

TwitterSentimentAnalysis

October 11, 2020

1 Twitter Sentiment Analysis

2 By Julian Murillo

```
[1]: import tweepy
from textblob import TextBlob

from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt

import os
os.getcwd()
```

```
[1]: '/Users/julianjr.'
```

```
[2]: # If you'd like to replicate this project go to:
# https://developer.twitter.com/en/docs/getting-started

'''
Also. Just as a recommendation...
I'd use environment variables to store this keys and tokens. Keep em safe.
'''

consumer_key = '<your consumer key here>'
consumer_secret = '<yout consumer secret here>'
access_token = '<your access token here>'
access_token_secret = '<your access token secret here>'
auth = tweepy.OAuthHandler(consumer_key,consumer_secret)
auth.set_access_token(access_token,access_token_secret)
api = tweepy.API(auth, wait_on_rate_limit = True)

query = "Donald Trump"
public_tweets = api.search(query, count = 100)
```

```
'''
# Old/quick way to grab analysis

for tweet in public_tweets:
    print(tweet.text)
    analysis = TextBlob(tweet.text)
    print(analysis.sentiment)
'''
```

```
[2]: '\n# Old/quick way to grab analysis\n\nfor tweet in public_tweets:\n
print(tweet.text)\n    analysis = TextBlob(tweet.text)\n
print(analysis.sentiment)\n'
```

```
[3]: # Extract just 100 tweets from twitter

# To look at Tweets from the account (Donalds Account)
# posts = api.user_timeline(screen_name = "Donald Trump", count = 100, lang = "en", text_mode = "extended")

# Print the 5 most recent PUBLIC tweets

i = 0
for tweet in public_tweets[0:5]:
    print(str(i) + '. ' + tweet.text + '\n')
    i+=1
```

0.) RT @acnewsitics: Hunter Biden was Chairman of the World Food Program,... which just won the Nobel Peace Prize.

Eric Trump, Donald J Trump...

1.) RT @charliekirk11: RT if you're thankful Donald Trump's Supreme Court nominee is getting a confirmation hearing tomorrow and NOT Hillary Cl...

2.) RT @SenDuckworth: Another week gone by and no word from Donald Trump on #BountyGate. Day 107.

3.) RT @shannonrwatts: Amy Thorn, the registered nurse from West Virginia who recently spoke at the Republican National Convention to defend Do...

4.) @DC_Draino How manynwomen need to accuse dump,of sex assault before they do something' <https://t.co/kxrlJXXxgT>

```
[4]: # Create DF with a column called tweets

tweets_df = pd.DataFrame([tweet.text for tweet in public_tweets], columns =
↳ ['tweets'])
tweets_df.head(5)
```

```
[4]:                                     tweets
0  RT @acnewsitics: Hunter Biden was Chairman of ...
1  RT @charliekirk11: RT if you're thankful Donal...
2  RT @SenDuckworth: Another week gone by and no ...
3  RT @shannonrwatts: Amy Thorn, the registered n...
4  @DC_Draino How manywomen need to accuse dump,...
```

```
[5]: # Clean tweets a little... remove @'s, links, etc.
# the little r in front tells Python that we're looking at a raw string
def clean_text(text):
    text = re.sub(r'@[A-Za-z0-9]+', '', text) # Remove @'s
    text = re.sub(r'#', '', text) # Remove hashtags
    text = re.sub(r'RT[\s+]', '', text) # Remove RT's
    text = re.sub(r'https?:\/\/\S+', '', text) # Remove Hyperlink
    text = re.sub(r':', '', text) # remove colons
    text = re.sub(r'https', '', text) # Remove Hyperlink

    return text

tweets_df['tweets'] = tweets_df['tweets'].apply(clean_text)

tweets_df.head()
```

```
[5]:                                     tweets
0  Hunter Biden was Chairman of the World Food P...
1  if you're thankful Donald Trump's Supreme Cou...
2  Another week gone by and no word from Donald ...
3  Amy Thorn, the registered nurse from West Vir...
4  _Draino How manywomen need to accuse dump,of ...
```

```
[6]: # Create a function that gets polarity of the text

def sentiment_grab(tweet):
    analysis = TextBlob(tweet)
    senti = analysis.sentiment.polarity

    return senti

tweets_df['polarity'] = tweets_df['tweets'].apply(sentiment_grab)

tweets_df.head()
```

```
[6]:
```

