**Sentiment Analysis of Craft Breweries and Beers   
from Twitter**

Ken Atkins

Virginia Tech  
Masters of Information Technology  
CS5664 – Social Media Analytics

katkins9@vt.edu

Seunghyuk Baek

Virginia Tech  
Masters of Information Technology  
CS5664 – Social Media Analytics

sbaek44@vt.edu

Nathaniel Grefe

Virginia Tech  
Masters of Information Technology  
CS5664 – Social Media Analytics

griffles@vt.edu

**ABSTRACT**

Craft breweries and craft beers have been dramatically increasing in popularity in the United States over the last few years. There are a lot of people tweeting about craft beer and the goal of this project is to see if we can mine useful sentiment information from the twittersphere about craft breweries. One of the most popular styles of craft beers is called the “India Pale Ale” or “IPA”. To narrow down the analysis into a simple question that can be answered in the scope of this class, we are going to focus on the sentiment of people’s tweets about the IPA beer style and compare it to the sentiment of craft beer in general. We will do this by collecting a set of tweets that specifically mention IPAs and compare the sentiment of that dataset to the sentiment of a set of tweets that are about “Craft Beer” in general. If IPAs are truly one of the most popular styles of craft beer, then there should be a more positive sentiment about IPAs than about craft beer in general. We will also explore the popularity of other popular craft beer styles, such as Sours and Stouts to see how they compare to IPAs.

**CCS Concepts**

**Information systems → World Wide Web → Web searching and information discovery → Social recommendation**

**Keywords**

Craft Beer; Craft Breweries; Breweries; India Pale Ale; IPA; Sour Beer; Stout Beer; Sentiment Analysis; Twitter;

# INTRODUCTION

Craft breweries have been growing dramatically in popularity in the United States over the past few years. There are many new craft breweries starting in every region, and even with the growing popularity of craft beer, the competition between the different breweries and craft beers available is intense. It would benefit greatly a new brewery to know which styles of beer they should start brewing first in order to attract a market and be competitive.

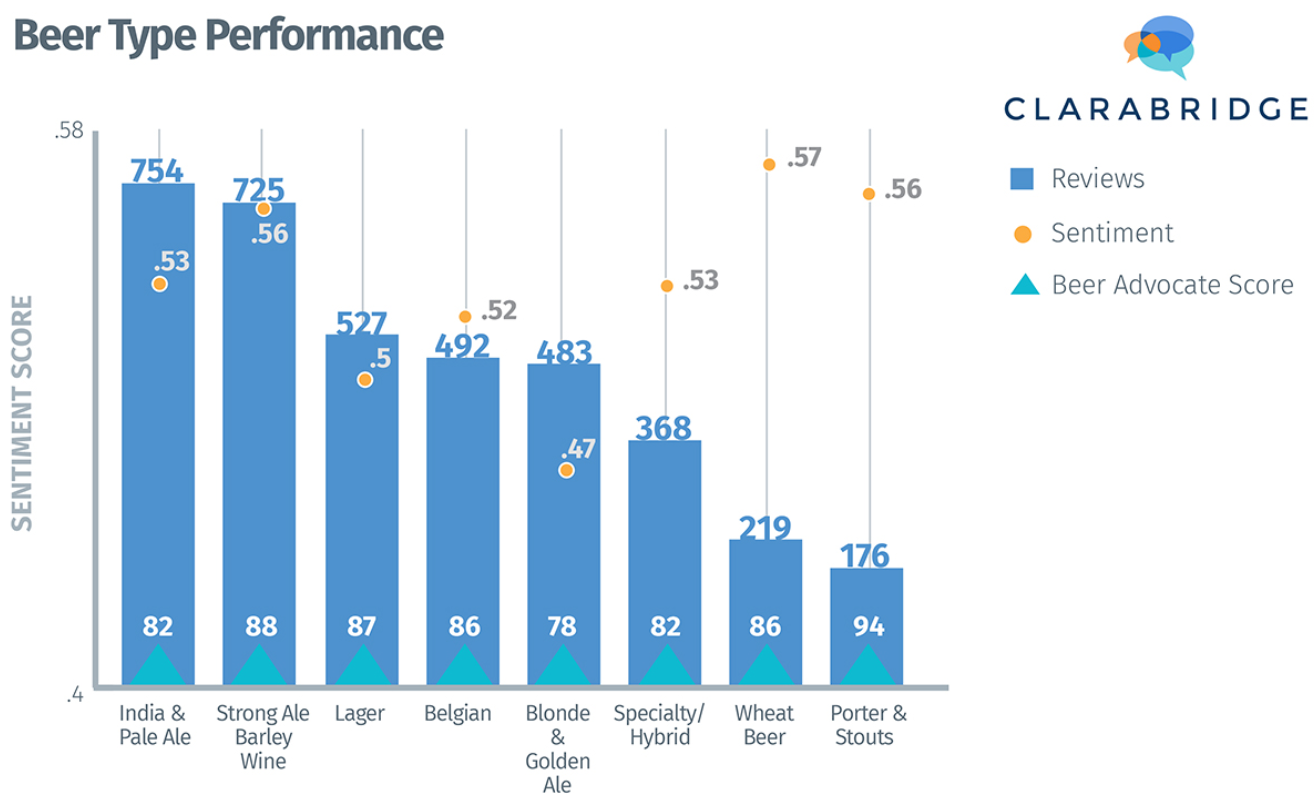
One of the most popular craft beer styles is the “India Pale Ale” or “IPA”, and many if not most craft breweries have started out with an IPA style beer as one of their launch styles. But is this really the best style for a new craft brewery to start with? Or is it a style whose popularity has waned? The primary goal of this project is to use social media to determine what the prevailing sentiment is toward the IPA style in particular and determine how it compares with the sentiment of craft beer overall or other styles.

The approach we will be taking is to mine twitter for tweets about craft beer in general and the IPA beer style in particular, and perform sentiment analysis on the text of the tweets about IPAs and compare the results to the overall sentiments about craft beer in general. We will also compare the sentiments of IPAs to sentiments of some other popular craft beer styles, such as “Sours” and “Stouts”. We will then make a recommendation about the desirability of including an IPA as an initial launch beer style for a potential new craft brewery.

# BACKGROUND / RELATED WORK

We did some web searches to see if there were any significant research available about using sentiment analysis to evaluate craft beer. We found some interesting articles and papers, but nothing exactly like what we are attempting, however, the search was not exhaustive (of course). One interesting online article was the following: [Sentiment of the Week: Prestige in Craft Beer](https://www.clarabridge.com/blog/whats-brewing-sxsw-analyzing-beer-popularity/) (Ramanan, 2015).

The source of the data used for the analysis in this paper was online reviews instead of twitter, but the article did analysis of sentiments of various styles and breweries based on the text of the reviews. One finding that validates our assumption that IPAs has been a popular beer style is shown in one of the charts from the article (Figure 1).



**Figure 1. Beer Type Performance (Sentiment)**

Another interesting article is [A craft beer made with social media, human sentiment and analytics](https://blogs.dxc.technology/2016/02/12/a-craft-beer-made-using-social-media-human-sentiment-and-analytics-and-low-on-calories/) (Nerney, 2016). This article is a brief description about a project where a team used IBM Watson to do sentiment analysis on tasting notes and reviews in combinations with the actual beer recipes to determine the “Perfect Beer Recipe”. Articles like this show that there is some interest in using sentiment analysis to determine how to brew more popular beer, which is the goal of our project.

# APPROACH

First, we collected tweets from Twitter with tweepy module. The topics used to collect tweets are ‘craft beer’, ‘IPA’, ‘stout’, and ‘sour beer’. The topic ‘craft beer’ resulted in 20158 unique tweets. The IPA topic initially resulted in 5951 unique tweets. Before began the sentiment analysis, the tweets from the topic ‘IPA’ were filtered with a collection of words (‘beer’ or ‘brewery’ or ‘drink’ or ‘brew’ or ‘bar’) to eliminate the tweets that were not discussing beer. After eliminating unrelated tweets from the collection, the number of tweets from the topic ‘IPA’ was reduced to 1713 tweets. There were also 2000 unique tweets about ‘sour beer’ and 3000 unique tweets about ‘stout beer’ collected.

