

```
In [1]: # 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.co
m
Collecting textblob
  Downloading textblob-0.17.1-py2.py3-none-any.whl (636 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 636.8/636.8 kB 24.5 MB/s eta 0:00:00
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: regex>=2021.8.3 in /home/ec2-user/anaconda3/envs/py
thon3/lib/python3.10/site-packages (from nltk>=3.1->textblob) (2022.10.31)
Requirement already satisfied: joblib in /home/ec2-user/anaconda3/envs/python3/li
b/python3.10/site-packages (from nltk>=3.1->textblob) (1.2.0)
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: tqdm in /home/ec2-user/anaconda3/envs/python3/lib/p
ython3.10/site-packages (from nltk>=3.1->textblob) (4.64.1)
Installing collected packages: textblob
Successfully installed textblob-0.17.1
Looking in indexes: https://pypi.org/simple, https://pip.repos.neuron.amazonaws.co
m
Collecting tweepy
  Downloading tweepy-4.14.0-py3-none-any.whl (98 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 98.5/98.5 kB 4.4 MB/s eta 0:00:00
Requirement already satisfied: requests<3,>=2.27.0 in /home/ec2-user/anaconda3/env
s/python3/lib/python3.10/site-packages (from tweepy) (2.28.1)
Collecting requests-oauthlib<2,>=1.2.0
  Downloading requests_oauthlib-1.3.1-py2.py3-none-any.whl (23 kB)
Collecting oauthlib<4,>=3.2.0
  Downloading oauthlib-3.2.2-py3-none-any.whl (151 kB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 151.7/151.7 kB 27.1 MB/s eta 0:00:00
Requirement already satisfied: certifi>=2017.4.17 in /home/ec2-user/anaconda3/env
s/python3/lib/python3.10/site-packages (from requests<3,>=2.27.0->tweepy) (2022.1
2.7)
Requirement already satisfied: charset-normalizer<3,>=2 in /home/ec2-user/anaconda
3/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/e
nvs/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/pytho
n3/lib/python3.10/site-packages (from requests<3,>=2.27.0->tweepy) (3.4)
Installing collected packages: oauthlib, requests-oauthlib, tweepy
Successfully installed oauthlib-3.2.2 requests-oauthlib-1.3.1 tweepy-4.14.0
Looking in indexes: https://pypi.org/simple, https://pip.repos.neuron.amazonaws.co
m
Collecting pycountry
  Downloading pycountry-22.3.5.tar.gz (10.1 MB)
  ━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 10.1/10.1 kB 81.5 MB/s eta 0:00:00
0:0100:01
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Requirement already satisfied: setuptools in /home/ec2-user/anaconda3/envs/python
3/lib/python3.10/site-packages (from pycountry) (65.6.3)
Building wheels for collected packages: pycountry
  Building wheel for pycountry (pyproject.toml) ... done
  Created wheel for pycountry: filename=pycountry-22.3.5-py2.py3-none-any.whl size
```

```
=10681832 sha256=cf7c8da095166c5f85109a71b9ef4b1f54dd018297640c634385960a650bddd8  
Stored in directory: /home/ec2-user/.cache/pip/wheels/62/4a/9c/7a46699df9efb845aa116fae5e52d8690fc442fef6d32213f7  
Successfully built pycountry  
Installing collected packages: pycountry  
Successfully installed pycountry-22.3.5  
Looking in indexes: https://pypi.org/simple, https://pip.repos.neuron.amazonaws.co  
m  
Collecting wordcloud  
  Downloading wordcloud-1.9.1.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x8  
6_64.whl (455 kB) 455.4/455.4 kB 21.2 MB/s eta 0:00:00  
Requirement already satisfied: matplotlib in /home/ec2-user/anaconda3/envs/python  
3/lib/python3.10/site-packages (from wordcloud) (3.6.2)  
Requirement already satisfied: pillow in /home/ec2-user/anaconda3/envs/python3/li  
b/python3.10/site-packages (from wordcloud) (9.4.0)  
Requirement already satisfied: numpy>=1.6.1 in /home/ec2-user/anaconda3/envs/pytho  
n3/lib/python3.10/site-packages (from wordcloud) (1.22.3)  
Requirement already satisfied: pyparsing>=2.2.1 in /home/ec2-user/anaconda3/envs/p  
ython3/lib/python3.10/site-packages (from matplotlib->wordcloud) (3.0.9)  
Requirement already satisfied: cycler>=0.10 in /home/ec2-user/anaconda3/envs/pytho  
n3/lib/python3.10/site-packages (from matplotlib->wordcloud) (0.11.0)  
Requirement already satisfied: python-dateutil>=2.7 in /home/ec2-user/anaconda3/en  
vs/python3/lib/python3.10/site-packages (from matplotlib->wordcloud) (2.8.2)  
Requirement already satisfied: packaging>=20.0 in /home/ec2-user/anaconda3/envs/py  
thon3/lib/python3.10/site-packages (from matplotlib->wordcloud) (21.3)  
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: 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: contourpy>=1.0.1 in /home/ec2-user/anaconda3/envs/p  
ython3/lib/python3.10/site-packages (from matplotlib->wordcloud) (1.0.6)  
Requirement already satisfied: six>=1.5 in /home/ec2-user/anaconda3/envs/python3/l  
ib/python3.10/site-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.  
16.0)  
Installing collected packages: wordcloud  
Successfully installed wordcloud-1.9.1.1  
Looking in indexes: https://pypi.org/simple, https://pip.repos.neuron.amazonaws.co  
m  
Collecting langdetect  
  Downloading langdetect-1.0.9.tar.gz (981 kB) 981.5/981.5 kB 33.1 MB/s eta 0:00:00  
    Preparing metadata (setup.py) ... done  
Requirement already satisfied: six in /home/ec2-user/anaconda3/envs/python3/lib/py  
thon3.10/site-packages (from langdetect) (1.16.0)  
Building wheels for collected packages: langdetect  
  Building wheel for langdetect (setup.py) ... done  
    Created wheel for langdetect: filename=langdetect-1.0.9-py3-none-any.whl size=99  
3225 sha256=834745fb9c2059bdd21534ed1fa94c5db902e82d6d40e6ac03ca1578d5cdf82  
    Stored in directory: /home/ec2-user/.cache/pip/wheels/c4/16/af/1889804d8b7c0c041  
caddee8e29673a938a332acbf2865c70a1  
Successfully built langdetect  
Installing collected packages: langdetect  
Successfully installed langdetect-1.0.9
```

In [2]: # 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...
```

Out[2]: True

Step 2: Authentication for Twitter API (tokens from my twitter account)

In [8]: # 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 [9]: #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 [10]: #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: 1000
positive number: 93
negative number: 158
neutral number: 749

```

In [11]: `tweet_list[0:20]`

Out[11]:

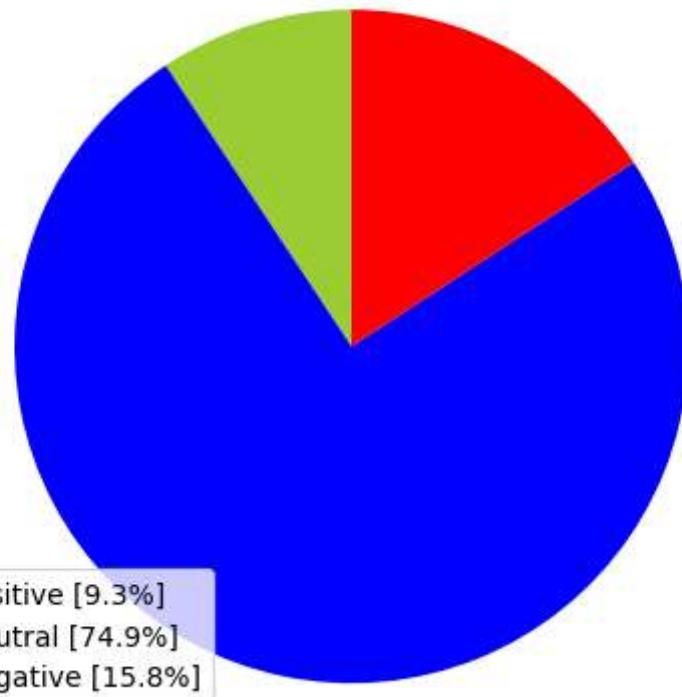
0

- 0** RT @HenryOmarVC1: La @ucmvccu presente jóvenes...
- 1** RT @Indie5051: Oficialmente la mejor trilogía ...
- 2** RT @Indie5051: Oficialmente la mejor trilogía ...
- 3** RT @Indie5051: Oficialmente la mejor trilogía ...
- 4** RT @Indie5051: Oficialmente la mejor trilogía ...
- 5** Guardianes de la galaxia vol 3. 9/10\nQué mal ...
- 6** RT @DepDocHppss: #SanctiSpiritusEnMarcha prese...
- 7** Descubre las claves de la Beca MEFP antes de...
- 8** RT @atorices_angie: El lunes 7 se abre la prei...
- 9** RT @Indie5051: Oficialmente la mejor trilogía ...
- 10** RT @JL_MarceloP: 16 ed. Jornadas contra la Hom...
- 11** RT @HenryOmarVC1: La @ucmvccu presente jóvenes...
- 12** Tenemos ya muy normalizado el UCM y el cine de...
- 13** RT @zebazcr: Ni se estrena y ya la odian. 😢\n...
- 14** RT @MarvelLatin: Top 5 VILLANOS del UCM https:...
- 15** @hpsscc @Aliosky72982660 @AvellaRivero @DirecS...
- 16** no puede ser posible que una sola película de ...
- 17** RT @Indie5051: Oficialmente la mejor trilogía ...
- 18** Sin pedos les puedo decir que "Guardianes de I...
- 19** Las dos mejores trilogías del #UCM sin duda ...

In [12]: `#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 [13]: tweet_list.drop_duplicates(inplace = True)
```

Extracting text values

```
In [14]: tw_list = pd.DataFrame(tweet_list)
tw_list["text"] = tw_list[0]
tw_list
```

Out[14]:

	0	text
0	RT @HenryOmarVC1: La @ucmvccu presente jóvenes...	RT @HenryOmarVC1: La @ucmvccu presente jóvenes...
1	RT @Indie5051: Oficialmente la mejor trilogía ...	RT @Indie5051: Oficialmente la mejor trilogía ...
5	Guardianes de la galaxia vol 3. 9/10\nQué mal ...	Guardianes de la galaxia vol 3. 9/10\nQué mal ...
6	RT @DepDocHppss: #SanctiSpiritusEnMarcha prese...	RT @DepDocHppss: #SanctiSpiritusEnMarcha prese...
7	💡 Descubre las claves de la Beca MEFP antes de...	💡 Descubre las claves de la Beca MEFP antes de...
...
982	Paloma Bel, @pbeldura , directora de @aecoop_u...	Paloma Bel, @pbeldura , directora de @aecoop_u...
991	#empathy #seminar #empathyseminar #EspacioRojo...	#empathy #seminar #empathyseminar #EspacioRojo...
992	@adoyma @DoloresCabra97 @amauthausen @IStolper...	@adoyma @DoloresCabra97 @amauthausen @IStolper...
996	@DerechoEspana Otra magnífica semblanza fue la...	@DerechoEspana Otra magnífica semblanza fue la...
997	Guardianes de la Galaxia es lo último que nos ...	Guardianes de la Galaxia es lo último que nos ...

337 rows × 2 columns

In [15]:

```
#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]+|([^\w-9A-Za-z \t])|(\w+:\//\w+)", " ", 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[15]:

	0	text
0	RT @HenryOmarVC1: La @ucmvccu presente jóvenes...	la presente j venes estudiantes trabajador...
1	RT @Indie5051: Oficialmente la mejor trilogía ...	oficialmente la mejor trilogía del ucm
5	Guardianes de la galaxia vol 3. 9/10\nQué mal ...	guardianes de la galaxia vol 3 9 10 qu mal l...
6	RT @DepDocHppss: #SanctiSpiritusEnMarcha prese...	sanctispiritusenmarcha presentes trabajadore...
7	💡 Descubre las claves de la Beca MEFP antes de...	descubre las claves de la beca mefp antes de...
8	RT @atorices_angie: El lunes 7 se abre la prei...	el lunes 7 se abre la preinscripción para el ...
10	RT @JL_MarceloP: 16 ed. Jornadas contra la Hom...	16 ed jornadas contra la homofobia y la tra...
12	Tenemos ya muy normalizado el UCM y el cine de...	tenemos ya muy normalizado el ucm y el cine de...
13	RT @zebazcr: Ni se estrena y ya la odian. 😢\n\...	ni se estrena y ya la odian themarvels ...
14	RT @MarvelLatin: Top 5 VILLANOS del UCM https:...	top 5 villanos del ucm

In [16]: *#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_8969/4219357815.py:4: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.
for index, row in tw_list['text'].iteritems():

	0	text	polarity	subjectivity	sentiment	neg	neu
0	RT @HenryOmarVC1: La @ucmvccu presente jóvenes...	la presente j venes estudiantes trabajador...	0.00	0.0	negative	0.155	0.845 0.
1	RT @Indie5051: Oficialmente la mejor trilogía ...	oficialmente la mejor trilogía del ucm	0.00	0.0	neutral	0.000	1.000 0.
5	Guardianes de la galaxia vol 3. 9/10\nQué mal ...	guardianes de la galaxia vol 3 9 10 qué mal l...	0.00	0.0	neutral	0.000	1.000 0.
6	RT @DepDocHppss: #SanctiSpiritusEnMarcha prese...	sanctispiritusenmarcha presentes trabajadore...	0.00	0.0	neutral	0.000	1.000 0.
7	💡 Descubre las claves de la Beca MEFP antes de...	descubre las claves de la beca mefp antes de...	0.00	0.0	neutral	0.000	1.000 0.
8	RT @atorices_angie: El lunes 7 se abre la prei...	el lunes 7 se abre la preinscripción para el ...	0.00	0.0	neutral	0.000	1.000 0.
10	RT @JL_MarceloP: 16 ed. Jornadas contra la Hom...	16 ed jornadas contra la homofobia y la tra...	0.00	0.0	positive	0.000	0.846 0.
12	Tenemos ya muy normalizado el UCM y el cine de...	tenemos ya muy normalizado el ucm y el cine de...	0.05	0.5	neutral	0.000	1.000 0.
13	RT @zebazcr: Ni se estrena y ya la odian. 😢\n...	ni se estrena y ya la odian themarvels ...	0.00	0.0	neutral	0.000	1.000 0.
14	RT @MarvelLatin: Top 5 VILLANOS del UCM https:	top 5 villanos del ucm	0.50	0.5	positive	0.000	0.625 0.

In [17]: `#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 [18]: `#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 [19]: `#Count_values for sentiment`
`count_values_in_column(tw_list,"sentiment")`

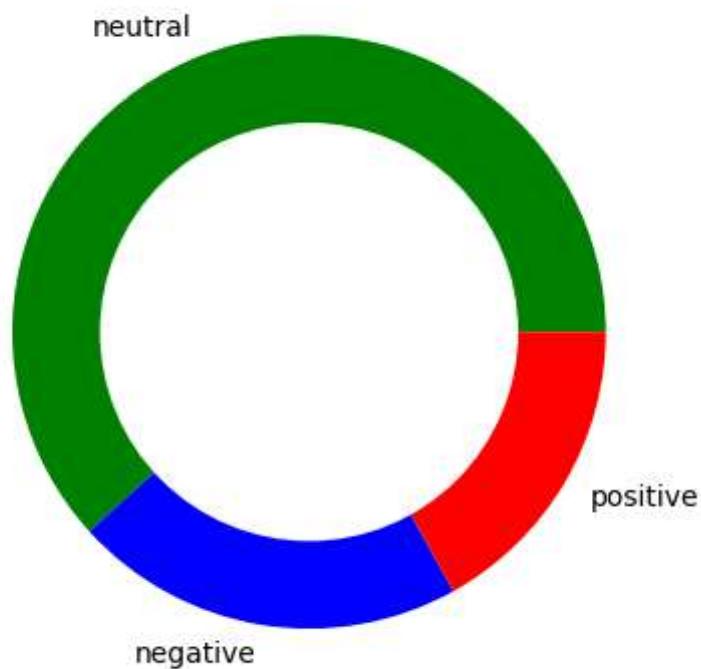
Out[19]:

	Total	Percentage
neutral	208	61.72
negative	72	21.36
positive	57	16.91

In [20]:

```
# 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 [21]: #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 [22]: #Creating wordcloud for all tweets  
create_wordcloud(tw_list[\"text\"].values)
```



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



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



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



In []: