

---

# Project Background

- We are all aware of biases in media today and how the media can politically incline readers towards one side. Without the knowledge of political attitude of an article or its source, readers can be manipulated in their opinion formation.
  - Political inclination towards left (liberal) vs right (conservative) is analyzed via Nobias.
  - Content engagement (popularity) is measured through Alexa
-

---

---

# Political Topics

- Amy Barrett
  - Biden
  - BLM
  - Build The Wall
  - Concede
  - Court Packing
  - Covid Vaccine
  - Election Fraud
  - Georgia Senate
  - Green New Deal
  - Stimulus Package
  - Trump
-

---

# Data Collection: Nobias

Nobias offers a browser extension that helps you explore the political slant, credibility, and readability of any article before you click.



[Download the Extension](#)

**Know the bias and credibility of any article**

Enter an article link

---

# Data Collection: Nobias

www.nytimes.com > news-event > 2020-election

 [2020 Presidential Election Results: Joe Biden Wins - The New](#)

## POLITICS

### POLITICAL SLANT

likely center-left

This source often slightly favors liberal ideas and causes. The article's leaning is not available at this time.

### CREDIBILITY

highly credible

This is a top international, national, or business news source.

To read more about how we determine slant and credibility, [click here](#).



resident of the United States on Nov. 7. See full  
election.

ernational ▼

mic Times

20: With the 2020 US presidential election pitting two  
the race to the White House this year ...

S ▼

Date, Results and Presidential ...

idential candidates, Results, Predictions, Opinion Poll  
idential Election on ...

election > us... ▼

mp falsely claims he won Georgia

elections 2020. US President Donald Trump gestures

---

# Data Collection: Nobias

As there is an absence of a concrete metric to determine left vs right leaning, Nobias follows the methodology of Matthew Gentzkow and Jesse Shapiro published in Gentzkow et al with a few exceptions.

- The 2005 congressional speeches is taken to identify phrases. Relevant phrases used in relevant times are added to this list such as “Obamacare”. Nobias uses 1318 such phrases as a benchmark.
  - Nobias then uses these 1318 phrases to identify the leaning of a news source by taking the ratio of democrat phrases matched in the article to that of the total phrases (Democrat + Republican).
  - If more than 90% of articles from a source are left or right leaning, the source is identified as either left or right.
-

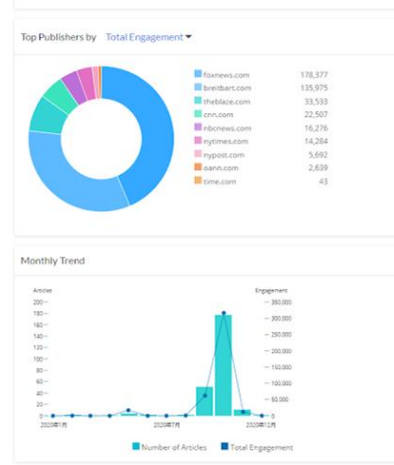
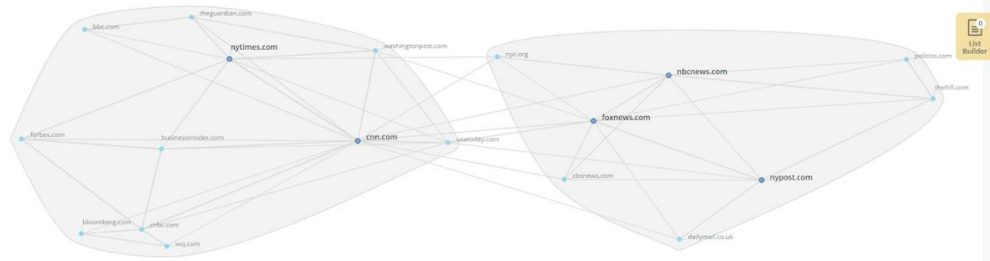
---

# Data Collection: Nobias - Selenium

- Nobias offers its data foremost via web extension
  - I resorted to finding the same info via <https://demo.nobias.com/>
  - Unfortunately, because the site was built on React, traditional methods (request) to web scrape HTML would not work (the HTML scraped would show default React code only)
  - Selenium allows the possibility to actually run a browser in the background to scrape the political lean results
-

# Data Collection: Alexa

1. Content Exploration
2. SEO ( Search Engine Optimization)
3. Audience
4. Competitors Analysis



---

# Data Collection

- News sources considered for this study: I used Alexa to find the most engaging articles from the following sources: CNN, NYTimes, FoxNews, Breitbart, NYPost, TheBlaze, NBCNews, ABCNews, Oann, and Time.

\* Engagement is measured by total posts from Twitter and Reddit

- Web Scraping: I used BeautifulSoup to scrape a total of 14,708 news articles from the above channels.

---



---

# Data Preparation

Removed punctuations, numbers, stop words. Vader works best with exclamation marks, caps, émoticons, etc. so, they were not removed.



Tokenized the text by splitting it into individual words.



Applied Vader for sentiment Analysis.

---

---

# Sentiment Analysis using Vader

Vader (Valence Aware Dictionary and Sentiment Reasoner) produces positive, neutral and negative scores which represent the proportion of the text that fall into each of the categories. It also produces compound score which is a normalized sum of all lexicon ratings of the article words. Compound evaluates the overall tone of the text and ranges from -1 (negative) to 1 (positive), with values between -0.05 and 0.05 classified as neutral.

---

---

# Execution

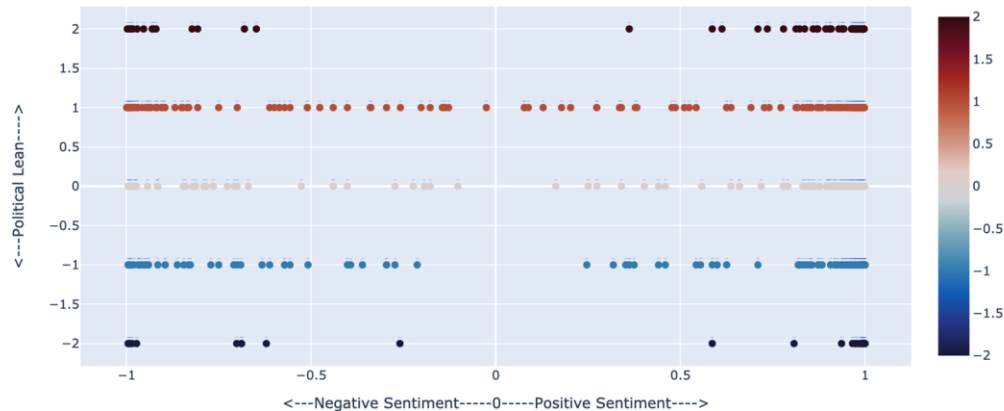
- I sadly ran into the problem that the Nobias demo site does not return results for all articles, whereas the extension does
  - To workaround this, I manually collected the result for the top 50 articles from each topic
  - The assumption is that those below the top 50 would not have as much of an impact in politics anyways
  - Once I have the sentiment results and political leans from our articles, I used plotly and Tableau to visualize the results
-

---

# Most Important Findings: Political Lean and Compound Sentiment

The more polar the political lean, the more extreme the sentiments become.

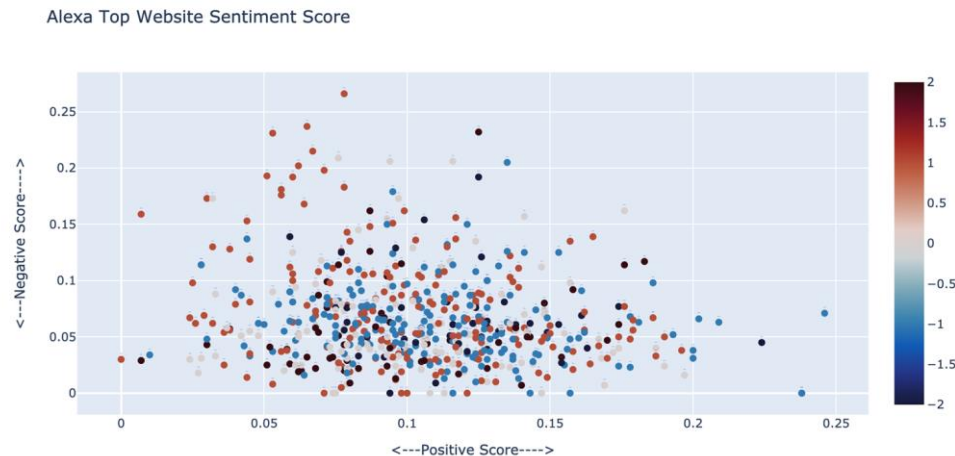
Alexa Top Website Sentiment Score



---

# Most Important Findings: Political Lean and Sentiments

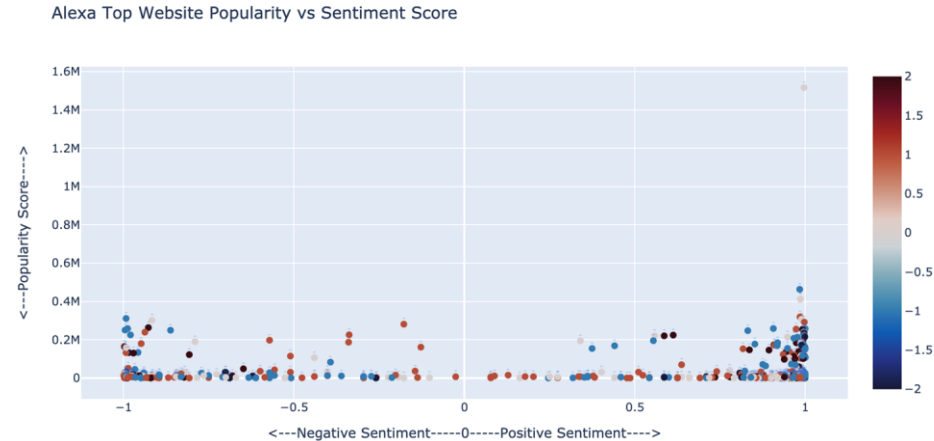
More high negative and low positive sentiment scoring articles are conservative, while more high positive and low negative sentiment scoring articles are liberal.



---

## Most Important Findings: Political Lean, Compound Sentiment, and Engagement/Popularity

Articles with more polarized sentiments tend to have greater popularity. Liberal articles tend to have more popularity for the topics that I explored.

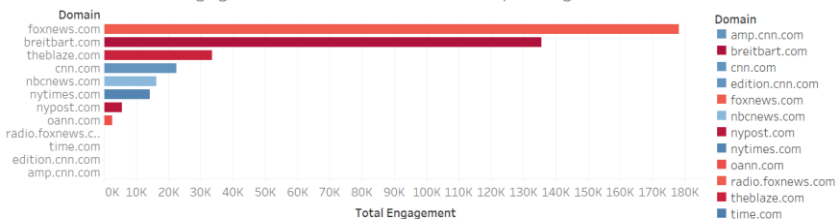


# Most Important Findings: Political Lean, Engagement/Popularity by Source and Topic

A	B	C
topic	Weighted political lean mean by number of engagemt	Column1
amy barrett	0.268136275	
georgia senate	0.176476197	
court packing	0.895181927	
biden	0.288079417	
blm	0.579386952	
build the wall	-0.576560805	
concede	0.036812265	
trump	-0.281464577	
stimulus package	-0.677015311	
green new deal	0.108189224	
election fraud	0.120745267	
covid vaccine	-0.769828214	

# Most Important Findings: Political Lean, Engagement/Popularity by Source and Topic

News source total engagement distribution for "court packing"



Sum of Total Engagement for each Domain. Color shows details about Domain. The data is filtered on Topic, which keeps court packing. The view is filtered on Domain, which excludes blogs.nytimes.com, messaging-custom-newsletters.nytimes.com and money.cnn.com.



---

---

# Q & A

Thank you!

---