Saira Chawla

Quantifying Abortion Notions - An analysis of tweets

Abstract

It is unquestionable that policy decisions are heavily influenced by the general population's beliefs. Furthermore, the internet allows for greater speed in the sharing of information than ever before. These facts are all the more important in light of the 2022 Dobbs V. Jackson case that overturned Roe V. Wade, which eliminated the right to have a safe, legal abortion by choice in many states (Mane et al.). The sharing of information through digital methods like social media has the ability to shape beliefs and thereby policy decisions, including those on abortion. My thesis aims to examine the information spread on social media platforms, specifically Twitter, regarding abortion in light of the overturning of Roe V. Wade. I want to explore this in order to understand how the content on social media has developed, dispersed, and built narratives around abortion. Furthermore, a key objective of this study is to add on to pre-existing state-level data regarding abortion-related features. I did this by comparing two datasets: one is a compilation of tweets and its features from each state via Twitter API, and another from a study that gathers abortion health, opinion, and access related data by each state from PLoS One. My research question is structured around four hypotheses that provides me with guidance to explore the two final datasets- the aggregated state-level dataset regarding abortion-related features and the other that is a compilation of over 30,000 tweets regarding the key abortion terms and necessary metrics. Through classifying networks, sentiment analysis, and gauging engagement rate, I can quantify the narratives around abortion.

Introduction

The debate around abortion did not begin with Roe. Before the early 19th century, abortion was legally practiced within the United States. At that time, approximately 25% of pregnancies would end in abortion (Mane et al.). However, by 1900, all US states had adopted laws criminalizing abortions. Throughout the years following up to Roe V. Wade's decision, it was estimated that approximately 200,000 to 1,200,000 illegal and unsafe abortions were happening per year, and all sorts of people began to advocate and protest for abortion law reform (Cates 25). It has been over four decades since the Supreme Court's 1973 Roe V. Wade decision, which legalized most safe-practice abortions in the United States. The decision established that: "a person may choose to have an abortion until a fetus becomes [viable, meaning] the ability to live outside the womb, which usually happens between 24 and 28 weeks after conception" ("Roe v. Wade, 410 U.S. 113 (1973)."). This class action was initiated by Roe, a single pregnant woman, to challenge Texas' criminal abortion laws. It was found that such laws—which did not consider a pregnant individual's timeline and other interests—were a violation of the "Due Process Clause of the 14th Amendment, which protects the right to privacy against state action, including a woman's qualified right to terminate her pregnancy" ("Roe v. Wade, 410 U.S. 113 (1973).").

Roe V. Wade was overturned by the Supreme Court in June 2022 by the Dobbs V. Jackson case, essentially revoking women's right to bodily privacy and choice over their own bodies (Mane et al.). Dobbs V. Jackson also overruled Planned Parenthood V. Casey. Because the Dobbs V. Jackson ruling states that Mississippi law banning abortions post 15 weeks is constitutional, there is now precedent to overturn Roe V. Wade (Mane et al.). The ruling essentially removed the constitutional restriction on legislation banning practice of abortion, which allows state laws to do almost whatever they want. Near total bans, six-week bans, and pending bans have now largely spread across mainly Southern and Midwestern states (Mane et al.). Since the ruling, there has been an increase in complications with pregnancy, and thus an increase in maternal mortality (Mane et al.). Once again, there were strong reactions to the ruling. This time, protests, discussions, and debates took many traditional and nontraditional forms. The internet and social media has played a greater role in politics, especially on topics such as abortion, as it has allowed users to easily share thoughts and consume content.

It is unquestionable that policy decisions are heavily influenced by the general population's beliefs. "The seminal work of communication scholar Gerard A. Hauser notes that, by coming together to freely discuss and identify contentious problems, individuals can reach a common judgment, form public opinion, and influence collective action, policy, and decision-making" (Sharma et al.). As social media has grown salient, more and more users express their views on platforms such as Twitter. Whereas traditional methods for gauging public opinions like surveys are costly, tedious, and introduce their own measurement errors, social media has become like a database for people's opinions. Opinion mining has become an alternative to doing surveys, "and promising results show that this could be even more effective than user surveys" (Jain). Twitter, which is a micro blogging website, allows people to share their thoughts using free-format and 140-character texts called tweets (Jain). Tweets at an aggregate level analysis "have been used to model things in many areas such as the stock market, earthquakes, and pandemics" (Cummings et al.).

My research question is: in light of the overturning of Roe V. Wade, what information is being spread on Twitter regarding abortion? The importance of answering this question is to give insight into the narratives and beliefs that circulate and spread, which can help better understand the problem of misinformation. My hypotheses are:

Hypothesis 1:

There will be an increase in the volume of tweets related to abortion after the leaking of the Supreme Court opinion and an even greater increase after the official announcement of the overturning of Roe V. Wade.

Hypothesis 2:

There will be more polarizing sentiment (stronger sentiment scores) across abortion-related tweets after the leaking of the overturning of Roe V. Wade and even more polarizing sentiment after the official announcement, regardless of whether the user is in favor, against, or neutral towards the legalization of abortion.

Hypothesis 3:

States with more tweets in favor of the legalization of abortion will have public survey opinion more in favor of the legalization of abortion. States with more tweets against the legalization of abortion will have public survey opinion more against the legalization of abortion.

Hypothesis 4:

There will be a lower volume of tweets regarding abortion in states with trigger abortion bans.

Furthermore, a key objective of this study is to add on to pre-existing state-level data regarding abortion-related features. My hypotheses provide me with guidance to explore the two final datasets- the aggregated state-level dataset regarding abortion-related features and the other that is a compilation of tweets regarding the key abortion terms.

Literature Review

The Turnaway study is the largest study in the United States that has examined women's experiences with abortion and unwanted pregnancy through longitudinal data. Approximately 1000 women–who either received or were denied an abortion–were recruited through 30+ abortion facilities across the US where they shared their experience through interviews. Before this study, there was little to no data collected on this topic ("The Turnaway Study."). Now, it has been validated through 50 other publications, journals, and annotations. "The main finding of The Turnaway Study is that receiving an abortion does not harm the health and wellbeing of women, while being denied an abortion results in worse financial, health and family outcomes" ("The Turnaway Study."). The study found many of the claims about why pregnant individuals should be denied an abortion to be negated by socioeconomic findings, meaning that those who were denied an abortion had poorer health, less income, etc, while 95% of those receiving an abortion were certain it was the right decision and had better standards of living ("The Turnaway Study."). Even though there is empirical evidence supporting Roe, "the economic implications of abortion and the stigmas that surround abortion are poorly understood" (Guendelman). Thus, I aim to use twitter data to figure out what beliefs and sentiments regarding abortion are circulating.

Twitter is a vast accessible database on various topics and is much more cost and time effective than traditional methods. Twitter has provided a platform/API through which researchers can access and use data at a public access level, elevated access level, and academic access level. Quantifying signals in tweets has been demonstrated through the use of natural language processing, neural networks, and other data science techniques, which is relatively new to research as social media has risen to be salient in only the last 10-20 years. There is still much to learn. However, there is some existing literature that uses social media and twitter to analyze the conversations around abortion.

In light of the overturning of Roe V. Wade, many journals began to examine the public's reaction on twitter. At the University of Iowa, researchers used tweets to analyze the reactions to the overturning of Roe V. Wade by using the Twitter API to create a "dataset [consisting] of 2,412,205 tweets and each tweet was sent on a date between May 1, 2022, and July 1, 2022" (Fan). They then separated the tweets into negative, neutral, and positive reactions. Approximately 43% of the tweets were negative, 56% were neutral, and 0.5% were positive (Fan). The study concludes by stating this is only their preliminary findings and that this is the first study done to analyze the overturning of Roe V. Wade. They further state that more geospatial research should be conducted using twitter data.

At Kenyon college, researchers conducted analysis on the distinct ways twitter users talk about abortion by querying through legal and emotionally-charged terms (Pham). The researchers cleaned the tweets to be simple text via the Pandas package in python. However, by removing emojis and other emotional combinations of indicators in text, the researchers limited themselves in language and emotional analysis. The researchers then looked at the frequency of popular words in response which is just simple text analysis. However, there is so much more research left to be done regarding abortion and conversations on twitter.

One study utilized "Twitter data to examine temporal, geographical, and sentiment patterns" (Mane et al.). It was revealed that there were "low levels of conversations on these topics until the leaked Supreme Court opinion in early May 2022" (Mane et al.). Whereas pro-choice tweets declined, pro-life tweets renewed interest. "Conversations were less prevalent in" states with more abortion trigger laws (Mane et al.). Through network analysis, the authors were able to find that “tweets mentioning woman/women, supreme court, and abortion spread faster and reached more Twitter users than those mentioning Roe Wade and SCOTUS” (Mane et al.).

Many studies chose to implement the Critical Discourse Analysis (CDA), which is a social science theory and multidisciplinary approach that is most generally applied to linguistics. The main objectives of CDA are “to analyze discourse practices that reflect or construct social problems; to investi- gate how ideologies can become frozen in language and find ways to break the ice; to increase awareness of how to apply these objectives to specific cases of injustice, prejudice, and misuse of power” (Měřička). Although applying CDA to online discourse is relatively new, works exist that apply it using tweets.

These studies classify their tweets into three categories: “*For Abortion, Against Abortion, and Neutral to Abortion”,* and then proceed to use classification frameworks (Sharma et al.) (Měřička). Some works chose to utilize the n-gram language model to look at top unigrams and bigrams; however, there are other features that can be explored which may allow the models to perform better. After preprocessing then vectorizing the raw data, many studies chose to put their tweets through a Support Vector Machine/Classifier, which is “a supervised machine learning algorithm that is used for text classification” (Mane et al.). Support vector machines are “one of the widely used supervised machine learning for textual polarity detection;” however, there is also evidence in support of neural networks and random forest classifiers (Ahmad).

Data

To answer the research question, explore my hypotheses, and construct additional data points, I did this by comparing two datasets: one is a compilation of tweets and its metrics from each state via Twitter API, and another from a study that gathers abortion health, opinion, and access related data by each state from a study published on PLoS One.

The first dataset, which I built upon, comes from a journal, ‘Shining the light on abortion: Drivers of online abortion searches across the United States in 2018’ by Sylvia Guendelman, Elena Yon, Elizabeth Pleasants, Alan Hubbard, Ndola Prata (Guendelman). It was published on PLOS One, where all datasets from journals are publicly accessible. This particular dataset was shared via Figma. Figma is a web-based design tool that helps create whatever a user may need. In this case, Figma was used to share a csv file of data. The dataset contains data collected from a Google API to collect search values for key terms (*abortion, abortion pill, birth control*) as well as ratings of health facilities, accessibility, affordability, demographics, and opinions on abortion for each state (Guendelman). The sample comes from all searches related to the key terms via this Google API (this data is publicly accessible), and the remainder of the data comes from a section of facilities review and survey data conducted by previous journals (Guendelman). Furthermore, all this data was collected in 2018 or earlier (Guendelman). This provides a good foundation for looking at abortion on twitter by querying through tweets for those key terms along with additional data from another media API.

For the second dataset, I built my own dataset using the Twitter API. Twitter provides academic research access level to those eligible which provides access to real-time and historical data with additional features and functionality. I filtered the dataset to only contain tweets from 2022 as that is the year in which the overturning of Roe V. Wade was announced, and queried for the key terms ‘abortion’, ‘abortion pill’, and ‘birth control’ by each state. The query returned a json dump in which I extracted the text, publication time, public metrics, and author public metrics of each tweet. Approximately 33,000 tweets were found. Below are five example rows of the final dataset for all the tweets queried with the additional columns Engagement Rate, VaderSentiment, and Predicted Stance (Tbl. 1). These additional columns were computed and will be gone over in the Methodology section.

