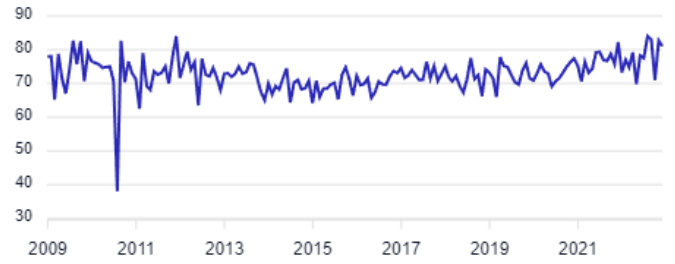

Fig. 3. The average Metacritic score of indie game releases from Jan 2009 to Dec 2022

Overall, the first two causes are promising leads for further causal investigation. On the other hand, Metacritic ratings may not be a great proxy for actual game quality. However, it is also plausible that the standards of the average gamer have risen over time.

V. What is the Profile of the Modern Indie Game?

To further understand indie games, we also want to come up with some ways to concretize the profile of the modern indie game. To do this, we ask some basic questions about the kinds of titles audiences might be familiar with. First, we take a look at the top 10 indie games by peak concurrent users in Table I.

We note that while they are a diverse group, they do share some common characteristics. First, most of these games are in the survival, crafting, or action-adventure genres. This suggests that these genres are particularly popular with indie game developers and audiences. Second, many of these games were released in the last few years, which indicates that the indie game scene is still very active. Third, the majority of these games are multiplayer, which suggests that indie games are

increasingly being designed for social play. Finally, all of these games are available on multiple platforms, which makes them more accessible to players.

TABLE I
TOP 10 INDIE GAMES BY PEAK CONCURRENT USERS

Category	Peak Concurrent Users
ARK: Survival Evolved	98215
Rust	96112
Wallpaper Engine	85643
Warhammer 40000: Darktide	63616
Unturned	54184
Cult of the Lamb	53272
Stardew Valley	39710
Raft	39404
Terraria	39165
Stray	36912

We also looked at the top 10 indie developers by number of games released, remarkable for their prolificness in publishing (see Table II). Like the top games, these developers also share some similarity between one another, with most of them focusing on casual or visual novel games. This suggests that these types of games may be easier or more appealing to develop. Many of these developers were active for many years, owing to a solid track record of releasing games to establish themselves. Finally, these developers are also mostly based in Europe, which suggests that the indie game development scene there may be more well-developed.

TABLE II
TOP 10 INDIE DEVELOPERS BY TOTAL GAMES RELEASED

Category	Number of Indie Games Developed
Choice of Games	142
Laush Dmitriy Sergeevich	130
Sokpop Collective	99
Boogygames Studios	97
Reforged Group	94
Hosted Games	89
Cristian Manolachi	88
Atomic Fabrik	87
Tero Lunkka	83
Blender Games	71

Going back to the heatmap of yearly game releases, we note that the most common subgenre of indie games are those in the "casual" subgenre. This suggests that it may be easier (or preferable) for small developer teams to focus on a lighter, more casual experience instead of a longer, heavier game that demands more technical expertise and quality control. We also note that this is followed by indie games in the action and adventure subgenres, suggesting that indie developers are interested in catering to younger audiences that are looking for a more fast-paced and exciting experience.

VI. CONCLUSION

To conclude our analysis, we generate three main insights. First, indie games are the backbone of the video game industry, having the most games out of any genre, and the most developers as well. Second price seems to be a major driver of indie game popularity, especially when you consider that the price has declined slightly over the years when it should really be going up due to inflation. Third, the largest subgenre of indie games is casual, suggesting that indie game audiences enjoy a more relaxed or leisurely gaming experience.

These insights can drive decision points for various audiences. For instance, indie developers considering which subgenre to make their new game in can choose to make a casual game, to tap into the already thriving casual indie game market. On the other hand, indie developers that are looking to stand out and find their own niche can explore other subgenres that do not have many games.

For investors, they may consider investing in indie games, given their significance in the video game industry as whole. Indie games show a lot of growth potential with a much lower production costs compared to AAA games.

Lastly, for video game enthusiasts, this can provide a new look at a genre of game that they might have taken for granted. This could even convince them to take a chance on an indie game they might have previously dismissed.

Nowadays, modern indie games are casual, often shorter and less complex than AAA games, making them more accessible to a wider audience. They are multiplayer, with many of the most popular indie games being designed for social play, allowing players to connect with friends and strangers online. They are versatile, often available on multiple platforms and in many genres, making them more accessible to players regardless of audience.

These characteristics have helped to make indie games a popular and vibrant part of the gaming landscape. They offer a unique and refreshing alternative to AAA games, and they continue to grow in popularity with each passing year.

REFERENCES

[1] J. Sirani, "The 10 Best-Selling Video Games of All Time," IGN, Jun. 13, 2023. https://www.ign.com/articles/best-selling-video-games-of-all-time-grand-theft-auto-minecraft-tetris (accessed Jul. 20, 2023).

[2] K. Stuart and A. Hern, "Minecraft sold: Microsoft buys Mojang for $2.5bn," The Guardian, Sep. 15, 2014. Accessed: Jul. 20, 2023. [Online]. Available: https://www.theguardian.com/technology/2014/sep/15/microsoft-buys-minecraft-creator-mojang-for-25bn

[3] C. Edwards, "Valve Lines Up Console Partners in Challenge to Microsoft, Sony - Bloomberg," Nov. 04, 2013. Accessed: Jul. 20, 2023. [Online]. Available: https://www.bloomberg.com/news/articles/2013-11-04/valve-lines-up-console-partners-in-challenge-to-microsoft-sony

[4] M. Bustos, "Steam Games Dataset." https://www.kaggle.com/datasets/fronkongames/steam-games-dataset (accessed Jul. 20, 2023).

[5] "Next.js by Vercel - The React Framework." https://nextjs.org/ (accessed Jul. 20, 2023).

[6] "MongoDB: The Developer Data Platform," MongoDB. https://www.mongodb.com (accessed Jul. 20, 2023).

[7] P. Tassi, "Steam Is Reportedly Rejecting Games Using AI Art," Forbes. https://www.forbes.com/sites/paultassi/2023/06/29/steam-is-reportedly-rejecting-games-using-ai-art/ (accessed Jul. 20, 2023).

[8] A. Shojaie and E. B. Fox, "Granger Causality: A Review and Recent Advances," Annu. Rev. Stat. Appl., vol. 9, no. 1, pp. 289–319, Mar. 2022, doi: 10.1146/annurev-statistics-040120-010930.

[9] P. Shahsavari Baboukani, C. Graversen, E. Alickovic, and J. Østergaard, "Estimating Conditional Transfer Entropy in Time Series Using Mutual Information and Nonlinear Prediction," Entropy (Basel), vol. 22, no. 10, p. 1124, Oct. 2020, doi: 10.3390/e22101124.