

A decorative graphic on the left side of the slide. It consists of a blue parallelogram and a light green parallelogram, both tilted at an angle. The blue shape is in the foreground, and the green shape is partially behind it. They are set against a dark blue background with faint, lighter blue diagonal stripes.

Up and Coming Video Games

Topic Discovery

We began exploring project topics with eSports, for there was a lot of data available within the video game industry.

As we explored different areas, we came upon the questions of how did video games get so big?

The industry has grown exponentially and there are many types of games available.



League of Legends (LOL) World Championship Finals in Paris, France, November 10, 2019.

Topic Selection

Our topic was further narrowed down to figuring out what makes a video game popular.

There are many features to focus on, such as the platform, type of game-play, genre, and various other categories.

By analyzing this data, it would help potential clients pick which creators to support, or help with decisions among their own creative teams to create a successful and popular video game.



Platform devices/controllers from left:
Nintendo Switch, Playstation 4, Xbox One, Personal Gaming Computer

Data Exploration

We discovered a lot of available data sources, but a promising one was the Steam service, which has a library of nearly 30,000 games of their own make and from third-party publishers.

Many open datasets from Steam are available on Kaggle to choose and work with.







Data Exploration - Sources

[Steam Store Games \(Clean dataset\)](#)




- Combined data of 27,000 games scraped from Steam and SteamSpy APIs (May 2019)
- Creator: Nik Davis

[Steam Games Complete Dataset](#)

- 40k Steam Games Dataset from Steam shop with detailed data. (June 2019)
- Creator: Alexander Antonov

A name	A platforms	A steamspy_tags	# positive_ratings	# negative_ratings
Title of app (game)	Semicolon delimited list of supported platforms. At most includes: windows;mac;linux	Semicolon delimited list of top steamspy game tags, similar to genres but community voted, e.g. action;adventure	Number of positive ratings, from SteamSpy	Number of negative ratings, from SteamSpy
27033 unique values	windows 68% windows;mac;linux 17% Other (4054) 15%	Action;Indie;Casual 3% Action;Adventure;In... 3% Other (25516) 94%	 0 2.64m	 0 487k
Counter-Strike	windows;mac;linux	Action;FPS;Multiplayer	124534	3339
Team Fortress Classic	windows;mac;linux	Action;FPS;Multiplayer	3318	633
Day of Defeat	windows;mac;linux	FPS;World War II;Multiplayer	3416	398

Focused columns from [Steam Store Games \(Clean dataset\)](#), 5 out of 18 columns displayed.

<u>A</u> name 	<u>A</u> popular_tags 	<u>A</u> genre 
Name of a game	tags	genre(s) of a game
40752 unique values	NaN 7% Action 3% Other (36808) 90%	Action 6% Action,Indie 5% Other (36318) 89%
DOOM	FPS, Gore, Action, Demons, Shooter, First-Person, Great Soundtrack, Multiplayer, Singleplayer, Fast-Paced, Sci...	Action
PLAYERUNKNOWN'S BATTLEGROUNDS	Survival, Shooter, Multiplayer, Battle Royale, PvP, FPS, Third-Person Shooter, Action, Online Co-Op, Tactical...	Action, Adventure, Massively Multiplayer
BATTLETECH	Mechs, Strategy, Turn-Based, Turn-Based Tactics, Sci-fi, Turn-Based Strategy, Tactical, Singleplayer, Robots...	Action, Adventure, Strategy

Focused columns from [Steam Games Complete Dataset](#), 3 out of 20 columns displayed.



Questions to Answer

What features determine a video game's popularity or success?

What relations are there between a game's rating and genre or popular tags?

	game_name [PK] character varying	percent_positive_reviews numeric (4,1)	popular_tags character varying	tag_1980s boolean	tag_1990s boolean	tag_2_5d boolean	tag_2d boolean	tag_2d_fighter boolean	tag_3d boolean	tag_3d_platformer boolean	tag_4_play boolean
1	#MONSTERCAKES	88.0	Casual,Indie,Puzzle,Ma...	false	false	false	false	false	false	false	false
2	\$1 RIDE	42.0	Casual,Action,Indie,Arc...	false	false	false	false	false	false	false	false
3	.HACK//G.U. LAST RECODE	88.0	JRPG,RPG,Anime,Singl...	false	false	false	false	false	false	false	false
4	//N.P.P.D. RUSH// - THE MILK...	33.0	Indie,Bullet Hell,Retro,A...	false	false	false	true	false	false	false	false
5	[THE SEQUENCE]	93.0	Indie,Puzzle,Minimalist...	false	false	false	false	false	false	false	false
6	~NECROMANCY~EMILY'S E...	71.0	Sexual Content,Nudity,...	false	false	false	false	false	false	false	false
7	【SCP】器閑 / 彷徨 -THE WI...	88.0	Free to Play,Violent,Gor...	false	false	false	false	false	false	false	false
8	√LETTER - ROOT LETTER -	84.0	Adventure,Violent,Visu...	false	false	false	false	false	false	false	false
9	001 GAME CREATOR	81.0	Game Development,De...	false	false	false	false	false	false	false	false

	game_name [PK] character varying	percent_positive_reviews numeric (4,1)	genre character varying	action boolean	adventure boolean	casual boolean	design_and_illustration boolean	early_access boolean	free_to_play boolean	indie boolean	massively_m boolean
1	#MONSTERCAKES	88.0	Casual,Indie	false	false	true	false	false	false	true	false
2	\$1 RIDE	42.0	Action,Casual,Indie	true	false	true	false	false	false	true	false
3	.HACK//G.U. LAST RECODE	88.0	RPG	false	false	false	false	false	false	false	false
4	//N.P.P.D. RUSH// - THE MILK...	33.0	Action,Indie	true	false	false	false	false	false	true	false
5	[THE SEQUENCE]	93.0	Indie	false	false	false	false	false	false	true	false
6	~NECROMANCY~EMILY'S E...	71.0	Action,Adventure,Casu...	true	true	true	false	false	false	true	false
7	【SCP】器閑 / 彷徨 -THE WI...	88.0	Adventure,Free to Play,...	false	true	false	false	false	true	false	false
8	√LETTER - ROOT LETTER -	84.0	Adventure	false	true	false	false	false	false	false	false
9	001 GAME CREATOR	81.0	Animation & Modeling,...	false	false	false	true	false	false	false	false

Cleaned Data Sets stored in PostgreSQL database.

Top Sample - Popular Tags data set

Below Sample - Genres data set

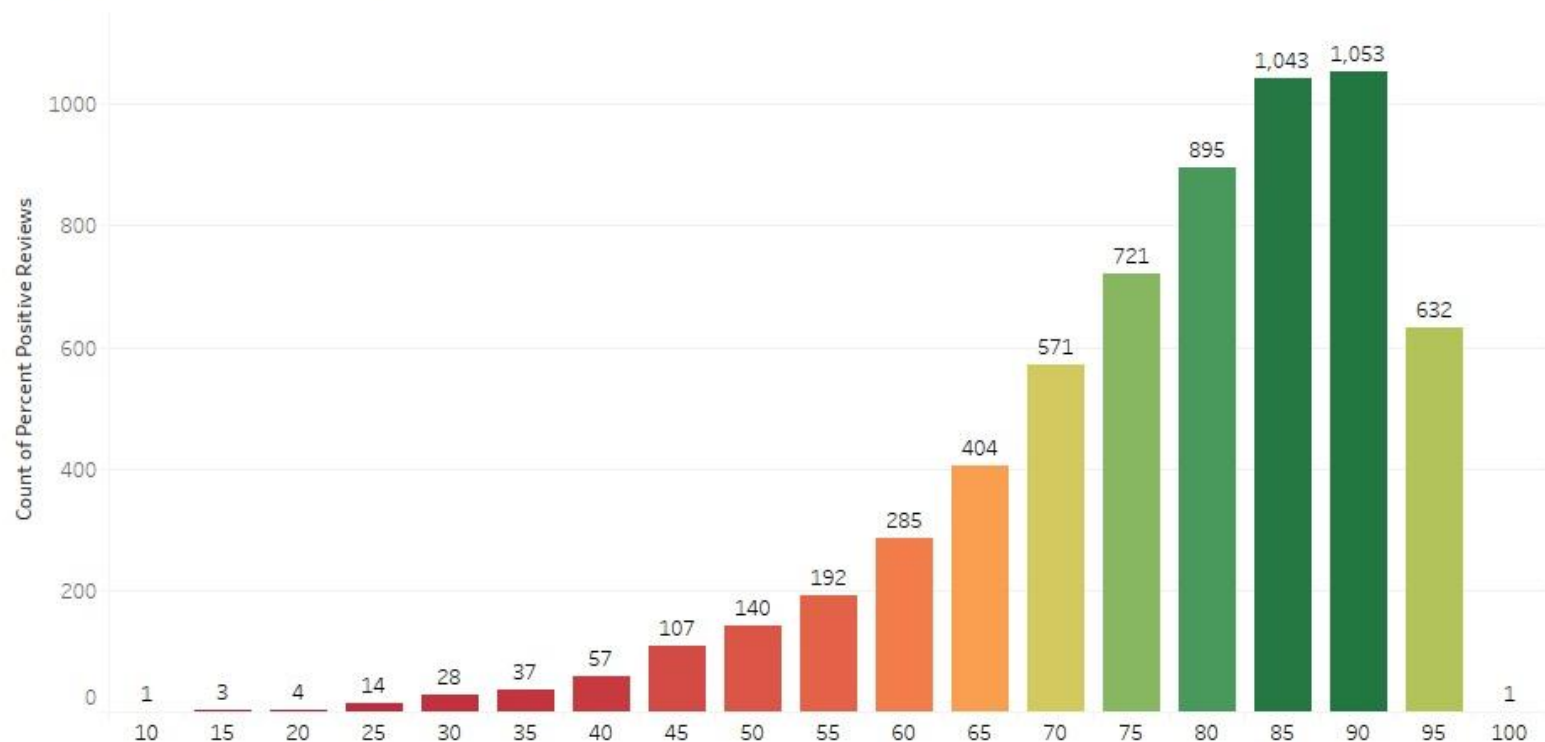
Video Game Ratings

SABBAT OF THE WITCH	100	^
ARCADE SPIRITS	99	
BACKBONE: PROLOGUE	99	
BLUE REVOLVER	99	
BYTEPATH	99	
COPY KITTY	99	
DING DONG XL	99	
DX-BALL 2: 20TH ANNIVERSARY EDITION	99	
FACTORIO	99	
FINAL ASSAULT	99	
FLOWERS -LE VOLUME SUR ETE-	99	
FURERABA ~FRIEND TO LOVER~	99	
GOOGLE SPOTLIGHT STORIES: AGE OF SAIL	99	
HEART OF THE WOODS	99	
HELLO CHARLOTTE EP3: CHILDHOOD'S END	99	
HIDDEN IN PLAIN SIGHT	99	
HYPNOSPACE OUTLAW	99	
ISOLAND 2 - ASHES OF TIME	99	
KAORI AFTER STORY	99	
LINELIGHT	99	
LITTLE BUSTERS! ENGLISH EDITION	99	

Using individual ratings answered a simple question: What games were rated the most popular?

This opened our next steps to determine the counts of games that fell into different rating brackets.

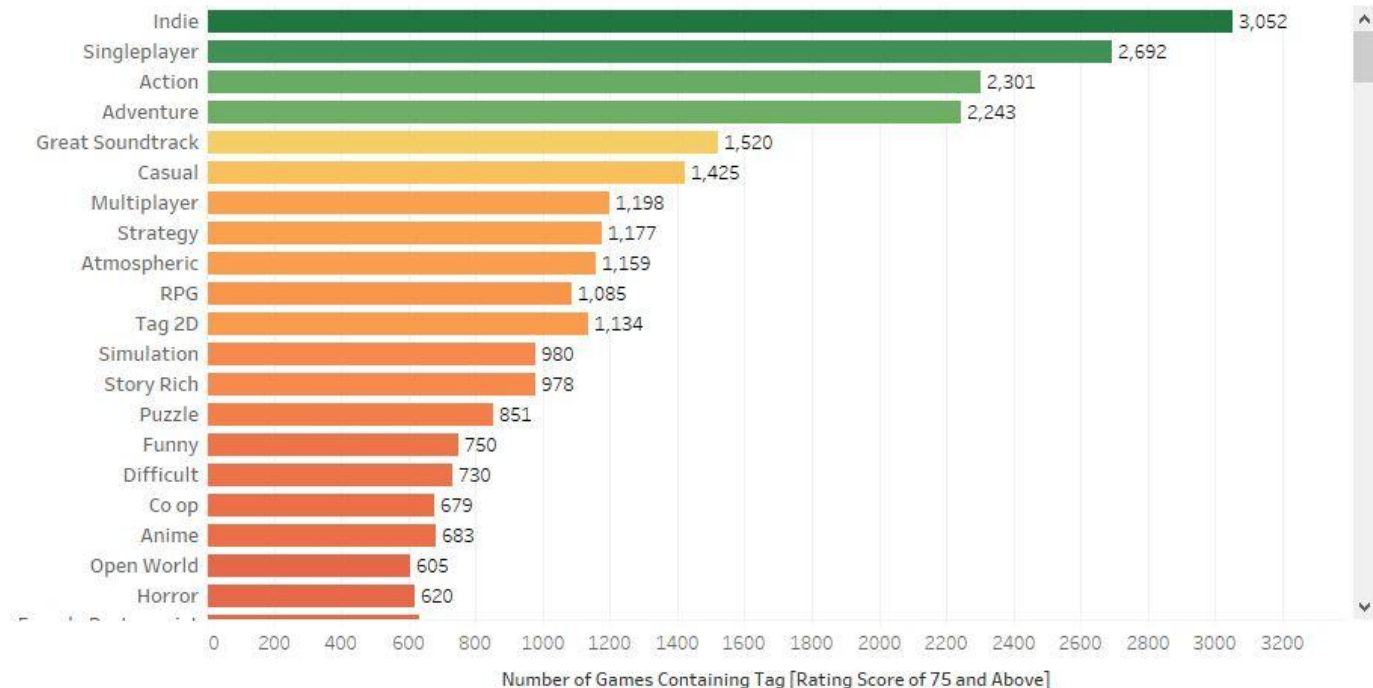
Count of Games per Rating



By grouping the data into average ratings, we were able to determine the minimum rating to use towards our machine learning model.

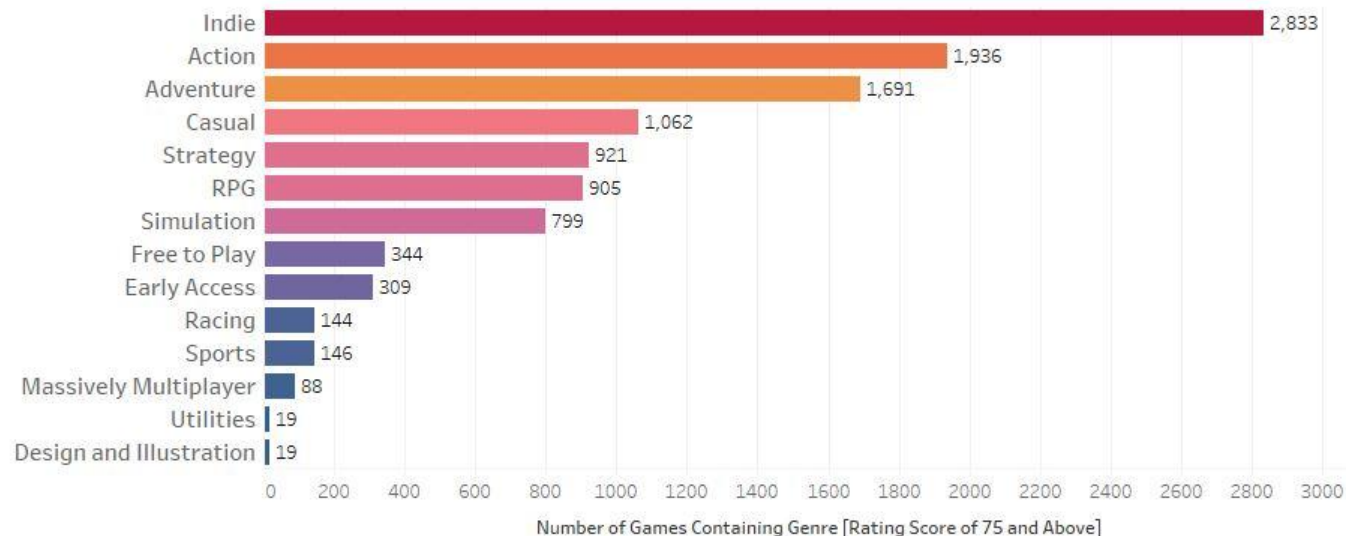
This defined what was considered a “Popular Game” and which was not. We declared a 70 minimum rating for popularity.

Popular Tag Counts



After declaring a score of 70 or above as a popular rating, we determined which Popular Tags were the top counts.

Genre Counts



After declaring a score of 70 or above as a popular rating, we determined which Genres were the top counts.



[Dashboard on Tableau](#) with charts updating with field and/or game selections.

Analysis Results - Machine Learning

Logistic Regression Model:
Popular Tags

	precision	recall	f1-score	support
0	0.63	0.41	0.50	494
1	0.76	0.89	0.82	1053
accuracy			0.73	1547
macro avg	0.70	0.65	0.66	1547
weighted avg	0.72	0.73	0.72	1547

Logistic Regression Model:
Genre

	precision	recall	f1-score	support
0	0.70	0.13	0.22	494
1	0.70	0.97	0.82	1053
accuracy			0.70	1547
macro avg	0.70	0.55	0.52	1547
weighted avg	0.70	0.70	0.63	1547




Conclusion

Top Popular Tags:

1. Indie
2. Singleplayer
3. Action
4. Adventure
5. Great Soundtrack
6. Casual
7. Multiplayer
8. Strategy
9. Atmospheric
10. RPG

Top Genres:

1. Indie
2. Action
3. Adventure
4. Casual
5. Strategy
6. RPG
7. Simulation



Recommendations for Future Analysis & Changes

- Connecting our data sources using Steam's API to provide updated library results, therefore continuing to evolve our learning model and allowing a client to track any trends over time.
- Expanding and including video games from other platforms.

** Machine Learning improvements:

Q & A

