Recommending Beers



Cartney Thompson, Ketki Potdar, Guy Duggins

IST 718 - Tuesday 9pm

December 2018

# 

# What is your new favorite beer?

Patrons are always wondering what beer to get at a bar. There are way too many choices for a casual beer drinker and many people usually know only a few beers out of the many choices available To help ease this decision making process, we are building a Beer Recommendation Engine which can present experienced as well as novice beer drinkers with optimal recommendations.

## Data science questions

As part of our data analysis we seek to discover the answers to the following data science questions:

* Given the beers that I have previously rated, what beer would you recommend next that I can try later?
* There are millions of reviews available. Can we provide some useful insights about users to Beer business owners, which can help their business?

# Key Findings

* From our dataset, we found that if users rate the beer highly overall, they also rated the other features of the beer - aroma, taste, appearance, and palate highly and vice versa.
* The overwhelming majority of beer reviews (73%) come from the Ale category
* 82% of beers reviewed were rated between 3.5 - 4.5 on a five point scale.
* Overwhelming majority of beers reviewed fall between 5% and 8% alcohol by volume
* Recommended beers via Matrix Factorization was based upon latent features derived from overall beer ratings from users

# Dataset

## Beer Reviews

This main dataset consists of beer reviews from Beeradvocate. The data spans a period of more than 10 years, including ~1.5 million reviews through November 2011. Each review includes ratings in terms of five "aspects": appearance, aroma, palate, taste, and overall impression using a 5-point scoring system with 1 being the lowest and 5 being the highest. Reviews include product and user information, followed by each of these five ratings, and a plaintext review. The table below itemizes the set of relevant properties defined for each review.

|  |  |
| --- | --- |
| **Property** | **Description** |
| ABV | The beer’s alcohol by volume |
| Beer Name | Name of the beer |
| Beer Style | The style of the beer (i.e. Hefeweizen, Stout, IPA, etc) |
| Appearance Rating | The appearance of the beer rated on a scale of 1-5 |
| Aroma Rating | Aroma rating on a scale of 1-5 |
| Overall Rating | Overall rating on a scale of 1-5 |
| Palate Rating | Palate rating on a scale of 1-5 |
| Taste Rating | Taste rating on a scale of 1-5 |
| Review Text | Custom user review of the beer |
| Review Date | The timestamp of when the review was submitted to beer advocate |
| Reviewer Profile Name | The unique username for the reviewer |

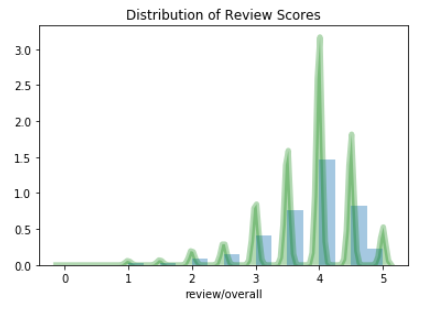
## Beer Category Hierarchy

With more than 66K unique beers and more than 100 beer styles, we used a data set defining beer hierarchies to help improve our ability to classify the beers during modeling.

|  |  |
| --- | --- |
|  | **Beer Style**  The style of the beer (i.e. Hefeweizen, Stout, IPA, etc)  **Beer Category**  The top level category for the beer (i.e. Ale, Lagers, etc)  **Beer Sub-Category**  The sub-category for the beer (i.e. Pale Ale, German Lager) |

