

```
In [1]: # Sentiment Analysis with Data from Twitter
        ## To replicate the sentiment analysis with data from twitter:
        ### A. Start a notebook instance in machine Learning Lab
        ### B. Open the example link and download the source code from Github repository.
        ### C. Sign up for <a href="https://developer.twitter.com/en/support/twitter-api/de
        ### D. Copy, paste and modify the python code file in the notebook
        ### E. Push the repository to GitHub
```

Step 1: Install and Import Libraries

```
In [2]: # Install Libraries
        !pip install textblob
        !pip install tweepy
        !pip install pycountry
        !pip install wordcloud
        !pip install langdetect
```

Looking in indexes: <https://pypi.org/simple>, <https://pip.repos.neuron.amazonaws.com>

Requirement already satisfied: textblob in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (0.17.1)

Requirement already satisfied: nltk>=3.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from textblob) (3.8.1)

Requirement already satisfied: joblib in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from nltk>=3.1->textblob) (1.2.0)

Requirement already satisfied: tqdm in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from nltk>=3.1->textblob) (4.64.1)

Requirement already satisfied: click in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from nltk>=3.1->textblob) (8.1.3)

Requirement already satisfied: regex>=2021.8.3 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from nltk>=3.1->textblob) (2022.10.31)

Looking in indexes: <https://pypi.org/simple>, <https://pip.repos.neuron.amazonaws.com>

Requirement already satisfied: tweepy in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (4.14.0)

Requirement already satisfied: requests-oauthlib<2,>=1.2.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from tweepy) (1.3.1)

Requirement already satisfied: requests<3,>=2.27.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from tweepy) (2.28.1)

Requirement already satisfied: oauthlib<4,>=3.2.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from tweepy) (3.2.2)

Requirement already satisfied: certifi>=2017.4.17 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests<3,>=2.27.0->tweepy) (2022.12.7)

Requirement already satisfied: charset-normalizer<3,>=2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests<3,>=2.27.0->tweepy) (2.1.1)

Requirement already satisfied: urllib3<1.27,>=1.21.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests<3,>=2.27.0->tweepy) (1.26.8)

Requirement already satisfied: idna<4,>=2.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from requests<3,>=2.27.0->tweepy) (3.4)

Looking in indexes: <https://pypi.org/simple>, <https://pip.repos.neuron.amazonaws.com>

Requirement already satisfied: pycountry in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (22.3.5)

Requirement already satisfied: setuptools in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from pycountry) (65.6.3)

Looking in indexes: <https://pypi.org/simple>, <https://pip.repos.neuron.amazonaws.com>

Requirement already satisfied: wordcloud in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (1.9.1.1)

Requirement already satisfied: matplotlib in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from wordcloud) (3.6.2)

Requirement already satisfied: pillow in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from wordcloud) (9.4.0)

Requirement already satisfied: numpy>=1.6.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from wordcloud) (1.22.3)

Requirement already satisfied: fonttools>=4.22.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from matplotlib->wordcloud) (4.38.0)

Requirement already satisfied: packaging>=20.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from matplotlib->wordcloud) (21.3)

Requirement already satisfied: cycler>=0.10 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from matplotlib->wordcloud) (0.10.0)

n3/lib/python3.10/site-packages (from matplotlib->wordcloud) (0.11.0)
 Requirement already satisfied: contourpy>=1.0.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from matplotlib->wordcloud) (1.0.6)
 Requirement already satisfied: pyparsing>=2.2.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from matplotlib->wordcloud) (3.0.9)
 Requirement already satisfied: kiwisolver>=1.0.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from matplotlib->wordcloud) (1.4.4)
 Requirement already satisfied: python-dateutil>=2.7 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from matplotlib->wordcloud) (2.8.2)
 Requirement already satisfied: six>=1.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
 Looking in indexes: https://pypi.org/simple, https://pip.repos.neuron.amazonaws.com
 Requirement already satisfied: langdetect in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (1.0.9)
 Requirement already satisfied: six in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (from langdetect) (1.16.0)

In [3]: *# 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
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.download('vader_lexicon')
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data] /home/ec2-user/nltk_data...
[nltk_data] Package vader_lexicon is already up-to-date!
```

Out[3]: True

Step 2: Authentication for Twitter API

In [5]: *# Authentication*

```
consumerKey = "2CJbp8m7Rg26LtZEKAeVXWceG"
consumerSecret = "IbULWK1S0uvhLnMTcFkzvMjDfZUCBM6eVbdLYMUbnafhieX8Y6"
accessToken = "1119095264961126402-8QDK0FvTjGmV7w1hUtGsTyCyH77bNE"
accessTokenSecret = "unheMaBnZsHrcBt3mK1bE8Ib4wtE3tPtAwSA65Y0zWIK0"
```

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

Step 3: Getting Tweets With Keyword or Hashtag

```
In [24]: #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')
```

In [25]: *#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: 20  
positive number: 2  
negative number: 0  
neutral number: 18
```

In [26]: tweet_list[0:20]

Out[26]:

0

- 0 RT @FiloSociedadUCM: Seminario Permanente de E...
- 1 Los amo personas con quien puedes ser tu versi...
- 2 RT @Peli_Comic: El UCM empezó hace 15 años. Yo...
- 3 RT @BComplutense: Inscríbete en el curso en lí...
- 4 RT @alaesterevista: ❤️ ¡Nueva revista de Creació...
- 5 ❤️ ¡Nueva revista de Creación Literaria! ❤️ \n\nAbr...
- 6 Minha linguagem de amor: obrigar a pessoa a as...
- 7 @SitoCinema En el UCM la cagaron por saltarse ...
- 8 RT @quetalquita: #Noticias 📰 🌋 | Vulcanólogo UCM...
- 9 @DisneyHaters Otros dicen que es Bryan singer ...
- 10 RT @digital_RGS: Learn more about Roy's work:\n...
- 11 CLUB MURANO SE ADJUDICÓ COPA "UCM 2023" https:...
- 12 RT @csicdanza: Nuestra compañera Irene López A...
- 13 RT @csicdanza: La doctora en filosofía @IbisAl...
- 14 #Noticias 📰 🌋 | Vulcanólogo UCM llama a la tranq...
- 15 RT @sshbysevensins: [SSH SEASON II vol.1 RUNWAY...
- 16 @Zack23304374 @elojaneiro @Handsome_Wolf @Cava...
- 17 RT @ujc_ucmc: Las Brigadas Técnicas Juveniles ...
- 18 RT @BiblioMadSalud: Ya está abierta la inscrip...
- 19 RT @ujc_ucmc: Las Brigadas Técnicas Juveniles ...

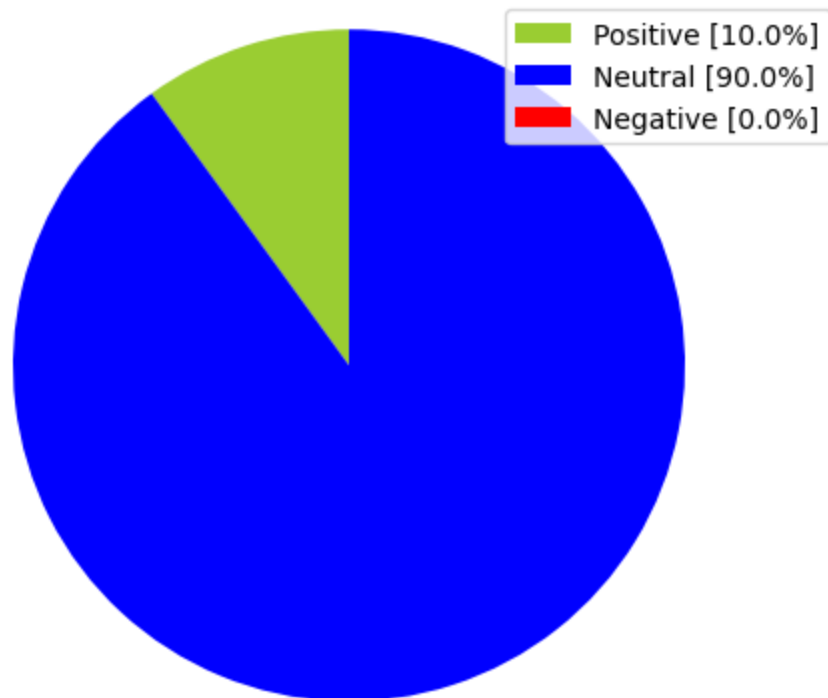
In [27]: #Creating PieCart

```

labels = ['Positive [' + str(positive) + '%]' , 'Neutral [' + str(neutral) + '%]', '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= ucm



```
In [28]: tweet_list.drop_duplicates(inplace = True)
```

Extracting text values

```
In [29]: tw_list = pd.DataFrame(tweet_list)
tw_list["text"] = tw_list[0]
tw_list
```

Out[29]:

0

text

0	RT @FiloSociedadUCM: Seminario Permanente de E...	RT @FiloSociedadUCM: Seminario Permanente de E...
1	Los amo personas con quien puedes ser tu versi...	Los amo personas con quien puedes ser tu versi...
2	RT @Peli_Comic: El UCM empezó hace 15 años. Yo...	RT @Peli_Comic: El UCM empezó hace 15 años. Yo...
3	RT @BComplutense: Insíbete en el curso en lí...	RT @BComplutense: Insíbete en el curso en lí...
4	RT @alaesterevista: ❤️ ¡Nueva revista de Creació...	RT @alaesterevista: ❤️ ¡Nueva revista de Creació...
5	❤️ ¡Nueva revista de Creación Literaria! ❤️ \n\nAbr...	❤️ ¡Nueva revista de Creación Literaria! ❤️ \n\nAbr...
6	Minha linguagem de amor: obrigar a pessoa a as...	Minha linguagem de amor: obrigar a pessoa a as...
7	@SitoCinema En el UCM la cagaron por saltarse ...	@SitoCinema En el UCM la cagaron por saltarse ...
8	RT @quetalquita: #Noticias 📰 🌋 Vulcanólogo UCM...	RT @quetalquita: #Noticias 📰 🌋 Vulcanólogo UCM...
9	@DisneyHaters Otros dicen que es Bryan singer ...	@DisneyHaters Otros dicen que es Bryan singer ...
10	RT @digital_RGS: Learn more about Roy's work:\n...	RT @digital_RGS: Learn more about Roy's work:\n...
11	CLUB MURANO SE ADJUDICÓ COPA "UCM 2023" https:...	CLUB MURANO SE ADJUDICÓ COPA "UCM 2023" https:...
12	RT @csicdanza: Nuestra compañera Irene López A...	RT @csicdanza: Nuestra compañera Irene López A...
13	RT @csicdanza: La doctora en filosofía @IbisAl...	RT @csicdanza: La doctora en filosofía @IbisAl...
14	#Noticias 📰 🌋 Vulcanólogo UCM llama a la tranq...	#Noticias 📰 🌋 Vulcanólogo UCM llama a la tranq...
15	RT @sshbysevensins: [SSH SEASON II vol.1 RUNWAY...	RT @sshbysevensins: [SSH SEASON II vol.1 RUNWAY...
16	@Zack23304374 @elojaneiro @Handsome_Wolf @Cava...	@Zack23304374 @elojaneiro @Handsome_Wolf @Cava...
17	RT @ujc_ucmc: Las Brigadas Técnicas Juveniles ...	RT @ujc_ucmc: Las Brigadas Técnicas Juveniles ...
18	RT @BiblioMadSalud: Ya está abierta la inscrip...	RT @BiblioMadSalud: Ya está abierta la inscrip...

In [11]: *#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)
```

Out[11]:

	0	text
0	RT @daxmonestime55: thanks for the camp invite...	thanks for the camp invite i hope i can get...
1	Fall in love with this 1 BD/ 2 BA in Warrensbu...	fall in love with this 1 bd 2 ba in warrensbu...
2	Revered Warrensburg coach/teacher recalled at ...	revered warrensburg coach teacher recalled at ...
4	RT @OGHS_Soccer: OG-5\nWarrensburg-1\n\nGoal S...	og 5 warrensburg 1 goal scorers kenzie baxl...
5	Spaces for lease at Santa Fe Trail Shopping Ce...	spaces for lease at santa fe trail shopping ce...
9	RT @iamstevencooper: #MidwestMade Record Label...	midwestmade record label launch party apri...
13	thanks for the camp invite @JonesgGreg i hope ...	thanks for the camp invite i hope i can get ...
14	RT @mrwagner25: Jarius Ingram was such a fun p...	jarius ingram was such a fun person to cover ...
15	OG-5\nWarrensburg-1\n\nGoal Scorers:\nKenzie B...	og 5 warrensburg 1 goal scorers kenzie baxle...
16	Jarius Ingram was such a fun person to cover a...	jarius ingram was such a fun person to cover a...

In [30]: *#Calculating Negative, Positive, Neutral and Compound values*

```
tw_list[['polarity', 'subjectivity']] = tw_list['text'].apply(lambda Text: pd.Series
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)
```

```
/tmp/ipykernel_7113/4219357815.py:4: FutureWarning: iteritems is deprecated and wi
ll be removed in a future version. Use .items instead.
    for index, row in tw_list['text'].iteritems():
```

Out[30]:

	0	text	polarity	subjectivity	sentiment	neg	
0	RT @FiloSociedadUCM: Seminario Permanente de E...	RT @FiloSociedadUCM: Seminario Permanente de E...	0.0	0.0	neutral	0.0	1.
1	Los amo personas con quien puedes ser tu versi...	Los amo personas con quien puedes ser tu versi...	0.0	0.0	neutral	0.0	1.
2	RT @Peli_Comic: El UCM empezó hace 15 años. Yo...	RT @Peli_Comic: El UCM empezó hace 15 años. Yo...	0.0	0.0	neutral	0.0	1.
3	RT @BComplutense: Insíbete en el curso en lí...	RT @BComplutense: Insíbete en el curso en lí...	0.0	0.0	neutral	0.0	1.
4	RT @alaesterevista: ¡Nueva revista de Creació...	RT @alaesterevista: ¡Nueva revista de Creació...	0.0	0.0	neutral	0.0	1.
5	¡Nueva revista de Creación Literaria! \n\nAbr...	¡Nueva revista de Creación Literaria! \n\nAbr...	0.0	0.0	neutral	0.0	1.
6	Minha linguagem de amor: obrigar a pessoa a as...	Minha linguagem de amor: obrigar a pessoa a as...	0.0	0.0	positive	0.0	0.
7	@SitoCinema En el UCM la cagaron por saltarse ...	@SitoCinema En el UCM la cagaron por saltarse ...	0.0	0.0	neutral	0.0	1.
8	RT @quetalquita: #Noticias Vulcanólogo UCM...	RT @quetalquita: #Noticias Vulcanólogo UCM...	0.0	0.0	neutral	0.0	1.
9	@DisneyHaters Otros dicen que es Brvan singer ...	@DisneyHaters Otros dicen que es Brvan singer ...	0.0	0.0	positive	0.0	0.

In [31]: *#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"]
```

In [32]: *#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)*
    return pd.concat([total,percentage],axis=1,keys=['Total','Percentage'])
```

In [33]: *#Count_values for sentiment*
count_values_in_column(tw_list,"sentiment")

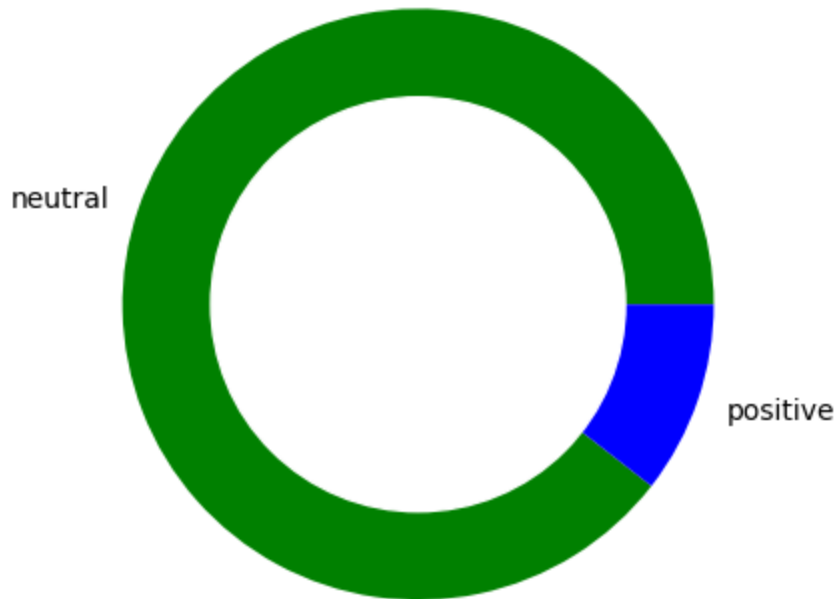
Out[33]:

	Total	Percentage
neutral	17	89.47
positive	2	10.53

In [34]: *# create data for Pie Chart*
pc = count_values_in_column(tw_list,"sentiment")
names= pc.index

```
size=pc["Percentage"]

# Create a circle for the center of the plot
my_circle=plt.Circle( (0,0), 0.7, color='white')
plt.pie(size, labels=names, colors=['green','blue','red'])
p=plt.gcf()
p.gca().add_artist(my_circle)
plt.show()
```



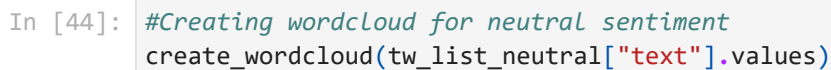
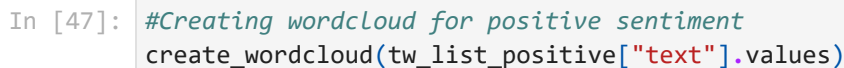
In [41]: *#Function to Create Wordcloud*

```
def create_wordcloud(text):
    stopwords = set(STOPWORDS)
    wc = WordCloud(background_color="white", width=600, height=400,
                    max_words=300,
                    stopwords=stopwords,
                    repeat=True)
    wc.generate(str(text))

    plt.imshow(wc)
    plt.show()
```

In [46]: *#Creating wordcloud for all tweets*

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





In []: