In [1]:
```
#pip install spaCy
```

```
Collecting spaCy
  Downloading spacy-3.4.3-cp39-cp39-win_amd64.whl (11.9 MB)
Collecting catalogue<2.1.0,>=2.0.6
  Downloading catalogue-2.0.8-py3-none-any.whl (17 kB)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\asus\anaconda3\lib\site-packages (from spaCy) (4.62.3)
Requirement already satisfied: numpy>=1.15.0 in c:\users\asus\anaconda3\lib\site-packages (from spaCy) (1.20.3)
Requirement already satisfied: packaging>=20.0 in c:\users\asus\anaconda3\lib\site-packages (from spaCy) (21.3)Note: you may ne
ed to restart the kernel to use updated packages.

Collecting wasabi<1.1.0,>=0.9.1
  Downloading wasabi-0.10.1-py3-none-any.whl (26 kB)
Requirement already satisfied: setuptools in c:\users\asus\anaconda3\lib\site-packages (from spaCy) (58.0.4)
Collecting langcodes<4.0.0,>=3.2.0
  Downloading langcodes-3.3.0-py3-none-any.whl (181 kB)
Requirement already satisfied: jinja2 in c:\users\asus\anaconda3\lib\site-packages (from spaCy) (2.11.3)
Collecting preshed<3.1.0,>=3.0.2
  Downloading preshed-3.0.8-cp39-cp39-win_amd64.whl (96 kB)
Collecting typer<0.8.0,>=0.3.0
  Downloading typer-0.7.0-py3-none-any.whl (38 kB)
Collecting spacy-legacy<3.1.0,>=3.0.10
  Downloading spacy_legacy-3.0.10-py2.py3-none-any.whl (21 kB)
Collecting thinc<8.2.0,>=8.1.0
  Downloading thinc-8.1.5-cp39-cp39-win_amd64.whl (1.3 MB)
Collecting srsly<3.0.0,>=2.4.3
  Downloading srsly-2.4.5-cp39-cp39-win_amd64.whl (481 kB)
Collecting murmurhash<1.1.0,>=0.28.0
  Downloading murmurhash-1.0.9-cp39-cp39-win_amd64.whl (18 kB)
Collecting spacy-loggers<2.0.0,>=1.0.0
  Downloading spacy_loggers-1.0.3-py3-none-any.whl (9.3 kB)
Collecting cymem<2.1.0,>=2.0.2
  Downloading cymem-2.0.7-cp39-cp39-win_amd64.whl (30 kB)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\asus\anaconda3\lib\site-packages (from spaCy) (2.26.0)
Collecting pathy>=0.3.5
  Downloading pathy-0.8.1-py3-none-any.whl (46 kB)
Collecting pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4
  Downloading pydantic-1.10.2-cp39-cp39-win_amd64.whl (2.1 MB)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\asus\anaconda3\lib\site-packages (from packaging>=20.0->spa
Cy) (3.0.4)
Collecting smart-open<6.0.0,>=5.2.1
  Downloading smart_open-5.2.1-py3-none-any.whl (58 kB)
Requirement already satisfied: typing-extensions>=4.1.0 in c:\users\asus\anaconda3\lib\site-packages (from pydantic!=1.8,!=1.8.
1,<1.11.0,>=1.7.4->spaCy) (4.3.0)
Requirement already satisfied: idna<4,>=2.5 in c:\users\asus\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spaCy)
(3.2)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\asus\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0
->spaCy) (1.26.7)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\asus\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->s
paCy) (2022.9.24)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\asus\anaconda3\lib\site-packages (from requests<3.0.0,>=2.
13.0->spaCy) (2.0.4)
Collecting confection<1.0.0,>=0.0.1
  Downloading confection-0.0.3-py3-none-any.whl (32 kB)
Collecting blis<0.8.0,>=0.7.8
  Downloading blis-0.7.9-cp39-cp39-win_amd64.whl (7.0 MB)
Requirement already satisfied: colorama in c:\users\asus\anaconda3\lib\site-packages (from tqdm<5.0.0,>=4.38.0->spaCy) (0.4.4)
Requirement already satisfied: click<9.0.0,>=7.1.1 in c:\users\asus\anaconda3\lib\site-packages (from typer<0.8.0,>=0.3.0->spaC
y) (8.0.3)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\asus\anaconda3\lib\site-packages (from jinja2->spaCy) (1.1.1)
Installing collected packages: catalogue, srsly, pydantic, murmurhash, cymem, wasabi, typer, smart-open, preshed, confection, b
lis, thinc, spacy-loggers, spacy-legacy, pathy, langcodes, spaCy
Successfully installed blis-0.7.9 catalogue-2.0.8 confection-0.0.3 cymem-2.0.7 langcodes-3.3.0 murmurhash-1.0.9 pathy-0.8.1 pre
shed-3.0.8 pydantic-1.10.2 smart-open-5.2.1 spaCy-3.4.3 spacy-legacy-3.0.10 spacy-loggers-1.0.3 srsly-2.4.5 thinc-8.1.5 typer-
0.7.0 wasabi-0.10.1
```

In [3]:
```
import nltk
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import re
import spacy
import string
%matplotlib inline

import warnings
warnings.filterwarnings('ignore')
```

In [4]:
```
full_df = pd.read_csv('twcs.csv',nrows = 5000)
df = full_df[['text']]
df['text'] = df['text'].astype(str)
full_df.head()
```

Out[4]:

	tweet_id	author_id	inbound	created_at	text	response_tweet_id	in_response_to_tweet_id
0	1	sprintcare	False	Tue Oct 31 22:10:47 +0000 2017	@115712 I understand. I would like to assist y...	2	3.0
1	2	115712	True	Tue Oct 31 22:11:45 +0000 2017	@sprintcare and how do you propose we do that	NaN	1.0
2	3	115712	True	Tue Oct 31 22:08:27 +0000 2017	@sprintcare I have sent several private messag...	1	4.0
3	4	sprintcare	False	Tue Oct 31 21:54:49 +0000 2017	@115712 Please send us a Private Message so th...	3	5.0
4	5	115712	True	Tue Oct 31 21:49:35 +0000 2017	@sprintcare I did.	4	6.0

In [5]:
```
df['text_lower'] = df['text'].str.lower()
```

Removal of Punctuations

In [6]:
```
df.drop(["text_lower"],axis = 1,inplace = True)
PUNCT_TO_REMOVE = string.punctuation
def remove_punctuation(text):
    return text.translate(str.maketrans('','',PUNCT_TO_REMOVE))
df['text_w/o_punct'] = df['text'].apply(lambda text: remove_punctuation(text))
df.head()
```

Out[6]:

	text	text_w/o_punct
0	@115712 I understand. I would like to assist y...	115712 I understand I would like to assist you...
1	@sprintcare and how do you propose we do that	sprintcare and how do you propose we do that
2	@sprintcare I have sent several private messag...	sprintcare I have sent several private message...
3	@115712 Please send us a Private Message so th...	115712 Please send us a Private Message so tha...
4	@sprintcare I did.	sprintcare I did

Removal of StopWords

In [9]:
```
from nltk.corpus import stopwords
", ".join(stopwords.words('english'))
```

Out[9]:
```
"i, me, my, myself, we, our, ours, ourselves, you, you're, you've, you'll, you'd, your, yours, yourself, yourselves, he, him, h
is, himself, she, she's, her, hers, herself, it, it's, its, itself, they, them, their, theirs, themselves, what, which, who, wh
om, this, that, that'll, these, those, am, is, are, was, were, be, been, being, have, has, had, having, do, does, did, doing,
a, an, the, and, but, if, or, because, as, until, while, of, at, by, for, with, about, against, between, into, through, during,
before, after, above, below, to, from, up, down, in, out, on, off, over, under, again, further, then, once, here, there, when,
where, why, how, all, any, both, each, few, more, most, other, some, such, no, nor, not, only, own, same, so, than, too, very,
s, t, can, will, just, don, don't, should, should've, now, d, ll, m, o, re, ve, y, ain, aren, aren't, couldn, couldn't, didn, d
idn't, doesn, doesn't, hadn, hadn't, hasn, hasn't, haven, haven't, isn, isn't, ma, mightn, mightn't, mustn, mustn't, needn, nee
dn't, shan, shan't, shouldn, shouldn't, wasn, wasn't, weren, weren't, won, won't, wouldn, wouldn't"
```

In [10]:
```
stop_words = set(stopwords.words('english'))
def remove_stopwords(text):
    return " ".join([word for word in str(text).split() if word not in stop_words])
df['text_w/o_sw'] = df['text'].apply(lambda text: remove_stopwords(text))
df.head()
```

Out[10]:

	text	text_w/o_punct	text_w/o_sw
0	@115712 I understand. I would like to assist y...	115712 I understand I would like to assist you...	@115712 I understand. I would like assist you....
1	@sprintcare and how do you propose we do that	sprintcare and how do you propose we do that	@sprintcare propose
2	@sprintcare I have sent several private messag...	sprintcare I have sent several private message...	@sprintcare I sent several private messages on...
3	@115712 Please send us a Private Message so th...	115712 Please send us a Private Message so tha...	@115712 Please send us Private Message assist ...
4	@sprintcare I did.	sprintcare I did	@sprintcare I did.

Removal of Frequent Keywords

In [28]:
```
from collections import Counter
cnt = Counter()
for text in df['text_w/o_sw'].values:
    for word in text.split():
        cnt[word] += 1

cnt.most_common(10)
```

Out[28]:
```
[('I', 1431),
 ('us', 658),
 ('DM', 436),
 ('Please', 372),
 ('We', 338),
 ('get', 277),
 ('like', 228),
 ('@Tesco', 222),
 ("I'm", 221),
 ('Hi', 218)]
```

In [31]:
```
freqwords = set(w for (w,wc) in cnt.most_common(10))
def remove_freqwords(text):
    return " ".join([word for word in str(text).split() if word not in freqwords])
df['text_w/o_freqwords'] = df['text'].apply(lambda text: remove_freqwords(text))
df.head()
```

Out[31]:

	text	text_w/o_punct	text_w/o_sw	text_w/o_freqwords
0	@115712 I understand. I would like to assist y...	115712 I understand I would like to assist you...	@115712 I understand. I would like assist you....	@115712 understand. would to assist you. would...
1	@sprintcare and how do you propose we do that	sprintcare and how do you propose we do that	@sprintcare propose	@sprintcare and how do you propose we do that
2	@sprintcare I have sent several private messag...	sprintcare I have sent several private message...	@sprintcare I sent several private messages on...	@sprintcare have sent several private messages...
3	@115712 Please send us a Private Message so th...	115712 Please send us a Private Message so tha...	@115712 Please send us Private Message assist ...	@115712 send a Private Message so that we can ...
4	@sprintcare I did.	sprintcare I did	@sprintcare I did.	@sprintcare did.

In [43]:
```
rarewords = set(w for (w,wc) in cnt.most_common()[:-11:-1])
def remove_rarewords(text):
    return " ".join([word for word in str(text).split() if word not in rarewords])
df['text_w/o_rarewords'] = df['text'].apply(lambda text: remove_rarewords(text))
df.head()
```

Out[43]:

	text	text_w/o_punct	text_w/o_sw	text_w/o_freqwords	text_w/o_rarewords
0	@115712 I understand. I would like to assist y...	115712 I understand I would like to assist you...	@115712 I understand. I would like assist you....	@115712 understand. would to assist you. would...	@115712 I understand. I would like to assist y...
1	@sprintcare and how do you propose we do that	sprintcare and how do you propose we do that	@sprintcare propose	@sprintcare and how do you propose we do that	@sprintcare and how do you propose we do that
2	@sprintcare I have sent several private messag...	sprintcare I have sent several private message...	@sprintcare I sent several private messages on...	@sprintcare have sent several private messages...	@sprintcare I have sent several private messag...
3	@115712 Please send us a Private Message so th...	115712 Please send us a Private Message so tha...	@115712 Please send us Private Message assist ...	@115712 send a Private Message so that we can ...	@115712 Please send us a Private Message so th...
4	@sprintcare I did.	sprintcare I did	@sprintcare I did.	@sprintcare did.	@sprintcare I did.

In [ ]: