Kyle Kuberski

SMU DDS Project 1

Professor Bivin

March 3, 2023

Project 1 Conclusion

From our EDA, Budweiser can engage in extrapolating powerful insights into the general areas of note, visually appealing, and intuitive inferences within the Brewery and Beer datasets. Overall, we have found that California and Colorado have a staggering amount of breweries and a large variation from state to state. Graphing the median and maximum ABV and IBU values allows us to summarize key information. American IPA’s and American Pale Ales lead the median amounts of beers in California and Colorado; uniformly high ABV maxes can be found in states like New York, Ohio and Oklahoma. Sixpoint Craft Ales in New York stands out with it’s beer, the “Abigale”, at 138 international bitterness units. For ABV, Colorado’s Upslope Brewing Company has the maximum ABV of all betters with its “Lee Hill Series Volume 5 – Belgian Style Quadrupel Ale”. The overall distribution of ABV tends to skew right, with outliers of low ABV values near .0001 units, and has an overall max of .128 units. The Median is around .0056 with the mean being .05826 units.

As for relationships between alcohol by volume and international bitterness units, there is an overall positive correlation between the two. This may mean that as bitterness increases, so does alcoholic contents. Although, this may not be true for all cases, as we can find outliers across the graph that tell us that there is some cases of high ABV and low IBU (and vice versa). With a KNN classifier our positive class was Ales, which means that more ales were identified as true positives than IPA’s. Through visualization of this data, we can see the relationship of the two which included the majority of both being centralized between .05 ABV and 50 IBU.

Conclusively, the data summarizations and visualizations will help to make clear inferences in the direction of many business intentions and interests. Budweiser may wish to engage with states with smaller amounts of breweries for a more blue-ocean (less competition) approach, or may wish to alter their beers in accordance with ABV and IBU relationships. These statistical findings can increase overall business knowledge gleaned from sampled data from just two datasets.