Analysis of Alcohol Sales in Iowa

ANTHONY R, JADE T, JASMINE G, AND MADISON B

MOTIVATION

HISTORY

- Before and during the prohibition era to 1966
- Current industry (2023)
  - 100 breweries
  - 120 wineries
  - 45 distilleries
- COVID-19
  - To-go cocktails
  - Spike in drinking
- Iowa has one of the highest binge drinking rates in the US
  - 22.6% vs 18.4%

ADDITIONAL MEASURES

- Detailed publication of alcohol sales transactions
- State policies:
  - Excise taxes and minimum unit pricing
  - Dram shop liability
  - Limiting days and hours of sales
- Local policies:
  - Conditional use permits
  - Noise ordinances
  - Alcohol advertising ordinances
  - Enforcement
  - Screening and intervention
  - RBS training

PROBLEM

- Use Iowa's past alcohol sales trends to track any changes and their possibilities:
  - Policies
  - Demographics
  - Population
  - COVID-19
  - Cultural movement

STATE OF FIELD

- Stores
  - Liquor sales trends
  - Store competition
  - Product variety
- Geographical maps
  - Density of alcohol outlets
  - Relationship with health, crime and consumption

DATASETS

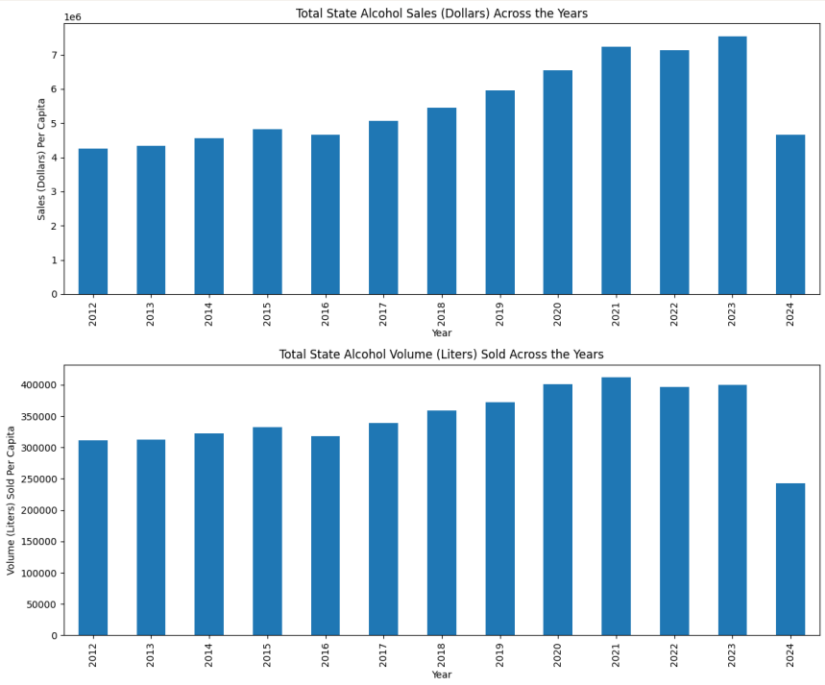
- Iowa Alcohol Sales
- Census data for population and income
- Iowa Geospatial Dataset

METHODS

EXPLORATORY DATA ANALYSIS

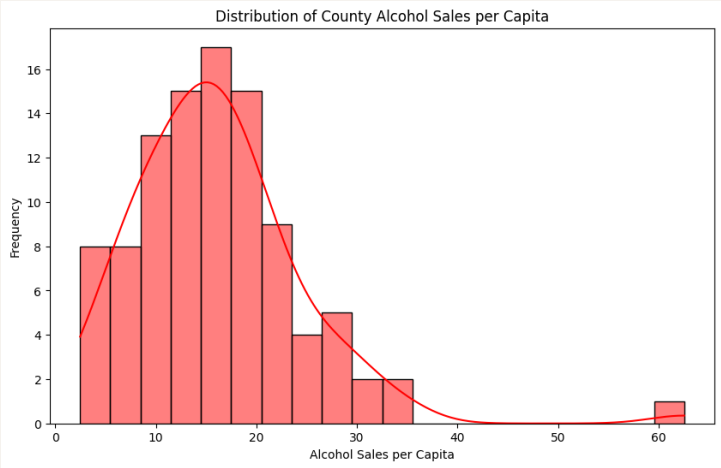
- Performed our analysis on random sample of liquor sales data
  - 29.9 million records -> 500,000 records
- Data munging for usability
  - Dropping null values, unneeded features, changing data types, renaming features
- Added in calculated features, such as ‘Average Population’ and ‘Average Income’
- Focused on understanding the inherit distributions within our different datasets

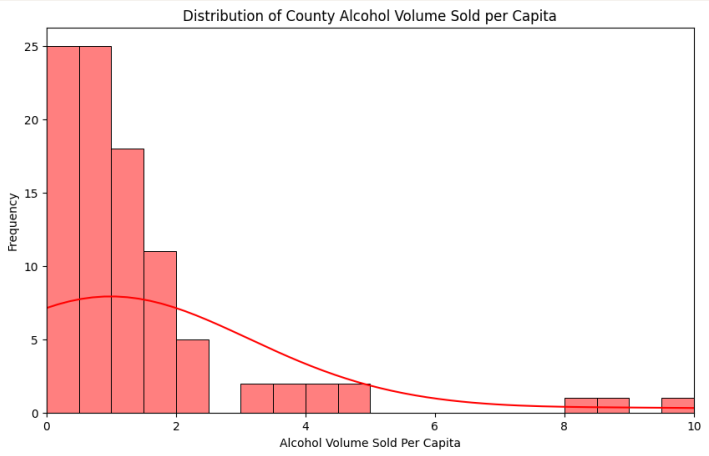
ALCOHOL SALES AND VOLUME



- Sales increasing however volume sold seems to be increasing at a much slower rate showing that alcohol has been getting more expensive.

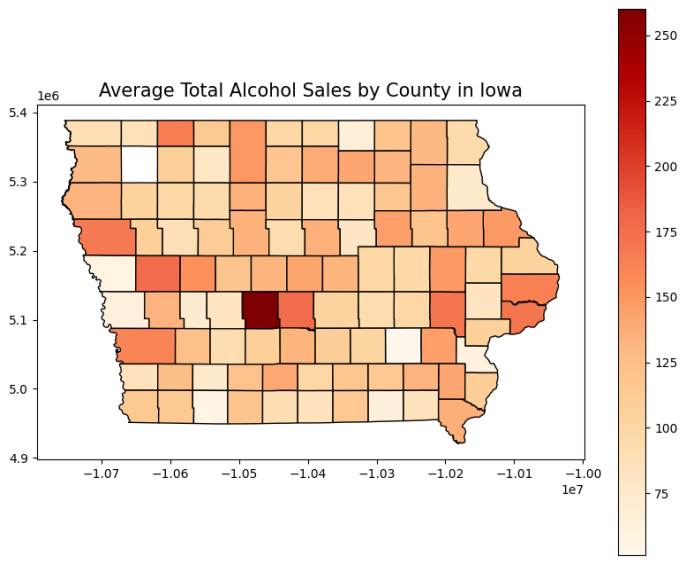
ALCOHOL SALES AND VOLUME PER CAPITA

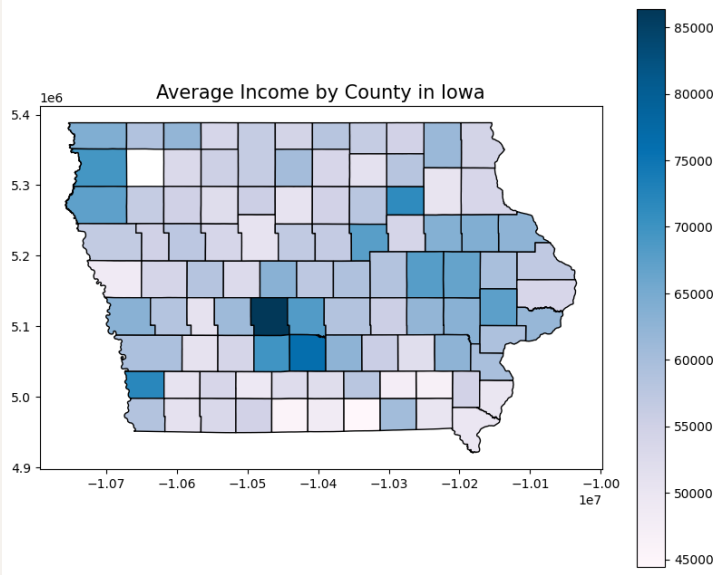




- Sales has a decent range where volume sold is concentrated to 2 liters or less.

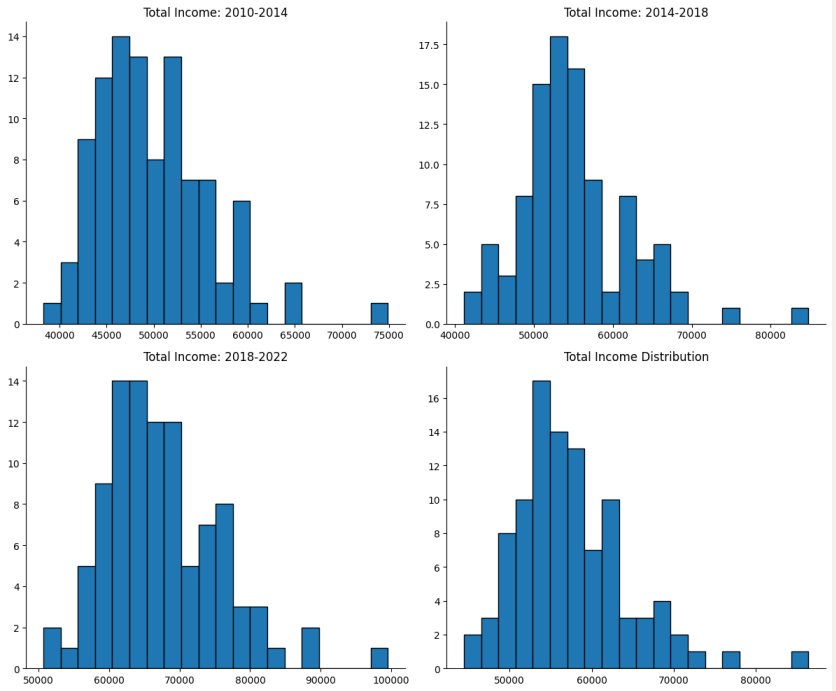
GEOSPATIAL VISUALIZATIONS



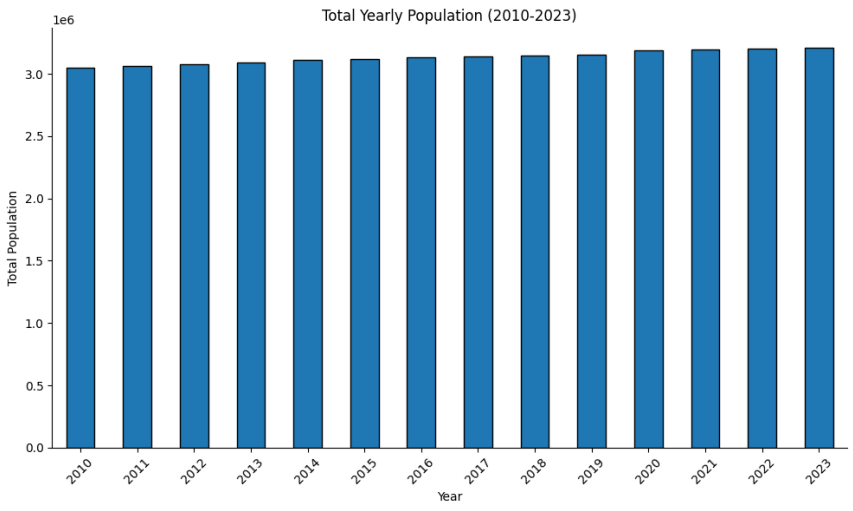


AVERAGE INCOME

- Income distribution moving more to the right as time goes on

