

In []:

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

from matplotlib.pyplot import imread
from wordcloud import WordCloud, STOPWORDS
%matplotlib inline
# if not installed scrapy : !pip install scrapy
```

```
In [2]: reviews=pd.read_csv("C:\\Users\\nishi\\Desktop\\Assignments\\Text_Mining\\1429_1.
reviews
```

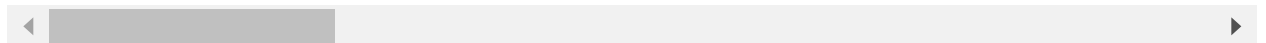
```
C:\Users\nishi\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:316
5: DtypeWarning: Columns (1,10) have mixed types.Specify dtype option on import
or set low_memory=False.
has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Out[2]:

		id	name	asins	brand	categories	
0	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon	
1	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon	
2	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon	
3	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon	
4	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon	
...	
34655	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets & Networking,Tablet & eBook ...	newamazonkindlef	
34656	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets & Networking,Tablet & eBook ...	newamazonkindlef	

	id	name	asins	brand	categories	
34657	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets & Networking,Tablet & eBook ...	newamazonkindlef
34658	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets & Networking,Tablet & eBook ...	newamazonkindlef
34659	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets & Networking,Tablet & eBook ...	newamazonkindlef

34660 rows × 21 columns



```
In [10]: review=reviews.iloc[0:5000]
print(review)
```

	id	name	
\			
0	AVqkIhwDv8e3D10-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	
1	AVqkIhwDv8e3D10-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	
2	AVqkIhwDv8e3D10-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	
3	AVqkIhwDv8e3D10-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	
4	AVqkIhwDv8e3D10-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	
...	
4995	AVphgVaX1cnluZ0-DR74	Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes...	
4996	AVphgVaX1cnluZ0-DR74	Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes...	
4997	AVphgVaX1cnluZ0-DR74	Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes...	
4998	AVphgVaX1cnluZ0-DR74	Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes...	
4999	AVphgVaX1cnluZ0-DR74	Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes...	
	asins	brand	categories
\			
0	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...
1	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...
2	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...
3	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...

```
In [11]: sample_list =review
converted_list = []

for element in sample_list:
    converted_list.append(element.strip())

print(converted_list)
```

```
['id', 'name', 'asins', 'brand', 'categories', 'keys', 'manufacturer', 'review
s.date', 'reviews.dateAdded', 'reviews.dateSeen', 'reviews.didPurchase', 'revie
ws.doRecommend', 'reviews.id', 'reviews.numHelpful', 'reviews.rating', 'review
s.sourceURLs', 'reviews.text', 'reviews.title', 'reviews.userCity', 'reviews.us
erProvince', 'reviews.username']
```

```
In [16]: review1=review.iloc[:,16]
print(review1)
```

```
0      This product so far has not disappointed. My c...
1      great for beginner or experienced person. Boug...
2      Inexpensive tablet for him to use and learn on...
3      I've had my Fire HD 8 two weeks now and I love...
4      I bought this for my grand daughter when she c...
...
4995    Amazing tablet for the price i love it is real...
4996    I love my fire... The games and the pictures g...
4997    Great tablet for kids but the screen isn't as ...
4998    Great product easy to use. Having different co...
4999    Love this tablet, great price and is really fa...
Name: reviews.text, Length: 5000, dtype: object
```

```
In [17]: review=[str(review1) for review1 in review.iloc[:,16]] # remove both the leading
review=[review1 for review1 in review if review1] # removes empty strings, because
review[0:10]
```

```
Out[17]: ['This product so far has not disappointed. My children love to use it and I
like the ability to monitor control what content they see with ease.',
'great for beginner or experienced person. Bought as a gift and she loves i
t',
'Inexpensive tablet for him to use and learn on, step up from the NABI. He w
as thrilled with it, learn how to Skype on it already...',
"I've had my Fire HD 8 two weeks now and I love it. This tablet is a great v
alue.We are Prime Members and that is where this tablet SHINES. I love being
able to easily access all of the Prime content as well as movies you can down
load and watch laterThis has a 1280/800 screen which has some really nice loo
k to it its nice and crisp and very bright infact it is brighter then the ipa
d pro costing $900 base model. The build on this fire is INSANELY AWESOME run
ning at only 7.7mm thick and the smooth glossy feel on the back it is really
amazing to hold its like the futuristic tab in ur hands.",
'I bought this for my grand daughter when she comes over to visit. I set it
up with her as the user, entered her age and name and now Amazon makes sure t
hat she only accesses sites and content that are appropriate to her age. Simp
le to do and she loves the capabilities. I also bought and installed a 64gig
SD card which gives this little tablet plenty of storage. For the price I thi
```

```
In [18]: review_text=' '.join(review)
review_text
```

```
Out[18]: 'This product so far has not disappointed. My children love to use it and I like the ability to monitor control what content they see with ease. great for beginner or experienced person. Bought as a gift and she loves it Inexpensive tablet for him to use and learn on, step up from the NABI. He was thrilled with it, learn how to Skype on it already... I\'ve had my Fire HD 8 two weeks now and I love it. This tablet is a great value.We are Prime Members and that is where this tablet SHINES. I love being able to easily access all of the Prime content as well as movies you can download and watch laterThis has a 1280/800 screen which has some really nice look to it its nice and crisp and very bright infact it is brighter then the ipad pro costing $900 base model. The build on this fire is INSANELY AWESOME running at only 7.7mm thick and the smooth glossy feel on the back it is really amazing to hold its like the futuristic tab in ur hands. I bought this for my grand daughter when she comes over to visit. I set it up with her as the user, entered her age and name and now Amazon makes sure that she only accesses sites and content that are appropriate to her age. Simple to do and she loves the capabilities. I also bought and installed a 64gig SD card which gives this little tablet plenty of storage. For the price I think this tablet is best one out there. You can spend hundred
```

```
In [19]: no_punc_text=review_text.translate(str.maketrans('','',string.punctuation))
no_punc_text
```

```
Out[19]: 'This product so far has not disappointed My children love to use it and I like the ability to monitor control what content they see with ease great for beginner or experienced person Bought as a gift and she loves it Inexpensive tablet for him to use and learn on step up from the NABI He was thrilled with it learn how to Skype on it already Ive had my Fire HD 8 two weeks now and I love it This tablet is a great valueWe are Prime Members and that is where this tablet SHINES I love being able to easily access all of the Prime content as well as movies you can download and watch laterThis has a 1280800 screen which has some really nice look to it its nice and crisp and very bright infact it is brighter then the ipad pro costing 900 base model The build on this fire is INSANELY AWESOME running at only 77mm thick and the smooth glossy feel on the back it is really amazing to hold its like the futuristic tab in ur hands I bought this for my grand daughter when she comes over to visit I set it up with her as the user entered her age and name and now Amazon makes sure that she only accesses sites and content that are appropriate to her age Simple to do and she loves the capabilities I also bought and installed a 64gig SD card which gives this little tablet plenty of storage For the price I think this tablet is best one out there You can spend hundreds of dollars more for additional speed and capacity but when it comes to the basics this tablets does everything that most people will ever need at a fraction of the cost This aw
```

```
In [20]: import nltk
nltk.download('punkt')
nltk.download('stopwords')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\nishi\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]   C:\Users\nishi\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[20]: True

```
In [21]: from nltk import word_tokenize
text_tokens=word_tokenize(no_punc_text)
print(text_tokens[0:50])
```

```
['This', 'product', 'so', 'far', 'has', 'not', 'disappointed', 'My', 'childre
n', 'love', 'to', 'use', 'it', 'and', 'I', 'like', 'the', 'ability', 'to', 'mon
itor', 'control', 'what', 'content', 'they', 'see', 'with', 'ease', 'great', 'f
or', 'beginner', 'or', 'experienced', 'person', 'Bought', 'as', 'a', 'gift', 'a
nd', 'she', 'loves', 'it', 'Inexpensive', 'tablet', 'for', 'him', 'to', 'use',
'and', 'learn', 'on']
```

```
In [22]: len(text_tokens)
```

Out[22]: 148661

```
In [23]: from nltk.corpus import stopwords
my_stop_words=stopwords.words('english')

sw_list=['I','The','It','A']
my_stop_words.extend(sw_list)

no_stop_tokens=[word for word in text_tokens if not word in my_stop_words]
print(no_stop_tokens)
```

```
['This', 'product', 'far', 'disappointed', 'My', 'children', 'love', 'use',
'like', 'ability', 'monitor', 'control', 'content', 'see', 'ease', 'great',
'beginner', 'experienced', 'person', 'Bought', 'gift', 'loves', 'Inexpensive',
'tablet', 'use', 'learn', 'step', 'NABI', 'He', 'thrilled', 'learn', 'Skype',
'already', 'Ive', 'Fire', 'HD', '8', 'two', 'weeks', 'love', 'This', 'tablet',
'great', 'valueWe', 'Prime', 'Members', 'tablet', 'SHINES', 'love', 'able',
'easily', 'access', 'Prime', 'content', 'well', 'movies', 'download',
'watch', 'laterThis', '1280800', 'screen', 'really', 'nice', 'look', 'nice',
'crisp', 'bright', 'infact', 'brighter', 'ipad', 'pro', 'costing', '900', 'base',
'model', 'build', 'fire', 'INSANELY', 'AWESOME', 'running', '77mm', 'thick',
'smooth', 'glossy', 'feel', 'back', 'really', 'amazing', 'hold', 'like',
'futuristic', 'tab', 'ur', 'hands', 'bought', 'grand', 'daughter', 'comes',
'visit', 'set', 'user', 'entered', 'age', 'name', 'Amazon', 'makes', 'sure',
'accesses', 'sites', 'content', 'appropriate', 'age', 'Simple', 'loves', 'capabilities',
'also', 'bought', 'installed', '64gig', 'SD', 'card', 'gives', 'little',
'tablet', 'plenty', 'storage', 'For', 'price', 'think', 'tablet', 'best',
'one', 'You', 'spend', 'hundreds', 'dollars', 'additional', 'speed', 'capacity',
'comes', 'basics', 'tablets', 'everything', 'people', 'ever', 'need',
'fraction', 'cost', 'This', 'amazon', 'fire', '8', 'inch', 'tablet', 'per
```

```
In [24]: lower_words=[comment.lower() for comment in no_stop_tokens]
print(lower_words)
```

```
['this', 'product', 'far', 'disappointed', 'my', 'children', 'love', 'use',
'like', 'ability', 'monitor', 'control', 'content', 'see', 'ease', 'great',
'beginner', 'experienced', 'person', 'bought', 'gift', 'loves', 'inexpensive',
'tablet', 'use', 'learn', 'step', 'nabi', 'he', 'thrilled', 'learn', 'skype',
'already', 'ive', 'fire', 'hd', '8', 'two', 'weeks', 'love', 'this', 'tablet',
'great', 'valuewe', 'prime', 'members', 'tablet', 'shines', 'love', 'able',
'easily', 'access', 'prime', 'content', 'well', 'movies', 'download',
'watch', 'laterthis', '1280800', 'screen', 'really', 'nice', 'look', 'nice',
'crisp', 'bright', 'infact', 'brighter', 'ipad', 'pro', 'costing', '900', 'base',
'model', 'build', 'fire', 'insanely', 'awesome', 'running', '77mm', 'thick',
'smooth', 'glossy', 'feel', 'back', 'really', 'amazing', 'hold', 'like',
'futuristic', 'tab', 'ur', 'hands', 'bought', 'grand', 'daughter', 'comes',
'visit', 'set', 'user', 'entered', 'age', 'name', 'amazon', 'makes', 'sure',
'accesses', 'sites', 'content', 'appropriate', 'age', 'simple', 'loves', 'capabilities',
'also', 'bought', 'installed', '64gig', 'sd', 'card', 'gives', 'little',
'tablet', 'plenty', 'storage', 'for', 'price', 'think', 'tablet', 'best',
'one', 'you', 'spend', 'hundreds', 'dollars', 'additional', 'speed', 'capacity',
'comes', 'basics', 'tablets', 'everything', 'people', 'ever', 'need',
'fraction', 'cost', 'this', 'amazon', 'fire', '8', 'inch', 'tablet', 'per
```

```
In [25]: from nltk.stem import PorterStemmer
ps=PorterStemmer()
stemmed_tokens=[ps.stem(word) for word in lower_words]
print(stemmed_tokens)
```

```
['thi', 'product', 'far', 'disappoint', 'my', 'children', 'love', 'use', 'lik
e', 'abil', 'monitor', 'control', 'content', 'see', 'eas', 'great', 'beginn',
'experienc', 'person', 'bought', 'gift', 'love', 'inexpens', 'tablet', 'use',
'learn', 'step', 'nabi', 'he', 'thrill', 'learn', 'skype', 'alreadi', 'ive',
'fire', 'hd', '8', 'two', 'week', 'love', 'thi', 'tablet', 'great', 'valuew',
'prime', 'member', 'tablet', 'shine', 'love', 'abl', 'easili', 'access', 'pri
me', 'content', 'well', 'movi', 'download', 'watch', 'laterthi', '1280800',
'screen', 'realli', 'nice', 'look', 'nice', 'crisp', 'bright', 'infact', 'bri
ghter', 'ipad', 'pro', 'cost', '900', 'base', 'model', 'build', 'fire', 'insa
n', 'awesom', 'run', '77mm', 'thick', 'smooth', 'glossi', 'feel', 'back', 're
alli', 'amaz', 'hold', 'like', 'futurist', 'tab', 'ur', 'hand', 'bought', 'gr
and', 'daughter', 'come', 'visit', 'set', 'user', 'enter', 'age', 'name', 'am
azon', 'make', 'sure', 'access', 'site', 'content', 'appropri', 'age', 'simp
l', 'love', 'capabl', 'also', 'bought', 'instal', '64gig', 'sd', 'card', 'giv
e', 'littl', 'tablet', 'plenti', 'storag', 'for', 'price', 'think', 'tablet',
'best', 'one', 'you', 'spend', 'hundr', 'dollar', 'addit', 'speed', 'capac',
'come', 'basic', 'tablet', 'everyth', 'peopl', 'ever', 'need', 'fraction', 'c
ost', 'thi', 'amazon', 'fire', '8', 'inch', 'tablet', 'perfect', 'size', 'pur
chas', 'husband', 'bigger', 'screen', 'phone', 'he', 'gotten', 'one', 'year',
```

```
In [26]: nlp=spacy.load('en_core_web_sm')
doc=nlp(' '.join(lower_words))
print(doc)
```

```
this product far disappointed my children love use like ability monitor contr
ol content see ease great beginner experienced person bought gift loves inexp
ensive tablet use learn step nabi he thrilled learn skype already ive fire hd
8 two weeks love this tablet great valuewe prime members tablet shines love a
ble easily access prime content well movies download watch laterthis 1280800
screen really nice look nice crisp bright infact brighter ipad pro costing 90
0 base model build fire insanely awesome running 77mm thick smooth glossy fee
l back really amazing hold like futuristic tab ur hands bought grand daughter
comes visit set user entered age name amazon makes sure accesses sites conten
t appropriate age simple loves capabilities also bought installed 64gig sd ca
rd gives little tablet plenty storage for price think tablet best one you spe
nd hundreds dollars additional speed capacity comes basics tablets everything
people ever need fraction cost this amazon fire 8 inch tablet perfect size pu
rchased husband bigger screen phone he gotten one years ago knew would good p
urchase great ereading go nice light weight price point given definitely wort
h purchase gave christmas gift inlaws husband uncle they loved easy use fanta
stic features great device read books like links borrowed library ebooks swit
ched another popular tablet brand happy choice made took time get books previ
ous nonkindle reader finally figured way love ordering books reading reader n
```



```
In [27]: lemmas=[token.lemma_ for token in doc]
print(lemmas)
```

```
['this', 'product', 'far', 'disappoint', 'my', 'child', 'love', 'use', 'lik
e', 'ability', 'monitor', 'control', 'content', 'see', 'ease', 'great', 'begi
nner', 'experienced', 'person', 'buy', 'gift', 'love', 'inexpensive', 'table
t', 'use', 'learn', 'step', 'nabi', 'he', 'thrill', 'learn', 'skype', 'alread
y', 'I', 've', 'fire', 'hd', '8', 'two', 'week', 'love', 'this', 'tablet', 'g
reat', 'valuewe', 'prime', 'member', 'tablet', 'shine', 'love', 'able', 'easi
ly', 'access', 'prime', 'content', 'well', 'movie', 'download', 'watch', 'lat
erthis', '1280800', 'screen', 'really', 'nice', 'look', 'nice', 'crisp', 'bri
ght', 'infact', 'bright', 'ipad', 'pro', 'cost', '900', 'base', 'model', 'bui
ld', 'fire', 'insanely', 'awesome', 'run', '77', 'mm', 'thick', 'smooth', 'gl
ossy', 'feel', 'back', 'really', 'amazing', 'hold', 'like', 'futuristic', 'ta
b', 'ur', 'hand', 'buy', 'grand', 'daughter', 'come', 'visit', 'set', 'user',
'enter', 'age', 'name', 'amazon', 'make', 'sure', 'access', 'site', 'conten
t', 'appropriate', 'age', 'simple', 'love', 'capability', 'also', 'buy', 'ins
tal', '64gig', 'sd', 'card', 'give', 'little', 'tablet', 'plenty', 'storage',
'for', 'price', 'think', 'tablet', 'good', 'one', 'you', 'spend', 'hundred',
'dollar', 'additional', 'speed', 'capacity', 'come', 'basic', 'tablet', 'ever
ything', 'people', 'ever', 'need', 'fraction', 'cost', 'this', 'amazon', 'fir
e', '8', 'inch', 'tablet', 'perfect', 'size', 'purchase', 'husband', 'big',
```

```
In [28]: clean_reviews=' '.join(lemmas)
clean_reviews
```

```
Out[28]: 'this product far disappoint my child love use like ability monitor control c
ontent see ease great beginner experienced person buy gift love inexpensive t
ablet use learn step nabi he thrill learn skype already I ve fire hd 8 two we
ek love this tablet great valuewe prime member tablet shine love able easily
access prime content well movie download watch laterthis 1280800 screen reall
y nice look nice crisp bright infact bright ipad pro cost 900 base model buil
d fire insanely awesome run 77 mm thick smooth glossy feel back really amazin
g hold like futuristic tab ur hand buy grand daughter come visit set user ent
er age name amazon make sure access site content appropriate age simple love
capability also buy instal 64gig sd card give little tablet plenty storage fo
r price think tablet good one you spend hundred dollar additional speed capac
ity come basic tablet everything people ever need fraction cost this amazon f
ire 8 inch tablet perfect size purchase husband big screen phone he get one y
ear ago know would good purchase great ereading go nice light weight price po
int give definitely worth purchase give christma gift inlaw husband uncle the
y love easy use fantastic feature great device read book like link borrow lib
rary ebook switch another popular tablet brand happy choice make take time ge
t book previous nonkindle reader finally figure way love order book read read
er not easy elderly user cease ad pop excellent product easy use large screen
```

```
In [35]: from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer()
reviewcv=cv.fit_transform(lemmas)
```

```
In [36]: print(cv.vocabulary_)
```

```
{'this': 4566, 'product': 3551, 'far': 1721, 'disappoint': 1369, 'my': 2988,
'child': 963, 'love': 2728, 'use': 4827, 'like': 2663, 'ability': 220, 'monit
or': 2924, 'control': 1126, 'content': 1121, 'see': 3976, 'ease': 1476, 'grea
t': 2093, 'beginner': 663, 'experienced': 1666, 'person': 3349, 'buy': 839,
'gift': 2000, 'inexpensive': 2363, 'tablet': 4436, 'learn': 2615, 'step': 428
6, 'nabi': 2991, 'he': 2175, 'thrill': 4580, 'skype': 4110, 'already': 370,
've': 4873, 'fire': 1793, 'hd': 2168, 'two': 4722, 'week': 4982, 'valuewe': 4
865, 'prime': 3525, 'member': 2827, 'shine': 4040, 'able': 222, 'easily': 148
0, 'access': 237, 'well': 4992, 'movie': 2949, 'download': 1430, 'watch': 495
9, 'laterthis': 2602, '1280800': 24, 'screen': 3945, 'really': 3683, 'nice':
3053, 'look': 2713, 'crisp': 1188, 'bright': 788, 'infact': 2366, 'ipad': 243
3, 'pro': 3537, 'cost': 1152, '900': 204, 'base': 617, 'model': 2906, 'buil
d': 818, 'insanely': 2377, 'awesome': 574, 'run': 3884, '77': 169, 'mm': 290
1, 'thick': 4552, 'smooth': 4148, 'glossy': 2024, 'feel': 1753, 'back': 588,
'amazing': 383, 'hold': 2232, 'futuristic': 1954, 'tab': 4431, 'ur': 4820, 'h
and': 2130, 'grand': 2073, 'daughter': 1233, 'come': 1038, 'visit': 4909, 'se
t': 4015, 'user': 4843, 'enter': 1576, 'age': 319, 'name': 2993, 'amazon': 38
6, 'make': 2765, 'sure': 4400, 'site': 4096, 'appropriate': 471, 'simple': 40
83, 'capability': 873, 'also': 372, 'instal': 2385, '64gig': 152, 'sd': 3956,
```

```
In [37]: print(cv.get_feature_names()[150:300])
```

```
['64', '6499', '64gig', '65', '68', '69', '6900', '695', '6999', '6th', '6yea
r', '6yr', '70he', '710', '72', '720p', '75', '750', '76', '77', '78', '79', '7
fire', '7inch', '7s', '7th', '7year', '7yearold', '7yr', '7â', '80', '800', '81
0', '83', '83year', '85', '86', '87', '88', '89', '8900', '8999', '89yearold',
'8hd', '8inch', '8inches', '8th', '8the', '8this', '8yearold', '8yo', '8yoshe',
'8yr', '90', '900', '92', '95', '96', '97', '99', '9900', '999', '9th', '9yearo
ld', '9yo', '9yr', 'a6', 'a8', 'abc', 'abcmouse', 'ability', 'abit', 'able', 'a
bount', 'about', 'absolute', 'absolutely', 'abundance', 'abundant', 'abuse', 'a
c', 'acc', 'acce', 'accent', 'accept', 'acceptable', 'acesory', 'access', 'acc
essibility', 'accessible', 'accessory', 'accident', 'accidental', 'accidentall
y', 'accidently', 'acclimate', 'accommodate', 'accomodate', 'accomplish', 'acco
mplished', 'accord', 'accordingly', 'account', 'accountpermission', 'accountsex
tra', 'accountsthe', 'accurate', 'accurateeasy', 'acer', 'acess', 'achieve', 'a
chievement', 'acquaint', 'acrobat', 'across', 'act', 'action', 'activate', 'act
ive', 'activity', 'actual', 'actually', 'ad', 'adaptable', 'adapter', 'adaptiv
e', 'adaptor', 'add', 'addict', 'addicted', 'addition', 'additional', 'addon',
'address', 'adequate', 'adequatelywould', 'adjust', 'adjustable', 'adjustment',
'adless', 'admit', 'admittedly', 'adopt', 'adorable', 'adore', 'adult', 'adults
super', 'adust', 'advance', 'advanced']
```

```
In [38]: print(reviewcv.toarray()[150:300])
```

```
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
```

```
In [39]: print(reviewcv.toarray().shape)
```

```
(81380, 5164)
```

```
In [40]: cv_ngram_range=CountVectorizer(analyzer='word',ngram_range=(1,3),max_features=10000)
bow_matrix_ngram=cv_ngram_range.fit_transform(lemmas)
```

```
In [41]: print(cv_ngram_range.get_feature_names())
print(bow_matrix_ngram.toarray())
```

```
['able', 'also', 'amazon', 'app', 'battery', 'be', 'big', 'black', 'book', 'bu
y', 'ca', 'charge', 'christmas', 'could', 'daughter', 'device', 'do', 'downloa
d', 'easy', 'enjoy', 'even', 'everything', 'far', 'fast', 'feature', 'find', 'f
ire', 'first', 'game', 'get', 'gift', 'give', 'go', 'good', 'great', 'happy',
'have', 'hd', 'he', 'internet', 'ipad', 'its', 'kid', 'kindle', 'life', 'ligh
t', 'like', 'little', 'long', 'look', 'lot', 'love', 'make', 'money', 'movie',
'much', 'my', 'need', 'new', 'nice', 'nt', 'old', 'one', 'perfect', 'play', 'pr
ice', 'prime', 'product', 'purchase', 'quality', 'read', 'reading', 'really',
'recommend', 'say', 'screen', 'set', 'she', 'size', 'small', 'son', 'store', 't
ablet', 'take', 'thing', 'think', 'this', 'time', 'use', 'very', 'video', 'wan
t', 'watch', 'we', 'well', 'wife', 'work', 'would', 'year', 'you']
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
```

```
In [42]: from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_ngram_max_features=TfidfVectorizer(norm='l2',analyzer='word',ngram_range=(
tfidf_matrix_ngram=tfidf_ngram_max_features.fit_transform(lemmas)
```

```
In [43]: print(tfidf_v_ngram_max_features.get_feature_names())
print(tfidf_matrix_ngram.toarray())
```

```
['10', '100', '16', '35', '50', 'ability', 'able', 'absolutely', 'access', 'acc
ount', 'actually', 'ad', 'add', 'adult', 'affordable', 'after', 'age', 'ago',
'alexa', 'all', 'allow', 'almost', 'already', 'also', 'always', 'amazing', 'ama
zon', 'amount', 'and', 'android', 'another', 'anyone', 'anything', 'anywhere',
'app', 'apple', 'around', 'as', 'ask', 'at', 'available', 'away', 'awesome', 'b
ack', 'bad', 'basic', 'battery', 'be', 'beat', 'beginner', 'big', 'birthday',
'bit', 'black', 'book', 'break', 'bright', 'bring', 'browse', 'browser', 'buc
k', 'budget', 'build', 'but', 'button', 'buy', 'ca', 'camera', 'card', 'carry',
'case', 'change', 'charge', 'charger', 'cheap', 'check', 'child', 'choice', 'ch
oose', 'christma', 'christmas', 'clear', 'close', 'color', 'come', 'comfortabl
e', 'compact', 'compare', 'complaint', 'computer', 'connect', 'connection', 'co
nsider', 'content', 'control', 'convenient', 'cost', 'could', 'couple', 'cove
r', 'customer', 'dad', 'daily', 'daughter', 'day', 'deal', 'decent', 'decide',
'definitely', 'design', 'device', 'difference', 'different', 'display', 'do',
'dollar', 'download', 'drop', 'due', 'durable', 'ease', 'easily', 'easy', 'eboo
k', 'electronic', 'else', 'email', 'end', 'enjoy', 'enough', 'entertainment',
'ereader', 'especially', 'etc', 'even', 'ever', 'every', 'everyday', 'everyon
e', 'everything', 'exactly', 'excellent', 'expect', 'expectation', 'expensive',
'experience', 'extra', 'extremely', 'eye', 'facebook', 'fact', 'family', 'far',
'fast', 'favorite', 'feature', 'feel', 'figure', 'finally', 'find', 'fine', 'fi
re', 'first', 'fit', 'for', 'free', 'friday', 'friend', 'friendly', 'front', 'f
ull', 'fun', 'function', 'game', 'gb', 'generation', 'get', 'gift', 'give', 'gl
ad', 'go', 'good', 'google', 'got', 'granddaughter', 'grandkid', 'grandson', 'g
raphic', 'great', 'hand', 'handle', 'handy', 'happy', 'hard', 'have', 'hd', 'hd
8', 'he', 'heavy', 'help', 'high', 'highly', 'hold', 'home', 'hope', 'hour', 'h
owever', 'huge', 'husband', 'if', 'impressed', 'improvement', 'in', 'inch', 'in
clude', 'inexpensive', 'install', 'instead', 'internet', 'ipad', 'issue', 'it',
'item', 'its', 'job', 'just', 'keep', 'keyboard', 'kid', 'kindle', 'know', 'lap
top', 'large', 'last', 'late', 'law', 'learn', 'leave', 'less', 'let', 'level',
'library', 'life', 'light', 'lightweight', 'like', 'limit', 'limited', 'liste
n', 'little', 'load', 'long', 'look', 'lot', 'love', 'loves', 'low', 'magazin
e', 'mainly', 'make', 'many', 'market', 'may', 'medium', 'meet', 'member', 'mem
ory', 'might', 'mind', 'mine', 'mini', 'minute', 'model', 'mom', 'money', 'mont
h', 'mostly', 'mother', 'movie', 'much', 'music', 'must', 'my', 'navigate', 'ne
ed', 'nephew', 'netflix', 'never', 'new', 'next', 'nice', 'niece', 'night', 'n
o', 'nook', 'not', 'nothing', 'now', 'nt', 'oasis', 'offer', 'often', 'ok', 'ol
d', 'one', 'online', 'only', 'open', 'operate', 'operating', 'option', 'order',
'original', 'os', 'other', 'overall', 'own', 'page', 'paperwhite', 'parent', 'p
arental', 'part', 'pay', 'people', 'perfect', 'perfectly', 'perform', 'performa
nce', 'person', 'phone', 'photo', 'pick', 'picture', 'place', 'play', 'playin
g', 'please', 'pleased', 'plenty', 'plug', 'plus', 'point', 'portable', 'powe
r', 'present', 'pretty', 'previous', 'price', 'prime', 'probably', 'problem',
'processor', 'product', 'purchase', 'purpose', 'purse', 'put', 'quality', 'quic
k', 'quickly', 'quite', 're', 'read', 'reader', 'reading', 'real', 'really', 'r
eason', 'reasonable', 'receive', 'recommend', 'regular', 'replace', 'replacemen
t', 'research', 'resolution', 'responsive', 'return', 'review', 'right', 'run',
'sale', 'samsung', 'satisfied', 'say', 'school', 'screen', 'sd', 'search', 'sec
ond', 'see', 'seem', 'sell', 'service', 'set', 'setup', 'several', 'sharp', 'sh
e', 'show', 'side', 'simple', 'since', 'sister', 'size', 'slow', 'small', 'so',
'someone', 'something', 'sometimes', 'son', 'sound', 'space', 'speaker', 'spee
d', 'spend', 'star', 'start', 'still', 'stop', 'storage', 'store', 'stream', 's
treaming', 'stuff', 'super', 'sure', 'surf', 'surfing', 'system', 'tab', 'tabl
e', 'tablet', 'take', 'tech', 'tell', 'thank', 'that', 'the', 'there', 'they',
```

```
'thing', 'think', 'third', 'this', 'though', 'thought', 'three', 'time', 'ton',
'top', 'touch', 'travel', 'trip', 'try', 'turn', 'tv', 'two', 'unit', 'update',
'upgrade', 'usb', 'use', 'user', 'value', 've', 'version', 'very', 'video', 'vi
ew', 'voyage', 'wait', 'want', 'watch', 'way', 'we', 'web', 'week', 'weight',
'well', 'what', 'when', 'white', 'wife', 'wifi', 'wish', 'without', 'wo', 'wond
erful', 'work', 'worth', 'would', 'wrong', 'year', 'yet', 'you', 'young', 'yout
ube', 'yr']
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
```

```
In [47]: def plot_cloud(wordcloud):
plt.figure(figsize=(40,30))
plt.imshow(wordcloud)
plt.axis('off')

# Generate word cloud

STOPWORDS.add('Pron')
wordcloud=WordCloud(width=3000,height=2000,background_color='white',max_words=100,
                    colormap='Set2',stopwords=STOPWORDS).generate(clean_reviews)
plot_cloud(wordcloud)
```



```
In [45]: nlp=spacy.load('en_core_web_sm')

one_block=clean_reviews
doc_block=nlp(one_block)
spacy.displacy.render(doc_block, style='ent', jupyter=True)
```

this product far disappoint my child love use like ability monitor control content see ease great
beginner experienced person buy gift love inexpensive tablet use learn step nabi he thrill learn
skype already I ve fire hd 8 two week **DATE** love this tablet great value we prime member
tablet shine love able easily access prime content well movie download watch later this
1280800 **CARDINAL** screen really nice look nice crisp bright in fact bright ipad pro cost 900
base model build fire insanely awesome run 77 mm **QUANTITY** thick smooth glossy feel
back really amazing hold like futuristic tab ur hand buy grand daughter come visit set user enter
age name amazon make sure access site content appropriate age simple love capability also
buy instal 64gig **CARDINAL** sd card give little tablet plenty storage for price think tablet good

```
In [48]: for token in doc_block[100:200]:  
         print(token,token.pos_)
```

```
visit VERB  
set NOUN  
user NOUN  
enter VERB  
age NOUN  
name NOUN  
amazon NOUN  
make VERB  
sure ADJ  
access NOUN  
site NOUN  
content NOUN  
appropriate ADJ  
age NOUN  
simple ADJ  
love NOUN  
capability NOUN  
also ADV  
buy VERB  
instal ADJ  
64gig NOUN  
sd ADP  
card NOUN  
give VERB  
little ADJ  
tablet NOUN  
plenty NOUN  
storage NOUN  
for ADP  
price NOUN  
think NOUN  
tablet NOUN  
good ADJ  
one NOUN  
you PRON  
spend VERB  
hundred NUM  
dollar NOUN  
additional ADJ  
speed NOUN  
capacity NOUN  
come VERB  
basic ADJ  
tablet NOUN  
everything PRON  
people NOUN  
ever ADV  
need VERB  
fraction NOUN  
cost VERB  
this DET  
amazon NOUN  
fire NOUN
```

8 NUM
inch NOUN
tablet NOUN
perfect ADJ
size NOUN
purchase NOUN
husband NOUN
big ADJ
screen NOUN
phone NOUN
he PRON
get VERB
one NUM
year NOUN
ago ADV
know VERB
would AUX
good ADJ
purchase VERB
great ADJ
ereading NOUN
go VERB
nice ADJ
light ADJ
weight NOUN
price NOUN
point NOUN
give VERB
definitely ADV
worth ADJ
purchase NOUN
give VERB
christma NOUN
gift NOUN
inlaw NOUN
husband NOUN
uncle NOUN
they PRON
love VERB
easy ADJ
use NOUN
fantastic ADJ
feature NOUN
great ADJ
device NOUN
read VERB
book NOUN


```
In [49]: nouns_verbs=[token.text for token in doc_block if token.pos_ in ('NOUN','VERB')]
print(nouns_verbs[100:200])
```

```
['purchase', 'ereading', 'go', 'weight', 'price', 'point', 'give', 'purchase',
'give', 'christma', 'gift', 'inlaw', 'husband', 'uncle', 'love', 'use', 'featur
e', 'device', 'read', 'book', 'link', 'borrow', 'library', 'ebook', 'switch',
'tablet', 'brand', 'choice', 'make', 'take', 'time', 'get', 'book', 'nonkindl
e', 'reader', 'figure', 'way', 'love', 'order', 'book', 'read', 'reader', 'use
r', 'cease', 'ad', 'pop', 'product', 'use', 'screen', 'make', 'watch', 'movie',
'read', 'want', 'father', 'tablet', 'value', 'watch', 'movie', 'play', 'game',
'use', 'need', 'thank', 'youand', 'silk', 'work', 'wonder', 'get', 'love', 'siz
e', 'screen', 'kindle', 'use', 'screen', 'color', 'reader', 'wantneed', 'bargai
n', 'tablet', 'give', 'star', 'push', 'start', 'time', 'unlock', 'screen', 'vid
eo', 'quality', 'lot', 'fun', 'app', 'fun', 'family', 'love', 'love', 'love',
'kindle', 'fire', 'yn']
```

```
In [50]: from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer()

X=cv.fit_transform(nouns_verbs)
sum_words=X.sum(axis=0)

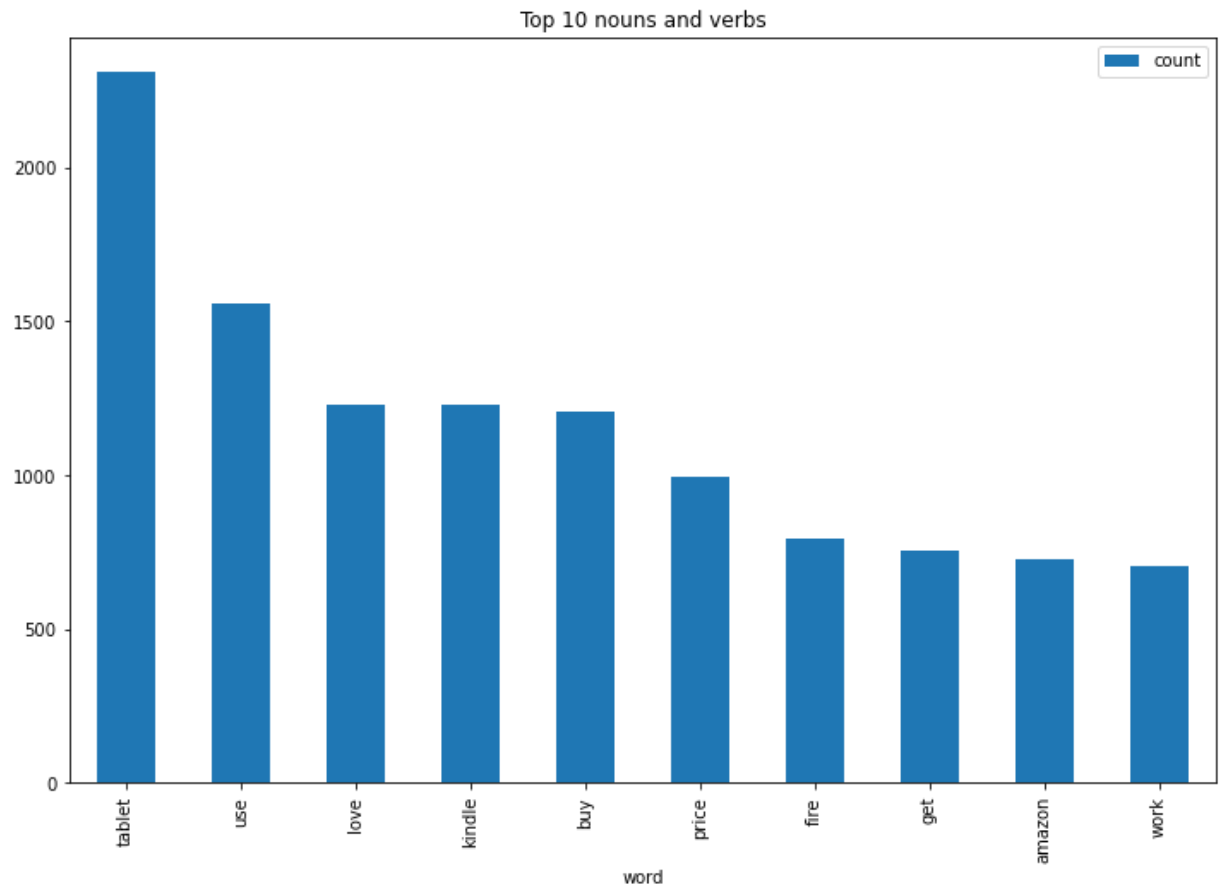
words_freq=[(word,sum_words[0,idx]) for word,idx in cv.vocabulary_.items()]
words_freq=sorted(words_freq,key=lambda x: x[1],reverse=True)

wd_df=pd.DataFrame(words_freq)
wd_df.columns=['word','count']
wd_df[0:10] # viewing top ten results
```

Out[50]:

	word	count
0	tablet	2309
1	use	1557
2	love	1229
3	kindle	1228
4	buy	1209
5	price	996
6	fire	794
7	get	754
8	amazon	728
9	work	705

```
In [51]: # Visualizing results (Barchart for top 10 nouns + verbs)
wd_df[0:10].plot.bar(x='word',figsize=(12,8),title='Top 10 nouns and verbs');
```



```
In [52]: from nltk import tokenize
sentences=tokenize.sent_tokenize(' '.join(review))
sentences
```

```
Out[52]: ['This product so far has not disappointed.',
'My children love to use it and I like the ability to monitor control what c
ontent they see with ease.',
'great for beginner or experienced person.',
'Bought as a gift and she loves it Inexpensive tablet for him to use and lea
rn on, step up from the NABI.',
'He was thrilled with it, learn how to Skype on it already...',
"I've had my Fire HD 8 two weeks now and I love it.",
'This tablet is a great value.We are Prime Members and that is where this ta
blet SHINES.',
'I love being able to easily access all of the Prime content as well as movi
es you can download and watch laterThis has a 1280/800 screen which has some
really nice look to it its nice and crisp and very bright infact it is bright
er then the ipad pro costing $900 base model.',
'The build on this fire is INSANELY AWESOME running at only 7.7mm thick and
the smooth glossy feel on the back it is really amazing to hold its like the
futuristic tab in ur hands.',
'I bought this for my grand daughter when she comes over to visit.',
'I set it up with her as the user, entered her age and name and now Amazon m
```

```
In [53]: sent_df=pd.DataFrame(sentences,columns=['sentence'])
sent_df
```

```
Out[53]:
```

	sentence
0	This product so far has not disappointed.
1	My children love to use it and I like the abil...
2	great for beginner or experienced person.
3	Bought as a gift and she loves it Inexpensive ...
4	He was thrilled with it, learn how to Skype on...
...	...
11582	The games and the pictures great Great tablet ...
11583	Charging port is very poorly built Great produ...
11584	Having different color choices wonderful.
11585	Love this tablet, great price and is really fast.
11586	Is worth every penny.

11587 rows × 1 columns

```
In [54]: affin=pd.read_csv("C:\\Users\\nishi\\Desktop\\Assignments\\Text_Mining\\Afinn.csv")
affin
```

```
Out[54]:
```

	word	value
0	abandon	-2
1	abandoned	-2
2	abandons	-2
3	abducted	-2
4	abduction	-2
...
2472	yucky	-2
2473	yummy	3
2474	zealot	-2
2475	zealots	-2
2476	zealous	2

2477 rows × 2 columns

```
In [55]: affinity_scores=affin.set_index('word')['value'].to_dict()
affinity_scores
```

```
Out[55]: {'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,
'absolving': 2,
'absorbed': 1
```

```
In [56]: 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 [57]: calculate_sentiment(text='good service')
```

```
Out[57]: 3
```

```
In [58]: sent_df['sentiment_value']=sent_df['sentence'].apply(calculate_sentiment)
         sent_df['sentiment_value']
```

```
Out[58]: 0      -2
         1       9
         2       3
         3       5
         4       0
         ..
        11582    8
        11583    4
        11584    4
        11585    6
        11586    2
         Name: sentiment_value, Length: 11587, dtype: int64
```

```
In [59]: sent_df['word_count']=sent_df['sentence'].str.split().apply(len)
         sent_df['word_count']
```

```
Out[59]: 0       7
         1      20
         2       6
         3      22
         4      12
         ..
        11582   16
        11583   11
        11584    5
        11585    9
        11586    4
         Name: word_count, Length: 11587, dtype: int64
```

```
In [60]: sent_df.sort_values(by='sentiment_value')
```

```
Out[60]:
```

	sentence	sentiment_value	word_count
2545	That Samsung tab lagged a lot but I haven't se...	-10	28
10623	So I've been wanting a fire since they origina...	-9	98
10317	I'm used to using iPads, but not a bad little ...	-6	24
6748	When I was holding my Paperwhite with my left ...	-6	51
10002	Overall the tablet is okay, I just paid 50\$ so...	-6	25
...
9303	Small and cheap gift for my nieces who like to...	20	55
4263	Definitely an amazing deal at an amazing price...	21	52
283	my kids enjoy the Fire HD and they love the pi...	21	142
9165	Like the product, takes a little time to resto...	26	131
38	Great video quality lots of fun apps fun for t...	31	76

11587 rows × 3 columns

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

```
In [61]: # negative sentiment score of the whole review  
sent_df[sent_df['sentiment_value']<=0]
```

```
Out[61]:
```

	sentence	sentiment_value	word_count
0	This product so far has not disappointed.	-2	7
4	He was thrilled with it, learn how to Skype on...	0	12
10	I set it up with her as the user, entered her ...	0	32
12	I also bought and installed a 64gig SD card wh...	0	17
14	You can spend hundreds of dollars more for add...	0	35
...
11569	Once you figure it out, it isn't bad.	-3	8
11572	Tablet was purchased by our daughter with her ...	0	10
11573	She worked her tail off for this tablet.	0	8
11575	Right size for travel.	0	4
11577	Everyone should own one.	0	4

4423 rows × 3 columns

```
In [62]: sent_df[sent_df['sentiment_value']>0]
```

```
Out[62]:
```

	sentence	sentiment_value	word_count
1	My children love to use it and I like the abil...	9	20
2	great for beginner or experienced person.	3	6
3	Bought as a gift and she loves it Inexpensive ...	5	22
5	I've had my Fire HD 8 two weeks now and I love...	3	13
6	This tablet is a great value.We are Prime Memb...	3	16
...
11582	The games and the pictures great Great tablet ...	8	16
11583	Charging port is very poorly built Great produ...	4	11
11584	Having different color choices wonderful.	4	5
11585	Love this tablet, great price and is really fast.	6	9
11586	Is worth every penny.	2	4

7164 rows × 3 columns

```
In [63]: sent_df['index']=range(0,len(sent_df))
sent_df
```

```
Out[63]:
```

	sentence	sentiment_value	word_count	index
0	This product so far has not disappointed.	-2	7	0
1	My children love to use it and I like the abil...	9	20	1
2	great for beginner or experienced person.	3	6	2
3	Bought as a gift and she loves it Inexpensive ...	5	22	3
4	He was thrilled with it, learn how to Skype on...	0	12	4
...
11582	The games and the pictures great Great tablet ...	8	16	11582
11583	Charging port is very poorly built Great produ...	4	11	11583
11584	Having different color choices wonderful.	4	5	11584
11585	Love this tablet, great price and is really fast.	6	9	11585
11586	Is worth every penny.	2	4	11586

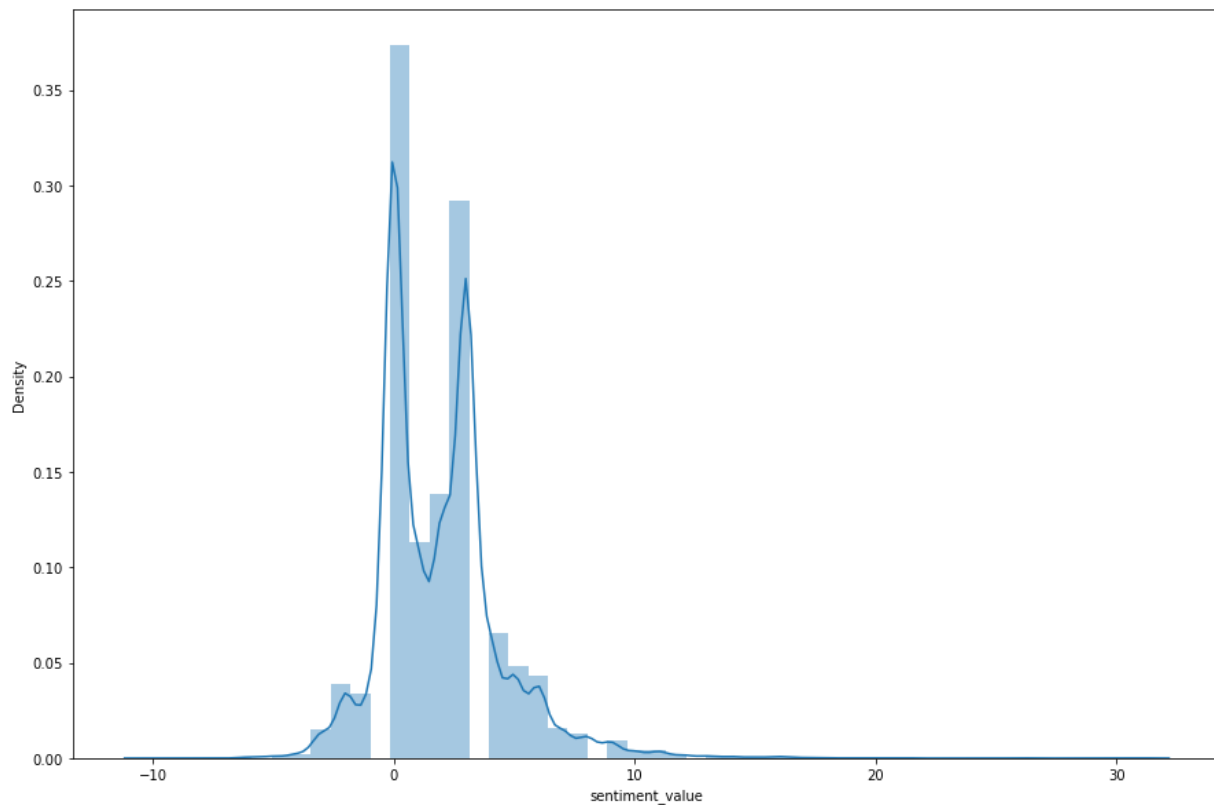
11587 rows × 4 columns

```
In [64]: import seaborn as sns
plt.figure(figsize=(15,10))
sns.distplot(sent_df['sentiment_value'])
```

C:\Users\nishi\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).

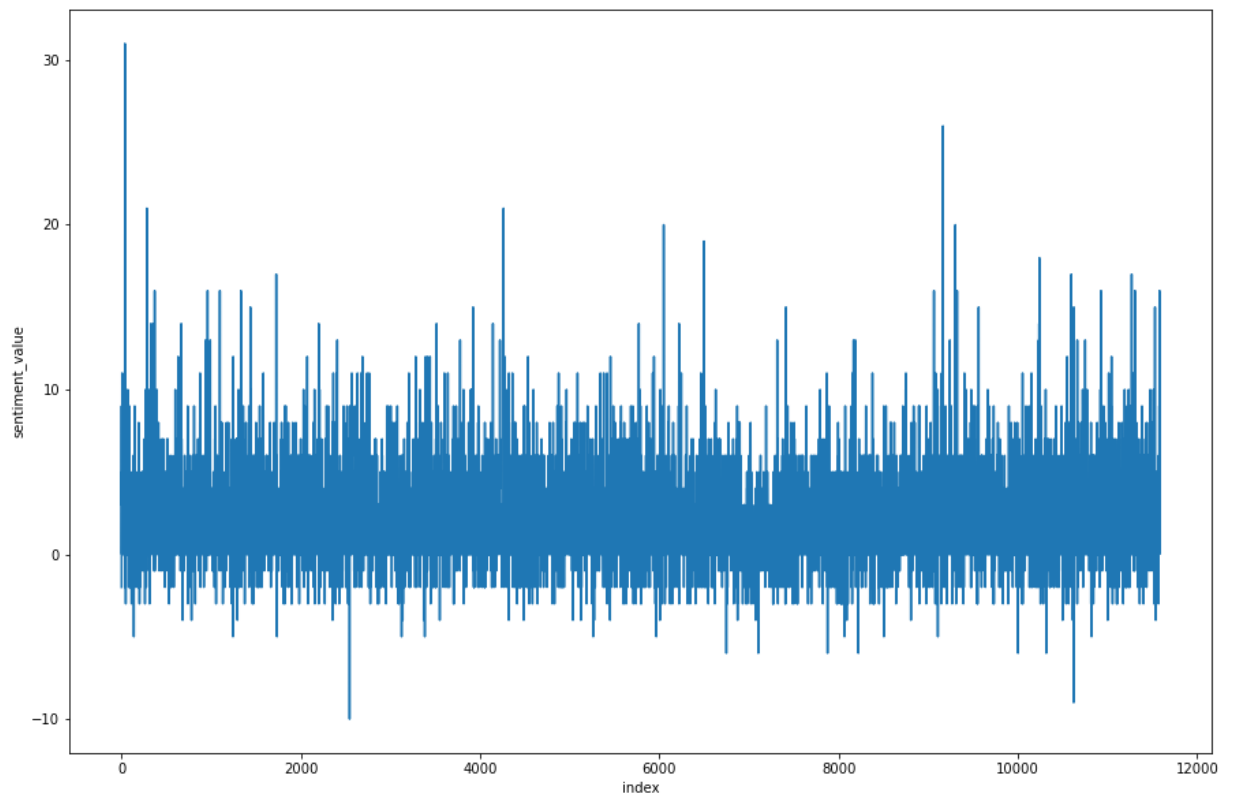
```
warnings.warn(msg, FutureWarning)
```

Out[64]: <AxesSubplot:xlabel='sentiment_value', ylabel='Density'>




```
In [65]: plt.figure(figsize=(15,10))
sns.lineplot(y='sentiment_value',x='index',data=sent_df)
```

```
Out[65]: <AxesSubplot:xlabel='index', ylabel='sentiment_value'>
```



```
In [ ]:
```