Video Game Analysis: Trends and Insights

Daniel Ethridge[1] Atharva Patil[2], Deep Shukhla[3], Sujith Battu[4]

University of Colorado, Boulder

Author Emails
[1]*Corresponding author: daet2304@colorado.edu*
[2]*atpa5127@colorado.edu*
[3]*desh3965@colorado.edu*
[4]*suba2075@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 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.

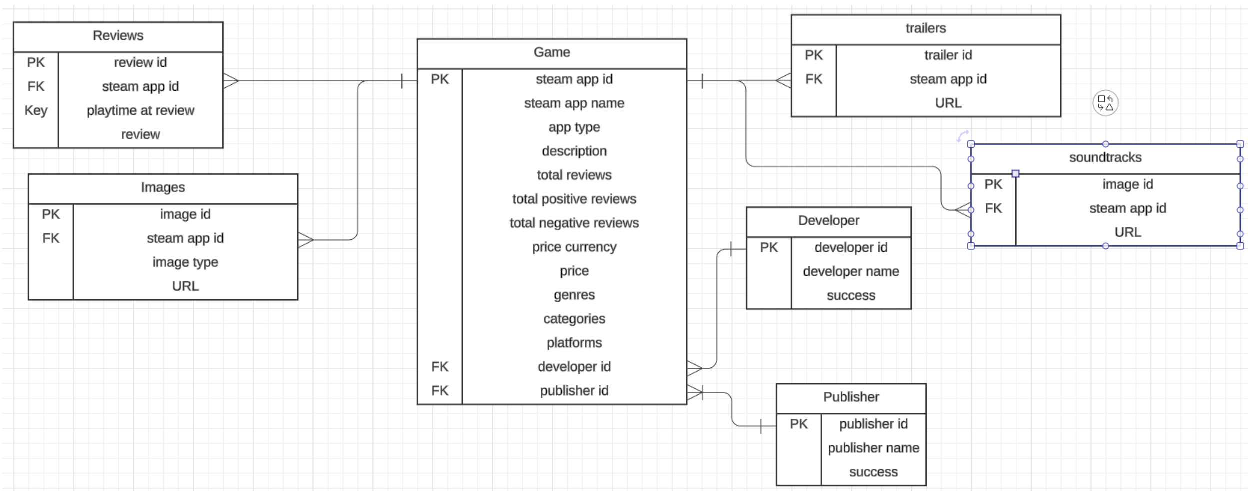
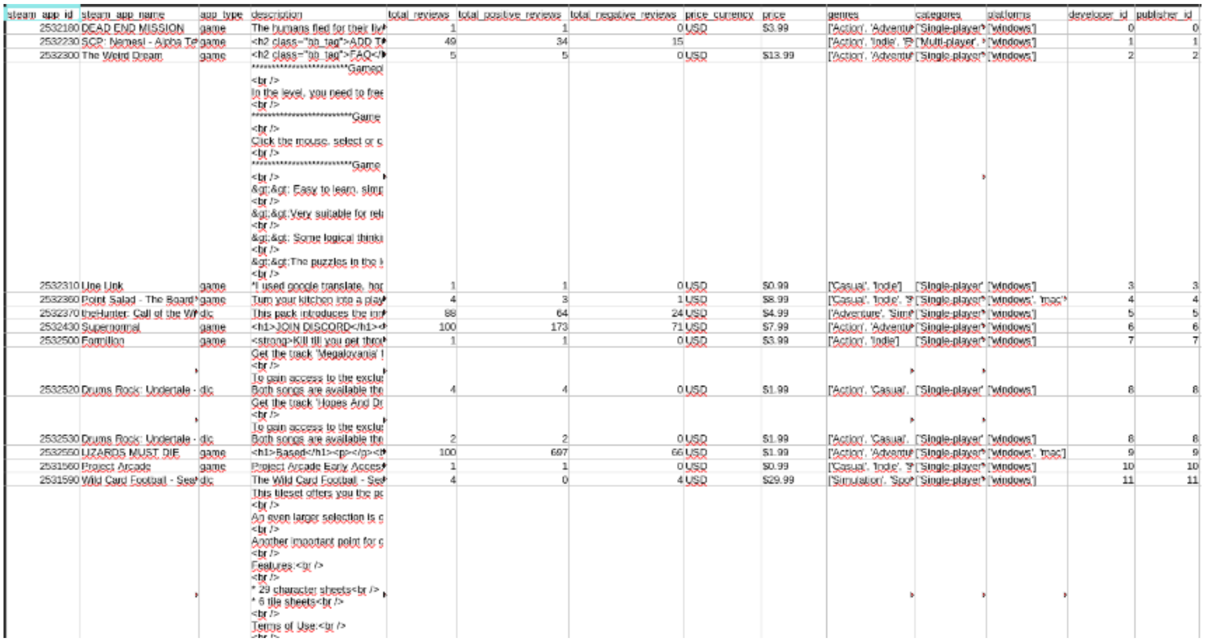


FIGURE 2. Data Schema

Fig.3 shows the exact data file we collected.



steam_app_id	steam_app_name	app_type	description	total_reviews	total_positive_reviews	total_negative_reviews	price_currency	price	genres	categories	platforms	developer_id	publisher_id
2532180	DEAD END MISSION	game	The humans fled for their liv	1	1	0	USD	$3.99	['Action', 'Adventu	['Single-player'	['windows']	0	0
2532230	SCP: Nemesi - Alpha Te	game	<h2 class="bb_tag">ADD T	49	34	15			['Action', 'Indie', 'R	['Multi-player', '	['windows']	1	1
2532300	The Weird Dream	game	<h2 class="bb_tag">FAQ</	5	5	0	USD	$13.99	['Action', 'Adventu	['Single-player'	['windows']	2	2
2532310	Line Link	game	************************Gamepl In the level, you need to free ************************Game Click the mouse, select or c ************************Game >&gt: Easy to learn, simp >&gt:Very suitable for rel >&gt: Some logical thinki >&gt:The puzzles in the l *I used google translate, hor	1	1	0	USD	$0.99	['Casual', 'Indie']	['Single-player'	['windows']	3	3
2532360	Point Salad - The Board	game	Turn your kitchen into a play	4	3	1	USD	$8.99	['Casual', 'Indie', 'S	['Single-player'	['windows', 'mac'	4	4
2532370	theHunter: Call of the W	dlc	This pack introduces the im	88	64	24	USD	$4.99	['Adventure', 'Sim	['Single-player'	['windows']	5	5
2532430	Supernormal	game	JOIN DISCORD <	100	173	71	USD	$7.99	['Action', 'Adventu	['Single-player'	['windows']	6	6
2532500	Formilion	game	<strong>Kill till you get thro	1	1	0	USD	$3.99	['Action', 'Indie']	['Single-player'	['windows']	7	7
2532520	Drums Rock: Undertale -	dlc	Get the track 'Megalovania' ! To gain access to the exclu Both songs are available thr	4	4	0	USD	$1.99	['Action', 'Casual',	['Single-player'	['windows']	8	8
2532530	Drums Rock: Undertale -	dlc	Get the track 'Hopes And Dr To gain access to the exclu Both songs are available thr	2	2	0	USD	$1.99	['Action', 'Casual',	['Single-player'	['windows']	8	8
2532550	LIZARDS MUST DIE	game	Based <t	100	697	66	USD	$1.99	['Action', 'Adventu	['Single-player'	['windows', 'mac']	9	9
2531560	Project Arcade	game	Project Arcade Early Access	1	1	0	USD	$0.99	['Casual', 'Indie', 'S	['Single-player'	['windows']	10	10
2531590	Wild Card Football - Sea	dlc	The Wild Card Football - Sea	4	0	4	USD	$29.99	['Simulation', 'Spo	['Single-player'	['windows']	11	11
			This tileset offers you the po An even larger selection is c Another important point for c Features: * 29 character sheets * 6 tile sheets Terms of Use: 										

FIGURE 3. Data file

Data Formatting

After collection, the data was organized into structured formats, primarily using CSV files. Each dataset included columns for game titles, release dates, sales figures, ratings, genres, and player demographics.

Data Integration

We tried to integrate multiple datasets but couldn't manage to do it at this point of the project timeline. But sure we were going to merged using common identifiers (e.g., game titles) to create a comprehensive dataset that includes sales, ratings, and demographic information.

