

Sentiment Analysis of Political Subreddits Rubric

Welcome to the Sentiment Analysis of Political Subreddits case study! In this project, you'll dive into real-world data from political subreddits (r/democrats, r/Republican, and r/PoliticalDiscussion) to uncover how online communities reacted to President Biden's withdrawal from the 2024 presidential election. You will use Python, sentiment analysis tools, and statistical methods to generate insights about online political discourse. Your mission is to clean and process the dataset, analyze sentiment changes before and after the event, and communicate your findings through clear visualizations and a written report. Use this rubric to guide your work, ensuring that each component meets the standards expected of a professional data scientist.

Deliverables include:

- A cleaned dataset ready for analysis.
- Sentiment analysis results, including key insights and trends.
- Statistical tests comparing sentiment before and after the event.
- At least two visualizations that effectively communicate your findings.
- A report summarizing your process, results, and conclusions.

Follow the detailed criteria below to structure and evaluate your work. Remember to document your process clearly and think critically about your findings. And, have fun!

EVALUATION CRITERIA

| Category | Description | Weight |
|-------------------------------|---|--------|
| Data Cleaning and Preparation | <p>Complete Documentation: Explain all cleaning steps, such as filtering for Biden-related posts and handling missing values.</p> <p>Output: Provide a cleaned dataset in CSV format, ensuring proper formatting.</p> <p>Justification: Justify each cleaning step to demonstrate its necessity for analysis.</p> | 25% |
| Sentiment Analysis | <p>Application of Vader: Use VADER to calculate sentiment scores (positive, neutral, negative, compound).</p> <p>Interpretation: Explain the meaning of sentiment scores</p> | 30% |

| | | |
|--------------------------|--|-----|
| | <p>and summarize the overall sentiment trends for each subreddit.</p> <p>Comparison: Highlight differences in sentiment before and after the event.</p> | |
| Statistical Analysis | <p>Methodology: Perform two-tailed t-tests to compare sentiment scores before and after the event for each subreddit.</p> <p>Results: Report key statistics, including means, standard deviations, and p-values.</p> <p>Interpretation: Discuss the significance of the results and their implications for understanding the data.</p> | 20% |
| Visualization | <p>Quality: Create at least two visualizations (e.g., bar charts, line plots) that effectively communicate your findings.</p> <p>Clarity: Ensure visualizations are labeled and accompanied by brief explanations.</p> <p>Relevance: Use visualizations to support key insights and comparisons.</p> | 15% |
| Presentation and Writing | <p>Clarity and Organization: Write a clear, concise, and well-organized report</p> <p>Engagement: Present findings in a way that is accessible and engaging.</p> <p>Grammar/Style: Use a professional format, free of grammatical errors.</p> | 10% |