AJ Garcia and Evan Price
IS 310: Computing in the Humanities
Lucian Li
12/15/2024

Indie in the Mainstream: What Makes an Indie Game Popular?

Repository: https://github.com/evanp28/is310-final-project
Datasets: https://uofi.box.com/s/130iae2hqoefxj8p3kovw3pjfymeyjj5

Introduction

Despite being around for a relatively short period of time, online gaming has become an incredibly popular and influential form of entertainment. More recently, indie games have achieved more mainstream fame and visibility, making them an important cultural phenomenon within the world of gaming. Indie games, short for independent, are games developed by independent developers or publishers, without the financial or technical support of a bigger publishing studio. The advent of platforms like Steam, which has provided a platform for independent developers to create and share their games, hase helped indie game creators achieve both financial and popular success. In our project, we attempted to understand why certain games become popular and how the culture perceives indie games that achieve mainstream success.

While it has historically not been recognized by the academy, gaming is an important cultural product with an immense amount of influence. One need not look far for evidence of this: almost 40% of the global population spends at least some time gaming.[1] Both of us spend our leisure time gaming, both socially and alone, and both of us have played indie games. In doing this research, we hoped to find what makes a game "indie," and which indie games are achieving the most mainstream success and why. We also wanted to find out why people played indie games, and if the reasoning behind choosing to purchase an indie game differed from the reasoning behind choosing to purchase a game produced by a AAA company.

This is a much narrower scope than our original project proposal. We originally wanted to look at whether indie games have become more or less popular with time, and who plays indie games. In the course of our research, we realized that indie is a difficult and ambiguous category to identify, and rather than looking at who plays indie games and how much, that it was much more interesting and relevant to research what the public perception of the category "indie" is, and how "indie" functions as a cultural

[1] Unravelling the complexity of the Video Game Industry:

marker within the universe of online gaming. That is to say, if and how the perception of indie games or a game as indie influences its popularity and ability to achieve mainstream success. By exploring these questions, we hope to investigate gaming culture and its contours more broadly, and whether or not "indie" is an important taxonomy within online gaming culture.

Scholars like Garry Crawford have explored the social impact of gaming, pointing out that the communal nature of online gaming is one of the most important motivations for gamers,[2] suggesting that options for multiplayer play may contribute to a game's success and demonstrating the importance of further scholarship that seeks to understand the social impact of gaming through the study of relationships and networks. Similarly, scholars have pointed out the social impact of indie games, especially as "indie" is one of the largest growing categories of games on Steam.[3] Scholar Óliver Pérez Latorre argues that the emergence of "mainstream-indie" games reflects how innovation and artistry can lead to commercial success in the industry.[4] But what makes a game indie? Is it innovation and artistry? Previous scholarship on the topic demonstrates that "indie" is not an easily defined term with concrete contours. "Indie" is an amorphous category with no concrete definition, and frequently reflects users' perception of a game, not its characteristics. Latorre argues that any definition of indie based on tangible markers like production and distribution excludes games that would be categorized as indie via cultural markers associated with indie production such as outsider status and experimental game design.[5] In our research, we attempt to understand whether it is the tangible or cultural markers of indie games that drive their popularity. In the paper "An Empirical Study of Game Reviews on the Steam Platform," researchers demonstrate that reviews tend to focus on game mechanics and overall gaming experiences, especially in early game[6], indicating that while game design is a key factor in a game's success, research is needed to understand if experimental game design appeals to players.

Data Collection & Curation

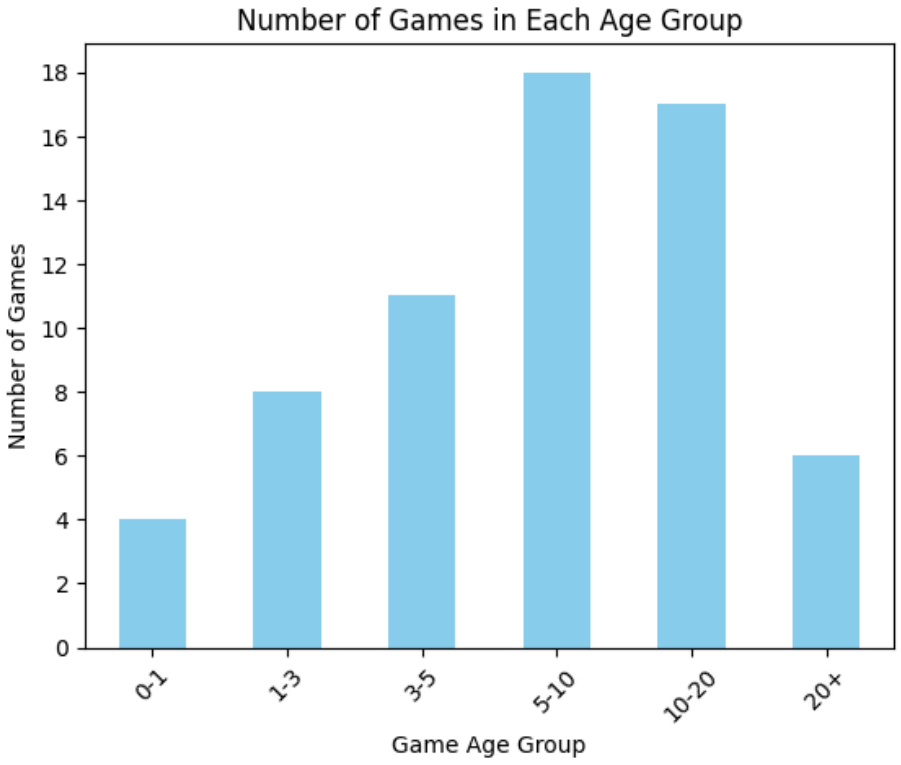
The initial dataset was collected using the Twitch API. We collected the top 100 games on the site at a moment in time. We used Twitch because the Steam API does not allow users to collect the top games on the site. As a result, our research will reflect games that have popularity and cultural sway because of their streamability in addition to normal playability. After generating the initial dataset, we removed any instances that weren't games, like "ASMR" or "Just chatting" which are popular categories of streaming on Twitch. This data included an id, the name of the game, a URL to the game's box art, and The ID that

[2] Crawford, "Online Gaming in Context."
[3] Mendes and Cunha, "Popularity of video games and collective memory."
[4] Latorre, "Indie or Mainstream? Tensions and Nuances between the Alternative and the Mainstream in Indie Games."
[5] Ibid.
[6] Lin, Bezemer, Zou, Hassan, "An empirical study of game reviews on the Steam Platform"

IGDB uses to identify the game. Then I manually added the developer, publisher, initial release date, genre (based on Twitch's genre system), engine, mode (whether a game is multiplayer, single-player, or both), and the base cost of the game on steam for each game in the dataset. Then I cleaned the data by removing the URL column and the IGDB id column as we did not plan to use them, and reordering the columns. After cleaning and hand-curating the data, the dataframe generated from it looked like this:

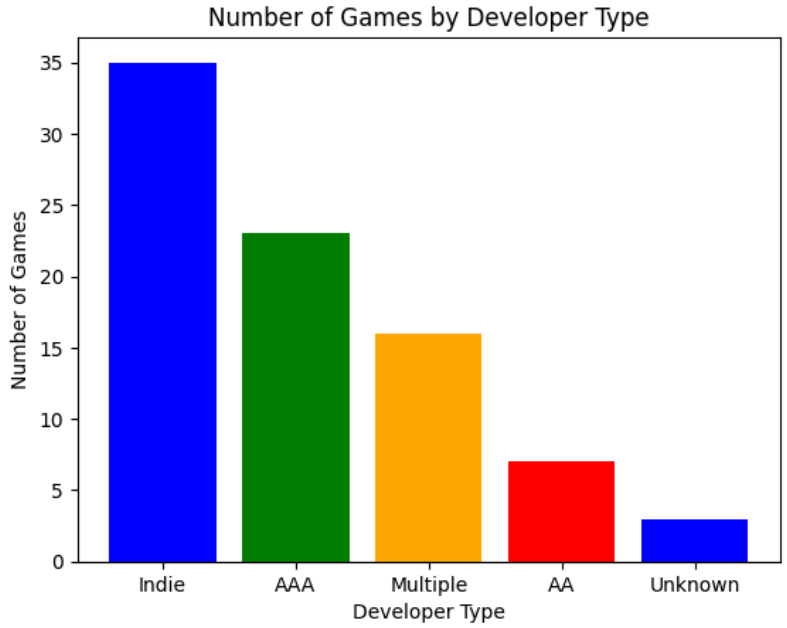
	id	name	platforms	developers	publishers	initial_release_date	genres	engine	modes	base_cost_on_steam
0	263490	Rust	PlayStation 4,Microsoft Windows,Linux,macOS,Xb...	Facepunch Studios,Double Eleven	Facepunch Studios,Double Eleven,Plaion	2013-12-11	FPS Shooter,RPG,Adventure Game,Indie Game,Surv...	Unity	Multiplayer	39.99
1	33214	Fortnite	macOS,Windows,PlayStation 4,Xbox One,iOS,Ninte...	Epic Games	Epic Games	2017-07-25	Shooter,RPG,Simulation,Strategy,Adventure Game...	Unreal Engine 5	Multiplayer	NaN
2	32982	Grand Theft Auto V	PlayStation 4,PlayStation 5,PlayStation 3,Xbox...	Rockstar North	Rockstar Games	2013-09-17	FPS,Shooter,Driving/Racing Game,Adventure Game...	RAGE	Single-player,Multiplayer	39.98
3	1672324422	Call of Duty: Black Ops 6	PlayStation 5,Xbox Series X and Series S,PlayS...	Treyarch,Raven Software	Activision	2024-10-25	FPS,Shooter,Action	IW	Single-player,Multiplayer	69.99
4	21779	League of Legends	macOS,Microsoft Windows,Mac operating systems	Riot Games	Riot Games	2009-10-27	RPG,Strategy MOBA,Action	Proprietary Engine	Multiplayer	0.00
...	...	...	...	...	...	...	...	...	...	...
79	32502	World of Warships	Microsoft Windows,iOS,Android,PlayStation 4,Pl...	Wargaming	Wargaming	2015-09-17	FPS,Shooter,RTS,Strategy,Simulation	BigWorld	Multiplayer	0.00
80	515024	Diablo IV	PlayStation 5,PlayStation 4,Xbox Series X and ...	Blizzard Entertainment,Team 3	Blizzard Entertainment	2023-06-05	RPG,Action,Horror,Adventure Game,MMO	Proprietary Engine	Single-player,Multiplayer	49.99
81	538054672	TEKKEN 8	PlayStation 5,Xbox Series X and Series S,Micro...	BANDAI NAMCO Studios,Tekken Project,ARIKA	Bandai Namco Entertainment	2024-01-26	Fighting,Action,Arcade	Unreal Engine 5	Single-player,Multiplayer	69.99
82	456845141	Zenless Zone Zero	PlayStation 5,Microsoft Windows,Android,iOS	miHoYo	miHoYo	2024-07-04	RPG,Adventure Game,Action	Unity	Single-player	NaN
83	20596	Dofus	Flash,Microsoft Windows,macOS	Ankama Games	Ankama Games	2004-09-01	RPG,Strategy,Adventure Game,MMO	Adobe Flash	Multiplayer	NaN

Based on this initial dataset, we were able to make some observations about the types of games that were most popular on Twitch. For example, the games on Twitch leaned older, with the median game age being 6 years old and the mean game age being 7.49 years old. This indicates that older games tend to be more popular because they have an established fanbase, especially on a site like Twitch, where people monetize playing games.



We also observed that of the 84 games we looked at, 75 offered multiplayer play, confirming our initial assumption that community drives online gaming as a phenomenon. We also found that the games tended to be affordable—the mean cost of a game was $24.98 and the median cost was $19.99, and 25% of the games were free (21 out of 84). This also may indicate that affordability is a factor in indie games popularity, as they tend to cost less than AAA games.

We then added a column indicating if the developer or developers involved were from Indie, AA, or AAA studios. We defined a studio as Indie, AA, or AAA using the schema developed by Video Game Insights, which uses the Steam API and IGDB API to research Market and Steam Analytics, such as average price and user ratings. They define a AAA publisher as a publisher who has "over $7.5m revenue per game released, over $250m lifetime revenues on Steam, average game price of over $17 and has an average game complexity above a VGI defined threshold," AA publishers as publishers with "$1-7.5m revenue per game released, $50-250m lifetime revenue on Steam, average game price of $10-17 and has an average game complexity above a VGI defined threshold," and Indie publishers as any publishers that don't fit the AA criteria. In adding developer information to our dataset, we found that many games had multiple types of developers involved. For example, *Call of Duty: Black Ops 6* was developed by two publishing houses: an Indie and AAA studio. We also found that many games developed by Indie studios were frequently published by AA or AAA studios, demonstrating the instability of Indie as a category. From this, we were able to observe that the vast majority of games had some level of involvement from an Indie studio, demonstrating the influence of indie studios on the gaming industry.



We then generated a second dataset by collecting reviews for each game in this top games list. To collect the reviews for each game we used the SteamWorks API, which by default provided 20 reviews

per game. For the future it would definitely be a good idea to include a lot more than 20 reviews for each game.

