The Twitter Presence of Hard Rock Hotel and Casino in Atlantic City

Nicholas Branco

10 August 2018

1 Abstract

This project examines the activity of Hard Rock Hotel and Casino in Atlantic City on Twitter, as well as investigates the users who are tweeting about Hard Rock Atlantic City. I found that the Twitter presence of Hard Rock Atlantic City was almost non-existent, with activity only happening on big days like opening. I also found that common words for the users included "New Jersey," "New York," and "love." The study was a success in showing the type of activity surrounding Hard Rock Atlantic City.

2 Executive Summary

After being hired by Hard Rock in Atlantic City, I took an interest in their social media presence. I scraped Twitter and collected tweets from May 25, 2018 to July 16, 2018 and found that Hard Rock's activity on the social media platform is not very expansive.

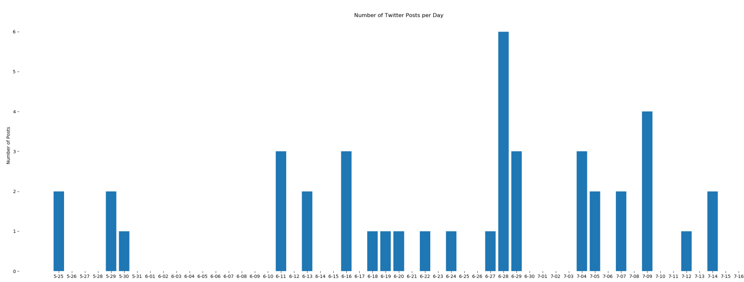

Figure 2.1: The frequency of tweets by @HardRockHCAC, Hard Rock Atlantic City's twitter account.

The above figure shows that the tweets came fairly sporadically, with a large increase on June 28, the day that the casino opened. A larger image of this figure is included later on in the document, however even with this smaller version one can see the massive gaps in activity.

I also collected data on the users, and used that information to form a list of words that are fairly common to users tweeting about Hard Rock. I obtained these words from the Twitter biographies of the users. The table on the right is a shortened version of another table later on in the document. Many people have the word ”Jersey” in their bio, indicating they are either from or living in New Jersey at the moment. Words such as ”love” and ”lover” as well as ”proud” show just what type of people are talking about the casino.

Word	**Count**
news	1105
love	629
new	343
world	340
lover	218
jersey	216
people	171
proud	165
...	...
school	53

Table 1: Popular words in Twitter bios.

Interestingly, the word ”school” does not appear very often, despite Hard Rock Atlantic City being in close proximity to Stockton University, as well as only around an hour away from Rowan University. This leads me to believe that there is not much of a base with students for Hard Rock on Twitter, contrary to what I initially thought. I believe that appealing to the small amount of students in nearby universities may result in the gradual building up of a base that the casino may be able to rely on in the off-seasons when students tend to be in the nearby colleges and vacationers are elsewhere.

3 Introduction

Social networks are a large part of our lives. Chances are, if you do not have a profile on a social network site, be it Facebook, Twitter, Instagram, or any other, you are still mentioned in a post, or in a picture posted on someone's page. For better or for worse, social media and social networks in general connect everybody on the planet in a way that was not possible fifty, thirty, or even just ten years ago.

Social media isn't just for the people either. Browse Twitter for five minutes, and you will probably see a tweet that either mentions a company or was written by a company. Businesses have taken to social media in an attempt to get people interested, and it works. You want everyone to know what kind of deals you offer today? It takes not even a minute for the thought to be turned into a post on social media, where it is immediately viewed by hundreds, thousands, or even (for the really big companies) millions of people. When I Google a restaurant or a store, it usually directs me to their Facebook page where I could find their menu, their hours, their deals for the day, and any other information a consumer needs to know. A strong social media presence is a sign of a strong company, at least from an advertising standpoint.

I began working for Hard Rock Hotel and Casino in Atlantic City in April of 2018. April 23, 2018 was my first day, in fact. It wasn't a particularly well-paying job, but it would get me through the summer. Hard Rock is, as you probably know, an international company based in Florida. Hard Rock has hotels, casinos, and cafes all over the globe. They are an ever-expanding entity, a true force in the hotel and casino worlds. Just the name Hard Rock lets people know that their experience will be one-of-a-kind.

Imagine my surprise, then, when I checked the Twitter for Hard Rock Atlantic City and found...nothing. I have an interest in Social Networks from a data science standpoint–they can tell you so much about a specific topic or a trend–so naturally, I was drawn to take a look at the casino's Twitter account.The hotel and casino I was working in wasn't open yet–they were set to open nearly two months after I began on June 28, 2018. Being only two months away from opening, I thought that Hard Rock Atlantic City would be pushing their brand all over social media in a push to sell rooms and reservations. This obviously baffled me. How could a casino brand known world wide not have at least a passable social media account to hype up their opening?

