

Prepared_data

January 10, 2019

1 Drop nulls and merge labeled DataSets

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from bs4 import BeautifulSoup
import re
```

1.1 First DataSet

```
In [2]: #Here we upload the users and the extract tweets to our DataFrame
cols = ['', 'Tweets', 'Name', 'Sentiment', '']
dataset = pd.read_excel("CLEANED_DATASET_1_Labeled.xlsx", header=None, names=cols, encoding='utf-8')
dataset.head()
```

```
Out[2]:
```

		Tweets	Name	Sentiment	
0	NaN	Tweets	Name	Sentiment	NaN
1	0.0	tweets	Name	NaN	NaN
2	1.0	NaN	AkincilarCW	0	NaN
3	2.0	NaN	AkincilarCW	0	NaN
4	3.0	NaN	AkincilarCW	0	NaN

```
In [3]: # Drop the columns that we don't need
dataset.drop(['', ''], axis=1, inplace=True)
dataset[dataset.Tweets == 0].head(5)
```

```
Out[3]: Empty DataFrame
Columns: [Tweets, Name, Sentiment]
Index: []
```

```
In [4]: #dealing with missing data
sufficientcolumns = dataset.isnull().sum() < 2
df = dataset.loc[:, sufficientcolumns] # keep only columns that have less than 2 missing values
df = dataset.drop(dataset.loc[dataset['Tweets'].isnull()].index) # remove the sample tweets with missing values
df.head()
```

```
Out[4]:
```

	Tweets	Name	Sentiment
0	Tweets	Name	Sentiment

	tweets	Name	NaN
1			
5	egypt s official news agency mena website hack...	AkincilarCW	1
7	middle east news agency the state news agency ...	AkincilarCW	1
11	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0

```
In [5]: dataset1 = df.ix[2:]
```

```
/opt/jupyterhub/anaconda/lib/python3.6/site-packages/ipykernel_launcher.py:1: DeprecationWarning:
.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing
```

See the documentation here:

<http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-deprecated>
 """Entry point for launching an IPython kernel.

```
In [6]: dataset1.head()
```

```
Out[6]:
```

	Tweets	Name	Sentiment
5	egypt s official news agency mena website hack...	AkincilarCW	1
7	middle east news agency the state news agency ...	AkincilarCW	1
11	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
12	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
13	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0

```
In [7]: #Here we upload the users and the extract tweets to our DataFrame
```

```
cols = ['', 'Tweets', 'Name', 'Sentiment', '', '', '', '', '', '']
dataset2 = pd.read_excel("CLEANED_DATASET_2_Labeled.xlsx", header=None, names=cols, encoding='utf-8')
dataset2.head()
```

```
Out[7]:
```

		Tweets	Name	\
0	NaN	Tweets	Name	
1	0.0	tweets	Name	
2	1.0	anonymousawake links	jacobwolf420_x	
3	2.0	ofc v	jacobwolf420_x	
4	3.0	kangna ranaut timesofindia idk so confused is ...	jacobwolf420_x	

	Sentiment
0	Sentiment NaN Related to hacking: 1 NaN NaN NaN NaN
1	NaN NaN Other: 0 NaN NaN NaN NaN
2	1 NaN NaN NaN NaN NaN NaN
3	0 NaN NaN NaN NaN NaN NaN
4	0 NaN NaN NaN NaN NaN NaN

```
In [8]: # Drop the columns that we don't need
```

```
dataset2.drop(['', '', '', '', '', '', '', ''], axis=1, inplace=True)
dataset2[dataset2.Tweets == 0].head(5)
```

```
Out[8]: Empty DataFrame
Columns: [Tweets, Name, Sentiment]
Index: []
```

```
In [9]: #dealing with missing data
sufficientcolumns = df_1.isnull().sum() < 2
df = dataset2.loc[:, sufficientcolumns] # keep only columns that have less than 2 miss
df = dataset2.drop(dataset2.loc[dataset2['Tweets'].isnull()].index) # remove the sampl
df.head()
```

```
-----
NameError                                Traceback (most recent call last)
```

```
<ipython-input-9-0112601c9730> in <module>
      1 #dealing with missing data
----> 2 sufficientcolumns = df_1.isnull().sum() < 2
      3 df = dataset2.loc[:, sufficientcolumns] # keep only columns that have less than 2 m
      4 df = dataset2.drop(dataset2.loc[dataset2['Tweets'].isnull()].index) # remove the s
      5 df.head()
```

```
NameError: name 'df_1' is not defined
```

```
In [10]: dataset2=df.ix[2:]
```

```
/opt/jupyterhub/anaconda/lib/python3.6/site-packages/ipykernel_launcher.py:1: DeprecationWarning:
.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing
```

See the documentation here:

<http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-deprecated>
"""Entry point for launching an IPython kernel.

```
In [11]: dataset2.head()
```

```
Out[11]:
```

	Tweets	Name	Sentiment
5	egypt s official news agency mena website hack...	AkincilarCW	1
7	middle east news agency the state news agency ...	AkincilarCW	1
11	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
12	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
13	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0

```
In [12]: Full_dataset = [dataset1,dataset2]
DataSet = pd.concat(Full_dataset)
DataSet.head()
```

```
Out[12]:
```

	Tweets	Name	Sentiment
5	egypt s official news agency mena website hack...	AkincilarCW	1
7	middle east news agency the state news agency ...	AkincilarCW	1
11	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
12	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
13	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0

```
In [13]: DataSet.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10208 entries, 5 to 5221
Data columns (total 3 columns):
Tweets      10208 non-null object
Name        10208 non-null object
Sentiment    10204 non-null object
dtypes: object(3)
memory usage: 319.0+ KB
```

2 Cleaned Labeled Dataset

```
In [14]: DataSet.to_csv('Labeled_Dataset.csv',encoding='utf-8')
csv = 'Labeled_Dataset.csv'
my_df = pd.read_csv(csv,index_col=0)
my_df
```

```
Out[14]:
```

	Tweets	Name \
5	egypt s official news agency mena website hack...	AkincilarCW
7	middle east news agency the state news agency ...	AkincilarCW
11	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW
12	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW
13	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW
14	t rk hackerlar m s r resmi ajans n n sitesini ...	AkincilarCW
15	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW
16	t rk hacker grubu cyberwarriortim taraf ndan m...	AkincilarCW
17	kurban bayram n n lkemize ve milletimize huzur...	AkincilarCW
18	kendilerini k rdish hackers olarak tan mlayan ...	AkincilarCW
19	bu lkede vatanseverler alimler hatipler as l p...	AkincilarCW
20	haziran tarihinde irak kuzeyi kandil b lgesine...	AkincilarCW
21	cyberwarriortim akincilar ter r rg t pkk opera...	AkincilarCW
22	cyberwarriortim oku ren payla kitap seferberli i	AkincilarCW
23	ba ta aziz ehitlerimizi rahmetle gazilerimizi ...	AkincilarCW
24	ak nc lar hacker grubu dirili postas na konu t...	AkincilarCW
25	selam naleyk m fsm hastanesinde yatan dedem i ...	AkincilarCW
26	lg n t rkler akincilar	AkincilarCW
27	avrupa da g ndem akincilar	AkincilarCW
28	t rk hackerler yunan ld rtt	AkincilarCW
29	yunanistan n g ndemi akincilar	AkincilarCW

30	siirt eruh da icra edilen hava destekli operas...	AkincilarCW
31	alt yap al malar m z tamamlanarak web sitemize...	AkincilarCW
32	ile saatleri aras nda cyberwarriortim sunucula...	AkincilarCW
33	mesele onlar i in milli birlik ve beraberlik d...	AkincilarCW
34	erdo an bah eli destici k l daro lu karamollao...	AkincilarCW
35	temmuz ha n darbe giri imini yapan nsanlar m z...	AkincilarCW
36	temmuz hain darbe giri imini yapan insanlar m ...	AkincilarCW
37	akincilar ypg operasyonu	AkincilarCW
38	vatan sevgisini s z nde de il z nde ya ayan pl...	AkincilarCW
...
5192	islamic state warns of black days for saudi ar...	IzzahHackers
5193	labienustitus anonopsaudix we dont forget our ...	IzzahHackers
5194	israel jails palestinians for facebook comments	IzzahHackers
5195	rohingya muslims found on boat rescued by myan...	IzzahHackers
5196	rohingya muslims found on boat rescued by myan...	IzzahHackers
5197	mass graves found in malaysia close to detenti...	IzzahHackers
5198	irfan death of humanity burma says muslims not...	IzzahHackers
5199	while syrian rebels fight hezbollah assad in e...	IzzahHackers
5200	just out of curiosity have hezbollah ever even...	IzzahHackers
5201	richard clarke s on iraq we destroyed iraq dae...	IzzahHackers
5202	day you know as whole the world is albeit slow...	IzzahHackers
5203	fox news cnn on the right wing deceivers and p...	IzzahHackers
5204	presstv iranian government influenced media ru...	IzzahHackers
5205	saudi regime involved in attack on qatif mosqu...	IzzahHackers
5206	saudi regime involved in attack on qatif mosqu...	IzzahHackers
5207	these politicians all the same everything when...	IzzahHackers
5208	down g n ration identitaire anti muslim hate g...	IzzahHackers
5209	breaking the senate has just passed its tpp fa...	IzzahHackers
5210	put israel on theunlist of serious rights viol...	IzzahHackers
5211	walts insight thank you for the follow	IzzahHackers
5212	opdarrienhunt anonrastaftp ted bingham susaa m...	IzzahHackers
5213	a malcolmite the bombing of a mosque will gene...	IzzahHackers
5214	america uk are protesting government corruptio...	IzzahHackers
5215	they don t want to take any responsibility for...	IzzahHackers
5216	then the nd gulf war happened and al qaida in ...	IzzahHackers
5217	al qaida was created after the st gulf war in ...	IzzahHackers
5218	the media who care a lot more about ratings an...	IzzahHackers
5219	us and nato occupations interventions led to t...	IzzahHackers
5220	the video next is tough to watch but it s a mu...	IzzahHackers
5221	months before isis was even formed reports of ...	IzzahHackers

Sentiment

5	1
7	1
11	0
12	0
13	0
14	0

15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0
26	0
27	0
28	0
29	0
30	0
31	0
32	0
33	0
34	0
35	0
36	0
37	0
38	0
...	...
5192	0
5193	0
5194	0
5195	0
5196	0
5197	0
5198	0
5199	0
5200	0
5201	0
5202	0
5203	0
5204	0
5205	0
5206	0
5207	0
5208	0
5209	0
5210	1
5211	0
5212	0
5213	0
5214	0

```

5215      0
5216      0
5217      0
5218      0
5219      0
5220      0
5221      1

```

```
[10208 rows x 3 columns]
```

```
In [15]: my_df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 10208 entries, 5 to 5221
Data columns (total 3 columns):
Tweets      10208 non-null object
Name        10208 non-null object
Sentiment    10204 non-null object
dtypes: object(3)
memory usage: 319.0+ KB

```

```
In [19]: Filter1 = my_df[my_df['Name']=='AkincilarCW']
```

```
In [22]: Filter1
```

```

Out[22]:

```

	Tweets	Name	Sentiment
5	egypt s official news agency mena website hack...	AkincilarCW	1
7	middle east news agency the state news agency ...	AkincilarCW	1
11	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
12	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
13	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
14	t rk hackerlar m s r resmi ajans n n sitesini ...	AkincilarCW	0
15	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW	0
16	t rk hacker grubu cyberwarriortim taraf ndan m...	AkincilarCW	0
17	kurban bayram n n lkemize ve milletimize huzur...	AkincilarCW	0
18	kendilerini k rdish hackers olarak tan mlayan ...	AkincilarCW	0
19	bu lkede vatanseverler alimler hatipler as l p...	AkincilarCW	0
20	haziran tarihinde irak kuzeyi kandil b lgesine...	AkincilarCW	0
21	cyberwarriortim akincilar ter r rg t pkk opera...	AkincilarCW	0
22	cyberwarriortim oku ren payla kitap seferberli i	AkincilarCW	0
23	ba ta aziz ehitlerimizi rahmetle gazilerimizi ...	AkincilarCW	0
24	ak nc lar hacker grubu dirili postas na konu t...	AkincilarCW	0
25	selam naleyk m fsm hastanesinde yatan dedem i ...	AkincilarCW	0
26	lg n t rkler akincilar	AkincilarCW	0
27	avrupa da g ndem akincilar	AkincilarCW	0
28	t rk hackerler yunan ld rtt	AkincilarCW	0
29	yunanistan n g ndemi akincilar	AkincilarCW	0
30	siirt eruh da icra edilen hava destekli operas...	AkincilarCW	0

31	alt yap al malar m z tamamlanarak web sitemize...	AkincilarCW	0
32	ile saatleri aras nda cyberwarriortim sunucula...	AkincilarCW	0
33	mesele onlar i in milli birlik ve beraberlik d...	AkincilarCW	0
34	erdo an bah eli destici k l daro lu karamollao...	AkincilarCW	0
35	temmuz ha n darbe giri imini yapan nsanlar m z...	AkincilarCW	0
36	temmuz hain darbe giri imini yapan insanlar m ...	AkincilarCW	0
37	akincilar ypg operasyonu	AkincilarCW	0
38	vatan sevgisini s z nde de il z nde ya ayan pl...	AkincilarCW	0
..
170	sabah b nyeniz de mevl t tezel denen m ptezeli...	AkincilarCW	0
171	akinsaglamas whatsapp tan konum yollar m s n	AkincilarCW	0
172	ruhu ad mekan cennet olsun ekimd nyakarizmag n	AkincilarCW	0
173	hay rl cumalar allah dualar n z kabul etsin	AkincilarCW	0
174	aydinerol ilyassalmanfan mervepekdemirrr melih...	AkincilarCW	0
175	ilyassalmanfan mervepekdemirrr melihgokcek do ...	AkincilarCW	0
176	basaranalper maalesef sa malam s n z	AkincilarCW	0
177	defacerdarkbat tebrikler yi idim ama a g z n u...	AkincilarCW	0
178	ki ili ini makam ndan alanlar makamdan sonra k...	AkincilarCW	0
179	b t n demo ullar g nahk rd r g nahk rlar n en ...	AkincilarCW	0
180	tarikkbjk ufukcc bu bikini giyip s z n n eri o...	AkincilarCW	0
181	kwonmusti tarikkbjk ufukcc elem tere fi kem g ...	AkincilarCW	0
182	tarikkbjk sercanarga bazen tehlikeli seyler ya...	AkincilarCW	0
183	sercanarga ma i in de il fenerbah e i in tarik...	AkincilarCW	0
184	sercanarga en son la geliyordum	AkincilarCW	0
185	ufukcc tarikkbjk kwonmusti daha ge en hafta nc...	AkincilarCW	0
186	tarikkbjk kwonmusti ufukcc bak bu oldu	AkincilarCW	0
187	kwonmusti ufukcc tarikkbjk bi daha da trabzon ...	AkincilarCW	0
188	ufukcc kwonmusti tarikkbjk bir fenerli ma m z ...	AkincilarCW	0
189	ufukcc kwonmusti tarikkbjk ta sessiz sedas z a...	AkincilarCW	0
190	kwonmusti ak am fener de b yle hezimete u ruyo...	AkincilarCW	0
191	kwonmusti oldu allah n seven defansa gelsin	AkincilarCW	0
192	kwonmusti m stehak abi	AkincilarCW	0
193	cwlazmania daha bu ne ki	AkincilarCW	0
194	ebubekir gereksizin biri medya maymunu	AkincilarCW	0
195	mazlumun umudu zalimin korkulu r yas tsk dlib ...	AkincilarCW	0
196	d nyahayat n nger e i l m var	AkincilarCW	0
197	ufukcc tarikkbjk negredo nun gole ula mas na s...	AkincilarCW	0
198	yildirim ankara n n ba kent ilan edili inin y ...	AkincilarCW	0
199	cuman z m barek olsun	AkincilarCW	0

[380 rows x 3 columns]

In [34]: Stat1 = Filter1[Filter1['Sentiment']=='1']

In [36]: Stat2 = Filter1[Filter1['Sentiment']=='0']

In [35]: Stat1

Out [35]:		Tweets	Name	Sentiment
5	egypt s official news agency mena website hack...	AkincilarCW		1

7	middle east news agency the state news agency ...	AkincilarCW	1
125	cyber warrior tim akincilar srail seferi intro...	AkincilarCW	1
126	so it appears the times of israel website was ...	AkincilarCW	1
127	has the timesofisrael homepage been hacked thi...	AkincilarCW	1
5	egypt s official news agency mena website hack...	AkincilarCW	1
7	middle east news agency the state news agency ...	AkincilarCW	1
125	cyber warrior tim akincilar srail seferi intro...	AkincilarCW	1
126	so it appears the times of israel website was ...	AkincilarCW	1
127	has the timesofisrael homepage been hacked thi...	AkincilarCW	1

In [37]: Stat2

Out [37]:

		Tweets	Name	Sentiment
11	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW		0
12	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW		0
13	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW		0
14	t rk hackerlar m s r resmi ajans n n sitesini ...	AkincilarCW		0
15	t rk hackerlar m s r resmi haber ajans n n sit...	AkincilarCW		0
16	t rk hacker grubu cyberwarriortim taraf ndan m...	AkincilarCW		0
17	kurban bayram n n lkemize ve milletimize huzur...	AkincilarCW		0
18	kendilerini k rdish hackers olarak tan mlayan ...	AkincilarCW		0
19	bu lkede vatanseverler alimler hatipler as l p...	AkincilarCW		0
20	haziran tarihinde irak kuzeyi kandil b lgesine...	AkincilarCW		0
21	cyberwarriortim akincilar ter r rg t pkk opera...	AkincilarCW		0
22	cyberwarriortim oku ren payla kitap seferberli i	AkincilarCW		0
23	ba ta aziz ehitlerimizi rahmetle gazilerimizi ...	AkincilarCW		0
24	ak nc lar hacker grubu dirili postas na konu t...	AkincilarCW		0
25	selam naleyk m fsm hastanesinde yatan dedem i ...	AkincilarCW		0
26	lg n t rkler akincilar	AkincilarCW		0
27	avrupa da g ndem akincilar	AkincilarCW		0
28	t rk hackerler yunan ld rtt	AkincilarCW		0
29	yunanistan n g ndemi akincilar	AkincilarCW		0
30	siirt eruh da icra edilen hava destekli operas...	AkincilarCW		0
31	alt yap al malar m z tamamlanarak web sitemize...	AkincilarCW		0
32	ile saatleri aras nda cyberwarriortim sunucula...	AkincilarCW		0
33	mesele onlar i in milli birlik ve beraberlik d...	AkincilarCW		0
34	erdo an bah eli destici k l daro lu karamollao...	AkincilarCW		0
35	temmuz ha n darbe giri imini yapan nsanlar m z...	AkincilarCW		0
36	temmuz hain darbe giri imini yapan insanlar m ...	AkincilarCW		0
37	akincilar ypg operasyonu	AkincilarCW		0
38	vatan sevgisini s z nde de il z nde ya ayan pl...	AkincilarCW		0
39	rt kirmizicehrel i ypg resmi websitesi a cyberw...	AkincilarCW		0
40	ter r rg t ypg operasyonu	AkincilarCW		0
..
170	sabah b nyeniz de mevl t tezel denen m ptezeli...	AkincilarCW		0
171	akinsaglamas whatsapp tan konum yollar m s n	AkincilarCW		0
172	ruhu ad mekan cennet olsun ekimd nyakarizmag n	AkincilarCW		0
173	hay rl cumalar allah dualar n z kabul etsin	AkincilarCW		0

174	aydinerol ilyassalmanfan mervepekdemirrr melih...	AkincilarCW	0
175	ilyassalmanfan mervepekdemirrr melihgokcek do ...	AkincilarCW	0
176	basaranalper maalesef sa malam s n z	AkincilarCW	0
177	defacerdarkbat tebrikler yi idim ama a g z n u...	AkincilarCW	0
178	ki ili ini makam ndan alanlar makamdan sonra k...	AkincilarCW	0
179	b t n demo ullar g nahk rd r g nahk rlar n en ...	AkincilarCW	0
180	tarikkbjk ufukcc bu bikini giyip s z n n eri o...	AkincilarCW	0
181	kwonmusti tarikkbjk ufukcc elem tere fi kem g ...	AkincilarCW	0
182	tarikkbjk sercanarga bazen tehlikeli seyler ya...	AkincilarCW	0
183	sercanarga ma i in de il fenerbah e i in tarik...	AkincilarCW	0
184	sercanarga en son la geliyordum	AkincilarCW	0
185	ufukcc tarikkbjk kwonmusti daha ge en hafta nc...	AkincilarCW	0
186	tarikkbjk kwonmusti ufukcc bak bu oldu	AkincilarCW	0
187	kwonmusti ufukcc tarikkbjk bi daha da trabzon ...	AkincilarCW	0
188	ufukcc kwonmusti tarikkbjk bir fenerli ma m z ...	AkincilarCW	0
189	ufukcc kwonmusti tarikkbjk ta sessiz sedas z a...	AkincilarCW	0
190	kwonmusti ak am fener de b yle hezimete u ruyo...	AkincilarCW	0
191	kwonmusti oldu allah n seven defansa gelsin	AkincilarCW	0
192	kwonmusti m stehak abi	AkincilarCW	0
193	cwlazmania daha bu ne ki	AkincilarCW	0
194	ebubekir gereksizin biri medya maymunu	AkincilarCW	0
195	mazlumun umudu zalimin korkulu r yas tsk dlib ...	AkincilarCW	0
196	d nyahayat n nger e i l m var	AkincilarCW	0
197	ufukcc tarikkbjk negredo nun gole ula mas na s...	AkincilarCW	0
198	yildirim ankara n n ba kent ilan edili inin y ...	AkincilarCW	0
199	cuman z m barek olsun	AkincilarCW	0

[370 rows x 3 columns]

In [41]: Value1=Stat1.count()

In [42]: Porcen = Value1/370

In [49]: print((Value1['Sentiment']/370)*100)

2.7027027027027026