The New York Times

OZEMPIC IN FOCUS: NYTIMES ARTICLES VS. COMMENTS

Applying NLP Across Media and Users

PRESENTED BY:
Akansha, Arpit, Olabanji, Nikhil & Safae

Problem Statement

The goal of this analysis is to examine NYTimes user comments on articles about Ozempic using various Natural Language Processing (NLP) techniques. This study aims to uncover patterns in reader discussions, including sentiment, key themes, and linguistic trends.

Objective # 1

Extract and preprocess NYTimes user comments related to Ozempic.

Objective # 2

Use NLP techniques like Sentiment Analysis, Topic Modeling, NER, and Word Frequency/N-gram Analysis to study text.

Objective # 3

Compare sentiment patterns and discussion trends across different articles.

Ozempic

What is it?

Ozempic is a prescription medication primarily used to treat type 2 diabetes, but it has gained significant attention for its off-label use in weight loss due to its ability to regulate appetite and blood sugar levels.

What are we trying to find?

By analyzing comments from NYT articles related to Ozempic, we aim to identify public sentiment, key concerns, and emerging themes surrounding its use. This includes understanding opinions on its effectiveness, side effects, accessibility, ethical considerations, and societal perceptions of weight-loss drugs.

NLP Technique: Topic Modelling

What is it?

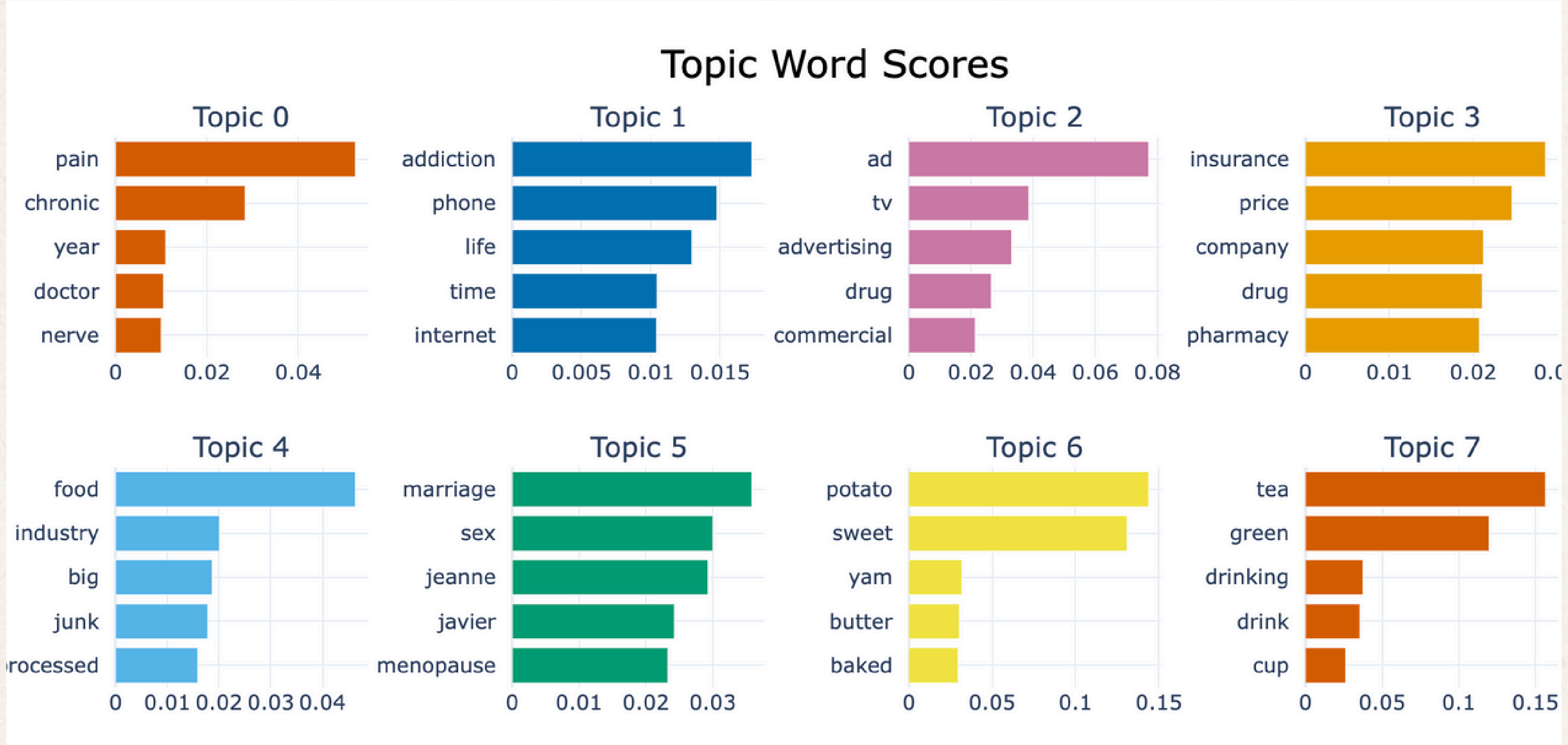
Topic modeling is an unsupervised Natural Language Processing (NLP) technique used to discover hidden topics in a large collection of text. It helps in organizing, summarizing, and structuring textual data by identifying groups of words that frequently occur together.

