

# Gathering Tweets on Super Bowl LVI

DS 3010 - Case Study 1 Report

Team 9: Abigail Albuquerque, Aria Yan, Isabel Herrero Estrada, Megan Sin, Sandra Phan

The Super Bowl has been an annual championship that traces back for many years. Over the last decade, the growing technology has allowed more people to tune in to this huge sports event connecting several fans across the globe. The Super Bowl not only has become an important yearly tradition in the United States, but it has grown financially as one of the most advertised and televised events. The first Super Bowl was played in 1967, and it was about \$75,000 for a 60-second advertisement spot. Today, commercials can cost up to \$5 million (Corey, 2021). Given its huge audience, we found it interesting to take a closer look at this “business” and how social media, such as Twitter has influenced its growth and its audience reach.

## Problems 1 and 2:

In problems 1 and 2, we used the Twitter Search API to sample a collection of tweets on a certain topic and analyze the sample based on frequency. We chose to collect data on the Super Bowl since the Super Bowl is taking place within the next month. It is one of the biggest sporting events in the USA and undoubtedly, there are a lot of people tweeting about it.

To analyze the data, we used the Twitter search API to sample a collection of tweets. Our search topic was ‘superbowl’. We extracted 300 tweets with a total word count of 5367. These tweets were saved in a .txt file on google drive.

The next step was to analyze tweets and tweet entities using frequency analysis.

We changed all the words to lowercase to remove duplicates of the same word just in different cases. We then removed all the stop words using a list from

[http://ir.dcs.gla.ac.uk/resources/linguistic\\_utils/stop\\_words](http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words). We then counted the remaining list of words into a dictionary and sorted them. Here are the top 30 words and their frequency.



Figure 1: Plot of top 30 words based on frequency

Many of these words were from the same tweet replied to multiple times in the list of popular tweets.

We repeated the same process but this time we extracted the retweet from the search results. We then sorted the list of tweets in order.

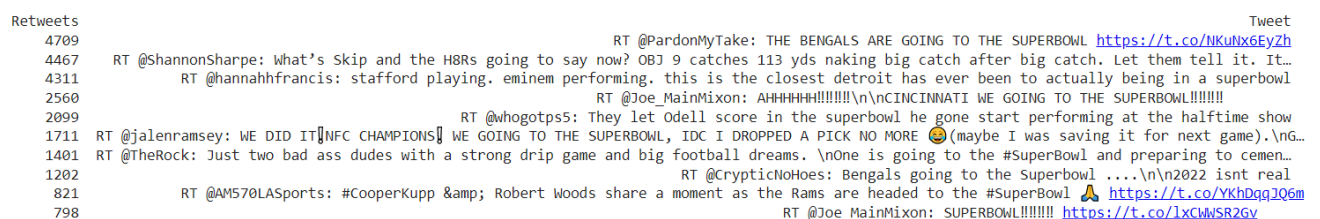


Figure 2: Plot of the top 10 tweets

Hashtag	Frequency	Hashtag
	90	SuperBowl
	8	RamsHouse
	6	Rams
	6	Bengals
	5	RuleTheJungle
	4	WhoDey
	4	NFL
	4	CooperKupp
	3	DevelopedHere
	2	Superbowl

*Figure 3: Plot of the most popular tweet entities*

We repeated the same process as word frequency but this time for the hashtags in each tweet.

User Mention	Frequency	Users
	54	hannahhfrancis
	21	whogotps5
	14	ShannonSharpe
	11	RamsNFL
	11	JoeyB
	11	Bengals
	10	TheRock
	7	Joe_MainMixon
	6	SportsCenter
	6	PardonMyTake

*Figure 4: The top 10 users mentioned*

We repeated the same process as word frequency but this time we found the frequency of user mentions. As you can see, most user mentions were of the popular tweets because they were replied to more.

### **Problem 3:**

For problem 3, we had to pick a popular Twitter user to sample 20 followers and friends. We chose celebrity Kendrick Lamar whose Twitter username is @kendricklamar. The images below show 20 of his followers and friends with their ID numbers as well as their screen names. Additionally, to find mutual followers, we iterated through the users he followed (friends) and looked for users who followed him back. Since the number of users Kendrick follows is a far

smaller number than users who follow him, it was a more efficient choice to iterate through friends rather than followers.

20 Followers:		
Screen Name	ID	
Samanth41861701	1266252418288762881	
I4gYZkPAfqYg5mP	1481760379802124288	
ISAACDA78059235	1487248409299927041	
Cameronjaylin05	1175612819921952769	
3xoticpradaz	1407216591150923777	
zxaixxxpp	1487004229931712512	
MkKimmie	1262139937484427265	
Thubale37785600	1344199392685916160	
thuGGer007	1175740155166167040	
FNeichze	1488376842054742017	
EgyptWilson6	1488374810816262145	
iluminattipa_	1460128378569248773	
ilovewomen1010	1488009209623027717	
majozie_zamo	1478943681918943233	
YEALeADemita	2199077709	
ChrisMe85733821	1442829200872910848	
lexi_is_ok	1374800786400641035	
HOWUCALLIT	60620686	
just_j4ii	1488369924326236161	
layjkunkyess	1488374237064749061	

Figure 5: Twenty of @kendricklamar's Twitter followers

20 Friends:		
Screen Name	ID	
Dude_Br0	255754440	
babykeem	2423213880	
redditSpacePorn	822825713615175680	
FEhrsam	315991624	
IamMRMOSELY	552086618	
Zenos_palace	2869425472	
ExavierTv	741518419208503297	
iamdesibanks	224468287	
1996Biggs	818249674108923907	
mikealiscool123	1148498270399811584	
KingOfQueenz	19298472	
ROKHOUSEMEDIA	81425158	
JordanPeele	63302020	
earlxsweat	486955518	
reasonTDE	3002054440	
DanielCaesar	1739829048	
spiceadams	30959290	
IamHWood	53206129	
SupremeDreams_1	305967569	
TheRaskinTwins	254128442	

Figure 6: Twenty of @kendricklamar's Twitter friends, users he follows

Mutuals:		
Screen Name		ID
Dude_Br0		255754440
babykeem		2423213880
redditSpacePorn	822825713615175680	
IamMRMOSELY		552086618
Zenos_palace		2869425472
ExavierTv	741518419208503297	
1996Biggs	818249674108923907	
mikealiscool123	1148498270399811584	
KingOfQueenz		19298472
ROKHOUSEMEDIA		81425158
JordanPeele		63302020
earlxsweat		486955518
reasonTDE		3002054440
DanielCaesar		1739829048
spiceadams		30959290
IamHWood		53206129
SupremeDreams_1		305967569

*Figure 7: @kendricklamar's mutual users*

#### **Problem 4:**

Twitter data can be very useful for businesses to identify their audience, optimize content, engage with customers, monitor their products etc. The Super Bowl is a big opportunity for businesses to advertise. Since the Super Bowl is such a large-scale event, viewed by such a wide and diverse audience, businesses that want to advertise may find it more beneficial and economical to target their advertisements. Thus, we wanted to see based on tweet density regarding the Super Bowl, which state is the most excited about the Super Bowl. However, we ran into an issue where none of our sample tweets had their geolocations turned on, thus not allowing us to perform this analysis.

Hence, we tweaked our question to see which athletes of each team from the current Super Bowl are the most popular, so brands could sponsor them and involve them in commercials. This idea is beneficial for several reasons. For instance, brand deals could use this information to rather invest in making merchandise for said most popular athlete and populate bigger amounts of the made merchandise throughout their store locations. In addition, the knowledge of which athlete is most popular leads to an incredible opportunity for several distinct businesses which can range from apparel to products. Finding out the most popular athlete from one of the most viewed sports events in history gives an edge to businesses that append themselves to their image. For

example, said athlete could promote specific apparel from a brand such as cleats, shirts, gloves, etc. which in hindsight are later associated with its popularity and exceeding performance therefore promoting the selling and production of said brand and apparel. In addition, this does not limit brand deals to being only apparel; this athlete could go even further to promoting a lifestyle and products to its fans. This could include things in the realm of cars, health, beauty, etc. allowing a broader range of advertisement and brand deals possible. This is extremely useful for any kind of level of business given that advertising helps raise your target's demographic awareness, therefore making this strategy a beneficial investment in the long run. A further step in this idea could be identifying different factors such as age, gender, etc that could narrow down the range of products to be promoted and therefore tailor the advertisement to this specific population. This would aid in creating even more specific targeted ads.

Frequency	Words
169	rt
99	#superbowl
79	superbowl
30	day
27	stafford

*Figure 8: Plot of 5 most popular words in 300 tweets*

Frequency	Words
625	rt
337	superbowl
333	#superbowl
121	day
108	bengals
106	win
89	just
82	santonio
79	td
69	spectacular
69	roethlisberger
69	linked
69	holmes
69	grabs
69	ben
69	@superbow...
69	@nfllegends:
69	2009,
64	stafford

*Figure 9: Plot of most popular words in 1000 tweets*

Based on the available data and our analysis Stafford is the rising star of this Super Bowl and companies should focus on sponsoring him. There is a lot of buzz about Santonio, a former football receiver, as people remember his catch from 10 years ago. There were also a lot of tweets about Roethlisberger, mostly because he is retiring. We do understand that our data is limited. Running an analysis on a large sample of data will give more accurate results with a larger scope.

### References

Corey, Author: Allie. "History of Super Bowl ADS: How and When They Became so Popular for Brands and Viewers." *Ksdk.com*, 26 Jan. 2021,  
<https://www.ksdk.com/article/sports/nfl/superbowl/super-bowl-ads-history/63-919a4d1c-121a-47d4-98f3-746564529b21>.