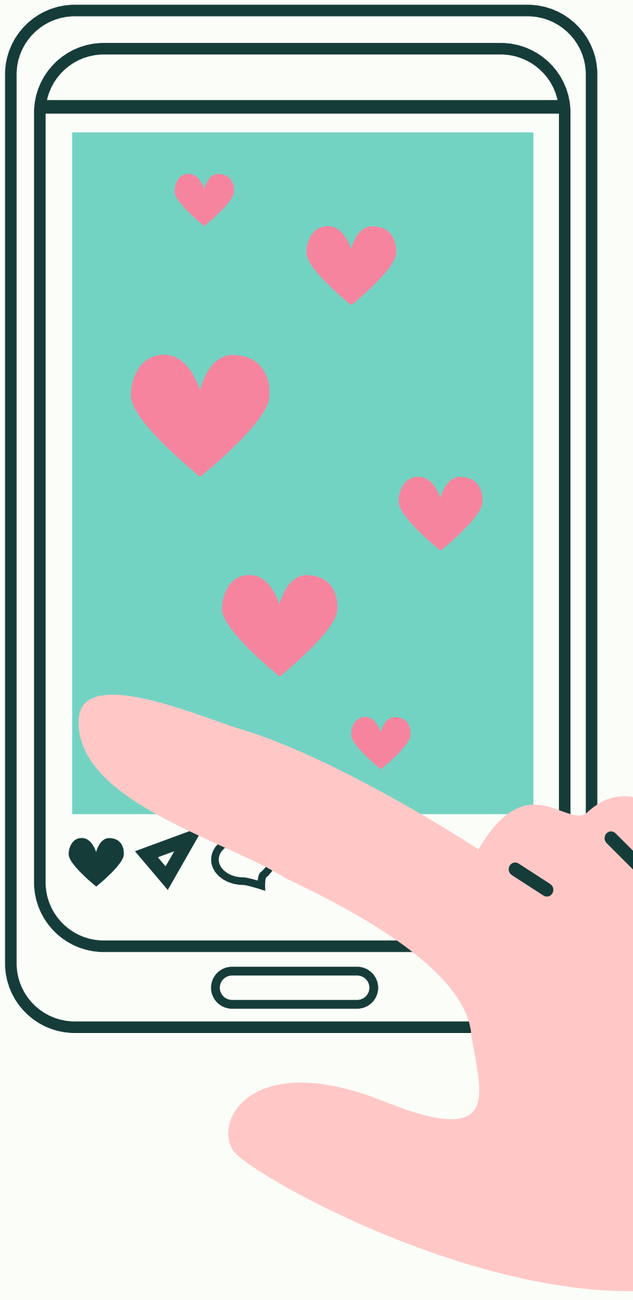
SOCIAL MEDIA

ANALYZE SOCIAL MEDIA SENTIMENT DATA TO UNDERSTAND PUBLIC PERCEPTION OF WELLNESS TRENDS.

Sarah AL Jamal



1. Objectives:

Analysis of the depression among students on Twitter

2.Overview

3.dataset

4.Clean and preprocessing of Tweet

5.Classify the data to Negative and Positive

6.Analysis

7.Future work

OVERVIEW

What is wellness Trends?

Wellness is the act of practicing healthy habits on a daily basis to attain better physical and mental health outcomes, so that instead of just surviving, you're thriving.

They include: social connectedness, exercise, nutrition, sleep and mindfulness.

Each one has an impact on your physical and mental health. By making simple and healthy choices on a daily basis, you will be well on your way towards reducing stress, having positive social interactions and achieving optimal wellness.

Social Connectedness

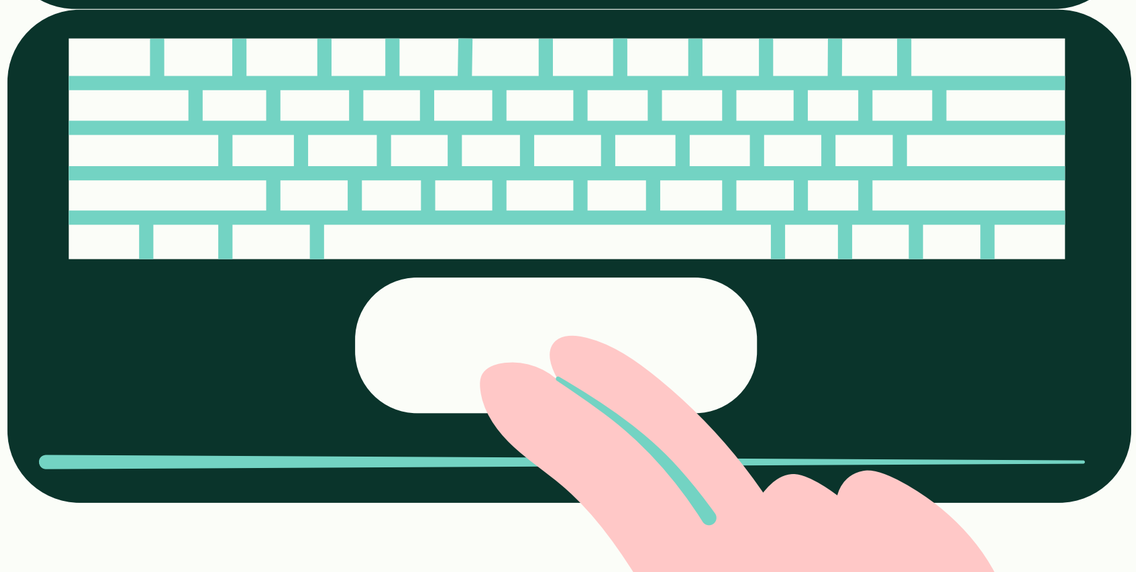
Sleep

Exercise

Nutrition

ABOUT KETO

Patients trying to lose weight often turn to popular diets, and in the past few years, the ketogenic diet has been a trendy option. With the increased interest in this diet, researchers are working to understand the impact of this pattern of eating on patients' health.



Keto diet

is one of the wellness trends that we will conduct analysis on how people perception towards it.

DAtAsEt

DATASET USED IS FROM REDDIT WEBSITE ON DIFFERENT TOPICS RELATED TO KETO DIET

COLUMNS NAMES POST TITLE', 'POST CONTENT', 'POST SCORE', 'COMMENT')

NUMBER OF RECORDS 19780
REDDIT DISCUSSION WE SCRAP FROM REDDIT.COM
FETCH COMMENTS BY EXPANDING MORE COMMENTS USING PYTHON CODE.
PUT THE TEXT INTO DATA FRAME MADE OF ROWS AND COLUMNS.
GET UNIQUE POST TITLES SUCH AS 'HAS ANYONE HERE SEEN SKIN IMPORVEMENTS WITH THE KETO DIET?',
'WHEN FOLLOWING A KETO DIET, IS IT NECESSARY TO ALWAYS EAT ORGANIC FOOD PRODUCTS?'

DATASET

TO CLASSIFY THE COMMENTS SCRAPPED INTO NEGATIVE AND POSITIVE WE USED TEXT BLOB

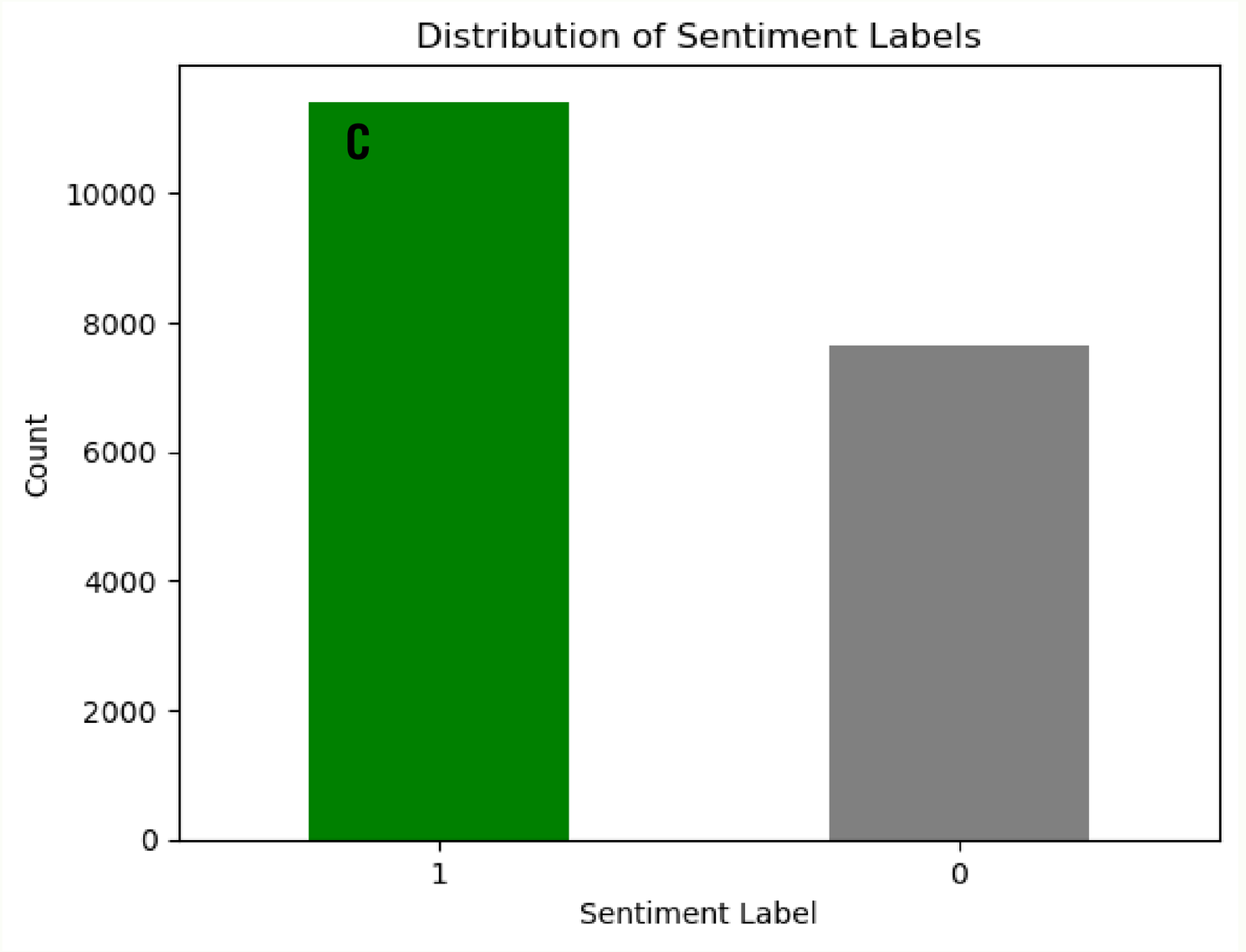
TO FIND THE FREQUENCIES OF WORDS WE USED WORD COUNT
WHICH SHOWED THAT THE WORD”MORE HAS THE MOST POSITIVE FREQUENCY = 29

5.Classify the data to Negative and Positive

Now you can count the occurrences of each label and plot them

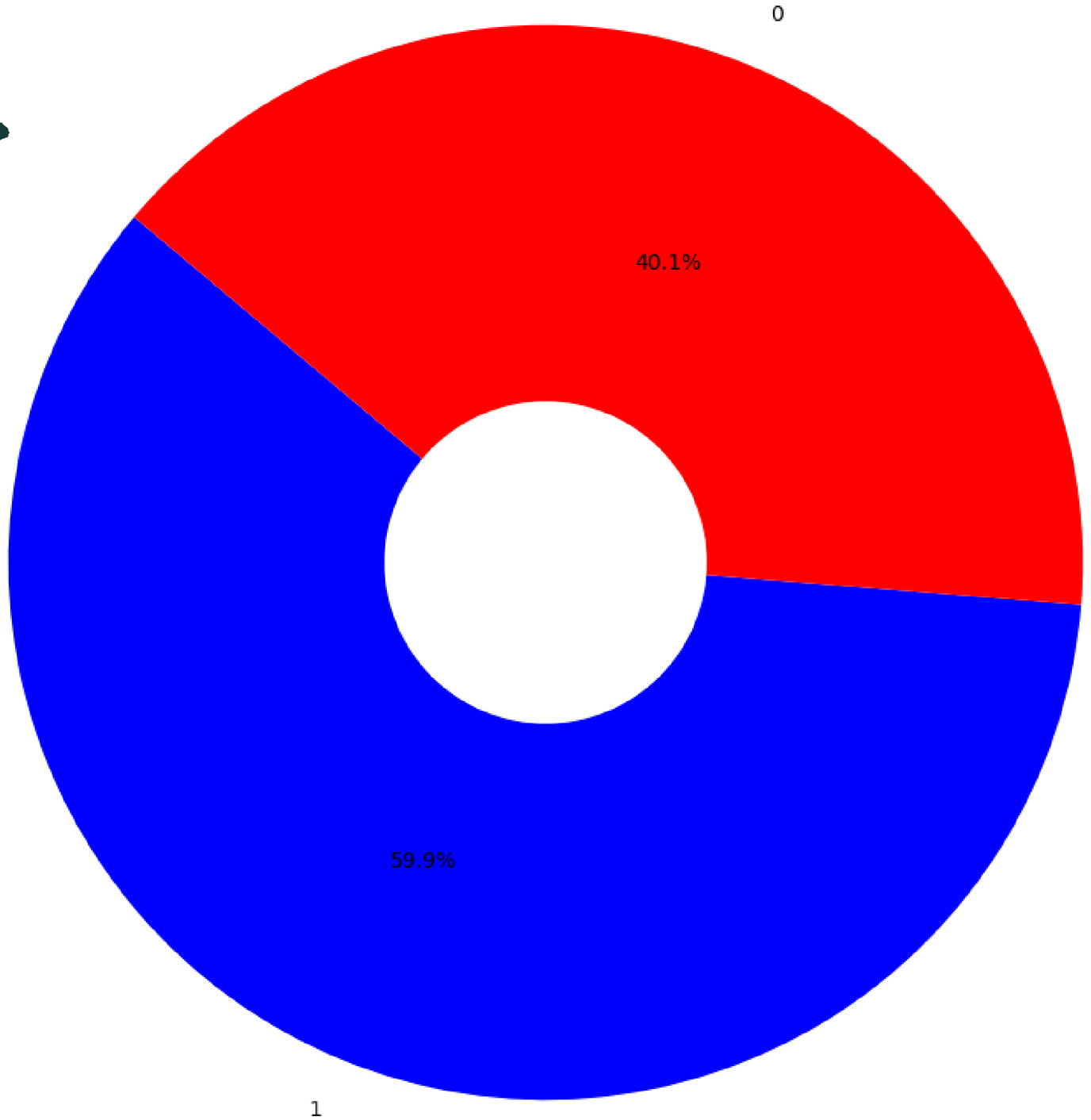
1 11407

0 7648



DATASET

THERE ARE TWO LABELS NEGATIVE (0)AND POSITIVE (1). NEGATIVE COMMENTS APPEAR TO OCCURE OF ABOUT 40% OF TOTAL COMMENTS OR DATASET AND THE POSITIVE LABEL SHOWS 59.9% OF THE DATASET AS ILLUSTRATED.



Analysis

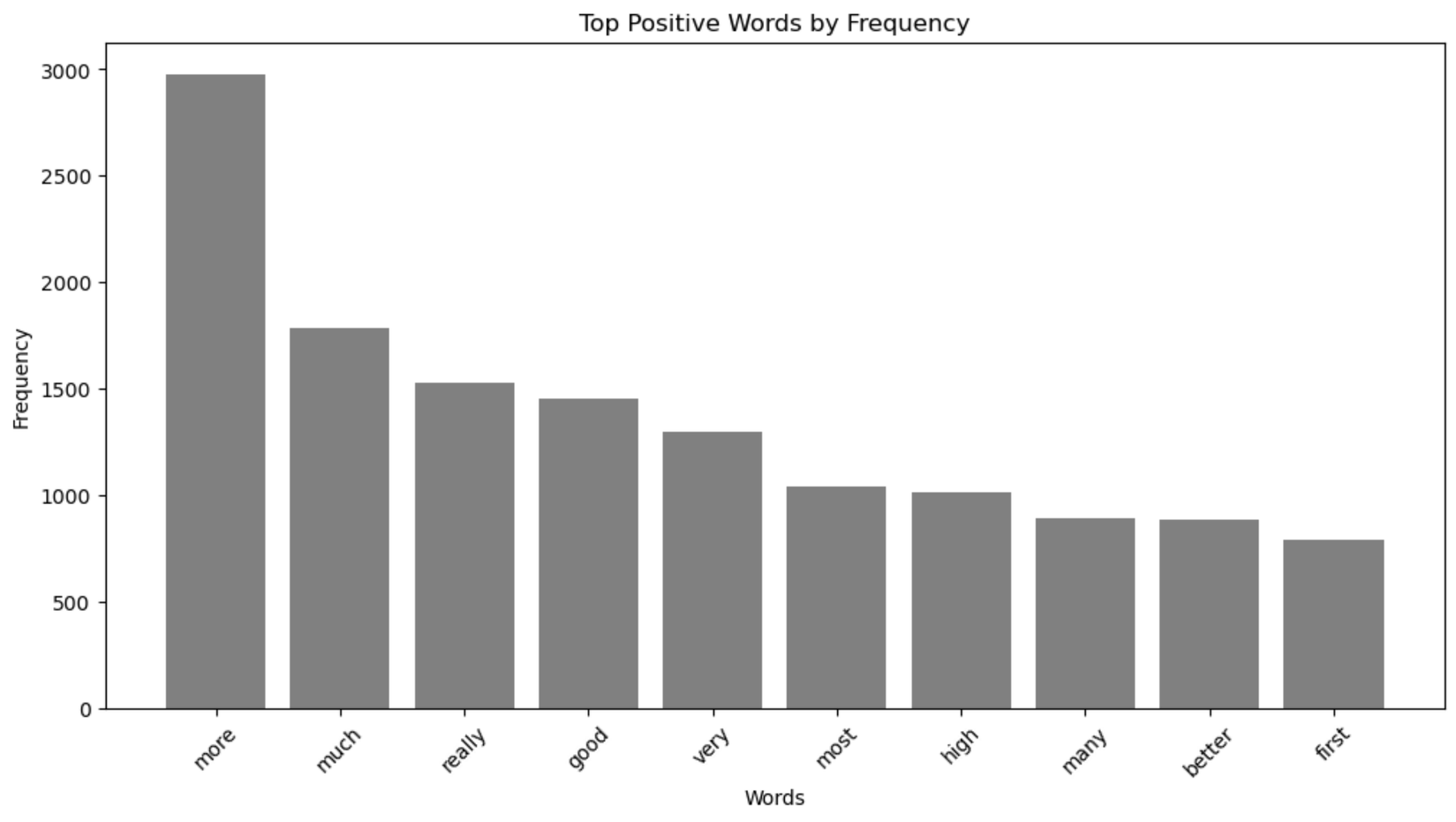
Word CLoud

A word cloud is a visualization technique used to represent text data, where the size of each word indicates its frequency or importance in the text. In a word cloud, more frequent words are displayed in larger fonts, while less frequent words appear in smaller fonts.

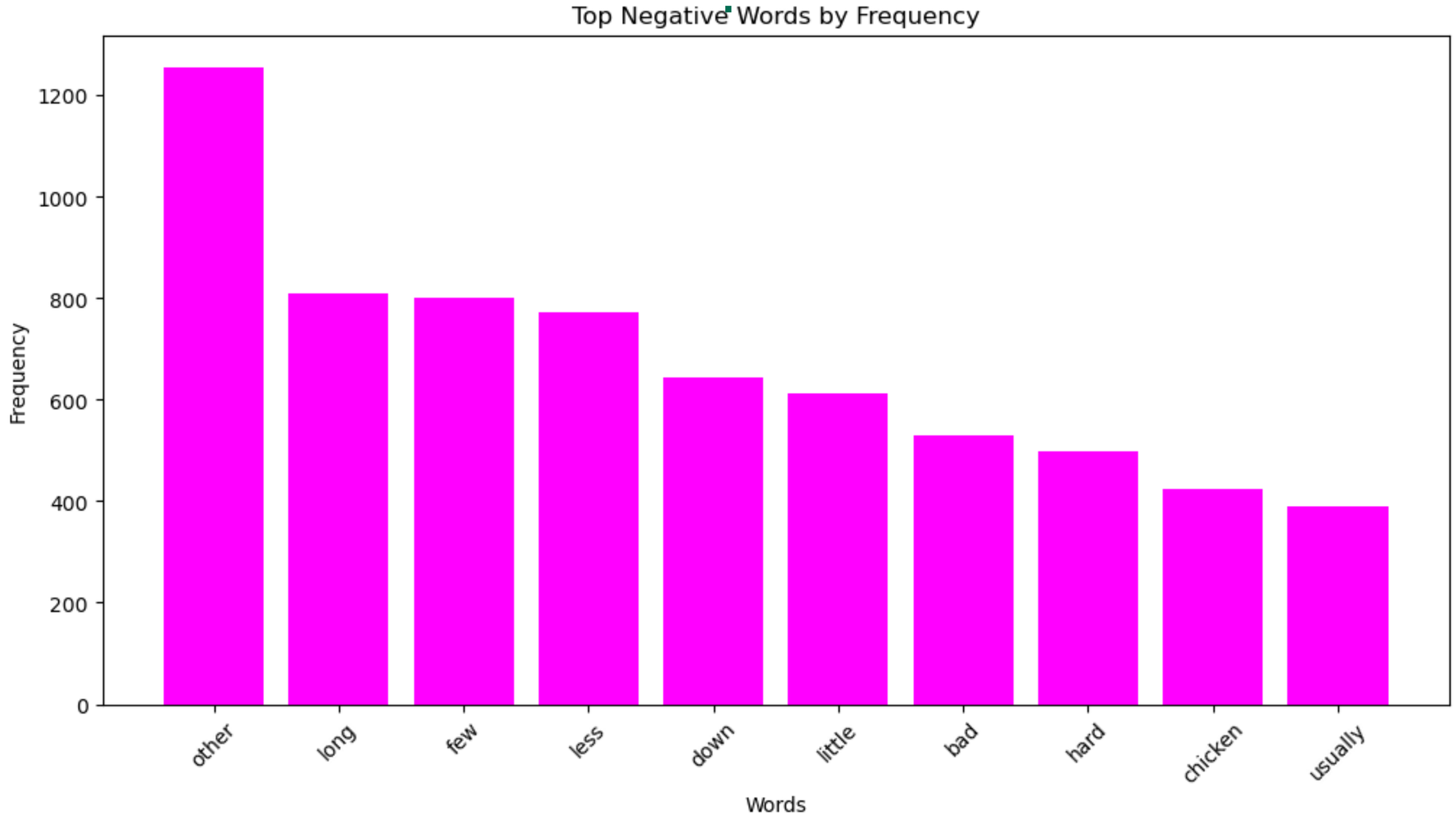
Word CLoud

This visualization shows the most frequent words and the less frequent words on our dataset for example among the Positive comments the words “good” and “More” are the most frequently used as they are written in a big font

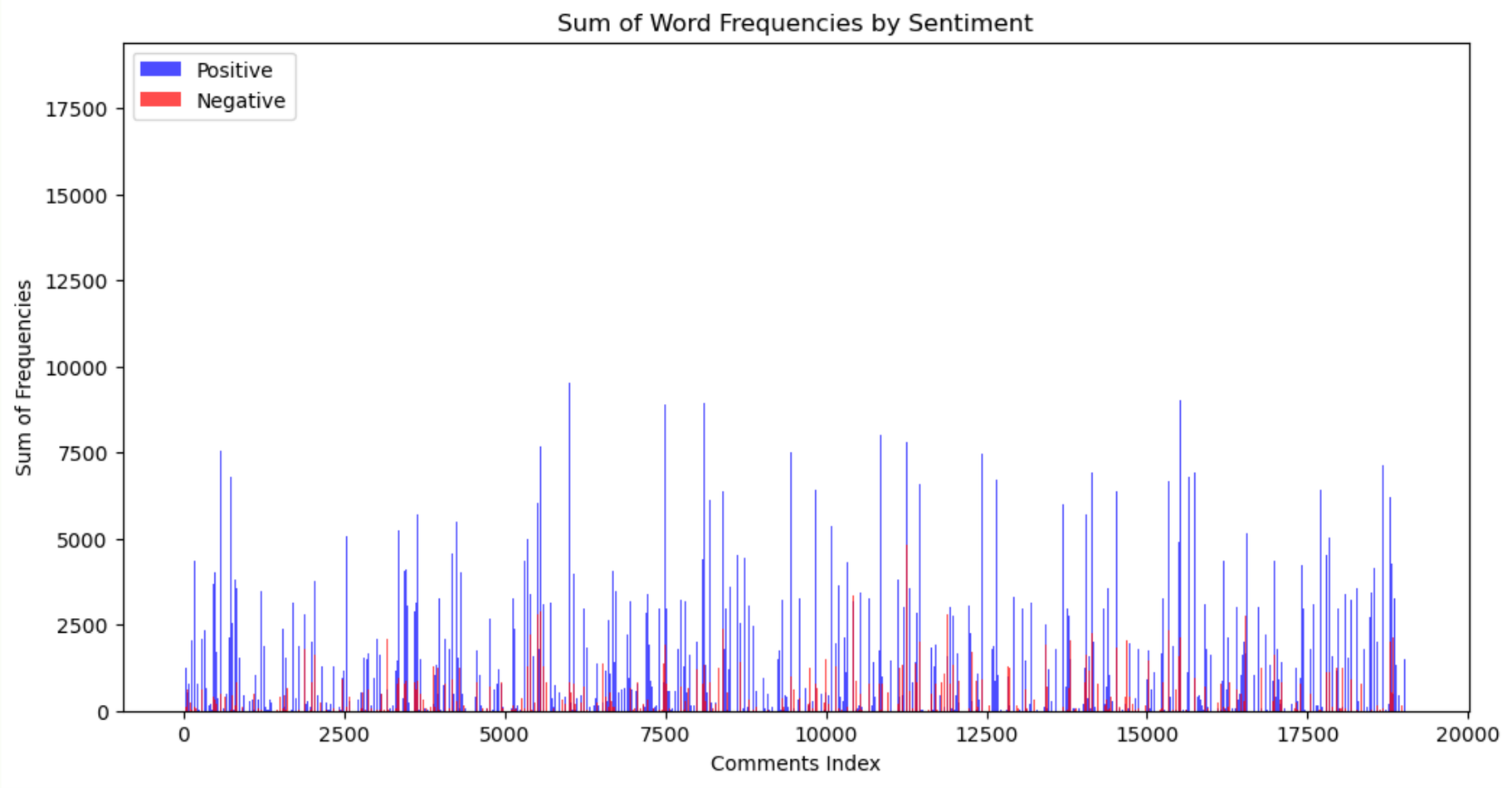
Word Frequency Distribution: Plot the most common positive words
The word ‘more’ is of the most occurrences about 3000 times .
The word ‘first ’ is the least frequent of about 700 occurrences .



Word Frequency Distribution: Plot the most Top 10common negative words
The word ‘other’ is of the most occurrences about 1300 times .
The word ‘usually ’ is the least frequent of about 380 occurrences .



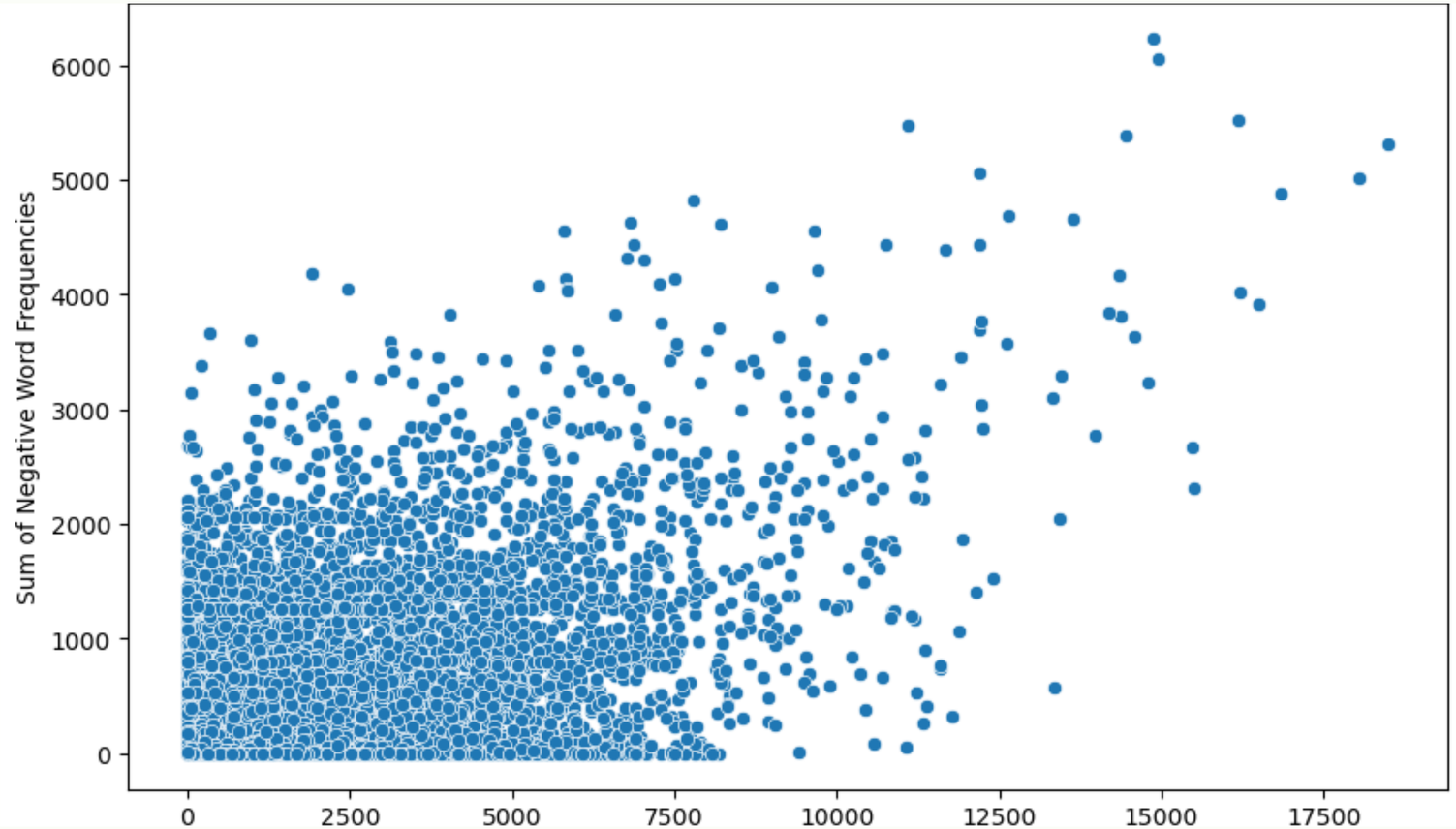
Comparing Positive and Negative Frequencies: For each comment, compare the sum of positive word frequencies to the sum of negative word frequencies.



Scatter Plot

Spread: The spread of points along the x-axis (Sum of Positive Word Frequencies) is wider than the y-axis (Sum of Negative Word Frequencies). This may imply that the variation in the frequency of positive words across the dataset is greater than that of negative words.

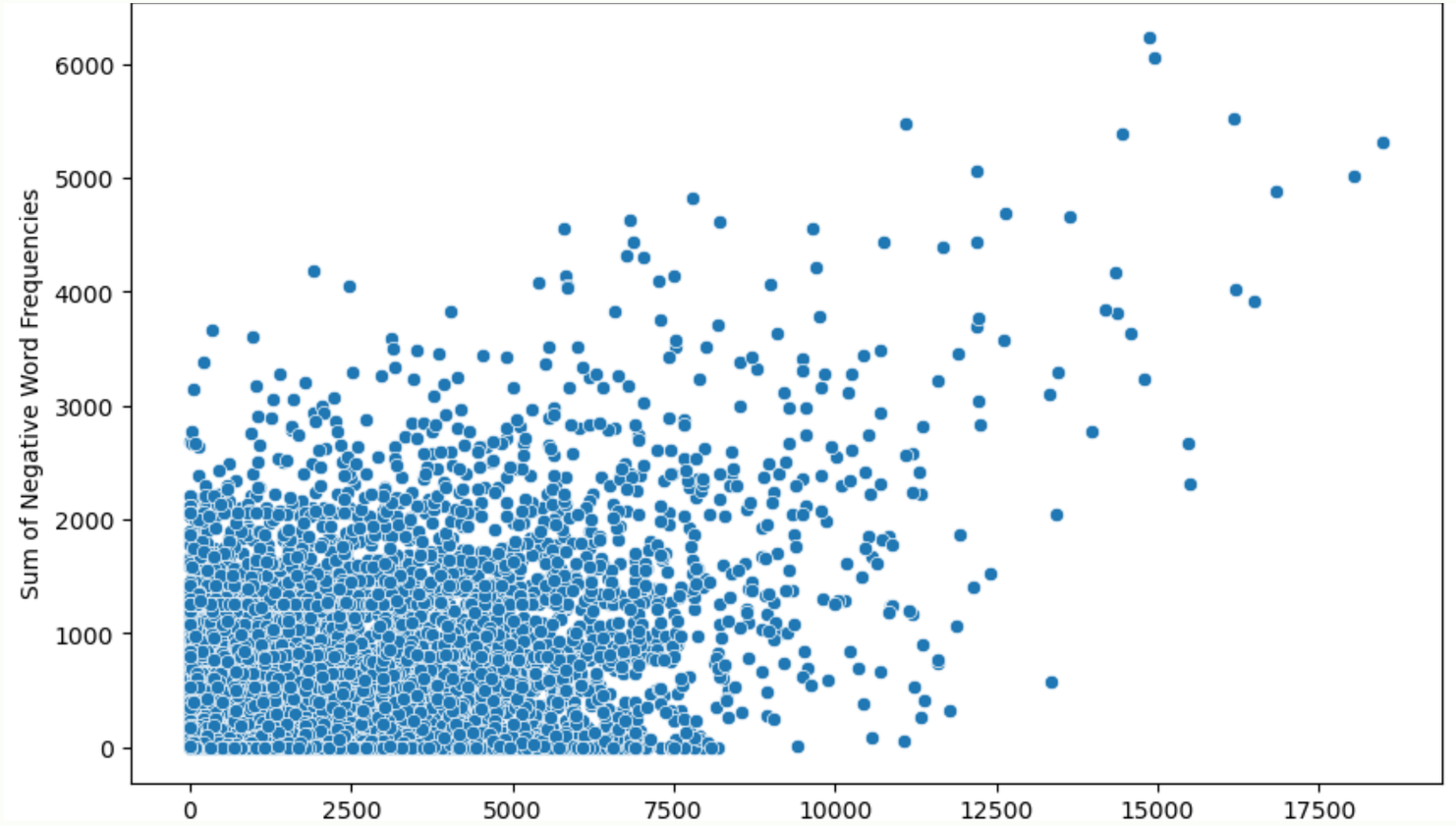
Outliers: There are a few data points far away from the dense cluster, especially along the x-axis. These points may represent texts with an unusually high frequency of positive sentiment words.



Scatter Plot

Correlation: There doesn't appear to be a strong linear relationship between positive and negative word frequencies. The points do not form a clear line or curve but are rather dispersed.

Interpretation: The plot might suggest that texts in your dataset typically either use sentiment words sparingly or tend to express sentiment using positive words more frequently than negative ones. The presence of points across the range of positive word frequencies, but not as much for negative frequencies, could mean that when sentiment is expressed, it's more often positive.



Correlation heatmap

These correlations suggest that there is not a strong linear relationship between the Post Score and the sum of positive or negative word frequencies, but there is a moderate relationship between the positive and negative word frequencies themselves.

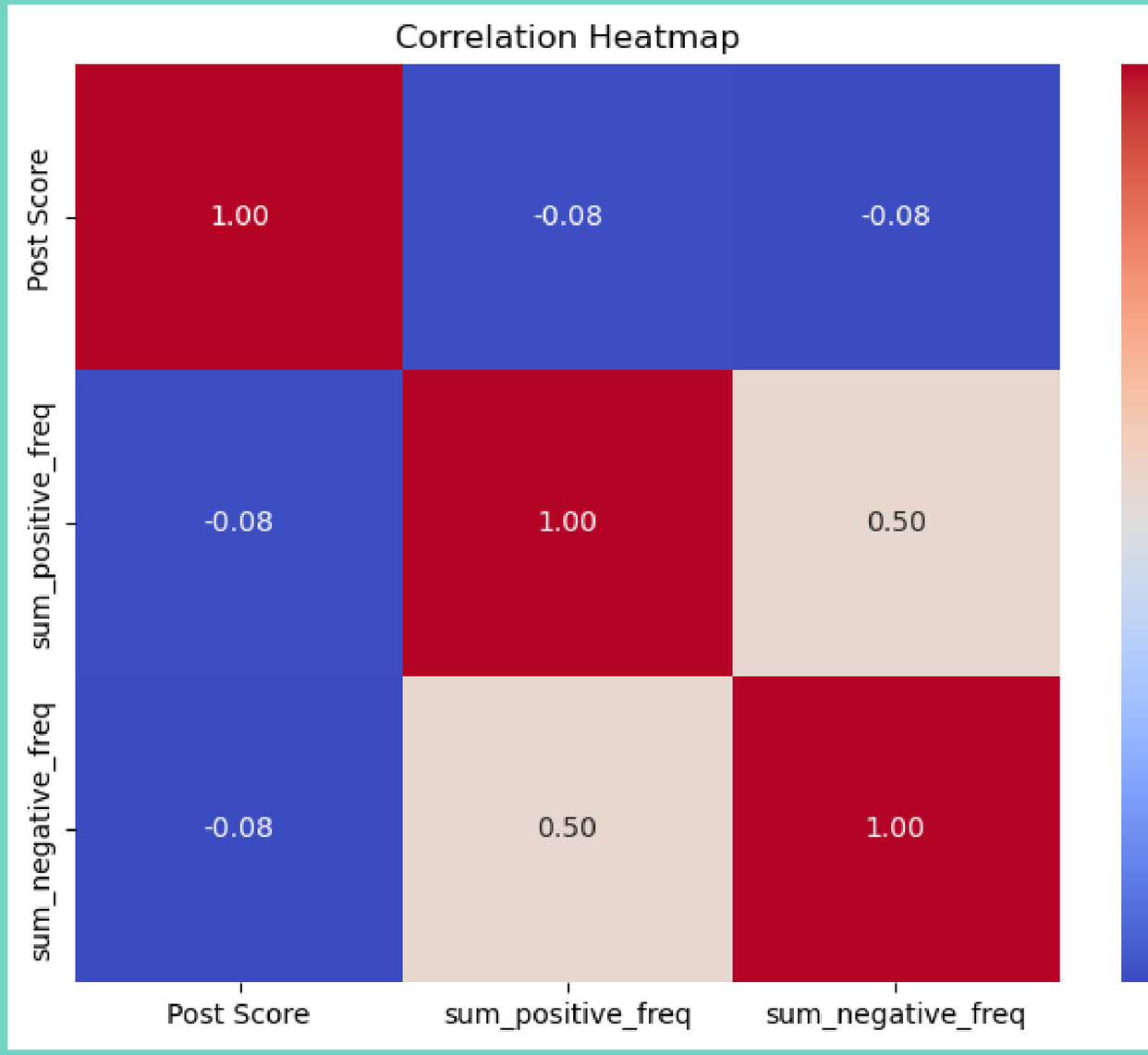
This could be due to various reasons, including the nature of the comments, the context in which words are used, or simply the dynamics of the discussion in the Reddit posts.

Correlation heatmap

Dynamics of Reddit Discussions:

- Reddit discussions can be complex and multi-faceted. A single comment thread can contain a range of opinions and sentiments, sometimes within a single comment. For example, a commenter might praise certain aspects of a post while criticizing others, leading to both positive and negative words being present.

- The sentiment of words can also be context-dependent. The word "sick" might be negative in a health context but positive when used as slang to express approval. Without deeper linguistic analysis, the sentiment scores might not fully capture these nuances.



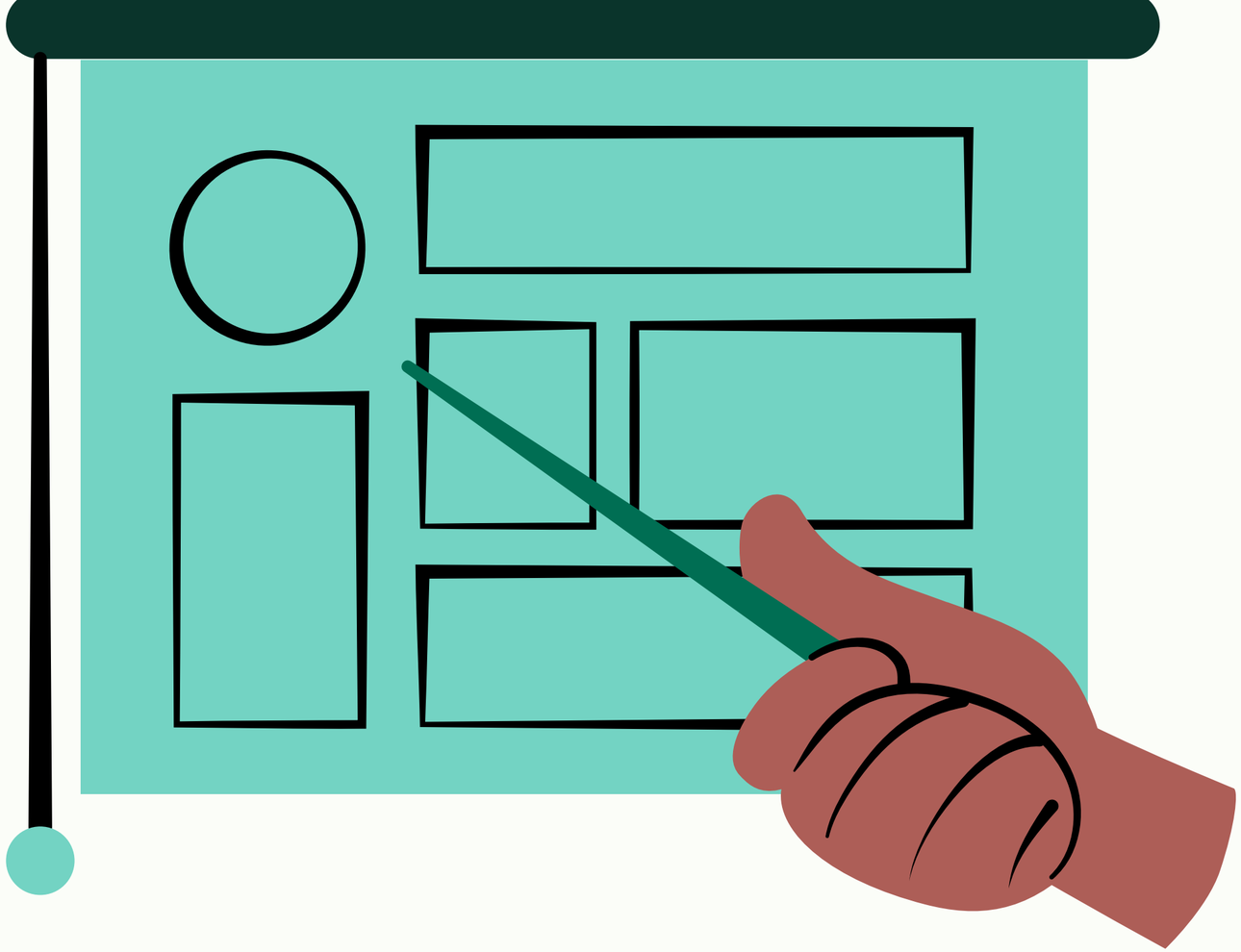
Future Work

- To gain more insight, we could consider additional analyses such as topic modeling to see if certain topics are more associated with higher post scores.

- we might also analyze the timing of comments and their relation to post scores, as the time of day or week might affect user engagement and sentiment.

- It could also be valuable to look at other features such as comment length, the use of question marks or exclamation points, or the presence of URLs, which could influence both sentiment and engagement.

- Reddit.com
- google.com
- https://www.healthline.com/nutrition/ketogenic-diet-101
- Wikipedia
- https://www.wellandgood.com/fitness-wellness-trends
-
- My notebook analysis using python
-



REFERENCES