Steam Video Game Recommender

Springboard Capstone Project 2 Milestone Report

Greg Gibson December 2020

Problem Statement:

Based on video game votes and reviews, can a recommendation system be built to assist sales growth?

Valve Corporation, an American video game development and digital distribution company (www.valvesoftware.com), is the developer of the market dominant software distribution platform Steam (www.steampowered.com), as well as many video games.

Steam noted 90 million monthly users at the end of 2018, and Statista.com states Steam revenue in 2017 was $4.3 billion and they had 18% of the video game market. Per Steam's news update on April 7, 2020, they had nearly 1,200 new games released in 2019 that earned $10,000 in their first two weeks. Therefore, any small uptick in sales would yield great results. What if a reliable recommendation system would help gamers make confident choices on which video game would give them the most entertainment for the value?

Dataset:

Video game developers can utilize the Steam API, called Steamworks, to allow social networking and community interaction, such as friend networks, game hosting, score rankings, and posting achievements unlocked and in-game snapshots. In addition, the website collects and shares user reviews and up or down votes.

Pypi.org has a library "steamreviews" developed by Wok, https://pypi.org/project/steamreviews/#files, that handles downloading reviews by game ID. For a provided list of game IDs, each ID will be downloaded in a separate JSON file. The library accesses Steamworks user reviews and more information can be found here: https://partner.steamgames.com/doc/store/getreviews

A list of 1,000 game IDs was obtained from Steamspy, a website dedicated to estimating game sales on Steam, https://steamspy.com/api.php. The parameters through steamreviews were restricted to purchased games, English language reviews, and created within the last three years. This yielded 748 games. Each JSON file had a nested "author" column with information regarding the reviewer that was subsequently flattened into the dataframe.

The primary features to be used are the voted_up boolean and review text columns. The recommendationid is the unique field. Game ID is appid and each reviewer has a unique number in the author.steamid column.

Exploratory Analysis:

Initial counts:

- 748 game IDs
- 2,324,565 customer IDs
- 3,264,320 reviews
- Avg. ~4,300 reviews per game

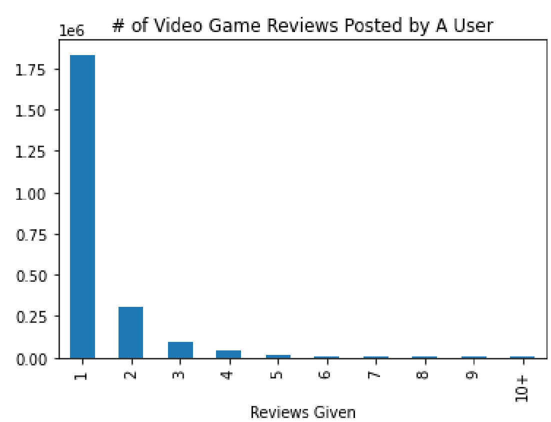
There are 6,969 blank reviews and these records are dropped. Some reviews are noted to be not in English, despite the download parameters. These rows need to be excluded as well prior to sentiment analysis.

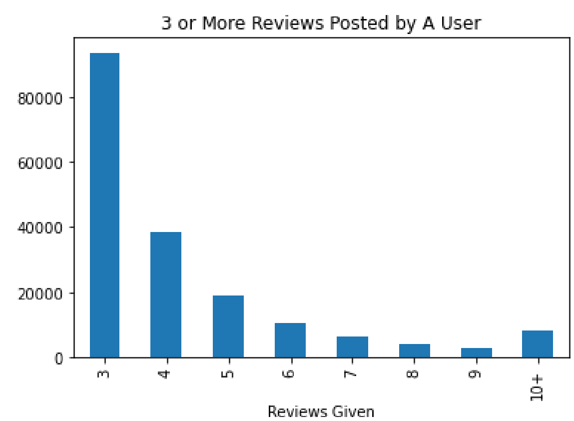
The players, per the review data:

- The average player bought 128 games from Steam
- The average player has nearly 16,400 minutes (273 hours) playing on Steam
- The average player spends about 100 minutes playing per week
- The number of reviews per player range from 1 to 125
- There are 1,832,307 players, or 79%, with only one review and 492,258 with multiple reviews
    - There are only 50,129 players with five or more reviews which is less than 2.2% of total
    - 182,631 players posted three or more reviews, or 7.9%

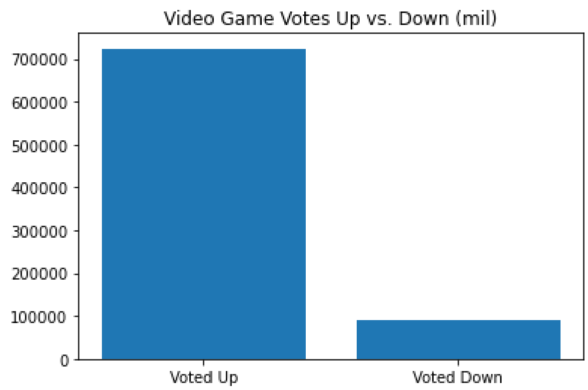
This sub-group of 183K players is selected to have sufficient records for a recommender system and for go-forward analysis.

- Just over half of those players, 93,943 or 51%, were exactly three reviews

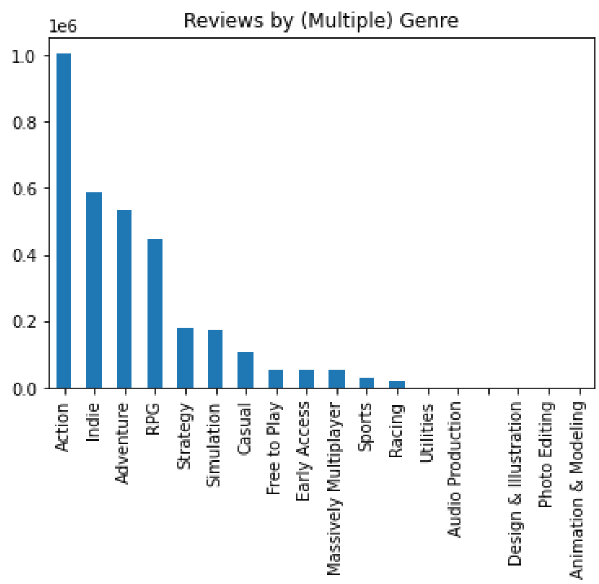




- The players approved of video games 8x more often than disapproved



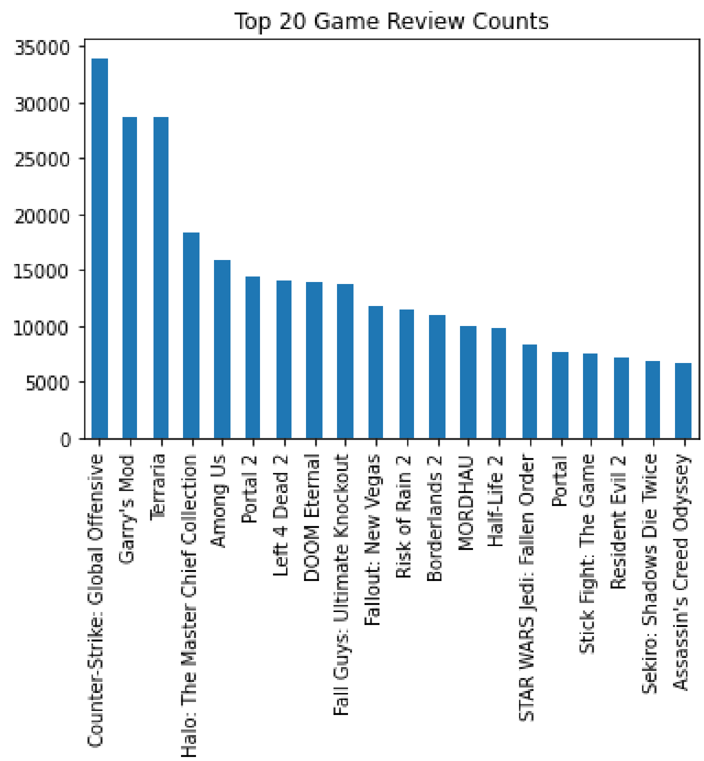
- Players primarily logged reviews of Action video games, double next categories
- (Games can be identified by multiple genres, the same review counts in each)



The games:

Constrained to players posting at least three reviews, the top game is Count-Strike: Global Offensive with nearly 34,000 reviews. The top 12 games are over the 10,000 review mark.

```
Counter-Strike: Global Offensive     33950
Garry's Mod                          28686
Terraria                             28598
Halo: The Master Chief Collection    18376
Among Us                             15900
Portal 2                             14330
Left 4 Dead 2                        14067
DOOM Eternal                         13807
Fall Guys: Ultimate Knockout         13680
Fallout: New Vegas                   11811
Risk of Rain 2                       11446
Borderlands 2                        10878
```



There should be a minimum number of reviews logged to allow a game to be recommended. When constrained to at least 1,000 reviews, the game selection dropped to only 189 choices. Arbitrarily, a minimum of 100 reviews was selected, leaving almost 500 different video games to recommend.

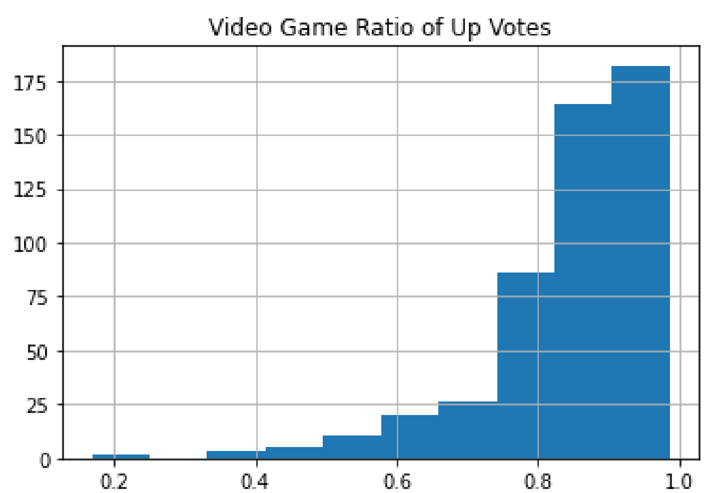
Of the remaining games, the lowest rated from up and down votes is Infestation: Survivor Stories 2020, just 17% voted up. Only ten games do not reach the 50% approval mark:

appid	name	reviews	up	down	up_rate
226700	Infestation: Survivor Stories 2020	107	18	89	17%
295110	Just Survive	569	138	431	24%
215280	Secret World Legends	158	59	99	37%
224540	Ace of Spades: Battle Builder	337	127	210	38%
424370	Wolcen: Lords of Mayhem	218	89	129	41%
1015500	WWE 2K20	165	70	95	42%
437220	The Culling	192	82	110	43%
834910	ATLAS	2509	1075	1434	43%
529180	Dark and Light	155	72	83	46%
841370	NBA 2K19	506	245	261	48%

For comparison, the top ten games in our analysis, led by the very popular Portal series:

appid	name	reviews	up	down	up_rate
620	Portal 2	14330	14148	182	99%
431960	Wallpaper Engine	291	287	4	99%
400	Portal	7583	7475	108	99%
427520	Factorio	108	106	2	98%
379720	DOOM	105	103	2	98%
312530	Duck Game	749	733	16	98%
250900	The Binding of Isaac: Rebirth	4868	4756	112	98%
294100	RimWorld	335	327	8	98%
519860	DUSK	2852	2783	69	98%
250320	The Wolf Among Us	976	951	25	97%

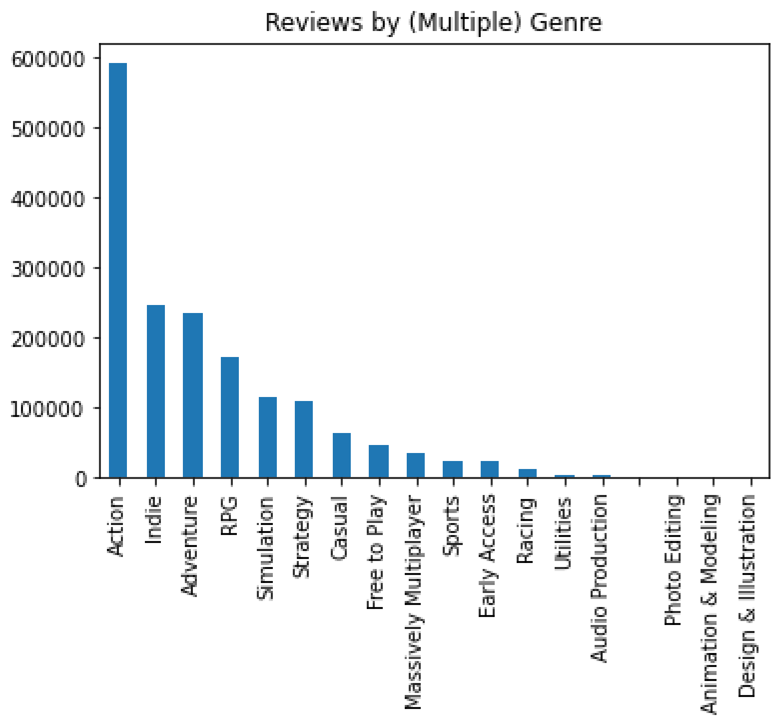
The video games skew toward highly rated, with the largest grouping over 90%:



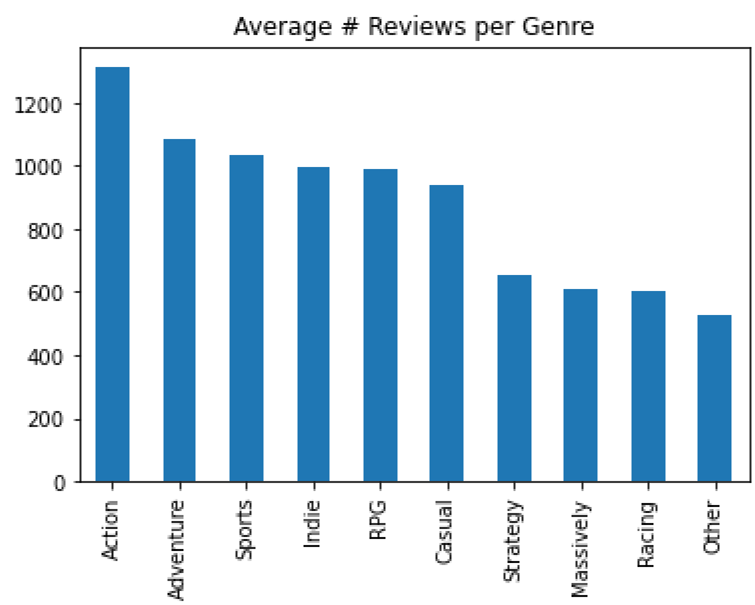
Our current dataset, limited to players who have posted at least three reviews, on games which have at least 100 reviews, has an average ratio of Up votes 85% of the time.

Genres:

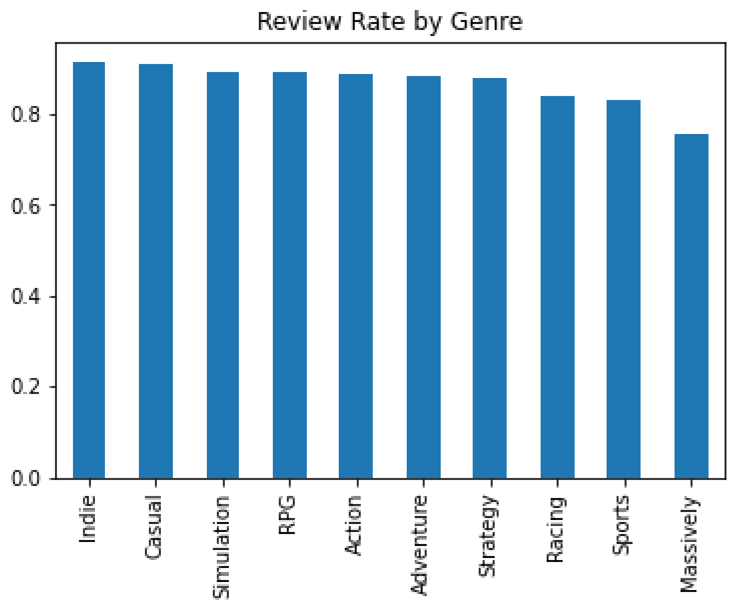
The video game reviews are led by Action genre, more than double the second and third categories of Indie and Adventure. Games can represent multiple genres per Steam's methods.



Led by Action, the genres split into an average of about 1,000 reviews per game for the top six, down to about 600 reviews on average for the remainder.



The genres overall mostly rate near 90% approval, led by the independent developers. Racing and Sports decline to low 80's and Massively Multiplayer is the lowest rated genre at 76%.



<u>Text Analysis:</u>

The textual reviews should be compared to the vote scores, to allow further refinement of video game ranking.

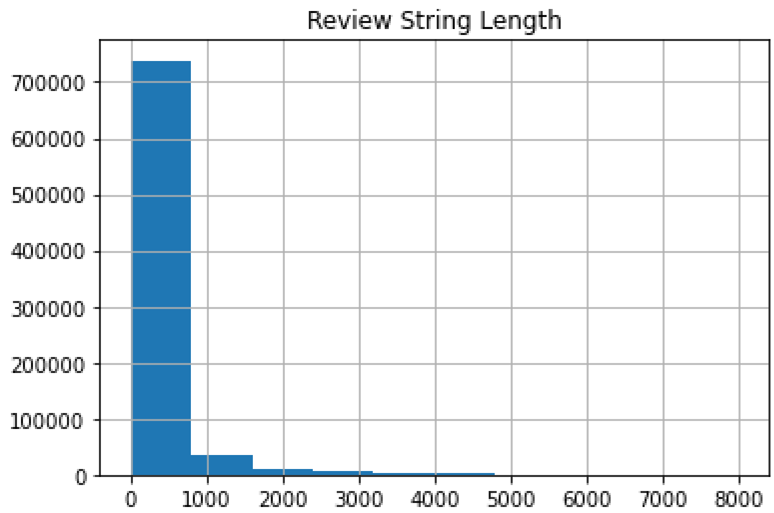
Cleaning:

To remove some not useful character language and symbols, Regex code was run to keep alphabetical characters and drop symbols, leaving 793,000 reviews in the dataframe.

Current distribution shows some wordy reviews up to 8,000 characters, apparently the maximum allowed. Most, but not all, appear to repeat a character or word to reach the maximum.

Several steps will be taken to correct comments with repetition and stretched words with repeating characters.

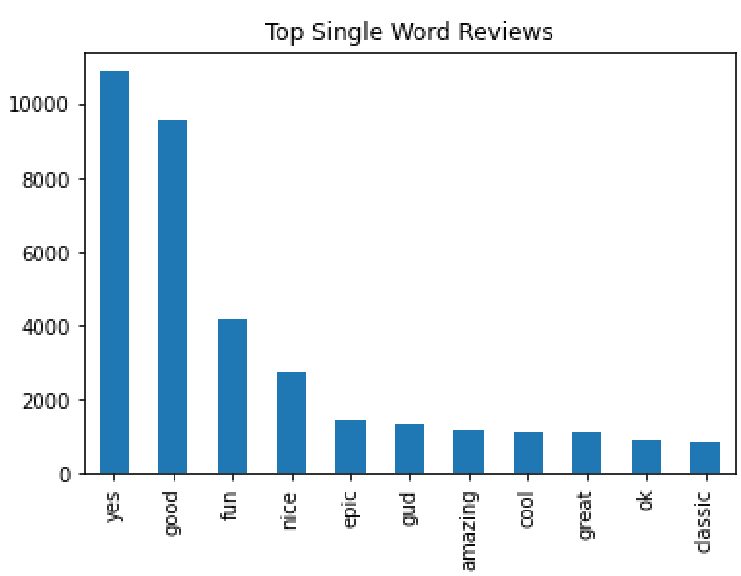
Any large reviews without appropriate white space counts will be removed.



