The purpose of the project is to conduct the analysis of key attributes of beers and their respective breweries, how they relate under the Budweiser umbrella. Whether we can leverage these relationships is still to be determined post EDA. Giving due course to this process, we have unveiled that ABV and IBU attributes have some correlation and this relationship can be deployed to predict whether a beer should be categorized as IPA or non-IPA for future beers, given IPAs are the most common in the USA. Statistical technics have empowered us to investigate the data in multiple lenses. For instance, we have realized through our analysis that KNN classification works well when it comes to predict the ALE vs IPA beers using their ABV and IBU as good predictors.

But we did not limit us in the one technique, rather we ventured into further analysis of ABV and IBU data to evaluate the normality and variance of each of these two dataset, and came to conclusion that we can use these data for multiple regression as the datasets do not corroborate with the assumptions required to perform the regression. Then, again we indulge ourselves with another technic, that is Naïve Bayes, recently learnt in the class, but resolved to evaluate KNN-Means; producing increased precision at the expense of recall.

We expanded our analysis beyond ABV and IBU, and strived to find how the container size, represented by ounce, is being favored by the different states, and breweries in general. Interestingly, we uncovered that 12-ounce container is predominant across types of beers, and most of the states. Off course, that puts the curiosity in our mind, and nudges us to find some answers to it. The answer we found is that 12-ounce constitutes one serving and mostly a standardized in the country. For example, if one goes to a pub and ask for a beer of their choice, he or she will be served with 12-ounce beer. Therefore, the breweries produce more of 12-ounce for the commercial purpose.

Furthermore, another attempt we made to pinpoint which city consumes the highest ABV beer by the state. Finding was not that intricate, but putting that city in the map with aesthetic become challenging as it requires to pull the coordinate for each city and linking it together.

Structure of the project work and related submission –

1. Data files
   1. Readme file stating the high level project objective and quick synopsis of the work we performed
   2. Beers.csv and Breweries.csv
   3. Rmarkdown (.Rmd) file containing all codes, graphs and analysis
   4. .html file created from .Rmd for the presentation for the work we have done for this project