Up and Coming Video Games

Team: Zubair Aslam, Eva Fuentes-Lopez, Ammad Rashid, Karyssa Tillis

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 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 Semicolon delimited list of supported platforms. At most includes: windows;mac;linux	A steamspy_tags Semicolon delimited list of top steamspy game tags, similar to genres but community voted, e.g. action;adventure	# positive_ratings = Number of positive ratings, from SteamSpy	# negative_ratings = 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;Multiplay	124534	3339
Team Fortress Classic	windows;mac;linux Action;FPS;Mult er		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.

A name =	<u>A</u> popular_tags	A genre = 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, Demo ns, Shooter, First- Person, Great Soundtrack, Multiplay er, Singleplayer, Fast -Paced, Sci	Action
PLAYERUNKNOWN'S BATTLEGROUNDS	Survival, Shooter, Mul tiplayer, Battle Royale, PvP, FPS, Third -Person Shooter, Action, Onlin e Co-Op, Tactical	Action,Adventure,Mas sively Multiplayer
BATTLETECH	Mechs, Strategy, Turn- Based, Turn-Based Tactics, Sci-fi, Turn- Based Strategy, Tactical, Si ngleplayer, Robots	Action,Adventure,Str ategy

Focused columns from <u>Steam Games Complete Dataset</u>, 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?

Data Analysis

Determined the following aspects to focus on from our data sources:

- PC based video games
- Popular Tags
- Genre
- Ratings

Combined both data sets based on name and reduced outliers by eliminating the following:

- Low rating counts
- Low counts of miscellaneous Popular Tags
- Low counts of Genre categories

	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	d_fighter an	tag_:	- 4	tag_3	d_platfean	ormer		ag_4_play oolean
1 :	#MONSTERCAKES	88.0	Casual,Indie,Puzzle,Ma	false	false	false		false	false		false		false			fa	alse
2	\$1 RIDE	42.0	Casual,Action,Indie,Arc	false	false	false		false	false		false		false			fa	alse
3 .	HACK//G.U. LAST RECODE	88.0	JRPG,RPG,Anime,Singl	false	false	false		false	false		false		false			fa	alse
4	//N.P.P.D. RUSH//- THE MILK	33.0	Indie,Bullet Hell,Retro,A	false	false	false		true	false		false		false			fa	alse
5	[THE SEQUENCE]	93.0	Indie,Puzzle,Minimalist	false	false	false		false	false		false	ĺ	false			fa	alse
6	~NECROMANCY~EMILY'S E	71.0	Sexual Content, Nudity,	false	false	false		false	false		false	Ę.	false			fa	alse
7	【SCP】器関ノ彷徨 -THE WI	88.0	Free to Play, Violent, Gor	false	false	false		false	false		false	6	false			fa	alse
8 -	VLETTER - ROOT LETTER -	84.0	Adventure,Violent,Visu	false	false	false		false	false		false		false			fa	alse
9 (001 GAME CREATOR	81.0	Game Development,De	false	false	false		false	false		false	ģ.	false			fa	alse
	game_name [PK] character varying	percent_positive_reviews numeric (4,1)	genre character varying			casual coolean	desi bool	i gn_and_illustrat ilean	tion	early_access boolean		free_to.		ind bo	lie olean 🎤	mas boo	ssively_m lean
1	#MONSTERCAKES	88.0	Casual,Indie	false 1	false t	rue	false	e		false		false		tru	е	false	е
2	\$1 RIDE	42.0	Action,Casual,Indie	true 1	false t	rue	false	e		false		false		tru	е	false	е
3	.HACK//G.U. LAST RECODE	88.0	RPG	false 1	false f	alse	false	е		false		false		fals	se	false	e
4	//N.P.P.D. RUSH//- THE MILK	33.0	Action,Indie	true 1	false f	alse	false	е		false		false		tru	е	false	e
5	[THE SEQUENCE]	93.0	Indie	false 1	false f	alse	false	e		false		false		tru	е	false	e
6	~NECROMANCY~EMILY'S E	71.0	Action,Adventure,Casu	true t	true t	rue	false	e		false		false		tru	е	false	e

true

true

false

false

false

false

false

false

false

true

false

false

false

true

false

false

false

false

false

false

false

false

Cleaned Data Sets stored in PostgreSQL database. Top Sample - Popular Tags data set Below Sample - Genres data set

88.0 Adventure, Free to Play,...

81.0 Animation & Modeling,...

84.0 Adventure

【SCP】器関/彷徨-THE WI...

√LETTER - ROOT LETTER -

001 GAME CREATOR

Data Analysis

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

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

Game	Rating	
Name	%	

Data Analysis

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.



Data Analysis - Machine Learning

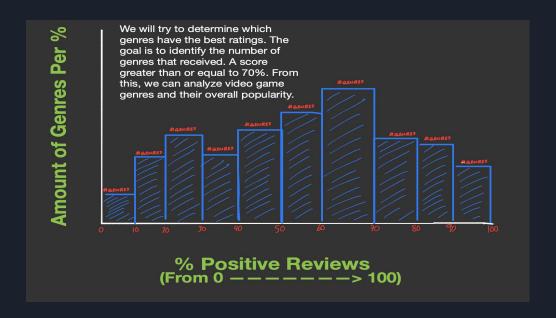
Data Analysis - Top Genres

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



Data Analysis - Top Popular Tags

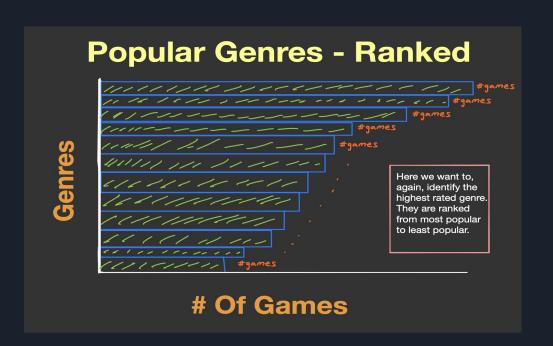
After declaring a score of 70 or above as a popular rating, we determined which popular tags were the top counts.



Data Analysis - Genre Ranks

Top Genres:

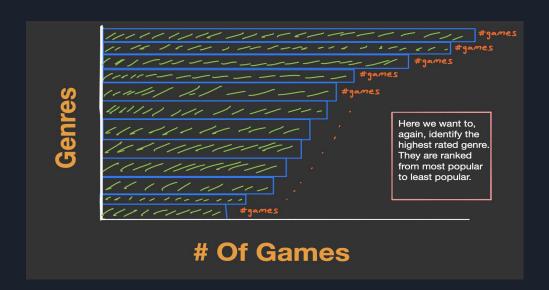
- 1. Item
- 2. Item
- 3. Item
- 4. Item
- 5. Item



Data Analysis - Popular Tag Ranks

Top Popular Tags:

- 1. Item
- 2. Item
- 3. Item
- 4. Item
- 5. Item



Data Analysis - Identifying Top Games

After identifying top genres and popular tags, found the specific games that fell under each category with the best ratings

Individual Game %

Game	Rating
Name	%

Sorting = Descending

After identifying the most popular genre (i.e. the genre with the highest rating percentage), we want to find the specific game(s) with the best ratings.

Tableau Storyboard & Dashboard

Results of Analysis

- Final top genres/top popular tags
- Top games falling under these categories

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: