

In [36]:

```
import numpy as np # Linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import string # special operations on strings
import spacy # Language models

from matplotlib.pyplot import imread
from matplotlib import pyplot as plt
from wordcloud import WordCloud
```

In [37]:

```
elon_musk=pd.read_csv('Elon_musk.csv',error_bad_lines =False ,encoding='latin-1')
elon_musk
```

Out[37]:

Unnamed: 0		Text
0	1	@kunalb11 I m an alien
1	2	@ID_AA_Carmack Ray tracing on Cyberpunk with H...
2	3	@joerogan @Spotify Great interview!
3	4	@gtera27 Doge is underestimated
4	5	@teslacn Congratulations Tesla China for amazi...
...	...	...
1994	1995	@flcnhvy True, it sounds so surreal, but the n...
1995	1996	@PPathole Make sure to read ur terms & con...
1996	1997	@TeslaGong @PPathole Samwise Gamgee
1997	1998	@PPathole Altho Dumb and Dumber is <U+0001F525...
1998	1999	Progress update August 28

1999 rows × 2 columns

In [38]:

```
elon_musk.Text
```

Out[38]:

```
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
Name: Text, Length: 1999, dtype: object
```

## Data preprocessing

In [39]:

```
elon_musk = [x.strip() for x in elon_musk.Text] # remove both the leading and the trailing
elon_musk = [x for x in elon_musk if x] # removes empty strings, because they are considere
elon_musk[0:10]
```

Out[39]:

```
['@kunalb11 I\x92m an alien',
 '@ID_AA_Carmack Ray tracing on Cyberpunk with HDR is next-level. Have you t
ried it?',
 '@joerogan @Spotify Great interview!',
 '@gtera27 Doge is underestimated',
 '@teslacn Congratulations Tesla China for amazing execution last year. Now
on to the next for even more!!',
 'Happy New Year of the Ox! https://t.co/9WFKMYu2oj', (https://t.co/9WFKMYu2oj'),
 'Frodo was the underdoge,\nAll thought he would fail,\nHimself most of all.
https://t.co/zGxJFDzzrM', (https://t.co/zGxJFDzzrM'),
 '@OwenSparks_ @flcnhvy @anonyx10 Haha thanks :)',
 '@flcnhvy @anonyx10 Indeed! Tweets definitely do not represent real-world t
ime allocation.',
 'The most entertaining outcome is the most likely']
```

In [40]:

```
from nltk import tokenize
sentences = tokenize.sent_tokenize(" ".join(elon_musk))
sentences[0:5]
```

Out[40]:

```
['@kunalb11 I\x92m an alien @ID_AA_Carmack Ray tracing on Cyberpunk with HDR
is next-level.',
 'Have you tried it?',
 '@joerogan @Spotify Great interview!',
 '@gtera27 Doge is underestimated @teslacn Congratulations Tesla China for a
mazing execution last year.',
 'Now on to the next for even more!!']
```

In [41]:

```
punc=string.punctuation
punc
```

Out[41]:

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

In [42]:

```
for i in elon_musk:
    if i in punc:
        elon_musk = punc.replace(i, " ")
```

In [45]:

```
elon_musk=str(elon_musk)
elon_musk.replace('@', '')
```

Out[45]:

```
'[\'kunalb11 I\\x92m an alien\\', \'ID_AA_Carmack Ray tracing on Cyberpunk
with HDR is next-level. Have you tried it?\\', \'joerogan Spotify Great int
erview!\\', \'gtera27 Doge is underestimated\\', \'teslacn Congratulations T
esla China for amazing execution last year. Now on to the next for even mo
re!!\\', \'Happy New Year of the Ox! https://t.co/9WFKMYu2oj\\', (https://t.
co/9WFKMYu2oj\\',) \'Frodo was the underdoge,\\nAll thought he would fai
l,\\nHimself most of all. https://t.co/zGxJFDzrM\\', (https://t.co/zGxJFDz
rM\\',) \'OwenSparks_ flcnhvy anonyx10 Haha thanks :)\\', \'flcnhvy anonyx1
0 Indeed! Tweets definitely do not represent real-world time allocatio
n.\\', \'The most entertaining outcome is the most likely\\', \'GiveDirectly
Just sent some\\', \'Just agree to do Clubhouse with kanyewest\\', \'http
s://t.co/3rWE9uHSTS\\', \'geoffkeighley UnrealEngine It\\x92s getting real
\\', \'Bought some Dogecoin for lil X, so he can be a toddler hodler\\', \'J
oshManMode He definitely has issues, but the sentencing seems a bit high
\\', \'freewalletorg Thanks for fixing\\', \'freewalletorg Please unlock my
account\\', \'AstroJordy <U+0001F923><U+0001F923>\\', \'This is true power
haha https://t.co/Fc9uhQSd70\\', (https://t.co/Fc9uhQSd70\\',) \'freewallet
org Any crvnto wallet that won\\x92t give you your private kevs should be
```

In [20]:

```
import re
elon_musk =re.sub('<[^>]+>', ' ', elon_musk)
```

In [21]:

```
sent_df = pd.DataFrame(sentences, columns=['sentence'])
sent_df
```

Out[21]:

	sentence
0	@kunalb11 I m an alien @ID_AA_Carmack Ray trac...
1	Have you tried it?
2	@joerogan @Spotify Great interview!
3	@gtera27 Doge is underestimated @teslacn Congr...
4	Now on to the next for even more!!
...	...
919	@kenyanwalstreet Not actually a payout, just a...
920	It may never pay out, as the stock can t b ht...
921	Details Aug 28.
922	AI symbiosis while u wait @vistacruiser7 @flcn...
923	@TeslaGong @PPathole Samwise Gamgee @PPathole ...

924 rows × 1 columns

## Sentiment analysis

In [22]:

```
afinn = pd.read_csv('Afinn.csv', sep=',', encoding='latin-1')
afinn.shape
```

Out[22]:

(2477, 2)

In [47]:

```
affinity_scores = afinn.set_index('word')['value'].to_dict()
affinity_scores
```

Out[47]:

```
{'abandon': -2,
 'abandoned': -2,
 'abandons': -2,
 'abducted': -2,
 'abduction': -2,
 'abductions': -2,
 'abhor': -3,
 'abhorred': -3,
 'abhorrent': -3,
 'abhors': -3,
 'abilities': 2,
 'ability': 2,
 'aboard': 1,
 'absentee': -1,
 'absentees': -1,
 'absolve': 2,
 'absolved': 2,
 'absolves': 2.
```

In [24]:

```
nlp = spacy.load("en_core_web_sm")
sentiment_lexicon = affinity_scores

def calculate_sentiment(text: str = None):
    sent_score = 0
    if text:
        sentence = nlp(text)
        for word in sentence:
            sent_score += sentiment_lexicon.get(word.lemma_, 0)
    return sent_score
```

In [25]:

```
calculate_sentiment(text = 'amazing')
```

Out[25]:

4

In [26]:

```
sent_df['sentiment_value'] = sent_df['sentence'].apply(calculate_sentiment)
```

In [27]:

```
sent_df['word_count'] = sent_df['sentence'].str.split().apply(len)
sent_df['word_count'].head(10)
```

Out[27]:

```
0      13
1       4
2       4
3      13
4       8
5       6
6      14
7      10
8       8
9     109
```

Name: word\_count, dtype: int64

In [28]:

```
sent_df.sort_values(by='sentiment_value').tail(10)
```

Out[28]:

	sentence	sentiment_value	word_count
817	Might be able to <a href="https://t.co/FB7e5nOq2E">https://t.co/FB7e5nOq2E</a> @Pri...	11	72
258	@TheOnion @ID_AA_Carmack Yup @flcnhvy Change o...	11	57
224	@PPathole @karpathy For a second there, I thou...	11	69
719	We just haven't observed the <a href="https://t.co/mez...">https://t.co/mez...</a>	11	72
184	Also, w <a href="https://t.co/IYdKtVsqq">https://t.co/IYdKtVsqq</a> @cleantechnic...	12	98
81	@teslaownersSV This is a good one @MrBeastYT I...	13	38
645	It s <a href="https://t.co/lxhyT6NuiC">https://t.co/lxhyT6NuiC</a> @Teslarati Tesla...	13	65
585	The open areas <a href="https://t.co/rabjKrtQlw">https://t.co/rabjKrtQlw</a> @Sav...	14	138
36	@ajtourville @Erdyastronaut @SpaceX Yes, but ...	15	231
105	@Erdyastronaut @SpaceX Was also thinking that...	16	94

In [29]:

```
sent_df['sentiment_value'].describe()
```

Out[29]:

```
count      924.000000
mean        1.392857
std         2.684249
min        -8.000000
25%         0.000000
50%         0.000000
75%         3.000000
max        16.000000
Name: sentiment_value, dtype: float64
```

In [30]:

```
sent_df[sent_df['sentiment_value']<=0].head()
```

Out[30]:

	sentence	sentiment_value	word_count
0	@kunalb11 I m an alien @ID_AA_Carmack Ray trac...	0	13
1	Have you tried it?	0	4
4	Now on to the next for even more!!	0	8
5	Happy New Year of the Ox!	0	6
6	https://t.co/9WFKMYu2oj Frodo was the underdog...	-2	14

In [31]:

```
sent_df[sent_df['sentiment_value']>=10].head()
```

Out[31]:

	sentence	sentiment_value	word_count
36	@ajtourville @Erdyastronaut @SpaceX Yes, but ...	15	231
81	@teslaownersSV This is a good one @MrBeastYT I...	13	38
105	@Erdyastronaut @SpaceX Was also thinking that...	16	94
151	@GUNMTLM3 @lexfridman No, but definitely this ...	10	45
159	@harsimranbansal @heydave7 It will most likely...	10	93

In [32]:

```
sent_df['index']=range(0,len(sent_df))
sent_df
```

Out[32]:

	sentence	sentiment_value	word_count	index
0	@kunalb11 I m an alien @ID_AA_Carmack Ray trac...	0	13	0
1	Have you tried it?	0	4	1
2	@joerogan @Spotify Great interview!	3	4	2
3	@gtera27 Doge is underestimated @teslacn Congr...	3	13	3
4	Now on to the next for even more!!	0	8	4
...	...	...	...	...
919	@kenyanwalstreet Not actually a payout, just a...	0	11	919
920	It may never pay out, as the stock can t b ht...	-4	31	920
921	Details Aug 28.	0	3	921
922	AI symbiosis while u wait @vistacruiser7 @flcn...	-2	47	922
923	@TeslaGong @PPathole Samwise Gamgee @PPathole ...	0	15	923

924 rows × 4 columns

In [33]:

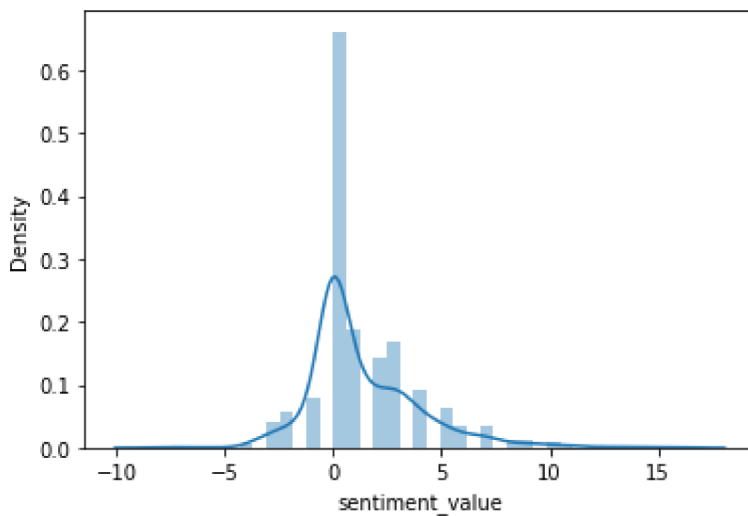
```
import seaborn as sns
import matplotlib.pyplot as plt
sns.distplot(sent_df['sentiment_value'])
```

C:\Users\ROOBA\anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)

Out[33]:

<AxesSubplot:xlabel='sentiment\_value', ylabel='Density'>

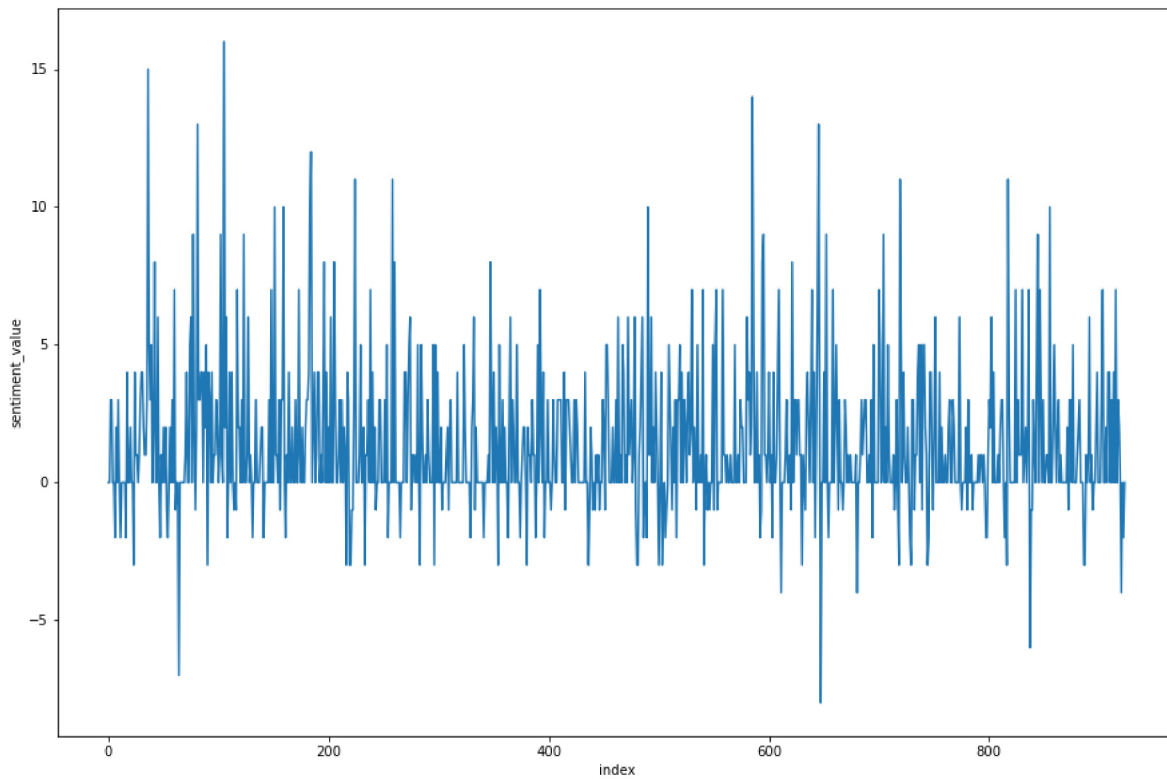


In [34]:

```
plt.figure(figsize=(15, 10))  
sns.lineplot(y='sentiment_value',x='index',data=sent_df)
```

Out[34]:

<AxesSubplot:xlabel='index', ylabel='sentiment\_value'>





In [35]:

```
sent_df.plot.scatter(x='word_count', y='sentiment_value', figsize=(8,8), title='Sentence se
```

Out[35]:

```
<AxesSubplot:title={'center': 'Sentence sentiment value to sentence word coun  
t'}, xlabel='word_count', ylabel='sentiment_value'>
```

