

Sentiment Analysis with Tweepy

Set up environment

```
In [1]: !pip install tweepy==4.9.0

Collecting tweepy==4.9.0
  Downloading tweepy-4.9.0-py3-none-any.whl (77 KB)
Collecting requests-oauthlib<2,>=1.2.0
  Downloading requests_oauthlib-1.1.1-py2.py3-none-any.whl (23 KB)
Collecting oauthlib<4,>=3.2.0
  Downloading oauthlib-3.2.2-py3-none-any.whl (151 KB)
Requirement already satisfied: requests<3,>=2.27.0 in c:\users\muska\anaconda3\lib\site-packages (from tweepy==4.9.0) (2.27.1)
Requirement already satisfied: certifi<=2017.4.17 in c:\users\muska\anaconda3\lib\site-packages (from requests<3,>=2.27.0->tweepy==4.9.0) (2021.10.8)
Requirement already satisfied: charset-normalizer<=2.0.0 in c:\users\muska\anaconda3\lib\site-packages (from requests<3,>=2.27.0->tweepy==4.9.0) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\muska\anaconda3\lib\site-packages (from requests<3,>=2.27.0->tweepy==4.9.0) (1.26.9)
Requirement already satisfied: idna<4,>=2.5 in c:\users\muska\anaconda3\lib\site-packages (from requests<3,>=2.27.0->tweepy==4.9.0) (1.3)
Installing collected packages: oauthlib, requests-oauthlib, tweepy
Successfully installed oauthlib-3.2.2 requests-oauthlib-1.3.1 tweepy-4.9.0


In [3]: !pip install textblob

Collecting textblob-0.17.1-py2.py3-none-any.whl (636 KB)
Requirement already satisfied: nltk<=3.1 in c:\users\muska\anaconda3\lib\site-packages (from textblob) (3.7)
Requirement already satisfied: regex<=2021.8.3 in c:\users\muska\anaconda3\lib\site-packages (from nltk<=3.1->textblob) (2022.3.15)
Requirement already satisfied: click in c:\users\muska\anaconda3\lib\site-packages (from nltk<=3.1->textblob) (8.0.4)
Requirement already satisfied: tqdm in c:\users\muska\anaconda3\lib\site-packages (from nltk<=3.1->textblob) (4.64.0)
Requirement already satisfied: joblib in c:\users\muska\anaconda3\lib\site-packages (from nltk<=3.1->textblob) (1.1.0)
Requirement already satisfied: colorama in c:\users\muska\anaconda3\lib\site-packages (from click->nltk<=3.1->textblob) (0.4.4)
Installing collected packages: textblob
Successfully installed textblob-0.17.1


In [5]: !pip install wordcloud

Collecting wordcloud
  Downloading wordcloud-1.8.2.2-cp39-cp39-win_and64.whl (153 KB)
Requirement already satisfied: matplotlib in c:\users\muska\anaconda3\lib\site-packages (from wordcloud) (3.5.1)
Requirement already satisfied: pillow in c:\users\muska\anaconda3\lib\site-packages (from wordcloud) (9.0.1)
Requirement already satisfied: regex<=2021.8.3 in c:\users\muska\anaconda3\lib\site-packages (from wordcloud) (1.2)
Requirement already satisfied: pyparsing<=2.2.1 in c:\users\muska\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.0.4)
Requirement already satisfied: python-dateutil<=2.7 in c:\users\muska\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: packaging<=20.0 in c:\users\muska\anaconda3\lib\site-packages (from matplotlib->wordcloud) (21.3)
Requirement already satisfied: fonttools<=4.22.0 in c:\users\muska\anaconda3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: kiwisolver<=1.0.1 in c:\users\muska\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.3.2)
Requirement already satisfied: cycler<=0.10 in c:\users\muska\anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: fonttools<=4.22.0 in c:\users\muska\anaconda3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: kiwisolver<=1.0.1 in c:\users\muska\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.3.2)
Requirement already satisfied: six<=1.5 in c:\users\muska\anaconda3\lib\site-packages (from python-dateutil<=2.7->matplotlib->wordcloud) (1.16.0)
Installing collected packages: wordcloud
Successfully installed wordcloud-1.8.2.2


In [6]: import tweepy
import pandas as pd
import re
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from textblob import TextBlob
import nltk
nltk.download('vader_lexicon')

import numpy as np
from PIL import Image
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt

[nltk_data] Downloading package vader_lexicon to
[nltk_data] c:\users\muska\appdata\local\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!


Twitter Authentication
```

```
In [10]: #Put your Bearer Token in the parenthesis below
client = tweepy.Client(bearer_token='AAAAAAAAAAAAAAAAAAOVWjVEAAAAaz3YAf8h621aWwrfBC9hKbH2B5fs302qKG4Pc7QN')
```

Get recent public tweets on a keyword

```
In [11]: # Pull tweets from twitter

query = 'apple -is:retweet lang=en'
paginator = tweepy.Paginator(
    client.search_recent_tweets,          # The method you want to use
    query=query,                         # Some argument for this method
    max_results=100,                    # How many tweets per page
    limit=10                             # How many pages to retrieve
)
# Get tweets that contain the hashtag #TypeKeywordHere
# is:retweet means I don't want retweets
# lang=en is asking for the tweets to be in english

tweet_list = []

for tweet in paginator.flatten(): # Total number of tweets to retrieve
    tweet_list.append(tweet)
    print(tweet)
```


[illegible]

[illegible]

[illegible]


```
#Apple #Message
Day 17 of sending a #Airtrag from Glasgow, Scotland to Belgrade, Serbia. Still in #Oslo Airport, will it get
y #Apple #Tracking #AppleAirtag from Glasgow, Scotland to Belgrade, Serbia. Still in #Oslo Airport, will it get
y! #Apple #Tracking #AppleAirtag from Glasgow, Scotland to Belgrade, Serbia. Still in #Oslo Airport, will it get
#NIBE: China, #newsbalance &mp; S.A.S shoes not china, #shark vacuum: China. #Miele vacuum: China. #Micheli
n: China, #Cooper &mp; #GoodYear not China. #Apple, #HP &mp; #Lenovo: China. #Samsung, #Sony &mp; #LG not Ch
ina. #USA #UTERANS #America #Christmas #christmas2022 #checkbeforeyoubuy
#RYDEINMGING #VIDEO #MICROBENCODING
#APPLE #UNITEDSTATES #JAPAN #AUSTRALIA #MEXICO #RUSSIA #CANADA #ISRAEL #CHINA #TAIWAN #UKRAINE #SWITZERLAND #NO
#RUSSIA #SOUTHKOREA #INDONESIA #EUROPE #INDIA
CLICK ON VIDEO COMBUND ALGORITHM MICROENCODING
https://t.co/bxYtYw0rkp
How that Apple has the ability to customize your lock screen, I think they should find an easier way to install
themes. I really like this aspect of #Android, but I love the fluid feel of iOS. @tim_cook can we figure somet
hing out? #Apple #themeselease #wishlist #tech
Chrome on Ventura 12.1 is a interesting experience, mine goes steady downhill until my maxed 14inch pro start ha
ating up 🤖, next stage is extreme lag and finally a total freeze #apple #chrome #mipr0 #buy v107.0.5304.121
Content analysis:
This article scores 21/100. The complete report can be viewed here: https://t.co/PDDd0un1JW
(I'm a bot)
```

```
#Apple #iPhone #https://t.co/2lYq99sFFR
#Apple App Tracking doesn't actually stop tracking and at the same time it does but against the competitors.
#Apple #Privacy
https://t.co/xL2m9ldS https://t.co/b6ryMlfY2B
#Apple Might Have Shaved Plans to Launch Reasonably priced Pencil: Report https://t.co/yiczA92oqK #ApplePencil
Sunday secrets. Onlyfans newbie. Fresh in the game ready to learn. Tell me what to do. I'll be your secret... Y
ou can be mine. #undies #secret #bunbun #50BBDwelly #adultchatting #eeet #billie #billiegies #transfom #vic
al #skinnytotes #Apple #photomemory #findom #bdm #all #of https://t.co/qTbhk2FCs
Content analysis:
This article scores 64/100. The complete report can be viewed here: https://t.co/BjULJ3jYWx
(I'm a bot)
```

```
#Apple #David https://t.co/bm3m3Zb19C
How that Apple has the ability to customize your lock screen, I think they should find an easier way to install
that Photo removed in 'Recents'. Current functionality only adds a photo to another album, it doesn't move it
#EnhancementRequest #iPhone #Apple #PhotoManagement
@elonmusk Of Apple or #google block you app I will switch to your phone service.
Weekend reading: there are two generations of Apple Pencil that offer similar experience, but only work with a
Apple Pencil
https://t.co/MXXxiQcQp2
Check out my #AAPL analysis on #TradingView: https://t.co/10aMR37oAs
#Spy #Tesla #StockMarket #Apple
High time for Tesla Mobile. @elonmusk
China #Apple #Twitter #ELONMUSK https://t.co/6vgfK239uB
New episode is live! We address some hate comments!
Youtube: https://t.co/c9h8f3Jooa
Apple: https://t.co/2aG0h2AllyJ
Spotify: https://t.co/eQ0dJ50Uto
Google Podcasts: https://t.co/Sydpq8h3u3
```

```
#podcast #spotify #apple #youtube #tiktok #canada #cancelled #internettrolls https://t.co/xlcmd4lY4s
Who made #nyuka because of Apple Pay?
https://t.co/6ENXQYcAvU
|
#Apple #Made
https://t.co/6ENXQYcAvU
Who made #nyuka because of Apple Pay?
https://t.co/XPD4oqV2ap
|
#Apple #Made https://t.co/FoG3p8dYdc
Best MacBook Cyber Monday 2022 deals: M2 MacBook Air/Pro $150 off, Mac Studio all-time low, more https://t.co/m
2TFLi4c0F #Apple #Mac #Technology https://t.co/HJHWPWtXglu
What Will Iger Do With ESPN and Hulu?
```

#AllianceBernstein #Apple #Bank #Binance #Coinbase #Disney #Walt #Twitter

```
https://t.co/3UUXkxvb6l https://t.co/9qzK7aTsdJ
#Spain vs Germany. 🇩🇪 #WU20cup2022 https://t.co/WUTW0UBU1J
Apple hobbled a crucial tool of dissent in China weeks before widespread protests broke out
```

#APPLE #CHINESE

```
https://t.co/MABaAKogLM
The ability to shapelifit comes with many temptations...
```

https://t.co/DxKxetSG26

```
#99c #bookdeals #a1a1 #bookboost #ASMSG #amazon #kindle #apple #nook #kobo #1ARTG #magic #fantasyromance #PNR
#readingcommunity #romanticfantasy #shifter #fatedmates https://t.co/KCB8BJALbD
Weekend reading: there are two generations of Apple Pencil that offer similar experiences, but only work with a
specific iPad's. Learn more about them here.
https://t.co/1lba7yab6
David Beckham #Apple and British Billion Air Jim Ratcliffe Are all Interested in Buying Manchester United Fo
otball Club
#apple cider vinegar
How to use apple cider vinegar for neuropathy
https://t.co/sa1q0tpnNM
```

```
In [12]: #Creating new dataframe and new features
tweet_list_df = pd.DataFrame(tweet_list)
tweet_list_df['text_len'] = tweet_list_df['text'].str.len()
tweet_list_df.head(5)
```

```
Out[12]:
```

	text
0	So good I had to share! Check out all the item...
1	#Covid-19: #Apple Will Lose 6 Million #PhoneL...
2	Pre-save my new album "C" on Spotify: https://...
3	It's #PodcastAndChill Monday/vr/#Wrestling Fan...
4	Samsung N10 plus A+ quality display, in stock...

Data preprocessing

```
In [13]: def preprocess_tweet(text):
    # Clean text: data up, leaving only 2 or more char long non-stopwords composed of A-Z & a-z only
    in lowercase
    sentence = text.lower()
    # Remove RT
    sentence = re.sub('RT @w+: ', '', sentence)
    # Remove special characters
    sentence = re.sub('[@|&a-z-9]+|([!@-9A-Za-z \\\s])|(\w+:/\|\\$+)', '', sentence)
    # Single character removal
    sentence = re.sub('['a-zA-Z]', '', sentence) # When we remove apostrophe from the word "Mark's", c
    # Remove multiple spaces
    sentence = re.sub('['\s+', ' ', sentence) # Next, we remove all the single characters and replace it by a s
    return sentence
```

```
In [14]: cleaned_tweets = []
for tweet in tweet_list_df['text']:
    cleaned_tweet = preprocess_tweet(tweet)
    cleaned_tweets.append(cleaned_tweet)
```

```
In [15]: tweet_list_df['cleaned'] = pd.DataFrame(cleaned_tweets)
tweet_list_df.head(5)
```

```
Out[15]:
```

	text	cleaned
0	So good I had to share! Check out all the item...	so good had to share check out all the items m...
1	#Covid-19: #Apple Will Lose 6 Million #PhoneL...	covid 19 apple will lose 6 million iphone14pr...
2	Pre-save my new album "C" on Spotify: https://...	pre save my new album on spotify powered by di...
3	It's #PodcastAndChill Monday/vr/#Wrestling Fan...	it podcastandchill monday wrestling fans downl...
4	Samsung N10 plus A+ quality display, in stock...	samsung n10 plus quality display in stock inte...

Generate Sentiment Labels

```
In [16]: #Calculating Negative, Positive, Neutral and Compound values
tweet_list_df[['polarity', 'subjectivity']] = tweet_list_df['cleaned'].apply(lambda text: pd.Series(TextBlob(te
for index, row in tweet_list_df['cleaned'].iteritems():
    score = SentimentIntensityAnalyzer().polarity_scores(row)
    neg = score['neg']
    neu = score['neu']
    pos = score['pos']
    comp = score['compound']
    if comp <= -0.05:
        tweet_list_df.loc[index, 'sentiment'] = "negative"
    elif comp >= 0.05:
        tweet_list_df.loc[index, 'sentiment'] = "positive"
    else:
        tweet_list_df.loc[index, 'sentiment'] = "neutral"
    tweet_list_df.loc[index, 'neg'] = neg
    tweet_list_df.loc[index, 'neu'] = neu
    tweet_list_df.loc[index, 'pos'] = pos
    tweet_list_df.loc[index, 'compound'] = comp
tweet_list_df.head(5)
```

```
Out[16]:
```

	text	cleaned	polarity	subjectivity	sentiment	neg	neu	pos	compound
0	So good I had to share! Check out all the item...	so good had to share check out all the items m...	0.650000	0.775000	positive	0.000	0.647	0.353	0.8516
1	#Covid-19: #Apple Will Lose 6 Million #PhoneL...	covid 19 apple will lose 6 million iphone14pr...	0.000000	0.000000	negative	0.251	0.749	0.000	-0.5719
2	Pre-save my new album "C" on Spotify: https://...	pre save my new album on spotify powered by di...	0.318182	0.477273	positive	0.000	0.842	0.158	0.4939
3	It's #PodcastAndChill Monday/vr/#Wrestling Fan...	it podcastandchill monday wrestling fans downl...	0.750000	0.600000	positive	0.000	0.789	0.211	0.7717
4	Samsung N10 plus A+ quality display, in stock...	samsung n10 plus quality display in stock inte...	0.250000	0.500000	positive	0.000	0.857	0.143	0.6124

Sentiment Visualisation

```
In [17]: #Creating new data frames for all sentiments (positive, negative and neutral)
tweet_list_df_negative = tweet_list_df[tweet_list_df['sentiment']=="negative"]
tweet_list_df_positive = tweet_list_df[tweet_list_df['sentiment']=="positive"]
tweet_list_df_neutral = tweet_list_df[tweet_list_df['sentiment']=="neutral"]
```

Donut Charts

```
In [18]: #Function for count_values in single column
def count_values_in_column(data, feature):
    total_data = count(data[feature], value_counts(dropna=False))
    percentage = round(data[feature].value_counts(dropna=False), normalize=True)*100, 2)
    return pd.concat([total, percentage, axis=1, keys=['Total', 'Percentage'])
```

```
In [19]: #Count values for sentiment
count_values_in_column(tweet_list_df, 'sentiment')
```

```
Out[19]:
```

	Total	Percentage
positive	505	50.70
neutral	338	33.87
negative	154	15.43

```
In [20]: # create data for pie chart
pchart = count_values_in_column(tweet_list_df, 'sentiment')
# names= pc.index
# size=pc["Percentage"]
names=pchart.index
size=pchart["Percentage"]
# Create a circle for the center of the plot
my_circle=plt.Circle( (0,0), 0.7, color='white')
p=plt.gcf()
plt.pie(size, labels=names, colors='green','blue','red')
p=plt.gcf()
p=plt.gcf().add_artist(my_circle)
plt.show()
```

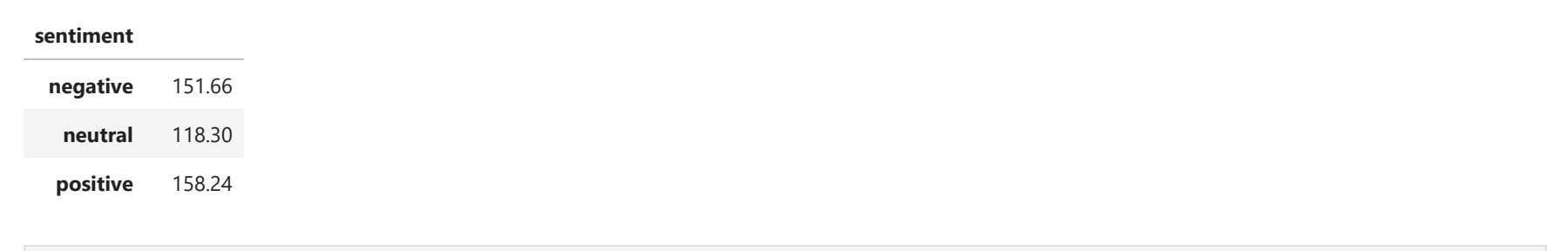


Word Clouds

```
In [22]: #Function to Create Wordcloud
def create_wordcloud(text):
    mask = np.array(Image.open("cloud.png"))
    stopwords = set(STOPWORDS)
    wc = WordCloud(background_color="white",
                    mask = mask,
                    max_words=100,
                    stopwords=stopwords,
                    repeat=True)
    wc.generate(text)
    wc.to_file("c1_wordcloud.png")
    print("Word cloud Saved Successfully")
    path=c1_wordcloud.png"
    display(Image.open(path))
```

```
In [23]: # Creating wordcloud for all tweets
create_wordcloud(tweet_list_df['cleaned'].values)
# create_wordcloud(tweet_list_df_positive['text'].values)
# create_wordcloud(tweet_list_df_negative['cleaned'].values)
# create_wordcloud(tweet_list_df_neutral['text'].values)
```

Word Cloud Saved Successfully



Text Insights

```
In [24]: #Calculating tweet's lenght and word count
tweet_list_df['text_len'] = tweet_list_df['cleaned'].astype(str).apply(len)
# create_wordcloud(tweet_list_df_negative['cleaned'].values)
# create_wordcloud(tweet_list_df_neutral['text'].values)
```

```
In [25]: round(pd.DataFrame(tweet_list_df.groupby("sentiment").text_len.mean()),2)
```

```
Out[25]:
```

sentiment	text_len
negative	151.66
neutral	118.30
positive	158.24

```
In [26]: round(pd.DataFrame(tweet_list_df.groupby("sentiment").text_word_count.mean()),2)
```

```
Out[26]:
```

sentiment	text_word count
negative	24.97
neutral	18.28
positive	25.55

Saving Output Tweets File to Drive

```
In [27]: tweet_list_df.to_csv("../c2_sentimentanalysis_output.csv", sep=';', encoding='UTF-8')
```

```
Out[27]:
```

	text	cleaned	polarity	subjectivity	sentiment	neg	neu	pos	compound	text_len	text_word_count
0	So good I had to share! Check out all the item...	so good had to share check out all the items m...	0.650000	0.775000	positive	0.000	0.647	0.353	0.8516	130	21
1	#Covid-19: #Apple Will Lose 6 Million #PhoneL...	covid 19 apple will lose 6 million iphone14pr...	0.000000	0.000000	negative	0.251	0.749	0.000	-0.5719	104	17
2	Pre-save my new album "C" on Spotify: https://...	pre save my new album on spotify powered by di...	0.318182	0.477273	positive	0.000	0.842	0.158	0.4939	110	18
3	It's #PodcastAndChill Monday/vr/#Wrestling Fan...	it podcastandchill monday wrestling fans downl...	0.750000	0.600000	positive	0.000	0.789	0.211	0.7717	180	27
4	Samsung N10 plus A+ quality display, in stock...	samsung n10 plus quality display in stock inte...	0.250000	0.500000	positive	0.000	0.857	0.143	0.6124	211	32