



IDENTIFYING POLITICAL PLATFORMS BASED ON CONGRESSIONAL TWEETS

using NLP TOPIC Modeling and Sentiment Analysis

UNSUPERVISED LEARNING PROJECT

MIKE BERNARDO

INTRODUCTION: WHY TWITTER?

450k+ Tweets from Congressional Accounts in 2021.



TWITTER AS BULLHORN

- followers as constituents



CONVERGENCE OF MEDIA

- same as traditional media*



TWITTER AS FOCUS GROUP

- interactive feedback

*** PEW Research: The News Agenda on Twitter vs. Traditional Platforms**

UNSUPERVISED LEARNING PROJECT GOAL:

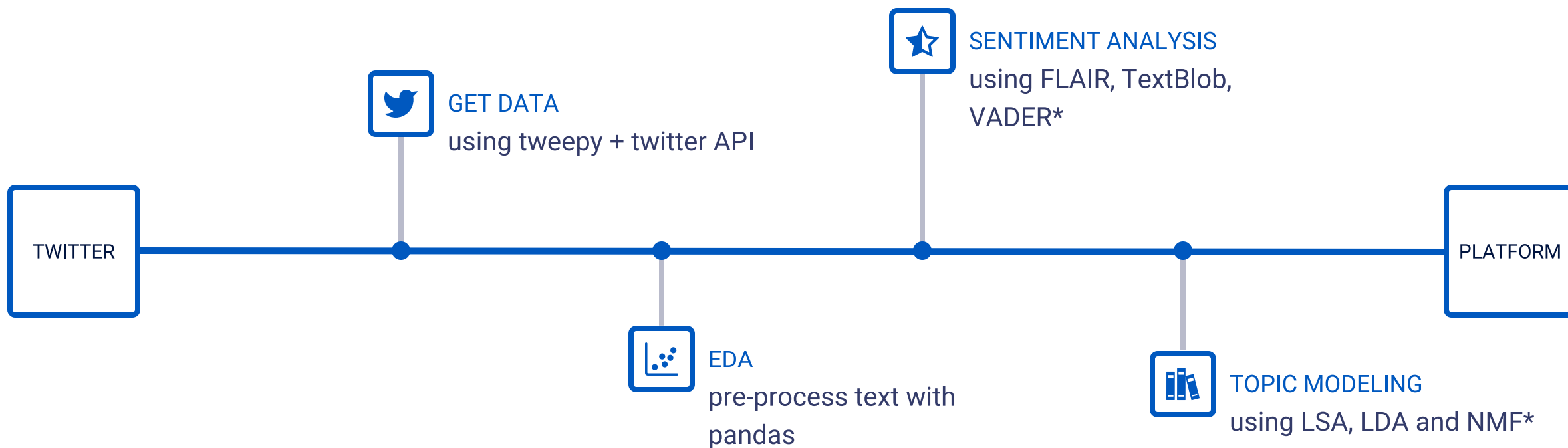
POLITICAL PLATFORMS

using NLP topic modeling on tweets



PROJECT WORKFLOW

From Tweets to Platform using NLP

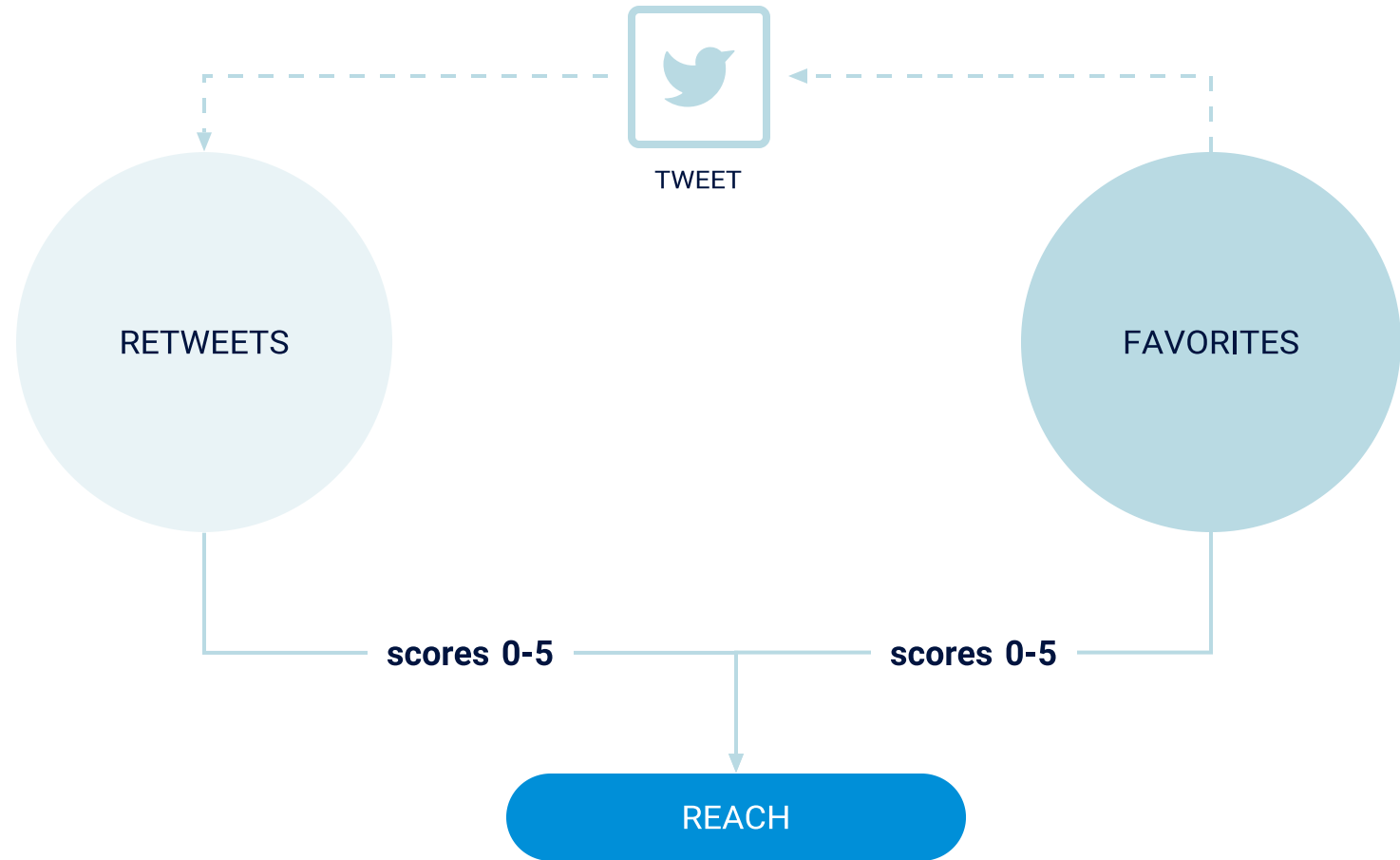


* INDICATES FINAL MODEL USED FOR ANALYSIS

FEATURE ENGINEERING:

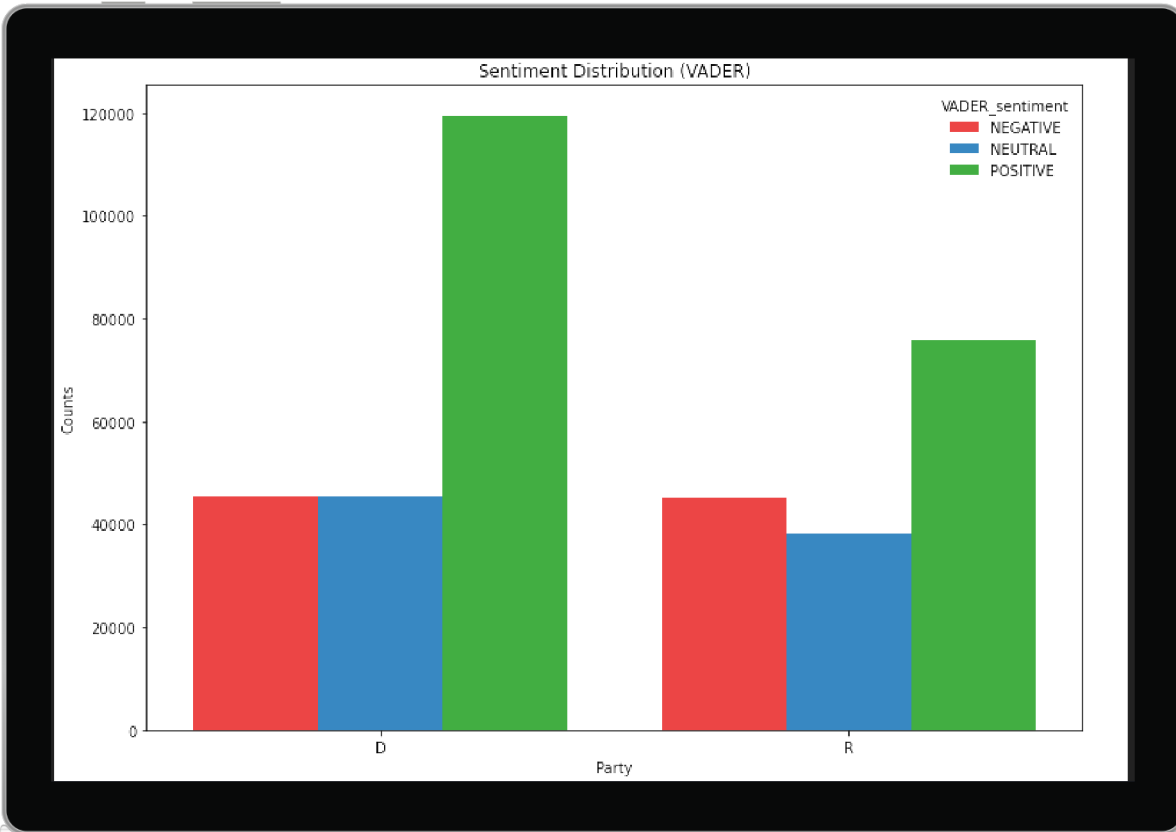
Reach is the aggregation of scaled favorites and retweets.

Values are from 0-10 that mimic how viral the tweet is.



SENTIMENT ANALYSIS: VADER

VADER rates sentiment as positive, neutral or negative and also gives a compound score.



● DEMOCRATS

- more positive tweets
- mostly about policies
- better 'Reach'

● REPUBLICANS

- 38% less positive tweets
- less retweets, less favorites
- mostly criticism of administration
- has suspended and banned users

NMF TOPIC MODELING

Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words	Topic 2 weights	Topic 3 words	Topic 3 weights	Topic 4 words	Topic 4 weights	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights
		biden	3.1										
american	2.2	president	2.1	infrastructure	2.6	back	2.3	house	1.5	covid	2.4	health	2.4
families	1.5			bipartisan	1.8	build	1.9	act	1.4	vaccine	1.6	care	2.1
people	1.4	president biden	1.6	bip					1.0	get	1.3	health care	1.4
great	1.3	border	1.0	infrastructure					1.0	covid vaccine	1.1	child	0.6
year	1.2	administration	0.9						0.8	vaccinated	0.7	child care	0.5
work	1.0	biden administration	0.8	infrastructure					0.7	get vaccinated	0.4	affordable	0.5
one	1.0	joe	0.7	investment	0.8	act	0.7	passed	0.7	vaccines	0.4	mental	0.5
time	1.0	joe biden	0.7	infrastructure					0.7	pandemic	0.3	mental health	0.5
country	0.9	crisis	0.7	investment					0.5	safe	0.3	access	0.4
working	0.9	southern	0.3	investment					0.5	free	0.3	coverage	0.3

Today, the Supreme Court confirmed what we all knew: @POTUS Biden's vaccine mandate on employees of private busines...
<https://t.co/TXdVdK0aSh>

Joe Biden is the president of high prices. It's a new year, and the same old record-high #Bidenflation

TOPIC 1: Biden Administration

TOP 7 TOPICS

Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	NMF_Topic_name	Polarity	Compound	Reach	Topic 5 words	Topic 5 weights	Topic 6 words	Topic 6 weights
american	2.2	biden	3.1	Covid	0.170293	0.236002	127055.5	covid	2.4	health	2.4
families	1.5	president	2.1	Voting Legislation	0.108495	0.123206	68327.5	vaccine	1.6	care	2.1
people	1.4	president biden	1.6	American Rescue Plan	0.096555	0.181038	34636.5	get	1.3	health care	1.4
great	1.3	border	1.0	Infrastructure	0.142306	0.239092	20022.5	covid vaccine	1.1	child	0.6
year	1.2	administration	0.9	Healthcare	0.118381	0.276057	18691.0	vaccinated	0.7	child care	0.5
work	1.0	biden administration	0.8	Biden Presidency	0.091025	0.067997	15144.5	get vaccinated	0.4	affordable	0.5
one	1.0	joe	0.7	Build Back Better Act	0.195808	0.381086	9433.0	vaccines	0.4	mental	0.5
time	1.0	joe biden	0.7					pandemic	0.3	mental health	0.5
country	0.9	crisis	0.7					safe	0.3	access	0.4
working	0.9	southern	0.3					free	0.3	coverage	0.3

REPUBLICAN TOPICS

Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights
joe	3.4	border	2.8
joe biden	3.3	crisis	1.6
biden	2.4	southern	1.2
america	0.8	southern border	1.2
biden america	0.6	border crisis	0.8
americans	0.4	crisis southern	0.4
inflation	0.3	illegal	0.4
presidency	0.2	biden border	0.4
crisis	0.2	security	0.3
failed	0.2	crisis border	0.3

ANALYSIS OF REPUBLICAN DATA

- corpus is from only GOP tweets
- higher weight values make us more certain of the topic purity
- required more tweaking of the hyper-parameters

REPUBLICAN TOPICS

	GOP_Topic_name	Polarity	Compound	Reach
5	Legislators	0.152556	0.229516	133514.0
3	Democratic Policies	0.041885	-0.054332	30635.0
6	Vaccine	0.066065	0.050272	18392.0
2	Border Crisis	-0.012526	-0.274275	14707.5
0	Biden Administration	0.004381	-0.109418	14572.0
1	Biden Policies	0.046002	-0.023681	13823.5
4	Joe Biden	0.015303	-0.104610	8787.5

TOP 7 TOPICS AND SENTIMENT

- compound values are mostly negative
- focused on 'personality politics'
- tweets are mostly criticism of the administration

DEMOCRAT TOPICS

Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights
families	2.2	infrastructure	2.3
working	1.4	bipartisan	1.6
tax	1.2	bipartisan infrastructure	1.4
child	1.2	jobs	0.8
work	1.2	investment	0.8
childtaxcredit	1.1	law	0.7
help	1.1	infrastructure investment	0.6
year	1.0	investment jobs	0.6
time	0.9	jobs act	0.6
great	0.9	infrastructure law	0.5

ANALYSIS OF DEMOCRAT DATA

- corpus is from only DEM tweets
- words are more spread out with lower topic weights
- looking at individual tweets helped name topics

DEMOCRAT TOPICS

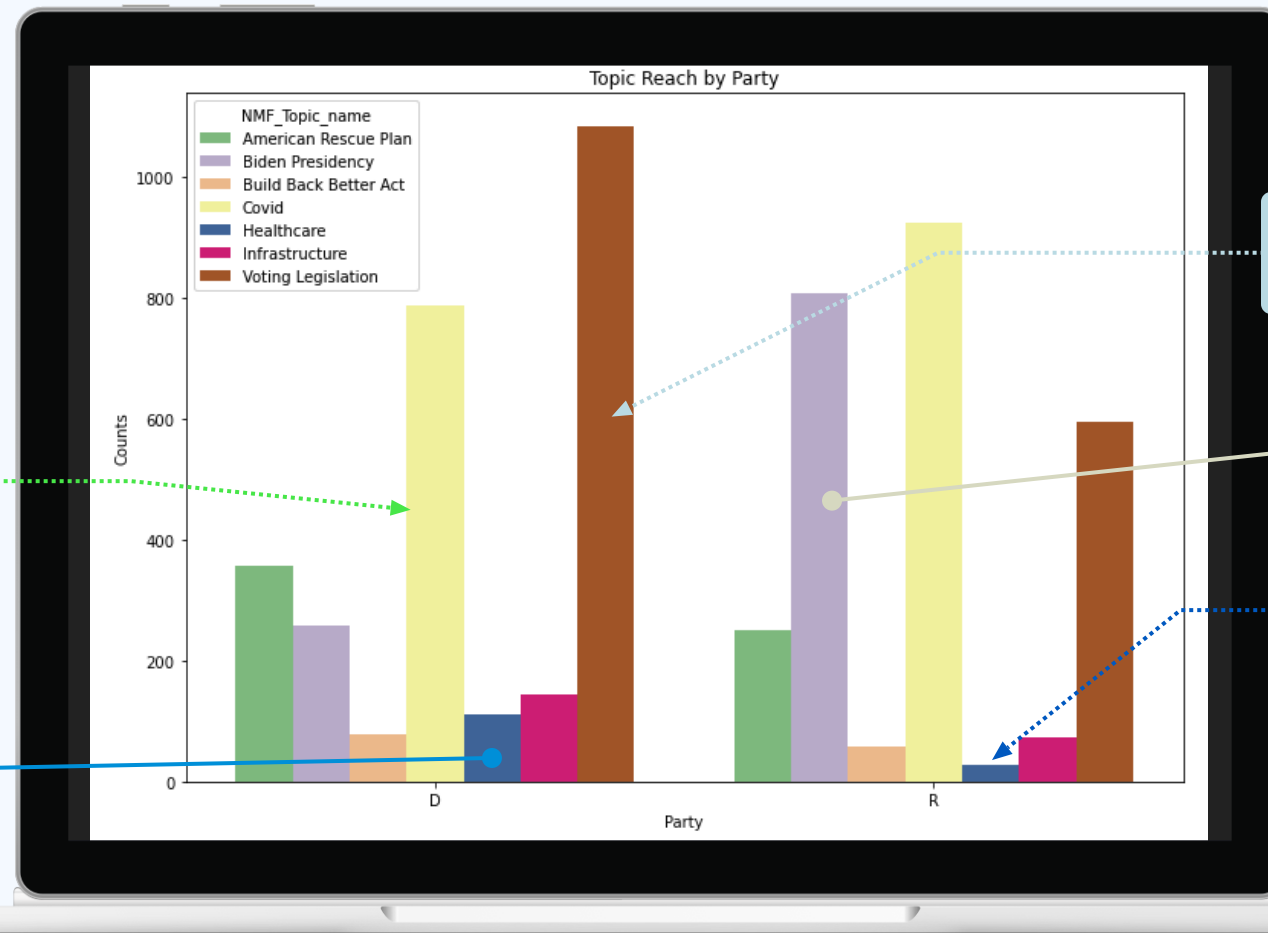
	DEM_Topic_name	Polarity	Compound	Reach
3	Family Support	0.143774	0.202108	158463.5
6	Voting Rights	0.114395	0.132944	50102.0
2	Covid	0.172595	0.191930	33996.5
5	Infrastructure	0.135326	0.227733	23058.5
1	Build Better Act	0.152016	0.220048	21217.5
4	Healthcare	0.121407	0.290886	18716.0
0	American Rescue Plan	0.086170	0.276870	13238.0

TOP 7 TOPICS AND SENTIMENT

- compound values are mostly positive
- focused on policy
- tweets are mostly about policy and legislation

THE TOPICS USING FULL CORPUS

VISUALIZATION OF 'REACH'



COVID
STILL DOMINANT TOPIC

Infrastructure
LEGISLATION PASSED

voting legislation
MAJOR PART OF AGENDA

Biden Presidency
MOSTLY NEGATIVE

healthcare
LOW PRIORITY FOR REPUBLICANS

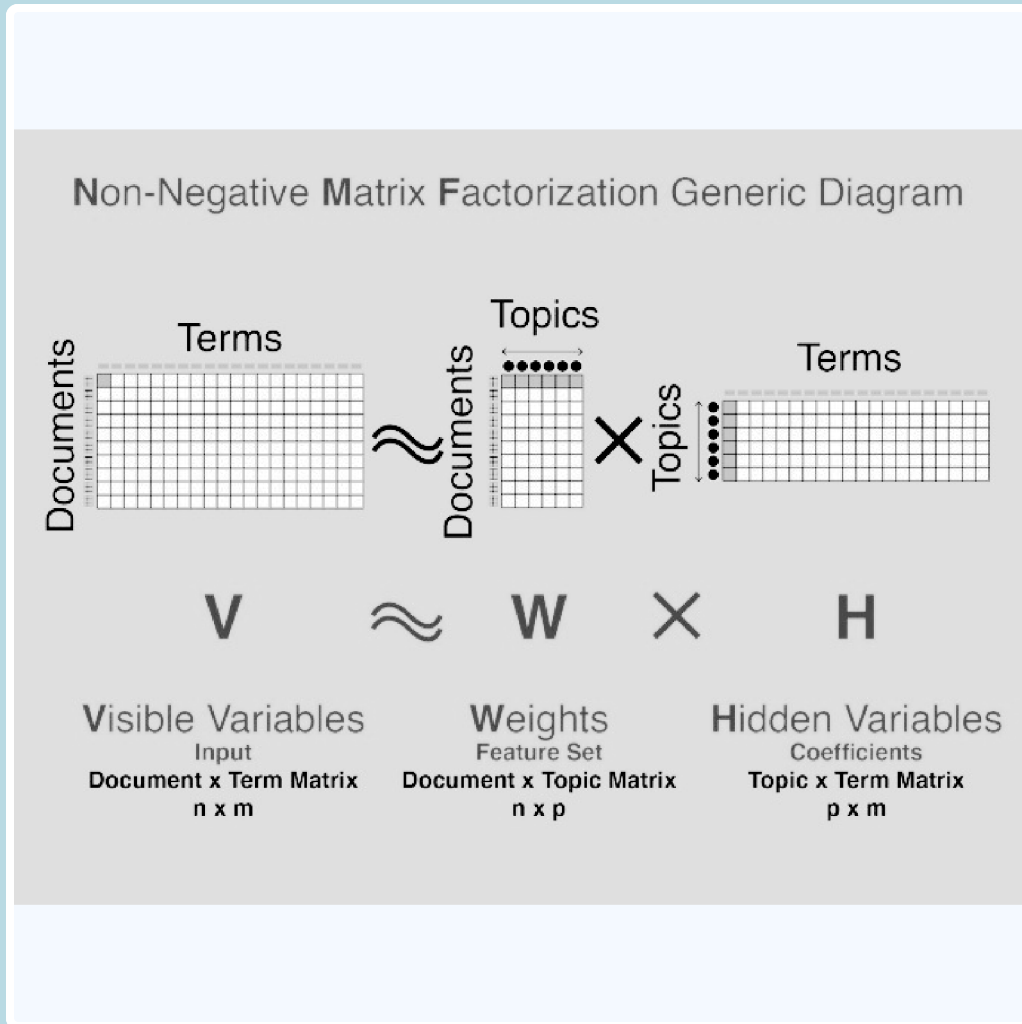
TRIGRAMS

	Republican Trigrams	count
0	crisis southern border	461
1	chinese communist party	405
2	brave men women	397
3	biden border crisis	354
4	biden vaccine mandate	345
5	critical race theory	321
6	supply chain crisis	312
7	tax spending spree	294
8	law enforcement officers	289
9	border patrol agents	229

	Democrat Trigrams	count
0	american rescue plan	2104
1	child tax credit	1449
2	build better act	1281
3	infrastructure investment jobs	923
4	investment jobs act	876
5	bipartisan infrastructure law	771
6	john lewis voting	643
7	lewis voting rights	602
8	good paying jobs	574
9	martin luther king	480



