For Text Mining assignment

ONE:

1) Perform sentimental analysis on the Elon-musk tweets (Exlon-musk.csv)

In [1]:
```
!pip install spacy
```

```
Requirement already satisfied: spacy in c:\users\admin\anaconda3\lib\site-packages (2.3.5)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in c:\users\admin\anaconda3\lib\site-packages (from spac
y) (0.7.4)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\admin\anaconda3\lib\site-packages (from spac
y) (2.0.5)
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in c:\users\admin\anaconda3\lib\site-packages (from spac
y) (1.0.5)
Requirement already satisfied: numpy>=1.15.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy) (1.
19.2)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in c:\users\admin\anaconda3\lib\site-packages (from spac
y) (1.1.3)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\admin\anaconda3\lib\site-packages (fro
m spacy) (1.0.5)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in c:\users\admin\anaconda3\lib\site-packages (from
spacy) (1.0.0)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\admin\anaconda3\lib\site-packages (from spac
y) (4.50.2)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\admin\anaconda3\lib\site-packages (from sp
acy) (3.0.5)
Requirement already satisfied: thinc<7.5.0,>=7.4.1 in c:\users\admin\anaconda3\lib\site-packages (from spac
y) (7.4.5)
Requirement already satisfied: setuptools in c:\users\admin\anaconda3\lib\site-packages (from spacy) (50.3.
1.post20201107)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in c:\users\admin\anaconda3\lib\site-packages (from spa
cy) (0.8.1)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\admin\anaconda3\lib\site-packages (from
spacy) (2.24.0)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in c:\users\admin\anaconda3\lib\site
-packages (from requests<3.0.0,>=2.13.0->spacy) (1.25.11)
Requirement already satisfied: chardet<4,>=3.0.2 in c:\users\admin\anaconda3\lib\site-packages (from reques
ts<3.0.0,>=2.13.0->spacy) (3.0.4)
Requirement already satisfied: idna<3,>=2.5 in c:\users\admin\anaconda3\lib\site-packages (from requests<3.
0.0,>=2.13.0->spacy) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\admin\anaconda3\lib\site-packages (from reque
sts<3.0.0,>=2.13.0->spacy) (2020.6.20)
```

In [3]:
```
!pip install WordCloud
```

```
Collecting WordCloud
  Downloading wordcloud-1.8.1-cp38-cp38-win_amd64.whl (155 kB)
Requirement already satisfied: matplotlib in c:\users\admin\anaconda3\lib\site-packages (from WordCloud)
(3.3.2)
Requirement already satisfied: numpy>=1.6.1 in c:\users\admin\anaconda3\lib\site-packages (from WordCloud)
(1.19.2)
Requirement already satisfied: pillow in c:\users\admin\anaconda3\lib\site-packages (from WordCloud) (8.0.
1)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\admin\anaconda3\lib\site-packages (from matplo
tlib->WordCloud) (1.3.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in c:\users\admin\anaconda3\lib\sit
e-packages (from matplotlib->WordCloud) (2.4.7)
Requirement already satisfied: certifi>=2020.06.20 in c:\users\admin\anaconda3\lib\site-packages (from matp
lotlib->WordCloud) (2020.6.20)
Requirement already satisfied: cycler>=0.10 in c:\users\admin\anaconda3\lib\site-packages (from matplotlib-
>WordCloud) (0.10.0)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\admin\anaconda3\lib\site-packages (from mat
plotlib->WordCloud) (2.8.1)
Requirement already satisfied: six in c:\users\admin\anaconda3\lib\site-packages (from cycler>=0.10->matplo
tlib->WordCloud) (1.15.0)
Installing collected packages: WordCloud
Successfully installed WordCloud-1.8.1
```

In [16]:
```
# importing libraries
import pandas as pd
import numpy as np
import string

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

import spacy
import nltk

%matplotlib inline
```

In [8]:
```
tweets_data = pd.read_csv('Elon_musk.csv', encoding = 'latin1', error_bad_lines=False)
tweets_data.head()
```

Out[8]:

	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...

In [9]:
```
tweets_data.shape
```

Out[9]: (1999, 2)

In [11]:
```
tweets_data = tweets_data.drop('Unnamed: 0', axis = 1)
tweets_data.head()
```

Out[11]:

	Text
0	@kunalb11 I m an alien
1	@ID_AA_Carmack Ray tracing on Cyberpunk with H...
2	@joerogan @Spotify Great interview!
3	@gtera27 Doge is underestimated
4	@teslacn Congratulations Tesla China for amazi...

In [12]:
```
# removing both the leading and the trailing characters such as spaces in tweets
tweets_data = [x.strip() for x in tweets_data.Text]
```

In [14]:
```
tweets_data[0:5]
```

Out[14]:
```
['@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 amazing execution last year. Now on to the next for even mor
e!!']
```

In [17]:
```
nltk.download('punkt')
```

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

Out[17]: True

In [19]:
```
from nltk import tokenize
sentences = tokenize.sent_tokenize(" ".join(tweets_data))
sentences[0:5]
```

Out[19]:
```
['@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 amazing execution last year.',
 'Now on to the next for even more!!']
```

In [54]:
```
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords
wordnet = WordNetLemmatizer()
import re

filtered_sent=[]
for i in range(len(sentences)):
    review = re.sub("[^A-Za-z" "]+"," ",sentences[i])
    review = re.sub("[0-9" "]+"," ",sentences[i])

    review = review.lower()
    review =review.split()
    review = [wordnet.lemmatize(word) for word in review if not word in set(stopwords.words('english'))]
    review = ' '.join(review)
    filtered_sent.append(review)
```

In [56]:
```
filtered_sent[0:5]
```

Out[56]:
```
['@kunalb i\x92m alien @id_aa_carmack ray tracing cyberpunk hdr next-level.',
 'tried it?',
 '@joerogan @spotify great interview!',
 '@gtera doge underestimated @teslacn congratulation tesla china amazing execution last year.',
 'next even more!!']
```

In [57]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
tf = TfidfVectorizer()
text_tf = tf.fit_transform(filtered_sent)
feature_names = tf.get_feature_names()
dense = text_tf.todense()
denselist = dense.tolist()
sentences_df =pd.DataFrame(denselist, columns=feature_names)
sentences_df.head()
```

Out[57]:

	bennettm	_extrachars	_rykllan	_t	_teslatom	a_howard	aayypcpp	ab	aber	able	...	zugaquvpy	zvm	zwiebelbach	zxd
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
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
2	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	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
4	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

5 rows × 5313 columns

In [98]:
```
sent_df = pd.DataFrame(sentences, columns=['sentence'])
sent_df.head()
```

Out[98]:

	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!!

In [47]:
```
with open("positive-words.txt","r") as pos:
    poswords = pos.read().split("\n")

with open("negative-words.txt","r") as neg:
    negwords = neg.read().split("\n")
```

In [48]:
```
pos_words = poswords[35:]
pos_words
```

Out[48]:
```
['a+',
 'abound',
 'abounds',
 'abundance',
 'abundant',
 'accessable',
 'accessible',
 'acclaim',
 'acclaimed',
 'acclamation',
 'accolade',
 'accolades',
 'accommodative',
 'accomodative',
 'accomplish',
 'accomplished',
 'accomplishment',
 'accomplishments',
 'accurate',
```

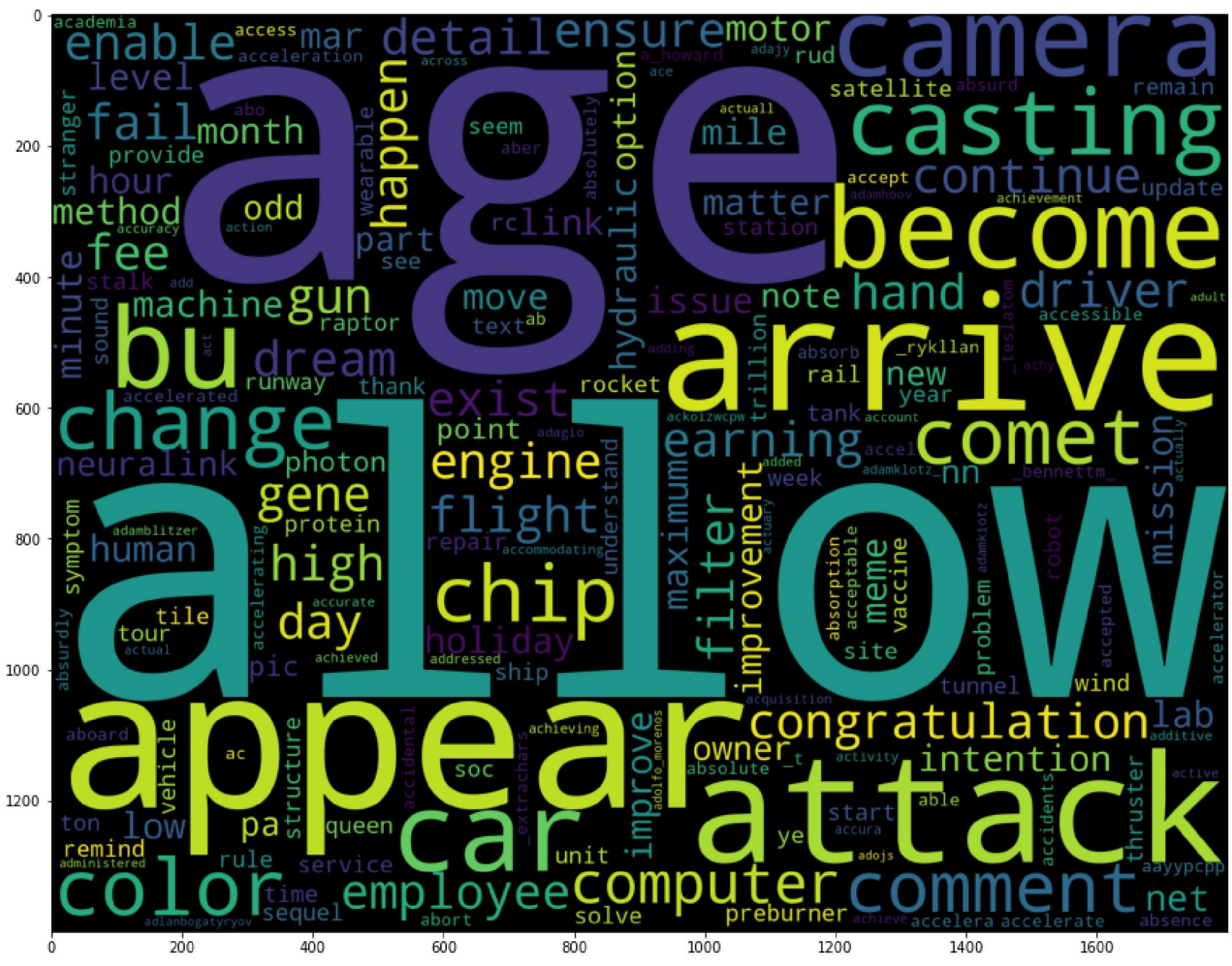
In [53]:
```
neg_words = negwords[35:]
neg_words
```

Out[53]:
```
['2-faced',
 '2-faces',
 'abnormal',
 'abolish',
 'abominable',
 'abominably',
 'abominate',
 'abomination',
 'abort',
 'aborted',
 'aborts',
 'abrade',
 'abrasive',
 'abrupt',
 'abruptly',
 'abscond',
 'absence',
 'absent-minded',
 'absentee',
```

In [90]:
```
#plotting wordcloud on TFIDF
from wordcloud import WordCloud
import matplotlib.pyplot as plt
cloud = ' '.join(sentences_df)
#cloud
```

In [74]:
```
f, axes = plt.subplots(figsize=(20,12))
wordcloud= WordCloud(
        background_color = 'black',
        width = 1800,
        height =1400).generate(cloud)
plt.imshow(wordcloud)
```

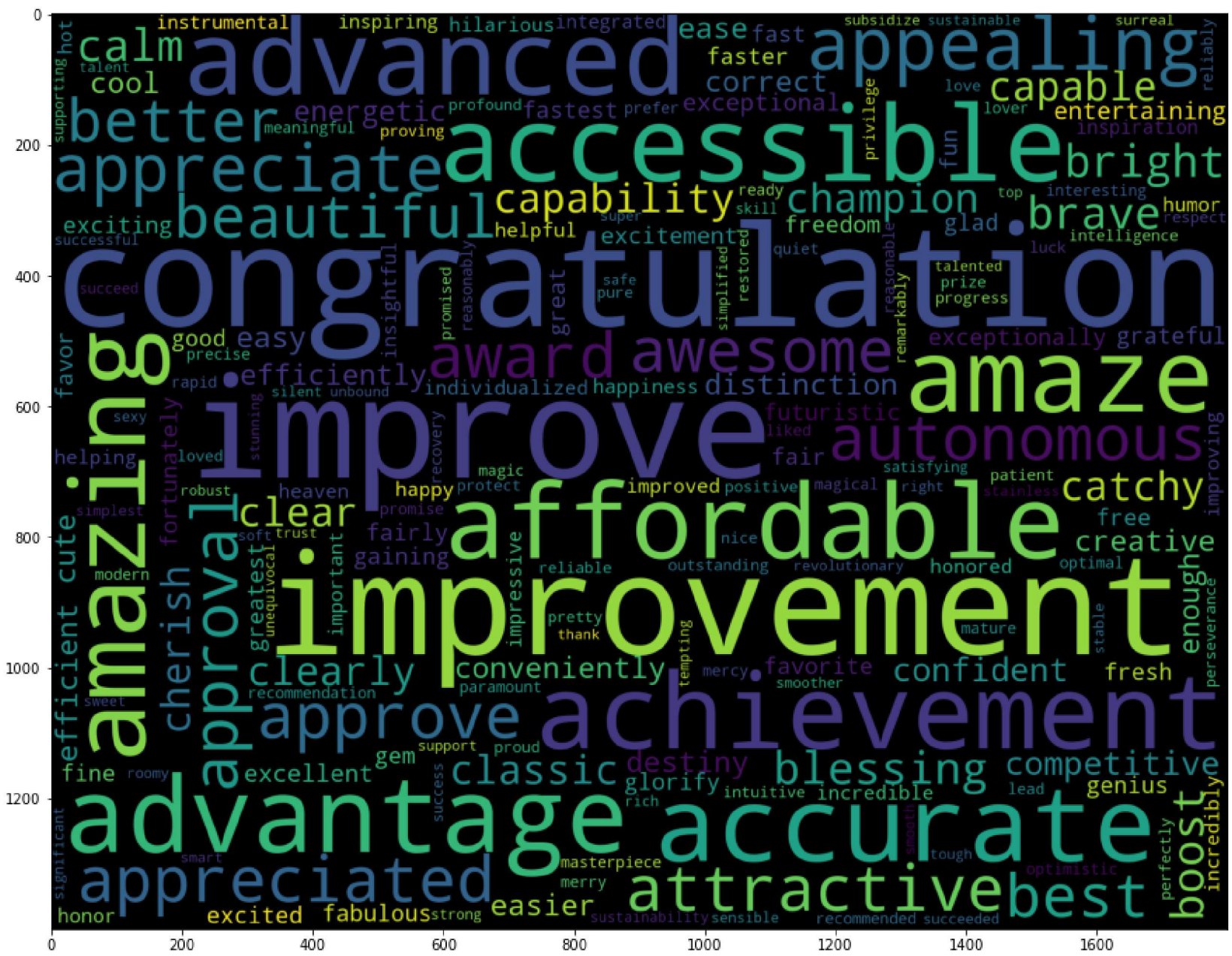
Out[74]: <matplotlib.image.AxesImage at 0x2197bd00310>



In [72]:
```
f, axes = plt.subplots(figsize=(20,12))
pos_words = ' '.join([w for w in sentences_df if w in poswords])

cloud_pos = WordCloud(
        background_color = 'black',
        width =1800,
        height=1400).generate(pos_words)
plt.imshow(cloud_pos)
```

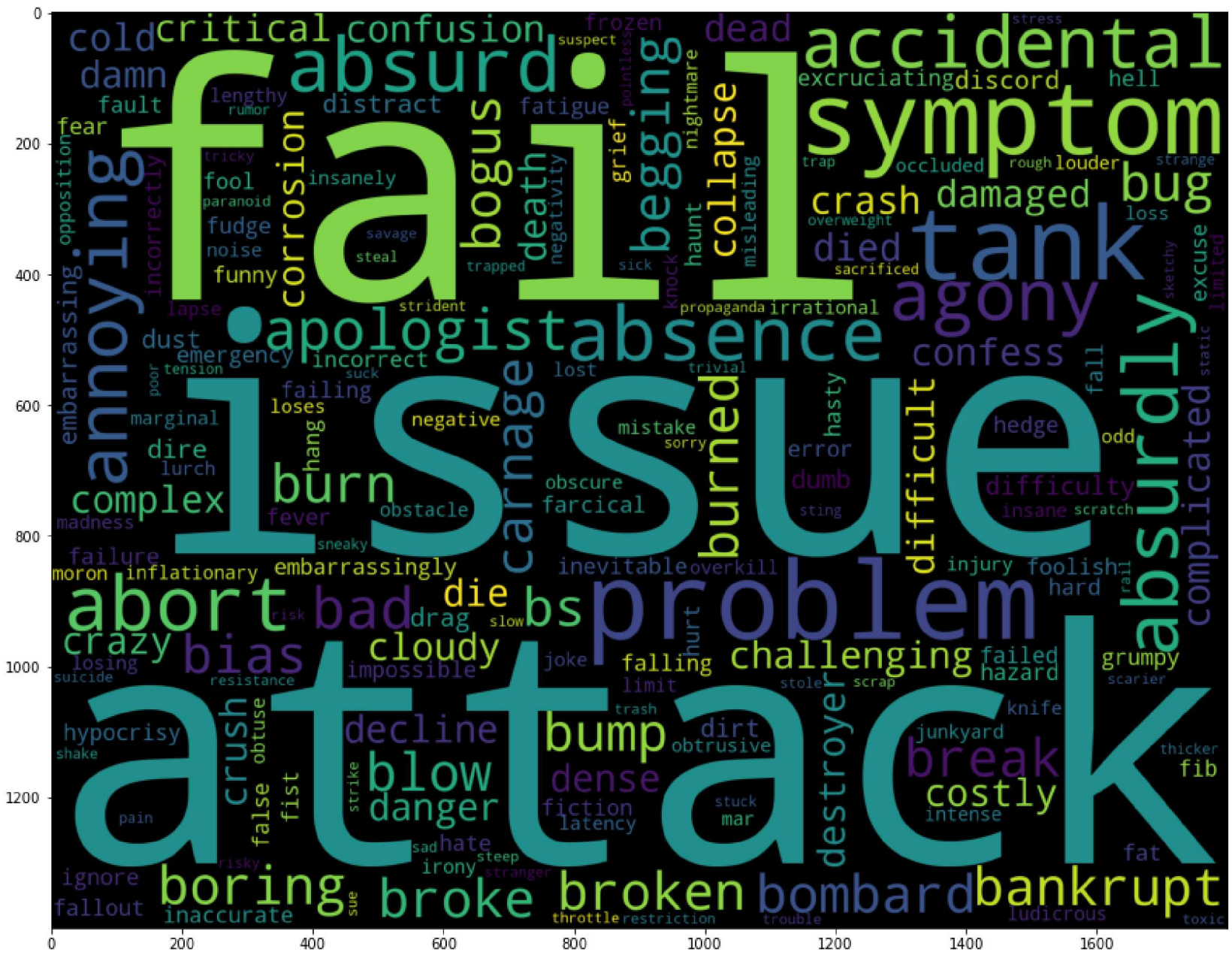
Out[72]: <matplotlib.figure.AxesImage at 0x219159bac70>



In [73]:
```
f, axes = plt.subplots(figsize=(20,12))
neg_words = ' '.join([w for w in sentences_df if w in negwords])

cloud_neg = WordCloud(
        background_color='black',
        width =1800,
        height =1400).generate(neg_words)
plt.imshow(cloud_neg)
```

Out[73]: <matplotlib.figure.AxesImage at 0x2191622f880>



Affinity scores:

In [77]:
```
with open("afinn2.txt","r") as affin:
    affinity = affin.read().split("\n")
```

In [81]:
```
affinity_data = pd.read_csv('afinn2.txt', sep="\t", header=None, names=["word", "value"])
affinity_data.head()
```

Out[81]:

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

In [85]:
```
affinity_scores = affinity_data.set_index('word')['value'].to_dict()
#affinity_scores
```

In [86]:
```
!python -m spacy download en_core_web_sm
```

```
Requirement already satisfied: en_core_web_sm==2.3.1 from https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-2.3.1/en_core_web_sm-2.3.1.tar.gz#egg=en_core_web_sm==2.3.1 (https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-2.3.1/en_core_web_sm-2.3.1.tar.gz#egg=en_core_web_sm==2.3.1) in c:\users\admin\anaconda3\lib\site-packages (2.3.1)
Requirement already satisfied: spacy<2.4.0,>=2.3.0 in c:\users\admin\anaconda3\lib\site-packages (from en_core_web_sm==2.3.1) (2.3.5)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (4.50.2)
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (1.0.5)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (1.0.0)
Requirement already satisfied: setuptools in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (50.3.1.post20201107)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (0.7.4)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (1.1.3)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (2.0.5)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (0.8.1)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (3.0.5)
Requirement already satisfied: thinc<7.5.0,>=7.4.1 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (7.4.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (1.0.5)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (2.24.0)
Requirement already satisfied: numpy>=1.15.0 in c:\users\admin\anaconda3\lib\site-packages (from spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (1.19.2)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in c:\users\admin\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (1.25.11)
Requirement already satisfied: idna<3,>=2.5 in c:\users\admin\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (2.10)
Requirement already satisfied: chardet<4,>=3.0.2 in c:\users\admin\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\admin\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy<2.4.0,>=2.3.0->en_core_web_sm==2.3.1) (2020.6.20)
[+] Download and installation successful
You can now load the model via spacy.load('en_core_web_sm')
```

In [87]:
```
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 [88]:
```
calculate_sentiment(text = 'good')
```

Out[88]: 3

In [92]:
```
sentences_df
```

Out[92]:

	bennettm	_extrachars	_rykllan	_t	_teslatom	a_howard	aayypcpp	ab	aber	able	...	zugaquvpy	zvm	zwiebelbach	zxc
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
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
2	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	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
4	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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
919	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
920	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
921	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
922	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
923	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

924 rows × 5313 columns

In [101]:
```
sent_df['sentiment_value'] = sent_df['sentence'].apply(calculate_sentiment)
sent_df.head()
```

Out[101]:

	sentence	sentiment_value
0	@kunalb11 I m an alien @ID_AA_Carmack Ray trac...	0
1	Have you tried it?	0
2	@joerogan @Spotify Great interview!	3
3	@gtera27 Doge is underestimated @teslacn Congr...	3
4	Now on to the next for even more!!	0

In [104]:
```
# how many words are in the sentence?
sent_df['word_count'] = sent_df['sentence'].str.split().apply(len)
sent_df.head()
```

Out[104]:

	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
2	@joerogan @Spotify Great interview!	3	4
3	@gtera27 Doge is underestimated @teslacn Congr...	3	13
4	Now on to the next for even more!!	0	8

In [106]:
```
sent_df.sort_values(by='sentiment_value')
sent_df.head()
```

Out[106]:

	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
2	@joerogan @Spotify Great interview!	3	4
3	@gtera27 Doge is underestimated @teslacn Congr...	3	13
4	Now on to the next for even more!!	0	8

In [107]:
```
sent_df['sentiment_value'].max()
```

Out[107]: 17

In [113]:
```
# tweet which is having max sentiment value
sent_df[sent_df['sentiment_value']==17]
```

Out[113]:

	sentence	sentiment_value	word_count
36	@ajtourville @Erdayastronaut @SpaceX Yes, but ...	17	231

In [114]:
```
# Full tweet at index 36
sent_df['sentence'][36]
```

Out[114]:
```
'@ajtourville @Erdayastronaut @SpaceX Yes, but engines have a min throttle point where there is flameout ri
sk, so la\x85 https://t.co/ThniYsSnWn (https://t.co/ThniYsSnWn) @Adamklotz_ @Erdayastronaut @SpaceX Yes @Er
dayastronaut @SpaceX It was foolish of us not to start 3 engines &amp;amp; immediately shut down 1, as 2 are ne
eded to land I am become meme,\nDestroyer of shorts @commander_cruz Might actually happen Sandstorm is a ma
sterpiece @MadOverlord We were too dumb Next time, we try pull *up* method https://t.co/J1M7WHLRDR (http
s://t.co/J1M7WHLRDR) Low-key Loki No highs, no lows, only Doge No need to be a gigachad to own Dogecoin is
 the people\x92s crypto ur welcome https://t.co/e2KF57KLxb (https://t.co/e2KF57KLxb) @HuobiGlobal The most
 entertaining outcome is the most likely Doge https://t.co/vviUzWhodT (https://t.co/vviUzWhodT) @Erdayastro
naut High seas &amp;amp; wind This was a tough one https://t.co/beY7peFpCj (https://t.co/beY7peFpCj) RT @Space
X: Watch Falcon 9 launch 60 Starlink satellites <U+2192> https://t.co/bJFjLCzWdK (https://t.co/bJFjLCzWdK)
https://t.co/Ln95rEesbW (https://t.co/Ln95rEesbW) Off Twitter for a while It will be filled with graffiti a
rt @eiraum <U+043C><U+044B> <U+0441><U+0434><U+0435><U+043B><U+0430><U+0435><U+043C> Giga Berlin progress h
ttps://t.co/ekpG5qcbUi (https://t.co/ekpG5qcbUi) @hamoon__ @neuralink Neuralink is working super hard to en
sure implant safety &amp;amp; is in close communication with the\x85 https://t.co/Yky1llPUMd (https://t.co/Yky1
llPUMd) RT @SpaceX: This mission enables access for everyday people who dream of going to space RT @SpaceX:
Announcing the first commercial astronaut mission to orbit Earth aboard Dragon <U+2192> https://t.co/MbESvn
akAD (https://t.co/MbESvnakAD) https://t.co/ukLsjFfRjk (https://t.co/ukLsjFfRjk) If you\x92ve worked on adv
anced wearables, phones or robots, those skills are needed @neuralink Feels weird helping make (hopefully g
ood version of) Cyberpunk come true Please consider working at Neuralink!'
```

In [115]:
```
# minimum sentiment value
sent_df['sentiment_value'].min()
```

Out[115]:
```
-8
```

In [116]:
```
# tweet which is having min sentiment value
sent_df[sent_df['sentiment_value']==-8]
```

Out[116]:

	sentence	sentiment_value	word_count
647	Very ba https://t.co/tJsh1Exz1Q @justpaulinel...	-8	60

In [118]:
```
sent_df["sentence"][647]
```

Out[118]:
```
'Very ba\x85 https://t.co/tJsh1Exz1Q (https://t.co/tJsh1Exz1Q) @justpaulinelol It\x92s too embarrassing Jus
t learned that my physics prof from Penn is head of Fermilab &amp;amp; still has some of my old homework <U+263
A><U+FE0F> @ruskin147 It is unfortunately common for many in academia to overweight the value of ideas &am
p; underweight bringing\x85 https://t.co/OFFoojCqNv (https://t.co/OFFoojCqNv) @l_vaux @kulpability @tlowdon
@EthicalSkeptic People are extremely unlikely to die of Covid/sars-cov2 alone.'
```

In [119]:
```
sent_df['sentiment_value'].describe()
```

Out[119]:
```
count    924.000000
mean       1.524892
std        2.819301
min       -8.000000
25%        0.000000
50%        0.000000
75%        3.000000
max       17.000000
Name: sentiment_value, dtype: float64
```

In [120]:
```
sent_df[sent_df["sentiment_value"]<0]
```

Out[120]:

	sentence	sentiment_value	word_count
6	https://t.co/9WFKMYu2oj Frodo was the underdog...	-2	14
11	However, if we don https://t.co/9Z8Ac6skqx @R...	-2	40
16	That s what you re missing.	-2	4
23	https://t.co/UEEocOfcTb The people have spoken...	-1	37
47	Would be better to have small fees than no fee...	-2	48
...	...	...	...
886	This is primarily an electrical/electronic (ch...	-3	69
887	Will those who write the algorithms ever reali...	-3	11
894	No expert installer required.	-1	4
920	It may never pay out, as the stock can t b ht...	-4	31
923	@TeslaGong @PPathole Samwise Gamgee @PPathole ...	-1	15

107 rows × 3 columns

In [121]:
```
sent_df[sent_df['sentiment_value']>=10]
```

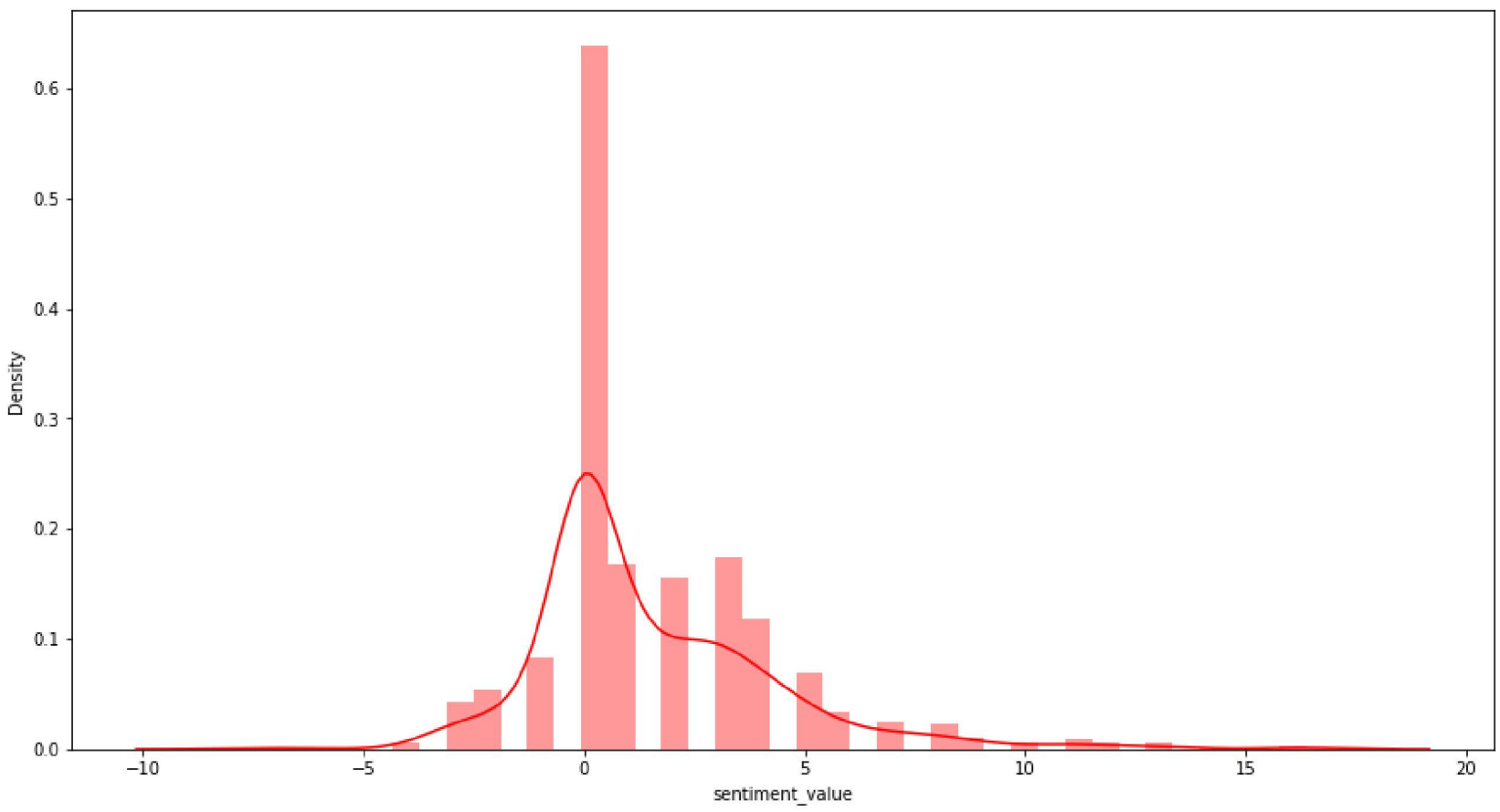
Out[121]:

	sentence	sentiment_value	word_count
36	@ajtourville @Erdayastronaut @SpaceX Yes, but ...	17	231
42	@thejackbeyer @NASASpaceflight Cryoproof, then...	12	84
81	@teslaownersSV This is a good one @MrBeastYT I...	13	38
102	@BoredElonMusk <U+0001F923><U+0001F923> @erigg...	11	34
105	@Erdayastronaut @SpaceX Was also thinking that...	16	94
159	@harsimranbansal @heydave7 It will most likely...	11	93
183	There s some great AI out ther https://t.co/v...	10	48
184	Also, w https://t.co/IYdKttVsqk @cleantechnic...	12	98
224	@PPathole @karpathy For a second there, I thou...	11	69
258	@TheOnion @ID_AA_Carmack Yup @flcnhvy Change o...	11	57
490	Gives a false sense of https://t.co/aIGaEnC8f...	10	41
535	Have you tried pushing on the c https://t.co/...	10	41
569	@WholeMarsBlog Haha true @oneandonlyheady Mayb...	13	68
585	The open areas https://t.co/rabjKrtQlw @Sav...	16	138
645	It s https://t.co/IxhyT6NuiC @Teslarati Tesla...	13	65
652	It s not useful for predicting the future, nor...	12	118
719	We just haven t observed the https://t.co/mez...	14	72
817	Might be able to https://t.co/FB7e5nOq2E @Pri...	11	72

In [122]:
```
sent_df['index']=range(0,len(sent_df))
```

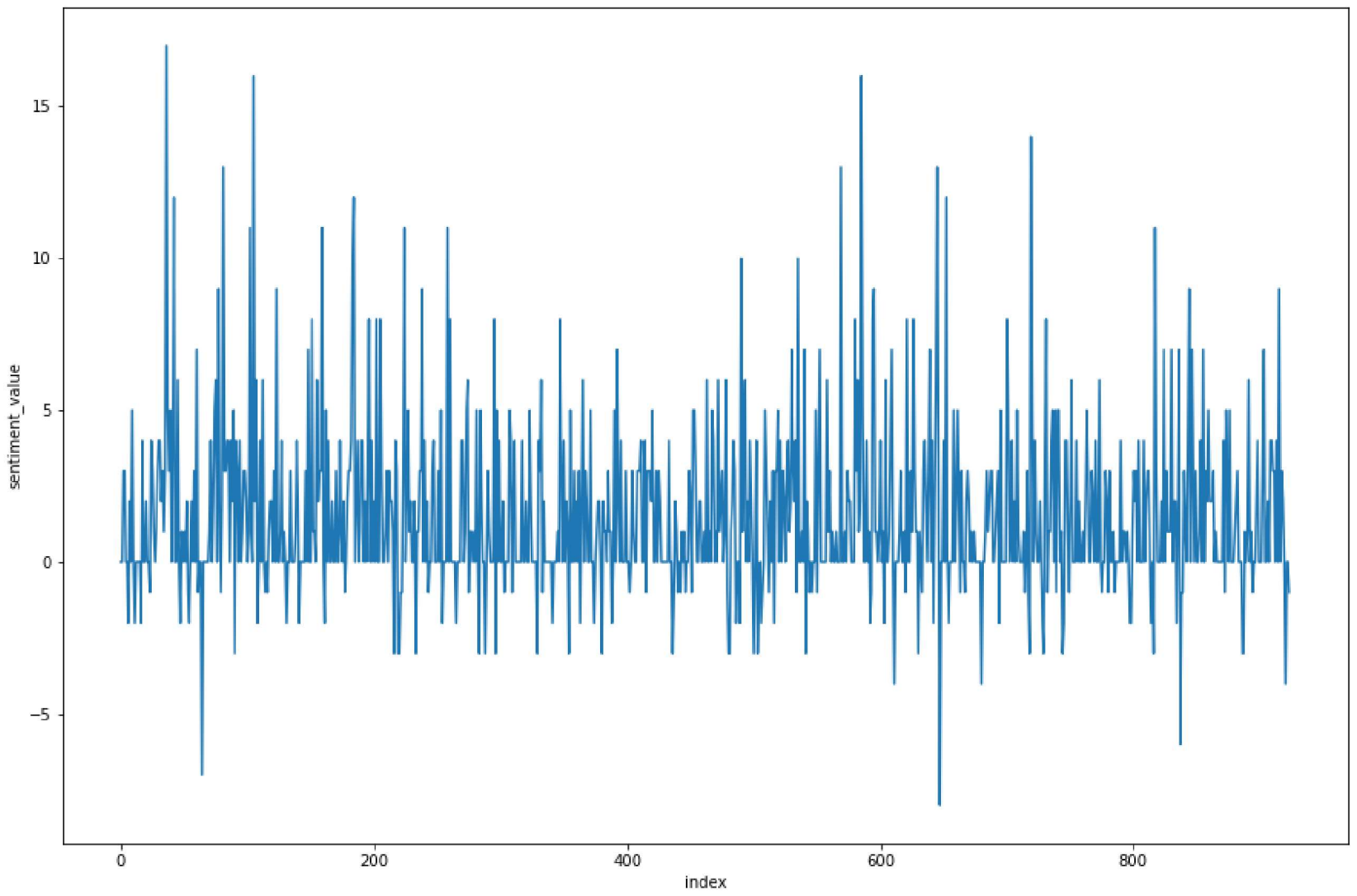
In [128]:
```
f, axes = plt.subplots(figsize= (15,8))
import seaborn as sns
sns.distplot(sent_df['sentiment_value'],color = "red")
plt.show()
```

```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\distributions.py:2551: FutureWarning: `distplot` is a de
precated 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)
```



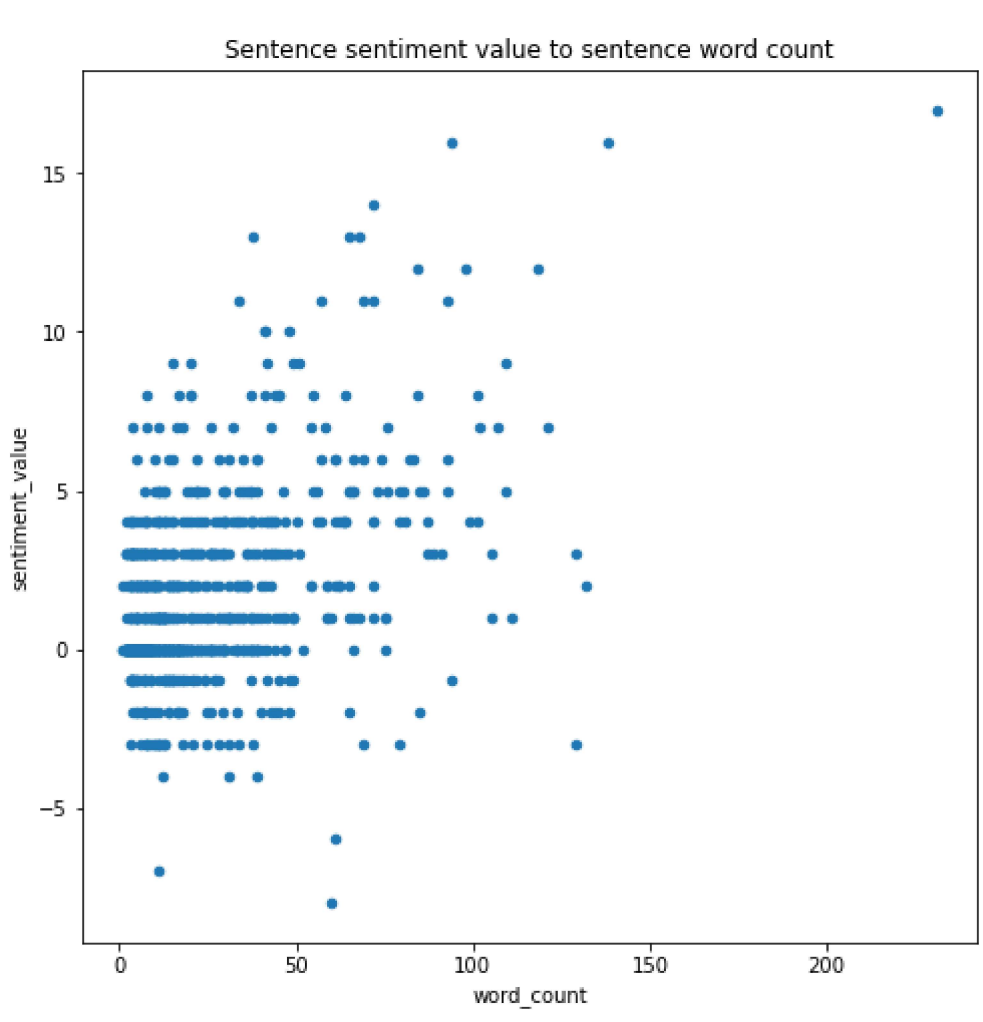
In [129]:
```
plt.figure(figsize=(15, 10))
sns.lineplot(y='sentiment_value',x='index',data=sent_df)
```

Out[129]: <AxesSubplot:xlabel='index', ylabel='sentiment_value'>



In [130]:
```
df.plot.scatter(x='word_count', y='sentiment_value', figsize=(8,8), title='Sentence sentiment value to senten
```

Out[130]: <AxesSubplot:title={'center':'Sentence sentiment value to sentence word count'}, xlabel='word_count', ylabel='sentiment_value'>



In [132]:
```
sent_df['Sentiment_Class'] = pd.cut(x=sent_df['sentiment_value'],bins=[-8, -1, 0, 17],
                                    labels=['Negative','Neutral','Positive'], right = True)
sent_df.head()
```

Out[132]:

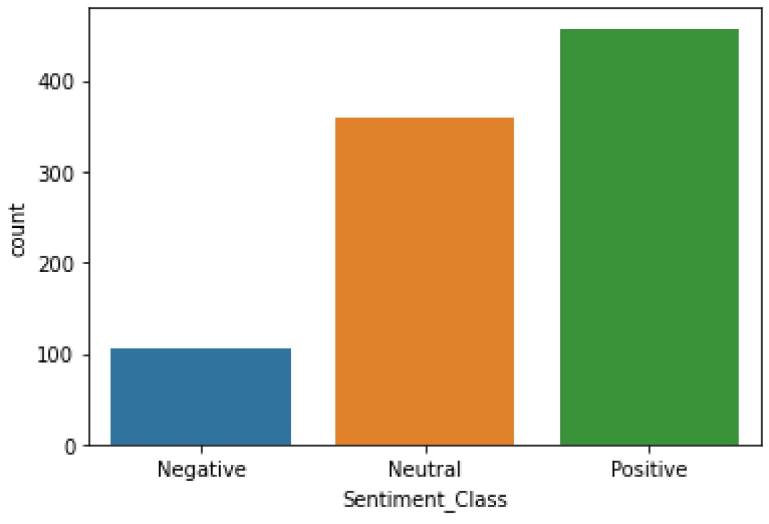
	sentence	sentiment_value	word_count	index	Sentiment_Class
0	@kunalb11 I m an alien @ID_AA_Carmack Ray trac...	0	13	0	Neutral
1	Have you tried it?	0	4	1	Neutral
2	@joerogan @Spotify Great interview!	3	4	2	Positive
3	@gtera27 Doge is underestimated @teslacn Congr...	3	13	3	Positive
4	Now on to the next for even more!!	0	8	4	Neutral

In [134]:
```
sent_df['Sentiment_Class'].value_counts()
```

Out[134]:
```
Positive    457
Neutral     360
Negative    106
Name: Sentiment_Class, dtype: int64
```

In [133]:
```
sns.countplot(x = 'Sentiment_Class', data = sent_df)
```

Out[133]: <AxesSubplot:xlabel='Sentiment_Class', ylabel='count'>



In [ ]: