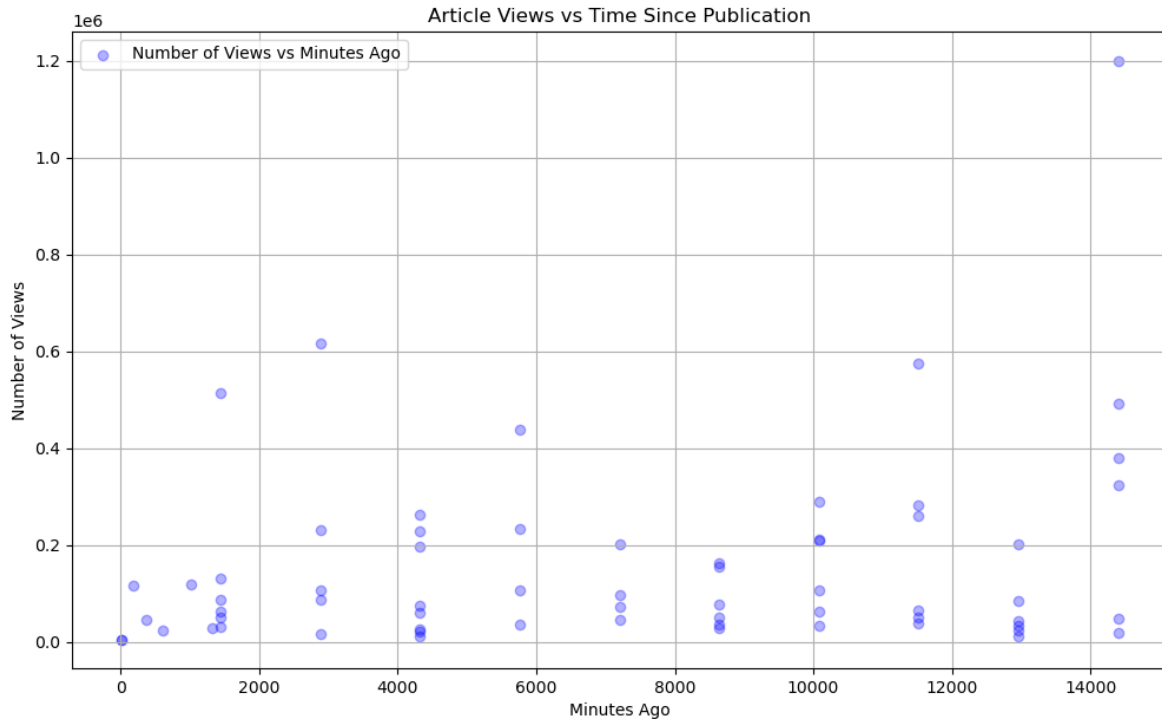


Analysis Report

This report analyzes a sample dataset of news articles from Fox News to clean the data file and perform some preliminary analysis.



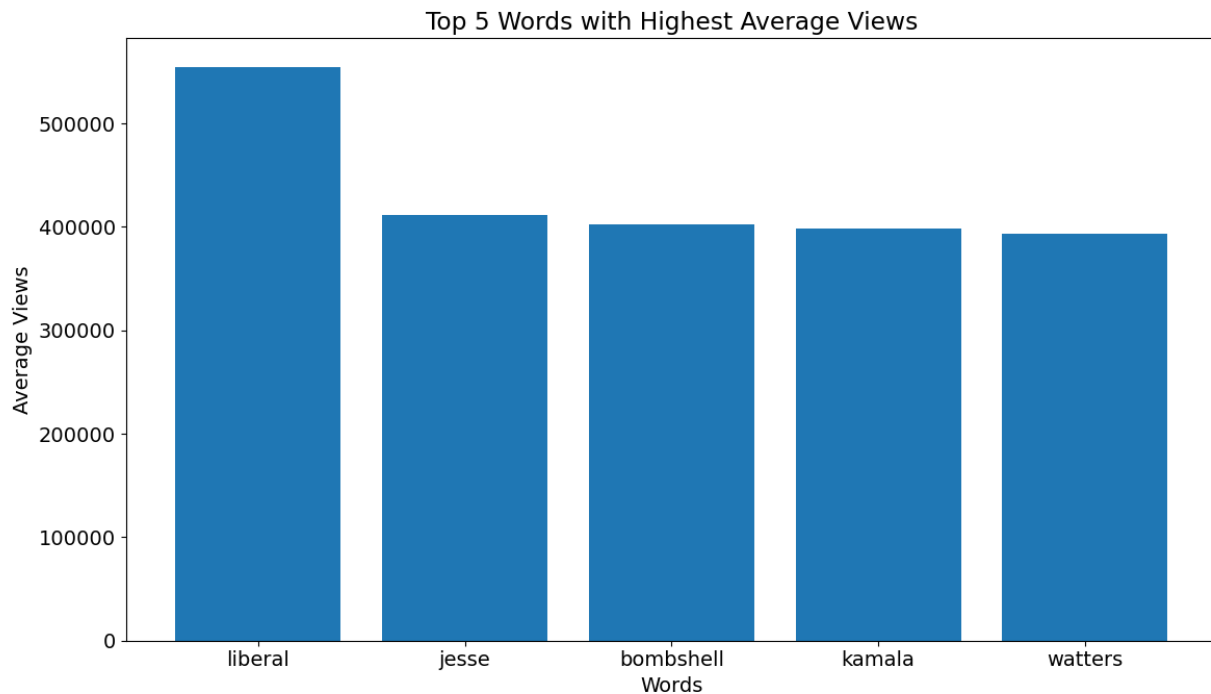
From the scatter plot, a distinct distribution pattern is evident: a significant cluster of points gathers in the lower section, while relatively fewer points are dispersed across the upper portion. This analysis reveals the dynamic nature of article viewership over time, reflecting the impact of various topics. Engaging themes tend to garner more views, whereas less captivating or ordinary ones fall behind. Over time, certain articles maintain a consistently low number of views, while others experience higher levels of engagement. There are a few rare “bombshell” articles that underscores the immediate resonance of specific subjects, driving increased views, while less notable topics show minimal growth. It's important to note that the x-axis represents discrete time, which might limit the accurate capture of interval-based trends.

The utilized text pre-processing steps contribute to the analysis in the following ways:

- Removing non-alphabetic characters, except spaces, is essential for text refinement. It streamlines content by discarding irrelevant characters, enhancing analysis.
- Converting text to lowercase maintains consistency in word format, mitigating case-related discrepancies and bolstering analysis reliability.
- Tokenization breaks titles into words, forming a structured basis for analysis by creating a word collection for further processing.

Suggestions for Improvement:

1. Removing common stop words like "to" and "of" from tokenized terms enhances focus on meaningful content by eliminating low-impact words.
2. Maintaining precise capitalization during the conversion from uppercase to lowercase is crucial. For instance, the distinction between "White House" and "white house" can lead to varying analytical interpretations.
3. Eliminating duplicates to ensure accuracy and prevent overestimation of word frequency.



From the bar plot, the word "liberal" has the highest number of average views, and is used to describe certain viewpoints or individuals. The words "Jesse" and "Watters" combined form the name "Jesse Watters," a conservative popular political commentator renowned for his criticism of liberals. "Kamala" stems from Vice President Kamala Harris, known for her affiliation with the liberal democratic party. The term "bombshell" implies that the articles may contain crucial and contentious information, which could lead to an increase in article views.

Hence, the following conclusions could be drawn:

- There might be a focus on conservative viewpoints or commentary in the articles
- Readers are drawn to articles that analyze and discuss current political figures and events