

Video Game Analysis: Trends and Insights

Daniel Ethridge¹ Atharva Patil², Deep Shukhla³, Sujith Battu⁴

University of Colorado, Boulder

Author Emails

¹Corresponding author: daet2304@colorado.edu

²atpa5127@colorado.edu

³desh3965@colorado.edu

⁴suba2075@colorado.edu

Abstract. This project examines patterns and trends within the video game sector, utilizing information from diverse sources to gain insights on game popularity, sales statistics, player demographics, reviews analysis and genre success. The study's goal is to pinpoint crucial factors that impact the success of video games and to represent these patterns through graphs and charts. This report offers a thorough summary of the video game market's current condition by gathering and organizing pertinent data, facilitating well-informed decision-making for stakeholders.

INTRODUCTION

The video game industry has quickly transformed from a specialized market to a worldwide force, propelled by technological progress and a variety of player demographics. This project examines patterns in the video game industry, with a specific emphasis on sales figures, genre trends, and player involvement.

Our goal is to discover important insights for developers and marketers by analyzing data from different sources like sales numbers and critical reviews. What types of genres are the most prevalent in the market? What is the relationship between game ratings and sales? By carefully gathering, scrubbing, and presenting data, this document offers a thorough examination of the gaming industry, emphasizing the correlation between consumer actions and market patterns. It is essential to grasp these dynamics in order to successfully navigate the future of the industry.

DATA COLLECTION/PREPARATION

This section discusses the procedure and measures we choose to collect data and the sources of our database. This section also details the issues and errors while preparing the data for visualization process.

Data Sources

1. SteamDB: Sales and user reviews for PC games.
2. API Access: The data utilized in this work comes directly from Steam via an API. There are three main API calls utilized to collect the data:
 - <https://api.steampowered.com/ISteamApps/GetAppList/v2/>. This is used to acquire all Steam app IDs.
 - <https://store.steampowered.com/api/appdetails?appids=<APPID>>, where APPID is the steam application ID number. This is used to collect details about a single Steam application.
 - https://store.steampowered.com/appreviews/<APPID>?json=1&num_per_page=100&cursor=<CUR>&filter=recent&purchase_type=all, where CUR is a string representing the next page of reviews if there are multiple pages.

Fig.1 is the image of the json file of the collected data.



FIGURE 1. Image of json file of data

Data Schema

Here is the schema of the final database we collected to make sure o the structure of the dataset.

Fig.2 Below shows the clear schema of the database.

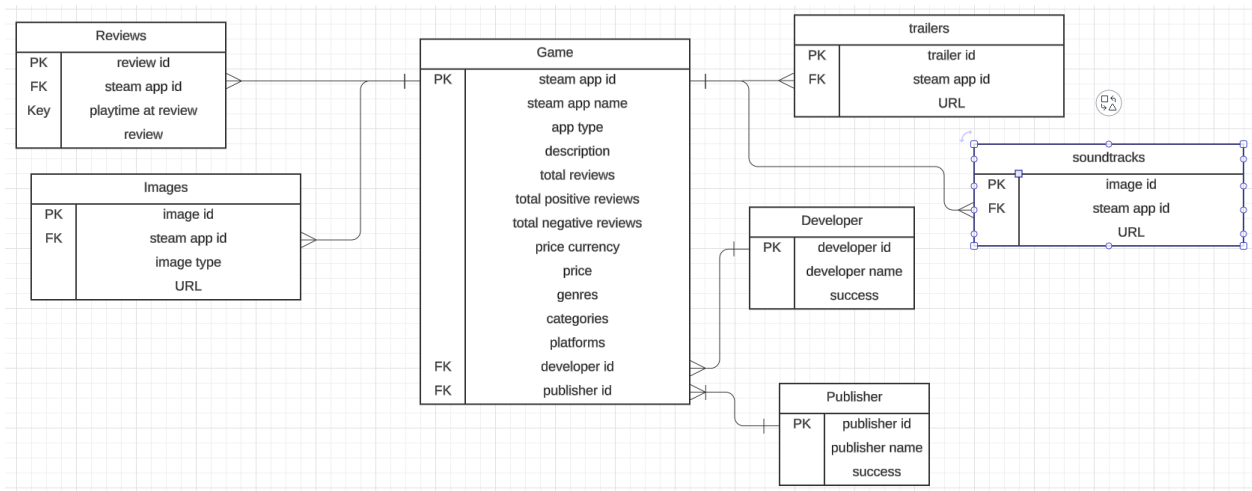


FIGURE 2. Data Schema

Fig.3 shows the exact data file we collected.

Tools Used

- Pandas: Used for data cleaning and manipulation in Python.
- NumPy is used for performing numerical computations and managing missing data.

Data Visualization

Plots used:

1. Bar Charts
2. Line Graphs
3. Heatmaps
4. Pie Charts
5. Bubble graph
6. Scatter plot

Visualizations:

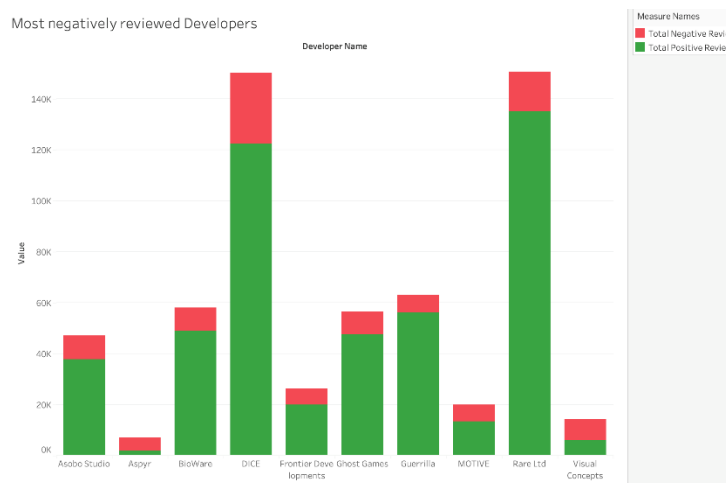


FIGURE 4. Negatively reviewed reviews 10 developers

The [Fig.4](#) below show initial insight into the data. Below each is a caption describing them. The x axis is the top 10 most negatively review game developers and y axis is the total positive and negative reviews. We can see that rare ltd and DICE are the two most negatively reviewed game developers and that most developers have almost nearly equal positive reviews as well.

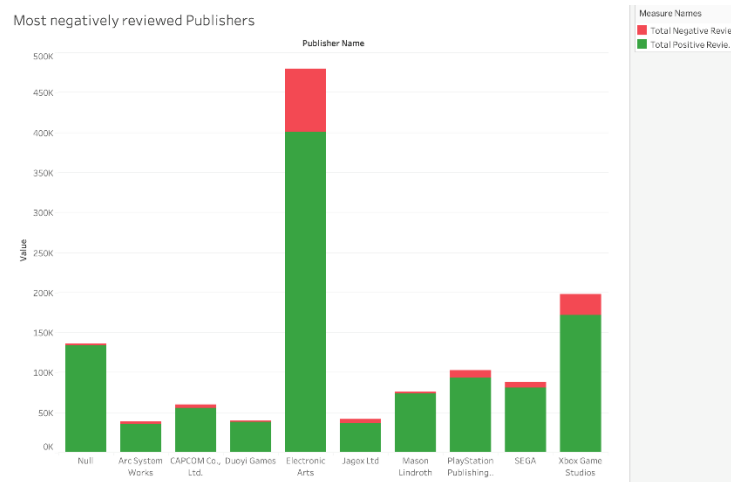


FIGURE 5. Most negatively reviewed 10 publishers

This graph in Fig.5 displays the top 10 most negatively reviewed game publishers and the number of positive (green) reviews and negative (red) reviews they get. We can observe that electronic arts has the most negative reviews as well, and that most of these publishers have comparable number of positive reviews to the negative reviews. Implying that people tend to have mixed opinions on game publisher games.

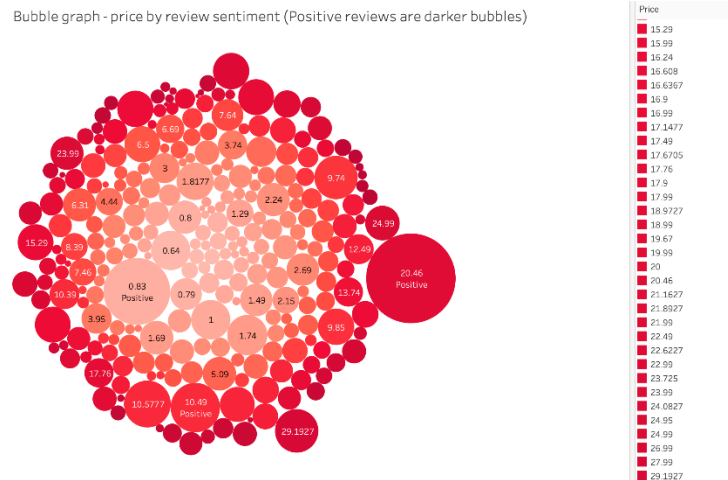


FIGURE 6. price vs reviews

The bubble plot Fig.6 shows the price vs review sentiment (which is the ratio of positive and negative reviews). The bubble size denotes the value of review sentiment (larger bubble implies positive review). While the colour denotes the price, more red implies more pricey game. We can observe that more pricey games tend to have more positive review sentiment, and vice versa. Which can be confirmed as less pricey games tend to be made by cheaper indie game developers who cannot make the best of games.

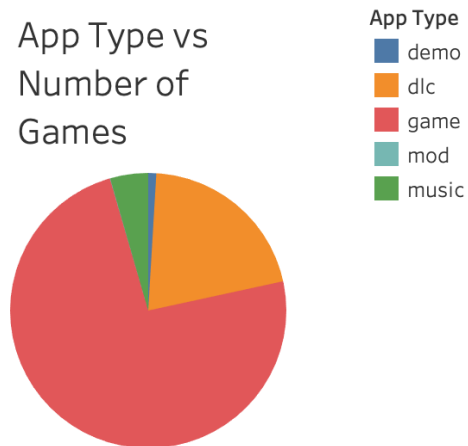


FIGURE 7. Game types

The pie chart [Fig.7](#) shows the app types like game, demo, dlc, etc vs the total number of games. We can observe that the game app type is most common followed by dlc and music.

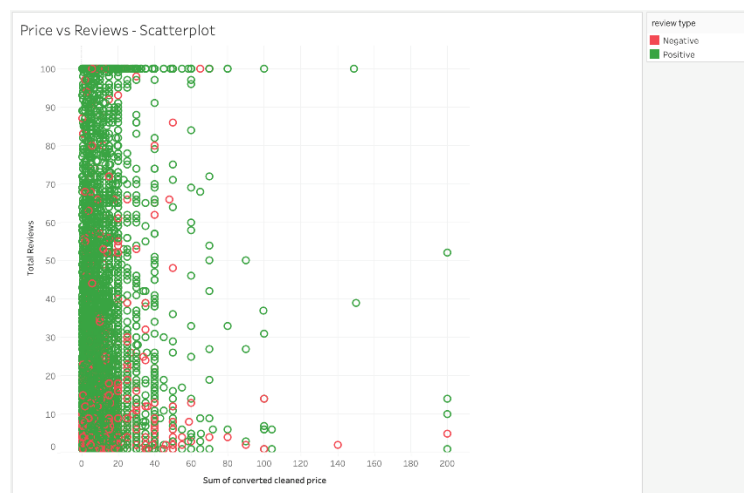


FIGURE 8. Price vs Reviews

The graph in [Fig.8](#) total number of reviews for each price range of each game. Red are net negatively reviewed games and green and net positively reviewed games. We have removed outliers to make the graph more digestible. We can see that most games are positively reviewed, only a few are negatively reviewed. This mean the data is more skewed towards positive reviews. We can also observe that the data is more concentrated towards the left end, meaning games tend to be less pricey. But on the y axis the total reviews are quite equally spread apart, meaning there is no visible trend with games and number of reviews.

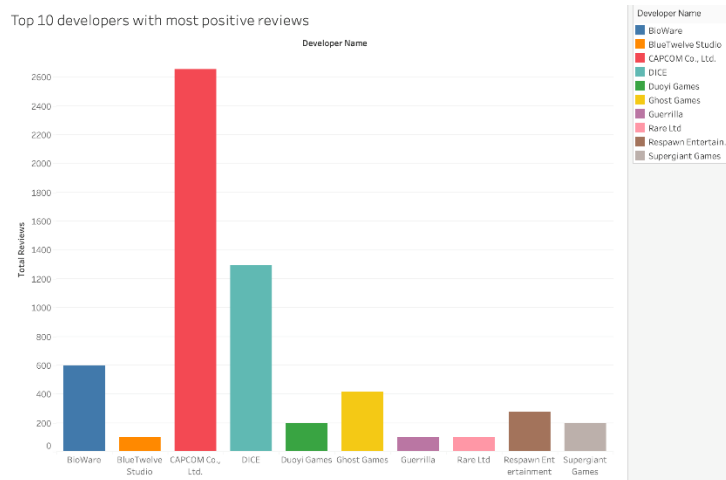


FIGURE 9. Most positively reviewed 10 developers

In above graph Fig.9 X axis is the top 10 game developers based on positive user reviews. Y axis is the total reviews received by each developer. We can see that capcom co has the highest reviews and is followed by dice , but there is a huge difference between the two. We can see that BioWare is the most positively reviewed developer.

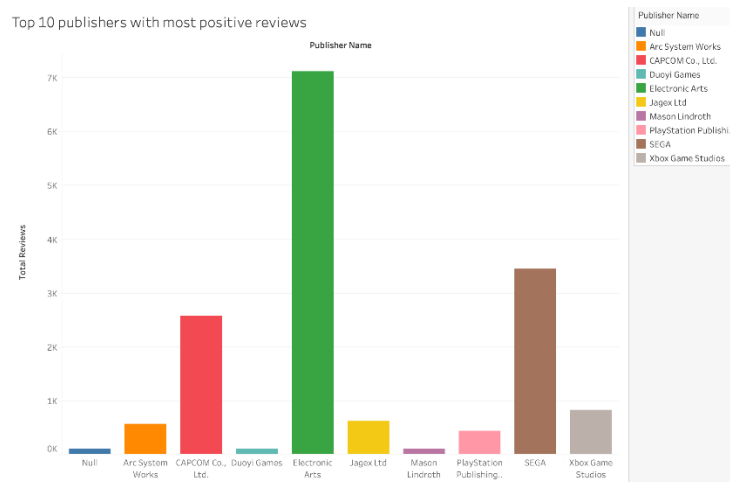


FIGURE 10. Most positively reviewed 10 developers

Above plot Fig.10 has X axis as the names of the top 10 publisher names, which were found using the most positive reviews. Y axis has the total number of reviews. This graph shows the top 10 publisher names, and that electronic arts has the highest number of reviews, and arc system works has the highest number of positive reviews.

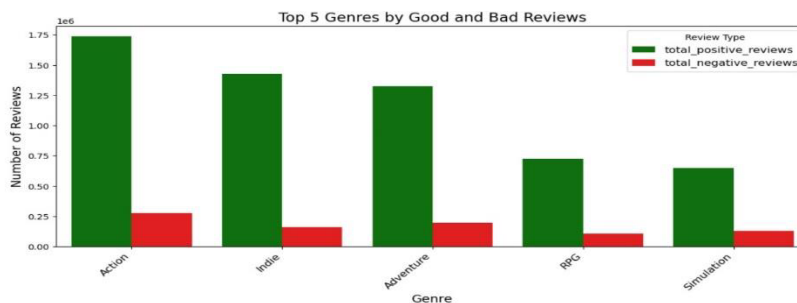


FIGURE 11. Genre analysis (Good vs Bad reviews)

Fig.11 X axis has the names of the top 5 game genres, which were found using the most positive reviews. Y axis has the total number of reviews. This graph shows the top 5 genre names, and that action is the most liked game genre and these have very little negative reviews.

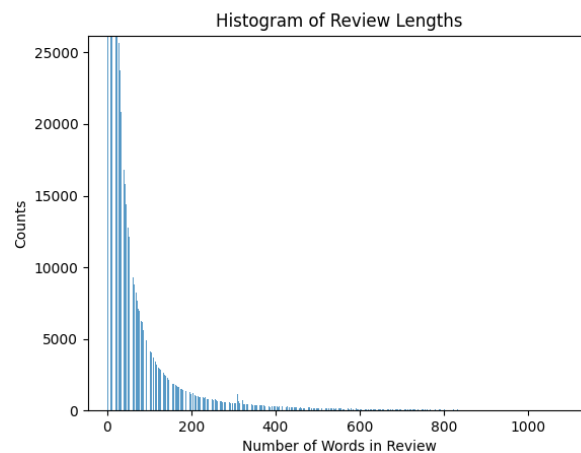


FIGURE 12. Number of words in reviews

The x axis in Fig.12 shows the number of words in a review and the y axis shows how many reviews had that many words. Perhaps unsurprisingly, words with fewer words are more common while reviews with more words are less common.

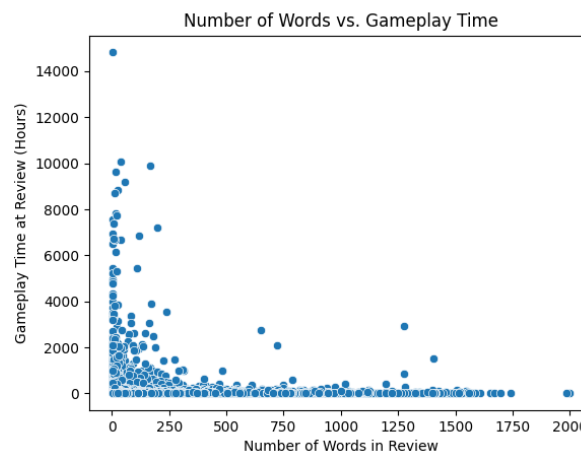


FIGURE 13. Number of words vs Gametime

The Fig.13 shows number of words in a review plotted against how long someone had played the game before leaviing a review. Intuitively, one may believe that longer, more indepth reviews would come from people who have played the game longer. However this is not the case. Longer reviews typically come from people with a shorter amount of playtime. This could present interesting ideas for analysis later on.

CONCLUSION

This project effectively studied patterns in the video game sector by gathering, refining, and presenting extensive data. The knowledge acquired can help developers, marketers, and stakeholders in making educated choices, matching product offerings with consumer preferences, and recognizing possible market opportunities. Future research might include a more in-depth examination of player involvement statistics and a wider investigation of upcoming gaming technologies.