Ex No.9: Implementation of Natural Language Processing programs

Natural Language Processing

NLP stands for Natural Language Processing, which is a part of Computer Science, Human language, and Artificial Intelligence. It is the technology that is used by machines to understand, analyse, manipulate, and interpret human's languages. It helps developers to organize knowledge for performing tasks such as translation, automatic summarization, Named Entity Recognition (NER), speech recognition, relationship extraction, and topic segmentation.

It is the ability to understand text and words in almost the same way human beings can. NLP strives to build machines that understand and respond to text or voice data with text or speech of their own. NLP combines computational linguistics with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data.

Natural Language Processing (NLP) is a component of AI in the field of linguistics that deals with **interpretation and manipulation of human speech or text using software**. It enables the computer to understand the natural way of human communication by combining machine learning, deep learning and statistical models.

NLP Applications

Here are a few applications of NLP:-

- 1. Speech Recognition
- 2. Voice Assistants and Chatbots
- 3. Auto Prediction or Grammar Correction
- 4. Filtering Emails
- 5. Text Summarization
- 6. Translation

NLP Algorithm

- 1. Enter the text (or sound converted to text)
- 2. Segmentation of text into components (segmentation and tokenization).
- 3. Text Cleaning (filtering from "garbage") removal of unnecessary elements.
- 4. Text Vectorization and Feature engineering.
- 5. Lemmatization and Steaming reducing inflections for words.
- 6. Using Machine Learning algorithms and methods for training models.
- 7. Interpretation of the result

Problem Statement:

To perform sentiment analysis using Natural Language Processing.

Sentiment analysis is the process of using natural language processing, text analysis, and statistics to analyze customer sentiments. The best business understand the sentiment of their customers- what people are saying, how they are saying it and what they mean.

customer sentiments can be found in tweets, comments, reviews or other places where people mention your brands. Sentiment analysis is the domain of understanding these emotions with software, and it it's a must understand topic for developers and business leaders in a modern workplace.

SENTIMENT ANALYSIS

sentiment analysis is the process of using natural language processing, text analysis, and statistics to analyse customer sentiments. The best business understand the sentiment of their customers- what people are saying, how they are saying it and what they mean.

customer sentiments can be found in tweets, comments, reviews or other places where people mention your brands. Sentiment analysis is the domain of understanding these emotions with software, and it it's a must understand topic for developers and business leaders in a modern workplace.

```
In [1]:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import re
import nltk
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from wordcloud import wordcloud
nltk.download('stopwords')
nltk.download('wordnet')
import warnings
warnings.filterwarnings('ignore')
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\NIKITHA\AppData\Roaming\nltk_data...
[nltk_data]
               Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]
                C:\Users\NIKITHA\AppData\Roaming\nltk_data...
[nltk_data]
                Package wordnet is already up-to-date!
In [2]:
from textblob import TextBlob
from tensorflow.keras.preprocessing.text import Tokenizer
from nltk.sentiment.util import *
In [3]:
trump = pd.read csv('hashtag donaldtrump.csv' , lineterminator='\n')
biden = pd.read csv('hashtag joebiden.csv', lineterminator='\n')
In [4]:
trump.head()
Out[4]:
   created_at
                tweet_id
                                tweet likes retweet_count
                                                                         user_id
                                                                                   user_name user_scree
                                                              source
                        #Elecciones2020
    2020-10-
                           l En #Florida:
                                                                                  El Sol Latino
         15 1.316529e+18
                                                           TweetDeck 3.60665e+08
                                       0.0
                                                    0.0
                                                                                                elsollati
                         #JoeBiden dice
                                                                                        News
     00:00:01
                        Usa 2020, Trump
    2020-10-
                                contro
         15
            1.316529e+18
                                      26.0
                                                                    3.316176e+08
                                                                                    Tgcom24 MediasetT
                                                    9.0
                            Facebook e
                                                            Mediaset
     00:00:01
                          Twitter: cop...
```

