In [3]:
```
#Eda feature engineering
import pandas as pd
import numpy as np

#visulazation packages
import matplotlib.pyplot as plt
import seaborn as sns

#text cleaning data
import neattext.functions as nfx

#load ml packages and estimators
from sklearn.linear_model import LogisticRegression

from sklearn.naive_bayes import MultinomialNB

#transformers
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
```

In [5]:
```
import os
os.chdir(r'C:\Users\91866\Documents\SENTIMENT ANALYSIS')
```

In [7]:
```
df = pd.read_csv('emotion_dataset_raw.csv')
```

In [9]:
```
df.head()
```

Out[9]:

	Emotion	Text
0	neutral	Why ?
1	joy	Sage Act upgrade on my to do list for tommorow.
2	sadness	ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN ...
3	joy	Such an eye ! The true hazel eye-and so brill...
4	joy	@Iluvmiasantos ugh babe.. hugggzzz for u .! b...

In [11]:
```
df.shape
```

Out[11]:
```
(34792, 2)
```

In [13]:
```
df.size
```

Out[13]:
```
69584
```

In [15]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34792 entries, 0 to 34791
Data columns (total 2 columns):
 #   Column   Non-Null Count  Dtype 
---  ------   --------------  ----- 
 0   Emotion  34792 non-null  object
 1   Text     34792 non-null  object
dtypes: object(2)
memory usage: 543.8+ KB
```

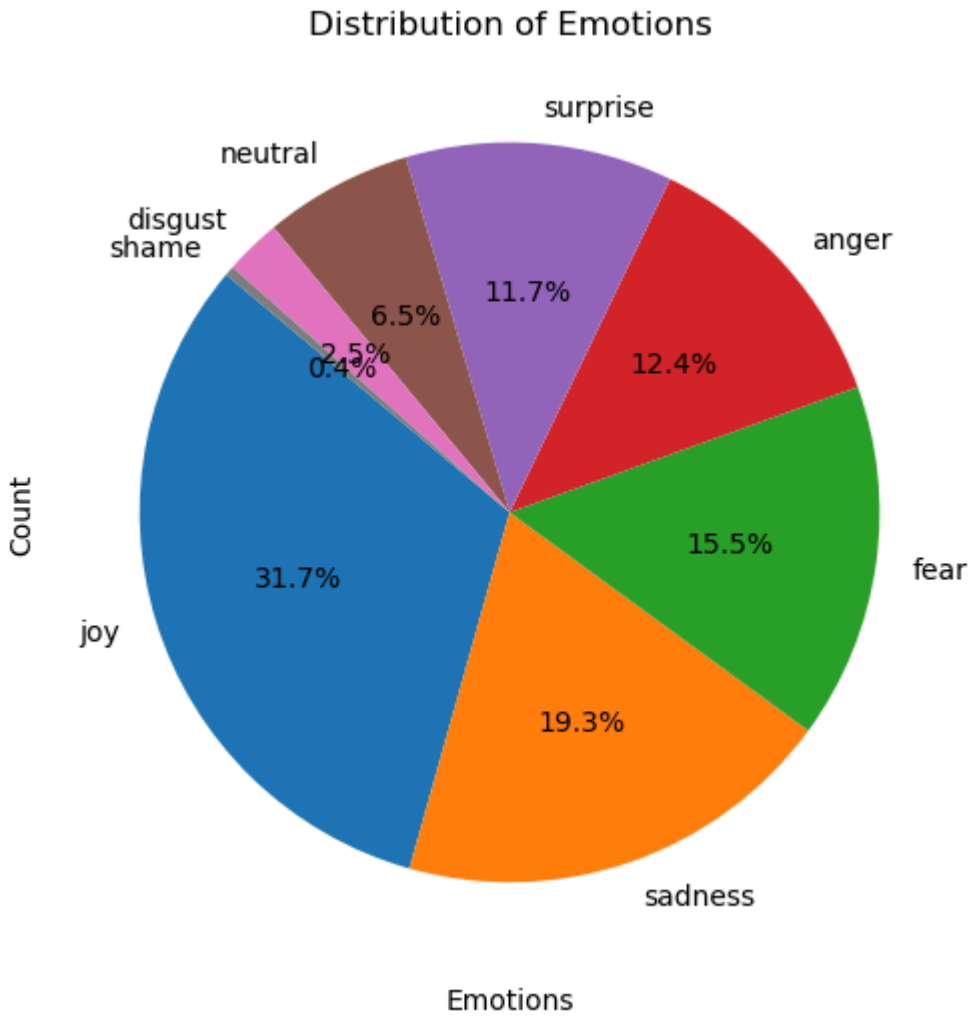
In [17]:
```
df.isna().any()
```

Out[17]:
```
Emotion    False
Text       False
dtype: bool
```

In [19]:
```
df['Emotion'].value_counts()
```

Out[19]:
```
Emotion
joy         11045
sadness      6722
fear         5410
anger        4297
surprise     4062
neutral      2254
disgust       856
shame         146
Name: count, dtype: int64
```

In [21]:
```
# Calculate value counts of emotions
emotion_counts = df['Emotion'].value_counts()

# Create a pie plot using Matplotlib
plt.figure(figsize=(6, 6))
plt.pie(emotion_counts, labels=emotion_counts.index, autopct='%1.1f%%', startang
plt.title('Distribution of Emotions')
plt.xlabel('Emotions')
plt.ylabel('Count')
plt.show()
```


Distribution of Emotions
surprise
neutral
disgust
shame
anger
11.7%
6.5%
2.5%
0.4%
12.4%
Count
15.5%
fear
31.7%
joy
19.3%
sadness
Emotions

In [23]:
```
dir(nfx)
```

Out[23]:
```
['BTC_ADDRESS_REGEX',
 'CURRENCY_REGEX',
 'CURRENCY_SYMB_REGEX',
 'Counter',
 'DATE_REGEX',
 'EMAIL_REGEX',
 'EMOJI_REGEX',
 'HASTAG_REGEX',
 'MASTERCard_REGEX',
 'MD5_SHA_REGEX',
 'MOST_COMMON_PUNCT_REGEX',
 'NUMBERS_REGEX',
 'PHONE_REGEX',
 'PoBOX_REGEX',
 'SPECIAL_CHARACTERS_REGEX',
 'STOPWORDS',
 'STOPWORDS_de',
 'STOPWORDS_en',
 'STOPWORDS_es',
 'STOPWORDS_fr',
 'STOPWORDS_ru',
 'STOPWORDS_yo',
 'STREET_ADDRESS_REGEX',
 'TextFrame',
 'URL_PATTERN',
 'USER_HANDLES_REGEX',
 'VISACard_REGEX',
 '__builtins__',
 '__cached__',
 '__doc__',
 '__file__',
 '__generate_text',
 '__loader__',
 '__name__',
 '__numbers_dict',
 '__package__',
 '__spec__',
 '_lex_richness_herdan',
 '_lex_richness_maas_ttr',
 'clean_text',
 'defaultdict',
 'digit2words',
 'extract_btc_address',
 'extract_currencies',
 'extract_currency_symbols',
 'extract_dates',
 'extract_emails',
 'extract_emojis',
 'extract_hashtags',
 'extract_html_tags',
 'extract_mastercard_addr',
 'extract_md5sha',
 'extract_numbers',
 'extract_pattern',
 'extract_phone_numbers',
 'extract_postoffice_box',
 'extract_shortwords',
 'extract_special_characters',
 'extract_stopwords',
 'extract_street_address',
```

```
'extract_terms_in_bracket',
'extract_urls',
'extract_userhandles',
'extract_visacard_addr',
'fix_contractions',
'generate_sentence',
'hamming_distance',
'inverse_df',
'lexical_richness',
'markov_chain',
'math',
'nlargest',
'normalize',
'num2words',
'random',
're',
'read_txt',
'remove_accents',
'remove_bad_quotes',
'remove_btc_address',
'remove_currencies',
'remove_currency_symbols',
'remove_custom_pattern',
'remove_custom_words',
'remove_dates',
'remove_emails',
'remove_emojis',
'remove_hashtags',
'remove_html_tags',
'remove_mastercard_addr',
'remove_md5sha',
'remove_multiple_spaces',
'remove_non_ascii',
'remove_numbers',
'remove_phone_numbers',
'remove_postoffice_box',
'remove_puncts',
'remove_punctuations',
'remove_shortwords',
'remove_special_characters',
'remove_stopwords',
'remove_street_address',
'remove_terms_in_bracket',
'remove_urls',
'remove_userhandles',
'remove_visacard_addr',
'replace_bad_quotes',
'replace_currencies',
'replace_currency_symbols',
'replace_dates',
'replace_emails',
'replace_emojis',
'replace_numbers',
'replace_phone_numbers',
'replace_special_characters',
'replace_term',
'replace_urls',
'string',
'term_freq',
'to_txt',
```

```
 'unicodedata',
 'word_freq',
 'word_length_freq']
```

In [25]:
```
df['Text']
```

Out[25]:
```
0                                                  Why ?
1            Sage Act upgrade on my to do list for tommorow.
2          ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN ...
3           Such an eye ! The true hazel eye-and so brill...
4          @Iluvmiasantos ugh babe.. hugggzzz for u .!  b...
                               ...
34787      @MichelGW have you gift! Hope you like it! It'...
34788      The world didnt give it to me..so the world MO...
34789                              A man robbed me today .
34790      Youu call it JEALOUSY, I call it of #Losing YO...
34791      I think about you baby, and I dream about you ...
Name: Text, Length: 34792, dtype: object
```

In [27]:
```
#user handles
df['clean_data']=df['Text'].apply(nfx.remove_userhandles)
```

In [29]:
```
df.head(10)
```

Out[29]:

	Emotion	Text	clean_data
0	neutral	Why ?	Why ?
1	joy	Sage Act upgrade on my to do list for tommorow.	Sage Act upgrade on my to do list for tommorow.
2	sadness	ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN ...	ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN ...
3	joy	Such an eye ! The true hazel eye-and so brill...	Such an eye ! The true hazel eye-and so brill...
4	joy	@Iluvmiasantos ugh babe.. hugggzzz for u .! b...	ugh babe.. hugggzzz for u .! babe naamazed ...
5	fear	I'm expecting an extremely important phonecall...	I'm expecting an extremely important phonecall...
6	sadness	.Couldnt wait to see them live. If missing th...	.Couldnt wait to see them live. If missing th...
7	surprise	maken Tip 2: Stop op een moment dat je het hel...	maken Tip 2: Stop op een moment dat je het hel...
8	surprise	En dan krijg je ff een cadeautje van een tweep...	En dan krijg je ff een cadeautje van een tweep...
9	surprise	@1116am Drummer Boy bij op verzoek van @BiemO...	Drummer Boy bij op verzoek van . : welk...

In [31]:
```
#stop words
df['clean_data']=df['clean_data'].apply(nfx.remove_stopwords)
#stopwords the coffe words the and it has is why where
df
```

Out[31]:

	Emotion	Text	clean_data
0	neutral	Why ?	?
1	joy	Sage Act upgrade on my to do list for tommorow.	Sage Act upgrade list tommorow.
2	sadness	ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN ...	WAY HOMEGIRL BABY FUNERAL!!! MAN HATE FUNERALS...
3	joy	Such an eye ! The true hazel eye-and so brill...	eye ! true hazel eye-and brilliant ! Regular f...
4	joy	@Iluvmiasantos ugh babe.. hugggzzz for u .! b...	ugh babe.. hugggzzz u .! babe naamazed nga ako...
...	...	...	...
34787	surprise	@MichelGW have you gift! Hope you like it! It'...	gift! Hope like it! hand wear ! It'll warm! Lol
34788	joy	The world didnt give it to me..so the world MO...	world didnt me..so world DEFINITELY cnt away!!!
34789	anger	A man robbed me today .	man robbed today .
34790	fear	Youu call it JEALOUSY, I call it of #Losing YO...	Youu JEALOUSY, #Losing YOU...
34791	sadness	I think about you baby, and I dream about you ...	think baby, dream time

34792 rows × 3 columns

In [33]:
```
#removing the special character
df['clean_data']=df['clean_data'].apply(nfx.remove_special_characters)
df
```

Out[33]:

	Emotion	Text	clean_data
0	neutral	Why ?	
1	joy	Sage Act upgrade on my to do list for tommorow.	Sage Act upgrade list tommorow
2	sadness	ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN ...	WAY HOMEGIRL BABY FUNERAL MAN HATE FUNERALS SH...
3	joy	Such an eye ! The true hazel eye-and so brill...	eye true hazel eyeand brilliant Regular feat...
4	joy	@Iluvmiasantos ugh babe.. hugggzzz for u .! b...	ugh babe hugggzzz u babe naamazed nga ako e b...
...	...	...	...
34787	surprise	@MichelGW have you gift! Hope you like it! It'...	gift Hope like it hand wear Itll warm Lol
34788	joy	The world didnt give it to me..so the world MO...	world didnt meso world DEFINITELY cnt away
34789	anger	A man robbed me today .	man robbed today
34790	fear	Youu call it JEALOUSY, I call it of #Losing YO...	Youu JEALOUSY Losing YOU
34791	sadness	I think about you baby, and I dream about you ...	think baby dream time

34792 rows × 3 columns

In [35]:
```
x=df.iloc[:,2]
y=df['Emotion']
```

In [37]:
```
x
```

Out[37]:
```
0
1                          Sage Act upgrade list tommorow
2        WAY HOMEGIRL BABY FUNERAL MAN HATE FUNERALS SH...
3        eye  true hazel eyeand brilliant  Regular feat...
4        ugh babe hugggzzz u  babe naamazed nga ako e b...
                              ...
34787           gift Hope like it hand wear  Itll warm Lol
34788           world didnt meso world DEFINITELY cnt away
34789                                    man robbed today
34790                             Youu JEALOUSY Losing YOU
34791                                think baby dream time
Name: clean_data, Length: 34792, dtype: object
```

In [39]:
```
y
```

Out[39]:
```
0         neutral
1             joy
2         sadness
3             joy
4             joy
           ...
34787    surprise
34788         joy
34789       anger
34790        fear
34791     sadness
Name: Emotion, Length: 34792, dtype: object
```

In [41]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.7,random_state=4
```

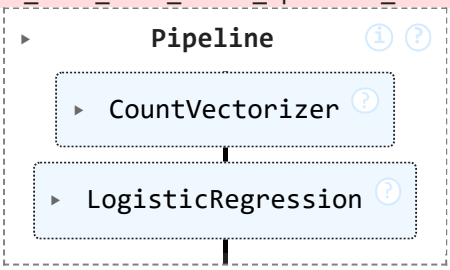
In [43]:
```
from sklearn.pipeline import Pipeline
```

In [45]:
```
#logistic regression pip line

pipe_lr=Pipeline(steps=[('cv',CountVectorizer()),('lr',LogisticRegression())])
```

In [47]:
```
pipe_lr.fit(x_train,y_train)
```

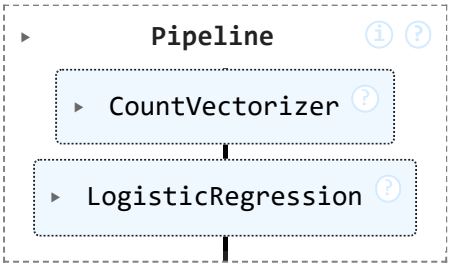
```
C:\Users\91866\anaconda3\Lib\site-packages\sklearn\linear_model\_logistic.py:469:
ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
```

Out[47]:



In [48]:
```
pipe_lr
```

Out[48]:



In [49]:
```
pipe_lr.score(x_test,y_test)
```

Out[49]:
```
0.6217666219582295
```

In [50]:
```
a='this book was so intersting'
pipe_lr.predict([a])
```

```
pipe_lr.predict_proba([a])
```

Out[50]:
```
array([[0.06701364, 0.04964737, 0.07291869, 0.4952967 , 0.02525901,
        0.2777449 , 0.00164602, 0.01047368]])
```

In [51]:
```
pipe_lr.predict_proba([a])# probability for each class
```

Out[51]:
```
array([[0.06701364, 0.04964737, 0.07291869, 0.4952967 , 0.02525901,
        0.2777449 , 0.00164602, 0.01047368]])
```

In [52]:
```
pipe_lr.classes_
```

Out[52]:
```
array(['anger', 'disgust', 'fear', 'joy', 'neutral', 'sadness', 'shame',
       'surprise'], dtype=object)
```

In [53]:
```
import joblib
pipeline_file=open('emotion_classifier.pkl','wb')
joblib.dump(pipe_lr,pipeline_file)
pipeline_file.close()
```

In [ ]:
```

```