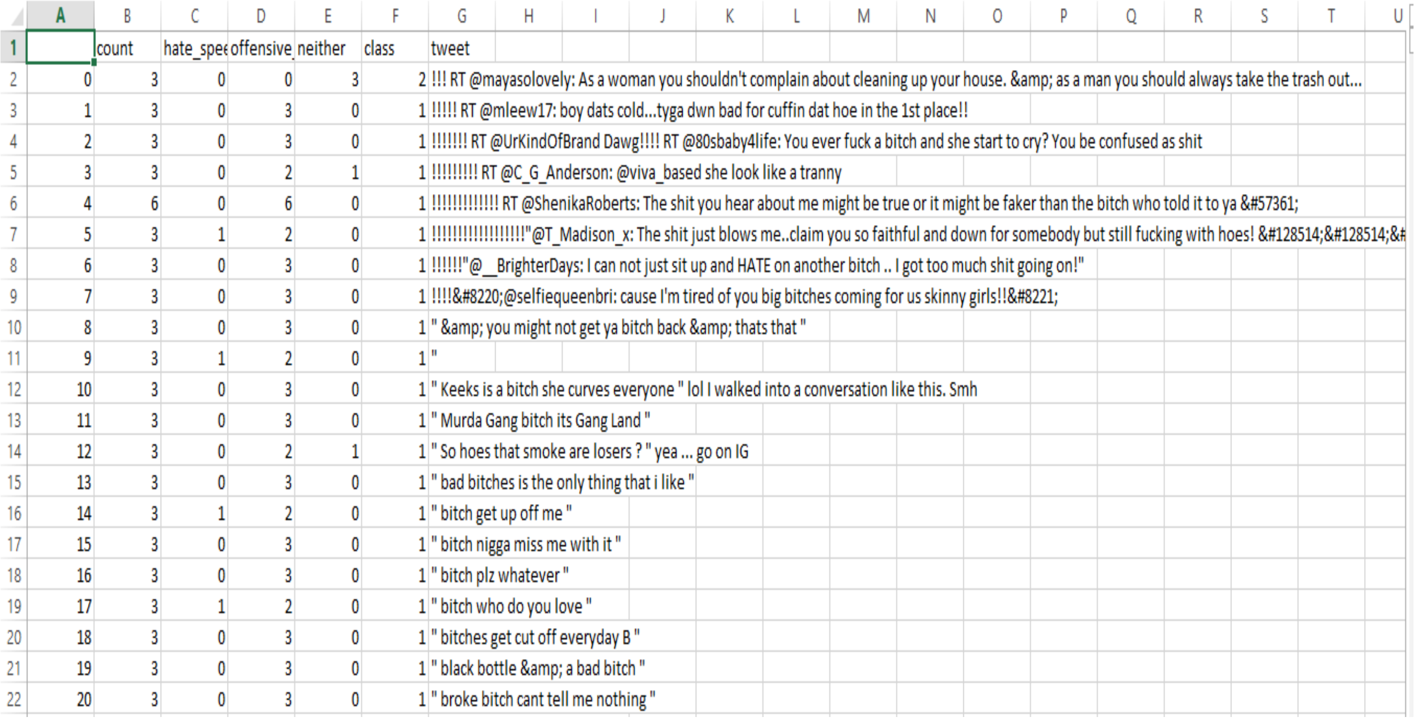
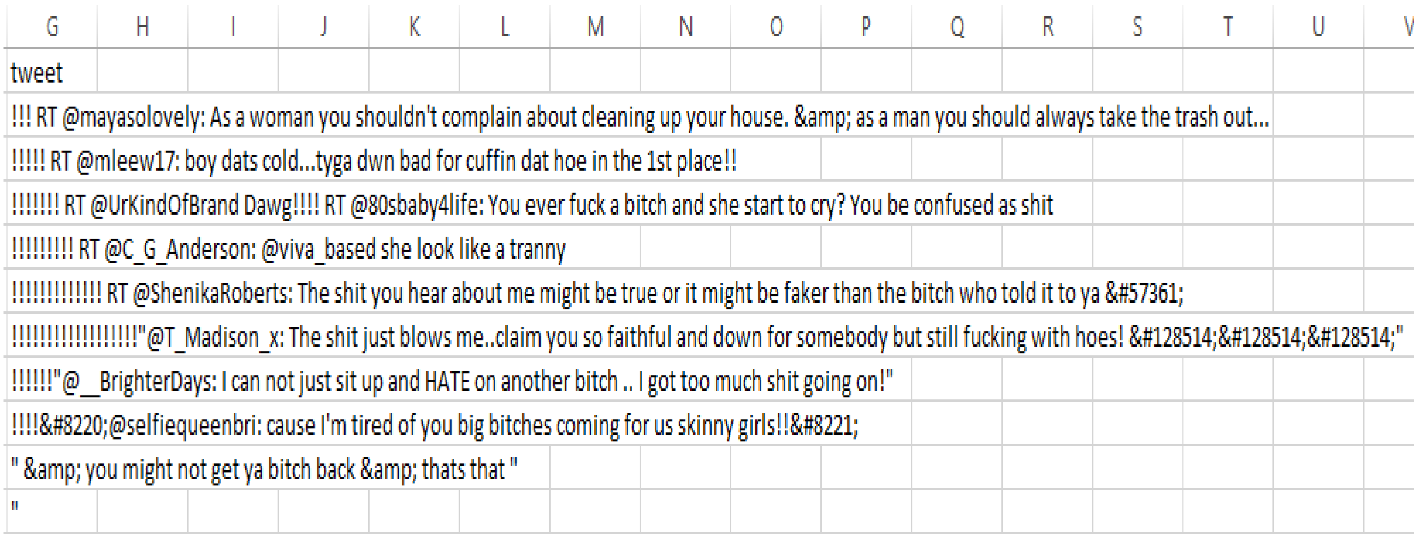
DOCUMENTATION FOR THE EVALUATION OF THE DEGREE OF PROFANITY IN THE TWITTER DATASET

IMPORTING THE DATASET

```
dataset = pd.read_csv("input_data.csv")
```



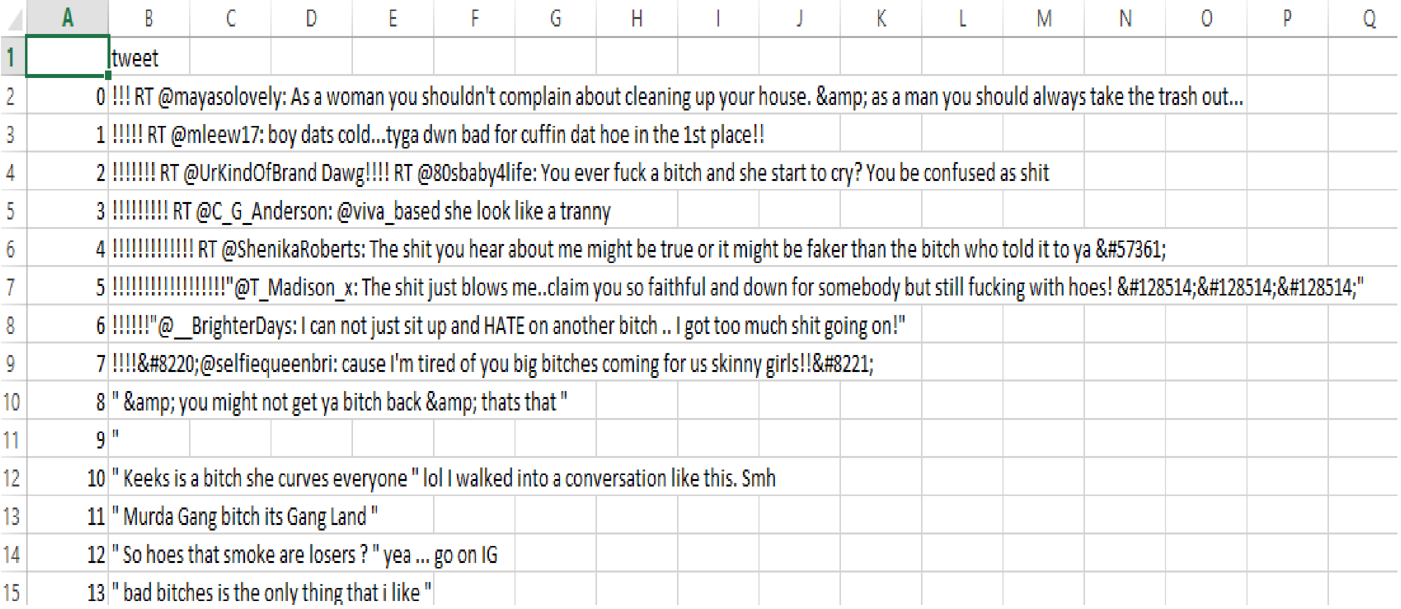
	A	B	C	D	E	F	G
1		count	hate_spee	offensive_	neither	class	tweet
2	0	3	0	0	3	2	!!! RT @mayasolovely: As a woman you shouldn't complain about cleaning up your house. &amp; as a man you should always take the trash out...
3	1	3	0	3	0	1	!!!!! RT @mleew17: boy dats cold...tyga dwn bad for cuffin dat hoe in the 1st place!!
4	2	3	0	3	0	1	!!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby4life: You ever fuck a bitch and she start to cry? You be confused as shit
5	3	3	0	2	1	1	!!!!!!!!! RT @C_G_Anderson: @viva_based she look like a tranny
6	4	6	0	6	0	1	!!!!!!!!!!!!! RT @ShenikaRoberts: The shit you hear about me might be true or it might be faker than the bitch who told it to ya &#57361;
7	5	3	1	2	0	1	!!!!!!!!!!!!!!!!!!"@T_Madison_x: The shit just blows me..claim you so faithful and down for somebody but still fucking with hoes! &#128514;&#128514;&#
8	6	3	0	3	0	1	!!!!!!"@__BrighterDays: I can not just sit up and HATE on another bitch .. I got too much shit going on!"
9	7	3	0	3	0	1	!!!!&#8220;@selfiequeenbri: cause I'm tired of you big bitches coming for us skinny girls!!&#8221;
10	8	3	0	3	0	1	" &amp; you might not get ya bitch back &amp; thats that "
11	9	3	1	2	0	1	"
12	10	3	0	3	0	1	" Keeks is a bitch she curves everyone " lol I walked into a conversation like this. Smh
13	11	3	0	3	0	1	" Murda Gang bitch its Gang Land "
14	12	3	0	2	1	1	" So hoes that smoke are losers ? " yea ... go on IG
15	13	3	0	3	0	1	" bad bitches is the only thing that i like "
16	14	3	1	2	0	1	" bitch get up off me "
17	15	3	0	3	0	1	" bitch nigga miss me with it "
18	16	3	0	3	0	1	" bitch plz whatever "
19	17	3	1	2	0	1	" bitch who do you love "
20	18	3	0	3	0	1	" bitches get cut off everyday B "
21	19	3	0	3	0	1	" black bottle &amp; a bad bitch "
22	20	3	0	3	0	1	" broke bitch cant tell me nothing "



G
tweet
!!! RT @mayasolovely: As a woman you shouldn't complain about cleaning up your house. &amp; as a man you should always take the trash out...
!!!!! RT @mleew17: boy dats cold...tyga dwn bad for cuffin dat hoe in the 1st place!!
!!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby4life: You ever fuck a bitch and she start to cry? You be confused as shit
!!!!!!!!! RT @C_G_Anderson: @viva_based she look like a tranny
!!!!!!!!!!!!! RT @ShenikaRoberts: The shit you hear about me might be true or it might be faker than the bitch who told it to ya &#57361;
!!!!!!!!!!!!!!!!!!"@T_Madison_x: The shit just blows me..claim you so faithful and down for somebody but still fucking with hoes! &#128514;&#128514;&#128514;"
!!!!!!"@__BrighterDays: I can not just sit up and HATE on another bitch .. I got too much shit going on!"
!!!!&#8220;@selfiequeenbri: cause I'm tired of you big bitches coming for us skinny girls!!&#8221;
" &amp; you might not get ya bitch back &amp; thats that "
"

If you look at the above dataset, then you will find the slur words and the usernames that we do not want for our analysis for profanity detection. There are many redundant words that we need to eliminate for easy computation.

Now, we will only work with column *tweet* and drop all the other columns as we don't need them for the analysis. Notice that the column *tweet* is at position B after removing the unwanted columns.



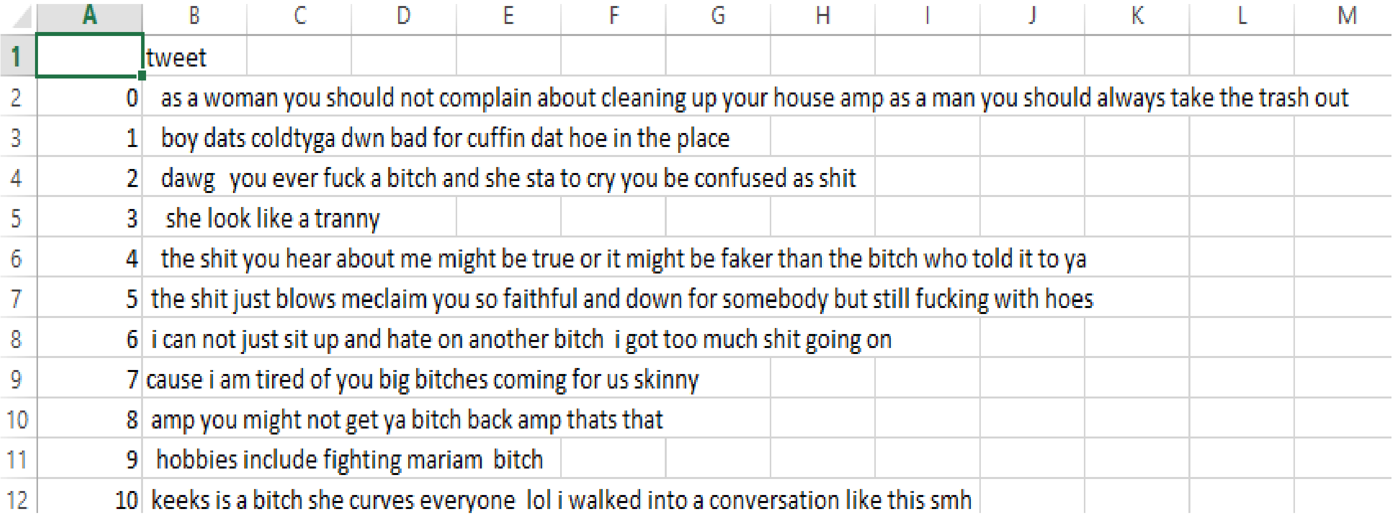
	A	B
1		tweet
2	0	!!! RT @mayasolovely: As a woman you shouldn't complain about cleaning up your house. &amp; as a man you should always take the trash out...
3	1	!!!!! RT @mleew17: boy dats cold...tyga dwn bad for cuffin dat hoe in the 1st place!!
4	2	!!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby4life: You ever fuck a bitch and she start to cry? You be confused as shit
5	3	!!!!!!!!! RT @C_G_Anderson: @viva_based she look like a tranny
6	4	!!!!!!!!!!!!! RT @ShenikaRoberts: The shit you hear about me might be true or it might be faker than the bitch who told it to ya &#57361;
7	5	!!!!!!!!!!!!!!!!!!"@T_Madison_x: The shit just blows me..claim you so faithful and down for somebody but still fucking with hoes! &#128514;&#128514;&#128514;"
8	6	!!!!!!"@__BrighterDays: I can not just sit up and HATE on another bitch .. I got too much shit going on!"
9	7	!!!!&#8220;@selfiequeenbri: cause I'm tired of you big bitches coming for us skinny girls!!&#8221;
10	8	" &amp; you might not get ya bitch back &amp; thats that "
11	9	"
12	10	" Keeks is a bitch she curves everyone " lol I walked into a conversation like this. Smh
13	11	" Murda Gang bitch its Gang Land "
14	12	" So hoes that smoke are losers ? " yea ... go on IG
15	13	" bad bitches is the only thing that i like "

CLEANING THE DATASET

We will now clean the data using clean text function.

```
def  clean_text(text):
  text =  text.lower()
  text = re.sub('@[^\s]+','',text)
  text = re.sub(r"i'm", "i am", text)
            . . . . .
            . . . . .
  text = re.sub(r"rt", "", text)
  return text
```

After using this function, we will get the following output which is much easier to work with.



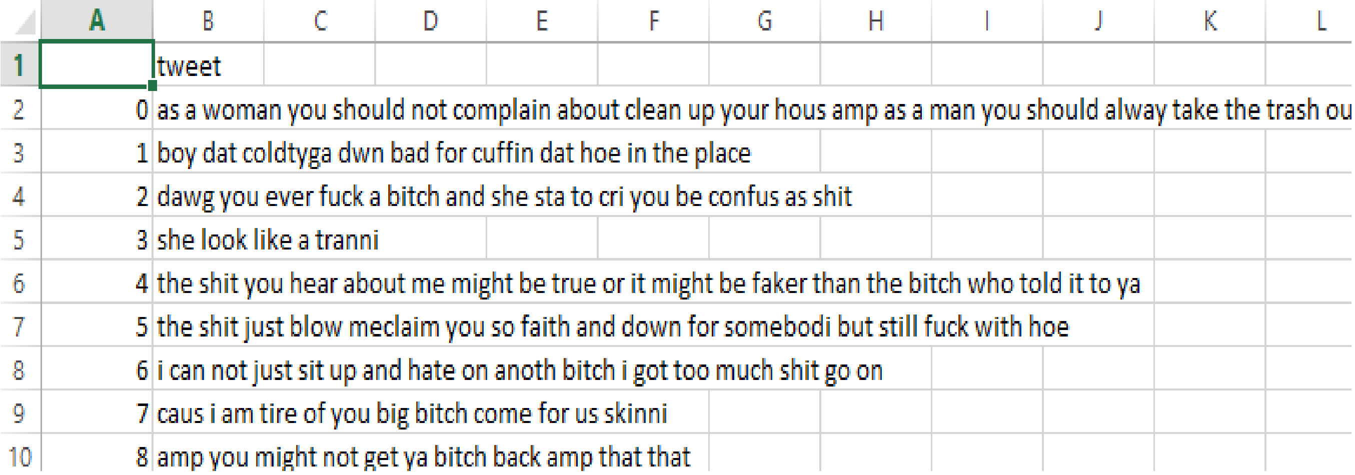
	A	B
1		tweet
2	0	as a woman you should not complain about cleaning up your house amp as a man you should always take the trash out
3	1	boy dats coldtyga dwn bad for cuffin dat hoe in the place
4	2	dawg you ever fuck a bitch and she sta to cry you be confused as shit
5	3	she look like a tranny
6	4	the shit you hear about me might be true or it might be faker than the bitch who told it to ya
7	5	the shit just blows meclaim you so faithful and down for somebody but still fucking with hoes
8	6	i can not just sit up and hate on another bitch i got too much shit going on
9	7	cause i am tired of you big bitches coming for us skinny
10	8	amp you might not get ya bitch back amp thats that
11	9	hobbies include fighting mariam bitch
12	10	keeks is a bitch she curves everyone lol i walked into a conversation like this smh

STEMMING

In this, we will stem each word of a tweet.

```
def stem_words(getDataset):
  sentences = []
  for index, row in getDataset.iterrows():
    sentence = ' '.join([snowball_stemmer.stem(word) for word in row["tweet"].split()])
    sentences.append(sentence)

  return pd.DataFrame({"tweet": sentences})
```



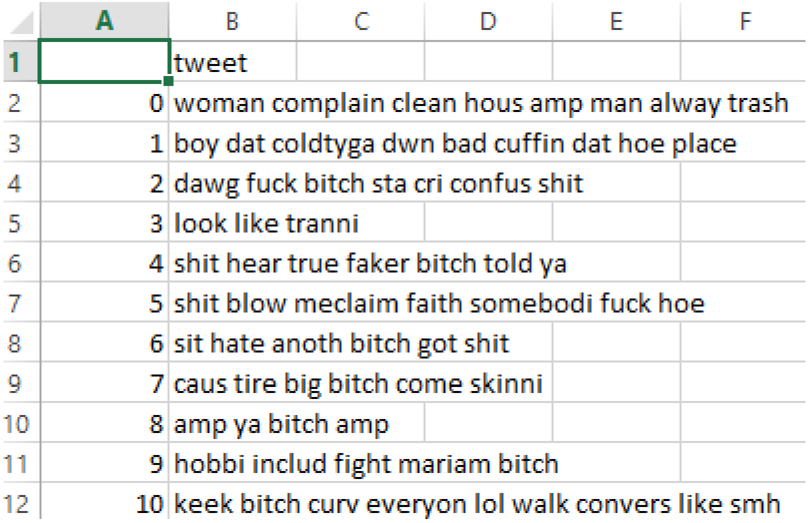
	A	B
1		tweet
2	0	as a woman you should not complain about clean up your hous amp as a man you should alway take the trash ou
3	1	boy dat coldtyga dwn bad for cuffin dat hoe in the place
4	2	dawg you ever fuck a bitch and she sta to cri you be confus as shit
5	3	she look like a tranni
6	4	the shit you hear about me might be true or it might be faker than the bitch who told it to ya
7	5	the shit just blow meclaim you so faith and down for somebodi but still fuck with hoe
8	6	i can not just sit up and hate on anoth bitch i got too much shit go on
9	7	caus i am tire of you big bitch come for us skinni
10	8	amp you might not get ya bitch back amp that that

STOPWORDS

Then, we will remove the stopwords.

```
def remove_stop_words(getDataset):
  sentences = []
  for index, row in getDataset.iterrows():
    sentence = ' '.join([word for word in row["tweet"].split() if word not in all_stopwords])
    sentences.append(sentence)

  return pd.DataFrame({"tweet": sentences})
```



	A	B	C	D	E	F
1		tweet				
2	0	woman complain clean hous amp man alway trash				
3	1	boy dat coldtyga dwn bad cuffin dat hoe place				
4	2	dawg fuck bitch sta cri confus shit				
5	3	look like tranni				
6	4	shit hear true faker bitch told ya				
7	5	shit blow meclaim faith somebodi fuck hoe				
8	6	sit hate anoth bitch got shit				
9	7	caus tire big bitch come skinni				
10	8	amp ya bitch amp				
11	9	hobbi includ fight mariam bitch				
12	10	keek bitch curv everyon lol walk convers like smh				

CENSORING THE SLUR WORDS AND CALCULATING THE DEGREE OF PROFANITY

In this, we will censor the slur words and count the **number of slur words per tweet as the degree of profanity**.

```
def censor_toxic_words(temp):
  sentences = []
  toxicity_degree = []
  for index, row in temp.iterrows():
    new_sentence = ""
    words_list = []
    toxicity_counter = 0
```

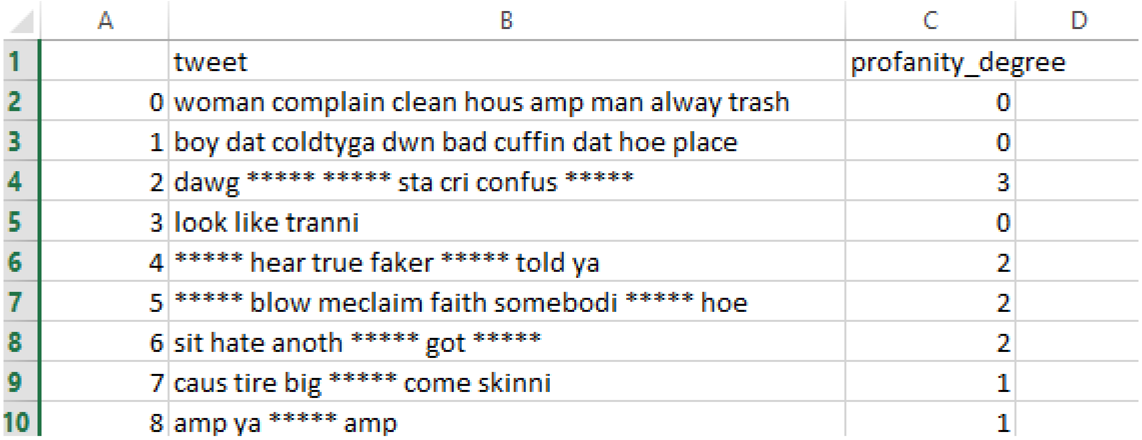
```
 for word in row["tweet"].split():
  if word in toxic_words:
    new_sentence += '***** '
    toxicity_counter += 1
  else:
    new_sentence += word + ' '

 words_list.append(new_sentence)
 toxicity_degree.append(toxicity_counter)
 sentence = ' '.join([word for word in words_list])
 sentences.append(sentence)
return pd.DataFrame({"tweet": sentences, "profanity_degree": toxicity_degree})
```

THE FINAL RESULT

Then we will output the result in the *output.csv* file.

```
# output the result
temp.to_csv("output_data.csv")
```



	A	B	C	D
1		tweet	profanity_degree	
2	0	woman complain clean hous amp man alway trash	0	
3	1	boy dat coldtyga dwn bad cuffin dat hoe place	0	
4	2	dawg ***** ***** sta cri confus *****	3	
5	3	look like tranni	0	
6	4	***** hear true faker ***** told ya	2	
7	5	***** blow meclaim faith somebodi ***** hoe	2	
8	6	sit hate anoth ***** got *****	2	
9	7	caus tire big ***** come skinni	1	
10	8	amp ya ***** amp	1	