Beer Recommendation Engine

Colin McMorrow

**Abstract**

The goal of this project was to build a recommendation engine for craft beers. There is an enormous amount of beer styles to choose from, along with thousands of different brands for each, and everyone has their own preferences as well. In the new age of personalized services, we also see all kinds of new businesses and applications catering to the individual, so creating a competitive recommendation system could be the difference in gaining or losing customers. By leveraging the natural language processing capabilities with python, we can analyze the tastes of individuals and use that to drive decesions.

**Design**

The data for this project was obtained from a repository of a professor at UCSD, who scraped this data from the website RateBeer.com. The primary task of this project was the utilize text data from beer reviews to help build a recommendation system. I used a variety of methods to process the text from these reviews down to meaningful words of value. I then created a topic model to reduce the dimensions of the sparse matrix of words. Using Latent Dirichlet Allocation (LDA), I consolidated the text into a topic distribution of 5 topics. With the information held in the text now represented in numerical values, I paired them with data about each beer and computed the distance between them. Using cosine similarity distance, the engine produced 3 recommendations.

**Data**

The data was originally scraped by a Graduate lab at UCSD, from RateBeer.com. Each review contained information on the beer, its style, the brewer it came from, and the user who wrote the review. In addition, each review had numerical ratings. One for aroma, taste, palate, appearance, and overall rating. Each row of the data set is one review by one user, for one beer.

**Algorithms**

Pandas, NumPy for data cleaning and manipulation

NLTK, SpaCy, RE for text preprocessing

Gensim for topic modeling

pyLDAvis for visualizing topics

Word Cloud for breakdown for term distribution

Sklearn for cosine similarity distance metric

**Communication**

