Iowa Liquor Sales

Jim Lung
05-15-2018

Abstract

- The goal of this project is to determine a geographic location (county) in Iowa that will yield the highest amount of liquor sales.
- To use the correlation and forward stepwise regression with linear method are applied to perform predication the bottles sold.
- To obtain the specific trend and predication, the dataset was subset into the highest volume sold location at DES MOINES City in 2017

Key words

- Liquor Sales,
- Naive Forecast,
- Linear Regression
- Inventory Forecast
- Prediction

Introduction

- To create a statistical model for the number of bottles sold of liquor which is within the state of Iowa.
- To make informed decisions on inventory prediction, sales, and assist wholesale distributors to plan for the predicted volume of distribution
- Large number of observations are found in Polk County, the city of Des Moines, and the zip code 50010 (Ames, Iowa). Ames is the home of Iowa State
- To set up the range of our analysis to the City of Des Moines in 2017

Literature review

An Overview of Forecasting Methodology:
David S. Walonick (1993)

- Trend extrapolation - These methods examine trends and cycles in historical data, and then use mathematical techniques to extrapolate to the future.
- The assumption of all these techniques is that the forces responsible for creating the past, will continue to operate in the future.
- This is often a valid assumption when forecasting short term horizons, but it falls short when creating medium and long term forecasts

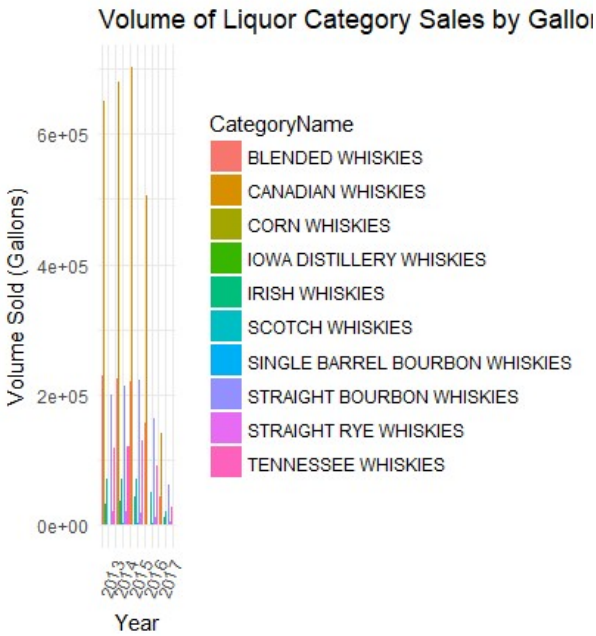
Data Exploration

Achieve the dataset from:
https://data.iowa.gov/Economy/Iowa-Liquor-Sales/m3tr-qhgy.

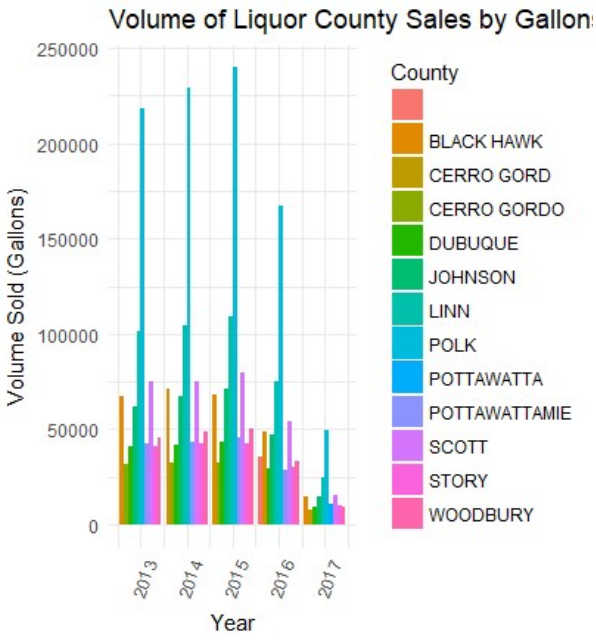
- The data set contains the spirits purchase information of Iowa Class “E” liquor licensees by product and date of purchase from January 2013 to December 2017
- The data set is provided by the Iowa Department of Commerce, Alcoholic Beverages Division
- data set is 3.3 GB in total size and much to large to use in a meaningful model

Data Exploration

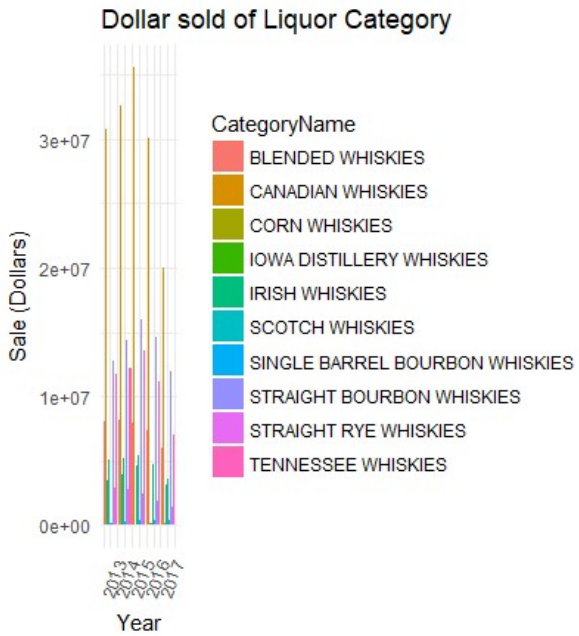
- Ensuring bottle size (e.g., 750 ml) x bottles sold = volume liters sold
- Ensuring bottle retail value x bottles sold = sale dollars
- I found no problems with the math, but it was good to check all the same



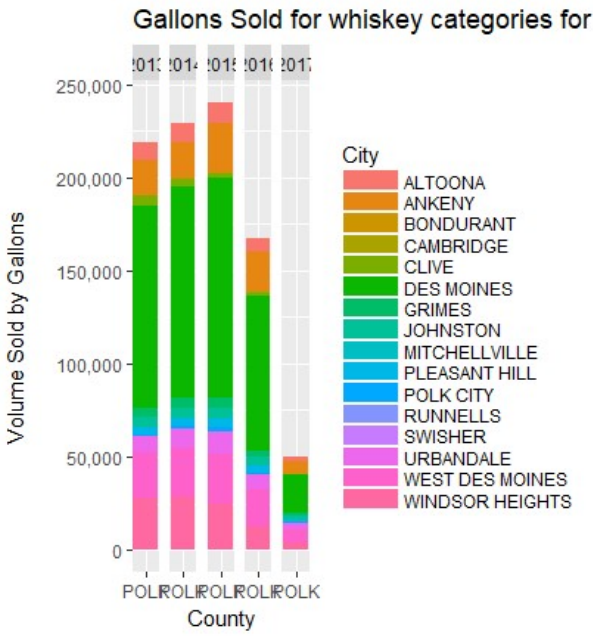
Volume of Liquor Category Sales by Gallons Sold



Volume of Liquor County Sales by Gallons Sold



Dollar sold of Liquor Category



Gallons Sold for whiskey categories for each City in Polk County

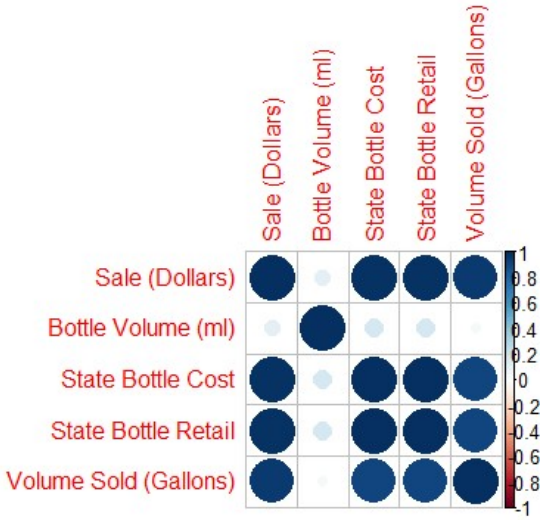
Build Models

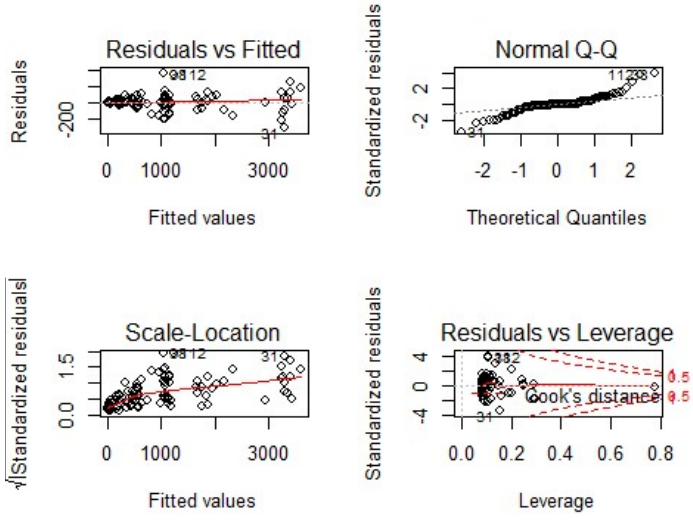
- Build model by selection the highest volume location at DES MOINES City

Month	Year	City	CategoryName	Bottles Sold	Sale (Dollars)
1	2017	DES MOINES	BLENDED WHISKIES	6.901226	35100.55
1	2017	DES MOINES	CANADIAN WHISKIES	7.926108	112301.42
1	2017	DES MOINES	CORN WHISKIES	3.170261	948.75
1	2017	DES MOINES	IOWA DISTILLERY WHISKIES	1.592121	716.73
1	2017	DES MOINES	IRISH WHISKIES	6.342434	31019.63
1	2017	DES MOINES	SCOTCH WHISKIES	6.065074	35540.77

Correlation (Model 1)

- model1 <- lm(`Bottles Sold` ~ CategoryName + `Sale (Dollars)` + `Bottle Volume (ml)` + `State Bottle Cost` + `State Bottle Retail` + `Volume Sold (Gallons)`, data=iowa_data_reduced2)





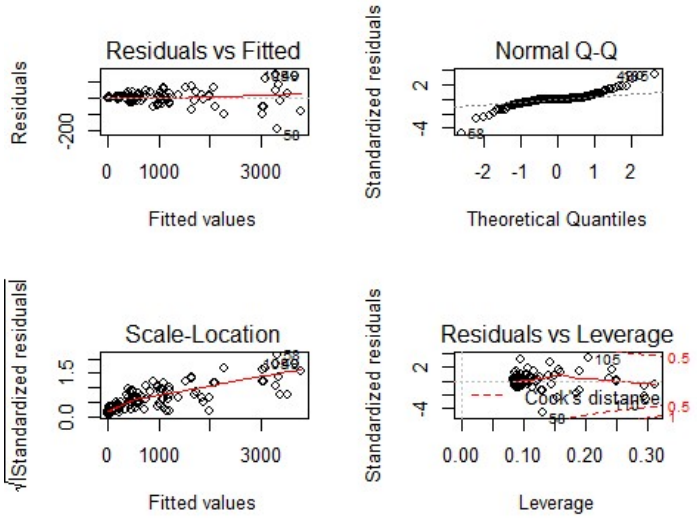
Correlation (Model 1)

Correlation (Model 1)

rn	GVIF	Df	$GVIF^{1/(2*Df)}$	Adjusted_GVIF
CategoryName	255.8244	9	1.360738	1.851608e+00
Sale (Dollars)	277.4708	1	16.657454	2.774708e+02
State Bottle Cost	1823143.9656	1	1350.238485	1.823144e+06
State Bottle Retail	1820815.5428	1	1349.375983	1.820816e+06

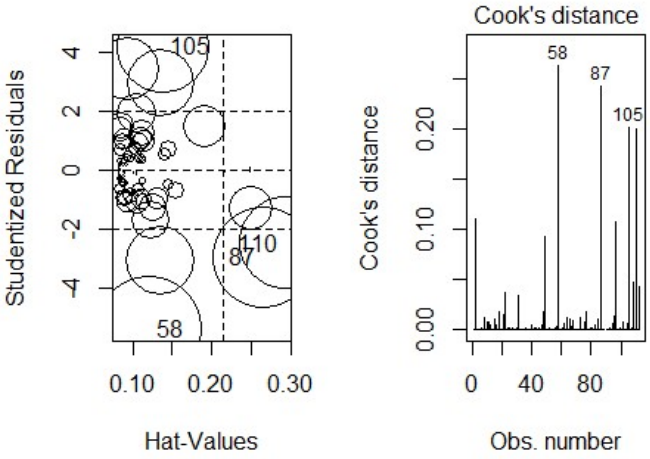
Forward regression (Model 2)

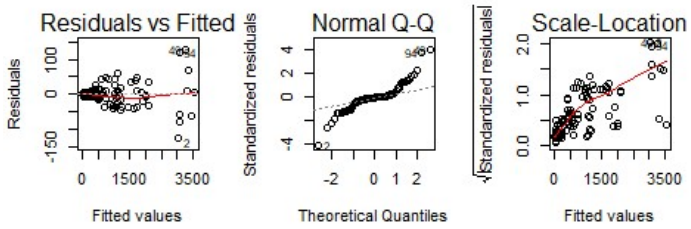
- forward <- step(lm(`Bottles Sold~1,data=iowa_data_reduced2),direction = "forward", scope=~CategoryName +Sale (Dollars)+ Bottle Volume (ml) +State Bottle Cost+State Bottle Retail+Volume Sold (Gallons)`,trace = FALSE)

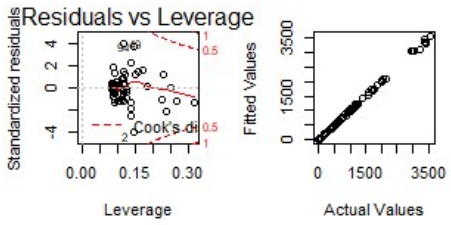


Forward regression (Model 2)

rn	GVIF	Df	$GVIF^{1/(2*Df)}$	Adjusted_GVIF
Volume Sold (Gallons)	60.38886	1	7.771027	60.388858
CategoryName	391.52906	9	1.393293	1.941265
State Bottle Retail	58.81699	1	7.669224	58.816992







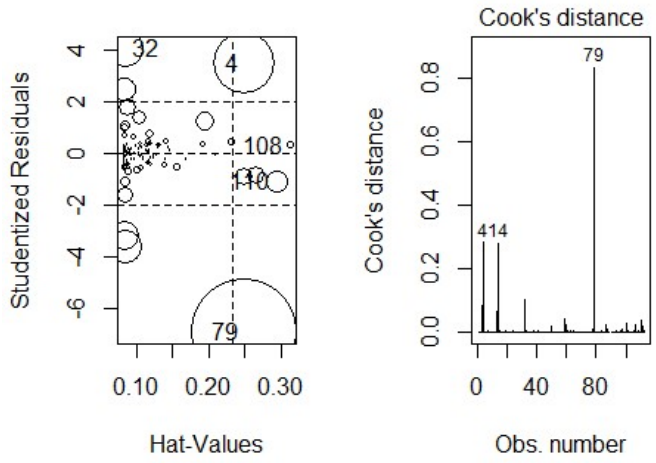
Model 2 remove influencePlot

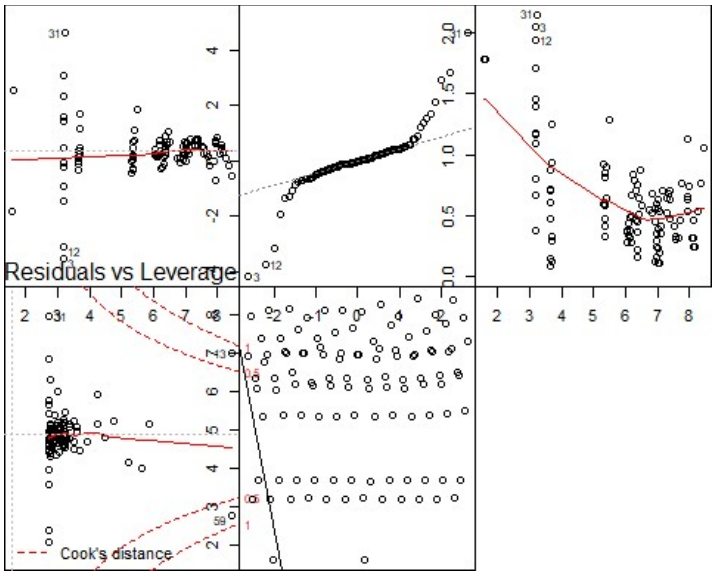
Bottles Sold Model with Log Transformation (Model 3)

- model3 <- step(lm(log(`Bottles Sold`) ~1,data=iowa_data_reduced2),direction = "forward", scope=~CategoryName +`Sale (Dollars)`+ `Bottle Volume (ml)` +`State Bottle Cost`+`State Bottle Retail`+`Volume Sold (Gallons)`,trace = FALSE)

rn	GVIF	Df	$GVIF^{1/(2*Df)}$	Adjusted_GVIF
CategoryName	6372.52413	9	1.626862	2.646679
`State Bottle Retail`	50.39835	1	7.099179	50.398346
`Volume Sold (Gallons)`	56.74351	1	7.532829	56.743511
`Bottle Volume (ml)`	22.88167	1	4.783479	22.881669

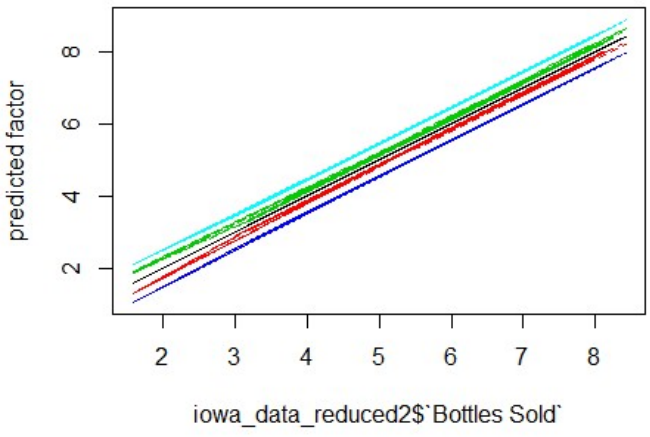
Bottles Sold Model with Log Transformation (Model 3)





Selection Model

Models	AIC	BIC	Deviance	df.residual
Correlation	1363.90365	1401.96263	9.928481e+05	99
Forward regression	1079.26792	1114.13563	1.087422e+05	96
Log transformation	-11.08298	26.72375	4.514541e+00	97



Prediction

Discussion and Conclusions

- resulting models allow us to model in Des Moines for both Bottles Sold. We can utilize a naive forecast, assuming that the prior year of 2017 is predictive of the year 2018.
- performance of the proposed method was evaluated using a real data set provided by Iowa Department of Commerce, Alcoholic Beverages Division. The results of the evaluation indicated that the proposed method can cope with the low number of past records while accurately forecasting sales.
- evaluation of the liquor data set using these techniques may provide greater insight as the vast number of records could produce a more accurate model.