Sentiment Analysis

Import Libraries

```
import pandas as pd
import matplotlib.pyplot as plt
import nltk, re
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from wordcloud import WordCloud
```

Read the Dataset

```
df= pd.read_csv("/content/recession_2023_india_tweets_dataset.csv")
df.head()
```

	date	id	content	username	likeCount	retwee
0	2022-11-30 04:17:49+00:00	1597807084934172672	📌 @elonmusk begins hiring for Twitter 2.0 afte...	workflexitweets	1	
			Global			

Data Preprocessing

```
df=df.drop(['date', 'id', 'username', 'likeCount', 'retweetCount'],axis=1)
```

```
df
```

	content
0	📌 @elonmusk begins hiring for Twitter 2.0 afte...
1	Global Recession से दुनिया हुई बेहाल, लेकिन In...
2	Will global recession impact India or not? #re...
3	Is Recession In India Coming Soon?\n\nWant to ...
4	A report by #realestate data collation & r...
...	...
5171	India beats recession blues, back on FDI radar...
5172	One lakh indian will return home to india in n...
5173	Puneet doubts that any economy (BRIC or not) i...
5174	Wow! So many guys who did their MS in US are b...
5175	Indians abandoning their cars at #Dubai airpor...

5176 rows × 1 columns

```
df.isnull().sum()
```

```
content    0
dtype: int64
```

```
def clean(text):
  text=str(text).lower()
  text=re.sub(r'https?:\/\/\S+','',text)
  text=re.sub(r'[^a-zA-Z\s]','',text)
  return text

df['content']=df['content'].apply(clean)
df
```

	content
0	elonmusk begins hiring for twitter after fir...
1	global recession india \nglobal...
2	will global recession impact india or not rece...
3	is recession in india coming soon\n\nwant to k...
4	a report by realestate data collation amp rese...
...	...
5171	india beats recession blues back on fdi radar ...
5172	one lakh indian will return home to india in n...
5173	puneet doubts that any economy bric or not is ...
5174	wow so many guys who did their ms in us are ba...

5176 rows × 1 columns

Install NLTK Library

```
pip install nltk
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk) (8.1.3)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk) (2022.10.31)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk) (4.65.0)
```

```
nltk.download('vader_lexicon')
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
True
```

Classifying Tweets

```
for index, row in df['content'].iteritems():
  score= SentimentIntensityAnalyzer().polarity_scores(row)
  neg=score['neg']
  neu=score['neu']
  pos=score['pos']
  if neg>pos:
    df.loc[index, 'Sentiment'] = 'Negative'
  elif pos>neg:
    df.loc[index, 'Sentiment'] = 'Positive'
  else:
    df.loc[index, 'Sentiment'] = 'Neutral'
```

```
df
```

```
<ipython-input-15-f6d5f818053d>:1: FutureWarning: iteritems is deprecated and will be removed in a futu
  for index, row in df['content'].iteritems():
```

	content	Sentiment
0	elonmusk begins hiring for twitter after fir...	Negative
1	global recession india \nglobal...	Negative
2	will global recession impact india or not rece...	Negative
3	is recession in india coming soon\n\nwant to k...	Negative
4	a report by realestate data collation amp rese...	Negative
...	...	...
5171	india beats recession blues back on fdi radar ...	Negative
5172	one lakh indian will return home to india in n...	Negative
5173	puneet doubts that any economy bric or not is ...	Negative
5174	wow so many guys who did their ms in us are ba...	Positive
5175	indians abandoning their cars at dubai airport...	Negative

5176 rows × 2 columns

Frequency of Classes

```
def count_tp_in_column(data,feature):
  total = data.loc[:,feature].value_counts(dropna=False)
  percentage = data.loc[:,feature].value_counts(dropna=False, normalize=True)*100
  return pd.concat([total,round(percentage,2)],axis=1,keys=['Total','Percentage'])

tp=count_tp_in_column(df,'Sentiment')
tp
```

	Total	Percentage
Negative	4286	82.81
Positive	802	15.49
Neutral	88	1.70

Data Visualization

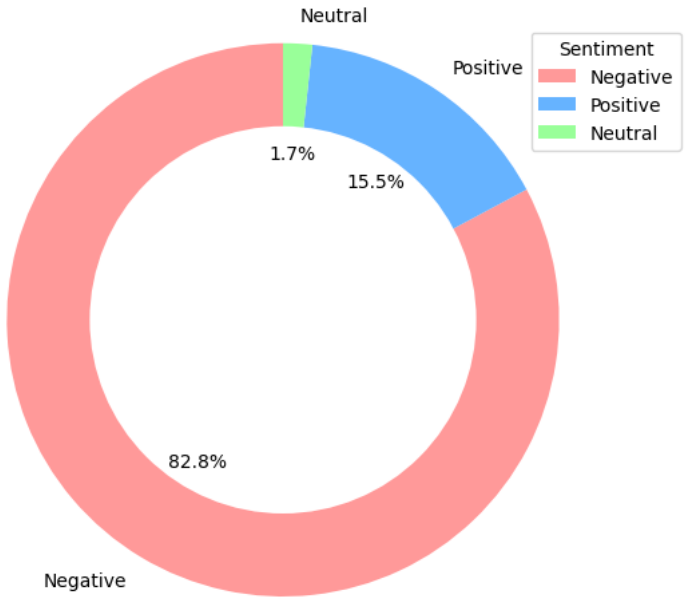
```
labels = tp.index
sizes = tp['Percentage']
colors= [ '#ff9999', '#66b3ff', '#99ff99']

fig1, ax1 = plt.subplots()
ax1.pie(sizes, colors=colors, labels=labels, autopct='%1.1f%%', startangle=90)

centre_circle = plt.Circle((0,0),0.70,fc='white')
fig=plt.gcf()
fig.gca().add_artist(centre_circle)

ax1.axis('equal')
plt.tight_layout()
plt.legend(title='Sentiment')
plt
```

```
<module 'matplotlib.pyplot' from '/usr/local/lib/python3.10/dist-packages/matplotlib/pyplot.py'>
```

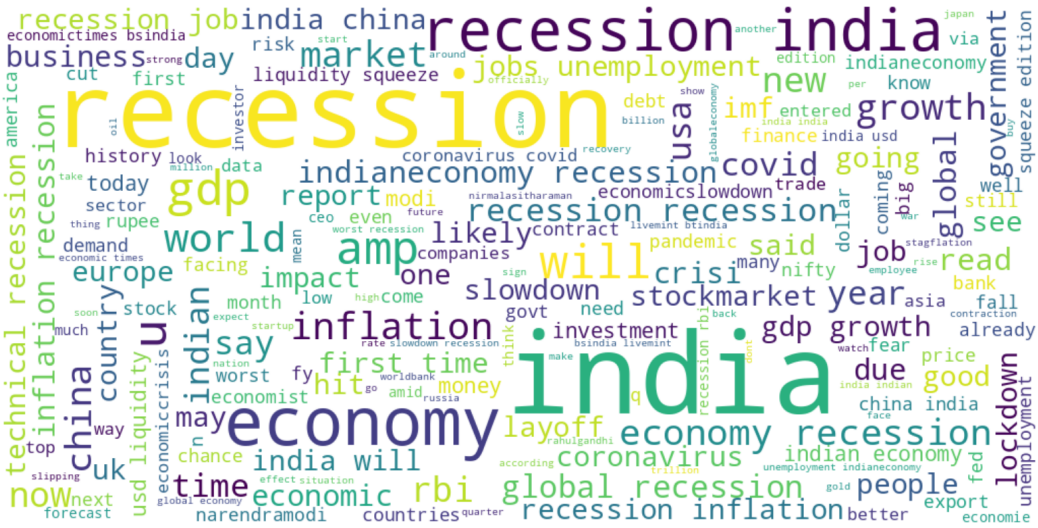



```
all_tweets = " ".join(tweet for tweet in df['content'])

wordcloud = WordCloud(width=1024, height=512, random_state=21, max_font_size=110, background_color='white').generate(all_tweets)

plt.figure(figsize=(20,10))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis('off')
plt
```

```
<module 'matplotlib.pyplot' from '/usr/local/lib/python3.10/dist-packages/matplotlib/pyplot.py'>
```




```
all_positive_tweets = " ".join(tweet for tweet in df[df['Sentiment'] == 'Positive']['content'])

wordcloud=WordCloud(width=1024, height=512, random_state=21, max_font_size=110, background_color='white').generate(all_positive_tweets)

plt.figure(figsize=(20,10))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis('off')
plt
```

```
<module 'matplotlib.pyplot' from '/usr/local/lib/python3.10/dist-packages/matplotlib/pyplot.py'>
```


```
all_negative_tweets=" ".join(tweet for tweet in df[df['Sentiment'] == 'Negative']['content'])

wordcloud=WordCloud(width=1024, height=512, random_state=21, max_font_size=110, background_color='white').generate(all_negative_tweets)
```

```
plt.figure(figsize=(20,10))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis('off')
plt
```

```
<module 'matplotlib.pyplot' from '/usr/local/lib/python3.10/dist-packages/matplotlib/pyplot.py'>
```