



# Analyzing Climate Sentiments

**A Big Data Approach to Understanding Public Opinion on Climate Change  
through Twitter**

**Group 7:**

Jyothi Prakash

Om Sri Rohith Raj Yadav Thalla

Shreya Sudini

# Introduction



- The Climate Change Twitter Dataset is a robust compilation of over 15 million tweets related to climate change, spanning 13 years. It provides a global perspective by including tweets from around the world, each associated with geolocation, user gender, climate change stance and sentiment, aggressiveness, temperature deviations, and topic modeling.
- This dataset can offer valuable insights to public opinions on climate change, sentiment analysis making it versatile for policymakers and government agencies to understand perception of general public to make improved policies.

# Project Objectives



- **Sentiment Analysis:** - Analyze the sentiment of tweets related to climate change to understand public opinion over time. Identify patterns or shifts in sentiment in response to significant climate-related events or disasters
- **Aggressiveness Analysis:** - Explore the relationship between aggressive language and climate change stances or sentiments.
- **Data Visualization** - Develop visualizations and dashboards to represent the analysis findings effectively, providing insights into public opinion and discourse on climate change.

## Data Source and Collection



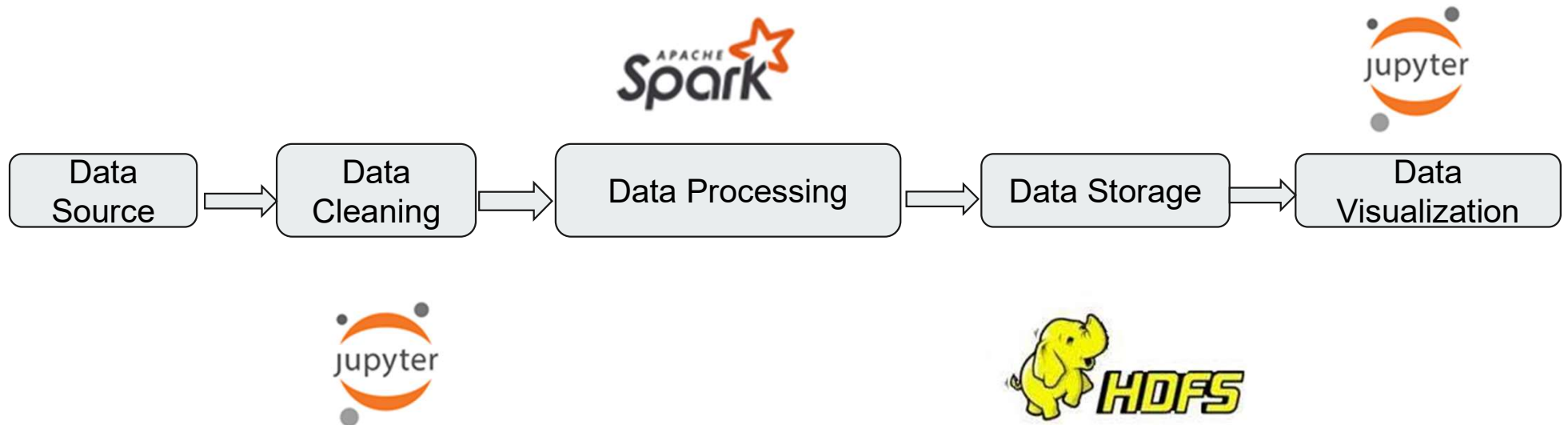
- Effrosynidis, Dimitrios (2022), “The Climate Change Twitter Dataset”, Mendeley Data, V1, doi: 10.17632/mw8yd7z9wc.1-
- Data size ~1.95 GB
- **Link to the source data:** <https://data.mendeley.com/datasets/mw8yd7z9wc/1>

## Data Attributes

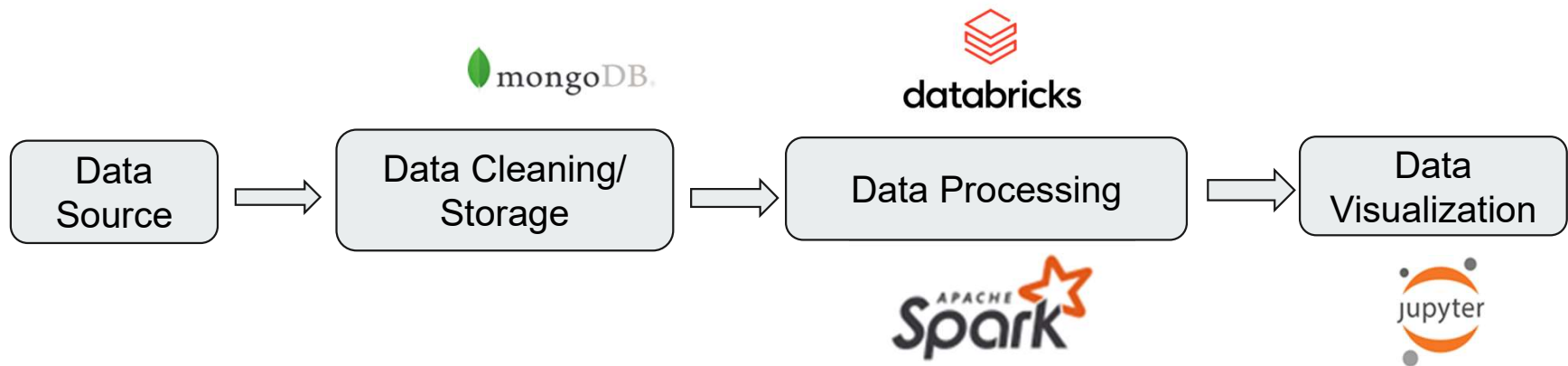


- **created\_at**: The timestamp of the tweet.
- **id**: The unique id of the tweet.
- **lng**: The longitude the tweet was written.
- **lat**: The latitude the tweet was written.
- **topic**: Categorization of the
- **Stance**: That is if the tweet supports the belief of man-made climate change (believer), if the tweet does not believe in man-made climate change (denier), and if the tweet neither supports nor refuses the belief of manmade climate change (neutral).
- **gender**: Whether the user that made the tweet is male, female, or undefined.
- **Temperature\_avg**: The temperature deviation in Celsius and relative to the January 1951-December 1980 average at the time and place the tweet was written.
- **aggressiveness**: That is if the tweet contains aggressive language or not.

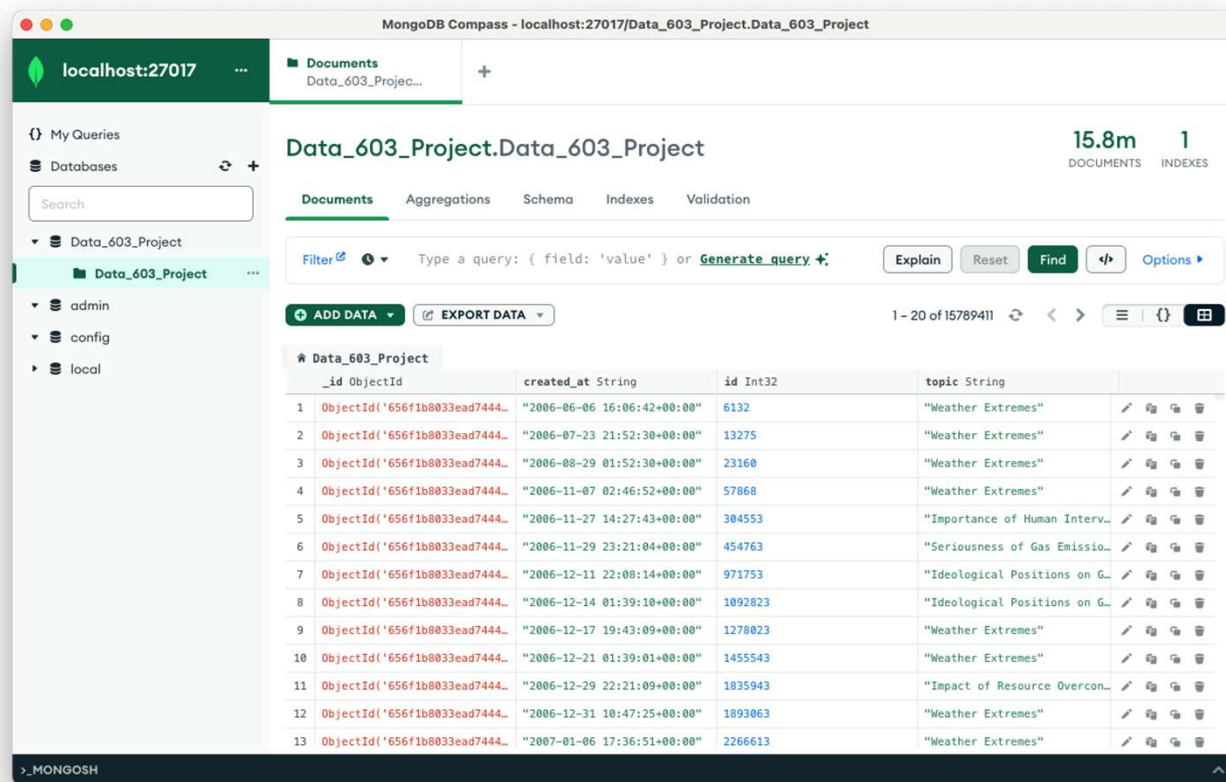
# Proposed Workflow



# Actual Workflow



# Data Feeding to MongoDB

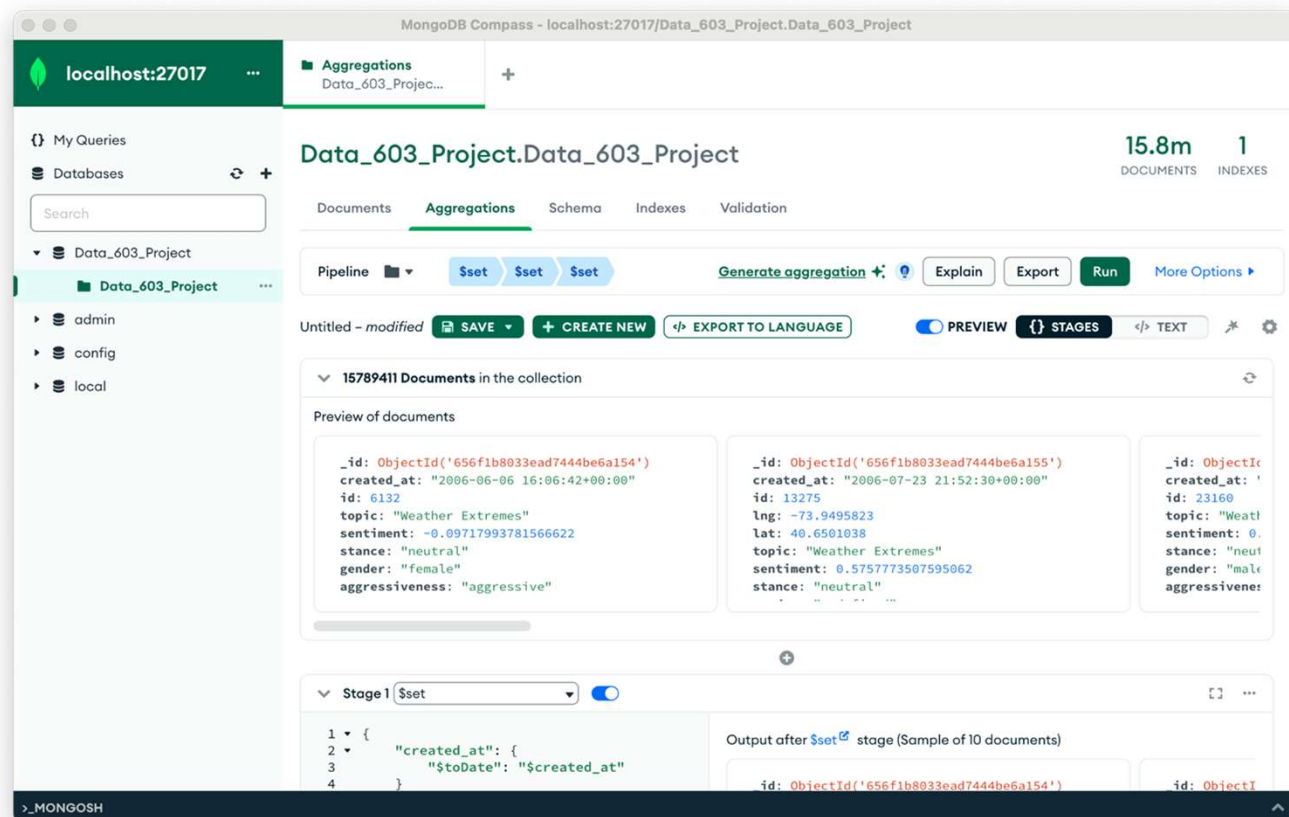


The screenshot displays the MongoDB Compass web interface. The top bar shows the connection to 'localhost:27017' and the selected database 'Data\_603\_Project'. The left sidebar lists the databases and collections, with 'Data\_603\_Project' selected. The main panel shows the 'Documents' tab for the 'Data\_603\_Project.Data\_603\_Project' collection, which contains 15.8 million documents and 1 index. A search bar is present with a filter icon and a query input field. Below the search bar, there are buttons for 'ADD DATA' and 'EXPORT DATA'. The document list is displayed in a table with columns: '\_id', 'ObjectId', 'created\_at', 'String', 'id', 'Int32', and 'topic', 'String'. The table shows 13 documents, each with a unique ObjectId, a timestamp, an integer ID, and a topic string. The topics include 'Weather Extremes', 'Importance of Human Interv...', 'Seriousness of Gas Emissio...', 'Ideological Positions on G...', and 'Impact of Resource Overcon...'. The bottom status bar shows the command prompt '>\_MONGOSH'.

	_id	ObjectId	created_at	String	id	Int32	topic	String
1	ObjectId('656f1b8033ead7444...	"2006-06-06 16:06:42+00:00"	6132	"Weather Extremes"				
2	ObjectId('656f1b8033ead7444...	"2006-07-23 21:52:30+00:00"	13275	"Weather Extremes"				
3	ObjectId('656f1b8033ead7444...	"2006-08-29 01:52:30+00:00"	23160	"Weather Extremes"				
4	ObjectId('656f1b8033ead7444...	"2006-11-07 02:46:52+00:00"	57868	"Weather Extremes"				
5	ObjectId('656f1b8033ead7444...	"2006-11-27 14:27:43+00:00"	304553	"Importance of Human Interv...				
6	ObjectId('656f1b8033ead7444...	"2006-11-29 23:21:04+00:00"	454763	"Seriousness of Gas Emissio...				
7	ObjectId('656f1b8033ead7444...	"2006-12-11 22:08:14+00:00"	971753	"Ideological Positions on G...				
8	ObjectId('656f1b8033ead7444...	"2006-12-14 01:39:10+00:00"	1092823	"Ideological Positions on G...				
9	ObjectId('656f1b8033ead7444...	"2006-12-17 19:43:09+00:00"	1278023	"Weather Extremes"				
10	ObjectId('656f1b8033ead7444...	"2006-12-21 01:39:01+00:00"	1455543	"Weather Extremes"				
11	ObjectId('656f1b8033ead7444...	"2006-12-29 22:21:09+00:00"	1835943	"Impact of Resource Overcon...				
12	ObjectId('656f1b8033ead7444...	"2006-12-31 10:47:25+00:00"	1893063	"Weather Extremes"				
13	ObjectId('656f1b8033ead7444...	"2007-01-06 17:36:51+00:00"	2266613	"Weather Extremes"				



# Data Cleansing using MongoDB



# Data in Databricks

The screenshot displays the Databricks web interface for a notebook titled "DATA\_603\_PROJECT". The interface includes a top navigation bar with the Databricks logo, a user profile, and a cluster selector showing "thalla's cluster (clone)". The notebook is written in Python and contains the following code:

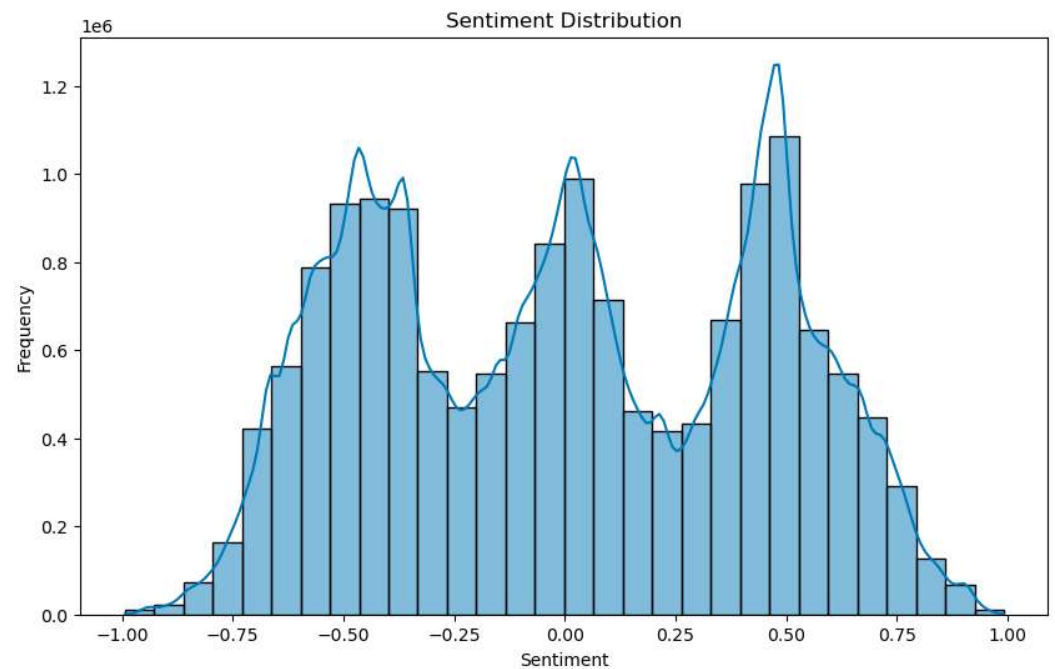
```
5 # CSV options
6 infer_schema = "false"
7 first_row_is_header = "false"
8 delimiter = ","
9
10 # The applied options are for CSV files. For other file types, these will be ignored.
11 df = spark.read.format(file_type) \
12     .option("inferSchema", infer_schema) \
13     .option("header", first_row_is_header) \
14     .option("sep", delimiter) \
15     .load(file_location)
16
17 display(df)
```

Below the code, the output shows "(2) Spark Jobs" and a summary of the DataFrame: "df: pyspark.sql.dataframe.DataFrame = [\_c0: string, \_c1: string ... 9 more fields]". A table view of the data is displayed below:

	_c0	_c1	_c2	_c3	_c4
1	_id	created_at	id	topic	score
2	656f1b8033ead7444be6a154	2006-06-06 16:06:42+00:00	6132	Weather Extremes	-0.5
3	656f1b8033ead7444be6a155	2006-07-23 21:52:30+00:00	13275	Weather Extremes	0.0
4	656f1b8033ead7444be6a156	2006-08-29 01:52:30+00:00	23160	Weather Extremes	0.0
5	656f1b8033ead7444be6a157	2006-11-07 02:46:52+00:00	57868	Weather Extremes	0.0
6	656f1b8033ead7444be6a158	2006-11-27 14:27:43+00:00	304553	Importance of Human Intervantion	-0.5

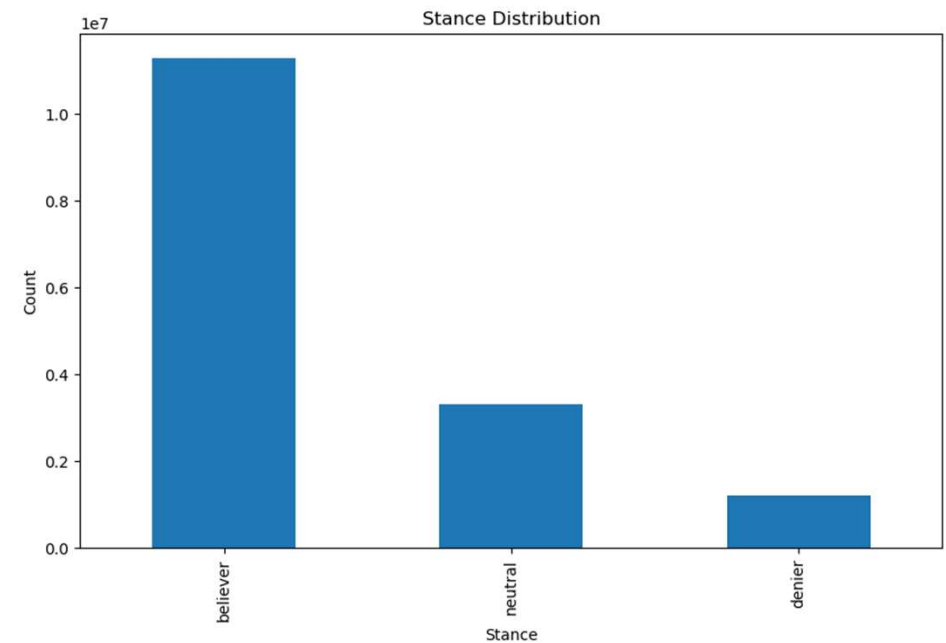
# Visualizations

- This histogram displays the frequency distribution of sentiment values across tweets. The distribution is bimodal, which indicates that there are two predominant groups in the dataset: one with a slightly negative sentiment and another with a slightly positive sentiment.
- The sentiment distribution suggests that while Twitter users express both positive and negative sentiments towards climate change, the sentiments are strongly felt, with fewer people being neutral.



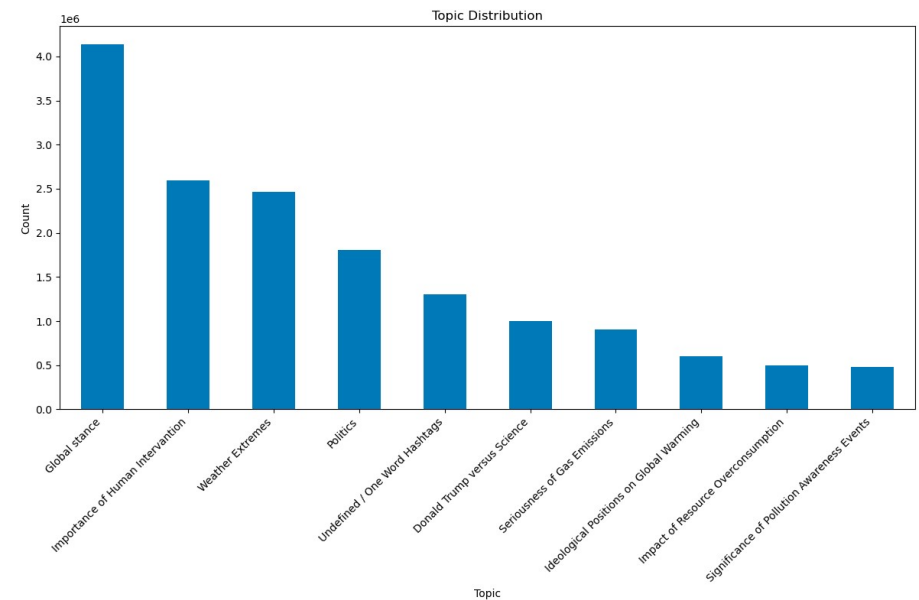
# Visualizations

- The bar chart shows the count of tweets classified into three categories based on the stance towards climate change: 'believer', 'neutral', and 'denier'.
- The stance distribution highlights that the conversation on climate change on Twitter may be led by those who believe in it, which could reflect a consensus in the Twitter community or a larger engagement by those who are concerned about climate change.



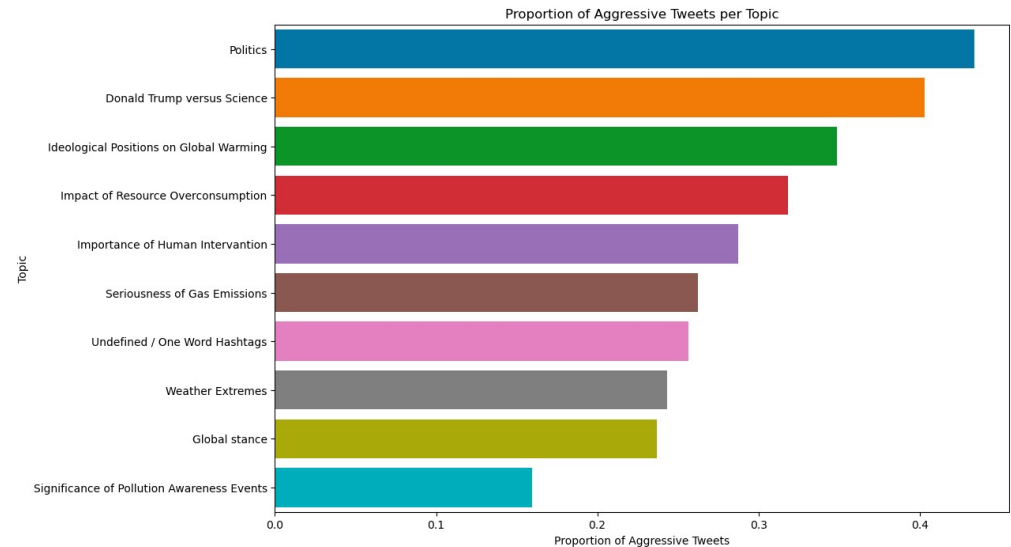
# Visualizations

- The high frequency of tweets about the global stance and human intervention might indicate that the public discourse on Twitter is not only active but also oriented towards acknowledging climate change and discussing solutions.
- The topic distribution can shed light on which aspects of climate change are most engaging for the public. This can be crucial for understanding public engagement and awareness.



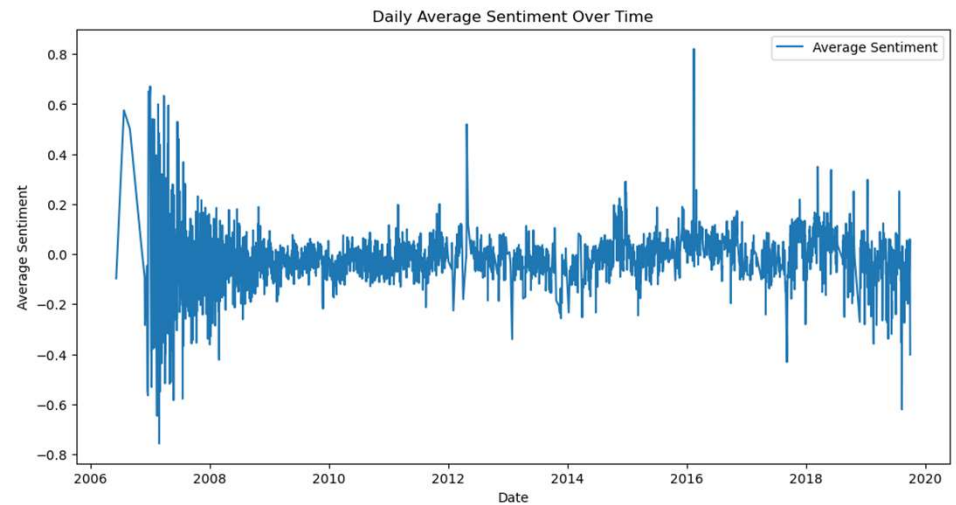
# Visualizations

- The topics "Politics" and "Donald Trump versus Science" appear to have the highest proportion of aggressive tweets. This suggests that political discussions and those directly involving political figures or comparisons with scientific perspectives tend to be more divisive or elicit stronger emotions, leading to more aggressive language.
- Topics that have a political or ideological charge, such as "Ideological Positions on Global Warming," also show higher levels of aggressiveness, which aligns with the idea that more contentious issues lead to more polarized and possibly aggressive online behavior.



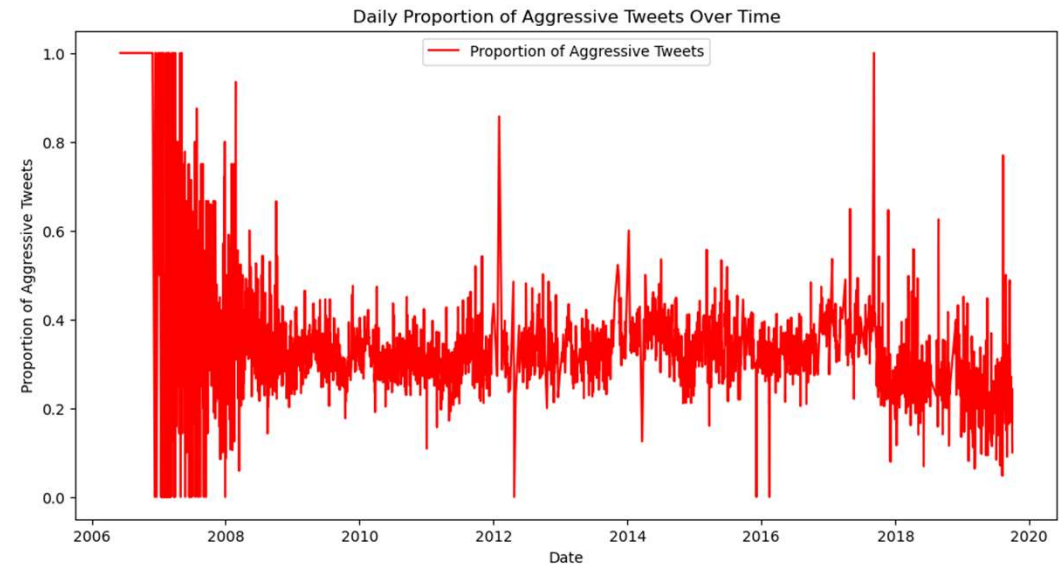
# Visualizations

- The sentiment analysis suggests that public opinion on climate change varies widely day by day, likely influenced by current events. The absence of a clear upward or downward trend in sentiment might indicate that public opinion is not becoming notably more positive or negative over time, but rather remains responsive to external stimuli.



# Visualizations

The proportion of aggressive tweets has its own unique pattern, distinct from sentiment. While it is not steadily increasing or decreasing, the data might indicate that public discourse has periods of intense debate or conflict.





# Analysis for Policy Makers and activists



1. **Promote Positive Discourse on Human Intervention:** Given the positive sentiment surrounding human intervention, policymakers and activists should capitalize on this by promoting success stories, creating more awareness about individual actions that make a difference, and encouraging positive reinforcement for sustainable practices.
2. **Address Controversial Topics with Nuanced Communication:** Topics that generate negative sentiment or aggressive discourse, such as "Ideological Positions on Global Warming" and "Donald Trump versus Science", should be approached with careful and nuanced communication strategies that aim to educate, not alienate. This could involve fact-based information campaigns that counter misinformation and foster understanding.
3. **Engage in Constructive Political Dialogue:** With "Politics" being a central topic of discussion, it's essential for policymakers to engage constructively with the public, emphasizing bipartisan solutions and common ground rather than divisive rhetoric. Activist groups can facilitate these discussions through community forums and debates.

## Future Advancements

---

- **Integration of AI in Real-time Sentiment Analysis:** Advanced artificial intelligence (AI) models could be developed to perform real-time sentiment analysis on social media streams. These models would be capable of understanding context, detecting sarcasm, and identifying subtle emotional cues.
- **Predictive Analytics for Policy Impact:** Machine learning algorithms could be used to predict public reactions to proposed climate policies or initiatives. By analyzing historical data, these models could forecast the effectiveness of policies and potential public pushback or approval, providing valuable insights for policymakers before implementation.
- **Comprehensive Multimodal Social Media Analysis:** Future advancements may include sophisticated multimodal analysis that goes beyond text to include images, videos, and audio. This approach would allow for a more holistic understanding of public discourse and sentiment on climate change.



## References

- Effrosynidis, Dimitrios (2022), “The Climate Change Twitter Dataset”, Mendeley Data, V1, doi: 10.17632/mw8yd7z9wc.1
- Databricks Documentation: <https://docs.databricks.com/en/index.html>
- MongoDB Documentation: <https://www.mongodb.com/docs/>



THANK YOU!