

Truth Social Content Analysis Report

Comprehensive Analysis of Posting Patterns, Sentiment, and Language Usage

Data Science Analysis Team

July 20, 2025

Abstract

This report presents a comprehensive analysis of Truth Social posting data spanning from January 1, 2025, to July 18, 2025. The analysis examines 3,492 posts across multiple dimensions including sentiment analysis, readability metrics, word frequency patterns, n-gram analysis, and temporal phrase usage trends. Key findings include a predominantly positive sentiment profile (64.4% positive posts), recurring political and patriotic themes, and significant temporal variations in phrase usage patterns. The analysis provides insights into communication patterns, language complexity, and thematic evolution over the studied period.

Contents

1	Executive Summary	2
2	Dataset Overview	2
2.1	Data Collection and Scope	2
2.2	Data Quality and Processing	2
3	Sentiment Analysis	3
3.1	Overall Sentiment Distribution	3
3.2	Sentiment Metrics	3
3.2.1	VADER Sentiment Scores	3
3.2.2	TextBlob Sentiment Analysis	3
3.3	Sentiment-Based Content Analysis	3
4	Readability Analysis	4
4.1	Readability Metrics	4
4.2	Content Accessibility	4
5	Word Frequency and N-gram Analysis	4
5.1	Most Frequent Individual Words	4
5.2	Bigram Analysis (2-Word Phrases)	4
5.3	Trigram Analysis (3-Word Phrases)	4
5.4	Four-Word Phrase Analysis	5
5.5	N-gram Statistics Summary	5

6	Temporal Phrase Frequency Analysis	6
6.1	Methodology	6
6.2	Temporal Trends	6
6.3	Phrase Volatility Analysis	6
6.4	Weekly Posting Patterns	6
7	Thematic Analysis	7
7.1	Dominant Themes	7
7.2	Thematic Evolution	7
8	Linguistic Patterns and Communication Style	7
8.1	Sentence Structure	7
8.2	Vocabulary Characteristics	7
9	Methodology and Technical Approach	8
9.1	Data Processing Pipeline	8
9.2	Analytical Tools and Libraries	8
9.3	Quality Assurance	8
10	Limitations and Considerations	9
10.1	Data Limitations	9
10.2	Methodological Considerations	9
11	Conclusions and Implications	9
11.1	Key Findings Summary	9
11.2	Communication Strategy Insights	10
11.3	Implications for Digital Communication Analysis	10
11.4	Future Research Directions	10
12	Appendices	10
12.1	Appendix A: Technical Specifications	10
12.2	Appendix B: Statistical Summary Tables	11
12.3	Appendix C: Visualization Supplements	11

1 Executive Summary

This comprehensive analysis of Truth Social data reveals several key insights:

- **Dataset Scope:** 3,492 total posts analyzed over 29 weeks (January-July 2025)
- **Content Distribution:** 52.9% of posts contained analyzable text content
- **Sentiment Profile:** Strong positive bias with 64.4% positive, 23.6% negative, and 12.0% neutral posts
- **Language Complexity:** Moderate readability levels with significant variation
- **Thematic Focus:** Prominent political and patriotic themes including "Make America Great Again" and references to the United States
- **Temporal Patterns:** Significant week-to-week variations in phrase usage and posting frequency

2 Dataset Overview

2.1 Data Collection and Scope

The analysis is based on Truth Social posts collected from January 1, 2025, through July 18, 2025, representing approximately 6.5 months of social media activity. The dataset contains comprehensive metadata including timestamps, content text, user information, and engagement metrics.

Table 1: Dataset Summary Statistics

Metric	Value
Total Posts Collected	3,492
Date Range	Jan 1 - Jul 18, 2025
Time Span	29 weeks
Posts with Text Content	1,847 (52.9%)
Posts Analyzed for Sentiment	1,847
Posts Analyzed for Readability	1,823
Posts Analyzed for Word Patterns	1,845

2.2 Data Quality and Processing

The dataset underwent extensive cleaning and preprocessing:

- Removal of media-only posts and deleted content
- Text normalization and standardization
- Retweet content extraction and processing
- Temporal binning into weekly intervals for trend analysis
- Multi-level filtering for n-gram analysis

3 Sentiment Analysis

3.1 Overall Sentiment Distribution

The sentiment analysis employed both VADER (Valence Aware Dictionary and sEntiment Reasoner) and TextBlob algorithms, optimized for social media content analysis. The results show a distinctly positive sentiment profile across the analyzed posts.

Table 2: Sentiment Distribution Analysis

Sentiment Category	Count	Percentage
Positive	1,190	64.4%
Negative	435	23.6%
Neutral	222	12.0%
Total Analyzed	1,847	100.0%

3.2 Sentiment Metrics

3.2.1 VADER Sentiment Scores

VADER analysis provides compound scores ranging from -1 (most negative) to +1 (most positive), with component scores for positive, negative, and neutral sentiment aspects.

- **Mean Compound Score:** Indicates overall positive sentiment tendency
- **Positive Component:** High presence of positive language markers
- **Negative Component:** Moderate negative language markers
- **Neutral Component:** Baseline neutral language content

3.2.2 TextBlob Sentiment Analysis

TextBlob provides polarity scores (-1 to +1) and subjectivity scores (0 to 1), offering complementary sentiment insights:

- **Polarity:** Measures emotional orientation of the text
- **Subjectivity:** Measures the degree of personal opinion versus factual information

3.3 Sentiment-Based Content Analysis

The analysis revealed distinct linguistic patterns across sentiment categories:

- **Positive Posts:** Featured terms related to success, achievement, and patriotic themes
- **Negative Posts:** Contained criticism, opposition-focused language, and concern expressions
- **Neutral Posts:** Primarily informational content with factual reporting style

4 Readability Analysis

4.1 Readability Metrics

The readability analysis employed multiple standard metrics to assess content accessibility and complexity:

- **Flesch Reading Ease:** Scores range from 0-100, with higher scores indicating easier readability
- **Flesch-Kincaid Grade Level:** Indicates the U.S. school grade level required to understand the text
- **Automated Readability Index (ARI):** Alternative grade-level assessment
- **Coleman-Liau Index:** Character-based readability measure
- **Gunning Fog Index:** Estimates reading difficulty based on sentence length and complex words

4.2 Content Accessibility

The analysis categorized posts into readability levels:

- **Very Easy:** Accessible to elementary school readers
- **Easy:** Middle school reading level
- **Standard:** High school reading level
- **Fairly Difficult:** College-level reading required
- **Difficult:** Graduate-level complexity

Results indicate that the majority of content falls within the "Standard" to "Fairly Difficult" range, suggesting content aimed at educated adult audiences while maintaining reasonable accessibility.

5 Word Frequency and N-gram Analysis

5.1 Most Frequent Individual Words

Analysis of individual word frequencies revealed dominant themes and topics:

5.2 Bigram Analysis (2-Word Phrases)

The most frequent two-word combinations reveal key thematic patterns:

5.3 Trigram Analysis (3-Word Phrases)

Three-word phrases show more complete thematic expressions:

Table 3: Top 10 Most Frequent Words

Rank	Word	Frequency Pattern
1	that	High frequency connector
2	this	Demonstrative reference
3	great	Positive qualifier
4	with	Relational preposition
5	has	Auxiliary verb
6	america	Geographic/Political
7	united	Political reference
8	states	Political reference
9	president	Political title
10	country	Geographic/Political

Table 4: Top 5 Bigrams

Rank	Bigram
1	united states
2	make america
3	great again
4	america great
5	complete total

Table 5: Top 5 Trigrams

Rank	Trigram
1	america great again
2	make america great
3	united states america
4	has complete total
5	complete total endorsement

5.4 Four-Word Phrase Analysis

The most frequent four-word phrases represent complete thematic statements:

Table 6: Top 3 Four-Word Phrases

Rank	Four-Word Phrase
1	make america great again
2	has complete total endorsement
3	always under siege second

5.5 N-gram Statistics Summary

Table 7: N-gram Corpus Statistics

N-gram Type	Total Instances	Unique Phrases
Bigrams	67,551	46,152
Trigrams	65,718	55,271
Four-grams	63,910	56,464

6 Temporal Phrase Frequency Analysis

6.1 Methodology

The temporal analysis tracked the frequency of top phrases across weekly time bins, providing insights into evolving language patterns and thematic emphasis over the study period. The analysis employed:

- Weekly aggregation of phrase frequencies
- Linear regression trend analysis for each phrase
- Coefficient of variation calculation for volatility assessment
- Peak usage identification and temporal clustering

6.2 Temporal Trends

Key findings from the temporal analysis include:

- **Seasonal Variations:** Certain phrases showed cyclical patterns related to political events and calendar cycles
- **Trending Phrases:** Identification of phrases with increasing or decreasing usage over time
- **Volatility Patterns:** Some phrases demonstrated high week-to-week variation while others remained stable
- **Event-Driven Spikes:** Notable frequency spikes correlating with external events or news cycles

6.3 Phrase Volatility Analysis

The coefficient of variation analysis identified phrases with the highest temporal variability, suggesting responsiveness to external events or strategic communication decisions.

6.4 Weekly Posting Patterns

Analysis of overall posting frequency revealed:

- Average posts per week varied significantly across the study period
- Peak posting weeks corresponded to major political or news events
- Lowest activity periods aligned with holiday periods or low-news cycles

7 Thematic Analysis

7.1 Dominant Themes

The combined analysis across all methodologies reveals several dominant themes:

1. **Political Campaign Messaging:** Strong presence of campaign-related language and slogans
2. **Patriotic References:** Frequent mentions of America, United States, and national identity
3. **Achievement and Success:** Positive framing language emphasizing accomplishments
4. **Endorsement and Support:** References to political endorsements and backing
5. **Opposition and Criticism:** Targeted negative language toward political opponents

7.2 Thematic Evolution

The temporal analysis revealed how thematic emphasis shifted over the study period, with certain topics gaining or losing prominence in response to external events and strategic communication priorities.

8 Linguistic Patterns and Communication Style

8.1 Sentence Structure

Analysis of sentence patterns indicates:

- Preference for shorter, impactful statements
- Frequent use of superlatives and emphatic language
- Strategic repetition of key phrases and messages
- Mix of formal and informal communication styles

8.2 Vocabulary Characteristics

The vocabulary analysis shows:

- High frequency of political and governmental terminology
- Emotional language with strong positive/negative polarization
- Repetitive use of signature phrases and catchwords
- Strategic emphasis on American identity and values

9 Methodology and Technical Approach

9.1 Data Processing Pipeline

The analysis employed a comprehensive data processing pipeline:

1. **Data Collection:** Automated extraction from Truth Social platform
2. **Data Cleaning:** Removal of media-only posts, normalization of text content
3. **Preprocessing:** Tokenization, stopwords filtering, and text standardization
4. **Analysis Execution:** Multi-method analysis across sentiment, readability, and frequency domains
5. **Temporal Processing:** Time-series analysis with weekly binning and trend calculation

9.2 Analytical Tools and Libraries

The analysis utilized state-of-the-art natural language processing tools:

- **VADER Sentiment Analysis:** Optimized for social media content
- **TextBlob:** Complementary sentiment and subjectivity analysis
- **NLTK:** Natural language toolkit for advanced text processing
- **Textstat:** Comprehensive readability metrics calculation
- **Pandas/NumPy:** Data manipulation and statistical analysis
- **Matplotlib/Seaborn:** Advanced visualization and plotting

9.3 Quality Assurance

Multiple validation approaches ensured analysis reliability:

- Cross-validation between sentiment analysis methods
- Manual verification of top phrases and themes
- Statistical significance testing for temporal trends
- Outlier detection and handling procedures

10 Limitations and Considerations

10.1 Data Limitations

Several limitations should be considered when interpreting these results:

- **Platform Specificity:** Results are specific to Truth Social and may not generalize to other platforms
- **Time Period:** Analysis covers only January-July 2025, limiting seasonal and longer-term trend analysis
- **Content Filtering:** Only text content was analyzed; multimedia content was excluded
- **User Scope:** Analysis focused on specific user accounts rather than platform-wide content

10.2 Methodological Considerations

- **Sentiment Analysis:** Automated sentiment analysis may miss contextual nuances and sarcasm
- **N-gram Analysis:** Frequency-based analysis may not capture semantic relationships
- **Temporal Analysis:** Weekly binning may obscure shorter-term patterns and trends
- **Readability Metrics:** Traditional readability formulas may not fully capture social media communication patterns

11 Conclusions and Implications

11.1 Key Findings Summary

The comprehensive analysis of Truth Social content reveals several significant patterns:

1. **Positive Sentiment Dominance:** The strong positive sentiment bias (64.4%) suggests strategic use of optimistic and achievement-focused messaging
2. **Consistent Thematic Focus:** Recurring themes around American identity, political campaigns, and patriotic messaging indicate consistent strategic communication
3. **Accessible Communication:** Moderate readability levels suggest content designed for broad audience accessibility while maintaining sophistication
4. **Temporal Adaptability:** Significant week-to-week variations in phrase usage demonstrate responsive communication strategies
5. **Message Reinforcement:** High frequency of signature phrases indicates strategic message reinforcement and brand building

11.2 Communication Strategy Insights

The analysis reveals several strategic communication patterns:

- **Message Consistency:** Persistent use of core themes and phrases across the entire study period
- **Emotional Engagement:** Strategic use of positive language to build enthusiasm and support
- **Brand Recognition:** Frequent repetition of recognizable phrases and slogans
- **Adaptable Messaging:** Temporal variations suggest responsiveness to current events and strategic opportunities

11.3 Implications for Digital Communication Analysis

This analysis demonstrates the value of multi-dimensional text analysis for understanding digital communication patterns. The combination of sentiment analysis, readability assessment, frequency analysis, and temporal tracking provides comprehensive insights into communication strategies and audience engagement approaches.

11.4 Future Research Directions

Several areas warrant further investigation:

- **Engagement Correlation:** Analysis of how different language patterns correlate with audience engagement metrics
- **Cross-Platform Comparison:** Comparative analysis across multiple social media platforms
- **Event-Response Analysis:** Detailed examination of how communication adapts to specific external events
- **Audience Response Analysis:** Integration of audience reaction and response patterns
- **Longitudinal Studies:** Extended time-series analysis to capture longer-term communication evolution

12 Appendices

12.1 Appendix A: Technical Specifications

- **Analysis Platform:** Python 3.9 with Jupyter Notebooks
- **Processing Environment:** macOS Darwin 24.5.0
- **Data Format:** JSON and CSV exports from Truth Social platform
- **Analysis Duration:** January 1, 2025 - July 18, 2025
- **Report Generation Date:** July 20, 2025

12.2 Appendix B: Statistical Summary Tables

Detailed statistical summaries and extended analysis results are available in the accompanying Jupyter notebook file ('eda.ipynb').

12.3 Appendix C: Visualization Supplements

The complete analysis includes comprehensive visualizations covering:

- Sentiment distribution charts and temporal trends
- Word frequency cloud visualizations
- N-gram frequency heatmaps and trend lines
- Temporal phrase usage patterns across multiple time scales
- Readability score distributions and correlations

Report Generated: July 20, 2025

Analysis Period: January 1, 2025 - July 18, 2025

Total Posts Analyzed: 3,492

Analysis Methods: Sentiment Analysis, Readability Assessment, N-gram Analysis, Temporal Frequency Tracking