



# FAN ENGAGEMENT ANALYSIS

NBA EASTERN CONFERENCE FINALS 2020

---

## PREPARED BY:

KATHY CHEN  
HUNG-LI CHUANG  
JI JIE  
SACHIN SHAH  
RACHEL SUNG

DSO 599: SPORTS PERFORMANCE ANALYTICS

# Table of Contents

Table of Contents	1
Executive Summary	2
Objective	2
Project Outline	2
<b>Data Preparation</b>	<b>3</b>
Overview	3
Data Scraping	4
Data Priming	7
Variable Creation	7
Description of Data	9
Overall Description	9
Description of Variables	9
Analysis	123
Sentiment Analysis	133
Statistical Modeling for Retweet Analysis	188
<b>Conclusions</b>	<b>255</b>
Results	255
Recommendations	255

## Executive Summary

---

### Objective

Nowadays, fans can engage in sports through a multitude of ways such as attending games, watching games through their league related TV subscriptions, enrolling in loyalty programs, or posting online real time reactions. With so many different avenues of fan interaction, it's becoming tougher for teams to track fan engagement and optimize their interaction with the fan.

This is especially true with online fan engagement. With an ever growing number of social media platforms being built on newer technologies, teams not only have to produce content on these platforms, but make sure that the content is engaging, unique, and interactive for the fan. Luckily, the analytical space has grown alongside social media, and it has become attainable to track and optimize online fan engagement.

In this project, our goal is to perform exploratory analysis of fan engagement via Twitter during Game 5 of the NBA Eastern Conference Finals between the Miami Heat and the Boston Celtics on September 25th, 2020 where the Celtics won 121 - 108. The Miami Heat were 1 win away from advancing to the NBA Finals but were not able to hold on after the Celtics made their run in the 3rd Quarter. By conducting sentiment analysis and statistical modeling, we can identify trends and provide recommendations to the two teams to help them maximize their fan engagement on Twitter as well as the NBA to enhance the overall game experience for the fans.

### Project Outline

Here is a high level overview of our work:

- Data Preparation:
  - We scrapped the data from Twitter for the Eastern Conference Finals 2020 Game 5 in Python. We searched for specific hashtags and extracted the data into an excel file, which contained over 12,145 unique tweets and retweets.
  - We removed duplicates and records which might not be related to Game 5. We then tagged the tweets to their respective team based on the hashtags and mapped them to the specific play of the games.
  - We created 10 additional variables in order to better interpret the data and conduct our further analysis.

- Sentiment Analysis:
  - We used the package “sentimentr” in R, which returned classified positive and negative sentiment as well as a numerical sentiment score for each tweet.
- Retweet Analysis:
  - We built multiple linear regression models and logistic regression models to see which variables in our dataset were significant to our dependent variable that indicated if the tweet was retweeted.
- Conclusion:
  - There were four variables that we found significant to our dependent variable: the counts of hashtags, negative tweets, positive tweets, and score differential.

## Data Preparation

---

### Overview

In collecting data for Game 5 for the Eastern Conference Finals between the Miami Heat and the Boston Celtics on September 25th, we looked to extract as many relevant & real-

time tweets as possible by focusing on specific team and NBA related hashtags. We then massaged the data within Excel and created new variables that would allow us to better gather insights in R.

### Data Scraping

The data was collected via web scraping in Python in which we repurposed code used in a previous class and tailored it to fit our needs. We were able to extract real time tweets by connecting to Twitter's API via keys and tokens and specified which variables we wanted to extract. We limited our scope by searching for specific hashtags and extracted the data into an excel file.

```
In [1]: import pandas as pd
import numpy as np
import requests
import json
import re, textwrap
import argparse
```

```
In [*]: !pip install TwitterAPI
```

```
In [73]: #NOTE: Must have TwitterAPI Installed
from TwitterAPI import TwitterAPI
from TwitterAPI import TwitterPager
```

```
In [74]: api_key="LhCSedmn8rIMMipticrS4g350"
api_secret_key="y4Ws7XPuV9cms0S3aId4LPIS5kMiA7nFzldBb10Zo5lxfmC032"
bearer="AAAAAAAAAAAAAAAAAB91GAEEAAAAA24kbznYamgCYk%2FRMiXyttUXILIU%3DIazdeos2q8xEOP4GkXAaK6gX37u71x0Gq6hrg9MT1z6mtAMDg0"
AccessToken="1285002674597851137-IXKri9EFCqh9EozdKZqTqKL7BSXDWI"
AccessTokenSecret="BjJM0o0X1hZ8hjUwYDUFeSzWrSPJgr1D5wXfGwQeZh8j"
```

```
In [75]: # Setup Twitter API Client
api = TwitterAPI(api_key,
                 api_secret_key,
                 AccessToken,
                 AccessTokenSecret)
```

```
In [78]: #Setup Initial Empty Data Frame
header = [
    "created_at",
    "tweet",
    "search_term",
    "hashtags",
    "language",
    "retweet_count",
    "user_id",
    "is_verified_user",
    "user_follower_count",
    "user_friend_count",
    "user_location",
]
index = np.arange(0)
df = pd.DataFrame(columns=header, index = index)
```

```
In [82]: count = 0 # counter
print("requested tweets for hashtag is limited to {} tweets".format(limit))
for item in r.get_iterator(wait=6):
    if 'text' in item:
        if count <= limit:
            if count % 10 == 0:
                print("collecting tweet {} of {}".format(count, limit))
            count += 1
            #Extract Tweet Info
            new_row = extract_tweet_info(item, count, HASHTAGS)
            df = df.append(new_row, ignore_index=True)

        else:
            print("requested tweet limit reached...")
            print("ending query for hashtag...")
            break

    elif 'message' in item and item['code'] == 88:
        print('SUSPEND, RATE LIMIT EXCEEDED: %s' % item['message'])
        break
```

```
requested tweets for hashtag is limited to 1000 tweets
collecting tweet 0 of 1000...
collecting tweet 10 of 1000...
collecting tweet 20 of 1000...
collecting tweet 30 of 1000...
collecting tweet 40 of 1000...
collecting tweet 50 of 1000...
collecting tweet 60 of 1000...
collecting tweet 70 of 1000...
```

In identifying which hashtags to filter on, we researched commonly used hashtags that each team were using as well as hashtags that casual NBA fans were using during the

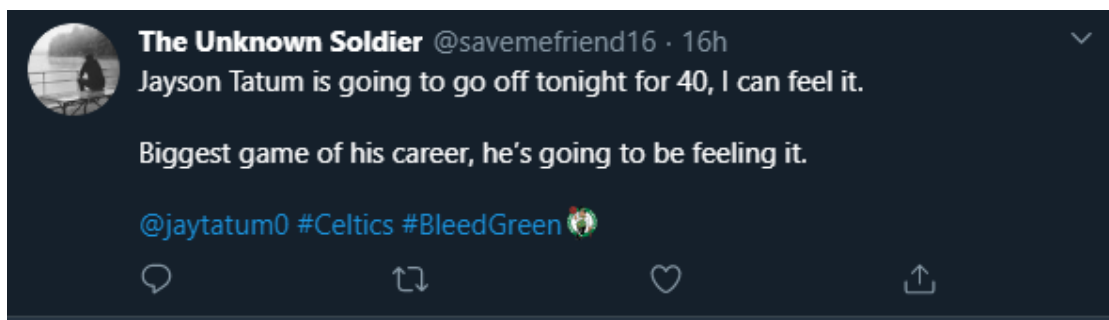


playoffs. We wanted our data to capture Miami Heat fans, Boston Celtics fans, and general NBA fans and came up with the distribution of hashtags per each segment:

Miami Heat	Boston Celtics	General (Mixed)
#HeatTwitter	#BleedGreen	#MIAsvsBOS
#HeatNation	#Celtics	#BOSvsMIA
#HeatCulture		#NBA
#Heat		#NBAPlayoffs
		#ECF

But in capturing the data, the Twitter API did have some limitations in terms of its capacity and frequency. We seemed to run into errors when we chose to extract more than 1000 tweets at a time or would run the protocol too many times in a small window. Therefore, we capped the amount of tweets extracted per scrape at 1000 and gave enough time between scrapes. We ran this many times throughout the game including pre-game, Q1 - Q4, and post-game, and ended up collecting 12,145 unique tweets and retweets.

Our data collection method was not perfect as it did have certain limitations. The Twitter API seemed to cut certain tweets if media, photos, videos, & gifs, was included and replaced the remaining characters with a hyperlink. Additionally, some hashtags did not filter on through to the tweet, although we were able to capture it in our Hashtag column, if emoticons immediately were included directly after the hashtag. An example is below.



This made it harder to classify each tweet's allegiance and is something we'd like to solve for in the future.

## Data Priming

In order to prime the data for analysis in R, we performed 4 steps in Excel.

- Step 1: Removed duplicate tweets as the data collection process had some overlaps
- Step 2: Removed tweets posted before 9/24 as those were more likely to pertain to Game 4.
- Step 3: Tagged tweets to their respective allegiance by analyzing hashtags used per each respective tweet. If a fan was to use only Miami Heat related hashtags, as specified in the table above, we classified the tweet as a Miami Heat tweet. We did the same for the Boston Celtics as well as general NBA fans. However, if any combination of Miami Heat and Boston Celtics hashtags were used within the same tweet, then we classified the tweet as general.
- Step 4: Lastly, we mapped the tweets directly to the plays of the game. We integrated a play by play data set and were able to determine which tweets occurred in which quarter and display its corresponding score.

## Variable Creation

We then looked to create more variables, most categorical but some numerical, in order to better interpret the data. The variables added were:

- Date - separated from a combined Date & Time field so we could better understand tweets by day
- Time - separated from a combined Date & Time field so we could better understand tweets by hour or time within the game
- Language - we had the abbreviations for each language a tweet occurred in but did not have the full name
  - Ex: En = English)
- Tweet or Retweet - classified each record as a tweet or retweet
- Retweet - classified if the record had been retweeted
- Category - tagged each tweet and retweet in accordance to its team preference



- Ex: Miami Heat, Boston Celtics, or General (Mixed)
- Time Frame - at what point did each tweet occur
  - Ex: Pre-Game, Q1, Q2, Halftime, Q3, Q4, & Post-Game
- Score Differential - identified the point differential for each respective team at time of tweet or retweet
  - Ex: 7:40 PM PST => MIA = -9, BOS = +9
- Timeout vs Instant Replay vs Non-Timeout - identified tweets and retweets that occurred during a timeout or during a replay
- Referee Referenced Tweet - identified tweets and retweets that specifically mentioned Referees

## Description of Data

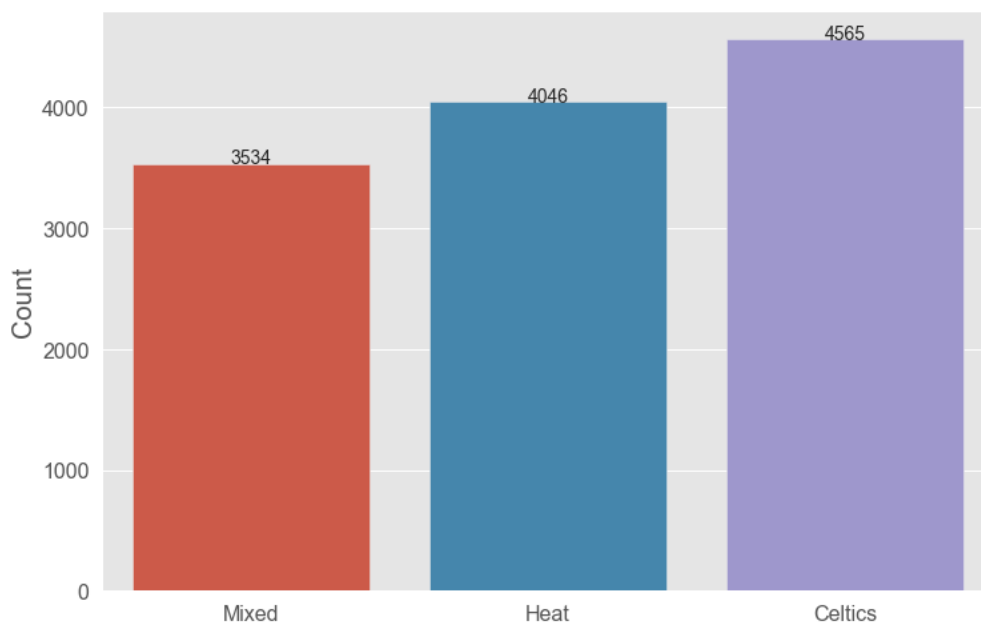
---

### Overall Description

After we cleansed our data, we ended up with 12,145 records and 22 features. Before we proceeded to the next step, we checked the data quality and looked into each column to understand the distribution of data and perform preliminary analysis.

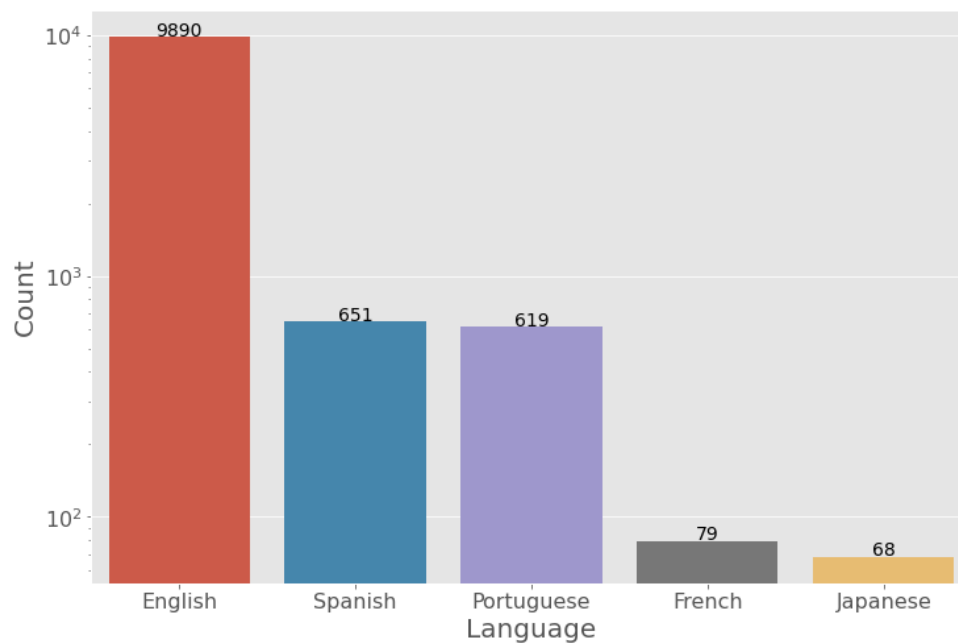
### Description of Variables

- **Category:** This column has three values: “Mixed”, “Heat” and “Celtics”. If the content of a tweet is related to only Boston Celtics, then we categorized it as “Celtics”; if only related to Miami Heat, we categorized it as “Heat”; if the tweet is related to both teams, we then categorized it as “Mixed”. The count of tweets was highest for the Boston Celtics.

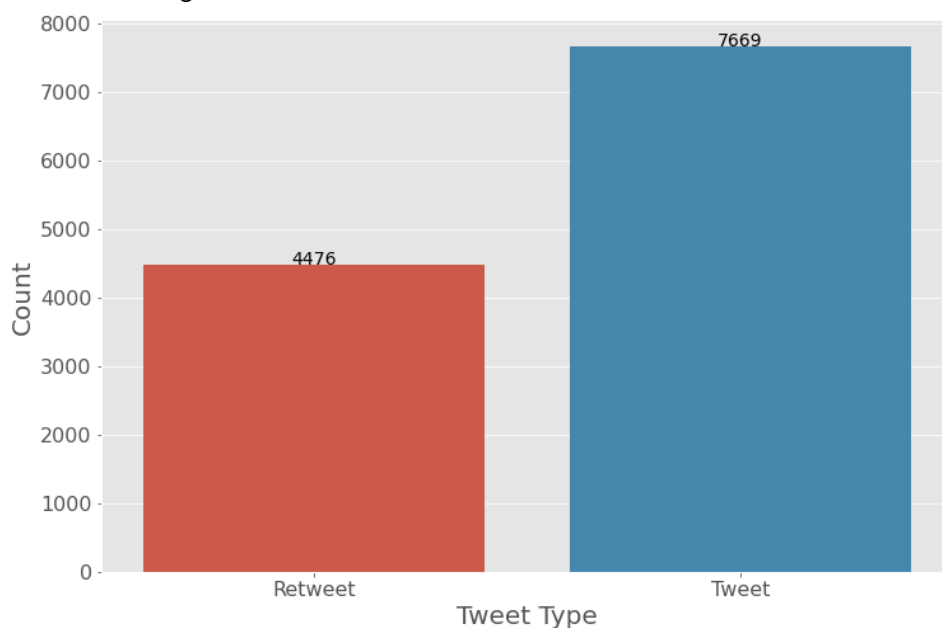


- **Full Date & Time:** This column specifies when did the user post this tweet. This feature is automatically converted to Greenwich Mean Time (GMT) rather than local time.
- **Tweet:** This column is the content of the tweet.
- **Hashtags:** This column indicates the hashtags that are included in the tweet.
- **Language code:** This column indicates the language of the tweet in abbreviation form.
- **Language:** This column spells out the language of the tweet. There are 28 different languages amongst all records. There are almost 10k tweets in English, more than 600 tweets in Spanish and Portuguese, and more than 50 tweets in French and

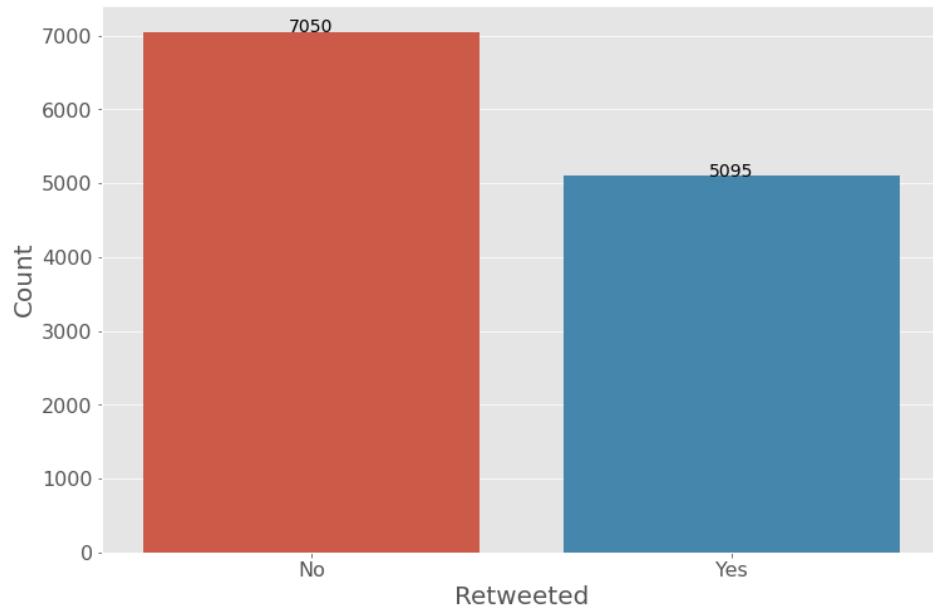
Japanese (note that the bar chart is in log scale). Other languages have no more than 50 tweets and 678 tweets were labelled as undefined.



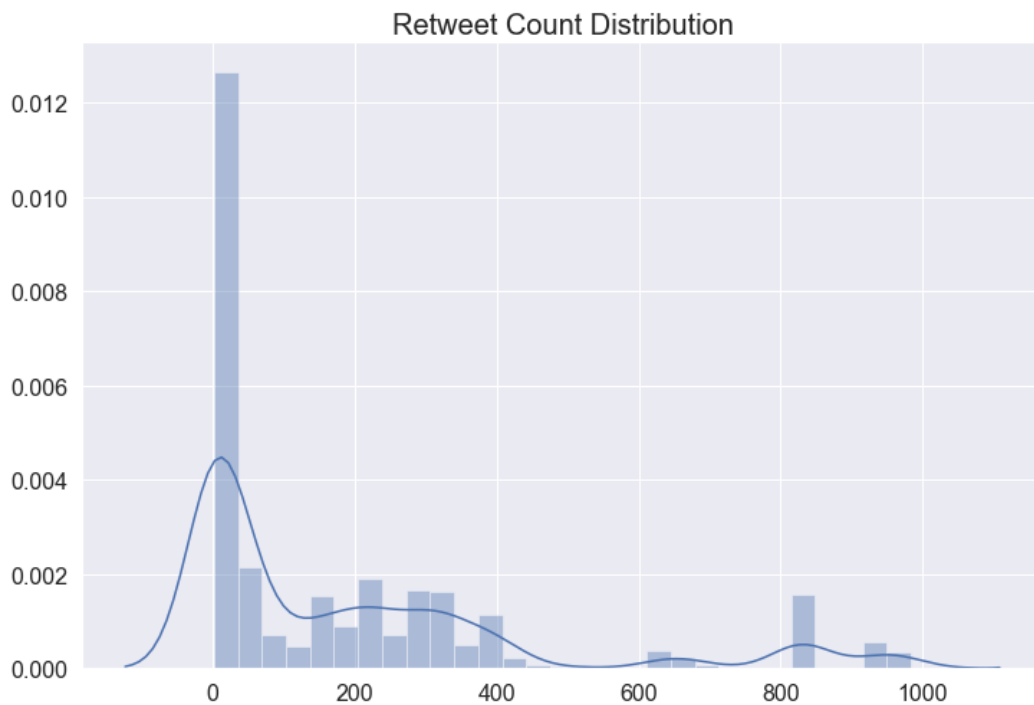
- Tweet/Retweet: This column indicates the tweet type, which means whether the record was an original tweet or a retweet. 37% of the tweets are retweets and 63% of the tweets are original tweets.



- Retweeted: This column indicates whether a tweet has been retweeted by any other user. 41.9% of all the tweets have been retweeted.

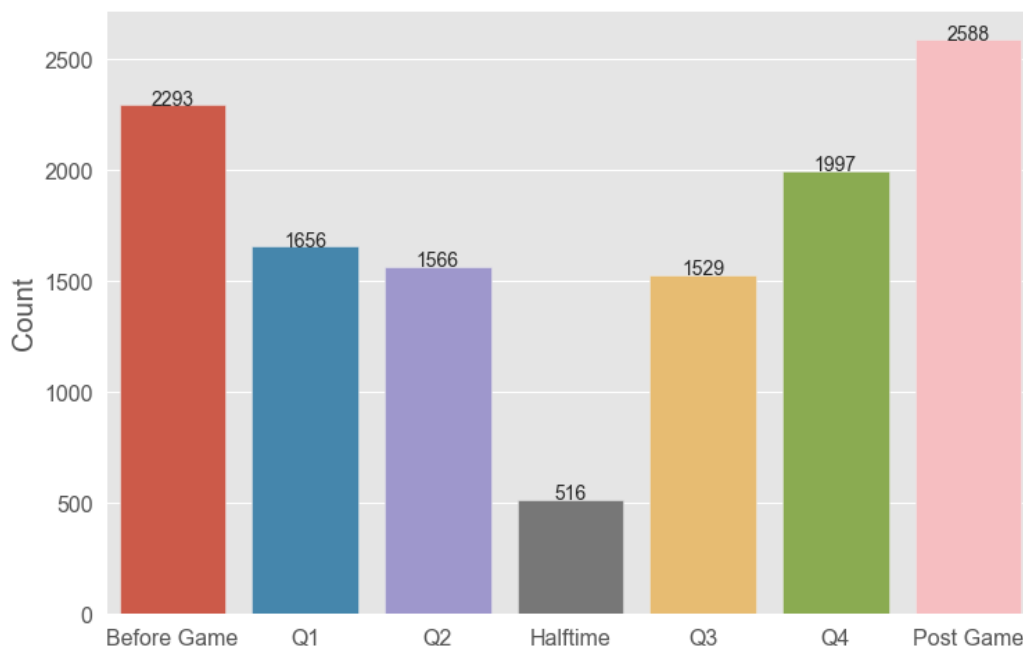


- Retweet count: This column shows how many times a tweet has been retweeted. The histogram below shows the distribution of retweet counts for all tweets that have been retweeted at least once and no more than 1,000 times.



- User follower count: This column indicates how many followers does a user have.
- User friend count: This column indicates how many friends does a user have.
- User location: This column shows the user location. Around 70% users shared their location.

- Time Frame: This column indicates what time frame does a tweet fall in. We have 7 different time frames: before game, during Q1/Q2/Q3/Q4, halftime and post game. There are more tweets before game and after game, mainly because the time window is longer; there are more tweets in quarter 4 than other three quarters because of the comeback win of the Celtics over the Heat. And surprisingly, not many people tweeted during halftime.



- MIA Score Differential: This variable indicates how many points the Miami Heat lead over Boston Celtics at a specific time.
- BOS Score Differential: Similar to the previous variable, this indicates how many points the Boston Celtics lead over Miami Heat at a specific time. It is simply the inverse of what the score differential would be for the Heat.
- Timeout/Instant Replay: This is another kind of time frame feature. This variable indicates whether a tweet was posted during a timeout or during an instant replay.
- Referee Related: This variable identifies tweets and retweets that specifically mentioned the referee. Only 2.9% of tweets mentioned the referee.

## Analysis

---

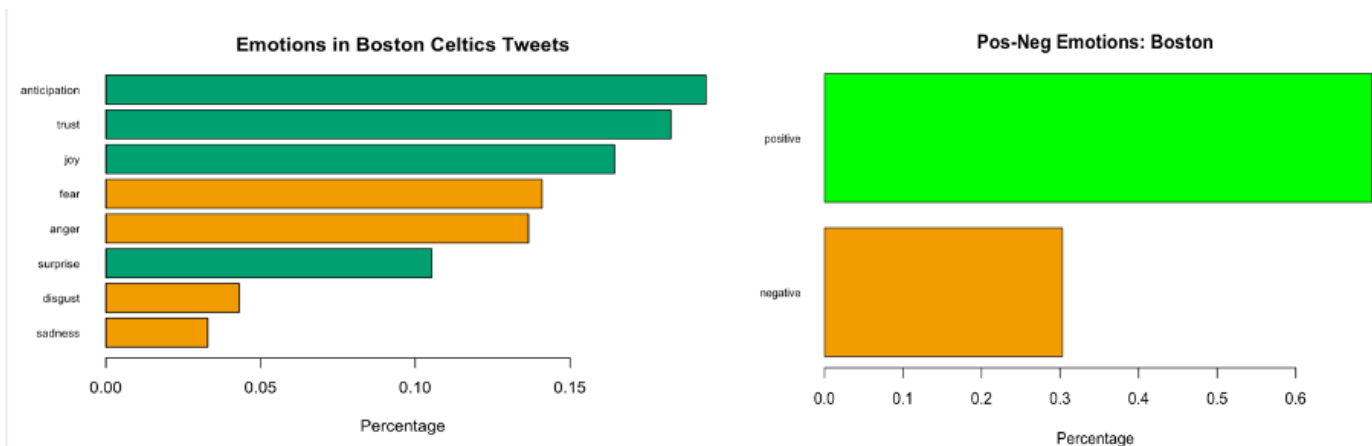
### Sentiment Analysis

With the data scraped from Twitter, our group looked to conduct sentiment analysis to better understand the emotions of fans as the course of the game occurred. By identifying trends, we could look to provide suggestions that could improve fan engagement.

For this analysis, we used a package called “sentimentr”, which is designed to take into account valence shifters like amplifiers and negators, which results in a more accurate sentiment score. This library not only categorizes each tweet as positive or negative, but also generates a numerical value for each tweet. The higher the sentiment value, the stronger the positive emotion in the post.

We separated the data by Category (Celtics, Heat and Mixed), and conducted sentiment analysis to discover the most common emotions in Tweets for each category. Below are our results:.

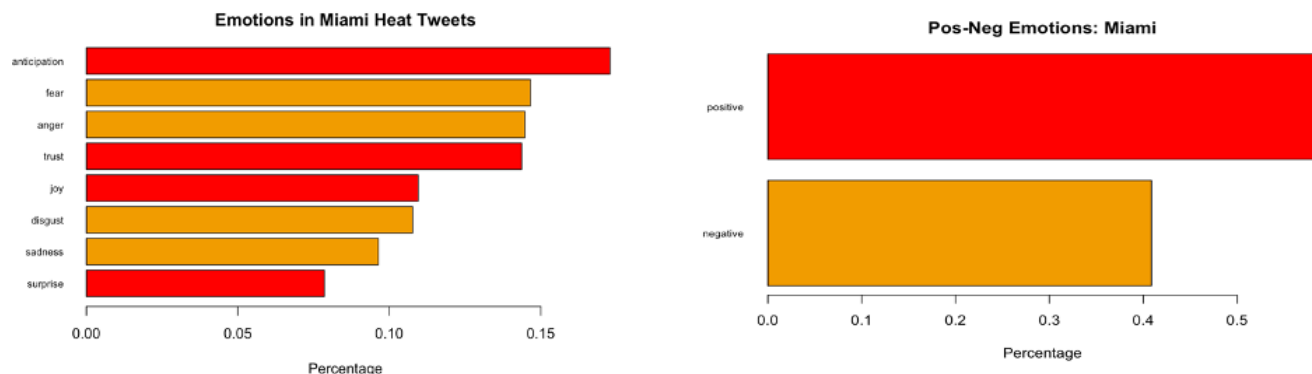
- Boston Celtics



From the bar plot above, we can see that the three most common emotions were “Anticipation”, “Trust” and “Joy”, all of which happened to be positive. Additionally, we can see that Boston positive tweets more than doubled their negative tweets.

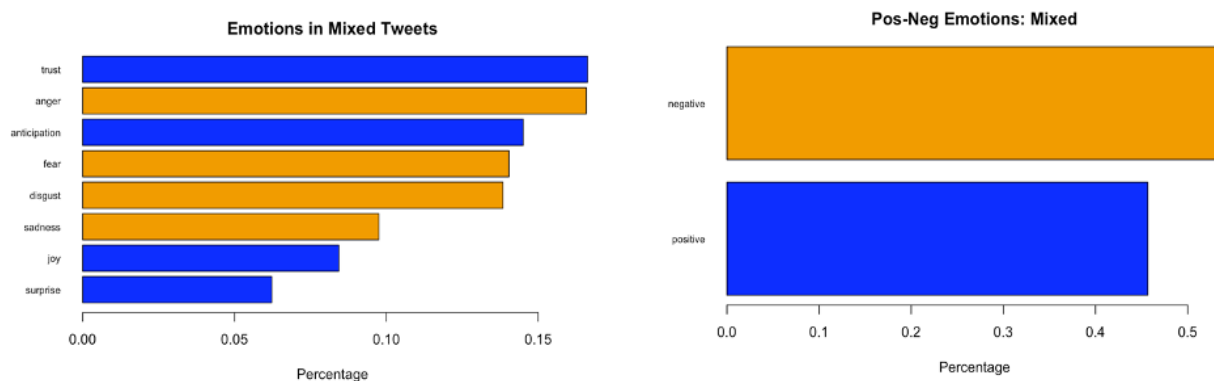
- Miami Heat

## Fan Engagement Analysis



For the Heat, the three most common emotions shown were “Anticipation”, “Fear” and “Anger” which makes sense given that they were one win away from the NBA Finals but blew a significant lead and lost the game. Although we see that Heat fans showed a more overall positive than negative attitude in their tweets, their percentage of negative tweets were significantly more than their Boston counterparts.

- Mixed

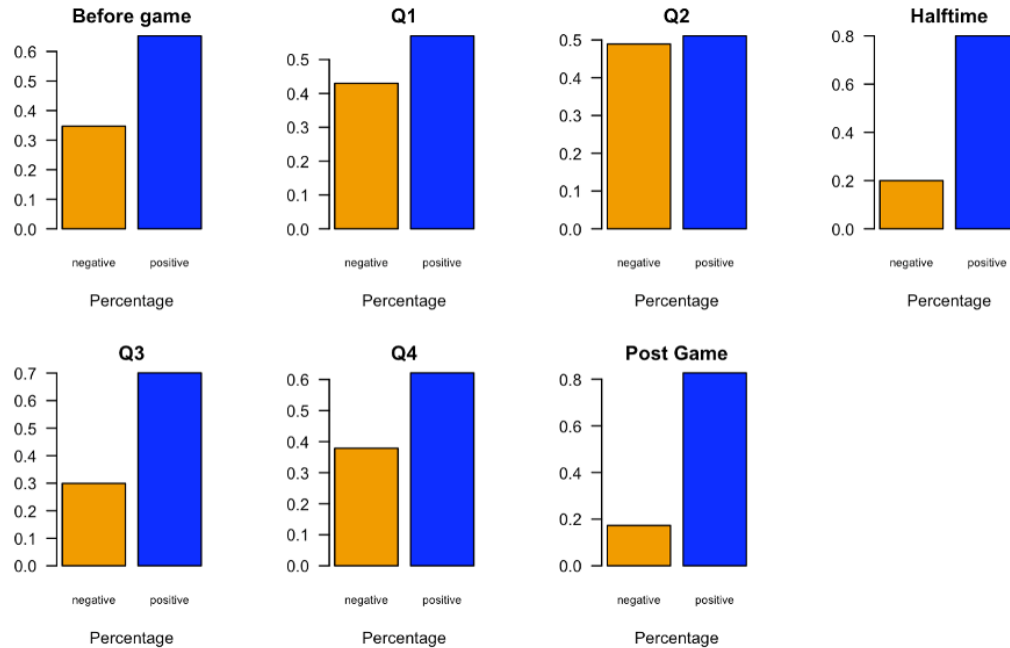


Lastly, for the category of mixed fans, we saw that they most exuded “Trust”, “Anger”, & “Anticipation”. However, it was surprising to see that negative emotions outweighed the positive emotions for a seemingly neutral crowd.

Our next set of analysis was to better understand how fans’ emotions may change over the course of a high stakes playoff game. So for both the Boston Celtics and Miami Heat, we analyzed the change in emotions through different time frames which includes: Before Game, Quarter 1, Quarter 2, Halftime, Quarter 3, Quarter 4, and Post Game.

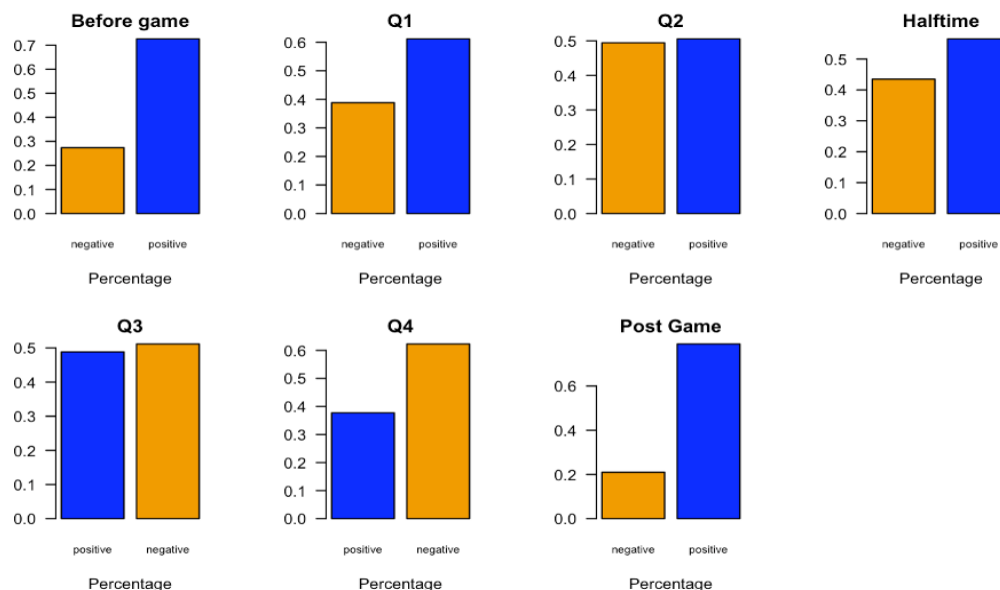
- Boston Celtics





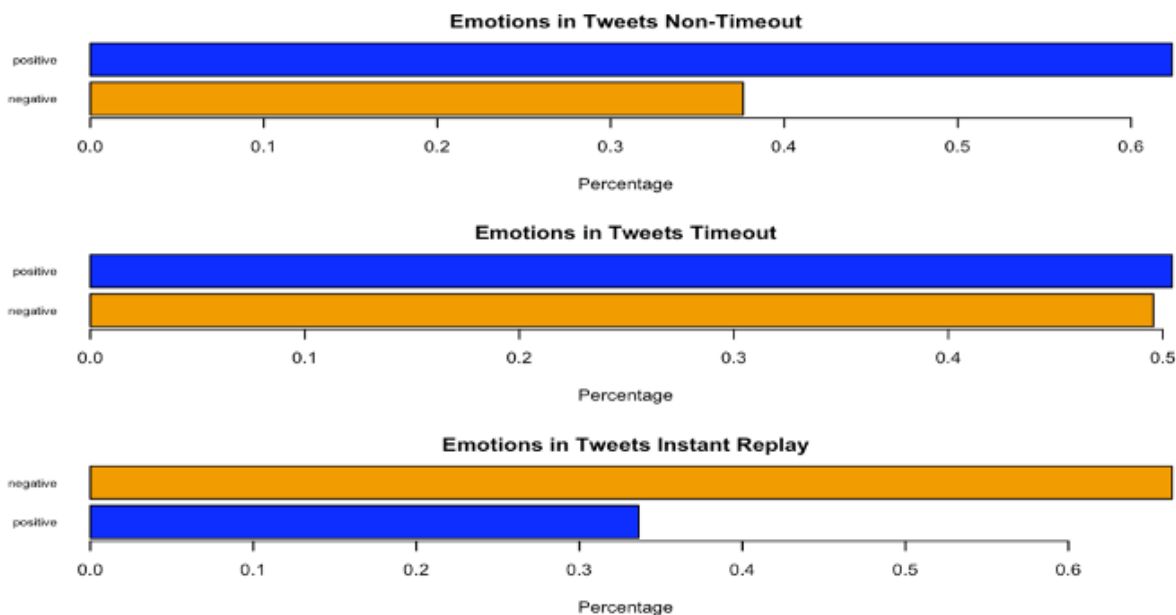
For the Boston Celtics, we found that fans never let the negative emotions outweigh the positive emotions and showed resilience within each time frame. Quarter 2 was the peak of frustration for Celtics fans. But sentiment shifted greatly during halftime and after the team secured the lead in the 3rd Quarter, Celtic fans remained happy throughout the rest of the game. It's also interesting to note how far the negative reactions fell from Q4 to Post Game, perhaps indicating that Celtic fans felt greater optimism and could overcome a 3-2 deficit in order to proceed to the NBA Finals.

