News Media Bias

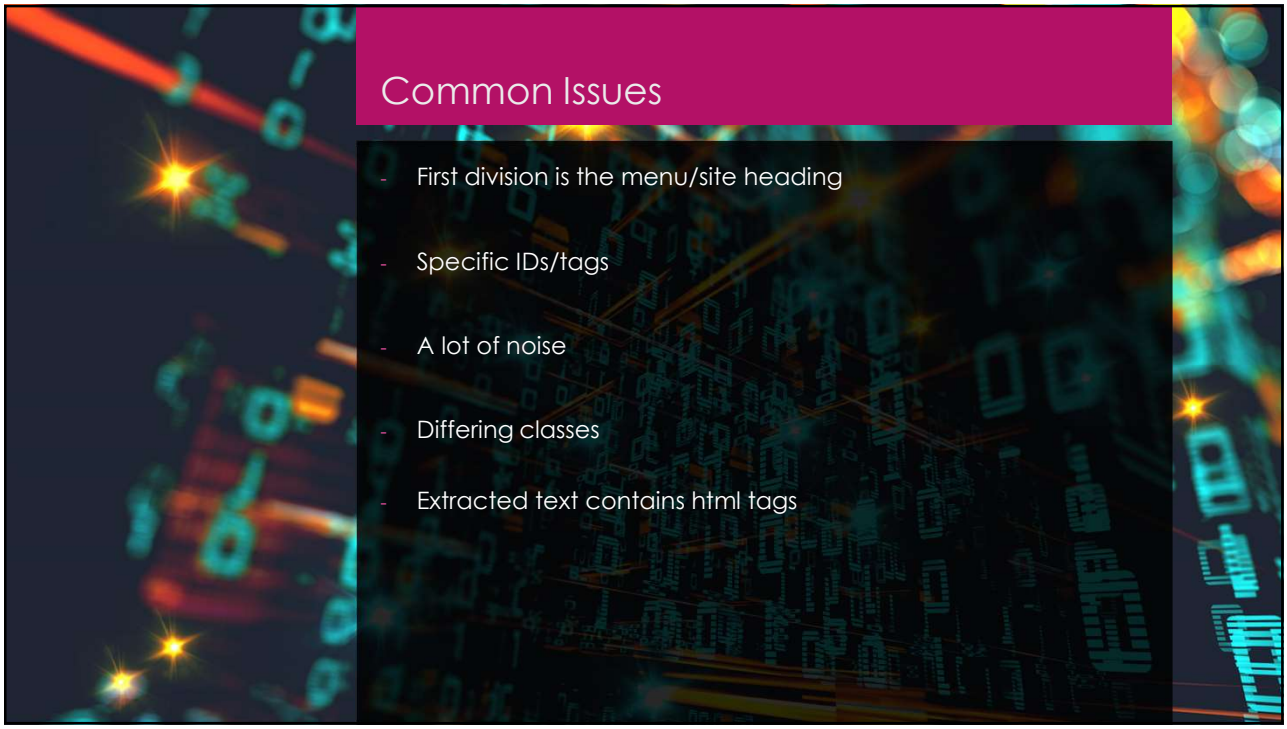
BRIAN PAN

1

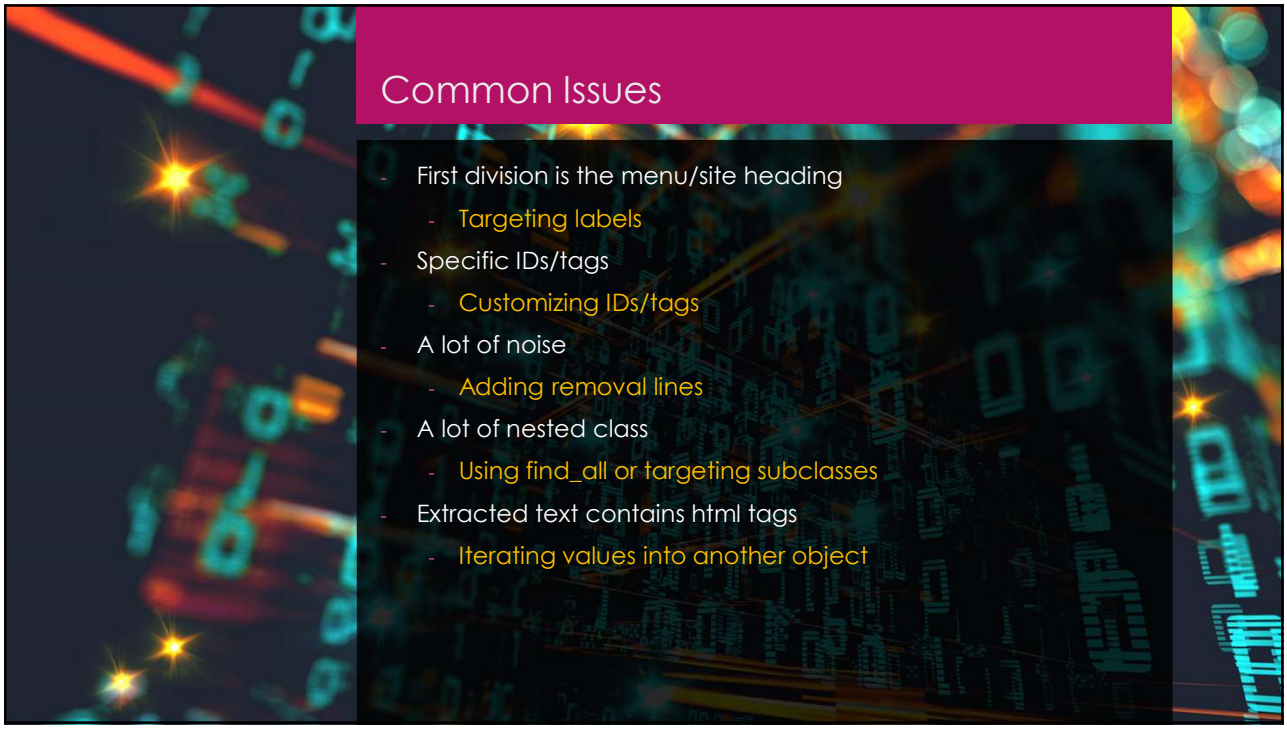
Outline

- Scraping/Coding Issues
- Left-leaning, Right-leaning
- Moderate
- Overall trends
- Findings

2


Common Issues
First division is the menu/site heading
Specific IDs/tags
A lot of noise
Differing classes
Extracted text contains html tags

3


Common Issues
First division is the menu/site heading
Targeting labels
Specific IDs/tags
Customizing IDs/tags
A lot of noise
Adding removal lines
A lot of nested class
Using find_all or targeting subclasses
Extracted text contains html tags
Iterating values into another object

4

Common Issues

```
# Extract article
article_text = ""
article_body = soup.find_all('div', class_='duet--article--article-body-component')
```

```
# Cleaning Texts
cleaned_text = re.sub(r'\bpic\.twitter\S*', '', text)
cleaned_text = re.sub(r'🔥', '', cleaned_text)
cleaned_text = re.sub(r'🚨', '', cleaned_text)
cleaned_text = re.sub(r'🤯', '', cleaned_text)
cleaned_text = re.sub(r'border@BretBaier', '', cleaned_text)
```

```
if image_url:
    # If srcset exists and contains multiple URLs, split by space and take the first one
    image_url = image_url.split()[0]
    print(image_url)  # Output the image URL
```

5

Resources

```
soup.find(class_="athing")
```

```
<tr class="athing" id="40866374">
<td align="right" class="title" valign="top"><span class="rank">1.</span></td> <td class="votelinks" valign="top"><center>
<a href="vote?id=40866374&amp;how=up&amp;goto=news" id="up_40866374"><div class="votearrow" title="upvote"></div></a></cen
ter></td><td class="title"><span class="titleline"><a href="https://www.mattkeeter.com/blog/2023-01-25-branch/">Do not tau
nt happy fun branch predictor (2023)</a><span class="sitebit comhead"> (<a href="from?site=mattkeeter.com"><span class="si
testr">mattkeeter.com</span></a>)</span></span></td></tr>
```

You can also find all elements with a specific class using:

```
soup.find_all(class_="athing")
```

```
[<tr class="athing" id="40866374">
 <td align="right" class="title" valign="top"><span class="rank">1.</span></td> <td class="votelinks" valign="top"><center><a
href="vote?id=40866374&amp;how=up&amp;goto=news" id="up_40866374"><div class="votearrow" title="upvote"></div></a></center></t
d><td class="title"><span class="titleline"><a href="https://www.mattkeeter.com/blog/2023-01-25-branch/">Do not taunt happy fu
n branch predictor (2023)</a><span class="sitebit comhead"> (<a href="from?site=mattkeeter.com"><span class="sitestr">mattkeet
er.com</span></a>)</span></span></td></tr>,
 <tr class="athing" id="40866155">
 <td align="right" class="title" valign="top"><span class="rank">2.</span></td> <td class="votelinks" valign="top"><center><a
href="vote?id=40866155&amp;how=up&amp;goto=news" id="up_40866155"><div class="votearrow" title="upvote"></div></a></center></t
d><td class="title"><span class="titleline"><a href="https://www.noemamag.com/living-in-a-lucid-dream/">Living in a Lucid Drea
m</a><span class="sitebit comhead"> (<a href="from?site=noemamag.com"><span class="sitestr">noemamag.com</span></a>)</span></s
pan></td></tr>,
```

6

Resources

SQLPad Questions AI Tools Blog Courses Pricing

```
# Perform the request
response = requests.get(url)

# Parse the content
soup = BeautifulSoup(response.text, 'html.parser')

# Find all job listings on the page
job_listings = soup.find_all('div', class_='job-listing')

# Loop through and extract details
jobs = []
for job in job_listings:
    title = job.find('h2').text
    company = job.find('div', class_='company').text
    location = job.find('div', class_='location').text
    jobs.append({'title': title, 'company': company, 'location': l

# Display the job listings
for job in jobs:
    print(f"Job Title: {job['title']}, Company: {job['company']},
```

7

Scraping

	jupyter Personal Code	Simplescraper	Instant Data Scraper	Browse AI
Setup Requirement	High	Low	Low	Medium – Sign up required
Cost	Free	Free	Free	Free until trial limit reached
Format Extraction	Any	CSV/ JSON	CSV / XLSX	CSV/JSON
Customizable	Yes	Yes	No	Yes
Guides Available	No	Yes	Yes	Yes
Support available	No	Yes	No	Yes
Unique Features	Customizable per scraping needs	Has recipes (formatting for specific scrapes)	Has option to scrape infinite scroll	Robots – custom recipes tailored for specified sites
Easy to scrape multiple	Yes with recipe	Yes with recipe	No	Yes with robot
Data Formatting	None	Yes	Yes	Yes

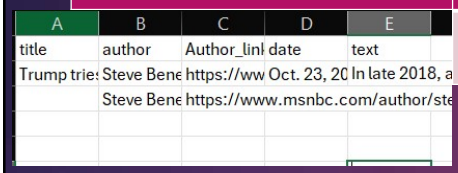
8

Scraping

	jupyter Personal Code	Simplescraper	Instant Data Scraper	Browse AI
Strengths	Free, easy to format, easy to import, easy to customize.	Free, easy to scrape, easy to create recipes	Free, easy to scrape, can do infinite-scroll websites	Professional support available. Easy to scrape and create 'robot' recipes.
Limits	Requires customization for each website, labor intensive,	A lot of formatting issues when importing data. Most pressing issue is the inability to encode utf-8.	Cannot choose which content to scrape. Data is disorganized.	Primarily a paid platform, but free trial is available until threshold data is reached.

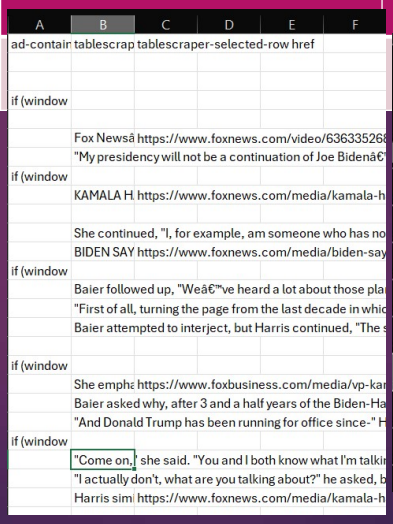
9

Simple Scraper



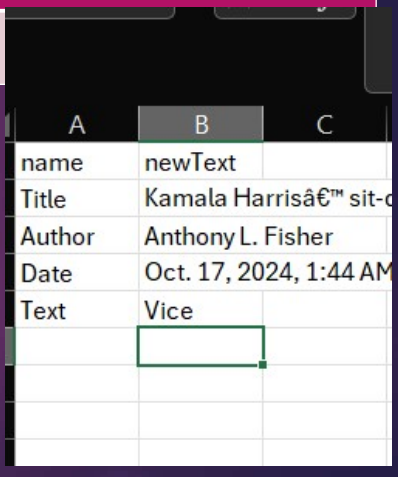
title	author	Author_link	date	text
Trump tries	Steve Bene	https://ww	Oct. 23, 20	In late 2018, a
	Steve Bene	https://www.msnbc.com/author/ste		

Instant Data Scraper



ad-contain	tablescrap	tablescraper-selected-row	href
if (window			
	Fox Newsâ	https://www.foxnews.com/video/63633526	
	"My presidency will not be a continuation of Joe Bidenâ€		
if (window			
	KAMALA H	https://www.foxnews.com/media/kamala-h	
	She continued, "I, for example, am someone who has no		
	BIDEN SAY	https://www.foxnews.com/media/biden-say	
if (window			
	Baier followed up, "Weâ€™ve heard a lot about those pla		
	"First of all, turning the page from the last decade in whic		
	Baier attempted to interject, but Harris continued, "The s		
if (window			
	She emphs	https://www.foxbusiness.com/media/vp-ka	
	Baier asked why, after 3 and a half years of the Biden-Ha		
	"And Donald Trump has been running for office since-" H		
if (window			
	"Come on," she said. "You and I both know what I'm talki		
	"I actually don't, what are you talking about?" he asked, b		
	Harris sim	https://www.foxnews.com/media/kamala-h	

Browse AI



name	newText
Title	Kamala Harrisâ€™ sit-c
Author	Anthony L. Fisher
Date	Oct. 17, 2024, 1:44 AM
Text	Vice

10

News Media Bias

BRIAN PAN

11

HOW CAN VOTERS EFFECTIVELY NAVIGATE THE PLETHORA OF NEWS TO MAKE GOOD DECISIONS AT THE BALLOT?

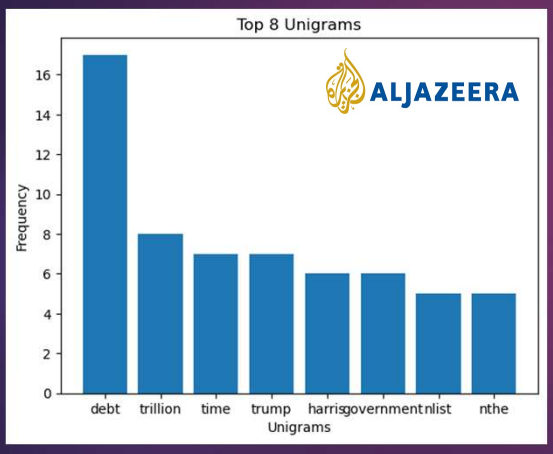
12

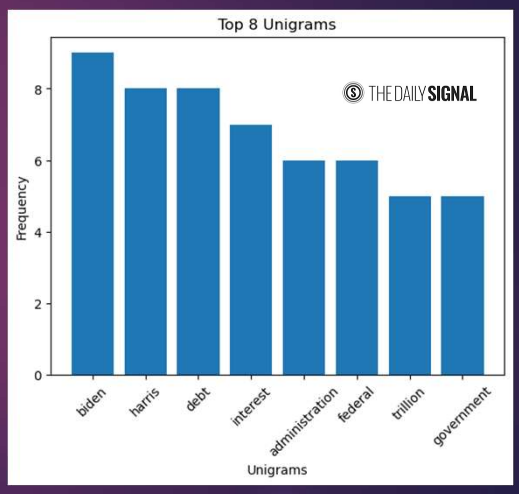
Comparing Articles

	Left-Leaning	Right-Leaning
Debt	ALJAZEERA	THE DAILY SIGNAL
Border	MSNBC	FOX NEWS
Guns	Vox	Townhall

13

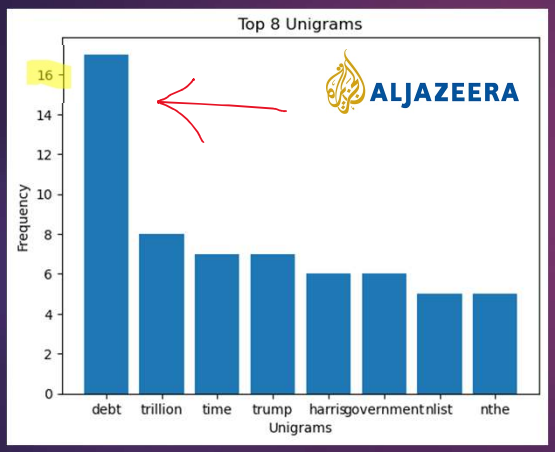
Comparing Debt

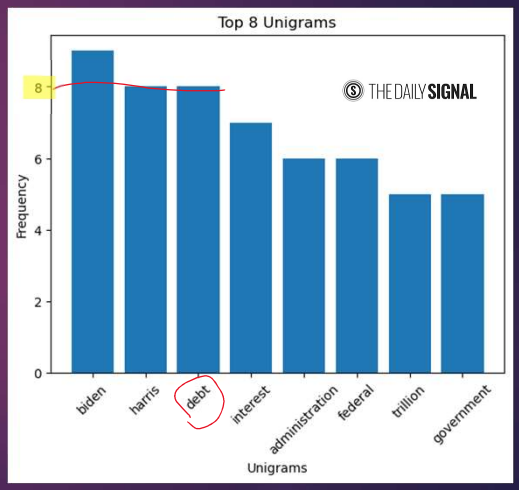




14

Comparing Debt

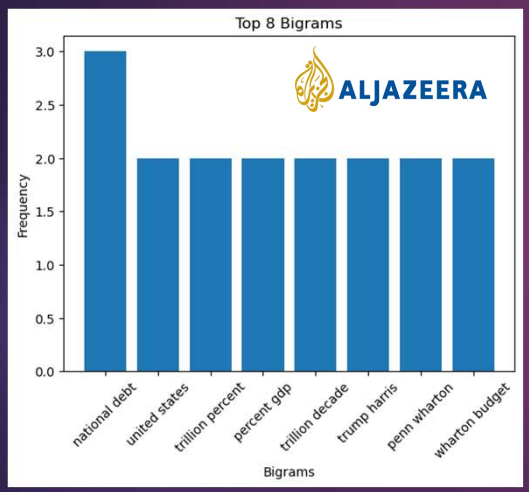


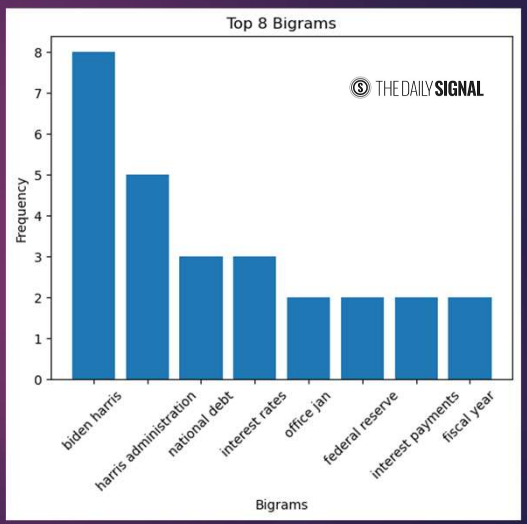


Noting the difference in most frequent unigrams here, The contrast is significant.

