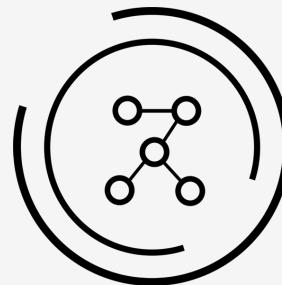


# Petition Data Analysis

DSS x Change.org



C.

# Data Science Society at UC Berkeley



# Project Managers



Nikki Trueblood



Natraj Vairavan

# Data Consultants



Kristen Vitolo



Preetha Kumar



Vikash Giritharan



Sooyeon Kim



Pujitha Nachuri

# Agenda

01

**Background**

02

**Petition Analysis**

03

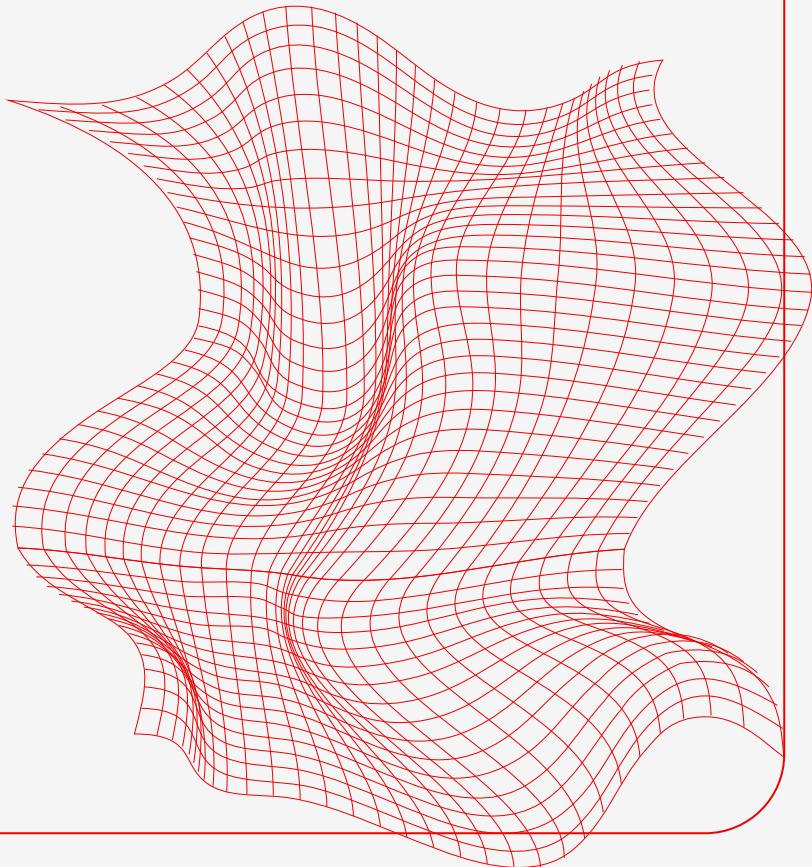
**Harmful Data**

04

**Challenges +  
Future Work**

01

# Background



# Our Tasks



## Analyze Petitions

Find factors that contribute to petition growth/signatures



## Predict Success

Use features to build a model that can predict a petition's # of signatures



## Examine Harmful Data

Investigate patterns in harmful/spam petitions to improve content regulation

# Key Terms

## Correlation

Strength of the relationship between two variables, between -1 and 1

## Modeling

Using data on existing petitions to predict how many signatures future petitions will get

## Overdispersion

When data is very spread out so the variance is greater than the mean

## NLP

Natural Language Processing → using code to analyze text

# Technical Limitations

## Our Barriers

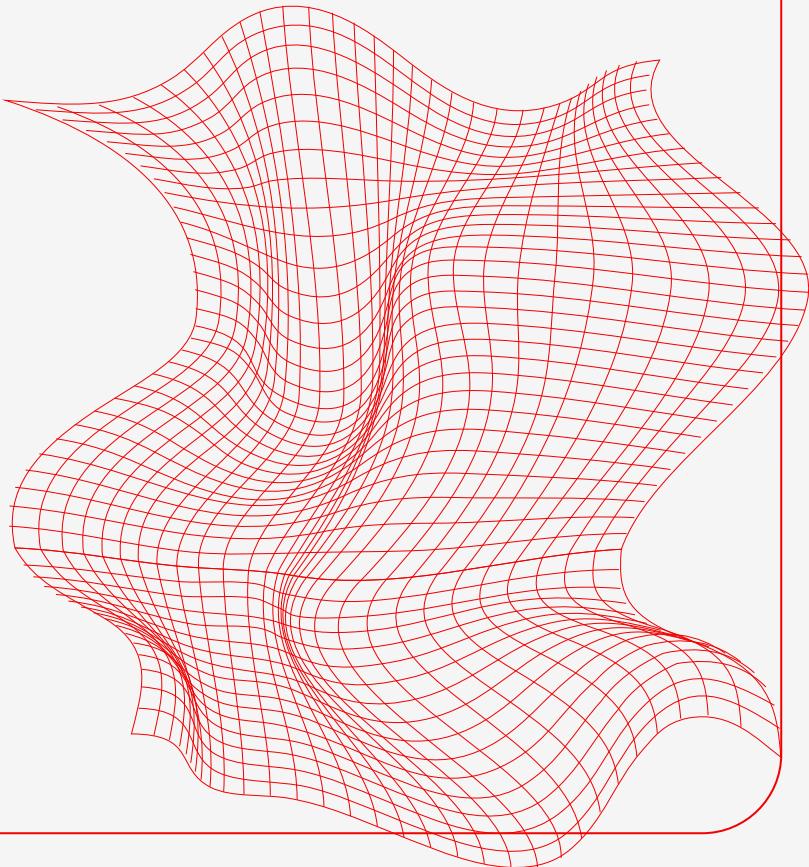
- Deepnote could not handle all 1.6 million rows of data
- Cannot download the data to run locally due to security concerns

## Our Solutions

- Initially ran analysis on 1000 rows, but then received 1.6 million later
- From 1.6 million rows, randomly sampled a smaller portion of data
- Think of our project as a test-run of data analysis on the full dataset!

02

# Petition Analysis



# Getting the Data

How we attempted using all of our data.

- 1.6 million rows of data total
- Full dataset could only be imported into Deepnote in sub-datasets
- Each dataset has around 100k-200k rows
- Sorted chronologically

2018

Dataset Rows 0-150k

Dataset Rows 150k-300k

Dataset Rows 300k-450k

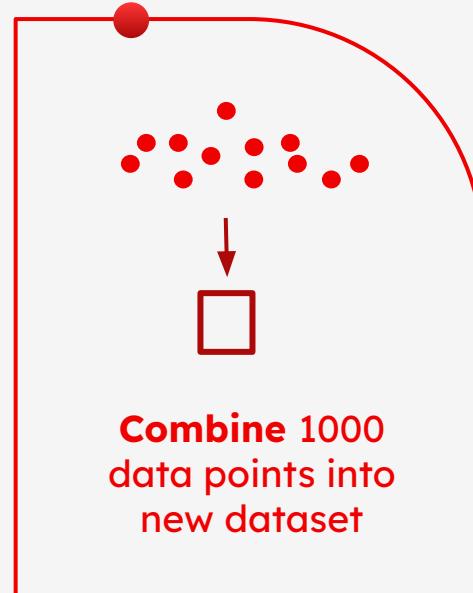
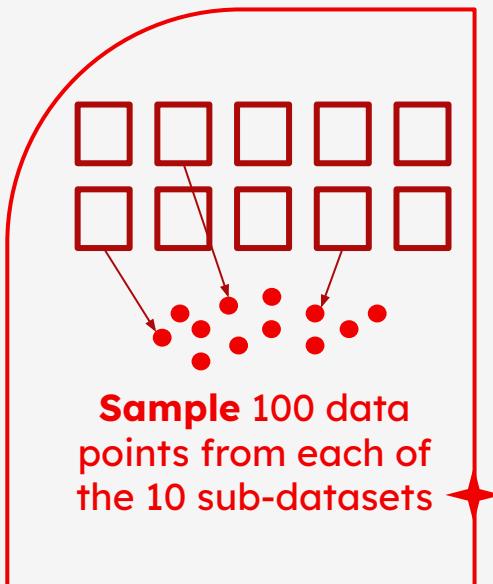
⋮

2023

Dataset Rows 1.4m-1.6m

# Random Sampling

On all the different combinations of data points

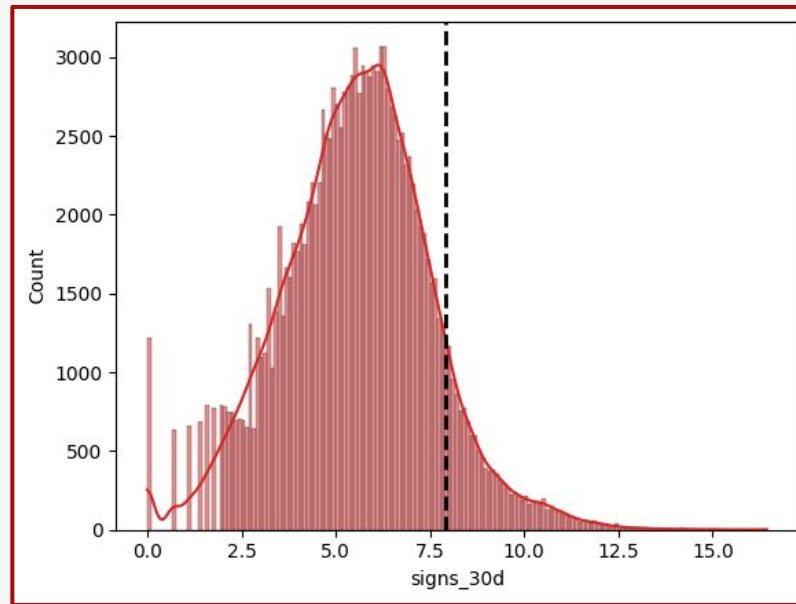


# Creating a Success Metric

Existing 'status' column does not always signify high success

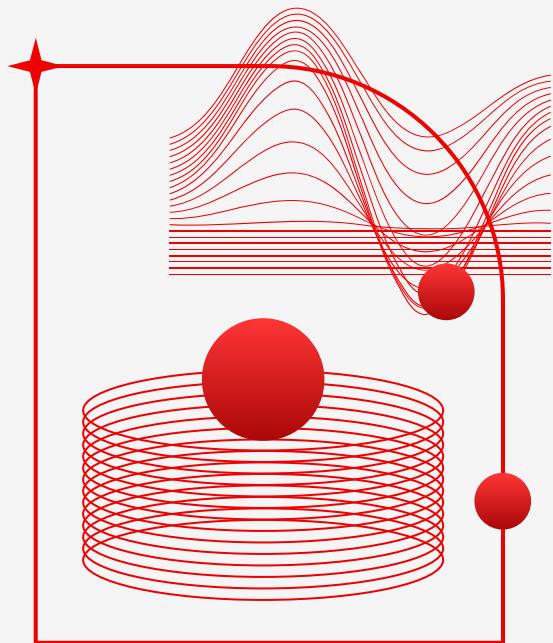
**Success metric:**

Petitions with # of signatures after 30 days that are above the 90th percentile



Histogram of Log of Number of Signatures on Day 30 Among All Petitions

# Factors of Growth



## Polarity

Calculated if petition descriptions had a “positive, negative, or neutral” sentiment

01

## Subject Matter

Identified the most important keywords amongst petitions

02

## Model

Experimented with various types of model to find the best predictor of petition success

03

# Polarity Score

Measures the overall sentiment of a particular text, ranged from [-1.0, 1.0] with -1.0 being negative and 1.0 positive.



# Polarity Scores

## POSITIVE SENTENCE:

"Petitions empower communities and foster change by enhancing voices of individuals hoping for a better world."

### Polarity score:

```
{'neg': 0.0, 'neu': 0.695, 'pos': 0.305, 'compound': 0.6908}
```

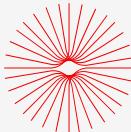
## NEGATIVE SENTENCE:

"Petitions can be disregarded and face obstacles, dampening the hopes of those seeking meaningful change."

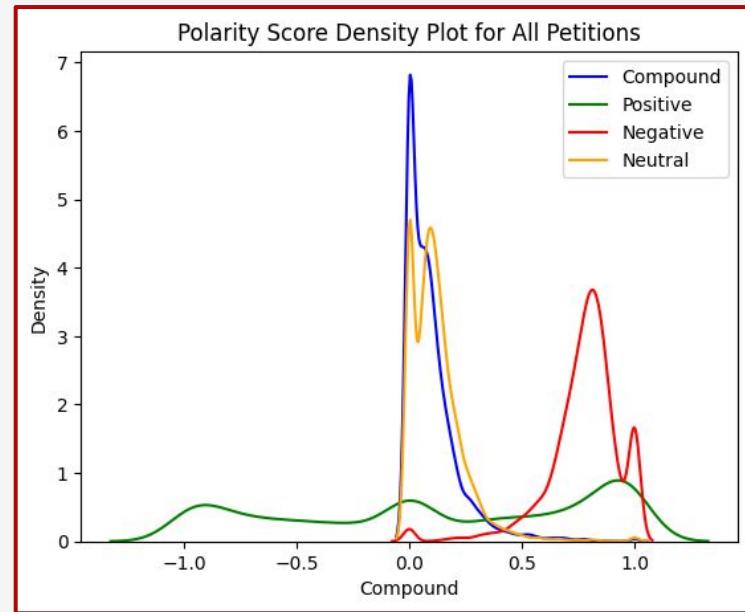
### Polarity score:

```
{'neg': 0.244, 'neu': 0.516, 'pos': 0.239, 'compound': -0.0258}
```

# Polarity Scores



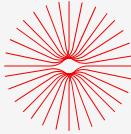
- **Compound:** central around 0, leaning towards more positive
- **Positive:** low, but spread throughout most petitions
- **Negative:** skewed to the right
- **Neutral:** similar to compound



Kernel Density Plot of Polarity Score on 10,000 randomly sampled rows



# Polarity Scores



- **10% of petition descriptions have polarity score = -1.0**
  - Changes with random sample
- **There existed petitions that contained spam content**
  - Might skew the polarity scores for petitions

Example Table:

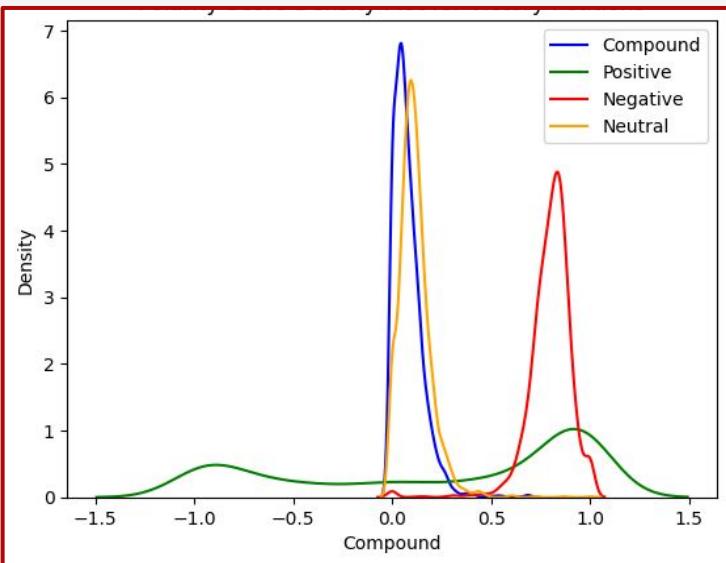
Title	Description
<: aaa	aaa
MLB: Get MLB to remove...	Let Pete back in baseball!
Yes	Its me

# Polarity Scores

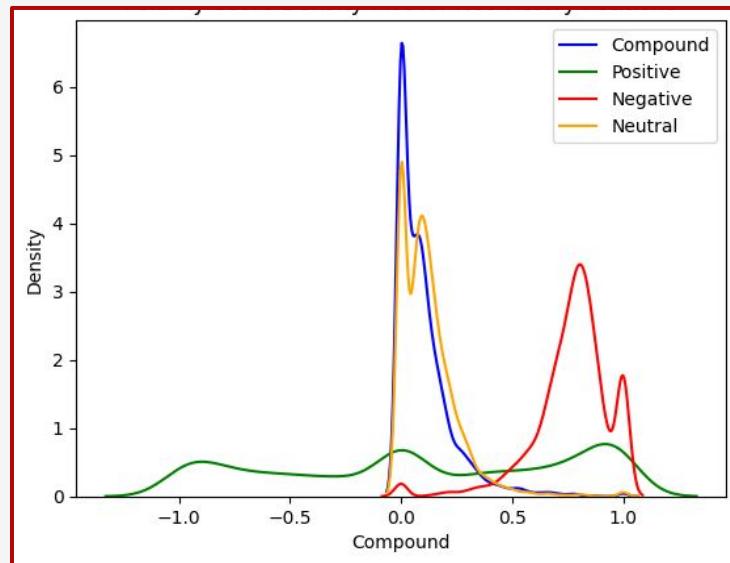
## Factors of High-Signature Petitions:

1. Strong wording
2. Acting *against* or preventing something

### 'Successful' Petitions



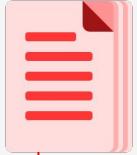
### 'Unsuccessful' Petitions



# Keyword Extraction with Yake:

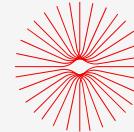
Yet Another Keyword Extractor

“unsupervised automatic keyword extraction method”



## 1. Text Pre-Processing

cleaning and normalizing the text,  
tokenizing text



## 2. Feature Calculation

calculating term frequency, position of  
word, degree of word



## 3. Keyword Scoring

calculate score based on steps 1 & 2

# Yake

in action

“Save the Gibson House Museum”

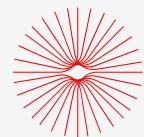
“Stop the “Adventure Park”

“Make BULLYING a criminal offense”



“Gibson House Museum”	0.0007
“Park”	0.206
“Criminal”	0.588

Deduplication Threshold: 0.1  
Max # of Words per Keyword: 3  
Language = English  
*\*low deduplication due to the smaller example*



# Yake

*ask column*

animal	0.175
kids	0.174
Home	0.169
Give	0.168
Parents	0.162

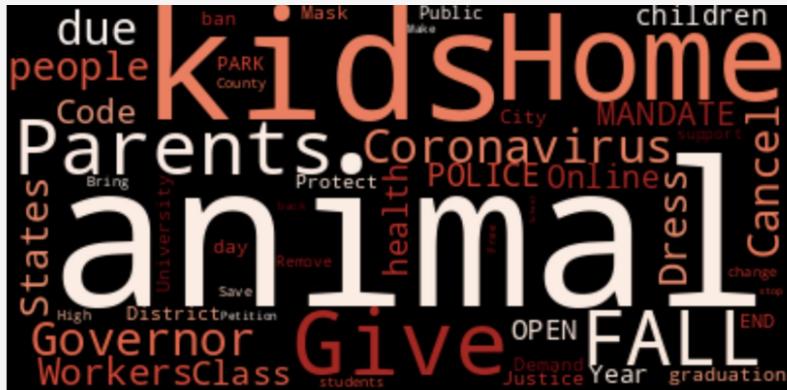
Top 5 Successful Keywords

MOD	0.0034
states	0.0034
time	0.0033
National	0.0032
ENDLESS	0.0032

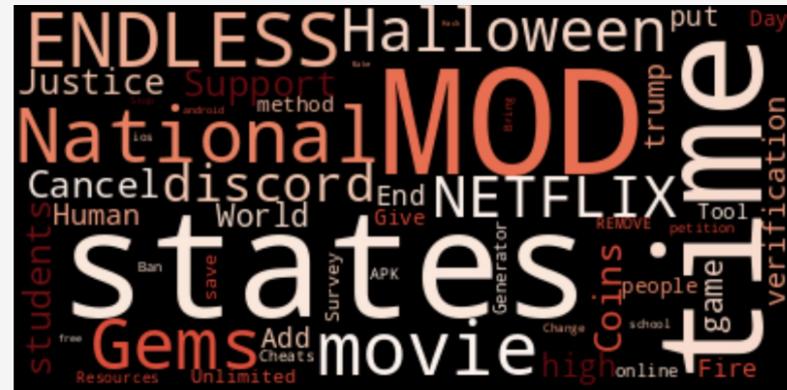
Top 5 Unsuccessful Keywords

# Yake WordClouds

## *ask column*



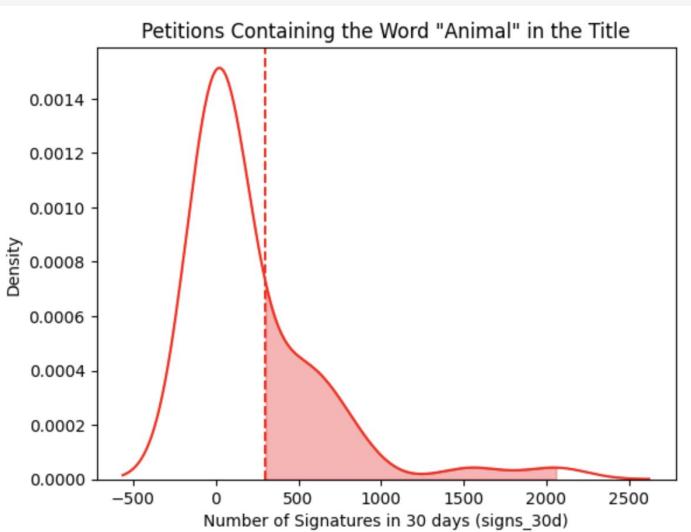
# Successful Petitions



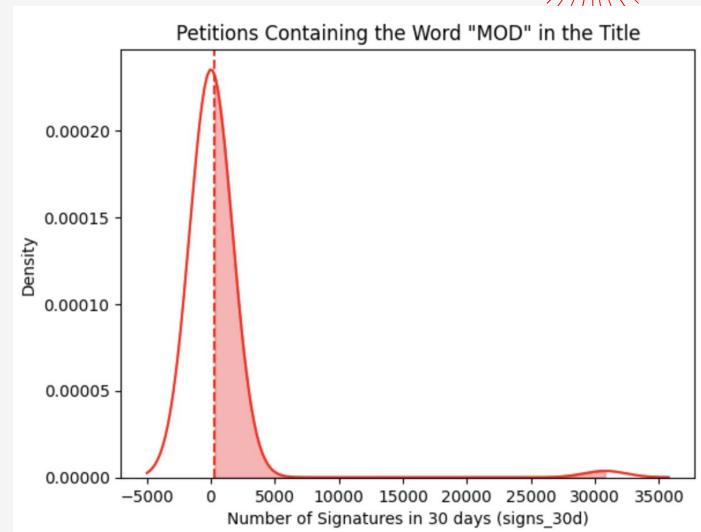
## Unsuccessful Petitions



# Validating Success w/ Yake



Type: **Successful Keyword**  
Success Rate: 0.255 or 25.5%



Type: **Unsuccessful Keyword**  
Success Rate: 0.034 or 3.4%

# Regression Models

Input petition features and predict the # of signatures after 30 days

## Linear

**Pro:**  
Very high accuracy  
Training  $R^2 = 0.8396$   
Testing  $R^2 = 0.8586$

**Con:**  
Could be overfitting

## Poisson

**Pro:**  
Useful for non-normal distributions

**Con:**  
Assumptions not met due to overdispersion

## Negative Binomial

**Pro:**  
Accounts for overdispersion

**Con:**  
Needs more data & text variables

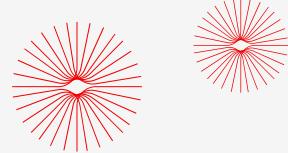
# Key Takeaways: Linear Regression



- Accuracy of predicting successful petition: 85.86%
- Training accuracy is smaller than testing indicating that the **model is good predictor for the data**
- High correlation between signs\_30d and signs\_7d compared to other predictors
  - Additional influential features needed to decrease weight of single predictor
- Used solely numerical variables, could improve accuracy through categorical (such as user/staff tags)



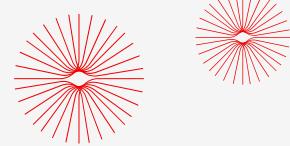
# Analysis of Features: Neg Binomial



Feature	7-Day Signatures	1-7 Day Signatures Ratio	Compound Ask Polarity Score	Number of User Tags	Number of Profanity Words	Amount of Punctuation	Ask Length
Weight in Model	0.002	0.050	0.285	0.027	-1.358	-0.001	0.014
p-value	~~0	~~0	~~0	0.109	~~0	0.868	~~0

Weights: average change of '30-Day Signatures' associated with a one unit increase in each feature

# Analysis of Features



## Strong Features

- 7 Day Signatures
- 1-7 Day Signatures Ratio
- Compound Ask Polarity Score
- # of Profanity Words
- Ask Length
- Has either a staff tag or user tag
- Has a photo

## Weak Features

- Number of User Tags
- Amount of Punctuation
- Description Length
- Having zero signatures at 7 Days

# Key Takeaways: Negative Binomial Regression

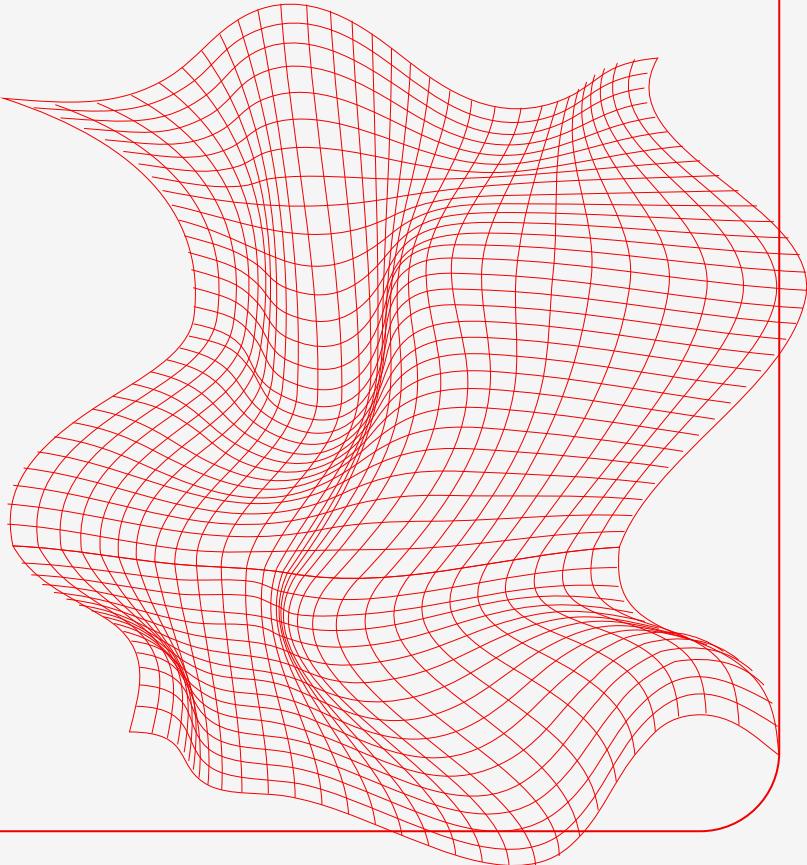


- Accuracy of predicting successful petition: 0.8128%
- We found some good features!
  - \*Not consistent with every random sample
- Log-likelihood is very small indicating that the **model is poor fit for the data**
- Used a Frequentist approach, could have more potential with a Bayesian approach
  - Updating beliefs using observed data



03

## Harmful Petitions



# Harmful Petitions - Background, Thought Process

**1** “Harmful” / “Abusive” Meaning?

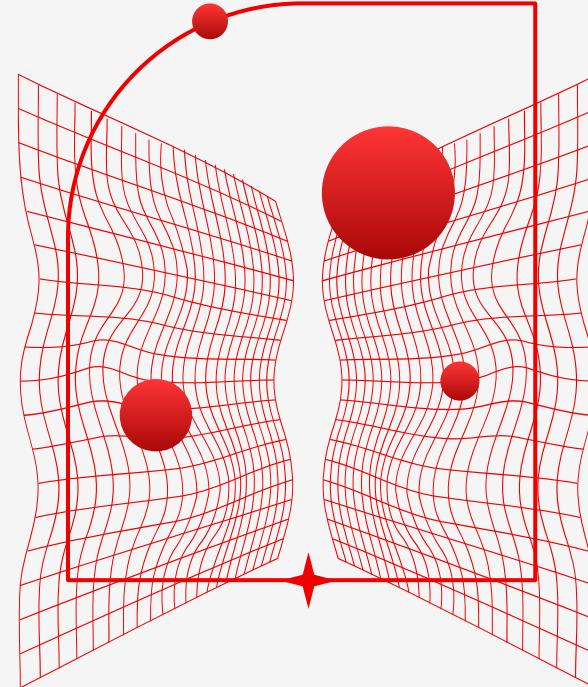
**2** The Complexity of “Harm”

**3** Freedom of Speech

**4** Minimizing Flagging of Good Petitions

# **Overview of Our Data Analysis**

# Perspective API



Background

Petition Analysis

Harmful Data

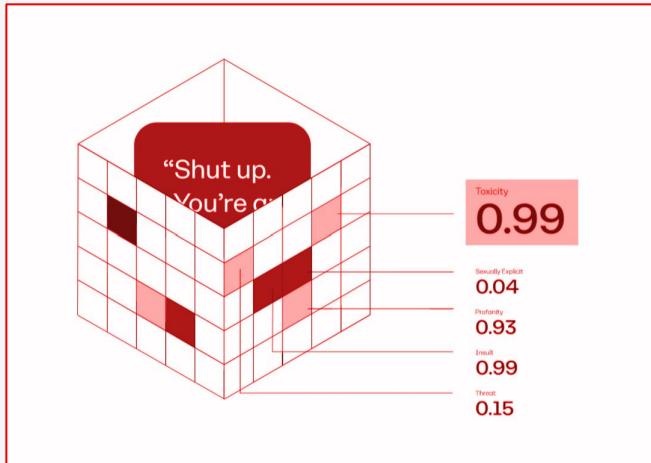
Challenges/Future Work

# Toxicity Scoring

- Identify and filter online content that may be considered "toxic" or harmful
- **Natural language processing (NLP)** algorithms to analyze text and assign a toxicity score
- Returns **probability (0-1)** of the content containing abusive, threatening, or offensive language

Text:	Toxicity Score:
“Safety in Rockingham County Public Schools”	<b>0.006345861</b>
“Derek should shut the f*ck up❤”	<b>0.9029226</b>

# How Perspective API Works



1 Severe toxicity

2 Insult

3 Profanity

4 Identity Attack

5 Threat

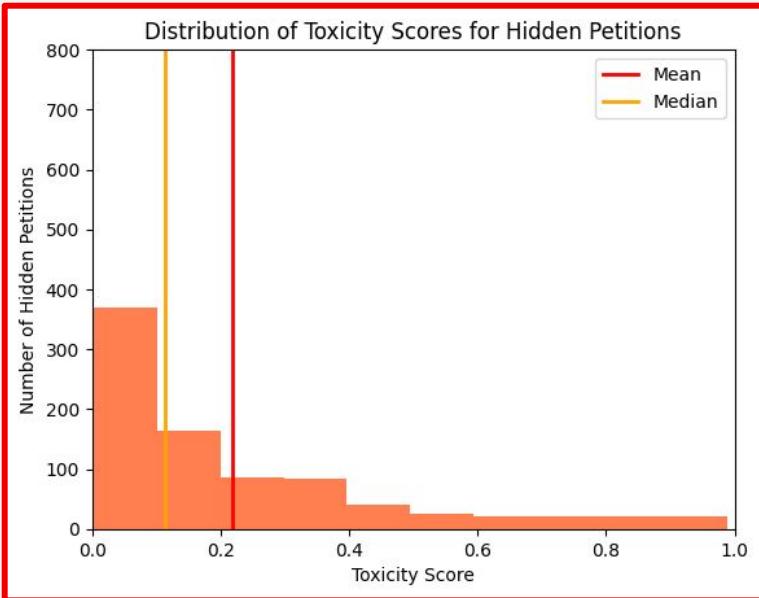
6 Sexually Explicit

# Categories as defined by Perspective API

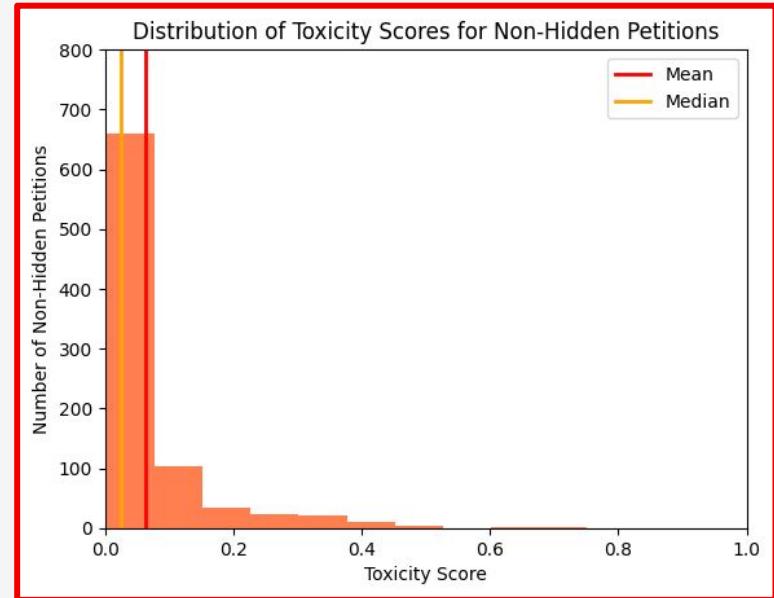
Toxicity Level	Description of level
Very Toxic	A comment that is very hateful, aggressive, disrespectful, or otherwise very likely to make a user leave a discussion or give up on sharing their perspective.
Toxic	A comment that is rude, disrespectful, unreasonable, or otherwise somewhat likely to make a user leave a discussion or give up on sharing their perspective.
Not Toxic	A neutral, civil, or even nice comment very unlikely to discourage the conversation.
I'm not sure	The comment could be interpreted as toxic depending on the context but you are not sure.

Category	Definition
Profanity/ Obscenity	Swear words, curse words, or other obscene or profane language.
Identity-based negativity	A negative, discriminatory, stereotype, or hateful comment against a group of people based on criteria including (but not limited to) race or ethnicity, religion, gender, nationality or citizenship, disability, age, or sexual orientation.
Insults	Inflammatory, insulting, or negative language towards a person or a group of people. Such comments are <b>not</b> necessarily identity specific.
Threatening	Language that is threatening or encouraging violence or harm, including self-harm.

# Hidden vs. Non-Hidden Toxicity Scores



**Mean: 0.218**  
**Median: 0.114**



**Mean: 0.062**  
**Median: 0.025**

**0.1549 - 0.1562**

Difference between hidden and  
non-hidden toxicity scores

## Challenges:

01

Not openly accessible to use: must request access from Google

02

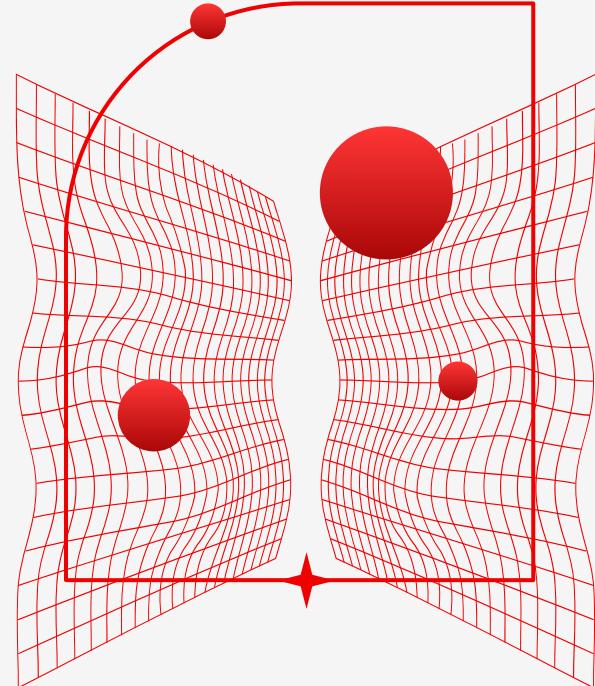
Server request error: standard API  
Package cannot process large-scale requests

## Future Exploration:



Use as a feature, linear modeling

# Profanity Exploration



Background

Petition Analysis

Harmful Data

Challenges/Future Work

# Petitions with 1+ Profanity Word (1 sample)

$46/826 = 5.7\%$

**hidden**

$44/9084 = 0.5\%$

**non-hidden**

# Petitions with 1+ Profanity Word: Bootstrap

~ 5.3-5.6% difference

[5.80%, 5.99%]

hidden

[0.37%, 0.42%]

non-hidden

Hidden petitions are ~**13-15 times** more likely to have 1+ profanity word (95% confidence, 200 samples, 10000 rows)

## Synthesizing Results: Perspective + Profanity Analysis

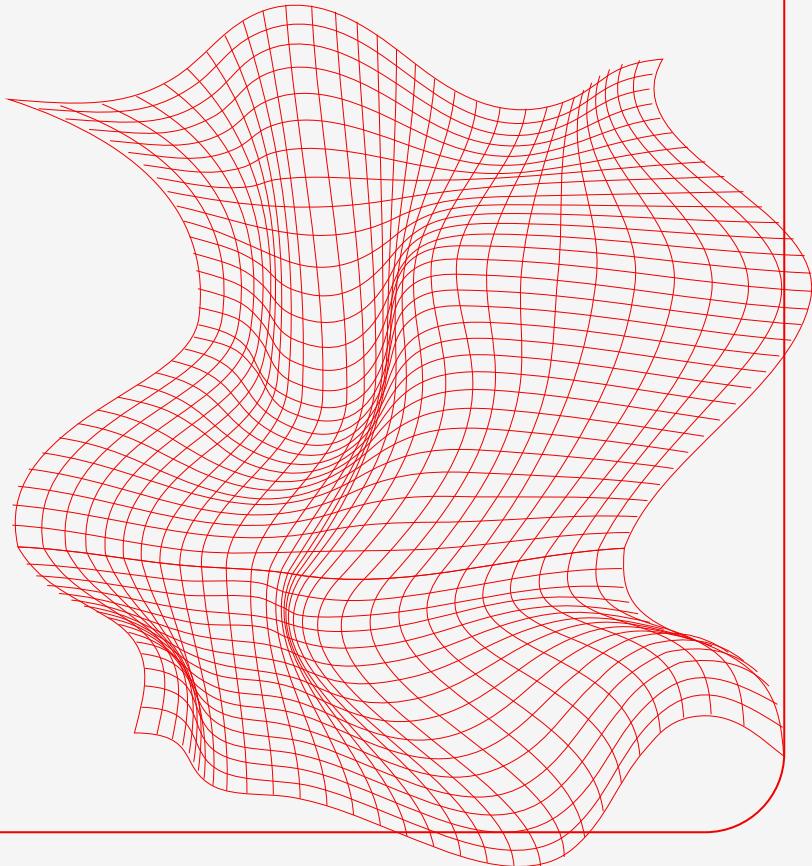
**Perspective Alone:** reasonably large toxicity difference for hidden vs non-hidden

**Profanity Analysis Alone:** somewhat large profanity difference for hidden vs non-hidden

**Perspective + Profanity Analysis:** other factors may play larger role into “toxicity”/harm than profanity. Profanity alone may just be a small factor.

04

## Challenges and Future Work



# Challenges



## Runtime

Large amount of data took an extremely long time to run through the cloud



## Updating Data Pipelines

Had to update code with every new file of data sent

## Modeling

Tested many different types of modeling before landing on final model



## Defining Harmful

Had to research and define the boundaries of what makes content harmful



# Next Steps



## Improve NLP

Create staff tags for every description

## Modeling

Unsupervised categorization to create tag categories for a semi-supervised model



## Server Change

Move work to platform with more computing power

## Location Analysis

Look more into how location impact petition signatures



# Key Takeaways

## Petition Analysis

### Best Features Associated with Success:

- 7 Day Signatures
- Compound polarity score

## Predicting Success

### Best Models:

- Negative Binomial
- Linear Regression

## Harmful Data

- Harm is multifaceted
- ‘Hidden’ should be examined further as a metric of harm
- Toxicity scores may be good indicators of harm

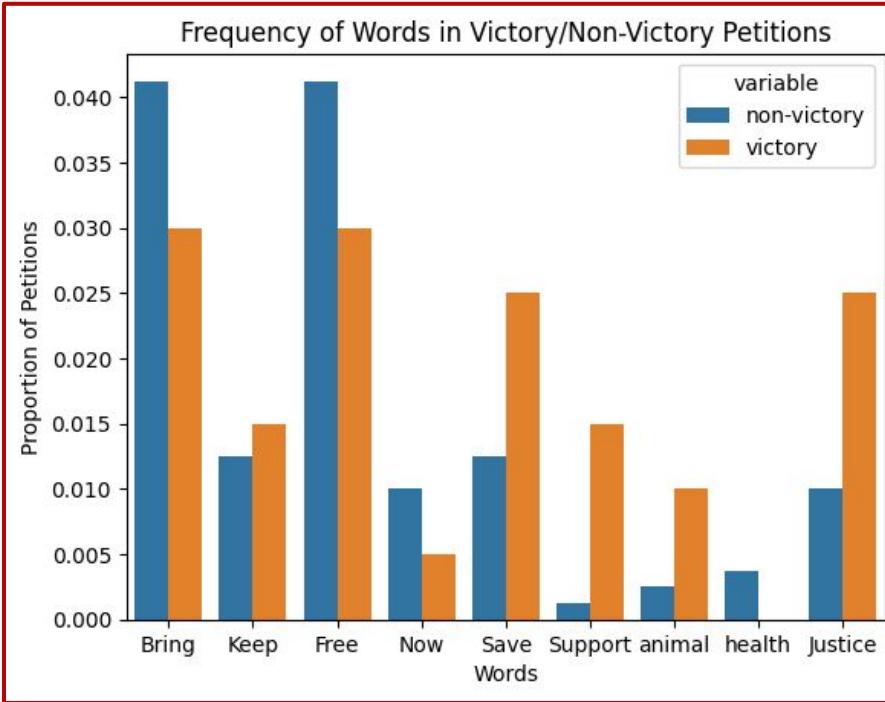
**Thank you!  
Questions?**



# Appendix

Slides We Didn't Present

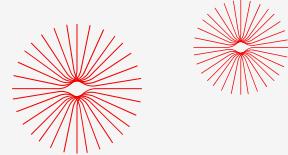
# Word Frequency



## Titles of Petitions: 'ask'

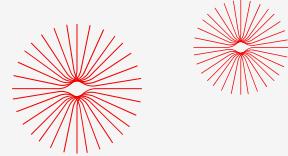
- Looking for visual comparisons between proportion of petitions
- Themes among successful petitions
  - **Keep, Save, Support, Animal, Justice**

# Analysis of Features: Neg Binomial



Feature	7-Day Signatures	1-7 Day Signatures Ratio	Compound Ask Polarity Score	Number of User Tags	Number of Profanity Words	Amount of Punctuation	Ask Length
Correlation w/ 30-Day Signatures	0.893	0.102	-0.005	0.166	0.007	0.036	0.060

# Analysis of Features: Neg Binomial



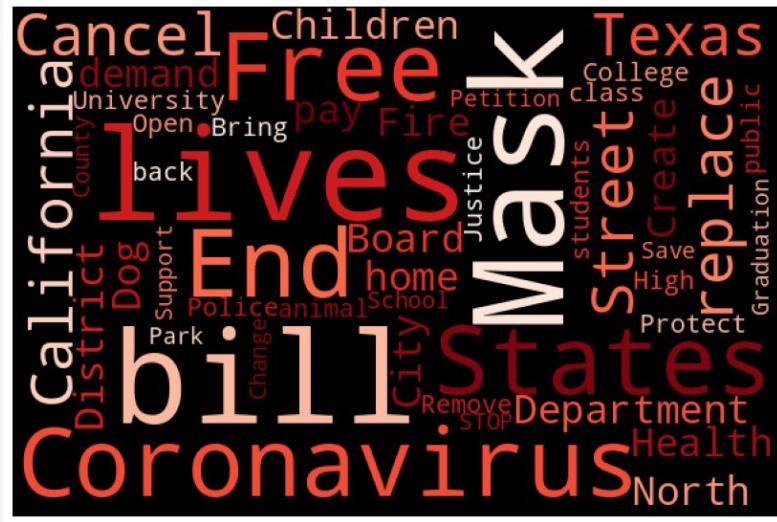
Generalized Linear Model Regression Results						
Dep. Variable:	signs_30d	No. Observations:	5020			
Model:	GLM	Df Residuals:	5008			
Model Family:	NegativeBinomial	Df Model:	11			
Link Function:	Log	Scale:	1.0000			
Method:	IRLS	Log-Likelihood:	-25408.			
Date:	Wed, 03 May 2023	Deviance:	1.0732e+08			
Time:	06:04:40	Pearson chi2:	1.80e+05			
No. Iterations:	100	Pseudo R-squ. (CS):	0.9675			
Covariance Type:	nonrobust					
		coef	std err	z	P> z	[0.025 0.975]
Intercept	1.8423	0.048	38.631	0.000	1.749	1.936
signs_7d	0.0020	5.44e-06	364.888	0.000	0.002	0.002
signs_ratio	0.0502	0.001	71.651	0.000	0.049	0.052
Compound_ask	0.2847	0.078	3.656	0.000	0.132	0.437
num_user_tags	0.0272	0.017	1.603	0.109	-0.006	0.060
profanity_count	-1.3582	0.189	-7.197	0.000	-1.728	-0.988
punctuation_count	-0.0005	0.003	-0.166	0.868	-0.007	0.006
description_len	3.69e-06	4.88e-06	0.756	0.450	-5.88e-06	1.33e-05
ask_len	0.0142	0.001	15.160	0.000	0.012	0.016
has_tag	0.7303	0.038	19.262	0.000	0.656	0.805
has_photo	0.6428	0.037	17.397	0.000	0.570	0.715
no_signs	-0.0033	0.076	-0.043	0.966	-0.153	0.146

Log-likelihood very low, indicating poorly fitted model.

# Word Cloud: Hidden vs. Non-Hidden



# Hidden



# Non-Hidden