For further analysis on how indie games are reviewed by consumers compared to other games we used the SteamWorks API, which provides reviews for games from Steam. To do this we first used the top games list we got from Twitch and cleaned it, so that only games that are actually available on Steam are included in the list. This is because some of the games on the list aren't available on Steam, Fortnite is an example of this. We then made requests to the SteamWorks API using the "GET store.steampowered.com/appreviews/<appid>?json=1" endpoint. This provided us with 20 reviews per game by default and we decided to use this because we thought it was sufficient enough for an initial analysis, but in the future using more reviews would provide a much better picture.

Data Analysis & Methods

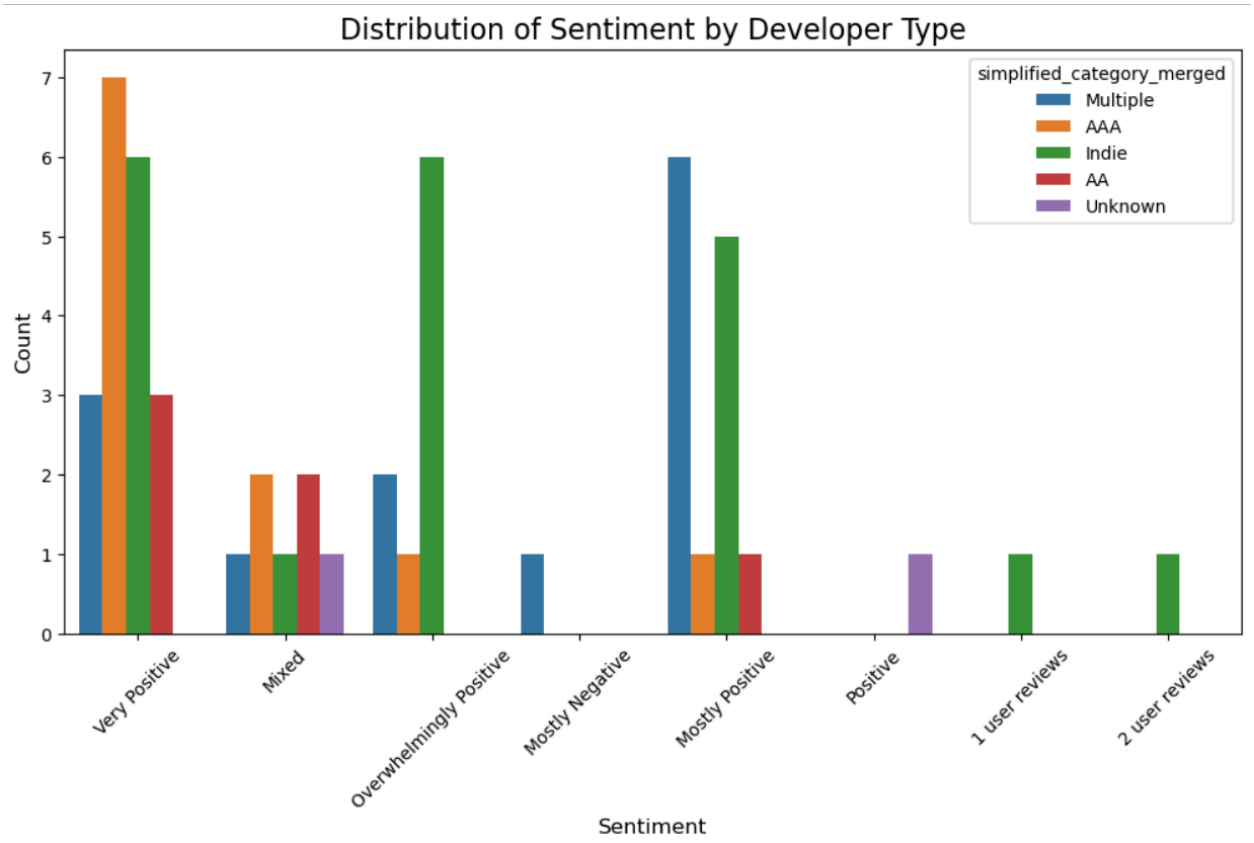
The first method we decided to use to determine how indie games are being reviewed is a BERT model that can perform sentiment analysis. The decision to use a BERT model was heavily influenced by the section in class we did on word embeddings and more specifically the article we read for class about BERT by Britney Muller.[7] This article explains a lot about how BERT works, but what was important to us is that it included a list of pre-trained BERT models with a sentiment analysis model being one of them. This sentiment analysis model is pre-trained on data from Twitter and classifies provided text using word embeddings into one of three sentiments (positive, negative, neutral). We decided to use sentiment analysis to determine the sentiment of each Steam review in our dataset and to do this we imported the model, cleaned the reviews (removed punctuation, numbers, etc.), and then added a column to our dataset that included the sentiment of the review.

	game_title	appid	recommendationid	review	cleaned_reviews	sentiment
0	Rust	252490	177409644	Rust is a great game. It is like real life if ...	rust is a great game it is like real life if t...	NEU
1	Rust	252490	177619447	This game makes me want to kill myself. It's s...	this game makes me want to kill myself its so ...	NEG
2	Rust	252490	178793028	Great game very unique and grindy, which I per...	great game very unique and grindy which i pers...	POS
3	Rust	252490	177844978	[h1]My [i]Rust[/i] Review - 223 Hours In[/h1]\...	hmy irusti review hours inh bratingb hour...	NEG
4	Rust	252490	178578421	Great fun, very addictive and highly toxic... ...	great fun very addictive and highly toxic prep...	NEG

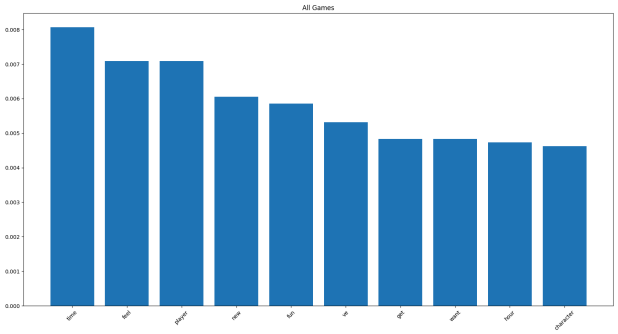
We then decided to create a visualization to show the sentiment of the reviews for indie games vs. other games. To do this we decided to use a bar graph and this is where the hardest challenge came because in order to do this we had to merge the Twitch and Steam review data frames in order to get the developer type for each game. The issue is that the data frames didn't have all the same games and some of the games included had slightly different names in each dataset, so because of this the dataset shrank even more since we had to drop any games with unknown values that resulted from the merge. The

[7] Britney Muller, "BERT 101 🤗 State Of The Art NLP Model Explained"

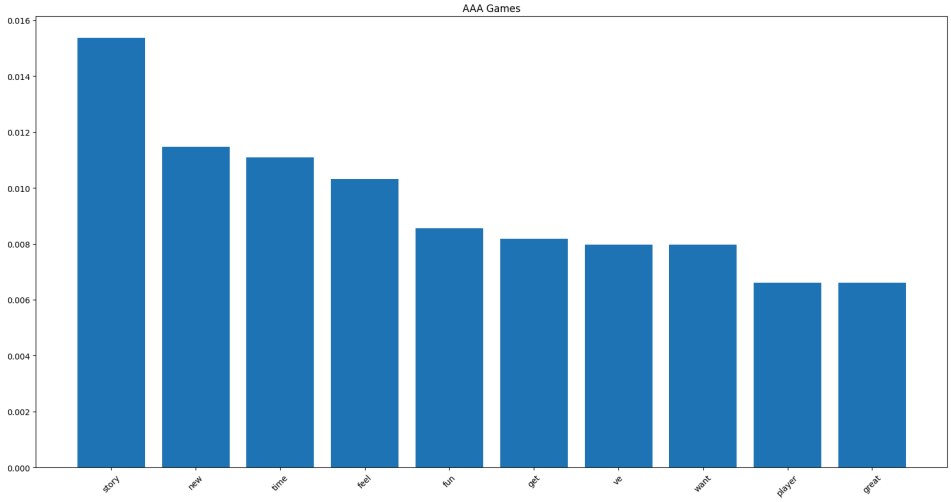
original visualization we created just showed the amount of positive, negative, and neutral games by developer type. Whether a game was considered positive, negative or neutral was based on what sentiment appeared the most among the 20 reviews for the game. This initial visualization was very basic and a little bit hard to understand unless you were working with our dataset, which was not ideal. To fix this we added more overall sentiments for the game, so instead of the sentiment just being whatever sentiment appeared most now the overall sentiment for a game can be overwhelmingly positive, mostly positive, mixed, etc. based on the ratio of the sentiments. To make the visualization easier to understand for each overall sentiment we included separate bars showing the count for each developer type. Looking at the finished visualization we can see that indie games were consistently highly represented in the positive sentiments, being the most represented in the overwhelmingly positive sentiment or the most positive sentiment, although this could be due to the fact that indie games were most represented in our dataset.



