

```
# Import extracted amazon reviews Dataset "extract_reviews_fire_bolt.csv"
reviews=pd.read_csv('extract_reviews_fire_bolt.csv')
reviews
```

	stars	comment
0	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n Writing this review...
1	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n I am writing this ...
2	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n Touch is not much ...
3	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n First of all desig...
4	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n Touch is not much ...
...
1587	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n First of all desig...
1588	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n Touch is not much ...
1589	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n I have found the t...
1590	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n Awesome product, d...
1591	4.0 out of 5 stars	\n\n\n\n\n\n\n\n\n\n\n\n I am using the wat...

1592 rows × 2 columns

▼ Text Preprocessing

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

["Writing this review after gifting it to my brother and using for 1 hour. Watch touch screen is awesome, no lag and smooth running. Charging is quite fast. 20 minutes and 20-60% charge. Watch is good looking too. I haven't test the sleep tracking yet. But my bro will surely tell the sleep tracking experience and I'll edit this review. The only bad (actually worst) thing about this watch is that the sensors are fully fake. I tested it on a table and it shows that the table has 78 bpm. If you wear it you will also get the same result. Watch is quite good, you could go for it if you are not looking for a heart beat, SpO2 sensing watch because they are worst.\n \nRead more",

'I am writing this detailed review after using this product for a week:Build Quality: It is metal body but not the pure metal material, overall build quality is good.Screen: Amazing experience with screen, even it is not Amoled display but you will not feel the

need as brightness is sufficient when wants to use in Sunlight. Touch is very smooth but you will get scratches on screen easily. Wakeup function was not working initially but worked pretty well after resetting. Accuracy: Most of the Smartwatches measure ghost reading so this does but overall accuracy is good in this budget. Battery: The average battery backup is 6-7 days, the battery backup will be 2-3 days if you are using calling function moderately. It takes 1.5 hour to full charge. Bluetooth Calling: Works far better than my expectations, the mic & speaker are good and you can talk without any background disturbance. The person on other side can also hear clear sound. Overall, this is good product in this price range. Go for it if you are looking budget friendly round dial watch.\n \nRead more',

"Touch is not much good, display size can be improved. It's a good product in budget. Don't expect too much from this product because in this price range you will get Bluetooth calling, SPO2, SLEEP TRACKING, etc in this watch. In this price range you can go for this watch to try it. The touch is good not seeing delay when it touch and yes it's a good product not wow but yes affordable. Built quality is ok and watch strap is good. Too light you don't feel that you are wearing something in your wrist. And speaker is decent you can talk in the room but not so clear in public.\n \nRead more",

'First of all design, and built quality is excellent. Beautiful watch in this price range. Sleep tracking, touch screen, Voice is clear, Mic is also good. And all other features are good. Good battery life minimum 4 to 5 days while using theatre mode or battery save mood. And Bluetooth calling mood 3 to 4 days working. Customizable watch faces are missing, No watch face storage option, always download and apply. There is no silent option for call. When a call or notification comes it rings very loudly and it cannot be changed. And I am not satisfied favourite contact list only 8. It will be good if fixed through software updates. Overall very nice product value for money.\n \nRead more',

"Touch is not much good, display size can be improved. It's a good product in budget. Don't expect too much from this product because in this price range you will get Bluetooth calling, SPO2, SLEEP TRACKING, etc in this watch. In this price range you can go for this watch to try it. The touch is good not seeing delay when it touch and yes it's a good product not wow but yes affordable. Built quality is ok and watch strap is good.\n \nRead more",

"I have found the the watch is good but there are some big flaws in the watch. The biggest of them is there is no silent option for call. When a call or notification comes it rings very loudly and it cannot be vibret. The battery life is more or less 5 to 6 days. After all these this is a good watch that's why I gave it 4 stars.\n \nRead more",

'Awesome product, does not betray at any point, also has motion sensors meaning when you move up your hand to see the time the screen automatically lights up which is a really cool and amazing feature not found in other watches of this range. Watch is good as compared to money. Touch is good and according to functions value for money. Personally I found it's bit heavy as compared to my another smart watch. Also using third party app for all functioning and allow permissions to access data, is not looks good in terms of data security. Absolutely loved this and highly recommend this.\n \nRead more',

"I am using the watch since 2 days. I have found the the watch is good but there are some big flaws in the watch. The biggest of them is there is no silent option for call. When a call or notification comes it rings very loudly and it cannot be changed. I am not able to add personalized wallpaper as well, I doesn't like the watch faces except a few of them. Regarding the battery life. I have charged it only once and it still has 85-90% of charge

```
# Joining the list into one string/text
reviews_text=' '.join(reviews)
reviews_text
```

'Writing this review after gifting it to my brother and using for 1 hour. Watch touch screen is awesome, no lag and smooth running. Charging is quite fast. 20 minutes and 20-60% charge. Watch is good looking too. I haven't test the sleep tracking yet. But my bro will surely tell the sleep tracking experience and I'll edit this review. The only bad (actually worst) thing about this watch is that the sensors are fully fake. I tested it on a table and it shows that the table has 78 bpm. If you wear it you will also get the same result. Watch is quite good, you could go for it if you are not looking for a heart beat, SpO2 sensing watch because they are worst \n \nRead more I am writing this detailed review after using this product for a week. Built

```
# Remove Punctuations
no_punc_text=reviews_text.translate(str.maketrans('', '', string.punctuation))
no_punc_text
```

'Writing this review after gifting it to my brother and using for 1 hour Watch touch screen is awesome no lag and smooth running Charging is quite fast 20 minutes and 2060 charge Watch is good looking too I haven't test the sleep tracking yet But my bro will surely tell the sleep tracking experience and I'll edit this review The only bad actually worst thing about this watch is that the sensors are fully fake I tested it on a table and it shows that the table has 78 bpm If you wear it you will also get the same result Watch is quite good you could go for it if you are not looking for a heart beat SpO2 sensing watch because they are worst\n \nRead more
I am writing this detailed review after using this product for a week Build Quality It is metal

```
# Tokenization
```

```
import nltk
```

```
nltk.download('punkt')
```

```
nltk.download('stopwords')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
```

```
[nltk_data] Package punkt is already up-to-date!
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
```

```
[nltk_data] Package stopwords is already up-to-date!
```

```
True
```

```
from nltk import word_tokenize
```

```
text_tokens=word_tokenize(no_punc_text)
```

```
print(text_tokens[0:50])
```

```
['Writing', 'this', 'review', 'after', 'gifting', 'if', 'to', 'my', 'brother', 'and', 'using',
```

```
len(text_tokens)
```

```
185667
```

```
# Remove stopwords
```

```
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)
```

```
['Writing', 'review', 'gifting', 'brother', 'using', '1', 'hour', 'Watch', 'touch', 'screen',
```

```
# Normalize the data
```

```
lower_words=[comment.lower() for comment in no_stop_tokens]
```

```
print(lower_words)
```

```
['writing', 'review', 'gifting', 'brother', 'using', '1', 'hour', 'watch', 'touch', 'screen',
```

```
# Stemming (Optional)
```

```
from nltk.stem import PorterStemmer
```

```
ps=PorterStemmer()
```

```
stemmed_tokens=[ps.stem(word) for word in lower_words]
```

```
print(stemmed_tokens)
```

```
['write', 'review', 'gift', 'brother', 'use', '1', 'hour', 'watch', 'touch', 'screen', 'awesome', 'lag', 'smooth', 'running', 'charge', 'quite', 'fast', '20', 'minute', '2060', 'charge', 'watch', 'good', 'looking', 'have', 'not', 'test', 'sleep', 'track', 'yet', 'but', 'bro', 'sure', 'ly', 'tell', 'sleep', 'track', 'experience', 'ill', 'edit', 'review', 'bad', 'actually', 'bad', 'thing', 'watch', 'sensor', 'fully', 'fake', 'test', 'table', 'show', 'table', '78', 'bpm', 'if', 'wear', 'also', 'get', 'result', 'watch', 'quite', 'good', 'could', 'go', 'look', 'heart', 'beat', 'spo2', 'sense', 'watch', 'worst', 'read', 'write', 'detailed', 'review', 'use', 'product', 'week', 'build', 'quality', 'metal', 'body', 'pure', 'metal', 'material', 'overall', 'build', 'quality', 'goods', 'screen', 'amazing', 'experience', 'screen', 'even', 'amole', 'display', 'feel', 'need', 'brightness', 'sufficient', 'want', 'use', 'sunlight', 'touch', 'smooth', 'get', 'scratch', 'screen', 'easily', 'wake', 'up', 'n', 'function', 'work', 'initially', 'work', 'pretty', 'well', 'resetting', 'accuracy', 'most', 'smartwatch', 'measure', 'ghost', 're
```

```
# Lemmatization
```

```
nlp=spacy.load('en_core_web_sm')
```

```
doc=nlp(' '.join(lower_words))
```

```
print(doc)
```

```
writing review gifting brother using 1 hour watch touch screen awesome lag smooth running charge
```

```
lemmas=[token.lemma_ for token in doc]
```

```
print(lemmas)
```

```
['write', 'review', 'gifting', 'brother', 'use', '1', 'hour', 'watch', 'touch', 'screen', 'awesome', 'lag', 'smooth', 'running', 'charge', 'quite', 'fast', '20', 'minute', '2060', 'charge', 'watch', 'good', 'looking', 'have', 'not', 'test', 'sleep', 'track', 'yet', 'but', 'bro', 'sure', 'ly', 'tell', 'sleep', 'track', 'experience', 'ill', 'edit', 'review', 'bad', 'actually', 'bad', 'thing', 'watch', 'sensor', 'fully', 'fake', 'test', 'table', 'show', 'table', '78', 'bpm', 'if', 'wear', 'also', 'get', 'result', 'watch', 'quite', 'good', 'could', 'go', 'look', 'heart', 'beat', 'spo2', 'sense', 'watch', 'worst', 'read', 'write', 'detailed', 'review', 'use', 'product', 'week', 'build', 'quality', 'metal', 'body', 'pure', 'metal', 'material', 'overall', 'build', 'quality', 'goods', 'screen', 'amazing', 'experience', 'screen', 'even', 'amole', 'display', 'feel', 'need', 'brightness', 'sufficient', 'want', 'use', 'sunlight', 'touch', 'smooth', 'get', 'scratch', 'screen', 'easily', 'wake', 'up', 'n', 'function', 'work', 'initially', 'work', 'pretty', 'well', 'resetting', 'accuracy', 'most', 'smartwatch', 'measure', 'ghost', 're
```

```
clean_reviews=' '.join(lemmas)
```

```
clean_reviews
```

```
'write review gifting brother use 1 hour watch touch screen awesome lag smooth running charge quite fast 20 minute 2060 charge watch good looking have not test sleep track yet but bro sure ly tell sleep track experience ill edit review bad actually bad thing watch sensor fully fake test table show table 78 bpm if wear also get result watch quite good could go look heart beat spo2 sense watch worst read write detailed review use product week build quality metal body pure metal material overall build quality goodscreen amazing experience screen even amole display feel need brightness sufficient want use sunlight touch smooth get scratch screen easily wake up n function work initially work pretty well resetting accuracy most smartwatch measure ghost re
```

Feature Extraction

1. Using CountVectorizer

```
from sklearn.feature_extraction.text import CountVectorizer
```

```
cv=CountVectorizer()
```

```
reviewscv=cv.fit_transform(lemmas)
```

```
print(cv.vocabulary_)
```

```
{'write': 252, 'review': 180, 'gifting': 101, 'brother': 48, 'use': 234, 'hour': 113, 'watch': 114, 'lag': 115, 'smooth': 116, 'running': 117, 'charge': 118, 'quite': 119, 'fast': 120, '20': 121, 'minute': 122, '2060': 123, 'charge': 124, 'watch': 125, 'good': 126, 'looking': 127, 'have': 128, 'not': 129, 'test': 130, 'sleep': 131, 'track': 132, 'yet': 133, 'but': 134, 'bro': 135, 'sure': 136, 'ly': 137, 'tell': 138, 'sleep': 139, 'track': 140, 'experience': 141, 'ill': 142, 'edit': 143, 'review': 144, 'bad': 145, 'actually': 146, 'bad': 147, 'thing': 148, 'watch': 149, 'sensor': 150, 'fully': 151, 'fake': 152, 'test': 153, 'table': 154, 'show': 155, 'table': 156, '78': 157, 'bpm': 158, 'if': 159, 'wear': 160, 'also': 161, 'get': 162, 'result': 163, 'watch': 164, 'quite': 165, 'good': 166, 'could': 167, 'go': 168, 'look': 169, 'heart': 170, 'beat': 171, 'spo2': 172, 'sense': 173, 'watch': 174, 'worst': 175, 'read': 176, 'write': 177, 'detailed': 178, 'review': 179, 'use': 180, 'product': 181, 'week': 182, 'build': 183, 'quality': 184, 'metal': 185, 'body': 186, 'pure': 187, 'metal': 188, 'material': 189, 'overall': 190, 'build': 191, 'quality': 192, 'goods': 193, 'screen': 194, 'amazing': 195, 'experience': 196, 'screen': 197, 'even': 198, 'amole': 199, 'display': 200, 'feel': 201, 'need': 202, 'brightness': 203, 'sufficient': 204, 'want': 205, 'use': 206, 'sunlight': 207, 'touch': 208, 'smooth': 209, 'get': 210, 'scratch': 211, 'screen': 212, 'easily': 213, 'wake': 214, 'up': 215, 'n': 216, 'function': 217, 'work': 218, 'initially': 219, 'work': 220, 'pretty': 221, 'well': 222, 'resetting': 223, 'accuracy': 224, 'most': 225, 'smartwatch': 226, 'measure': 227, 'ghost': 228, 're': 229}
```

```
print(cv.get_feature_names()[50:250])
```

```
['budget', 'battery', 'build', 'but', 'call', 'calling', 'change', 'charge', 'charge', 'bluetooth', 'c',  
/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function  
warnings.warn(msg, category=FutureWarning)
```

```
print(reviewscv.toarray()[50:250])
```

```
[[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]]
```

```
print(reviewscv.toarray().shape)
```

```
(102684, 255)
```

2. CountVectorizer with N-grams (Bigrams & Trigrams)

```
cv_ngram_range=CountVectorizer(analyzer='word',ngram_range=(1,3),max_features=100)
bow_matrix_ngram=cv_ngram_range.fit_transform(lemmas)
```

```
print(cv_ngram_range.get_feature_names())
print(bow_matrix_ngram.toarray())
```

```
['affordable', 'also', 'amazing', 'and', 'awesome', 'bad', 'battery', 'big', 'bluetooth', 'budg
[[0 0 0 ... 0 1 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]]
```

3. TF-IDF Vectorizer

```
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_v_ngram_max_features=TfidfVectorizer(norm='l2',analyzer='word',ngram_range=(1,3),max_features=5
tfidf_matrix_ngram=tfidf_v_ngram_max_features.fit_transform(lemmas)
```

```
print(tfidf_v_ngram_max_features.get_feature_names())
print(tfidf_matrix_ngram.toarray())
```

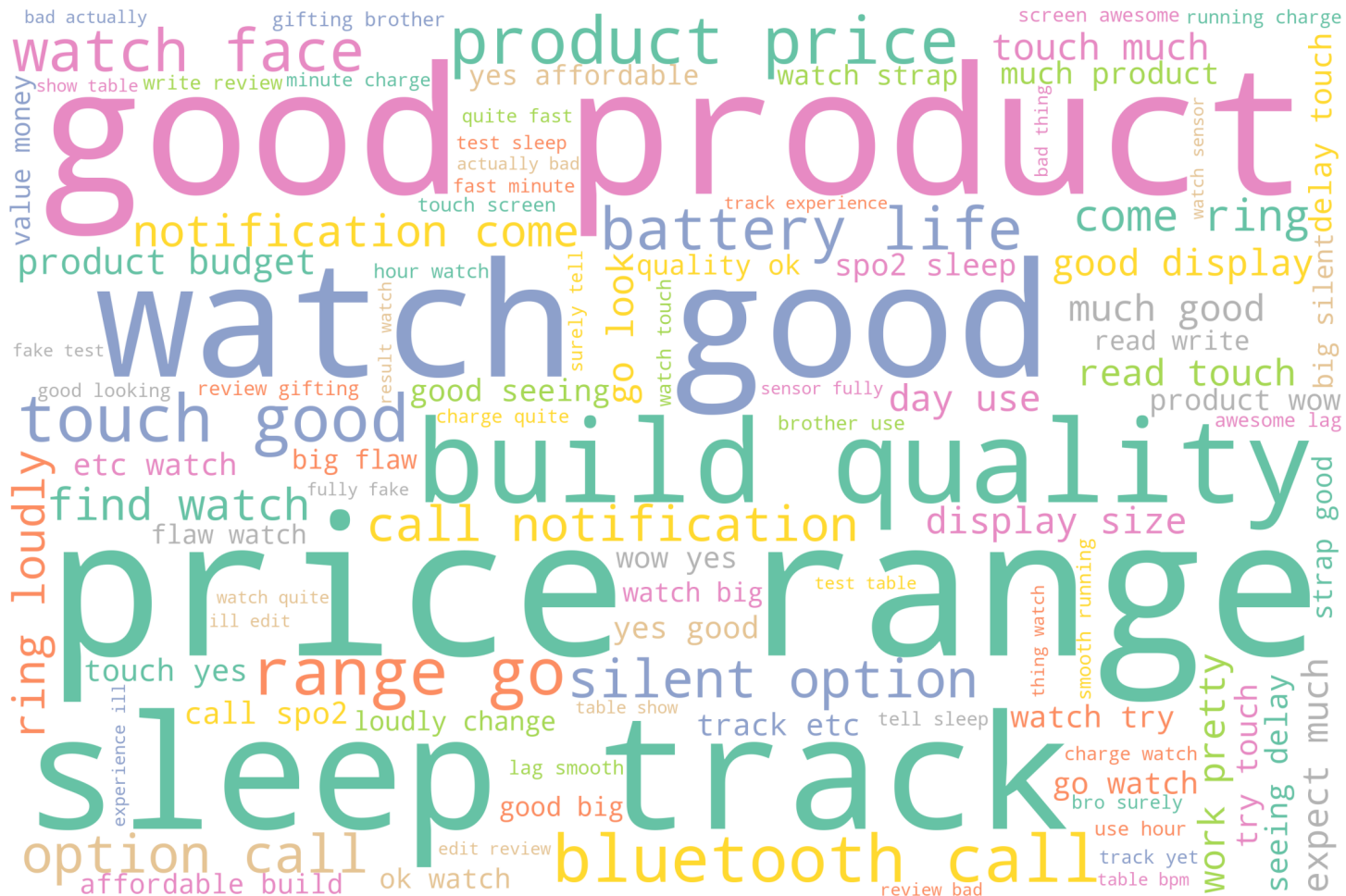
```
['15', '20', '2060', '23', '67days', '78', '8590', 'able', 'absolutely', 'access', 'accord', 'a
[[0. 0. 0. ... 1. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
```

Generate Word Cloud

```
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)
```



▼ Named Entity Recognition (NER)

```
# Parts of speech (POS) tagging
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)
```

write review gifting brother use 1 hour **TIME** watch touch screen awesome lag smooth running charge quite
fast 20 minute **TIME** 2060 **DATE** charge watch good looking have not test sleep track yet but bro surely tell
sleep track experience ill edit review bad actually bad thing watch sensor fully fake test table show table 78

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

sunlight NOUN
touch VERB
smooth ADJ
get VERB
scratch NOUN
screen NOUN
easily ADV
wakeup VERB
function NOUN
work NOUN
initially ADV
work VERB
pretty ADV
well ADV
resettingaccuracy NOUN
most ADJ
smartwatche NOUN
measure NOUN
ghost NOUN
read VERB
overall ADJ
accuracy NOUN
good ADJ
budgetbattery NOUN
average ADJ
battery NOUN
back ADV
67days NUM
battery NOUN
backup NOUN
23 NUM
day NOUN
use NOUN
call NOUN
function NOUN
moderately ADV
take VERB
15 NUM
hour NOUN
full ADJ
chargebluetooth NOUN
calling VERB
work NOUN
far ADV
well ADV
expectation NOUN
mic ADJ
speaker NOUN
good ADJ
talk NOUN
without ADP
background NOUN
disturbance NOUN
person NOUN

side NOUN
also ADV
hear VERB
clear ADJ

```
# Filtering the nouns and verbs only
nouns_verbs=[token.text for token in doc_block if token.pos_ in ('NOUN','VERB')]
print(nouns_verbs[100:200])
```

['hear', 'soundoverall', 'product', 'price', 'range', 'go', 'look', 'budget', 'dial', 'watch',

```
# Counting the noun & verb tokens
from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer()
```

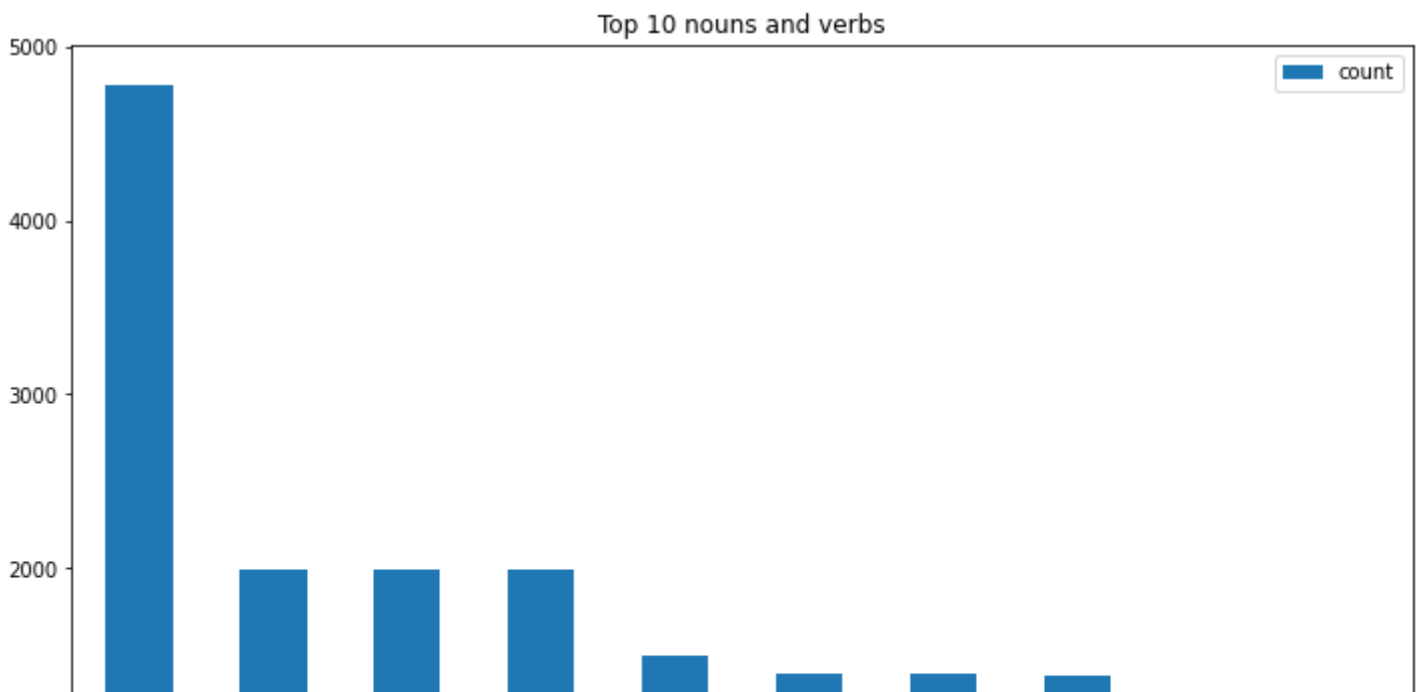
```
X=cv.fit_transform(nouns_verbs)
sum_words=X.sum(axis=0)
```

```
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
```

