ISCTE IUL
Instituto Universitário de Lisboa

Does Mixed Martial Arts (MMA) fighters' trash-talk go viral?

Duarte Silva Tereso

A Dissertation presented in partial fulfillment of the Requirements for the Degree of Master's in Computer Science and Management

Co-Advisor:

Dr. Sérgio Moro, Assistant Professor at ISCTE-IUL, Instituto Universitário de Lisboa, Department of Information Science and Technology (ISTA)

Co-Advisor:

Dr. Pedro Ramos, Associate Professor at ISCTE-IUL, Instituto Universitário de Lisboa, Department of Information Science and Technology (ISTA)

September 2019

Abstract

Since the beginning of the 2000s, Mixed Martial Arts (MMA) and, more specifically, the Ultimate Fighting Championship organization (UFC), have been experiencing exponential growth. In recent years, there have been some differences in what fans want to see. Having fights with a lot of drama between the fighters generates large numbers of pay-per-view compared to fights between highly technical and dominant fighters, even if they are title fights. The purpose of our study is to understand if trash-talk between the fighters before the fight tends to attract more fans to the sport and what kind of sentiments does it generate in them. We extracted tweets from fans and answers from fighters and cross these data. We found that fights involving a lot of drama between the fighters are the ones that have the biggest pay per view numbers, 2 of them generating 2.5 times more pay per view than 7 fights between lower profile fighters. However, we did not find a direct influence of each fighter's negative expressions used during the conference and the fans' tweets but, we did find that the context of the event in itself plays a more significant role as the rivalry between fighters is expressed during the press conference. Regarding the engagement of fans', while we found no evidence of a cause-effect relationship with the negative sentiment of the fighters, we confirmed that higher levels of profanity speech are associated with higher engagement, both on pay-per-view and on the number of tweets.

Keywords: Mixed martial arts (MMA); Ultimate Fighting Championship (UFC); Fandom; Trash-Talk; sentiment analysis; fighting events.

Resumo

Desde o início dos anos 2000, as Artes Marciais Mistas (MMA) e, mais especificamente, a organização Ultimate Fighting Championship (UFC), têm tido um crescimento exponencial. Nos últimos anos, parece haver algumas diferenças no que os fãs querem ver, tendo combates que envolvem muito drama entre os lutadores gerado números de *pay-per-view* muito maiores do que combates entre lutadores altamente técnicos e dominantes, mesmo sendo estes combates pelo título. O objetivo do nosso estudo é entender se o *trash-talk* entre os lutadores antes da luta tende a atrair mais fãs para o desporto e que tipo de sentimentos provoca nestes fãs. Extraímos tweets dos fãs e respostas dos lutadores e cruzámos a informação. Os resultados mostram que os combates que envolvem muito drama entre os lutadores são os que têm os maiores números de *pay-per-view,* gerando 2 deles 2,5 vezes mais pay per view do que 7 combates entre lutadores com um perfil mais discreto. No entanto, não encontrámos uma influência direta das expressões negativas dos lutadores nas conferência de imprensa nos tweets dos seus fãs, mas descobrimos que o contexto do evento em si desempenha um papel mais significativo, pois a rivalidade entre os lutadores é expressa durante a conferência de imprensa do evento. Em relação ao envolvimento dos fãs, embora não tenhamos encontrado evidências de uma relação de causa-efeito com o sentimento negativo dos lutadores, verificámos que níveis mais altos de palavrões estão associados a um maior envolvimento, tanto no *pay-per-view* como no número de tweets.

Palavras-chave: Artes marciais mistas (MMA); Ultimate Fighting Championship (UFC); *Fandom*; *Trash-Talk*; análise de sentimentos; desportos de combate.

Acknowledgements

I would like to thank, first and foremost, my advisor Professor Sérgio Moro for guiding me through this master thesis. I am forever thankful for his constant availability, support and continuous feedback, who helped me a lot on this walk.

I also would like to thank my Co-Advisor Professor Pedro Ramos that, whenever I needed, always was available to give me his opinion and Professor Teresa Calapez for joining us and for providing essential help in the analysis and conclusions of the data.

I also want to thank my family, my martial arts family and my friends, who always supported me. Specially, to my cousin Rafael Nunes and my friend David Guapo for all the help they gave me to review my dissertation and constructive criticism that helped me a lot to improve the final result.

Index

Index of figures

Index of tables

List of abbreviations

UFC	Ultimate Fighting Championship
MMA	Mixed Martial Arts
SA	Sentiment analysis
PPV	Pay-per-view

1. Introduction

Mixed Martial Arts is a sport with a relatively recent history and, since its inception, it has experienced an exponential growth. The number of practitioners, either at an amateur level, or at a professional level, and the number of fans who support fighters and attend events, either live or through pay-per-view, have been increasing, considerably, over the last few years. This exponential growth has aroused the interest of several researchers in trying to understand what motivates fans to consume, increasingly, a sport like MMA and what factors influence this consumption the most.

Studies show that there are different motivations for fans to consume MMA, such as, for example, escape and entertainment (Wann, 1995), sport interest, drama and aesthetics (Seungmo et al., 2008) or interest in specific fighters and specific weight classes (Tainsky et al., 2013). Although many studies suggest different motivations for fans to consume MMA, very few analyze the weight of the fighter's profile has in the motivations of the fans. The most notorious fighters have legions of fans which they influence with both their positive and negative character qualities, as well as their behavior (Brown et al., 2013). Although many adopt positive values and behaviors, there are also many whose destructive behaviors are adopted by their fans.

Does the way how fighters express themselves influences their fans' behavior? This is the main question that guided the present study. Specifically, we aim to understand how the negative behavior of the fighters influence the negative behavior of his fans. For this, we collected a sample of information from two dimensions: Fighters and Fans. To validate this information, we collected data from social media Twitter related to MMA, such as tweets from fans, and from events' press conferences, such as the answers that fighter's gave. Secondly, we did a sentiment analysis in the collected data and crossed this information to understand the patterns of behavior from fans who are influenced by the fighters they support. We focused on the negative sentiment and profanity expressed from both Fighters and Fans on their comments (tweets or answers) and crossed the information to understand if a cause-effect relation exists or not. Thus, by comparing the sentiments and profanity expressions generated between both, we contribute to existing literature by unveiling the propagation effect of trash-talk in fighting sports, which reflects aggressiveness among individuals within a society (Workman, 2012).

The next chapter describes in more detail the theoretical background which supports this study, with a detailed description of the chosen methodology and approach to extract data and perform a sentiment analysis. Then the chosen path is explained, describing the methods and criteria that we used to treat the data. Next, the results are discussed and interpreted in order to extract knowledge out of the data. After this, the conclusions are drawn together with a description of the future work and, finally, the limitations that we found during the study are explained.

2. Literature review

2.1 The history of MMA

In November 1993 MMA was introduced to the world, when the first Ultimate Fighting Championship (UFC) event debuted. However, the origin of MMA dates to 649 B.C, when the sport of Pankration, a mix of boxing and wrestling, was introduced into the Olympic Games and to about 80 years ago when Vale Tudo, a Brazilian form of MMA, began to draw some attention. In present day, UFC athletes are skilled in many forms of martial arts, including boxing, wrestling, jiu-jitsu, kickboxing, muay thai and other fighting sports and compete at an elite level in a regulated environment where safety is paramount. Athletes win by either submitting the opponent, forcing him to physical or verbal tap out, by knocking him out (athlete is knocked unconscious due to strikes or impact, KO), by interruption of the referee (TKO), or by judges' decision (Seungmo et al., 2008; UFC, 2018). Although in the beginning MMA was seen as a violent sport and for some years was banned from most US states and PPV channels (Schumacher-Dimech et al., 2012), over the years, through a process of mainstreaming MMA, it becomes possible for almost anyone to participate in and enjoy this sport (Andreasson & Johansson, 2019).

Since the beginning of the 2000s, MMA and the UFC have been experiencing exponential growth, generating interest from fans and participants around the world. UFC television events began to have higher cable ratings than events of big sports federations, such as National Basketball Association (NBA), National Hockey League (NHL), and Major League Baseball (MLB), and pay-per-view revenues comparable to major boxing and wrestling (Seungmo et al., 2008). Today, the UFC is progressing at a rate never seen in the professional sports world and holds the distinction of the largest live Pay-Per-View event provider in the world. UFC produces more than 40 live events annually, is broadcast in over 129 countries and territories, to nearly 800 million TV households worldwide, in 28 different languages (UFC, 2018).

In 2001, the Fertitta brothers bought the Ultimate Fighting, now known as UFC, for $2 million and, in 2008, Forbes valued the organization at $1 billion (Miller, 2008; Bearak, 2011). This shows the exponential growth of the sport and the UFC around the world. Co-owner Lorenzo Fertitta asserted that the company is worth "more than Manchester United, more than the New York Yankees, more than the Dallas Cowboys.", what, if it becomes true, puts the UFC in the $2 billion mark (Bearak, 2011).

2.2 Fans motivation for MMA consumption

Given the exponential growth of MMA and its number of fans, it became important to understand why people are attending these events and what is driving them to consume the sport through media and merchandise. Different sports and different consumer segments call for different motives to be appreciated. The motives of sport consumers can be different depending on the type of sport, whether they are artistic sports (e.g., gymnastics and synchronized swimming) or combative sports (e.g., wrestling, mixed martial arts and boxing); therefore, individual motives should be rationalized for each sport. Researchers have studied and identified different key motivation factors and have developed scales to measure the motives of sport consumers (Funk et al., 2002; Seungmo et al., 2008; Wann et al., 2008).

In the case of fighting sports, previous research has identified several factors, either related with personal purpose, such as eustress and self-esteem (Wann, 1995; Wann et al., 2008), empathy (James and Ross, 2004), achievement (Funk et al., 2002; Seungmo et al., 2008), entertainment (Funk et al., 2002; James and Ross, 2004; Wann et al., 2008) or drama (James and Ross, 2004; Seungmo et al., 2008; Andrew et al., 2009). Related with the sport, such as interest in sport and teams (Funk et al., 2002; James and Ross, 2004; Seungmo et al., 2008; Wann et al., 2008). Or related with the fighters, such as, aesthetic (Wann, 1995; Wann et al., 2008; Andrew et al., 2009), role modeling (Funk et al., 2002), skill (James and Ross, 2004) or interest in specific fighters and in specific weight classes (Tainsky et al., 2013). The last two could be explained because people develop psychological relationships with celebrities and see them as role models, so they tend to adopt the perceived values and behaviors they see in them (Fraser & Brown, 2002).

Thus, physical violence by itself has not been identified as the main reason for following and attending MMA events. As a result, managers invest in advertising and organize press conferences where fighters aim to draw the attention of their fans and the wider MMA community (Andreasson & Johansson, 2019).

2.3 Fighters and trash-talk

As we saw in section 2.1, MMA is a full-contact fighting sport, allowing striking and grappling techniques, both standing and on the ground, derived originally from various styles of martial arts. Therefore, athletes must be very well prepared physically, technically and mentally. Many researchers have studied the MMA fighters' profile from different points of view.

For example, Massey et al. (2013), concluded that creating and maintaining a routine that deliberately induces pain and distress and increases the levels of stress and fatigue are essential factors to training and better performance for MMA athletes and are seen as normal elements of training. Fear was also one of the main concerns noted in MMA athletes, both internally as suppressing a fear of failure and externally as creating fear in opponents. To manage that fear, fighters use different mechanisms such as, for example, approaching the fight as a normal day in the gym or looking at the opponents as inferior. With that vision they can manage their emotions, suppress fear and evoke confidence, which demonstrates motivation and mental toughness, two other factors noted in MMA fighters (Vaccaro et al., 2011; Chen & Cheesman, 2013). This mental toughness plays a central role in the spectacle of MMA because the violence has a considerable impact on the athletes. Although this violence is clear and present, it is often seen as something good and productive, as a part of the show, leading athletes to develop different strategies to handle the physical force and violence in the cage. Thus, instead of focusing on violence, fighters aim to draw attention to the formation of the MMA and positive aspects of the sport (Andreasson & Johansson, 2019).

The next two studies present interesting points for our study. Robbins and Zemanek (2017) analyzed the influence of high-profile celebrity fighters in the PPV numbers. They compared the value and impact generated for the UFC by these fighters with the highest ranked pound-for-pound fighter and find that celebrity has far more economic value than fighting skill. Vaccaro et al. (2011) shown that in order to manage emotions and cause fear in their opponents, fighters enact intimidating personas by using language and their bodies strategically. Nowadays, competition involves more than just the physical contests between athletes, and fighters enact these personas, before and during the competition, aiming to affect their opponent's performance. This is called Trash-talk (Kniffin & Palacio, 2018).

Trash-talk is expressed in a competitive context in which, at least, one person is competing for recognition or status. In other words, trash-talk consists of the communication exchanged between people in which one side, or both, will make proud comments about themselves and insulting comments about the other, trying to boost their ego and demean the other side. Although, every day, in organizational life, it is common to see trash-talk between individuals, is much more common in sports or politics. Yip et al. (2018, p. 126), defined trash-talking as "boastful comments about the self or insulting comments about an opponent that are delivered by a competitor typically before or during competition" and found out that it can motivate both constructive and destructive behavior and increase competition between people.

Contact sports, such as football or wrestling, for example, are more closely associated with trash talk than other sports and, usually, athletes use their skills or physical appearance to trash-talk their opponents (Kniffin & Palacio, 2018). In short, competitors use trash-talk to intimidate, distract or humiliate their opponents and boost their own morale. The former boxing champion Muhammad Ali is an excellent example of a top fighting athlete that notoriously adopted trash-talk against his opponents (Zirin, 2005).

2.4 Analyzing fans and athletes behavior through social media

Fans and athletes influence each other in different ways and social media have a big weight on this relation. In the last two decades, social media became highly popular to support relationships and keep people in contact (Ramos et al., 2019). The opportunity to communicate between fans and athletes is clearly important, with the adoption of online social media becoming the mainstream of such communication. Currently a massive amount of data is generated daily by the millions of users of such systems and this data has been leveraged to study various aspects of human behavior (Ellison et al., 2007; Raacke & Bonds-Raacke, 2008).

Previous researches on users' social media behavior can be used to predict traits, such as, for example, depression and mental illnesses (Guntuku et al., 2017), age and gender (Sap et al., 2014; Schwartz et al. 2013; Jaidka et al., 2018), personality (Schwartz et al. 2013), stress (Lin et al. 2014; Jaidka et al., 2018), empathy (Jaidka et al., 2018) or predict wins (Schumaker et al., 2016). In most of these studies, Facebook or Twitter were used to extract data, so we will focus on these two social media platforms. People use Facebook and Twitter, most of the times, for different purposes. Researches about social media users suggest that they use Facebook to connect with things that are most important for them such as, for example, friends and family (Joinson, 2008; Jaidka et al., 2018) and they use Twitter to connect with athletes or topics of interest due to the unique insights and the proximity between athletes and fans that Twitter provides (Hambrick et al., 2010; Kassing & Sanderson, 2010). Through the users' posts, we can try to predict how they feel. For example, users with higher stress are more likely to express negative emotion, while posting messages with positive emotions seems to be associated with a high state of well-being (Kim & Lee, 2011).

For professional athletes, it is very important to have a positive exposure, either to interact and create engagement with their fans or attract lucrative contracts (Pegoraro, 2010). Since 2008, Twitter has become a popular online social network for professional athletes, and it seems positioned to have a large impact on sports communication because it offers interactivity between athletes and fans and also has a rapid adoption by the sport industry. Athletes use Twitter for many reasons, although surprisingly many of their tweets were not sports-related. They can share and address what they care about most with their fans, such as changes on their professional career, endorsement of products as marketing strategy, personal experiences and moments to keep them informed and increase the engagement (Hambrick et al., 2010; Kassing & Sanderson, 2010; Cunningham and Bright, 2012; Kassing & Sanderson, 2015).

With the valuable opinions of many people available online, it became important to analyze and extract the meaning of that information for different purposes. Social media are valuable sources of information to better understand the motivations and strengthen the relationship with fans and athletes. With this need, many studies have analyzed data in a way that attempts to perceive and explain the fans' and athlete's behavior. Through a Twitter available API (application programming interface), it is possible to collect a large number of tweets which can be analyzed using sentiment analysis tools. Studies benefiting from such approach can help to understand how athlete's use social media for promotional purposes (Hambrick and Mahoney, 2011) and the effectiveness of their recommendations (Cunningham and Bright, 2012), to understand fans' emotional responses in their tweets (Yu and Wang, 2015) or how people respond to highly emotional events and how these emotions vary depending on the context (Gratch et al., 2015). More related with MMA, to analyze the different levels of fan identification related to athletes and the UFC organization (Brown et al., 2013) or to analyze fans reactions after the fighter that they support lose (Salles et al., 2013).

From the studies used to support this research, the ones referenced below were chosen to resume the work done so far and the work we intend to add to the current research. Despite a greater identity with the matter under study, the research of Salles et al. (2013) was not included in the table as we perceive their study credibility does not warrant much certainty. However, it was also analyzed and referenced. All these studies share the goal of analyzing data on fans and athletes. Andrew et al. (2009) and Yip et al. (2018) used questionnaires to obtain the participant's data. Vaccaro et al. (2011) performed interviews with participants and analyzed their behavior during classes and competitions. Yu and Wang (2015) used Twitter to collect data. Tainsky et al. (2013) collect data from all UFC payper-view events from UFC 33 through

UFC 132 and Robbins & Zemanek (2017) collect data from all UFC payper-view events between 2005 and 2016.

For our study, we used data mining to extract data from fans from Twitter (as we see above, this social media is one of the most, if not the most, used platforms by fans and athletes to connect with each other) and SA on this data to analyze their sentiment. The last press conferences with the main event fighters before the fights were analyzed as well to extract the answers from the fighters and, also was used SA to analyze their sentiment.

Regarding fans, their motivations to consume MMA (Andrew et al., 2009; Tainsky et al., 2013) and the sentiments they display during competitions (Yu & Wang, 2015; Salles et al., 2013) were the object of study. As for athletes, attention was given to profile features (Vaccaro et al., 2011), the importance of social media in communication (Hambrick, 2010), the influence of celebrity fighters in PPV numbers (Robbins and Zemanek, 2017) and the influence of trash-talk (Yip et al., 2018). Andrew et al. (2009) studied the drama related to the outcome of the fight. We aim to understand influence that fighters hold on their fans, more specifically, if thrash-talk between athletes, before and during the fight, tends to attract more fans to the sport and what kind of sentiments does it generate in them. In short, we tried to understand if fans consume MMA for the drama and if this drama is related to the trash-talk between the fighters before the fight and if this trash-talk increases the engagement and the negative sentiment shown by their fans. For this, these are our main hypotheses:

Hypothesis 1: Fans demonstrate a greater aggressiveness in the way they express their support in fights between fighters with a greater tendency to provoke their opponents before the fight (Trash-Talk).

Hypothesis 2: Fights between fighters with a greater tendency to provoke their opponents before the fight (Trash-Talk), can create a greater engagement of the fans.

Both hypotheses are grounded on the seven studies summarized in Table 1, with our study contributing to understanding the influence of athlete's behavior on the behavior and engagement of their fans.

Reference	Goal	Method	Findings
Andrew et al. 2009	Explored how nine motives impact media and merchandise consumption among consumers of MMA.	It was used a 43-item questionnaire to study some information about 162 consumers at a professional MMA event.	Contrary to what was believed, violence was not the strongest motive for fans attending MMA events. Drama and aesthetic are the strongest motives.
Hambrick (2010)	Examined Twitter use among professional athletes who use Twitter to communicate with fans and other players.	A sample of 1962 tweets from 101 Twitter athletes accounts was analyzed.	Twitter helps users meet needs and may lead to increased fans' identification with the athletes and their teams.
Vaccaro et al. (2011)	Analyze MMA fighters' fears, how they managed them, and how they adopted intimidating personas to evoke fear in opponents.	121 interviews were made, and 100 practices were observed in a gym. The fighters' behavior was also observed in 10 competitions.	Fighters suppressed fear and evoked confidence. Fighters also use language and their bodies to enact intimidating personas to instill fear in opponents.
Tainsky et al. (2013)	Used a consumer-theory modeling to estimate UFC pay-per-view purchases.	It was collected data from all UFC payper-view events from UFC 33 through UFC 132. 93 observations were included in the study.	There are consumers preferences for specific fighters and specific weight classes. It was noted that consumption rates increase substantially with title fights between the most popular fighters.
Yu & Wang (2015)	Analyzed the sentiments in U.S. sports fans' tweets during five 2014 FIFA World Cup games.	The Twitter search API was used as well as SA to examine U.S. soccer fans' emotional responses in their real-time tweets	Sports fans use Twitter for emotional purposes; big data approach was used to analyze sports fans' sentiment and showed generally consistent results.
Robbins & Zemanek (2017)	Analyze the influence of high-profile celebrity fighters in the PPV numbers and compare it with highest ranked pound-for-pound fighter.	Extract PPV numbers from UFC events from 2005 to 2016. 155 events and 204 fighters were analyzed.	The results shown that celebrity has far more economic value than fighting skill.
Yip et al. (2018)	Study of how trash-talking increases the psychological stakes of competition and motivates targets to outperform their opponents.	The research was divided in six studies, each one analyzing the participants' behavior in different situations and had, at least, 142 participants.	Trash-talking is a common workplace behavior that can foster rivalry and motivate both constructive and destructive behavior.

Table 1 - Literature review of athletes' influence on fans and trash-talk.

3. Methods

3.1 Approach

To meet our hypothesis, we had to find a way to select the UFC events to be analyzed, extract information from the tweets and press conferences revolving around the events, perform a SA on tweets from fans and answers from fighters and analyze the results.

We start by looking for a way that allows us to find and select the most searched for UFC events. For that purpose, we choose Google trends. Over the years, Google's services have been used to extract data to help predict a different variety of outcomes and, according to Choi and Varian (2012) claiming's "Google Trends may help in predicting the present". Google trends provides the volume of searches that users have done on a particular topic, within a geographic area, and over a period of time (Choi & Varian, 2012). The volume of searches (or interest over time) is calculated using the following formula:

$$\text{Volume} = \frac{Total\ number\ \ of\ queries\ for\ the\ keyword\ in\ the\ region\ during\ the\ time\ period}{Total\ number\ of\ queries\ in\ the\ region\ during\ the\ time\ period}$$

The data goes back to January 1, 2004 (Choi & Varian, 2012), which is ideal for us because the oldest event we have access to, happened in 2014.

Next, we choose Twitter to extract the data. As we see in the literature review, Twitter is a widely used social media platform that users use to share their real-time reactions and emotions and, can be used to explain, detect or predict various traits (Yu & Wang, 2015). However, after a little research, we found that Twitter Official API has a time limitation, that does not allow us to get tweets older than a week, unless you subscribe to the premium account (Twitter, n.d.). After searching for a solution for this problem, we found that some people have the same issue and solved it with a script that uses the Twitter Search. Basically, when you do a search on Twitter a scroll loader starts and, if you scroll down you start to get more and more tweets, all through calls to a JSON provider, that allow us to search and extract the deepest, oldest tweets (Henrique, 2016).

The SA was performed using the VADER algorithm, that has been already applied to analyze text from social media (Costa et al., 2019). VADER uses a human-validated ground, based on twitter, movies, product reviews and opinion new articles, to perform SA. Regarding performance, the authors conclude that, in most cases, VADER performed better than other highly regarded SA tools. One of its advantages it is that VADER takes into consideration several factors, such as capital letters, excess of punctuation or emojis, among others, that usually are forgotten and that help to improve the accuracy of the analysis (Hutto & Gilbert, 2014).

For each sentence (tweets and answers), the method "polarity_scores (String)" returns 4 metrics: Positive, Negative, Neutral, and Compound. The Positive, Negative and Neutral scores represent the proportion of sentence that falls in these categories, in other words, represent the percentage of words that have a positive, negative and neutral sentiment. The Compound score is a metric that calculates the sum of all the lexicon ratings which have been normalized between -1 (most negative) and +1 (most positive) (Hutto & Gilbert, 2014). Basically, we used compound to determine if a sentence is positive or not. Figure 1 shows some examples of VADER sentiment classification:

```
#Baseline sentence
sentiment_analyzer_scores('The service here is good')
```

```
The service here is good---------------- {'neg': 0.0, 'neu': 0.58, 'pos': 0.42,
'compound': 0.4404}
```

```
#Degree Modifiers
print(sentiment_analyzer_scores('The service here is extremely good'))
print(sentiment_analyzer_scores('The service here is marginally good'))
```

```
The service here is extremely good------ {'neg': 0.0, 'neu': 0.61, 'pos': 0.39,
'compound': 0.4927}
None
The service here is marginally good----- {'neg': 0.0, 'neu': 0.657, 'pos': 0.34
3, 'compound': 0.3832}
None
```

```
#Conjunctions
sentiment_analyzer_scores('The food here is great, but the service is horrible')
```

```
The food here is great, but the service is horrible {'neg': 0.31, 'neu': 0.523,
'pos': 0.167, 'compound': -0.4939}
```

Figure 1 - Examples of a result from VADER sentiment analysis.

As we can see, the compound value depends on the other three metrics. If the higher value that we have is positive, the compound will be higher than zero, if it's neutral, the compound value will be zero and if it's negative, the compound value will be lower than zero.

We also analyzed the profanity in each sentence. For that purpose, we chose the Python library, "profanity-check". According the author, Zhou (2019), this library was developed with the objective of filling some flaws, like limited number of words used to detect profanity, found in existing libraries, such as "profanity", "profanity-filter", “profanityfilter” and “better-profanity” libraries.

Profanity-check uses a linear support vector machine model, trained on 200k human-labeled samples of text strings. Comparing with the other libraries, “profanity” uses a 32 words list, “better-profanity” uses a 140 words list and “profanityfilter” uses a 418 words list. In addition, “profanity-check” presents a 95% accuracy test against 91.8% and 85.6% of the "profanity-filter" and "profanity" libraries, respectively. As far as prediction is concerned, profanity-check is anywhere from 1.5 - 7 times faster than “profanity” and 300 - 4000 times faster than “profanity-filter”. Profanity-check bases its classifications on data to avoid being subjective. To do that, it combines a dataset from two sources, comments from Wikipedia’s pages and tweets scraped from Twitter. From this library, we used the “predict([string])” and “predict_prob([string])” methods to analyze each sentence. Predict takes an array and returns 1 if the string have profanity or 0 if not. Predict_prob takes an array and returns the probability each string is offensive.

3.2 Data selection

As explained above, we used Google Trends to determine the events we were to review and the inclusion criterion was: 25 Events with the highest volume in the last 5 years, worldwide. The designation of the event usually consists in the word "UFC” plus the event Number, (e.g., UFC229), with the exception of non-pay-per-view events, which have the words "UFC Fight Night” or “UFC On Fox” plus the event number (e.g., UFC Fight night 83 or UFC On Fox 16),

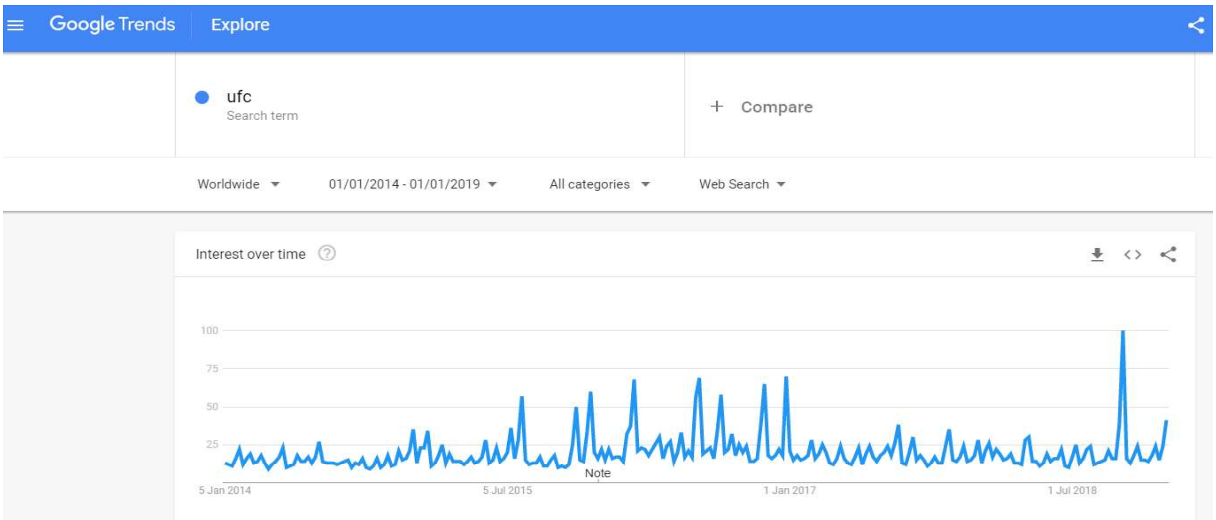

Figure 2 - Google's trends search about UFC from the last five years.

but these were not analyzed. That said, Figure 2 shows our first search, the query was "UFC", the region was "Worldwide" and the time period was between 01/01/2014 and 01/01/2019:

The numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means that there was not enough data for this term.

After we get the chart, we extracted its .csv file with the data and selected the 25 most interesting events. However, Google trends only gives us the date (year and month) and the value of the interest, not of the day nor the description associated with each value, which makes it difficult to understand the event that refers to that volume.

To work around this problem, for each value selected, we did a new search (figure 3) but this time only in the time period, year and month, corresponding to the value. As an example, if the date was 06/10/2018, we did a new search with the time period between 01/10/2018 and 31/10/2018:

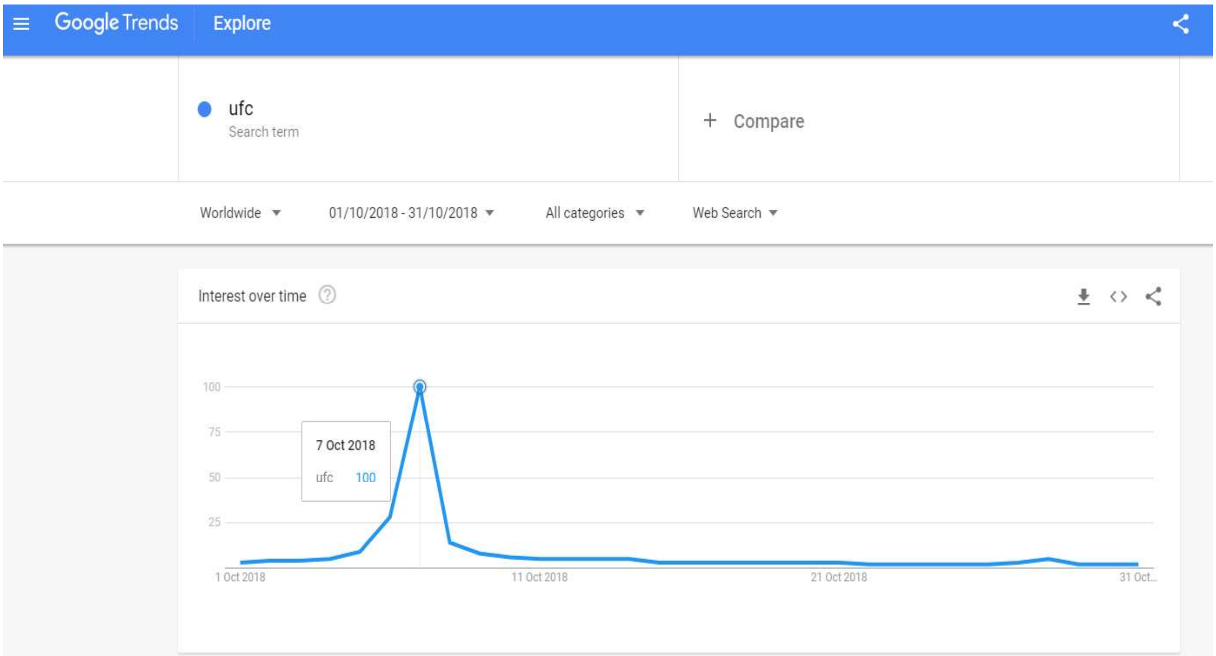

Figure 3 - Google's trends search about UFC from the month of the fight.

With this we get a new chart of interest this month, but as the search is smaller, Google trends already tells us the day corresponding to each volume. Once we had the day, we just had to conduct a search to find out what event it refers to. Table 2 describes all 25 events in terms of date, volume, event and main event (name of the fighters from the main event):

Date	Volume	Event	Main event
06/10/2018	100	UFC229	Conor McGregor vs Khabib Nurmagomedov
25/12/2016	69	UFC 207	Amanda Nunes vs Ronda Rousey
05/03/2016	68	UFC 196	Conor McGregor vs Nate Diaz
10/07/2016	68	UFC 200	Daniel Cormier vs Jon Jones II
12/11/2016	64	UFC 205	Conor McGregor vs Eddie Alvarez
12/12/2015	60	UFC 194	Jose Aldo vs Conor McGregor
01/08/2015	58	UFC 190	Ronda Rousey vs Bethe Correira
20/08/2016	57	UFC 202	Conor McGregor vs Nate Diaz II
15/11/2015	50	UFC 193	Ronda Rousey vs Holly Holm
29/12/2018	41	UFC 232	Jon Jones vs Alexander Gustafsson II
29/07/2017	38	UFC 214	Jon Jones vs Daniel Cormier III
11/07/2015	36	UFC 189	Conor McGregor vs Chad Mendes
28/02/2016	36	UFC Fight Night 84	Anderson Silva vs Michael Bisping
04/11/2017	35	UFC217	Michael Bisping vs George St. Pierre
03/01/2015	34	UFC182	Jon Jones vs Daniel Cormier
01/02/2015	34	UFC183	Anderson Silva vs Nick Diaz
10/09/2016	33	UFC203	Stipe Miocic vs Alistair Overeem
04/06/2016	32	UFC199	Luke Rockhold vs Michael Bisping II
09/09/2017	31	UFC215	Amanda Nunes vs Valentina Shevchenko
21/02/2016	31	UFC Fight night 83	Donald Cerrone vs Alex Oliveira
23/04/2016	30	UFC197	Jon Jones vs Ovince St. Preux
08/04/2018	30	UFC223	Khabib Nurmagomedov vs Max Holloway
30/12/2017	28	UFC219	Cris Cyborg vs Holly Holm
26/07/2015	28	UFC On Fox 16	Tj Dillashaw vs Renan Barão II
24/05/2015	28	UFC187	Anthony Johnson vs Daniel Cormier

Table 2 - 25 events with the highest interest in the last five years.

After we had the events, we applied several exclusion criteria that will be explained next (Figure 4 helps to better understand the whole process). The first exclusion criterion was that if the event is not a pay-per-view event because we want to use the pay-per-view numbers as a metric to perceive the engagement in our analysis.

Next, we went to research the last pre-fight conference (either press conference or media conference call) for each event. In case we can't find the conference, or the event did not have a conference before the fight, the respective event was excluded (second exclusion criterion).

The third exclusion criterion was if any of the fighters of the main event were replaced before the fight and we don't have a pre-fight press conference with his substitute. In some cases, fighters are unable to fight due to injuries, health problems or test positive for performance enhancing drugs (or doping).

The fourth exclusion criterion was that the press conference was too far from the fight because, for some events, the conferences we found took place quite a long way from the fight, in some cases almost two months earlier and that may affect the fans emotions regarding the fight, because emotions may decay over time (Garrett & Maddock, 2001). With this in mind, we exclude all events that the pre-fight press conference didn't occur in the fight week (between Monday and Sunday).

The fifth exclusion criterion was the number of tweets that we extract related to the event. For a better analysis, we set a minimum threshold of 1,000 tweets by event, so all events with the total number of tweets below that were excluded. In short, these were the exclusion criteria that we used:

(1) Excluded because is not a pay-per-view event;

(2) Excluded because there is no pre-fight press conference;

(3) Excluded because one of the fighters was replaced;

(4) Excluded because the pre-fight press conference is too far from the event;

(5) Excluded because the number of tweets is too low.

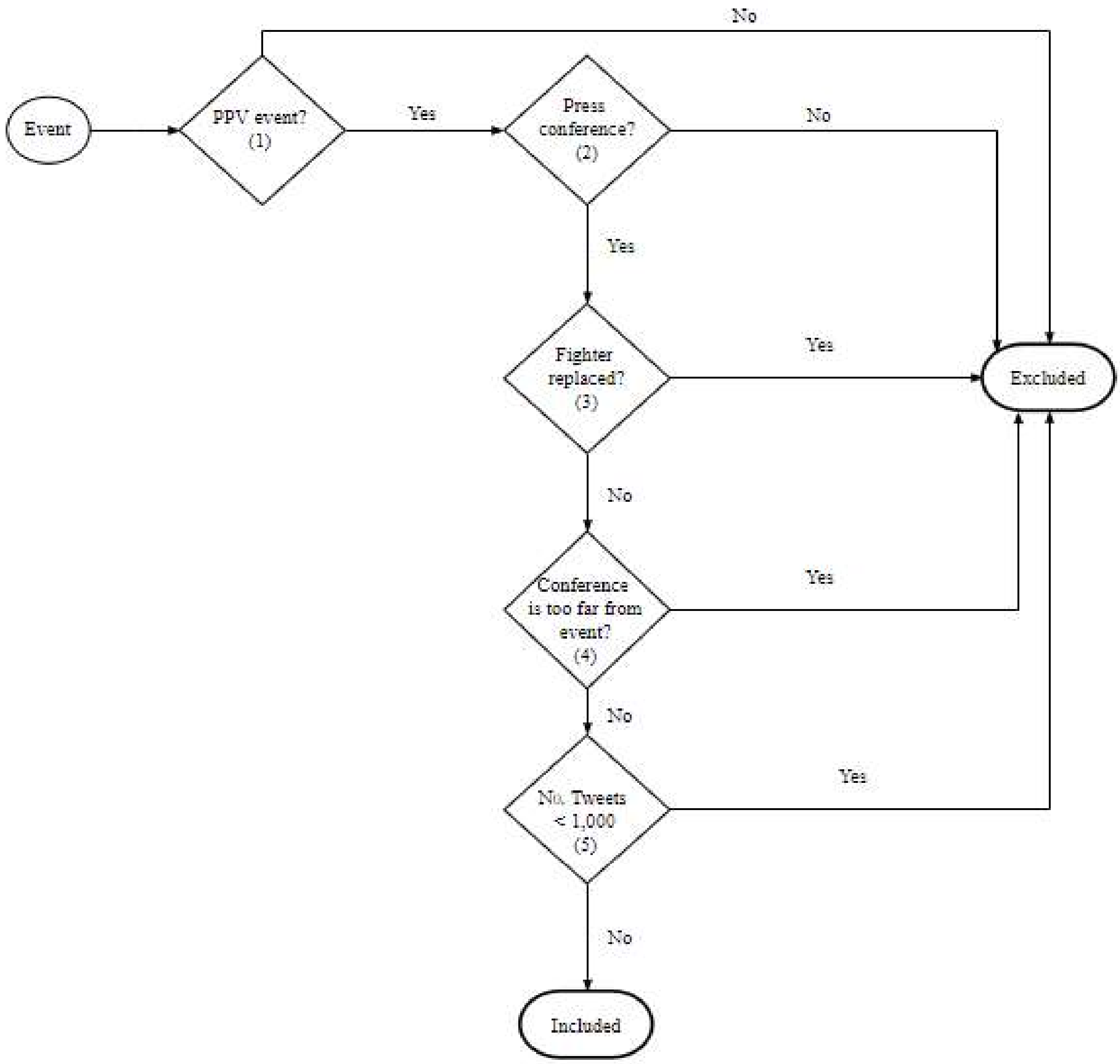

Figure 4 - Flow chart with a simple representation of the exclusion process.

After applying the exclusion criterion, from the initial 25 events, we were left with 12 events (highlighted in gray in table 3). Table 3 describes all events and their status (the numbers refer to the five-exclusion criterion), whether or not they had a press conference and whether they were included in the analysis or were excluded and the reason for the exclusion.

Date	Volume	Event	Main event	Press Conference?	Conference date	Days (*)	Status
06/10/2018	100	UFC229	Conor McGregor vs. Khabib Nurmagomedov	Yes	04/10/2018	2	Included
25/12/2016	69	UFC 207	Amanda Nunes vs. Ronda Rousey	No	N/A	N/A	Excluded (2)
05/03/2016	68	UFC 196	Conor McGregor vs. Nate Diaz	Yes	03/03/2016	3	Included
10/07/2016	68	UFC 200	Daniel Cormier vs. Jon Jones II	Yes	06/07/2016	4	Excluded (3)
12/11/2016	64	UFC 205	Conor McGregor vs. Eddie Alvarez	Yes	10/11/2016	5	Included
12/12/2015	60	UFC 194	Jose Aldo vs. Conor McGregor	Yes	09/12/2015	3	Included
01/08/2015	58	UFC 190	Ronda Rousey vs. Bethe Correira	Yes	27/07/2015	5	Included
20/08/2016	57	UFC 202	Conor McGregor vs. Nate Diaz II	Yes	17/08/2016	3	Included
15/11/2015	50	UFC 193	Ronda Rousey vs. Holly Holm	Yes	16/09/2015	60	Excluded (4)
29/12/2018	41	UFC 232	Jon Jones vs. Alexander Gustafsson II	Yes	27/12/2018	2	Included
29/07/2017	38	UFC 214	Jon Jones vs. Daniel Cormier III	Yes	26/07/2017	3	Included
11/07/2015	36	UFC 189	Conor McGregor vs. Chad Mendes	Yes	09/07/2015	2	Included
28/02/2016	36	UFC Fight night 84	Anderson Silva vs. Michael Bisping	Yes	29/12/2015	61	Excluded (1)
04/11/2017	35	UFC217	Michael Bisping vs. George St. Pierre	Yes	02/11/2017	2	Included
03/01/2015	34	UFC182	Jon Jones vs. Daniel Cormier	Yes	29/12/2014	5	Included
01/02/2015	34	UFC183	Anderson Silva vs. Nick Diaz	Yes	17/11/2014	76	Excluded (4)
10/09/2016	33	UFC203	Stipe Miocic vs. Alistair Overeem	Yes	08/09/2016	2	Excluded (5)
04/06/2016	32	UFC199	Luke Rockhold vs. Michael Bisping II	Yes	02/06/2016	2	Included
09/09/2017	31	UFC215	Amanda Nunes vs. Valentina Shevchenko	Yes	31/08/2017	9	Excluded (4)
21/02/2016	31	UFC Fight night 83	Donald Cerrone vs. Alex Oliveira	No	N/A	N/A	Excluded (1)
23/04/2016	30	UFC197	Jon Jones vs. Ovince St. Preux	Yes	17/04/2016	6	Excluded (4)
08/04/2018	30	UFC223	Khabib Nurmagomedov vs Max Holloway	Yes	04/04/2018	4	Excluded (3)
30/12/2017	28	UFC219	Cris Cyborg vs. Holly Holm	Yes	21/12/2017	9	Excluded (4)
26/07/2015	28	UFC On Fox 16	Tj Dillashaw vs. Renan Barão II	No	N/A	N/A	Excluded (1)
24/05/2015	28	UFC187	Anthony Johnson vs. Daniel Cormier	Yes	N/A	N/A	Excluded (3)

Table 3 - 12 final events after applying the exclusion criteria.

(*) Difference of days between the pre-fight press conference and the fight.

For the 12 events, we reviewed the conferences and extracted the answers from the two main event fighters. The main event is the last fight of the main card of the event and is usually the most expected fight by the fans because it always evolves a title fight, where the champion of a category is decided or a fight between fighters who are having a hype of recognition and interest from the media and fans (Robbins and Zemanek, 2017; Mazique, 2018). Figure 5 shows an example of a main card and main event from a UFC event:

Figure 5 – UFC229 main card.

After reviewing the conference, we extracted the tweets from the fans. For this we used the name, nickname and the official twitter account of each fighter, hashtags related with the event and with fighters, in the search queries (Annex A).

The conference usually takes place 3 to 5 days before the event and the main card of the event starts around 3am and ends around 6am (GMT+1 Portugal time) on Sunday. With this, we extracted tweets (Figure 6) that relate to the week of the event, that is, from Monday at 00:00 until Sunday at 23:59. In this way, we are able to extract tweets from different heights in time, before the conference, after the conference, before, during and after the fight (only english tweets were considered).

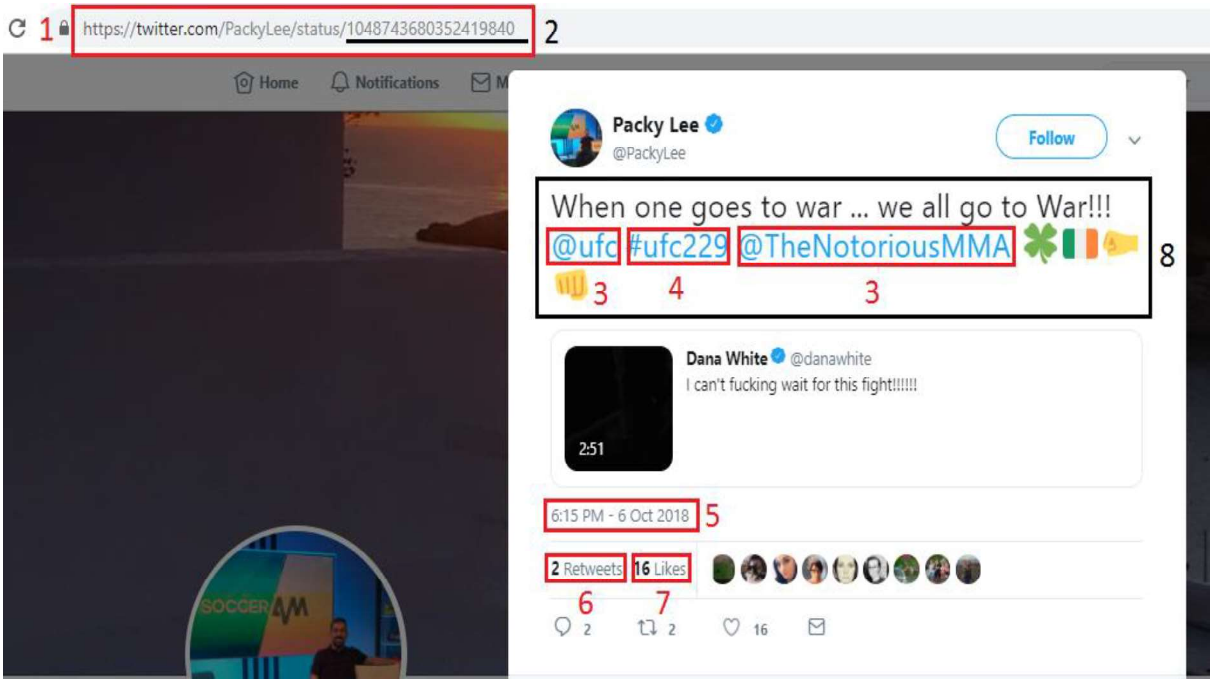

Figure 6 - Example of a tweet.

Regarding pay-per-view numbers, we looked at the official UFC website and found nothing that showed the official pay-per-view number of the events we want to review. We also sent two emails to UFC requesting this information (Annex C), but we did not get any answer. Having said that, after analyzing several sites related to MMA, we realized that the pay-per-view numbers they provide do not vary much between them. Thus, we chose to cross-data between the sites (Rosenman, 2018; “Pay-per-view”, n.d.; “Pay Per View Buys”, n.d.; GypsyGold, 2018; Fox, 2018) and a scientific article (Robbins & Zemanek, 2017) and make an average of the pay-per-view for each event.

Table 4 describes the input and output variables by variable name, source, data type, variable type, description and status (the first 8 variables with numbers correspond to the numbers on the figure 6):

Variable Name	Source	Data Type	Variable Type	Description	Status
permanentlink [1]	Twitter	url	Output	Tweet link	Used
id [2]	Twitter	Integer	Output	Tweet identification	Used
mentions [3]	Twitter	String	Output	Mentions to other pages made in the tweet	Not used
hashtags [4]	Twitter	String	Output	Hashtags mentioned in tweet	Not used
hour [5]	Twitter	Hour	Output	Hour	Used
date [5]	Twitter	Date	Output	Date	Used
retweets [6]	Twitter	Integer	Output	Number of times the tweet was shared	Not used
likes [7]	Twitter	Integer	Output	Number of likes	Not used
text [8]	Twitter	String	Output/Input	Tweet (text, hashtags, mentions, symbols etc ...)	Used
answer	Press conference	String	Output/Input	Answers from fighters in the press conferences	Used
predict	Profanity_check	Integer	Output/Input	Describe if sentence is offensive or not	Not used
predict_prob	Profanity_check	Double	Output	Probability each string is offensive	Used
neg	VADER	Double	Output/Input	% of the sentence with negative sentiment	Not used
pos	VADER	Double	Output/Input	% of the sentence with positive sentiment	Not used
neu	VADER	Double	Output/Input	% of the sentence with neutral sentiment	Not used
compound	VADER	Double	Output/Input	Sentiment final result	Used
sentiment	Calculated	Percentage	Output	Describe the type of sentiment	Used
negSAConfer	Calculated	Percentage	Output	% of negative sentiment in conferences	Used
negSATweets	Calculated	Percentage	Output	% of negative sentiment of tweets	Used
avgProfanityConf	Calculated	Percentage	Output	Average profanity of the conferences	Used
avgProfanityTweets	Calculated	Percentage	Output	Average profanity of the tweets	Used
fighter	Calculated	String	Output	Name of the fighter that the tweet refers to	Used
averagePPV	Calculated	Percentage	Output	Average pay-per-view by event	Used
numTotalweets	Calculated	Integer	Output	Number total of tweets by event	Used

Table 4 – Input and output variables.

3.3 Data extraction and cleaning

We had to create a script that uses Twitter Search to extract tweets from a specific time period (greater than one week). The Twitter search url always starts the same way, only changing the search we want to do, which makes our work easier. We used the "HttpClient" package from java to access the url and we gave as input a string that contained the query search and the time period from which we intended to extract tweets.

Here is an example of the search and the url:

querysearch="conor mcgregor" since=2018-10-01 until=2018-10-08

```
String url =
String.format("https://twitter.com/i/search/timeline?lang=engb&f=realtime&q=%s&src=typd&max_position=%s", URLEncoder.encode(querysearch, "UTF-8"), scrollCursor);
```

We access the url and save the pages in a JSON variable. The "scrollCursor" is initially passed as null. However, after we have the JSON we use it to get the position and increment the "scrollCursor". Here is an example of the url:

"https://twitter.com/i/search/timeline?lang=en-gb&f=realtime&q=+since%3A2018-10-01+until%3A2018-10-08+conor+mcgregor&src=typd&max_position=null"

After we used the url, a list of tweets with their information is retrieved and exported to a .csv file. Figure 7 shows an example of an extraction result:

	A	B	C	D	E	F	G	H
1	date	retweets	favorites	text	mentions	hashtags	id	permalink
2	08/10/2018 00:52	5	6	Here's my latest article for @SportsTalkFLA on the many options @TheNotoriousMMA has next after his loss last night to @TeamKhabib. #UFC #UFC229 #ConorvsKhabib #ConorKhabib #ConorMcGregor #KHABIBvsMcGregor #KhabibTime #KhabibVsConorhttps://www.sportstalkflorida.com/more/mma/whats-next-for-conor-mcgregor-2/ …	@SportsTalkFLA @TheNotoriousMMA @TeamKhabib	#UFC #UFC229 #ConorvsKhabib #ConorKhabib #ConorMcGregor #KHABIBvsMcGregor #KhabibTime #KhabibVsConorhttps	1,04909E+18	https://twitter.com/Greg_LaFountain/status/1049085204160294912

Figure 7 - Example of an extracted tweet.

After the tweets were extracted, we had to clean the data. This cleaning consisted in dividing the date into date and time, removing duplicates, and detecting the language in which the tweet was written.

We used the tweet “id” to delete the duplicates (in some cases the “id” came as a null values, so we used the number at the end of the permanent link) and as we looked at a sample of 4000 tweets, almost all duplicates were removed. However, there are cases where tweets are the same, changing only the link from the video or picture, but the script assumes it is different and assigns different ids. We found out that this happens when is a different user to post the same tweet, probably a retweet, so we did not consider it as duplicate (Annex B, figure 24). Or when it’s official pages like UFC, MMA blogs, Sports Channels, etc… that post the same tweet multiple times, but at different times (Annex B, figure 25). These we considered as duplicates. We found 38 examples of duplicated tweets, which represents less than 1% of the total analyzed tweets, so we accepted these duplicates.

To identify the language of each tweet we used the Google spreadsheet function, “detectlanguage”. After applying the function, we analyzed a sample of 1000 tweets identified as “en” (which corresponds to English) and all were correct. However, some tweets classified with other languages or undefined seem to have some errors. For example, the tweet “He’s baaaaack. # UFC229 pic.twitter.com/nynJQdv4Uy “was classified as” hi “(Hindi) and clearly is an english tweet that was written in the wrong way with hashtags and an image, which may confuse the function. Since the number of wrongly classified tweets was low, we chose to exclude all the remaining languages and only consider tweets identified as “en”.

Once we cleaned all data, we ran the python script for each sentence to get the SA and profanity in each one. The script read a .csv file, that contains the sentences that we want to analyze, and returns other .csv file with the sentences and their respective SA and profanity results. The sentence and it’s results were separated by “Â»” (that is, sentence Â» SA results Â» profanity results), which made it easier to divide the data into columns. Figure 8 shows two examples of the output:

	A
1	he knew what he signed up for. ive been late. the traffic is heavy. hes better off running anyway. you know yourself. i bet you he was saying he didnt say anything the last time so, i mean, whatever. it is what it is. these things happen. Â»{'neg': 0.0, 'neu': 0.938, 'pos': 0.062, 'compound': 0.4404}Â»[0.04087689]Â»[0]
2	you will never beat the fucking irish. the irish are back in town. Â»{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}Â»[0.68870803]Â»[1]

Figure 8 - Example of a VADER output.

After we separate the data in columns, we extract the fighter that the tweet mentions, whether it’s just about one of the fighters, about both or none of them (just tweets about the event), this helped us to. Figures 9 shows two examples of tweets and answers on which we worked:

Date	Hour	Tweet	Compound	Sentiment	Avg Profanity	Fighter
07/10/2018	06:23:00	All you McGregor fans are so hurt. Unprofessional or not he got murdered on live TV.	-0.6628	Neg	0.12955496	McGregor
07/10/2018	07:19:00	Very classy message from Conor McGregor's coach... #UFC229pic.twitter.com/2j1Zzygee1	0.4927	Pos	0.1314912	McGregor

Answer	Compound	Sentiment	Avg Profanity
he knew what he signed up for. ive been late. the traffic is heavy. hes better off running anyway. you know yourself. i bet you he was saying he didnt say anything the last time so, i mean, whatever. it is what it is. these things happen.	0.4404	Pos	0.04087689
he's afraid, make no mistake. he has nervous reactions, he's a flincher we called him. he's easily backed up. i'm very prepared and i know what to expect. it's nothing that phases me.	-0.2516	Neg	0.03219235

Figure 9 - Examples of tweets and answers.

4. Results and discussions

4.1 First hypothesis

We extracted a total of 72,919 tweets but, considering we only reviewed the conferences of the main event fighters, we opted only to consider tweets directly related to these fighters. Excluding either the tweets related to other fights of the event or the tweets that mention the two fighters, we got a total of 32,360 tweets. Table 5 describe the SPSS data that we used to study the first hypothesis:

Event Number	Event	Fighter	NegSA Conf	AvgPro fanConf	NegSA Tweets	AvgProfa nTweets	AvgPPV	TotalT weets
1	UFC232	Jon Jones	16.67	22.36	27.13	16.28	675,000	1,209
2	UFC232	Alexander Gustafsson	39.29	25.7	28	15.74	675,000	100
3	UFC229	Conor McGregor	22.22	30.31	27.11	19.36	2,233,333	3,073
4	UFC229	Khabib Nurmagomedov	30	14.41	32.75	22.82	2,233,333	3,527
5	UFC217	George St-Pierre	20	11.16	15.41	14.89	875,000	1,181
6	UFC217	Michael Bisping	51.72	51.07	22.19	15.15	875,000	302
7	UFC214	Jon Jones	15	15.73	14.41	14.14	855,000	1,020
8	UFC214	Daniel Cormier	22.22	13.22	19.49	13.98	855,000	313
9	UFC205	Conor McGregor	14.71	23.48	13.21	14.37	1,300,000	3,331
10	UFC205	Eddie Alvarez	15.38	15.38	20.78	13.77	1,300,000	154
11	UFC202	Conor McGregor	60	63.46	15.98	14.38	1,641,667	2,935
12	UFC202	Nate Diaz	31.25	32.76	16.76	18.27	1,641,667	859
13	UFC199	Luke Rockhold	17.65	23.18	22.28	13.51	310,750	184
14	UFC199	Michael Bisping	16.13	34.33	10.53	12.94	310,750	817
15	UFC196	Conor McGregor	14.71	18.69	22.09	14.93	1,394,667	3,074
16	UFC196	Nate Diaz	37.04	33.15	19.27	20.68	1,394,667	633
17	UFC194	Conor McGregor	7.14	4.02	13.1	14.49	1,112,500	2,564
18	UFC194	Jose Aldo	30	8.49	20.25	11.94	1,112,500	395
18	UFC190	Ronda Roussey	8.7	16.9	17.56	17.16	900,000	3,639
20	UFC190	Bethe Correia	14.29	7.38	25.28	18.62	900,000	443
21	UFC189	Conor McGregor	21.43	16.12	15.19	13.71	825,000	1,284
22	UFC189	Chad Mendes	22.22	29.49	23.65	12.5	825,000	148
23	UFC182	Jon Jones	7.69	10.7	20.05	16.39	620,000	823
24	UFC182	Daniel Cormier	9.09	5.06	21.31	11.96	620,000	352

Table 5 – SPPS data set used for the first hypothesis.

In order to study the first hypothesis "Fans demonstrate greater aggressiveness in the way they express their support in fights between fighters with greater tendency to provoke their opponents before the fight – Trash-Talk”, first, we began by analyzing the results of the SA and profanity of the answers of the fighters in the conferences, because they are much less than the tweets, to see if the results make sense.

We saw that in many cases, when fighters used the word “Fight”, VADER classified their answer with negative sentiment. Given that we are talking about MMA, which is a fight sport, in this context, the word “Fight” has the same meaning as “game” or “match”, which have a positive sentiment. Thus, we looked for the file that VADER uses to classify the words (vader_lexicon.txt), we changed the feeling of the word “Fight” from negative to positive and we performed the analysis of sentiments and profanity again. The results of UFC202 and UFC217 caught our attention as well, because some answers seemed to be misclassified taking into account their context and the fighter’s posture. Therefore, we looked again at the VADER lexicon file and we realized that in the case of UFC202, VADER was not taking into consideration 3 phrases as negative because it did not have the word “f***ing” in the file. In the case of UFC217, the word “a**hole” was not in the file either. Since both of these words are normally used in a negative sense (which is also seen after reviewing the conferences and the context in which they are used), we edited the file and added both words with a negative feeling (we attribute the same sentiment to the word “ f***ing “ as to the word “f**k”, and to the word “a**hole” as to the word “jack**s”, which is a synonym).

We began by comparing the negative sentiment (Figure 10) and average profanity (Figure 11) between fighters and events to confirm there is diversity in each fighter’s approach to press conferences in using trash-talk.

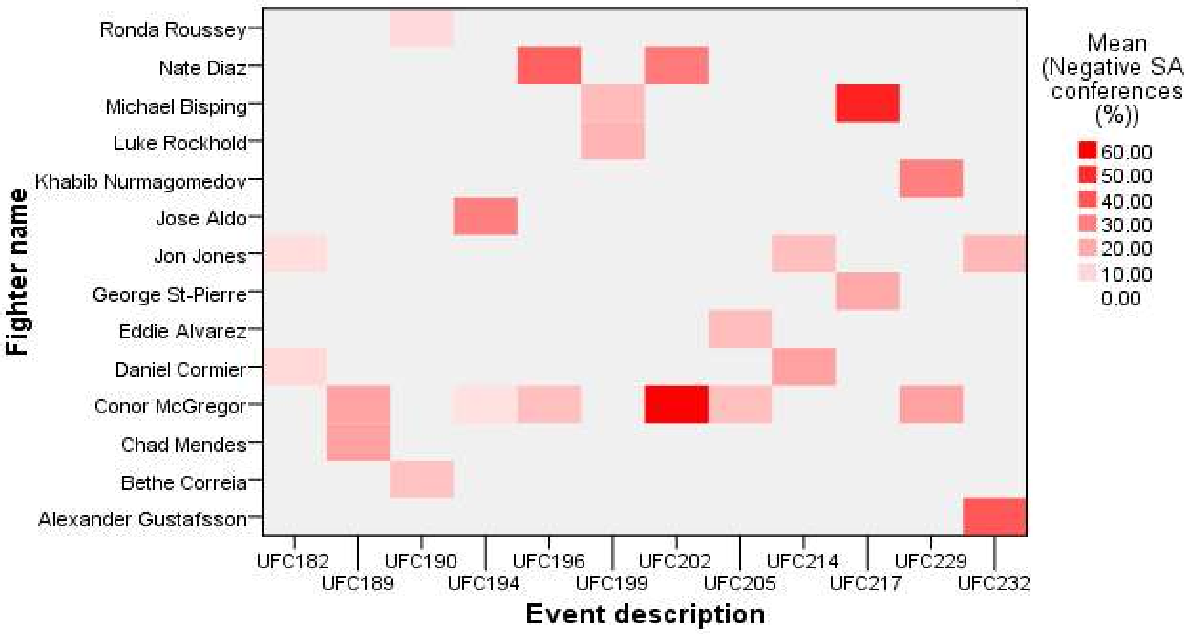

Figure 10 - Negative SA in conferences by event and fighter.

