Beer and Breweries – Case Study 1

Team - Hops and Dreams

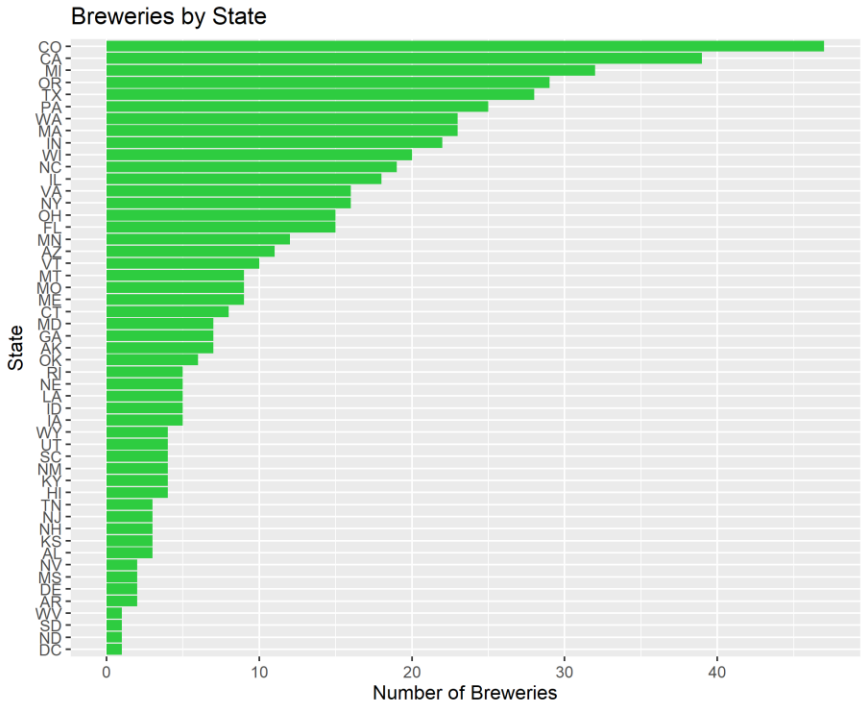
athobani@mail.smu.edu
spurvis@mail.smu.edu
senthil@mail.smu.edu

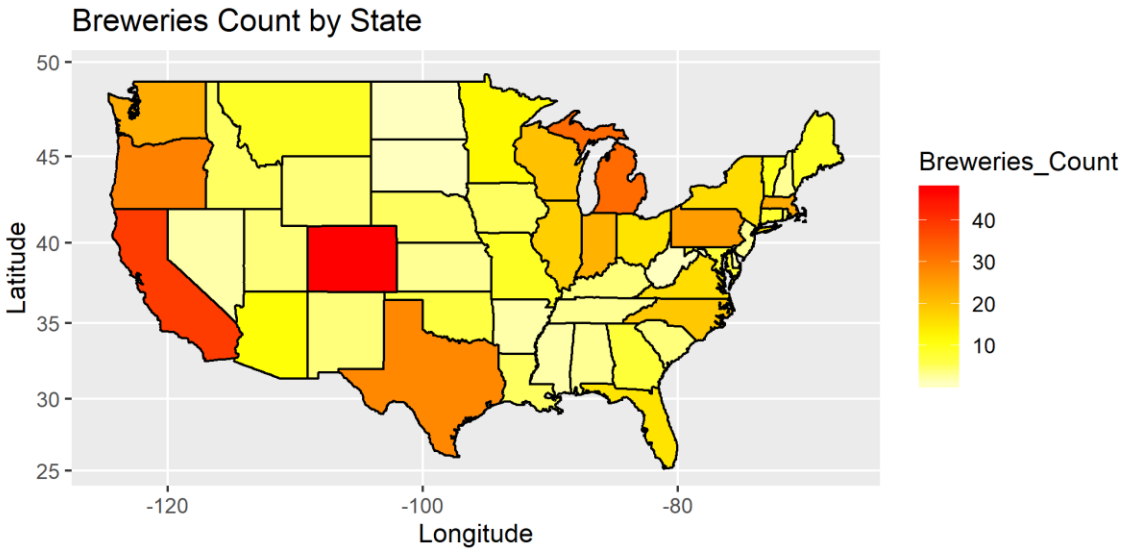
Introduction

In order to assist Budweiser®. in determining the proper product line of craft brews, we performed this study and analyzed the beer and brewery data and collected metrics based on Alcohol by Volume (ABV) and International Bitterness Units (IBU). The success of Budweiser products will also depend on this and variety of geographical, demographic and product data analysis. This project is limited to the analysis of several metrics of Beer and Brewery data from several breweries in the United States. The goal of our case study is to find any obvious relationship with Beer and Brewery data and any new insights that could be exposed for further product development.

Data Handling

a. We first found how the breweries were spread across different states of USA. This would give us an insight of how the Beer popularity exists between these states.





Above bar graph and heat map shows Colorado tops in number of breweries and District of Columbia the lowest. Its interesting to know there's huge variation in the breweries across these states and further study could be made to understand this geographical distribution and the influencing factors. There are several possible assumptions on the influencing factors like beer popularity, demographics, geographical location related to land mass etc.

b. We merged the Beers and Breweries dataset by joining them on Brew_ID column and reviewed the first 6 and last 6 rows from the resulting dataset to ensure the data is merged correctly.

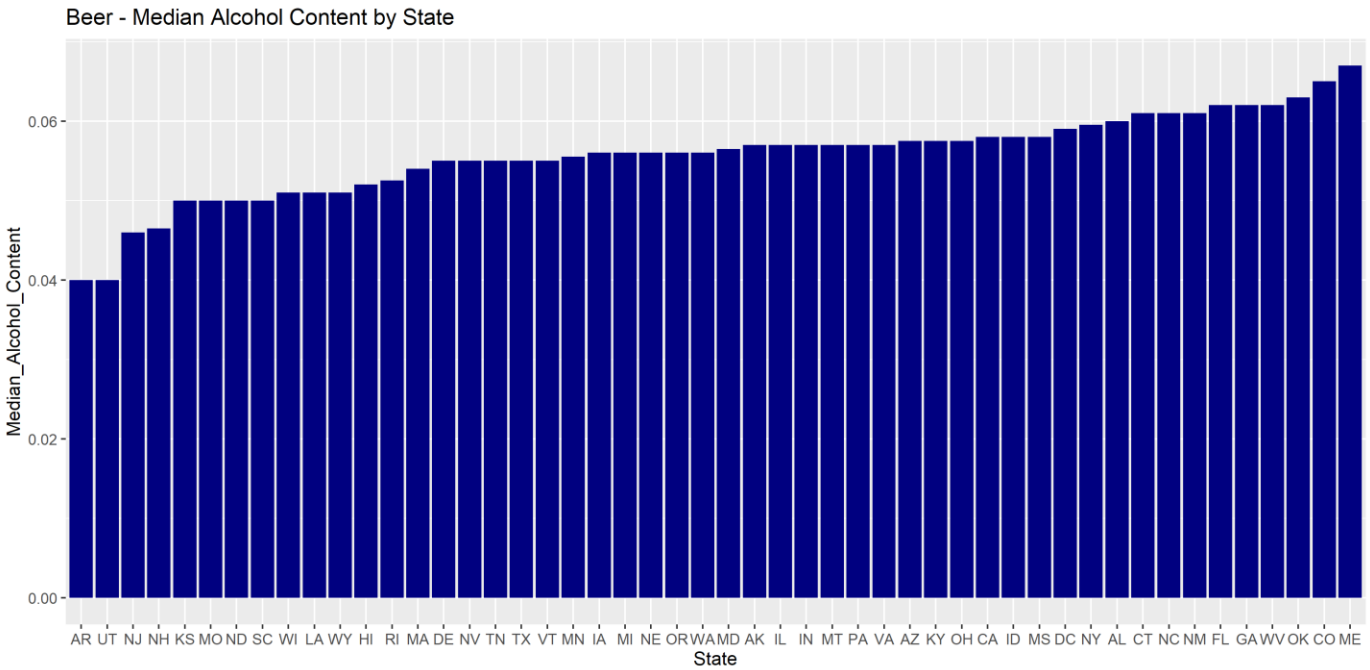
```
> head(BeerMerge1)
  Brew_ID      Beer_Name Beer_ID   ABV IBU                        Styl Ounces       Brewery_Name        City State
1       1  Get Together    2692 0.045  50                   American IP     16 NorthGate Brewing  Minneapolis    MN
2       1 Maggie's Leap    2691 0.049  26               Milk / Sweet Stou     16 NorthGate Brewing  Minneapolis    MN
3       1    Wall's End    2690 0.048  19               English Brown Al     16 NorthGate Brewing  Minneapolis    MN
4       1       Pumpion    2689 0.060  38                   Pumpkin Al     16 NorthGate Brewing  Minneapolis    MN
5       1    Stronghold    2688 0.060  25                American Porte     16 NorthGate Brewing  Minneapolis    MN
6       1   Parapet ESB    2687 0.056  47 Extra Special / Strong Bitter (ESB     16 NorthGate Brewing  Minneapolis    MN
```

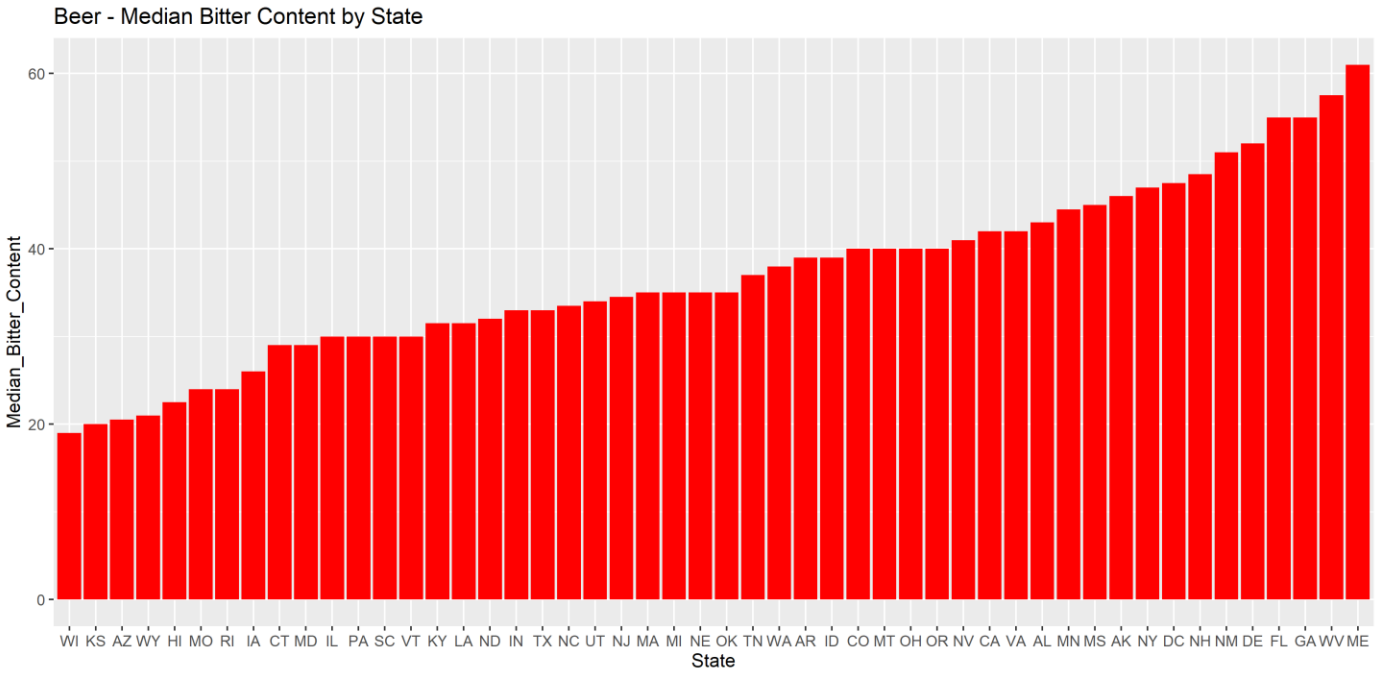
```
> tail(BeerMerge1)
     Brew_ID                 Beer_Name Beer_ID   ABV IBU                 Style Ounces                 Brewery_Name        City State
2405     556            Pilsner Ukiah      98 0.055  NA        German Pilsener     12          Ukiah Brewing Company      Ukiah    CA
2406     557 Heinnieweisse Weissebier      52 0.049  NA             Hefeweizen     12          Butternuts Beer and Ale Garrattsville    NY
2407     557        Snapperhead IPA        51 0.068  NA            American IPA     12          Butternuts Beer and Ale Garrattsville    NY
2408     557        Moo Thunder Stout      50 0.049  NA      Milk / Sweet Stout     12          Butternuts Beer and Ale Garrattsville    NY
2409     557        Porkslap Pale Ale      49 0.043  NA American Pale Ale (APA)     12          Butternuts Beer and Ale Garrattsville    NY
2410     558 Urban Wilderness Pale Ale      30 0.049  NA          English Pale Ale     12 Sleeping Lady Brewing Company    Anchorage    AK
```

c. After merge we analyzed the dataset for consistency and found there are high number of missing values(NA,<NA>,blank). We also found duplicate rows that required cleanup. With a total result set of 2410 observations, there were 1005 missing values for International Bitterness Unit(IBU) and 62 missing values for Alcohol volume by content(ABV), that's 42% of data for IBU and 2.6% of database for ABV. The data corresponding to these missing values were excluded from the dataset for our further case study. Any predictions that's resulting from this analysis may not be accurate as there's higher % of supporting data could not be used due to inconsistency.

```
> s
Brew_ID  Beer_Name  Beer_ID  ABV  IBU  Style
      0          0        0   62 1005      0
Ounces Brewery_Name     City State
     0            0        0     0
```

d. We looked at Median Alcohol volume(ABV) by State and Median Bitterness Unit(IBU) by State to find any correlation with these parameters.





The above results show there's a positive correlation between median alcohol content by volume and bitterness unit, however this correlation is not the same between states and there's wide variation on some of the states. Interestingly, this also shows there's some other factor(s) that influences this trend.

e. We also looked at the state that has the maximum alcohol content beer and state with highest bitterness beer.

   Colorado topped the spot with the beer "Lee Hill Series Vol. 5 - Belgian Style Quadrupel Ale" having maximum alcohol content by volume(ABV) of 0.128 (12.8%).

   On the bitterness side Oregon took the top spot with the beer "Bitter Bitch Imperial IPA" having maximum bitterness unit of 138.

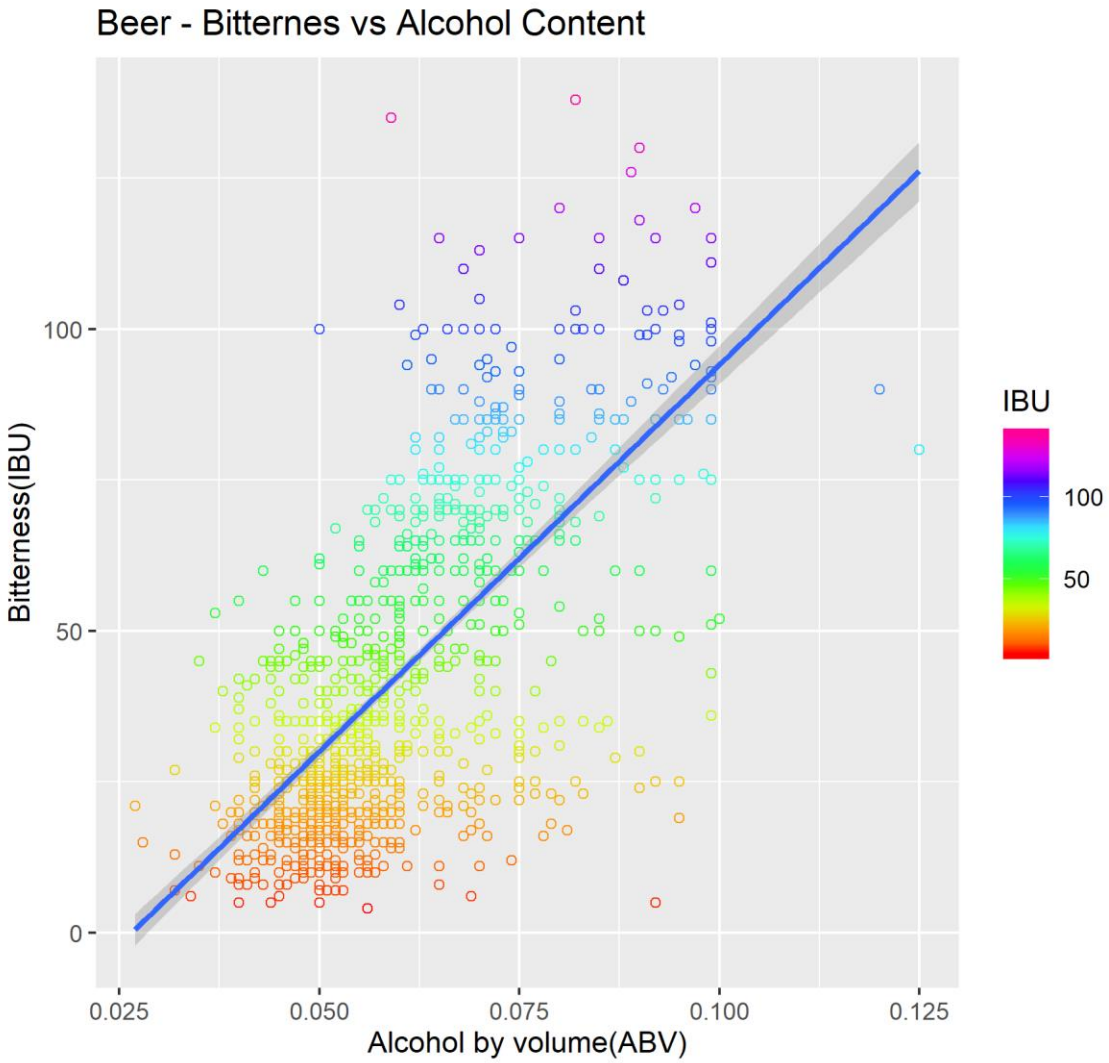
f. We looked at the summary statistics of ABV (Alcohol volume by content)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.02700	0.05000	0.05700	0.05991	0.06800	0.12500

   From the summary statistics, it can be inferred that the ABV of our craft beers dataset follows a normal distribution because the median and the mean values are very close. There is a slight right skew to the data. Also from the Median One thing to note is that 75% of all the beers have

a 5% ABV or more. However, the interquartile distance is only 1.8% suggesting that 50% of the data is bounded between 5% and 6.8% ABV. It is also important to remember that missing values are not included in the summary statistics, and there's a possibility that this representation may not be accurate.

g. We created a scatter plot to study any apparent relationship between Alcohol content and Bitterness in beer.



The scatterplot above provides visual evidence for a positive linear relationship between the alcohol content and the IBU values (R-Squared 0.64), meaning that 64% of the variation in the IBU of craft beers is explained by the Alcohol by Volume variable. Though this result might be satisfactory for a positive

correlation there can also be numerous other factors that may affect the IBU but are not accounted for in this linear regression.

There is a caveat to keep in mind for this regression analysis that 42% of the IBU data is missing. This could be strongly influencing the regression coefficients since currently there is no way of knowing exactly where that 42% falls within the overall spectrum of IBU values for all craft beers in the US.

Conclusion

From the case study we could say there's moderate positive correlation between bitterness and alcohol content of beer. It suggests there are other factors that's not accounted for in this case study to be influencing this factor. It would be interesting to know what other factors that's involved in the brewing process and if those data can be brought in to see if there's any change in this relationship. Also further study could be made by filling in the missing IBU data to see what influence this database might produce in this relationship.