

# Google Tweets Analysis

Salma BARKETALLAH, Ndjaidoukmi KEDA, Noémie GAUTIER

Feb 04, 2021

## Contents

<b>Introduction:</b>	<b>3</b>
<b>Part 1- Methodology:</b>	<b>3</b>
<b>PART 2 - Tweets from Google:</b>	<b>4</b>
1. Summary of Tweets: . . . . .	4
1.1. The Categories of Tweets: . . . . .	4
1.2. The Sources of Tweets: . . . . .	4
1.3. The Frequency of Tweets: . . . . .	5
1.4.The Tweets' Engagement: . . . . .	6
2. Content Analysis: . . . . .	6
2.1. Word cloud analysis: . . . . .	6
2.2. Sentiment Analysis: . . . . .	8
2.3. Topic Analysis: . . . . .	9
<b>PART 3 - Tweets about Google:</b>	<b>11</b>
1. Summary of Tweets: . . . . .	11
1.1. The Categories of Tweets: . . . . .	11
1.2. The Sources of Tweets: . . . . .	11
1.3. The Tweets' Engagement: . . . . .	12
2. Content Analysis: . . . . .	13
2.1. Word cloud analysis: . . . . .	13
2.2. Sentiment Analysis: . . . . .	14
2.3. Topic Analysis: . . . . .	15

<b>Part 4 - Tweets with #google:</b>	<b>17</b>
1. Summary of Tweets: . . . . .	17
1.1. The Categories of Tweets: . . . . .	17
1.2. The Source of Tweets: . . . . .	17
1.3. The Tweets' Engagement: . . . . .	18
2. Content Analysis: . . . . .	19
2.1. Word cloud analysis: . . . . .	19
2.2. Sentiment Analysis: . . . . .	20
2.3. Topic Analysis: . . . . .	21
<b>Part 5 - Tweets with @google:</b>	<b>22</b>
1. Summary of Tweets: . . . . .	22
1.1 The Categories of Tweets: . . . . .	22
1.2. The Source of Tweets: . . . . .	23
1.3. The Tweets' Engagement: . . . . .	24
2. Content Analysis: . . . . .	24
2.1. Word Cloud Analysis: . . . . .	24
2.2. Sentiment Analysis: . . . . .	26
2.3. Topic Analysis: . . . . .	27
<b>PART 6 - Users Analysis:</b>	<b>29</b>
1. Followers Analysis: . . . . .	29
2. Friends Analysis: . . . . .	31
3. Friends vs Followers: . . . . .	33
<b>Part 7 - Limitations:</b>	<b>36</b>
<b>Conclusion:</b>	<b>37</b>

## **Introduction:**

This report is part of the final project of Social Media Analytics Course of the master Big Data analytics for Business at IESEG School of Management. The objective of the project is to conduct an in-depth analysis of a tweeter account including different aspects to be analyzed such as but not limited to: Tweets generated from that accounts and tweets mentioning the account. The account chosen for this report is Google.

## **Part 1- Methodology:**

As important quantity of data should be gathered for a relevant text mining project, the choice to go for a big and famous company was obvious. Thus, the chosen tweeter account is Google. Additionally, the project is conducted on information collected in January 2021. Only tweets in English were taken into consideration to make the analysis easier to understand, and respect the amount of Tokens we had. Indeed, as the analysis is conducted in an academic environment, tweeter standard API accounts were used to extract data from Twitter. The analysis was conducted using R and mainly the package ‘RTweet’. Other packages were also used such as ‘tidytext’ or ‘dplyr’ for data processing and visualization.

The whole project followed these steps:

1. Data was collected by calling the tweeter API. Three tokens were used to collect the maximum amount of data.
2. Data was saved in csv files.
3. Data was reloaded in R in various notebooks.
4. Data was pre-processed. In this step, tweets' text was first tokenized and cleaned by removing punctuation, stop words, unwanted symbols and words...
5. First level analysis was conducted by analyzing the frequency of words, plotting the word clouds, and looking at the type of tweets.
6. Engagement analysis was conducted to understand how followers behave with Google Tweets.
7. Sentiment analysis was conducted to detect positive or negative sentiment in tweets.
8. Topic analysis was conducted to get insights on the tweets topics.
9. A user analysis was conducted to gather information on the friends, followers and users of the Google account.
10. The following report was generated using RMarkdown and all the R script that was used in the previous steps is handed with this report.

A shiny application could be created based on the data collected but the team decided not to. We wanted to focus on explaining the insights gathered through our data in this report. The following parts of this reports will be on the findings and outcomes of this research on ‘Google’ Tweets.

## PART 2 - Tweets from Google:

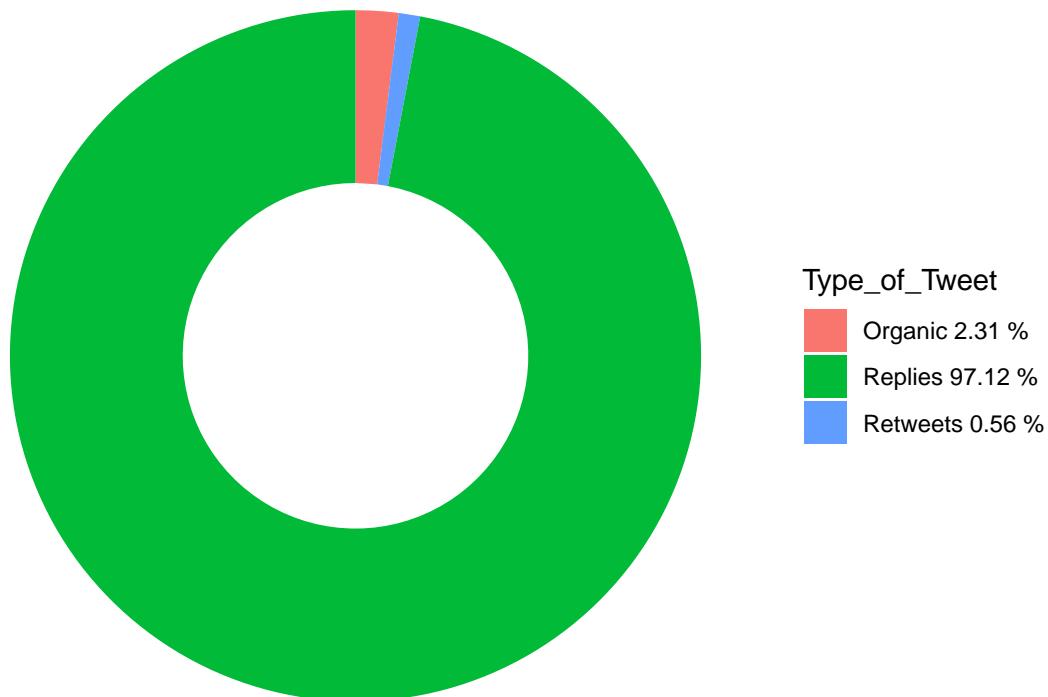
This part is going to focus only on tweets that are issued directly from the Google official account in 2020.

### 1. Summary of Tweets:

The total number of tweets analyzed in this category is: 3200.

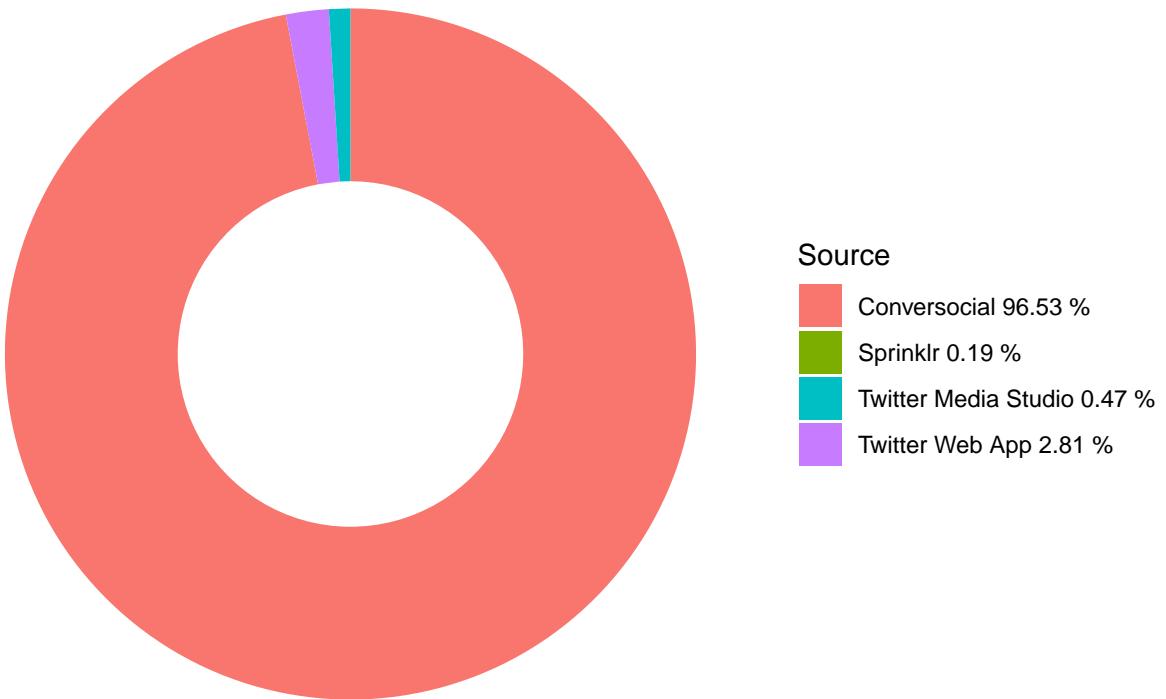
#### 1.1. The Categories of Tweets:

The big majority (97%) of these tweets are replies to other tweets. Organic and retweets represent less than 1%. This means that the Google twitter account mostly replies to posts, the account rarely posts or retweet posts in comparison. One could argue that this is common for most twitter accounts as people tend to answer tweets more than post them. But, 97% seems to be a very high percentage of replies in comparison to the rest. This highlights that Google mainly uses Twitter to answer questions about their products, and reply to tweets from other users. Therefore they use Twitter to communicate and answers others more than to share new information.



#### 1.2. The Sources of Tweets:

The majority of these tweets are from conversocial that is a social customer service software Google and other big companies use to manage the flow of customer service inquiries and discussions on social media networks (97%). As shown in the pie chart below, the remaining part are published from the tweeter web application or Tweeter Media Studio.

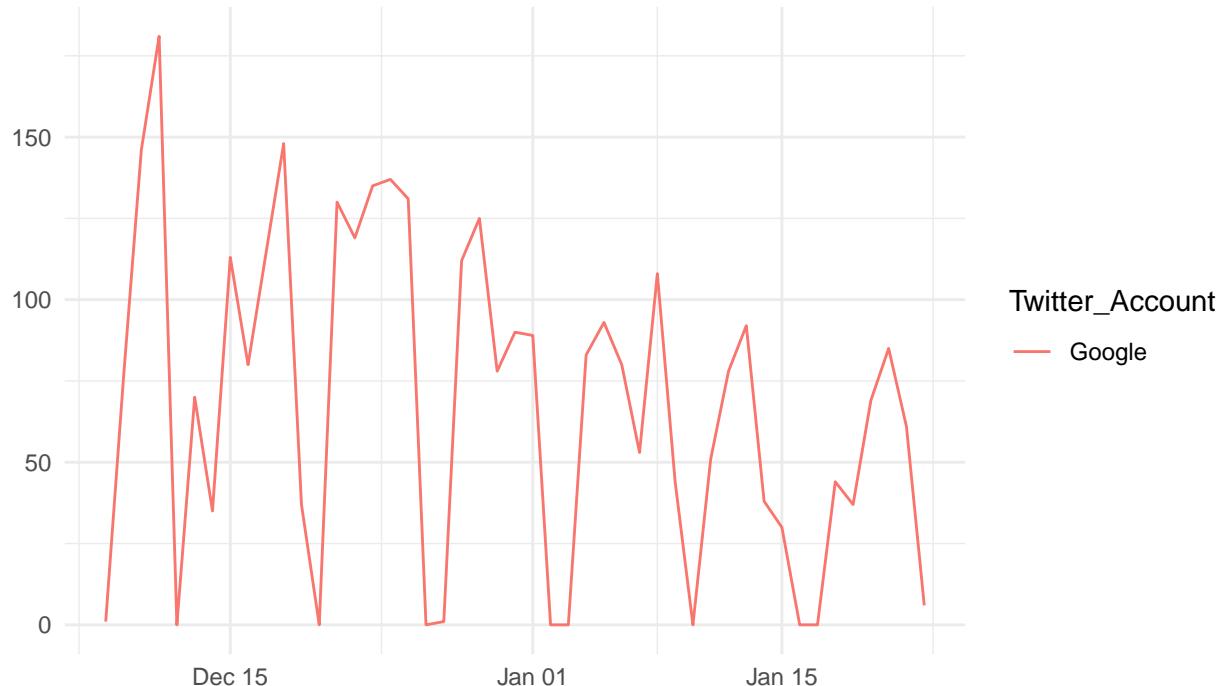


### 1.3. The Frequency of Tweets:

We also looked into when tweets are posted from the Google Twitter account. The graph bellow illustrates the frequency of tweets in the last two months. As we can see the frequency is not a straight stable line which would have indicated that Google posted the same amount of tweets daily. Instead of graph highlights that some days up to 175 tweets are posted from the Google account, while on other days 0 tweets are published. But overall we can see that Google is quite present on twitter and there is rarely more than a day or two without them posting a tweet.

## Frequency of Tweets from Googles

Tweet counts aggregated by days



Source: Data collected from Twitter's REST API via rtweet

### 1.4. The Tweets' Engagement:

Concerning the organic tweets of Google, they have on average 669 likes and are retweeted 137 times. The most liked one and the most retweeted ones are related to the new year 2021 and to christmas celebrations both created in December 2020. The increased engagement of these tweets can be correlated to the topic: the end of year holidays. Indeed, this topic can relate to many users, and with the difficult year most people faced in 2020, the end of year posts seem to be very appreciated.

## 2. Content Analysis:

After conducting a summary analysis, in this part, we will undertake text mining and look at the text and words inside tweets.

### 2.1. Word cloud analysis:

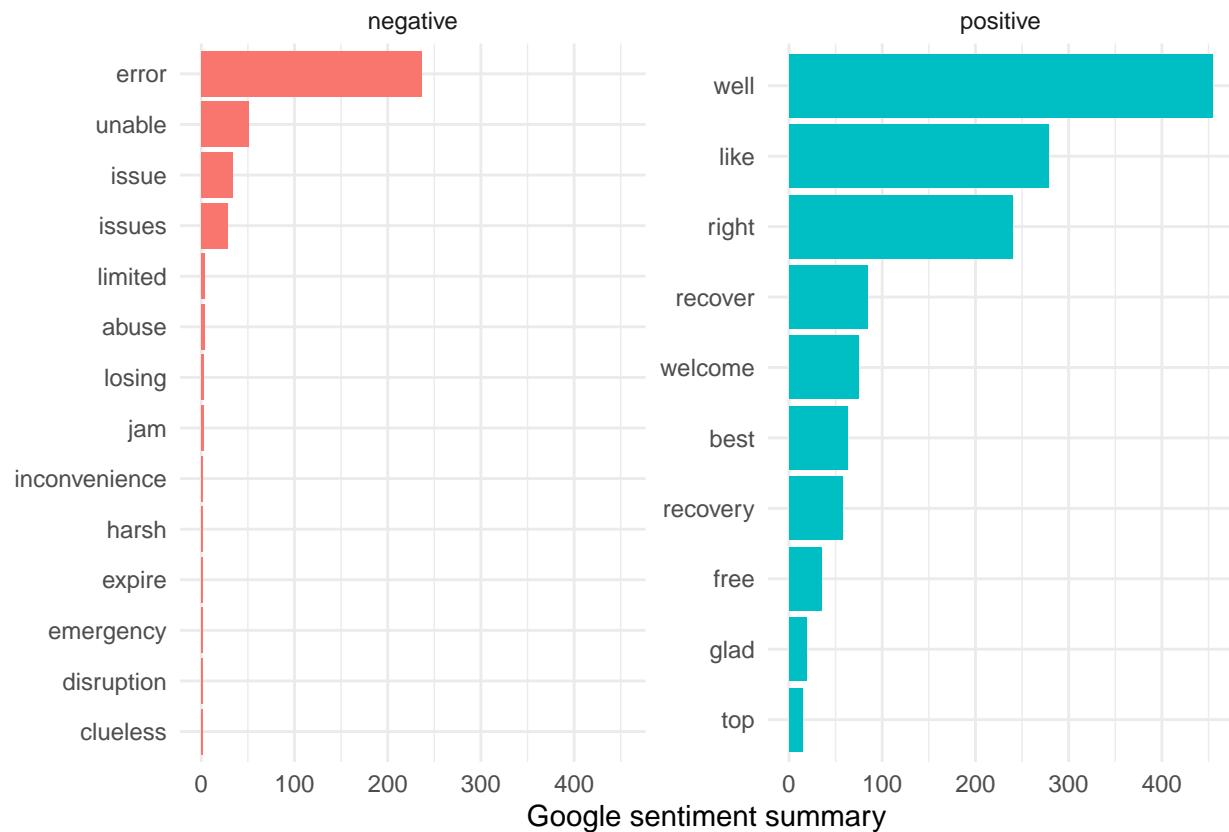
This analysis has been made twice for the text constituting the tweets. The first one is for the text before being preprocessed and the second time, when the text was cleaned and some words were deleted such as the stopwords that are the most common words or the word Google. Looking at these two word clouds below we can see how preprocessing was helpful in making the word cloud analysis more relevant. We can also detect an underlying trend in Google tweets. The lexical field of help and sharing is predominant with words such as 'help', 'please', 'suggest', 'guide'... This is logical when we know that Google shares more tweets as replies. So we can assume that Google is mainly responding to help questions related to Google products and services when using the tweeter platform.

httpstcouhhudctnqf  
look share  
there account  
tips YOUR may are  
confirm when the help these  
hmm next you and through  
have out see can know well  
please steps what just  
try follow this lets for google  
httpstconrurccasgj here  
completing



## 2.2. Sentiment Analysis:

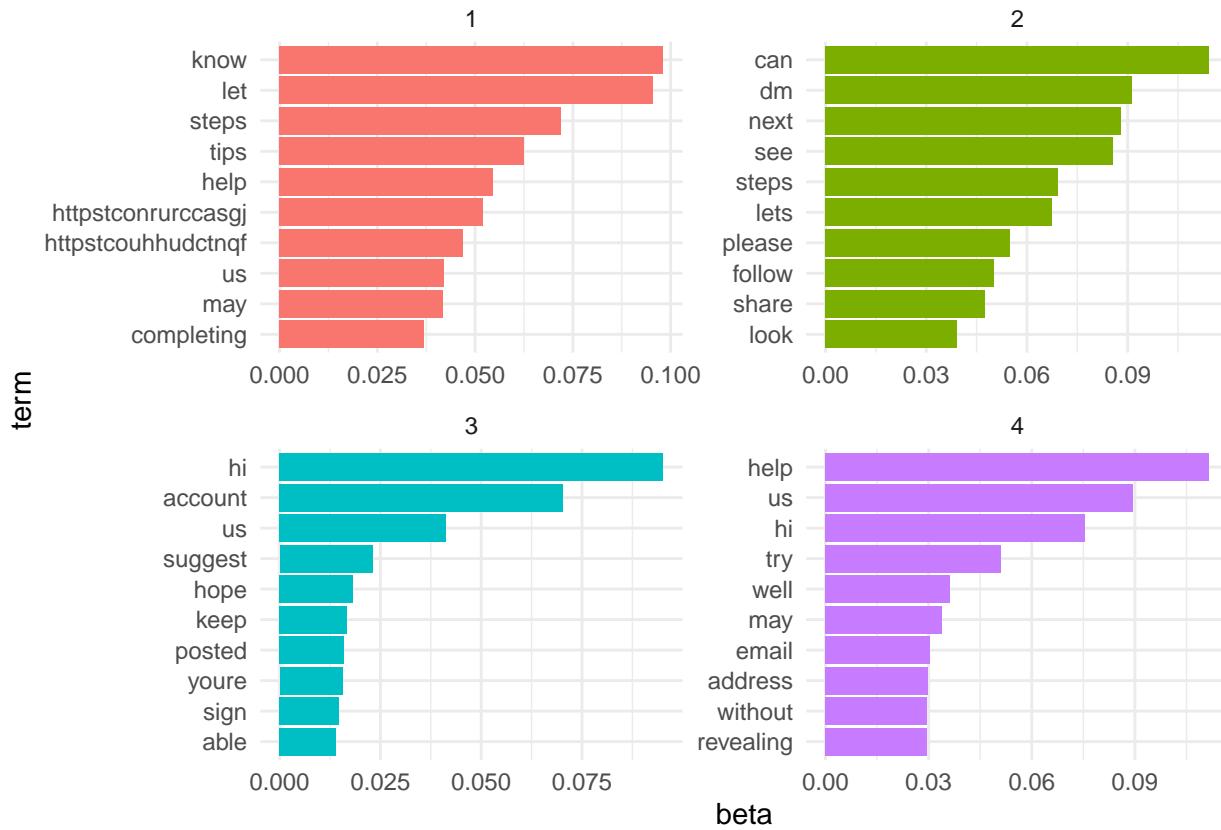
Regarding the sentiment analysis of tweets from the Google account we can see that positive words are used more than negative ones. The graph bellow highlights that the positive word used the most such as “well”, “like”, and “right” are used between 250 and 450 times. On the other hand the negative words the most used: “error”, “unable”, and “issue” are used between 50 and 250 times. This highlights that Google tends to use a positive vocabulary in its tweets. As mentioned above Google tends to reply to tweets the most and in these replies they use mostly positive words, most likely to promote a good image of the company and their products. The negative words used are all related to issues or problems most likely faced by their products and that they are trying to answer to.



### 2.3. Topic Analysis:

The tweets from google timeline can be split into 4 different topics which are: - First topic is on the help request from google product users. In this topic we have frequent words such as help, try, email, address - The second topic is on direct mail (DM) request from google. We have on this topic: dm(direct mail), can, steps, please, follow as frequent words - The third topics is on tips provided by google to his customers. Frequent words in this topic are: know, tips, help, completing - Last topic is on helps for the account and most common words are: account, helps, suggest, hope

The graph bellow highlights the words most used for each topic.



This topic analysis allows us to see what type of tweets are posted from the Google account. As mentioned above, Google mostly replies to tweets, and one of the topics most used is to help requests from product users. The second topic, Direct Mail, shows that Google is open to having direct conversations with users that seem to have difficult issues that most likely cannot be solved by one reply. The third topic, tips provided, highlights that Google also uses their account to share information to its customers. Finally, the last topic reflects that Google also posts regarding help or suggestions for their account. The company may be gathering feedback about their tweets or products.

## PART 3 - Tweets about Google:

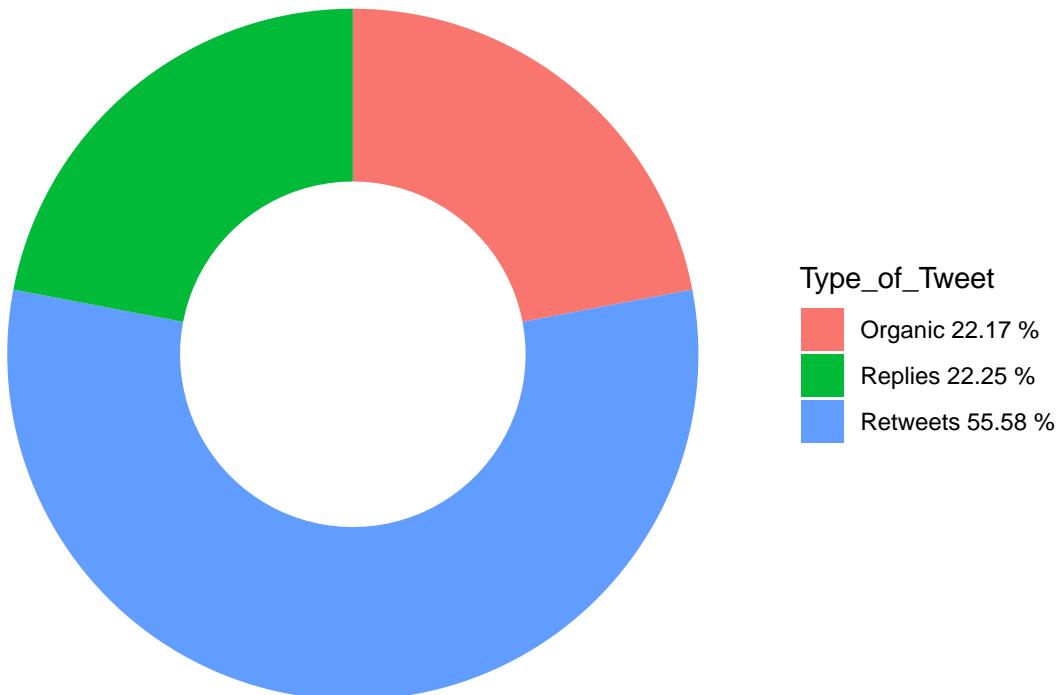
After analyzing all the data coming from the Google account, it is important to analyze what people share on Tweeter about Google. This means this analysis takes into consideration all the tweets mentioning Google both by the @ and # or directly using the word “google”.

### 1. Summary of Tweets:

The total number of tweets considered is 1200. This is the maximum of tweets we could get with our token.

#### 1.1. The Categories of Tweets:

This type of tweets is mainly mentioning ‘google’ in retweets (56%). This makes sense as most Twitter users retweet posts more than they post themselves. Replies to tweets and organic tweets are basically equal (22% each). The following pie chart shows the proportion of each category in this type of tweets.

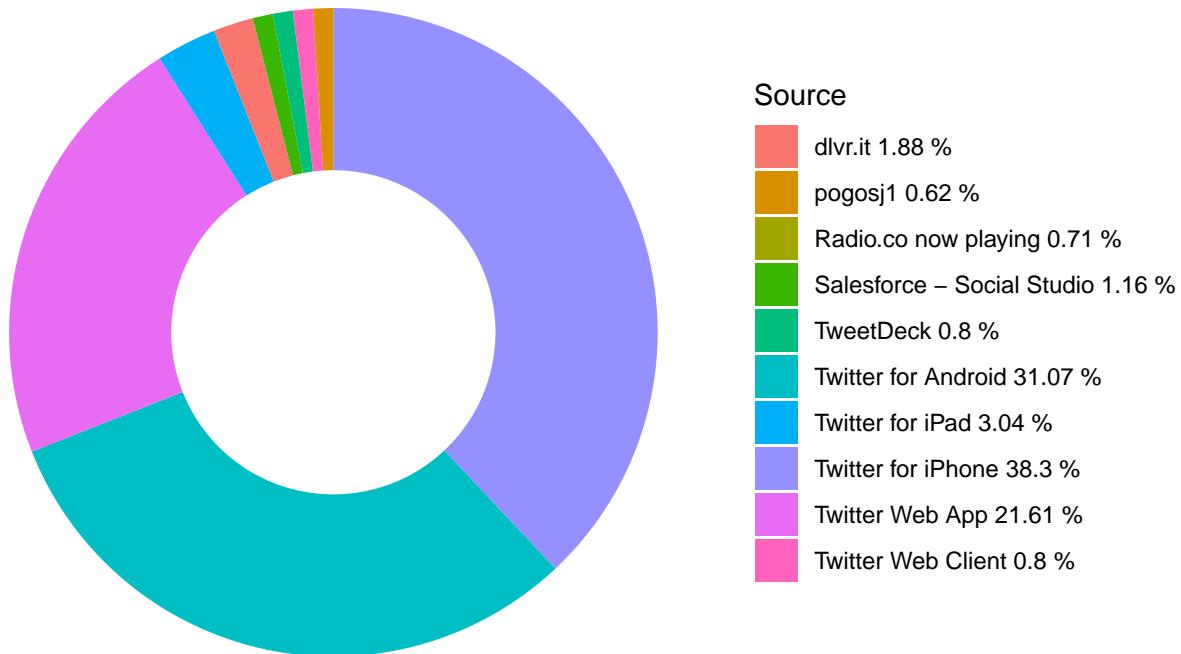


#### 1.2. The Sources of Tweets:

As the graph bellow highlights, there are various origins where tweets mentioning Google come from. The first source used to publish these tweets is a phone device (70%). As most users today use their phones it makes sense that most tweets come from there. We can also see that Iphones are more used than Android devices (7% more). This can reflect the type of users that mention Google and where they are from. Indeed, Apple phones are used more in the USA, or in Europe than Android. But in Asia they are used less. The

fact that most tweets come from Iphones suggest that most users who mention Google are from the USA or Europe.

On the third position, the tweets are shared from the web application. The graph illustrates that 21.61% of users use the web app. This is interesting as the usage of the web app could have a link with the users' age, indeed the younger population may tend to use their phone more. As you can see in the following pie chart, other sources are used but are a minority.



### 1.3. The Tweets' Engagement:

We looked into the most liked tweets and the most retweeted tweets mentioning Google. We found that one tweet was both the most liked and the most retweeted. This tweet is from an account 'pannchoa'. It has 547 likes and was retweeted 210 times. This tweet is very simple, it is one line long and contains two weblinks. The tweet is as follows: "Inner Circle protest against YG's mistreatment towards Winner with a LED truck <https://t.co/PBdE9u9Fj1> <https://t.co/qDblWQ8KVf>".

The structure of this tweet can influence its engagement, as it is easy to read, short and with additional links to look into. The user can also have an impact on the engagement depending on the amount of friends and followers they have. Finally, the topic, an inner circle protest can also influence the engagement, as protests are often discussed and debated online.

On average a tweet mentioning Google is liked 4 times and retweeted once. This is interesting as we see that the average engagement is quite low. This could be explained by a lot of tweets mentioning Google, some may get retweeted, and liked but most are from everyday people and get 0 engagement.

## 2. Content Analysis:

### **2.1. Word cloud analysis:**

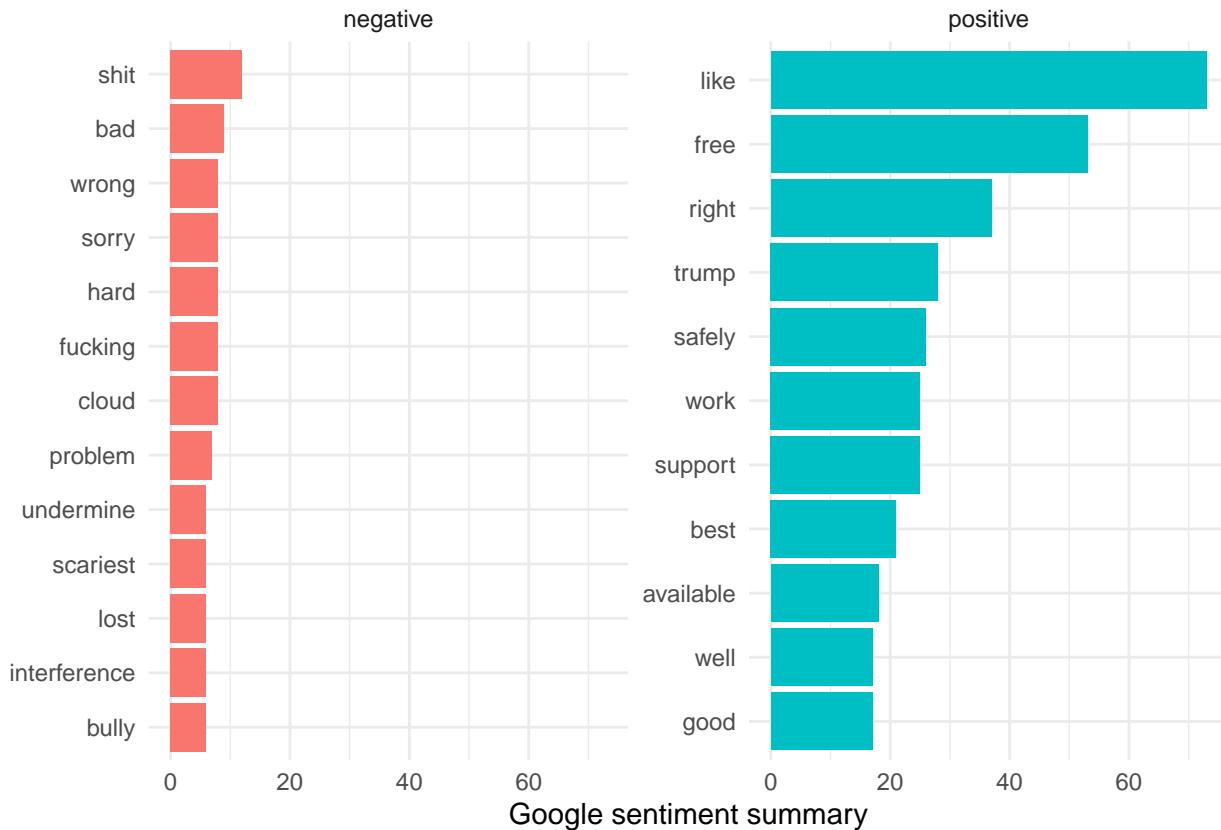
As for the previous category, we undertook a word cloud analysis before and after pre-processing. As we can see, pre-processing was again helpful since the first word cloud does not seem relevant for analysis and that it contains many stop words. For the second word cloud, we can see that tweets that talk about ‘google’ mention as well the other GAFA companies naming Amazon, Apple and Facebook. So we can assume that these tweets are related to the information technology industry. The lexical field of help is also present with frequent words such as ‘please’ and ‘need’.





## 2.2. Sentiment Analysis:

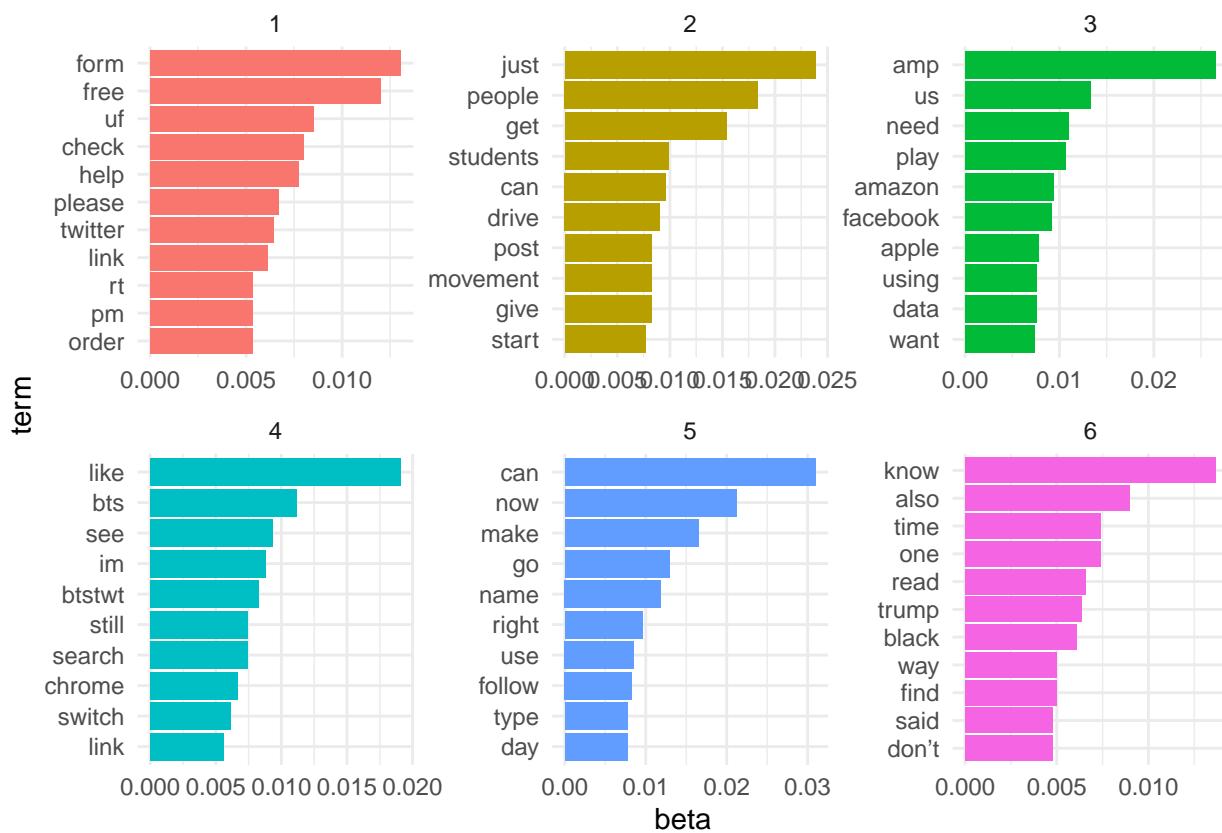
We did a simple sentiment analysis of the tweets mentioning Google. We found, as shown in the graph below that the negative words used were insults or disapprovals, most likely regarding a post, product or action related to Google. The positive words mention approval, and compliments regarding Google. We can also see that one of the positive words is “trump”, which could be qualified as positive or negative, this is quite subjective depending on the political opinion of the person writing the tweet. Finally, we can see that the positive words are mentioned much more than the negative ones. Excluding the word “Trump” which if qualified as negative would be at the top. Overall, positive words seem to be used more often than negative words, suggesting that tweets mentioning google are mostly positive.



### 2.3. Topic Analysis:

During our analysis, we have classified tweets about google into 6 different major topics. Indeed we could have extend topics to a higher number but this will make interpretation really difficult. The topics identifies are: - The first topic is on George Floyd's murder (black life matter movement) with frequent word like free, black, help - The second is on Trump having Covid. Frequent word in the topic are: know, one, trump , covid - The third topic is on the online/distance learning as we have frequent words like student, people, zoom - The fourth topic is on AMP (Accelerator Mobile Page) on chrome with frequent words like AMP, please, need, get, chrome - The fifth topic is related to the 2020 election in united state, frequent words used are :now, make, vote, right - The sixth topic is on the release of new BTS (korean pop) album 2020. Frequent words used for this topic are: BTS, search, like, search

The key words of the topics are illustrated in the graph bellow.



Topic about google on twitter are various they are not concerning only google activities but also general topics related to the society. This illustrates that Google is a multinational company often related to news and topics that they should not be directly related to, but due to their impact on society, they are often linked to every topic on the news. Additionally, we can see that all the topics mentioned above made the news recently, but they are about quite different contexts. Finally, most of these topics are related to the USA, which could suggest a lot of the tweets mentioning Google are from Americans. On the other hand, most of these topics also made the world wide news.

## Part 4 - Tweets with #google:

Unlike the previous part, the following one will only analyze tweets mention '#google'. This will allow to detect if there are trends related to the use of the #.

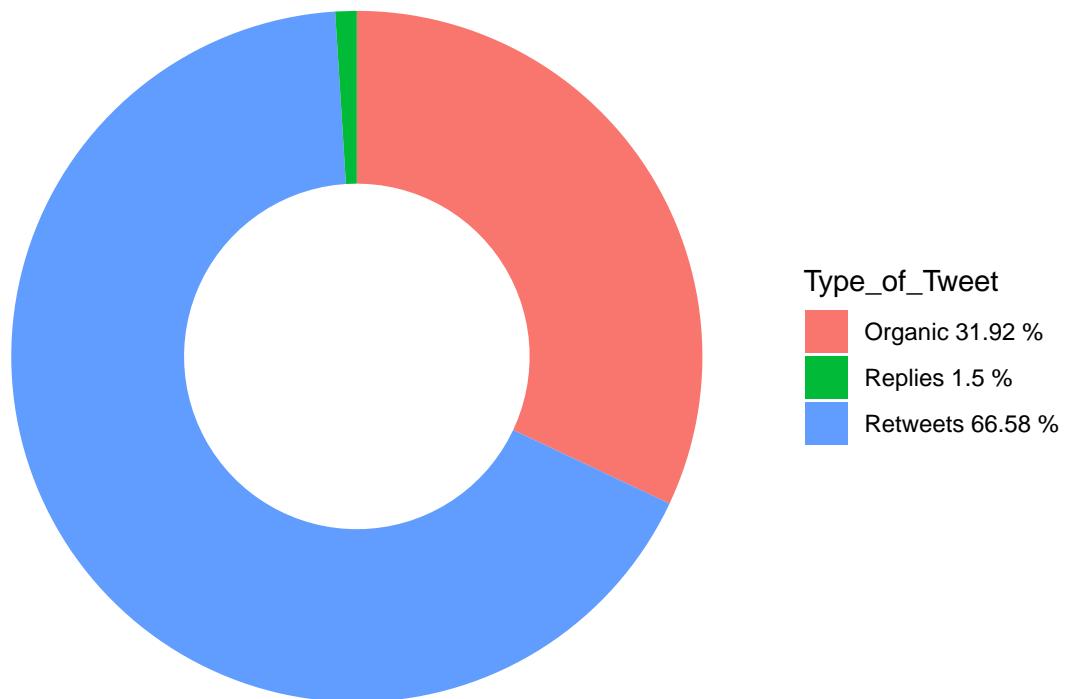
### 1. Summary of Tweets:

Once again the total number of tweets considered is 1200. This is the maximum of tweets we could get with our token.

#### 1.1. The Categories of Tweets:

Tweets with #google are mainly mentioning Google in retweets (66%). This can once again be explained by the proportions of retweets people do on twitter in comparison of organic tweets or replies they post. But here we can see that there are basically no #google in replies (1.5%). On the other hand organic tweets represent 32% of tweets with #google. This means users tend to not use # in replies, but more in retweets or organic tweets.

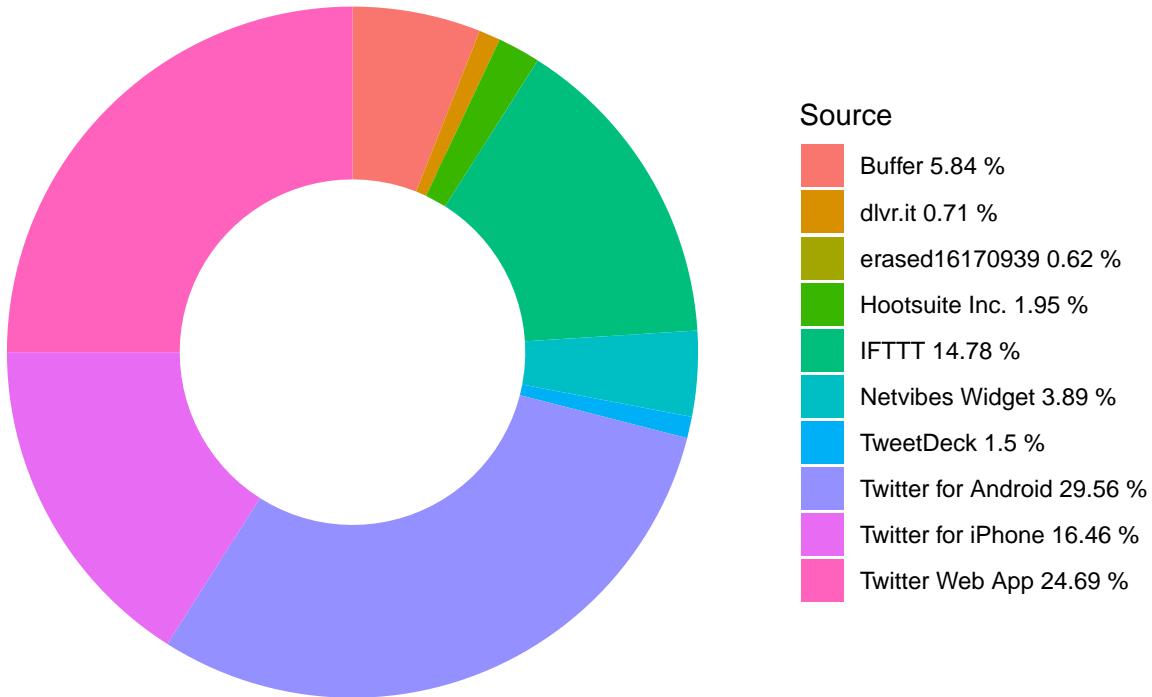
The following pie chart shows the proportion of each category in this type of tweets.



#### 1.2. The Source of Tweets:

Regarding the source of tweets, we can see that similarly to source of all tweets mentioning Google most tweets come from phones (45%). But there are quite a few changes between the source of tweets with #google

and the overall source of tweets for all tweets mentioning Google. Here we can see that there are more tweets from Android devices than Iphones (10%), and that the proportion of tweets coming from phones is smaller in comparison to the rest. Indeed, there are 25% of these tweets that come from the web app, but also 15% from IFTTT, which is a web service. This data highlights that the population of users using # is different from the whole population that mentions Google. They may not come from the same countries, based on the Iphone vs Android, and may not be of the same age group either. Further research into the users of # could be interesting to confirm some of these hypothesis.



### 1.3. The Tweets' Engagement:

Engagement for this category of tweets is also low. They have on average 4 likes and are retweeted on average 2 times. This low engagement can be explained by the high frequency of using #google in tweets. In this category, the most liked tweet is also the most retweeted one. It was issued in October 2020 from the account ‘theblkclst’ and is about a partnership with google. The text of the tweet contains ‘Great news for your Thursday! We’ve partnered with Google Assistant to create the inaugural Black List x Google Assistant Storytelling Fellowship! #heygoogle Read more in @deadline: <https://t.co/iNwbH4Fawn> <https://t.co/3LEuzyWv6d>’. This tweet gathered 336 likes and was retweeted 134 times. Looking at the text of this tweet we can see that it is a positive one celebrating a good news. So can we assume that tweeter users are more engaging with good news? Further investigation on this could lead to interesting results.

## **2. Content Analysis:**

### **2.1. Word cloud analysis:**

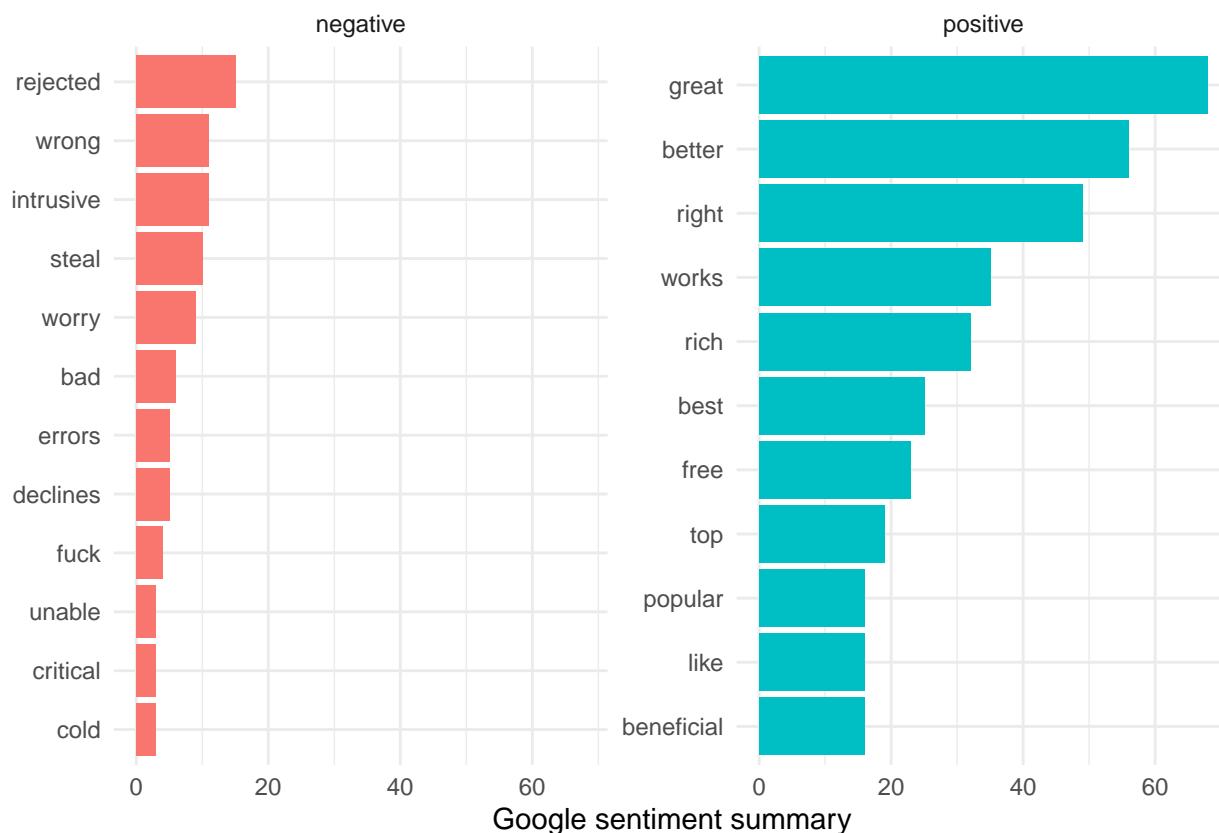
Here again, the two word clouds are about tweet text before and after preprocessing. For this category of tweets, we can see that the lexical field of digital marketing is present with words such as ‘digital marketing’, ‘seo’ and ‘searchengineoptimization’ or ‘ranking’. We can assume here that people mentioning #google in their tweets are looking at ways to improve their visibility on the search engine of Google.





