Drinking Excess Alcohol is Dangerous (DEAD)

Consultants: Brendan Callender, Jadyn Ellis, Kyle Lew, Soren Paetau, Anagha Sikha
California Polytechnic State University - San Luis Obispo

Abstract

In this proposal, you will find descriptions of the data we will be using within our analysis. Moreover, we have included the methods we will use to validate and select our final model to provide you with the most significant results. Our resulting conclusions will include seasonal trends and detailed feature effects data that will allow DEAD to better plan when and where to focus advocacy efforts.

Data

The data was pulled from the Iowa Liquor Sales dataset provided by the Iowa Department of Revenue, Alcoholic Beverages section. The data includes the purchase information for Iowa Class "E" liquor licenses on a store level. An example row from the data is shown below.

Iowa Liquor Sales Data:

- Each row represents the store's order of single alcohol category on a given day
    - *Store Number* uniquely identifies each store
    - *Bottles Sold* tracks the number of bottles purchased by the store
    - *State Bottle* retail tracks the per bottle price the store paid
    - *Sale (Dollars)* tracks the total money spent by the store

Date	Store Number	Store Name	City	Category Name	Bottles Sold	State Bottle Retail	Sale (USD)
01/23/2018	4310	Fareway Stores #502 / Cherokee	Cherokee	Canadian Whiskies	3	19.65	58.95

So, above we see that Store Number 4130 purchased 3 bottles of Canadian Whiskey for $58.95

We also pulled external data relating to the population for Iowa cities in order to group cities into different population brackets to better predict alcohol demand for stores in those cities. (https://www.iowa-demographics.com/cities_by_population)

Aggregated Liquor Sales Data

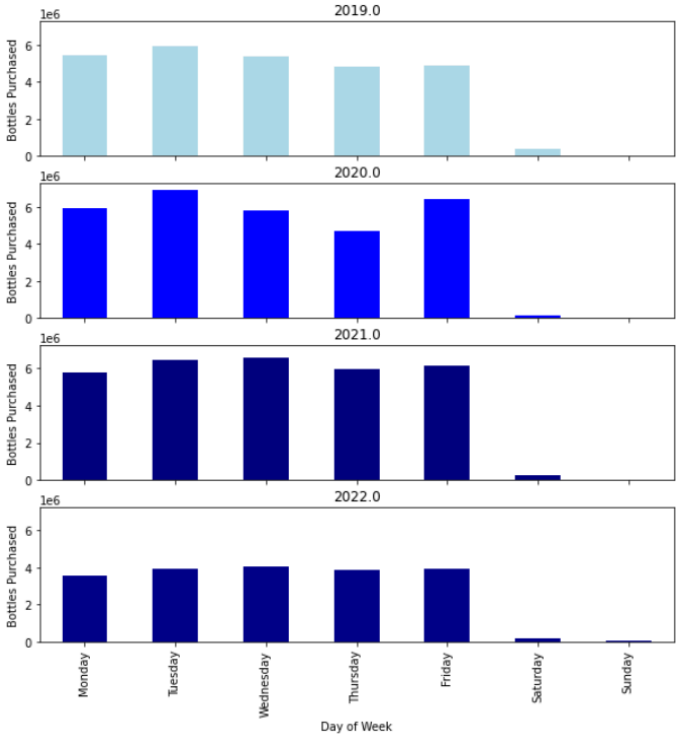
- Each row represents the order (to the distributor) of a single alcohol category for a store on a given day
    - *isWeekend* is True if it is a Friday, Saturday, or Sunday, and is False if it is a Monday, Tuesday, Wednesday, and Thursday

- *College Town* is True if the store is in Iowa City for University of Iowa and Ames for Iowa State and False if otherwise
- *isHoliday* is True if the date was a holiday, and False if the date was not on a holiday
- *Bottles Sold* is the number of bottles of the given liquor category that the store purchased on that day

Date	Store Name	Category Name	Bottle Volume (ml)	State Retail Cost	Bottles Sold	Sale (Dollars)	Day of Week	isWeekend	College Town	isHoliday
2019-03-19	Hy-Vee	Whiskey	750	14.67	10	150	Monday	F	T	F

So, as seen above, on 03/19/2019, a Hy-Vee store bought 10 bottles of whiskey for $150 total in a college town.

While the day of the week itself may not be important in determining the store purchases of liquor, larger purchases during the week may indicate increased demand for alcohol over the weekend. To explore the Day of the Week variable, we plotted the bottles purchased for each day where more bottles were purchased on Monday to Friday compared to Saturday and Sunday.



Model Selection & Validation

The goal with creating a model on our selected parameters is to explore the effects certain variables have on alcohol consumption. With this goal in mind, we plan to fit a multiple linear regression model, using the concluded coefficients to quantify these effects. As we move through the model selection process, we will use R^2 as our validation metric and cross validation as our training/test splits to test our models as we decide what variables to include in the model.

The main model selection choice we will be exploring is which parameters to include in our model. With our list of potential explanatory variables, we will test multiple models, including different combinations to try to find the best predictive model. With our selected explanatory variables, we will prioritize and ensure that our resulting model fit is interpretable and able to be utilized when determining what drives higher liquor sales.

Conclusions

Once developed, we will provide you with information and interpretations of the resultant coefficients for each of our selected features when predicting bottles purchased by stores. We hope by the end of our analysis, that we will be able to say with confidence which factors influence the purchasing habits of alcohol and the degree to which it influences said purchasing habits.

What you can expect from our model:

- Coefficients quantifying the effect for each variable
- Significance of seasonal trends of alcohol consumption
- Interpretations regarding the meaning of the model coefficients
    - Evidence to support where to focus advocacy efforts