

Interference with Democracy?

Twitter Scraping for Conclusions

Scott Nelson, Quinn Dillard,
Samir Fridhi, and Will Judy

01

HYPOTHESIS & BACKGROUND

Establishing our original hypothesis along with background information.

02

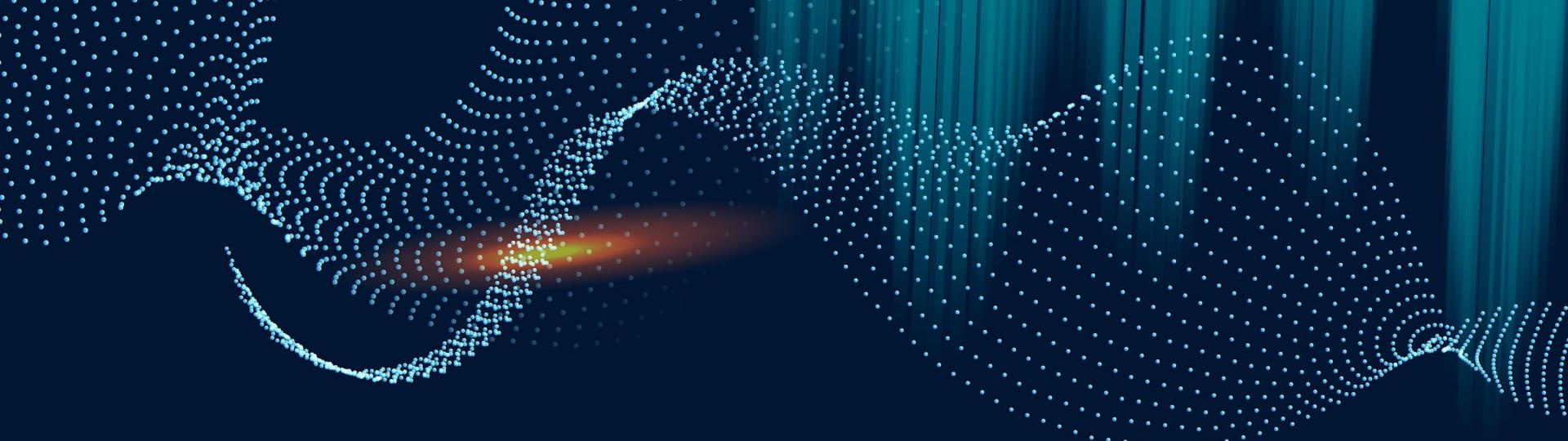
COLLECTION & ANALYSIS

Explaining and demonstrate our collection code and methods employed in our analysis.

03

RESULTS

We will revisit our hypothesis in light of our analysis



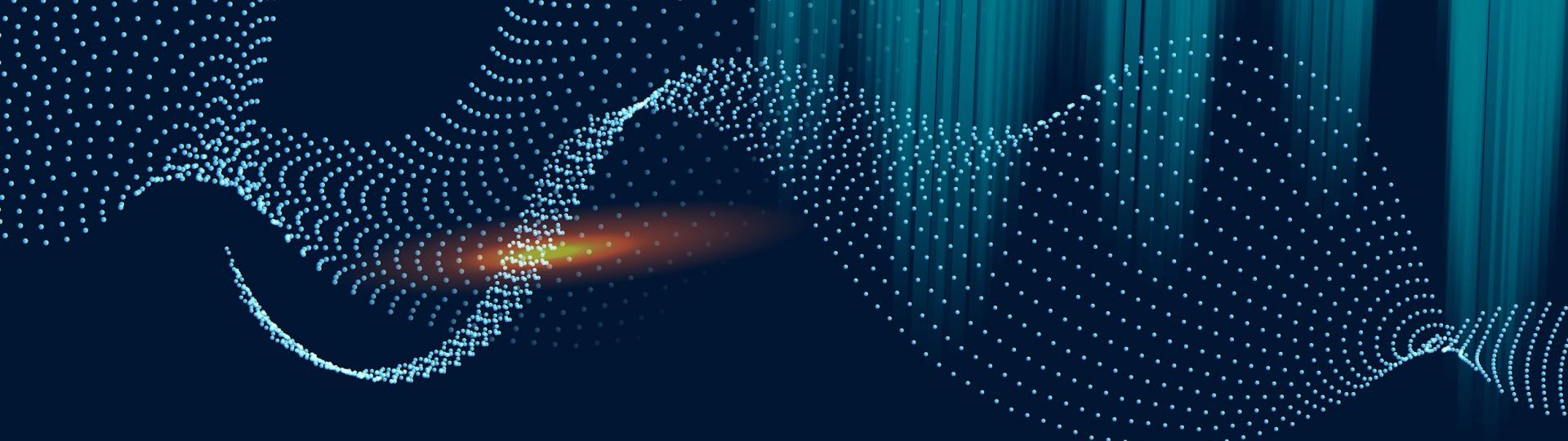
01 | Background & Hypothesis

Background Information:

- Ukraine has been defending their country from a hostile Russian Invasion for the last ten weeks, which has displaced approximately 2.3 million people and has cost the lives 13,000 people, many of them innocent civilians (Reuters).
- This crisis prompted immediate sanctions on Russia, with many companies choosing to suspend business in the country until peace has been restored to the region, such as Apple, Microsoft, and McDonald's.
- Russia has retaliated against many of these companies due to the support for the Ukraine and, as of February 24th, Russia has blocked access to Twitter in the country overall.
- If you may recall from the 2016 elections, it was found that Russia attempted to influence the election results via bots programmed to spread misinformation through this same social media platform.

HYPOTHESIS

“There will be a decline in tweets classified as negative containing the word ‘Democracy’ after Russia was cut-off from Twitter in February 2022.”



02 | Collection & Analysis

Kite: indexing LSP Python: ready conda (Python 3.9.7) Line 106, Col 24 ASCII CRLF RW Mem 20%

File Edit Search Source Run Debug Consoles Projects Tools View Help

C:\Users\Samir Fridhi\BigBoiLocalFlairRatios.py

BigBoiLocalFlairRatios.py x

```
1 import snscrape.modules.twitter as sntwitter
2 import pandas as pd
3 from flair.models import TextClassifier
4 from flair.data import Sentence
5 from segtok.segmenter import split_single
6 import re
7
8 from pathlib import Path
9 classifier = TextClassifier.load('en-sentiment')
10
11 #Flair Setup:https://rileymjones.medium.com/sentiment-analysis-with-the-flair-nlp-library-cfe830bfdf0f4
12 #Flair Advantages:https://www.analyticsvidhya.com/blog/2019/02/flair-nlp-library-python/
13 #Flair Comparison:https://towardsdatascience.com/the-best-python-sentiment-analysis-package-1-huge-common-mistake-d6da9ad6cddeb
14
15 def clean(raw):
16     """ Remove hyperlinks and markup """
17     result = re.sub('<[a][^>]*>(.+)</[a]>', 'Link.', raw)
18     result = re.sub('&gt;', '', result)
19     result = re.sub('&gt;?', '', result)
20     result = re.sub('&quot;', '', result)
21     result = re.sub('&#xF;', '', result)
22     result = re.sub('<p>', '', result)
23     result = re.sub('<i>', '', result)
24     result = re.sub('&#62;', '', result)
25     result = re.sub('<br>', '', result)
26     result = re.sub("\n", ' ', result)
27     return result
28
29 def make_sentences(text):
30     """ Break apart text into a list of sentences """
31     if len(str(tweet.content))==0 or str(tweet.content).strip()=="":
32         return "NaN"
33     #sentences = [sent for sent in split_single(text)]
34     sentences=[]
35     for sent in split_single(text):
36         if sent=="":
37             sent="Empty Sentence"
38         else:
39             sentences.append(sent)
40     return sentences
41
42 def predict(sentence):
43     """ Predict the sentiment of a sentence """
44     text = Sentence(sentence)
45     if sentence=="" or text.text == "":
46         value=0
47     else:
48         print(tweet.date)
49         print(tweet.id)
50         print(text)
51         classifier.predict(text)
```

x -02- Replace with: -01- no matches Find next In selection All

Source Console Object

Usage

Here you can get help of any object by pressing CtrlH in front of it, either on the Editor or the Console.

Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.

New to Spyder? Read our tutorial

Help Variable Explorer Plots Files

Console 1/A x

```
those backwards savages to bring democracy !
2020-01-06 02:34:05+00:00
1214012227956727809
Sentence: "# JNUattack our we going to leave this type of legacy behind in the
world 's biggest democracy and secular"
2020-01-06 02:33:09+00:00
121401199654842776
Sentence: "On Sunday morning talk shows, Mike Pompeo stated he believes escalating
tensions with Iran will help pro democracy Iranians seize power. In truth , Iranian
hardliners are suddenly much more popular. Pompeo is either a liar , a fool , or
both "
2020-01-06 02:32:52+00:00
1214011923869655040
Sentence: "Early 00 's - We invade Iraq based on dubious evidence of WMDs , claim
to be bringing democracy ."
2020-01-06 02:32:52+00:00
1214011923869655040
Sentence: "Now - Trump wants to sanction the country if they exhibit their
democracy to boot the U.S. out because we assassinated a foreign country 's
general on their soil ."
2020-01-06 02:32:52+00:00
1214011923869655040
Sentence: "Crazy ."
```

IPython console History

Collection Method:

- Imports:
 - Pandas: Used to create a DataFrame of our totals and proportions of positive and negative tweets.
 - SNScrape Twitter Module: Used to scrape Twitter for tweets containing “Democracy” in a particular time frame.
 - Flair’s Text Classifier & Sentence:
 - Sentence: Used to represent a sentence or text fragment as a list of tokens which will be used to classify the sentiment of the sentence or text fragment.
 - Text Classifier: Used to classify text as positive or negative after it has been broken up into sentences and passed through “Sentence”
 - Segtok’s split_single: Used in the “make_sentences” function to break up each tweet into individual sentences.
 - PathLib’s Path: Used to save the data in csv file locally.
 - Regular Expression: Used to strip raw text of any markups or hyperlinks.

Collection Method (Cont.):

- Functions:
 - clean(raw): Passes in raw text, stripping it of any hyperlinks and/or markups using the Regular Expression operations. Passes to the make_sentences function.
 - make_sentences(text): Splits the text into sentences, stores in an array. Passes to the predict function.
 - predict(sentences): Employs Flair's Sentence which translates each sentence into tokens and classifies each sentence in the sentence array.
 - get_scores(sentences): Scores each classified Sentence in the sentence array and appends the results to an empty results array.
 - get_sum(scores): Sums the scores of each sentence of a tweet & converts to proportions between 0 & 1

Collection Method (Cont.):

- For Loop/Filter:
 - Due to the sheer volume of tweets mentioning Democracy, from the beginning we made a calculated decision about which tweets we should filter out as to limit our scope, control for context, and collect only those tweets necessary for our analysis.
 - As such, we used SNScrape Twitter Module attribute, such as “.inReplyToUser” and ”.lang”, to limit ourselves to “original” tweets, i.e not in response to another tweet, and specifically to tweets in English.
 - Furthermore, we had to limit our tweets even further, filtering out any tweets that are empty in tweet.content, i.e tweets with no content but a photo and “#Democracy”, and tweets which either mention another user or just have the “@” symbol in the tweet.

The background features a dark blue gradient with a subtle dotted texture. A prominent feature is a diagonal beam of light that transitions from orange at the bottom left to yellow at the top right, with a slight lens flare effect.

Questions?

What we collected: 20 CSV files

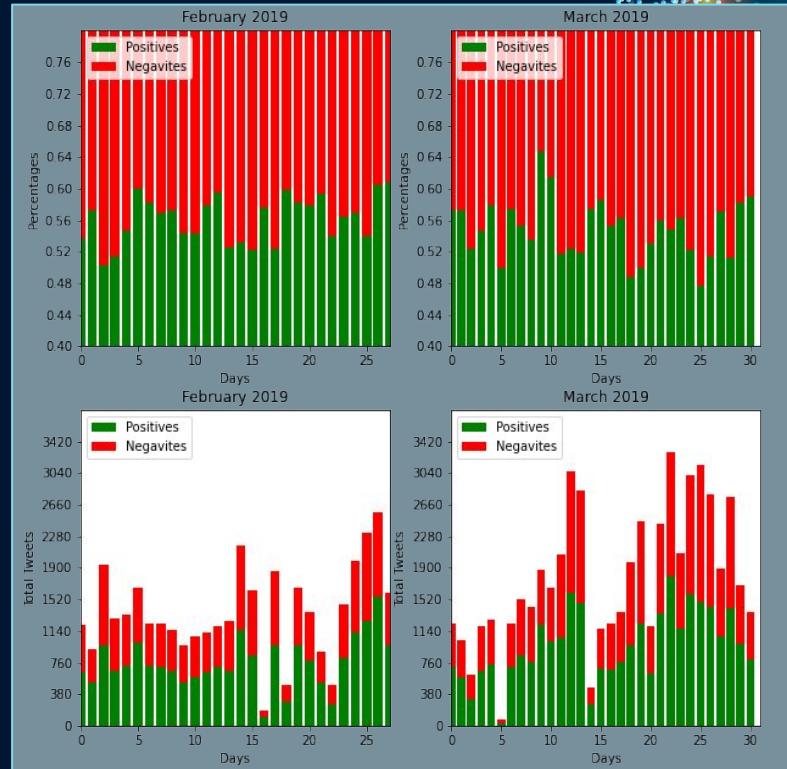
“Democracy” - March & February of 2018, 2019, 2020, 2021, & 2022

“Democracy” - January of 2020 & 2021

“Fake News” - March & February of 2018, 2020, 2021

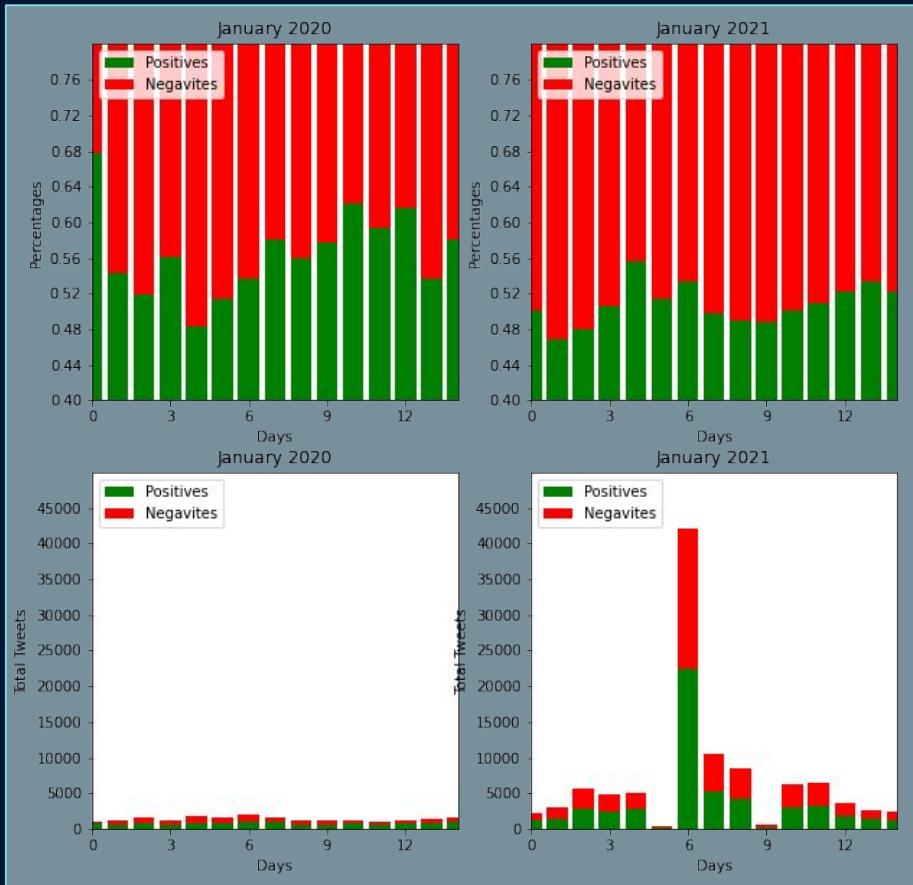
“Conservatives” & “Liberals” - March of 2020

Barplots - Percentages & Total Tweets Over Time



*Spikes indicate significant events related to democracy: Here, Parkland School Shooting

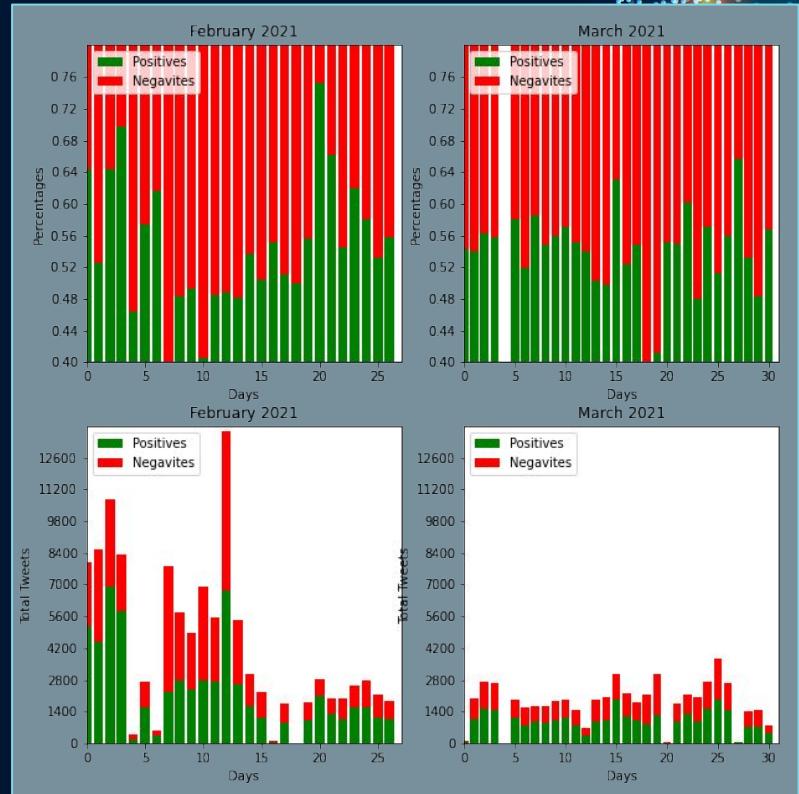
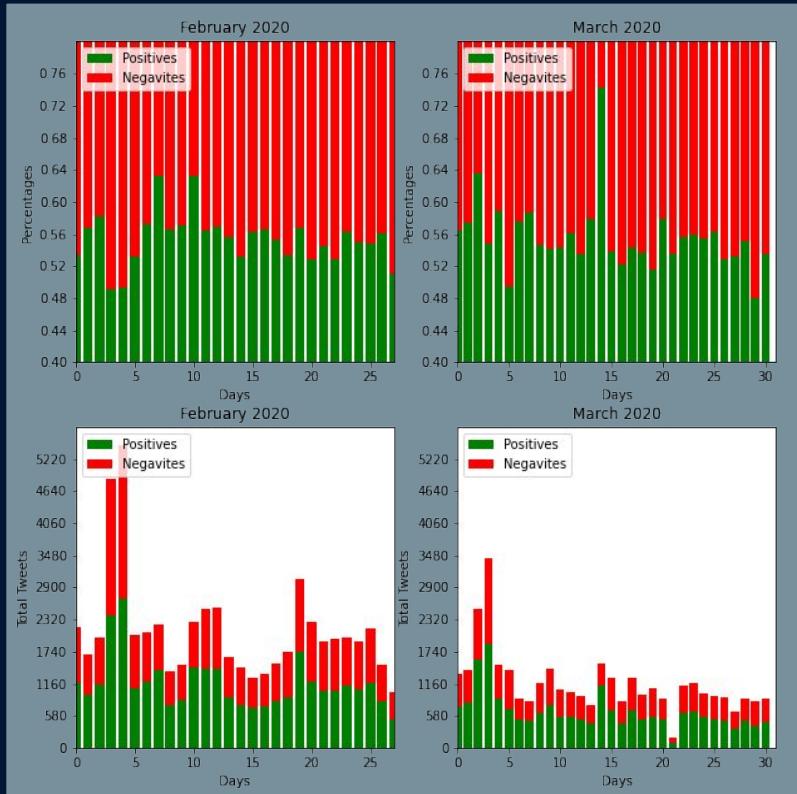
Barplots - January 2020 & January 2021



Insurrection Day!
January 6th 2021

43,000 Tweets!
(Relating to Democracy)

Barplots - Percentages & Total Tweets Over Time



Barplots - Percentages & Total Tweets Over Time



FEBRUARY AND
MARCH OF 2022

There does look to be
some general decline
in frequency, but is
there statistical
significance?

Statistical Significance

FEBRUARY

Totals:

```
[[1.      0.22533455 0.16908162 0.06097507 0.03899181]
 [0.22533455 1.      0.11624394 0.0958885 0.3831999 ]
 [0.16908162 0.11624394 1.      0.38153203 0.23483536]
 [0.06097507 0.0958885 0.38153203 1.      0.30645536]
 [0.03899181 0.3831999 0.23483536 0.30645536 1.      ]]
```

Positives:

```
[[1.      0.25233103 0.17426943 0.11043279 0.1108268 ]
 [0.25233103 1.      0.21742536 0.08776487 0.07352754]
 [0.17426943 0.21742536 1.      0.24188112 0.16207258]
 [0.11043279 0.08776487 0.24188112 1.      0.07219655]
 [0.1108268 0.07352754 0.16207258 0.07219655 1.      ]]
```

Negatives:

```
[[1.      0.18914532 0.05874983 0.11464815 0.0780033 ]
 [0.18914532 1.      0.21742536 0.08776487 0.07352754]
 [0.05874983 0.21742536 1.      0.24188112 0.16207258]
 [0.11464815 0.08776487 0.24188112 1.      0.07219655]
 [0.0780033 0.07352754 0.16207258 0.07219655 1.      ]]
```

Total Averages

```
[[0.29887661 0.36413338 0.38033859 0.36897019 0.39269649]]
```

Positive Averages

```
[[0.32957201 0.32620976 0.3591297 0.30245506 0.28372469]]
```

Positive Averages

```
[[0.28810932 0.31357262 0.33602578 0.30329814 0.27715999]]
```

MARCH

Totals:

```
[[1.      0.05325158 0.07595482 0.09022659 0.13439658]
 [0.05325158 1.      0.43536688 0.18993099 0.29187471]
 [0.07595482 0.43536688 1.      0.2383682 0.04904017]
 [0.09022659 0.18993099 0.2383682 1.      0.22577531]
 [0.13439658 0.29187471 0.04904017 0.22577531 1.      ]]
```

Positives:

```
[[1.      0.16432366 0.03323714 0.05008119 0.01667178]
 [0.16432366 1.      0.07518646 0.04002844 0.26933002]
 [0.03323714 0.07518646 1.      0.13571556 0.03322068]
 [0.05008119 0.04002844 0.13571556 1.      0.0893605 ]
 [0.01667178 0.26933002 0.03322068 0.0893605 1.      ]]
```

Negatives:

```
[[1.      0.10384811 0.24184245 0.18754881 0.10837899]
 [0.10384811 1.      0.07518646 0.31416093 0.26933002]
 [0.24184245 0.07518646 1.      0.09550872 0.03322068]
 [0.18754881 0.31416093 0.09550872 1.      0.05776274]
 [0.10837899 0.26933002 0.03322068 0.05776274 1.      ]]
```

Total Averages

```
[[0.27076592 0.39408483 0.35974601 0.34886022 0.34021735]]
```

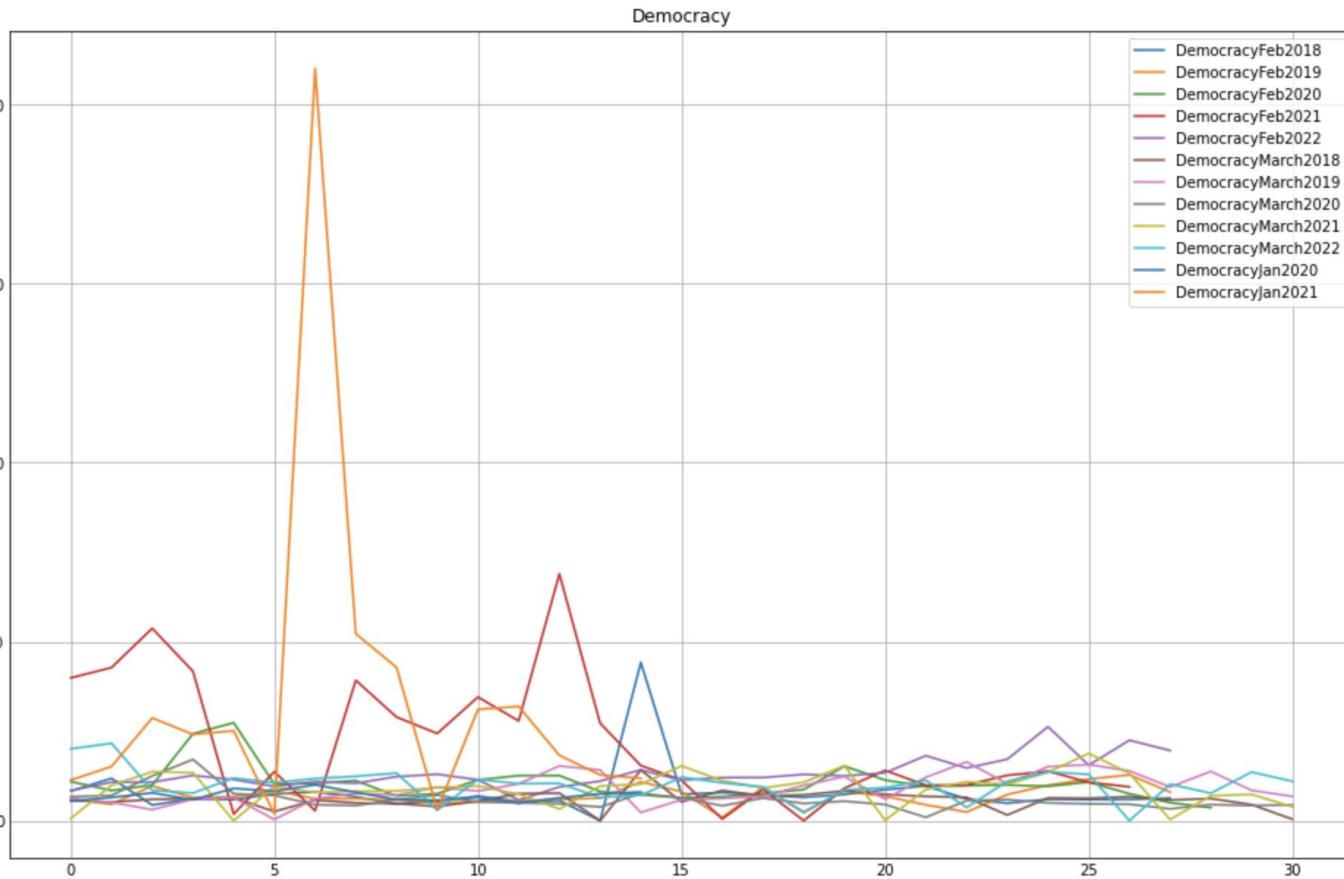
Positive Averages

```
[[0.25286275 0.30977372 0.25547197 0.26303714 0.2817166 ]]
```

Positive Averages

```
[[0.32832367 0.3525051 0.28915166 0.33099624 0.29373849]]
```

