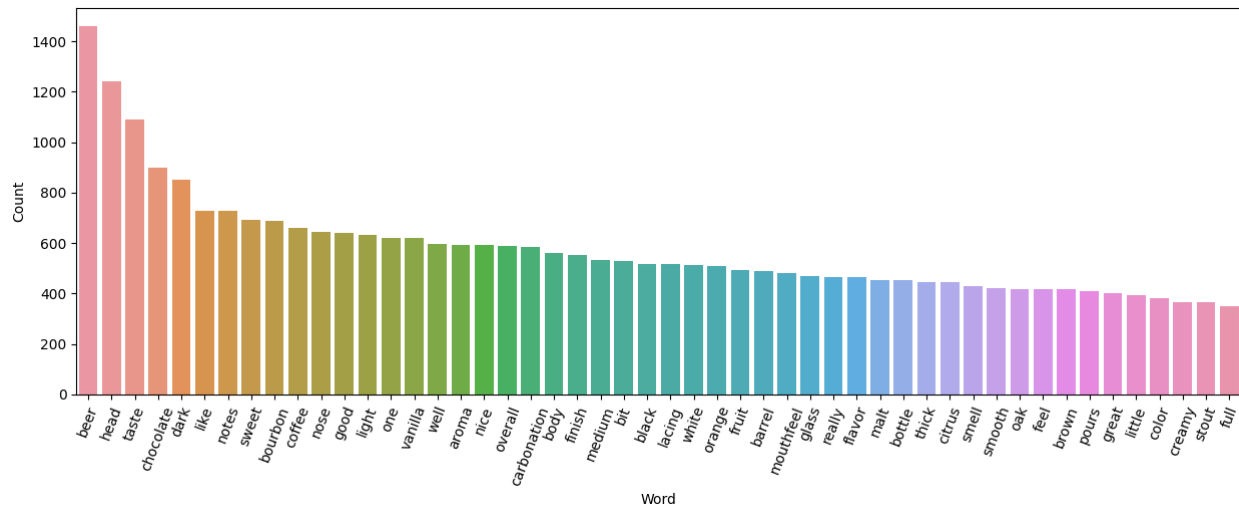


Crowdsourced Beer Recommender System

This beer recommender system relies on ~2500 reviews scraped from beeradvocate.com.

Different brands of beer are scored based on 5 attributes provided by the user. These attributes are what the user looks for in a brand of beer, in no order of priority



Scoring

For the purpose of demonstration, sample attributes are taken from a list of the most frequently occurring words across all reviews. Sample attributes are:

- Chocolate
- Dark
- Sweet
- Bourbon
- Coffee

To begin with, the cosine similarity scores between the desired attributes and reviews are calculated for each review. Then, the similarity scores are averaged across reviews for each brand and stored as `similarity_score`.

	similarity_score
product_name	
A Deal With The Devil - Double Oak-Aged	0.095450
A Deal With The Devil - Triple Oak-Aged	0.138721
Abner	0.023313
Abrasive Ale	0.015972
Abraxas	0.095875

The same process is repeated with sentiment analysis that uses each individual attribute as a reference, i.e. each review is scored on each of the 5 attributes using VADER sentiment analysis. At this point, we have similarity scores as well as the attribute scores for each brand. The average of each brand's attribute scores is calculated and stored as `avg_score`.

	similarity_score	chocolate_score	dark_score	sweet_score	bourbon_score	coffee_score	avg_score
product_name							
A Deal With The Devil - Double Oak-Aged	0.095450	NaN	0.000000	0.48415	0.075433	0.0000	0.279792
A Deal With The Devil - Triple Oak-Aged	0.138721	NaN	0.000000	0.45880	0.000000	NaN	0.458800
Abner	0.023313	NaN	NaN	0.45880	NaN	NaN	0.458800
Abrasive Ale	0.015972	NaN	NaN	0.45880	NaN	NaN	0.458800
Abraxas	0.095875	0.079629	0.069675	0.53690	0.273200	0.0602	0.203921

Lastly, the final evaluation metric `evaluation_metric` is calculated by taking the average of the `similarity_score` and `avg_score` for each brand. In order to produce the top recommendations, we sort by the highest `evaluation_metric`.

Cosign Similarity Scores:

Beer	Rating	Similarity Score	Sentiment Score	Evaluation Score
Black Tuesday - Rum Barrel-Aged	4.224	.0965	.7964	.4465
Crusher	4.4511	.0520	.7650	.4085
JJJuiceeee Machine	4.256	.0566	.7351	.3958

Evaluating Cosine Similarity Calculations Against SpaCy Cosine Similarity Scores

As above, similarity scores are calculated between the attributes and brand reviews, however this time we use the SpaCy library instead of the `get_cosine()` function defined earlier. The scores do deviate a bit, yet the final brand recommendations come out to be the same, just in a slightly different order.

SpaCy Scores:

Beer	Rating	Similarity Score	Sentiment Score	Evaluation Score
Crusher	4.4511	.4669	.7650	.6159
JJJuiceeee Machine	4.256	.4561	.7351	.5956
Black Tuesday - Rum Barrel-Aged	4.224	.3734	.7964	.5844

Group Members: Domitille Chambon, Sam D'Avila, Steven Hobson, Saiansh Raizada, Johann Thomas, & Purva Tiwari