- Miami Heat

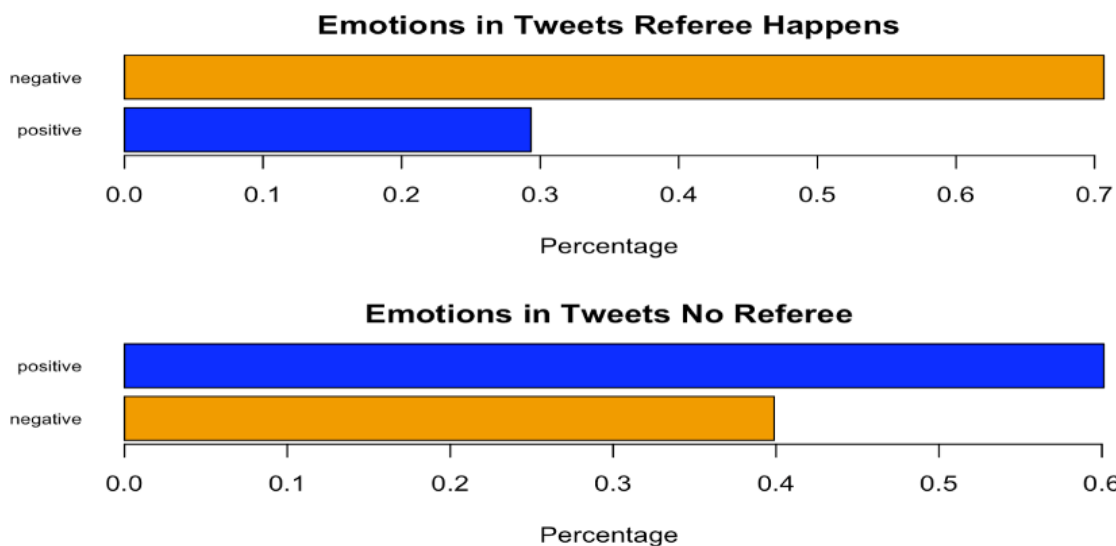


For the Heat, a drastic shift of sentiment occurred during Quarter 2 where the Celtics put themselves back in the game. The optimism increased during halftime, however, as the game progressed in Quarters 3 and 4, fans became increasingly negative. Interestingly enough, post game sentiment levels reverted back to pre-game levels indicating that Heat fans still felt that they had the best chance in advancing to the finals.

Our third and final sentiment analysis looked at other elements besides sentiment versus time frame. We were curious to see how fan sentiment was affected by timeouts and instant replays. Additionally, we wanted to see if there were significant sentiment shifts when a tweet mentioned Referees.



In comparing sentiment scores of game play vs timeouts, it's no surprise that fans posted or retweeted more negative tweets. Timeouts could serve as a general venting timeframe where fans unload their thoughts and frustrations without having to miss any basketball action. However, during Instant Replays, fans became overwhelmingly negative. This is perhaps due to the nature of the replay where it tremendously slows down the pace of the game. However, this feature is a newer one and has shown value in moments of crucial game play.



In this analysis, we looked at tweets and retweets that specifically mentioned Referees. We initially thought there would be somewhat of a negative view towards Referees, however we did not expect the magnitude of negativity when referees were mentioned. Although the number of Referee related tweets were low in relation

to the total number of tweets, when fans did mention the referees, it did not bode well for sentiment.

### Statistical Modeling for Retweet Analysis

As social media has become more and more important for businesses, Twitter in particular can help the business develop a stronger following because it provides a closer connection with a broader audience. One of the most crucial aspects of Twitter is the retweet. The power of the retweet is extremely influential as when a business gets retweeted, it can indicate that:

- The retweet shows somebody appreciated its content;
- The retweet spreads the content and increases the probability of it going viral;
- The retweet signals the brand to the retweeter's followers. Therefore, the business is not just promoting the work to its followers, but promoting to their followers.
- The retweet increases the amount of influence by the business.

Because retweets allow content to spread quicker and to more people, we would like to further analyze what are the factors that would affect if a tweet to be retweeted not. As a result, we conducted linear regression and logistic regression models in order to have a deeper understanding from a statistical perspective.

#### Linear Regression Models

Linear regression is a common statistical data analysis technique. It is used to determine the extent to which there is a linear relationship between a dependent variable ( $x$ ) and one or more independent variables ( $y$ ). There are two types of linear regression, simple linear regression and multiple linear regression. In simple linear regression, a single independent variable is used to predict the value of a dependent variable. In multiple linear regression, two or more independent variables are used to predict the value of a dependent variable.

We built multiple regression models to see what variables are statistically significant to retweets. The independent variables are numerical variables, including the counts of hashtags, binary values if the user is verified or not, followers counts of the user, friends counts of the user, sentiment scores indicating if the tweet includes positive/negative emotion, and the difference in team scores. The dependent variable is the counts of

retweets. We also built separate models to see how each team (Boston Celtics, Miami Heat, or Mixed) performed.

Dependent variable (y)	retweet_count	numerical
Independent variable (x)	hashtags_count	numerical
	verified_bi	binary
	user_follower_count	numerical
	user_friend_count	numerical
	negative	numerical
	positive	numerical
	abs(score_differential)	numerical

For whole dataset:

R-squared: 9.35%

Variables	Coefficient	P value
(Intercept)	2.161e+02	<2e-16
hashtags_count	-3.903e+01	<2e-16
verified_bi	-1.880e+01	0.452
user_follower_count	-1.541e-03	0.135
user_friend_count	3.730e-06	0.641
negative	-6.574e+01	<2e-16
positive	9.876e+01	<2e-16
abs(score_differential)	-7.389e+00	<2e-16

For Boston Celtics:

R-squared: 27.05%

Variables	Coefficient	P value
(Intercept)	3.585e+01	<2e-16
hashtags_count	-5.084e+01	<2e-16
verified_bi	-8.012e+01	0.02989
user_follower_count	-6.177e-03	0.00111
user_friend_count	1.126e-05	0.67543
negative	-1.699e+02	<2e-16
positive	1.367e+02	<2e-16
abs(score_differential)	-1.649e+01	<2e-16

For Miami Heat:

R-squared: 0.34%

Variables	Coefficient	P value
(Intercept)	7.372e+01	1.16e-13
hashtags_count	-1.595e+01	6.55e-05
verified_bi	2.258e+01	0.4960
user_follower_count	7.611e-05	0.9484
user_friend_count	3.329e-06	0.68347
negative	-8.654e+00	0.2798
positive	1.533e+01	0.0406
abs(score_differential)	2.049e+00	0.0113

For Mixed:

R-squared: 36.94%

Variables	Coefficient	P value
(Intercept)	2.6417024	0.0106
hashtags_count	-0.138185	0.5386
verified_bi	NA	NA
user_follower_count	0.0009134	0.1835
user_friend_count	0.0003088	0.2337
negative	-1.1908983	0.2470
positive	-0.9558371	0.1434
abs(score_differential)	NA	NA

In general, there are 4 variables that are statistically significant to the count of retweets: the count of hashtags, negative tweets, positive tweets, and score differential. Based on the coefficients for hashtags, we can conclude that using less hashtags may not necessarily lead to a higher retweet count. Additionally, posting more positive tweets is more likely to help increase the amount of retweets as we can see that be the case for both the Boston Celtics and Miami Heat. Lastly, from additional analysis we were able to determine that when the scores of the 2 teams are close, fans are more likely to retweet to thus increasing the count of retweets. However, it is surprising that the counts of both user followers and user friends are not statistically significant to the count of retweets.

### Logistic Regression Models

Like linear regression, logistic regression is a predictive analysis. Logistic regression estimates the parameters of a logistic model. A binary logistic model has a dependent value with two possible values, where the two values are labeled “0” and “1”. It is used to describe data and to explain the relationship between one dependent binary variable and one or more variables.

We built logistic models to see what variables are statistically significant to retweets. The independent variables are numerical variables, including the counts of hashtags, binary



values if the user is verified or not, followers counts of the user, friends counts of the user, sentiment scores indicating if the tweet includes positive/negative emotion, and the difference in team scores. The dependent variable is the binary value of retweets (that is, 1 to be retweeted and 0 to be not retweeted). We also built separate models to see how each team (Boston Celtics, Miami Heat, or Mixed) performs.

Dependent variable (y)	retweet_bi	binary
Independent variable (x)	hashtags_count	numerical
	verified_bi	binary
	user_follower_count	numerical
	user_friend_count	numerical
	negative	numerical
	positive	numerical
	abs(score_differential)	numerical

For whole dataset:

Variables	Coefficient	P value
(Intercept)	5.283e-02	0.35001
hashtags_count	-6.784e-02	0.00234
verified_bi	2.261e-01	0.30795
user_follower_count	5.996e-06	0.58316
user_friend_count	2.325e-06	0.12846
negative	-3.598e-01	2.44e-12
positive	7.335e-01	<2e-16
abs(score_differential)	-2.701e-02	1.34e-10

For Boston Celtics:

Variables	Coefficient	P value
(Intercept)	4.904e-01	1.04e-08
hashtags_count	-7.497e-02	0.01834
verified_bi	-1.275e-01	0.74611
user_follower_count	5.626e-05	0.00671
user_friend_count	5.378e-06	0.04253
negative	-5.338e-01	5.99e-11
positive	8.706e-01	<2e-16
abs(score_differential)	-6.565e-02	<2e-16

For Miami Heat:

Variables	Coefficient	P value
(Intercept)	-4.259e-01	1.26e-07
hashtags_count	-3.703e-02	0.2547
verified_bi	3.925e-01	0.1446
user_follower_count	-1.696e-05	0.2841
user_friend_count	1.183e-06	0.3417
negative	-3.158e-01	3.82e-06
positive	4.656e-01	1.44e-14
abs(score_differential)	1.671e-02	0.0101

For Mixed:

Variables	Coefficient	P value
(Intercept)	-2.550e-01	0.783
hashtags_count	1.061e-02	0.960
verified_bi	NA	NA
user_follower_count	1.186e-03	0.197
user_friend_count	8.758e-05	0.813
negative	-1.845e+01	0.997
positive	6.069e-01	0.372
abs(score_differential)	NA	NA

Similar to the linear regression result, in general, there were 4 variables that were statistically significant to a tweet being retweeted: the counts of hashtags, negative tweets, positive tweets, and score differential. We can make the same inferences when it comes to hashtag usage, the power of the positive tweet, and how close games causes users to retweet more.

## Conclusions

---

### Results

In regards to our sentiment analysis, we found:

- Celtics Fans:
  - Displayed significantly more positive emotions than negative
  - Showed resilience as positive emotions never outweighed negative emotions, even when the team was falling behind
- Heats Fans:
  - Overall emotions were generally positive but negative emotions were greater than Celtics fans
  - Negative emotions were predominant during the second half of the game as the Celtics overtook the Heat
  - Post-game optimism levels reverted back to pre-game levels, even after a loss
- Mixed Category:
  - For a seemingly neutral crowd, it was surprising to see that negative emotions outweighed the positive
- Sentiment scores dropped during Timeouts and dropped even further during Instant Replays
- Tweets and retweets that mentioned Referees elicited more negative emotions

In regards to our retweet analysis we found:

- For both multiple linear regression models and logistic regression models, there are four variables that are significant:
  - Count of hashtags
  - Negative tweets
  - Positive tweets
  - Score differential
- Less hashtag usage may not lead to higher retweets
- Negative tweets will cause fewer amounts of retweets while positive tweets will cause more

### Recommendations

There are several recommendations we can make per organization based on our analysis:

- Celtics
  - Posting more positive tweets that indicate optimism and hope are more likely to be retweeted and have higher retweet counts
  - Post more tweets when score difference is low as tweets are more likely to be retweeted

- Continue to use 1 hashtag, #BleedGreen, but include hashtag on all tweets as majority of Celtics tweets during game did not include any hashtags
- Heat
  - Posting more positive tweets that indicate optimism and hope are more likely to be retweeted and have higher retweet counts
  - Post more tweets when score difference is low as tweets are more likely to be retweeted
  - Consolidate number of hashtags used from 4 to 1 and focus on most popular hashtag, #HeatTwitter
- NBA
  - Timeouts
    - Reduce length of timeout until it adversely affects the risk of the player's health
    - Replace traditional ads with exclusive content and show more in game ads throughout the game to make up for lost revenue
      - Exclusive content can include player interviews, highlight reels, expert analysis, mic'd up players
      - In game ads can occur more frequently during fouls and free throws
    - Switch to an ad free mode that in exchange for no ads, the user will pay a small fee
      - In lieu of the ads, fans can have the ability to explore different cameras and angles
  - Challenges/Instant Replays
    - Sentiment shift was so severe that the NBA should look to
      - Cap the length of how long a challenge can be analyzed
      - Couple timeouts with challenges
      - Only have challenges occur during final minutes of game or remove challenges all together
  - Referees
    - Provide more training so less incorrect calls are made
    - Understand which calls cause the most negative sentiment and adjust appropriately in order to enhance flow of game

## Conclusion

By analyzing just 1 NBA game, we were able to derive a lot of insight into the behavior of the online fan. We can identify certain trends and recommend actionable items that can

allow teams to maximize their fan engagement and the league to make the proper adjustments in order to enhance the flow of the game.