That was in the late days of April. I checked in a couple of weeks later and, interestingly, there were a few more tweets. It seemed like the account was getting a bit more active the closer we got to opening, but it was still unimpressive. I then had my goal: I wanted to see how the casino's social media presence will evolve from the days of being a little baby account. My expectations were simple: the social media presence will be fairly week in the weeks leading up to opening, then, about a week or two away from opening, the casino's presence on Twitter would drastically increase in order to drum up hype for the opening. Finally, as the days went by after opening, I expected the presence to level off.

The activity of Hard Rock wasn't the only thing I was interested in finding out, though. I wanted to know what kind of people come to the casino. After all, knowing who shows up allows the company to adjust their business accordingly. As a college student in the area, I wondered if people tweeting about Hard Rock were in college. While I myself don't participate in the nightlife, I know plenty of others I went to school with who do. Poor college kids or not, they will always find a way to afford alcohol and gambling. Knowing this about colleges, I wondered if there was enough support for colleges in the Hard Rock Twitter Network to justify partnering with local colleges and universities for deals on tickets, rooms, and meals. Just giving college students a discount on tickets could drastically increase business if the framework is already there.

I now had a purpose for my project. I would look into Hard Rock's Twitter network, finding out how often they post from the weeks leading into opening to the weeks after. I would also investigate what kind of people are talking about Hard Rock, whether they're college students, liberals, conservatives, or whatever else a user decides to throw into their Twitter biography. With that, let's get into it.

4 Methods

4.1 Collecting the Data

Every study begins with collecting data. For this particular project, I decided to use the R library rtweet in order to collect tweets. The R code I used was very short, provided to me by one of my professors, Dr. Clifton Baldwin. I will go line-by-line and detail how rtweet works so that you have an understanding. A complete appendix of my code is located at the end of this document.

```
library(rtweet)

appname <- "hard_rock"
key <- "XXXXXXXXXXXXXXXXX"
secret <- "XXXXXXXXXXXX"
```

To use rtweet, you must first create a Twitter account. You then have to visit https://developer.twitter.com/content/developer-twitter/en.html or just Google "Twitter API" to turn your account into a developer account. You can then create an app (mine here is called hard_rock) and obtain your key and secret key. In the code block above, I blocked out my key and secret key as they are account-specific and I don't want anyone using my account for their projects.

Loading the app name, key, and secret key in will allow me to create a Twitter token:

```
twitter_token <- create_token(
  app = appname,
  consumer_key = key,
  consumer_secret = secret
```

```
)

saveRDS(twitter_token, "~/.rtweet-oauth.rds")
```

This Twitter token will allow us to do multiple scrapes of Twitter without having to verify the program every time. One and done.

Once the verification was complete, I then began to scrape Twitter.

```
rstats_tweets <- search_tweets2(
    q = "Hard␣Rock␣Hotel␣OR␣Hard␣Rock␣Casino␣OR␣HardRockHCAC",
    n = 15000,
    parse = TRUE,
    type="mixed")

save(rstats_tweets, file="rtweetsyyyymmdd.RData")
```

Using rtweet, I scraped Twitter for every tweet that contained the words ”Hard Rock Hotel”, ”Hard Rock Casino”, and/or ”HardRockAC”, which is the Twitter handle of Hard Rock Atlantic City. I limited my search to 15000 tweets (more than enough wiggle room). I then saved my tweets to a file, naming them for the day that I ran them. The dates I ran this code are as follows:

1. June 3, 2018
2. June 9, 2018
3. June 14, 2018
4. June 19, 2018
5. June 25, 2018
6. July 1, 2018
7. July 7, 2018
8. July 11, 2018 and
9. July 16, 2018

This gives me nine total datasets. The program will collect Tweets from between a week and ten days prior to the code being ran. I chose the dates to allow for overlap and did what I could to make sure I was collecting every tweet I needed. With the overlap, I was easily able to rectify this by using the unique command in R, after loading in the data sets and merging them.

```
load(file="rtweets20180603.RData")
users1 <- users_data(rstats_tweets)
tweets1 <- rstats_tweets[,
    c("screen_name","text","created_at")]
load(file="rtweets20180609.RData")
```

```
users2 <- users_data(rstats_tweets)
tweets2 <- rstats_tweets[,
    c("screen_name","text","created_at")]
load(file="rtweets20180614.RData")
users3 <- users_data(rstats_tweets)
tweets3 <- rstats_tweets[,
    c("screen_name","text","created_at")]
load(file="rtweets20180619.RData")
users4 <- users_data(rstats_tweets)
tweets4 <- rstats_tweets[,
    c("screen_name","text","created_at")]
load(file="rtweets20180625.RData")
users5 <- users_data(rstats_tweets)
tweets5 <- rstats_tweets[,
    c("screen_name","text","created_at")]
load(file="rtweets20180701.RData")
users6 <- users_data(rstats_tweets)
tweets6 <- rstats_tweets[,
    c("screen_name","text","created_at")]
load(file="rtweets20180707.RData")
users7 <- users_data(rstats_tweets)
tweets7 <- rstats_tweets[,
    c("screen_name","text","created_at")]
load(file="rtweets20180711.RData")
users8 <- users_data(rstats_tweets)
tweets8 <- rstats_tweets[,
    c("screen_name","text","created_at")]
load(file="rtweets20180716.RData")
users9 <- users_data(rstats_tweets)
tweets9 <- rstats_tweets[,
    c("screen_name","text","created_at")]
```

The above code is what I used to load in the data necissary for this study. My goal was to have two data sets: one for user data and the other for data on the actual tweets. Thus, I merged each users dataframe and each tweets dataframe so that I only had two, and made them unique so as to get rid of any doubles that were collected:

```
users <- rbind(
    users1,
    users2,
    users3,
    users4,
    users5,
    users6,
    users7,
```

```
    users8,
    users9
)
tweets <- rbind(
    tweets1,
    tweets2,
    tweets3,
    tweets4,
    tweets5,
    tweets6,
    tweets7,
    tweets8,
    tweets9
)

users <- unique(users)
tweets <- unique(tweets)
```

And with that, I had my data. The only thing left was to analyze it.

4.2 The Move to Python

Since I am more familiar with Python than R, I decided this project would be a great opportunity to attempt moving the data sets from R to Python for me to perform the analysis in my language of choice. To do this, I used an R library and a Python package called "feather." After installing it, it takes a simple three lines of code to save my data sets in a format that Python can read.

```
library(feather)
write_feather(tweets, "tweets")
write_feather(users, "users")
```

To read these data sets in Python, it was a simple three more lines of code:

```
import feather

tweets = feather.read_dataframe( "tweets" )
users = feather.read_dataframe( "users" )
```

Feather works both ways–so If I decided to mess around with the data and send it back to R, I could. However, that is not a part of this study, though I look forward to seeing what possibilities feather, and a sort of integration of Python and R, can do for me in the future.

4.3 Hard Rock AC's Tweets

Since I saved my data sets as data frames, I needed some way to work with data frames in Python. I chose Pandas for this, as it is universally the package used to work with data frames.

To get Hard Rock AC's tweets over the period of time I wanted to study, I isolated the tweets created by @HardRockHCAC, the twitter account for the property. Then, I took the timestamps for those tweets and put them in a list called ”times”

The next step was to create a list of all of the dates and count up how many times Hard Rock tweeted out on those days.

```
dates = []

for post in times:
    post = str(post)
    dates.append(post[:post.find("␣")])

dates_posted = list(set(dates))
dates_posted = sorted(
    dates_posted,
    key=lambda d: tuple(map(int, d.split('-')))
)

number_of_posts = []
new_dates = []
for date in dates_posted:
    y = dates.count(date)
    number_of_posts.append(y)
    date = date[6:]
    new_dates.append(date)

dates_posted = new_dates
```

I performed the same steps with all of the tweets, not just HardRockHCAC in order to get an understanding of the social media activity of the whole network, rather than just the central user.

4.4 What are People Tweeting About Hard Rock?

In order to find out who was tweeting (and what they were tweeting about), I had to get a list of the users I wanted to look at.

```
user_data = list(tweets["screen_name"])
unique_users = list(set(user_data))
unique_users = sorted(unique_users)

number_of_posts = []

for user in unique_users:
    y = user_data.count(user)
```

```
    number_of_posts.append(y)

import pandas as pd

df = pd.DataFrame(
    {'Screen_Name': unique_users,
     'Number␣of␣Tweets': number_of_posts
    })

df = df.sort_values(by='Number␣of␣Tweets', ascending=False)
```

I decided to sort the users by the number of Tweets they had. I then looked at the top twenty tweeters and removed sixteen of them due to their accounts being journals, news sites, or Hard Rock. I wanted to look at only the citizens tweeting about Hard Rock, and what they were tweeting about.

```
df_negs = df[:20]
df_negs = df_negs.drop([3423,3446,1432,4419])

for name in df_negs.Screen_Name:
    tweets = tweets[tweets.screen_name.str.contains(name) == False
        ↪ ]
```

I used the package "re" in order to separate each tweet into a list of words, then I used "itertools" to pull all of those lists of words into one giant list.

```
import re

allwords = []
for tweet in tweets.text:
    string = tweet
    wordlist = re.sub("[^\w]", "␣", string).split()
    allwords.append(wordlist)

import itertools

allwords = list(itertools.chain.from_iterable(allwords))
```

Finally, using a similar code as the one in the section above, I counted every word as it appeared and found the most common words.

```
words = list(set(allwords))

number_of_times_written = []
for word in words:
    y = allwords.count(word)
    number_of_times_written.append(y)
```

```python
languages = language[language.number >= 1000]

for idx in languages.index:
    language = language.drop([idx])

languages = language[language.number >= 300]

common_words = [
    'and',
    'but',
    'a',
    'me',
    'my',
    'it',
    'for',
    'you',
    'be',
    'so',
    'to',
    'too',
    'is',
    'if',
    'be',
    'will',
    'that',
    'what',
    'from'
]


for x in common_words:
    language = language[
        language.word.str.contains(x) == False
    ]

languages = language[language.number >= 200]
```

This required a lot of trial and error...the "common_words" list was created to remove the clutter from the table that I printed out. I also removed anything with over 1000 uses as, looking at the list, it was mostly prepositions, "RT" (which signifies that the tweet was a retweet), and the words "Hard", "Rock", "Hotel", "Casino", "Atlantic", and "City.

4.5 Who are the People Tweeting About Hard Rock?

For this section in my project, I wanted to look only at English-speaking users. Since Hard Rock is a worldwide corporation, it could be interesting to expand this section to all languages, provided I could read them all.

Similar to the section above, I took all of the biographies of the English Twitter users and counted up the words used in them, omitting some common words using trial and error as I did.

```
en_users = users[users.account_lang == 'en']

desc = []
for bio in en_users.description:
    wordlist = re.sub("[^\w]", " ", bio).split()
    for word in wordlist:
        word = word.lower()
        desc.append(word)

bio = list(set(desc))

number = []
for word in bio:
    y = desc.count(word)
    number.append(y)

language = pd.DataFrame({'word' : bio,
 'number' : number,
  }, columns=['word', 'number'])
```

From there, it is similar to the code above. The common words I omitted were a more extensive list for reasons I will expand on in the Results section. The entire list can be found in the appendices.

5 Results

The results of my study give two graphs and two tables. The graphs I will present rotated 90 degrees since they are too wide to actually fit on the page.

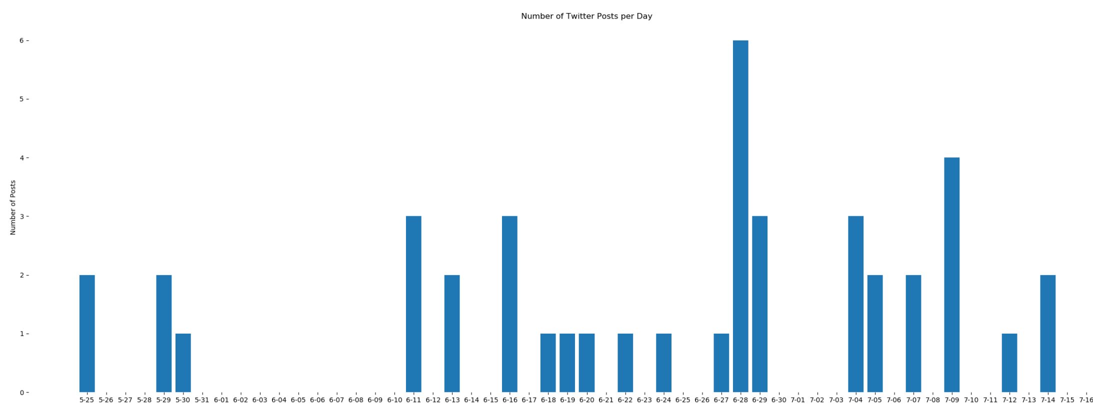

Figure 5.1: Tweets by Hard Rock Atlantic City's twitter account from May 25, 2018 to July 16, 2018.

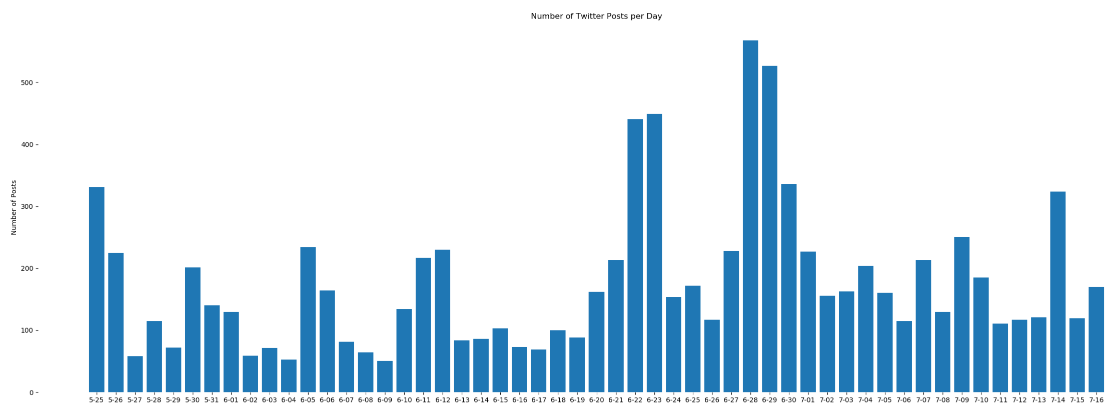

Figure 5.2: Tweets about Hard Rock from May 25, 2018 to July 16, 2018. This includes users from all over the globe.

Table 2: Popular Words Tweeted

Word	**Count**
Seminole	669
new	619
open	431
Hollywood	381
word	352
EXCLUSIVE	341
July	338
Opening	336
World	334
Trump	329
Event	262
friends	255
Jersey	249
FL	248
Hulk	244
show	215
Big	214
2018	208

This table displays the popular words that were in the tweets collected. Many other common words, such as ”Hard” and ”Rock” were removed from this list, and I omitted many from the list printed out that were not interesting, such as the word ”up.”

Table 3: Popular Words in Twitter Bios

Word	**Count**
news	1105
love	629
new	343
world	340
lover	218
jersey	216
people	171
proud	165
producer	160
food	144
...	...
nyc	67
18	63
york	60
shore	57
blogger	57
school	53

This table displays the popular words that were in the biographies of the users collected. I included the words that were interesting to find and, as noted by the ellipses, skipped a large amount in between. Many of the words omitted are prepositions, pronouns, conjunctions, determiners, etc., that are not exactly what I am looking for in this study. Most of what is in the list above is a verb, noun, or an adjective.

6 Analysis

6.1 Hard Rock Atlantic City's Twitter Presence

Figure 5.1 was the information that I was looking for initially when I began this project. A complete graph of Hard Rock AC's twitter presence from the weeks before opening to the weeks after opening. Initially, I hypothesized that there would be a small amount of Twitter presence up until the week or days leading up to opening where there would be a spike in activity and then the presence would then level off to a moderate amount.

Needless to say, this is not what was found. The first part is partially correct–there was a very small amount of posts made in the beginning of the time frame used for this study. By small amount, I mean almost none! Between May 25, 2018 and June 10, 2018 there were five posts from Hard Rock Atlantic City's Twitter account. However, in the days leading up to the opening, between June 16 and June 27, there were around 9 posts which then gives way to a spike of six instances of activity on the 28th; opening day.

6.2 The Rest of the World

While the Hard Rock AC Twitter was a bit sporadic, the rest of the world did not rest. Hard Rock's time in the spotlight seems to be fluctuating as indicated by the periodic nature of Figure 5.2. The dates where there are more activity are dates where shows are playing at a Hard Rock somewhere in the world. Note that this figure is for anyone talking about Hard Rock as a company, not just the Atlantic City property.

On June 22 and 23, there was a much larger spike in activity. I initially thought this was due to the Friends and Family day at Hard Rock AC, but that began on the 25th. This was the same day as a Pitbull concert in Sioux City. The much larger spike could be attributed to Pitbull being a major name and that he would be playing a show in Atlantic City the next week.

The back end of this data, from the 28th onward, matches my initial expectations for Hard Rock AC, where there was a very large spike that was followed by a more level amount of activity. The amount of activity on the 28th reached well over 500 instances which was more than double what other, smaller increases in activity reached.

6.3 Popular Words in Tweets

After removing all of the prepositions and whatnot, I ended up with Table 2. The table shows the words that are popular in tweets, anything above 200 uses. Looking at the table, it doesn't give much. We can see that Seminole is one of the most popular words, which makes sense since the Seminole Tribe are the owners of Hard Rock. We can also see words like Hollywood and FL. Hollywood, Florida is the location of another Hard Rock, and pretty much the "home base" of the company.

The words ”open,” ”Opening,” and ”Trump” all appear on this list along with ”Jersey” and ”2018.” This is most likely due to the opening of Hard Rock in Atlantic City, New Jersey in 2018 in the building that used to be called the Trump Taj Mahal.

Some words, words like ”friends” and ”show,” are words that can tell us a bit about the people who are tweeting about Hard Rock. Unfortunately, there aren't many of these kinds of words on this list, nor do they have a large presence under 200 uses. The word I am most interested in, though, is ”Hulk.” Is this referring to Hulk Hogan, or the Incredible Hulk? I am not sure, but the fact it beat out 2018 in uses is baffling.

6.4 Who Are the People Tweeting About Hard Rock?

Unlike the previous table, Table 3 gives me exactly what I was looking for. This table lays out the most popular words in Twitter users' biographies. People tend to post their jobs, hobbies, and identities within these bios, so this portion of the study was crucial in understanding what type of people are talking about Hard Rock on Twitter.

The top word is ”news.” This is to be expected–many news sites and reporters have Twitter accounts and would be talking about the brand new opening of a major casino. I did my best to eliminate this from the table, but ultimately decided to keep it in to compare it to other words. It is nearly double the second word, so one can assume that people affiliated with news are talking about Hard Rock more than others. We can also see this with ”blogger” also being on this list, though ”blogger” invokes a less-professional vision of a news source than the word ”news” does.

A grouping of words I find interesting are ”love,” ”lover,” and ”proud.” Whether ”proud” is referring to a proud parent, the Pride Movement, or anything else, these three words still tell a lot about the users. These words evoke a sense of general happiness and bliss, and is a clear indication of the type of users that Hard Rock attracts.

Some locations are on the list as well. We have ”jersey” and ”york” along with the word ”new,” which obviously means ”New York” and ”New Jersey” form a large part of these biographies. This is very interesting, as this means that many people talking about Hard Rock are in New York or New Jersey rather than other places in the world.

Coming in last place, we have ”school.” The inspiration for this project was partly to see what kinds of people are interested in Hard Rock. I proposed that students would be a large part of the users, however they were nowhere on this list, leaving only ”school” with 53 uses.

7 Discussion

In a small way, I was correct about Hard Rock AC's Twitter activity. However, I am truly surprised at just how little activity overall was recorded. A spike in Twitter activity is what I expected, but I hoped for more than six posts being the high point of the social media presence. I am also baffled at the days leading up to and the days immediately after the opening day. Little to no activity. One would think that a company would want to constantly drum up hype in the days before opening, but alas, that was not the case. Even the weeks after opening were a bit dull as well. This is important, since social media is a great way to connect with consumers and let them know what is happening. Perhaps Twitter was not Hard Rock's main social media platform, but they could stand to increase their profile to reach those who only use Twitter for social media, such as myself.

I am a bit disappointed in the results of the popular words tweeted out. I am not sure what exactly I was expecting, but I was hoping to at least get a glimpse into the lives of the users. In hindsight, this was not the best way to do this; rather, looking through their biographies is what I needed for this type of project. That investigation led into some interesting conclusions and provided a bit more insight as to who tweets about Hard Rock.

I was hoping to see students tweeting about Hard Rock, but unfortunately they weren't a large presence on Twitter. If they had been, I would have suggested that Hard Rock use that to their advantage by offering deals to students and partnering with colleges and universities to appeal to a base that was already there. Even though this was not the case, I still think this would be a cool idea to attempt in an effort to possibly bring in more students, and create the base themselves. Considering Hard Rock Atlantic City is close to Stockton University and within an hour of Rowan University, I think this would be an idea that the company should at least entertain, especially in the autumn and winter when students will be around and vacationers will be absent.

Future studies into this topic would include a sentiment analysis of the tweets. I made an attempt at doing this, however I could not get it working in time for the due date of this project. A sentiment analysis could show the overall emotions behind the tweets, though I expect it to be mostly good based on what we found here. Many posts for Hard Rock are generally relaying information about shows. I would also, in the future, expand the analysis of bio words to all languages if I could translate them, or if I could read most of them. Another endeavor that could follow this study is looking at the locations of the tweeters and seeing if there are common places that Hard Rock guests reside in.

As with all social media analysis, this study relies on collecting user data. Since the data is totally subject to any mistakes users make, be it spelling errors, mistyping, forgetting to add a handle, it is likely that this analysis has missed some users and words. I certainly missed many people by only focusing on Twitter, and separate investigations into Facebook and other social media sites may provide very different results. The code itself also wasn't the most efficient way to perform this, as it required some manual removal and trial and error for the results I was looking for. Despite these small hiccups, I have shown that social media analysis can show just who is talking about a particular subject, or observe the activity of users. I firmly believe that this type of analysis is very important to not only a company, but the consumer as well. Knowing how one interacts with the other is crucial to creating and maintaining a working business model.

Appendices

A R Code to Collect Tweets

```
# load twitter library - the rtweet library is recommended now
    ↪ over twitteR
library(rtweet)

# Enter your Twitter credentials
appname <- "hard_rock" # name I assigned my app in Twitter
key <- "XXXXXX"
secret <- "XXXXXXXXXXX"

# create token named "twitter_token"
twitter_token <- create_token(
  app = appname,
  consumer_key = key,
  consumer_secret = secret)

saveRDS(twitter_token, "~/.rtweet-oauth.rds")

# Scrape tweets
rstats_tweets <- search_tweets2(
    q = "Hard␣Rock␣Hotel␣OR␣Hard␣Rock␣Casino␣OR␣HardRockHCAC",
    n = 15000,
    parse = TRUE,
    type="mixed"
)

# Save tweets to a RData file
save(rstats_tweets, file="rtweetsyyyymmdd.RData")
```

B R Code to Clean Data

```
library(rtweet)

load(file="rtweets20180603.RData")
users1 <- users_data(rstats_tweets) # Gets data relevant to the
    ↪ users from the rstats_tweets dataframe
tweets1 <- rstats_tweets[,c("screen_name","text","created_at")] #
    ↪  I keep these in one dataframe for later.

load(file="rtweets20180609.RData")
users2 <- users_data(rstats_tweets)
tweets2 <- rstats_tweets[,c("screen_name","text","created_at")]

load(file="rtweets20180614.RData")
users3 <- users_data(rstats_tweets)
tweets3 <- rstats_tweets[,c("screen_name","text","created_at")]

load(file="rtweets20180619.RData")
users4 <- users_data(rstats_tweets)
tweets4 <- rstats_tweets[,c("screen_name","text","created_at")]

load(file="rtweets20180625.RData")
users5 <- users_data(rstats_tweets)
tweets5 <- rstats_tweets[,c("screen_name","text","created_at")]

load(file="rtweets20180701.RData")
users6 <- users_data(rstats_tweets)
tweets6 <- rstats_tweets[,c("screen_name","text","created_at")]

load(file="rtweets20180707.RData")
users7 <- users_data(rstats_tweets)
tweets7 <- rstats_tweets[,c("screen_name","text","created_at")]

load(file="rtweets20180711.RData")
users8 <- users_data(rstats_tweets)
tweets8 <- rstats_tweets[,c("screen_name","text","created_at")]

load(file="rtweets20180716.RData")
users9 <- users_data(rstats_tweets)
tweets9 <- rstats_tweets[,c("screen_name","text","created_at")]

users <- rbind(users1,users2,users3,users4,users5,users6,users7,
    ↪ users8,users9) # Merge all of the users data frames into
```

```
    ↪ one
tweets <- rbind(tweets1,tweets2,tweets3,tweets4,tweets5,tweets6,
    ↪ tweets7,tweets8,tweets9) # Merge all of the tweets data
    ↪ frames into one

rm(tweets1,tweets2,tweets3,tweets4,tweets5,tweets6,tweets7,
    ↪ tweets8,tweets9,users1,users2,users3,users4,users5,users6,
    ↪ users7,users8,users9,rstats_tweets) # Clean up a bit

users <- unique(users)
tweets <- unique(tweets)

library(feather)
write_feather(tweets, "tweets")
write_feather(users, "users")
```

C Python Code to Analyze Data

```
import feather

tweets = feather.read_dataframe( "tweets" ) # Read in tweet data
users = feather.read_dataframe( "users" ) # Read in user data

times = tweets["created_at"] # list of all timestamps a tweet was
    ↪ created using

dates = []

# Convert timestamp into dates
for post in times:
    post = str(post)
    dates.append(post[:post.find("␣")])

dates_posted = list(set(dates))
dates_posted = sorted(dates_posted, key=lambda d: tuple(map(int,
    ↪ d.split('-'))))

# Count the amount of time a date shows up
number_of_posts = []
new_dates = []
for date in dates_posted:
    y = dates.count(date)
    number_of_posts.append(y)
    date = date[6:] # Remove the 2018 from the dates
    new_dates.append(date)

# Get rid of the 2018
dates_posted = new_dates

# Make a Bar Graph
import matplotlib.pyplot as plt; plt.rcdefaults()
import numpy as np
import matplotlib.pyplot as plt

y_pos = np.arange(len(dates_posted))

fig, ax = plt.subplots(figsize=(30,10))
plt.plot(dates_posted, number_of_posts)
plt.xticks(y_pos, dates_posted)
plt.ylabel('Number␣of␣Posts')
```

```python
plt.title('Number of Twitter Posts per Day')

# Remove the box...it adds nothing to the visualization
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)
ax.spines['bottom'].set_visible(False)
ax.spines['left'].set_visible(False)

plt.savefig("total_activity.png") # Save the graph

plt.show() # Show the graph

# List of all of the screen names
user_data = list(tweets["screen_name"])
unique_users = list(set(user_data))
unique_users = sorted(unique_users)

number_of_posts = []

for user in unique_users:
    y = user_data.count(user)
    number_of_posts.append(y)

import pandas as pd

df = pd.DataFrame(
    {'Screen_Name': unique_users,
     'Number of Tweets': number_of_posts
    })

df = df.sort_values(by='Number of Tweets', ascending=False)

df_negs = df[:20] # Take a look at the top twenty in the data
    ↪ frame. This was done through trial and error.

df_negs = df_negs.drop([3423,3446,1432,4419]) # We want to keep
    ↪ these accounts

# Remove the rest of the top 20 screen names
for name in df_negs.Screen_Name:
    tweets = tweets[tweets.screen_name.str.contains(name) == False
        ↪ ]

# Split up the tweets into words
import re
allwords = []
```

```
for tweet in tweets.text:
    string = tweet
    wordlist = re.sub("[^\w]", "␣", string).split()
    allwords.append(wordlist)

import itertools

allwords = list(itertools.chain.from_iterable(allwords))

words = list(set(allwords))

# Count up all of the words
number_of_times_written = []
for word in words:
    y = allwords.count(word)
    number_of_times_written.append(y)

# Organize into a table
language = pd.DataFrame({'word' : words,
 'number' : number_of_times_written,
  }, columns=['word', 'number'])

# Some trial and error...this was all done in Jupyter Notebook.
languages = language[language.number >= 1000]

for idx in languages.index:
    language = language.drop([idx])\

languages = language[language.number >= 300]

# Common words to remove
common_words = [
    'and',
    'but',
    'a',
    'me',
    'my',
    'it',
    'for',
    'you',
    'be',
    'so',
    'to',
    'too',
    'is',
    'if',
```

```
    'be',
    'will',
    'that',
    'what',
    'from'
]


for x in common_words:
    language = language[language.word.str.contains(x) == False] #
        ↪ Remove common words

languages = language[language.number >= 200]


## Same thing as above, but for English accounts with biographies
    ↪  rather than tweets
en_users = users[users.account_lang == 'en']

desc = []

# Split the bio into words
for bio in en_users.description:
    wordlist = re.sub("[^\w]", "␣", bio).split()
    for word in wordlist:
        word = word.lower()
        desc.append(word)

bio = list(set(desc))

# Count the words
number = []
for word in bio:
    y = desc.count(word)
    number.append(y)

language = pd.DataFrame({'word' : bio,
 'number' : number,
  }, columns=['word', 'number'])

print(languages)

# Common words to remove
common_words = [
    'and',
    'but',
```

```
    'a',
    'me',
    'my',
    'it',
    'for',
    'you',
    'be',
    'so',
    'to',
    'too',
    'is',
    'if',
    'be',
    'will',
    'that',
    'what',
    'from',
    'the',
    'of',
    'i',
    'https',
    'in',
    'co',
    't',
    'on'
    's',
    'we',
    'm'
]

for x in common_words:
    language = language[language.word.str.contains(x) == False] #
        ↪ Remove common words

language = language.sort_values(by=['number'],ascending=False) #
    ↪ Sort decending

print(language[language.number >= 50]) # Display the table

tweets = feather.read_dataframe( "tweets" ) # Re-load the data

hardrock = tweets[tweets.screen_name == 'HardRockHCAC'] # Take
    ↪ only Hard Rock's data

times = hardrock["created_at"] # Take the timestamps
```

```
dates = []

# Split the timestamps into dates
for post in times:
    post = str(post)
    dates.append(post[:post.find("␣")])
dates_posted = list(set(dates))
dates_posted = sorted(dates_posted, key=lambda d: tuple(map(int,
    ↪ d.split('-'))))

# Count the dates
number_of_posts = []
new_dates = []
for date in dates_posted:
    y = dates.count(date)
    number_of_posts.append(y)
    date = date[6:] # Get rid of the 2018
    new_dates.append(date)

dates_posted = new_dates

# Plot the graph
import matplotlib.pyplot as plt; plt.rcdefaults()
import numpy as np
import matplotlib.pyplot as plt

y_pos = np.arange(len(dates_posted))

fig, ax = plt.subplots(figsize=(30,10))
plt.bar(dates_posted, number_of_posts)
plt.xticks(y_pos, dates_posted)
plt.ylabel('Number␣of␣Posts')
plt.title('Number␣of␣Twitter␣Posts␣per␣Day')

#Remove the Spines
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)
ax.spines['bottom'].set_visible(False)
ax.spines['left'].set_visible(False)

plt.savefig("hard_rock_posts.png") # Save

plt.show() # Show
```