Table 1.

	State	Phrase	Tweet	Timestamp	Engagement Rate	Vader Sentiment	Predicted Stance
8049	Illinois	abortion	@SenWarren No. Abortion is murder.	2022-07-10	0.000000	-0.6908	AGAINST
15165	New York	abortion	There is absolutely nothing, NOTHING that is justifiably #prolife denying a 1O YEAR OLD an abortion. This child is already traumatized for life by OH BY THE WAY WHERE'S THE RAPIST?	2022-07-03	0.000000	-0.9095	AGAINST
31578	North Carolina	birth+control	@calgal93 @Beverly61751063 This issue is really about power and control over women, especially strong women. They hate having to deal with us, but we will never cease to exercise our rights to body autonomy, birth control, education, working outside the home, and making decisions which are best 4 them!	2022-05-05	0.000011	0.7928	FAVOR
16413	North Carolina	abortion	Abortion rights supporters gather in Atlanta after Supreme Court decision https://t.co/S1FnEsp4KG via @GeorgiaRecorder	2022-06-25	0.000005	0.7579	FAVOR
13893	Nevada	abortion	Damn Economy, Trump, North Korea, China, Afghan, Russia, Ukraine, immigration, Abortion,Virus, Climate Change can't the American people catch a break, This is why we voted Biden into Oval. Calm in our country! American People are very upset with all parties	2022-10-04	0.000000	-0.2228	NONE

I selected each example to reflect some nuances with my classification model. As can be seen in Tbl. 1, not every tweet is classified correctly. For instance, the tweet indexed at 15165 is not actually against the legalization of abortion, and the tweet indexed at 16413 actually has no stance on the legalization of abortion.

Once the collection was done, I used natural language processing to formulate a variable that stands for whether the overall stance of tweets regarding the key phrases from a given state was neutral, for, or against the legalization of abortion, which allowed me to classify twitter users into three networks. Furthermore, other variables that capture average sentiment and engagement rate of abortion-related tweets were constructed to further answer the research question.

Once this data collection was complete, the final dataset appears with the 16 original columns from the first dataset: state, RSV for ‘abortion’, RSV for ‘abortion pill’, RSV for ‘birth control’, overall health systems performance score, cost of care, access to care, health outcomes, number of abortion facilities, number of women per abortion facility, number abortion restrictions, number abortion protections, % opine that abortion should be illegal, % unintended pregnancy, % population is 18-24 years old, and % rural population, and three additional columns derived from our twitter data. These additional columns are Vader Sentiment, Engagement Rate, and Numerical Predicted Stance. The Vader Sentiment column represents the average sentiment of all abortion-related tweets of the given state. The Engagement Rate column represents the average engagement rate of abortion-related tweets from the given state. The Numerical Predicted Stance represents the average stance of all-abortion related tweets of the given state numerically, with 100 being against the legalization of abortion, 0 being in favor of the legalization of abortion, and 50 being none. Below represents the first five rows of the final state-level dataset (Tbl. 2) and each column's respective definitions (Tbl. 3).

Table 2.

	State	RSV for "abortion" (2018)	RSV for "abortion pill" (2018)	RSV for "birth control" (2018)	Overall Health Systems Performance Score	Cost of Care	Access to Care	Health Outcomes	Number of Abortion Facilities	Number of women ages 15-49 per abortion facility	Number abortion restrictions	Number abortion protections	% Opine that abortion should be illegal	% Unintended pregnancy	% Population 18-24 years old	% Rural population	Vader Sentiment	Engagement Rate	Numerical Predicted Stance
0	Alabama	71	80	93	46	44	44	44	5	223458	15	0	58.0	55	8.910	40.6	-0.145169	0.004059	80.600000
1	Alaska	51	52	67	49	51	37	22	6	27969	7	2	34.0	48	9.021	34.0	0.008690	0.000152	66.129032
2	Arizona	55	74	74	36	40	43	27	8	190750	17	2	46.0	51	8.276	10.1	-0.095629	0.012688	81.083845
3	Arkansas	56	60	82	48	37	31	50	3	222577	20	1	60.0	55	8.476	43.4	-0.037166	0.021559	80.630631
4	California	51	77	73	29	17	45	19	152	61740	4	8	38.0	48	9.262	4.9	-0.063165	0.004762	79.898911

Table 3.

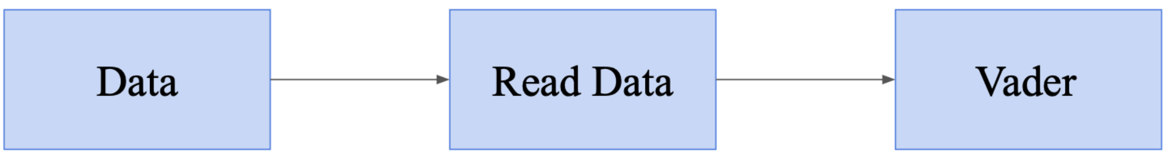
	Column Name	Definition
0	State	
1	RSV For "abortion" (2018)	relative search value for "abortion" in 2018
2	RSV For "abortion pill" (2018)	relative search value for "abortion pill" in 2018
3	RSV For "birth control" (2018)	relative search value for "birth control" in 2018aggregate performance score of health systems
4	Overall Health Systems Performance Score	
5	Cost of Care	
6	Access to Care	
7	Health Outcomes	
8	Number of Abortion Facilities	
9	Number of women ages 15-49 per abortion facility	
10	Number abortion restrictions	
11	Number abortion protections	survey opinion from US 2014 Religious Census Survey
12	% Opine that abortion should be illegal	
13	% Unintended pregnancy	
14	% Population 18-24 years old	
15	% Rural population	
16	Vader Sentiment	average sentiment of tweets from given state
17	Engagement Rate	average engagement rate of tweets for given state
18	Numerical Predicted Stance	average opinion on whether or not abortion should be illegal of tweets for given state

Just like all datasets, there are some aspects of this dataset that are less than ideal. For instance, this original data from PLoS One is limited as it was collected in 2018 or earlier and a more recent data collection would have been ideal to see what changed with the overturning of Roe V. Wade. Another is that we are assuming the sample on Twitter is representative of the entire population, where in actuality, we cannot conclude it is representative as a subset of certain people may use Twitter. Platforms such as Twitter "are biased in several aspects, from their algorithms to the population participating in them". In fact, a study "[showed] that around 25% of Twitter users in the US produce around 97% of all tweets" (Hutchinson). Thus, the conversations happening on Twitter may not be representative of beliefs of the general population, resulting in a measurement error. We must also assume the same for polls as "polls often diverge from reality due to systematic error" (Cummings et al.). However, since social media is by nature *social* we can assume that twitter reflects many of the general population's social patterns (Coppersmith). Furthermore, we continue to make this assumption for google residual search values. Some people may not have access to wifi or platforms such as the google search engine. Thus, we assume that sample search values reflect the same desire for resources and information regarding the topic for the entire population. All these factors affect our variables of interest.

Methodology

Sentiment Analysis

To perform sentiment analysis, I used VADER (Valence Aware Dictionary and sEntiment Reasoner) Sentiment, which "is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media" (Hutto and Gilbert). VADER has the ability to handle vocabularies, abbreviations, capitalizations, repeated punctuations, as well as emoticons. Thus, I performed no preprocessing on the text of each tweet at this step. I used VADER sentiment's compound score which is a summation of its neutral, positive, and negative scores to find the overall sentiment of the given tweet.

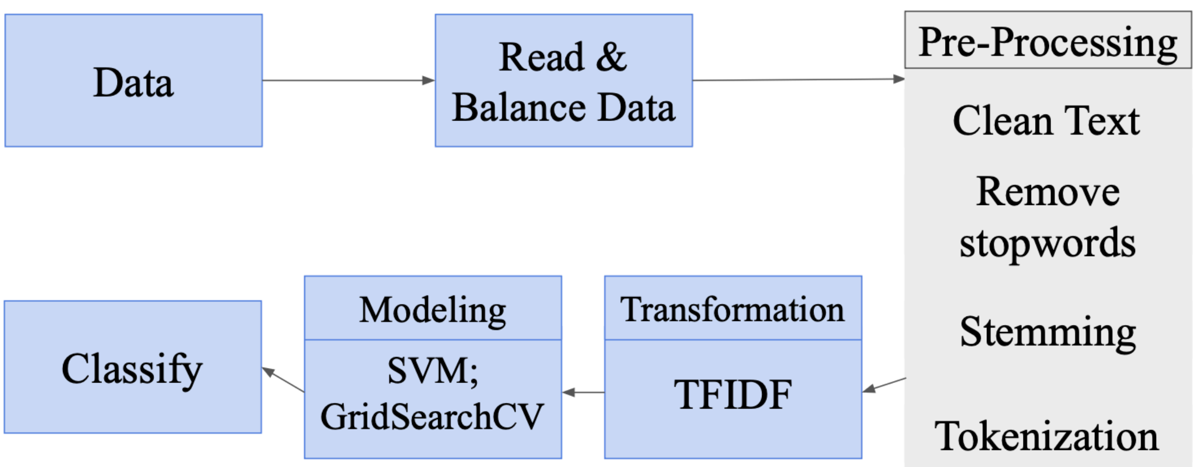


Modeling

The purpose of modeling is to classify each tweet's stance. To perform the actual classification, many data wrangling steps were taken and additional data to build and test our model needed to be found. Task 6 from the SemEval-2016 journal presents a dataset for detecting stance on tweets for 5 different topics, including the legalization of abortion (Mohammad et al.). This is the dataset used to build my model upon. The data wrangling steps performed upon this dataset were designed to help improve my model's metrics. For instance, to each tweet, I applied a clean_text function that removed unnecessary characters; a remove_stopwords function that removed words such as "the", "a", etc; and a stemming function that found each word's root word. Next, the stance of the tweets were unbalanced and the signal for *AGAINST* tweets was greater. Using resampling, I was able to balance the data and signal between all three stance values.

The next step was to prepare the textual data in a way my model could understand. After splitting the data into a training set and test set, I performed TF-IDF vectorization with an ngram range of one to three. Thus, after vectorizing the textual data, it was finally ready to be implemented in the model. It is important to note that I played around with multiple preprocessing, tokenization, and feature selection steps to figure out which model performed best.

For the modeling, I used a support vector classifier, which has the objective "to find hyperplane(s) in an N(the number of features)-dimensional space that distinctly classifies the data points" (Gandhi). To tune this model for its best parameters and performance, I performed GridSearchCV, which "means trying all possible combinations of parameters of interest" (Müller



and Guido). The best parameters were C = 10, gamma = scale, and kernel = linear with a test accuracy of 91.0%, a precision score of 91.1%, a recall score of 91.0%, and a cross validation mean score of 86.5%. The cross-validation score indicates how the model would perform on unseen data.

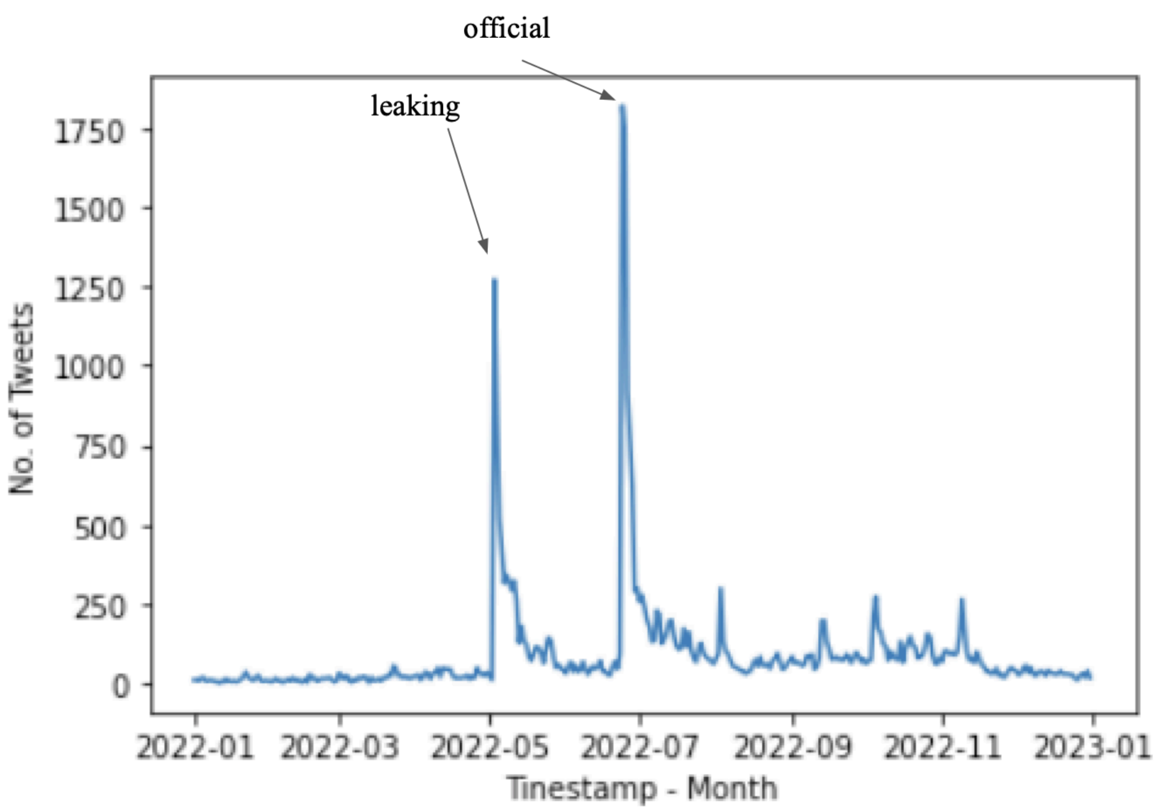
Engagement Rate

The engagement rate is a formula created by Mention that summarizes the interaction and virtual impact of the tweet. The formula "is calculated as the sum of: (Likes + Retweets + Quotes + Replies) divided by the number of tweets, then by the total number of followers, then multiplied by 100" ("Twitter Engagement Calculator (Free Tool).").

Results

Revisiting the first hypothesis, the leaking of the Supreme Court opinion was in May of 2022 and as seen in the figure below (Fig. 1), the volume of tweets related to abortion increased. Furthermore, the official announcement of the overturning of Roe V. Wade was in July of 2022 and Twitter experienced an even greater increase in the volume of tweets related to abortion during that time period. Before the initial leaking, Twitter had near zero volume rates for abortion-related tweets, but after the initial leaking and official announcement spikes, there seemed to be a slightly renewed interest in the topic of abortion on Twitter.

Figure 1.



Classification of Tweets

Of the approximately 33,000 abortion-related tweets in 2022, approximately 77% were against the legalization of abortion, 18% were in favor, and the rest were neutral (Fig. 2). This is a similar finding to previous studies mentioned in the Literature Review section. All three groups of tweets renewed interest after the announcements of the overturning of Roe. However, whereas *NEUTRAL* and *FAVOR* tweets declined to a maximum volume of a few hundred throughout the remainder of the year, *AGAINST* tweets remained a minimum volume of approximately 1,000 or higher (Fig. 3). A key assumption here is that my Support Vector Classifier classified every tweet correctly with 100% accuracy.

Figure 2.

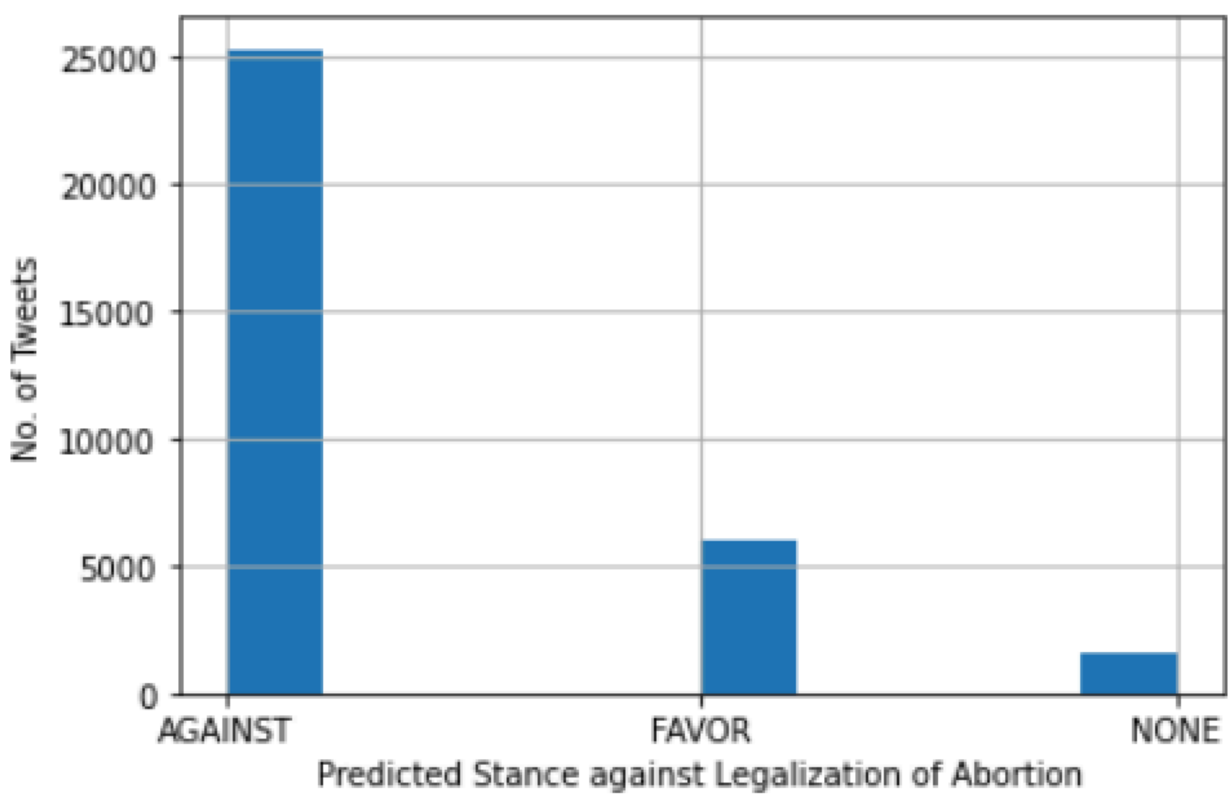
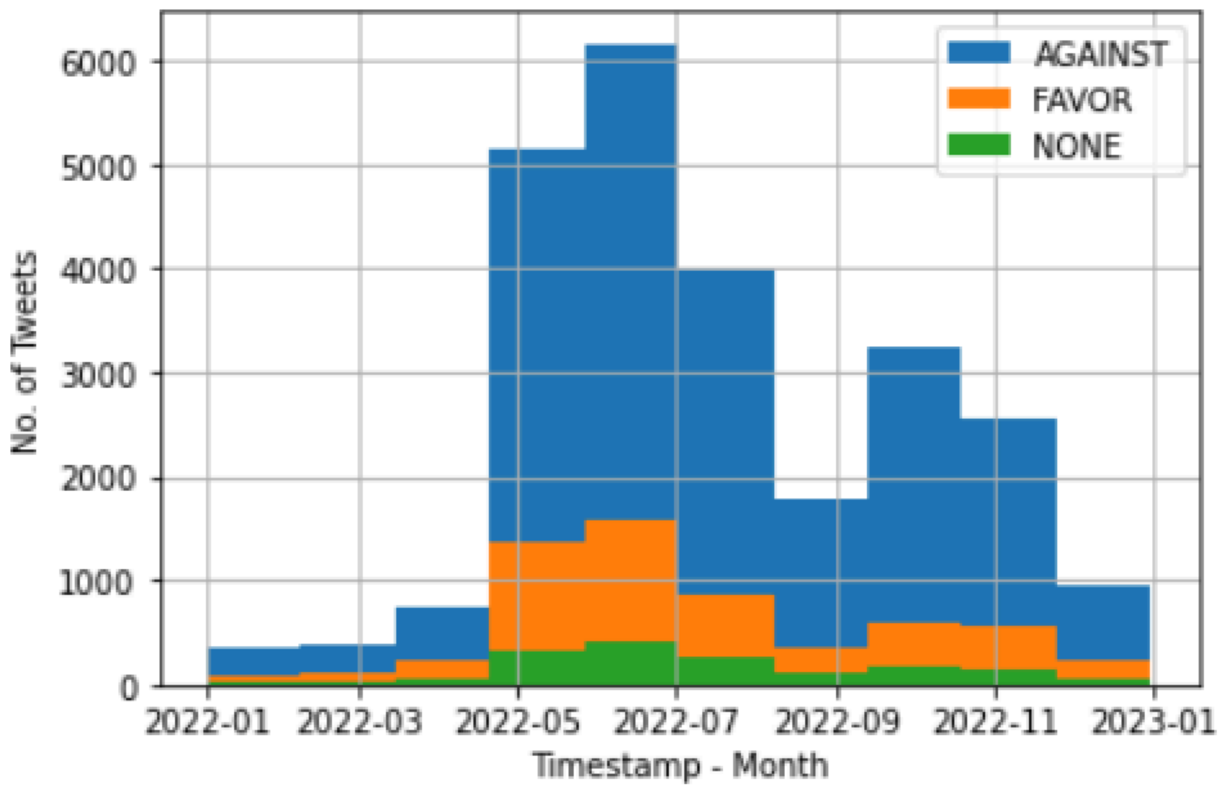


Figure 3.

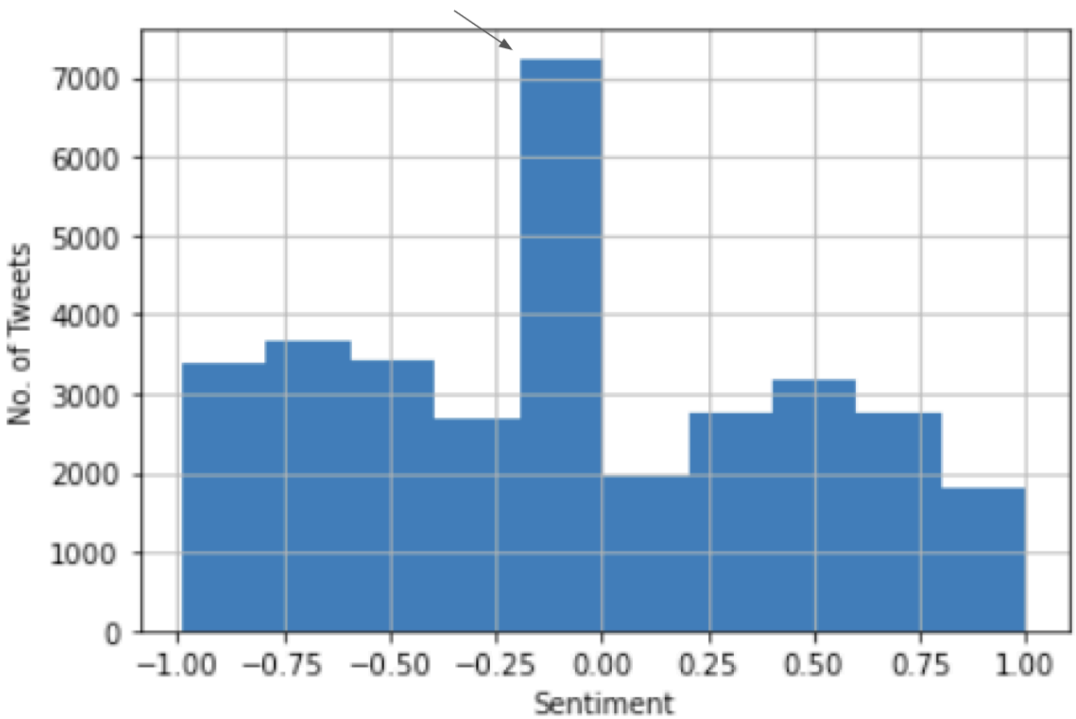


The mean engagement rate for tweets against abortion was 0.0147, for tweets in favor of abortion was 0.0078, and for tweets that were neutral was 0.0013. The overall average engagement rate on Twitter for all tweets is 0.0037 ("Twitter Engagement Calculator (Free Tool)."). The median engagement rate for all three networks was significantly lower than their average, implying that tweets with extremely high engagement rates are skewing the average. Thus, looking at maximum engagement rates for all three networks, *AGAINST* had the highest engagement at a rate of 61.538, followed by *FAVOR* with a maximum engagement rate of 7.692, and finally followed by *NEUTRAL* with a maximum engagement rate of 1.818.

Sentiment

Sentiment across all abortion-related tweets seemed to follow a uniform distribution with a few exceptions. Majority of the tweets had a sentiment between -0.25 and 0, and it seems that there are more tweets with negative sentiment than positive sentiment (Fig. 4).

Figure 4.

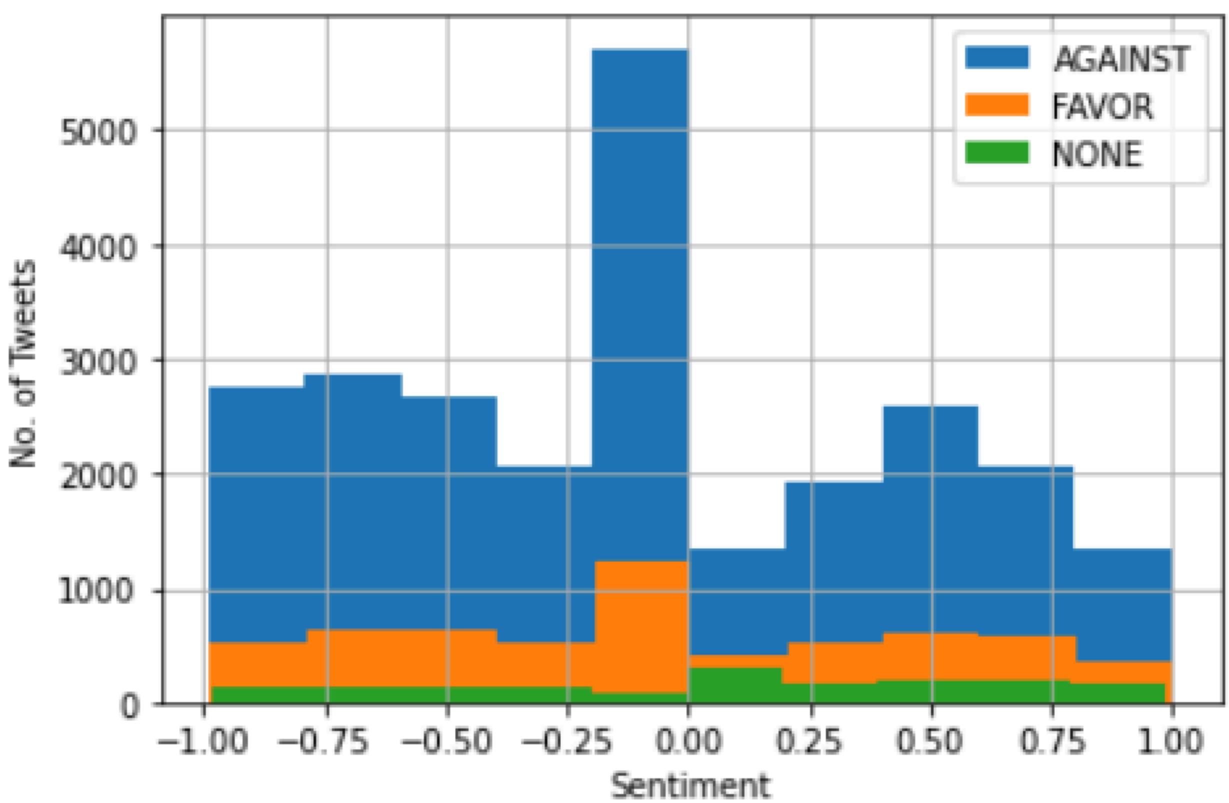


To get a more in depth look, I compared sentiment across the three classified networks: *AGAINST, NEUTRAL,* and *FAVOR.* Majority of *AGAINST* and *FAVOR* tweets had a sentiment between -0.25 and 0, whereas the majority of *NEUTRAL* tweets had a sentiment between 0 and 0.25 (Fig. 5). Furthermore, tweets *AGAINST* the legalization of abortion had the most negative sentiment on average (Tbl. 5).

Table 5.

	Vader Sentiment
Predicted Stance	
AGAINST	-0.072378
FAVOR	-0.024382
NONE	0.068042

Figure 5.



Looking across time, sentiment varies quite a bit, but one noticeable peak is the dip in sentiment in March of 2022 (Fig. 6). However, after calculating the rolling z-scores of 30 days by also calculating the rolling average and rolling standard deviation of 30 days to determine whether or not there was a significant magnitude of change, there were mainly four significant points as can be seen in Fig. 7. The arrows point to significant z-scores which indicates a change in sentiment that is unrelated to randomness. To further check these numbers weren't skewed, I checked to make sure there were more than 100 tweets the given month with the significant change. These months were March, June, and December. June, specifically, had two stark changes in sentiment.

Figure 6.

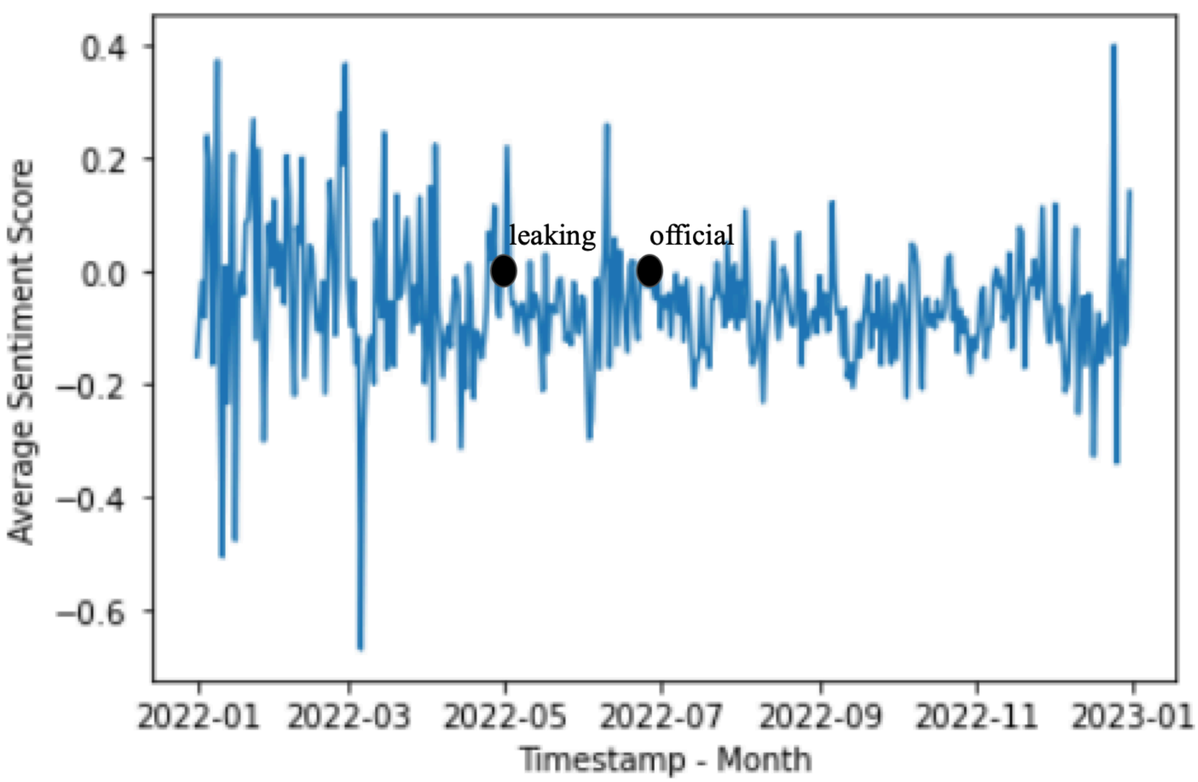
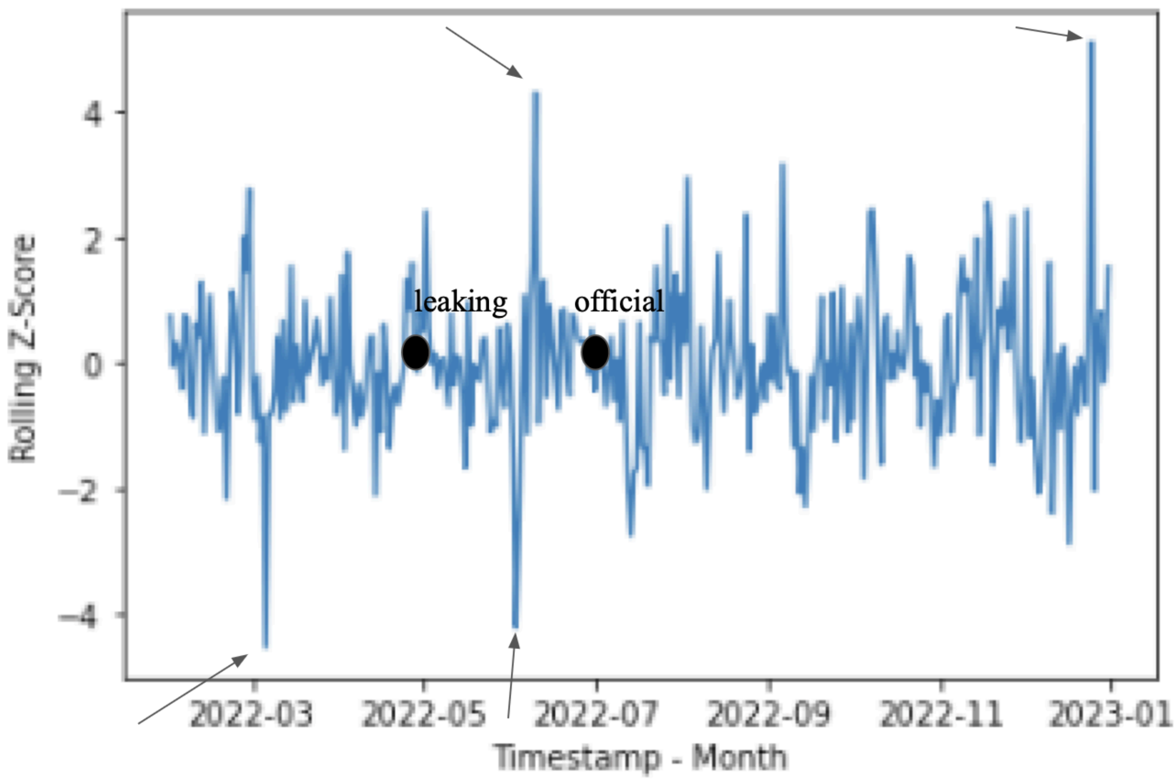


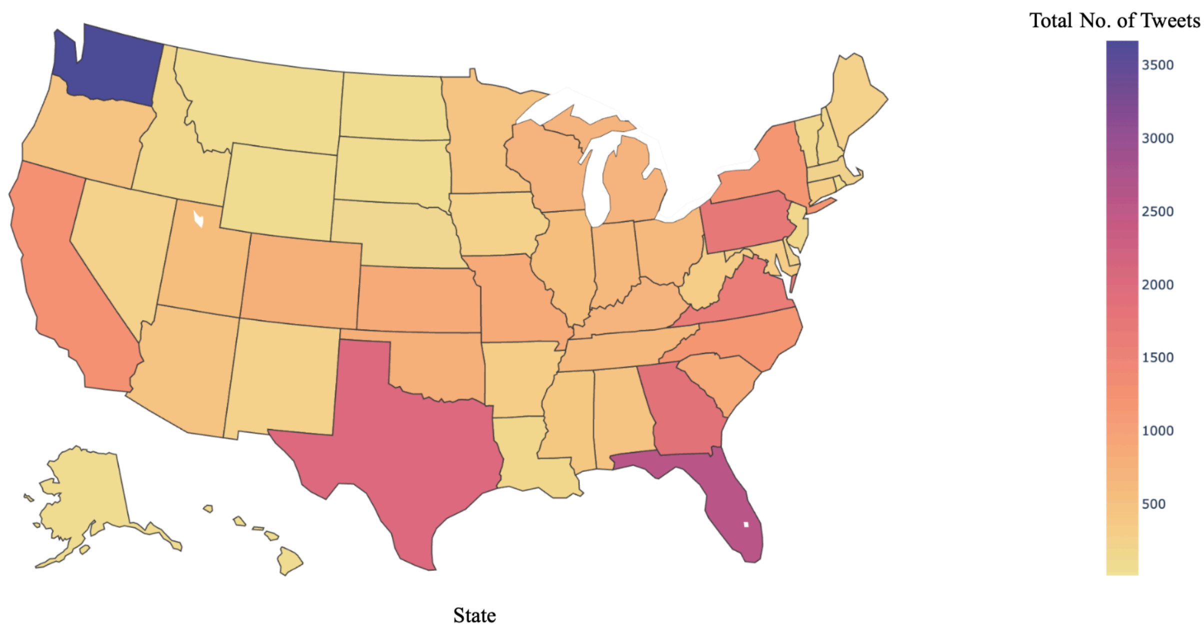
Figure 7.



State-Level

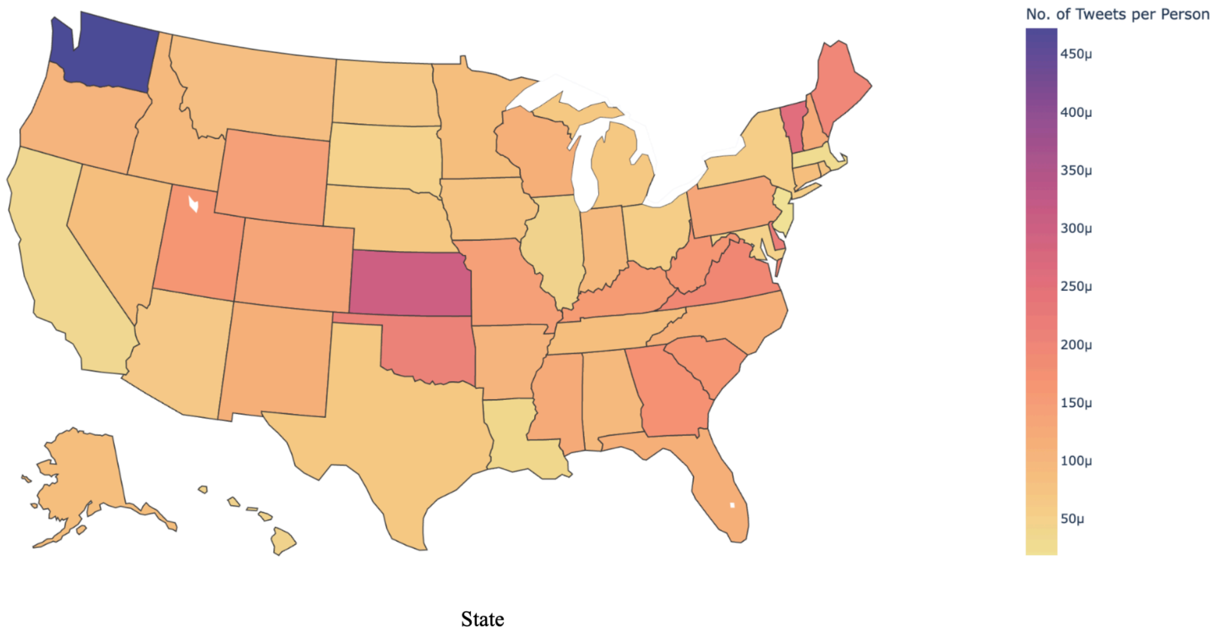
The graph below depicts how many tweets came from each state, not accounting for population size (Fig. 8). Washington has the most number of tweets, 3,667, in the year 2022, followed by Florida at 2586 tweets, and Texas at 2010 tweets.

Figure 8.



The graph below demonstrates the number of tweets from each state, adjusted for population size (Fig. 9). Therefore, the graph depicts the number of tweets per person of each state. Once again, Washington has the most number of tweets per person, but now followed by Kansas and then Vermont.

Figure 9.



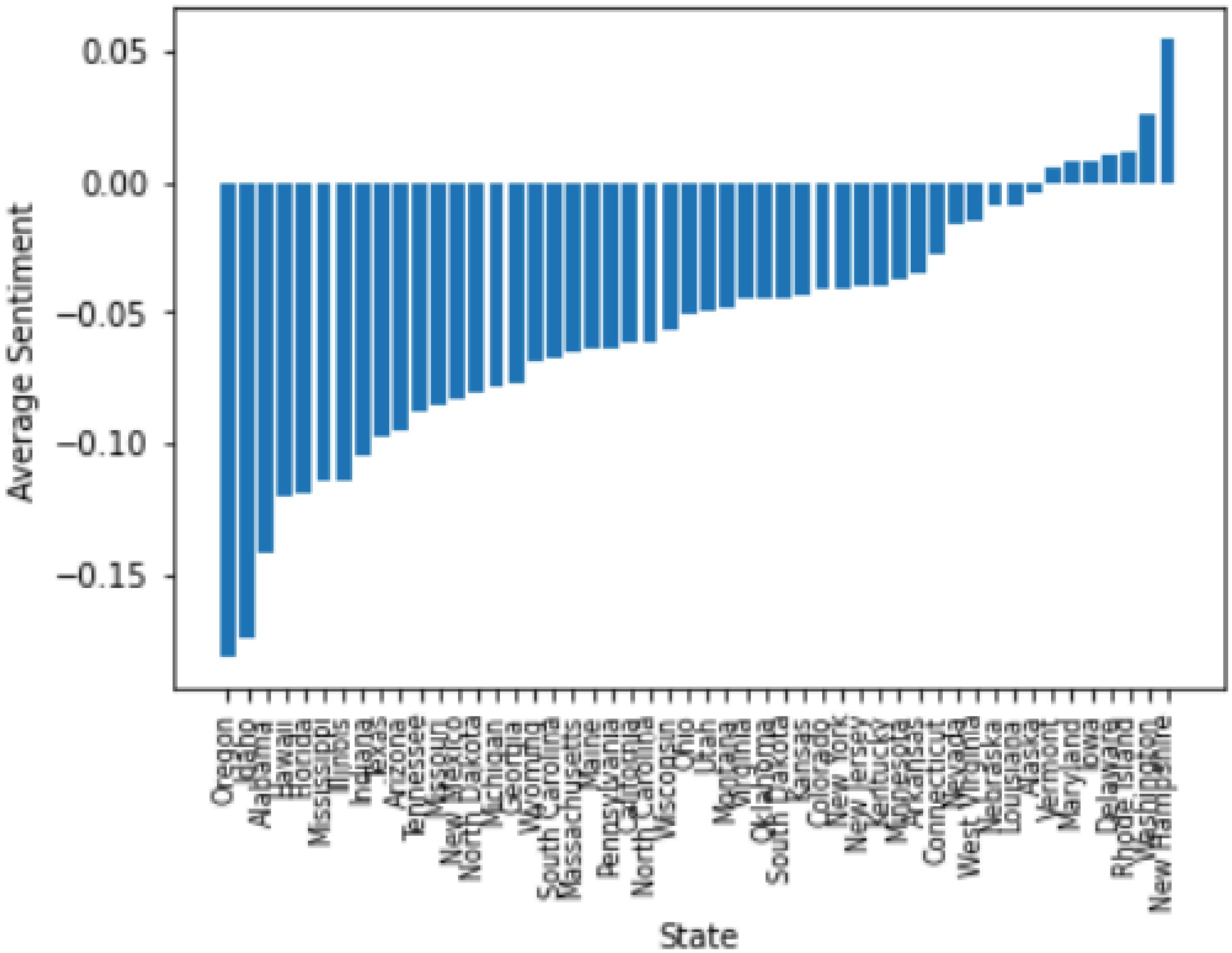
The table below (Tbl. 6) shows descriptive statistics for each column/feature of the final dataset (which includes the data points I compiled from Twitter). For my purposes, I will focus mainly on the columns: % Opine that abortion should be illegal, Vader, Engagement Rate, and Numerical Predicted Stance.

Table 6.

	RSV for "abortion" (2018)	RSV for "abortion pill" (2018)	RSV for "birth control" (2018)	Overall Health Systems Performance Score	Cost of Care	Access to Care	Health Outcomes	Number of Abortion Facilities	Number of women ages 15-49 per abortion facility	Number abortion restrictions	Number abortion protections	% Opine that abortion should be illegal	% Unintended pregnancy	% Population 18-24 years old	% Rural population	Vader Sentiment	Engagement Rate	Numerical Predicted Stance
count	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	5.100000e+01	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000
mean	57.509804	61.705882	76.568627	26.000000	26.000000	26.000000	26.000000	15.294118	1.961390e+05	11.078431	1.882353	42.843137	50.490196	8.376627	25.650980	-0.057539	0.014331	79.221677
std	9.847584	20.058209	8.052962	14.866069	14.866069	14.866069	14.866069	25.804879	2.422532e+05	6.137893	2.084678	9.757812	5.964470	0.613852	14.843751	0.049063	0.033101	4.230645
min	40.000000	0.000000	63.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.390500e+04	0.000000	0.000000	22.000000	36.000000	7.292000	0.000000	-0.182335	0.000152	66.129032
25%	51.000000	52.000000	69.000000	13.500000	13.500000	13.500000	13.500000	3.000000	6.505650e+04	6.000000	0.000000	34.500000	46.000000	7.967500	12.250000	-0.083225	0.001186	77.275748
50%	58.000000	64.000000	76.000000	26.000000	26.000000	26.000000	26.000000	6.000000	1.201350e+05	12.000000	1.000000	45.000000	51.000000	8.276000	25.600000	-0.052587	0.003933	80.153105
75%	62.000000	77.000000	81.000000	38.500000	38.500000	38.500000	38.500000	18.000000	2.230175e+05	16.500000	3.000000	50.000000	55.000000	8.774000	34.550000	-0.035922	0.011279	81.865614
max	100.000000	100.000000	100.000000	51.000000	51.000000	51.000000	51.000000	152.000000	1.365575e+06	23.000000	8.000000	60.000000	62.000000	10.773000	61.600000	0.054297	0.188334	86.274510

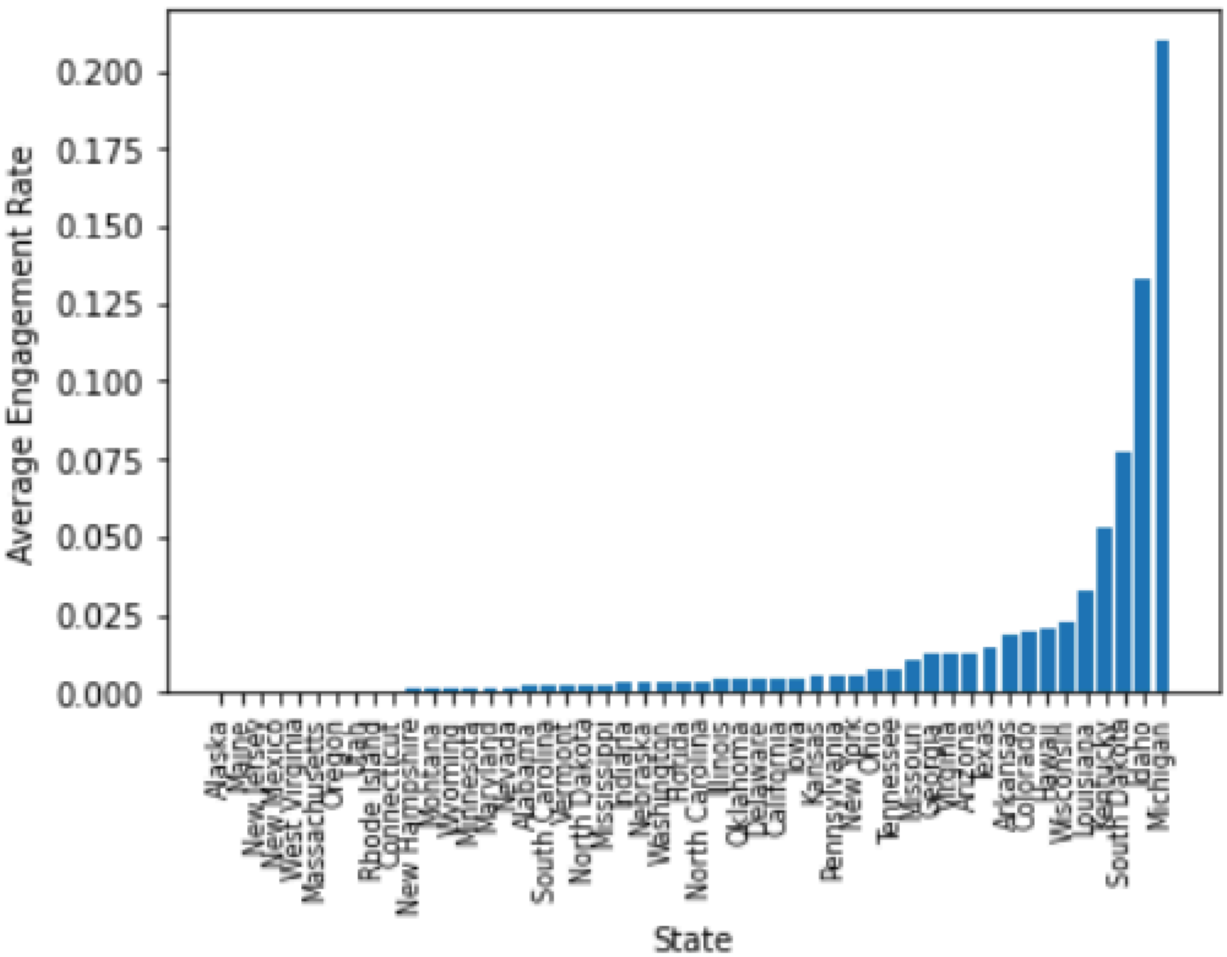
As can be seen in the graph below, most states had a negative sentiment on average (Fig. 10). The three states with the most positive sentiment were New Hampshire, Washington, and Rhode Island. The states with the most negative sentiment were Oregon, Idaho, and Alabama.

Figure 10.



The average engagement rate was relatively low for all states, but the states with high engagement rates were Michigan, Idaho, and South Dakota in that order (Fig. 11).

Figure 11.



States with Abortion Trigger Laws

Below are descriptive statistics for the thirteen states with abortion trigger ban laws- Arkansas, Idaho, Kentucky, Louisiana, Mississippi, Missouri, North Dakota, Oklahoma, South Dakota, Tennessee, Texas, Utah, and Wyoming (Tbl. 7).

Table 7.

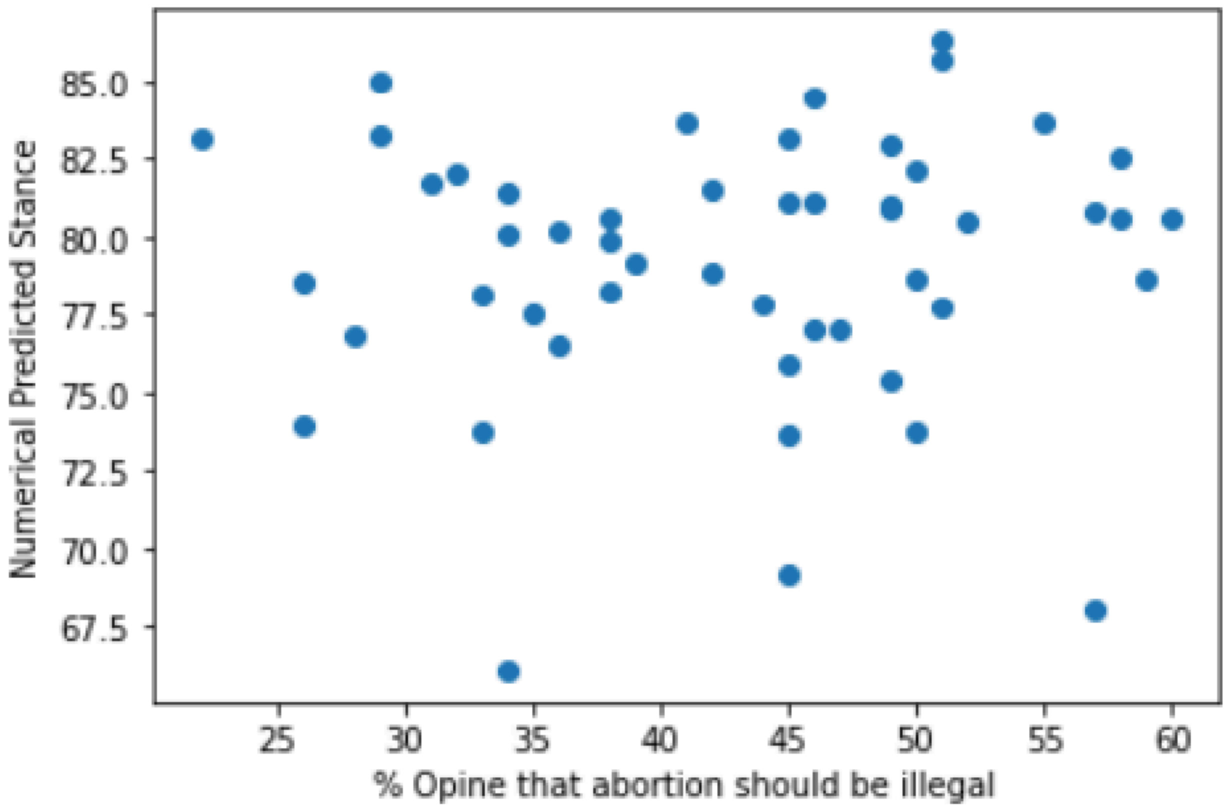
	Engagement Rate	Vader Sentiment	Numerical Predicted Stance
count	5.901000e+03	5908.000000	5908.000000
mean	2.057507e-02	-0.080345	80.966486
std	5.249641e-01	0.521633	38.260778
min	0.000000e+00	-0.984700	0.000000
25%	0.000000e+00	-0.533400	100.000000
50%	5.788989e-07	0.000000	100.000000
75%	2.518167e-05	0.352625	100.000000
max	2.857143e+01	0.981300	100.000000

Revisiting the final hypothesis, I calculated the number of tweets per person in trigger states vs non-trigger states in order to standardize the volume of tweets between these two groupings of states. The number of tweets per person in trigger states is 9.067e-05, and the number of tweets per person in non-trigger states is 9.186e-08.

Survey V.S. Twitter Opinion

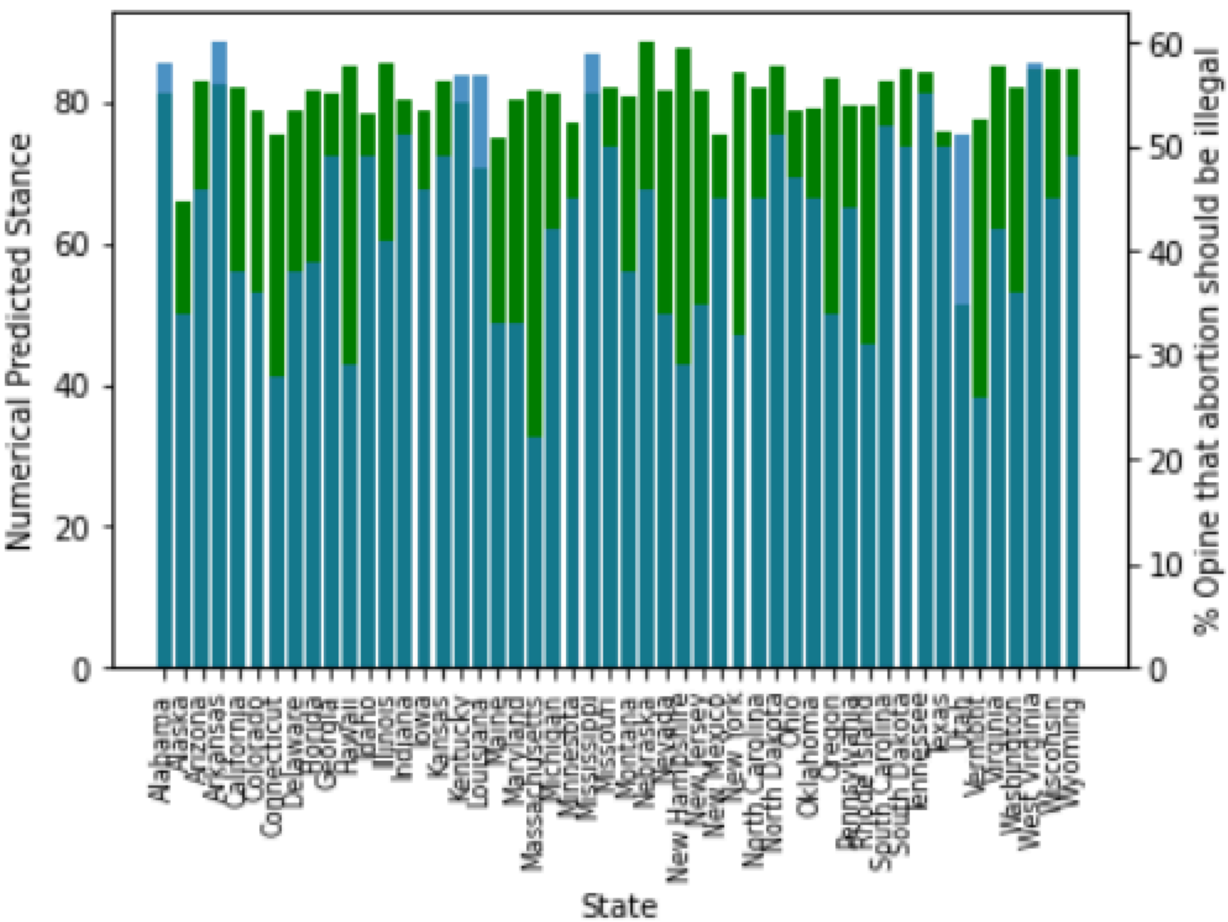
Revisiting the third hypothesis, evidence supporting this would be like a scatterplot with a positive relationship between public survey opinion and average predicted stance of tweets of each state. Thus, to see if there was any form of a relationship between % Opine that abortion should be illegal and Numerical Predicted Stance, I plotted a scatterplot and used pandas correlation function to find the correlation between the two variables. The correlation was 0.0534, representing a slight positive relationship between the two variables (Fig. 12).

Figure 12.



Below is a visualization comparing the Numerical Predicted Stance (green) and % Opine that abortion should be illegal (blue) of each state (Fig. 13). The green gaps represent how much more the overall stance on Twitter is against the legalization of abortion compared to public survey opinion of the given state. To further investigate this, I performed a t-test which is a statistical test used to determine if there is a significant difference between the means of two groups. The t-value here was approximately -25.36 and the p-value was 7.479738574225256e-45. Both values indicated a significant difference between the means of the two groups. The negative sign of the t-value indicated that the mean of one group is lower than the other group. This p-value indicates that there is an extremely low probability of observing such an extreme t-value (or more extreme) if there is no true difference between the two groups. Typically, a p-value less than 0.05 is considered statistically significant, meaning that the result is not likely to be due to chance.

Figure 13.



Discussion

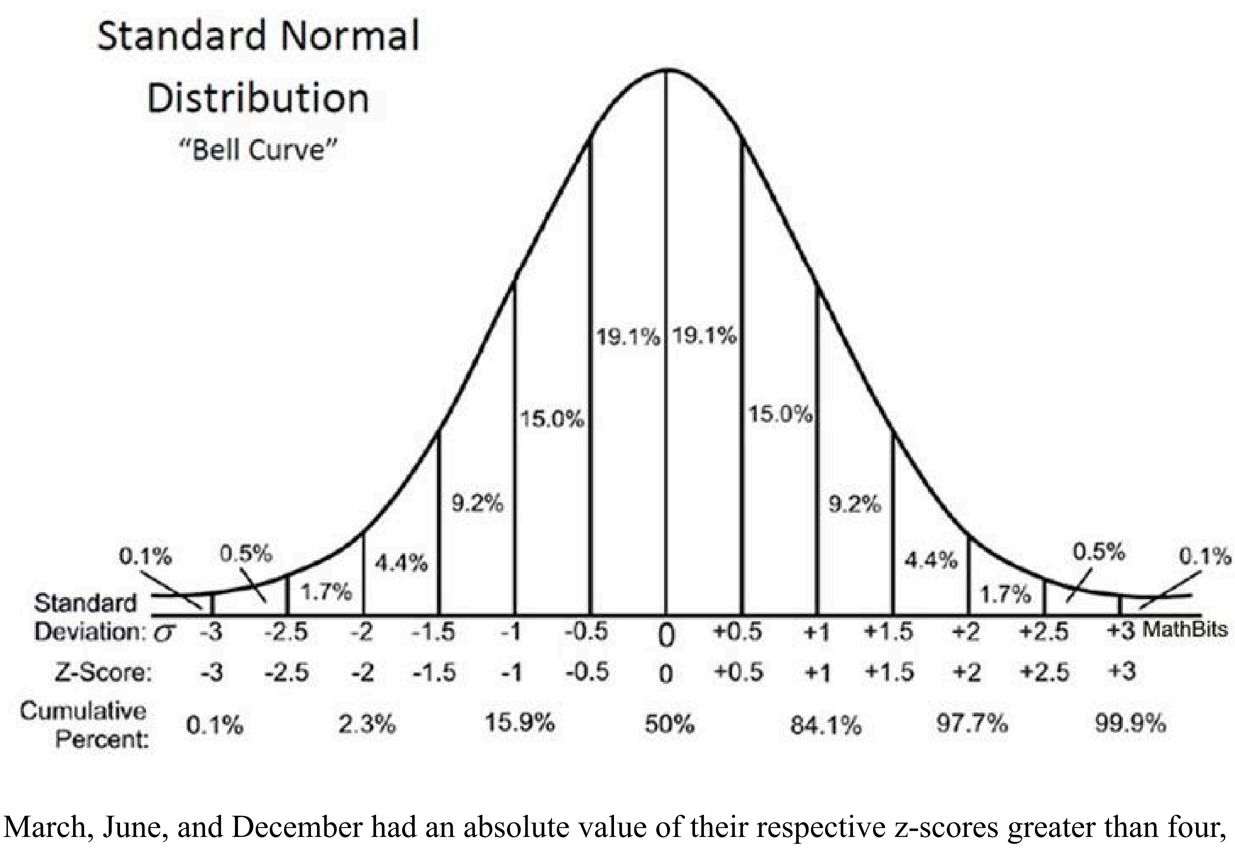
It is important to analyze the results using Critical Discourse Analysis theory in order to answer the research question. Using CDA, I can understand how ideologies and narratives perform in language and how language and narratives combined affect injustice and misuse of power, especially regarding human rights.

Aligning with the first hypothesis, as can be seen in Fig. ?, there was an increase in the volume of tweets related abortion after the leaking of the Supreme Court opinion and an even greater increase after the official announcement of the overturning of Roe V. Wade. Thus, this is consistent with the first hypothesis as well as similar hypotheses from previous studies. Twitter is a platform designed for people to share, react, and respond to news, events, and information such as the overturning of Roe V. Wade. Since the legalization of abortion is known to be such a hot topic of controversy for historical amounts of years, Twitter reflected the popularity of the abortion debate with a significantly increased volume of abortion-related tweets from users. This also demonstrates how communication has changed over time from word-of-mouth to social media because instead of just verbally sharing their thoughts, users are able to put it out into the world via the Internet (Nazzaro). Because the Internet has revolutionized the way humans communicate, platforms such as Twitter have become crucial to starting movements, especially in response to human rights issues. As can be seen in the example tweets and from the number of tweets in each class, most tweets took a stance regarding Roe V. Wade and the legalization of abortion.

From approximately 33,000 abortion-related tweets, over 75% was classified as *AGAINST* the legalization of abortion. Once again, this is under the assumption that the model has classified everything correctly although in actuality, there is some margin of error. This finding is conclusive with previous journals' results. However, to conclude that because the majority of abortion-related tweets in the US were *AGAINST* the legalization of abortion, the general public population is against the legalization of abortion would be a leap. In fact, many reputable institutions such as the Pew Research Center find that the majority of the public disagrees with the overturning of Roe V. Wade (Nadeem). There could be many reasons for this. For instance, one could be that the movement *AGAINST* the legalization of abortion is more vocal, active, and better at mobilizing compared to the movement in *FAVOR* of the legalization of abortion. As seen, engagement rate is on average higher for AGAINST tweets and AGAINST tweets had a higher volume after the official announcement. Furthermore, the history of the creation of alt-right social media platforms demonstrates the mobilization power possessed. In American politics, to be rightwing is to typically be pro-life and thus, *AGAINST* the legalization of abortion. Another reason for why the opinion towards the legalization of abortion on Twitter is majority *AGAINST* is that Twitter is a more right leaning platform. In 2021, Twitter found that their platform's algorithm "statistically significantly [...] favour[s] the political right wing" ("Twitter Admits Bias in Algorithm for Rightwing Politicians and News Outlets."). Another potential reason could be that public survey opinion was surveyed in 2014 and opinions of states have truly changed since. Using this information, it makes sense why *AGAINST* tweets had a higher engagement rate and were greater in amount.

Looking at the sentiment analysis, over half of the tweets had a negative sentiment which can possibly confirm that abortion is a negatively charged topic. When looking at sentiment across class variation, both parties that have a stance (*AGAINST* or *in FAVOR*) spread tweets with negative sentiment on average with those *AGAINST* being slightly more charged. Majority of sentiment was negative because many people were distraught, regardless of their stance against the legalization of abortion. It is important to remember that Twitter is an inherently impulsive platform as the affects of the technology make it very easy and quick to tweet. Thus, people are more likely to share their true opinions on Twitter rather than Instagram that allows several steps to review a post.

Revisiting the second hypothesis, it was essential to perform a time-series analysis. The months that had stark changes in magnitude of sentiment were March, June, and December, noting that June had two significant changes. Referring to Fig. 7, the y-axis labels the z-score of the sentiment, and the example below of a standard normal curve represents how likely that z-score is to occur.



March, June, and December had an absolute value of their respective z-scores greater than four, implying that this change in sentiment is unlikely due to a random instance. For March, I was unable to find any news articles related to the US or US States. Thus, a literal event may not have ignited a stark change in sentiment, but something else may have. Perhaps, a controversial tweet or reaction to other worldly news related to abortion could have caused the change. Further investigation into the month of March on Twitter is required. June was when the official announcement of the Dobbs V. Jackson case came out, and June had two significant z-scores. This insinuates that there were a lot of strong and polarizing feelings regarding the overturning of Roe V. Wade. This partially confirms my second hypothesis. It confirms that there were stronger and more polarizing sentiment scores after the official announcement, but fails to confirm there were stronger and more polarizing sentiment scores after the leaking in May. This could be due to the fact that a rumor does not have as strong of an influence as the official announcement. This can also be seen in the fact that we have higher volumes of tweets after the official announcement than the volume of tweets after the leak. In December, there was one significant stark change in sentiment. December is the time for end of the year reflection. A few tweets, articles, and journals came out post the November 2022 election reflecting the year, especially in regards to abortion. These articles could have reignited the debate on twitter, reflecting more polarizing sentiment scores due to heightened emotion.

A key goal of this journal is to build upon and provide more abortion-related data. In doing this, I was able to depict state-level trends. During an initial analysis, Washington stood out in its volume of abortion-related tweets. Abortion has been a legal right in Washington since 1970 and this did not change with the overruling of Roe. In June of 2022, Washington Department of Health put out a statement on reaffirming access to abortion, restating rights to abortion, and movements to lead the fight in favor of the legalization of abortion. Furthermore, organizations like Legal Voice, which fights for gender autonomy for all, have made residence in Washington fighting for greater reproductive health and healthcare access (“Legislative

Updates."). One of which includes Washington to give its best support to those seeking an abortion from abortion-banned states. Furthermore, Washington is #8 in the least amount of abortion restrictions and #4 in the most amount of abortion protections. Washington taking an active stance on the legalization of abortion could have incited many reactions from all sides of the debate which would be reflected on Twitter.

The next state that stood out in overall high volume of tweets was Kansas. Following the overturning of Roe, in August of 2022, a constitutional amendment in Kansas was voted on that would determine whether or not "Kansas Constitution does not guarantee a right to abortion, giving the Kansas state government power to prosecute individuals involved in abortions, and further declared that the Kansas government is not required to fund abortions" ("2022 Kansas Abortion Referendum."). It was the "first abortion-related constitutional amendment on the ballot since the Supreme Court of the United States overturned the landmark Roe v. Wade decision in Dobbs v. Jackson Women's Health Organization" ("2022 Kansas Abortion Referendum."). This constitutional amendment would overturn a 2019 Kansas ruling that stated abortion was a right under the state constitution. With a percentage of approximately 59% and the highest turn out of Kansas women voters ever seen, Kansas voters voted *NO* against the constitutional amendment. This was a historic moment which was spread and talked about via platforms such as Twitter. Thus, the volume of tweets from Kansas can be confirmed.

To get a more in depth look at how the two main stances towards the legalization of abortion (*AGAINST* or *in FAVOR*) use language to build narratives, I am going to examine the language used in two states from each stance (New Hampshire, Rhode Island, Alabama, Idaho). Of these states, Idaho is a state with trigger ban laws. I am pulling from the states that had the most polarizing sentiment scores and/or highest engagement rates. Furthermore, I narrowed down the selection of states by using public survey opinion to select two states with low opinion that abortion should be illegal (New Hampshire and Rhode Island) and two states with high opinion that abortion should be illegal (Alabama and Idaho). The word frequencies of the other remaining significant states can be found in the Appendix.

Below are the top 10 common words in abortion-related tweets for Alabama (Tbl. 8), Idaho (Tbl. 9), Rhode Island (Tbl. 10), and New Hampshire (Tbl. 11).

Table 8.

	0	1
0	abort	502
1	http	123
2	co	123
3	birth	107
4	control	106
5	women	80
6	state	68
7	get	62
8	life	60
9	use	59
10	right	58

Table 9.

	0	1
0	abort	196
1	http	52
2	co	52
3	control	50
4	birth	50
5	state	46
6	right	41
7	women	27
8	ban	25
9	idaho	22
10	want	22

Table 10.

	0	1
0	abort	89
1	co	43
2	http	41
3	right	15
4	go	13
5	law	12
6	state	12
7	women	11
8	want	10
9	like	10
10	control	9

Table 11.

	0	1
0	abort	191
1	http	91
2	co	91
3	right	58
4	state	41
5	ban	28
6	babi	28
7	time	25
8	nhpolit	25
9	get	23
10	want	21

The most common root word in abortion-related tweets across all selected states was ‘abort’. Then, the next two root words for these states were ‘http’ and ‘co’. These indicate that a lot of tweets referred to linked content. It was important to showcase this as our model is unable to take into account what the linked content is and how the content may indicate the user's stance. In the states that are traditionally against the legalization of abortion (Alabama and Idaho), the conversations were centered around birth, control, women, life, and rights. In the states that are traditionally for the legalization of abortion (New Hampshire and Rhode Island), the conversations were centered around rights, laws, state, women, and control. It seems as if states against the legalization of abortion talk more about the birth and pregnancy of life whereas states in favor of the legalization of abortion talk more about human rights and control of women.

Revisiting the fourth hypothesis, it was essential to standardize the volume of tweets between trigger states and non trigger states because population sizes vary across different states. On the contrary to my hypothesis and previous research, the volume of tweets per person in trigger states was significantly greater than the volume of tweets in non trigger states. Thus, more information regarding abortion in light of the overturning of Roe was circulating on Twitter in abortion trigger states. This could reflect a shift in education on public policy, especially regarding human rights, in states with trigger bans, meaning that the population in these states are talking more and becoming more aware. This discrepancy can be further explored.

Revisiting the third hypothesis, the t-test performed indicated a highly significant difference between the public survey opinion on the legalization of abortion and the twitter opinion on the legalization of abortion and the scatter plot demonstrated no significant correlation between the two. This suggests I reject the third hypothesis that twitter opinion would align with public survey opinion of the given state. The discrepancy between public survey opinion and twitter opinion could be due to many things. As stated earlier, Twitter's algorithm is more right leaning which could be why twitter opinion is more right leaning as well. It could also be that the movement *AGAINST* the legalization of abortion is stronger as cited earlier. The discrepancy is unlikely due to chance and needs to be further explored.

Conclusion

The information being spread on Twitter regarding abortion is vast, mainly negative, not reflective of real public survey opinion (although this is just one methodological approach), and more prevalent in states with trigger abortion bans. Just as the debates around abortion are intense in real life, they are intense on Twitter, and tweets with high engagement rate ignite a lot of pushback from the other side as well. Although Twitter reflected key movements regarding abortion across important states like Kansas and Washington, discrepancies between opinion on the legalization of abortion between Twitter and survey opinion need to be explored further.

The increased volume of tweets after the leaking and then official announcement of Dobbs V. Jackson case overturning Roe demonstrates the intensity people have around the debate. It sparked conversations from all sides of the debate. Furthermore, the polarizing sentiment around abortion, especially after the announcement, demonstrates the emotional charge behind this topic and the debate around it.

The discrepancy seen between Twitter opinion and public survey opinion could be due to many things throughout the study. First, in reality, the model is performing not as well as the metrics say on new and unseen data. Hashtags, slang, and formatting have all changed since the training data was collected. It may have been more useful to classify a sample of tweets collected in 2022 by hand on what their stance towards the legalization of abortion is for the training data. Next, the choice of modeling impacts the boundaries found. In this study, the best model found was a support vector machine, but it would have been interesting to see what relationships could be found from other models such as neural networks. It could also have been possible to use three well-performing supervised learning models and use the majority votes of those three models. These are just examples on how modeling choice affects classification. Finally, it could be true that Twitter has a much stronger stance against the legalization of abortion, but it should be further explored why that is.

Each state had its own internal conversation around the overturning of Roe V. Wade, and I was able to showcase some of the narratives these states were developing. The states showcased were Washington, Kansas, Alabama, Idaho, Rhode Island, and New Hampshire. States with trigger abortion bans are becoming more invested and vocal about how policy decisions affect them. Each state can have a study dedicated to itself on its narrative around abortion.

To further build upon this study, it would be interesting to see how different modeling choices perform on the unseen data that I pulled from the Twitter API. Furthermore, content from URLs should be taken into account as many tweets make a stance regarding that linked content. I also believe some social network analysis to look at how one tweet from one user can spread to a community can give insight into how the spread of information works.

While the debate around abortion may seem like a healthcare issue, it has become centered around political, ideological, economical, and religious issues as well. This study utilized the Twitter API to examine the public's reaction to overturning of Roe V. Wade and what narratives were spread around abortion in that time. It is important to explore reactions and narratives as it can influence others' beliefs and ideas, and thus their vote and policy decisions. This study provides a foundation that can be utilized to look into the spread of misinformation regarding abortion on social media platforms. Leveraging social media has become more prevalent in research, and it is vigilant that further research is conducted to investigate the importance of social media and how it impacts the general public, especially regarding public policies.

Bibliography

"Abortion." *Washington State Department of Health*, doh.wa.gov/you-and-your-family/sexual-and-reproductive-health/abortion. Accessed 9 May 2023.

Cates, Willard, et al. "The Public Health Impact of Legal Abortion: 30 Years Later." *Guttmacher Institute*, Guttmacher Institute, 25 Aug. 2022, https://www.guttmacher.org/journals/psrh/2003/01/public-health-impact-legal-abortion-30-years-later.

Coppersmith, Glen, et al. "Quantifying Mental Health Signals in Twitter." *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2014, https://doi.org/10.3115/v1/w14-3207.

Cummings, David, et al. *Who Needs Polls? Gauging Public Opinion from Twitter Data*, https://nlp.stanford.edu/courses/cs224n/2011/reports/nwang6-davidjc-harukioh.pdf.

Fan, Jonathan, et al. "Analyzing Online User Reactions to the Overturning Decision of Roe vs. Wade." *Geo-Social*.

Gandhi, Rohith. "Support Vector Machine - Introduction to Machine Learning Algorithms." *Medium*, Towards Data Science, 5 July 2018, https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.

Guendelman, Sylvia, et al. “Shining the Light on Abortion: Drivers of Online Abortion Searches across the United States in 2018.” *PLOS ONE*, vol. 15, no. 5, 2020, https://doi.org/10.1371/journal.pone.0231672.

Hutchinson, Andrew. “New Study Shows That 25% of Twitter Users Produce 97% of All Tweets.” *Social Media Today*, 16 Nov. 2021, https://www.socialmediatoday.com/news/new-study-shows-that-25-of-twitter-users-produce-97-of-all-tweets/610146/.

Hutto, C., and Eric Gilbert. “Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.” *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, 2014, pp. 216–225., https://doi.org/10.1609/icwsm.v8i1.14550.

“Legislative Updates.” *Legal Voice*, 15 Feb. 2023, legalvoice.org/leg-updates/.

Mane, Heran, et al. “Examination of the Public’s Reaction on Twitter to the over-Turning of Roe v Wade and Abortion Bans.” *Healthcare*, vol. 10, no. 12, 2022, p. 2390., https://doi.org/10.3390/healthcare10122390.

Měřička, Bc. Vít. *Critical Discourse Analysis of the Abortion Debate in the USA*, 2AD.

Mohammad, Saif, et al. “Semeval-2016 Task 6: Detecting Stance in Tweets.” *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, https://doi.org/10.18653/v1/s16-1003.

Müller, Andreas C., and Sarah Guido. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Inc., 2018.

Nadeem, Reem. "Majority of Public Disapproves of Supreme Court's Decision to Overturn Roe v. Wade." *Pew Research Center - U.S. Politics & Policy*, 21 Feb. 2023, www.pewresearch.org/politics/2022/07/06/majority-of-public-disapproves-of-supreme-courts-decision-to-overturn-roe-v-wade/#:~:text=About%20six%2Din%2Dten%20Americans,majority%20(57%25)%20of%20men.

Nazzaro, M., & Blackshaw, P. (2006). Consumer-generated media (CGM) 101: Word-of-month in the age of the web-fortified consumer. A Nielsen BuzzMetrics white paper (2nd ed.). Spring.

Pham, Tiana. "Abortion Talks: The Distinct Ways Twitter Discusses Abortion in 2021." *Kenyon University*.

"Roe v. Wade, 410 U.S. 113 (1973)." *Justia Law*, https://supreme.justia.com/cases/federal/us/410/113/.

Sharma, Eva, et al. "Analyzing Ideological Discourse on Social Media." *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas*, 2017, https://doi.org/10.1145/3145574.3145577.

"The Turnaway Study." *ANSIRH*, ANSIRH, 15 Aug. 2022, https://www.ansirh.org/research/ongoing/turnaway-study.

"Twitter Admits Bias in Algorithm for Rightwing Politicians and News Outlets." *The Guardian*, 22 Oct. 2021, www.theguardian.com/technology/2021/oct/22/twitter-admits-bias-in-algorithm-for-rightwing-politicians-and-news-outlets.

“Twitter Engagement Calculator (Free Tool).” *Mention.com*, https://mention.com/en/twitter-engagement-calculator/#:~:text=The%20median%20engagement%20rate%20on,is%20considered%20a%20good%20result.&text=0.098%25%20or%20higher%20%2D%20Awesome.

“2022 Kansas Abortion Referendum.” *Wikipedia*, 26 Apr. 2023, en.wikipedia.org/wiki/2022_Kansas_abortion_referendum.

Appendix

Table 12. *Top 10 Words and Frequencies for the state of Washington*

	0	1
0	abort	3835
1	co	2058
2	http	2053
3	right	994
4	state	509
5	women	447
6	get	403
7	amp	382
8	control	357
9	ban	347
10	birth	342

Table 13. *Top 10 Words and Frequencies for the state of Oregon*

	0	1
0	abort	489
1	co	162
2	http	161
3	right	81
4	birth	76
5	control	69
6	want	64
7	state	61
8	women	57
9	babi	54
10	amp	51

Table 14. *Top 10 Words and Frequencies for the state of Michigan*

	0	1
0	abort	693
1	co	280
2	http	276
3	right	152
4	state	112
5	birth	106
6	control	99
7	women	87
8	want	84
9	peopl	81
10	get	79

Table 15. *Top 10 Words and Frequencies for the state of South Dakota*

	0	1
0	abort	51
1	http	16
2	co	16
3	right	13
4	law	8
5	state	7
6	good	7
7	like	5
8	one	5
9	rape	5
10	statist	5