	reated 18t	tweet_id	#TrumptwAset student I used		retweet_count		source Twitter Web	user_id		me user_scre
2	15 00:00:02	1.316529e+18	to hear for years,	2.0	1.0)	Арр	8.436472e+06	sna	rke
3	2020-10- 15 00:00:02	1.316529e+18	2 hours since last tweet from #Trump! Maybe he	0.0	0.0) Tri	umpytweeter	8.283556e+17	Trumpytwee	eter trump
4	2020-10- 15 00:00:08	1.316529e+18	You get a tie! And you get a tie! #Trump 's ra	4.0	3.0)	Twitter for iPhone	4.741380e+07	Rana Abt بتر	ar- F رناأ
5 ro	ws × 21 c	olumns								
4)
In	[5]:									
oid	en.head	()								
Out	[5]:									
•	created_at	tweet_id	twe	et lik	es retweet_co	ount	source	user_id	user_name	user_screen_r
0	2020-10- 15 00:00:01	1.316529e+18	#Elecciones202 En #Florid #JoeBiden dice	la: (0.0	0.0	TweetDeck	3.60665e+08	El Sol Latino News	elsollatino
1	2020-10- 15 00:00:18	1.316529e+18	#HunterBide #HunterBidenEmai #JoeBiden #Joe	ils (0.0	0.0	Twitter for iPad	8.099044e+08	Cheri A . []	Biloximee
2	2020-10- 15 00:00:20	1.316529e+18	@IslandGirlPF @BradBeauregard @MeidasTouch T	dJ (0.0	0.0	Twitter Web App	3.494182e+09	Flag Waver	Flag_Wa
3	2020-10- 15 00:00:21	1.316529e+18	©chrislongvie Watching ar setting dvr. Let's	nd (0.0	0.0	Twitter for iPhone	8.242596e+17	Michelle Ferg	Michelle
4	2020-10- 15 00:00:22	1.316529e+18	#censorsh #HunterBide #Bide #BidenEmails #	en en	1.0	0.0	Twitter Web App	1.032807e+18	the Gold State	theegold
5 ro	ws × 21 c	olumns								
اا										
In	[6]:									
	nt(trump									
(97	0919, 21 6886, 21	1)								
	[7]:	- /								
	mp.info	()								
Ran	geIndex:	: 970919 en ns (total 2 n	rame.DataFram tries, 0 to 9 1 columns): Non-Nul	7091 1 Co	unt Dtyp					
0 1 2	create tweet_ tweet	ed_at	970919 970919 970919	non- non-	null obje null floa	ct it64				

```
970919 non-null float64
 4
    retweet count
 5
    source
                          970043 non-null object
    user_id
                          970919 non-null float64
    user_name
                          970903 non-null object
    user screen name
                          970919 non-null object
    user_description
                          869653 non-null object
                          970919 non-null object
10 user_join_date
 11 user_followers_count 970919 non-null float64
 12 user location
                          675966 non-null object
 13 lat
                          445719 non-null float64
 14 long
                          445719 non-null
                                           float64
 15 city
                          227187 non-null object
 16
    country
                          442748 non-null object
 17 continent
                          442765 non-null object
 18 state
                          320620 non-null object
                          300425 non-null object
 19 state code
                          970919 non-null object
20 collected at
dtypes: float6\overline{4}(7), object(14)
memory usage: 155.6+ MB
In [8]:
biden.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 776886 entries, 0 to 776885
Data columns (total 21 columns):
# Column
                          Non-Null Count
                                           Dtype
___
0 created at
                          776886 non-null object
    tweet id
                          776886 non-null
                                           float64
                          776886 non-null object
    tweet
                          776886 non-null
                                           float64
3
    likes
                          776886 non-null float64
    retweet_count
                          776173 non-null
 5
    source
                                           object
                          776886 non-null float64
 6
    user_id
 7
    user name
                          776868 non-null object
 8
    user_screen_name
                          776886 non-null
                                           object
    user description
                          694880 non-null object
                          776886 non-null object
    user_join_date
 11 user_followers_count 776886 non-null float64
 12 user_location
                          543095 non-null object
                          355293 non-null float64
 13 lat
                         355293 non-null float64
 14 long
 15
    city
                          186872 non-null object
 16
    country
                         353779 non-null object
 17
    continent
                          353797 non-null object
                          260195 non-null object
 18 state
 19 state code
                          244609 non-null object
 20 collected at
                          776886 non-null object
dtypes: float64(7), object(14)
memory usage: 124.5+ MB
In [9]:
# concat both dataframes for easier analysis
trump['Candidate'] = 'Trump'
In [10]:
trump
Out[10]:
      created at
                  tweet id
                                 tweet likes retweet count
                                                           source
                                                                     user id
                                                                              user name user
                         #Elecciones2020
       2020-10-
                            | En #Florida:
                                                                             El Sol Latino
           15 1.316529e+18
                                       0.0
                                                  0.0
                                                        TweetDeck 3.60665e+08
                          #JoeBiden dice
                                                                                  News
        00:00:01
```

	2020-10-		Usa 2020, Trump							
1	00:00:01	1.316529e+18	contro Facebook e Twitter: cop	ikes 26.0	retweet_count	s 8oc Medias	et 3.3	user_id 316176e+08	user_name Tgcom24	МŠ
	00:00:01		e i witter: cop							
2	2020-10- 15	1.316529e+18	#Trump: As a student I used to	2.0	1.0	Twitter W		36472e+06	snarke	
_	00:00:02	110100200110	hear for years,	2.0		A	ор О.Ч		onarno	
			2 hours since							
3	2020-10- 15	1.316529e+18	last tweet from	0.0	0.0	Trumpytweet	er 8.2	283556e+17	Trumpytweeter	
-	00:00:02		#Trump! Maybe he							
			You get a tie!							
4	2020-10- 15	1.316529e+18	And you get a	4.0	3.0	Twitter		41380e+07	Rana Abtar -	
•	00:00:08	110100200110	tie! #Trump 's ra		0.0	iPho	ne	110000101	رنا أبتر	
-		695			•••					
970914	2020-11- 08	1.325589e+18	@PelleX @Mikerger	0.0	0.0	Twitter W	eb 24	27951e+09	Alexander P.	
970914	23:59:43	1.3233690+16	@spiegelbilderz_	0.0	0.0	A	op 3.4	1279510+09	Vogt	
			@BertiFranz							
070015	2020-11-	1.325589e+18	JOE BIDEN WHY JOE BIDEN DID	4.0	0.0	Twitter W	eb 。	245258e+09	sh	
970915	08 23:59:48	1.3255890+18	WINNED ??	4.0	0.0	A	op 3.2	452566+09	ahzy	
			TRUMP TR							
070040	2020-11-	4.005500	#AfDller reagieren	0= 0		Twitter W	eb 🖺	000100 - 4=	and the de	
970916	08 23:59:53	1.325589e+18	panisch bis	25.0	4.0	A	op 9.8	866108e+17	watchdog	
			hysterisch auf							
	2020-11-		@sammelbis1998 @iheartmindy			Twitter	or			
970917	08 23:59:54	1.325589e+18	@bnorthg First,	0.0	0.0	iP		87811e+08	Debra	
			уо							
	2020-11-		OK just had to do it !\n#Trump			Twitter W	eb		Dunken K	
970918	08 23:59:55	1.325589e+18	#CatapultTrump	105.0	28.0		op 4.5	519791e+08	Bliths	
70919	rows × 22 (columns								
In [11										
]:									
oiden[te']= 'bide	en'							
	'Candidat	te']= 'bide	en'							
oiden[In [12	'Candidat	te']= 'bide	en'							
	'Candidat	te']= 'bide	en'							
In [12	'Candidat	te']= 'bide	en'							
In [12 piden	'Candidat		en'							
In [12 piden	'Candidat	te']= 'bide tweet_id	en'		tweet	likes retweet	count	sourc	e user_id	u
In [12 piden Dut[12	'Candidat]: created_at 2020-10-	tweet_id		nes202	(Marin 1970)				2000 2000 2000 2000	
In [12 piden	'Candidat]: created_at 2020-10- 15				tweet 0 En #Florida: beBiden dice	likes retweet			e user_id k 3.60665e+08	
In [12 piden Dut[12	'Candidat]: created_at 2020-10-	tweet_id			0 En #Florida:				2000 2000 2000 2000	
In [12 piden Dut[12	'Candidat]: created_at 2020-10- 15 00:00:01	tweet_id	#Eleccio	#Jo	0 En #Florida: oeBiden dice			TweetDec	k 3.606665e+08	
In [12 piden Dut[12	'Candidat]: created_at 2020-10- 15 00:00:01 2020-10- 15	tweet_id	#Eleccio	#Jo	0 En #Florida: beBiden dice rerBidenEmails			TweetDec	k 3.606665e+08	
In [12 piden Out[12	'Candidat]: created_at 2020-10- 15 00:00:01	tweet_id 1.316529e+18	#Eleccio	#Jo	0 En #Florida: oeBiden dice	0.0	0.0	TweetDec	k 3.606665e+08	
In [12 piden Out[12	'Candidat]: created_at 2020-10- 15 00:00:01 2020-10- 15	tweet_id 1.316529e+18	#Eleccio	#Jo	0 En #Florida: beBiden dice rerBidenEmails	0.0	0.0	TweetDec	k 3.606665e+08	
In [12 poiden Out[12 0	'Candidat]: created_at 2020-10- 15 00:00:01 2020-10- 15 00:00:18	tweet_id 1.316529e+18 1.316529e+18	#Eleccio #HunterBiden	#Jo #Hunt #Jo	0 En #Florida: beBiden dice rerBidenEmails	0.0	0.0	TweetDec	k 3.606665e+08	
In [12 piden Out[12	'Candidat]: created_at 2020-10- 15 00:00:01 2020-10- 15 00:00:18	tweet_id 1.316529e+18	#Eleccio #HunterBiden	#Jo #Hunt #Jo V @Bra	0 En #Florida: beBiden dice serBidenEmails beBiden #Joe	0.0	0.0	TweetDec	k 3.60665e+08	
In [12 poiden Out[12 0	'Candidat]: created_at 2020-10- 15 00:00:01 2020-10- 15 00:00:18 2020-10- 15 00:00:20	tweet_id 1.316529e+18 1.316529e+18	#Eleccio #HunterBiden	#Jo #Hunt #Jo V @Bra	0 En #Florida: beBiden dice cerBidenEmails beBiden #Joe	0.0	0.0	TweetDec Twitter for iPa	k 3.60665e+08	100
In [12 poiden Out[12 0	'Candidat]: created_at 2020-10- 15 00:00:01 2020-10- 15 00:00:18 2020-10- 2020-10- 2020-10- 2020-10-	tweet_id 1.316529e+18 1.316529e+18	#Eleccio #HunterBiden	#Jo #Hunt #Jo V @Bra @Me	0 En #Florida: beBiden dice erBidenEmails beBiden #Joe adBeauregardJ bidasTouch T	0.0	0.0	TweetDec Twitter for iPa Twittee Web Ap	k 3.606665e+08	FI

