INTERPRETIVE AND PREDICTIVE MODELS ON IOWA ALCOHOL SALES PER CAPITA

DUNCAN APPLEGARTH
Department of Computer Engineering
duncana513@gmail.com

SARAH ELLWEIN
Department of Mathematics
sellwein@calpoly.edu

CHARLIE WARD
Department of Computer Science and Software Engineering
cward11@calpoly.edu

October 20, 2022

ABSTRACT

The Iowa Department of Commerce, Alcoholic Beverages Division, provides a dataset of spirits purchase information of liquor licensed establishments from January 1, 2012 to September 1, 2021 [3]. The liquor sales dataset, with additional community demographic datasets, can serve as a proxy in analyzing and predicting alcohol consumption in the state of Iowa. This project aims to produce both predictive and interpretive linear models on different measures of alcohol consumption per capita using a ridge regression estimator. All models, both predictive and interpretive, indicate that county and alcohol category, among other variables, are strong predictors of alcohol consumption. This report will also cover ethics of using these models and further suggestions in development.

***Keywords*:** Linear Regression · Interpretive Model · Predictive Model · Alcohol Sales

1 Introduction

Alcohol is a big part of the modern-day lifestyle of many people – especially when considering the college demographic that we are in today. Whether it's to celebrate a special occasion or just to wind down with some friends over the weekend, the alcohol industry is quite profitable in the United States and around the world. With this context, the group was tasked to investigate the relationships between alcohol purchases that were made by liquor stores in the state of Iowa and various dependent variables to get a better understanding of the alcohol trends in Iowa.

This report details the analysis of the Iowa Liquor Sales 2012 - 2022 data in combination with Iowa population, income, and unemployment datasets to analyze relationships between these data and construct models for these relationships. You will find an explanation of the datasets used, data cleaning techniques employed, features engineered, and dataset aggregation in the relevant sections below. After preparing and aggregating these datasets, analysis of regression techniques to (1) build predictive models and (2) understand the relationship between the engineered independent variables and the per capita Number of Bottles Sold, Sales Volume of Alcohol ($), and Volume of Sold Alcohol (L) dependent variables will be presented. For each target variable listed above, the goals are twofold: (1) to find the model that will best predict Iowa alcohol sales data, and (2) to explain the factors that have influenced existing Iowa alcohol sales. This report will also cover ethics of using these models with the aggregated data and future suggestions to improve these results.

2 Methodology

The method in developing and implementing linear regression models is separated into steps:

1. **Data Sourcing**: External datasets were gathered and cleaned to measure capita values and add features of interest.
2. **Data Integration**: All datasets were aggregated by at least two dimensions of data: time and geographic location. Additionally, the alcohol category was added as a third dimension for aggregation.
3. **Feature Engineering**: Additional features were extracted from the final aggregated dataset for the linear regression model.
4. **Implementation**: Linear regression models (along with choice of loss function) were developed and employed for each of the target variables.
5. **Evaluation**: Models were evaluated based upon their performance on R^2 and root mean squared error (RMSE) metrics. Decisions on whether to apply further feature engineering, choose different predictors, or finalize results were based on these metrics.

Although these steps may appear chronological, the process is far more cyclical in nature. Multiple models were created to find the best predictors for predictive models and impactful predictors for interpretive models. Figure 1 depicts the true nature of the project's method.

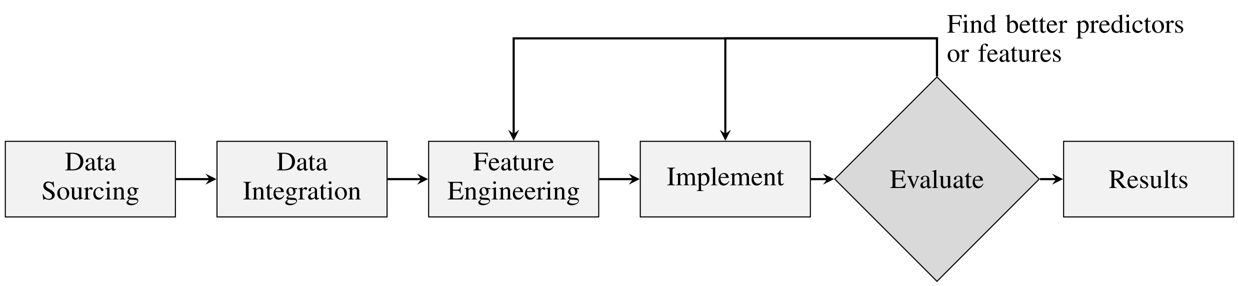


Figure 1: Flowchart depicting workflow of this project. Feature engineering, and implementation steps were repeated based on evaluation. Results are finalized once model meets satisfaction.

Decisions regarding aggregation and data integration components needed to be made prior to seeking external datasets to complement the provided Iowa Liquor Sales data. In particular, the specifics regarding the granularity for the time and geolocation aggregation had to be made to inform the data sourcing effort.

We decided the resolution for time and location aggregation should be by week and county, respectively. The rationale for aggregating by week is to maintain granularity of the transactional data while capturing overall trends, as opposed to per day basis which would potentially make trends harder to interpret or per month basis which may obfuscate potential trends due to its coarse nature. The rationale for aggregating by county for the geographic location dimension is similar – per town, per zip code, or per city may have been too fine to glean any interesting overall trends, where anything more generalized would obfuscate interesting trends. With these aggregation decisions made, additional data sources were collected and combined with the Iowa Liquor Sales data.

2.1 Data Sourcing

Additional datasets were required to calculate per capita measures and to provide additional features that increased predictive and interpretive power for the linear regression models. Because this project's target variables are per capita, population data needed to be gathered. Various county demographics and statistics may provide useful predictors and insights on alcohol purchases per location. For this purpose, per-county income and unemployment rate datasets were chosen to investigate the associations between individual spending power and alcohol purchases per capita. This section will describe the datasets used for the model and data cleaning techniques applied for each dataset.

Iowa Liquor Sales

The Iowa Liquor Sales dataset is provided by the Iowa Department of Commerce, Alcoholic Beverages Division in Iowa Data. This dataset contains 24,847,481 records of transactional commercial establishment liquor purchases in the state of Iowa that span from January 1, 2012 to September 1, 2022 [3]. There are 24 columns providing store metadata and store alcohol transaction information, some of which are purchase date, county, liquor category, bottles sold, volume sold (in gallons/liters), and sale amount (in USD).

The Iowa Department of Commerce published the dataset under the Creative Commons Zero license. In essence, reusers may build upon the material in any medium with no conditions. Credit does not have to be given to the creator, commercial uses of the dataset are permitted, and derivatives/adaptations of the dataset are permitted [2].

These Iowa Liquor Sales data contained errors, missing, and unnecessary data. The following techniques were used to clean this dataset:

- Missing county data was imputed based upon the maximum value count of the county name for equivalent store IDs, and dropped otherwise.
- Records containing spelling errors for county names were corrected, and records containing counties outside of Iowa were dropped.
- Records during the year 2022 were dropped due to associated missing data from other data sources.

County Population in Iowa

The County Population dataset is provided by the U.S. Census Bureau, Population Division in Iowa Data. This dataset contains 3,069 records of county population in Iowa spanning from years 1990 to 2020 [4]. The 2021 population estimates, directly sourced from the U.S. Census Bureau, were used to fill the missing 2021 population data.

Annual Iowa County Personal Income

The Annual Personal Income dataset is provided by the U.S. Department of Commerce, Bureau of Economic Analysis. This dataset contains 4,752 records of both personal income and personal income per capita (PIPC) information for each Iowa county spanning from years 1997 to 2020 [5].

The personal income data was dropped in favor of PIPC. The PIPC for 2021 was imputed using a linear regression for each county where year was the dependent variable and PIPC was being predicted. The ordinary least squares method was utilized for this specific regression and optimized using QR factorization and back substitution.

Unemployment Rates per Iowa County

The Unemployment Rates dataset was sourced from the U.S. Bureau of Labor Statistics. Unlike the other external datasets mentioned, these data were collected by concatenating 99 individual datasets, one for each of the 99 Iowa counties from the BLS website [1]. The final dataset used for Unemployment Rates contains 38,808 records of unemployment rate data for each Iowa county spanning from January 1990 to August 2022. Unemployment Rate data for 2022 were dropped due to associated missing data from other data sources.