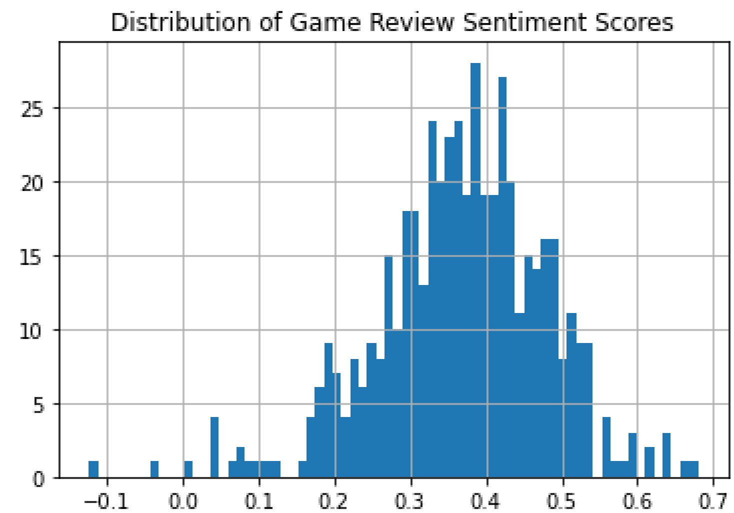
Upon observation, many reviews are not using the spacebar, making long, single-text entries, frequently gibberish, that will not be useful for identifying sentiment. The longest words observed were 14 characters: disappointment and oversimplified. At length 15 characters and up, the reviews appeared not useful. Remove any reviews of one ‘word’ entry, greater than 14 characters, or 1-2 characters with no meaning.

The reviews were tokenized into lists of words and compared to the set of words from the nltk library. Any reviews with zero matches to the set of words, about 4,300, were deleted. There are 783,686 reviews remaining, with an average review length of 29 words.

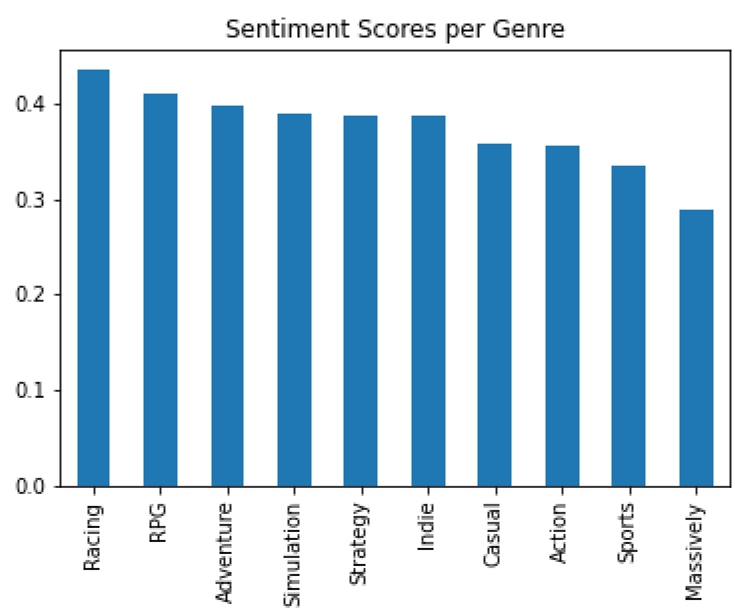
Top single word reviews are led by “yes” and “good”, the latter being first if including the purposeful typo “gud”, which happens to exist as a match in the nltk set.



Python's sentiment rating tool, VADER, was applied to the reviews to obtain compound scores. The rating is from -1 to 1, and the mean score of these reviews is 0.36. Why is the VADER score low? Words used in gaming can have a very negative connotation outside gaming: Shoot -0.34, Battle -0.38, Destroy -0.54, Enemy -0.54, Die -0.60, Terrorist -0.69, etc.



The sentiment rating by genre rates Sports and Massively Multiplayer games at the bottom, the same as simple Up/Down review ratios. But the rest are very different, Racing games rate far better, likely due to not using combat words, Indie and Casual move to the middle and Action drops toward the bottom. This should help the recommender system suggest more games outside of a player's usual choice.



Here is how the sentiment scores changed the top and bottom ten game rankings. The two games with negative rates are the same two lowest games as Up/Down voting, as well as Ace of Spades. There is no overlap in the top ten games, the two Portal versions are replaced with two Trine versions.

name	game_score
Machinarium	68%
Unheard	66%
Trine Enchanted Edition	64%
GRIS	64%
Trine 2: Complete Story	64%
Brothers - A Tale of Two Sons	62%
Bastion	62%
Ori and the Will of the Wisps	60%
The Tiny Bang Story	60%
A Plague Tale: Innocence	59%

name	game_score
Hurtworld	7%
BROKE PROTOCOL: Online City RPG	7%
Getting Over It with Bennett Foddy	7%
Sniper: Ghost Warrior	5%
F.E.A.R. 3	5%
Ace of Spades: Battle Builder	4%
Doom 3: BFG Edition	4%
Kane & Lynch 2: Dog Days	1%
Just Survive	-3%
Infestation: Survivor Stories 2020	-12%

Can "gaming words" be accounted for, to improve sentiment scores? There are 177,000 unique words used within the game reviews. 2,700 of those were ranked negatively by VADER, the worst being "rapist" at -0.71.

	word	score
129983	rapist	-71%
125745	slavery	-70%
38086	raping	-70%
42409	murder	-69%
97485	fu	-69%
...	...	...
136760	longing	-3%
144661	battleship	-3%
105570	sentenced	-3%
161756	freeloaders	-3%
66003	lowland	-3%

Reviewing the 2,700 negative words, about 300 words and their variations were identified for video game use, such as: evil, villain, violent, blood, threat, attack, fought, defeat, prisoner, casualties, hostiles, apocalyptic, etc.

Excluding these words from sentiment scoring, the mean only changed 13%, improving from 0.36 to 0.41. Something else is restraining sentiment scores. The remaining words were then weighted by multiplying their frequency of occurrence among the reviews.

	word	freq	score	weight
23	no	83,455	-30%	(24,703)
137	bad	44,849	-54%	(24,322)
942	shit	14,698	-56%	(8,193)
1145	hell	10,656	-68%	(7,255)
2404	doom	18,041	-40%	(7,251)

The frequency of use of the words 'no' and 'bad' appear to put most of the negative sentiment into reviews. Hopefully, video games with little use of these words will rise to the top of our recommendations.

What happens if we replace “bad guy” to remove the negative connotation? The rescoring improves slightly at the 4th decimal place.

Collaborative Filtering:

As the sentiment analysis is subjective, let us start with a baseline based on the simple up-vote provided by the users for each game they have played. For binary simplicity, only the positive up-vote will be used(1) and ignore the down-vote. Games disliked or not played will be zero.

First, an item based filter will take similarities between game consumption histories, using cosine similarity because of its simplicity with sparse vectors.

The user information is dropped, and the matrix has the video game names as the index and the column names. Games that intersect with themselves have a cosine similarity of 1. For the remaining games, the closer to 1, the greater the frequency those two games are liked together.

For each game, the cosine similarity values can be sorted in descending order, identifying the most similarly liked games, led by matching itself. This first column of matching itself can be ignored.

Checking on a few examples, video game series with many iterations would match themselves, i.e. Resident Evil 1 through 6, Resident Evil Revelations and Revelations 2. Tomb Raider’s various titles would match, such as TR: Legend, TR: Anniversary, TR: Underworld, but I was pleased to see the connection to the differently named “Lara Croft and the Guardian of Light”. There was also a game title, 古剑奇谭三(Gujian3), preceded by Chinese characters, and its closest matching game was titled “Chinese Parents”.

Next is to create the user based filter by computing all pairwise vector similarities. This matrix has user IDs for the index and game names for columns. For each user-game intersection, if the user has already up-voted this game, set it to zero to avoid recommendation. Otherwise, obtain the top nearest matching games and similarity score from the prior table, ignoring the same game itself. For any of these similar games, if the user has up-voted them, sum their similarity scores and divide by the number of up-voted games to create a weighted score estimating how closely the new, unplayed game might score. Sort these newly scored games in descending order to make the sequence of recommended video games for the user to purchase.

Recommendation Accuracy:

Future Improvements:

Due to time restrictions, there are additional steps to improve this model in the future as time allows.

1. Replace the very slow for-loops that populate the user-game matrix for user based filtering with a Jaccard similarity on a scipy sparse matrix, scipy.sparse.csc_matrix.

2. Cluster users into groups and restrict calculations to those sub-groups.

3. Do not calculate the scores of the entire user-based matrix, limit first to only the video games' nearest neighbors.

4. Incorporate the sentiment analysis scores, probably scaled, as an additional refinement to recommending the best matching games, in a similar manner to a movie recommender.

5. Incorporate a magnitude measurement to normalize the user vectors such that a person who up-voted 100 games doesn't overwhelm users who up-voted 10 games.

6. To make recommendations more interesting, consolidate series into one title, where the user has played none of them, and they can determine whether to start at the beginning or the latest version, etc.

7. Instill variety using the genres column, as similar neighbor games are likely the same genre, the recommender can replace lower rated neighbor games and offer high scoring games from other genres.