Case Study 1

Team TAM

Authors: Andrew Heroy, Martin Garcia and Tej Tenmattam

MSDS 6306: Doing Data Science

Introduction:

Slurm cola sales have matured and by expanding into the growing beer industry we can continue to grow. Consumers tastes are changing and are drinking less cola. Whereas according to our research, Beer sales are on the rise. The purpose of this project was to investigate the trends in the beer industry. Beer drinkers are seeking quality over quantity, evidenced by continued growth in premium, super-premium and craft categories; while an increasing number of health-conscious consumers are turning to low and no alcohol varieties. In this project, we have identified how alcohol percentage in beer varies by type of beer and region the beer is brewed in. Our research also shows the various ABU and IBV values of craft beers across the US.

Research Methodology:

At Slurm beverages, we have an awesome team of Data Scientists. Our team follows an agile methodology with weekly scrum meetings and follows the DevOps principles and tools for our research and development.

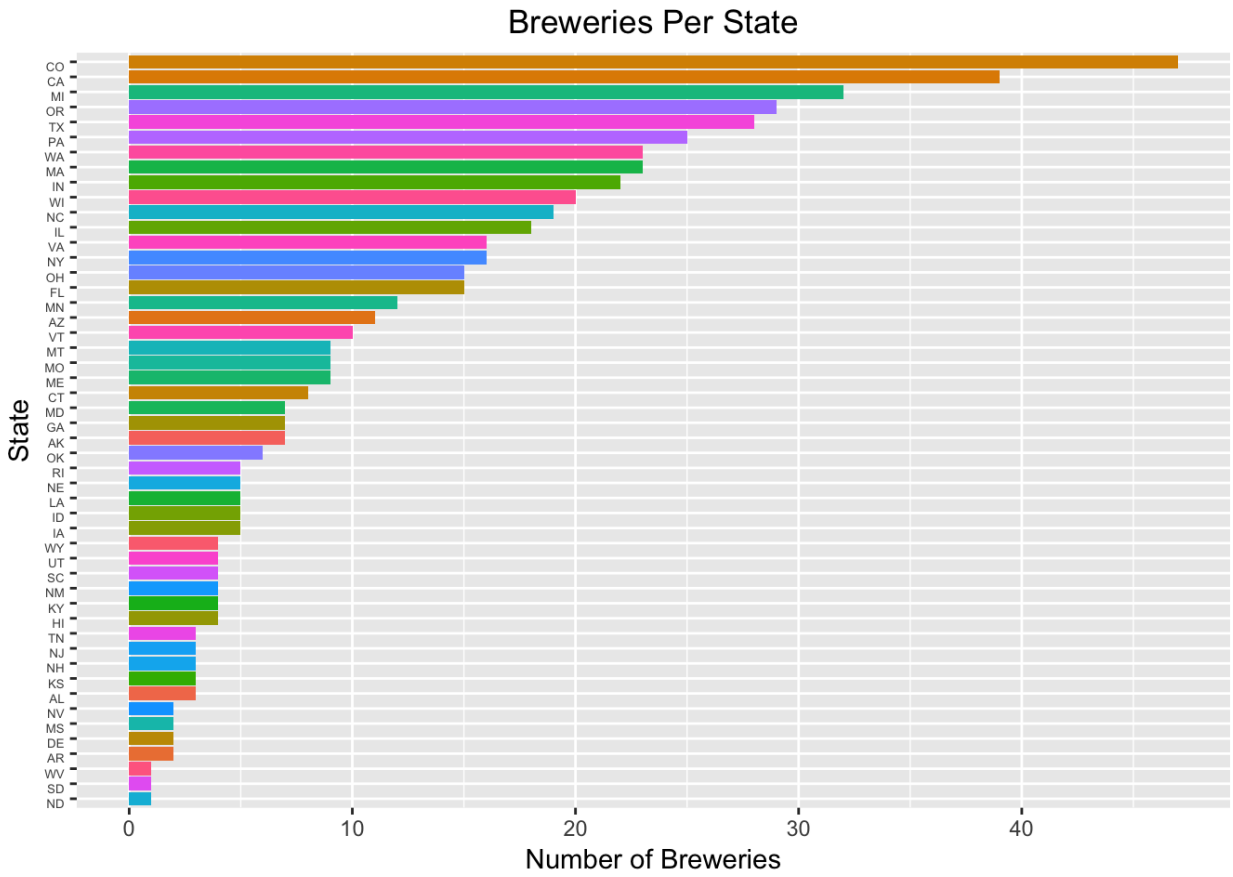
Our research and findings are easily reproducible. We have a public git repository here: https://github.com/tejten/team_tam_case_study0. There are four branches, one master and three branches for each one of the data scientist who has worked on this project. The directories are layered as follows:

- The “Data” directory consists of the data sets. We got our Beer and Brewery data sets from the “Alcohol & Tobacco Tax and Trade Bureau” https://www.ttb.gov/index.shtml
- The “Code” directory is where you will find the source R code to support our claims and an R markdown file as well as a Codebook explaining our table structures are located.
- The “Presentation” directory consists of the “html” output of our research as well as the presentation materials (ppt) and a PowerBI dashboard that you can use for some fun exploratory perusing at your own leisure.

Questions Addressed:

Question 1: How many breweries are present in each state?

Here is our bar chart of breweries by state. It is to no one's surprise that Colorado is leading the pack, closely trailed by California. Two states well known for producing vast varieties and quantities of beer.



Question 2: Display the first 6 observations and the last six observations of the beer data in an alphabetical order.

	Name <fctr>	Beer_ID <int>	ABV <dbl>	IBU <int>	Brew_ID <int>	Style <fctr>	Ounces <dbl>	Brewery <fctr>	City <fctr>	State <fctr>
1	Pub Beer	1436	0.050	NA	409	American Pale Lager	12	10 Barrel Brewing Company	Bend	OR
2	Devil's Cup	2265	0.066	NA	178	American Pale Ale (APA)	12	18th Street Brewery	Gary	IN
3	Rise of the Phoenix	2264	0.071	NA	178	American IPA	12	18th Street Brewery	Gary	IN
4	Sinister	2263	0.090	NA	178	American Double / Imperial IPA	12	18th Street Brewery	Gary	IN
5	Sex and Candy	2262	0.075	NA	178	American IPA	12	18th Street Brewery	Gary	IN
6	Black Exodus	2261	0.077	NA	178	Oatmeal Stout	12	18th Street Brewery	Gary	IN

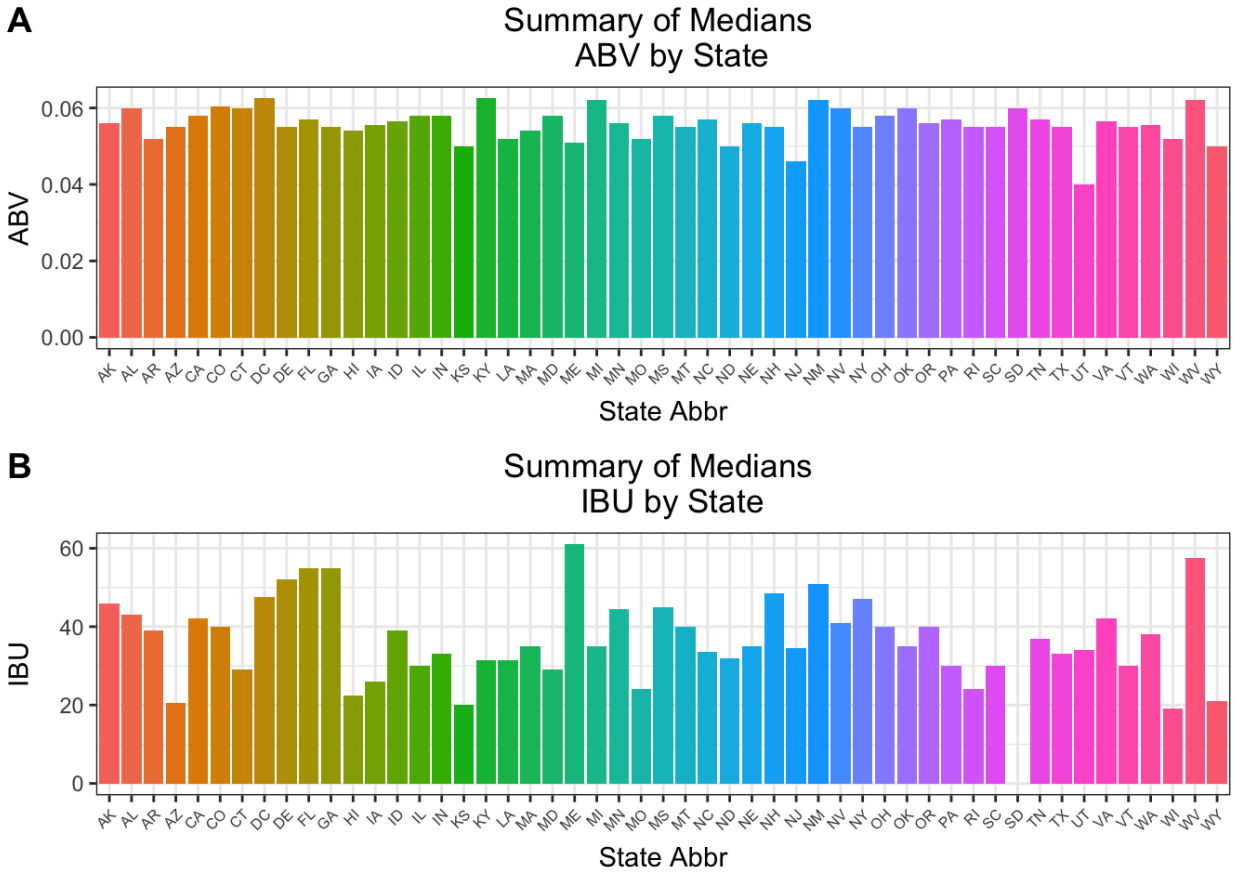
	Name <fctr>	Beer_ID <int>	ABV <dbl>	IBU <int>	Brew_ID <int>	Style <fctr>	Ounces <dbl>	Brewery <fctr>	City <fctr>	State <fctr>
2405	Rocky Mountain Oyster Stout	1035	0.075	NA	425	American Stout	12	Wynkoop Brewing Company	Denver	CO
2406	Belgorado	928	0.067	45	425	Belgian IPA	12	Wynkoop Brewing Company	Denver	CO
2407	Rail Yard Ale	807	0.052	NA	425	American Amber / Red Ale	12	Wynkoop Brewing Company	Denver	CO
2408	B3K Black Lager	620	0.055	NA	425	Schwarzbier	12	Wynkoop Brewing Company	Denver	CO
2409	Silverback Pale Ale	145	0.055	40	425	American Pale Ale (APA)	12	Wynkoop Brewing Company	Denver	CO
2410	Rail Yard Ale (2009)	84	0.052	NA	425	American Amber / Red Ale	12	Wynkoop Brewing Company	Denver	CO

Question 3: Identify the breweries who haven't listed their ABV and IBU ratings?

```
Name Beer_ID     ABV     IBU Brew_ID   Style  Ounces Brewery    City   State
   0       0      62    1005       0       0       0       0       0       0
```

Parsing the dataset for NA values we see that our ABV has a total of 62 NA value's while the ABV column shows 1005 NA values. It would appear that breweries haven't fully adopted the IBU rating scale for their beers, This won't give us as complete a picture of the market, but there's not much we can do about that.

Question 4: Display the Summary of Medians of ABV and IBU by State



Question 5: Which state has the maximum alcoholic (ABV) beer? Which state has the most bitter (IBU) beer?

```
##   State Max_ABV Max_IBU
## 1 " CO"   0.128     104
```

```
##   State Max_ABV Max_IBU
## 1 " OR"   0.088     138
```

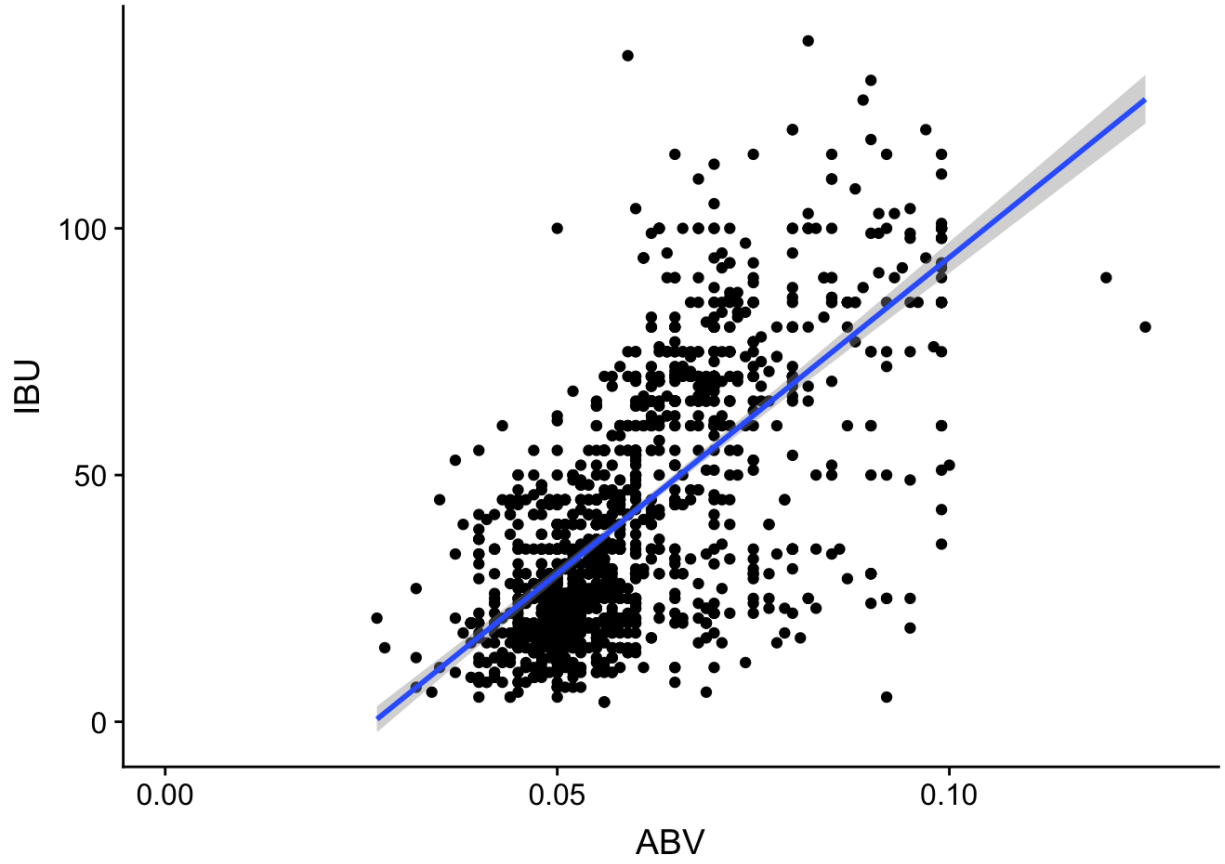
It appears that if we're looking for the highest alcohol content in our beer, Colorado is the place to go! Now if we were looking for the most bitter beer available, then Oregon would be the destination of choice.

Question 6: Summary statistics for the ABV variable.

```
##       ABV
##  Min.   :0.00100
##  1st Qu.:0.05000
##  Median :0.05600
##  Mean   :0.05977
##  3rd Qu.:0.06700
##  Max.   :0.12800
##  NA's   :62
##        ABV
## 0.01354173
```

Looking at the table above we can see the values of ABV range from [0.001 - .1238], With a Median of 0.056 ABV, the data shows us a standard deviation of 0.014.

Question 7: Is there an apparent relationship between the bitterness of the beer and its alcoholic content?



You can see from the plot above that there is a slight correlation of ABV to IBU. A low p value and R correlation of 0.67 tells us we should target a mildly bitter beer to best mirror the clusting of the scatterplot.

Conclusion:

From this research project we can conclude that Colorado and California are leading in the number breweries per state. If we're looking for the highest alcohol content in our beer, Colorado is the place to go! Now if we were looking for the most bitter beer available, then Oregon would be the destination of choice. And finally we noticed a slight correlation of ABV to IBU. A low p value and R correlation of 0.67 tells us we should target a mildly bitter beer as number of health-conscious consumers are turning to low alcohol varieties.