

# *The Popularity Puzzle: A Deep Dive into Video Game Success Factors*

Ouwen Li<sup>1,a,\*,†</sup>, Ruishen Guo<sup>2,b,†</sup>

<sup>1</sup>International Center, Jinan Foreign Language School, Jinan, 250108, China

<sup>2</sup>Shenzhen Foreign Language School, Shenzhen, 518028, China

a. Owen121523@gmail.com, b. g2383414784@gmail.com

\*corresponding author

†These authors contributed equally.

**Abstract:** In the digital age characterized by technological advancements, video games have risen to prominence as a leading form of global entertainment. Beyond their primary role as a source of amusement, they have seamlessly integrated into various aspects of our lives, from social interactions and education to mental well-being. However, despite their widespread influence, there exists a noticeable void in comprehensive research that delves into the determinants of a game's success and the evolving trends that shape the industry. This study embarked on a journey to bridge this gap. Utilizing regression analysis, we meticulously examined key parameters such as publishers, ratings, release dates, and game genres. Our systematic exploration of extensive video game datasets revealed intriguing patterns: certain game types, styles, and themes consistently resonate with audiences; there's a tangible correlation between game ratings and their commercial success; and historical data offers a glimpse into the potential future trajectories of popular games. Our findings indicate that specific game styles, such as adventure and indie genres, have a heightened popularity among users. A discernible relationship was observed between game ratings and indicators of success, emphasizing the pivotal role of user feedback in shaping a game's market performance. Additionally, by analyzing past trends, we've identified potential directions for the industry's evolution, hinting at a surge in narrative-driven and innovative gameplay mechanics in the coming years. These insights not only provide a comprehensive understanding of the current gaming landscape but also offer a roadmap for future developments, making this study invaluable for developers, analysts, and policymakers.

**Keywords:** Video games, One-hot encoding, Top-labels method, Marketing strategies

## 1. Introduction

In the past few decades, the technological landscape has undergone a significant transformation. This rapid evolution has not only reshaped our daily lives but has also revolutionized the entertainment industry. Among the myriad of entertainment options available today, video games stand out as a dominant force. With the proliferation of digital devices and platforms, video games have transitioned from being mere recreational activities to a global phenomenon, captivating audiences of all ages across the globe [1].

The significance of video games extends far beyond their entertainment value. They have become an integral component of the broader entertainment ecosystem, intertwining with various facets of our lives. Video games, in today's context, are not just about gameplay; they serve as platforms for social interaction, tools for education, therapeutic interventions for mental health, and even as mediums for storytelling and artistic expression.

However, the meteoric rise of video games in popular culture presents an intriguing paradox. While their influence is undeniable and their reach expansive, there exists a noticeable gap in comprehensive research exploring the underlying factors that contribute to a game's success and its broader implications on the industry. The intricacies of what makes a game resonate with its audience, the dynamics between game ratings and their commercial success, and the evolving trends that shape the industry's future are areas that remain relatively uncharted.

This research endeavors to bridge this gap. By employing a meticulous regression analysis, we aim to dissect the developmental trends and characteristics that define popular video games. Our focus encompasses key parameters like publishers, ratings, release dates, and game genres. Through an in-depth examination of extensive video game datasets, our objective is to shed light on several pivotal aspects: 1) Common Traits of Popular Games: Delving into the nuances of game design and content, we seek to identify the elements, be it game type, style, theme, or other factors, that resonate with audiences and contribute to a game's widespread appeal [2]. 2) Association between Ratings and Success: In an industry where success is often quantified by sales figures and popularity metrics, understanding the correlation between game ratings and these indicators is crucial. This segment explores the dynamics between critical acclaim, user reviews, and commercial success. 3) Prediction of Developmental Trends: By harnessing historical data, we aim to forecast the potential trajectories that popular games might take in the coming years, providing a roadmap for the industry's future evolution.

In addition to these core areas, our research has broader implications. For stakeholders like game developers, market analysts, and policymakers, the insights gleaned from this study can serve as a valuable compass, guiding strategies and decisions to nurture the video gaming industry's sustainable growth [3]. Furthermore, by leveraging cutting-edge data analysis techniques, we not only aim to demystify the current landscape of the video gaming world but also to draw parallels and offer insights into related domains such as digital media, user behavior patterns, and evolving marketing strategies [4].

In essence, this research is a deep dive into the world of video games, aiming to unravel the complexities that shape its present and forecast the innovations that will define its future.

## **2. Data Pre-Processing**

### **2.1. Data Cleaning**

In our pursuit of refining our dataset for academic analysis, we meticulously conducted a comprehensive data cleansing process to ensure the dataset's accuracy and reliability, which are critical for the integrity of our academic study. This process involved several key phases:

**Data Reduction and Parsing:** We began by removing extraneous summary information from the original dataset to focus our analysis. Additionally, we addressed missing values in the rating column ('NA') with careful consideration to prevent unintended biases.

**Enhanced Field Segmentation:** To handle the challenge of certain fields, such as developers and categories, being separated by commas, we employed a structured approach in R. Commas were replaced with semicolons to enable accurate segmentation.

**Managing Symbol Complexities:** Symbol-related complexities, including non-English characters and quotation marks within titles and developer names, were meticulously addressed. We recognized

and utilized ASCII codes to systematically remove non-standard characters, ensuring data consistency.

**Numerical Consistency and Formatting:** Numerical values underwent standardized formatting. Entries with 'K' suffixes were adjusted, with 'K' removed and non-'K' values divided by 1000. This standardized the scale of all numerical data for equitable analysis.

**Transformation of Textual Data:** A critical transformation step involved converting textual data into numerical format, facilitating quantitative analysis and correlation studies.

In conclusion, our multi-faceted and comprehensive data cleansing process ensures the dependability and suitability of our dataset for scholarly investigation in our academic study.

## 2.2. One-hot Encoding

One-hot encoding is a method to represent categorical data numerically [5]. It converts categories into binary vectors, where each category corresponds to a unique column with '1' or '0' values indicating the presence or absence of the category. This technique is valuable in machine learning. For instance, in a movie dataset where genres are encoded using one-hot encoding:

### **Original Data:**

Movie A: Action, Adventure

Movie B: Romance, Drama

### **One-Hot Encoded:**

Movie A: Action[1], Adventure[1], Romance[0], Drama[0]

Movie B: Action[0], Adventure[0], Romance[1], Drama[1]

This transformation allows for efficient analysis, such as personalized recommendations based on genres a user prefers, studying the impact of genres on ratings, and grouping similar movies using clustering algorithms.

The one-hot encoding provides us with a more comprehensive numerical representation of data, enabling the feasibility of various machine-learning methods and analyses. Here are three potential innovative analytical points utilizing such encoded data:

### **Prediction Models based on Developers, Platforms, and Genres:**

Through one-hot encoding, we can employ developers, platforms, and genres as features to train machine learning models, predicting aspects like game ratings, user counts, etc. For instance, we can explore which specific developers, platforms, or genres correlate most with high-rated games. This aids in understanding the factors that might influence a game's success in the market.

### **Association Rule Mining:**

Post one-hot encoding, association rule mining techniques can be employed to uncover patterns within the data. For example, we could discover certain platforms and genres that frequently co-occur, or that specific developers tend to produce popular games of particular styles. This provides insights into market trends and consumer preferences.

## 3. Data Analysis

### 3.1. Association Rules Mining

By leveraging one-hot encoded features, we conducted a cluster analysis to categorize games based on developers, platforms, and genres. This method aids in identifying submarkets within the gaming industry and potentially reveals unexplored market niches. Such analytical insights are invaluable for developers, marketers, and strategic decision-makers, offering a clearer view of the gaming landscape.

From our mining of association rules, frequent style combinations with support exceeding 5% are:

Adventure and Indie: 14.64%  
Adventure and RPG: 10.20%  
Adventure and Puzzle: 7.72%  
Adventure and Platform: 7.64%  
Indie and Puzzle: 5.82%

Furthermore, similar methodologies can discern associations between platforms, genres, and developers. Our exploration of gaming platforms and genres revealed patterns indicating certain genres' resonance with specific platforms. For instance, adventure games' popularity on Windows PC could stem from gameplay mechanics, user interface, or the platform's immersive capabilities. Indie games, often innovative and experimental, flourish on Windows PC due to the platform's accessibility for independent developers, fostering a conducive environment for innovative and experimental game designs [6]. Moreover, platform-genre popularity might also mirror socio-cultural trends.

When examining the relationship between developers and genres, we find patterns that speak to developers' alignment with specific genres, reflecting their creative and technical strengths and responding to market demands.

The results from frequent platform and genre combinations exceeding 5% support include:

Adventure and Windows PC: 26.54%  
Indie and Windows PC: 24.34%  
Mac and Windows PC: 17.74%  
PlayStation 4 and Windows PC: 14.84%  
Adventure and Indie: 14.64%

A notable trend from our results is the significant alignment of various genres with the Windows PC platform. The platform's architecture, user base, and hardware diversity enable a myriad of gaming experiences. Its legacy as an innovation hub fosters both mainstream and niche genres. Windows PC's capabilities align naturally with genres requiring deep narratives and user engagement, making genres like adventure and indie stand out. This suitability positions these games, known for their narrative depth and innovative gameplay, as natural contenders for high-support combinations [7].

While Windows PC enjoys a broad spectrum of genres, platforms like PlayStation 4 and Mac have unique attributes that cater to specific genres that resonate with their user base and functionalities.

In summary, Windows PC's versatility in supporting diverse gaming experiences sets it apart in the industry, emphasizing its role as a crucial player in the gaming ecosystem. This signifies not only the platform's popularity but also its capacity to support a rich tapestry of gaming experiences that cater to a broad spectrum of players [8].

## 3.2. Clustering

### 3.2.1. Introduction of K-Means

Given a dataset  $x = \{x_1, x_2, \dots, x_m\}$ , where each  $x_i$  is an n-dimensional data point. Let  $C = \{C_1, C_2, \dots, C_K\}$  be the K cluster centers. The goal of K-means is to minimize the sum of the distances between every data point and its assigned cluster center. Mathematically, this can be expressed as:

$$J(C, X) = \sum_{i=1}^m \min_{j=1, \dots, K} \|x_i - c_j\|^2$$

where  $\| \ \|$  represents the Euclidean distance and C is the assigned cluster center to be updated. In each iteration, K-means performs two main steps. For each data point  $x_i$ , find the nearest cluster center  $c_j$  and assign  $x_i$  to that center. Then for each cluster center  $c_j$ , recompute it as the mean of all data points assigned to it.

### 3.2.2. Clustering Analysis

We conducted K-means clustering on the data using different quantities. Data clustering helps to divide a large dataset into meaningful groups, making it easier to understand the structure and characteristics of the data. Through clustering, we can identify patterns, trends, and correlations within the data, which aids in revealing the inherent properties of the data.

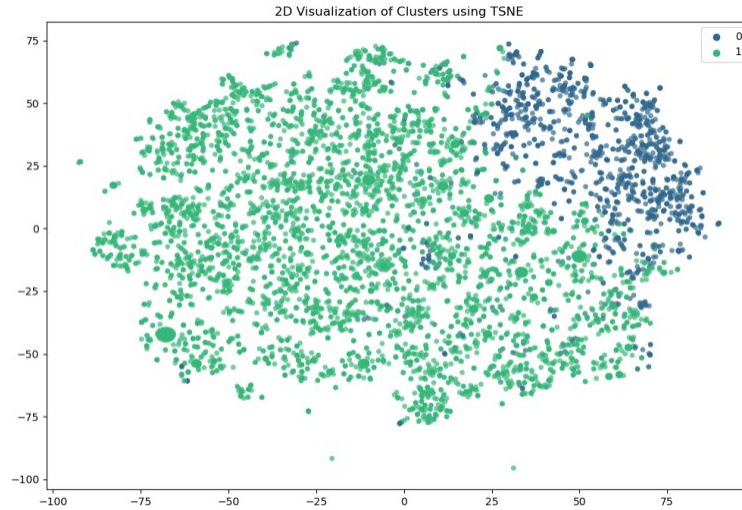


Figure 1: 2D Visualization of 2 Clusters using TSNE

Table 1: Statistics of 2-dimension Clustering

		Genres_0	Developers_0	Platforms_0
Cluster				
0	Top_1	Adventure	Square Enix	Windows PC
	Top_2	Indie	Capcom	PlayStation 4
	Top_3	RPG	Bandai Namco Entertainment	Xbox One
1	Top_1	Adventure		Windows PC
	Top_2	RPG	Nintendo	Mac
	Top_3	Indie	Sega	PlayStation 3

As indicated by the binary clustering graph, it is evident that the predominant determinant for the formation of the two clusters is the age of the gaming platform (see Figure 1). Table 1 shows the main features of each cluster in the binary clustering graph, namely the terms with the highest frequency. In the category associated with higher ratings (Category 0), the principal gaming platforms include PS4 and Xbox, which rank closely behind PC platforms in prominence during this particular era. In the context of the relatively average rated category (Category 1), the primary games featured are end games. However, following the advent of the PS4, console-based gaming swiftly gained a substantial foothold in the market [9].



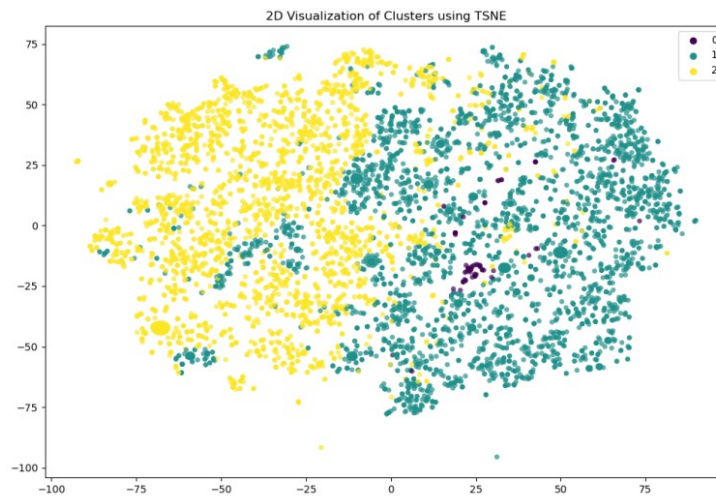


Figure 2: 2D Visualization of 3 Clusters using TSNE

Table 2: Statistics of 3-dimension Clustering

		Genres_0	Developers_0	Platforms_0
Cluster				
0	Top_1	Adventure	Crystal Dynamics	SteamVR
	Top_2	Indie	Devolver Digital	Oculus Quest 2
	Top_3	Puzzle	Oculus Quest	
1	Top_1	Adventure		Windows PC
	Top_2	Indie	Electronic Arts	Mac
	Top_3	RPG	Square Enix	PlayStation 4
2	Top_1	Adventure		PlayStation 3
	Top_2	RPG	Nintendo	PlayStation 2
	Top_3	Platform	Capcom	Nintendo Switch

Figure 2 shows the result of the 3-dimension clustering, which shows clear characteristics. As shown in Table 2, the majority of games categorized as 0 in the provided table primarily fall within the realm of sensory gaming experiences. Sensory games have emerged as a prominent trend in the contemporary gaming industry, and there is a noteworthy anticipation that they may exert a significant influence on the market in the foreseeable future.

The initial category of games listed in the table predominantly aligns with gaming platforms such as PC, Mac, and PS4, all of which have gained substantial popularity in recent years. In contrast, the mainstream platforms within the second category comprise PS3, PS2, and Switch, representing earlier phases of popularity in the gaming landscape.

Upon conducting a comparative analysis, it becomes evident that SQUARE ENIX holds a dominant position in the primary market segment for higher-rated games within the binary clustering. In the third category, SQUARE ENIX is notably positioned within the first tier of companies, highlighting its consistent presence and prominence across different game categories.

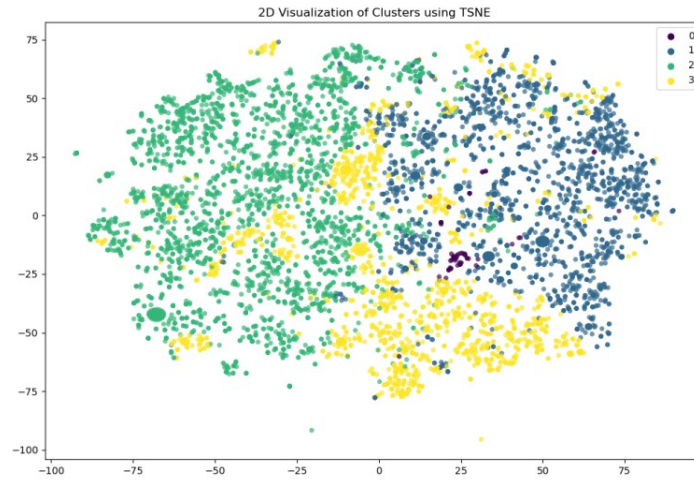


Figure 3: 2D Visualization of 4 Clusters using TSNE

Table 3: Statistics of 4-dimension Clustering

Cluster		Genres_0	Developers_0	Platforms_0
0	Top_1	Indie	Devolver Digital	SteamVR
	Top_2	Adventure		Oculus Quest 2
	Top_3	Puzzle	Valve	Oculus Quest
1	Top_1	Adventure		Windows PC
	Top_2	Indie	Square Enix	PlayStation 4
	Top_3	RPG	Capcom	Nintendo Switch
2	Top_1	Adventure		PlayStation 3
	Top_2	RPG	Nintendo	PlayStation 2
	Top_3	Platform	Capcom	Arcade
3	Top_1	Strategy		Windows PC
	Top_2	Indie	Electronic Arts	Mac
	Top_3	Shooter	Sega	Linux

From the comprehensive data presented in Table 3, the characteristics of 4-dimension clustering are also obvious (see Figure 3). It becomes clear that both shooting and strategy games have experienced a surge in popularity, marking their prominence within the gaming sector. For a significant duration, these game genres not only attracted a vast number of players but also maintained a sterling reputation for delivering high-quality gaming experiences. They became emblematic of what players sought in terms of gameplay depth, visual appeal, and narrative strength.

Yet, as with many industries, the gaming arena is not static. Recent observations indicate a discernible shift in the trajectory of these game genres. Specifically, there has been a gradual but concerning dip in the quality of titles under the shooting and strategy categories. While the reasons for this decline might be multifaceted, one cannot dismiss the influence of contemporary market dynamics.

The gaming industry, being in a constant state of flux, is shaped by an interplay of technological advancements, player expectations, and market saturation. Given the previous dominance of shooting and strategy games, it is plausible that an influx of developers attempted to capitalize on this trend,

potentially leading to a crowded market. This, in turn, could have led to a rush in game production, prioritizing quantity over quality.

Another angle to consider is the evolving tastes of the modern gamer. With the rise of alternative gaming genres and platforms, traditional genres might find themselves in a position to innovate or risk obsolescence. In the end, the decline in quality observed in shooting and strategy games serves as a critical reminder of the industry's dynamic nature and the perpetual need for adaptation and evolution in response to ever-shifting market demands [10].

This expanded version retains the necessary citation and delves deeper into potential reasons for the observed trends.

### 3.3. Linear Regression

Based on our dataset, we have determined that predicting the rating through Developers, Platforms, and Genres, or predicting the rating based on player-related information, holds the most significant potential value. After conducting extensive experimentation with various methodologies, we ultimately concluded that employing Platforms and Genres for multiple linear regression analysis is the most suitable approach.

After experimenting with various regression approaches, we observed that due to the diversity and randomness inherent in gaming market data, numerous lesser-known games could introduce significant uncontrollable variability into overall trend predictions. This rendered comprehensive forecasting nearly unattainable. Therefore, we made the decision to mitigate the impact of these random factors within our dataset. We employed an innovative approach called "top-labels".

"Top-labels" is the method we use to extract the key features of the data. After arranging the data in descending order of rating, we only allow one label to appear in the top 100 data that contain it. This method offers distinct advantages in gaming market analysis, as it effectively encapsulates the characteristics of various Platforms and Genres, while also providing insights into the current industry status based on label distribution density. Following the preprocessing using this approach, we achieved highly accurate prediction results through conventional linear regression techniques.

The model's performance can be gauged using several statistical metrics, among which the R-squared ( $R^2$ ) index and the mean squared error (MSE) stand out for their significance. In our analysis, the  $R^2$  value achieved was 0.6500. In layman's terms, this suggests that the model can explain approximately 65% of the variance in the dependent variable based on the independent variables provided. This is a noteworthy achievement, especially in the complex realm of gaming market dynamics, where numerous factors interplay. Further delving into the model's metrics, the reported MSE is 0.0784. This metric offers insights into the prediction accuracy of the model. An MSE of 0.0784 indicates that, on average, the model's predictions deviate from the actual values by a margin of approximately 0.28. While no model is infallible, such a low MSE in a domain rife with uncertainties underscores the reliability of the model's predictions. In essence, these statistical values highlight the robustness of our model in deciphering the intricacies of the gaming market. The substantial  $R^2$  index, coupled with a relatively low MSE, positions our model as a valuable tool for stakeholders looking to understand and predict trends within the gaming industry. This success can be attributed to the "top-label" approach's ability to filter out random influences, allowing the model to focus on the most influential and representative data points, leading to improved accuracy and robustness in its predictions.



## 4. Insights and Discussions

Video games have significantly impacted modern society, extending their reach beyond entertainment into areas such as social interaction, education, and mental health. Recognizing their profound influence, this study delves into understanding the complex dynamics of the gaming industry.

### 4.1. The Role of Data Analysis

Our rigorous research methodology was rooted in precision and clarity. Faced with the challenges of categorical data, we utilized one-hot encoding to transform this data into a format amenable to numerical analysis. This transformation was instrumental in ensuring our dataset was compatible with advanced machine learning techniques. By converting qualitative information into binary columns, we were able to leverage sophisticated analytical approaches. This enabled a deeper dive into the intricacies of the gaming industry, allowing for a more nuanced understanding of its trends and dynamics.

### 4.2. Common Traits of Popular Games

Through our meticulous use of advanced data analysis techniques, we delved deeply into the complex interplay between game genres, platforms, and the minds behind them - the developers. This analysis aimed to shed light on patterns and tendencies that seem to strike a chord with the vast gaming community. A notable revelation from our research was the pronounced association of adventure and indie games with the Windows PC platform. This isn't merely a random correlation. The Windows PC platform, renowned for its open environment and flexibility, provides an inviting canvas for developers to experiment and innovate. It has become a hub for groundbreaking game designs that push the boundaries of storytelling and player immersion. Indie developers, often operating without the financial backing of large studios, find the Windows ecosystem especially conducive. It offers them both a relatively low entry barrier and access to a vast global audience. This symbiotic relationship between the platform and these game genres ensures that players are continually presented with rich, immersive narratives and innovative gameplay mechanics, enhancing their overall gaming experience.

### 4.3. Predicting Developmental Trends

Our research, meticulously centered on the present landscape of the gaming industry, is replete with insights that resonate beyond the current moment, potentially shaping the blueprints for future game development. A standout observation is the clear gamer inclination toward titles offering holistic, immersive sensory experiences. These aren't just about graphical or auditory fidelity; they encapsulate a deeper connection, almost blurring the delineation between reality and the virtual realm. This trend signifies a pivotal direction for the industry's evolution, suggesting a future where immersion is paramount. Concurrently, the discernible dip in quality among shooting and strategy games warrants attention. This decline could be a manifestation of market saturation, with myriad titles diluting unique value, or it might spotlight an evolution in player tastes, as conventional game dynamics possibly wane in appeal. For developers and industry stakeholders, understanding and navigating these multifaceted shifts are crucial for sustained relevance in this ever-evolving gaming cosmos.

