Sophia Chung, Lana Huynh, Jamie Luna, Rachel Roggenkemper, James Rounthwaite
California Polytechnic State University - San Luis Obispo

Drinking Excess Alcohol is Dangerous (DEAD)

13th October 2023

OVERVIEW

This project aims to find the most impactful patterns that drive alcohol purchases in Iowa. There are many factors that can affect when people buy and drink alcohol. These complex patterns are not always immediately apparent. We plan on providing a tool showcasing the most impactful patterns and variables that go into alcohol sales.

DATA PLAN

The current data does not enclose the liquor stores exact sales, so we will be using various metrics to represent purchases in Iowa. The easiest and most reliable way to accomplish this is by observing when and how much alcohol is purchased by stores from vendors. Furthermore, we plan on separating the types of alcohol's inventory fluctuations to gain greater insight into what circumstances impact buying different types of alcohol. We plan on defining month-to-month inventory changes by alcohol category type to watch how they fluctuate for our model to use.

In order to make this dataset more manageable, there are several alterations we have made. We have added a Month and a Year column so that we have the option to use these as model predictors rather than only the date as a whole.

Additionally, we chose to use a subset of the data which includes 300,000 rows with a fairly equal distribution of purchases between the years 2012 and 2015. This allows us to quickly build lots of different models and features with relative ease. However, when we start narrowing down the best model to deliver we will use the bulk of the data to provide the most accurate results.

Lastly, other convenient changes have been amended to the dataset to make the process easier. For example, month and year columns have been added.

METHODS AND MODELS

Figure 1. Bottles of Alcohol Bought from Vendor by Central City Liquors

Bottles of Alcohol Sold from Vendor by Central City Liquors

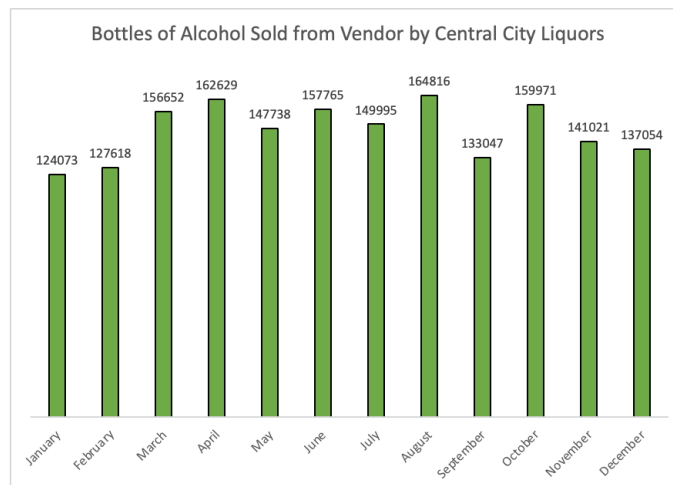| Month	Bottles
January	124073
February	127618
March	156652
April	162629
May	147738
June	157765
July	149995
August	164816
September	133047
October	159971
November	141021
December	137054

Figure 1 shows us the distribution of bottles of alcohol bought by month. As expected, January has the lowest amount of bottles sold, which makes sense because it is after the rush of holiday sales.

Figure 2. Mean Bottles Sold by Day of Week

Mean Bottles Sold by Day of Week

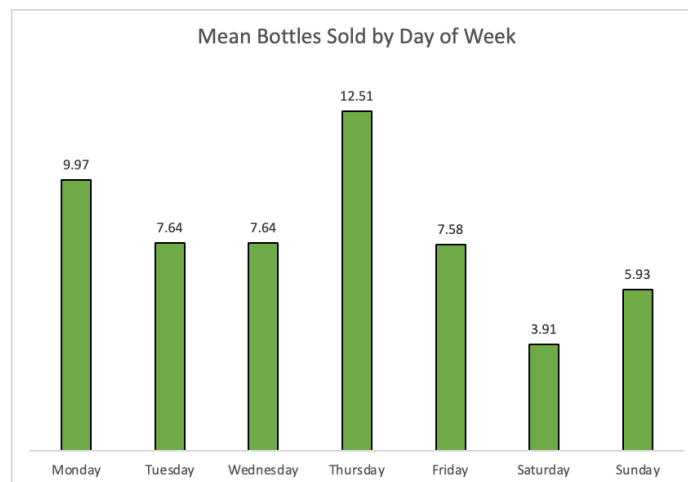| Day	Mean Bottles
Monday	9.97
Tuesday	7.64
Wednesday	7.64
Thursday	12.51
Friday	7.58
Saturday	3.91
Sunday	5.93

Figure 2 shows us the distribution of bottles of alcohol bought by week. As expected, Thursday has the highest amount of bottles sold, which makes sense as people are getting prepared for the weekend.

Since we don't have sales as a feature on this data, to fit the models, we will be inventing some kind of metric to represent the sales. Our strategy to achieve this involves monitoring the

purchasing patterns of alcohol from liquor stores. We can do this by summing up the alcohol acquired by each store and examining the monthly patterns.

Finding the most impactful factors that go into alcohol sales requires a wide range of methods to ensure the results are grounded in truth. Alcohol sales can fluctuate for many reasons and often without good cause. We want to focus on finding factors that have predictable impact on sales to best inform D.E.A.D. One way we will be doing this is taking a robust approach towards outliers in our data. In a technical sense this means using objective functions like log-cosh loss and penalties such as ridge and elastic net to ensure the results we give are the most helpful. All the models we make will be geared towards looking at the relationship between patterns and current inventory of numerous alcohol stores.

Two end results models will be selected. The first will be using an "aggregated inventory", which looks at the inventory items fluctuations irrespective of alcohol category. The second model will look at inventory changes per alcohol category. Our categories will focus on broad categories like hard liquor, beer, and wine to provide insights in when particular spirits are bought.

Our models will have an emphasis on finding accurate correlations over anything else.

For choosing the best model we plan on completing an extensive model search with various combinations of penalty weights and model specifications. Each model will be judged on how complicated it is, how well it does compared to an average guess, and how far away our predictions are. All of these will aid us in creating a successful model to find the most impactful factors that go into alcohol sales.

CONCLUSIONS

Upon completion you will be able to see the most impactful factors in alcohol sales by category to better orient your campaigns against drinking. By honing in on reliable implications we plan on providing a tool to best support your campaign that can be continually updated with current fluctuations. Beyond the model itself, we plan on providing immediate analysis driven from the model's results to be used immediately.

While there are many ways to find the driving factors behind alcohol purchases, we are confident that this process will produce accurate and nuanced information to aid your campaign. Everything from how we plan on counting alcohol purchases to our model selection are geared towards building a model to answer your questions about Iowa's alcohol purchases.