

Drinking Excess Alcohol is Dangerous (DEAD)

Project Proposal

DATA 403: Project 1

Justin Mai, Libby Brill, Maxwell Dubow, Rachel Hartfelder

Introduction

Thank you for choosing us to perform this analysis. Provided below is detailed information about the data we will be using, the model selection process, and the conclusions we aim to provide. We look forward to working with you.

Data

For the model, we will be using the alcohol sales transactions from the [Iowa Liquor Sales dataset](#). The data it contains spans from 2012 to 2025. However, we will be using the section of data beginning on September 1st, 2020 and ending on August 31st, 2025 as we believe this more recent, post-covid data is more relevant to your goals. We also excluded any unincorporated places as they represent only a minute chunk (~0.5%) of the data and they often do not appear in other data sources we chose to bring in. For example, in addition to alcohol sales, we were interested in the population and demographics of places where Iowans purchased alcohol, so we brought in data from the US Census Bureau on [Incorporated Places in Iowa from 2020 to 2024](#).

Using this data, we also created new variables for the analysis. We brought in data about the location of universities and colleges in Iowa to create a binary college town variable that is true if the city contains a college and false otherwise. In addition, we divided the total liters sold by the population of that city to create a liters per person variable since we were interested in exploring the amount of alcohol sold per capita.

To understand the nature of the data, we performed an initial exploratory analysis and found that for most types of alcohol, college students are the primary consumer purchasing more liters per person than non college cities. We also found that whisky and vodka alcohols are more popular than other types.

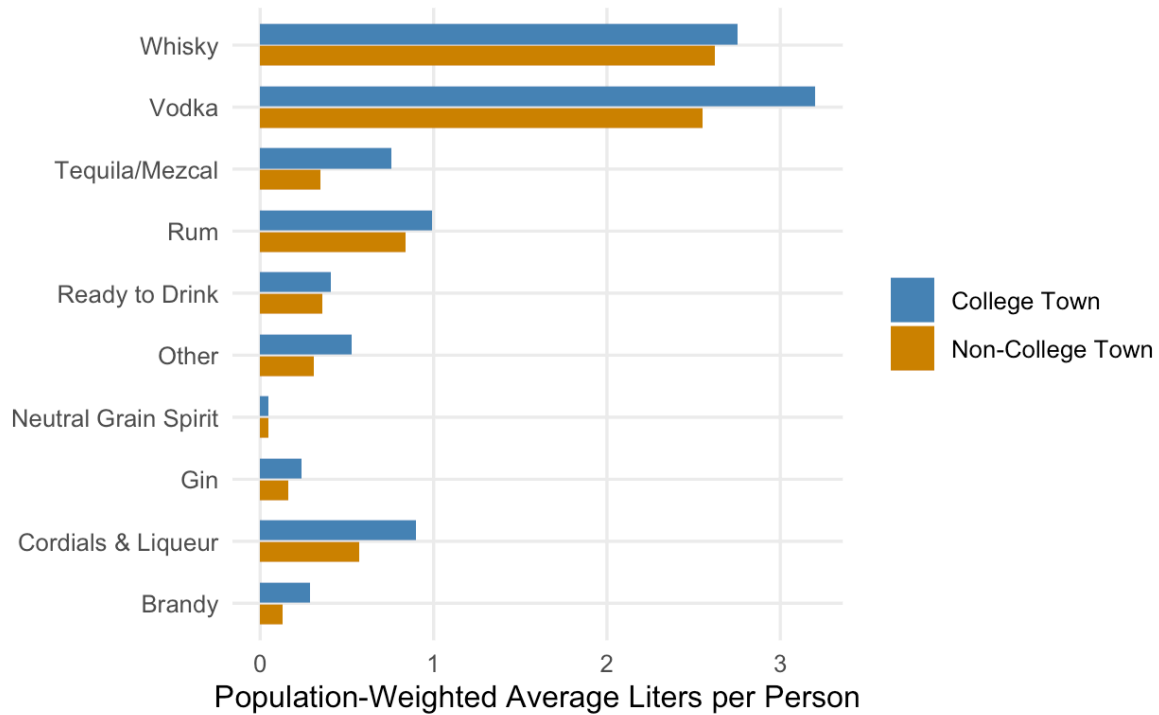


Figure 1: Graph comparing the liters of alcohol sold per person between types of alcohol and college town presence

Model Selection

To meet the goals of your organization, Drinking Excess Alcohol is Dangerous, we chose to create a linear regression model that will predict the alcohol sales in volume per person in a city. The model will use city demographics, college town information, and alcohol sales to make its predictions. Since you are looking to understand what factors drive higher or lower alcohol purchases to make alcohol culture safer in Iowa, the model will also contain Iowa specific predictors such as the county name. Our goal with the model is to discover which factors are the best predictors of high drinking rates, which allow your organization to have more targeted campaigning in those areas. To achieve this, we will be testing multiple models with various predictors using cross validation and the appropriate scoring metric.

Conclusion

Drinking Excess Alcohol is Dangerous is a crucial organization in the fight for safe alcohol culture in the state of Iowa. We are confident that with our analytical tools and industry expertise, we will be able to assist you in this endeavor. Through our analysis, we aim to

provide you with a robust linear regression model that will empower your company to determine the locales with higher risk for dangerous alcohol consumption, providing a data-driven foundation for your advocacy efforts.

We appreciate the opportunity to support you in this initial phase and look forward to continuing our collaboration as you fight to keep Iowa safe.