Beers and Breweries Analysis

Anish Bhandari & Carolina Craus

2022-10-04

Introduction

The main objective of this analysis is to add value to Budweiser company through our exploratory data analysis. We have discovered several interesting opportunities throughout this study. We are positive that this analysis will enhance Budweiser's strategic decision making and open a few doors for future opportunities.

Datasets

1. Beers.csv: The Beers dataset contains 2,410 beers that are being produced in the United States along with metrics for ABV (Alcohol by Volume), IBU (International Bitterness Units), Style, and ounces and includes the Brewery ID associated with that beer.
2. Breweries.csv: The Breweries dataset contains 558 Breweries in the United States along with the Brewery ID (which matches with the one from the Beers dataset) and which city and state it is located in.

Preliminary Data Collection & Cleaning

```
# Importing and Cleaning the data
Beer <- read.csv("https://raw.githubusercontent.com/BivinSadler/MSDS_6306_Doing-Data-Science/Mas
ter/Unit%208%20and%209%20Case%20Study%201/Beers.csv",header = TRUE)

Brewery <- read.csv("https://raw.githubusercontent.com/BivinSadler/MSDS_6306_Doing-Data-Science/
Master/Unit%208%20and%209%20Case%20Study%201/Breweries.csv", header = TRUE)

colnames(Beer)[5]="Brew_ID"# Renamed Brewery_id to Brew_ID so that it has the same name in both
 the dataframe. This will help us merge by the brewery identification
colnames(Brewery)[2] = "Brewery_Name"

# Question # 2 Merge the beer data with brewery data
beerfinal <- merge(Beer,Brewery,by.x = "Brew_ID", by.y = "Brew_ID" ) # outer merged the 2 datafr
ames by brewery identification
beerfinal$State <- as.factor(beerfinal$State) # changed state into factor
beerfinal$Style <- as.factor(beerfinal$Style) # changed beer style into a factor so that it can
 be grouped
beerfinal$Brewery <- as.factor(beerfinal$Brewery_Name)
```

Load Libraries

```
library(usmap)
library(ggplot2)
library(dplyr) # pipe function
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(maps)
library(mapproj)
library(usdata)
library(stringi)
library(ggpubr)
```

Exploratory Data Analysis

1. Breweries by State

Every state has at least 2 brewery, ranging from 2 (Delaware and West Virginia) to 265 (Colorado).Washington DC also has one brewery which has been excluded from the heat map as DC is not a state. The median breweries across the state is 27.Regionally, West has the most breweries with Colorado, California and Oregon in top 6.Midwest is close second with Michigan, Indiana, Illinois, and Wisconsin in top 10.

```
#head(beerfinal)
#summary(beerfinal)
#colnames(beerfinal)
# In order to use the maps and usdata libraray, we need to remove one brewery which is in DC
grep("DC",beerfinal$State,ignore.case = TRUE)
```

```
## [1] 1286 1287 1288 1289 1290 1291 1292 1293
```

```
beerfinal[1286:1293,]
```

```
##      Brew_ID                      Name Beer_ID   ABV IBU
## 1286     228          Stone of Arbroath    2078 0.080  NA
## 1287     228              The Tradition    1809 0.050  15
## 1288     228             El Hefe Speaks    1263 0.053  11
## 1289     228        Penn Quarter Porter    1092 0.055  NA
## 1290     228 On the Wings of Armageddon     851 0.092 115
## 1291     228             The Corruption     186 0.065  80
## 1292     228                The Citizen     185 0.070  NA
## 1293     228                 The Public     184 0.060  NA
##                                Style Ounces           Brewery_Name       City
## 1286         Scotch Ale / Wee Heavy     12 DC Brau Brewing Company Washington
## 1287            American Blonde Ale     12 DC Brau Brewing Company Washington
## 1288                     Hefeweizen     12 DC Brau Brewing Company Washington
## 1289                American Porter     12 DC Brau Brewing Company Washington
## 1290 American Double / Imperial IPA     12 DC Brau Brewing Company Washington
## 1291                   American IPA     12 DC Brau Brewing Company Washington
## 1292               Belgian Pale Ale     12 DC Brau Brewing Company Washington
## 1293        American Pale Ale (APA)     12 DC Brau Brewing Company Washington
##      State                 Brewery
## 1286    DC DC Brau Brewing Company
## 1287    DC DC Brau Brewing Company
## 1288    DC DC Brau Brewing Company
## 1289    DC DC Brau Brewing Company
## 1290    DC DC Brau Brewing Company
## 1291    DC DC Brau Brewing Company
## 1292    DC DC Brau Brewing Company
## 1293    DC DC Brau Brewing Company
```

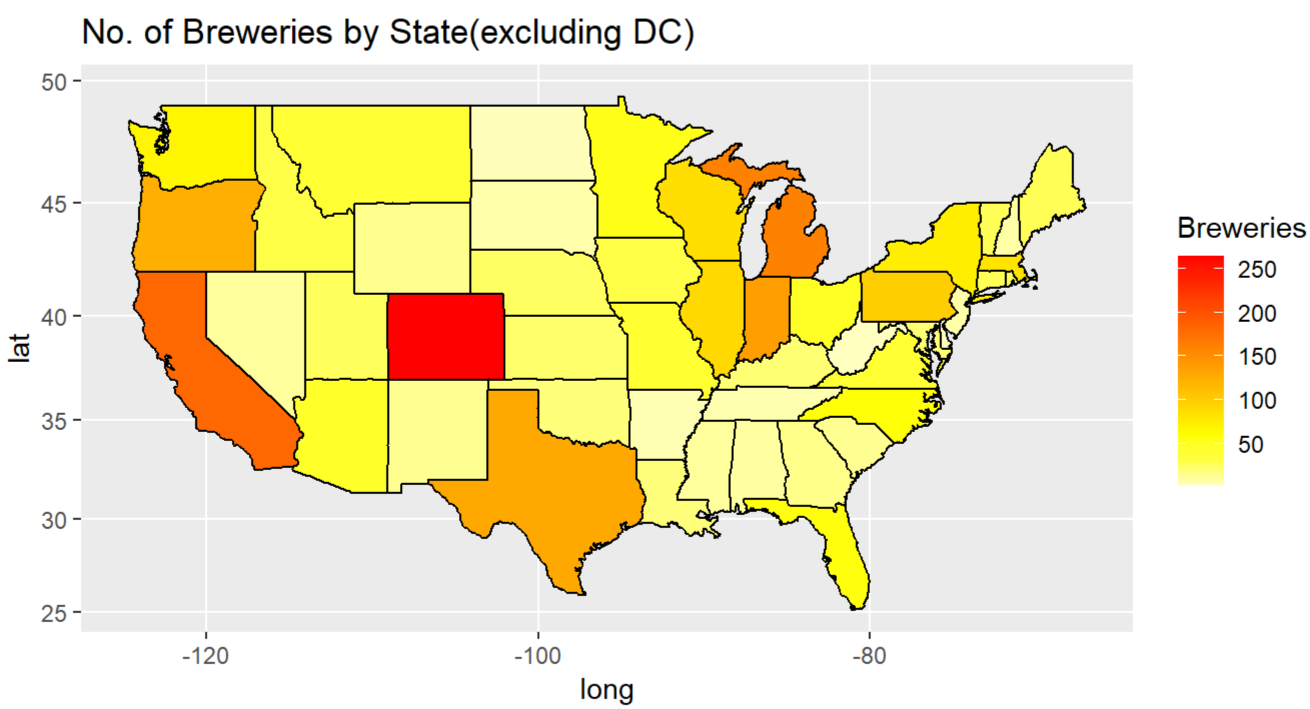
```
beerfinal_wodc = beerfinal[-(1286:1293),]
grep("DC",beerfinal_wodc$abb,ignore.case = TRUE)
```

```
## integer(0)
```

```
# remove 1 factor level of Dc to match the map
beerfinal_wodc$State <- factor(beerfinal_wodc$State, exclude = "DC") # remove 1 factor level of
 Dc to match the map
beerfinal_wodc$State = as.factor(beerfinal_wodc$State)


#makes a data frame with State name and abbreviation
lookup = data.frame(abb = state.abb, State = state.name)

# used stri_sub function to remove a "space" in front of the state abbreviations
beerfinal_wodc$State = stri_sub(beerfinal_wodc$State,-2)

#changed state to lower case so that it can be used to match the location(longitude,latitude)
beerfinal_wodc$State = tolower(abbr2state(beerfinal_wodc$State))

# ake one data set with state names and abb
beerheatData = count(beerfinal_wodc,State) #count up the occurrence of each state.


colnames(beerheatData)[2] = "Breweries" #change "n" to "Breweries"
colnames(beerheatData)[1] = "region" # changed the 1st column to "region"

# Arrange the States by Breweries
beerheatData1 = arrange(beerheatData,desc(Breweries))

#This shows the number of brewery excluding 1 in DC
beerheatData1
```

```
##            region Breweries
## 1        colorado       265
## 2      california       183
## 3        michigan       162
## 4         indiana       139
## 5           texas       130
## 6          oregon       125
## 7    pennsylvania       100
## 8        illinois        91
## 9       wisconsin        87
## 10  massachusetts        82
## 11       new york        74
## 12     washington        68
## 13 north carolina        59
## 14        florida        58
## 15      minnesota        55
## 16           ohio        49
## 17        arizona        47
## 18       missouri        42
## 19        montana        40
## 20       virginia        40
## 21          idaho        30
## 22           iowa        30
## 23    connecticut        27
## 24         hawaii        27
## 25          maine        27
## 26   rhode island        27
## 27        vermont        27
## 28           utah        26
## 29         alaska        25
## 30       nebraska        25
## 31         kansas        23
## 32       kentucky        21
## 33       maryland        21
## 34      louisiana        19
## 35       oklahoma        19
## 36        georgia        16
## 37        wyoming        15
## 38     new mexico        14
## 39 south carolina        14
## 40    mississippi        11
## 41         nevada        11
## 42        alabama        10
## 43  new hampshire         8
## 44     new jersey         8
## 45   south dakota         7
## 46      tennessee         6
## 47       arkansas         5
## 48   north dakota         3
## 49       delaware         2
## 50  west virginia         2
```

```
#summary(beerheatData1)

# Generating a heat map
states <- map_data("state")
map.df <- merge(states,beerheatData, by="region", all.x=T)
map.df <- map.df[order(map.df$order),]
ggplot(map.df, aes(x=long,y=lat,group=group))+
  geom_polygon(aes(fill=Breweries))+
  geom_path()+
  scale_fill_gradientn(colours=rev(heat.colors(10)),na.value="grey90")+ggtitle("No. of Breweries
by State(excluding DC)")+
  coord_map()
```



2. First 6 observations & last 6 observations of merged file

Merging was done in the first section. Observations are shown below.

```
head(beerfinal, n=6)
```

```
##   Brew_ID          Name Beer_ID   ABV IBU                               Style
## 1       1  Get Together    2692 0.045  50                         American IPA
## 2       1 Maggie's Leap    2691 0.049  26                   Milk / Sweet Stout
## 3       1    Wall's End    2690 0.048  19                    English Brown Ale
## 4       1       Pumpion    2689 0.060  38                          Pumpkin Ale
## 5       1    Stronghold    2688 0.060  25                      American Porter
## 6       1   Parapet ESB    2687 0.056  47 Extra Special / Strong Bitter (ESB)
##   Ounces       Brewery_Name        City State           Brewery
## 1     16 NorthGate Brewing  Minneapolis    MN NorthGate Brewing
## 2     16 NorthGate Brewing  Minneapolis    MN NorthGate Brewing
## 3     16 NorthGate Brewing  Minneapolis    MN NorthGate Brewing
## 4     16 NorthGate Brewing  Minneapolis    MN NorthGate Brewing
## 5     16 NorthGate Brewing  Minneapolis    MN NorthGate Brewing
## 6     16 NorthGate Brewing  Minneapolis    MN NorthGate Brewing
```

```
tail(beerfinal,n=6)
```

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

3. Missing Values

We can see that there are 62 missing values in ABV (2.57%) and 1005 in IBU (41.7%). We are not going to delete the missing values as we may have to do data analysis on individual properties. We will use filter na function to study any relations in future.

```
summary(beerfinal)
```

```
##     Brew_ID         Name              Beer_ID           ABV         
##  Min.   :  1.0   Length:2410        Min.   :   1.0   Min.   :0.00100  
##  1st Qu.: 94.0   Class :character   1st Qu.: 808.2   1st Qu.:0.05000  
##  Median :206.0   Mode  :character   Median :1453.5   Median :0.05600  
##  Mean   :232.7                      Mean   :1431.1   Mean   :0.05977  
##  3rd Qu.:367.0                      3rd Qu.:2075.8   3rd Qu.:0.06700  
##  Max.   :558.0                      Max.   :2692.0   Max.   :0.12800  
##                                                      NA's   :62       
##       IBU                                   Style         Ounces     
##  Min.   :  4.00   American IPA                 : 424   Min.   : 8.40  
##  1st Qu.: 21.00   American Pale Ale (APA)      : 245   1st Qu.:12.00  
##  Median : 35.00   American Amber / Red Ale     : 133   Median :12.00  
##  Mean   : 42.71   American Blonde Ale          : 108   Mean   :13.59  
##  3rd Qu.: 64.00   American Double / Imperial IPA: 105   3rd Qu.:16.00  
##  Max.   :138.00   American Pale Wheat Ale      :  97   Max.   :32.00  
##  NA's   :1005     (Other)                      :1298                  
##  Brewery_Name           City               State     
##  Length:2410        Length:2410        CO     : 265  
##  Class :character   Class :character   CA     : 183  
##  Mode  :character   Mode  :character   MI     : 162  
##                                        IN     : 139  
##                                        TX     : 130  
##                                        OR     : 125  
##                                       (Other):1406  
##                          Brewery    
##  Brewery Vivant            :  62  
##  Oskar Blues Brewery       :  46  
##  Sun King Brewing Company  :  38  
##  Cigar City Brewing Company:  25  
##  Sixpoint Craft Ales       :  24  
##  Hopworks Urban Brewery    :  23  
##  (Other)                   :2192
```

```
#colSums(is.na(beerfinal))
```

```
colMeans(is.na(beerfinal)) * 100
```

```
##      Brew_ID         Name      Beer_ID          ABV          IBU        Style 
##     0.000000     0.000000     0.000000     2.572614    41.701245     0.000000 
##       Ounces Brewery_Name         City        State      Brewery 
##     0.000000     0.000000     0.000000     0.000000     0.000000
```

4. Median Alcohol & Bitterness

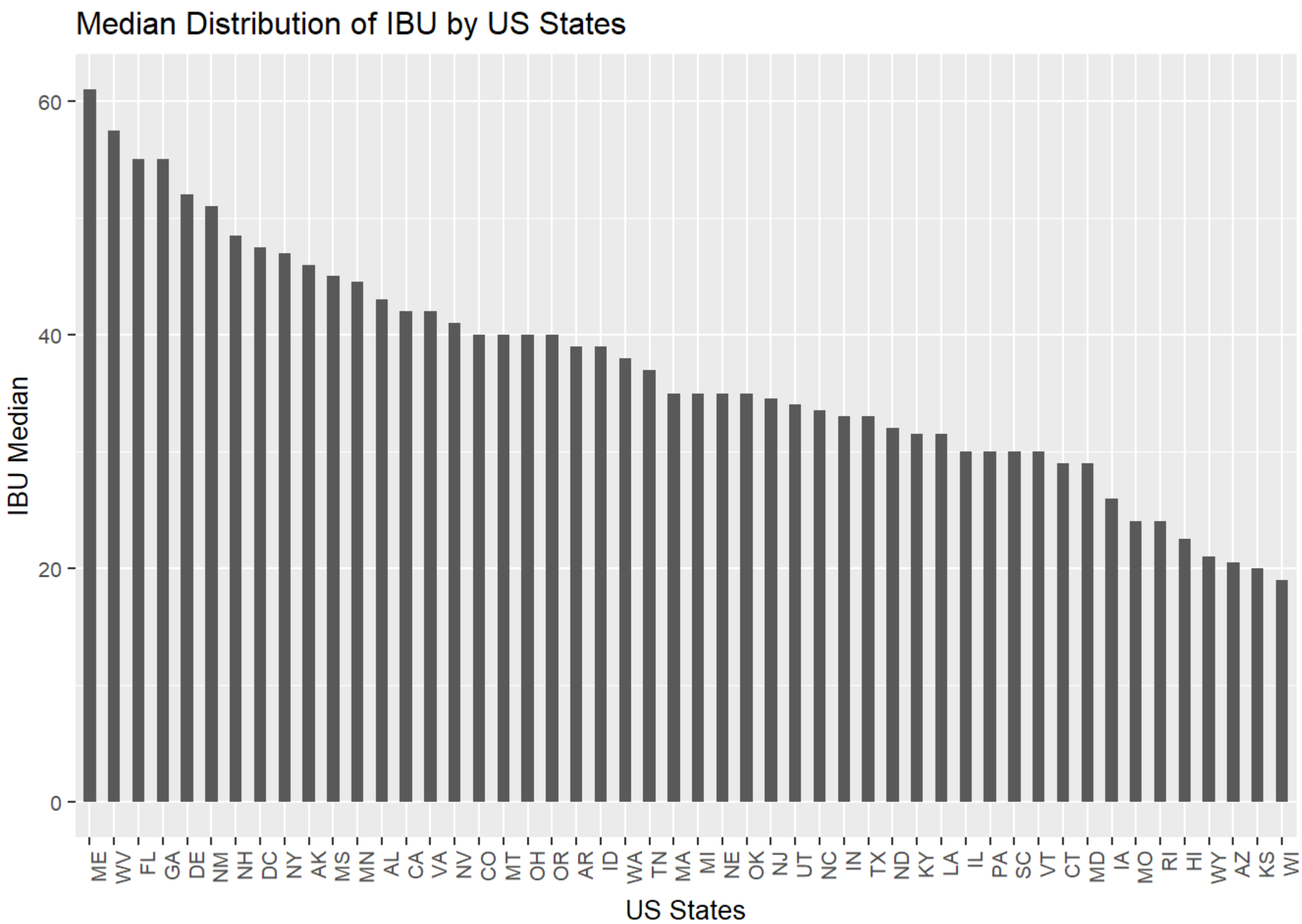
The median IBU is highest in Maine(61) followed by West Virginia(57.5) and Florida(55). The lowest median IBU is in Wisconsin (19) followed by Kansas(20), and Arizona(20.5). The median highest ABV is in Washington DC and Kentucky(6.25%) closely followed by a 3 way tie between Michigan, New Mexico, and West Virginia(6.2%). The lowest median ABV is in Utah (4%) followed by New Jersey(4.6%).The low ABV in Utah is due to the state regulation, where only 4% ABV was allowed in grocery stores. However, this law changed in 2019 and currently grocery stores in Utah can sell up-to 5% ABV.

```
# 4 Median Alcohol and Bitterness Unit by State
#calculating Median for IBU
sumdata = beerfinal %>% filter(!is.na(beerfinal$IBU)) %>% group_by(State)%>% summarise(Mean=mean
(IBU), Max=max(IBU), Min=min(IBU), Median=median(IBU), Std=sd(IBU))
sumdata1 = arrange(sumdata,desc(Median))
print (sumdata1, n=51)
```

```
## # A tibble: 50 × 6
##    State  Mean   Max   Min Median   Std
##    <fct> <dbl> <int> <int>  <dbl> <dbl>
##  1 " ME"  52.9    70    28   61    17.4
##  2 " WV"  57.5    71    44   57.5  19.1
##  3 " FL"  46.8    82    10   55    22.5
##  4 " GA"  46.3    65    17   55    16.2
##  5 " DE"  52      52    52   52    NA
##  6 " NM"  57     100    15   51    36.7
##  7 " NH"  48.5    82    15   48.5  47.4
##  8 " DC"  55.2   115    11   47.5  50.9
##  9 " NY"  46     111     7   47    23.4
## 10 " AK"  40.9    71    10   46    23.3
## 11 " MS"  46.5    80    18   45    22.3
## 12 " MN"  50.0   120    10   44.5  26.0
## 13 " AL"  51.2   103    25   43    24.7
## 14 " CA"  46.3   115     4   42    27.4
## 15 " VA"  45.4   135    12   42    26.9
## 16 " NV"  46.5    90    15   41    28.2
## 17 " CO"  47.4   104     9   40    26.1
## 18 " MT"  41.7    80    11   40    20.5
## 19 " OH"  44.2   126    11   40    24.8
## 20 " OR"  47.9   138    13   40    28.1
## 21 " AR"  39      39    39   39    NA
## 22 " ID"  55.1   100     9   39    35.5
## 23 " WA"  45.0    83    18   38    19.9
## 24 " TN"  41.6    61    23   37    15.9
## 25 " MA"  38     130     7   35    24.6
## 26 " MI"  36.7   115     6   35    24.8
## 27 " NE"  30.7    65    10   35    17.2
## 28 " OK"  40.7   100     9   35    31.1
## 29 " NJ"  46.4   100     9   34.5  31.9
## 30 " UT"  45.5    83    10   34    27.8
## 31 " NC"  43.3    98     5   33.5  25.3
## 32 " IN"  43.0   115     8   33    25.1
## 33 " TX"  40.4   118     5   33    26.3
## 34 " ND"  40.3    70    19   32    26.5
## 35 " KY"  40.7    80    13   31.5  23.9
## 36 " LA"  33      60    13   31.5  16.4
## 37 " IL"  41.5   100     8   30    27.4
## 38 " PA"  42.4   113     8   30    29.2
## 39 " SC"  30.2    65     5   30    22.8
## 40 " VT"  42.3   120     8   30    34.4
## 41 " CT"  40.8    85     6   29    36.1
## 42 " MD"  36.8    90    10   29    24.1
## 43 " IA"  33.2    99     5   26    21.1
## 44 " MO"  32.5    89     7   24    22.9
## 45 " RI"  31.6    75    10   24    17.9
## 46 " HI"  32.7    75    12   22.5  22.0
## 47 " WY"  32.1    75    15   21    20.4
## 48 " AZ"  35.2    99     9   20.5  25.7
```

```
## 49 " KS"  36.7   110    12   20    30.0
## 50 " WI"  26.5    80     7   19    20.7
```

```
# Median Distribution of IBU by US States
sumdata %>% ggplot(aes(x=reorder(State,-Median), y=Median)) + geom_bar(stat = "identity",width = 0.5) + xlab("US States") + ylab("IBU Median") + ggtitle("Median Distribution of IBU by US States") + theme(text = element_text(size=10),axis.text.x = element_text(angle=90, hjust=1))
```

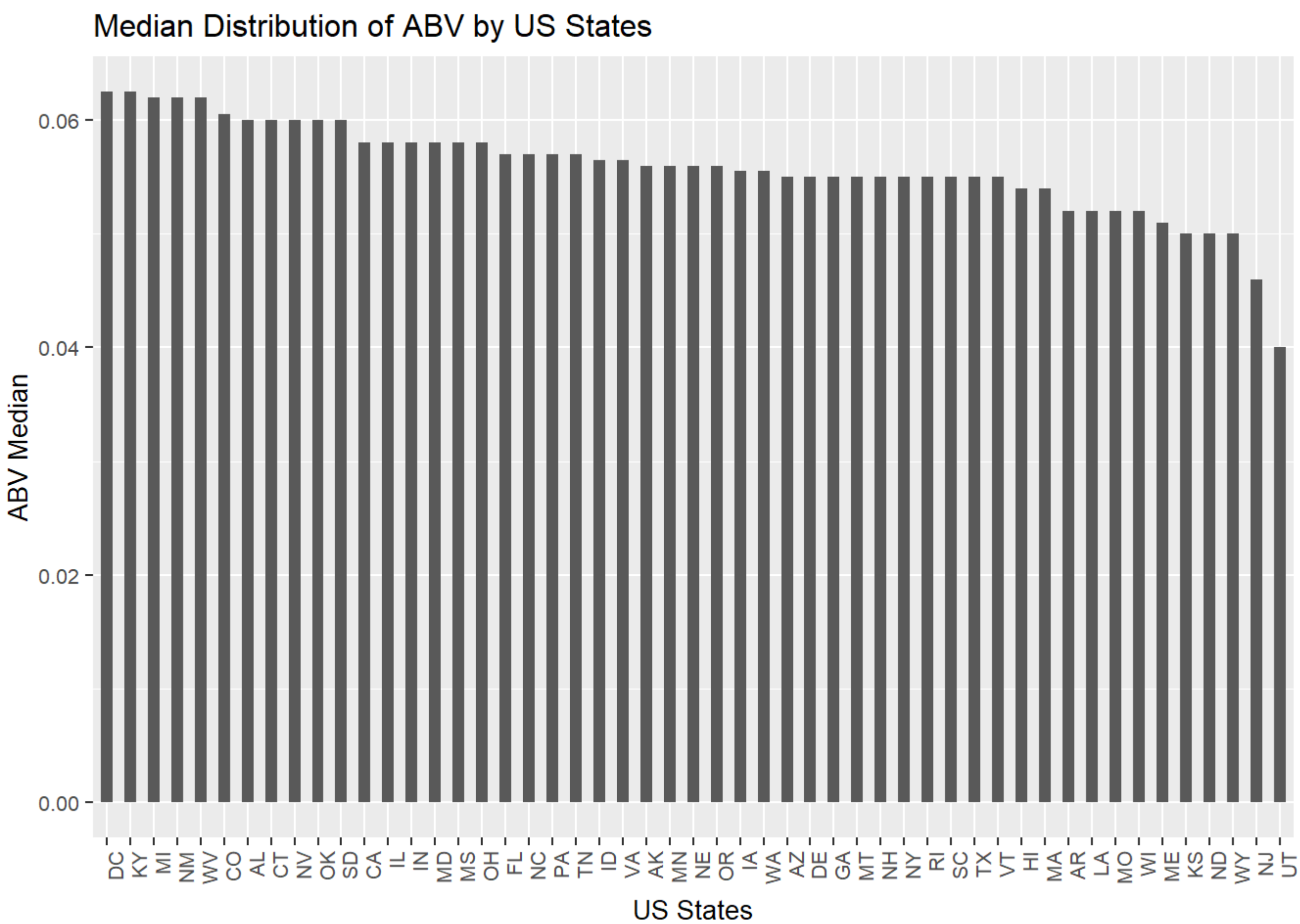



```
#calculating Median for ABV
sumdataabv = beerfinal %>% filter(!is.na(beerfinal$ABV)) %>% group_by(State)%>% summarise(Mean=mean(ABV), Max=max(ABV), Min=min(ABV), Median=median(ABV), Std=sd(ABV))
sumdataabv = arrange(sumdataabv,desc(Median))
print(sumdataabv, n=51)
```

```
## # A tibble: 51 × 6
##    State   Mean   Max   Min Median      Std
##    <fct>  <dbl> <dbl> <dbl>  <dbl>    <dbl>
##  1 " DC" 0.0656 0.092 0.05  0.0625  0.0145
##  2 " KY" 0.0646 0.125 0.04  0.0625  0.0202
##  3 " MI" 0.0634 0.099 0.038 0.062   0.0138
##  4 " NM" 0.0611 0.08  0.045 0.062   0.0100
##  5 " WV" 0.062  0.067 0.057 0.062   0.00707
##  6 " CO" 0.0634 0.128 0.041 0.0605  0.0146
##  7 " AL" 0.062  0.093 0.05  0.06    0.0124
##  8 " CT" 0.0611 0.09  0.034 0.06    0.0161
##  9 " NV" 0.0669 0.099 0.05  0.06    0.0184
## 10 " OK" 0.0595 0.085 0.04  0.06    0.0131
## 11 " SD" 0.0593 0.069 0.045 0.06    0.00830
## 12 " CA" 0.0611 0.099 0.001 0.058   0.0152
## 13 " IL" 0.0620 0.096 0.04  0.058   0.0141
## 14 " IN" 0.0634 0.12  0.04  0.058   0.0149
## 15 " MD" 0.0590 0.085 0.042 0.058   0.0115
## 16 " MS" 0.0589 0.08  0.038 0.058   0.0131
## 17 " OH" 0.0620 0.099 0.043 0.058   0.0135
## 18 " FL" 0.0599 0.082 0.038 0.057   0.0102
## 19 " NC" 0.0600 0.099 0.032 0.057   0.0130
## 20 " PA" 0.0601 0.099 0.032 0.057   0.0133
## 21 " TN" 0.0552 0.062 0.045 0.057   0.00655
## 22 " ID" 0.0607 0.099 0.042 0.0565  0.0157
## 23 " VA" 0.0570 0.088 0.04  0.0565  0.0106
## 24 " AK" 0.0556 0.068 0.048 0.056   0.00569
## 25 " MN" 0.0602 0.099 0.04  0.056   0.0144
## 26 " NE" 0.0580 0.096 0.042 0.056   0.0129
## 27 " OR" 0.0571 0.088 0.027 0.056   0.0118
## 28 " IA" 0.0595 0.095 0.048 0.0555  0.0122
## 29 " WA" 0.0578 0.084 0.04  0.0555  0.0107
## 30 " AZ" 0.0602 0.095 0.042 0.055   0.0124
## 31 " DE" 0.055  0.055 0.055 0.055  NA
## 32 " GA" 0.0564 0.072 0.045 0.055   0.00790
## 33 " MT" 0.0565 0.075 0.045 0.055   0.00780
## 34 " NH" 0.0524 0.065 0.028 0.055   0.0107
## 35 " NY" 0.0572 0.1   0.027 0.055   0.0144
## 36 " RI" 0.0570 0.086 0.037 0.055   0.0125
## 37 " SC" 0.0607 0.097 0.04  0.055   0.0175
## 38 " TX" 0.0598 0.099 0.04  0.055   0.0133
## 39 " VT" 0.0604 0.096 0.04  0.055   0.0156
## 40 " HI" 0.0573 0.083 0.042 0.054   0.0113
## 41 " MA" 0.0557 0.099 0.035 0.054   0.0111
## 42 " AR" 0.052  0.061 0.04  0.052   0.00797
## 43 " LA" 0.0555 0.088 0.039 0.052   0.0124
## 44 " MO" 0.0548 0.08  0.035 0.052   0.00979
## 45 " WI" 0.0541 0.099 0.035 0.052   0.0102
## 46 " ME" 0.0578 0.099 0.035 0.051   0.0153
## 47 " KS" 0.0561 0.085 0.044 0.05    0.0125
## 48 " ND" 0.054  0.067 0.045 0.05    0.0115
## 49 " WY" 0.0549 0.072 0.046 0.05    0.00884
```

```
## 50 " NJ" 0.0574 0.099 0.039 0.046   0.0222
## 51 " UT" 0.0519 0.09  0.04  0.04    0.0165
```

```
# Median Distribution of ABV by US States

sumdataabv %>% ggplot(aes(x=reorder(State,-Median), y=Median)) + geom_bar(stat = "identity",widt
h =0.5) + xlab("US States") + ylab("ABV Median") + ggtitle("Median Distribution of ABV by US Sta
tes") + theme(text = element_text(size=10),axis.text.x = element_text(angle=90, hjust=1))
```



5. Maximum IBU & ABV

The maximum ABV(12.8%) is Lee Hill Series Vol.5 beer from Upslope Brewing Company in Boulder, Colorado. The maximum IBU (138) is Bitter Bitch Imperial beer from Astoria Brewing Company located in Astoria, Oregon.

```
sumdata2 = arrange(sumdata,desc(Max))
head(sumdata2)
```

```
## # A tibble: 6 × 6
##   State  Mean   Max   Min Median   Std
##   <fct> <dbl> <int> <int>  <dbl> <dbl>
## 1 " OR"  47.9   138    13   40    28.1
## 2 " VA"  45.4   135    12   42    26.9
## 3 " MA"  38     130     7   35    24.6
## 4 " OH"  44.2   126    11   40    24.8
## 5 " MN"  50.0   120    10   44.5  26.0
## 6 " VT"  42.3   120     8   30    34.4
```

```
max_abv = beerfinal[which.max(beerfinal$IBU),]
max_abv
```

```
##      Brew_ID                      Name Beer_ID   ABV IBU
## 1857     375 Bitter Bitch Imperial IPA     980 0.082 138
##                                Style Ounces            Brewery_Name    City
## 1857 American Double / Imperial IPA     12 Astoria Brewing Company Astoria
##      State                 Brewery
## 1857    OR Astoria Brewing Company
```

```
sumdataabv2 = arrange(sumdataabv,desc(Max))
head(sumdataabv2)
```

```
## # A tibble: 6 × 6
##   State   Mean   Max   Min Median    Std
##   <fct>  <dbl> <dbl> <dbl>  <dbl>  <dbl>
## 1 " CO" 0.0634 0.128 0.041 0.0605 0.0146
## 2 " KY" 0.0646 0.125 0.04  0.0625 0.0202
## 3 " IN" 0.0634 0.12  0.04  0.058  0.0149
## 4 " NY" 0.0572 0.1   0.027 0.055  0.0144
## 5 " MI" 0.0634 0.099 0.038 0.062  0.0138
## 6 " NV" 0.0669 0.099 0.05  0.06   0.0184
```

```
most_bitter = beerfinal[which.max(beerfinal$ABV),]
most_bitter
```

```
##     Brew_ID                                                 Name Beer_ID   ABV
## 375      52 Lee Hill Series Vol. 5 - Belgian Style Quadrupel Ale    2565 0.128
##     IBU            Style Ounces            Brewery_Name    City State
## 375  NA Quadrupel (Quad)   19.2 Upslope Brewing Company Boulder    CO
##                     Brewery
## 375 Upslope Brewing Company
```

6. Summary of ABV

The ABV distribution is right skewed with the tail around 7 to 10 %. There are number of factors which causes this skeweness. Some states require higher alcohol beer to only be sold in liquor stores. This causes breweries to restrict alcohol percent by stopping the fermentation early.Another factor is the cost associated with higher alcohol. Higher alcohol, in general, costs higher to produce due to the higher fermentation time and extra raw materials.

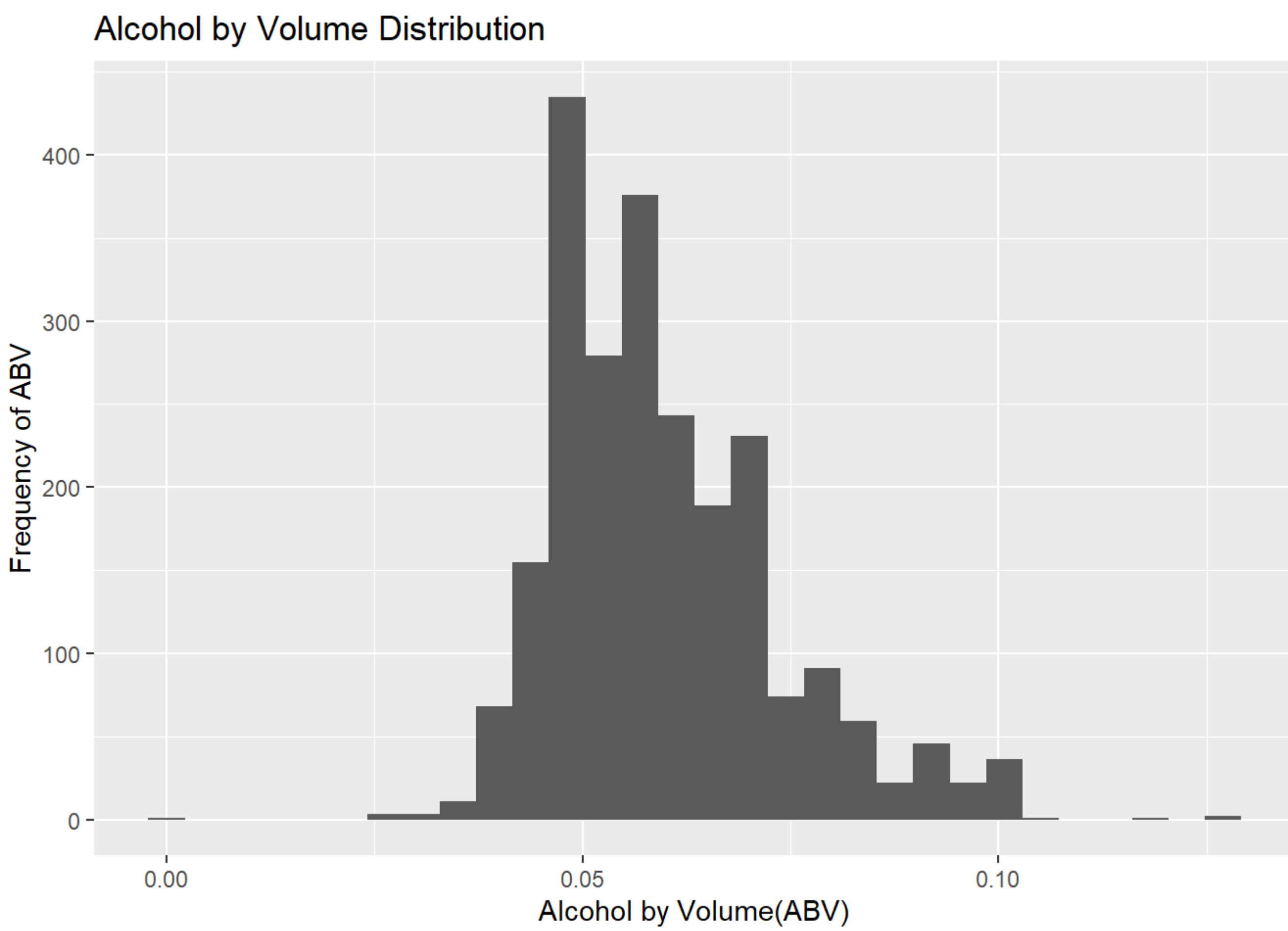
```
# 6 Summary of ABV

summary(beerfinal$ABV)
```

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

```
beerfinal %>% ggplot(aes(x=ABV)) + geom_histogram()+ xlab("Alcohol by Volume(ABV)") + ylab("Freq
uency of ABV")+ ggtitle("Alcohol by Volume Distribution")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 62 rows containing non-finite values (stat_bin).
```

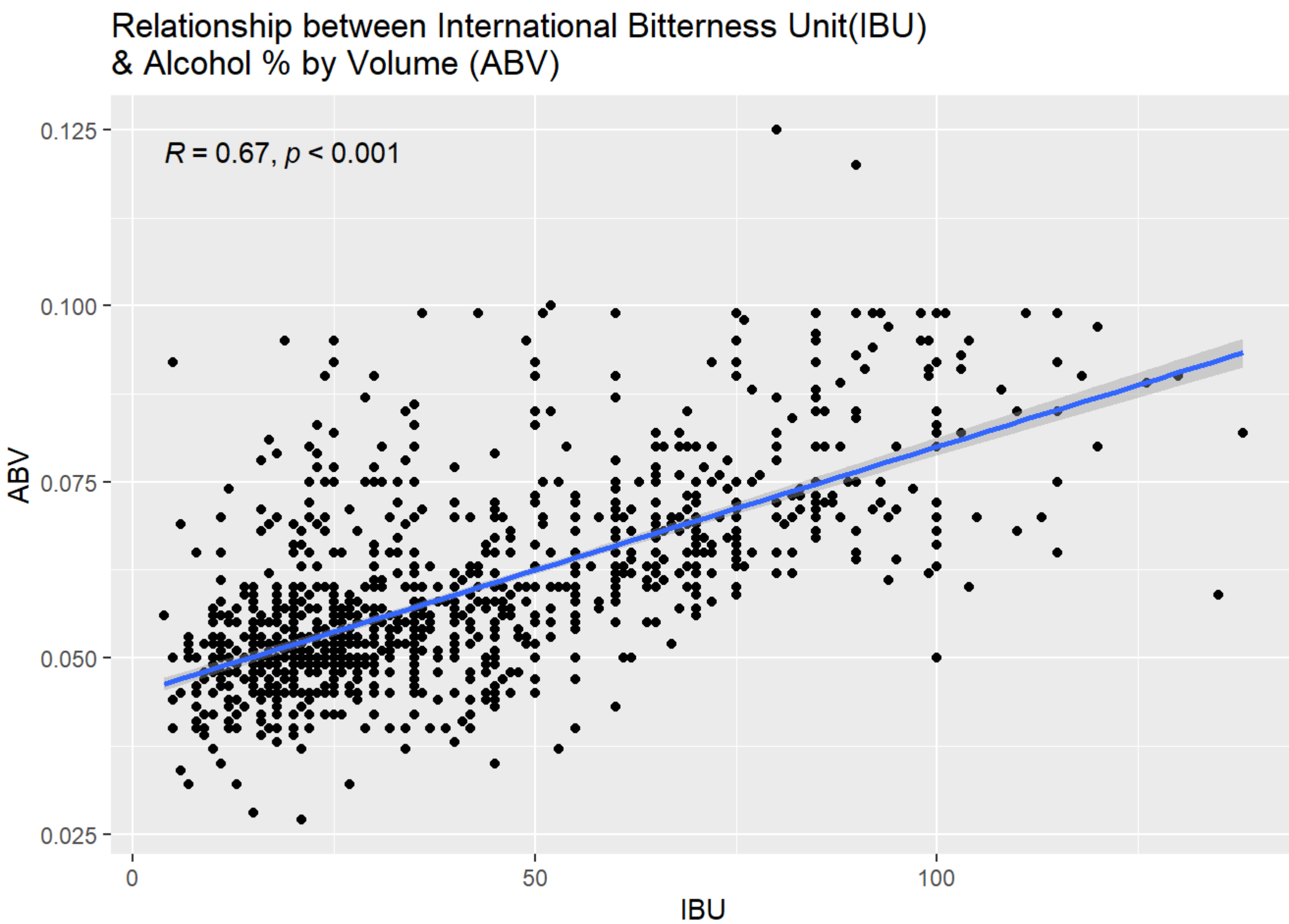


7. Relationship between ABV & IBU

The correlation coefficient between IBU and ABV is 0.67. This evidence suggests that there is a positive correlation between IBU and ABV. A correlation below and over the median IBU(35) was also run. The relationship gets worse at lower IBU (IBU <35, R=0.25) and comparable when IBU is higher (IBU < 35, R = 0.64). There is an opportunity to optimize the relationship based on the style of beer.

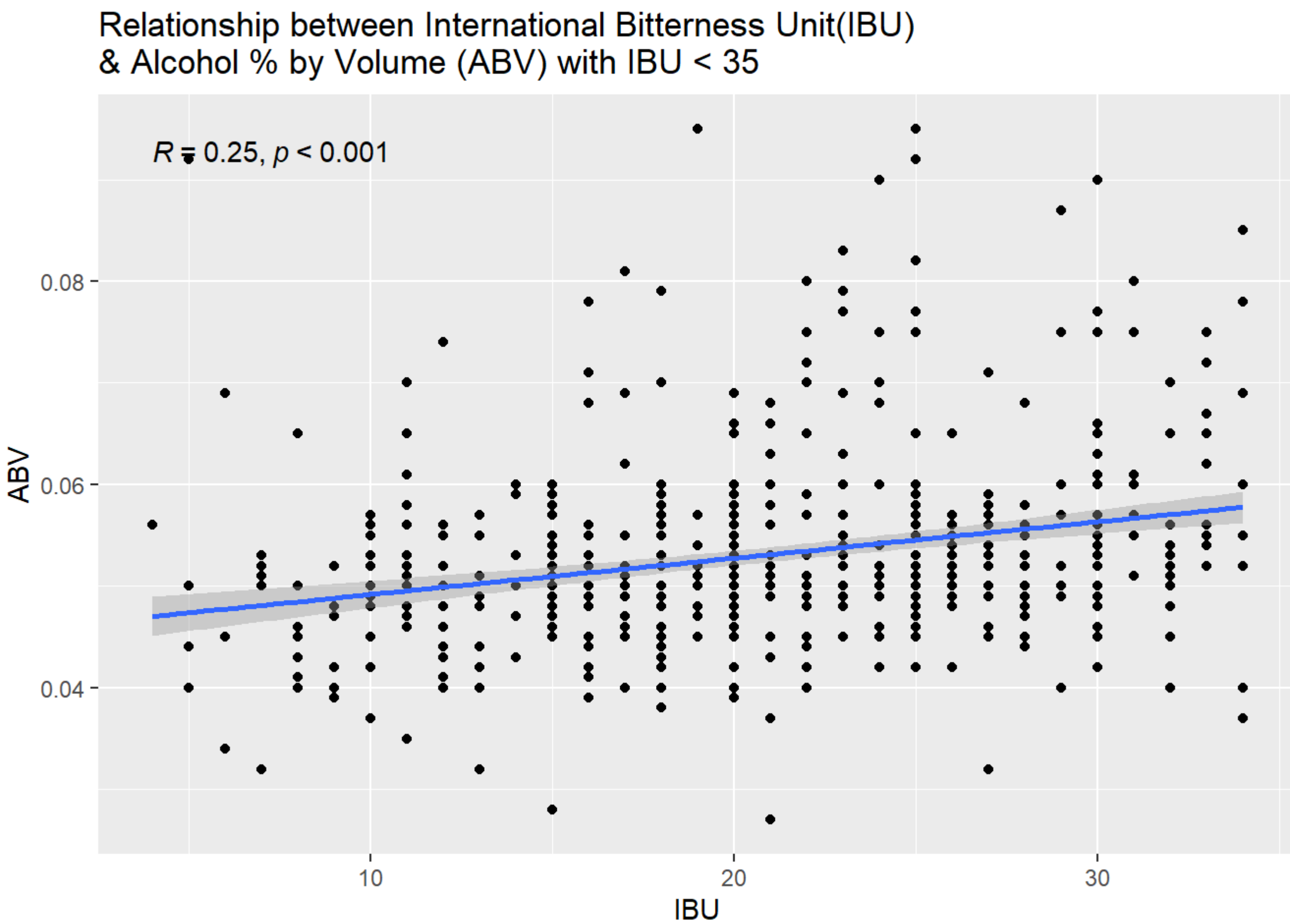
```
# 7 Relationship between ABV and IBU using Scatterplot
beerfinal %>% filter(!is.na(beerfinal$ABV) & !is.na(beerfinal$IBU)) %>% ggplot(aes(x=IBU, y=AB
V))+ geom_point() + geom_smooth(method="lm") + stat_cor(p.accuracy = 0.001) + ggtitle("Relations
hip between International Bitterness Unit(IBU) \n& Alcohol % by Volume (ABV)")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



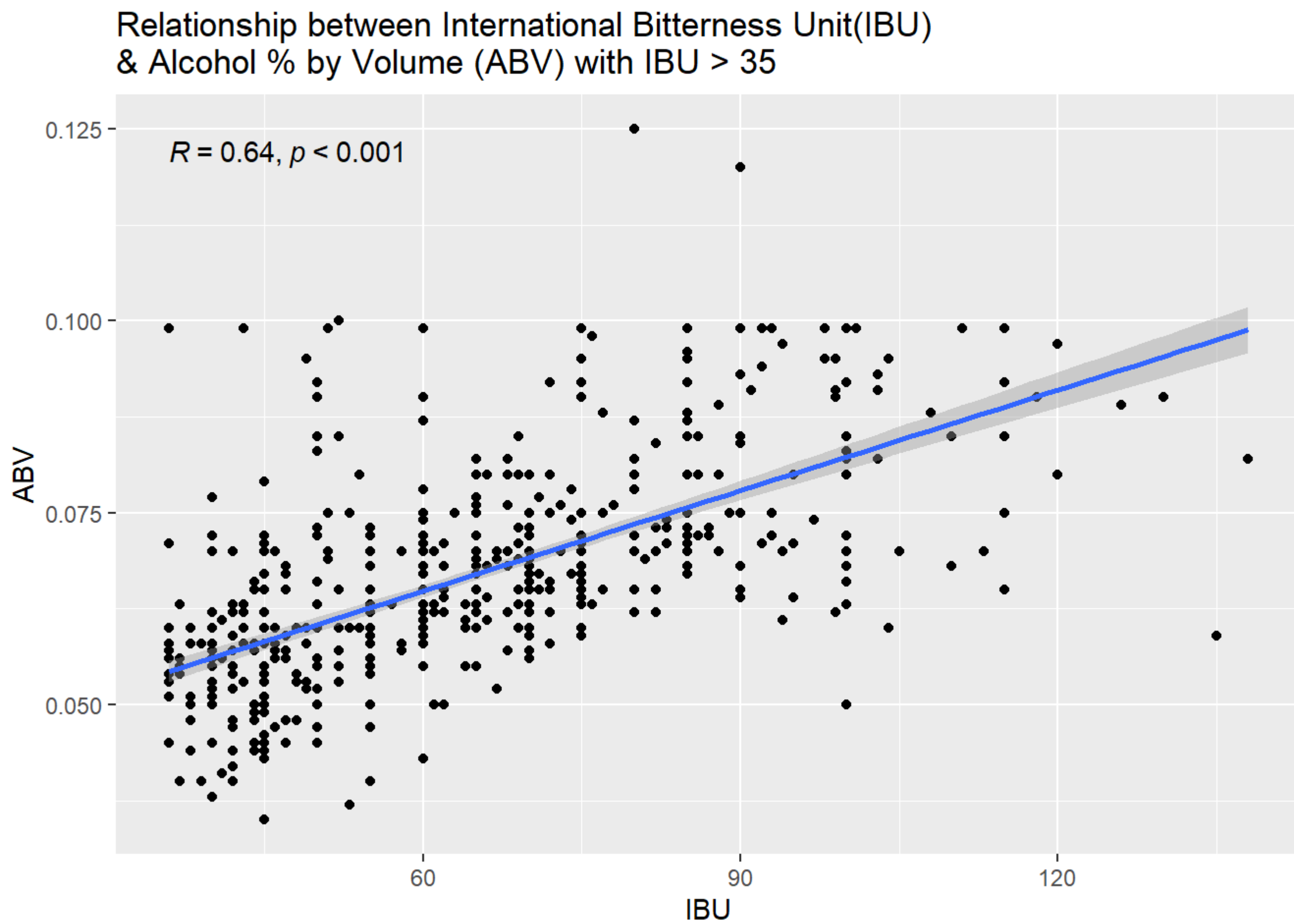

```
# Relationship between ABV and IBU using Scatterplot with IBU < & > 35 (Median IBU)
beerfinal %>% filter(!is.na(beerfinal$ABV) & !is.na(beerfinal$IBU) & IBU < 35) %>% ggplot(aes(x=
IBU, y=ABV))+ geom_point() + geom_smooth(method="lm") + stat_cor(p.accuracy = 0.001) + ggtitle(
"Relationship between International Bitterness Unit(IBU) \n& Alcohol % by Volume (ABV) with IBU
 < 35 ")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
beerfinal %>% filter(!is.na(beerfinal$ABV) & !is.na(beerfinal$IBU) & IBU > 35) %>% ggplot(aes(x=
IBU, y=ABV))+ geom_point() + geom_smooth(method="lm") + stat_cor(p.accuracy = 0.001) + ggtitle(
"Relationship between International Bitterness Unit(IBU) \n& Alcohol % by Volume (ABV) with IBU
 > 35 ")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



8. Classification of IPA & ALE Using KNN

Filter & Clean data

```
library(dplyr)
library(caret)
```

```
## Loading required package: lattice
```

```
library(class)
# filtering IPA & Ale style beers only
both_df  <- dplyr::filter(beerfinal, grepl('IPA|Ale', beerfinal$Style))
sum(is.na(both_df))
```

```
## [1] 632
```

```
# normalize beer style -- add type column for classification
both_df$Type <- if_else(grepl('IPA', both_df$Style), "IPA", "ALE")
head(both_df)
```

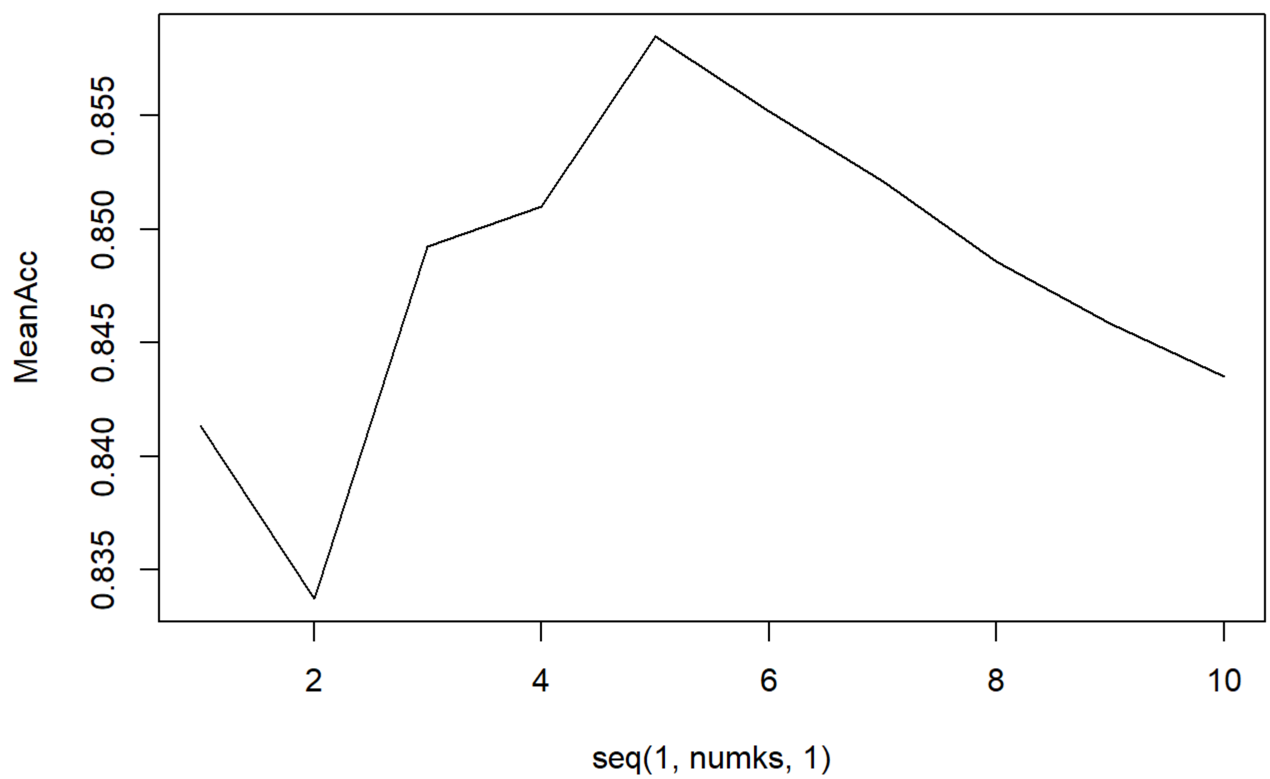
```
##   Brew_ID           Name Beer_ID   ABV IBU                           Style
## 1       1   Get Together    2692 0.045  50                    American IPA
## 2       1     Wall's End    2690 0.048  19               English Brown Ale
## 3       1        Pumpion    2689 0.060  38                     Pumpkin Ale
## 4       2 Citra Ass Down    2686 0.080  68 American Double / Imperial IPA
## 5       2         A Beer    2683 0.042  42         American Pale Ale (APA)
## 6       2  Flesh Gourd'n    2681 0.066  21                     Pumpkin Ale
##   Ounces              Brewery_Name        City State                Brewery
## 1     16         NorthGate Brewing  Minneapolis    MN        NorthGate Brewing
## 2     16         NorthGate Brewing  Minneapolis    MN        NorthGate Brewing
## 3     16         NorthGate Brewing  Minneapolis    MN        NorthGate Brewing
## 4     16 Against the Grain Brewery  Louisville    KY Against the Grain Brewery
## 5     16 Against the Grain Brewery  Louisville    KY Against the Grain Brewery
## 6     16 Against the Grain Brewery  Louisville    KY Against the Grain Brewery
##   Type
## 1  IPA
## 2  ALE
## 3  ALE
## 4  IPA
## 5  ALE
## 6  ALE
```

```
# remove missing values
both_clean <- remove_missing(both_df)
```

```
## Warning: Removed 590 rows containing missing values.
```

KNN Model

```r
# split of 70/30 for training and test sets
set.seed(6)
splitPerc = .70
iterations = 100
numks = 10

# randomly split the train and test sets


masterAcc = matrix(nrow = iterations, ncol=numks)

# using only IBU & ABV values & Style as the class against which knn will search
# iterate through 100 values of k to hypertune the k parameter
for (j in 1:iterations) {
  trainIndices = sample(1:dim(both_clean)[1], round(splitPerc * dim(both_clean)[1]))
  train = both_clean[trainIndices,]
  test = both_clean[-trainIndices,]
  for (i in 1:numks) {
    classifications = knn(train[,c(4,5)], test[,c(4,5)], train$Type, prob=TRUE, k=i)
    table(classifications, test$Type)
    CM = confusionMatrix(table(classifications,test$Type)) # confusion matrix
    masterAcc[j, i] = CM$overall[1]
  }
}

MeanAcc = colMeans(masterAcc)


# plot the Mean knn value
plot(seq(1, numks, 1), MeanAcc, type="l")
```




```
which.max(MeanAcc)
```

```
## [1] 5
```

```
max(MeanAcc)
```

```
## [1] 0.8584806
```

Using knn classification to explore the relationship between ABV and IBU and test whether we can accurately classify a beer as an IPA or an ALE using these two metrics. Iterating through values of k from a range of 1 to 10, we can see from a visual inspection of the plot that the most optimal value for k is 5. This tells us we will be looking at the 5 nearest neighbors ABV & IBU values of a specified beer in order to determine its classification.

Confusion Matrix

```
CM
```

```
## Confusion Matrix and Statistics
##
##
## classifications ALE IPA
##             ALE 147  26
##             IPA  13  97
##
##                Accuracy : 0.8622
##                  95% CI : (0.8165, 0.9001)
##     No Information Rate : 0.5654
##     P-Value [Acc > NIR] : < 2e-16
##
##                   Kappa : 0.7161
##
##  Mcnemar's Test P-Value : 0.05466
##
##             Sensitivity : 0.9187
##             Specificity : 0.7886
##          Pos Pred Value : 0.8497
##          Neg Pred Value : 0.8818
##              Prevalence : 0.5654
##          Detection Rate : 0.5194
##    Detection Prevalence : 0.6113
##       Balanced Accuracy : 0.8537
##
##        'Positive' Class : ALE
##
```

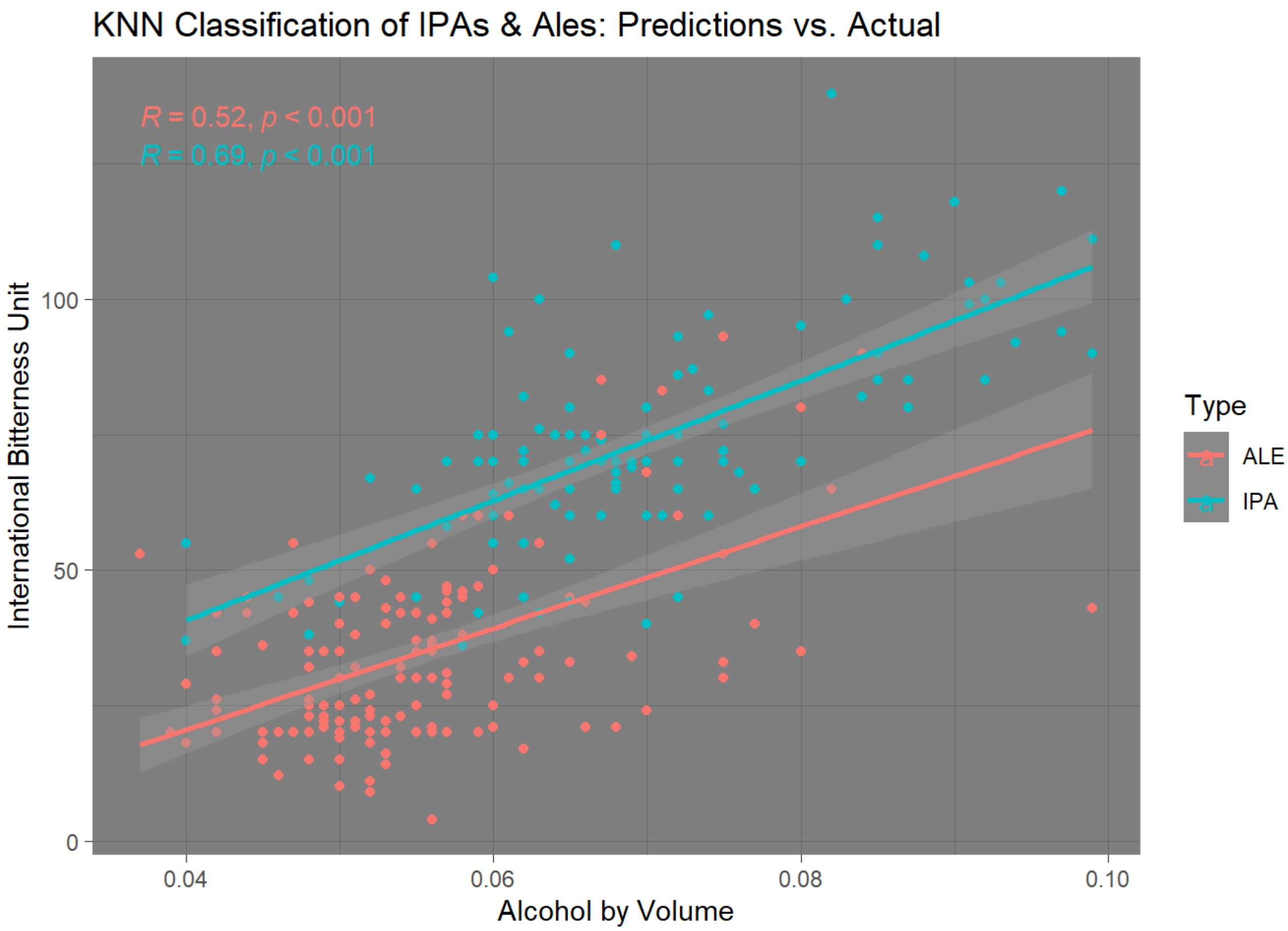
The confusion matrix and model statistics above is calculated from our predictions using the k-nn classifier with k=5 and tells us our model had an Accuracy of 86% in predicting the correct classification of beer, which is adequate for a classification model, but there are opportunities for optimization we can explore. The confusion matrix also tells us that 147 Ales were predicted correctly and 13 ales were predicted incorrectly (91.9% accuracy for Ales) and 97 IPAs were predicted correctly and 26 IPAs incorrectly (86.2% accuracy of IPAs).

Scatterplot of KNN Results

The scatter plot of the predicted values of Ale & IPA based on their IBU & ABV shows us that there is a strong relationship between IBU & ABV and the classification of Ale or IPA. As we can see, our prediction lines for both Ales and IPas follows their respective data points well. As can be seen from the plot, the Ale Threshold for IBU is below around 50 units and for ABV it is below around 6%. The IPA Threshold for IBU is above around 60 units and for ABV it above around 6%.

```
# scatter plot of the predicted values of ALE & IPA based on their IBU & ABV
test  %>% filter(!is.na(test$ABV) & !is.na(test$IBU))%>% ggplot(aes(x = ABV, IBU,color = Type))
 + geom_point() + geom_smooth(method = "lm") + theme_dark() + xlab("Alcohol by Volume") + ylab(
"International Bitterness Unit")+ stat_cor(p.accuracy = 0.001) + ggtitle("KNN Classification of
 IPAs & Ales: Predictions vs. Actual")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



9. Focus on IPA and Acquisition

Background

Budweiser's parent company is Anheuser-Busch. Anheuser-Busch owns various popular commercial brands such as Budweiser, Bud Light, Michelob Ultra, Natural Light etc. However, most are unaware of their acquisition of many microbreweries within US. Most of the bigger breweries such Anheuser-Busch, Miller, Coors etc have been strategically purchasing shares in microbreweries over the last 20 years or so, as popularity in microbreweries has grown. IPA is one of the most popular styles of beer in US. With this analysis, we will propose future microbrewery acquisition opportunities to the Budweiser's CEO and CFO, based on IPA's ABV and IBU.

Potential Future Acquisition

Data collected from https://www.anheuser-busch.com/brands/ (https://www.anheuser-busch.com/brands/)

We have collected most IPA brands from microbreweries owned by Anheuser-Busch. Our median ABV from this dataset is 6.9 % and IBU is 50. For comparison, the median ABV for IPA from our brewery data is 6.8% and IBU is 70. For potential acquisition, we filtered IPA with similar hop profile (picked 45 to 55 IBU with median 50 in the middle), slightly lower ABV (6.2 to median 6.9%, to find an efficient and cost effective beer). We came across 5 breweries using this filter. However, there are only 2 breweries that are not in the states that Anheuser-Busch operates. This is important for the new market penetration. The 2 breweries are Big Wood Brewery in MN and Abita Brewing Company in LA. We believe that this analysis can supplement any future acquisition decision by Anheuser-Busch.

```
#data manually transferred from website into a csv file
AB_IPA = read.csv(file.choose(), header = TRUE)
head(AB_IPA)
```

```
##                         Company State Name.of.Beer..IPA. IPA_ABV IPA_IBU
## 1         10 Barrel Brewing Co    OR     Apocalypse IPA     6.8      70
## 2         10 Barrel Brewing Co    OR       Nature Calls     6.5      55
## 3 Appalachian Mountain Brewery    NC   Low and Hazy IPA     4.0      29
## 4 Appalachian Mountain Brewery    NC      Rain Drop IPA     6.7      65
## 5 Appalachian Mountain Brewery    NC      LONG LEAF IPA     7.1      74
## 6         Breckenridge Brewery    CO         JUICE DROP     7.0      60
```

```
summary(AB_IPA)
```

```
##    Company             State           Name.of.Beer..IPA.    IPA_ABV
##  Length:49          Length:49          Length:49          Min.   :0.050
##  Class :character   Class :character   Class :character   1st Qu.:6.500
##  Mode  :character   Mode  :character   Mode  :character   Median :6.900
##                                                           Mean   :6.879
##                                                           3rd Qu.:7.500
##                                                           Max.   :9.900
##
##     IPA_IBU
##  Min.   :20.00
##  1st Qu.:41.50
##  Median :50.00
##  Mean   :49.53
##  3rd Qu.:57.75
##  Max.   :85.00
##  NA's   :13
```

```
### ABV and IBU(IPA only) from the Breweries dataset

IPA_only <- dplyr::filter(both_clean, grepl('IPA', both_clean$Type))
summary(IPA_only)
```

```
##     Brew_ID          Name              Beer_ID           ABV         
##  Min.   :  1.0   Length:392         Min.   :   4.0   Min.   :0.03800  
##  1st Qu.: 99.0   Class :character   1st Qu.: 805.8   1st Qu.:0.06200  
##  Median :210.0   Mode  :character   Median :1558.5   Median :0.06800  
##  Mean   :223.4                      Mean   :1461.0   Mean   :0.06914  
##  3rd Qu.:337.2                      3rd Qu.:2039.5   3rd Qu.:0.07500  
##  Max.   :547.0                      Max.   :2692.0   Max.   :0.09900  
##                                                                       
##       IBU                                    Style         Ounces     
##  Min.   : 30.00   American IPA                  :301   Min.   :12.00  
##  1st Qu.: 60.00   American Double / Imperial IPA: 75   1st Qu.:12.00  
##  Median : 70.00   English India Pale Ale (IPA)  :  7   Median :12.00  
##  Mean   : 71.95   American White IPA            :  6   Mean   :13.57  
##  3rd Qu.: 85.00   Belgian IPA                   :  3   3rd Qu.:16.00  
##  Max.   :138.00                                 :  0   Max.   :32.00  
##                   (Other)                       :  0                  
##  Brewery_Name           City               State    
##  Length:392         Length:392         CA     : 49  
##  Class :character   Class :character   CO     : 37  
##  Mode  :character   Mode  :character   OR     : 25  
##                                        IN     : 22  
##                                        MA     : 18  
##                                        TX     : 18  
##                                       (Other):223  
##                         Brewery        Type          
##  Cigar City Brewing Company:  9   Length:392        
##  Golden Road Brewing       :  9   Class :character  
##  Oskar Blues Brewery       :  9   Mode  :character  
##  Sun King Brewing Company  :  7                     
##  Sixpoint Craft Ales       :  6                     
##  Two Beers Brewing Company :  6                     
##  (Other)                   :346                     
```

```
valuebeer  <- IPA_only %>% filter(IBU > 45 & IBU < 55 &  ABV > 0.062 & ABV < 0.069)
valuebeer
```

```
##   Brew_ID                           Name Beer_ID   ABV IBU
## 1     161                   Hop Knot IPA     358 0.067  47
## 2     384 Northern Lights India Pale Ale    1205 0.065  52
## 3     384 Northern Lights India Pale Ale     368 0.065  52
## 4     438            East India Pale Ale    1279 0.068  47
## 5     438            East India Pale Ale     566 0.068  47
## 6     445                  Bark Bite IPA     985 0.066  50
## 7     534                    Jockamo IPA     516 0.065  52
##                          Style Ounces            Brewery_Name
## 1                 American IPA     12 Four Peaks Brewing Company
## 2                 American IPA     16         Starr Hill Brewery
## 3                 American IPA     12         Starr Hill Brewery
## 4 English India Pale Ale (IPA)     16           Brooklyn Brewery
## 5 English India Pale Ale (IPA)     12           Brooklyn Brewery
## 6                 American IPA     16           Big Wood Brewery
## 7                 American IPA     12      Abita Brewing Company
##              City State                    Brewery Type
## 1           Tempe    AZ Four Peaks Brewing Company  IPA
## 2          Crozet    VA         Starr Hill Brewery  IPA
## 3          Crozet    VA         Starr Hill Brewery  IPA
## 4        Brooklyn    NY           Brooklyn Brewery  IPA
## 5        Brooklyn    NY           Brooklyn Brewery  IPA
## 6 Vadnais Heights    MN           Big Wood Brewery  IPA
## 7   Abita Springs    LA      Abita Brewing Company  IPA
```

Conclusion

Based on this dataset, there is evidence to suggest that there is a linear positive relationship between Bitterness Unit and Alcohol. There's also evidence to suggest that this relationship differs across different type of beer. There is a big opportunity to optimize Bitterness Unit and Alcohol at the Budweiser breweries across US. This optimization can be enhanced further through knn model, which showed evidence of high accuracy in determining type of beer (IPA & Ale). The biggest opportunity for Anheuser-Busch and Budweiser is to create a model for micro brewery acquisition. We realize that there are multiple parameters important for acquisition and most of them are financial. The financial inputs can be paired alongside beer properties to create a strategic model for Anheuser-Busch's brewery acquisition.