

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
%matplotlib inline

import re
import nltk
from nltk.corpus import stopwords
from wordcloud import WordCloud, STOPWORDS
import warnings
warnings.filterwarnings('ignore')
```

```
In [2]: raw_data = pd.read_csv("Elon_musk.csv", encoding='cp1252').drop('Unnamed: 0', axis=1)
df = raw_data.copy()
df
```

Out[2]:

	Text
0	@kunalb11 I'm an alien
1	@ID_AA_Carmack Ray tracing on Cyberpunk with H...
2	@joerogan @Spotify Great interview!
3	@gtera27 Doge is underestimated
4	@teslacn Congratulations Tesla China for amazi...
...	...
1994	@flcnhvy True, it sounds so surreal, but the n...
1995	@PPathole Make sure to read ur terms & con...
1996	@TeslaGong @PPathole Samwise Gamgee
1997	@PPathole Altho Dumb and Dumber is <U+0001F525...
1998	Progress update August 28

1999 rows × 1 columns

```
In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1999 entries, 0 to 1998
Data columns (total 1 columns):
#   Column  Non-Null Count  Dtype
---  -
0    Text    1999 non-null      object
dtypes: object(1)
memory usage: 15.7+ KB
```

```
In [4]: df.Text = df.Text.astype(str)
```

Preprocessing data

```
In [5]: # Clean The Data using RegEx
#result = re.sub(pattern, repl, string, count=0, flags=0)

def cleantext(text):
    text = re.sub(r"@[A-Za-z0-9]+", "", text) # Remove Mentions
    text = re.sub(r"#", "", text) # Remove Hashtags Symbol
    text = re.sub(r"RT[\s]+", "", text) # Remove Retweets
    text = re.sub(r"https?:\/\/\/\S+", "", text) # Remove The Hyper L
    text = re.sub(r"_[A-Za-z0-9]", "", text) # Removing underscores

    return text
# Clean The Text
df["Text"] = df["Text"].apply(cleantext)
df.head()
```

Out [5]:

	Text
0	I'm an alien
1	Aarmack Ray tracing on Cyberpunk with HDR is n...
2	Great interview!
3	Doge is underestimated
4	Congratulations Tesla China for amazing execu...

```
In [6]: df.tail()
```

Out [6]:

	Text
1994	True, it sounds so surreal, but the negative ...
1995	Make sure to read ur terms & conditions b...
1996	Samwise Gamgee
1997	Altho Dumb and Dumber is <U+0001F525><U+0001F...
1998	Progress update August 28

Tweet Sentiment Values

```
In [11]: nltk.download('vader_lexicon')
```

```
[nltk_data] Downloading package vader_lexicon to  
[nltk_data] /Users/nehachavan/nltk_data...
```

```
Out[11]: True
```

```
In [12]: from nltk.sentiment.vader import SentimentIntensityAnalyzer  
sentialz = SentimentIntensityAnalyzer()
```

```
In [13]: sentialz.polarity_scores(df.Text.iloc[2])
```

```
Out[13]: {'neg': 0.0, 'neu': 0.185, 'pos': 0.815, 'compound': 0.6588}
```

```
In [14]: #Function to retrieve sentiment class  
def get_senti_class(review):  
    analysis = sentialz.polarity_scores(review)  
    # set sentiment  
    if analysis["compound"] > 0:  
        return 'positive'  
    elif analysis["compound"] < 0:  
        return 'negative'  
    else:  
        return 'neutral'  
  
def get_sentiment(review):  
    return pd.Series([sentialz.polarity_scores(review), get_senti_cl
```

```
In [15]: #Applying function to all tweets
df[["Scores", "Sentiment"]] = df.Text.apply(get_sentiment)
df
```

Out [15]:

		Text	Scores	Sentiment
0		I'm an alien	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...}	neutral
1		Aarmack Ray tracing on Cyberpunk with HDR is n...	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...}	neutral
2		Great interview!	{'neg': 0.0, 'neu': 0.185, 'pos': 0.815, 'comp...}	positive
3		Doge is underestimated	{'neg': 0.512, 'neu': 0.488, 'pos': 0.0, 'comp...}	negative
4		Congratulations Tesla China for amazing execu...	{'neg': 0.0, 'neu': 0.628, 'pos': 0.372, 'comp...}	positive
...	
1994		True, it sounds so surreal, but the negative ...	{'neg': 0.249, 'neu': 0.561, 'pos': 0.19, 'com...}	negative
1995		Make sure to read ur terms & conditions b...	{'neg': 0.0, 'neu': 0.634, 'pos': 0.366, 'comp...}	positive
1996		Samwise Gamgee	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...}	neutral
1997		Altho Dumb and Dumber is <U+0001F525> <U+0001F...	{'neg': 0.592, 'neu': 0.408, 'pos': 0.0, 'comp...}	negative
1998		Progress update August 28	{'neg': 0.0, 'neu': 0.517, 'pos': 0.483, 'comp...}	positive

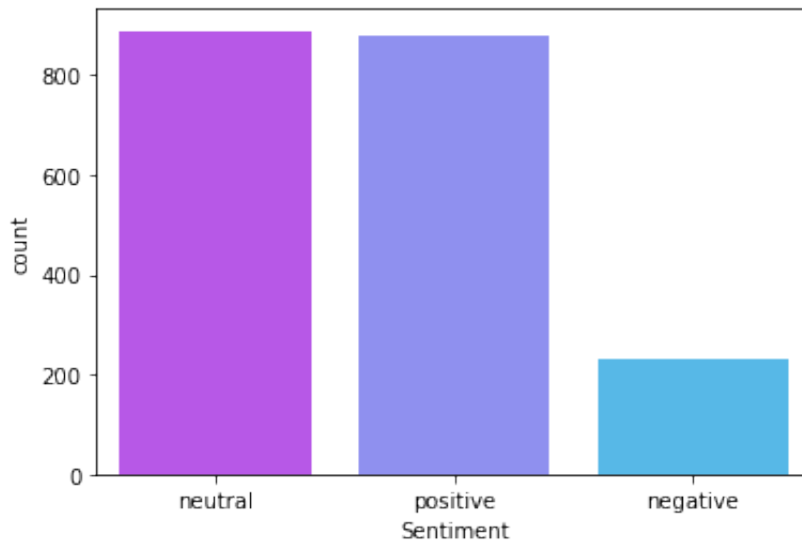
1999 rows × 3 columns

```
In [16]: comp_values = [d.get('compound') for d in df.Scores]
df.Sentiment.value_counts()
```

```
Out [16]: neutral      888
positive    880
negative    231
Name: Sentiment, dtype: int64
```

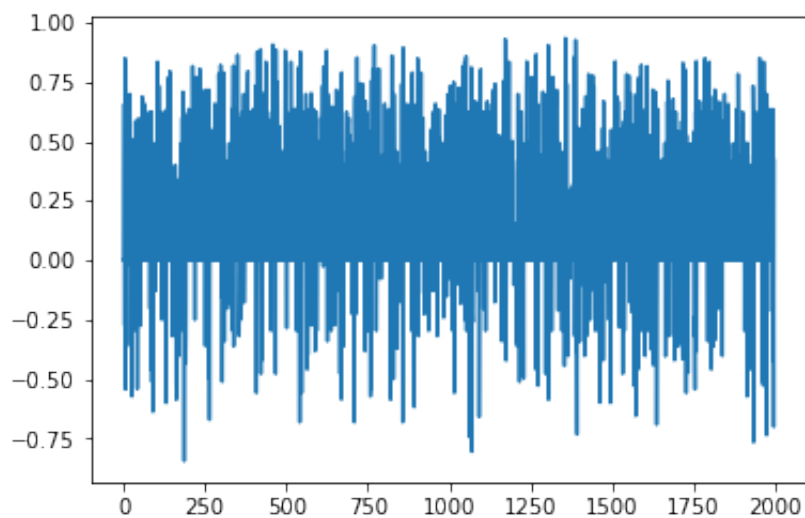
```
In [17]: sns.countplot(df.Sentiment,palette='cool_r') #There are very less n
```

```
Out[17]: <AxesSubplot:xlabel='Sentiment', ylabel='count'>
```



```
In [18]: sns.lineplot(data = comp_values, palette='cool_r') #Most are positi
```

```
Out[18]: <AxesSubplot:>
```



In [19]: `df.loc[df.Sentiment=='negative']`

Out[19]:

	Text	Scores	Sentiment
3	Doge is underestimated	{'neg': 0.512, 'neu': 0.488, 'pos': 0.0, 'comp...	negative
6	Frodo was the underdoge,\nAll thought he would...	{'neg': 0.226, 'neu': 0.774, 'pos': 0.0, 'comp...	negative
15	He definitely has issues, but the sentencing ...	{'neg': 0.162, 'neu': 0.681, 'pos': 0.157, 'co...	negative
20	Any crypto wallet that won't give you your pr...	{'neg': 0.138, 'neu': 0.862, 'pos': 0.0, 'comp...	negative
21	Your app sucks	{'neg': 0.556, 'neu': 0.444, 'pos': 0.0, 'comp...	negative
...
1965	That is the near-term danger of AI	{'neg': 0.362, 'neu': 0.638, 'pos': 0.0, 'comp...	negative
1975	Death is the loss of information	{'neg': 0.608, 'neu': 0.392, 'pos': 0.0, 'comp...	negative
1991	No easy way to answer this in a tweet, but he...	{'neg': 0.205, 'neu': 0.622, 'pos': 0.174, 'co...	negative
1994	True, it sounds so surreal, but the negative ...	{'neg': 0.249, 'neu': 0.561, 'pos': 0.19, 'com...	negative
1997	Altho Dumb and Dumber is <U+0001F525> <U+0001F...	{'neg': 0.592, 'neu': 0.408, 'pos': 0.0, 'comp...	negative

231 rows × 3 columns

In []: