Sentiment Analysis on Republicans and Democrats focused on US presidential elections

Department of Computer Science

Oklahoma State University

Sai Tharuni Samineni

A20346521

Aim:

The final goal of the project is to categorize the sentiment of various states towards each of the Democrats and Republicans parties.

Methodology:

The research is focused on the US Presidential Elections and aims to comprehend amount of public sentiment and attitude toward the Republican and Democratic parties in various US states.

I gathered tweets using the search word republicans and Democrats using twitter api and scrapped the tweets from the starting of this year to present. For this analysis, I have collected tweets made over the last two years using Twitter's publicly available API and performed sentiment analysis of those tweets using VADER Sentiment Analysis. Used multiple data frames to manipulate the data to get the results of the sentiment.

Algorithm Choice:

For the purpose of sentiment analysis in the project I have tried working with some rule-based sentiment analysis algorithms and went ahead to use VADER. I will discuss how I concluded the algorithm below.

Text blob:

It is a basic Python package that provides access to APIs for many NLP activities, including sentiment analysis, spelling correction, and other things.

For a given input, the Text blob sentiment analyzer returns two properties

Polarity is a float with a value between [-1,1], where -1 denotes a negative emotion and +1, a positive emotion.

Additionally, subjectivity is a float with a [0, 1] range. In general, subjective sentences refer to a person's feelings, opinions, or judgment.

Flair:

A straightforward foundation for cutting-edge NLP is Flair. It included a variety of features, including text embeddings, NER, pre-trained sentiment analysis models, and more.

You could use your dataset to train a classifier using flair to create a custom sentiment analyzer for a particular domain. But for the current project, I wouldn't need it.

VADER:

The text sentiment is calculated using a collection of lexical features (such as words) that are classified as positive or negative based on their semantic orientation. The Vader Sentiment function returns the likelihood that a particular input sentence will be Positive, negative, and neutral.

We use the VADER Sentiment Analyzer to perform the sentiment analysis. Formally, VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. A sentiment lexicon is a list of lexical features, in simpler terms words which are generally labeled according to their semantic orientation as either positive or negative. In addition to classifying feelings as positive or negative, VADER also identifies how positive or negative a feeling is.

- The following are some benefits of employing VADER, which facilitates several tasks:
- There is no need for training data.
- It is quite capable of comprehending the tone of a text that uses emoticons, slang, conjunctions, capital letters, punctuation, and many other stylistic devices.
- It functions well on text for social media.
- VADER is compatible with a variety of domains.

After looking into some of the possible options, I went ahead and used Vader as my main NLP algorithm.

Procedure:

Data Wrangling:

For using the Twitter API, I made a twitter developer account which gave me all the tokens to scrape the tweets data. I noted these parameters and went ahead and scrapped the data. I used python confirmation parser to read those data to use the twitter API.

Now, we can't just send the parameters into the method because the Cursor is a passed callable. Instead, the Cursor constructor method receives the parameters. The following parameters are passed to our API to search tweets:

1. q: This is the tweet's searchable keyword. We transmit the party name (Republicans/Democrats) for our project.

2. lang: The language of the tweets we wish to get from the API is indicated here. We collect tweets written in English because it is the official language of the USA and English is widely spoken there.

3. since: This is the starting date for the tweet retrieval. For our purposes, the data from the previous eight months.

A cursor item iterator object is the outcome of the search API request. We repeatedly iterate through this iterator object to retrieve the geo, text, username, and location information. Some information, such as geo- and location-specifics, depend on whether the specific user is sharing this information or not. We will receive blank data for these columns if the user is not providing these details.

Then the obtained list is converted into a data frame

Data Cleaning:

When we look at the tweet messages, we see that there are many characters that are useless for sentiment analysis. Many tweets start with the letter RT, signifying that they have been retweeted. The user's identity from whom the tweet was retweeted is also listed before each tweet, although we are

unconcerned with this information. Additionally, we eliminate any html website links that may have been included in the tweet. With the aid of Python's built-in "re" package, all these data cleaning tasks are completed.

Data Division:

Finding the present political climate in each US state using data from Twitter is our main goal in this investigation. We need to perform additional filtering on the initial data that we downloaded from the tweepy API to accomplish this. As we previously covered, the tweepy API gives us access to the user's location if they are willing to share it with us. Most of this location information is blank. Others may only include the city name, the country name, the city and state separated by a comma, and so forth.

To extract tweets that contain these in the location record, we form a list of US states and another list of US state codes in the following section. Next, I generated a corpus of twitter data on which to run our sentiment analyzer by appending the cleaned data to an existing csv file.

Sentiment Analysis

Then I separated the two-party data into separate data frames.

The dataset is not evenly distributed, like many other datasets. This indicates that some states have more data than others, like Florida, California, and Texas. Additionally, it appears that Republicans have slightly more data available than Democrats do.

I continued with the sentiment analysis after that. To begin with, I used the polarity scores() method to analyze each tweet's text to determine its emotion. The output of this method call is a dictionary that displays the degree to which the tweet was negatively, neutrally, or positively sentiment. The fourth number, which represents the overall compound sentiment of the tweet, is calculated using all three of these variables. This number will be used to determine whether a tweet has a positive, negative, or neutral sentiment.

Figure showing the bar graph democrats overall sentiment by state

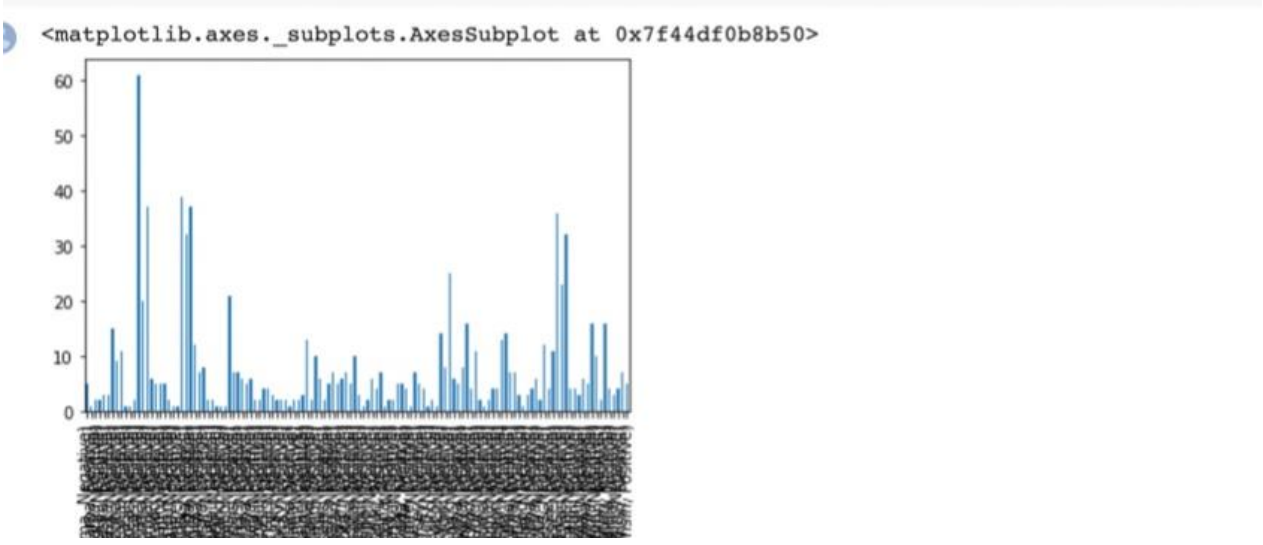


Figure showing the table of democrats sentiment by state

state	sentiment_overall	
Alabama	Negative	5
	Neutral	1
	Positive	2
Alaska	Negative	2
	Neutral	3
	Positive	3
Arizona	Negative	15
	Neutral	9
	Positive	11
Arkansas	Negative	1
	Neutral	1
	Positive	2
California	Negative	61
	Neutral	20
	Positive	37
Colorado	Negative	6
	Neutral	5
	Positive	5
Connecticut	Negative	5
	Neutral	2
Delaware	Negative	1
	Positive	1
Florida	Negative	39
	Neutral	32
	Positive	37
Georgia	Negative	12
	Neutral	7
	Positive	8
Hawaii	Negative	2
	Neutral	2
Idaho	Negative	1
	Neutral	1
	Positive	1
Illinois	Negative	21
	Neutral	7
	Positive	7
Indiana	Negative	6
	Neutral	5
	Positive	6

Figure showing the bar graph Republicans overall sentiment by state

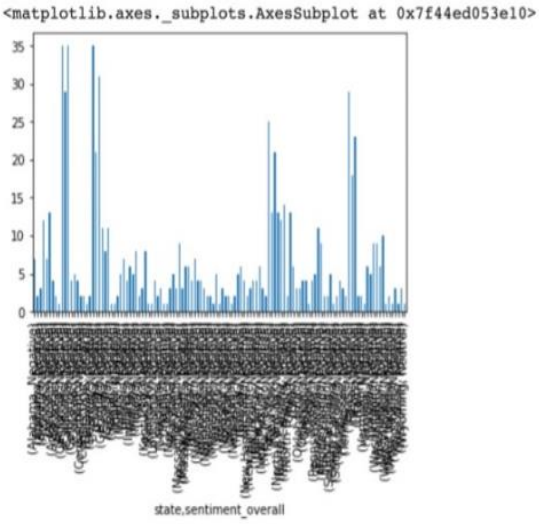


Figure showing the republicans data by location and getting their sentiment

	text	user	location	Subject	state	sentiment	sentiment_overall
785	I'm absolutely thrilled about the Mid Terms.. But I can't get past the fact that MILLIONS of Americans watched the Re…	TalkingAboutTh4	New York, USA	Republicans	New York	{'neg': 0.0, 'neu': 0.901, 'pos': 0.099, 'compound': 0.2724}	Positive
786	Republicans refusing to engage in mail-in voting, ballot harvesting, and early voting operations effectively means we are goi…	my3sonsmonti	MN, USA	Republicans	Minnesota	{'neg': 0.117, 'neu': 0.652, 'pos': 0.23, 'compound': 0.3818}	Positive
787	The sad reality is that Republicans are nowhere near as committed to saving America as the Democrats are to killing Americ…	Rundannyboy123	Tennessee	Republicans	Tennessee	{'neg': 0.341, 'neu': 0.659, 'pos': 0.0, 'compound': -0.8524}	Negative
788	So, the dude a bunch of Senate Republicans want as their "leader" is the same one whose big 'plan' for the country is to…	DebAma818	Texas, USA	Republicans	Texas	{'neg': 0.0, 'neu': 0.944, 'pos': 0.056, 'compound': 0.0772}	Positive
789	BREAKING: Rick Scott just announced he is running to be the Senate minority leader. This is the same guy whose "Rescue A…	susanreininger	Gulf Breeze, FL	Republicans	Florida	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	Neutral
790	BREAKING: Decision Desk officially projects that Republicans will seize the House but with an embarrassingly slim marg…	marymatt718	Barrow/Utqiaġvik, AK	Republicans	Alaska	{'neg': 0.182, 'neu': 0.818, 'pos': 0.0, 'compound': -0.5499}	Negative
791	Matt Gaetz is attacking Kevin McCarthy. Rick Scott is attacking Mitch McConnell.Republicans are eating their own an…	LaurenDownSouth	Georgia, USA	Republicans	Georgia	{'neg': 0.286, 'neu': 0.714, 'pos': 0.0, 'compound': -0.7184}	Negative
792	The sad reality is that Republicans are nowhere near as committed to saving America as the Democrats are to killing Americ…	thedoveness	Florida	Republicans	Florida	{'neg': 0.341, 'neu': 0.659, 'pos': 0.0, 'compound': -0.8524}	Negative
793	Now that it looks like Republicans are almost certainly going to take back the House, I see people declaring democracy d…	TheLadyValk	Neptune Beach, FL	Republicans	Florida	{'neg': 0.0, 'neu': 0.796, 'pos': 0.204, 'compound': 0.5584}	Positive

After categorizing each tweet's data as either positive, negative, or neutral, we can now organize the data by state and determine how the people of that state feel generally about a certain candidate. This is the step we will take next.

Figure showing the Democrats data by location and getting their sentiment

	text	user	location	Subject	state	sentiment	sentiment_overall
0	The Trump announcement is not scaring us Democrats, especially after the Red Wave was as effective as Rudy Giuliani's…	sobertj	Mound, MN	Democrats	Minnesota	{'neg': 0.0, 'neu': 0.755, 'pos': 0.245, 'compound': 0.6711}	Positive
1	It's a win-win for Democrats. Trump thumps DeSantis for the 2024 Republican Presidential nomination and then Presid…	ByronJHoldenJr2	Silver Springs, Florida	Democrats	Florida	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	Neutral
2	11/15: US Tells Ukraine It Doesn't Need to Talk With Russia, Xi to Biden: Tensions Benefit No One I Julian Assange Activis…	SantinoCortez99	Gainesville, FL	Democrats	Florida	{'neg': 0.189, 'neu': 0.695, 'pos': 0.116, 'compound': -0.2263}	Negative
3	Why are Democrats celebrating the election as if they didn't lose the House? Joe Biden is officially a lame duck preside…	JustHanginOn22	Allentown, PA	Democrats	Pennsylvania	{'neg': 0.21, 'neu': 0.649, 'pos': 0.141, 'compound': -0.2023}	Negative
4	This is the picture Meghan McCain posted on Instagram an hour ago.Anyone who doesn't think that the McCain RINO machine…	MrSpargel	Honolulu, HI	Democrats	Hawaii	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	Neutral
...	...	...	...	...	...	...	...
95	BREAKING: Decision Desk officially projects that Republicans will seize the House but with an embarrassingly slim marg…	sbeisler18	Bronx, NY	Democrats	New York	{'neg': 0.182, 'neu': 0.818, 'pos': 0.0, 'compound': -0.5499}	Negative
96	The overwhelming majority of Americans have no idea Democrats just funded their candidates with a massive ponzi scheme…	Jo11635271	Minnesota, USA	Democrats	Minnesota	{'neg': 0.121, 'neu': 0.879, 'pos': 0.0, 'compound': -0.296}	Negative
97	Democratic official in Connecticut sentenced for ballot fraud	WashTimes	Washington, D.C.	Democrats	Washington	{'neg': 0.45, 'neu': 0.55, 'pos': 0.0, 'compound': -0.5994}	Negative
98	Weird how all toss-up seats and many red seats went totally blue this year in a historically unpopular environment for De…	MARIEKHLAIRE	Dal, Texas	Democrats	Texas	{'neg': 0.082, 'neu': 0.918, 'pos': 0.0, 'compound': -0.1779}	Negative
99	Wow Democrats, Gen. Flynn throwing it down... Will you accept his challenge?	1Bobby72	Washington, USA	Democrats	Washington	{'neg': 0.0, 'neu': 0.539, 'pos': 0.461, 'compound': 0.7717}	Positive

Figure showing the table of Republicans sentiment by state

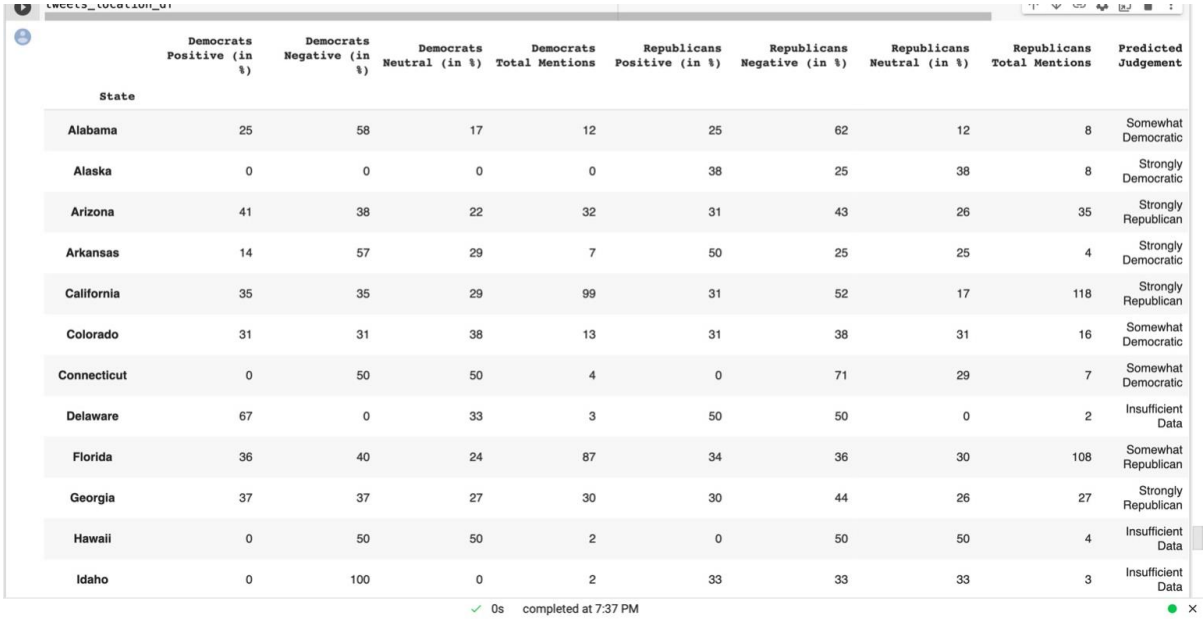
state	sentiment_overall	
Alabama	Negative	7
	Neutral	2
	Positive	3
Arizona	Negative	12
	Neutral	7
	Positive	13
Arkansas	Negative	4
	Neutral	2
	Positive	1
California	Negative	35
	Neutral	29
	Positive	35
Colorado	Negative	4
	Neutral	5
	Positive	4
Connecticut	Negative	2
	Neutral	2
Delaware	Neutral	1
	Positive	2
Florida	Negative	35
	Neutral	21
	Positive	31
Georgia	Negative	11
	Neutral	8
	Positive	11
Hawaii	Negative	1
	Neutral	1
Idaho	Negative	2
Illinois	Negative	5
	Neutral	7
	Positive	4
Indiana	Negative	6
	Neutral	5
	Positive	8
Iowa	Negative	2
	Neutral	3
Kentucky	Negative	8
	Neutral	1
	Positive	1
Louisiana	Negative	4

Finally, to form the conclusion we know that the Republicans have a better positive sentiment in certain states. Final results in the figure below.

```
tweets_location_df.groupby('Predicted Judgement').size()
```

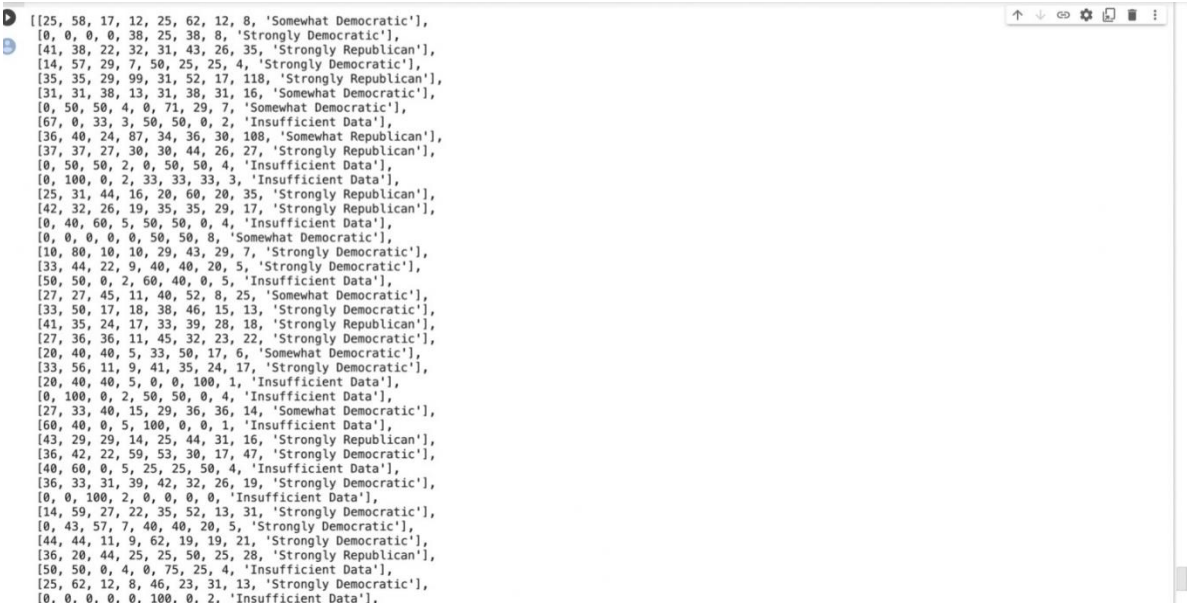
```
Predicted Judgement
Insufficient Data      14
Somewhat Democratic    10
Somewhat Republican     2
Strongly Democratic    16
Strongly Republican     8
dtype: int64
```

Figure showing the sentiments for each party in percentage



State	Democrats Positive (in %)	Democrats Negative (in %)	Democrats Neutral (in %)	Democrats Total Mentions	Republicans Positive (in %)	Republicans Negative (in %)	Republicans Neutral (in %)	Republicans Total Mentions	Predicted Judgement
Alabama	25	58	17	12	25	62	12	8	Somewhat Democratic
Alaska	0	0	0	0	38	25	38	8	Strongly Democratic
Arizona	41	38	22	32	31	43	26	35	Strongly Republican
Arkansas	14	57	29	7	50	25	25	4	Strongly Democratic
California	35	35	29	99	31	52	17	118	Strongly Republican
Colorado	31	31	38	13	31	38	31	16	Somewhat Democratic
Connecticut	0	50	50	4	0	71	29	7	Somewhat Democratic
Delaware	67	0	33	3	50	50	0	2	Insufficient Data
Florida	36	40	24	87	34	36	30	108	Somewhat Republican
Georgia	37	37	27	30	30	44	26	27	Strongly Republican
Hawaii	0	50	50	2	0	50	50	4	Insufficient Data
Idaho	0	100	0	2	33	33	33	3	Insufficient Data

✓ 0s completed at 7:37 PM



```
[[25, 58, 17, 12, 25, 62, 12, 8, 'Somewhat Democratic'],
 [0, 0, 0, 0, 38, 25, 38, 8, 'Strongly Democratic'],
 [41, 38, 22, 32, 31, 43, 26, 35, 'Strongly Republican'],
 [14, 57, 29, 7, 50, 25, 25, 4, 'Strongly Democratic'],
 [35, 35, 29, 99, 31, 52, 17, 118, 'Strongly Republican'],
 [31, 31, 38, 13, 31, 38, 31, 16, 'Somewhat Democratic'],
 [0, 50, 50, 4, 0, 71, 29, 7, 'Somewhat Democratic'],
 [67, 0, 33, 3, 50, 50, 0, 2, 'Insufficient Data'],
 [36, 40, 24, 87, 34, 36, 30, 108, 'Somewhat Republican'],
 [37, 37, 27, 30, 30, 44, 26, 27, 'Strongly Republican'],
 [0, 50, 50, 2, 0, 50, 50, 4, 'Insufficient Data'],
 [0, 100, 0, 2, 33, 33, 33, 3, 'Insufficient Data'],
 [25, 31, 44, 16, 20, 60, 20, 35, 'Strongly Republican'],
 [42, 32, 26, 19, 35, 35, 29, 17, 'Strongly Republican'],
 [0, 40, 60, 5, 50, 50, 0, 4, 'Insufficient Data'],
 [0, 0, 0, 0, 0, 50, 50, 8, 'Somewhat Democratic'],
 [10, 80, 10, 10, 29, 43, 29, 7, 'Strongly Democratic'],
 [33, 44, 22, 9, 40, 40, 20, 5, 'Strongly Democratic'],
 [50, 50, 0, 2, 60, 40, 0, 5, 'Insufficient Data'],
 [27, 27, 45, 11, 40, 52, 8, 25, 'Somewhat Democratic'],
 [33, 50, 17, 18, 38, 46, 15, 13, 'Strongly Democratic'],
 [41, 35, 24, 17, 33, 39, 28, 18, 'Strongly Republican'],
 [27, 36, 36, 11, 45, 32, 23, 22, 'Strongly Democratic'],
 [20, 40, 40, 5, 33, 50, 17, 6, 'Somewhat Democratic'],
 [33, 56, 11, 9, 41, 35, 24, 17, 'Strongly Democratic'],
 [20, 40, 40, 5, 0, 0, 100, 1, 'Insufficient Data'],
 [0, 100, 0, 2, 50, 50, 0, 4, 'Insufficient Data'],
 [27, 33, 40, 15, 29, 36, 36, 14, 'Somewhat Democratic'],
 [60, 40, 0, 5, 100, 0, 0, 1, 'Insufficient Data'],
 [43, 29, 29, 14, 25, 44, 31, 16, 'Strongly Republican'],
 [36, 42, 22, 59, 53, 30, 17, 47, 'Strongly Democratic'],
 [40, 60, 0, 5, 25, 25, 50, 4, 'Insufficient Data'],
 [36, 33, 31, 39, 42, 32, 26, 19, 'Strongly Democratic'],
 [0, 0, 100, 2, 0, 0, 0, 0, 'Insufficient Data'],
 [14, 59, 27, 22, 35, 52, 13, 31, 'Strongly Democratic'],
 [0, 43, 57, 7, 40, 40, 20, 5, 'Strongly Democratic'],
 [44, 44, 11, 9, 62, 19, 19, 21, 'Strongly Democratic'],
 [36, 20, 44, 25, 25, 50, 25, 28, 'Strongly Republican'],
 [50, 50, 0, 4, 0, 75, 25, 4, 'Insufficient Data'],
 [25, 62, 12, 8, 46, 23, 31, 13, 'Strongly Democratic'],
 [0. 0. 0. 0. 0. 100. 0. 2. 'Insufficient Data'].
```