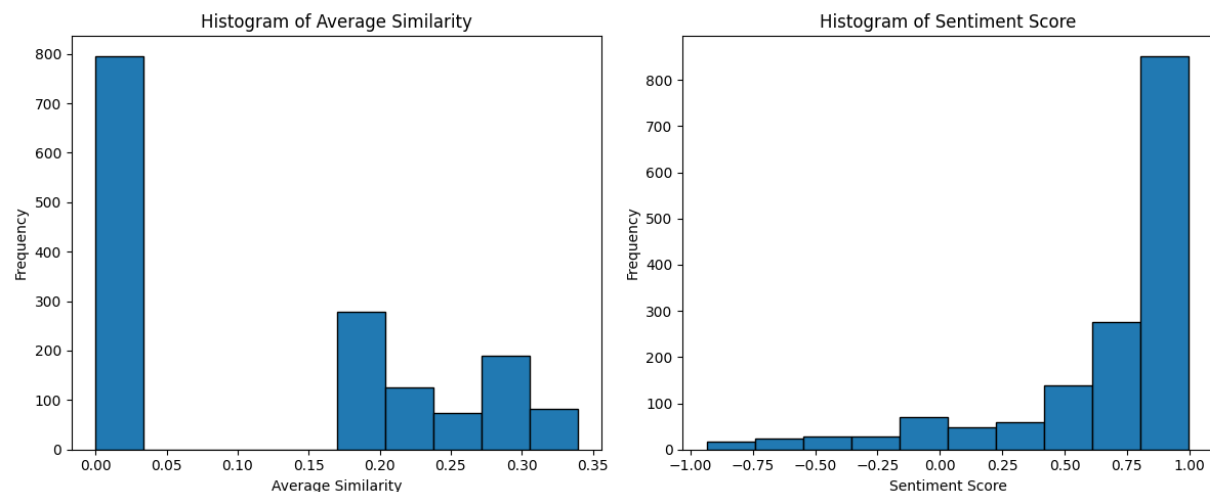


Executive Summary:

The Goal of this project is to create a function recommender system that takes 3 user specified attributes as inputs and then returns 3 specific beer recommendations based on user-generated review data from the beeradvocate.com.

To start this project we utilized BeautifulSoup to scrape user reviews of the Most Popular beers on Beeradvocate.com. We then tokenized the user reviews and removed stop words to come up with a ranking of the most frequently used words. We then utilized the concept of cosine similarity and the TfidfVectorizer of the sklearn package to assign each review a similarity score to the user given attribute. Then using VADER a Sentiment Score was computed and assigned to each review in our dataframe. This data allowed us to gain a better understanding of the data at hand as portrayed in the plots below.



We then computed a Combined Score of our Similarity and sentiment score for our final recommendation by taking a weighted sum consisting of 70% of the Similarity and 30% of the Sentiment Score. This was done to account for the sentiment present in the review to penalize negative and reward positive reviews that are similar to our given attributes. We output our 3 recommended beers by ranking our df in descending order for our combined score at printing the 3 beers with the highest score.

We also explored the impact of using word vectors (e.g., spaCy with pretrained word vectors) versus traditional bag-of-words cosine similarity in the recommendation process. This analysis will involve assessing the percentage of reviews mentioning the user's preferred attributes and comparing the two approaches to determine which is more effective for this specific application. Moreover, we investigated how

recommendations change when we opt to select the three highest-rated products from the entire dataset, factoring in similarity and sentiment scores along with overall ratings. Lastly, to better our understanding of beer attributes and their relationships, we conducted an attribute lift analysis. This involves identifying the top four attributes of beer through word frequency analysis and calculating attribute lifts between these attributes and a selection of ten beers in our dataset. By using lift values, we will identify the most similar beer to a given reference beer. With that, we determined that Imperial Stout was the most similar beer with a lift difference of about 2.354