UC San Diego

Analysis of Sentiment around Election Debates

Jaeyoung Kang
Anye Shi
Jasmine Chiang
Aaron Luo

Motivation and Objective

Performance of Candidates

Evaluate connotation of each candidate's debate topics

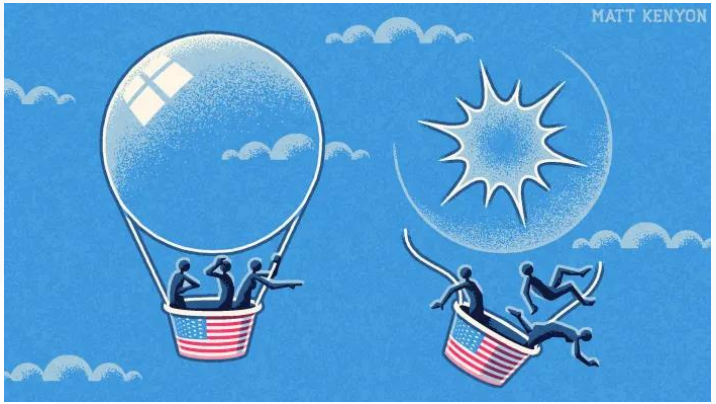
Tweets to try to understand public opinion and draw correlation to debate topics

Dataset

1. Debate Transcript
   - Debate dialogue
   - Speaker
   - Timestamps

2. Tweets by the General Public
   - Used Keywords:

     "Trump", "Biden", "Wallace", "Debate", "Election", "Harris", "Tax", "Pence", "National Security", "Race in America", "Climate Change", "US Economy", "Supreme Court"



Methodology

Data Crawling

- Based on snscrape library
- Generate URL with search queries
- Given URL, parse json response from Twitter

Basic Data Filtering

- Get tweets only written in English
- Exclude short tweets (<20 characters)

Methodology

Tokenizing

- Breaking down a corpus to smaller chunks
  - Sentences, phrases or individual words
- Features used to detect the sentiment

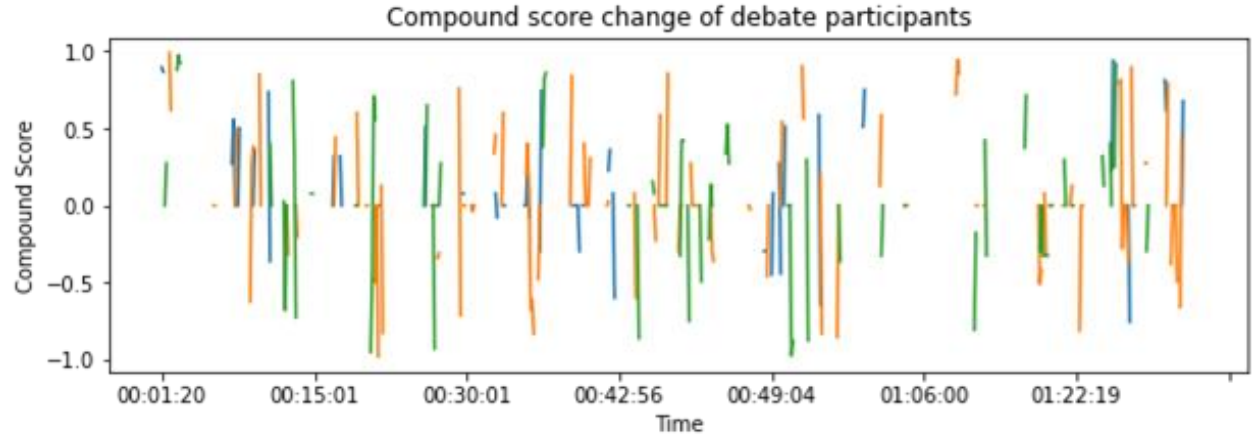
Stopwords

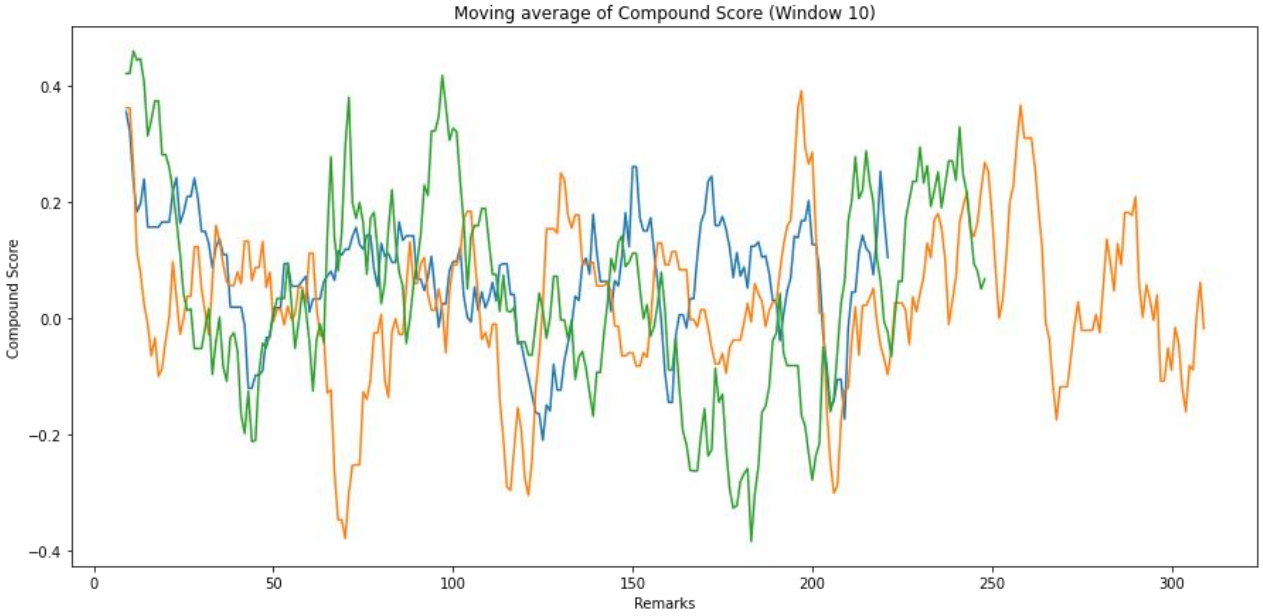
- Words that add little to no extra meaning to the sentence
- EX: a, the, he, she, etc
- Also remove extra punctuation, links, etc

Transcript Analysis Debate 1 [1]

Words Spoken During Debate 1

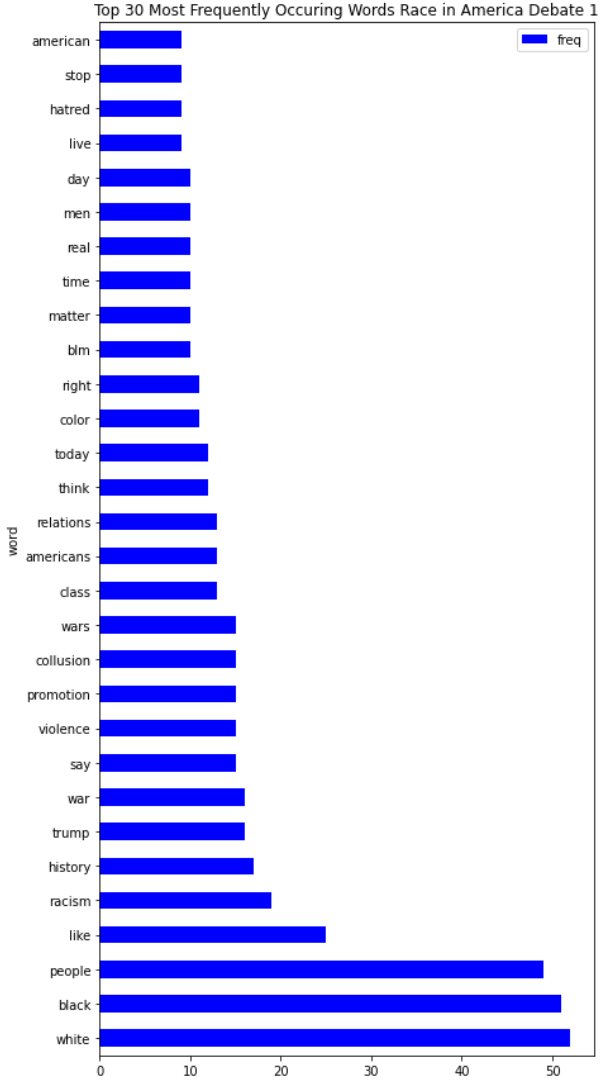
- **Chris Wallace 4552 words**
- **President Trump 7234**
- **President Elect Biden 6620**

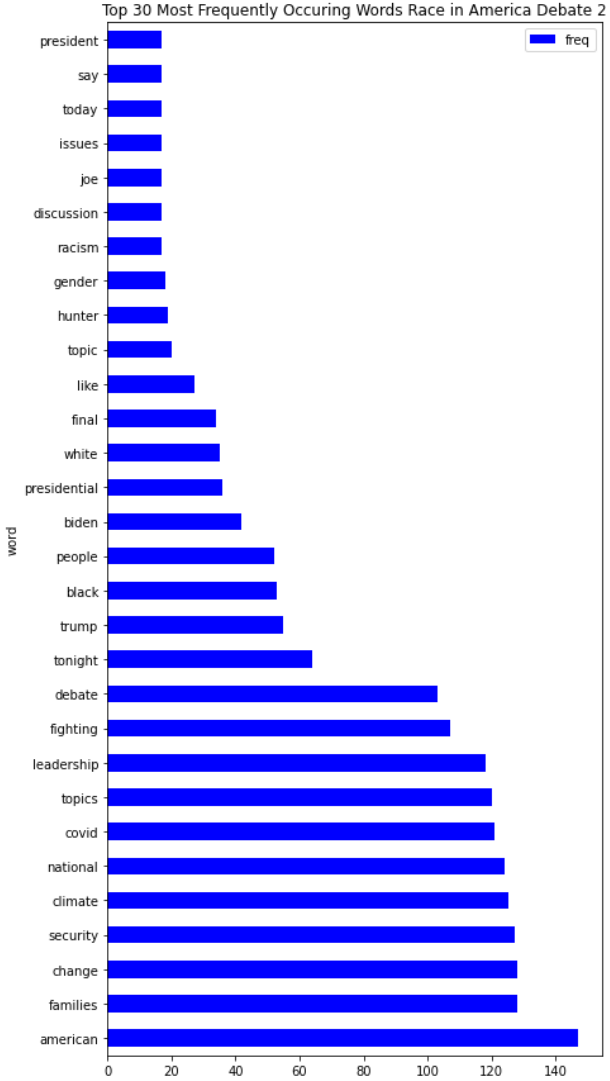




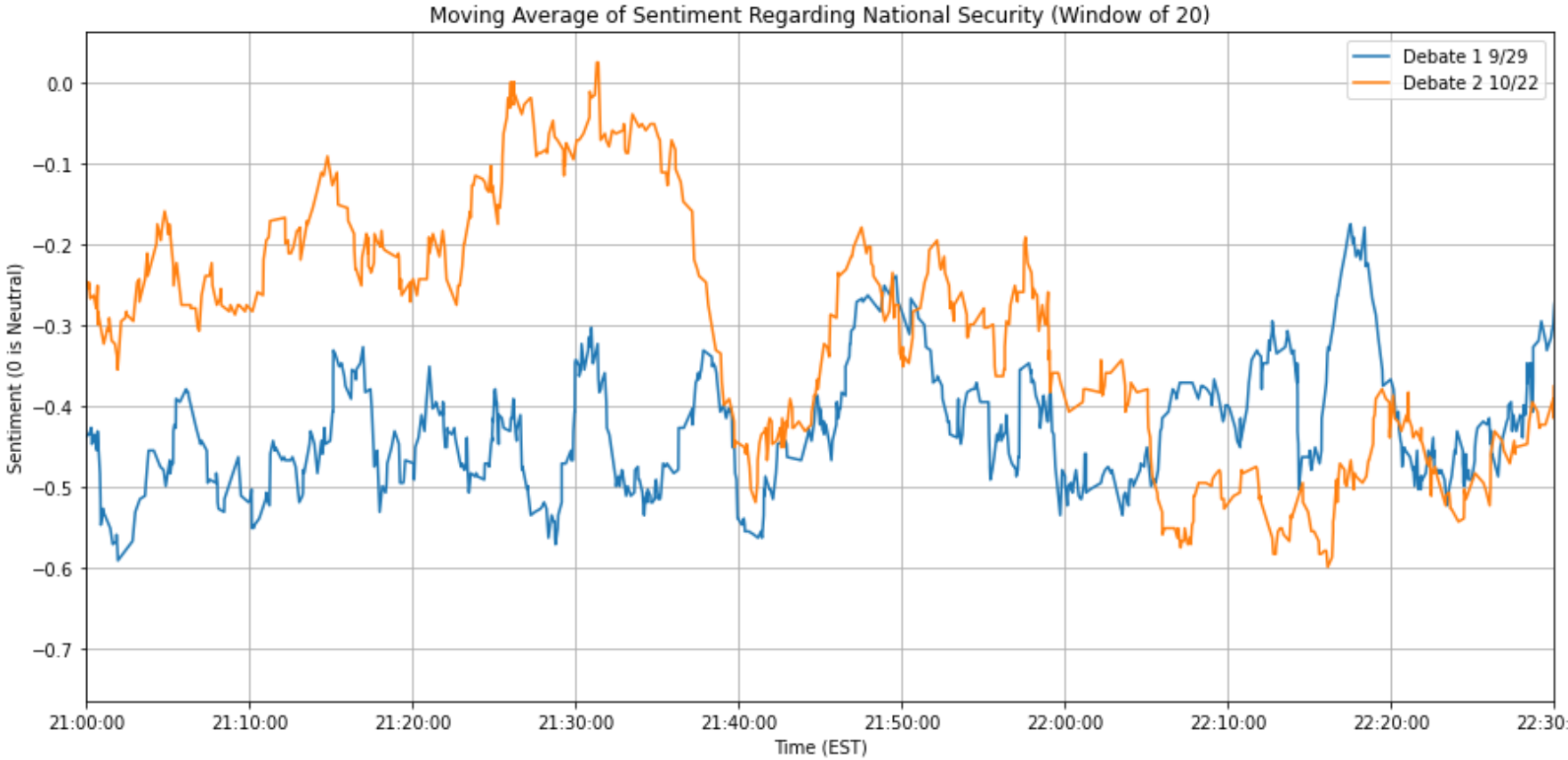
Race in America

Debate 1		Debate 2	
@torrainewalker	4	@realdonaldtrump	20
@realdonaldtrump	3	@joebiden	10
@joebiden	3	@kwelkernbc	7
@thehill	3	@debates	4
@seanhannity	3	@donaldjtrumpjr	3
@bernlennials	3	@nytimes	3
@townsmarquis	3	@emilywestfahey	3
@rffunke	3	@wftate4	3

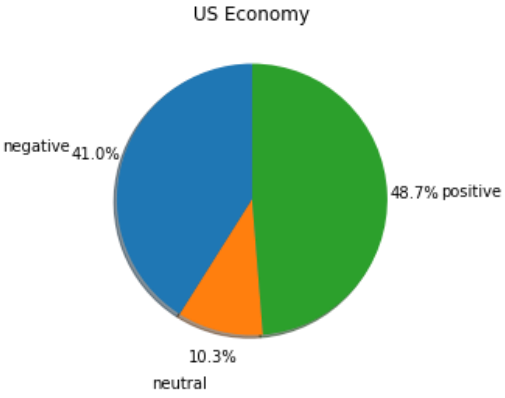


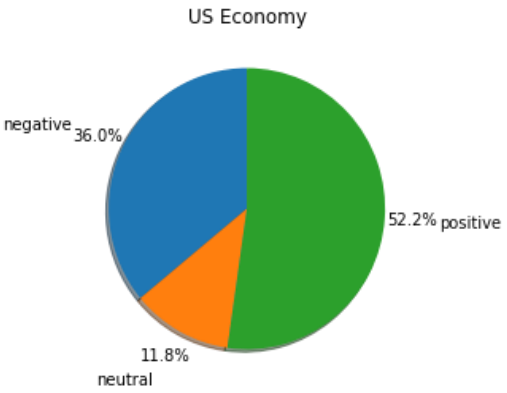


General Opinion: National Security



Debate 1 US Economy Debate 2



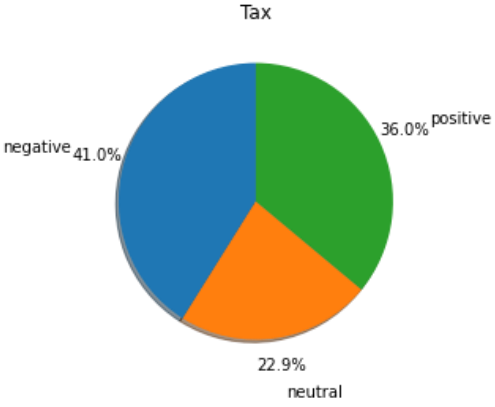


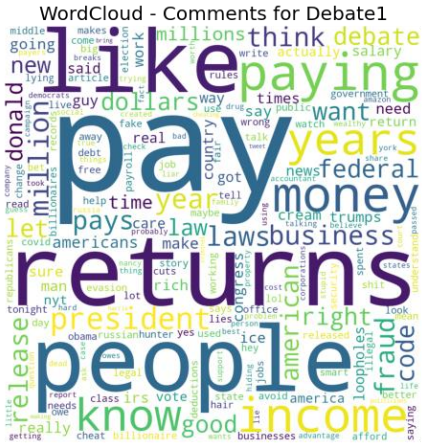
WordCloud - Comments for Debate1

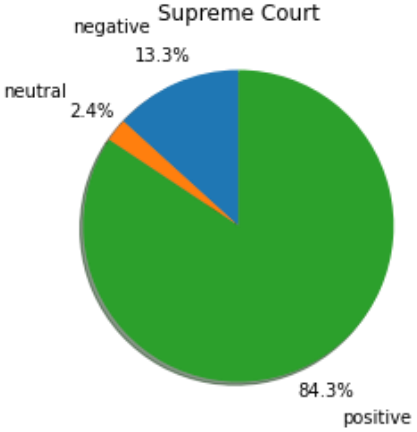
WordCloud - Comments for Debate2

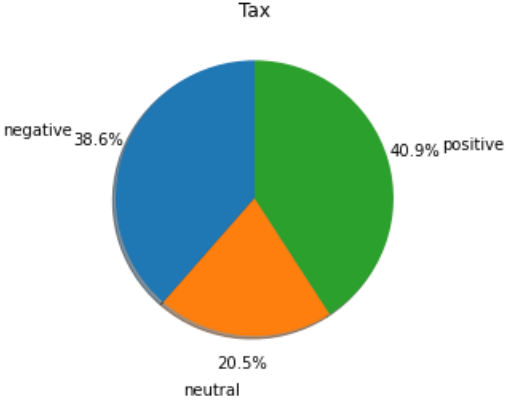
Taxes

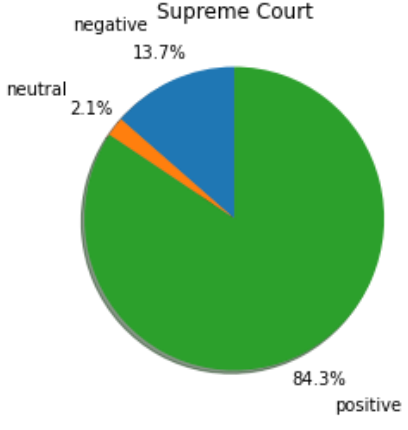
Supreme Court





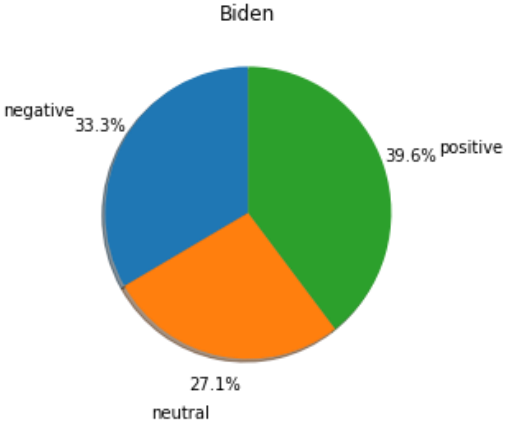


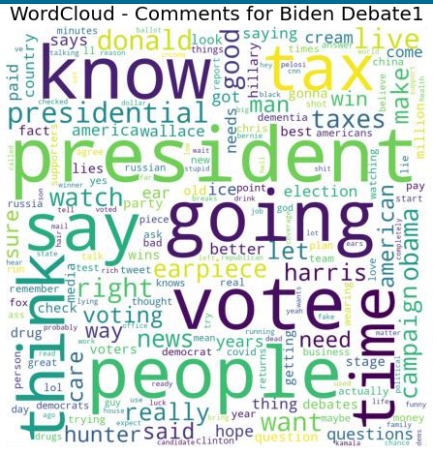


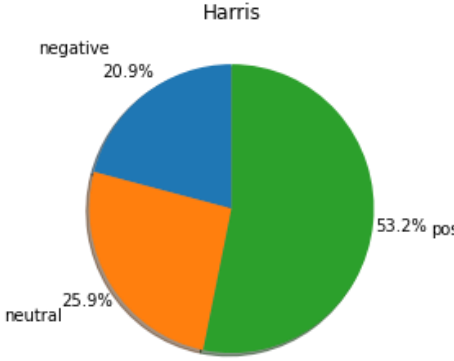


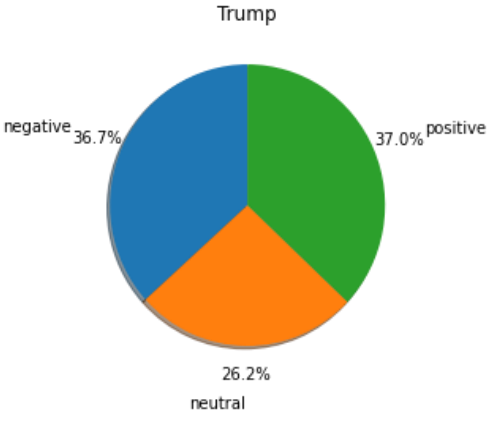
Debate 1 Candidates

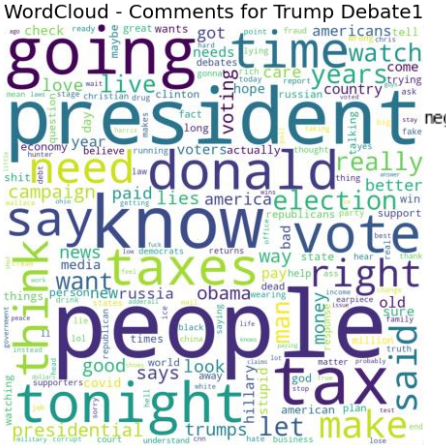
Running Mates

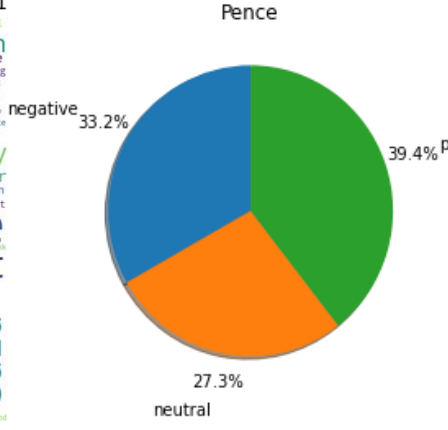






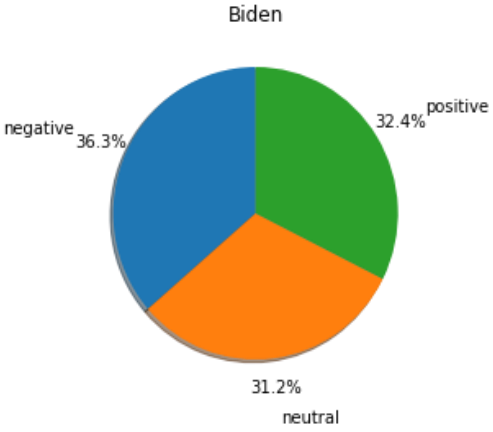


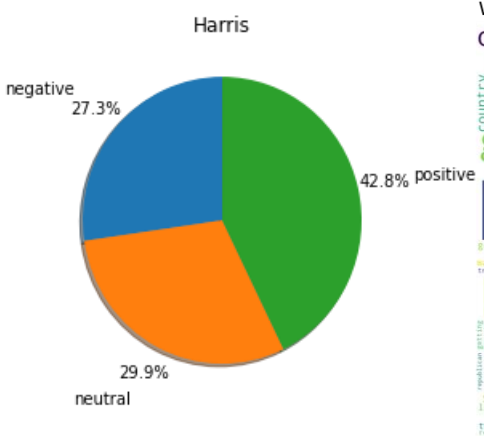


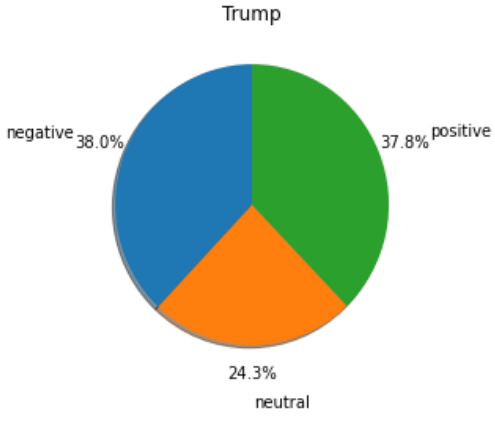


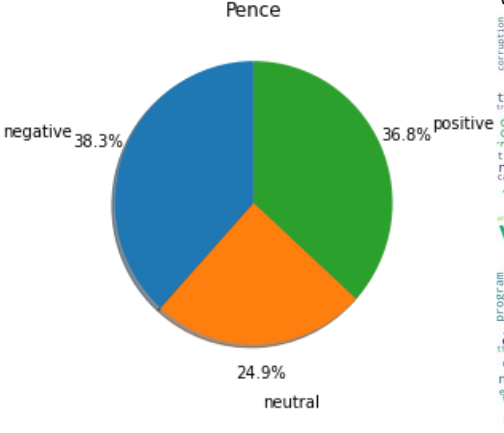
Debate 2 Candidates

Running Mates



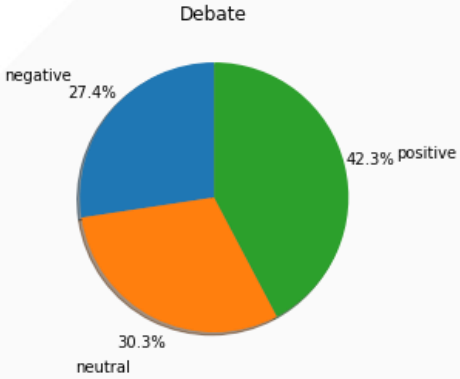


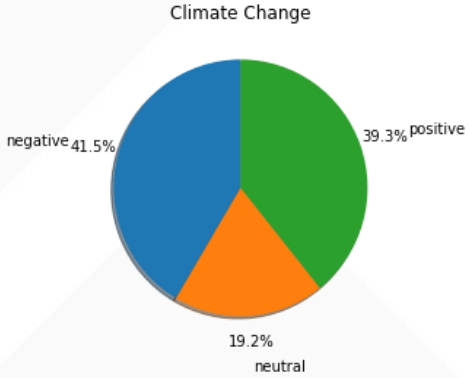


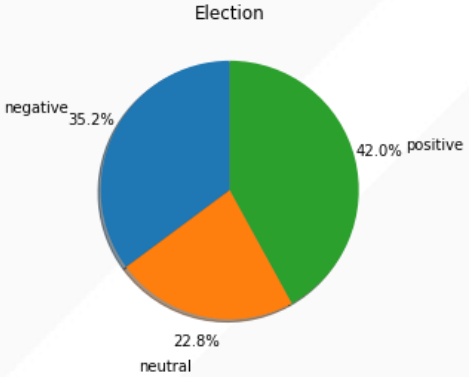


Conclusion

- Sentiment Analysis is hard for complex problems
- The actual spoken words during the debate was generally more negative
- Candidates seemed to have been mentioned in all the topics we searched for
- Sentiment was not overwhelmingly indicative of the results of the election







Citations

[1] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

UC San Diego

Thank You

Any Questions?

Prediction of Sentiment

Classification	Biden	Trump	Debate	Election	ClimateChange	Economy	Race	SupremeCourt	Tax	Harris	Pence	Wallace	Walker
Randome Forest	0.49	0.46	0.51	0.53	0.42	0.60	0.33	0.40	0.50	0.47	0.46	0.56	0.51
Logistic Regression	0.58	0.51	0.46	0.60	0.42	0.42	0.45	0.40	0.54	0.71	0.54	0.58	0.61
Decision Tree	0.48	0.52	0.39	0.55	0.46	0.46	0.43	0.53	0.47	0.46	0.51	0.62	0.47
SVM	0.63	0.62	0.67	0.65	0.60	0.67	0.68	0.69	0.62	0.78	0.65	0.73	0.66
CNN	0.68	0.71	0.75	0.72	0.73	0.76	0.68	0.70	0.72	0.79	0.74	0.73	0.69