A Longitudinal Study of Global Warming Discussion on Social Media

Student: CONNOR EMERY, Bryant University
cemery1@Bryant.edu

Mentor: DR. SUHONG LI, Bryant University
sli@Bryant.edu

Project owner: DR. GAURAV KHANNA, University of Rhode Island gkhanna@uri.edu

JULIE MA
jma@mghpcc.org

03/22/2024

CAREERS CYBERTEAM

A Longitudinal Study of Global Warming Discussion on Social Media

Description

- Study based on 26 million tweets collected from 2010-2022
- Project aims to analyze the change in global warming sentiment over the three stages of the COVID-19 pandemic
- Used word clouds, topic modeling and emotion analysis in order to evaluate the discussion
- Using python, mainly pyspark and associated packages in order to perform analysis

A Longitudinal Study of Global Warming Discussion on Social Media

Project goals

1. Identify patterns in conversation in history of global warming discussion

2. Use topic modeling in order to identify major topics that emerge in conversation

3. Use emotion scores in order to summarize change in emotion through history and determine how they have changed due to the pandemic

A Longitudinal Study of Global Warming Discussion on Social Media

October 2023 - March 2024

CAREERS CYBERTEAM

A Longitudinal Study of Global Warming Discussion on Social Media

What we accomplished

- Found patterns in global warming discussion using given dataset
- Completed emotion analysis as well as topic modeling
- Summarized results for future publication

CAREERS CYBERTEAM

Why Study Climate Change Discourse?

Climate change is a wide-reaching global issue that has devastating impacts on human life, from increasing temperatures to changing landscapes and food insecurity (NOAA)

Many people around the world are skeptical regarding climate change, instigated from former president Donald Trump

Social media platforms such as twitter have grown to be important for analyzing public sentiment

Where has the conversation been before?

Hedonometer and appraisal theories have been used in order to perform sentiment analysis along with global warming discussion. Hedonometer quantifies polarity of tweets while appraisal theory attempts to identify linguistic features. It was found that the discussion was mostly negative and expressed mostly with appreciation (Jiang et al., 2022).

A study based on 15 million tweets over 13 years found that there was a difference in discussion between climate change believers and deniers, notably with believers using the word “global warming” 5x less often compared to “climate change” (Effrosynidis et al., 2022)

CAREERS CYBERTEAM

Methodologies

Data collection

Tweets were collected by Dr. Li between 2010 and 2022, aggregating 26 million tweets in total using the keyword “global warming.”

Data cleaning

Special characters were stripped from the tweets and tokenized for analysis in order to maximize efficiency while processing. Words were converted to lowercase and lemmatization applied to get words to their base root

Data analysis

Data was analyzed using access granted from CAREERS cyberteam where clusters were used on MGHPCC computing cluster. Analysis performed consisted of top hashtags, top tweeters, word clouds, topic modeling and emotion analysis

CAREERS CYBERTEAM

A Longitudinal Study of Global Warming Discussion on Social Media

Period	Timeline	Tweet count
Pre	2010-2020	18,917,228
During	2020-2022	4,871,587
Post	2022-2022	3,175,376

Exploratory Data Analysis

Top hashtags

Pre-COVID

Word	Count
#tcot	120981
#science	79854
#actonclimate	77279

Top hashtags

During COVID

Word	Count
#climatecrisis	67058
#climateaction	48713
#climateemergency	42134

Top hashtags

Post-COVID

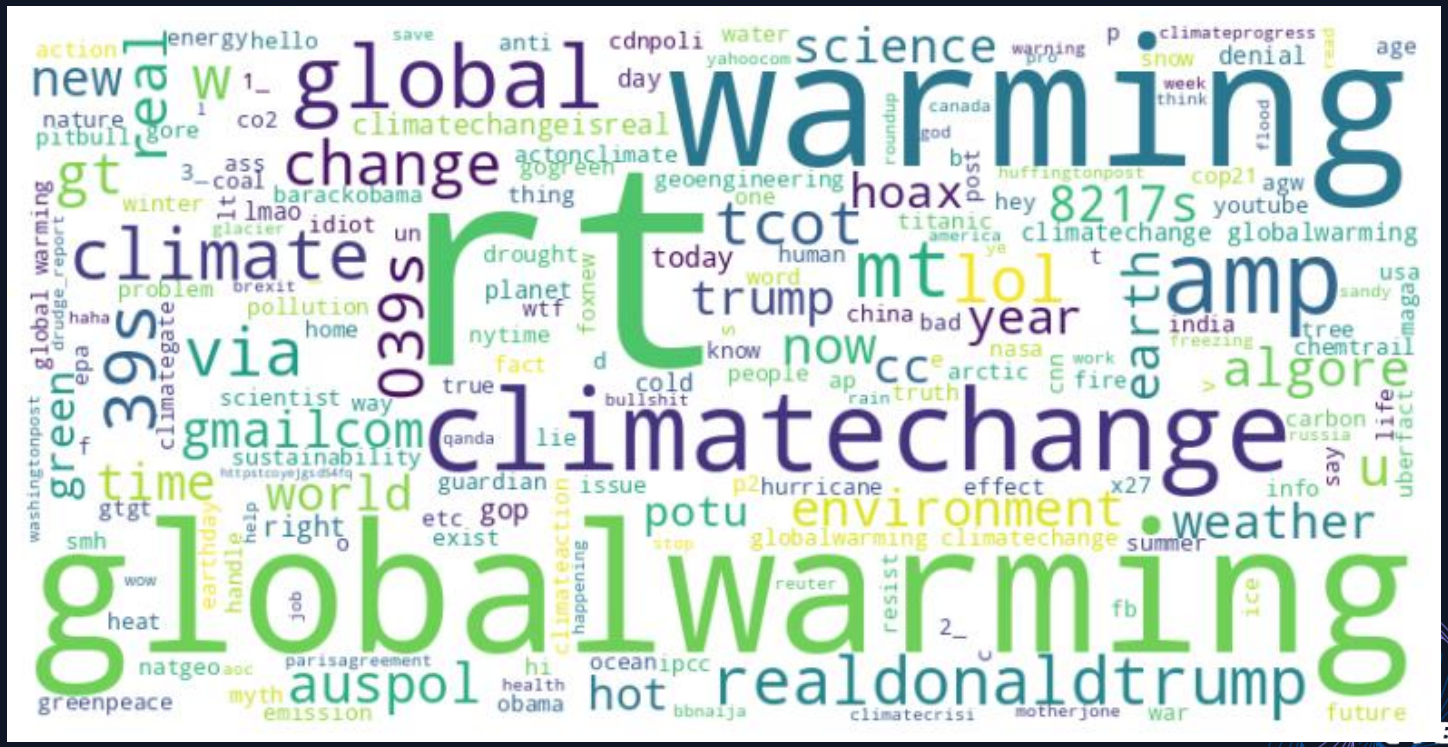
Word	Count
#climatecrisis	122342
#climate	95532
#climateaction	85867

CAREERS CYBERTEAM

Identifying dominant words in each phase

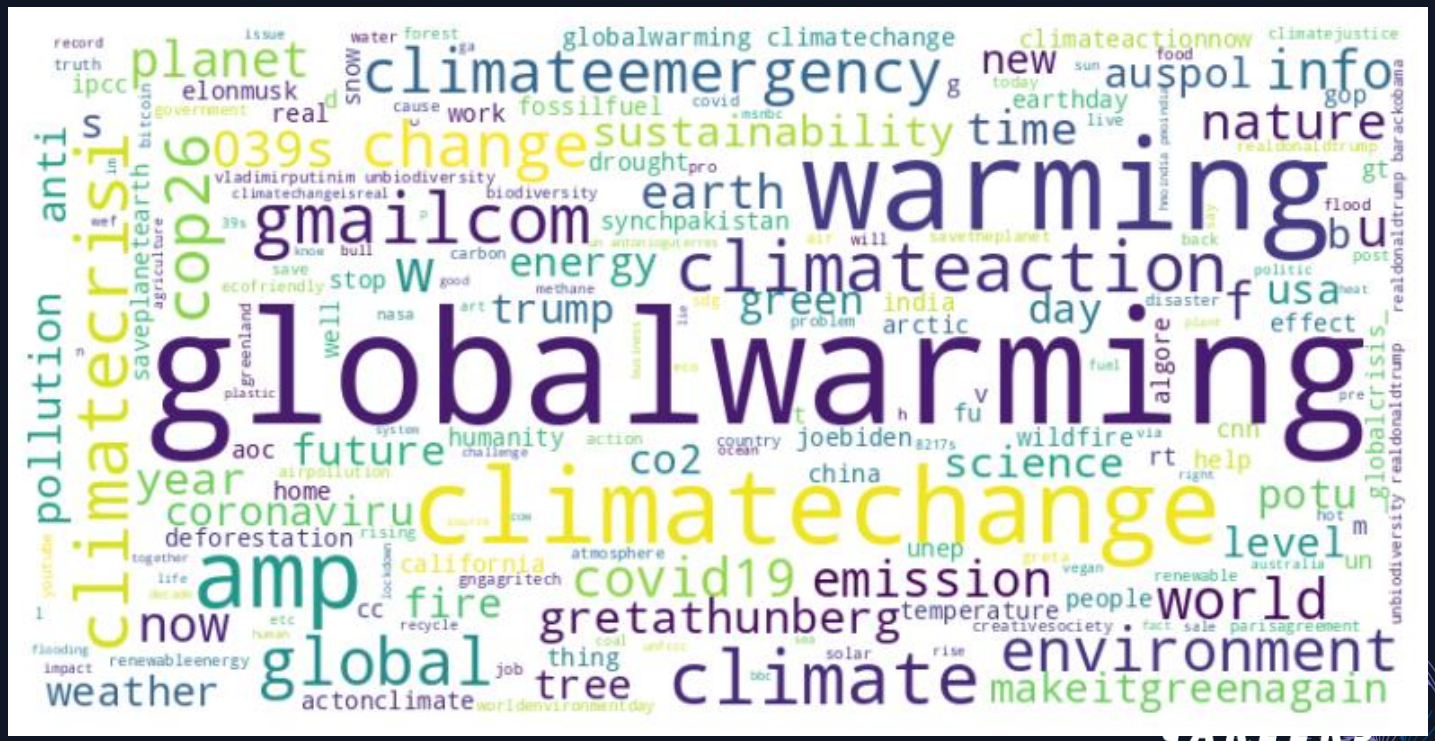
A Longitudinal Study of Global Warming Discussion on Social Media

Pre-COVID word cloud



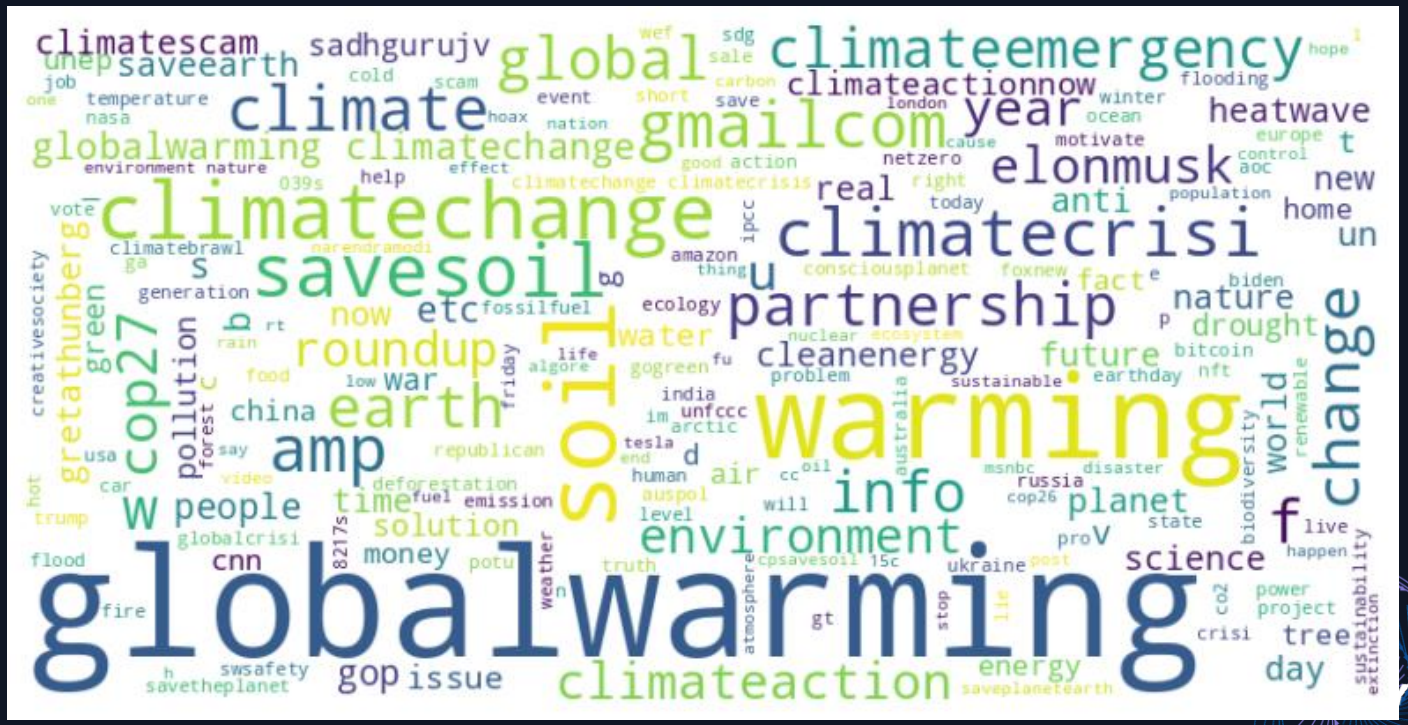
A Longitudinal Study of Global Warming Discussion on Social Media

During COVID word cloud



A Longitudinal Study of Global Warming Discussion on Social Media

Post-COVID word cloud

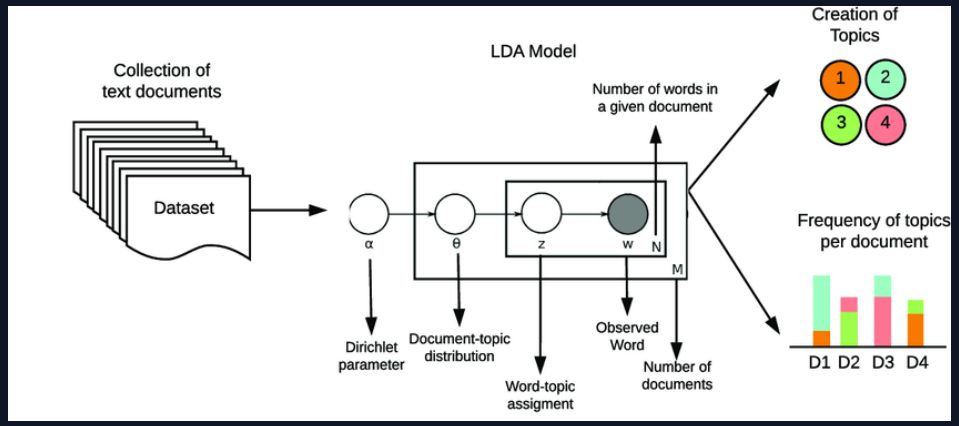


Identifying Dominant Topics

What is LDA topic modeling

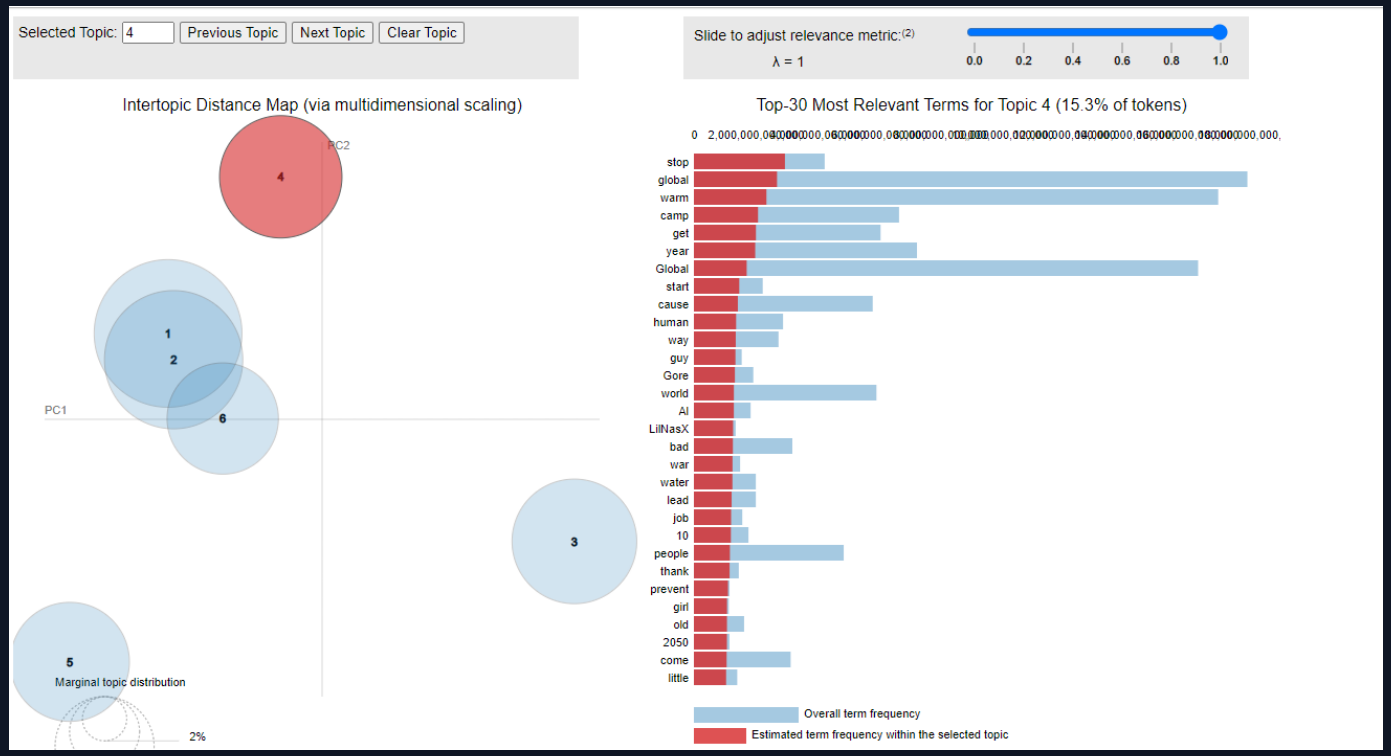
Unsupervised machine learning technique that scans a group of documents and finds patterns to cluster groups of words

LDA topic modeling is a generative probabilistic model that assumes that each document is a distribution over a set of words. Words are paired to other words that show up in similar contexts, forming a topic



CAREERS CYBERTEAM

Example of topic modeling output



CAREERS CYBERTEAM

A Longitudinal Study of Global Warming Discussion on Social Media

Pre-COVID Topics

Topic	Key terms
Observing	Global, temperature, cool, ice
Worry and anxious	Global, end, something, freak
Solutions	Think, global, extreme, mean

CAREERS CYBERTEAM

A Longitudinal Study of Global Warming Discussion on Social Media

During COVID Topics

Topic	Key terms
Climate action	Global, temperature, emission, increase
Impact and awareness	Human, fire, end, seriously
Solutions	Global, think, fact, pandemic

A Longitudinal Study of Global Warming Discussion on Social Media

Post-COVID Topics

Topic	Key terms
Worry	Global, CO2, tree, worry
Impact	Condition, temperature, effect, increase
Denier	Crazy, NASA, junkscience

CAREERS CYBERTEAM

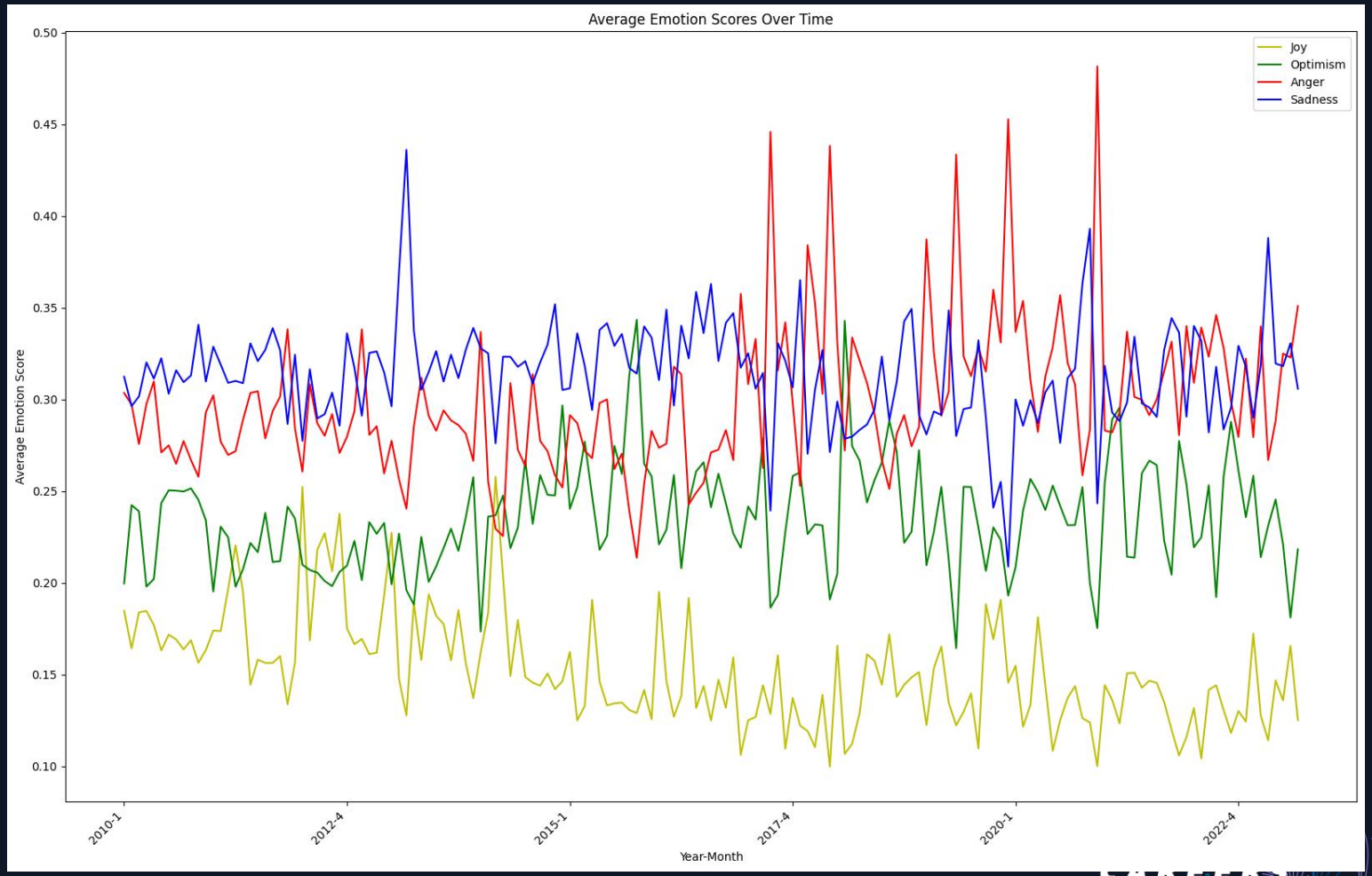
Emotion Analysis

Emotion analysis

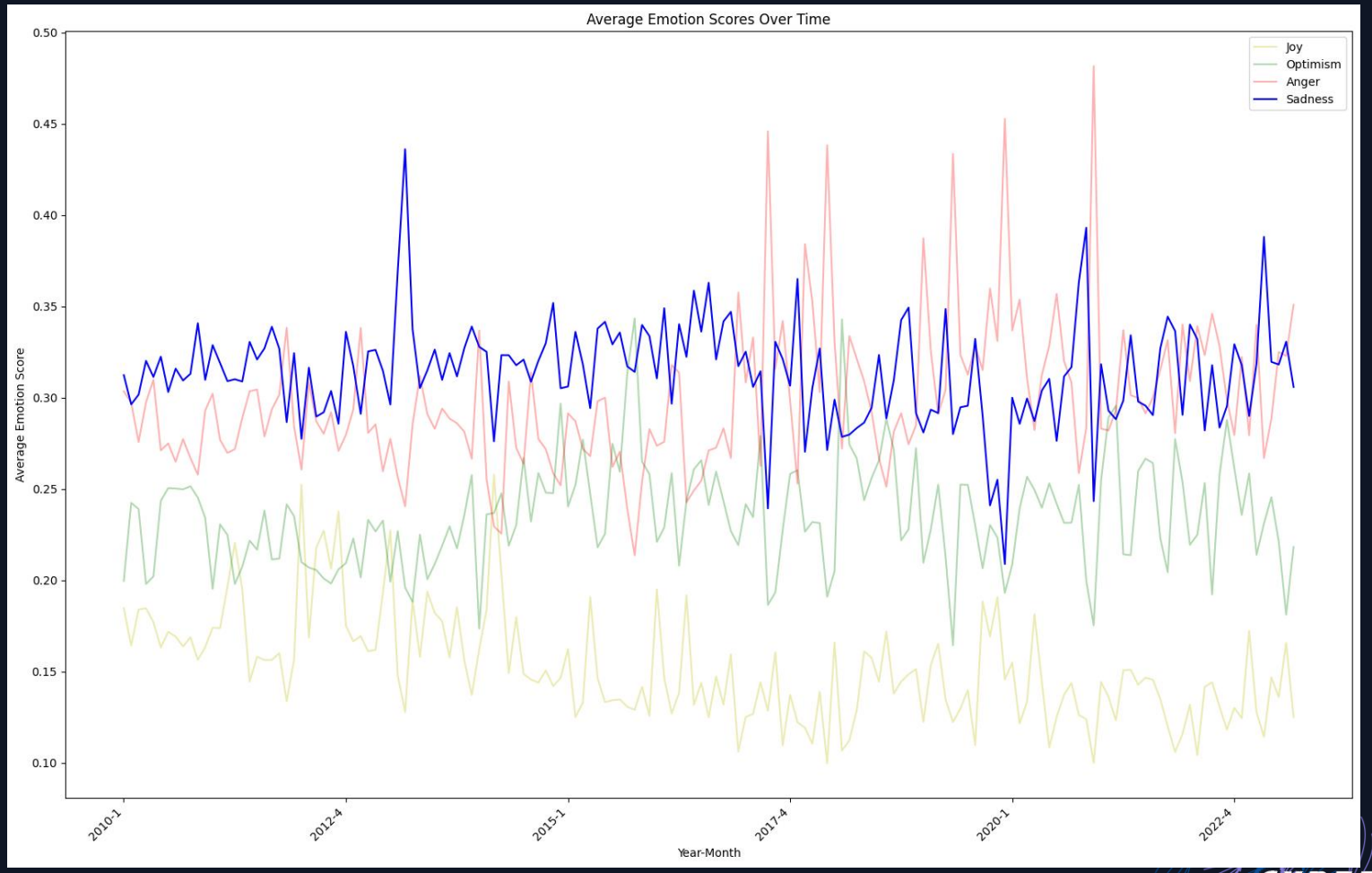
- Hugging Face (Twitter-roBERTa-base for Emotion Recognition).
- roBERTa-base model trained on ~58M tweets with the TweetEval benchmark.
- Each tweet is classified into four emotions (joy, optimism, anger, and sadness) with a confidence score.

CAREERS CYBERTEAM

Tweet emotion over time

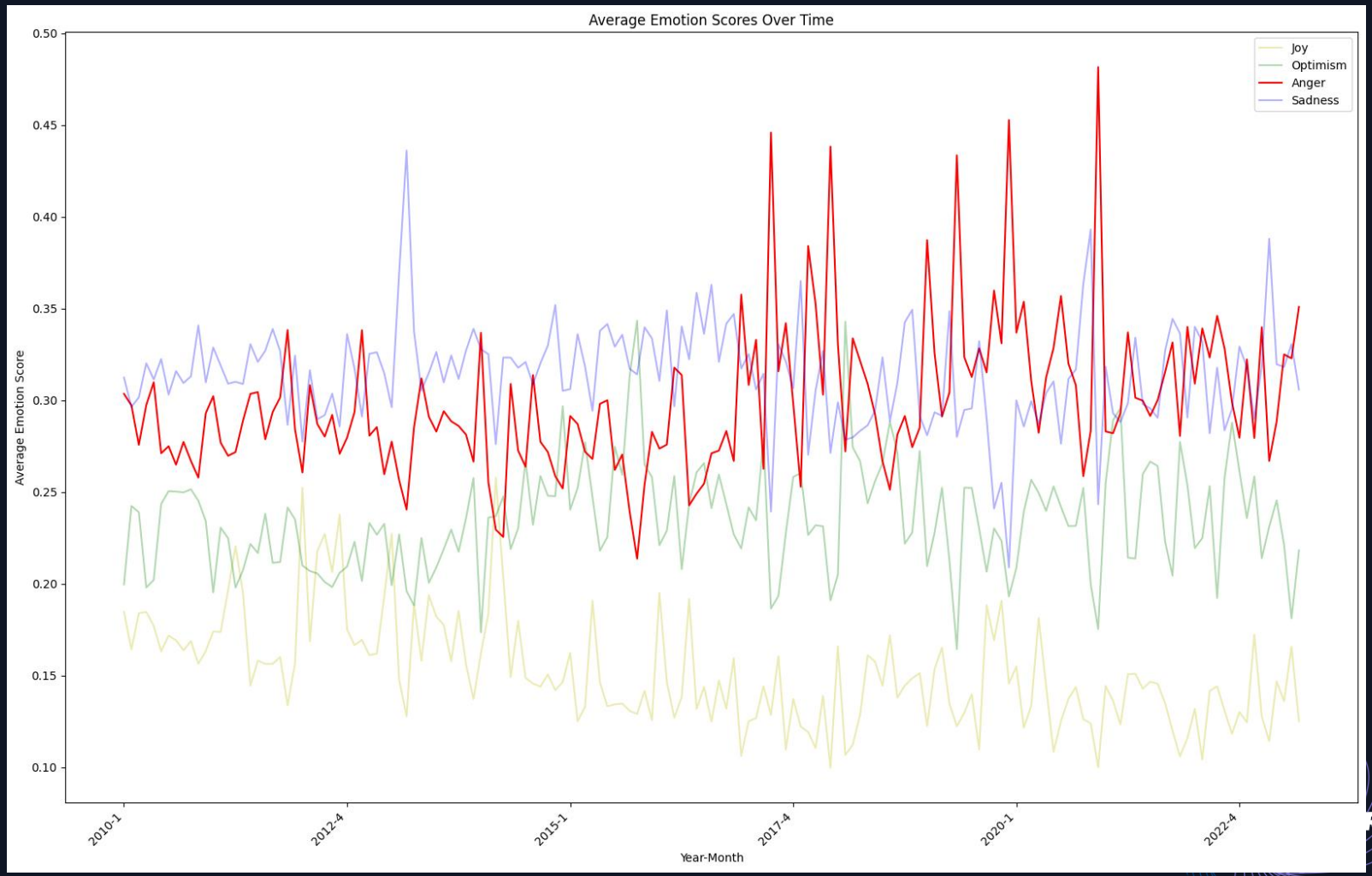


Sadness over time

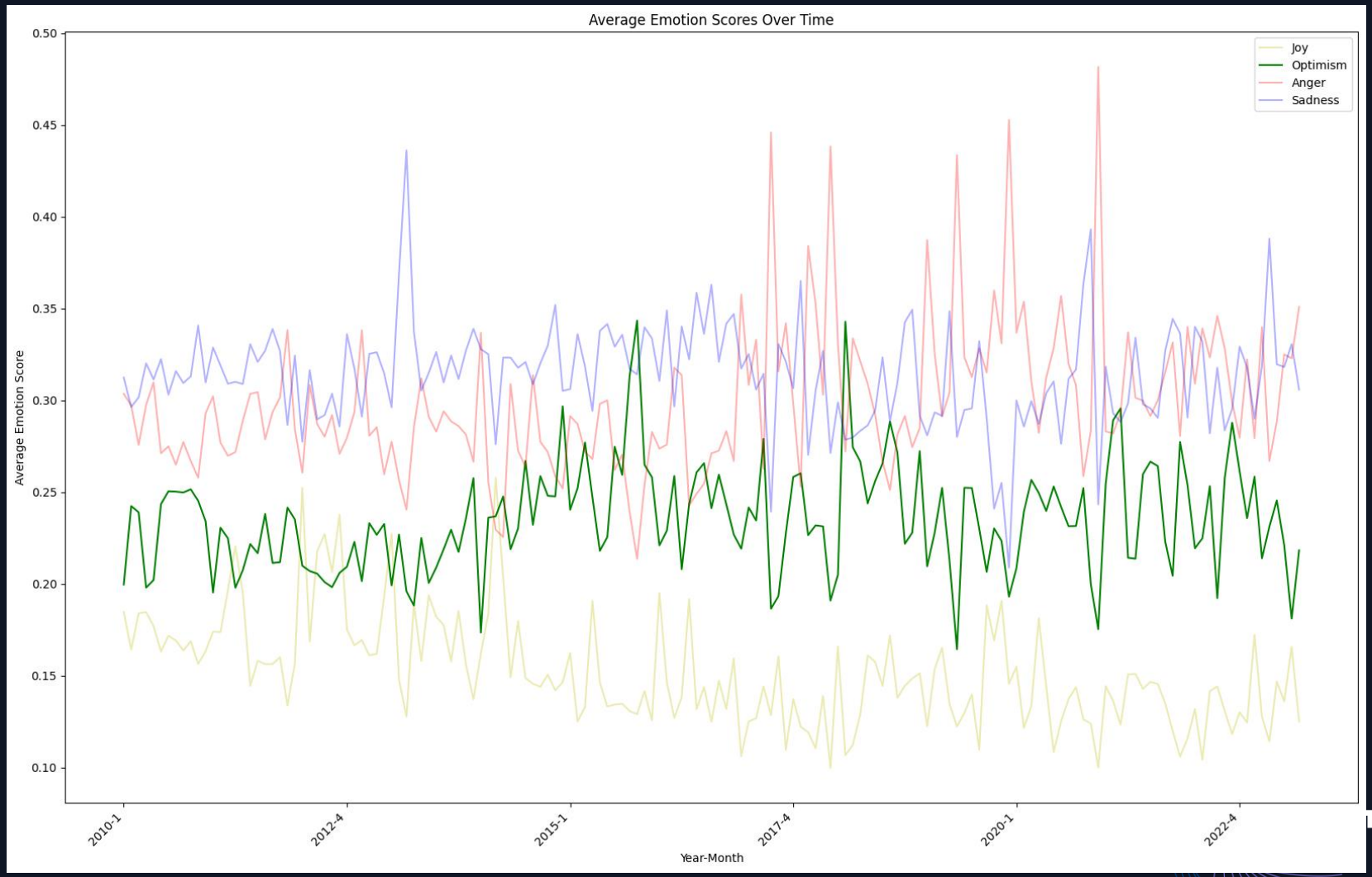


CYBERTEAM

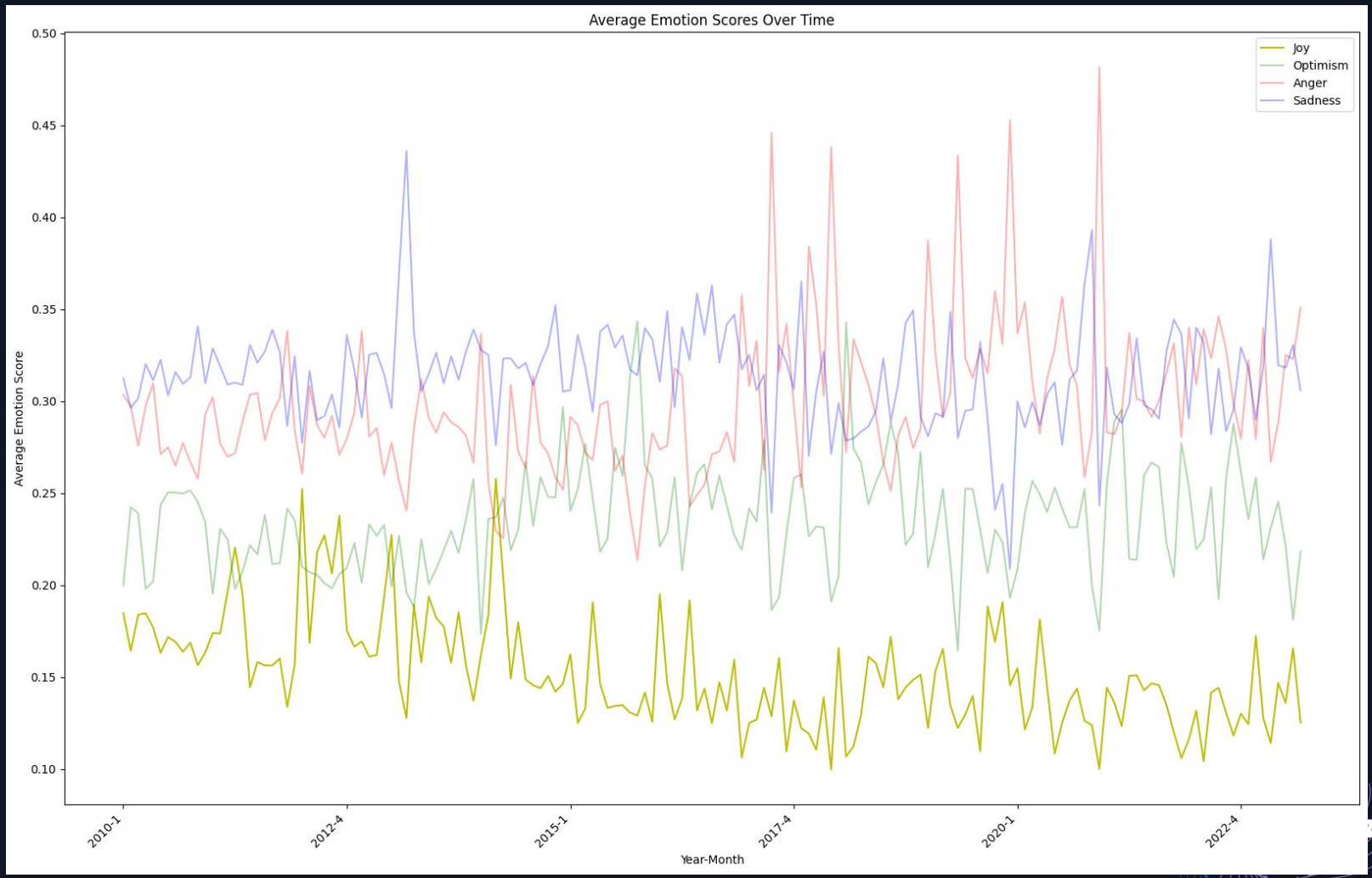
Anger over time



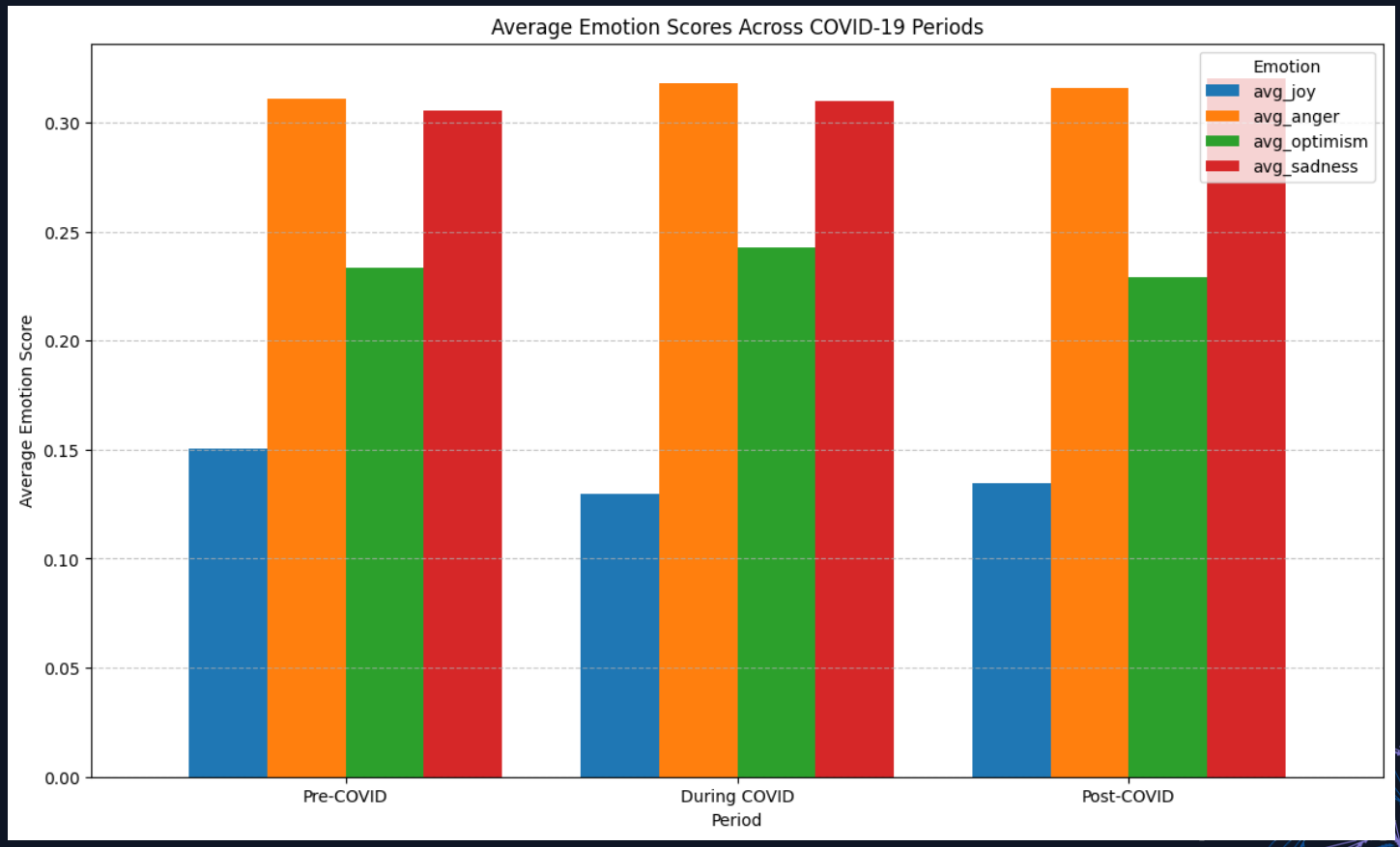
Optimism over time



Joy over time



Emotion of Climate Change Across Pandemic Phases



CYBERTEAM

Results and conclusion

Topic modeling suggests that people are more worried regarding the impact of climate change now compared to before the pandemic

Within the past 5 years the data suggests there has been a rise in anger, part of which can be contributed to the pandemic

Given emotion analysis, clear that climate discussion happens in peaks, likely aligned with extreme weather events that prompt a reaction

A Longitudinal Study of Global Warming Discussion on Social Media

What went well and what could have been done differently?

- Using HPC environment made analysis easier, however also came with its challenges
- Analysis was easy to interpret
- Could have narrowed focus of the research more

Limitations and Future Study

- Identify which emotions are the most prevalent over each topic and how these emotions change across periods
- Split dataset into probability of users being believers and deniers to determine underlying patterns
- Study was restricted to tweets that were in English, therefore restricting the number of tweets pulled

Lessons Learned

- Learned how to use UNITY effectively
- Successfully ran analysis and found patterns in discussion
- Could have made the research more specific

Publications/Contributions

Attended

- Analytics Without Borders, 2024

Plan to attend

- NEDSI 2024 April 4-6th
- PEARC 2024

Contributions to Research Computing Community

Please include links and topics!

- No contributions yet to the computing community
- Planning to contribute on GitHub soon
- 3/22-5/1

Thank you!
Any questions?