**Beer Recommendation Shiny App**

**Nilofer Rajpurkar, Jacob Collins, Mitch Lamm**

collin85@purdue.edu; lammm@purdue.edu; nrajpurk@purdue.edu;

**Abstract**

Our Shiny app takes a list of favored beers by the end-user, and recommends beers within the user’s comfort zone, as well as recommending beers with a style differing from what the user inputted. The craft beer industry was estimated to be worth $89.25 Billion in 2019 and is projected to be at 190B by 2027. Fans of craft beer love to explore and try new beers, and a beer recommendation app is a great way to quickly share possible brews that the user may also enjoy. Our app takes into consideration the beer name, style, various user ratings, and the profile who left a review, and uses this information to group the end-user into a group (cluster) of beers they make like. The app takes the profile names who also rated the same beer input by the end-user and suggests similar beers that those users also liked. It then will also recommend beers of different styles, but also rated highly by those same users.

**Business Problem**

Beer fans come in many different stages of fandom, from the novice all the way up to the most experience connoisseur of beer. The novice, on one hand, will be new to the many options of craft beer and could become overwhelmed. The experienced beer drinker will have had many different beers in their time and could be looking for new options other than their old go-to’s. To address both issues, one would normally perform extensive research online or consult friends and family about different options. To streamline this time intensive research process, we will create a tool that the user can quickly use to find new beers to try based on other beers that they know they have enjoyed in the past.

**Analytics Problem**

Someone for shopping for a craft beer needs an easy way to create a list of beers that they like and get a list of suggested beers based on their initial list. There are multiple beer review websites that currently exist and utilizing that review data is one way to find other beers to recommend the viewer. The problem is that some of these datasets are over a million review, and users need a quick way to see a “top 5” list of beers to try based on what they already know they like.

**Data**

To make our app successful, we need a dataset that has beer name, brewery name, and various review data from users. Having the username of the reviewer is also helpful, so we can use that to determine if the app user has a similar flavor palate as other reviewers.

We found a dataset of Kaggle, which has 1.5 million beer reviews, which were published by BeerAdvocate, which is one of the most popular beer review sites. Our dataset was almost all populated, except for the beer ABV column. To fill in this missing data, we used a prediction model using caret to populate the missing data, by predicting based on the beer style, as well as users reviews of the taste, aroma, appearance, and palate. We added in these review scores because some users may rate a taste and/or palate lower if the beer has a higher ABV. Also, the aroma of a beer can be less pleasing if a beer has a higher ABV. We scaled these ratings using caret, and we also one-hot encoded the beer styles to get an accurate prediction of the beers who were missing their ABV data. Once we populated this, our dataset was ready to work with

**Methodology Selection**

We used predictive analytics to fill our missing ABV data with caret. We felt that this was better than using the standard average ABV of each beer style, because using an average just reinforces the average even further, for each beer style. Considering user review ratings into our prediction, this helped to prevent this issue. Once the data was cleaned and prepped, we then used descriptive analytics to filter and manipulate the data to show the end-user a recommended beer list. We used clustering to group different reviewers into sub-groups and based on the app-users inputted beer list, they were placed into one of the existing clusters. The descriptive analytics continue further as the user is then placed into an existing cluster to recommend future beer choices.

**Model Building**

Placing the user in a subgroup will be critical as we want to group people with others of similar taste. To find an optimal number of clusters within the data, we performed a k-means analysis and graphed it to an elbow chart to find that the optimal number of clusters would be 5 within the data. However, we wanted to enable the user to define how closely they could associate with one group vs another. So, we also applied k-means to the data to place them within k= 10 & k=16 different clusters to define how picky of a subgroup they wanted to be placed in.

**Functionality**

Functionally, the app utilizes background data that has already been calculated and updated by the team. Rather than trying to run the full amount of code available (1.5million lines) we wanted to make the user experience smooth and not time intensive. The simple inputs from the user will quickly sort through the massive amount of data and enable a quick selection of beers for the user to pick from. There for four separate sorts that is done based on the clustering of the user’s input, historical clustering of beers, and finally how risky of a beer it is for the palette of the user. The lists that are generated are randomized every time and are sent to a text output section.

**GUI Design and Functionality**

Functionally, there three different sections of the GUI. First, in the input section, the user is allowed to select three minimum to 5 max different beers. In the informational output section, the user will be recommended beers that they will love based on the type of beer they chose, risky beers that they may have normally not picked but would be a good fit, a beer that will be high in ABV and will make intoxication easier, and finally a beer that would be good for preventing over intoxication (low ABV). The approach to make the GUI only text-based is to simplify the visual output for the user, rather than overwhelming them with visuals that they will have to decipher.

**Conclusions**

Beer enthusiasts new and old are always on the look out for new beer to try, but there are many options available to them. We created a tool that takes the many options of beer and analyzes millions of reviews to find the best beer selections for the user. The users will then be presented with several personally curated options. Finally, the user will be informed of new beers that they can sample and find their future favorite beer.

**References (if any)**

Code reference is given to Github user lauraarossi <https://github.com/lauraarossi>