To perform the tweet analysis, we used the NLTK module in python along with basic hashtag counting. The hashtag function of Twitter allows a user to add a word or phrase, designated by the “#” symbol. This hashtag is often added at the end of the tweet to represent the topic of the tweet or reference similar topics. Therefore, by looking at top hashtags, we can grasp what subjects are popular among the users. For example, if ‘Beer Party’ was the most frequent hashtag for the collected tweets for ‘IPA,’ we can assume a relationship between ‘Beer Party’ and IPA. The primary approach to evaluating the collected tweets was sentiment analysis. Sentiment analysis or opinion mining is computer analytic technique that determines whether a given text is positive or negative towards to the topic. The purpose of sentiment analysis is vary, therefore, researchers choose either machine learning or lexicon-based approach. In machine learning sentiment analysis, set of pre-labeled sample is provided to learning algorithm after the algorithm studied sample, new data set is provided to analysis. On the other hand, lexicon-based approach uses dictionaries. The dictionaries contains set of words and their polarity. The subject text is parsed into collection of words and trimmed by removing stop words. Then each parsed words are compared with the dictionary[4].

# EXPERIMENT

The main experiment we performed was sentiment analysis on tweets about Craft Beer and IPAs in particular.

## Source Datasets

The datasets were generated using the tweepy python library extracting tweets from our development twitter accounts using the credentials we generated as part of a class exercise. The python code used was basically the code that was shared with the class. We ran the extracts a couple of times over a two month period to get a spectrum of tweets over time. The results of each run was saved into a csv file. (An example of the python code for the extract process is included in the “FinalProject\_IPA\_Tweet\_Extraction.ipynb” file with the project submission).

We ran similar scripts with different search strings to generate the following datasets:  
  
 **Dataset Input Queries**

Craft Beer “Craft Beer”  
 IPA “IPA”, “India Pale Ale”  
 Sours “Sour beer”  
 Stouts “Stout”

For each dataset the various extract files were merged into single files, and duplicates were eliminated using the twitter ID (example code: Merge\_IPA\_datasets.ipynb). Additional work was done to eliminate unrelated data (for instance the concept of “International Public Affair” showed up in some of the tweets).   
  
 **Dataset Number of Tweets** Craft\_Beer\_Tweets.csv 21,792  
 IPA\_Tweets\_ALL.csv 1,827  
 sourBeer.csv 1,999  
 stoutBeer.csv 2,999

**Entity Extraction (Abandoned approach)**

Initially we tried to perform entity extraction on the datasets to determine if we could identify specific beers or breweries directly from the tweets. We found that while the entities extracted from the tweets did show some different beer names and brewery names, it was very difficult to be sure which entities were either beer names or brewery names. Here were the top 25 frequent entities:  


As you can see, most of the entities were generic like “Craft Beer”, “American”, “New” or locations. We determined that this was not really very useful.

Next, we located an official list of breweries and screen scraped a list of breweries from their website. This list was used to attempt to identify the brewery mentioned in the tweets, and this also did not produce useful results. There were only 279 tweets where we found an exact match on the brewery name from our list. Here are the top 24 breweries with matches:  


Only about 18 of these were actual breweries, and most of them had only 2 or 3 tweets. There were an additional 35 breweries that had only a single tweet in our data set.

After these disappointing results for entity extraction, we decided to change the approach of our analysis and simply do a sentiment analysis on the tweets in general using subsets of tweets for a few different standard craft brew styles.

The code we used for the entity extraction attempt is included with this paper in the following file::

CraftBeer\_Tweet\_Entity\_Extraction.ipynb)

**Hashtag Analysis**

Top 10 hashtag were extracted from each topics. In addition, hashtag ‘nan’ and hashtag ‘photo’ were omitted from the result since those hashtags were considered as non-related subjects.

