Regional Liquor Sales in Des Moines, Iowa

Christophe Hunt[*,a], Senthil Dhanapal[a], Yadu Chittampalli[a]

[a] *CUNY School of Professional Studies, Data Analytics, New York, NY*

Abstract

This is the abstract.

It consists of two paragraphs

Keywords: Liquor Sales, Naive Forecast*

Problem

The objective of this report is to create a statistical model for the volume sold of liquor in gallons and the profit dollars in the City of Des Moines which is within the state of Iowa. This can help us make informed decisions on inventory production, sales, and assist wholesale distributors to plan for the predicted volume of distribution.

Introduction

In February, the Distilled Spirits Council (DISCUS), announced that spirits had an estimated retail sales of nearly $72 billion in 2015. Additionally, DISCUS credits the continuous growth of the distilled spirits industry to several key factors - continuous fascination with American Whiskeys in the United States and abroad, innovations in flavors, permutation across all spirits categories leading to consumer interest, improved regulatory and tax environment resulting in expanded market access and a relatively low number of state tax threats, and the growth of small distillers, which expanded grassroots and overall interest in the spirits category Del Buono (2016).

This establishes that spirit sales in the Unites States is a valuable market worth exploring for a more detailed and statistical understanding of sales and volume. We hope to more thoroughly understand what impact specific store sights may have accounting for the seasonal impact in November that might effect liquor sales. We will limit the analysis to the City of Des Moines for only whisky sales in the month of November. In 2000 the State of Iowa reported sales at a record pace during the last half of 2000 Boshart (2001). The later part of the year has an increase in sales so planning to meet capacity is a suitable goal for any company. Our years of interest for this analysis will be the month of November for 2015 and 2016.

Research Background (Literature Review)

Our goal is inventory prediction.

[*] Authors

Email addresses: christophe.hunt@spsmail.cuny.edu (Christophe Hunt), senthil.dhanapal@spsmail.cuny.edu (Senthil Dhanapal), yadu.chittampalli@spsmail.cuny.edu (Yadu Chittampalli)

Methodology

The initial data set contained many variables and is sufficiently large as it includes sales by location and each individual invoice since 2012. The reason for the large size of the initial data set is due to it including every liquor transaction from 2012 to present in Iowa, so it approaches 2.68 GB. For the purposes of this analysis, to analyze a data set this large is not feasible. Therefore, we reduced the number of variables and summarized to a regional aggregate.

Additionally, we looked into the top 10 liquor categories for each year by number of bottles sold. In 2015, the top categories were American Cocktails, Blended Whiskies, Canadian Whiskies, Imported Vodka, Puerto Rico & Virgin Islands Rum, Spiced Rum, Straight Bourbon Whiskies, Tequila, Vodka 80 Proof, and Whiskey Liqueur. Interestingly straight bourbon appears to have more sales in 2015 than 2014 which coincides with the literature of strong growing whiskey sales for every whiskey segment (Anonymous (2016)). We decided to focus on whiskey do to its strong sales and growing interest in the US.

Furthermore, we looked into the sales by `County` for Iowa

So we reduced the data set to a subset of variables as follows; `Vendor Name`, `Pack` (pack size of bottles sold) `Bottle Volume`, `State Bottle Cost`, `State Bottle Retail`, `Sales Dollars` (Total sales), and our dependent variables are `Volume sold in Gallons` and `Profit Dollars`.

We initially attempted to model for a dependent variable of `Volume Sold in Gallons` by `County`, however, the distribution of this variable becomes over-dispersed and negatively skewed when aggregating the data set to a manageable size and therefore we were unsuccessful in modeling this variable across regions. Our hope was that if we could model for the gallons sold by region we could more accurately predict our planned inventory and anticipate production goals. We were more successful in limiting our regional analysis to one city and the stores within that city.

In Des Moines in Iowa, we want to predict bottles sold of whiskey (as a whiskey company) in the holiday month of November, the stores (storeID) and we have Vendor Name, Pack (in case pack size influences sales), bottle volume (people looking for deals) state bottle cost (how much to procure), state bottle retail (how much we sell for), Sale Dollars (total sales) Volume (gallons), profit margin (calculate for the business problem).

We then sought to model the mean price per bottle, by modeling the price per bottle we can make pricing decisions based on region and the volume we plan to sell.

discuss the key aspects of your problem, data set and regression model(s). Given that you are working on real-world data, explain at a high-level your exploratory data analysis, how you prepared the data for regression modeling, your process for building regression models, and your model selection.

Experimentation and Results

Model 1

Forward step

```
## Start:  AIC=4270.71
## Bottles.Sold ~ 1
##
##                          Df Sum of Sq      RSS    AIC
## + Volume.Sold..Gallons.   1  31263820   777424 2874.4
## + Sale..Dollars.          1  29649270  2391974 3297.0
## + State.Bottle.Cost       1  17914640 14126604 3964.8
## + State.Bottle.Retail     1  17910359 14130885 3964.9
## + Store.Number           72  15883862 16157382 4157.3
## + Category.Name           7   1729576 30311668 4263.8
```

```
## + Bottle.Volume..ml.     1    191055 31850189 4270.5
## <none>                               32041244 4270.7
## + Pack                   1     12595 32028649 4272.6
##
## Step:  AIC=2874.44
## Bottles.Sold ~ Volume.Sold..Gallons.
##
##                       Df Sum of Sq    RSS    AIC
## + Category.Name        7    120760 656663 2825.0
## + Pack                 1     85890 691534 2832.4
## + State.Bottle.Retail  1     67111 710313 2842.5
## + State.Bottle.Cost    1     67106 710318 2842.5
## + Bottle.Volume..ml.   1     52773 724651 2850.0
## + Sale..Dollars.       1     35590 741834 2858.8
## <none>                             777424 2874.4
## + Store.Number        72    228710 548714 2887.4
##
## Step:  AIC=2824.97
## Bottles.Sold ~ Volume.Sold..Gallons. + Category.Name
##
##                       Df Sum of Sq    RSS    AIC
## + State.Bottle.Retail  1     71786 584877 2783.4
## + State.Bottle.Cost    1     71739 584925 2783.5
## + Pack                 1     66826 589837 2786.6
## + Bottle.Volume..ml.   1     49810 606854 2797.3
## + Store.Number        72    230477 426186 2806.4
## + Sale..Dollars.       1     12286 644378 2819.9
## <none>                             656663 2825.0
##
## Step:  AIC=2783.44
## Bottles.Sold ~ Volume.Sold..Gallons. + Category.Name + State.Bottle.Retail
##
##                      Df Sum of Sq    RSS    AIC
## + Pack                1     75578 509299 2733.4
## + Bottle.Volume..ml.  1     52157 532721 2750.3
## + Store.Number       72    190258 394619 2779.5
## + Sale..Dollars.      1      4072 580805 2782.8
## <none>                            584877 2783.4
## + State.Bottle.Cost   1      1748 583130 2784.3
##
## Step:  AIC=2733.41
## Bottles.Sold ~ Volume.Sold..Gallons. + Category.Name + State.Bottle.Retail +
##     Pack
##
##                      Df Sum of Sq    RSS    AIC
## + Sale..Dollars.      1      6620 502680 2730.5
## <none>                            509299 2733.4
## + Bottle.Volume..ml.  1      2188 507111 2733.8
## + State.Bottle.Cost   1      1672 507628 2734.2
## + Store.Number       72    153610 355689 2742.4
##
## Step:  AIC=2730.49
## Bottles.Sold ~ Volume.Sold..Gallons. + Category.Name + State.Bottle.Retail +
##     Pack + Sale..Dollars.
```

```
##
##                       Df Sum of Sq    RSS    AIC
## <none>                               502680 2730.5
## + Bottle.Volume..ml.  1      2354 500326 2730.7
## + State.Bottle.Cost   1      1016 501663 2731.7
## + Store.Number       72    147694 354985 2743.7
```

```
##
## Call:
## lm(formula = Bottles.Sold ~ Volume.Sold..Gallons. + Category.Name +
##     State.Bottle.Retail + Pack + Sale..Dollars., data = dfLiquorSales)
##
## Coefficients:
##                                   (Intercept)
##                                     -31.761507
##                          Volume.Sold..Gallons.
##                                       4.093187
##                Category.NameCANADIAN WHISKIES
##                                     36.253685
##                   Category.NameIRISH WHISKIES
##                                     -4.873548
##                  Category.NameSCOTCH WHISKIES
##                                     -8.116166
## Category.NameSINGLE BARREL BOURBON WHISKIES
##                                     12.384335
##        Category.NameSTRAIGHT BOURBON WHISKIES
##                                     -3.416345
##             Category.NameSTRAIGHT RYE WHISKIES
##                                     17.005140
##                Category.NameTENNESSEE WHISKIES
##                                     -1.310738
##                            State.Bottle.Retail
##                                       0.079326
##                                           Pack
##                                       2.759534
##                                Sale..Dollars.
##                                     -0.003802
```

```
fit1 <- lm(formula = Bottles.Sold ~ Volume.Sold..Gallons. + Category.Name +
     State.Bottle.Retail + Pack + Sale..Dollars., data = dfLiquorSales)
```

Volume.Sold..Gallons=One unit of increase in volume sold will increase the bottles sold by 4.10 units

Category.Name

CANADIAN WHISKIES is the only one that is significant.

CANDIAN WHISKIES vs BLENDED WHISKIES coefficient is 5.492

State.Bottle.Retail=One unit of increase in State.Bottle.Retail will increase the bottles sold by 7.133 units

Pack=One unit of increase in Pack will increase the bottles sold by 7.133 units

Sale..Dollars.=It is slightly significant. One unit of increase in Sale..Dollars. will decrease the bottles sold by 2.189 units

Formula = -4.196 + 25.314*Volume.Sold..Gallons. + 5.492* (Category.Name="CANADIAN WHISKIES") +

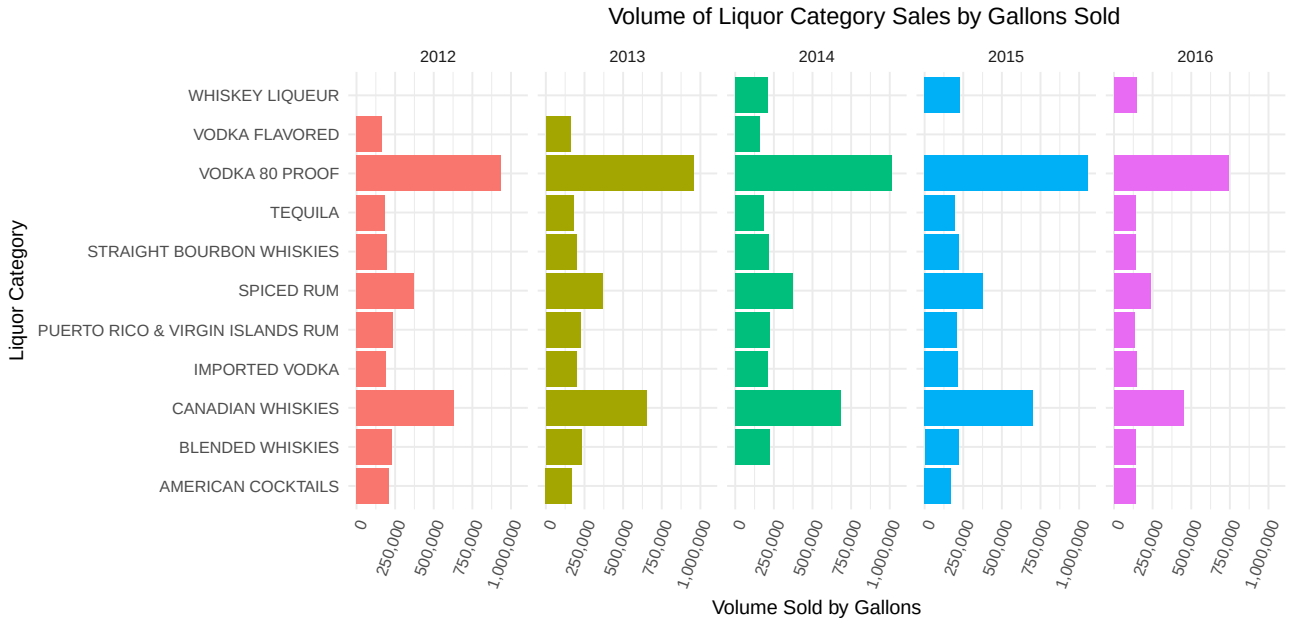
7.133 * (State.Bottle.Retail) + 7.521 * (Pack) - 2.189 * Sale..Dollars.