# Analysis

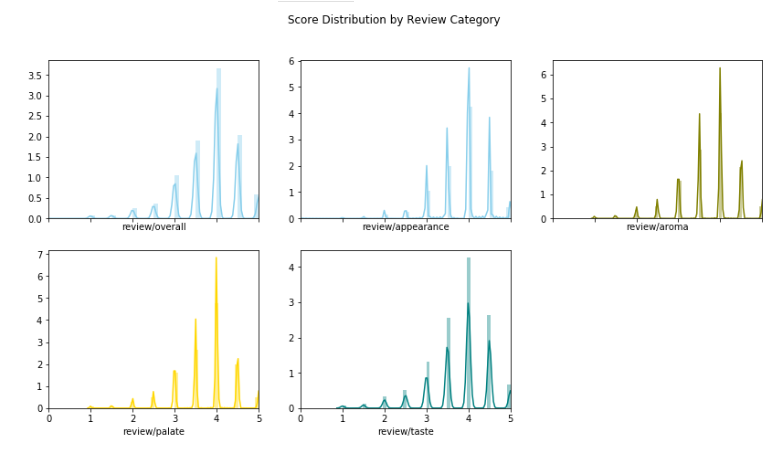
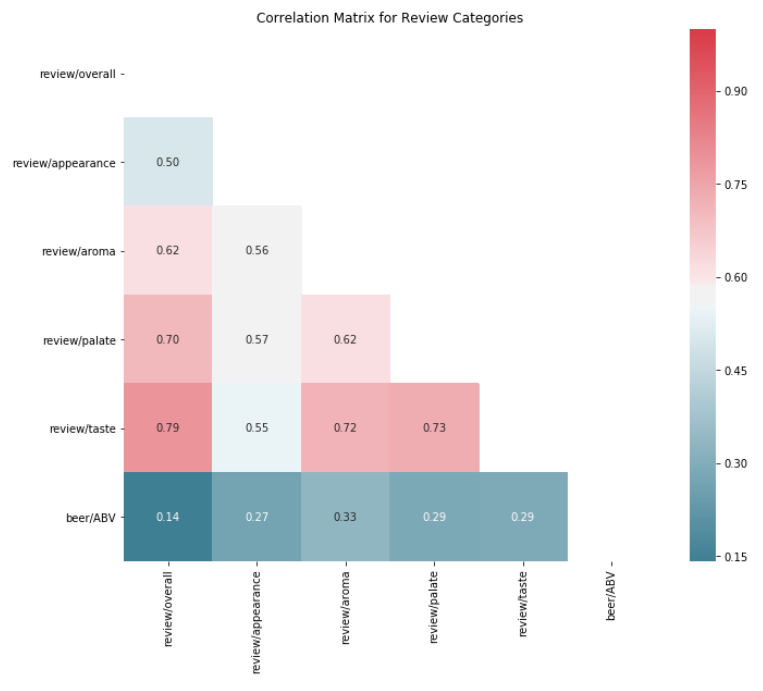
## Beer Review Scores



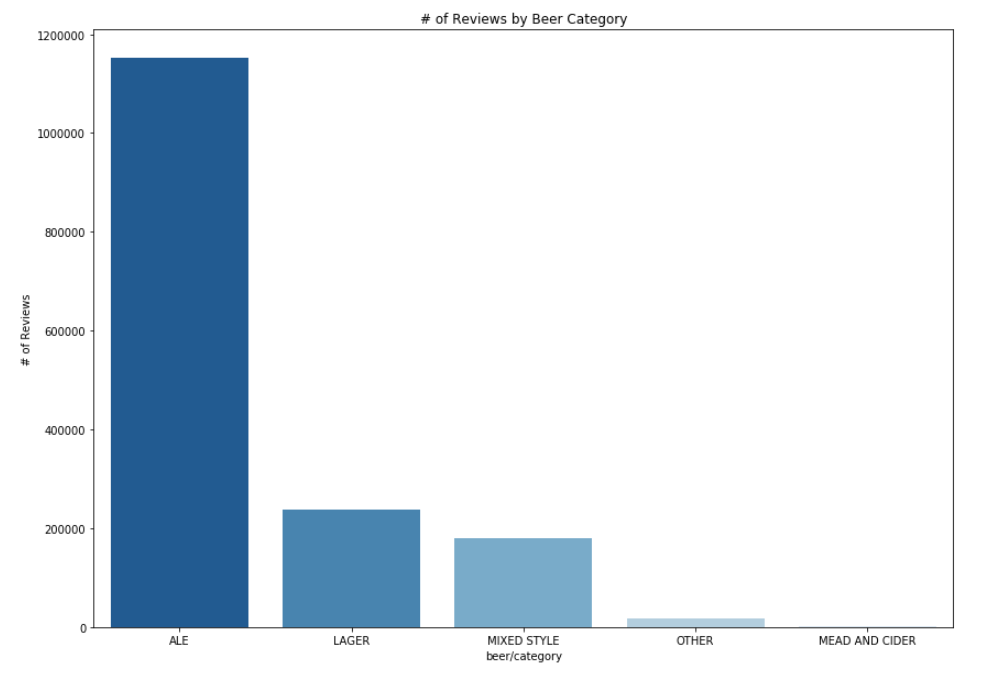
The overall distribution of user beer reviews tends to skew more towards 3.5 - 4.5 on the 5-point scale. Reviews between 3.5 and 4.5 make up 82% of all reviews in the dataset. In addition to overall review score, beers in the dataset are scored in five categories;