## 5. Conclusion

This academic journey is more than just another research paper; it represents a significant stride in closing the knowledge gap that has long loomed over the determinants of success in the gaming

industry. The insights unearthed go beyond merely highlighting the prevailing trends. They act as a compass, illuminating the path for future strategies in game development and marketing.

A dominant takeaway is the rising prominence of the Windows PC platform. Historically viewed as a tool for productivity, PCs have metamorphosed into formidable gaming juggernauts, largely owing to their versatility and adaptability. This revelation offers both budding and established game developers a clear indication of where to focus their energies. Moreover, the burgeoning interest in sensory gaming experiences—games that transcend traditional visual and auditory boundaries to offer a more tactile and immersive experience—underscores the industry's shift towards holistic player immersion.

However, it's not all smooth sailing. The observed wane in quality across certain game genres should be a clarion call for the industry. It's a stark reminder that, in a field as dynamic as gaming, resting on laurels is not an option. Today's innovation could easily become tomorrow's redundancy if not continually evolved and refined. The ephemeral nature of trends emphasizes the indispensability of perpetual research and the agility to pivot based on market feedback.

In culmination, this exhaustive exploration does more than just augment our understanding of the gaming universe; it serves as a reservoir of knowledge for anyone involved or invested in the industry. The multifaceted world of gaming is akin to a living organism, constantly evolving and adapting to its environment. The findings from our research act as a lighthouse, guiding stakeholders through the sometimes-murky waters of game development, platform selection, and player preference. With its blend of art, technology, and commerce, the gaming industry's horizon is teeming with opportunities. Armed with these informed insights and a commitment to innovation, its trajectory toward reshaping global entertainment seems unstoppable.

## References

- [1] Sabadello M. *History of electronic games and an interactive installation at the "Technisches Museum" in Vienna, Austria*[D]. 2006.
- [2] Vedenoja P. *Roleplaying Games in Visual Media: How to make tabletop roleplaying games engaging for an outside audience?*[J]. 2016.
- [3] Deterding S, Andersen MM, Kiverstein J, Miller M. *Mastering uncertainty: A predictive processing account of enjoying uncertain success in video game play*. *Front Psychol*. 2022 Jul 26;13:924953. doi: 10.3389/fpsyg.2022.924953. PMID: 35959012; PMCID: PMC9363017.
- [4] Li Lin, Dirong Cao and Xu Yin, "A Study on China's online games development trend," 2010 International Conference on Mechanic Automation and Control Engineering, Wuhan, 2010, pp. 1725-1728, doi: 10.1109/MACE.2010.5536050.
- [5] Rodríguez, P., Bautista, M. A., González, J., & Escalera, S. (2018). *Beyond one-hot encoding: Lower dimensional target embedding*. *Image and Vision Computing*, 75, 21-31. doi: <https://doi.org/10.1016/j.imavis.2018.04.004>.
- [6] *International Handbook of Children, Media and Culture* / Livingstone, Sonia; Drotner, Kirsten - London: Sage, 2008 - 560 p. - ISBN: 9781446206645 - Permalink: <http://digital.casalini.it/9781446206645> - Casalini id: 4912719.
- [7] Ryan, *Video games are fun. Here's why, and how they hook us.*, 2019 Nov 22
- [8] Adair C, *15 Reasons People Play Video Games*
- [9] Sirani J, *Where Switch, PS5 Rank Among the Best-Selling Video Game Consoles of All Time*, 2023 Aug 3
- [10] Kolbe R, *An Analysis of First Person Shooters: What Makes a Successful FPS and Where the Market is Headed*, 2021 Aug 19