Op-Ed Biases in Israel-Palestine Conflict

How Political Sentiment Shifts in a Fast-Moving Crisis

Project Objective

Intro

- I looked at how the Op-Ed sections of three major U.S. media outlets leaned.
  - How sympathetic to Israel and its suffering
  - How sympathetic to Palestine and its suffering
- Using allsides.com:
  - NYTimes (Op-Ed) → extreme left of center
  - CNN.com → left of center
  - WSJ.com → right of center

Objective

Could I use sentiment analysis (NLP) to pick up on this "Political Sympathy?"

Were there different "political sympathies" within the same Op-Ed section?

Did this sentiment change over each week?

Did some of the same author's change their sentiments over the weeks?

Methodology

- Use Op-Eds from NYTimes.com, CNN.com, and WSJ.com for a 5-week period
- Analyze their sentiment score using NLP techniques (TextBlob)
  - Did not pick up on political biases
- BERT too sophisticated for my CPU

- Devise my own metric "Sympathy Score" using two LLMs
  - ChatGPT 4
  - Claude 2.1
- Calibrate the LLMs and my own ratings

Limitations

- Small dataset
  - 124 Op-Eds
- Not all of the spectrum covered
- I am "informed layperson", not expert

- My own biases
- Issues with LLMs
- Not scalable

Training LLMs

Prompt engineering and tuning

Step #1 - Create input

I would like to come up with a number, between -1 and 1, that captures to what extent an article sympathizes with Israel and to what extent it sympathizes with Palestine. This is going to be tricky, though, because what exactly does 'sympathize' mean. To give you a sense: an article that is, say, -0.8 (so strongly sympathetic to Israel) will focus almost exclusively on Israel's suffering. It will only mention Palestine in the context of Hamas. In essence, it paints a good vs. evil dichotomy in which it is Israel vs. Palestine (and it will often conflate Hamas with Palestine). On the other hand, an article is a positive 0.8 will focus only on the current destruction and death in Gaza and Israel as the aggressor . . . how does this description work? Is it clear? Is there something I potentially left out? Is there potentially something I focused too much on?

Prompt engineering and tuning

Step #2 - Incorporate feedback

ChatGPT

Your approach to quantifying sympathy in articles about the Israel-Palestine conflict using a numerical scale from -1 to 1 is an interesting and challenging endeavor. The description you've provided is fairly clear in its intent to measure the degree of sympathy towards either Israel or Palestine in an op-ed piece. The key elements you've described include:

This scale seems to capture a broad spectrum of possible biases in reporting and opinion pieces. However, there are a few considerations that could further refine your approach:

- **Subjectivity in Interpretation**: The scale relies on the reader's interpretation of what constitutes sympathy or bias, which could vary greatly. Establishing clear criteria for what constitutes a -0.8 versus a -0.2, for example, might be necessary.
- **Nuanced or Mixed Viewpoints**: Some articles might present a more balanced view or acknowledge valid points on both sides. It's important to consider how such articles would be scored.

Prompt engineering and tuning

Step #2 - Incorporate feedback

included. One thing that should also be include is a lack of sympathy (for the other side), whether through explicitly discounting it or not mentioning it all. For nuances and context, let's say two articles sympathize with one side at the exact same levels, if we include your three criteria--emotional support, justification of actions, allocation of blame--but one article does acknowledge the other side's perspective whereas the other doesn't. The article that does acknowledge should have a score closer to zero. By how much depends on the level of discussing and to whatever extent empathizing with the other side. Differentiation Between Political Entities and People: Yes, how it distinguishes between groups should be included in this analysis. Yes, historical and cultural context are important. The choice of words, the tone, etc. are all important to include in the "Sympathy Score." Let's not account for source and author bias. While these certainly contribute to one's perspective, I want us to engage with the raw text as much as possible. I'm taking for reputable news sources, so I don't anticipate misinformation.

Prompt engineering and tuning

Step #2 - Incorporate feedback

Thank you for the additional context and clarification on key aspects of assessing sympathy in these articles. Based on your details, here is an updated summary of the key factors to consider:

- Level of emotional support, justification of actions, and allocation of blame towards Israel vs Palestine

- Presence and degree of acknowledging the other side's perspective and suffering

- Differentiation between governments, political groups (e.g. Hamas), and the people

- Tone and word choice used when discussing each side (e.g. use of "oppressor" vs more neutral terms)

- Inclusion of historical and cultural context and which narrative it aligns with

- Lack of sympathy through omission of the other side's perspective or suffering

- Expresses concerns about potential ethnic cleansing and genocide of Palestinians

The article does acknowledge wrongdoing by Hamas and warns against anti-Israel extremist views. But overall it is still clearly sympathetic to the Palestinian side, while portraying Israel's actions and rhetoric as severely disproportionate and dangerous. Hence a score of 0.6 rather than a full 0.8.

Please let me know if I have applied the criteria appropriately again in assessing the sympathy level here. Looking forward to your feedback.

Copy

I had it between 0.5 and 0.6 for much of the same reasoning you did. Specifically, the lines acknowledging the horrific, dehumanizing acts against Israel. These lines, while brief, are unequivocal and lead off the article. They don't just say Hamas attack; they equate them with the horrors unfolding in Gaza. Secondly, it shows a little bit of nuance
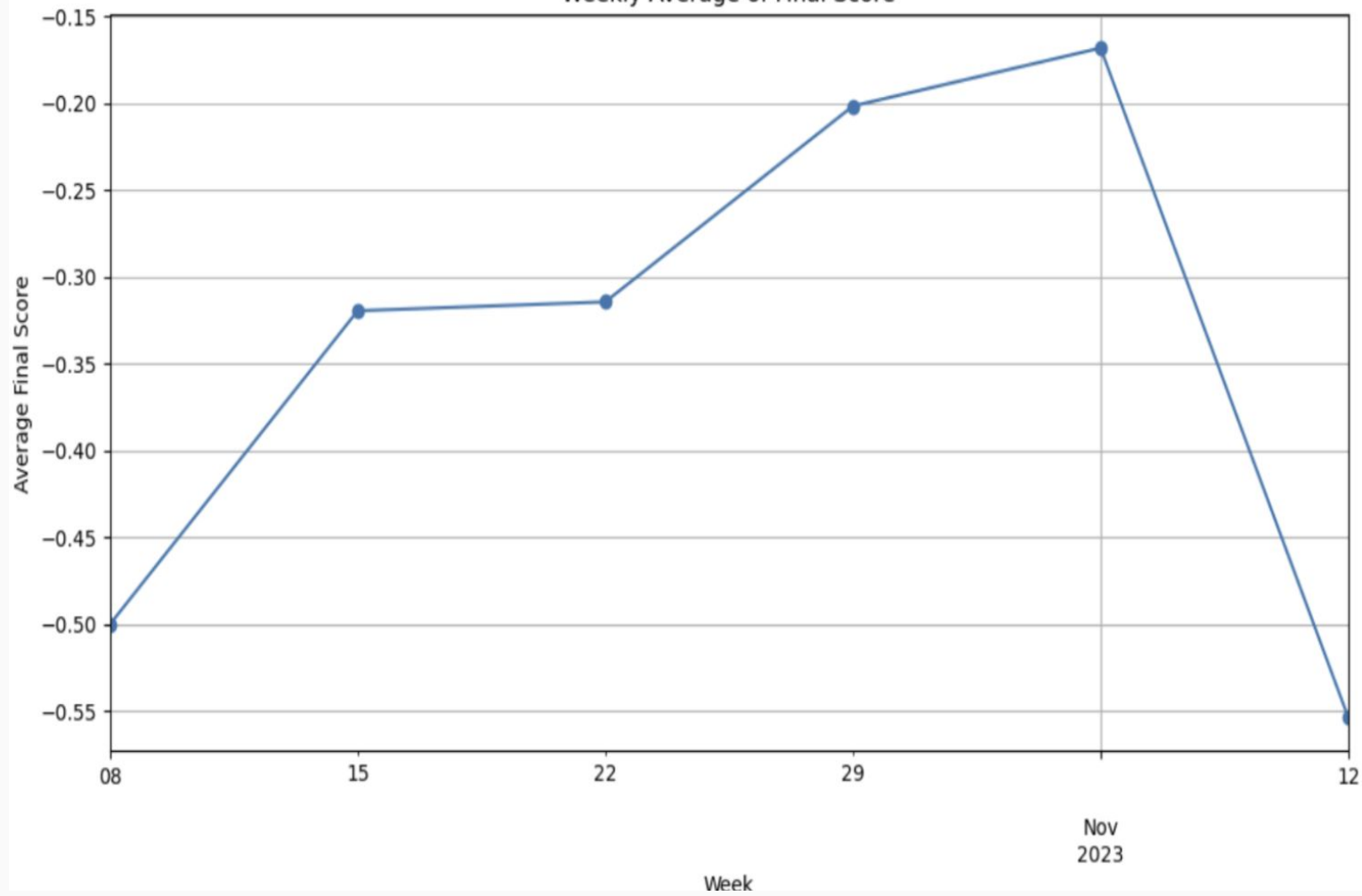
Prompt engineering and tuning
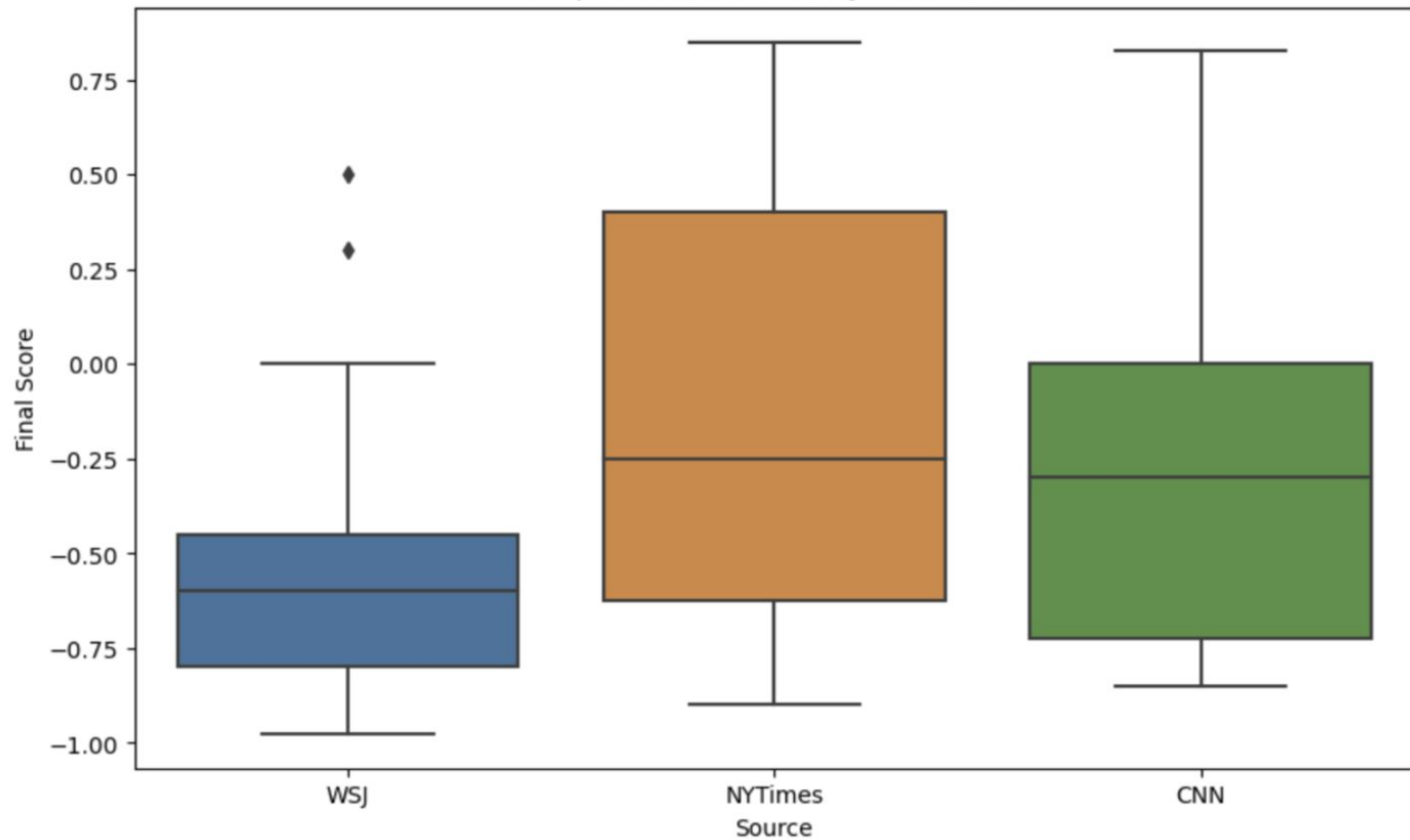
Step #3 - Calibration

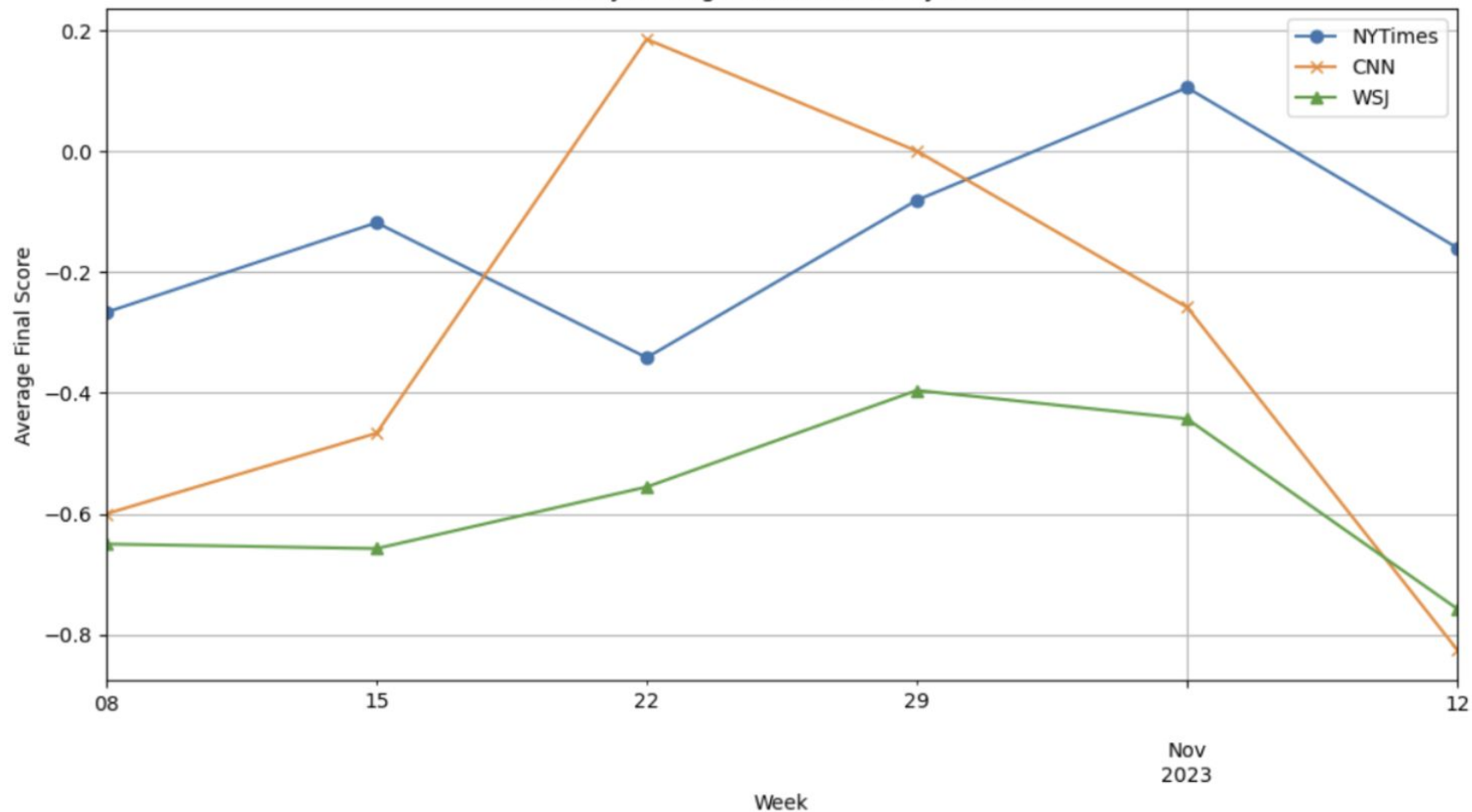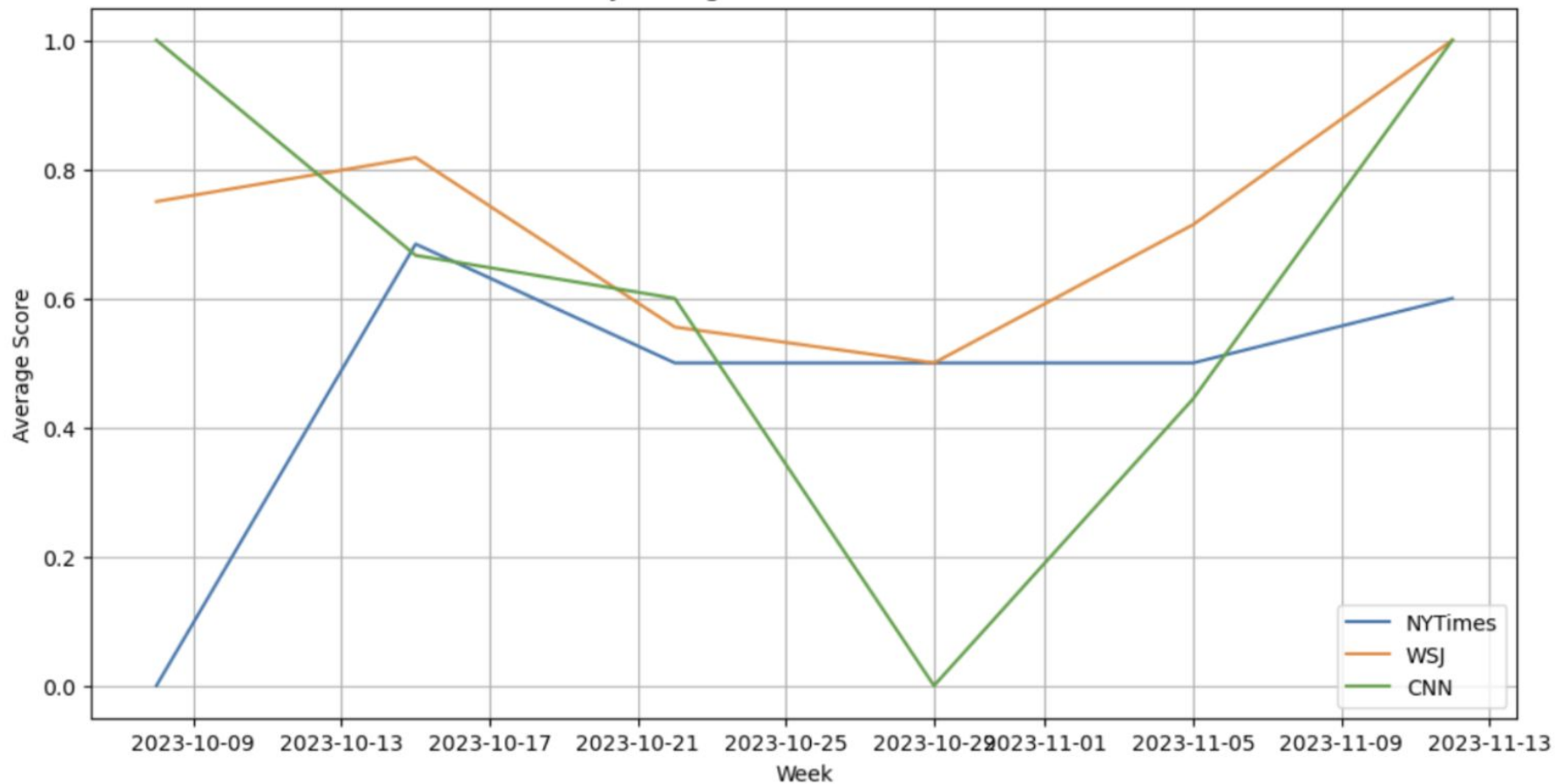| SS (Human)	SS (GPT)	SS (Claude)
-0.8	-0.8	-0.8
0.7	0.8	0.7
-0.5	-0.2	-0.65

Exploratory Data Analysis (EDA)

Weekly Average of Final Score
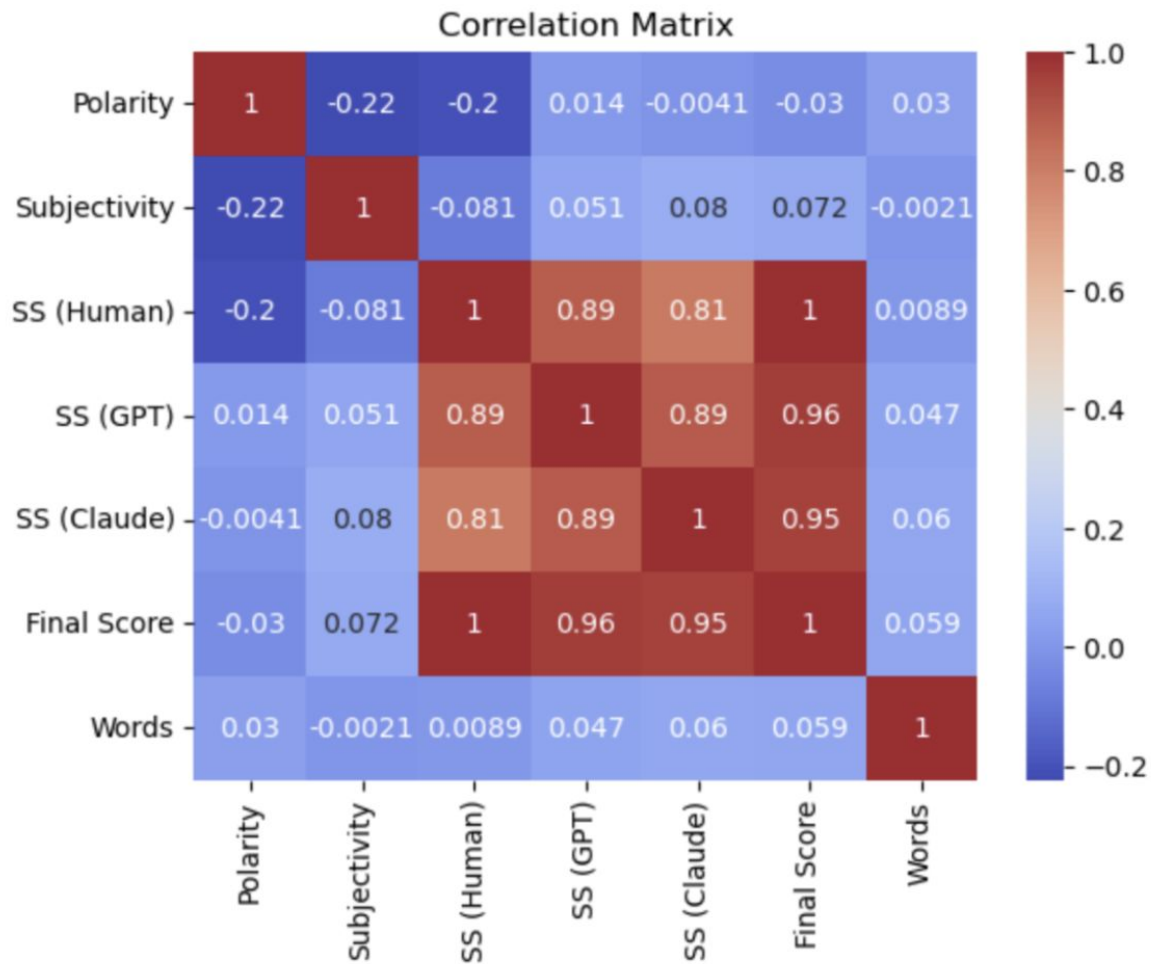
Boxplot of Final Score by Publication

Weekly Average of Final Score by Source

Weekly Average Score for Each Publication

Correlation Matrix

|	Polarity	Subjectivity	SS (Human)	SS (GPT)	SS (Claude)	Final Score	Words
Polarity	1	-0.22	-0.2	0.014	-0.0041	-0.03	0.03
Subjectivity	-0.22	1	-0.081	0.051	0.08	0.072	-0.0021
SS (Human)	-0.2	-0.081	1	0.89	0.81	1	0.0089
SS (GPT)	0.014	0.051	0.89	1	0.89	0.96	0.047
SS (Claude)	-0.0041	0.08	0.81	0.89	1	0.95	0.06
Final Score	-0.03	0.072	1	0.96	0.95	1	0.059
Words	0.03	-0.0021	0.0089	0.047	0.06	0.059	1

Modeling

Baseline model

Model MSE: 0.34134798999999993
Model RMSE: 0.5842499379546393
Model MAE: 0.46062
Model R-squared: -0.022134622525122882

K-Means Clustering

PCA Plot with Cluster Labels
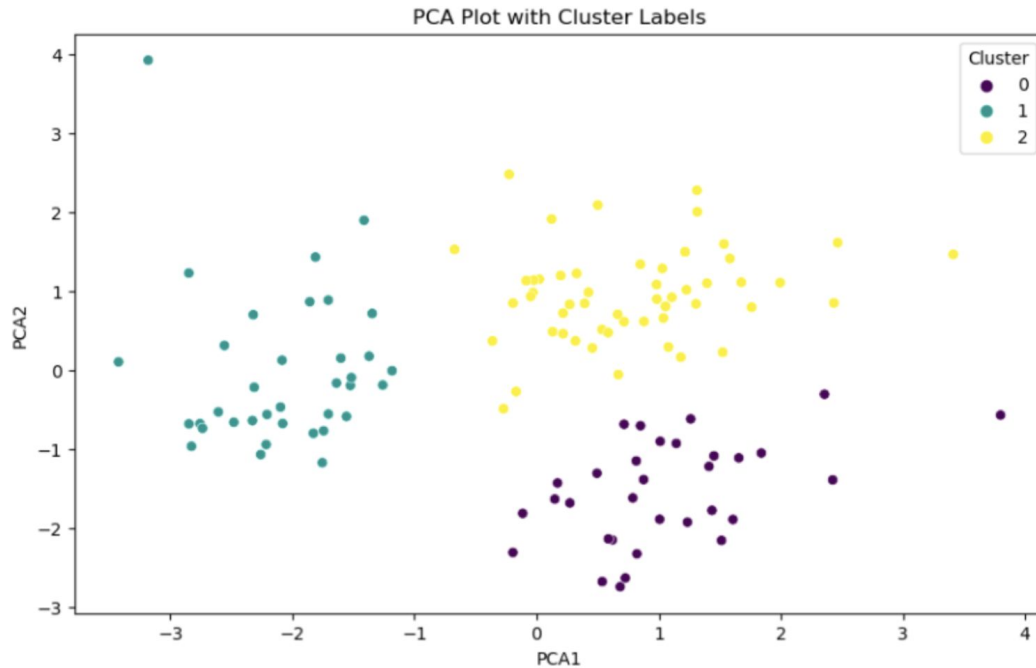
K-Means Clustering

Cluster	0	1	2
Source			
CNN	0.190476	0.238095	0.571429
NYTimes	0.172414	0.465517	0.362069
WSJ	0.428571	0.095238	0.476190

Additional Use of LLMs

LLM	Success Rate
ChatGPT	6 of 8 articles/75%
Claude	6 of 8 articles/75%

Future Steps…

- Use BERT or other advanced NLP tools to do similar analysis
  - Compare results with this analysis
  - Will not replace BERT but will act as a substitute ("Poor man's version)
- Run with more advanced LLMs in future to see if they are more accurate