Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications.

Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis. Seaborn works easily with dataframes and the Pandas library. The graphs created can also be customized easily.

Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance.

Model_selection is a method for setting a blueprint to analyze data and then using it to measure new data. Selecting a proper model allows you to generate accurate results when making a prediction. To do that, you need to train your model by using a specific dataset.

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.

A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. Stop Words: A stop word is a commonly used word (such as “the”, “a”, “an”, “in”) that a search engine has been programmed to ignore

It's a built-in module and we have to import it before using any of its constants and classes.

This module implements specialized container datatypes providing alternatives to Python's general purpose built-in containers, dict , list , set , and tuple .

In [16]:
```
import os
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from wordcloud import WordCloud
from sklearn.model_selection import train_test_split
import numpy as np
import re
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords
import string
from nltk.stem.porter import PorterStemmer
from collections import Counter
import nltk
```

In [17]:
```
df = pd.read_json('sentiment_anlysis_twitter2.json')
```

In [18]:
```
df
```

Out[18]:

	in_reply_to_screen_name	quoted_status_id	source	text	quoted_status_id_str	id_str	in_reply_to_status
0	None	NaN	Techie Wiz bot	RT @SalesTechStar: Anova Elevates The Bar On C...	NaN	1468178182335520768	
1	None	NaN	Twitter Web App	RT @shivKR007: Main Stream Media is silent Aft...	NaN	1468178182012502016	
2	None	NaN	Twitter for Android	RT @RealDLHughley: Former Restaurant Workers A...	NaN	1468178180762656768	
3	None	1.467952e+18	Twitter for Android	RT @datingdecisions: One time early in my care...	1.467952e+18	1468178180548837376	
4	None	NaN	Twitter Web App	RT @shayararar: Reporters from all major news ...	NaN	1468178179550367744	
...	...	...	...	...	...	...	
795	AmandaDupont	NaN	Twitter for Android	@AmandaDupont calling @official_jubjub into or...	NaN	1468177949803397120	
796	None	NaN	Twitter for iPad	RT @paddydocherty: PROPOSAL: we abolish billio...	NaN	1468177949669171200	
797	None	NaN	Twitter for iPad	RT @realTuckFrumper: Why Trump hasn't been cha...	NaN	1468177949597782016	
798	None	NaN	Twitter Web App	RT @MForstater: Transgender critics will be pr...	NaN	1468177949203615744	
799	None	NaN	Twitter for Android	RT @SIPTU: Strike begins at Job Clubs in Offal...	NaN	1468177949027414016	

800 rows × 15 columns

In [19]:
```
df['text'][0]
```

Out[19]:
```
'RT @SalesTechStar: Anova Elevates The Bar On Consumer Engagement And Empowerment With The Release Of Anova Co
nnect™ https://t.co/7IBSPQhPGG…' (https://t.co/7IBSPQhPGG…')
```

In [20]:
```
df['text'][2]
```

Out[20]:
```
'RT @RealDLHughley: Former Restaurant Workers Are Sharing Just How "Damaging" The Service Industry Is, And I
\'m Equal Parts Enraged And Heart…'
```

In [21]:
```
df
```

Out[21]:

	in_reply_to_screen_name	quoted_status_id	source	text	quoted_status_id_str	id_str	in_reply_to_status
0	None	NaN	Techie Wiz bot	RT @SalesTechStar: Anova Elevates The Bar On C...	NaN	1468178182335520768	
1	None	NaN	Twitter Web App	RT @shivKR007: Main Stream Media is silent Aft...	NaN	1468178182012502016	
2	None	NaN	Twitter for Android	RT @RealDLHughley: Former Restaurant Workers A...	NaN	1468178180762656768	
3	None	1.467952e+18	Twitter for Android	RT @datingdecisions: One time early in my care...	1.467952e+18	1468178180548837376	
4	None	NaN	Twitter Web App	RT @shayararar: Reporters from all major news ...	NaN	1468178179550367744	
...	...	...	...	...	...	...	
795	AmandaDupont	NaN	Twitter for Android	@AmandaDupont calling @official_jubjub into or...	NaN	1468177949803397120	
796	None	NaN	Twitter for iPad	RT @paddydocherty: PROPOSAL: we abolish billio...	NaN	1468177949669171200	
797	None	NaN	Twitter for iPad	RT @realTuckFrumper: Why Trump hasn't been cha...	NaN	1468177949597782016	
798	None	NaN	Twitter Web App	RT @MForstater: Transgender critics will be pr...	NaN	1468177949203615744	
799	None	NaN	Twitter for Android	RT @SIPTU: Strike begins at Job Clubs in Offal...	NaN	1468177949027414016	

800 rows × 15 columns

Stemming is the process of producing morphological variants of a root/base word.

In [22]:
```
ps = PorterStemmer()
```

The isalnum() method returns True if all the characters are alphanumeric, meaning alphabet letter (a-z) and numbers (0-9).

In [23]:
```
def tranform_text(text):

    #Convert into
    text = text.lower()
    text = nltk.word_tokenize(text)

    y = []
    for i in text:
        if i.isalnum():
            y.append(i)


    text = y[:]
    y.clear()

    custom_list = ['rt']
    for i in text:
        if i not in stopwords.words('english') and i not in string.punctuation:
            y.append(i)
    print(y)

    return " ".join(y)
```

In [24]:
```
df['transformed_text'] = df['text'].apply(tranform_text)
```

```
ps']
['news24', 'bisouthafrica', 'yes', 'old', 'news']
['rt', 'alanpps', 'men', 'killed', 'bradley', 'gledhill', 'batley', 'jail', 'terms', 'extended', '28', 'yea
rs', 'https']
['chronopost', 'lemondefr', 'least', 'chronopost', 'able', 'make', 'délivery', 'worst', 'company', 'franc
e']
['gayeonjunie', 'news']
['vajihaqureshi', 'hello', 'vajiha', 'thank', 'reaching', 'us', 'sincerely', 'sorry', 'read', 'please', 'ht
tps']
['rt', 'bharadwajspeaks', 'situation', 'worse', 'pakistan', 'next', 'day', 'mobs', 'destroyed', 'ancient',
'hindu', 'temple', 'rawalpindi']
['prosecute', 'crook', 'cameron', 'crimes', 'https']
['int', 'l', 'commission', 'allows', '15', 'rise', 'pacific', 'bluefin', 'tuna', 'catch', 'limit', 'https']
['rt', 'valuesoffrench', 'beyond', 'devastated', 'passing', 'dear', 'friend', 'simon', 'always', 'joy', 'wo
rk', 'generous']
['rt', 'chadocl', 'news', 'forthcoming', 'us', 'diplomatic', 'boycott', 'makes', 'already', 'small', 'chanc
e', 'winter', 'olympic', 'breakthrough', 'north']
['rt', 'bollyhungama', 'salmankhan', 'gave', 'atrangire', 'title', 'aanandlrai', 'reason', 'https']
['rt', 'variety', 'colin', 'farrell', 'penguin', 'return', 'hbo', 'max', 'series', 'thebatman', 'https']
['5', 'promises', 'people', 'scheduled', 'caste', 'provide', 'free', 'education', 'children', 'free', 'coac
```

In [25]:
```
df
```

Out[25]:

	in_reply_to_screen_name	quoted_status_id	source	text	quoted_status_id_str	id_str	in_reply_to_s
0	None	NaN	Techie Wiz bot	RT @SalesTechStar: Anova Elevates The Bar On C...	NaN	1468178182335520768	
1	None	NaN	Twitter Web App	RT @shivKR007: Main Stream Media is silent Aft...	NaN	1468178182012502016	
2	None	NaN	Twitter for Android	RT @RealDLHughley: Former Restaurant Workers A...	NaN	1468178180762656768	
3	None	1.467952e+18	Twitter for Android	RT @datingdecisions: One time early in my care...	1.467952e+18	1468178180548837376	
4	None	NaN	Twitter Web App	RT @shayararar: Reporters from all major news ...	NaN	1468178179550367744	
...	...	...	...	...	...	...	
795	AmandaDupont	NaN	Twitter for Android	@AmandaDupont calling @official_jubjub into or...	NaN	1468177949803397120	
796	None	NaN	Twitter for iPad	RT @paddydocherty: PROPOSAL: we abolish billio...	NaN	1468177949669171200	
797	None	NaN	Twitter for iPad	RT @realTuckFrumper: Why Trump hasn't been cha...	NaN	1468177949597782016	
798	None	NaN	Twitter Web App	RT @MForstater: Transgender critics will be pr...	NaN	1468177949203615744	

	in_reply_to_screen_name	quoted_status_id	source	text	quoted_status_id_str	id_str	in_reply_to_s
799	None	NaN	Twitter for Android	RT @SIPTU: Strike begins at Job Clubs in Offal...	NaN	1468177949027414016	

800 rows × 16 columns

In [26]:
```
df
```

Out[26]:

	in_reply_to_screen_name	quoted_status_id	source	text	quoted_status_id_str	id_str	in_reply_to_s
0	None	NaN	Techie Wiz bot	RT @SalesTechStar: Anova Elevates The Bar On C...	NaN	1468178182335520768	
1	None	NaN	Twitter Web App	RT @shivKR007: Main Stream Media is silent Aft...	NaN	1468178182012502016	
2	None	NaN	Twitter for Android	RT @RealDLHughley: Former Restaurant Workers A...	NaN	1468178180762656768	
3	None	1.467952e+18	Twitter for Android	RT @datingdecisions: One time early in my care...	1.467952e+18	1468178180548837376	
4	None	NaN	Twitter Web App	RT @shayararar: Reporters from all major news ...	NaN	1468178179550367744	
...	...	...	...	...	...	...	
795	AmandaDupont	NaN	Twitter for Android	@AmandaDupont calling @official_jubjub into or...	NaN	1468177949803397120	
796	None	NaN	Twitter for iPad	RT @paddydocherty: PROPOSAL: we abolish billio...	NaN	1468177949669171200	
797	None	NaN	Twitter for iPad	RT @realTuckFrumper: Why Trump hasn't been cha...	NaN	1468177949597782016	
798	None	NaN	Twitter Web App	RT @MForstater: Transgender critics will be pr...	NaN	1468177949203615744	

	in_reply_to_screen_name	quoted_status_id	source	text	quoted_status_id_str	id_str	in_reply_to_s
799	None	NaN	Twitter for Android	RT @SIPTU: Strike begins at Job Clubs in Offal...	NaN	1468177949027414016	

800 rows × 16 columns

In [27]:
```
df.columns
```

Out[27]:
```
Index(['in_reply_to_screen_name', 'quoted_status_id', 'source', 'text',
       'quoted_status_id_str', 'id_str', 'in_reply_to_status_id_str',
       'in_reply_to_user_id_str', 'retweet_count', 'favorite_count',
       'in_reply_to_status_id', 'in_reply_to_user_id', 'lang', 'id',
       'source_url', 'transformed_text'],
      dtype='object')
```

In [28]:
```
df2 = df.drop(columns=[ 'lang', 'id_str', 'text', 'in_reply_to_screen_name',
       'source', 'in_reply_to_status_id_str', 'quoted_status_id', 'id',
       'in_reply_to_user_id_str', 'in_reply_to_user_id', 'favorite_count',
       'in_reply_to_status_id', 'quoted_status_id_str', 'source_url'],axis=1)
```

In [29]:
```
df2
```

Out[29]:

	retweet_count	transformed_text
0	1	rt salestechstar anova elevates bar consumer e...
1	174	rt shivkr007 main stream media silent watching...
2	3	rt realdlhughley former restaurant workers sha...
3	2	rt datingdecisions one time early career forwa...
4	11	rt shayararar reporters major news channels wa...
...	...	...
795	0	amandadupont calling amandadupont macgpodcasta...
796	2	rt paddydocherty proposal abolish billionaires...
797	57	rt realtuckfrumper trump charged obstruction e...
798	159	rt mforstater transgender critics protected ne...
799	15	rt siptu strike begins job clubs offaly dealt ...

800 rows × 2 columns

In [30]:
```
df2.to_json("Sentiment_anlysis_file.json")
```

In [ ]:
```

```