Data Cleaning

Handling Missing Values: Utilized imputation methods for handling missing numerical data and excluded rows with significant missing values.

Removing redundants:
Recognized and eliminated repeated entries in order to maintain data accuracy.

Standardizing Formats:
Guaranteed uniformity across date formats, genre names, and other categories.

Outlier Detection:
Identified outliers in sales data and ratings using Z-scores and IQR methods, and assessed their impact on overall analysis.

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:

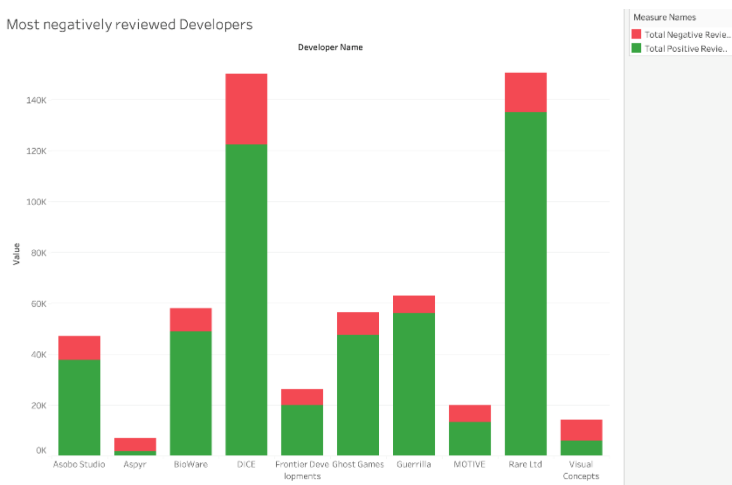


FIGURE 4. Negatively reviewed reviews 10 developers

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.

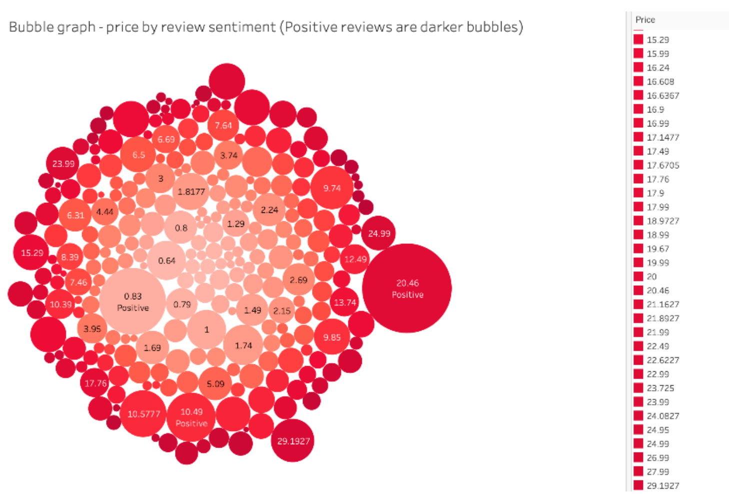


FIGURE 6. price vs reviews

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.

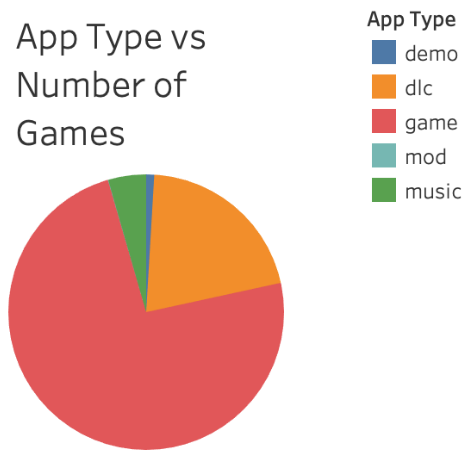


FIGURE 7. Game types

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.

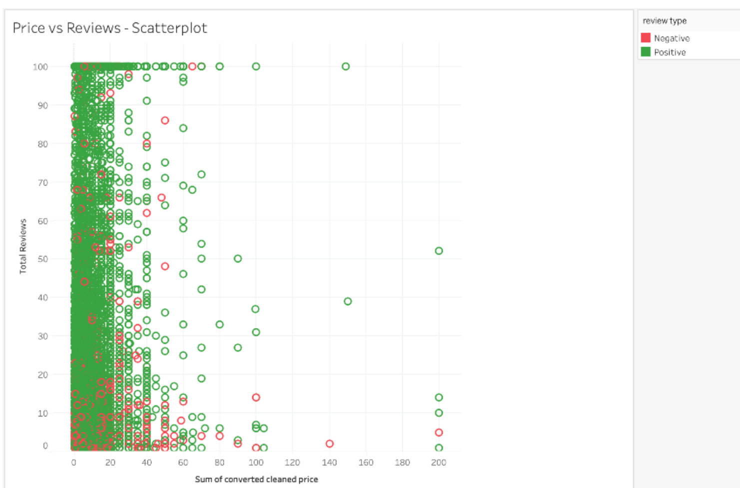


FIGURE 8. Price vs Reviews

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.

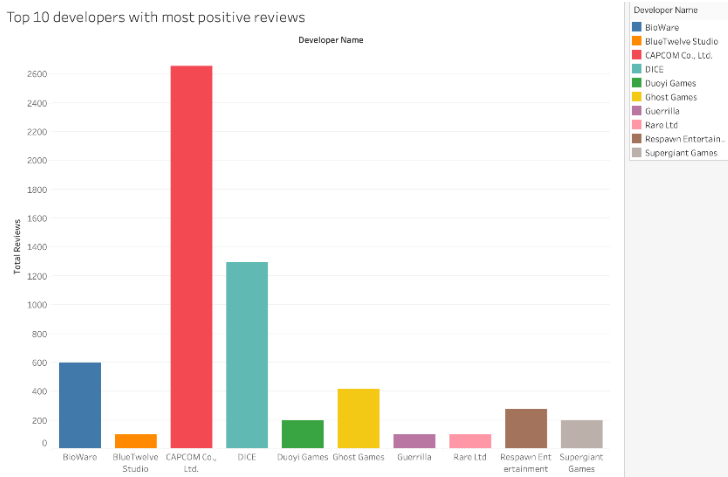


FIGURE 9. Most positively reviewed 10 developers

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.

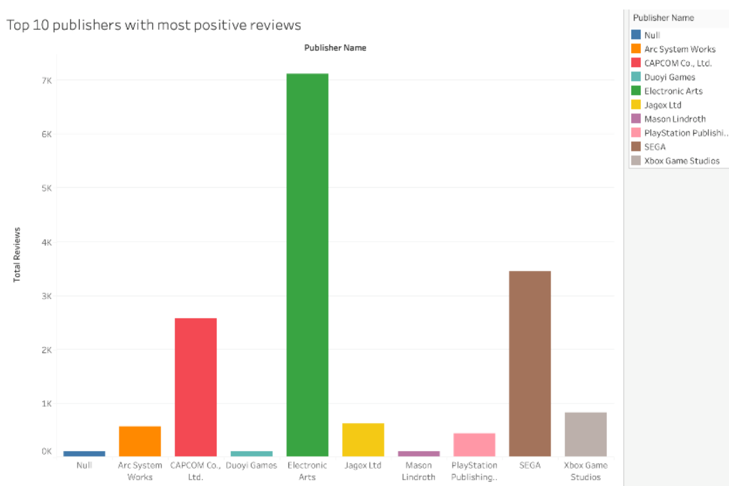


FIGURE 10. Most positively reviewed 10 developers

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.

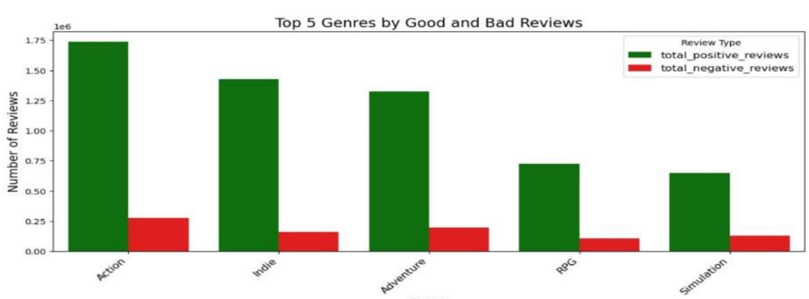


FIGURE 11. Genre analysis (Good vs Bad reviews

[Fig.11](#fig11) 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.

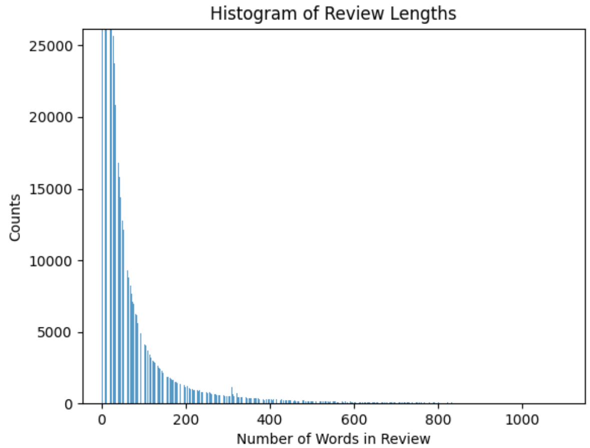


FIGURE 12. Number of words in reviews

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.

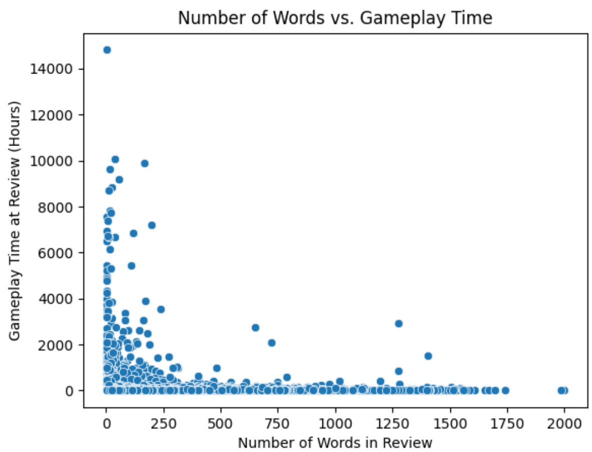


FIGURE 13. Number of words vs Gametime

The Fig.13 shows number of words in a review plottted 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.

MODEL IMPLEMENTATION

Variables in the dataset are, reviews, price, developer, publisher, genre, currency, platforms, images, number of reviews, positive reviews, negative reviews, brightness, contrast, saturation and hue.

1. CLIP MODEL: CLIP is a multimodal model, proposed by OpenAI, that aligns textual and image embeddings into a joint space. It was trained to predict which caption fits best for an image among candidates, allowing it to learn associations of visual and textual data. Trained on diverse image-text pairs, CLIP can execute tasks such as image classification, retrieval, and zero-shot learning without fine-tuning specifically for individual tasks. Using a dual encoder architecture, it processes images through a vision model-e.g., ResNet or ViT-and processes text using a transformer to produce embeddings to be compared. Such flexibility and generalization make CLIP a very powerful model for different visual-linguistic applications.

WHY WE USED IT:

We utilized the CLIP model in our project to examine and find associations between game images and their textual descriptions. The ability of this model to create a common embedding space for text and images allowed us to efficiently perform tasks like image classification and retrieval, helping uncover patterns, understand context, and improve insight into the relationships between visuals and game descriptions.

DATA CLEANING:

How data looked before cleaning:

steam_app_id	steam_app_name	app_type	description
2532180	DEAD END MISSION	game	The humans fled for their lives! Now only aliens remain in the city. The alien queen lays eggs, and the aliens become more numerous and slowly mutate over t
2532230	SCP: Nemesi - Alpha Testing	game	<h2 class="bb_tag">ADD TO YOUR WISHLIST!</h2><img class="bb_img" src="https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532
2532300	The Weird Dream	game	<h2 class="bb_tag">FAQ</h2>Q: My controller isn't responding? A: Please confirm the controller is connected to your pc, or re-plug i
2532310	Line Link	game	************************Gameplay************************ In the level, you need to freely choose a marker point as your starting point, and plan a line based on this starting point. This line needs to contain all the mar ************************Game Control************************ Click the mouse, select or connect the direction with the left mouse button, and cancel the operation with the right button ************************Game Features************************
2532360	Point Salad - The Board Game	game	Turn your kitchen into a playground with Point Salad - the delightful digital take on a fast-paced, card-drafting, family board game! Craft the ultimate salad by
2532370	theHunter: Call of the Wild™ - Ambusher Pack	dlc	This pack introduces the innovative FOXPRO X360 electronic caller, the powerful Moradi Model 1894 lever-action rifle, and the sleek Stenberg Takedown Rec
2532430	Supernormal	game	<h1>JOIN DISCORD</h1><p> <img class="bb_img" src="https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532430/extras/G
2532500	Formilion	game	Kill till you get through. In Formilion, you will have to unravel the cause of a missing person and go through a challenging journey uncoverin
2532520	Drums Rock: Undertale - 'MEGALOVANIA'	dlc	Get the track 'Megalovania' from the Undertale game! This track is part of the Undertale DLC.
2532530	Drums Rock: Undertale - 'Hopes And Dreams'	dlc	Get the track 'Hopes And Dreams' from the Undertale game! This track is part of the Undertale DLC.
2532550	LIZARDS MUST DIE	game	Based VIKTOR - BASED CHARACTER About the Game <img class="bb_img" src="https://shared.
2531580	Project Arcade	game	Project Arcade Early Access release! Currently, it has a Memory Cards game, but the plan is to add more in the future.<h2 class="bb_tag">Memory
2531590	Wild Card Football - Season Pass	dlc	The Wild Card Football - Season Pass includes a day one Blazing Bones outfit unlock and 3 future DLC packs with legacy players and cosmetic outfits all at
2531660	RPG Maker MV - Winlu Fantasy Tileset - Overworld	dlc	This tileset offers you the possibility to create overworld or map maps. Numerous biomes such as different grasslands, deserts, wastelands, black sand dese An even larger selection is offered by the tileset with normal sprites such as withered trees, white trees, cactuses, palms, and flower blankets. In addition, all t Another important point for creating overworld maps are mountains. These are available in numerous sizes and colours from 2*2 tiles to 7*8 tiles. The tileset a Features:

After cleaning:

	steam_app_id	cleaned_description
22135	1097960	br game party skills unique team idle gems cla...
22714	1084860	br risk system deluxe content original light e...
14642	1266070	stage healthy box vr fitness game people willi...
21109	1120461	lord zedd character power rangers battle grid
1045	2484670	chess game games remix world rules new play sh...
20667	1130990	br roman romans evil britons level welcome sea...
2714	2417700	unique steam extras strong malice platformer a...
16973	1212690	borg instructors mountain faulty apprentice dl...
5312	2325880	game heroes players battle deck zodiac abiliti...
10495	1375270	maze steam extras shadow reborn ego shape expe...
9939	1389830	electronic components power steam extras robot...
15236	1254700	format
2533	2425050	steam extras person night fang mysterious new ...
14438	1270620	chemistry steam extras molecules simulation re...
7047	2264940	dlc valentine xi churchill
14618	1266440	br treatment dungeon health way tests obstacle...
13771	1285610	people horror game city person thrill feelings...
24033	731180	coffee run game upgrade graphics ui gameplay c...
4059	2370540	love note steam extras notebook japanese roman...
7346	2254780	steam game mini keys games points characters k...
23391	1070400	br artefacts player mode maldrin journey game ...
20063	1145360	supergiant god like steam extras strong underw...
15538	1246390	br steve human silent hill game average day in...
20807	1127620	picture bylo quick horror game experience back...
10436	1376470	bear ears cave lantern footsteps hunted large ...
18571	1178870	br characters battlers rpg maker hero bust pac...
18092	1189060	steam extras game explosions floor weapons str...
7819	2239140	grappling players physics hook movement active...
20363	1138260	steam extras level action strategy battlefield...

This would serve to clean game descriptions of noise like irrelevant words and characters, hence making the text data concise and meaningful. The preprocessing step helps in improving the quality of the textual embeddings that are generated by the CLIP model, therefore improving its ability to relate game descriptions with their corresponding images accurately. Cleaned data reduces computational complexity and improves overall model performance.

Cleaning was also done for image dataset, such as removing 404 error URLs and only filtering screenshots for more consistent training.

MODEL EVALUATION (using similarity scores):

```
Similarity scores:
[[0.1638 0.2218 0.1312 ... 0.169  0.1782 0.1259]
 [0.1913 0.2017 0.1605 ... 0.1886 0.2017 0.1562]
 [0.22   0.232  0.2188 ... 0.2515 0.23   0.208 ]
 ...
 [0.2034 0.251  0.2646 ... 0.2292 0.2242 0.2021]
 [0.2034 0.251  0.2646 ... 0.2292 0.2242 0.2021]
 [0.2034 0.251  0.2646 ... 0.2292 0.2242 0.2021]]
```

The similarity scores indicate how well text embeddings (e.g., game descriptions) align with image embeddings in the CLIP model. Each value represents the similarity between a text-image pair where higher scores indicate stronger associations. Rows correspond to text embeddings, columns correspond to image embeddings, and together this allows for an evaluation of CLIP's effectiveness at multimodal alignment. The scores are cosine similarity, between 0 and 1. Since we are finding custom descriptions for completely new images, even a score above 0.2 is good.

OUTPUT:

The CLIP is model is very effective in describing what each game image is. In a matrix of 10 images, it was able to correctly describe each image. Thus proving CLIP is a effective image to text model. The examples of results provided show the capability of the CLIP model in analyzing and describing game screenshots in high detail. For each of the nine game images, CLIP identified relevant tags or genres such as "space combat simulation," "stealth game," "resource management game," and "visual novel," among others. It successfully aligned visual elements from the images with meaningful semantic labels, showcasing its capability to understand and interpret visual data in the gaming context. The CLIP is model is very effective in describing what each game image is. In a matrix of 10 images, it was able to correctly describe each image. Thus proving CLIP is a effective image to text model.

WHAT WE LEARNED:
This functionality has many applications in the analysis of games. CLIP enables automation in categorizing and tagging games with screenshots, reducing human effort and improving scalability for large gaming datasets. This can be useful for game developers and platforms in streamlining their metadata creation processes, therefore making it easier for end-users to discover or search for games. It also improves the quality of the recommendation systems by allowing them to suggest games with appealing visuals or gameplay themes that appeal to the users' likings.

Moreover, the ability of CLIP to extract meaningful insights from images will help in content moderation, ensuring the proper classification of the games. By pairing visual data with descriptive text, CLIP bridges the gap between visual representation and semantic understanding and is a powerful tool for indexing, filtering, and analysis of games efficiently and in a scalable manner.

2. **GRADIENT BOOSTING:** Gradient Boosting is a powerful machine learning technique that generates models sequentially, each new model correcting errors of the previous ones. A weak learner, usually decision trees, is combined into a strong predictive model through minimizing a loss function. Gradient Boosting has a wide range of applications in regression and classification problems, where it maintains accuracy and flexibility.

WHY WE USED IT:
We applied Gradient Boosting to the image data set for an effective modeling of dependencies between features including brightness, saturation, hue and contrast, impacting the visual characteristics. Its ability to deal with non-linear patterns and emphasize misclassified examples ensures accurate predictions. Gradient Boosting flexibility and robustness make this approach ideal for capturing subtle variations and improvement of image-based classification or regression results in general.

DATA CLEANING:
The dataset we created-

steam_app_id	image_url	brightness	contrast	saturation	dominant_hue
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/header.jpg?t=1702866992	32.97723963599600	61.234595150441	43.19052578361980	170
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/capsule_231x87.jpg?t=1702866992	46.12300343334830	81.78831055491310	27.78558988903820	170
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/extras/BESKRIVNING4.png	1.44405555555600	16.22203340619160	2.14661458333330	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/extras/BESKRIVNING1.png	30.54166815476190	68.07168243577570	68.92479017857140	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/extras/BESKRIVNING5.png	25.55749652777780	65.13446715684070	6.68501041666670	170
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/extras/BESKRIVNING3.png	37.21125595238100	44.55414775842530	79.11767857142860	170
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/extras/BESKRIVNING6.png	40.59772916866670	93.00265705692210	16.35094791666670	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_2c6dc8ae40359ce1ef0a21a94ebc7eeeff05e92f.1920x1080.jpg?t=1702866992	4.17516637731482	41.14737879054750	10.1420597029321	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_95baa8d4fb0a15945450c96b14ddcfc58b278c22.1920x1080.jpg?t=1702866992	28.63883551954730	38.20771616913970	69.22583705632720	170
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_b81cd3dc65c739424c5cc23b3aa17df36e3cc454.1920x1080.jpg?t=1702866992	7.16655076517490	21.93952475397360	16.77673514660490	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_349eb120de5444d7971658775bc3aa6e7da72129.1920x1080.jpg?t=1702866992	25.48999967849790	47.46920675190940	52.60092978395060	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_72daa0ee37ffecaaed290ff739103b917aa02ab.1920x1080.jpg?t=1702866992	2.97132298096708	27.10779105870650	5.17849344135803	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_f8802784eef225948b3b5530fd08a00d0e214cc4.1920x1080.jpg?t=1702866992	15.42449250900210	51.61181536658390	35.13280574845680	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_a06725dbf6c1232179f81df788eb6cfad9008b98.1920x1080.jpg?t=1702866992	35.86553546167700	46.9126016047621	75.98064429012350	170
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_2f40ad3013341d75fecf07dbf8451cadae785769.1920x1080.jpg?t=1702866992	4.71946871784979	30.85245820035470	5.96573543595679	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_c565f985398bebc2eca689b757430d67d05fadd4.1920x1080.jpg?t=1702866992	1.41883375128601	22.37126880852940	2.06995177469136	0
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_bf3676a2a4db8ef2fc6a33fd4918ed97c9969fd3.1920x1080.jpg?t=1702866992	27.59357687114200	46.01344948186440	67.80410011574070	170
2532180	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532180/ss_fb4cab3c641cf2bc79fffd7e3bfcc856eb9b4328.1920x1080.jpg?t=1702866992	26.39344939557610	68.61886178824790	42.53997733410490	0
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/header.jpg?t=1705252659	80.59465453319850	81.86884514091230	29.49451971688570	0
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/capsule_231x87.jpg?t=1705252659	47.3431855500821	79.56920006421580	9.52570035328656	0
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/extras/steampage_wishlist.gif	44.40688311688310	28.11480557987010	32.13272727272730	151
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/extras/SCP_Nemesi_-_Year_1_roadmap_final_touches_Steam.png	42.17158825804360	54.71798208629220	5.40558684831019	170
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_b2db4bbda497188b104f6a6d7db6d26df94f23ce.1920x1080.jpg?t=1705252659	163.57769354423900	95.254394290151	12.73402536651240	0
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_7e6969f9b2f8f737babe8dce98be1c67b25ec20b.1920x1080.jpg?t=1705252659	70.7524017811214	17.75461327236450	13.45898051697530	146
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_5448a4978973734f0ff555461714018719646111.1920x1080.jpg?t=1705252659	128.17139065715000	69.0286855542233	10.65032986111110	29
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_0e1a4969e1ae996561618fabb3f19919819f3e11.1920x1080.jpg?t=1705252659	90.32507716049380	46.9634908586714	7.91615933641975	146
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_07a1101c7448a372971086f935e6c7af703a1363.1920x1080.jpg?t=1705252659	128.08000369727400	55.97917117465220	0.89925202546295	127
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_80b592fccfef0b0f4946fdc484ee2fc36cd42958.1920x1080.jpg?t=1705252659	12.90215587129630	28.11848220238750	7.91594569830247	42
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_c8f875e8dc7d1d94ddfb6f0888089fa09461e1f6.1920x1080.jpg?t=1705252659	41.1475898598251	67.95644477395900	47.12513647762350	0
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_efc7487f72648c7011c79999f93ab44db1281e44.1920x1080.jpg?t=1705252659	24.76990644290120	59.01512890526380	19.72720968364200	0
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_bd1b299de9ef94e9e4c2e3312ef952443338a5e4.1920x1080.jpg?t=1705252659	75.18153822659470	30.02768929563620	13.92270833333330	146
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_83af6a0a535e496c5d296b794e70d5d99d927343.1920x1080.jpg?t=1705252659	108.19512345679000	52.54250595032660	7.38229938271604	146
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_ba9d03a127b742f6241844df7a154557bfdab4cf.1920x1080.jpg?t=1705252659	70.29527263374490	31.42932425910740	18.71817660108020	152
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_5e5fef44d960970aac3fa7fb5577df30459f90fd.1920x1080.jpg?t=1705252659	15.26698768647120	42.92624276810920	15.14083670910490	0
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_3558ed2d05aca481ab2fc790aab397f360b68fa0.1920x1080.jpg?t=1705252659	24.88651507844650	56.02811277671470	26.63240499614200	0
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_9fb0f7a422d97f8202dc82cb4f77cc4c5cbdeb31.1920x1080.jpg?t=1705252659	16.7304793595679	48.79478912614250	11.78755208333330	0
2532230	https://shared.akamai.steamstatic.com/store_item_assets/steam/apps/2532230/ss_d0eb124128bd206c958dd4ec02cfb12f2f9fe1c1.1920x1080.jpg?t=1705252659	65.40706211419750	35.37122443314660	21.32321469907410	147

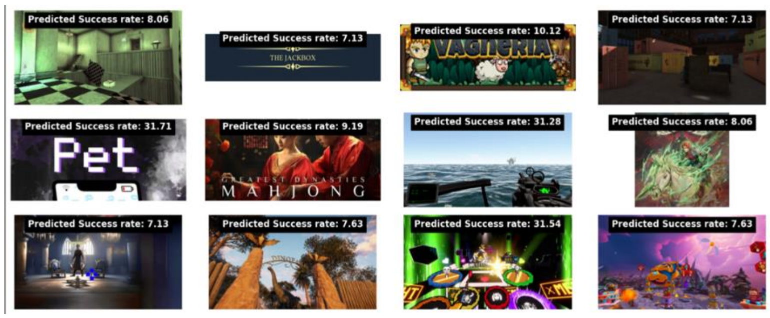
To train the gradient boosting model, we had to create a new dataset from the image URLs, which contained the app ids and image values such as brightness, contrast, saturation and hue. We made this so that we can find patterns in underlying aspects of the images and game success.

MODEL EVALUATION (using Mean Square Error):

```
Gradient Boosting Mean Squared Error: 130.68720097122053
```

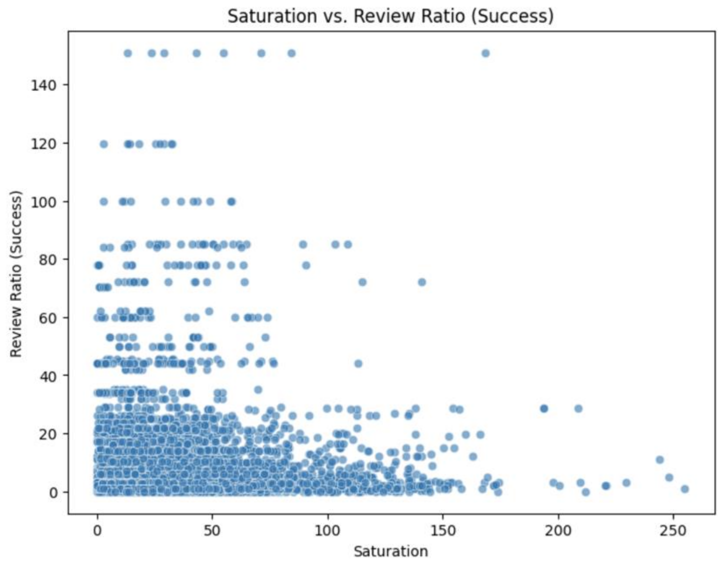
We use Mean Squared Error to calculate an average of the squared difference between predicted and actual values. The lower the MSE, the better the accuracy of the model. At a value of 130, the model performance is good, showing that the predictions are relatively near to the actual values; thus, effective learning takes place.

OUTPUT:

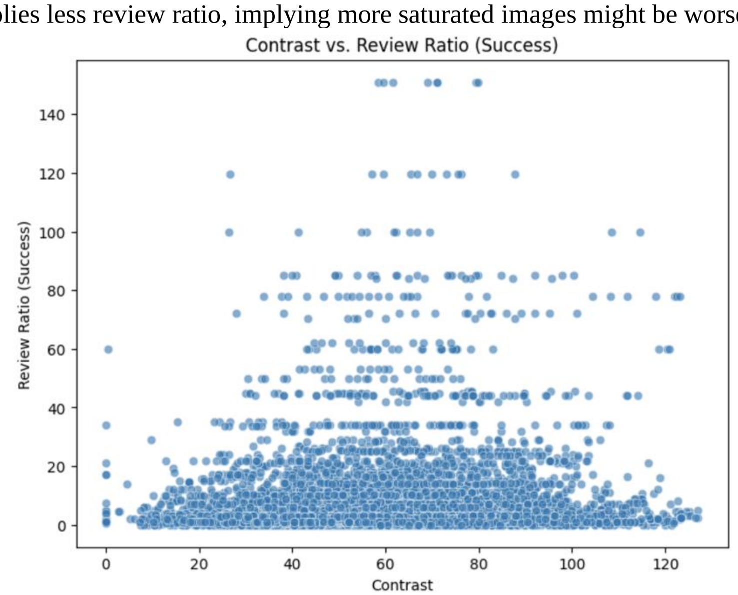


As we can see, the model predicts the success rate , which is the ratio of number of positive reviews to number of negative reviews. It rates higher success rate to games with more interesting game image screenshots and lesser success rates to dull or confusing game image screenshots.

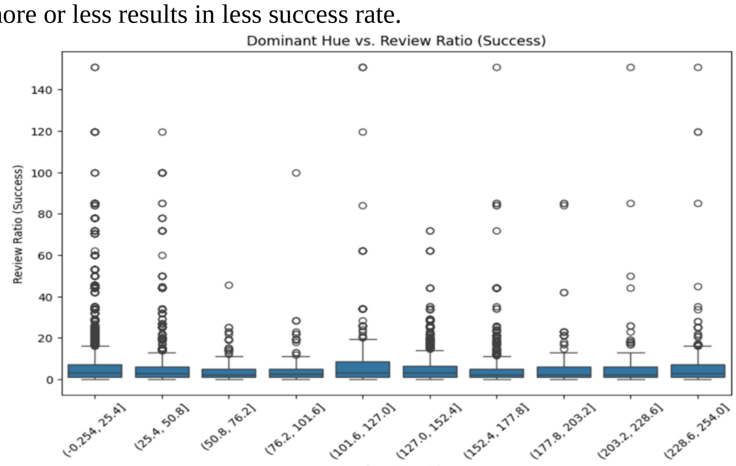
This is a correlation heatmap, showing correlation between various features, as we can see most features are unrelated except contrast and brightness which have some correlation.



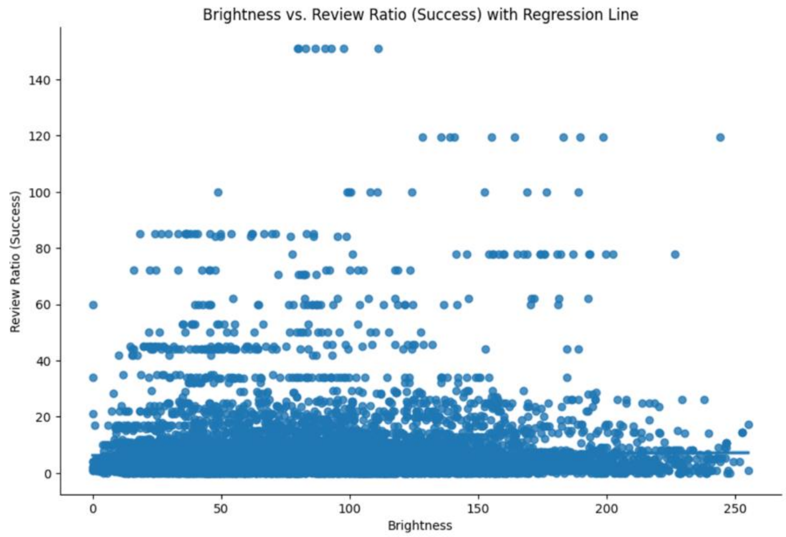
This scatterplot shows the relationship between review ratio (success rate) and saturation of images. As we can see the more saturation implies less review ratio, implying more saturated images might be worse for game success.



This scatterplot is the relation between contrast and success rate. As we can see a contrast between 50 and 80 is the sweet spot. Anything more or less results in less success rate.



This graph is the boxplot relationship between hue and success rate. It shows that all colors have similar success rate and there is no correlation between color/Hue and success of a game. It means that consumers don't prefer one color or the other. Hues between 101-127 (green-yellow) have slightly higher success rate, but nothing can be said with certainty.



This graph shows the relationship between brightness of game images and success rate of those games. As we can see, there is no clear relationship between brightness and game success.

WHAT WE LEARNED:
From this model, we extracted a few important points about how far image features such as brightness, contrast, and saturation will reflect success rates of games through reviews. Thus, the model predicted an optimal range of 50 to 80 for contrast for high success rates, while very low or high values of contrast will negatively affect the review score. Interestingly, saturation is negatively correlated with the success rate, which might indicate that highly saturated images could turn players off. On the other hand, hue did not correlate significantly with success, which would indicate that the consumer's preference is not strongly connected with specific colors. Also, no obvious influence of brightness could be found on the success rate, meaning it's not a determining factor. The model performs well with an MSE of 130, showing that it learns the relationship between the image features and game success quite effectively. This shows the importance of specific image attributes such as contrast in predicting game success.

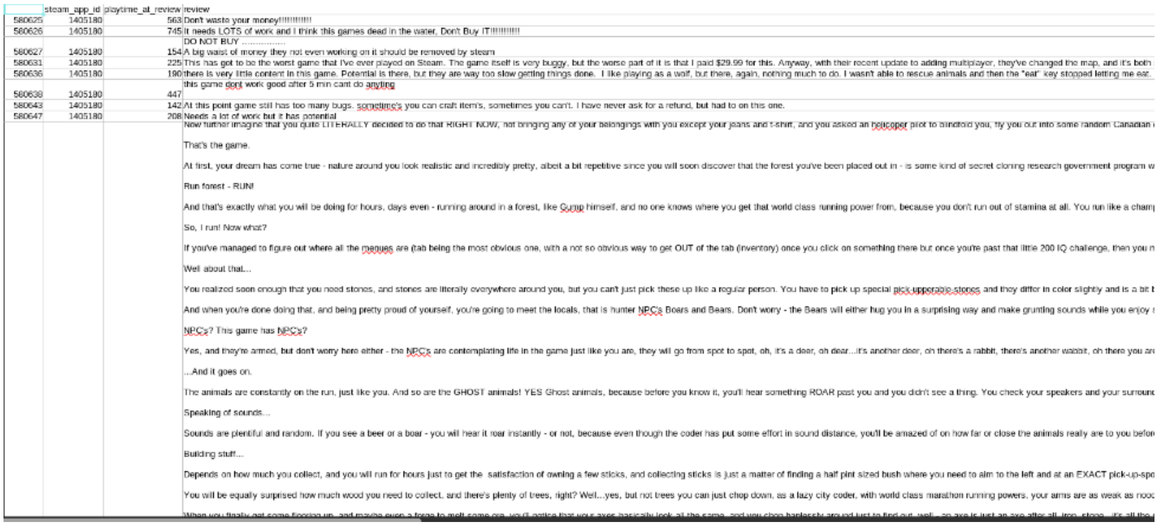
3. ASSOCIATION RULE MINING:

WHY WE USED IT**:**
We applied association rule mining to the reviews of the games. We looked at Steam Apps that were classified as a "game" (in other words, we didn't look at DLC or other non-game Steam Apps), and we looked at differences among the top 10 genres. We filtered game reviews by the ratio of positive reviews to negative reviews of a game. If a game had more positive reviews than negative, it was considered a positive and vice versa. We acknowledge that this is not the best method, especially considering Steam reviews are tagged as either positive or negative. A bug in code caused that review tag to not be collected. Grouping reviews by how positively overall the game they are associated with is the best approximation.

ARM requires transaction data. To create the transaction data, we first extracted the necessary reviews according to liked and disliked games and genre. We removed any line break characters, removed capital letters, punctuation, numbers, non-latin characters, duplicate words, and a set of words (such as "game", "ever", "it", "etc") that provide no meaning. Afterwards, we saved the cleaned words as basket data. The two images below show messy data (top) and clean data (right).

DATA PREPROCESSING:

Before:

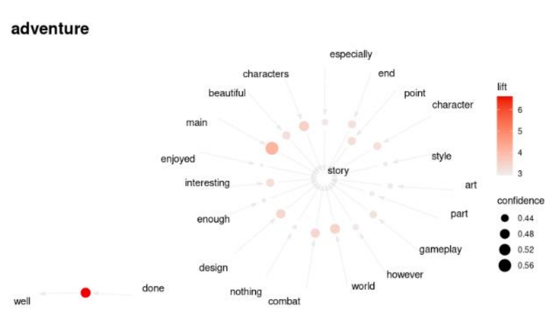


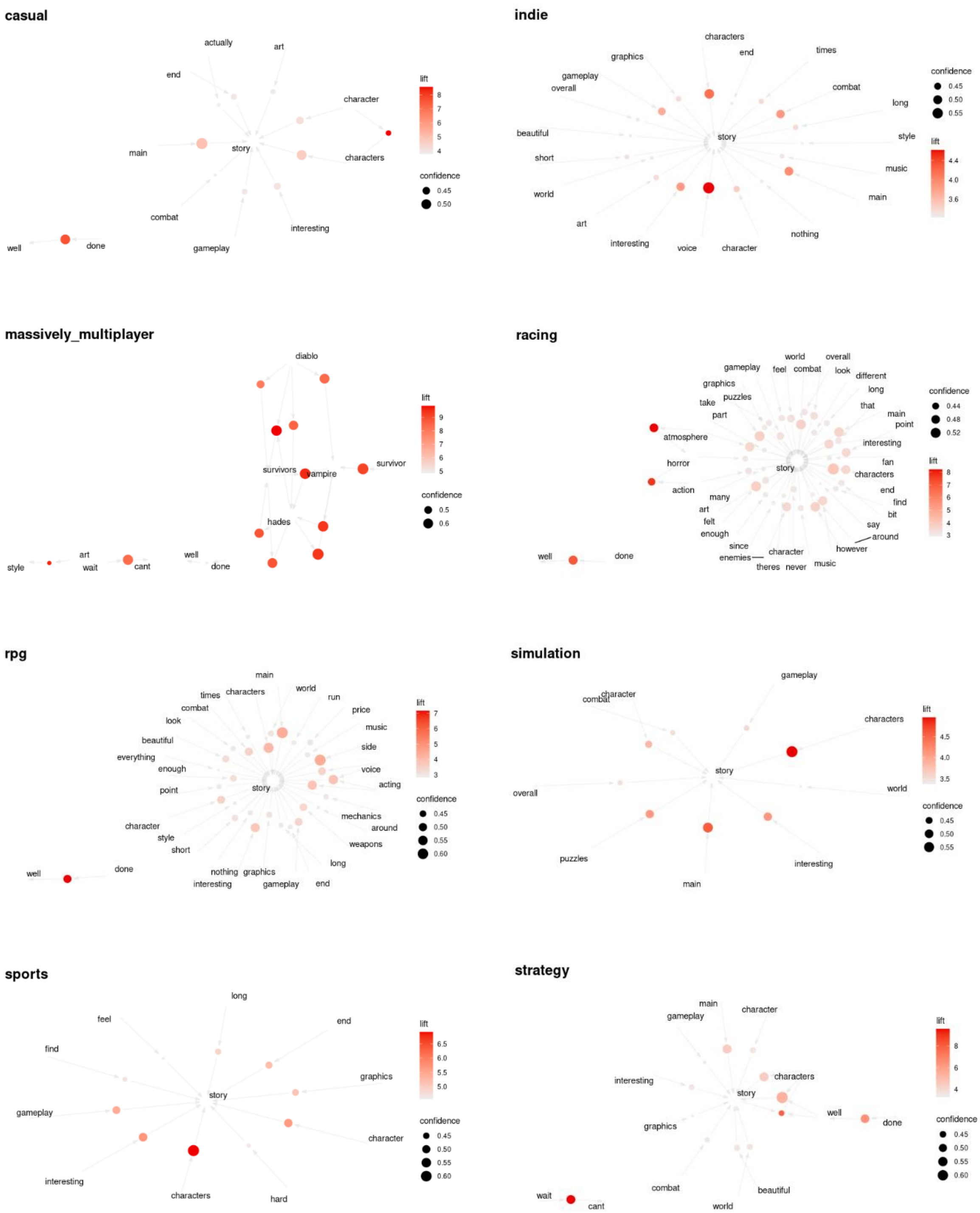
	steam_app_id	playtime_at_review	review
580625	1405180	563	Don't waste your money!!!!!!!!!!!!
580626	1405180	745	It needs LOTS of work and I think this games dead in the water, Don't Buy IT!!!!!!!!!!!
580627	1405180	154	DO NOT BUY .... A big waist of money they not even working on it should be removed by steam
580631	1405180	225	This has got to be the worst game that I've ever played on Steam. The game itself is very buggy, but the worse part of it is that I paid $29.99 for this. Anyway, with their recent update to adding multiplayer, they've changed the map, and it's both
580636	1405180	190	there is very little content in this game. Potential is there, but they are way too slow getting things done. I like playing as a wolf, but there, again, nothing much to do. I wasn't able to rescue animals and then the "eat" key stopped letting me eat.
580638	1405180	447	this game dont work good after 5 min cant do anyting
580643	1405180	142	At this point game still has too many bugs. sometime's you can craft item's, sometimes you can't. I have never ask for a refund, but had to on this one.
580647	1405180	208	Needs a lot of work but it has potential

