Sentiment Dynamics in Professional News Reporting After the 2024 U.S. Presidential Election
Jingwen Wang
USC email: jwang426@usc.edu
USC ID: 7179708388

This project examines sentiment patterns in professional news reporting following the 2024 U.S. presidential election. While news outlets aim for neutrality, subtle emotional framing may still appear, especially in attention-grabbing headlines. Using BBC News articles from the World, Business, and Technology sections, the study compares sentiment in headlines versus article bodies and tracks how these patterns evolve over time during the post-election period. Through computational sentiment analysis, the project provides insight into how professional journalism balances factual reporting with reader engagement during a politically sensitive period.

The primary research questions addressed in this project are:

- Are news headlines more emotionally charged than article body text?
- How does sentiment in headlines and article bodies change over time during the aftermath of the 2024 U.S. presidential election?
- Do sentiment patterns differ across news sections such as World, Business, and Technology?

3.1 Data Sources

The data used in this project were collected from the international edition of BBC News, specifically from the website https://www.bbc.com/news. Articles were scraped directly from the BBC News website across three sections: World, Business, and Technology. All articles were accessed through BBC News search result pages and individual article pages hosted on bbc.com/news, which share a consistent HTML structure across sections.

The data were collected using custom Python web-scraping scripts written with the requests and BeautifulSoup libraries. Rather than downloading a pre-existing dataset, my project programmatically queried BBC News search pages using section-specific keywords related to the 2024 U.S. presidential election and its aftermath.

For each article, the following information was collected and stored: Article URL, News section, Publication date, Headline text, Article body preview, and Raw HTML file (saved locally for reproducibility)

All raw HTML files were saved under data/raw/html/, and a structured metadata file was created at data/raw/metadata.csv.

3.2 Data Collection Approach

Data collection was fully automated using the script get_data.py. For each news section, election-related keywords were used to query BBC News search pages, from which article URLs were extracted and visited. The script downloaded each article's HTML, parsed key metadata (headline, date, section, and text), and consolidated all records into a structured CSV file for subsequent cleaning and analysis.

3.3 Number of Data Samples

After completing data collection and subsequent cleaning, the final dataset used for analysis contains: 274 news articles. Time period: December 1, 2024 to December 15, 2025 (post-election aftermath window) Distribution across sections are: World: 96 articles, Business: 90 articles, Technology: 88 articles

4. Data Cleaning, Analysis & Visualization

After collecting raw BBC News pages and metadata, I constructed a clean, analysis-ready dataset aligned with the project's research questions: comparing headline versus body sentiment, tracking post-election sentiment over time, and examining differences across sections (World, Business, Technology). The processing pipeline is fully reproducible, with intermediate outputs saved under data/processed/ and results/ for transparency and verification.

For data cleaning, the goal was to transform the scraped raw files into a consistent structured format while removing noise that could bias sentiment measurement. I first consolidated raw article metadata (URL, section, publication time, headline text, and a short body preview extracted from HTML) into a single table. Then I standardized the time fields by parsing the original published_at and converting it into a normalized published_date column (YYYY-MM-DD. I also removed duplicates using URLs as a unique identifier to avoid counting the same story multiple times. In addition, I filtered the dataset to the election aftermath window (2024-12-01 to 2025-12-15), which keeps the analysis aligned with the project premise (sentiment changes during the aftermath of the 2024 U.S. presidential election). The outputs of this cleaning step were stored as data/processed/articles_clean.csv and data/processed/articles_clean.jsonl, so the dataset can be reused for analysis without re-scraping.

For analysis, I used VADER sentiment scoring to compute emotional tone separately for headlines and for the article body text (represented by the cleaned body preview). VADER returns a compound score in the range [-1, 1], where values closer to -1 indicate more negative sentiment and values closer to +1 indicate more positive sentiment. This design directly supports the main research question of whether headlines are more emotionally charged than body text. For each article, I computed two key features: headline_compound and body_compound, and then assigned categorical sentiment labels (negative / neutral / positive) using consistent thresholds so I could compare not only average sentiment but also the proportion of each sentiment class. I also created a comparison variable headline_minus_body to measure direction and magnitude of difference within the same article, which makes the headline-vs-body comparison more interpretable than comparing raw averages alone.

To connect results to the "section" dimension in the proposal, I aggregated the sentiment features by section and computed descriptive statistics such as the number of articles per section, average headline sentiment and average body sentiment within each section, and the distribution of sentiment labels. To connect results to the "time" dimension, I aggregated sentiment by date (and section when needed), then used rolling averages to smooth daily fluctuations and better show trends over the event window. All summary outputs were written to results/summary_section.csv and results/summary_time.csv, and the full scored dataset was written to data/processed/articles_with_sentiment.csv so that anyone can rerun visualizations without recomputing sentiment.

For visualization, I created figures that directly correspond to the planned visualizations described in the proposal, and I used Matplotlib to generate reproducible PNG outputs saved under results/. First, I plotted the number of articles per section to show the dataset composition and to make sure later comparisons are interpreted in context, shown in figure 1

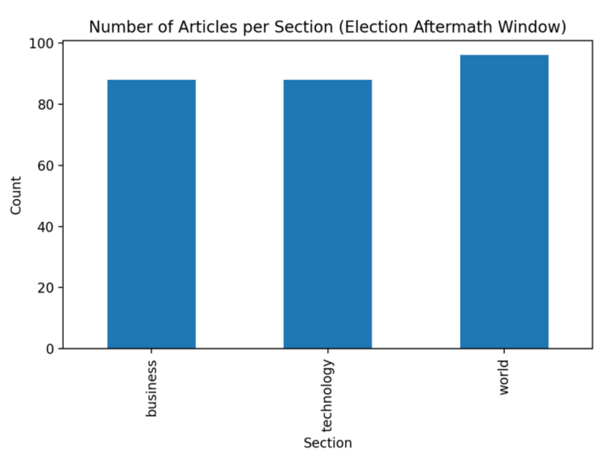
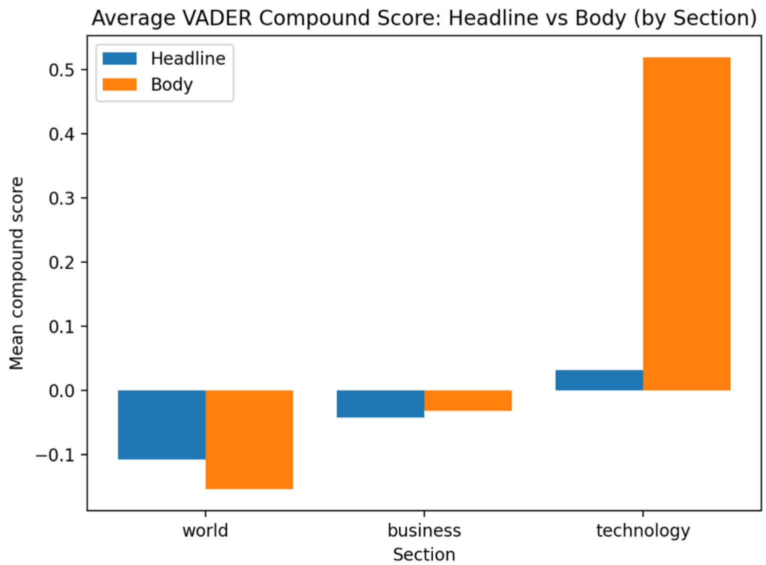


Figure 1

Second, I created a grouped bar chart comparing the average VADER compound sentiment scores of headlines and article bodies across sections, as shown in Figure 2.

This visualization reveals clear section-level differences in how sentiment is expressed in headlines versus article bodies.

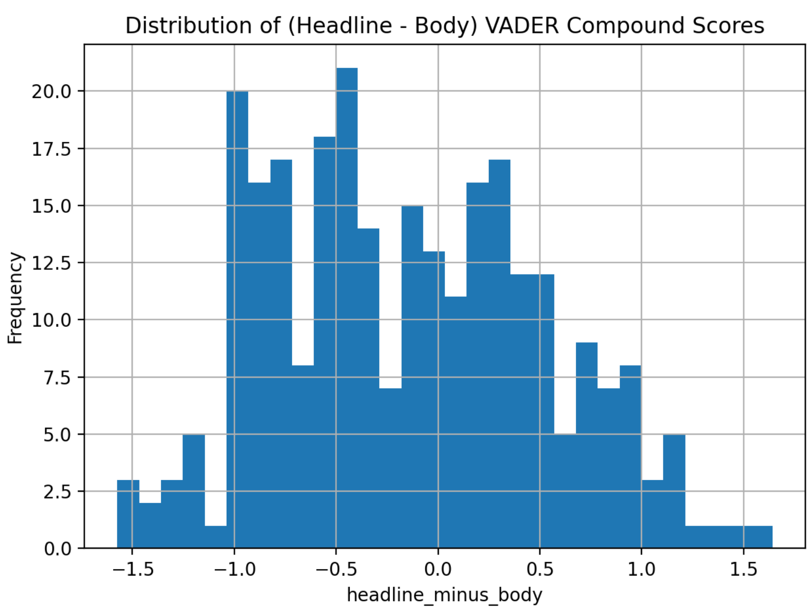
In the technology section, body text exhibits substantially higher average sentiment than headlines, suggesting that while technology headlines remain relatively restrained, the article content itself tends to emphasize positive developments and outcomes. In contrast, both the world and business sections show more muted sentiment overall, with headlines and bodies clustered closer to neutral or slightly negative values. Notably, in the world section, headlines are slightly less negative than the body text, which may reflect editorial framing that softens complex or adverse global events at the headline level.



Third, I plotted a histogram of the difference between headline and body sentiment scores (headline − body), as shown in Figure 3, to examine how sentiment diverges within the same article. In this figure, the x-axis represents the difference in VADER compound scores, where positive values indicate that the headline is more positive than the body text, and negative values indicate that the headline is more negative. The y-axis shows the frequency of articles falling into each difference range.

Figure 2

The distribution is spread across both positive and negative values, rather than being tightly centered around zero, indicating that headline and body sentiment frequently diverge instead of matching closely. A noticeable concentration of values appears on the negative side of the distribution, suggesting that headlines are often more negative than their corresponding body text. At the same time, the presence of a right tail with positive values shows that the opposite pattern also occurs, with some headlines adopting a more positive tone than the article content.



Fourth, I visualized the proportions of sentiment labels (negative, neutral, and positive) for both headlines and article bodies across sections, as shown in Figure 4 below.

The headline plot shows clear section-level differences: Technology headlines are more often neutral, while World headlines contain a higher share of negative sentiment, indicating topic-specific framing.

Figure3

In contrast, body text is more polarized,with fewer neutral articles and stronger positive or negative sentiment, especially in the World and Business sections.

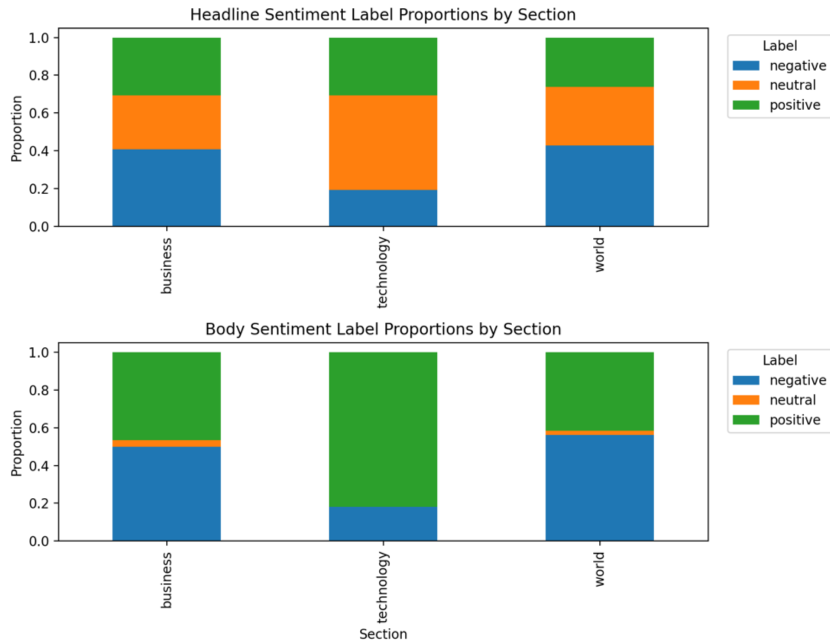


Figure 4

Finally, I visualized sentiment over the post-election period using a 7-day rolling mean of VADER compound scores (Figure 5). The time-series shows that sentiment varies substantially over time across all sections, indicating that emotional tone responds dynamically to events.

The relationship between headline and body sentiment also changes over time: in some periods the two closely align, while in others they diverge, suggesting selective emotional framing in headlines. Section-level differences are evident as well, with Technology showing higher and more volatile body sentiment and World exhibiting sharper negative swings near the end of the window.

Overall, this temporal analysis reinforces the project's premise that sentiment patterns evolve over time, differ by section, and are not always aligned between headlines and body text during the election aftermath.

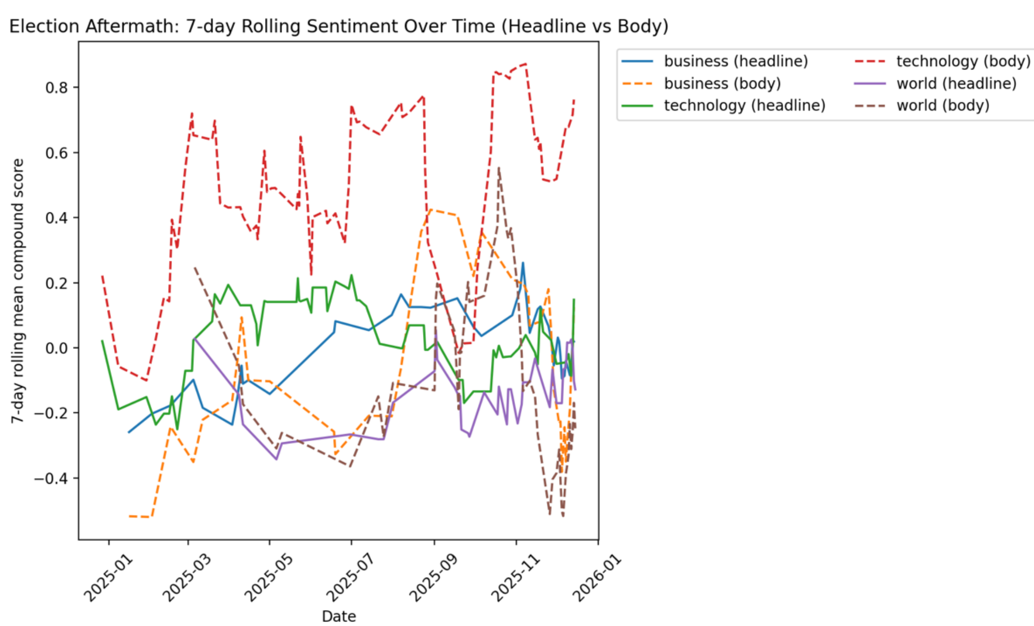


Figure 5

Overall, the results support the project premise. Sentiment varies across sections, and headline sentiment often differs from body sentiment in both magnitude and direction, consistent with headlines serving a framing and attention-grabbing role while body text is more informational. Time-series analyses further show that sentiment changes over the post-election period rather than remaining constant. Together, the analyses and visualizations demonstrate that news tone differs by section, diverges between headlines and bodies, and evolves over time during the election aftermath.

5. Changes from Original Proposal

During the course of the project, several adjustments were made to ensure that the analysis was technically feasible, analytically meaningful, and aligned with the course expectations. Early in the process, the scope of the study was refined to focus on a clearly defined and event-centered time window—the aftermath of the 2024 U.S. presidential election (2024-12-01 to 2025-12-15). This refinement provides a concrete temporal context for the analysis, strengthens interpretability of time-series patterns, and avoids overly broad or ambiguous conclusions that could arise from an unspecified time period.

To strengthen the analytical depth of the project, the final design incorporated a direct comparison between headline sentiment and article body sentiment. Rather than assigning a single sentiment score per article, sentiment was computed separately for headlines and body text, allowing the analysis to examine differences in emotional framing within the same article.

In addition, my original proposal did not specify dataset size. In the completed project, the dataset was expanded and structured to include several hundred articles across the World, Business, and Technology sections, with reasonably balanced coverage. This ensured a sufficient sample size for section-level comparisons and reliable aggregation over time.

6. Future Work

If more time or resources were available, several extensions could further strengthen this project. One natural direction would be to expand the data sources beyond BBC News and include additional outlets such as CNN, Reuters, or The Guardian. Comparing sentiment patterns across different news organizations would allow the analysis to examine whether headline framing and emotional tone vary systematically by outlet, rather than being driven by a single editorial style.

Another potential extension would be to enrich the text analysis beyond lexicon-based sentiment scoring. While VADER provides a transparent and interpretable baseline, future work could explore more advanced language models or topic-aware methods to better capture context, nuance, and framing in professional news writing. This could help distinguish whether sentiment differences are driven by specific topics, narrative structures, or event-driven language.

Finally, the time-series component of the project could be extended by collecting data over a longer horizon or by focusing on multiple major events. This would allow for more robust trend analysis and enable comparisons between different types of events, such as elections, geopolitical crises, or economic announcements. Together, these extensions would deepen the analytical scope of the project and provide a more comprehensive view of how news sentiment evolves across outlets, topics, and time.