CS171 | BM1

Assignment 2.1

Pre-processing of Tweets

After gaining Elevated access to the Twitter API Credentials, we were finally able to download the dataset needed for performing the following pre-processing techniques.

A. Pre-processing technique: Lower-casing

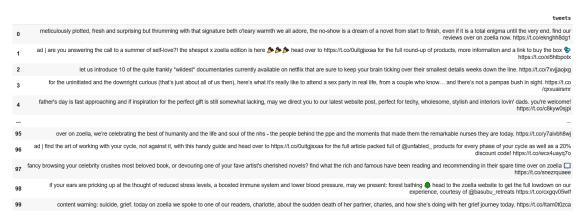
The input:

The code/process:

```
# Create a function to clean the tweets. Remove profanity, unnecessary characters, spaces, and stopwords.
def clean_tweet(tweet):
    if type(tweet) == float:
        return ""
    r = tweet.lower()  # convert text to lower-case
    return r

# cleaning text
df['tweets'] = df['tweets'].apply(clean_tweet)
# show cleaned text
df
```

The output:



B. Pre-processing technique: Punctuation free text

The input:

```
meticulously plotted, fresh and surprising but thrumming with that signature beth o'leary warmth we all adore, the no-show is a dream of a novel from start to finish, even if it is a total enigma untill the very enf. find our reviews over on zoellan ow. https://t.co/leghgh8dg1

ad | are you answering the call to a summer of self-love?! the shespot x zoella edition is here head over to https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box of https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box of https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box of https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box of https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box of https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box of https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box of https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box of https://t.co/louitgjsxaa for the full round-up of products, more information and a link to buy the box of https://t.co/louitgjsxaa for the full round-up of products for every phase down the line. https://t.co/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsxaa/louitgjsysaa/louitgjsxaa/louitgjsxaa/louit
```

The code/process:

The output:

C. Pre-processing technique: Tokenization

The input:

```
meticulously plotted fresh and surprising but thrumming with that signature beth oleary warmth we all adore the noshow is a dream of a novel from start to finish even if it is a total enigma until the very end find our reviews
            ad are you answering the call to a summer of selflove the shespot x zoella edition is here head over to httpstco0uitgisxaa for the full roundup of products more information and a link to buy the box httpstcoxi5htbpotx
2
                               let us introduce 10 of the quite frankly wildest documentaries currently available on netflix that are sure to keep your brain ticking over their smallest details weeks down the line httpstco7xvjjaojxg
            for the uninitiated and the downright curious thats just about all of us then heres what its really like to attend a sex party in real life from a couple who know and theres not a pampas bush in sight httpstcoqxvuairsmr
           fathers day is fast approaching and if inspiration for the perfect gift is still somewhat lacking may we direct you to our latest website post perfect for techy wholesome stylish and interiors lovin dads youre welcome
95
                      over on zoella were celebrating the best of humanity and the life and soul of the nhs the people behind the ppe and the moments that made them the remarkable nurses they are today httpstcoy7aivbh8wj
       ad find the art of working with your cycle not against it with this handy guide and head over to httpstco0uitgjsxaa for the full article packed full of unfabled_products for every phase of your cycle as well as a 20 discount
          fancy browsing your celebrity crushes most beloved book or devouring one of your fave artists cherished novels find what the rich and famous have been reading and recommending in their spare time over on zoella
       if your ears are pricking up at the thought of reduced stress levels a boosted immune system and lower blood pressure may we present forest bathing head to the zoella website to get the full lowdown on our experience
                         content warning suicide grief today on zoella we spoke to one of our readers charlotte about the sudden death of her partner charles and how shes doing with her grief journey today httpstcottam0t0zca
```

The code/process:

```
# Create a function to clean the tweets. Remove profanity, unnecessary characters, spaces, and stopwords.
def clean tweet(tweet):
   if type(tweet) == float:
       return ""
   r = tweet.lower()
                                             # convert text to lower-case
   r = re.sub(r'[^\w\s]','', r)
                                             # remove punctuation
   r = r.split()
                                             # tokenization
   return r
# cleaning text
df['tweets'] = df['tweets'].apply(clean_tweet)
# show cleaned text
df
```

The output:

[meticulously, plotted, fresh, and, surprising, but, thrumming, with, that, signature, beth, oleary, warmth, we, all, adore, the, noshow, is, a, dream, of, a, novel, from, start, to, finish, even, if, it, is, a, total, enigma, until, the, very, end, find, our, reviews, over, on, zoella, now, https://doi.org/10.1003/j.com/1 [ad, are, you, answering, the, call, to, a, summer, of, selflove, the, shespot, x, zoella, edition, is, here, head, over, to, httpstco0uitgjsxaa, for, the, full, roundup, of, products, more, information, and, a, link, to, buy, the, box [let, us, introduce, 10, of, the, quite, frankly, wildest, documentaries, currently, available, on, netflix, that, are, sure, to, keep, your, brain, ticking, over, their, smallest, details, weeks, down, the, line, httpstco7xvjjaojxgj [for, the, uninitiated, and, the, downright, curious, thats, just, about, ali, of, us, then, heres, what, its, really, like, to, attend, a, sex, party, in, real, life, from, a, couple, who, know, and, theres, not, a, pampas, bush, in, sight, [fathers, day, is, fast, approaching, and, if, inspiration, for, the, perfect, gift, is, still, somewhat, lacking, may, we, direct, you, to, our, latest, website, post, perfect, for, techy, wholesome, stylish, and, interiors, lovin, dads, [over, on, zoella, were, celebrating, the, best, of, humanity, and, the, life, and, soul, of, the, nhs, the, people, behind, the, ppe, and, the, moments, that, made, them, the, remarkable, nurses, they, are, today, [ad, find, the, art, of, working, with, your, cycle, not, against, it, with, this, handy, guide, and, head, over, to, httpstco0uitgisxaa, for, the, full, article, packed, full, of, unfabled_products, for, every, phase, of, your, cycle, as well, as, a, 20, discount, code, httpstcowcx4uayq7o [fancy, browsing, your, celebrity, crushes, most, beloved, book, or, devouring, one, of, your, fave, artists, cherished, novels, find, what, the, rich, and, famous, have, been, reading, and, recommending, in, their, spare, time, over, on, zoella, httpstcosnezrquaee] [if, your, ears, are, pricking, up, at, the, thought, of, reduced, stress, levels, a, boosted, immune, system, and, lower, blood, pressure, may, we, present, forest, bathing, head, to, the, zoella, website, to, get, the, full, lowdown, on, our, experience, courtesy, of, basubu_retreats, https://doi.org/10.1006/j.com/10.100 99 [content, warning, suicide, grief, today, on, zoelia, we, spoke, to, one, of, our, readers, charlotte, about, the, sudden, death, of, her, partner, charles, and, how, shes, doing, with, her, grief, journey, today, https://doi.org/10.1003/j.coa.j.

D. Pre-processing technique: Stop words removal

The input:

```
tweets

freeltculously, plotted, fresh, and, surprising, but, thrumming, with, that, signature, beth, oleary, warmth, we, all, adore, the, noshow, is, a, dream, of, a, novel, from, start, to, finish, even, if, it, is, a, total, enigma, until, the, very, end, find, our, reviews, over, on, zoella, now, httpstcoeknghin&dgr]

flet, us, introduce, 10, of, the, quite, frankly, wildest, documentaries, currently, available, on, netflix, that, are, sure, to, keep, your, brain, ticking, over, their, smallest, details, weeks, down, the, line, httpstco7xyljaojxg]

flor, the, uninitiated, and, the, downright, curious, thats, just, about, all, of, us, then, heres, what, its, really, like, to, attend, a, sex, party, in, real, life, from, a, couple, who, know, and, theres, not, a, pampas, bush, in, sight, httpstcoavairsmr]

flathers, day, is, fast, approaching, and, if, inspiration, for, the, perfect, gift, is, still, somewhat, lacking, may, we, direct, you, to, our, latest, website, post, perfect, for, techy, wholesome, stylish, and, interiors, lovin, dads, youre, welcome, httpstcoavairsmr]

flover, on, zoella, were, celebrating, the, best, of, humanity, and, the, life, and, soul, of, the, nhs, the, people, behind, the, ppe, and, the, moments, that, made, them, the, remarkable, nurses, they, are, today, httpstcoavairsmr, well, as, a, 20, discount, code, httpstcoov/aivabwij

flor, browsing, your, celebrity, crushes, most, beloved, book, or, devouring, one, of, your, fave, arlists, cherished, novels, find, what, the, rich, and, famous, have, been, reading, and, recommending, in, their, spare, time, over, on, zoella, were, celebrating, the, best, of, near, pricking, up, at, the, thought, of, reduced, stress, levels, a, boosted, immune, system, and, lower, blood, pressure, may, we, present, forest, bathling, head, ohe, zoella, httpstcooxequowell lowdown, on, our, experience, courtesy, of, basubur_refreats, httpstcocxequowell lowdown, on, our, experience, courtesy, of, basubur_refreats, httpstcocxequowell lowdown, on, our, experie
```

The code/process:

```
# Create a function to clean the tweets. Remove profanity, unnecessary characters, spaces, and stopwords.
def clean_tweet(tweet):
   if type(tweet) == float:
        return ""
   r = tweet.lower()
                                               # convert text to lower-case
   r = re.sub(r'[^\w\s]','', r) # remove punctuation
   r = r.split()
                                               # tokenization
    stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to", "from"]
    r = [w \text{ for } w \text{ in } r \text{ if not } w \text{ in stopwords}] \# remove stopwords}
    r = " ".join(word for word in r)
   return r
# cleaning text
df['tweets'] = df['tweets'].apply(clean_tweet)
# show cleaned text
```

The output:

	tweets
0	meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now httpstcoeknghh8dg1
1	ad are you answering call summer selflove shespot x zoella edition is here head over httpstco0uitgjsxaa full roundup products more information link buy box httpstcoxi5htbpotx
2	let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line httpstco7xvjjaojxg
3	uninitiated downright curious thats just about all us then heres what its really like attend sex party real life couple who know theres not pampas bush sight httpstcoqxvuairsmr
4	fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome httpstcoc8kyw0sjpi
95	over zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today httpstcoy7aivbh8wj
96	ad find art working with your cycle not against it with this handy guide head over httpstco0uitgisxaa full article packed full unfabled_products every phase your cycle as well as 20 discount code httpstcowcx4uayq7o
97	fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella httpstcosnezrquaee
98	if your ears are pricking up at thought reduced stress levels boosted immune system lower blood pressure may we present forest bathing head zoella website get full lowdown our experience courtesy basubu_retreats httpstcocxgqv05wlf
99	content warning suicide grief today zoella we spoke one our readers charlotte about sudden death her partner charles how shes doing with her grief journey today httpstcollam010zca

E. Pre-processing technique: Removing duplicate characters from the word The input:

meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now httpstcoeknghh8dg1

da are you answering call summer selflove shespot x zoella edition is here head over httpstco0ultgisxaa full roundup products more information link buy box httpstcoeknghh8dg1

let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line httpstco7xyljaojxsr

distinct and initiated downright curious thats just about all us then heres what its really like attend sex party real life couple who know theres not pampas bush sight httpstcoqxvuaignet

fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome httpstcoc8kyw0sjpi

fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome httpstcoc8kyw0sjpi

ad find art working with your cycle not against it with this handy guide head over httpstcoOultgisxaa full article packed full unfabled_products every phase your cycle as well as 20 discount code httpstcooxxx4uayq7o

fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella httpstcooxexqu05wff

fly our ears are pricking up at thought reduced stress levels boosted immune system lower blood pressure may we present forest bathing head zoella website get full lowdown our experience courtesy basubu_retraces

httpstcooxyqu05wff

content warning suicide grief today zoella we spoke one our readers charlotte about sudden death her partner charles how shes doing with her grief journey today httpstco

The code/process:

```
# Create a function to clean the tweets. Remove profanity, unnecessary characters, spaces, and stopwords.
def clean tweet(tweet):
   if type(tweet) == float:
       return ""
                                          # convert text to lower-case
   r = tweet.lower()
   r = r.split()
                                          # tokenization
   stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to", "from"]
   r = [w \text{ for } w \text{ in } r \text{ if not } w \text{ in stopwords}] # remove stopwords
   r = " ".join(word for word in r)
   return r
# cleaning text
df['tweets'] = df['tweets'].apply(clean tweet)
# show cleaned text
```

The output:

F. Pre-processing technique: Remove profanity

The input:

```
meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now httpstcooknghh8dg1

da are you answering call summer selflove shespot x zoella edition is here head over httpstcoolitgisxaa full roundup products more information link buy box httpstcooknghh8dg1

let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line httpstco7xvijaojkg

uninitiated downright curious thats just about all us then heres what its really like attend sex party real life couple who know theres not pampas bush sight httpstcookyvalisrym

fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome httpstcoo8kyw0sjpi

over zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today httpstcoo/latybn8wy

ad find art working with your cycle not against it with this handy guide head over httpstcoolitgisxaa full article packed full unfabled_products every phase your cycle as well as 20 discount code httpstcooxx4uayq7o

fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella httpstcooxgqv05wf

fly your ears are pricking up at thought reduced stress levels boosted immune system lower blood pressure may we present forest bathing head zoella website get full lowdown our experience courtesy basubur_retraced shiftstocoxgqv05wf

content warning suicide grief today zoella we spoke one our readers charlotte about sudden death her partner charles how shes doing with her grief journey today httpstcottam0t0zca
```

The code/process:

```
# Create a function to clean the tweets. Remove profanity, unnecessary characters, spaces, and stopwords.
def clean_tweet(tweet):
   if type(tweet) == float:
       return ""
   r = tweet.lower()
                                          # convert text to lower-case
   r = profanity.censor(r)
                                        # remove profanity
                                          # tokenization
   r = r.split()
   stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to", "from"]
   r = [w \text{ for } w \text{ in } r \text{ if not } w \text{ in stopwords}] \# remove stopwords}
   r = " ".join(word for word in r)
   return r
# cleaning text
df['tweets'] = df['tweets'].apply(clean_tweet)
# show cleaned text
```

The output:

	tweets
0	meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now httpstcoeknghh8dg1
1	ad are you answering call summer selflove shespot x zoella edition is here head over httpstco0ultgisxaa full roundup products more information link buy box httpstcoxi5htbpotx
2	let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line httpstco7xvjjaojxg
3	uninitiated downright curious thats just about all us then heres what its really like attend **** party real life couple who know theres not pampas bush sight httpstcoqxvuairsmr
4	fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome httpstcoc8kyw0sjpi
95	over zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today httpstcoy/Taivbh8wj
96	ad find art working with your cycle not against it with this handy guide head over httpstcoowcx4uayq7o
97	fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella httpstcosnezrquaee
98	if your ears are pricking up at thought reduced stress levels boosted immune system lower blood pressure may we present forest bathing head zoella website get full lowdown our experience courtesy basubu_retreats https://docs.org/v105wif
99	content warning suicide grief today zoella we spoke one our readers charlotte about sudden death her partner charles how shes doing with her grief journey today httpstcoltam0t0zca

G. Pre-processing technique: Remove mentions

The input:

meliculously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now https://documentaries.com/https://docu

The code/process:

```
# Create a function to clean the tweets. Remove profanity, unnecessary characters, spaces, and stopwords.
def clean tweet(tweet):
     if type(tweet) == float:
         return ""
    r = tweet.lower()
                                                        # convert text to lower-case
    r = re.sub(r'[^\w\s]','', r) # remove punctuation

r = re.sub("(.)\1{2,}", "\1", r) # remove duplicate characters in a word

r = profanity.censor(r) # remove duplicate r = re.sub(r'[^\w])
    r = r.split()
                                                      # tokenization
    r = re.sub("@[A-Za-z0-9_]+","", r)
                                                      # remove mentions
     stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to", "from"]
     r = [w \text{ for } w \text{ in } r \text{ if } not \text{ } w \text{ in } stopwords] # remove stopwords
     r = " ".join(word for word in r)
     return r
# cleaning text
df['tweets'] = df['tweets'].apply(clean tweet)
# show cleaned text
df
```

The output:

meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now 0 ad are you answering call summer selflove shespot x zoella edition is here head over httpstco0uitgisxaa full roundup products more information link buy box httpstcoxi5htbpotx 2 let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line httpstco7xvjjaojxg uninitiated downright curious thats just about all us then heres what its really like attend **** party real life couple who know theres not pampas bush sight httpstcoqxvuairsmr 4 fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome httpstcoc8kyw0sjpi 95 over zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today httpstcoy7aivbh8wj 96 ad find art working with your cycle not against it with this handy guide head over httpstco0uitgjsxaa full article packed full unfabled_products every phase your cycle as well as 20 discount code httpstcowcx4uayq7o 97 fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella httpstcosnezrquaee if your ears are pricking up at thought reduced stress levels boosted immune system lower blood pressure may we present forest bathing head zoella website get full lowdown our experience courtesy basubu_retreats 98 content warning suicide grief today zoella we spoke one our readers charlotte about sudden death her partner charles how shes doing with her grief journey today httpstcoltam0t0zca

H. Pre-processing technique: Remove hashtags

The input:

```
tweets
                                meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now https://doi.org/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.com/10.1003/j.
 0
                                                                                                   ad are you answering call summer selflove shespot x zoella edition is here head over httpstcoOuitgjsxaa full roundup products more information link buy box httpstcoxi5htbpotx
 2
                                                                                            let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line httpstco7xvjjaojxg
                                                                                                     uninitiated downright curious thats just about all us then heres what its really like attend **** party real life couple who know theres not pampas bush sight httpstcoqxvuairsmr
 4
                                fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome httpstcoc8kyw0sjpi
95
                                                                                                                                       over zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today httpstcov7aivbh8wi
96
                       ad find art working with your cycle not against it with this handy quide head over httpstcoOuitgisxaa full article packed full unfabled products every phase your cycle as well as 20 discount code httpstcowcx4uayq7o
97
                    fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella httpstcosnezrquaee
                   If your ears are pricking up at thought reduced stress levels boosted immune system lower blood pressure may we present forest bathing head zoella website get full lowdown our experience courtesy basubu retreats
                                                                                    content warning suicide grief today zoella we spoke one our readers charlotte about sudden death her partner charles how shes doing with her grief journey today httpstcoltam0t0zca
```

The code/process:

```
# Create a function to clean the tweets. Remove profanity, unnecessary characters, spaces, and stopwords.
def clean_tweet(tweet):
    if type(tweet) == float:
        return ""
    r = tweet.lower()
                                                   # convert text to lower-case
    r = profanity.censor(r)
                                                  # remove profanity
    r = re.sub("'", "", r)
                                                  # This is to avoid removing contractions in english
    r = re.sub("@[A-Za-z0-9_]+","", r)
                                                # remove mentions
                                               # remove the # in #hashtags
    r = re.sub("#[A-Za-z0-9_]+","", r)
    r = re.sub(r'[^\w\s]','', r)
r = re.sub("(.)\\1{2,}", "\\1", r)
r = re.sub("[^a-z0-9]"," ", r)
                                                  # remove punctuation
                                                   # remove duplicate characters in a word
    r = r.split()
    stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to", "from"]
    r = [w \text{ for } w \text{ in } r \text{ if not } w \text{ in stopwords}] \# remove stopwords}
    r = " ".join(word for word in r)
    return r
# cleaning text
df['tweets'] = df['tweets'].apply(clean tweet)
# show cleaned text
```

The output:

meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zcoella now httpstcocknghh8dg1

1 ad are you answering call summer selflove shespot x zoella edition is here head over httpstcoOultgisxaa full roundup products more information link buy box httpstcoxisfhtbpotx

2 let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line httpstco7xvjjaojxg

3 uninitiated downright curious thats just about all us then heres what its really like attend party real life couple who know theres not pampas bush sight httpstcooxyvuairsmr

4 fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome httpstcoc8kyw0sjpi

...

95 over zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today httpstcoy7aivbh8wj

96 ad find art working with your cycle not against it with this handy guide head over httpstcoOultgisxaa full article packed full products every phase your cycle as well as 20 discount code httpstcooxx4uayq7o

97 fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella httpstcooxezquoes

98 if your ears are pricking up at thought reduced stress levels boosted immune system lower blood pressure may we present forest bathing head zoella website get full lowdown our experience courtesy httpstcooxing voices and the partner charles how shes doing with her grief journey today httpstcooltam0f0zca

I. Pre-processing technique: Remove links

The input:

```
meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now httpstcocknghh8dg1

ad are you answering call summer selflove shespot x zoella edition is here head over httpstcoolutgisxaa full roundup products more information link buy box httpstcockshghh8dg1

let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line httpstco7xvjjaojxg

uninitiated downright curious thats just about all us then heres what its really like attend party real life couple who know theres not pampas bush sight httpstcoqxvuairsmr

fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome httpstcoc8kyw0sjpi

...

prefect techy wholesome stylish interiors lovin dads youre welcome httpstcoc8kyw0sjpi

over zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today httpstcoy7aivh8wj

ad find art working with your cycle not against it with this handy guide head over httpstcoolutgjsxaa full article packed full products every phase your cycle as well as 20 discount code httpstcooxx4uayq7o

fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella httpstcooxezqu05wf

fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella httpstcooxezqu05wf

fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella h
```

The code/process:

```
# Create a function to clean the tweets. Remove profanity, unnecessary characters, spaces, and stopwords.
def clean tweet(tweet):
    if type(tweet) == float:
        return ""
    r = tweet.lower()
                                                    # convert text to lower-case
    r = profanity.censor(r)
                                                    # remove profanity
    r = re.sub("'", "", r)
                                                   # This is to avoid removing contractions in english
    r = re.sub("@[A-Za-z0-9_]+","", r)
r = re.sub("#[A-Za-z0-9_]+","", r)
                                                   # remove mentions
                                                   # remove the # in #hashtags
    r = re.sub(r'[^\w\s]','', r)
r = re.sub(r'http\s+', '', r)
r = re.sub("(.)\\1{2,}", "\\1", r)
                                                    # remove punctuation
                                                    # remove links
                                                    # remove duplicate characters in a word
    r = re.sub("[^a-z0-9]"," ", r)
    r = r.split()
    stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to", "from"]
    r = [w \text{ for } w \text{ in } r \text{ if not } w \text{ in stopwords}] # remove stopwords
    r = " ".join(word for word in r)
    return r
# cleaning text
df['tweets'] = df['tweets'].apply(clean_tweet)
# show cleaned text
```

The output:

tweets

neticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now

ad are you answering call summer selflove shespot x zoella edition is here head over full roundup products more information link buy box

let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line

uninitiated downright curious thats just about all us then heres what its really like attend party real life couple who know theres not pampas bush sight

fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome

conver zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today

ad find art working with your cycle not against it with this handy guide head over full article packed full products every phase your cycle as well as 20 discount code

fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella

if your ears are pricking up at thought reduced stress levels boosted immune system lower blood pressure may we present forest bathing head zoella website get full lowdown our experience courtesy

content warning suicide grief today zoella we spoke one our readers chariotte about sudden death her partner charles how shes doing with her grief journey today

II. Getting the Polarity of the Tweets

sentiment df

These lines of codes were used to add Polarity columns which are derived from the tweet text using the TextBlob library in Python. It adds one column for the Polarity value. We create a list of polarity values and tweet text and create a dataframe.

```
# Define the sentiment objects using TextBlob
sentiment objects = [TextBlob(tweet) for tweet in cleaned]
sentiment objects[0].polarity, sentiment objects[0]
# Create a list of polarity values and tweet text
sentiment_values = [[tweet.sentiment.polarity, str(tweet)] for tweet in sentiment_objects]
# Print the value of the 0th row.
sentiment values[0]# Print all the sentiment values
sentiment_values[0:99]
# Create a dataframe of each tweet against its polarity
sentiment df = pd.DataFrame(sentiment values, columns=["polarity", "tweet"])
# Save the polarity column as 'n'.
n=sentiment df["polarity"]
# Convert this column into a series, 'm'.
m=pd.Series(n)
m
# show the dataframe with the polarity column
```

	polarity	tweet
0	0.300000	meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshow is dream novel start finish even if it is total enigma until very end find our reviews over zoella now
1	0.425000	ad are you answering call summer selflove shespot x zoella edition is here head over full roundup products more information link buy box
2	0.248148	let us introduce 10 quite frankly wildest documentaries currently available netflix that are sure keep your brain ticking over their smallest details weeks down line
3	0.100000	uninitiated downright curious thats just about all us then heres what its really like attend party real life couple who know theres not pampas bush sight
4	0.512500	fathers day is fast approaching if inspiration perfect gift is still somewhat lacking may we direct you our latest website post perfect techy wholesome stylish interiors lovin dads youre welcome
95	0.450000	over zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today
96	0.433333	ad find art working with your cycle not against it with this handy guide head over full article packed full products every phase your cycle as well as 20 discount code
97	0.518750	fancy browsing your celebrity crushes most beloved book or devouring one your fave artists cherished novels find what rich famous have been reading recommending their spare time over zoella
98	0.175000	if your ears are pricking up at thought reduced stress levels boosted immune system lower blood pressure may we present forest bathing head zoella website get full lowdown our experience courtesy
99	-0.533333	content warning suicide grief today zoella we spoke one our readers charlotte about sudden death her partner charles how shes doing with her grief journey today

III. Sentiment Analysis and Classification of the Tweets

To start with the sentiment analysis, we must initialize first the variables for negative, positive, and neutral.

```
# Initialize variables, 'pos', 'neg', 'neu'.

pos=0
neg=0
neu=0
```

We then create a loop to classify the tweets as Positive, Negative, or Neutral. This loop determines if the tweet is positive if the polarity is greater than zero. If polarity is lesser than zero, then it is negative. Lastly, if the value is equal to zero, then it is neutral. We then print the total positive, negative, and neutral tweets there are.

```
for items in m:
    if items>0:
         print("Positive")
         pos=pos+1
    elif items<0:
         print("Negative")
         neg=neg+1
    else:
         print("Neutral")
         neu=neu+1
print(pos,neg,neu)
Positive
Positive
Negative
Positive
Positive
Positive
Positive
Neutral
Positive
Positive
Neutral
Positive
Positive
Positive
Neutral
Positive
Positive
Negative
Positive
Positive
Positive
Positive
Positive
Positive
Positive
Positive
Positive
Negative
72 15 13
```

IV. Getting the Sentiment Classification Percentage

These lines of code display the percentage of positive, negative, and neutral tweets in the dataset. They are calculated using the Polarity value, which has a range of -1 to 1. If the value is less than zero, the tweet is negative, if it is greater than zero, the tweet is positive, and if it is equal to zero, the tweet is neutral.

```
# Display the number of tweets who feel a certain way.
print("%f percent of the tweets are positive."%(pos))
print("%f percent of the tweets are negative."%(neg))
print("%f percent of the tweets are neutral."%(neu))

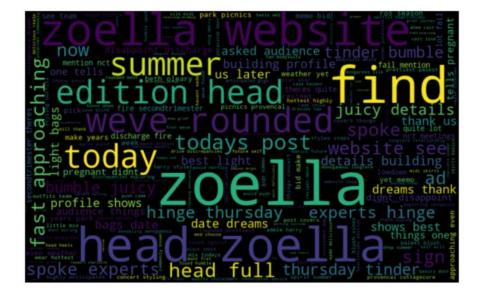
72.000000 percent of the tweets are positive.
15.000000 percent of the tweets are negative.
13.000000 percent of the tweets are neutral.
```

V. Wordcloud

WordCloud is a technique to show which words are the most frequent among the tweets. The WordCloud was used to check how the tweets changed when pre-processing techniques have been applied. There should be no URLs, mentions, punctuation, and so on.

```
# Create a Wordcloud from the tweets
all_words = ' '.join([text for text in cleaned])
wordcloud = WordCloud(width=800, height=500, random_state=21, max_font_size=110).generate(all_words)

plt.figure(figsize=(10, 7))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis('off')
plt.show()
```



VI. Plotting the Sentiment Analysis Classification

A pie graph was used to visualize the results of our sentiment analysis.

```
# We will create a pie chart to graphically display the results of our analysis:
pieLabels=["Positive","Negative","Neutral"]

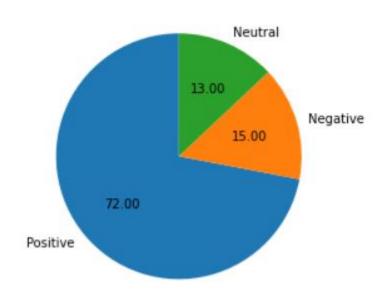
populationShare=[pos,neg,neu]

figureObject, axesObject = plt.subplots()

axesObject.pie(populationShare,labels=pieLabels,autopct='%1.2f',startangle=90)

axesObject.axis('equal')

plt.show()
```



VII. Saving the cleaned dataset

Last step is then to save the final and clean dataset to a CSV file (cleaned_tweets.csv).

```
# Saving the dataset to a CSV

df.to_csv('cleaned_tweets.csv', index = False)
```