

```
In [1]: 1 import re
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5 import seaborn as sns
6 import string
7 import nltk
8 import warnings
9 warnings.filterwarnings("ignore", category=DeprecationWarning)
10
11 %matplotlib inline
```

Read in training set of tweets

```
In [2]: 1 train = pd.read_csv('final_impossible_text_training_tweets.csv')
```

```
In [3]: 1 train
```

Out[3]:

	url	date_and_time	tweet	tweet_id	reply_count	retweet_count	like_count	lang	negative	neutral	positive	compound
0	https://twitter.com/Kkh291/status/140275690219...	2021-06-09 22:38:02	@me1stVegan2nd They have 2 make a smaller vers...	1402756902191210497	0	0	0	en	0.079	0.726	0.195	0.6249
1	https://twitter.com/VictoriaH1962/status/14027...	2021-06-09 22:37:37	@brooklin0000 @DavidMazzieAE @catfishyak Impos...	1402756797979582465	2	0	2	en	0.000	1.000	0.000	0.0000
2	https://twitter.com/NlIteNinc2/status/14027525...	2021-06-09 22:20:39	@aminorjourney 🐔\n\nMake a fake hen out of Imp...	1402752527892766724	1	0	0	en	0.100	0.769	0.131	0.3612
3	https://twitter.com/SonidoMatinal/status/14027...	2021-06-09 22:19:59	🔔 Impossible Burger, la hamburguesa sin carne ...	1402752359986499584	0	0	0	ca	0.231	0.769	0.000	-0.5574
4	https://twitter.com/PanXchange/status/14027504...	2021-06-09 22:12:31	In case you missed our recent #blog post, chec...	1402750481257033731	0	1	0	en	0.099	0.901	0.000	-0.2960
...
29995	https://twitter.com/e4rthmover/status/13057646...	2020-09-15 07:05:27	Lomi is officially the impossible burger https...	1305764650080968704	0	0	0	en	0.000	1.000	0.000	0.0000
29996	https://twitter.com/edgesportnfit/status/13057...	2020-09-15 06:36:45	@michaelharriot Biggie Smalls doing remix vers...	1305757428101730304	0	0	0	en	0.000	1.000	0.000	0.0000
29997	https://twitter.com/cennsith/status/1305736363...	2020-09-15 05:13:03	@t_white_no12 You're a good dude torn, I promis...	1305736363388604417	1	0	0	en	0.000	0.794	0.206	0.8100
29998	https://twitter.com/fathueyfreeman/status/1305...	2020-09-15 05:12:02	i can get a impossible/turkey/chicken burger a...	1305736107347378176	0	0	0	en	0.000	0.841	0.159	0.2235
29999	https://twitter.com/Oinkpoolooenga/status/1305...	2020-09-15 05:05:34	@saqib_z @nsfwllpisces @cryst6l I never said ...	1305734480225730561	1	0	0	en	0.038	0.908	0.054	-0.0387

30000 rows x 12 columns

```
In [4]: 1 train_original=train.copy()
```

Read in tweets for test set

```
In [5]: 1 test = pd.read_csv('impossible_text_training_tweets.csv', skiprows=range(1, 30001))
```

```
In [6]: 1 test.head()
```

Out[6]:

	url	date_and_time	tweet	tweet_id	reply_count	retweet_count	like_count	lang
0	https://twitter.com/sucertamere/status/1305722...	2020-09-15 04:19:23	@Stardogkilledme Thoughts on the impossible bu...	1305722857654226945	0	0	0	en
1	https://twitter.com/chinchlady701/status/13057...	2020-09-15 03:59:43	@jasminelydia17 the thing is one place can hav...	1305717912125218816	0	0	2	en
2	https://twitter.com/coffieluvr/status/13057167...	2020-09-15 03:55:13	@syluwuv RIP WENDYS .. order an impossible bur...	1305716778207719425	0	0	0	en
3	https://twitter.com/BlobCostas/status/13057162...	2020-09-15 03:53:04	@nymillenials What the hell? I ordered an impo...	1305716237318545410	1	0	2	en
4	https://twitter.com/Kirra_Whatever_/status/130...	2020-09-15 03:35:57	the lady at burger king just asked me if i wan...	1305711929219055618	0	0	1	en

```
In [7]: 1 test_original=test.copy()
```

Save test set of tweets (to be used in part II)

```
In [8]: 1 test_original.to_csv('test_tweets.csv', index=False)
```

Combine training set and test set

```
In [9]: 1 combine = train.append(test,ignore_index=True)
```

```
In [10]: 1 combine
```

Out[10]:

	url	date_and_time	tweet	tweet_id	reply_count	retweet_count	like_count	lang	negative	neutral	positive	compound
0	https://twitter.com/Kkh291/status/140275690219...	2021-06-09 22:38:02	@me1stVegan2nd They have 2 make a smaller vers...	1402756902191210497	0	0	0	en	0.079	0.726	0.195	0.6249
1	https://twitter.com/VictoriaH1962/status/14027...	2021-06-09 22:37:37	@brooklin0000 @DavidMazzieAE @catfishyak Impos...	1402756797979582465	2	0	2	en	0.000	1.000	0.000	0.0000
2	https://twitter.com/NliteNinc2/status/14027525...	2021-06-09 22:20:39	@aminorjourney 🥰\n\nMake a fake hen out of Imp...	1402752527892766724	1	0	0	en	0.100	0.769	0.131	0.3612
3	https://twitter.com/SonidoMatinal/status/14027...	2021-06-09 22:19:59	🔔 Impossible Burger, la hamburguesa sin carne ...	1402752359986499584	0	0	0	ca	0.231	0.769	0.000	-0.5574
4	https://twitter.com/PanXchange/status/14027504...	2021-06-09 22:12:31	In case you missed our recent #blog post, chec...	1402750481257033731	0	1	0	en	0.099	0.901	0.000	-0.2960
...
39994	https://twitter.com/DerQuotenossi/status/14027...	2021-06-09 22:56:14	@maulendemiri ihre Impossible Burger werden au...	1402761483759624192	1	0	1	de	NaN	NaN	NaN	NaN
39995	https://twitter.com/estebanj3/status/14027590...	2021-06-09 22:46:24	@thehauer Like an impossible burger?	1402759010479181827	1	0	1	en	NaN	NaN	NaN	NaN
39996	https://twitter.com/WilmaDickfit6/status/14027...	2021-06-09 22:45:24	@cerebralsymphoy @AnimalJustice6 @AlanAlan5240...	1402758758716092419	0	0	0	en	NaN	NaN	NaN	NaN
39997	https://twitter.com/xiancommie/status/14027585...	2021-06-09 22:44:24	@TheAmberPicota Yeah, I was pretty impressed b...	1402758505627455488	1	0	1	en	NaN	NaN	NaN	NaN
39998	https://twitter.com/TheAmberPicota/status/1402...	2021-06-09 22:40:07	The main thing I love about eating vegan is th...	1402757429616648199	3	0	13	en	NaN	NaN	NaN	NaN

39999 rows × 12 columns

Remove Twitter handles

```
In [11]: 1 def remove_pattern(text,pattern):
2
3         r = re.findall(pattern,text)
4
5         for i in r:
6             text = re.sub(i,"",text)
7
8         return text
```

```
In [12]: 1 combine['Tidy_Tweets'] = np.vectorize(remove_pattern)(combine['tweet'], "@[\w]*")
```

Remove punctuation, numbers, special characters

```
In [13]: 1 combine['Tidy_Tweets'] = combine['Tidy_Tweets'].str.replace("[^a-zA-Z#]", " ")

<ipython-input-13-cd036dace8e1>:1: FutureWarning: The default value of regex will change from True to False in a future version.
  combine['Tidy_Tweets'] = combine['Tidy_Tweets'].str.replace("[^a-zA-Z#]", " ")
```

Remove short words

```
In [14]: 1 combine['Tidy_Tweets'] = combine['Tidy_Tweets'].apply(lambda x: ' '.join([w for w in x.split() if len(w)>3]))
```

Tokenize tweets

```
In [15]: 1 tokenized_tweet = combine['Tidy_Tweets'].apply(lambda x: x.split())
```

Stem tweets

```
In [17]: 1 from nltk import PorterStemmer
2 ps = PorterStemmer()
```

```
In [18]: 1 tokenized_tweet = tokenized_tweet.apply(lambda x: [ps.stem(i) for i in x])
```

Recombine tokens

```
In [19]: 1 for i in range(len(tokenized_tweet)):
2         tokenized_tweet[i] = ' '.join(tokenized_tweet[i])
```

```
In [20]: 1 combine['Tidy_Tweets'] = tokenized_tweet
```

Tidy Tweets

In [21]:

1 combine

Out[21]:

	url	date_and_time	tweet	tweet_id	reply_count	retweet_count	like_count	lang	negative	neutral	positive	compound	Tidy_Tweets
0	https://twitter.com/Kkh291/status/140275690219...	2021-06-09 22:38:02	@me1stVegan2nd They have 2 make a smaller vers...	1402756902191210497	0	0	0	en	0.079	0.726	0.195	0.6249	they have make smaller version imposs burger s...
1	https://twitter.com/VictoriaH1962/status/14027...	2021-06-09 22:37:37	@brooklin0000 @DavidMazzieAE @catfishyak impos...	1402756797979582465	2	0	2	en	0.000	1.000	0.000	0.0000	imposs burger
2	https://twitter.com/NliteNinc2/status/14027525...	2021-06-09 22:20:39	@aminorjourney 🐔\n\nMake a fake hen out of Imp...	1402752527892766724	1	0	0	en	0.100	0.769	0.131	0.3612	make fake imposs burger meat coyot goe chicken...
3	https://twitter.com/SonidoMatinal/status/14027...	2021-06-09 22:19:59	🔔 Impossible Burger, la hamburguesa sin carne ...	1402752359986499584	0	0	0	ca	0.231	0.769	0.000	-0.5574	imposs burger hamburguesa carn creada laborato...
4	https://twitter.com/PanXchange/status/14027504...	2021-06-09 22:12:31	In case you missed our recent #blog post, chec...	1402750481257033731	0	1	0	en	0.099	0.901	0.000	-0.2960	case miss recent #blog post check #hemp probab...
...
39994	https://twitter.com/DerQuotenossi/status/14027...	2021-06-09 22:56:14	@maulendemiri ihre Impossible Burger werden au...	1402761483759624192	1	0	1	de	NaN	NaN	NaN	NaN	ihr imposs burger werden selben grill rinderbu...
39995	https://twitter.com/estebanj3/status/14027590...	2021-06-09 22:46:24	@thehauer Like an impossible burger?	1402759010479181827	1	0	1	en	NaN	NaN	NaN	NaN	like imposs burger
39996	https://twitter.com/WilmaDickfit6/status/14027...	2021-06-09 22:45:24	@cerebralsymphoy @AnimalJustice6 @AlanAlan5240...	1402758758716092419	0	0	0	en	NaN	NaN	NaN	NaN	should backsid honey look like imposs burger p...
39997	https://twitter.com/xiancommie/status/14027585...	2021-06-09 22:44:24	@TheAmberPicota Yeah, I was pretty impressed b...	1402758505627455488	1	0	1	en	NaN	NaN	NaN	NaN	yeah pretti impress their imposs burger
39998	https://twitter.com/TheAmberPicota/status/1402...	2021-06-09 22:40:07	The main thing I love about eating vegan is th...	1402757429616648199	3	0	13	en	NaN	NaN	NaN	NaN	main thing love about eat vegan that give damn...

39999 rows x 13 columns

In [22]:

1 from wordcloud import WordCloud
2 from PIL import Image
3 import urllib
4 import requests

In [32]:

1 stopwords = ['imposs burger', 'imposs', 'burger', 'beyond', 'beyond burger', 'http', 'thi', 'carn', 'that', 'hamburguesa', 'impossibleburg', 'tri', 'they']

Create word cloud of tweets with positive compound sentiment

In [33]:

1 all_words_positive = ' '.join(text for text in combine['Tidy_Tweets'][combine['compound']>0])

In [34]:

1 wc_positive = WordCloud(background_color='white', height=1500, width=4000, stopwords=stopwords).generate(all_words_positive)

[illegible]

```
Out[36]: <wordcloud.wordcloud.WordCloud at 0x7fed23442cd0>
```

```
In [37]: 1 all_words_negative = ' '.join(text for text in combine['Tidy_Tweets'][combine['compound']<0])
```

```
In [37]: 1 all_words_negative = ' '.join(text for text in combine['Tidy_Tweets'][combine['compound']<0])
```

In [38]:

In [39]:



In [40]:

Out[40]:

In [41]:

Extract hashtags from tweets with positive compound sentiment

In [42]:


```
In [51]: 1 ht_negative_unnest

Out[51]: ['blog',
'hemp',
'impossibleburg',
'burgerk',
'impossibleburg',
'impossibleburg',
'burger',
'food',
'impossibleburg',
'artoftheburg',
'themetropolitanmuseum',
'themetro',
'dupontcircl',
'willheeverreturn',
'doghau',
'burgerk',
'mcdonald',
'maggi',
'noodl',
..]
```

Frequency of hashtags from tweets with positive compound sentiment

```
In [52]: 1 word_freq_positive = nltk.FreqDist(ht_positive_unnest)

In [53]: 1 word_freq_positive

Out[53]: FreqDist({'impossibleburg': 469, 'vegan': 138, 'plantbas': 93, 'burger': 73, 'imposs': 52, 'vegetarian': 50, 'podcast': 48, 'marri': 48, 'chat': 48, 'life': 48, ...})

In [54]: 1 df_positive = pd.DataFrame({'Hashtags':list(word_freq_positive.keys()), 'Count':list(word_freq_positive.values())})

In [65]: 1 sorted_df_positive = df_positive.sort_values(by='Count', ascending=False)

In [66]: 1 sorted_df_positive

Out[66]:
```

	Hashtags	Count
8	impossibleburg	469
15	vegan	138
14	plantbas	93
1	burger	73
0	imposs	52
...
583	makingmemori	1
582	bt	1
581	mothersday	1
580	veganfoodi	1
1481	cancelledaf	1

1482 rows × 2 columns

Frequency of hashtags from tweets with negative compound sentiment

```
In [69]: 1 word_freq_negative = nltk.FreqDist(ht_negative_unnest)

In [70]: 1 word_freq_negative

Out[70]: FreqDist({'gmo': 108, 'impossibleburg': 93, 'vegan': 40, 'plantbas': 23, 'wewereher': 19, 'takeact': 17, 'burger': 14, 'vegetarian': 14, 'burgerk': 11, 'imposs': 10, ...})

In [71]: 1 df_negative = pd.DataFrame({'Hashtags':list(word_freq_negative.keys()), 'Count':list(word_freq_negative.values())})

In [72]: 1 sorted_df_negative = df_negative.sort_values(by='Count', ascending=False)
```

```
In [74]: 1 sorted_df_negative

Out[74]:
```

	Hashtags	Count
84	gmo	108
2	impossibleburg	93
49	vegan	40
40	plantbas	23
399	wewereher	19
...
181	radio	1
180	dotheimpossiblelav	1
179	burnvegan	1
178	supercarnivor	1
500	realfood	1

501 rows x 2 columns

```
In [77]: 1 pip install dataframe_image

In [78]: 1 import dataframe_image as dfi

In [80]: 1 dfi.export(sorted_df_positive, 'impossible_df_positive.png', max_rows=30)

In [81]: 1 dfi.export(sorted_df_negative, 'impossible_df_negative.png', max_rows=30)
```