Analysis of Alcohol Sales Trends in Iowa

Anthony Rosamond[1], Jade Towgood[1], Jasmine Gomez[1], and Madison Brown[1]

[1]Arizona State University, Tempe, AZ 85281, USA

October 10, 2024

1 Abstract/Executive Summary

This study investigates Iowa's liquor sales trends and their connection to broader lifestyle changes and population shifts. By leveraging datasets such as Iowa Liquor Sales and Iowa Census data, the analysis sought to uncover correlations between alcohol consumption, population size, and income distribution. Through exploratory data analysis, supervised learning models, and time series forecasting, we examined sales patterns and predicted future consumption trends.

The results revealed that urbanized counties like Polk and Linn consistently reported higher alcohol sales than rural areas. Population size, rather than income, emerged as the key driver of liquor consumption in Iowa. This finding aligns with previous studies highlighting the influence of urbanization and population density on alcohol availability and social behaviors [23].

Seasonal trends highlighted peaks in sales during holiday periods, while the SARIMA time series analysis forecasted a potential decline in future sales [11]. This projected decrease is likely influenced by a combination of health-conscious behaviors and societal shifts post-COVID-19, which have led to increased awareness of alcohol consumption's health impacts [18, 26].

These findings carry significant implications for policymakers and public health advocates in Iowa. By providing insights into the role of population density and urbanization in driving alcohol consumption, the study lays a foundation for further research into how vendor accessibility, public health policies, and demographic factors influence consumption patterns. These insights will be valuable in crafting policies aimed at regulating alcohol sales while balancing public health concerns and economic interests [5, 25].

2 Introduction and Background

2.1 Introduction

The impact of alcohol plays a significant role in the social and economical landscape of the world. One place that gives us ample information on this background is the state of Iowa. As of 2022, Iowa has one of the highest binge, or excessive, drinking rates of alcohol in the United states, with 22.6% of the adult population reporting binge drinking habits whereas the national average is 18.4% in the U.S. [24]. This is the percentage of adults who reported binge drinking, which is classified as either: four or more drinks for females and five or more drinks for males on one occasion in the past 30 days; or as heavy drinking of eight or more drinks for females and 15 or more for males per week [10]. Due to the effects of binge drinking on public health, the state has started efforts towards addressing and limiting the excessive consumption of alcohol by their residents [13, 20]. As seen during the prohibition era, an entire ban on liquor is impossible for a multitude of reasons so a balance between promoting the sale of alcohol and still safeguarding public health is imperative.

When examining alcohol sales across Iowa there are a couple different variables we need to take into account that can affect liquor sales beyond individual consumption. For example, population density, average

age, urban versus rural communities, whether there are universities located nearby, churches, or if it is a tourist area. Separating the state into different areas, such as by county, can help shed light onto regions of high sales and help with analyzing the alcohol sales throughout it. This can allow us to see how any policies that Iowa has implemented have influenced the purchase and consumption of alcohol on a smaller level instead of statewide, whether those policies were to help increase sales, like allowing the sale of to-go cocktails during COVID-19 [21], or policies to restrict alcohol consumption. A few of the latter policies that Iowa has implemented statewide in this regard are: excise taxes and minimum unit pricing, and dram shop liability [20]. Using Iowa's alcohol sales dataset [7], will also help us with predicting future sales too.

Studying liquor sales trends allows us to view the complex relationship between consumer behavior, retailers, and public health. We seek to use Iowa's past alcohol sales to track any changes in trends and investigate the factors the may have led to them, such as whether it is the policies implemented, changes in income and population, lingering effects from the COVID-19 pandemic [18, 26], or just time influencing any changes we identify [16].

2.2 Background

2.2.1 History and Context

Alcoholic beverages have been consumed for thousands of years, and in the United States, alcohol consumption is prevalent across various regions and populations [24]. Throughout U.S. history, attitudes and laws regarding alcohol sales and consumption have fluctuated significantly due to cultural, social, and legal factors. In comparison to the rest of the world's standards on alcohol and its regulation, the U.S. differs in our unique approach over the years. When looking at the 194 member countries of the WHO, there are only 64 that have drinking recommendations, but only four have comprehensive ones, including the US making them stand out, along with their minimum age requirement of 21 compared to other countries being 18 or even 16 in some [4]. Despite this unique approach, we score in the middle of the study that included 30 countries, when it comes to alcohol policies based on a study done in 2007 [2]. There have been multiple moments where either individual states or the federal government have imposed severe restrictions on the sales and consumption alcohol and the state of Iowa has been no exception.

Since 1846, when Iowa was granted admission to the Union, the topic of alcohol regulation has been a significant political issue. Early efforts to regulate alcohol consumption was spearheaded by the Women's Christian Temperance Union of Iowa, which led to members of the Iowa Republican Party to introduce and pass a state amendment imposing a total ban on alcohol [17] in 1882. However, this would eventually be dismissed by the state's Supreme Court.

Alcohol remained tightly regulated within Iowa, and in 1920 the Eighteenth Amendment to the U.S. Constitution was passed which banned the production, transportation, and sale of alcohol nationwide, also know as Prohibition. This would remain in effect until the Twenty-First Amendment was passed in 1933, repealing the Eighteenth Amendment. Even with this repeal, some states still continued to enforce their own prohibition laws; however by 1966, all states, including Iowa, had repealed their prohibition laws [9].

Now, the modern alcohol industry in the state of Iowa is booming with 100 breweries, 120 wineries and 45 distilleries as of 2023 [17]. Excessive drinking has historically been more predominant in Midwestern states, with Iowa ranking number 4 in the nation when it comes to binge drinking rates [24]. As binge drinking is directly linked to health issues such as chronic disease and compromised immune systems, this has the potential to become a large public health issue [24]. Part of the state of Iowa's focus on addressing alcohol consumption has led to detailed publication of liquor sale transactions across the entire state [7], which is a major focus of our study.

In addition to the publication of sales, Iowa has enacted some policies to help with limiting the consumption/sale of alcohol on the state and local level. On the state level there are three main policies they have put in place. The first is an excise taxes and minimum unit pricing. This has shown to have a targeted effect on high-risk drinkers [20]. The second is dram shop liability to ensure that establishments can be held liable for over-serving a patron and third, limiting the days and hours of sale, which has been shown to reduce excessive alcohol consumption [20]. On the local level some communities have enacted conditional use permits, nuisance ordinances, alcohol advertising ordinances, enforcement, screening and intervention, and responsible beverage service (RBS) training [20]. Conditional use permits are mainly used to reduce establishment clustering, and sometimes to place restrictions on alcohol establishments with a history of problem incidents [20]. Enforcement is used in the form of bar checks for selling to minors and has been shown to reduce excessive drinking [20]. Screening and intervene, also known as e-SBI, is used to identify individuals that may have or are at risk of developing an alcohol use disorder and intervene by describing the risks of excessive drinking [20]. Finally, RBS training are for all employees selling and/or serving alcohol to identify ID cards correctly, help recognize intoxicated patrons, and how to intervene and not allow them to drive while under the influence [20].

2.2.2 Key Sources and Papers

In order to analyze the alcohol sales trends in Iowa and the broader implications they may have, we need to delve into several external factors and trends that have influenced consumption patterns. Looking into other key sources and papers that already exists will help us in this study. The Iowa Liquor Sales Data is an essential tool for understanding alcohol consumption patterns across various regions, offering a detailed view of how factors like urbanization and economic conditions influence liquor sales [7]. Similarly, the Iowa Enterprise Data Liquor Sales Snapshot provides real-time, monthly data from January 2012 through August 2024 that tracks sales by product type and region [6]. This dataset is valuable for analyzing seasonal shifts in drinking habits, especially around holidays or local events.

When analyzing population dynamic and urbanization, the shifts between rural and urban areas tend to have a significant impact on alcohol sales, according to the study published by the Public Science Collaborative on Iowa [22]. This study found that urban regions tend to have more stores selling alcohol and a higher sales volume in turn due to a denser population and accessibility. Furthermore, they found that there was a direct correlation between the density of liquor stores and consumption patterns [22]. In addition, a separate study also found that urbanization in Iowa has contributed to the increased alcohol availability with more stores, while more rural areas faced reduced access [12].

Looking at the effects of COVID-19, a study done by Columbia University found that the pandemic led to a surge in at-home consumption of alcohol [26]. The pandemic created a disruption in traditional consumption patterns, and is an external factor that needs to be taken into account. A study published by the NYC Data Science Academy also noted how the seasonal sales trends were disrupted during the pandemic and became steady instead of spiking during the holidays due to the restrictions [12]. This study also noted how online alcohol sales became more prevalent during this period [12].

Taking into account the "Sober Curious" movement [18], it emphasizes that younger generations are exploring low to no alcohol content drinks for health and wellness concerns. This article also notes that this trend is evident in some regions of the state where you can see a decline in sales by the younger demographic [18].

In conclusion, the alcohol sales trends in Iowa are shaped by multiple factors that are always evolving. Taking these factors into account are imperative for business owners, policy makers and for public health and wellness in general to understand the long-term implications of these trends.

2.2.3 Key Controversies

Several key controversies related to the study of alcohol consumption including alcohol outlet density, the regulation of alcohol sales, and the increasing amount of alcohol consumption in rural areas. The first is an essential concept within the study of alcohol sales trends and consumption and tends to be related to the density alcohol outlets [5]. A key assumption is that increased availability leads to higher consumption and, subsequently, alcohol related issues [1]. This leads to public health concerns about alcohol abuse and according to the Iowa Alcohol Outlet Density report, the rapid growth of liquor stores in neighborhoods has created fears of higher alcohol related harm [22].

Another controversy is binge drinking, which is an issue as touched upon earlier. Iowa has one of the highest rates of binge drinking in the country and public health officials have been broadcasting the negative long-term effects from it [22]. They have tried to mitigate this amount by implementing new or updated policies. These local policies include conditional use permits, noise ordinances, alcohol advertising ordinances, enforcement in the establishments and outside, screening and intervention, and responsible beverage service (RBS) training [20]. On a state level, Iowa has excise taxes and minimum unit pricing to target the highest-risk drinkers, dram shop liability policies for establishments over serving a patron, and limiting days and hours of sales from 6:00 AM to 2:00 AM everyday [20]. With the implementation of these policies, which were described in more depth previously, there has been arguments that these unfairly penalizes responsible drinkers and harms local businesses.

In addition, rural areas in Iowa are seeing shifts in alcohol consumption patterns that are seemingly linked to social isolation and economic stress. This has led to more alcohol related issues, like drunk driving and alcohol abuse in older adults. These trends raise questions on whether or not the current policies are actually helping address the unique challenges in both urban and rural areas [12].

2.2.4 Open Ended Questions

Researching the topic of alcohol sales raises questions on if and how Iowa's demographics, such as population size and spread, ties into it's high alcohol consumption. Another relevant question is how liquor sales over the years have changed, especially in relation to the implementation of different regulatory policies in Iowa. The goal of this question is to determine if these new changes are creating the desired effect on consumption patterns among residents.

Other questions were raised that may require more study outside of our project's scope. One is surrounding the role of mental health on one's likelihood of binge drinking and alcoholism. Whether these factors also influence alcohol-related mortality rates may cause Iowa to start implementing mental-health related changes to their laws and policies to address this related challenge. Another potential area of investigation is surrounding Iowa resident's access to programs or facilities for recovering from alcoholism, as well as the quality of these resources. Regional liquor sales may reveal where residents buy less alcohol, leading to a potential area of study surrounding what social neighborhood influences cause consumers to buy less alcohol. Similarly, a final open-ended question is whether there are any major societal trends over the years that impacted alcohol sales.

2.3 Methodologies in Literature

The two most current studies related to studying the alcohol sales trends and consumption in Iowa were published by the NYC Data Science Academy and the Public Science Collaborative. The NYC Data Science study, which was done from 2012-2020, focused primarily on liquor sales trends, store dynamics and product distribution across different stores. This study used k-means clustering and ARIMA forecasting models to analyze the variables: sales trends, store competition, and product variety [12]. The NYC Data Science Academy found that with the increase in stores, there has been a total increase in overall sales, but also a decrease in sales per store, most likely due from the increased competition. They also found that both liquor and convenience stores sell similar products adding to the competition. This analysis also showed seasonal peaks in certain types of liquor—tequila and cocktails in summer, brandy and liqueurs in winter. Finally with their forecasting, they noted predictable seasonality and demand around the holidays and summer [12].

The Iowa State University study took a more social perspective to their analysis and used geographical maps to analyze their variables. They took into account the density of liquor outlets and their relationship with health, crime, and consumption patterns in Iowa. They created visual geographic distributions to display the implications for public health and crime [22]. These visualizations show that the areas with higher alcohol outlet densities correlated to higher alcohol consumption and related social issues, such as crime and health risks. They also found the urban areas had more alcohol outlets per capita compared to rural areas, drawing them to conclude that this factor influences local consumption patterns and societal behaviors [22]. These reports both complement each other with their findings, with the NYC study looking at detailed sales trends and Iowa State's study focusing on outlet density's effects on different communities.

2.4 Project Aim

Due to alcohol's impact on society and the economy as a whole, our motivation for this project is to understand how societal and lifestyle changes, forced and not, are reflected in these sales trends and possibly predict future sales. We aim to equip policymakers in the government, public health advocates and professionals, and business owners and makers with the knowledge they may need to create informed decisions about alcohol sales as well as other decisions regarding the people of Iowa and their evolving dynamics.

Our main focus is: **How can analyzing alcohol sales trends provide insights into broader lifestyle changes and shifts in Iowa's residents?** We specifically are focusing our study on uncovering the relationships between alcohol sales data [7] and a number of other factors—demographics [3], economical changes, and cultural movements [16]—to better understand the underlying societal shifts in Iowa.

Another question we want to answer in addition to our main focus is: *How have external events, such as the COVID-19 pandemic, affected alcohol trends in Iowa?* One could assume that during and in the aftermath of the pandemic the sales and volume of alcohol sold has increased significantly with people being forced to stay home during the lockdown and becoming accustomed to not going out as often to drink socially [26]. With this study, we should hopefully be able to see if this assumption is founded.

One final questions we look to answer as well is: *What role do demographics, such as income and age, play in influencing alcohol sales and consumption trends?* We assume that both income and age will play a significant role on these trends. For example, we would imagine that residents with more disposable income are more likely to spend more money on purchasing alcohol—in quality and quantity [14]. Also that young adults, aged 21 to 28, would have a higher consumption rate than other age ranges, so not necessarily in price but quantity.

3 Methods

3.1 Data Sources

The two main datasets for this project will be the Iowa Liquor Sales Data and the Iowa Census Data. The first dataset, Iowa Liquor Sales Data [7], is published and maintained through the Iowa Department of Revenue, Alcoholic Beverages and contains data starting from January 1, 2012 to the current month. It is typically updated the first of every month. As of September 1, 2024, there are 29.9M instances representing individual product purchases in this dataset with 24 different variables. The 24 variables are as follows: Invoice/Item Number, Date, Store Number, Store Name, Address, City, Zip Code, Store Location, County Number, County, Category, Category Name, Vendor Number, Vendor Name, Item Number, Item Description, Pack, Bottle Volume (ml), State Bottle Cost, State Bottle Retail, Bottles Sold, Sale (Dollars), and Volume Sold (Gallons).

The variables that we will be focusing on this dataset are Date, County, Category Name, Item Description, Sale (Dollars), and Volume Sold (Liters). The Date is the date of the order in month/day/year format. The county is where the store is located. Category Name describes the type of the liquor that was purchased. The Item Description is the description/name of the individual liquor product purchased. The Sale (Dollars) encompasses the total cost of the liquor ordered, which is calculated using the number of bottles multiplied by the state bottle retail. Finally, the Volume Sold (Liters) is the total volume of liquor ordered in liters, which is calculated by the bottle volume (ml) by bottles sold divided by 1,000. We dropped the other columns from the dataframe in order to have more workable data.

Using `https://data.census.gov/` we were able to gather data from multiple datasets [3] regarding different Iowa demographics. For this project, we gathered information about population size, income, and age. It is separated by year and can be filtered through individual county, which we will be using to merge with our alcohol sales dataframe. We also acquired a county boundaries dataset that is on Iowa's geodata website that contains the county FIPS codes for creating geographic visualization.

3.2 Exploratory Data Analysis

The initial steps of our exploratory data analysis was to clean the datasets to make sure we had no irregularities such as null values. We then properly formatted the data so we could use it with Python. This included renaming columns, dropping unneeded features, and changing data types from type "object" to numerical. For the population data, we also appended a new column that was a calculated average of the other columns. Since each column in this dataset recorded the population of a county for a given year, this new column is the average population for a given county over the time frame from 2010-2023.

For the alcohol sales data, we decided to trim down the number of instances we would be working with since it contained 29.9 million records. We do not have the computational power to efficiently analyze such a large dataset, so we took a random sample of 500,000 records. This would be the alcohol sales data that we would be working with moving forward. This data was actually fairly clean and only a little over 2,000 records had missing/null values, so we decided to remove them since they constituted statistically negligible portion of the data.

For the actual exploratory data analysis, we focused on understanding the inherit distributions within our different datasets. For instance, we created histograms to observe the distribution of certain features, such as income. We also focused on plotting trend lines of the average population for counties, ultimately focusing on the top 10 most populous counties. We then created scatter plots to see the visual relationship between alcohol sales and different variables, such as population and income. Finally, we used the geopandas library to create geographic representations of these distributions. The results of this exploratory data analysis are featured in section 4.1.

3.3 Supervised Learning

For our supervised learning methods, we focused our study on time series analysis for forecasting sales and two different regression models to understand the relationship between different features and alcohol sales. To forecast alcohol sales using time series analysis, we decided to utilize a SARIMA model. We first plotted the monthly average sales in order to visually inspect the sales data and see what kinds of trends were present, such as general growth or whether certain months had spikes in sales. We then checked for stationarity using the Augmented Dickey-Fuller test so we can ensure a SARIMA model could be used with our data. We realized that we needed to make the data stationary by removing the trend component. We then used autocorrelation and partial autocorrelation in order to determine the parameters for our model. Once we had the optimal paramters selected, we trained the model and then plotted its results compared to our actual sales data.

We also employed both a linear regression and a random forest regression model in order to see if we can predict alcohol sales given certain demographic parameters. In order to create these models, we had to transform the data in order to prevent any inconsistencies. We scaled all numeric data using the standard scaler. This ensures that the varying scales of numeric data don't interfere with each other. We also needed to transform the categorical data into numeric data. This was done using the one hot encoder, which is suited to handle categorical data where we wish to reduce the issue of ordinality. Model performance was evaluated using the root-mean-squared error which is calculated by taking the square root of the average of the difference squared of the calculated and actual value.

3.4 Unsupervised Learning

For unsupervised learning we performed a cluster analysis on the data. The goal was to see if any patterns would emerge that wouldn't have been apparent through traditional EDA or the supervised learning methods. The main clustering technique we used was the K-Means method. This method clusters data points around specific points called centroids, where the variance within each cluster is minimized. In order to get good clusters we performed dimensionality reduction using principal component analysis (PCA). Before that was done the data was put through the standard scaler since it contains different scales of measurements. This scaled data was then transformed using PCA which reduced it to two columns of data.

Once we had this transformed dataset, we needed to decided how many clusters were appropriate for the analysis. We used the elbow method, which involves plotting the within-cluster-sum-of-squares vs. the number of clusters. Then, we performed the cluster analysis. We then took the cluster assignment for each data point and added it to the data frame from before the PCA transformation. From there we then analyzed the summary statistics for each cluster to see how each one differed from each other.

4 Results

4.1 Exploratory Data Analysis

4.1.1 Liquor Sales

Since a portion of our project investigation was to determine which external features or demographics influence sales, we identified two key features early on within the Iowa liquor sales dataset: alcohol sales and volume of alcohol sold. In Figure 1, we plotted the distribution of both of the features after scaling by population.

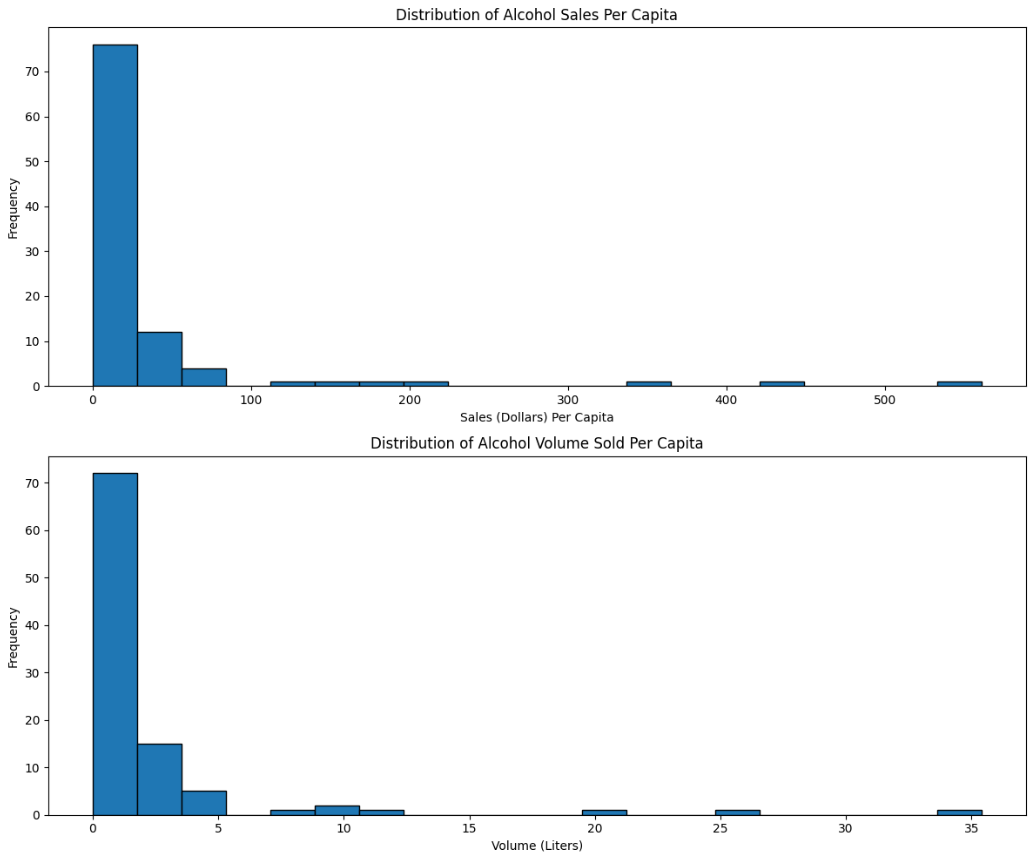

Figure 1: Distribution of alcohol sales per capita and alcohol volume sold per capita.

This resulted in the distribution of alcohol sales per capita for the top figure and the distribution of alcohol volume sold per capita for the bottom figure. Though each has a few outliers, both figures depict high frequency at the lower end on the scale.

To understand this distribution better, we recreated the distribution plot of alcohol sales per capita on a smaller scale to better understand the spread. This can be seen in Figure 2, which shows that the average sales transaction per capita is $15. Similarly for the distribution of alcohol volume sold per capita, we recreated the distribution plot on a smaller scale. This is shown in Figure 3, which shows that the average volume of alcohol sold per transaction per capita is one liter or less.

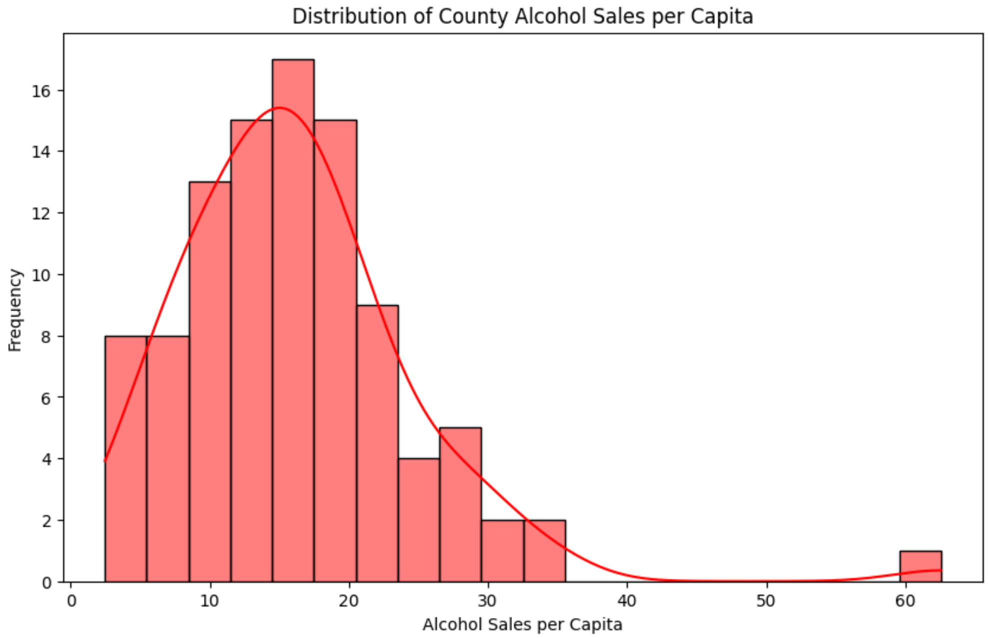

Figure 2: Distribution of alcohol sales per capita.

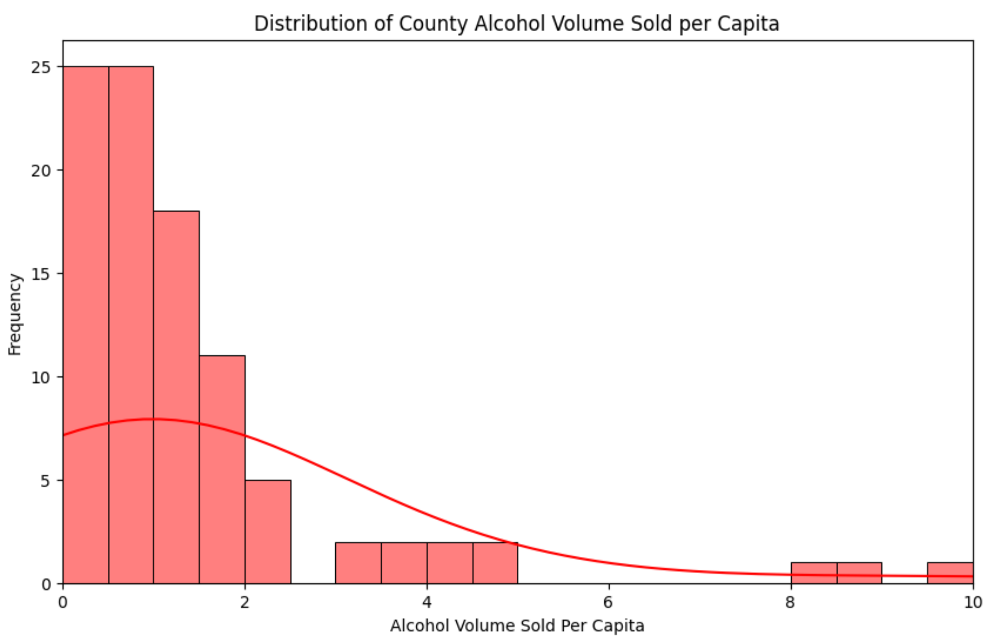

Figure 3: Distribution of alcohol volume per capita.

We also created two geographic visualizations to display the similarities between the alcohol sales per capita and volume of alcohol sold per capita. Figure 4 and Figure 5 showcase the visual correlation between the two variables.

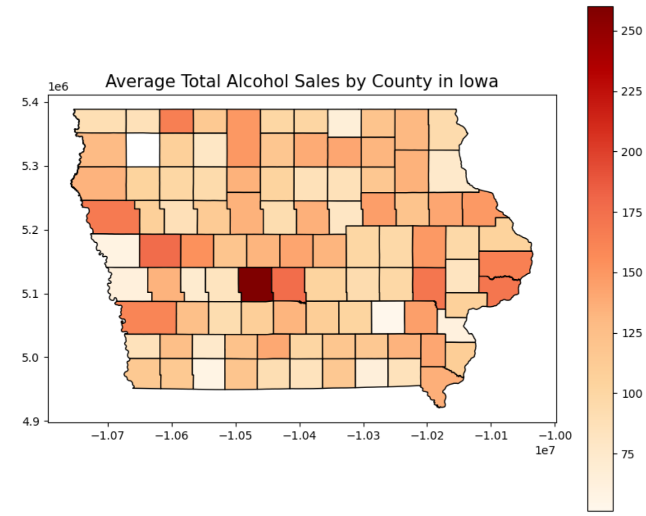

Figure 4: A geospacial map to show the spread ofalcohol sales across counties in Iowa percapita/average

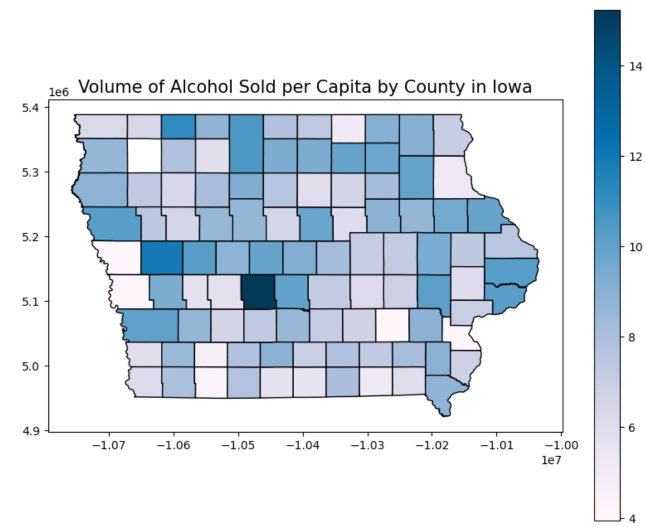

Figure 5: A geospacial map to show the spread of volume of alcohol sold across counties in Iowa per capita/average

4.1.2 Income

For the Iowa income data, we plotted in Figure 6 the distribution of average total income across all counties for 2010-2014, 2014-2018, and 2018-2022. These periods were inherited from the Census data, which averaged yearly income data from the American Community Survey (ACS) across four-year periods [3]. The average income for each period is \$50,000, \$55,000, and \$67,000 respectively, displaying an increase in average income across the twelve-year period in Iowa.

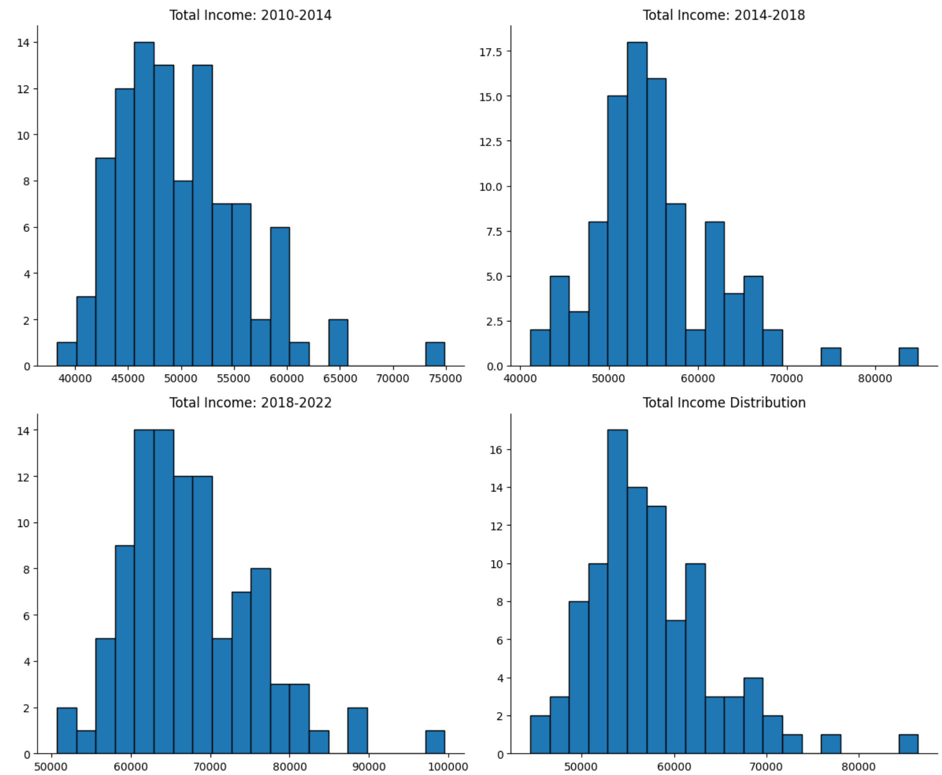

Figure 6: Four graphs displaying the distribution of total income for a) 2010-2014, b) 2014-2018, c) 2018-2022, d) across all years.

The last histogram is the calculated distribution of average total income across all counties and all years. This same histogram is depicted in Figure 7 with an additional distribution curve. For this twelve-year period and across all Iowa counties, the average income is \$58,000.

Our final visualization for the income dataset is Figure 8, which is a geospacial map of the income per capita across all years. The highest concentration within this map is near Dallas County, whereas across the rest of the counties is relatively similar.

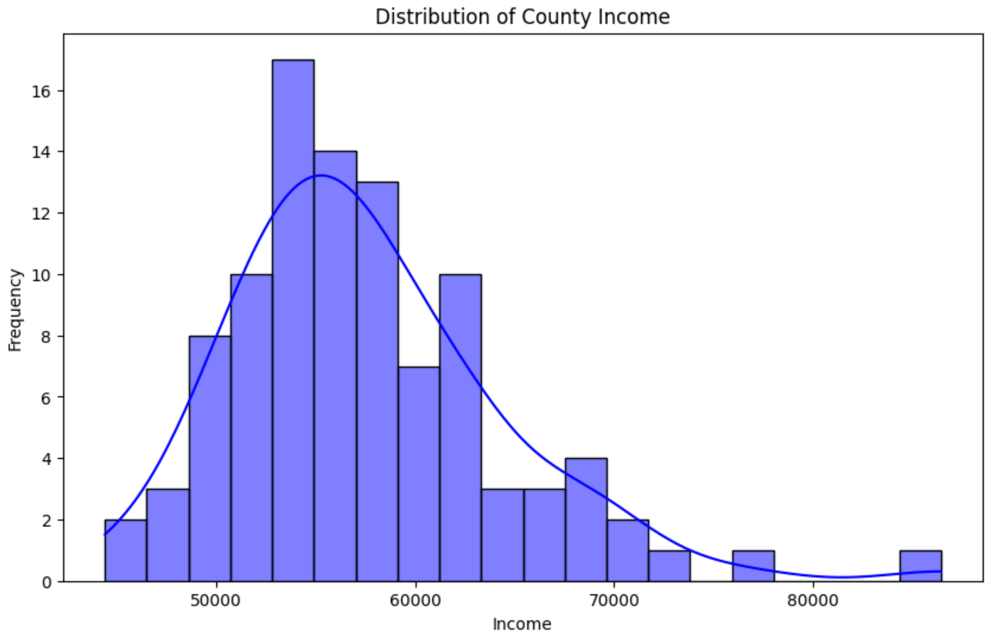

Figure 7: Distribution of income per capita for all counties across 2010-2022.

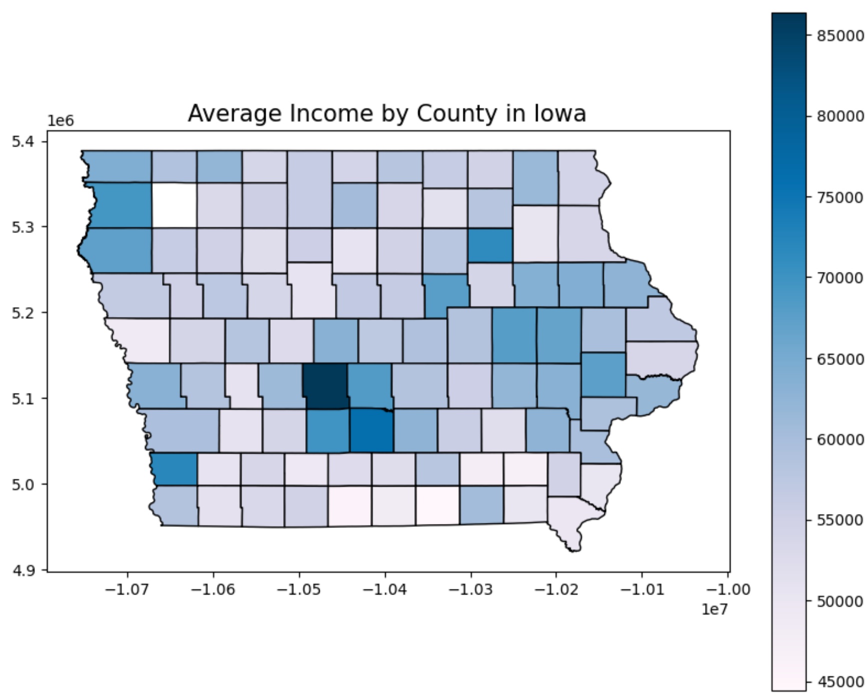

Figure 8: A geospacial map to show the spread of income across counties in Iowa per capita/average

4.1.3 Population

As part of the initial investigation into Iowa population within each county, we decided to see how the population has changed. Initially, we plotted the average population for each county for each year as shown in Figure 9.

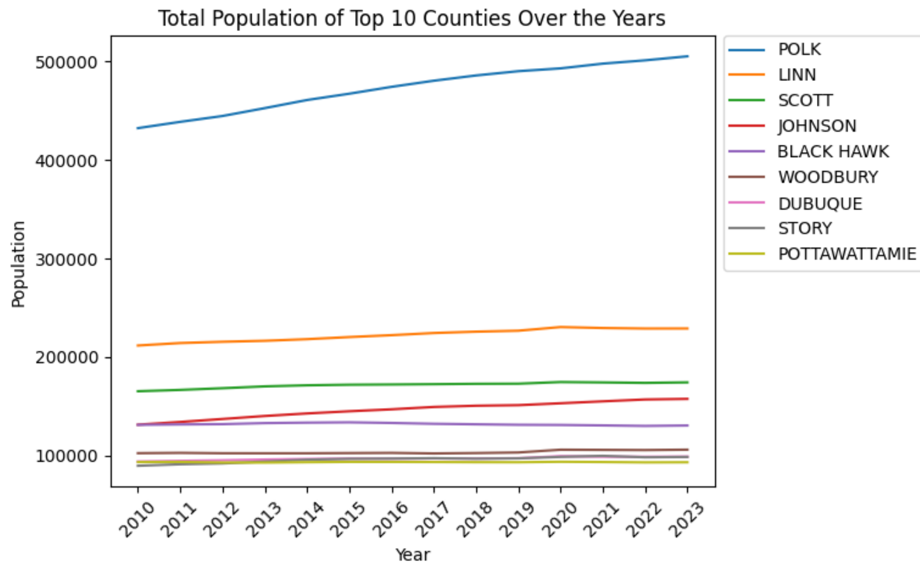

Figure 9: Total population count for the top ten counties.

As displayed in Figure 10, we calculated the growth rate across all counties across 2010-2023. The distribution is right skewed towards the negative values, which means population in Iowa counties overall slightly decreased.

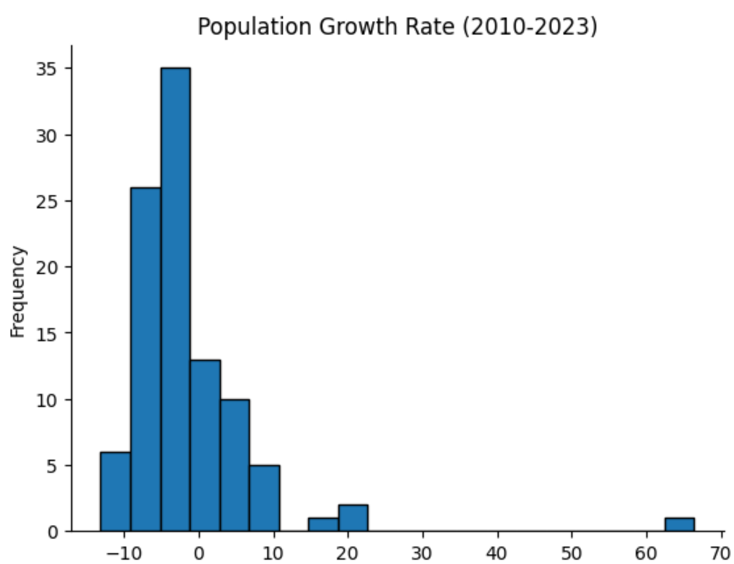

Figure 10: Distribution of population growth rate across Iowa counties.

However, Figure 11 displays the total population for every year and does display an increase in total population, suggesting that the increase in Iowa's total population was concentrated to few individual counties, such as Polk county.

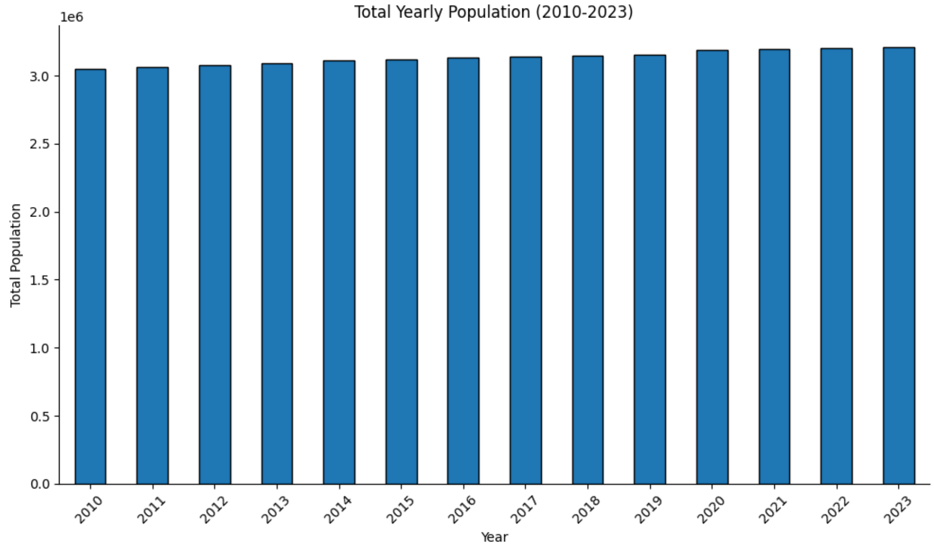

Figure 11: Total population count across Iowa counties.

This is demonstrated within Figure 12, which shows the top ten counties that did see an increase in population during the time period. There was a total of twenty-five counties which did see a positive increase in their population size, though only the top ten shown in Figure 12 showed more than a 5% increase in growth.

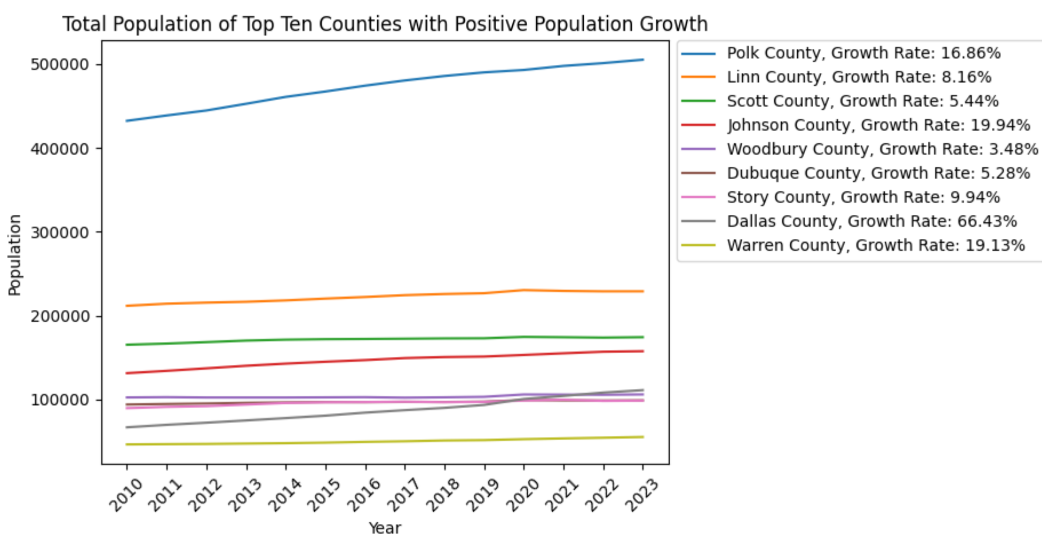

Figure 12: Total population count for the top ten counties with a positive population growth, including the growth rate.

4.1.4 Multivariable Analysis

Additionally, we decided to analyze the correlation between alcohol sales with population and income. The scatter plot of alcohol sales and population is shown in Figure 13. The correlation between these two variables is 0.46, which is a moderately positive correlation. Figure 14 displays the scatterplot of alcohol sales and income. The correlation between these two variables is 0.05, which is a very weak positive correlation.

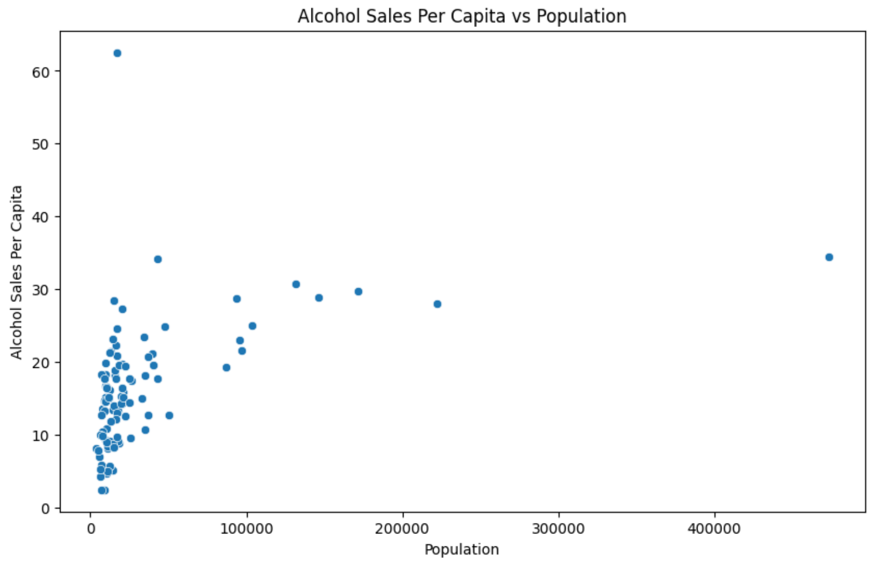

Figure 13: The correlation between Alcohol Sales per Capita and Population is 0.46.

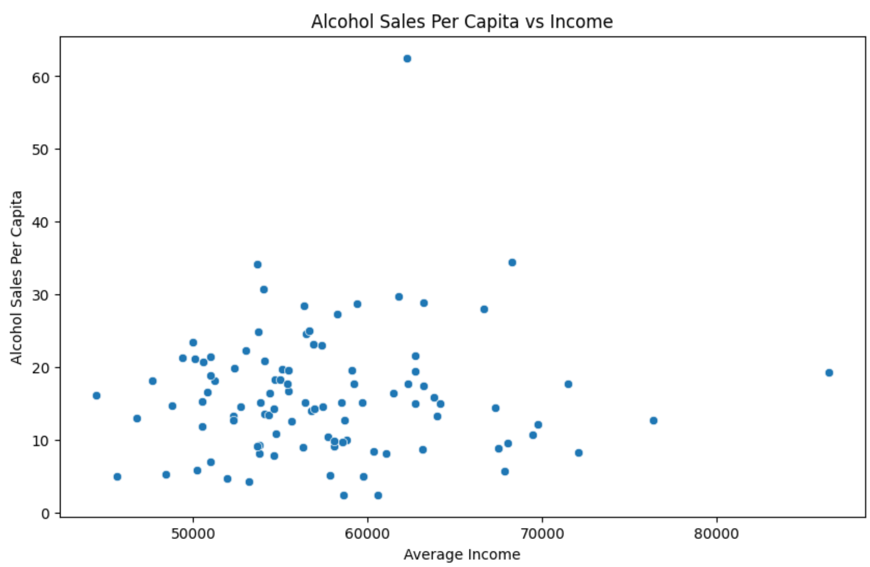

Figure 14: The correlation between Alcohol Sales per Capita and Income is 0.05.

4.2 Results: Time Series Analysis

The purpose of the time series analysis was to investigate any potential trends within the monthly sales data and to develop a model to predict future sales. The first step was to visually inspect the sales data, which is plotted in Figure 15.

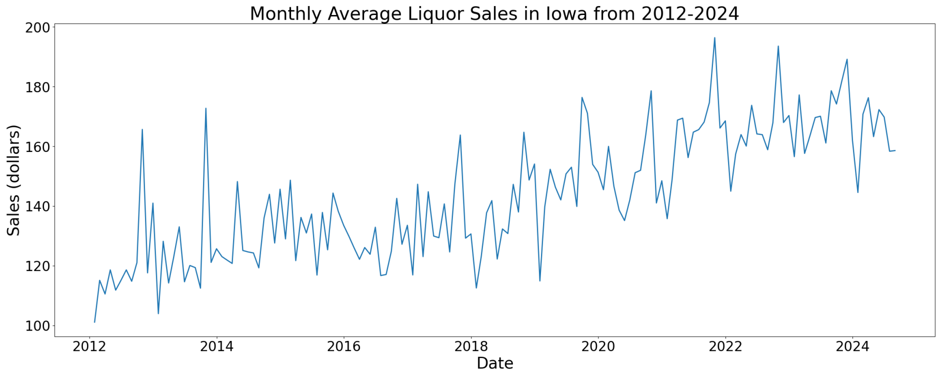

Figure 15: Monthly average liquor sales in Iowa.

To better visualize the components that make up this sales data, the seasonal_decompose function from the statsmodels library was used to break down the sales data into seasonal components. Figure 16 confirms that the dataset contains both trend and seasonal components. In particular, the sales data shows an upwards trend. The seasonal component confirms that there is a yearly spike surrounding the end of year holidays.

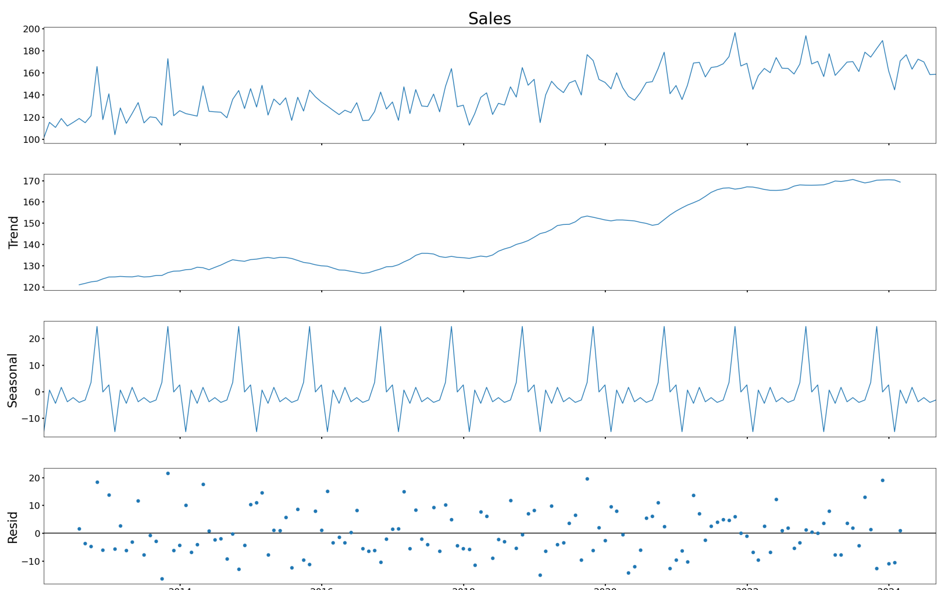

Figure 16: Trend, seasonality, and residuals of the time series against the full liquor sales.

After determining the best parameters for a SARIMA model utilizing the auto_arima function from the pmdarima library, the model was created and trained on the sales data. Then, the most recent twelve months on data was used to test its performance. The model's forecasted sales data compared to its training sales data is shown in Figure 17, and the model's forecasted sales data compared to its testing sales data is shown in Figure 18. The mean absolute error of the model was 9.539 and the root mean squared error of the model was 10.953.

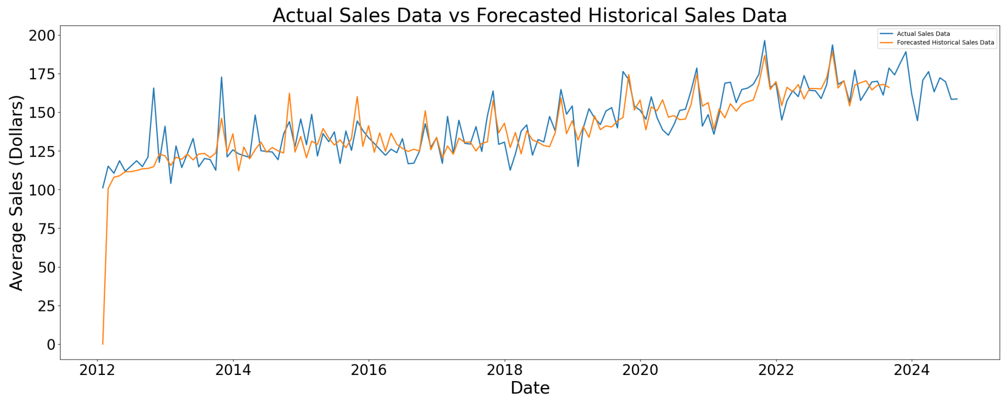

Figure 17: SARIMA model's forecasted historical sales over entire sales data.

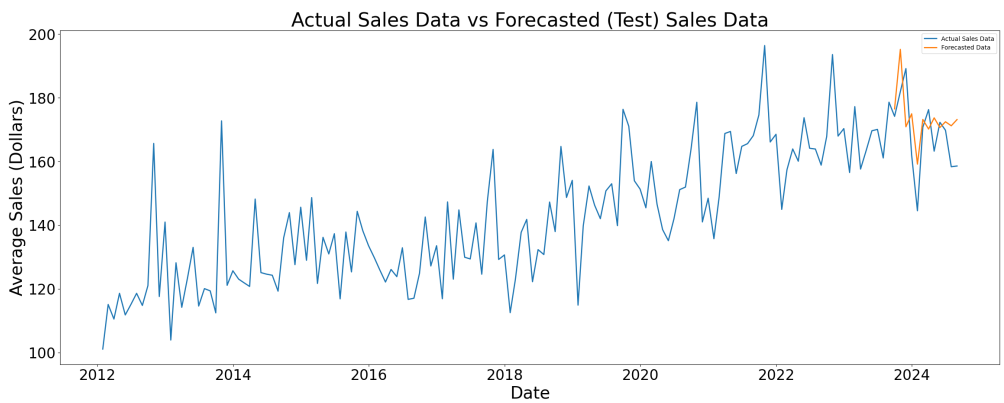

Figure 18: SARIMA model's forecasted sales over testing sales data.

The final portion of the time series analysis was to use the model to predict future liquor sales in Iowa. As shown in Figure 19, the SARIMA model predicts a decrease in future liquor sales.

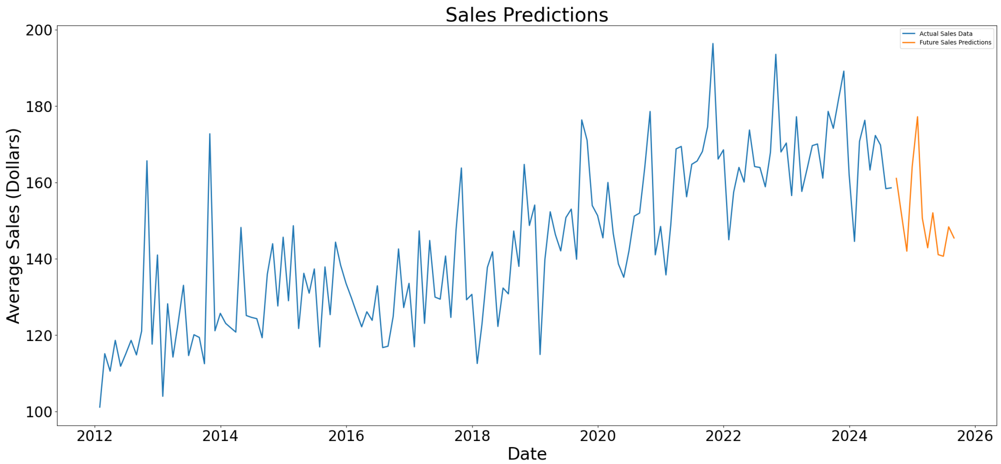

Figure 19: SARIMA model's future sales predictions for the next twelve months.

However, in Figure 20 there had been a slight decrease in sales in 2022 compared to 2021, yet the increasing trend in sales continued in 2023.

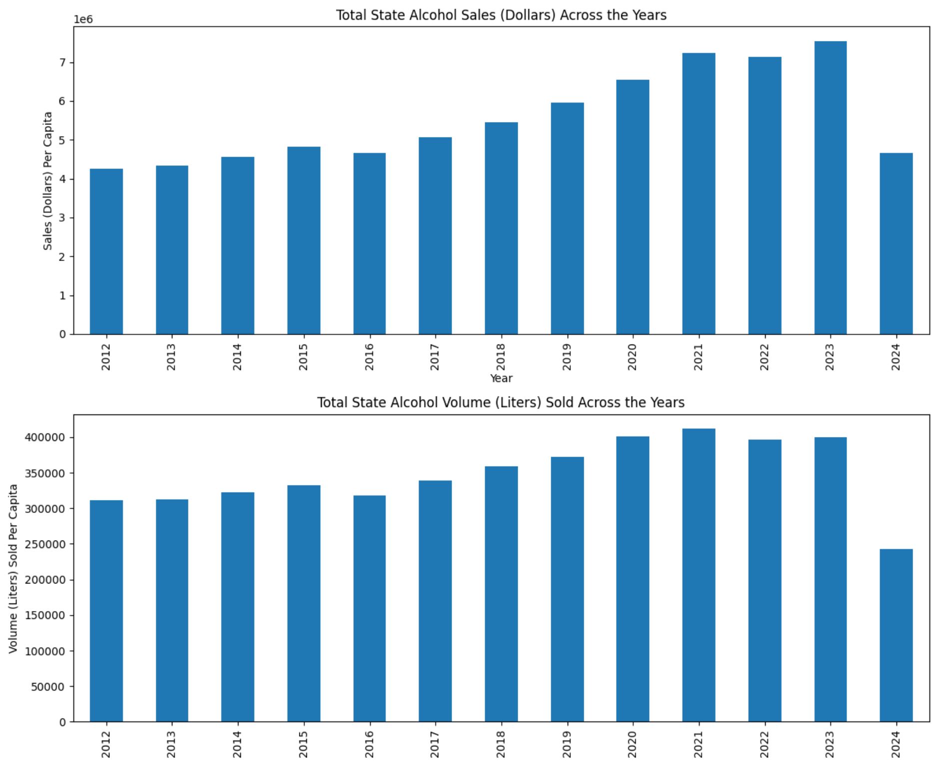

Figure 20: Total state alcohol sales and total state alcohol volume sold across the years.

4.3 Results: Supervised Machine Learning Models

Beyond the previously discussed features of population and income, we identified additional features within our existing sales dataset and additional census data to use within the supervised regression models. These relevant features include:

- average income
- average population
- average volume (liters) of alcohol sold
- total number of alcohol vendors
- average cost of one bottle of alcohol
- total number of different alcohol categories sold
- total number of religion congregations
- percent of population that adhered to a religious congregation,
- median age
- total number of universities
- majority urban versus rural classification
- percentage of population in urban areas
- percentage of population in rural areas

.

The aim of the regression models was to predict the Average Liquor Sales (dollars). Before creating the models, the features were split between numeric and categorical, and scaled using a Standard Scaler for the former and encoded with One Hot Encoding for the latter.

The two types of regression models utilized were linear regression and random forest regression. The linear regression model had a root mean squared error of 11.650 and an R2 score of 0.819. The random forest regression model had a root mean squared error of 11.705 and an R2 score of 0.817. The random forest feature_importance method was used to identify the most influential feature in this model, which was the Average Volume Sold (liters) as shown in Figure 21.

Since this feature significantly overshadowed the other features in this model's predictions, the models were recreated dropping the average volume (liters) feature. The idea was to see which other features impacted liquor sales.

This significantly impacted the model's performance in prediction sales based on its features. The second linear regression model had a root mean squared error of 25.399 and an R^2 score of 0.140. The second random forest regression model had a root mean squared error of 27.072 and an R^2 score of 0.0224. For the second random forest regression model, the feature_importance function revealed that the total number of alcohol vendors, the total number of alcohol categories sold, and the average cost of a bottle of alcohol to be the top three features. The new ranking is depicted in Figure 22.

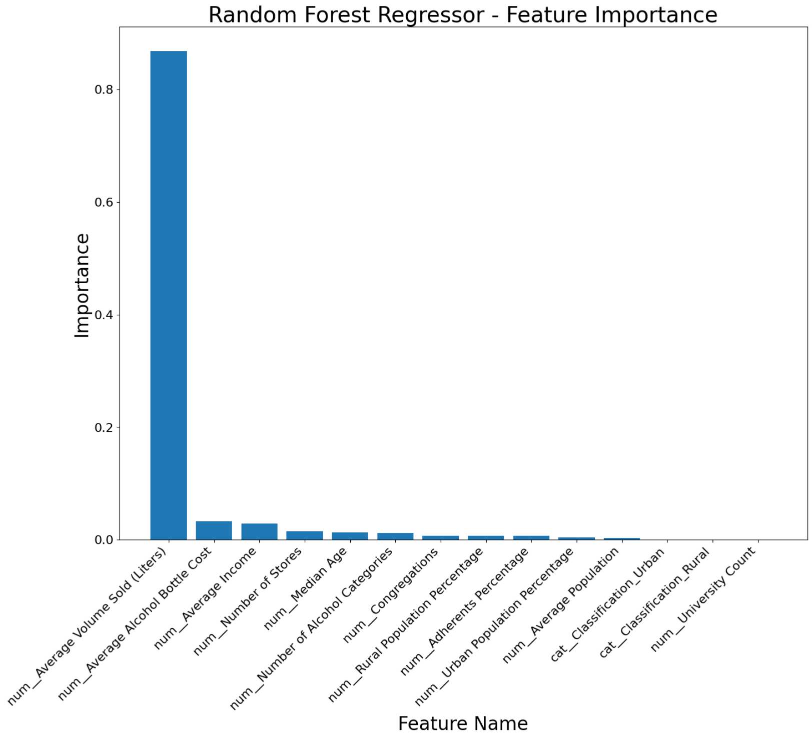

Figure 21: Random Forest Regression feature importance ranking including Average Volume (liters) of Alcohol Sold as a feature.

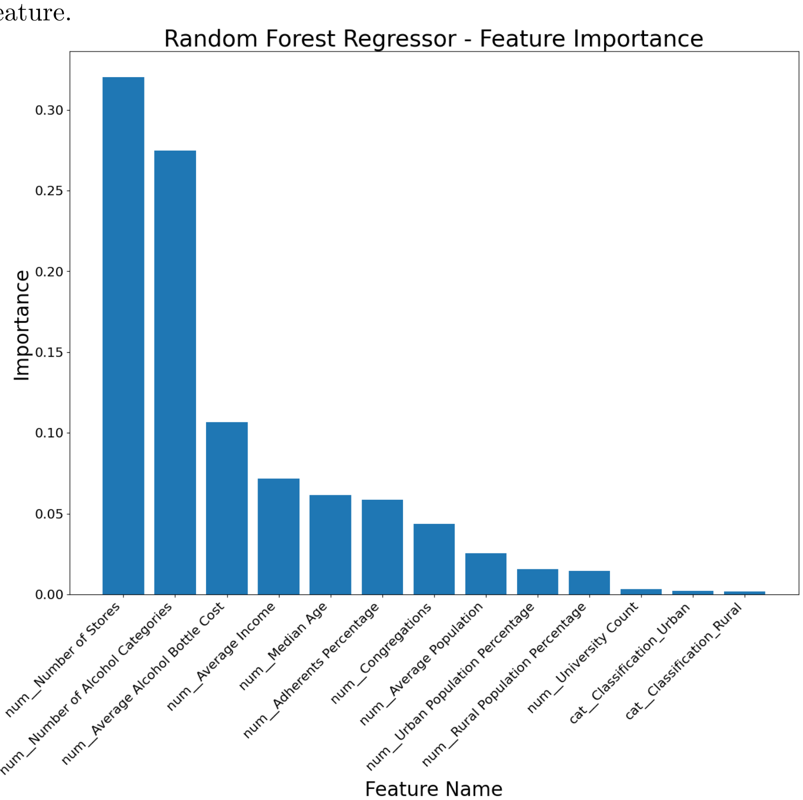

Figure 22: Random Forest Regression feature importance ranking after removing Average Volume (liters) of Alcohol Sold as a feature.

4.4 Results: Unsupervised Machine Learning Model

Our clustering analysis didn't quite yield as much as we had hoped for, though it did give us a little deeper insight into the sales of alcohol. The clusters that were formed mainly segmented the distribution on the total cost per transaction. We had one cluster of a small amount of really large transactions, and then a cluster of much smaller transactions, which constituted most of the data. We then looked at the average cost per bottle within each cluster. One of the clusters had an average bottle retail price of $99, which was over three times the price of the next highest bottle retail price. There were only 5206 records within this cluster and roughly half of them were sales within the 5 most populated counties. Densely populated areas are more likely to contain high income individuals, who then might purchase higher quality and more expensive alcohol. Within this cluster, there is a single record of a bottle of alcohol selling for $3000 in Polk county. Though this might skew the average slightly, the quartile values are still higher than the other clusters. The presence of this cluster shows that there is a market for high end and expensive alcohol within Iowa in its more populous counties.

5 Discussion

This project set out to examine Iowa's liquor sales trends and the broader implications of these trends on lifestyle and population dynamics. To do so, we utilized Iowa liquor sales data, population demographics, and income distribution to understand the socio-economic drivers of alcohol consumption. The results of our analysis align with previous research on alcohol consumption and urbanization, providing insights that deepen our understanding of how population growth and accessibility to vendors drive liquor sales.

5.1 Summary of Approach and Results

The analysis revealed that population size, rather than income, plays a significant role in determining alcohol sales. Urban counties, such as Polk and Linn, consistently reported higher total liquor sales compared to rural counties. This finding is consistent with previous research that demonstrates how urbanization and higher population density can lead to increased alcohol consumption due to greater accessibility to alcohol vendors and social drinking opportunities [23]. This suggests that policies aimed at controlling alcohol consumption must take into account the density and distribution of vendors in urban areas.

In contrast to expectations based on economic theories, the weak correlation between income and alcohol sales challenges the traditional assumption that higher income leads to increased discretionary spending on alcohol. As found in similar studies, such as by Johnson (2021), income alone is not a strong predictor of alcohol consumption [15]. Cultural and social factors, rather than strictly economic conditions, appear to be stronger drivers of consumption in Iowa, especially in urbanized regions where social drinking and gatherings may be more common. These results are also consistent with Lee's (2021) work, which emphasizes the role of social norms in influencing drinking habits [19].

The time series analysis using the SARIMA model provided key insights into potential future trends. The forecasted decline in alcohol sales points to broader lifestyle changes, such as an increasing emphasis on health-conscious behavior and a possible shift in public attitudes toward alcohol consumption. This decline is likely influenced by long-term public health campaigns and the social changes prompted by the COVID-19 pandemic. Similar patterns have been observed in other regions where health awareness has increased, leading to reduced alcohol consumption, as noted by Garcia (2022) [11]. The impact of these societal shifts highlights the need for ongoing research into how public health interventions influence consumption behaviors over time.

5.2 Scientific and Ethical Implications

From a scientific perspective, this study's findings contribute to the ongoing debate about the relationship between urbanization, vendor accessibility, and alcohol consumption. The strong correlation between population density and alcohol sales supports the argument that urban centers are more likely to foster higher alcohol consumption due to the concentration of vendors and social opportunities. However, the absence of individual-level data limits our understanding of how these trends translate into per capita consumption. Future research that incorporates per capita data could help disentangle the effects of population size from individual consumption habits, offering a more granular view of alcohol-related behaviors.

Ethically, the findings raise important questions about the balance between economic growth, public health, and the accessibility of alcohol. The concentration of vendors in urban areas, as well as the availability of high-end alcohol products, raises concerns about the potential for increased alcohol-related harm in densely populated communities. Thompson (2020) argues that regulating vendor density is an effective way to mitigate public health risks while still supporting the local economy [25]. However, such measures must be carefully considered to ensure that they do not disproportionately affect businesses that rely on alcohol sales, particularly in areas where tourism or social nightlife is an economic driver.

Another important ethical consideration is the discrepancy in healthcare resources between urban and rural areas. As Davis and Peterson (2019) point out, rural areas often have limited access to healthcare services, making it more difficult to address issues related to alcohol misuse [8]. While our data suggest lower total alcohol sales in rural counties, these regions may still face significant challenges in managing the social and health consequences of alcohol consumption. Tailored interventions that address the unique needs of rural populations, such as targeted public health campaigns or mobile health services, could help bridge this gap and ensure that both urban and rural areas benefit from alcohol-related health initiatives.

5.3 Social and Cultural Context

Beyond the scientific and ethical implications, the cultural and social context of alcohol consumption in Iowa must also be considered. The seasonal peaks in sales, particularly around holiday periods, reflect the deep-rooted role of alcohol in social gatherings and celebrations. Johnson (2021) discusses how alcohol consumption tends to spike during culturally significant events, such as holidays and festivals, as part of larger social and familial traditions [15]. This pattern suggests that alcohol consumption is not merely a function of economic or demographic factors, but also a reflection of social identity and cultural norms.

Furthermore, the rise of health-conscious movements such as the "sober curious" trend is beginning to reshape public perceptions of alcohol consumption. As noted by Lee (2021), younger generations are increasingly adopting behaviors that prioritize health and wellness, which may explain the forecasted decline in future sales [19]. These shifts point to an evolving social landscape in which alcohol is no longer as central to socializing as it once was, particularly among younger, more health-aware demographics. Public health campaigns that promote alternative forms of socializing, along with policies that encourage moderation, could help accelerate this cultural shift.

5.4 Next Steps

Building on these findings, future research should focus on addressing the limitations of this study, particularly in terms of per capita consumption analysis. By disaggregating data by age, education, and other socio-economic factors, researchers can provide a clearer understanding of how individual consumption patterns are evolving. Additionally, the role of public health campaigns, such as those promoting "sober curious" lifestyles, should be investigated to assess their long-term effectiveness in reducing alcohol dependency and promoting healthier alternatives [19].

Policymakers should also explore regulating alcohol vendor density, particularly in urban areas where consumption is higher. Examining the relationship between vendor accessibility and alcohol sales will provide important insights into how public health risks can be managed while balancing economic interests. Further research should also assess the effectiveness of policies that restrict alcohol availability in curbing excessive consumption.

6 Conclusion

This study set out to address the question: *How can analyzing liquor sales trends provide insights into broader lifestyle changes and shifts in Iowa's population?* Through a comprehensive analysis of Iowa's liquor sales data, census information, and time series forecasting, our findings indicate that population size and urbanization, rather than income, are the most significant factors influencing alcohol consumption patterns across the state. Urban counties like Polk and Linn consistently reported higher alcohol sales, underscoring the critical role of population density and the concentration of alcohol vendors in driving these trends [23, 22].

While the relationship between income and alcohol sales was found to be relatively weak, the findings suggest that cultural and social factors, particularly those associated with urbanization, play a more prominent role in shaping consumption behaviors [15, 19]. The impact of urbanization on total alcohol consumption is consistent with previous studies, which show that greater vendor accessibility and social opportunities in urban areas contribute to increased consumption [23]. However, these results must be interpreted with caution, as they reflect total alcohol sales and not individual consumption.

Furthermore, our SARIMA forecast projects a potential decline in future alcohol sales, reflecting broader lifestyle shifts toward health-conscious behaviors, potentially accelerated by public health interventions and changing societal norms around alcohol consumption [11, 18]. The ongoing influence of the COVID-19 pandemic may also have contributed to these changing consumption patterns, as individuals have reevaluated their drinking habits in light of increased health awareness [26].

In conclusion, Iowa's liquor sales trends provide valuable insights into the broader demographic and lifestyle shifts occurring within the state. Urbanization has been a key driver of alcohol consumption, with population density and vendor availability significantly influencing total sales. As public health policies evolve and society becomes increasingly health-conscious, these trends offer a foundation for future research and policy making aimed at promoting responsible alcohol consumption and balancing public health concerns with economic interests [25, 5]. The results of this study highlight the need for targeted interventions, particularly in urban areas, where alcohol-related harm may be more prevalent due to higher consumption levels. Policymakers should consider the potential benefits of regulating vendor density and continuing to promote health-focused public campaigns that encourage moderation and responsible drinking behaviors.

References

[1] T. Babor. *Alcohol: no ordinary commodity: research and public policy*. Oxford University Press, USA, 2010.

[2] S. M. R.-L. A. P. F. . L. A. B. Brand, D. A. Comparative analysis of alcohol control policies in 30 countries. *PLoS medicine*, 2007.

[3] U. S. C. Bureau. State iowa, 2022. Accessed: 09-01-2024.

[4] C. S.-H. U. F. P. T. P. R. N. T. C Stockley, CI Grande and P. Struder. Comparison of international alcohol drinking guidelines. *OIV - Internationsl Organization of Vine and Wine*, 2019.

[5] C. A. Campbell, R. A. Hahn, R. Elder, R. Brewer, S. Chattopadhyay, J. Fielding, T. S. Naimi, T. Toomey, B. Lawrence, J. C. Middleton, et al. The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms. *American journal of preventive medicine*, 37(6):556–569, 2009.

[6] A. B. D. Commerce. Iowa enterprise data liquor sales snapshot, 2021. Accessed: 2024-09-12.

[7] A. B. D. Commerce. Iowa liquor sales, 2024. Accessed: 09-01-2024.

[8] M. Davis and R. Peterson. Rural vs. urban alcohol consumption patterns: A comparative analysis. *Rural Health Journal*, 28(2):78–85, 2019.

[9] H. Editors. Prohibition, 2009.

[10] U. H. Foundation. Excessive drinking in united states. *America's Health Rankings*, 2022.

[11] L. Garcia. Time series analysis of alcohol sales using sarima models. *Time Series Analysis Journal*, 11(3):201–210, 2022.

[12] J. Gerstel. Data analysis on liquor sales in iowa: Insights into stores. *NYC Data Science Academy*, 2021.

[13] I. HHS. Your life iowa, 2024.

[14] D. Jernigan. The failures of alcohol regulation. *Vital City*, 2023.

[15] E. Johnson. The impact of income on alcohol consumption: A review. *Economic Behavioral Studies*, 42:123–130, 2021.

[16] K. Kelley. Alcoholic beverage consumption and purchasing trends, 2024. Accessed: 2024-09-12.

[17] A. Kutz. 'fascinating history': Iowa's relationship with alcohol, explained, 2023.

[18] S. Lamb. Can the sober curious trend change u.s. alcohol consumption?, 2023. Accessed: 2024-09-12.

[19] M. Lee.

[20] K. V. S. . C. J. D. Matthew J Voss, Shawn F Dorius. Local alcohol policies to improve community health safety in iowa. *Public Science Collaborative*, 2023.

[21] I. D. of Revenue. To-go carryout cocktails, 2020.

[22] K. V. S. Shawn Dorius and C. Dorius. Geospatial analysis of health, safety and alcohol outlets in iowa. *Public Science Collaborative*, 2021.

[23] J. Smith and J. Doe. Urban alcohol consumption trends in the u.s. *Journal of Urban Studies*, 34(4):456–467, 2023.

[24] L. Ta. Binge drinking rates are higher in iowa than other states. *Axios Des Moines*, 2024.

[25] A. Thompson. Alcohol vendor density and its impact on public health policy. *Public Health Policy Review*, 15(1):67–80, 2020.

[26] C. University. Study shows uptick in u.s. alcohol beverage sales during covid-19 pandemic, 2021. Accessed: 2024-09-12.

A Code Appendix

Project Code: https://github.com/jtowgood/DAT490Capstone