

Data Analysis with R Term Project

Beer Reviews

Jugal Chauhan

Student ID: 0755856

Problem Statement

Customers' preference for a type of beer is usually based on several factors ranging from the appearance to even the smell of the drink. It is often difficult to choose a drink from variety of available options. A metric based on the reviews of the products can help in selecting a drink that is related to the customer's choice. The primary objective of this project is to construct a resourceful and accessible beer review application that ranks the beer on a scale of 0 to 5 based on five constructs, namely appearance, palette, aroma, taste, and the overall rating of the beer. The goal of this project is to present a reactive visualization application that communicates user ratings of the product. The purpose is to present the capabilities of visualization techniques through R shiny application, and in no way intends to promote the consumption of alcohol.

Background

There are researches regarding different aspects of beer that help customers to identify the right choice of drink and assist brewing companies to manage the expectation of customers. A brief focus is put on the nutritional aspects of beer that appear to be beneficial in combating coronary heart disease (Bamforth, 2002). Also considering the taste of beer, the chemical compounds used have a substantial effect on the quality of beer and the flavour control should be measured on the thresholds of chemical ingredients (Harrison, 1970). According to (Lukinac et al., 2019) beers are differentiated primarily based on visual appearance and fermentation process. It is understood that customer choice is driven by psychological, and socio-cultural factors, and with the sensory aspects of beer including the external sensory attributes like packaging and presentation, however, more research is required for the relative contribution of each factor on the customer choice. (Betancur et al., 2020). It is also found that online ratings and the strength of reviews play an important role in understanding product growth in the marketplace (Clemons et al., 2006). With this, brewing companies have come to realize the factors that influence the choice of beer and aim to expand their industries to meet consumer requirements (Habschied et al., 2020). This lays the groundwork for the requirement of a platform where beer reviews based on multiple aspects of the drink are easily accessible for both consumers and enterprises.

Instructions

The instructions to access and use the application are quite simple. The application can be accessed through this [URL](#) that is hosted on the shinyapps domain. The ratings are based on four individual factors: appearance, palette, aroma, and taste of the beer. Additionally, the overall rating of the beer is also included. Bar plot is used as visualization for average ratings of each component. The reactive bar plot and search bar in the application is developed using Shiny package available in R. The application can be used to easily access the average user rating of beers on the bar chart with a simple search and select option available on the side panel. The center panel is utilized to display the bar plot of the selected beer. The search bar can be used to select the beer of choice, a drop-down menu is also provided to browse through the available beer choices. Figure 1 showcases one of the instances of the application.

Beer Reviews

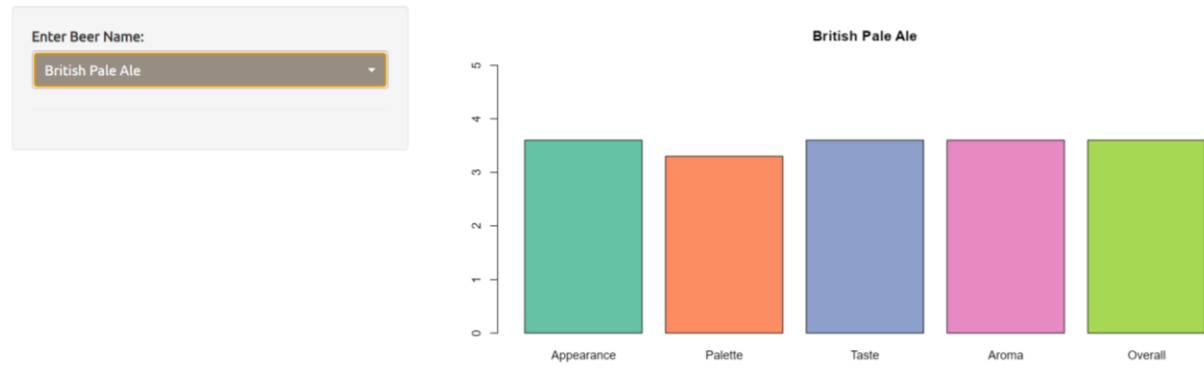


Figure 1: Beer Review Application

Dataset Description

The Beer Review project visualizes the average user ratings of 18,339 unique beers from 22,801 unique users obtained from [Beer Advocate](#) data available on [kaggle](#). The dataset used for this application originally contains 13 variables and 528870 observations, and included NA values primarily in the (“beer_ABV”) variable. The variables their description is provided in table 1. The data cleaning process involved the following steps:

1. Removal of unnecessary “review_time” variable.
2. Generating new variables with average rating score based on (appearance, palette, aroma, taste, and overall) from multiple reviews of the same beer.
3. Creating a new variable “review_count” that stores the total number of reviews recorded for each beer.
4. Dropping the original review_appearance, review_aroma, review_palette, review_taste, and review_overall columns and only keeping the columns with average rating score for each component.
5. Ignoring the 4471 NA values from beer_ABV variable as it is not needed on the bar plot.
6. Transforming data into a long format using the gather() function.

Variable Name	Variable Description
beer_ABV	The alcohol weight by volume of the beer in per cent.
beer_beerId	A unique id assigned to each beer name.
beer_brewerId	A unique id assigned to each brewing company
beer_name	Name of the beer
beer_style	Style of the beer
review_appearance	Rating on the appearance of beer on the scale of 0 to 5
review_palette	Rating on the palette of beer on the scale of 0 to 5
review_taste	Rating on the taste of beer on the scale of 0 to 5
review_aroma	Rating on the aroma of beer on the scale of 0 to 5
review_overall	Overall of beer on the scale of 0 to 5
review_profilename	Profilename of the reviewer
review_text	Full review of the beer
review_time	Time of the review

Table 1: Dataset with original variables

The long format of the data is ideal to generate bar plots of average ratings for each beer, the dataset after the data cleaning process contains the type of review (whether appearance, aroma, palette, taste, or overall) in the “review” column and the corresponding rating in the “value” column. The long format of the data is used in ui and server functions of the shiny package to generate bar plots based on input value.

Limitations

The primary limitation of the Beer Reviews application is that the number of reviews for each beer differ, hence the average rating score can be proved to be misleading sometimes (a beer can have only one reviewer that gives 5 ratings against a beer that has more than 100 reviews has lower average rating) this can be mitigated using the “review_count” variable that shows the number of times a beer has been reviewed, providing a clearer understanding of the ratings.

Future Implementations

The current version of the Beer Reviews application is a very basic one, and there are few things that can include more information related to the user review of a selected beer. It is possible to display information such as Alcohol weight by volume of the selected beer beside the bar plot. Furthermore, full user review of the beer can be displayed below the bar plot along with the user

profile name that can be helpful in providing detailed experience of the beer. R shiny widgets can be explored more thoroughly to apply sort and filter options on the list of available beer names which allows the user to find specific type of beer based on the style of the beer or the brewing company, or even based on the percentage of alcohol weight by volume. The visualization part can be improved by implementing a side-by-side comparison of two different beer ratings. More advanced visualization can be applied such as Radar chart that showcases each aspect of the review.

References

1. Charles W. Bamforth, Nutritional aspects of beer—a review, *Nutrition Research*, Volume 22, Issues 1–2, 2002, Pages 227-237, ISSN 0271-5317
2. Harrison, G.A.F. (1970), THE FLAVOUR OF BEER—A REVIEW. *Journal of the Institute of Brewing*, 76: 486-495.
3. Maria Isabel Betancur, Kosuke Motoki, Charles Spence, Carlos Velasco, Factors influencing the choice of beer: A review, *Food Research International*, Volume 137, 2020, 109367, ISSN 0963-9969, <https://doi.org/10.1016/j.foodres.2020.109367>.
4. Habschied, K., Živković, A., Krstanović, V., & Mastanjević, K. (2020). Functional Beer—A Review on Possibilities. *Beverages*, 6(3), 51. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/beverages6030051>
5. Lukinac, J., Mastanjević, K., Mastanjević, K., Nakov, G., & Jukić, M. (2019). Computer Vision Method in Beer Quality Evaluation—A Review. *Beverages*, 5(2), 38. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/beverages5020038>
6. Eric K. Clemons, Guodong Gordon Gao & Lorin M. Hitt (2006) When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry, *Journal of Management Information Systems*, 23:2, 149-171, DOI: [10.2753/MIS0742-1222230207](https://doi.org/10.2753/MIS0742-1222230207)