Overall score, appearance, aroma, palate, and taste.

The review categories are individually scored by each user.

For our recommendation project, our preference is for the attributes to be fixed based upon characteristics of each beer, similar to how alcohol by volume is fixed. For example, Caldera Ginger Beer has different aroma and palate scores from its nine reviewers.. As a result, each score category is highly correlated with the overall score. The review category that’s least correlated with the overall score is appearance, and even that score is still above 0.50. Also, the individual score distribution for each review category follows a similar distribution to other review categories.

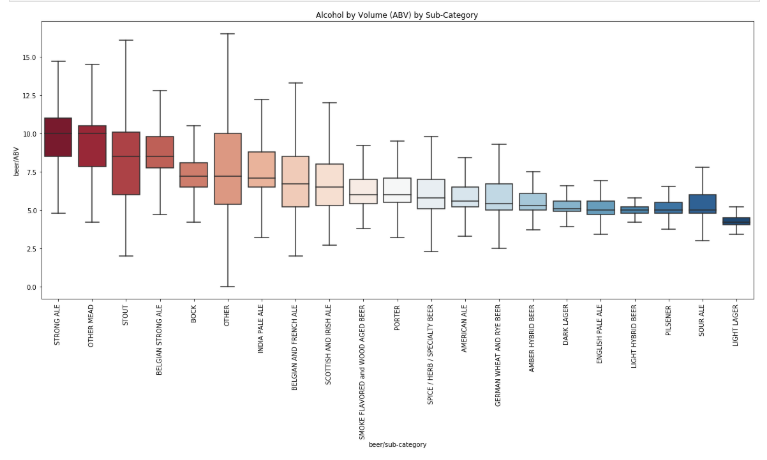
## Beer Categories

Most reviews come from beers in the Ale category. In our data set, roughly 73% of all reviews came from beers categorized as ale, followed by 15% as lagers, 11% as mixed styles, and 1% as other style of beer. For sub-categories, India Pale Ale (14%), American Ale (13%), Stout (11%), Belgian and French Ale (10%), and Pilsener (9%) made up most reviews. Combined, those five sub-categories make up roughly 24% of beer sub-categories, however they constitute more than 57% of total beer reviews. 

## Distribution of Beer Styles

### Alcohol By Volume (AVB)

Typically, most beers have an alcohol by volume (ABV) in the 5% - 8% range. This distribution varies by sub-category.



Stouts and “other” sub-categories of beer tend to have the widest range in alcohol by volume percentage. This is because the stouts in our data set tend to have a high number of variations; from milk and oatmeal stouts (that tend to have lower ABV, in the 3-4% range) to imperial stouts that typically have a ABV content over 9%. Not surprisingly, the lowest alcohol by volume content beer sub-categories tend to have a lighter appearance (Sours, Pilseners, and Lagers).

## The Reviewers

## User profile characteristics are largely unknown, but in order to generate some attributes, we classified reviewers into categories that reflected how active they were as reviewers.

56% of reviewers are considered newbies, with another 21% considered sippers. Yet, 46% of the reviews were performed by 2% of our reviewers considered to be connoisseurs.

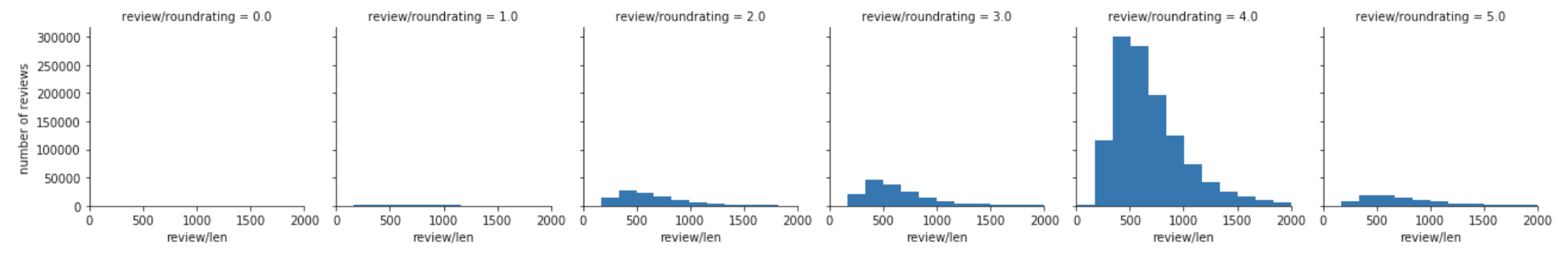
## Beer Reviews Over Time

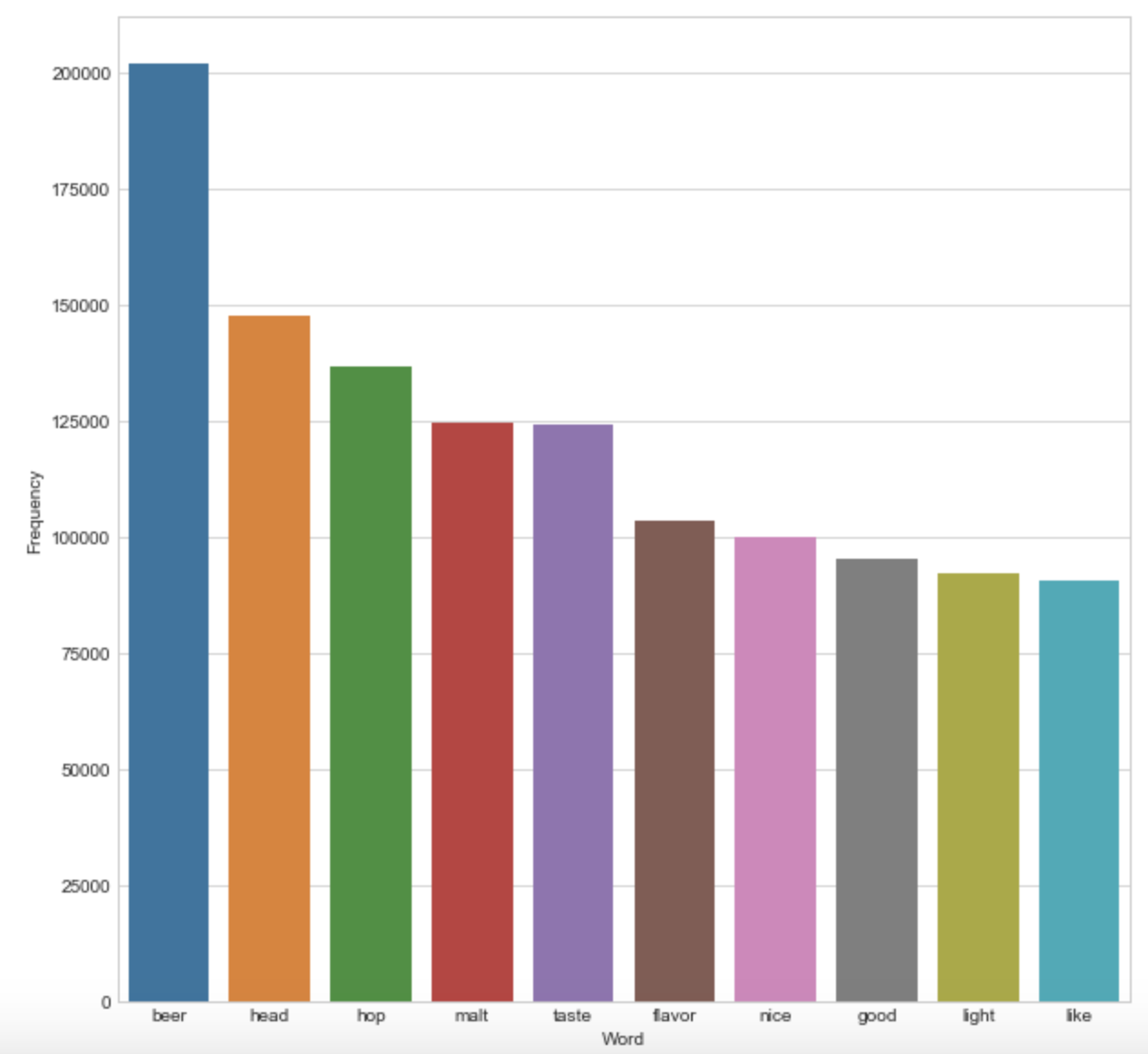
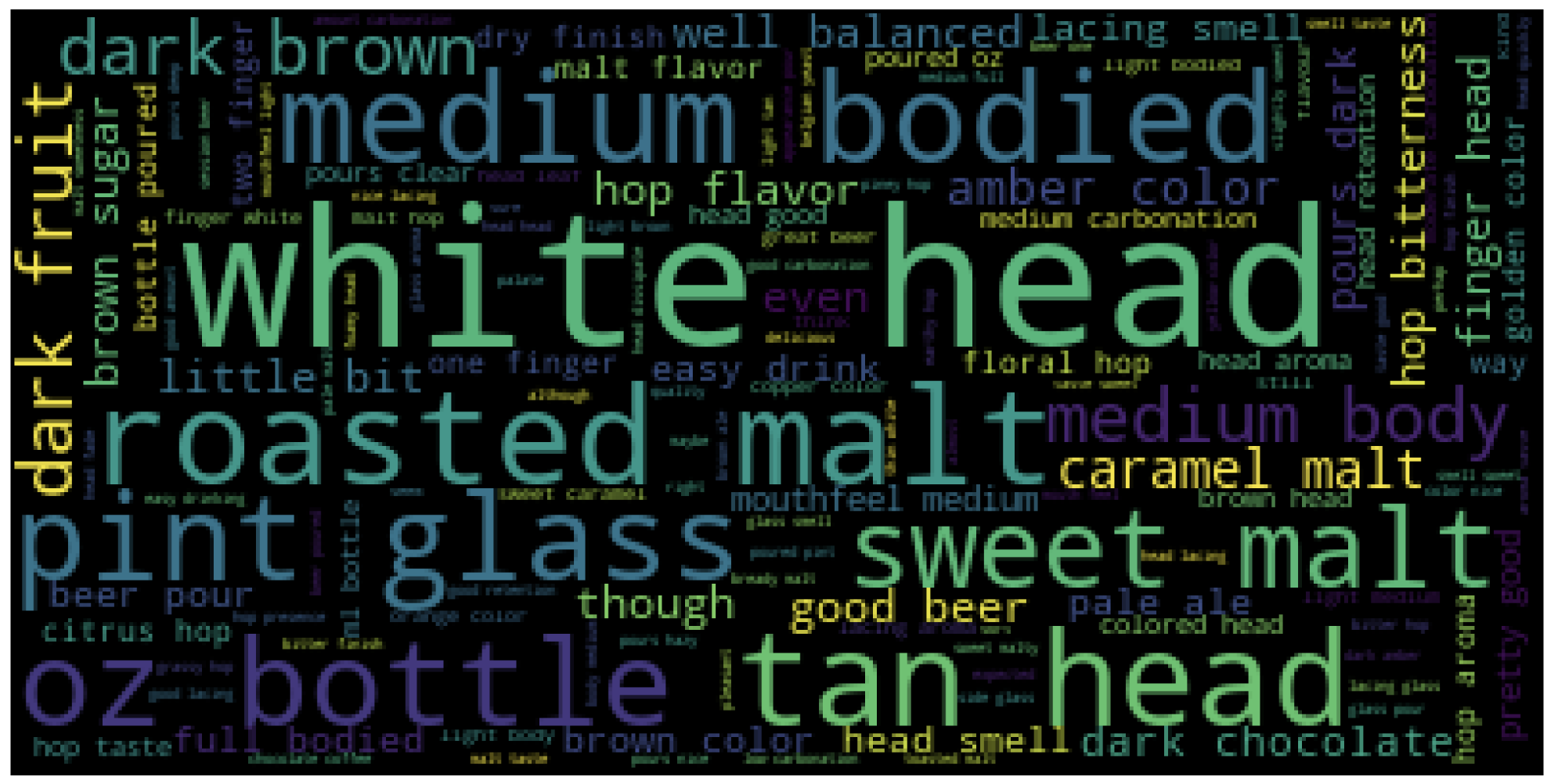
As the number of reviews and reviewers increased over the years, the average ratings across the 5 categories of beer remained relatively consistent. In more recent years, reviews of lagers have seen a modest decrease in their overall ratings. Ales however, have consistently had the best average rating of any category.

## Analysis of the Review Text

The dataset has ~1.5 millions of user reviews for beers belonging to 104 categories. This information can provide good insights to beer business owners about beer drinkers and their preferences. But it’s almost impossible to view and process all these reviews manually. The text mining of the review text can help.

As apparent from the exploratory data analysis, the dataset is biased consisting mostly positive reviews with rating 4. Most of reviews are 500 to 1000 words long.





The word cloud and word frequency map shows the reviewers have mentioned beer head in the reviews quite a lot. Also taste and flavor as well as malt play important role while reviewing the beer.

As a first step in text analysis, the reviews are categorized as positive reviews and negative reviews based on the user rating with reviews having rating > 3.5 stars as positive reviews and reviews with rating < 2.5 stars as negative reviews. The reviews data is then cleaned to remove stop words and non-text words and tokenized with TFIDF and 1-grams. TF-IDF (Term Frequency – Inverse Document Frequency) is numeric statistic used to find how important the word is in document. The tf-idf value increases proportionally to number of times the word appears in the document and is offset by number of documents in the corpus that contain the word.

NMF (Non negative matrix factorization) is then applied to positive and negative sparse matrices obtained from TF-IDF. NFM is algorithm in multivariate analysis where a matrix is factorized into two matrices with the property that all three matrices have non negative elements. A document term matrix is factored to obtain top topics and their corresponding weights in reviews text. This makes reviews much more easier to understand.

### 

### Top Positive topics

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Legend**   |  |  | | --- | --- | | 0 | Hop / malt / citrus | | 1 | Coffee / chocolate / roasted | | 2 | Sweet / caramel | | 3 | Like / smell | | 4 | Yeast / wheat / lemon | | 5 | Malty / color | |

The chart on left shows association of top positive topics with top 6 positive reviews. Almost all 6 reviewers have mentioned about the smell. They also seem to prefer wheat beer.

### Top Negative topics

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Legend**   |  |  | | --- | --- | | 0 | Hop / malt / bitter | | 1 | Pale / watery | | 2 | Bad / drink | | 3 | Sweet / syrupy | | 4 | Coffee / chocolate / roasted | | 5 | Cherry / fruit / tart / vinegar | |

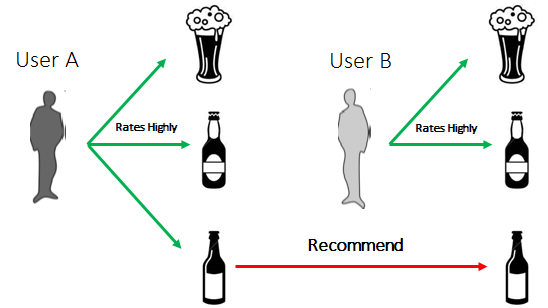
Though ‘bad drink’ is the common phrase in thee reviews, these reviewers have more specifically mentioned that some beers are pale and watery.

These top topics help to summarize positive as well as negative reviews. Our recommendations for beer business owners would be based on these topics.

# Model

## Overview

We will build a model based on collaborative filtering leveraging user beer ratings. How collaborative filtering works is as follows:



**For Example**

User A rates highly three beers:

* Pilsener
* Light Lager
* Sour Ale

User B likes two of those same beers:

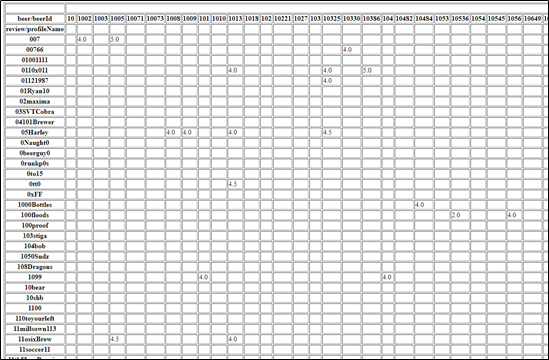
* Pilsener
* Light Lager

Therefore, we will recommend the Sour Ale to that user. That’s a simplistic view, but that’s the end goal. In order to hold the users and their rating, we then create a matrix to hold the data.

Our constructed matrix is very sparse as we cannot expect reviewers to review all of the more than 66K beers defined. In order to build a recommendation system, we need to fill in those empty cells with predicted values. We can get those values by Matrix Factorization U x B where:

U = User attributes

B = Beer attributes

We do not have explicit values for user attributes (this would be user preferences for beer i.e. User A rate beers with certain characteristics highly) or for beer attributes. For us, these are latent attributes which we could get the values of from the actual beer rating from users. In order to get U and B latent feature values, we will use the alternating least squares algorithm in spark to learn the latent factors.

## Collaborative Filtering: User Recommendations

We built our model utilizing the ALS (alternating least squares) module in Spark. To prepare our model, we utilize the following options:

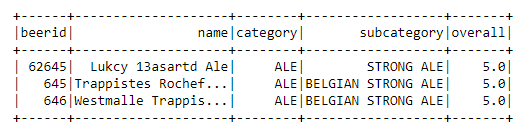
* *implicitPrefs* (False): Users provide explicit beer ratings so we will utilize false for this option.
* *nonnegative* (True)
* *rank* (4): We will assume four latent factors
* regParam (Default): We will keep our regParam at the default value.

After building our recommendation engine, we have a root mean square error (RMSE) of 0.55. Below is an example recommendation use case.

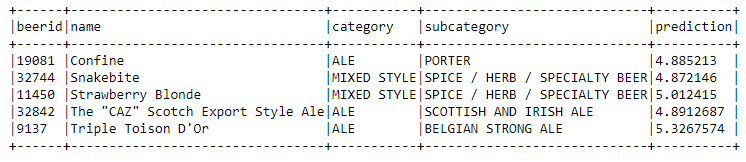
### User Recommendation Use Case

**User Profile:** Rifugium

**Top Beers Rated**



**Recommended Beers**



# Recommendations

* Most beer drinkers care a lot about beer aroma, it would help to invest more into improving beer aroma
* Flavors like coffee, chocolate as well as sweet taste are either hit or miss, some people prefer these flavors but others might not like those. Careful targeting of these flavors can prove beneficial without negatively impacting the brewer’s brand.
* Hops with citrus flavor are a user favorite while bitter flavor is least preferred
* Pale watery beer is also least favorite among users
* Some users don’t like cherry and tart flavors

# Areas of Further Analysis

After building our recommendation engine, we identified the following areas where further analysis would be beneficial.

## Additional Data

We would like additional data related to user preferences and beer characteristics to improve the results of our collaborative filtering recommendation engine

*We used the ALS method to get user and beer attributes to predict ratings. However our recommendations would be a lot stronger if we knew what users preferred about a beer (i.e. hoppy, foamy, dark, strong smell, fruity, etc) and the characteristics of each beer via a scored record.*

## New Real World System

We would want to test our recommendation system in the real world to help tune our algorithm to make better predictions.

*Our recommendation system gets better once we get more feedback. Has a user tried a new beer? If so, how would they rate it? How does it compare to our predicted rating? This is valuable information as if you remember, our matrix was very sparse. The more information we know about user beer preferences, the better our recommendation algorithm can be.*