

```
In [1]: # sentiment analysis
####
####
####
```

```
In [2]: # VADER is a VALENCE based sentiment analyzer
# Valence based approach taken into consideration the "intensity" of a word as
# opposed to only the polarity (+ve or -ve). For ex. "Great" is more treated as
# more +ve as opposed to "Good"
# suggested scale
# Neutral -0.5 < and > 0.5
# Positive >=0.5
# Negative <=-0.5
```

```
In [3]: import pandas as pd
import numpy as np
%matplotlib inline

import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer

analyser = SentimentIntensityAnalyzer()
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data] /Users/Jacek/nltk_data...
[nltk_data] Package vader_lexicon is already up-to-date!

/Users/Jacek/opt/anaconda3/lib/python3.7/site-packages/nltk/twitter/__init__.
py:20: UserWarning: The twython library has not been installed. Some function
ality from the twitter package will not be available.
  warnings.warn("The twython library has not been installed. "
```

```
In [4]: # Airline Tweets database
sentences = pd.read_csv('TSMMS_class7_Airline_Tweets.csv')
len(sentences)
```

Out[4]: 14640

```
In [5]: # inspect the data
sentences.columns
```

```
Out[5]: Index(['tweet_id', 'airline_sentiment', 'airline_sentiment_confidence',
               'negativereason', 'negativereason_confidence', 'airline',
               'airline_sentiment_gold', 'name', 'negativereason_gold',
               'retweet_count', 'text', 'tweet_coord', 'tweet_created',
               'tweet_location', 'user_timezone'],
              dtype='object')
```

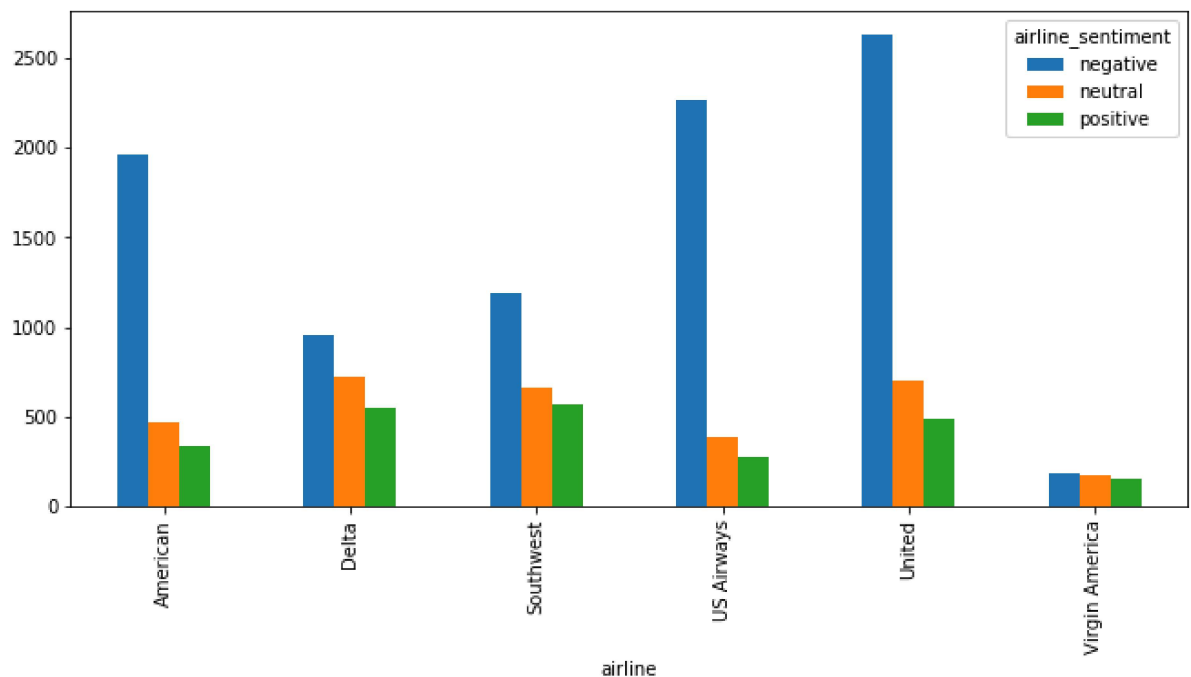
```
In [6]: sentences.head()
```

Out[6]:

	tweet_id	airline_sentiment	airline_sentiment_confidence	negativereason	negativerc
0	570306133677760513	neutral	1.0000	NaN	
1	570301130888122368	positive	0.3486	NaN	
2	570301083672813571	neutral	0.6837	NaN	
3	570301031407624196	negative	1.0000	Bad Flight	
4	570300817074462722	negative	1.0000	Can't Tell	

```
In [7]: # compare the airlines
sentences.groupby(['airline', 'airline_sentiment']).size().unstack().plot(kind='bar',figsize=(11, 5))
```

Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc12c3c9c90>



```
In [8]: # how about the tweets?
sentences = sentences[['airline_sentiment', 'airline', 'text' ]]
sentences.head()
```

```
Out[8]:
```

	airline_sentiment	airline	text
0	neutral	Virgin America	@VirginAmerica What @dhepburn said.
1	positive	Virgin America	@VirginAmerica plus you've added commercials t...
2	neutral	Virgin America	@VirginAmerica I didn't today... Must mean I n...
3	negative	Virgin America	@VirginAmerica it's really aggressive to blast...
4	negative	Virgin America	@VirginAmerica and it's a really big bad thing...

```
In [9]: #filtering dataset for United
sentences = sentences[sentences['airline']=='United']
print(len(sentences))
sentences = sentences.reset_index(drop = True)
sentences.head(10)
```

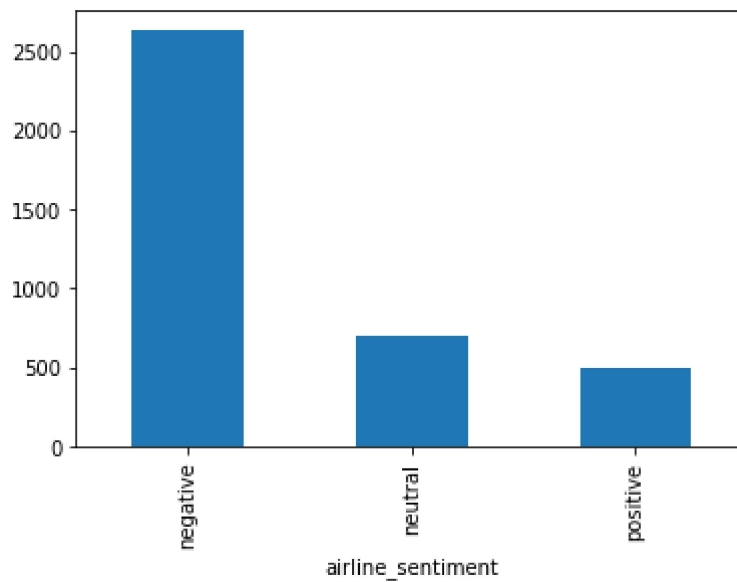
3822

```
Out[9]:
```

	airline_sentiment	airline	text
0	positive	United	@united thanks
1	positive	United	@united Thanks for taking care of that MR!! Ha...
2	negative	United	@united still no refund or word via DM. Please...
3	negative	United	@united Delayed due to lack of crew and now de...
4	positive	United	@united thanks -- we filled it out. How's our ...
5	negative	United	@united Your ERI-ORD express connections are h...
6	neutral	United	@united even on international flight Glasgow t...
7	negative	United	@united you think you boarded flight AU1066 to...
8	neutral	United	@united I understand, but it's tough when ther...
9	negative	United	@united Gate agent hooked me up with alternate...

```
In [10]: sentences.groupby('airline_sentiment').size().plot(kind='bar')
```

```
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc129885f90>
```



```
In [11]: # move to sentiment analysis
def print_sentiment_scores(sentence):
    snt = analyser.polarity_scores(sentence) #call the polarity analyzer
    print("{:-<40} {}".format(sentence, str(snt)))
```

```
In [12]: # compound value scale = -1 to 1 (-ve to +ve)
print_sentiment_scores("United flight was a bad experience")
```

```
United flight was a bad experience----- {'neg': 0.376, 'neu': 0.323, 'pos': 0.301, 'compound': -0.1779}
```

```
In [13]: %time # to calculate the time it takes the algorithm to compute a VADER score
i=0 # counter

compval1 = [ ] # empty list to hold our computed 'compound' VADER scores

while (i<len(sentences)):

    k = analyser.polarity_scores(sentences.iloc[i]['text'])
    compval1.append(k['compound'])

    i = i+1

# convert sentiment values to numpy for easier usage
compval1 = np.array(compval1)
len(compval1)
```

```
CPU times: user 2 µs, sys: 0 ns, total: 2 µs
Wall time: 5.25 µs
```

```
Out[13]: 3822
```

```
In [14]: sentences['VADER score'] = compval1
```

```
In [15]: # how it looks like?
sentences.head(30)
```

Out[15]:

	airline_sentiment	airline	text	VADER score
0	positive	United	@united thanks	0.4404
1	positive	United	@united Thanks for taking care of that MR!! Ha...	0.8856
2	negative	United	@united still no refund or word via DM. Please...	-0.2732
3	negative	United	@united Delayed due to lack of crew and now de...	-0.2960
4	positive	United	@united thanks -- we filled it out. How's our ...	0.7399
5	negative	United	@united Your ERI-ORD express connections are h...	0.7495
6	neutral	United	@united even on international flight Glasgow t...	0.0000
7	negative	United	@united you think you boarded flight AU1066 to...	0.0000
8	neutral	United	@united I understand, but it's tough when ther...	-0.8957
9	negative	United	@united Gate agent hooked me up with alternate...	0.2103
10	neutral	United	@united flying gla-mco in a few weeks. How lon...	0.0000
11	positive	United	@united thanks	0.4404
12	positive	United	@united Thanks. It is on the same ticket.	0.4404
13	neutral	United	@united why not? Is it a law or a policy?	0.0000
14	negative	United	@united is the worst. Nonrefundable First clas...	-0.6249
15	negative	United	@united @CheerTymeDad So I can buy tix 3 days ...	0.4767
16	negative	United	@united I did start a claim but 8-10 weeks is ...	0.0000
17	positive	United	@united follow me please this airline is beaut...	0.3182
18	neutral	United	@united no I don't, but I'm sure United have m...	0.1672
19	negative	United	@united if the car seat is lost @united should...	-0.3182
20	neutral	United	@united Just submitted my response on the link...	0.0000
21	neutral	United	@united yes, David Allan send an email with th...	0.4723
22	negative	United	@united @highbuddyyy that totally sucks my cou...	-0.4201
23	positive	United	@united just wanted to let you know how wonder...	0.5719
24	neutral	United	@united yes, a paper voucher that I got on Jan...	0.4019
25	negative	United	@united beginning of Feb I called United they ...	0.4215
26	negative	United	@United the internet is a great thing. I am e...	0.6249
27	neutral	United	@united Do I need to use a Chase United Club p...	0.4215
28	neutral	United	@united and what am I suppose to do with that ...	0.0772
29	negative	United	@united thats weak. See ya 🙌\nHey @VirginAmeri...	-0.5399

```
In [16]: %time

# assign score categories and logic
i = 0

predicted_value = [ ] # empty series to hold our predicted values

while(i<len(sentences)):
    if ((sentences.iloc[i]['VADER score'] >= 0.7)):
        predicted_value.append('positive')
        i = i+1
    elif ((sentences.iloc[i]['VADER score'] > 0) & (sentences.iloc[i]['VADER s
core'] < 0.7)):
        predicted_value.append('neutral')
        i = i+1
    elif ((sentences.iloc[i]['VADER score'] <= 0)):
        predicted_value.append('negative')
        i = i+1
```

CPU times: user 3 µs, sys: 1 µs, total: 4 µs
Wall time: 5.25 µs

```
In [17]: sentences['predicted sentiment'] = predicted_value
```

```
In [18]: len(sentences['predicted sentiment'])
```

Out[18]: 3822

In [19]: `sentences.head(30)`

Out[19]:

	airline_sentiment	airline	text	VADER score	predicted sentiment
0	positive	United	@united thanks	0.4404	neutral
1	positive	United	@united Thanks for taking care of that MR!! Ha...	0.8856	positive
2	negative	United	@united still no refund or word via DM. Please...	-0.2732	negative
3	negative	United	@united Delayed due to lack of crew and now de...	-0.2960	negative
4	positive	United	@united thanks -- we filled it out. How's our ...	0.7399	positive
5	negative	United	@united Your ERI-ORD express connections are h...	0.7495	positive
6	neutral	United	@united even on international flight Glasgow t...	0.0000	negative
7	negative	United	@united you think you boarded flight AU1066 to...	0.0000	negative
8	neutral	United	@united I understand, but it's tough when ther...	-0.8957	negative
9	negative	United	@united Gate agent hooked me up with alternate...	0.2103	neutral
10	neutral	United	@united flying gla-mco in a few weeks. How lon...	0.0000	negative
11	positive	United	@united thanks	0.4404	neutral
12	positive	United	@united Thanks. It is on the same ticket.	0.4404	neutral
13	neutral	United	@united why not? Is it a law or a policy?	0.0000	negative
14	negative	United	@united is the worst. Nonrefundable First clas...	-0.6249	negative
15	negative	United	@united @CheerTymeDad So I can buy tix 3 days ...	0.4767	neutral
16	negative	United	@united I did start a claim but 8-10 weeks is ...	0.0000	negative
17	positive	United	@united follow me please this airline is beaut...	0.3182	neutral
18	neutral	United	@united no I don't, but I'm sure United have m...	0.1672	neutral
19	negative	United	@united if the car seat is lost @united should...	-0.3182	negative
20	neutral	United	@united Just submitted my response on the link...	0.0000	negative
21	neutral	United	@united yes, David Allan send an email with th...	0.4723	neutral
22	negative	United	@united @highbuddyyy that totally sucks my cou...	-0.4201	negative
23	positive	United	@united just wanted to let you know how wonder...	0.5719	neutral

	airline_sentiment	airline	text	VADER score	predicted sentiment
24	neutral	United	@united yes, a paper voucher that I got on Jan...	0.4019	neutral
25	negative	United	@united beginning of Feb I called United they ...	0.4215	neutral
26	negative	United	@United the internet is a great thing. I am e...	0.6249	neutral
27	neutral	United	@united Do I need to use a Chase United Club p...	0.4215	neutral
28	neutral	United	@united and what am I suppose to do with that ...	0.0772	neutral
29	negative	United	@united thats weak. See ya 🙌 InHey @VirginAmeri...	-0.5399	negative

```
In [20]: # which classifications were correct?
madeit = sentences[sentences['airline_sentiment'] == sentences['predicted sentiment']]
```

```
In [21]: len(madeit)/len(sentences)
```

```
Out[21]: 0.586865515436944
```

In [22]: `madeit.head(30)`

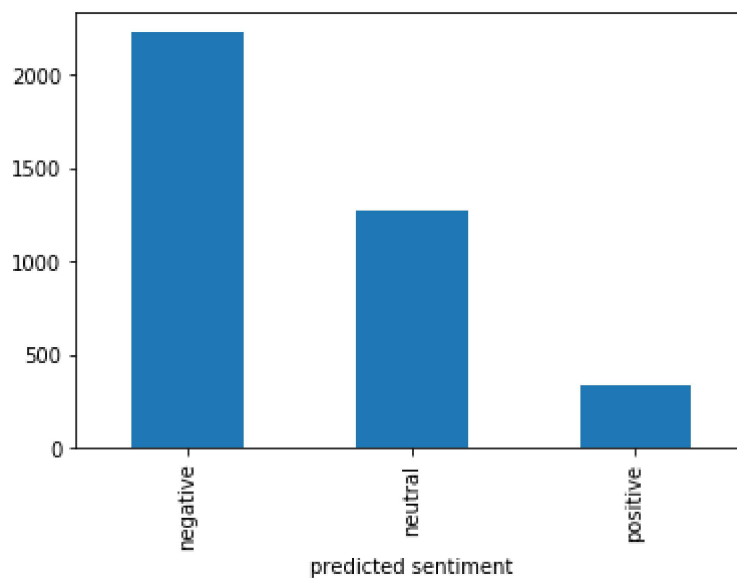
Out[22]:

	airline_sentiment	airline	text	VADER score	predicted sentiment
1	positive	United	@united Thanks for taking care of that MR!! Ha...	0.8856	positive
2	negative	United	@united still no refund or word via DM. Please...	-0.2732	negative
3	negative	United	@united Delayed due to lack of crew and now de...	-0.2960	negative
4	positive	United	@united thanks -- we filled it out. How's our ...	0.7399	positive
7	negative	United	@united you think you boarded flight AU1066 to...	0.0000	negative
14	negative	United	@united is the worst. Nonrefundable First clas...	-0.6249	negative
16	negative	United	@united I did start a claim but 8-10 weeks is ...	0.0000	negative
18	neutral	United	@united no I don't, but I'm sure United have m...	0.1672	neutral
19	negative	United	@united if the car seat is lost @united should...	-0.3182	negative
21	neutral	United	@united yes, David Allan send an email with th...	0.4723	neutral
22	negative	United	@united @highbuddyyy that totally sucks my cou...	-0.4201	negative
24	neutral	United	@united yes, a paper voucher that I got on Jan...	0.4019	neutral
27	neutral	United	@united Do I need to use a Chase United Club p...	0.4215	neutral
28	neutral	United	@united and what am I suppose to do with that ...	0.0772	neutral
29	negative	United	@united thats weak. See ya 🙌InHey @VirginAmeri...	-0.5399	negative
33	negative	United	@united I sure did. I had to drive a total o...	-0.1027	negative
35	negative	United	@united See? We were told repeatedly that the ...	0.0000	negative
37	negative	United	@united #Newarkliberty Airport need to indicat...	0.0000	negative
39	negative	United	@united How does United refuse to reissue a \$4...	-0.1779	negative
40	negative	United	@united also during the run I was instructed t...	-0.4767	negative
43	neutral	United	@united Have clients with an 11 hr layover at ...	0.4939	neutral
46	negative	United	@united iah to charlotte. Baggage claim rep la...	-0.3182	negative
50	negative	United	@united A refund and flight vouchers or anothe...	-0.4005	negative

	airline_sentiment	airline	text	VADER score	predicted sentiment
51	negative	United	@united then why have I not received my call b...	0.0000	negative
55	negative	United	@united No, I need you guys to not over book p...	-0.2960	negative
57	negative	United	@united When will email address/username sign-...	0.0000	negative
59	negative	United	@united rude rude	-0.7184	negative
60	negative	United	@united - Why can't I get a boarding pass for ...	-0.2500	negative
62	negative	United	@united just flew to #TelAviv paid \$100 from a...	-0.4389	negative
63	negative	United	@United My bag is still in Colo Springs. I am...	-0.3412	negative

```
In [23]: sentences.groupby('predicted sentiment').size().plot(kind='bar')
```

```
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc12a6fb610>
```



```
In [24]: # how about wrong predictions?
didntmakeit = sentences[sentences['airline_sentiment'] != sentences['predicted sentiment']]
```

```
In [25]: didntmakeit.reset_index(drop=True, inplace=True)  
         didntmakeit.head(30)
```

Out[25]:

	airline_sentiment	airline	text	VADER score	predicted sentiment
0	positive	United	@united thanks	0.4404	neutral
1	negative	United	@united Your ERI-ORD express connections are h...	0.7495	positive
2	neutral	United	@united even on international flight Glasgow t...	0.0000	negative
3	neutral	United	@united I understand, but it's tough when ther...	-0.8957	negative
4	negative	United	@united Gate agent hooked me up with alternate...	0.2103	neutral
5	neutral	United	@united flying gla-mco in a few weeks. How lon...	0.0000	negative
6	positive	United	@united thanks	0.4404	neutral
7	positive	United	@united Thanks. It is on the same ticket.	0.4404	neutral
8	neutral	United	@united why not? Is it a law or a policy?	0.0000	negative
9	negative	United	@united @CheerTymeDad So I can buy tix 3 days ...	0.4767	neutral
10	positive	United	@united follow me please this airline is beaut...	0.3182	neutral
11	neutral	United	@united Just submitted my response on the link...	0.0000	negative
12	positive	United	@united just wanted to let you know how wonder...	0.5719	neutral
13	negative	United	@united beginning of Feb I called United they ...	0.4215	neutral
14	negative	United	@United the internet is a great thing. I am e...	0.6249	neutral
15	negative	United	@united I flew back w other company since Unit...	0.4215	neutral
16	neutral	United	@united I lost a voucher that I was given in M...	-0.3182	negative
17	negative	United	@united I am trying to book awards for Septemb...	0.7750	positive
18	neutral	United	@united @CheerTymeDad Gee that's like almost c...	0.8180	positive
19	negative	United	@united why no preferred security line anymore...	0.1260	neutral
20	negative	United	@united yes please! I am newly married and try...	0.8528	positive
21	negative	United	@united yes, we've been with the agents for th...	0.8176	positive
22	neutral	United	@united that's right- with an overnight Miami.	0.0000	negative
23	neutral	United	@united I am trying to find out if the loaner ...	-0.2023	negative

	airline_sentiment	airline	text	VADER score	predicted sentiment
24	neutral	United	@united i DMed you the details	0.0000	negative
25	neutral	United	@united Kewl. Will also let them know the over...	0.8555	positive
26	positive	United	@united - thanks for your help...got me what l...	0.2382	neutral
27	negative	United	@united @UCtraveladvisor - I would have loved ...	0.1027	neutral
28	neutral	United	@united thanks again for your concern. I will ...	0.7269	positive
29	negative	United	@United well sitting on the ground 'on time' b...	0.1406	neutral

```
In [26]: # example mistake
didntmakeit.iloc[8]
```

```
Out[26]: airline_sentiment      neutral
airline      United
text      @united why not? Is it a law or a policy?
VADER score      0
predicted sentiment      negative
Name: 8, dtype: object
```

```
In [27]: didntmakeit.iloc[8]['text']
```

```
Out[27]: '@united why not? Is it a law or a policy?'
```

```
In [28]: # some basic visualization
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
```

```
In [29]: df = madeit[madeit['predicted sentiment']=='negative']

words = ' '.join(df['text'])
cleaned_word = " ".join([word for word in words.split()
                          if 'http' not in word
                          and not word.startswith('@')
                          and word != 'RT'
                          ])

stopwords = set(STOPWORDS)
stopwords.add("amp")
stopwords.add("flight")
stopwords.add("united")
stopwords.add("plane")
stopwords.add("now")

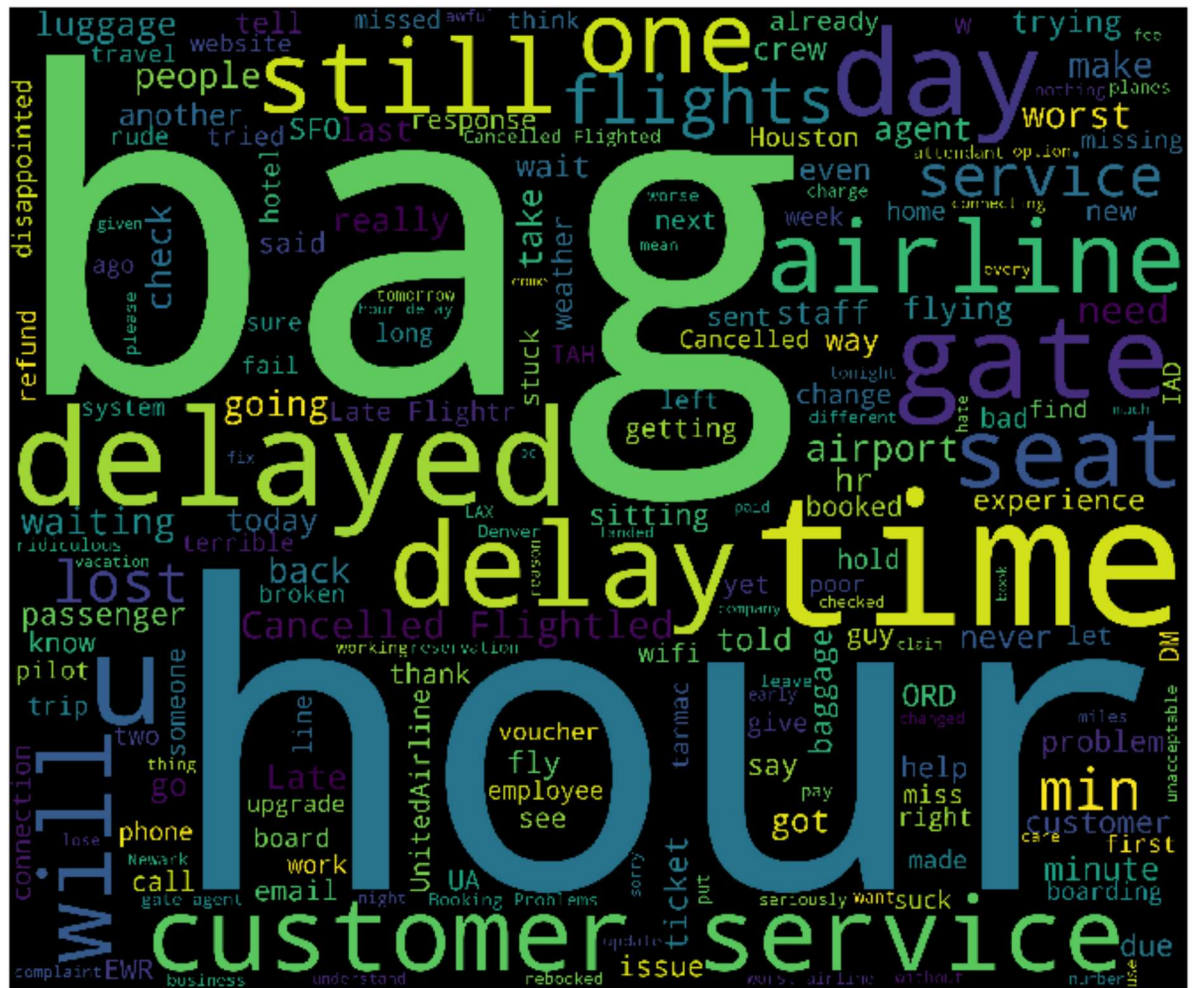
wordcloud = WordCloud(stopwords=stopwords,
                      background_color='black',
                      width=3000,
                      height=2500
                      ).generate(cleaned_word)
```

```
In [30]: type(cleaned_word)
```

```
Out[30]: str
```



```
plt.figure(1,figsize=(12, 12))
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```



```
In [32]: df = madeit[madeit['predicted sentiment']=='positive']

words = ' '.join(df['text'])
cleaned_word = " ".join([word for word in words.split()
                          if 'http' not in word
                          and not word.startswith('@')
                          and word != 'RT'
                          and word != '&']
                        ])

stopwords = set(STOPWORDS)
stopwords.add("amp")
stopwords.add("flight")
stopwords.add("flights")
stopwords.add("united")
stopwords.add("plane")

wordcloud = WordCloud(stopwords=stopwords,
                      background_color='black',
                      width=3000,
                      height=2500
                      ).generate(cleaned_word)
```

```
plt.figure(1,figsize=(12, 12))
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```

