Climate Change Discourse: NASA Social Media Analysis Report

Comprehensive Analysis of Social Media Engagement on Climate Change Topics

Project Details

Project Prepared By: Gurram Sree Pavani

Date of Preparation: 2025-12-03

Contact Information: sreepavani8112001@gmail.com

1. Problem Statement

NASA maintains a strong digital presence across major social media platforms, sharing scientific insights, research updates, and educational content related to climate change. In recent years, public engagement with climate-related posts has grown substantially, attracting diverse audiences including researchers, activists, educators, and general followers. Social media interactions—particularly comments and likes—serve as valuable indicators of public interests, concerns, and discourse patterns.

The dataset analyzed for this project consists of 522 social media comments and engagement metrics recorded between September 7, 2022, and December 22, 2022. Across the dataset, over 2,000 likes were recorded, with an average of 4.72 likes per comment. This dataset offers meaningful insights but also presents considerable challenges. The commentsCount field contains 278 missing values, accounting for 53.26% of the data. Furthermore, the likesCount column includes 65 outliers, indicating instances of highly disproportionate engagement. Engagement is unevenly distributed across 92 unique profiles, making user-based comparisons difficult. Additional data type inconsistencies were identified in the text field, where mixed datatypes appear.

The primary objective of this research is to analyze NASA's climate change discourse on social media by identifying engagement patterns, influential contributors, and temporal trends. Through structured data preprocessing, exploratory analysis, and dashboard visualization, the study aims to uncover meaningful insights about how climate change discussions evolve online.

2. Techniques & Tools Used

2.1 Data Cleaning & Preprocessing

- Tools: Python (Pandas, NumPy)
- Loaded raw JSON dataset
- Parsed timestamps, standardized data types
- Handled missing values: 278 missing in commentsCount, 18 missing in text
- Hash-anonymized profile names
- Removed duplicates

2.2 Exploratory Data Analysis (EDA)

- Tools: Pandas profiling, descriptive statistics
- likesCount: Mean = 4.72, Std = 12.05, Range = 0–126
- commentsCount: Mean = 8.70, Std = 12.27, Range = 1–93
- Outliers detected using IQR
- Extracted temporal features and profile-wise metrics

2.3 Data Quality Assessment

- Identified issues: outliers, missing values, mixed types
- Applied capping and imputation strategies

2.4 Dashboard Creation

- Tool: Microsoft Power BI
- Created KPI cards, bar charts, and monthly trend visualizations
- Added interactive filters

3. Tools & Technologies Used

Category	Tools/Technologies	Purpose
Data Processing	Python, Pandas, NumPy	Data cleaning, transformation, EDA
Data Quality	Pandas Profiling	Automated quality assessment
Visualization	Power BI	Interactive dashboards
Data Storage	CSV/JSON	Source dataset
Environment	Jupyter Notebook	Analysis & documentation

4. Key Findings & Insights

- Engagement skewed across profiles
- September peak activity
- Outlier comments with up to 126 likes
- Correlation between comments and likes
- Data requires cleaning for advanced modeling

5. Dashboard Screenshots

Dashboard Overview:

6. Conclusion & Recommendations

Analysis reveals strong engagement patterns around climate change topics on NASA social media. Data cleaning improved dataset reliability, and Power BI visualizations enabled clear trend identification. Future work may include sentiment analysis, NLP-based topic extraction, and predictive modeling to understand audience sentiment and engagement drivers.