Climate Change Modeling – Sentiment Analysis Report

Tools Used:

Jupyter Notebook, Python, pandas, matplotlib, seaborn, sklearn, nltk

1. Project Goal

The primary goal of this project is to analyze public sentiment regarding climate change by using user comments collected from NASA's Facebook page between 2020 and 2023. The project explores the relationship between public sentiment and user engagement (likes and replies), and also builds a basic machine learning model to predict comment engagement.

2. Dataset Overview

- Source: NASA Climate Change Facebook Page (2020–2023)
- File Used: climate_nasa.csv
- Total Records: ~500 comments
- Columns:
    - Date: Timestamp of the comment
    - LikesCount: Number of likes
    - CommentsCount: Number of replies
    - ProfileName: Hashed user identifier
    - Text: The actual comment

3. Exploratory Data Analysis (EDA)

Comments Over Time

- Monthly comment volume was analyzed by grouping by year and month.
- Observations show steady user interaction across the dataset period.

Most Liked Comments

- Several comments received a high number of likes, usually reflecting positive or supportive sentiment.

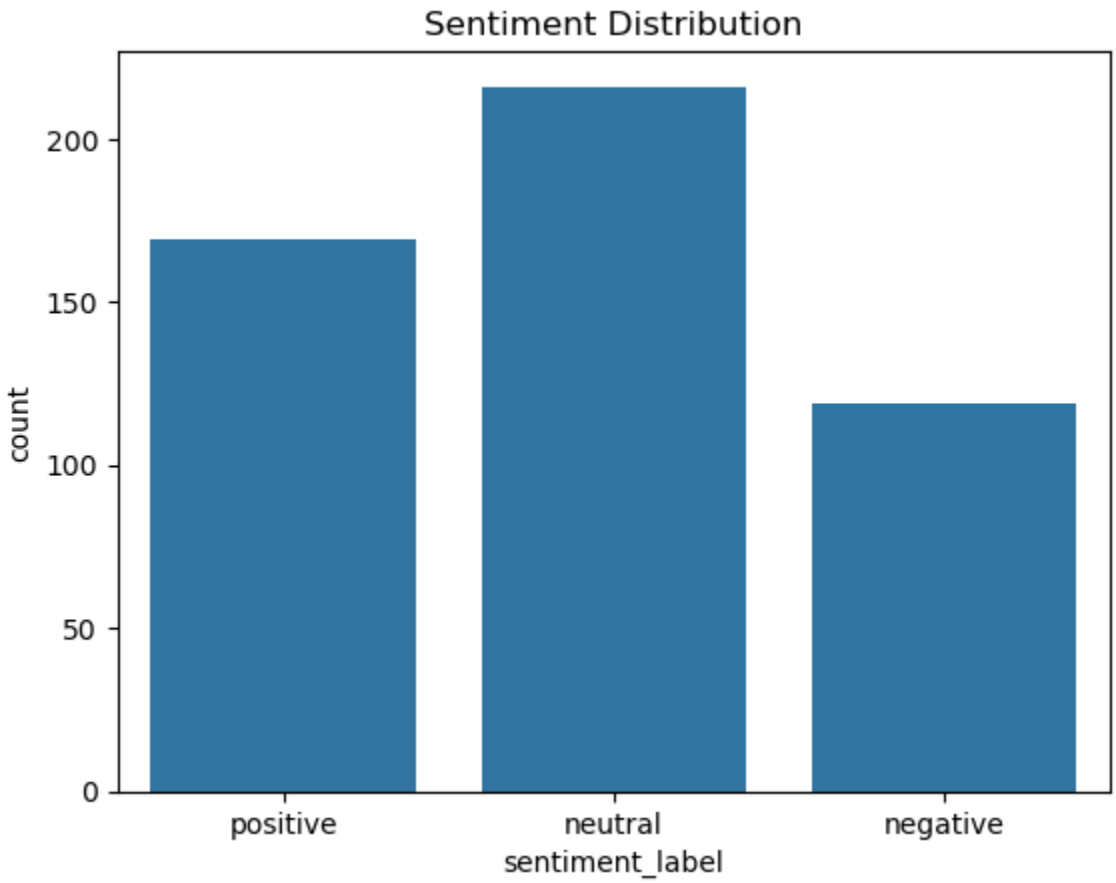
Data Cleaning

- Converted Date to datetime format.
- Removed null or empty comments.
- Added new column comment_length for scenario analysis.

4. Sentiment Analysis

Method:

- Used **NLTK's VADER** sentiment analyzer.
- Calculated compound score and categorized as:
    - positive (> 0.2)
    - neutral (-0.2 to 0.2)
    - negative (< -0.2)



- Positive comments were the most common, followed by neutral and negative.

Likes by Sentiment:

- **Positive comments** had the highest average likes.
- **Negative comments** were more polarizing—some had high engagement, others very low.

5. Machine Learning Model

Goal:

Predict LikesCount based on comment text.

Preprocessing:

- Used TfidfVectorizer to convert text into numerical features (max 500 features).

Model:

- **Algorithm:** Random Forest Regressor
- **Train/Test Split:** 80/20
- **Metric:** Mean Squared Error (MSE)

Results:

- **MSE:** 327.0826241651351
- **Conclusion:** The model had moderate performance. Adding more features (e.g., comment timing or sentiment) may improve accuracy.

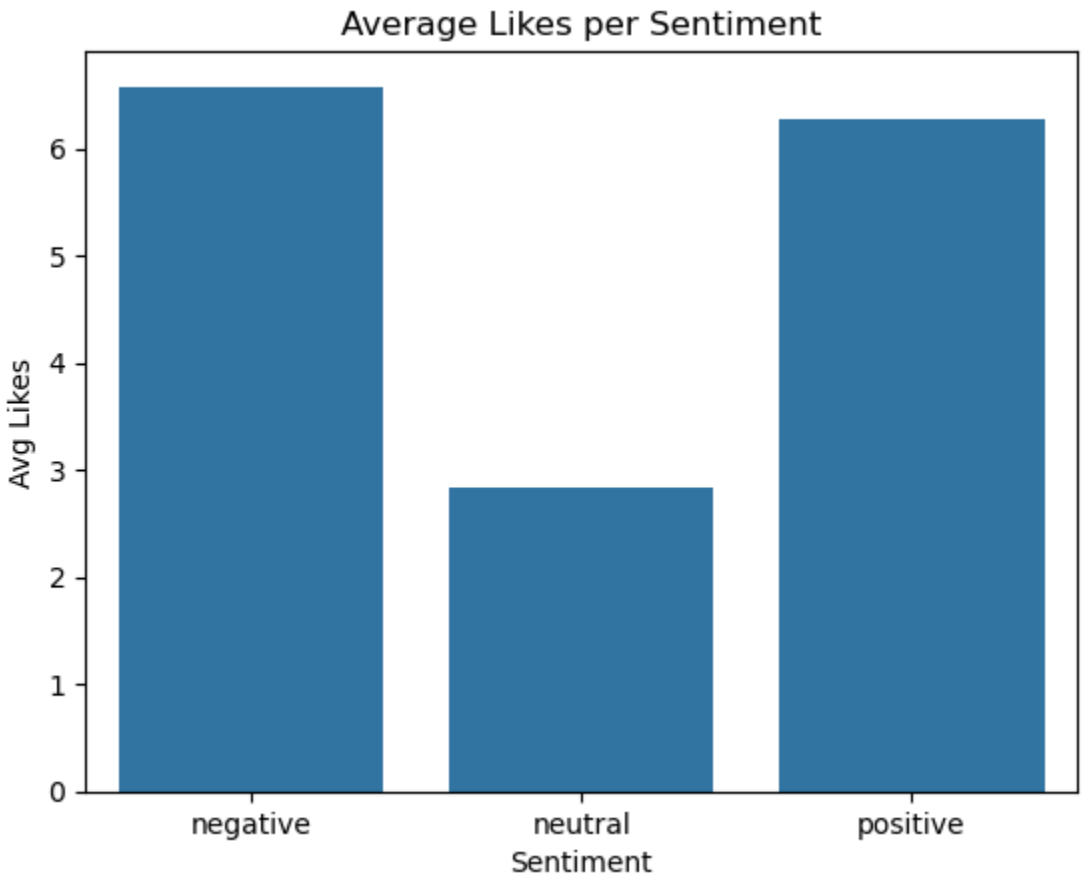
6. Scenario Analysis

Scenario 1: Sentiment vs Likes

Positive : 6.266272

Neutral : 2.833333

Negative : 6.579832



Positive comments consistently get more likes, suggesting the audience favors optimism or support.

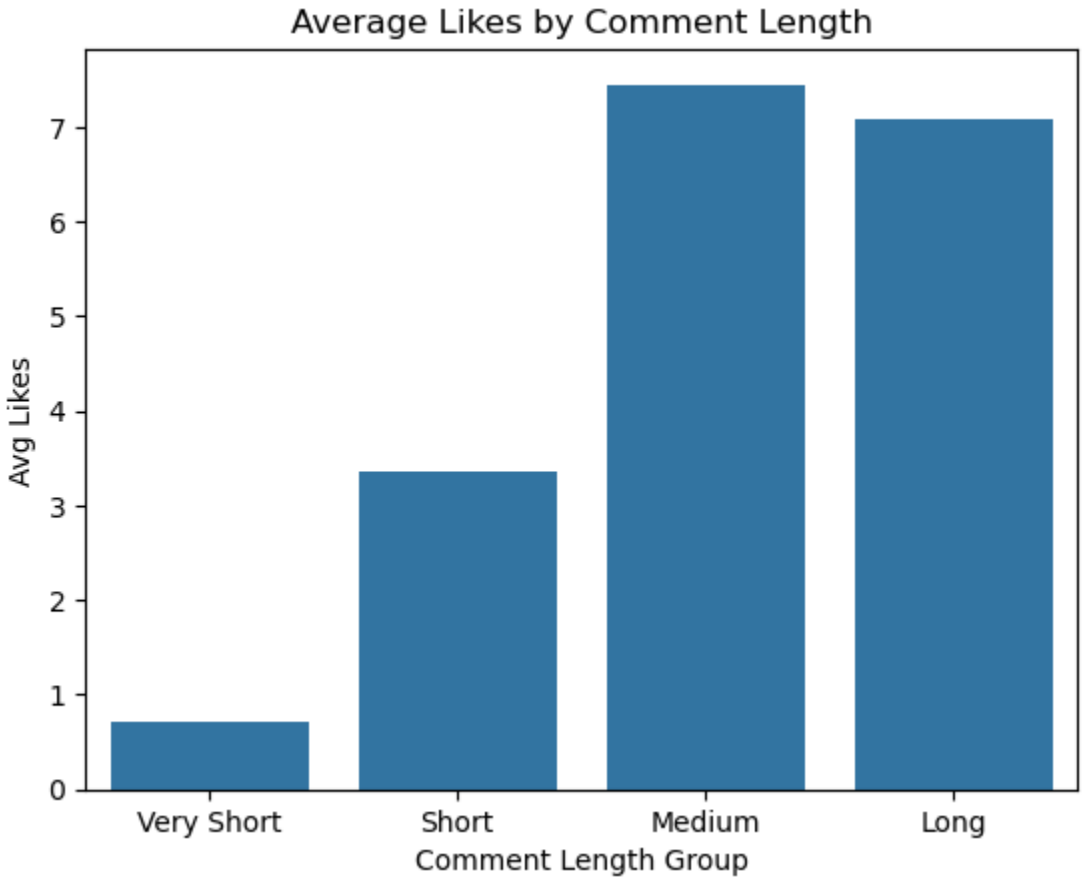
Scenario 2: Comment Length vs Likes

Very Short : 0.710000

Short : 3.357664

Medium : 7.446154

Long : 7.446154



Medium and long comments received more likes, likely due to richer content and explanation.

7. Future Projections

Goal:

Forecast the sentiment distribution for the next month based on historical sentiment trends.

Method:

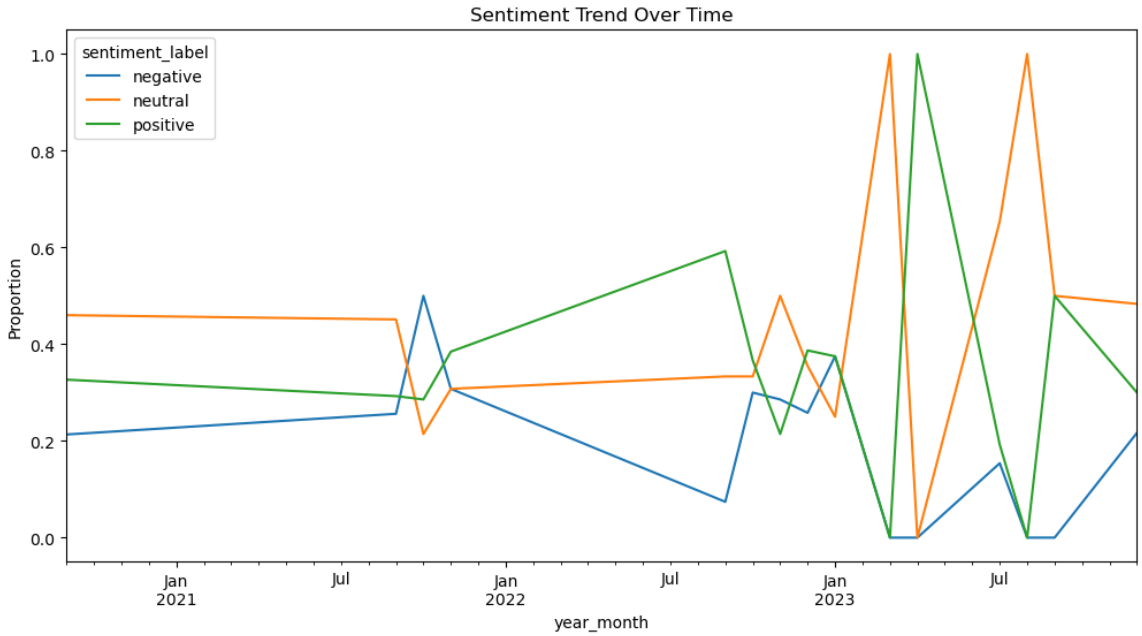
- Grouped past sentiments by month.
- Calculated proportions of each sentiment label.
- Applied linear regression to project sentiment trends forward.

Projected Sentiment Distribution

Positive : 0.33

Neutral : 0.62

Negative : 0.04



Based on recent trends, public sentiment is expected to remain predominantly positive with a slight decrease in neutral comments.

8. Conclusion

- **Public sentiment** about NASA's climate efforts is largely **positive**.
- **Engagement** (likes) tends to be higher for **positive** and **medium/long** comments.
- A basic ML model was able to predict engagement to some extent, though further feature engineering could improve it.

- **Future sentiment projections** suggest continued positive outlook among commenters.
- This kind of analysis can help NASA and other organizations **optimize communication strategies** by understanding how the public responds to different tones and types of messages.