**Craft Beer Tweets Top Hashtags**  
[ ('craftbeer', 751), ('CraftBeer', 185), ('beer,craftbeer', 150), ('FreedomBeerAintFree', 120), ('beer', 75), ('craftbeer,beer', 51), ('ThreeKingsPub,craftbeer', 41), ('Budget2018', 40)]

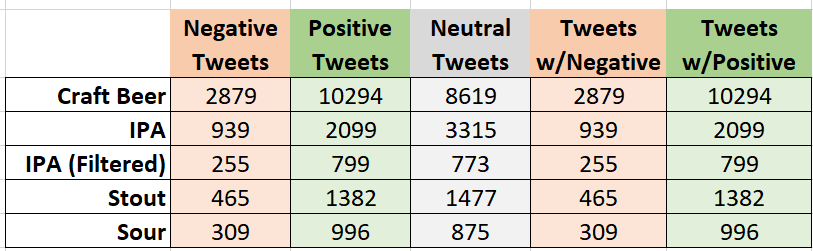
**IPA Tweets Top Hashtags**  
[ ('ibelieveinIPA', 104), ('BeerMenus', 31), ('IPA', 25), ('WorldBeerCup2018,GreenBattlesAPA', 25), ('auspol', 24), ('Workchoices,IPA, 21)(publicservice', 21), ('TexasBeer', 15), ('GreatBarrierReef', 15)]

**Sour Beer Top Hashtags**  
[('beer', 9), ('2RoadsBrewing', 6), ('OhioCraftBeer, 5),(DrinkBeerMadeHere', 5), ('sour,UnFiltered,vegan', 4), ('firstfriday', 4), ('blending,collab', 4), ('CraftBeer', 4), ('poty', 4)]

**Stout Top Hashtags**  
[('GoUtes', 12), ('DogCelebration,HappyHour, 10),(dogsoftwitter', 10), ('BeerMenus', 10), ('TopBrewsTues', 10), ('Elementary', 8), ('ICYMI,NFLDraft', 7), ('BYUvUTAH', 6), ('madbeerweek', 4)]

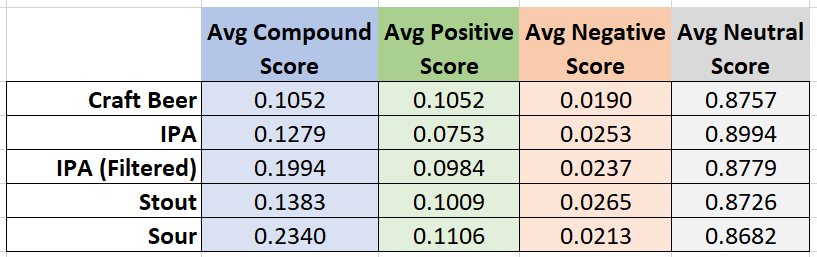
**Sentiment Analysis**

The raw counts of the tweet sentiments are shown in Table 1 for comparison and further reference. Interestingly enough, it seems that IPA’s suffer from comparatively low positivity in the collected tweets.



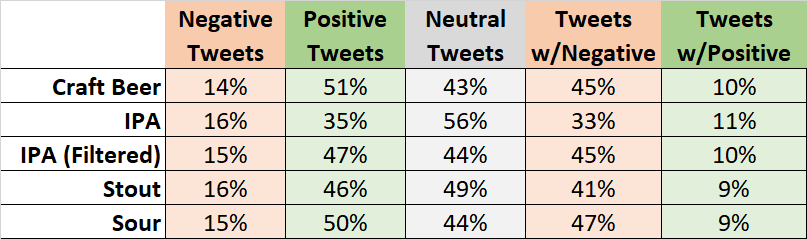
**Table 1. Sentiment Tweet Counts**

IPA received the lowest Average Positive Score and the second lowest Average Negative Score. It also had the lowest Average Compound Score at 0.1279 (see Table 2) of any dataset. By comparison, the Craft Beer Average Compound Score was 0.2335, the Sour Beer score was 0.2340, and the Stout Beer score was 0.1383. Filtering the IPA tweets did shift the sentiment in a more positive direction. The Average Compound Score rose to 0.1994 and the Average Positive Score rose to 0.0984. The Average Negative Score fell from 0.0253 to 0.0237.

**Table 2. Sentiment Scores Table**

Craft Beer showed more positive sentiment than IPA by all measures. The Average Compound score was 0.2335, the Average Positive Score was 0.1052, and the Average Negative Score was a lowly 0.0190. The tweet-set with the highest Average Compound Score was the Sour Beer set at 0.2340, which also had the highest Average Positive Score at 0.1106. The Stout Beer dataset showed generally negative sentiment, although not as extreme as the IPA set. Stout Beer had the second lowest Average Compound Score at 0.1383, the second lowest Average Positive Score at 0.1009, and the highest Average Negative Score at 0.0265.

For a clearer assessment of the total number of tweets per dataset showing positivity or negativity, Table 2 displays the percentages of the totals.

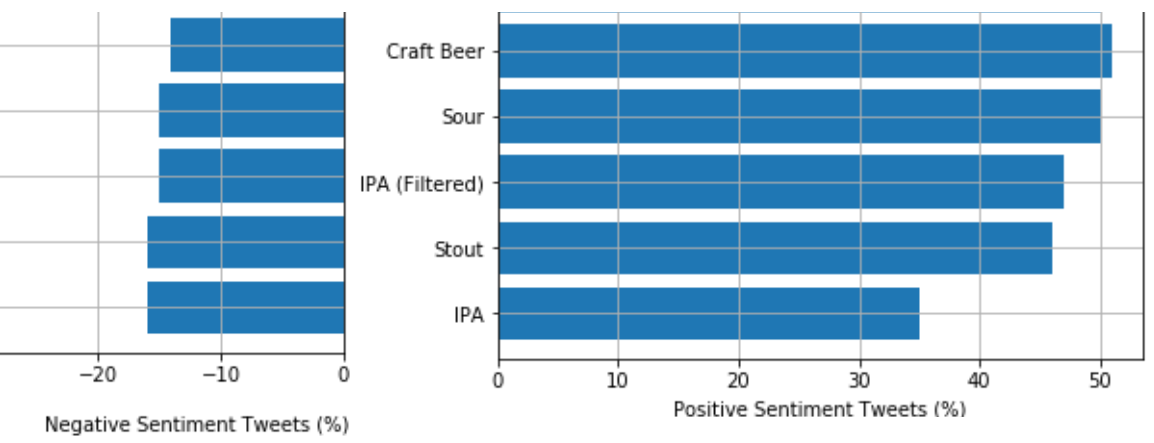
****

**Table 3. Percentage Table**

These ratios tell a slightly different story than the averages shown in Table 1. Number of Positive Sentiment Tweets refers to the whether the tweet displays any positive sentiment; it can also display negative sentiment; recorded in Number of Negative Sentiment Tweets. Here IPA shows the lowest percent of tweets that show some degree of positive sentiment (35%), but it also has the highest percent of tweets that show a positive compound score (11%). One explanation for this is that overall IPA’s are not tweeted positively about, but the tweets that are positive are very positive.

Craft Beer has the highest percentage of tweets showing some positive sentiment (51%) and the lowest percentage of tweets with a positive compound score (7%). This implies that the opposite of the IPA set - that overall Craft Beer is tweeted positively about but there are fewer highly positive tweets. Both Sour Beer and Stout Beer show higher percentages of positive compound scores (9%) than Craft Beer. They also both have high percentages of occurrence of positive sentiment, with Sour Beer at 50% and Stout Beer at 46%.

Figure 2 shows a chart with the negative sentiment at the left and the positive sentiment at the right



**Figure 2. Positive and Negative Sentiment by Style**

# CONCLUSION

## What did we learn?

Comparing the percentages between IPA and Craft Beer, there are some interesting conclusions that can be drawn. Overall, IPA’s are not generally tweeted about positively but the fewer positive tweets are more enthusiastic. This implies that may be a niche market: a smaller group of enthusiasts who feel very strongly and positively about IPA’s. Craft Beer, on the other hand, has broader positive support but a smaller dedicated group that feels positively and strongly. Knowing that Craft Beer is a general description for many types of beer, IPA included, one could deduce that Craft Beer is popular but that tweeters feel more passionately when discussing specific types of craft beer.

As a vendor looking at these numbers, it would seem that Craft Beer is popular and worth investing in. However, people seem to feel more passionately when discussing specific types of beer. Perhaps advertising the specific categories of beer, like “Sour’s Sold Here” would be more likely to attract interest than simply advertising “Craft Beer Sold Here.”

Furthermore, from hashtag analysis, it seems that people relate their tweets with social media event, local event or sport event. For example, 'GoUtes' and ‘NFL’ draft are sport events and interestingly both were found only in stout beer tweets. 'WorldBeerCup2018’ is local beer festival and ‘TopBrewsTues’ is the social media event where people post tweets about the beer they like. Another interesting hashtag is ‘BeerMenus’ which is a website for public to share and rate beer products.

If a vendor is looking for social media marketing via Twitter, it can be recommended to participate various event not only in local but also in social media. In addition, earning good reputation via reviewing website will help the company to grow.

## Ideas for Future Work

Here are some ideas for future work and enhancements that could be made to this analysis.

### More Data!

Because of the limitations of the free twitter feed and the fact that we only had access to a few weeks of historical twitter data, we did not end up with a very large dataset. We could really enhance this analysis if we were able to capture a larger number of tweets.

### Continue with Brewery Identification

I think with a larger dataset and some additional work the entity extraction approach to identify specific breweries for detailed sentiment analysis on each brewery would be possible. One of the problems with tweets is that since they are so short, people mention breweries in an abbreviated fashion by a nickname or one word of a longer brewery Name. If more work was put into a dictionary of breweries that included common nicknames or shortened names we might end up with a higher frequency of brewery identification in the

### Include Additional Beer Styles

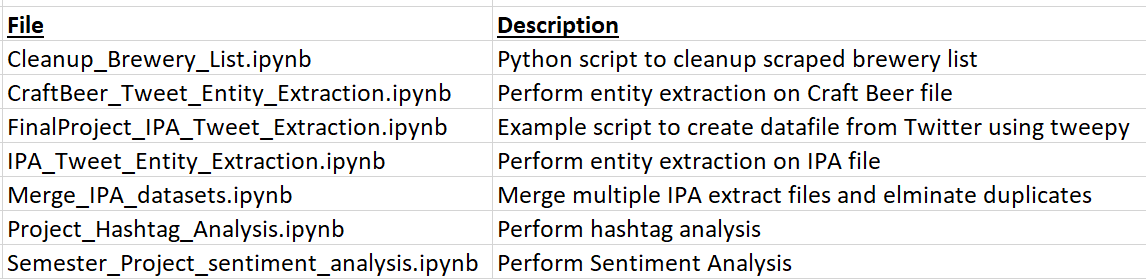
There are many different beer styles, and we only chose a few for this analysis. It might be interesting to add additional sytles (i.e. Lager, Steam, Pilsner, Belgian, Farmhouse, Red, etc.)

# REFERENCES

1. Ramanan, Dheepan, 2015. Sentiment of the Week: Prestige in Craft Beer. <https://www.clarabridge.com/blog/whats-brewing-sxsw-analyzing-beer-popularity/>
2. Nerney, Chris, 2016. A craft beer made with social media, human sentiment and analytics. <https://blogs.dxc.technology/2016/02/12/a-craft-beer-made-using-social-media-human-sentiment-and-analytics-and-low-on-calories/>
3. Brewers Association for Small and Independent Craft Brewers, 2018. Brewery Directory. <https://www.brewersassociation.org/directories/breweries/>
4. Li, Qin, 2015, Examining the Accuracy of Sentiment Analysis by Brand Monitoring Companies. University of Twente, 5th IBA Bachelor Thesis Conference.

# SUPPLEMENTARY MATERIAL

## Jupyter Notebooks



## Data Files

