

## Introduction:

Sentiment analysis is a very important technique in business when it comes to measuring marketing effort success on a product release or press release from a major corporation. Being able to gauge consumer reaction before, during, and/or after an event such as this on a global scale can really optimize and drive further business decision making.

In this project, I gauge consumer sentiment of a AAA video game reboot of a 20+ year old cult classic – Final Fantasy 7. The “Remake” game was just released to consumers on April 10<sup>th</sup>, 2020 after a roughly 5-year development cycle. The company, SquareEnix, took a risk with the release – as they noted there are elements of the original story that have changed slightly (not the major story line), as well as deciding to release the game in an episodic format (meaning the whole game won’t be released all at once). This demonstrates a risk the company took with their fan base, and within this project I will analyze if the consumer’s initial reception to this strategy has paid off.

Next, I dig into the content by analyzing how many times the name of a character shows up in a sample of tweets within the game itself. Here we’re interested in if there’s a general sense of “popularity” among the rank and file, which in turn may provide indicators as to which character could generate the most profit in alternative revenue streams such as merchandising. Given the nature of the episodic release and SquareEnix’s propensity to modify content from the original, this may drive further game development as a way to meet consumer demand – for example, if the crowd “demands more *Barret Wallace*” then he may show up more or have a bigger role to play in the upcoming episodes!

## Data Collection & Cleaning:

To get the process started, I’ll first need to collect tweets from Twitter about the game. I look at the official twitter account for SquareEnix, and identify the main hashtag being used is #FF7R, therefore this will be my main search string when collecting the tweets. I also notice a large amount of tweets using the #FinalFantasy7Remake hashtag, therefore it’s included in the collection process.

Using Twitter’s REST API, over the course of three sessions I attempt to pull two sets of 1,000 tweets using the hashtags mentioned above. Default behavior for the search is to provide a rolling 7-day window to return tweets, so in attempt to avoid duplication I make use of the “since” (date) parameter which allows the return of tweets on, or newer than, the date specified. For example, I queried tweets on 4/25 using the since date of 4/24 (one day after). The use of this parameter along with filtering of retweets is a strategy to minimize or avoid duplicate data altogether. It’s also important to note that the focus is just on the “English” language tweets, therefore this analysis may not be appropriate to extrapolate into geographic areas where English is not the primary language (note: SquareEnix’s headquarters is in Japan, and the sentiment there may be different).

Next, the raw tweet data must go through a cleaning process prior to being analyzed. Tweets can contain items like URLs, emojis & icons, and other non-essential punctuation or characters that may interfere or bias the sentiment analysis, therefore I remove URLs and non-essential punctuation, and convert emojis & icons to textual representations of itself<sup>[1](#)</sup>. A total of 5,836 tweets are collected, cleaned, and provided a sentiment score between -1 to +1, where negative represents a poor sentiment and positive represents a favorable sentiment. I also count the number of times the name of a character appears in this sample of tweets, as it will help us answer our second set of questions about character popularity. This information is written to .CSV, and is used in R for the analysis.

## Analysis:

To begin our sentiment analysis, we first decide how to treat those tweets that were labeled as “neutral” sentiment, score 0 - meaning they were neither positive nor negative. Because they do not help in answering our question relating to positive or negative, I decide to remove them from the data set before proceeding with the analysis, leaving 4,171 tweets for the analysis.

Next we check the distribution of the data using a histogram. We want to see a “bell-shaped” curve to the data which confirms our sample is normally distributed, and therefore we can perform the types of statistical tests that will help us address our questions. The histogram displayed in fig. 1 generally demonstrates this bell-shaped curve, and is slightly right skewed toward the positive sentiment end.

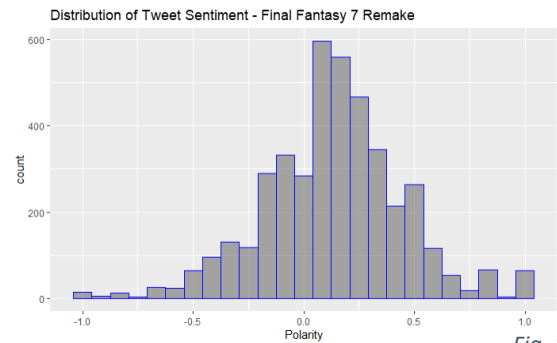


Fig. 1

Now that the assumption of normality is met, I perform a one-sample proportion test to check if the number of tweets that scored positive via the sentiment analysis is greater than 50% of the sample – i.e. the “majority”. At a 95% confidence level, there is enough evidence to claim that the majority of tweets in the entire Twittersphere regarding this video game are positive ( $p < 2.2e^{-16}$ ) – great news for SquareEnix since their game currently appears to be well received!