15

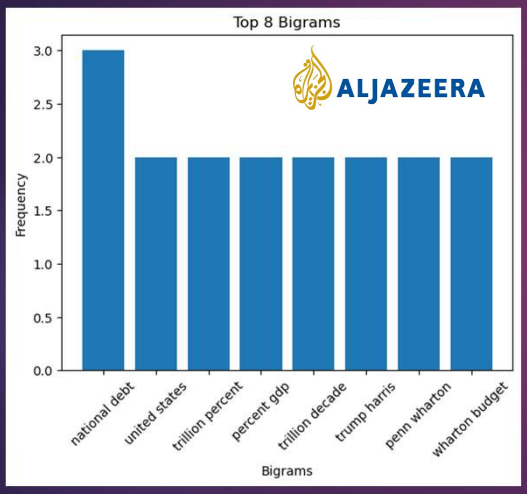
Comparing Debt

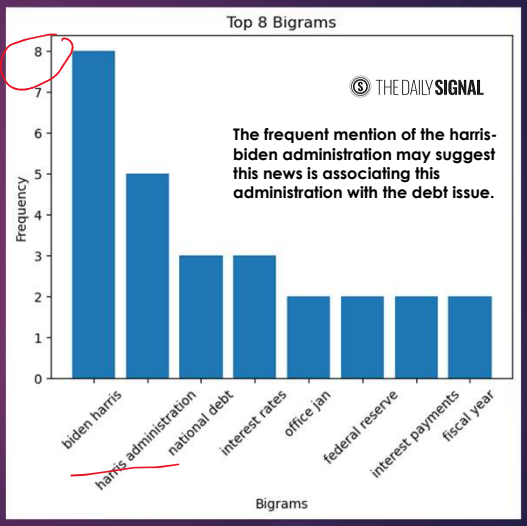




16

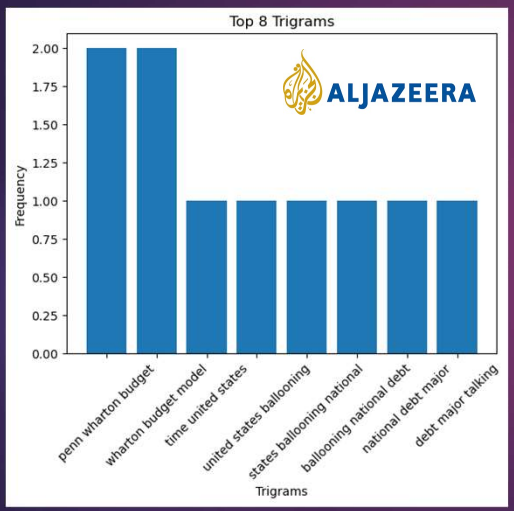
Comparing Debt

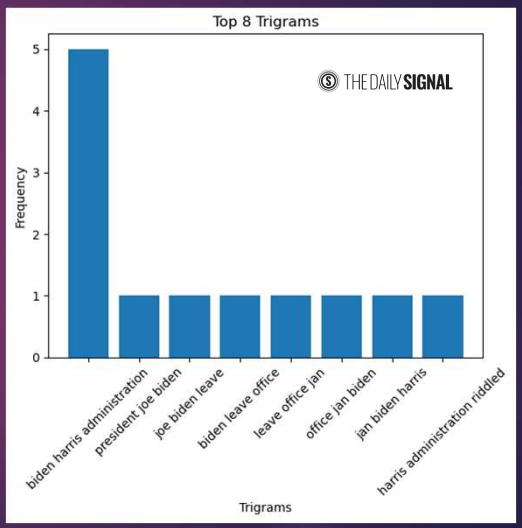




17

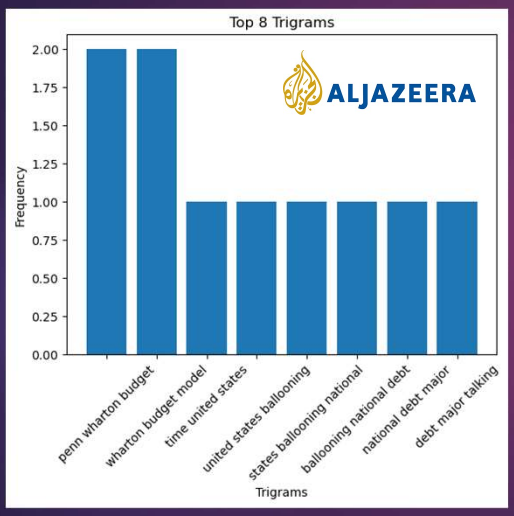
Comparing Debt

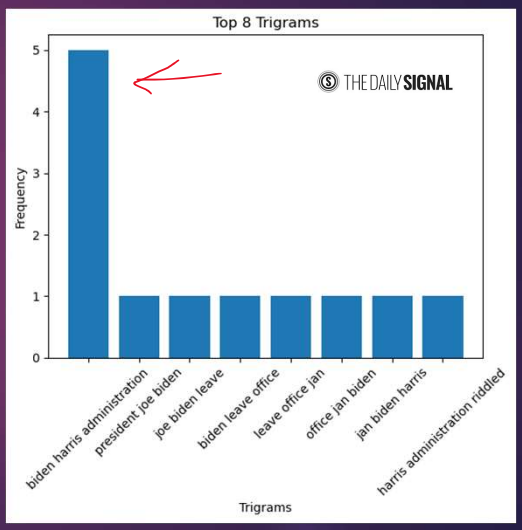




18

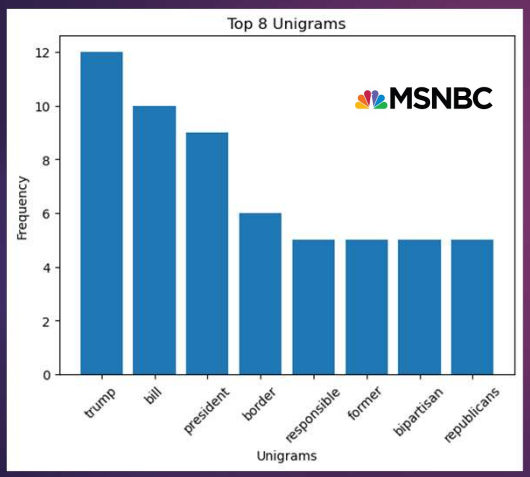
Comparing Debt

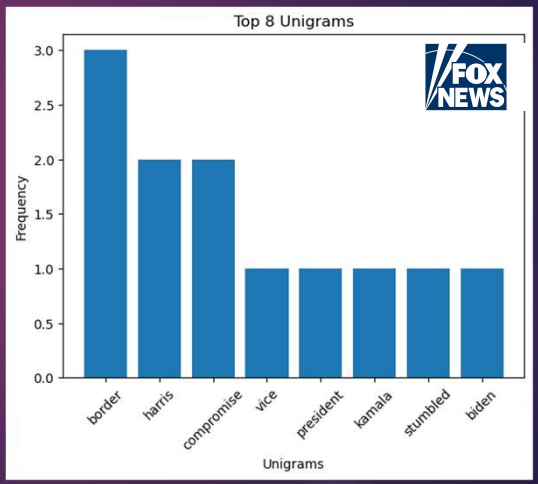




19

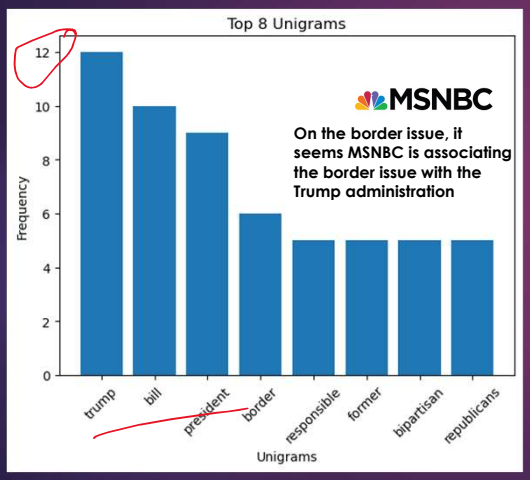
Comparing Border

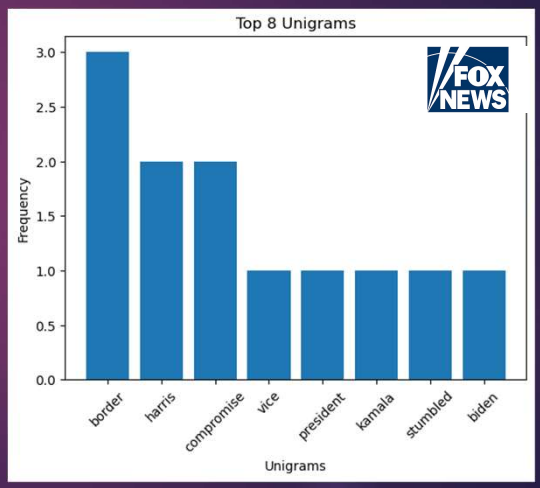




20

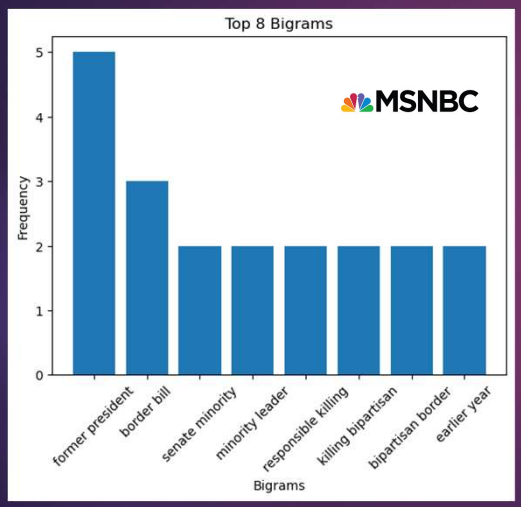
Comparing Border

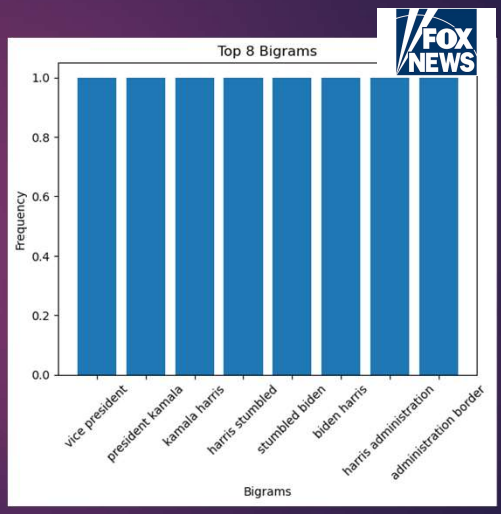




21

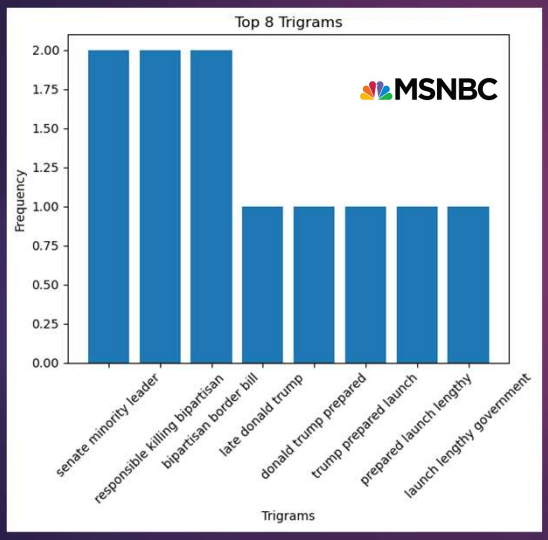
Comparing Border

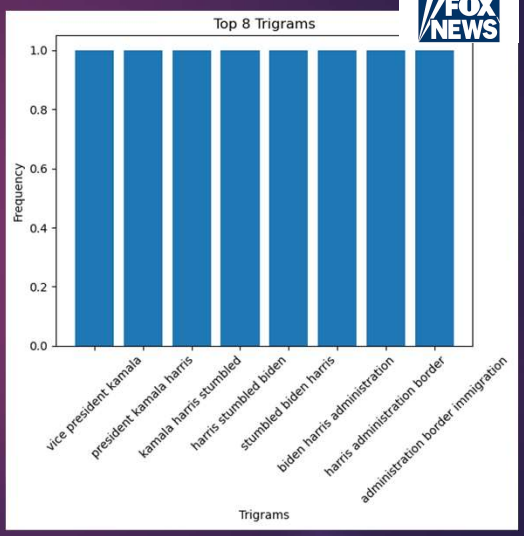




22

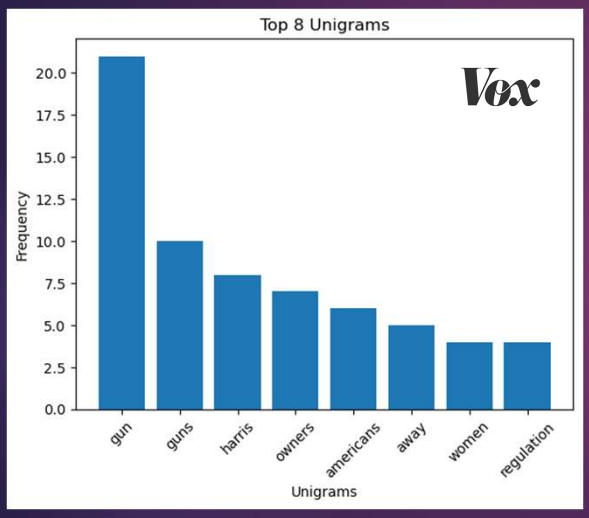
Comparing Border

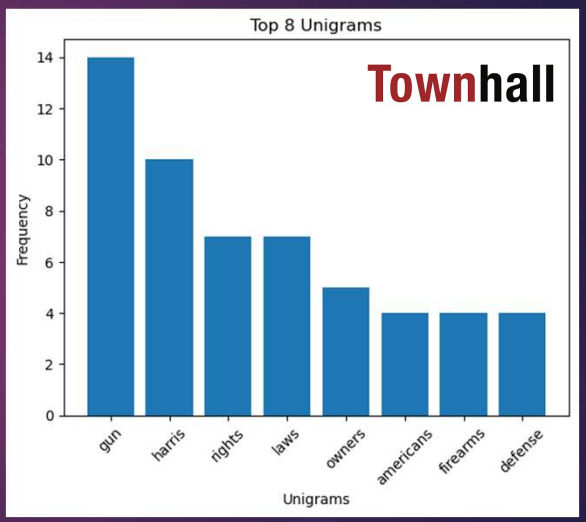




23

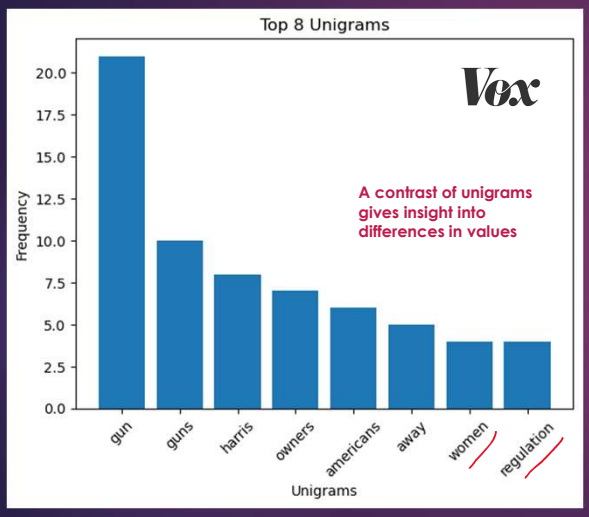
Comparing Guns

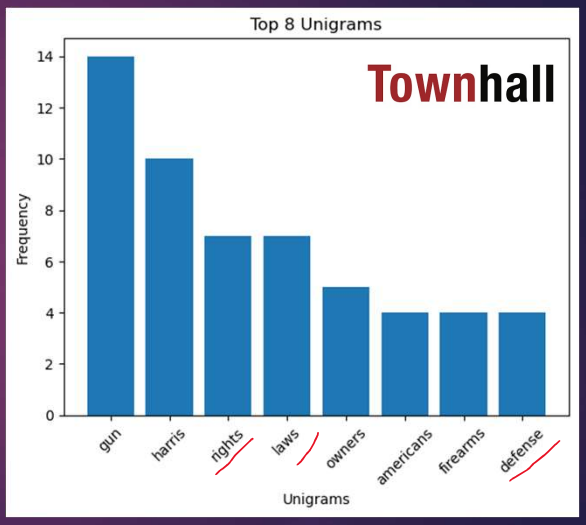




24

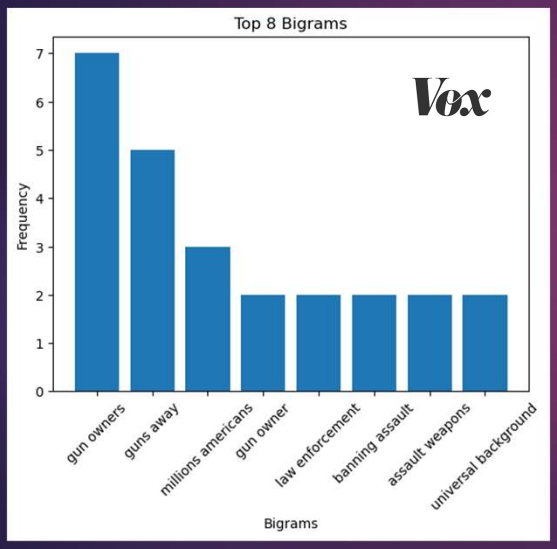
Comparing Guns

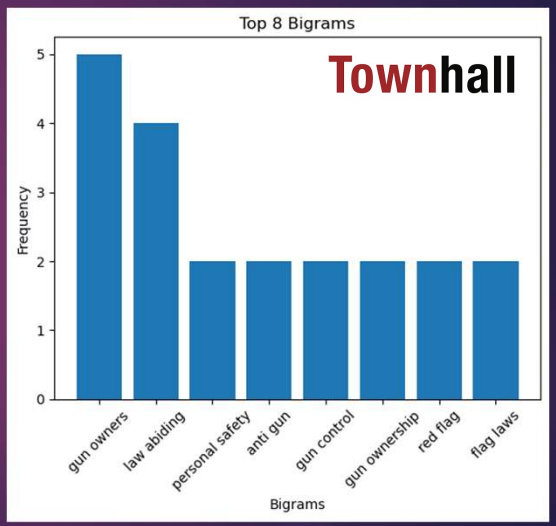




25

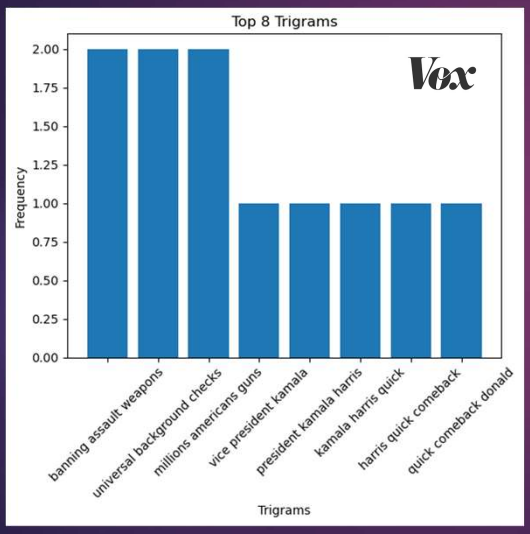
Comparing Guns

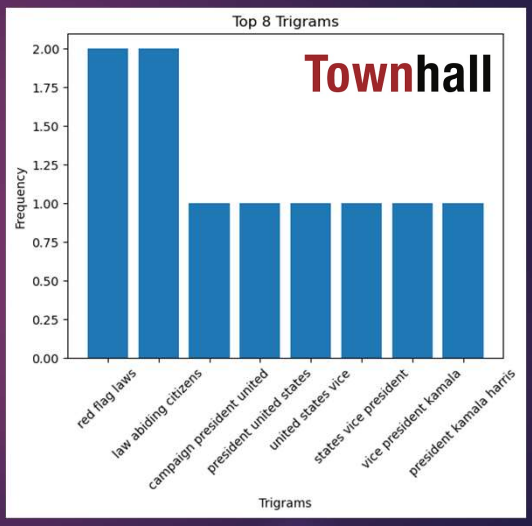




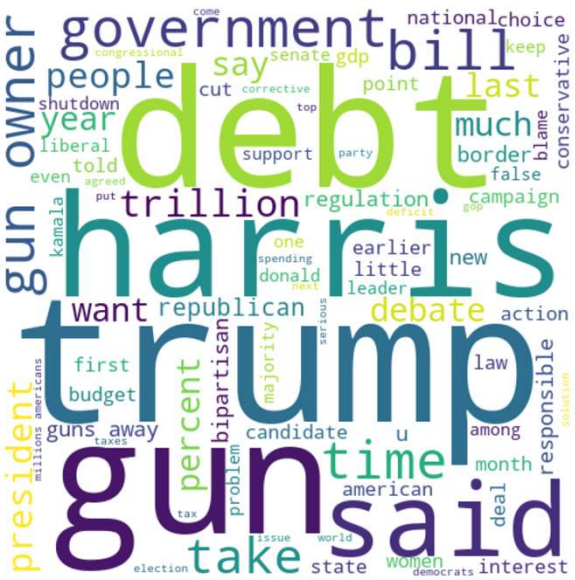
26

Comparing Guns





27



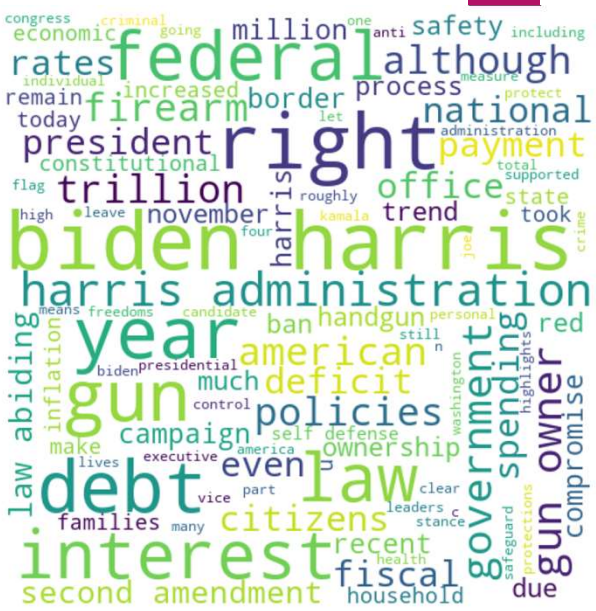
Liberal

- Several words are more frequent than others
- Certain words are emphasized a lot more than others
- Key issues around ‘debt’,’harris’,’trump’ and ‘guns’

28

Conservative

- Few words are pronounced
- Word disparity is not as pronounced, words are mentioned at roughly same frequency
- Little mention of Trump, strong focus on Biden-Harris



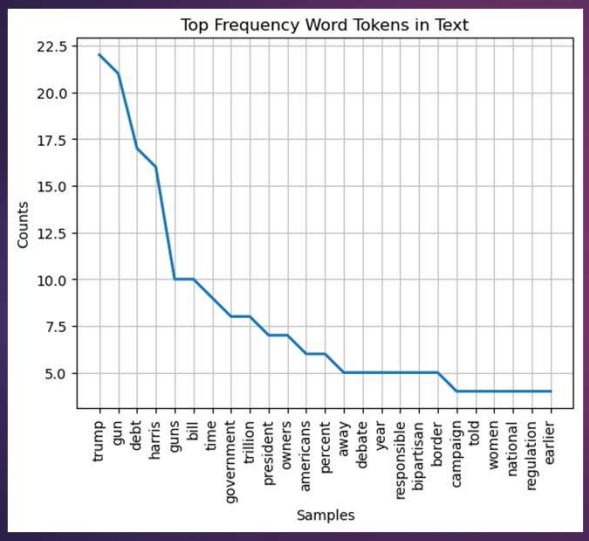
29

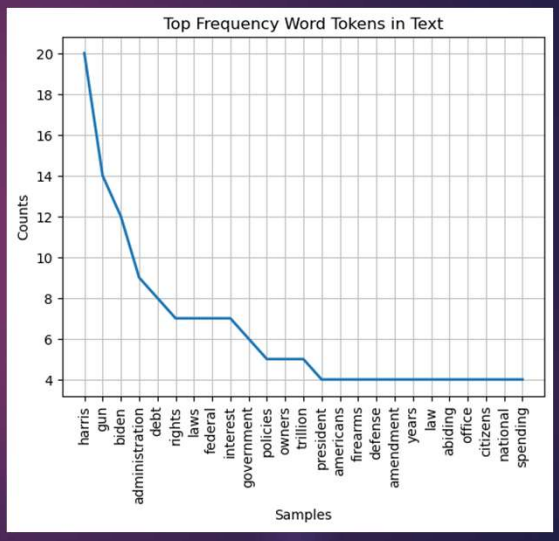
Conservative

- Few words are pronounced
- Word disparity is not as pronounced, words are mentioned at roughly same frequency
- Little mention of Trump, strong focus on Biden-Harris

30

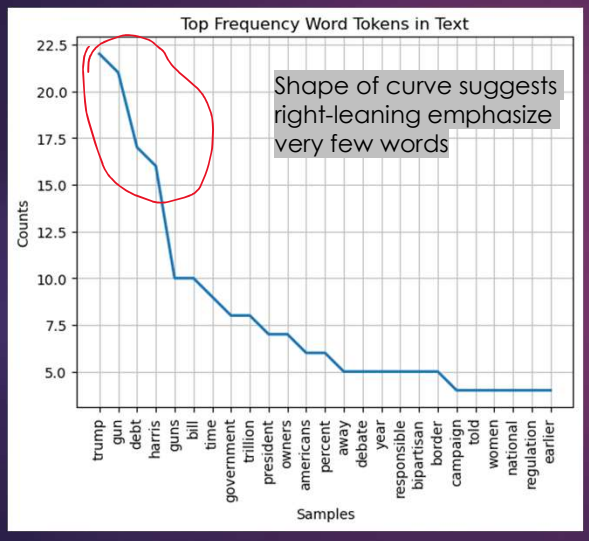
Word Frequencies

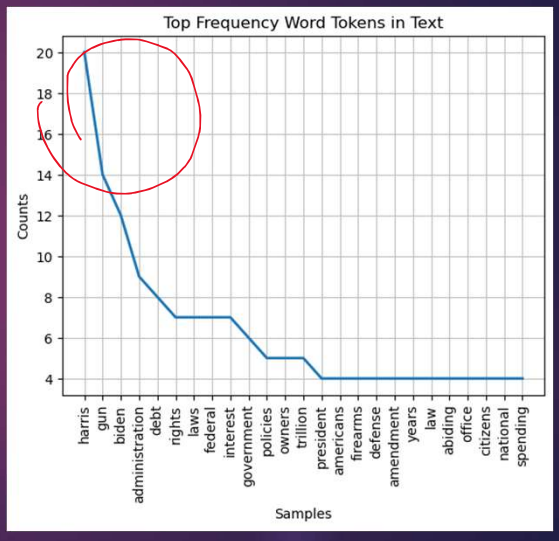




31

Word Frequencies





32

Findings

	Left-Leaning	Right-Leaning	Both
Actors	Significant mention of Trump	Partial against Biden-Harris, little mention of Trump	Emphasize political regimes
Debt	More emphasis on spending	More emphasis on Biden-Harris	Feature some mention of parties
Border	Significant emphasis on Trump	Some emphasis on politics, but no mention of Trump	Feature some mention of parties
Guns	Emphasis on regulation and women	Emphasis on laws, rights, and defense	Mention 'Americans'
Word Association	A couple words are regularly emphasized	Very few words stand out	Heavily stresses certain word/rhetoric

33

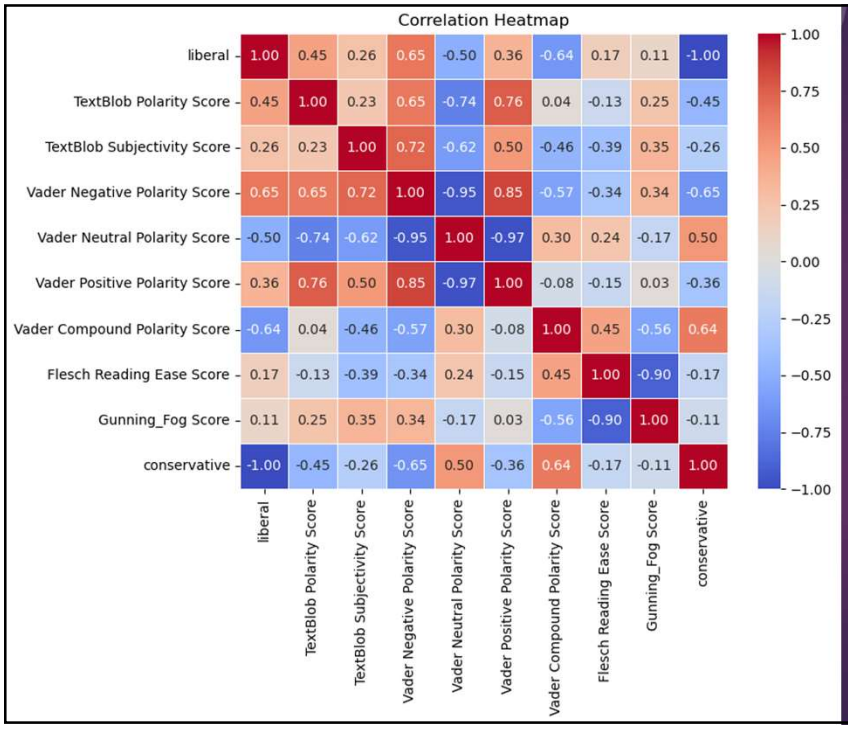
Methodology

	Unigrams	Bigrams	Trigrams	WordCloud
Contextual Understanding	Effective in identifying values/frame	Moderate	Not effective. Not many trigrams present	Effective in identifying values/frame
Complexity	Simple – easy to understand	Simple – easy to understand	Simple – easy to understand	Complex – difficult to interpret
Information Density	Moderate	Moderate	Light – little information attributed to few trigrams	Dense
Advantage	Simple, easy to interpret, highlights details quickly	Same as unigram; highlights compound words	Effective in finding phrases or emphasized clusters	Birds-eye view of data. Easy to spot textual emphasis and values
Disadvantage	Strong bias for adjectives (e.g. 'gun' owners)	Leaves out important words that are not bigrams	Very sparse findings due to infrequency of trigrams	Difficult to extract further information

34

Text Analysis

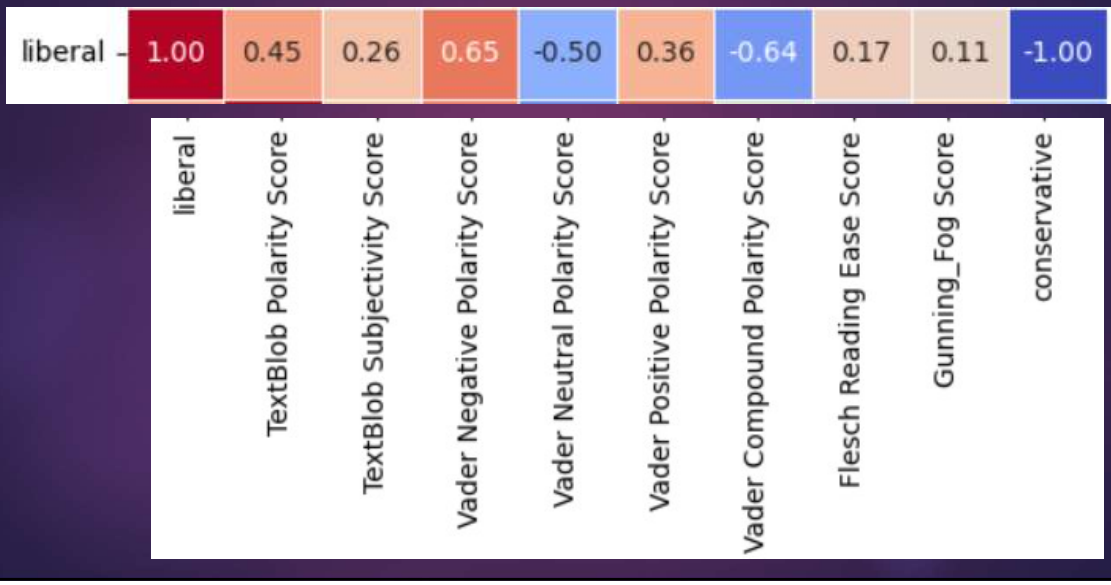
35



Correlation

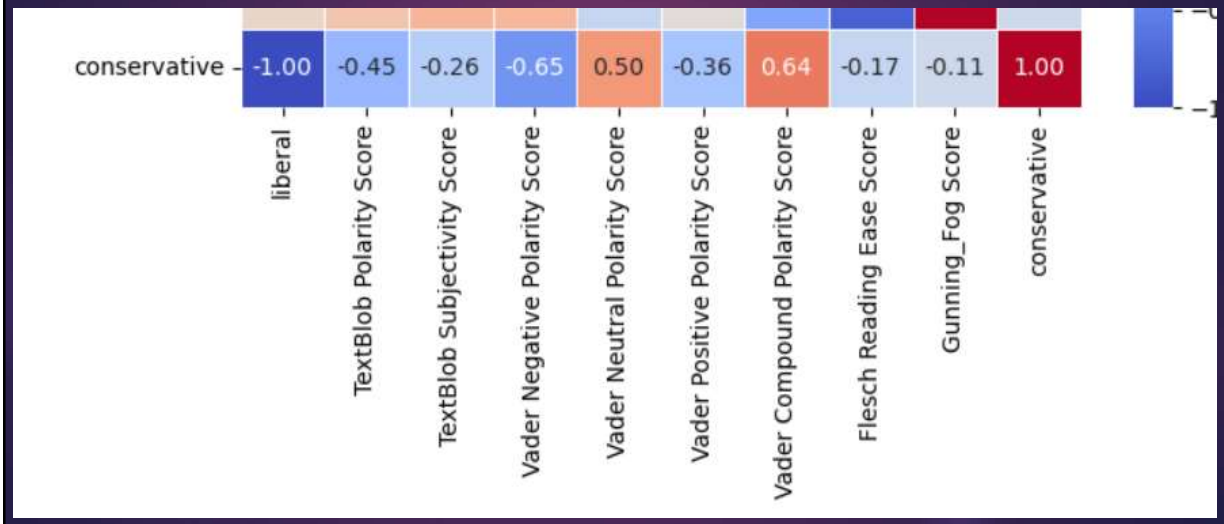
36

Correlation



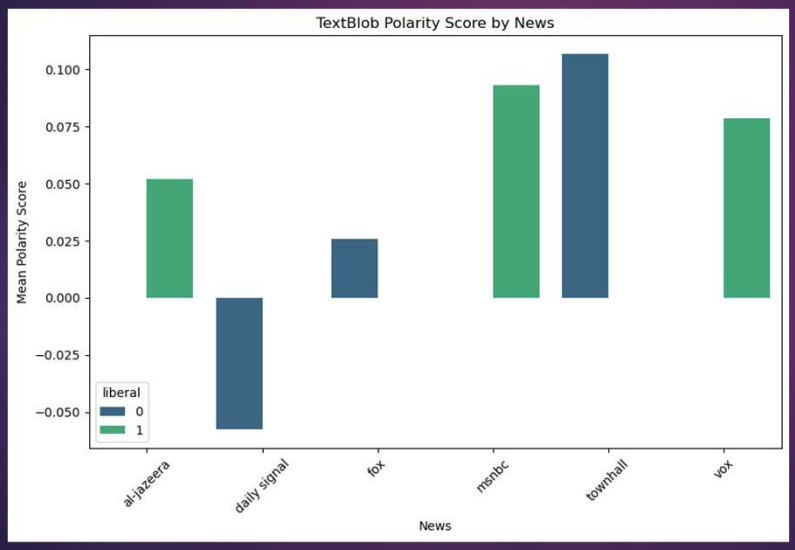
37

Correlation



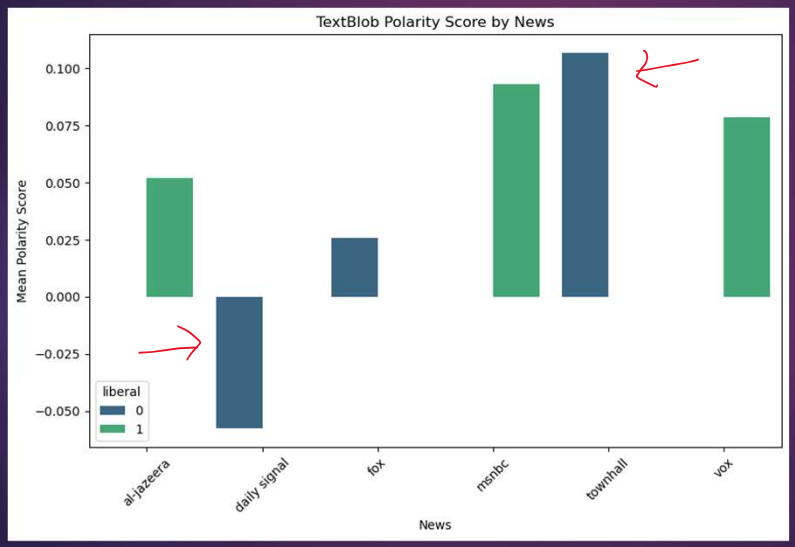
38

News Comparison



39

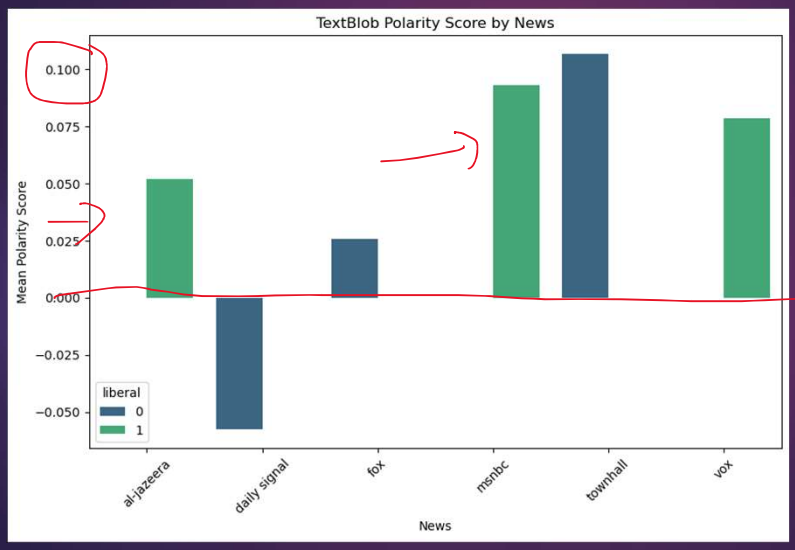
News Comparison



Conservative leaning has most positive sentiment and most negative sentiment

40

News Comparison

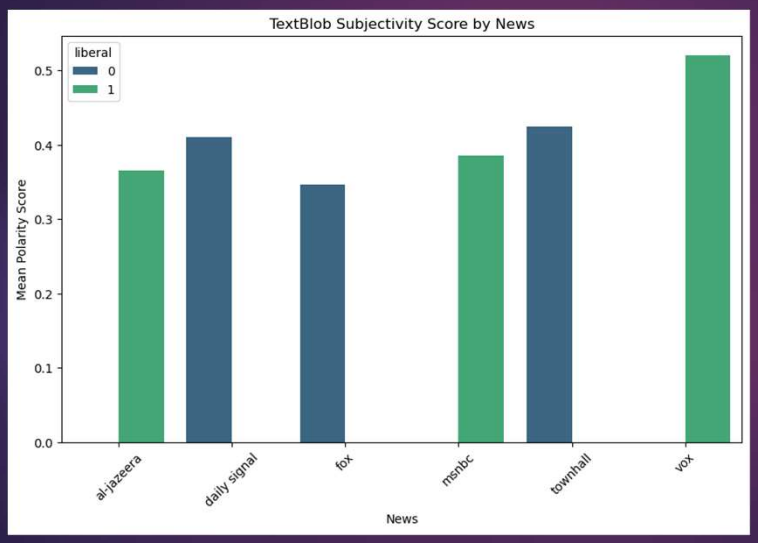


Left-Leaning news are generally more positive leaning, as shown by its positive value.

Important to still note that the range of TextBlob polarity is [-1,1]. While positive trending, actually closer to neutral.

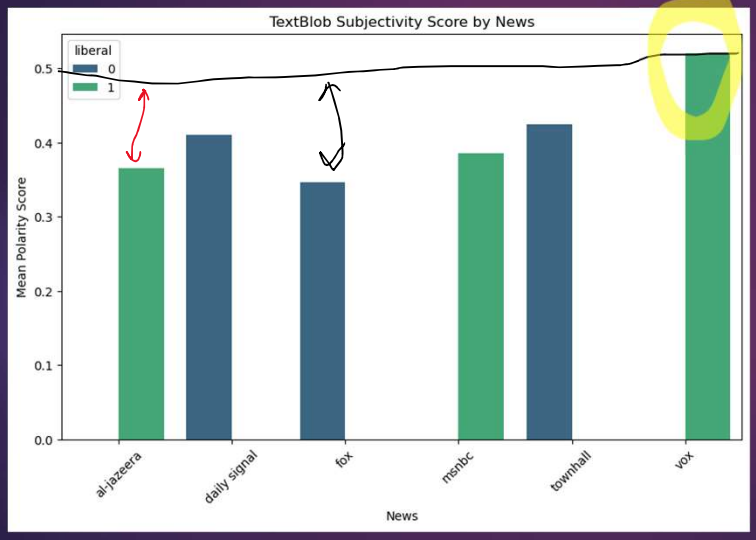
41

News Comparison



42

News Comparison



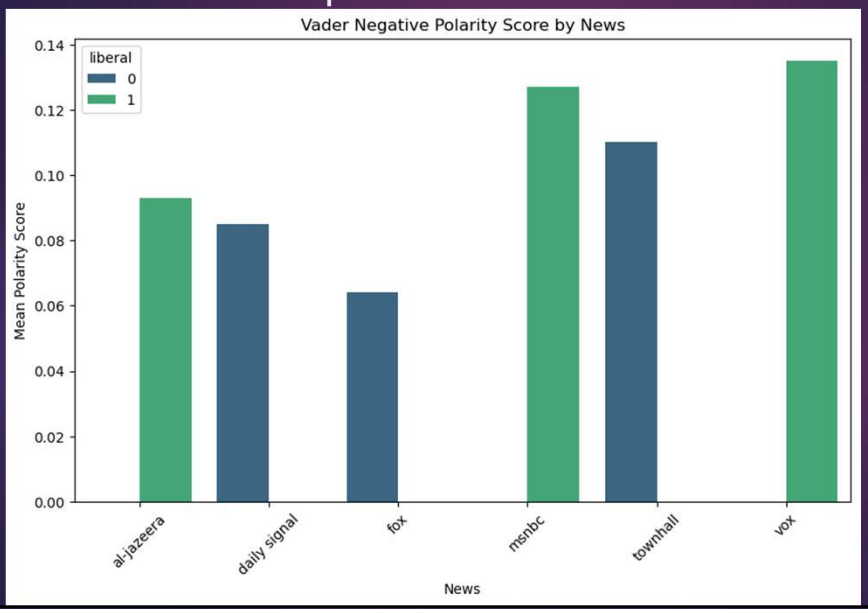
TextBlob Subjectivity Range [0,1]

Most are generally closer to 0.4, which suggests a balance between subjective and objective.

Al-Jazeera and Fox seem to be relatively more objective while Vox is the most subjective.

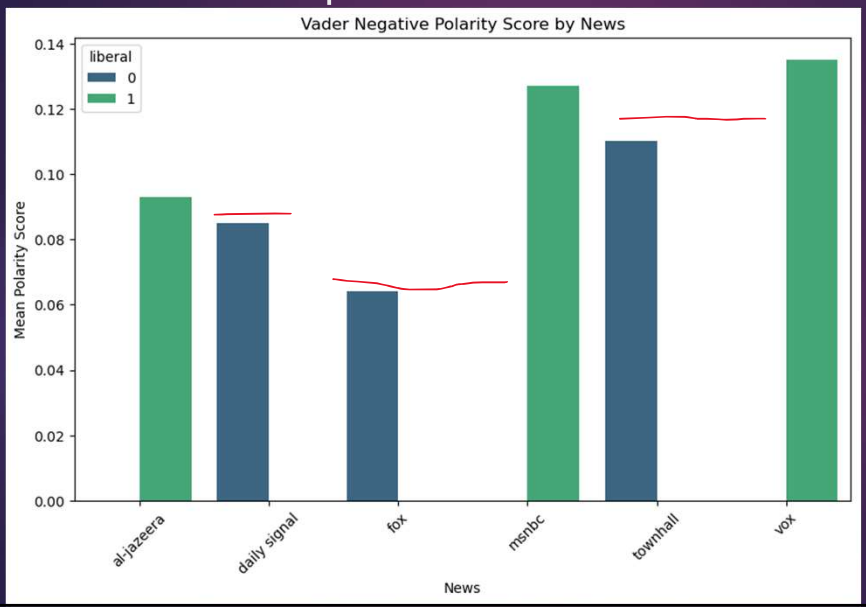
43

News Comparison



44

News Comparison



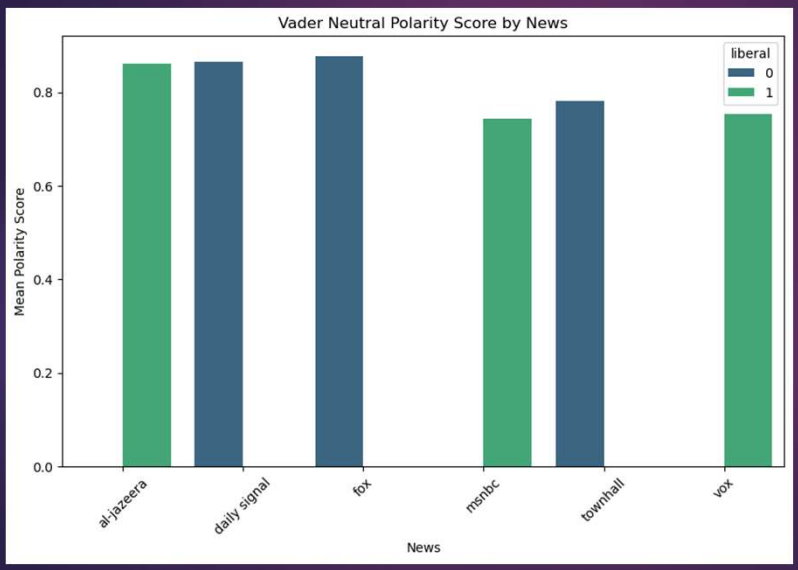
Comparatively, left-leaning outlets are more negative based on vader negative polarity.

You can see from the chart that the leftist media are comparatively higher.

While Al-Jazeera is less negative that townhall, it still more negative than fox or daily signal.

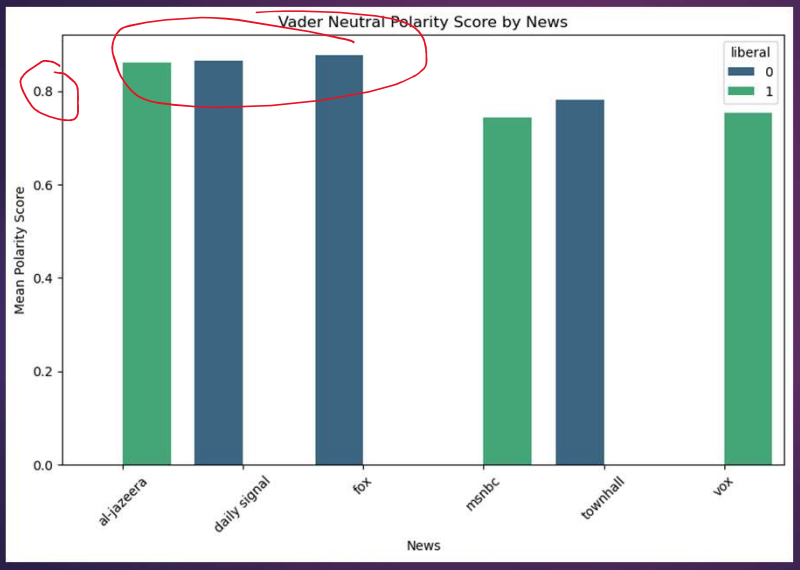
45

News Comparison



46

News Comparison

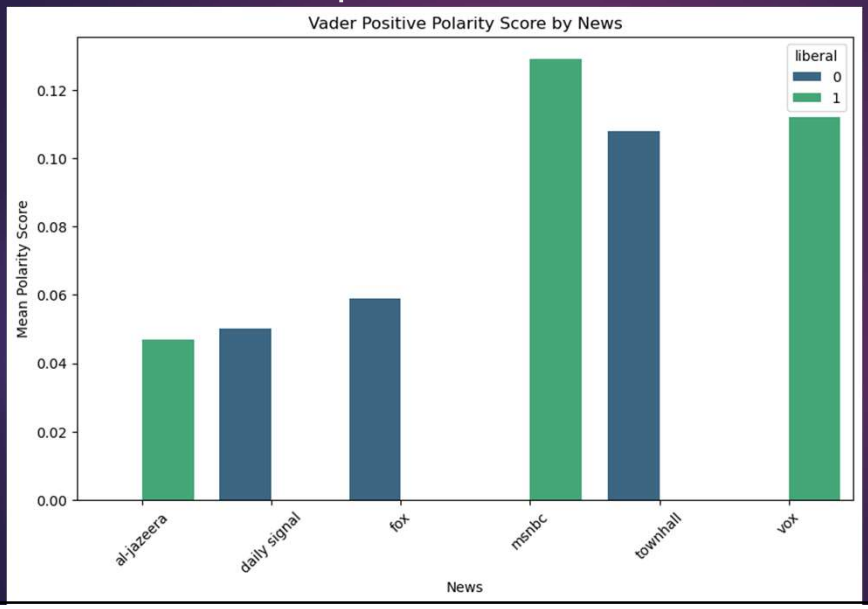


Vader Neutral Polarity range [0,1] with 1 being most neutral.

Overall, most news sources are quite neutral, with the Al Jazeera, Daily Signal, and Fox being close to 1.

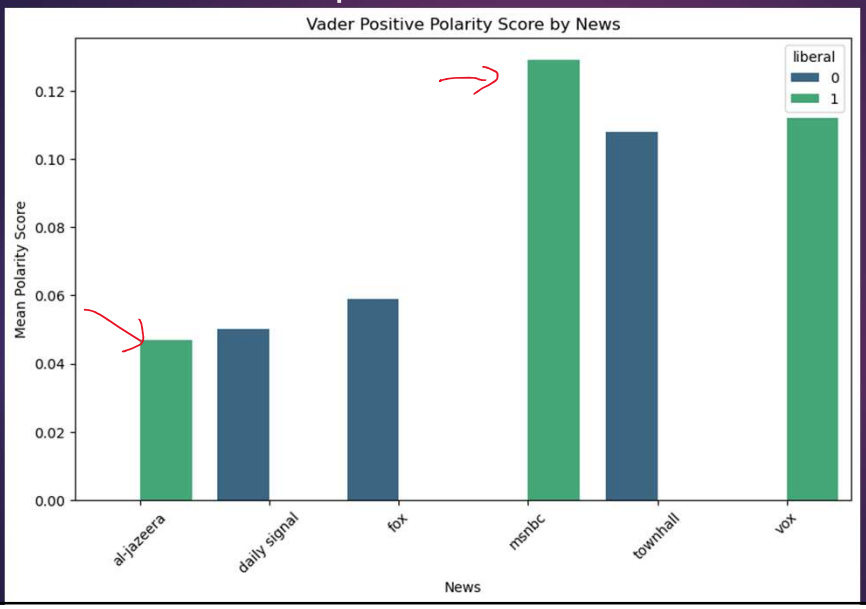
47

News Comparison



48

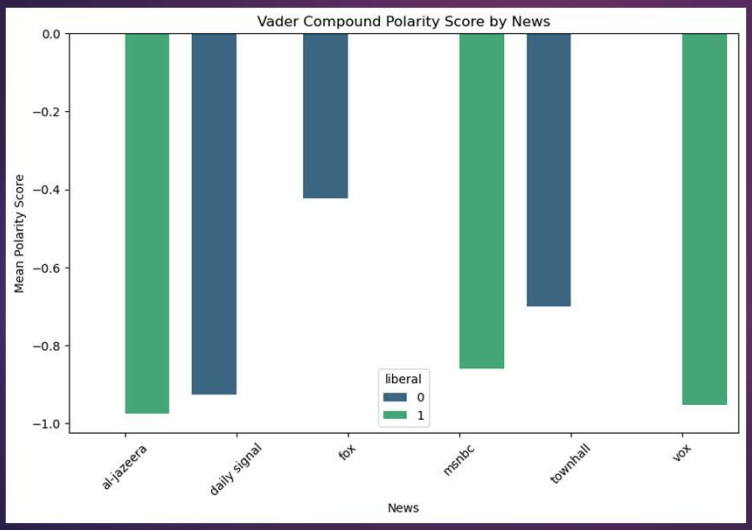
News Comparison



MSNBC is most positive polarity while Al-Jazeera is least positive.

49

News Comparison



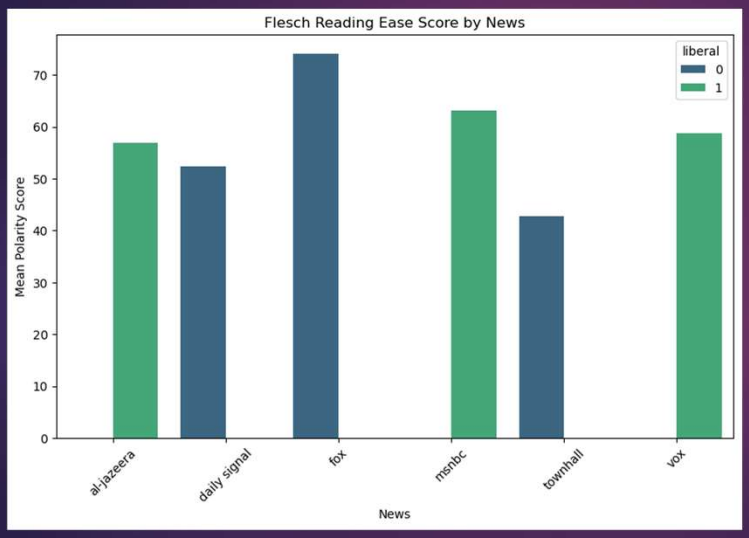
Vader Compound aggregates the previous scores into a range between [-1,1].

This aggregate score suggests that both (and potentially any) news are generally negatively.

Fox surprisingly is relatively less negative.

50

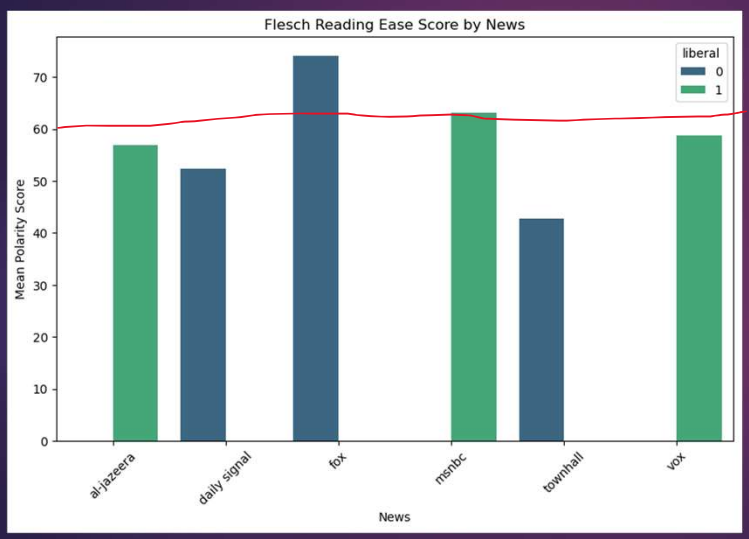
News Comparison



Score	School level (US)
100.00–90.00	5th grade
90.0–80.0	6th grade
80.0–70.0	7th grade
70.0–60.0	8th & 9th grade
60.0–50.0	10th to 12th grade
50.0–30.0	College
30.0–10.0	College graduate
10.0–0.0	Professional

51

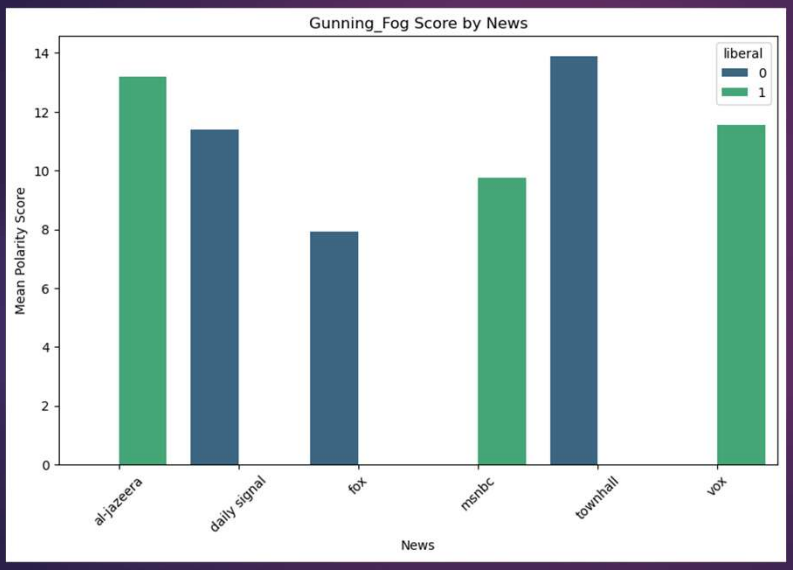
News Comparison



Score	School level (US)
100.00–90.00	5th grade
90.0–80.0	6th grade
80.0–70.0	7th grade
70.0–60.0	8th & 9th grade
60.0–50.0	10th to 12th grade
50.0–30.0	College
30.0–10.0	College graduate
10.0–0.0	Professional

52

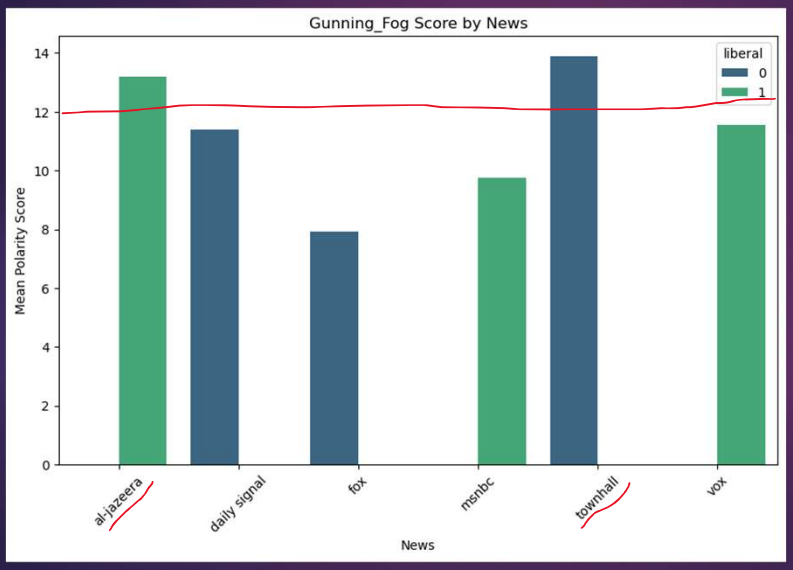
News Comparison



Fog Index	Reading level by grade
17	College graduate
16	College senior
15	College junior
14	College sophomore
13	College freshman
12	High school senior
11	High school junior
10	High school sophomore
9	High school freshman
8	Eighth grade
7	Seventh grade
6	Sixth grade

53

News Comparison



Fog Index	Reading level by grade
17	College graduate
16	College senior
15	College junior
14	College sophomore
13	College freshman
12	High school senior
11	High school junior
10	High school sophomore
9	High school freshman
8	Eighth grade
7	Seventh grade
6	Sixth grade

54

Results

	Meaning	Findings	Interpretations
TextBlob Polarity Range [-1,1]	-1 = Negative Sentiment 0 = Neutral Sentiment 1 = Positive Sentiment	Conservative have highest and lowest score. Liberals have generally positive scores. Overall, centered around neutral.	Sentiments are mostly neutral. However, this score disagrees with Vader Score, as left-leaning is negative for Vader, suggesting a different scoring method.
TextBlob Subjectivity Range[0,1]	0 = Objective 1 = Subjective	Al-Jazeera and Fox are most objective. Vox is most subjective.	Al-Jazeera and Fox are objective, but all are balanced.
Vader Negative Polarity Score Range[0,1]	0 = Least Negative 1 = Most Negative	Left-leaning media overall more negative than right-leaning.	This disagrees with textblob, which found left to be more positive.
Vader Neutral Polarity Score Range[0,1]	0 = Least Neutral 1 = Most Neural	Al-Jazeera, Daily Signal and Fox are most neutral.	This agrees with TextBlob. Al-Jazeera and Fox are neutral, objective.
Vader Positive Polarity Range[0,1	0 = Least Positive 1 = Most Positive	MSNBC most positive while Al-Jazeera is least positive.	Some disagreement with Textblob, where Al-Jazeera was positive.
Vader Compound Polarity Range[0,1	-1 = Strong Negative 0 = Neural 1 = Strong Positive	All news tend towards negative, but left-leaning has stronger stronger negative.	Disagrees with textblob considerably, which found all news slightly positive but close to neutral.
Flesch Reading Ease Score Range[0,100]	0 = Hard to Read 100 = Easy to Read	Fox is at a middle school reading level, while MSNBC is at high school.	This agrees with Gunning Fog.
Gunning Fog Index Range[0,17]	0 = No requirements 17 = Number years of education required to read	Al-Jazeera and townhall are at higher education reading level.	This agrees with Flesch Reading Ease.

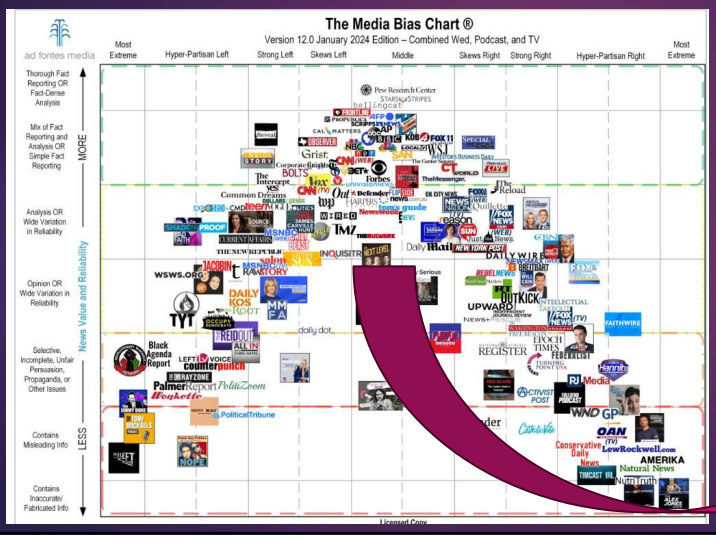
55

Findings

- Most news outlets are generally neutral and balanced between subjective and objective, but it is inconclusive if they are leaning more positive or more negative.
- Big news networks like Fox and MSNBC have a lower reading level, likely to reach popular appeal.

56

Investigating Moderates News



Selected Moderate Media Sources

SCRIPPS NEWS

DeseretNews.

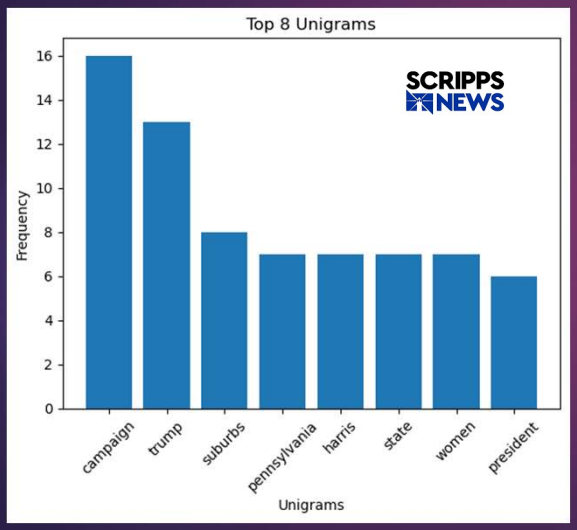
Newsweek

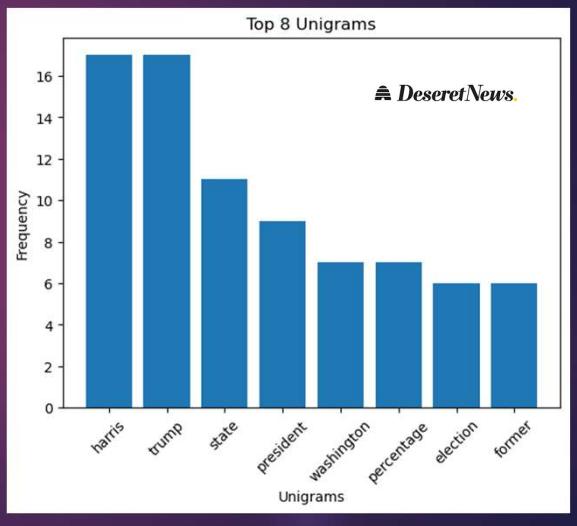
THE GLOBE AND MAIL

Chart from Ad Fontes Media

57

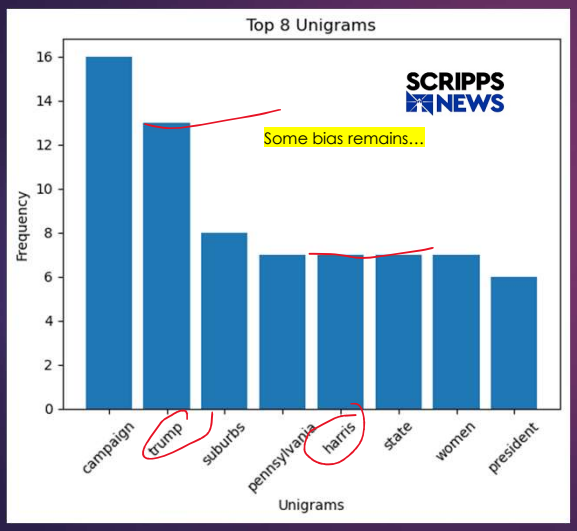
Comparing Unigrams

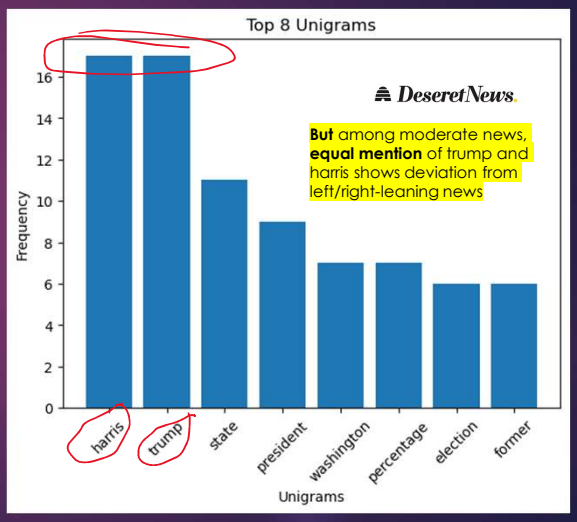




58

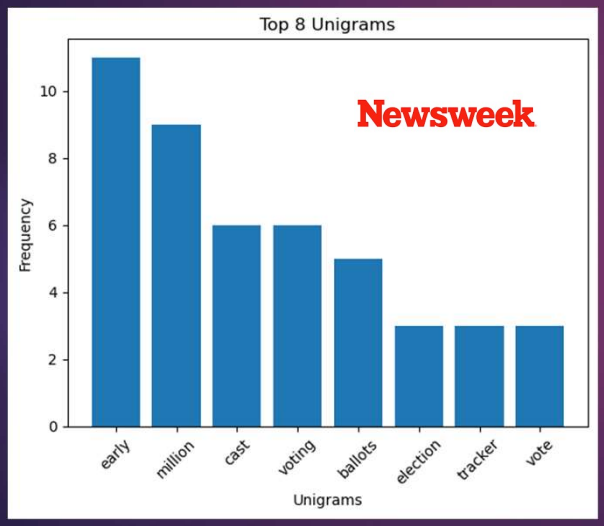
Comparing Unigrams

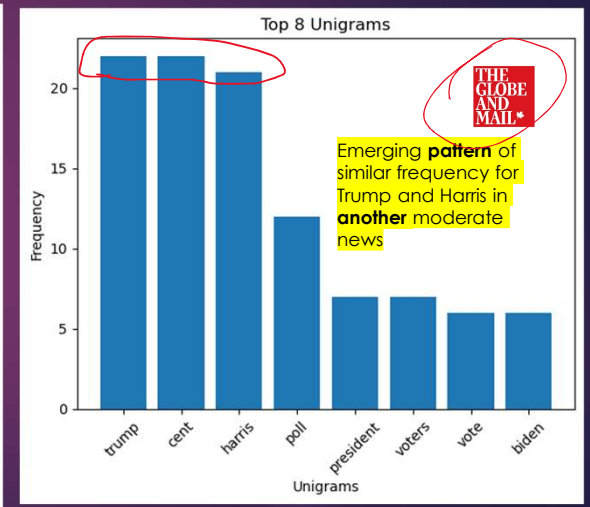




59

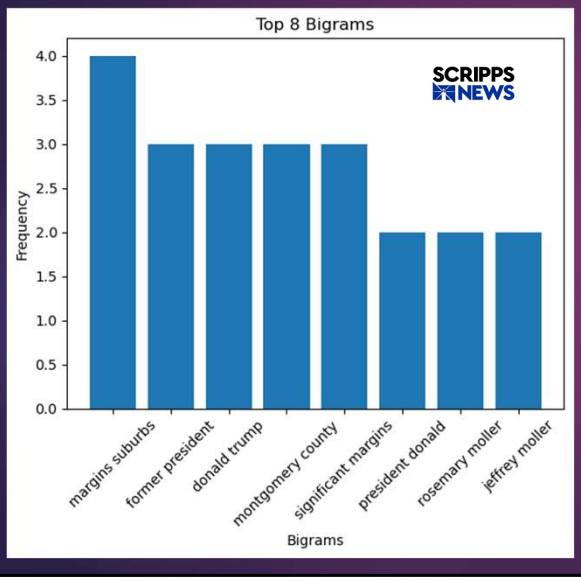
Comparing Unigrams

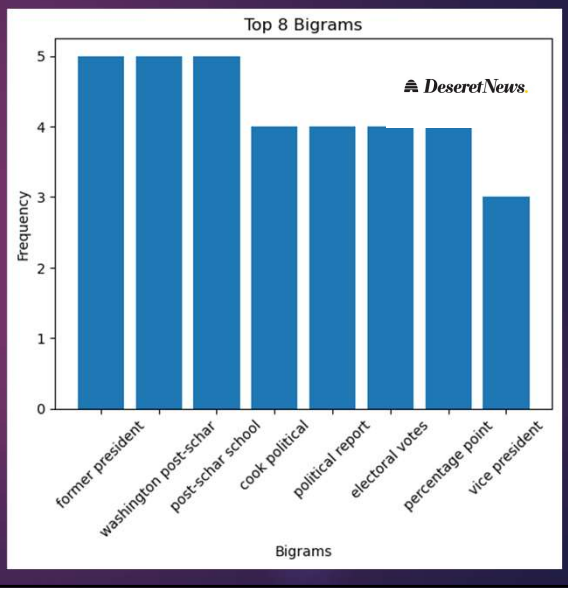




60

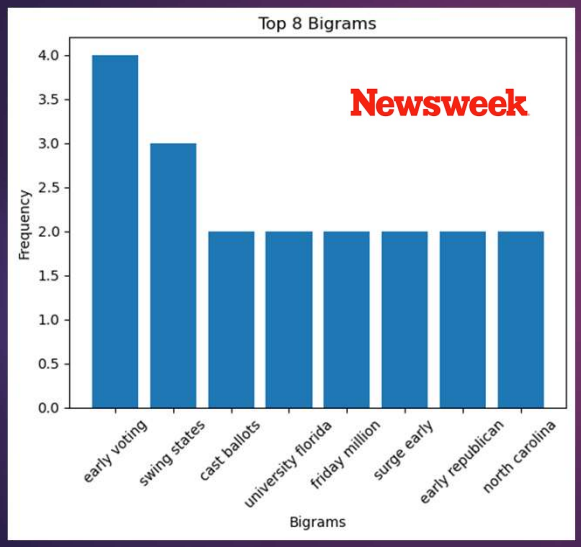
Comparing Bigrams

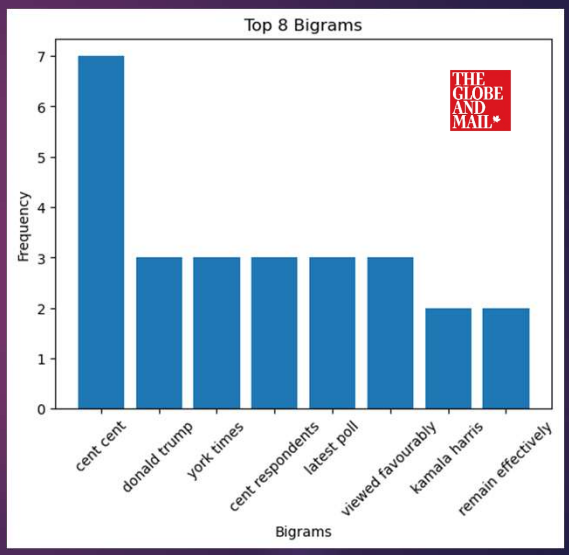




61

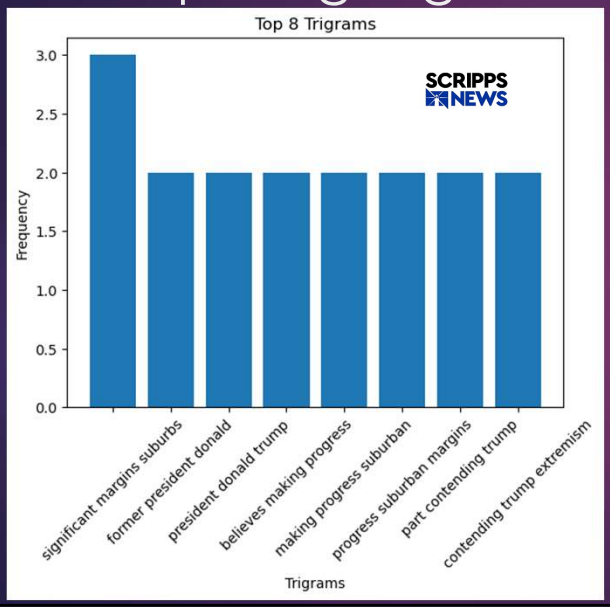
Comparing Bigrams

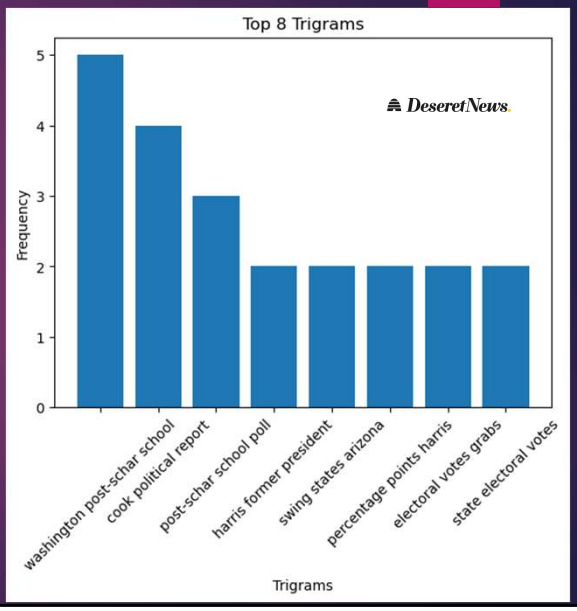




62

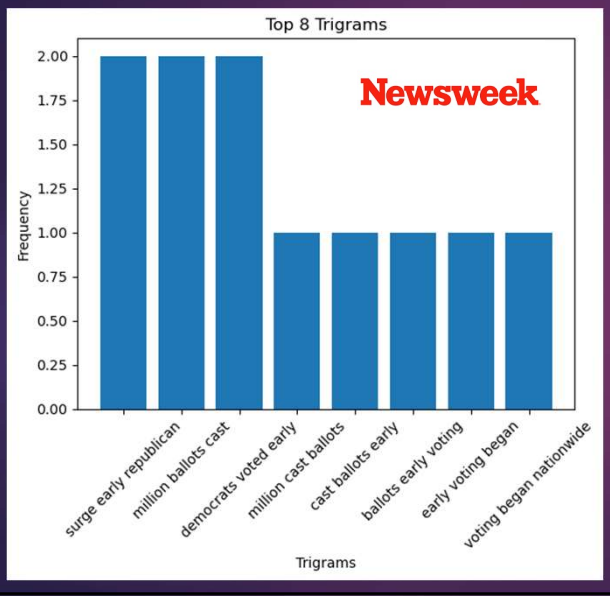
Comparing Trigrams

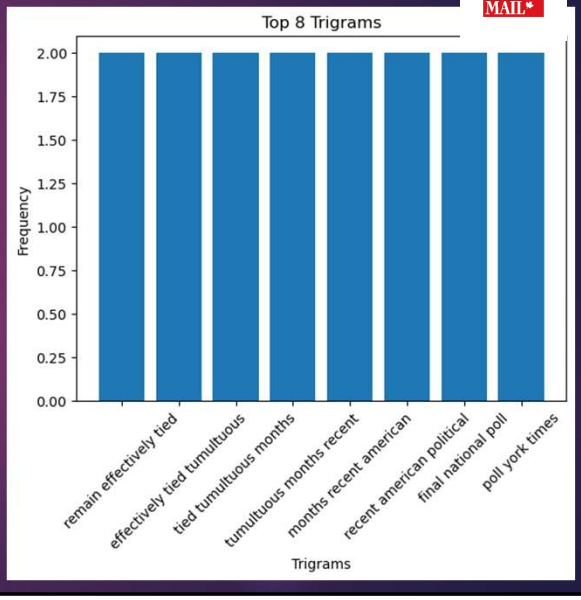




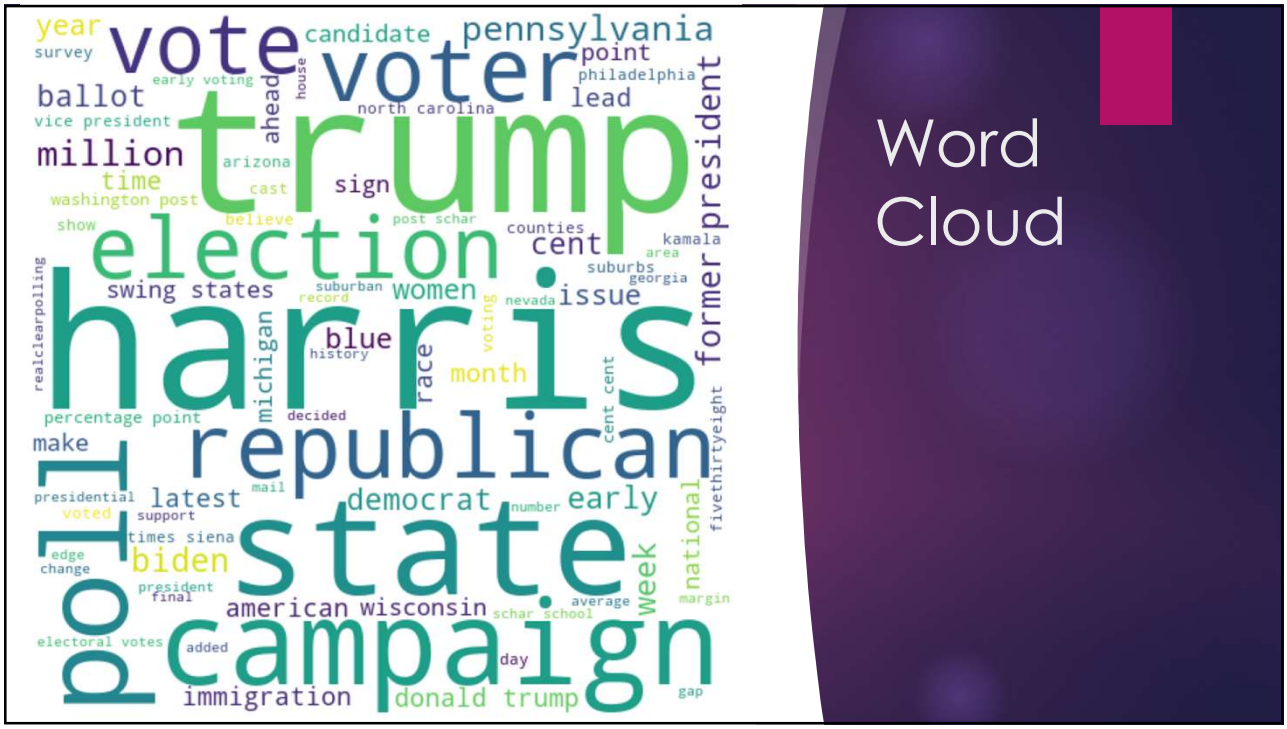
63

Comparing Trigrams



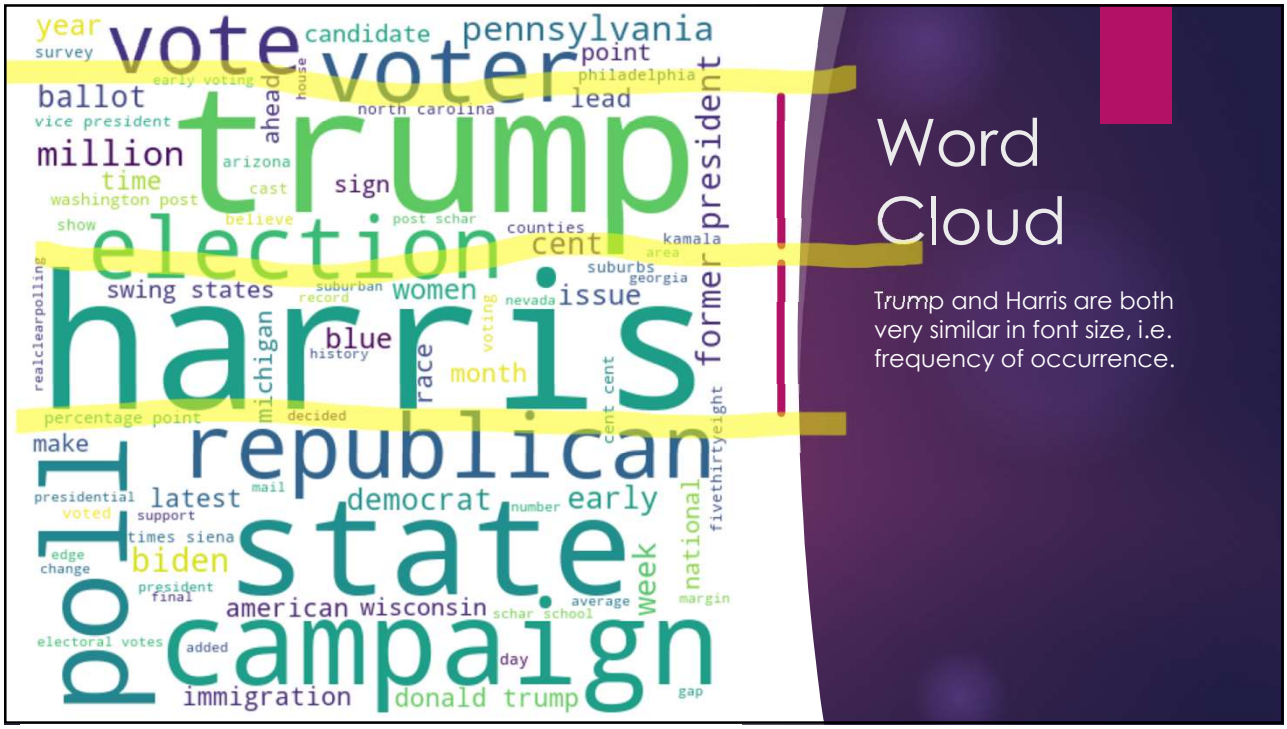


64



Word Cloud

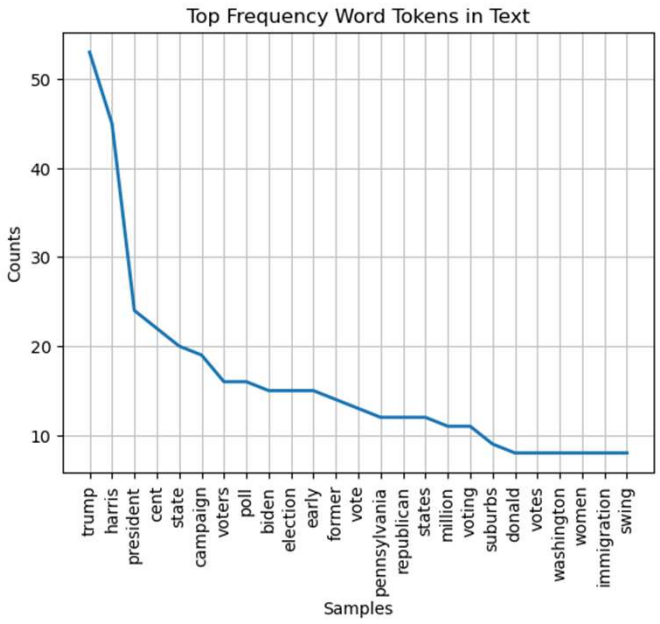
65



Word Cloud

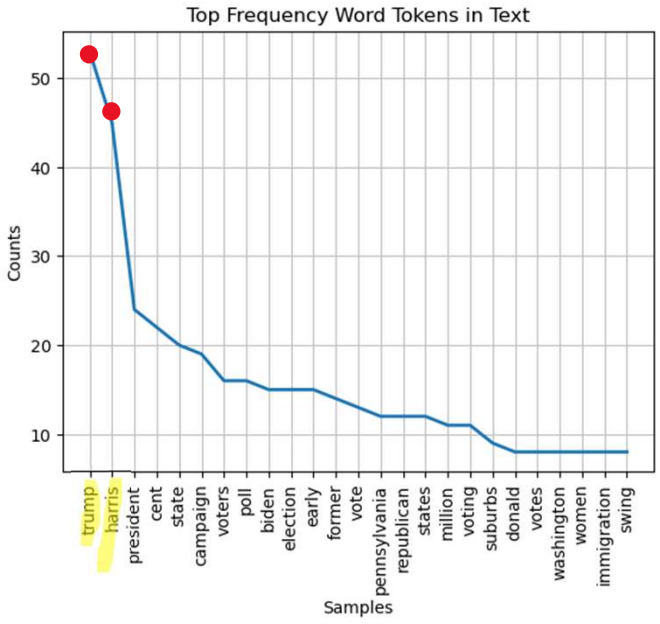
Trump and Harris are both very similar in font size, i.e. frequency of occurrence.

66



Word Frequency

67



Word Frequency

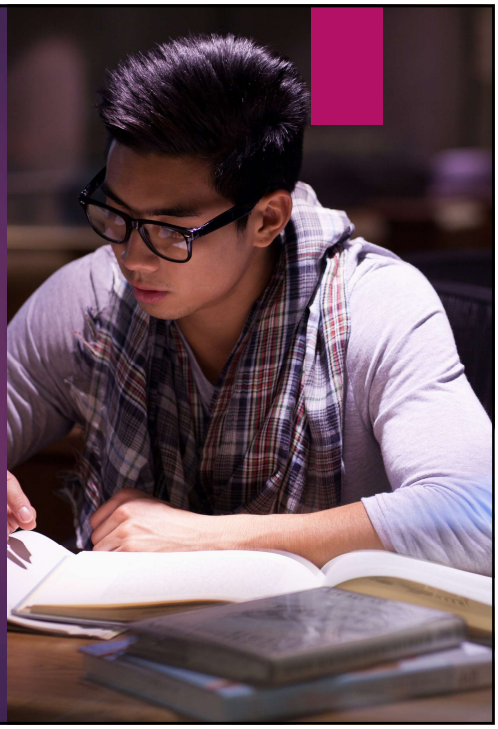
Trump seems to be most frequent, but Harris is second highly mentioned.

This supports findings of equal representation within moderate media outlets.

68

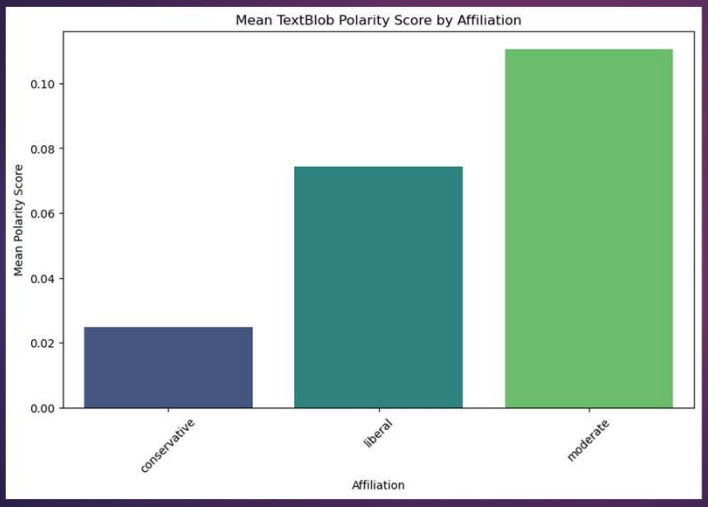
Findings

- Among moderate news outlets, certain news still retain a considerable level of bias. However, interesting emergence is the equal representation of both candidates on certain issues.
- This is suggested by both the unigrams and word cloud, where you can see both candidates level in word frequency gram-charts and relatively similar font size in the word cloud.



69

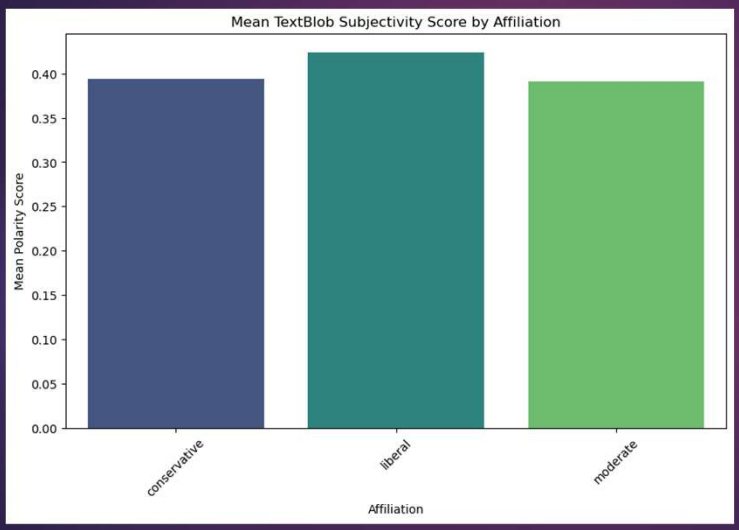
TextBlob Polarity



- Moderate Media are relatively more positive, per TextBlob
- All news are closer to being neutral with the highest polarity score of 0.1 (polarity of zero denotes neutral)

70

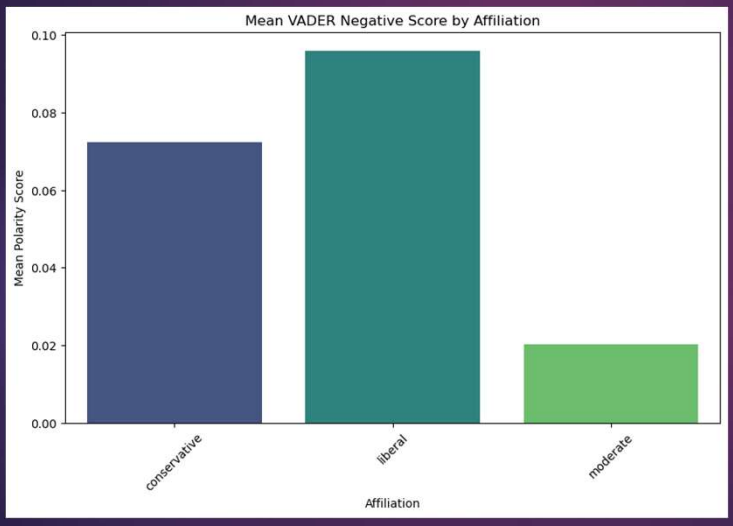
TextBlob Subjectivity



- Moderate Media roughly same level of subjectivity, around 50:50 between purely objective and purely subjective with slight lean towards objective .

71

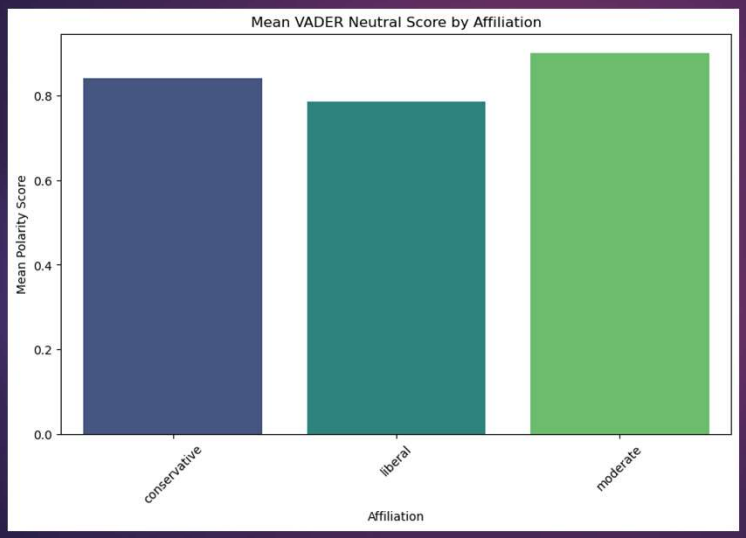
Vader Negative Score



- Vader Score shows that Moderate News is comparatively less negative

72

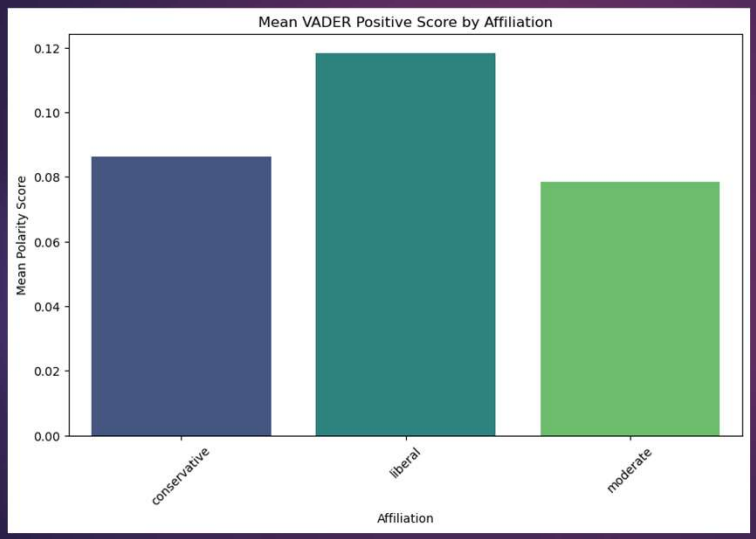
Vader Neutral Score



- Vader determines that Moderate is comparatively more neutral, but generally at similar levels with other news platforms

73

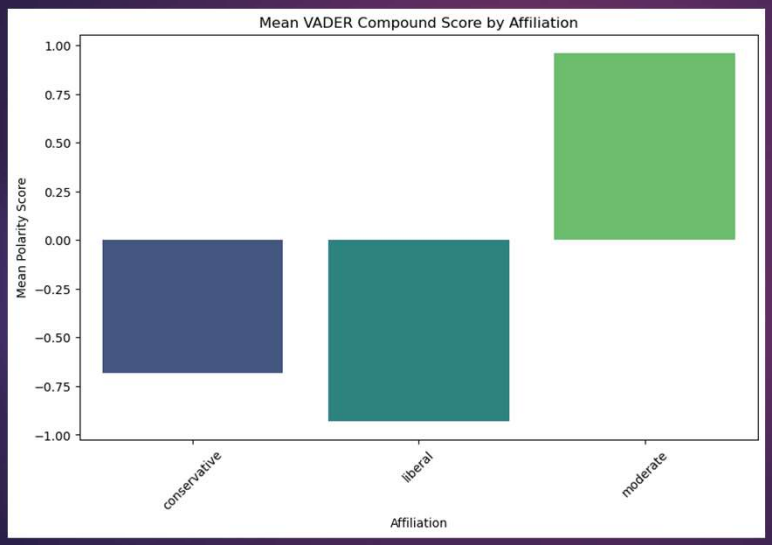
Vader Positive Score



- Vader determines that moderates are less positive than liberal or conservative media
- This finding is conflicts with TextBlob, likely due to scoring methodology

74

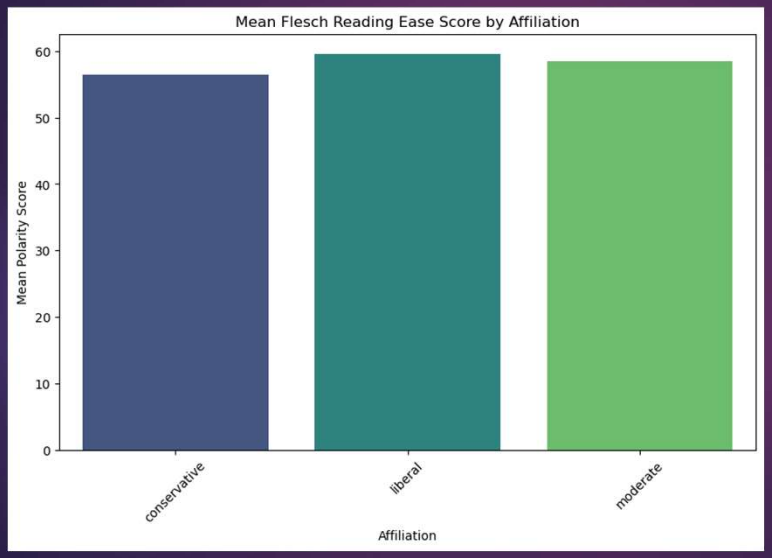
Vader Compound Score



- Vader Compound Scores, which aggregates Positive, Neutral, and Negative determines that moderate is positive.
- This is a strong contrast to liberal and conservative news.
- This finding is also consistent with TextBlob.

75

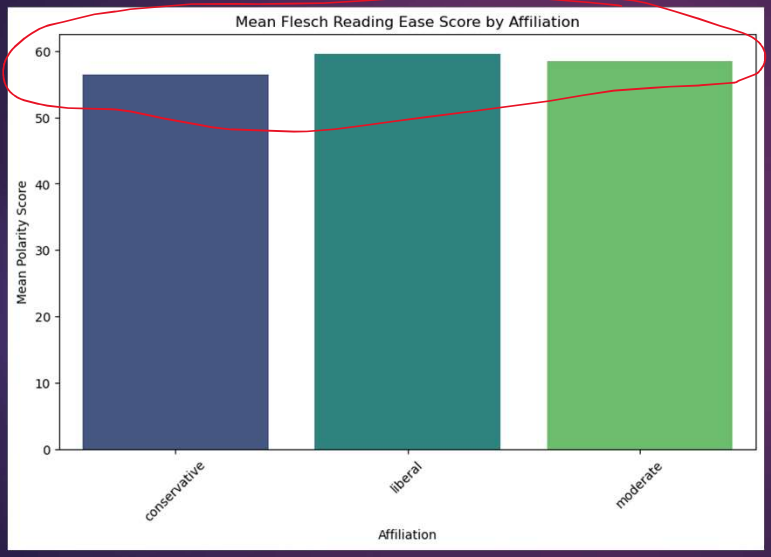
Flesch Reading Ease Score



Score	School level (US)
100.00–90.00	5th grade
90.0–80.0	6th grade
80.0–70.0	7th grade
70.0–60.0	8th & 9th grade
60.0–50.0	10th to 12th grade
50.0–30.0	College
30.0–10.0	College graduate
10.0–0.0	Professional

76

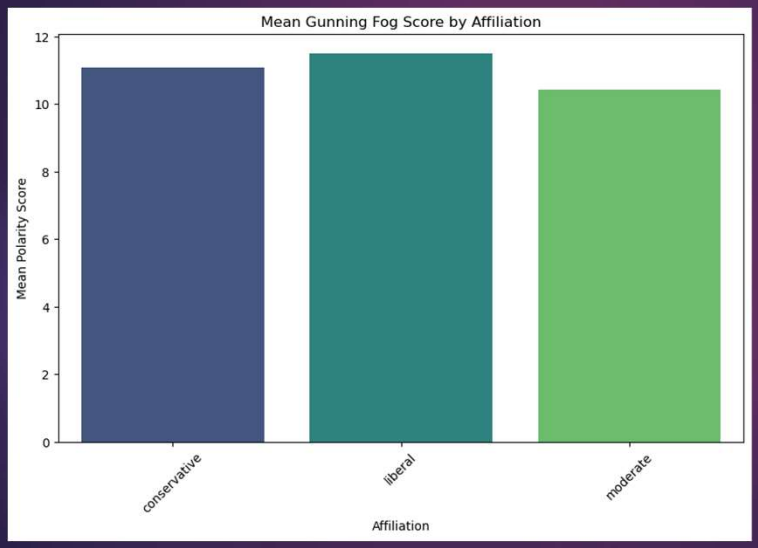
Flesch Reading Ease Score



Score	School level (US)
100.00–90.00	5th grade
90.0–80.0	6th grade
80.0–70.0	7th grade
70.0–60.0	8th & 9th grade
60.0–50.0	10th to 12th grade
50.0–30.0	College
30.0–10.0	College graduate
10.0–0.0	Professional

77

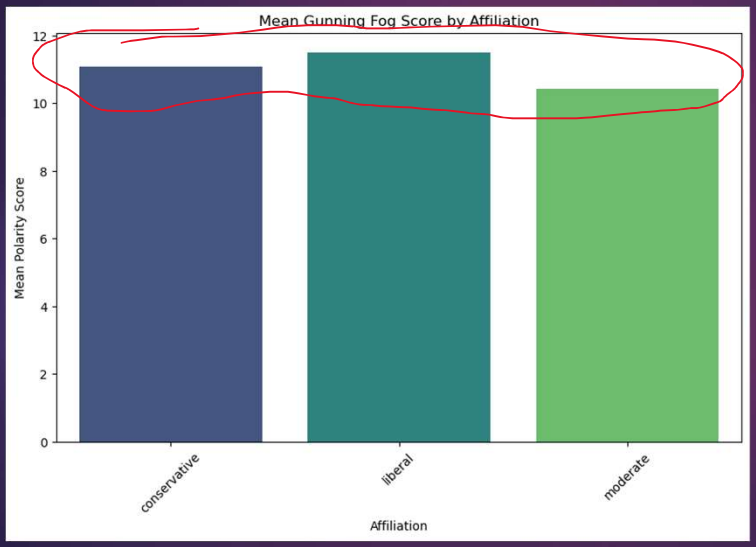
Gunning Fog



Fog Index	Reading level by grade
17	College graduate
16	College senior
15	College junior
14	College sophomore
13	College freshman
12	High school senior
11	High school junior
10	High school sophomore
9	High school freshman
8	Eighth grade
7	Seventh grade
6	Sixth grade

78

Gunning Fog



Fog Index	Reading level by grade
17	College graduate
16	College senior
15	College junior
14	College sophomore
13	College freshman
12	High school senior
11	High school junior
10	High school sophomore
9	High school freshman
8	Eighth grade
7	Seventh grade
6	Sixth grade

79

Findings

Affiliation	Subjectivity-Objectivity	Positive-Negative	Reading Level
Left-Leaning	50:50	Somewhere between neutral and negative	High School
Right-Leaning	50:50	Somewhere between neutral and negative	High School
Moderate	50:50	Mostly positive	High School

80

Implications

- Biased news media may influence polarization among voters in the upcoming election. The negative-leaning sentiments for left-leaning and right-leaning news outlets may influence voting along party lines rather than proper evaluation of candidates and platforms.
- While most news outlets aim to write at a high school reading level to garner mass appeal, this may oversimplify high-level details or conceptual implications that are important to key issues.
- The industry-wide finding of being 50:50 ratio for subjectivity to objectivity implies that no one news outlet will provide adequate comprehensive covered with the 50% subjectivity bias. It will be important for voters to consume a variety of reputable sources to have a complete picture.

81