

SENTIMENT ANALYSIS SOCIAL MEDIA PLATFORM

In [1]:

```
import pandas as pd
```

In [2]:

```
# Loading the data set
social_media = pd.read_csv('Twitter dataset.csv')
social_media
```

Out[2]:

	Tweet_ID	Username	Text	Retweets	Likes	Timestamp
0	1	julie81	Party least receive say or single. Prevent pre...	2	25	2023-01-30 11:00:51
1	2	richardhester	Hotel still Congress may member staff. Media d...	35	29	2023-01-02 22:45:58
2	3	williamsjoseph	Nice be her debate industry that year. Film wh...	51	25	2023-01-18 11:25:19
3	4	danielsmary	Laugh explain situation career occur serious. ...	37	18	2023-04-10 22:06:29
4	5	carlwarren	Involve sense former often approach government...	27	80	2023-01-24 07:12:21
...
9995	9996	ntate	Agree reflect military box ability ever hold. ...	81	86	2023-01-15 11:46:20
9996	9997	garrisonjoshua	Born which push still. Degree sometimes contro...	73	100	2023-05-06 00:46:54
9997	9998	adriennejackson	You day agent likely region. Teacher data mess...	10	62	2023-02-27 14:55:08
9998	9999	kcarlson	Guess without successful save. Particular natu...	21	60	2023-01-09 16:09:35
9999	10000	vickerson	Body onto understand team about product beauti...	65	54	2023-04-19 01:35:56

10000 rows x 6 columns

In [3]:

```
social_media.drop(['Timestamp'],axis=1)
```

Out[3]:

	Tweet_ID	Username	Text	Retweets	Likes
0	1	julie81	Party least receive say or single. Prevent pre...	2	25
1	2	richardhester	Hotel still Congress may member staff. Media d...	35	29
2	3	williamsjoseph	Nice be her debate industry that year. Film wh...	51	25
3	4	danielsmary	Laugh explain situation career occur serious. ...	37	18
4	5	carlwarren	Involve sense former often approach government...	27	80
...
9995	9996	ntate	Agree reflect military box ability ever hold. ...	81	86
9996	9997	garrisonjoshua	Born which push still. Degree sometimes contro...	73	100
9997	9998	adriennejackson	You day agent likely region. Teacher data mess...	10	62
9998	9999	kcarlson	Guess without successful save. Particular natu...	21	60
9999	10000	vickerson	Body onto understand team about product beauti...	65	54

10000 rows x 5 columns

In [4]:

```
# data cleaning
social_media.shape
```

Out[4]:

```
(10000, 6)
```

In [5]:

```
social_media.info()
```

Out[5]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype
---  --
 0   Tweet_ID    10000 non-null  int64
 1   Username    10000 non-null  object
 2   Text        10000 non-null  object
 3   Retweets    10000 non-null  int64
 4   Likes       10000 non-null  int64
 5   Timestamp   10000 non-null  object
dtypes: int64(3), object(3)
memory usage: 488.9+ KB
```

In [6]:

```
social_media.describe()
```

Out[6]:

	Tweet_ID	Retweets	Likes
count	10000.000000	10000.000000	10000.000000
mean	5000.500000	49.721200	49.929300
std	2586.895568	28.944856	28.877193
min	1.000000	0.000000	0.000000
25%	2500.750000	25.000000	25.000000
50%	5000.500000	49.000000	50.000000
75%	7500.250000	75.000000	75.000000
max	10000.000000	100.000000	100.000000

In [7]:

```
social_media['Likes'].mean()
```

Out[7]:

```
49.9293
```

In [8]:

```
social_media.median()
```

Out[8]:

```
Tweet_ID    5000.5
Retweets     49.0
Likes       50.0
dtype: float64
```

In [9]:

```
# how many null values
social_media.isnull().sum()
```

Out[9]:

```
Tweet_ID      0
Username      0
Text          0
Retweets      0
Likes         0
Timestamp     0
dtype: int64
```

In [10]:

```
# stopwords removal
import nltk
from nltk.corpus import stopwords
```

In [10]:

```
nltk_stopword =set(stopwords.words('english'))
print(nltk_stopword)
```

In [10]:

```
{'no', 'hadn't', 'with', 'during', 'here', 'weren', 'the', 'against', 'don', 'can', 'few', 'its', 'them', 'into', 'this', 'between', 'down', 'shoul
d', 'again', 'before', 'through', 'were', 'mightn', 'are', 'i', 'yours', 'in', 'same', 'whom', 'at', 'has', 'and', 'hadn', 'under', 'was',
'hasn't', 'couldn', 'from', 'of', 'out', 'he', 'needn't', 'who', 'did', 'my', 'because', 'shouldn't', 're', 'you're', 'how', 'as', 'ours', 'so', 'm
e', 'ourselves', 'you'd', 'theirs', 'if', 'that', 'it', 'own', 'it's', 'be', 'having', 'but', 'shan', 'for', 'which', 'why', 'most', 'off', 'furthe
r', 'hers', 'now', 'herself', 'any', 'these', 'other', 'is', 'too', 'than', 'have', 'a', 'once', 'such', 'to', 'while', 'a', 'has n
t', 'needn', 'just', 'by', 'we', 'mightn't', 'do', 'll', 'than', 'isn't', 'doesn', 'have', 'haven', 'ma', 'a', 'hasn', 'once', 'didn't', 'such', 'di
d n't', 'while', 'an', 'there', 'then', 'when', 'should've', 'been', 'does', 'until', 'wasn', 'our', 'she', 'after', 'over', 'not', 'doesn't', 't
he', 'weren't', 'both', 'more', 'o', 'where', 'itself', 'stagn', 'y', 'am', 'your', 'haven't', 'i', 'aren't', 'some', 'shouldn', 'being',
'had', 'their', 'him', 'haven't', 'nor', 'hisself', 'only', 'on', 'doing', 'myself', 'those', 'won't', 'above', 'won', 'm', 'ain', 'you', 'all', 'i
sn', 'each', 's', 'wouldn', 'his', 'yourself', 'up', 'yourselves', 'aren', 'wouldn't', 'or', 'very', 'will', 'couldn't', 'she's', 'about', 'themsel
ves'}
```

In [10]:

```
len(nltk_stopword)
```

Out[10]:

```
179
```

In [10]:

```
from sklearn.feature_extraction.text import ENGLISH_STOP_WORDS
stop = set(ENGLISH_STOP_WORDS)
sklearn
```

Out[10]:

```
{',',
 'about',
 'above',
 'across',
 'after',
 'afterwards',
 'again',
 'against',
 'all',
 'almost',
 'alone',
 'along',
 'already',
 'also',
 'although',
 'always',
 'am',
 'among',
 'amongst',
 'amount',
 'an',
 'and',
 'another',
 'any',
 'anyhow',
 'anyone',
 'anything',
 'anywhere',
 'are',
 'are',
 'as',
 'at',
 'back',
 'became',
 'because',
 'become',
 'becomes',
 'becoming',
 'been',
 'between',
 'beforehand',
 'behind',
 'being',
 'below',
 'beside',
 'besides',
 'between',
 'beyond',
 'bill',
 'both',
 'bottom',
 'but',
 'by',
 'call',
 'can',
 'cannot',
 'cant',
 'co',
 'could',
 'couldn',
 'couldn',
 'de',
 'describe',
 'detail',
 'do',
 'does',
 'down',
 'due',
 'during',
 'each',
 'eg',
 'either',
 'else',
 'elsewhere',
 'empty',
 'enough',
 'etc',
 'ever',
 'every',
 'everyone',
 'everything',
 'everywhere',
 'except',
 'few',
 'fifteen',
 'fifty',
 'fill',
 'find',
 'fire',
 'first',
 'first',
 'five',
 'for',
 'former',
 'formerly',
 'forty',
 'four',
 'from',
 'front',
 'full',
 'further',
 'get',
 'give',
 'go',
 'had',
 'has',
 'hasn',
 'he',
 'here',
 'hereafter',
 'hereby',
 'herein',
 'hereupon',
 'hers',
 'herself',
 'him',
 'himself',
 'his',
 'how',
 'however',
 'hundred',
 'i',
 'if',
 'in',
 'into',
 'indeed',
 'interest',
 'into',
 'is',
 'it',
 'itself',
 'keep',
 'last',
 'latter',
 'latterly',
 'least',
 'less',
 'ltd',
 'made',
 'may',
 'me',
 'meanwhile',
 'might',
 'mill',
 'more',
 'moreover',
 'most',
 'move',
 'much',
 'must',
 'myself',
 'name',
 'neither',
 'never',
 'nevertheless',
 'next',
 'nine',
 'no',
 'nobody',
 'nor',
 'not',
 'nothing',
 'now',
 'nowhere',
 'of',
 'off',
 'often',
 'on',
 'once',
 'only',
 'onto',
 'or',
 'other',
 'others',
 'otherwise',
 'our',
 'ours',
 'ourselves',
 'out',
 'over',
 'own',
 'perhaps',
 'per',
 'please',
 'put',
 'rather',
 're',
 'same',
 'saw',
 'seem',
 'seemed',
 'seeming',
 'seems',
 'serious',
 'several',
 'she',
 'should',
 'show',
 'side',
 'since',
 'sincere',
 'six',
 'sixty',
 'so',
 'some',
 'someone',
 'something',
 'sometimes',
 'someday',
 'sometimes',
 'still',
 'such',
 'sure',
 'take',
 'than',
 'that',
 'the',
 'their',
 'them',
 'themselves',
 'there',
 'thereafter',
 'thereby',
 'therein',
 'thereupon',
 'these',
 'they',
 'thick',
 'thin',
 'third',
 'this',
 'those',
 'though',
 'three',
 'through',
 'throughout',
 'thru',
 'thus',
 'to',
 'together',
 'too',
 'toward',
 'towards',
 'twenty',
 'two',
 'um',
 'under',
 'until',
 'up',
 'upon',
 'us',
 'very',
 'via',
 'was',
 'we',
 'were',
 'what',
 'whatever',
 'when',
 'whence',
 'whenever',
 'where',
 'whereafter',
 'whereas',
 'whereby',
 'wherein',
 'whereupon',
 'wherever',
 'whether',
 'which',
 'while',
 'whither',
 'who',
 'whoever',
 'whom',
 'whose',
 'why',
 'will',
 'with',
 'within',
 'without',
 'would',
 'yet',
 'you',
 'your',
 'yours',
 'yourself',
 'yourselves'}
```

In [10]:

```
len(sklearn)
```

Out[10]:

```
318
```

In [10]:

```
# find the comma stopwords NLTK and sklearn
print(nltk_stopword.intersection(sklearn))
```

Out[10]:

```
{',',
 'about',
 'above',
 'across',
 'after',
 'afterwards',
 'again',
 'against',
 'all',
 'almost',
 'alone',
 'along',
 'already',
 'also',
 'although',
 'always',
 'am',
 'among',
 'amongst',
 'amount',
 'an',
 'and',
 'another',
 'any',
 'anyhow',
 'anyone',
 'anything',
 'anywhere',
 'are',
 'are',
 'as',
 'at',
 'back',
 'became',
 'because',
 'become',
 'becomes',
 'becoming',
 'been',
 'between',
 'beforehand',
 'behind',
 'being',
 'below',
 'beside',
 'besides',
 'between',
 'beyond',
 'bill',
 'both',
 'bottom',
 'but',
 'by',
 'call',
 'can',
 'cannot',
 'cant',
 'co',
 'could',
 'couldn',
 'couldn',
 'de',
 'describe',
 'detail',
 'do',
 'does',
 'down',
 'due',
 'during',
 'each',
 'eg',
 'either',
 'else',
 'elsewhere',
 'empty',
 'enough',
 'etc',
 'ever',
 'every',
 'everyone',
 'everything',
 'everywhere',
 'except',
 'few',
 'fifteen',
 'fifty',
 'fill',
 'find',
 'fire',
 'first',
 'first',
 'five',
 'for',
 'former',
 'formerly',
 'forty',
 'four',
 'from',
 'front',
 'full',
 'further',
 'get',
 'give',
 'go',
 'had',
 'has',
 'hasn',
 'he',
 'here',
 'hereafter',
 'hereby',
 'herein',
 'hereupon',
 'hers',
 'herself',
 'him',
 'himself',
 'his',
 'how',
 'however',
 'hundred',
 'i',
 'if',
 'in',
 'into',
 'indeed',
 'interest',
 'into',
 'is',
 'it',
 'itself',
 'keep',
 'last',
 'latter',
 'latterly',
 'least',
 'less',
 'ltd',
 'made',
 'may',
 'me',
 'meanwhile',
 'might',
 'mill',
 'more',
 'moreover',
 'most',
 'move',
 'much',
 'must',
 'myself',
 'name',
 'neither',
 'never',
 'nevertheless',
 'next',
 'nine',
 'no',
 'nobody',
 'nor',
 'not',
 'nothing',
 'now',
 'nowhere',
 'of',
 'off',
 'often',
 'on',
 'once',
 'only',
 'onto',
 'or',
 'other',
 'others',
 'otherwise',
 'our',
 'ours',
 'ourselves',
 'out',
 'over',
 'own',
 'perhaps',
 'per',
 'please',
 'put',
 'rather',
 're',
 'same',
 'saw',
 'seem',
 'seemed',
 'seeming',
 'seems',
 'serious',
 'several',
 'she',
 'should',
 'show',
 'side',
 'since',
 'sincere',
 'six',
 'sixty',
 'so',
 'some',
 'someone',
 'something',
 'sometimes',
 'someday',
 'sometimes',
 'still',
 'such',
 'sure',
 'take',
 'than',
 'that',
 'the',
 'their',
 'them',
 'themselves',
 'there',
 'thereafter',
 'thereby',
 'therein',
 'thereupon',
 'these',
 'they',
 'thick',
 'thin',
 'third',
 'this',
 'those',
 'though',
 'three',
 'through',
 'throughout',
 'thru',
 'thus',
 'to',
 'together',
 'too',
 'toward',
 'towards',
 'twenty',
 'two',
 'um',
 'under',
 'until',
 'up',
 'upon',
 'us',
 'very',
 'via',
 'was',
 'we',
 'were',
 'what',
 'whatever',
 'when',
 'whence',
 'whenever',
 'where',
 'whereafter',
 'whereas',
 'whereby',
 'wherein',
 'whereupon',
 'wherever',
 'whether',
 'which',
 'while',
 'whither',
 'who',
 'whoever',
 'whom',
 'whose',
 'why',
 'will',
 'with',
 'within',
 'without',
 'would',
 'yet',
 'you',
 'your',
 'yours',
 'yourself',
 'yourselves'}
```

In [10]:

```
len(nltk_stopword.intersection(sklearn))
```

Out[10]:

```
119
```

In [10]:

```
# combine the stop words sklearn and NLTK(NLP)
combine_stopword =nltk_stopword.union(sklearn)
print(combine_stopword)
```

Out[10]:

```
{',',
 'about',
 'above',
 'across',
 'after',
 'afterwards',
 'again',
 'against',
 'all',
 'almost',
 'alone',
 'along',
 'already',
 'also',
 'although',
 'always',
 'am',
 'among',
 'amongst',
 'amount',
 'an',
 'and',
 'another',
 'any',
 'anyhow',
 'anyone',
 'anything',
 'anywhere',
 'are',
 'are',
 'as',
 'at',
 'back',
 'became',
 'because',
 'become',
 'becomes',
 'becoming',
 'been',
 'between',
 'beforehand',
 'behind',
 'being',
 'below',
 'beside',
 'besides',
 'between',
 'beyond',
 'bill',
 'both',
 'bottom',
 'but',
 'by',
 'call',
 'can',
 'cannot',
 'cant',
 'co',
 'could',
 'couldn',
 'couldn',
 'de',
 'describe',
 'detail',
 'do',
 'does',
 'down',
 'due',
 'during',
 'each',
 'eg',
 'either',
 'else',
 'elsewhere',
 'empty',
 'enough',
 'etc',
 'ever',
 'every',
 'everyone',
 'everything',
 'everywhere',
 'except',
 'few',
 'fifteen',
 'fifty',
 'fill',
 'find',
 'fire',
 'first',
 'first',
 'five',
 'for',
 'former',
 'formerly',
 'forty',
 'four',
 'from',
 'front',
 'full',
 'further',
 'get',
 'give',
 'go',
 'had',
 'has',
 'hasn',
 'he',
 'here',
 'hereafter',
 'hereby',
 'herein',
 'hereupon',
 'hers',
 'herself',
 'him',
 'himself',
 'his',
 'how',
 'however',
 'hundred',
 'i',
 'if',
 'in',
 'into',
 'indeed',
 'interest',
 'into',
 'is',
 'it',
 'itself',
 'keep',
 'last',
 'latter',
 'latterly',
 'least',
 'less',
 'ltd',
 'made',
 'may',
 'me',
 'meanwhile',
 'might',
 'mill',
 'more',
 'moreover',
 'most',
 'move',
 'much',
 'must',
 'myself',
 'name',
 'neither',
 'never',
 'nevertheless',
 'next',
 'nine',
 'no',
 'nobody',
 'nor',
 'not',
 'nothing',
 'now',
 'nowhere',
 'of',
 'off',
 'often',
 'on',
 'once',
 'only',
 'onto',
 'or',
 'other',
 'others',
 'otherwise',
 'our',
 'ours',
 'ourselves',
 'out',
 'over',
 'own',
 'perhaps',
 'per',
 'please',
 'put',
 'rather',
 're',
 'same',
 'saw',
 'seem',
 'seemed',
 'seeming',
 'seems',
 'serious',
 'several',
 'she',
 'should',
 'show',
 'side',
 'since',
 'sincere',
 'six',
 'sixty',
 'so',
 'some',
 'someone',
 'something',
 'sometimes',
 'someday',
 'sometimes',
 'still',
 'such',
 'sure',
 'take',
 'than',
 'that',
 'the',
 'their',
 'them',
 'themselves',
 'there',
 'thereafter',
 'thereby',
 'therein',
 'thereupon',
 'these',
 'they',
 'thick',
 'thin',
 'third',
 'this',
 'those',
 'though',
 'three',
 'through',
 'throughout',
 'thru',
 'thus',
 'to',
 'together',
 'too',
 'toward',
 'towards',
 'twenty',
 'two',
 'um',
 'under',
 'until',
 'up',
 'upon',
 'us',
 'very',
 'via',
 'was',
 'we',
 'were',
 'what',
 'whatever',
 'when',
 'whence',
 'whenever',
 'where',
 'whereafter',
 'whereas',
 'whereby',
 'wherein',
 'whereupon',
 'wherever',
 'whether',
 'which',
 'while',
 'whither',
 'who',
 'whoever',
 'whom',
 'whose',
 'why',
 'will',
 'with',
 'within',
 'without',
 'would',
 'yet',
 'you',
 'your',
 'yours',
 'yourself',
 'yourselves'}
```

In [10]:

```
len(combine_stopword)
```

Out[10]:

```
378
```

In [10]:

```
# test Normalization stemming and lematization
# stemming is faster as it cuts words without knowing the context,while lematization is slower as it knows the context
# of words before
from nltk.stem import wordnetLemmatizer
lemmatizer =wordnetLemmatizer()
```

Out[10]:

```
['Indicate',
 'memory',
 'manage',
 'fact',
 'Ready',
 'fund',
 'law',
 'scientist',
 'federal',
 'case',
 'According',
 'market',
 'anything',
 'them',
 'provide',
 'Newspaper',
 'he',
 'social',
 'more',
 'size',
 'paper',
 'morning',
 'report',
 'National',
 'up',
 'date',
 'many',
 'Staff',
 'position',
 'statement',
 'partner',
 'Organization',
 'vote',
 'beyond',
 'not',
 'security',
 'box. ']
```

In [10]:

```
new_text=""
for token in social_media['Text'][:100].split():
    new_set = new +lemmatizer.lemmatize(token.lower())+' '
new_set=new
```

Out[10]:

```
'box. '
```

In [10]:

```
!pip install contractions
```

Out[10]:

```
Collecting contractions
  Downloading contractions-0.1.73-py2.py3-none-any.whl (8.7 kB)
Collecting textsearch>=0.0.21
  Downloading textsearch-0.0.24-py2.py3-none-any.whl (7.6 kB)
Collecting anyascii
  Downloading anyascii-0.3.2-py3-none-any.whl (289 KB)
Collecting pyahocorasick
  Downloading pyahocorasick-2.0.0-cp38-cp38-win amd64.whl (39 KB)
Installing collected packages: pyahocorasick, anyascii, textsearch, contractions
Successfully installed anyascii-0.3.2 contractions-0.1.73 pyahocorasick-2.0.0 textsearch-0.0.24
```

In [10]:

```
import contractions
social_media['Text'][:50]
```

Out[10]:

```
'Card left quickly college return. Speak with account here thousand turn.\nSurface play their state space born. Boy firm arm seat little later. The
n chance indicate between thus impact nor.'
```

In [10]:

```
contractions.fix(social_media['Text'][:50])
```

Out[10]:

```
'Card left quickly college return. Speak with account here thousand turn.\nSurface play their state space born. Boy firm arm seat little later. The
n chance indicate between thus impact nor.'
```

In [10]:

```
def cleaner_function(text):
    new_text=re.sub("s|o|is",text)
    new_text=re.sub("g|u",new_text)
    new_text=contractions.fix(new_text)
    new_text=new_text.lower().strip()

    clean_text=""
    for token in new_text.split():
        cleaned_text=clean_text+lemmatizer.lemmatize(token.lower())+' '
        cleaned_text
```

In [10]:

```
# data visualization
all_word=[]
for k in social_media['Text']:
    all_word.extend(k.split())
print(all_word[:48])
len(set(all_word))
```

Out[10]:

```
['Party', 'least', 'receive', 'say', 'or', 'single', 'Prevent', 'prevent', 'husband', 'affect.', 'May', 'himself', 'cup', 'style', 'evening', 'pro
tect', 'Effect', 'another', 'themselves', 'stage', 'perform', 'possible', 'try', 'tax', 'share', 'television', 'with.', 'Successful', 'm
uch', 'sell', 'development', 'economy', 'effect', 'Metel', 'still', 'Congress', 'may', 'member', 'staff. ']
```

In [10]:

```
2986
```


In [10]:

```
import matplotlib.pyplot as plt
```

In [10]:

```
# frequency distribution and clean data in tweets
frequency =nltk.FreqDist(all_word)
plt.figure(figsize=(10,5))
plt.title('most common word')
plt.xticks(fontsize=12)
frequency.plot(20,cumulative=False)
plt.show()
```

Out[10]:



In [10]: