Iowa Liquor Sales
and
Socioeconomic Status

Liam Fruzyna, Katy Weathington, and Sarah Graupman

https://mail929.github.io/MU-COSC4610-Project

Introduction

Alcohol consumption and alcoholism is often associated with the poor and homeless. We are looking to find if there is an association between income and alcohol purchases by looking at data within counties in Iowa. This research can be used to help find a way to guide people out of alcoholism by determining a potential cause of the problem.

The goal of this project is to determine if there are any connections between local liquor sales and the socioeconomic status of a county in Iowa. Our focuses for this project is to see if there is a correlation between a county's income and the amount, count, and cost of alcohol purchases in that county. We also want to look at popularity and density of stores within the counties. We will evaluate our conclusions by comparing similar counties to each other and looking for consistent patterns as well as comparing our results to those found in our research.

Data Set

We are currently planning on using two data sets. The first is the Iowa Liquor Sales[1] set. This is a dataset of all individual spirits purchases in the state of Iowa from 2012 to present (for some reason it only appears to be 2013 through 2015). This is a very large data set with over 13 million entries. Each entry contains 24 attributes including the purchase data, information about the store, and a description of the purchase including name, costs, size, and count.

Our second data set is the US Census Bureau's SAIP Iowa County Estimates for 2016[2]. We are interested in this data primarily because it provides estimated median income by county for all states including Iowa. It also includes information about children and families in poverty which could also be a useful factor in our progress.

Tools

We will do most of our work in Python, utilizing Jupyter Notebook and Anaconda. Any front end we will do will be in html5 and JavaScript. Libraries we are likely to use include:

- Pandas, allowing us to easily convert our data from a tsv to a dataframe
- MatPlotLib, to allow for advanced graphs
- Plotly, to allow for advanced graphs
- Cufflinks, to facilitate the use of plotly with our pandas dataframe
- Seaborn, to allow for advanced graphs
- NumPy, for linear algebra and vector calculations
- Scikit-Learn, for general classification algorithms

Though each library enhances python in some way, we will likely not use every single one of these libraries, especially not all the graph enhancers. However, we will utilize Pandas and

[1] https://data.iowa.gov/Economy/Iowa-Liquor-Sales/m3tr-qhgy
[2] https://www.census.gov/data/datasets/2016/demo/saipe/2016-state-and-county.html

Scikit-learn and at least one of MatPlotLib, Plotly, or Seaborn. We may find other libraries that prove useful throughout the duration of the project.

Literature Review

The Relationship Between Neighborhood Poverty and Alcohol Use: Estimation by Marginal Structural Models[3]

The goal of this paper was to model the relationship between neighborhood poverty and alcohol use while accounting for neighborhoods changing over time, because most do not. They wanted to then determine if there is a larger influence on alcohol use from the neighborhood choice or the selection of people who ended up living in the neighborhood. In the study they determined that the status of the individual does affect neighborhood choice and alcohol consumption, but it also determined that alcohol use was a product of the opportunities or conditions of the neighborhood. Overall the study determined that while longer term exposure to poverty increased alcohol use, short term exposure increased binging. This provides insight into our project as we may want at previous years census data to determine how long neighborhoods have been exposed to poverty.

Socioeconomic Determinants of Exposure to Alcohol Outlets[4]

The goal of this paper was to first test is there are more alcohol outlets in areas with high alcohol demand and secondly test if there are less outlets in areas with high land and structured rents. The study used a Bayesian conditional autoregressive Poisson models to estimate alcohol outlets and mark potential income and zoning ordinances. The result of the study was that there is a concentration of outlets in low-income areas and fewer in high-income areas. This aligns with our hypothesis that there is increased alcohol consumption and density areas of poverty.

Is there a link between alcohol consumption and the level of poverty? [5]

The goal of this paper was to predict poverty levels of Sri Lankan household based on the type of alcohol consumed in that household. They used survey and census data to build a logistic regression that would predict the poverty level of a household. The study found that if a household consumed illegal beverages like kasippu or toddy, they were more likely to be in poverty. They also found that most non-poor households consumed legal beverages. This is important to note in that we may find a relationship between different type of liquors and poverty levels. It is also important because we must realize that recorded alcohol sales are not the only indication of alcoholic consumption.

[3] http://www.jstor.org/stable/pdf/25680578.pdf
[4] https://0-www-jsad-com.libus.csd.mu.edu/doi/full/10.15288/jsad.2015.76.439
[5] https://www.tandfonline.com/doi/full/10.1080/00036846.2015.1114574