

Web Scraping And Sentiment Analysis Project

The issue of racism in public education has been a problem in the United States for years, with marginalized communities receiving unequal access to quality education. Recently, the Republican Party has been accused of waging a nationwide war on public education through policies such as funding cuts and promoting private schools, which could worsen existing racial and economic disparities.

Understanding the background and context of this issue is important for analyzing data and developing solutions.

A web scrape and sentiment analysis was carried out on the reddit post ["Not just Florida: The entire GOP is waging a nationwide racism-fueled war on public education."](#) Web scraping was used to gather data on the conversation under the post. After gathering the data, sentiment analysis was performed to analyze the tone and sentiment of the language used in the posts. This analysis aimed to identify whether the sentiment was positive, negative, or neutral towards the GOP's actions and policies.

Text Collection and Preprocessing

The reddit post had an initial total of 1376 comments. However, after scrapping and removal of duplicates, it had 1217 comments showing that it would have sufficient data to analyze. To prepare the data for analysis, we performed several pre-processing steps, including dropping unnecessary columns, removing symbols and stop words, tokenizing the comments, and performing stemming.

By dropping unnecessary columns , we focused solely on the comment text. We also removed symbols, punctuations, and stop words from the text to reduce noise and dimensionality. Tokenization was performed to split the comments into individual words, and stemming was performed to reduce variations of the same word to a common root form.

These pre-processing steps helped to clean and prepare the data for sentiment analysis. The cleaned data was ready for analysis and provided valuable insights into the public's perception of the GOP's actions and policies towards public education.

Sentiment Analysis

A thorough analysis of the sentiment of the comments in the Reddit post "Not just Florida: The entire GOP is waging a nationwide racism-fueled war on public education" was conducted. The goal was to gain insights into the opinions and attitudes of the participants towards public education and GOP policies.

To begin the analysis, a word cloud (**Appendix Image 1**) that visualized the most frequently occurring words in the comments was created. The top 10 words were then identified using a bar chart (**Appendix Image 2**), which reinforced the findings from the word cloud. These top 10 words included "people," "school," "education," "schools," "public," "want," "like," "kids," "don't," and "even." It was evident from the top words that the comments were largely centered around people's concerns about public education and the impact of GOP policies on schools, students, and communities.

Next, we used the textblob library to analyze the polarity and subjectivity of the comments. Polarity is a measure of the sentiment of a statement, and it ranges from -1 (negative sentiment) to 1 (positive sentiment). Subjectivity, on the other hand, measures the degree to which a statement is based on personal opinion rather than factual information, and it ranges from 0 (objective) to 1 (subjective).

The analysis revealed that most of the comments had a neutral to slightly positive sentiment. There were 507 comments with a positive sentiment, 374 with a neutral sentiment, and 336 with a negative sentiment. The highest frequency of sentiments occurred between 0.0 and 0.25, as seen in the Polarity Distribution Histogram (**Appendix Image 3**), indicating that the majority of the comments had a neutral to very slight positive sentiment. This suggests that people may not be entirely satisfied with the current state of public education, but they may not be completely opposed to GOP policies either. For subjectivity, the majority of the comments were objective. Specifically, out of the 1217 analyzed comments, 844 were objective and 373 were subjective. The highest frequency of subjectivity occurred between 0.0 and 0.1, as seen in the Subjectivity Distribution Histogram (**Appendix Image 4**). This could be an indication that the

participants in the discussion were interested in presenting a rational and well-informed perspective on the issue of the GOP's policies towards public education.

To compare the summary statistics between the two groups, polarity and subjectivity, would be an unfair comparison as they have different ranges. However, the summary statistics can be found in the **Appendix Image 5**. One interesting observation from the Sentiment Analysis was that the cross tabulation (**Appendix Image 6**) of the Polarity and Subjectivity classification showed that the majority of the comments were Objective with Neutral Sentiment followed by Objective with Positive Sentiment, which is similar to the findings from the histograms.

Finally, the scatterplot (**Appendix Image 7**) confirmed the results of the histograms and cross tabulation. We observed a high density of comments between 0.0 and 0.25 on the polarity axis (x-axis), confirming the neutral to slightly positive sentiment. On the subjectivity sentiment (y-axis), we saw a high density between 0 to 0.5, confirming that most comments were objective. When combined, the comments were likely to be Objective with Neutral Sentiment or Positive sentiment, as shown by the cross tabulation.

Recommendation/Conclusion

Overall, these findings suggest that there is a need for a more comprehensive and inclusive approach towards public education that takes into account the concerns and needs of marginalized communities. Rather than waging a war on public education, policymakers should focus on improving and investing in the public education system to provide equitable and quality education to all students. This may involve increasing funding for public schools, reducing the achievement gap, providing more resources to teachers, and promoting diversity and inclusion in schools. The findings of this sentiment analysis suggest that there is a need for a more inclusive and equitable approach towards public education that promotes diversity, equity, and social justice.

Appendix

[illegible]

Word	Word Count
people	340
schools	330
school	305
public	255
education	245
like	195
dont	180
would	175
want	165
think	145

Image 3 - Polarity Distribution Histogram

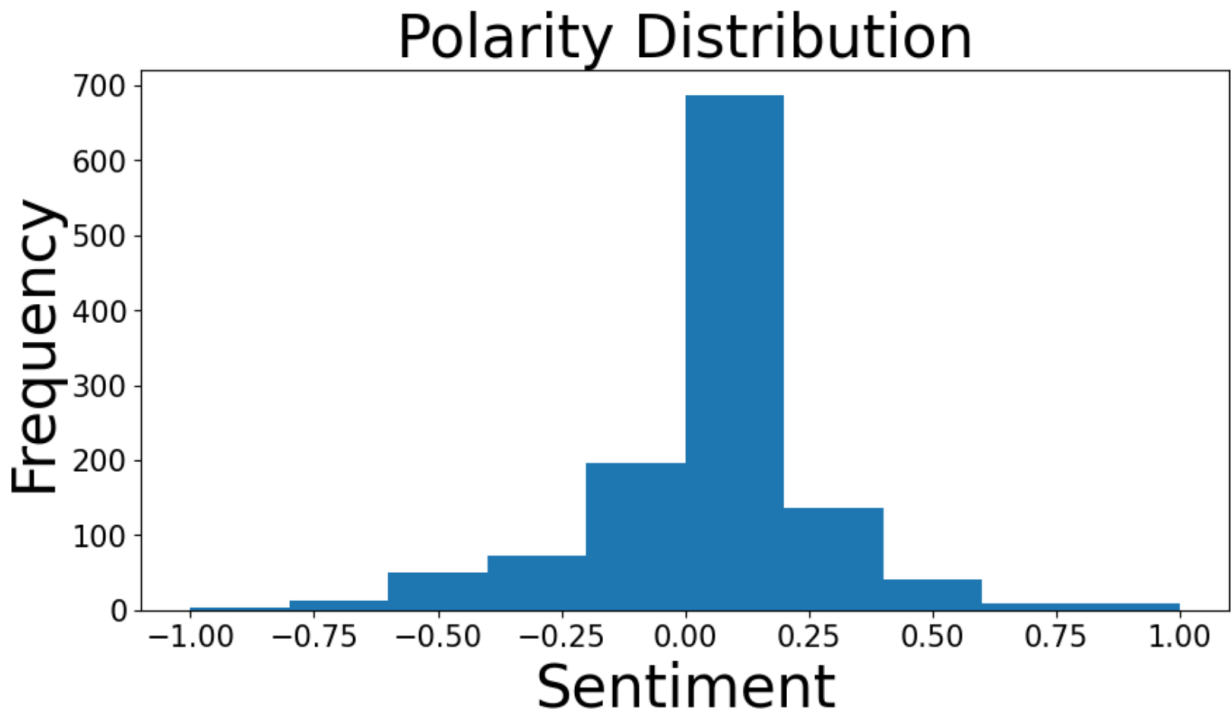


Image 4 - Subjectivity Distribution Histogram

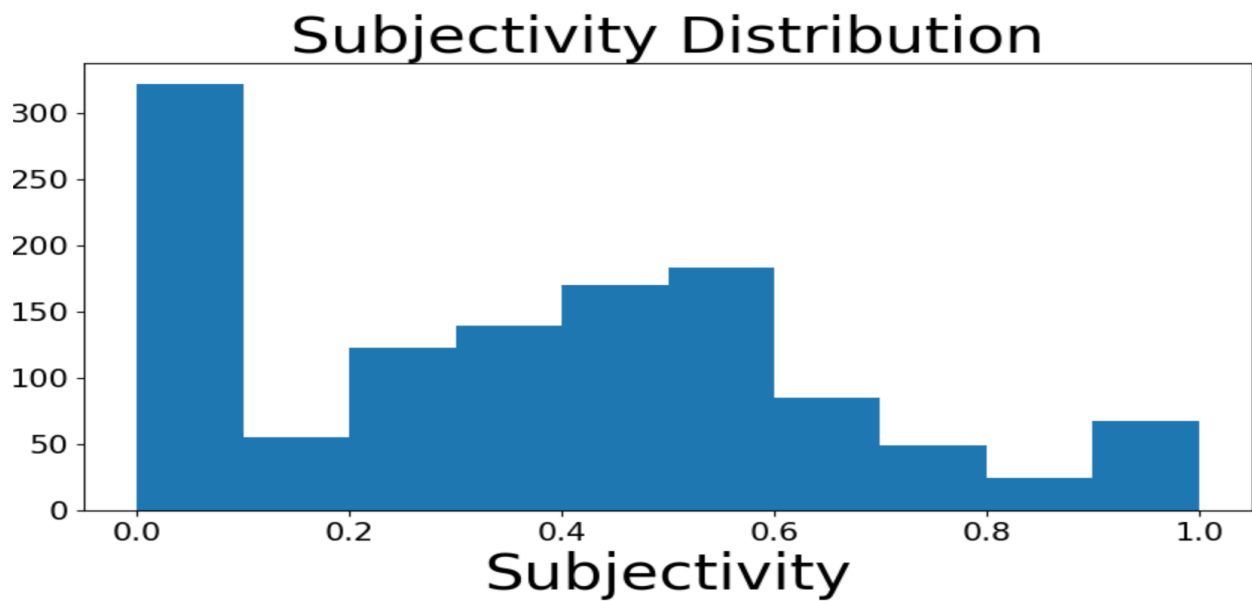


Image 5 - Summary Statistics Of Polarity And Subjectivity

	polarity	subjectivity
count	1217.000000	1217.000000
mean	0.024115	0.360694
std	0.235935	0.282046
min	-1.000000	0.000000
25%	-0.030612	0.050000
50%	0.000000	0.380000
75%	0.125000	0.536111
max	1.000000	1.000000

Image 6 - Cross Tabulation Of Polarity And Subjectivity

SubjectivityClassification	Objective	Subjective
sentiment		
Negative Sentiment	169	167
Neutral Sentiment	360	14
Positive Sentiment	315	192

Image 7 - Scatterplot Of Polarity And Subjectivity

