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

Read in training set of tweets

```
In [2]:   1  train = pd.read_csv('final_beyond_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/libraryjogirl/status/14357...	2021-09-08 23:57:23	@AWCanada Hi A&W! I'm seeing commercials f...	1435754169965613057	1	0	0	en	0.000	0.912	0.088	0.4738
1	https://twitter.com/eliseelara/status/14357513...	2021-09-08 23:46:04	their supreme sauce?!? THEIR SEASONED FRIES?!?...	1435751320439361543	0	0	0	en	0.097	0.723	0.180	0.5615
2	https://twitter.com/t_kelly15/status/143574800...	2021-09-08 23:32:53	Someone accidentally sent their A&W dinner...	1435748002925326336	1	0	5	en	0.130	0.870	0.000	-0.5859
3	https://twitter.com/marlenemdiaz08/status/1435...	2021-09-08 23:24:47	Like beyond nasty for a fast food joint. No ke...	1435745963189014532	0	0	0	en	0.242	0.690	0.067	-0.9083
4	https://twitter.com/divinaxo/status/1425869375...	2021-08-12 17:18:45	Beyond burger 😋 😋 😋 https://t.co/eiqlclnsnQ	1425869375941267463	0	1	1	en	0.000	1.000	0.000	0.0000
...	...	...	...	...	...	...	...	...	...	...	...	...
29995	https://twitter.com/_HiZel_/status/13277190531...	2020-11-14 21:04:24	Tell me why I called a Beyond burger an Infini...	1327719053171933184	0	0	7	en	0.000	1.000	0.000	0.0000
29996	https://twitter.com/400lez/status/132771789520...	2020-11-14 20:59:48	feel like absolute garbage. just want a beyond...	1327717895208460288	0	0	5	en	0.000	0.648	0.352	0.4215
29997	https://twitter.com/njroute22/status/132771443...	2020-11-14 20:46:03	More Reasons to Avoid Beyond Meat Fake Food ht...	1327714432378695683	0	0	0	en	0.218	0.782	0.000	-0.8865
29998	https://twitter.com/aviosAdventurer/status/132...	2020-11-14 20:40:46	@GazRich88 I'm trying to get my hands on some ...	1327713105720336385	1	0	1	en	0.000	0.870	0.130	0.6369
29999	https://twitter.com/ReginaBanali/status/132771...	2020-11-14 20:30:14	Beyond Meat Reveals It Is Behind the McDonald'...	1327710455666790401	0	0	0	en	0.000	1.000	0.000	0.0000

30000 rows × 12 columns

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

Read in tweets for test set

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

```
In [6]:   1  test
```

Out[6]:

	url	date_and_time	tweet	tweet_id	reply_count	retweet_count	like_count	lang
0	https://twitter.com/PathanShekib/status/132770...	2020-11-14 20:01:34	@aneelfassa beyond meat burgers are diff	1327703240729796609	0	0	0	en
1	https://twitter.com/galwaybae/status/132770305...	2020-11-14 20:00:50	dennys plant based burgers a close second to b...	1327703053777199106	0	0	0	en
2	https://twitter.com/TheHeroes_HERO/status/1327...	2020-11-14 19:32:53	Who ever said Beyond Meat tastes like meat FUC...	1327696020923420673	0	0	2	en
3	https://twitter.com/cherrymorello/status/13276...	2020-11-14 19:09:58	@feebee79 I've just had a beyond burger with f...	1327690255366053888	0	0	2	en
4	https://twitter.com/downlowbambi/status/132768...	2020-11-14 19:05:29	@themetdtd Yeah it's great that all I have to ...	1327689125328158720	1	0	0	en
...	...	...	...	...	...	...	...	...
9994	https://twitter.com/Intofurler/status/12956694...	2020-08-18 10:30:42	@O0Rocker0O they started selling the beyond bu...	1295669444891615233	1	0	0	en
9995	https://twitter.com/TravelOceans/status/129566...	2020-08-18 10:24:45	Reduce your meat consumption, start by cutting...	1295667947550253056	0	0	0	en
9996	https://twitter.com/TNarinen/status/1295646808...	2020-08-18 09:00:45	@mattimolari Tässähän tämä homma kärjistettynä...	1295646808992096256	0	0	0	fi
9997	https://twitter.com/surf_panda/status/12956346...	2020-08-18 08:12:20	@niksy @NebojsaG @polojaci @josephbt @borisrad...	1295634621636509696	0	0	2	und
9998	https://twitter.com/UPlantATree/status/1295620...	2020-08-18 07:17:11	Just eat salads instead of burgers to save the...	1295620742806228996	0	0	0	en

9999 rows × 8 columns

```
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('beyond_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/libraryjogirl/status/14357...	2021-09-08 23:57:23	@AWCanada Hi A&W! I'm seeing commercials f...	1435754169965613057	1	0	0	en	0.000	0.912	0.088	0.4738
1	https://twitter.com/eliseelara/status/14357513...	2021-09-08 23:46:04	their supreme sauce?!? THEIR SEASONED FRIES?!?...	1435751320439361543	0	0	0	en	0.097	0.723	0.180	0.5615
2	https://twitter.com/t_kelly15/status/143574800...	2021-09-08 23:32:53	Someone accidentally sent their A&W dinner...	1435748002925326336	1	0	5	en	0.130	0.870	0.000	-0.5859
3	https://twitter.com/marlenemdiaz08/status/1435...	2021-09-08 23:24:47	Like beyond nasty for a fast food joint. No ke...	1435745963189014532	0	0	0	en	0.242	0.690	0.067	-0.9083
4	https://twitter.com/divinaxo/status/1425869375...	2021-08-12 17:18:45	Beyond burger 😊 😊 😊 https://t.co/eiqlclnsnQ	1425869375941267463	0	1	1	en	0.000	1.000	0.000	0.0000
...	...	...	...	...	...	...	...	...	...	...	...	...
39994	https://twitter.com/Intofurler/status/12956694...	2020-08-18 10:30:42	@O0Rocker0O they started selling the beyond bu...	1295669444891615233	1	0	0	en	NaN	NaN	NaN	NaN
39995	https://twitter.com/TravelOceans/status/129566...	2020-08-18 10:24:45	Reduce your meat consumption, start by cutting...	1295667947550253056	0	0	0	en	NaN	NaN	NaN	NaN
39996	https://twitter.com/TNarinen/status/1295646808...	2020-08-18 09:00:45	@mattimolari Tässähän tämä homma kärjistettynä...	1295646808992096256	0	0	0	fi	NaN	NaN	NaN	NaN
39997	https://twitter.com/surf_panda/status/12956346...	2020-08-18 08:12:20	@niksy @NebojsaG @polojaci @josephbt @borisrad...	1295634621636509696	0	0	2	und	NaN	NaN	NaN	NaN
39998	https://twitter.com/UPlantATree/status/1295620...	2020-08-18 07:17:11	Just eat salads instead of burgers to save the...	1295620742806228996	0	0	0	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#]", " ")
```

...

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 [16]:  1  from nltk import PorterStemmer
          2  ps = PorterStemmer()
```

```
In [17]:    1  tokenized_tweet = tokenized_tweet.apply(lambda x: [ps.stem(i) for i in x])
```

Recombine tokens

```
In [18]:    1  for i in range(len(tokenized_tweet)):
            2      tokenized_tweet[i] = ' '.join(tokenized_tweet[i])
```

```
In [19]:    1  combine['Tidy_Tweets'] = tokenized_tweet
```

Tidy Tweets

```
In [20]:    1  combine
```

Out[20]:

	url	date_and_time	tweet	tweet_id	reply_count	retweet_count	like_count	lang	negative	neutral	positive	compound	Tidy_Tweets
0	https://twitter.com/libraryjogirl/status/14357...	2021-09-08 23:57:23	@AWCanada Hi A&W! I'm seeing commercials f...	1435754169965613057	1	0	0	en	0.000	0.912	0.088	0.4738	see commerci your beyond meat burger vegetaria...
1	https://twitter.com/eliseelara/status/14357513...	2021-09-08 23:46:04	their supreme sauce?!? THEIR SEASONED FRIES?!?...	1435751320439361543	0	0	0	en	0.097	0.723	0.180	0.5615	their suprem sauc their season fri their beyon...
2	https://twitter.com/t_kelly15/status/143574800...	2021-09-08 23:32:53	Someone accidentally sent their A&W dinner...	1435748002925326336	1	0	5	en	0.130	0.870	0.000	-0.5859	someon accident sent their dinner apart will u...
3	https://twitter.com/marlenemdiaz08/status/1435...	2021-09-08 23:24:47	Like beyond nasty for a fast food joint. No ke...	1435745963189014532	0	0	0	en	0.242	0.690	0.067	-0.9083	like beyond nasti fast food joint ketchup smas...
4	https://twitter.com/divinaxo/status/1425869375...	2021-08-12 17:18:45	Beyond burger 😋😋😋 https://t.co/eiqlclnsnQ	1425869375941267463	0	1	1	en	0.000	1.000	0.000	0.0000	beyond burger http eiqlclnsnq
...	...	...	...	...	...	...	...	...	...	...	...	...	...
39994	https://twitter.com/Intofurler/status/12956694...	2020-08-18 10:30:42	@O0Rocker0O they started selling the beyond bu...	1295669444891615233	1	0	0	en	NaN	NaN	NaN	NaN	they start sell beyond burger supermarket migh...
39995	https://twitter.com/TravelOceans/status/129566...	2020-08-18 10:24:45	Reduce your meat consumption, start by cutting...	1295667947550253056	0	0	0	en	NaN	NaN	NaN	NaN	reduc your meat consumpt start cut beef then s...
39996	https://twitter.com/TNarinen/status/1295646808...	2020-08-18 09:00:45	@mattimolari Tässähän tämä homma kärjistettynä...	1295646808992096256	0	0	0	fi	NaN	NaN	NaN	NaN	homma rjistettyn miksei siell esim beyond burg...
39997	https://twitter.com/surf_panda/status/12956346...	2020-08-18 08:12:20	@niksy @NebojsaG @polojaci @josephbt @borisrad...	1295634621636509696	0	0	2	und	NaN	NaN	NaN	NaN	beyond burger lar sven
39998	https://twitter.com/UPlantATree/status/1295620...	2020-08-18 07:17:11	Just eat salads instead of burgers to save the...	1295620742806228996	0	0	0	en	NaN	NaN	NaN	NaN	just salad instead burger save planet imposs b...

39999 rows × 13 columns

```
In [21]:    1  from wordcloud import WordCloud
            2  from PIL import Image
            3  import urllib
            4  import requests
```

```
In [22]:    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 [23]:    1  all_words_positive = ' '.join(text for text in combine['Tidy_Tweets'][combine['compound']>0])
```

```
In [24]:    1  wc_positive = WordCloud(background_color='white', height=1500, width=4000, stopwords=stopwords).generate(all_words_positive)
```

```
In [25]:  1  plt.figure(figsize=(50,50))
          2  plt.imshow(wc_positive)
          3  plt.axis('off')
          4  plt.show()
```



```
In [26]:  1  wc_positive.to_file('beyond_wc_positive.png')
```

Out[26]: <wordcloud.wordcloud.WordCloud at 0x7ff54e0458e0>

Create word cloud of tweets with negative compound sentiment

```
In [27]:  1  all_words_negative = ' '.join(text for text in combine['Tidy_Tweets'][combine['compound']<0])
```

```
In [28]:  1  wc_negative = WordCloud(background_color='white', height=1500, width=4000, stopwords=stopwords).generate(all_words_negative)
```

```
In [29]:   1  plt.figure(figsize=(50,50))
           2  plt.imshow(wc_negative)
           3  plt.axis('off')
           4  plt.show()
```



```
In [31]:   1  wc_negative.to_file('beyond_wc_negative.png')
```

Out[31]: <wordcloud.wordcloud.WordCloud at 0x7ff538b2cd60>

```
In [32]:   1  def Hashtags_Extract(x):
           2      hashtags=[]
           3
           4      for i in x:
           5          ht = re.findall(r'#(\w+)',i)
           6          hashtags.append(ht)
           7
           8      return hashtags
```

Extract hashtags from tweets with positive compound sentiment

```
In [33]:   1  ht_positive = Hashtags_Extract(combine['Tidy_Tweets'][combine['compound']>0])
```

```
In [34]:   1  ht_positive
```

```
Out[34]: [[],
          [],
          [],
          [],
          ['moobyspopup'],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          ['comfortfood'],
          ['fundrais', 'dine', 'nonprofit'],
          [],
          ..
```

```
In [35]:   1  ht_positive_unnest = sum(ht_positive,[])
```

```
In [36]:   1  ht_positive_unnest
```

```
Out[36]: ['moobyspopup',
          'comfortfood',
          'fundrais',
          'dine',
          'nonprofit',
          'onthisdateinhiphop',
          'odog',
          'larenzt',
          'portlandburgerweek',
          'meatless',
          'burger',
          'notanad',
          'vegan',
          'burger',
          'sausag',
          'plantbas',
          'healthyfood',
          'vegan',
          'dentistemaxwithgulf',
          ..
```

Extract hashtags from tweets with negative compound sentiment

```
In [37]:   1  ht_negative = Hashtags_Extract(combine['Tidy_Tweets'][combine['compound']<0])
```

```
In [38]:   1  ht_negative
```

```
Out[38]: [[],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          [],
          ..
```

```
In [39]:   1  ht_negative_unnest = sum(ht_negative,[])
```

```
In [40]: 1 ht_negative_unnest
```

```
Out[40]: ['ufo',
         'tictac',
         'uap',
         'alien',
         'leburgerweek',
         'beyondburg',
         'plantbas',
         'yeahright',
         'ripoffbritain',
         'essen',
         'porlosanimal',
         'porelplaneta',
         'portusalud',
         'vegan',
         'vegano',
         'plantbas',
         'beyondburg',
         'beyondmeat',
         'followyourheart',
```

Frequency of hashtags from tweets with positive compound sentiment

```
In [41]: 1 word_freq_positive = nltk.FreqDist(ht_positive_unnest)
```

```
In [42]: 1 word_freq_positive
```

```
Out[42]: FreqDist({'vegan': 315, 'beyondburg': 234, 'plantbas': 108, 'veganiseyourmenu': 107, 'beyondmeat': 82, 'burger': 75, 'vegetarian': 56, 'food': 37, 'beyond': 36, 'gobeyond': 26, ...})
```

```
In [43]: 1 df_positive = pd.DataFrame({'Hashtags':list(word_freq_positive.keys()),'Count':list(word_freq_positive.values())})
```

```
In [44]: 1 sorted_df_positive = df_positive.sort_values(by='Count', ascending=False)
```

```
In [45]: 1 sorted_df_positive
```

Out[45]:

	Hashtags	Count
12	vegan	315
46	beyondburg	234
14	plantbas	108
99	veganiseyourmenu	107
44	beyondmeat	82
...	...	...
648	makemoneyonlin	1
647	affiliatemarket	1
646	jasonabalo	1
645	dfwre	1
1636	ohiocraftb	1

1637 rows × 2 columns

Frequency of hashtags from tweets with negative compound sentiment

```
In [46]: 1 word_freq_negative = nltk.FreqDist(ht_negative_unnest)
```

```
In [47]: 1 word_freq_negative
```

```
Out[47]: FreqDist({'beyondburg': 53, 'vegan': 52, 'urbanfantasi': 32, 'scifi': 32, 'recip': 32, 'plantbas': 19, 'beyond': 19, 'food': 14, 'beyondmeat': 13, 'new': 12, ...})
```

```
In [48]: 1 df_negative = pd.DataFrame({'Hashtags':list(word_freq_negative.keys()),'Count':list(word_freq_negative.values())})
```

```
In [49]: 1 sorted_df_negative = df_negative.sort_values(by='Count', ascending=False)
```

```
In [50]:    1  sorted_df_negative
```

Out[50]:

	Hashtags	Count
5	beyondburg	53
13	vegan	52
144	recip	32
143	scifi	32
142	urbanfantasi	32
...	...	...
178	rasselbocklb	1
177	meatabl	1
176	newproduct	1
175	canada	1
480	checker	1

481 rows × 2 columns

```
In [51]:    1  import dataframe_image as dfi
```

```
In [52]:    1  dfi.export(sorted_df_positive, 'beyond_df_positive.png', max_rows=30)
```

```
In [53]:    1  dfi.export(sorted_df_negative, 'beyond_df_negative.png', max_rows=30)
```