

▼ Text Mining on Elon Musk



▼ After seeing this Elon Musk be like :

"Looks like my tweets are finally getting the attention they deserve. I always knew my words had power, but I didn't know they could be analyzed like this. Maybe next time I'll tweet in code to keep these text miners on their toes. 🤦"

```
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

# load the dataset
tweets=pd.read_csv('Elon_musk.csv',encoding='Latin-1')
tweets.drop(['Unnamed: 0'],inplace=True, axis=1)
tweets
```

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...
...	...
1994	@flcnhv True, it sounds so surreal, but the n...
1995	@PPathole Make sure to read ur terms & con...
1996	@TeslaGong @PPathole Samwise Gamgee
1997	@PPathole Altho Dumb and Dumber is <U+0001F525...
1998	Progress update August 28

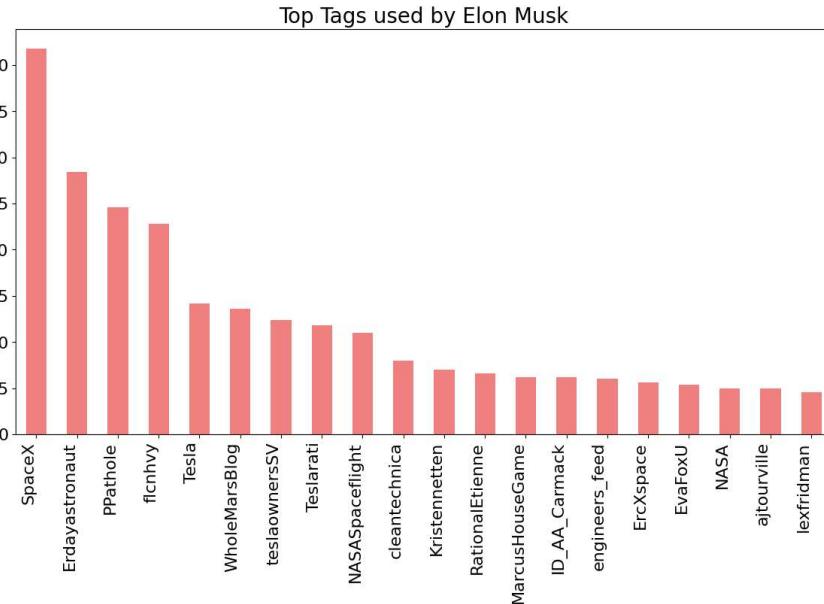
1999 rows × 1 columns

▼ Data Exploration

```
import itertools

c = list(itertools.chain(*tweets.Text.map(lambda t: [handle.replace(":", "")[1:] for handle in t.split(" ") if '@' in handle.replace(":", "")]).tolist()))

pd.Series(c).value_counts().head(20).plot.bar(figsize=(14, 7), fontsize=16, color='lightcoral')
plt.gca().set_title('Top Tags used by Elon Musk', fontsize=20)
plt.gca().set_xticklabels(plt.gca().get_xticklabels(), rotation=90, ha='right', fontsize=16)
pass
```



▼ Cleaning the text by removing irrelevant information

```
tweets=[Text.strip() for Text in tweets.Text] # remove both the leading and the trailing characters
tweets=[Text for Text in tweets if Text] # removes empty strings, because they are considered in Python as False
tweets[0:10]

['@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!!',
 'Happy New Year of the Ox! https://t.co/9WFKMYu2oj',
 'Frodo was the underdoge,\nAll thought he would fail,\nHimself most of all. https://t.co/zGxJFDzzrM' ,
 '@OwenSparks_ @flcnhv @anonyx10 Haha thanks :)',
 '@flcnhv @anonyx10 Indeed! Tweets definitely do not represent real-world time allocation.',
 'The most entertaining outcome is the most likely']

# Joining the list into one string/text
tweets_text=' '.join(tweets)
tweets_text

 '@kunalb11 I\x92m an alien @ID_AA_Carmack Ray tracing on Cyberpunk with HDR is next-level. Have you tried it? @joerogan @Spotify G
reat interview! @gtera27 Doge is underestimated @teslacn Congratulations Tesla China for amazing execution last year. Now on to th
e next for even more!! Happy New Year of the Ox! https://t.co/9WFKMYu2oj Frodo was the underdoge,\nAll thought he would fail,\nHim
self most of all. https://t.co/zGxJFDzzrM @OwenSparks_ @flcnhv @anonyx10 Haha thanks :) @flcnhv @anonyx10 Indeed! Tweets defin
itely do not represent real-world time allocation. The most entertaining outcome is the most likely @GiveDirectly Just sent some Jus
t agree to do Clubhouse with @kanwest https://t.co/3rWF9uHSTS @enoffkeishlev @InrealEngine Tt\x92s getting real Bought some Doge
```

```
# remove Twitter username handles from a given twitter text. (Removes @usernames)
from nltk.tokenize import TweetTokenizer
tknizer = TweetTokenizer(strip_handles=True)
tweets_tokens=tknizer.tokenize(tweets_text)
print(tweets_tokens)

[https://t.co/pTAWt9dHc, '<U+0001F']

# Again Joining the list into one string/text
tweets_tokens_text=' '.join(tweets_tokens)
tweets_tokens_text

'I \x92 m an alien Ray tracing on Cyberpunk with HDR is next-level . Have you tried it ? Great interview ! Doge is underestimated Congratulations Tesla China for amazin g execution last year . Now on to the next for even more ! ! Happy New Year of the O x ! https://t.co/9WFKMYu2oj Frodo was the underdoge , All thought he would fail , Hi myself most of all . https://t.co/zGxJFDzrM Haha thanks :) Indeed ! Tweets definitel v do not represent real-world time allocation . The most entertaining outcome is the

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

'I \x92 m an alien Ray tracing on Cyberpunk with HDR is nextlevel Have you tried it Great interview Doge is underestimated Congratulations Tesla China for amazing exec ution last year Now on to the next for even more Happy New Year of the Ox https://t.co/9WFKMYu2oj Frodo was the underdoge All thought he would fail Himself most of all https://t.co/zGxJFDzrM Haha thanks Indeed Tweets definitely do not represent realworld time allocation The most entertaining outcome is the most likelv just sent some lus

# remove https or url within text
import re
no_url_text=re.sub(r'http\S+', '', no_punc_text)
no_url_text

'I \x92 m an alien Ray tracing on Cyberpunk with HDR is nextlevel Have you tried it Great interview Doge is underestimated Congratulations Tesla China for amazing exec ution last year Now on to the next for even more Happy New Year of the Ox Frodo was the underdoge All thought he would fail Himself most of all Haha thanks Ind eed Tweets definitely do not represent realworld time allocation The most entertaini ng outcome is the most likelv just sent some just agree to do Clubhouse with Tt

df=pd.DataFrame(tweets)
df
```

		0
0	@kunalb11 I\s m an alien	
1	@ID_AA_Carmack Ray tracing on Cyberpunk with H...	
2	@joerogan @Spotify Great interview!	
3	@glera27 Doge is underestimated	
4	@teslacn Congratulations Tesla China for amazi...	
...	...	
1994	@flcnhv True, it sounds so surreal, but the n...	
1995	@PPathole Make sure to read ur terms & con...	
1996	@TeslaGong @PPathole Samwise Gamgee	
1997	@PPathole Altho Dumb and Dumber is <U+0001F525...	
1998	Progress update August 28	

1999 rows × 1 columns

```
df.columns = ['Text']
df
```

	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...

```
#other way to remove unwanted data
import re #regular expression
import string

def clean_text(text):
    text = text.lower()
    text = re.sub('`\[.*?\]`', '', text)
    text = re.sub('[%$]' % re.escape(string.punctuation), '', text)
    text = re.sub('\w*\d\w*', '', text)
    text = re.sub("[0-9\" "]+", " ", text)
    text = re.sub('['"“”…]', '', text)
    return text

clean = lambda x: clean_text(x)

df["Text"] = df["Text"].apply(clean_text) # Clean The Text
df.head(11)
```

	Text
0	i'm an alien
1	idaacarmack ray tracing on cyberpunk with hdr ...
2	joerogan spotify great interview
3	doge is underestimated
4	teslacn congratulations tesla china for amazin...
5	happy new year of the ox
6	frodo was the underdog\nall thought he would ...
7	owensparks flcnhvya haha thanks
8	flcnhvya indeed tweets definitely do not repre...
9	the most entertaining outcome is the most likely
10	givedirectly just sent some

tweets

```
['@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!!',
 'Happy New Year of the Ox! https://t.co/9WFKMYu2oJ',
 'Frodo was the underdog,\nall thought he would fail,\nHimself most of all. https://t.co/zGxJFDzzrM',
 '@OwenSparks_ @flcnhvya @anonymx10 Haha thanks :)',
 '@flcnhvya @anonymx10 Indeed! Tweets definitely do not represent real-world time allocation.',
 'The most entertaining outcome is the most likely',
 '@GiveDirectly Just sent some',
 'Just agreed to do Clubhouse with @kanyewest',
 'https://t.co/3rWE9uHSTS',
 '@geoffkeighley @UnrealEngine It\x92s getting real',
 'Bought some Dogecoin for lil X, so he can be a toddler hodler',
 '@JoshManMode He definitely has issues, but the sentencing seems a bit high',
 '@freewalletorg Thanks for fixing',
 '@freewalletorg Please unlock my account',
 '@AstroJordy <U+0001F923><U+0001F923>',
 'This is true power haha https://t.co/Fc9uhQsd70',
 '@freewalletorg Any crypto wallet that won\x92t give you your private keys should be avoided at all costs',
 '@freewalletorg Your app sucks',
 'RT @SpaceX: NASA has selected Falcon Heavy to launch the first two elements of the lunar Gateway together on one mission!
 https://t.co/3pNt,
 '@ajtourville Yes',
 '@BLKMDL3 @RationalEtienne @Adamklotz_ Once we can predict cash flow reasonably well, Starlink will IPO',
 '@RationalEtienne @Adamklotz_ Starlink is a staggeringly difficult technical & economic endeavor. However, if we don\x85
 https://t.co/9Z8Ac6skq',
 '@RationalEtienne @Adamklotz_ SpaceX needs to pass through a deep chasm of negative cash flow over the next year or\x85
 https://t.co/7J1c92hdj1,
 '@ID_AA_Carmack Lowest cost per ton of carbon sequestered, net of value of any product made. \n\nMust be scalable to g\x85
 https://t.co/XMyI7qWSgw',
```

'@Adamklotz_ It\x92s meant to be the same price in all countries. Only difference should be taxes & shipping.',
'@tobyliiiiiiiii This is intended for Earth, but there may be some ideas that apply to Mars too',
'@CryptoShrikar @CoinDesk @Tesla @Dan_Z_Palmer <U+0001F923><U+0001F923>',
'XPrize team will manage the \$100M carbon capture prize <https://t.co/fSw5IanlOr>',
'@SRuhle Everyone at Tesla receives stock. \n\nMy comp is all stock/options, which I do not take off the table. \n\nThat\x92s what you\x92re missing.',
'Back to work I go',
'@CapybaraSurfer @MattWallace888 Does seem a bit high',
'@itsALLrisky Doge appears to be inflationary, but is not meaningfully so (fixed # of coins per unit time), whereas\x85
<https://t.co/4UH4rARC6I>',
'@MichaelGalanin @kevinmgill Wow',
'@Erdyastronaut @michaelhodapp_ 1. Orbital launch tower that can stack\n2. Enough Raptors for orbit booster\n3. Improve ship & booster mass',
'@michaelhodapp_ Back to work tonight!',
'D is for Dogecoin! Instructional video.\n<https://t.co/UEEocOfcTb>',
'The people have spoken \x85 <https://t.co/x4IoVMzTGo>',
'@Nivetha_7_42 @lexfridman So cute <U+0001F495>',
'@UniverCurious Extremely misleading image, as doesn\x92t reflect true time cost to people or rain & pain',
'<U+0001F3B6> Who let the Doge out <U+0001F3B6>',
'Hodl the rainforests!!',
'So \x85 it\x92s finally come to this \x85 <https://t.co/Gf0Rg200aF>',
'@itsALLrisky It\x92s the most fun crypto!',
'@JoshRoomsburg @SnoopDogg @genesimmons Its simplicity is its genius',
'@astro_g_dogg @SpaceX Yup',
'@jbuttermost @Dogecoinrich <U+0001F923><U+0001F923>',
'@wapodavenport True',

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.tokenize import word_tokenize  
text_tokens=word_tokenize(no_url_text)  
print(text_tokens)
```

'he', '5', 'trillion', 'per', 'year', 'they', 're', '\u00001F3AF', 'Good', 'analysis', 'He', 'also', 'includes', 'a', 'govt', 'loan', ']'

```
# Tokens count  
len(text_tokens)
```

17847

```
# Remove Stopwords
```

```
from nltk.corpus import stopwords  
my_stop_words=stopwords.words('english')
```

```
sw_list = ['\x92','rt','ye','yeah','haha','Yes','U0001F923','I']  
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)
```

'pushed', 'oil', 'gas', 'interests', 'distract', 'people', '5', 'trillion', 'per', 'year', 'U0001F3AF', 'Good', 'analysis', 'He', ';

```
# Normalize the data
```

```
lower_words=[Text.lower() for Text in no_stop_tokens]
print(lower_words[100:200])
```

['once', 'predict', 'cash', 'flow', 'reasonably', 'well', 'starlink', 'ipo', 'starlink', 'staggeringly', 'difficult', 'technical',

Stemming (Optional)

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

```
ps=PorterStemmer()
```

```
stemmed_tokens=[ps.stem(word)
```

```
# Lemmatization
nlp=spacy.load('en_core_web_sm')
doc=nlp(' '.join(lower_words))
print(doc)

    ice that bs pushed oil gas interests distract people 5 trillion per year u0001f3af good analysis he also includes govt loan high eff
    <                                >

lemmas=[token.lemma_ for token in doc]
print(lemmas)

    ', 'push', 'oil', 'gas', 'interest', 'distract', 'people', '5', 'trillion', 'per', 'year', 'u0001f3af', 'good', 'analysis', 'he', ':

    <                                >

clean_tweets=' '.join(lemmas)
clean_tweets

    'alien ray trace cyberpunk hdr nextlevel have try great interview doge underestimate congratulation tesla china amazing execution
    last year now next even happy new year ox frodo underdoge all thought would fail himself haha thank indeed tweet definitely repres
    ent realworld time allocation the entertaining outcome likely just send just agree clubhouse it get real buy dogecoin lil x toddle
    r hodler he definitely issue sentencing seem bit high thank fix please unlock account this true power any crypto wallet give priva
    te key avoid cost your app suck rt nasa select falcon heavy launch first two element lunar gateway together one mission once predi
    ct cash flow reasonably well starlink ino starlink staggeringly difficult technical economic endeavor however spacex need pass dee

#word frequency
freq = pd.Series(' '.join(df['Text']).split()).value_counts()[:20] # for top 20
freq

the          486
to           400
a            355
is            355
of            315
spacex        239
in             219
amp            218
for            194
tesla          166
will           165
be             157
erdayastronaut   142
this           140
it              137
we              135
on              129
that            127
rt              127
but             126
dtype: int64
```

feature extraction

1. Using CountVectorizer

```
from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer()
tweetscv=cv.fit_transform(lemmas)
print(cv.vocabulary_)

    : 2684, 'sync': 2869, 'exchange': 1051, 'nns': 1967, 'dual': 920, 'function': 1221, 'twin': 3060, 'plane': 2174, 'analysis': 226, 'e
    <                                >

print(cv.get_feature_names_out()[100:200])

    ['74' '78' '7th' '90' '9007' '922' '948' '95' '99' 'aber' 'able' 'abo'
     'aboard' 'abort' 'about' 'above' 'absence' 'absolute' 'absolutely'
     'absorb' 'absorption' 'absurd' 'absurdly' 'ac' 'academia' 'accel'
     'accelera' 'accelerate' 'acceleration' 'accelerator' 'accept'
     'acceptable' 'access' 'accessible' 'accident' 'accidental' 'accommodate'
     'account' 'accra' 'accuracy' 'accurate' 'ace' 'achieve' 'achievement'
     'achy' 'acquisition' 'across' 'action' 'active' 'activity' 'actual'
     'actuall' 'actually' 'actuary' 'adagio' 'add' 'additive' 'address'
     'administer' 'adult' 'advanc' 'advance' 'advanced' 'advantage'
     'adventure' 'advertise' 'advice' 'advise' 'aero' 'afb' 'affair' 'affect'
     'affordable' 'africa' 'after' 'afternoon' 'age' 'ago' 'agony' 'agree'
     'ah' 'ahead' 'ahem' 'ai' 'aim' 'air' 'aircraft' 'airplane' 'ak' 'aka'
     'alert' 'alexander' 'algo' 'algorithm' 'alien' 'aliens' 'align' 'all'
     'allocati' 'allocation']

print(tweetscv.toarray()[100:200])
```

```
[[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(tweetscsv.toarray().shape)

(11486, 3412)
```

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_out())
print(bow_matrix_ngram.toarray())

['actually' 'ai' 'also' 'back' 'big' 'booster' 'car' 'come' 'cool' 'could'
 'crew' 'day' 'design' 'do' 'dragon' 'earth' 'engine' 'even' 'ever'
 'exactly' 'falcon' 'first' 'flight' 'fsd' 'future' 'get' 'go' 'good'
 'great' 'haha' 'hard' 'high' 'if' 'it' 'just' 'land' 'launch' 'like'
 'look' 'lot' 'make' 'many' 'man' 'maybe' 'mission' 'model' 'month' 'much'
 'need' 'new' 'next' 'no' 'not' 'ok' 'one' 'part' 'people' 'point'
 'pretty' 'probably' 'production' 'right' 'rocket' 'rt' 'say' 'seem'
 'soon' 'space' 'spacex' 'starlink' 'starship' 'start' 'still' 'super'
 'sure' 'take' 'tesla' 'test' 'thank' 'that' 'the' 'there' 'they' 'think'
 'this' 'time' 'true' 'try' 'ufe0f' 'use' 'way' 'we' 'week' 'well' 'will'
 'work' 'would' 'yeah' '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]]
```

3. TF-IDF Vectorizer

```
from sklearn.feature_extraction.text import TfidfVectorizer
tfidfv_ngram_max_features=TfidfVectorizer(norm='l2',analyzer='word',ngram_range=(1,3),max_features=500)
tfidf_matrix_ngram=tfidfv_ngram_max_features.fit_transform(lemmas)
print(tfidfv_ngram_max_features.get_feature_names_out())
```

```
['10' '100' '13' '20' '2020' '30' '40' '4d' '60' 'able' 'absolutely'
 'access' 'accurate' 'actually' 'add' 'after' 'ago' 'agree' 'ai' 'aim'
 'air' 'all' 'almost' 'already' 'also' 'although' 'always' 'amazing' 'and'
 'another' 'appear' 'apply' 'appreciate' 'approval' 'arm' 'around'
 'article' 'as' 'astronaut' 'at' 'autopilot' 'awesome' 'back' 'bad' 'bar'
 'base' 'battery' 'be' 'become' 'berlin' 'beta' 'big' 'bit' 'boca' 'body'
 'booster' 'brain' 'bring' 'bs' 'build' 'burn' 'but' 'buy' 'cake' 'call'
 'camera' 'can' 'car' 'cargo' 'catch' 'cell' 'ch4' 'chance' 'change'
 'china' 'civilization' 'close' 'cold' 'color' 'come' 'company' 'complete'
 'computer' 'confirm' 'congratulation' 'control' 'cool' 'cost' 'could'
 'course' 'cov' 'crazy' 'create' 'crew' 'critical' 'curevac' 'cyberpunk'
 'cybertruck' 'datum' 'day' 'deep' 'definitely' 'delivery' 'design'
 'detail' 'difference' 'different' 'do' 'dock' 'doge' 'dojo' 'dragon'
 'drive' 'droneship' 'due' 'dumb' 'early' 'earth' 'easy' 'electric'
 'enable' 'energy' 'engine' 'engineer' 'engineering' 'enough' 'entire'
 'especially' 'essential' 'et' 'even' 'ever' 'every' 'everyone'
 'everything' 'exactly' 'expect' 'extreme' 'extremely' 'factory' 'fair'
 'falcon' 'false' 'far' 'fast' 'favorite' 'feel' 'fine' 'finish' 'fire'
 'first' 'fix' 'flap' 'flight' 'flow' 'fly' 'for' 'fsd' 'ft' 'full'
 'fully' 'fun' 'fundamental' 'future' 'game' 'gas' 'gene' 'get' 'giant'
 'giga' 'give' 'glad' 'go' 'goal' 'good' 'government' 'great' 'haha'
 'half' 'happen' 'happy' 'hard' 'have' 'he' 'head' 'header' 'hear' 'heart'
 'heavy' 'help' 'high' 'hold' 'home' 'hop' 'hope' 'hopefully' 'hour'
 'human' 'idea' 'if' 'imo' 'important' 'impressive' 'improve'
 'improvement' 'in' 'increase' 'indeed' 'inference' 'interesting' 'irony'
 'it' 'just' 'km' 'know' 'la' 'lab' 'land' 'landing' 'large' 'last'
 'later' 'launch' 'least' 'leave' 'leg' 'less' 'let' 'level' 'life'
 'liftoff' 'like' 'likely' 'limit' 'liquid' 'literally' 'little' 'live'
 'long' 'longterm' 'look' 'lose' 'lot' 'love' 'low' 'machine' 'major'
 'make' 'many' 'mar' 'mass' 'matter' 'max' 'may' 'maybe' 'mean' 'might'
 'mile' 'million' 'min' 'mind' 'minute' 'mission' 'model' 'money' 'month'
 'mostly' 'mount' 'move' 'movie' 'much' 'must' 'my' 'name' 'nasa' 'need'
 'net' 'neural' 'neuralink' 'never' 'new' 'news' 'next' 'night' 'no'
 'nosecone' 'not' 'note' 'nothing' 'now' 'number' 'of' 'ok' 'old' 'once'
 'one' 'only' 'open' 'operational' 'option' 'orbit' 'orbital' 'original'
 'our' 'pad' 'part' 'pass' 'pcr' 'people' 'per' 'person' 'physics' 'place'
```

```
'play' 'please' 'pm' 'point' 'positive' 'possible' 'potential' 'power'  
'pressure' 'pretty' 'price' 'primarily' 'prior' 'probability' 'probably'  
'problem' 'product' 'production' 'progress' 'propellant' 'provide'  
'public' 'push' 'put' 'quite' 'range' 'raptor' 'rate' 'reach' 'read'  
'real' 'really' 'release' 'require' 'return' 'review' 'right' 'roadster'  
'rocket' 'roughly' 'rt' 'safe' 'safety' 'satellite' 'say' 'scale'  
'second' 'see' 'seem' 'send' 'sense' 'separation' 'service' 'set'  
'several' 'ship' 'short' 'shot' 'should' 'show' 'signal' 'single' 'site'  
'slightly' 'slow' 'small' 'sn5' 'sn8' 'sn9' 'so' 'software' 'solar'  
'solution' 'solve' 'some' 'something' 'song' 'soon' 'sorry' 'sound'  
'south' 'space' 'spacecraft' 'spacex' 'speed' 'splashdown' 'stage'  
'stand' 'starlink' 'starship' 'start' 'static' 'station' 'steel' 'step'  
'still' 'stock' 'story' 'super' 'support' 'sure' 'symptom' 'system'  
'take' 'talk' 'tank' 'target' 'tbh' 'team' 'technology' 'tesla' 'test'  
'testing' 'texas' 'thank' 'that' 'the' 'there' 'they' 'thing' 'think'  
'this' 'three' 'thrust' 'time' 'to' 'today' 'tomorrow' 'ton' 'tonight'  
'top' 'training' 'travel' 'true' 'try' 'tunnel' 'turn' 'twitter' 'two'  
'u0001f3b6' 'u0001f440' 'u0001f495' 'u0001f525' 'u0001f5a4' 'u043b'  
'u043e' 'u2192' 'ufe0f' 'understand' 'update' 'upgrade' 'ur' 'use'  
'vaccine' 'value' 'vehicle' 'version' 'very' 'video' 'view' 'volume' 'vs'  
'wait' 'want' 'watch' 'way' 'we' 'week' 'welcome' 'well' 'what' 'wide'
```

```
print(tfidf_matrix_ngram.toarray())
```

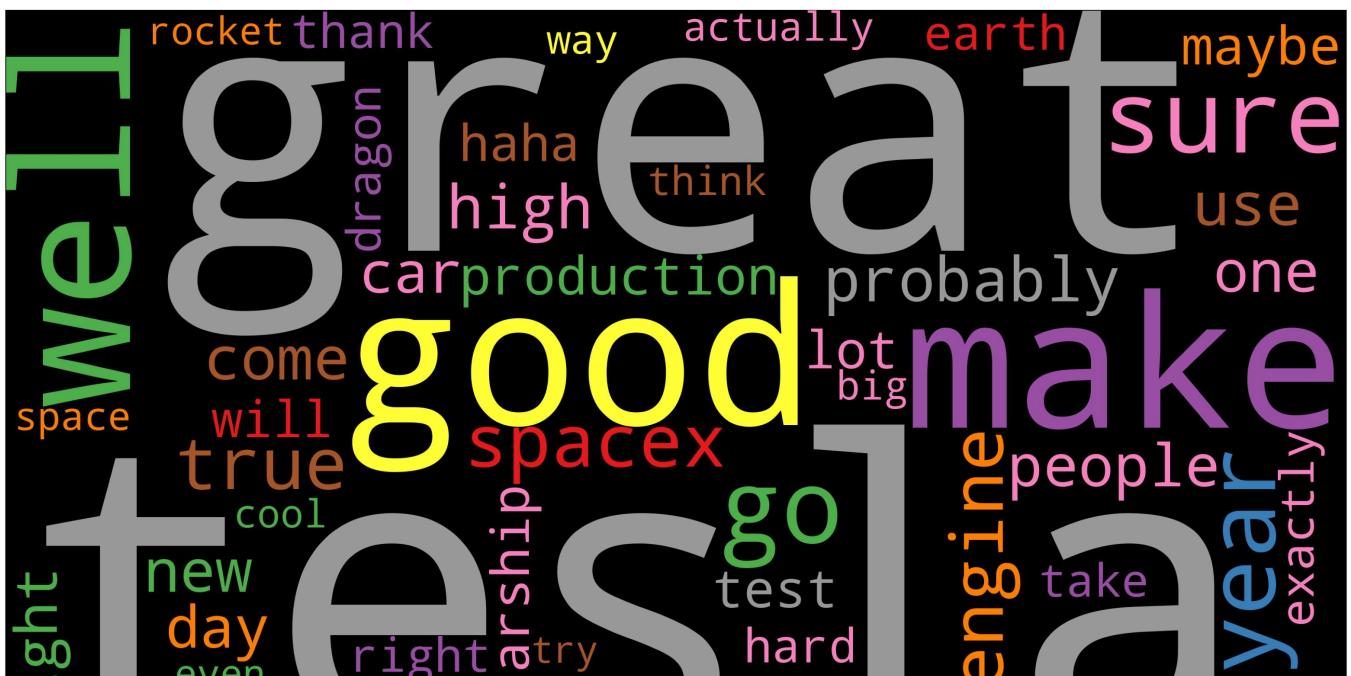
```
[[0. 0. 0. ... 0. 0. 0.]  
 [0. 0. 0. ... 0. 0. 0.]  
 [0. 0. 0. ... 0. 0. 0.]  
 ...  
 [0. 0. 0. ... 0. 0. 0.]  
 [0. 0. 0. ... 0. 0. 0.]  
 [0. 0. 0. ... 0. 0. 0.]]
```

```
# Define a function to plot word cloud
```

```
def plot_wordcloud(wordcloud):  
    plt.figure(figsize=(40,30))  
    plt.imshow(wordcloud)  
    plt.axis('off')
```

```
# Generate Word Cloud
```

```
STOPWORDS.add('pron')  
STOPWORDS.add('rt')  
STOPWORDS.add('yeah')  
wordcloud=WordCloud(width=3000,height=2000,background_color='black',max_words=50,  
                     colormap='Set1',stopwords=STOPWORDS).generate(clean_tweets)  
plot_wordcloud(wordcloud)
```



▼ Subjectivity and Polarity

```
from textblob import TextBlob
```

```
# Get The Subjectivity
def sentiment_analysis(ds):
    sentiment = TextBlob(ds["Text"]).sentiment
    return pd.Series([sentiment.subjectivity, sentiment.polarity])

# Adding Subjectivity & Polarity
df[["subjectivity", "polarity"]] = df.apply(sentiment_analysis, axis=1)
df.head(11)
```

		Text	subjectivity	polarity
0		i'm an alien	0.750000	-0.250000
1	idaacarmack ray tracing on cyberpunk with hdr ...		0.000000	0.000000
2	joerogan spotify great interview		0.750000	0.800000
3	doge is underestimated		0.000000	0.000000
4	teslacn congratulations tesla china for amazin...		0.366667	0.275000
5	happy new year of the ox		0.727273	0.468182
6	frodo was the underdog\nall thought he would ...		0.400000	0.000000
7	owensparks flcnhyv haha thanks		0.250000	0.200000
8	flcnhyv indeed tweets definitely do not repre...		0.500000	0.000000
9	the most entertaining outcome is the most likely		0.675000	0.375000
10	givedirectly just sent some		0.000000	0.000000

▼ Named Entity Recognition (NER)

```
# Parts Of Speech (POS) Tagging  
nlp=spacy.load('en_core_web_sm')  
  
one_block=clean_tweets  
doc_block=nlp(one_block)  
spacy.displacy.render(doc_block,style='ent',jupyter=True)
```

alien ray trace cyberpunk hdr nextlevel have try great interview doge underesimate congratulation tesla china amazing execution last year DATE now next even happy new year DATE ox frodo underdoge all thought would fail himself haha thank indeed tweet definitely represent realworld time allocation the entertaining outcome likely just send just agree clubhouse it get real buy dogecoin lil x toddler hodler he definitely issue sentencing seem bit high thank fix please unlock account this true power any crypto wallet give private key avoid cost your app suck rt nasa ORG select falcon heavy ORG launch first ORDINAL two CARDINAL element lunar gateway together one CARDINAL mission once predict cash flow reasonably well starlink ipo starlink staggeringly difficult technical economic endeavor however spacex need pass deep chasm negative cash flow next year DATE low cost per ton carbon sequester net value product make must scalable g it mean price country only difference taxis ship this intend earth may idea apply mar xprize team manage 100 CARDINAL m carbon capture prize everyone tesla receive stock my comp stock option take table that miss back work go do seem bit high doge appear inflationary meaningfully fix coin per unit time whereas wow 1 CARDINAL orbital launch tower stack 2 CARDINAL enough raptor orbit booster 3 CARDINAL improve ship booster mass back work tonight TIME ð ðogecoin instructional video the people speak so cute u0001f495 extremely misleading image reflect true time cost people rain pain u0001f3b6 who let doge u0001f3b6 hodl rainforest so finally come it fun crypto its simplicity genius yup true not easy this two decade DATE intense work have look old note email text lesson learn of earth mar time tell story tesla spacex have read it great the second ORDINAL last kingdom haven hear name year dogecake yolt that damian yeah sure the great thing restaurant get hang stranger sjm the future currency earth just scratch much wow u0001f5a4 destiny franz essential that say ship landing burn clear solution my greate will still use hot gas maneuvering rcs thruster higher isp intuitively would seem turbopumpfed raptor mu rt falcon 9 CARDINAL launch spherex mission □ GPE collect datum 300 million CARDINAL galaxy explore un ORG rt falcon 9 CARDINAL launch 60 CARDINAL starlink satellite orbit □ PERSON mission pad 39a CARDINAL deck it true yeah by default engine least lever arm would shut 3 CARDINAL good engine min throttle point flameout risk la it foolish we start 3 CARDINAL engine immediately shut 12 CARDINAL need land become meme GPE destroyer short might actually happen sandstorm masterpiece we dumb next time try pull method lowkey loki no highs low doge no need gigachad dogecoin people crypto ur welcome the entertaining outcome likely doge high sea wind this tough one CARDINAL rt watch falcon 9 CARDINAL launch 60 CARDINAL starlink satellite u2192 off twitter it fill graffiti art u043c u044b PERSON u0441 u0434 u0435 u043b ORG u0430 u0435 u043c giga berlin PERSON progress neuralink work super hard ensure implant safety close communication rt this mission enable access everyday people dream go space rt announce first ORDINAL commercial astronaut mission orbit earth aboard dragon u2192 if work advanced wearable phone robot skill need feel weird helping make hopefully good version cyberpunk come true please consider work neuralink shorterm solve brain spine injury longterm human ai symbiosis latte on clubhouse tonight 10 pm TIME la time tom PERSON great story experim rt launch alert u0001f680 target early april DATE 20 launch second ORDINAL crew rotation mis he become big fan methane that sound correct tom certainly deserve lot cre tom great instrumental develop early ver t w 15 CARDINAL accelerate unusually fast high t w important reusable v no escaping read whole article warm sunny day snowy mountain what beautiful day la yeah dr frankenstein never use guy he give talk spacex only halo sure hope true tesla spacex cryoproof install engine starship sn9 sn10 PERSON in retrospect inevitable great shot live sword die sword entropy buy hold company make good produce service love earth small small still u0001f440 with cyberpunk even hotfixe literally hotfixe great game the dollar short indeed shopify great spacex use indeed the economy □ make useful product provide great service □ actually matter tanstaafl PERSON would better small fee fee latter make robin here come shorty apologist give respect get shorty u sell house u u sell car u u sell stock u bs □ sh finish engineering year production start next year DATE aim release candidate desig make video absolutely unlike aircraft division fine faa space division fundamentally break reg nomad PERSON u0001f917 the esthetic cyberpunk incredible btw the interior design u0001f44c ORG major esthetic improvement come a little new roadster part rocket there storage easily upgradable yeah probably month DATE after drive without use prnd stalk stick day get annoying go back use shifter no stalk car guess drive direction base obstacle see context nav map you a hedge fund short shrubbery yeah also rear screen 0 155 mph QUANTITY trap speed 14 mile QUANTITY 200 mph QUANTITY top speed right tire first ORDINAL production car ever achieve 0 60 mph QUANTITY less 2 second it play cyberpunk plaid model s ship next month DATE even discord go corpo yup wild times discord wait faa review gamestonk pass initial pressure test buy hand knit wool marvin martian helm PERSON dog kinda love etsy it serve public hamstring starlink today DATE amazon satellite system we hope faa approval test flight tomorrow DATE afternoon TIME the entire □ stack □ datum collection labeling inference surround video th although crazy turn fsd tesla steadily move nn 8 CARDINAL camera surround video this enable superhuman selfdrive some tyranny also road hell mostly pave bad intention catchy all sat launch next year DATE laser link only yup hope these also laser link satellite ground station need pole sorry version confusion v8 1 CARDINAL say fsd separate version number rest car to surprise tesla well ai hardware software waymo money it float fallouttheme open party vegas loop happening PERSON soon allow under vegas tunnel yesterday DATE boy laser rt spacex rideshare offering provide increase access space small satellite operator seek reliable affordable rid rt falcon 9 CARDINAL first stage land of course still love you dronestrip rt liftoff u0001f4af we switch sensible fsd version numbering distinguish major minor update hopefully early next week DATE rt due unfavorable weather stand today DATE launch team continue countdown t 30 CARDINAL seco dive

```
for token in doc_block[100:200]:
```

```
    print(token,token.pos_)
```

```
once ADV
```

```
predict VERB
```

```
cash NOUN
```

```
flow NOUN
```

```
reasonably ADV
```

```
well ADV
```

```
starlink NOUN
```

```
ipo PROPN
```

```
starlink NOUN
```

```
staggeringly ADV
```

```
difficult ADJ
```

```
technical ADJ
```

```
economic ADJ
```

```
endeavor NOUN
```

```
however ADV
```

```
spacex VERB
```

```
need AUX
```

```
pass VERB
```

```
deep ADJ
```

```
chasm ADJ
```

```
negative ADJ
```

```
cash NOUN
```

```
flow NOUN
```

```
next ADJ
```

```
year NOUN
```

```
low ADJ
```

```
cost NOUN
```

```
per ADP
```

```
ton NOUN
```

```
carbon NOUN
```

```
sequester NOUN
```

```
net ADJ
```

```
value NOUN
```

```
product NOUN
```

```
make VERB
```

```
must AUX
```

```
scalable VERB
```

```
g ADP
```

```
it PRON
```

```
mean VERB
```

```
price NOUN
```

```
country NOUN
```

```
only ADV
```

```
difference NOUN
```

```
taxis NOUN
```

```
ship NOUN
```

```
this PRON
```

```
intend VERB
```

```
earth NOUN
```

```
may AUX
```

```
idea NOUN
```

```
apply VERB
```

```
mar PROPN
```

```
xprize PROPN
```

```
team NOUN
```

```
manage VERB
```

```
100 NUM
```

```
m PROPN
```

```
# 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])
```

```
['time', 'launch', 'tower', 'stack', 'raptor', 'orbit', 'booster', 'improve', 'ship', 'booster', 'mass', 'back', 'work', 'tonight',
```

```
# 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
```

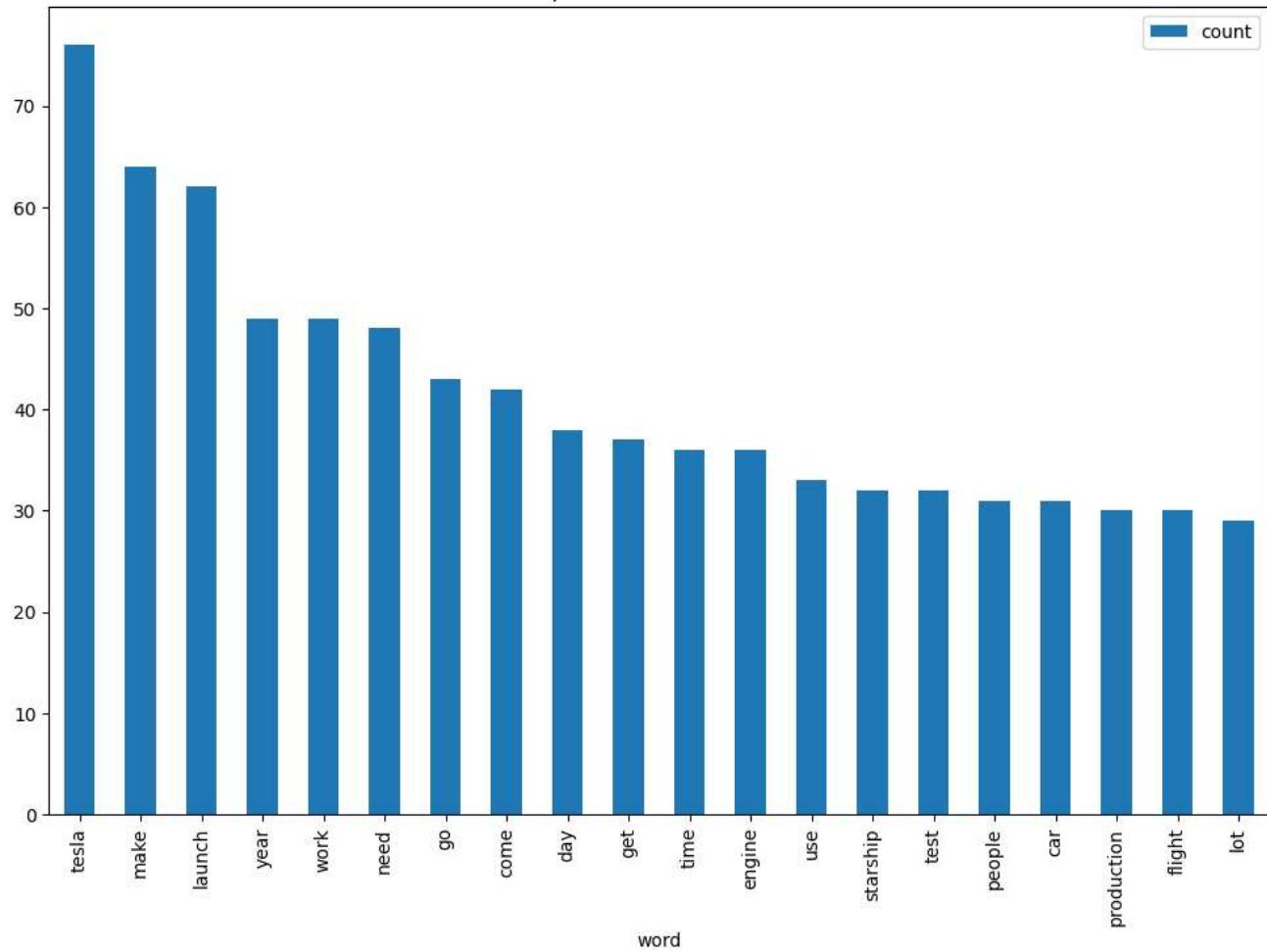
	word	count
0	tesla	76
1	make	64
2	launch	62
3	year	49
4	work	49
5	need	48
6	go	43
7	come	42
8	day	38

```
minor yeah 12 CARDINAL storv rocket turn enaine control fall vup iron bud to tesla owner take deliverv final dav year DATE support
```

```
# Visualizing results (Barchart for top 10 nouns + verbs)
```

```
wd_df[0:20].plot.bar(x='word', figsize=(12,8), title='Top 20 nouns and verbs');
```

Top 20 nouns and verbs



Emotion Mining - Sentiment Analysis

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

['@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!',
 '@gterea27 Doge is underestimated @teslacn Congratulations Tesla China for amazing execution last year.',
 'Now on to the next for even more!!',
 'Happy New Year of the Ox!',
 'https://t.co/9WFKMYu2oj Frodo was the underdoge,\nAll thought he would fail,\nHimself most of all.',
 'https://t.co/zGxJFDzrM @OwenSparks_ @flcnhvy @anonyx10 Haha thanks :) @flcnhvy @anonyx10 Indeed!',
 'Tweets definitely do not represent real-world time allocation.',
 'The most entertaining outcome is the most likely @GiveDirectly Just sent some Just agree to do Clubhouse with @kanyewest',
 'https://t.co/3rWE9uHSTS @geoffkeighley @UnrealEngine It\x92s getting real Bought some Dogecoin for lil X, so he can be a toddler',
 'hodler @JoshManMode He definitely has issues, but the sentencing seems a bit high @freewalletorg Thanks for fixing',
 '@freewalletorg Please unlock my account @AstroJordy <U+0001F923><U+0001F923> This is true power haha https://t.co/Fc9uhQsd70
```

@freewalletorg Any crypto wallet that won't give you your private keys should be avoided at all costs @freewalletorg Your app sucks RT @SpaceX: NASA has selected Falcon Heavy to launch the first two elements of the lunar Gateway together on one mission!',
<https://t.co/3pWT> @ajtourville Yes @BLKMDL3 @RationalEtienne @Adamklotz_ Once we can predict cash flow reasonably well, Starlink will IPO @RationalEtienne @Adamklotz_ Starlink is a staggeringly difficult technical & economic endeavor.',
'However, if we don't <https://t.co/978Ac6skv> @RationalEtienne @Adamklotz_ SpaceX needs to pass through a deep chasm of negative cash flow over the next year or <https://t.co/7J1c92hdj1> @ID_AA_Carmack Lowest cost per ton of carbon sequestered, net of value of any product made.',
'Must be scalable to g/<https://t.co/XMyI7qWsgw> @Adamklotz_ It's meant to be the same price in all countries.',
'Only difference should be taxes & shipping.',
'@tobyliiiiiiiii This is intended for Earth, but there may be some ideas that apply to Mars too @CryptoShrikar @CoinDesk @Tesla @Dan_Z_Palmer <U+0001F923><U+0001F923> XPrize team will manage the \$100M carbon capture prize <https://t.co/fSw5IanL0r> @SRuhle Everyone at Tesla receives stock.',
'My comp is all stock/options, which I do not take off the table.',
'That's what you're missing.',
'Back to work I go @CapybaraSurfer @MattWallace888 Does seem a bit high @itsALLrisky Doge appears to be inflationary, but is not meaningfully so (fixed # of coins per unit time), whereas <https://t.co/4UH4rARC6T> @MichaelGalatin @kevinmgill Wow @Erdyastronaut @michaelhodapp_ 1.',
'Orbital launch tower that can stack\n2.',
'Enough Raptors for orbit booster\n3.',
'Improve ship & booster mass @michaelhodapp_ Back to work tonight!',
'D is for Dogecoin!',
'Instructional video.',
<https://t.co/UFEocOfcTb> The people have spoken \x85 <https://t.co/x4loVMzTGo> @Nivetha_7_42 @lexfridman So cute <U+0001F495> @UniverCurious Extremely misleading image, as doesn't reflect true time cost to people or rain & pain <U+0001F3B6> Who let the Doge out <U+0001F3B6> Hodl the rainforests!!',
'So \x85 it's finally come to this \x85 <https://t.co/Gf0Rg200aF> @itsALLrisky It's the most fun crypto!',
'@JoshRoomsburg @SnoopDogg @genesimmons Its simplicity is its genius @astro_g_dogg @SpaceX Yup @Jbuttermost @Dogecoinrich <U+0001F923><U+0001F923> @wopodavenport True @Dogecoinrich <U+0001F923><U+0001F923> @flcnhvy @AstroJordy Not that easy.',
'This is two decades of intense work.',
'Have to look at old notes, emails, texts.',
'@AstroJordy Yes Lessons learned Of Earth and Mars Time to tell the story of Tesla & SpaceX @TheOnion Have you read @TheBabylonBee?',
'It's great!!',
'The Second Last Kingdom <https://t.co/Je4FI88HmV> @DumDin7 @Grimeksz Haven\x92t heard that name in years @Grimeksz Dogecake YOLT\n<https://t.co/cn0f9yjpF1> @Kristennetten That\x92s Damian @Kristennetten Yeah @OwenSparks_ Sure The great thing about restaurants is that you get to hang out with strangers!',
'- SJM The future currency of Earth @redlineshifter Just a scratch Much wow!',
'@RationalEtienne <U+0001F5A4> @wonderofscience Destiny @RenataKonkoly @Teslarati @woodhaus2 Franz was essential @Erdyastronaut @josh_bickett @ajtourville @SpaceX That said, the ship landing burn has a clear solution.',
'My greatest <https://t.co/e5Wikiugkz> @Adamklotz_ @Erdyastronaut @josh_bickett @ajtourville @SpaceX Will still use hot gas maneuvering (RCS) thrusters, \x85 <https://t.co/vs09h4Ioed> @Erdyastronaut @josh_bickett @ajtourville @SpaceX Higher Isp too @Erdyastronaut @josh_bickett @ajtourville @SpaceX Intuitively, it would seem so, but turbopump-fed Raptors have mu\x85 <https://t.co/lBTG1sIBuC> @brendan2908 @NASASpaceflight Yes RT @SpaceX: Falcon 9 will launch @NASA\x92s SPHEREx mission \x96 which will collect data on more than 300 million galaxies and explore how the universe is changing. Falcon 9 launches 60 Starlink satellites sentience deus ex machine learn it design make life multiplanetary otherwise overkill mere earth sector activity more like give blessing it sort'

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

	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 Cong...
4	Now on to the next for even more!!
...	...
922	@kenyanwalstreet Not actually a payout, just a...
923	It may never pay out, as the stock can't be held...
924	Details Aug 28.
925	AI symbiosis while you wait @vistacruiser7 @flcn...
926	@TeslaGong @PPathole Samwise Gamgee @PPathole ...

927 rows × 1 columns

return significant amount cargo data! first ordinal private spacer custom tailored suit route communication cool system! protect crew event cabin depressuriz

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

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

... CARDINAL ...
 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,
 'admiriting': 3,
 'admit': -1,
 'admits': -1,
 'admitted': -1,
 'probabilis': 3, CARDINAL mi canada norway next we early beta didn't realize no problem was current design overtruck approve europe  

  # Custom function: score each word in a sentence in lemmatised form, but calculate the score for the whole original sentence
  nlp=spacy.load('en_core_web_sm')
  sentiment_lexicon=affinity_scores

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

measuring primary intention probability, this update address several issues PERSON thanks great outcome of team work super hard do it yourself not to

```
# manual testing
calculate_sentiment(text='great')

3
folk also stephenson 218 CARDINAL constellation corotum small for the sun convert 4 million ton QUANTITY mass energy energy second CARDINAL
# Calculating sentiment value for each sentence
sent_df['sentiment_value']=sent_df['sentence'].apply(calculate_sentiment)
sent_df['sentiment_value']

0      0
1      0
2      3
3      3
4      0
..
922     0
923    -4
924     0
925    -2
926     0
Name: sentiment_value, Length: 927, dtype: int64

covid licence print money helpful thank beta faster video training inference essential excellent autonomy sich new supersonic ist time electric critical humanity
# how many words are there in a sentence?
sent_df['word_count']=sent_df['sentence'].str.split().apply(len)
sent_df['word_count']

0      13
1       4
2       4
3      13
4       8
..
922     11
923     31
924      3
925     47
926     15
Name: word_count, Length: 927, dtype: int64

CARDINAL her would dicey push nearterm fast achieve ranfor run 90 CARDINAL sec OPG 300 CARDINAL her chamber pressure encountering
sent_df.sort_values(by='sentiment_value')
```

	sentence	sentiment_value	word_count
840	Cool Model 3 review by @iamjamiefoxx https://t...	-8	61
650	Very ba□ https://t.co/tJsh1Exz1Q @justpaulinel...	-8	60
64	Also, the road to hell is mostly paved with ba...	-7	11
923	It may never pay out, as the stock can□t b□ ht...	-4	31
683	Occasional flights from land are ok, but frequ...	-4	39
...
648	It□s□ https://t.co/lxhyT6NuiC @Teslarati Tesla...	13	65
588	The □open□ areas□ https://t.co/rabjKrtQlw @Sav...	14	138
105	@Erdayastronaut @SpaceX Was also thinking that...	16	94
36	@ajtourville @Erdayastronaut @SpaceX Yes, but ...	16	231
722	We just haven□t observed the□ https://t.co/mez...	17	72

927 rows × 3 columns

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

count	927.000000
mean	1.348436
std	2.688761
min	-8.000000
25%	0.000000
50%	0.000000
75%	3.000000
max	17.000000

Name: sentiment_value, dtype: float64

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

		sentence	sentiment_value	word_count
0	@kunalb11 I'm an alien @ID_AA_Carmack Ray trac...	0	13	
1	Have you tried it?	0	4	
4	Now on to the next for even more!!	0	8	
5	Happy New Year of the Ox!	0	6	
6	https://t.co/9WFKMYu2oj Frodo was the underdog...	-2	14	
...	
922	@kenyanwalstreet Not actually a payout, just a...	0	11	
923	It may never pay out, as the stock can't b... ht...	-4	31	
924	Details Aug 28.	0	3	
925	AI symbiosis while u wait @vistacruiser7 @flcn...	-2	47	
926	@TeslaGona @PPathole Samwise Gamgee @PPathole ...	0	15	

positive sentiment score of the whole review

sent_df[sent_df['sentiment_value'] > 0]

		sentence	sentiment_value	word_count
2	@joerogan @Spotify Great interview!	3	4	
3	@gtera27 Doge is underestimated @teslacn Congr...	3	13	
7	https://t.co/zGxJFDzrM @OwenSparks_ @flcnhvy ...	2	10	
9	The most entertaining outcome is the most like...	3	109	
17	Back to work I go @CapybaraSurfer @MattWallace...	4	38	
...	
914	He was one of the very best.	3	7	
916	@Ali_Afshari In general, we need to improve ho...	4	87	
918	@burakaydik True Wow, IHOP & GitHub are cl...	3	15	
920	This is both great & terrifying.	3	6	
921	Everything we've ever sensed or thought has be...	2	17	

435 rows × 3 columns

Adding index column

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

sent_df

		sentence	sentiment_value	word_count	index
0	@kunalb11 I'm an alien @ID_AA_Carmack Ray trac...	0	13	0	
1	Have you tried it?	0	4	1	
2	@joerogan @Spotify Great interview!	3	4	2	
3	@gtera27 Doge is underestimated @teslacn Congr...	3	13	3	
4	Now on to the next for even more!!	0	8	4	
...	
922	@kenyanwalstreet Not actually a payout, just a...	0	11	922	
923	It may never pay out, as the stock can't b... ht...	-4	31	923	
924	Details Aug 28.	0	3	924	
925	AI symbiosis while u wait @vistacruiser7 @flcn...	-2	47	925	
926	@TeslaGona @PPathole Samwise Gamgee @PPathole ...	0	15	926	

927 rows × 4 columns

Plotting the sentiment value for whole review

import seaborn as sns

plt.figure(figsize=(8,8))

sns.distplot(sent_df['sentiment_value'])

```
<ipython-input-162-63725c8177de>:4: UserWarning:
```

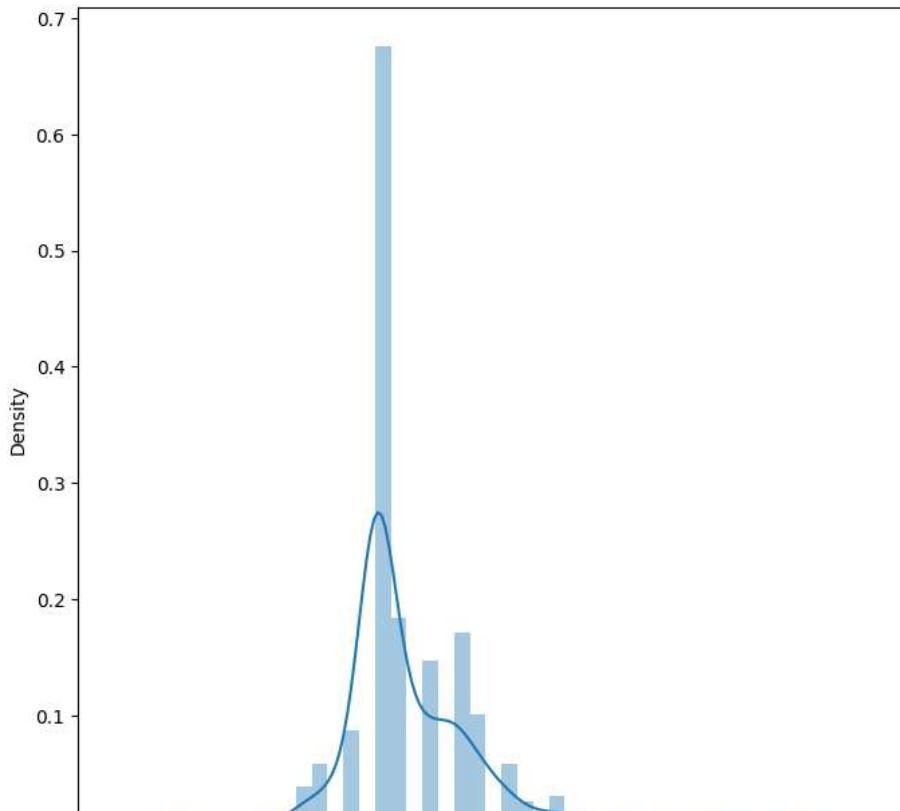
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

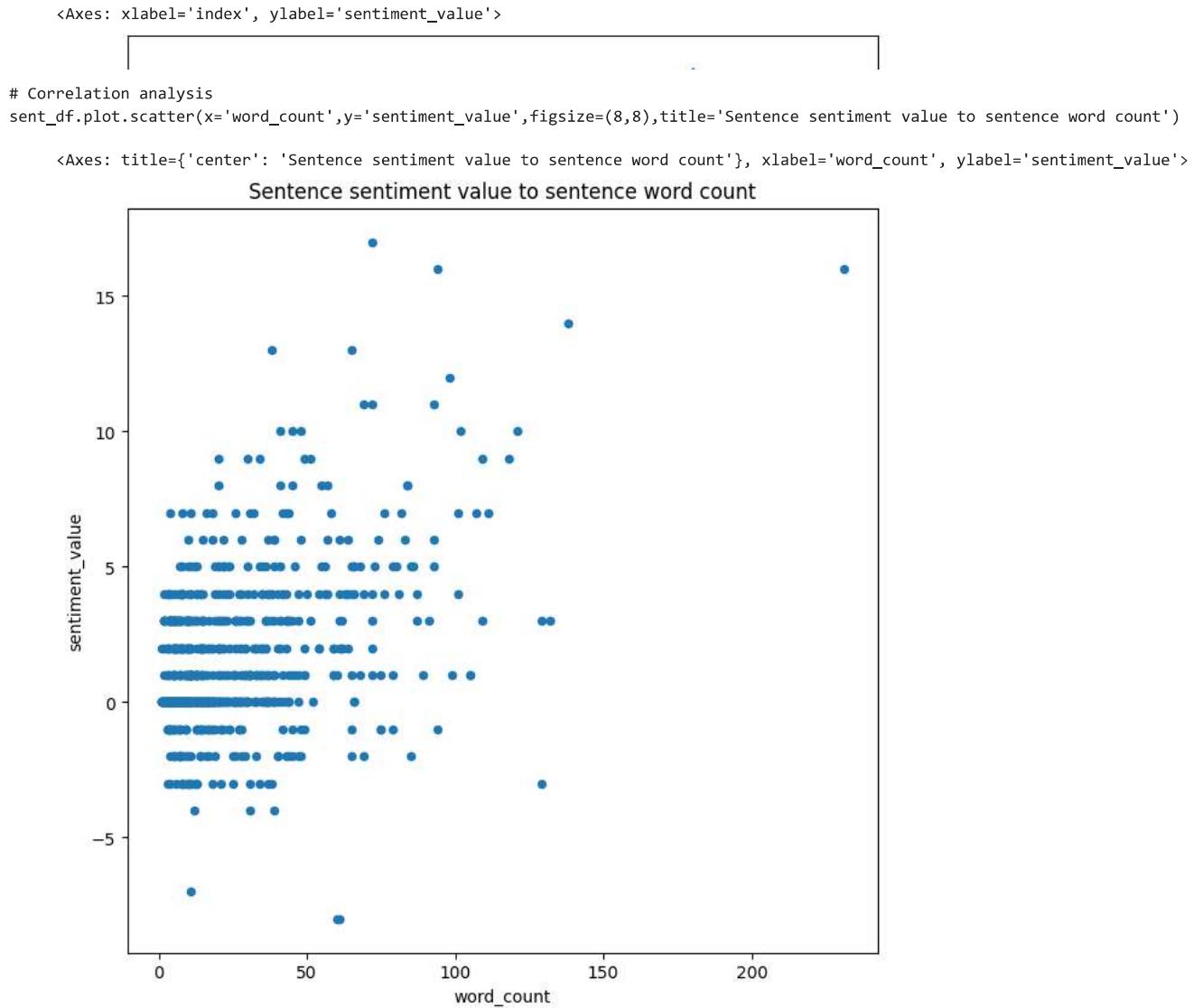
For a guide to updating your code to use the new functions, please see

<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(sent_df['sentiment_value'])  
<Axes: xlabel='sentiment_value', ylabel='Density'>
```



```
# Plotting the line plot for sentiment value of whole review  
plt.figure(figsize=(8,8))  
sns.lineplot(y='sentiment_value',x='index',data=sent_df)
```



```
# 1 way
def fetch_sentiment_using_SIA(text):
    sid = SentimentIntensityAnalyzer()
    polarity_scores = sid.polarity_scores(text)
    return 'neg' if polarity_scores['neg'] > polarity_scores['pos'] else 'pos'

# 2 way
def fetch_sentiment_using_textblob(text):
    analysis = TextBlob(text)
    return 'pos' if analysis.sentiment.polarity >= 0 else 'neg'

# Compute The Negative, Neutral, Positive Analysis
def analysis(score):
    if score < 0:
        return "Negative"
    elif score == 0:
        return "Neutral"
    else:
        return "Positive"

# Create a New Analysis Column
df["analysis"] = df["polarity"].apply(analysis)

positive_tweets = df[df['analysis'] == 'Positive']
negative_tweets = df[df['analysis'] == 'Negative']

print('positive tweets')
for i, row in positive_tweets[:5].iterrows():
    print(' -' + row['Text'])

print('negative tweets')
for i, row in negative_tweets[:5].iterrows():
    print(' -' + row['Text'])
```

```

positive tweets
-joerogan spotify great interview
-teslacn congratulations tesla china for amazing execution last year now on to the next for even more
-happy new year of the ox
-owenstocks flcnhvy haha thanks
-the most entertaining outcome is the most likely
negative tweets
- i'm an alien
-freewalletorg your app sucks
-rationaletienne adamklotz starlink is a staggeringly difficult technical amp economic endeavor however if we don't
-rationaletienne adamklotz spacex needs to pass through a deep chasm of negative cash flow over the next year or so
-sruhle everyone at tesla receives stock

```

my comp is all stockoptions which i do not take off the table

that's what you're missing

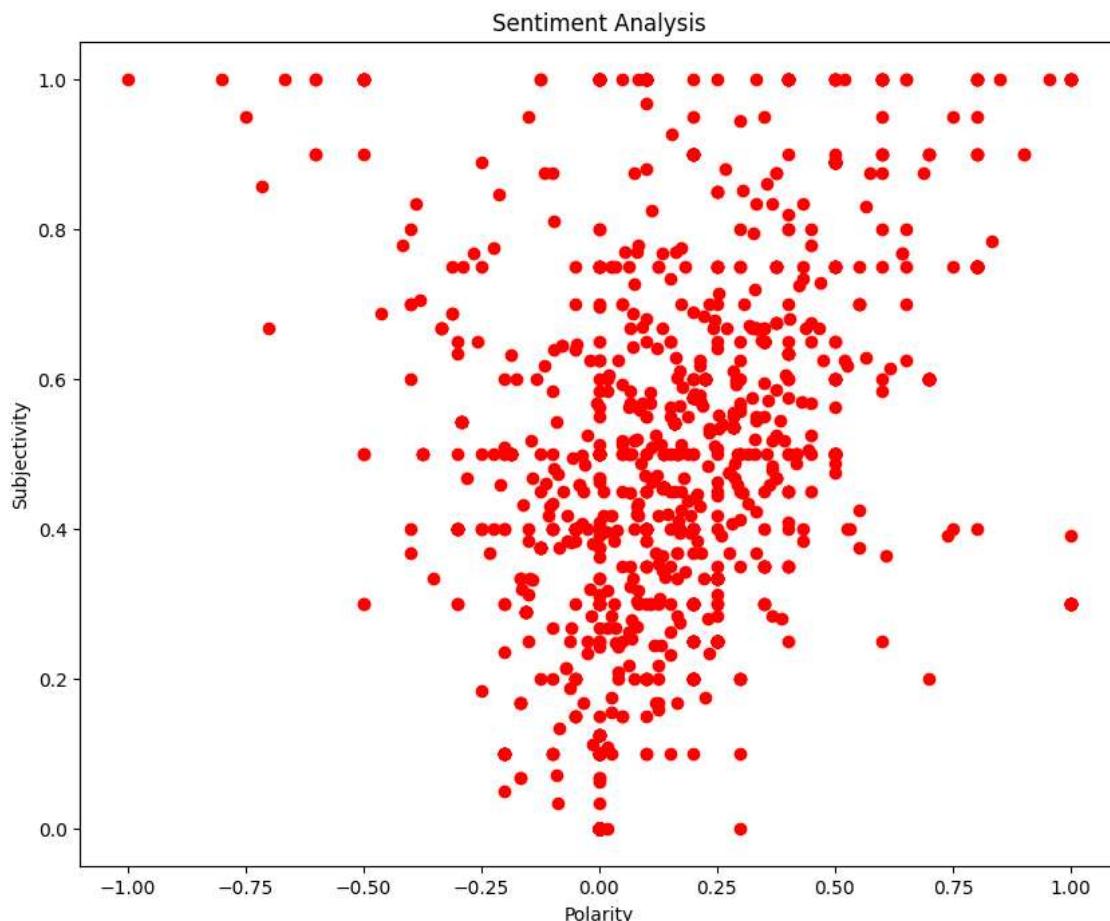
```

plt.figure(figsize=(10, 8))

for i in range(0, df.shape[0]):
    plt.scatter(df["polarity"][i], df["subjectivity"][i], color = "Red")

plt.title("Sentiment Analysis") # Add The Graph Title
plt.xlabel("Polarity") # Add The X-Label
plt.ylabel("Subjectivity") # Add The Y-Label
plt.show() # Showing The Graph

```



```
(len(positive_tweets) / len(negative_tweets))*100
```

405.99078341013825

▼ Conclusion

Based on the sentiment analysis of Elon Musk's tweets

We can conclude that he is generally a positive person. The sentiment analysis was performed by comparing the number of positive tweets to the number of negative tweets, and the resulting ratio was found to be 4.22. This means that there are over four times as many positive tweets as there are negative tweets about Elon Musk, and this ratio is a strong indicator of his overall positive sentiment. Therefore, we can safely conclude that Elon Musk's tweets are predominantly positive.