barbie_data.csv:

- text: full written review written by IMDB user
- rating: number rating from 1-10 assigned by IMDB reviewer
- full_date: date of review written in day/month/year format

barbie_Cleaned.csv:

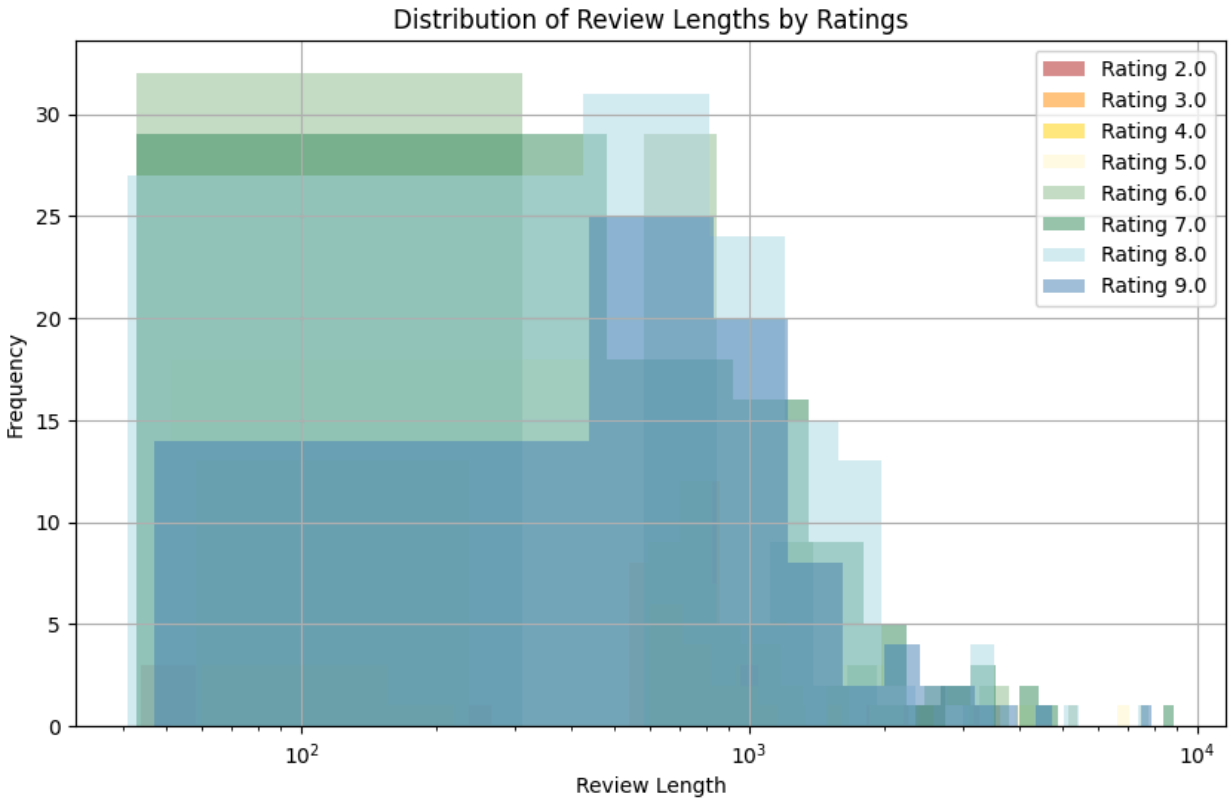
- initial data from Kaggle
    - 797 unique rows
    - last updated 6 mos ago
- text: full written review written by IMDB user
- rating: number rating from 1-10 assigned by IMDB reviewer

barbie_data_vader_scores.csv:

- barbie_data dataset with addition of vader_scores column

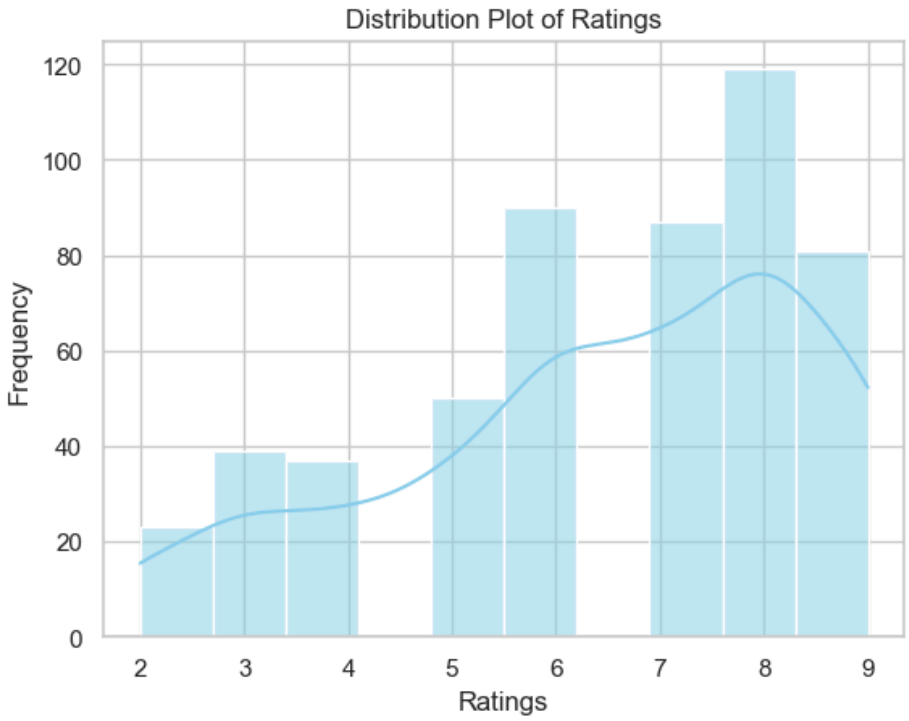
- vader_scores: numerical value ranging from -1 to 1, denoting sentiment analysis
- text: full written review written by IMDB user
- rating: number rating from 1-10 assigned by IMDB reviewer
- full_date: date of review written in day/month/year format

a. Histogram visualizing the distribution of review lengths by number ratings



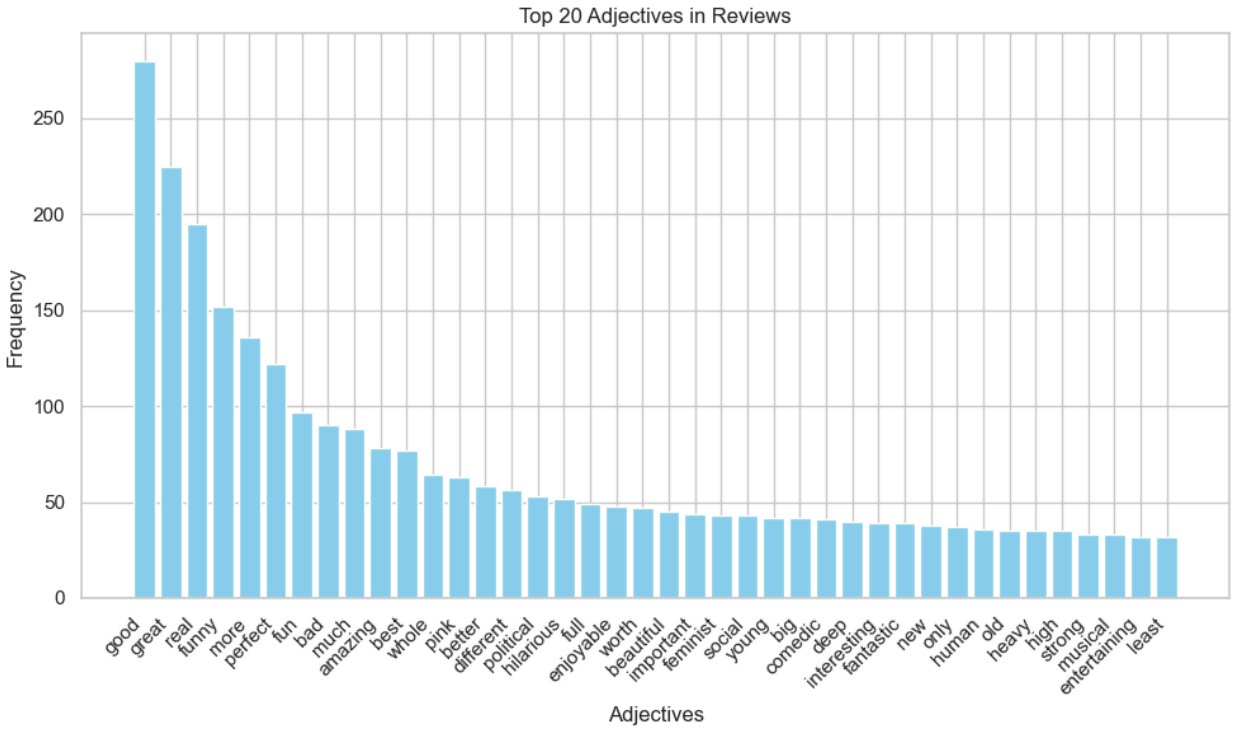
Ratings of 6, 7, and 8 are the most frequent and most ratings are 1000 characters or less.

b. Histogram of distribution of ratings by frequency



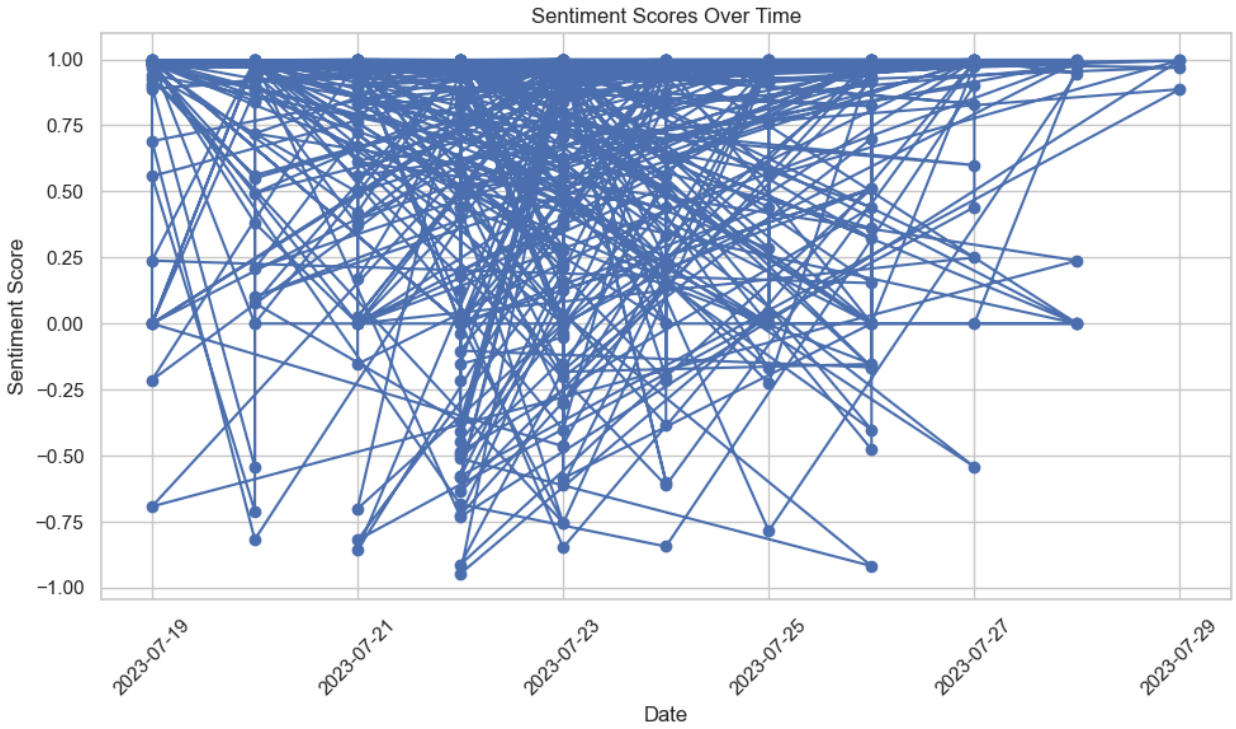
Most frequent ratings are 6, 7, and 8, with 8 being the most frequent.

c. Bar plot of top 20 adjectives by frequency



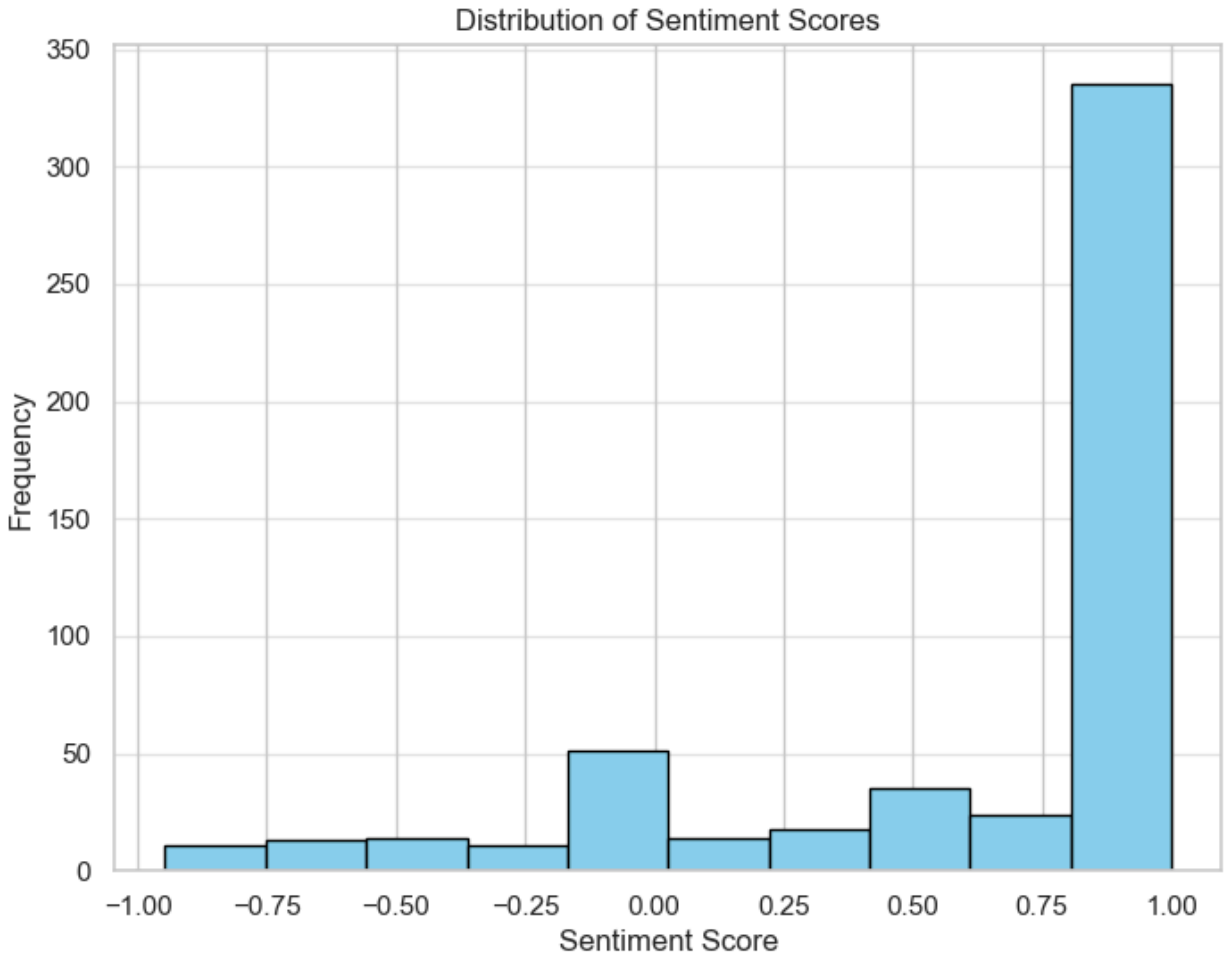
Good is the most common adjective, followed by great, really, and funny. Positive adjectives are the most frequent words featured in reviews.

d. Distribution of sentiment scores over time



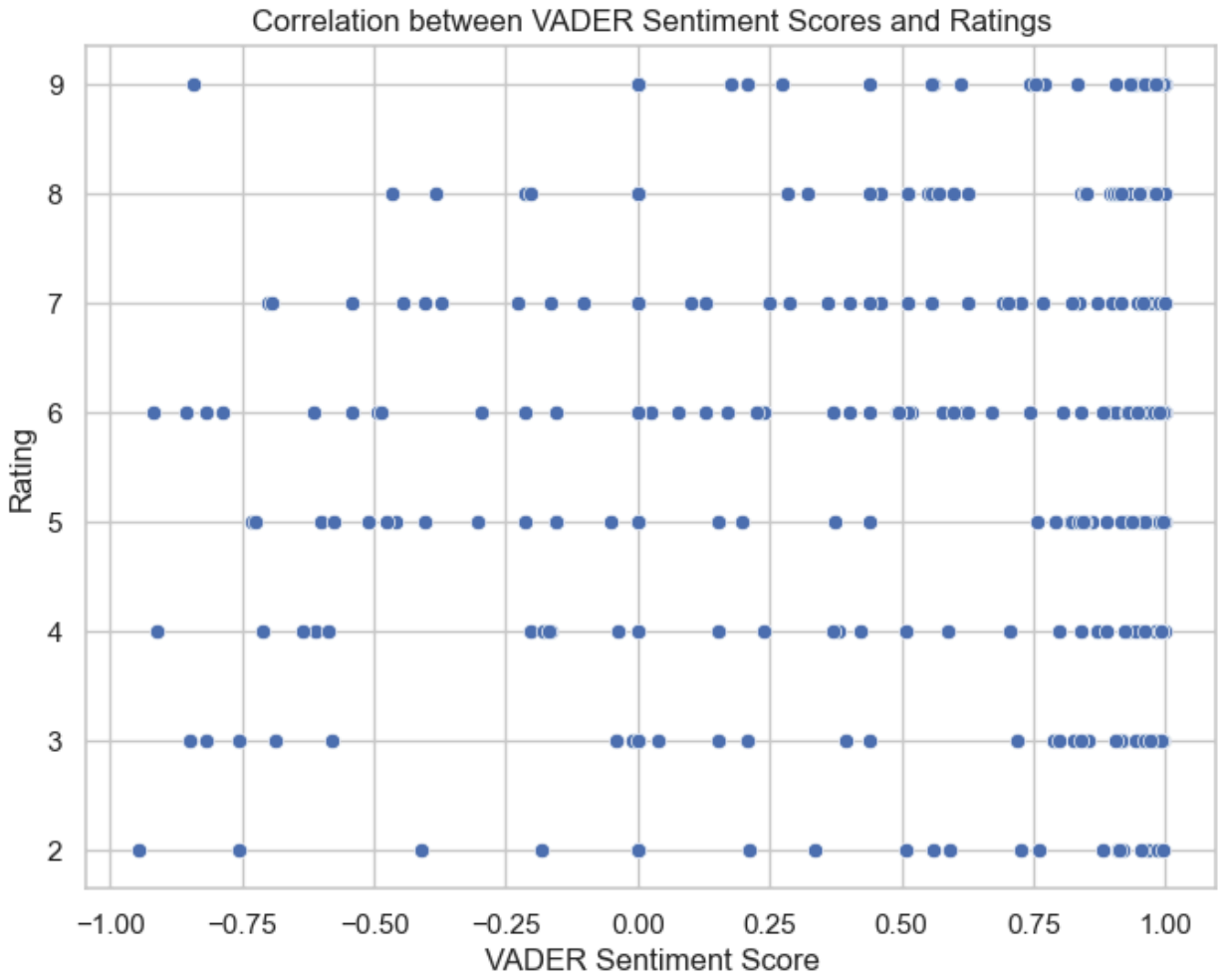
This plot was intended to see if there was a change in sentiment score over time. However, because the time period was only one week, we see that there is no correlation between time frame and sentiment score.

e. Bar plot of distribution of sentiment score by frequency



The most common VADER sentiment score is 1. There is also a spike in frequency around the sentiment score of 0, meaning neutral.

f. Scatter plot of correlation between VADER sentiment scores and numerical ratings



This scatter plot shows that there is not significant correlation between VADER sentiment score and rating.