	word	count
0	watch	4776
1	touch	1990
2	product	1990
3	call	1990
4	read	1497
5	price	1393
6	range	1393
7	use	1380
8	screen	995
9	quality	995

```
# 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');
```



Emotion Mining - Sentiment Analysis

```
from nltk import tokenize
sentences=tokenize.sent_tokenize(' '.join(reviews))
sentences
```

```
['Writing this review after gifting if to my brother and using for 1 hour.',
 'Watch touch screen is awesome, no lag and smooth running.',
 'Charging is quite fast.',
 '20 minutes and 20-60% charge.',
 'Watch is good looking too.',
 'I haven't test the sleep tracking yet.",
 'But my bro will surely tell the sleep tracking experience and I'll edit this review.',
 'The only bad (actually worst) thing about this watch is that the sensors are fully
fake.',
 'I tested it on a table and it shows that the table has 78 bpm.',
 'If you wear it you will also get the same result.',
 'Watch is quite good, you could go for it if you are not looking for a heart beat, SpO2
sensing watch because they are worst.',
 'Read more I am writing this detailed review after using this product for a week:Build
Quality: It is metal body but not the pure metal material, overall build quality is
good.Screen: Amazing experience with screen, even it is not AMOLED display but you will not
feel the need as brightness is sufficient when wants to use in Sunlight.',
 'Touch is very smooth but you will get scratches on screen easily.',
 'Wakeup function was not working initially but worked pretty well after
resetting.Accuracy: Most of the Smartwatches measure ghost reading so this does but overall
accuracy is good in this budget.Battery: The average battery back up is 6-7days, the
battery backup will be 2-3 days if you are using calling function moderately.',
 'It takes 1.5 hour to full chargeBluetooth Calling: Works far better than my expectations,
the mic & speaker are good and you can talk without any background disturbance.',
 'The person on other side can also hear clear sound.Overall, this is good product in this
price range.',
 'Go for it if you are looking budget friendly round dial watch.',
 'Read more Touch is not much good, display size can be improvedI t's a good product in
budget.',
 'Don't expect too much from this product because in this price range you will get
Bluetooth calling, SpO2, SLEEP TRACKING, etc in this watch.',
 'In this price range you can go for this watch to try it.',
 'The touch is good not seeing delay when it touch and yes it's a good product not wow but
yes affordable.',
```

'Built quality is ok and watch strap is good.',
 "Too light you don't feel that you are wearing something in your wrist.",
 'And speaker is decent you can talk in the room but not so clear in public\n \nRead more
 First of all design,and built quality is excellent.',
 'Beautiful watch in this price range.Sleep tracking,touch screen, Voice is clear, Mic is
 also good.',
 'And all other features are good.Good battery life minimum 4 to 5 days while using theatre
 mode or battary save mood.',
 'And Bluetooth calling mood 3 to 4 days working.Customizable watch faces are missing, No
 watch face storage option, always download and apply.there is no silent option for call.',
 'When a call or notification comes it rings very loudly and it cannot be changed.',
 'And I am not satisfied favourite contact list only 8.',
 'It will Be good if fixed through software updates.',
 'Overall very nice product value for money.',
 "Read more Touch is not much good, display size can be improved it's a good product in
 budget.",
 "Don't expect too much from this product because in this price range you will get
 Bluetooth calling, SPO2, SLEEP TRACKING, etc in this watch.",
 'In this price range you can go for this watch to try it.',
 "The touch is good not seeing delay when it touch and yes it's a good product not wow but
 yes affordable.",
 'Built quality is ok and watch strap is good.',
 'Read more I have found the the watch is good but there are some big flaws in the watch.',
 'The biggest of them is there is no silent option for call.'

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

	sentence
0	Writing this review after gifting if to my bro...
1	Watch touch screen is awesome, no lag and smoo...
2	Charging is quite fast.
3	20 minutes and 20-60% charge.
4	Watch is good looking too.
...	...
11339	There are no lagging issues.
11340	The touch response is were good.
11341	Speaker is very loud.
11342	Except the wallpapers all other features are g...
11343	Read more

11344 rows × 1 columns

```
# Emotion Lexicon - Affin
affin=pd.read_csv('Afinn.csv',sep=',',encoding='Latin-1')
affin
```

	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

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

```
{'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,
 'abuse': -3,
 'abused': -3,
 'abuses': -3,
 'abusive': -3,
 'accept': 1,
 'accepted': 1,
 'accepting': 1,
 'accepts': 1,
 'accident': -2,
 'accidental': -2,
 'accidentally': -2,
 'accidents': -2,
 'accomplish': 2,
 'accomplished': 2,
 'accomplishes': 2,
 'accusation': -2,
 'accusations': -2,
 'accuse': -2,
 'accused': -2,
 'accuses': -2,
 'accusing': -2,
 'ache': -2,
```

```

'achievable': 1,
'aching': -2,
'acquit': 2,
'acquits': 2,
'acquitted': 2,
'acquitting': 2,
'acrimonious': -3,
'active': 1,
'adequate': 1,
'admire': 3,
'admired': 3,
'admires': 3,
'admiring': 3,
'admit': -1,
'admits': -1,
..

```

Custom function: score each word in a sentence in lemmatised form, but calculate the score for the

```

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

```

```

# manual testing
calculate_sentiment(text='good service')

```

```

3

```

```

# Calculating sentiment value for each sentence
sent_df['sentiment_value']=sent_df['sentence'].apply(calculate_sentiment)
sent_df['sentiment_value']

```

```

0      2
1      2
2      0
3      0
4      3
..
11339  -3
11340   3
11341   0
11342   3
11343   0
Name: sentiment_value, Length: 11344, dtype: int64

```

```

# how many words are there in a sentence?
sent_df['word_count']=sent_df['sentence'].str.split().apply(len)
sent_df['word_count']

```

```

0      14
1     10
2       4
3       5
4       5

```

```

..
11339      5
11340      6
11341      4
11342     12
11343      2
Name: word_count, Length: 11344, dtype: int64

```

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

	sentence	sentiment_value	word_count
176	The only bad (actually worst) thing about this...	-9	16
6499	The only bad (actually worst) thing about this...	-9	16
953	The only bad (actually worst) thing about this...	-9	16
6488	The only bad (actually worst) thing about this...	-9	16
942	The only bad (actually worst) thing about this...	-9	16
...
2414	The touch is good not seeing delay when it tou...	11	21
8798	The touch is good not seeing delay when it tou...	11	21
2371	The touch is good not seeing delay when it tou...	11	21
8658	The touch is good not seeing delay when it tou...	11	21
8413	The touch is good not seeing delay when it tou...	11	21

11344 rows × 3 columns

```

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

```

```

count      11344.000000
mean         2.192789
std          3.390812
min         -9.000000
25%          0.000000
50%          2.000000
75%          4.000000
max         11.000000
Name: sentiment_value, dtype: float64

```

```

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

```

	sentence	sentiment_value	word_count
2	Charging is quite fast.	0	4
3	20 minutes and 20-60% charge.	0	5
5	I haven't test the sleep tracking yet.	0	7
6	But my bro will surely tell the sleep tracking...	0	15
7	The only bad (actually worst) thing about this...	-9	16
...
11335	When a call or notification comes it rings ver...	0	15

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

	sentence	sentiment_value	word_count
0	Writing this review after gifting if to my bro...	2	14
1	Watch touch screen is awesome, no lag and smoo...	2	10
4	Watch is good looking too.	3	5
11	Read more I am writing this detailed review af...	9	58
13	Wakeup function was not working initially but ...	3	51
...
11333	I have found the the watch is good but there a...	4	17
11336	I am not able to add personalized wallpaper as...	2	21
11338	All the sensor seem to be working pretty good ...	4	12
11340	The touch response is were good.	3	6
11342	Except the wallpapers all other features are g...	3	12

6567 rows × 3 columns

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

	sentence	sentiment_value	word_count	index
0	Writing this review after gifting if to my bro...	2	14	0
1	Watch touch screen is awesome, no lag and smoo...	2	10	1
2	Charging is quite fast.	0	4	2
3	88 minutes and 88.88% battery	0	5	3

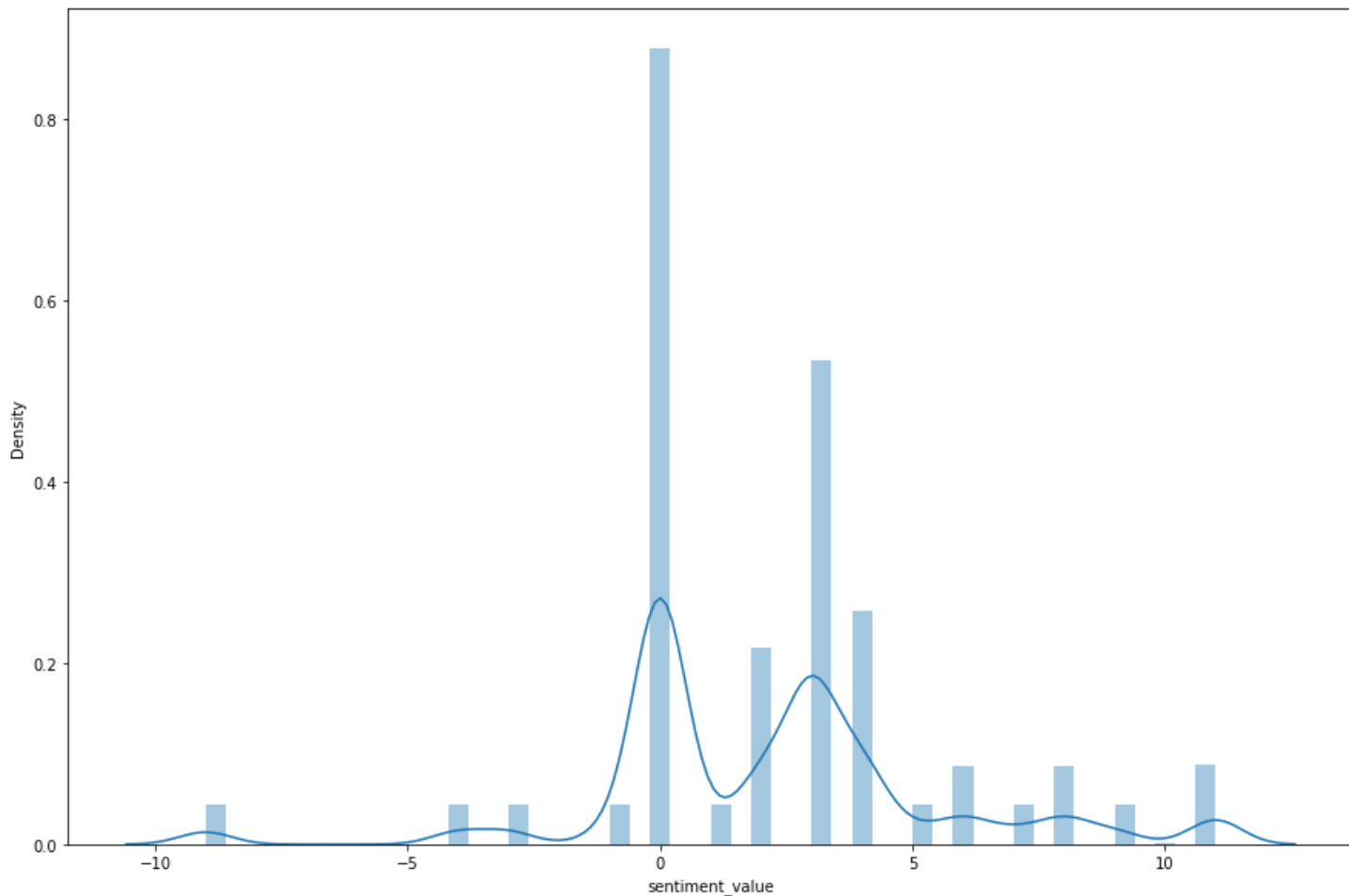
```
# Plotting the sentiment value for whole review
```

```
import seaborn as sns
```

```
plt.figure(figsize=(15,10))
```

```
sns.distplot(sent_df['sentiment_value'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot`
warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f194d1628d0>
```



```
# Plotting the line plot for sentiment value of whole review
```

```
plt.figure(figsize=(15,10))
```

```
sns.lineplot(y='sentiment_value',x='index',data=sent_df)
```


<matplotlib.axes._subplots.AxesSubplot at 0x7f1945a2c710>

