Booze 'R' Us 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 sale transactions from the Iowa Liquor Sales dataset. The orignal dataset spans from 2012 to 2025. However, we will be using the data from September 1st, 2020 to August 31st, 2025 as we believe this more recent, post-covid data is more relevant to your goals. We also excluded unincorporated areas as they represent a minute chunk of the data (~0.5%) and we believe they don't have enough additional data to result in accurate predictions. In addition to alcohol sales, we will be using Iowa population count from the Iowa Incorporated Places: 2020 to 2024 dataset. This dataset spans from July 1st 2020 to July 1st 2024.

Based on the existing the data, we created new variables to aid in the analysis. Using sale price and population, we calculated the sales per person. We also created a binary college town variable that is true if the city contains a college and false otherwise.

To better understand the data, we performed some initial exploratory analysis on data from September 2024 to September 2025. From our exploration, we found that college towns in Iowa tend to sell more alcohol per person than non college towns. Alcohol sales also vary by time of the year. More alcohol tends to be sold during the latter half of the year than the beginning of the year.

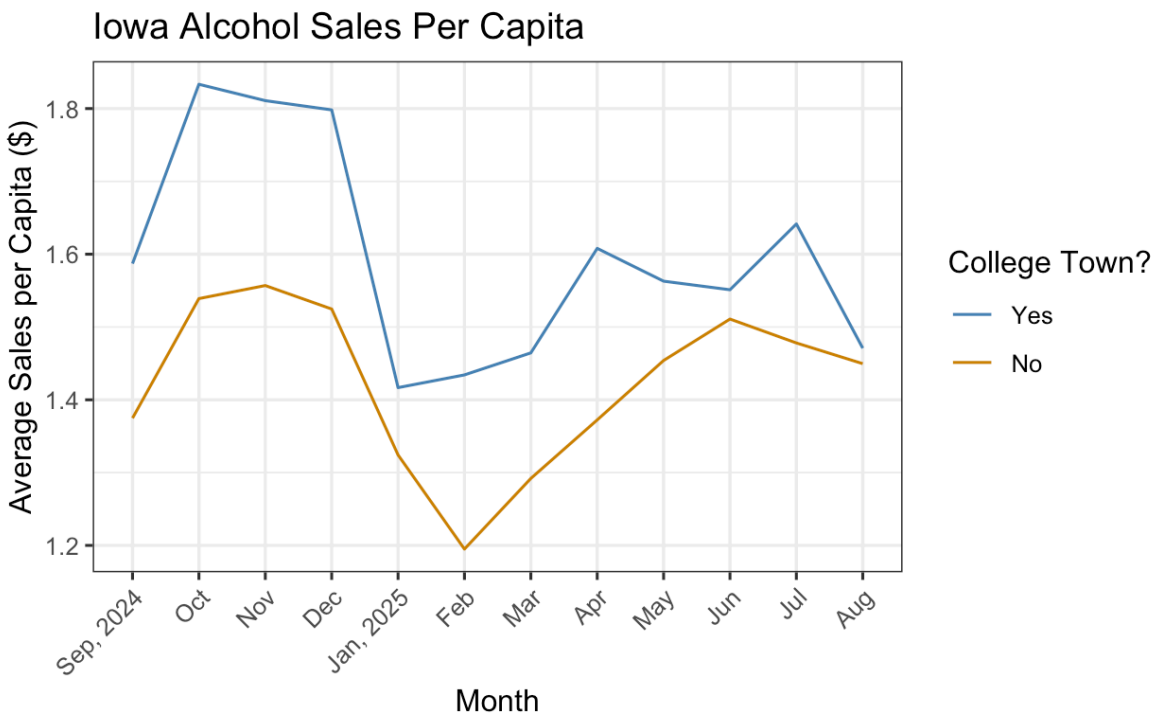

Figure 1: Graph showing average alcohol sales per capita by college presence over time

Model Selection

Using the data above, we will be creating a linear regression model to predict the alcohol sales per person of a city. Our model will use city demographics, college town information, and alcohol sales to make its predictions. We won't be including data specific to Iowan cities as we want our predictors to translate effectively to other states you are interested in expanding to.

To provide you with the best performing model, we will be testing various predictors with cross validation and a scoring metric.

Conclusions

Booze 'R' Us has a strong foothold in the state of Iowa, so we agree that this is the ideal time to explore expansion opportunities in new markets. We are confident that we have the analytical tools and industry expertise 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 predict the sales of alcohol in cities across the country, providing a data-driven foundation for your expansion goals.

We appreciate the opportunity to support you in this initial phase and look forward to continuing our collaboration as you expand your reach.