Beer Recommender

Mario Burstein

December 20, 2017

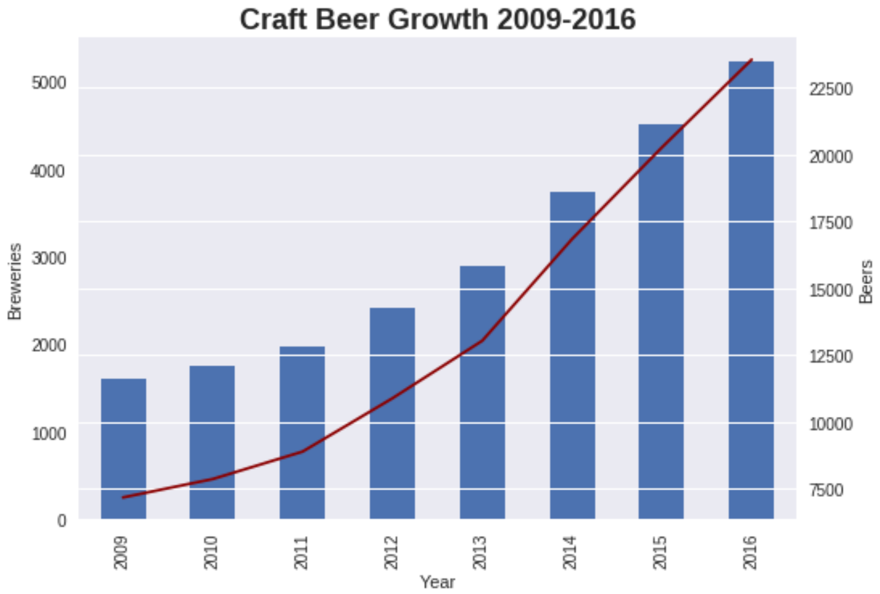
Background & Overview

Background

The U.S. craft beer industry has exploded with the number of craft breweries growing at an annual rate of 19% from 2009 to 2016. As a result the number of craft beers produced in the U.S. has increased from 7.5K to 22.5K over the eight year period. The increase in available beers has left consumers asking the following questions:

1) **What beer do you recommend?**

2) **How do I know if I will like the recommendation?**



Source: https://www.craftbeer.com/editors-picks/craft-brewing-growth-statistics-2016-ba-report

Project Overview

In response to the adjacent questions I have developed a beer recommendation system for American Ales[1], which includes popular beer styles such as American IPA, DIPA, Stouts, Imperial Stouts, and Ambers among others[2]. Additionally Fruit Beers have been included given their increased prevalence in craft breweries. The recommendation system was developed based on beer drinker ratings. The project was completed in the following phases:

1) **Data Collection and Storage** – Data sources, methods of extraction, and storage.

2) **Exploratory Analysis** – Understanding the underlying data and developing inferences/assumptions.

3) **Similarity Measurements** - Beer drinker ratings are used to determine drinker and beer similarities.

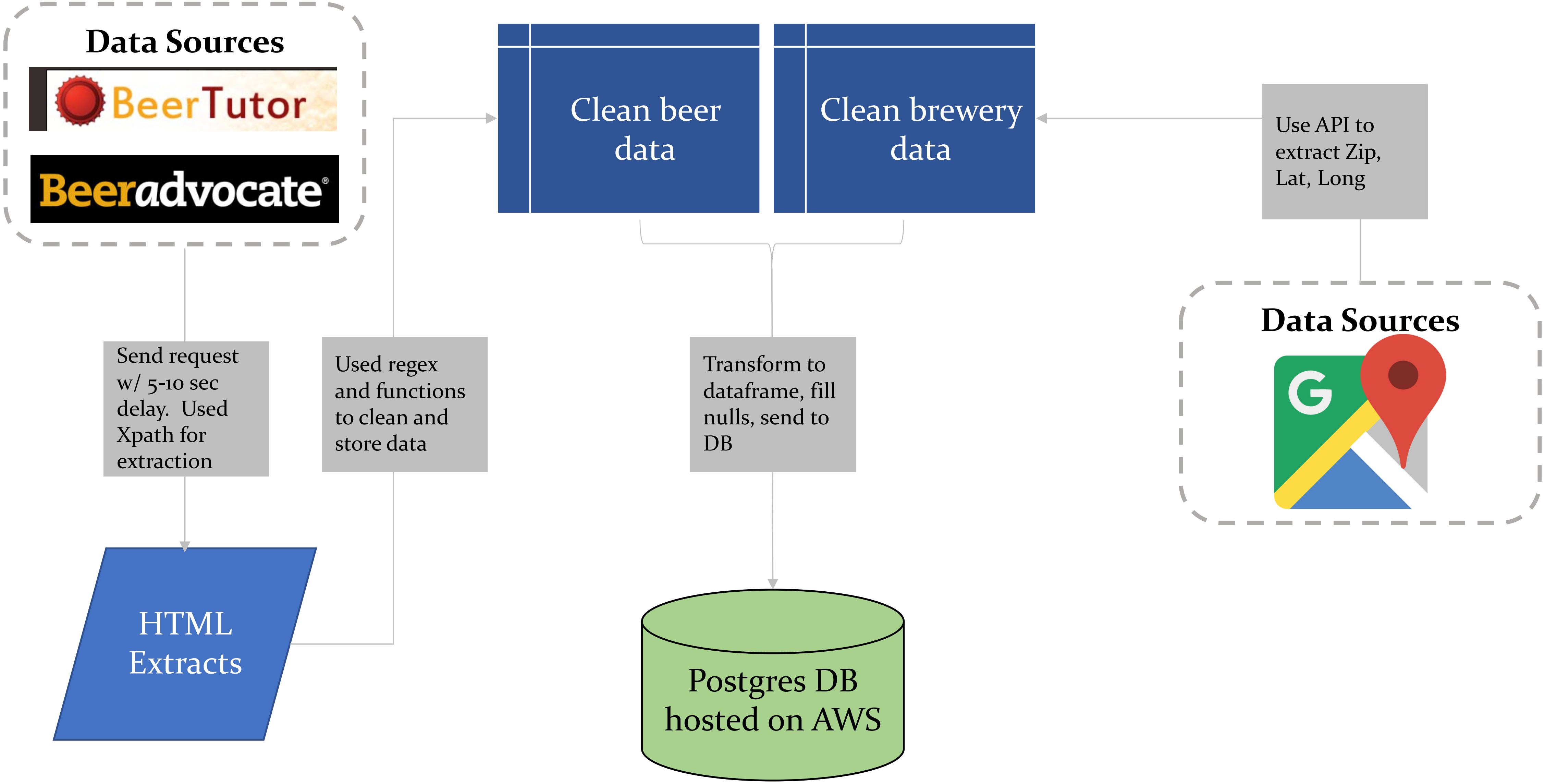
4) **Classification Modeling** – Specific beer attributes are used to determine whether a drinker will like the beer.

5) **Final Recommendation** - Putting it all together.

[1] This category of beer uses yeast that ferments at the "top" of the fermentation vessel, and typically at higher temperatures than lager yeast (60°-75°F), which, as a result, makes for a quicker fermentation period (7-8 days, or even less). Ale yeast are known to produce by-products called esters, which are "flowery" and "fruity" aromas ranging, but not limited to apple, pear, pineapple, grass, hay, plum, and prune.

[2] Others include American Pale Ale (APA), Wild Ales, Cream Ales, Black Ale, Strong Ale, and Brown Ale

Phase 1: Data Collection & Storage

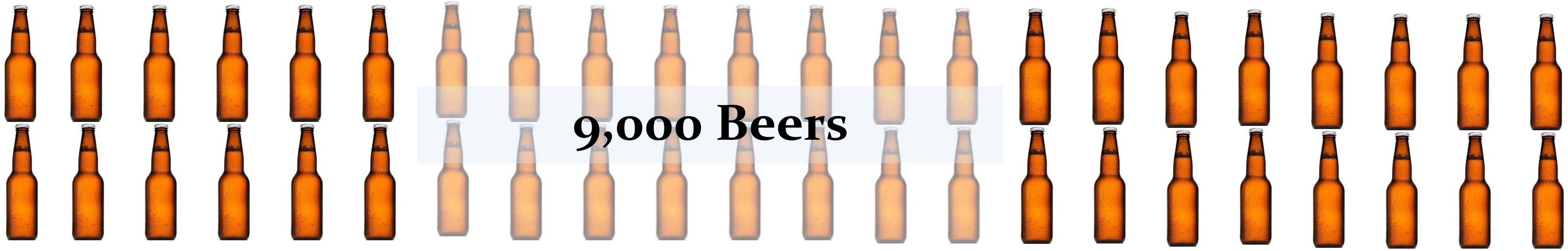


Phase 1: Data Overview

Brewery Table

2,000 Breweries

Beer Table



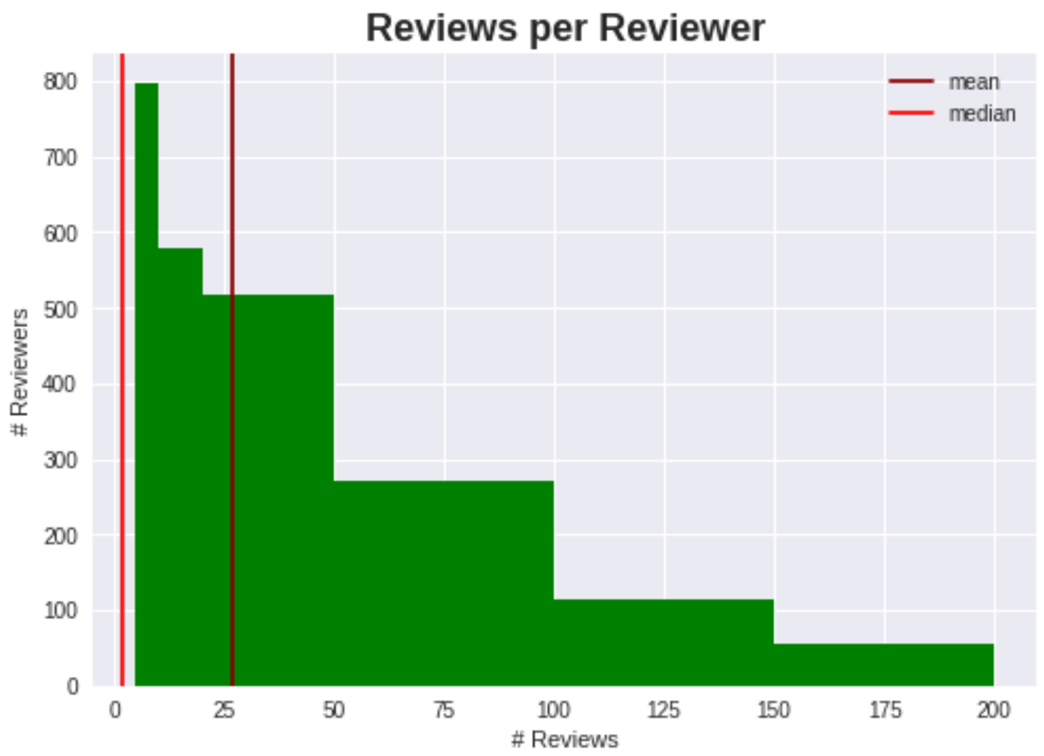
9,000 Beers

Ratings Table

211,000 Ratings
from 8,000 Reviewers

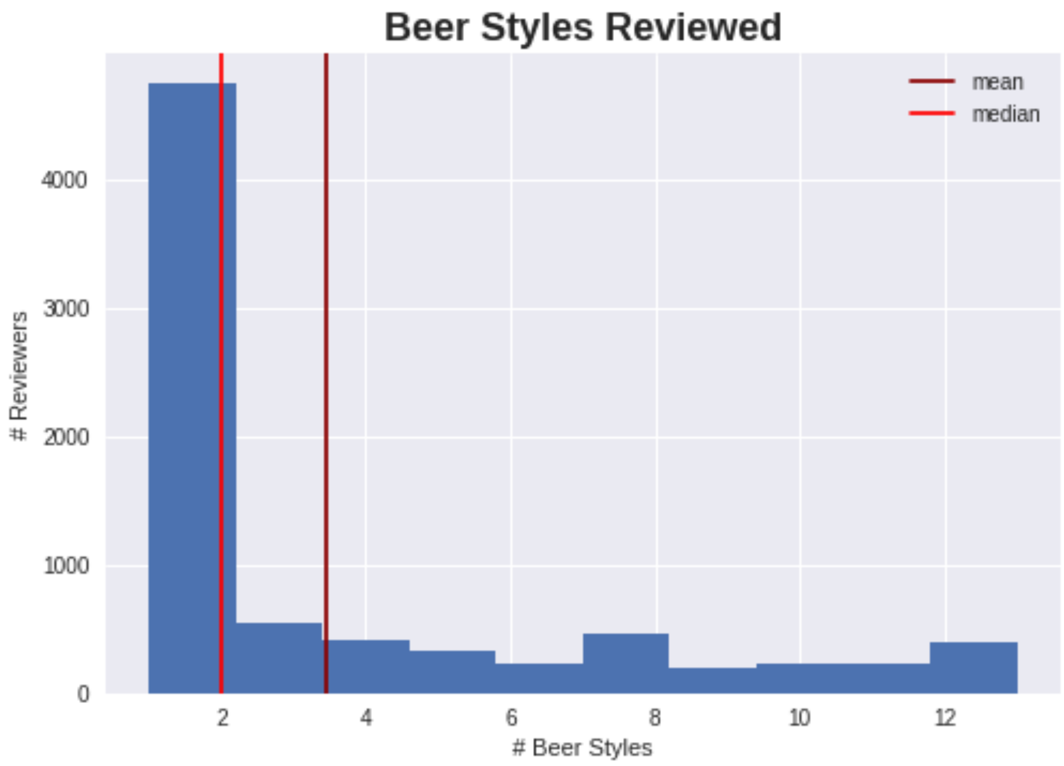
3.75, 4. , 3.75, 4.09, 4.21, 3.94, 3.75, 3.75, 3.99, 4. , 4.33, 4. , 4.5 , 4.25, 3.5 , 3.5 , 3.77, 4.25, 4. , 3.88, 4. , 4.25, 4. , 3.61, 3.79, 3.94, 4. , 4. , 4.75, 3.98, 3.66, 3.99, 3.4 , 4.25, 4.25, 3.75, 4. , 4. , 4.25, 3.65, 3.98, 4.08, 4. , 4.02, 4.25, 3.75,
3.71, 4.02, 3.71, 4. , 2.5 , 4. , 4.35, 4.13, 4.35, 3.83, 3.75, 3.75, 4.18, 4. , 3.75, 3.75, 4. , 3.92, 4. , 3.75, 3.75, 3.5 , 3.75, 3.25, 3.25, 4.25, 3.5 , 3.08, 4.09, 3.45, 3.53, 3.98, 3.82, 3.36, 4.18, 3.58, 3.75, 3.7 , 3.28, 3.42, 3.39, 3.77, 3.68, 2.54,
3.33, 4.21, 3.8 , 3.5 , 4. , 4.25, 3.39, 4. , 2.5 , 3.75, 3.63, 4. , 4.5 , 3.5 , 3.75, 4.25, 3.75, 4. , 4.25, 4.25, 4. , 4.25, 3.75, 3.79, 3.75, 4.18, 3.67, 3.51, 3.16, 3.38, 3.16, 3.58, 2.22, 3.39, 3.26, 4.25, 3. , 4. , 4.03, 4.34, 3.93, 4.5 , 3.75, 4. , 3.5 ,
3.76, 3.77, 4.25, 4. , 3.75, 3.5 , 4.38, 3.75, 4.42, 3.65, 3.67, 3.75, 4.25, 3.65, 4.25, 3.92, 3.5 , 3.79, 4. , 3.25, 3.5 , 3.75, 4. , 4.25, 4.5 , 4. , 4. , 4.21, 4.37, 3.75, 3.5 , 4. , 4.13, 4. , 4.17, 4.23, 4.15, 4.2 , 3.75, 4.11, 4. , 3.74, 3.95, 4.21, 3.48,
3.45, 4.2 , 4.21, 3.5 , 3.93, 4. , 4.06, 3.45, 4. , 3.81, 3.77, 3.52, 3.75, 3.78, 4.06, 4.25, 3.68, 3.62, 3.74, 4.1 [illegible], 3.77, 4.25, 4.11, 3.03, 3.53, 4.25, 4.08, 4.02, 4. , 4.2 , 4.31, 4.11, 3.56, 2.27, 3.02, 3. , 1.5 , 3. , 3.5 , 3. , 4. , 3.75, 3.75, 3.93,
4.19, 3.93, 4. , 4. , 4.08, 3.2 , 4.25, 3.77, 3.5 , 3.75, 3.83, 4.34, 4. , 3.89, 3.25, [illegible] 3.93, 2.75, 3.96, 4.16, 4.1 , 3.77, 3.5 , 3.5 , 3.25, 3.85, 4.08, 4. , 3.5 , 3.75, 3.2 , 2.5 , 4.09, 3.51, 3.58,
3.87, 3.75, 3.52, 3.5 , 3.87, 4.4 , 3.47, 3.55, 2.74, 3.63, 4.04, 3.94, 3.69, 4. , 3.98, 3.38, 3.62, 3.71, 3.72, 3.75, 3.86, 4.23, 3.5 [illegible] .69, 3.94, 3.69, 3.75, 4. , 4.5 , 3.75, 4.36, 3.6 , 4.5 , 3.88, 3.6 , 3.8 , 3.75, 3.25, 3.55, 4. , 4.02, 3.5 , 4.23, 4. ,
3.75, 3.75, 4.25, 3.25, 3.75, 3.77, 3.8 , 4.42, 3.36, 3.03, 4.01, 3.5 , [illegible] 4. , 4.08, 3.85, [illegible] 3.49, 4.03, [illegible] 4. , [illegible] 3.59, 4.5 , 3. , 3.63, 3.76, 3.84, 3.66, 3.69, 3.79, 3.45, 3.1 , 2.75, 3.5 , 3.75, 3.39, 4.24, 4.76, 3.69, 3. , 3.74,
3.75, 3.39, 3.75, 3.77, 4. , 4. , 4. , 3.68, 3.73, 4.19, 3.74, 3.75, 4.49 [illegible] 33, 3.85, 4.3 , 3.75, 4. , 4.6 , 3.37, 3.63, 3.93, 4.25, 3.8 , 4. , 3.75, 4. , 2.75, 3.75, 4.25, 3.34,
4.25, 4.13, 4.25, 4. , 3.5 , 3.9 , 3.54, 4.25, 3.3 , 4.1 , 4.2 , 3.75, 4.25, 4. , 3.75, 3.75, 4.25, 4. , 4.21, 4.02, 3.98, 4. , 3.99, 4. , 4. , 3.83, 3.94, 4. , 4. , 4.02, 3.83, 4. , 3.77, 3.89, 4.19, 4. , 4. , 4. , 3.75, 3.83, 3.83, 3.5 , 3.75, 4. , 4.25, 3.75, 3.98,
3.74, 3.83, 4.04, 3.74, 3.75, 3.75, 4.08, 4.06, 3.74, 3.98, 3.5 , 4.02, 4.25, 4. , 3.75, 3.6 , 4.12, 4.3 , 3.75, 3.3 , 3.9 , 3.68, 3.76, 3.79, 3.47, 4.19, 3.66, 3.94, 4.08, 4.09, 3.25, 1.51, 4.5 , 4.56, 4. , 4. , 4. , 2.77, 3.98, 3.23, 4.14, 3.74, 4.25,
3.76, 3.77, 4.25, 4. , 3.75, 3.5 , 4.38, 3.75, 4.42, 3.65, 3.67, 3.75, 4.25, 3.65, 4.25, 3.92, 3.5 , 3.79, 4. , 3.25, 3.5 , 3.75, 4. , 4.25, 4.5 , 4. , 4. , 4.21, 4.37, 3.75, 3.5 , 4. , 4.13, 4. , 4.17, 4.23, 4.15, 4.2 , 3.75, 4.11, 4. , 3.74, 3.95, 4.21, 3.48,

Phase 2: Who is the typical beer reviewer?

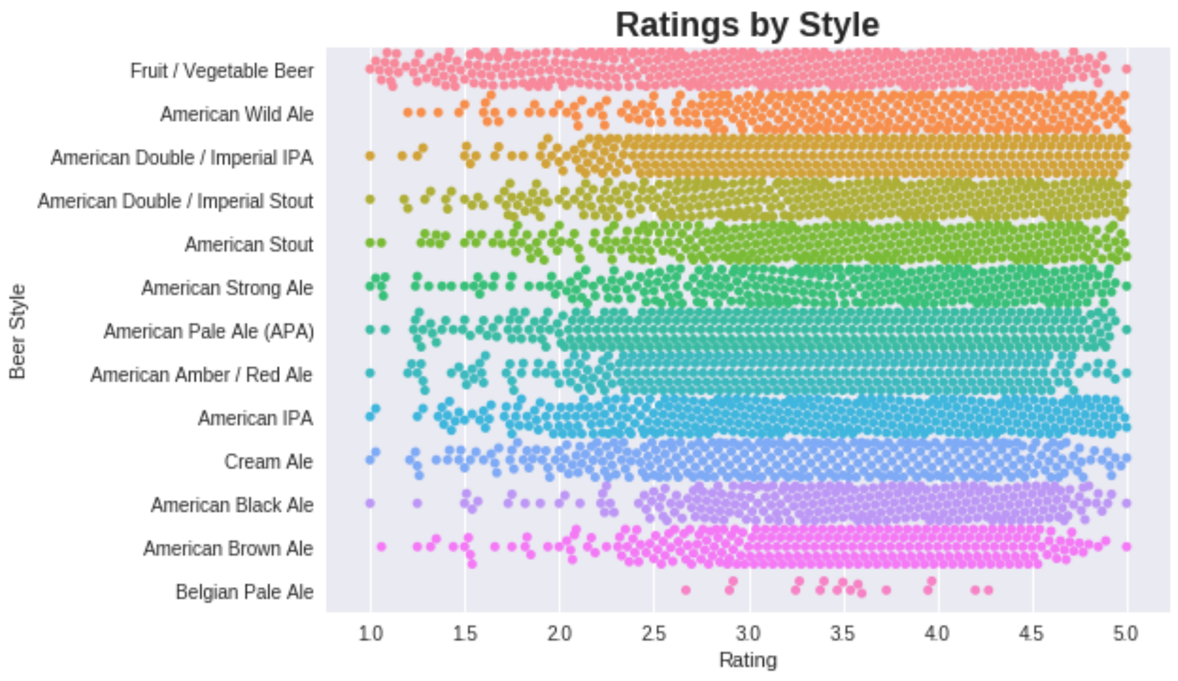


The above graph presents the distribution of the number of reviews per reviewer. The distribution is skewed, meaning the right tail is elongated, due to 350+ reviewers having reviewed more than 100 beers. The median reviews per reviewer is 2, however due to the skewed distribution the mean reviews per reviewer is 25. **For purposes of recommending beer I have assumed the typical beer drinker reviews 25 beers.** Note although the median is only 2 reviews 25 provides more observations for classification modeling. *Refer to Phase 4 for further discussion.*

The below graph presents number of beer styles reviewed per reviewer. The typical reviewer has rated 2-4 different beer styles, indicating beer drinkers are willing to try different styles. **For purposes of recommending beer I have assumed the typical beer drinker would accept recommendations on multiple styles of beer.**

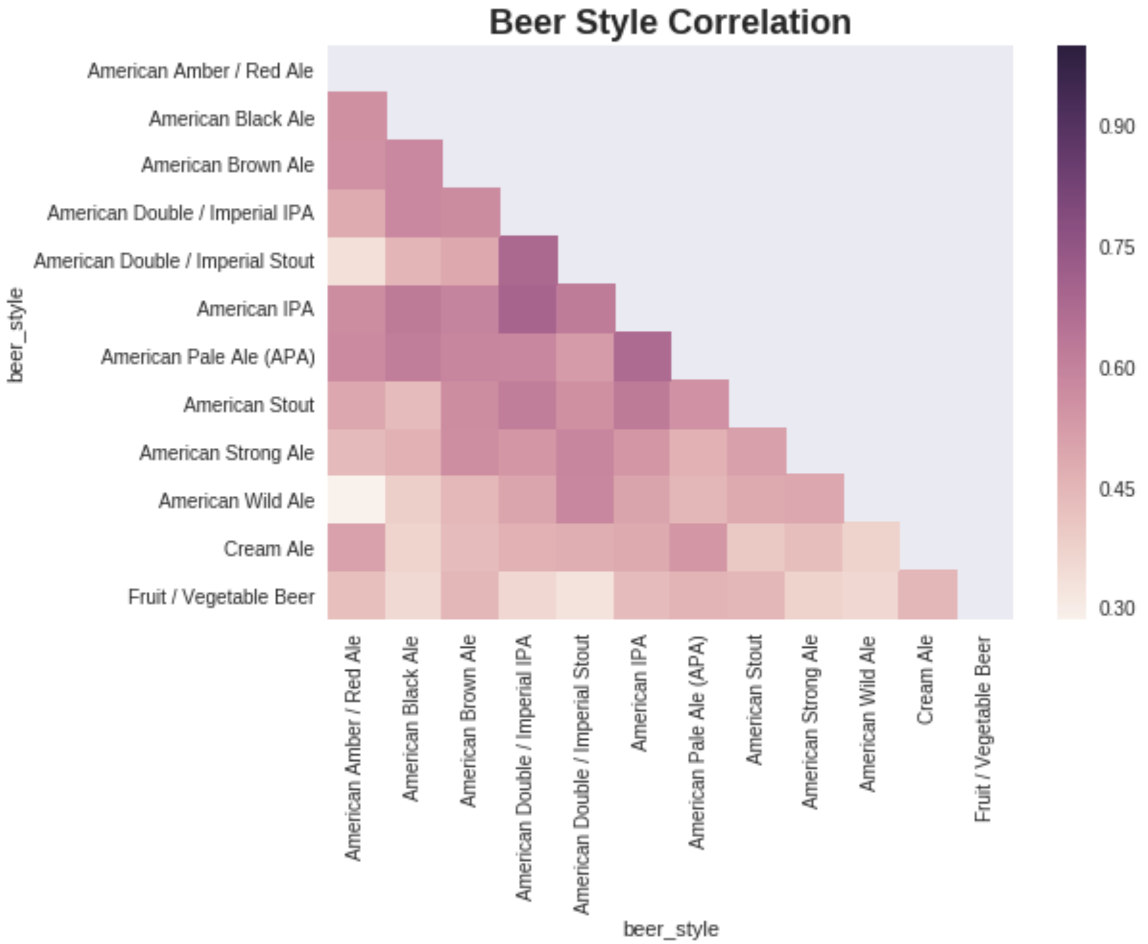


Phase 2: How do ratings vary among styles?



The above graph presents the distribution of ratings for each beer style. Nearly every style has the majority of its ratings between 3 and 5, indicating 1) people only rate beers they like or 2) people are more likely to rate beers they like a 4 or 5 and beers they do not like a 3. Further the overall rating mean is 3.7. **Given the scale for most beer styles is 3-5 versus 1-5, I have assumed 3.7 is the benchmark for liking a beer (>=3.7) versus not liking a beer (<3.7), unless otherwise noted.**

The below graph presents the correlation of ratings between each beer style. There is a strong correlation between Double IPAs and IPAs and Imperial Stouts and Double IPAs. As these are considered the more popular beers the finding was unsurprising.



Phase 3: Selecting a similarity measurement

Collaborative Filtering

The similarity measurements evaluated for the recommendation system were based on collaborative filtering methods, which are based on drinker rating patterns.

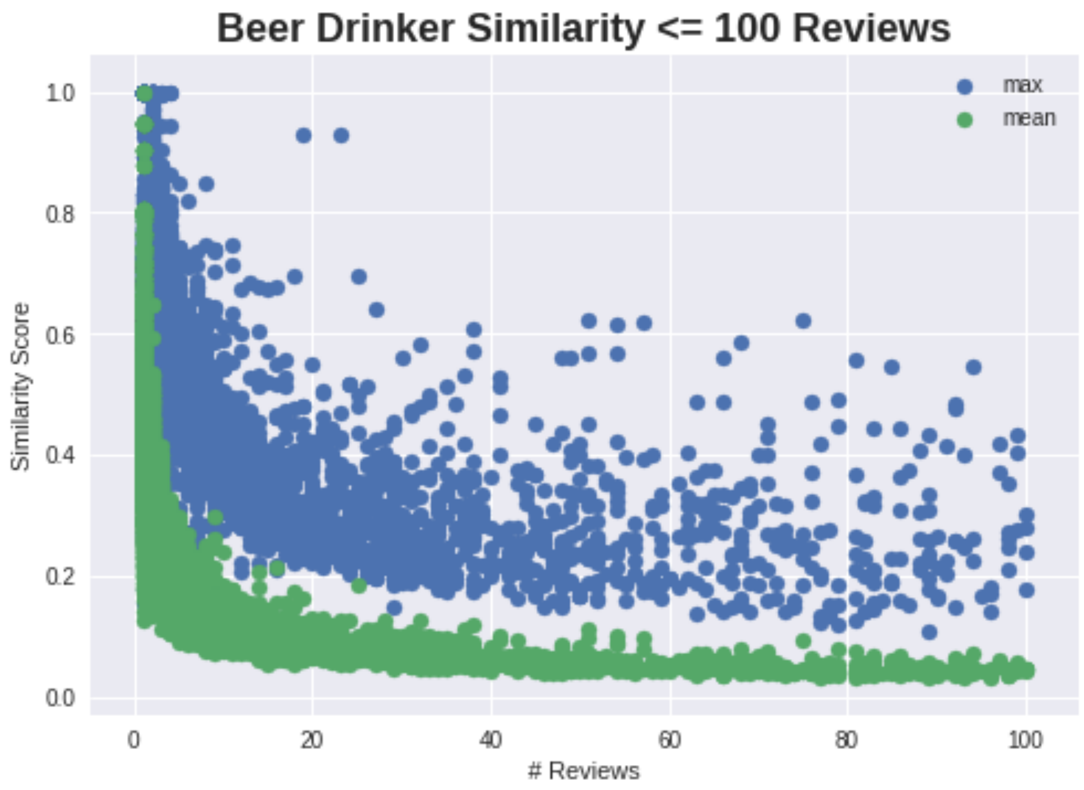
- **Drinker-based collaborative filtering** – Cosine similarity was used to determine the "closeness" of two drinkers based on 1) the beers they have rated and 2) the ratings of such beers.
  - Cosine similarity is bound by 0 and 1, meaning 1 indicates users are exactly alike and 0 means users have no similarity.
  - The method answers the question "What drinkers are similar to me and what beers do they like?"
- **Item-based collaborative filtering** – Euclidean distance was used to determine how different two beers are to each other based on the drinker ratings.
  - The method answer the question "People that like beer A also like beer B?"

Selecting A Method

Drinker-based collaborative filtering was chosen as the similarity measurement for the beer recommendation system. Due to ratings being condensed between 3 and 5 the distance measure used in item-based filtering did not create enough dispersion between beers. For example Imperial Stouts would return only Cream Ales. Further given that users typical rate 2-4 beer styles I did not want to limit the recommendation by style.

Drinker-based Collaborative Filter Results

The below graph presents the max and mean similarity score for each user with less than 100 reviews. As the number of reviews approaches 20 the similarity scores level out, which may be indicative of needing more drinkers with 20 or more ratings. Although the scores between individual users appear low by applying aggregation and and weighting we can generate recommendations.



Phase 3: Generate recommendations based on user similarities

What beer should I have?

Active Drinker

Use similarity relative to active drinker to adjust ratings

Drinker	Similarity (A)	Beer	Rating (B)	Adj. Rating (AxB)
JSmoltz	0.5	Pliny	4.0	2.0
GMaddux	0.2	Pliny	3.5	0.7
SAvery	0.3	2x4	5.0	1.5
MWholers	0.2	2x4	5.0	1.0

1) Aggregate the adjusted ratings and similarity scores

2) Standardize the adjusted rating back to the 1-5 scale

Similarity (A)	Beer	Adj. Rating (C)	Total Rating (A/C)
0.7	Pliny	2.7	3.9
0.5	2x4	2.5	5.0

Aggregating takes into account similarity across all drinkers

Recommend highest total rating

Enjoy a Beer

2x4 is better than Pliny

Active Drinker

Phase 4: Like or not like

Attribute Selection

The similarity measurement provides a recommendation based on the active drinker's relation to other drinkers, but what specific attributes related to a specific beer can be used to predict whether the active beer will actually like the beers recommended? The following attributes were collected for each beer:

1) **ABV** – Represents percentage of alcohol content in each beer. Nearly every beer contained this field, and thus it was ***included***.
2) **Beer Description** – Represents the commercial description of the beer. Seasonal beers, which were typically rated highly, did not have this field. Further ~50% of the 8K beers contained the field thus this attribute was excluded.
3) **Beer Name** – Represents the commercial name for the beer. Every beer contained this field, and thus it was ***included.***
4) **Brewery name and location** – These attributes were intentionally excluded to avoid brand preferences and the "bubble" effect, i.e. people only trying beers near them.

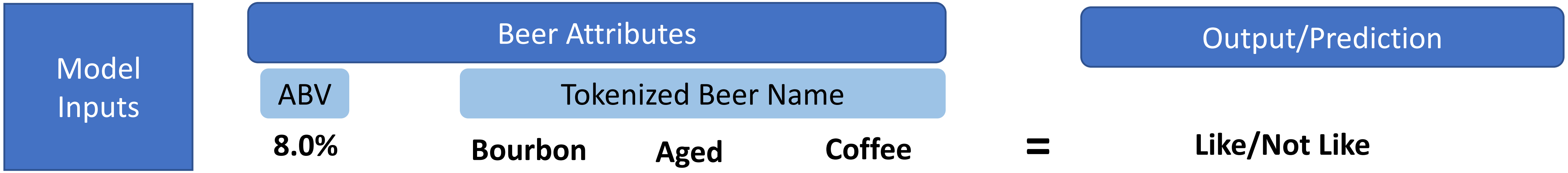
Classification Modeling

A Like was considered any rating greater than or equal to 3.7 and and Not Like was defined as any rating <3.7, based on the rating meaning of the total population.

- Drinkers with less than 25 reviews were excluded from the classification modeling, as there were not enough ratings to train and evaluate the model.
- The focus was on creating a predictive model that generalized well to new beers.
- Beer Name was tokenized to determine which words in a beer name indicate whether the active drinker will like the beer or not.

Two classification models were evaluated for prediction:

1) **Random Forest** – Represents the mean of several decision trees, which classify Likes/Not Likes based on splits in the attributes.
2) **Logistic Regression** – Measures the relationship between Likes/Note Likes and the beer attributes (ABV and beer name). Derives a probability to determine the Like/Not Like.



Phase 4: Classification Evaluation

Model Evaluation

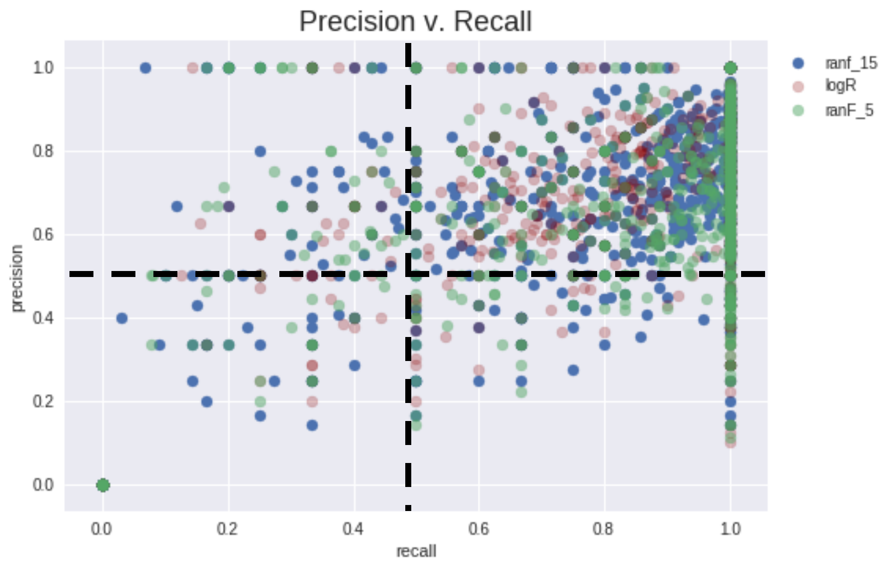
The adjacent charts measure 1) the precision versus the recall for each model and 2) the recall for scores based on the number reviewer reviews.

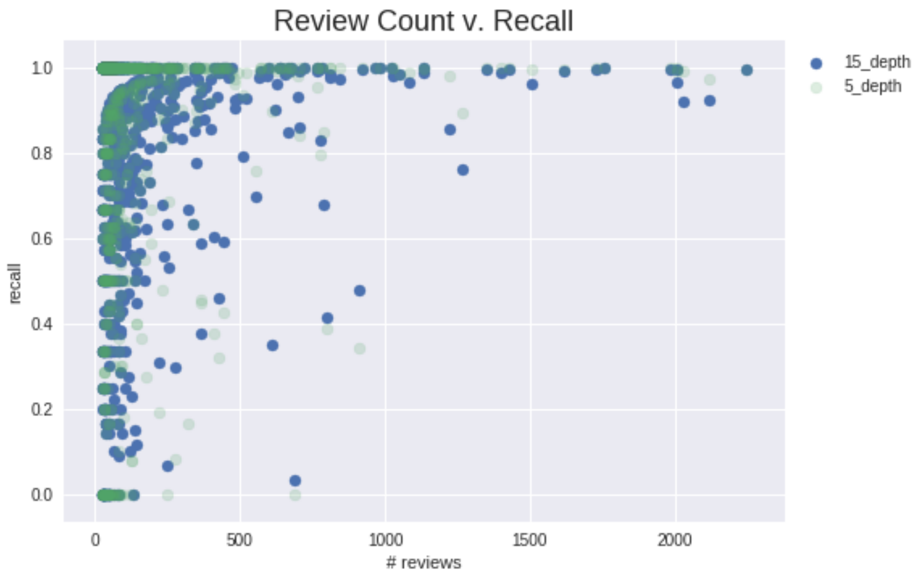
The classification models were evaluated using the following metrics:

1) **Precision** measures the percentage of predicted beer Likes that were predicted correctly.
   i. Predicted True Likes / Total Predicted Likes
   ii. Precision is penalized for misclassifying Not Likes as Likes, therefore the consequence is drinking a beer that you would not enjoy.

2) **Recall** measures the percentage of actual beer Likes that were predicted as Likes.
   i. Predicted True Likes / Total Actual Likes
   ii. Recall is penalized for misclassifying Likes as Not Likes, therefore the consequence is missing out on beers the active drinker may actually like.

Since the penalty for low precision is drinking a bad beer, which is inconsequential, the model was built to maximize recall. By maximizing recall the active drinker is less likely to miss out on a delicious beer.





Phase 5: Optimized Beer Recommendations