Data Science Research Portfolio

Adil Gokturk

School of Information Studies, Syracuse University, 2019 - 2020

Portfolio Contents

Syracuse University

DataScience@Syracuse

Learning Objectives

- **MS in Applied Data Science Program Learning Goals**
  - A broad overview of the major practice areas in data science
  - Collect and organize data
  - Identify patterns in data via visualization, statistical analysis, and data mining
  - Develop alternative strategies based on the data
  - Develop a plan of action to implement the business decisions derived from the analyses
  - Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization
  - Synthesize the ethical dimensions of data science practice

Key Projects

Project I: An Analysis of Viral Tweets, # Yellow Vests

IST707 Data Analytics, Dr. Ami Gates

Fall 2018

Syracuse University

Data: Tweets…

- Need to set up a twitter developer account
- Get an access to use API: application programming interface
- R library (twitteR)
- Keyword: ***#YellowVests***
  - n=1000
- Cleaning and preparing
- TEXT Mining Applications
- Publishing the results in a reproducible way as an open data access on my website
  - ***http://hagprojects.com***

```
> dim(TweetDF)
[1]  1000 13637
>
```

Gokturk_IST707 Project_Viral_... * | TweetDF[1:10,] | Search_YW_DF$text | Search_YW_DF | HW4_Te...

Filter

	text
6	RT @akihheikkinen: #Putin's #tourists participating to #YellowVests #GiletsJeunes today? https://t.co/IKEBwdhJtA
7	RT @enough14: 19:28 #Livestream from #Toulouse, where clashes continue. #GiletsJaunes #8Decembre #YellowVests #YellowJacket...
8	RT @Fnews_english: #French protesters in central #Paris: "#Macron leave" . #France #FranceProtest #ParisProtest #ParisRiots #Yellow...
9	RT @worldnewsevery: Paris anti-Macron movements. #GiletsJaunes #YellowVests https://t.co/HWCxV8YPQL
10	Policías de paisano infiltrados entre los manifestantes practican detenciones, una práctica habitual también en las... https://t.co/4b...
11	RT @NBbreaking: Peaceful yellow vest protester with arms in the air gets shot in the stomach with a Flash Ball by a Police officer. Prot...
12	RT @Myasiryaman: This is france! This is democratic europe ! #YellowVests #GiletsJaunes #MacronDemission #Ancona #8 صباحات_الهلال...
13	RT @EpochTimes: HAPPENING NOW: Riot police in #Paris move towards #YellowVests protesters at Saint Lazare Train Station. As nigh...
14	RT @doc_hal: Most Wanted in FRANCE 🇫🇷 #giletsjaunes #YellowVests # Yellow Jackets https://t.co/wtCv7JO2h6
15	RT @realSi_jeff: Check the flag out on this armoured Veichle which is in #France to scare and threaten protesters. This is how it start...
16	RT @peterk88: discover END PLAY WINNING BLACKJACK ... it's go time today #ArmyvsNavy #bbcfootball #BETBreaks #bettingtwitter ...
17	RT @PrisonPlanet: Looting clothing stores has nothing to do with standing up against the EU or carbon taxes. Usual suspects have infi...
18	RT @realSi_jeff: Check the flag out on this armoured Veichle which is in #France to scare and threaten protesters. This is how it start...
19	RT @MarkSleboda1: So it's OK for Western govts, media & commentariat to openly support fringe political opposition protests i...
20	RT @PorteTonAme: 📷🇫🇷#France : Bon bah #Toulouse est en train de brûler. #8Decembre #YellowVests #GiletsJaunes https://t.co/B...
21	RT @enough14: 19:36 Burning barrivades in the Avenue de Grande-Bretagne in #Toulouse. #8Decembre #Act4 #GiletsJaunes #Yello...
22	RT @Umut_Sendikasi: #MacronDefol! Adalet ve eşitlik istiyoruz! #SefaletEkenÖfkeBiçer #SarıYeleklıler #GiletsJaunes #8Decembre #YeL...
23	#YellowVests protests: protests continue on the streets of #Toulouse https://t.co/7pV4n3U7Y4
24	RT @rs_sputnik: IZ MINUTA U MINUT Broj povređenih u protestima u Parizu povećan na 55 #FranceProtest #France #YellowVests http...

Showing 5 to 25 of 1,000 entries

Console ~/Desktop/Fall 18/IST 707 Data Analytics/IST 707/

#YellowVests?

Syracuse University

Transaction sets and summary

```
 [85] "like"
 [86] "look"
 [87] "work"
 [88] "presid"
 [89] "social"
 [90] "target"
 [91] "conflictsw"
 [92] "video"
 [93] "show"
 [94] "arrest"
 [95] "stop"
 [96] "use"
 [97] "gas"
 [98] "tear"
 [99] "break"
[100] "call"
[101] "dialogu"
[102] "fresh"
[103] "pm"
[104] "right"
[105] "fire"
[106] "forc"
[107] "dessus"
[108] "drapeau"
[109] "d…"
[110] "fait"
[111] "françai"
[112] "genoux"
[113] "les"
[114] "littéral"
[115] "marcher"
[116] "par"
```

```
> df <- data.frame(term=names(term.freq), freq=term.freq)
> str(df)
'data.frame':   250 obs. of  2 variables:
 $ term: Factor w/ 250 levels "…","\U0001f4a5now\U0001f4a5",..: 24 27 34 67 68 107 108
139 155 179 ...
 $ freq: num  40 32 53 42 31 48 31 31 86 280 ...
```

```
> summary(df)
                                                 term          freq
 …                                                  :  1   Min.   : 10.00
 \U0001f4a5now\U0001f4a5                            :  1   1st Qu.: 12.25
 \U0001f534\U0001f4f9\U0001f1eb\U0001f1f7           :  1   Median : 17.00
 \U0001f645\U0001f3fb♀ \U0001f1eb\U0001f1f7:  1   Mean   : 32.61
 à                                                  :  1   3rd Qu.: 31.00
 abdullahciftcib                                    :  1   Max.   :641.00
 (Other)                                            :244
```

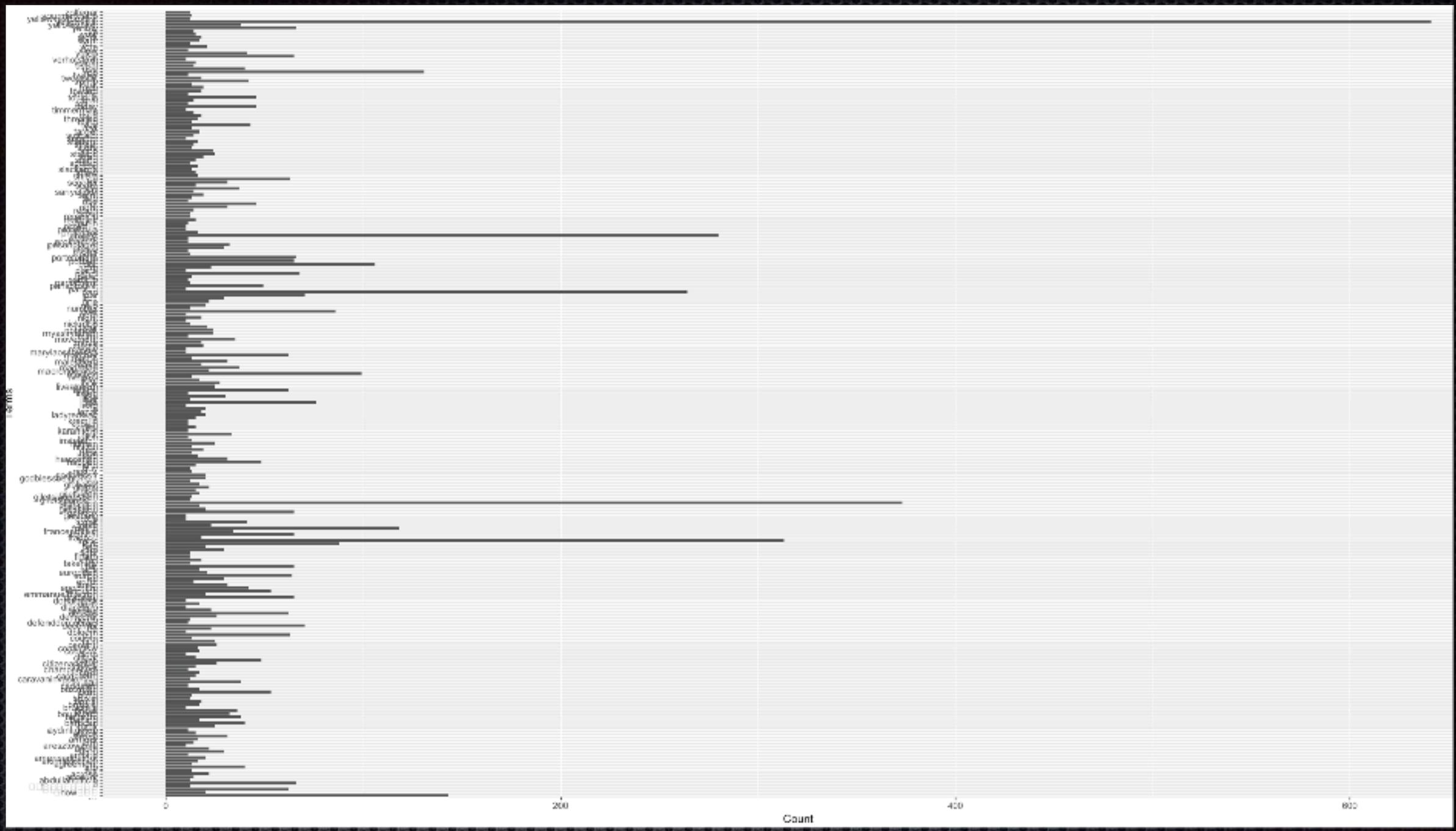
```
writing ... Error in arules::apriori(TweetTrans, parameter = list(support = 0.01,  :
  not enough memory. Increase minimum support!
In addition: Warning message:
In arules::apriori(TweetTrans, parameter = list(support = 0.01,  :
  Mining stopped (time limit reched). Only patterns up to a length of 4 returned!
```

	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1	c(rt	epochtimes	happening	now	yellowvests	protesters	burn	scooter	erecting	
2		umut_sendikasi	kapitalizme	ve	faşizme	karşı	direnen	paris	halkının	
3		citizenanalyst	scene	champs	élysées	stunning	guy	just	ran	
4		ruptly	yellow	vest	movement	protesters	continue	fourth	week	
5		ladyredwave	voted	trump	americanpatriots	say	nwo	amp	now	
6		akihheikkinen	putin's	tourists	participating	yellowvests	giletsjeunes	today		
7					livestream	toulouse	clashes	continue	giletsjaunes	
8		fnews_english	french	protesters	central	paris	macron	leave	france	
9		worldnewsevery	paris	anti	macron	movements	giletsjaunes	yellowvests		
10	policías	de	paisano	infiltrados	entre	los	manifestantes	practican	detenciones	
11		nbbreaking	peaceful	yellow	vest	protester	arms	air	gets	
12		myasiryaman	france	democratic	europe	yellowvests	giletsjaunes	macrondemission	ancona	
13		epochtimes	happening	now	riot	police	paris	move	towards	
14		doc_hal	wanted	france	giletsjaunes	yellowvests	yellow	jackets		
15		realsi_jeff	check	flag	armoured	velchle	france	scare	threaten	
16			discover	end	play	winning	blackjack	it's	go	
17		prisonplanet	looting	clothing	stores	nothing	standing	eu	carbon	
18		realsi_jeff	check	flag	armoured	velchle	france	scare	threaten	
19			ok	western	govts	media	amp	commentariat	openly	

Showing 1 to 20 of 20 entries

Cleaned Data

Syracuse University



Cleaned Data, *word counts*

Syracuse University

Frequency words and association

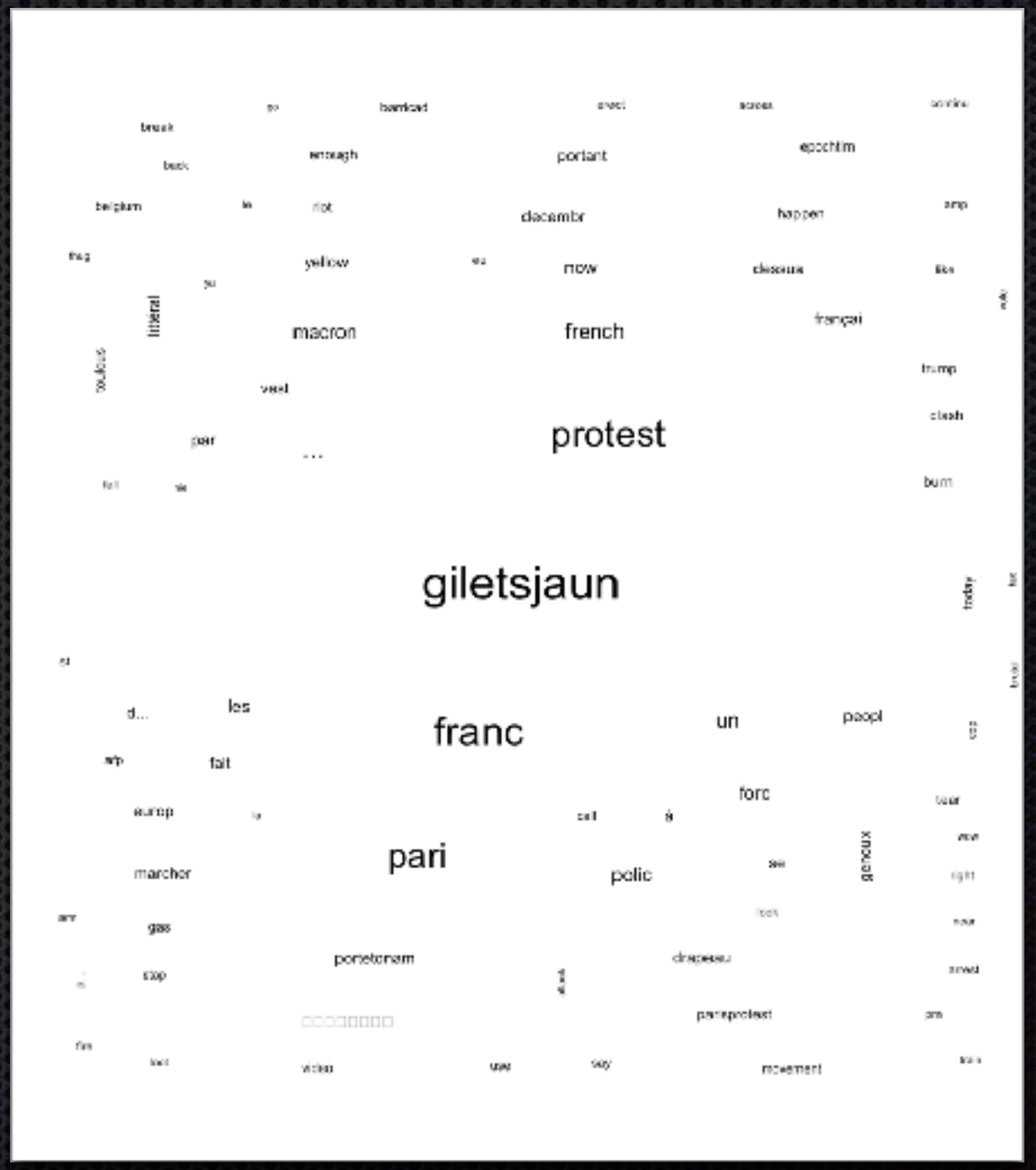
```
> ## which words are associated with 'yellowvest'
> findAssocs(tdm, "yellowvest", corlimit = 0.20)
$yellowvest
alexandrosm     bruxell        demo     brussel     belgium     gelehe…
       0.25        0.24        0.24        0.23        0.22        0.22
    belgiqu   giletjaun
       0.21        0.21
```

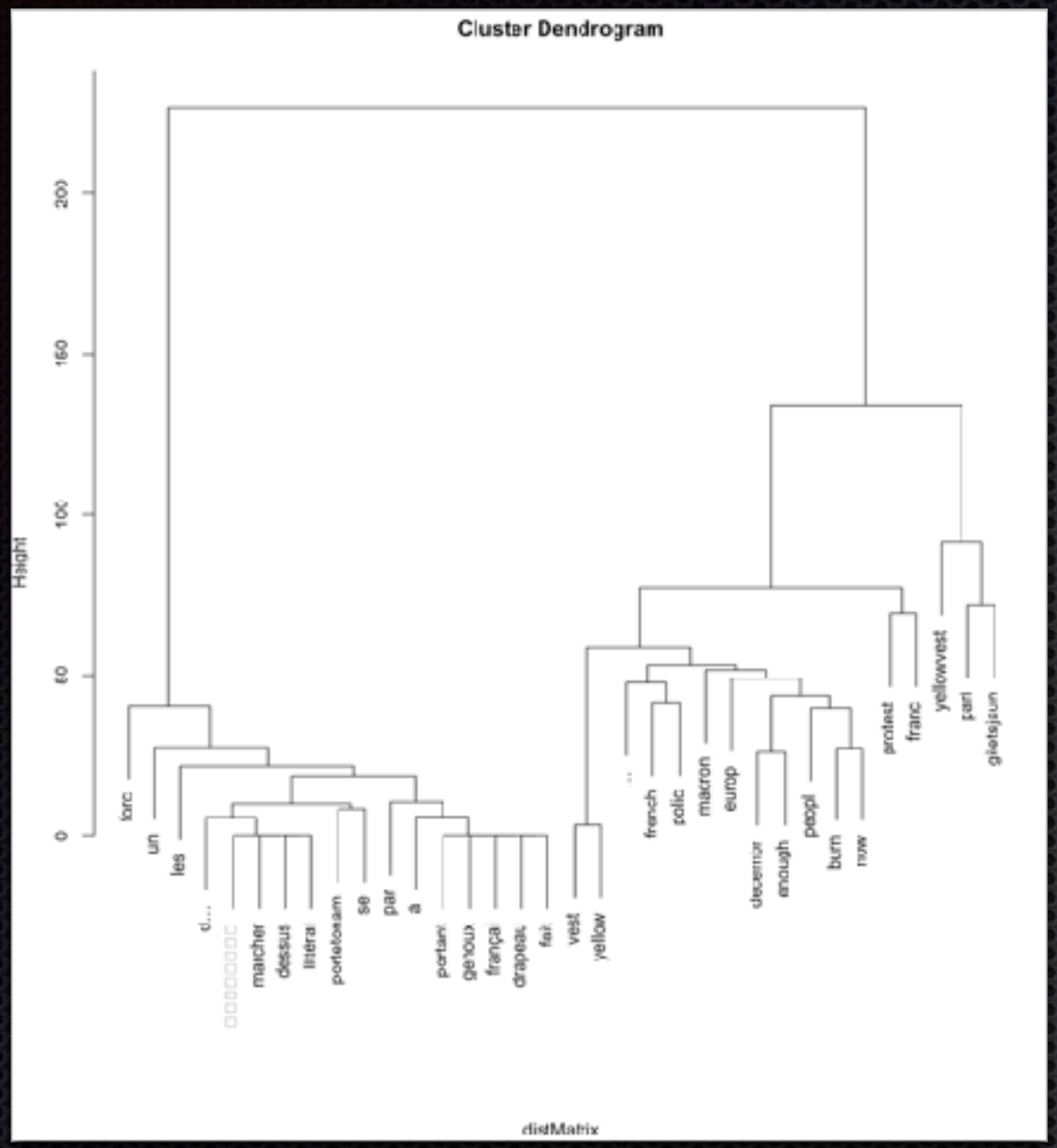
```
> findAssocs(tdm, "macron", corlimit = 0.15)
$macron
      prefer         one   tweetsdjt      presid       trump
        0.60        0.47        0.46        0.40        0.34
      social          w…      among defenddemocraci       known
        0.32        0.30        0.28        0.28        0.28
     kremlin  prokremlin      promot        top…     twitter
        0.28        0.28        0.28        0.28        0.28
        view        cake         eat       thank     account
        0.28        0.26        0.26        0.26        0.25
      target     violenc      dbloom        act…     legitim
        0.24        0.24        0.24        0.23        0.23
      result           s       media     attempt   globalist
        0.23        0.23        0.20        0.19        0.19
        want      listen       armor  champelysé   markvallen
        0.17        0.17        0.17        0.17        0.17
yourmarklubb          m…
        0.15        0.15
```

Word Cloud

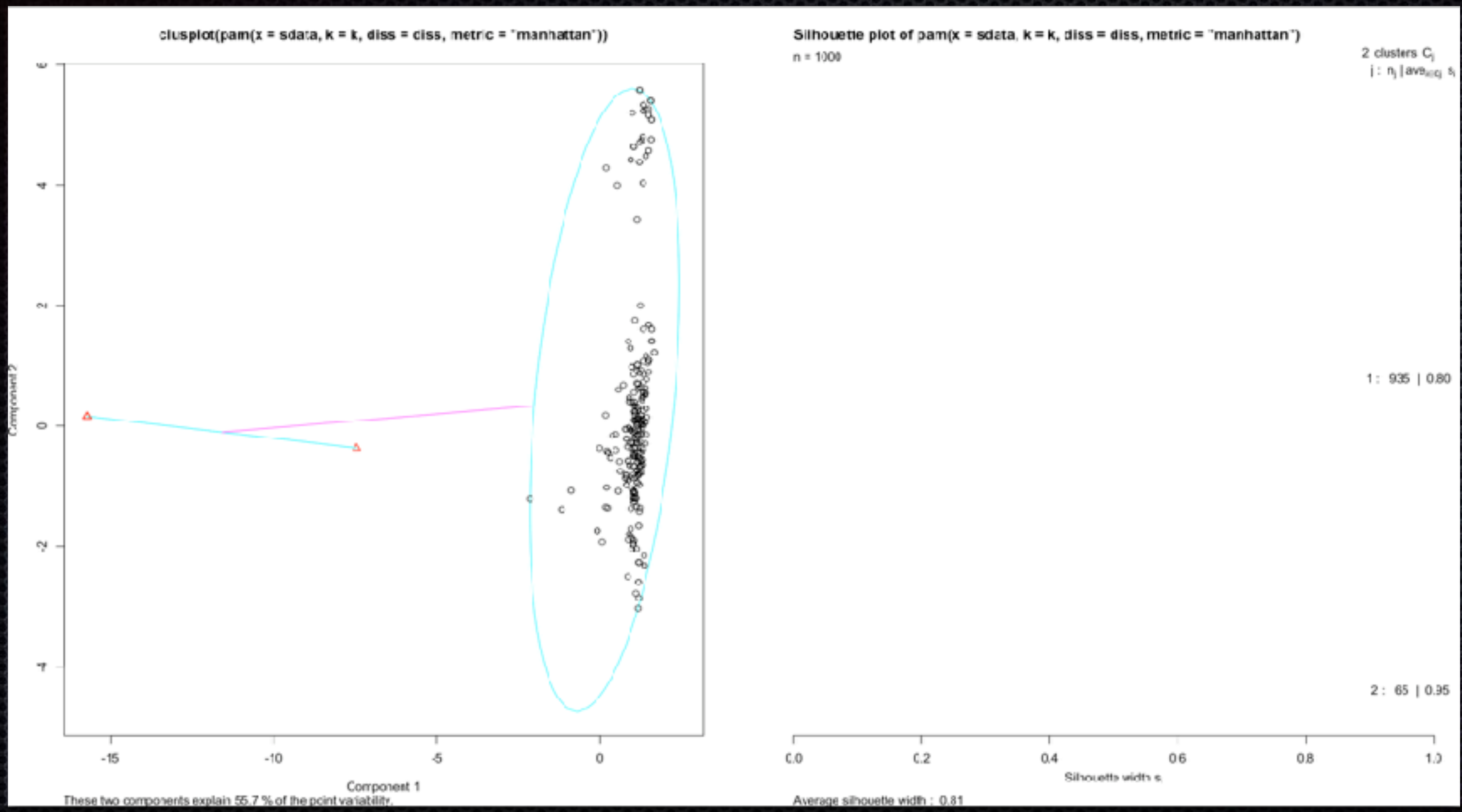
```
> ## which words are associated with 'yellowvest'
> findAssocs(tdm, "yellowvest", corlimit = 0.20)
$yellowvest
alexandrosm     bruxell        demo     brussel     belgium     gelehe…
       0.25        0.24        0.24        0.23        0.22        0.22
    belgiqu   giletjaun
       0.21        0.21
```

```
> dim(TweetDF)
[1]  1000 13637
>
```




Cluster Dendrogram
Height
0
50
100
150
200
distMatrix
hclust (*, "ward.D")

Clustering

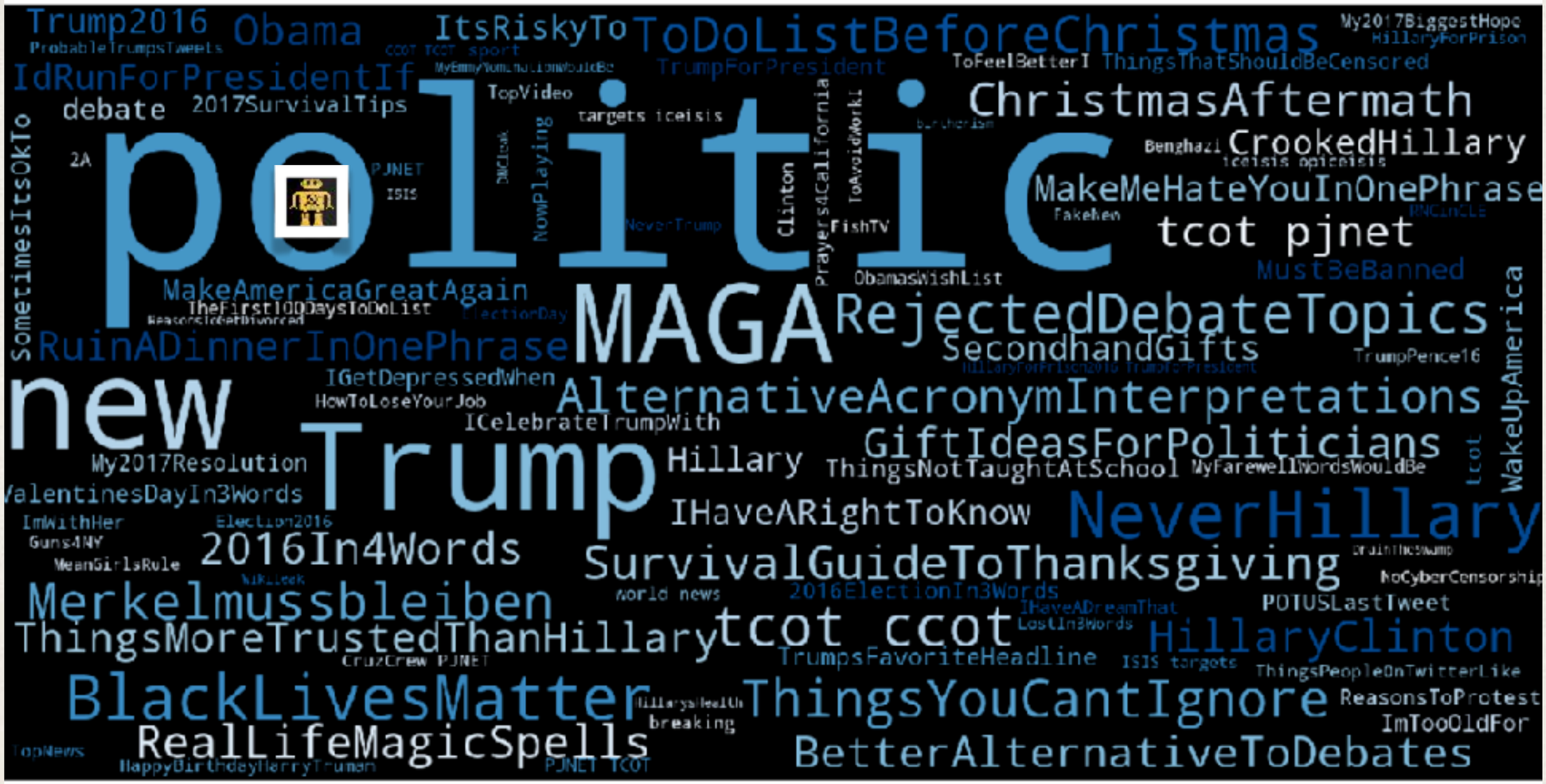


Clustering

Syracuse University

Conclusion

- TheText mining is a powerful data science tool to analyze social media such as Twitter and Facebook.
- Text Mining can be classified into
  - text clustering, text categorization, association rule mining and trend analysis
  - text mining has progressed well on #YellowVests.



Project II: Identifying Russian Troll Accounts on Twitter

IST 718 – Big Data Analytics, Course Instructor: Dr. Jon Fox

Project Partners: Drew Howell & Scott Snow, ***Spring 2019***

Story & Hypothesis

- According to the House Intelligence Committee investigation, Russia's Internet Research Agency attempted to interfere with the 2016 U.S. election by running fake accounts on twitter known as **"Russian trolls"**.

- Our hypothesis Questions
  - Would it be possible to demonstrate whether these tweets used to manipulate/target the 2016 U.S. election to favor on any presidential candidate?
  - Would it be possible by developing machine learning models to predict whether a Twitter account is a Russian troll/fake?

Russian Troll Tweets

Data

Source:

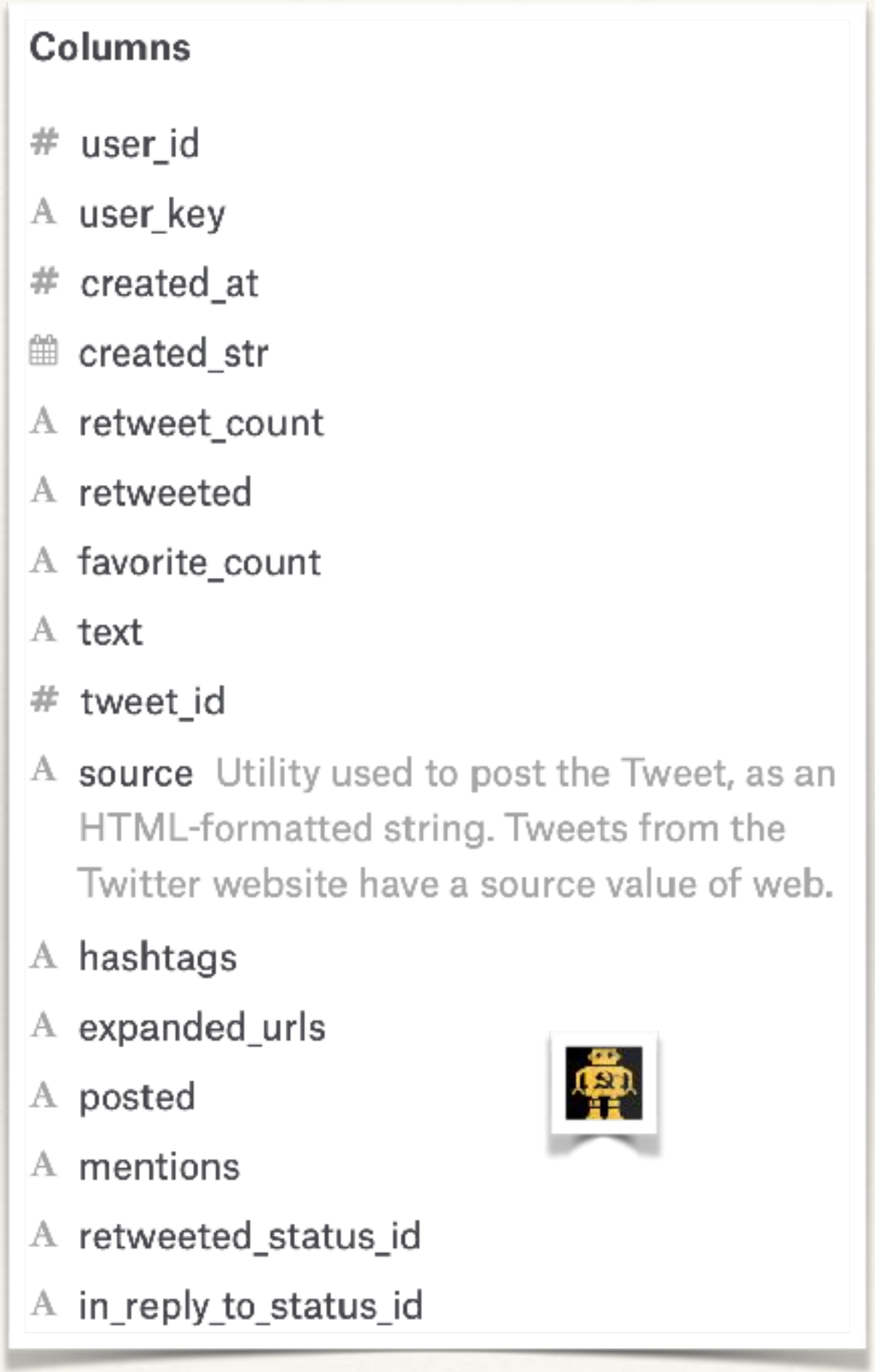
- NBC News
  - https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731
- Kaggle - Russian Troll Tweets
  - https://www.kaggle.com/vikasg/russian-troll-tweets

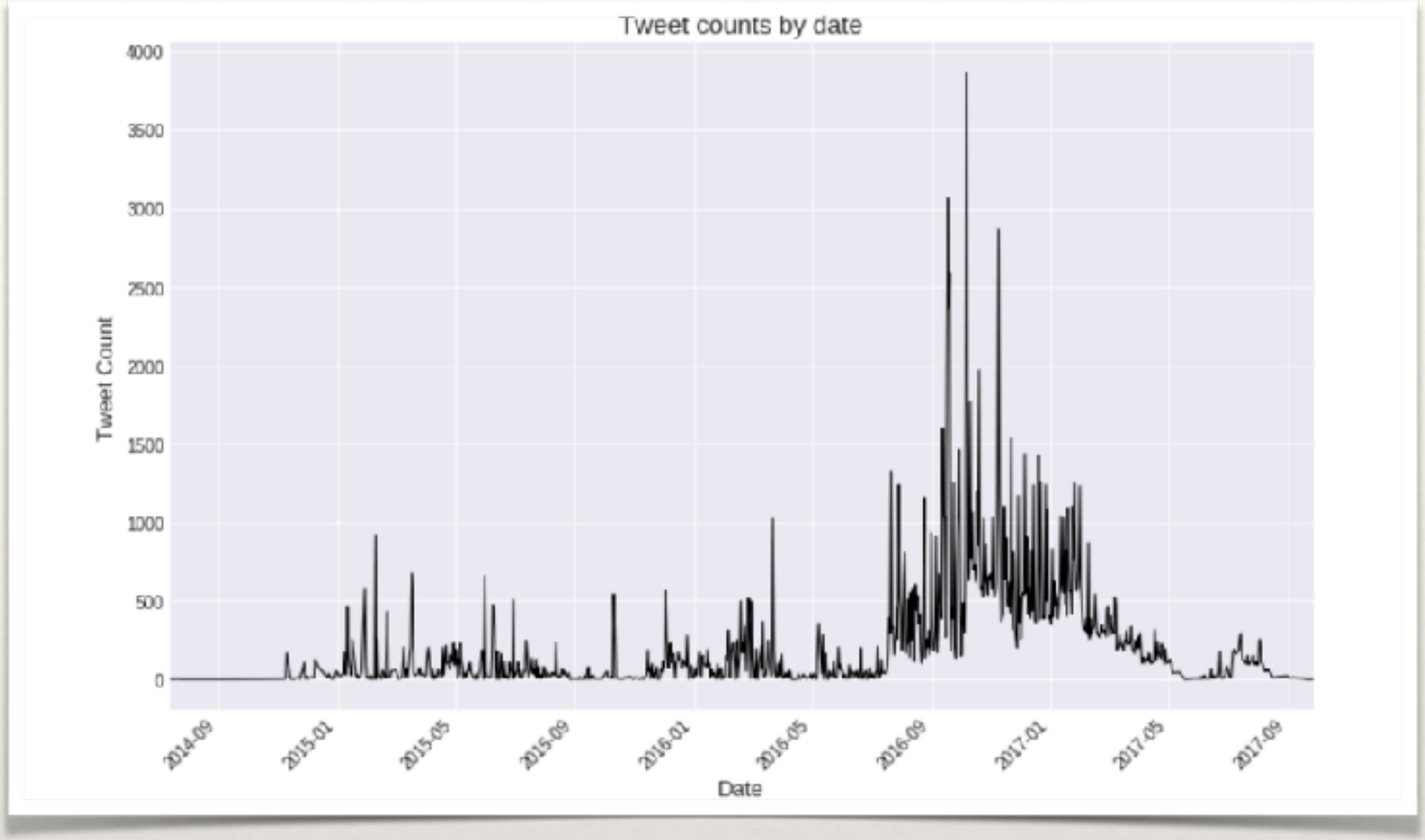
Dimension:

- 203,482 tweets
- 16 variable columns

Data Sets:

- tweets.csv
- users.csv

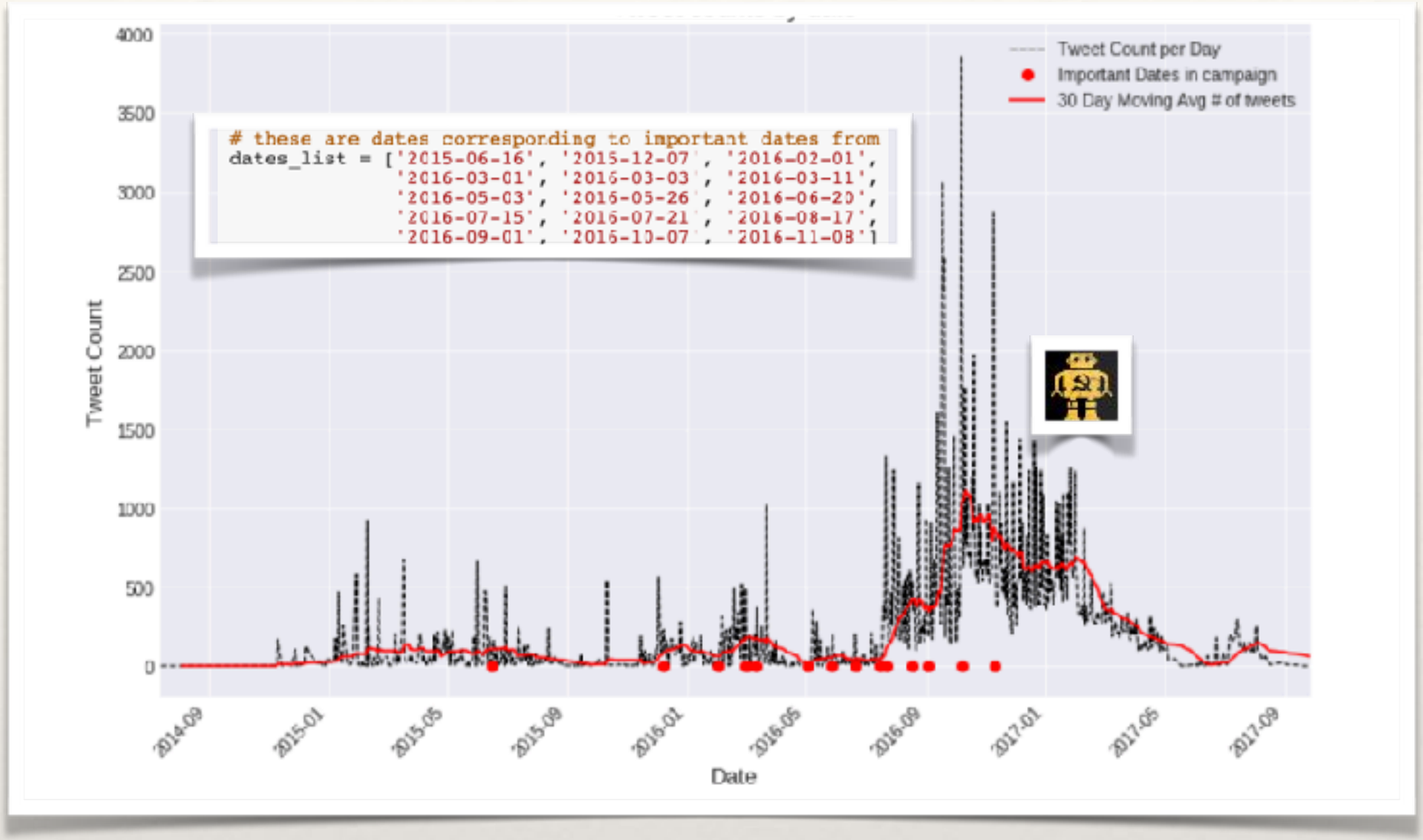




Time line of the tweets

Exploratory Analysis

3 years of tweets starts on **July 14th, 2014 and ends on of September 26th, 2017**



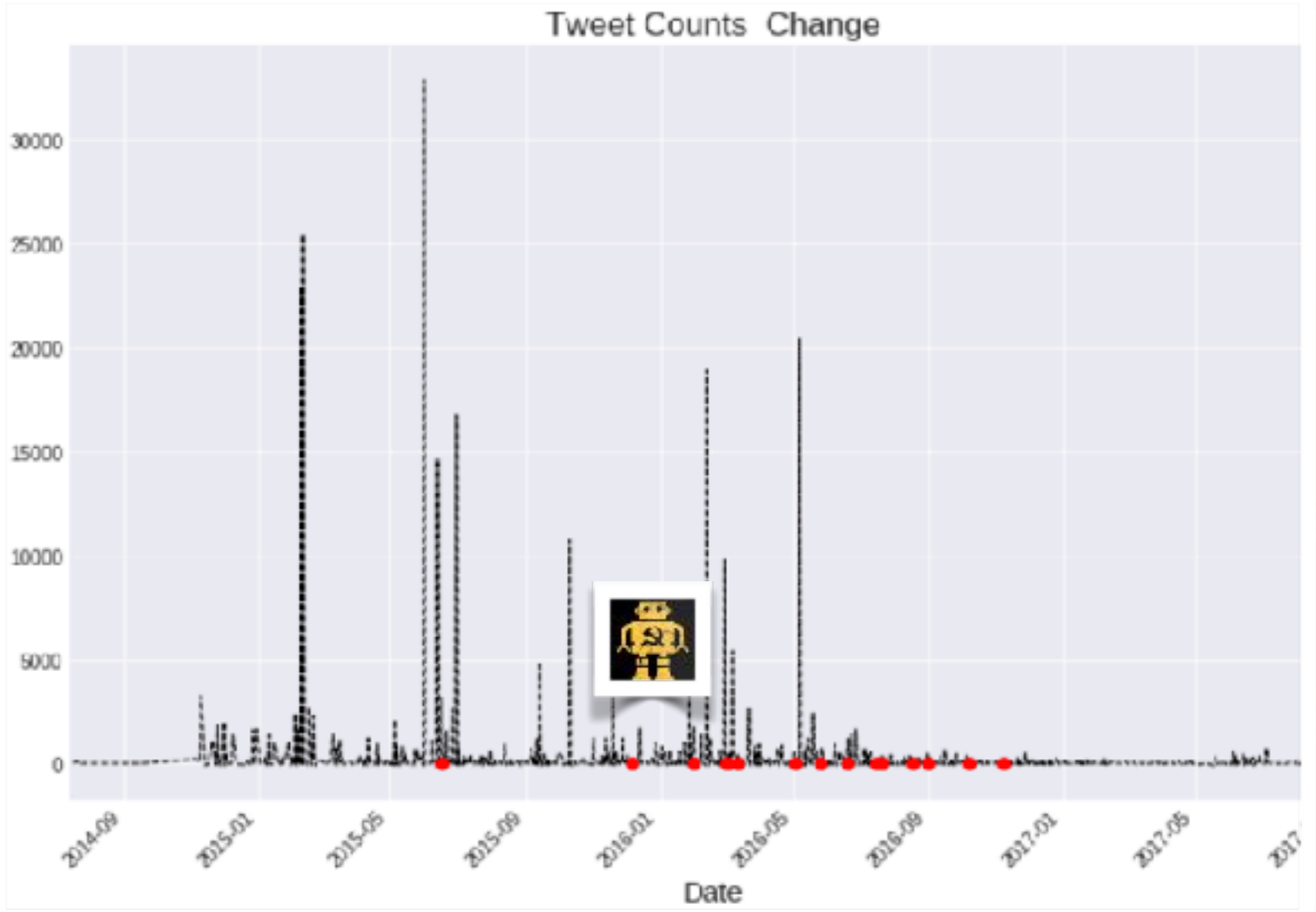

```
# these are dates corresponding to important dates from
dates_list = ['2015-06-16', '2015-12-07', '2016-02-01',
              '2016-03-01', '2016-03-03', '2016-03-11',
              '2016-05-03', '2016-05-26', '2016-06-20',
              '2016-07-15', '2016-07-21', '2016-08-17',
              '2016-09-01', '2016-10-07', '2016-11-08']
```

Time line of the tweets comparison to the Trump's campaign important event days

Time Line Comparison

Relatively similar tweet trend activity between the Campaign and the troll accounts.

	tweet_count	Pct_Chg_tweets
2015-06-16	3	50.000000
2015-12-07	219	204.166667
2016-02-01	18	1700.000000
2016-03-01	143	-71.052632
2016-03-03	6	-92.105263
2016-03-11	64	-69.523810
2016-05-03	38	216.666667
2016-05-26	6	-50.000000
2016-06-20	201	1156.250000
2016-07-15	47	17.500000
2016-07-21	1327	349.830508
2016-08-17	534	20.270270
2016-09-01	337	-63.918630
2016-10-07	2222	-42.450142
2016-11-08	2867	145.042735



Time line of the tweets comparison to the Trump's campaign important event days

Percentage Change in Tweet Counts

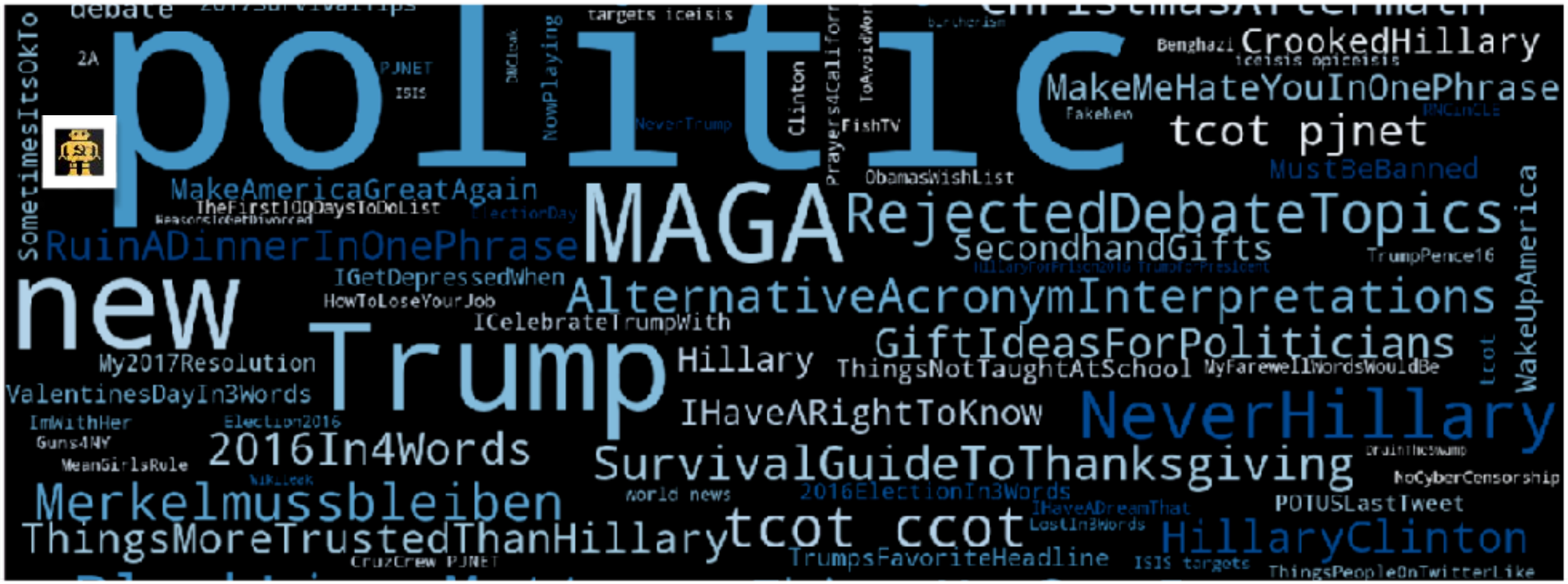
The US president was elected on **November 8th, 2016** - the last red dot on the chart.
The tweet activity near the end of the campaign was **increased**.

```
0    #IslamKills Are you trying to say that there w...
1    Clinton: Trump should‘ve apologized more, atta...
2    RT @ltapoll: Who was/is the best president of ...
3    RT @jww372: I don't have to guess your religio...
4    RT @Shareblue: Pence and his lawyers decided w...
5                                  @ModicaGiunta me, too!
6    RT @MDBlanchfield: You’ll never guess who twee...
7    RT @100PercFEDUP: New post: WATCH: DIAMOND AND...
8    RT @AriaWilsonGOP: 3 Women Face Charges After ...
9    One of the ways to remind that #BlackLivesMatt...
```

Sample troll tweets

Text Analytics

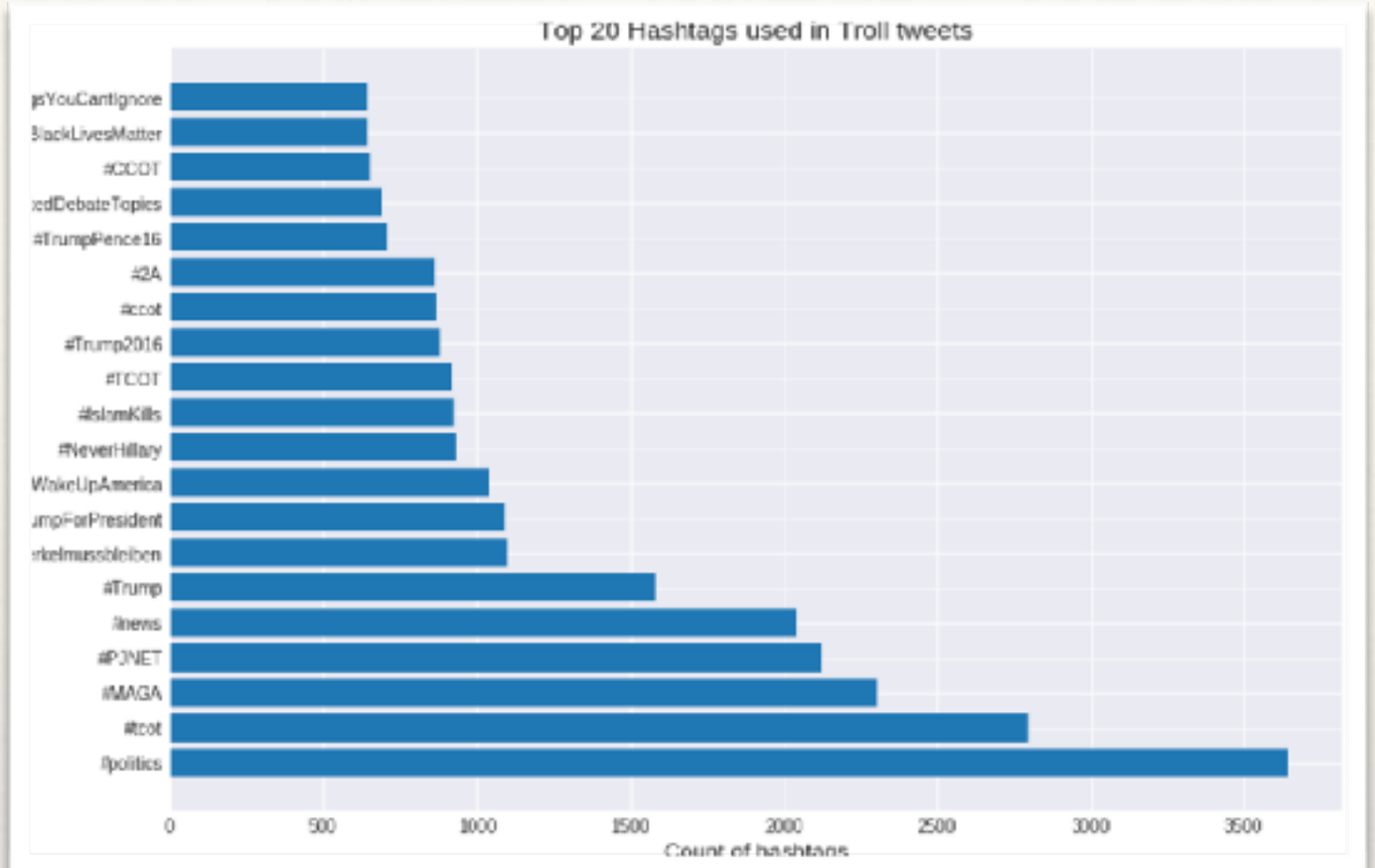
Need to **scrub** some column features from the data, such as **RT mentions, links, hashtags, extra spaces**



Most Commonly used words

Word Cloud

Word Cloud provides a general context of 200,000 troll tweets.



1. #POLITICS
2. #TCOT
3. #MAGA
4. #PJNET
5. #news
6. #Trump

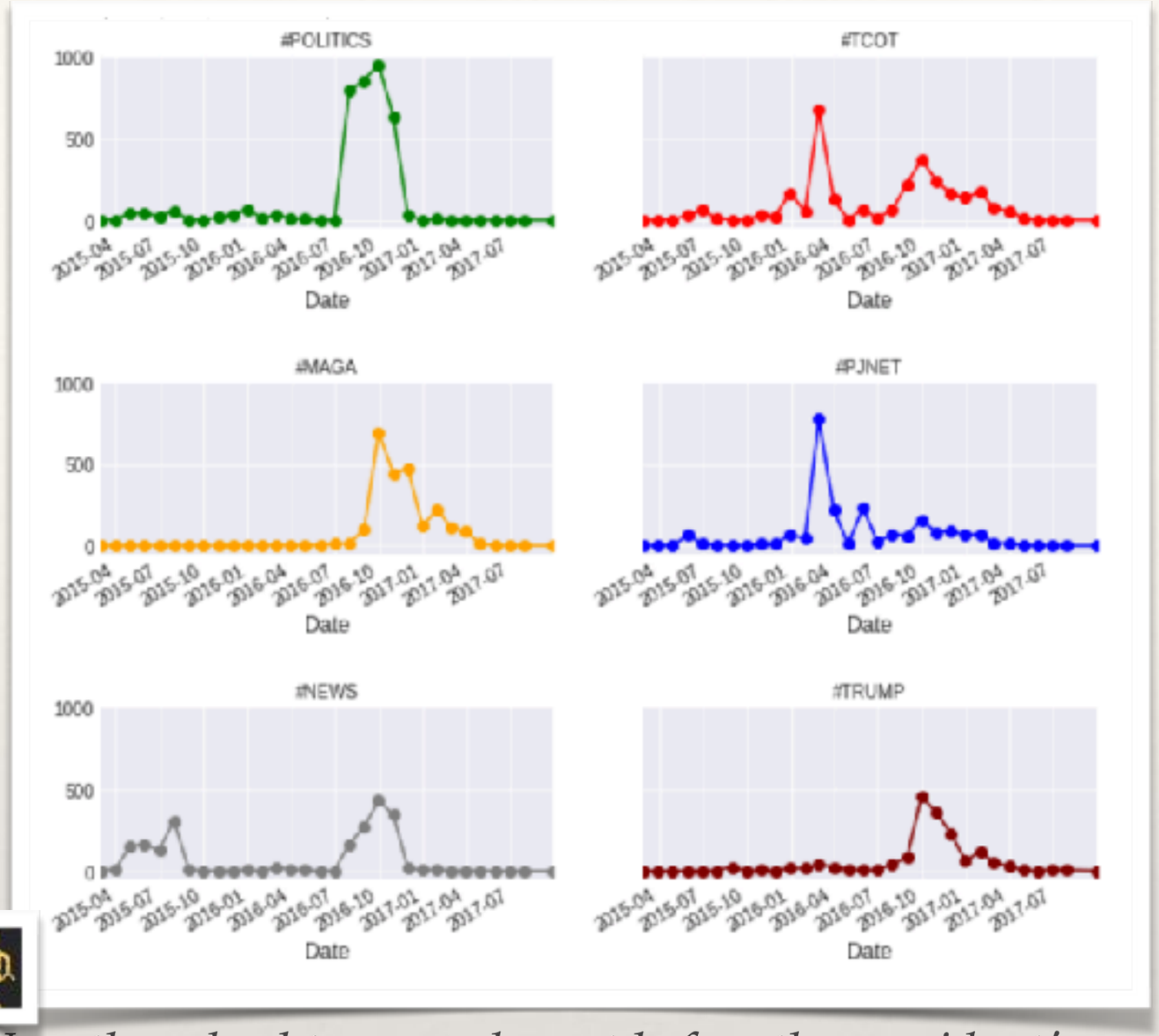
Most Commonly used hashtags

Top 20 Hashtag used in Troll Tweets

Trolls seems highly supported to the Trump's Presidential Campaign

TCOT:Top conservatives on Twitter

CCOT: Christian conservatives on twitter



1. #POLITICS
2. #TCOT
3. #MAGA
4. #PJNET
5. #news
6. #Trump

Were these hashtags used most before the president's campaign?

Prior usage of the most common hashtags

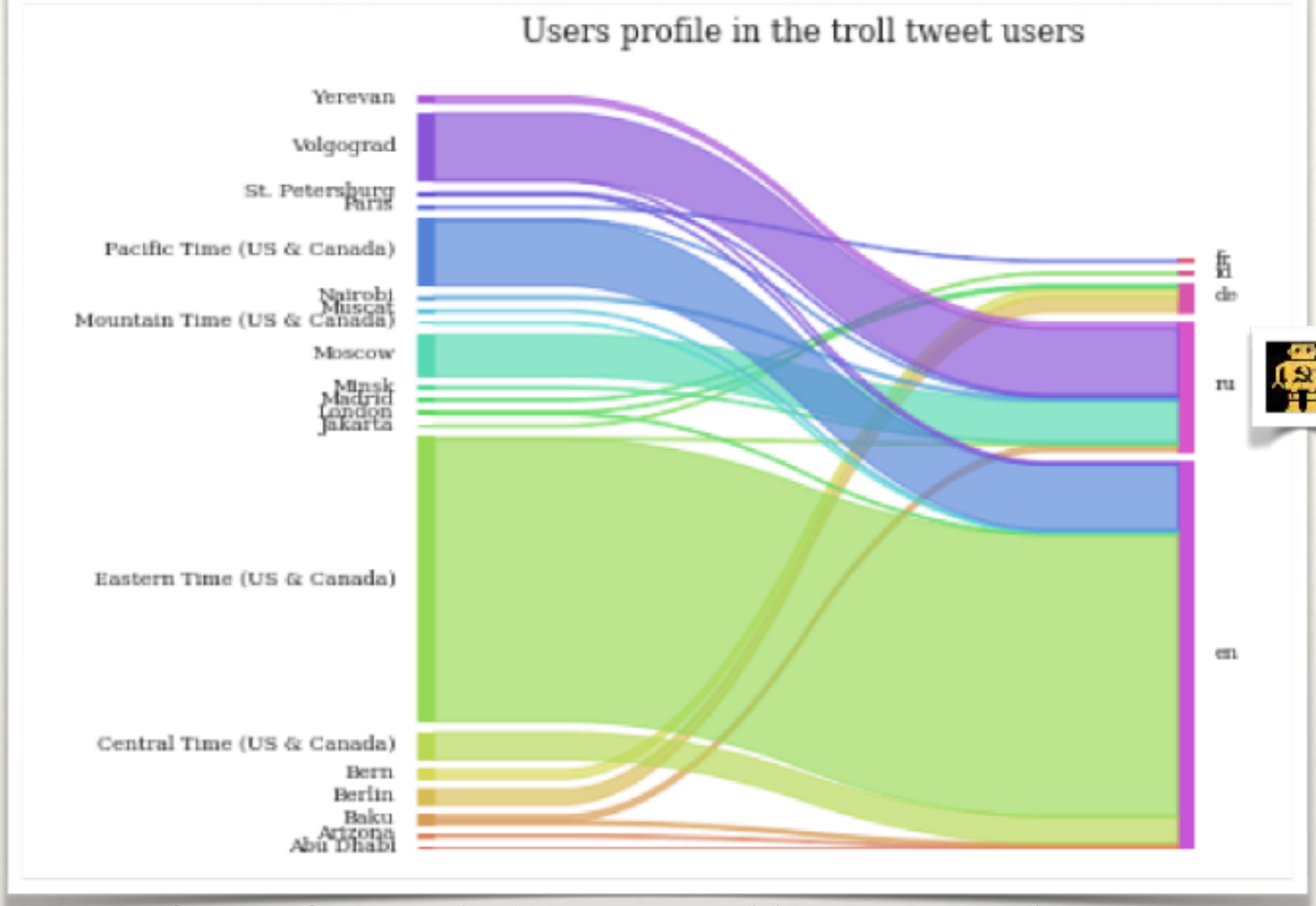
Apparently, most of these hashtags picked up in the year **2016 near March or later in July, close to the elections.**

Most used Hashtags

Clustering by HashTags

- More than 28,000 unique hashtags
- Qualified hashtags
  - tweeted > 50 times
  - 435 hashtags tweeted

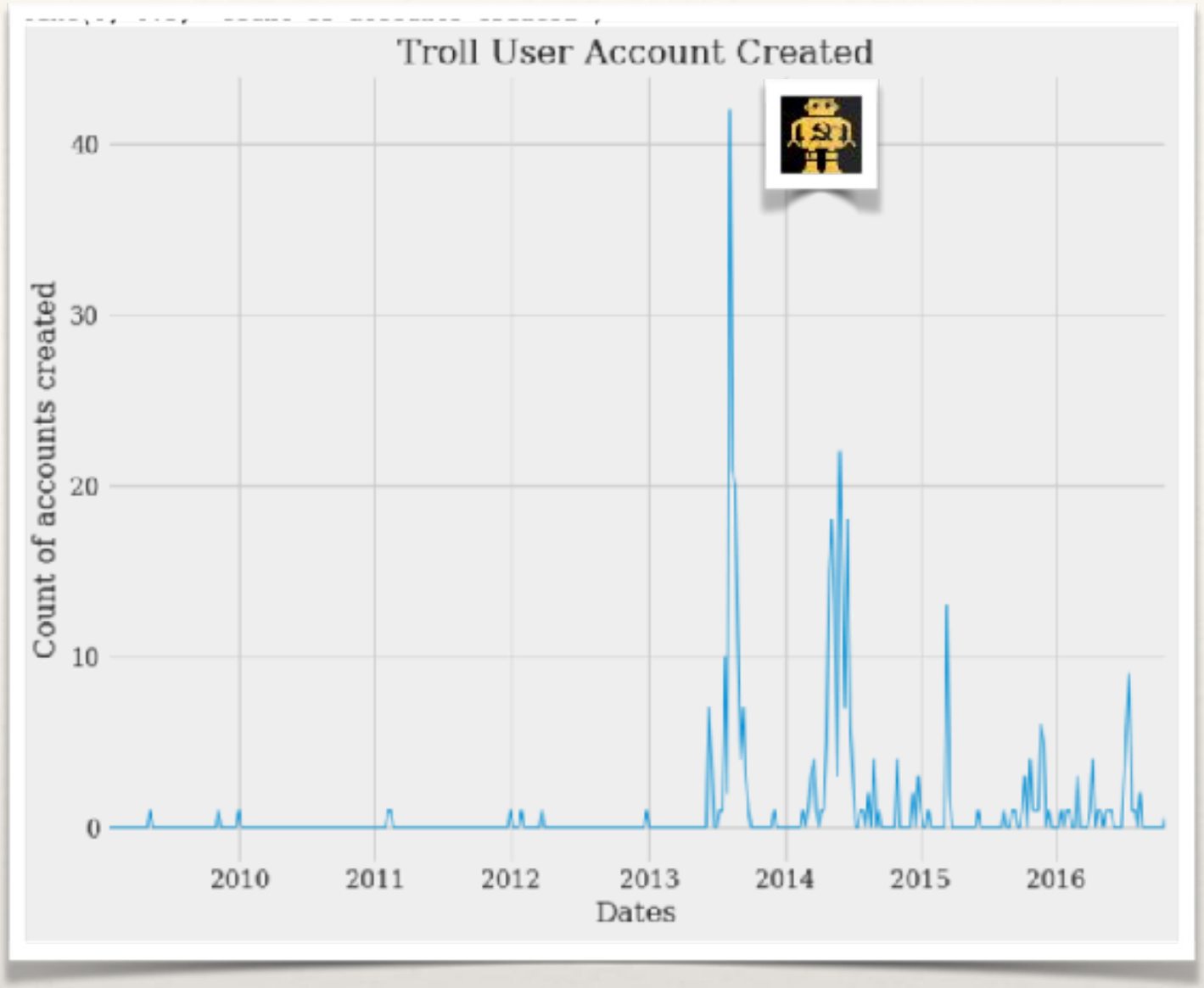
#politics	3638
#tcot	2799
#MAGA	2306
#PJNET	2121
#news	2046
#Trump	1583
#Merkelmussbleiben	1104
#TrumpForPresident	1088
#WakeUpAmerica	1038
#NeverHillary	932
#IslamKills	926
#TCOT	921
#Trump2016	882
#ccot	867
#2A	865
#TrumpPence16	710
#RejectedDebateTopics	691
#CCOT	651
#BlackLivesMatter	648
#ThingsYouCantIgnore	643



Sankey Plot - Count of users from each time-zone and language combination

Users Locations and Languages

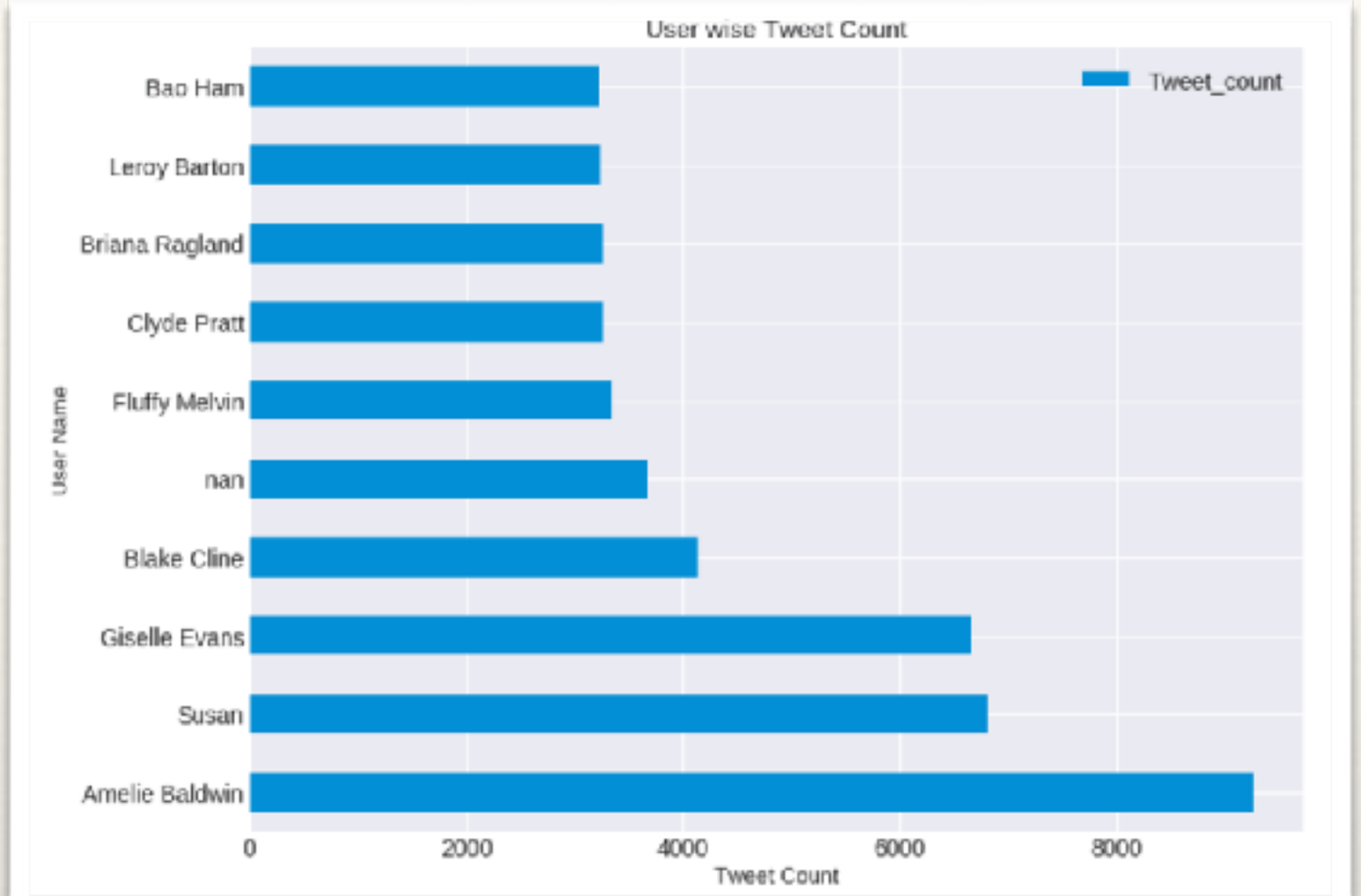
Apparently, English speaking users come from **US & Canada .**
Russian speaking users come from **Moscow, Volgograd, Yerevan and Minsk**.

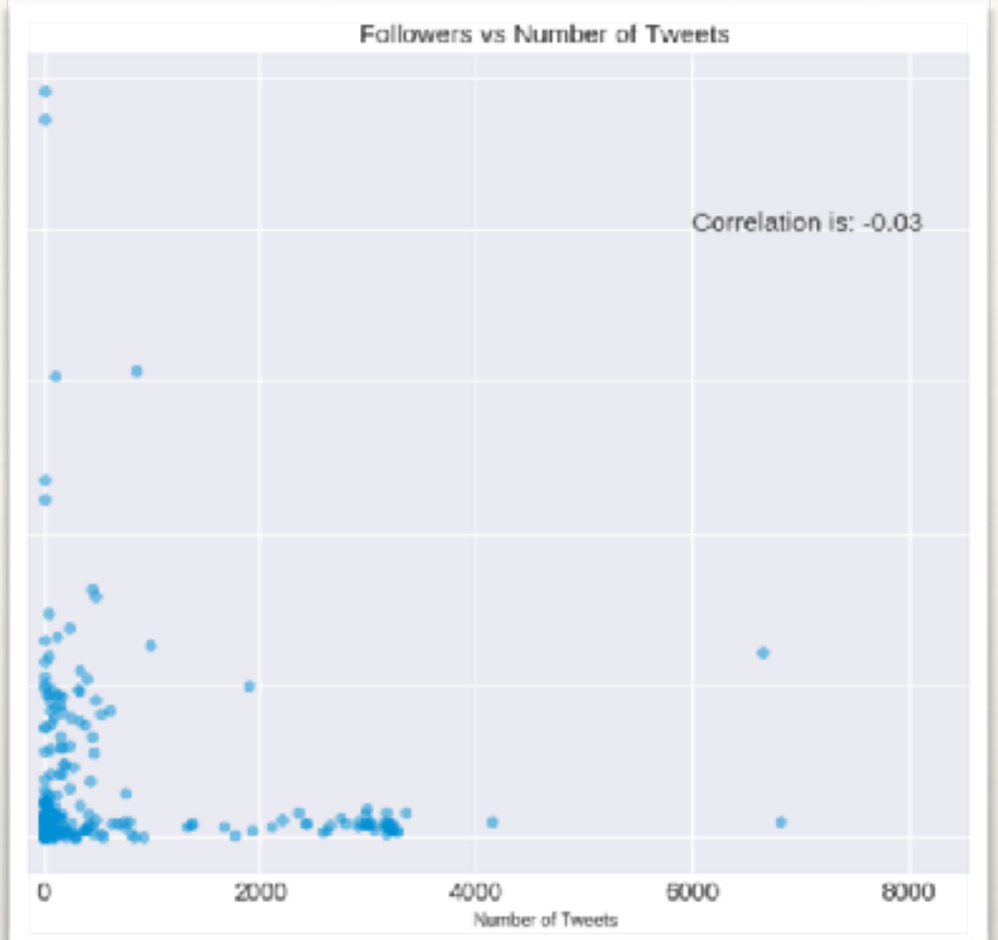


The created_at column in the users dataframe provides this information

Twitter Accounts creation dates

Most accounts were created in the second half of 2013 or first half of 2014





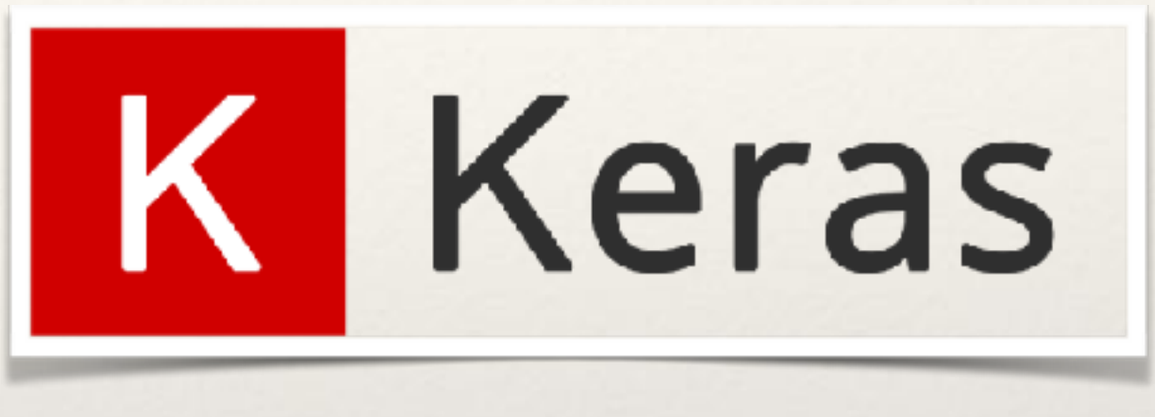
Many users have very low tweets count

Correlation between higher number of followers and large number of tweets

No such correlation exists.

Deep Learning Learning Library

Keras Model

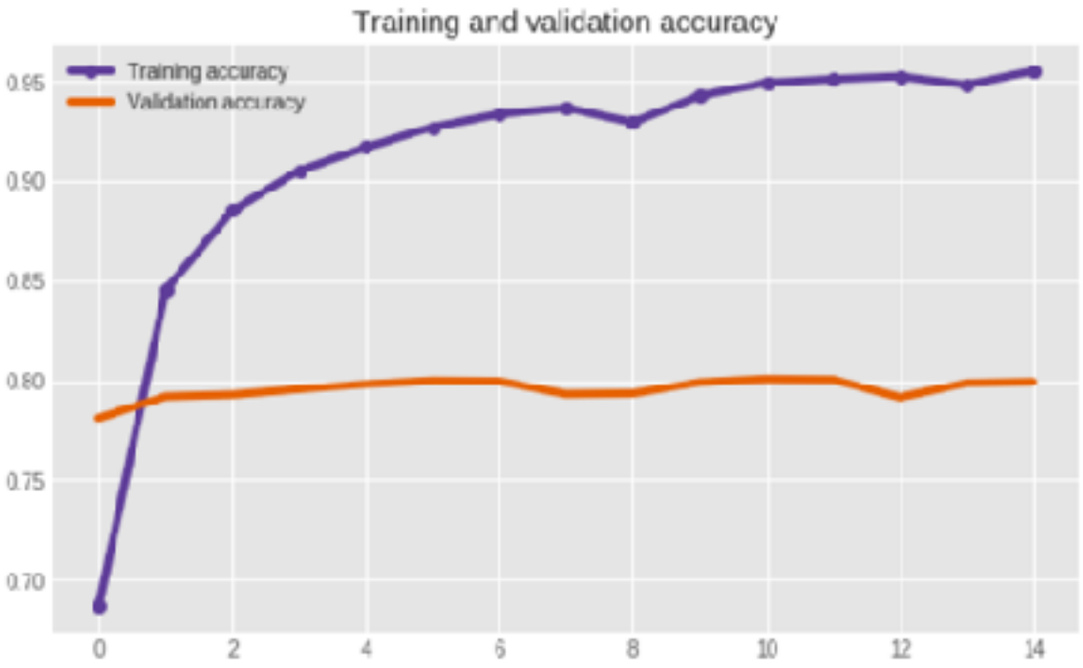


- Ran in Google Colab
- Combined Troll tweets data with generic political tweets from Election day
- Original combined size: 601,090 tweets
- Sampled due to RAM Restrictions on Colab
- Sample Size 75,136
- Train/Test Split – 50%
- 7304 features

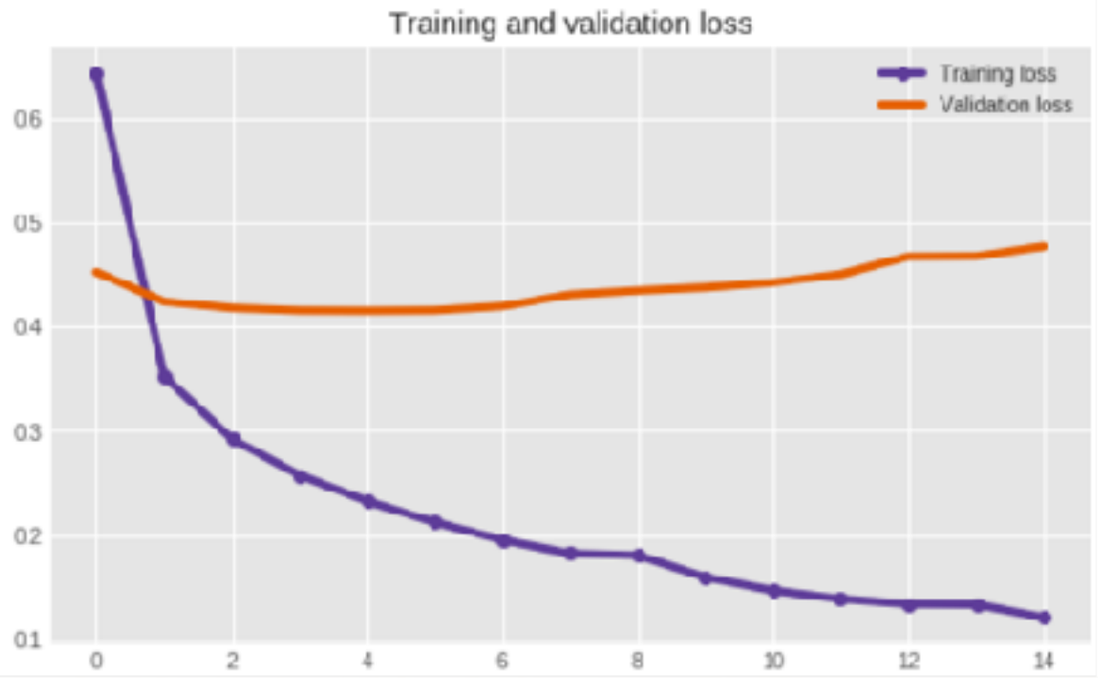
Keras Sequential Configuration

- 1 Inner Layer
  - Activation – Rectified Linear Unit
  - Input Dimensions – 7304
- Final Layer
  - Activation – Sigmoid
  - Input Dimensions – 1
- Fit configuration
  - Batch Size – 10000
  - Epochs – 16
  - Val_loss callback patience - 10

Accuracy



Loss



Keras Deep Learning Learning Library

Model Performance by Epoch

Accuracy & Loss

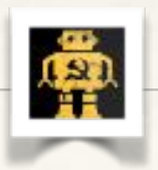
22162	2656
4880	7870

	precision	recall	f1-score	support
Real	0.82	0.89	0.85	24818
Troll	0.75	0.62	0.68	12750
accuracy			0.80	37568
macro	0.78	0.76	0.77	37568
weighted	0.80	0.80	0.79	37568

Keras Deep Learning Learning Library

Model Accuracy

- Having the additional tweets likely reduces the potential for accuracy.
- There is unfortunately a time mismatch between the data sets. All of the Real twitter posts were entirely from election day.
- More activation models and more layers will be attempted.

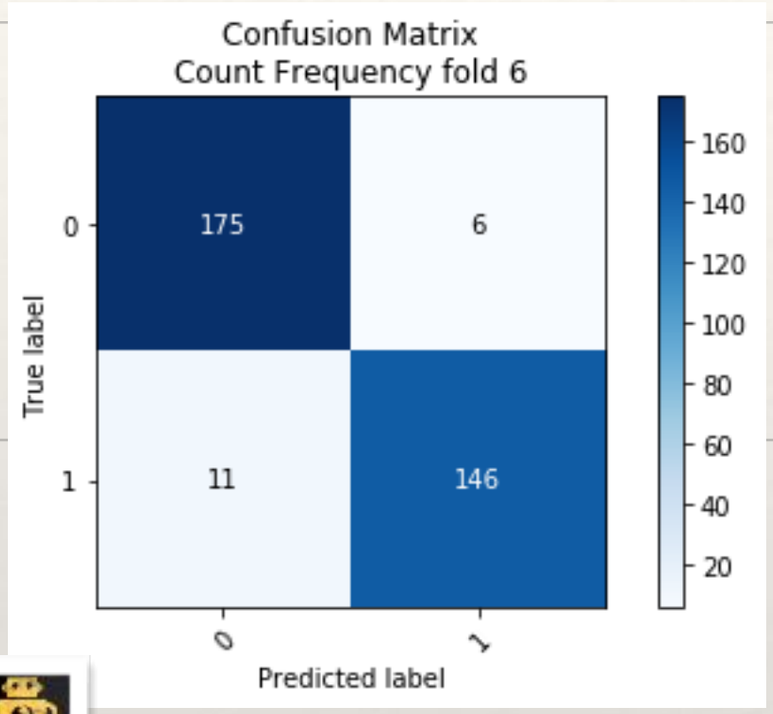


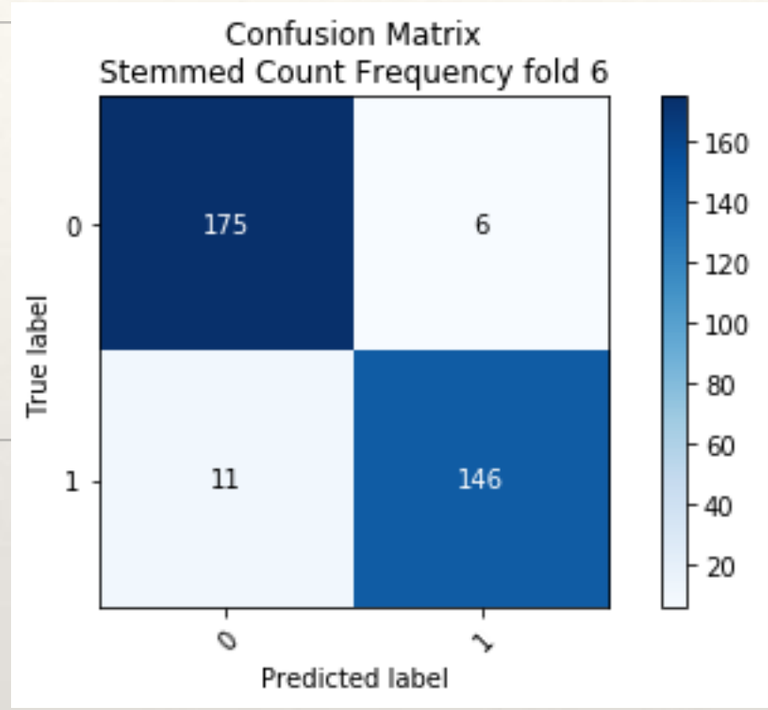
Multinominal Naive Bayes

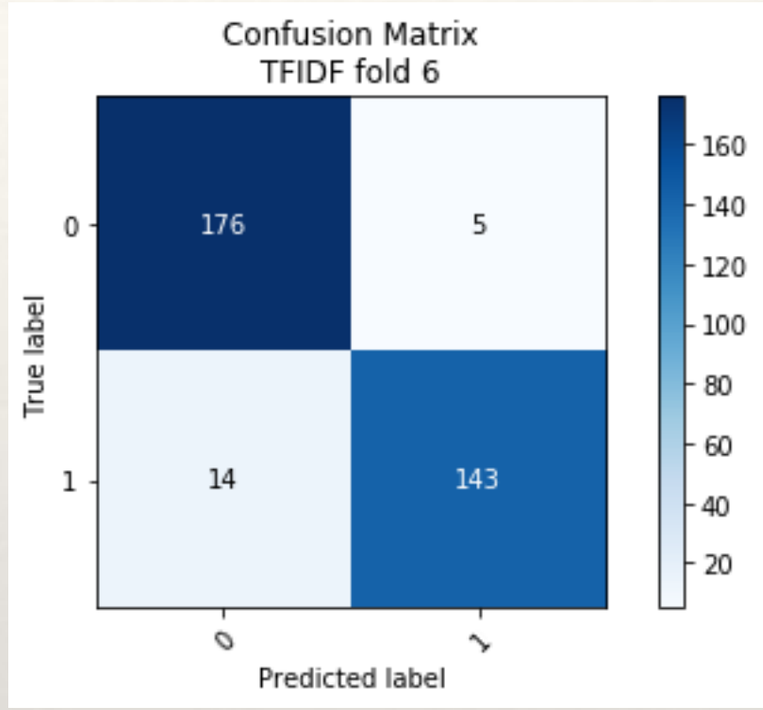
Real/Fake Prediction Average Accuracy:

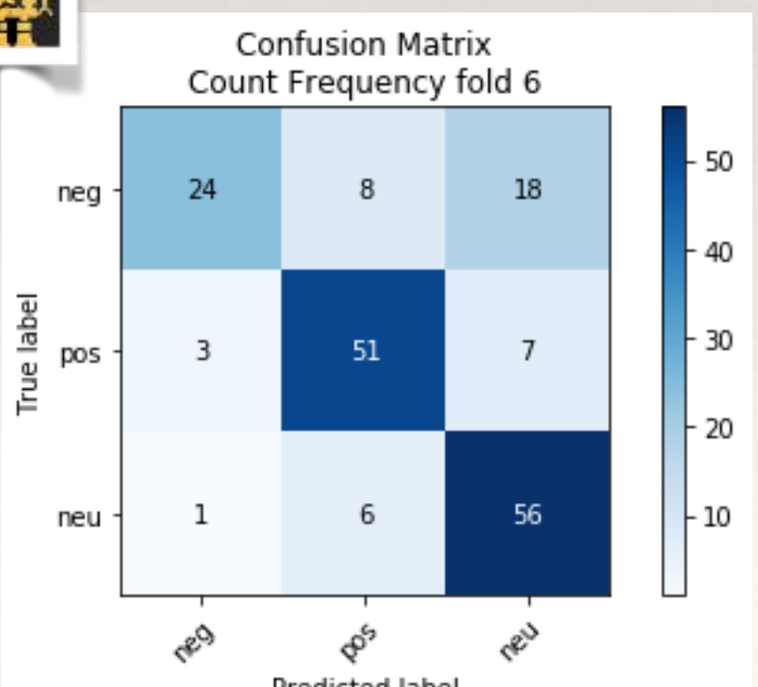
- Count frequency - 94.9%
- TF-IDF - 92.7%
- Stemmed count freq. - 94.6%

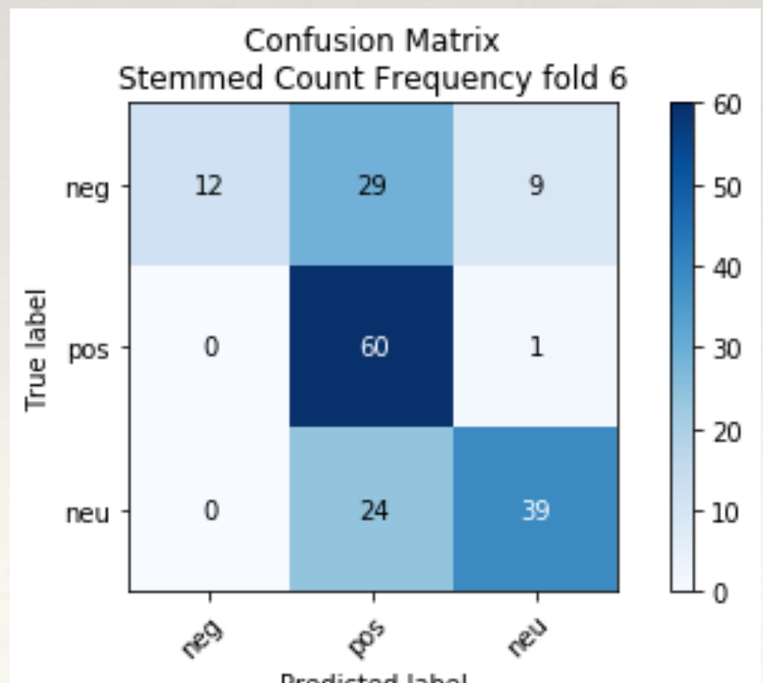
- Sentiment Prediction Average Accuracy:
  - Count frequency - 73.1%
  - TF-IDF - 73.9%
  - Stemmed count freq. - 68.6%

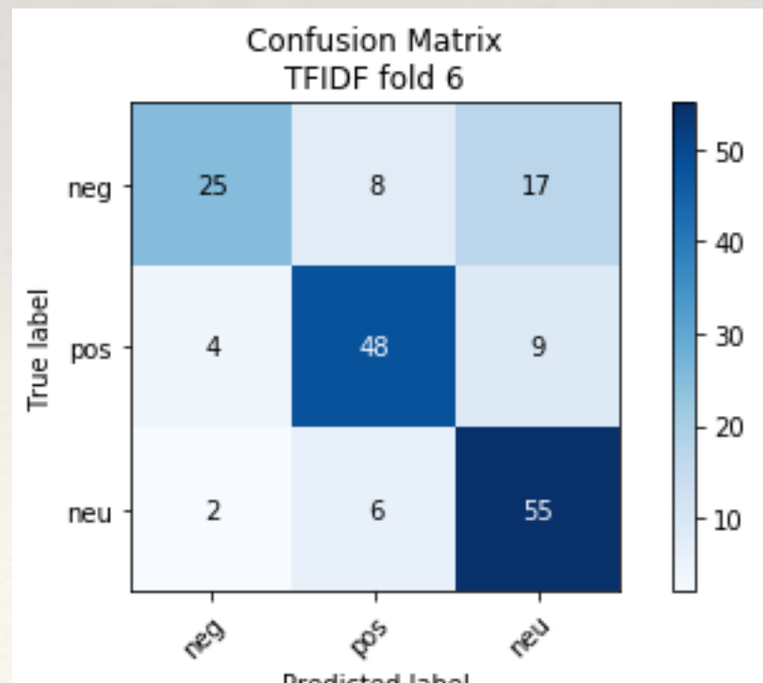










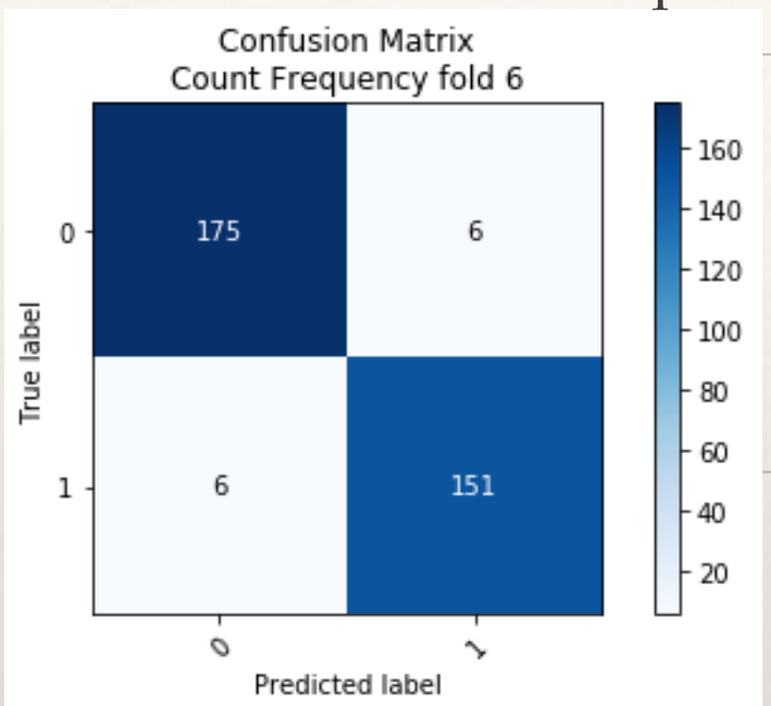


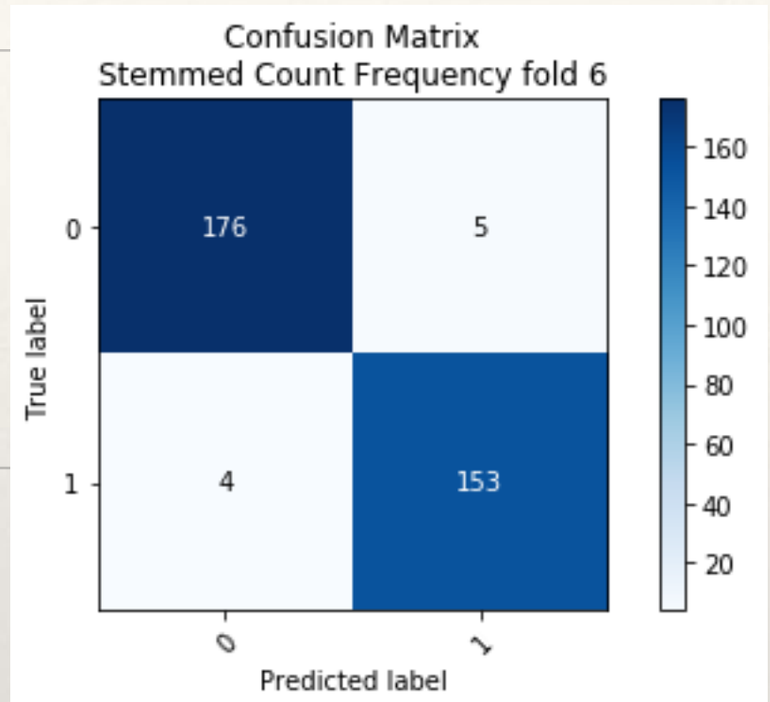
Linear SupportVector Machine

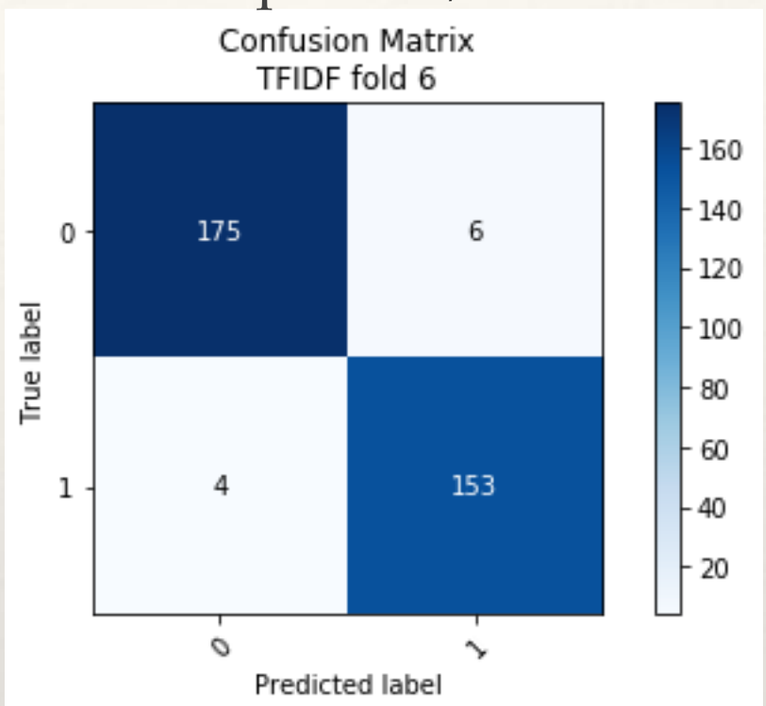
Real/Fake Prediction Average Accuracy:

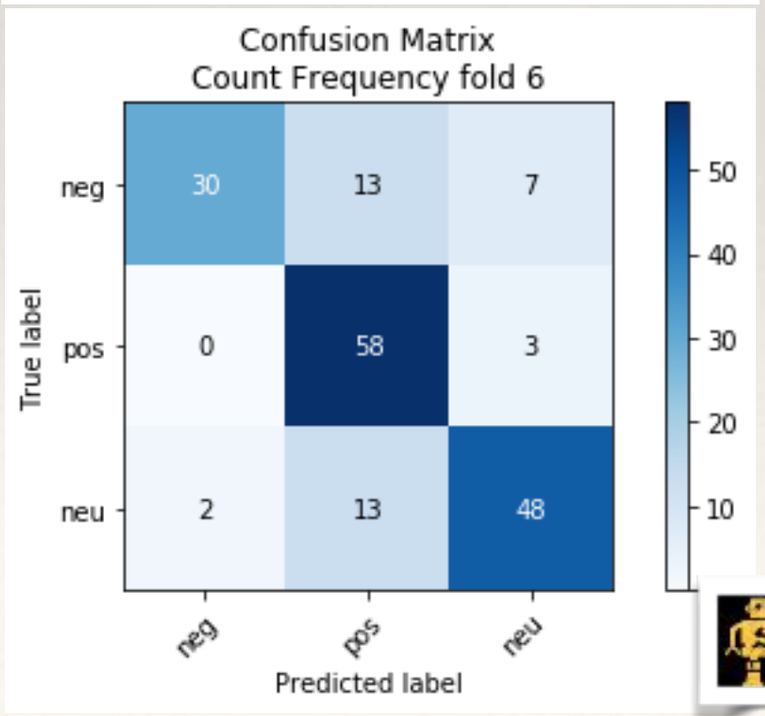
- Count frequency - 96.2%
- TF-IDF - 96.4%
- Stemmed count freq. - 96.3%

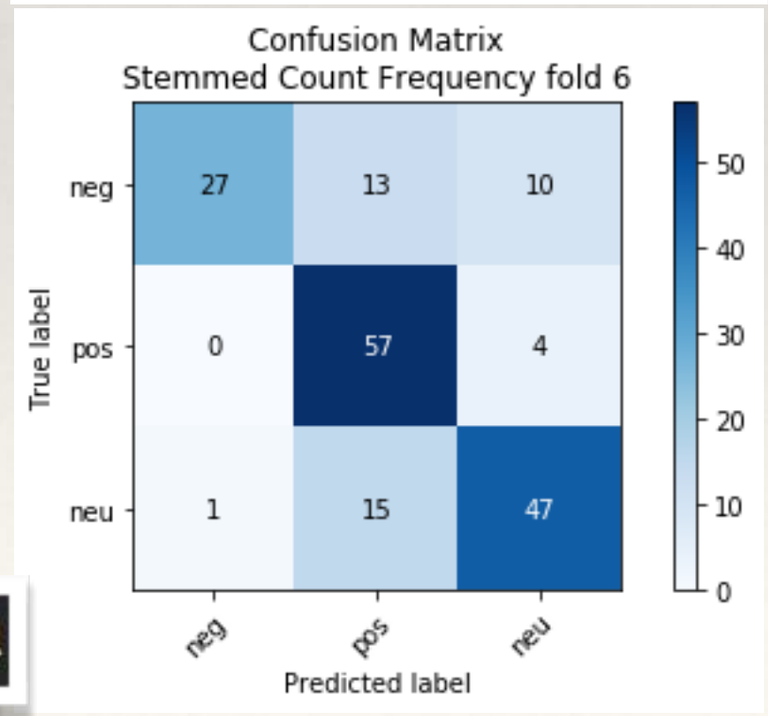
- Sentiment Prediction Average Accuracy:
  - Count frequency - 78.4%
  - TF-IDF - 80.0%
  - Stemmed count freq. - 77.2%

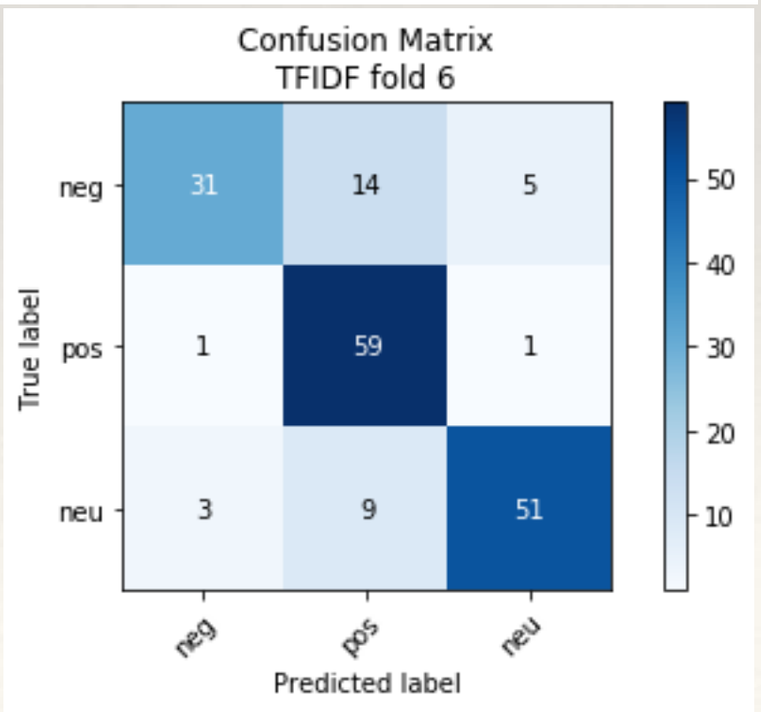






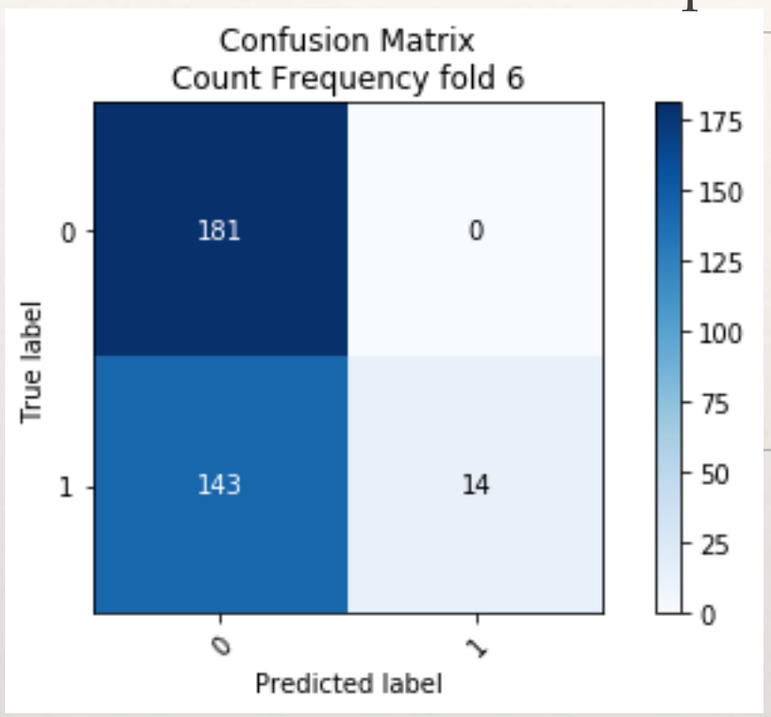


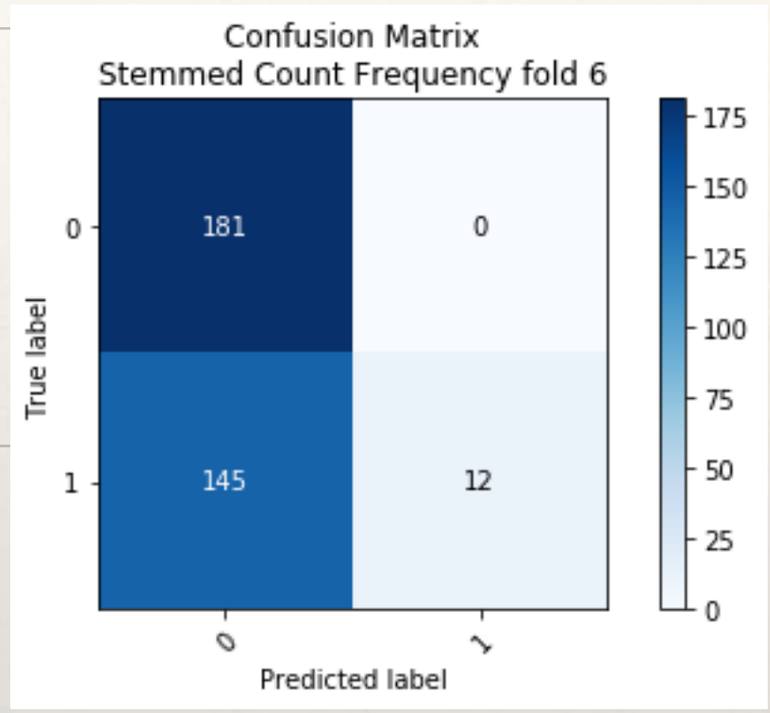


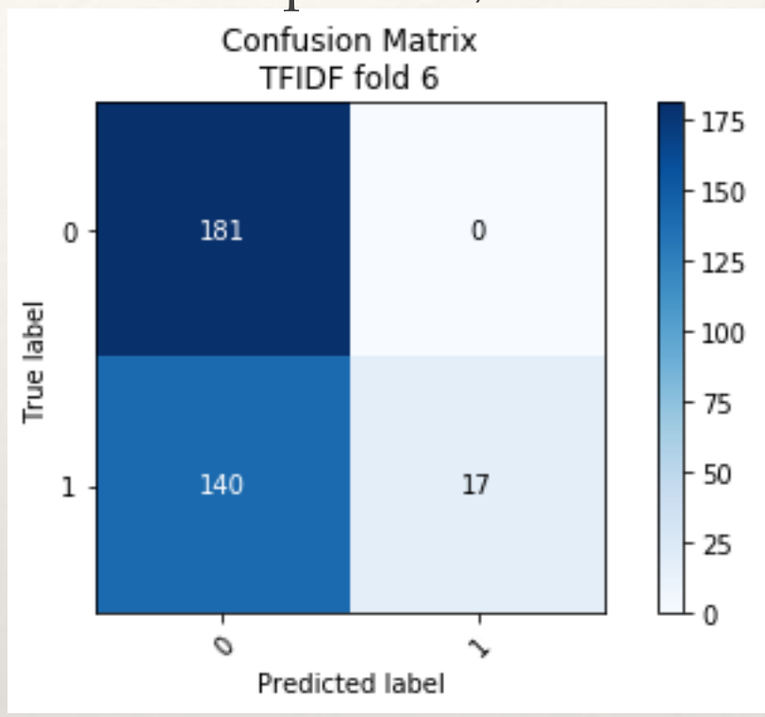


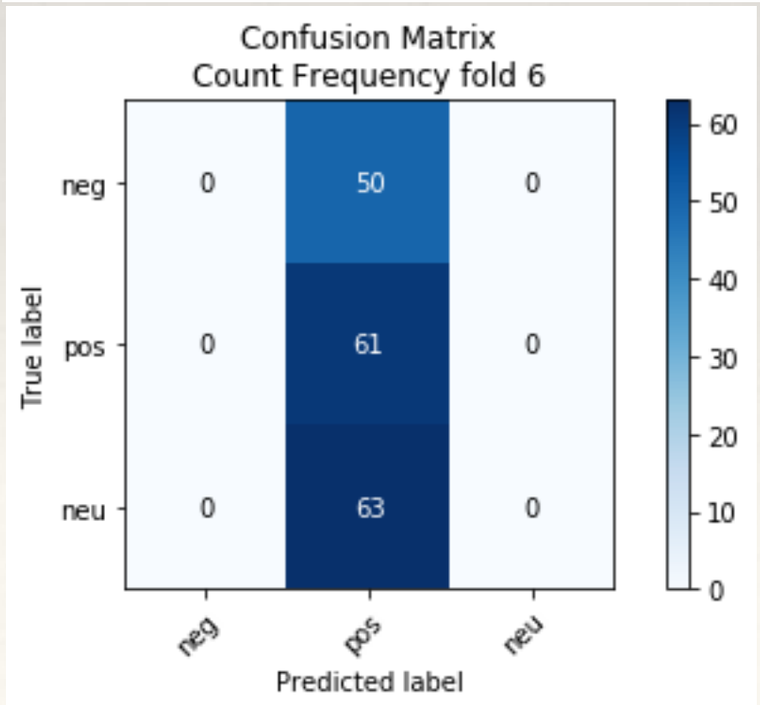
Random Forest

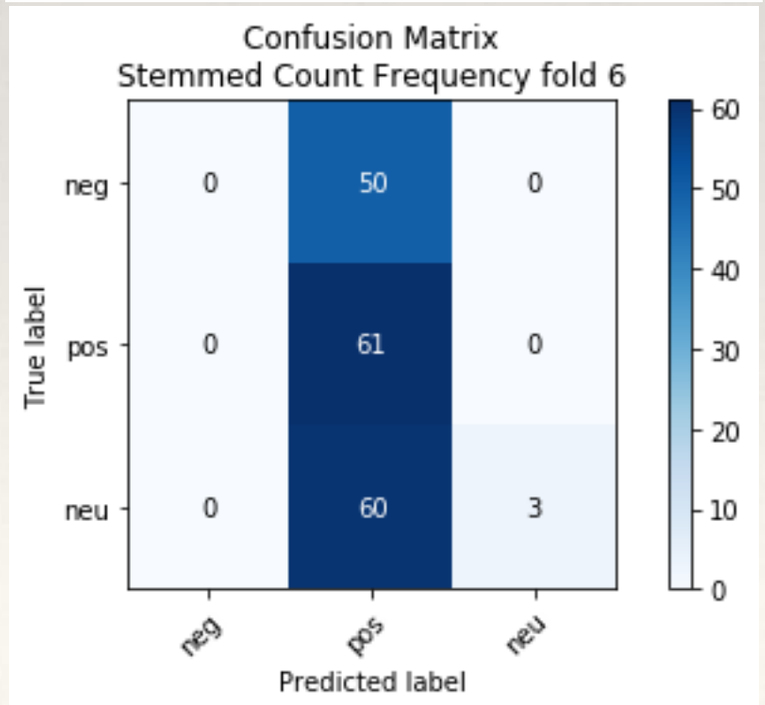
- Real/Fake Prediction Average Accuracy:
  - Count frequency - 70.4%
  - TF-IDF - 70.5%
  - Stemmed count freq. - 67.1%
- Sentiment Prediction Average Accuracy:
  - Count frequency - 42.4%
  - TF-IDF - 42.8%
  - Stemmed count freq. - 42.6%

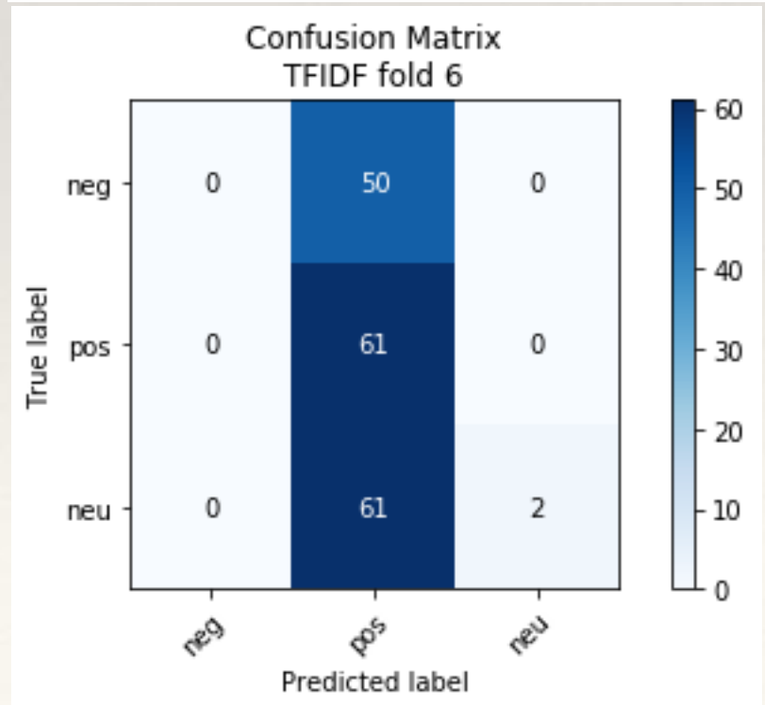






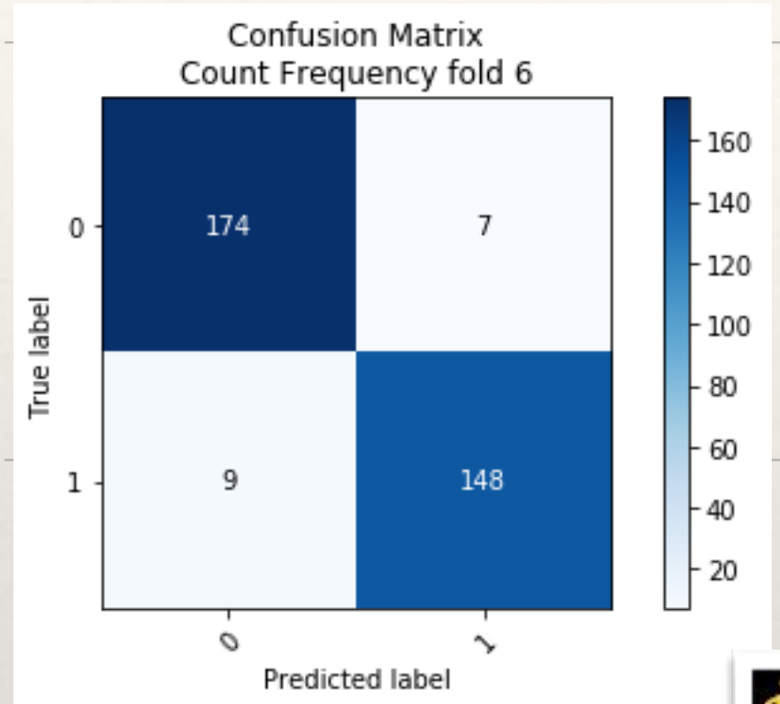


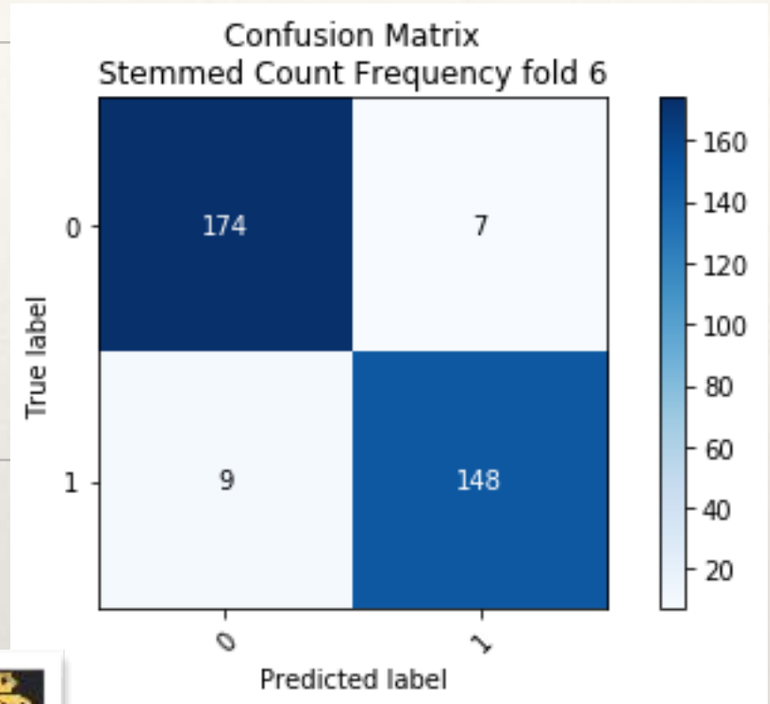


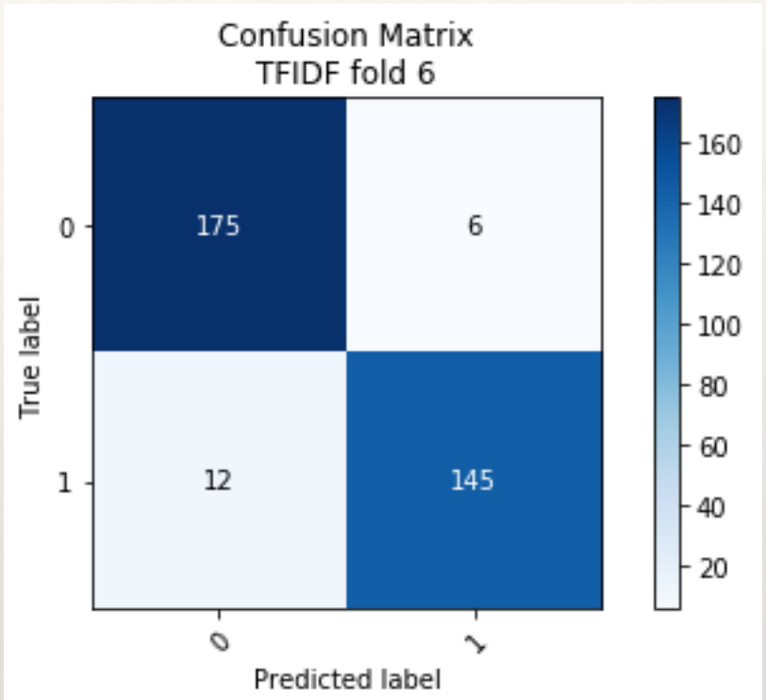


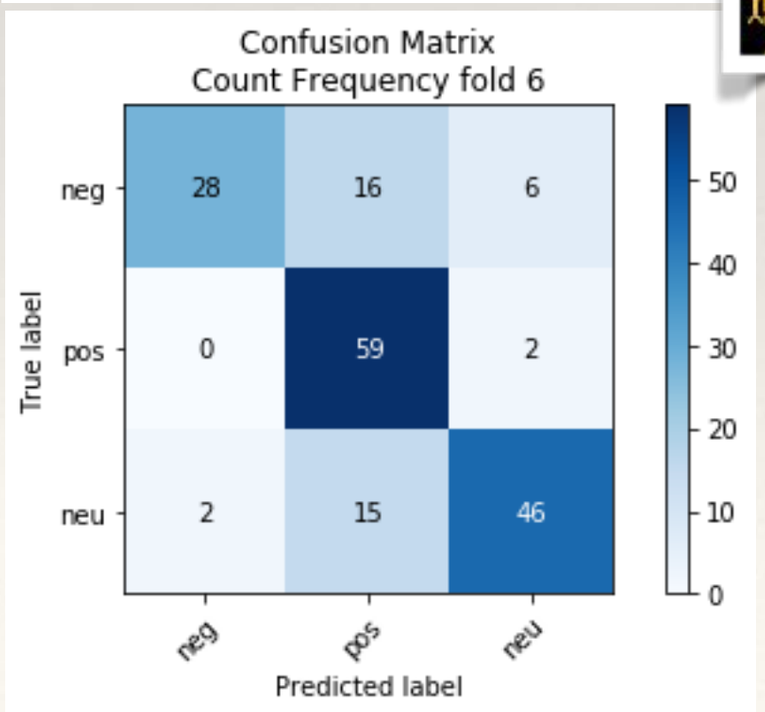
Logistic Regression

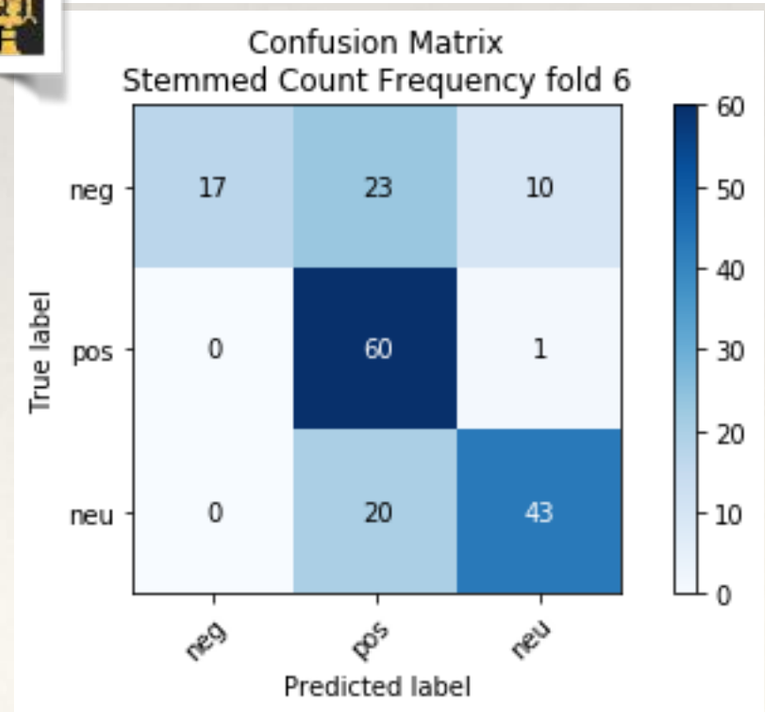
- Real/Fake Prediction Average Accuracy:
  - Count frequency - 96.2%
  - TF-IDF - 95.8%
  - Stemmed count freq. - 96.3%
- Sentiment Prediction Average Accuracy:
  - Count frequency - 76.0%
  - TF-IDF - 78.1%
  - Stemmed count freq. - 70.7%

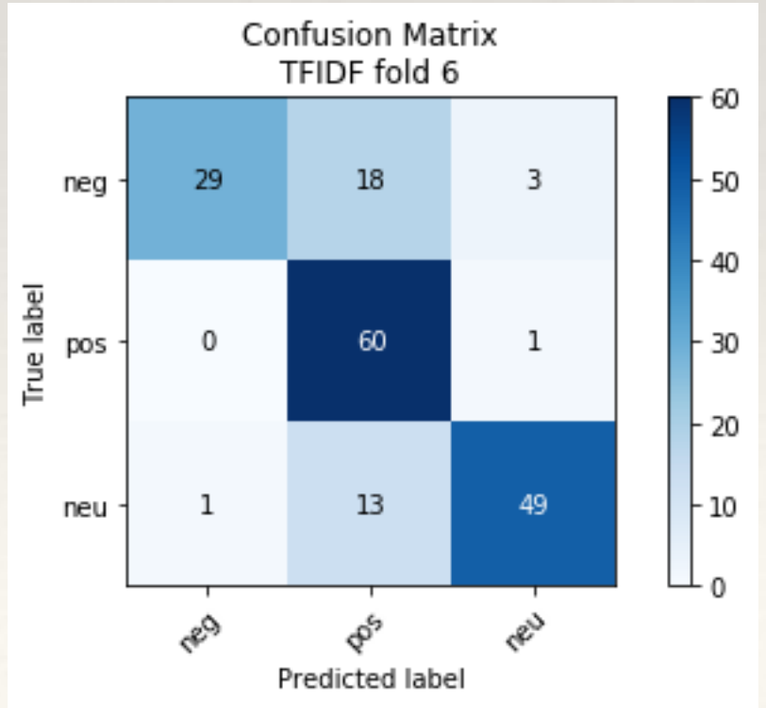












Results

MNB Real	SVM Real	RF Real	Logistic Real	MNB Sent	SVM Sent	RF Sent	Logistic Sent	VecType
0.948817	0.962722	.704142	0.961538	0.73127	0.78421	.424646	0.760026	Count Frequency
0.927219	0.964497	.70503	0.957692	0.738748	0.799744	.428088	0.781324	TFIDF
0.94645	0.962722	.671302	0.962722	0.685822	0.772131	.425796	0.706544	Stemmed Count Frequency

MNB Real Top 5 Feature Importance


```
{'Count Frequency': [{'congress': 5.759759701563032,
   '2016electionin3words': 4.892782663471337,
   'trumpforpresident': 4.8038351774548405,
   'thingspeopleontwitterlike': 4.607461867205235,
   'rt': 4.143370644325254},
  {'congress': 5.765685446818899,
   'trumpforpresident': 5.483495801988313,
   '2016electionin3words': 4.957714087068672,
   'thingspeopleontwitterlike': 4.608836453589971,
   'rt': 4.140772689162062},
  {'congress': 5.780156408593282,
   '2016electionin3words': 4.8593462247299675,
   'thingspeopleontwitterlike': 4.489986121263919,
   'trumpforpresident': 4.3735319894575895,
   'rt': 4.152449488586919},
  {'congress': 5.746033046902089,
   '2016electionin3words': 4.917746183823315,
   'trumpforpresident': 4.750380718183011,
   'thingspeopleontwitterlike': 4.519107040785551,
   'rt': 4.06770483749036},
  {'congress': 5.760270831488239,
   'trumpforpresident': 5.4621285205784735,
   '2016electionin3words': 4.969298241031561,
   'thingspeopleontwitterlike': 4.604400317912397,
   'rt': 4.2325897302644995},
```

Conclusion

- Social media and society in general are negatively impacted by trolls. It's helpful to know who is real and who isn't.
  - The trolls attempted to influence and disrupt a broad range of local and global political debates
    - Most but not all of the content favors the extremist right wing
- Some of our models were surprisingly good at predicting bots. Three models consistently *predicted with mid-90% accuracy* in cross-validation tests *(10 folds),* independent of vectorization methods
- They were also good at predicting sentiment (around 75%, kappa = 0.73).
- Keras model predicted with 80% accuracy . We are still looking at the best optimization for the Keras model

Project III: Social Media and The Music Festivals

IST 736 – Text Mining, Course Instructor: Dr. Ami Gates

Project Partner: Scott Snow, ***Spring 2019***

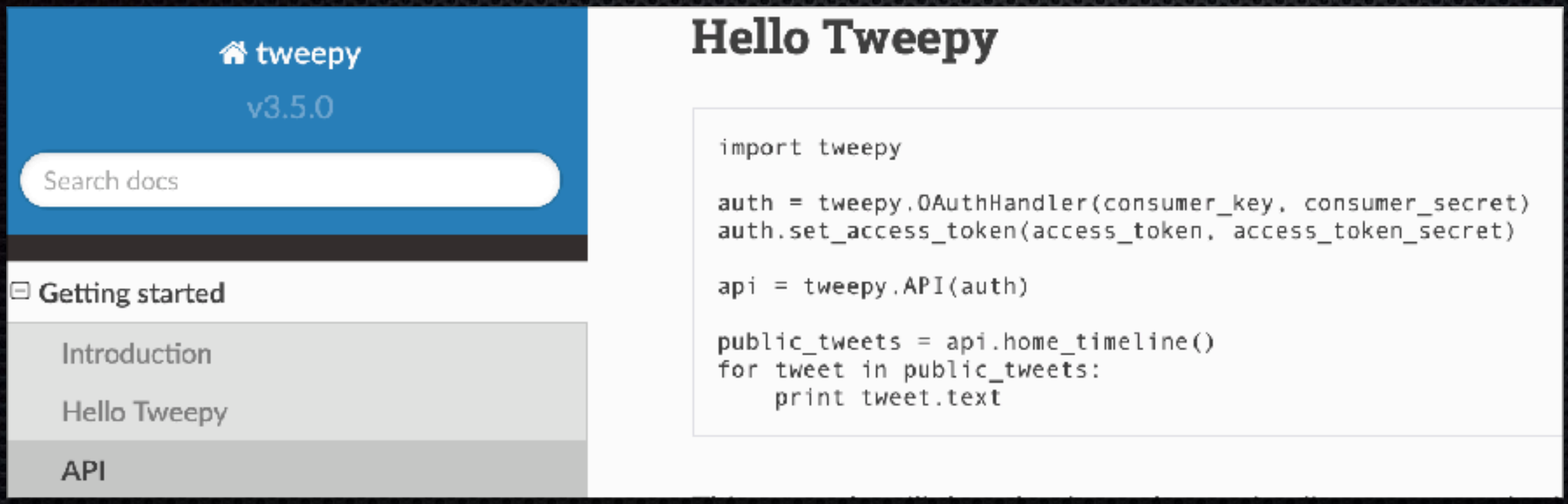
Research Question:

- In what ways can we utilize data-driven text mining algorithms of Twitter data—tweets to gain valuable insights about music festivals?

Research Goals:

- A Twitter text mining project
- We download tweets as data set to investigate the phenomenas of the Music Festivals
- Categories:
  - Sentimental analysis
  - Musical Trends and Genres
  - Specific language structures
    - Festival slang!
  - Sponsors
  - Consumptions
    - Soft/hard liquor
    - Costume/fashion trends
    - Others (could be personal and restrictive!)

Python Library and Data: Tweets



- Need to set up a twitter developer account
- Get an access to use API: application programming interface
- Python libraries
  - nltk, tweeps, textblob
- Hash-tag: ***#musicfestival***
  - n=500
- Cleaning and preparing Data
- Applying Text Mining Applications
  - Finding applicable working model

- Cleaning and prep of the data: tweets ***#musicfestival***

```
RT @cassidyaspinall: Can’t wait to walk into festivals with the music banging, sun shining and my best
girls with me! Summer hurry up alrea…
The Premier's war on music has to stop. We need to protect our arts and culture in NSW, not put more
barriers in th… https://t.co/bENpmZa5pF
RT @ElPasoTXGov: El Paso is not only amongst the safest cities in the nation, we are also thriving. From
new restaurants in downtown, cultu…
Can’t wait to walk into festivals with the music banging, sun shining and my best girls with me! Summer
hurry up already 😭
I think I need to start going to music festivals. People seem to find real love there https://t.co/
01PCAJj2tJ
To help you choose where to get your groove on during your #holiday, we have compiled a list for you of
the Top 5 M… https://t.co/0yP6vbQfnw
RT @UVMLondon: Sun, Sea, Cider, Great Live Music— sounds like heaven - we’re there on 26th May. #festivals
#campingbeciderseaside #devon #t…
```

Unstructured Data—a sample tweet

{"created_at":"Mon Jan 15 01:14:44 +0000 2018","id":952710610915090432,"id_str":"952710610915090432","text":"RT @JamesMelville: D
onald Trump\u2019s State of the Union speech is on January 30th. \n\nHe\u2019s obsessed by TV ratings.\n \nAmerica - don\u2019t w
atch his\u2026","source":"\u003ca href=\"http:\/\/twitter.com\/download\/iphone\" rel=\"nofollow\"\u003eTwitter for iPhone\u003c\
/a\u003e","truncated":false,"in_reply_to_status_id":null,"in_reply_to_status_id_str":null,"in_reply_to_user_id":null,"in_reply_to
_user_id_str":null,"in_reply_to_screen_name":null,"user":{"id":743212353546817540,"id_str":"743212353546817540","name":"Mariah","
screen_name":"mariahrash003","location":null,"url":null,"description":null,"translator_type":"none","protected":false,"verified":
false,"followers_count":75,"friends_count":81,"listed_count":0,"favourites_count":4195,"statuses_count":1206,"created_at":"Wed Ju
n 15 22:43:28 +0000 2016","utc_offset":null,"time_zone":null,"geo_enabled":false,"lang":"en","contributors_enabled":false,"is_tra
nslator":false,"profile_background_color":"F5F8FA","profile_background_image_url":"","profile_background_image_url_https":"","pro
file_background_tile":false,"profile_link_color":"1DA1F2","profile_sidebar_border_color":"C0DEED","profile_sidebar_fill_color":"D
DEEF6","profile_text_color":"333333","profile_use_background_image":true,"profile_image_url":"http:\/\/pbs.twimg.com\/profile_ima
ges\/943920250780901377\/e8_hLGL1_normal.jpg","profile_image_url_https":"https:\/\/pbs.twimg.com\/profile_images\/943920250780901

- **Tweet Data Dictionary**
  - **created_at**
  - **id**
  - id_str
  - **text**
  - source
  - truncated
  - in_reply_to_status_id
  - in_reply_to_status_id_str
  - in_reply_to_user_id
  - in_reply_to_user_id_str
  - in_reply_to_screen_name
  - **user**
  - **coordinates**
  - **place**
  - quoted_status_id
  - quoted_status_id_str
  - **geo**

- Bag of Words
- ***#musicfestivals***

```
In [14]: len(BagOfWords)
Out[14]: 4149
```

```
'June', 'magia', 'existe', 'Solo', 'tienes', 'que', 'abrir', 'los', 'ojos',
'annual', 'Calle', 'Ocho', 'event', 'Miami', 'Blu', 'New', 'Single', 'CAR',
'AWFUL', 'shredded', 'Fringe', 'Vest', 'Festival', 'vest', 'gypsy', 'boho',
'Tassel', 'vest', 'WHITE', 'Earthy', 'retro', 'inspired', 'vest', 'made',
'elegant', 'fearsome', 'from', 'grand', 'Just', 'click', 'here', 'Show',
'organizer', 'and', 'post', 'your', 'show', 'with', 'all', 'the', 'details',
'and', 'images', 'tienes', 'que', 'disfrutar', 'que', 'musica', 'guie', 'sure
'and', 'get', 'your', 'tickets', 'for', 'They', 'are', 'running', 'out', 'fas
'Amo', 'Amas', 'June', "we'll", 'celebrate', 'years', 'family', 'pride',
'tradition', 'and', 'puro', 'mariachi', 'Will', 'you', 'there', 'Best', 'seat
'Colorful', 'Austin', 'THE', 'FOO', 'FIGHTERS', 'Contiki', 'Sounds', 'returns
'with', 'trip', 'the', "UK's", 'Reading', 'Festival', 'SCARF', 'beach', 'cove
'gypsy', 'sarong', 'scarf', 'festival', 'clothing', 'NAVY', 'blue', 'boho',
'chic', 'hippie', 'beach', 'skirt', 'the', 'building', 'performing', 'live',
'Official', 'Hats', 'MOHAWK', 'BENDS', 'killed', 'travels', 'April',
'documentary', 'about', 'toda', 'energía', 'música', 'year', 'benefiting',
'support', 'vital', 'mental', 'health', 'services', 'for', 'youth', 'and',
'families', 'Robot', 'work', 'progress', 'Wave', "I'm", 'Surfing', 'Right',
'Now', 'Dude', 'TUNE']
```

- Bag of Hash-Tags
  - ***#musicfestivals***

```
In [16]: len(BagOfHashes)
Out[16]: 7950
```

```
'musicians', 'lovemusic', 'NewMusic', 'Ultra2019', 'album', 'newsong', 'live',
'np', 'music', 'productor', 'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic',
'NewMusic', 'Ultra2019', 'album', 'newsong', 'live', 'np', 'music', 'productor',
'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic', 'NewMusic', 'Ultra2019',
'album', 'newsong', 'live', 'np', 'music', 'productor', 'TOPTUNE',
'musicfestival', 'musicians', 'lovemusic', 'NewMusic', 'Ultra2019', 'album',
'newsong', 'live', 'np', 'music', 'productor', 'TOPTUNE', 'musicfestival',
'musicians', 'lovemusic', 'NewMusic', 'Ultra2019', 'album', 'newsong', 'live',
'np', 'music', 'productor', 'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic',
'NewMusic', 'Ultra2019', 'album', 'newsong', 'live', 'np', 'music', 'productor',
'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic', 'NewMusic', 'Ultra2019',
'album', 'newsong', 'live', 'np', 'music', 'productor', 'TOPTUNE', 'musicfestival',
'musicians', 'lovemusic', 'NewMusic', 'Ultra2019', 'album', 'newsong', 'live',
'np', 'music', 'productor', 'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic',
'NewMusic', 'Ultra2019', 'album', 'newsong', 'live', 'np', 'music', 'productor',
'TOPTUNE', 'musicfestival', 'musicians', 'lovemusic', 'NewMusic', 'Ultra2019',
'album', 'newsong', 'live', 'np', 'music', 'show', 'edclasvegas', 'lasvegas',
'electro', 'musicismylife', 'rappers', 'goodmusic', 'pop', 'studio', 'livemusic',
```

- Bag of Links
- ***#musicfestivals***

```
In [18]: len(BagOfLinks)
Out[18]: 823
```

```
9G2vhHH#DJ', 'https://t.co/ohF108A34r#DJ', 'https://t.co/QgWLiRGajv#DJ',
tps://t.co/WDf6fEM7lb#DJ', 'https://t.co/Q8lrUnuwN6#DJ', 'https://t.co/
ERoq5sC#DJ', 'https://t.co/QPZM1h8AHZ#DJ', 'https://t.co/sMCY27n0fX#DJ',
tps://t.co/xJs8QJtas7#DJ', 'https://t.co/RyelbKqFQt#DJ', 'https://t.co/
8j4yax0#DJ', 'https://t.co/qDAvEMCI7oRegister', 'https://t.co/9sGRPa1ttL',
tps://t.co/nGgSo6zBRC😍⚡🎇solo', 'https://t.co/jTAfiGi756🔥🔥📣#instagood'
tps://t.co/Uj0sDwccE1Make', 'https://t.co/bGZZX591H3💖💏🎡Yo', 'https://t.
GMrbDoXOn', 'https://t.co/Q6A7otvmzehttps://t.co/grgFkr3hhW', 'https://t.co
AD3H60ART', 'https://t.co/JUOFWjXswkSEE', 'https://t.co/SGPusa26bP…',
tps://t.co/wLKPTd6DrFCROCHET', 'https://t.co/TY9JABL4dNRT', 'https://t.co/
z27UJER.#AfroRootsFest', 'https://t.co/prTeA6h7VP😍liberar', 'https://t.co/
Hok1nM3#bohochic', 'https://t.co/5ouy9TNCg0This', 'https://t.co/
1IfmG5NEnvious', 'https://t.co/T4bbFTgx2X[redacted]Catch', 'https://t.co/37bMEtqPx
```

- BigBag = BagOfWords + BagOfHashes
  - ***#musicfestivals***

```
In [27]: len(BigBag)
Out[27]: 12099
```

```
'Ultra2019', 'album', 'newsong', 'live', 'np', 'music', 'show',
'edclasvegas', 'lasvegas', 'electro', 'musicismylife', 'rappers',
'goodmusic', 'pop', 'studio', 'livemusic', 'band', 'like', 'rock',
'musicfestival', 'coachella2019', 'coachella', 'coachellastyle',
'ravecouples', 'electro', 'DJ', 'productor', 'musicfestival',
'musicians', 'album', 'newsong', 'music', 'EDM', 'production',
'SXSW2019', 'Texas', 'austintexas', 'city', 'SouthbySouthWest',
'musicfestival', 'filmfestival', 'musicfestival', 'SXSW2019',
'blaqpyrates', 'Aust', 'artist', 'musician', 'artistandrepertoire',
'magmileshows', '6thstreetaustin', 'independent', 'KeyWest',
'burningman', 'blackrockcity', 'ravebooty', 'electro', 'bohemian',
'festival', 'festivalfashion', 'music', 'musicfestival', 'sunday',
'sundayfunday', 'bohostyle', 'craft', 'music', 'instafun',
'hiphopblog', 'hiphopnation', 'hiphoplife', 'bhfyp', 'hiphophead',
'instagood', 'musicismylife', 'rappers', 'goodmusic', 'pop']
```

- Ignored Words
- ***#musicfestivals***

```
In [28]: len(IgnoreThese)
Out[28]: 264
```

```
ティバル","ラジオ","フェスティバル","ラジオ","フェスティバル","ラジオ","フェ
","フェスティバル","ラジオ","صخب_غربة","ملبورن","自動運転中","صخب_غربة","おい
","제1회빛이나예술제","제1회빛이나예술제","لحظات","___","حفل","ختام","ختام","ائي
ティバル","ラジオ","フェスティバル","بإمكانكم","قضاء","عيشها","Muslim","Holi(
ramsNagpur","Haldirams","제1회빛이나예술제","जगह","حفل","ختام","السينمائي","ي
ゃんと間違われる","川崎東田商店街その","زال","بإمكانكم","قضاء","لحظات",
田商店街その","ممتعة","آخر","يوم","لمهرجان","الدولي","الثالث","للطائرات","الورقية",
みでお祭り","リカちゃんと間違われる","川崎東田商店街その","제1회빛이나예술제","着
ゃんと間違われる","川崎東田商店街その","の開催が決定","クラブミュージック","現
つりをお楽しみください","گردشگری","فصل","زمستان","سود","بیشتری","برای","دولت","
chanchan","ようこそ女子大エリア","一年一度的節日","今天剛好早下班","就難得出門
霏打給我","跟她一起看煙火","覺得很開心滿足","偕楽園","제1회빛이나예술제","偕楽園
ram","Nagpur","BOTH","menunjukkan","HOLIKA","BLACK","SLAB","PDF","U
```

- Words Dictionary
- ***#musicfestivals***

```
In [43]: len(WordDict)
Out[43]: 1871
```

```
stival2019': 6, 'WeLiveTheBeachLi': 3, 'BeachLifeFestival2': 3, 'Bea': 3,
: 3, 'JimLindberg': 3, 'BRATSONTHEBEAT': 3, 'RAMONESFORKIDS': 3, 'stall': 3
y': 3, 'oughtringtonprimaryschool': 3, 'lymm': 3, 'helpneeded': 3, 'oughtyf
': 3, 'RootsPicnic': 6, 'SingerSongwriter': 6, 'JUNO': 6, 'CFMA': 6, 'BC':
 3, 'BeachLifeFestLA': 3, 'LeylandFe': 6, 'MarlayPark': 6, 'psymusic': 6,
': 6, 'musi': 9, 'breakoutwest': 3, 'canadianmusic': 3, 'coachellaoutfit':
shion': 3, 'Franklin': 3, 'talent': 6, 'feisliverpool': 9, 'Etsy': 3, 'etsy
: 6, 'festivalstyle': 3, 'model': 3, 'studio': 6, 'light': 3, 'bed': 3, 'wa
musicfestivalRT': 6, 'fdnwhour': 6, 'manchestermusicscene': 3, 'musicphotog
 3, 'girlzrawk': 3, 'girlzrawkmanchester': 3, 'PushitChallenge': 3, 'Beatbl
Tone': 9, '80s': 9, 'MUSICFestival': 9, 'musicman': 3, 'musicianlife': 3,
': 3, 'HereWeGoLo': 3, 'cuzeventsrock': 3, 'rootspicnic': 3, 'philly': 3, '
sBTQ': 3, 'qotd': 3, 'quote': 3, 'wild': 3, 'flowerchild': 3, 'LeylandFesti
ure': 3, 'dayfornight': 3, 'shades': 3, 'eprom': 3, 'alixperez': 3, 'PalNor
: 6, 'FreddieMercury': 3, 'goodmusic': 9, 'familyfriendly': 3, 'pr': 9, 'mu
list': 6, 'spotify': 6, 'Spotifyplays': 6, 'SpotifyMusicPromotion': 6,
opid': 3, 'reggae': 3, 'iration': 3, 'ska': 3, 'rebelution': 3, 'pepper': 3
```

- Word Cloud - top 50 words
  - ***#musicfestivals***

Results: Sentiment Analysis

- Python textblob library

```
Positive tweets percentage: 41.935483870967744 %
Negative tweets percentage: 3.2258064516129032 %
```

Frequency Distribution

- Used raw tweets.
- tokenized and lowered the tweeds
- **34886 words**
- Created Frequency distribution

```
('musicfestival', 1965)
('music', 1848)
('musicians', 1599)
('album', 1545)
('newsong', 1545)
('productor', 1536)
('live', 1464)
('edm', 1449)
('toptune', 1413)
('np', 1413)
('ff', 1095)
('deephouse', 687)
('onairnow', 654)
('lovemusic', 648)
('newmusic', 615)
('dj', 522)
('festival', 339)
('top', 219)
('production', 153)
('tickets', 120)
('crochet', 120)
('djxexer', 96)
('newmusicalert', 93)
('electronicmusic', 90)
```

```
'summer', 75)
'day', 69)
'show', 60)
'get', 60)
'april', 60)
'clothing', 57)
'halter', 57)
'check', 57)
'stage', 57)
'airplay', 57)
'see', 54)
'art', 54)
'electro', 51)
'ultra2019', 51)
'june', 48)
'wave', 45)
'boho', 45)
'love', 45)
'days', 45)
'one', 45)
'musician', 45)
'party', 42)
'event', 42)
'fringe', 42)
'march', 39)
'may', 39)
'weekend', 39)
```

Syracuse University

Bigram

- Used raw tweets.
- tokenized and lowered the tweeds
- **3329 bigrams**
- List of top 50 scored

```
(('album', 'newsong'), 0.044287106575703 72)
(('musicfestival', 'musicians'), 0.044287106575703 72)
(('live', 'np'), 0.04050335378088631)
(('newsong', 'live'), 0.04050335378088631)
(('productor', 'toptune'), 0.04050335378088631)
(('np', 'music'), 0.040417359399185 92)
(('music', 'edm'), 0.039901393108983545)
(('toptune', 'musicfestival'), 0.036203634695866536)
(('edm', 'ff'), 0.031387949320644384)
(('musicians', 'album'), 0.025712320128418276)
(('deephouse', 'productor'), 0.019606719027690192)
(('musicians', 'lovemusic'), 0.018574786447285444)
(('ff', 'onairnow'), 0.018488792065585048)
(('lovemusic', 'newmusic'), 0.017628848248581092)
(('newmusic', 'album'), 0.015221005560970016)
(('dj', 'deephouse'), 0.00894341569684114)
(('ff', 'productor'), 0.008599438170039558)
(('onairnow', 'deephouse'), 0.006965544917732041)
(('dj', 'productor'), 0.005417646047124921)
(('edm', 'production'), 0.003783752794817405)
(('newsong', 'music'), 0.003783752794817405)
(('productor', 'musicfestival'), 0.0035257696497162185)
(('onairnow', 'dj'), 0.0029238089778134495)
(('production', 'djxexer'), 0.0027518202144126585)
(('music', 'festival'), 0.002665825832712263)
(('newmusicalert', 'musicfestival'), 0.002665825832712263)
(('toptune', 'newmusicalert'), 0.002665825832712263)
(('djxexer', 'electronicmusic'), 0.0024938370693114715)
(('edm', 'productor'), 0.0024938370693114715)
(('onairnow', 'productor'), 0.002235853924210285)
(('music', 'productor'), 0.0017198876340079115)
(('ff', 'dj'), 0.0016338932523075 16)
(('toptune', 'airplay'), 0.0016338932523075 16)
(('halter', 'top'), 0.0015478988706071203)
(('top', 'crochet'), 0.0014619044889067248)
(('music', 'deephouse'), 0.0013759101072063293)
(('newmusic', 'ultra2019'), 0.0013759101072063293)
(('ultra2019', 'album'), 0.0013759101072063293)
(('airplay', 'musicfestival'), 0.0011179269621051425)
(('coub', 'album'), 0.0010319325804047 47)
(('festival', 'top'), 0.0010319325804047 47)
(('newmusic', 'coub'), 0.0010319325804047 47)
(('top', 'festival'), 0.0010319325804047 47)
(('check', 'out'), 0.0009459381987043513)
(('festival', 'clothing'), 0.0009459381987043513)
(('lovemusic', 'album'), 0.0009459381987043513)
(('music', 'wave'), 0.0009459381987043513)
(('musicfestival', 'art'), 0.0009459381987043513)
(('onairnow', 'prod'), 0.0009459381987043513)
(('wave', 'musicfestival'), 0.0009459381987043513)
```

Bigram - PMI Pointwise Mutual Information

- tf-idf vectorizer
- A measure of how often two events x and y occur, compared with what we would expect if they were independent
- Frequency filter = 3
- **3329 bigrams**
- List of top 50 scored

```
(('able', 'look'), 13.505398067619186)
(('ableton', 'logicpro'), 13.505398067619186)
(('abril', 'formali'), 13.505398067619186)
(('abrir', 'los'), 13.505398067619186)
(('acid', 'psytribe'), 13.505398067619186)
(('addition', 'comes'), 13.505398067619186)
(('adrian', 'kosky'), 13.505398067619186)
(('alto', 'mare'), 13.505398067619186)
(('american', 'icon'), 13.505398067619186)
(('amo', 'amas'), 13.505398067619186)
(('amped', 'spend'), 13.505398067619186)
(('américacantat9', 'panamá'), 13.505398067619186)
(('andorra', 'five'), 13.505398067619186)
(('andy', 'track-a-day'), 13.505398067619186)
(('anniversary', 'design'), 13.505398067619186)
(('arizona', 'final'), 13.505398067619186)
(('arrive', 'trust'), 13.505398067619186)
(('avoid', 'disappointment'), 13.505398067619186)
(('awesome', 'session'), 13.505398067619186)
(('awful', 'shredded'), 13.505398067619186)
(('bald', 'eagle'), 13.505398067619186)
(('bamboo', 'head'), 13.505398067619186)
(('bastard', 'drops'), 13.505398067619186)
(('bathing', 'swim'), 13.505398067619186)
(('bea', 'nowplaying'), 13.505398067619186)
(('beard', 'entered'), 13.505398067619186)
(('bed', 'wall'), 13.505398067619186)
(('bends', 'killed'), 13.505398067619186)
(('better', 'heading'), 13.505398067619186)
(('beyondwonderland', 'insomniac'), 13.505398067619186)
(('bhfyp', 'hiphophead'), 13.505398067619186)
(('bigtakeoverband', 'thebigtakeover'), 13.505398067619186)
(('blackrockcity', 'ravebooty'), 13.505398067619186)
(('blaqpyrates', 'aust'), 13.505398067619186)
(('blossom', 'jamboree'), 13.505398067619186)
(('booking', 'fort'), 13.505398067619186)
(('bourbonandbeyond', 'bourbon'), 13.505398067619186)
(('brampton', 'mississauga'), 13.505398067619186)
(('bratsonthebeat', 'ramonesforkids'), 13.505398067619186)
(('breakout', 'west'), 13.505398067619186)
(('breakoutwest', 'canadianmusic'), 13.505398067619186)
(('brimmed', 'beanie'), 13.505398067619186)
(('build-your-own', 'subscriptions'), 13.505398067619186)
(('burning', 'man'), 13.505398067619186)
(('busy', 'tones'), 13.505398067619186)
(('caledon', 'orangeville'), 13.505398067619186)
(('calle', 'ocho'), 13.505398067619186)
(('cap', 'brimmed'), 13.505398067619186)
(('cardrossestate', 'dtrhfest'), 13.505398067619186)
(('carla', 'maxwell'), 13.505398067619186)
```

Bigram - PMI Pointwise Mutual Information - II

- Frequency filter = 50
  - Significantly different than freq_filter = 3
- **3329 bigrams**
- List of top 50 scored

```
(('djxexer', 'electronicmusic'), 8.456488467138236)
(('production', 'djxexer'), 7.83297272564768 75)
(('halter', 'top'), 7.23757099673789)
(('top', 'crochet'), 6.0811082551021425)
(('lovemusic', 'newmusic'), 5.750510565455716)
(('ff', 'onairnow'), 4.9736539386643415)
(('dj', 'deephouse'), 4.923690501814605)
(('toptune', 'airplay'), 4.625814818006404)
(('toptune', 'newmusicalert'), 4.625814818006 4005)
(('edm', 'ff'), 4.589518688783411)
(('live', 'np'), 4.574660730056298)
(('productor', 'toptune'), 4.505398067619183)
(('deephouse', 'productor'), 4.4990842936869 77)
(('album', 'newsong'), 4.496969445548601)
(('musicians', 'lovemusic'), 4.447406344860006)
(('newsong', 'live'), 4.4458153575985015)
(('edm', 'production'), 4.37652496544921)
(('newmusic', 'album'), 4.28509489612615)
(('onairnow', 'deephouse'), 4.23785995762994)
(('np', 'music'), 4.2355452238765)
(('music', 'edm'), 4.180713143216078)
(('newmusicalert', 'musicfestival'), 4.150046971194371)
(('musicfestival', 'musicians'), 4.100483870505773)
(('toptune', 'musicfestival'), 3.988140144647982)
(('musicians', 'album'), 3.66297939698752 73)
(('onairnow', 'dj'), 3.381733088243866)
(('dj', 'productor'), 3.0397344952703733)
(('ff', 'productor'), 2.6375016036265286)
(('music', 'festival'), 2.3726288748959696)
(('ff', 'dj'), 1.7986294314466633)
(('onairnow', 'productor'), 1.437653460983352)
(('newsong', 'music'), 0.6896145234909952)
(('productor', 'musicfestival'), 0.5075989758124528)
(('edm', 'productor'), 0.44749968391098 35)
(('music', 'productor'), -0.439460378188353 93)
```

Bigram - PMI
Pointwise Mutual Information - III

- Frequency filter = 100
  - Significantly different than freq_filter = 3
- **3329 bigrams**
- List of top 50 scored

```
('lovemusic', 'newmusic'), 5.750510565455716)
('ff', 'onairnow'), 4.9736539386643415)
('dj', 'deephouse'), 4.923690501814605)
('edm', 'ff'), 4.5895186887834 11)
('live', 'np'), 4.574660730056298)
('productor', 'toptune'), 4.505398067619183)
('deephouse', 'productor'), 4.4990842936869 77)
('album', 'newsong'), 4.496969445548601)
('musicians', 'lovemusic'), 4.447406344860006)
('newsong', 'live'), 4.4458153575985015)
('edm', 'production'), 4.376524965449 21)
('newmusic', 'album'), 4.28509489612615)
('onairnow', 'deephouse'), 4.23785995762994)
('np', 'music'), 4.2355452238765)
('music', 'edm'), 4.180713143216078)
('musicfestival', 'musicians'), 4.100483870505773)
('toptune', 'musicfestival'), 3.988140144647982)
('musicians', 'album'), 3.6629793969875273)
('onairnow', 'dj'), 3.381733088243866)
('dj', 'productor'), 3.0397344952703733)
('ff', 'productor'), 2.6375016036265286)
('newsong', 'music'), 0.6896145234909952)
('productor', 'musicfestival'), 0.507598975 8124528)
```

djxexer', 'electronicmusic'), 8.456488467138236)
production', 'djxexer'), 7.83297272256476875)
halter', 'top'), 7.23757099673789)
top', 'crochet'), 6.0811082551021425)
lovemusic', 'newmusic'), 5.750510565455716)
ff', 'onairnow'), 4.97365393866643415)
dj', 'deephouse'), 4.923690501814605)
toptune', 'airplay'), 4.625814818006404)
toptune', 'newmusicalert'), 4.6258148180064005)
edm', 'ff'), 4.589518688783411)
live', 'np'), 4.574660730056298)
productor', 'toptune'), 4.505398067619183)
deephouse', 'productor'), 4.4990842936869 77)
album', 'newsong'), 4.496969445548601)
musicians', 'lovemusic'), 4.447406344860006)
newsong', 'live'), 4.4458153575985015)
edm', 'production'), 4.37652496544921)
newmusic', 'album'), 4.28509489612615)
onairnow', 'deephouse'), 4.23785995762994)
np', 'music'), 4.2355452238765)
music', 'edm'), 4.180713143216078)
newmusicalert', 'musicfestival'), 4.15004697119437
musicfestival', 'musicians'), 4.100483870505773)
toptune', 'musicfestival'), 3.9881401446479 82)
musicians', 'album'), 3.6629793969875273)
onairnow', 'dj'), 3.3817330882438 66)
dj', 'productor'), 3.0397344952703733)
ff', 'productor'), 2.6375016036265286)
music', 'festival'), 2.3726288748959696)
ff', 'dj'), 1.7986294314466633)
onairnow', 'productor'), 1.4376534609833 52)
newsong', 'music'), 0.6896145234909952)
productor', 'musicfestival'), 0.5075989758124528)
edm', 'productor'), 0.4474996839109835)
music', 'productor'), -0.4394603781883 5393)

Results:

- Musical Trends and Genres
- EDM
  - (('music', 'edm'), 0.039901393108983545)
- DeepHouse

  - (('music', 'deephouse'), 0.0013759101072063293)

djxexer', 'electronicmusic'), 8.456488467138236)
production', 'djxexer'), 7.83297272256476875)
halter', 'top'), 7.237570996737789)
top', 'crochet'), 6.0811082551021425)
lovemusic', 'newmusic'), 5.750510565455716)
ff', 'onairnow'), 4.9736539386643415)
dj', 'deephouse'), 4.923690501814605)
toptune', 'airplay'), 4.625814818006404)
toptune', 'newmusicalert'), 4.6258148180064005)
edm', 'ff'), 4.589518688783411)
live', 'np'), 4.574660730056298)
productor', 'toptune'), 4.505398067619183)
deephouse', 'productor'), 4.4990842936869 77)
album', 'newsong'), 4.496969445548601)
musicians', 'lovemusic'), 4.447406344860006)
newsong', 'live'), 4.4458153575985015)
edm', 'production'), 4.37652496544921)
newmusic', 'album'), 4.28509489612615)
onairnow', 'deephouse'), 4.23785995762994)
np', 'music'), 4.2355452238765)
music', 'edm'), 4.1807131432160 78)
newmusicalert', 'musicfestival'), 4.15004697119437)
musicfestival', 'musicians'), 4.1004838705057 73)
toptune', 'musicfestival'), 3.988140144647982)
musicians', 'album'), 3.6629793969875273)
onairnow', 'dj'), 3.3817330882438 66)
dj', 'productor'), 3.0397344952703733)
ff', 'productor'), 2.6375016036265286)
music', 'festival'), 2.3726288748959696)
ff', 'dj'), 1.7986294314466633)
onairnow', 'productor'), 1.4376534609833 52)
newsong', 'music'), 0.6896145234909952)
productor', 'musicfestival'), 0.5075989758124528)
edm', 'productor'), 0.4474996839109835)
music', 'productor'), -0.4394603781883 5393)

Results:

- Specific language structures, Slang
  - np: no problem, now playing
  - ff: fully functional
  - crochet: body piercing

Syracuse University

djxexer', 'electronicmusic'), 8.456488467138236)
production', 'djxexer'), 7.83297272256476875)
halter', 'top'), 7.23757099673789)
top', 'crochet'), 6.0811082551021425)
lovemusic', 'newmusic'), 5.750510565455716)
ff', 'onairnow'), 4.97365393866643415)
dj', 'deephouse'), 4.923690501814605)
toptune', 'airplay'), 4.625814818006404)
toptune', 'newmusicalert'), 4.6258148180064005)
edm', 'ff'), 4.589518688783411)
live', 'np'), 4.574660730056298)
productor', 'toptune'), 4.505398067619183)
deephouse', 'productor'), 4.4990842936869 77)
album', 'newsong'), 4.496969445548601)
musicians', 'lovemusic'), 4.447406344860006)
newsong', 'live'), 4.44581535759850 15)
edm', 'production'), 4.37652496544921)
newmusic', 'album'), 4.28509489612615)
onairnow', 'deephouse'), 4.23785995762994)
np', 'music'), 4.2355452238765)
music', 'edm'), 4.180713143216078)
newmusicalert', 'musicfestival'), 4.15004697119437
musicfestival', 'musicians'), 4.100483870505773)
toptune', 'musicfestival'), 3.9881401446479 82)
musicians', 'album'), 3.66297939698752 73)
onairnow', 'dj'), 3.3817330882438 66)
dj', 'productor'), 3.0397344952703733)
ff', 'productor'), 2.6375016036265286)
music', 'festival'), 2.3726288748959696)
ff', 'dj'), 1.7986294314466633)
onairnow', 'productor'), 1.4376534609833 52)
newsong', 'music'), 0.6896145234909952)
productor', 'musicfestival'), 0.5075989758124528)
edm', 'productor'), 0.4474996839109835)
music', 'productor'), -0.439460378188353 93)

Results:

- Festivals
- burning man
- Austin Music
- ultra2019
- coachella
- halsburywest
- microfest2018

Syracuse University

djxexer', 'electronicmusic'), 8.456488467138236)
production', 'djxexer'), 7.83297272256476875)
halter', 'top'), 7.237570996737789)
top', 'crochet'), 6.0811082551021425)
lovemusic', 'newmusic'), 5.750510565455716)
ff', 'onairnow'), 4.9736539386643415)
dj', 'deephouse'), 4.923690501814605)
toptune', 'airplay'), 4.625814818006404)
toptune', 'newmusicalert'), 4.6258148180064005)
edm', 'ff'), 4.589518688783411)
live', 'np'), 4.574660730056298)
productor', 'toptune'), 4.505398067619183)
deephouse', 'productor'), 4.4990842936869 77)
album', 'newsong'), 4.496969445548601)
musicians', 'lovemusic'), 4.447406344860006)
newsong', 'live'), 4.44581535759850 15)
edm', 'production'), 4.37652496544921)
newmusic', 'album'), 4.28509489612615)
onairnow', 'deephouse'), 4.23785995762994)
np', 'music'), 4.2355452238765)
music', 'edm'), 4.1807131432160 78)
newmusicalert', 'musicfestival'), 4.15004697119437
musicfestival', 'musicians'), 4.1004838705057 73)
toptune', 'musicfestival'), 3.988140144647982)
musicians', 'album'), 3.6629793969875273)
onairnow', 'dj'), 3.3817330882438 66)
dj', 'productor'), 3.0397344952703733)
ff', 'productor'), 2.6375016036265286)
music', 'festival'), 2.3726288748959696)
ff', 'dj'), 1.7986294314466633)
onairnow', 'productor'), 1.437653460983352)
newsong', 'music'), 0.6896145234909952)
productor', 'musicfestival'), 0.5075989758124528)
edm', 'productor'), 0.4474996839109835)
music', 'productor'), -0.4394603781883 5393)

Results:

- Costume/ fashion trends
- bikini
- boho
- bohemian

Syracuse University

Conclusion

- The Music Festivals are the rising stars in entertainment industry, worth almost $ 3billion globally.
- While the concept of a music festival hasn't changed in decades, technology and social media are making them more data-driven, more efficient and more accountable.
- Twitter text mining is one of the most crucial tools helping decision makers to reach their goals in the industry.

MS in Applied Data Science Program Learning Goals

Syracuse University

Learning Objective I

A broad overview of the major practice areas in data science

- **Project I - Yellow Vest Movement**
  - Global Viral Event
  - Social Media
  - Text Mining
  - R Libraries
  - ML Applications
    - K-Means Clustering
    - EM—Expectation Maximization
    - HAC—Hierarchical agglomerative clustering
    - Support Vector Machine—SVM
    - Decision Tree
    - Random Forest

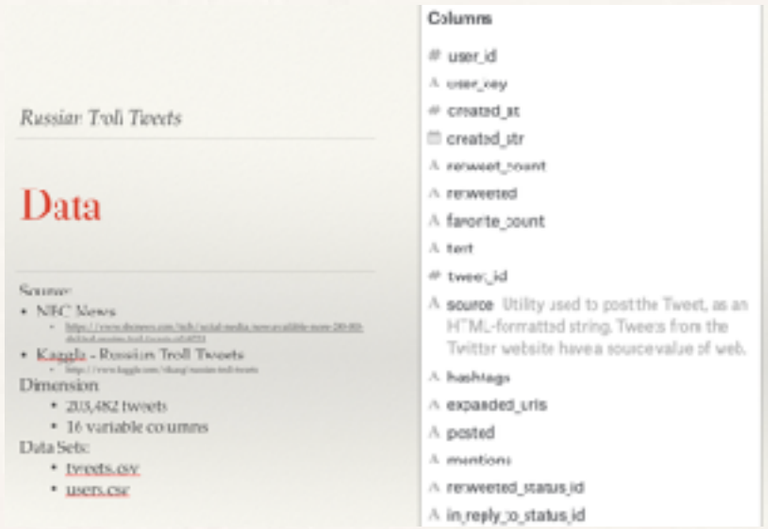
- **Project II - Russian Troll Tweets**
  - Global Viral Event
  - Social Media
  - Python Libraries
  - Text Mining
  - ML Applications
    - K-Means Clustering
    - Support Vector Machine—SVM
    - Decision Tree
    - Random Forest
    - naïve Base algorithms
    - Keras Model

- **Project III - Music Festivals**
  - Global Viral Event
  - Social Media
  - Python Libraries
  - Text Mining
  - NLP
  - Sentiment Analysis

Learning Objective II

Collect and organize data

- Project I:
  - Twitter mining
  - Cleaning
  - Preparation
    - Corpus

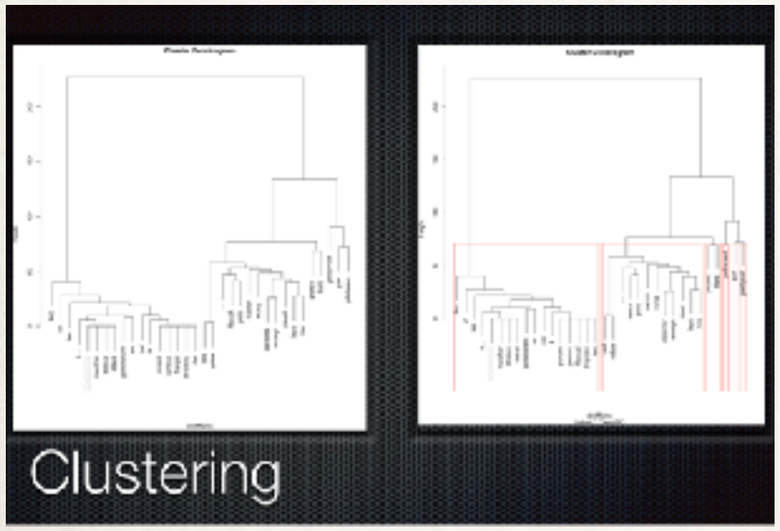


- Project II:
  - Aggregating similar data from many files
  - Sampling

- Project III:
  - Twitter mining
  - Cleaning
  - Preparation
    - Corpus

Learning Objective III

Identify patterns in data via visualization, statistical analysis, and data mining



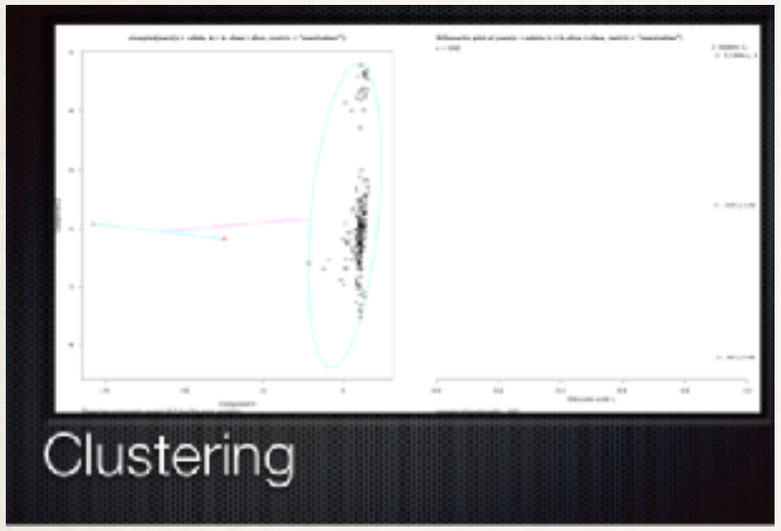
- Project I:
- Word Clouds
- Frequency Word Association table
- Clustering Diagrams

- Project II: Word Clouds
- Frequency Word Association table
- Hashtags
- User Location Charts
- Timeline graphs
- Bigram charts

- Project III:
- Word Clouds
- Frequency Word Association table
- Bigram charts

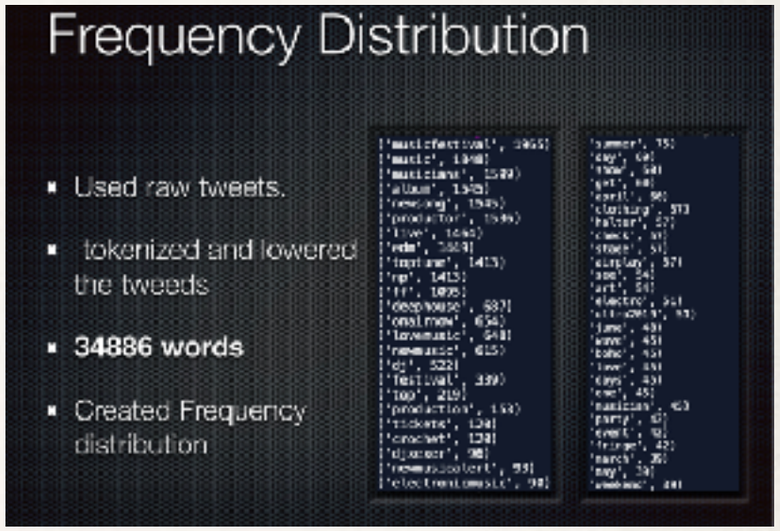
Learning Objective IV

Develop alternative strategies based on the data



- Project I:
- Other Clustering ML application

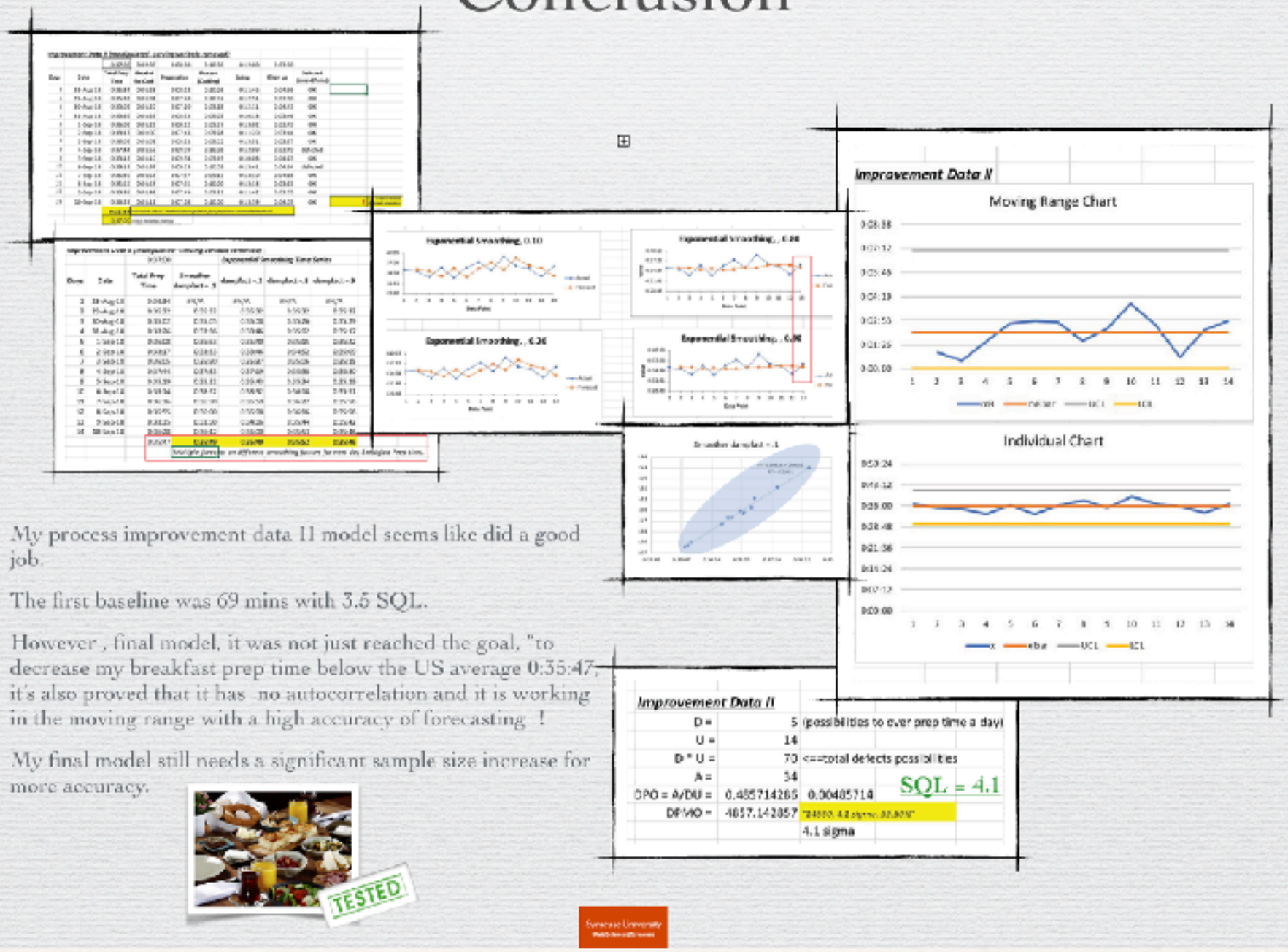
- Project II:
- User Location and Language analysis



- Project III:
- Frequency distribution

Syracuse University

Conclusion



- My process improvement data II model seems like did a good job.
- The first baseline was 69 mins with 3.5 SQL.
- However , final model, it was not just reached the goal, "to decrease my breakfast prep time below the US average 0:35:47, it's also proved that it has no autocorrelation and it is working in the moving range with a high accuracy of forecasting !
- My final model still needs a significant sample size increase for more accuracy.

Develop a plan of action to implement the business decisions derived from the analyses

Learning Objective V

MBC 638, Summer 2018

Privacy - Opening Statement

It is a fundamental human right and essential to a Democratic society

"Privacy is the constitutional core of human dignity."

It is not a fiction, reality

Part III: Surveillance State–China

Demonstrate communication skills regarding data and its analysis for managers, IT professionals, programmers, statisticians, and other relevant professionals in their organization

Learning Objective VI

Learning Objective VII

Synthesize the ethical dimensions of data science practice

- Consideration of Privacy and data usage
- No personal identifier usage
- Honest Portrayal of Result, despite personal opinion
- Ethical responsibility on intellectual property right

Conclusion

- I believe, I've achieved all the objectives of the MS in Applied Data Science Program at Syracuse University.

- Big data is just like a block of marble and information science is the tool that allows me to answer questions that have never been answered before. The application of information science and technology is creating a second renaissance, which will revolutionize everything humans do, forever changing human history.

- As an entrepreneur, I am a life-long learner aspiring to a career in consultant, academia and research.

—Thank you!

Syracuse University