Psychographic Segmentation of Gamers: An Analysis of Gameplay Time, Sales, and Review Score

Aidan Olson

May 2023

Abstract

In this project I initially sought to understand what makes a video game successful according to Review Score, Sales, Genre, and various measurements of Gameplay Time. I found that depending on how the data is represented, there emerge two completely opposing conclusions backed by equally convincing data. I found that Review Score and Gameplay Time both do and do not matter in the success of a game. I later determined this constituted a psychographic segmentation among gamers.

I found two player bases who appear to behave in opposite ways: one player base buys games according to how they score on a rating scale. Consequently, these players buy a game solely based on reviews and, subsequently, the quality of its gameplay. The other player base cares less about reviews and gameplay quality. These players seem to favor games with multiplayer features and shorter gameplay time. We can even draw connections between how these differing behaviors cause them to gravitate toward different genres.

Dataframe Information

The data provided features many elements including the time required to beat certain video games from 2004 to 2010 in various playstyles. These playstyles include:

- Completionist: 100% game completion, including all game extensions and hidden quests, also known as Easter eggs. This metric is useful in determining the overall "size" of a game.
- Main Story: Involves only the content required to beat the core game, without any extensions or Easter eggs.

- Main Story + Extra: This section is slightly more vague, but it includes beating the main game and any major expansion packs, which are large downloadable content packs that often come after initial release.
- All PlayStyles: This section is like an average of the other three. As such, All PlayStyles is a more standardized measurement of gameplay time. It is a measurement of how much time people invest in the game, rather than how much time it takes someone to beat it. Therefore, this metric encompasses both casual players and more committed players.

Other elements in the dataframe include:

- Review Score: A typical review score for this game, out of 100.
- Max Players: The maximum number of players that can play this game.
- Sales: The total sales made on this game, measured in millions of dollars.

Introduction

While it's challenging to measure how "exciting" or "unique" a game is, through the given variables it's possible to establish what makes a game successful.

1. How does the Gameplay Time of a game impact its Sales?
2. Do people seek longer games to get more "bang for their buck"?
3. What about shorter games designed around "replayability"?
4. How do Review Score, Sales, and Gameplay Time differ by Genre?
5. Do certain Genres generally have longer games that appeal more to consumers?

Identifying Trends in Gameplay Time and Sales

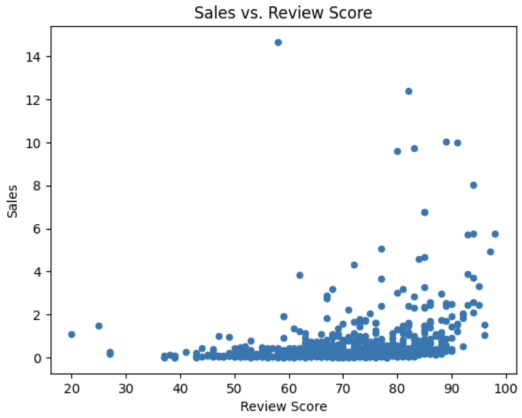
To begin with, I used scatter plots with the three numerical variables:

- Gameplay Time
- Sales
- Review Score

Sales and Review Score

The first figure is a two dimensional visualization of an individual game's Sales and Review Score. Each point represents a game from our dataframe. The x axis displays the game's Review Score and the y axis displays the game's Sales performance.

Figure 1: Sales ($ millions) vs. Review Score



This is a trivial idea and, for the most part, Figure 1 tells us that high Review Score slightly correlates to higher Sales. This correlation is not strong, however.

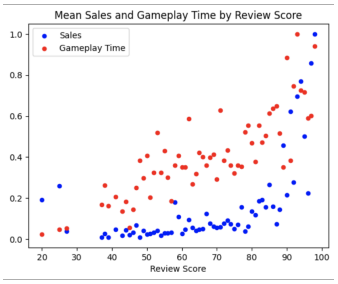
Gameplay Time and Sales

There are two ways to visualize the relationship between Gameplay Time and Sales. Since Review Score is already discretized, I think it would be effective to group by review score and find the mean values of Gameplay Time and Sales per each review score. So, we can group games by their Review Score then average Sales and Gameplay Time, then normalize the two plots and compare them on the same graph where the x value identifies the Review Score grouping and the y value denotes the average relative Sales or Gameplay Time value compared to other games (Figure 2a-d). Or, we can simply plot each game as a single scatter plot where the x axis displays how the game measures in Gameplay Time and the y axis displays how the game's Sales performance (Figure 3).

The first method—grouping by review score—would help answer the simple question: "do better reviewing and or longer games make more money?" A positive trend in either Sales or Gameplay Time would indicate a "yes" to the respective question. On the other hand, the second method—direct comparison—would show us more raw data similar to Figure 1.

Grouping By Review Score

Figure 2: Sales (normalized) and Gameplay Time (normalized) by Review Score



(a) All PlayStyles

(b) Main Story

(c) Main Story + Extras

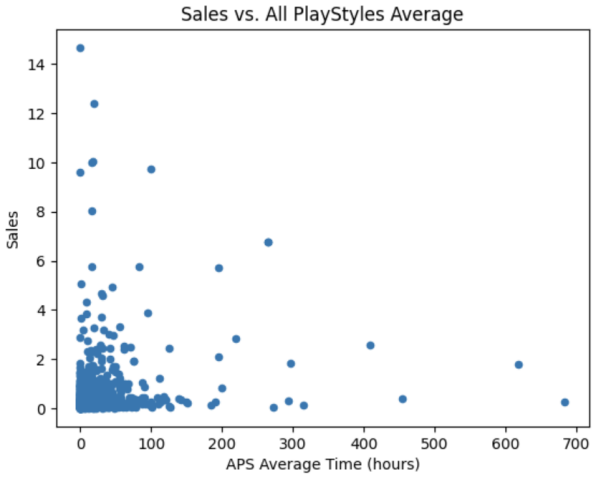
(d) Completionists

Overall, these figures suggest there is a correlation between Review Score, Gameplay Time, and Sales. We can observe that high review scoring games sell more, and are longer in general. These figures lead us to believe that games with longer Gameplay Time often have better Sales. Hence, this supports the "bang for you buck" hypothesis.

As we cycle through the playstyles, we can see that the Time plot slowly converges to the sales scatter plot. We have to keep in mind that these plots are normalized, so I suspect that there is at least one game (at the top) that has an overwhelmingly long Completionists Time. Therefore, when dividing by the max gameplay time, most of the shorter games get squished together and take the shape of the Sales plot. Since these figures all show positive trends regardless of which Gameplay Time metric is used, we can safely say that using any of these metrics will provide equally significant results when looking for trend patterns. Thus, since it is the most average of the four, I will use All PlayStyles as the Gameplay Time variable for future analyses. That being said, it is notable to mention that the two plots in Figure 2d take a very similar shape. It may be a coincidence due to the nature of the normalization process described above, but the shapes are too similar.

Direct Comparison

Figure 3: All PlayStyles Average Time (hours) vs. Sales ($ millions)



This Sales vs. All PlayStyles Average scatter plot provides an opposing argument that Sales and Gameplay Time are indirectly related. A notable difference between this plot and the last few is that earlier figures were grouped by Review Score. When we ungroup by Review Score and directly compare Sales and APS Average Time, we get a figure that supports the "replayability" hypothesis. What exactly about grouping by Review Score causes this dramatic split in our data visualization? Also, how could this happen to such a strong degree?

Earlier we saw that Review Score showed a slight correlation with Sales. By that logic, we would assume that Sales and APS Average Time would be directly related. So, why does this figure show the opposite correlation?

Depending on how we compare these three variables, we can get different results and draw completely opposite conclusions with fair certainty. Hence, there is reason to believe these figures are highlighting something else. Since we're mainly concerned about video game sales, I suspect we've identified two cases where consumers in the video game market behave differently.

Psychographic Segmentation Hypothesis

Our two plots provide the basis for a psychographic segmentation among gamers. This is the idea that there are distinct groups of gamers that consume video games differently. I think the data outlines the behavior of two groups: Multiplayer Centered Players and Singleplayer "Grind" Players.

Multiplayer Centered Players (Figure 3)

These players like investing their money on multiplayer games where they can be social. These games are shorter (as evidenced by the figures) and promote replayability by sporting simple game features that make the games more digestible. These players support the "replayability" idea. Since they tend to reach a bigger player base, I suspect this is why Figure 2 shows they get more sales.

Multiplayer Centered Players care more about the overall experience of playing a game rather than the game itself. This can include who you're with, when you play, where you play, and other factors. So, much of the experience is independent from the game which makes playing "bad" games not very different from "good" games. Similar to how we enjoy watching bad, corny movies, these players can enjoy playing bad games with low Review Score. As such, Review Score has no correlation to Sales for these games. This is also another explanation for how games like Wii Play made a lot of money. It wasn't a particularly 'good' game with a ground breaking idea, but many people fun playing couch games with friends and family.

This analysis paired with the fact that data in Figures 1a-d is averaged is why we see a gentle, positive slope from $20 <$ Review Score < 80 of Figures 1a-d.

Singleplayer "Grind" Players (Figures 2a-d)

These players like investing their money in longer games that are often single player. Such games require a lot more time to complete because of long stories with complicated lore. As such, these games are less digestible for the average player and are likely less popular. I suspect this is why Figure 2 shows they get less Sales.

Since these games require such a large time investment, Review Score is important in determining which games get played. For games with bad Review Score, the "Grind" Players will ignore it. According to our movie analogy, these players would be most likely to watch films nominated to win an Academy Award. As such, most of the games bought and played by this player base will have high Review Score. When we examine game Sales, these long, singleplayer games that depend heavily on Review Score will generate more Sales.

These players support the "bang for their buck" idea. As such, it's players like these who contribute to the strong correlation between completion time and sales since they seek to complete the entire game. Perhaps this is why the Time plot converged to the Sales plot in Figure 1a-d.

This analysis paired with the fact that data in Figures 1a-d is averaged is why we see a steep, positive slope from Review Score > 80 of Figures 1a-d. Since longer games often have more complicated and interesting lore, this could be why we see this correlation between Gameplay Time and Sales. In short, long games have better reviews that weigh the average sales data when reported by review score.

Proving the Psychographic Segmentation by Genre Distinctions

Now that we have a hypothesis for the two different playstyles, let's conduct a similar analysis as performed above and explore how the introduction of Genre as a new variable will impact our figures. I will start by identifying which genres are predominately singleplayer and multiplayer. By filtering the Max Players category, we can get a better idea.

To prove the psychographic segmentation, I want to identify two genres that are on the opposite ends of the multiplayer spectrum. I first need to count how many games there are by genre with only 1 max player (SP) and then the games with strictly more than 1 max player (MP). Then, the MP/SP Ratio will give me an idea of where each genre lies on the multiplayer spectrum. A higher number means it is more multiplayer centered.

Game Genre	**MP/SP Ratio**
Action	0.379630
Adventure	0.111111
Educational	NaN
Racing / Driving	0.5
Role-Playing (RPG)	0.091667
Simulation	0.347826
Sports	1.275862
Strategy	0.238095

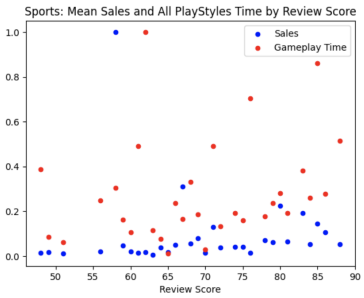
By this ratio test, most sports games are multiplayer and most RPGs are singleplayer. This makes sense as adventure games and RPGs are meant to be an individual experience while sports are more team focused. Let's do another sales analysis and see what we get.

If our hypothesis is correct, we should find little to no Sales to Gameplay Time correlation in our Sports graph but find some remnants of a positive trend for Sales to Gameplay Time as Review Score increases in our RPG graph. Perhaps when conducting a similar analysis as above by genre, we can answer the questions we asked earlier: what exactly about grouping by Review Score causes this dramatic split in our data visualization?

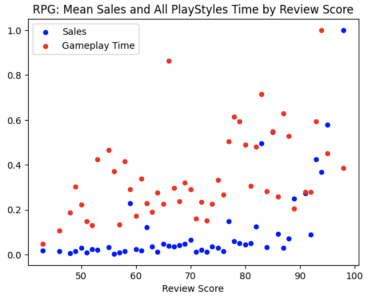
As I discussed in the "Gameplay Time and Sales" section introduction, grouping by review score assumes that if a game is good, it will make money. Per our psychographic segmentation hypothesis, this distinction should separate the games that rely on having good reviews from those that rely on multiplayer features to drive market success. This is an attempt to disprove the notion that good games are always successful games and vice versa. If we can use the same two data visualization methods as before but while separating multiplayer and singleplayer games, we can hopefully pull similar results from both methods of data visualization and prove our pyschographic segmentation hypothesis.

Grouping By Review Score (Distinct Genre Adaptation)

Figure 4: Sales (normalized) and APS Time (normalized) vs. Review Score



(a) Sports

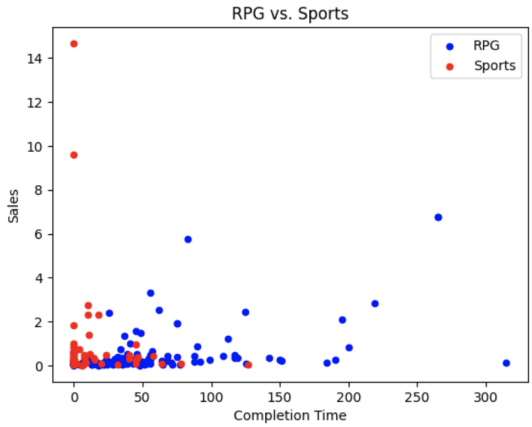


(b) RPG

As predicted in our hypothesis, the Sports graph data is much more scattered compared to our RPG graph. Thus, it is clear that Sales is not dependent on Review Score for Multiplayer Centered games like Sports games. That is not the case for our Singleplayer representative. RPG games show a slight trend. Let's the direct comparison analysis and see if we get similar results. We should expect RPG to occupy mostly the right and Sports to occupy mostly the left.

Direct Comparison (Distinct Genre Adaptation)

Figure 5: Sales ($ millions) and Completion Time (hours) of RPG and Sports Games



Most of the extreme left points are red while all of the extreme right points are blue. This convincingly validates the psychographic segmentation hypothesis.

Additional work is required to identify whether other explanatory values may be involved.

Conclusion

In this project, we identified the two types of gamers. Those who seek multiplayer games to have fun with friends, and those who seek singleplayer games to get lost in gameplay with engaging storytelling. Using this pyschographic segmentation, we found two distinct and opposite game genres, Sports and RPG, whose game features uniquely serve the two player types. Through the interpretation of Gameplay Time, Sales, and Review Score, we established a psychological segmentation on two player bases and determine how their different needs are satisfied by different Genres.

Singleplayer players value buying a high quality game with long gameplay times to ensure they spend their money as efficiently as possible, only buying the best games. These players support the "bang for your buck" hypothesis. The Multiplayer players value short games with multiplayer features to optimize replayability. They like shorter games because they're simple to understand and easier to use as a tool to socialize.

However convincing this argument may be, the hypothesis was likely incomplete. These variables have shown to depend on each other, but they may depend on others that were not provided by this dataframe. Variables that measure Total Lifetime Players or Maximum Live Players would give another indication of game popularity in addition to Review Score and Sales. Maybe other dataframes that focus less on games themselves, but rather gamers would help conduct a similar analysis.