Next we perform a Chi-Squared Goodness of Fit test to identify if some characters are more “popular” than others – we want to know if they are all being discussed equally, or if some are being discussed more frequently than others. At a 95% confidence level, there is enough evidence to claim that there is an unequal use of character names in tweets ( $p < 2.2e^{-16}$ ), thus indicating that some characters are more “popular” than others (note: sentiment is not considered here; “popular” for good or bad reasons). The bar chart in fig. 2 reflects this unequal distribution, and indicates our heroes Cloud, Tifa, and Aerith round out the top 3 spots.

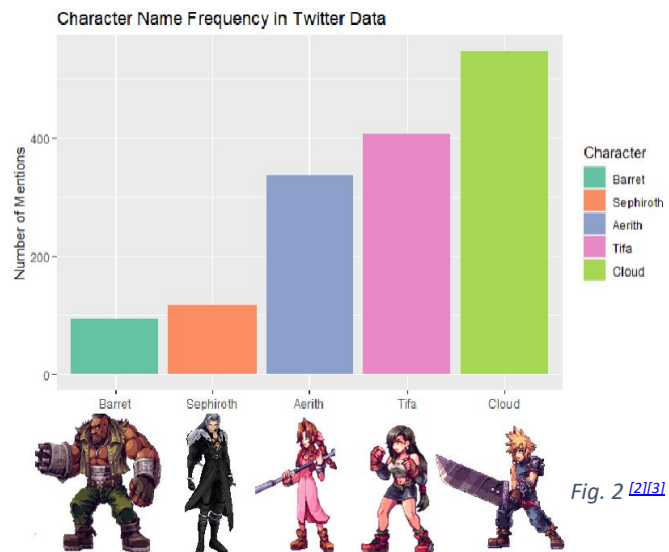


Fig. 2 [\[2\]\[3\]](#)

Our last question is if the character at the top (Cloud) is a “lone-wolf” in popularity – i.e. does he stand out as the people’s favorite? For this, we perform another one-sample proportion test to check if the frequency of his name use is greater than 50% in the sample collected. At a 95% confidence level, there is not enough evidence to claim that Cloud is the people’s favorite ( $p = 1$ ) – sorry Cloud, looks like you have some competition as the overwhelming majority favorite. Marketing for merchandising should lean toward the top 3 – Aerith, Tifa, and Cloud.

## Results & Conclusion:

The results of this project show that SquareEnix’s risk to do something different with a AAA remake of a cult-classic appear to be paying off. The majority of people are enjoying the game, and while there is no overwhelmingly popular character, there are also some clear favorites – sorry Barret, maybe next episode?

### References:

[1] user3082900 (2017, May 5). Python- replace Unicode emojis with ASCII characters. Stack Overflow. <https://stackoverflow.com/questions/43797500/python-replace-unicode-emojis-with-ascii-characters>

[2] Snaps. "Transparent Tifa Lockhart Png - Final Fantasy 7 Character Art, Png Download". *PNGITME*. [https://www.pngitem.com/middle/ibowhJx\\_transparent-tifa-lockhart-png-final-fantasy-7-character/](https://www.pngitem.com/middle/ibowhJx_transparent-tifa-lockhart-png-final-fantasy-7-character/). Accessed 28 Apr. 2020

[2] Gabriel Alcaraz. "Sephiroth Transparent Image - Sephiroth Smash Ultimate, HD Png Download". *PNGITME*. [https://www.pngitem.com/middle/iRhhiwi\\_sephiroth-transparent-image-sephiroth-smash-ultimate-hd-png/](https://www.pngitem.com/middle/iRhhiwi_sephiroth-transparent-image-sephiroth-smash-ultimate-hd-png/). Accessed 28 Apr. 2020