BeER ratings ANALYSIS

Forrest Fallon

Forrest Fallon

IST 707 – Final Project

## Introduction

According to researchers at the History Channel (Andrews, 2014), the first brewed beer is believed to have been created roughly 12,000 years ago. Since then, little to no progress has been made on a grading scale that is equal parts universal and transparent. In this project, we will explore one dataset consisting of 3000+ unique beers, 900+ breweries, and a litany of variables assigned to each entry. This dataset contains enough variables to do deep-dives into questions such as how mouthfeel affects review score, or if a certain brewer creates beer consistently above a certain ABV threshold; however, to keep this project at an appropriate length, we will focus mainly on the beer-styles at hand and their respective review scores and ABV measurement.

Upon completion of the planned tests, we should have a better understanding of which styles of beer are often rated the highest and lowest, along with how well machine learning programs can handle a dataset such as this. With as many variables and types of beer style within this set, there are a few preliminary steps to take before we begin our analysis.

## Data Prep

To avoid wasting time sifting through error messages R-studio would inevitably throw at us, the following actions must be performed upon importing our dataset:

* Check for NAs



We are in luck, this dataset was well managed and contains no missing values.

* Rename columns



* Assign our “Style” column to factor.

Upon completing the above, we must now think critically about the integrity of the data we are handling. A quick survey of the “Number of reviews” column would show that there are many entries with lower than 100 total reviews. This does not allow for much confidence to be placed in the overall review scores; to remedy this, we will create a new dataset with only beers that have been reviewed more than 100 times.



As mentioned in the introduction, we will need to slim down our dataset even further to remove variables that are not of interest to us at this time: flavor, mouthfeel, aroma, etc. To do this, we will select the specific columns that we need:



Another column worth adjusting at this point would be “Styles,” seeing as though each style has its own sub-style; i.e. Lager – American, Lager – European. Because of prior experience, we can assume that these differences in naming conventions may cause some problems down the line with algorithms such as RandomForest. To get ahead of these potential problems meanwhile allowing for more accurate analysis of each style within the set, we will eliminate the sub-styles. A few examples of this change shown here:

Text

Description automatically generated

The above block of code will change all IPA sub-styles to just “IPA,” which will allow all of them to be grouped together properly for tests and descriptive statistics. Once this method is applied to all styles available within the dataset, we can move on to our final necessary preparations.

Finally, we will create factor-labels for each review overall and ABV. This will allow for our machine learning programs to better predict values when they are not burdened with decimal-specific values. The following lines of code will create the breaks necessary, and set them to factors:

Text

Description automatically generated

Our prep is now complete! We can safely move on with analysis and expect minimal push-back from R-Studio. We will first answer a few questions about the dataset with descriptive statistics.

## Descriptive Statistics

First, we will return a summary of the overall review scores and ABV values of the dataset:





As seen above, with a minimum of 1.98 and maximum of 4.63, we can rest assured that no beer within our dataset has been deemed the “worst beer in existence,” or the “best beer ever created.” In fact, it seems as though our dataset is somewhat generous in its grading scale seeing as though a “true middle” of this scale would be 2.5 out of 5, and the mean score being 3.8. This begs the question, **do people feel more generous about the review score of an alcoholic beverage because having an alcoholic beverage at all is considered a “good time,” meanwhile inhibiting judgement?**

Regardless of the answer to the above, we can move on and find what this dataset’s best and worst rated beers are. The following code will provide those results for us:





Max score  


Min Score  


Unfortunately, R-Studio does not handle foreign text-characters very well. Upon searching the original database for the highest scoring beer, we can see that Blåbær Lambik, from Brasserie Cantillon holds a stellar **rating of 4.63** with 156 total reviews. High praise! On the other side of the beer-aisle, we have an entry that is much more recognizable: Corona Light, with over 340 reviews, boasts an average **review of 1.96**.

With the Sour (lambic) style taking the top spot the power ratings, we should next find out what styles as a whole rank the highest. Here is how we can accomplish that within R:



Which provides the following:

Text

Description automatically generated

Wild Ale has taken the lead! It is worth noting here that the differences between each style’s average review scores are very small; meanwhile, the Wild Ale scores 0.11 higher than the next closest option.

The power of descriptive statistics should not be underestimated, many of the questions one may have about a dataset can be answered with just a few commands from R-Studio. The information provided above about beer styles and rankings would be great info to have when choosing what beer to be on the lookout for; one could also use this info to correct those with false information about what the “world’s favorite beer” is.

Next in our analysis, we will use some more powerful offerings within R-Studio that can provide us with insights not able to be spotted with the naked eye.

## Apriori Rules

The Apriori rules analysis reads a dataset’s variables, and determines which ones are most closely related. Upon running the test, we will be presented with a list of the most commonly occurring rules that will tell us more information about our dataset:

Text

Description automatically generated

It is worth noting that we are inspecting rules 14 through 24, due to the fact that the first 13 rules provide no left-hand-side argument. This is an issue specific to this dataset, and those blank rules are essentially telling us that all beers have a rank between 1.9-4.6, and have an ABV of 0-28. Below are the rules that tell us something worthwhile:

Text

Description automatically generated

A few key findings from this list are as follows:

* Pumpkin Beer: Grades = B
  + Pumpkin Beer is closely related with being given a review score of B.
* Style = Old Ale: ABV = [7.5, 28]
  + Old Ales are typically found to have ABV’s within 7.5 and the max values of our dataset.
* Brewery = Anheuser – Busch: Review\_Overall = [1.96, 3.74]
  + Anheuser Busch beers are typically rated on the lower side of the spectrum.

In specific, the final highlighted rule we listed above had a confidence factor of 1.0, with a count of 23. This means that the program is 100% confident in the fact that Anheuser Busch beers do not exceed a rating of 3.74, and there were 23 instances of that happening within our data. American Domestic beers are not holding up well in our rankings!

The Apriori Rules analysis can help us find associations within our data that would likely take many more hours of analysis to find ourselves; it is a great tool to point us in the right direction of further analysis. Our final tool will show us if any of these factors previously discussed can be used for predictive measures.

## RandomForest

The RandomForest algorithm is a complex system of creating and merging many different decision trees with random splits, as opposed to the more traditional approach to decision trees that relies on the next “most important node” to create a split. What we aim to do with RandomForest is simple, we want the algorithm to take the data we have and be able to predict a beer’s overall review score or ABV value based on the other variables available. The classifications we created for each review score and ABV earlier will come in handy here.

Our first pass of this algorithm will focus on the overall review score value. In order to correctly run this test, we will first create a test and train dataset via the following code:

Text

Description automatically generated

Once this is done, we can run the RandomForest command, followed by a confusion matrix that combines the prediction command and our test results to show us the accuracy of the test:

Text

Description automatically generated

Text, calendar

Description automatically generated

As seen above, the test resulted in an accuracy of 69%; while this is by no means an amazing accuracy, it is worthy of at least paying some attention to. The program was able to guess a beer’s overall rating based off its name, style, brewery, ABV, and number of reviews. In fact, we can command R to show us a variable importance plot to see which variables were the most influential to the algorithms ability to guess the review scores:

Table

Description automatically generated

What this graph indicates is how much the accuracy of the test goes down when each of the variables are removed. In specific, if number\_of\_reviews is removed from the test, the accuracy is affected the most. We can infer that the number of reviews has the largest impact on review score guess-accuracy. Reminder: this test does not tell us that as number of reviews goes up, so does the review score itself. Rather, this test tells us that it can more accurately guess the review score when there are more score entries.

Next, we will do the same testing on the ABV grades; let’s see if the program performs better or worse:

Text

Description automatically generated

Table

Description automatically generated

Based on the accuracy rating, the RandomForest algorithm had a harder time guessing at ABV values than review scores, with an accuracy of 63%. This makes sense as an ABV value can vary wildly from beer to beer, and a single brewery can have many different ABV values within its catalog. The Style variable having the most sway on the test also makes sense here; style names are often considered “rule sets” for beer flavors and attributes. It would make sense that the program found the most safety in guessing ABV values from a beer’s style.

## Conclusion

Although this subject matter is rife with subjectivity, we were able to successfully answer a few burning questions surrounding the beer world. This dataset provided the info we needed, and the tools developed within class allowed for thorough and accurate analysis. The framework of this project is often the best way to start any analysis of a broad subject area: clean the data, start with descriptive statistics, use rule-analysis to find consistencies between variables, and test the predictability of the data with advanced machine learning algorithms. When performed in this order, datasets that seem daunting at first can easily be dissected and inferred upon.

What was most interesting about this data in specific was the confirmation of the fact that many American Domestic beers are looked down upon when it comes to quality. Many of the highest scoring beers and styles were specialty wild ales and foreign sours, not your typical American beer drinker’s first choice. With that being said, these findings may point to a larger problem within the dataset.

Without knowing for sure where these ratings came from, it is hard to say with confidence that a proper study-group was chosen for the data gathering. We have no way of knowing the country of origin of each of these entries, which would offer a massive insight into the reasoning behind each review score. American domestic brands may have performed better if the study population had a stronger American representation; however, it may already be a purely American study group!

Finally, the method bias that surrounds any study involving alcoholic beverages is essentially ignored here. Without knowing whether these beers were merely tasted rather than consumed, there is no telling whether the consumption of alcohol impacted test scores at all. Until a world-wide survey can be completed with proper data-gathering protocols enacted, the results of our testing should be used with caution.

# References

Andrews, E. (2014, January 8). *Who Invented Beer?* Retrieved from history.com: https://www.history.com/news/who-invented-beer