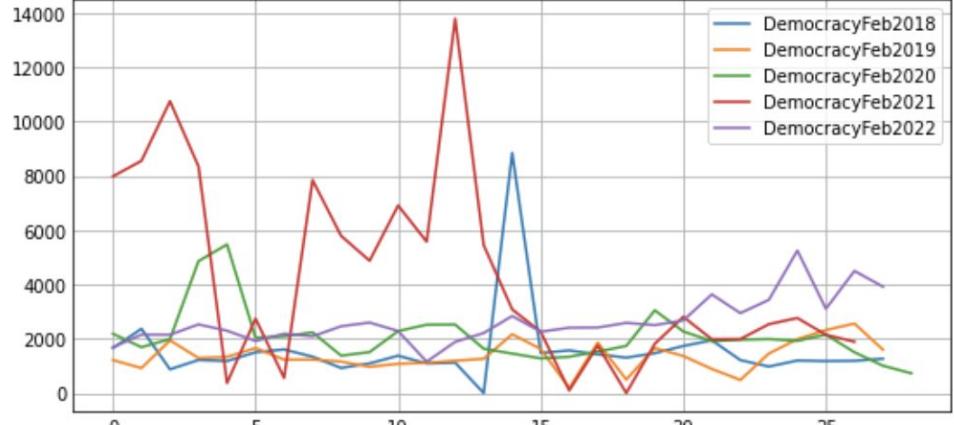
Total Tweets over all months



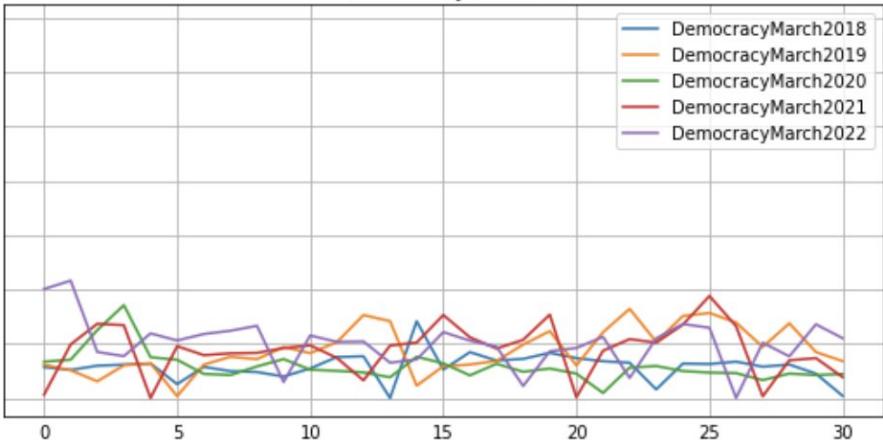
Democracy totals from February to March

- Years 2018-2022

Democracy February

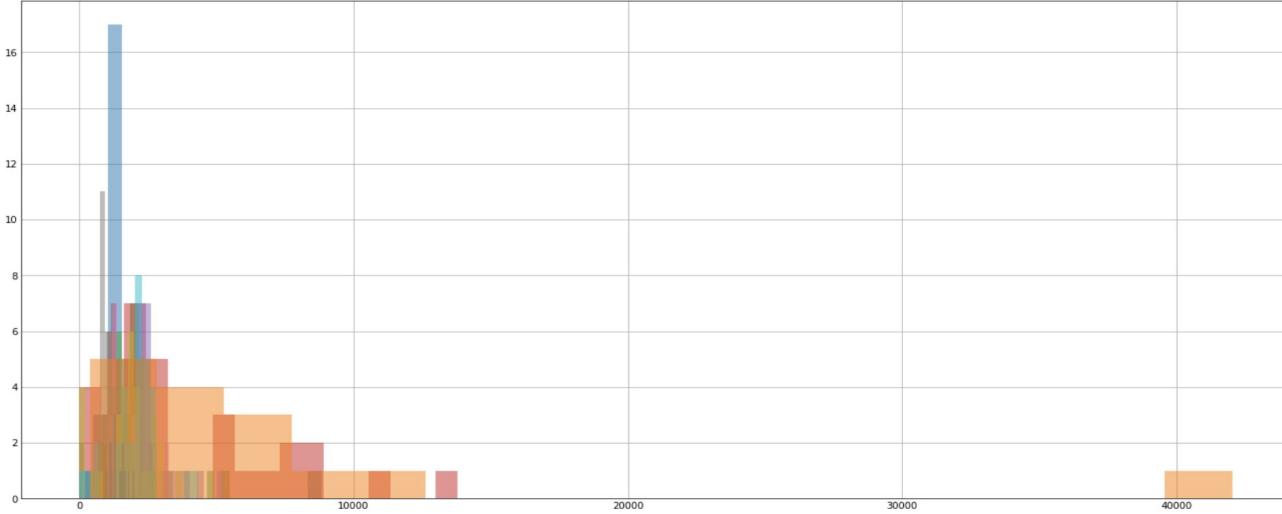


Democracy March

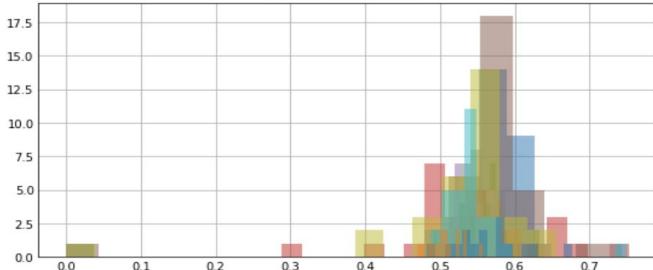


Distribution for Total, Positive, Negative

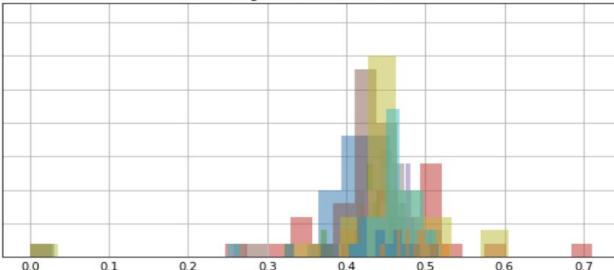
Total TWEET distribution



Positive Sentiment %



Negative Sentiment %



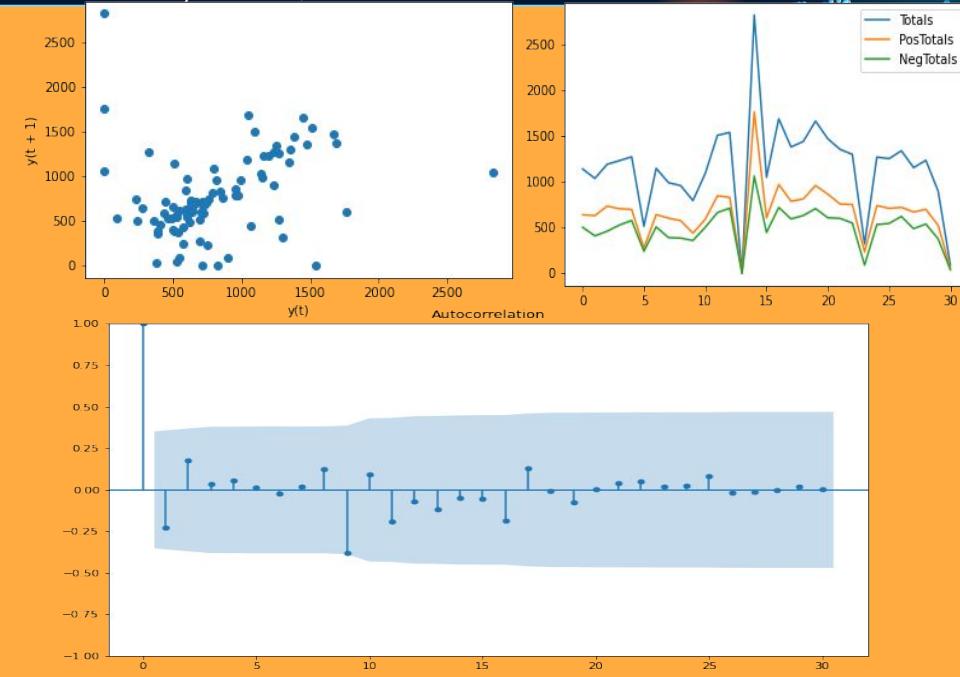
- DemocracyFeb2018
- DemocracyFeb2019
- DemocracyFeb2020
- DemocracyFeb2021
- DemocracyFeb2022
- DemocracyMarch2018
- DemocracyMarch2019
- DemocracyMarch2020
- DemocracyMarch2021
- DemocracyMarch2022
- DemocracyJan2020
- DemocracyJan2021
- DemocracyFeb2018
- DemocracyFeb2019
- DemocracyFeb2020
- DemocracyFeb2021
- DemocracyFeb2022
- DemocracyMarch2018
- DemocracyMarch2019
- DemocracyMarch2020
- DemocracyMarch2021
- DemocracyMarch2022
- DemocracyJan2020
- DemocracyJan2021

Analysis Method:

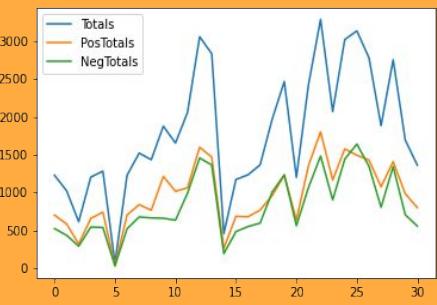
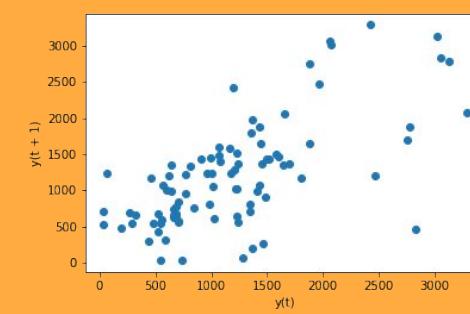
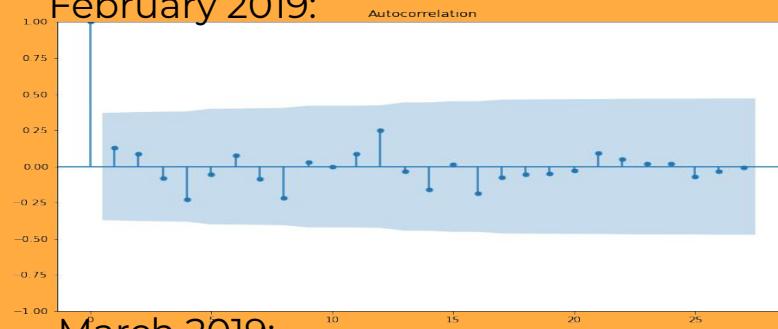
*Feb is left, March is right

- Time Series:

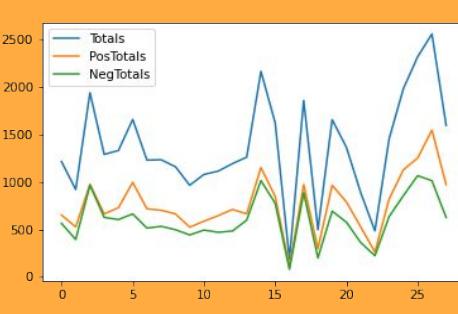
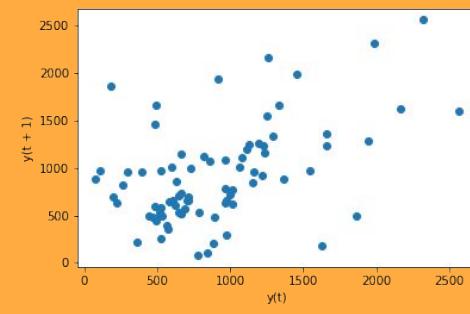
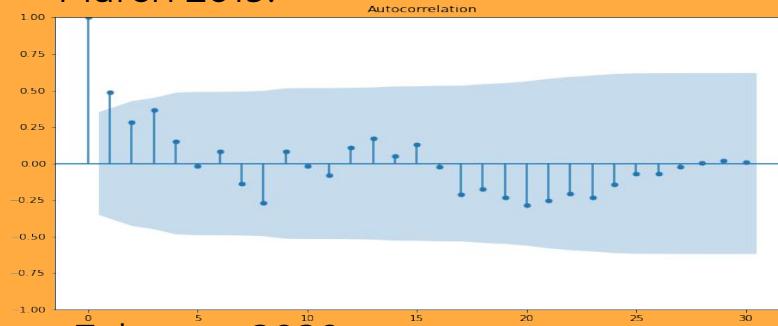
- As our first analysis technique, we used Time Series Analysis to produce Series, Lag, and Autocorrelation plots for each month we collected; seen below for 2018*



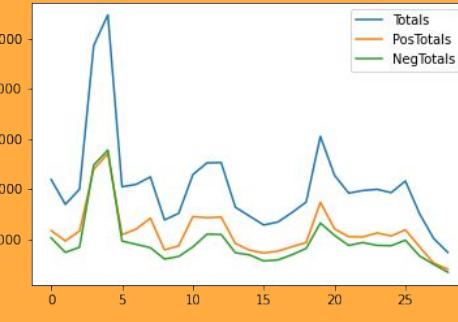
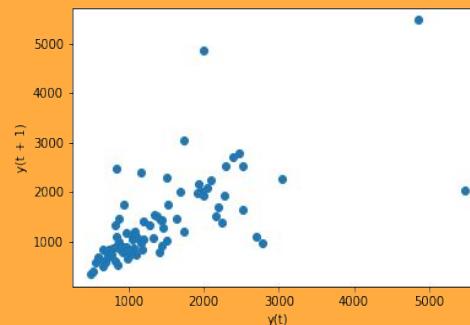
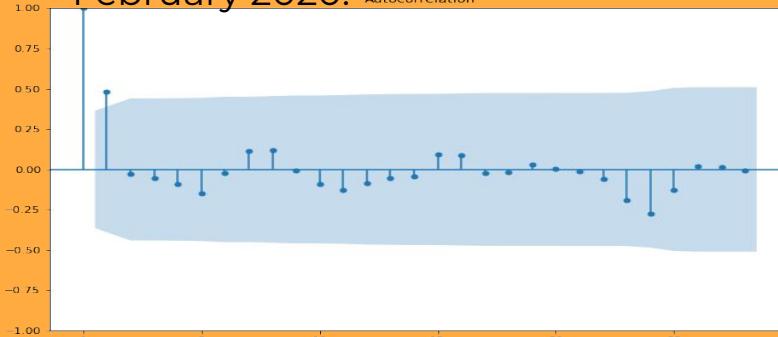
February 2019:



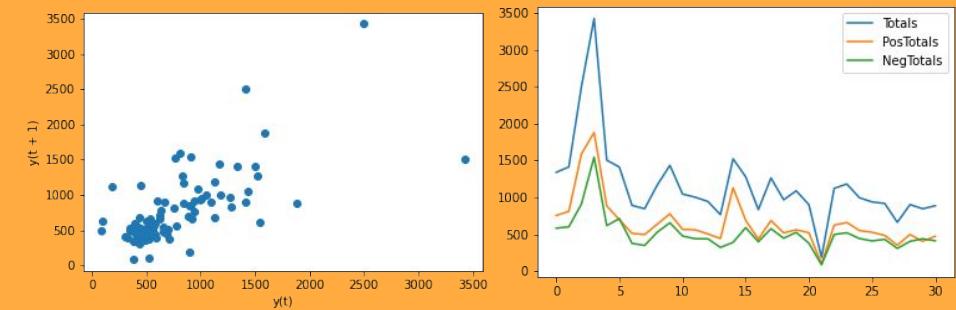
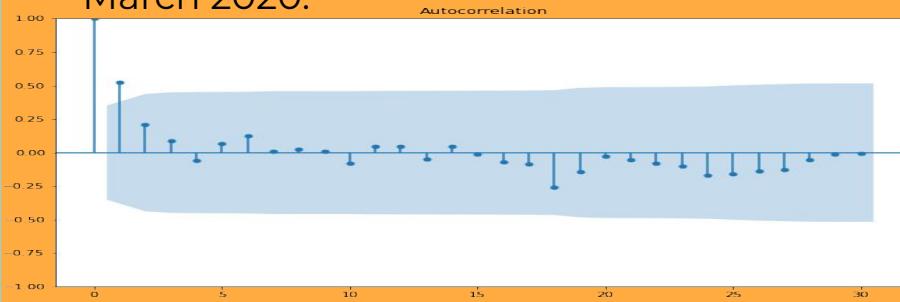
March 2019:



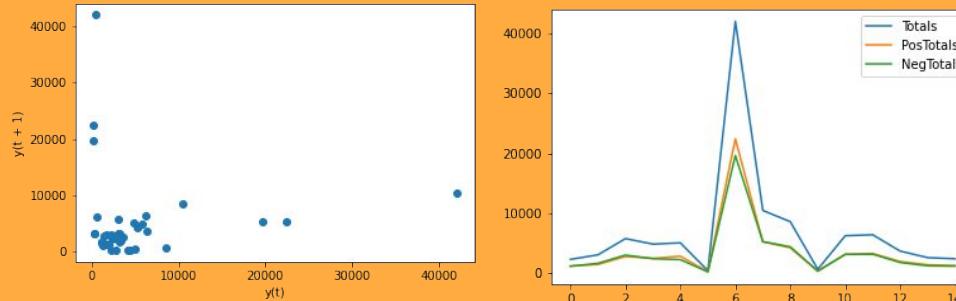
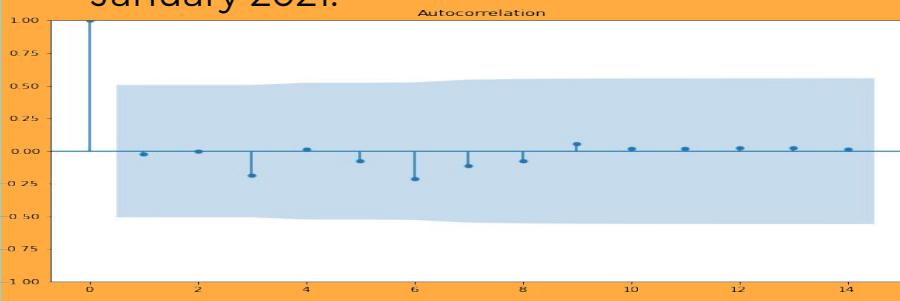
February 2020:



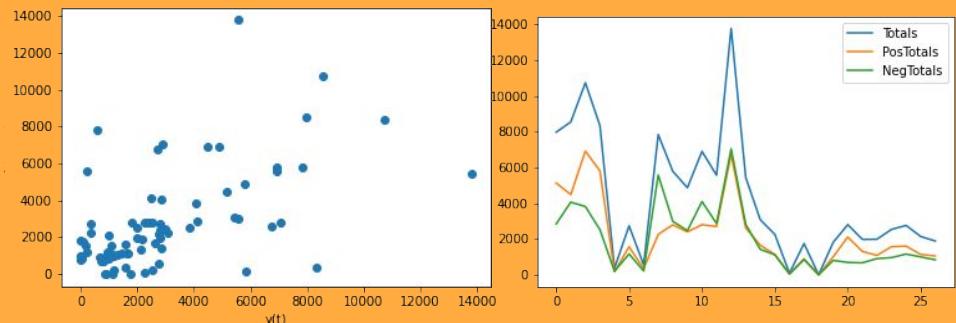
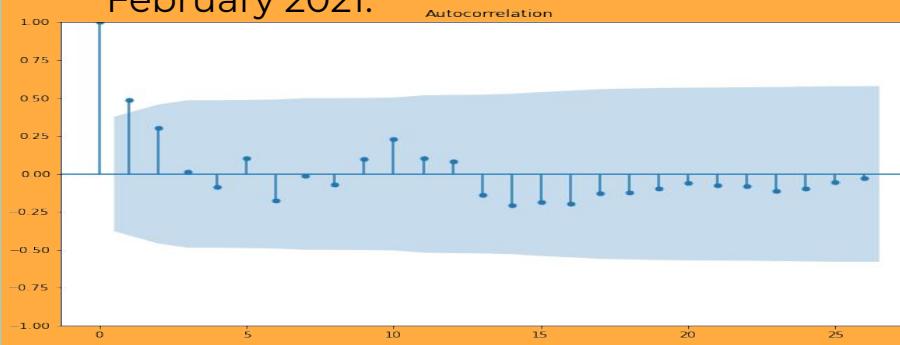
March 2020:



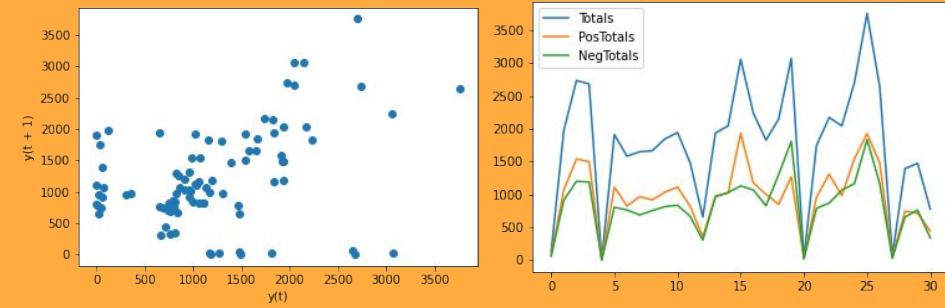
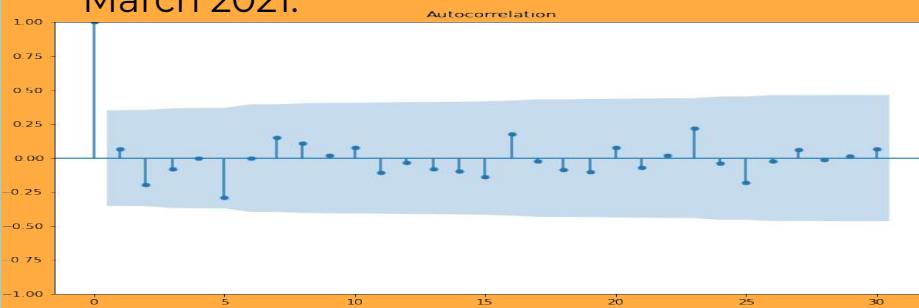
January 2021:



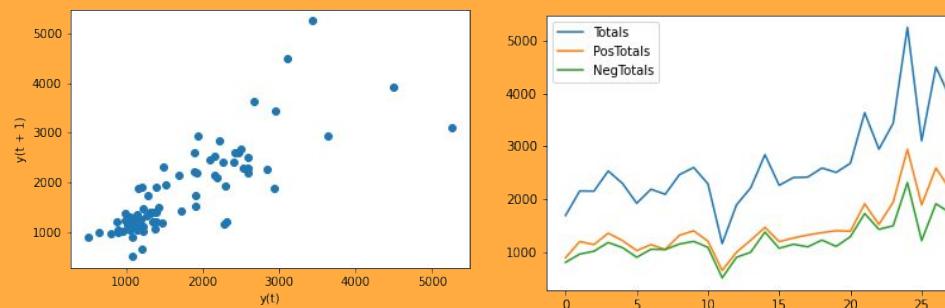
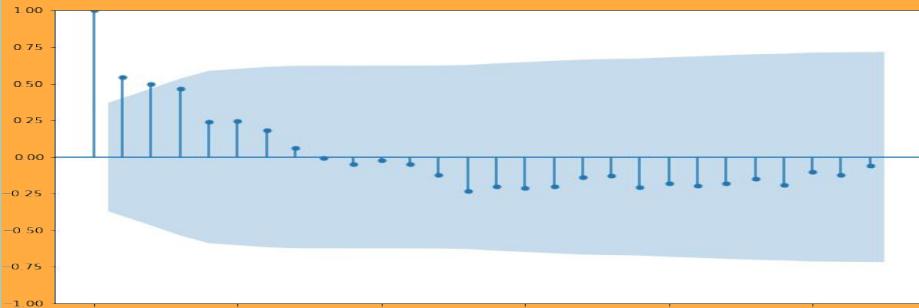
February 2021:



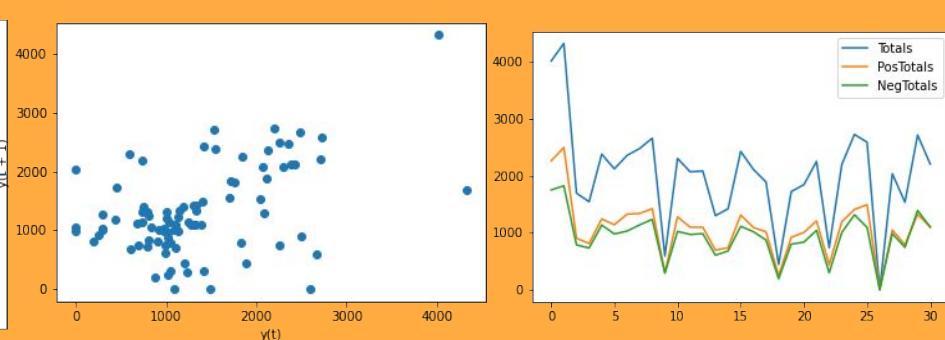
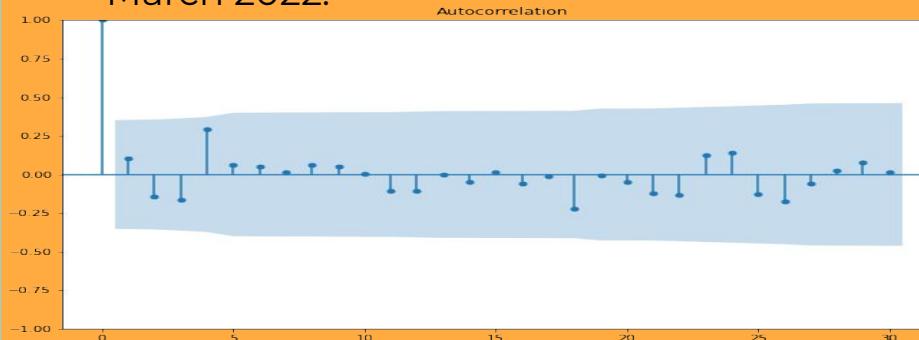
March 2021:

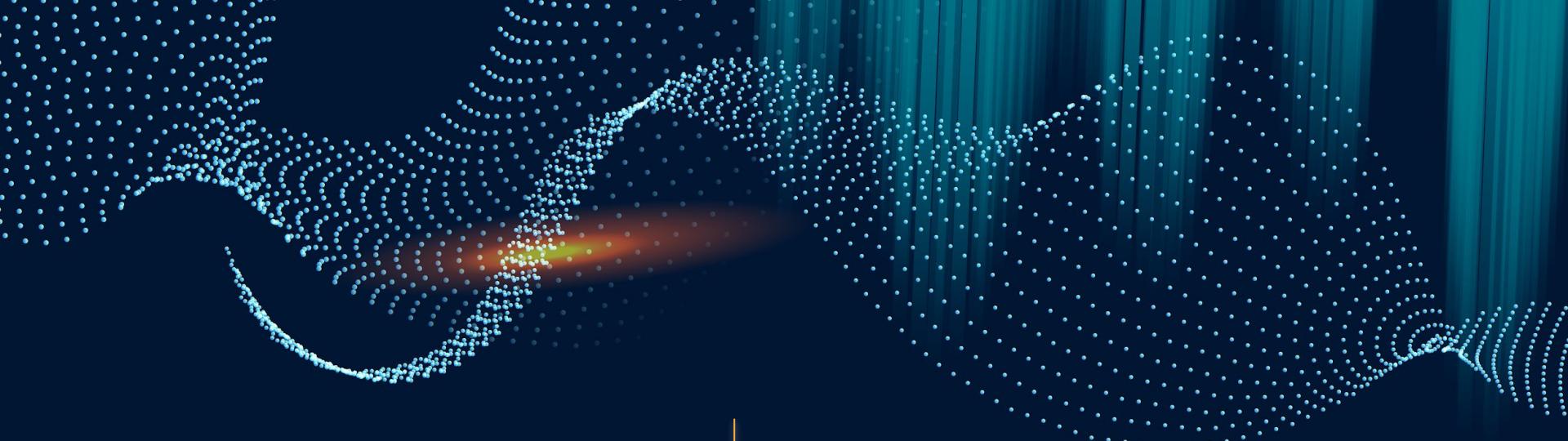


February 2022:



March 2022:





03 | Results

~~HYPOTHESIS~~

“There will be a decline in tweets classified as negative containing the word ‘Democracy’ after Russia was cut-off from Twitter in February 2022.”

We found no correlation!

Sources

<https://github.com/JustAnotherArchivist/sns scrape>

<https://rileymjones.medium.com/sentiment-analysis-with-the-flair-nlp-library-cfe830bfd0f4>

<https://www.analyticsvidhya.com/blog/2019/02/flair-nlp-library-python/>

<https://towardsdatascience.com/the-best-python-sentiment-analysis-package-1-huge-common-mistake-d6da9ad6cdeb>

Slidesgo.com for theme :)



**Thank You!
Any Questions?**
