

THE CURIOSITY CUP 2025

A Global SAS® Student Competition

Team: SASquatch

Market Dynamics and Success Factors of Indie Games on Steam

Vladyslav Horbatenko, Jonas Clausen, Salar Komeyshi, Roskilde University

ABSTRACT

Indie games have become a major force in the gaming industry, offering unique and innovative experiences that differ from A-list titles. This paper analyses the market dynamics of indie games on Steam by analyzing game metadata, sales insights, user reviews, and descriptive tags using SAS tools such as SAS Studio and SAS Viya Visual Analytics. Our findings reveal that indie games represent a significant portion of the platform's library at around 63%, often receiving high user ratings despite having smaller budgets and limited marketing resources. In spite of this, mainstream games continue to dominate in terms of volume of reviews and visibility. The paper highlights potential key factors contributing to indie game success, including community-driven promotion and quality-driven engagement. These insights provide valuable guidance for developers and industry stakeholders aiming to navigate the competitive digital gaming landscape.

INTRODUCTION

Indie games have reshaped the gaming industry by offering innovative, creatively driven experiences that differ from blockbuster franchises. Unlike mainstream titles, indie releases operate on smaller budgets and rely heavily on community-driven promotion, making their market dynamics fundamentally distinct. This study seeks to understand how these dynamics influence the place of indie games on the market.

We examine Steam's platform by analyzing game metadata, sales insights, user reviews, and descriptive tags, applying advanced techniques in SAS Studio and SAS Viya Visual Analytics. Our focus includes release strategies, consumer reception, and engagement patterns, aiming to uncover the key factors that drive indie success. These findings will illuminate how indie developers can thrive in a competitive digital entertainment landscape, ultimately shaping their market position and long-term viability.

ANALYSIS

In the following section we will approach our analysis in multiple steps. It is also possible to check out our fork of the original dataset in [github \[1\]](#).

DATA PREPARATION AND PREPROCESSING

We source our data from the *steam-insights* repository on GitHub [2]. While the dataset has a good level of completeness, it also comes with significant inconsistencies. We encountered blank (\n) cells, varying currencies, and duplicated entries within csv files. Additionally, some cells contain HTML and JSON within them. This poses challenges since SAS software

couldn't properly read those cells in pure form, as their longer code often spilled into multiple rows.

To analyze our data, we clean and process it using SAS Studio and SAS OnDemand for Python. The dataset consisted of seven separate tables, all in CSV format, with data linked together through *app_id* – a unique identifier assigned to each game. The tables were:

1. categories.csv: List of the categories of games.
2. games.csv: List of a general overview of games
3. genres.csv: List of all the genres of games
4. reviews.csv: List of all the game reviews
5. steam_insights.csv: List of all Steam game metadata
6. tags.csv: List of all the tags of the games.

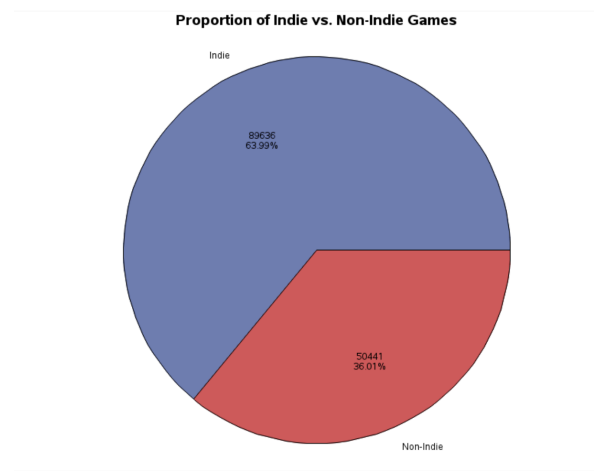
We begin by reading each CSV into Pandas DataFrames, using parameters to handle special characters, encoding, and any malformed rows. After successfully loading the data, we extract JSON-based price details from **games.csv**, ensuring we preserve the structure of fields. We also remove multi-line text fields (such as *languages* or user *reviews*) when they disrupt CSV parsing or cause formatting issues in SAS.

Finally, we load these cleaned datasets into SAS Studio. We run a series of *PROC SQL* steps to select only the columns we need, merge tables by *app_id* using an *INNER JOIN*, convert date fields to a proper SAS date format, and print a sample of the combined data. This approach – cleaning with Python and merging with SAS – takes advantage of Python's flexibility in data parsing and SAS's robust data manipulation and reporting capabilities.

DATA ANALYSIS

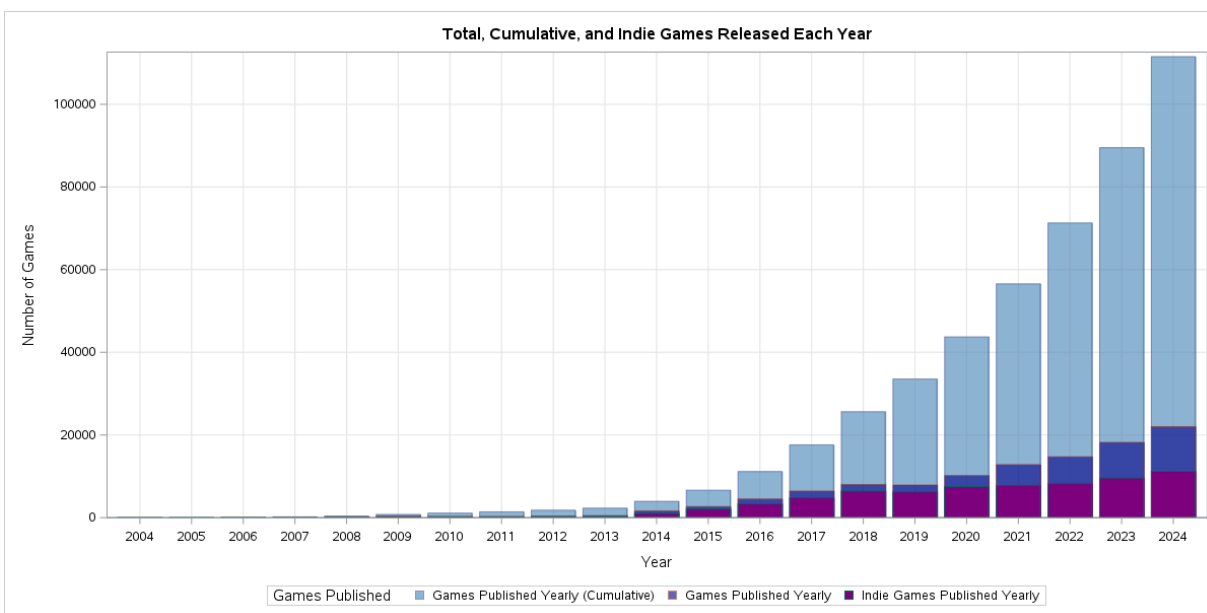
Once our data was prepared we were then able to analyze these databases. To do this, we utilized both SAS Studio and SAS visual analytics. Starting from the top, we were able to see that there are around 180 thousand games on the Steam platform, of which we were able to acquire 140 thousand. The size of Steam's platform supports the idea that gaming is known to be the biggest in the entertainment industry [3]. Furthermore, this number also showcases that the gaming industry at large is constantly growing with no signs of stopping. Within our data, 63.99% of these games are all classified as 'indie' shorthand for 'independant'. This indicates a large growth in small independent studios or solo game developers.

This demonstrates that indie games play a significant role in the gaming industry. While the absolute number of indie game releases remains high, we do see a slowing trend after 2020, when indie development appears to have peaked.

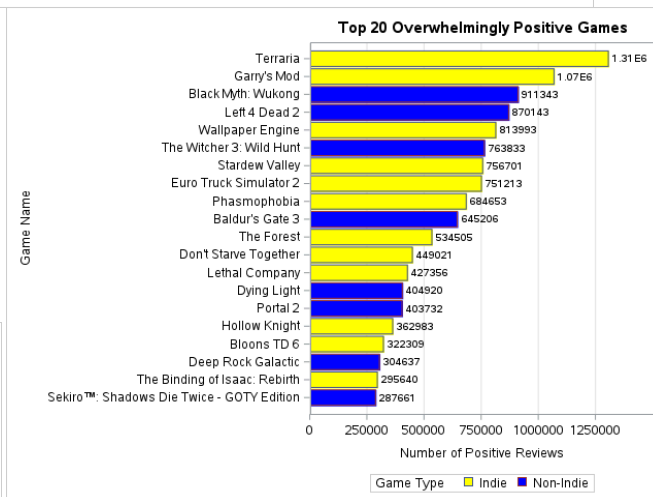
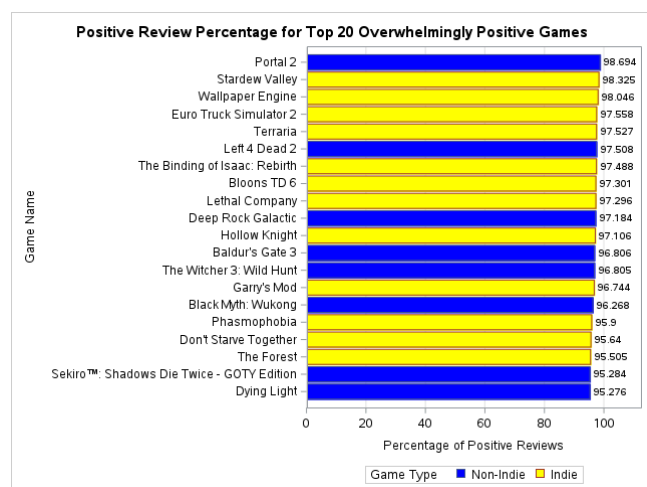


One factor may be the impact of COVID-19: during 2020, with more people staying at home, there was a surge of interest in gaming. Indie developers kept releasing titles, but larger, well-resourced studios also ramped up production to meet rising demand. Over time, that post-2020 shift seems to have resulted in a slightly slower growth rate for indie releases relative to non-indie titles.

One can speculate that technology becoming more accessible, and the increasing power and ease of use of game engines like Unity and Unreal Engine has benefited developers of all sizes.



This data becomes more interesting when you take a look at the customer reception of indie games. One would assume that indie games, having far fewer resources compared to those of a bigger studio, would lack the quality and in turn make worse games, but the data



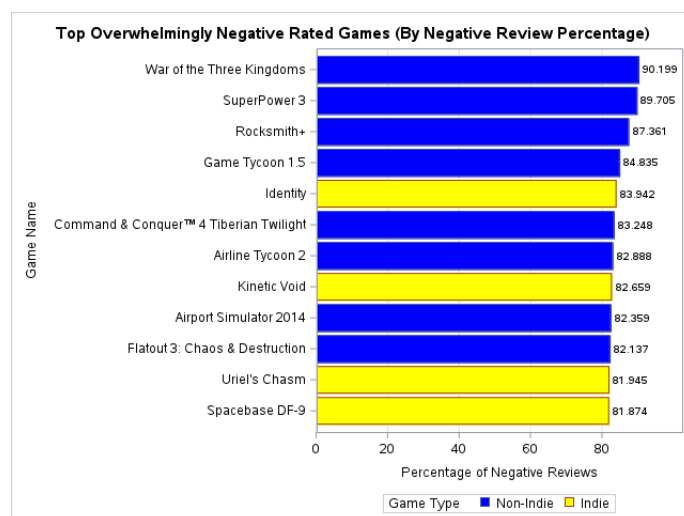
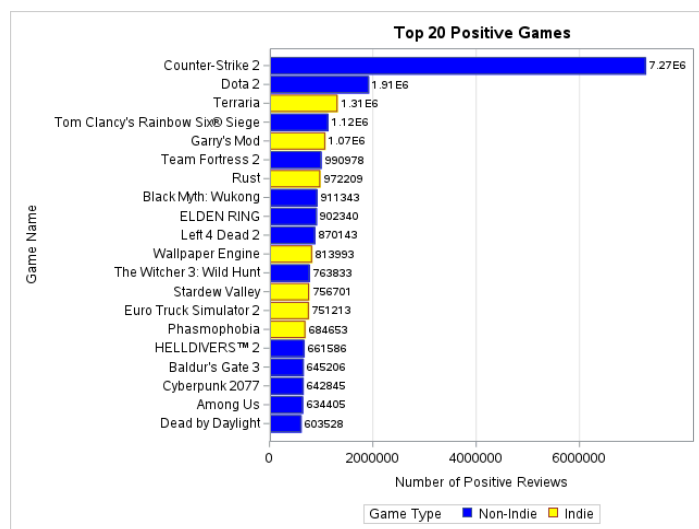
shows otherwise.

According to the reviews on Steam, 11 of the top 20 games that have received the overwhelmingly positive reviews (highest

positive review rating) are indie. This showcases the size and love for indie games by the masses. The interesting thing here is that the opposite is true when you look for the negative reviews. When looking at the lowest rating possible, only 4 of the 20 games are indie. Based on this data alone, we can extrapolate that if a company is to spend more money on game development, it won't automatically become a better game with higher positive reviews.

Typically, indie games also do not have marketing to the same degree of bigger studios, so it is impressive when an indie game that is usually marketed by word of mouth ends up securing a higher position than a major game. An interesting point that the data showed, however, is that when we go a tier below overwhelmingly positive, which is simply called positive. The story is very different.

When looking at the data here, we can see Counter-Strike 2, is ranked the highest, massively overshadowing every other game's positive review count by several magnitudes. In addition to this, several A-list games appear at the highest ranking, where just 7 of them are indie games. This indicates that A-list games and popular games in general get the highest amount of reviews, where a lot if not most of them are positive.



This high amount of review total could well be the cause for the missing titles in the overwhelmingly positive list, as with a higher base of negative comments it becomes increasingly difficult, if not impossible to beat the positive to negative comment rate as seen in the positive review percentage graph. For example, 'Portal 2' is the game with the highest rate of positive to negative reviews on the platform, but on the graph of total positive reviews it is the 15th entry.

CONCLUSION

On the premise of our data and the analysis, the significant role indie games play in the modern gaming industry can be highlighted making up a substantial portion of Steam's game library. Whilst indie titles often operate with fewer resources compared to A-list games, their success can be shown not in the numbers of user engagement, but rather the

quality of it. Our findings indicated that indie games consistently receive high user ratings, many of which rank amongst the best-reviewed titles on the platform.

However, mainstream games still dominate in terms of sheer review volume, suggesting that marketing budgets and brand recognition play a role in driving user engagement and/or attention. The differences in review patterns between indie and mainstream games emphasize that financial investment alone does not guarantee higher user satisfaction, but it will compel users to interact with and buy the product.

These insights can help indie-game developers better navigate the competitive landscape by focusing on quality, creativity, and community-driven marketing strategies. In addition, this shows that indie developers have less to fear with smaller numbers of user reviews, as the rate of positive to negative comments is ultimately more important in the grand scheme of Steam's marketplace that has so far been largely spearheaded by the surprising quality of low budget indie-games.

In the end, our study underscores the unique market dynamics of indie games and their continued influence on the gaming industry.

REFERENCES

[1] Rifolio. "Steam Insights." GitHub. February 21, 2025. Available at <https://github.com/rifolio/steam-insights>.

[2] NewbieIndieGameDev. "Steam Insights." GitHub. February 21, 2025. Available at <https://github.com/NewbieIndieGameDev/steam-insights>.

