In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import nltk
```

In [2]:
```
nltk.download('punkt')
```

```
[nltk_data] Error loading punkt: <urlopen error [Errno 11001]
[nltk_data]     getaddrinfo failed>
```

Out[2]:
```
False
```

Modi DataSet

In [3]:
```
modi_df =pd.read_csv('./modi_reviews.csv')
```

In [4]:
```
modi_df.shape
```

Out[4]:
```
(25688, 3)
```

In [5]:
```
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
```

In [6]:
```
modi_df.head()
```

Out[6]:

	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 [7]:
```
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 [8]:
```
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']
```

In [9]:
```
nltk.download('vader_lexicon')
```

```
[nltk_data] Error loading vader_lexicon: <urlopen error [Errno 11001]
[nltk_data]     getaddrinfo failed>
```

Out[9]:
```
False
```

Rahul Dataset

In [10]:
```
rahul_df = pd.read_csv('./rahul_reviews.csv')
```

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

Out[11]:
```
(14261, 3)
```

In [12]:
```
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 [13]:
```
rahul_df.head()
```

Out[13]:

	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 [14]:
```
from nltk.sentiment import SentimentIntensityAnalyzer
from tqdm.notebook import tqdm

sia = SentimentIntensityAnalyzer()
```

In [15]:
```
sia.polarity_scores('I am so happy!')
```

Out[15]:
```
{'neg': 0.0, 'neu': 0.318, 'pos': 0.682, 'compound': 0.6468}
```

In [16]:
```
sia.polarity_scores('This is the worst thing ever.')
```

Out[16]:
```
{'neg': 0.451, 'neu': 0.549, 'pos': 0.0, 'compound': -0.6249}
```

In [17]:
```
sia.polarity_scores(modi_df['Tweet'][0])
```

Out[17]:
```
{'neg': 0.0, 'neu': 0.741, 'pos': 0.259, 'compound': 0.8126}
```

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

Out[18]:
```
{'neg': 0.131, 'neu': 0.816, 'pos': 0.054, 'compound': -0.5106}
```

Run the polarity score on the entire modi dataset

In [19]:
```
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 [00:13<00:00, 2173.76it/s]

Run the polarity score on the entire rahul dataset

In [20]:
```
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 [00:05<00:00, 2403.45it/s]
```

In [21]:
```
print(res1)
```

```
0, 'pos': 0.0, 'compound': 0.0}, 89: {'neg': 0.028, 'neu': 0.972, 'pos':
0.0, 'compound': -0.0516}, 90: {'neg': 0.0, 'neu': 0.909, 'pos': 0.091, 'c
ompound': 0.3818}, 91: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.
0}, 92: {'neg': 0.0, 'neu': 0.523, 'pos': 0.477, 'compound': 0.8074}, 93:
{'neg': 0.0, 'neu': 0.878, 'pos': 0.122, 'compound': 0.7488}, 94: {'neg':
0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 95: {'neg': 0.0, 'neu': 0.8
87, 'pos': 0.113, 'compound': 0.5719}, 96: {'neg': 0.0, 'neu': 1.0, 'pos':
0.0, 'compound': 0.0}, 97: {'neg': 0.036, 'neu': 0.857, 'pos': 0.108, 'com
pound': 0.5574}, 98: {'neg': 0.053, 'neu': 0.786, 'pos': 0.161, 'compoun
d': 0.7269}, 99: {'neg': 0.15, 'neu': 0.633, 'pos': 0.217, 'compound': 0.
0}, 100: {'neg': 0.239, 'neu': 0.662, 'pos': 0.099, 'compound': -0.7916},
101: {'neg': 0.119, 'neu': 0.881, 'pos': 0.0, 'compound': -0.6124}, 102:
{'neg': 0.0, 'neu': 0.797, 'pos': 0.203, 'compound': 0.6249}, 103: {'neg':
0.0, 'neu': 0.847, 'pos': 0.153, 'compound': 0.5859}, 104: {'neg': 0.178,
'neu': 0.696, 'pos': 0.126, 'compound': -0.4588}, 105: {'neg': 0.1, 'neu':
0.833, 'pos': 0.066, 'compound': -0.34}, 106: {'neg': 0.0, 'neu': 0.727,
'pos': 0.273, 'compound': 0.9517}, 107: {'neg': 0.078, 'neu': 0.615, 'po
s': 0.307, 'compound': 0.941}, 108: {'neg': 0.0, 'neu': 0.701, 'pos': 0.29
9, 'compound': 0.9256}, 109: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compoun
d': 0.0}, 110: {'neg': 0.0, 'neu': 0.846, 'pos': 0.154, 'compound': 0.636
```

In [22]:
```
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 [23]:
```
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 [24]:
```
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 [25]:
```
# Now we have sentiment score and metadata
modi_vaders.head()
```

Out[25]:

	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 [26]:
```
rahul_vaders.head()
```

Out[26]:

	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 [27]:
```
# 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_14396\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_14396\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 [28]:
```
modi_vaders
```

Out[28]:

	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...	negativ
2	2	0.170	0.830	0.000	-0.6520	PVenkatGandhi	#LokSabhaElections2019 \n23rd May 2019 will re...	negativ
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...	negativ
25684	25684	0.262	0.694	0.044	-0.8509	SONUPINKOO	I am seriously worried about future of Indian ...	negativ
25685	25685	0.134	0.866	0.000	-0.6696	SONUPINKOO	If @narendramodi is so Bad then Why does @INCI...	negativ
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 ...	negativ

25688 rows × 8 columns

In [29]:
```
rahul_vaders
```

Out[29]:

	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 [30]:
```
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 [31]:
```
modi_vaders
```

Out[31]:

	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 [32]:
```
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 [33]:
```
rahul_vaders
```

Out[33]:

	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 [34]:
```
print(modi_vaders.shape)
print(rahul_vaders.shape)
print(21194-10194)
print(11654-654)
```

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

In [35]:
```
#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 [36]:
```
modi.shape
```

Out[36]: (11000, 8)

In [37]:
```
#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 [38]:
```
rahul.shape
```

Out[38]: (11000, 8)

In [39]:
```
print(modi.shape)
print(rahul.shape)
```

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

In [40]:
```
#  Counting Positive and Negative Sentiment
modi_count = modi.groupby('Label').count()
rahul_count = rahul.groupby('Label').count()
```

In [41]:
```
modi_count
```

Out[41]:

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

In [42]:
```
rahul_count
```

Out[42]:

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

In [43]:
```
# 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 [44]:
```
# 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 [45]:
```
import plotly.graph_objects as go
import plotly.express as xp
```

In [46]:
```
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()
```