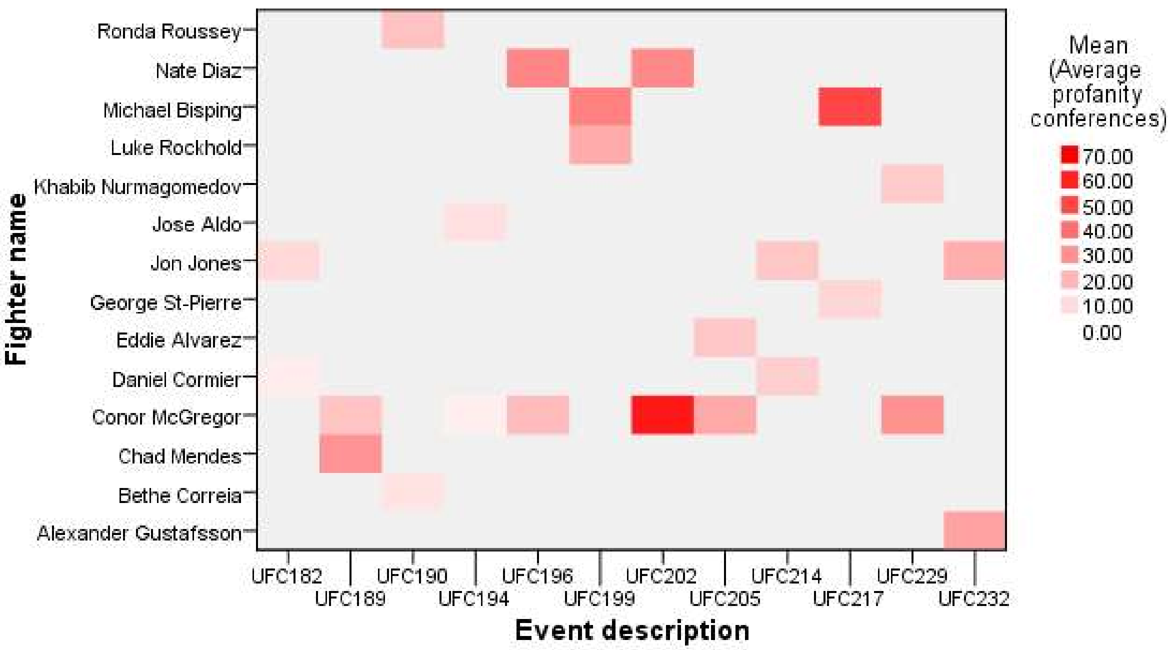

Figure 11 - Average of profanity by event and fighter.

We saw that there are big differences between fighters, and between events. There are fighters who show much higher values of negative SA and profanity in conference than others (for example, Conor McGregor on UFC 202 or Michael Bisping on UFC217); there are fighters who show differences (in some cases, big differences) in the negative SA and profanity values depending on the event (for example, Conor McGregor on UFC194 and UFC202), which may suggest that the behavior of the fighter changes according to the context of the event.

Secondly, we considered as trash-talk all the answers given by the fighters in the conferences which have a negative SA or profanity. To assess H1, we computed the Pearson correlation between the negative sentiment in the answers with the negative sentiment in the tweets (Figure 12), and between the average profanity in the answers with the average profanity in the tweets (Figure 14) and, to better understand these relations, we also drew their respective dispersion chart (Figures 13 and 15, respectively):

		Negative SA tweets (%)	Negative SA conferences (%)
Negative SA tweets (%)	Pearson Correlation	1	0.165
	Sig. (2-tailed)		0.440
	N	24	24
Negative SA conferences (%)	Pearson Correlation	0.165	1
	Sig. (2-tailed)	0.440	
	N	24	24

Figure 12 - Correlation between the negative SA in conferences and tweets.

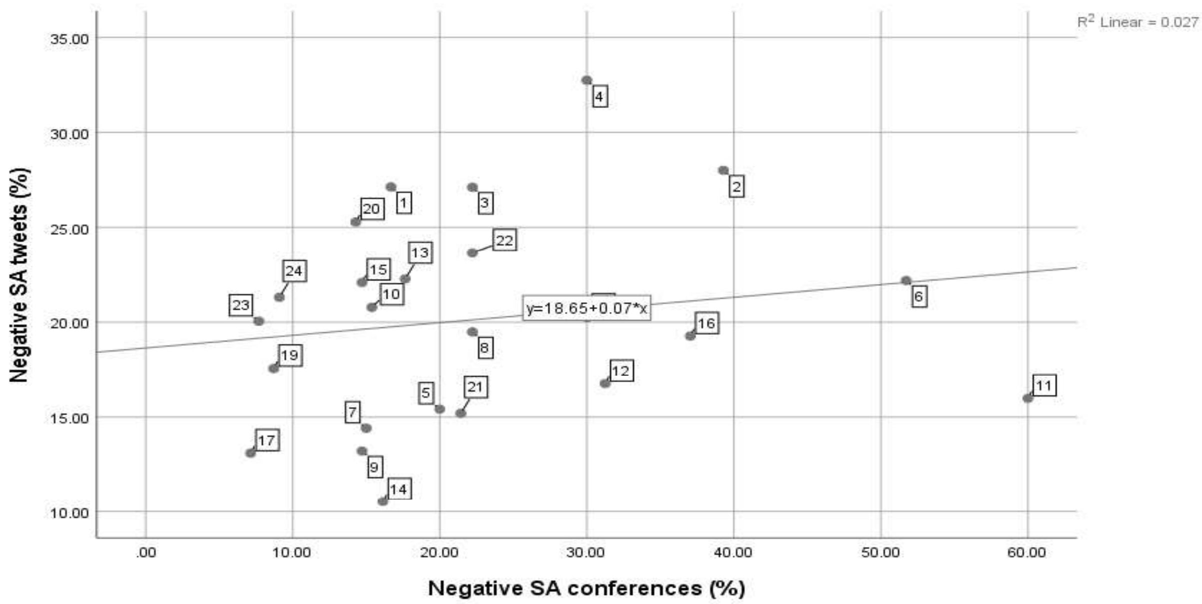

Figure 13 - Dispersion chart to relate negative SA in conferences and tweets.

		Average profanity conferences	Average profanity tweets
Average profanity conferences	Pearson Correlation	1	0.072
	Sig. (2-tailed)		0.739
	N	24	24
Average profanity tweets	Pearson Correlation	0.072	1
	Sig. (2-tailed)	0.739	
	N	24	24

Figure 14 - Correlation between the average profanity in conferences and tweets.

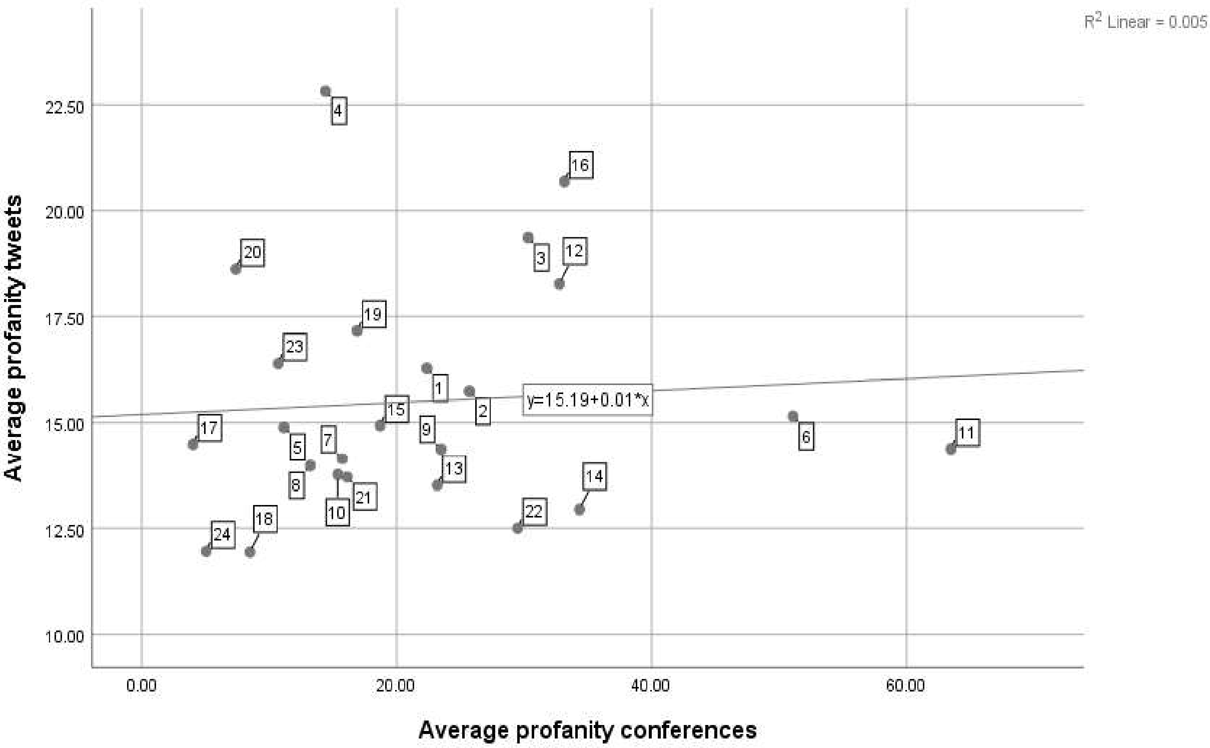

Figure 15 - Dispersion chart to relate average profanity in conferences and tweets.

For the former (figures 12 and 13), no significant linear relation was found between the negative sentiment in the tweets and the negative sentiment in the answers (R=0.165, p-value=0.440) and for the latter (figures 14 and 15), no significant linear relation was found between the avg profanity in the tweets and the avg profanity in the answers (R=0.072, p-value=0.739). Thus, in both cases, we can't prove the relation between fighters' trash-talk and fans' tweets, therefore H1 is not confirmed. However, the results have drawn our attention to something that we were not considering. As we saw in the early charts, there are large differences in profanity between the fighters but also between the same fighter in different events, which may be explained because events occurred in different time periods, often with different fighters and in different contexts, leading to different behaviors in fighters and this might have attracted fans differently. That said, it makes sense to think that there is another variable, a moderator variable, that can influence the relation between our variables and could help us to better understand the fans' behavior and this relation (Sharma et al., 1981).

The results also showed another interesting point. The correlation that we saw in Figures 12 and 14 is from an analysis of all fighters from all events individually and, as we saw, the analyses of the figures 10 and 11 may suggest that fighter's change their behavior in the context of the fight. Thus, starting from this assumption, we analyzed each fighter individually, by event (Figure 16), to compare the results:

Event / Fighter		Negative SA conferences (%)	Negative SA tweets (%)
UFC182	Daniel Cormier	9.09	21.31
	Jon Jones	7.69	20.05
UFC189	Chad Mendes	22.22	23.65
	Conor McGregor	21.43	15.19
UFC190	Bethe Correia	14.29	25.28
	Ronda Roussey	8.70	17.56
UFC194	Conor McGregor	7.14	13.10
	Jose Aldo	30.00	20.25
UFC196	Conor McGregor	14.71	22.09
	Nate Diaz	37.04	19.27
UFC199	Luke Rockhold	17.65	22.28
	Michael Bisping	16.13	10.53
UFC202	Conor McGregor	60.00	15.98
	Nate Diaz	31.25	16.76
UFC205	Conor McGregor	14.71	13.21
	Eddie Alvarez	15.38	20.78
UFC214	Daniel Cormier	22.22	19.49
	Jon Jones	15.00	14.41
UFC217	George St-Pierre	20.00	15.41
	Michael Bisping	51.72	22.19
UFC229	Conor McGregor	22.22	27.11
	Khabib Nurmagomedov	30.00	32.75
UFC232	Alexander Gustafsson	39.29	28.00
	Jon Jones	16.67	27.13

Figure 16 - Analysis of negative SA variables of each fighter individually, by event.

We observed that in 83.33% of the cases (10/12, highlighted in yellow) the fighter who had a higher negative SA in the conference is generated more SA negative in the tweets of his fans which showed a strong cause-effect relation between the variables in this context. Only in two cases (highlighted in red), this did not happen. However, we only compared two fighters per event, i.e., 2 sets with 2 values each, which is a very small sample. We also did the same analysis but with the average profanity instead (Figure 17) and the results are quite different:

Event / Fighter		Average profanity conferences	Average profanity tweets
UFC182	Daniel Cormier	5.06	11.96
	Jon Jones	10.70	16.39
UFC189	Chad Mendes	29.49	12.50
	Conor McGregor	16.12	13.71
UFC190	Bethe Correia	7.38	18.62
	Ronda Roussey	16.90	17.16
UFC194	Conor McGregor	4.02	14.49
	Jose Aldo	8.49	11.94
UFC196	Conor McGregor	18.69	14.93
	Nate Diaz	33.15	20.68
UFC199	Luke Rockhold	23.18	13.51
	Michael Bisping	34.33	12.94
UFC202	Conor McGregor	63.46	14.38
	Nate Diaz	32.76	18.27
UFC205	Conor McGregor	23.48	14.37
	Eddie Alvarez	15.38	13.77
UFC214	Daniel Cormier	13.22	13.98
	Jon Jones	15.73	14.14
UFC217	George St-Pierre	11.16	14.89
	Michael Bisping	51.07	15.15
UFC229	Conor McGregor	30.31	19.36
	Khabib Nurmagomedov	14.41	22.82
UFC232	Alexander Gustafsson	25.70	15.74
	Jon Jones	22.36	16.28

Figure 17 - Analysis of average profanity variables of each fighter individually, by event.

The results only unveiled 41.66% of cases (5/12, highlighted in yellow) in which the fighter who had a higher profanity in the conference was the one that generated more negative sentiment in the tweets of his fans, which corresponds to half of the relation found previously. Therefore, although the negative sentiment shown by the fighters can influence the negative sentiment in the tweets of the fans, regarding to profanity, this influence is smaller, as only in less than half of the cases the fighter who had a higher profanity in his answers generated more profanity in the tweets of his fans.

4.2 Second hypothesis

To assess the second hypothesis "Fights between fighters with a greater tendency to provoke their opponents before the fight (Trash-Talk), can create a greater engagement of the fans", we measured the engagement of the fans through the pay-per-view numbers and numbers of tweets, that each event generated. Table 6 describe the SPSS data that we used to study the second hypothesis:

EventNumber	Event	NegSAConf	AvgProfanityConf	AvgPPV	TotalTweets
1	UFC232	26.56	24.03	675,000	1,309
2	UFC229	25.00	22.36	2,233,333	6,600
3	UFC217	43.59	31.12	875,000	1,483
4	UFC214	18.24	14.48	855,000	1,333
5	UFC205	15.00	19.43	1,300,000	3,485
6	UFC202	42.31	48.11	1,641,667	3,794
7	UFC199	16.92	1.21	310,750	1,001
8	UFC196	24.59	25.92	1,394,667	3,707
9	UFC194	16.67	6.25	1,112,500	2,959
10	UFC190	10.00	12.14	900,000	4,082
11	UFC189	21.82	22.81	825,000	1,432
12	UFC182	8.33	7.88	620,000	1,175

Table 6 - SPPS data set used for the second hypothesis.

In order to study this hypothesis, we computed the Pearson correlation (Figure 18) between both the negative sentiment and profanity in conferences, the average PPV, and the total number of tweets and we drew their respective dispersion chart as well:

		Negative SA conferences (%)	Average profanity conferences	Average pay per view	Total number of tweets
Negative SA conferences (%)	Pearson Correlation	1	0.844**	0.332	0.021
	Sig. (2-tailed)		0.001	0.292	0.949
	N	12	12	12	12
Average profanity conferences	Pearson Correlation	0.844**	1	0.525	0.247
	Sig. (2-tailed)	0.001		0.080	0.439
	N	12	12	12	12
Average pay per view	Pearson Correlation	0.332	0.525	1	0.896**
	Sig. (2-tailed)	0.292	0.080		0.000
	N	12	12	12	12
Total number of tweets	Pearson Correlation	0.021	0.247	0.896**	1
	Sig. (2-tailed)	0.949	0.439	0.000	
	N	12	12	12	12

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 18 - Correlation between negative SA conferences, average PPV and tweets.

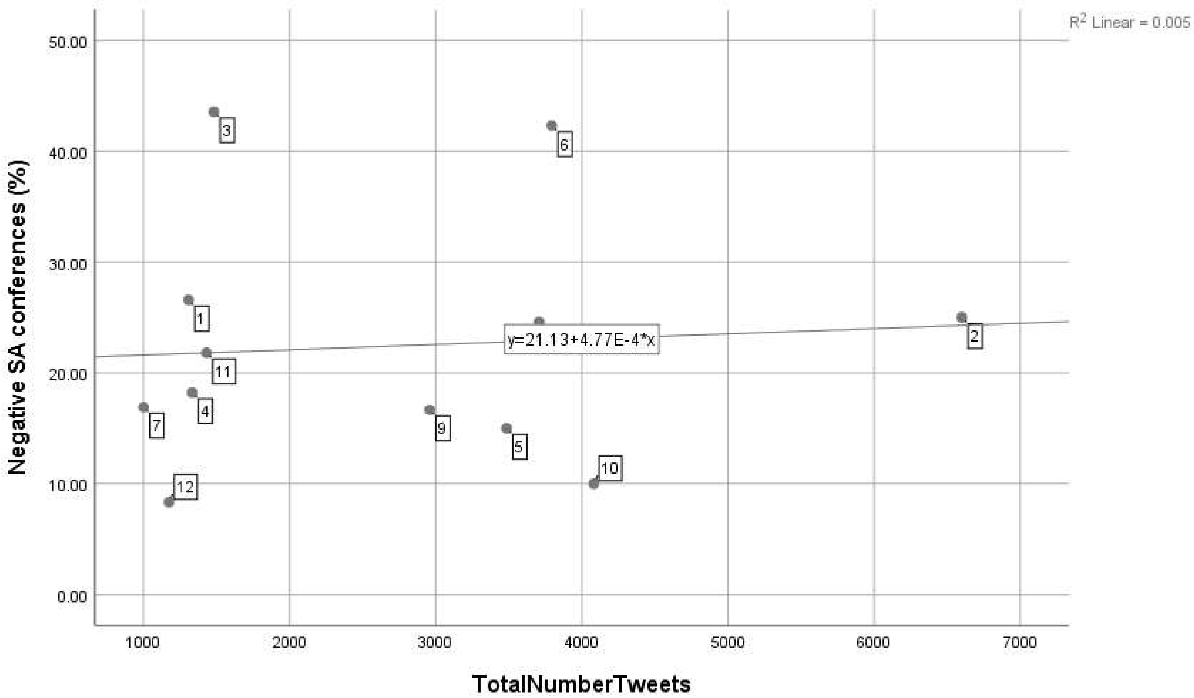

Figure 19 - Dispersion chart to relate negative sentiment with total number of tweets.

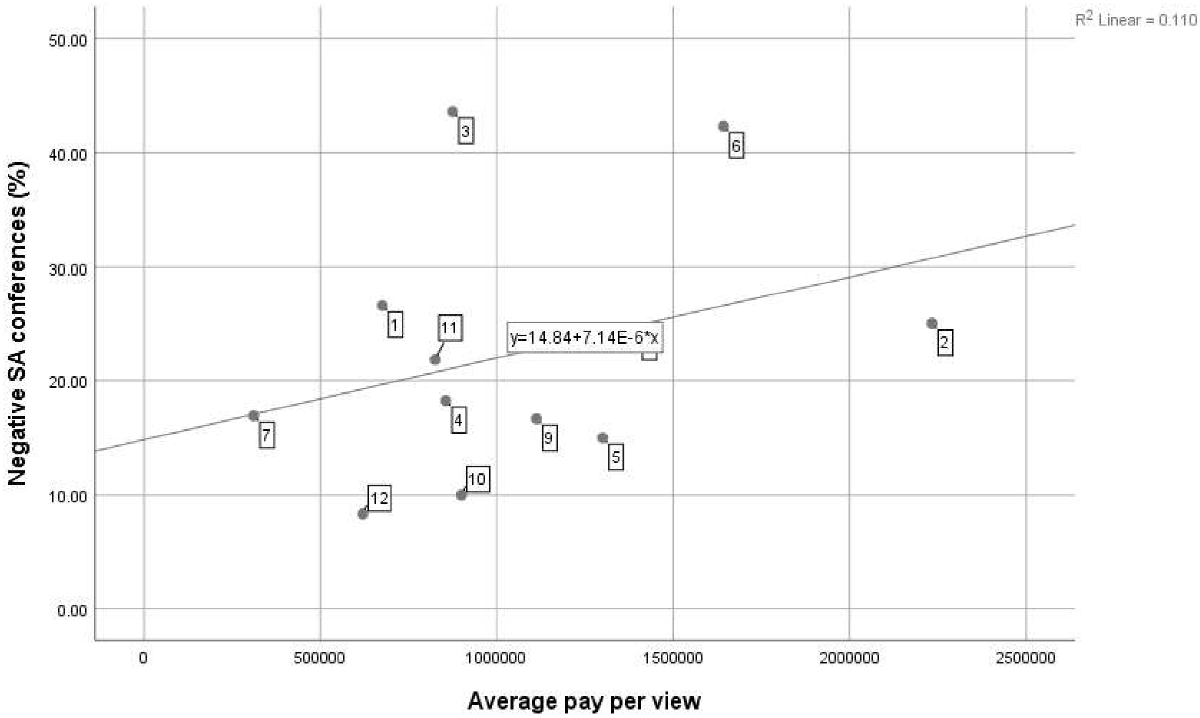

Figure 20 - Dispersion chart to relate negative sentiment with average PPV.

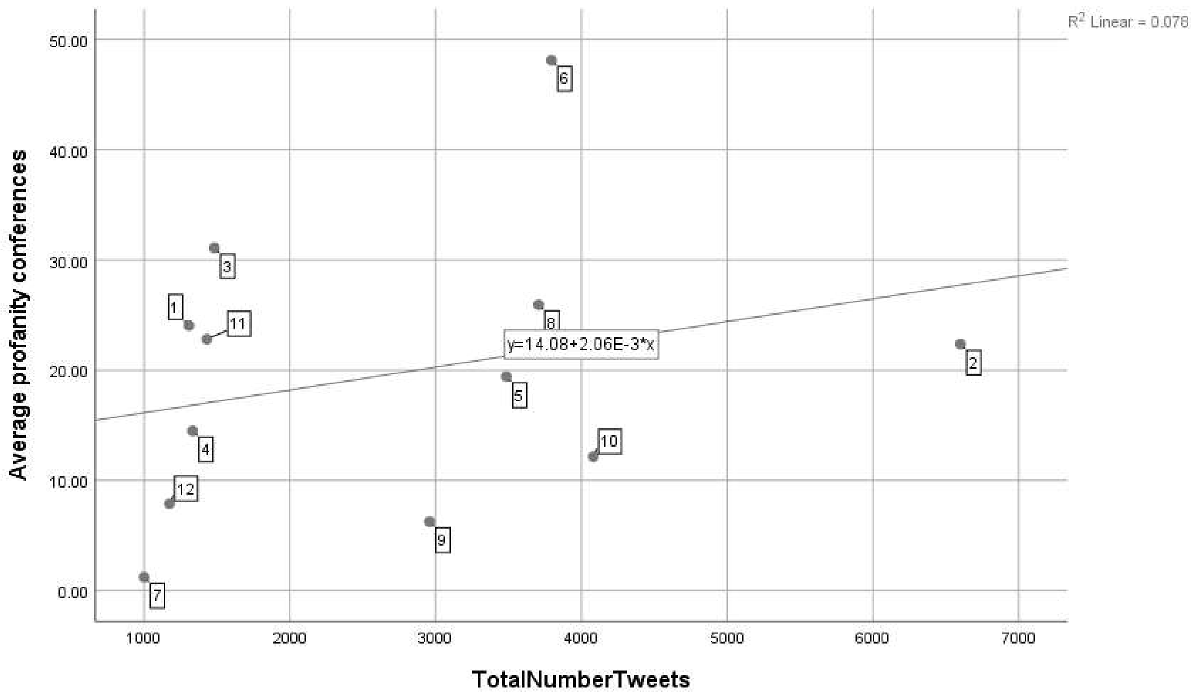

Figure 21 - Dispersion chart to relate average profanity with total number of tweets.

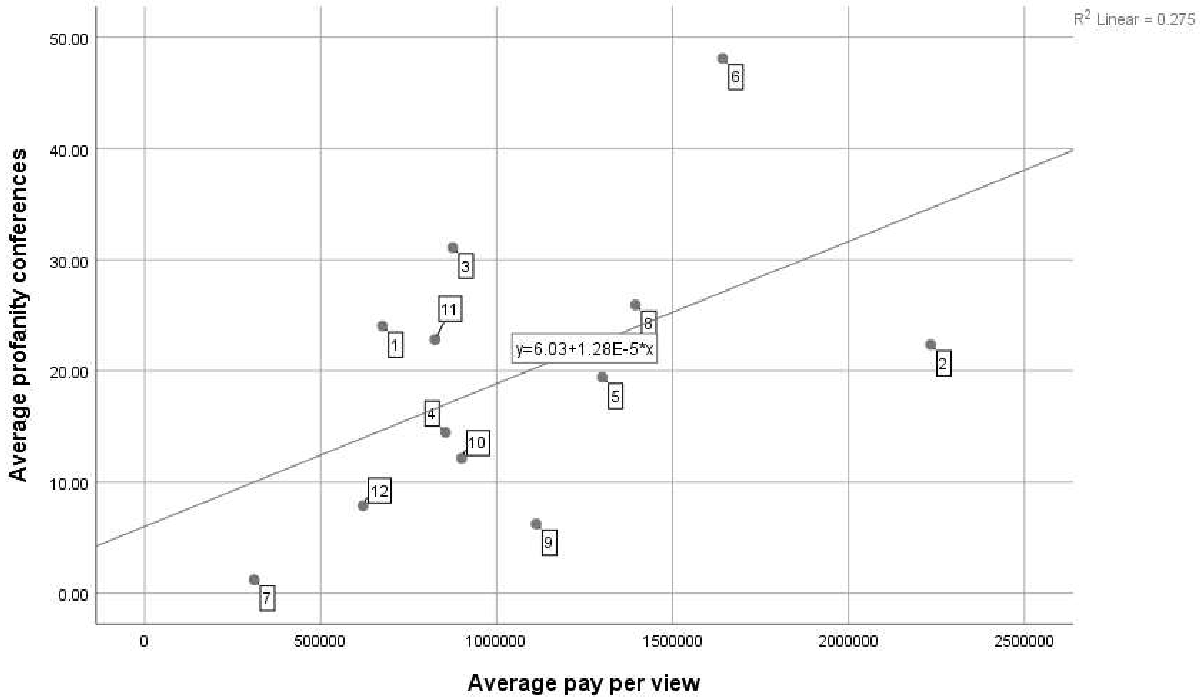

Figure 22 - Dispersion chart to relate average profanity with average PPV.

No significant linear relations were found either between the negative sentiment in the conferences and average PPV (R=0.332, p-value=0.292), the negative sentiment in the conferences and total number tweets (R=0.021, p-value=0.949) or between the average profanity in the conferences and total number tweets (R=0.247, p-value=0.439). However, the analysis between the average profanity in the conferences and average PPV showed a 10% level statistically significant correlation (R=0.525, p-value=0.080), which suggest that higher average profanity values are also associated with higher PPV values. These results seems to indicate that trash-talk in the form of profanity speech does have an influence to some degree in fan engagement. This is an interesting contribution to existing body of knowledge by discerning the impact of using solely negative speech and using profanity speech. We also saw a very strong correlation, 89.6%, between the total number of tweets and the average PPV. Once again, the analysis we conducted above considered only the tweets that refer to one of the two main event fighters. Thus, we did correlate (Figure 23) with all generated tweets for each event and found that both correlations involving the total number of tweets significantly increase, this may suggest that it might be interesting to analyze everything about an event and not just the main event fight.

To complement the abovementioned analyses, we also conducted an analysis by fighter (Table 7) to understand the influence of each one in the engagement of his/her fans and divided in: number of fights (1), number of PPV (2), average PPV (3), % of negative sentiment in the conferences (4), % in the total PPV (5), % in total number of tweets (6) and % in total number of fights (7), % in total number of negative sentiment in the conferences (8).

		Negative SA conferences (%)	Average profanity conferences	Average pay per view	Total number of tweets
Negative SA conferences (%)	Pearson Correlation	1	0.844**	0.332	0.162
	Sig. (2-tailed)		0.001	0.292	0.615
	N	12	12	12	12
Average profanity conferences	Pearson Correlation	0.844**	1	0.525	0.345
	Sig. (2-tailed)	0.001		0.080	0.273
	N	12	12	12	12
Average pay per view	Pearson Correlation	0.332	0.525	1	0.958**
	Sig. (2-tailed)	0.292	0.080		0.000
	N	12	12	12	12
Total number of tweets	Pearson Correlation	0.162	0.345	0.958**	1
	Sig. (2-tailed)	0.615	0.273	0.000	
	N	12	12	12	12

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 23 - Correlation between negative SA conferences, average PPV and all tweets.

Fighter Name	#Fights (1)	#PPV (2)	AvgPPV (3)	%NegSA (4)	%TotalPPV (5)	%TotalTweets (6)	%TotalFights (7)	%TotalNegSA (8)
Conor McGregor	6	8,507,167	1,417,861	23.37%	33.38%	50.25%	50.00%	45.76%
Nate Diaz	2	3,036,333	1,518,167	34.14%	11.91%	4.61%	16.67%	4.12%
Khabib Nurmagomedov	1	2,233,333	2,233,333	30.00%	8.76%	10.90%	8.33%	17.90%
Jon Jones	3	2,150,000	716,667	13.12%	8.44%	9.43%	25.00%	9.92%
Daniel Cormier	2	1,475,750	592,875	15.66%	4.65%	2.06%	16.67%	2.11%
Eddie Alvarez	1	1,300,000	1,300,000	15.38%	5.10%	0.48%	8.33%	0.50%
Michael Bisping	2	1,185,750	450,000	33.93%	3.53%	1.97%	16.67%	2.37%
Jose Aldo	1	1,112,500	1,112,500	30.00%	4.37%	1.22%	8.33%	1.24%
Ronda Rousey	1	900,000	900,000	8.70%	3.53%	11.25%	8.33%	9.91%
Bethe Correia	1	900,000	900,000	14.29%	3.53%	1.37%	8.33%	1.74%
George St. Pierre	1	875,000	875,000	20.00%	3.43%	3.65%	8.33%	2.82%
Chad Mendes	1	825,000	825,000	21.43%	3.24%	0.46%	8.33%	0.54%
Alexander Gustafsson	1	675,000	675,000	39.29%	2.65%	0.11%	8.33%	0.43%
Luke Rockhold	1	310,750	310,750	17.65%	1.22%	0.57%	8.33%	0.64%

Table 7 - Analysis of the influence of each fighter on the engagement of his fans.

We ordered the information by the fighters with the highest % of PPV generated and some interesting results emerged. The first four fighters (underlined in blue) appear 12 times out of a total of 24, spread through nine of the 12 fights. The PPV generated by these four fighters represents 62.49% of the total PPV. These fighters are also responsible for 75.19% of the total number of tweets and for 77.70% of the number of tweets with negative sentiment. Three of these fighters not only generate more PPV, but also more tweets and more tweets with negative sentiment than the remaining fighters.

Some of these fights, as in the case of both McGregor-Diaz fights, are not even title fights, contrary to the common sense that title fights would attract more fans and generate more PPV. These two fights are the 2nd and 3rd fights generating more PPV ever in the UFC, only surpassed by the fight between McGregor-Khabib. McGregor and Diaz have a bad relationship and whenever the two are together at conferences there is a lot of trash-talk and a big rivalry up to the point of, for example, in the end of the UFC202 press conference, after engaging in profanity exchanges, both began throwing objects at each other (MMAFightingonSBN, 2016, Segura, 2016).

Aside from Khabib being the actual champion (as of September 2019) and an undefeated UFC fighter, winning his 12 fights in the UFC, there was a lot of trash-talk and drama around the McGregor fight, which caused a huge brawl at the end of the fight (Guardian sport, 2019; Global News, 2018). Jon Jones and Cormier also have a big rivalry between each other and, consequently, a bad relationship, which also causes a lot of drama and trash-talk between them (the first time both fighters met, at the UFC178, in one of the pre-fight press conferences, both were involved in physical confrontations - MMAWeekly, 2014; MMA Fighting, 2014). Two of Jon Jones's fights are with Cormier.

As a comparison, we looked for two other top-level fighters, Tj. Dillashaw and Demetrious Johnson, who were champions in their categories. During the period we analyzed, 2014-2019, both fighters fought nine times each against different opponents and both only lost one time. They performed, and won, 3 and 4 main event fights, respectively, during this period, with all of the events being title fights. However, after we analyzed their engagement numbers, the results did not quite match the outstanding sports results they achieved. Tj. Dillashaw generated 640,000 in PPV buys in all three fights and Demetrious Johnson generated 560,000 in PPV buys in all four fights. This may suggest that fans prefer drama between fighters and emphasizes one of the results of Andrew et al. (2009; p. 207) who stated that "… the fact that consumers rated drama so highly in each study indicates that people desire close fights with uncertain outcomes regardless of the level of competition".

Our findings also corroborate a finding by Brown et al. (2013, p. 29), who stated that “UFC should begin to more heavily promote individual fighters and storylines that may arise (rivalries, alliances, etc.) with them. This establishes more of an emotional, individual connection to the fighters, allowing the UFC to expand ratings”.

5. Conclusions and recommendations

5.1 Conclusions

Fans consume MMA for different reasons. Recent studies have highlighted important motives for fans to engage and support athletes. Our study unveiled that in fight sports events such as MMA, fights with a lot of drama and rivalries between the fighters generate large numbers of PPV compared to fights between highly technical and dominant fighters, even if they are title fights. For example, the two non-title fights between McGregor and Diaz generated 2.5 times more PPV than the seven title fights combined where champions Tj. Dillashaw and Demetrious Johnson participated. Therefore, we sought to unfold the influence that trash-talk between fighters in the conferences has on fans' engagement and support of athletes. That said, our study focused on the influence that fighters have on their fans. More specifically, in the influence that the negative sentiment and profanity expressed by the fighters in the conferences has on the engagement of the fans and in the way in which they express their support.

Regarding the first hypothesis, "fans demonstrate more aggressiveness in the way they express their support in fights between fighters with a higher tendency to provoke their opponents before the fight (trash-talk)", our results don't indicate a correlation between each fighter's negative expressions used during the conference and the fans' tweets. Rather, we found that the context of the event in itself plays a more significant role as the rivalry between fighters is expressed during the press conference and, in this context, it is very likely that fans react with negative sentiment if fighters show negative sentiment, but the use or not of profanity does not add much to this effect.

The second hypothesis assessed if "events (fights) between fighters with a higher tendency to provoke their opponents before the fight (trash-talk) might create a greater engagement of the fans". Our contribution to existing literature regarding this hypothesis can be summarized in two folds. First, we discovered a moderate association between profanity usage in the conference press by both fighters and the average PPV, and a weak association between profanity usage and the number of tweets posted by fans. Second, we unveiled only a weak association between negative expressions used in press conferences and PPV and, regarding the number of tweets written by fans, the correlation is almost null. This implies that, while profanity is connected to negative sentiments, it's even more connected to user engagement. Athletes' managers can use such knowledge to instruct them to use profanity instead of just negatively connotated words, which results in a higher engagement and subsequent revenue.

5.2 Limitations

The inexistence of pre-fight press conference or the occurrence of the press conference more than one week before the event represents an important limitation of this study, since it left us with a small set of events. The Twitter API also posed the challenge of only allowing data extraction up to a week from current date (considering the 2014-2019 timeframe and that data was collected during the first semester of 2019).

Another important constraint was collecting the PPV numbers for each event, because no single official source was available. We needed to collect such data from several sources and average the results to mitigate eventual inaccurate information. Finally, our findings suggest that current theory on trash-talk propagation from athletes to fans is missing important moderation variables that still remain undisclosed. Therefore, the refutation of both hypothesis may denote that other variables need to be accounted for the influence in the sentiments expressed by fans and their engagement.

5.3 Future work

The results of both hypotheses only analyzed sentences that refer to the two main event fighters and, if we relate all generated tweets for each event, both correlations involving the total number of tweets significantly increase. Therefore, as a suggestion for future work, it might be interesting to look at the conference of all the fighters of that event (rather than just the two of the main event) and see the reaction of the fans, especially to assess if the prominence of the fighter moderates trash-talk propagation and, specifically, profanity, from athletes to fans. This does not mean that all correlations will increase, but at least we will have more fighters, answers and tweets to compare by event.

Additionally, the theoretical model between trash-talk and engagement could be further developed if other variables are incorporated. Such variables can be borrowed from celebrity studies and include the exposition level of fighters on the distinct social media platforms and on mass media, especially considering Robbins and Zemanek (2017) classified fighters as celebrities.

References

Andreasson, J., & Johansson, T. (2019). Negotiating violence: Mixed martial arts as a spectacle and sport. *Sport in Society, 22(7)*, 1183-1197.

Andrew, D. P., Kim, S., O'Neal, N., Greenwell, T. C., & James, J. D. (2009). The Relationship Between Spectator Motivations and Media and Merchandise Consumption at a Professional Mixed Martial Arts Event. *Sport Marketing Quarterly, 18(4)*, 199-209.

BEARAK, B. (2011). *Ultimate fighting dips a toe into the mainstream*. Retrieved from https://www.nytimes.com/2011/11/12/sports/ultimate-fighting-championship-comes-of-age-financially.html

Brown, N. A., Devlin, M. B., & Billings, A. (2013). Fan Identification Gone Extreme: Sports Communication Variables Between Fans and Sport in the Ultimate Fighting Championship. *International Journal of Sport Communication, 6*(1), 19-32.

Chen, M. A., & Cheesman, D. J. (2013). Mental toughness of mixed martial arts athletes at different levels of competition. *Perceptual & Motor Skills, 116(3)*, 905-917.

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record, 88*, 2-9.

Costa, A., Guerreiro, J., Moro, S., & Henriques, R. (2019). Unfolding the characteristics of incentivized online reviews. *Journal of Retailing and Consumer Services, 47*, 272-281.

Cunningham, N., & Bright, L. F. (2012). The Tweet Is in Your Court: Measuring Attitude Towards Athlete Endorsements in Social Media. *International Journal of Integrated Marketing Communications, 4(2)*, 73-87.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook ‘’ Friends :’’ Social Capital and College Students ’ Use of Online Social Network Sites. *Computer-Mediated Communication, 12*, 1143-1168.

Fox, J. (2018, February 16). *All-Time UFC PPV Buyrates*. Retrieved from The sport daily: https://thesportsdaily.com/2018/02/16/all-time-ufc-ppv-sales-data-fox11/

Fraser, B. P., & Brown, W. J. (2002). Media , Celebrities , and Social Influence : Identification With Elvis Presley. *Mass Communication & Society, 5(2)*, 183-206.

Funk, D., Mahony, D., & Ridinger, L. (2002). Characterizing Consumer Motivation as Individual Difference Factors: Augmenting the Sports Interest Inventory (SII) to Explain Level of Spectator Support. *Sport Marketing Quarterly, 11(1)*, 33-43.

Garrett, A. S., & Maddock, R. J. (2001). Time course of the subjective emotional response to aversive pictures : relevance to fMRI studies. *Psychiatry Research: Neuroimaging Section, 108(1)*, 39-48.

Global News. (2018, October 7). *Brawl breaks out at Khabib vs McGregor UFC 229 fight*. Retrieved from Global News: https://globalnews.ca/video/4524653/brawl-breaks-out-at-khabib-vs-mcgregor-ufc-fight

Gratch, J., Lucas, G., Malandrakis, N., Szablowski, E., Fessler, E., & Nichols, J. (2015). GOAALLL!: Using sentiment in the World Cup to explore theories of emotion. *International Conference on Affective Computing and Intelligent Interaction (ACII)*, (pp. 898-903).

Guardian sport. (2019, January 29). *'Politics forever': McGregor fined $50,000, Nurmagomedov $500,000 for UFC 229 brawl*. Retrieved from The Guardian: https://www.theguardian.com/sport/2019/jan/29/conor-mcgregor-banned-six-months-khabib-nurmagomedov-brawlufc-229

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media : an integrative review. *Current Opinion in Behavioral Sciences, 18*, 43-49.

GypsyGold. (2018, January 11). *Every single PPV listed in order of "Buy-Rate!"*. Retrieved from Reddit: https://www.reddit.com/r/MMA/comments/7pmlgl/every_single_ppv_listed_in_order_of_buyrate/

Hambrick, M. E., & Mahoney, T. Q. (2011). 'It's incredible–trust me': exploring the role of celebrity athletes as marketers in online social networks. *International Journal of Sport Management and Marketing, 10(3-4)*, 161-179.

Hambrick, M. E., Simmons, J. M., Greenhalgh, G. P., & Greenwell, T. (2010). Understanding Professional Athletes’ Use of Twitter: A Content Analysis of Athlete Tweets. *International Journal of Sport Communication, 3*(4), 454-471.

Henrique, J. (2016, April 15). Retrieved from Github: https://github.com/Jefferson-Henrique/GetOldTweets-java

Hutto, C. J., & Gilbert, E. (2014). VADER : A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *In Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI.*

Jaidka, K., Guntuku, S., & Ungar, L. (2018). Facebook vs . Twitter : Cross-Platform Differences in Self-disclosure and Trait Prediction. *International AAAI Conference on Web and Social Media 2018*.

James, J. D., Ross, S., D. (2004). Motivations across multiple sports. *Sport Marketing Quarterly, 13(1)*, 17-25.

Joinson, A. N. (2008). ‘ Looking at ’, ‘ Looking up ’ or ‘ Keeping up with ’ People ? Motives and Uses of Facebook. *In Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 1027-1036). ACM.

Kassing, J. W., & Sanderson, J. (2010). Fan–Athlete Interaction and Twitter Tweeting through the Giro: A Case Study. *International Journal of Sport Communication, 3*(1), 113-128.

Kassing, J. W., & Sanderson, J. (2015). Playing in the New Media Game or Riding the Virtual Bench : Confirming and Disconfirming Membership in the Community of Sport. *Journal of Sport and Social Issues, 39(1)*, 3-18.

Kim, J., & Lee, J. E. R. (2011). The Facebook Paths to Happiness : Effects of the Number of Facebook Friends and Self-Presentation. *Cyberpsychology, Behavior, and Social Networking, 14(6)*, 359-364.

Kniffin, K. M., & Palacio, D. (2018). Trash-Talking and Trolling. *Human Nature, 29(3)*, 353-369.

Lin, H., Jia, J., Guo, Q., Xue, Y., Li, Q., Huang, J., ... & Feng, L. (2014). User-Level Psychological Stress Detection from Social Media Using Deep Neural Network. *In Proceedings of the 22nd ACM international conference on Multimedia* (pp. 507-516). ACM.

Massey, W. V., Meyer, B. B., & Naylor, A. (2013). Toward a grounded theory of self-regulation in mixed martial arts. *Psychology of Sport & Exercise, 14*(1), 12-20.

Mazique, B. (2018, October 5). *UFC 229 Card: Conor McGregor Vs. Khabib Nurmagomedov Odds, Predictions And DraftKings Picks*. Retrieved from Forbes: https://www.forbes.com/sites/brianmazique/2018/10/05/ufc-229-card-conor-mcgregor-vs-khabib-nurmagomedov-odds-predictions-and-draftkings-picks/#bffb97a892e5

Miller, M. (2008). *Ultimate Cash Machine*. Retrieved from https://www.forbes.com/forbes/2008/0505/080.html#1ce5ecc03e9e

MMA Fighting. (2014, August 4). *Jon Jones, Daniel Cormier have wild brawl at UFC 178 media day*. Retrieved from MMA Fighting: https://www.mmafighting.com/2014/8/4/5968683/jon-jones-daniel-cormier-have-wild-brawl-at-ufc-178-media-day

MMAFightingonSBN. (2016, August 17). *Chaos Breaks Out at UFC 202 Press Conference*. Retrieved from Youtube: https://www.youtube.com/watch?v=PFBVhdEDjrg

MMAWeekly. (2014, August 4). *Jon Jones and Daniel Cormier Brawl (Complete Fight)*. Retrieved from Youtube: https://www.youtube.com/watch?v=rf_WWMTR9LU

Pay-Per-View Buys. (n.d.). Retrieved from Tapology: https://www.tapology.com/search/mma-event-figures/ppv-pay-per-view-buys-buyrate

Pay-per-view. (n.d.). Retrieved from MMA Payout: http://mmapayout.com/blue-book/pay-per-view/

Pegoraro, A. (2010). Look Who’s Talking—Athletes on Twitter: A Case Study. *International Journal of Sport Communication, 3*(4), 501-514.

Raacke, J., & Bonds-Raacke, J. (2008). MySpace and Facebook : Applying the Uses and Gratifications Theory to Exploring. *Cyberpsychology & Behavior, 11*(2), 169-174.

Ramos, R. F., Rita, P., & Moro, S. (2019). From institutional websites to social media and mobile applications: A usability perspective. *European Research on Management and Business Economics, 25(3)*, 138-143.

Robbins, T., & Zemanek, J. E. . (2017). UFC pay-per-view buys and the value of the celebrity fighter. *Innovative Marketing (hybrid), 13(4)*, 35-46.

Rosenman, D. (2018, January 15). *UFC PPV Sales*. Retrieved from Kaggle: https://www.kaggle.com/daverosenman/ufc-ppv-sales

Salles, S., Affonso, C., & Rocco Jr, A. J. (2013). O comportamento do consumidor de atletas de alto rendimento nas mídias sociais: uma análise do caso Anderson Silva e seus fãs. *Revista Intercontinental de Gestão Desportiva, 3*.

Sap, M., Park, G., Eichstaedt, J., Kern, M., Stillwell, D., Kosinski, M., ... & Schwartz, H. A. (2014). Developing Age and Gender Predictive Lexica over Social Media. *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (pp. 1146-1151).

Schumacher-Dimech, A. M., Brechbühl, A., Kohlbrenn. (2012). An exploratory study about the perception and justification of violence in Mixed Martial Arts and kickboxing athletes. *The British Psychological Society, 107*, 19-24.

Schumaker, R. P., Jarmoszko, A. T., & Labedz Jr, C. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems, 88*, 76-84.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLoS ONE, 8(9), e73791*.

Segura, D. (2016, August 17). *Pros react to UFC 202 press conference chaos between Conor McGregor and Nate Diaz*. Retrieved from MMA Fighting: https://www.mmafighting.com/2016/8/17/12525188/pros-react-to-ufc-202-press-conference-chaos-between-conor-mcgregor

Seungmo, K., Greenwell, T. C., Andrew, D. P., Lee, J., & Mahony, D. (2008). An Analysis of Spectator Motives in an Individual Combat Sport : A Study of Mixed Martial Arts Fans. *Sports Management Commons, 17*, 109-119.

Sharma, S., Durand, R. M., & Gur-Arie, O. (1981). Identification and analysis of moderator variables. *Journal of marketing research, 18(3)*, 291-300.

Tainsky S., Salaga S., & Santos C. A. (2013). Determinants of Pay-Per-View Broadcast Viewership in Sports : The Case of the Ultimate Fighting Championship. *Sport Management, 27(1)*, 43-58.

Twitter. (n.d.). *Developer tools and APIs built for innovation and scale*. Retrieved from https://developer.twitter.com/en/pricing.html

UFC. (2018). *INTRO TO MMA*. Retrieved from https://www.ufc.com/intro-mma

Vaccaro, C. A., Schrock, D. P., & McCabe, J. M. (2011). Managing Emotional Manhood : Fighting and Fostering Fear in Mixed Martial Arts. *Social Psychology Quarterly, 74(4)*, 414-437.

Wann, D. L. (1995). Prelimiary validation of the sport fan motivation scale. *Journal of Sport & Social Issues, 19(4)*, 377-396.

Wann, D. L., Grieve, F. G., Zapalac, R. K., & Pease, D. G. (2008). Motivational Profiles of Sport Fans of Different Sports. *Sport Marketing Quarterly, 17(1)*, 6-19.

Workman, M. (2012). Rash impulsivity, vengefulness, virtual-self and amplification of ethical relativism on cyber-smearing against corporations. *Computers in Human Behavior, 28(1)*, 217-225.

Yip, J. A., Schweitzer, M. E., & Nurmohamed, S. (2018, 1 1). Trash-talking: Competitive incivility motivates rivalry, performance, and unethical behavior. *Organizational Behavior and Human Decision Processes, 144*, 125-144.

Yu, Y., & Wang, X. (2015). World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets. *Computers in Human Behavior, 48*, 392-400.

Zhou, V. (2019, February 4). *Building a Better Profanity Detection Library with scikit-learn*. Retrieved from https://victorzhou.com/blog/better-profanity-detection-with-scikit-learn/.

Zirin, D. (2005). *What's my name, fool?: Sports and resistance in the United States.* Haymarket Books.

Annexes

Annex A

UFC232

querysearch="UFC232" since=2018-12-24 until=2018-12-31

querysearch="#UFC232" since=2018-12-24 until=2018-12-31

querysearch="jon jones" since=2018-12-24 until=2018-12-31

querysearch="#JonJones" since=2018-12-24 until=2018-12-31

querysearch="#Jones" since=2018-12-24 until=2018-12-31

querysearch="@JonnyBones" since=2018-12-24 until=2018-12-31

querysearch="alexander gustafsson" since=2018-12-24 until=2018-12-31

querysearch="@AlexTheMauler" since=2018-12-24 until=2018-12-31

querysearch="#AlexanderGustafsson" since=2018-12-24 until=2018-12-31

querysearch="#Gustafsson" since=2018-12-24 until=2018-12-31

UFC229

querysearch="UFC229" since=2018-10-01 until=2018-10-08

querysearch="#UFC229" since=2018-10-01 until=2018-10-08

querysearch="mcgregor" since=2018-10-01 until=2018-10-08

querysearch="#mcgregor" since=2018-10-01 until=2018-10-08

querysearch="conor mcgregor" since=2018-10-01 until=2018-10-08

querysearch="#ConorMcGregor" since=2018-10-01 until=2018-10-08

querysearch="@TheNotoriousMMA" since=2018-10-01 until=2018-10-08

querysearch="khabib" since=2018-10-01 until=2018-10-08

querysearch="#khabib" since=2018-10-01 until=2018-10-08

querysearch="@TeamKhabib" since=2018-10-01 until=2018-10-08

querysearch="#nurmagomedov" since=2018-10-01 until=2018-10-08

querysearch="nurmagomedov" since=2018-10-01 until=2018-10-08

UFC217

querysearch="UFC217" since=2017-10-30 until=2017-11-06

querysearch="#UFC217" since=2017-10-30 until=2017-11-06

querysearch="michael bisping" since=2017-10-30 until=2017-11-06
querysearch="bisping" since=2017-10-30 until=2017-11-06
querysearch="@bisping" since=2017-10-30 until=2017-11-06
querysearch="george st-pierre" since=2017-10-30 until=2017-11-06
querysearch="@GeorgesStPierre" since=2017-10-30 until=2017-11-06
querysearch="GSP" since=2017-10-30 until=2017-11-06
querysearch="#GSP" since=2017-10-30 until=2017-11-06
querysearch="st-pierre" since=2017-10-30 until=2017-11-06

UFC214
querysearch="UFC214" since=2017-07-24 until=2017-07-31
querysearch="#UFC214" since=2017-07-24 until=2017-07-31
querysearch="jon jones" since=2017-07-24 until=2017-07-31
querysearch="#JonJones" since=2017-07-24 until=2017-07-31
querysearch="#Jones" since=2017-07-24 until=2017-07-31
querysearch="@JonnyBones" since=2017-07-24 until=2017-07-31
querysearch="#DanielCormier" since=2017-07-24 until=2017-07-31
querysearch="#Cormier" since=2017-07-24 until=2017-07-31
querysearch="daniel cormier" since=2017-07-24 until=2017-07-31
querysearch="@dc_mma" since=2017-07-24 until=2017-07-31

UFC 205
querysearch="UFC205" since=2016-11-07 until=2016-11-14
querysearch="#UFC205" since=2016-11-07 until=2016-11-14
querysearch="#mcgregor" since=2016-11-07 until=2016-11-14
querysearch="#ConorMcGregor" since=2016-11-07 until=2016-11-14
querysearch="mcgregor" since=2016-11-07 until=2016-11-14
querysearch="@TheNotoriousMMA" since=2016-11-07 until=2016-11-14
querysearch="conor mcgregor" since=2016-11-07 until=2016-11-14
querysearch="eddie alvarez" since=2016-11-07 until=2016-11-14
querysearch="#EddieAlvarez" since=2016-11-07 until=2016-11-14
querysearch="#Alvarez" since=2016-11-07 until=2016-11-14

querysearch="@Ealvarezfight" since=2016-11-07 until=2016-11-14

UFC203

querysearch="UFC203" since=2016-09-05 until=2016-09-12

querysearch="#UFC203" since=2016-09-05 until=2016-09-12

querysearch="miocic" since=2016-09-05 until=2016-09-12

querysearch="@stipemiocic" since=2016-09-05 until=2016-09-12

querysearch="#Miocic" since=2016-09-05 until=2016-09-12

querysearch="@Alistairovereem" since=2016-09-05 until=2016-09-12

querysearch="overeem" since=2016-09-05 until=2016-09-12

querysearch="#Overeem" since=2016-09-05 until=2016-09-12

UFC202

querysearch="UFC202" since=2016-08-15 until=2016-08-22

querysearch="#UFC202" since=2016-08-15 until=2016-08-22

querysearch="mcgregor" since=2016-08-15 until=2016-08-22

querysearch="conor mcgregor" since=2016-08-15 until=2016-08-22

querysearch="#mcgregor" since=2016-08-15 until=2016-08-22

querysearch="#ConorMcGregor" since=2016-08-15 until=2016-08-22

querysearch="@TheNotoriousMMA" since=2016-08-15 until=2016-08-22

querysearch="nate diaz" since=2016-08-15 until=2016-08-22

querysearch="#NateDiaz" since=2016-08-15 until=2016-08-22

querysearch="@NateDiaz209" since=2016-08-15 until=2016-08-22

querysearch="#Diaz" since=2016-08-15 until=2016-08-22

UFC199

querysearch="UFC199" since=2016-05-31 until=2016-06-06

querysearch="#UFC199" since=2016-05-31 until=2016-06-06

querysearch="#Rockhold" since=2016-05-31 until=2016-06-06

querysearch="luke rockhold" since=2016-05-31 until=2016-06-06

querysearch="rockhold" since=2016-05-31 until=2016-06-06

querysearch="@LukeRockhold" since=2016-05-31 until=2016-06-06

querysearch="michael bisping" since=2016-05-31 until=2016-06-06

querysearch="bisping" since=2016-05-31 until=2016-06-06

querysearch="#Bisping" since=2016-05-31 until=2016-06-06
querysearch="@bisping" since=2016-05-31 until=2016-06-06

UFC196
querysearch="UFC196" since=2016-02-29 until=2016-03-07
querysearch="#UFC196" since=2016-02-29 until=2016-03-07
querysearch="mcgregor" since=2016-02-29 until=2016-03-07
querysearch="conor mcgregor" since=2016-02-29 until=2016-03-07
querysearch="#mcgregor" since=2016-02-29 until=2016-03-07
querysearch="#ConorMcGregor" since=2016-02-29 until=2016-03-07
querysearch="@TheNotoriousMMA" since=2016-02-29 until=2016-03-07
querysearch="nate diaz" since=2016-02-29 until=2016-03-07
querysearch="#NateDiaz" since=2016-02-29 until=2016-03-07
querysearch="#Diaz" since=2016-02-29 until=2016-03-07
querysearch="@NateDiaz209" since=2016-02-29 until=2016-03-07

UFC194
querysearch="UFC194" since=2015-12-07 until=2015-12-14
querysearch="#UFC194" since=2015-12-07 until=2015-12-14
querysearch="mcgregor" since=2015-12-07 until=2015-12-14
querysearch="#mcgregor" since=2015-12-07 until=2015-12-14
querysearch="#ConorMcGregor" since=2015-12-07 until=2015-12-14
querysearch="conor mcgregor" since=2015-12-07 until=2015-12-14
querysearch="@TheNotoriousMMA" since=2015-12-07 until=2015-12-14
querysearch="jose aldo" since=2015-12-07 until=2015-12-14
querysearch="#JoseAldo" since=2015-12-07 until=2015-12-14
querysearch="#Aldo" since=2015-12-07 until=2015-12-14
querysearch="@josealdojunior" since=2015-12-07 until=2015-12-14

UFC190
querysearch="UFC190" since=2015-07-27 until=2015-08-03
querysearch="#UFC190" since=2015-07-27 until=2015-08-03

querysearch="ronda rousey" since=2015-07-27 until=2015-08-03
querysearch="#Ronda" since=2015-07-27 until=2015-08-03
querysearch="@RondaRousey" since=2015-07-27 until=2015-08-03
querysearch="bethe" since=2015-07-27 until=2015-08-03
querysearch="bethe correia" since=2015-07-27 until=2015-08-03
querysearch="#bethe" since=2015-07-27 until=2015-08-03
querysearch="@bethecorreia" since=2015-07-27 until=2015-08-03

UFC189
querysearch="UFC189" since=2015-07-06 until=2015-07-13
querysearch="#UFC189" since=2015-07-06 until=2015-07-13
querysearch="mcgregor" since=2015-07-06 until=2015-07-13
querysearch="#mcgregor" since=2015-07-06 until=2015-07-13
querysearch="#ConorMcGregor" since=2015-07-06 until=2015-07-13
querysearch="conor mcgregor" since=2015-07-06 until=2015-07-13
querysearch="@TheNotoriousMMA" since=2015-07-06 until=2015-07-13
querysearch="chad mendes" since=2015-07-06 until=2015-07-13
querysearch="@chadmendes " since=2015-07-06 until=2015-07-13
querysearch="#ChadMendes" since=2015-07-06 until=2015-07-13

UFC182
querysearch="UFC182" since=2014-12-29 until=2015-01-05
querysearch="#UFC182" since=2014-12-29 until=2015-01-05
querysearch="jon jones" since=2014-12-29 until=2015-01-05
querysearch="#JonJones" since=2014-12-29 until=2015-01-05
querysearch="#Jones" since=2014-12-29 until=2015-01-05
querysearch="@JonnyBones" since=2014-12-29 until=2015-01-05
querysearch="daniel cormier" since=2014-12-29 until=2015-01-05
querysearch="#DanielCormier" since=2014-12-29 until=2015-01-05
querysearch="#Cormier" since=2014-12-29 until=2015-01-05
querysearch="@dc_mma" since=2014-12-29 until=2015-01-05

Annex B

Tweet	Permanentlink
For all the pricks talking about cowardice and disgusting behaviour #UFC229 pic.twitter.com/49byBjce4W	https://twitter.com/JJJ26317501/status/1048804838526476288
For all the pricks talking about cowardice and disgusting behaviour #UFC229 pic.twitter.com/ISzCYhRfyE	User https://twitter.com/Roofio_ThatsAll/status/1048806473680261120

Figure 24 - Same tweet from different users.

Tweet	Permanentlink
Conor McGregor, 'I Will Give Nate Diaz His Rematch' http://dlvr.it/QmHgVx	https://twitter.com/TMZ/status/1047857140717633536
Conor McGregor, 'I Will Give Nate Diaz His Rematch' http://tmz.me/d13S48D	User https://twitter.com/TMZ/status/1047943194044973057
Conor McGregor, 'I Will Give Nate Diaz His Rematch' http://tmz.me/SlXI2Hr	https://twitter.com/TMZ/status/1048139486809530369

Figure 25 - Same tweets from the same user in different times.

Annex C

Information about PPV numbers

Duarte Tereso <duarte.s.tereso@gmail.com>
to Support

11 Apr 2019, 22:35

Hello,

My name is Duarte Tereso, I'm sending this email to ask for your help.

I'm finalising my dissertation, which is related to Mixed Martial Arts and the influence of fighters on their fans. I require the pay-per-view numbers of 8 events (UFC232, UFC229, UFC205, UFC202, UFC196, UFC194, UFC193, UFC190).

I've tried searching the Internet but no website or forum that I found can be considered a trusted source. I've also searched your oficial website but also haven't found anything. Is it possible for you to send me this information ?

If there is another email I should try, please let me know and sorry to bother you. This information would be really important for my work.

I thank you for your time.

Best Regards,
Duarte Tereso

Information about PPV numbers

Duarte Tereso <duarte.s.tereso@gmail.com>
to info

Fri, 30 Aug, 16:10

Hello,

My name is Duarte Tereso, i'm a Master student at ISCTE and i'm sending this email to ask for your help.

I'm finalising my dissertation, which is related to Mixed Martial Arts and the influence of fighters on their fans and i need the pay-per-view numbers from some UFC events to support my study.

I've tried searching the Internet but no website or forum that I found can be considered a trusted source. I've also searched your oficial website but also haven't found anything.

I need the pay-per-view numbers from the events: UFC232, UFC229, UFC217, UFC214 UFC205, UFC202, UFC199, UFC196, UFC194, UFC190, UFC189, UFC182. Is it possible for you to send me this information ?

If there is another email I should try, please let me know and sorry to bother you. This information would be really important for my work.

I thank you for your time.

Best Regards,
Duarte Tereso

Figure 26 - Emails sent to UFC requesting PPV information.