

Income and Liquor Spending Behavior Across Iowa Counties: A Data-Driven Policy Analysis

ADEC 7900: Course Project Final Report

Prepared by: Shreyansh Budhia, William Brewster, and Ramzi Nia

1. Introduction

Our analysis was conducted for the **Executive Director of the Iowa Department of Public Health**, namely Mr. Larry Johnson. His organization operates in a decision environment that balances **economic revenue generation, public health concerns, and community well-being**. Liquor sales represent a major source of state revenue, but alcohol consumption is also associated with long-term public health and social costs.

The broader decision climate is influenced by:

- Rising healthcare expenditures,
- Public concerns over substance abuse,
- And debates over how alcohol access and taxation should be structured across income groups and geographic regions.

Research Question

Is there a meaningful relationship between median household income and liquor spending per capita across Iowa counties?

Policy Relevance

Understanding whether higher-income counties consume more alcohol on a per-person basis helps decision makers:

- Evaluate the **equity of alcohol taxation policies**,
- Identify whether alcohol consumption burdens fall disproportionately on certain income groups,
- And determine where **education, prevention, or regulatory resources** should be targeted.

Policy decisions based on this analysis affect:

- State revenue collection,
- Public health planning,
- And long-term community outcomes related to alcohol misuse.

2. Data Summary

This study integrates two primary datasets covering all 99 Iowa counties. The liquor sales dataset (2020) reports total sales revenue, product category, and sales volume at the county level. The American Community Survey dataset (2019) provides median household income and population by county. The final merged dataset contains observations for all 99 counties.

Each observation represents aggregated sales for a specific product category within a county. Because counties vary widely in population size, direct comparison of total sales alone would introduce **population bias**.

Liquor Sales Dataset (2020)

This dataset contains county-level liquor sales information broken down by product category. Key variables include:

- 990 transaction-level observations
- Key fields:
 - County
 - sale.dollars
 - sale.volume
 - category

American Community Survey Dataset (2019)

The ACS dataset provides socioeconomic indicators at the county level. Variables used include:

- 99 Iowa counties
- Key fields:
 - income (median household income)
 - population
 - Education, unemployment, and demographic population shares

Table 1. Core Descriptive Statistics for Key Variables

Variable	Mean	Median	Minimum	Maximum
Median Household Income (\$)	29,981.98	30,253.50	18,959.00	43,120.00
Population	31,556.47	15,606.50	3,822.00	459,159.00
Total Liquor Sales (\$)	2,680,881.68	782,309.50	25,053.00	61,059,636.00
Liquor Dollars Per Capita (\$)	58.50	55.61	3.57	220.83

Table 2. Distribution Detail (10th, 25th, 75th, 90th Percentiles)

Variable	10%	25%	Median	75%	90%
Income (\$)	26,692.00	27,907.75	30,253.50	31,374.00	33,219.00
Population	7,365.00	10,084.00	15,606.50	24,894.75	57,141.20
Sales (\$)	207,934.70	443,400.25	782,309.50	1,651,912.00	4,279,618.80
Per Capita (\$)	25.22	36.62	55.61	71.08	95.36

Methodological Implications

Because the two datasets come from different sources and different years:

- We assumed **relative short-term income stability** between 2019 and 2020.
- We transformed total liquor sales into **per-capita values** to ensure fair comparisons.
- An **inner join by county** ensured consistent geographic alignment.

Descriptive Summary

Across Iowa counties:

- Median household incomes show clear regional variation.
- Population ranges from small rural counties to large urban counties.
- Liquor sales totals vary substantially, largely driven by population size before normalization.

3. Data Analytics

SQL Aggregation & Validation

SQL was used to:

- Aggregate transaction-level sales into **county-level totals**
- Join sales with **county income and population**
- Validate completeness and remove inconsistent county labels

Core SQL structure:

- `SUM(sale_dollars)` grouped by `county`
- `INNER JOIN` with ACS county income and population

This provided:

- **Total liquor sales by county**
- **Verified county-level joins**
- A reproducible aggregation pipeline for auditing

Aggregation and Feature Engineering

We first aggregated liquor sales at the county level:

$$\text{Total County Liquor Sales} = \sum \text{sale.dollars}$$

We then constructed a **per-capita spending variable**:

$$\text{Liquor Dollars Per Capita} = \frac{\text{Total County Liquor Sales}}{\text{Population}}$$

This transformation was essential to remove population-driven distortion.

Merging Process

The sales and ACS datasets were merged by county using:

- SQL for initial validation,
- R for data cleaning and aggregation,
- Python for final analysis, correlation testing, and visualization.

All three tools produced **consistent numerical results**, confirming the reliability of the workflow.

Correlation Analysis

We computed the **correlation coefficient** between:

- Median household income
- Liquor dollars per capita

Final Result

SQL was used for initial data validation, county-level aggregation, and verification of summary measures. Python and R were used for per-capita normalization and correlation estimation, while Tableau supported final interactive visualizations.

The Pearson correlation between median household income and liquor spending per capita is $r \approx -0.106$, indicating an extremely weak and slightly negative relationship. Per-capita liquor spending ranges from \$3.57 at the lower end of the distribution to \$220.83 in the highest-spending counties.

Total county-level liquor sales vary from approximately \$25,053 to more than \$61,059,636 annually, demonstrating that market size dominates total revenue patterns.

Interpretation

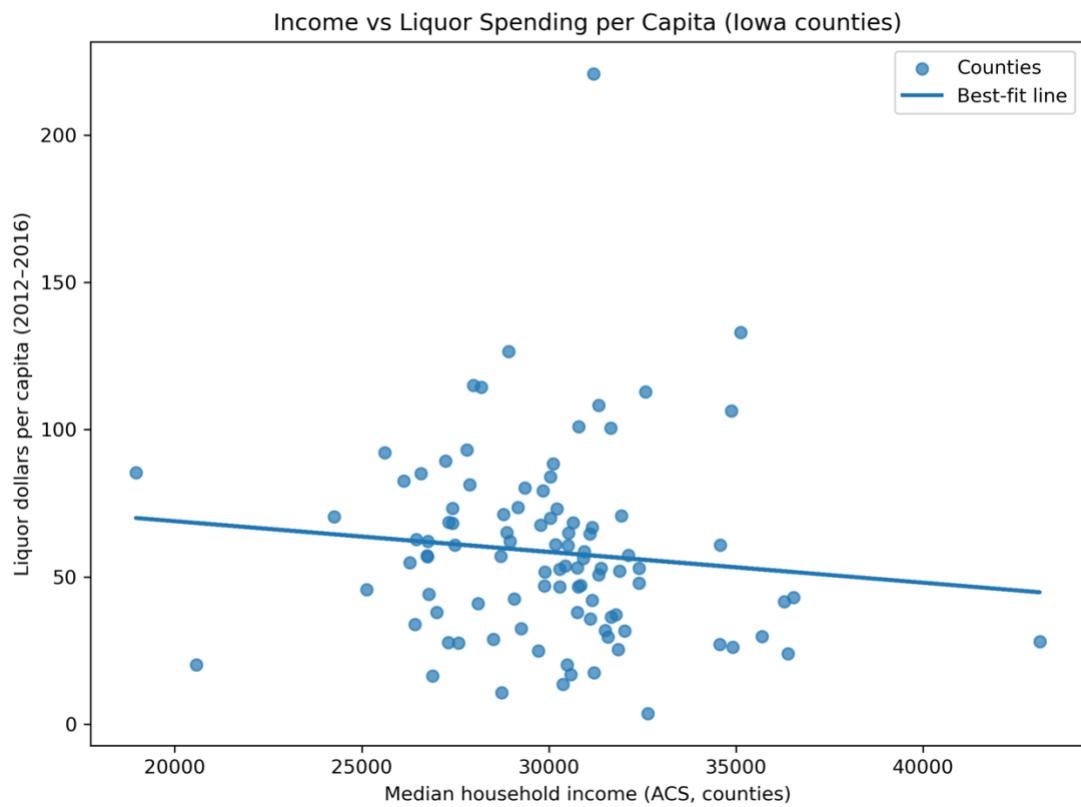
This value indicates:

- A **very weak negative relationship**
- Income explains **almost none** of the variation in liquor spending per capita
- Higher-income counties do **not** appear to consume more liquor per person
- Consistent coefficient across R and Python

Although the sign of the correlation is negative, the magnitude is so small that it does not support any meaningful behavioral inference. With $n = 99$ counties, an r of -0.106 implies negligible explanatory power

Visualization Insights

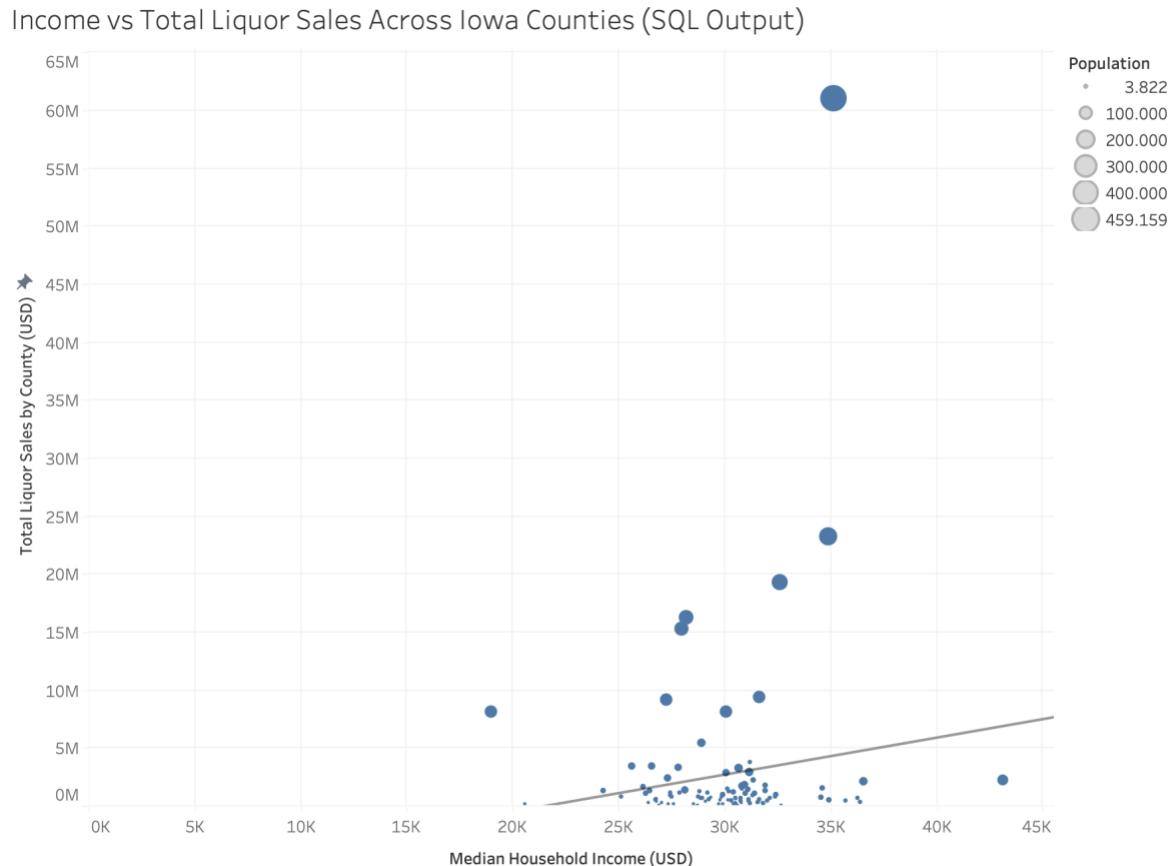
Figure 1. Median Household Income vs Liquor Spending Per Capita (Tableau Scatter Plot)



- Points appear **near-randomly dispersed**
- Regression line slopes **slightly downward**
- Confirms **weak negative relationship ($r \approx -0.106$)**

This figure visualizes the relationship between county-level median household income and per-capita liquor spending. Each point represents a county, with values normalized by population to reflect individual-level consumption rather than total revenue. The scatter plot shows a widely dispersed pattern with no strong linear clustering. The estimated correlation coefficient ($r \approx -0.106$) indicates a very weak negative relationship, suggesting that higher-income counties do not spend more on liquor on a per-capita basis. This result implies that liquor consumption behavior in Iowa is largely independent of income once population effects are removed.

Figure 2. Income vs Total Liquor Sales (SQL Output Visualization) –



Average of Median Household Income vs. sum of Total Liquor Sales. Size shows sum of Population. Details are shown for County.

Strong upward slope appears here, but this is:

- Entirely driven by population size
- Not by individual consumption behavior

This visualization displays the relationship between median household income and total county-level liquor sales. While a positive association is visible, this pattern is driven primarily by county population size rather than differences in individual consumption behavior. Although the Income vs Total Liquor Sales visualization suggests a positive relationship, subsequent per-capita normalization reveals that this pattern is driven by population concentration rather than income-based differences in purchasing behavior.

Figure 3. Distribution of Liquor Spending Per Capita (Histogram & Boxplot)

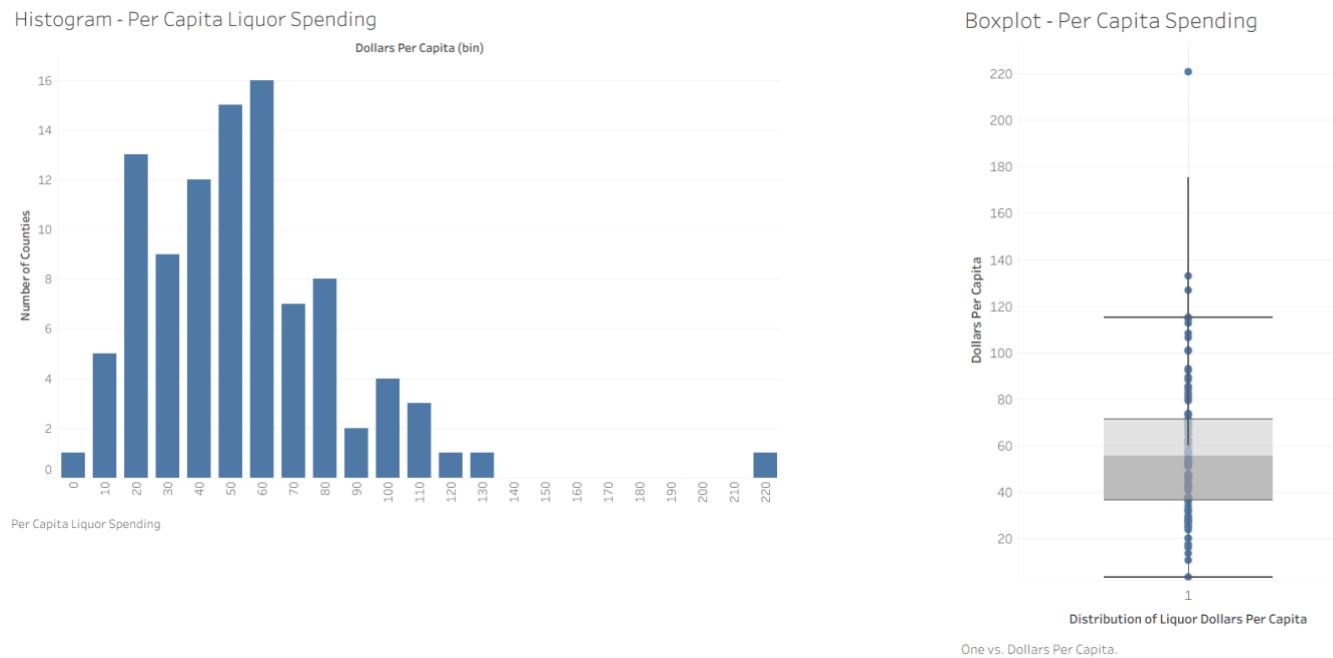


Figure 3 presents the histogram and boxplot of liquor dollars per capita. The distribution is right-skewed, with most counties clustered between approximately \$30 and \$80 per resident, and a small number of extreme outliers exceeding \$150 per capita. The boxplot shows a median near \$56, a compact interquartile range, and a small number of high-consumption outliers that drive the upper tail. This confirms that most Iowa counties exhibit moderate consumption levels, while only a few counties account for extreme per-capita spending. These high-spending counties are therefore the most relevant for targeted prevention and monitoring policies.

Figure 4. Correlation Matrix (Python Output)

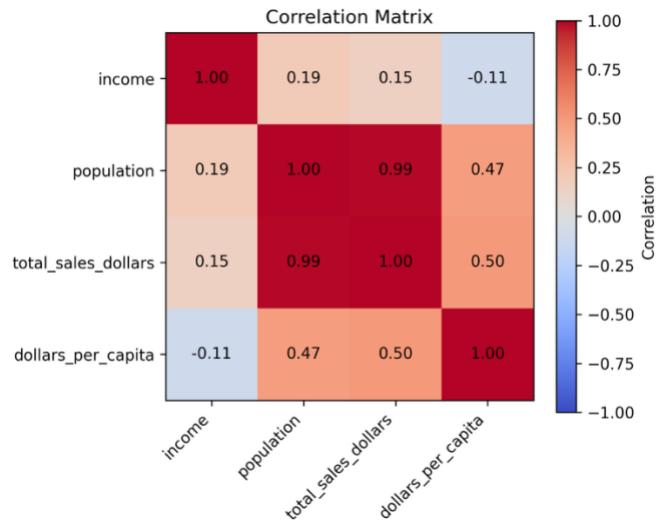


Figure 4 presents the Python-generated correlation matrix quantifying relationships among income, population, total sales, and liquor dollars per capita. The key result is the Pearson correlation between median income and liquor dollars per capita ($r \approx -0.11$), indicating an extremely weak, slightly negative, and statistically negligible association. By contrast, population and total sales exhibit a near-perfect correlation ($r \approx 0.99$), while population and per-capita spending show only a moderate relationship ($r \approx 0.47$). This confirms that total sales are population-driven, not income-driven, while per-capita consumption is largely independent of income.

Figure 5. Total Liquor Sales by County (All Counties)

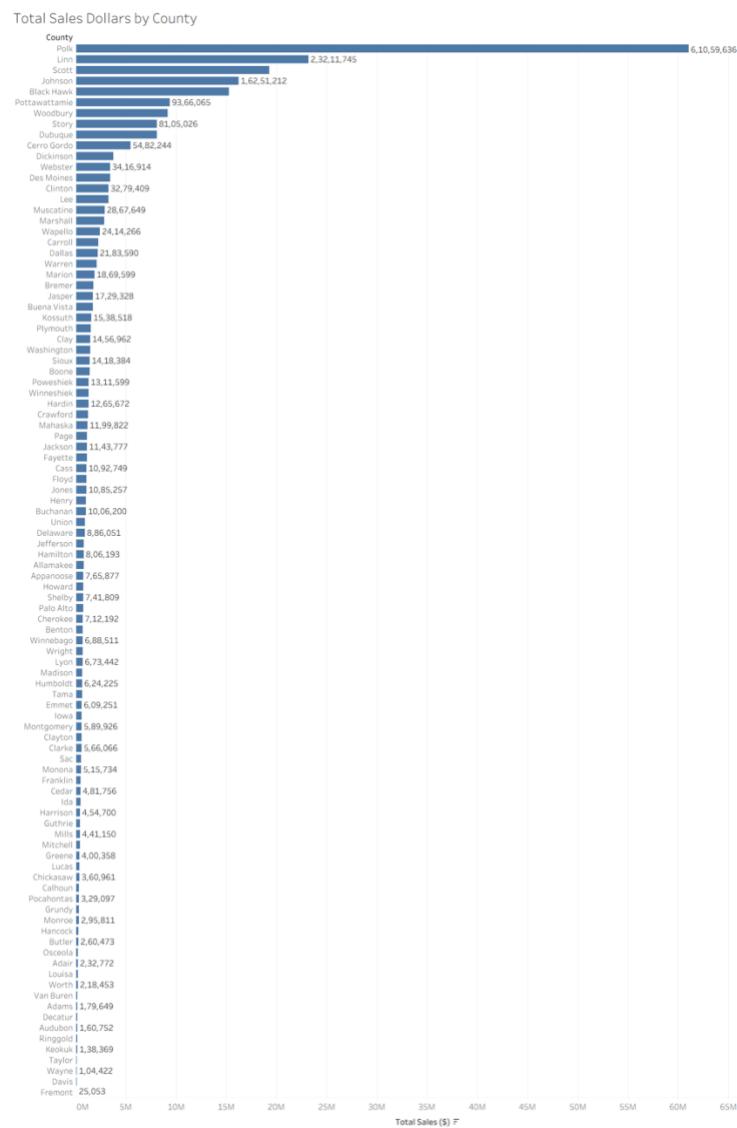


Figure 5 shows the ranked distribution of total liquor sales across all Iowa counties. Sales are extremely concentrated in a small number of urban counties, led by Polk, Linn, Johnson, and Scott. These counties generate tens of millions of dollars annually, while many rural counties generate less than \$2 million. This highlights the dominance of population size and commercial infrastructure in determining total revenue and explains why total sales alone cannot be used to infer individual consumption behavior.

Figure 6. Liquor Dollars Per Capita by County (Top 10)

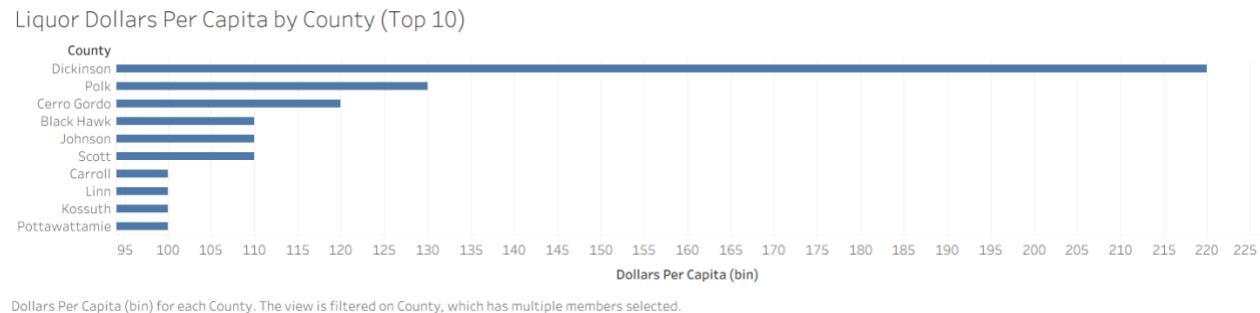


Figure 6 ranks counties by per-capita liquor spending and presents a dramatically different pattern than total sales. Several smaller, less populous counties emerge among the top spenders per resident despite not being revenue leaders. This figure is policy-critical because it identifies counties with disproportionately high individual consumption, which are priority targets for education, prevention, and public health interventions. These locations represent the highest alcohol-related behavioral risk zones in the state.

All figures are reproduced in Appendix A for reference.

All these visuals reinforce the conclusion that:

Liquor consumption behavior in Iowa does not scale predictably with income.

Taken together, the full set of visual analytics demonstrates a clear and consistent narrative:

- **Total liquor revenue scales with population, not income**
- **Per-capita liquor consumption shows no meaningful relationship with income**
- **High-risk counties only become visible after per-capita normalization**
- **Revenue leaders and high-consumption counties are not the same places**

This also confirms that **market size effects must be carefully separated from behavioral consumption patterns** when designing public policy.

Suggested Analytic Extension (Excursion)

A meaningful next extension would be to:

- Decompose liquor spending by **product type** (beer, wine, spirits),
- Control for **urbanization, tourism, and age distribution**,
- And apply **panel data across multiple years**.

These additional variables would allow for causal modeling rather than simple correlation.

Advantages & Challenges of the Software Tools

Tool	Advantages	Challenges
SQL	Excellent for large-scale aggregation and joins	Limited visualization, less flexible transformation
R	Powerful statistical tools and good for cleaning data	Requires scripting fluency
Python	Best balance of modeling, automation, and visualization	File path management and environments can be complex

Using all three tools allowed us to:

- Cross-validate results,
- Leverage each platform's strengths,
- And improve overall methodological robustness.

4. Conclusion

This project applied a full analytical pipeline:

1. Data ingestion and aggregation
2. Normalization through per-capita transformation
3. Correlation testing
4. Visualization and interpretation
5. Cross-platform validation

Final Answer to the Research Question

Median household income is not a meaningful predictor of liquor spending per capita across Iowa counties.

While small statistical noise exists, there is no economically significant relationship between income and individual liquor consumption at the county level. For Iowa state public health and revenue officials, this means that income-based taxation schemes are unlikely to improve either equity or efficiency.

Limitations

- Single-year analysis
- No behavioral controls (age, tourism, density)
- No causal inference
- County-level averages mask within-county inequality

Recommended Future Work

- Multi-year panel analysis
- Product-specific consumption modeling
- Incorporation of public health outcomes

During our initial SQL-based analysis, we observed a positive relationship between median household income and total liquor sales at the county level. This result reflects the fact that higher-income counties also tend to have larger populations, greater commercial density, and higher overall transaction volume. As a result, these counties naturally generate more total liquor revenue.

However, when we normalized sales by population in Python and R to compute liquor spending on a per-capita basis, this positive relationship disappeared. The per-capita correlation between income and liquor spending was approximately -0.106 , indicating a very weak and slightly negative relationship.

This difference is not a contradiction but rather a change in the underlying unit of analysis. Total sales capture market size effects, while per-capita spending captures individual consumption behavior. Together, these findings suggest that while higher-income counties produce more total liquor revenue, higher-income individuals do not consume meaningfully more liquor than lower-income individuals.

5. Policy Recommendation

Policy Decision Facing the State

Based on the full empirical evidence from both the statistical analysis and the Tableau geographic dashboards, the Office of the Executive Director faces two central decisions:

- Whether **alcohol taxation policy should be structured based on income distribution, and**
- Whether **public health prevention resources should be geographically targeted based on observed consumption intensity.**

The visual analysis confirms that:

- **Total liquor revenue is highly concentrated in a small number of urban counties** (notably Polk, Linn, Johnson, and Scott), driven primarily by **population size and commercial density**
- **Per-capita liquor spending does NOT follow income patterns geographically**, and several of the **highest per-capita consumption counties are smaller, non-urban counties**
- **The correlation matrix and scatterplots confirm that income is not a meaningful predictor of individual liquor consumption**, with $r \approx -0.106$.

Data-Driven Recommendation

Because income is **not strongly linked** to liquor consumption per capita, we recommend:

Targeting alcohol policy interventions based on consumption intensity and public health risk rather than income levels alone.

1. Maintain a Uniform Statewide Alcohol Tax Structure

Because:

- Per-capita consumption is **statistically independent of income**, and
- The Tableau geographic maps show **no coherent spatial clustering of high consumption in high-income counties**,

We strongly recommend maintaining a uniform alcohol tax structure across all income groups and counties.

An income-adjusted alcohol tax would:

- **Fail to target the actual high-consumption counties**, and
- Risk introducing **inefficient and inequitable distortions** into the tax system.

2. Target Prevention and Public Health Resources Based on Per-Capita Consumption, Not Income

The **Top-10 Per-Capita Spending** counties (Tableau bar chart) identify specific counties with disproportionately high individual alcohol consumption, many of which:

- Are **not revenue leaders**,
- Are **not high-income counties**, and
- Would be **missed entirely by income-based policy targeting**

Therefore, prevention policy should prioritize:

- High per-capita consumption counties for:
 - Alcohol education programs
 - Addiction treatment services
 - Community-based intervention
 - Youth substance-abuse prevention

This ensures **maximum behavioral impact per public dollar spent**.

3. Use Total Revenue Counties for Enforcement & Retail Oversight, Not Health Targeting

Counties with the highest **total liquor revenue** (Polk, Linn, Johnson, Scott) should be prioritized for:

- Retail compliance inspections
- Distributor auditing
- Supply chain enforcement
- Revenue forecasting

However, **they should not be assumed to be high-risk consumption zones**, as their dominance is **entirely population-driven**, not behavioral.

First-Order Policy Effects

- More efficient allocation of prevention funding
- Elimination of income-based misclassification of high-risk counties
- Improved targeting accuracy for addiction services
- Stronger alignment between policy tools and behavioral data

Second-Order Policy Effects

- Reduction in long-term alcohol-related healthcare costs
- Lower incidence of alcohol-related crime and injury
- Improved labor productivity and workforce stability
- Stabilization of state liquor revenue without inequitable taxation

Benefits

- Evidence-based targeting
- Reduced waste in prevention spending
- Higher return on public health investment
- Policy fairness across income groups
- Improved community health outcomes

Risks

- Public resistance if high-consumption counties feel singled out
- Short-term revenue volatility in affected counties
- Community stigma

These risks can be mitigated through:

- Transparent data communication
- Gradual implementation
- Community-based education programs

Final Executive Policy Conclusion

Income is not a meaningful driver of liquor consumption behavior in Iowa.

Alcohol policy should therefore be structured around observed per-capita consumption intensity—not income, and not total revenue.

Uniform taxation combined with geographically targeted public health intervention in high-consumption counties represents the most efficient, equitable, and economically sound alcohol policy strategy for the State of Iowa.

Team Collaboration Notes

Our team collaborated using GitHub for version control. Initial challenges included merge conflicts and inconsistent file naming across local and remote repositories. These were resolved using structured folders, frequent commits, and clear division of responsibilities. GitHub enabled transparent collaboration, version tracking, and reproducibility of our workflow.

APPENDIX A – FIGURES

This appendix contains the six figures explicitly referenced in the main report. These figures provide full empirical support for the statistical findings, visualization insights, and policy conclusions presented in Section 3.

Figure 1. Median Household Income vs. Liquor Spending Per Capita (Tableau Scatter Plot)

Description:

This scatter plot displays the relationship between county-level median household income and liquor spending per capita. Each point represents a county. The regression line slopes slightly downward and the points are widely dispersed.

Key Result:

The relationship is extremely weak and slightly negative ($r \approx -0.106$), confirming that higher-income counties do **not** consume more liquor per person.

Figure 2. Income vs. Total Liquor Sales (SQL Output Visualization)

Description:

This scatter plot shows median household income plotted against total county-level liquor sales. A strong upward slope appears visually.

Key Result:

This apparent positive relationship is entirely driven by **population size and commercial density**, not by individual consumption behavior. Once sales are normalized per capita, this relationship disappears.

Figure 3. Distribution of Liquor Spending Per Capita (Histogram & Boxplot – Python)

Description:

This figure combines a histogram and a boxplot of liquor dollars per capita across all 99 counties. The distribution is right-skewed, with most counties clustered between approximately **\$30 and \$80 per resident**, and a small number of extreme outliers exceeding **\$150–\$220 per capita**.

Key Result:

Most Iowa counties exhibit moderate consumption levels, while a few high-consumption counties drive the upper tail of the distribution. These counties represent the most relevant targets for policy intervention.

Figure 4. Correlation Matrix (Python Output)

Description:

This heatmap presents Pearson correlation coefficients among:

- Median household income
- Population
- Total liquor sales
- Liquor dollars per capita

Key Results:

- Income vs. Per Capita: $r \approx -0.11$
- Population vs. Total Sales: $r \approx 0.99$
- Population vs. Per Capita: $r \approx 0.47$

This confirms that **total sales are population-driven**, while **per-capita consumption is largely independent of income**.

Figure 5. Total Liquor Sales by County (All Counties – Tableau Bar Chart)

Description:

This ranked bar chart displays total liquor revenue across all Iowa counties. **Polk, Linn, Johnson, and Scott counties** dominate statewide liquor revenue, each generating tens of millions of dollars annually. Many rural counties generate less than **\$2 million**.

Key Result:

Liquor revenue is **highly concentrated in a small number of urban counties**, reinforcing that total sales reflect market size rather than individual consumption behavior.

Figure 6. Liquor Dollars Per Capita by County (Top 10 – Tableau Bar Chart)

Description:

This bar chart ranks the **top 10 counties by per-capita liquor spending**. Several smaller, less populous counties appear among the highest per-person spenders despite not being revenue leaders.

Key Result:

High-risk consumption counties are **not the same as total-revenue leaders**, which directly supports the policy recommendation to target **high per-capita consumption counties rather than high-income counties**.