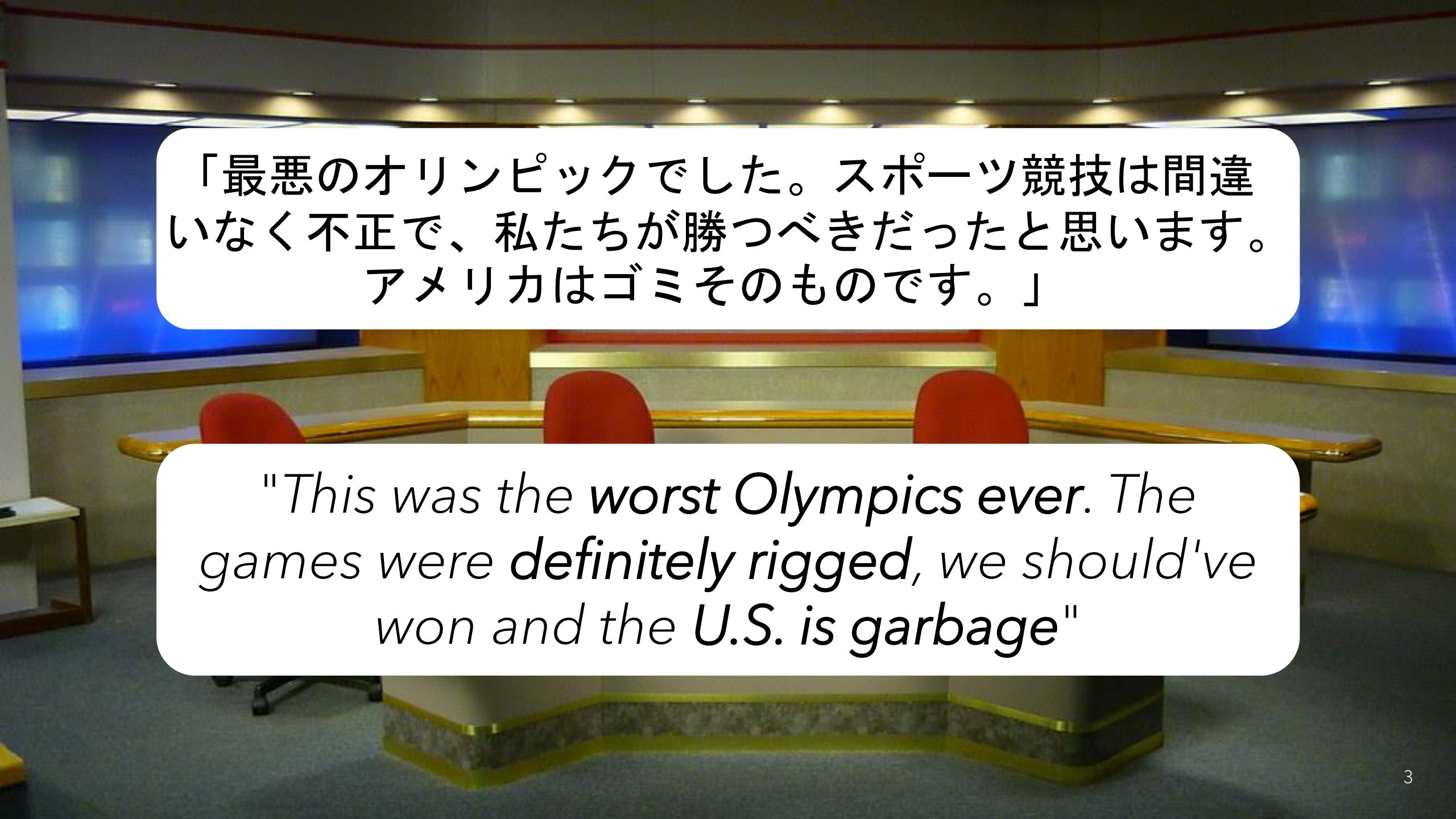
JANAS

Japan and America News Analysis through Sentiment

SAMYUKTA VAKKALANKA, SATOMI ITO, TREY ROARK

LIVE
BREAKING
NEWS


「最悪のオリンピックでした。スポーツ競技は間違いなく不正で、私たちが勝つべきだったと思います。アメリカはゴミそのものです。」
"This was the worst Olympics ever. The games were definitely rigged, we should've won and the U.S. is garbage"

JANAS

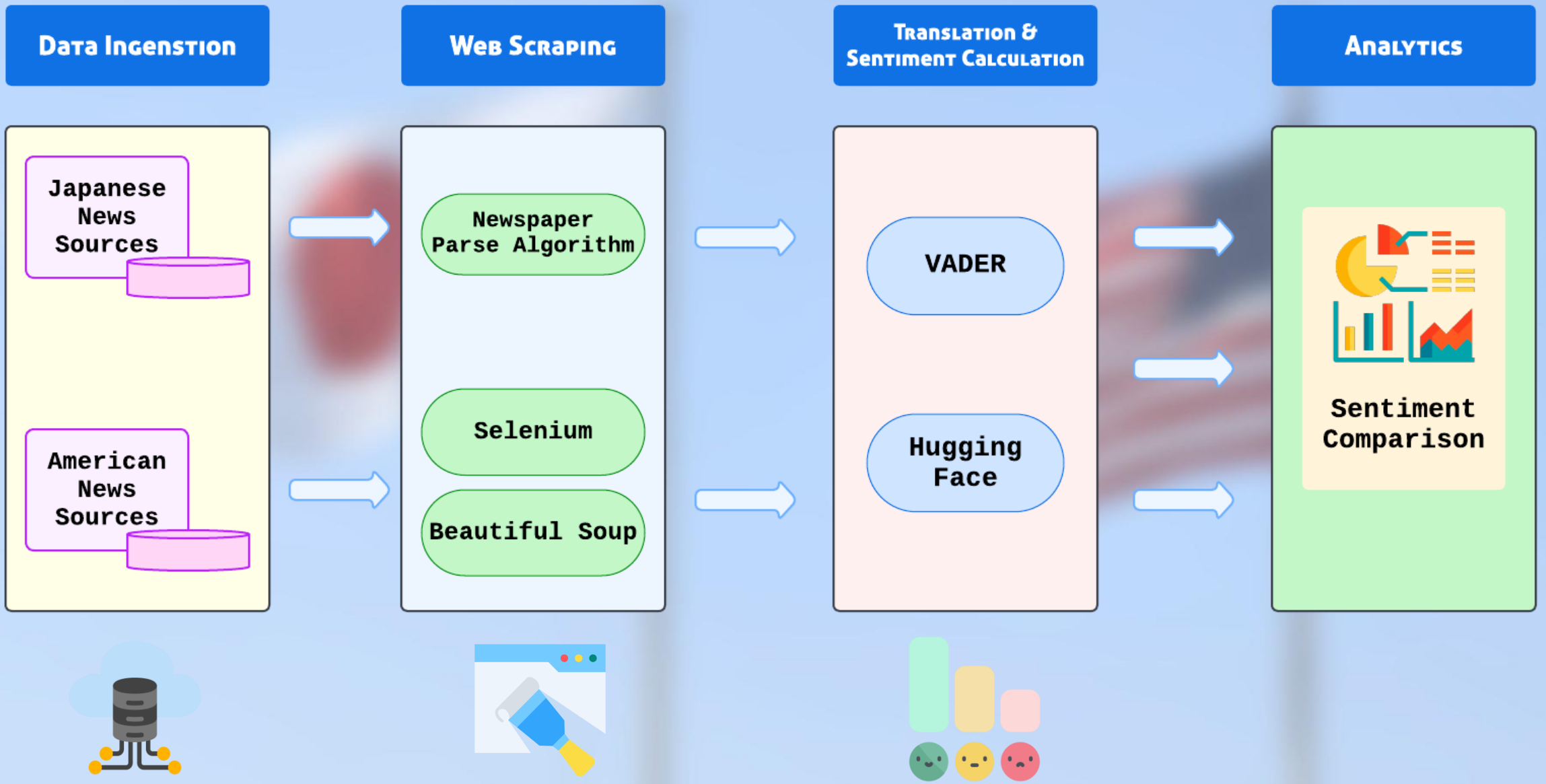
Japan and America News Analysis through Sentiment

Samyukta Vakkalanka, Satomi Ito, Trey Roark

Objective

Create an empirical review of the differences between how certain topics are displayed between American and Japanese News Media

Project Pipeline



Collecting News Articles

News Sources

- Left Leaning
- Central
- Right Leaning

Web Scraping

- Selenium
- Beautiful Soup
- Newspaper

Web Scraping - America

Sources

CNN

FOX NEWS channel

HUFFPOST

Topics

Trending
Business
International
Politics
Japan

Algorithm

Newspaper3k

Web Scraping - Japan

Sources

NHK

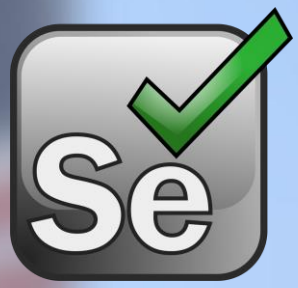
THE SANKEI SHIMBUN

毎日新聞

Topics

Trending
Business
International
Politics
America

Algorithms



BeautifulSoup

Translating Japanese Articles to English

- Hugging Face
- BERT Japanese Tokenizer
- Pre trained Encoder Decoder Model for translation
  - Manually checked sample of translations for accuracy

Sentiment Analysis

If English

- Valence Aware Dictionary and sEntiment Reasoner (VADER)

If Japanese

- Hugging face
- Tokenization: BertJapaneseTokenizer
- Logits were converted into probabilities using softmax function
- Compound score is calculated as the weighted sum of the probabilities

JANAS Data

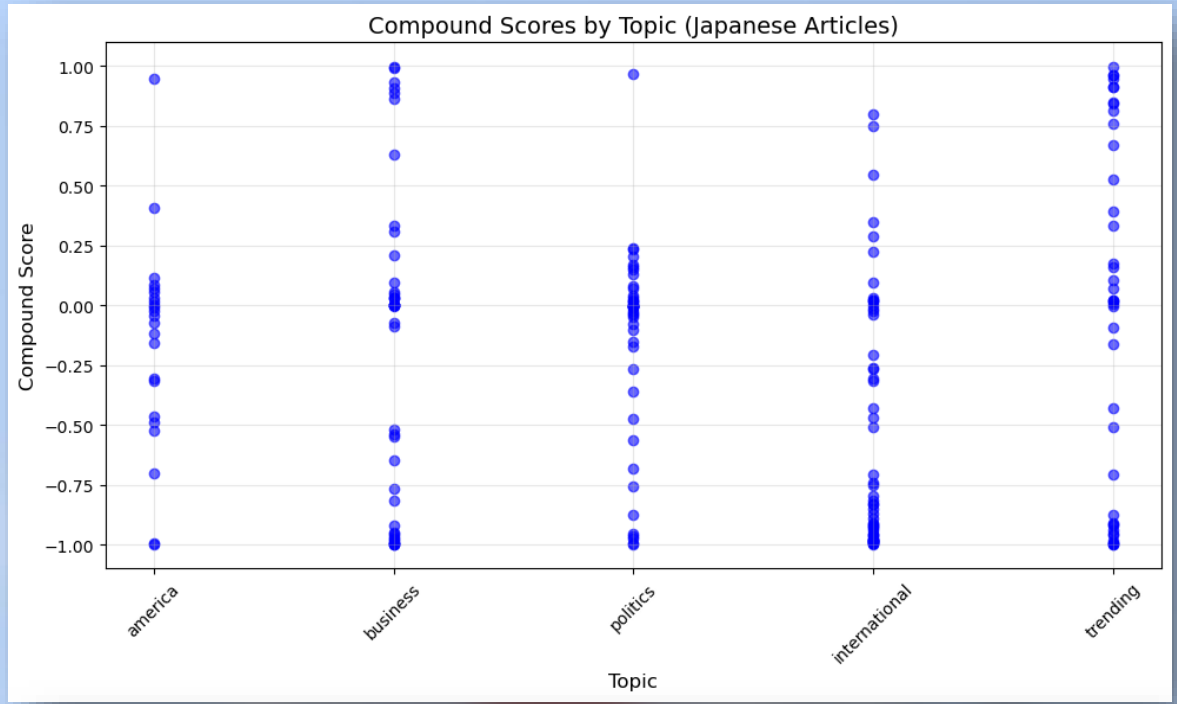
~ 250
Japanese
Articles

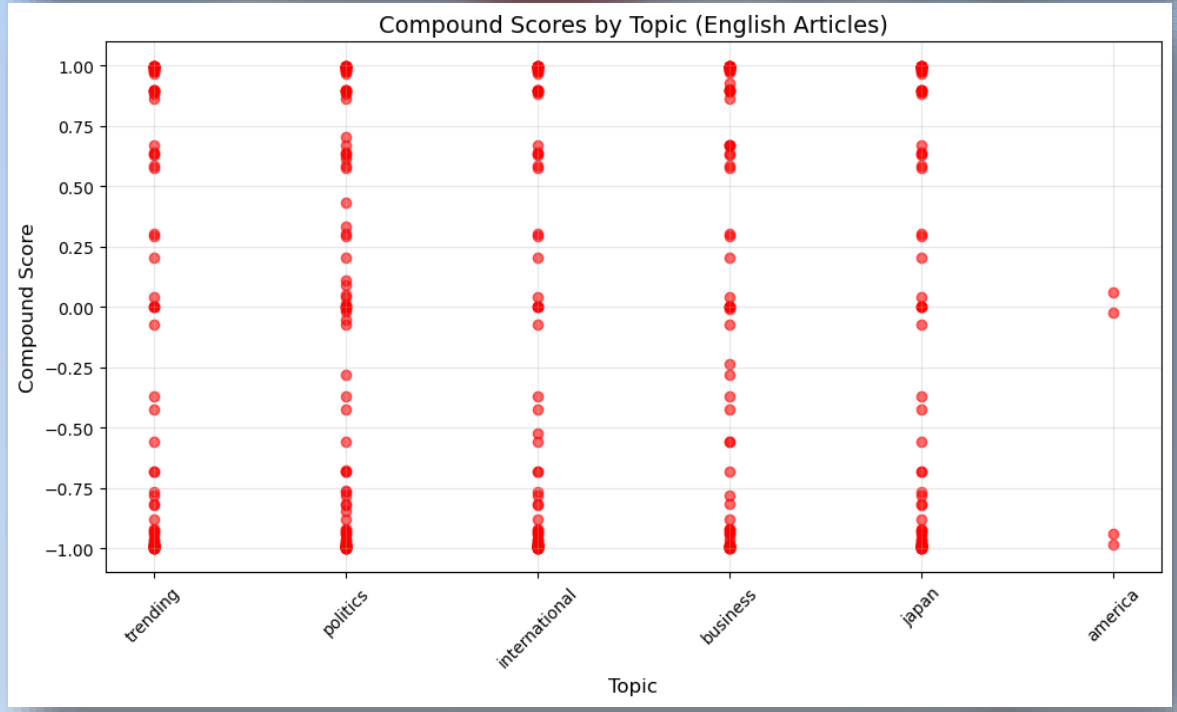
7
Features

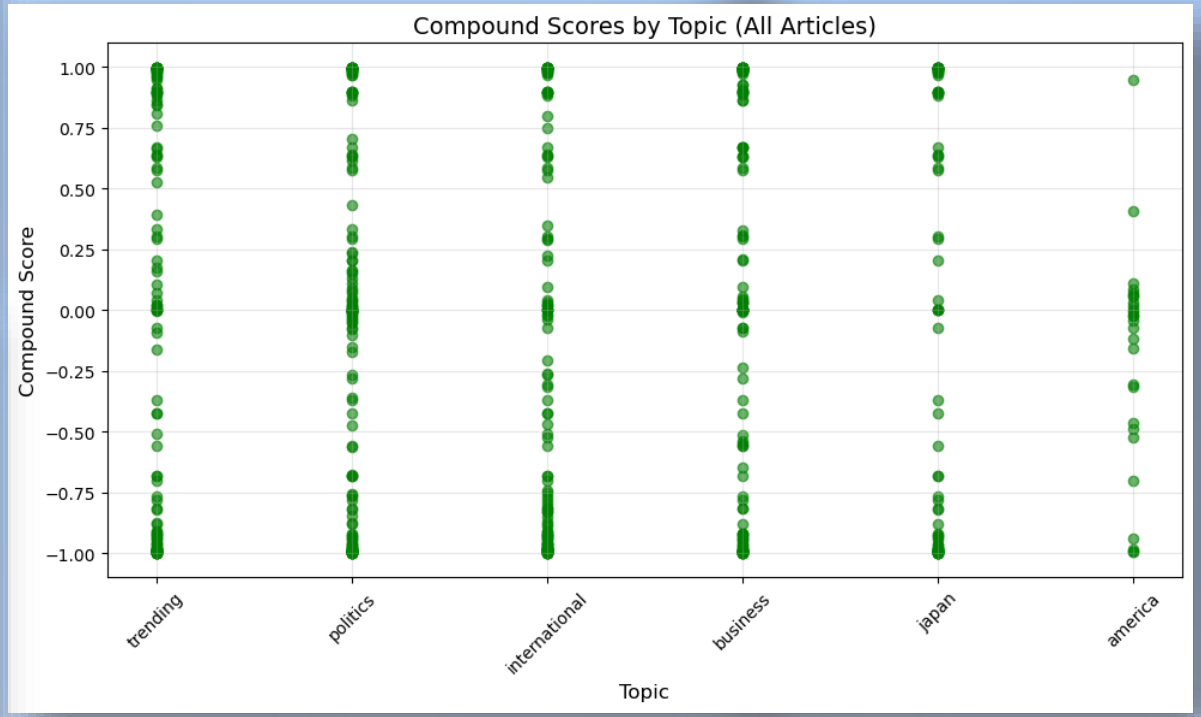
~ 450
American
Articles

- **Link**
- **Source**
- **Bias**
- **Topic**
- **Title**
- **Content**
- **Sentiment**

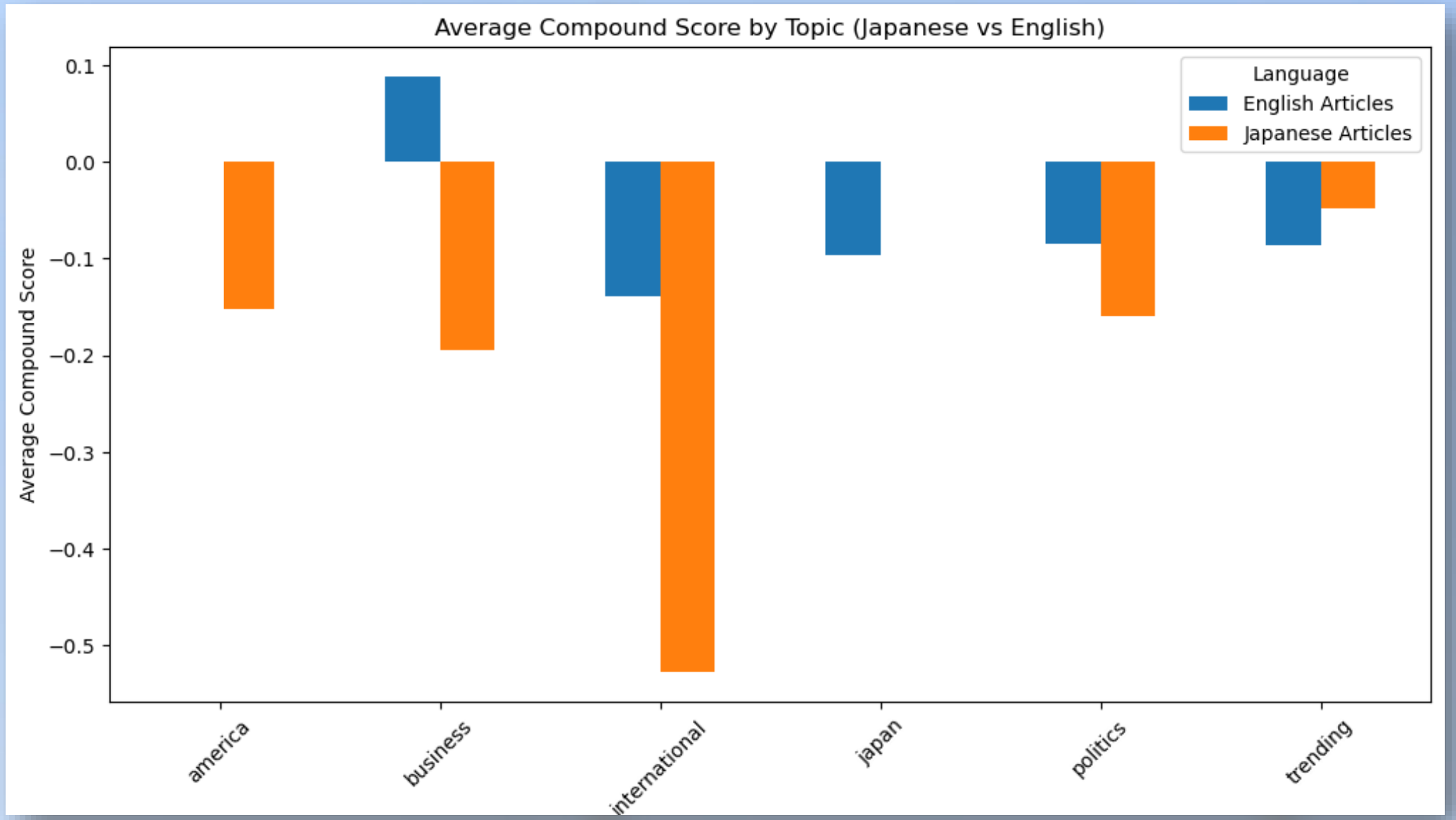
Sentiment Analysis: Compound Scores



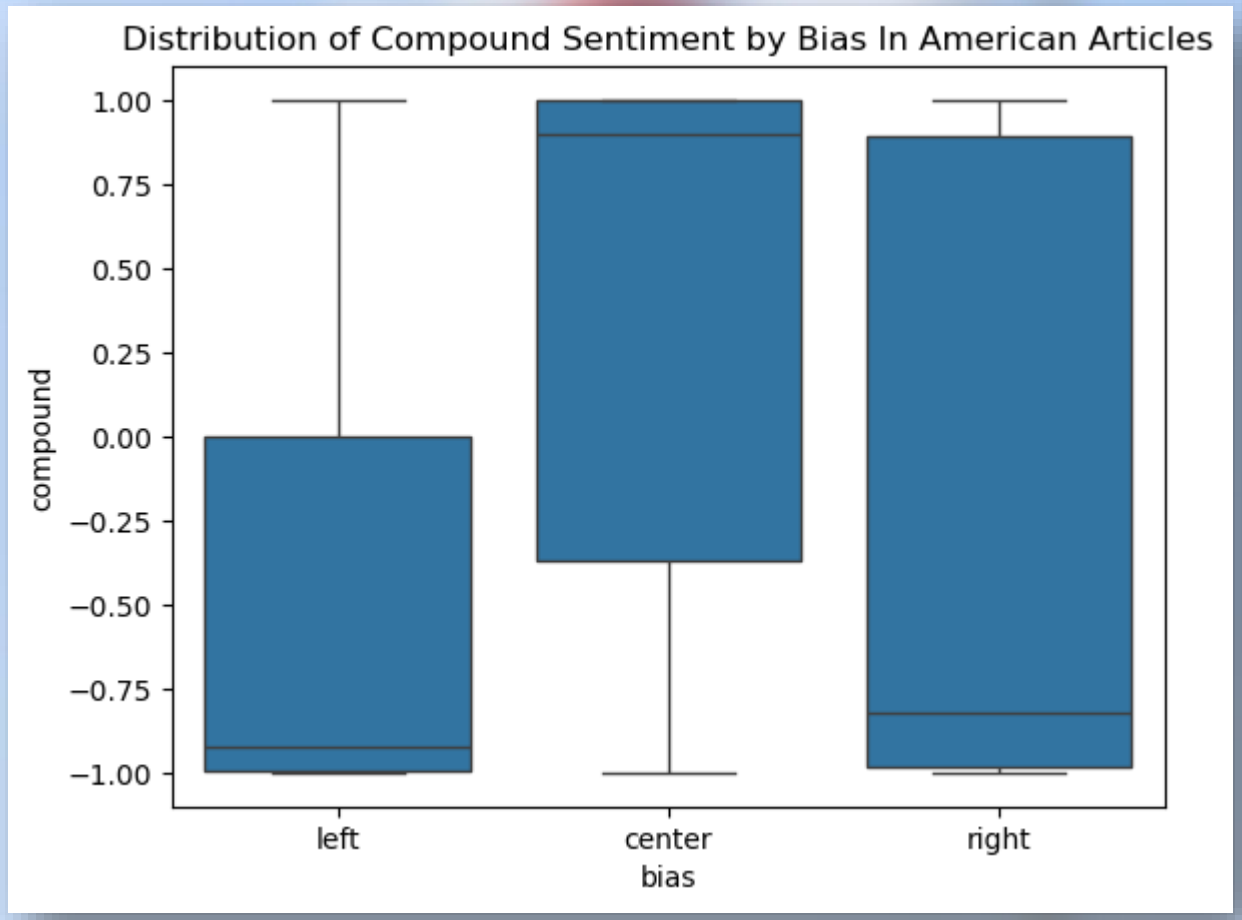


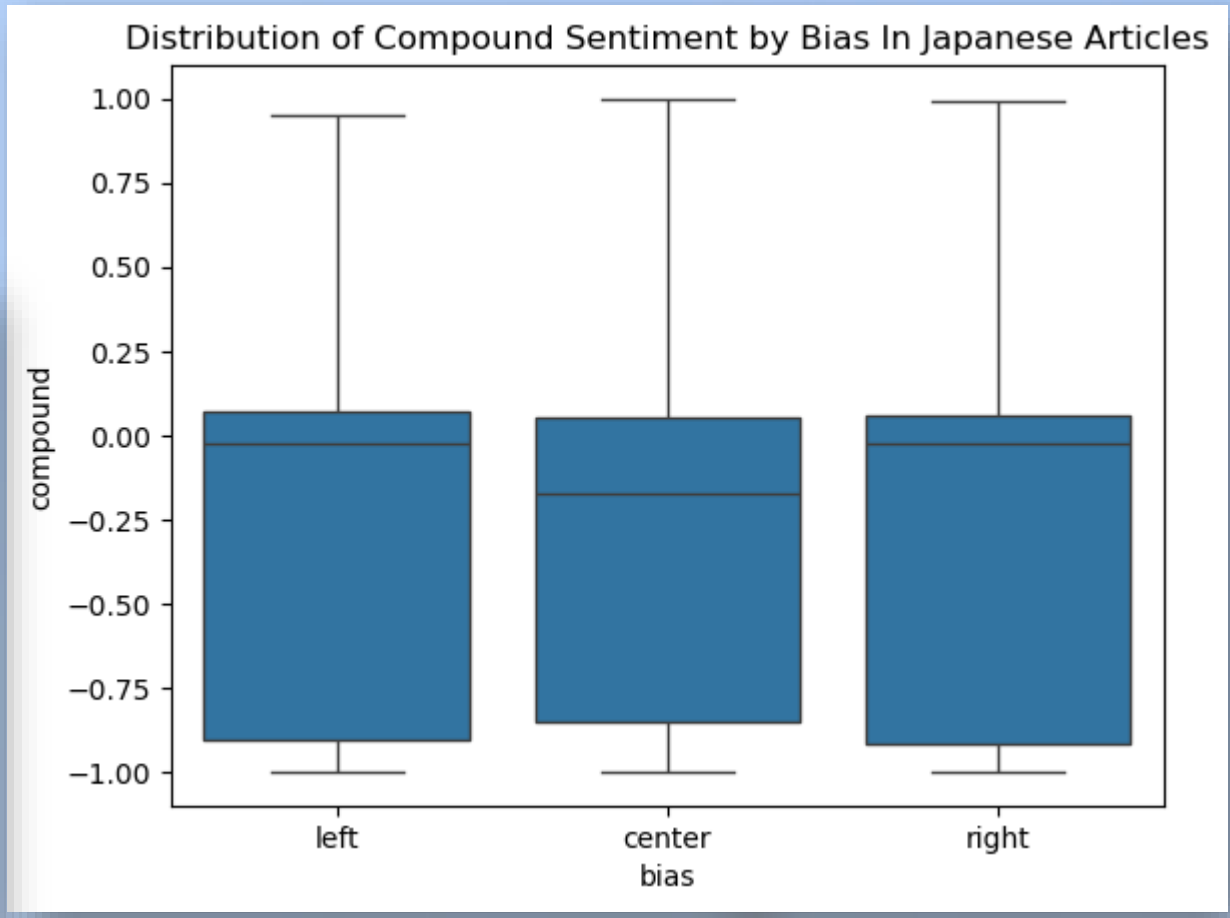


Sentiment Analysis: Across Languages

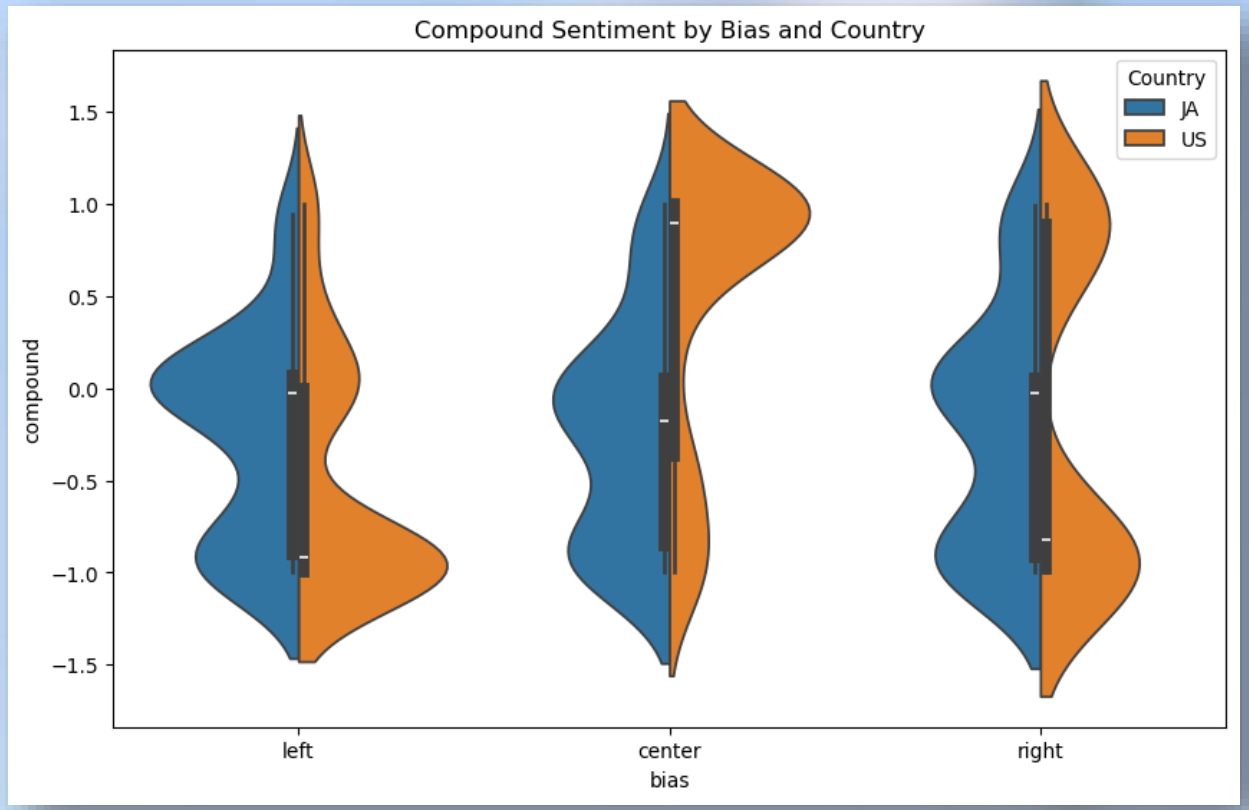


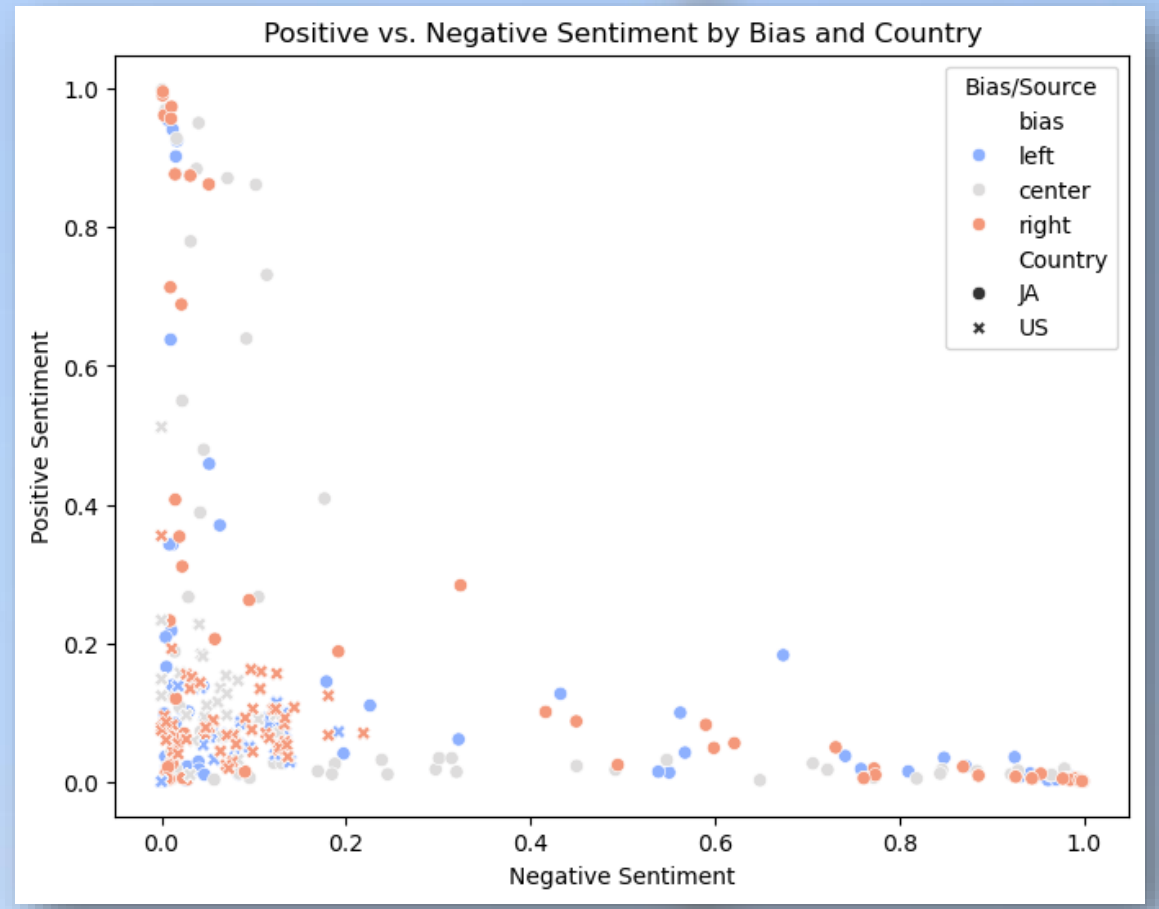
Sentiment Analysis: Bias In Each Country



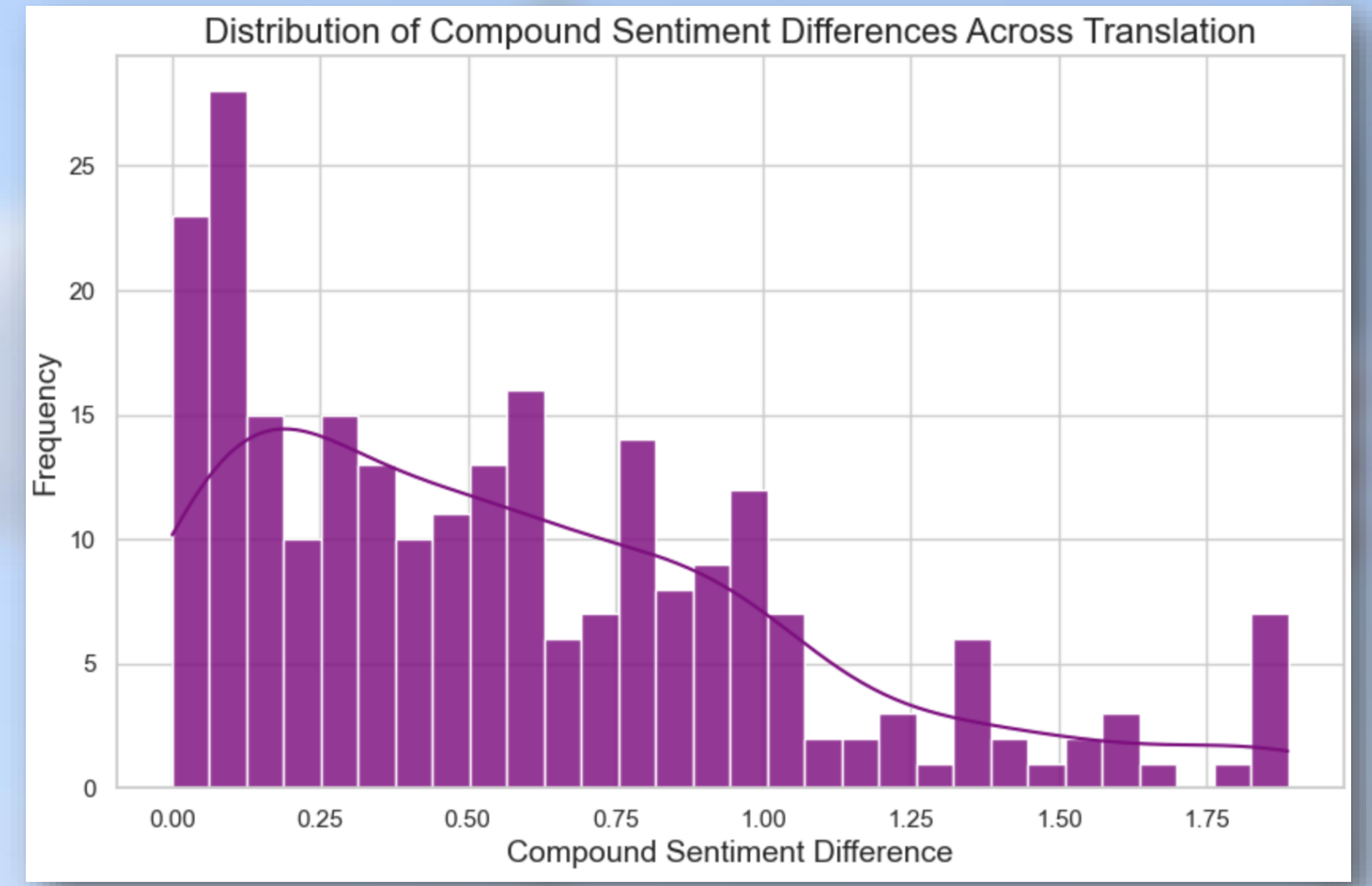


Sentiment Analysis: Bias Between Countries

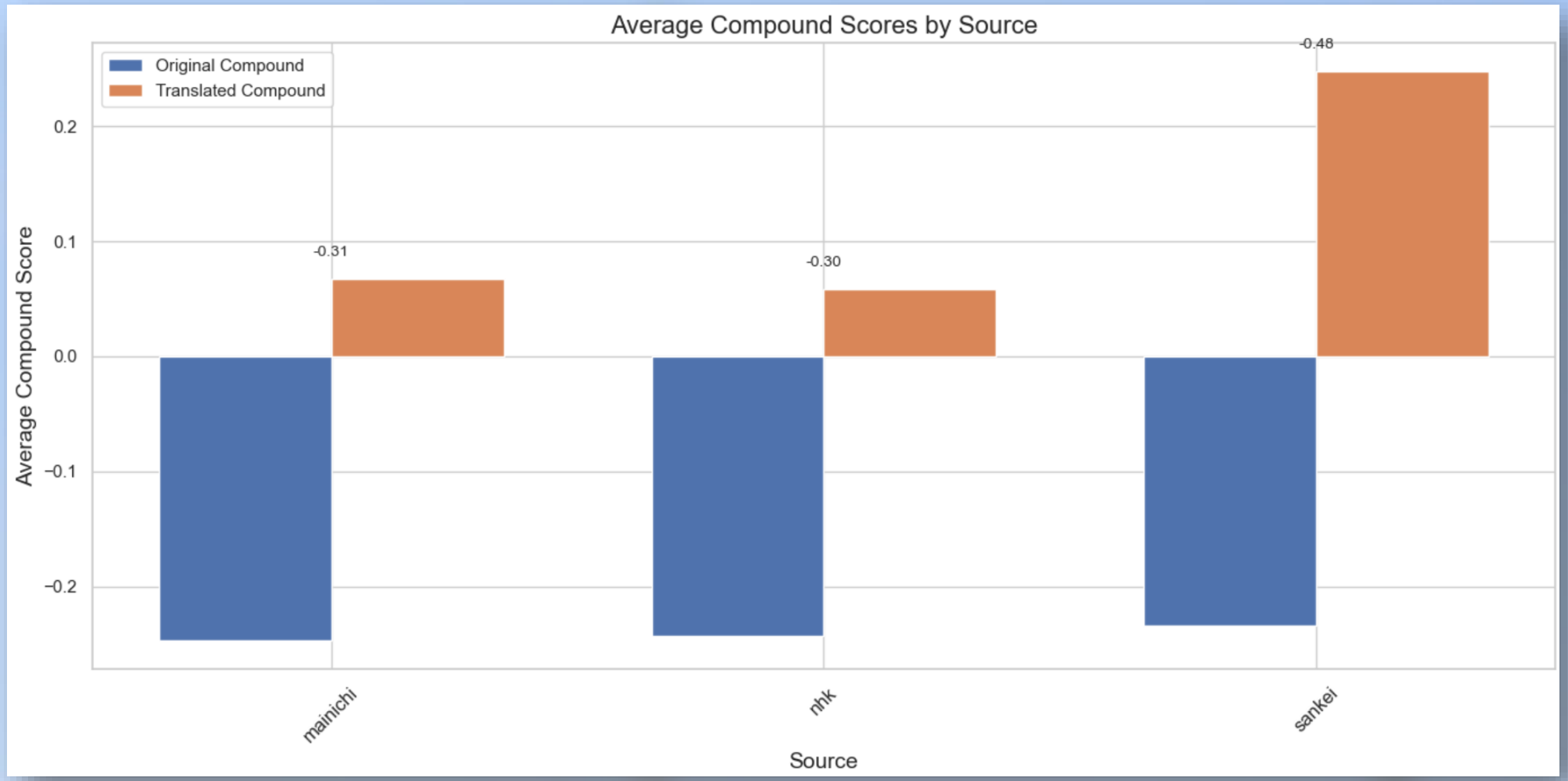




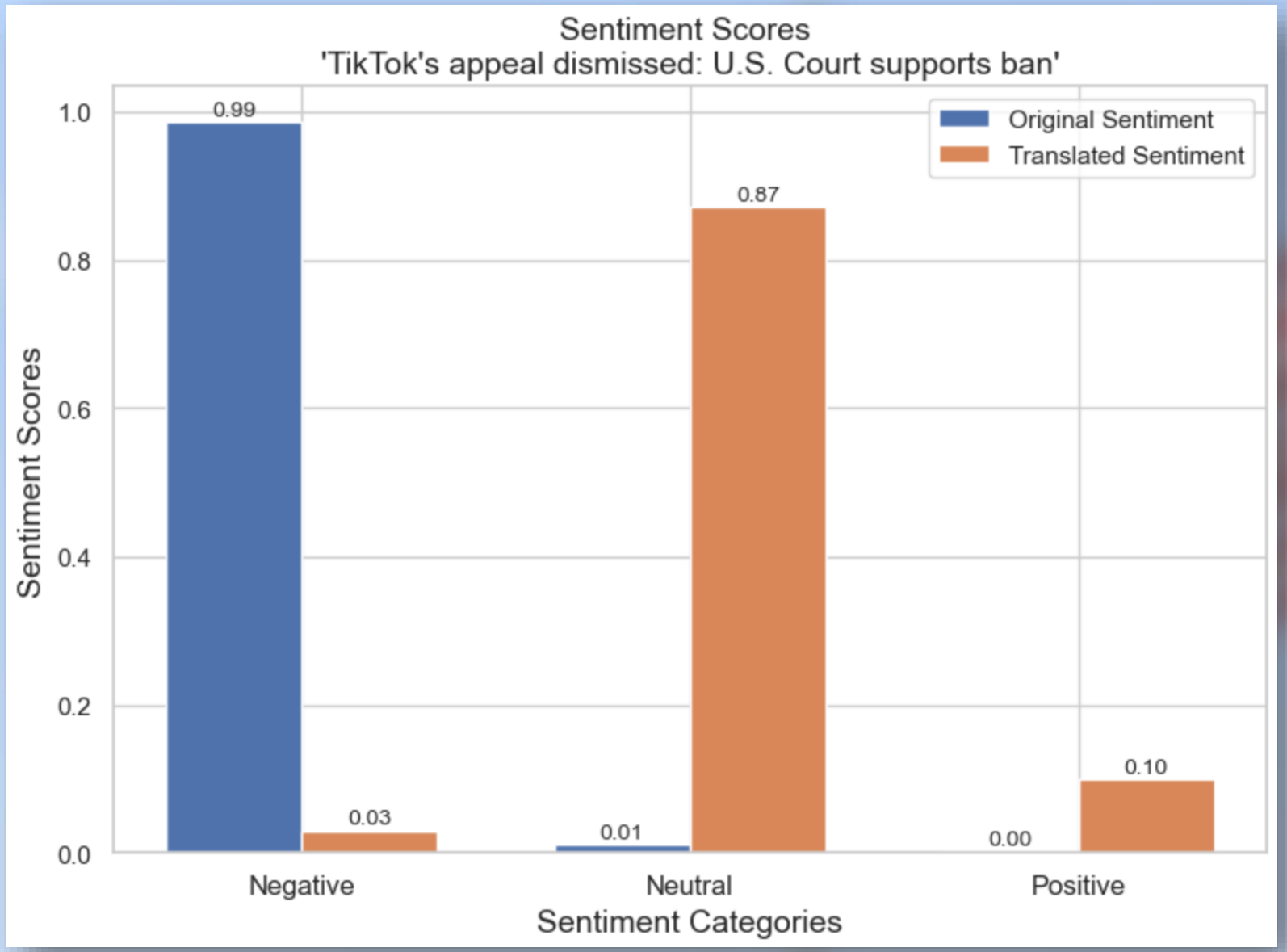
Sentiment Analysis: Across Translation



Sentiment Analysis: Across Translation



Translation Dilemma: 毎日新聞 - TikTok Ban



TikTokの訴え却下 米高裁、事業禁止法を支持

Takeaways

Translated documents need to be taken with a grain of salt even those with high accuracy

News can tend to be emotional, regardless of source

It's imperative to analyze these sources in their original language

Future Work

Website Construction

Further Analysis

Automation through Organization

References

"Christian-Phu/Bert-Finetuned-Japanese-Sentiment · Hugging Face." *Christian-Phu/Bert-Finetuned-Japanese-Sentiment · Hugging Face*, huggingface.co/christian-phu/bert-finetuned-japanese-sentiment. Accessed 9 Dec. 2024.

"VaderSentiment." PyPI, 23 Apr. 2018, pypi.org/project/vaderSentiment/.

Ou-Yang, Lucas . "Newspaper3k: Article Scraping & Curation — Newspaper 0.0.2 Documentation." Newspaper.readthedocs.io, 2013, newspaper.readthedocs.io/en/latest/.

Mitchell, Amy, et al. "Media Sources: Distinct Favorites Emerge on the Left and Right." Pew Research Center, 21 Oct. 2014, www.pewresearch.org/journalism/2014/10/21/section-1-media-sources-distinct-favorites-emerge-on-the-left-and-right/.

Gilroy, Susan. "Research Guides: News Media across the Political Spectrum: Starting Point: 1. "the Chart."" Guides.library.harvard.edu, 2024, guides.library.harvard.edu/newsleans/thechart.

Ghasiya, P., Okamura, K. (2020). Comparative Analysis of Japan and the US Cybersecurity Related Newspaper Articles: A Content and Sentiment Analysis Approach. In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds) Advanced Information Networking and Applications. AINA 2020. Advances in Intelligent Systems and Computing, vol 1151. Springer, Cham. https://doi.org/10.1007/978-3-030-44041-1_39

"Media Bias Chart." AllSides, 27 Feb. 2019, www.allsides.com/media-bias/media-bias-chart .