# PURDUE UNIVERSITY | Krannert School of Management

# Predicting Preference of Craft Beer Styles

**Ananth Srinath, Jake Winkle, Saben Fletcher, Derek Gaswirth, Tyler Meyer, Matthew Lanham**

Purdue University, Krannert School of Management

srinatha@purdue.edu; winkle0@purdue.edu; fletch29@purdue.edu; dgaswirt@purdue.edu; meyer288@purdue.edu; lanhamm@purdue.edu

## ABSTRACT

In this project, we evaluated three predictive models to determine the most accurate model for beer drinkers, sellers, and bars to use to better predict drink preferences. Based on misclassification rate, we evaluated K-Nearest Neighbors (KNN), K-Means Clustering with Decision Trees, and Decision Tree models without prior clustering to try and find the most accurate model to forecast stylistic craft beer preferences of 2,049 observations of historical data for a given consumer.

## INTRODUCTION

In the rapidly growing craft beer industry, many customers are lost in the sea of new options available to them. Even through detailed menu descriptions, beer drinkers face the difficulty of translating this information into taste and may order something they do not enjoy. If a restaurant or brewery knows the style of craft beer their customers enjoy most, they can predict what the next customer is likely to order. These forecasts save beer drinkers money by having options tailored to their most favorable styles of beer and help businesses to maintain happy customers. Our project addresses this issue by using past beer preferences to predict which new beers a given customer is likely to enjoy based on stylistic preferences.
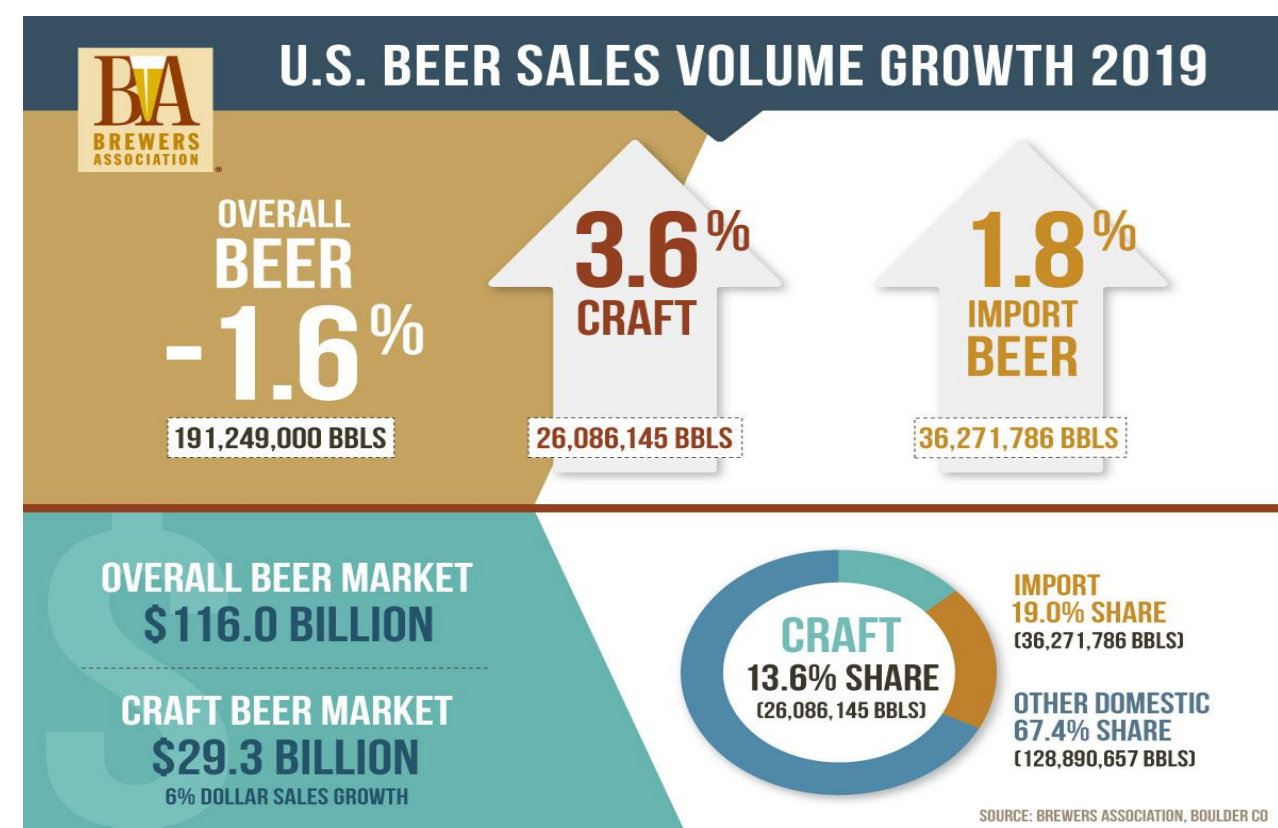


Fig 1. Comparison of U.S. Craft Beer Growth vs. Overall
Source: *Brewersassocation.org*

- What predictive model would be best in assisting sellers and bars in predicting similar preferences among beer users?
- How accurately can we predict customers' beer preferences?

## LITERATURE REVIEW

Similar studies to ours, that are referenced below, have previously analyzed craft beer preferences based on variable factors such as age segmentation, consumer behavior based on labeling and branding, socioeconomic status, and consumer personality traits. This differs from our study where we observed IBU and ABV levels as beer determinants while using, KNN, decision trees, and clustering.

| Study | Tukey | Anova | L&V model | Probit | H. Cluster | Euclidean |
|---|---|---|---|---|---|---|
| (2020) Jaeger | ✓ | ✓ | | | ✓ | ✓ |
| (2020) Higgens | | | ✓ | | | |
| (2014) Aquilani | | ✓ | | ✓ | ✓ | |
| (2020) Betancur | | | | | ✓ | |
| (2020) Jaeger | ✓ | | | | ✓ | ✓ |

## METHODOLOGY

**Raw Data**
- Raw data sourced from Kaggle competition (2,049 observations)

**Data Cleaning**
- Using the Beers dataset
- Standardization of IBU as well as ABV
- Correctly name each category
- Remove geographical data

**Data Preparation**
- One-hot encoded name and style
- Impute missing IBU/ABV data

**Data Partitioning**
- Train the Beers dataset (validation set cross-validation)

**Predictive Modeling**
- Need to decide which predictive model would work best deciding beer preference
- Three predictive models used are shown to the right

**Analysis and Conclusion**
- Assess the predictive models and decide which model is most accurate.
- Using misclassification rate to predict strength of predictive model

**Decision**
- Classification Trees had the lowest misclassification rate (.14%)

- 70% Train 30% Test
- Classification Trees
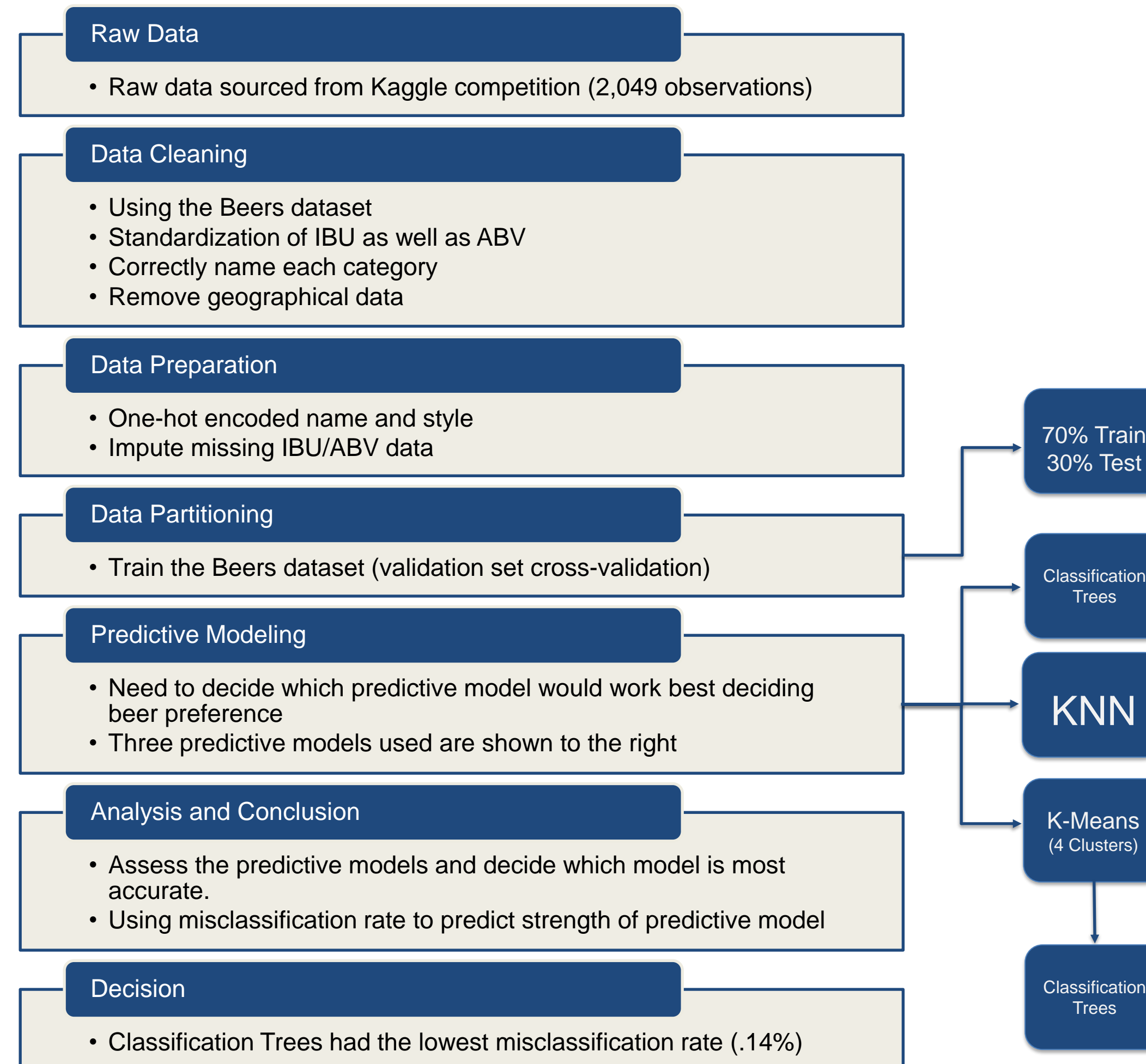- KNN
- K-Means (4 Clusters)
- Classification Trees

Fig 2. Study Design

## STATISTICAL RESULTS

After looking at the misclassification rates, we have decided that using Classification Trees would be the best option because the misclassification rate of the next best model is about 25% worse than Classification Trees.
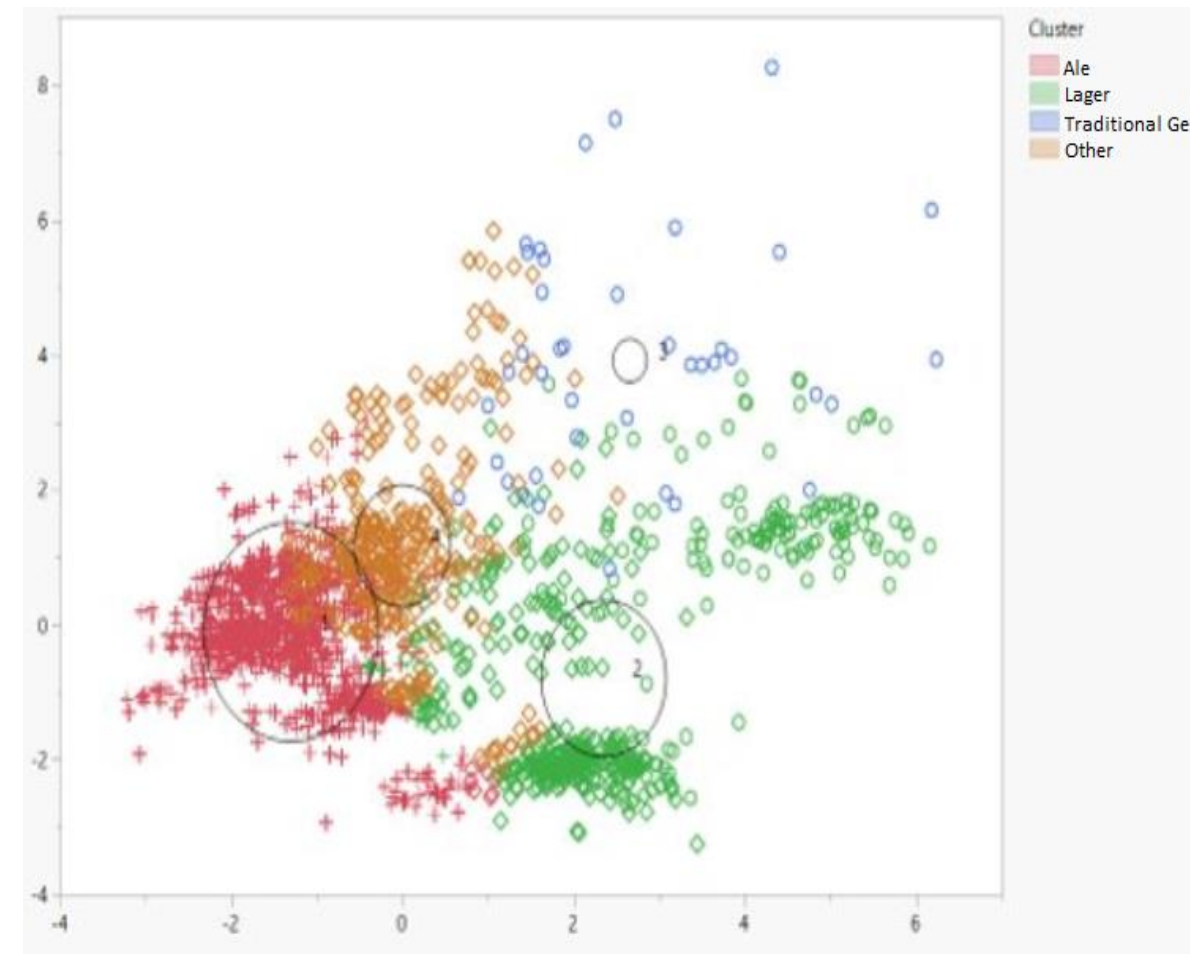


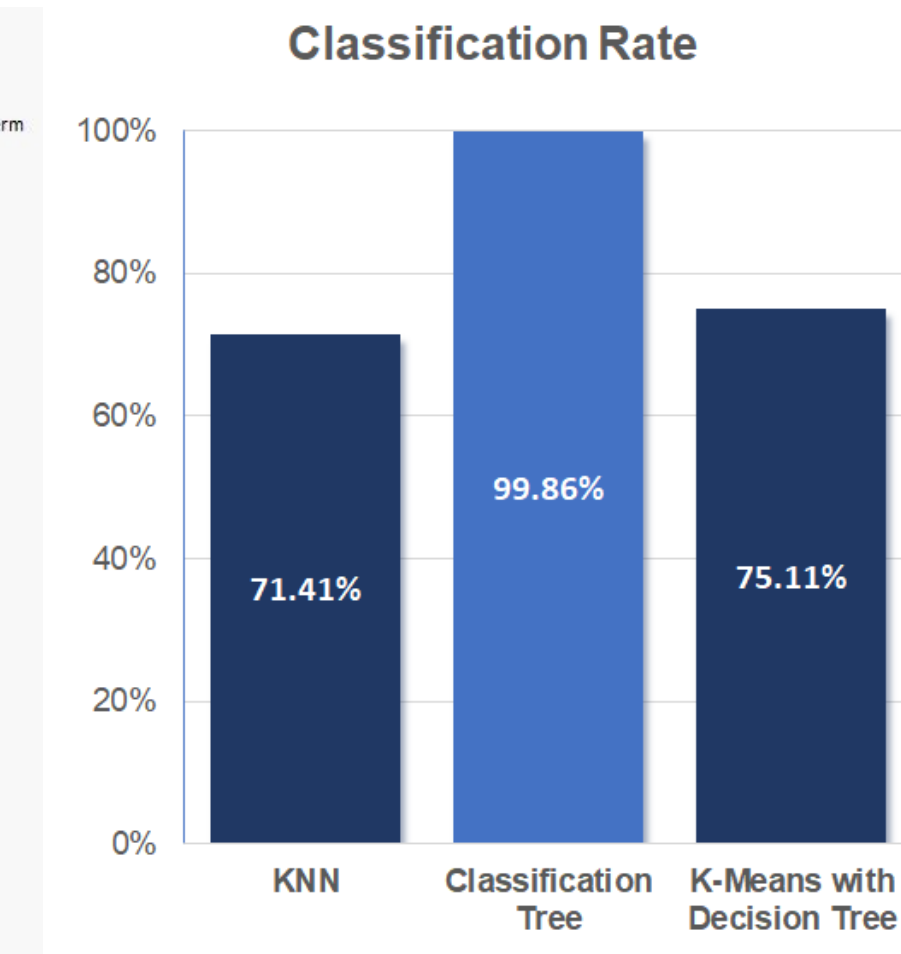Fig 3. Model Results for K-Means Clusters



Fig 4. Model Comparison

## EXPECTED BUSINESS IMPACT

From a corporate perspective, we can use this classification model to source beers to stock the shelves and allow beer consumers to identify similar tasting beers to source from craft breweries. Thus, helping expand the beer market and increasing profits of both the breweries and bars at the same time.

**Misclassification Rate: (FP + FN) / Total**

With the COVID-19 pandemic it has become more imperative to accurately forecast consumer trends within the craft beer industry. Craft beer sellers have had a decrease in sales by more than 50% throughout the country because of local restrictions. Our model can increase the efficiency of products produced per consumer by effectively forecasting overall consumer demand and restoring industry performance to normal levels.
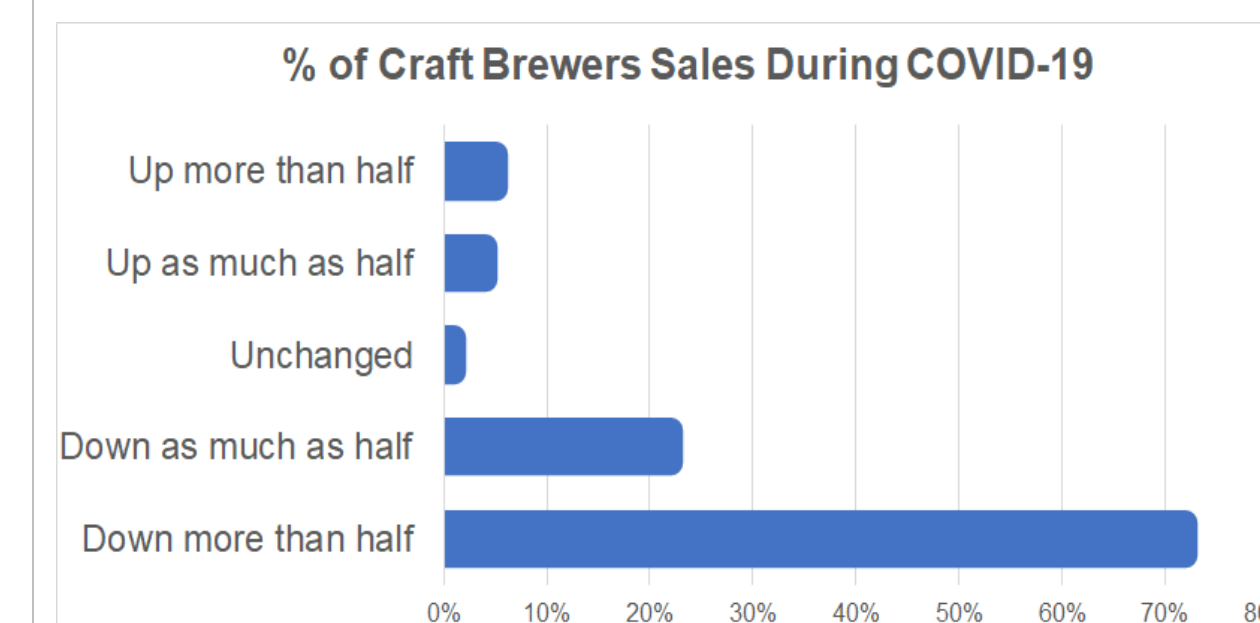


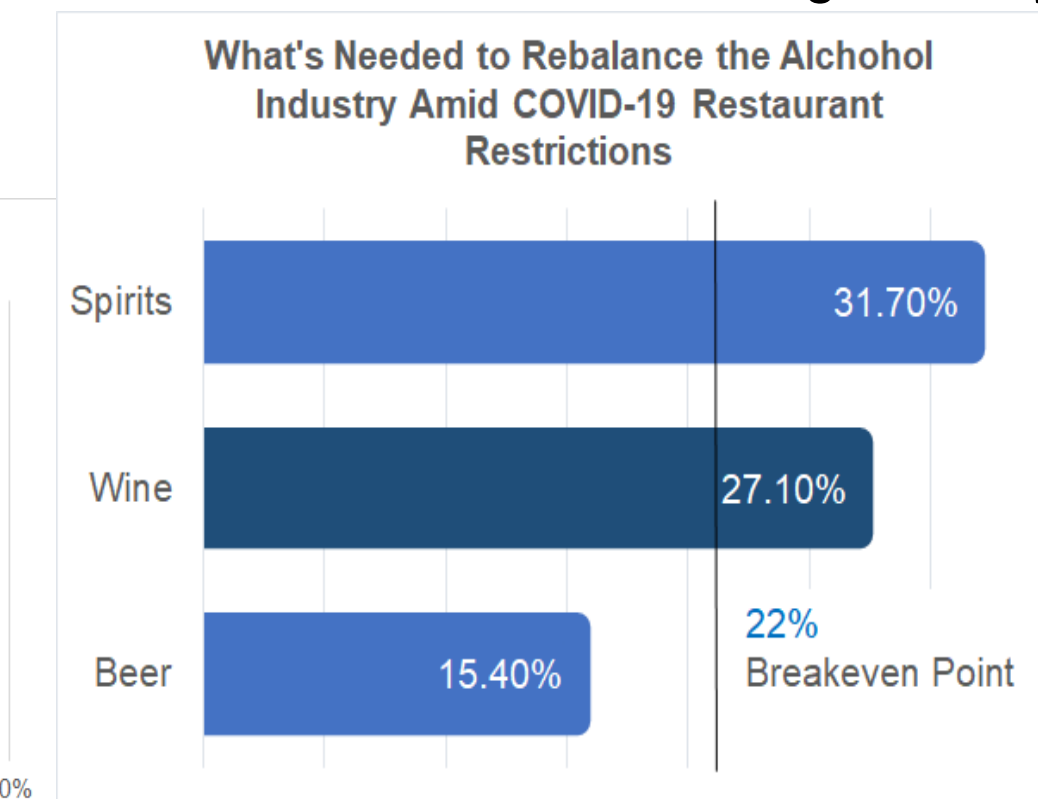Fig 5. Craft Brewers Sales During The Pandemic
Source: Brewers Association



Fig 6. Alcohol Industry Restrictions Impact
Source: Nielsen Retail MS

## CONCLUSIONS

**What predictive model would be best in assisting sellers and bars in predicting similar preferences among beer users?**
- Classification trees have the lowest misclassification rate.
- Pruning the decision tree and making less branches leads to higher misclassification, but simpler statistics.

**How accurately can we predict customers' beer preferences?**
- On a test set, we found that our model predicted the consumers beer preferences with over 99% accuracy, with a misclassification rate below 1%.
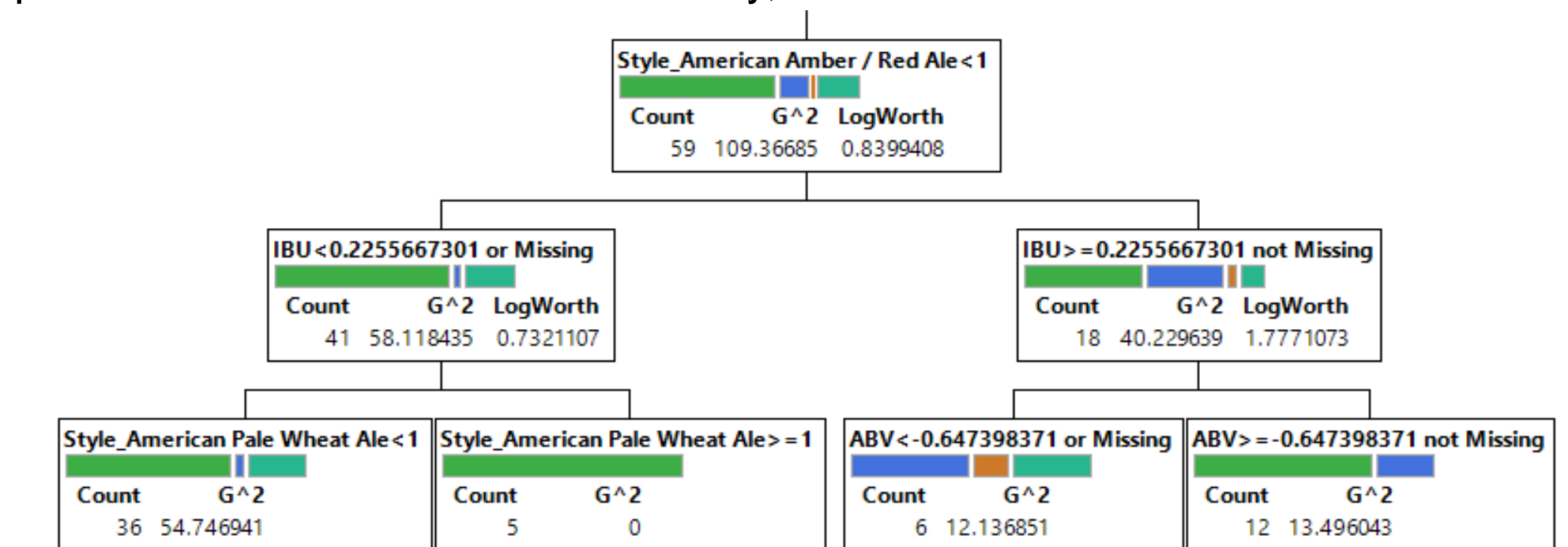


Fig 7. Portion of Our Classification Tree

## ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham, Xinyu Wang, Theo Ginting, and our graduate student mentors Gerardo Gutierrez and Laura Rossi for their guidance and support on this project.