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
Requirement already satisfied: textblob in /home/ec2-user/anaconda3/envs/python3/l
ib/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: tqdm in /home/ec2-user/anaconda3/envs/python3/lib/p
ython3.10/site-packages (from nltk>=3.1->textblob) (4.64.1)
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: click in /home/ec2-user/anaconda3/envs/python3/lib/
python3.10/site-packages (from nltk>=3.1->textblob) (8.1.3)
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)
Looking in indexes: https://pypi.org/simple, https://pip.repos.neuron.amazonaws.co
m
Requirement already satisfied: tweepy in /home/ec2-user/anaconda3/envs/python3/li
b/python3.10/site-packages (4.14.0)
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)
Requirement already satisfied: requests-oauthlib<2,>=1.2.0 in /home/ec2-user/anaco
nda3/envs/python3/lib/python3.10/site-packages (from tweepy) (1.3.1)
Requirement already satisfied: oauthlib<4,>=3.2.0 in /home/ec2-user/anaconda3/env
s/python3/lib/python3.10/site-packages (from tweepy) (3.2.2)
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: 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: 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)
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)
Looking in indexes: https://pypi.org/simple, https://pip.repos.neuron.amazonaws.co
m
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/python
3/lib/python3.10/site-packages (from pycountry) (65.6.3)
Looking in indexes: https://pypi.org/simple, https://pip.repos.neuron.amazonaws.co
m
Requirement already satisfied: wordcloud in /home/ec2-user/anaconda3/envs/python3/
lib/python3.10/site-packages (1.9.1.1)
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: 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: 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: packaging>=20.0 in /home/ec2-user/anaconda3/envs/py
thon3/lib/python3.10/site-packages (from matplotlib->wordcloud) (21.3)
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: 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: 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: 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: 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)
Looking in indexes: https://pypi.org/simple, https://pip.repos.neuron.amazonaws.co
m
Requirement already satisfied: langdetect in /home/ec2-user/anaconda3/envs/python
3/lib/python3.10/site-packages (1.0.9)
Requirement already satisfied: six in /home/ec2-user/anaconda3/envs/python3/lib/py
thon3.10/site-packages (from langdetect) (1.16.0)
```

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...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

Out[2]: True

Step 2: Authentication for Twitter API

In [4]:
```
# Authentication
consumerKey = "2CJbp8m7Rg26LtZEKAeVXWceG"
consumerSecret = "IbULWKlS0uvhLnMTcFkzvMjDfZUcBM6eVbdLYMUbNafhieX8Y6"
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 [6]:
```
#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 [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:  1000
positive number:  258
negative number:  250
neutral number:  492
```

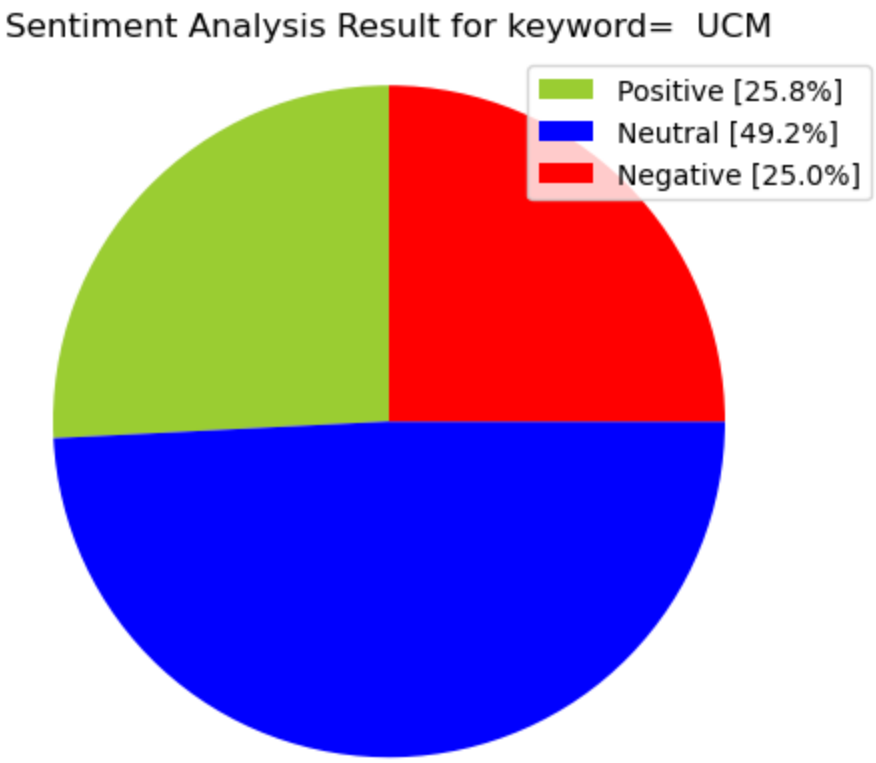
In [8]:
```
tweet_list[0:20]
```

Out[8]:

	0
0	@ss_ucm @JL_MarceloP @JLBrocheLorenzo @DeivyPr...
1	@fernandemiguels Pior que pensando no todo, de...
2	RT @peterjordan100: Saiu video do Raluca e a g...
3	RT @habitaciondcine: Gunn deja el UCM por la p...
4	RT @Multiverso_GK: Y así cierra, una de las me...
5	@FalandodeHerois O UCM nunca vai acabar e desi...
6	Las 31 películas y las 12 series del UCM 🤣 htt...
7	@FalandodeHerois Adeus é um exagero. Tenho cer...
8	@JesusTeAbomina @MarvelBRNews é interessante s...
9	@FalandodeHerois O que de empolgaria então? Sé...
10	RT @Multiverso_GK: Y así cierra, una de las me...
11	RT @peterjordan100: Geral falando bem de Guard...
12	#GuardiansOfTheGalaxyVol3 es un viaje hermoso ...
13	@DanielsMaru A quien le importa el trato justo...
14	RT @Multiverso_GK: Y así cierra, una de las me...
15	RT @peterjordan100: Geral falando bem de Guard...
16	RT @OS_BirdHouse: Huge congrats to our 4 FFC m...
17	RT @Multiverso_GK: Y así cierra, una de las me...
18	RT @zebazcr: Ni se estrena y ya la odian. 😂\n\...
19	¿Mejor trilogía del UCM de Marvel?

In [9]:
```
#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()
```



In [11]:
```
tweet_list.drop_duplicates(inplace = True)
```

In [ ]:
```
#Extracting text values
```

In [27]:
```
tw_list = pd.DataFrame(tweet_list)
tw_list["text"] = tw_list[0]
tw_list
```

Out[27]:

	0	text	polarity	subjectivity	sentiment	neg	
0	@ss_ucm @JL_MarceloP @JLBrocheLorenzo @DeivyPr...	@ss_ucm @JL_MarceloP @JLBrocheLorenzo @DeivyPr...	0.0	0.0	neutral	0.000	1.
1	@fernandemiguels Pior que pensando no todo, de...	@fernandemiguels Pior que pensando no todo, de...	0.0	0.0	negative	0.121	0.
2	RT @peterjordan100: Saiu video do Raluca e a g...	RT @peterjordan100: Saiu video do Raluca e a g...	0.0	0.0	neutral	0.000	1.
3	RT @habitaciondcine: Gunn deja el UCM por la p...	RT @habitaciondcine: Gunn deja el UCM por la p...	0.0	0.0	neutral	0.000	1.
4	RT @Multiverso_GK: Y así cierra, una de las me...	RT @Multiverso_GK: Y así cierra, una de las me...	0.0	0.0	neutral	0.000	1.
...	...	...	...	...	...	...	
990	RT @ETECSA_Cuba: ATENCIÓN!!\nNuevamente circul...	RT @ETECSA_Cuba: ATENCIÓN!!\nNuevamente circul...	0.0	0.0	neutral	0.000	1.
992	RT @centrofidel: #Fidel:"Existe todavía el imp...	RT @centrofidel: #Fidel:"Existe todavía el imp...	0.0	0.0	neutral	0.000	1.
994	No he visto la pelicula pero dicen muchos que ...	No he visto la pelicula pero dicen muchos que ...	0.0	0.0	negative	0.162	0.
996	#ALaPatriaManosYCorazón https://t.co/WsUdrhmf7J	#ALaPatriaManosYCorazón https://t.co/WsUdrhmf7J	0.0	0.0	neutral	0.000	1.
998	RT @dsn: HOY ➡️ 04mayo2023 «Jornadas sobre el #S...	RT @dsn: HOY ➡️ 04mayo2023 «Jornadas sobre el #S...	0.0	0.0	neutral	0.000	1.

349 rows × 9 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]+)|([^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[15]:

	0	text
0	@ss_ucm @JL_MarceloP @JLBrocheLorenzo @DeivyPr...	ucm marcelop
1	@fernandemiguels Pior que pensando no todo, de...	pior que pensando no todo devo concordar a...
2	RT @peterjordan100: Saiu video do Raluca e a g...	saiu video do raluca e a galera ta mais hypad...
3	RT @habitaciondcine: Gunn deja el UCM por la p...	gunn deja el ucm por la puerta grande situan...
4	RT @Multiverso_GK: Y así cierra, una de las me...	y as cierra una de las mejores trilog as de...
5	@FalandodeHerois O UCM nunca vai acabar e desi...	o ucm nunca vai acabar e desistir assim um...
6	Las 31 películas y las 12 series del UCM 🤣 htt...	las 31 pel culas y las 12 series del ucm
7	@FalandodeHerois Adeus é um exagero. Tenho cer...	adeus um exagero tenho certeza de que o u...
8	@JesusTeAbomina @MarvelBRNews é interessante s...	interessante sim e sem d vida o melhor ...
9	@FalandodeHerois O que de empolgaria então? Sé...	o que de empolgaria ent o s rio que ainda t...

In [16]:

```
#Calculating Negative, Positive, Neutral and Compound values

tw_list[['polarity', 'subjectivity']] = tw_list['text'].apply(lambda Text: pd.Serie
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_18559/4219357815.py:4: FutureWarning: iteritems is deprecated and w
ill be removed in a future version. Use .items instead.
  for index, row in tw_list['text'].iteritems():
```

Out[16]:

	0	text	polarity	subjectivity	sentiment	neg	neu	pos	compound
0	@ss_ucm @JL_MarceloP @JLBrocheLorenzo @DeivyPr...	ucm marcelop	0.0	0.0	neutral	0.000	1.000	0.000	0.0000
1	@fernandemiguels Pior que pensando no todo, de...	pior que pensando no todo devo concordar a...	0.0	0.0	negative	0.121	0.879	0.000	-0.2960
2	RT @peterjordan100: Saiu video do Raluca e a g...	saiu video do raluca e a galera ta mais hypad...	0.0	0.0	neutral	0.000	1.000	0.000	0.0000
3	RT @habitaciondcine: Gunn deja el UCM por la p...	gunn deja el ucm por la puerta grande situan...	0.0	0.0	neutral	0.000	1.000	0.000	0.0000
4	RT @Multiverso_GK: Y así cierra, una de las me...	y as cierra una de las mejores trilog as de...	0.0	0.0	neutral	0.000	1.000	0.000	0.0000
5	@FalandodeHerois O UCM nunca vai acabar e desi...	o ucm nunca vai acabar e desistir assim um...	0.0	0.0	neutral	0.000	1.000	0.000	0.0000
6	Las 31 películas y las 12 series del UCM 🤣 htt...	las 31 pel culas y las 12 series del ucm	0.0	0.0	neutral	0.000	1.000	0.000	0.0000
7	@FalandodeHerois Adeus é um exagero. Tenho cer...	adeus um exagero tenho certeza de que o u...	0.0	0.0	neutral	0.000	1.000	0.000	0.0000
8	@JesusTeAbomina @MarvelBRNews é interessante s...	interessante sim e sem d vida o melhor ...	0.0	0.0	neutral	0.000	1.000	0.000	0.0000
9	@FalandodeHerois O que de empolgaria então? Sé...	o que de empolgaria ent o s rio que ainda	0.0	0.0	positive	0.000	0.843	0.157	0.4215

