Twitter Sentiment Analysis

A Case Study for Mcdonalds and KFC

BAIT 508 BA1 2022W1 Business Analytics Programming

Group 11

Group Members:

Christopher Lin	xup65k6t@student.ubc.ca	Python Specialist	49213176
Kristy Yu	yuyiner.yy@gmail.com	Data Analyst	18105156
Sonal Bimbra	sonalbimbra01@gmail.com	Data Analyst	20160867

Introduction

When it comes to fast food, the top two options that jump into people's mind are McDonalds and KFC. McDonalds and KFC have become famous fast-food brands worldwide for many decades. They are the main competitors for each other. Each year, both will create different marketing campaigns to increase awareness of the brand and attract customers' attention.

For example, McDonalds is holding its "Coast to Coast Monopoly" game recently, and KFC is providing "free meals" to the flood victims in Pakistan. Such marketing campaigns will take a lot of effort and it is very important for the companies to understand customers' feelings about their brand and how well their marketing strategies perform.

In today's world, with the advance of the internet, people have many places to express their feelings and opinions. Twitter is one of the most popular social media platforms that allow people to share information and opinions effectively. In this project, we will collect tweets data using Twitter API for McDonalds and KFC, then use text analysis techniques and create word clouds to discover the trending topics between the companies. Further, we will also do a sentiment analysis to find out how people feel about them.

A. Keyword Selection and Data Collection

For collecting the tweets, we mainly used the streaming API, Tweepy and collaborated with the Python-based script provided by the instructor team. We retrieved 5000 tweets using the keyword "mcdonalds'' and 5000 tweets using keyword "kfc", setting the language to english only. The data was collected on a random day Oct 3, 2022, and it includes the tweets that were sent from Oct 1, 2022, to Oct 3, 2022.

The raw data contained a lot of noises such as stop words and URLs. For the preliminary analysis, we will analyze the dataset with and without the stop word and remove the stop word and URLs for the word cloud visualization. To remove the url addresses, we build a function to search whether the input is an url or not. The function takes string as input data structure and returns True or False. We use the function in the double for loops within the list comprehension technique to remove every url. The outer for loop is used to iterate every "text" column values, putting the output into the inner for loop. The inner for loop, on the other hand, splits the values from outer for loop, tests whether it is url, and concatenates non-url values together.(See reference code 1) To clean the words, we splitted the word and changed all the words into lowercase, and then selected the non-stop words into a list. (See reference code 2)

Author information of the collected tweets is also collected as well. Tweepy has a protection scheme that only allows to extract 300 authors' data every 15 minutes. In order to collect the data, we have used the sleep function in the time package. There are 4406 unique author id in Mcdonalds and 4396 author id in KFC. We have fetched 4390 unique and valid author information for Mcdondals and 4373 for KFC. There are a few authors whose ids are not found which might be due to the cancellation of the account. From the author data, we will be able to find the most influential author in these topics.

B. Preliminary Analysis

In the preliminary Analysis, we have explored the dataset from many different aspects, such as popular words, influential authors etc. We have divided our findings into different sub topics which help to understand the recent trends in Mcdonalds and KFC.

Top ten most popular words with and without stop words:

For McDonald, the ten most popular words with stopwords are:

mcdonalds, the, to, a, rt, and, in, i, is, you (graph shown in Figure 1)

Ten most popular words with stopwords are:

mcdonalds, @mcdonalds, like, get, make, chicken, happy, used, one, nuggets (graph shown in Figure 2)

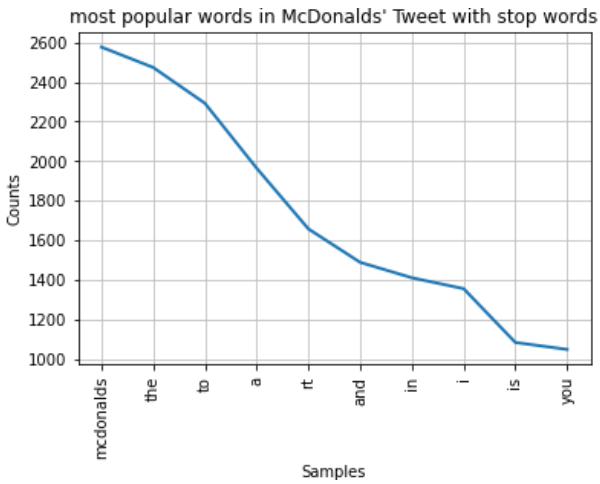

Figure 1

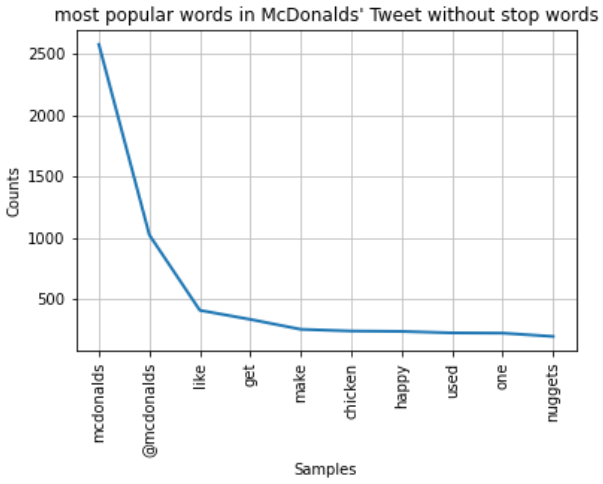

Figure 2

For KFC, the ten most popular words with stopwords are:

kfc, the, to, rt, i, and, a, in, for, of (graph shown in Figure 3)

Ten most popular words with stopwords are:

kfc, like, get, would, chicken, food, want, go, good, it's (graph shown in Figure 4)

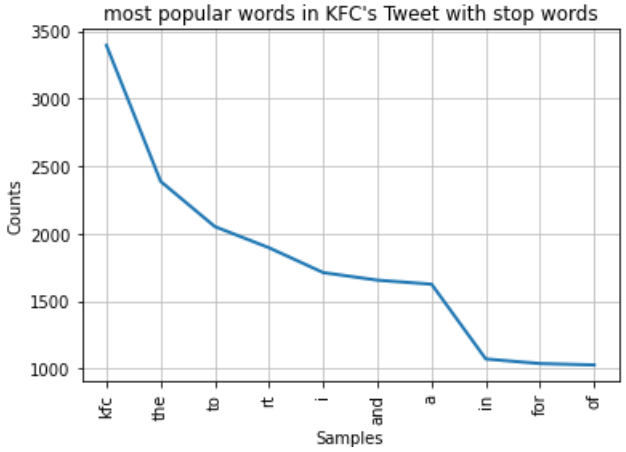

Figure 3

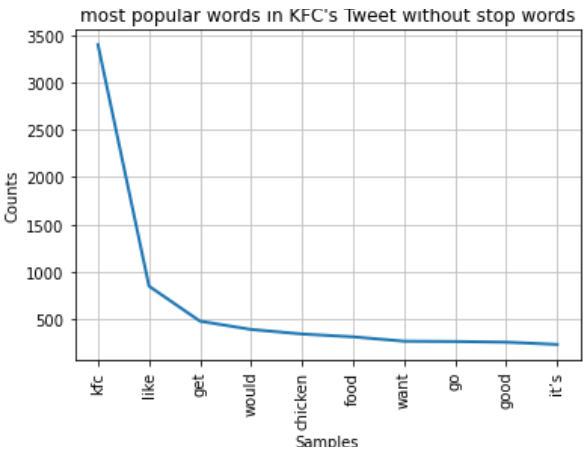

Figure 4

From Figure 1 and Figure 3, observations with stop words are meaningless so that stop word will be dropped for the remaining analysis. Figure 2 and Figure 4 captured many positive words such as “like”, “happy”, “good” which show the positive aspects of both brands.

Ten most popular hashtags (#hashtag):

The ten most popular hashtags and their corresponding counts for both McDonald and KFC are shown in the pictures below.

McDonald 10 most popular hashtags	KFC 10 most popular hashtags
#mcdonalds, 109	**#kfc, 150**
#dreamteamfanart, 49	**#gulfinmanila202…, 146**
#dreamfanart, 49	**#gulfinmanila2022, 72**
#georgenotfoundfanart, 49	**#princewilliam, 55**
#sapnapfanart, 49	**#princeofwales, 55**
#mccafemotivation, 44	**#pakvseng2022, 32**
#ymornings, 38	**#colonelsogbirthday, 29**
#rctid, 22	**#oc_tober2022, 24**
#mcdonaldschickensandwich, 21	**#food, 14**
#crypto, 17	**#jobs, 12**

From the popular hashtag, we are able to find the ongoing discussion topics in these brands.

- KFC has a popular hashtag **#pakvseng2022** because during the recent cricket match between Pakistan VS England, KFC on each wicket by Pakistan, pledged to donate Rs 50,000/- to the flood hit victims in Pakistan.
- **#georgenotfoundfanart** is a popular Youtuber which went viral recently, so people have been tagging him too in the Mcdonalds related tweets.
- Some users love the McCafe that there is a popular hashtag **#mccafemotivation**.

Ten most frequently mentioned usernames (@username):

The ten most frequently mentioned usernames and their usage frequency for both McDonalds and KFC are shown in the pictures below.

McDonalds 10 most frequently mentioned usernames

@mcdonalds, 1022

@thehealthb0t, 162

@bonisile_rms, 142

@marcellacomedy, 92

@mcdonaldscanada, 66

@jeremycom, 54

@actingthegom, 54

@d04_exe, 49

@mcdonalds_sa, 42

@yfm, 36

KFC 10 most frequently mentioned usernames

@kfc_es, 201

@kfc, 192

@kfc_uki, 156

@98_josua, 146

@phantomarcade3k, 79

@gulfkeepersph, 77

@headofmeduzaa, 66

@happeehour, 65

@ash_candygulf, 63

@kfc_ghana, 58

The McDonalds username was mentioned the most, probably because they host a lot of marketing campaigns for people to mention them on Twitter.

Three most common sources of the tweets:

The three most common sources of both McDonald and KFC are the same:

- Twitter for iPhone
- Twitter for Android
- Twitter Web App

Their counts for both Mcdonalds and KFC:

McDonald 3 most common sources		KFC 3 most common sources	
Twitter for iPhone	**2145**	**Twitter for iPhone**	**1936**
Twitter for Android	**1307**	**Twitter for Android**	**1913**
Twitter Web App	**902**	**Twitter Web App**	**796**

Time trend of tweet counts:

As we have collected the latest tweets, below are the distribution of time for tweets of both KFC and Mcdonalds, we see that users have been regularly posting about these brands. From the distribution plots, we also can find that there are more people who have been tweeting about Mcdonalds these days.

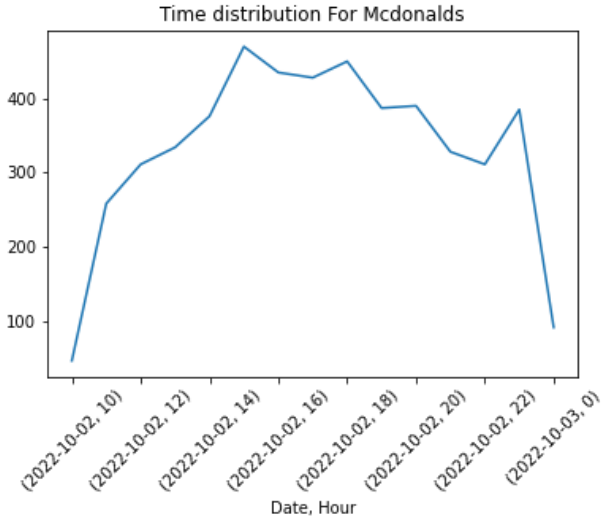

Figure 5

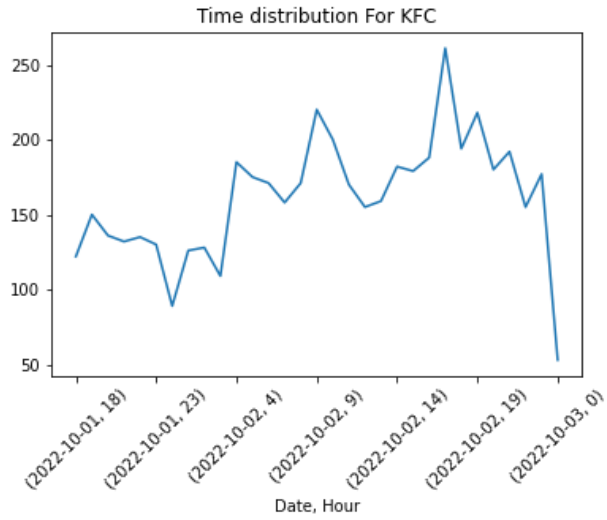

Figure 6

Three most influential tweets:

We have scored the tweets' influence based on "quote_count", "reply_count", "retweet_count", "like_count". Below tweets have the top attention:

McDonald 3 most influential tweets:

RT @McDonalds: do u need a ride to the FFVII 25th anniversary party

RT @yosupjt: yo who tf installed doom on the mcdonalds kiosk 😭

RT @stanzipotenza: how good your McDonalds tastes depends on the state of your mental health

KFC 3 most influential tweets:

RT @LilNasX: your employee just slapped me, called me a bottom, and refused to give me my change @kfc

RT @iNabber69: imagine having a family member brutally murdered, you have 10 years of therapy to cope and then one day you go onto youtube…

RT @DailyLoud: They said he got an interview at KFC today 👀🤣

Three most vocal authors/influencers:

We found the top most vocal authors for both Mcdonalds and KFC, listed below:

For Mcdonalds and their tweet counts recently:

1) Mcdonald's : 144
2) Coommanman : 29
3) mcdsnackwrap : 12

For KFC and their tweet counts:

1) KFC South Africa : 16
2) StuffBot : 13
3) Totes3d : 12

- Mcdonald's twitter user seems to be a very active user and has tweeted about 144 times about the Mcdonald's brand in a day whereas there is no such a vocal author for the KFC brand.
- KFC in South Africa has recently launched new product variations and hence people are excited about them and have been actively posting about it.

Three most influential authors

We have scored the author's influence by its "followers_count", "following_count", "listed_count". "tweet_count".

Top most influential authors for Mcdonalds are:

1) McDonald's
2) PulpNews Crime
3) Major League Soccer

Top most influential authors for KFC are:

1) BuzzFeed
2) GPOSERS_FFXIV
3) AnimatorsPal

- Buzzfeed, Pulpnews is a popular news and entertainment channel which justifies their name being here in the list.
- Mcdonalds twitter page too is quite influential as their marketing campaigns have been quite famous.

C. Word Cloud Analysis

Before creating the word cloud, we add the Lemmatization method to make every word transform back into their base or dictionary form. For example, we can use the method to transform the words from "cats" into "cat". By doing so, we can reduce the repeated words showing up in our word cloud graph. (See reference code 3)

There are various words related to the brands in each of the below word clouds.

Classification of words seen in the word cloud:

1) Products they sell. For example: Happy meal in Mcdonalds, Chicken in KFC
2) Locations of the outlets. For instance: KFC in Thailand
3) Taglines: 'I'm lovin it' for Mcdonalds and 'Finger lickin good' for KFC
4) Other competitive brands: Mcdonalds in KFC
5) About the brand: Generic words like fast food, eat, good and so on.

There are a few negative words for both the brands which are more valuable to the stakeholders as they would help them identify improvement opportunity areas. For instance, word cloud of Mcdonalds has some negative words such as: “synthetic material” which could be a customer review the Mcdonald's team should address. On the contrary, there are some positive words too such as ‘Happy meal' which shows that the happy meal is actually loved by the customers. Similarly KFC word cloud has some positive words as well such as their advertising tagline ‘Finger Licking good’ . The word ‘Filipino’ is used quite a lot in KFC tweets as KFC Filipino rice chicken combo gained a lot of attention and people really liked it. Hence, these are some observations that would help the stakeholders of these companies to look into what people are actually saying about their brand.

Figure 7 McDonalds Word Cloud

Figure 8 KFC Word Cloud

D. Sentiment Analysis

Average polarity and subjectivity scores:

Brand	Mcdonalds	KFC
Average Subjectivity Score	0.360	0.368
Average Polarity Score	0.082	0.027

Polarity and Subjectivity score distributions:

We got the sentimental distributions for Mcdonalds and KFC in our sentimental analysis. The subjectivity score of McDonald (Figure 5) and KFC (Figure 6) is approximately 0.3 which shows that the majority of people tweet facts about the products and the brand rather than their personal opinions. However, there are some people who do post their personal opinions about these brands and that is more for McDonalds than KFC. Now these sentiments could be negative or positive, which we will evaluate further in this case study by checking the polarity of the sentiments.

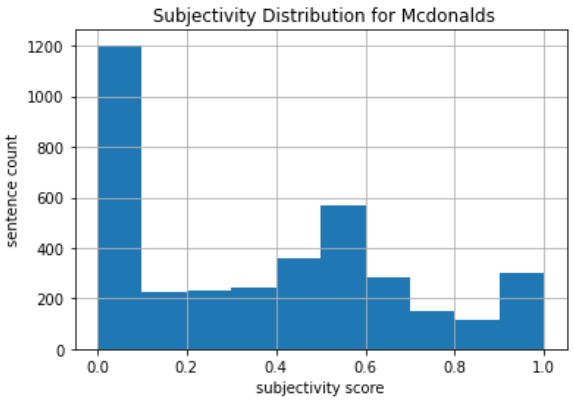

Figure 9

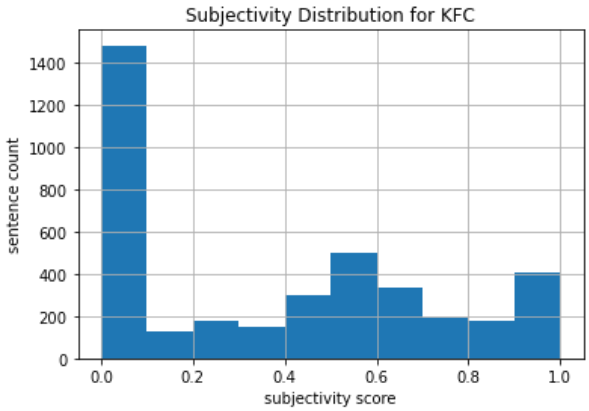

Figure 10

The average polarity score for Mcdonalds is 0.082 (score distribution shown in Figure 7) and KFC has an average polarity of 0.027(score distribution shown in Figure 8) , which shows that most people have a neutral sentiment about these two brands. However, more people have positive reviews about McDonalds compared to KFC. There could be various reasons why people have more positive polarity towards Mcdonalds, some are:

1) Better taste
2) Better customer satisfaction
3) Better Quality
4) Better food choices

But the difference in polarity is very small because of which we will need more customer reviews to evaluate on which brand performance is better.

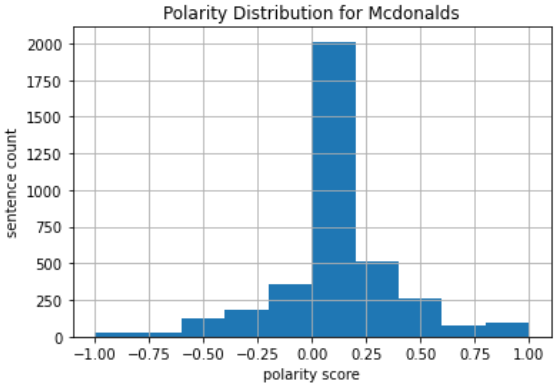

Figure 11

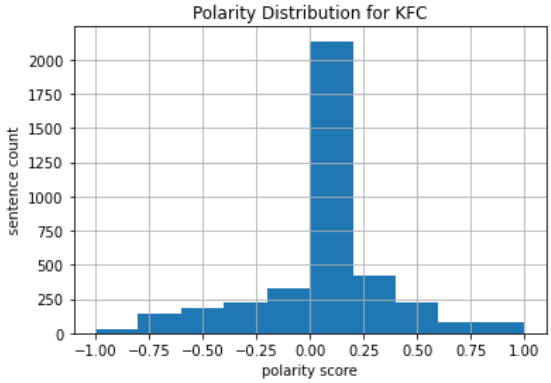

Figure 12

Most positive and negative tweets on the keyword:

Three most positive tweets about Mcdonalds and KFC:

Mcdonald's:

"woman sues mcdonald's for $13 million after accusing staff of serving coffee containing chemicals • hollywood unlocked and it was awesome!"

"hit us up with more info on this here: rt @angrycrazyzebra: lets goooo!!!!"

"rt @casmz5: 🚨🚨new video🚨🚨 👇👇👇👇👇👇👇 🔗- 🔗- the best two things in a frappe."

KFC:

"@chikn_kfc good times!!"

"it was delicious."

"good morning!!"

Positive tweets analysis:

1) People are happy with the taste and quality of food being served
2) Users love the Mcdonalds coffee
3) There is one post where the user has mentioned about a recent negative experience another person had and believes that it is a fake stunt. This user comes out in support of the brand and a loyal Mcdonald customer.

Three most negative tweets about Mcdonalds and KFC:

Mcdonald's:

"absolutely disgusting!!"

"don't know why but it's terrible that it does that."

"— ali francis (@apwf… rt @femaieboyfriend: got she'd at mcdonalds but then the cashier said 'oh my bad' rt @ianlovesfilm: bruh i came to mcdonalds specifically for the cars toys and i got lebron james?.!.?!:!;?"

KFC:

"subway is nasty."

"starbucks coffee tastes nasty af."

"you know that if i did order it, it would be disgusting."

Negative tweet analysis:

1) Some people do not like the food being served at these restaurants.
2) Certain tweets show low customer satisfaction. For instance: A customer is unhappy about the unavailability of Mcdonald happy meal toys.
3) There are some tweets where users are actually posting negative comments about other brands but mentioning McDonalds or KFC in those posts, probably to gain more viewers.

E. Insights

The world has changed a lot with the development of the technologies, but people still love McDonalds and KFC. The McDonalds word clouds show that people are interested in its happy meal and chicken nuggets. And KFC is known for its chicken. The word clouds also show more positive words than negative words in both McDonalds and KFC, which indicate people's positive attitude toward the brands. The attitude of the individuals can also be analyzed by sentiment analysis. The subjectivity distribution shows that on Twitter, most people prefer to tweet facts rather than express their personal feelings. But for those who do post their personal opinions, McDonalds' positive polarity (8.3%) is slightly higher than KFC's positive polarity (2.7%). It is very significant to know people's attitude or opinions toward the brands as it can help to determine companies' marketing strategies. The difference between McDonalds and KFC is quite small in this analysis, a larger data set will be needed to further analyze the difference in the future.

Other than tweets, there are many other pieces of data that are available to discover nowadays. One of the interesting ones can be customer reviews and transactions in E-commerce platforms. People's shopping habits are changing in recent years. Rather than shopping in the malls, people now prefer to shop online as it is cheaper and more options to choose from. Many merchants or individuals see this as a great opportunity to earn money. A lot of new E-commerce stores are opening in different shopping platforms such as Amazon, eBay and Shopee. Most of the time operators want to know what the recent trending products are and how people feel like it. As the trends are changing rapidly, if one catches these trends and launches products accordingly, they can take over most of the market shares. It would be interesting to have a project on developing tools to capture the customer transactions and reviews on a daily or weekly basis and analyze the trends and people's preferences. The data-driven insight can help the E-commerce operators to discover more market opportunities.

Reference code

1. Dealing with url

```
# Remove 'https://' keyword from the tweets as it was being counted as the
# top used words in the tweets and had no importance in our analysis

from urllib.parse import urlparse
#https://stackoverflow.com/a/52455972
def is_url(url):
  ''' Input string to test whether it is url (i.e. starting from https://)
  '''
  try:
    result = urlparse(url)
    return all([result.scheme, result.netloc])
  except ValueError:
    return False

df_tweets_mcd['text'] = [' '.join(y for y in x.split() if not is_url(y)) for x in df_tweets_mcd['text']] # double for loop in list comprehension
df_tweets_kfc['text'] = [' '.join(y for y in x.split() if not is_url(y)) for x in df_tweets_kfc['text']]
# Outer for loop to iterate each 'text' value ; inner for loop to split the value from sectence to words

# this cell is revised from the website below
# https://stackoverflow.com/questions/56358888/how-to-remove-https-links-from-a-string-column-in-pandas
```

0.7s

2. Dealing with letter case

```
words_mcd = df_tweets_mcd['text'].apply(str.lower).apply(str.split).sum() # change every text value to lower case, and split them from sentence to words
words_kfc = df_tweets_kfc['text'].apply(str.lower).apply(str.split).sum()
# output is the list of words
```

3.8s

```
words_mcd_woStopWords = [w for w in words_mcd if w not in stopwords and len(w) > 1] # list comprehension to include non-stop words
words_kfc_woStopWords = [w for w in words_kfc if w not in stopwords and len(w) > 1]
print(len(words_mcd_woStopWords))
print(len(words_kfc_woStopWords))
```

0.4s

```
55915
57125
```

3. Lemmatizing

```
from nltk.stem import WordNetLemmatizer
wnl = WordNetLemmatizer()
text_mcd = '' # our string accumulator
for word in words_mcd_woStopWords:
    if len(word) == 1 or word in stopwords:
        continue
    text_mcd = text_mcd + ' ' + wnl.lemmatize(word) #lemmatize for returning the base or dictionary form of a word

text_kfc = '' # our string accumulator
for word in words_kfc_woStopWords:
    if len(word) == 1 or word in stopwords:
        continue
    text_kfc = text_kfc + ' ' + wnl.lemmatize(word)
```

10.2s Python