[3] Bain & Company. "Global Video Game Revenue to Reach \$257 Billion by 2028, Outpacing Combined Revenues of Other Media Types, Finds Bain & Company." Bain & Company. February 21, 2025. Available at <https://www.bain.com/about/media-center/press-releases/2024/global-video-game-revenue-to-reach-%24257-billion-by-2028-outpacing-combined-revenues-of-other-media-types-finds-bain--company/>

CONTACT INFORMATION

Vladyslav Horbatenko

vladyslav@ruc.dk

Jonas Clausen

jwsc@ruc.dk

Salar Komeyshi

salar@ruc.dk

APPENDIX

[A] Reading CSV Files into Pandas

```
df_categories = pd.read_csv('steam_data/categories.csv')
df_games = pd.read_csv('steam_data/games.csv', quotechar='"', escapechar="\\", encoding="utf-8",
on_bad_lines="skip")
df_genres = pd.read_csv('steam_data/genres.csv')
df_reviews = pd.read_csv('steam_data/reviews.csv', quotechar='"', escapechar="\\", encoding="utf-8",
on_bad_lines="skip")
df_tags = pd.read_csv('steam_data/tags.csv')
df_steamspy_insights = pd.read_csv('steam_data/steamspy_insights.csv', quotechar='"', escapechar="\\",
encoding="utf-8", on_bad_lines="skip")
```

[B] Extracting JSON Price Details from games.csv

```
def extract_price_info(price_str):
    try:
        data = json.loads(price_str)
    except (TypeError, json.JSONDecodeError):
        return pd.Series({
            'final': None,
            'initial': None,
            'currency': None,
            'final_formatted': None,
            'discount_percent': None,
            'initial_formatted': None,
        })
    return pd.Series({
        'final': data.get('final'),
        'initial': data.get('initial'),
        'currency': data.get('currency'),
        'final_formatted': data.get('final_formatted'),
        'discount_percent': data.get('discount_percent'),
        'initial_formatted': data.get('initial_formatted'),
    })

price_details = df_games['price_overview'].apply(extract_price_info)
df_games = pd.concat([df_games, price_details], axis=1)
```

[C] Selecting Columns in SAS

```
LIBNAME mydata "/home/u63174806/sasuser.v94/permanent_data";
```

```
PROC SQL;
    CREATE TABLE mydata.selected_games AS
    SELECT app_id, name, release_date, is_free, type
    FROM mydata.games;
QUIT;
```

```
PROC SQL;
    CREATE TABLE mydata.selected_steamspy AS
    SELECT app_id, developer, publisher, owners_range
    FROM mydata.steamspy_insights;
QUIT;
```

```
PROC SQL;
    CREATE TABLE mydata.selected_reviews AS
    SELECT app_id, review_score, review_score_description, positive, negative, total
    FROM mydata.reviews;
QUIT;
```

[D] Merging Datasets, Converting Dates, and Verifying Output in SAS

```
PROC SQL;
  CREATE TABLE mydata.combined AS
  SELECT g.app_id, g.name, g.release_date, g.is_free, g.type,
         s.developer, s.publisher, s.owners_range,
         r.review_score, r.review_score_description,
         INPUT(r.positive, COMMA12.) AS positive,
         INPUT(r.negative, COMMA12.) AS negative,
         INPUT(r.total, COMMA12.) AS total
  FROM mydata.selected_games AS g
  INNER JOIN mydata.selected_steamspy AS s ON g.app_id = s.app_id
  INNER JOIN mydata.selected_reviews AS r ON g.app_id = r.app_id;
QUIT;

/* Step 3: Convert `release_date` to SAS date format */
DATA mydata.combined;
  SET mydata.combined;

  IF VTYPE(release_date) = 'C' THEN
    release_date_fixed = INPUT(release_date, YYMMDD10.);
  ELSE
    release_date_fixed = release_date;

  FORMAT release_date_fixed DATE9.;
  DROP release_date;
  RENAME release_date_fixed = release_date;
RUN;

/* Step 4: Print a sample of the combined data */
PROC PRINT DATA=mydata.combined (OBS=10);
  TITLE "Sample of Combined Data";
RUN;
```