

Reddit NLP project

Author: Jake Peavy

r/politics

r/the_donald

- I scraped 1993 posts total
- both are echo chambers for opinionated individuals with contrasting political (and moral) views
- my hunch is that they will be easily discernible, even by a dumb computer



★ 598 days in office and despite the endless mud slinging by media and book writers, Trump remains 6 points ahead of where Obama was. ★ (i.imgur.com)

submitted an hour ago by riidii - stickied
45 comments share save hide give gold deport crosspost



★ Look at All of These Human Trafficking Arrests Done By Obama! ★ (i.redd.it)

submitted 46 minutes ago by nitroyoshi9 FL - stickied
54 comments share save hide give gold deport crosspost



President Trump on Twitter: Sour Obummer Trump Tweet (twitter.com)

submitted 4 hours ago by derangedwinchester
246 comments share save hide give gold deport crosspost



Peter Strzok, Lisa Page conspired to leak anti-Trump stories to mainstream media CORRUPTION (washingtontimes.com)

submitted 7 hours ago by Stickrell SUPER ELITE
308 comments share save hide give gold deport crosspost



Republicans release second round of proposed tax cuts - which follow comprehensive tax legislation December. DRAIN THE SWAMP (reuters.com)

submitted 5 hours ago by asiatrails SCT
share save hide give gold deport crosspost



God-Emperor nearly STARED A MAN TO DEATH when someone screamed GUN and the Secret kept into action GOD-EMPEROR (redd.it)

submitted 5 hours ago by onewalleee 1776
share save hide give gold deport crosspost

one by @SecNielsen and @FEMA @FEMA_Brock, along with @VP Mike
off. This is a very dangerous form which is very dangerous. Heed the directions of your State
you. Be SAFE! Hurricane Florence (twitter.com)

ing and Trumo-Kim summit STABLE GENIUS (thehill.com)

ed this on Twitter "The city of San Francisco has released a Guide to Registration and Voting

search



648,416 Patriots
13,273 winners online!



SUBMIT A LINK



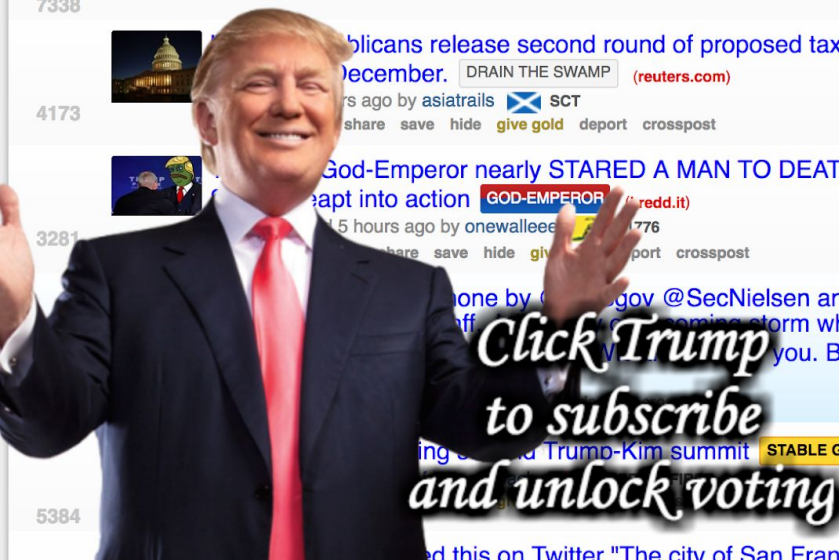
SUBMIT A TEXT POST

subscribe

☒ Show my flair on this subreddit. It looks like:
nnnnnnnnnnnnnnnn (edit)



WELCOME



2



The Creepiest Ghost Towns & Abandoned Cities Around The World Guaranteed To Give You Chills

(American Upbeat)



promoted by graydown_ja

[promoted](#) [save](#) [give gold](#) [report](#)

179



The "What happened in your state last week?" Megathread, Week 37

self.politics



submitted 12 hours ago by [optimalg](#) [M] - [POLITICS MOD POST](#)

[74 comments](#) [share](#) [save](#) [hide](#) [give gold](#) [report](#) [crosspost](#)

27.8k



Trump reportedly exploded at his ex-lawyer after he heard Mueller was looking into his relationship with Deutsche Bank: 'This is bulls---!' [businessinsider.com](#)

submitted 5 hours ago by [roku44](#)

[1694 comments](#) [share](#) [save](#) [hide](#) [give gold](#) [report](#) [crosspost](#)

11.5k



Bombs Away: Trump Has the I.Q. of an Inbred Tanning Bed, Says a Liberated Gary Cohn [vanityfair.com](#)

submitted 5 hours ago by [_basquiat](#) [District Of Columbia](#)

[797 comments](#) [share](#) [save](#) [hide](#) [give gold](#) [report](#) [crosspost](#)

22.8k



Ted Cruz Takes Brave Anti-Tofu Stance, Despite Texas Being a Soybean State [esquire.com](#)

submitted 9 hours ago by [Mysteriagant](#) [Texas](#)

[1781 comments](#) [share](#) [save](#) [hide](#) [give gold](#) [report](#) [crosspost](#)

10.1k



What boycott? Nike sales are up 31 percent since the Kaepernick campaign [nbcnews.com](#)

submitted 8 hours ago by [wonderingsocrates](#)

[797 comments](#) [share](#) [save](#) [hide](#) [give gold](#) [report](#) [crosspost](#)

Opinion | Contractors get \$750 per child per day to detain migrants. Something's wrong here. [washingtonpost.com](#)

EDA

The top 15 words used by r/politics:

```
In [169]: list(df_cv[df_cv['is_donald'] == 0].drop(['is_donald'], axis=1).sum().sort_values(ascending=False)[:15].index)
```

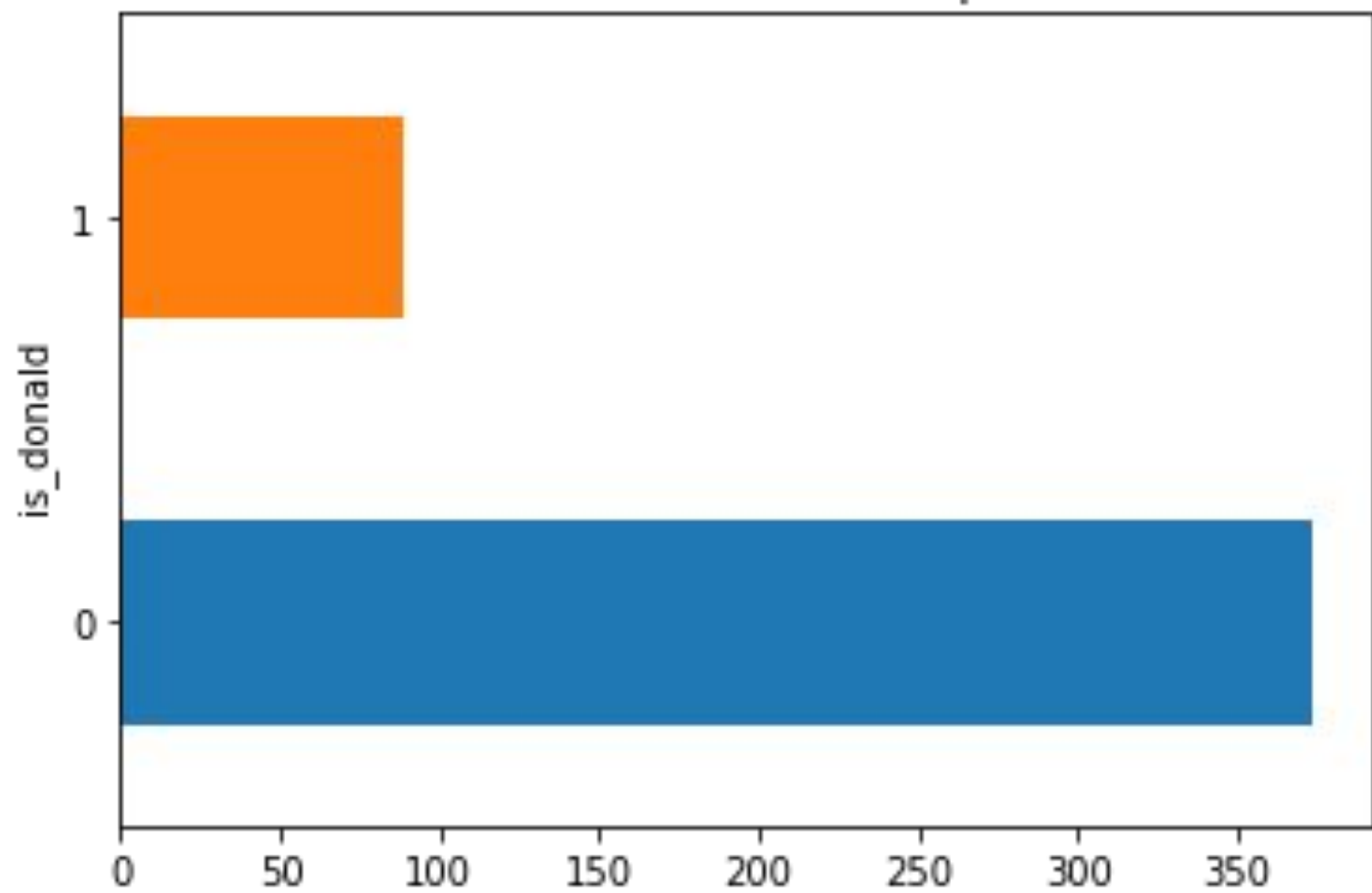
```
Out[169]: ['trump',  
          'kavanaugh',  
          'woodward',  
          'book',  
          'hearing',  
          'bob',  
          'brett',  
          'new',  
          'says',  
          'president',  
          'white',  
          'house',  
          'donald',  
          'court',  
          'democrats']
```

The top 15 words used by r/the_donald:

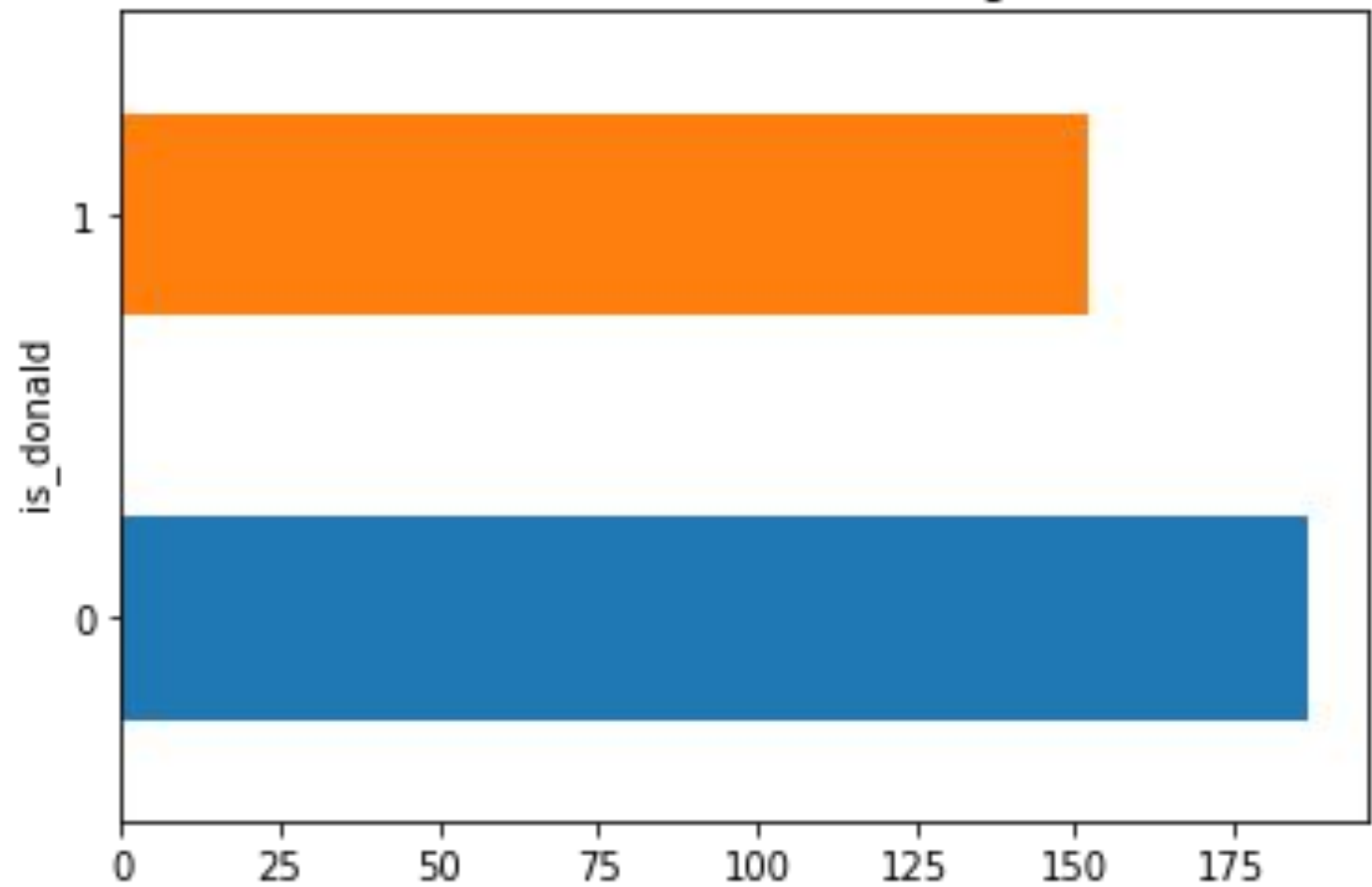
```
In [170]: list(df_cv[df_cv['is_donald'] == 1].drop(['is_donald'], axis=1).sum().sort_values(ascending=False)[:15].index)
```

```
Out[170]: ['kavanaugh',  
           'hearing',  
           'trump',  
           'just',  
           'democrats',  
           'brett',  
           'twitter',  
           'alex',  
           'jones',  
           'day',  
           'don',  
           'white',  
           'people',  
           'confirmation',  
           'president']
```

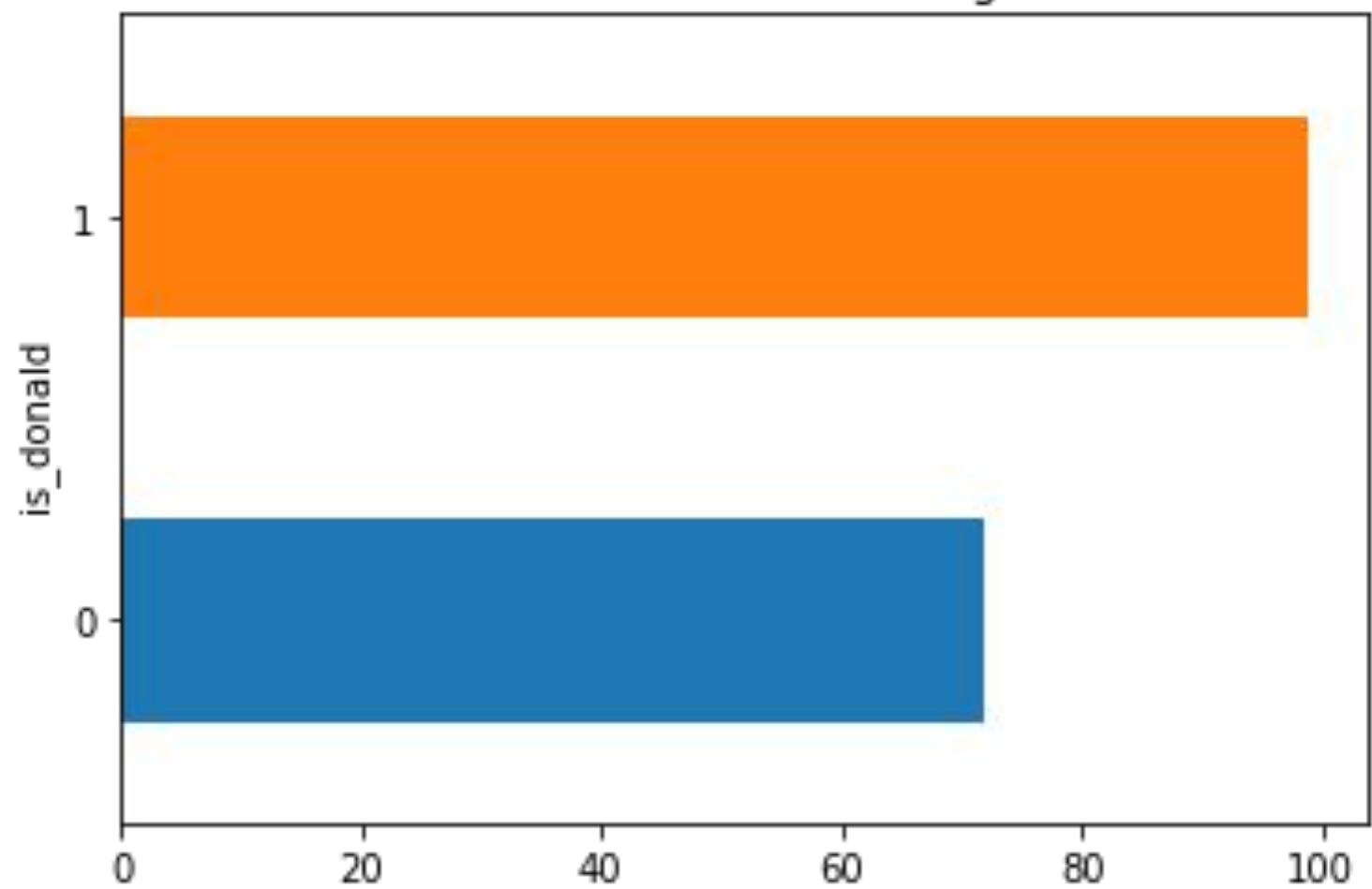
Occurences of Trump



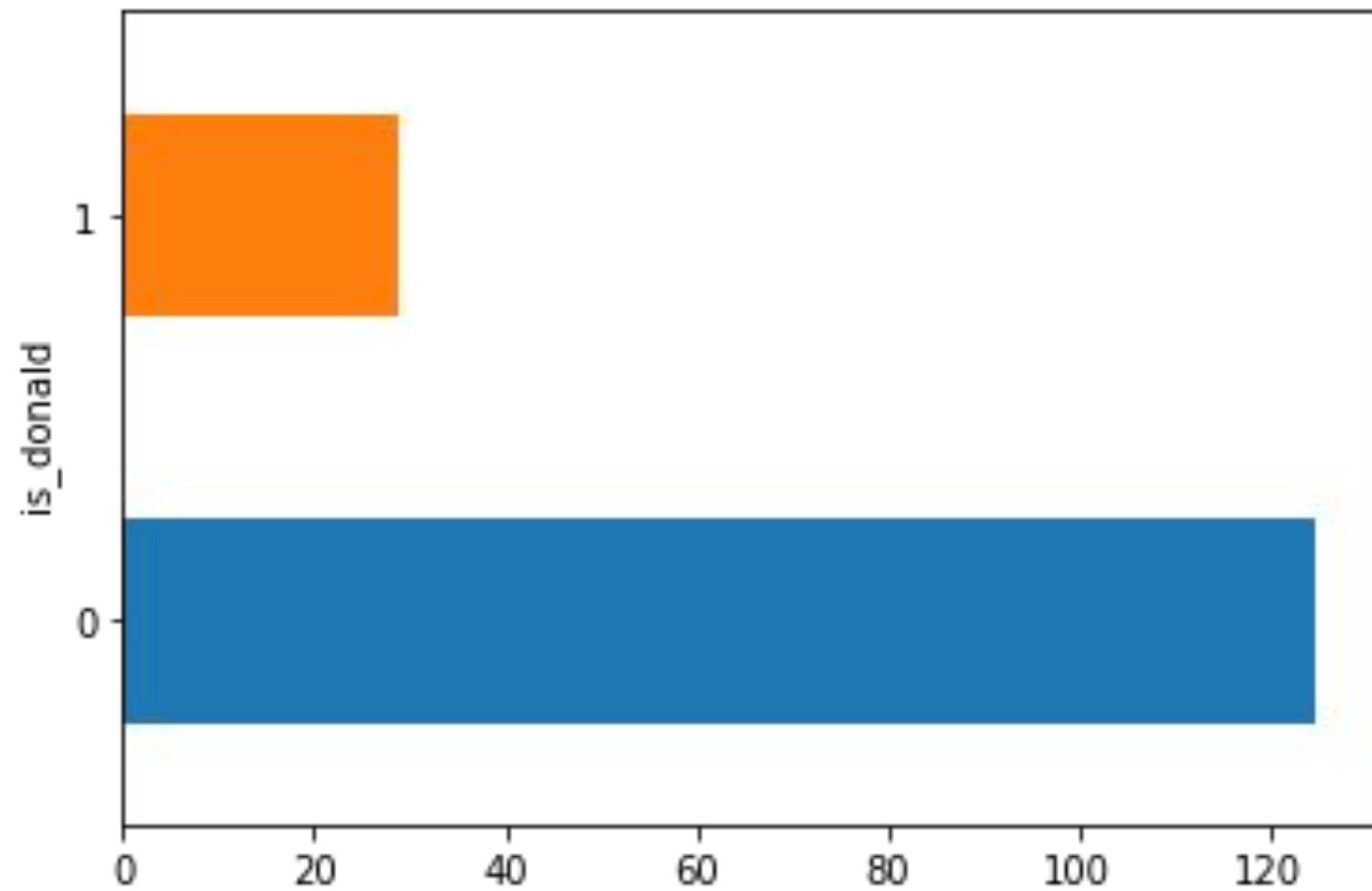
Occurences of Kavanaugh



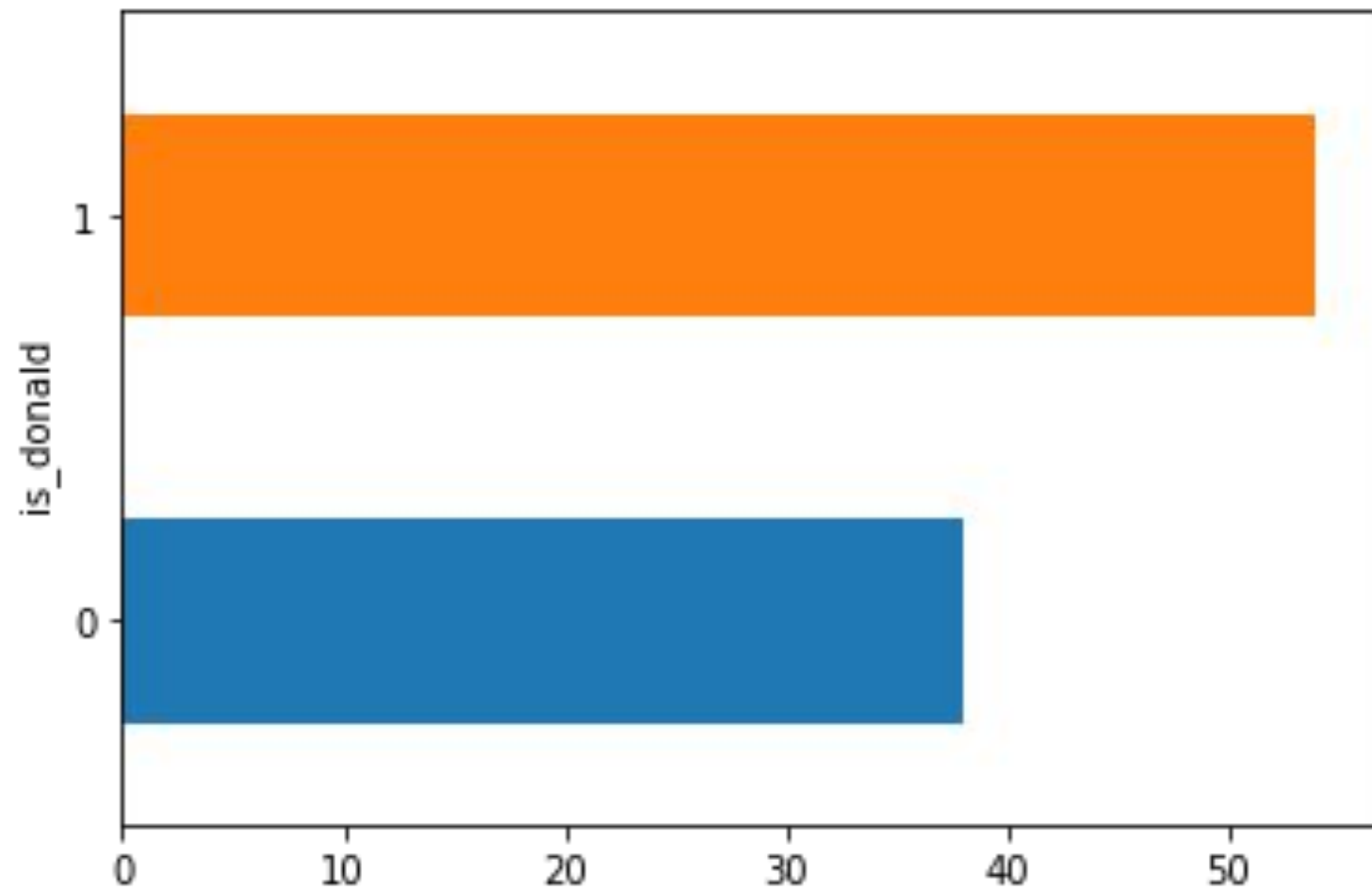
Occurences of Hearing



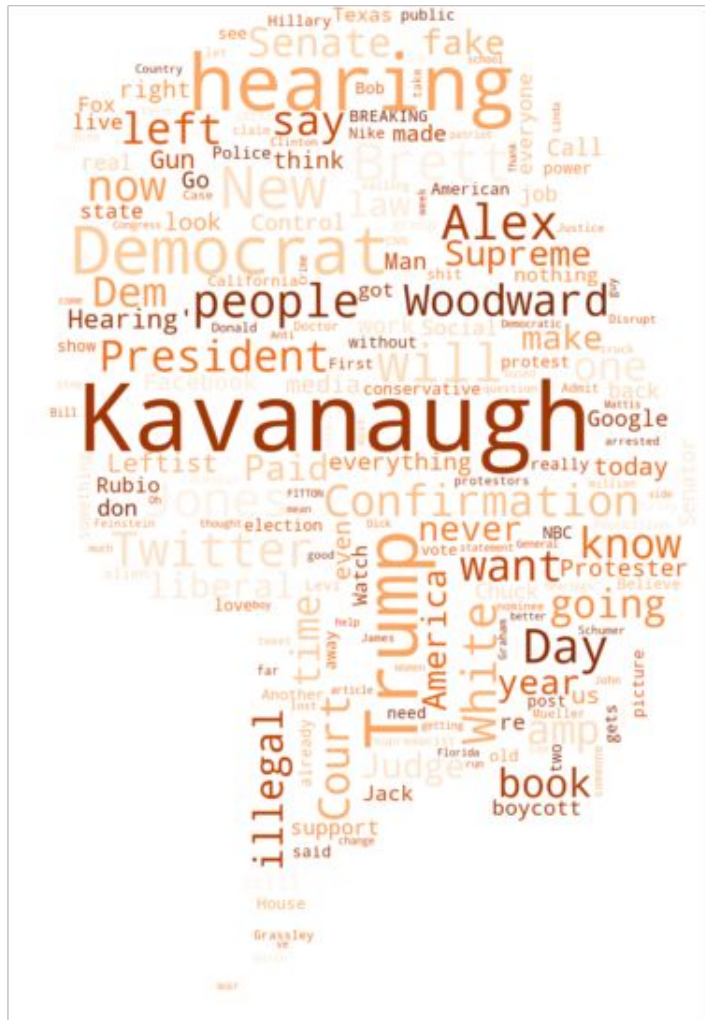
Occurences of Woodward



Occurences of Democrats



Word Clouds



Liberal Snowflake

- ■ ■ made from liberal tears (words)

gay frog



Sentiment Analysis



	compound	neg	neu	pos
is_donald				
0	-0.092988	0.124289	0.806154	0.069574
1	-0.059031	0.115859	0.785260	0.098876

62.55% of the top 50% most "positive" posts are from `r/the_donald`

It seems that `r/the_donald` has a higher rate of positivity overall

this is probably due to the fact that `r/the_donald` is a subreddit that serves as an echo chamber for staunch trump supporters - they have less to complain about

and `r/politics` shows up a lot more often in the top 50% of negative sentiments

in the exact opposite fashion of `r/the_donald`, `r/politics` has A LOT to complain about

Models

Random Forest Classifier

```
pipe = Pipeline([
    ('fu', FeatureUnion([
        ('cv', CountVectorizer(min_df=1, stop_words=None)),
        ('awl', AverageWordLength()),
        ('cs', CompoundScore()),
        ('ns', NegScore()),
        ('ps', PosScore()),
        ('ne', NeuScore())
    ])),
    ('rf', RandomForestClassifier(min_samples_split=2,
                                max_features=5,
                                n_estimators=500,
                                criterion='entropy',
                                max_depth=15))
])
```

In [40]: gs.best_score_

Out[40]: 0.8012048192771084

Random Forest Classifier Confusion Matrix

213	35
53	198

Accuracy rate = $(411/499) = 82.4\%$

Extra Trees Classifier

```
pipe = Pipeline([
    ('fu', FeatureUnion([
        ('cv', CountVectorizer()),
        ('awl', AverageWordLength()),
        ('cs', CompoundScore()),
        ('ns', NegScore()),
        ('ps', PosScore()),
        ('ne', NeuScore())
    ])),
    ('et', ExtraTreesClassifier())
])

pipe.fit(X_train, y_train)
```

```
pipe.score(X_train, y_train), pipe.score(X_test, y_test)
```

```
(0.9979919678714859, 0.811623246492986)
```

Extra Trees Classifier Confusion Matrix

202	46
48	203

Accuracy rate = $(405/499) = 81.2\%$

Predictions

Random Forest predictions with new posts

```
def rf_predict_sub(title):  
    pred = gs.predict(pd.Series(title))  
    if pred == 1:  
        print("This sounds like a r/the_donald post!")  
    else:  
        print("This sounds like a r/politics post!")
```

```
rf_predict_sub('That time God-Emperor nearly STARED A MAN TO DEATH when '+  
               'someone screamed GUN and the Secret Service leapt into action')
```

This sounds like a r/the_donald post!

```
rf_predict_sub('Trump reportedly exploded at his ex-lawyer after he heard Mueller'+  
               'was looking into his relationship with Deutsche Bank: ')
```

This sounds like a r/politics post!

More Random Forest Predictions

```
rf_predict_sub('white people')
```

This sounds like a r/politics post!

```
rf_predict_sub('god-emperor')
```

This sounds like a r/the_donald post!

Extra Trees predictions with new posts

```
def et_predict_sub(title):  
    preds = pipe.predict(pd.Series(title))  
    if preds == 1:  
        print("This sounds like a r/the_donald post!")  
    else:  
        print("This sounds like a r/politics post!")
```

```
et_predict_sub('Trump reportedly exploded at his ex-lawyer after he heard Mueller was'+  
               'looking into his relationship with Deutsche Bank: ')
```

This sounds like a r/politics post!

```
et_predict_sub('That time God-Emperor nearly STARED A MAN TO DEATH when '+  
               'someone screamed GUN and the Secret Service leapt into action')
```

This sounds like a r/the_donald post!

More Extra Trees Predictions

```
et_predict_sub('💩')
```

This sounds like a r/the_donald post!

```
et_predict_sub('trump')
```

This sounds like a r/politics post!