

In [241]:

```
#import pandas package to process the data
#import os package to be able to locate WD
import pandas as pd
import os
import matplotlib
from matplotlib import figure
import numpy as np
from scipy import stats
import tweepy
import math
import time
import sys
import datetime
import copy
from datetime import datetime
import json
import re
import matplotlib.pyplot as plt
import plotly.plotly as py
from statsmodels.sandbox.regression.predstd import wls_prediction_std
from __future__ import print_function
from statsmodels.compat import lzip
import statsmodels.api as sm
from statsmodels.formula.api import ols
import seaborn as sns
import itertools
from itertools import chain, combinations
import statsmodels.formula.api as smf
import scipy.stats as scipystats
import statsmodels.api as sm
import statsmodels.stats.stattools as stools
import statsmodels.stats as stats
from statsmodels.graphics.regressionplots import *
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cross_validation import train_test_split
from statsmodels.formula.api import logit, probit, poisson, ols
```

In [242]:

```
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from nltk.sentiment.util import *
nltk.download('vader_lexicon')
from nltk import tokenize
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]      /Users/trishaharjani/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

Reading, filter for english, sort chronologically

In [243]:

```
#Assigning the value 'data' to my CSV file
data = pd.read_csv('ref_data.csv', low_memory=False, dtype=None, encoding = 'utf_8')
```

In [244]:

```
pd.set_option('display.max_columns', None)
pd.options.display.max_colwidth = 1000
data.head(10)
```

Out[244]:

	coordinates	created_at	hashtags	
0	NaN	Mon May 21 17:16:37 +0000 2018	Together4Yes	NaN
1	NaN	Mon May 21 17:18:05 +0000 2018	NaN	NaN
2	NaN	Mon May 21 17:16:13 +0000 2018	repeal loveboth	NaN
3	NaN	Mon May 21 17:17:22 +0000	NaN	NaN

		2018		
4	NaN	Mon May 21 17:17:40 +0000 2018	repealthe8th together4yes	NaN
5	NaN	Mon May 21 17:16:20 +0000 2018	NaN	NaN
6	NaN	Mon May 21 17:16:50 +0000 2018	repealthe8th	NaN
7	NaN	Mon May 21 17:16:31 +0000 2018	NaN	NaN
8	NaN	Mon May 21 17:17:42 +0000 2018	Together4Yes	https://twitter.com/DBNRepeal/status/998610
9	NaN	Mon May 21 17:16:57 +0000 2018	ITSTIME	NaN

In [245]:

```
pd.set_option('display.max_columns', None)
pd.options.display.max_colwidth = 1000
data['rtt'] = data['text'].str.contains('RT')==True
```

In [246]:

```
data.shape
```

Out[246]:

```
(2027025, 35)
```

In [247]:

```
data_users = data.drop_duplicates(subset = 'user_screen_name.1')
data_users.shape
```

Out[247]:

```
(426353, 35)
```

In [248]:

```
head = list(data)
head
df_headings = pd.DataFrame({'HEADINGS':head})#list to dataframe
print(df_headings)
```

```

                                HEADINGS
0                                coordinates
1                                created_at
2                                hashtags
3                                media
4                                urls
5                                favorite_count
6                                id
7                                in_reply_to_screen_name
8                                in_reply_to_status_id
9                                in_reply_to_user_id
10                               lang
11                               place
12                               possibly_sensitive
13                               retweet_count
14                               reweet_id
15                               retweet_screen_name
16                               source
17                               text
18                               tweet_url
19                               user_created_at
20                               user_screen_name
21 user_default_profile_image
22                               user_description
23                               user_favourites_count
24                               user_followers_count
25                               user_friends_count
26                               user_listed_count
27                               user_location
28                               user_name
29                               user_screen_name.1
30                               user_statuses_count
31                               user_time_zone
32                               user_urls
33                               user_verified
34                               rtt

```

In [249]:

```

#filtering out the non-english tweets
data2 = data[data['lang'].str.contains('en')==True]
data2.shape

```

Out[249]:

```
(1883121, 35)
```

In [250]:

```

data2.columns = data2.columns.str.replace('user_screen_name.1','Screenname') #
Renaming two of the columns
data2.columns = data2.columns.str.replace('created_at','Date_Time')
data2.head(3)

```

Out[250]:

	coordinates	Date_Time	hashtags	media	urls	favorite_count	
0	NaN	Mon May 21 17:16:37 +0000 2018	Together4Yes	NaN	NaN	7	998613546119
1	NaN	Mon May 21 17:18:05 +0000 2018	NaN	NaN	NaN	0	998613914956
2	NaN	Mon May 21 17:16:13 +0000 2018	repeal loveboth	NaN	NaN	0	998613444529

In [251]:

```
data2['date1'] = data2['Date_Time'].apply(lambda x: datetime.strptime(x, '%a %b %d %H:%M:%S +0000 %Y')) #reformat time/date  
#Split date and time into 2 columns  
data2['DATE'] = data2['date1'].dt.date  
data2['TIME'] = data2['date1'].dt.time  
data2.shape
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:3: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:4: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

Out[251]:

(1883121, 38)

In [252]:

```
data_sorted = data2.sort_values(by=['DATE','TIME']) #Organise the data chronologically  
#data_sorted
```

In [253]:

```
data_sorted.shape
```

Out[253]:

(1883121, 38)

In [254]:

```
d_sorted_users = data_sorted.drop_duplicates(subset='Screenname')
d_sorted_users.shape
```

Out[254]:

```
(381662, 38)
```

Spam Classifier

In [255]:

```
data_sorted.shape
```

Out[255]:

```
(1883121, 38)
```

In [256]:

```
#Create a column of the following to follow ratio
data_sorted['friend_foll_ratio'] = (data_sorted.user_followers_count / data_sorted.user_friends_count)
data_sorted.shape
```

Out[256]:

```
(1883121, 39)
```

In [257]:

```
#the number of users in the full dataset
user_data_names = data_sorted.drop_duplicates(subset='Screenname')
user_data_names.shape
```

Out[257]:

```
(381662, 39)
```

In [258]:

```
#THE SPAM CLASSIFIER - create a bool column
data_sorted['is_spam'] = data_sorted['friend_foll_ratio'] < 0.01
data_sorted.shape
```

Out[258]:

```
(1883121, 40)
```


In [259]:

```
#THRESHOLDS  
#number of tweets classified as spam:  
#0.001 - 1199 (for every 1 person I follow 0.01 people follow me back: 5000:5)  
##0.01 - 1357 (for every 1 person I follow 0.01 people follow me back: 5000:50  
)  
##0.02 - 1991 (for every 1 person I follow 0.02 people follow me back: 2500:50  
)  
##0.10 - 3380  
##0.5 - 608912
```

In [260]:

```
#HOW MANY TWEETS ARE SPAM?  
data_sorted['is_spam'].value_counts()
```

Out[260]:

```
False    1881764  
True       1357  
Name: is_spam, dtype: int64
```

In [261]:

```
data_sorted_users = data_sorted.drop_duplicates(subset='Screenname')  
data_sorted_users['is_spam'].value_counts() #HOW MANY USERS ARE SPAM
```

Out[261]:

```
False    380903  
True        759  
Name: is_spam, dtype: int64
```

In [262]:

```
USERDATA_spam = data_sorted[data_sorted['is_spam']] #CHECKING  
USERDATA_spam.shape
```

Out[262]:

```
(1357, 40)
```

In [263]:

```
#DELETE THE SPAM ACCOUNTS BY SELECTING ALL FALSE VALUES FROM IS SPAM COLUMN
```

In [264]:

```
data_clean = data_sorted[data_sorted['is_spam'] == False]
```

In [265]:

```
data_clean.shape
```

Out[265]:

```
(1881764, 40)
```

In [266]:

```
data_clean['is_spam'].value_counts()
```

Out[266]:

```
False      1881764
```

```
Name: is_spam, dtype: int64
```

Filtering by Location and Tweets with Hashtags

In [267]:

```
data_clean.isnull().sum() #any empty variables?
```

Out[267]:

```
coordinates      1881554
Date_Time        0
hashtags         845926
media            1684070
urls             1595296
favorite_count   0
id               0
in_reply_to_screen_name 1809581
in_reply_to_status_id 1818759
in_reply_to_user_id 1809577
lang             0
place            1850360
possibly_sensitive 1410335
retweet_count     0
reweet_id        399085
retweet_screen_name 399085
source           0
text             0
tweet_url        0
user_Date_Time   0
user_screen_name 0
user_default_profile_image 0
user_description 268610
user_favourites_count 0
user_followers_count 0
user_friends_count 0
user_listed_count 0
user_location    454881
user_name        216
Screenname       0
user_statuses_count 0
user_time_zone   1881764
user_urls        1274204
user_verified    0
rtt              0
date1            0
DATE             0
TIME             0
friend_foll_ratio 351
is_spam          0
dtype: int64
```

In [268]:

```
data_clean.head(5)
```

Out[268]:

	coordinates	Date_Time	hashtags	media	urls	favorite_count
		Fri Apr 13 14:16:35				

1457159	NaN	+0000 2018	NaN	NaN	NaN	0
1457143	NaN	Fri Apr 13 14:16:49 +0000 2018	NaN	NaN	NaN	0
1457121	NaN	Fri Apr 13 14:16:55 +0000 2018	NaN	NaN	NaN	0
1457174	NaN	Fri Apr 13 14:16:57 +0000 2018	RepealThe8th	NaN	NaN	0
1457164	NaN	Fri Apr 13 14:16:59 +0000 2018	TrustWomenVoteYes RepealThe8th	NaN	NaN	3

In [269]:

```
data_clean['hash'] = data_clean['text'].str.contains('#')
data_clean['hash'].value_counts()
```

```
/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packag
es/ipykernel_launcher.py:1: SettingWithCopyWarning:
```

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

Out[269]:

```
True      1083227
False      798537
Name: hash, dtype: int64
```

In [270]:

```
data_clean = data_clean[data_clean['hash']] #only include the tweets with hash tags in them
```

In [271]:

```
data_clean.shape
```

Out[271]:

```
(1083227, 41)
```

In [272]:

```
#tweets = data_clean[data_clean.hashtags.notnull()] #only consider data with hashtags
tweets = data_clean
tweets.shape
```

Out[272]:

```
(1083227, 41)
```

In [273]:

```
#location_filtering
df_Ireland=tweets[tweets['user_location'].str.contains('Ireland|Munster|Connacht|Ulster|Antrim|Armagh|Carlow|Cavan|Clare|Cork|Derry|Donegal|Down|Dublin|Fermanagh|Galway|Kerry|Kildare|Kilkenny|Laois|Leitrim|Limerick|Longford|Louth|Mayo|Meath|Monaghan|Offaly|Roscommon|Sligo|Tipperary|Tyrone|Waterford|Westmeath|Wexford|Wicklow',case=False)==True]
print(df_Ireland.shape)
```

```
(433653, 41)
```

In [274]:

```
df_IRE_users = df_Ireland.drop_duplicates(subset='Screenname')
df_IRE_users.shape
```

Out[274]:

```
(42326, 41)
```

In [275]:

```
tweetsuser = tweets.drop_duplicates(subset='Screenname') #how many users in the dataset?
tweetsuser.shape
```

Out[275]:

```
(259379, 41)
```

RT & URL Boolean

In [276]:

```
pd.set_option('display.max_colwidth', -1)
#tweets
```

In [277]:

```
tweets['contains_url'] = tweets['urls'].notnull()
#tweets
tweets['contains_url'].value_counts()
```

Out[277]:

```
False      898024
True       185203
Name: contains_url, dtype: int64
```

In [278]:

```
tweets['hashtags'] = tweets.hashtags.str.replace(' ', ' ', ') #Remove the @ mark
#tweets
```

In [279]:

```
tweets['is_retweet'] = tweets['retweet_count']>0
```

In [280]:

```
tweets['retweet_count'] = tweets['retweet_count'].astype(float)
#tweets.dtypes
```

In [281]:

```
tweets['Logged_RT_count'] = np.log(tweets.retweet_count) #take the nautral log of the retweet count
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:1: RuntimeWarning:

divide by zero encountered in log

News Accounts

In [282]:

```
pd.set_option('display.max_columns', None)  
tweets['user_verified'].value_counts()
```

Out[282]:

```
False    1051612  
True      31615  
Name: user_verified, dtype: int64
```

In [283]:

```
V = tweets[tweets['user_verified']]  
NAMES_V = V.drop_duplicates(subset='user_screen_name', keep='first')  
NAMES_V.shape #how many verified users are there
```

Out[283]:

```
(4756, 44)
```

In [284]:

```
V_SORTED_FOLL = NAMES_V.sort_values(by=['user_followers_count'], ascending = F  
alse)  
V_SORTED_FOLL  
VB = V_SORTED_FOLL[V_SORTED_FOLL['user_description'].str.contains('news|broadcast|newspapers',case=False)==True]  
VB.shape  
#GOT RID OF 'WRITER'- took away 500 values  
#GOT RID OF 'journalist/journalism' - took away 350 accounts
```

Out[284]:

```
(724, 44)
```

In [285]:

```
tweets['desc'] = tweets['user_description'].str.contains('news|broadcast|newspapers',case=False)==True
```

In [286]:

```
tweets['screen'] = tweets['user_screen_name'].str.match('AfloatMagazine|NSAI_S  
tandards|MidlandTrib|thecontel|TippFM|TV3Ireland|ESAI_Environ|Longford_Leader|  
radiomidwest|MissingMigrants|IrishTimes|offalyindo|Metro_Eireann|ClareFM|CTrib  
une|shannonsidenews|ShannonsideFM|BHRRC|KKPeopleNews|tippstar|Corks96FM|highla  
ndradio|OceanFmIreland|Independent_ie|IrishMirror|LMFMRADIO|meathchronicle|iri  
shexaminer|DundalkDemocrat|iamamigrant|TodayFM|AgrilandIreland|TheNationalist|  
LeitrimToday|IrishSunOnline|theirishworld|siliconrepublic|themayonews|HeraldNe  
wsdesk|Herald_ie|IOMIreland|TheEchoOnline|westmeathindo|dgldemocrat|FineGael|I  
rishEcho|teagasc|EUvsDisinfo|NEU_wind|FarmersGuardian|theangocelt|Limerick_Lea  
der|leinsleadernews|derryjournal|CorkEveningEcho|rte|Independent_ie|WaterfordN  
S|kclr96fm|breakingnewsie|LimLive95fm|thejournal_ie|NewstalkFM|Offaly_Express|  
theNSMonaghan|WesternPeople|LaoisNews|clarechampion|dublinpeople|CWnationalist  
|worldcrunch|beat102103|SouthernStarIRL')==True  
#boolean specifying whether something is a list from EMM
```

In [287]:

```
tweets['news_g'] = (tweets['desc']) & (tweets['user_verified'])  
#the news sources that are verified and contains key words in their descriptio  
n
```

In [288]:

```
tweets2 = tweets.drop_duplicates(subset='user_screen_name')  
tweets2['news_g'].value_counts()
```

Out[288]:

```
False    258655  
True       724  
Name: news_g, dtype: int64
```

In [289]:

```
tweets2['screen'].value_counts()
```

Out[289]:

```
False    259324  
True       55  
Name: screen, dtype: int64
```

In [290]:

```
tweets['news'] = (tweets['news_g']) | (tweets['screen'])  
#create the final news boolean; verified + desc filtering OR EMM list
```


In [291]:

```
tweets['news'].value_counts()
```

Out[291]:

```
False    1077044
True      6183
Name: news, dtype: int64
```

In [292]:

```
tweets2 = tweets.drop_duplicates(subset='user_screen_name')
#Drop duplicates to see how many news screennames there are
```

In [293]:

```
tweets2['news'].value_counts()
```

Out[293]:

```
False    258614
True      765
Name: news, dtype: int64
```

In [294]:

```
VB['user_screen_name']
```

Out[294]:

```
412978    nytimes
173806    BBCWorld
1404065    TheEconomist
560221    XHNews
175221    BBCNews
504272    AJEnglish
145842    SkyNews
627718    Newsweek
923146    EconomicTimes
1628934    Independent
441885    Telegraph
228697    itvnews
349070    nytimesworld
842837    bsindia
766662    abcnews
1900288    benshapiro
1046933    HuffPostPol
279102    Refinery29
1014311    welt
275579    BuzzFeedNews
493800    NBCNewsHealth
615172    Salon
191192    mailandguardian
569929    voxdotcom
1128895    BostonGlobe
1560492    rtenews
607279    smitaprakash
967719    NYDailyNews
```

```

967710      NYDailyNews
316634      democracynow
198287      UKLabour

...
1312501     ExtraIRL
255658      KirstyHickey
909716      andreleslie
170738      TheaHalpin
2014091     JessRapana
174047      joshkelliott
380214      Janine__Graham
367027      JSinnerton
446253      Sinclair_H
845023      raji47ram
1597754     karadaniellefox
777344      luke__cooper
878467      RJKDuggan
531891      lovelylnlovie
433178      CatherineNQE
1903555     molcranenewman
919694      AnnieeKane
806531      tobybakare
520922      DaisyMcCorggray
189906      joejsmiddleton
460330      mglynn720
1078101     thisisrachels
515433      JessCiara
835266      AngelaThomps0n
273300      OliviaNiland
337395      MaijaKappler
263444      ArielScotti
1699585     maccytothedee
1051636     FionaDillon
468054      Legeiser
Name: user_screen_name, Length: 724, dtype: object

```

Identifying News Tweets

In [295]:

```
#Create a new column and ultimately a new data set of the tweets continaing news content
```

In [296]:

```
tweets['RT'] = tweets['retweet_screen_name'].str.contains('AfloatMagazine|NSAI
_Standards|MidlandTrib|thecontel|TippFM|TV3Ireland|ESAI_Environ|Longford_Leade
r|radiomidwest|MissingMigrants|IrishTimes|offalyindo|Metro_Eireann|ClareFM|CTr
ibune|shannonsidenews|ShannonsideFM|BHRRC|KKPeopleNews|tippstar|Corks96FM|high
landradio|OceanFmIreland|Independent_ie|IrishMirror|LMFMRADIO|meathchronicle|i
rishexaminer|DundalkDemocrat|iamamigrant|TodayFM|AgrilandIreland|TheNationalis
t|LeitrimToday|IrishSunOnline|theirishworld|siliconrepublic|themayonews|Herald
Newsdesk|Herald_ie|IOMIreland|TheEchoOnline|westmeathindo|dgldemocrat|FineGael
|IrishEcho|teagasc|EUvsDisinfo|NEU_wind|FarmersGuardian|theangocelt|Limerick_L
eader|leinsleadernews|derryjournal|CorkEveningEcho|rte|Independent_ie|Waterfor
dNS|kclr96fm|breakingnewsie|LimLive95fm|thejournal_ie|NewstalkFM|Offaly_Expres
```

st|theNSMonaghan|WesternPeople|LaoisNews|clarechampion|dublinpeople|CWnationali
st|worldcrunch|beat102103|SouthernStarIRL|nytimes|BBCWorld|TheEconomist|XHNews
|BBCNews|AJEnglish|SkyNews|Newsweek|EconomicTimes|Independent|Telegraph|itvnew
s|nytimesworld|bsindia|abcnews|benshapiro|HuffPostPol|Refinery29|welt|BuzzFeed
News|NBCNewsHealth|Salon|mailandguardian|voxdotcom|BostonGlobe|rtenews|smitapr
akash|NYDailyNews|democracynow|UKLabour|CNBCTV18News|CNBCTV18Live|hugoclement|
standardnews|thejournal_ie|abhihitmajumder|krishgm|theprojecttv|Independent_ie
|FinancialXpress|macleans|STVNews|KayBurley|HuffPostCanada|TheArtNewspaper|ETC
anada|DailyCaller|qz|cnalive|marieclaireuk|euronews|globalnews|MetroUK|birming
ham_live|FOX10Phoenix|NBCNewsWorld|faisalislam|cgtnamerica|davidmcw|TelegraphW
orld|BBCNewsNI|MumbaiMirror|SputnikInt|NEWS1130|MishalHusainBBC|jheil|SaimaMoh
sin|NewstalkFM|StephenNolan|MaajidNawaz|dwnews|GuardianUS|skysarahjane|France2
4_en|channeltennews|emmafreud|GuardianAus|MaverickSabre|CatholicNewsSvc|Kevin_
Maguire|globalnewsto|BBCNewsbeat|ABCthedrum|morningireland|fannicanelles|arabn
ews|TheRebelTV|MiriamOCal|thecableng|paulwaugh|Sandra_Sully|PA|RTERadio1|cnsne
ws|maitlis|TV3Ireland|KatyKatopodis|helenlewis|samanthamaiden|CKNW|GettyImages
News|Fabulousmag|EveningStandard|Cambslive|eoghanmcdermo|sundaybusiness|garyyo
unge|marklittlenews|CorksRedFM|BreakfastNT|SocietyGuardian|7NewsAdelaide|irish
_news|MirrorPolitics|csmonitor|RTENewsNow|ScotNational|SecondCaptains|SeanMonc
rieff|HumanHeadline|AndrewSparrow|AngusRobertson|LeftFootFwd|siliconrepublic|A
lbertoNardelli|stephenkb|FemiOke|abcperth|Lkhagva|grainne_seoige|MiriamElde|va
nbadham|rickoshea|wendy_harmer|drivetimerte|calamur|Tuko_co_ke|gavreilly|valmc
dermid|PoliticoRyan|bairdjudia|KNX1070|davidmackau|owen_g|MikeStuchbery_|Willi
amsJon|BethRigby|broderick|headlinepg|GargiRawat|breakingnewsie|QMUL|yahoolife
style|annabosch|ShelaghFogarty|Irish_TechNews|dabeard|misterhsk|CaitrionaPerry
|PatKennyNT|RidgeOnSunday|maryashakil|NorahCasey|3NewsIreland|Limerick_Leader|
legsidelizzy|HayesBrown|BuzzFeedUKPol|arthur_affect|IrishCentral|IrishMirror|i
mage_magazine|craigwilde|pritheworld|KateWilliamsme|RandaHabib|BabitaBBC|amnes
typress|jonoread|davidpreecel2|talkRADIO|fiannafailparty|merrionstreet|rachsha
bi|SkyNewsEditor|WomensAgenda|JenSiebelNewsom|SteveCannane|jessbrammar|NewsFro
mAmnesty|lucysiegle|PINsykes|ciarakellydoc|metronewyork|NewshubNZ|An_Phoblacht
|MarshallCohen|LaureJames|cmarinucci|sabizak|aljwhite|tweeter_anita|PeninsulaQ
atar|jonathanhealy|SarahAMcInerney|liamyoun|BBCJohnBeattie|MolloyJoe|IAMMGrah
am|SiobhanFenton|AlanFisher|CerianJenkins|SJAMcBride|aishagani|YukariWatanabe|
TodayFMNews|Simon_Cullen|minihammary|bprerna|mikenesbittni|bindelj|tcarmody|Ch
arleneWhite|karengier|johnsweeneyroar|MartinaFitzg|noreenahertz|derryjournal|
louisemcsharry|WatchCTVNews|MarisaKabas|NiallSF|HerFamilydotie|BrendanCourtney
|haynesdeborah|sarayasin|malonebarry|skynewssarah|TomNamako|MattMcGrathBBC|Plu
toPress|AlisonTodayFM|AM2DM|KeelinShanley|am640|sarahditum|AllisonMorris1|stee
letalk|ellievhall|samueloakford|DMcCaffreySKY|newschambers|jljacobson|Walldo|f
ranmcnulty|SamanthaLibreri|kclr96fm|janemartinson|NadineORegan|asteris|AaronMe
hta|Howard_Goodall|NehaPoonia|indyfromspace|garethoconnor|dinabass|PaulBrandIT
V|stuartmillar159|katiejmbaker|EVOKE|RhiannaDhillon|Dr_Ellie|ianguider|stephhe
garty|psmith|paigelav|gnel|skydavidblevins|JillKrop|StephLauren|latestly|doyle
_kevin|abchobart|alicefeiring|LondonLabour|AntoniaZ|jesskellynt|MarryERegan|amb
iej|PippaCrerar|BBCNuala|drlfarrell|dw_europe|EoinBearla|juliareinstein|Sally_
Jackson|lewis_goodall|luciemorrismarr|AlixFox|NishSwish|IrishTimesPol|LouiseRi
dley|WikiTribune|RositaBoland|profkeithdevlin|philipbromwell|TJ_Politics|sswin
kgma|MarksLarks|davidmacdougall|KevDoyle_Indo|TheMaggieDoyle|VeraMBergen|Harri
etMinter|Creative_Dundee|HelenKennedy|macdac|thomas_mohr|emilylhauser|elenacre
sci|MSpicuzzaMJS|IrishSunOnline|JulietteGash|BrynLucas|Orlaodo|ameliargh|sharo
ntobin|action|meganspecia|MikeSegalov|byJenAMiller|PaschalSheehy|SineadOCarrol
l|ColetteFitzIRL|Ankaman616|HenryMcKean|tdwilliamsrome|Juanita_Phillip|LiamHal
ligan|AdamParsons1|AMTomchak|omar_aok|evan_greer|PAImages|RebeccaRideal|La__Cu
en|siobhanheanue|rowhoop|NiamhHoran|maxinemawhinney|sayoniaiyar|universitytim
es|roridonaghy|KirstyStricklan|uscsocialwork|rosieswash|BBCKasiaMadera|jtes|An
neClaireCNN|tara_mulholland|ginarush|willyleeadams|KFYRTV|sweetoblivion26|Les

Glorieuses|CathyReisenwitz|aislingmkeenan|yukokato1701|pennymitv|CBedfordDC|ju
liaccarpenter|FredianiITV|sineadspain|k_yoder|NikiBlasina|MaureenShaw|matthewc
hampion|Fi_Rutherford|thetimesIE|sesmith|SportforBusines|FergalOBrienTV3|colet
te_sexton|wearerabble|VinnyMcAv|GlasgowGuardian|mollymhunter|victoriasanusi|mc
bride_caitlin|TonyTassell|laurafleur|christineestima|anguswalkertalk|earleyedi
tion|orla_barry|carlfranzen|Anoosh_C|matbreen|ImtiazTyab|alastairjam|mrjoshz|K
GNSnews|DCampoamor|chrisapplegate|AmyKazmin|vickyyyf|tomphillipsEG|c_crampton|
SuzannePhan|versharma|ajreid|eleanormcevoy|kassyapple|horton_official|AilbheCo
nneely|jonathansamuels|rachel_handler|EmmaVardyTV|reider|liberalinternat|sd_sa
urav|ODonnellDale|yodaberg|stephencomedy|suskeogh|robdougan|camillard|IvanYate
sNT|RachaelKrishna|MarkHennesy|mofitzmaurice|JoanaRamiroUK|fpjindia|AmritaChe
emaTV|terischultz|GraemeDemiany|davidlevin123|shamindernahal|jessicawheeler3|D
eannaO_Connor|acatherwoodnews|paddylogue|ClodaghLRice|DionneGrant|KarenDSouza4
|richjamesuk|kiliandoyle|CurtisHouck|cristinleach|RayKennedyHack|maxhofmann|ka
ylawebley|The_Newsmakers|JuliaRampenMM|bspeed8|JoshuaHoyos|EmilyHeward|CAugust
enborg|aspeciale|brentgofftv|byrnemaria|lucymcdonald|GuthrieGF|JessReports|ale
xbrucesmith|kirstiemccrum|JamesERothwell|JoeLeogue|yessibelloperetz|JackieLeona
rd0|colincoyle|Sachk0|GilloArvos|JoanScalesIT|JoelTaylorhack|daithaigilbert|Re
miAdekoyal|katielannan|YvetteCaster|harriepw|jc_cummins|tomasclancy|matthewhaa
g|KeithWalkerNew|deefinnerty|SandyRashty|LibbyWienerITV|andreagilligan|Zoah_HS
|amandacconn|emmareyn|Newschannel6No|_esaliba|joelmgunter|AnnaCollinson|stanch
ers|JohnODowdIRL|Nicola_Slawson|msmegwatson|PaulineBock|AnnaCafolla|SallyBiddu
lph|dellakilroy|kierancuddihy|SeanPrevil|GenevieveCarber|JessicaMudditt|CathyK
oo|asvokos|TheProject_NZ|kamalasripada|RachLoxton|ziadramley|kashmiragander|Da
vid_Batty|JaydeLovell|LakshmiGandhi|rachelblundy|ruddick|nevali|richardaedy|ev
elynwoodsen|metamythos|SineadStubbins|yvonneredmond|SteBreen|AnnaVDesigns|abi
gailrieley|Blackwell|AlineBarros2|RobertCorp|marnieoneill7|salimah_shivji|powe
lldennis|fiona_day|elizabeth_shell|dlabanyi|HattyCollier|alicia_ms|alex_zee|te
ssa5news|SianElvin|jack_sommers|alisonsandy|conor_feehan|kathryn42|callumlesli
e92|BobBrigham|CJAMcMahon|rooITV|IrishEmbMalta|suematthias|BryonyEHopkins|Kath
riller|heatheranne9|IanJSherwood|freyadro|BernardOToole|ingredient_x|jrud|thec
inemile|LauraHoganTV3|matildbattersby|GingerOtis|DuncanHooper|RCarrollTV3|Dian
aRusk|jimmymacRTE|jameshendicott|tendailyau|alisonnulty|johndeguzman|PJ_Kirkha
m|colmsmyth|ColleenLHarris|buileshuibhne|kat_b_hawkins|jane_last|samcrogers|ed
jsandova|jqbilbao|kirancmoodley|basialcummings|Leila_MA|mhdelaney|a_zabjek|Ann
eBranigin|Mr_LukeBenson|Raftastico|denise_denise|Leanna_Byrne|sanyaburgess|Kie
ranG77|Tapeterman|Ben_oharabyrne|bradleysecker|DaraghBroph|vcaivano|DwyerCraig
|HRW_Brussels|ocrowcroft|anibundel|LucieMcInerney|ryan_padraic|mikewsc1|Inessa
Nova|conorhallahan|sentascar|lousetooney|gallagherfergal|rkennedy228|Jamesaa|c
iarariordan|KevinCGriffin|CatharinaMoh|katieforster|morgtait|CJMortimer|Suchan
drikaC|HelenRumbelow|LouisAlexDore|KiiimBuckley|MichaelRyanTV3|maithreyi_s|Nia
mhTowey1|meghas|laloeffelstiel|DebsNaylor|shannonjpower|davemolloy|RichALWilli
ams|nivie|seanmstephens|timhatfield87|IrlEmbDenmark|jofahy|JustinRohrlich|Jaso
nLeMiere|ByClare|SiobhanMorris|linnytayls|jupadhye|carrollste|ChristinaMcS|Chr
isCreegan|ExSaminer|NicoleHDeSmet|HendersStu|MikaelaConley|alesbon|conormcmaho
n|HolstaT|RDKoncienzy|RobMcD|c00perM|jvanevra|lucysween|laurabundock|DevikaBh
at|holly_monery|kcassowary|EwanPalmer|anniewhiteabc|IrlEmbPrague|AlysonHenry_
KathrynRiddell|SincerelyAde|TheGregMurphy|ambhack|mahamabedi|tayknopf|alliehum
phries|ExtraIRL|KirstyHickey|andreleslie|TheaHalpin|JessRapana|joshkelliott|Ja
nine__Graham|JSinnerton|raji47ram|Sinclair_H|karadaniellefox|luke__cooper|RJKD
uggan|lovelynlovie|CatherineNQE|molcranenewman|tobybakare|AnnieeKane|DaisyMcCo
rgray|joejsmiddleton|mglynn720|thisisrachels|JessCiara|OliviaNiland|MaijaKappl
er|ArielScotti|maccytothede|FionaDillon|Legeiser')==True
tweets['RT'].value_counts() #variable displaying retweets of news accounts

Out[296]:

```
False      1039329
True       43898
Name: RT, dtype: int64
```

In [297]:

```
tweets['text_news'] = tweets['text'].str.contains('AfloatMagazine|NSAI_Standards|MidlandTrib|thecontel|TippFM|TV3Ireland|ESAI_Environ|Longford_Leader|radiomidwest|MissingMigrants|IrishTimes|offalyindo|Metro_Eireann|ClareFM|CTribune|shannonsidenews|ShannonsideFM|BHRRC|KKPeopleNews|tippstar|Corks96FM|highlandradio|OceanFmIreland|Independent_ie|IrishMirror|LMFMRADIO|meathchronicle|irisht Examiner|DundalkDemocrat|iamamigrant|TodayFM|AgrilandIreland|TheNationalist|LeitrimToday|IrishSunOnline|theishworld|siliconrepublic|themayonews|HeraldNewsdesk|Herald_ie|IOMIreland|TheEchoOnline|westmeathindo|dgldemocrat|FineGael|IrishEcho|teagasc|EUvsDisinfo|NEU_wind|FarmersGuardian|theangocelt|Limerick_Leader|leinsleadernews|derryjournal|CorkEveningEcho|rte|Independent_ie|WaterfordNS|kclr96fm|breakingnewsie|LimLive95fm|thejournal_ie|NewstalkFM|Offaly_Express|theNSMonaghan|WesternPeople|LaoisNews|clarechampion|dublinpeople|CWnationalist|worldcrunch|beat102103|SouthernStarIRL|nytimes|BBCWorld|TheEconomist|XHNews|BBCNews|AJEnglish|SkyNews|Newsweek|EconomicTimes|Independent|Telegraph|itvnews|nytimesworld|bsindia|abcnews|benshapiro|HuffPostPol|Refinery29|welt|BuzzFeedNews|NBCNewsHealth|Salon|mailandguardian|voxdotcom|BostonGlobe|rtenews|smitaprakash|NYDailyNews|democracynow|UKLabour|CNBCTV18News|CNBCTV18Live|hugoclement|standardnews|thejournal_ie|abhijitmajumder|krishgm|theprojecttv|Independent_ie|FinancialXpress|macleans|STVNews|KayBurley|HuffPostCanada|TheArtNewspaper|ETCanada|DailyCaller|qz|cnalive|marieclaireuk|euronews|globalnews|MetroUK|birmingham_live|FOX10Phoenix|NBCNewsWorld|faisalislam|cgtnamerica|davidmcw|TelegraphWorld|BBCNewsNI|MumbaiMirror|SputnikInt|NEWS1130|MishalHusainBBC|jheil|SaimaMohsin|NewstalkFM|StephenNolan|MaajidNawaz|dwnews|GuardianUS|skysarahjane|France24_en|channeltennews|emmafreud|GuardianAus|MaverickSabre|CatholicNewsSvc|Kevin_Maguire|globalnewsto|BBCNewsbeat|ABCthedrum|moringireland|fannicanelles|arabnews|TheRebelTV|MiriamOCal|thecableng|paulwaugh|Sandra_Sully|PA|RTERadio1|cnsnews|maitlis|TV3Ireland|KatyKatopodis|helenlewis|samanthamaiden|CKNW|GettyImagesNews|Fabulousmag|EveningStandard|Cambslive|eoghanmcdermo|sundaybusiness|garyyounge|marklittlenews|CorksRedFM|BreakfastNT|SocietyGuardian|7NewsAdelaide|irish_news|MirrorPolitics|csmonitor|RTENewsNow|ScotNational|SecondCaptains|SeanMoncrieff|HumanHeadline|AndrewSparrow|AngusRobertson|LeftFootFwd|siliconrepublic|AlbertoNardelli|stephenkb|FemiOke|abcperth|Lkhagva|grainne_seoige|MiriamElde|vanbadham|rickoshea|wendy_harmer|drivetimerte|calamur|Tuko_co_ke|gavreilly|valmcdermid|PoliticoRyan|bairdjudia|KNX1070|davidmackau|owen_g|MikeStuchbery|WilliamsJon|BethRigby|broderick|headlinepg|GargiRawat|breakingnewsie|QMUL|yahoolifestyle|anabosch|ShelaghFogarty|Irish_TechNews|dabeard|misterhsk|CaitrionaPerry|PatKennedyNT|RidgeOnSunday|maryashakil|NorahCasey|3NewsIreland|Limerick_Leader|legsidelizzy|HayesBrown|BuzzFeedUKPol|arthur_affect|IrishCentral|IrishMirror|image_magazine|craigwilde|pritheworld|KateWilliamsme|RandaHabib|BabitaBBC|amnestypress|jonoread|davidpreecel2|talkRADIO|fiannafailparty|merrionstreet|rachshabi|SkyNewsEditor|WomensAgenda|JenSiebelNewsom|SteveCannane|jessbrammar|NewsFromAmnesty|lucysiegle|PINsykes|ciarakellydoc|metronewyork|NewshubNZ|An_Phoblacht|MarshallCohen|LaureJames|cmarinucci|sabizak|aljwhite|tweeter_anita|PeninsulaQatar|jonathanhealy|SarahAMcInerney|liamyoun|BBCJohnBeattie|MolloyJoe|IAMMGraham|SiobhanFenton|AlanFisher|CerianJenkins|SJAMcBride|aishagani|YukariWatanabe|TodayFMNews|Simon_Cullen|minihanmary|bprerna|mikenesbittni|bindelj|tcarmody|CharleneWhite|karengerier|johnsweeneyroar|MartinaFitzg|noreenahertz|derryjournal|louisemcsharry|WatchCTVNews|MarisaKabas|NiallSF|HerFamilydotie|BrendanCourtney|haynesdeborah|sarayasin|malonebarry|skynewssarah|TomNamako|MattMcGrathBBC|PlutoPress|
```

AllisonTodayFM|AM2DM|KeelinShanley|am640|sarahdittum|AllisonMorris1|steelcalk|ellievhall|samueloakford|DMcCaffreySKY|newschambers|jlljacobson|Walldo|franmcnulty|SamanthaLibreri|kclr96fm|janemartinson|NadineORegan|asteris|AaronMehta|Howard_Goodall|NehaPoonia|indyfromspace|garethoconnor|dinabass|PaulBrandITV|stuartmillar159|katiejmbaker|EVOKE|RhiannaDhillon|Dr_Ellie|ianguider|stephhegarty|psmith|paigelav|gnel|skydavidblevins|JillKrop|StephLauren|latestly|doyle_kevin|abchobart|alicefeiring|LondonLabour|AntoniaZ|jesskellynt|MaryERegan|ambiej|PippaCrerar|BBCNuala|drlfarrell|dw_europe|EoinBearla|juliareinstein|Sally_Jackson|lewis_goodall|luciemorrismarr|AlixFox|NishSwish|IrishTimesPol|LouiseRidley|WikiTribune|RositaBoland|profkeithdevlin|philipbromwell|TJ_Politics|sswinkgma|MarksLarks|davidmacdougall|KevDoyle_Indo|TheMaggieDoyle|VeraMBergen|HarrietMinter|Creative_Dundee|HelenKennedy|macdac|thomas_mohr|emilylhauser|elenacresci|MSpicuzzaMJS|IrishSunOnline|JulietteGash|BrynLucas|Orlaodo|ameliargh|sharontobin|action|meganspecia|MikeSegalov|byJenAMiller|PaschalSheehy|SineadOCarroll|ColetteFitzIRL|Ankaman616|HenryMcKean|tdwilliamsrome|Juanita_Phillip|LiamHalligan|AdamParsons1|AMTomchak|omar_aok|evan_greer|PAImages|RebeccaRideal|La__Cuen|siobhanheanue|rowhoop|NiamhHoran|maxinemawhinney|sayoniaiyar|universitytimes|roridonaghy|KirstyStricklan|uscsocialwork|rosieswash|BBCAsiaMadera|jtes|AnneClaireCNN|tara_mulholland|ginarush|willyleeadams|KFYRTV|sweetoblivion26|Les_Glorieuses|CathyReisenwitz|aislingmkeen|yukokato1701|pennymitv|CBedfordDC|juliaccarpenter|FredianiITV|sineadspain|k_yoder|NikiBlasina|MaureenShaw|matthewchampion|Fi_Rutherford|thetimesIE|sesmith|SportforBusines|FergalOBrienTV3|colette_sexton|wearerabble|VinnyMcAv|GlasgowGuardian|mollymhunter|victoriasanusi|mcbride_caitlin|TonyTassell|laurafleur|christineestima|anguswalkertalk|earleyedition|orla_barry|carlfranzen|Anoosh_C|matbreen|ImtiazTyab|alastairjam|mrjoshz|KGNSNews|DCampoamor|chrisapplegate|AmyKazmin|vickyyyf|tomphillipsEG|c_crampton|SuzannePhan|versharma|ajreid|eleanormcevoy|kassyapple|horton_official|AilbheConneely|jonathansamuels|rachel_handler|EmmaVardyTV|reider|liberalinternat|sd_saurav|ODonnellDale|yodaberg|stephencomedy|suskeogh|robdougan|camillard|IvanYatesNT|RachaelKrishna|MarkHennessy|mofitzmaurice|JoanaRamiroUK|fpjindia|AmritaCheemaTV|terischultz|GraemeDemian|davidlevin123|shamindernahal|jessicawheeler3|DeannaO_Connor|acatherwoodnews|paddylogue|ClodaghLRice|DionneGrant|KarenDSouza4|richjamasuk|kiliandoyle|CurtisHouck|cristinleach|RayKennedyHack|maxhofmann|kaylawebley|The_Newsmakers|JuliaRampenMM|bspeed8|JoshuaHoyos|EmilyHeward|CAugustenborg|aspeciale|brentgofftv|byrnemaria|lucymcdonald|GuthrieGF|JessReports|alexbrucesmith|kirstiemccrum|JamesERothwell|JoeLeogue|yessibelloperez|JackieLeonard0|colincoyle|Sachk0|GilloArvos|JoanScalesIT|JoelTaylorhack|daithaigilbert|RemiAdekoyal|katielannan|YvetteCaster|harriepw|jc_cummins|tomasclancy|matthewhaag|KeithWalkerNew|deefinnerty|SandyRashty|LibbyWienerITV|andreagilligan|Zoah_HS|amandacconn|emmareyn|Newschannel6No|_esaliba|joelmgunter|AnnaCollinson|stanchers|JohnODowdIRL|Nicola_Slawson|msmegwatson|PaulineBock|AnnaCafolla|SallyBiddulph|dellakilroy|kierancuddihy|SeanPrevil|GenevieveCarber|JessicaMudditt|CathyKoo|asvokos|TheProject_NZ|kamalasripada|RachLoxton|ziadramley|kashmiragander|David_Batty|JaydeLovell|LakshmiGandhi|rachelblundy|ruddick|nevali|richardaedy|evelynwoodsen|metamythos|SineadStubbins|yvonneredmond|SteBreen|AnnaVDesigns|abigailrielley|Blackwell|AlineBarros2|RobertCorp|marnieoneill17|salimah_shivji|powelldennis|fiona_day|elizabeth_shell|dlabanyi|HattyCollier|alicia_ms|alex_zee|tessa5news|SianElvin|jack_sommers|alisonsandy|conor_feehan|kathryn42|callumleslie92|BobBrigham|CJAMcMahon|rooITV|IrishEmbMalta|suematthias|BryonyEHopkins|Kathriller|hetheranne9|IanJSherwood|freyadro|BernardOTOole|ingredient_x|jrud|thecinemile|LauraHoganTV3|matildbattersby|GingerOtis|DuncanHooper|RCarrollTV3|DianaRusk|jimmacRTE|jameshendicott|tendailyau|alisonnulty|johndeguzman|PJ_Kirkham|colmsmyth|ColleenLHarris|buileshuibhne|kat_b_hawkins|jane_last|samcrogers|edjsandova|jqbilbao|kirancmoodley|basialcummings|Leila_MA|mhdelaney|a_zabjek|AnneBranigin|Mr_LukeBenson|Raftastico|denise_denise|Leanna_Byrne|sanyaburgess|KieranG77|Tapeterman|Ben_oharabyrne|bradleysecker|DaraghBroph|vcaivano|DwyerCraig|HRW_Brussels|ocrowcroft|anibundel|LucieMcInerney|ryan_padraic|mikewsc1|InessaNova|conorhallahan|sentascar|lousetooney|gallagherfergal|rkennedy228|Jamesaa|ciararior

```

dan|KevinCGriffin|CatharinaMoh|katieforster|morgtait|CJMortimer|SuchandrikaC|H
elenRumbelow|LouisAlexDore|KiiimBuckley|MichaelRyanTV3|maithreyi_s|NiamhTowey1
|meghas|lalloeffelstiel|DebsNaylor|shannonjpower|davemolloy|RichALWilliams|nivi
e|seanmstephens|timhatfield87|IrlEmbDenmark|jofahy|JustinRohrlich|JasonLeMiere
|ByClare|SiobhanMorrin|linnytayls|jupadhye|carrollste|ChristinaMcS|ChrisCreega
n|ExSaminer|NicoleHDeSmet|HendersStu|MikaelaConley|alesbon|conormcmahon|Holsta
T|RDKoncienzczy|RobMcD|c00perM|jvanevra|lucysween|laurabundock|DevikaBhat|holly
_monery|kcassowary|EwanPalmer|anniewhiteabc|IrlEmbPrague|AlysonHenry_|KathrynR
iddell|SincerelyAde|TheGregMurphy|ambhack|mahamabedi|tayknopf|alliehumphries|E
xtraIRL|KirstyHickey|andreleslie|TheaHalpin|JessRapana|joshkelliott|Janine__Gr
aham|JSinnerton|raji47ram|Sinclair_H|karadaniellefox|luke__cooper|RJKDuggan|lo
velynlovie|CatherineNQE|molcranenewman|tobybakare|AnnieeKane|DaisyMcCorggray|jo
ejsmiddleton|mglynn720|thisisrachels|JessCiara|OliviaNiland|MaijaKappler|Ariel
Scotti|maccytothedee|FionaDillon|Legeiser')==True
tweets['text_news'].value_counts() #tweets that MENTION news acc
#cancelled b/c not looking at mentions

```

Out[297]:

```

False      984129
True       99098
Name: text_news, dtype: int64

```

In [298]:

```

tweets['screen_news'] = tweets['Screenname'].str.contains('AfloatMagazine|NSAI
_Standards|MidlandTrib|thecontel|TippFM|TV3Ireland|ESAI_Environ|Longford_Leade
r|radiomidwest|MissingMigrants|IrishTimes|offalyindo|Metro_Eireann|ClareFM|CTr
ibune|shannonsidenews|ShannonsideFM|BHRRC|KKPeopleNews|tippstar|Corks96FM|high
landradio|OceanFmIreland|Independent_ie|IrishMirror|LMFMRADIO|meathchronicle|i
rishexaminer|DundalkDemocrat|iamamigrant|TodayFM|AgrilandIreland|TheNationalis
t|LeitrimToday|IrishSunOnline|theirishworld|siliconrepublic|themayonews|Herald
Newsdesk|Herald_ie|IOMIreland|TheEchoOnline|westmeathindo|dgldemocrat|FineGael
|IrishEcho|teagasc|EUvsDisinfo|NEU_wind|FarmersGuardian|theangocelt|Limerick_L
eader|leinsleadernews|derryjournal|CorkEveningEcho|rte|Independent_ie|Waterfor
dNS|kclr96fm|breakingnewsie|LimLive95fm|thejournal_ie|NewstalkFM|Offaly_Expres
s|theNSMonaghan|WesternPeople|LaoisNews|clarechampion|dublinpeople|CWnationali
st|worldcrunch|beat102103|SouthernStarIRL|nytimes|BBCWorld|TheEconomist|XHNews
|BBCNews|AJEnglish|SkyNews|Newsweek|EconomicTimes|Independent|Telegraph|itvnew
s|nytimesworld|bsindia|abcnews|benshapiro|HuffPostPol|Refinery29|welt|BuzzFeed
News|NBCNewsHealth|Salon|mailandguardian|voxdotcom|BostonGlobe|rtenews|smitapr
akash|NYDailyNews|democracynow|UKLabour|CNBCTV18News|CNBCTV18Live|hugoclement|
standardnews|thejournal_ie|abhijitmajumder|krishgm|theprojecttv|Independent_ie
|FinancialXpress|macleans|STVNews|KayBurley|HuffPostCanada|TheArtNewspaper|ETC
anada|DailyCaller|qz|cnalive|marieclaireuk|euronews|globalnews|MetroUK|birming
ham_live|FOX10Phoenix|NBCNewsWorld|faisalislam|cgtnamerica|davidmcw|TelegraphW
orld|BBCNewsNI|MumbaiMirror|SputnikInt|NEWS1130|MishalHusainBBC|jheil|SaimaMoh
sin|NewstalkFM|StephenNolan|MaajidNawaz|dwnews|GuardianUS|skysarahjane|France2
4_en|channeltennews|emmafreud|GuardianAus|MaverickSabre|CatholicNewsSvc|Kevin_
Maguire|globalnewsto|BBCNewsbeat|ABCthedrum|morningireland|fannicanelles|arabn
ews|TheRebelTV|MiriamOCal|thecableng|paulwaugh|Sandra_Sully|PA|RTERadio1|cnsne
ws|maitlis|TV3Ireland|KatyKatopodis|helenlewis|samanthamaiden|CKNW|GettyImages
News|Fabulousmag|EveningStandard|Cambslive|eoghanmcdermo|sundaybusiness|garyyo
unge|marklittlenews|CorksRedFM|BreakfastNT|SocietyGuardian|7NewsAdelaide|irish
_news|MirrorPolitics|csmonitor|RTENewsNow|ScotNational|SecondCaptains|SeanMonc
rieff|HumanHeadline|AndrewSparrow|AngusRobertson|LeftFootFwd|siliconrepublic|A
lbertoNardelli|stephenkb|FemiOke|abcperth|Lkhagva|grainne_seoige|MiriamElde|va

```


nbadam|rickoshea|wendy_harmer|drivetimerte|calamur|Tuko_co_ke|gavreilly|valmc
dermid|PoliticoRyan|bairdjudia|KNX1070|davidmackau|owen_g|MikeStuchbery_|Willi
amsJon|BethRigby|broderick|headlinepg|GargiRawat|breakingnewsie|QMUL|yahoolife
style|annabosch|ShelaghFogarty|Irish_TechNews|dabeard|misterhsk|CaitrionaPerry
|PatKennyNT|RidgeOnSunday|maryashakil|NorahCasey|3NewsIreland|Limerick_Leader|
legsidelizzy|HayesBrown|BuzzFeedUKPol|arthur_affect|IrishCentral|IrishMirror|i
mage_magazine|craigwilde|pritheworld|KateWilliamsme|RandaHabib|BabitaBBC|amnes
typress|jonoread|davidpreecel2|talkRADIO|fiannafailparty|merrionstreet|rachsha
bi|SkyNewsEditor|WomensAgenda|JenSiebelNewsom|SteveCannane|jessbrammar|NewsFro
mAmnesty|lucysiegle|PINsykes|ciarakellydoc|metronewyork|NewshubNZ|An_Phoblacht
|MarshallCohen|LaureJames|cmarinucci|sabizak|aljwhite|tweeter_anita|PeninsulaQ
atar|jonathanhealy|SarahAMcInerney|liamyong|BBCJohnBeattie|MolloyJoe|IAMMGrah
am|SiobhanFenton|AlanFisher|CerianJenkins|SJAMcBride|aishagani|YukariWatanabe|
TodayFMNews|Simon_Cullen|minihanmary|bprerna|mikenesbittni|bindelj|tcarmody|Ch
arleneWhite|karengier|johnsweeneyroar|MartinaFitzg|noreenahertz|derryjournal|
louisemcsharry|WatchCTVNews|MarisaKabas|NiallSF|HerFamilydotie|BrendanCourtney
|haynesdeborah|sarayasin|malonebarry|skynewssarah|TomNamako|MattMcGrathBBC|Plu
toPress|AlisonTodayFM|AM2DM|KeelinShanley|am640|sarahditum|AllisonMorris1|stee
letalk|ellievhall|samueloakford|DMcCaffreySKY|newschambers|jljacobson|Walldo|f
ranmcnulty|SamanthaLibreri|kclr96fm|janemartinson|NadineORegan|asteris|AaronMe
hta|Howard_Goodall|NehaPoonia|indyfromspace|garethoconnor|dinabass|PaulBrandIT
V|stuartmillar159|katiejmbaker|EVOKE|RhiannaDhillon|Dr_Ellie|ianguider|stephhe
garty|psmith|paigelav|gnel|skydavidblevins|JillKrop|StephLauren|latestly|doyle
_kevin|abchobart|alicefeiring|LondonLabour|AntoniaZ|jesskellynt|MaryERegan|amb
iej|PippaCrerar|BBCNuala|drlfarrell|dw_europe|EoinBearla|juliareinstein|Sally_
Jackson|lewis_goodall|luciemorrismarr|AlixFox|NishSwish|IrishTimesPol|LouiseRi
dley|WikiTribune|RositaBoland|profkeithdevlin|philipbromwell|TJ_Politics|sswin
kgma|MarksLarks|davidmacdougall|KevDoyle_Indo|TheMaggieDoyle|VeraMBergen|Harri
etMinter|Creative_Dundee|HelenKennedy|macdac|thomas_mohr|emilylhauser|elenacre
sci|MSpicuzzaMJS|IrishSunOnline|JulietteGash|BrynLucas|Orlaodo|ameliargh|sharo
ntobin|action|meganspecia|MikeSegalov|byJenAMiller|PaschalSheehy|SineadOCarrol
l|ColetteFitzIRL|Ankaman616|HenryMcKean|tdwilliamsrome|Juanita_Phillip|LiamHal
ligan|AdamParsons1|AMTomchak|omar_aok|evan_greer|PAImages|RebeccaRideal|La__Cu
en|siobhanheanue|rowhoop|NiamhHoran|maxinemawhinney|sayoniaiyar|universitytim
es|roridonaghy|KirstyStricklan|uscsocialwork|rosieswash|BBCKasiaMadera|jtes|An
neClaireCNN|tara_mulholland|ginarush|willyleeadams|KFYRTV|sweetoblivion26|Les_
Glorieuses|CathyReisenwitz|aislingmkeenan|yukokato1701|pennymitv|CBedfordDC|ju
liaccarpenter|FredianiITV|sineadspain|k_yoder|NikiBlasina|MaureenShaw|matthewc
hampion|Fi_Rutherford|thetimesIE|sesmith|SportforBusines|FergalOBrienTV3|colet
te_sexton|wearerabble|VinnyMcAv|GlasgowGuardian|mollymhunter|victoriasanusi|mc
bride_caitlin|TonyTassell|laurafleur|christineestima|anguswalkertalk|earleyedi
tion|orla_barry|carlfranzén|Anoosh_C|matbreen|ImtiazTyab|alastairjam|mrjoshz|K
GNSnews|DCampoamor|chrisapplegate|AmyKazmin|vickyyyf|tomphillipsEG|c_crampton|
SuzannePhan|versharma|ajreid|eleanormcevoy|kassyapple|horton_official|AilbheCo
nneely|jonathansamuels|rachel_handler|EmmaVardyTV|reider|liberalinternat|sd_sa
urav|ODonnellDale|yodaberg|stephencomedy|suskeogh|robdougan|camillard|IvanYate
sNT|RachaelKrishna|MarkHennessy|mofitzmaurice|JoanaRamiroUK|fpjindia|AmritaChe
emaTV|terischultz|GraemeDemian|davidlevin123|shamindernahal|jessicawheeler3|D
eannaO_Connor|acatherwoodnews|paddylogue|ClodaghLRice|DionneGrant|KarenDSouza4
|richjamesuk|kiliandoyle|CurtisHouck|cristinleach|RayKennedyHack|maxhofmann|ka
ylawebley|The_Newsmakers|JuliaRampenMM|bspeed8|JoshuaHoyos|EmilyHeward|CAugust
enborg|aspeciale|brentgofftv|byrnenmaria|lucymcdonald|GuthrieGF|JessReports|ale
xbrucesmith|kirstiemccrum|JamesERothwell|JoeLeogue|yessibelloperéz|JackieLeona
rd0|colincoyle|Sachk0|GilloArvos|JoanScalesIT|JoelTaylorhack|daithaigilbert|Re
miAdekoya1|katielannan|YvetteCaster|harriepw|jc_cummins|tomasclancy|matthewhaa
g|KeithWalkerNew|deefinnerty|SandyRashty|LibbyWienerITV|andreagilligan|Zoah_HS
|amandacconn|emmareyn|Newschannel6No|esaliba|joelmgunter|AnnaCollinson|stanch


```

ers|JohnODowdIRL|Nicola_Slawson|msmegwatson|PaulineBock|AnnaCafolla|SallyBiddu
lph|dellakilroy|kierancuddihy|SeanPrevil|GenevieveCarber|JessicaMudditt|CathyK
oo|asvokos|TheProject_NZ|kamalasripada|RachLoxton|ziadramley|kashmiragander|Da
vid_Batty|JaydeLovell|LakshmiGandhi|rachelblundy|ruddick|nevali|richardaedy|ev
elynvwoodsen|metamythos|SineadStubbins|yvonneredmond|SteBreen|AnnaVDesigns|abi
gailrieley|Blackwell|AlineBarros2|RobertCorp|marnieoneill17|salimah_shivji|powe
lldennis|fiona_day|elizabeth_shell|dlabanyi|HattyCollier|alicia_ms|alex_zee|te
ssa5news|SianElvin|jack_sommers|alisonsandy|conor_feehan|kathryn42|callumlesli
e92|BobBrigham|CJAMcMahon|rooITV|IrishEmbMalta|suematthias|BryonyEHopkins|Kath
riller|heatheranne9|IanJSherwood|freyadro|BernardOToole|ingredient_x|jrud|thec
inemile|LauraHoganTV3|matildbattersby|GingerOtis|DuncanHooper|RCarrollTV3|Dian
aRusk|jimmymacRTE|jameshendicott|tendailyau|alisonnulty|johndeguzman|PJ_Kirkha
m|colmsmyth|ColleenLHarris|buileshuibhne|kat_b_hawkins|jane_last|samcrogers|ed
jsandova|jqbilbao|kirancmoodley|basialcummings|Leila_MA|mhdelaney|a_zabjek|Ann
eBranigin|Mr_LukeBenson|Raftastico|denise_denise|Leanna_Byrne|sanyaburgess|Kie
ranG77|Tapeterman|Ben_oharabyrne|bradleysecker|DaraghBroph|vcaivano|DwyerCraig
|HRW_Brussels|ocrowcroft|anibundel|LucieMcInerney|ryan_padraic|mikewsc1|Inessa
Nova|conorhallahan|sentascar|lousetooney|gallagherfergal|rkenedy228|Jamesaa|c
iarariordan|KevinCGriffin|CatharinaMoh|katieforster|morgtait|CJMortimer|Suchan
drikaC|HelenRumbelow|LouisAlexDore|KiiimBuckley|MichaelRyanTV3|maithreyi_s|Nia
mhTowey1|meghas|laloeffelstiel|DebsNaylor|shannonjpower|davemolloy|RichALWilli
ams|nivie|seanmstephens|timhatfield87|IrlEmbDenmark|jofahy|JustinRohrlich|Jaso
nLeMiere|ByClare|SiobhanMorrin|linnytayls|jupadhye|carrollste|ChristinaMcS|Chr
isCreegan|ExSaminer|NicoleHDeSmet|HendersStu|MikaelaConley|alesbon|conormcmaho
n|HolstaT|RDKoncienzczy|RobMcD|c00perM|jvanevra|lucysween|laurabundock|DevikaBh
at|holly_monery|kcassowary|EwanPalmer|anniewhiteabc|IrlEmbPrague|AlysonHenry_|
KathrynRiddell|SincerelyAde|TheGregMurphy|ambhack|mahamabedi|tayknopf|alliehum
phries|ExtraIRL|KirstyHickey|andreleslie|TheaHalpin|JessRapana|joshkelliott|Ja
nine__Graham|JSinnerton|raji47ram|Sinclair_H|karadaniellefox|luke__cooper|RJKD
uggan|lovelynlovie|CatherineNQE|molcranenewman|tobybakare|AnnieeKane|DaisyMcCo
rgray|joejsmiddleton|mglynn720|thisisrachels|JessCiara|OliviaNiland|MaijaKappl
er|ArielScotti|maccytothede|FionaDillon|Legeiser')==True
tweets['screen_news'].value_counts()

```

Out[298]:

```

False      1074318
True        8909
Name: screen_news, dtype: int64

```

In [299]:

```

tweets['news_content'] = (tweets['RT']) | (tweets['text_news']) | (tweets['sc
reen_news'])
tweets['news_content'].value_counts() #~10% of the data

```

Out[299]:

```

False      976890
True       106337
Name: news_content, dtype: int64

```

In [300]:

```
news_data = tweets[tweets['news_content']]
news_data.shape
```

Out[300]:

(106337, 52)

Identifying Politicians

In [301]:

```
tweets['pol'] = tweets['text'].str.contains('PresidentIRL|greenparty_ie|cathma
rtingreen|EamonRyan|SocDems|CathMurphyTD|RoisinShortall|labour|BrendanHowlin|W
illiePenroseTD|workersparty|FineGael|campaignforleo|simoncoveney|fiannafailpar
ty|MichealMartinTD|daracalleary|irelandcp|sinnfeinireland|MaryLouMcDonald|mone
illsf|pb4p|solidarityie|SeamusHealyTD|eirigi1916|FisNuaAnnounce|ddi|Greene_DDI
|2940478|kensmollen|RENUAIreland|johnleahyOY|ModowdMichael|UnitedIreland|jeffr
udd31|AitheantasEire|SocialistParty|JoeHigginsSP',case=False)==True
tweets['pol'].value_counts()
```

Out[301]:

```
False    1043723
True      39504
Name: pol, dtype: int64
```

In [302]:

```
tweets['scpol'] = tweets['Screenname'].str.match('PresidentIRL|greenparty_ie|c
athmartingreen|EamonRyan|SocDems|CathMurphyTD|RoisinShortall|labour|BrendanHow
lin|WilliePenroseTD|workersparty|FineGael|campaignforleo|simoncoveney|fiannafa
ilparty|MichealMartinTD|daracalleary|irelandcp|sinnfeinireland|MaryLouMcDonald
|moneillsf|pb4p|solidarityie|SeamusHealyTD|eirigi1916|FisNuaAnnounce|ddi|Green
e_DDI|2940478|kensmollen|RENUAIreland|johnleahyOY|ModowdMichael|UnitedIreland|
jeffrudd31|AitheantasEire|SocialistParty|JoeHigginsSP')==True
tweets['scpol'].value_counts()
```

Out[302]:

```
False    1080398
True      2829
Name: scpol, dtype: int64
```

In [303]:

```
#Count
polz = tweets.drop_duplicates(subset='Screenname')
polz['scpol'].value_counts()
```

Out[303]:

```
False    259321
True      58
Name: scpol, dtype: int64
```

In [304]:

polz[polz['scpol']]

Out[304]:

	coordinates	Date_Time	hashtags	
1457279	NaN	Fri Apr 13 14:24:01 +0000 2018	8thref	NaN
1457359	NaN	Fri Apr 13 14:30:08 +0000 2018	8thRef , repealthe8th , RegisterToVote	https://twitter.com/greenparty_ie/
1457643	NaN	Fri Apr 13 14:55:46 +0000 2018	NaN	NaN
1457729	NaN	Fri Apr 13 15:12:04 +0000 2018	8thRef	NaN
1457689	NaN	Fri Apr 13 15:13:38 +0000 2018	8thRef	NaN
1458279	NaN	Fri Apr 13 16:18:57 +0000 2018	repealthe8th , 8thref	NaN
1460192	NaN	Fri Apr 13 19:24:40	socdems ,	NaN

		+0000 2018	repealthe8th	
1461056	NaN	Fri Apr 13 20:22:44 +0000 2018	together4yes	NaN
1461622	NaN	Fri Apr 13 20:42:25 +0000 2018	NaN	NaN
1461824	NaN	Fri Apr 13 20:52:12 +0000 2018	socdems , repealthe8th	NaN
1462955	NaN	Fri Apr 13 21:53:44 +0000 2018	socdems , repealthe8th	NaN
1463077	NaN	Fri Apr 13 21:58:16 +0000 2018	socdems , repealthe8th	NaN
1464687	NaN	Sat Apr 14 04:20:07 +0000 2018	Meltdown	NaN
1465671	NaN	Sat Apr 14 08:24:54 +0000 2018	repealthe8th , repeal , Together4Yes , TogetherForYes	https://twitter.com/pb4p/status/984040404040404040
1473317	NaN	Sat Apr 14 19:37:36 +0000 2018	Limerick , repealthe8th	NaN

1485565	NaN	Sun Apr 15 18:43:00 +0000 2018	NewRoss	NaN
1491021	NaN	Mon Apr 16 09:05:25 +0000 2018	repealthe8th , forwomenshealth , Together4Yes	https://twitter.com/deirdrekingston
1491633	NaN	Mon Apr 16 10:30:35 +0000 2018	repealthe8th , TogetherForYes	https://twitter.com/ConorSheehar
1492002	NaN	Mon Apr 16 11:04:20 +0000 2018	savethe8th , loveboth	NaN
1496104	NaN	Mon Apr 16 18:22:38 +0000 2018	Together4Yes	https://twitter.com/workersparty/s
1501391	NaN	Tue Apr 17 08:16:48 +0000 2018	Repealthe8th	NaN
1501826	NaN	Tue Apr 17 08:58:56 +0000 2018	Together4YES , SocDems	https://twitter.com/SocDemsFing
1503952	NaN	Tue Apr 17 12:38:34 +0000 2018	together4yes	NaN
1507485	NaN	Tue Apr 17 18:25:33 +0000	8thref	NaN

		2018		
1530711	NaN	Thu Apr 19 11:48:04 +0000 2018	SexEdBill	NaN
1532640	NaN	Thu Apr 19 14:34:59 +0000 2018	TogetherForYes	https://twitter.com/sinnfeinireland
1533568	NaN	Thu Apr 19 16:08:26 +0000 2018	JoinTheRebellion	NaN
1536640	NaN	Thu Apr 19 20:14:45 +0000 2018	Repealing	NaN
1544338	NaN	Fri Apr 20 11:30:57 +0000 2018	Together4Yes	NaN
1565812	NaN	Sat Apr 21 16:06:01 +0000 2018	SocDemsforYes , Repealthe8th	https://twitter.com/SocDemsDL/s
1569605	NaN	Sat Apr 21 20:04:20 +0000 2018	Repealthe8th , Repeal	https://twitter.com/SocDemsMW/
1586392	NaN	Sun Apr 22 22:18:57 +0000 2018	SocDems	NaN

1590273	NaN	Mon Apr 23 09:17:05 +0000 2018	pknt , repealthe8th	NaN
1606126	NaN	Tue Apr 24 07:28:16 +0000 2018	hometovote	https://twitter.com/LdnIrishARC/s
1615945	NaN	Tue Apr 24 17:48:13 +0000 2018	REPEALTHE8TH	NaN
1616593	NaN	Tue Apr 24 18:25:55 +0000 2018	RepealThe8th	NaN
1636123	NaN	Wed Apr 25 18:45:52 +0000 2018	MayDay2018	NaN
1693345	NaN	Sat Apr 28 12:41:37 +0000 2018	VoteYes , repealthe8th , VoteYes	NaN
1741098	NaN	Tue May 01 12:49:56 +0000 2018	RepealThe8th , TogetherForYes	https://twitter.com/eirigi1916/stati
1747500	NaN	Tue May 01 19:51:48 +0000 2018	8thRef	https://twitter.com/FineGael/statu

1815266	NaN	Sat May 05 14:31:12 +0000 2018	Repeal8th , Together4YES , SocDems	https://twitter.com/SocDemsDBN
1848657	NaN	Mon May 07 10:20:12 +0000 2018	together4yes	NaN
1895435	NaN	Wed May 09 11:06:48 +0000 2018	8thref	NaN
1899229	NaN	Wed May 09 14:10:13 +0000 2018	Together4Yes	NaN
1944504	NaN	Fri May 11 08:24:28 +0000 2018	Together4Yes	https://twitter.com/SocDems/stati
1187417	NaN	Wed May 16 17:54:09 +0000 2018	SocDems	NaN
1201055	NaN	Thu May 17 06:40:25 +0000 2018	indyref2	NaN
1253662	NaN	Fri May 18 15:56:42 +0000 2018	FarmBill2018 , somuchdrama , agchat , voteYES	https://twitter.com/ddills88/status
27144	NaN	Tue May 22 08:32:59 +0000 2018	8thRef	NaN
		Tue May		

35851	NaN	22 11:51:32 +0000 2018	NaN	NaN
89079	NaN	Tue May 22 23:40:49 +0000 2018	Together4YES	NaN
115221	NaN	Wed May 23 11:59:58 +0000 2018	8thRef	NaN
212754	NaN	Thu May 24 10:44:51 +0000 2018	repealthe8th	NaN
435489	NaN	Fri May 25 07:56:28 +0000 2018	VoteYes , Repeal8th , HometoVote	NaN
464933	NaN	Fri May 25 09:02:10 +0000 2018	referendum2018 , 8thamendment	NaN
479894	NaN	Fri May 25 09:34:53 +0000 2018	Repeal8th , repealheroes , VoteYes	NaN
1008755	NaN	Sat May 26 17:24:00 +0000 2018	Repealthe8th	NaN
1090968	NaN	Sun May 27 09:46:34 +0000	Repealthe8th	NaN

In [305]:

```
tweets['rtpol'] = tweets['retweet_screen_name'].str.match('PresidentIRL|greenp
arty_ie|cathmartingreen|EamonRyan|SocDems|CathMurphyTD|RoisinShortall|labour|B
rendanHowlin|WilliePenroseTD|workersparty|FineGael|campaignforleo|simoncoveney
|fiannafailparty|MichealMartinTD|daracalleary|irelandcp|sinnfeinireland|MaryLo
uMcDonald|moneillsf|pb4p|solidarityie|SeamusHealyTD|eirigil1916|FisNuaAnnounce|
ddi|Greene_DDI|2940478|kensmollen|RENUAIreland|johnleahyOY|ModowdMichael|Unite
dIreland|jeffrudd31|AitheantasEire|SocialistParty|JoeHigginsSP',case=False)==T
rue
tweets['rtpol'].value_counts()
```

Out[305]:

```
False      1065443
True        17784
Name: rtpol, dtype: int64
```

In [306]:

```
tweets['poltweet'] = (tweets['scpol']) | (tweets['rtpol']) | (tweets['pol'])
tweets['poltweet'].value_counts()
```

Out[306]:

```
False      1041911
True        41316
Name: poltweet, dtype: int64
```

Identifying Campaign Tweets

Yes Campaign

In [307]:

```
#Is the tweet posted by the yes campaign?
tweets['screen_yc'] = tweets['Screenname'].str.contains('Together4yes|RepealWa
terford|RosForChoice|CorkTogether4Y|westcork4choice|repealglobal|artistsrepeal
8|NWCi|repealeight|repealproject|IrishFPA|atheistie|choiceireland|freesafelega
l|ARCDonegal|ArcOffaly|LeitrimARC|MayoArc|MeathForChoice|TippForChoice|KKforCh
oice|KerryForChoice|ClareARC1|SligoARC|LdnIrishARC|ScotIrishARC|Rebels4Choice|
FingalARC|repealdubwest|GalwayProChoice|GWYEast4Choice|DingleRepeal|NorthKilda
reARC|LaoisForChoice|Repeal_LK|arcwestmeath|RepealArklow|Alliance4Choice|ProCh
oiceWX|Lawyers4Choice|Doctors4Choice|IASW_IRL|FarmersR8th|adopted4choice|Scien
ceForYes|DW_Ireland|sibealnetwork|InclusionIre|SpunOut|SIPTU|AIMSIreland|AkiDw
A|BeLong_To|CommWorkIreland|dublinwellwoman|ICCLtweet|IrishFPA|NTWFIRL|1family
ireland|RCNIreland|SAFEIreland|TFMRIRE|TheUSI|Womens_Aid|parentsforYES|All4Cho
ice|ARNirl|IRLbi|gender_ucd|es_irl|HumanismIreland|HunRealIssues|LINCwomen|lab
ouryouth|MSFC|Midwives4Choice|merj_ireland|NursesForChoice|RosaWomen|SWAIrela
nd|solidarityie|tcdsu|Turepealthe8th|TENI_Tweets|MSFC|Voices_4_Choice|thexilep
roject|YCHREIreland|younggreens|IrishFPA|WGNRR|UCDSU|nxfie')==True
tweets['screen_yc'].value_counts()
```

Out[307]:

```
False      1071729
True        11498
Name: screen_yc, dtype: int64
```

In [308]:

```
camp = tweets.drop_duplicates(subset='Screenname')
camp['screen_yc'].value_counts()
```

Out[308]:

```
False      259296
True         83
Name: screen_yc, dtype: int64
```

In [309]:

```
#Is the tweet a retweet of a yes campaigner's tweet?
```

```
tweets['retweet_yc'] = tweets['retweet_screen_name'].str.contains('Together4yes|RepealWaterford|RosForChoice|CorkTogether4Y|westcork4choice|repealglobal|artistsrepeal8|NWCi|repealeight|repealproject|IrishFPA|atheistie|choiceireland|freesafelegal|ARCDonegal|ArcOffaly|LeitrimARC|MayoArc|MeathForChoice|TippForChoice|KKforChoice|KerryForChoice|ClareARC1|SligoARC|LdnIrishARC|ScotIrishARC|Rebels4Choice|FingalARC|repealdubwest|GalwayProChoice|GWYEast4Choice|DingleRepeal|NorthKildareARC|LaoisForChoice|Repeal_LK|arcwestmeath|RepealArklow|Alliance4Choice|ProChoiceWX|Lawyers4Choice|Doctors4Choice|IASW_IRL|FarmersR8th|adopted4choice|ScienceForYes|DW_Ireland|sibealnetwork|InclusionIre|SpunOut|SIPTU|AIMSIreland|AkiDwA|BeLong_To|CommWorkIreland|dublinwellwoman|ICCLtweet|IrishFPA|NTWFI RL|1familyireland|RCNIreland|SAFEIreland|TFMRIRE|TheUSI|Womens_Aid|parentsfor yes|All4Choice|ARNirl|IRLbi|gender_ucd|es_irl|HumanismIreland|HunRealIssues|LINCwomen|labouryouth|MSFC|Midwives4Choice|merj_ireland|NursesForChoice|RosaWomen|SWAIreland|solidarityie|tcdsu|Turepealthe8th|TENI_Tweets|MSFC|Voices_4_Choice|thexileproject|YCHREIreland|younggreens|IrishFPA|WGNRR|UCDSU|nxfie')==True
tweets['retweet_yc'].value_counts()
```

Out[309]:

```
False      1034271
True        48956
Name: retweet_yc, dtype: int64
```

In [310]:

```
#Does the text mention a yes campaign account?
```

```
tweets['text_yc'] = tweets['text'].str.contains('Together4yes|RepealWaterford|RosForChoice|CorkTogether4Y|westcork4choice|repealglobal|artistsrepeal8|NWCi|repealeight|repealproject|IrishFPA|atheistie|choiceireland|freesafelegal|ARCDonegal|ArcOffaly|LeitrimARC|MayoArc|MeathForChoice|TippForChoice|KKforChoice|KerryForChoice|ClareARC1|SligoARC|LdnIrishARC|ScotIrishARC|Rebels4Choice|FingalARC|repealdubwest|GalwayProChoice|GWYEast4Choice|DingleRepeal|NorthKildareARC|LaoisForChoice|Repeal_LK|arcwestmeath|RepealArklow|Alliance4Choice|ProChoiceWX|Lawyers4Choice|Doctors4Choice|IASW_IRL|FarmersR8th|adopted4choice|ScienceForYes|DW_Ireland|sibealnetwork|InclusionIre|SpunOut|SIPTU|AIMSIreland|AkiDwA|BeLong_To|CommWorkIreland|dublinwellwoman|ICCLtweet|IrishFPA|NTWFI RL|1familyireland|RCNIreland|SAFEIreland|TFMRIRE|TheUSI|Womens_Aid|parentsfor yes|All4Choice|ARNirl|IRLbi|gender_ucd|es_irl|HumanismIreland|HunRealIssues|LINCwomen|labouryouth|MSFC|Midwives4Choice|merj_ireland|NursesForChoice|RosaWomen|SWAIreland|solidarityie|tcdsu|Turepealthe8th|TENI_Tweets|MSFC|Voices_4_Choice|thexileproject|YCHREIreland|younggreens|IrishFPA|WGNRR|UCDSU|nxfie')==True
tweets['text_yc'].value_counts()
```

Out[310]:

```
False      994482
True        88745
Name: text_yc, dtype: int64
```

In [311]:

```
#Variable 'yes_camp' shows tweets that are either posted by the yes campaign, r  
etweeted from yes campaign  
#or mentions the campaign  
tweets['Yes_camp'] = (tweets['screen_yc']) | (tweets['retweet_yc']) | (tweets[  
'text_yc'])  
tweets['Yes_camp'].value_counts()
```

Out[311]:

```
False      983398  
True        99829  
Name: Yes_camp, dtype: int64
```

No campaign

In [312]:

```
#Is the tweet posted by the no campaign?  
tweets['screen_nc'] = tweets['Screenname'].str.contains('prolifecampaign|proli  
feireland|loveboth8|SaveLivesAlwayslifeinstitute|YouthDefence|theRallyforLife|  
doctors4lifeire|NationalPartyIE|CatholicBishops|Islamireland|pciassembly|spucp  
rolife|PreciousLifeCom|ThomasMoreSoc|SJYPS|ionainstitute|OneDayMore|lifefronts  
quare')==True  
tweets['screen_nc'].value_counts()
```

Out[312]:

```
False      1078699  
True        4528  
Name: screen_nc, dtype: int64
```

In [313]:

```
camp2 = tweets.drop_duplicates(subset='Screenname')  
camp2['screen_nc'].value_counts()
```

Out[313]:

```
False      259367  
True        12  
Name: screen_nc, dtype: int64
```

In [314]:

```
#Is the tweet a retweet of a no campaigner's tweet?
tweets['retweet_nc'] = tweets['retweet_screen_name'].str.contains('prolifecampaign|prolifeireland |loveboth8|SaveLivesAlwayslifeinstitute|YouthDefence|theRallyforLife|doctors4lifeire|NationalPartyIE|CatholicBishops|Islamireland|pciassembly|spucprolifePreciousLifeCom|ThomasMoreSoc|SJYPS|ionainstitute|OneDayMore|lifefrontsquare')==True
tweets['retweet_nc'].value_counts()
```

Out[314]:

```
False    1067815
True      15412
Name: retweet_nc, dtype: int64
```

In [315]:

```
#Does the text mention a no campaign account?
tweets['text_nc'] = tweets['text'].str.contains('prolifecampaign|prolifeireland|loveboth8|SaveLivesAlwayslifeinstitute|YouthDefence|theRallyforLife|doctors4lifeire|NationalPartyIE|CatholicBishops|Islamireland|pciassembly|spucprolifePreciousLifeCom|ThomasMoreSoc|SJYPS|ionainstitute|OneDayMore|lifefrontsquare')==True
tweets['text_nc'].value_counts()
```

Out[315]:

```
False    1063455
True      19772
Name: text_nc, dtype: int64
```

In [316]:

```
#Variable 'no_camp' shows tweets that are either posted by the no campaign, retweeted from no campaign
#or mentions the campaign
tweets['No_camp'] = (tweets['screen_nc']) | (tweets['retweet_nc']) | (tweets['text_nc'])
tweets['No_camp'].value_counts()
```

Out[316]:

```
False    1059294
True      23933
Name: No_camp, dtype: int64
```

Dividing Tweets by Support

In [317]:

```
tweets.shape
```

Out[317]:

```
(1083227, 64)
```

In [318]:

```
tweets.head(2)
```

Out[318]:

	coordinates	Date_Time	hashtags	media	urls	favorite_count
1457174	NaN	Fri Apr 13 14:16:57 +0000 2018	RepealThe8th	NaN	NaN	0
1457164	NaN	Fri Apr 13 14:16:59 +0000 2018	TrustWomenVoteYes , RepealThe8th	NaN	NaN	3

In [319]:

```
#TWEETS SUPPORTING YES VOTE
YES_T = tweets[tweets['text'].str.contains('healthcarenotairfare|repeal|repeal
forbetterbirth|wetrustwomen|voteyes|ivotedyes|togetherforyes|together4yes|repe
althe8th|mybodymychoice|trustwomen|men4yes|roadtorepeal|lawyersforyes|academic
sforyes|farmersforyes|lawyersforchoice|studentsforchoice|nowforNI|time4choice|
whyimvotingyes|deaftogetherforyes|register4yes|whoneedsyouryes',case=False)==T
rue]
```

In [320]:

```
#NUMBER OF ACCOUNTS SUPPORTING YES VOTE
YES_U = YES_T.drop_duplicates(subset='user_screen_name')
```

In [321]:

```
#NUMBER OF TWEETS IN YES VOTE SUBSET THAT HASHTAG NO HASHTAGS
YES_E = YES_T[YES_T['text'].str.contains('lovebothvoteno|hometovoteno|lifecanv
ass|menforno|votenotoabortion|rallyforlife|votenotoabortion|standupforlife|men
4no|savethe8th|rallyforlife|protectthe8th|Togetherforno',case = False)==True]
```

In [322]:

```
#NUMBER OF USERS IN YES VOTE SUBSET THAT HASHTAG NO HASHTAGS
YES_E_U = YES_E.drop_duplicates(subset='user_screen_name')
```

In [323]:

```
#DIMENSIONS; tweets supporting yes, users supporting yes, selective exposure i
n tweets, selective exposure in users
YES_T.shape, YES_U.shape, YES_E.shape, YES_E_U.shape
```

Out[323]:

```
((677458, 64), (194157, 64), (30132, 64), (13004, 64))
```

In [324]:

```
#TWEETS SUPPORTING NO VOTE
NO_T = tweets[tweets['text'].str.contains('lovebothvoteno|hometovoteno|lifecan
vass|votenotoabortion|rallyforlife|votenotoabortion|standupforlife|men4no|menf
orno|savethe8th|rallyforlife|protectthe8th|Togetherforno',case = False)==True]
```

In [325]:

```
#USERS SUPPORTING NO VOTE
NO_U = NO_T.drop_duplicates(subset='user_screen_name')
```

In [326]:

```
NO_E = NO_T[NO_T['text'].str.contains('healthcarenotairfare|repeal|repealforbe
tterbirth|wetrustwomen|voteyes|ivotedyes|togetherforyes|together4yes|repealthe
8th|mybodymychoice|trustwomen|men4yes|roadtorepeal|lawyersforyes|academicsfory
es|farmersforyes|lawyersforchoice|studentsforchoice|nowforNI|time4choice|whyim
votingyes|deaftogetherforyes|register4yes|whoneedsyouryes')==True]
NO_E_U = NO_E.drop_duplicates(subset='user_screen_name')
```

In [327]:

```
#DIMENSIONS; tweets supporting NO, users supporting NO, error in tweets, error
in users
NO_T.shape, NO_U.shape, NO_E.shape, NO_E_U.shape
```

Out[327]:

```
((137027, 64), (27077, 64), (20952, 64), (11122, 64))
```

In [328]:

```
#PUT IT ALL IN A DATAFRAME AND CALC ERROR
```


In [329]:

```
tweets['Yes_VOTE'] = tweets['text'].str.contains('healthcarenotairfare|repeal|
repealforbetterbirth|wetrustwomen|voteyes|ivotedyes|togetherforyes|together4ye
s|repealthe8th|mybodymychoice|trustwomen|men4yes|roadtorepeal|lawyersforyes|ac
ademicsforyes|farmersforyes|lawyersforchoice|studentsforchoice|wemadehistory|n
owforNI|time4choice|whyimvotingyes|deaftogetherforyes|register4yes|whoneedsyou
ryes',case = False)==True
```

In [330]:

```
tweets['No_VOTE'] = tweets['text'].str.contains('lovebothvoteno|menforno|homet
ovoteno|lifecanvass|votenotoabortion|rallyforlife|votenotoabortion|standupforl
ife|men4no|savethe8th|rallyforlife|protectthe8th|Togetherforno',case = False)=
=True
```

In [331]:

```
tweets['Yes_VOTE'].value_counts()
```

Out[331]:

```
True      677532
False     405695
Name: Yes_VOTE, dtype: int64
```

In [332]:

```
tweets['No_VOTE'].value_counts()
```

Out[332]:

```
False     946200
True      137027
Name: No_VOTE, dtype: int64
```

Sentiment Scoring Entire Dataset With Vader

In [333]:

```
sid = SentimentIntensityAnalyzer()
tweets['sentiment_compound_polarity']=tweets.text.apply(lambda x:sid.polarity_
scores(x)['compound'])
tweets['sentiment_neutral']=tweets.text.apply(lambda x:sid.polarity_scores(x)[
'neu'])
tweets['sentiment_negative']=tweets.text.apply(lambda x:sid.polarity_scores(x)
['neg'])
tweets['sentiment_pos']=tweets.text.apply(lambda x:sid.polarity_scores(x)['pos
'])
tweets['sentiment_type']=''
tweets.loc[tweets.sentiment_compound_polarity>0,'sentiment_type']='POSITIVE'
tweets.loc[tweets.sentiment_compound_polarity==0,'sentiment_type']='NEUTRAL'
tweets.loc[tweets.sentiment_compound_polarity<0,'sentiment_type']='NEGATIVE'
tweets.head()
```

Out[333]:

	coordinates	Date_Time	hashtags	
1457174	NaN	Fri Apr 13 14:16:57 +0000 2018	RepealThe8th	NaN
1457164	NaN	Fri Apr 13 14:16:59 +0000 2018	TrustWomenVoteYes , RepealThe8th	NaN
1457160	NaN	Fri Apr 13 14:17:10 +0000 2018	savethe8th , savethe8th	NaN
1457133	NaN	Fri Apr 13 14:17:27 +0000 2018	Together4Yes	NaN
1457132	NaN	Fri Apr 13 14:18:21 +0000 2018	LoveBothVoteNO	https://twitter.com/loveboth8/s

In [334]:

```
pd.set_option('display.max_columns', None)
tweets.head()
```

Out[334]:

	coordinates	Date_Time	hashtags	
1457174	NaN	Fri Apr 13 14:16:57 +0000 2018	RepealThe8th	NaN
1457164	NaN	Fri Apr 13 14:16:59 +0000 2018	TrustWomenVoteYes , RepealThe8th	NaN
1457160	NaN	Fri Apr 13 14:17:10 +0000 2018	savethe8th , savethe8th	NaN
1457133	NaN	Fri Apr 13 14:17:27 +0000 2018	Together4Yes	NaN
1457132	NaN	Fri Apr 13 14:18:21 +0000 2018	LoveBothVoteNO	https://twitter.com/loveboth8/s

In [335]:

```
tweets.shape
```

Out[335]:

```
(1083227, 71)
```

Delete all unnecessary variables

In [336]:

```
tweets.shape
```

Out[336]:

```
(1083227, 71)
```

In [337]:

```
tweets.columns
```

Out[337]:

```
Index(['coordinates', 'Date_Time', 'hashtags', 'media', 'urls',
      'favorite_count', 'id', 'in_reply_to_screen_name',
      'in_reply_to_status_id', 'in_reply_to_user_id', 'lang', 'place',
      'possibly_sensitive', 'retweet_count', 'retweeted_id',
      'retweeted_screen_name', 'source', 'text', 'tweet_url', 'user_date_time',
      'user_screen_name', 'user_default_profile_image', 'user_description',
      'user_favourites_count', 'user_followers_count', 'user_friends_count',
      'user_listed_count', 'user_location', 'user_name', 'Screenname',
      'user_statuses_count', 'user_time_zone', 'user_urls', 'user_verified',
      'rtt', 'date1', 'DATE', 'TIME', 'friend_foll_ratio', 'is_spam', 'hash',
      'contains_url', 'is_retweet', 'Logged_RT_count', 'desc', 'screen',
      'news_g', 'news', 'RT', 'text_news', 'screen_news', 'news_content',
      'pol', 'scpol', 'rtpol', 'poltweet', 'screen_yc', 'retweet_yc',
      'text_yc', 'Yes_camp', 'screen_nc', 'retweet_nc', 'text_nc', 'No_camp',
      'Yes_VOTE', 'No_VOTE', 'sentiment_compound_polarity',
      'sentiment_neutral', 'sentiment_negative', 'sentiment_pos',
      'sentiment_type'],
      dtype='object')
```

In [338]:

```
tweets3 = tweets[['DATE','TIME','hashtags','urls','contains_url','favorite_count',  
  
'is_retweet','retweet_count','Logged_RT_count','retweet_screen_name','retweet_yc',  
'retweet_nc','rtpol','RT',  
  
                'text','text_news','sentiment_compound_polarity','sentiment_type','Yes_VOTE',  
                'No_VOTE',  
  
                'news_content', 'Yes_camp','No_camp','poltweet','friend_foll_ratio','user_statuses_count',  
                'user_location',  
  
                'Screenname','user_description','user_favourites_count','user_followers_count',  
                'user_friends_count',  
                'screen_yc','screen_nc','scpol','news']]
```

In [339]:

```
tweets3.shape
```

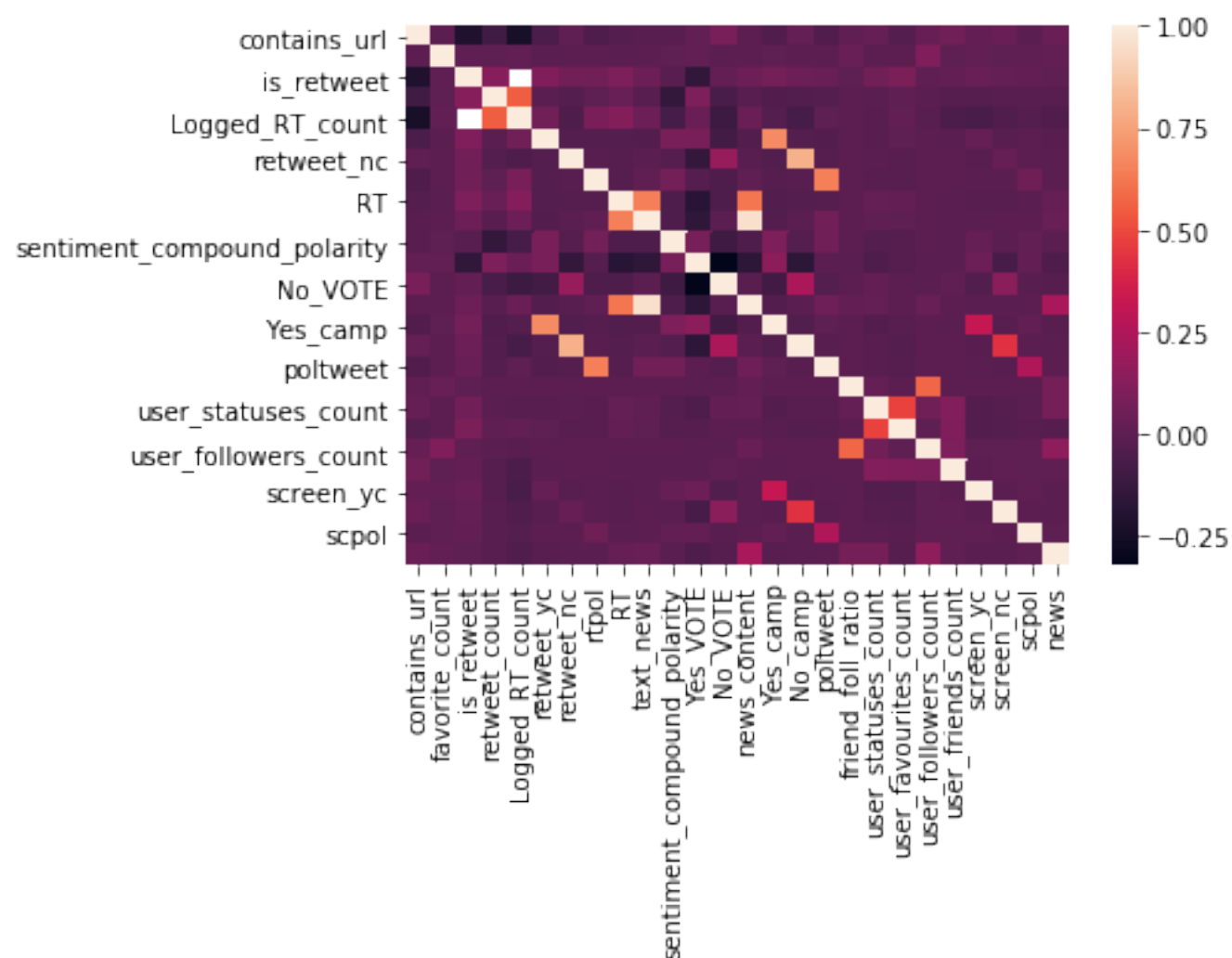
Out[339]:

```
(1083227, 36)
```

Preparing Data for Linear Regression

In [340]:

```
sns.heatmap(tweets3.corr()) #show the correlation between my variables
plt.show()
```



In [393]:

```
corr = tweets3.corr()
corr.style.background_gradient()
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/matplotlib/colors.py:504: RuntimeWarning:

invalid value encountered in less

Out[393]:

	contains_url	favorite_count	is_retweet	retweet_c
contains_url	1	-6.71037e-05	-0.208056	-0.093492
favorite_count	-6.71037e-05	1	0.0106504	0.0023840
is_retweet	-0.208056	0.0106504	1	0.12519
retweet_count	-0.0934928	0.00238404	0.12519	1
Logged_RT_count	-0.232073	0.0113523	nan	0.552111
retweet_yc	-0.057299	-0.0076044	0.104847	-0.006830
retweet_nc	0.00618991	-0.00419914	0.0578964	-0.030043
rtpol	-0.0489658	-0.00451574	0.0622616	0.0029883
RT	-0.0275239	-0.00718331	0.0990413	0.0320494
text_news	-0.0112621	-0.00403429	0.0447472	-0.015459
sentiment_compound_polarity	-0.0106737	0.00311565	-0.0101497	-0.143015
Yes_VOTE	0.0139629	0.0125009	-0.144327	0.0998593
No_VOTE	0.0839018	-0.0045624	0.0182687	-0.065801
news_content	-0.000599938	-0.00033758	0.0426695	-0.016867
Yes_camp	-0.042735	0.00212026	0.0688517	-0.042183
No_camp	0.0220216	-0.0024229	0.0445265	-0.037831
poltweet	-0.0421177	-7.711e-05	0.0321024	-0.025346
friend_foll_ratio	0.0071685	0.0362897	0.0027281	-0.001396
user_statuses_count	0.0266442	-0.000863839	0.0579242	-0.004784
user_favourites_count	-0.0254218	-0.00290484	0.094784	0.0089095
user_followers_count	0.0289445	0.106539	0.00952381	-0.004735
user_friends_count	0.0639394	0.00421902	0.0165159	-0.020188
screen_yc	0.0207966	0.0135544	0.0353583	-0.024432
screen_nc	0.0367831	0.00269556	0.0225885	-0.016454
scpol	-0.00464512	0.0106352	0.0208676	-0.011852
news	0.0410724	0.0150246	0.00157661	-0.011609
campaign_support	0.0418751	0.0105751	0.0398089	-0.028172
support	0.0911215	0.00439635	-0.0840853	0.0066814

In [341]:

```
#Convert all true-false variables into 0 1:
tweets3['Yes_VOTE'] = tweets3.Yes_VOTE.astype(int)
tweets3['No_VOTE'] = tweets3.No_VOTE.astype(int)
tweets3['No_VOTE'] = tweets3['No_VOTE'].apply(lambda x: x*2)
tweets3['news'] = tweets3.news.astype(int)
tweets3['news_content'] = tweets3.news_content.astype(int)
tweets3['text_news'] = tweets3.text_news.astype(int)
tweets3['screen_nc'] = tweets3.screen_nc.astype(int)
tweets3['screen_yc'] = tweets3.screen_yc.astype(int)
tweets3['retweet_nc'] = tweets3.retweet_nc.astype(int)
tweets3['retweet_yc'] = tweets3.retweet_yc.astype(int)
tweets3['scpol'] = tweets3.scpol.astype(int)
tweets3['rtpol'] = tweets3.rtpol.astype(int)
#tweets3['sentiment_type'] = tweets3['sentiment_type'].astype(int)
tweets3['contains_url'] = tweets3['contains_url'].astype(int)
tweets3['sentiment_type'].value_counts()
#change senti-type to numerical variable
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:3: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:4: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:5: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:6: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:7: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:8: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:9: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:10: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:11: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:12: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:13: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:15: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

Out[341]:

POSITIVE	559010
NEUTRAL	284965
NEGATIVE	239252

Name: sentiment_type, dtype: int64

In [342]:

```
tweets3['screen_nc'] = tweets3['screen_nc'].apply(lambda x: x*2) #change no c  
ampaign support to 2
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packag
es/ipykernel_launcher.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

In [343]:

```
tweets3['screen_nc'].value_counts()
```

Out[343]:

```
0    1078699  
2     4528  
Name: screen_nc, dtype: int64
```

In [344]:

```
tweets3['campaign_support'] = (tweets3.screen_nc + tweets3.screen_nc)  
tweets3['campaign_support'].value_counts()
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packag
es/ipykernel_launcher.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

Out[344]:

```
0    1067201  
1     11498  
2     4528  
Name: campaign_support, dtype: int64
```

In [345]:

```
#Combine yes vote and no vote into one supporters column
tweets3['support'] = (tweets3.Yes_VOTE + tweets3.No_VOTE)
#data_sorted['friend_foll_ratio'] = (data_sorted.user_followers_count / data_s
orted.user_friends_count)
#tweets_reg = tweets4[tweets4['support']==1]
pd.set_option('display.max_colwidth', -1)
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packag
es/ipykernel_launcher.py:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

In [346]:

```
tweets3['support'].value_counts()
#0 = neither
#1 = yes
#2 = no
#3 = both
```

Out[346]:

```
1    647400
0    298800
2    106895
3     30132
Name: support, dtype: int64
```

In [347]:

```
tweets3['support'] = tweets3['support'] + 1
tweets3['support'].value_counts()
#1 = neither
#2 = yes
#3 = no
#4 = both
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

Out[347]:

```
2    647400
1    298800
3    106895
4     30132
Name: support, dtype: int64
```

In [348]:

```
tweets4 = tweets3[tweets3['retweet_count']>0] #get rid of the tweets with 0 re
tweets so the log is not infinite

          #linear reg looks at the relationship between amount of retweets a
nd variables
```

In [349]:

```
tweets4['support'].value_counts()
#1 = neither
#2 = yes
#3 = no
#4 = both
```

Out[349]:

```
2    497169
1    268127
3     90699
4     23075
Name: support, dtype: int64
```

In [350]:

```
tweets.shape, tweets3.shape, tweets4.shape
```

Out[350]:

((1083227, 71), (1083227, 38), (879070, 38))

LINEAR REGRESSION

In [351]:

```
tweets4['is_retweet'].value_counts()  
#tweets_reg.columns
```

Out[351]:

True 879070
Name: is_retweet, dtype: int64

In [352]:

```
tweets_NEG_N = tweets4[tweets4['sentiment_compound_polarity']<=0]  
tweets_POS_N = tweets4[tweets4['sentiment_compound_polarity']>=0]  
tweets_NEG_POS = tweets4[tweets4['sentiment_compound_polarity']!=0]  
tweets_NEG_N.shape, tweets_POS_N.shape, tweets_NEG_POS.shape
```

Out[352]:

((431223, 38), (689457, 38), (637460, 38))

In [353]:

```
tweets4.head()
```

Out[353]:

	DATE	TIME	hashtags	urls	contains_url	favorite_count	is_ret
1457174	2018-04-13	14:16:57	RepealThe8th	NaN	0	0	True
1457160	2018-04-13	14:17:10	savethe8th , savethe8th	NaN	0	0	True

1457133	2018-04-13	14:17:27	Together4Yes	NaN	0	0	True
1457132	2018-04-13	14:18:21	LoveBothVoteNO	NaN	0	0	True
1457201	2018-04-13	14:18:22	RepealThe8th	NaN	0	0	True

In [354]:

```
#Divide the data by endorsement and run separate regressions
tweets_neither = tweets4[tweets4['support']==1]
tweets_yes = tweets4[tweets4['support']==2]
tweets_no = tweets4[tweets4['support']==3]
tweets_both = tweets4[tweets4['support']==4]
tweets_neither.shape, tweets_yes.shape, tweets_no.shape, tweets_both.shape
```

Out[354]:

```
((268127, 38), (497169, 38), (90699, 38), (23075, 38))
```


In [355]:

```
#how many users in each
tweets_nei_users = tweets_neither.drop_duplicates(subset='Screenname')
tweets_yes_users = tweets_yes.drop_duplicates(subset='Screenname')
tweets_no_users = tweets_no.drop_duplicates(subset='Screenname')
tweets_both_users = tweets_both.drop_duplicates(subset='Screenname')
tweets_nei_users.shape, tweets_yes_users.shape, tweets_no_users.shape, tweets_
both_users.shape
```

Out[355]:

```
((100131, 38), (159036, 38), (15795, 38), (11979, 38))
```

In [356]:

```
tweets4['sentiment_type'].value_counts()
```

Out[356]:

```
POSITIVE      447847
NEUTRAL       241610
NEGATIVE      189613
Name: sentiment_type, dtype: int64
```

In [357]:

```
# WITHIN THE SUBSET OF TWEETS THAT ENDORSE YES, NO, NEITHER OR BOTH CAMPAIGNS,
HOW DOES SENTIMENT POLARITY AFFECT THE OUTCOME? HOW DOES A NEWS TWEET AFFECT I
T? A TWEET FROM A POLITICIAN? A TWEET FROM THE YES CAMP? NO CAMP?
```

In [453]:

```
reg_test = smf.ols(formula = 'Logged_RT_count ~ sentiment_compound_polarity +
C(support) + news + scpol + screen_yc + screen_nc', data= tweets4).fit()
reg_test.summary()
```

Out[453]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.033
Model:	OLS	Adj. R-squared:	0.033
Method:	Least Squares	F-statistic:	3778.
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	04:00:38	Log-Likelihood:	-2.0701e+06
No. Observations:	879070	AIC:	4.140e+06
Df Residuals:	879061	BIC:	4.140e+06
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.1645	0.005	825.326	0.000	4.155	4.174
C(support)[T.2]	-0.0798	0.006	-13.005	0.000	-0.092	-0.068
C(support)[T.3]	-1.2535	0.010	-126.445	0.000	-1.273	-1.234
C(support)[T.4]	0.0660	0.018	3.765	0.000	0.032	0.100
sentiment_compound_polarity	-0.4776	0.006	-75.435	0.000	-0.490	-0.465
news	-1.6147	0.036	-44.866	0.000	-1.685	-1.544
scpol	-1.3195	0.049	-27.068	0.000	-1.415	-1.224
screen_yc	-1.5600	0.025	-63.234	0.000	-1.608	-1.512
screen_nc	-0.3813	0.020	-19.295	0.000	-0.420	-0.343

Omnibus:	80886.207	Durbin-Watson:	1.742
Prob(Omnibus):	0.000	Jarque-Bera (JB):	33002.171
Skew:	0.264	Prob(JB):	0.00
Kurtosis:	2.211	Cond. No.	21.5

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [358]:

```
reg_yes = smf.ols(formula = 'Logged_RT_count ~ sentiment_compound_polarity + news + scpol + screen_yc +screen_nc', data = tweets_yes).fit()  
reg_yes.summary()
```

Out[358]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.022
Model:	OLS	Adj. R-squared:	0.022
Method:	Least Squares	F-statistic:	2247.
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	00:57:11	Log-Likelihood:	-1.2080e+06
No. Observations:	497169	AIC:	2.416e+06
Df Residuals:	497163	BIC:	2.416e+06
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.1446	0.004	943.482	0.000	4.136	4.153
sentiment_compound_polarity	-0.7578	0.009	-83.969	0.000	-0.776	-0.740
news	-0.9010	0.071	-12.760	0.000	-1.039	-0.763
scpol	-1.3776	0.060	-23.133	0.000	-1.494	-1.261
screen_yc	-1.6179	0.029	-55.303	0.000	-1.675	-1.561
screen_nc	-0.7654	0.103	-7.411	0.000	-0.968	-0.563

Omnibus:	88317.718	Durbin-Watson:	1.707
Prob(Omnibus):	0.000	Jarque-Bera (JB):	24677.177
Skew:	0.280	Prob(JB):	0.00
Kurtosis:	2.064	Cond. No.	27.2

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [359]:

```
reg_no = smf.ols(formula = 'Logged_RT_count ~ sentiment_compound_polarity + news + scpol + screen_yc +screen_nc', data = tweets_no).fit()  
reg_no.summary()
```

Out[359]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.016
Model:	OLS	Adj. R-squared:	0.016
Method:	Least Squares	F-statistic:	289.2
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	4.38e-308
Time:	00:57:11	Log-Likelihood:	-1.7379e+05
No. Observations:	90699	AIC:	3.476e+05
Df Residuals:	90693	BIC:	3.476e+05
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	2.8923	0.006	510.393	0.000	2.881	2.903
sentiment_compound_polarity	-0.3881	0.012	-32.669	0.000	-0.411	-0.365
news	-0.5312	0.260	-2.043	0.041	-1.041	-0.022
scpol	-0.1111	0.171	-0.648	0.517	-0.447	0.225
screen_yc	0.7151	0.439	1.627	0.104	-0.146	1.576
screen_nc	-0.2662	0.014	-18.481	0.000	-0.294	-0.238

Omnibus:	1493.135	Durbin-Watson:	1.593
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1578.858
Skew:	0.309	Prob(JB):	0.00
Kurtosis:	3.190	Cond. No.	81.2

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [360]:

```
reg_both = smf.ols(formula = 'Logged_RT_count ~ sentiment_compound_polarity +
news + scpol + screen_yc +screen_nc', data = tweets_both).fit()
reg_both.summary()
```

Out[360]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.008
Model:	OLS	Adj. R-squared:	0.008
Method:	Least Squares	F-statistic:	37.97
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	6.17e-39
Time:	00:57:11	Log-Likelihood:	-58220.
No. Observations:	23075	AIC:	1.165e+05
Df Residuals:	23069	BIC:	1.165e+05
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.2272	0.020	211.321	0.000	4.188	4.266
sentiment_compound_polarity	0.4865	0.053	9.232	0.000	0.383	0.590
news	-1.3254	0.328	-4.042	0.000	-1.968	-0.683
scpol	-0.6054	0.910	-0.665	0.506	-2.389	1.178
screen_yc	-2.6161	1.006	-2.601	0.009	-4.588	-0.644
screen_nc	-0.9953	0.108	-9.243	0.000	-1.206	-0.784

Omnibus:	66.615	Durbin-Watson:	0.520
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2101.679
Skew:	0.132	Prob(JB):	0.00
Kurtosis:	1.545	Cond. No.	50.7

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [361]:

```
reg_neither = smf.ols(formula = 'Logged_RT_count ~ sentiment_compound_polarity
+ news + scpol + screen_yc +screen_nc', data = tweets_neither).fit()
reg_neither.summary()
```

Out[361]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.012
Model:	OLS	Adj. R-squared:	0.012
Method:	Least Squares	F-statistic:	630.3
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	00:57:12	Log-Likelihood:	-6.0982e+05
No. Observations:	268127	AIC:	1.220e+06
Df Residuals:	268121	BIC:	1.220e+06
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.0981	0.005	837.346	0.000	4.088	4.108
sentiment_compound_polarity	-0.0305	0.011	-2.811	0.005	-0.052	-0.009
news	-1.9531	0.040	-48.263	0.000	-2.032	-1.874
scpol	-1.1853	0.105	-11.310	0.000	-1.391	-0.980
screen_yc	-1.2041	0.055	-21.926	0.000	-1.312	-1.097
screen_nc	-0.7452	0.051	-14.598	0.000	-0.845	-0.645

Omnibus:	32873.454	Durbin-Watson:	1.701
Prob(Omnibus):	0.000	Jarque-Bera (JB):	9437.077
Skew:	0.148	Prob(JB):	0.00
Kurtosis:	2.130	Cond. No.	23.4

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [362]:

```
reg_yes.params
```

Out[362]:

Intercept	4.144577
sentiment_compound_polarity	-0.757818
news	-0.900978
scpol	-1.377561
screen_yc	-1.617866
screen_nc	-0.765431

dtype: float64

In [363]:

```
reg_no.params
```

Out[363]:

Intercept	2.892325
sentiment_compound_polarity	-0.388132
news	-0.531217
scpol	-0.111084
screen_yc	0.715057
screen_nc	-0.266217

dtype: float64

In [364]:

```
reg_both.params
```

Out[364]:

Intercept	4.227175
sentiment_compound_polarity	0.486538
news	-1.325450
scpol	-0.605369
screen_yc	-2.616086
screen_nc	-0.995287

dtype: float64

In [365]:

```
reg_neither.params
```

Out[365]:

Intercept	4.098057
sentiment_compound_polarity	-0.030490
news	-1.953141
scpol	-1.185304
screen_yc	-1.204136
screen_nc	-0.745199

dtype: float64

In [366]:

```
reg2 = ols(formula = 'Logged_RT_count ~ C(support)*sentiment_compound_polarity
+ C(news) + C(scpol) + C(screen_yc) + C(screen_nc)', data= tweets4)
```

In [367]:

```
res = reg2.fit()
```

In [368]:

```
print(res.summary())
```

OLS Regression Results					
=====					
=====					
Dep. Variable:	Logged_RT_count		R-squared:		
0.037					
Model:	OLS		Adj. R-squared:		
0.037					
Method:	Least Squares		F-statistic:		
3039.					
Date:	Fri, 24 Aug 2018		Prob (F-statistic):		
0.00					
Time:	00:57:32		Log-Likelihood:		
-2.0686e+06					
No. Observations:	879070		AIC:		
4.137e+06					
Df Residuals:	879058		BIC:		
4.137e+06					
Df Model:	11				
Covariance Type:	nonrobust				
=====					
=====					
				coef	std err
t	P> t	[0.025	0.975]		

Intercept				4.0944	0.005
777.319	0.000	4.084	4.105		
C(support)[T.2]				0.0506	0.007
7.636	0.000	0.038	0.064		
C(support)[T.3]				-1.1915	0.010
-116.975	0.000	-1.211	-1.172		
C(support)[T.4]				0.1233	0.018
7.016	0.000	0.089	0.158		
C(news)[T.1]				-1.6133	0.036
-44.903	0.000	-1.684	-1.543		
C(scpol)[T.1]				-1.2999	0.049
-26.713	0.000	-1.395	-1.205		
C(screen_yc)[T.1]				-1.5452	0.025
-62.739	0.000	-1.594	-1.497		
C(screen_nc)[T.2]				-0.7674	0.039
-19.447	0.000	-0.845	-0.690		
sentiment_compound_polarity				-0.0277	0.012
-2.359	0.018	-0.051	-0.005		


```

C(support)[T.2]:sentiment_compound_polarity    -0.7311    0.014
-50.759    0.000    -0.759    -0.703
C(support)[T.3]:sentiment_compound_polarity    -0.3585    0.022
-16.432    0.000    -0.401    -0.316
C(support)[T.4]:sentiment_compound_polarity     0.5102    0.046
11.098    0.000    0.420    0.600
=====
=====

```

```

Omnibus:                84156.703    Durbin-Watson:
1.743
Prob(Omnibus):          0.000    Jarque-Bera (JB):
32743.205
Skew:                   0.251    Prob(JB):
0.00
Kurtosis:               2.199    Cond. No.
21.7
=====
=====

```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [369]:

```
parameters = print(res.params)
```

```

Intercept                4.094413
C(support)[T.2]          0.050630
C(support)[T.3]          -1.191466
C(support)[T.4]          0.123294
C(news)[T.1]             -1.613258
C(scpol)[T.1]            -1.299907
C(screen_yc)[T.1]        -1.545249
C(screen_nc)[T.2]        -0.767369
sentiment_compound_polarity -0.027677
C(support)[T.2]:sentiment_compound_polarity -0.731109
C(support)[T.3]:sentiment_compound_polarity -0.358488
C(support)[T.4]:sentiment_compound_polarity  0.510196
dtype: float64

```

In [396]:

```

reg2d = smf.ols(formula = 'Logged_RT_count ~ C(support)*sentiment_compound_polarity ', data = tweets4).fit()
reg2d.summary()

```

Out[396] :

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.029
Model:	OLS	Adj. R-squared:	0.029
Method:	Least Squares	F-statistic:	3742.
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	02:22:07	Log-Likelihood:	-2.0721e+06
No. Observations:	879070	AIC:	4.144e+06
Df Residuals:	879062	BIC:	4.144e+06
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.0602	0.005	771.302	0.000	4.050	4.070
C(support)[T.2]	0.0501	0.007	7.541	0.000	0.037	0.063
C(support)[T.3]	-1.1875	0.010	-117.286	0.000	-1.207	-1.168
C(support)[T.4]	0.1439	0.018	8.158	0.000	0.109	0.178
sentiment_compound_polarity	-0.0341	0.012	-2.892	0.004	-0.057	-0.011
C(support) [T.2]:sentiment_compound_polarity	-0.7431	0.014	-51.392	0.000	-0.771	-0.715
C(support) [T.3]:sentiment_compound_polarity	-0.3594	0.022	-16.412	0.000	-0.402	-0.316
C(support) [T.4]:sentiment_compound_polarity	0.5095	0.046	11.040	0.000	0.419	0.600

Omnibus:	90131.160	Durbin-Watson:	1.736
Prob(Omnibus):	0.000	Jarque-Bera (JB):	34614.871
Skew:	0.265	Prob(JB):	0.00
Kurtosis:	2.185	Cond. No.	20.7

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [454]:

```
reg2 = smf.ols(formula = 'Logged_RT_count ~ C(support)*sentiment_compound_polarity + C(news) + C(scpol) + C(screen_yc) + C(screen_nc) - 1', data = tweets4).fit()
reg2.summary()
```

Out[454]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.037
Model:	OLS	Adj. R-squared:	0.037
Method:	Least Squares	F-statistic:	3039.
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	04:35:22	Log-Likelihood:	-2.0686e+06
No. Observations:	879070	AIC:	4.137e+06
Df Residuals:	879058	BIC:	4.137e+06
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
C(support)[1]	4.0944	0.005	777.319	0.000	4.084	4.105
C(support)[2]	4.1450	0.004	1021.063	0.000	4.137	4.153
C(support)[3]	2.9029	0.009	332.397	0.000	2.886	2.920
C(support)[4]	4.2177	0.017	251.503	0.000	4.185	4.251
C(news)[T.1]	-1.6133	0.036	-44.903	0.000	-1.684	-1.543
C(scpol)[T.1]	-1.2999	0.049	-26.713	0.000	-1.395	-1.205
C(screen_yc)[T.1]	-1.5452	0.025	-62.739	0.000	-1.594	-1.497
C(screen_nc)[T.2]	-0.7674	0.039	-19.447	0.000	-0.845	-0.690
sentiment_compound_polarity	-0.0277	0.012	-2.359	0.018	-0.051	-0.005
C(support)[T.2]:sentiment_compound_polarity	-0.7311	0.014	-50.759	0.000	-0.759	-0.703
C(support)[T.3]:sentiment_compound_polarity	-0.3585	0.022	-16.432	0.000	-0.401	-0.316
C(support)[T.4]:sentiment_compound_polarity	0.5102	0.046	11.098	0.000	0.420	0.600

Omnibus:	84156.703	Durbin-Watson:	1.743

Prob(Omnibus):	0.000	Jarque-Bera (JB):	32743.205
Skew:	0.251	Prob(JB):	0.00
Kurtosis:	2.199	Cond. No.	14.5

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [371]:

```
print(reg2.params)
#1 = neither
#2 = yes
#3 = no
#4 = both
```

```
Intercept                                4.094413
C(support)[T.2]                          0.050630
C(support)[T.3]                         -1.191466
C(support)[T.4]                          0.123294
C(news)[T.1]                             -1.613258
C(scpol)[T.1]                           -1.299907
C(screen_yc)[T.1]                       -1.545249
C(screen_nc)[T.2]                       -0.767369
sentiment_compound_polarity              -0.027677
C(support)[T.2]:sentiment_compound_polarity -0.731109
C(support)[T.3]:sentiment_compound_polarity -0.358488
C(support)[T.4]:sentiment_compound_polarity  0.510196
dtype: float64
```

In [372]:

```
reg3 = smf.ols(formula = 'Logged_RT_count ~ C(support)*sentiment_type + C(news) + C(scpol) + C(screen_yc) + C(screen_nc)', data = tweets4).fit()
reg3.summary()
```

Out[372]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.055
Model:	OLS	Adj. R-squared:	0.055
Method:	Least Squares	F-statistic:	3440.
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	00:58:18	Log-Likelihood:	-2.0599e+06
No. Observations:	879070	AIC:	4.120e+06
Df Residuals:	879054	BIC:	4.120e+06
Df Model:	15		
Covariance Type:	nonrobust		

Covariance Type:	nonrobust		
------------------	-----------	--	--

	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.7167	0.011	353.538	0.000	3.696	3.737
C(support)[T.2]	1.1413	0.013	86.053	0.000	1.115	1.167
C(support)[T.3]	-0.5384	0.018	-29.228	0.000	-0.575	-0.502
C(support)[T.4]	-0.4060	0.032	-12.809	0.000	-0.468	-0.344
sentiment_type[T.NEUTRAL]	0.9079	0.014	65.374	0.000	0.881	0.935
sentiment_type[T.POSITIVE]	0.2265	0.013	18.019	0.000	0.202	0.251
C(news)[T.1]	-1.6616	0.036	-46.693	0.000	-1.731	-1.592
C(scpol)[T.1]	-1.2641	0.048	-26.234	0.000	-1.359	-1.170
C(screen_ye)[T.1]	-1.5276	0.024	-62.643	0.000	-1.575	-1.480
C(screen_nc)[T.2]	-0.7486	0.039	-19.160	0.000	-0.825	-0.672
C(support) [T.2]:sentiment_type[T.NEUTRAL]	-1.7745	0.017	-101.658	0.000	-1.809	-1.740
C(support) [T.3]:sentiment_type[T.NEUTRAL]	-1.2204	0.026	-46.484	0.000	-1.272	-1.169
C(support) [T.4]:sentiment_type[T.NEUTRAL]	-1.5116	0.052	-29.016	0.000	-1.614	-1.409
C(support) [T.2]:sentiment_type[T.POSITIVE]	-1.4325	0.016	-91.036	0.000	-1.463	-1.402
C(support) [T.3]:sentiment_type[T.POSITIVE]	-0.7584	0.023	-32.395	0.000	-0.804	-0.712
C(support) [T.4]:sentiment_type[T.POSITIVE]	1.7157	0.040	43.185	0.000	1.638	1.794

Omnibus:	112585.519	Durbin-Watson:	1.751
Prob(Omnibus):	0.000	Jarque-Bera (JB):	32432.772
Skew:	0.170	Prob(JB):	0.00
Kurtosis:	2.123	Cond. No.	33.1

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [373]:

```
print(reg3.params)
```

```
Intercept                                3.716655
C(support)[T.2]                          1.141334
C(support)[T.3]                         -0.538400
C(support)[T.4]                         -0.405967
sentiment_type[T.NEUTRAL]                0.907926
sentiment_type[T.POSITIVE]              0.226494
C(news)[T.1]                            -1.661575
C(scpol)[T.1]                           -1.264066
C(screen_yc)[T.1]                       -1.527618
C(screen_nc)[T.2]                       -0.748593
C(support)[T.2]:sentiment_type[T.NEUTRAL] -1.774492
C(support)[T.3]:sentiment_type[T.NEUTRAL] -1.220397
C(support)[T.4]:sentiment_type[T.NEUTRAL] -1.511573
C(support)[T.2]:sentiment_type[T.POSITIVE] -1.432508
C(support)[T.3]:sentiment_type[T.POSITIVE] -0.758370
C(support)[T.4]:sentiment_type[T.POSITIVE]  1.715706
dtype: float64
```

In [374]:

```
reg4 = smf.ols(formula = 'Logged_RT_count ~ C(support) + sentiment_compound_polarity + C(support):sentiment_compound_polarity + C(news) + C(scpol) + C(screen_yc) + C(screen_nc)', data = tweets4).fit()
reg4.summary()
```

Out[374]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.037
Model:	OLS	Adj. R-squared:	0.037
Method:	Least Squares	F-statistic:	3039.
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	00:58:39	Log-Likelihood:	-2.0686e+06
No. Observations:	879070	AIC:	4.137e+06
Df Residuals:	879058	BIC:	4.137e+06
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.0944	0.005	777.319	0.000	4.084	4.105
C(support)[T.2]	0.0506	0.007	7.636	0.000	0.038	0.064
C(support)[T.3]	-1.1915	0.010	-116.975	0.000	-1.211	-1.172

C(support)[T.4]	0.1233	0.018	7.016	0.000	0.089	0.158
C(news)[T.1]	-1.6133	0.036	-44.903	0.000	-1.684	-1.543
C(scpol)[T.1]	-1.2999	0.049	-26.713	0.000	-1.395	-1.205
C(screen_yc)[T.1]	-1.5452	0.025	-62.739	0.000	-1.594	-1.497
C(screen_nc)[T.2]	-0.7674	0.039	-19.447	0.000	-0.845	-0.690
sentiment_compound_polarity	-0.0277	0.012	-2.359	0.018	-0.051	-0.005
C(support)[T.2]:sentiment_compound_polarity	-0.7311	0.014	-50.759	0.000	-0.759	-0.703
C(support)[T.3]:sentiment_compound_polarity	-0.3585	0.022	-16.432	0.000	-0.401	-0.316
C(support)[T.4]:sentiment_compound_polarity	0.5102	0.046	11.098	0.000	0.420	0.600

Omnibus:	84156.703	Durbin-Watson:	1.743
Prob(Omnibus):	0.000	Jarque-Bera (JB):	32743.205
Skew:	0.251	Prob(JB):	0.00
Kurtosis:	2.199	Cond. No.	21.7

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [375]:

```
#interacting with just news and sentiment
reg5 = smf.ols(formula = 'Logged_RT_count ~ C(news)*sentiment_compound_polarit
y + C(scpol) + C(screen_yc) + C(screen_nc) + C(support)', data=tweets4).fit()
reg5.summary()
```

Out[375] :

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.033
Model:	OLS	Adj. R-squared:	0.033
Method:	Least Squares	F-statistic:	3363.
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	00:58:59	Log-Likelihood:	-2.0701e+06
No. Observations:	879070	AIC:	4.140e+06
Df Residuals:	879060	BIC:	4.140e+06
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.1650	0.005	825.344	0.000	4.155	4.175
C(news)[T.1]	-1.7123	0.039	-43.822	0.000	-1.789	-1.636
C(scpol)[T.1]	-1.3197	0.049	-27.073	0.000	-1.415	-1.224
C(screen_yc)[T.1]	-1.5597	0.025	-63.222	0.000	-1.608	-1.511
C(screen_nc)[T.2]	-0.7625	0.040	-19.295	0.000	-0.840	-0.685
C(support)[T.2]	-0.0799	0.006	-13.010	0.000	-0.092	-0.068
C(support)[T.3]	-1.2537	0.010	-126.469	0.000	-1.273	-1.234
C(support)[T.4]	0.0656	0.018	3.746	0.000	0.031	0.100
sentiment_compound_polarity	-0.4802	0.006	-75.691	0.000	-0.493	-0.468
C(news) [T.1]:sentiment_compound_polarity	0.6391	0.100	6.410	0.000	0.444	0.835

Omnibus:	80793.365	Durbin-Watson:	1.742
Prob(Omnibus):	0.000	Jarque-Bera (JB):	32982.118
Skew:	0.264	Prob(JB):	0.00
Kurtosis:	2.211	Cond. No.	44.5

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

OLS NEWS

In [376]:

```
news_df = tweets[tweets['news_content']]
news_df = news_df[news_df['retweet_count']>0] #get rid of the tweets with 0 re
tweets so the log is not infinite
news_df.columns
```

Out[376]:

```
Index(['coordinates', 'Date_Time', 'hashtags', 'media', 'urls',
      'favorite_count', 'id', 'in_reply_to_screen_name',
      'in_reply_to_status_id', 'in_reply_to_user_id', 'lang', 'place',
      'possibly_sensitive', 'retweet_count', 'retweeted_id',
      'retweeted_screen_name', 'source', 'text', 'tweet_url', 'user_
      _Date_Time',
      'user_screen_name', 'user_default_profile_image', 'user_des
      cription',
      'user_favourites_count', 'user_followers_count', 'user_frie
      nds_count',
      'user_listed_count', 'user_location', 'user_name', 'Screenn
      ame',
      'user_statuses_count', 'user_time_zone', 'user_urls', 'user
      _verified',
      'rtt', 'date1', 'DATE', 'TIME', 'friend_foll_ratio', 'is_sp
      am', 'hash',
      'contains_url', 'is_retweet', 'Logged_RT_count', 'desc', 's
      creen',
      'news_g', 'news', 'RT', 'text_news', 'screen_news', 'news_c
      ontent',
      'pol', 'scpol', 'rtpol', 'poltweet', 'screen_yc', 'retweet_
      yc',
      'text_yc', 'Yes_camp', 'screen_nc', 'retweet_nc', 'text_nc'
      , 'No_camp',
      'Yes_VOTE', 'No_VOTE', 'sentiment_compound_polarity',
      'sentiment_neutral', 'sentiment_negative', 'sentiment_pos',
      'sentiment_type'],
      dtype='object')
```

In [377]:

```
#Convert all true-false variables into 0 1:
news_df['Yes_VOTE'] = news_df.Yes_VOTE.astype(int)
news_df['No_VOTE'] = news_df.No_VOTE.astype(int)
news_df['No_VOTE'] = news_df['No_VOTE'].apply(lambda x: x*2)
news_df['news'] = news_df.news.astype(int)
news_df['news_content'] = news_df.news_content.astype(int)
news_df['text_news'] = news_df.text_news.astype(int)
news_df['screen_nc'] = news_df.screen_nc.astype(int)
news_df['screen_yc'] = news_df.screen_yc.astype(int)
news_df['retweet_nc'] = news_df.retweet_nc.astype(int)
news_df['retweet_yc'] = news_df.retweet_yc.astype(int)
news_df['scpol'] = news_df.scpol.astype(int)
news_df['rtpol'] = news_df.rtpol.astype(int)
news_df['contains_url'] = news_df['contains_url'].astype(int)
news_df['sentiment_type'].value_counts()
```

Out[377]:

```
POSITIVE      41843
NEUTRAL       30076
NEGATIVE      19755
Name: sentiment_type, dtype: int64
```

In [378]:

```
#Combine yes vote and no vote into one supporters column
news_df['support'] = (news_df.Yes_VOTE + news_df.No_VOTE)
pd.set_option('display.max_colwidth', -1)
news_df = news_df
news_df['support2'] = news_df['support'] + 1
news_df['support1'].value_counts()
#1 = neither
#2 = yes
#3 = no
#4 = both
```

Out[378]:

```
1      51306
2      30600
3       8572
4       1196
Name: support2, dtype: int64
```

In [455]:

```
news_df['support'].value_counts()
```

Out[455]:

```
0      51306
1      30600
2       8572
3       1196
Name: support, dtype: int64
```

In [456]:

```
#Considering all tweets that are news, what is the relationship between the re
tweet count
news1 = smf.ols(formula = 'Logged_RT_count ~ C(support2)*sentiment_compound_po
larity + scpol + screen_yc + screen_nc ', data = news_df).fit()
news1.summary()
```

Out[456]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.062
Model:	OLS	Adj. R-squared:	0.062
Method:	Least Squares	F-statistic:	607.3
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	04:59:58	Log-Likelihood:	-2.1128e+05
No. Observations:	91674	AIC:	4.226e+05
Df Residuals:	91663	BIC:	4.227e+05
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	4.6423	0.011	408.415	0.000	4.620	4.665
C(support2)[T.2]	-0.9087	0.019	-48.637	0.000	-0.945	-0.872
C(support2)[T.3]	-0.7264	0.029	-25.265	0.000	-0.783	-0.670
C(support2)[T.4]	-2.8127	0.071	-39.494	0.000	-2.952	-2.673
sentiment_compound_polarity	-0.3850	0.032	-11.917	0.000	-0.448	-0.322
C(support2) [T.2]:sentiment_compound_polarity	-0.2606	0.046	-5.662	0.000	-0.351	-0.170
C(support2) [T.3]:sentiment_compound_polarity	-1.1413	0.067	-16.987	0.000	-1.273	-1.010
C(support2) [T.4]:sentiment_compound_polarity	0.3449	0.157	2.191	0.028	0.036	0.653
scpol	-1.2587	0.167	-7.529	0.000	-1.586	-0.931
screen_yc	-1.2348	0.083	-14.789	0.000	-1.398	-1.071
screen_nc	-1.3979	0.137	-10.176	0.000	-1.667	-1.129

Omnibus:	20381.696	Durbin-Watson:	1.410
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3785.200

Skew:	0.030	Prob(JB):	0.00
Kurtosis:	2.006	Cond. No.	22.6

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [380]:

```
#Divide the data by endorsement and run separate regressions
news_neither = news_df[news_df['support']==0]
news_yes = news_df[news_df['support']==1]
news_no = news_df[news_df['support']==2]
news_both = news_df[news_df['support']==3]
news_neither.shape, news_yes.shape, news_no.shape, news_both.shape
```

Out[380]:

```
((51306, 73), (30600, 73), (8572, 73), (1196, 73))
```

In [381]:

```
news_yes = smf.ols(formula = 'Logged_RT_count ~ sentiment_compound_polarity +
scpol + screen_yc +screen_nc', data = news_yes).fit()
news_yes.summary()
```

Out[381]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.021
Model:	OLS	Adj. R-squared:	0.021
Method:	Least Squares	F-statistic:	163.3
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	1.45e-138
Time:	00:59:29	Log-Likelihood:	-68199.
No. Observations:	30600	AIC:	1.364e+05
Df Residuals:	30595	BIC:	1.364e+05
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.7307	0.014	269.776	0.000	3.704	3.758
sentiment_compound_polarity	-0.6471	0.030	-21.297	0.000	-0.707	-0.588
scpol	-0.8468	0.223	-3.798	0.000	-1.284	-0.410
screen_yc	-1.1524	0.090	-12.763	0.000	-1.329	-0.975
screen_nc	-1.2119	0.441	-2.748	0.006	-2.076	-0.348

Omnibus:	5062.086	Durbin-Watson:	1.408
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1122.117
Skew:	0.022	Prob(JB):	2.17e-244
Kurtosis:	2.063	Cond. No.	34.8

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [382]:

```
news_no = smf.ols(formula = 'Logged_RT_count ~ sentiment_compound_polarity + s
cpol + screen_yc +screen_nc', data = news_no).fit()
news_no.summary()
```

Out [382] :

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.084
Model:	OLS	Adj. R-squared:	0.084
Method:	Least Squares	F-statistic:	196.9
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	8.48e-162
Time:	00:59:29	Log-Likelihood:	-19639.
No. Observations:	8572	AIC:	3.929e+04
Df Residuals:	8567	BIC:	3.932e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.9155	0.026	149.368	0.000	3.864	3.967
sentiment_compound_polarity	-1.5260	0.058	-26.247	0.000	-1.640	-1.412
scpol	-1.3657	0.664	-2.056	0.040	-2.668	-0.064
screen_yc	0.7367	1.197	0.616	0.538	-1.609	3.082
screen_nc	-1.4078	0.164	-8.587	0.000	-1.729	-1.086

Omnibus:	2933.660	Durbin-Watson:	0.802
Prob(Omnibus):	0.000	Jarque-Bera (JB):	427.480
Skew:	-0.112	Prob(JB):	1.49e-93
Kurtosis:	1.929	Cond. No.	46.3

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [383]:

```
news_neither = smf.ols(formula = 'Logged_RT_count ~ sentiment_compound_polarity + scpol + screen_yc +screen_nc', data = news_neither).fit()
news_yes.summary()
```

Out[383]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.021
Model:	OLS	Adj. R-squared:	0.021
Method:	Least Squares	F-statistic:	163.3
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	1.45e-138
Time:	00:59:29	Log-Likelihood:	-68199.
No. Observations:	30600	AIC:	1.364e+05
Df Residuals:	30595	BIC:	1.364e+05
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.7307	0.014	269.776	0.000	3.704	3.758
sentiment_compound_polarity	-0.6471	0.030	-21.297	0.000	-0.707	-0.588
scpol	-0.8468	0.223	-3.798	0.000	-1.284	-0.410
screen_yc	-1.1524	0.090	-12.763	0.000	-1.329	-0.975
screen_nc	-1.2119	0.441	-2.748	0.006	-2.076	-0.348

Omnibus:	5062.086	Durbin-Watson:	1.408
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1122.117
Skew:	0.022	Prob(JB):	2.17e-244
Kurtosis:	2.063	Cond. No.	34.8

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [384]:

```
news_both = smf.ols(formula = 'Logged_RT_count ~ sentiment_compound_polarity +
scpol + screen_yc +screen_nc', data = news_both).fit()
news_both.summary()
```

Out[384]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.001
Model:	OLS	Adj. R-squared:	-0.002
Method:	Least Squares	F-statistic:	0.2761
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.894
Time:	00:59:30	Log-Likelihood:	-2408.6
No. Observations:	1196	AIC:	4827.
Df Residuals:	1191	BIC:	4853.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	1.7980	0.053	33.843	0.000	1.694	1.902
sentiment_compound_polarity	-0.0465	0.116	-0.403	0.687	-0.273	0.180
scpol	1.0023	1.819	0.551	0.582	-2.567	4.572
screen_yc	0.1479	1.817	0.081	0.935	-3.418	3.714
screen_nc	0.3255	0.410	0.794	0.427	-0.479	1.130

Omnibus:	435.571	Durbin-Watson:	1.303
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1505.465
Skew:	1.788	Prob(JB):	0.00
Kurtosis:	7.174	Cond. No.	34.7

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

IRELAND Regression

In [385]:

```
#location_filtering
Ireland=tweets4[tweets4['user_location'].str.contains('Ireland|Munster|Connacht|Ulster|Antrim|Armagh|Carlow|Cavan|Clare|Cork|Derry|Donegal|Down|Dublin|Fermanagh|Galway|Kerry|Kildare|Kilkenny|Laois|Leitrim|Limerick|Longford|Louth|Mayo|Meath|Monaghan|Offaly|Roscommon|Sligo|Tipperary|Tyrone|Waterford|Westmeath|Wexford|Wicklow',case=False)==True]
print(Ireland.shape)
```

(343098, 38)

In [386]:

```
tweets4['support'].value_counts()
```

Out[386]:

```
2    497169
1    268127
3     90699
4     23075
Name: support, dtype: int64
```

In [387]:

```
reg4 = smf.ols(formula = 'Logged_RT_count ~ C(support)*sentiment_compound_polarity + C(news) + C(scpol) + C(screen_yc) + C(screen_nc)', data = Ireland).fit()
reg4.summary()
```

Out[387]:

OLS Regression Results

Dep. Variable:	Logged_RT_count	R-squared:	0.028
Model:	OLS	Adj. R-squared:	0.028
Method:	Least Squares	F-statistic:	912.5
Date:	Fri, 24 Aug 2018	Prob (F-statistic):	0.00
Time:	00:59:48	Log-Likelihood:	-7.5358e+05
No. Observations:	343098	AIC:	1.507e+06
Df Residuals:	343086	BIC:	1.507e+06
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.5859	0.007	482.816	0.000	3.571	3.600
C(support)[T.2]	-0.3814	0.009	-41.172	0.000	-0.400	-0.363

C(support)[T.3]	-1.0016	0.014	-69.095	0.000	-1.030	-0.973
C(support)[T.4]	-0.9749	0.026	-37.431	0.000	-1.026	-0.924
C(news)[T.1]	-1.5257	0.038	-39.764	0.000	-1.601	-1.450
C(scpol)[T.1]	-0.7190	0.043	-16.729	0.000	-0.803	-0.635
C(screen_yc)[T.1]	-0.8917	0.023	-39.548	0.000	-0.936	-0.847
C(screen_nc)[T.2]	-0.3215	0.037	-8.756	0.000	-0.393	-0.250
sentiment_compound_polarity	0.1512	0.016	9.316	0.000	0.119	0.183
C(support) [T.2]:sentiment_compound_polarity	-0.1993	0.020	-10.077	0.000	-0.238	-0.161
C(support) [T.3]:sentiment_compound_polarity	-0.2090	0.030	-6.937	0.000	-0.268	-0.150
C(support) [T.4]:sentiment_compound_polarity	-0.4069	0.059	-6.895	0.000	-0.523	-0.291

Omnibus:	15165.635	Durbin-Watson:	1.852
Prob(Omnibus):	0.000	Jarque-Bera (JB):	14387.822
Skew:	0.452	Prob(JB):	0.00
Kurtosis:	2.566	Cond. No.	19.7

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Average Share RT yes UPDATE

```
In [471]:  
  
#Get all RETWEETS that are endorsing the yes campaign from regression data  
tweets_yesRT = tweets_yes[tweets_yes['is_retweet']]  
tweets_yes.shape  
  
Out[471]:  
  
(497169, 38)  
  
In [472]:  
  
#divide these retweets by sentiment  
tweets_yes_negative = tweets_yesRT[tweets_yesRT['sentiment_type'].str.contains  
( 'NEGATIVE' )==True]  
tweets_yes_positive = tweets_yesRT[tweets_yesRT['sentiment_type'].str.contains  
( 'POSITIVE' )==True]  
tweets_yes_neutral = tweets_yesRT[tweets_yesRT['sentiment_type'].str.contains(  
'NEUTRAL' )==True]
```

In [473]:

```
#How many are there?
```

In [474]:

```
#Proportion of tweets that are positive and supporting YES  
tweets_yes_positive.shape, tweets_yesRT.shape
```

Out[474]:

```
((263141, 38), (497169, 38))
```

In [475]:

```
#Proportion of tweets that are negative and supporting YES  
tweets_yes_neutral.shape, tweets_yesRT.shape
```

Out[475]:

```
((136966, 38), (497169, 38))
```

In [476]:

```
#Proportion of tweets that are neutral and supporting YES  
tweets_yes_negative.shape, tweets_yesRT.shape #
```

Out[476]:

```
((97062, 38), (497169, 38))
```

In [477]:

```
#27.1% neutral ; 20% negative ; 52.8% positive
```

Average share RT no UPDATE

In [478]:

```
#Get all RETWEETS that are endorsing the no campaign from above  
tweets_noRT = tweets_no[tweets_no['is_retweet']]  
tweets_no.shape
```

Out[478]:

```
(90699, 38)
```

In [479]:

```
#Get all RETWEETS that are endorsing the yes campaign  
tweets_no_negative = tweets_noRT[tweets_noRT['sentiment_type'].str.contains('NEGATIVE')==True]  
tweets_no_positive = tweets_noRT[tweets_noRT['sentiment_type'].str.contains('POSITIVE')==True]  
tweets_no_neutral = tweets_noRT[tweets_noRT['sentiment_type'].str.contains('NEUTRAL')==True]
```

In [480]:

```
#Proportion of tweets that are positive and supporting NO  
tweets_no_positive.shape, tweets_noRT.shape
```

Out[480]:

```
((39154, 38), (90699, 38))
```

In [481]:

```
#Proportion of tweets that are negative and supporting NO  
tweets_no_neutral.shape, tweets_noRT.shape
```

Out[481]:

```
((23637, 38), (90699, 38))
```

In [482]:

```
#Proportion of tweets that are neutral and supporting NO  
tweets_no_negative.shape, tweets_noRT.shape
```

Out[482]:

```
((27908, 38), (90699, 38))
```

In [483]:

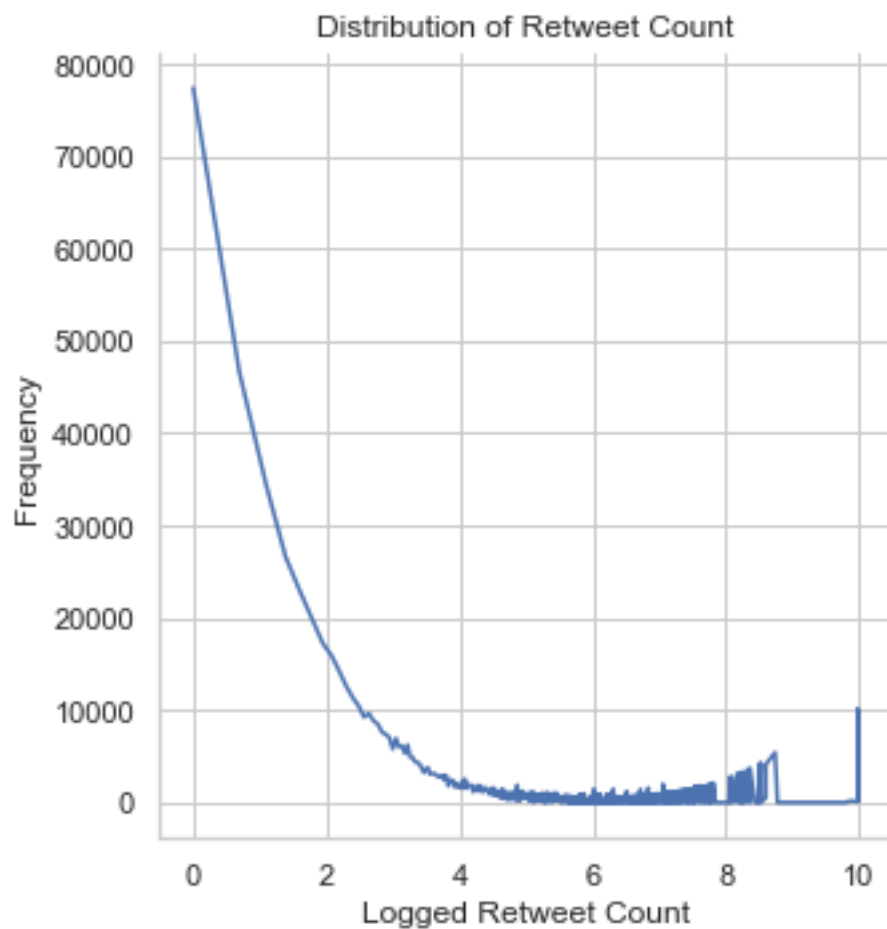
```
#24.2% neutral ; 30.7% negative ; 45.0% positive
```

Analysis of All Tweets

Distribution of Retweet Count by Endorsement

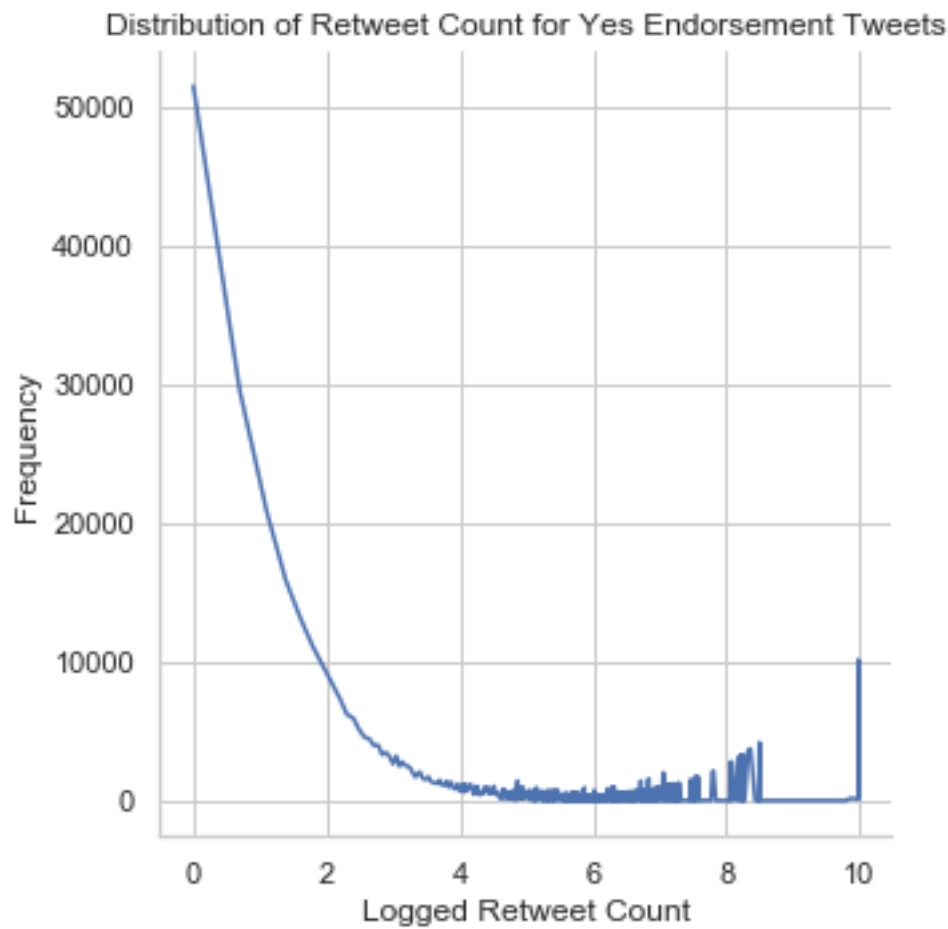
In [468]:

```
freq_RT = pd.DataFrame(tweets4['Logged_RT_count'].value_counts().reset_index()
)
FRT = sns.relplot(x='index',y='Logged_RT_count',data = freq_RT, kind="line",ma
rkers = True)
FRT.set(xlabel = 'Logged Retweet Count', ylabel = 'Frequency')
FRT = plt.title("Distribution of Retweet Count")
```



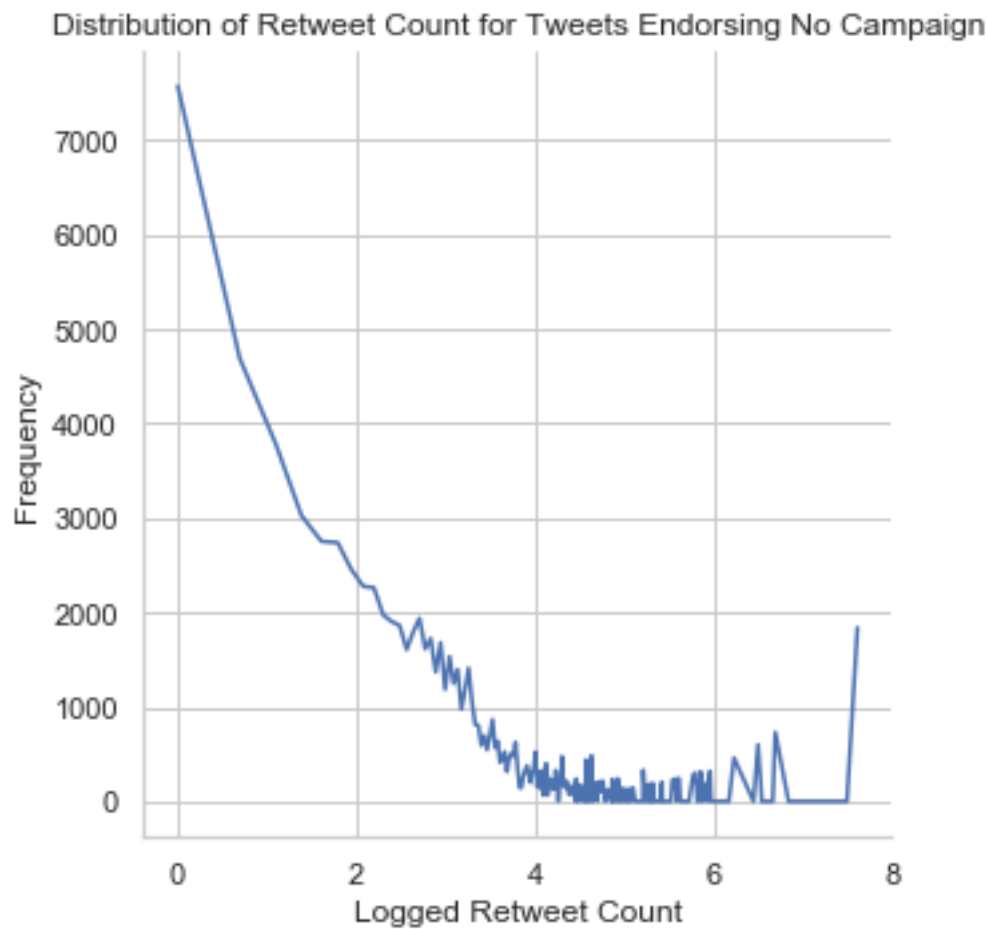
In [469]:

```
freq_RT_Y = pd.DataFrame(tweets_yes['Logged_RT_count'].value_counts().reset_index())
FRTY = sns.relplot(x='index',y='Logged_RT_count',data = freq_RT_Y, kind="line",markers = True)
FRTY.set(xlabel = 'Logged Retweet Count', ylabel = 'Frequency')
FRTY = plt.title("Distribution of Retweet Count for Yes Endorsement Tweets")
```



In [470]:

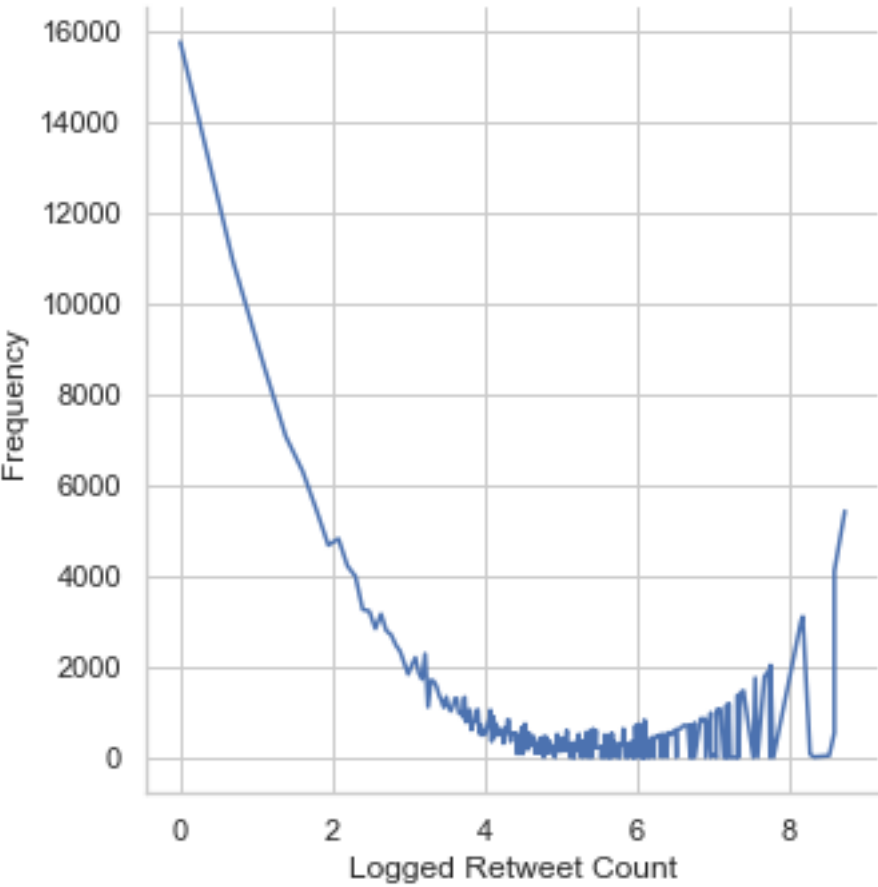
```
RT_dis_N = pd.DataFrame(tweets_no['Logged_RT_count'].value_counts().reset_index())
FRTN = sns.relplot(x='index',y='Logged_RT_count',data = RT_dis_N, kind="line",
markers = True)
FRTN.set(xlabel = 'Logged Retweet Count', ylabel = 'Frequency')
FRTN = plt.title("Distribution of Retweet Count for Tweets Endorsing No Campaign")
```



In [447]:

```
RT_dis_Neither = pd.DataFrame(tweets_neither['Logged_RT_count'].value_counts().reset_index())
FRTNei = sns.relplot(x='index',y='Logged_RT_count',data = RT_dis_Neither, kind="line",markers = True)
FRTNei.set(xlabel = 'Logged Retweet Count', ylabel = 'Frequency')
FRTNei = plt.title("Distribution of Retweet Count for Tweets Endorsing Neither Campaign")
```

Distribution of Retweet Count for Tweets Endorsing Neither Campaign



In []:

```
RT_dis_both = pd.DataFrame(tweets_both['Logged_RT_count'].value_counts().reset_index())
FRTB = sns.relplot(x='index',y='Logged_RT_count',data = RT_dis_both, kind="line",markers = True)
FRTB.set(xlabel = 'Logged Retweet Count', ylabel = 'Frequency')
FRTB = plt.title("Distribution of Retweet Count for Tweets Endorsing Both Campaigns")
```

In [450]:

```
sent_dist = pd.DataFrame(tweets4['sentiment_compound_polarity'].value_counts().reset_index())
sent_dist
```

Out[450]:

	index	sentiment_compound_polarity
0	0.0000	241610
1	-0.3182	22932
2	0.4019	22781

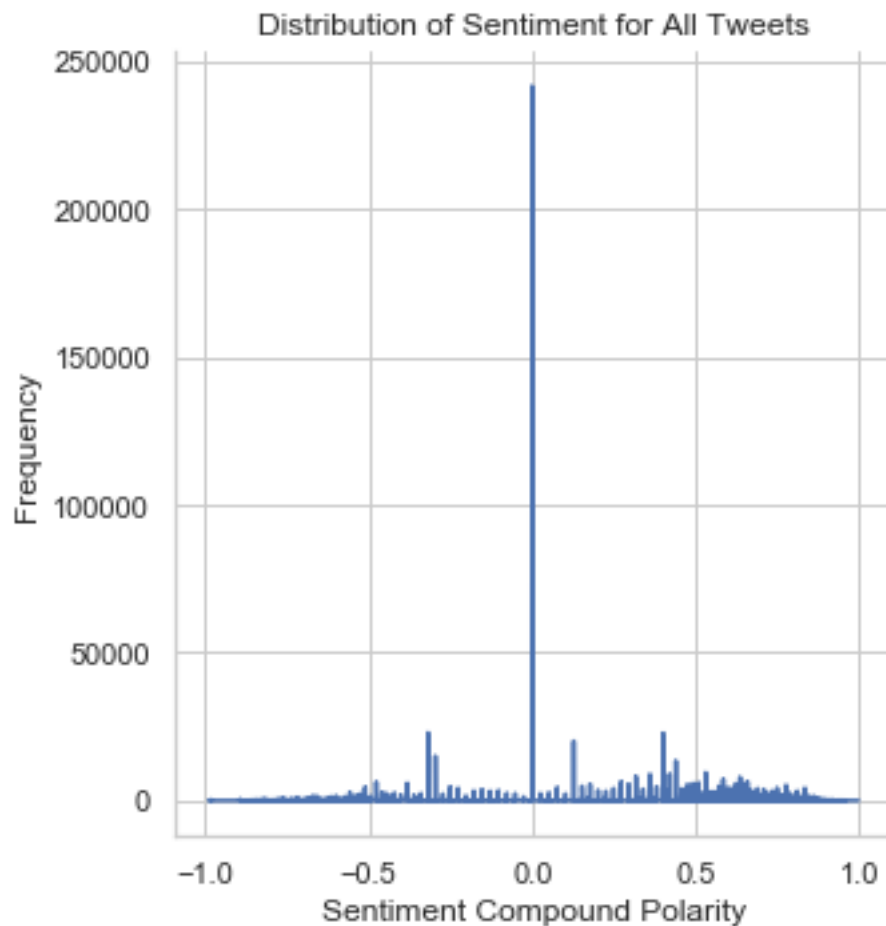
3	0.1280	20108
4	-0.2960	15054
5	0.4404	13463
6	0.5319	9447
7	0.4215	9001
8	0.3612	8949
9	0.3182	8215
10	0.6369	7985
11	0.5859	7247
12	0.2732	6459
13	0.6597	6414
14	-0.4767	6294
15	0.5106	6107
16	-0.3818	5892
17	0.4939	5823
18	0.6486	5778
19	0.6249	5742
20	0.2960	5585
21	0.1779	5541
22	0.4767	5475
23	0.7783	5259
24	0.5719	4889
25	0.1531	4876
26	-0.2500	4749
27	0.3818	4741
28	-0.5106	4586
29	0.0772	4490
...
10822	0.3886	1
10823	0.5477	1
10824	0.4611	1
10825	0.6415	1
10826	-0.7673	1
10827	-0.4792	1

10828	-0.7520	1
10829	-0.8490	1
10830	-0.9421	1
10831	-0.8543	1
10832	-0.9820	1
10833	0.0286	1
10834	0.5643	1
10835	-0.7838	1
10836	-0.6758	1
10837	-0.8701	1
10838	0.8048	1
10839	-0.0942	1
10840	0.8354	1
10841	0.9844	1
10842	-0.7133	1
10843	0.2549	1
10844	0.1608	1
10845	0.5734	1
10846	-0.7466	1
10847	-0.3524	1
10848	-0.5890	1
10849	-0.5205	1
10850	-0.5551	1
10851	-0.0946	1

10852 rows × 2 columns

In [452]:

```
SD = sns.relplot(x='index',y='sentiment_compound_polarity',data = sent_dist, k
ind="line",markers = True)
SD.set(xlabel = 'Sentiment Compound Polarity', ylabel = 'Frequency')
SD = plt.title("Distribution of Sentiment for All Tweets")
```



Summary of Data

In [400]:

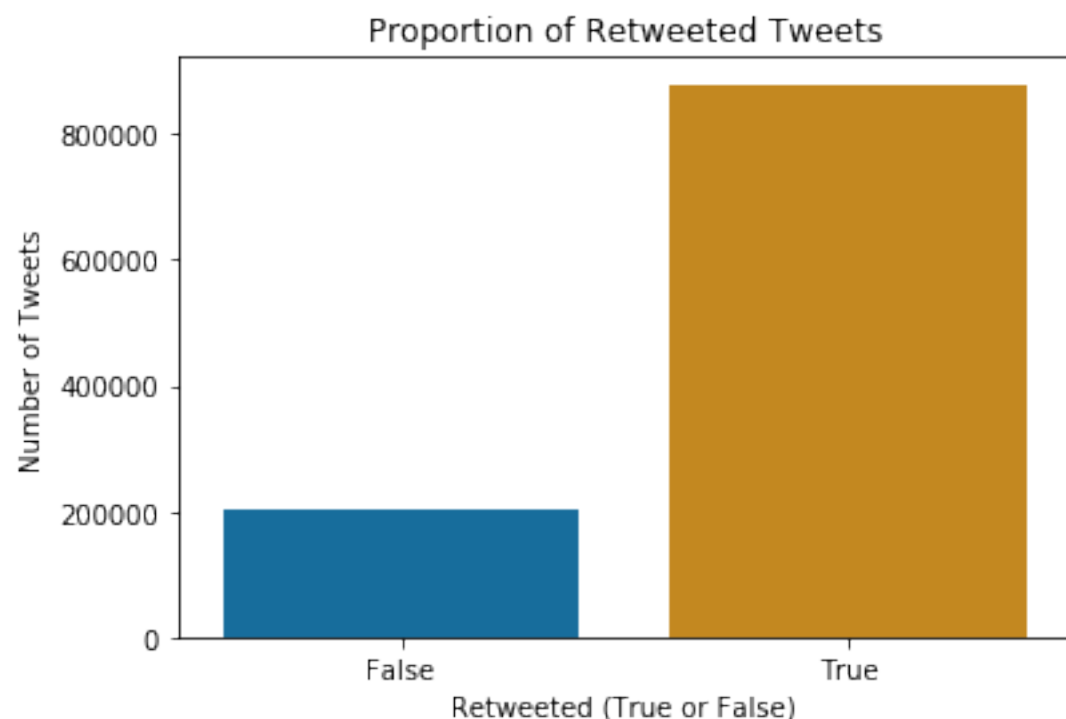
```
#how many retweets in the dataset
tweets3['is_retweet'].value_counts()
```

Out[400]:

```
True      879070
False     204157
Name: is_retweet, dtype: int64
```

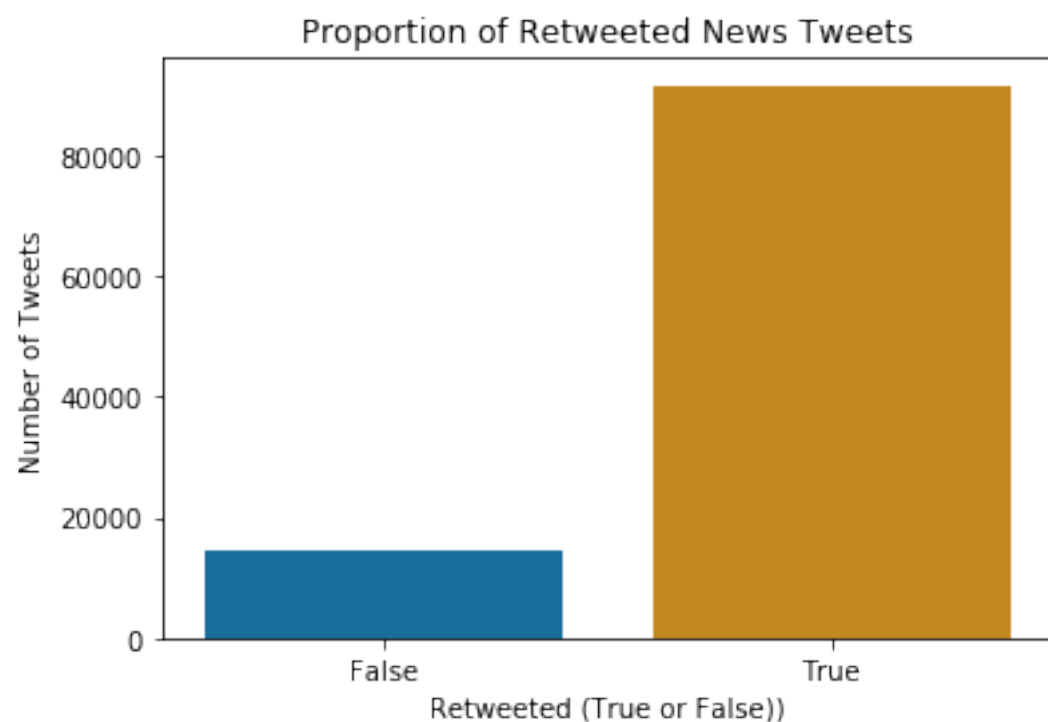
In [401]:

```
ax = sns.countplot(x='is_retweet', data= tweets3, palette = 'colorblind')
ax.set(xlabel = 'Retweeted (True or False)', ylabel = 'Number of Tweets')
ax = plt.title("Proportion of Retweeted Tweets")
plt.show()
```



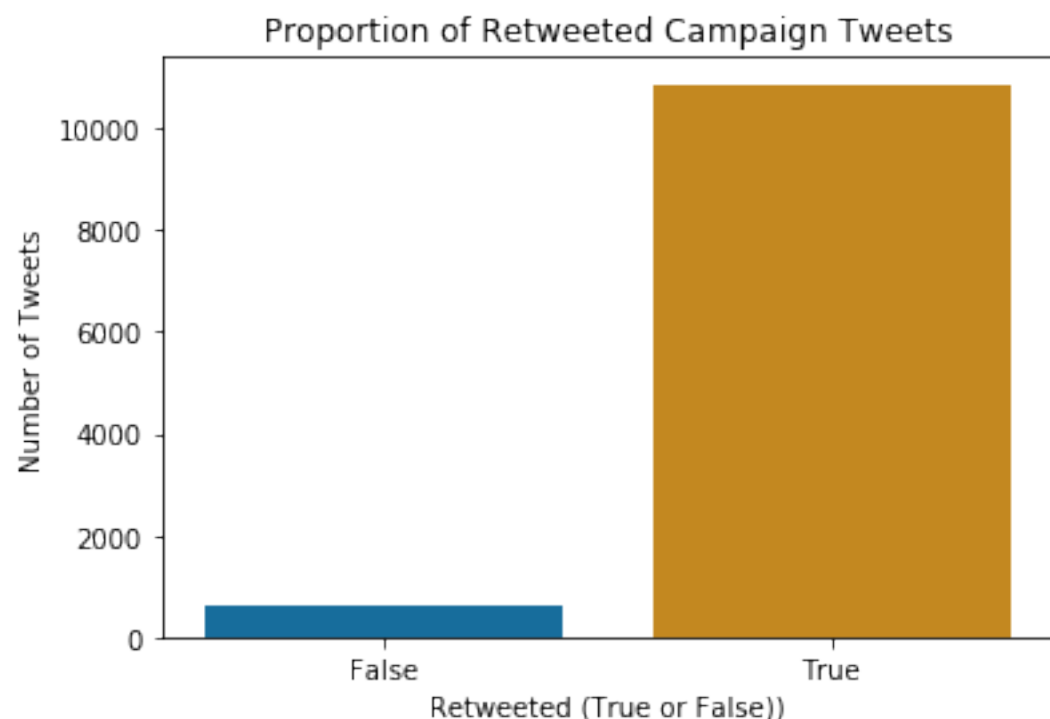
In [402]:

```
ax = sns.countplot(x='is_retweet', data= news_data, palette = 'colorblind')
ax.set(xlabel = 'Retweeted (True or False))', ylabel = 'Number of Tweets')
ax = plt.title("Proportion of Retweeted News Tweets")
```



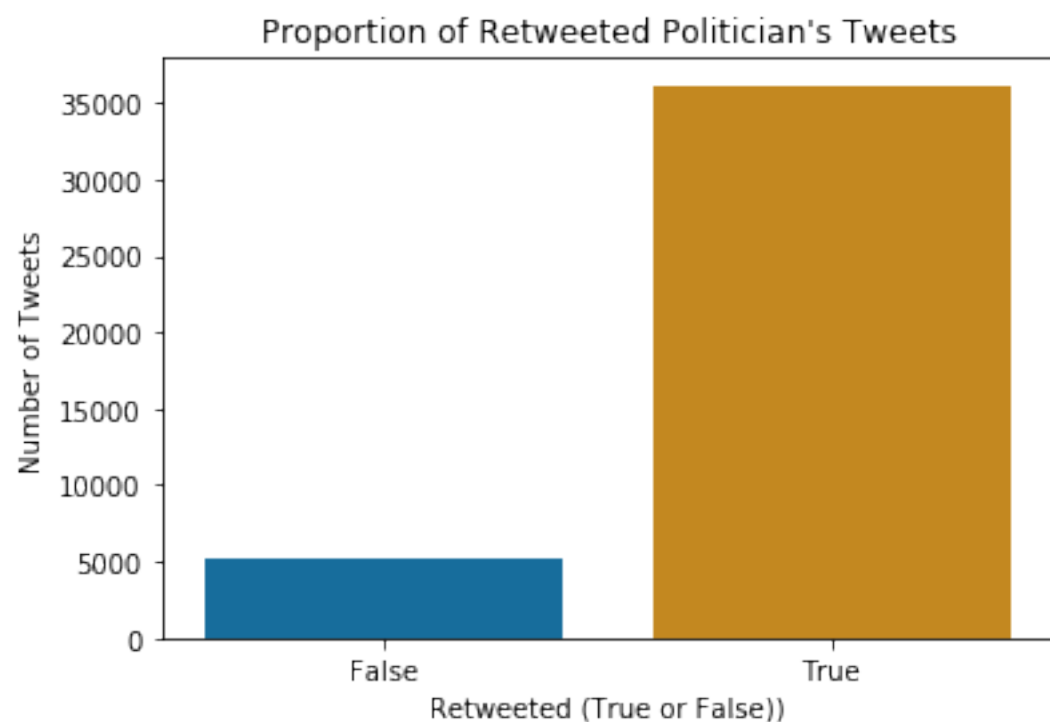
In [403]:

```
ax = sns.countplot(x='is_retweet', data= campaign_data, palette = 'colorblind')
ax.set(xlabel = 'Retweeted (True or False)', ylabel = 'Number of Tweets')
ax = plt.title("Proportion of Retweeted Campaign Tweets")
```



In [404]:

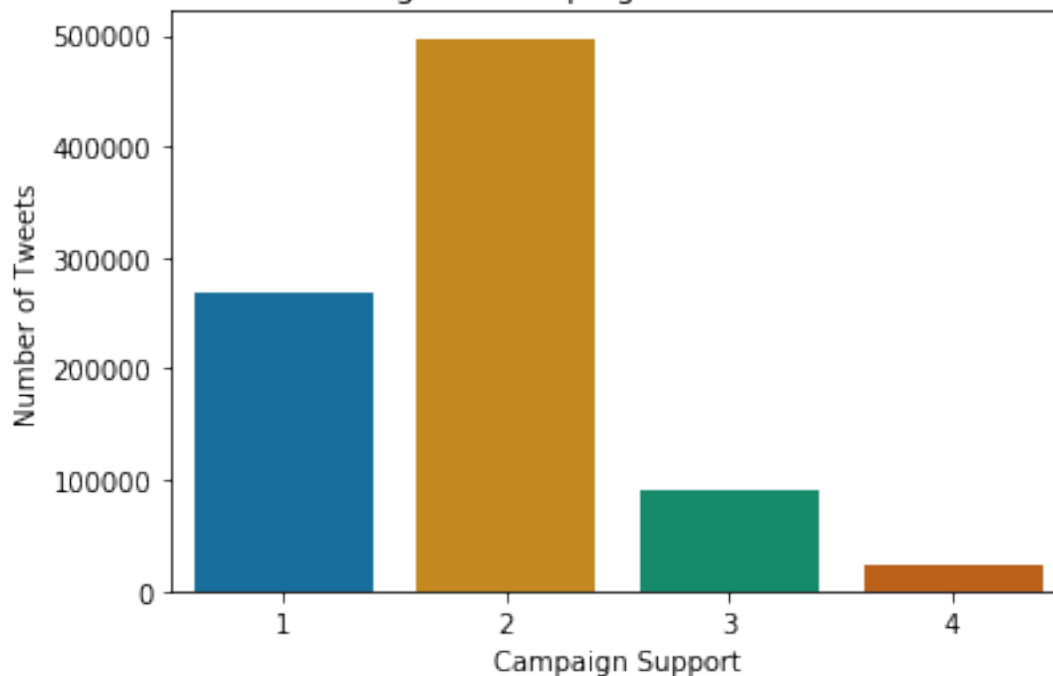
```
ax = sns.countplot(x='is_retweet', data= pol_data, palette = 'colorblind')
ax.set(xlabel = 'Retweeted (True or False)', ylabel = 'Number of Tweets')
ax = plt.title("Proportion of Retweeted Politician's Tweets")
```



In [412]:

```
ax2 = sns.countplot(x='support', data= tweets4, palette = 'colorblind')
ax2.set(xlabel = 'Campaign Support', ylabel = 'Number of Tweets')
ax2 = plt.title("Proportion of tweets endorsing Yes Campaign (1 = Neither, 2 = Yes, 3 = No, 4 = Both)")
```

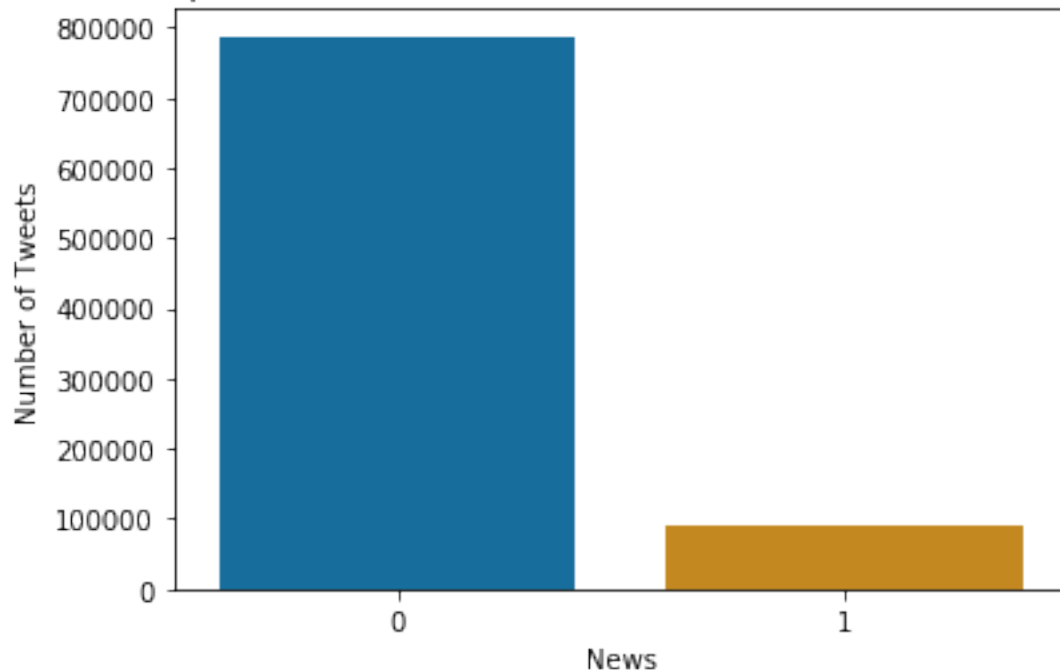
Proportion of tweets endorsing Yes Campaign (1 = Neither, 2 = Yes, 3 = No, 4 = Both)



In [406]:

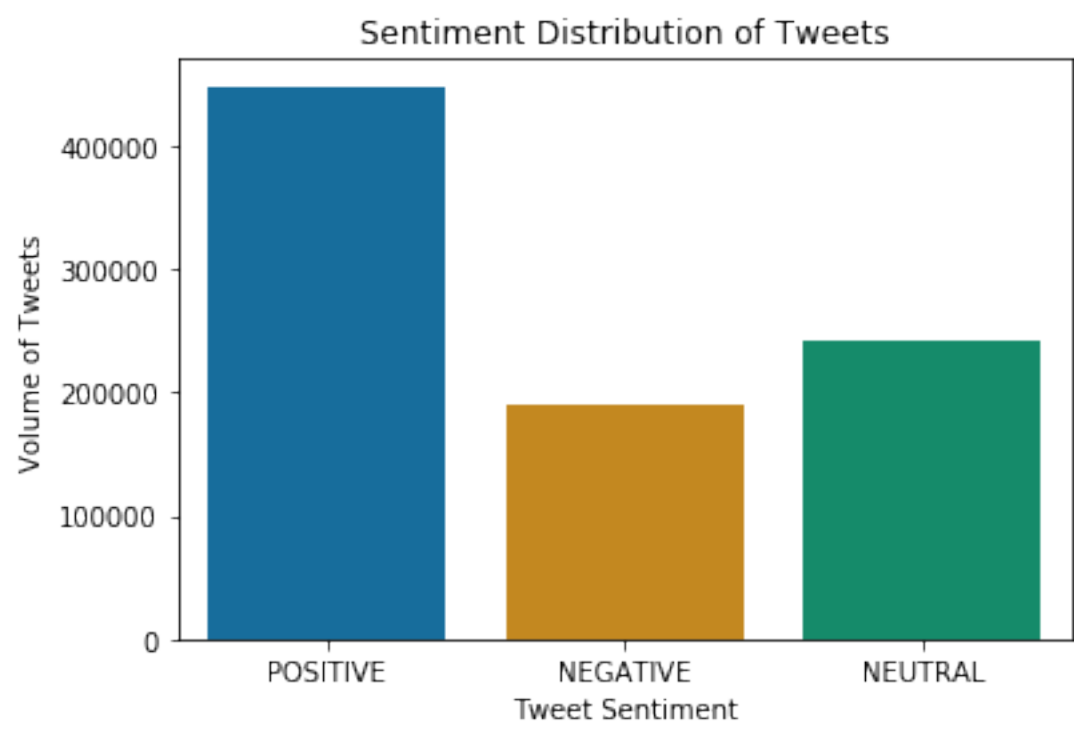
```
ax2 = sns.countplot(x='news_content', data= tweets4, palette = 'colorblind')
ax2.set(xlabel = 'News', ylabel = 'Number of Tweets')
ax2 = plt.title("Proportion of tweets that are news (0=non-news, 1 = news)")
```

Proportion of tweets that are news (0=non-news, 1 = news)



In [407]:

```
ax3 = sns.countplot(x='sentiment_type', data= tweets4, palette = 'colorblind')
ax3.set(xlabel = 'Tweet Sentiment', ylabel = 'Volume of Tweets')
ax3 = plt.title("Sentiment Distribution of Tweets")
```



In [408]:

```
dates = pd.DataFrame(tweets4['DATE'].value_counts().reset_index())
dates.columns = dates.columns.str.replace('index','Date') #Renaming two of the columns
dates.columns = dates.columns.str.replace('DATE','Frequency')
dates = dates.sort_values(by=['Date'])
dates
```

Out[408]:

	Date	Frequency
49	2018-04-13	2906
43	2018-04-14	4879
44	2018-04-15	4788
45	2018-04-16	4382
41	2018-04-17	5229
36	2018-04-18	6515
38	2018-04-19	6101
32	2018-04-20	7240
34	2018-04-21	6654
40	2018-04-22	6017
27	2018-04-23	8014
19	2018-04-24	9396

25	2018-04-25	8636
35	2018-04-26	6607
13	2018-04-27	12535
26	2018-04-28	8038
39	2018-04-29	6067
42	2018-04-30	5167
31	2018-05-01	7248
28	2018-05-02	7848
29	2018-05-03	7801
30	2018-05-04	7262
33	2018-05-05	6765
24	2018-05-06	8709
22	2018-05-07	9228
23	2018-05-08	8818
16	2018-05-09	10823
17	2018-05-10	10191
20	2018-05-11	9390
15	2018-05-12	11391
21	2018-05-13	9236
7	2018-05-14	16235
9	2018-05-15	14827
14	2018-05-16	12007
8	2018-05-17	15756
12	2018-05-18	13375
11	2018-05-19	13515
10	2018-05-20	14506
6	2018-05-21	19689
4	2018-05-22	29205
3	2018-05-23	37054
2	2018-05-24	81487
0	2018-05-25	207616
1	2018-05-26	111747
5	2018-05-27	26475
18	2018-05-28	9465

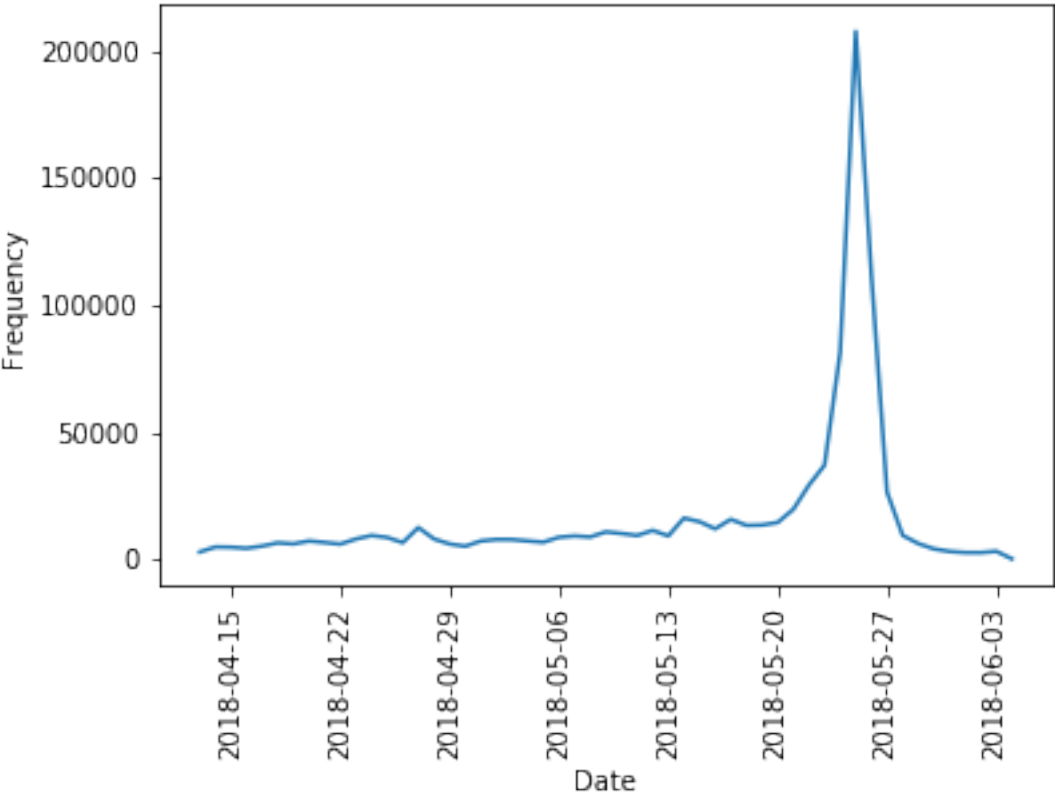
37	2018-05-29	6220
46	2018-05-30	4126
48	2018-05-31	3128
50	2018-06-01	2693
51	2018-06-02	2680
47	2018-06-03	3186
52	2018-06-04	197

In [409]:

```
volumeover_time = sns.lineplot(x=dates.Date, y=dates.Frequency)
#g.set_xticklabels(g.get_xticklabels(), rotation=90)
plt.xticks(rotation=90)
```

Out[409]:

(array([736799., 736806., 736813., 736820., 736827., 736834., 736841.,
736848.]), <a list of 8 Text xticklabel objects>)



In [410]:

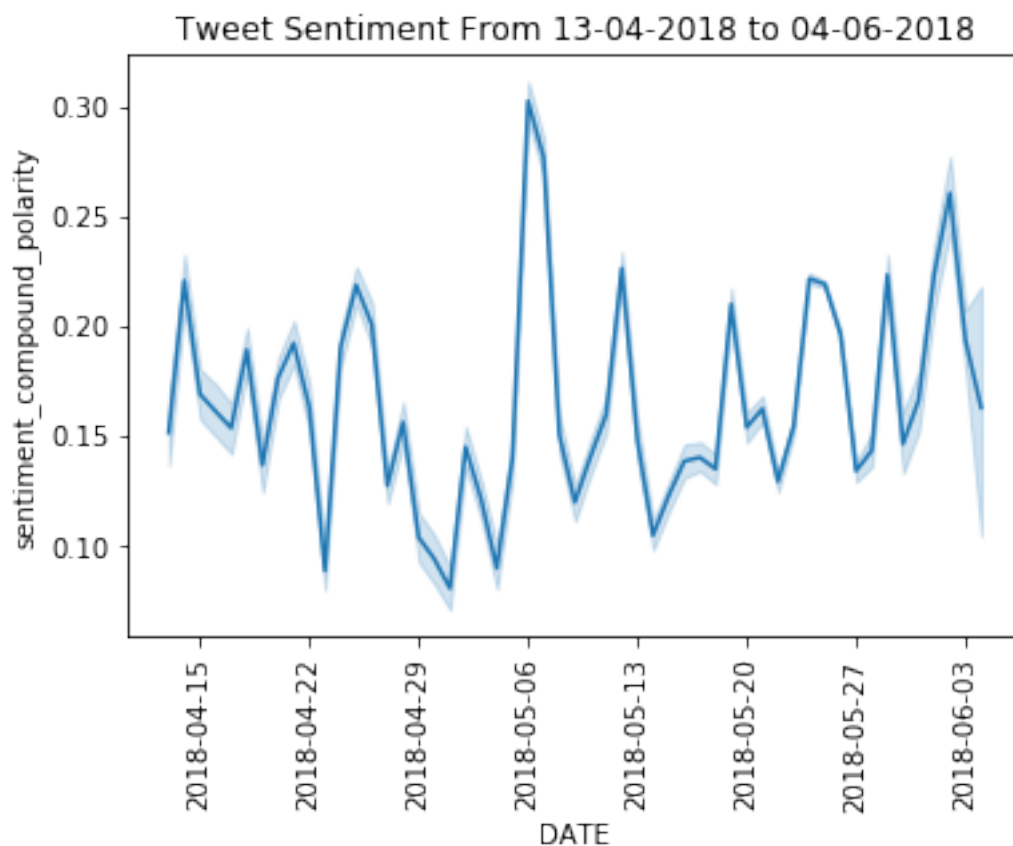
```
#Show the sentiment distribution based on sentiment_compound_polarity OVER TIME
H = sns.lineplot(x=tweets3.DATE, y=tweets3.sentiment_compound_polarity, data=tweets3)
plt.xticks(rotation=90)
H.set_title('Tweet Sentiment From 13-04-2018 to 04-06-2018')
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

Out[410]:

Text(0.5,1,'Tweet Sentiment From 13-04-2018 to 04-06-2018')



In [424]:

```
palette = sns.color_palette("mako_r", 4)
```

In [425]:

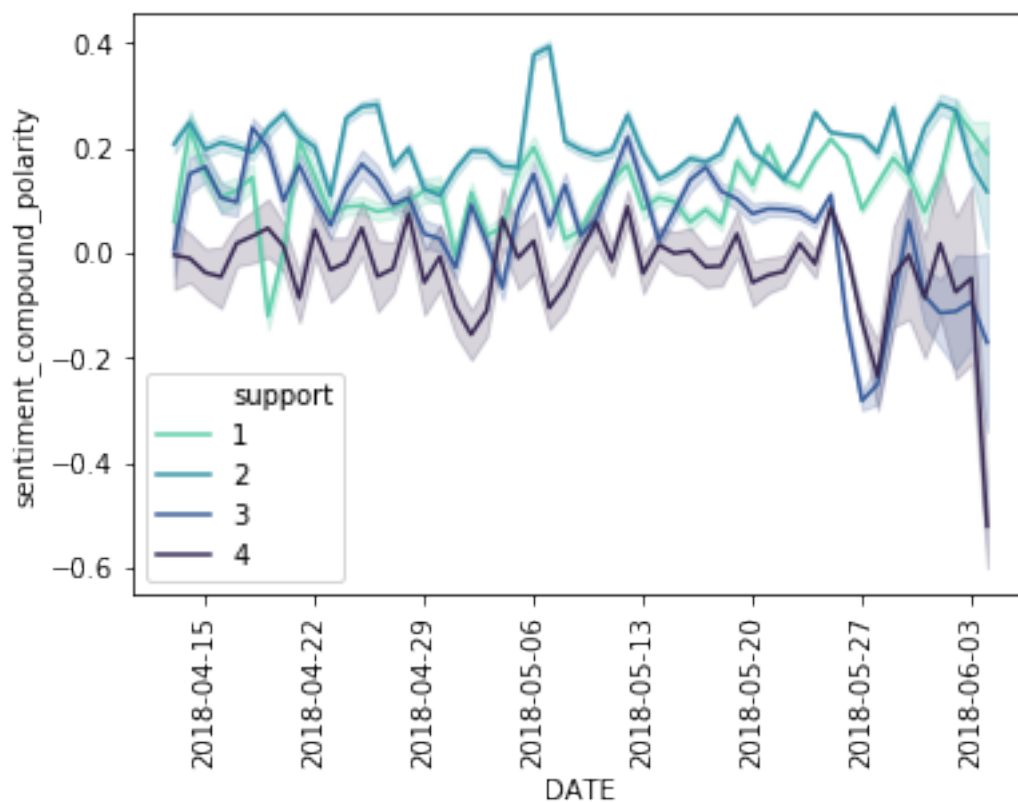
```
g = sns.lineplot(x=tweets3.DATE, y=tweets3.sentiment_compound_polarity, hue =
tweets3.support, palette = palette)
plt.xticks(rotation=90)
H.set_title('Tweet Sentiment From 13-04-2018 to 04-06-2018 by Support')
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

Out[425]:

Text(0.5,1,'Tweet Sentiment From 13-04-2018 to 04-06-2018 by Support')



In [426]:

```
palette = sns.color_palette("mako_r", 2)
```

In [428]:

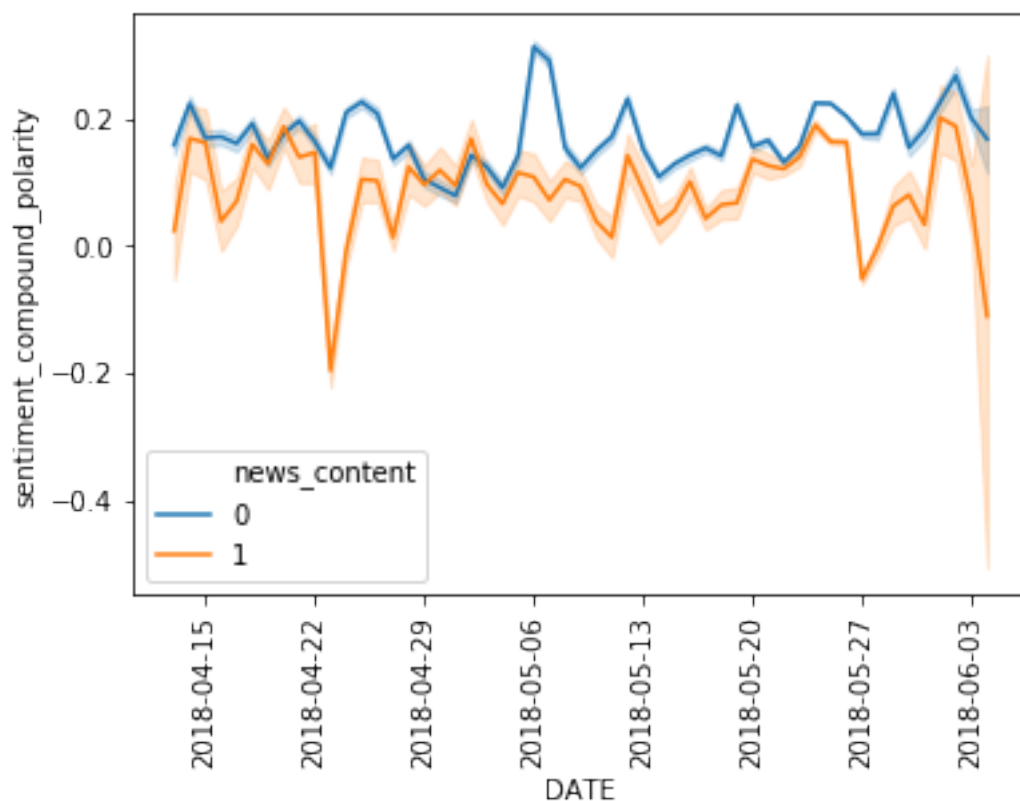
```
gn = sns.lineplot(x=tweets3.DATE, y=tweets3.sentiment_compound_polarity, hue =
tweets3.news_content)
plt.xticks(rotation=90)
H.set_title('Tweet Sentiment From 13-04-2018 to 04-06-2018 by News')
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

Out[428]:

Text(0.5,1,'Tweet Sentiment From 13-04-2018 to 04-06-2018 by News')
)



Network of Campaign Retweets

In [593]:

```
tweets3['screen_nc'].value_counts()
```

Out[593]:

```
0    1078699
2      4528
Name: screen_nc, dtype: int64
```

In [599]:

```
tweets3['camp_total'] = (tweets3.screen_nc + tweets.screen_yc)
no = tweets3[tweets3['screen_nc']==2]
tweets3['camp_total'].value_counts()
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

Out[599]:

```
0    1067201
1     11498
2      4528
Name: camp_total, dtype: int64
```

In [601]:

```
campaign_data = tweets3[tweets3['camp_total']>=1]
```

In [606]:

```
campaign_data['camp_total'].value_counts()
```

Out[606]:

```
1     11498
2      4528
Name: camp_total, dtype: int64
```

In [607]:

```
retweets = campaign_data[campaign_data['is_retweet']]
retweets['original_twitter'] = retweets['text'].str.extract('RT @([a-zA-Z0-9]\w{0,}):', expand=True)
retweets.columns
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

Out[607]:

```
Index(['DATE', 'TIME', 'hashtags', 'urls', 'contains_url', 'favorite_count',
      'is_retweet', 'retweet_count', 'Logged_RT_count', 'retweet_screen_name',
      'retweet_yc', 'retweet_nc', 'rtpol', 'RT', 'text', 'text_news',
      'sentiment_compound_polarity', 'sentiment_type', 'Yes_VOTE', 'No_VOTE',
      'news_content', 'Yes_camp', 'No_camp', 'poltweet', 'friend_foll_ratio',
      'user_statuses_count', 'user_location', 'Screenname', 'user_description',
      'user_favourites_count', 'user_followers_count', 'user_friends_count',
      'screen_yc', 'screen_nc', 'scpol', 'news', 'campaign_support', 'support',
      'camp_total', 'original_twitter'],
      dtype='object')
```

In [622]:

```
edges = retweets[['Screenname', 'original_twitter', 'DATE']]
edges.columns = ['Source', 'Target', 'Strength']
```

In [623]:

```
strengthLevel = 1 # Network connection strength level: the number of times in
total each of the tweeters RTed the other
# If you have 1 as the level, then all tweeters who mentioned or replied to an
other at least once will be displayed. But if you have 5, only those who have
mentioned or responded to a particular tweeter at least 5 times will be displayed,
which means that only the strongest bonds are shown.
edges2 = edges.groupby(['Source', 'Target'])['Strength'].count()
edges2 = edges2.reset_index()
edges2.head = edges2[edges2['Strength'] >= strengthLevel]
```

In [624]:

```
users = campaign_data[['Screenname','user_followers_count','screen_yc','screen_nc']]
users = users.sort_values(['Screenname','user_followers_count'], ascending=[True, False])
users = users.drop_duplicates(['Screenname'], keep='first')

ids = edges2['Source'].append(edges2['Target']).to_frame()
ids['Label'] = ids
ids.columns = ['Screenname', 'Label']
ids = ids.drop_duplicates(['Screenname'], keep='first')
nodes = pd.merge(ids, users, on='Screenname', how='left')
```

In [625]:

```
print(nodes.shape) #user-based data
print(edges2.shape)#tweet-based data
#(213390, 7)
#(558796, 3)

(2065, 5)
(4111, 3)
```

In [628]:

```
edges2.head
```

Out[628]:

	Source	Target	Strength
0	1familyireland	AndyPbridge	1
1	1familyireland	Doctors4Choice	1
2	1familyireland	HSPodNetwork	1
3	1familyireland	IrishFPA	1
4	1familyireland	Longford4Repeal	1
5	1familyireland	SenLynnRuane	1
6	1familyireland	SparkIreland	3
7	1familyireland	Together4yes	3
8	1familyireland	dublinwellwoman	2
9	1familyireland	jenoconnell	1
10	1familyireland	nycinews	1
11	AIMSIreland	ARCDonegal	1
12	AIMSIreland	Aidan_Regan	1
13	AIMSIreland	AilbheConneely	1
14	AIMSIreland	ArtimusFoul	1

15	AIMSIreland	BrennanGaye	1
16	AIMSIreland	BrianLeonardFR	1
17	AIMSIreland	Catholics4yes	1
18	AIMSIreland	Ciara_OC	1
19	AIMSIreland	Colmogorman	1
20	AIMSIreland	CorkTogether4Y	1
21	AIMSIreland	DLTogether4Yes	1
22	AIMSIreland	DroghedaT4Y	36
23	AIMSIreland	FireyMrs	1
24	AIMSIreland	HazelKLarkin	1
25	AIMSIreland	InHerIrishShoes	1
26	AIMSIreland	JaneMahony	1
27	AIMSIreland	KilkennyForYes	1
28	AIMSIreland	Marytcurtin	4
29	AIMSIreland	MeghanJRice	3
...
4081	thexileproject	mummakeupbag	2
4082	thexileproject	newschambers	1
4083	thexileproject	pb4p	1
4084	thexileproject	philipbromwell	1
4085	thexileproject	rlbowie	1
4086	thexileproject	roemcdermott	1
4087	thexileproject	roisiningle	3
4088	thexileproject	rte	1
4089	thexileproject	ruairimckiernan	1
4090	thexileproject	thexileproject	7
4091	thexileproject	unapower	1
4092	thexileproject	veganperiod	1
4093	westcork4choice	CorkTogether4Y	2
4094	westcork4choice	FintanMcKahey	1
4095	westcork4choice	JJ5000	1
4096	westcork4choice	Jenrok06	1
4097	westcork4choice	Katebyrne7	2

4098	westcork4choice	LdnIrishARC	1
4099	westcork4choice	Men4Yes	1
4100	westcork4choice	Rebels4Choice	1
4101	westcork4choice	Together4yes	1
4102	westcork4choice	TrishDeseine	1
4103	westcork4choice	ailbhes	1
4104	westcork4choice	dararyder	1
4105	westcork4choice	disabled_choice	1
4106	westcork4choice	famron	1
4107	westcork4choice	gavreilly	1
4108	westcork4choice	lianbell	1
4109	westcork4choice	oneilllo	2
4110	westcork4choice	parentsforYES	1

4111 rows × 3 columns

In [618]:

```
nodes.head()
```

Out[618]:

	Screenname	Label	user_followers_count	screen_yc	screen_nc
0	1familyireland	1familyireland	2741.0	1.0	0.0
1	AIMSIreland	AIMSIreland	3975.0	1.0	0.0
2	ARCDonegal	ARCDonegal	1381.0	1.0	0.0
3	ARNirl	ARNirl	3271.0	1.0	0.0
4	All4Choice	All4Choice	8433.0	1.0	0.0

In [619]:

```
nodes.to_csv('campaign_tweetsnodes.csv', encoding='utf-8', index=False)
edges2.to_csv('campaign_tweets.csv', encoding='utf-8', index=False)
```

Network of Politicians

In [629]:

```
pol_data = tweets3[tweets3['scpol']==1]
pol_data.columns
```

Out[629]:

```
Index(['DATE', 'TIME', 'hashtags', 'urls', 'contains_url', 'favorite_count',
      'is_retweet', 'retweet_count', 'Logged_RT_count', 'retweet_screen_name',
      'retweet_yc', 'retweet_nc', 'rtpol', 'RT', 'text', 'text_news',
      'sentiment_compound_polarity', 'sentiment_type', 'Yes_VOTE', 'No_VOTE',
      'news_content', 'Yes_camp', 'No_camp', 'poltweet', 'friend_foll_ratio',
      'user_statuses_count', 'user_location', 'Screenname', 'user_description',
      'user_favourites_count', 'user_followers_count', 'user_friends_count',
      'screen_yc', 'screen_nc', 'scpol', 'news', 'campaign_support', 'support',
      'camp_total'],
      dtype='object')
```

In [630]:

```
retweets = pol_data[pol_data['is_retweet']]
retweets['original_twitter'] = retweets['text'].str.extract('RT @([a-zA-Z0-9]\w{0,}):', expand=True)
retweets.columns
```

```
/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
```

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

Out[630]:

```
Index(['DATE', 'TIME', 'hashtags', 'urls', 'contains_url', 'favorite_count',
      'is_retweet', 'retweet_count', 'Logged_RT_count', 'retweet_screen_name',
      'retweet_yc', 'retweet_nc', 'rtpol', 'RT', 'text', 'text_news',
      'sentiment_compound_polarity', 'sentiment_type', 'Yes_VOTE', 'No_VOTE',
      'news_content', 'Yes_camp', 'No_camp', 'poltweet', 'friend_foll_ratio',
      'user_statuses_count', 'user_location', 'Screenname', 'user_description',
      'user_favourites_count', 'user_followers_count', 'user_friends_count',
      'screen_yc', 'screen_nc', 'scpol', 'news', 'campaign_support', 'support',
      'camp_total', 'original_twitter'],
      dtype='object')
```

In [631]:

```
edges = retweets[['Screenname', 'original_twitter', 'DATE']]
edges.columns = ['Source', 'Target', 'Strength']
```

In [632]:

```
strengthLevel = 1 # Network connection strength level: the number of times in total each of the tweeters RTed the other
# If you have 1 as the level, then all tweeters who mentioned or replied to an other at least once will be displayed. But if you have 5, only those who have mentioned or responded to a particular tweeter at least 5 times will be displayed, which means that only the strongest bonds are shown.
edges2 = edges.groupby(['Source', 'Target'])['Strength'].count()
edges2 = edges2.reset_index()
edges2.head = edges2[edges2['Strength'] >= strengthLevel]
```

In [633]:

```
users = pol_data[['Screenname','user_followers_count','support','news_content',
'sentiment_type']]
users = users.sort_values(['Screenname','user_followers_count'], ascending=[True, False])
users = users.drop_duplicates(['Screenname'], keep='first')

ids = edges2['Source'].append(edges2['Target']).to_frame()
ids['Label'] = ids
ids.columns = ['Screenname', 'Label']
ids = ids.drop_duplicates(['Screenname'], keep='first')
nodes = pd.merge(ids, users, on='Screenname', how='left')
```

In [634]:

```
print(nodes.shape) #user-based data
print(edges2.shape)#tweet-based data
#(213390, 7)
#(558796, 3)
```

(571, 6)
(854, 3)

In [635]:

```
edges2.head
```

Out[635]:

	Source	Target	Strength
0	BrendanHowlin	BrianODriscoll	1
1	BrendanHowlin	BridinM	1
2	BrendanHowlin	DawnButlerBrent	1
3	BrendanHowlin	DrMarkMurphy	1
4	BrendanHowlin	Together4yes	1
5	BrendanHowlin	ivanabacik	1
6	BrendanHowlin	labour	14
7	CathMurphyTD	1GaryGannon	2
8	CathMurphyTD	1KenCampbell	3
9	CathMurphyTD	AnMailleach	1
10	CathMurphyTD	AoifeM1601	1
11	CathMurphyTD	CarlyBee25	1
12	CathMurphyTD	Ciaraioch	1
13	CathMurphyTD	Claire_McGing	1

14	CathMurphyTD	ColetteKelleher	1
15	CathMurphyTD	Colmogorman	1
16	CathMurphyTD	CorkSocDems	1
17	CathMurphyTD	DelsolClaire	1
18	CathMurphyTD	DevinToner4	1
19	CathMurphyTD	GalwayTFY	1
20	CathMurphyTD	Ian_madigan	1
21	CathMurphyTD	KathleenCHarris	1
22	CathMurphyTD	Kildare4Repeal	1
23	CathMurphyTD	KilkennyForYes	1
24	CathMurphyTD	LandersIE	1
25	CathMurphyTD	MartinaFitzg	1
26	CathMurphyTD	MaynoothForYes	2
27	CathMurphyTD	MeathForChoice	1
28	CathMurphyTD	NWtogether4YES	1
29	CathMurphyTD	NaasForRepeal	2
...
824	sinnfeinireland	BBCTalkback	1
825	sinnfeinireland	CArchibald_SF	1
826	sinnfeinireland	CMcGin2006	1
827	sinnfeinireland	CorkSinnFein	1
828	sinnfeinireland	DMitchellTD	1
829	sinnfeinireland	DeclanKearneySF	1
830	sinnfeinireland	GerryAdamsSF	1
831	sinnfeinireland	MaryLouMcDonald	4
832	sinnfeinireland	PearseDoherty	1
833	sinnfeinireland	ROMurchu	1
834	sinnfeinireland	RonanLowry	1
835	sinnfeinireland	c1aranquinn	1
836	sinnfeinireland	johnbradysf	1
837	sinnfeinireland	loreillysf	1
838	sinnfeinireland	moneillsf	1
839	sinnfeinireland	pdquinn7	1
840	solidarityie	MadMarx__	1

841	solidarityie	Rita702	1
842	solidarityie	RuthCoppingerTD	3
843	solidarityie	cllrmobrien	1
844	solidarityie	paulmurphy_TD	2
845	solidarityie	pb4p	1
846	workersparty	CathieDoherty85	1
847	workersparty	DavidGardinerWP	1
848	workersparty	EoghanGardiner	4
849	workersparty	GavinMGleason	1
850	workersparty	adammccarthy	1
851	workersparty	eilistweets	10
852	workersparty	gemmacweir1	1
853	workersparty	ownohmanny	2

854 rows × 3 columns

In [636]:

```
nodes.head(1000)
```

Out[636]:

	Screenname	Label	user_followers_count	support	news_c
0	BrendanHowlin	BrendanHowlin	3672.0	2.0	1.0
1	CathMurphyTD	CathMurphyTD	17675.0	2.0	0.0
2	EamonRyan	EamonRyan	21285.0	2.0	0.0
3	FineGaelBrux	FineGaelBrux	173.0	1.0	1.0
4	JoeHigginsSP	JoeHigginsSP	20310.0	2.0	0.0
5	MaryLouMcDonald	MaryLouMcDonald	51195.0	2.0	0.0
6	ModowdMichael	ModowdMichael	464.0	3.0	0.0
7	RENUAIreland	RENUAIreland	5405.0	1.0	0.0
8	RoisinShortall	RoisinShortall	11784.0	2.0	1.0
9	SocDems	SocDems	13897.0	2.0	0.0
10	SocDemsDBN	SocDemsDBN	756.0	2.0	0.0
11	SocDemsDBS	SocDemsDBS	370.0	1.0	0.0
12	SocDemsDL	SocDemsDL	432.0	2.0	0.0

13	SocDemsDubCen	SocDemsDubCen	725.0	2.0	0.0
14	SocDemsDublinSW	SocDemsDublinSW	280.0	2.0	0.0
15	SocDemsFingal	SocDemsFingal	2171.0	2.0	1.0
16	SocDemsGEast	SocDemsGEast	459.0	2.0	0.0
17	SocDemsGalway	SocDemsGalway	1333.0	2.0	0.0
18	SocDemsKN	SocDemsKN	547.0	2.0	0.0
19	SocDemsLK	SocDemsLK	467.0	2.0	0.0
20	SocDemsLW	SocDemsLW	307.0	2.0	0.0
21	SocDemsLouth	SocDemsLouth	294.0	1.0	1.0
22	SocDemsRosGal	SocDemsRosGal	342.0	1.0	0.0
23	SocDemsWX	SocDemsWX	278.0	2.0	0.0
24	campaignforleo	campaignforleo	167523.0	2.0	0.0
25	cathmartingreen	cathmartingreen	5403.0	2.0	0.0
26	ddi	ddi	1177.0	2.0	0.0
27	fiannafailparty	fiannafailparty	30732.0	1.0	1.0
28	greenparty_ie	greenparty_ie	21779.0	2.0	0.0
29	irelandcp	irelandcp	4747.0	1.0	0.0
...
541	PBPproChoice	PBPproChoice	NaN	NaN	NaN
542	RBoydBarrett	RBoydBarrett	NaN	NaN	NaN
543	alternaTiVeIRE	alternaTiVeIRE	NaN	NaN	NaN
544	bridsmithTD	bridsmithTD	NaN	NaN	NaN
545	edfitzgeraldpbp	edfitzgeraldpbp	NaN	NaN	NaN
546	madeleineswp	madeleineswp	NaN	NaN	NaN
547	maevomc	maevomc	NaN	NaN	NaN
548	pbp_gillian	pbp_gillian	NaN	NaN	NaN
549	DeirdreCluneMEP	DeirdreCluneMEP	NaN	NaN	NaN
550	FineGael	FineGael	31673.0	1.0	1.0
551	ArnaldoOtegi	ArnaldoOtegi	NaN	NaN	NaN
552	BBCTalkback	BBCTalkback	NaN	NaN	NaN
553	CArchibald_SF	CArchibald_SF	NaN	NaN	NaN
554	CMcGin2006	CMcGin2006	NaN	NaN	NaN
555	CorkSinnFein	CorkSinnFein	NaN	NaN	NaN
556	DeclanKearneySF	DeclanKearneySF	NaN	NaN	NaN

557	GerryAdamsSF	GerryAdamsSF	NaN	NaN	NaN
558	ROMurchu	ROMurchu	NaN	NaN	NaN
559	pdquinn7	pdquinn7	NaN	NaN	NaN
560	MadMarx__	MadMarx__	NaN	NaN	NaN
561	Rita702	Rita702	NaN	NaN	NaN
562	cllrmobrien	cllrmobrien	NaN	NaN	NaN
563	CathieDoherty85	CathieDoherty85	NaN	NaN	NaN
564	DavidGardinerWP	DavidGardinerWP	NaN	NaN	NaN
565	EoghanGardiner	EoghanGardiner	NaN	NaN	NaN
566	GavinMGleason	GavinMGleason	NaN	NaN	NaN
567	adammccarthy	adammccarthy	NaN	NaN	NaN
568	eilistweets	eilistweets	NaN	NaN	NaN
569	gemmacweir1	gemmacweir1	NaN	NaN	NaN
570	ownohmanny	ownohmanny	NaN	NaN	NaN

571 rows × 6 columns

In [637]:

```
nodes.to_csv('poltweetnodes.csv', encoding='utf-8', index=False)
edges2.to_csv('poledges.csv', encoding='utf-8', index=False)
```

News Network

In [638]:

```
users_news = tweets3[tweets3['news']==1]
users_news.shape
```

Out[638]:

(6183, 39)

In [639]:

```
retweets = users_news[users_news['is_retweet']]
retweets['original_twitter'] = retweets['text'].str.extract('RT @([a-zA-Z0-9]\w{0,}):', expand=True)
retweets.columns
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

Out[639]:

```
Index(['DATE', 'TIME', 'hashtags', 'urls', 'contains_url', 'favorite_count',
      'is_retweet', 'retweet_count', 'Logged_RT_count', 'retweet_screen_name',
      'retweet_yc', 'retweet_nc', 'rtpol', 'RT', 'text', 'text_news',
      'sentiment_compound_polarity', 'sentiment_type', 'Yes_VOTE', 'No_VOTE',
      'news_content', 'Yes_camp', 'No_camp', 'poltweet', 'friend_foll_ratio',
      'user_statuses_count', 'user_location', 'Screenname', 'user_description',
      'user_favourites_count', 'user_followers_count', 'user_friends_count',
      'screen_yc', 'screen_nc', 'scpol', 'news', 'campaign_support', 'support',
      'camp_total', 'original_twitter'],
      dtype='object')
```

In [640]:

```
edges = retweets[['Screenname', 'original_twitter', 'DATE']]
edges.columns = ['Source', 'Target', 'Strength']
```

In [641]:

```
strengthLevel = 1 # Network connection strength level: the number of times in
total each of the tweeters RTed the other
# If you have 1 as the level, then all tweeters who mentioned or replied to an
other at least once will be displayed. But if you have 5, only those who have
mentioned or responded to a particular tweeter at least 5 times will be displayed,
which means that only the strongest bonds are shown.
edges2 = edges.groupby(['Source', 'Target'])['Strength'].count()
edges2 = edges2.reset_index()
edges2.head = edges2[edges2['Strength'] >= strengthLevel]
```

In [642]:

```
users = users_news[['Screenname','user_followers_count','news','screen_yc','screen_nc','scpol']]
users = users.sort_values(['Screenname','user_followers_count'], ascending=[True, False])
users = users.drop_duplicates(['Screenname'], keep='first')

ids = edges2['Source'].append(edges2['Target']).to_frame()
ids['Label'] = ids
ids.columns = ['Screenname', 'Label']
ids = ids.drop_duplicates(['Screenname'], keep='first')
nodes = pd.merge(ids, users, on='Screenname', how='left')
```

In [643]:

```
print(nodes.shape) #user-based data
print(edges2.shape)#tweet-based data
#(213390, 7)
#(558796, 3)

(1240, 7)
(1577, 3)
```

In [644]:

```
edges2.head
```

Out[644]:

	Source	Target	Strength
0	1395FIVEaa	FIVEaaBreakfast	1
1	345triangle	MorabitoCM	1
2	3NewsIreland	ericgclarke	1
3	3NewsIreland	gavreilly	2
4	5_News	tessa5news	1
5	AMTomchak	gavreilly	1
6	AMTomchak	pollycurtis	1
7	AaronMehta	britainelects	1
8	AilbheConneely	Dyaneconnor	1
9	AilbheConneely	morningireland	1
10	AlbertoNardelli	NCPoliticsEU	1
11	AlineBarros2	nowthisnews	1
12	AlisonTodayFM	KiiimBuckley	1
13	AlisonTodayFM	SamanthaELong	1
14	AlisonTodayFM	gavreilly	1

15	AlisonTodayFM	rtenews	1
16	AlixFox	MorabitoCM	1
17	AllisonMorris1	DarranMarshall	1
18	AllisonMorris1	irish_news	1
19	AllisonMorris1	niallcarsonpa	1
20	AmritaCheemaTV	MorabitoCM	1
21	AmyKazmin	MorabitoCM	1
22	An_Phoblacht	SIPTU	1
23	An_Phoblacht	Together4yes	1
24	An_Phoblacht	joeslynch	1
25	An_Phoblacht	sinnfeinireland	4
26	AndrewSparrow	MarinaHyde	1
27	AndrewSparrow	jackcevans	1
28	AngelaThomps0n	parkinsbrea	1
29	AngusRobertson	MarkHennesy	1
...
1547	vanbadham	neetagreen	1
1548	vanbadham	qwikpix	1
1549	vanbadham	sheraciara	1
1550	vanbadham	tea_and_biccies	1
1551	vanbadham	vanbadham	1
1552	vcaivano	SkyNews	1
1553	versharma	irishexaminer	1
1554	versharma	jiatolentino	1
1555	victoriasanusi	RossalynWarren	1
1556	victoriasanusi	cp_conlan	1
1557	victoriasanusi	lindsayreports	1
1558	wearerabble	Repealshortfilm	1
1559	wearerabble	Taraustralis	1
1560	wearerabble	Together4yes	1
1561	wearerabble	dontstoprepeal	1
1562	wearerabble	hazelnolan	1
1563	wearerabble	ronronzo	1

1564	welt	StefanieBolzen	1
1565	yessibelloperez	BBCNews	1
1566	yodaberg	TrooDee	1
1567	yodaberg	matthewcarson	1
1568	yodaberg	sarahditum	1
1569	yukokato1701	MorabitoCM	1
1570	yukokato1701	NCPoliticsEU	1
1571	yukokato1701	newschambers	1
1572	yukokato1701	ryan_padraic	1
1573	yvonneredmond	AilbheConneely	1
1574	yvonneredmond	patomahony1	1
1575	yvonneredmond	stellarmagazine	1
1576	ziadramley	gavreilly	1

1577 rows × 3 columns

In [647]:

```
nodes.head(4)
```

Out[647]:

	Screenname	Label	user_followers_count	news	screen_yc	screen_nc	s
0	1395FIVEaa	1395FIVEaa	34441.0	1.0	0.0	0.0	(
1	345triangle	345triangle	7360.0	1.0	0.0	0.0	(
2	3NewsIreland	3NewsIreland	38562.0	1.0	0.0	0.0	(
3	5_News	5_News	20901.0	1.0	0.0	0.0	(

In [646]:

```
nodes.to_csv('newsnodes.csv', encoding='utf-8', index=False)
edges2.to_csv('newsedges.csv', encoding='utf-8', index=False)
```

Analysis of YES Supporters Tweets

In [429]:

```
tweets_yes.shape
```

Out[429]:

```
(497169, 38)
```

In [430]:

```
tweets_yes['news'].value_counts()
```

Out[430]:

```
0    495650
1     1519
Name: news, dtype: int64
```

In [431]:

```
users_yes = tweets_yes.drop_duplicates(subset='Screenname')
users_yes.shape
```

Out[431]:

```
(159036, 38)
```

In [432]:

```
Sentiment_tweets_yes = pd.DataFrame(tweets_yes['sentiment_type'].value_counts(
normalize=True).reset_index())
Sentiment_tweets_yes
```

Out[432]:

	index	sentiment_type
0	POSITIVE	0.529279
1	NEUTRAL	0.275492
2	NEGATIVE	0.195229

In [433]:

```
Sentiment_tweets_yes.columns = Sentiment_tweets_yes.columns.str.replace('index', 'Sentiment_Type')
Sentiment_tweets_yes.columns = Sentiment_tweets_yes.columns.str.replace('sentiment_type', 'Number_of_Tweets')
Sentiment_tweets_yes
```

Out[433]:

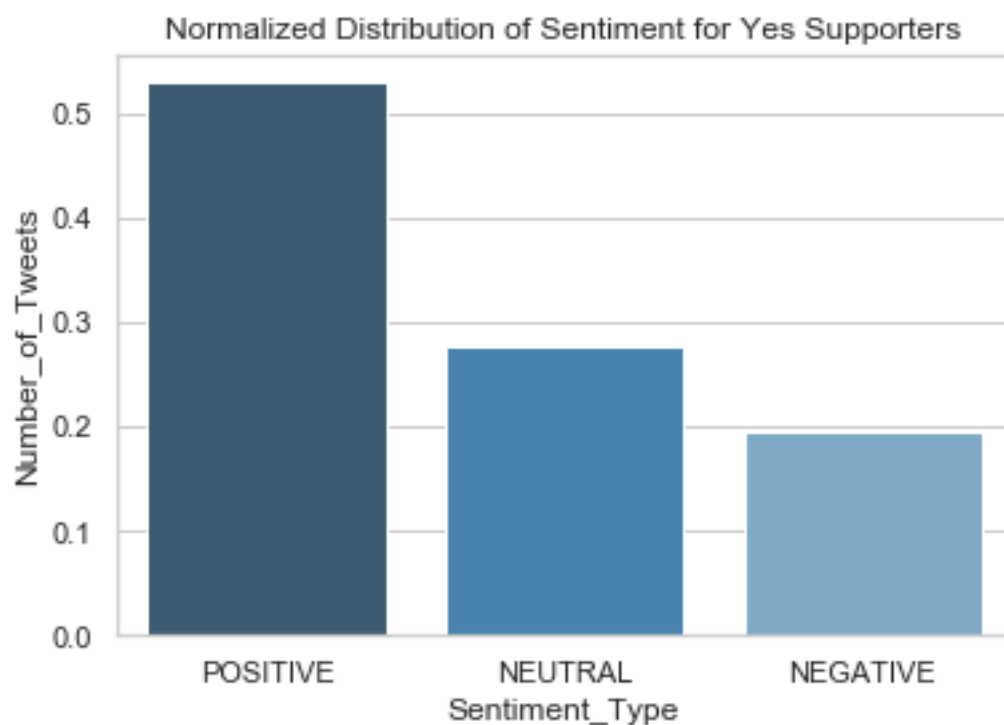
	Sentiment_Type	Number_of_Tweets
0	POSITIVE	0.529279
1	NEUTRAL	0.275492
2	NEGATIVE	0.195229

In [434]:

```
#Show the sentiment distribution based on sentiment_type
sns.set(style="whitegrid")
Sentiment_type_Y = sns.barplot(x=Sentiment_tweets_yes.Sentiment_Type, y=Sentiment_tweets_yes.Number_of_Tweets, palette="Blues_d")
Sentiment_type_Y.set_title('Normalized Distribution of Sentiment for Yes Supporters')
```

Out[434]:

```
Text(0.5,1,'Normalized Distribution of Sentiment for Yes Supporters')
```



In [435]:

```
#SUBSET OF TWEETS ('YES_VOTE') WHEN SENTIMENT ANALYSIS WAS DONE ON THE ENTIRE DATASET
All_T_YES = tweets[tweets['Yes_VOTE']]
```

In [436]:

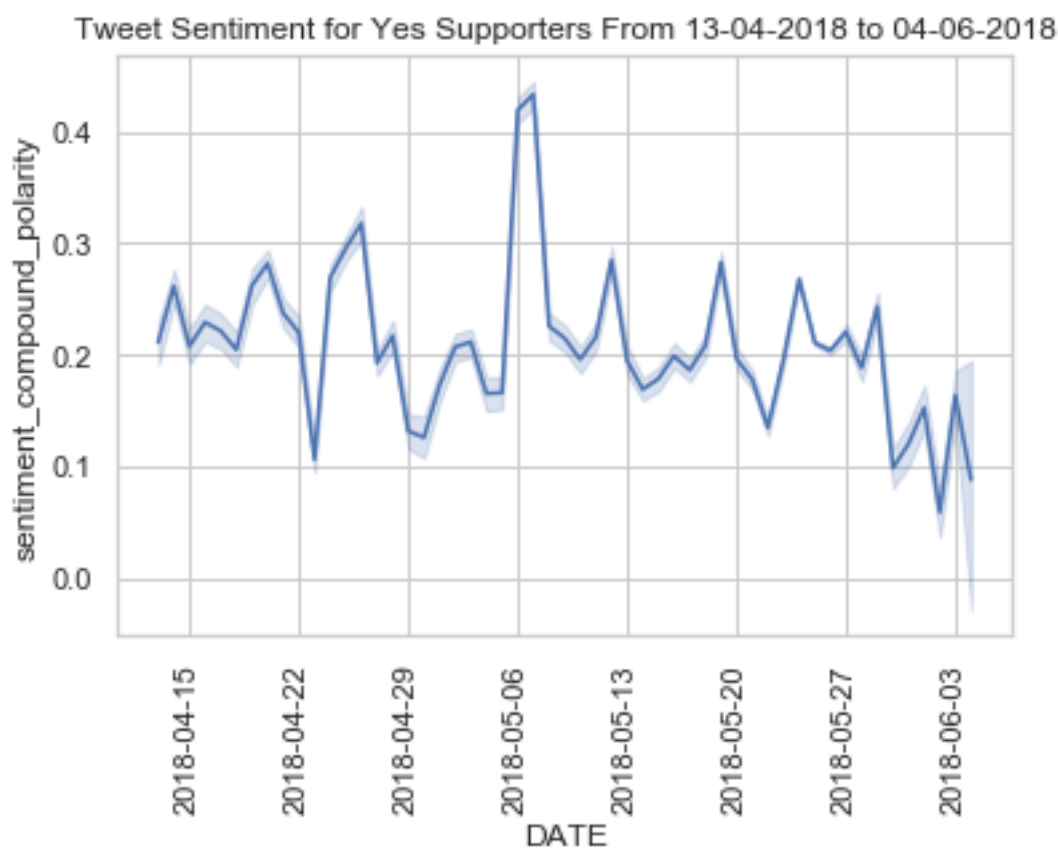
```
#Analysis of Sentiment compound polarity for yes supporters using the full dataset analysis
H = sns.lineplot(x=tweets_yes.DATE, y=tweets_yes.sentiment_compound_polarity, data=tweets_yes)
plt.xticks(rotation=90)
H.set_title('Tweet Sentiment for Yes Supporters From 13-04-2018 to 04-06-2018')
)
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

Out[436]:

Text(0.5,1,'Tweet Sentiment for Yes Supporters From 13-04-2018 to 04-06-2018')



In [437]:

```
tweets_yes['sentiment_type'].value_counts()
```

Out[437]:

```
POSITIVE    263141
NEUTRAL     136966
NEGATIVE     97062
Name: sentiment_type, dtype: int64
```

Analysis of No Supporters

In [438]:

```
tweets_no.shape
```

Out[438]:

```
(90699, 38)
```

In [439]:

```
tweets_no['news'].value_counts()
```

Out[439]:

```
0    90659
```

```
1     40
```

```
Name: news, dtype: int64
```

In [440]:

```
users_no = tweets_no.drop_duplicates(subset='Screenname')
users_no.shape
```

Out[440]:

```
(15795, 38)
```

In [441]:

```
Sentiment_tweets_no = pd.DataFrame(tweets_no['sentiment_type'].value_counts(normalize=True).reset_index())
Sentiment_tweets_no
```

Out[441]:

	index	sentiment_type
0	POSITIVE	0.431692
1	NEGATIVE	0.307699
2	NEUTRAL	0.260609

In [442]:

```
Sentiment_tweets_no.columns = Sentiment_tweets_no.columns.str.replace('index',  
'Sentiment_Type')  
Sentiment_tweets_no.columns = Sentiment_tweets_no.columns.str.replace('sentime  
nt_type', 'Number_of_Tweets')  
Sentiment_tweets_no
```

Out[442]:

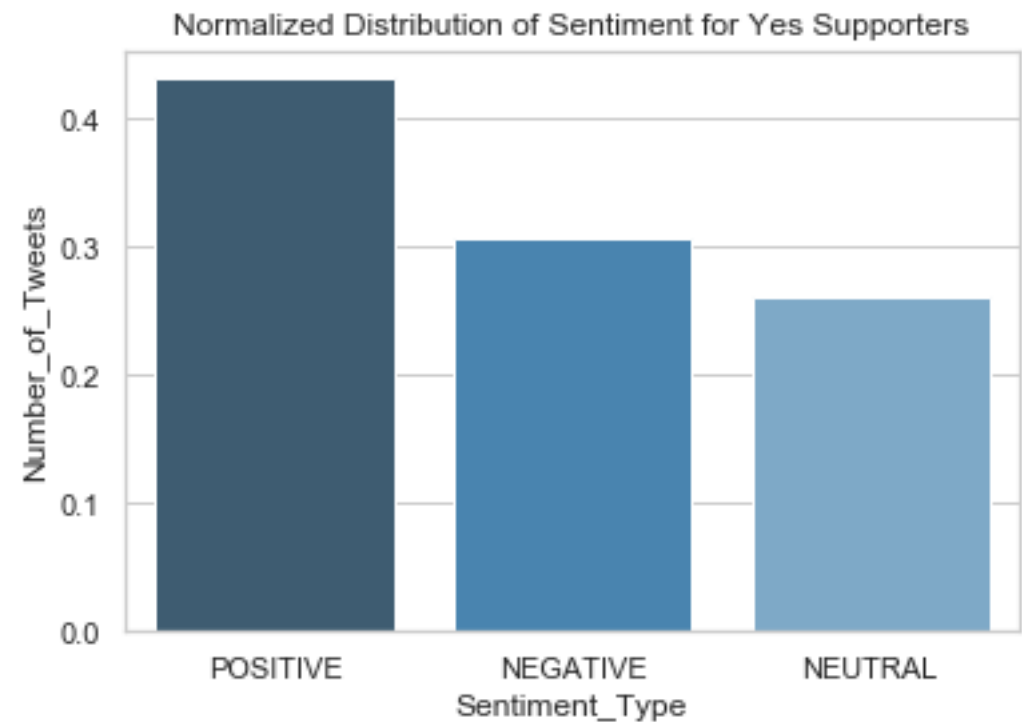
	Sentiment_Type	Number_of_Tweets
0	POSITIVE	0.431692
1	NEGATIVE	0.307699
2	NEUTRAL	0.260609

In [443]:

```
#Show the sentiment distribution based on sentiment_type  
sns.set(style="whitegrid")  
Sentiment_type_N = sns.barplot(x=Sentiment_tweets_no.Sentiment_Type, y=Sentime  
nt_tweets_no.Number_of_Tweets, palette="Blues_d")  
Sentiment_type_N.set_title('Normalized Distribution of Sentiment for Yes Suppo  
rters')
```

Out[443]:

Text(0.5,1,'Normalized Distribution of Sentiment for Yes Supporter
s')



In [444]:

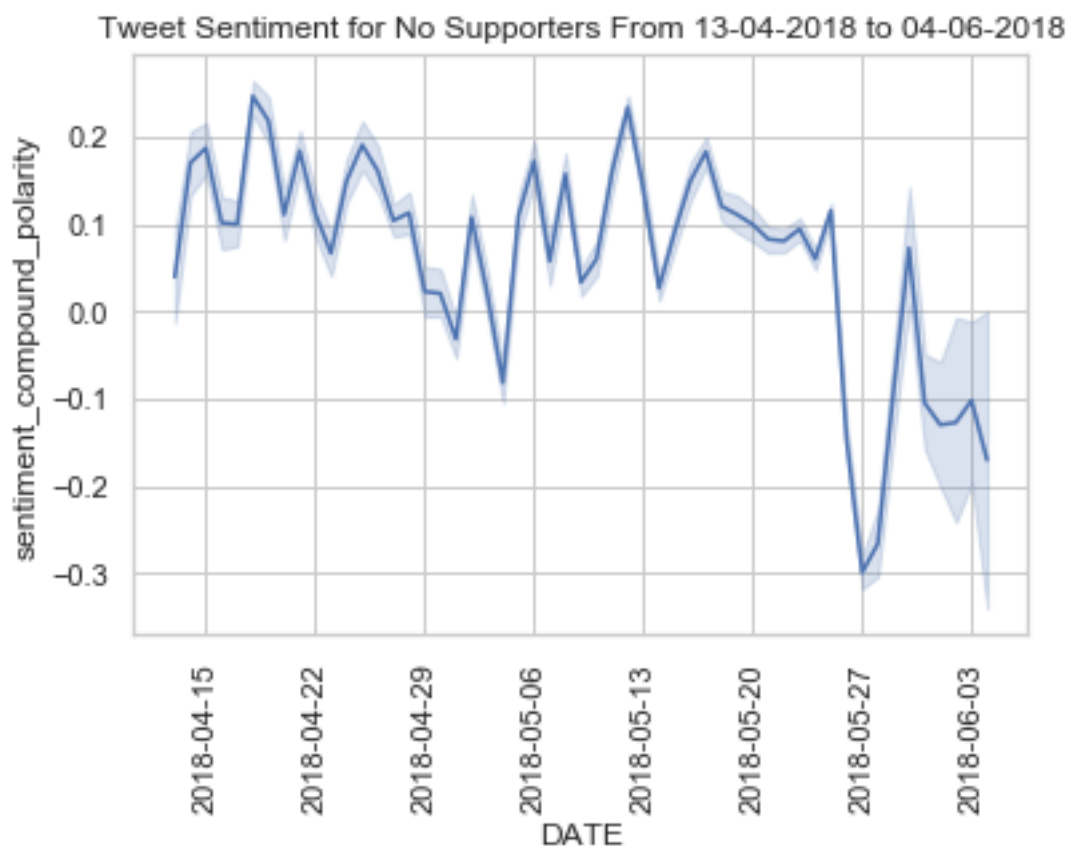
```
#Analysis of Sentiment compound polarity for yes supporters using the full dataset analysis
H = sns.lineplot(x=tweets_no.DATE, y=tweets_no.sentiment_compound_polarity, data=tweets_no)
plt.xticks(rotation=90)
H.set_title('Tweet Sentiment for No Supporters From 13-04-2018 to 04-06-2018')
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

Out[444]:

Text(0.5,1,'Tweet Sentiment for No Supporters From 13-04-2018 to 04-06-2018')



In [445]:

```
tweets_no['sentiment_type'].value_counts()
```

Out[445]:

```
POSITIVE      39154
NEGATIVE      27908
NEUTRAL       23637
Name: sentiment_type, dtype: int64
```

In []:

```
RT_Sentiment_all = tweets[['sentiment_type','text','favorite_count','sentiment_compound_polarity','Screenname','retweet_count','news']]
RT_Sentiment_all = RT_Sentiment_all.sort_values(by=['retweet_count'],ascending=False)
RT_Sentiment_all
```

In []:

```
sns.barplot(x=RT_Sentiment_all.sentiment_type, y=RT_Sentiment_all.retweet_count,hue=RT_Sentiment_all.news)
```

Top Favourite, Retweeted and Followed

In [484]:

```
pd.set_option('display.max_columns', None)
pd.options.display.max_colwidth = 1000
TopRT = tweets3.sort_values(by=['retweet_count'],ascending=False)
TopRT = TopRT [['DATE','retweet_count','text','Screenname','retweet_screen_name','user_followers_count','sentiment_type','support','screen_nc','screen_yc','news','scpol']]
Unique_TOPRT = TopRT.drop_duplicates(subset='text')
Unique_TOPRT.head(10)
```

Out[484]:

	DATE	retweet_count	text	Screenname	retweet_s
568402	2018-05-25	21556.0	RT @MorabitoCM: As Ireland goes #hometovote to #RepealThe8th today, please remember Savita Halappanavar who died of sepsis after being deni...	_aditya_kumar_	MorabitoC
384271	2018-05-25	21555.0	As Ireland goes #hometovote to #RepealThe8th today, please remember Savita Halappanavar who died of sepsis after being denied an abortion for a pregnancy even after doctors said miscarriage was inevitable. Her death was a catalyst for this referendum. She was 31.	MorabitoCM	NaN

			https://t.co/EcZm95xrEI		
1959541	2018-05-11	19223.0	RT @NicolaSturgeon: Should Scotland be an independent country? RT if you intend to #voteYES	stevie_lindsay	NicolaStu
1194172	2018-05-16	17838.0	Today in Ireland, if a woman is raped and impregnated, who then takes an abortion pill, will find herself liable to get a jail sentence longer than the man who raped her.\n\nLet that sink in.\n\nWe need to get rid of the 8th\n\nThis is barbaric\n\n#repealthe8th #together4yes #voteeyes #YES	becksk	NaN
198430	2018-05-24	8616.0	I have just been transferred to a hospice as my treatment for stage 4 breast cancer has stopped. I have just lost power in my legs so will be going to vote in a wheelchair for first time. If I can you can. Well I hope I can 😊 #VoteYes #Together4Yes	KAYDO6	NaN
326721	2018-05-24	7333.0	A vote for the freedom to choose, a vote for women's rights, a vote for women's control over their bodies, a vote for women's health & safety, a vote towards equality. Thinking of Ireland #together4yes #repealthe8th https://t.co/CDNj9KoSGt	EmmaWatson	NaN
137560	2018-	6615.0	Just collected eldest son from Dublin Airport. In 2015 Ireland gave him his right - the right to get married. \nOn Friday he'll return the favour and vote	MchughNoeleen	NaN

	05-23		to give women the right to make decisions about their own bodies. #Together4Yes #hometovote #soproud		
1681708	2018-04-27	6470.0	RT @DanielAndrewsMP: The Golden Girls got it in the '80s.\n\nHopefully Australia agrees in 2017.\n\n#VoteYes. https://t.co/URIYpzp7hF	MLQVH3928	DanielAnc
870303	2018-05-26	6230.0	RT @laurafleur: Irish women who have come #hometovote arriving at Dublin airport. https://t.co/jTMm1xg9a1	BrownsKitchen	laurafleur
289719	2018-05-24	6230.0	Irish women who have come #hometovote arriving at Dublin airport. https://t.co/iTMm1xd9a1	laurafleur	NaN

In [485]:

```
pd.set_option('display.max_columns', None)
pd.options.display.max_colwidth = 1000
TopFav = tweets3.sort_values(by=['favorite_count'], ascending=False)
TopFav = TopFav [['DATE', 'favorite_count', 'text', 'Screenname', 'user_followers_
count', 'sentiment_type', 'support']]
TopFav.head(10)
```

Out[485]:

	DATE	favorite_count	text	Screenname	user_follc
198430	2018-05-24	51516	I have just been transferred to a hospice as my treatment for stage 4 breast cancer has stopped. I have just lost power in my legs so will be going to vote in a wheelchair for first time. If I can you can. Well I hope I can 😊 #VoteYes #Together4Yes	KAYDO6	1227
			Just collected eldest son from Dublin Airport. In 2015 Ireland gave him his		

137560	2018-05-23	46517	right - the right to get married. \nOn Friday he'll return the favour and vote to give women the right to make decisions about their own bodies. #Together4Yes #hometovote #soproud	MchughNoeleen	291
384271	2018-05-25	43812	As Ireland goes #hometovote to #RepealThe8th today, please remember Savita Halappanavar who died of sepsis after being denied an abortion for a pregnancy even after doctors said miscarriage was inevitable. Her death was a catalyst for this referendum. She was 31. https://t.co/EcZm95xrEI	MorabitoCM	815
707703	2018-05-25	37331	My granny is as old as the Irish State. She just voted #yes #Repealthe8th #Together4Yes 🥺💕 https://t.co/b1JTLxuusO	hannahmdeasy	1213
467277	2018-05-25	35688	Big day for Ireland today. Not normally one for asking people to vote one way or another. But as an Irish son and brother, no state should have the right to tell a woman what to do with her body. #Repealthe8th	radioleary	2829221
326721	2018-05-24	35385	A vote for the freedom to choose, a vote for women's rights, a vote for women's control over their bodies, a vote for women's health & safety, a vote towards equality. Thinking of Ireland #together4yes #repealthe8th https://t.co/CDNj9KoSGt	EmmaWatson	28560343

289719	2018-05-24	33562	Irish women who have come #hometovote arriving at Dublin airport. https://t.co/jTMm1xg9a1	laurafleur	7766
267640	2018-05-24	33408	Last minute decision to go #hometovote to #repealthe8th No flights left from London, so I have a 4 hour train, 4 hour wait and 3 hour ferry to make it home to vote- which is a walk in the park in comparison to the journey that Irish women are making every day to the UK	MichelleMarleyy	284
39501	2018-05-22	33340	Keep your nose out of my dick. #TogetherForYes https://t.co/QDPQuG8hlc	Kathbum	405462
1194172	2018-05-16	32791	Today in Ireland, if a woman is raped and impregnated, who then takes an abortion pill, will find herself liable to get a jail sentence longer than the man who raped her.\n\nLet that sink in.\n\nWe need to get rid of the 8th\n\nThis is barbaric\n\n#repealthe8th #together4yes #voteyes #YES	becksk	1070

In [486]:

```
top_users = tweets3.drop_duplicates(subset='Screenname')
top_users.shape
```

Out[486]:

(259379, 38)

In [487]:

```
top_users = top_users.sort_values(by=['user_followers_count'],ascending=False)
top_users = top_users[['Screenname','user_statuses_count','user_followers_count','support','screen_nc','screen_yc','news','scpol','friend_foll_ratio']]
top_10 = top_users.head(10)
top_10.to_csv('top_users.csv')
```

In [488]:

```
#location_filtering
df_Ire=tweets3[tweets3['user_location'].str.match('Ireland|Munster|Connacht|Ulster|Antrim|Armagh|Carlow|Cavan|Clare|Cork|Derry|Donegal|Down|Dublin|Fermanagh|Galway|Kerry|Kildare|Kilkenny|Laois|Leitrim|Limerick|Longford|Louth|Mayo|Meath|Monaghan|Offaly|Roscommon|Sligo|Tipperary|Tyrone|Waterford|Westmeath|Wexford|Wicklow',case=False)==True]
print(df_Ire.shape)
```

(358979, 38)

In [489]:

```
df_Ire.columns
```

Out[489]:

```
Index(['DATE', 'TIME', 'hashtags', 'urls', 'contains_url', 'favorite_count',
      'is_retweet', 'retweet_count', 'Logged_RT_count', 'retweet_screen_name',
      'retweet_yc', 'retweet_nc', 'rtpol', 'RT', 'text', 'text_news',
      'sentiment_compound_polarity', 'sentiment_type', 'Yes_VOTE', 'No_VOTE',
      'news_content', 'Yes_camp', 'No_camp', 'poltweet', 'friend_foll_ratio',
      'user_statuses_count', 'user_location', 'Screenname', 'user_description',
      'user_favourites_count', 'user_followers_count', 'user_friends_count',
      'screen_yc', 'screen_nc', 'scpol', 'news', 'campaign_support', 'support'],
      dtype='object')
```

In [490]:

```
top_users_ire = df_Ire.drop_duplicates(subset='Screenname')
```

In [491]:

```
top_users_ire = top_users_ire[['Screenname','user_location','user_statuses_count','user_description','user_followers_count','support','screen_nc','screen_yc','news','scpol','friend_foll_ratio']]
top_users_ire = top_users_ire.sort_values(by='user_followers_count',ascending=False)
topusers = top_users_ire.head(10)
topusers.to_csv('topusersIre.csv')
```


In [492]:

```
URLS = tweets3[tweets3['contains_url']==1]
topRT_url = URLS.sort_values(by='retweet_count',ascending=False)
topRT_url[['DATE','urls','retweet_count','Screenname','retweet_screen_name','retweet_yc','retweet_nc','rtpol','RT','text','sentiment_type','screen_yc','support','screen_nc','scpol','news']]
topRT_url.head(30)
```

Out[492]:

	DATE	TIME	hashtags	u
326721	2018-05-24	21:27:11	together4yes , repealthe8th	https://www.togetherforyes.ie
737008	2018-05-25	21:08:01	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752790	2018-05-25	21:31:23	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
736409	2018-05-25	21:07:05	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

				1.3508861
881395	2018-05-26	09:01:18	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752811	2018-05-25	21:31:24	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881272	2018-05-26	09:01:03	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752850	2018-05-25	21:31:24	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881227	2018-05-26	09:00:53	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

				1.3508861
752804	2018-05-25	21:31:28	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752816	2018-05-25	21:31:28	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881615	2018-05-26	09:02:35	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752852	2018-05-25	21:31:28	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752954	2018-05-25	21:31:32	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

752887	2018-05-25	21:31:33	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752914	2018-05-25	21:31:33	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752902	2018-05-25	21:31:35	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
753048	2018-05-25	21:31:41	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
737665	2018-05-25	21:09:01	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

880595	2018-05-26	08:58:34	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881563	2018-05-26	09:02:16	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881758	2018-05-26	09:02:53	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752989	2018-05-25	21:31:43	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752780	2018-05-25	21:31:18	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

736411	2018-05-25	21:07:04	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
882824	2018-05-26	09:07:16	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
882577	2018-05-26	09:06:15	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752554	2018-05-25	21:31:04	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752535	2018-05-25	21:31:05	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

752667	2018-05-25	21:31:08	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
---------------	------------	----------	--	---

In [493]:

```
topRT_url.drop_duplicates(subset='text')
topRT_url.head(1000)
```

Out[493]:

	DATE	TIME	hashtags	u
326721	2018-05-24	21:27:11	together4yes , repealthe8th	https://www.togetherforyes.ie
737008	2018-05-25	21:08:01	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752790	2018-05-25	21:31:23	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

736409	2018-05-25	21:07:05	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881395	2018-05-26	09:01:18	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752811	2018-05-25	21:31:24	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881272	2018-05-26	09:01:03	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752850	2018-05-25	21:31:24	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

881227	2018-05-26	09:00:53	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752804	2018-05-25	21:31:28	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752816	2018-05-25	21:31:28	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881615	2018-05-26	09:02:35	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752852	2018-05-25	21:31:28	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

752954	2018-05-25	21:31:32	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752887	2018-05-25	21:31:33	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752914	2018-05-25	21:31:33	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752902	2018-05-25	21:31:35	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
753048	2018-05-25	21:31:41	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
				https://www.irishtimes.com/news/politics/iris

737665	2018-05-25	21:09:01	Referendum2018 , repealthe8th , savethe8th	times-exit-poll-projects-ireland-has-voted-by landslide-to-repeal-eighth-amendment-1.3508861
880595	2018-05-26	08:58:34	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881563	2018-05-26	09:02:16	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
881758	2018-05-26	09:02:53	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752989	2018-05-25	21:31:43	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
			Referendum2018	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

752780	2018-05-25	21:31:18	, repealthe8th , savethe8th	landslide-to-repeal-eighth-amendment-1.3508861
736411	2018-05-25	21:07:04	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
882824	2018-05-26	09:07:16	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
882577	2018-05-26	09:06:15	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
752554	2018-05-25	21:31:04	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
	2018-		Referendum2018 , repealthe8th ,	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

752535	05-25	21:31:05	savethe8th	landslide-to-repeal-eighth-amendment-1.3508861
752667	2018-05-25	21:31:08	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
...
757973	2018-05-25	21:39:50	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
841798	2018-05-26	04:41:37	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
841737	2018-05-26	04:41:23	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
				https://www.irishtimes.com/news/politics/iris

841702	2018-05-26	04:40:19	Referendum2018 , repealthe8th , savethe8th	times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758074	2018-05-25	21:39:55	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
737480	2018-05-25	21:08:40	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758009	2018-05-25	21:39:57	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758224	2018-05-25	21:40:15	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
			Referendum2018	https://www.irishtimes.com/news/politics/iris

840673	2018-05-26	04:26:13	, repealthe8th , savethe8th	times-exit-poll-projects-ireland-has-voted-by landslide-to-repeal-eighth-amendment-1.3508861
840663	2018-05-26	04:26:03	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
839632	2018-05-26	04:11:22	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758334	2018-05-25	21:40:29	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
839585	2018-05-26	04:10:09	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
	2018-		Referendum2018	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by

839591	05-26	04:09:45	, repealthe8th , savethe8th	landslide-to-repeal-eighth-amendment-1.3508861
758380	2018-05-25	21:40:32	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758377	2018-05-25	21:40:33	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758414	2018-05-25	21:40:33	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758438	2018-05-25	21:40:37	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
839271	2018-	04:05:49	Referendum2018 , repealthe8th ,	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

	05-26		savethe8th	1.3508861
839337	2018-05-26	04:05:32	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
839266	2018-05-26	04:05:24	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
839318	2018-05-26	04:04:26	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758505	2018-05-25	21:40:40	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758562	2018-05-25	21:40:41	Referendum2018 , repealthe8th ,	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

			savethe8th	1.3508861
839165	2018-05-26	04:02:26	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758472	2018-05-25	21:40:42	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758481	2018-05-25	21:40:44	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
838990	2018-05-26	03:59:36	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861
758353	2018-05-25	21:40:26	Referendum2018 , repealthe8th ,	https://www.irishtimes.com/news/politics/iris-times-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

			savethe8th	1.3508861
737493	2018-05-25	21:08:39	Referendum2018 , repealthe8th , savethe8th	https://www.irishtimes.com/news/politics/irishtimes-exit-poll-projects-ireland-has-voted-by-landslide-to-repeal-eighth-amendment-1.3508861

1000 rows × 38 columns

NEWS TWEETS ANALYSIS

In []:

```
pd.set_option('display.max_columns', None)
pd.options.display.max_colwidth = 1000
RT_Sentiment_compound = tweets[tweets['news']]
RT_Sentiment_compound.sort_values(by=['retweet_count'],ascending=False)
```

In []:

```
RT_Sentiment_compound.sentiment_compound__polarity.describe()
```

In []:

```
RT_Sentiment_compound.dtypes
```

In []:

```
RT_Sentiment_compound['Hfollowers'] = RT_Sentiment_compound['user_followers_count'] > 319670
RT_Sentiment_compound['Hfollowers'].value_counts()
```

In []:

```
RT_Sentiment_compound['log_retweet'] = np.log(RT_Sentiment_compound.retweet_count)
#USERDATA['log_friends'] = np.log(USERDATA.user_friends_count)
RT_Sentiment_compound
```

In []:

```
sns.scatterplot(x=RT_Sentiment_compound.sentiment_compound_polarity, y=RT_Sentiment_compound.log_retweet)
```

In []:

```
RT_Sentiment_Type = data_news[['sentiment_type', 'retweet_count', 'user_followers_count']]
```

In []:

```
sns.barplot(x=RT_Sentiment_Type.sentiment_type, y=RT_Sentiment_Type.retweet_count)
```

In []:

```
group = data_news.groupby('Screenname')
```

In []:

```
data_news['log_retweets'] = np.log(data_news.retweet_count)
data_news
```

Sentiment Scoring Yes Supporters

In []:

```
sid = SentimentIntensityAnalyzer()
tweets_yes['sentiment_compound_polarity']=tweets_yes.text.apply(lambda x:sid.polarity_scores(x)['compound'])
tweets_yes['sentiment_neutral']=tweets_yes.text.apply(lambda x:sid.polarity_scores(x)['neu'])
tweets_yes['sentiment_negative']=tweets_yes.text.apply(lambda x:sid.polarity_scores(x)['neg'])
tweets_yes['sentiment_pos']=tweets_yes.text.apply(lambda x:sid.polarity_scores(x)['pos'])
tweets_yes['sentiment_type']=''
tweets_yes.loc[tweets_yes.sentiment_compound_polarity>0,'sentiment_type']='POSITIVE'
tweets_yes.loc[tweets_yes.sentiment_compound_polarity==0,'sentiment_type']='NEUTRAL'
tweets_yes.loc[tweets_yes.sentiment_compound_polarity<0,'sentiment_type']='NEGATIVE'
tweets_yes.head()
```

In []:

```
pd.set_option('display.max_columns', None)
tweets_yes.head()
```

In []:

```
tweets_yes.dtypes
```

Sentiment Scoring No Supporters

In []:

```
sid = SentimentIntensityAnalyzer()
tweets_no['sentiment_compound_polarity']=tweets_no.text.apply(lambda x:sid.polarity_scores(x)['compound'])
tweets_no['sentiment_neutral']=tweets_no.text.apply(lambda x:sid.polarity_scores(x)['neu'])
tweets_no['sentiment_negative']=tweets_no.text.apply(lambda x:sid.polarity_scores(x)['neg'])
tweets_no['sentiment_pos']=tweets_no.text.apply(lambda x:sid.polarity_scores(x)['pos'])
tweets_no['sentiment_type']=''
tweets_no.loc[tweets_no.sentiment_compound_polarity>0,'sentiment_type']='POSITIVE'
tweets_no.loc[tweets_no.sentiment_compound_polarity==0,'sentiment_type']='NEUTRAL'
tweets_no.loc[tweets_no.sentiment_compound_polarity<0,'sentiment_type']='NEGATIVE'
tweets_no.head()
```

In []:

```
pd.set_option('display.max_columns', None)
tweets_no.head()
```

In []:

```
tweets_no.dtypes
```

In []:

```
#according to this I lose 439745 tweets: get a dataset of this as well and do
a) topic modelling b) check how many news
#accounts within the dataset

#FOR the entire dataset (b/c the above is also a robustness check; create a bar
chart of the sentiment type
#filtered by YES and NO; this will have to be two barcharts since the filters
are separate)
```

In []:

```
tweets.to_csv('tweets_regression.csv')
```

In []:

```
retweets = tweets[tweets['is_retweet']]
```

In []:

```
retweets.to_csv('retweets.csv')
```

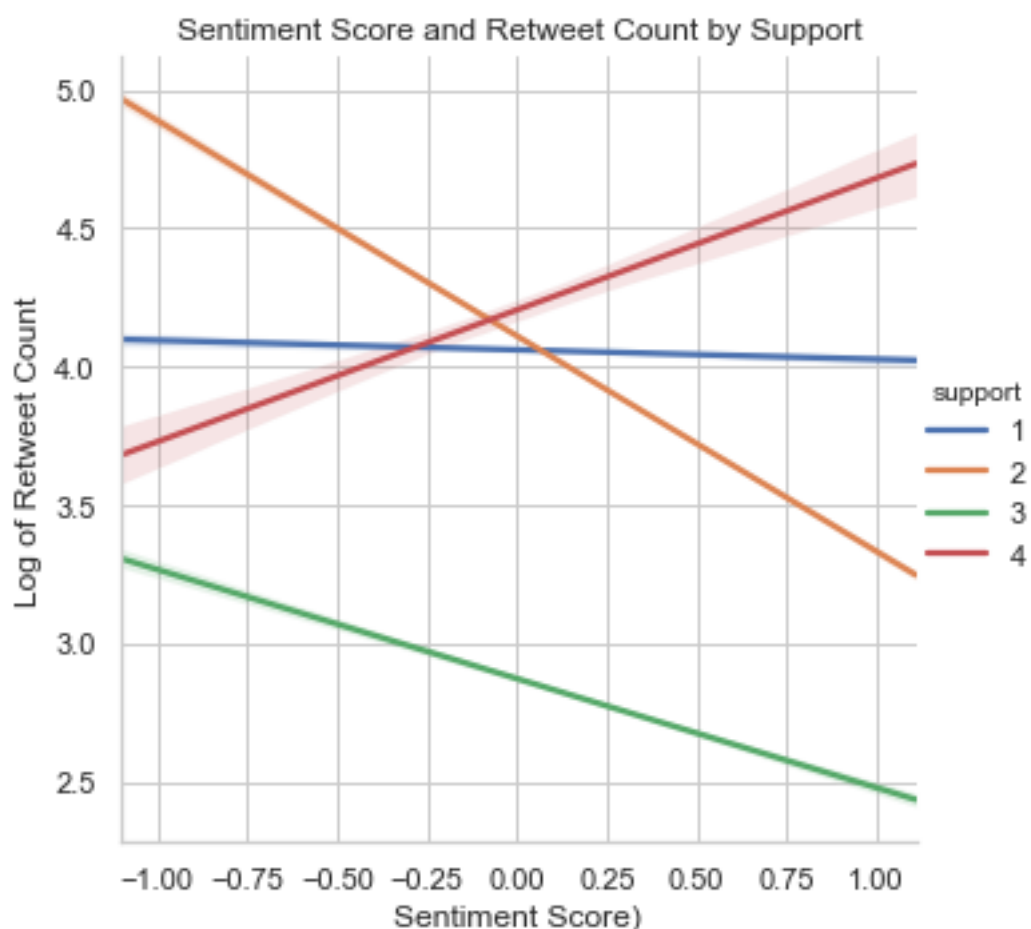
Regression Plots

In [459]:

```
reg1 = sns.lmplot(x="sentiment_compound_polarity",y="Logged_RT_count",data=twe  
ets4,hue="support", scatter = False)  
reg1.set(xlabel = 'Sentiment Score)', ylabel = 'Log of Retweet Count')  
reg1 = plt.title("Sentiment Score and Retweet Count by Support")
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packag
es/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprec
ated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future t
his will be interpreted as an array index, `arr[np.array(seq)]`, w
hich will result either in an error or a different result.

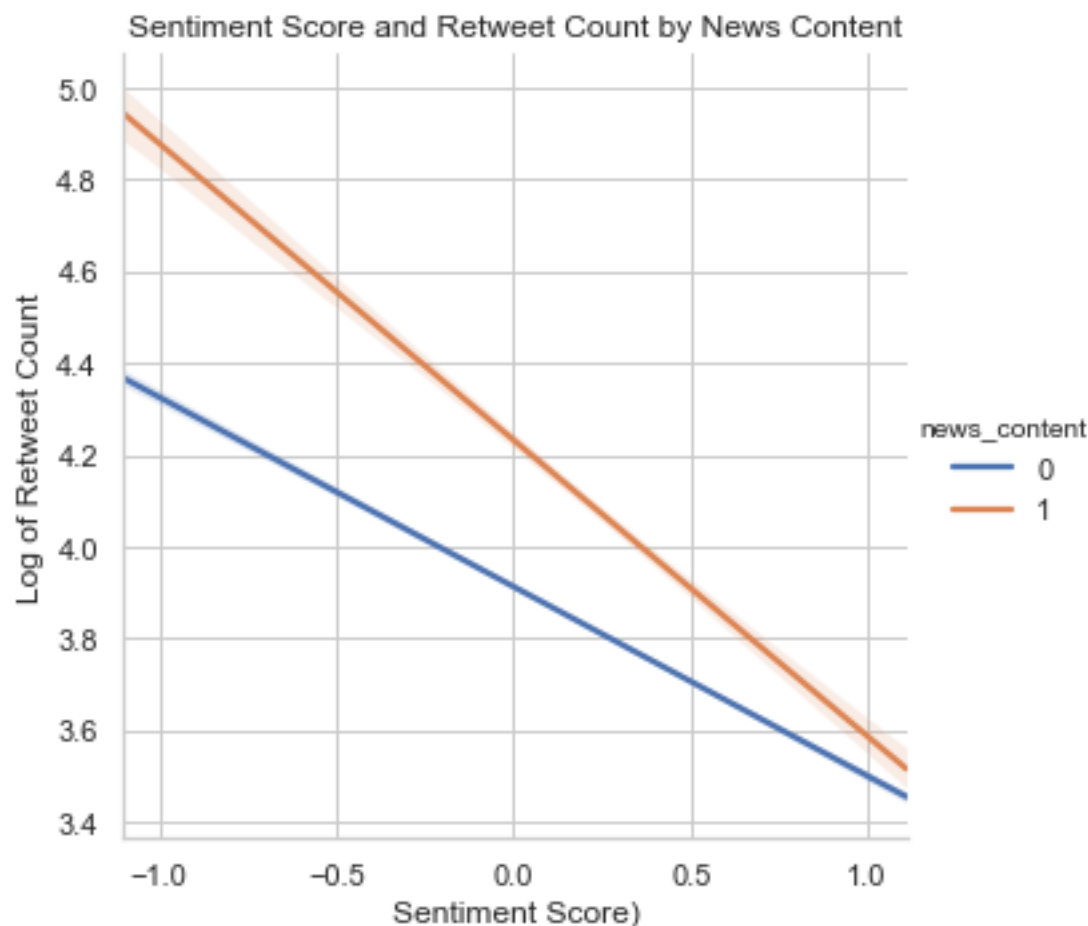


In [460]:

```
reg_news = sns.lmplot(x="sentiment_compound_polarity", y = "Logged_RT_count",  
data = tweets4, hue = "news_content", scatter=False)  
reg_news.set(xlabel = 'Sentiment Score)', ylabel = 'Log of Retweet Count')  
reg_news = plt.title("Sentiment Score and Retweet Count by News Content")
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

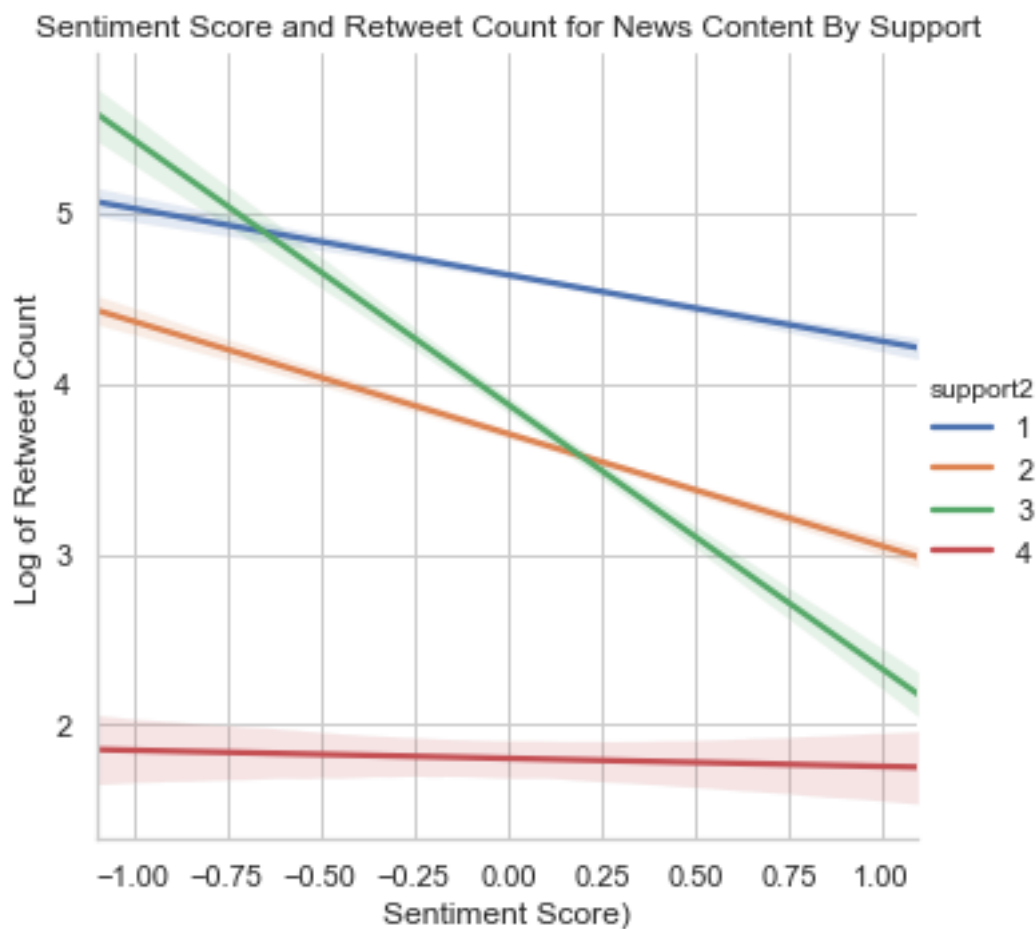


In [467]:

```
reg3 = sns.lmplot(x="sentiment_compound_polarity",y="Logged_RT_count",data=news_df,hue="support2", scatter = False)
reg3.set(xlabel = 'Sentiment Score)', ylabel = 'Log of Retweet Count')
reg3 = plt.title("Sentiment Score and Retweet Count for News Content By Support")
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.



In [462]:

```
campaign_data2 = campaign_data[campaign_data['is_retweet']]
campaign_data.shape, campaign_data2.shape
```

Out[462]:

```
((11498, 39), (10866, 39))
```

In [463]:

```
pol_data2 = pol_data[pol_data['is_retweet']]
pol_data.shape, pol_data2.shape
```

Out[463]:

```
((41316, 38), (36134, 38))
```

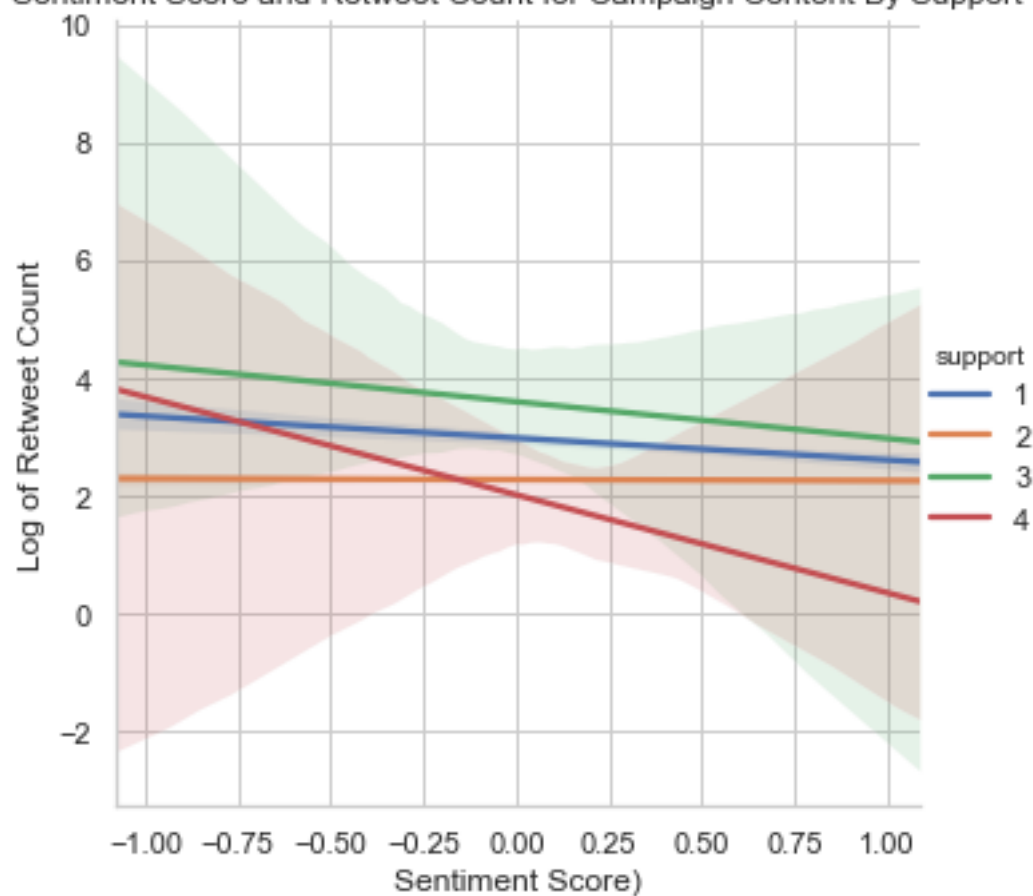

In [464]:

```
camp = sns.lmplot(x="sentiment_compound_polarity",y="Logged_RT_count",data=campaign_data2,hue="support",scatter=False)
camp.set(xlabel = 'Sentiment Score)', ylabel = 'Log of Retweet Count')
camp = plt.title("Sentiment Score and Retweet Count for Campaign Content By Support")
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

Sentiment Score and Retweet Count for Campaign Content By Support

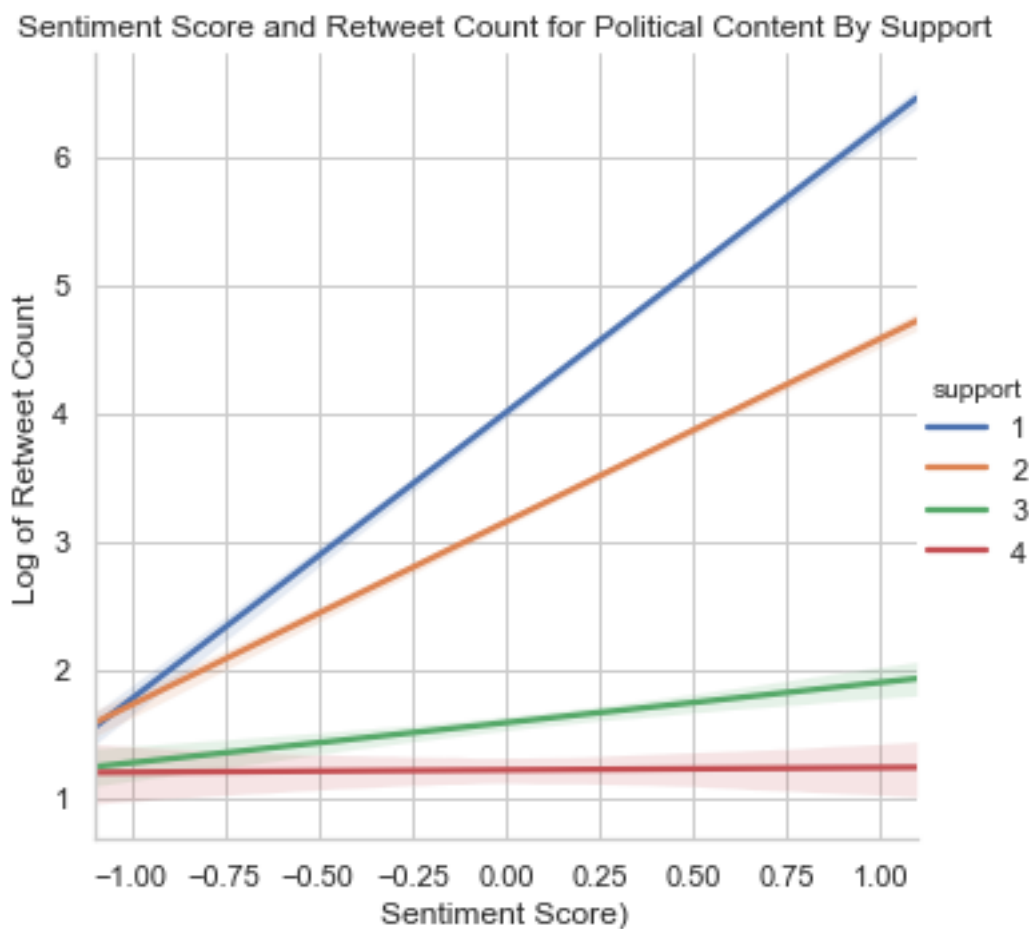


In [465]:

```
pol = sns.lmplot(x="sentiment_compound_polarity",y="Logged_RT_count",data=pol_data2,hue="support",scatter=False)
pol.set(xlabel = 'Sentiment Score)', ylabel = 'Log of Retweet Count')
pol = plt.title("Sentiment Score and Retweet Count for Political Content By Support")
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.



In [466]:

```
Ire = sns.lmplot(x="sentiment_compound_polarity",y="Logged_RT_count",data=Ireland,hue="support",scatter=False)
Ire.set(xlabel = 'Sentiment Score)', ylabel = 'Log of Retweet Count')
Ire = plt.title("Sentiment Score and Retweet Count for News Content By Support In Ireland")
```

/Users/trishaharjani/anaconda2/envs/py36/lib/python3.6/site-packages/scipy/stats/stats.py:1713: FutureWarning:

Using a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will result either in an error or a different result.

Sentiment Score and Retweet Count for News Content By Support In Ireland