After:

OUTPUT:

Game analysis:

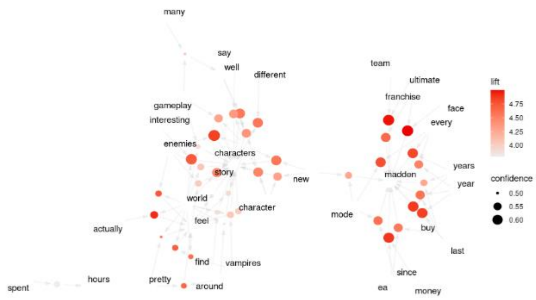




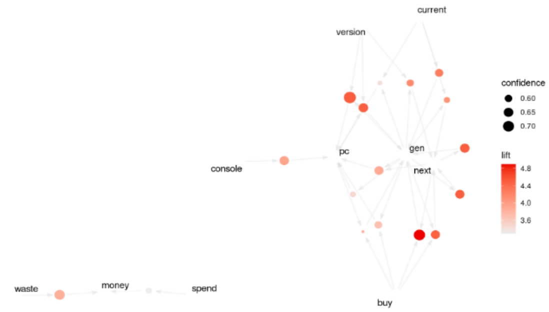
casual
actually
art
end
character
lift
main
story
characters
confidence
0.45
0.50
combat
interesting
well
done
gameplay
indie
characters
end
times
graphics
gameplay
overall
combat
long
confidence
0.45
0.50
0.55
beautiful
story
style
short
lift
music
world
main
art
interesting
voice
character
nothing
massively_multiplayer
diablo
lift
survivors
vampire
survivor
confidence
0.5
0.6
hades
style
art
wait
cant
well
done
racing
world
overall
gameplay
feel
combat
look
different
graphics
long
take
puzzles
that
part
main
point
atmosphere
interesting
horror
fan
story
characters
action
end
many
find
art
bit
felt
say
enough
around
since
character
however
enemies
theres
never
music
well
done
confidence
0.44
0.48
0.52
lift
rpg
main
times
characters
world
run
combat
price
look
music
beautiful
side
everything
enough
voice
acting
point
story
confidence
0.45
0.50
0.55
0.60
mechanics
around
character
style
weapons
short
long
well
done
nothing
graphics
interesting
gameplay
end
simulation
gameplay
combat
character
characters
lift
story
overall
world
confidence
0.45
0.50
0.55
puzzles
interesting
main
sports
long
feel
end
lift
find
graphics
story
gameplay
confidence
0.45
0.50
0.55
0.60
character
interesting
characters
hard
strategy
main
character
gameplay
lift
interesting
characters
story
well
done
graphics
confidence
0.45
0.50
0.55
0.60
combat
beautiful
wait
cant
world

Game reviews analysis:

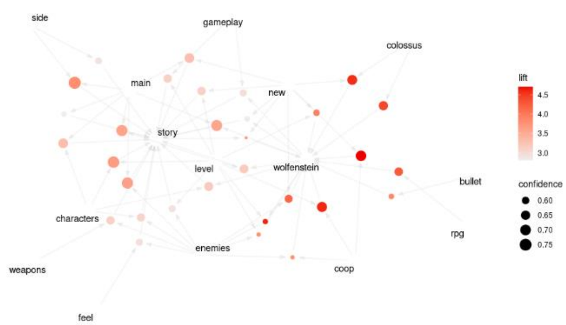
action



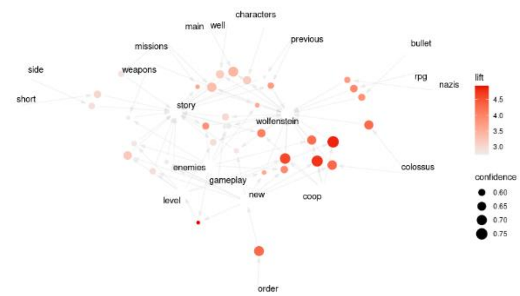
casual



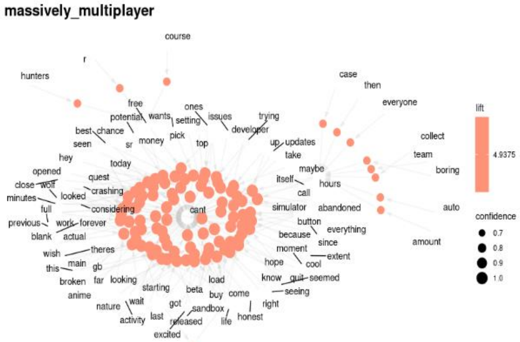
adventure



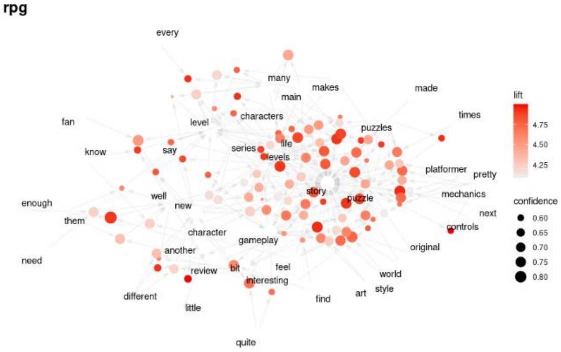
indie



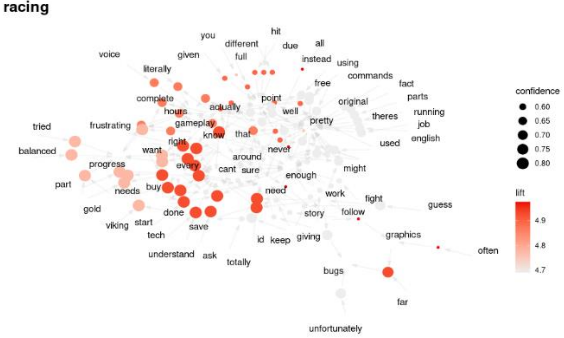
massively_multiplayer



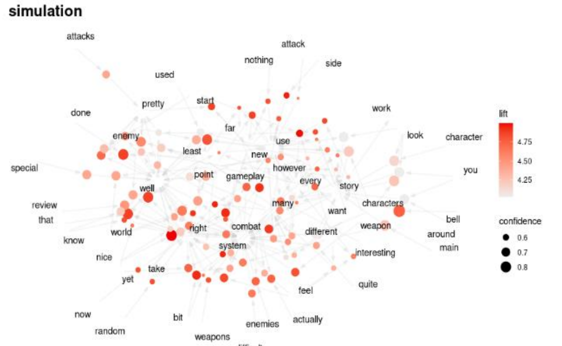
rpg

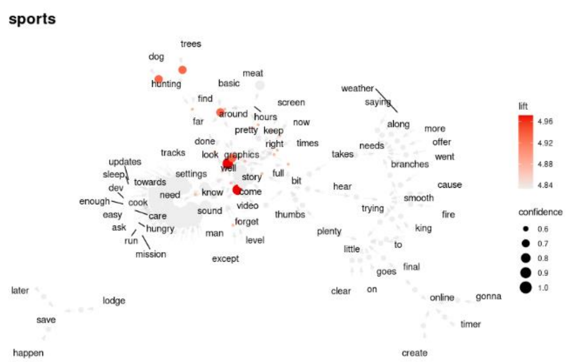


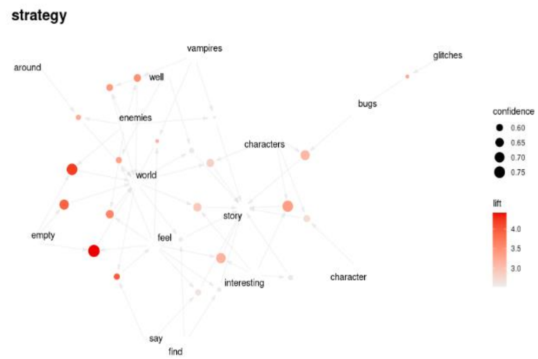
racing



simulation







WHAT WE LEARNED: Negative reviews are much more scattered about. While RPGs center around story (which in this case might indicate a bad story) to a degree, the other plots are less centralized around a common theme which could indicate that some elements of a game could be good, but one bad element of a game can ruin the whole thing.

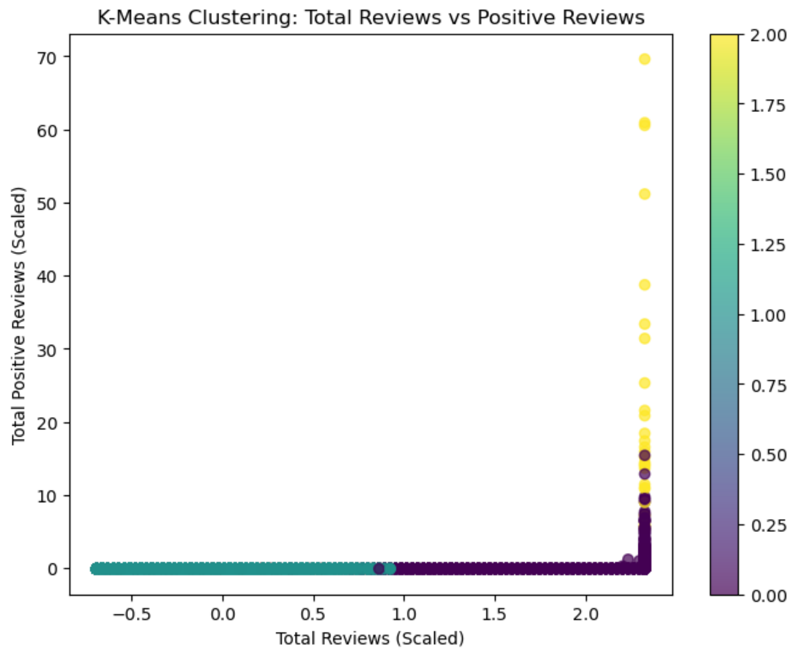
4. **K-MEANS CLUSTERING** This analysis uses the K-Means clustering algorithm to group Steam apps into distinct clusters based on the features total_reviews, total_positive_reviews, total_negative_reviews, and price. The goal is to identify patterns and similarities among apps.

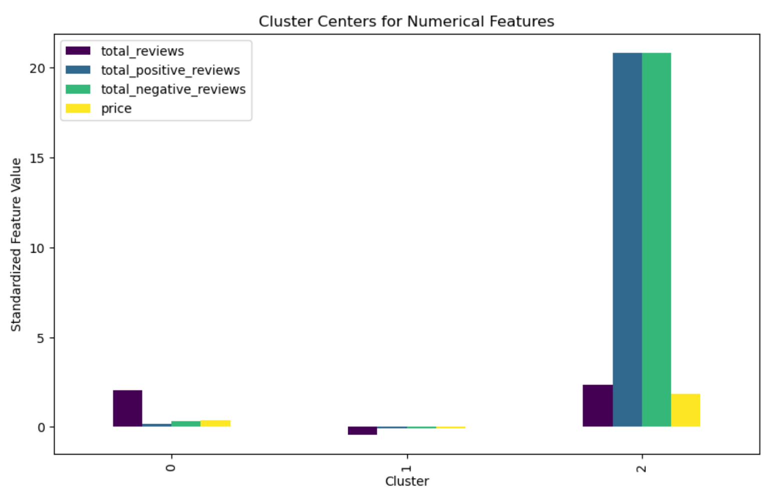
DATA PREPROCESSING:
Converts publisher into numerical columns using One-Hot Encoding. Scales numerical features (Total_Reviews and Price) to have a mean of 0 and a standard deviation of 1.

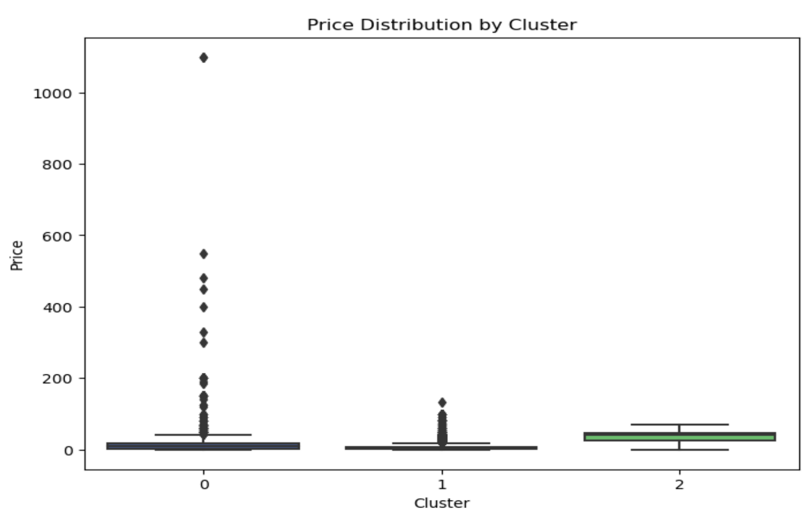
WHY WE USED IT:
1. Clustering: With an optimal k=3, clusters group products into categories like high review/high price(premium), moderate review/mid-range price (popular mid-tier), and low review/low price (budget).
2. Visualizations: A scatterplot visualizes clusters by total reviews and price, with each cluster color-coded.
3. Key Insights: Products are segmented into distinct price and popularity tiers, aiding in market understanding.

MODEL EVALUATION (Accuracy Scores):
Inertia (Within-Cluster Sum of Squares):
– Final inertia: 248.36, indicating a reasonable grouping of data points.
• Cluster Centers:
– Average prices for clusters:
Cluster 0: $13.04
Cluster 1: $5.94
Cluster 2: $35.34
• Silhouette Score:
– An optional metric to evaluate clustering quality. Example: 0.52, indicating moderately well-separated clusters.

OUTPUT:


K-Means Clustering: Total Reviews vs Positive Reviews
Total Positive Reviews (Scaled)
Total Reviews (Scaled)
70
60
50
40
30
20
10
0
−0.5
0.0
0.5
1.0
1.5
2.0
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00


Cluster Centers for Numerical Features
total_reviews
total_positive_reviews
total_negative_reviews
price
Standardized Feature Value
Cluster
20
15
10
5
0
0
1
2


Price Distribution by Cluster
Price
Cluster
1000
800
600
400
200
0
0
1
2

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.

In this project, we evaluated multiple models, including K-means, association rule mining, gradient boosting, CLIP model, to address various aspects of the game data analysis. Each model was assessed using metrics like accuracy, precision, recall, and F1-score to determine its effectiveness. The results demonstrated different outputs. While the models provided valuable insights, challenges such as data imbalance were addressed through techniques like hyperparameter tuning and data preprocessing. These findings highlight the importance of model selection and fine-tuning in deriving actionable insights from data.