Data Acquisition

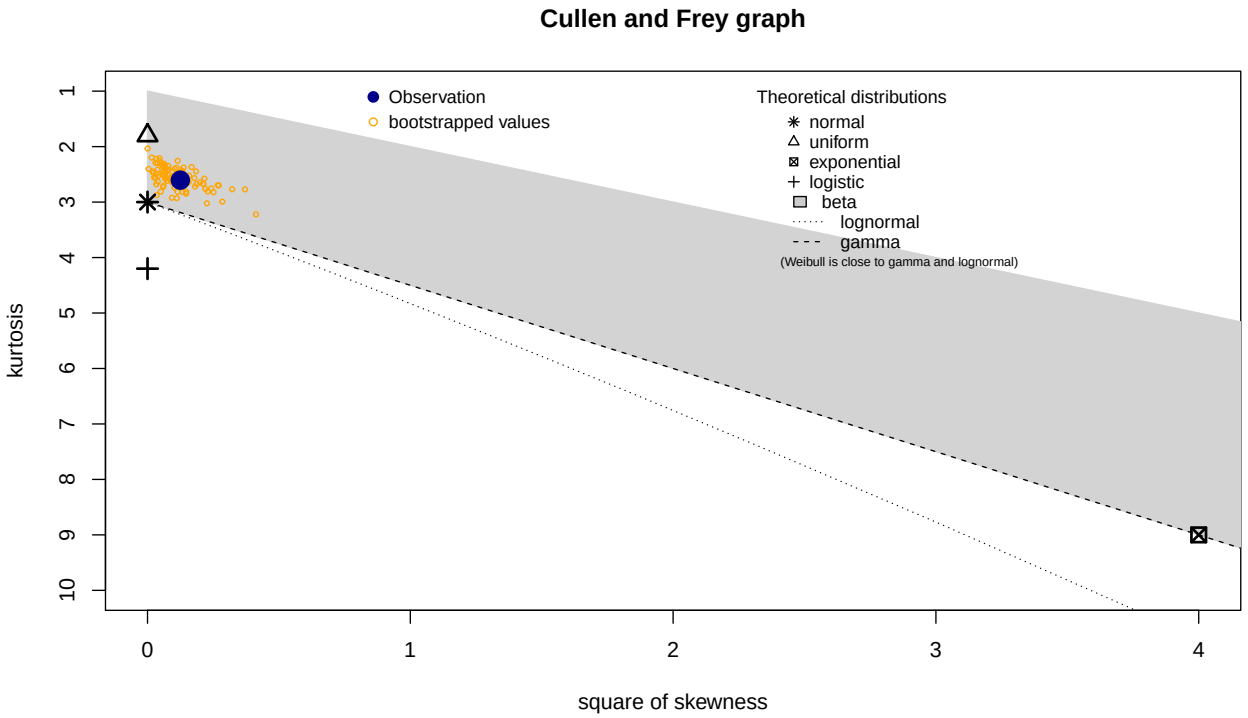
The data set contains the spirits purchase information of Iowa Class "E" liquor licensees by product and date of purchase from January 1, 2012 to current. The data set is provided by the Iowa Department of Commerce, Alcoholic Beverages Division, click here to view the data set at Data.Iowa.Gov.

As discussed in our methodogly section, the data set is 2.68 GB in total size and much to large to process with our current techiniques. A greater understanding of working with data sets of this size with R w

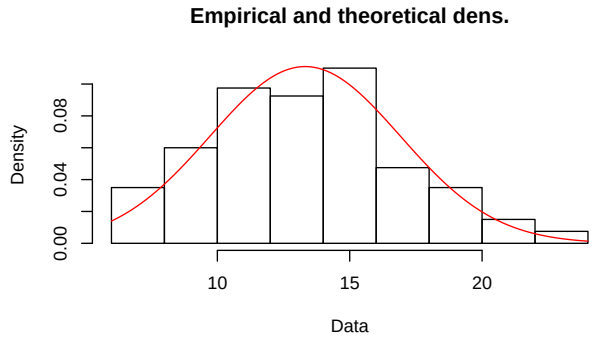
We reviewed the liquor sales by gallons sold per year by Liquor Category. Initially, we viewed the top 5 Liquor Categories by volume sold but there were large disparate between years, suggesting that the top 5 change often and likely due to changing consumer tastes. We do see a more stable set of liquor categories for the top 10 category which suggests that while tastes may change we don't see large movements in liquor categories. For this analysis we will focus on the whiskey categories.

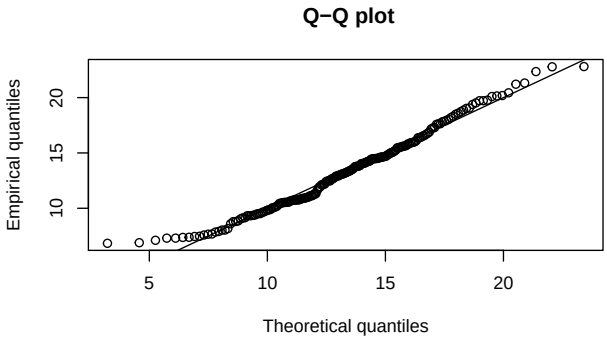


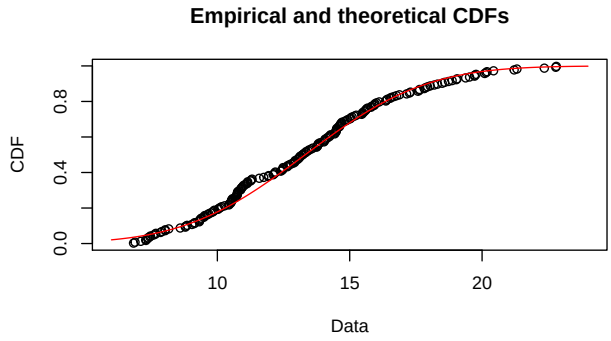
Our first attempt was to use a Poisson regression due to the over-dispersion created by aggregating the data set. However, the distribution was far to negatively skewed to fit a poission distribution. We therefore chose

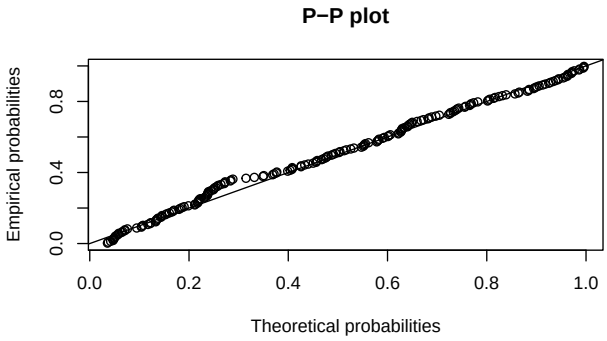


min	max	median	mean	sd	skewness	kurtosis	method
6.834324	22.80474	13.21952	13.31821	3.604091	0.3534382	2.604675	unbiased









dfLiquorSales
14 Variables 376 Observations

X

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
376	0	376	1	188.5	125.7	19.75	38.50	94.75	188.50	282.25	338.50	357.25

```
lowest :   1   2   3   4   5, highest: 372 373 374 375 376
```

Month

n	missing	distinct	Info	Mean	Gmd
376	0	1	0	11	0

```
Value        11
Frequency   376
Proportion    1
```

Year

n	missing	distinct	Info	Mean	Gmd
376	0	1	0	2015	0

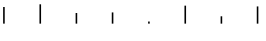
```
Value      2015
Frequency   376
Proportion    1
```

City

n	missing	distinct	value
376	0	1	DES MOINES

```
Value      DES MOINES
Frequency         376
Proportion          1
```

Category.Name



n	missing	distinct
376	0	8

BLENDED WHISKIES (61, 0.162), CANADIAN WHISKIES (71, 0.189), IRISH WHISKIES (39, 0.104), SCOTCH WHISKIES (41, 0.109), SINGLE BARREL BOURBON WHISKIES (8, 0.021), STRAIGHT BOURBON WHISKIES (66, 0.176), STRAIGHT RYE WHISKIES (27, 0.072), TENNESSEE WHISKIES (63, 0.168)

Store.Number

n	missing	distinct
376	0	73

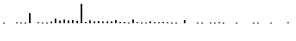
lowest : 2190 2248 2527 2528 2532, highest: 5131 5132 5137 5145 5169

Store.Name

n	missing	distinct
376	0	72

lowest : AV Superstop Best Food Mart / Des Moines C Fresh Market Cash Saver / E Euclid Ave C
highest: Walgreens #05852 / Des Moines Walgreens #07452 / Des Moines Walgreens #07453 / Des Moines Walgreens #07833 / Des Moines W

Pack



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
376	0	141	0.99	13.38	6.201	6.00	6.00	9.75	12.00	16.00	21.37	24.00

lowest : 3.00000 4.50000 5.00000 5.25000 5.40000, highest: 30.00000 32.00000 32.57143 33.81818 36.00000

Bottles.Sold

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
376	0	153	0.999	116.1	173.3	3.00	5.00	9.75	32.00	96.50	247.00	444.25

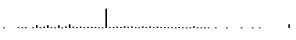
lowest : 1 2 3 4 5, highest: 1450 1704 1961 2431 3228

Sale..Dollars.

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
376	0	343	1	1891	3024	44.98	69.84	162.63	371.25	1078.08	2937.44	6963.48

lowest : 7.20 21.74 22.49 26.25 27.14, highest: 35818.26 38945.46 40135.98 54923.22 71157.84

Bottle.Volume..ml.



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
376	0	180	0.986	803.8	345.4	375.0	435.4	573.8	750.0	953.8	1250.0	1564.3

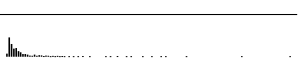
lowest : 200.000 270.000 287.500 300.000 310.000, highest: 1416.667 1500.000 1550.000 1607.143 1750.000

State.Bottle.Cost

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50
376	0	319	1	101.7	127.4	7.262	10.670	18.495	47.375
.75	.90	.95							
103.537	225.940	338.075							

lowest : 3.21 3.46 3.50 4.10 4.40, highest: 887.95 918.77 1045.11 1362.12 1640.44

State.Bottle.Retail



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
376	0	322	1	152.6	191.3	10.90	16.01	27.75	71.09	155.34	338.93	507.33

```
lowest :    4.82    5.19    5.25    6.15    6.60, highest: 1332.12 1378.66 1568.12 2049.35 2467.91
```

Volume.Sold..Gallons.

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
376	0	254	1	24.4	38.04	0.4225	0.7350	1.7900	5.1500	18.2400	47.0350	88.4975

```
lowest :   0.13   0.20   0.30   0.32   0.39, highest: 383.96 417.74 483.05 623.52 830.55
```

```
describe(dfLiquorSales$Store.Number)
```

```
## dfLiquorSales$Store.Number
##        n  missing distinct
##      376        0       73
##
## lowest : 2190 2248 2527 2528 2532, highest: 5131 5132 5137 5145 5169
```

```
describe(dfLiquorSales$Category.Name)
```

```
## dfLiquorSales$Category.Name
##        n  missing distinct
##      376        0        8
##
## BLENDED WHISKIES (61, 0.162), CANADIAN WHISKIES (71, 0.189), IRISH
## WHISKIES (39, 0.104), SCOTCH WHISKIES (41, 0.109), SINGLE BARREL BOURBON
## WHISKIES (8, 0.021), STRAIGHT BOURBON WHISKIES (66, 0.176), STRAIGHT RYE
## WHISKIES (27, 0.072), TENNESSEE WHISKIES (63, 0.168)
```

describe the specifics of what you did (data exploration, data preparation, model building, model selection, model evaluation, etc.), and what you found out (statistical analyses, interpretation and discussion of the results, etc.).

Discussion and Conclusions

In another study conducted in 2012 in Idaho, the monthly revenue generated was examined rather than the yearly revenue generated. The continued growth was rather owed to the number of weekends a month has (five instead of four) and to the higher prices in neighboring states. In Washington, the voters approved an initiative that led the state to sell its liquor stores and add new distributor and retail fees, making prices in the neighboring states (Idaho and Oregon) look better. There were no changes made in marketing or pricing in response to the regulatory shift in Washington (**???**). Further research into the proximity of our counties to states and towns with higher prices and regulation may provide more insight into sales and volume of liquor sold. Additionally, reviewing the data by identifying months that has 5 weekends instead of four could provide further insights.

conclude your findings, limitations, and suggest areas for future work.

Appendices

Supplemental tables and/or figures.

Session Info

- R version 3.3.2 (2016-10-31), `x86_64-w64-mingw32`
- Locale: `LC_COLLATE=English_United States.1252`, `LC_CTYPE=English_United States.1252`, `LC_MONETARY=English_United States.1252`, `LC_NUMERIC=C`, `LC_TIME=English_United States.1252`
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: dplyr 0.5.0, fitdistrplus 1.0-7, Formula 1.2-1, ggplot2 2.2.0, Hmisc 4.0-1, lattice 0.20-34, logspline 2.1.9, magrittr 1.5, MASS 7.3-45, pacman 0.4.1, pander 0.6.0, purrr 0.2.2, readr 1.0.0, survival 2.40-1, tibble 1.2, tidyr 0.6.0, tidyverse 1.0.0
- Loaded via a namespace (and not attached): acepack 1.4.1, assertthat 0.1, backports 1.0.4, base64 2.0, cluster 2.0.5, codetools 0.2-15, colorspace 1.3-1, data.table 1.10.0, DBI 0.5-1, digest 0.6.10, evaluate 0.10, foreign 0.8-67, grid 3.3.2, gridExtra 2.2.1, gtable 0.2.0, htmlTable 1.7, htmltools 0.3.5, knitr 1.15.1, labeling 0.3, latticeExtra 0.6-28, lazyeval 0.2.0, Matrix 1.2-7.1, munsell 0.4.3, nnet 7.3-12, openssl 0.9.5, plyr 1.8.4, R6 2.2.0, RColorBrewer 1.1-2, Rcpp 0.12.8, reshape2 1.4.2, rmarkdown 1.2, rpart 4.1-10, rprojroot 1.1, rticles 0.2, scales 0.4.1, splines 3.3.2, stringi 1.1.2, stringr 1.1.0, tools 3.3.2, yaml 2.1.14

R statistical programming code.

Please see Final Project.rmd on GitHub for source code.

https://github.com/ChristopheHunt/DATA-621-Group-1/blob/master/Final%20Project/Final%20Project.Rmd

References

Anonymous. 2016. “Specialty Products Grow in Wine, Spirits.” *Beverage Industry* 107(7).

Boshart, Rod. 2001. “Liquor Sales in Iowa Set Record.” *Gazette*.

Del Buono, Amanda. 2016. “Keeping Spirits High.” *Beverage Industry* 107.4: 14–16, 18.