-	create	90 <u>:</u> 21	twe	et_id			dvr. Le	et's weet	likes re	tweet_count	iPhone source		ser_id	us
4	ŀ	0-10- 15 00:22	1.316529	le+18	#cens	orship	#HunterBiden #E #BidenEmai		1.0	0.0	Twitter Web App	1.03280)7e+18	3
776881	ı	0-11- 08 59:38	1.325589	e+18	Stop	alaying	g @CNN !\n#Paris #London don		0.0	0.0	Twitter for iPhone	3.87362	20e+08	
776882	2	0-11- 08 59:38	1.325589	e+18	Ωχ ελπίζ	ωναμ	μη μας βγει σαν οπαδούς τ		0.0	0.0	Twitter for Android	4.03281	9e+08	к
776883	3	0-11- 08 59:41	1.325589	e+18	L'OTAI	N va so	ortir de sa létharç redeve		48.0	14.0	Twitter for Android	7.81918	33e+17	(
776884	ı	0-11- 08 59:52	1.325589	e+18]\n\n"#cono	jiuntifu	ıoriregione"\n\n" feli	Sono ice	1.0	1.0	Twitter for iPhone	5.29331	5e+08	
776885	5	0-11- 08 59:58	1.325589	e+18	lk mod	et zegg	gen dat ik #Biden lesser of		0.0	0.0	Twitter for Android	5.86386	63e+08	
76886	o rows	× 22 (columns	í										
3	3]:													
in [13] lata_n lata_n lata_n lata_n	mixed mixed mixed	.sort	z_value		rump,bid ='create	d_at'								
in [13] lata_n lata_n lata_n lata_n	mixed mixed mixed	.sort	_value	es (by=	e'create	d_at'	retweet_count		sourc	e user	id use	r_name	user_s	6Cr(
in [1: lata_nlata_nlata_nlata_n crea	mixed mixed mixed 3]: ated_at	.sort	z_value	#Elecc	='create	d_at'		Т		e user k 3.60665e+	FLSo	r_name I Latino News	user_s els	465
in [1: lata_ilata_	mixed	. sort	t_value d() tweet_id	#Elecc I Er #Joel Usa 20	tweet siones2020 n #Florida: Biden dice	d_at'	retweet_count	T		k 3.60665e+	₀₈ El So	I Latino	els	olla
Creation (1.1) Creation (1.1)	mixed	1.316	tweet_id	#Elecc IEI #Joel Usa 20 F Twi #TI stuc	tweet siones2020 n #Florida: Biden dice 020, Trump contro	likes	retweet_count		weetDec	al 3.316176e+	08 El So	l Latino News	els	olla
(1	mixed	1.316 1.316	z_value d() tweet_id 529e+18	#Elecc IEI #Joel Usa 20 F Twi #TI stuc	tweet ciones2020 n #Florida: Biden dice 020, Trump contro acebook e ditter: cop rump: As a dent I used to hear for	likes 0.0 26.0	0.0 9.0	Tv	Socia Mediase vitter We	al 3.316176e+	08 Tg	I Latino News gcom24 snarke	els Media	solla
Create 1 1 20 20 20 20 20 20 3 3 0 20 4 4	mixed	1.316 1.316 1.316	z_value d() tweet_id 5529e+18	#Elecc IEI #Joel Usa 20 FTwi #Tri stuc 2 h last t #Trur	tweet ciones2020 n #Florida: Biden dice contro acebook e itter: cop rump: As a dent I used to hear for years, ours since tweet from mp! Maybe	likes 0.0 26.0 2.0	0.0 9.0	Tv	Socia Mediase vitter We	k 3.606665e+ al 3.316176e+ b 8.436472e+ er 8.283556e+	08 Tg 06 Trumpy	I Latino News gcom24 snarke	els Media	ise
Create Control of the	mixed	1.316 1.316 1.316	z_value d() tweet_id 5529e+18 5529e+18	#Elecc IEI #Joel Usa 20 FTwi #Tri stuc 2 h last t #Trur	tweet siones2020 n #Florida: Biden dice 020, Trump contro acebook e ditter: cop rump: As a dent I used to hear for years, ours since tweet from mp! Maybe he u get a tie! d you get a #Trump 's	likes 0.0 26.0 0.0	0.0 9.0 1.0	Tv	weetDec Socia Mediase vitter We Ap pytweete	k 3.606665e+ al 3.316176e+ b 8.436472e+ er 8.283556e+	08 Tg 06 Trumpy	I Latino News gcom24 snarke tweeter	els Media	olla

```
... . . . . .
# check the number of tweets received for each of them
data_mixed.groupby('Candidate')['tweet'].count().plot.bar()
plt.ylabel('number of tweets')
plt.title('Trump vs Biden')
plt.bar()
plt.show()
  1.0 <del>|</del>
                         Trump vs Biden
   0.8
 number of tweets
   0.2
   0.0
                             Candidate
In [15]:
# comparison of likes
data mixed.groupby('Candidate')['likes'].sum().plot.bar()
plt.ylabel('number of likes')
plt.title('trump vs biden')
plt.show()
                        trump vs biden
   6
 number of likes
   2
                           Candidate
In [16]:
data_mixed.country.value_counts()
Out[16]:
United States of America
                                     332495
United States
                                      61905
United Kingdom
                                      58051
                                      40091
India
                                      35379
Germany
Samoa
Saint Lucia
Northern Mariana Islands
East Timor
Lesotho
Name: country, Length: 189, dtype: int64
```

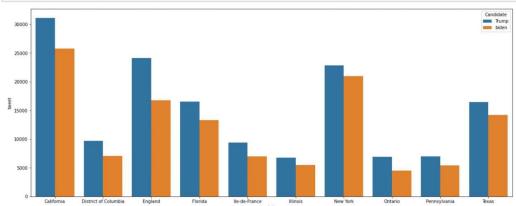
```
In [17]:
data mixed['country'] = data mixed['country'].replace({"United States of America":"US","Un
ited States":"US"})
In [18]:
data_mixed.country.value_counts()
Out[18]:
                              394400
United Kingdom
                               58051
India
                               40091
                               35379
Germany
                               35299
France
Cape Verde
Saint Lucia
East Timor
Lesotho
                                   1
Northern Mariana Islands
Name: country, Length: 188, dtype: int64
In [19]:
# countries with highest number of tweets
plt.figure(figsize=(10,5))
data_mixed.groupby('country')['tweet'].count().sort_values(ascending=False).head(10).plo
t.bar()
plt.ylabel('Number of tweets')
plt.title('top countries with highest number of tweets')
plt.show()
                         top countries with highest number of tweets
   400000
  350000
   300000
  250000
 of
  200000
 ₹ 150000
  100000
   50000
          3
                 United Kingdom
                                        country
In [20]:
# tweets for biden and trump from each country
top10countries = data_mixed.groupby('country')['tweet'].count().sort_values(ascending=Fa
lse) .head(10) .index.tolist()
tweet_df = data_mixed.groupby(['country', 'Candidate'])['tweet'].count().reset_index()
tweet df = tweet df[tweet df['country'].isin(top10countries)]
tweet df
Out [20]:
          country Candidate tweet
```

	, 	Candidate	t-0205
	Australia	biden	5534
54	Canada	Trump	16996
55	Canada	biden	10809
12	France	Trump	20689
13	France	biden	14610
18	Germany	Trump	22618
19	Germany	biden	12761
46	India	Trump	18559
47	India	biden	21532
60	Italy	Trump	11206
61	Italy	biden	8870
:07	Mexico	Trump	5811
08	Mexico	biden	5092
36	Turkey	Trump	5594
37	Turkey	biden	4774
40	US	Trump	213263
141	US	biden	181137
48 U	nited Kingdom	Trump	34743
49 U	nited Kingdom	biden	23308
	arplot(data now()		
lt.sl	now()		
200000 -			
175000 - 150000 -			
175000 - 150000 - 125000 - 100000 -			
175000 - 150000 - 125000 - 100000 - 75000 -			
175000 - 150000 - 125000 - 100000 - 75000 -	Australia	Canada	France
175000 - 150000 - 125000 - 100000 - 75000 - 50000 -	Australia	Canada	France
175000 - 150000 - 125000 - 100000 - 75000 - 25000 -		Canada	France
175000 - 150000 - 125000 - 100000 - 75000 - 25000 -			
175000 - 150000 - 125000 - 125000 - 75000 - 75000 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0	2]: states = da	ata_mixectolist()	d.group
175000 - 150000 - 150000 - 125000 - 1000000 - 75000 - 25000 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0	2]: states = da 10).index. df = data df = state	ata_mixec tolist() _mixed.gr	d.group
175000 - 150000 - 125000 - 125000 - 75000 - 75000 - 10000 - 100	2]: states = da 10).index. df = data df = state	ata_mixec tolist() _mixed.gr	d.group
175000 - 150000 - 150000 - 125000 - 1000000 - 75000 - 25000 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0 - 0	estates = da 10).index. df = data df = state	ata_mixec tolist() _mixed.gr	d.group
175000 - 150000 - 125000 - 125000 - 125000 - 125000 - 1000000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 1000000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 1000000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 1000000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 1000000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 1000000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 1000000 - 100000 - 100000 - 100000 - 100000 - 100000 - 100000 - 10000	estates = da 10).index.i df = data df = state df	ata_mixec tolist() _mixed.gr	d.group coupby e_df['

207	Calif state	Candidata	260:EST
349	District of Columbia	Trump	9684
350	District of Columbia	biden	7055
386	England	Trump	24111
387	England	biden	16743
410	Florida	Trump	16568
411	Florida	biden	13284
518	Ile-de-France	Trump	9393
519	Ile-de-France	biden	7001
5 20	Illinois	Trump	6777
521	Illinois	biden	5500
794	New York	Trump	22881
795	New York	biden	21005
869	Ontario	Trump	6907
870	Ontario	biden	4492
899	Pennsylvania	Trump	7011
900	Pennsylvania	biden	5403
1154	Texas	Trump	16491
1155	Texas	biden	14191

In [23]:

```
plt.figure(figsize=(20,8))
sns.barplot(data=state_df,x='state',y='tweet',hue='Candidate')
plt.show()
```



Sentiment analysis of presidential elections

```
In [24]:

lm = WordNetLemmatizer()

In [25]:

def clean(text):
    text = str(text).lower()
    text = re.sub('[^a-z]',' ', str(text))
    text = text.split()
```

```
text = [lm.lemmatize(j) for j in text if j not in set(stopwords.words('english'))]
text = ' '.join(str(k) for k in text)
return text
```

create function to get polarity and subjectivity #subjectivity---- if the sentence is subjective or objective---- subjective sentence refer to personal opinions, emotions or people's judgment whereas objective sentence refers to factual information.lt ranges from [0,1]. ## Polarity-- Float value that lies between [-1,1] where 1 means positive sentiment and -1 means negative sentiment.

```
In [26]:
```

```
def getsubjectivity(text):
    return TextBlob(text).sentiment.subjectivity
def getpolarity(text):
    return TextBlob(text).sentiment.polarity
def getAnalysis(score):
    if score <0:
        return 'negative'
    elif score == 0:
        return 'neutral'
    else:
        return 'positive'</pre>
```

DonaldTrump tweets sentiment analysis

```
In [27]:
```

```
trump_tweets = data_mixed.query('(Candidate == "Trump")').sort_values('user_followers_cou
nt',ascending=False).drop_duplicates(['user_name'])[['tweet','country']]
trump_tweets = trump_tweets.dropna().loc[trump_tweets.country =='US']
trump_tweets.reset_index(inplace = True ,drop =True)
```

```
In [28]:
```

```
trump_tweets
```

Out[28]:

tweet country

US	Without a doubt, the #FourSeasons debacle is t	0
US	President Donald #Trump's youngest daughter, T	1
US	Woah. Have you read this article?\n\nAll the p	2
US	The latest episode of #SNL tackled #DonaldTrum	3
US	${\small \square}{\small \square}$ #Decision2020: @JoeBiden firmará una serie	4
US	@NBCNews You wanted it over by election nights	45518
US	Even if #Trump DID win this election, he would	45519
US	Twisted, pressed & Dried almost empty too	45520
US	${\tt \#Election 2020 \ lection Results 2020 \ $	45521
US	Watching this poor guy on @FoxNews explain rea	45522

45523 rows × 2 columns

```
In [29]:
```

```
trump_tweets['cleantext'] = trump_tweets['tweet'].apply(clean)
In [30]:
trump_tweets['subjectivity'] = trump_tweets['cleantext'].apply(getsubjectivity)
```

```
Tn [31]:
```

```
--- .--.
trump tweets['polarity'] = trump tweets['cleantext'].apply(getpolarity)
In [32]:
trump_tweets['analysis'] = trump_tweets['polarity'].apply(getAnalysis)
In [33]:
trump_tweets.head()
Out[33]:
                                tweet country
                                                                       cleantext subjectivity
                                                                                          polarity analysis
                                          US without doubt fourseasons debacle best
        Without a doubt, the #FourSeasons
                                                                                  0.483333 0.333333 positive
                          debacle is t...
       President Donald #Trump's youngest
                                                   president donald trump youngest
                                          US
                                                                                  0.000000 0.000000
1
                                                                                                    neutral
                          daughter, T...
                                                                  daughter tiffa...
    Woah. Have you read this article?\n\nAll
                                              woah read article president debt donald
                                          US
                                                                                  0.000000 0.000000
                                                                                                   neutral
         The latest episode of #SNL tackled
                                               latest episode snl tackled donaldtrump
                                                                                  0.800000 _____ negative
                        #DonaldTrum...
                                                                       loss co...
   III #Decision2020: @JoeBiden firmará una
                                                decision joebiden firmar una serie de
                                          US
                                                                                  0.000000 0.000000 neutral
                               serie ...
                                                                      rdenes e...
In [34]:
trump_tweets.analysis.value_counts(normalize=True)*100
Out[34]:
neutral
              45.627485
positive
             34.389210
              19.983305
negative
Name: analysis, dtype: float64
In [35]:
# Plot graph for analysis of Trump Tweets
plt.figure(figsize=(10,5))
(trump_tweets.analysis.value_counts(normalize=True)*100).plot.bar()
plt.ylabel('%age of tweets')
plt.show()
   40
%age of tweets
02 0
   10
In [37]:
from wordcloud import WordCloud, STOPWORDS
```

```
eed
PLEASE VOIE
        Q
                                GOP
        0
                   media
        O
                 nothing
                  don t
                       said
                   American
                                                   OTUS
                      de
     JoeBiden CNN Trump
  never Americanti Trump coronavirus
                                         really
                                                know
```

Joe Biden Sentiment Analysis

```
In [38]:
```

```
biden_tweets = data_mixed.query('(Candidate=="biden")').sort_values('user_followers_coun
t',ascending=False).drop_duplicates(['user_name'])[['tweet','country']]
biden_tweets = biden_tweets.dropna().loc[biden_tweets.country =='US']
biden_tweets.reset_index(inplace = True ,drop =True)
```

In [39]:

biden tweets

Out[39]:

country	tweet	
US	#JoeBiden is enlisting A-list star power to he	0
US	While we were campaigning across battleground	1
US	The big fight TODAY!!!!! @realDonaldTrump vs #	2
US	News outlets have projected that #JoeBiden has	3
US	During his victory speech, #JoeBiden said he u	4
US	46th President of The Unites States! Congratul	57545
US	Biden IS going to win tonight. #biden #Electio	57546
US	All 238k deaths recorded in the United States	57547
us	Joe Biden'in başa gelmesi bizim bence işimize	57548
	O 15 117 W 11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	

```
5/549 @realDonaldTrump It's like monopoly.. #JoeBide... US tweet country
57550 rows × 2 columns
In [40]:
biden tweets['cleantext'] = biden tweets['tweet'].apply(clean)
In [41]:
biden_tweets['subjectivity'] = biden_tweets['cleantext'].apply(getsubjectivity)
biden_tweets['polarity'] = biden_tweets['cleantext'].apply(getpolarity)
In [43]:
biden_tweets['analysis'] = biden_tweets['polarity'].apply(getAnalysis)
In [44]:
biden_tweets.head()
Out[44]:
                                 tweet country
                                                                        cleantext subjectivity polarity analysis
    #JoeBiden is enlisting A-list star power to
                                                 joebiden enlisting list star power help
                                           US
                                                                                        0.00
                                                                                                     neutral
                                                                          close.
         While we were campaigning across
                                                campaigning across battleground state
                                           US
                                                                                       0.75
                                                                                               0.80
                                                                                                     positive
                         battleground ...
                                                                        ran grea...
                   The big fight TODAY!!!!!
                                                     big fight today realdonaldtrump v
                                           US
 2
                                                                                       0.10
                                                                                               0.00
                                                                                                     neutral
                 @realDonaldTrump vs #...
                                                                    joebiden got...
           News outlets have projected that
                                                news outlet projected joebiden secured
 3
                                                                                       0.00
                                                                                               0.00
                                                                                                     neutral
                        #JoeBiden has...
                                                                         elector...
    During his victory speech, #JoeBiden said
                                               victory speech joebiden said understood
                                                                                       0.75
                                                                                               -0.75 negative
In [45]:
biden_tweets.analysis.value_counts(normalize=True)*100
Out[45]:
neutral
             46.997394
positive
             38.748914
             14.253692
negative
Name: analysis, dtype: float64
In [46]:
plt.figure(figsize=(10,5))
(biden tweets.analysis.value counts(normalize=True) *100).plot.bar()
plt.ylabel('%age of tweets')
plt.show()
   40
%age of tweets
```

```
negative - negative -
```

In [48]:

```
still president elect watchdemocrat michige called already made
usatelljoebiden kamalaharris
plan said take need vote biden
country biden harris biden votinglong
                                                                            good
                                                                            back
     realdonaldtrump <sup>1</sup>
                                                      year
                                                         tonight<sup>senat</sup>
biden bidenharris condemocracy
          votershowamericaseep help president
                                                                              hink president united
 rally win keep Want - twitter Containing the old win keep Want - twitter Containing the old will be clection result georgia live Winter Containing the old will be clection result georgia live Winter Containing the old will be clection result georgia.
                                                       vice president
                                                                                           biden win
                                                                                   eguy
                                                                                                              potus
  <u>vote</u>
                                                                                                                 poll
                                                2
bidenharris
                                                                                                                ust de
                                              pa let
                                              may O
                                                                                               trumpman cnn
          thank come
```

According to sentiment analysis Biden has an upper hand over Trump in US which actually matters in presedential elections.

RESULT

Sentiment analysis was applied on tweets using Natural Language Processing.

In []: