Analysis of Midterm Election Discussion on Twitter

By Kyle Arbide and Zhaoyang Chen

Problem

It is impossible to understand the full scope of discussion about elections on Twitter

Solution

Provide accessible statistics and visualizations of topics being discussed on Twitter, powered by Natural Language Processing

Project Breakdown

Data Collection	Natural Language Processing	Interface
Twitter API	Text Preprocessing	Streamlit
	Exploratory Analysis	
	Topic Modeling	

Data Collection

- Tweets were collection continuously from September 20th - November 10th
- Terms and hashtags were specified identify tweets regarding the 2022 Midterm Elections
  - Ex: “#Midterm2022, #Senaterace” etc.
- License was granted through GWU with Social Feed Manager
  - https://gwu-libraries.github.io/sfm-ui/

Cleaning and EDA

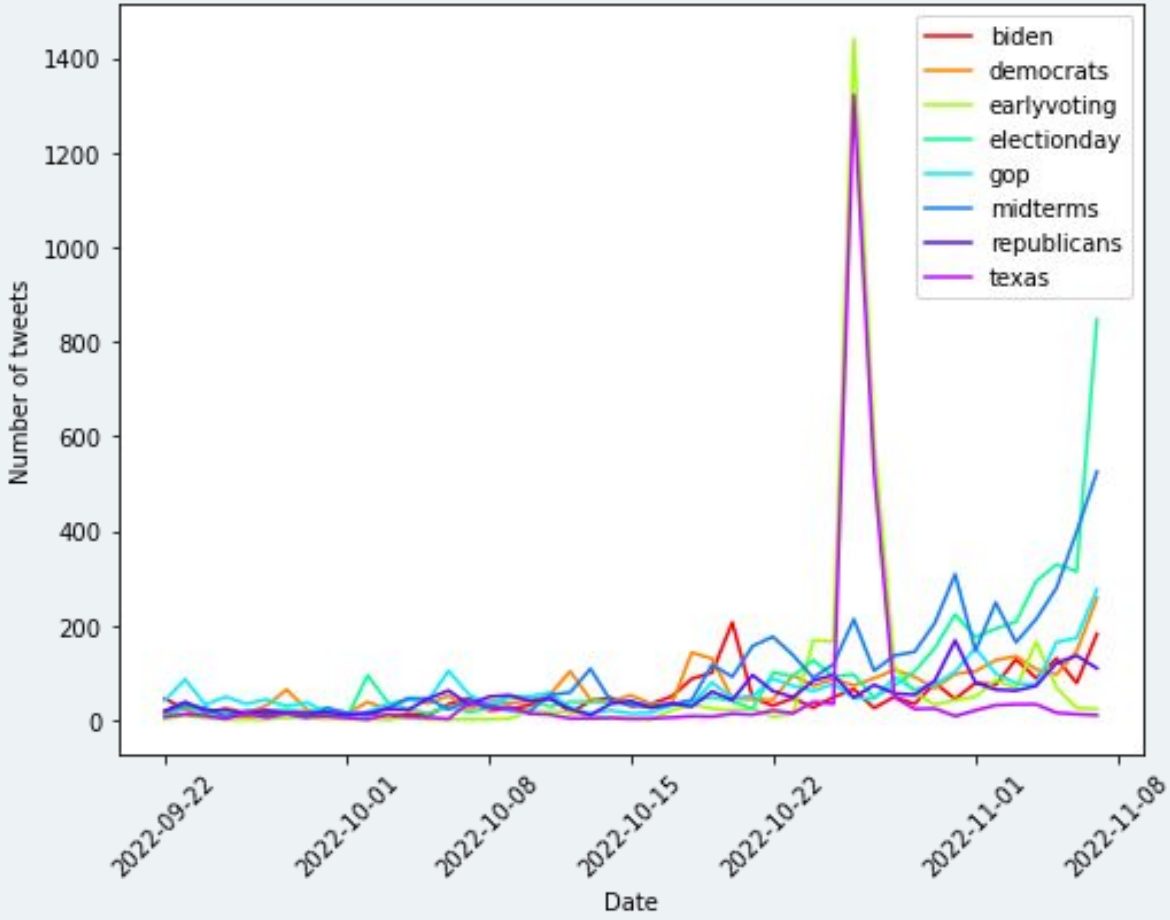
Data Cleaning

- Removing Emojis, Links, Number, Punctuation, and Stop words
- Removing Non-English Tweets

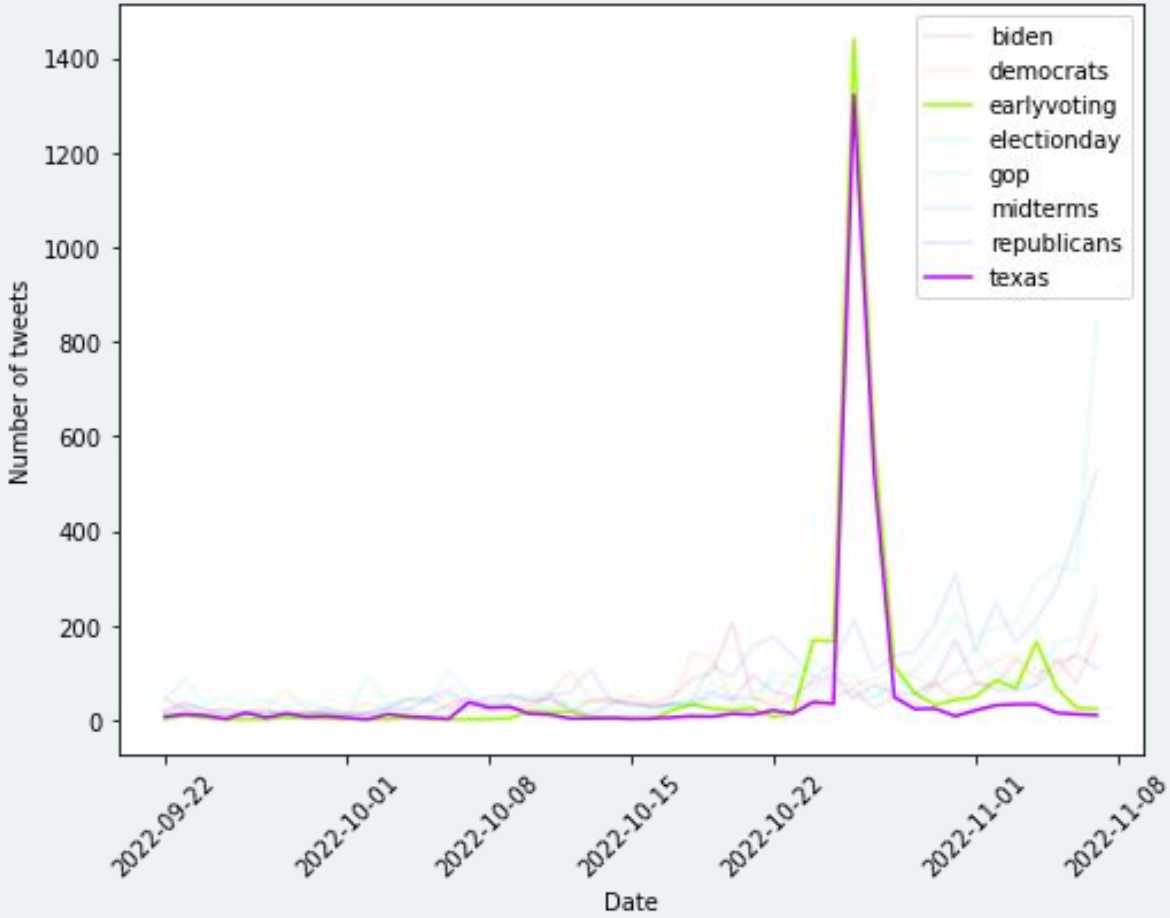
Exploratory Analysis

- Finding Correlation between the use of different hashtags/mentions
- Plotting/Visualization

EDA cont.


biden
democrats
earlyvoting
electionday
gop
midterms
republicans
texas
Number of tweets
0
200
400
600
800
1000
1200
1400
2022-09-22
2022-10-01
2022-10-08
2022-10-15
2022-10-22
2022-11-01
2022-11-08
Date

EDA cont.



Pearson Correlation: 0.988

Topic Modeling

BERT

- Transformer Based Machine Learning Model
- Developed by Google AI

Topic Modeling

- Allows us to create highly interpretable topics for our tweets
- These topics can be used to generalize

Streamlit UI

- Streamlit allows you to create basic web-apps in Python
- Markdown based/Easy to use

```
streamlit run Welocome_Page.py
```

Takeaways

Innovative topic modeling using modern transformer models allow us to create highly interpretable topics. This combined with some basic interfacing and visualization provides much greater context into the election discussion on Twitter