TOPIC MODELING: NON-NEGATIVE MATRIX FACTORIZATION (NMF)



MATRICES

- V = frequency of Terms aka a count of each word's occurrence.
- H = topic or semantic feature composed of frequencies
- W = weights or a semantic feature that recurs throughout the corpus

TOPIC MODELING: NON-NEGATIVE MATRIX FACTORIZATION (NMF)

Topic 0 words	Topic 0 weights	Topic 1 words	Topic 1 weights	Topic 2 words
american	2.2	biden	3.1	infrastructure
families	1.5	president	2.1	bipartisan
people	1.4	president biden	1.6	bipartisan infrastructure
great	1.3	border	1.0	bill
year	1.2	administration	0.9	infrastructure bill
work	1.0	biden administration	0.8	jobs
one	1.0	joe	0.7	investment
time	1.0	joe biden	0.7	infrastructure investment
country	0.9	crisis	0.7	law
working	0.9	southern	0.3	investment jobs

TOPICS BASED ON WORDS:

- topic 0 : American Families
- topic 1: President Biden

TOPIC MODELING: NON-NEGATIVE MATRIX FACTORIZATION (NMF)

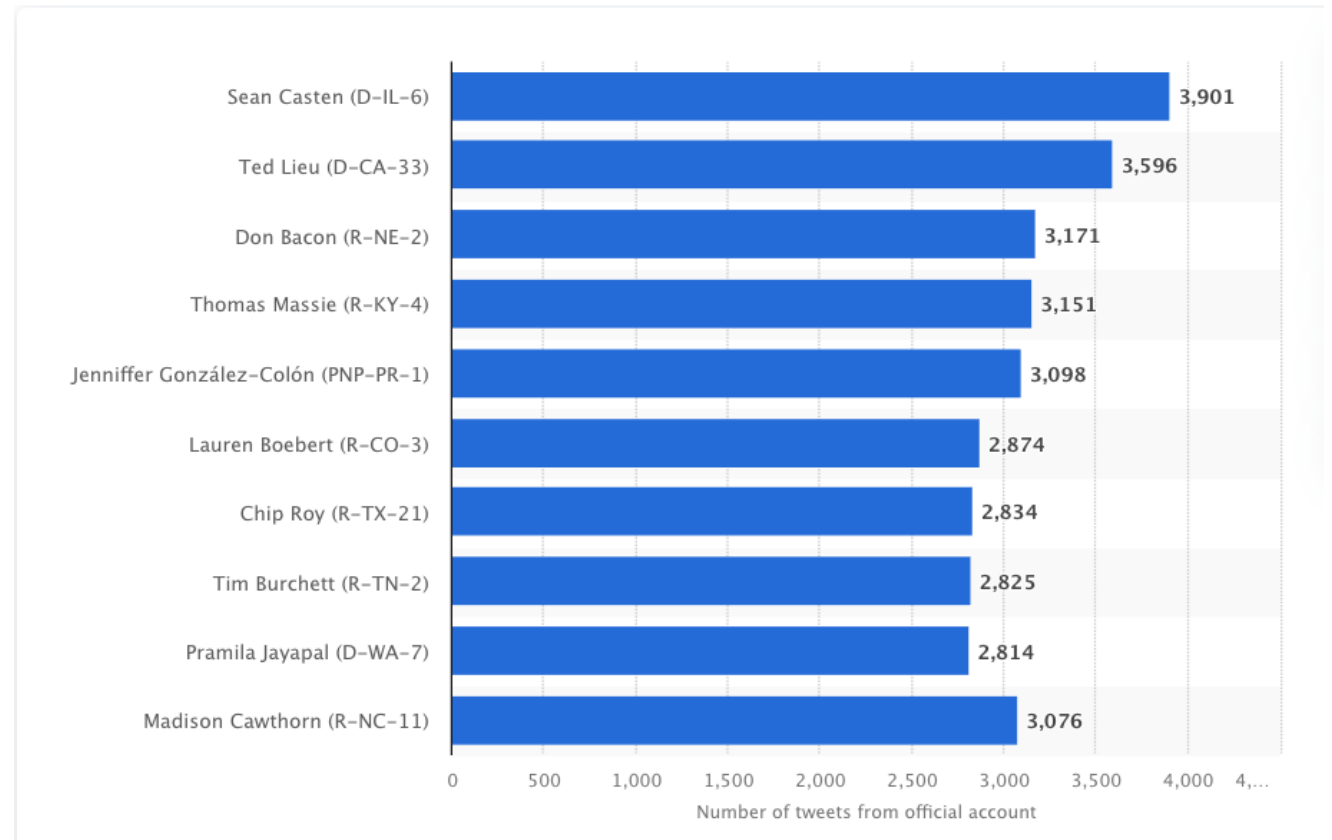
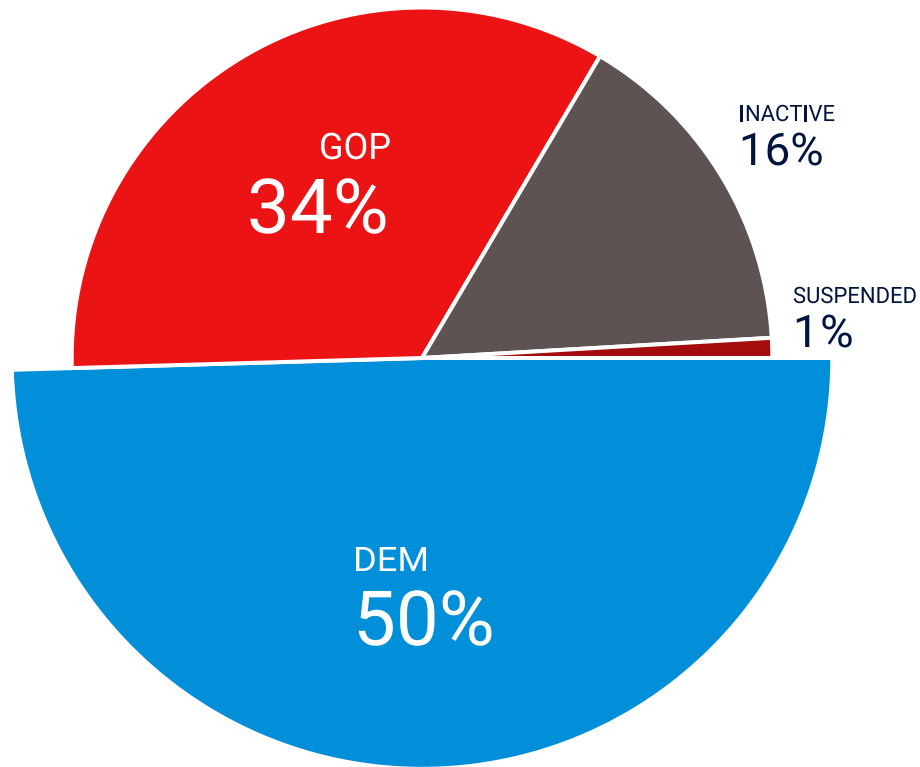
Covid	167226
Voting Legislation	70383
Biden Presidency	40490
American Rescue Plan	40158
Infrastructure	23414
Healthcare	17900
Build Back Better Act	9685

TOP 7 TOPICS BY TWEET COUNT

- election debate topics
- party platform

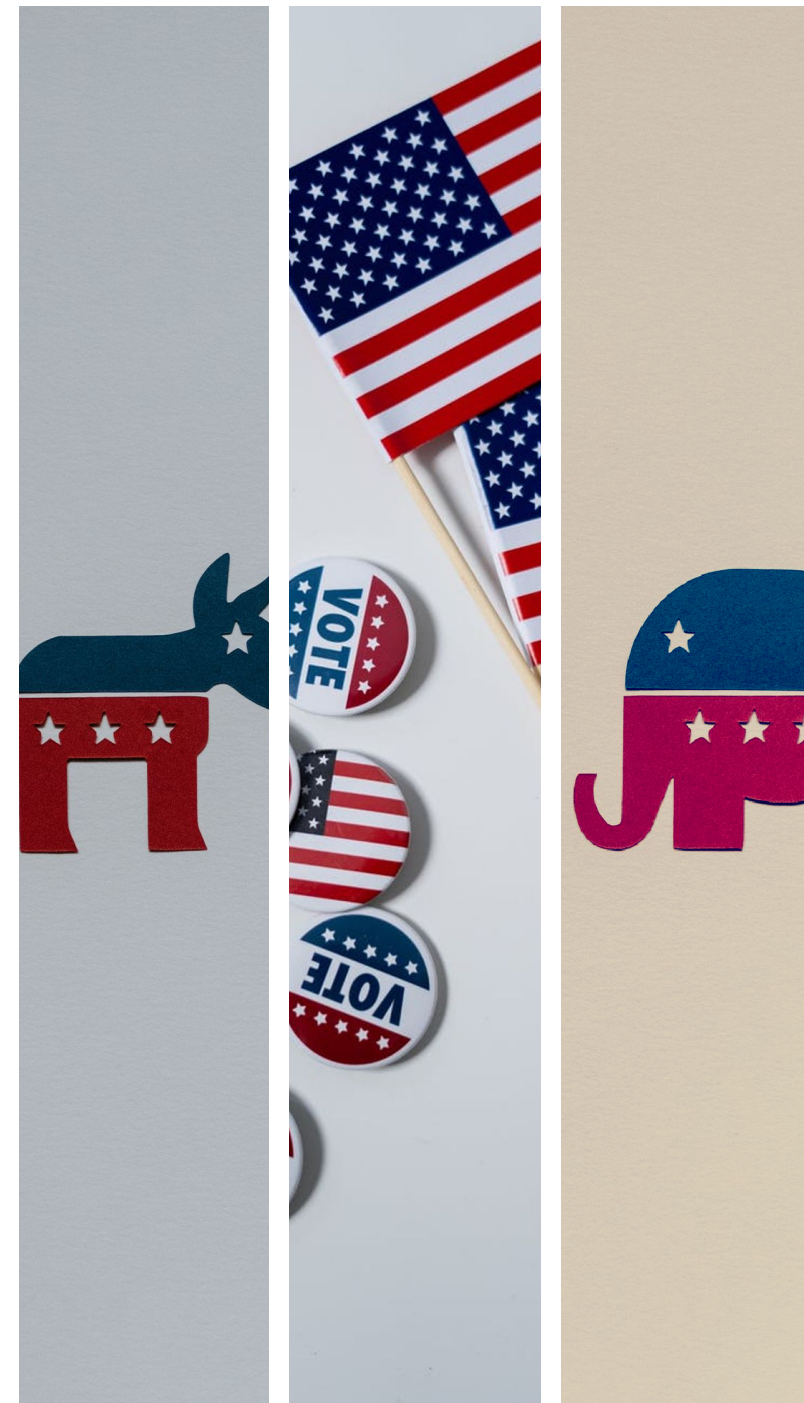
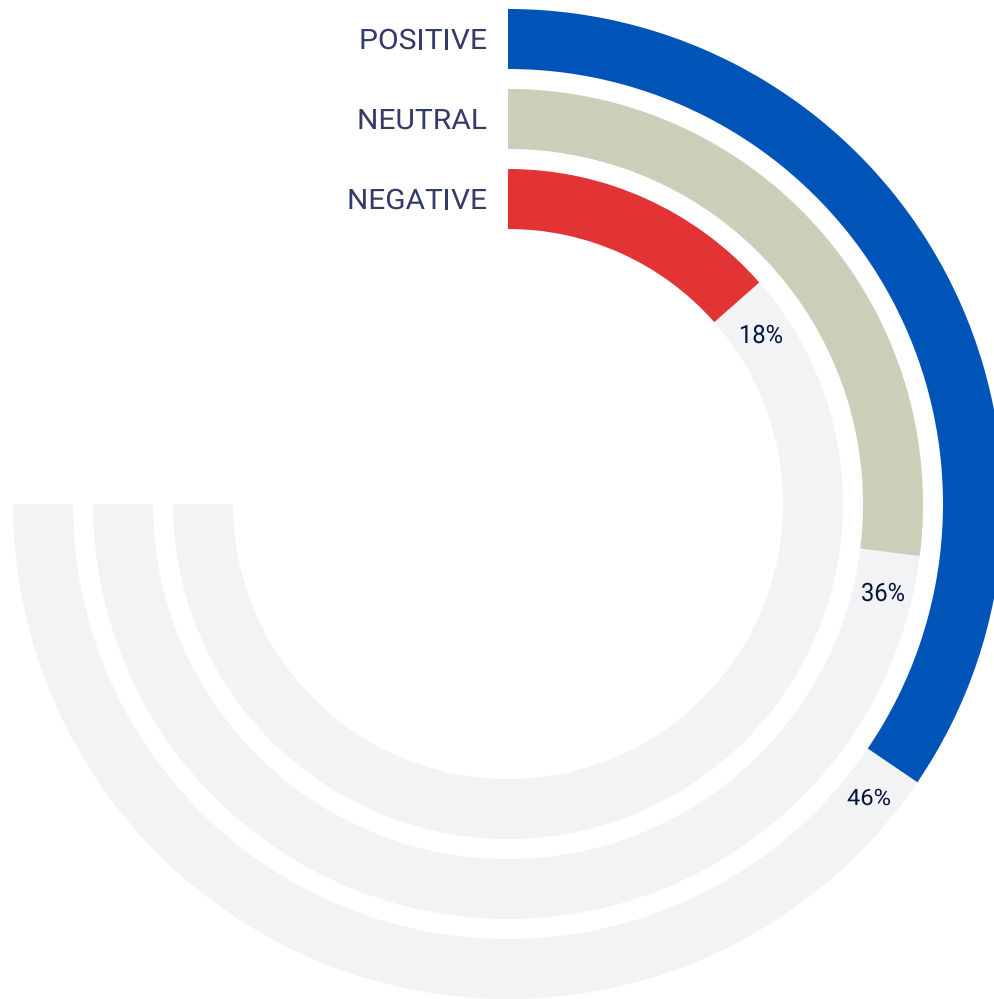
DATA: CONGRESSIONAL ACCOUNTS

456k Tweets in from 1 Jan 2021 - 20 Jan 2022



OVERALL SENTIMENT

The Political vibe on Twitter



CLEANING TWITTER DATA

to reduce noise and prepare for NLP modeling