	tweets	polarity
0	Hunter Biden was Chairman of the World Food P...	0.0
1	if you're thankful Donald Trump's Supreme Cou...	0.0
2	Another week gone by and no word from Donald ...	0.0
3	Amy Thorn, the registered nurse from West Vir...	0.0
4	_Draino How manywomen need to accuse dump,of ...	0.0

```
[7]: # Create a function that gets subjectivity of the text... a little quicker than
      ↪ 1st function
def subjectivity_grab(tweet):
    return TextBlob(tweet).sentiment.subjectivity

tweets_df['subjectivity'] = tweets_df['tweets'].apply(subjectivity_grab)

tweets_df.head()
```

```
[7]:
```

	tweets	polarity	subjectivity
0	Hunter Biden was Chairman of the World Food P...	0.0	0.00
1	if you're thankful Donald Trump's Supreme Cou...	0.0	0.00
2	Another week gone by and no word from Donald ...	0.0	0.00
3	Amy Thorn, the registered nurse from West Vir...	0.0	0.25
4	_Draino How manywomen need to accuse dump,of ...	0.0	0.00

```
[8]: # Visualization: WordCloud
all_words = ' '.join([tw for tw in tweets_df['tweets']])

MyCloud = WordCloud(width = 500, height = 300, random_state = 25, max_font_size=
      ↪ 105).generate(all_words)

plt.imshow(MyCloud, interpolation = "bilinear")
plt.axis('off')
plt.show()
```


98	Helmut Norpoth Model has correctly predicted ...	0.000	0.000
99	Psychiatrists warn voters 'things will get wo...	-0.400	0.600

```

analysis
95 Neutral
96 Negative
97 Neutral
98 Neutral
99 Negative

```

```

[11]: # Coerce a boolean to an int by just multiplying it by one

# Binary for Positives
tweets_df['IsPositive'] = (tweets_df["analysis"] == 'Positive') * 1

# Binary for Negatives
tweets_df['IsNegative'] = (tweets_df["analysis"] == 'Negative') * 1

# Binary for Neutrals
tweets_df['IsNeutral'] = (tweets_df["analysis"] == 'Neutral') * 1

tweets_df.head()

```

```

[11]:
           tweets  polarity  subjectivity \
0  Hunter Biden was Chairman of the World Food P...      0.0      0.00
1  if you're thankful Donald Trump's Supreme Cou...      0.0      0.00
2  Another week gone by and no word from Donald ...      0.0      0.00
3  Amy Thorn, the registered nurse from West Vir...      0.0      0.25
4  _Draino How manywomen need to accuse dump,of ...      0.0      0.00

```

	analysis	IsPositive	IsNegative	IsNeutral
0	Neutral	0	0	1
1	Neutral	0	0	1
2	Neutral	0	0	1
3	Neutral	0	0	1
4	Neutral	0	0	1

```

[12]: tweets_df.tail() # I don't thiknk we csan get more than 100 search results for_
      ↪ a single query

```

```

[12]:
           tweets  polarity  subjectivity \
95  Hello media, the  broke a story that Donald T...      0.000      1.000
96  Let's not forget, whatever other distractions...     -0.125      0.375
97  Omar Bula Escobar Donald Trump estorba los pl...      0.000      0.000
98  Helmut Norpoth Model has correctly predicted ...      0.000      0.000
99  Psychiatrists warn voters 'things will get wo...     -0.400      0.600

```

	analysis	IsPositive	IsNegative	IsNeutral
95	Neutral	0	0	1
96	Negative	0	1	0
97	Neutral	0	0	1
98	Neutral	0	0	1
99	Negative	0	1	0

```
[13]: # Print all the positive tweets from query

j = 1
organized_df = tweets_df.sort_values(by = ['polarity'])
for i in range(organized_df.shape[0]):
    if (organized_df['analysis'][i] == "Positive"):
        print(str(j) + ') ' + organized_df['tweets'][i])
        print()
        j = j+1
```

1) The Taliban, Russia, and China all want Donald Trump to win this election.

When three of America's biggest adversarie...

2) Donald Trump is showing off his Trump University medical degree. Only the best physicians graduate from...

3) Donald Trump is going to win again... because he has too.

4) Trump is holding campaign rallies this week in various states. Tragically for our country, Trump cares more about getti...

5) Joe Biden's most recent endorsements Cindy McCain, Scientific American, Greta Thunberg, the New England Journal of Medic...

6)_2020 Exactly, China just drops the value of their currency if Trump increases Tariffs. Donal... //t.co/FhA75BPRFC

7)_Anderson Media "Statistically, the coronavirus is more likely to cost the than his...

8) Joe Biden's most recent endorsements Cindy McCain, Scientific American, Greta Thunberg, the New England Journal of Medic...

9)_nyc Trump already posted this nonsense last week - it's a copy/paste job from an old tweet! And for those who...

10)Holy Mary, pray for wisdom, judgement, guidance, protection, & success for our Donald Trump.
St Michael, t... //t.co/WFp3Xw40hU

11) The Taliban endorsed Donald Trump today.

No. Seriously. They really did. This is not a joke.

That's it. That's the whole...

12) New from In 2016, as campaign funds dwindled, Trump orchestrated a windfall,
"over \$21 million in highly unus...

13) Joe Biden's most recent endorsements Cindy McCain, Scientific American, Greta Thunberg, the New England Journal of Medic...

14)_nyc More specifically, how do people feel about Donald Trump's performance? Only 43.6% approve while 53.2% dis...

15) Joe Biden's most recent endorsements Cindy McCain, Scientific American, Greta Thunberg, the New England Journal of Medic...

16)_Tail BREAKING

LATEST POLL SHOWS HILLARY CLINTON HAS MOVED INTO SECOND PLACE IN THE PRESIDENTIAL RACE PASSING DONALD TRUMP!

17) It is positively hilarious that Donald Trump would be fear tweeting that Dems will pack the court. It is exactly precisely e...

18) Joe Biden's most recent endorsements Cindy McCain, Scientific American, Greta Thunberg, the New England Journal of Medic...

19) Joe Biden's most recent endorsements Cindy McCain, Scientific American, Greta Thunberg, the New England Journal of Medic...

20) We should all be able to see Trump's tax returns
[//t.co/u8AbenlkZP](https://t.co/u8AbenlkZP)

21)_Tail BREAKING

LATEST POLL SHOWS HILLARY CLINTON HAS MOVED INTO SECOND PLACE IN THE PRESIDENTIAL RACE PASSING DONALD TRUMP!

22) Garrett Watson "Social Security could be funded using general fund revenue or alternative revenue source,... [//t.co/WtGgdvSFHn](https://t.co/WtGgdvSFHn)

23) Joe Biden's most recent endorsements Cindy McCain, Scientific American, Greta Thunberg, the New England Journal of Medic...


```
[20]: # Print all negative tweets from query
j = 1
organized_df = tweets_df.sort_values(by = ['polarity'])
for i in range(organized_df.shape[0]):
    if (organized_df['analysis'][i] == "Negative"):
        print(str(j) + ' ' + organized_df['tweets'][i])
        print()
        j = j+1
```

1) One out of every 1,640 Americans alive before the pandemic hit, is now dead because of... [//t.co/lyw7Ddlt54](https://t.co/lyw7Ddlt54)

2) "I don't have to shake hands with these disgusting people." - President Donald J. Trump.

Baskets of deplorables, perhaps?

3) "I don't have to shake hands with these disgusting people." - President Donald J. Trump.

Baskets of deplorables, perhaps?

4) _Graham Listen. I hate Donald Trump. Christians hate me because I do. Who is the righteous? I'll call it a draw.

5) ...and here is _Osborne 's previous investigation (w/ +) into Donald Trump's history with...

6) "Maybe COVID is a good thing? That way I don't have to shake hands with these disgusting people."

- Donald Trump on his su...

7) There comes a time where we have to choose hope over fear, kindness over hate, and country over party. That's exactly w...

8) The evidence and the science show that Donald Trump has badly damaged the U.S. and its people-because he rejects evidence and sc...

9) Twitter makes sure that Trending on Twitter is anything bad, Fake or not, about President Donald Trump. So obvious wha...

10) In 2012, Newsmax wanted a Republican debate moderated by Donald Trump. I killed it because who would want a debate mode...

11) In 2012, Newsmax wanted a Republican debate moderated by Donald Trump. I killed it because who would want a debate mode...

12) This morning Donald Trump spread the dangerous "antifa-wildfires" disinformation campaign-which has resulted in people...

13) Donald Trump's catastrophic failure as President of the United States, combined with his dishonesty, cr...

14) Twitter makes sure that Trending on Twitter is anything bad, Fake or not, about President Donald Trump. So obvious wha...

15) "I don't have to shake hands with these disgusting people." - President Donald J. Trump.

Baskets of deplorables, perhaps?

16)TrumpIsALoser ha. donald trump is a fucking trifling chump.
//t.co/Q00lBuUWBF

17) Donald Trump's catastrophic failure as President of the United States, combined with his dishonesty, cr...

18) "I don't have to shake hands with these disgusting people." - President Donald J. Trump.

Baskets of deplorables, perhaps?

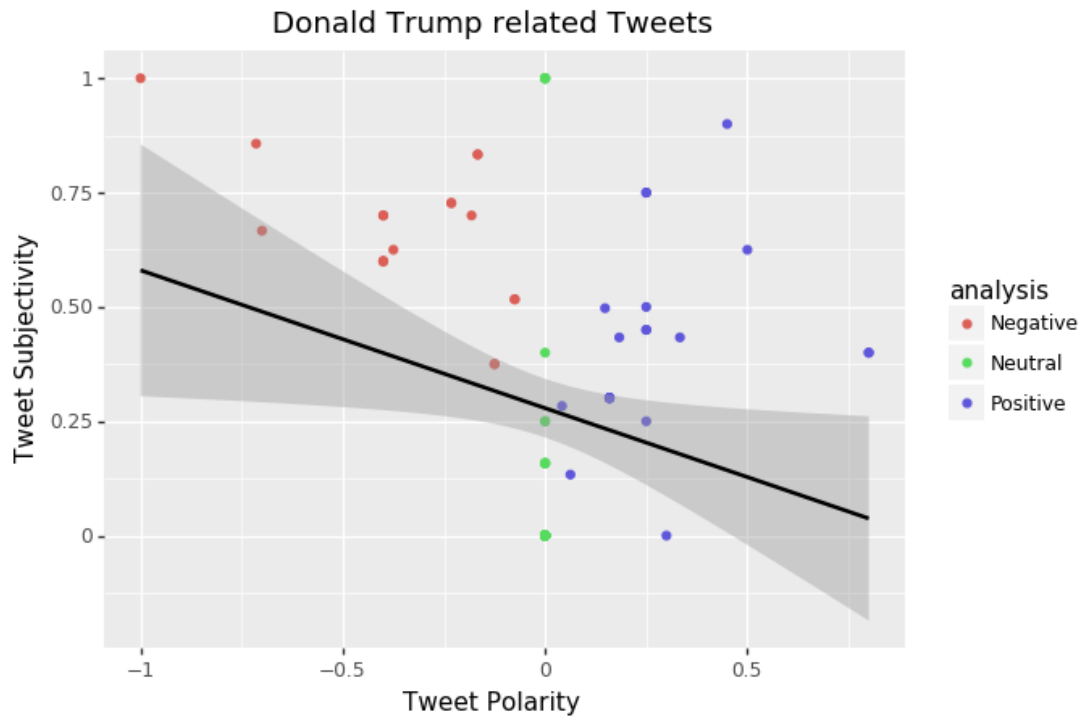
19) Why hasn't every single major newspaper in the country called on Donald Trump to resign?

20) Now do trump billy boy? He repeats things at least 3 times. Hmm.. wonder why ?

Watch Unfit The psycholo...

```
[14]: # Plot Polarity and Subjectivity using GGplot2
from plotnine import *
from matplotlib import pyplot

(ggplot(tweets_df, aes(x = 'polarity', y = 'subjectivity')) +
  ↳geom_point(mapping = aes(color = 'analysis')) +
  labs(x = "Tweet Polarity", y = "Tweet Subjectivity", title = "Donald Trump
  ↳related Tweets") + stat_smooth(method='lm'))
```



[14]: <ggplot: (8765806713909)>

Interestingly, based off of this search, the more subjective the tweets are... the more negative they tend to be! The colors of the dots in the scatter plot indicate the classification of the tweet (negative, neutral, positive).

```
[15]: # Lets see what % of positive and negative tweets there are
pos_tweets = tweets_df[tweets_df.analysis == 'Positive']

pos_perc = round((pos_tweets.shape[0]/tweets_df.shape[0])* 100,2)
print(pos_perc)
```

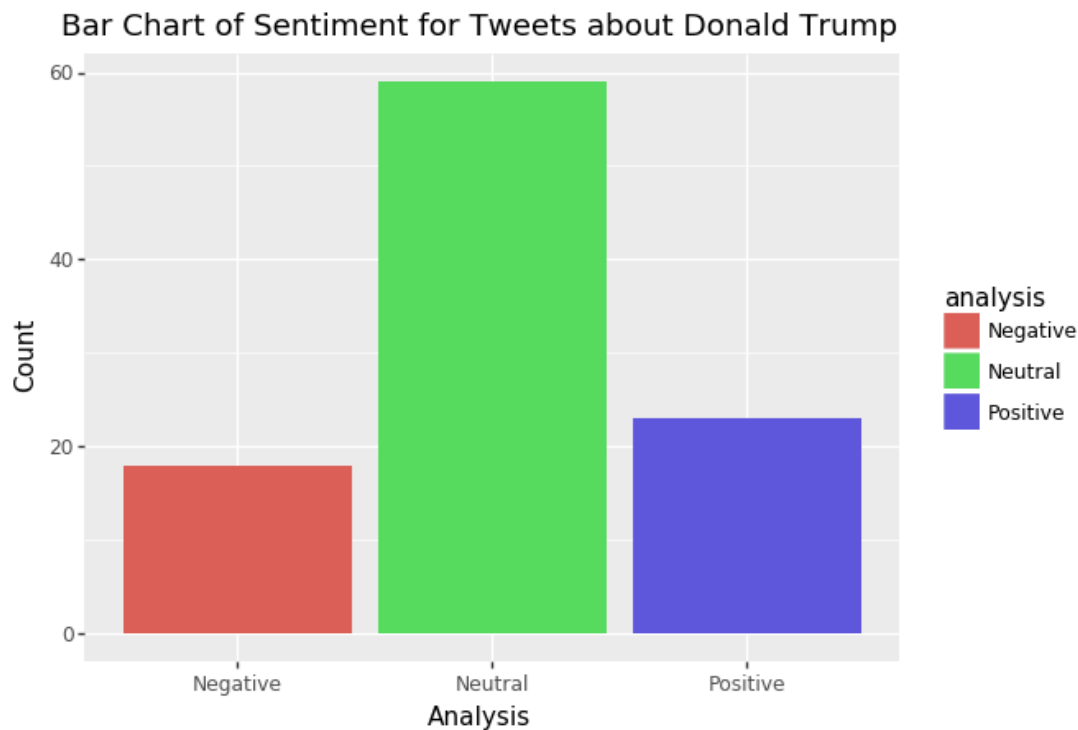
23.0

```
[16]: # Now % of negative tweets
neg_tweets = tweets_df[tweets_df.analysis == 'Negative']
neg_perc = round(neg_tweets.shape[0]/tweets_df.shape[0] * 100,2)
print(neg_perc)
```

18.0

```
[17]: # Last super quick plot... should illustrate our percentages perfectly since we
      ↳ only have 100 Tweets
(ggplot(tweets_df , aes(x = 'analysis')) + geom_bar(aes(fill = 'analysis')) +
```

```
labs(x = 'Analysis', y = 'Count', title = 'Bar Chart of Sentiment for Tweets_
↳about Donald Trump'))
```



[17]: <ggplot: (8765806721201)>

Based off of this visual it weirdly looks like most Donald Trump subjected tweets are Neutral. Might have to do with the way that TextBlob calculates Polarity & Subjectivity.

```
[29]: '''
Store Twitter Data into CSV file for more robust analysis later on/time series_
↳analysis
Although we only get 100 search results per query execution... we can easily_
↳just save each query output to a CSV
then read it back at a later time.
'''
os.chdir("/Users/julianjr./Desktop/GitHub/TwitterSentimentAnalysis")

tweets_df.to_csv('TwitterData9.17.20.csv')
```

Above is code to store the most relevant/Top 100 “Donald Trump” Tweets. Unfortunately I don’t think that the Twitter API (the free one) will allow us to get more than 100 tweets at a time. But you could just running your query and store over and over in the same .csv file to do more comrehensive analysis with more than 100 tweets.