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	source	user_id	user_name	user_scree
0	2020-10-15 00:00:01	1.316529e+18	#Elecciones2020 | En #Florida: #JoeBiden dice ...	0.0	0.0	TweetDeck	3.606665e+08	El Sol Latino News	elsollati
1	2020-10-15 00:00:01	1.316529e+18	Usa 2020, Trump contro Facebook e Twitter: cop...	26.0	9.0	Social Mediaset	3.316176e+08	Tgcom24	MediasetT

	created_at	tweet_id	#Trump: As a student I used to hear for years,...	likes	retweet_count	source	user_id	user_name	user_scree
2	2020-10-15 00:00:02	1.316529e+18		2.0	1.0	Twitter Web App	8.436472e+06	snarke	
3	2020-10-15 00:00:02	1.316529e+18	2 hours since last tweet from #Trump! Maybe he...	0.0	0.0	Trumpytweeter	8.283556e+17	Trumpytweeter	trumpy
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 Abtar - رنا أبتر	Ra

5 rows × 21 columns

In [5]:

```
biden.head()
```

Out[5]:

	created_at	tweet_id	tweet	likes	retweet_count	source	user_id	user_name	user_screen_na
0	2020-10-15 00:00:01	1.316529e+18	#Elecciones2020 | En #Florida: #JoeBiden dice ...	0.0	0.0	TweetDeck	3.606665e+08	El Sol Latino News	elsollatinone
1	2020-10-15 00:00:18	1.316529e+18	#HunterBiden #HunterBidenEmails #JoeBiden #Joe...	0.0	0.0	Twitter for iPad	8.099044e+08	Cheri A. □ □□	Biloximeem
2	2020-10-15 00:00:20	1.316529e+18	@IslandGirlPRV @BradBeauregardJ @MeidasTouch T...	0.0	0.0	Twitter Web App	3.494182e+09	Flag Waver	Flag_Wav
3	2020-10-15 00:00:21	1.316529e+18	@chrislongview Watching and setting dvr. Let's...	0.0	0.0	Twitter for iPhone	8.242596e+17	Michelle Ferg	MichelleFe
4	2020-10-15 00:00:22	1.316529e+18	#censorship #HunterBiden #Biden #BidenEmails #...	1.0	0.0	Twitter Web App	1.032807e+18	the Gold State	theegoldst

5 rows × 21 columns

In [6]:

```
print(trump.shape)
print(biden.shape)
```

```
(970919, 21)
(776886, 21)
```

In [7]:

```
trump.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 970919 entries, 0 to 970918
Data columns (total 21 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   created_at            970919 non-null  object
 1   tweet_id              970919 non-null  float64
 2   tweet                 970919 non-null  object
 3   likes                 970919 non-null  float64
```

```
 4   retweet_count          970919 non-null  float64
 5   source                 970043 non-null  object
 6   user_id                970919 non-null  float64
 7   user_name              970903 non-null  object
 8   user_screen_name       970919 non-null  object
 9   user_description       869653 non-null  object
 10  user_join_date         970919 non-null  object
 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
 19  state_code             300425 non-null  object
 20  collected_at           970919 non-null  object
dtypes: float64(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
 1   tweet_id               776886 non-null  float64
 2   tweet                  776886 non-null  object
 3   likes                  776886 non-null  float64
 4   retweet_count          776886 non-null  float64
 5   source                 776173 non-null  object
 6   user_id                776886 non-null  float64
 7   user_name              776868 non-null  object
 8   user_screen_name       776886 non-null  object
 9   user_description       694880 non-null  object
 10  user_join_date         776886 non-null  object
 11  user_followers_count   776886 non-null  float64
 12  user_location          543095 non-null  object
 13  lat                    355293 non-null  float64
 14  long                   355293 non-null  float64
 15  city                   186872 non-null  object
 16  country                353779 non-null  object
 17  continent              353797 non-null  object
 18  state                  260195 non-null  object
 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
0	2020-10-15 00:00:01	1.316529e+18	#Elecciones2020 | En #Florida: #JoeBiden dice ...	0.0	0.0	TweetDeck	3.606665e+08	El Sol Latino News	e

	created_at	tweet_id	tweet	likes	retweet_count	source	user_id	user_name	user
1	2020-10-15 00:00:01	1.316529e+18	Usa 2020, Trump contro Facebook e Twitter: cop...	26.0	9.0	Social Mediaset	3.316176e+08	Tgcom24	Mec
2	2020-10-15 00:00:02	1.316529e+18	#Trump: As a student I used to hear for years,...	2.0	1.0	Twitter Web App	8.436472e+06	snarke	
3	2020-10-15 00:00:02	1.316529e+18	2 hours since last tweet from #Trump! Maybe he...	0.0	0.0	Trumpytweeter	8.283556e+17	Trumpytweeter	
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 Abtar - رنا أبتر	
...	...	...	...	...	...	...	...	...	
970914	2020-11-08 23:59:43	1.325589e+18	@PelleX @Mikerger @spiegelbilderz_ @BertiFranz...	0.0	0.0	Twitter Web App	3.427951e+09	Alexander P. Vogt	
970915	2020-11-08 23:59:48	1.325589e+18	JOE BIDEN WHY JOE BIDEN DID WINNED ?? TRUMP TR...	4.0	0.0	Twitter Web App	3.245258e+09	ahzy	
970916	2020-11-08 23:59:53	1.325589e+18	#AfDller reagieren panisch bis hysterisch auf ...	25.0	4.0	Twitter Web App	9.866108e+17	watchdog	
970917	2020-11-08 23:59:54	1.325589e+18	@sammelbis1998 @iheartmindy @bnorthg First, yo...	0.0	0.0	Twitter for iPad	7.687811e+08	Debra	
970918	2020-11-08 23:59:55	1.325589e+18	OK just had to do it !\n#Trump #CatapultTrump ...	105.0	28.0	Twitter Web App	4.519791e+08	Dunken K Bliths	

970919 rows × 22 columns

In [11]:

```
biden['Candidate']= 'biden'
```

In [12]:

```
biden
```

Out[12]:

	created_at	tweet_id	tweet	likes	retweet_count	source	user_id	use
0	2020-10-15 00:00:01	1.316529e+18	#Elecciones2020 | En #Florida: #JoeBiden dice ...	0.0	0.0	TweetDeck	3.606665e+08	
1	2020-10-15 00:00:18	1.316529e+18	#HunterBiden #HunterBidenEmails #JoeBiden #Joe...	0.0	0.0	Twitter for iPad	8.099044e+08	Cl
2	2020-10-15 00:00:20	1.316529e+18	@IslandGirlPRV @BradBeauregardJ @MeidasTouch T...	0.0	0.0	Twitter Web App	3.494182e+09	Flag
3	2020-10-15	1.316529e+18	@chrislongview Watching and setting	0.0	0.0	Twitter for	8.242596e+17	M

	created_at	tweet_id	tweet	likes	retweet_count	source	user_id	use
			dvr. Let's...			iPhone		
4	2020-10-15 00:00:22	1.316529e+18	#censorship #HunterBiden #Biden #BidenEmails #...	1.0	0.0	Twitter Web App	1.032807e+18	t
...	...	...	...	...	...	...	...	
776881	2020-11-08 23:59:38	1.325589e+18	Stop laying @CNN !\n#Paris and #London dont gi...	0.0	0.0	Twitter for iPhone	3.873620e+08	
776882	2020-11-08 23:59:38	1.325589e+18	Ωχ ελπίζω να μη μας βγει σαν τους οπαδούς του...	0.0	0.0	Twitter for Android	4.032819e+08	ο κά
776883	2020-11-08 23:59:41	1.325589e+18	L'OTAN va sortir de sa léthargie et redevenir ...	48.0	14.0	Twitter for Android	7.819183e+17	(m 6
776884	2020-11-08 23:59:52	1.325589e+18	\n\n“#congiuntifuoriregione”\n\n‘Sono felice ...	1.0	1.0	Twitter for iPhone	5.293315e+08	
776885	2020-11-08 23:59:58	1.325589e+18	Ik moet zeggen dat ik #Biden "the lesser of tw...	0.0	0.0	Twitter for Android	5.863863e+08	

776886 rows × 22 columns

In [13]:

```
data_mixed = pd.concat([trump,biden])
data_mixed.sort_values(by='created_at')
data_mixed.head()
```

Out[13]:

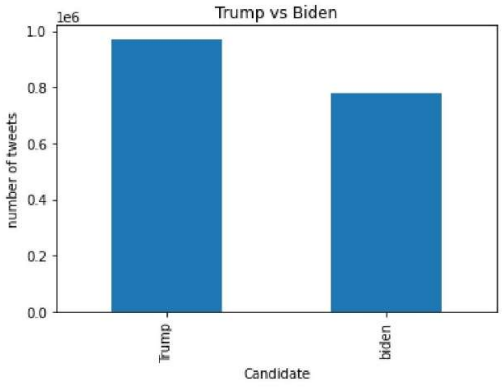
	created_at	tweet_id	tweet	likes	retweet_count	source	user_id	user_name	user_scree
0	2020-10-15 00:00:01	1.316529e+18	#Elecciones2020 | En #Florida: #JoeBiden dice ...	0.0	0.0	TweetDeck	3.606665e+08	El Sol Latino News	elsollati
1	2020-10-15 00:00:01	1.316529e+18	Usa 2020, Trump contro Facebook e Twitter: cop...	26.0	9.0	Social Mediaset	3.316176e+08	Tgcom24	MediasetT
2	2020-10-15 00:00:02	1.316529e+18	#Trump: As a student I used to hear for years,...	2.0	1.0	Twitter Web App	8.436472e+06	snarke	
3	2020-10-15 00:00:02	1.316529e+18	2 hours since last tweet from #Trump! Maybe he...	0.0	0.0	Trumpytweeter	8.283556e+17	Trumpytweeter	trumpy
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 Abtar - رنا أبتر	Ra

5 rows × 22 columns

Exploratory Data Analysis

In [14]:

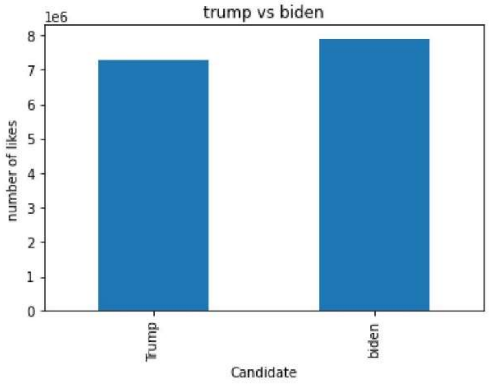
```
# check the number of tweets received for each of them
data_mixed.groupby('Candidate')['tweet'].count().plot.bar()
plt.ylabel('number of tweets')
plt.title('Trump vs Biden')
plt.show()
```



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()
```



In [16]:

```
data_mixed.country.value_counts()
```

Out[16]:

```
United States of America    332495
United States                61905
United Kingdom               58051
India                        40091
Germany                      35379
                             ...
Samoa                            2
Saint Lucia                      1
Northern Mariana Islands         1
East Timor                       1
Lesotho                          1
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]:

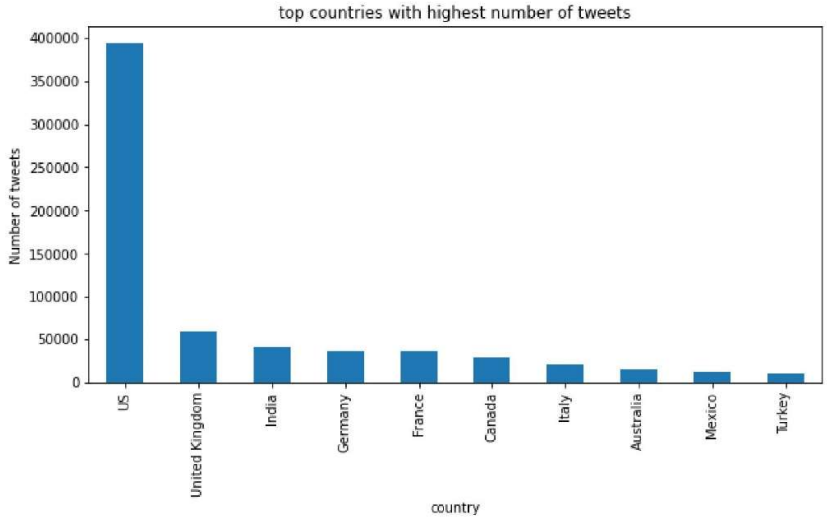
```
data_mixed.country.value_counts()
```

Out[18]:

```
US                            394400
United Kingdom                 58051
India                          40091
Germany                        35379
France                         35299
                               ...
Cape Verde                         2
Saint Lucia                        1
East Timor                         1
Lesotho                            1
Northern Mariana Islands           1
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()
```



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
```

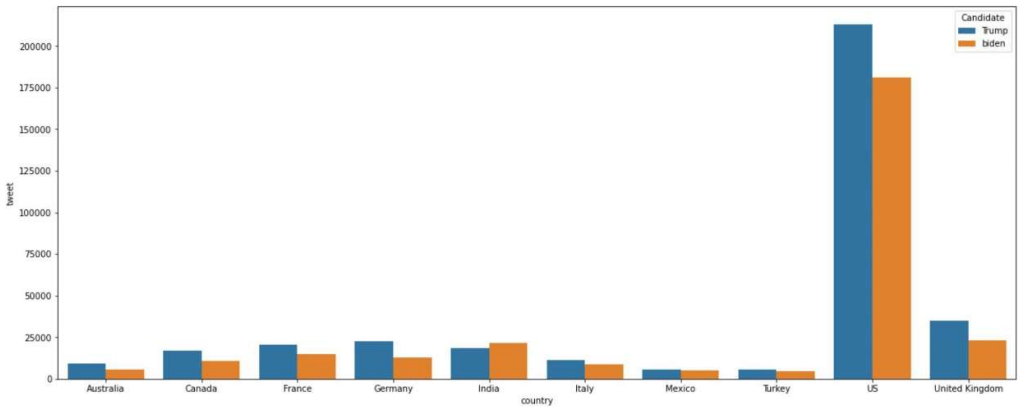
Out[20]:

	country	Candidate	tweet

16	Australia	biden	5534
17	Australia	biden	5534
54	Canada	Trump	16996
55	Canada	biden	10809
112	France	Trump	20689
113	France	biden	14610
118	Germany	Trump	22618
119	Germany	biden	12761
146	India	Trump	18559
147	India	biden	21532
160	Italy	Trump	11206
161	Italy	biden	8870
207	Mexico	Trump	5811
208	Mexico	biden	5092
336	Turkey	Trump	5594
337	Turkey	biden	4774
340	US	Trump	213263
341	US	biden	181137
348	United Kingdom	Trump	34743
349	United Kingdom	biden	23308

In [21]:

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



In [22]:

```
top10states = data_mixed.groupby('state')['tweet'].count().sort_values(ascending=False).
head(10).index.tolist()
state_df = data_mixed.groupby(['state','Candidate'])['tweet'].count().reset_index()
state_df = state_df[state_df['state'].isin(top10states)]
state_df
```

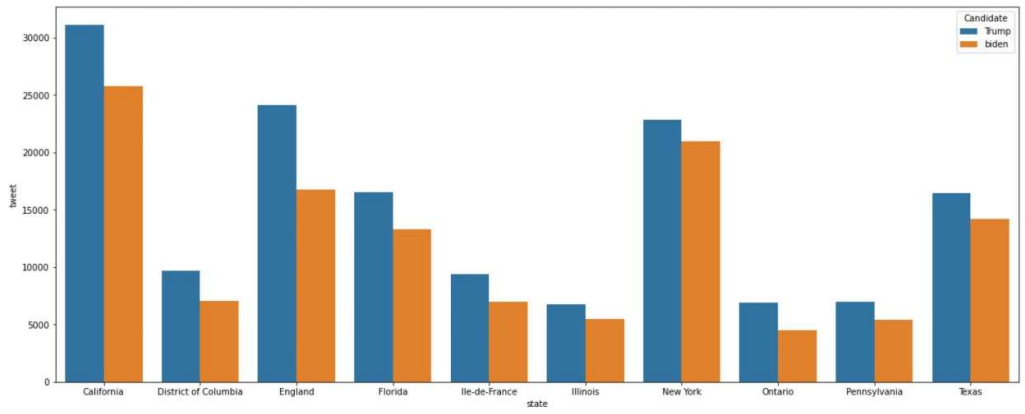
Out[22]:

	state	Candidate	tweet
206	California	Trump	31149

207	**California**	**Candidate**	**26013**
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**
520	**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.It 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'
```

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
0	Without a doubt, the #FourSeasons debacle is t...	US
1	President Donald #Trump's youngest daughter, T...	US
2	Woah. Have you read this article?\n\nAll the p...	US
3	The latest episode of #SNL tackled #DonaldTrum...	US
4	□□ #Decision2020: @JoeBiden firmará una serie ...	US
...	...	...
45518	@NBCNews You wanted it over by election nights...	US
45519	Even if #Trump DID win this election, he would...	US
45520	Twisted, pressed & curled almost empty too...	US
45521	#Election2020\n#USElectionResults2020\n#Donald...	US
45522	Watching this poor guy on @FoxNews explain rea...	US

45523 rows × 2 columns

In [29]:

```
trump_tweets['cleantext'] = trump_tweets['tweet'].apply(clean)
```

In [30]:

```
trump_tweets['subjectivity'] = trump_tweets['cleantext'].apply(getsubjectivity)
```

In [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
0	Without a doubt, the #FourSeasons debacle is t...	US	without doubt fourseasons debacle best part do...	0.483333	0.333333	positive
1	President Donald #Trump's youngest daughter, T...	US	president donald trump youngest daughter tiffa...	0.000000	0.000000	neutral
2	Woah. Have you read this article?\n\nAll the p...	US	woah read article president debt donald trump ...	0.000000	0.000000	neutral
3	The latest episode of #SNL tackled #DonaldTrum...	US	latest episode snl tackled donaldtrump loss co...	0.800000	-0.033333	negative
4	□□ #Decision2020: @JoeBiden firmará una serie ...	US	decision joebiden firmar una serie de rdenes e...	0.000000	0.000000	neutral

In [34]:

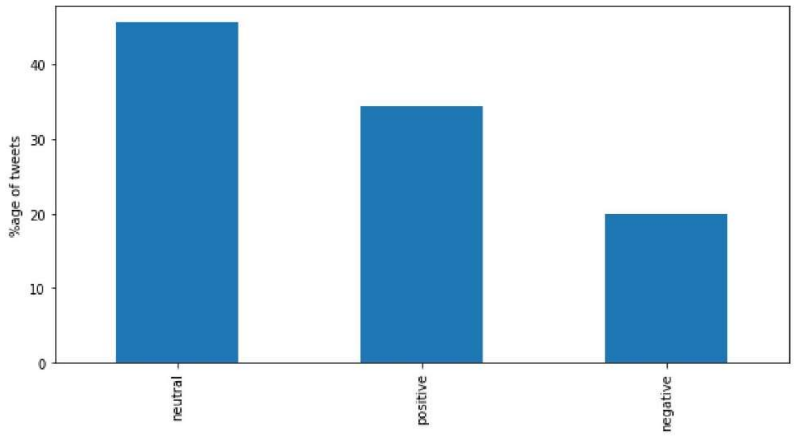
```
trump_tweets.analysis.value_counts(normalize=True)*100
```

Out[34]:

```
neutral     45.627485
positive    34.389210
negative    19.983305
Name: analysis, dtype: float64
```

In [35]:

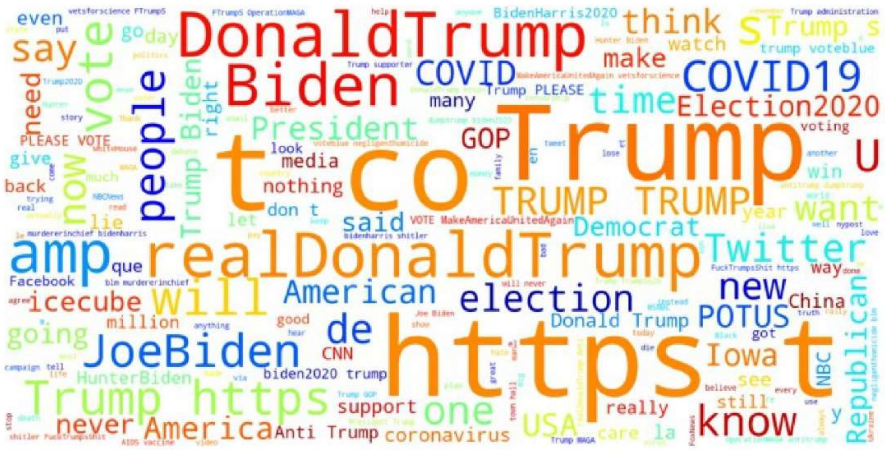
```
# Plot graph for analysis of Trump Tweets
plt.figure(figsize=(10,5))
(trump_tweets.analysis.value_counts(normalize=True)*100).plot.bar()
plt.ylabel('%age of tweets')
plt.show()
```



In [37]:

```
from wordcloud import WordCloud, STOPWORDS
```

```
def word_cloud(wd_list):
    stopwords = set(STOPWORDS)
    all_words = ' '.join(text for text in wd_list)
    wordcloud = WordCloud(background_color='white',
                          stopwords= stopwords,
                          width=1600, height=800,
                          random_state=1,
                          colormap='jet',
                          max_words=200,
                          max_font_size=200).generate(all_words)
    plt.figure(figsize=(12,10))
    plt.axis('off')
    plt.imshow(wordcloud)
word_cloud(trump_tweets['cleantext'][:5000])
```



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]:

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

	tweet	country
57549	@realDonaldTrump It's like monopoly.. #JoeBide...	US

57550 rows × 2 columns

In [40]:

```
biden_tweets['cleantext'] = biden_tweets['tweet'].apply(clean)
```

In [41]:

```
biden_tweets['subjectivity'] = biden_tweets['cleantext'].apply(getsubjectivity)
```

In [42]:

```
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
0	#JoeBiden is enlisting A-list star power to he...	US	joebiden enlisting list star power help close ...	0.00	0.00	neutral
1	While we were campaigning across battleground ...	US	campaigning across battleground state ran grea...	0.75	0.80	positive
2	The big fight TODAY!!!!! @realDonaldTrump vs #...	US	big fight today realdonaldtrump v joebiden got...	0.10	0.00	neutral
3	News outlets have projected that #JoeBiden has...	US	news outlet projected joebiden secured elector...	0.00	0.00	neutral
4	During his victory speech, #JoeBiden said he u...	US	victory speech joebiden said understood trump ...	0.75	-0.75	negative

In [45]:

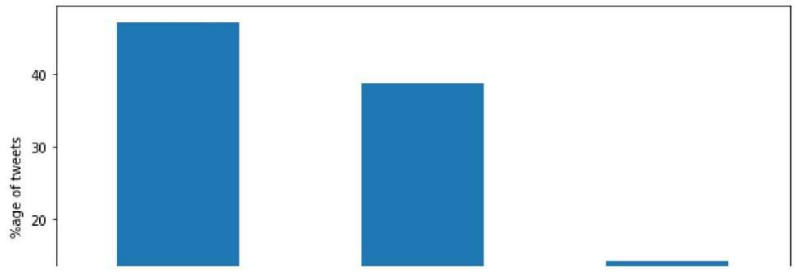
```
biden_tweets.analysis.value_counts(normalize=True)*100
```

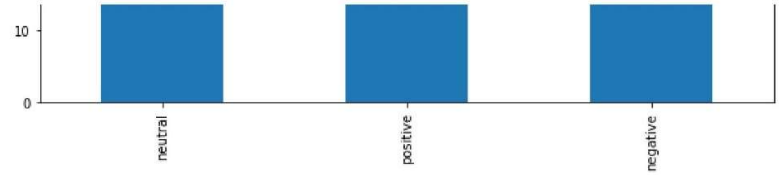
Out[45]:

```
neutral     46.997394
positive    38.748914
negative    14.253692
Name: analysis, dtype: float64
```

In [46]:

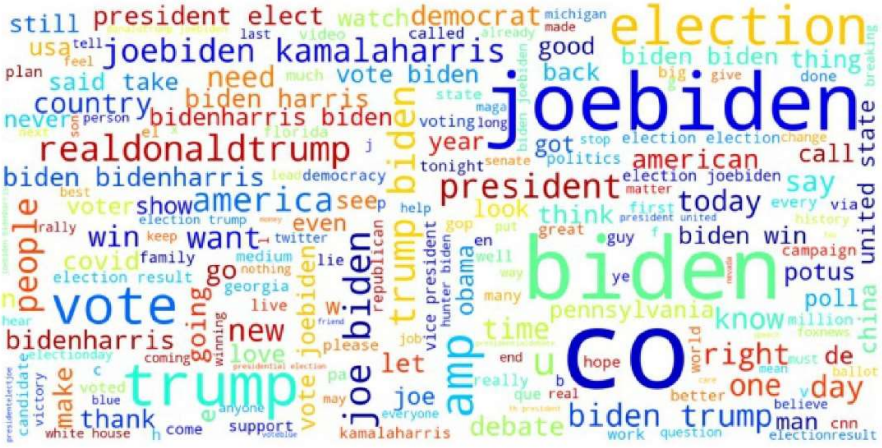
```
plt.figure(figsize=(10,5))
(biden_tweets.analysis.value_counts(normalize=True)*100).plot.bar()
plt.ylabel('%age of tweets')
plt.show()
```





In [48]:

```
from wordcloud import WordCloud, STOPWORDS
def word_cloud(wd_list):
    stopwords = set(STOPWORDS)
    all_words = ' '.join(text for text in wd_list)
    wordcloud = WordCloud(background_color='white',
                          stopwords= stopwords,
                          width=1600, height=800,
                          random_state=1,
                          colormap='jet',
                          max_words=200,
                          max_font_size=200).generate(all_words)
    plt.figure(figsize=(12,10))
    plt.axis('off')
    plt.imshow(wordcloud)
word_cloud(biden_tweets['cleantext'][:5000])
```



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 [ ]: