

Research on the Commercial Application of Video Games Based on Data Analysis

Yu Xiao

Tianjin University of Science and Technology, Artificial Intelligence Academy, Tianjin, 300450, China
xiaoyuyy6@gmail.com

ABSTRACT

In recent years, the video game market is favored by users and capital, and the number of video game products is showing an exponential growth trend. In the face of the massive video game products, how to use reasonable marketing strategies to increase product revenue, and how to choose products to get more benefits from users are important issues to be tackled. In the era of big data, the combination of data analysis technology for video game business research can get large-scale, diversified, personalized, and real-time research results, and is more conducive to the mining of various data information of users. In this paper, the author crawls vgchartz.com to collect video game data samples. The statistical processing and exploratory analysis are mainly performed on the data of game product distribution platforms, genres, and years of distribution. The K-Means algorithm is used to classify the game products according to sales, and the corresponding business strategies are proposed for platforms, publishers, and users in the context of the market. The results show that platforms can prioritize the purchase of games in the 'Action' genre, and need to pay attention to game releases that focus more on quality than quantity; publishers should choose platforms and game genres with high sales to sell and invent; users should focus on platforms with high sales and experience to choose a certain type of game products. This study can stimulate the marketing and economy and make the market enter a virtuous cycle.

Keywords: Video game, data analysis, business strategy, K-Means

1. INTRODUCTION

Video games came into the limelight with the birth of the original game console. In 1889, the creation of the Nintendo Company brought video games into a climax era, such as the well-known Mario, which has sold more than 373 million copies to date [1]. The video game entered the multimedia era with the competition of 16-bit consoles and the 3D era with the competition of 32-bit consoles. After entering the 21st century, Microsoft, Nintendo, and Sony became the three dominant players of video games. Video game entered China in the late 20th century and has experienced four eras so far: the preparatory era, the leap forward era, the stable era and the multi-platform era [2]. In recent years, with the rapid development of e-sports industry [3], major businesses have started to join the game industry, and video game has entered the golden period of commercial development. Public demand for video games has also increased dramatically, and due to the global outbreak of the New Coronary Pneumonia in 2020, video games have become a mechanism for coping with anxiety in an epidemic setting and one of the most attractive

technological interventions to address the mental health effects of New Coronary Pneumonia [4]. Video games are favored by users and capital. However, in the face of the huge number of video game products, how merchants can use reasonable marketing strategies to increase product revenue and how users can choose products to get more benefits are important issues to be tackled. In addition, data analysis is needed to conduct commercial research on video games and obtain large-scale, diversified, personalized and real-time research results [5, 6], which is more conducive to the mining of various data information of users. This shows that it is a worthwhile direction to investigate the commercial application of video games based on data analysis. In this paper, the author crawls vgchartz.com to collect video game data samples. The statistical processing and exploratory analysis are mainly conducted on the game product release platform, genre, and release year data. The K-Means algorithm is used to classify game products according to sales, and the corresponding business strategies are proposed for platforms, publishers, and users in conjunction with the market to stimulate marketing and market economy.

2. ANALYSIS

2.1. Data Collection.

The dataset in this paper is generated by crawling vgchartz.com and contains a list of video games with more than 100,000 copies sold [7]. The collected data include the ranking of game product sales worldwide, game product name, game product release platform, game product release year (1980-2020), game product genre, game product publisher, and game product sales. There are 16,598 records in total, and 2 records were deleted due to incomplete information.

2.2. Data loading and data cleaning.

In this paper, the author uses Jaime Becerra Guerrero's Sales Cluster & Analysis, Notebook from the Kaggle website [8] as the basis for data analysis and processing. After data reading, it is found that there are 271 and 58 missing values in the columns of 'Year' and 'Publisher', respectively, and the missing data only accounts for 1.63% and 0.35% of the total data, so the amount of missing data is very small, so the rows corresponding to the missing data can be deleted.

2.3. Descriptive analysis.

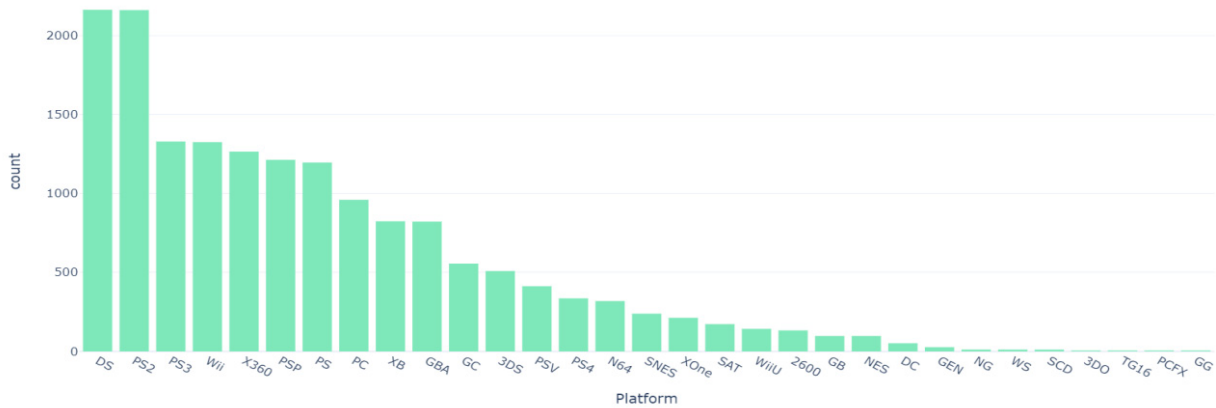


Figure 1. Bar chart of the number of game releases by platform

Figure 1 takes 31 different platforms as horizontal coordinates and the total number of games released for each platform from 1980-2020 as vertical coordinates.

The results show that the five platforms DS, PS2, PS3, Wii, and X360 have the highest number of games released in the period 1980-2020.

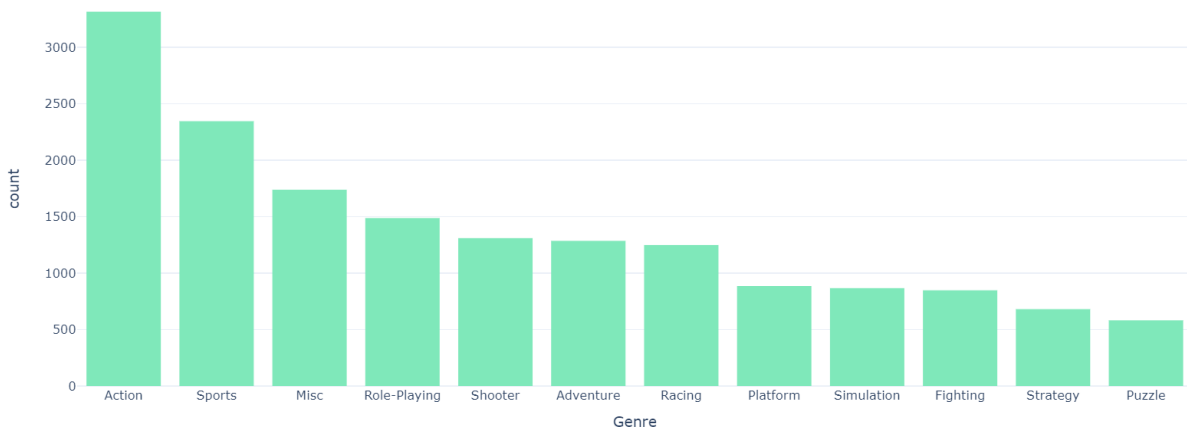


Figure 2. Histogram of the number of game releases by genre

Figure 2 takes 12 different game genres as horizontal coordinates and the total number of games released in each game genre from 1980 to 2020 as vertical coordinates. It can be obtained that the released number of 5 game genres, namely, Action, Sports, Misc, Role-Playing, Shooter, are the highest during 1980-2020

2.4. Data Analysis

2.4.1. Platform

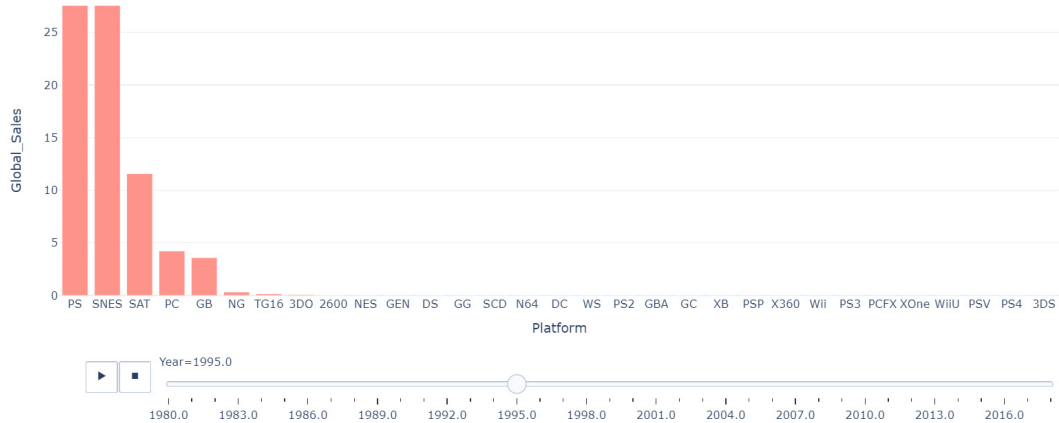


Figure 3. Global sales by platform over the years

NA_Sales: Sales in North America (in millions); EU_Sales: Sales in Europe; JP_Sales: Sales in Japan; Other_Sales: Sales in the rest of the world; Global_Sales: Total worldwide sales.

To display the global sales of each platform in order by year and rank them from largest to smallest, it is necessary to remove the rows corresponding to the

missing data in the 'Year' column and explore figure 5 to see that some of the dominant platforms for the period 1980-2020.

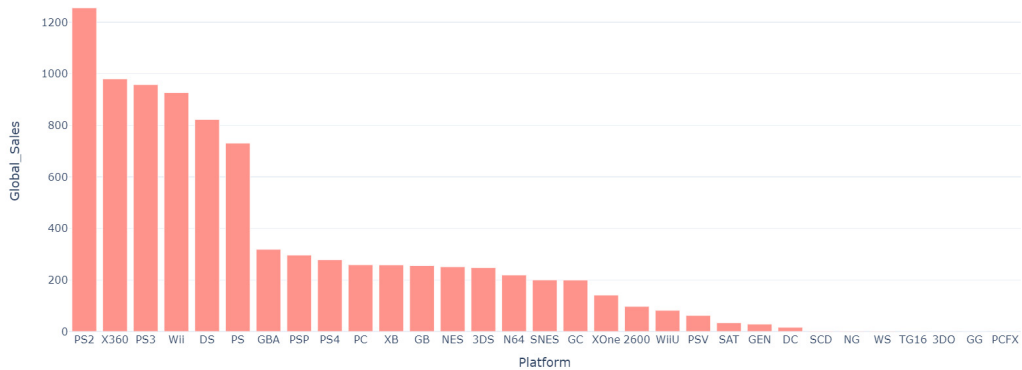


Figure 4. Global sales sorted by platform

The total global sales volume of each platform from 1980 to 2020 is sorted from largest to smallest. As can be seen from Figure 6, PS2 sales volume is the first, amounting to 1200 million.

2.4.2. Genre

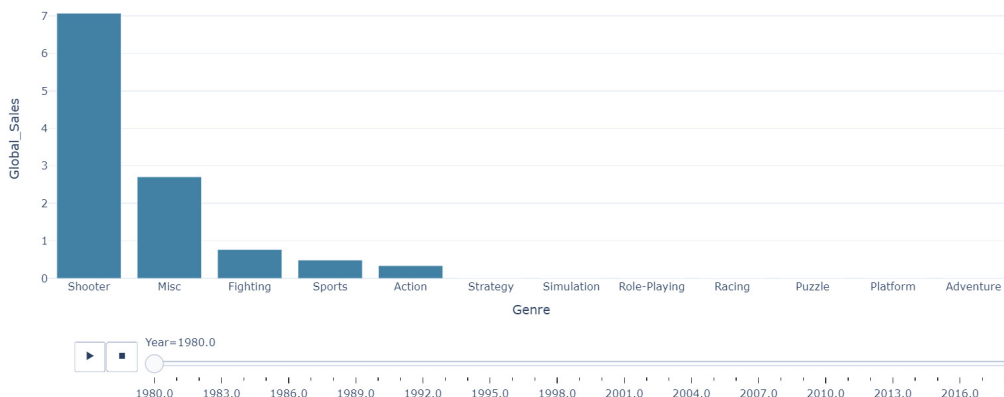


Figure 5. Global sales by genre by calendar year

Figure 7 shows the global sales of each game genre in order by year from the largest to the smallest. As is shown in figure 7, no game genre has dominated for a long time

before 2001, and since 2001, the 'Action' genre has dominated for a long time.

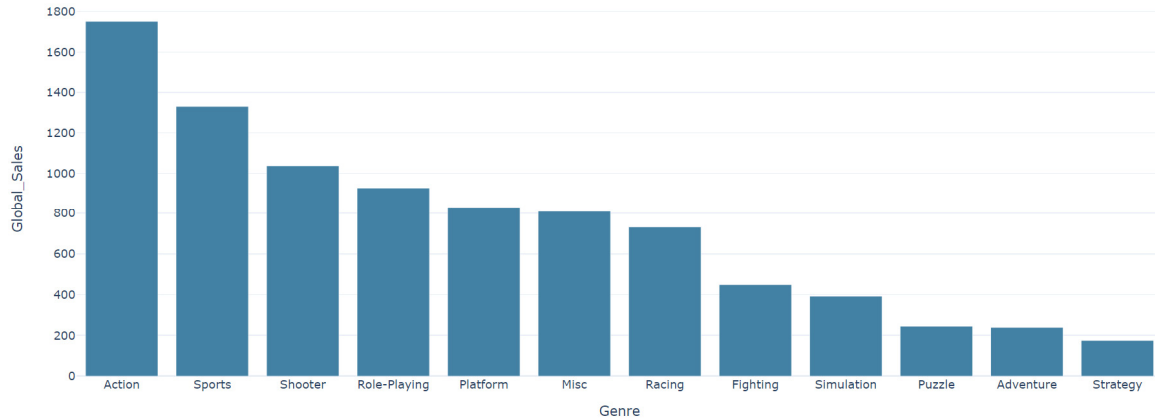


Figure 6. Global sales sorted by genre

In figure 8, the total worldwide sales of each game genre from 1980 to 2020 is ranked in descending order. It can be seen from this figure that the 'Action' genre is the first, with sales of nearly 1,800 million.

In addition, the clustering results are presented in a two-dimensional form with NA_Sales, EU_Sales, JP_Sales, and Other_Sales as the horizontal coordinates respectively and Global_Sales as the vertical coordinates. Figures 10 shows NA_Sales as examples, it can be seen that the '0' class represents low sales, '2' class represents medium sales, and '1' class represents high sales.

2.5. Clustering model.

By using the K-Means algorithm, the data are divided into 3 categories: low sales, medium sales, and high sales.

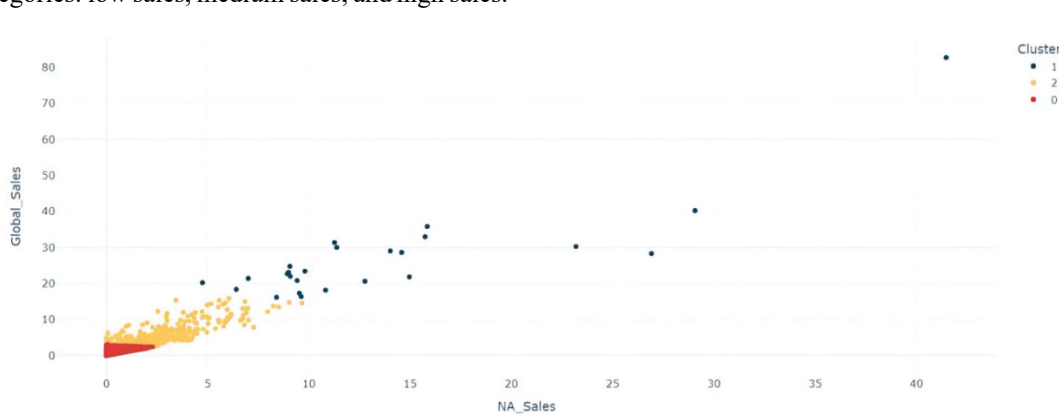


Figure 7. NA_Sales two-dimensional clustering chart

Comparing the number of games with 3 different sales types, the following chart shows that the number of games with low sales is the highest.

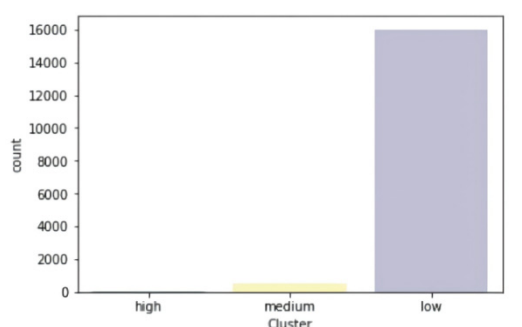


Figure 8. Statistical chart of the number of games in 3 categories

2.5.1. Low Sales Type

Run through the corresponding code to get 5 most frequent Platform in the low category are: DS, PS2, Wii, PS3, PSP. 5 most frequent Genre in the low category are: Action, Sports, Misc, Role-Playing, Adventure. 5 most frequent Publisher in the low category are: Electronic Arts, Namco Bandai Games, Activision, Ubisoft, Konami Digital Entertainment.