## 2.2. Sentiment Analysis:

Similarly to the other sentiment analysis, we can see on the graph bellow that negative words are used less often than positive words in the tweets with #google. But here we can see that on the contrary to the negative words from all tweets mentioning Google, there are less insults, and more disappointment in the words used. This may reflect on the topic of the tweets which will be discussed below. The words used seem to be less personal and maybe more product related. This is also seen in the positive words, which seem to give an opinion about Google or their products such as the word “works” would suggest.

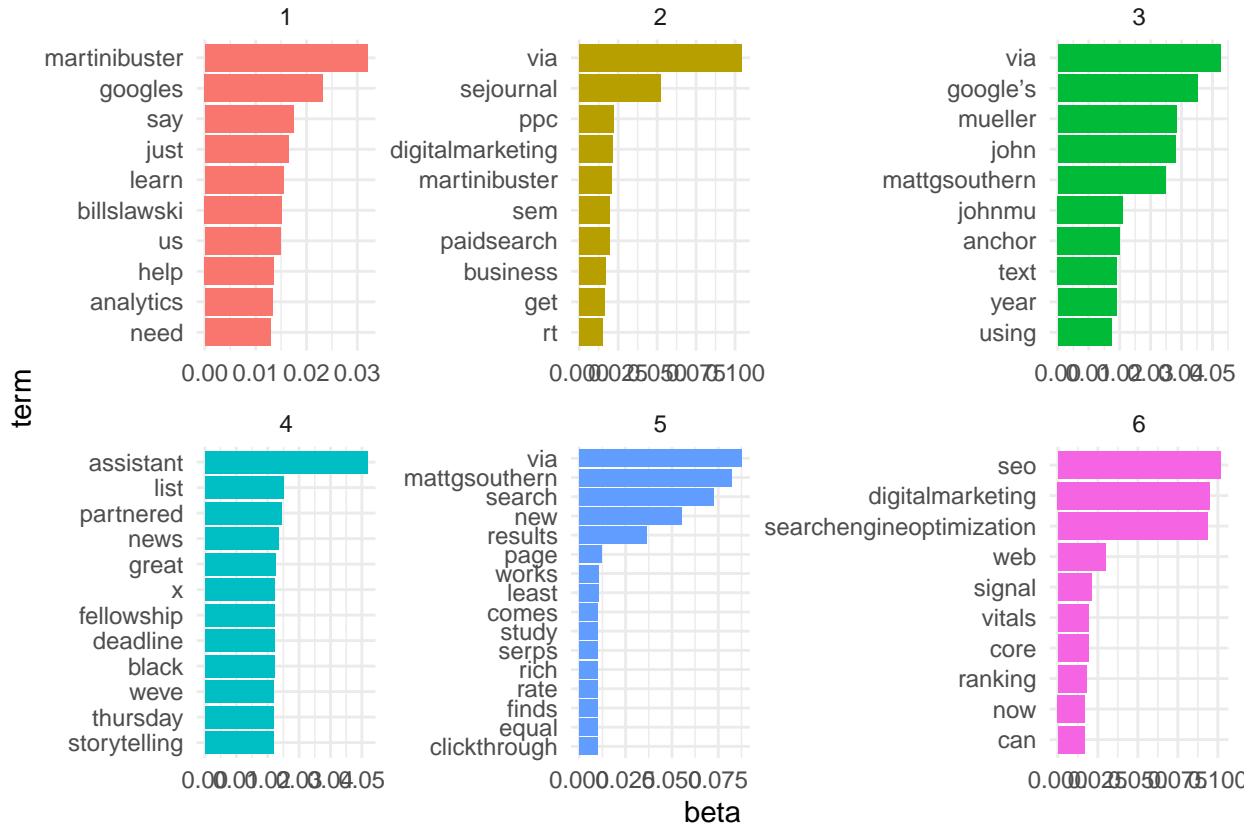


### 2.3. Topic Analysis:

Regarding topics on google hashtags, they are also separate into six different topics - First topics is on search engine optimization and digital marketing with frequent words as seo(search engine optimization), search engine optimization, digital marketing - Second topic is on the new year most frequent words used are new, year, user, better - Third is about media and journalist with frequent words like john muller,an american television journalist, Matt Southern, author at Search Engine - The fourth topics is related to the assistance request made by google product user. Frequent words in this topic is assistance, great, fellowship, news - Fifth topic is on google analytics and most common words are analytics, google analytics, engine, search - Sixth topic is on martini buster, leader company in digital marketing

Contrary to tweets about google, google hashtags tweets are more about google products and assistance request made by users. This once again shows their is a real difference between the use of Google or #google. The hashtag is used more to gather information regarding new products and hightech.

The graph bellow shows the key words used for each topic.



## Part 5 - Tweets with @google:

This section will be dedicated to tweets tagging google with the @ sign. Like the previous part, the objective is to detect trends related to the tweets tagging google.

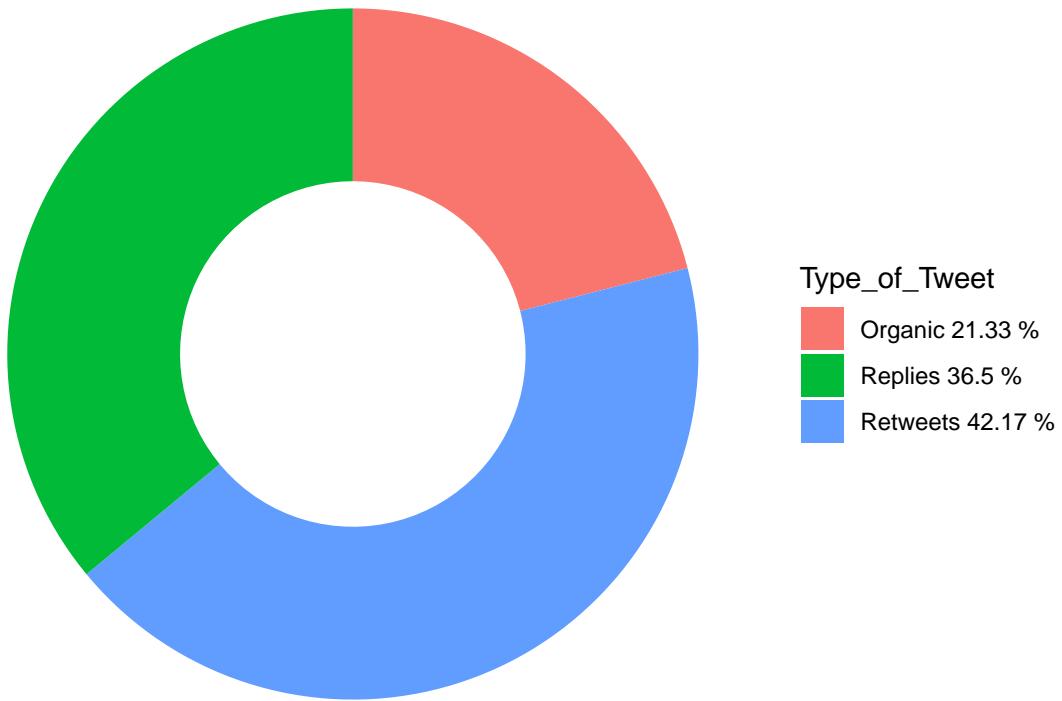
### 1. Summary of Tweets:

The total number of tweets considered is 1200. This is the maximum of tweets we could get with our token.

#### 1.1 The Categories of Tweets:

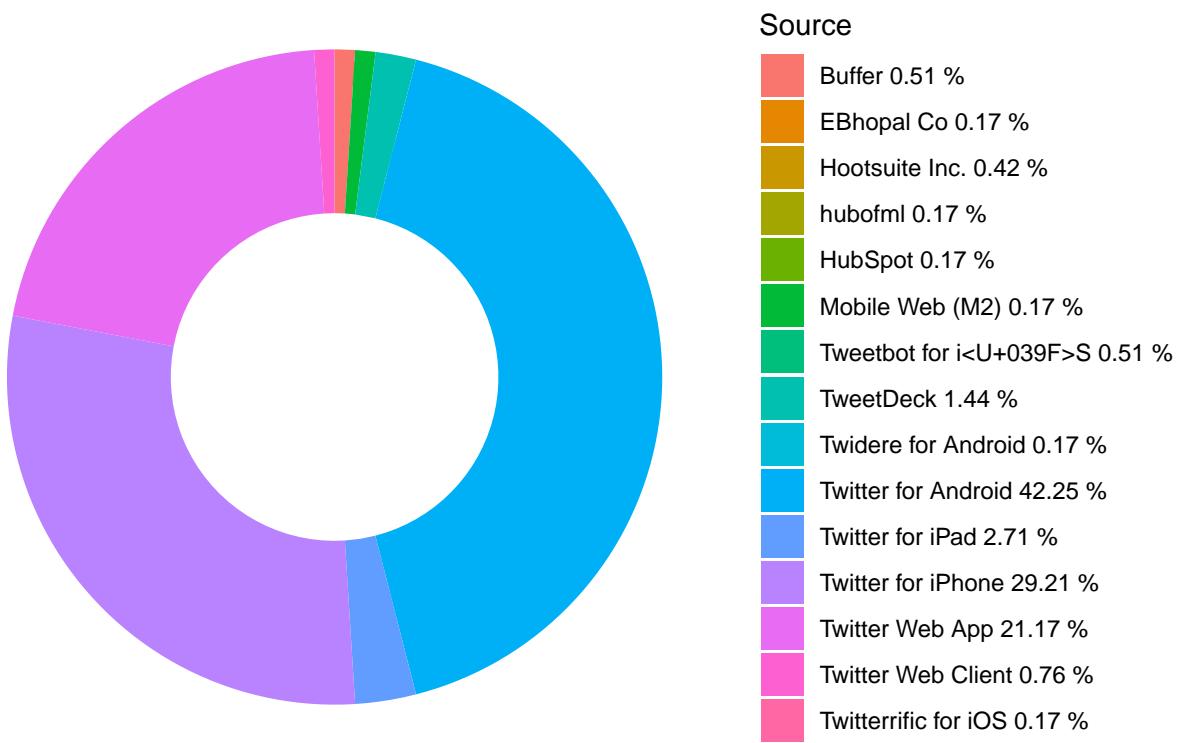
Tweets with @google are mainly mentioning Google in retweets (42%). But the difference between the proportions of retweets and the other types of tweets is smaller than for tweets containing #google, and all tweets mentioning google. Here we can see that 36% of tweets containing @google are replies and 21% are organic tweets. This highlights that the @, which is a direct mention and link to the google account, is more used in replies and organic tweets than the # or just the word google. This is most likely to attract Google attention as mentioning them will send them a notification. Additionally, if users are replying to the Google account they may tag them with the @google, which can explain why there are more replies here.

The following pie chart shows the proportion of each category for tweets containing @google.



## 1.2. The Source of Tweets:

Once again, the source of tweets here looks different from the overall source of tweets mentioning Google, and the source of tweets with #google. In the graph below we can see that 42% of tweets are from an Android device. In regards, 29% are from Iphones, which highlights a higher usage of Androids. Overall, about 70% of tweets are from phones, and 21% from the web app, leaving very little place for the other options. This highlights that the source of tweets with #google and @google are very different, suggesting that the users are too.



### **1.3. The Tweets' Engagement:**

For this category of tweets, engagement is also low as for the two previous categories but is even lower when we see that the average number of likes is 1 and tweets are on average not retweeted (0 retweets). The most liked tweet in this category gathered only 55 likes and is also the most retweeted one (23 retweets only). The tweet comes from the account 'CHIPSAlliance' and was published in June 2020. The text of this tweet is 'Don't miss out on @mithro's talk tomorrow, June 30 at 4 p.m. GMT (9 a.m. PT)! Tim will be talking about an exciting joint project from CHIPS Alliance member company @Google, along with SkyWater, to provide a fully open source Process Design Kit. <https://t.co/zKAZ48OP8S> <https://t.co/acfhNizh0j>' Looking at the words of this tweet, we can see that it also contains positive words such as 'exciting'. This goes in hand with the assumption on engagement provided in the previous part.

## **2. Content Analysis:**

### **2.1. Word Cloud Analysis:**

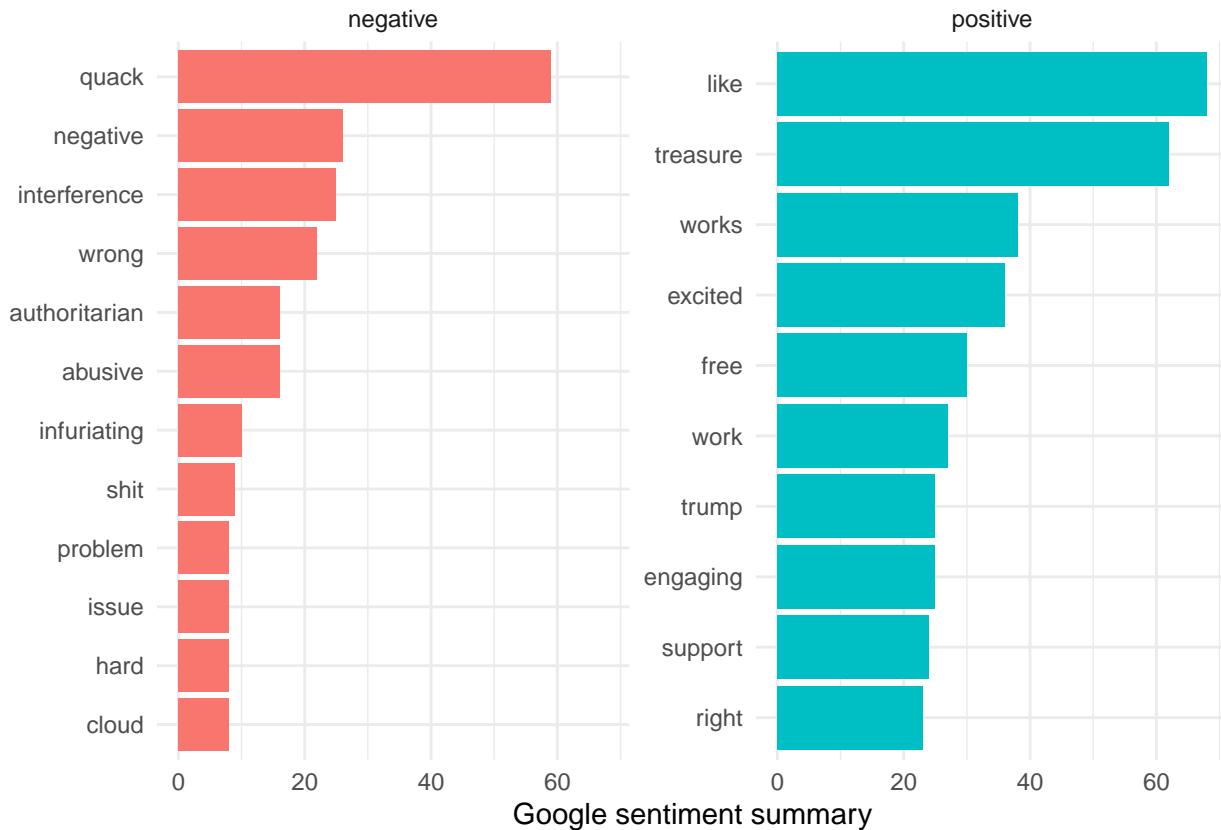
Again, word cloud analysis was conducted for both text without and with preprocessing. Before preprocessing, many irrelevant words appear such as 'the' or 'you' and obviously the word 'Google'. Yet, even after preprocessing for this category and since spelling was not corrected for several reasons, we have words that appear and that are not meaningful or that are in Spanish such as 'poder', 'por'... There is also no underlying trend in this word cloud.

members facebook  
just when treasure  
want not search  
like but you out  
from your for and all its  
have can google  
more por they the what are  
they that this fix has been  
please now dont youtube  
treasuremembers could with research



## 2.2. Sentiment Analysis:

In the graph bellow we can see the results from the sentiment analysis of tweets containing @google. It should be mentioned that the first negative wors “quack” is very used in comparaison to the negative words used in our other sentiment analysis. Here we can see that the firt negative words is used almost as much as the first positive word, suggesting that tweets with @google may be more negative than the rest with #google or just the word google. But we can see that the rest of the negative words are used less than the positive ones, meaning the tweets are most likely not all negative. The words themsleves seem to vary a lot in the topic, this can be explained by the topic analysis bellow.

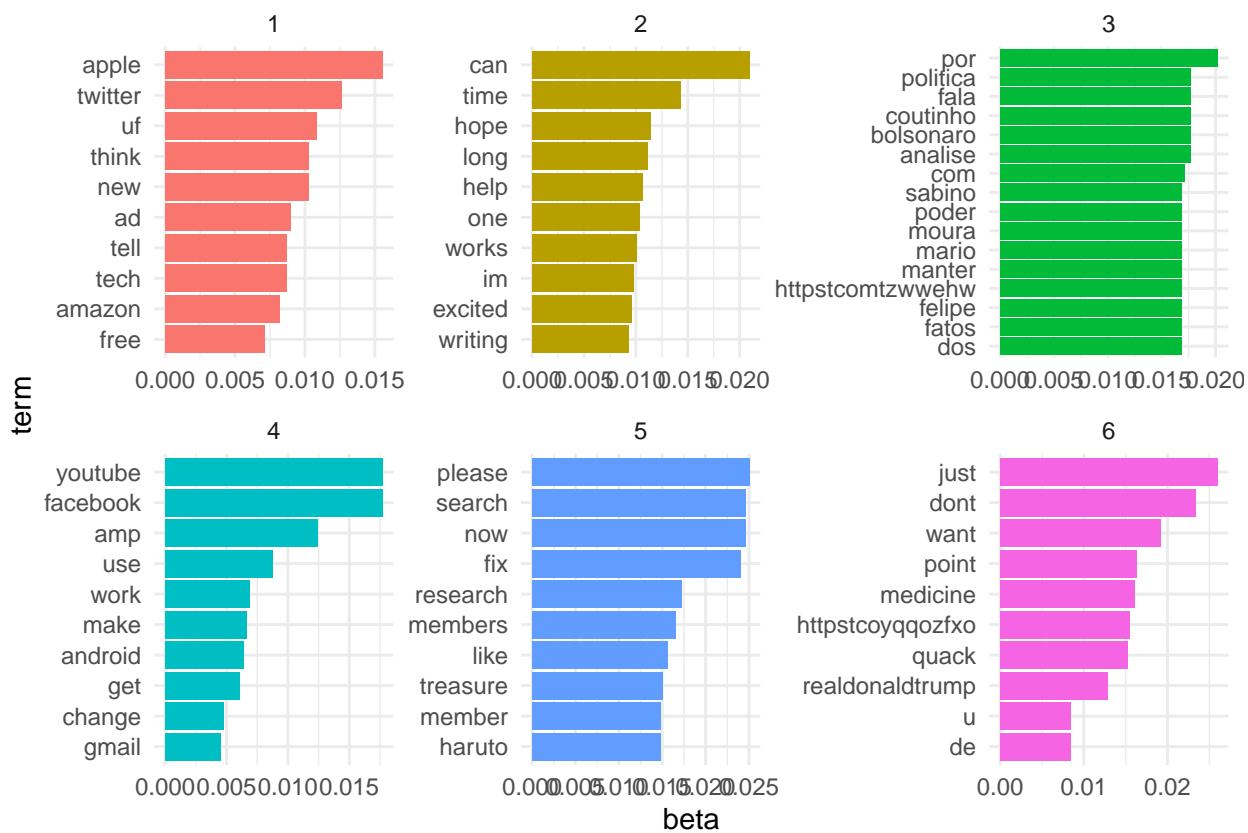


### 2.3. Topic Analysis:

The at google tweets are also divided into 6 different topics: - First topics is about brazilian politics as we have Bolsonero, politica, countinho as frequent words on this topic - Second topics on Donald trump politics toward CNN, Apple and twitter. Frequent words used in this topics are realdonaldtrump, apple, cnn - Third topics is on the medicine and most used words in this topics are medicine, UF(university of Florida) - Fourth topics is on Haruto, rapper of Treasure. Common words on this topics are Haruto, treasure, treasure members - Fifth is on help, technical assistance to google product users. Frequent words used are help, works, technical, engineers - Sixth and last topics is on the use of youtube on android phone. Frequent word in this topics are: youtube, use, android, work, facebook

Some of the topic trends of tweets with @google are general politics, pops and assistance request by Google product users. We can see that @google is used a lot to discuss political topics, which can be related to the influence of Google worldwide. Here we also see that topic 5 and 6 are about products and assistance. The @ is therefore used as a mix for various topics.

The graph bellow highlights the key words used for each topic mentioned above.



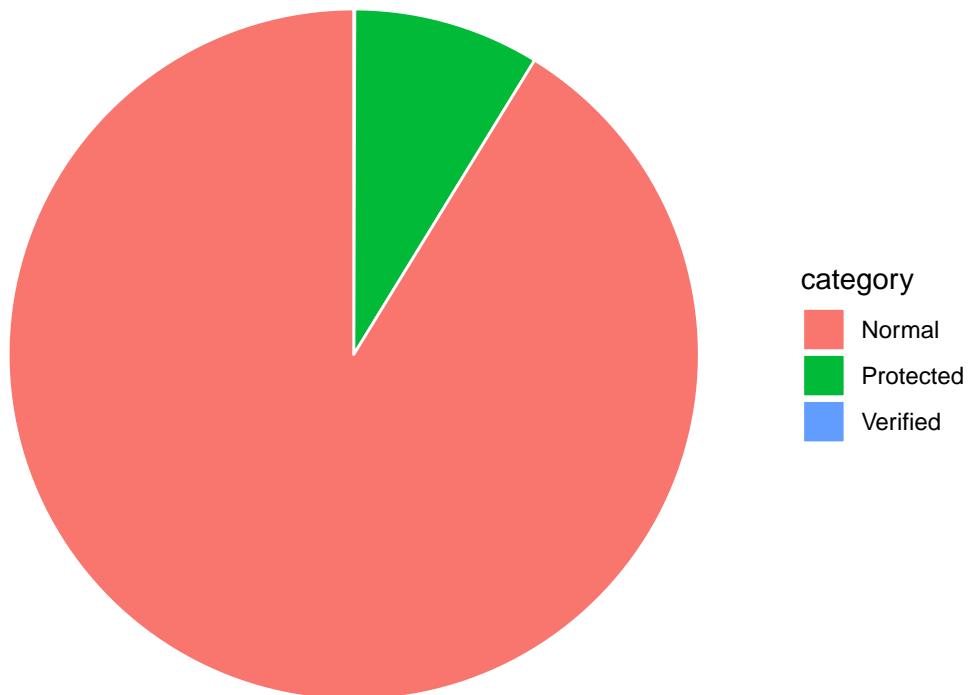
## PART 6 - Users Analysis:

In this part, we will analyze google followers and the accounts that Google are following that we will refer to, in this analysis, as friends of Google.

### 1. Followers Analysis:

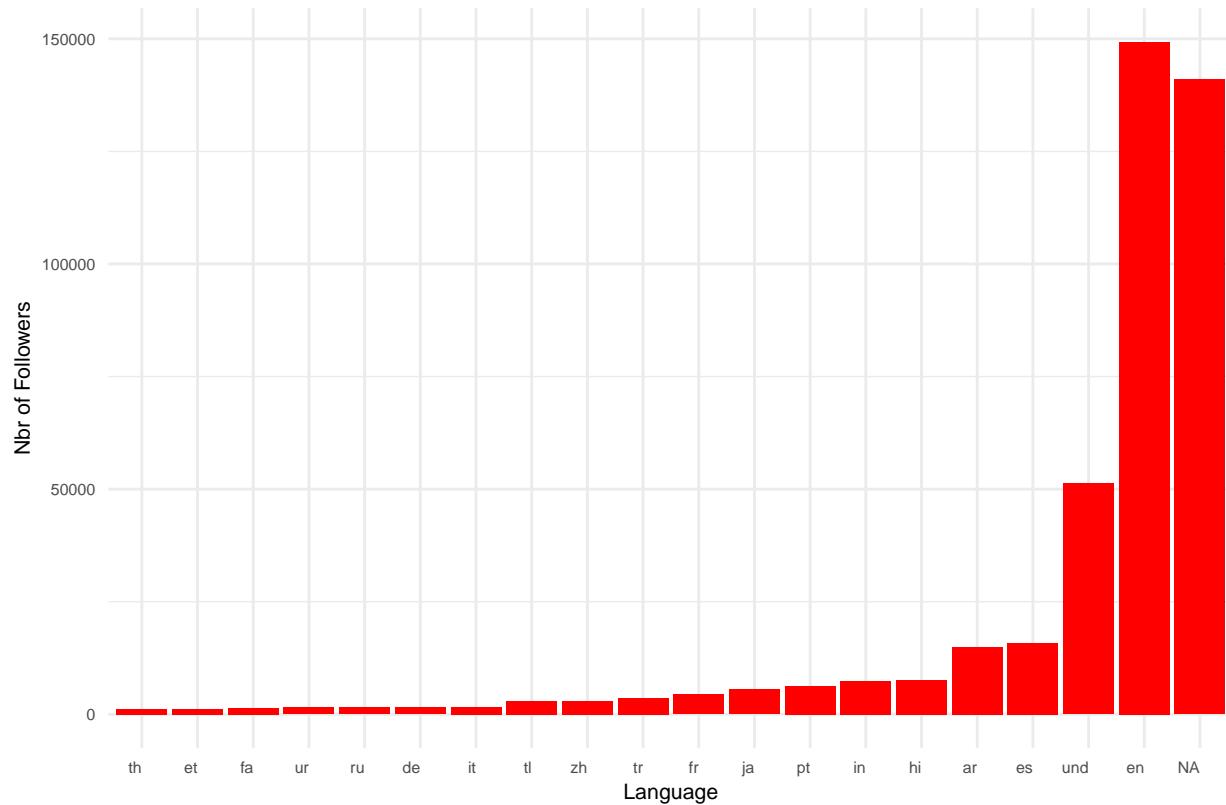
Starting with the followers of Google that are more than 22.6M Followers. It was impossible to download their complete list. For this analysis, the API randomly selected 434780 accounts. Amongst them, 201 were verified accounts. According to the help center of tweeter, ‘To receive the blue badge, your account must be authentic, notable, and active.’ 396670 of these accounts are regular accounts and 37909 are protected accounts.

The pie chart here after shows the distribution of these accounts into these categories:



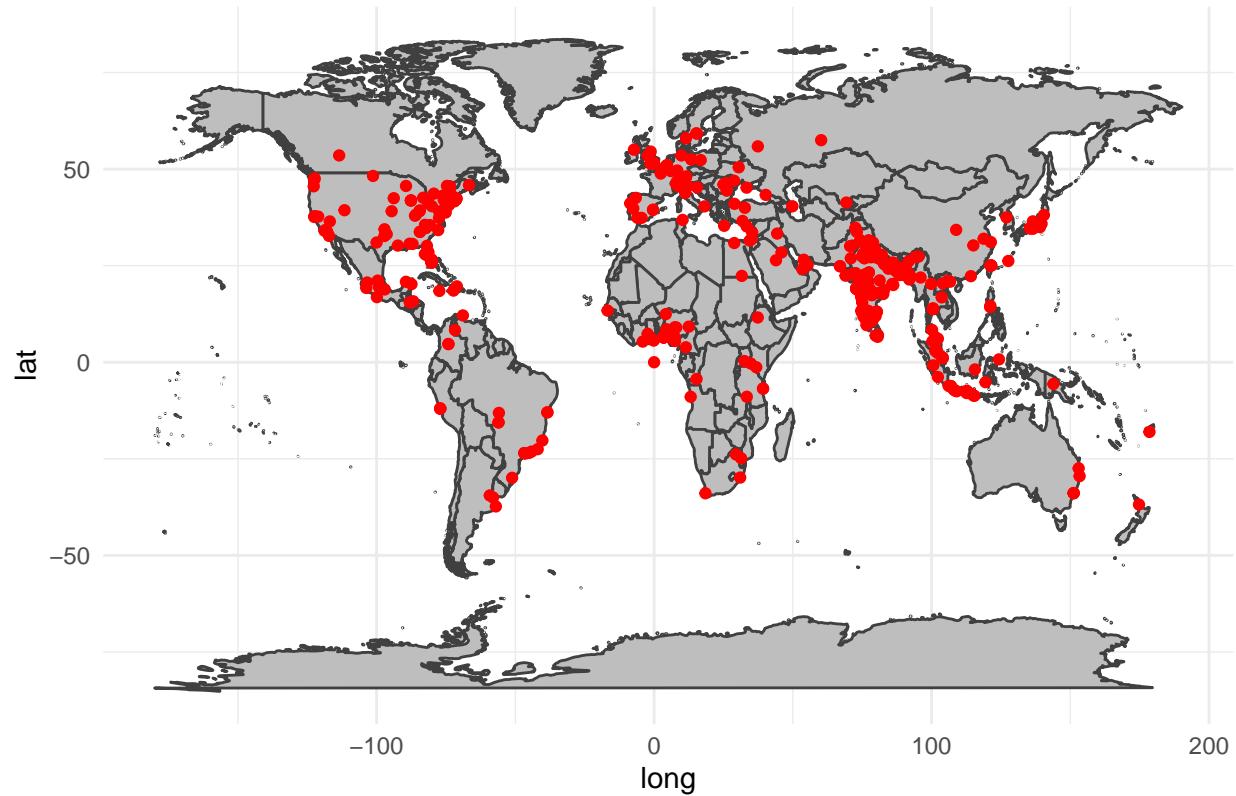
In this sample of accounts, the main language of users is english directly followed by accounts not providing their official languages (NA and Und). The bar chart below shows the top 20 languages of Google followers

### Top 20 Language Distribution



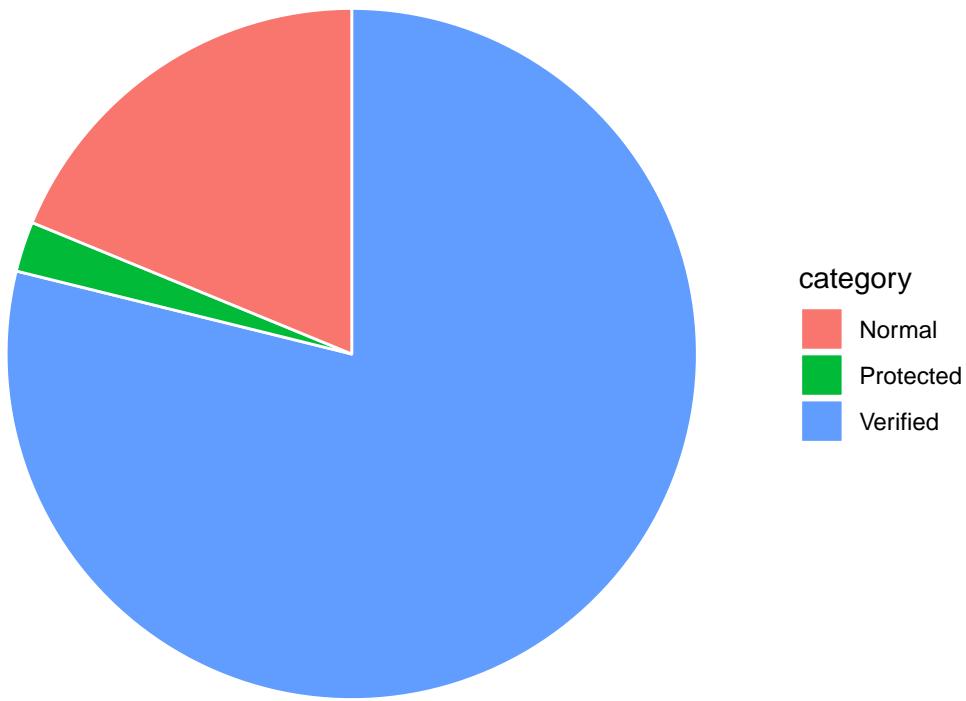
It was also important to see where these accounts are located but it was challenging since location is not an information that all tweeter accounts provide. Yet, we decided to conduct the analysis for the information available. As you can see in the map here after, followers are located worldwide but mainly in South East Asia and in North America. This can be explained by several assumptions: Google is known worldwide In these areas where followers are concentrated, people may use more Tweeter than in other locations. In fact, according to Statistic official website, the top 3 countries of tweeter users are USA, Japan and India. This explains our findings.

Google followers locations



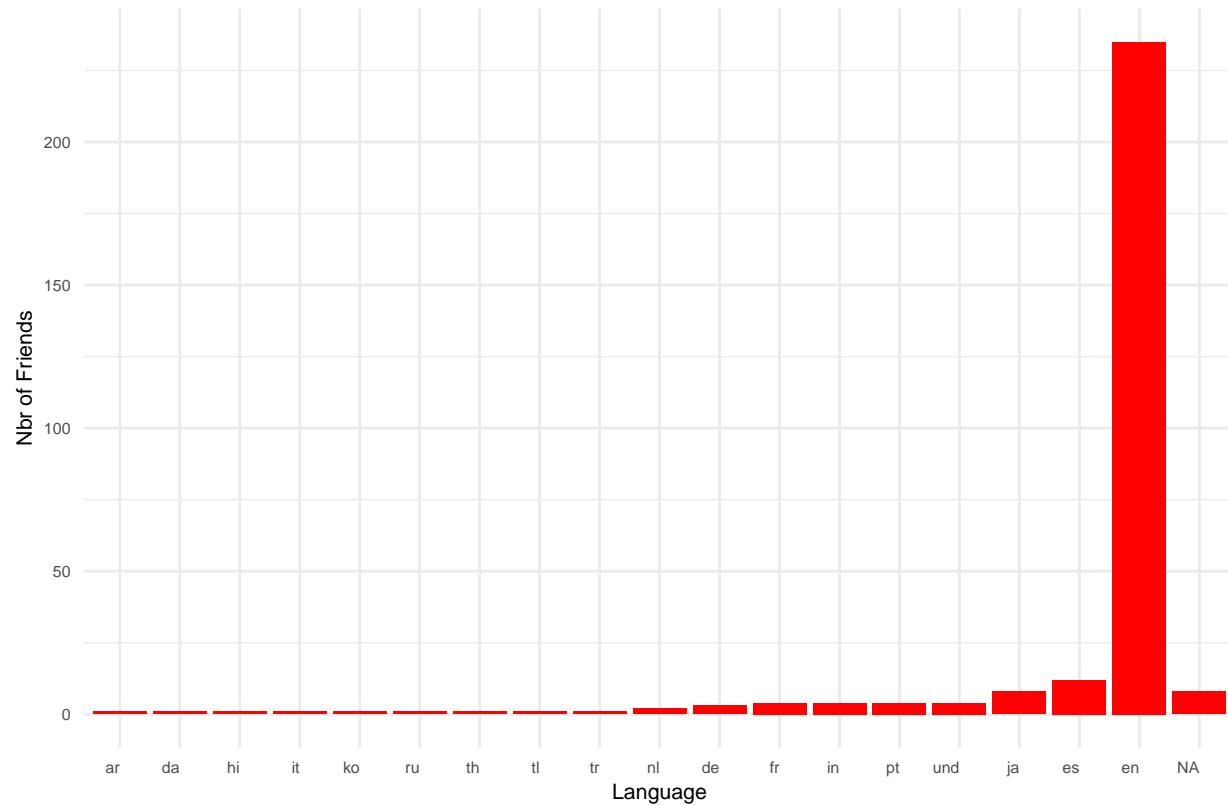
## 2. Friends Analysis:

Concerning the friends of Google, it was easier to conduct a comprehensive analysis since the API could easily extract the 293 accounts Google is following. Among these accounts, the majority are verified (235). Only 7 are protected and 61 were regular.



Concerning the languages of these accounts and as we can see in the graph below, English is the main language of account followed (majority of accounts). When investigating these accounts, we can see that a big majority are Google affiliated accounts or important people working for Google.

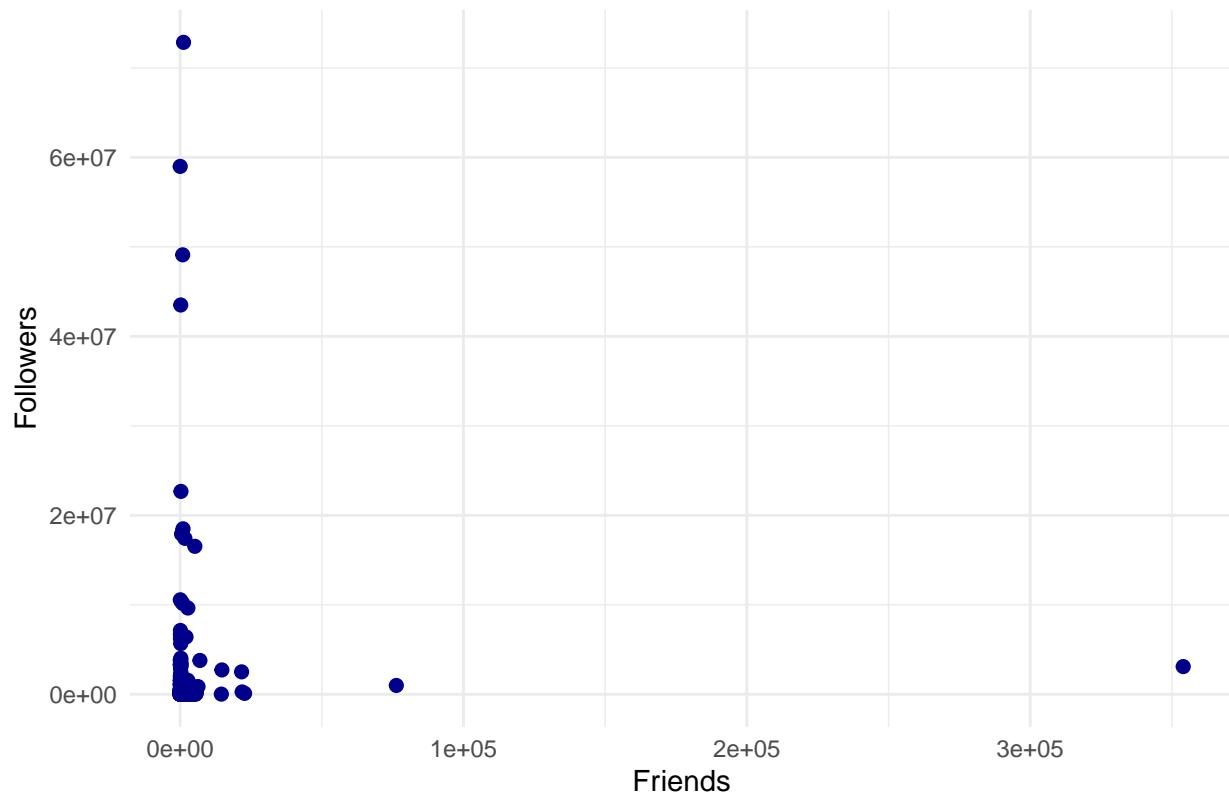
Top 20 Language Distribution



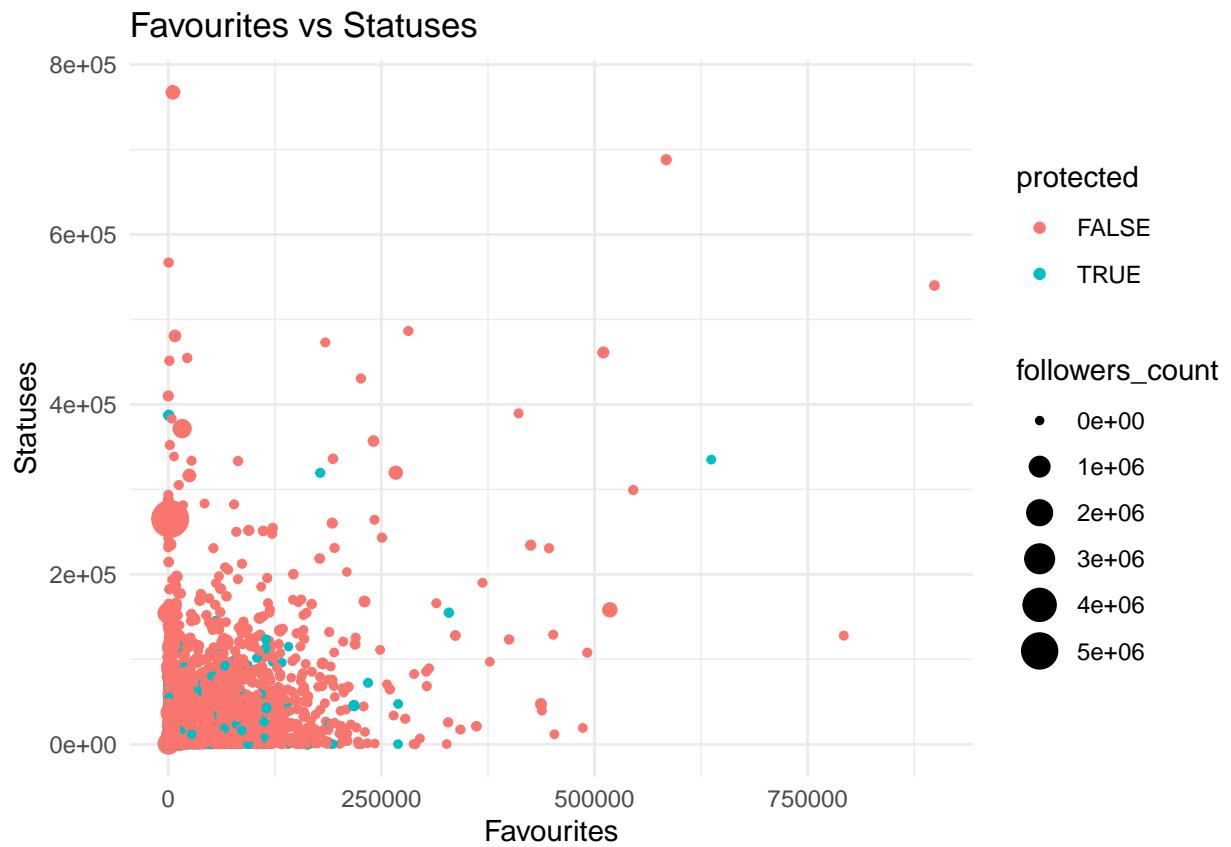
### 3. Friends vs Followers:

We also compared Google's friends and their followers. The scatter plot bellow highlights that even though Google has a limited amount of friends, these friends have a lot of followers. This is interesting as it means that when Google's friends share, like or comment a post from Google, all of their followers will also see it. Google's friends have more followers than Google's followers do. This can be a criteria for Google's choice in friends vs followers. It could also be explained by the amount of 'ordinary' users that follow Google, when the friends have to be specifically accepted by Google and are generally other big accounts.

## Followers vs Friends



The graph bellow illustrates that Google followers tweets a lot in comparison to Google's friends. Additionally, Google's followers gets more engagement on their posts than Google's friends do. This highlights that Google followers are important to spread their tweets, with every like, comment and so on.



These two plots highlight that Google friends are chosen and verified, whilst followers can be anyone. But both are very important for the Google Twitter account to share their posts to the largest amount of users possible.

## **Part 7 - Limitations:**

Despite the valuable content of this research, the team is aware of the limitations that are present due to several constraints and that are presented hereafter:

First, as the project used a tweeter standard API, limited data could be gathered. Using a premium or enterprise APIs could have improved the value of data extracted and thus the outcome of the analysis. Additionally, we limited ourselves with public data, but with additional APIs we could potentially gather additional data on users.

Second, the project only focused on tweets written in English, if we gathered all tweets used in multiple languages, the preprocessing would have been too complex. But this could be done in another project to gather more information.

Third, preprocessing could have been improved to delete for instance non-meaningful words from the word clouds. This was not done intentionally since it requires a heavy code and may delete important content when people for instance write without space such as the ‘searchengineoptimization’ that was frequent in tweets mentioning #google.

Additionally, regarding the sentiment analysis, we decided to use the dictionary look-up method, but the ‘bing’ dictionary is not always accurate. Indeed, we can see that the word “trump” is seen as positive, when in reality it most likely refers to President Trump which cannot be characterized as positive or negative without further information on the tweet or the political beliefs of the user. To be more accurate we could have used the Machine Learning method, but we decided to focus on other aspects of our analysis.

Finally, to make a comprehensive analysis of a giant company such as Google, it would also be important to analyze the other official accounts of the company naming for instance in our case ‘Google Play’ and ‘Gmail’ but since this analysis is made in an academic context and in a short time period, it was impossible to conduct a such in-depth analysis.

## **Conclusion:**

Throughout this project, we decided to focus not only on the tweets posted by Google but also the ones mentioning Google. This decision was taken to get a more accurate and comprehensive idea of how Google is seen on Twitter. Regarding the tweets posted from Google, we found that they are mostly replies, and use a positive vocabulary, as well as topics about help, service and direct messages. This suggest that Google uses Twitter as a communication tool to help its users.

Regarding the tweets mentioning Google we decided to split them in multiple categories of tweets and overall found that tweets mentioning Google are very different when they use @google, #google or just the word google. Their topics differ, and so do their sentiment, and engagement. Surprisingly, the tweets with @ are not the ones that get the most engagement even though they mention Google directly. We also found that Google is linked to many topics such as political, and social ones, not just high tech.

There is a lot more that could be done on this analysis as suggested in the limitations. Yet, throughout this project, we learned how a multinational Company like Google uses Twitter and how it is seen by the public.

This project was an opportunity for the team to conduct a real and comprehensive social media analytics project and apply concepts and theories discussed in class.

Overall, the project allowed the team members to:

- Use twitter developer account and call APIs using the RTweet package in R;
- Work with a big amount of data;
- Conduct an in-depth text mining project with pre-processing, topic analysis, and sentiment analysis;
- Improve skills related to data visualization using Ggplot package;
- Use Rmarkdown to generate a PDF report;
- Learn concepts related to social media communication and digital marketing;
- Provide constructive analysis on topics related to social media.