


 [saipranay190198](#) / [My-personal-files](#) Private


[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Security](#) [Insights](#)


 main 



[My-personal-files](#) / [twitter\\_sentiment Analysis.ipynb](#)

 **saipranay190198** Add files via upload History

 1 contributor

1801 lines (1801 sloc) | 906 KB 

```
In [2]: import re
import nltk
import string
import warnings
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

pd.set_option("display.max_colwidth", 200)
warnings.filterwarnings("ignore", category = DeprecationWarning)

%matplotlib inline
```

```
In [3]: train = pd.read_csv("train_tweets.csv")
        test = pd.read_csv("test_tweets.csv")
```

```
In [4]: train
```

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run
1	2	0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthankd
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in urðŸ™±!!! ðŸŽšðŸ‘žðŸ‘... ðŸ‘!ðŸ‘!ðŸ‘!
4	5	0	factsguide: society now #motivation
...	...	...	...
31957	31958	0	ate @user isz that youuu? ðŸ˜~ ðŸ˜~ ðŸ˜~ ðŸ˜~ ðŸ˜~ ðŸ˜~ ðŸ˜~ ðŸ˜~ ðŸ˜~ â€” ¤ı,
31958	31959	0	to see nina turner on the airwaves trying to wrap herself in the mantle of a genuine hero like shirley chisolm. #shame #imwithher
31959	31960	0	listening to sad songs on a monday morning otw to work is sad
31960	31961	1	@user #sikh #temple vandalised in in #calgary, #wso condemns act
31961	31962	0	thank you @user for you follow

31962 rows x 3 columns

```
In [5]: test
```

```
Out[5]:
```

id	tweet
----	-------

17197 rows x 2 columns

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run
1	2	0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disappointed #getthankd
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in urðŸ™±!!! ðŸ™ˆ~ðŸ™ŽðŸ™„ðŸ™…ðŸ™¡ðŸ™¡ðŸ™¡
4	5	0	factsguide: society now #motivation
5	6	0	[2/2] huge fan fare and big talking before they leave. chaos and pay disputes when they get there. #allshowandnogo
6	7	0	@user camping tomorrow @user @user @user @user @user @user @user dannyâ€¢
7	8	0	the next school year is the year for exams.ðŸ™ˆ- can't think about that ðŸ™ˆ- #school #exams #hate #imagine #actorslife #revolutionschool #girl
8	9	0	we won!!! love the land!!! #allin #cavs #champions #cleveland #clevelandcavaliers â€¢
9	10	0	@user @user welcome here ! i'm it's so #qr8 !

3/12

Out[7]:

	id	label	tweet
13	14	1	@user #cnn calls #michigan middle school 'build the wall' chant '' #tcot
14	15	1	no comment! in #australia #opkillingbay #seashepherd #helpcovedolphins #thecove #helpcovedolphins
17	18	1	retweet if you agree!
23	24	1	@user @user lumpy says i am a . prove it lumpy.
34	35	1	it's unbelievable that in the 21st century we'd need something like this. again. #neverump #xenophobia
56	57	1	@user lets fight against #love #peace
68	69	1	ðŸ˜™@the white establishment can't have blk folx running around loving themselves and promoting our greatness
77	78	1	@user hey, white people: you can call people 'white' by @user #race #identity #medâ€™
82	83	1	how the #altright uses & insecurity to lure men into #whitesupremacy
111	112	1	@user i'm not interested in a #linguistics that doesn't address #race & . racism is about #power. #raciolinguistics bringsâ€™

In [8]: `train.shape , test.shape`

Out[8]: `((31962, 3), (17197, 2))`

In [9]: `train["label"].value_counts()`

Out[9]:

