

From Niche to Mainstream: Evaluating the Cross-Platform Success of High-quality Steam Games

Wen Chen

2023-12-13

Contents

Introduction	1
Data	2
Results:	4
Discussion	9
References	9

Introduction

The gaming industry is a dynamic and rapidly evolving sector, which has consistently captured a diverse audience worldwide. Understanding the trajectory of a video game's success is crucial for developers, marketers, and gaming enthusiasts alike. This research paper, titled "From Niche to Mainstream: Evaluating the Cross-Platform Success of High-quality Steam Games," embarks on an exploratory journey to decipher the key elements that propel high-quality games from being niche favorites to achieving mainstream success. Through a meticulous analysis of the top 150 high-rating games released on Steam in 2021, this study sheds light on the pivotal role of sales volume, pricing strategies, and Google search interest levels in determining a game's popularity and success across various platforms.

The digital gaming market, characterized by fierce competition and rapidly changing consumer preferences, presents a unique set of challenges and opportunities. In this context, the research aims to comprehensively understand the market dynamics that influence a game's transition from a niche favorite to a widely recognized and financially successful product. By employing advanced data analytics techniques, including K-means clustering and regression analysis, the study seeks to offer valuable insights into the interplay between financial performance and public interest, as measured by Google search trends.

This paper also delves into the nuanced relationship between game pricing and sales volume. It investigates how pricing strategies impact a game's accessibility and appeal, and consequently, its market performance (Brodie, 2016). Furthermore, by analyzing Google search interest levels - both average and peak - the research provides an in-depth understanding of public curiosity and engagement with these games. The study explores how these interest levels correlate with actual sales, revealing intriguing patterns and trends that defy conventional wisdom.

Data

Data from Website: Best 150 games of 2021-2022

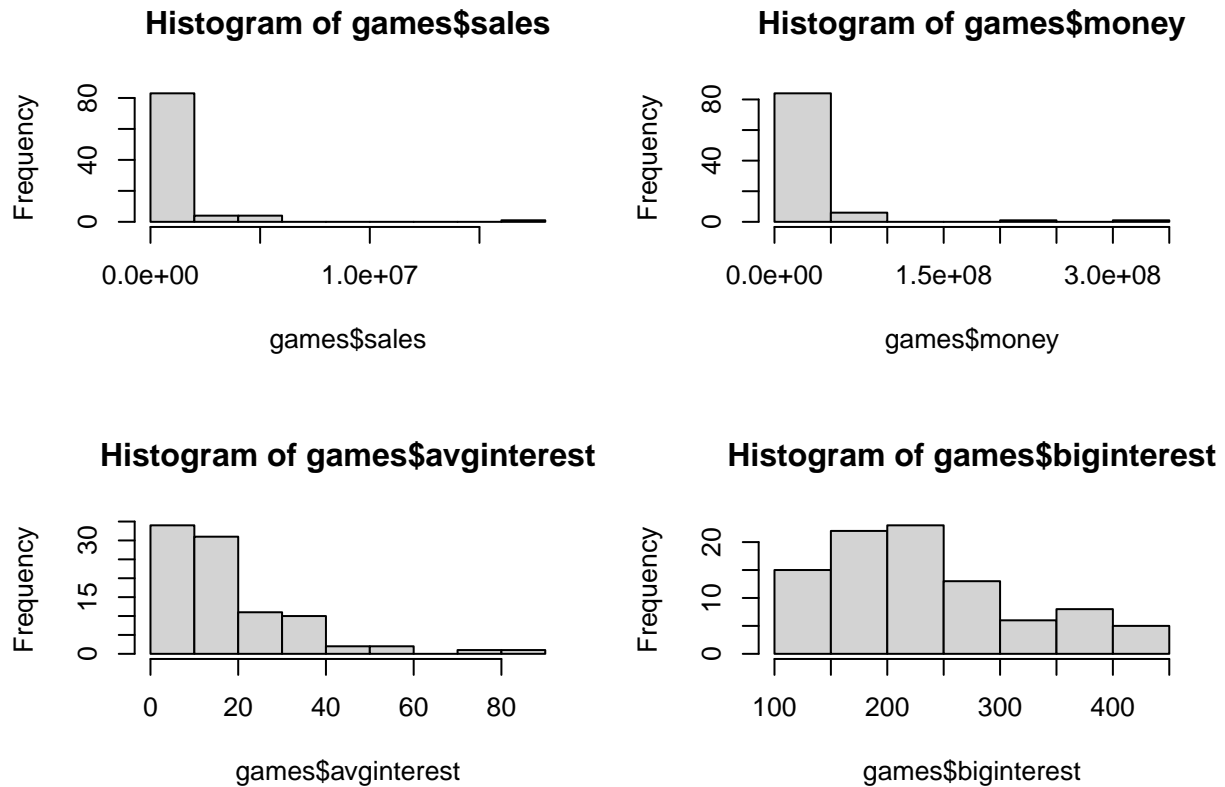
The study primarily utilized data from Steam250, a reputable website listing the top 150 Steam games for the year 2021. From this source, I extracted crucial variables for each game: ‘rank’ (the game’s position in the rating list), ‘name’ (title of the game), ‘score’ (rating received on Steam), ‘price’ (retail price in USD), and ‘votes’ (number of player reviews on Steam). This data provided a foundational understanding of each game’s market performance and popularity.

To translate these variables into more tangible business metrics, I employed a method outlined on VG Insights for estimating sales volume and total revenue (2021). The ‘sales’ variable was inferred by multiplying the ‘votes’ (player reviews) by a factor of 40, a method supported by industry insights and analysis. This approach provided a reasonable estimate of total sales volume for each game. Further, I calculated the ‘money’ variable, representing total sales revenue, by multiplying the game’s ‘price’ by its estimated sales volume. This transformation process facilitated a comprehensive dataset, enabling us to analyze the intricate relationship between market performance, pricing strategies, and public interest in these top-rated games.

Google Trends Data Scraping

To complement the Steam data, I scraped Google Trends for public search interest in each game. A custom R function, `gtrends_with_backoff`, was developed to reliably gather data despite potential server errors or request limits (2021). I created functions to fetch and calculate average and peak monthly search interests for each game between January 1, 2021, and December 1, 2023. The ‘avginterest’ variable represents the average monthly search interest, while ‘biginterest’ denotes the peak monthly search interest. Games with ambiguous or misleading names were excluded to ensure data accuracy, such as “Firework”. This meticulous process resulted in a refined dataset of 92 games, providing a comprehensive view of public engagement and interest trends in relation to the Steam games under study.

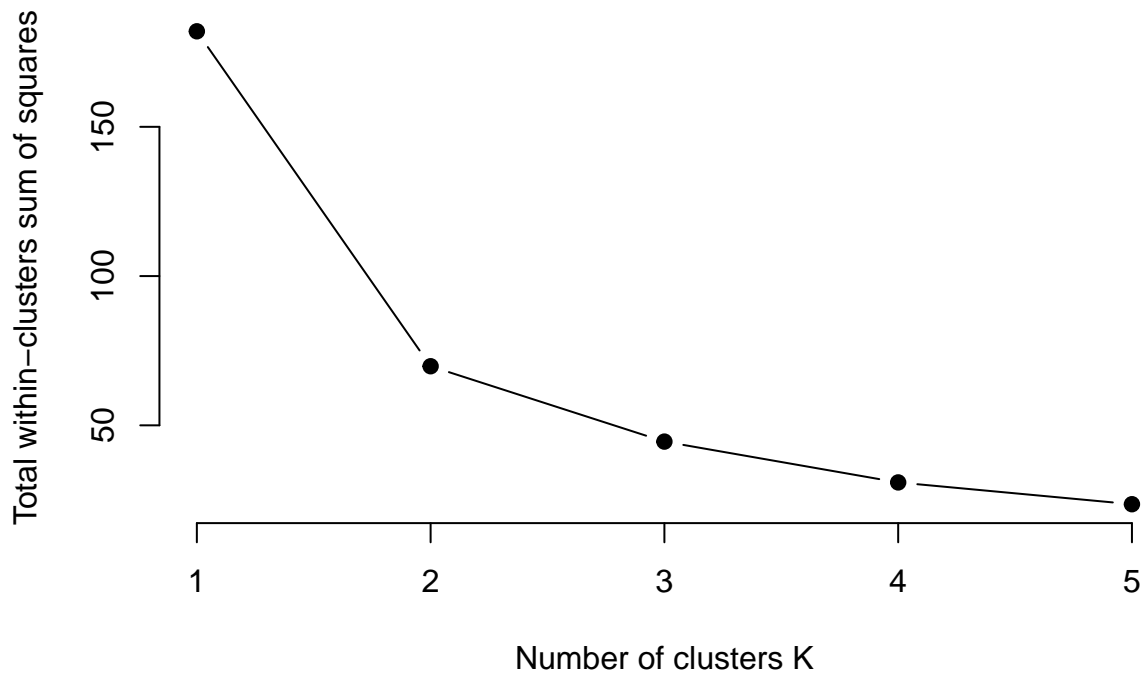
Data Summary:



The histograms for the four key continuous variables—sales volume, sales revenue, average monthly search interest, and peak monthly search interest—reveal distinct distributions among high-ranking Steam games. The heavily right-skewed distribution of total sales volume, total sales revenue, and average monthly search interest shows: although although these are all high-ranking high-quality games, a few achieve significantly higher sales volumes, financial success and monthly average interest level on the Steam platform. The distribution of the peak monthly search interest has a more even spread across its range but still exhibits a slight right skew.

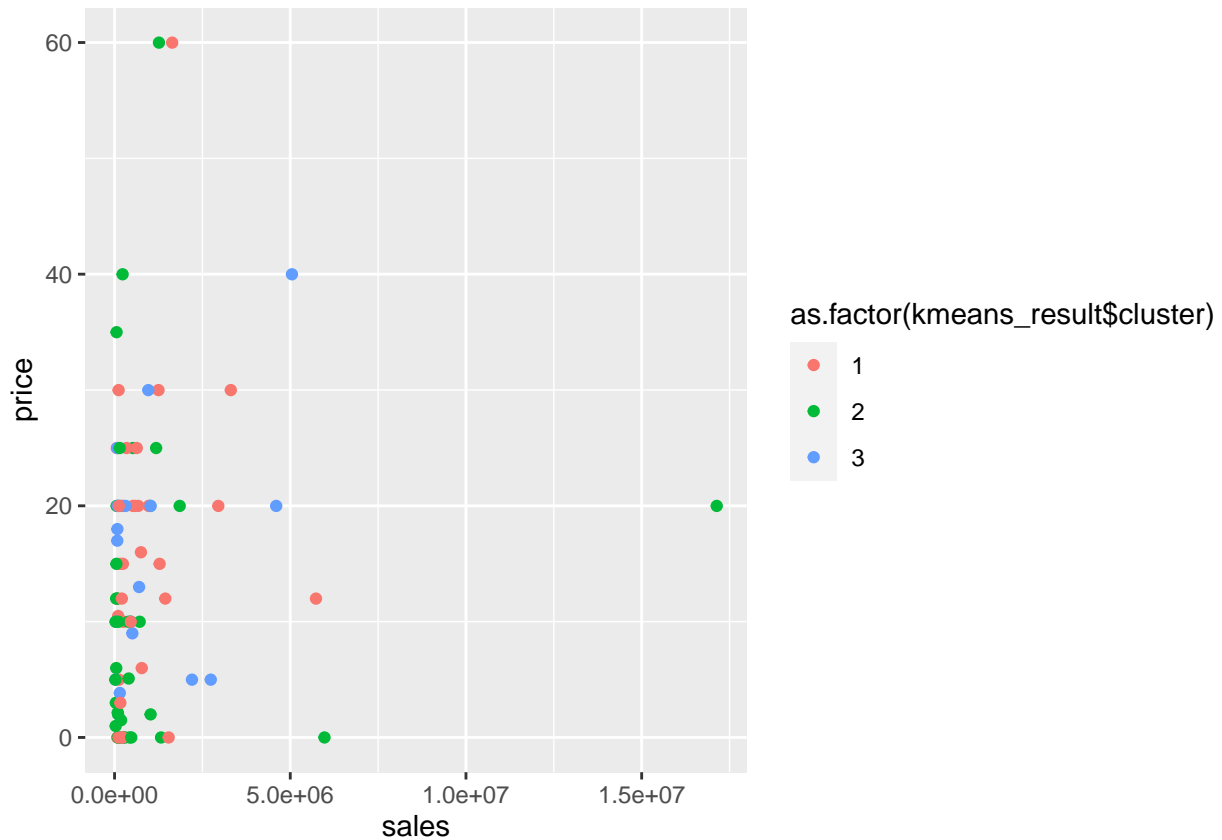
Results:

Integration of Search Interest and Sales Data through K-means Clustering



```
## K-means clustering with 3 clusters of sizes 35, 38, 19
##
## Cluster means:
##   avginterest biginterest
## 1  -0.2134067   0.2310353
## 2  -0.5976333  -0.9049789
## 3   1.5883842   1.3843665
##
## Clustering vector:
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
##  3  1  1  3  2  1  2  1  2  3  3  2  3  2  1  2  1  2  1  1  1  2  2  3  2  2
## 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
##  2  1  1  1  2  1  1  2  1  1  1  2  1  1  2  1  2  1  2  1  2  2  1  2  1  2
## 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78
##  3  1  2  1  2  2  1  3  2  2  3  3  2  2  2  1  2  3  1  3  3  2  1  2  3  1
## 79 80 81 82 83 84 85 86 87 88 89 90 91 92
##  2  2  2  1  1  3  3  3  2  3  3  1  2  1
##
## Within cluster sum of squares by cluster:
## [1] 15.532834  6.431799 27.530122
## (between_SS / total_SS =  72.8 %)
```

```
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"       "
```



To discern the relationship between Google search interest and sales data, the study first undertook the standardization of monthly average interest levels and peak monthly search interest. This normalization allowed for an equitable comparison, mitigating the scale differences between the two variables. Upon integrating these standardized metrics, I applied the K-means clustering algorithm to segment the dataset into distinct groups. Utilizing the Elbow Method, I determined that the optimal number of clusters for the dataset was three.

The analysis then shifted focus to the characteristics of these clusters. Cluster 3 emerged as particularly noteworthy, encompassing games that not only sustained high average interest levels but also reached substantial peaks in search interest. This cluster represents titles that consistently captured public attention and experienced marked surges in popularity, presumably coinciding with significant events such as game releases, updates, or community-driven activities.

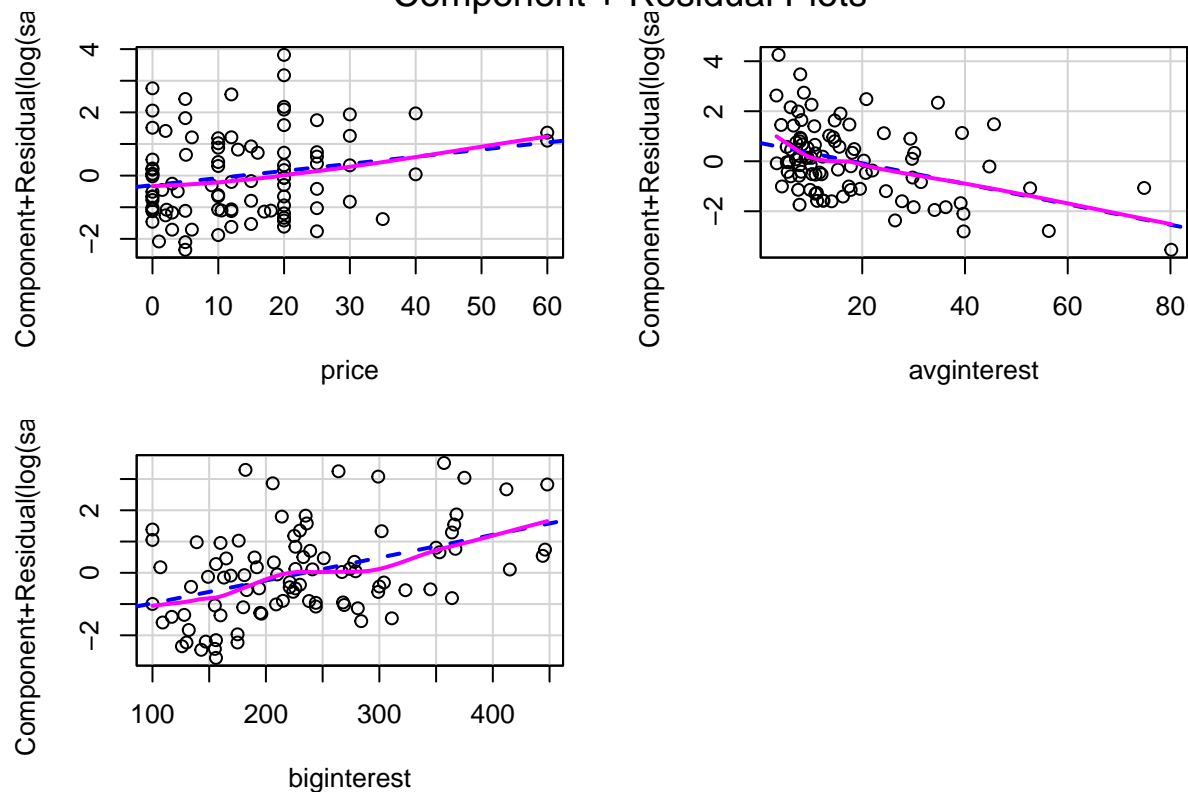
Within this distinguished group, I observed a diverse range of sales and price points. However, a common thread among these games was their tendency to command higher prices, correlating with their robust sales volumes. This phenomenon suggests a market dynamic where elevated public interest and pricing power converge, potentially signaling a premium value placed on these games by consumers.

These findings illustrate a segment of the gaming market where high visibility and consumer willingness to pay intersect, leading to remarkable commercial success. The insights drawn from Cluster 3 provide a compelling narrative on the attributes of hit games that manage to break through the noise and capture both the interest and the wallets of gamers.

Regression Analysis

```
##
## Call:
## lm(formula = log(sales) ~ price + avginterest + biginterest,
##     data = games)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1544 -1.0079 -0.2074  0.8358  3.6707
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.358220   0.445516  25.495  < 2e-16 ***
## price        0.022538   0.011587   1.945  0.05495 .
## avginterest -0.040929   0.014975  -2.733  0.00758 **
## biginterest  0.007302   0.002598   2.811  0.00609 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.325 on 88 degrees of freedom
## Multiple R-squared:  0.1529, Adjusted R-squared:  0.124
## F-statistic: 5.295 on 3 and 88 DF,  p-value: 0.002107
```

Component + Residual Plots

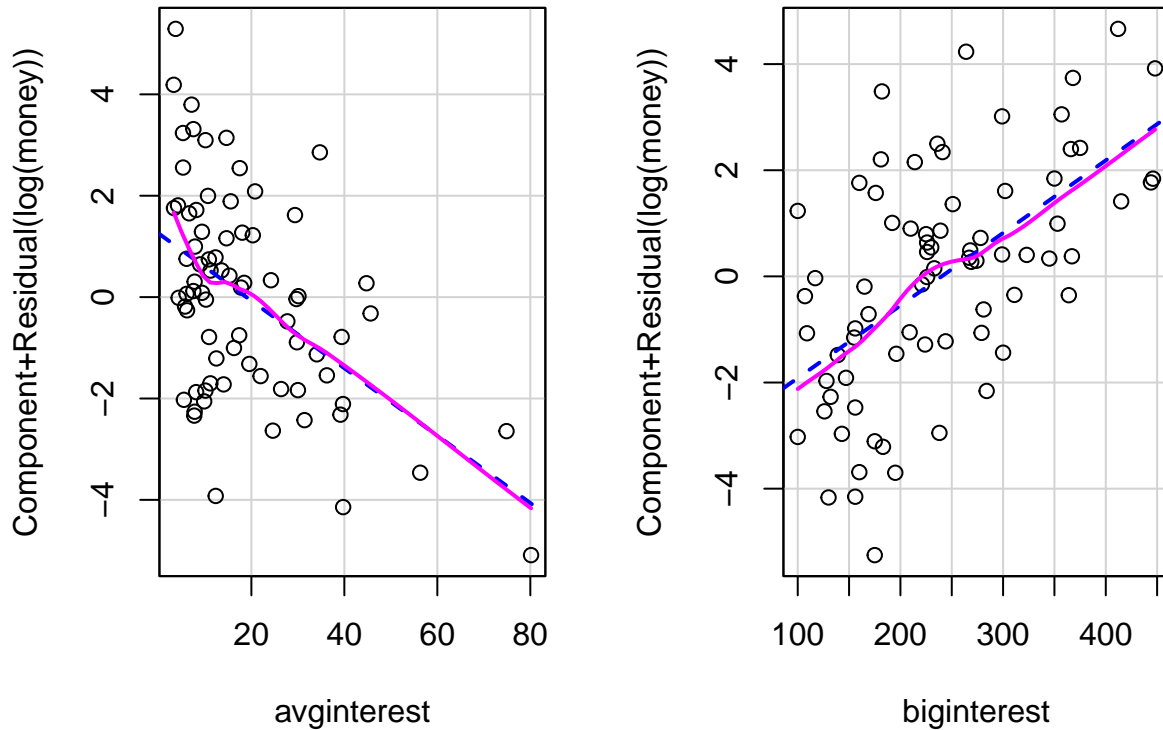


First regression analysis focused on understanding the relationship between the sales volume of video games and several key predictors: the price of the games, their average monthly search interest (avginterest),

and their peak monthly search interest (biginterest). The model, “mod_log <- lm(log(sales) ~ price + avginterest + biginterest, data = games)”, provided insightful results into how these factors correlate with game sales. Price Influence: The coefficient for price was found to be positive (Estimate: 0.022538, p = 0.05495), suggesting a slight tendency for higher-priced games to have increased sales. This might indicate that customers associate price with quality or value in the gaming market. Interestingly, the avginterest variable showed a negative relationship with sales (Estimate: -0.040929, p = 0.00758). This indicates that games with higher average search interest might not necessarily translate into higher sales volumes. This finding could reflect a phenomenon where certain games, despite piquing public curiosity or interest, fail to convert this interest into actual purchases. This aspect could be due to various factors like the nature of the interest (e.g., controversy or short-term trends) not leading to buying decisions. Peak Monthly Search Interest: On the contrary, biginterest displayed a positive relationship with sales (Estimate: 0.007302, p = 0.00609), suggesting that games reaching peak search interest levels are more likely to see higher sales volumes. This could imply that surges in search interest, potentially reflecting marketing campaigns, product launches, or significant updates, are effective in driving sales.

```
##
## Call:
## lm(formula = log(money) ~ avginterest + biginterest, data = games_omit0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3614 -1.1033  0.0112  1.0400  4.2795
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.13122    0.65048  20.187 < 2e-16 ***
## avginterest  -0.06652    0.02115  -3.145 0.002413 **
## biginterest   0.01367    0.00363   3.766 0.000336 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.794 on 72 degrees of freedom
## Multiple R-squared:  0.1659, Adjusted R-squared:  0.1427
## F-statistic:  7.16 on 2 and 72 DF,  p-value: 0.001459
```

Component + Residual Plots



In my second model, I aimed to delve deeper into the dynamics of sales revenue in the video game industry, particularly in relation to online search interest. To refine the analysis, I focused exclusively on games generating revenue by excluding free games from the dataset. This allowed us to better assess the impact of consumer interest on actual revenue generation.

Using the modified dataset, I constructed a linear model (`mod2_log`) to investigate the relationship between the logarithm of sales revenue (`log(money)`) and two key predictors: average monthly search interest (`avginterest`) and peak monthly search interest (`biginterest`).

The model revealed a significant negative relationship between `avginterest` and sales revenue (Estimate: -0.06652, $p = 0.002413$). This suggests that games with consistently high search interest do not necessarily translate into higher revenue. This finding may indicate that a sustained general interest in a game does not always lead to purchasing behavior. It could reflect interest driven by factors other than purchase intent, such as media coverage, social media trends, or curiosity about the game without the intention to buy.

Contrarily, `biginterest` showed a significant positive correlation with sales revenue (Estimate: 0.01367, $p = 0.000336$). This implies that spikes in search interest, possibly tied to specific marketing campaigns, major updates, or significant events, are more effective in driving sales. These peaks may represent moments of heightened consumer engagement and intent to purchase.

The results underscore the complex dynamics between online interest in video games and actual sales. While high levels of sustained interest do not necessarily equate to higher sales, peaks in interest, possibly reflecting targeted events or promotions, appear to have a more direct impact on sales. This insight could be pivotal for marketing strategies, highlighting the importance of creating moments that capture peak audience attention.

Further, the positive but relatively weak correlation between price and sales offers an interesting perspective for pricing strategies in the gaming industry. It suggests a cautious approach where pricing decisions should consider the perceived value and market positioning of the product.

Discussion

This research has provided valuable insights into the dynamics of financial success and online visibility in the realm of high-quality Steam games. By analyzing the interplay between sales revenue, price points, and Google search trends, I have uncovered nuanced relationships that govern the cross-platform success of these games. Key findings include the notable influence of peak search interests on sales, as well as the less straightforward relationship between sustained average search interest and financial performance.

These insights are particularly valuable for strategizing in the video game market. While average interest levels provide a measure of general awareness and popularity, they may not be reliable indicators of revenue potential. Instead, peaks in search interest seem to be more indicative of periods where potential buyers are actively considering a purchase. This distinction can guide marketing and sales strategies. Focusing on creating and capitalizing on moments that generate peak interest may be more effective than maintaining a consistent level of visibility. This approach could involve targeted promotional activities, strategic release timing, and leveraging events or updates to create buzz and drive sales.

However, it is crucial to acknowledge the limitations of the study. Firstly, focus on Google search trends as a primary indicator of online interest might not capture the full spectrum of digital engagement. Platforms like Reddit, known for their vibrant gaming communities, can offer additional insights into the popularity and discussion surrounding games, potentially revealing trends and sentiments not fully captured by search data. Secondly, the analysis was conducted within the ecosystem of Steam, a major but not exclusive platform for game distribution. The gaming industry is diverse and multi-faceted, with various other platforms contributing to a game's overall success. Therefore, the findings might not be entirely generalizable across different distribution channels. Additionally, the influence of game characteristics, such as genre, game-play mechanics, and narrative elements, was not explicitly factored into the analysis. These elements can significantly affect a game's appeal and success and warrant consideration in a more comprehensive study.

Looking forward, several avenues for further research present themselves. An expanded analysis incorporating data from other platforms, such as console-specific stores or other PC game marketplaces, could provide a more holistic view of a game's success. Similarly, exploring the impact of social media platforms and gaming forums could yield a richer understanding of consumer engagement and its correlation with sales. Investigating the role of game characteristics, including genre, design, and player engagement strategies, could also offer deeper insights. Such an analysis could unravel how different types of games resonate with varying audience segments and how these preferences translate into sales and online trends.

References

- Brodie, S. (2016, January 21). What your game's Price Really Means. GamesIndustry.biz. <https://www.gamesindustry.biz/what-your-games-price-really-means>
- Karoliskoncevicus. (2022, July 26). GitHub gtrends_with_backoff.R. https://github.com/trendecon/trendecon/blob/master/R/gtrends_with_backoff.R
- Top 150 best steam games released in 2021. Steam 250. (n.d.). <https://steam250.com/2021>
- VG Insights. (2021, August 1). Games industry data and analysis. How to Estimate Steam Video Game Sales? <https://vginsights.com/insights/article/how-to-estimate-steam-video-game-sales>