2.2 Data Integration

After collecting and cleaning all datasets, both provided and external, they all needed to be merged and aggregated into a single dataset that is ready for use with linear regression models. For the merging step, all external datasets were inner joined, with the Population and Annual Personal Income datasets being merged with Iowa Liquor Sales on year and county, and the Unemployment Rates dataset merged with these three on month, year, and county.

These data were then aggregated by year, week, county, and category, to create aggregated features on the number of bottles sold, sales, state profit, and volume sold: resulting in 678,476 records ($\approx 2.7\%$ the size of the original Iowa Liquor Sales dataset). This aggregation has reduced the number of total records while maintaining granularity on the time and geographic location dimensions that our group deemed acceptable for analysis, as well as increasing data portability. Further feature engineering could then be performed on this aggregated dataset.

2.3 Feature Engineering

Pre-Aggregation

The following features were extracted prior to aggregation:

- Dates were broken up into year, month, and week values.

- Liquor categories containing multiple names for general liquors (e.g. White Rum, Spiced Rum, etc.) were mapped to a general liquor category (e.g. White Rum $\rightarrow$ rum).
- A separate weeks column was created and numbered from January 1, 2012 to December 31 (now labeled 0 to 513).

Post-Aggregation

Per capita values were calculated and added to the final dataset: bottles sold per capita, volume (in liters) sold per capita, sales (in USD) per capita, and state profit (in USD) per capita. These per capita values were the model's target variables.

When initially testing different linear regression models, it became evident that county and alcohol category were strong predictors for the desired dependent variables. It was hypothesized that county and category had interactions with each other, so this became an additional feature. An interaction term between county and alcohol category (named *county:category*) was a vital addition to the final models.

The structure of our final dataset is explained in figure 2.

Attribute Name	Data Type	Description
Year	Categorical	Year purchase was made (labeled 2012 to 2021)
Week	Categorical	Week of the year purchase was made (labeled 1 to 52)
County	Categorical	County of establishment where purchase was made
Category	Categorical	Category of alcohol purchased
County:Category	Categorical	String *county* * *category*
Month	Categorical	Month of the year purchase was made
WeeksFromJan12012	Numerical	Weeks starting from January 1, 2012 to December 31, 2021
Population	Numerical	Annual population of county at the given time
IncomePerCap	Numerical	Annual Personal Income Per Capita of county at the given time
UnemploymentRate	Numerical	Unemployment rate of that county at the given time
*BottlesSoldPerCap**	Numerical	Bottles sold per capita
*Sales$PerCap**	Numerical	Sales (in USD) per capita
*VolumeLSoldPerCap**	Numerical	Volume (in liters) per capita
*StateProfitPerCap**	Numerical	State profit (in USD) per capita

Figure 2: Column description of the final dataset used (* marks target variables).

2.4 Implementation

Once the data had been cleaned, features were engineered, and variables were aggregated, we began our regression implementation. We created dummification and analysis pipelines that allowed us to discover the most important variables, tune the penalty coefficient, and evaluate the models against the R^2 and RMSE metrics.

At the outset, 80% of the final dataset was selected at random and reserved for training, 10% of the dataset was randomly selected as a holdout set to compare models against each other with, and the remaining 10% was held as a test set for reporting R^2 and RMSE statistics about the final models.

Pipeline

Prior to training the model, categorical variables (*county*, *category*, *county:category*, *week*, *year*, *month*) were one-hot encoded and numerical variables (*PIPC*, *population*) were standardized using standard scalers. The post-pipeline design matrix is named X with an array of target values Y. One-hot encoded variables are backed by sparse matrices to take advantage of this data structure optimization.

Loss Function

Ordinary least squares proved to be a challenging method since X is a sparse matrix, making it singular or near singular:

$$\beta = (X^T X)^{-1} X^T Y$$

Therefore, the ridge function was chosen as the model's loss function to avoid the issue of matrix invertibility:

$$\beta = (X^T X + \lambda I)^{-1} X^T Y$$

Different values of the penalty coefficient λ were tested in each model: 0.5, 1, 10, 100, and 1000.

2.5 Evaluation

For each of the three target variables (*BottlesSoldPerCap*, *SalesPerCap*, and *VolumeLSoldPerCap*) a ridge regression with every dependent variable (*Year*, *Week*, *Month*, *County*, *Category Name*, *County:Category*, *IncomePerCap*, *UnemploymentRate*, and *WeeksFromJan12012*) was run with a λ of 10,000. The sorted order of magnitude for each feature coefficient was then examined to get an approximate idea of which features were most important.

Next, a search for the best lambda value was run using a trimmed feature set (features that are obviously unimportant thrown out). Each λ values are tested using the holdout set and the best λ is recorded.

The best predictive model was determined by taking the best λ determined in the previous step and by iteratively removing parameters from the previous feature set. The model with the best R^2 on the holdout set was chosen. The best interpretive model was chosen by taking the most important feature(s) determined in the initial ridge regression test and running brute force tests (using the same 'best' λ) to determine which features were required to achieve a moderate R^2, in comparison to the best predictive model, whilst restricting dimensionality. Tests were run to determine whether or not standardization had a large effect on model performance. If it did not, no standardization was used for the best interpretive model. The model with the best dimensionality/performance to interpretability tradeoff was chosen.

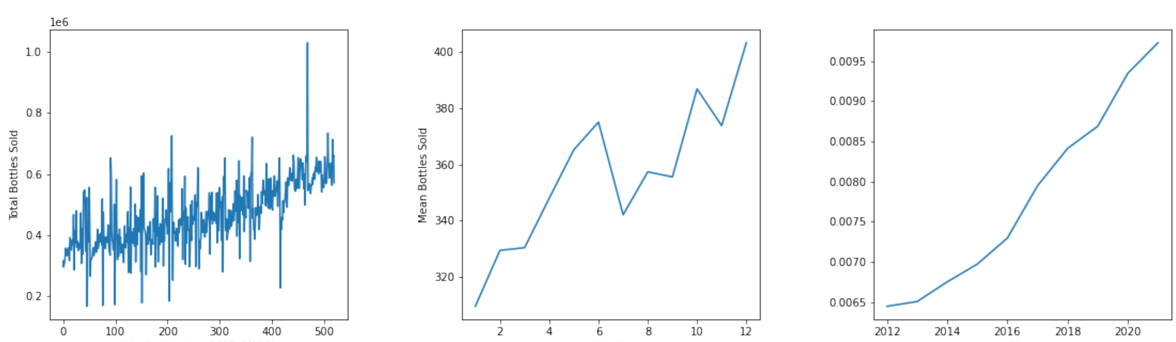
This feature selection methodology boils down to using a top-down approach in determining the best predictive model and a bottom-up approach for determining the best interpretive model.

Limitations

There is one caveat: due to computational resource limitations we were not able to include the *County:Category* interaction term in the interpretive model calculations. Our interpretive models all included *County* and *Category* separately and the addition of this interaction term introduces a very large number of very sparse dimensions that is difficult to compute the inverse of. To put it simply, we were not able to train any models with purely high dimensional sparse data. This was not a problem with any of the predictive models as these were trained with quantitative features. A possible approach for the future is provided in the discussions section.

3 Observations

In exploring the final dataset, we found a very interesting periodic trend occurring between weeks and bottles sold per capita (Figure 3a). Each period of the trend appear to match shape with the upward trend in mean bottles sold per capita vs. month (Figure 3b). In general, the mean number of bottles sold per capita increases every year (Figure 3c).



(a) Total Bottles Sold vs. Weeks (b) Mean Bottles Sold vs. Month (c) Mean Bottles Sold Per Capita vs. Year

Figure 3: Bottles Sold over Time

4 Results

Number of Bottles Sold Per Capita

The initial run (with all dependent variables, including county:category interactions, and a lambda of 10,000) showed that the Category Name, County, and County:Category features are the most important. Notably, Week 51, 26, and 21 have large, positive coefficients as well. *Year*, *Month*, and *WeeksFromStart* features were eliminated as they did not contribute to predictive power, shown in Figure 4a.

Characteristic	
Penalty λ	0.5
Features	*County* *Category Name* *County:Category* *IncomePerCap* *UnemploymentRate* *Week*
R^2	0.7366
$RMSE$	0.0068

(a) Predictive Model (standardization applied)

Characteristic	
Penalty λ	1
Features	*County* *Category Name*
R^2	0.6000
$RMSE$	0.0083

(b) Interpretive Model (no standardization applied)

Figure 4: Linear regression models predicting bottles sold per capita.

Notably, *IncomePerCap* and *UnemploymentRate* did not contribute much to model performance. We assume this is due to this data being inherently contained by the *county* attribute. We chose this as the best interpretive model as one can easily analyze how predicted bottle sales are affected by the *county* and *category*, shown in Figure 4b. The largest coefficients (Whiskey, Vodka, and Dickinson) show that, when all else held constant, we predict bottle sales to increase by 0.0263 for Whiskey sales, 0.0224 for Vodka Sales, and 0.0153 for Dickinson County. Including the *Week* variable did not increase R^2 noticeably and thus it was not added to the interpretive model.

Sales ($) Volume Per Capita

Our initial run taught us that Year, WeeksFromStart, and UnemploymentRate are not very important features in predicting Sales Volume Per Capita. These predictors were still included in the predictive model, shown in Figure 5a.

Characteristic	
Penalty λ	0.5
Features	*County* *Week* *Month* *Year* *Category Name* *IncomePerCap* *UnemploymentRate* *County:Category*
R^2	0.6829
$RMSE$	0.0995

(a) Predictive Model (standardization applied)

Characteristic	
Penalty λ	0.5
Features	*County* *Category Name*
R^2	0.5422
$RMSE$	0.1180

(b) Interpretive Model (no standardization applied)

Figure 5: Linear regression models predicting sales per capita.

The chosen best interpretive model matches the feature set of the previous target variable, shown in Figure 5b. Here, Whiskey by itself has the largest coefficient of 0.3645, followed by Dickinson County and Vodka with 0.236

and 0.22 coefficients respectively. Interestingly, Fremont County and Mezcal have some of the largest coefficients of -0.125 and -0.117 respectively. This would indicate that we expect Sales Volume Per Capita to decrease by $\approx \$0.12$ for Mezcal sold in a given week. This model obviously gives poor predictive results in certain scenarios: we would predict negative sales per capita for Mezcal in Fremont County (this is a scenario in which not being able to include the *County:Category* interaction term was important).

Volume (L) of Sold Alcohol Per Capita

The first harsh ridge penalization taught us that *IncomePerCap*, *Year*, and *UnemploymentRate* are not important features in predicting Volume of Sold Alcohol Per Capita. These predictors were still included in the predictive model, shown in Figure 6a.

Characteristic	
Penalty λ	0.5
Features	*County* *Week* *Month* *Year* *Category Name* *IncomePerCap* *UnemploymentRate* *County:Category*
R^2	0.6790
$RMSE$	0.0075

(a) Predictive Model (standardization applied)

Characteristic	
Penalty λ	1
Features	*County* *Category Name* *Month*
R^2	0.5560
$RMSE$	0.0087

(b) Interpretive Model (no standardization applied)

Figure 6: Linear regression models predicting volume sold per capita.

County and *Category Name* still continue to be vital predictors, hence included in the final interpretive model, shown in Figure 6b. Additionally, *Month* became another significant predictor as well. Whiskey and Vodka had the largest coefficients of 0.025 and 0.023 followed by Dickinson County with 0.017. Thus, we expect the Volume of Alcohol Sold to be 0.042 Liters for Whiskey in Dickinson County.

4.1 *Comparative Analysis*

The difference between the most accurate and interpretable models is the inclusion of time variables such as *Week* and *IncomePerCap* or *UnemploymentRate*. Our predictive model for the number of bottles sold per capita had an impressive $R^2 \approx 0.74$ while the interpretive model had an $R^2 \approx 0.60$. From this we can tell that the number of bottles sold is mostly a function of the county and category of alcohol. It is a similar story for the other two target variables.

The best predictive models always employed the *County*, *Category*, *County:Category*, and *Week* attributes. Including these features always led to a model with a better R^2 on the holdout set.

4.2 *Model Observations*

Dickinson County by itself, or as an interaction term with Vodka or Whiskey, always had one of the largest coefficients. After inspection, we see that Dickinson County is a vacation destination and thus these alcohol sales numbers (as well as the large sales of Vodka and Whiskey) make sense. Other counties that often have large coefficients are Polk and Cerro Gordo. Both of these are large metropolitan areas. This seems to indicate that larger cities have larger per capita liquor sales than smaller ones. On its own, the category Whiskey often had one of the largest coefficients. This indicates that liquor stores in Iowa sell more Whiskey than other alcohol types.

When the Week feature showed large coefficient values it did so for Weeks 51, 26, and 21. After some light research, we see that Week 51 and 26 line up with New Years and Independence Day, respectively. However, we were not able to come up with an explanation for why liquor stores in Iowa may be purchasing more in Week 21 (all 3 have positive coefficients). It may be that Week 21 is near the start of summer (early June) and liquor stores are stocking up for upcoming summertime events.

Each of the target values seem to have a sinusoidal trend over time. We account for this partially by treating *Week* as a categorical variable (large/small sales weeks will have large magnitudes) but this would be improved by incorporating seasonality into the model.

5 Discussion

This section discusses the ethics of our model and investigating potential improvements for possible future development.

5.1 Ethics

The predictive models may serve to benefit licensed liquor store owners who pursue opening more establishments in Iowa. Additionally, it can optimize how much liquor per category should be purchased to make profit while avoiding overstocking. The interpretive models may give insight to researchers and/or commercial establishments on variables that affect alcohol consumption.

On the other hand, the predictive model can be exploitative in nature. Commercial establishments may take advantage of communities negatively impacted by alcohol consumption. These models only view counties as predictors of alcohol consumption irregardless of community vulnerabilities to alcoholism. Using these models to understand where and how liquor should be advertised could lead to people in these communities becoming alcoholics or cause current alcoholics to relapse.

5.2 Suggestions

As previously stated, we believe interaction terms can make a significant improvement to the models given. We no longer pursued incorporating interactions in our models due to limited computational power and time. An approach we attempted is developing a ridge regression model that **estimates** the best coefficients by reducing the dimension of the original predictors using singular value decomposition. This dimensional reduction strategy is a modification of Dr. Wieringen's strategy to optimize computing ridge regression estimators [6].

Rather than inverting the a large sparse design matrix X, decompose X using singular value decomposition where Σ is the diagonal of the top k singular values of largest magnitude (k is chosen arbitrarily) and V is its column of corresponding components (the resulting matrices should still be multiplication compatible). Then let $R = U\Sigma$

$$X \Rightarrow U\Sigma V^T = RV^T$$

Then substitute X with RV^T in the ridge estimator to arrive at this result:

$$\beta(\lambda) = V(R^T R + \lambda I)^{-1} R^T Y$$

Where I is the same dimension as $R^T R$. Although matrix inversion is still required, inverting a $k \times k$ matrix where k is less than the dimension of $X^T X$ can be a potentially vital improvement to save computation power and time.

Additionally, we would be interested in fitting a seasonality model for sales as we observed data trending in a rising, periodic nature. Some interesting feature engineering can be incorporated into this model that encapsulates the periodic trend, such as holiday indicators.

6 Conclusion

This project produced compelling insights on predictors of alcohol consumption per capita utilizing the Iowa Liquor Sales dataset as a proxy. Our models suggest that county and category of liquor are the largest predictors in alcohol sales, with its interaction variable suggesting it's more than just the sum of its parts. These models were created in hopes of developing meaningful insights with ethics taken into consideration.

References

[1] Bls data finder. https://beta.bls.gov/dataQuery/find?q=iowa+unemploymentq=iowa+unemployment%20rate. Public Domain U.S. Government License.

[2] COMMONS, C. About cc licenses. https://creativecommons.org/about/cclicenses/.

[3] IOWA DEPARTMENT OF COMMERCE, A. B. D. Iowa liquor sales. https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy, November 2014. Creative Commons License (CC0).

[4] U.S. CENSUS BUREAU, P. D. County population in iowa by year. https://data.iowa.gov/Community-Demographics/County-Population-in-Iowa-by-Year/qtnr-zsrc, July 2015. Public Domain U.S. Government License.

[5] U.S. DEPARTMENT OF COMMERCE, B. O. E. A. Annual personal income for state of iowa by county. https://data.iowa.gov/Economic-Statistics/Annual-Personal-Income-for-State-of-Iowa-by-County/st2k-2ti2, January 2020. Public Domain U.S. Government License.

[6] WIERINGEN, W. N. V. Lecture notes on ridge regression. https://arxiv.org/pdf/1509.09169.pdf, May 2021. Creative Commons License (Attribution-Non Commercial-Share Alike).

7 Appendix

This appendix holds information on all datasets used for this project: dataset name, column name, column description, and licensing.

Iowa Liquor Sales

Licensing: Creative Commons Zero
URL: https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy

Attribute Name	Date Type	Description
Invoice/Item Number	Text	Concatenated invoice and line number associated with the liquor order.
Date	Date Time	Date of order.
Store Number	Text	Unique number assigned to the store who ordered the liquor.
Store Name	Text	Name of store who ordered the liquor.
Address	Text	Address of store who ordered the liquor.
City	Text	City where the store who ordered the liquor is located.
Zip Code	Text	Zip code where the store who ordered the liquor is located.
Store Location	Point	Location of store who ordered the liquor.
County Number	Text	Iowa county number for the county where store who ordered the liquor is located.
County	Text	County where the store who ordered the liquor is located.
Category	Text	Category code associated with the liquor ordered.
Category Name	Text	Category of the liquor ordered.
Vendor Number	Text	The vendor number of the company for the brand of liquor ordered.
Vendor Name	Text	The vendor name of the company for the brand of liquor ordered.
Item Number	Text	Item number for the individual liquor product ordered.
Item Description	Text	Description of the individual liquor product ordered.
Pack	Number	The number of bottles in a case for the liquor ordered.
Bottle Volume (mL)	Number	Volume of each liquor bottle ordered in milliliters.
State Bottle Cost	Number	The amount that Alcoholic Beverages Division paid for each bottle of liquor ordered.
State Bottle Retail	Number	The amount the store paid for each bottle of liquor ordered.
Bottles Sold	Number	The number of bottles of liquor ordered by the store.
Sale (Dollars)	Number	Total cost of liquor order (number of bottles multiplied by the state bottle retail).
Volume Sold (Liters)	Number	Total volume of liquor ordered in liters. (i.e. (Bottle Volume (ml) x Bottles Sold)/1,000).
Volume Sold (Gallons)	Number	Total volume of liquor ordered in gallons. (i.e. (Bottle Volume (ml) x Bottles Sold)/3785.411784).

County Population in Iowa by Year

Licensing: Public Domain U.S. Government
URL: https://data.iowa.gov/Community-Demographics/County-Population-in-Iowa-by-Year/qtnr-zsrc

Attribute Name	Date Type	Description
FIPS	Number	A five-digit code for counties based on Federal Information Processing Standards Publication.
County	Text	County name
Year	Date Time	For population estimates the date of reference is always July 1. The decennial census has a date of April 1.
Population	Number	Population
Primary Point	Point	Primary latitude and longitude in decimal degrees for the county.

Annual Personal Income for State of Iowa by County

Licensing: Public Domain U.S. Government
URL: https://data.iowa.gov/Economic-Statistics/Annual-Personal-Income-for-State-of-Iowa-by-County/st2k-2ti2

Attribute Name	Date Type	Description
Row ID	Text	Unique row identifier for the variable and time period.
Geography ID	Text	A federal processing information series five-digit code for the identification of counties and county equivalents of the United States.
Name	Text	Name of the geography.
Variable Code	Text	Unique identifier for the variable reported.
Variable	Text	Description of the variable.
Value	Number	Value reported for the time period.
Variable Unit	Text	Units associated with the value reported.
Date	Date Time	Date the annual period ended that the reported value is for.
Location	Point	Primary lat and long for the geography.

Unemployment Rate

Licensing: Public Domain U.S. Government
URL: https://beta.bls.gov/dataQuery/find?q=iowa+unemploymentq=iowa+unemployment%20rate

Attribute Name	Date Type	Description
County	Text	County Name
Year	Text	Year
Month	Text	Month
UnemploymentRate	Number	Unemployment rate