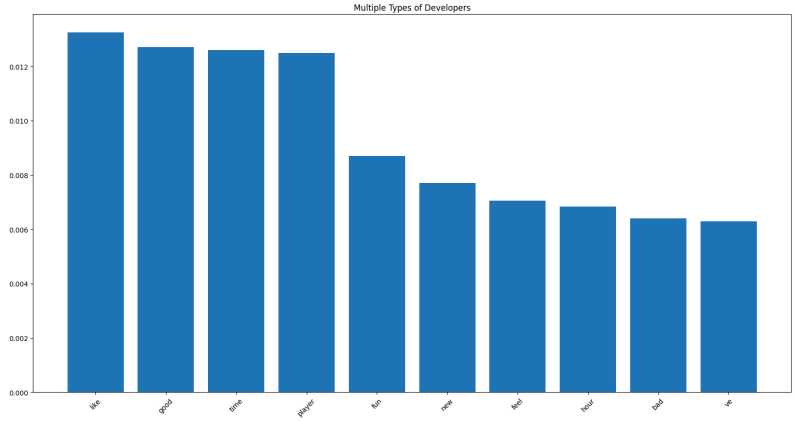
We then used a bag of words analysis to look for themes that re-occurred in reviews for games with Indie developers. After lemmatizing the reviews, we looked at the ten most common words in all reviews, and then the ten most common words in games with indie, multiple, and AA developers. For all games, the ten most common words were: time, feel, player, new, fun, ve, get , want, hour, and character.



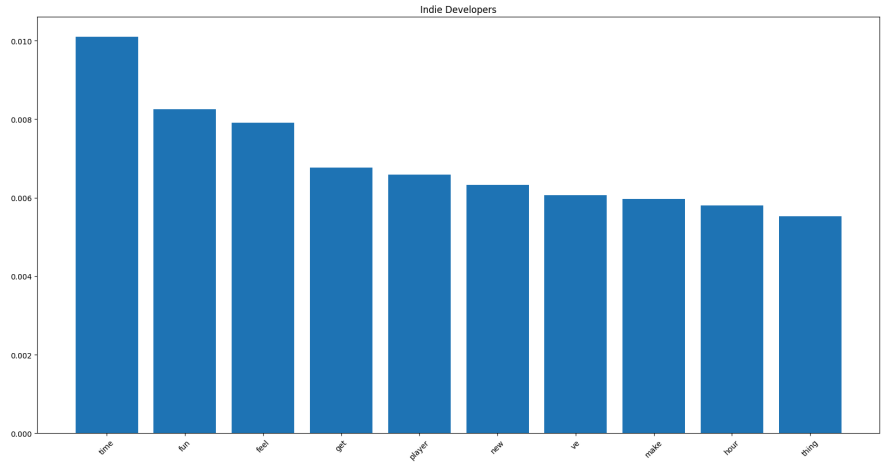
For AAA games, the ten most common words were: story, new, time, feel, fun, get, ve, want, player, and great.



For games with multiple kinds of developers, the ten most common words were like, good, time, player, fun, new, feel, hour, bad, and ve.



For games with Indie developers, the ten most common words were time, fun, feel, get, player, new, ve make, hour, and thing.

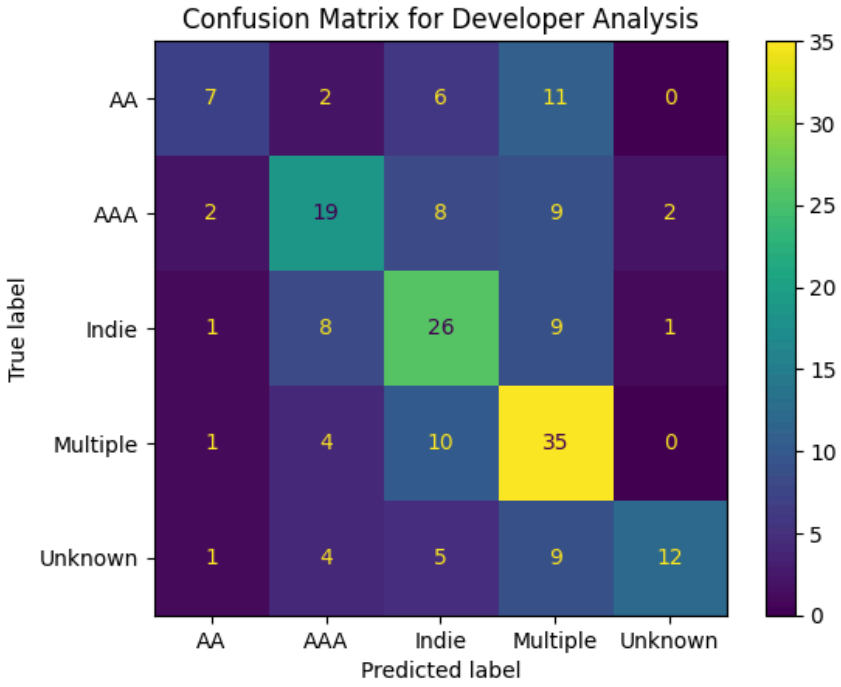


Some words, like time, feel, fun, new, and player appear in each list, indicating that regardless of type of developer, games are judged on their level of engagement and the player experience. Interestingly, the word "story" only appears in the top ten words for AAA games, suggesting that part of the appeal of larger development companies is that they are able to craft more interesting narratives and storylines into their games. The word "bad" also only appears in games with multiple types of developers involved, suggesting that collaborations across different types of development studios have not necessarily been well-received. "Make" also only appears in the top ten words for indie games, perhaps reflecting the unique mechanics associated with indie games.

Next, we trained two logistic regression models, one with the sentiment column as the y variable, and another with the simplified_category column as the y variable. By using a Wordcloud to visualize the logistic regression model that used the sentiment column as the y variable, one can see that technical issues are a big driver for negative reviews, with the inclusion of words like "crash," "unplayable," "fix," and "shit."

For the logistic regression where the developer type was used as the sentiment label, a confusion matrix indicated that the model struggled to differentiate between types of games, especially games by 'AA' developers, who it only identified correctly 7 out of 26 times. It was most successful at guessing

which games were made by multiple types of developers, with a 70% success rate (35/50). This further reaffirms the weakness of indie as a fixed category.



We then did a Gensim analysis of the reviews on indie games, and generated five topics. The first topic, which included words like “fun,” “way,” “time,” “feel,” and “player,” likely reflects user experience in shaping reviews for games. The second topic, which had words like “time,” “make,” “vr,” and “gameplay,” likely reflects reviews that focus heavily on game mechanics and gameplay. The third topic, which included words like “fun,” “feel,” “community,” and “experience,” may highlight the ways in which games are engaging to players either through features or community. The fourth topic, which includes “experience,” “new,” “recommend,” “love,” and “great” most likely reflects positive feedback. The last topic, which has the words “feel,” “time,” “hour,” and “update” most likely reflects reviews from players who have played for longer periods of time.

Conclusion & Reflection

Our research demonstrated that while indie games may be a point of interest in the zeitgeist, with indie compromising a large and growing category of games on Steam, and indie developers dominating the top category on Twitch, its status as a category collapses when attempting to define a game as indie. In part, because the creation of video games often involves large teams and multiple studios, and an indie developer does not guarantee an indie publisher. Games do not fall neatly into categories of indie and AAA and there is significant collaboration between indie and AAA studios. We also found there to be

significant overlap in how indie versus AAA games were discussed in user reviews. Reviews for games from all kinds of developers focused heavily on player experience, with the words “time,” “feel,” “fun,” “new,” and “player” represented heavily in reviews for games from all types of developer studios.

Overall, we found that people are drawn to games because of playability, affordability, and user engagement. Negative sentiment for games was largely correlated with reviews that mentioned bugs, crashes, and other technical issues. Affordability also played a large role in which games were popular. The most popular games on Twitch tended to be on average $20-$25, with free games representing 25% of games being streamed. One review highlighted the importance of playability and affordability to players:

> *sorry guys but after a whole year its still not worth it i know i should probably spend more time with this but its genuinely so fucking boring i cant get behind you guys if theres no content or consideration for the average player not paying for every piece of content payday already had huge content drops throughout its first year cycle no crimefest no nothing horribly optimised still both in game and networking and a ui fix wont bring the game around you guys seriously dropped the ball on this*

Users want games to be worth the price paid for the experience. User engagement also played a large role in what drew users to games. Reviews for games across developer types included words like “fun,” “want,” and “character,” suggesting engagement in the world of the game itself. In our logistic regression analysis, we found that reviews that correlated with positive or neutral sentiment included words like “pretty” and “look,” suggesting that engaging graphics are sought after features in games.

Indie games are stereotyped and associated with outsider status and experimental game design. Frustration with the perceived greed of large companies and unique and engaging design may drive the emergence of “mainstream-indie” as a genre, with users perceiving indie games like Sifu, Stardew Valley, and Hades, which are all at the top of the “Indie category” on Steam overwhelmingly positively. Yet what makes Steam categorize a game as Indie is opaque, and our own findings demonstrate that it is not tangible markers associated with indie developers that drive these games popularity, but rather appeal based in intangible things—playability, design, and engagement.

Throughout this project we faced a lot of different challenges as we worked toward completing the project. The main challenge we ran into was where to collect our data from, which led to other challenges. Early on we decided that we wanted to use reviews for games to help us answer our research questions and to do that we wanted to use Steam as it is the largest digital distribution platform for PC games, which would hopefully give us a more representative dataset. Unfortunately the Steam API doesn’t have an endpoint to get reviews for games and we wanted to look at the top games for our dataset, but again Steam did not have an endpoint to get the top games by downloads, playtime, or anything else.

To overcome this challenge we had to look elsewhere. In order to get the list of top games for our dataset we used the Twitch API instead, which had an endpoint that provided the top games streamed on the platform. To get the reviews we used the Steamworks API, which is meant for game developers, but it does have an endpoint that allows you to get the reviews for a game on Steam. We did overcome this challenge, but it did lead to a lot of other challenges when combining these datasets later on, such as some games from Twitch not being available on Steam, Steam and Twitch game ids not matching, and more. In the end even if we didn't always come up with the best solution we did overcome all of the challenges in our way.

Future directions you could go with this project is just expanding it to include more than just the data from Steam and Twitch. When looking at our dataset we can see that even though we did try to make it representative of all players in the end there definitely could be slight biases toward PC players. This is because we got our game reviews from Steam, which mainly distributes PC games, and our list of top games came from Twitch, where most of the top gaming streamers play on PC. This bias could reduce the reliability of our results when applied to all players since not all players are involved in the Steam or Twitch ecosystem. That being said this also leads into another future direction, which compares how indie games perform on consoles, mobile devices, and on PC since this could be a possible big factor in determining the popularity of indie games.

Works Cited

"Behind the Game: Exploring the Twitch Streaming Platform | IEEE Conference Publication | IEEE Xplore." n.d.-a. https://ieeexplore.ieee.org/document/7382994.

Crawford, G., Gosling, V. K., & Light, B. (2011). The Social and Cultural Significance of Online Gaming. In *Online Gaming in Context*. Routledge.

Cunha, L. R., Pessa, A. A. B., & Mendes, R. S. (2024). Shape patterns in popularity series of video games. Chaos, Solitons & Fractals, 185, 115081. https://doi.org/10.1016/j.chaos.2024.115081

Deng, J., Cuadrado, F., Tyson, G., & Uhlig, S. (2015). Behind the game: Exploring the twitch streaming platform. *2015 International Workshop on Network and Systems Support for Games (NetGames)*, 1–6. https://doi.org/10.1109/NetGames.2015.7382994

Goh, Edward, Omar Al-Tabbaa, and Zaheer Khan. 2023. "Unravelling the Complexity of the Video Game Industry: An Integrative Framework and Future Research Directions." *Telematics and Informatics Reports* 12 (December):100100. https://doi.org/10.1016/j.teler.2023.100100.

"Indie." 2021. October 27, 2021. https://store.steampowered.com/tags/en/Indie/?snr=1_5_9__409.

Jenson, Jennifer, Suzanne de Castell, and Stephanie Fisher. 2007. "Girls Playing Games: Rethinking Stereotypes." In *Proceedings of the 2007 Conference on Future Play*, 9–16. Future Play '07. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1328202.1328205.

Latorre, Óliver Pérez. 2016. "Indie or Mainstream? Tensions and Nuances between the Alternative and the Mainstream in Indie Games." *Anàlisi*, June, 15–30. https://doi.org/10.7238/a.v0i54.2818.

Lin, D., Bezemer, C.-P., Zou, Y., & Hassan, A. E. (2018). An empirical study of game reviews on the Steam Platform. Empirical Software Engineering, 24(1), 170–207. https://doi.org/10.1007/s10664-018-9627-4

Mendes, L. O., Cunha, L. R., & Mendes, R. S. (2022). Popularity of video games and collective memory. Entropy, 24(7), 860. https://doi.org/10.3390/e24070860

Pereira, L. S., & Bernardes, M. M. (2018). Aspects of independent game production. Computers in Entertainment, 16(4), 1–16. https://doi.org/10.1145/3276322

Muller, B. (2022, March 2). Bert 101 - state of the art NLP model explained. BERT 101 - State Of The Art NLP Model Explained. https://huggingface.co/blog/bert-101