Technique Used: BERTopic

BERTopic leverages BERT embeddings along with UMAP for dimensionality reduction and HDBSCAN for clustering. This approach has helped me extract more meaningful and coherent topics from unstructured text data, making it particularly effective for analyzing short texts like customer feedback, social media posts, and business documents.

The New York Times

Topic Modelling



What do these topics represent?

- Topic 0 - Chronic Pain & Healthcare: Ozempic side effects like nausea and long-term health impacts.
- Topic 1 - Technology & Addiction: Viral social media attention on Ozempic for weight loss reliance.
- Topic 2 - Advertising & Media: Ozempic ads shape consumer views (ad, TV, drug keywords).
- Topic 3 - Pharma & Pricing: High cost and insurance debates (insurance, price, company keywords).
- Topic 4 - Food & Junk Food: Ozempic linked to reduced junk food cravings.
- Topic 5 - Marriage & Relationships: Weight loss affects relationships and body image.
- Topic 6 - Sweet Potatoes & Cooking: Ozempic shifts diets, reducing carb cravings like sweet potatoes.
- Topic 7 - Tea & Beverages: Users note changes in habits, favoring tea and lighter foods.

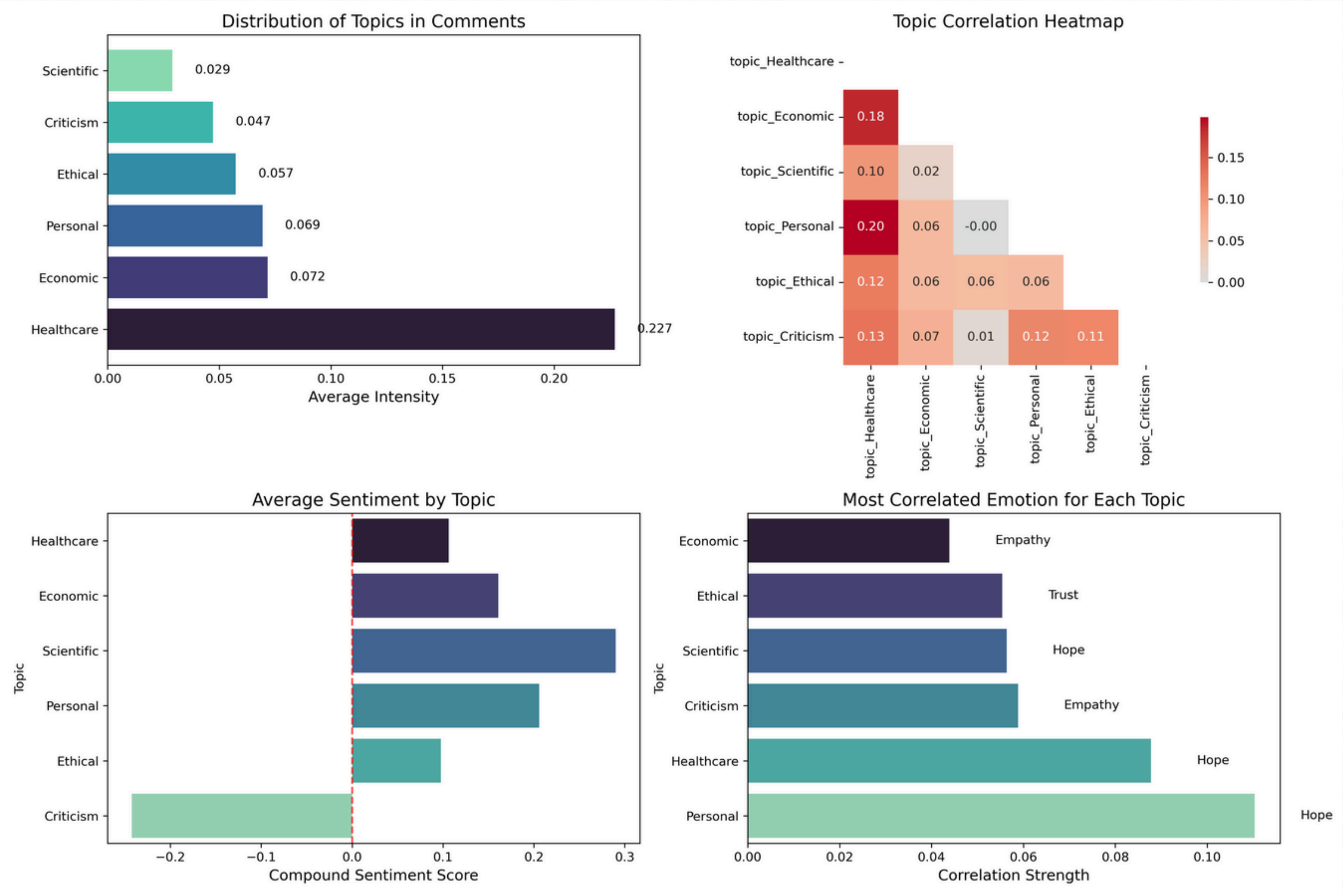
Why 'drug' appears twice

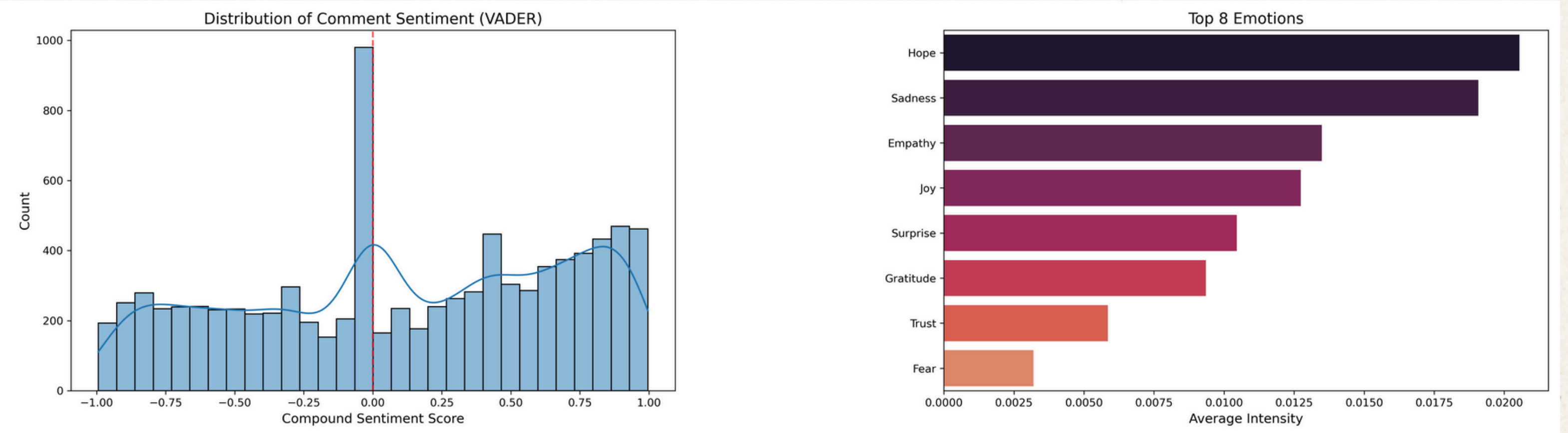
- Topic 2 (Advertising & Media) → How Ozempic is marketed and promoted.
- Topic 3 (Pharmaceutical Industry & Pricing) → How Ozempic is priced and regulated.

This reinforces the dual nature of Ozempic's popularity—its media hype and its cost/availability challenges.

The New York Times

NLP Technique: Emotional Analysis





Key Findings

Emotion Insights

Hope (0.021) and Sadness (0.019) lead in comments, followed by Empathy and Joy, showing emotional engagement. Anger and Disgust are minimal, suggesting constructive discourse.

Emotional Distribution

VADER shows a bimodal split: neutral (0) and positive (0.75). Scientific topics are most positive (0.25), Criticism slightly negative (-0.05), with most topics leaning positive.

Topic-Emotion Relationships

Hope ties to Healthcare, Personal, and Scientific topics. Empathy links to Economic and Criticism, showing connection. Trust aligns with Ethical topics, tied to healthcare morals.

Topic Analysis

Healthcare (0.227) dominates, followed by Economic (0.072) and Personal (0.069) topics. Scientific content gets less focus, indicating emphasis on access over research.

NLP Technique: Sentiment Analysis

What is it?

Sentiment analysis is a Natural Language Processing (NLP) technique used to determine the emotional tone behind a body of text. It helps identify whether the expressed opinion is positive, negative, or neutral, providing quick insights into the overall attitude or mood in the text.

Technique Used: Hugging Face Transformers

We utilized a pretrained Hugging Face model (DistilBERT fine-tuned on SST-2) for sentiment classification. This model reads each article's text and outputs a sentiment label (Positive, Negative, or sometimes Neutral) along with confidence score.

Why It's Useful:

- Quickly Gauges Tone: We can see if coverage around Ozempic skews positive or negative.
- Scalable: Works on multiple articles at once, letting us track sentiment trends across a dataset.
- Easy to Implement: Hugging Face pipeline allow us to integrate powerful language models with just a few lines of code.

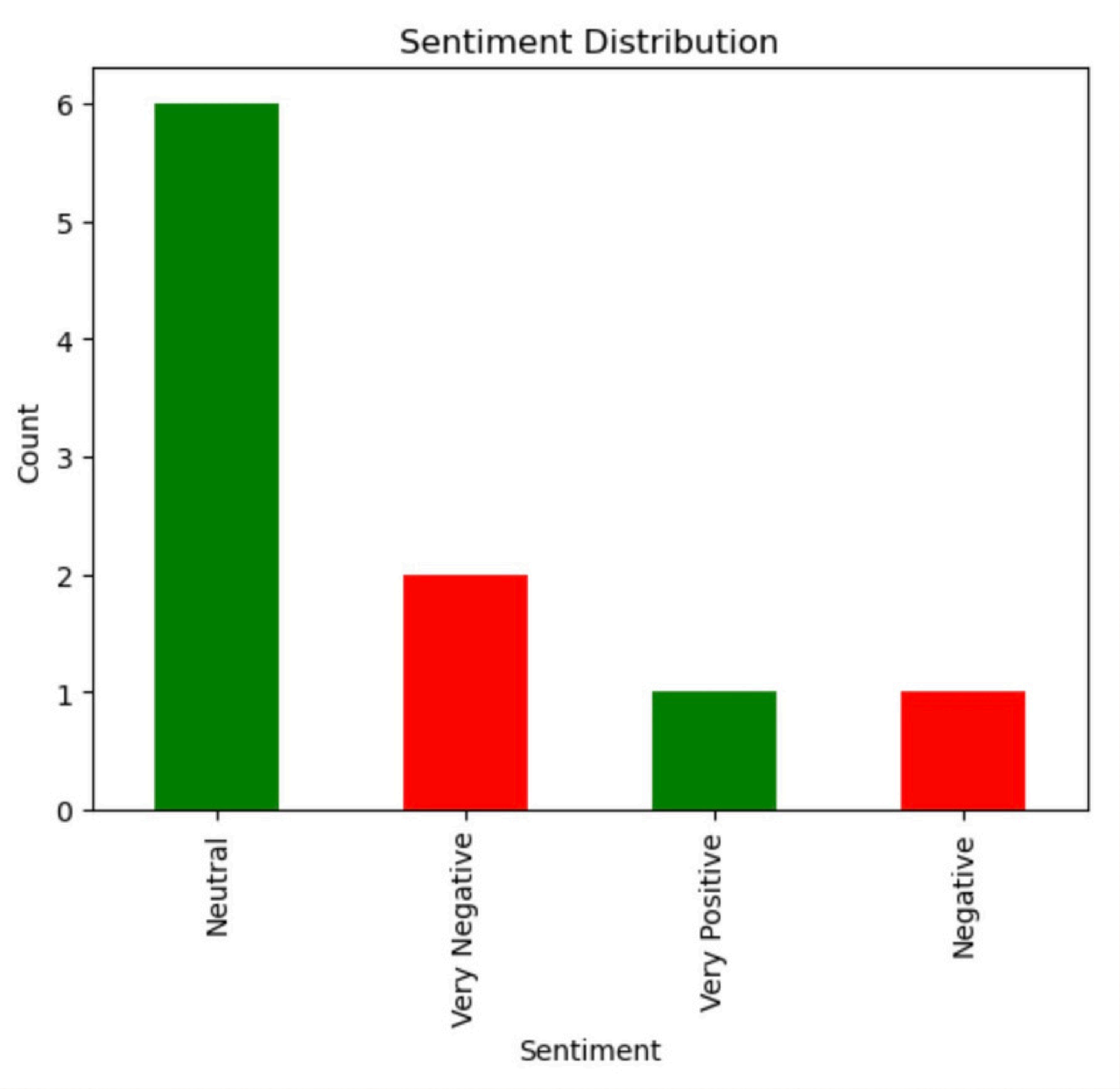
NLP Technique: Sentiment Analysis

Main Observations

- Majority are Neutral: Most articles present a balanced or informational tone rather than a strongly positive or negative stance.
- Noticeable Negative Proportion: A significant share of articles lean negative, possibly discussing concerns (e.g., side effects, cost, controversies).
- Smaller Positive Portion: Fewer articles highlight positive aspects (e.g., effectiveness for treating conditions, success stories).

Limitations

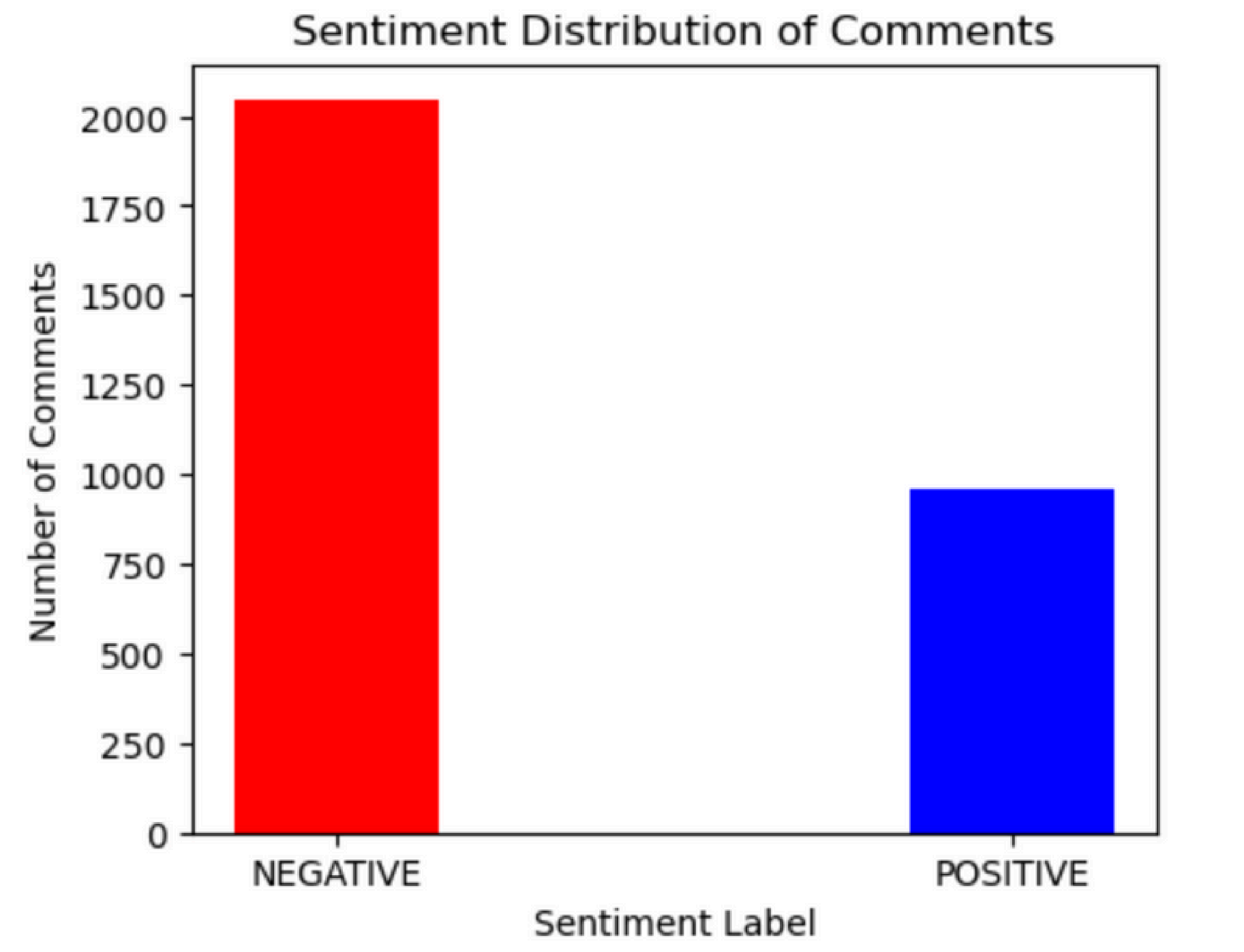
- Context Matters: Medical/health-related topics often mix facts and opinions. A "neutral" label may indicate straightforward reporting.
- Model Limitations: Sentiment Analysis can oversimplify nuanced topics–especially in medical or policy-related articles.
- Further Analysis: We could dive deeper into each sentiment category to see which themes (e.g., cost, accessibility, side effects) drive these sentiments.



The New York Times

DistilBERT vs. T5: A Comparative Analysis

comment	sentiment	topic
True, anti-immigrant politics are broadly popular in many societies, including Denmark. But it also ties center left parties to its right wing partners, and their politics. Instead of the usual red block coalition, Mette Frederiksen decided to make a coalition with centrist and center right parties for her latest government. As a consequence, her party's popularity has dropped significantly and looks to lose to	: True, anti-immigrant	: Right wing and center right parties don't want right wing policies.
ocessing workers.) But there,Äôs nothing inherently anti progressive about fair and reasonable restrictions on immigration.		



Observations and Limitations

- Upon conducting sentiment analysis with both models, DistilBERT proved to be fast and reliable, assigning clear Positive/Negative labels that align well with structured data visualization. However, it lacks depth in explaining sentiment beyond the classification.
- T5, on the other hand, generates contextualized sentiment explanations, capturing nuances in the text. While this makes it more human-like, it introduces variability and is computationally heavier, making structured analysis more challenging .
- Key Takeaway: DistilBERT is ideal for efficient, large-scale classification, while T5 offers richer sentiment insights but requires more refinement for structured use.

NLP Technique: Name Entity Recognition

What is it?

Named Entity Recognition (NER) is a Natural Language Processing (NLP) technique that identifies key entities such as people, organizations, locations, and products in a given text. It helps categorize and analyze which subjects are frequently mentioned.

Technique Used: spaCy

We used spaCy, a state-of-the-art NLP library, to extract named entities from both NYT articles and user comments. Additionally, a custom rule was added to ensure Ozempic is classified as a PRODUCT.

Why It's Useful:

- Identifies Key Subjects: Analyzes entities linked to Ozempic (e.g., companies, public figures, locations).
- Compares Sources: Reveals differences between media coverage and public discussions.
- Enables Deeper Analysis: Supports sentiment and topic modeling for further insights.

The New York Times

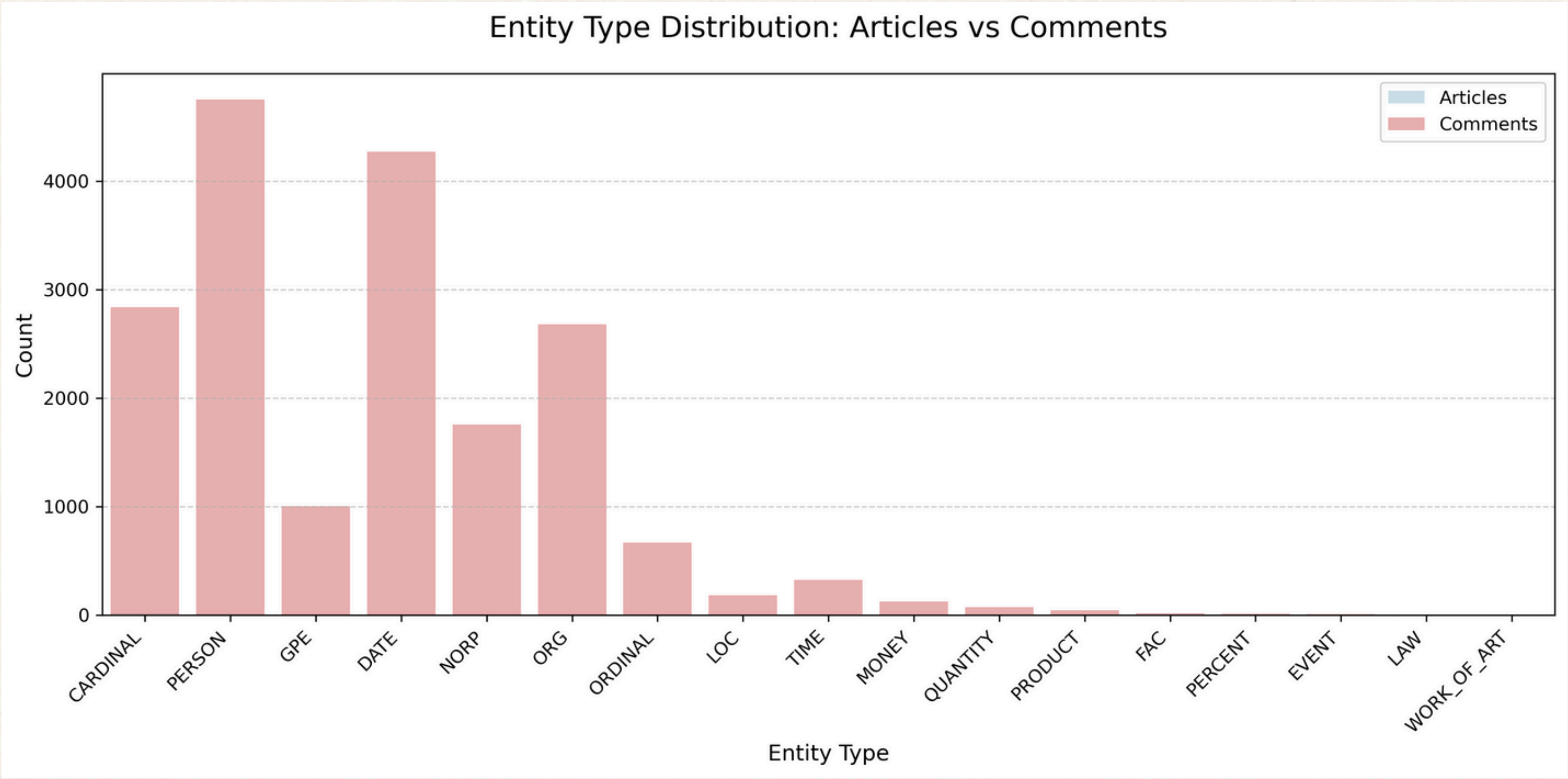
NLP Technique: Sentiment Analysis

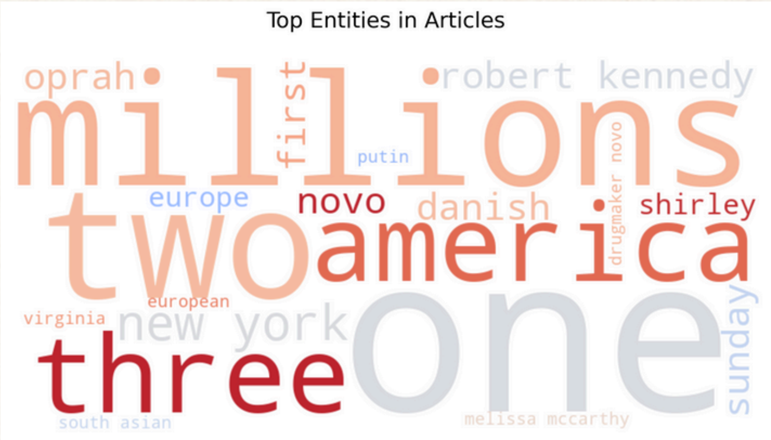
Main Observations

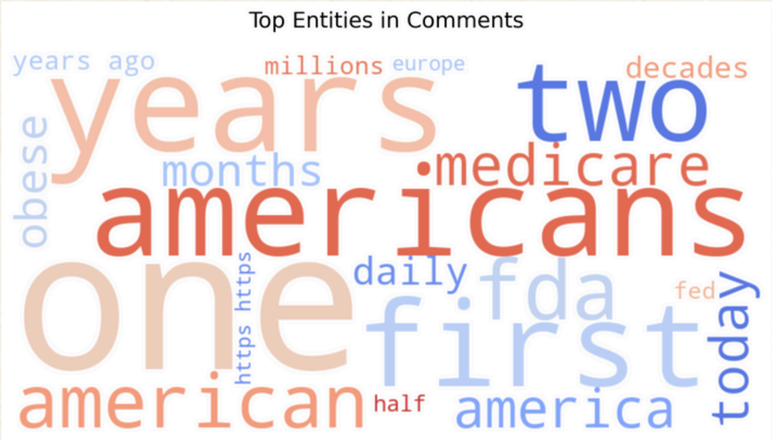
- Articles vs. Comments: Articles focus on organizations (Novo Nordisk, FDA) and geographic locations (America, Europe), reflecting a policy-driven narrative.
- Comments highlight personal concerns, with frequent mentions of Medicare, cost, and obesity, showing a focus on accessibility and impact.
- Word Cloud Insights: Articles emphasize corporate and regulatory entities.
- Comments reflect public discourse on health, affordability, and policy.

Limitations

- Different Perspectives: Media focuses on policy and companies, while comments emphasize personal stories and healthcare concerns.
- NER Challenges: Some detected words (e.g., "one," "two") lack context—filtering needed for better insights.
- Further Analysis: Co-occurrence analysis could reveal how entities relate to concerns like cost or regulation.







The New York Times

Results

The analysis shows a clear contrast between media coverage and public discussion on Ozempic. Articles focus on corporate and policy aspects, frequently mentioning Novo Nordisk and the FDA, while comments highlight personal concerns like Medicare, cost, and accessibility. Sentiment trends reveal mostly neutral coverage, but comments lean more negative, reflecting affordability and healthcare frustrations.

Conclusion

Media presents a structured, policy-driven narrative, while public discussions focus on personal impact and healthcare costs. This contrast highlights the gap between reporting and real-life concerns. Future research could explore entity relationships and sentiment trends to better understand public concerns over time.