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

IMPORTING THE DATASET

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

4	Α	В	С	D	E	F G	Н	I	J	K	L	М	N	0	Р	Q	R	S	Т	U
1	С	ount	hate_specoffe	nsive neit	her clas															
2	0	3	0	0	3			vely: As a w							amp; as a n	nan you sho	uld alway	s take the	trash out	
3	1	3	0	3	0			7: boy dats o												
4	2	3	0	3	0			OfBrand Da	_				itch and sl	ne start to	cry? You b	e confused	as shit			
5	3	3	0	2	1			Anderson: (
6	4	6	0	6	0			nenikaRobe				_		_						F4 4 . O.
/	5 6	3	0	3	0			T_Madison_ rDays: I can								ut still tuck	ing with h	0es: €	5514;€	514;&
9	7	3	0	3	0			equeenbri:	-				_							
0	8	3	0	3	0			t not get ya				conning it	/ us skilli	y girisaac	10221,					
1	9	3	1	2	0	1 "	p, you mign	r not get yu	DITON DUCK	xump, mu	its that									
2	10	3	0	3	0	_	s is a bitch s	she curves e	vervone " l	ol I walke	d into a con	versation I	ike this. S	mh						
3	11	3	0	3	0			h its Gang La	•											
14	12	3	0	2	1	1 " So h	oes that smo	oke are lose	rs ? " yea	go on IG										
15	13	3	0	3	0	1 " bad	bitches is th	e only thing	that i like '	•										
16	14	3	1	2	0	1 " bitch	get up off r	me"												
17	15	3	0	3	0			me with it "												
18	16	3	0	3	0		ı plz whatev													
19	17	3	1	2	0		who do yo													
20	18	3	0	3	0			ff everyday												
21	19	3	0	3	0			np; a bad bit												
22	20	3	U	3	0	1 DIOK	e bitch cant	tell me not	ning											
G		Н		J		K	L	М	N		0	Р)	R	S	Т		U	
				-							_		_						_	
twee	-																			
			/ely: As a v				•				ıse. &an	ıp; as a r	nan yo	ı shoul	d always	take the	trash c	out		
IIIII F	T @m	leew1	7: boy dats	coldty	yga dwn	n bad for c	uffin dat	hoe in th	ne 1st pl	ace!!										
	RT @	UrKind	OfBrand Da	awg!!!!	RT @80s	sbaby4life	: You ev	er fuck a	bitch an	d she st	art to cr	/? You b	e confu	sed as	shit					
	!! RT	@C G	Anderson:	@viva	based s	he look lil	ke a tranı	ny												
	!!!!!!	RT @Sh	enikaRobe	erts: The	shit yo	u hear ab	out me n	night be t	true or it	might	be faker	than th	e bitch	who to	d it to y	a ᙨ	1;			
	!!!!!!!	!!!!!"@ [:]	T Madison	x: The	shit just	t blows m	eclaim	you so fa	ithful an	d down	for som	ebody k	out still	fucking	with ho	es!	28514;&	#128514	;ԅ	514;"
			- Days: I can	_															-	
			equeenbri:									-								
			t not get ya			•	-			7 6		,								
ı	.6110				an wall	py critical														

IMPORTING THE LIST OF BAD WORDS

```
text_file = open("badwords.txt", "rt")
```

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. Note that the column *tweet* will be at position B after removing the unwanted columns.

1	Α	В	С	D	E	F	G	Н	- 1	J	K	L	М	N	0	Р	Q
1		tweet															
2	0	!!! RT @m	!! RT @mayasolovely: As a woman you shouldn't complain about cleaning up your house. & as a man you should always take the trash out														
3	1	!!!!! RT @	!!!!! RT @mleew17: boy dats coldtyga 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	!!!!!!!!! R7	!!!!!!!! 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 														
7	5	!!!!!!!!!!@T_Madison_x: The shit just blows meclaim you so faithful and down for somebody but still fucking with hoes! 😂 😂 😂 "											514;"				
8	6	!!!!!!"@_	_BrighterDa	ays: I can n	ot just sit (up and HAT	E on anoth	ner bitch	I got too m	uch shit go	ing on!"						
9	7	!!!!“	0;@selfieq	ueenbri: c	ause I'm ti	red of you	big bitche	s coming fo	or us skinny	/ girls!!	221;						
10	8	" & y	ou might n	ot get ya b	itch back 8	kamp; that	s that "										
11	9	II															
12	10	" Keeks is	a bitch she	curves ev	eryone " l	ol I walked	into a con	versation l	ike this. Sr	nh							
13	11	" Murda G	'Murda Gang bitch its Gang Land "														
14	12	" So hoes	So hoes that smoke are losers ? " yea go on IG														
15	13	" bad bitc	hes is the o	only thing t	hat i like "												

CLEANING THE DATASET

We will now clean the data using clean text function.

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

4	Α	В	C	D	E	F	G	Н	1	J	K	L	M	
1		tweet												
2	0	as a wor	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 date	boy dats coldtyga dwn bad for cuffin dat hoe in the place											
4	2	dawg y	ou ever fu	ck a bitch a	nd she sta	to cry you	be confuse	ed as shit						
5	3	she lool	k like a trar	nny										
6	4	the shit	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 j	ust blows r	neclaim yo	u so faithf	ul and dov	n for some	ebody but	still fucking	g with hoe	S			
8	6	i can not	just sit up a	and hate o	n another l	bitch i got	too much s	hit going o	n					
9	7	cause i an	n tired of y	ou big bitc	hes comin	g for us ski	nny							
10	8	amp you	might not	get ya bitcl	h back amp	thats that								
11	9	hobbies	include fig	hting mari	am bitch									
12	10	keeks is a	a bitch she	curves eve	ryone lol	i walked in	to a conve	rsation 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})
```

4	Α	В	С	D	Е	F	G	Н	1	J	K	L	
1		tweet											
2	0	as a woma	as a woman you should not complain about clean up your hous amp as a man you should alway take the trash										
3	1	boy dat co	boy dat coldtyga dwn bad for cuffin dat hoe in the place										
4	2	dawg you	dawg you ever fuck a bitch and she sta to cri you be confus as shit										
5	3	she look l	ike a trann	i									
6	4	the shit yo	ou hear ab	out me mig	ght be true	or it migh	t be faker t	than the bi	tch who to	ld it to ya			
7	5	the shit ju	ist blow me	eclaim you	so faith ar	nd down fo	r someboo	di but still f	fuck with h	oe			
8	6	i can not j	ust sit up a	nd hate or	anoth bit	ch i got too	much shit	go on					
9	7	caus i am	tire of you	big bitch c	ome for us	skinni							
10	8	amp you r	night not g	et 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})
```

Δ	Α	В	С	D	Е	F						
1		tweet										
2	0	woman co	mplain cle	an hous ar	np man alv	way trash						
3	1	boy dat co	ooy dat coldtyga dwn bad cuffin dat hoe place									
4	2	dawg fuck	bitch sta c	ri confus s	hit							
5	3	look like t										
6	4	shit hear t	shit hear true faker bitch told ya									
7	5	shit blow	meclaim fa	ith someb	odi fuck h	oe						
8	6	sit hate ar	oth bitch g	got shit								
9	7	caus tire b	ig bitch co	me skinni								
10	8	amp ya bi	tch amp									
11	9	hobbi incl	ud fight m	ariam bitch	1							
12	10	keek bitch	curv ever	yon lol wal	k convers	like smh						

CENSORING THE SLUR WORDS AND CALCULATING THE DEGREE OF PROFANITY

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

THE FINAL RESULT

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

output the result

temp.to_csv("output_data.csv")

1	Α	В	С	D
1		tweet	profanity_deg	gree
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	