```python
In [3]:  import nltk
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         from collections import Counter
         from nltk.tokenize import word_tokenize
         from nltk.corpus import stopwords
         from nltk.sentiment.vader import SentimentIntensityAnalyzer
```

Modi DataSet

```python
In [4]:  modi_df =pd.read_csv('./modi_reviews.csv')
```

```python
In [5]:  modi_df.shape
```

```
Out[5]:  (25688, 3)
```

```python
In [6]:  modi_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25688 entries, 0 to 25687
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  25688 non-null  int64
 1   User        25683 non-null  object
 2   Tweet       25683 non-null  object
dtypes: int64(1), object(2)
memory usage: 602.2+ KB
```

```python
In [7]:  modi_df.head()
```

Out[7]:

	Unnamed: 0	User	Tweet
0	0	advosushildixit	@anjanaomkashyap I am seeing you as future #bj...
1	1	jiaeur	#LokSabhaElections2019 \n23rd May 2019 will re...
2	2	PVenkatGandhi	#LokSabhaElections2019 \n23rd May 2019 will re...
3	3	TheNirbhay1	PM Modi creates a new record of being the only...
4	4	ShakeChilli	@abhijitmajumder Appointment of Successor! \n\...

In [8]:
```python
modi_df['Tweet'] = modi_df['Tweet'].astype(str)
modi_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25688 entries, 0 to 25687
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  25688 non-null  int64
 1   User        25683 non-null  object
 2   Tweet       25688 non-null  object
dtypes: int64(1), object(2)
memory usage: 602.2+ KB
```

In [10]:
```python
tokens = nltk.word_tokenize(modi_df['Tweet'][0])
print(tokens)
```

```
['@', 'anjanaomkashyap', 'I', 'am', 'seeing', 'you', 'as', 'future', '#', 'bj
p', 'spokesperson', '..', 'Good', 'luck', '.', 'Anjana', 'Om', 'Modi', 'oop
s', 'Kashyap', '.', 'Journalists', 'like', 'you', 'changed', 'the', 'meanin
g', 'of', 'journalism', '.', 'Janta', 'maaf', 'nai', 'karege']
```

Rahul Dataset

In [11]:
```python
rahul_df = pd.read_csv('./rahul_reviews.csv')
```

In [12]:
```python
rahul_df.shape
```

Out[12]: (14261, 3)

In [13]:
```python
rahul_df['Tweet'] = rahul_df['Tweet'].astype(str)
rahul_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14261 entries, 0 to 14260
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  14261 non-null  int64
 1   User        14261 non-null  object
 2   Tweet       14261 non-null  object
dtypes: int64(1), object(2)
memory usage: 334.4+ KB
```

In [14]:
```python
rahul_df.head()
```

Out[14]:

	Unnamed: 0	User	Tweet
0	0	Sunnysweet16	Wonder why no academic or journalist asks INC ...
1	1	drnitinchaube	Congrats for the change #australiavotes2019 an...
2	2	mrvivek07	Peopel Say "Govt Ne 70 Years Kya kiya".\nUnse ...
3	3	JosephPravinP	@ajaymaken @RahulGandhi And as a final touch, ...
4	4	VandanaMegastar	#LokSabhaElections2019 Anyone not having mass ...

VADER

In [15]:
```python
from nltk.sentiment import SentimentIntensityAnalyzer
from tqdm.notebook import tqdm

sia = SentimentIntensityAnalyzer()
```

In [16]:
```python
sia.polarity_scores('I am so happy!')
```

Out[16]: {'neg': 0.0, 'neu': 0.318, 'pos': 0.682, 'compound': 0.6468}

In [17]:
```python
sia.polarity_scores('This is the worst thing ever.')
```

Out[17]: {'neg': 0.451, 'neu': 0.549, 'pos': 0.0, 'compound': -0.6249}

In [18]:
```python
sia.polarity_scores(modi_df['Tweet'][0])
```

Out[18]: {'neg': 0.0, 'neu': 0.741, 'pos': 0.259, 'compound': 0.8126}

In [19]:
```python
sia.polarity_scores(rahul_df['Tweet'][0])
```

Out[19]: {'neg': 0.131, 'neu': 0.816, 'pos': 0.054, 'compound': -0.5106}

Run the polarity score on the entire modi dataset

In [20]:
```python
res1 = {}
for i, row in tqdm(modi_df.iterrows(), total=len(modi_df)):
    text = row['Tweet']
    myid = row['Unnamed: 0']
    res1[myid] = sia.polarity_scores(text)
```

100%   25688/25688 [00:14<00:00, 2466.15it/s]

Run the polarity score on the entire rahul dataset

In [21]:
```python
res2 = {}
for i, row in tqdm(rahul_df.iterrows(), total=len(rahul_df)):
    text = row['Tweet']
    myid = row['Unnamed: 0']
    res2[myid] = sia.polarity_scores(text)
```

100%   14261/14261 [00:09<00:00, 1782.52it/s]

In [22]:
```python
print(res1)
```

```
{'neg': 0.0, 'neu': 0.756, 'pos': 0.244, 'compound': 0.875}, 312: {'neg':
0.0, 'neu': 0.611, 'pos': 0.389, 'compound': 0.9022}, 313: {'neg': 0.076,
'neu': 0.924, 'pos': 0.0, 'compound': -0.34}, 314: {'neg': 0.128, 'neu':
0.798, 'pos': 0.074, 'compound': -0.5562}, 315: {'neg': 0.249, 'neu': 0.75
1, 'pos': 0.0, 'compound': -0.6492}, 316: {'neg': 0.141, 'neu': 0.797, 'po
s': 0.062, 'compound': -0.4885}, 317: {'neg': 0.219, 'neu': 0.781, 'pos':
0.0, 'compound': -0.8924}, 318: {'neg': 0.0, 'neu': 0.923, 'pos': 0.077,
'compound': 0.4632}, 319: {'neg': 0.068, 'neu': 0.781, 'pos': 0.15, 'compo
und': 0.5187}, 320: {'neg': 0.215, 'neu': 0.652, 'pos': 0.133, 'compound':
-0.5267}, 321: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 322:
{'neg': 0.051, 'neu': 0.949, 'pos': 0.0, 'compound': -0.4215}, 323: {'ne
g': 0.06, 'neu': 0.94, 'pos': 0.0, 'compound': -0.3182}, 324: {'neg': 0.0,
'neu': 0.929, 'pos': 0.071, 'compound': 0.3818}, 325: {'neg': 0.0, 'neu':
0.912, 'pos': 0.088, 'compound': 0.4114}, 326: {'neg': 0.116, 'neu': 0.76
8, 'pos': 0.117, 'compound': -0.2422}, 327: {'neg': 0.035, 'neu': 0.746,
'pos': 0.219, 'compound': 0.8126}, 328: {'neg': 0.0, 'neu': 0.781, 'pos':
0.219, 'compound': 0.818}, 329: {'neg': 0.0, 'neu': 0.891, 'pos': 0.109,
'compound': 0.4019}, 330: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound':
0.0}, 331: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 332: {'n
eg': 0.13, 'neu': 0.87, 'pos': 0.0, 'compound': -0.4278}, 333: {'neg': 0.
```

In [23]:
```python
print(res2)
```

```
{0: {'neg': 0.131, 'neu': 0.816, 'pos': 0.054, 'compound': -0.5106}, 1:
{'neg': 0.0, 'neu': 0.89, 'pos': 0.11, 'compound': 0.5707}, 2: {'neg': 0.
0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 3: {'neg': 0.0, 'neu': 0.742,
'pos': 0.258, 'compound': 0.916}, 4: {'neg': 0.081, 'neu': 0.887, 'pos':
0.032, 'compound': -0.3252}, 5: {'neg': 0.0, 'neu': 0.94, 'pos': 0.06, 'co
mpound': 0.2023}, 6: {'neg': 0.252, 'neu': 0.709, 'pos': 0.039, 'compoun
d': -0.8689}, 7: {'neg': 0.0, 'neu': 0.782, 'pos': 0.218, 'compound': 0.51
71}, 8: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 9: {'neg':
0.261, 'neu': 0.557, 'pos': 0.182, 'compound': -0.6447}, 10: {'neg': 0.0,
'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 11: {'neg': 0.139, 'neu': 0.789,
'pos': 0.072, 'compound': -0.516}, 12: {'neg': 0.0, 'neu': 0.921, 'pos':
0.079, 'compound': 0.3612}, 13: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'comp
ound': 0.0}, 14: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 1
5: {'neg': 0.0, 'neu': 0.423, 'pos': 0.577, 'compound': 0.6249}, 16: {'ne
g': 0.0, 'neu': 0.898, 'pos': 0.102, 'compound': 0.6249}, 17: {'neg': 0.0,
'neu': 0.667, 'pos': 0.333, 'compound': 0.91}, 18: {'neg': 0.0, 'neu': 0.9
03, 'pos': 0.097, 'compound': 0.4215}, 19: {'neg': 0.0, 'neu': 1.0, 'pos':
0.0, 'compound': 0.0}, 20: {'neg': 0.149, 'neu': 0.565, 'pos': 0.286, 'com
pound': 0.7035}, 21: {'neg': 0.0, 'neu': 0.923, 'pos': 0.077, 'compound':
```

In [24]:
```python
modi_vaders = pd.DataFrame(res1).T
modi_vaders = modi_vaders.reset_index().rename(columns={'index': 'Unnamed: 0'}
modi_vaders = modi_vaders.merge(modi_df, how='left')
```

In [25]:
```python
rahul_vaders = pd.DataFrame(res2).T
rahul_vaders = rahul_vaders.reset_index().rename(columns={'index': 'Unnamed: 0
rahul_vaders = rahul_vaders.merge(rahul_df, how='left')
```

In [26]:
```python
# Now we have sentiment score and metadata
modi_vaders.head()
```

Out[26]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet
0	0	0.000	0.741	0.259	0.8126	advosushildixit	@anjanaomkashyap I am seeing you as future #bj...
1	1	0.170	0.830	0.000	-0.6520	jiaeur	#LokSabhaElections2019 \n23rd May 2019 will re...
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	#LokSabhaElections2019 \n23rd May 2019 will re...
3	3	0.000	0.750	0.250	0.8402	TheNirbhay1	PM Modi creates a new record of being the only...
4	4	0.086	0.764	0.151	0.4312	ShakeChilli	@abhijitmajumder Appointment of Successor! \n\...

In [27]:
```
rahul_vaders.head()
```

Out[27]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet
0	0	0.131	0.816	0.054	-0.5106	Sunnysweet16	Wonder why no academic or journalist asks INC ...
1	1	0.000	0.890	0.110	0.5707	drnitinchaube	Congrats for the change #australiavotes2019 an...
2	2	0.000	1.000	0.000	0.0000	mrvivek07	Peopel Say "Govt Ne 70 Years Kya kiya".\nUnse ...
3	3	0.000	0.742	0.258	0.9160	JosephPravinP	@ajaymaken @RahulGandhi And as a final touch, ...
4	4	0.081	0.887	0.032	-0.3252	VandanaMegastar	#LokSabhaElections2019 Anyone not having mass ...

Peaple with netural Sentiment in modi and Rahul

In [28]:
```python
# Peaple with netural Sentiment in modi
modi_vaders['Label']= np.where(modi_vaders['compound']>0,'positive','negitive'
modi_vaders['Label'][modi_vaders['compound']==0]='Neutral'

# Peaple with netural Sentiment in Rahul
rahul_vaders['Label']= np.where(rahul_vaders['compound']>0,'positive','negitiv
rahul_vaders['Label'][rahul_vaders['compound']==0]='Neutral'
```

```
C:\Users\arnav\AppData\Local\Temp\ipykernel_17456\3578675575.py:3: SettingWit
hCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  modi_vaders['Label'][modi_vaders['compound']==0]='Neutral'
C:\Users\arnav\AppData\Local\Temp\ipykernel_17456\3578675575.py:7: SettingWit
hCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  rahul_vaders['Label'][rahul_vaders['compound']==0]='Neutral'
```

In [29]: `modi_vaders`

Out[29]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Labe
0	0	0.000	0.741	0.259	0.8126	advosushildixit	@anjanaomkashyap I am seeing you as future #bj...	positiv
1	1	0.170	0.830	0.000	-0.6520	jiaeur	#LokSabhaElections2019 \n23rd May 2019 will re...	negitiv
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	#LokSabhaElections2019 \n23rd May 2019 will re...	negitiv
3	3	0.000	0.750	0.250	0.8402	TheNirbhay1	PM Modi creates a new record of being the only...	positiv
4	4	0.086	0.764	0.151	0.4312	ShakeChilli	@abhijitmajumder Appointment of Successor! \n\...	positiv
...	...	...	...	...	...	...	...	.
25683	25683	0.115	0.885	0.000	-0.5267	shv_indian	#1DDrive #ModiSpeaksToNews18 #LokSabhaElection...	negitiv
25684	25684	0.262	0.694	0.044	-0.8509	SONUPINKOO	I am seriously worried about future of Indian ...	negitiv
25685	25685	0.134	0.866	0.000	-0.6696	SONUPINKOO	If @narendramodi is so Bad then Why does @INCI...	negitiv
25686	25686	0.000	1.000	0.000	0.0000	SONUPINKOO	On this 23rd March (Pakistan Day), I Urge all ...	Neutra
25687	25687	0.129	0.703	0.168	-0.3348	parthajit0101	@timesofindia #ArvindKejriwal, #AAP sirji pls ...	negitiv

25688 rows × 8 columns

In [30]: `rahul_vaders`

Out[30]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	L
0	0	0.131	0.816	0.054	-0.5106	Sunnysweet16	Wonder why no academic or journalist asks INC ...	neg
1	1	0.000	0.890	0.110	0.5707	drnitinchaube	Congrats for the change #australiavotes2019 an...	pos
2	2	0.000	1.000	0.000	0.0000	mrvivek07	Peopel Say "Govt Ne 70 Years Kya kiya".\nUnse ...	Ne
3	3	0.000	0.742	0.258	0.9160	JosephPravinP	@ajaymaken @RahulGandhi And as a final touch, ...	pos
4	4	0.081	0.887	0.032	-0.3252	VandanaMegastar	#LokSabhaElections2019 Anyone not having mass ...	neg
...	...	...	...	...	...	...	...	
14256	14256	0.000	0.834	0.166	0.7142	SunjayJK	@quizzicalguy In this #LokSabhaElections2019, ...	pos
14257	14257	0.056	0.840	0.104	0.2187	SunjayJK	@AnumaVidisha @RahulGandhi @ArvindKejriwal whe...	pos
14258	14258	0.046	0.891	0.063	0.1872	RAMANKAIRA	@sherryontopp Early morning #political move af...	pos
14259	14259	0.045	0.893	0.062	0.1872	RAMANKAIRA	@mayankgandhi04 @mallesh_2004 Early morning #p...	pos
14260	14260	0.046	0.891	0.063	0.1872	RAMANKAIRA	@DrKumarVishwas Early morning #political move ...	pos

14261 rows × 8 columns

In [31]:
```python
neutral_modi = modi_vaders[modi_vaders['compound']==0.0000]
remove_neutral_modi = modi_vaders['compound'].isin(neutral_modi['compound'])
modi_vaders.drop(modi_vaders[remove_neutral_modi].index,inplace=True)
print(neutral_modi.shape)
print(modi_vaders.shape)
```

```
(4494, 8)
(21194, 8)
```

In [32]: `modi_vaders`

Out[32]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Labe
0	0	0.000	0.741	0.259	0.8126	advosushildixit	@anjanaomkashyap I am seeing you as future #bj...	positiv
1	1	0.170	0.830	0.000	-0.6520	jiaeur	#LokSabhaElections2019 \n23rd May 2019 will re...	negitiv
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	#LokSabhaElections2019 \n23rd May 2019 will re...	negitiv
3	3	0.000	0.750	0.250	0.8402	TheNirbhay1	PM Modi creates a new record of being the only...	positiv
4	4	0.086	0.764	0.151	0.4312	ShakeChilli	@abhijitmajumder Appointment of Successor! \n\...	positiv
...	...	...	...	...	...	...	...	.
25682	25682	0.087	0.913	0.000	-0.2960	CrazySatire	No one calling that 'BJP is communal' in this ...	negitiv
25683	25683	0.115	0.885	0.000	-0.5267	shv_indian	#1DDrive #ModiSpeaksToNews18 #LokSabhaElection...	negitiv
25684	25684	0.262	0.694	0.044	-0.8509	SONUPINKOO	I am seriously worried about future of Indian ...	negitiv
25685	25685	0.134	0.866	0.000	-0.6696	SONUPINKOO	If @narendramodi is so Bad then Why does @INCI...	negitiv
25687	25687	0.129	0.703	0.168	-0.3348	parthajit0101	@timesofindia #ArvindKejriwal, #AAP sirji pls ...	negitiv

21194 rows × 8 columns

In [33]:
```python
neutral_rahul = rahul_vaders[rahul_vaders['compound']==0.0000]
remove_neutral_rahul = rahul_vaders['compound'].isin(neutral_rahul['compound']
rahul_vaders.drop(rahul_vaders[remove_neutral_rahul].index,inplace=True)
print(neutral_rahul.shape)
print(rahul_vaders.shape)
```

```
(2607, 8)
(11654, 8)
```

In [34]: `rahul_vaders`

Out[34]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	L
0	0	0.131	0.816	0.054	-0.5106	Sunnysweet16	Wonder why no academic or journalist asks INC ...	neg
1	1	0.000	0.890	0.110	0.5707	drnitinchaube	Congrats for the change #australiavotes2019 an...	pos
3	3	0.000	0.742	0.258	0.9160	JosephPravinP	@ajaymaken @RahulGandhi And as a final touch, ...	pos
4	4	0.081	0.887	0.032	-0.3252	VandanaMegastar	#LokSabhaElections2019 Anyone not having mass ...	neg
5	5	0.000	0.940	0.060	0.2023	RGspeak	@INCIndia should release a video of @RahulGand...	pos
...	...	...	...	...	...	...	...	
14256	14256	0.000	0.834	0.166	0.7142	SunjayJK	@quizzicalguy In this #LokSabhaElections2019, ...	pos
14257	14257	0.056	0.840	0.104	0.2187	SunjayJK	@AnumaVidisha @RahulGandhi @ArvindKejriwal whe...	pos
14258	14258	0.046	0.891	0.063	0.1872	RAMANKAIRA	@sherryontopp Early morning #political move af...	pos
14259	14259	0.045	0.893	0.062	0.1872	RAMANKAIRA	@mayankgandhi04 @mallesh_2004 Early morning #p...	pos
14260	14260	0.046	0.891	0.063	0.1872	RAMANKAIRA	@DrKumarVishwas Early morning #political move ...	pos

11654 rows × 8 columns
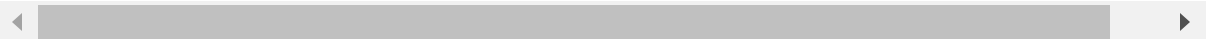
In [35]:
```python
print(modi_vaders.shape)
print(rahul_vaders.shape)
print(21194-10194)
print(11654-654)
```

```
(21194, 8)
(11654, 8)
11000
11000
```

In [36]:
```python
#modi
np.random.seed(10)
remove_n = 10194
drop_indices = np.random.choice(modi_vaders.index,remove_n,replace=False)
modi = modi_vaders.drop(drop_indices)
```

In [37]:
```python
modi.shape
```

Out[37]: (11000, 8)

In [38]:
```python
#rahul
np.random.seed(10)
remove_n = 654
drop_indices = np.random.choice(rahul_vaders.index,remove_n,replace=False)
rahul = rahul_vaders.drop(drop_indices)
```

In [39]:
```python
rahul.shape
```

Out[39]: (11000, 8)

In [40]:
```python
print(modi.shape)
print(rahul.shape)
```

```
(11000, 8)
(11000, 8)
```

In [41]:
```python
#  Counting Positive and Negative Sentiment
modi_count = modi.groupby('Label').count()
rahul_count = rahul.groupby('Label').count()
```

In [42]:
```python
modi_count
```

Out[42]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet
Label							
negitive	4042	4042	4042	4042	4042	4042	4042
positive	6958	6958	6958	6958	6958	6958	6958

In [43]:
```python
rahul_count
```

Out[43]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet
Label							
negitive	4496	4496	4496	4496	4496	4496	4496
positive	6504	6504	6504	6504	6504	6504	6504

In [44]:
```python
# Calculation negative and Positive Sentiment in modi_count
neg_modi =( modi_count['compound'][0]/1000)*100
pos_modi =( modi_count['compound'][1]/1000)*100
```

In [45]:
```python
# Calculation negative and Positive Sentiment in modi_count
neg_rahul =( rahul_count['compound'][0]/1000)*100
pos_rahul =( rahul_count['compound'][1]/1000)*100
```

In [46]:
```python
import plotly.graph_objects as go
import plotly.express as xp
```

```
In [47]: politicians = ['Modi','Rahul']

         neg_list = [neg_modi,neg_rahul]
         pos_list = [pos_modi,neg_modi]


         fig = go.Figure(
         data = [
             go.Bar(name='Negative',x=politicians,y=neg_list),
             go.Bar(name='Positive',x=politicians,y=pos_list)
         ]
         )
         fig.update_layout(barmode='group')
         fig.show()
```
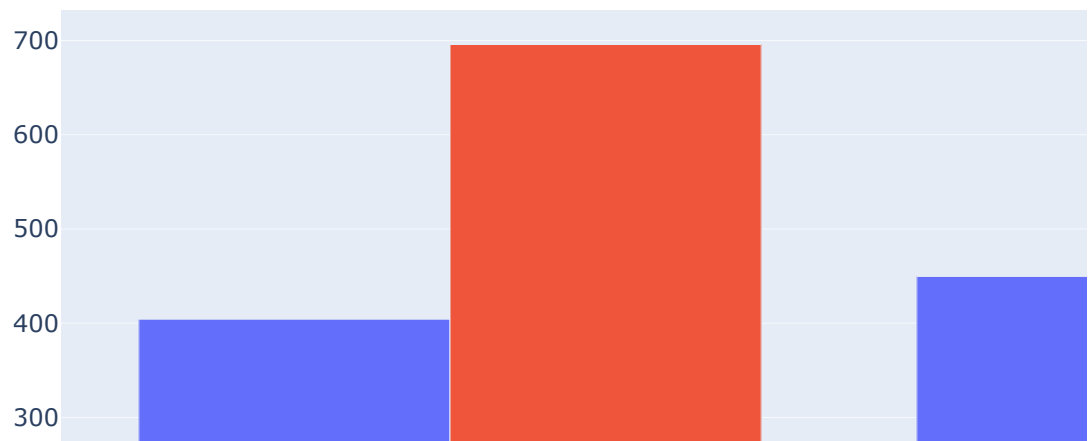

Analysis of Emotion

- modi dataframe
- rahul dataframe

In [48]: `modi.head()`

Out[48]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label
1	1	0.170	0.830	0.000	-0.6520	jiaeur	#LokSabhaElections2019 \n23rd May 2019 will re...	negitive
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	#LokSabhaElections2019 \n23rd May 2019 will re...	negitive
3	3	0.000	0.750	0.250	0.8402	TheNirbhay1	PM Modi creates a new record of being the only...	positive
4	4	0.086	0.764	0.151	0.4312	ShakeChilli	@abhijitmajumder Appointment of Successor! \n\...	positive
5	5	0.000	0.870	0.130	0.2023	UttarrPradesh	Which of the following should be top priority ...	positive

In [51]: `modi.shape`

Out[51]: `(11000, 8)`

In [60]: `modi.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11000 entries, 1 to 25684
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  11000 non-null  int64
 1   neg         11000 non-null  float64
 2   neu         11000 non-null  float64
 3   pos         11000 non-null  float64
 4   compound    11000 non-null  float64
 5   User        11000 non-null  object
 6   Tweet       11000 non-null  object
 7   Label       11000 non-null  object
dtypes: float64(4), int64(1), object(3)
memory usage: 773.4+ KB
```

In [56]: `rahul.head()`

Out[56]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label
0	0	0.131	0.816	0.054	-0.5106	Sunnysweet16	Wonder why no academic or journalist asks INC ...	negitive
1	1	0.000	0.890	0.110	0.5707	drnitinchaube	Congrats for the change #australiavotes2019 an...	positive
3	3	0.000	0.742	0.258	0.9160	JosephPravinP	@ajaymaken @RahulGandhi And as a final touch, ...	positive
4	4	0.081	0.887	0.032	-0.3252	VandanaMegastar	#LokSabhaElections2019 Anyone not having mass ...	negitive
5	5	0.000	0.940	0.060	0.2023	RGspeak	@INCIndia should release a video of @RahulGand...	positive

In [58]: `rahul.shape`

Out[58]: `(11000, 8)`

In [59]: `rahul.info`

Out[59]:
```
<bound method DataFrame.info of         Unnamed: 0    neg    neu    pos  compo
und         User  \
0                0  0.131  0.816  0.054  -0.5106    Sunnysweet16
1                1  0.000  0.890  0.110   0.5707    drnitinchaube
3                3  0.000  0.742  0.258   0.9160    JosephPravinP
4                4  0.081  0.887  0.032  -0.3252  VandanaMegastar
5                5  0.000  0.940  0.060   0.2023          RGspeak
...            ...    ...    ...    ...      ...              ...
14256        14256  0.000  0.834  0.166   0.7142          SunjayJK
14257        14257  0.056  0.840  0.104   0.2187          SunjayJK
14258        14258  0.046  0.891  0.063   0.1872        RAMANKAIRA
14259        14259  0.045  0.893  0.062   0.1872        RAMANKAIRA
14260        14260  0.046  0.891  0.063   0.1872        RAMANKAIRA

                                                    Tweet      Label
0      Wonder why no academic or journalist asks INC ...   negitive
1      Congrats for the change #australiavotes2019 an...   positive
3      @ajaymaken @RahulGandhi And as a final touch, ...   positive
4      #LokSabhaElections2019 Anyone not having mass ...   negitive
5      @INCIndia should release a video of @RahulGand...   positive
...                                                  ...        ...
14256  @quizzicalguy In this #LokSabhaElections2019, ...   positive
14257  @AnumaVidisha @RahulGandhi @ArvindKejriwal whe...   positive
14258  @sherryontopp Early morning #political move af...   positive
14259  @mayankgandhi04 @mallesh_2004 Early morning #p...   positive
14260  @DrKumarVishwas Early morning #political move ...   positive

[11000 rows x 8 columns]>
```

change the data type of tweet from object to str

In [61]:
```python
modi['Tweet'] = modi['Tweet'].astype(str)
modi.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11000 entries, 1 to 25684
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  11000 non-null  int64
 1   neg         11000 non-null  float64
 2   neu         11000 non-null  float64
 3   pos         11000 non-null  float64
 4   compound    11000 non-null  float64
 5   User        11000 non-null  object
 6   Tweet       11000 non-null  object
 7   Label       11000 non-null  object
dtypes: float64(4), int64(1), object(3)
memory usage: 773.4+ KB
```

In [62]:
```python
rahul['Tweet'] = rahul['Tweet'].astype(str)
rahul.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11000 entries, 0 to 14260
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  11000 non-null  int64
 1   neg         11000 non-null  float64
 2   neu         11000 non-null  float64
 3   pos         11000 non-null  float64
 4   compound    11000 non-null  float64
 5   User        11000 non-null  object
 6   Tweet       11000 non-null  object
 7   Label       11000 non-null  object
dtypes: float64(4), int64(1), object(3)
memory usage: 773.4+ KB
```

Lower Case the Tweet

In [63]:
```python
modi['Tweet'] = modi['Tweet'].str.lower()
modi.head()
```

Out[63]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label
1	1	0.170	0.830	0.000	-0.6520	jiaeur	#loksabhaelections2019 \n23rd may 2019 will re...	negitive
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	#loksabhaelections2019 \n23rd may 2019 will re...	negitive
3	3	0.000	0.750	0.250	0.8402	TheNirbhay1	pm modi creates a new record of being the only...	positive
4	4	0.086	0.764	0.151	0.4312	ShakeChilli	@abhijitmajumder appointment of successor! \n\...	positive
5	5	0.000	0.870	0.130	0.2023	UttarrPradesh	which of the following should be top priority ...	positive

In [64]:
```python
rahul['Tweet'] = rahul['Tweet'].str.lower()
rahul.head()
```

Out[64]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label
0	0	0.131	0.816	0.054	-0.5106	Sunnysweet16	wonder why no academic or journalist asks inc ...	negitive
1	1	0.000	0.890	0.110	0.5707	drnitinchaube	congrats for the change #australiavotes2019 an...	positive
3	3	0.000	0.742	0.258	0.9160	JosephPravinP	@ajaymaken @rahulgandhi and as a final touch, ...	positive
4	4	0.081	0.887	0.032	-0.3252	VandanaMegastar	#loksabhaelections2019 anyone not having mass ...	negitive
5	5	0.000	0.940	0.060	0.2023	RGspeak	@incindia should release a video of @rahulgand...	positive

Remove Punctutaion from Tweet

In [65]:
```python
import string
print(string.punctuation)
```

```
!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
```

In [66]:
```python
#  cleaned the string woth punctions (Three Parameter)
# str1 = specifies the list of charcater that needed to be reaplaced
# str2 = specifies the list of charcater with which  the character needed to b
# str1 = specifies the list of charcater that needed to be deleted

modi['Tweet'] = modi['Tweet'].apply(lambda x: x.translate(str.maketrans('',''
modi.head()
```

Out[66]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label
1	1	0.170	0.830	0.000	-0.6520	jiaeur	loksabhaelections2019 \n23rd may 2019 will rev...	negitive
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	loksabhaelections2019 \n23rd may 2019 will rev...	negitive
3	3	0.000	0.750	0.250	0.8402	TheNirbhay1	pm modi creates a new record of being the only...	positive
4	4	0.086	0.764	0.151	0.4312	ShakeChilli	abhijitmajumder appointment of successor \n\ng...	positive
5	5	0.000	0.870	0.130	0.2023	UttarrPradesh	which of the following should be top priority ...	positive

In [67]:
```python
rahul['Tweet'] = rahul['Tweet'].apply(lambda x: x.translate(str.maketrans('','
rahul.head()
```

Out[67]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label
0	0	0.131	0.816	0.054	-0.5106	Sunnysweet16	wonder why no academic or journalist asks inc ...	negitive
1	1	0.000	0.890	0.110	0.5707	drnitinchaube	congrats for the change australiavotes2019 and...	positive
3	3	0.000	0.742	0.258	0.9160	JosephPravinP	ajaymaken rahulgandhi and as a final touch mod...	positive
4	4	0.081	0.887	0.032	-0.3252	VandanaMegastar	loksabhaelections2019 anyone not having mass b...	negitive
5	5	0.000	0.940	0.060	0.2023	RGspeak	incindia should release a video of rahulgandhi...	positive

tokenising Tweet

In [68]:
```python
def textToToken(text):
    return nltk.word_tokenize(text);



modi['Tweet_Tokens'] = modi['Tweet'].apply(textToToken)
modi.head()
```

Out[68]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label	
1	1	0.170	0.830	0.000	-0.6520	jiaeur	loksabhaelections2019 \n23rd may 2019 will rev...	negitive	[loks 23r
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	loksabhaelections2019 \n23rd may 2019 will rev...	negitive	[loks 23r
3	3	0.000	0.750	0.250	0.8402	TheNirbhay1	pm modi creates a new record of being the only...	positive	[p new
4	4	0.086	0.764	0.151	0.4312	ShakeChilli	abhijitmajumder appointment of successor \n\ng...	positive	
5	5	0.000	0.870	0.130	0.2023	UttarrPradesh	which of the following should be top priority ...	positive	f

In [69]:
```python
rahul['Tweet_Tokens'] = rahul['Tweet'].apply(textToToken)
rahul.head()
```

Out[69]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label	
0	0	0.131	0.816	0.054	-0.5106	Sunnysweet16	wonder why no academic or journalist asks inc ...	negitive	a
1	1	0.000	0.890	0.110	0.5707	drnitinchaube	congrats for the change australiavotes2019 and...	positive	
3	3	0.000	0.742	0.258	0.9160	JosephPravinP	ajaymaken rahulgandhi and as a final touch mod...	positive	r
4	4	0.081	0.887	0.032	-0.3252	VandanaMegastar	loksabhaelections2019 anyone not having mass b...	negitive	[l
5	5	0.000	0.940	0.060	0.2023	RGspeak	incindia should release a video of rahulgandhi...	positive	

Removing Stop Words in the modi and rahul dataframe in Tweet_Tokens column

```
In [70]:  def stopToWords(tokenized_words):
              final_words = []
              for word in tokenized_words:
                  if word not in stopwords.words('english'):
                      final_words.append(word)
              return final_words


          modi['Tweet_Tokens'] = modi['Tweet_Tokens'].apply(stopToWords)
          modi.head()
```

Out[70]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label	
1	1	0.170	0.830	0.000	-0.6520	jiaeur	loksabhaelections2019 \n23rd may 2019 will rev...	negitive	[loks
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	loksabhaelections2019 \n23rd may 2019 will rev...	negitive	[loks
3	3	0.000	0.750	0.250	0.8402	TheNirbhay1	pm modi creates a new record of being the only...	positive	
4	4	0.086	0.764	0.151	0.4312	ShakeChilli	abhijitmajumder appointment of successor \n\ng...	positive	
5	5	0.000	0.870	0.130	0.2023	UttarrPradesh	which of the following should be top priority ...	positive	[fc m

In [71]:
```python
rahul['Tweet_Tokens'] = rahul['Tweet_Tokens'].apply(stopToWords)
rahul.head()
```

Out[71]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label	
0	0	0.131	0.816	0.054	-0.5106	Sunnysweet16	wonder why no academic or journalist asks inc ...	negitive	
1	1	0.000	0.890	0.110	0.5707	drnitinchaube	congrats for the change australiavotes2019 and...	positive	
3	3	0.000	0.742	0.258	0.9160	JosephPravinP	ajaymaken rahulgandhi and as a final touch mod...	positive	
4	4	0.081	0.887	0.032	-0.3252	VandanaMegastar	loksabhaelections2019 anyone not having mass b...	negitive	[l...
5	5	0.000	0.940	0.060	0.2023	RGspeak	incindia should release a video of rahulgandhi...	positive	[i...

Adding EmotionWords Column Showing the Different Types of Emotions

```
In [72]: def emotionAndWords(text):
             word_list =[]
             emotion_list = []
             with open('./emotion','r') as file:
                 for line in file:
                     clear_line = line.replace(',','')\
                     .replace('\n','')\
                     .replace("'",'')\
                     .strip()
                     word,emotion = clear_line.split(':')
         #           print("Words : "+word+" Emotion:"+emotion)

                     if word in text:
                         word_list.append(word)
                         emotion_list.append(emotion)
             return emotion_list


         modi["EmotionWords"] = modi["Tweet_Tokens"].apply(emotionAndWords)
         modi.head()
```

Out[72]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label	
1	1	0.170	0.830	0.000	-0.6520	jiaeur	loksabhaelections2019 \n23rd may 2019 will rev...	negitive	[loks
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	loksabhaelections2019 \n23rd may 2019 will rev...	negitive	[loks
3	3	0.000	0.750	0.250	0.8402	TheNirbhay1	pm modi creates a new record of being the only...	positive	
4	4	0.086	0.764	0.151	0.4312	ShakeChilli	abhijitmajumder appointment of successor \n\ng...	positive	
5	5	0.000	0.870	0.130	0.2023	UttarrPradesh	which of the following should be top priority ...	positive	[fc m

In [84]:
```python
rahul["EmotionWords"] = rahul["Tweet_Tokens"].apply(emotionAndWords)
rahul.head()
```

Out[84]:

	Unnamed: 0	neg	neu	pos	compound	User	Tweet	Label	
0	0	0.131	0.816	0.054	-0.5106	Sunnysweet16	wonder why no academic or journalist asks inc ...	negitive	
1	1	0.000	0.890	0.110	0.5707	drnitinchaube	congrats for the change australiavotes2019 and...	positive	
3	3	0.000	0.742	0.258	0.9160	JosephPravinP	ajaymaken rahulgandhi and as a final touch mod...	positive	
4	4	0.081	0.887	0.032	-0.3252	VandanaMegastar	loksabhaelections2019 anyone not having mass b...	negitive	[l
5	5	0.000	0.940	0.060	0.2023	RGspeak	incindia should release a video of rahulgandhi...	positive	[i

Joining the EmotionWords list of Lists into a Single List

or

Flatten the EmotionWords List of Lists

In [102]:
```python
emotion_list_modi = modi['EmotionWords'].tolist()
print(emotion_list_modi)
```

```
[[], [], [], [' sad', ' sad'], [], [], [], [], [], [], [], [], [], [], [],
[], [' apathetic'], [' sad'], [], [], [], [], [], [], [], [' independe
nt'], [' happy'], [], [], [' adequate'], [], [' independent'], [], [], [],
[], [], [], [], [], [' happy'], [], [' attracted', ' fearful'], [], [],
[], [], [], [], [], [], [' happy'], [' alone'], [], [], [], [], [], [' sa
d'], [], [], [' lost'], [], [], [], [], [], [], [], [], [' focused', '
independent'], [], [], [], [], [], [' alone'], [], [], [], [' angry', ' fe
arful'], [], [' fearless', ' adequate', ' fearful', ' powerless'], [' atta
ched', ' loved'], [], [], [], [], [], [], [], [], [' happy'], [], [], [],
[], [], [], [], [], [], [], [], [], [], [], [' happy'], [], [], ['
happy'], [], [' adequate'], [], [], [], [], [], [], [], [], [], [' sad', '
sad'], [' independent'], [' fearful', ' fearful'], [], [], [' powerless'],
[' powerless'], [' attached', ' loved'], [' happy'], [], [], [], [], [],
[], [' fearful'], [' free'], [], [' independent'], [' fearful'], [' hate
d'], [' anxious'], [], [' anxious'], [' powerless'], [], [' free'], [],
[], [], [], [], [' angry'], [' happy'], [], [], [], [], [' powerless'],
[], [], [], [], [], [], [], [], [], [], [], [' adequate', ' love
d'], [' cheated'], [], [], [], [], [' average', ' happy'], [' cheated', '
fearless'], [' alone'], [], [], [], [' free'], [], [], [], [' cheated'],
```

In [103]:
```python
final_emotion_list_modi = []
for llist in emotion_list_modi:
    final_emotion_list_modi = final_emotion_list_modi + llist

print(final_emotion_list_modi)
```

```
ppy ,   happy ,    rree ,   attractea ,   surprise ,   coaepenaent ,   powe
rless', ' fearful', ' average', ' happy', ' surprise', ' happy', ' powerle
ss', ' happy', ' happy', ' sad', ' angry', ' happy', ' happy', ' happy', '
happy', ' attracted', ' independent', ' sad', ' sad', ' cheated', ' free',
' sad', ' sad', ' fearful', ' free', ' powerless', ' esteemed', ' happy',
' lost', ' esteemed', ' esteemed', ' happy', ' independent', ' lost', ' av
erage', ' free', ' happy', ' lost', ' happy', ' lost', ' esteemed', ' este
emed', ' happy', ' average', ' happy', ' fearless', ' happy', ' happy', '
esteemed', ' sad', ' fearful', ' sad', ' sad', ' esteemed', ' powerless',
' lost', ' happy', ' sad', ' adequate', ' entitled', ' attracted', ' fre
e', ' independent', ' fearful', ' powerless', ' fearful', ' singled out',
' loved', ' powerless', ' cheated', ' happy', ' fearless', ' fearless', '
happy', ' sad', ' happy', ' lost', ' free', ' cheated', ' happy', ' happ
y', ' sad', ' cheated', ' happy', ' independent', ' powerless', ' happy',
' alone', ' happy', ' esteemed', ' esteemed', ' sad', ' happy', ' powerles
s', ' attracted', ' derailed', ' fearless', ' cheated', ' happy', ' cheate
d', ' cheated', ' ecstatic', ' belittled', ' cheated', ' lost', ' fearles
s', ' fearful', ' focused', ' sad', ' esteemed', ' adequate', ' happy', '
attracted', ' happy', ' angry', ' fearful', ' independent', ' independen
t', ' independent', ' cheated', ' sad', ' independent', ' happy', ' powerl
```

In [104]:
```python
emotion_list_rahul = rahul['EmotionWords'].tolist()
print(emotion_list_rahul)
```

```
[], [], [], [], [], [], [], [], [  independent ], [], [], [  happy ], [],
[], [], [], [], [' fearless'], [' esteemed'], [], [' happy'], [], [' i
ndependent'], [' anxious', ' fearful'], [], [' loved'], [], [], [], [], ['
happy'], [], [], [], [], [], [' fearful'], [], [], [], [' angry'], [], ['
happy'], [], [], [], [], [], [' powerless'], [], [], [], [], [' fearful'],
[], [' happy'], [' adequate'], [], [], [], [], [], [], [' happy'], [], [],
[], [], [' adequate'], [], [], [' happy'], [], [], [], [], [], [], [],
[' average', ' free'], [], [], [], [], [' powerless'], [], [], [' focuse
d'], [], [], [], [], [], [], [], [], [' esteemed'], [], [], [], [], [], ['
attached'], [], [], [], [], [], [], [], [], [], [], [], [], [], [],
[], [' attracted'], [], [], [], [], [], [], [], [], [' sad', ' sad'],
[], [], [], [], [' lost', ' independent', ' fearful'], [], [], [], [' 
lost'], [' fearless'], [], [], [' happy'], [], [], [], [], [], [], [], [],
[], [' powerless'], [], [], [], [], [' sad', ' surprise', ' sad'], [], [],
[], [' average'], [], [' lost'], [], [], [], [], [], [], [], [], [], [' sa
d', ' sad'], [' attached'], [], [], [], [], [' anxious'], [], [], [' alon
e', ' hated'], [], [], [], [], [], [], [' sad'], [], [' cheated'], [], [],
[], [' happy'], [], [], [], [' angry'], [], [' esteemed'], [], [], [], [],
[], [], [], [], [], [], [], [' happy'], [], [' surprise', ' surpris
e'], [], [], [], [' happy'], [], [], [], [], [' sad', ' sad'], [], [' chea
ted'   'happy'], [], [' attracted'], [], [' fearless'], [], [], [' fearfu
```

In [105]:
```python
final_emotion_list_rahul = []
for llist in emotion_list_rahul:
    final_emotion_list_rahul = final_emotion_list_rahul + llist

print(final_emotion_list_rahul)
```

[' happy', ' sad', ' happy', ' happy', ' independent', ' sad', ' sad', ' l
ost', ' average', ' attached', ' loved', ' independent', ' attracted', ' f
earful', ' fearless', ' happy', ' happy', ' sad', ' derailed', ' happy', '
powerless', ' esteemed', ' independent', ' entitled', ' happy', ' happy',
' happy', ' fearless', ' adequate', ' fearful', ' powerless', ' happy', '
happy', ' attached', ' happy', ' loved', ' independent', ' lost', ' attach
ed', ' happy', ' cheated', ' attracted', ' powerless', ' loved', ' happy',
' free', ' esteemed', ' happy', ' adequate', ' sad', ' sad', ' fearless',
' fearful', ' fearful', ' lost', ' cheated', ' fearless', ' attached', ' e
steemed', ' loved', ' cheated', ' free', ' lost', ' focused', ' embarrasse
d', ' attracted', ' angry', ' surprise', ' focused', ' happy', ' happy', '
happy', ' lost', ' adequate', ' surprise', ' safe', ' happy', ' loved', '
happy', ' sad', ' surprise', ' sad', ' singled out', ' hated', ' sad', ' s
ingled out', ' hated', ' sad', ' attached', ' surprise', ' independent', '
happy', ' fearless', ' esteemed', ' happy', ' independent', ' anxious', '
fearful', ' loved', ' happy', ' fearful', ' angry', ' happy', ' powerles
s', ' fearful', ' happy', ' adequate', ' happy', ' adequate', ' happy', '
average', ' free', ' powerless', ' focused', ' esteemed', ' attached', ' a
ttracted', ' sad', ' sad', ' lost', ' independent', ' fearful', ' lost', '

Counting Emotion Using Counter

In [106]:
```python
emotion_list_modi_counter = Counter(final_emotion_list_modi)
print(emotion_list_modi_counter)
```

Counter({' happy': 819, ' sad': 342, ' independent': 226, ' fearful': 216, '
free': 172, ' attracted': 157, ' powerless': 156, ' lost': 154, ' adequate':
148, ' esteemed': 137, ' fearless': 116, ' average': 96, ' angry': 90, ' chea
ted': 74, ' attached': 67, ' surprise': 67, ' alone': 64, ' hated': 59, ' lov
ed': 55, ' anxious': 37, ' singled out': 32, ' focused': 27, ' embarrassed':
27, ' ecstatic': 21, ' safe': 16, ' demoralized': 16, ' codependent': 14, ' e
ntitled': 14, ' obsessed': 10, ' belittled': 7, ' bored': 5, ' apathetic': 4,
' lustful': 3, ' derailed': 2, ' burdened': 1})

In [107]:
```python
emotion_list_rahul_counter = Counter(final_emotion_list_rahul)
print(emotion_list_rahul_counter)
```

Counter({' happy': 519, ' sad': 297, ' fearful': 230, ' lost': 160, ' adequat
e': 142, ' powerless': 140, ' free': 136, ' attracted': 133, ' independent':
103, ' average': 78, ' esteemed': 73, ' fearless': 70, ' loved': 69, ' cheate
d': 66, ' angry': 63, ' surprise': 61, ' attached': 57, ' alone': 53, ' hate
d': 44, ' anxious': 43, ' focused': 42, ' embarrassed': 33, ' singled out': 2
8, ' entitled': 22, ' safe': 21, ' demoralized': 16, ' ecstatic': 13, ' obses
sed': 12, ' apathetic': 10, ' belittled': 9, ' codependent': 8, ' bored': 6,
' derailed': 4, ' lustful': 3})

Ploting Graph

```
In [111]:  # plt.bar(emotion_counter.keys() , emotion_counter.values() ,color="purple")

           fig ,ax1 = plt.subplots()
           ax1.bar(emotion_list_modi_counter.keys() , emotion_list_modi_counter.values()
           ax1.bar(emotion_list_rahul_counter.keys() , emotion_list_rahul_counter.values(
           fig.autofmt_xdate()
           plt.xticks(rotation=90)
           plt.ylabel("Score of Emotions")
           plt.legend()
           plt.show()
```