Assignment-11-Text Mining-02-Amazon-Product-Reviews

Named Entity Recognition

In [1]:

```
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
%matplotlib inline
```

In [2]:

```
import pandas
book=pd.read_csv("C:/Users/LENOVO/Documents/Custom Office Templates/Xiaomi11_Lite.txt",erro
book
```

```
aw 3\nSkipping line 119: expected 1 fields, saw 3\nSkipping line 120: expe
cted 1 fields, saw 2\nSkipping line 121: expected 1 fields, saw 3\nSkippin
g line 127: expected 1 fields, saw 2\nSkipping line 135: expected 1 field
s, saw 3\nSkipping line 137: expected 1 fields, saw 3\nSkipping line 139:
expected 1 fields, saw 10\nSkipping line 146: expected 1 fields, saw 5\nSk
ipping line 151: expected 1 fields, saw 2\nSkipping line 155: expected 1 f
ields, saw 2\nSkipping line 157: expected 1 fields, saw 2\nSkipping line 1
59: expected 1 fields, saw 2\nSkipping line 161: expected 1 fields, saw 4
\nSkipping line 163: expected 1 fields, saw 4\nSkipping line 167: expected
1 fields, saw 3\nSkipping line 169: expected 1 fields, saw 2\nSkipping lin

e 173: expected 1 fields, saw 2\nSkipping line 175: expected 1 fields, saw
3\nSkipping line 177: expected 1 fields, saw 2\n'
```

Out[2]:

	Xiaomi
0	As per this price range it is absolute patheti...
1	Camera quality is just like 10000 to 15000 rs ...

In [3]:

```
book = [Xiaomi.strip() for Xiaomi in book.Xiaomi]   # remove both the leading and the trail
book = [Xiaomi for Xiaomi in book if Xiaomi ]        # removes empty strings, because they
book[0:10]
```

Out[3]:

```
['As per this price range it is absolute pathetic quality of front and rear
camera...after going through review video I had purchased this phone ...but
it is absolutely one of the biggest mistake.',
 'Camera quality is just like 10000 to 15000 rs range.it is totally not wort
h the money of 26000 rs.',
 'After a month I am writing phone is great mostly by weight performance.',
 "Granted most of the reviewers on YouTube were overtly heaping praises on t
his phone (mostly paid by Xiaomi to do so). Probably that's the reason many
of you were disappointed by this phone when you finally got it in your hand
s.",
 'Think of this phone as iPhone SE of the Android world and you will underst
and the true value for money of this phone.',
 'Storage becomes inaccessible after disabling work profile for all apps in
personal profile.',
 "Work profile button on notification bar doesn't light up or grey out.",
 'Disabling work profile disables application running in dual mode as wel
l.',
 'DnD does not disable notifications nor silence calls',
 'Pictures/screenshot will not be stored because storage is inaccessible']
```

In [4]:

```
##Part Of Speech Tagging
nlp = spacy.load('en_core_web_sm')

one_block = book[2]
doc_block = nlp(one_block)
spacy.displacy.render(doc_block, style='ent', jupyter=True)
```

```
C:\Users\LENOVO\anaconda3\lib\site-packages\spacy\displacy\__init__.py:200:
UserWarning: [W006] No entities to visualize found in Doc object. If this is
surprising to you, make sure the Doc was processed using a model that suppor
ts named entity recognition, and check the `doc.ents` property manually if n
ecessary.
  warnings.warn(Warnings.W006)
```

After a month I am writing phone is great mostly by weight performance.

In [5]:

```
one_block
```

Out[5]:

```
'After a month I am writing phone is great mostly by weight performance.'
```

In [6]:

```
for token in doc_block[0:20]:
    print(token, token.pos_)
```

```
After ADP
a DET
month NOUN
I PRON
am AUX
writing VERB
phone NOUN
is AUX
great ADJ
mostly ADV
by ADP
weight NOUN
performance NOUN
. PUNCT
```

In [7]:

```
#Filtering for nouns and verbs only
nouns_verbs = [token.text for token in doc_block if token.pos_ in ('NOUN', 'VERB')]
print(nouns_verbs[0:10])
```

```
['month', 'writing', 'phone', 'weight', 'performance']
```

In [8]:

```
#Counting tokens again
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer()

X = cv.fit_transform(nouns_verbs)
sum_words = X.sum(axis=0)
words_freq = [(word, sum_words[0, idx]) for word, idx in cv.vocabulary_.items()]
words_freq =sorted(words_freq, key = lambda x: x[1], reverse=True)
wf_df = pd.DataFrame(words_freq)
wf_df.columns = ['word', 'count']
```

In [9]:

```
wf_df[0:10]
```

Out[9]:

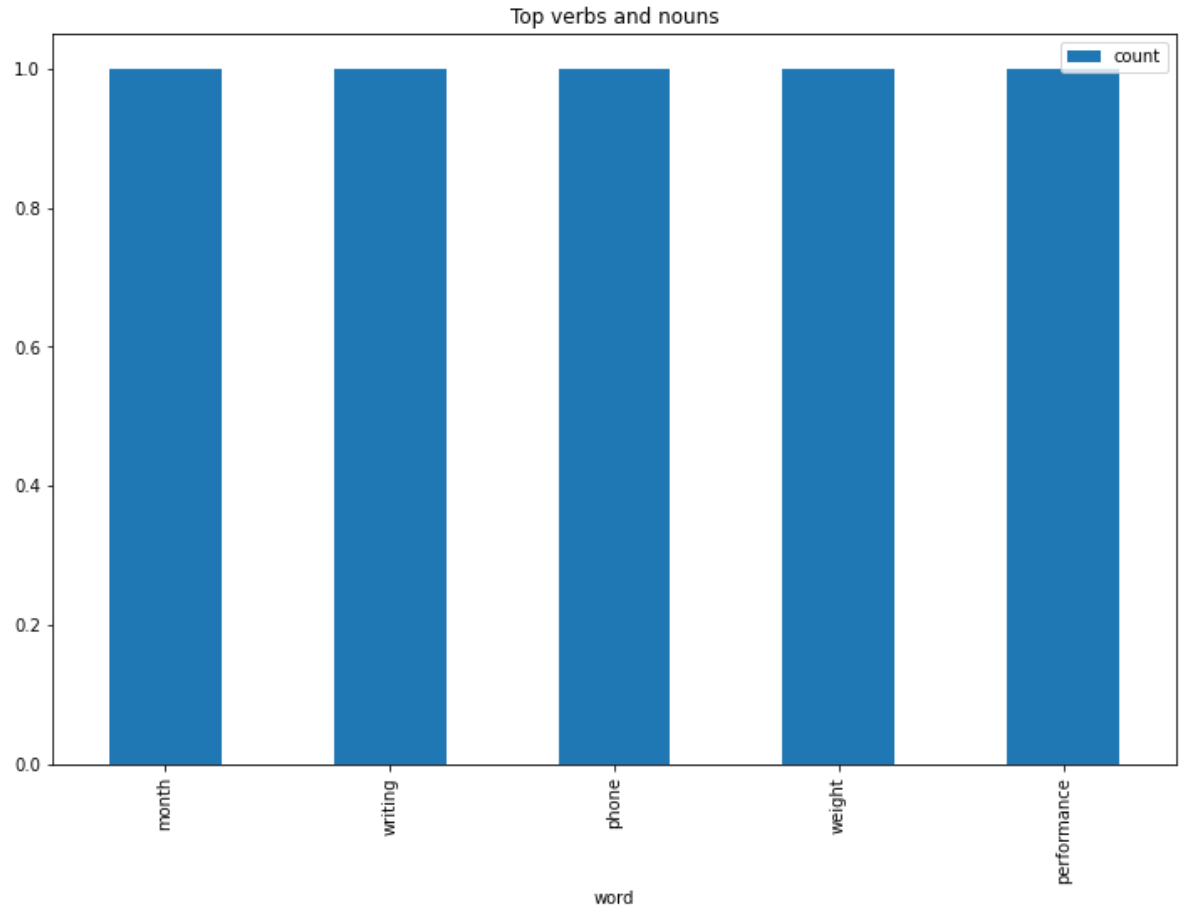
	word	count
0	month	1
1	writing	1
2	phone	1
3	weight	1
4	performance	1

In [10]:

```
##Visualizing results
#Barchart for top 10 nouns + verbs
wf_df[0:10].plot.bar(x='word', figsize=(12,8), title='Top verbs and nouns')
```

Out[10]:

```
<AxesSubplot:title={'center':'Top verbs and nouns'}, xlabel='word'>
```



Emotion Mining

In [11]:

```
#Sentiment analysis
afinn = pd.read_csv('C:/Users/LENOVO/Documents/Custom Office Templates/Afinn.csv', sep=',',
afinn.shape
```

Out[11]:

```
(2477, 2)
```

In [12]:

```
afinn.head()
```

Out[12]:

	word	value
0	abandon	-2
1	abandoned	-2
2	abandons	-2
3	abducted	-2
4	abduction	-2

In [13]:

```
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
%matplotlib inline

import pandas
book=pd.read_csv("C:/Users/LENOVO/Documents/Custom Office Templates/Xiaomi11_Lite.txt",erro
book = [Xiaomi.strip() for Xiaomi in book.Xiaomi] # remove both the leading and the trailin
book = [Xiaomi for Xiaomi in book if Xiaomi] # removes empty strings, because they are cons
```

```
b'Skipping line 7: expected 1 fields, saw 3\nSkipping line 8: expected 1 fie
lds, saw 2\nSkipping line 10: expected 1 fields, saw 2\nSkipping line 17: ex
pected 1 fields, saw 2\nSkipping line 21: expected 1 fields, saw 3\nSkipping
line 22: expected 1 fields, saw 3\nSkipping line 23: expected 1 fields, saw
4\nSkipping line 25: expected 1 fields, saw 2\nSkipping line 29: expected 1
fields, saw 5\nSkipping line 31: expected 1 fields, saw 2\nSkipping line 32:
expected 1 fields, saw 2\nSkipping line 34: expected 1 fields, saw 3\nSkippi
ng line 40: expected 1 fields, saw 2\nSkipping line 41: expected 1 fields, s
aw 2\nSkipping line 42: expected 1 fields, saw 2\nSkipping line 51: expected
1 fields, saw 2\nSkipping line 56: expected 1 fields, saw 4\nSkipping line 6
0: expected 1 fields, saw 3\nSkipping line 77: expected 1 fields, saw 2\nSki
pping line 78: expected 1 fields, saw 2\nSkipping line 80: expected 1 field
s, saw 2\nSkipping line 82: expected 1 fields, saw 2\nSkipping line 83: expe
cted 1 fields, saw 2\nSkipping line 89: expected 1 fields, saw 3\nSkipping l
ine 90: expected 1 fields, saw 2\nSkipping line 92: expected 1 fields, saw 6
\nSkipping line 95: expected 1 fields, saw 3\nSkipping line 96: expected 1 f
ields, saw 2\nSkipping line 98: expected 1 fields, saw 2\nSkipping line 101:
expected 1 fields, saw 2\nSkipping line 103: expected 1 fields, saw 2\nSkipp
ing line 105: expected 1 fields, saw 2\nSkipping line 106: expected 1 field
s, saw 2\nSkipping line 112: expected 1 fields, saw 3\nSkipping line 113: ex
pected 1 fields, saw 2\nSkipping line 115: expected 1 fields, saw 6\nSkippin
g line 118: expected 1 fields, saw 3\nSkipping line 119: expected 1 fields,
saw 3\nSkipping line 120: expected 1 fields, saw 2\nSkipping line 121: expec
ted 1 fields, saw 3\nSkipping line 127: expected 1 fields, saw 2\nSkipping l
ine 135: expected 1 fields, saw 3\nSkipping line 137: expected 1 fields, saw
3\nSkipping line 139: expected 1 fields, saw 10\nSkipping line 146: expected
1 fields, saw 5\nSkipping line 151: expected 1 fields, saw 2\nSkipping line
155: expected 1 fields, saw 2\nSkipping line 157: expected 1 fields, saw 2\n
Skipping line 159: expected 1 fields, saw 2\nSkipping line 161: expected 1 f
ields, saw 4\nSkipping line 163: expected 1 fields, saw 4\nSkipping line 16
7: expected 1 fields, saw 3\nSkipping line 169: expected 1 fields, saw 2\nSk
ipping line 173: expected 1 fields, saw 2\nSkipping line 175: expected 1 fie
lds, saw 3\nSkipping line 177: expected 1 fields, saw 2\n'
```

In [14]:

```
from nltk import tokenize
sentences = tokenize.sent_tokenize(" ".join(book))
sentences[5:15]
```

Out[14]:

```
['Think of this phone as iPhone SE of the Android world and you will underst
and the true value for money of this phone.',
 'Storage becomes inaccessible after disabling work profile for all apps in
personal profile.',
 "Work profile button on notification bar doesn't light up or grey out.",
 'Disabling work profile disables application running in dual mode as wel
l.',
 "DnD does not disable notifications nor silence calls Pictures/screenshot w
ill not be stored because storage is inaccessible The phone is good with ama
zing specs But the MIUI is very buggy eg many PDFs don't open from many apps
30 people found this helpful It's a mid premium smartphone and it's so light
and slim.",
 'The phone is powered by Qualcomm Snapdragon 778G which is good and you pre
tty decent performance for the price.',
 "It have a triple camera setup it's okish not so great not so bad.",
 'It have a decent camera.',
 'The telephoto macro camera is awesome and allows you to take great macro s
hots.',
 'The networking and WiFi performance is also good and also have carrier agg
regation.']
```

In [15]:

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

Out[15]:

	sentence
0	As per this price range it is absolute patheti...
1	Camera quality is just like 10000 to 15000 rs ...
2	After a month I am writing phone is great most...
3	Granted most of the reviewers on YouTube were ...
4	Probably that's the reason many of you were di...
...	...
82	Sometimes the photo looks over-sharpened espec...
83	And the photo has too much contrast and noise.
84	Sometimes The phone picks up the wrong colour.
85	Xiaomi needs to work on camera optimisation (f...
86	The front camera is OKAY.

87 rows × 1 columns

In [16]:

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

In [17]:

```
#Custom function :score each word in a sentence in lemmatised form,
#but calculate the score for the whole original sentence.
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 [18]:

```
# test that it works
calculate_sentiment(text = 'amazing')
```

Out[18]:

```
4
```

In [19]:

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

In [20]:

```
# how many words are in the sentence?
sent_df['word_count'] = sent_df['sentence'].str.split().apply(len)
sent_df['word_count'].head(10)
```

Out[20]:

```
0    33
1    19
2    13
3    21
4    20
5    23
6    13
7    12
8    11
9    54
Name: word_count, dtype: int64
```

In [21]:

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

Out[21]:

	sentence	sentiment_value	word_count
26	Design: lite weight and sleek design amazing l...	4	8
78	Stereo speakees are also awesome.	4	5
27	Got this phone on great indian sale with my ol...	5	16
19	Display: beautiful screen edge to edge display...	6	11
38	I have experienced hot body temperature only w...	8	49
13	The telephoto macro camera is awesome and allo...	8	14
9	DnD does not disable notifications nor silence...	9	54
17	21 people found this helpful Good Looking (Dia...	10	45
32	Pros Crisp and vibrant display Dolby vision su...	15	82
79	I have purchased this mobile in offer so basic...	16	157

In [22]:

```
# Sentiment score of the whole review
sent_df['sentiment_value'].describe()
```

Out[22]:

```
count    87.000000
mean      1.793103
std       3.267664
min      -3.000000
25%       0.000000
50%       0.000000
75%       3.000000
max      16.000000
Name: sentiment_value, dtype: float64
```

In [23]:

```
# negative sentiment score of the whole review
sent_df[sent_df['sentiment_value']<=0]
```

Out[23]:

	sentence	sentiment_value	word_count
0	As per this price range it is absolute patheti...	-3	33
4	Probably that's the reason many of you were di...	-2	20
5	Think of this phone as iPhone SE of the Androi...	0	23
6	Storage becomes inaccessible after disabling w...	0	13
7	Work profile button on notification bar doesn'...	-1	12
8	Disabling work profile disables application ru...	0	11
11	It have a triple camera setup it's okish not s...	0	14
12	It have a decent camera.	0	5
21	Camera: lots of feature in camera.	0	6
22	You can record 4k videos upto 30fps.	0	7
29	Camera is descent.. And it has got 35W fast ch...	0	10
30	If you are looking for light weight and slim m...	0	18
34	My phone is easily lasting more than one day w...	0	14
36	I see some users have raised heating issue whi...	-2	31
40	My phone is easily lasting more than one day w...	0	14
42	I see some users have raised heating issue whi...	-2	31
44	I have experienced hot body temperature only w...	-2	49
45	So my verdict for this phone would be is that ...	-3	36
46	I don't think my next phone would be a Xiaomi.	0	10
52	12 5G bands...	0	3
54	It stops responding.	-1	3
55	Even turned off suddenly.	0	4
56	You can not use sbi mobile banking yono app.	0	9
58	App categorization is not working.	0	5
59	All these issues are with in 3 days of puchase.	0	10
60	Now outlook app is not working on this phone.	0	9
61	It sometimes open emails sometimes don't.	0	6
62	Most of the time keyboard will not appear for ...	0	10
63	You need to reopen the app.	0	6
65	Light and fast.	0	3
66	Earlier i was using realme x2 pro with sd855+ ...	0	10
67	I find no difference in speed.	-1	6
68	All the features compressed into this small li...	0	9
71	Literally twice as better than my earlier phon...	0	11

	sentence	sentiment_value	word_count
72	Only con is camera.	0	4
74	So 4 stars only.	0	4
75	I must mention the very even and small bezels ...	0	13
76	Makes it a treat for the eyes.	0	7
77	Disign wise 10/10.	0	3
81	It is not consistent.	0	4
82	Sometimes the photo looks over-sharpened espec...	0	12
83	And the photo has too much contrast and noise.	0	9
84	Sometimes The phone picks up the wrong colour.	-2	8
85	Xiaomi needs to work on camera optimisation (f...	0	11
86	The front camera is OKAY.	0	5

In [24]:

```
# Positive sentiment score of the whole review
sent_df[sent_df['sentiment_value']>0]
```

Out[24]:

	sentence	sentiment_value	word_count
1	Camera quality is just like 10000 to 15000 rs ...	4	19
2	After a month I am writing phone is great most...	3	13
3	Granted most of the reviewers on YouTube were ...	3	21
9	DnD does not disable notifications nor silence...	9	54
10	The phone is powered by Qualcomm Snapdragon 77...	4	19
13	The telephoto macro camera is awesome and allo...	8	14
14	The networking and WiFi performance is also go...	3	13
15	The MIUI is surprisingly smooth will optimized...	3	50
16	The introductory price with Amazon offer was a...	3	11
17	21 people found this helpful Good Looking (Dia...	10	45
18	Sharing this review so that you can decide if ...	2	15
19	Display: beautiful screen edge to edge display...	6	11
20	4k videos are looking beautiful in this screen.	3	8
23	Good camera setup.	3	3
24	Battery: Nice battery backup depends on your u...	3	8
25	Fast charging is good.	3	4
26	Design: lite weight and sleek design amazing l...	4	8
27	Got this phone on great indian sale with my ol...	5	16
28	Build quality good...	3	3
31	Performance is good but less when compared snp...	3	10
32	Pros Crisp and vibrant display Dolby vision su...	15	82
33	Battery and charging speed are also good.	3	7
35	And if you are a fan boy of ultra superfast da...	3	37
37	It gets a little warm during prolonged usage(w...	1	12
38	I have experienced hot body temperature only w...	8	49
39	Battery and charging speed are also good.	3	7
41	And if you are a fan boy of ultra superfast da...	3	37
43	It gets a little warm during prolonged usage(w...	1	12
47	Processor is amazing at this price range.	4	7
48	Display is beautiful.	3	3
49	Battery life is good.	3	4
50	Sound is stereo and pretty good.	4	6
51	Camera is good for the price range.	3	7
53	Even samsung GALAXY 20 doesn't offer so many b...	2	20

	sentence	sentiment_value	word_count
57	Not sure how many more apps are not supported.	2	9
64	Looks like has memory management issues.	2	6
69	Impressed with battery and display.	3	5
70	Diaplay is amazing and so many colours to see.	4	9
73	Could be improved with software updates.	2	6
78	Stereo speakees are also awesome.	4	5
79	I have purchased this mobile in offer so basic...	16	157
80	I think the biggest con of this phone is the c...	1	11

In [25]:

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

Out[25]:

	sentence	sentiment_value	word_count	index
0	As per this price range it is absolute patheti...	-3	33	0
1	Camera quality is just like 10000 to 15000 rs ...	4	19	1
2	After a month I am writing phone is great most...	3	13	2
3	Granted most of the reviewers on YouTube were ...	3	21	3
4	Probably that's the reason many of you were di...	-2	20	4
...	...	...	...	...
82	Sometimes the photo looks over-sharpened espec...	0	12	82
83	And the photo has too much contrast and noise.	0	9	83
84	Sometimes The phone picks up the wrong colour.	-2	8	84
85	Xiaomi needs to work on camera optimisation (f...	0	11	85
86	The front camera is OKAY.	0	5	86

87 rows × 4 columns

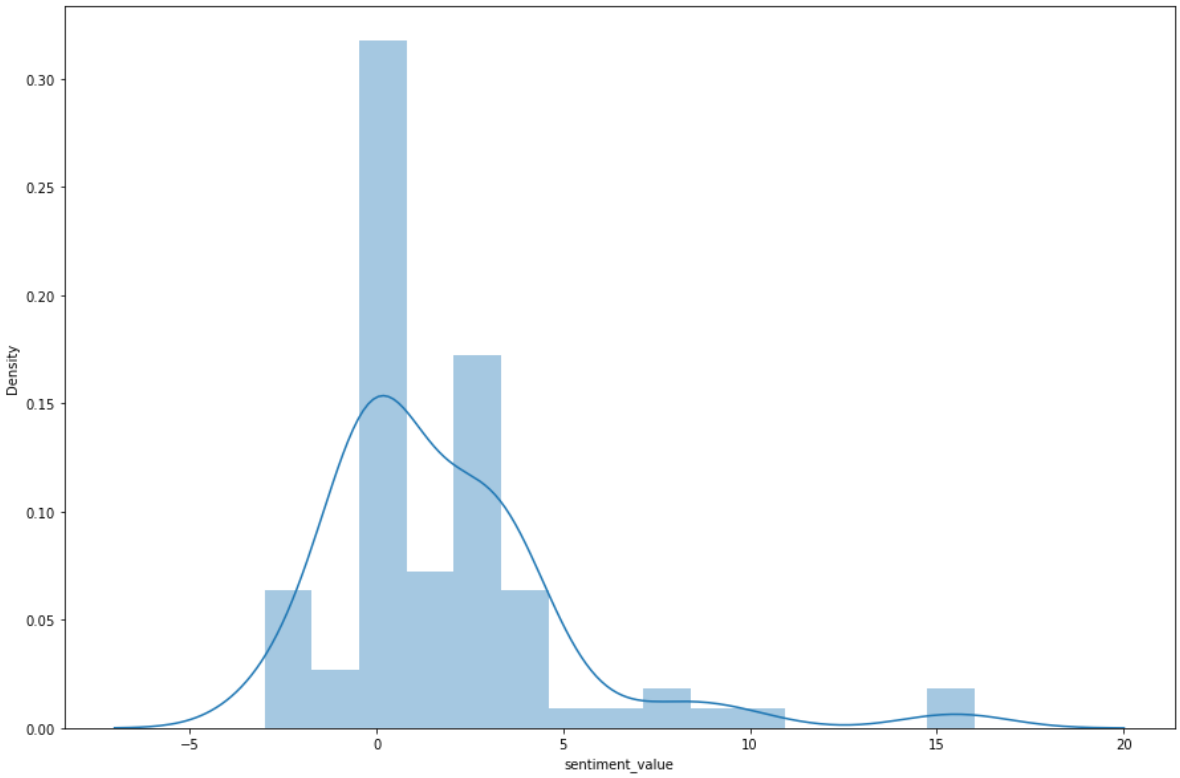
In [26]:

```
# Plotting the sentiment value for whole review
import seaborn as sns
plt.figure(figsize=(15,10))
sns.distplot(sent_df['sentiment_value'])
```

```
C:\Users\LENOVO\anaconda3\lib\site-packages\seaborn\distributions.py:2557: F
utureWarning: `distplot` is a deprecated function and will be removed in a f
uture version. Please adapt your code to use either `displot` (a figure-leve
l function with similar flexibility) or `histplot` (an axes-level function f
or histograms).
  warnings.warn(msg, FutureWarning)
```

Out[26]:

```
<AxesSubplot:xlabel='sentiment_value', ylabel='Density'>
```

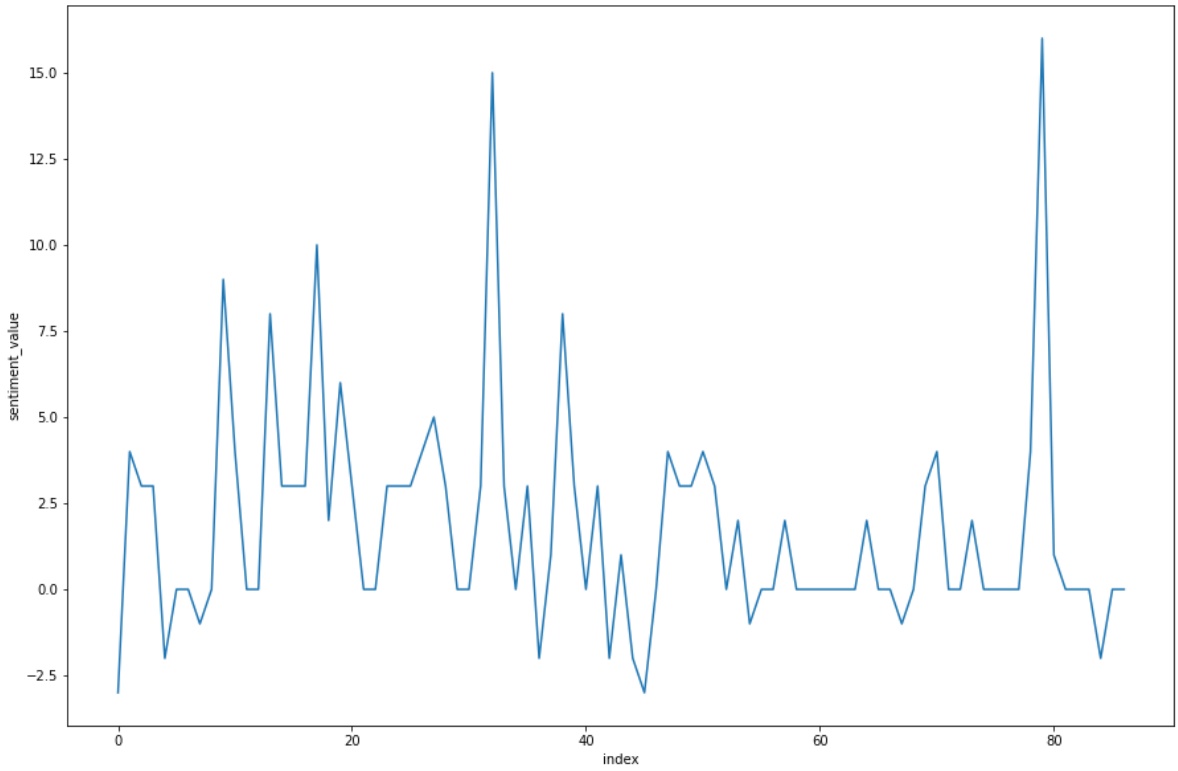


In [27]:

```
# Plotting the line plot for the sentiment value of whole review
plt.figure(figsize=(15,10))
sns.lineplot(y='sentiment_value',x='index',data=sent_df)
```

Out[27]:

```
<AxesSubplot:xlabel='index', ylabel='sentiment_value'>
```

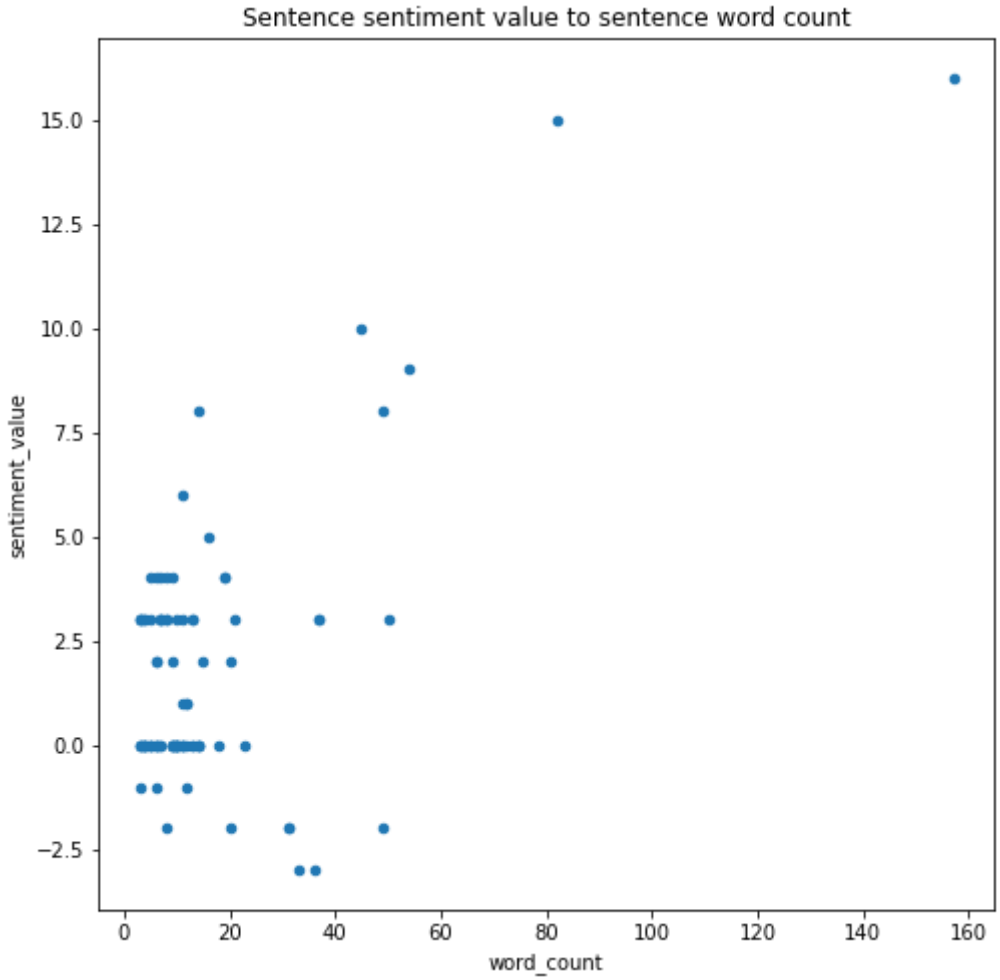


In [28]:

```
# Correlation anlysis
sent_df.plot.scatter(x='word_count', y='sentiment_value',figsize=(8,8),title='Sentence sent
```

Out[28]:

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



In [ ]:

```
```