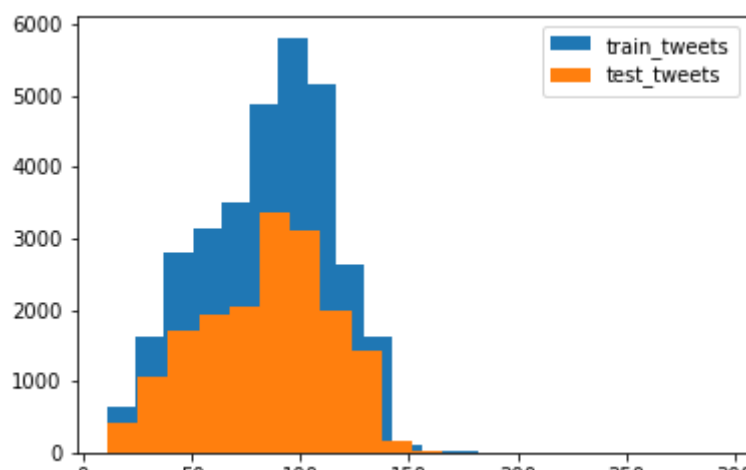
```
0    29720
1     2242
Name: label, dtype: int64
```

In [10]:

```
length_train = train["tweet"].str.len()
length_test = test["tweet"].str.len()

plt.hist(length_train, bins = 20, label = "train_tweets")
plt.hist(length_test, bins = 20, label = "test_tweets")

plt.legend()
plt.show()
```



```
C:\Users\rsaipranayreddy\AppData\Local\Temp\ipykernel_12188\3655217207.py:1: FutureWarning: The frame.append method is deprecated and will be removed from pandas in a future version. Use pandas.concat instead.
  combine = train.append(test, ignore_index = True)
(49159, 3)
```

```
combine['tidy_tweet'] = np.vectorize(remove_pattern)(combine['tweet'], "(
combine.head()
```

```
combine['tidy_tweet'] = combine['tidy_tweet'].str.replace("[^a-zA-Z#]", ' ')
combine.head(10)
```

	id	label	tweet	tidy_tweet
0	1	0.0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run	when a father is dysfunctional and is so selfish he drags his kids into his dysfunction #run

1	2	0.0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked	thanks for #lyft credit i can t use cause they don t offer wheelchair vans in pdx #disapointed #getthanked
2	3	0.0	bihday your majesty	bihday your majesty
3	4	0.0	#model i love u take with u all the time in urðŸ™±!!! ðŸ™ðŸ™ŽðŸ™ðŸ™...ðŸ™ðŸ™ðŸ™	#model i love u take with u all the time in ur
4	5	0.0	factsguide: society now #motivation	factsguide society now #motivation
5	6	0.0	[2/2] huge fan fare and big talking before they leave. chaos and pay disputes when they get there. #allshowandnogo	huge fan fare and big talking before they leave chaos and pay disputes when they get there #allshowandnogo
6	7	0.0	@user camping tomorrow @user @user @user @user @user @user dannyâ€	camping tomorrow danny
7	8	0.0	the next school year is the year for exams.ðŸ™ can't think about that ðŸ™- #school #exams #hate #imagine #actorslife #revolutionschool #girl	the next school year is the year for exams can t think about that #school #exams #hate #imagine #actorslife #revolutionschool #girl
8	9	0.0	we won!!! love the land!!! #allin #cavs #champions #cleveland #clevelandcavaliers â€	we won love the land #allin #cavs #champions #cleveland #clevelandcavaliers
9	10	0.0	@user @user welcome here ! i'm it's so #gr8 !	welcome here i m it s so #gr

In [15]:

```
combine['tidy_tweet'] = combine['tidy_tweet'].apply(lambda x: ' '.join([v
combine.head()
```

Out[15]:

	id	label	tweet	tidy_tweet
0	1	0.0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run	when father dysfunctional selfish drags kids into dysfunction #run
1	2	0.0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disapointed #getthanked	thanks #lyft credit cause they offer wheelchair vans #disapointed #getthanked
2	3	0.0	bihday your majesty	bihday your majesty
3	4	0.0	#model i love u take with u all the time in urðŸ™±!!! ðŸ™ðŸ™ŽðŸ™ðŸ™...ðŸ™ðŸ™ðŸ™	#model love take with time
4	5	0.0	factsguide: society now #motivation	factsguide society #motivation

In [16]:

```
tokenized_tweet = combine['tidy_tweet'].apply(lambda x: x.split())
tokenized_tweet.head()
```

Out[16]:

```
0      [when, father, dysfunctional, selfish, drags, kids, int
o, dysfunction, #run]
1      [thanks, #lyft, credit, cause, they, offer, wheelchair, vans, #disap
ointed, #getthanked]
2      [bihday, your, majesty]
-      -      -      -
```

```

3                                     [#model, lov
e, take, with, time]
4                                     [factsguide, s
ociety, #motivation]
Name: tidy_tweet, dtype: object

```

In [17]:

```

from nltk.stem.porter import *

Stemmer = PorterStemmer()
tokenized_tweet = tokenized_tweet.apply(lambda x: [Stemmer.stem(i) for i

```

In [18]:

```
tokenized_tweet
```

Out[18]:

```

0                                     [when, father, dysfun
ct, selfish, drag, kid, into, dysfunct, #run]
1                                     [thank, #lyft, credit, caus, they, o
ffer, wheelchair, van, #disapoint, #getthank]
2
3 [bihday, your, majesti]
4 [#model, love, take, with, time]
5
6 [factsguid, societi, #motiv]
7
8                                     ...
49154                                [thought, factori, left, right, polaris, #trump,
#uselect, #leadership, #polit, #brexit, #blm]
49155                                [feel, like, mermaid, #hairflip, #neverr
eadi, #formal, #wed, #gown, #dress, #mermaid]
49156                                [#hillari, #campaign, today, #ohio, use, word, like, asset, liab
il, never, onc, #clinton, thee, word, #radic]
49157                                [happi, work, confer, right, mindset, l
ead, cultur, develop, organ, #work, #mindset]
49158                                [song, glad,
free, download, #shoegaz, #newmus, #newsong]
Name: tidy_tweet, Length: 49159, dtype: object

```

In [19]:

```

import nltk
nltk.download('omw-1.4')
from nltk.corpus import wordnet
from nltk.stem.wordnet import WordNetLemmatizer

Lemmatizer = WordNetLemmatizer()
tokenized_tweet = tokenized_tweet.apply(lambda x: [Lemmatizer.lemmatize(i

[nltk_data] Downloading package omw-1.4 to
[nltk_data] C:\Users\rsaipranayreddy\AppData\Roaming\nltk_data...
[nltk_data] Package omw-1.4 is already up-to-date!

```

In [20]:

```
tokenized_tweet
```

Out[20]:

```

0                                     [when, father, dysfun
ct, selfish, drag, kid, into, dysfunct, #run]
1                                     [thank, #lyft, credit, caus, they, o
ffer, wheelchair, van, #disapoint, #getthank]
2
3 [bihday, your, majesti]
4

```

```
[#model, love, take, with, time]
4
[factsguid, societi, #motiv]

...
49154 [thought, factori, left, right, polaris, #trump,
#uselect, #leadership, #polit, #brexit, #blm]
49155 [feel, like, mermaid, #hairflip, #neverr
eadi, #formal, #wed, #gown, #dress, #mermaid]
49156 [#hillari, #campaign, today, #ohio, use, word, like, asset, liab
il, never, onc, #clinton, thee, word, #radic]
49157 [happi, work, confer, right, mindset, l
ead, cultur, develop, organ, #work, #mindset]
49158 [song, glad,
free, download, #shoegaz, #newmus, #newsong]
Name: tidy_tweet, Length: 49159, dtype: object
```

```
In [21]: for i in range(len(tokenized_tweet)):
        tokenized_tweet[i] = ' '.join(tokenized_tweet[i])
        combine['tidy_tweet'] = tokenized_tweet
```

```
In [22]: tokenized_tweet
```

```
Out[22]: 0 when father dysfunct selfis
h drag kid into dysfunct #run
1 thank #lyft credit caus they offer wheelc
hair van #disapoint #getthank
2
bihday your majesti
3
#model love take with time
4
factsguid societi #motiv

...
49154 thought factori left right polaris #trump #uselect #l
eadership #polit #brexit #blm
49155 feel like mermaid #hairflip #neverreadi #form
al #wed #gown #dress #mermaid
49156 #hillari #campaign today #ohio use word like asset liabil never
onc #clinton thee word #radic
49157 happi work confer right mindset lead cultur
develop organ #work #mindset
49158 song glad free downl
oad #shoegaz #newmus #newsong
Name: tidy_tweet, Length: 49159, dtype: object
```

```
In [23]: combine
```

```
Out[23]:
```

	id	label	tweet	tidy_tweet
0	1	0.0	@user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction. #run	when father dysfunct selfish drag kid into dysfunct #run
1	2	0.0	@user @user thanks for #lyft credit i can't use cause they don't offer wheelchair vans in pdx. #disappointed #getthank	thank #lyft credit caus they offer wheelchair van #disapoint #getthank
2	3	0.0	bihdav vour maiestv	bihdav vour maiesti



...	...	...	...	...
3	4	0.0	#model i love u take with u all the time in urðŸ"±!!! ðŸ~™ðŸ~ŽðŸ„ðŸ'...ðŸ'!ðŸ'!ðŸ'!	#model love take with time
4	5	0.0	factsguide: society now #motivation	factsguid societi #motiv
...	...	...	...	...
49154	49155	NaN	thought factory: left-right polarisation! #trump #uselections2016 #leadership #politics #brexit #blm &gt;3	thought factori left right polaris #trump #uselect #leadership #polit #brexit #blm
49155	49156	NaN	feeling like a mermaid ðŸ~~ #hairflip #neverready #formal #wedding #gown #dresses #mermaid â€;	feel like mermaid #hairflip #neverreadi #formal #wed #gown #dress #mermaid
49156	49157	NaN	#hillary #campaigned today in #ohio((omg)) & used words like "assets&liability" never once did #clinton say thee(word) #radicalization	#hillary #campaign today #ohio use word like asset liabil never onc #clinton thee word #radic
49157	49158	NaN	happy, at work conference: right mindset leads to culture-of-development organizations #work #mindset	happi work confer right mindset lead cultur develop organ #work #mindset
49158	49159	NaN	my song "so glad" free download! #shoegaze #newmusic #newsong	song glad free download #shoegaz #newmus #newsong

49159 rows x 4 columns

In [24]:

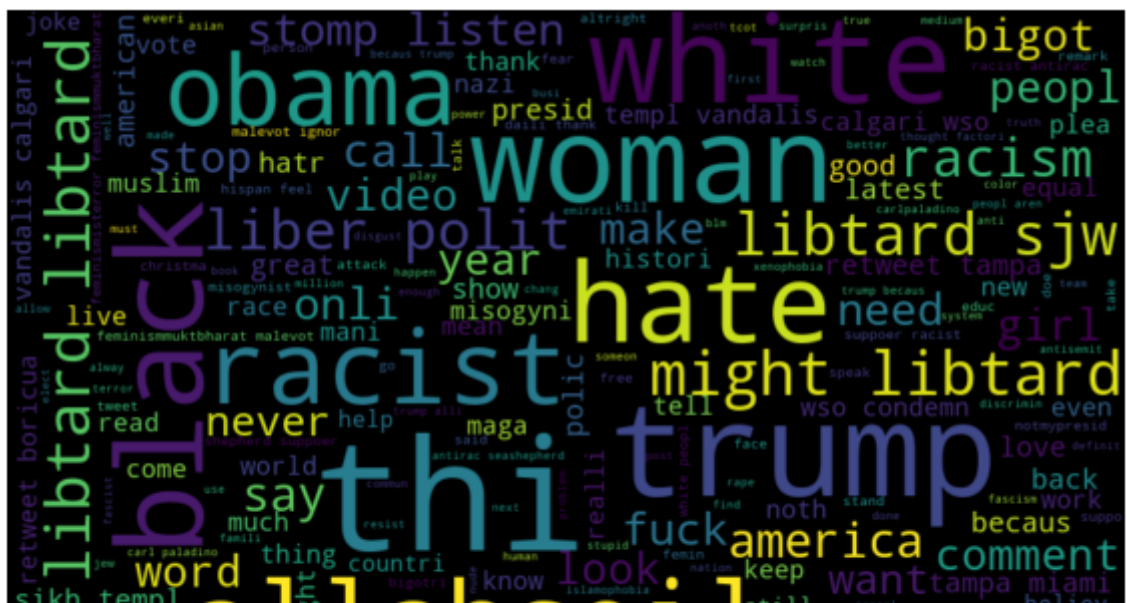
```
from wordcloud import WordCloud
all_words = ' '.join([text for text in combine['tidy_tweet']])
wordcloud = WordCloud(width= 800, height = 500, random_state = 21, max_fo
plt.figure(figsize=(10,7))
plt.imshow(wordcloud, interpolation = "bilinear")
plt.axis('off')
plt.show()
```



```
normal_words = ' '.join([text for text in combine['tidy_tweet']][combine['wordcloud = WordCloud(width= 800, height = 500, random_state = 21, max_fr
plt.figure(figsize=(10,7))
plt.imshow(wordcloud, interpolation = "bilinear")
plt.axis('off')
plt.show()
```



```
negative_words = ' '.join([text for text in combine['tidy_tweet'][combine
wordcloud = WordCloud(width= 800, height = 500, random_state = 21, max_fc
plt.figure(figsize=(10,7))
plt.imshow(wordcloud, interpolation = "bilinear")
plt.axis('off')
plt.show()
```





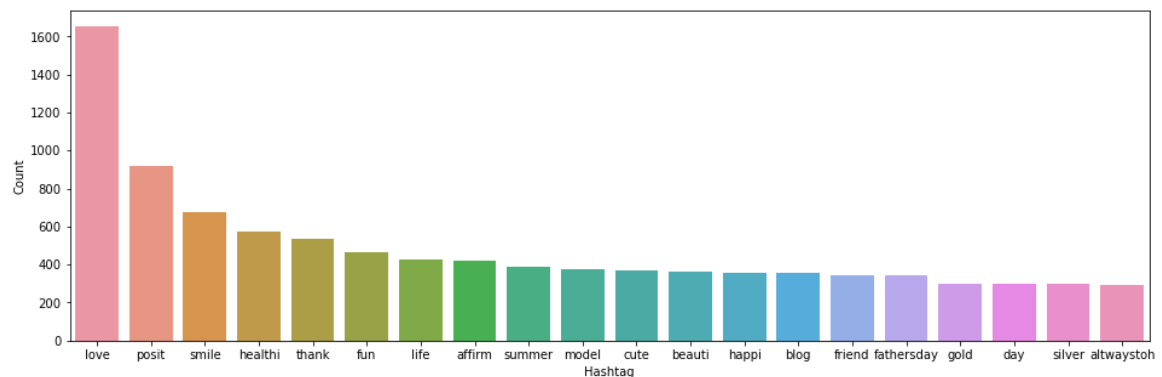
```
In [27]: def hashtag_extract(x):
          hashtags = []
          for i in x:
              ht = re.findall(r"#(\w+)", i)
              hashtags.append(ht)
          return hashtags
```

```
In [28]: HT_regular = hashtag_extract(combine['tidy_tweet'][combine['label'] == 0])
          HT_regular = sum(HT_regular,[])
```

```
In [29]: HT_negative = hashtag_extract(combine['tidy_tweet'][combine['label'] == 1])
          HT_negative = sum(HT_negative,[])
```

```
In [30]: a = nltk.FreqDist(HT_regular)
          d = pd.DataFrame({'Hashtag': list(a.keys()), 'Count': list(a.values())})

          d = d.nlargest(columns = "Count", n = 20)
          plt.figure(figsize = (16,5))
          ax = sns.barplot(data = d, x = "Hashtag", y = "Count")
          ax.set(ylabel = 'Count')
          plt.show()
```



```
In [31]: b = nltk.FreqDist(HT_negative)
          d = pd.DataFrame({'Hashtag': list(b.keys()), 'Count': list(b.values())})

          d = d.nlargest(columns = "Count", n = 20)
          plt.figure(figsize = (16,5))
          ax = sns.barplot(data = d, x = "Hashtag", y = "Count")
          ax.set(ylabel = 'Count')
          plt.show()
```



