

Twitter_TFM

December 10, 2021

```
[1]: !pip install tweepy
      !pip install textblob
      !pip install pycountry
      !pip install langdetect
      !pip install pandas
      !pip install sklearn
```

```
Requirement already satisfied: tweepy in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (4.4.0)
Requirement already satisfied: requests-oauthlib<2,>=1.0.0 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from tweepy) (1.3.0)
Requirement already satisfied: requests<3,>=2.11.1 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from tweepy) (2.23.0)
Requirement already satisfied: oauthlib>=3.0.0 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from requests-
oauthlib<2,>=1.0.0->tweepy) (3.1.1)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from
requests<3,>=2.11.1->tweepy) (1.25.8)
Requirement already satisfied: chardet<4,>=3.0.2 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from
requests<3,>=2.11.1->tweepy) (3.0.4)
Requirement already satisfied: idna<3,>=2.5 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from
requests<3,>=2.11.1->tweepy) (2.9)
Requirement already satisfied: certifi>=2017.4.17 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from
requests<3,>=2.11.1->tweepy) (2021.10.8)
Requirement already satisfied: textblob in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (0.17.1)
Requirement already satisfied: nltk>=3.1; python_version >= "3" in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from textblob) (3.6.5)
Requirement already satisfied: click in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from nltk>=3.1;
python_version >= "3"->textblob) (8.0.3)
Requirement already satisfied: joblib in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from nltk>=3.1;
python_version >= "3"->textblob) (1.1.0)
```

```

Requirement already satisfied: regex>=2021.8.3 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from nltk>=3.1;
python_version >= "3"->textblob) (2021.11.10)
Requirement already satisfied: tqdm in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from nltk>=3.1;
python_version >= "3"->textblob) (4.46.0)
Requirement already satisfied: pycountry in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (20.7.3)
Requirement already satisfied: langdetect in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (1.0.9)
Requirement already satisfied: six in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from langdetect)
(1.14.0)
Requirement already satisfied: pandas in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (1.3.4)
Requirement already satisfied: numpy>=1.17.3; platform_machine != "aarch64" and
platform_machine != "arm64" and python_version < "3.10" in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from pandas) (1.21.4)
Requirement already satisfied: python-dateutil>=2.7.3 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from pandas) (2021.3)
Requirement already satisfied: six>=1.5 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from python-
dateutil>=2.7.3->pandas) (1.14.0)
Requirement already satisfied: sklearn in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (0.0)
Requirement already satisfied: scikit-learn in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from sklearn) (1.0.1)
Requirement already satisfied: numpy>=1.14.6 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from scikit-
learn->sklearn) (1.21.4)
Requirement already satisfied: joblib>=0.11 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from scikit-
learn->sklearn) (1.1.0)
Requirement already satisfied: scipy>=1.1.0 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from scikit-
learn->sklearn) (1.7.3)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/Users/anadeory/miniconda3/lib/python3.8/site-packages (from scikit-
learn->sklearn) (3.0.0)

```

```
[ ]: conda install -c conda-forge wordcloud=1.6.0
```

```
[33]: # Import Libraries

from textblob import TextBlob
```

```

import sys
import tweepy
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import os
import nltk
nltk.download('vader_lexicon')
nltk.download('stopwords')
import pycountry
import re
import string

from wordcloud import WordCloud, STOPWORDS
from PIL import Image
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from langdetect import detect
from nltk.stem import SnowballStemmer
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from sklearn.feature_extraction.text import CountVectorizer

```

```

[nltk_data] Downloading package vader_lexicon to
[nltk_data]    /Users/anadeory/nltk_data...
[nltk_data] Package vader_lexicon is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]    /Users/anadeory/nltk_data...
[nltk_data] Package stopwords is already up-to-date!

```

```

[3]: # Authentication
consumerKey = "a0cLYyjRSHFcXcTTU7RvYGgHh"
consumerSecret = "uQVZt6LIcgpqrqJ3iAjDpehFufKgAfqYzPpbreIEHUcN7OrXPw"
accessToken = "1461438773078769666-uAYj5L4fiB5lzaOT5fj1VtW5gdckie"
accessTokenSecret = "UBGFL3v5kFJZm8VTXMoecGQ2eerYteQ3rRa4ia7vVbX2b"

auth = tweepy.OAuthHandler(consumerKey, consumerSecret)
auth.set_access_token(accessToken, accessTokenSecret)
api = tweepy.API(auth)

```

```

[5]: #Sentiment Analysis

def percentage(part,whole):
    return 100 * float(part)/float(whole)

keyword = input("Please enter keyword or hashtag to search: ")
noOfTweet = int(input("Please enter how many tweets to analyze: "))

```

```

tweets = tweepy.Cursor(api.search_tweets, q=keyword).items(noOfTweet)
positive = 0
negative = 0
neutral = 0
polarity = 0
tweet_list = []
neutral_list = []
negative_list = []
positive_list = []

for tweet in tweets:

    #print(tweet.text)
    tweet_list.append(tweet.text)
    analysis = TextBlob(tweet.text)
    score = SentimentIntensityAnalyzer().polarity_scores(tweet.text)
    neg = score['neg']
    neu = score['neu']
    pos = score['pos']
    comp = score['compound']
    polarity += analysis.sentiment.polarity

    if neg > pos:
        negative_list.append(tweet.text)
        negative += 1

    elif pos > neg:
        positive_list.append(tweet.text)
        positive += 1

    elif pos == neg:
        neutral_list.append(tweet.text)
        neutral += 1

positive = percentage(positive, noOfTweet)
negative = percentage(negative, noOfTweet)
neutral = percentage(neutral, noOfTweet)
polarity = percentage(polarity, noOfTweet)
positive = format(positive, '.1f')
negative = format(negative, '.1f')
neutral = format(neutral, '.1f')

```

Please enter keyword or hashtag to search: bitcoin
Please enter how many tweets to analyze: 500

```

[7]: #Number of Tweets (Total, Positive, Negative, Neutral)
tweet_list = pd.DataFrame(tweet_list)

```

```

neutral_list = pd.DataFrame(neutral_list)
negative_list = pd.DataFrame(negative_list)
positive_list = pd.DataFrame(positive_list)
print("total number: ",len(tweet_list))
print("positive number: ",len(positive_list))
print("negative number: ", len(negative_list))
print("neutral number: ",len(neutral_list))

```

```

total number: 500
positive number: 160
negative number: 64
neutral number: 276

```

```
[8]: tweet_list
```

```

[8]:
0      RT @EMPOWR_us: Empower Oversight Sues SEC For ...
1      RT @MinusWells: Make it or break it week... Wh...
2      RT @CoinCornerMolly: Paying for a coffee with ...
3      RT @HPOS10INU: Coinmarketcap has announced it ...
4      RT @CryptoYodaGuru: En lo que esperamos la sub...
..
495    RT @saylor: #Bitcoin on #Lightning is the fu...
496                                @fckboislayer bitcoin?
497    RT @Crypto_Woof: Our 3rd Giveaway Is Live, Thi...
498    RT @ODELL: bitcoin will be perceived as a risk...
499    RT @Willbrandblock: I made my biggest gains by...

[500 rows x 1 columns]

```

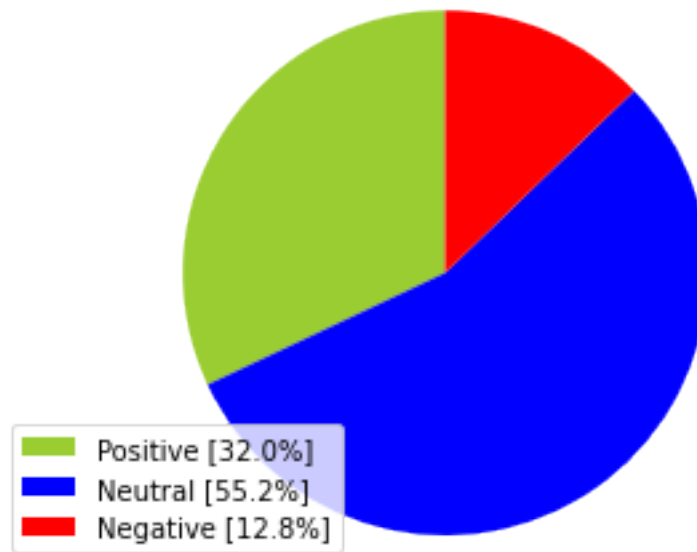
```
[9]: #Creating PieCart
```

```

labels = ['Positive ['+str(positive)+'%]' , 'Neutral_
↪['+str(neutral)+'%'],'Negative ['+str(negative)+'%]']
sizes = [positive, neutral, negative]
colors = ['yellowgreen', 'blue','red']
patches, texts = plt.pie(sizes,colors=colors, startangle=90)
plt.style.use('default')
plt.legend(labels)
plt.title("Sentiment Analysis Result for keyword= "+keyword+" ")
plt.axis('equal')
plt.show()

```

Sentiment Analysis Result for keyword= bitcoin



```
[10]: tweet_list.drop_duplicates(inplace = True)
```

```
[11]: tw_list = pd.DataFrame(tweet_list)
tw_list["text"] = tw_list[0]
tw_list
```

```
[11]:
```

	0 \	text
0	RT @EMPOWR_us: Empower Oversight Sues SEC For ...	RT @EMPOWR_us: Empower Oversight Sues SEC For ...
1	RT @MinusWells: Make it or break it week... Wh...	RT @MinusWells: Make it or break it week... Wh...
2	RT @CoinCornerMolly: Paying for a coffee with ...	RT @CoinCornerMolly: Paying for a coffee with ...
3	RT @HPOS10INU: Coinmarketcap has announced it ...	RT @HPOS10INU: Coinmarketcap has announced it ...
4	RT @CryptoYodaGuru: En lo que esperamos la sub...	RT @CryptoYodaGuru: En lo que esperamos la sub...
..
494	@100trillionUSD @Birdson64827145 How much more...	@100trillionUSD @Birdson64827145 How much more...
495	RT @saylor: #Bitcoin on #Lightning is the fu...	RT @saylor: #Bitcoin on #Lightning is the fu...
496	@fckboislayer bitcoin?	@fckboislayer bitcoin?
498	RT @ODELL: bitcoin will be perceived as a risk...	RT @ODELL: bitcoin will be perceived as a risk...
499	RT @Willbrandblock: I made my biggest gains by...	RT @Willbrandblock: I made my biggest gains by...

```

494 @100trillionUSD @Birdson64827145 How much more...
495 RT @saylor: #Bitcoin on #Lightning is the fu...
496 @fckboislayer bitcoin?
498 RT @ODELL: bitcoin will be perceived as a risk...
499 RT @Willbrandblock: I made my biggest gains by...

```

[397 rows x 2 columns]

[12]: tweet_list

[12]: 0 \

```

0 RT @EMPOWR_us: Empower Oversight Sues SEC For ...
1 RT @MinusWells: Make it or break it week... Wh...
2 RT @CoinCornerMolly: Paying for a coffee with ...
3 RT @HPOS10INU: Coinmarketcap has announced it ...
4 RT @CryptoYodaGuru: En lo que esperamos la sub...
..
494 @100trillionUSD @Birdson64827145 How much more...
495 RT @saylor: #Bitcoin on #Lightning is the fu...
496 @fckboislayer bitcoin?
498 RT @ODELL: bitcoin will be perceived as a risk...
499 RT @Willbrandblock: I made my biggest gains by...

```

text

```

0 RT @EMPOWR_us: Empower Oversight Sues SEC For ...
1 RT @MinusWells: Make it or break it week... Wh...
2 RT @CoinCornerMolly: Paying for a coffee with ...
3 RT @HPOS10INU: Coinmarketcap has announced it ...
4 RT @CryptoYodaGuru: En lo que esperamos la sub...
..
494 @100trillionUSD @Birdson64827145 How much more...
495 RT @saylor: #Bitcoin on #Lightning is the fu...
496 @fckboislayer bitcoin?
498 RT @ODELL: bitcoin will be perceived as a risk...
499 RT @Willbrandblock: I made my biggest gains by...

```

[397 rows x 2 columns]

[13]: *#Cleaning Text (RT, Punctuation etc)*

#Creating new dataframe and new features

```
tw_list = pd.DataFrame(tweet_list)
```

```
tw_list["text"] = tw_list[0]
```

#Removing RT, Punctuation etc

```
remove_rt = lambda x: re.sub('RT @\w+: ', "", x)
```

```
rt = lambda x: re.sub("(@[A-Za-z0-9]+)|([\^0-9A-Za-z \t])|(\w+:\/\/\S+)", "", x)
```

```
tw_list["text"] = tw_list.text.map(remove_rt).map(rt)
tw_list["text"] = tw_list.text.str.lower()
tw_list.head(10)
```

[13]: 0 \

```
0 RT @EMPOWR_us: Empower Oversight Sues SEC For ...
1 RT @MinusWells: Make it or break it week... Wh...
2 RT @CoinCornerMolly: Paying for a coffee with ...
3 RT @HPOS10INU: Coinmarketcap has announced it ...
4 RT @CryptoYodaGuru: En lo que esperamos la sub...
5 RT @Rare_Metaverse: Rocket Arena Airdrop is li...
6 What if, for the first time this entire cycle,...
7 When in doubt, Zoom _____. #Bitcoin https:/...
8 RT @javierbitcoin: Lo que se avecina, es una c...
9 RT @WhyBitcoinJ: Gratz if you grabbed some of ...

                                text
0 empower oversight sues sec for access to foia...
1 make it or break it week what do you fine ...
2 paying for a coffee with bitcoin lightning n...
3 coinmarketcap has announced it will be sponso...
4 en lo que esperamos la subida de bitcoin un ...
5 rocket arena airdrop is live now earn 200 r...
6 what if for the first time this entire cycle ...
7 when in doubt zoom bitcoin
8 lo que se avecina es una crisis del modelo m...
9 gratz if you grabbed some of the first 90 of...
```

[14]: *#Calculating Negative, Positive, Neutral and Compound values*

```
tw_list[['polarity', 'subjectivity']] = tw_list['text'].apply(lambda Text: pd.
    ↳Series(TextBlob(Text).sentiment))
for index, row in tw_list['text'].iteritems():
    score = SentimentIntensityAnalyzer().polarity_scores(row)
    neg = score['neg']
    neu = score['neu']
    pos = score['pos']
    comp = score['compound']
    if neg > pos:
        tw_list.loc[index, 'sentiment'] = "negative"
    elif pos > neg:
        tw_list.loc[index, 'sentiment'] = "positive"
    else:
        tw_list.loc[index, 'sentiment'] = "neutral"
    tw_list.loc[index, 'neg'] = neg
    tw_list.loc[index, 'neu'] = neu
    tw_list.loc[index, 'pos'] = pos
```



```
tw_list.loc[index, 'compound'] = comp

tw_list.head(10)
```

```
[14]:
```

	text	polarity	subjectivity
0	empower oversight sues sec for access to foia...	-0.100000	0.600000
1	make it or break it week what do you fine ...	0.416667	0.500000
2	paying for a coffee with bitcoin lightning n...	0.250000	0.333333
3	coinmarketcap has announced it will be sponso...	0.000000	0.000000
4	en lo que esperamos la subida de bitcoin un ...	0.000000	0.000000
5	rocket arena airdrop is live now earn 200 r...	0.289899	0.744444
6	what if for the first time this entire cycle ...	-0.013889	0.500000
7	when in doubt zoom bitcoin	0.000000	0.000000
8	lo que se avecina es una crisis del modelo m...	0.000000	0.000000
9	gratz if you grabbed some of the first 90 of...	0.183333	0.200000

	sentiment	neg	neu	pos	compound
0	positive	0.119	0.725	0.155	0.1779
1	positive	0.000	0.878	0.122	0.2023
2	negative	0.236	0.764	0.000	-0.5719
3	neutral	0.000	1.000	0.000	0.0000
4	positive	0.000	0.840	0.160	0.2732
5	positive	0.000	0.758	0.242	0.6249
6	negative	0.091	0.909	0.000	-0.1027
7	negative	0.385	0.615	0.000	-0.3612
8	negative	0.409	0.591	0.000	-0.8860
9	positive	0.176	0.627	0.197	0.1779

```
[15]: #Creating new data frames for all sentiments (positive, negative and neutral)
```

```
tw_list_negative = tw_list[tw_list["sentiment"]=="negative"]
tw_list_positive = tw_list[tw_list["sentiment"]=="positive"]
tw_list_neutral = tw_list[tw_list["sentiment"]=="neutral"]
```

```
[16]: #Function for count_values_in single columns
```

```
def count_values_in_column(data,feature):  
    total=data.loc[:,feature].value_counts(dropna=False)  
    percentage=round(data.loc[:,feature].  
↪value_counts(dropna=False,normalize=True)*100,2)  
    return pd.concat([total,percentage],axis=1,keys=['Total','Percentage'])
```

```
[17]: #Count_values for sentiment
```

```
count_values_in_column(tw_list,"sentiment")
```

```
[17]:
```

	Total	Percentage
neutral	179	45.09
positive	160	40.30
negative	58	14.61

```
[21]: #Function to Create Wordcloud
```

```
def create_wordcloud(text):  
    stopwords = set(STOPWORDS)  
    wc = WordCloud(background_color="white",  
                    max_words=3000,  
                    stopwords=stopwords,  
                    repeat=True)  
    wc.generate(str(text))  
    wc.to_file("wc.png")  
    print("Word Cloud Saved Successfully")  
    path="wc.png"  
    display(Image.open(path))
```

```
[22]: #Creating wordcloud for all tweets
```

```
create_wordcloud(tw_list["text"].values)
```

Word Cloud Saved Successfully



```
[23]: #Creating wordcloud for positive sentiment
create_wordcloud(tw_list_positive["text"].values)
```

Word Cloud Saved Successfully



```
[24]: #Creating wordcloud for negative sentiment
create_wordcloud(tw_list_negative["text"].values)
```

Word Cloud Saved Successfully



```
[25]: #Creating wordcloud for neutral sentiment
create_wordcloud(tw_list_neutral["text"].values)
```

Word Cloud Saved Successfully



```
[26]: #Calculating tweet's lenght and word count
tw_list['text_len'] = tw_list['text'].astype(str).apply(len)
tw_list['text_word_count'] = tw_list['text'].apply(lambda x: len(str(x).
↳split()))
```

```
[27]: round(pd.DataFrame(tw_list.groupby("sentiment").text_len.mean()),2)
```

```
[27]:          text_len
      sentiment
      negative      93.62
      neutral       91.42
      positive     105.77
```

```
[28]: round(pd.DataFrame(tw_list.groupby("sentiment").text_word_count.mean()),2)
```

```
[28]:          text_word_count
      sentiment
      negative      15.97
      neutral       14.73
      positive      16.37
```

```
[29]: #Removing Punctuation
      def remove_punct(text):
          text = "".join([char for char in text if char not in string.punctuation])
          text = re.sub('[0-9]+', '', text)
          return text

      tw_list['punct'] = tw_list['text'].apply(lambda x: remove_punct(x))
```

```
[30]: #Applying tokenization
      def tokenization(text):
          text = re.split('\W+', text)
          return text

      tw_list['tokenized'] = tw_list['punct'].apply(lambda x: tokenization(x.lower()))
```

```
[34]: #Removing stopwords
      stopword = nltk.corpus.stopwords.words('english')
      def remove_stopwords(text):
          text = [word for word in text if word not in stopword]
          return text

      tw_list['nonstop'] = tw_list['tokenized'].apply(lambda x: remove_stopwords(x))
```

```
[35]: #Applying Stemmer
      ps = nltk.PorterStemmer()

      def stemming(text):
          text = [ps.stem(word) for word in text]
          return text

      tw_list['stemmed'] = tw_list['nonstop'].apply(lambda x: stemming(x))
```

```
[36]: #Cleaning Text
def clean_text(text):
    text_lc = "".join([word.lower() for word in text if word not in string.
    ↳punctuation]) # remove punctuation
    text_rc = re.sub('[0-9]+', '', text_lc)
    tokens = re.split('\W+', text_rc) # tokenization
    text = [ps.stem(word) for word in tokens if word not in stopwords] # remove_
    ↳stopwords and stemming
    return text
```

```
[37]: tw_list.head()
```

```
[37]:
```

```
0 RT @EMPOWR_us: Empower Oversight Sues SEC For ...
1 RT @MinusWells: Make it or break it week... Wh...
2 RT @CoinCornerMolly: Paying for a coffee with ...
3 RT @HPOS10INU: Coinmarketcap has announced it ...
4 RT @CryptoYodaGuru: En lo que esperamos la sub...
```

	text	polarity	subjectivity
0	empower oversight sues sec for access to foia...	-0.100000	0.600000
1	make it or break it week what do you fine ...	0.416667	0.500000
2	paying for a coffee with bitcoin lightning n...	0.250000	0.333333
3	coinmarketcap has announced it will be sponso...	0.000000	0.000000
4	en lo que esperamos la subida de bitcoin un ...	0.000000	0.000000

	sentiment	neg	neu	pos	compound	text_len	text_word_count
0	positive	0.119	0.725	0.155	0.1779	116	16
1	positive	0.000	0.878	0.122	0.2023	75	14
2	negative	0.236	0.764	0.000	-0.5719	97	15
3	neutral	0.000	1.000	0.000	0.0000	126	16
4	positive	0.000	0.840	0.160	0.2732	75	12

	punct
0	empower oversight sues sec for access to foia...
1	make it or break it week what do you fine ...
2	paying for a coffee with bitcoin lightning n...
3	coinmarketcap has announced it will be sponso...
4	en lo que esperamos la subida de bitcoin un ...

	tokenized
0	[, empower, oversight, sues, sec, for, access,...
1	[, make, it, or, break, it, week, what, do, yo...
2	[, paying, for, a, coffee, with, bitcoin, ligh...
3	[, coinmarketcap, has, announced, it, will, be...
4	[, en, lo, que, esperamos, la, subida, de, bit...

```

                                nonstop \
0  [, empower, oversight, sues, sec, access, foia...
1  [, make, break, week, fine, people, thing, btc...
2  [, paying, coffee, bitcoin, lightning, nfc, pa...
3  [, coinmarketcap, announced, sponsoring, tomfe...
4  [, en, lo, que, esperamos, la, subida, de, bit...

```

```

                                stemmed
0  [, empow, oversight, sue, sec, access, foia, d...
1  [, make, break, week, fine, peopl, thing, btc,...
2  [, pay, coffe, bitcoin, lightn, nfc, pair, ano...
3  [, coinmarketcap, announc, sponsor, tomfelton,...
4  [, en, lo, que, esperamo, la, subida, de, bitc...

```

```

[38]: #Applying Countvectorizer
countVectorizer = CountVectorizer(analyzer=clean_text)
countVector = countVectorizer.fit_transform(tw_list['text'])
print('{} Number of reviews has {} words'.format(countVector.shape[0],
↪countVector.shape[1]))
#print(countVectorizer.get_feature_names())

```

397 Number of reviews has 1964 words

[]: