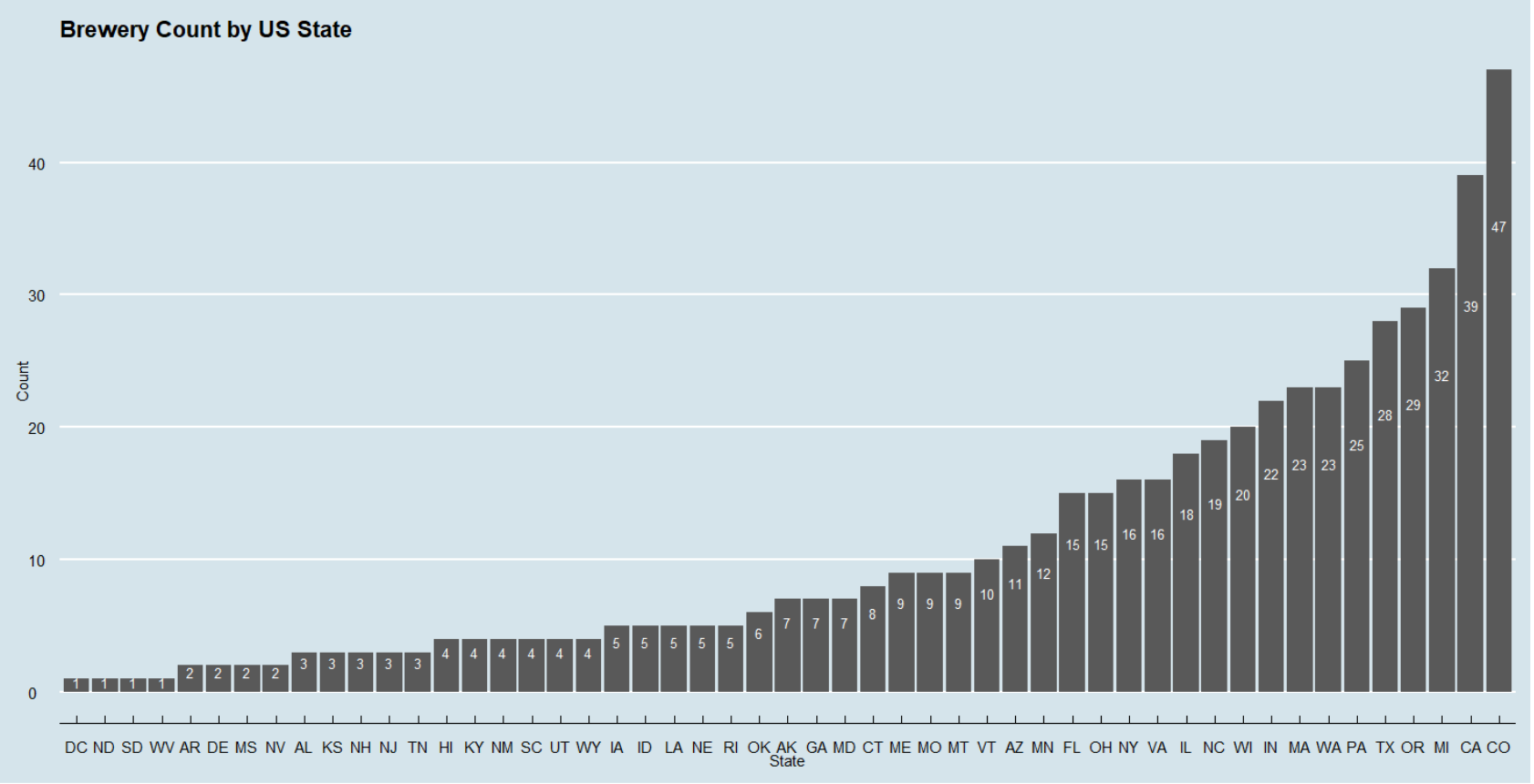
Analyzing US Craft Beers and Breweries

ALMUHANNAD QNEIS & AHMAD SALAMA

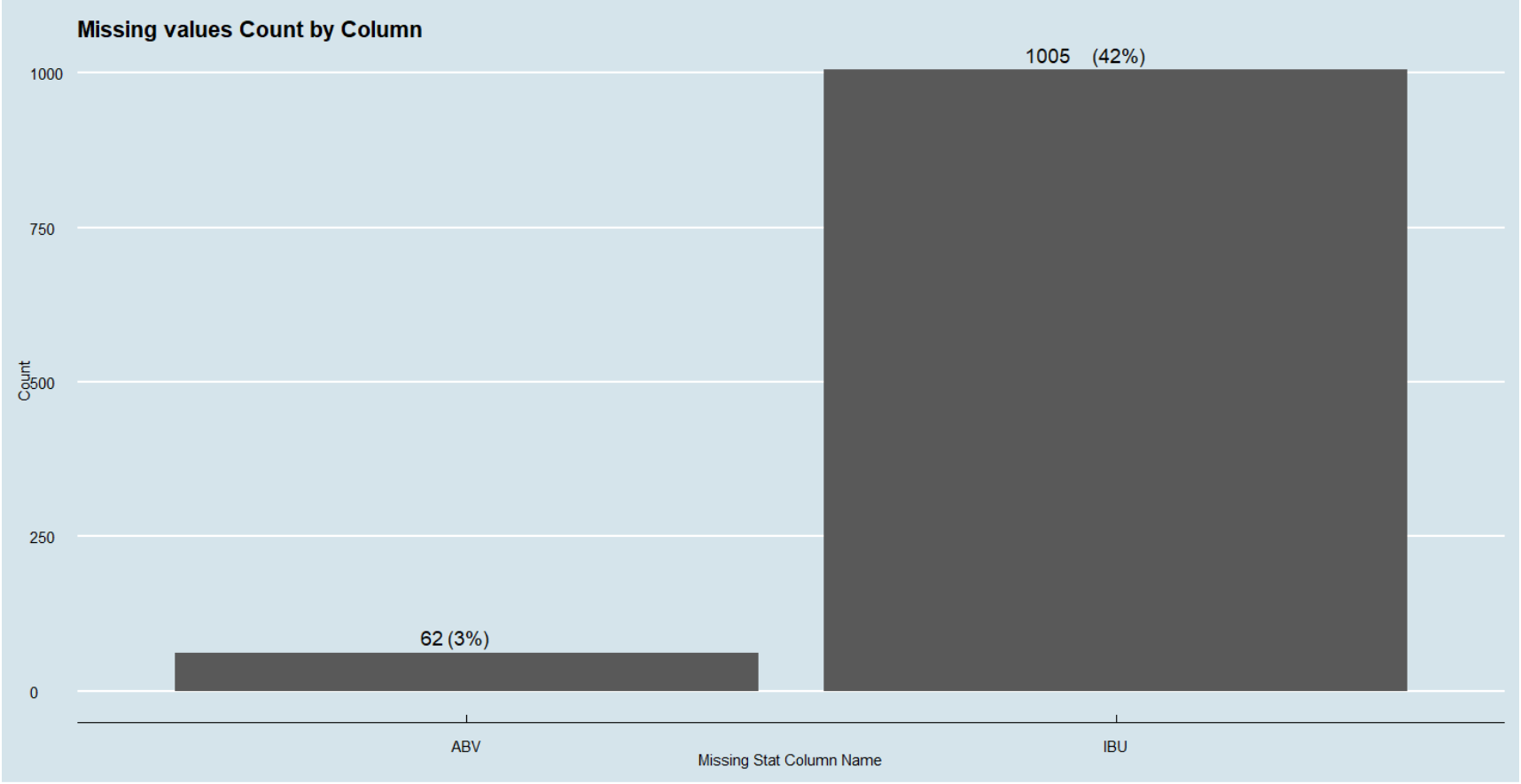
Brewery Count by State

- The brewery data set is complete, and there are no missing values.
- Colorado has the most craft breweries. The state has 47 in total, compared to 2nd and 3rd place California and Minnesota, which have 39 and 32 breweries, respectively.
- Washington D.C., North Dakota, South Dakota, and West Virginia is in the last place with only 1 breweries in each state.



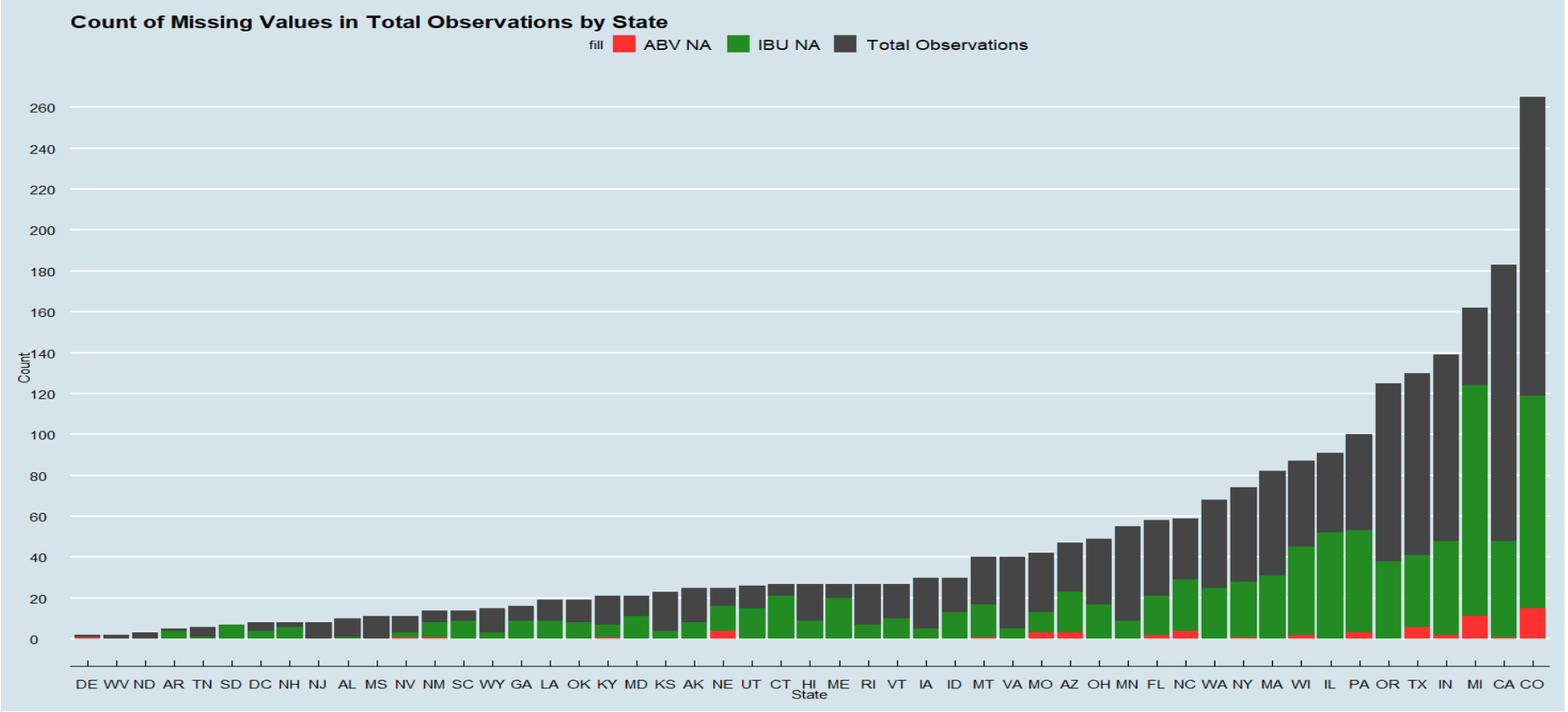
Missing Values

- Data quality and completeness are critical for decision-making.
- There are 62 missing values of the Beer Alcohol Content (ABV).
- There are 1005 (42%) missing values of Beer bitterness (IBU).
- Trying to collect the missing values could be complex and expensive in this case.



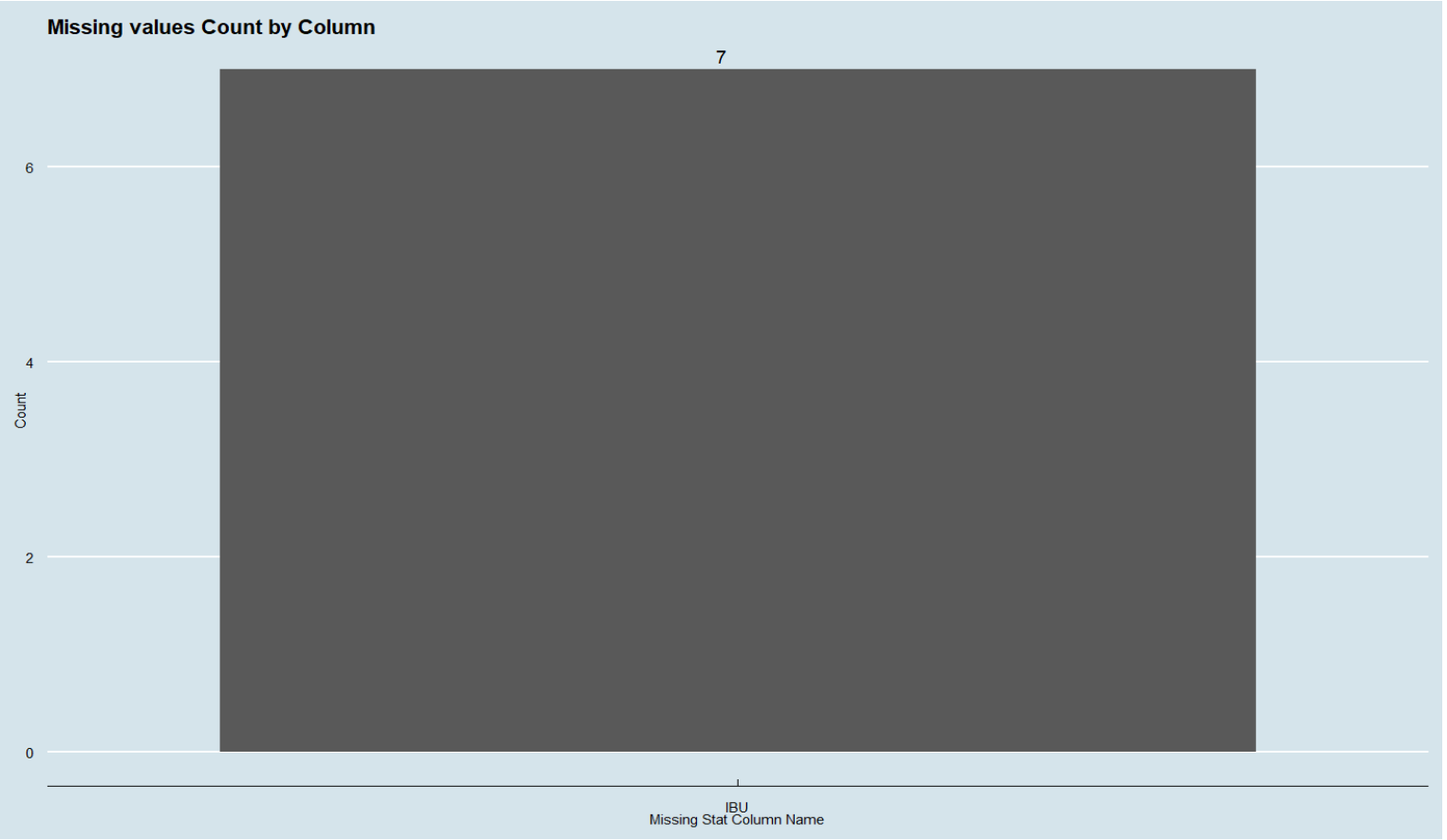
Count of Missing Values in Total Observations by State

- visually representations of the frequency of missing data (NA) within the "ABV" and "IBU" columns, along with the overall count of observations categorized by state.
- Impute the missing values by assigning the mean value of available data grouped by City and Style. Then by style, and for the remaining values impute by the mean of values grouped by state.



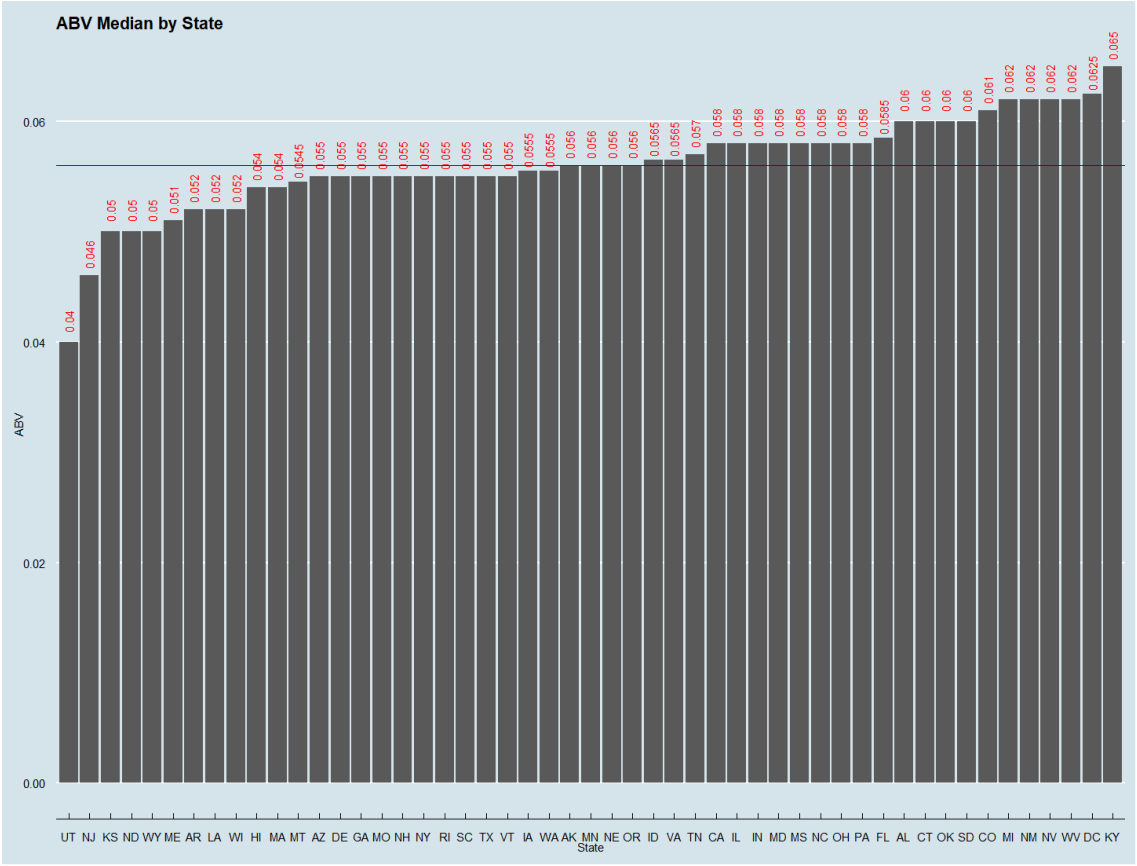
Missing Values

- Finally, for the remaining missing values, we assigned the mean value of available data grouped by State.
- As a result, through the three rounds, we imputed all ABV missing values, and we ended up with only 7 IBU missing values.



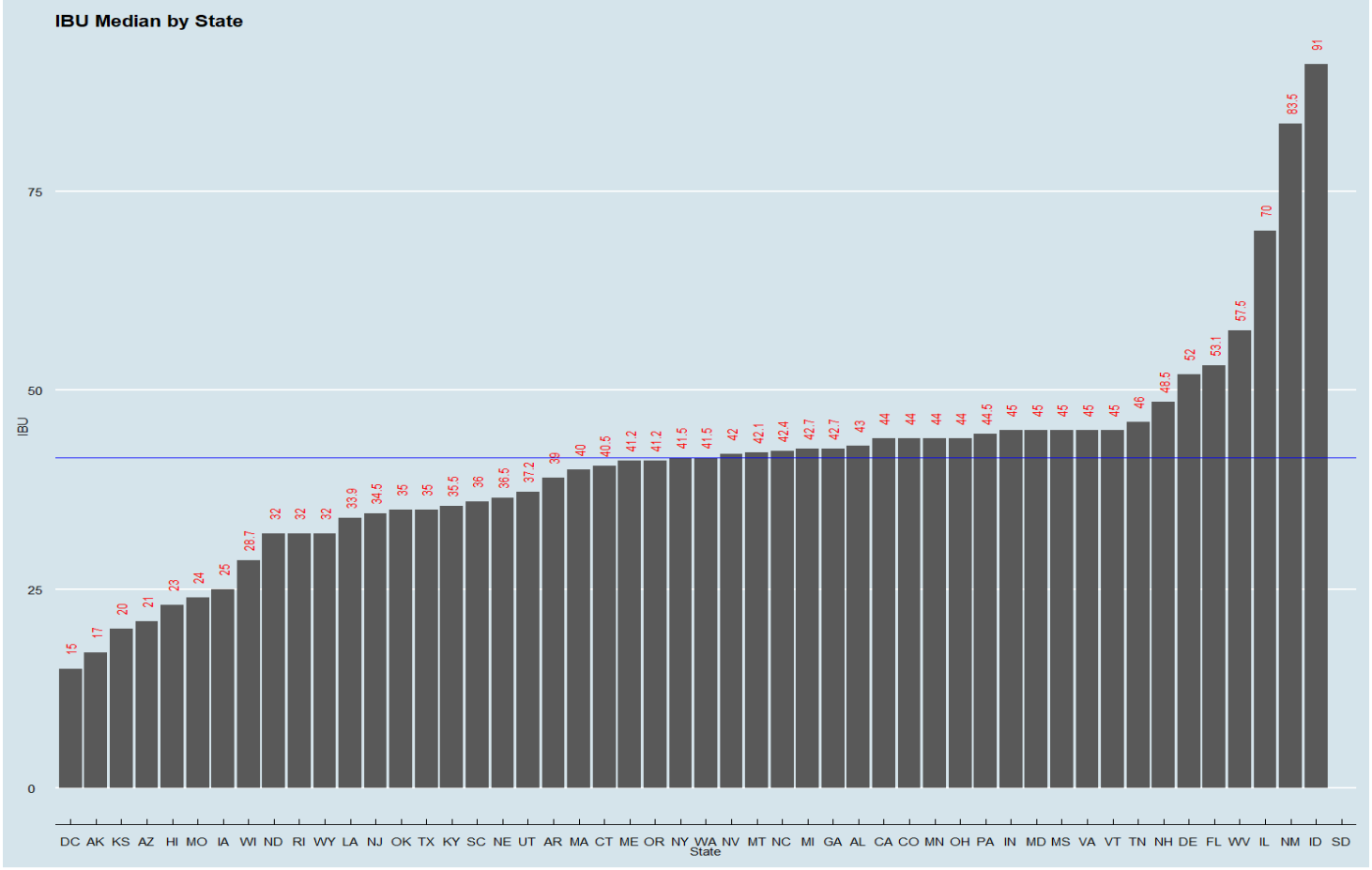
ABV Median Analysis

- Kentucky have Highest ABV Median value: 0.065. Followed by DC & WV
- Utah has the lowest ABV median value: 0.04. Second & third from last is NJ & KS.
- There are about 11 states with ABV median equal or greater than 0.06



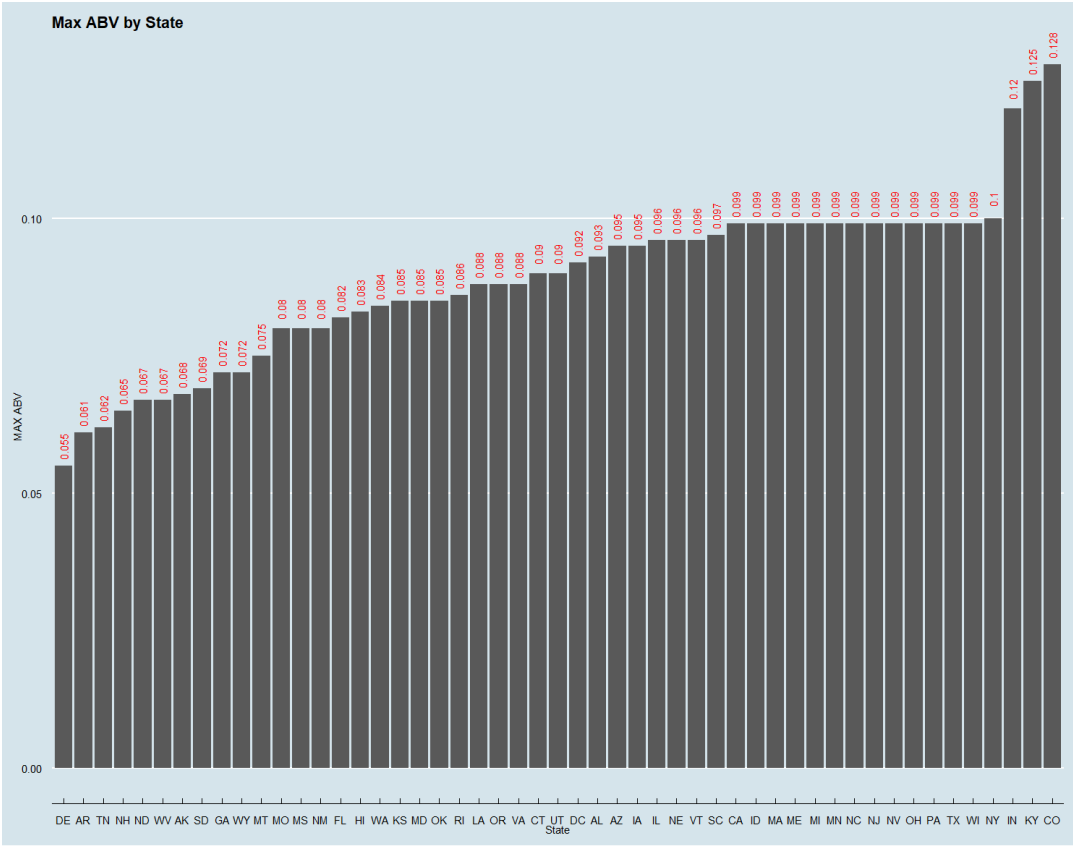
IBU Median Analysis

- Idaho has the highest IBU median value: 91. Second & third are New Mexico & Illinois.
- DC has the lowest median value: 15
- Data Suggest that States have distinct tastes regarding the bitterness of their beers.



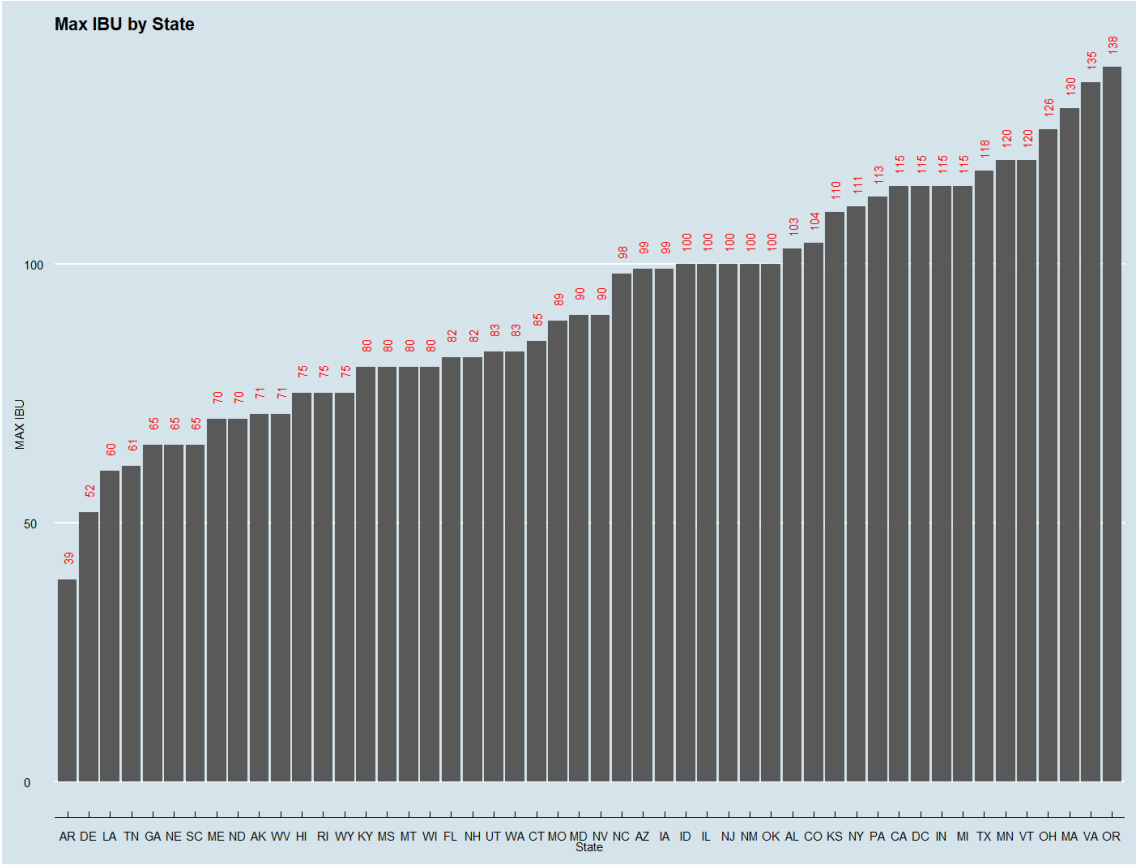
Max ABV by state

- Maximum ABV is in Colorado: 0.128
- The lowest Max ABV value is in Delaware: 0.055
- There is an opportunity for Budweiser to work with communities to install breathalyzers and other safety measures to mitigate DUI in states & cities with highest ABV.



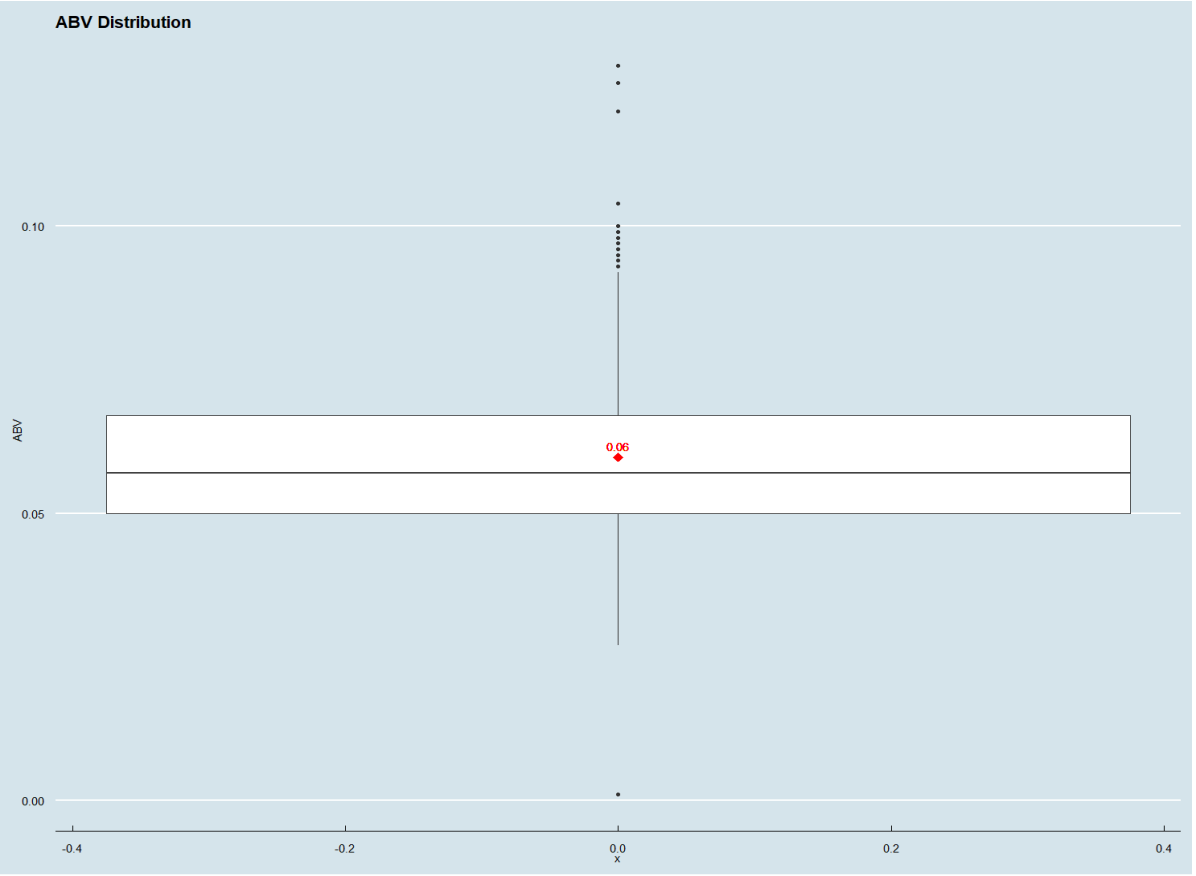
Max IBU by state

- The highest Maximum IBU is in Oregon: 138
- The lowest Maximum IBU value is in Arkansas: 39
- Further studies regarding IBU distribution within state might help in tailoring and marketing product.



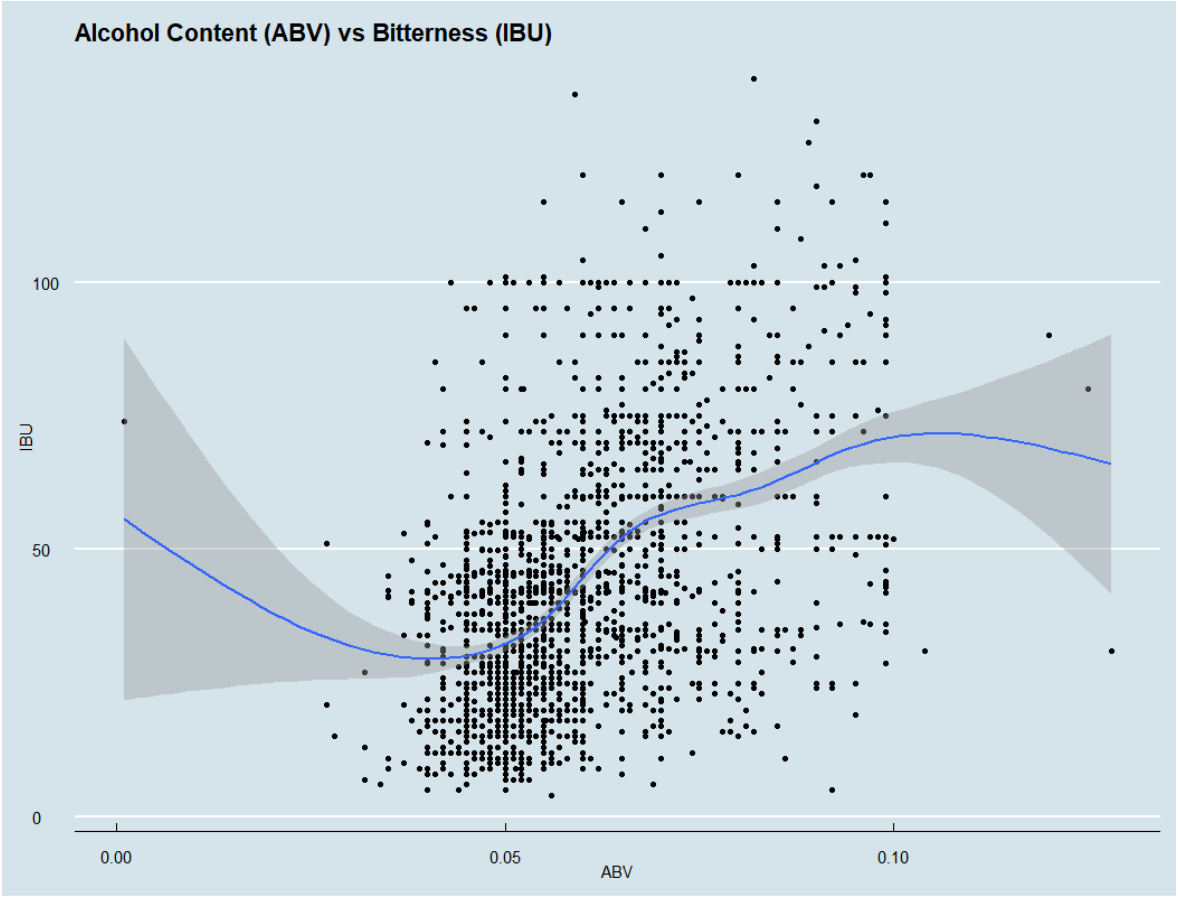
ABV Summary Statistics and Distribution

Summary Analysis					
Min	1st Quartile	Median	Mean	3rd Quartile	Max
0.001	0.05	0.057	0.0598	0.067	0.128



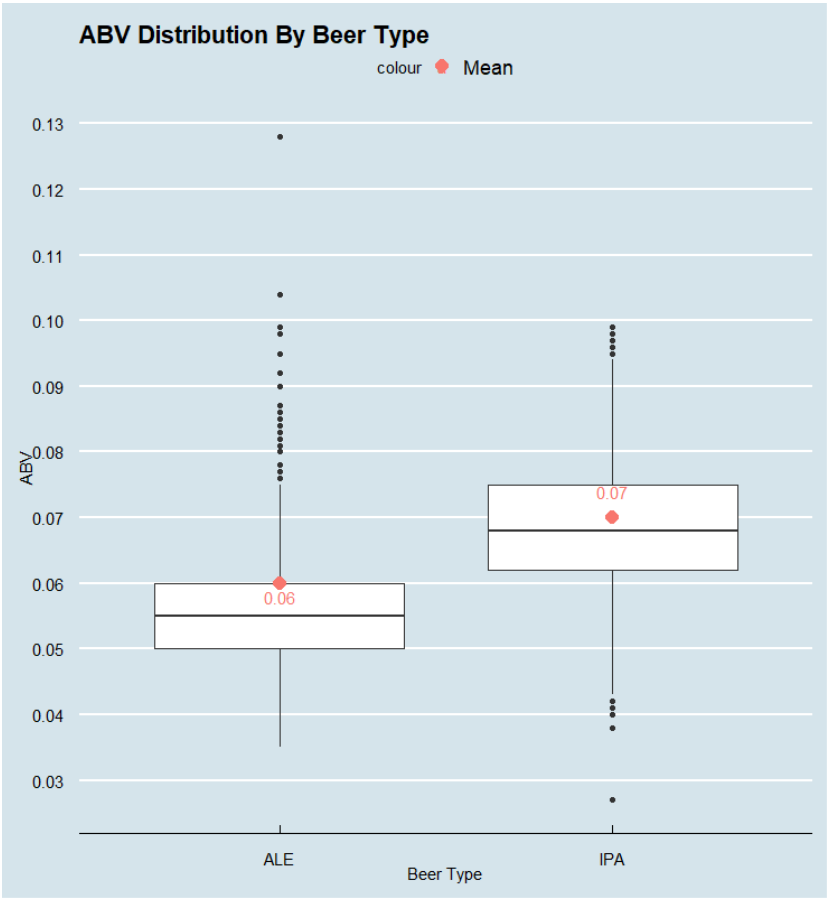
Relationship between IBU and ABV

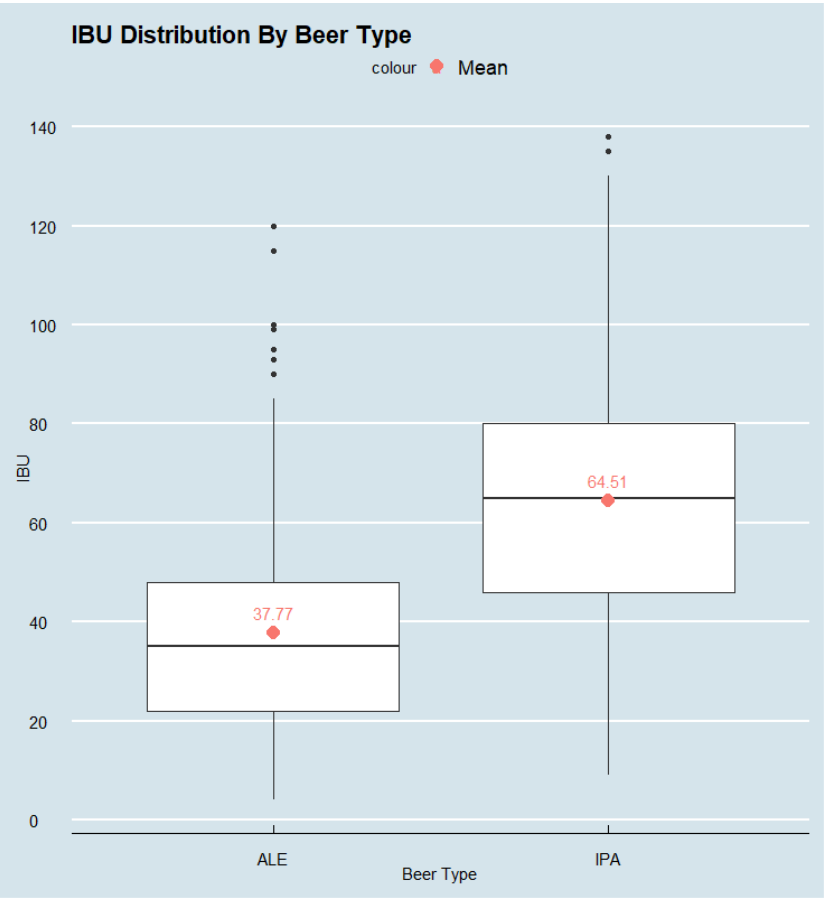
- Visually, it looks like there is a moderate positive association between ABV & IBU.
- The correlation value is: 0.481
- Evidence suggests that an increase in ABV moderately associated with an increase in IBU.
- The relationship between IBU & ABV might be stronger if we incorporate other features like the Beer Type and City and\or State.



IBU and ABV Distribution By Beer Type

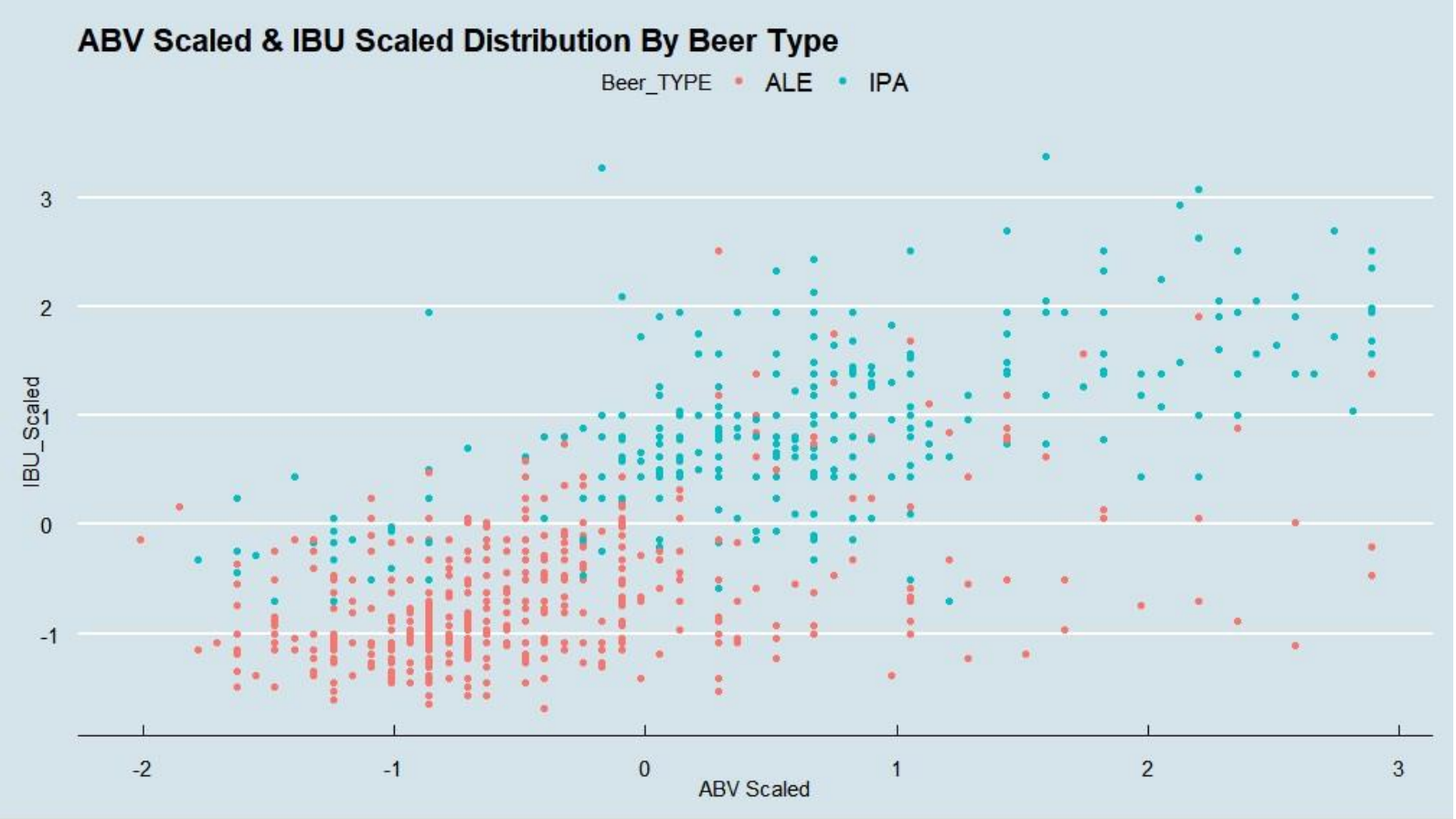
- Data Suggest that the IPA IBU mean & median are higher than ALE mean and median
- Data Suggest that the IPA ABV mean & median are higher than the ALE mean and median
- ALE ABV has a wider distribution. This might decrease the accuracy of our prediction model and will require further optimizations.





Scaling of IBU and ABV Distribution By Beer Type

- Most of ALE & IPA beers are concentrated in 2 regions which make it a good candidate for k-NN modeling.
- This plot chart shows a scaled ABV and IBU distribution between the beer types Ale and IPA



KNN Model

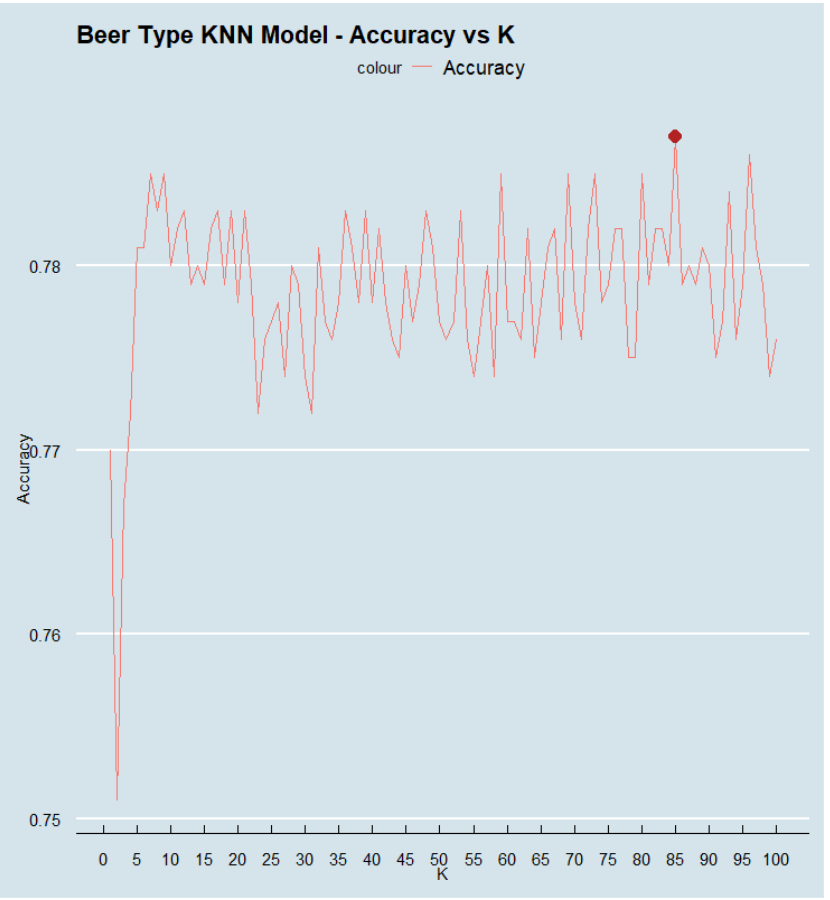
- **K-Nearest Neighbors** is the supervised machine-learning algorithm used for classification or predictions based on existing data.
- It classifies data based on distance and majority votes. It finds the k-nearest neighbors to the test data, and then classification is performed by the majority vote.
- Maximum Accuracy could be achieved by selecting the optimal K value.

Figure 1: Photo via educba.com

Optimizing KNN Model Accuracy: Finding the best k Value

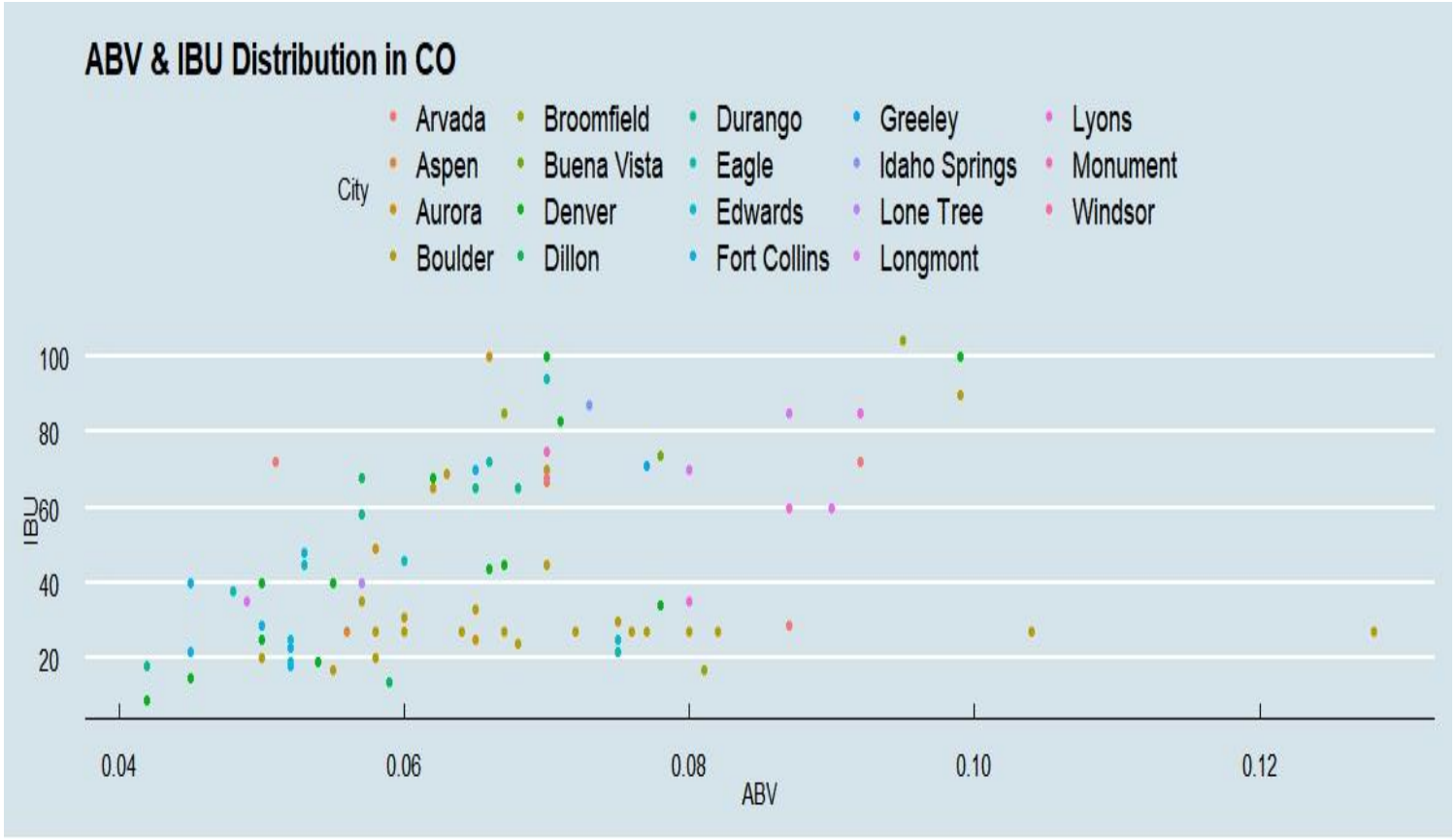
- Find optimal K for kNN model to achieve the best accuracy.
- Tested K values: 1: 100 with 20 iterations for each k and calculate the average accuracy for each k.
- The plot to the right shows how the accuracy of the kNN model changes with different values of k
- identify the best value of k that maximizes accuracy.
- K = 85 provides the highest accuracy on our testing data set

```
Confusion Matrix and Statistics

          Reference
Prediction ALE IPA
       ALE 300  55
       IPA  32 107

               Accuracy : 0.8239
                 95% CI : (0.7874, 0.8565)
    No Information Rate : 0.6721
    P-Value [Acc > NIR] : 2.551e-14

                  Kappa : 0.5854

 Mcnemar's Test P-Value : 0.01834

            Sensitivity : 0.9036
            Specificity : 0.6605
         Pos Pred Value : 0.8451
         Neg Pred Value : 0.7698
             Prevalence : 0.6721
         Detection Rate : 0.6073
   Detection Prevalence : 0.7186
      Balanced Accuracy : 0.7821

       'Positive' Class : ALE
```



ABV & IBU Distribution By City in Colorado

- Looking deeper into ABV and IBU distribution in Colorado, there looks like there is also distinct IBU and ABV distribution by each city.
- Visually, data suggest that Cities within the CO State have distinct tastes regarding the bitterness and alcohol content of their beers.
- We advise that Budweiser use the data to tailor products based on cities within states.



Questions

References

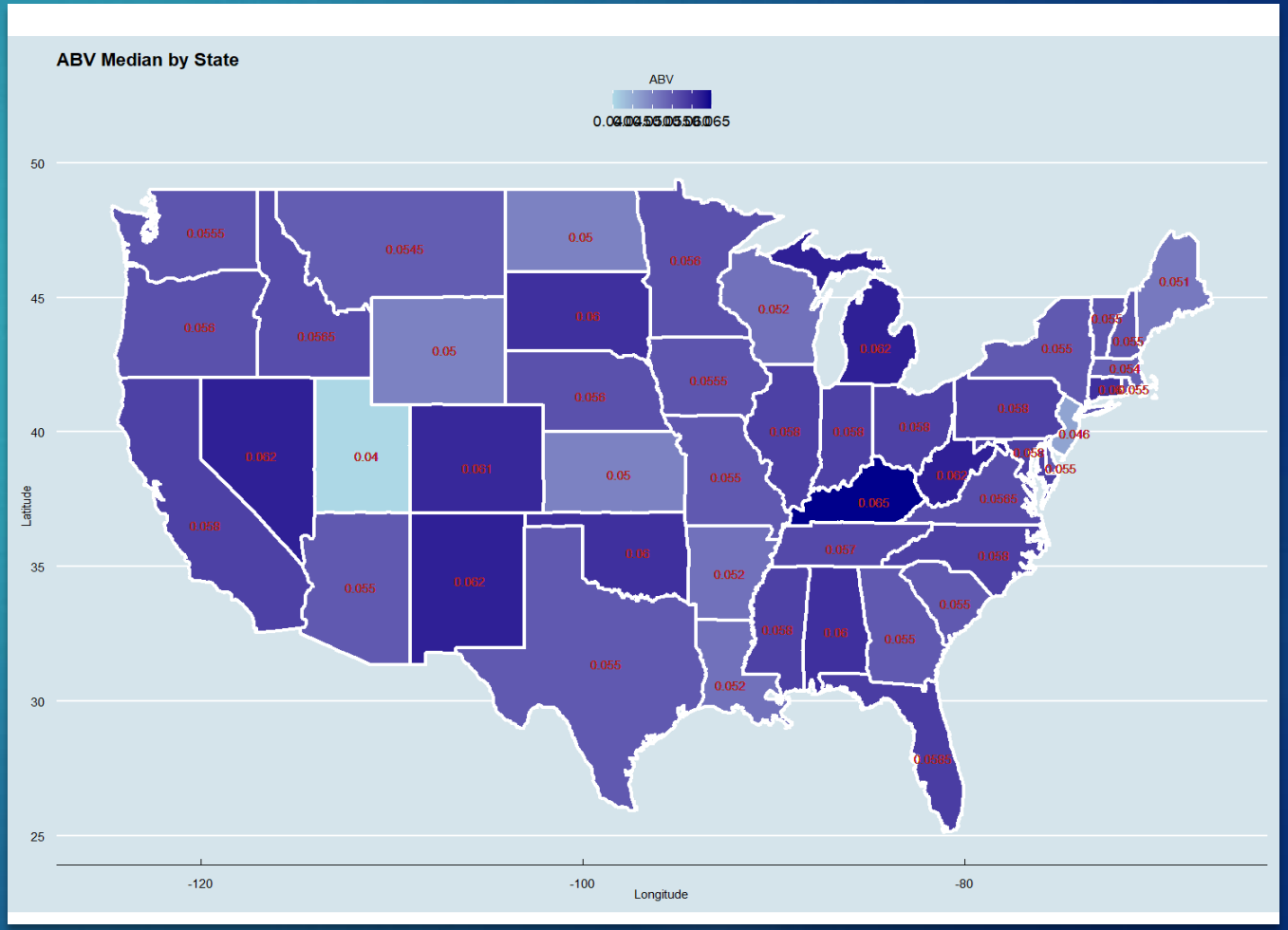
- https://www.educba.com/nearest-neighbors-algorithm/

Appendix

- The following slides are extra, and we couldn't find the time to go over.

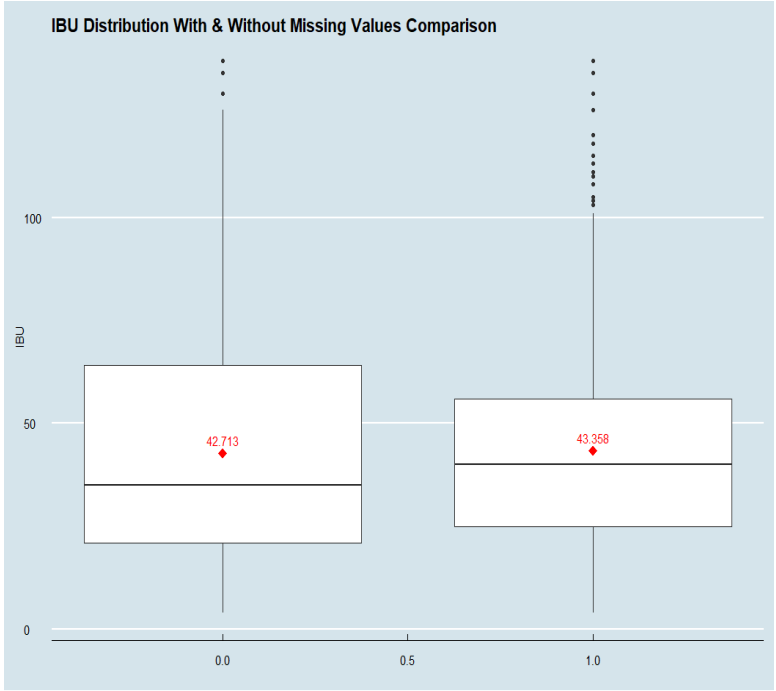
ABV Median Analysis

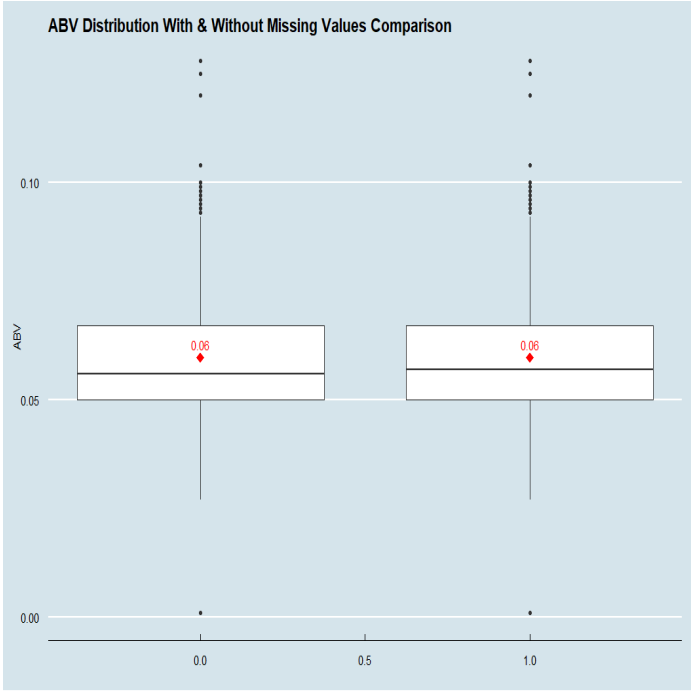
- Washington DC & Kentucky have Highest Median value: 0.065
- Utah has the lowest ABV median value: 0.04
- Only 11 states with ABV median equal or greater than 0.06



Missing Values

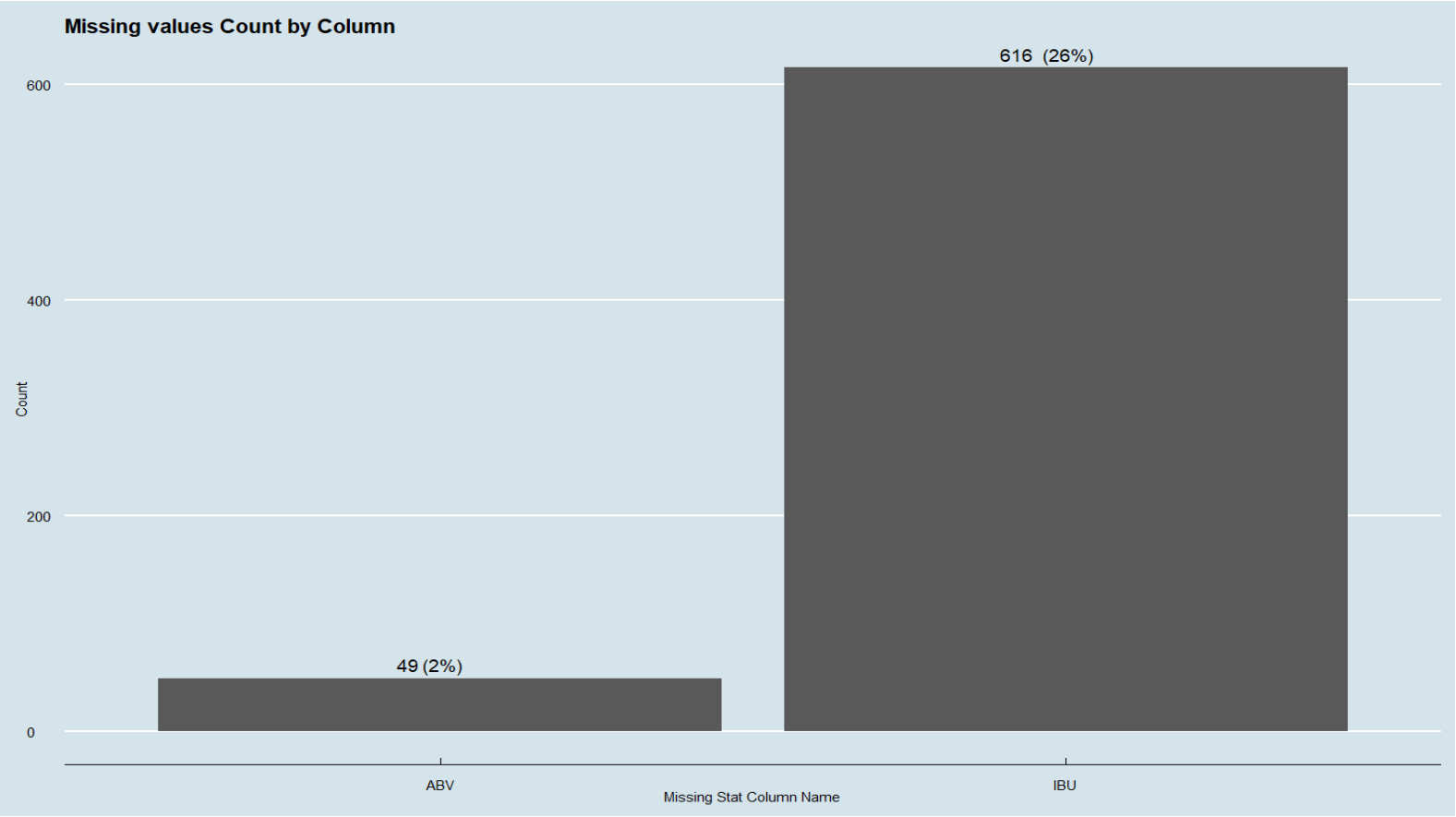
- After the handling the missing values using our approach, we didn't notice significant changes in the Mean of ABV and IBU.





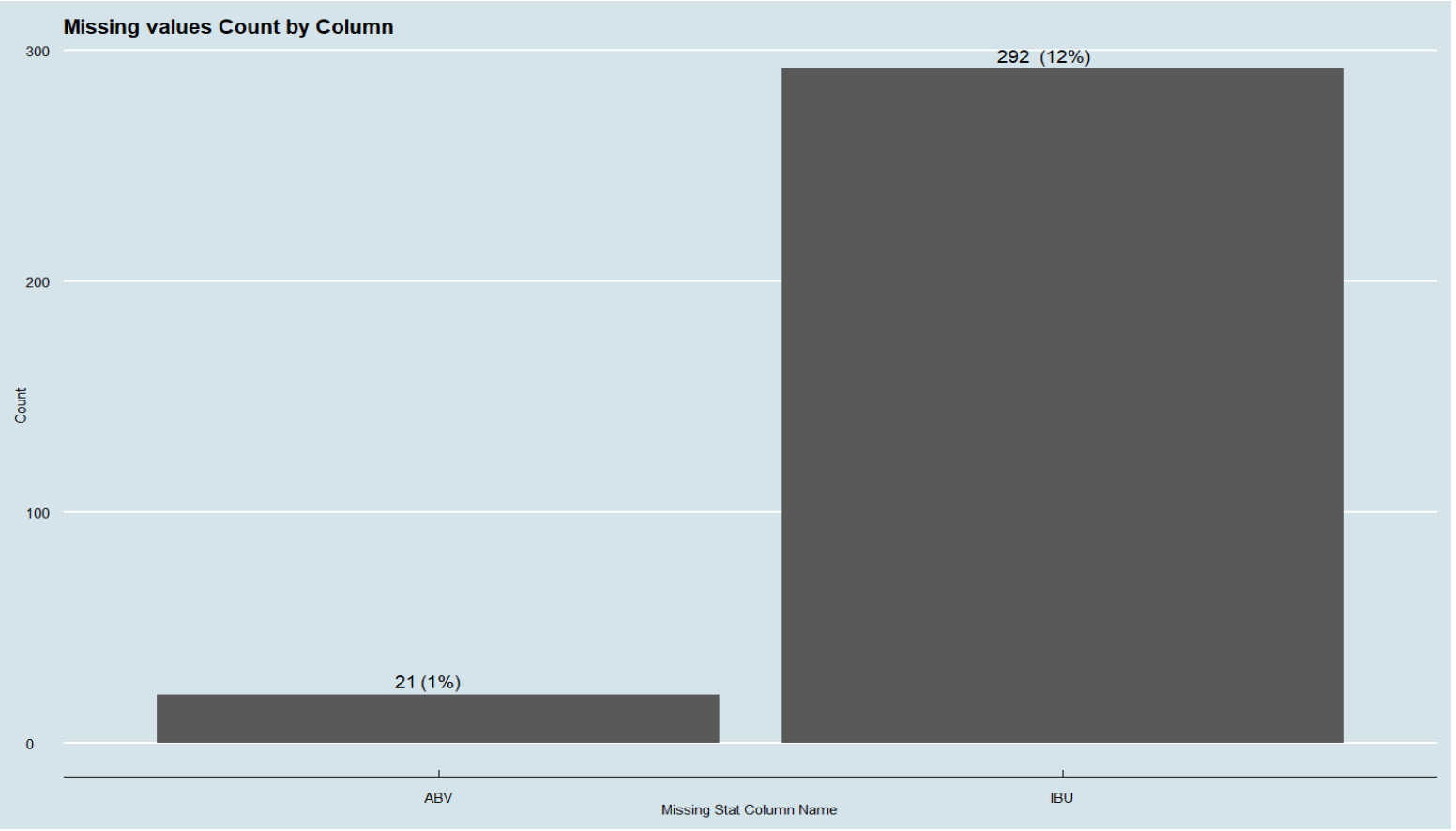
Missing Values

- Fix the missing values by assigning the mean value of available data grouped by City and Style.
- These combined factors we used would indicate a more precise mean to input for missing values.
- As a result, The count of missing values was reduced significantly.



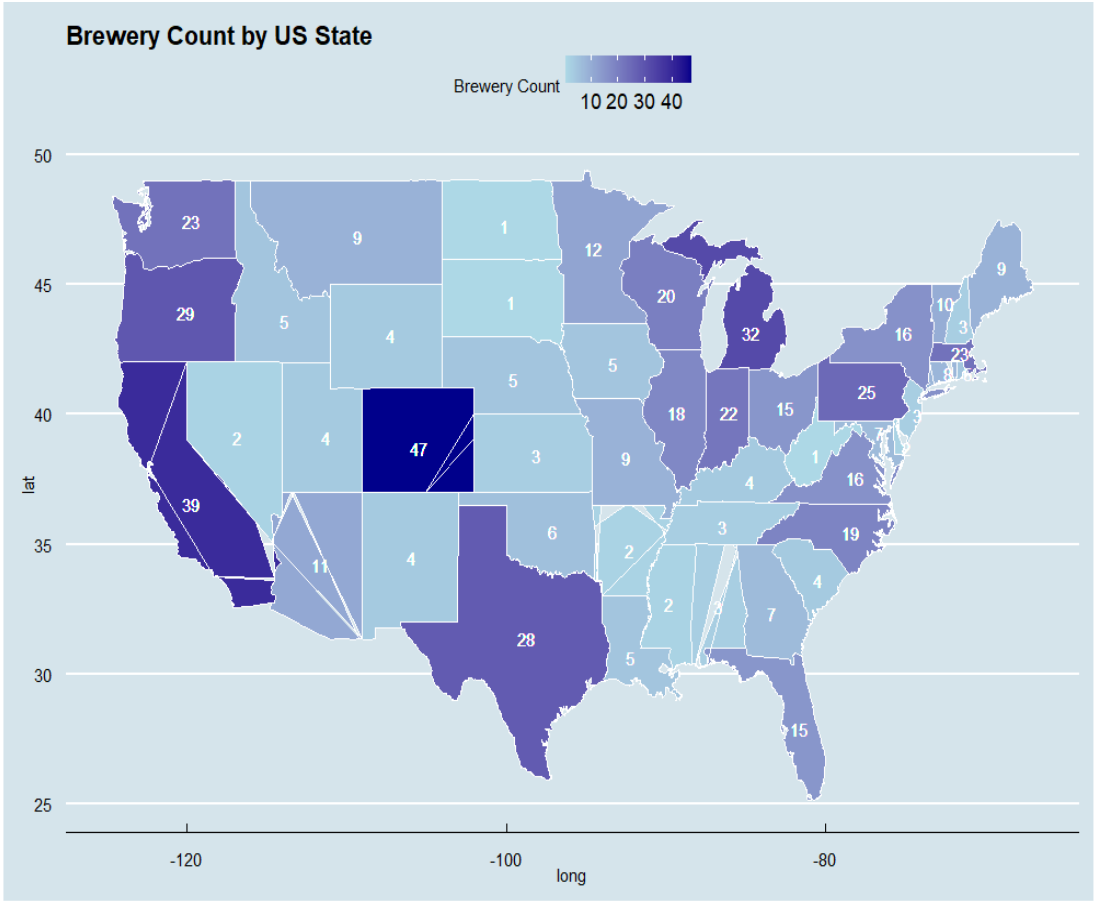
Missing Values

- For remaining missing values, assign the mean of available data grouped by style.
- As a result, The count of missing values was reduced significantly as well, about 50%.



Brewery Count by State

- The States with more breweries are likely to be highly competitive.
- Budweiser may need to focus and invest on expanding its marketing and promotional campaigns to maintain a competitive edge.
- Spending more on Ads and discounts might be a good thing to consider.
- There is an opportunity for Budweiser to establish and expand its presence in states with a low number of breweries. Potential Markets



IBU and ABV Distribution By Beer Type

- Most of ALE & IPA beers are concentrated in 2 regions which make it a good candidate for k-NN modeling.