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'])
```

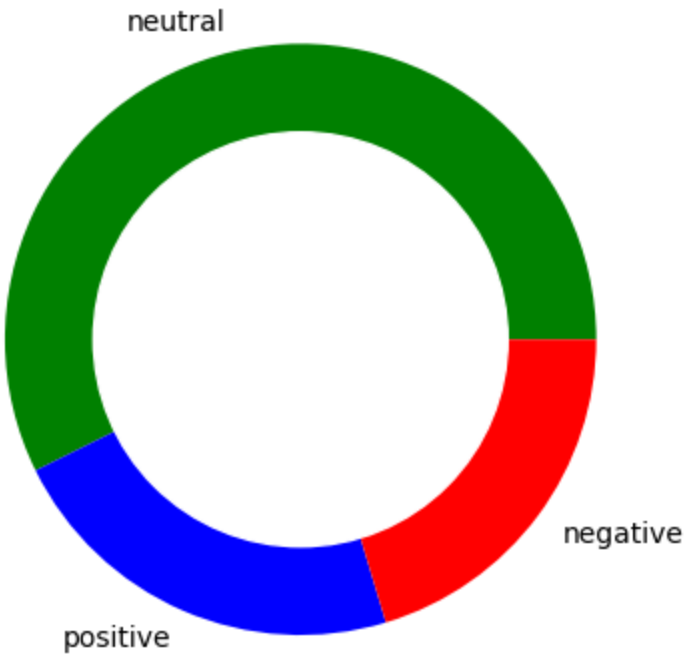
In [19]:
```
#Count_values for sentiment
count_values_in_column(tw_list,"sentiment")
```

Out[19]:

	Total	Percentage
neutral	200	57.31
positive	78	22.35
negative	71	20.34

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()
```



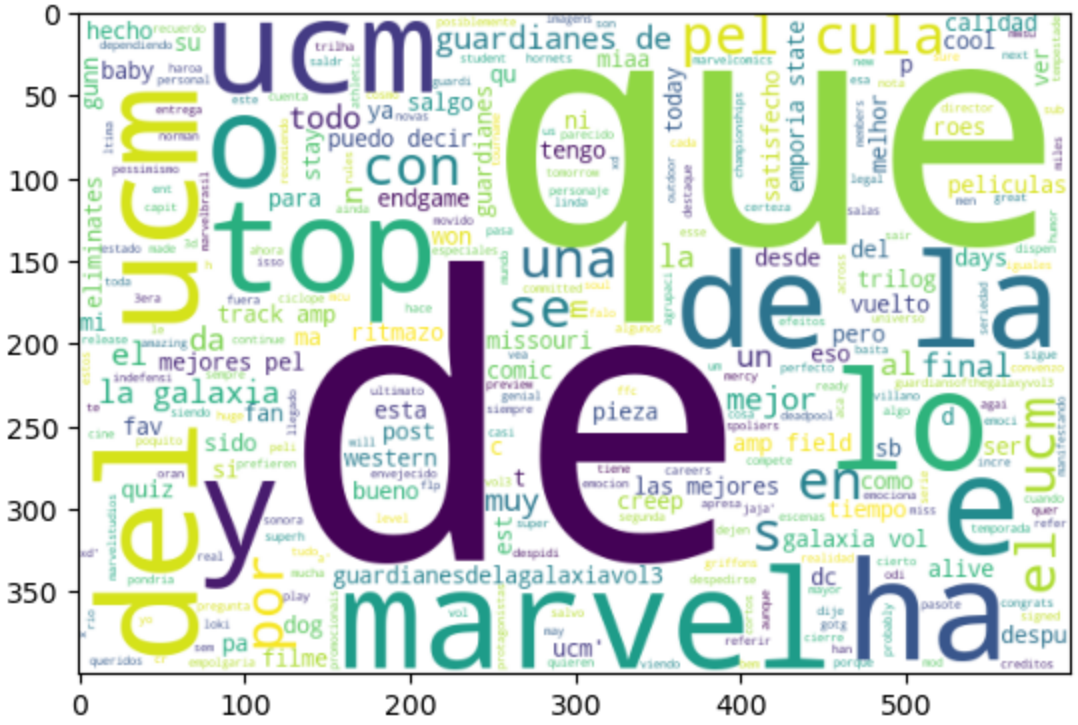
In [21]:
```
#Function to Create Wordcloud

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 [ ]: