

Assignment 2.1

## I. 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:

	tweets
0	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 until the very end. Find our reviews over on Zoella now. <a href="https://t.co/eKNghH8DG1">https://t.co/eKNghH8DG1</a>
1	Ad   Are you answering the call to a summer of self-love?! The SheSpot X Zoella edition is here 🍌🍌🍌 Head over to <a href="https://t.co/0uitGJsXaA">https://t.co/0uitGJsXaA</a> for the full round-up of products, more information and a link to buy the box <a href="https://t.co/xi5HTbpOIX">https://t.co/xi5HTbpOIX</a>
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. <a href="https://t.co/7xvJJaOjXg">https://t.co/7xvJJaOjXg</a>
3	For the uninitiated and the downright curious (that's just about all of us then), here's what it's really like to attend a sex party in real life, from a couple who know... and there's not a pampas bush in sight. <a href="https://t.co/QXVuaiRsmr">https://t.co/QXVuaiRsmr</a>
4	Father's 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. You're welcome! <a href="https://t.co/C8Kyw0Sjpl">https://t.co/C8Kyw0Sjpl</a>
...	...
95	Over on Zoella, we're 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. <a href="https://t.co/Y7aIVBH8wJ">https://t.co/Y7aIVBH8wJ</a>
96	Ad   Find the art of working with your cycle, not against it, with this handy guide and head over to <a href="https://t.co/0uitGJsXaA">https://t.co/0uitGJsXaA</a> for the full article packed full of @unfabled_ products for every phase of your cycle as well as a 20% discount code! <a href="https://t.co/WCX4uAYq7O">https://t.co/WCX4uAYq7O</a>
97	Fancy browsing your celebrity crushes most beloved book, or devouring one of your fave artist's cherished novels? Find what the rich and famous have been reading and recommending in their spare time over on Zoella <a href="https://t.co/snEzrquAEe">https://t.co/snEzrquAEe</a>
98	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 <a href="https://t.co/CXgQv05Wlf">https://t.co/CXgQv05Wlf</a>
99	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 she's doing with her grief journey today. <a href="https://t.co/lTam0T0Zca">https://t.co/lTam0T0Zca</a>

## 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()
    return r

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

# show cleaned text
df
```

## The output:

	tweets
0	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 until the very end. find our reviews over on zoella now. <a href="https://t.co/ekngth8dg1">https://t.co/ekngth8dg1</a>
1	ad   are you answering the call to a summer of self-love?! the shespot x zoella edition is here 🍌🍌🍌 head over to <a href="https://t.co/0uitgjsxaa">https://t.co/0uitgjsxaa</a> for the full round-up of products, more information and a link to buy the box <a href="https://t.co/xi5htbpotx">https://t.co/xi5htbpotx</a>
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. <a href="https://t.co/7xvjjaojxg">https://t.co/7xvjjaojxg</a>
3	for the uninitiated and the downright curious (that's just about all of us then), here's what it's really like to attend a sex party in real life, from a couple who know... and there's not a pampas bush in sight. <a href="https://t.co/qxvuairsmr">https://t.co/qxvuairsmr</a>
4	father's 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. you're welcome! <a href="https://t.co/c8kyw0sjpl">https://t.co/c8kyw0sjpl</a>
...	...
95	over on zoella, we're 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. <a href="https://t.co/y7aivbh8wj">https://t.co/y7aivbh8wj</a>
96	ad   find the art of working with your cycle, not against it, with this handy guide and head over to <a href="https://t.co/0uitgjsxaa">https://t.co/0uitgjsxaa</a> for the full article packed full of @unfabled_ products for every phase of your cycle as well as a 20% discount code! <a href="https://t.co/wcx4uayq7o">https://t.co/wcx4uayq7o</a>
97	fancy browsing your celebrity crushes most beloved book, or devouring one of your fave artist's cherished novels? find what the rich and famous have been reading and recommending in their spare time over on zoella <a href="https://t.co/snezrquae">https://t.co/snezrquae</a>
98	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 <a href="https://t.co/cxgqv05wlf">https://t.co/cxgqv05wlf</a>
99	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 she's doing with her grief journey today. <a href="https://t.co/ltam0t0zca">https://t.co/ltam0t0zca</a>

## B. Pre-processing technique: Punctuation free text

### The input:

tweets	
0	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 until the very end. find our reviews over on zoella now. <a href="https://t.co/ekngnh8dg1">https://t.co/ekngnh8dg1</a>
1	ad   are you answering the call to a summer of self-love?! the shespot x zoella edition is here 🐘🐘🐘 head over to <a href="https://t.co/0uitgjsxaa">https://t.co/0uitgjsxaa</a> for the full round-up of products, more information and a link to buy the box <a href="https://t.co/xi5htbpotx">https://t.co/xi5htbpotx</a>
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. <a href="https://t.co/7xvjaojxg">https://t.co/7xvjaojxg</a>
3	for the uninitiated and the downright curious (that's just about all of us then), here's what it's really like to attend a sex party in real life, from a couple who know... and there's not a pampas bush in sight. <a href="https://t.co/qxvuairsmr">https://t.co/qxvuairsmr</a>
4	father's 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. you're welcome! <a href="https://t.co/c8kyw0sjpi">https://t.co/c8kyw0sjpi</a>
...	...
95	over on zoella, we're 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. <a href="https://t.co/y7aivbh8wj">https://t.co/y7aivbh8wj</a>
96	ad   find the art of working with your cycle, not against it, with this handy guide and head over to <a href="https://t.co/0uitgjsxaa">https://t.co/0uitgjsxaa</a> for the full article packed full of @unfabled_ products for every phase of your cycle as well as a 20% discount code! <a href="https://t.co/wcx4uayq7o">https://t.co/wcx4uayq7o</a>
97	fancy browsing your celebrity crushes most beloved book, or devouring one of your fave artist's cherished novels? find what the rich and famous have been reading and recommending in their spare time over on zoella <a href="https://t.co/snezrquae">https://t.co/snezrquae</a>
98	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 <a href="https://t.co/cxgqv05wff">https://t.co/cxgqv05wff</a>
99	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 she's doing with her grief journey today. <a href="https://t.co/ltam0t0zca">https://t.co/ltam0t0zca</a>

### 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()
    r = re.sub(r'^\w\s', '', r)
    return r

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

# show cleaned text
df
```

### The output:

tweets	
0	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 <a href="https://t.co/ekngnh8dg1">httpstcoekngnh8dg1</a>
1	ad are you answering the call to a summer of selflove the shespot x zoella edition is here head over to <a href="https://t.co/0uitgjsxaa">httpstco0uitgjsxaa</a> for the full roundup of products more information and a link to buy the box <a href="https://t.co/xi5htbpotx">httpstcoxi5htbpotx</a>
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 <a href="https://t.co/7xvjaojxg">httpstco7xvjaojxg</a>
3	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 <a href="https://t.co/qxvuairsmr">httpstcoqxvuairsmr</a>
4	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 <a href="https://t.co/c8kyw0sjpi">httpstcoc8kyw0sjpi</a>
...	...
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 <a href="https://t.co/y7aivbh8wj">httpstcoy7aivbh8wj</a>
96	ad find the art of working with your cycle not against it with this handy guide and head over to <a href="https://t.co/0uitgjsxaa">httpstco0uitgjsxaa</a> for the full article packed full of unfabled_ products for every phase of your cycle as well as a 20 discount code <a href="https://t.co/wcx4uayq7o">httpstcowcx4uayq7o</a>
97	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 <a href="https://t.co/snezrquae">httpstcosnezrquae</a>
98	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 <a href="https://t.co/cxgqv05wff">httpstcocxgqv05wff</a>
99	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 <a href="https://t.co/ltam0t0zca">httpstcoltam0t0zca</a>

# C. Pre-processing technique: Tokenization

## The input:

	tweets
0	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 httpstcoeknghh8dg1
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 httpstcoxi5htbpbx
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 httpstco7xvjjaoxg
3	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
4	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 httpstcoc8kyw0sjpl
...	...
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
96	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 code httpstcowcx4uayq7o
97	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 httpstcosnezrquae
98	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 httpstcocxqv05wlf
99	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 httpstcoitam0t0zca

## 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:

	tweets
0	[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, httpstcoeknghh8dg1]
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, httpstcoxi5htbpbx]
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, httpstco7xvjjaoxg]
3	[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]
4	[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, httpstcoc8kyw0sjpl]
...	...
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]
96	[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, code, httpstcowcx4uayq7o]
97	[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, httpstcosnezrquae]
98	[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, httpstcocxqv05wlf]
99	[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, httpstcoitam0t0zca]

## D. Pre-processing technique: Stop words removal

### The input:

tweets	
0	[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, httpstcoeknghh8dg1]
1	[ad, are, you, answering, the, call, to, a, summer, of, selflove, the, shespot, x, zoella, edition, is, here, head, over, to, httpstco0ultgjsxaa, 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, httpstco7xvjaojxg]
3	[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]
4	[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, httpstcoc8kyw0sjpi]
...	...
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]
96	[ad, find, the, art, of, working, with, your, cycle, not, against, it, with, this, handy, guide, and, head, over, to, httpstco0ultgjsxaa, for, the, full, article, packed, full, of, unfabled_, products, for, every, phase, of, your, cycle, as, well, as, a, 20, discount, code, httpstcowcx4uayq7o]
97	[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, httpstcosnezrquae]
98	[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, httpstcocxgqv05wif]
99	[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, 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 = 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 for w in r if not w 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:

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 httpstco0ultgjsxaa 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 httpstco7xvjaojxg
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 httpstco0ultgjsxaa 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 httpstcosnezrquae
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 httpstcocxgqv05wif
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

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

### The input:

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 httpstco0ultgjsxaa 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 httpstco7xvjaojxg
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 httpstco0ultgjsxaa 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 httpstcosnezrquae
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 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 = re.sub(r'[^\w\s]', '', r) # remove punctuation
    r = re.sub(r"(\.|\{|\}|\,)", "\\1", r) # remove duplicate characters in a word
    r = r.split() # tokenization
    stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to", "from"]
    r = [w for w in r if not w 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:

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 httpstco0ultgjsxaa 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 httpstco7xvjaojxg
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 httpstco0ultgjsxaa 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 httpstcosnezrquae
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 httpstcoltam0t0zca



## F. Pre-processing technique: Remove profanity

### The input:

		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 httpstco0ultgjsxaa 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 httpstco0ultgjsxaa 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 httpstcosnezrquae	
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 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 = re.sub(r'[\w\s]', '', r) # remove punctuation
    r = re.sub(r"(\.)\\1{2,}", "\\1", r) # remove duplicate characters in a word
    r = profanity.censor(r) # remove profanity
    r = r.split() # tokenization
    stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to", "from"]
    r = [w for w in r if not w 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:

		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 httpstco0ultgjsxaa 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 httpstcoy7aivbh8wj	
96	ad find art working with your cycle not against it with this handy guide head over httpstco0ultgjsxaa 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 httpstcosnezrquae	
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 httpstcoltam0t0zca	

## G. Pre-processing technique: Remove mentions

### The input:

	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 httpstco7xvjaojxg
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 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 httpstcosnezrquae
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 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()
    r = re.sub(r'^\w\s',' ', r)
    r = re.sub(r"(\.|\!|\?|,)", "\1", r)
    r = profanity.censor(r)
    r = r.split()
    r = re.sub("@[A-Za-z0-9_]+", "", r)
    stopwords = ["for", "on", "an", "a", "of", "and", "in", "the", "to", "from"]
    r = [w for w in r if not w in 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:

	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 httpstco7xvjaojxg
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 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 httpstcosnezrquae
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 httpstcoltam0t0zca

## H. Pre-processing technique: Remove hashtags

### The input:

	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 httpstco7xvjaojxg
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 httpstcoc8kyw0sjpl
...	...
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 httpstcosnezrquae
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 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
    r = re.sub("#[A-Za-z0-9_]+", "", r) # remove the # in #hashtags
    r = re.sub(r'[\W\s]', '', r) # remove punctuation
    r = re.sub("(\.|\!{2,})", "\1", r) # 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 for w in r if not w 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:

	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 httpstco7xvjaojxg
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 httpstcoc8kyw0sjpl
...	...
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 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 httpstcosnezrquae
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 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 httpstcoltam0t0zca



## I. Pre-processing technique: Remove links

### The input:

		tweets
0	meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshov is dream novel start finish even if it is total enigma until very end find our reviews over zoella now	httpstcoeknggh8dg1
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	httpstco7xvjaojxg
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	httpstcoy7aivbh8wj
96	ad find art working with your cycle not against it with this handy guide head over	httpstco0uitgjsxaa full article packed full 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	httpstcosnezrquae
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	httpstcocxqv05wlf
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

### 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
    r = re.sub("#[A-Za-z0-9_]+", "", r) # remove the # in #hashtags
    r = re.sub(r'^\w\s', '', r) # remove punctuation
    r = re.sub(r'http\S+', '', r) # remove links
    r = re.sub("(.)\\1{2,}", "\\1", r) # 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 for w in r if not w 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:

		tweets
0	meticulously plotted fresh surprising but thrumming with that signature beth oleary warmth we all adore noshov is dream novel start finish even if it is total enigma until very end find our reviews over zoella now	
1	ad are you answering call summer selflove shespot x zoella edition is here head over	full roundup products more information link buy box
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	
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	
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	
...	...	
95	over zoella were celebrating best humanity life soul nhs people behind ppe moments that made them remarkable nurses they are today	
96	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	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	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	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	

## II. Getting the Polarity of the Tweets

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
sentiment_df
```

	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



#### 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))
```

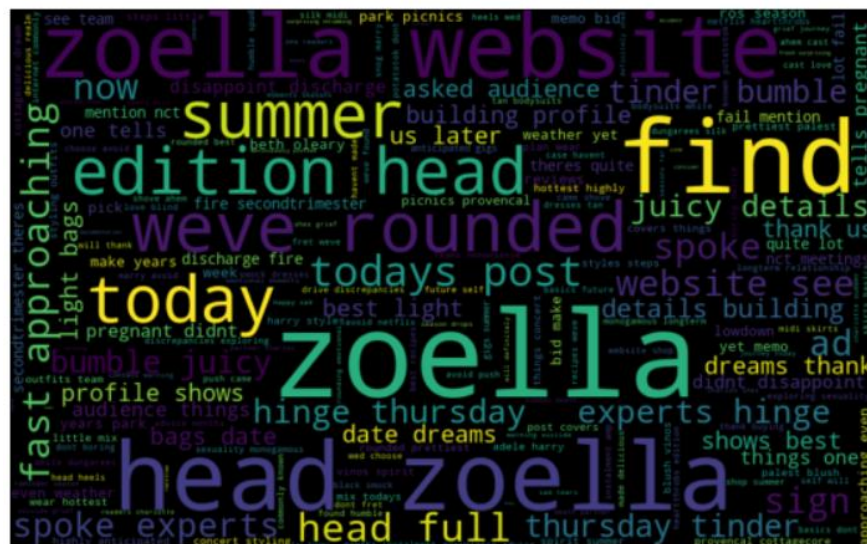
```
72.000000 percent of the tweets are positive.
15.000000 percent of the tweets are negative.
13.000000 percent of the tweets are neutral.
```

## 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"]

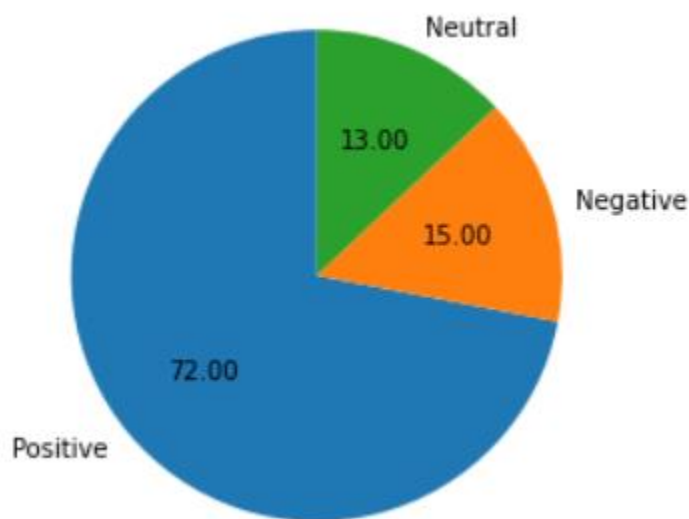
populationShare=[pos,neg,neu]

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)
```