Amir Yaacoobi

Kian Ghaffari

Tristan Kao

Gaming Trends Analysis

Introduction: Video games have swiftly become a go-to escape for many individuals. A majority of them flock to Steam, known as the biggest and most prominent video game distribution service on PC, Mac, and Linux. Unless they are playing on Xbox and Playstation, most customers use Steam to buy their games. While there are other smaller competitors to Steam, such as EpicGames Store, they do not have nearly the same consumer base as Steam and they are known for giving out a free game every week so the frequency of free games will inflate the number of games sold, which will in turn skew the data. Steam provides various services, including the purchase and download of video games, game updates, community engagement features, and multiplayer gaming functionality. The platform was formally launched in September 2003 by Valve Entertainment, and was created as a means to perform automatic updates for their own video games, but eventually expanded to include games from third-party publishers.

Hypothesis: Using the data, we wanted to first discover what video game genres would be the most popular and how popularity shifted for those genres from 2000 to 2019. We expect to see Free-To-Play video games having the most popularity throughout the decade due to their accessibility. We also examined each genre's profitability throughout the period to see whether there is an economic incentive to make certain kinds of games and how it has changed

throughout it. With this in mind, we expect to see First-Person-Shooters (FPS) as the most profitable due to their versatility as a genre.

API Usage, Processing, and Methodology: In this exploration, we accessed an API to extract the Steam database. The API allowed us to look at the different parameters, removing the need to sort raw data. Although the data was extremely messy, cleaning the data was manageable. However, the cleaning and simplification of the various columns were particularly extensive. We accessed two separate APIs to get the necessary information to answer the questions and support our hypotheses. The first is an API for SteamSpy, an independent API that uses public information from Steam users. It accepts requests in a GET string and returns data in JSON arrays. Using the same method, the other API we used was a CSV file from the Steam app itself. In addition, we had to use a CSV file to supplement and merge the data for visualization purposes since the API did not include data from 2009 and onwards.

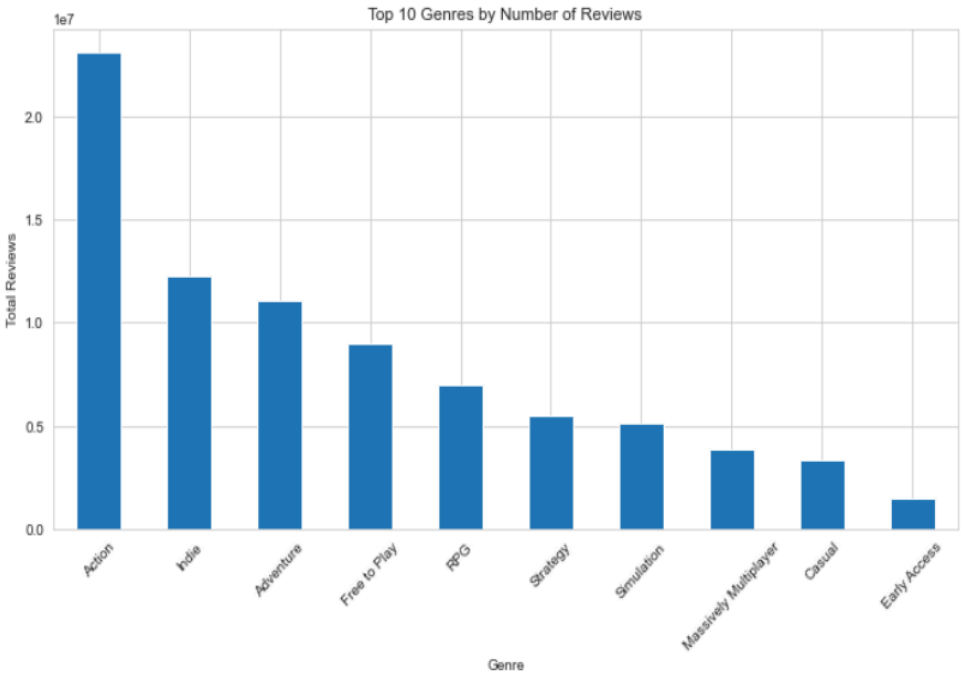
	steam_appid	name	release_date	developer	publisher	platforms	categories	genres	steamspy_tags	positive_ratings	negative_ratings	owners	price	reviews
0	10	Counter-Strike	2000-11-01	Valve	Valve	windows;mac;linux	Multi-player;Online Multi-Player;Local Multi-P...	Action	Action;FPS;Multiplayer	124534	3339	10000000-20000000	7.19	127873
1	20	Team Fortress Classic	1999-04-01	Valve	Valve	windows;mac;linux	Multi-player;Online Multi-Player;Local Multi-P...	Action	Action;FPS;Multiplayer	3318	633	5000000-10000000	3.99	3951
2	30	Day of Defeat	2003-05-01	Valve	Valve	windows;mac;linux	Multi-player;Valve Anti-Cheat enabled	Action	FPS;World War II;Multiplayer	3416	398	5000000-10000000	3.99	3814
3	40	Deathmatch Classic	2001-06-01	Valve	Valve	windows;mac;linux	Multi-player;Online Multi-Player;Local Multi-P...	Action	Action;FPS;Multiplayer	1273	267	5000000-10000000	3.99	1540
4	50	Half-Life: Opposing Force	1999-11-01	Gearbox Software	Valve	windows;mac;linux	Single-player;Multi-player;Valve Anti-Cheat en...	Action	FPS;Action;Sci-fi	5250	288	5000000-10000000	3.99	5538

sum(diag(tab1))/sum(tab1)

Before visualizations, we first started by cleaning the data so that a lot of the useless variables were removed. We made sure to filter out irrelevant data that would either clutter or skew our

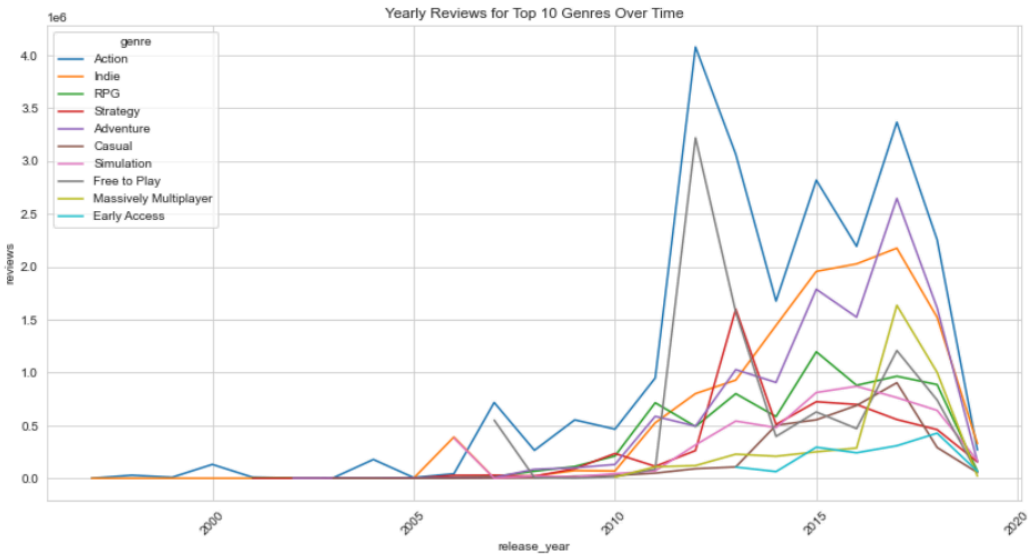
visualizations, leading us to make inaccurate conclusions about our exploration. Additionally, we got rid of all the unnecessary columns in the table so that the data could be more focused on the visualizations we needed. With game revenue, we had to estimate the revenue using what is called the Boxleiter Method, which takes the number of reviews in each game, multiplies it by a constant of 20, and finally multiplies it by its initial price. This had to be done due to Steam not allowing access to the actual sales numbers of all the games. After cleaning each of the datasets we got from accessing the SteamAPI and the Steam Spy, we combined them. Then, we supplemented the dataset that has a full range of what we want to use for our exploration and merged the API data and the Kaggle dataset to make a complete database to create the table above to begin our data visualization and explore the data.

Analysis + Visualizations: We first began by working on which genre is the most popular. To begin, popularity is a very subjective metric and can be defined by many variables. In our case, we would be using total reviews from all games in our data frame as our metric for popularity. With this in mind, we created a bar graph looking at the 10 most popular video game genres. Through this bar chart, action was shown to be the most popular among the different genres.



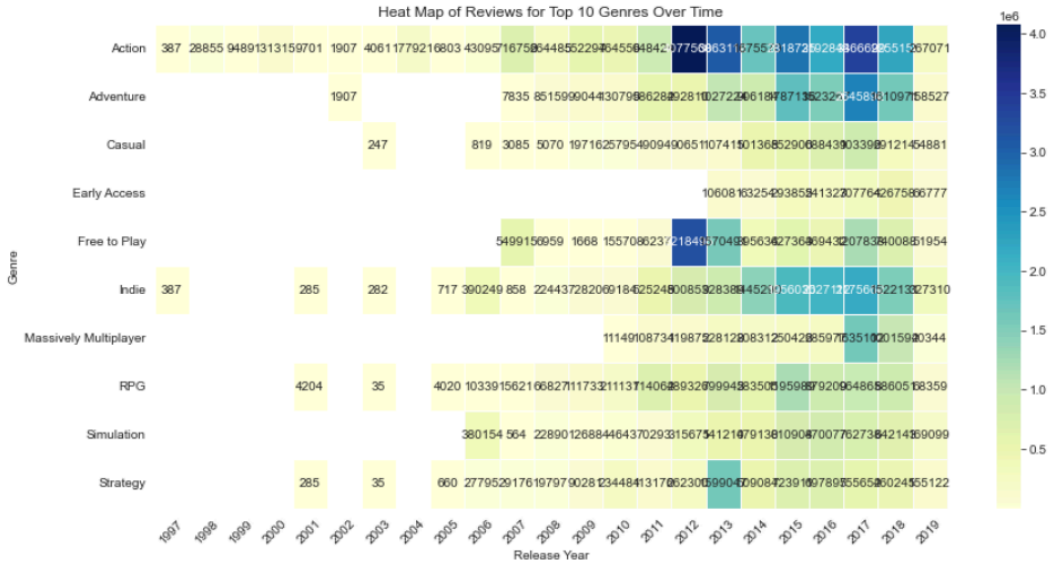
With this in mind, we moved forward in answering our question by looking at the 10 most popular genres over time from 2000 to 2019. We made a scatter plot with the year of release and popularity (total reviews) and color-coded the data points to the 10 most popular genres. Viewing the legend, it is quite clear that action games and action-adjacent games are popular for the whole time frame. Furthermore, people tend to not only lean toward the action genre, they lean toward the free-to-play genre. The combination of the two genres shows that gamers tend to want something engaging and affordable. This can be seen with spikes in 2007 and 2013 with those years coming out with free-to-play action games that have thriving player bases that continue to play it.

We also felt it important to convert this into a line graph to look for a more clear flow of time as well as examine the trends in more general genres besides just the action genre. With the line graph showing the 10 most popular general genres, it

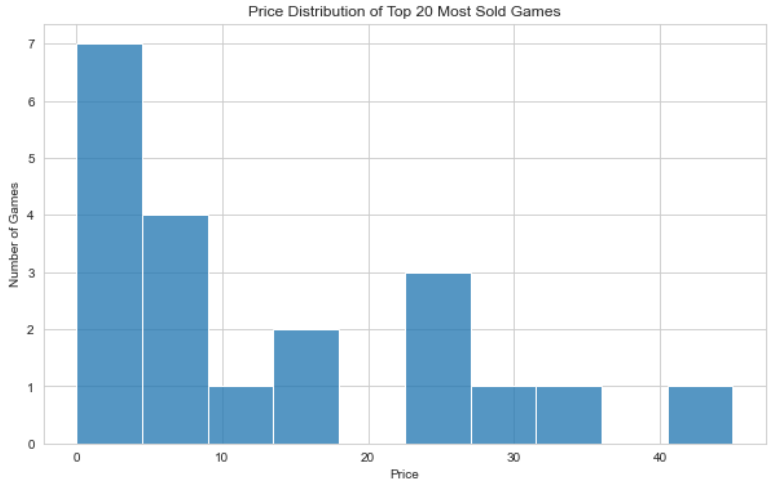


still presents action games as having the highest points in popularity from 2007 to 2019. However, it is also important to note that Strategy had a spike in 2014 and then plummeted in popularity, demonstrating a lack of interest in this type of genre. It is also important to note that certain years had spikes from the fact that older games can be bought at a later time, especially through special Steam sales. Additionally, many people can get nostalgic about a popular game in the past.

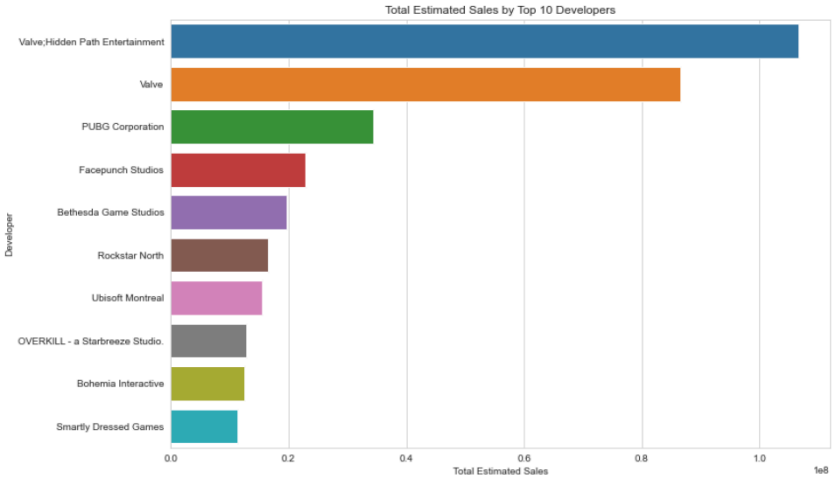
We also created a heat map that looked for saturation of reviews over time using genre as different categories. Once again, the action and free-to-play genre comes up on top. This time, it is in terms of reviews. This shows that games with a higher reception tend to perform better than the games that did not do as well in reception.



Before we asked the second question, we were curious about how much the top 20 most-sold games cost so we implemented our data into a histogram that presented that distribution. As it's shown, there is a fairly right-skewed distribution with a majority of the games being less than 20 dollars and the peak number of the most sold games being from free games.

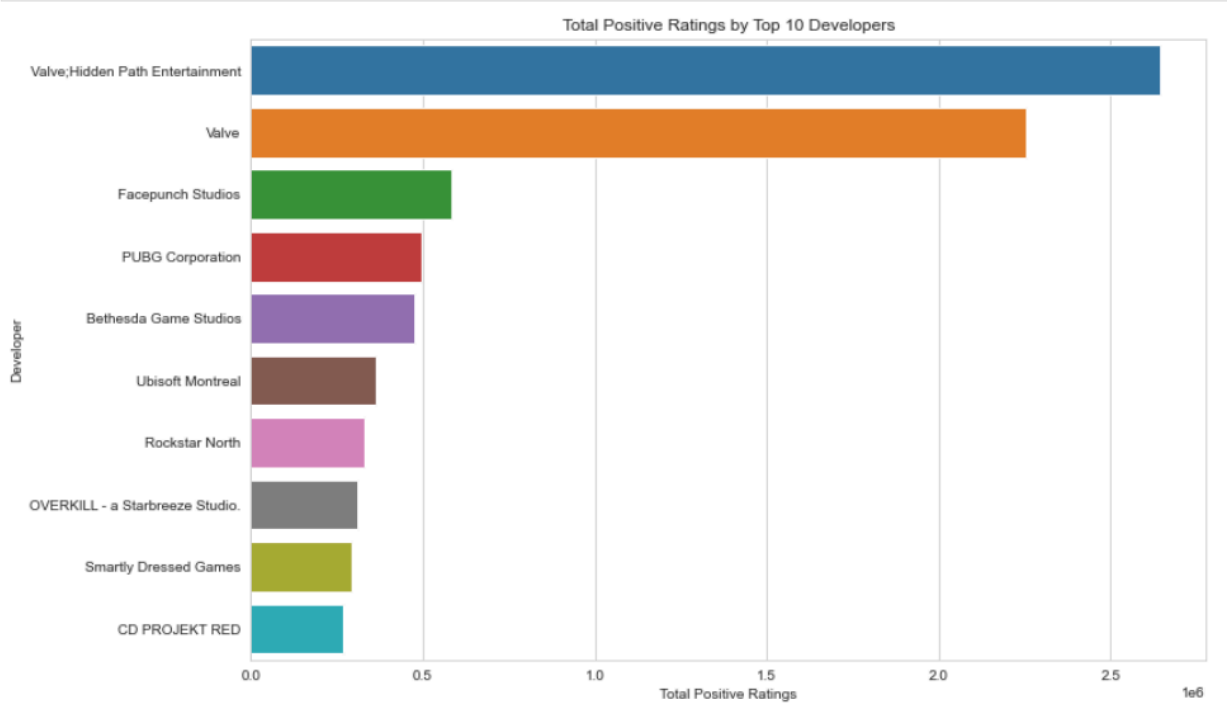


When looking at answering the second question, we wanted to actually take a look at the total estimated revenue for the 10 most popular developers instead of games since this has a more impactful understanding of the gaming industry and in turn the sales. We thus made a horizontal bar graph of the 10 most popular developers and the total estimated sales. Through the bar graph above, Value-Hidden Path Entertainment was found to have the highest sales out of all the studios. Also, it is good to note that all these studios



focus on developing low to medium-cost action games for the gaming masses.

The result from the previous graph made us consider a correlation between positive rating and estimated sales, keeping other variables constant. Examining the horizontal bar graph, there



is a huge correlation between developers (and in turn games) that have positive reviews and their sales numbers. In other words, people tend to only buy highly rated games. Additionally, it demonstrates that Valve, the owners and creators of Steam itself, are widely popular among the gaming community.

Conclusion: Throughout the years, the action genre has exponentially increased its dominance in the gaming industry. This implies that gamers prefer a more stimulating experience with a more fast mechanic. More importantly, gamers tend to also prefer to spend on a lower budget, since low to medium-costing games are higher selling. This is supported by the fact that free-to-play action games ended up being the most popular genre in all the graphs and visualization. This trend has been further emphasized since all the top studios are known for

making games that require some sort of aggression such as FPS (first-person shooter) and RPGs (role-playing games). It also shows that studios have found other ways to monetize their games without making customers pay an upfront fee. It is hard to know how the market will evolve in the future, but we know that gaming studios will always have ways to keep their games affordable. Overall, the gaming industry continues to provide consumers with the thrill that they expect without having to charge such a hefty price.