2.5.2. Medium Sales Type

Run through the corresponding code to get 5 most frequent Platform in the medium category are: PS2, X360, PS3, PS, DS. 5 most frequent Genre in the medium category are: Action, Shooter, Sports, Role-Playing, Platform. 5 most frequent Publisher in the medium category are: Nintendo, Electronic Arts, Sony Computer Entertainment, Activision, Ubisoft.

2.5.3. High sales type

Run through the corresponding code to get 5 most frequent Platform in the high category are: Wii, DS, GB, NES, PS2. 5 most frequent Genre in the high category are: Platform, Action, Sports, Role-Playing, Misc. 5 most frequent Publisher in the high category are: Nintendo, Take-Two Interactive, Microsoft Game Studios.

3. RESULTS

3.1. Platforms

From Figure 8, it can be obtained that the sales of 'Action' type games have been at a high level, therefore, major platforms can give priority to 'Action' type games when buying video games. From figure 1, people can see that although the DS platform has released the largest number of games, the revenue earned is only the 5th among the major platforms, indicating that the number of games released is not proportional to the revenue earned, and the major platforms still need to carefully measure and filter to decide on the purchase of games.

3.2. Publishers

From figure 6, people can get that publishers put their games into PS2, X360, PS3, Will, DS platforms to sell, and make more profit with less risk. From figure 8, people can get that publishers inventing 'Action' and 'Sports' type of games will cater to the market and gain more revenue, but there is also a risk of competition in the same industry.

3.3. Users

From figure 6, it can be obtained that users focus on PS2, X360, PS3, Will, DS platforms and may get cost-

effective products. When specifying a certain game genre to select the games for the platform, people can first conduct the corresponding statistical analysis. For example, when specifying the 'Action' type of game for the platform selection, figure 13 provides statistics. It can be concluded that on the basis of the 'Action' type of games, the number of PS3 platform games ranked first, accounting for 11%, indicating that the PS3 platform is experienced in running the market for 'Action' type games, and users may get a better sense of experience. Therefore, users can look for 'Action' type games for PS3 platform first.

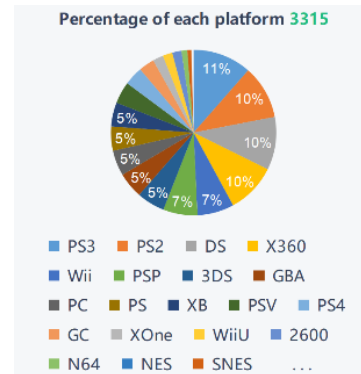


Figure 9. Pie chart of the number of games by platform under the 'Action' game genre

4. CONCLUSION

This paper focuses on cleaning, descriptive analysis, data analysis of related contents and data visualization of the video game dataset to derive the corresponding business application strategies for different groups in the market. The results show that platforms can prioritize the purchase of games in the 'Action' genre and need to focus more on game distribution quality than quantity; publishers should choose platforms and game genres with high sales to sell and invent; users should focus on platforms with high sales and experience to choose a certain type of game product. This study can stimulate game marketing and economy and bring the market into a virtuous cycle. Although this paper uses K-Means algorithm to divide the sales of video game products, it does not dig deeper and analyze the products in each division, and more commonality of products under each division can be explored for improving and proposing better business strategies. The video game market will have broader sales prospects and is worth promoting and propagating vigorously.

REFERENCES

[1] Kaiser McDonald. Super Mario: the legend of the top plumber[J]. Windfall Generation,2021(05):46-47.
 [2] Hua Xia. Study on the development history of online games in China[D]. Liaoning University,2018.

- [3] Yu Weitao. Research on the development of China's e-sports industry [J]. *Enterprise Reform and Management*,2020(01):201-202.
- [4] Pallavicini Federica, Chicchi Giglioli Irene Alice, Kim Gerard Jounghyun, Alcañiz Mariano, Rizzo Albert. Editorial: Virtual Reality, Augmented Reality and Video Games for Addressing the Impact of COVID-19 on Mental Health[J]. *Frontiers in Virtual Reality*,2021:
- [5] Liang Feng. Research on the development trend of intelligent leisure sports based on big data analysis[J]. *Journal of Physics: Conference Series*,2021,1941(1):
- [6] Gao Ting. Research on short video recommendation strategy based on big data analysis[J]. *Journal of Physics: Conference Series*,2021,1941(1):
- [7] GregorySmith, video game sales,2018, <https://www.kaggle.com/gregorut/videogamesales/>
- [8] Jaime Becerra Guerrero, Sales Cluster & Analysis,2021.3, <https://www.kaggle.com/jaimebecerraguerrero/sales-cluster-analysis/>