

# BUSI4370 - Analytics Specialization and Applications

Coursework 2  
Student ID: 20492770

## EXECUTIVE SUMMARY

This report analyses Twitter data related to the FA Cup, one of the most prestigious soccer competitions in the world. The data was collected over a period of one week, the last week of February, and includes over 2500 tweets. The report begins with an overview of the data collection process and an approach summary of the processes undertaken to extract insights from the data. The report then analyses the top hashtags, mentions, and keywords associated with FA Cup on Twitter. The report also includes a sentiment analysis, which measures the positive or negative sentiment expressed in tweets related to the FA Cup. The sentiment analysis shows that the majority of tweets are neutral and positive, with fans expressing excitement and anticipation for the upcoming matches. It also uses topic modelling to get the prevalent topics in the tweets.

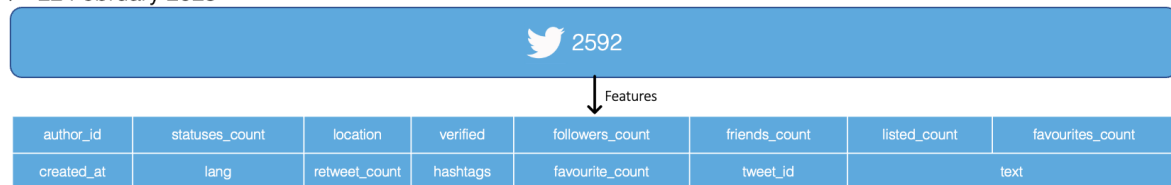
Using the different analyses of tweets, the report suggests a Twitter user as a micro-influencer who could help increase engagement of FA Cup topics on Twitter. Finally, the report concludes with other recommendations that could be used to increase engagement.

## DATA DESCRIPTION

The Football Association Challenge Cup, also known as the FA Cup, is an annual knockout football competition in men's domestic English football. It is organized by the Football Association. The huge community of fans and sponsors discuss about the competition on various social platforms including Twitter, Facebook, and Instagram. This report uses tweets to analyze different aspects of engagement on Twitter.

✓ 22 February 2023

1 March 2023 ✓



Data has been extracted as a csv file by querying Twitter API, for a date range and the topic "FA Cup". The query fetched 2592 tweets that have been posted during the last week of February 2023, i.e., between 22 February 2023 and 1 March 2023, both dates inclusive. The twitter API provides a lot of details regarding a tweet of which the following features will be used for the analysis.

Features	Short description
author_id	Unique ID of the user
location	Geo location of user
followers_count	The number of followers the user has
favourites_count	The number of likes the user has got

Features	Short description
created_at	Timestamp of tweet
retweet_count	The number of times the tweet has been retweeted
favourite_count	Number of likes for the tweet

The other features include statuses\_count, verified, friends\_count, listed\_count, lang, hashtags, text, and tweet\_id.

## APPROACH SUMMARY

To establish the public's view on FA cup using twitter data, different types of analysis can be performed on the tweets. This report uses the following types of analysis and preprocessing steps. Once the analytics are performed to explore the brand's engagement on twitter, a micro influencer will be chosen.

## 1. Pre-processing tweets

Before raw twitter data can be used for analysis or machine learning, it must first go through a number of preprocessing procedures. Tokenizing the text into words, stemming, or lemmatizing the words, reducing any leftover noise in the text data, deleting URLs, mentions, hashtags, punctuation, and stop words are some of these stages.

- **Cleaning the text:** The tweet texts are cleaned from URLs, emoticons, and punctuations which will not be useful for analysis. Also stop words that don't carry any meaning for analytics purposes such as "the", "a", "an", "in", "on", "at", are removed from the text.
- **Tokenization:** It is the process of breaking down a text paragraph into smaller chunks such as words or sentences. It also helps in removing unwanted spaces and punctuation marks from the text. It helps in improving the accuracy of text analysis by providing a more accurate representation of the text.
- **Lemmatization:** It is a technique that reduces words to their base or root form by taking into consideration the context of a word and making use of the context to convert the word to its meaningful base form which is known as lemma.

## 2. Temporal analysis

Using timestamp of the tweets we can visualise where the number of tweets have peaked and analyse the causes behind it. It provides details when an influencer or a marketing campaign ad can be published so that it reaches maximum number of users.

## 3. Engagement, prevalence of mentions and key descriptive words

The report also discusses on the engagement of users on the topic, prevalence of mentions and using Natural Language Toolkit (NLTK), extract key descriptive words with respect to FA cup.

## 4. Topic modelling

Topic modelling is an unsupervised natural language processing technique that identifies topics and structures in large text datasets. It clusters similar documents and summarizes large volumes of text data. It's used in document classification, sentiment analysis, and recommendation systems. This report uses Bertopic, machine learning algorithm that combines the power of BERT language modelling with topic modelling techniques. It is an extension of the popular topic modelling algorithm called "Latent Dirichlet Allocation" (LDA), but it uses BERT embeddings to represent the text data and cluster them into topics.

## 5. Sentiment analysis

Sentiment analysis is a natural language processing technique that identifies subjective information in text as positive, negative, or neutral. This report utilizes the Roberta model (Refer link provided in notebook attached for details.) to analyse tweet sentiment.

## 6. Choosing a micro influencer

Finally, using all the different analysis performed mentioned above and studying users with 10000 to 100000 followers, a micro influencer will be chosen based on how frequent and involved the user is regarding the topic, how positive are the tweets and how many likes the user gets on an average.

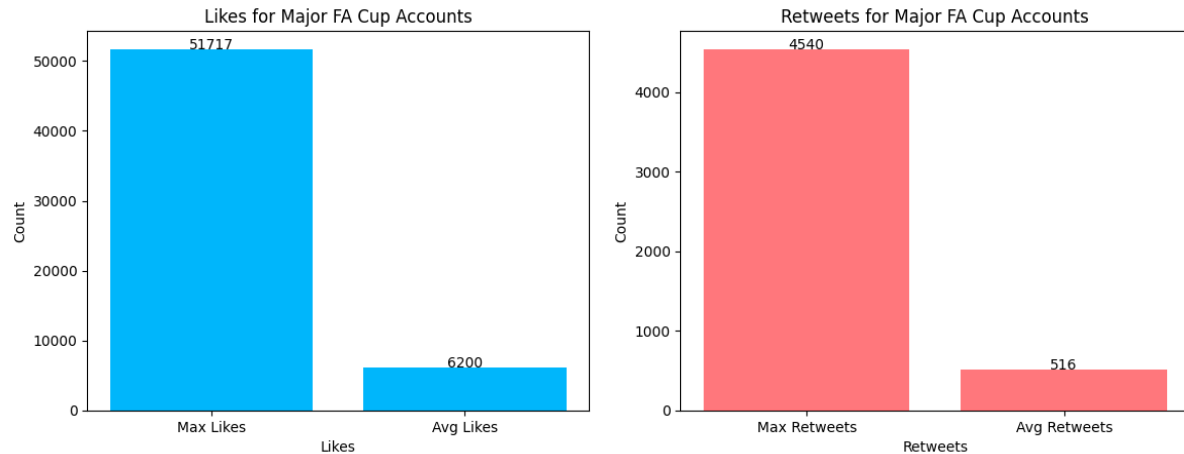
# BRAND EXPLORATION

---

In this exploration, we will delve into the world of the FA Cup on Twitter and examine how users engage with the tournament, share their opinions, and discuss the latest news and updates. We will also analyze the impact of the tournament on social media.

## 1. Engagement

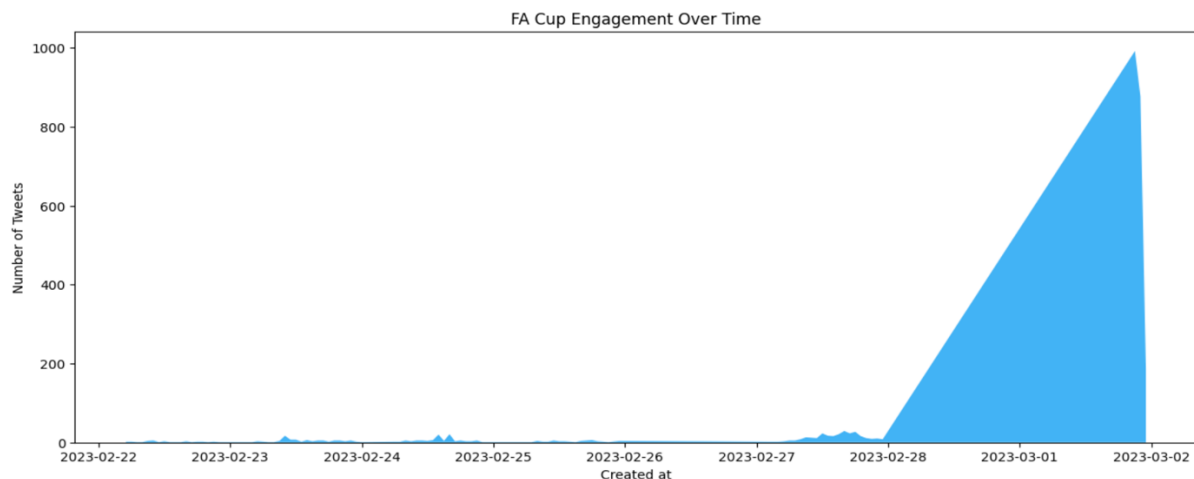
Based on the data collected, it was found that 1974 unique users tweeted about the FA Cup during a week in which the games were held mid-week. Of these users, 17 accounts had over 100,000 followers, and each tweet from these accounts received a minimum of 100 likes. While the number of users directly tweeting about the tournament was limited, most users engaged with FA Cup-related content by liking and retweeting tweets from these major accounts. The chart below summarizes the maximum and average like and retweet counts received by these accounts.



Further engagement with respect to time and sentiment of the users around the tweet are reported in their respective sections.

## 2. Time series analysis

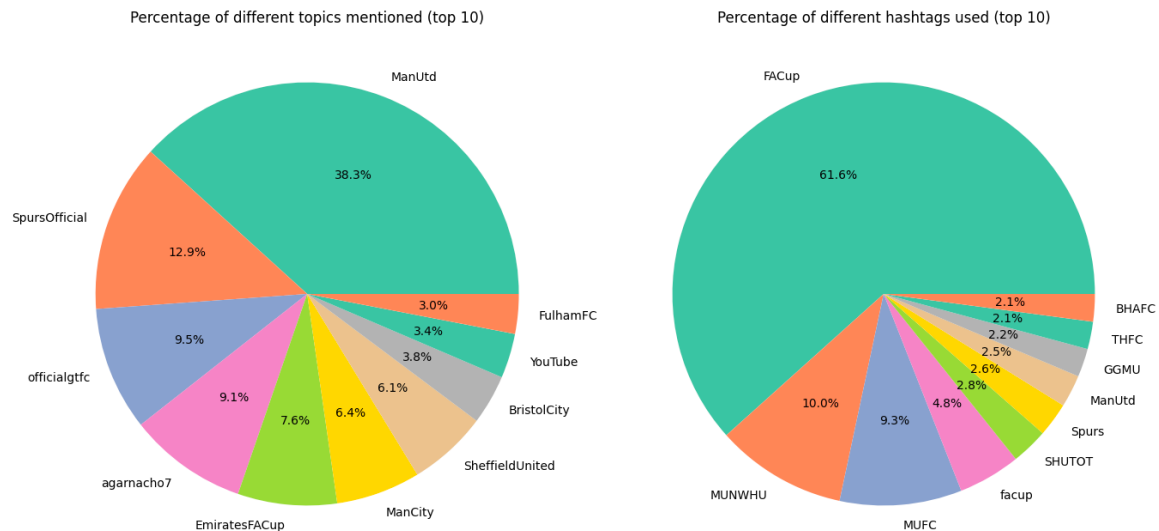
A basic analysis of number of tweets talking about FA cup over the timeline can give us information as to when the users mostly engage in sharing their opinions and it can also provide insights in choosing a micro influencer who remains active consistently during those periods of time.



As shown in the above chart, people mostly engage in tweeting and liking the posts during the match day. FA cup matches are held mid-week and there is a clear spike in the number of tweets on a Wednesday i.e., 1<sup>st</sup> March 2023.

## 3. Prevalence of mentions and hashtags

Through an analysis of mentions and hashtags related to the FA Cup on Twitter, valuable insights can be obtained on popular accounts and keywords associated with this topic. Such information can be useful for brands, media outlets, and other entities looking to engage with the FA Cup audience on Twitter.

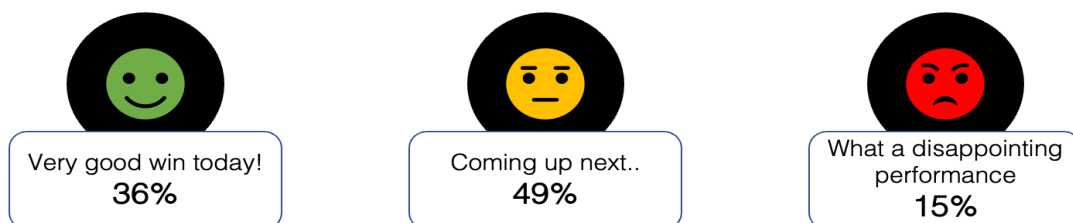


Two charts were created to showcase the most mentioned topics and hashtags associated with the FA Cup. The results indicated that Manchester United and Tottenham Hotspurs were the most mentioned clubs, with mentions of Manchester United players also featuring prominently. Furthermore, the hashtags used in these mentions revealed that matches involving Manchester United and Tottenham Hotspurs were most frequently mentioned, with popular hashtags such as #MUNWHU, #MUFC, and #Spurs.

This analysis helps us choose a micro influencer who engages in tweeting mostly related to these topics. Or advice the influencer to promote the same.

#### 4. Sentiment analysis

According to the sentiment analysis conducted on tweets related to FA Cup, they can be categorised into positive, negative, and neutral emotions. This analysis provides insights into the major sentiments around the topic and the reasons behind them, which can be useful for influencers, brands, and media outlets to tailor their content strategy.



The above image shows the percentage of tweets classified under different sentiments after running them through the sentiment analysis model. Analysis of the tweets with respect to these sentiments reveals that neutral tweets are mostly news updates and it is the most prevalent emotion. On the other hand, positive and negative tweets are mostly related to the results of a match held. Some negative tweets also discuss the system of choosing who plays against whom.

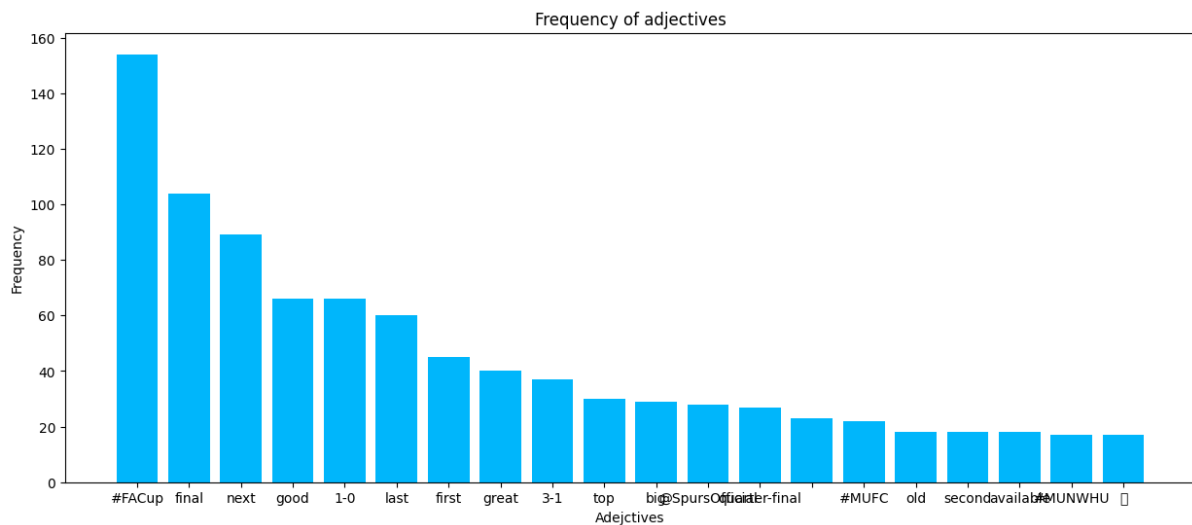
Based on these insights, an influencer who mostly updates the news and results surrounding FA Cup is likely to get more reach.

#### 5. Key descriptive words

Key descriptive words used regarding the FA Cup on Twitter include the names of the participating teams, the match results, the performance of individual players, the stage of the tournament, and the

overall excitement and anticipation surrounding the event. Other descriptive words may also depend on the context in which the FA Cup is being discussed, such as the role of social media influencers.

The below chart provides a list of the most common descriptive words and how frequent they are mentioned.



As we can see, the words such 'final' and 'quarter-final' represent that users mostly are excited to follow matches from quarter-final stage. Words like '1-0' and '3-0' discuss about the results of a match. There are mentions of MUFC and MUNWHU talking about the teams involved in a match. This helps us choose an influencer who is really active during final stages of the tournament, and posts scores of different games going on. Also tweets that invite people to predict final contestants can be very engaging.

## 6. Topic modelling

The output of topic modelling algorithm is a list of topics, each represented by a set of keywords or phrases that are most strongly associated with that topic. These topics can then be analysed to gain insights into the most common themes and conversations surrounding the FA Cup on Twitter.

After running the tweets collected through the algorithm, the following were the top 5 topics in the tweets and also an explanation of the theme of the topic.


Topic	Theme
<b>fulham_home_trafford_mufc</b>	Suggests a conversation related to Manchester United playing at home against Fulham, with the game taking place at Old Trafford
<b>grimsby_southampton_town_grimsbytown</b>	Indicates a conversation related to a match between Grimsby Town and Southampton, with Grimsby Town being the home team
<b>erik_europaleague_hag_europa</b>	Conversation related to the Europa League, a European soccer competition. This topic likely includes discussions of the teams, players, and matches taking place in the Europa League
<b>garnacho_alejandro_goal_munwhu</b>	Indicates a conversation related to a goal scored by Alejandro Garnacho, likely during a match between Manchester United and West Ham United
<b>munwhu_weghorst_muftc_facup</b>	Conversation related to a match between Manchester United and Wolfsburg in the FA Cup

## CHOOSING A MICRO INFLUENCER

Based on the information gathered on the engagement of the topic FA cup on twitter, we can now choose a micro influencer that has the right audience, content, and engagement rate for promoting events related to FA cup on Twitter. A micro influencer is an individual with a social media following of between 1,000 to 100,000 people. Also, when choosing a micro-influencer, it's important to consider not only their popularity within their network, but also their ability to connect different groups or individuals, known as their betweenness centrality.

Based on the time-series analysis conducted, it is recommended to choose a micro-influencer who is active during the FA Cup games to maximize engagement. Additionally, selecting an influencer whose content is related to Manchester United is recommended since Manchester United is the most mentioned account related to the FA Cup. Sentiment analysis further suggests that the search for a suitable influencer should focus on those who tend to update their followers on news and events with a neutral emotional tone. Key descriptive words in the analysis highlight the importance of tweeting about the results of highly significant matches, particularly from the quarter-final stage, to increase engagement.

Applying these filters on the available data will fetch us the influencer mentioned below with some of his tweet stats and sample tweets.



User ID: 1406215235829180000  
Followers: 31,283

THE THEME AROUND TWEET CONTENT: News updates  
EMOTION: Neutral or Positive  
TARGET FOOTBALL CLUB: Manchester United

"OFFICIAL: #MUFC will face Fulham in the #FACup quarter-final at Old Trafford."

"Steve McClaren tells Erik ten Hag "you did it" and points out the #MUFC manager's "winning mentality" 🏆🔴 \n\n @kway\_bumaye #FACup #MUNWHU"

Some stats:

Most active time: Game day (1<sup>st</sup> March 2023)  
Maximum retweets for a tweet: 36  
Maximum likes for a tweet: 402

## CONCLUSION AND FURTHER RECOMMENDATIONS

Using data and text analytics a lot of insights can be gathered from social media comments as performed in this report. The overall role of social media and how people's emotions are brought together on the platform was studied. It helped us choose a micro influencer who has the correct audience and content and further informed us to assist the influencer to increase engagement.

### Further recommendations for increasing engagement:

1. Advice chosen influencer to tweet on other followed topics such as major clubs playing in FA cup that are not Manchester United.
2. Check if the system of choosing game fixtures can be updated and made fairer. Perform an initial survey on twitter through the influencer.
3. Use data more than a week span and perform similar analysis to understand behavior of people across different stages of the tournament.