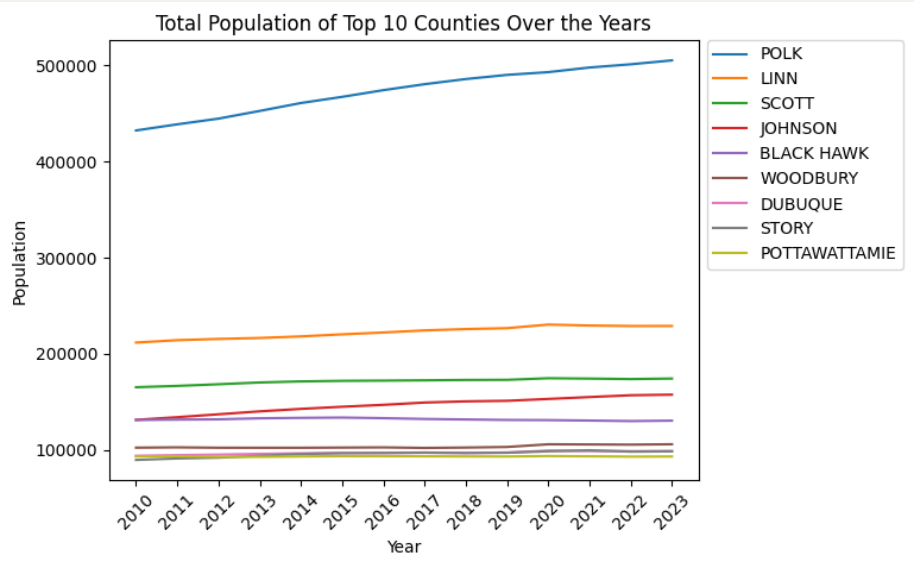


IOWA POPULATION DATASET



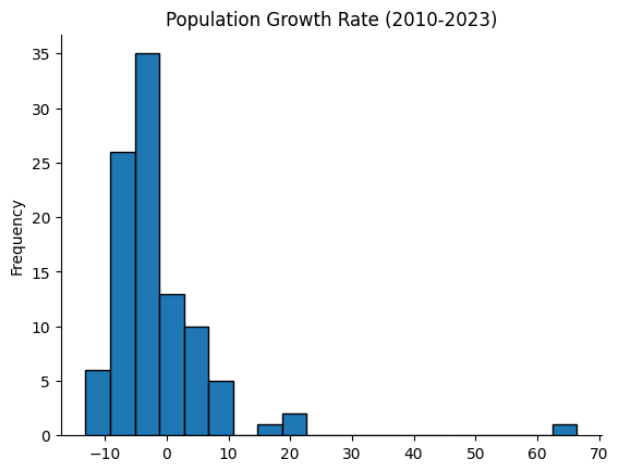
- Population increasing at a modest rate

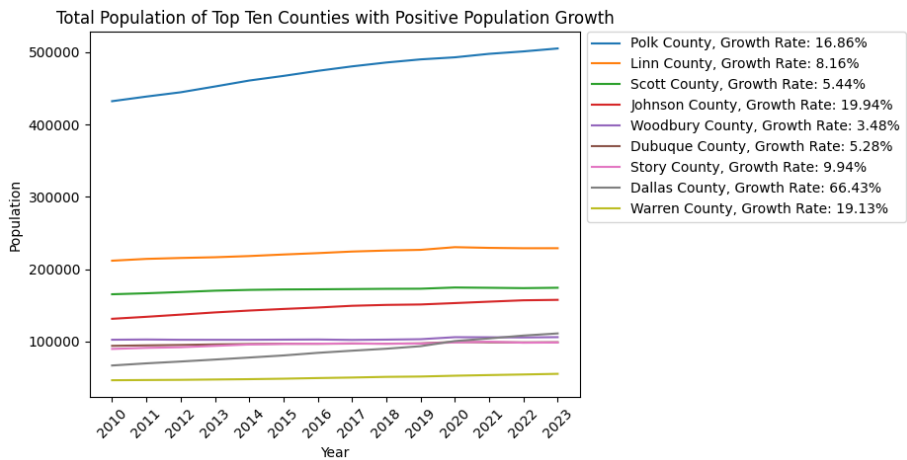
COUNTY POPULATION



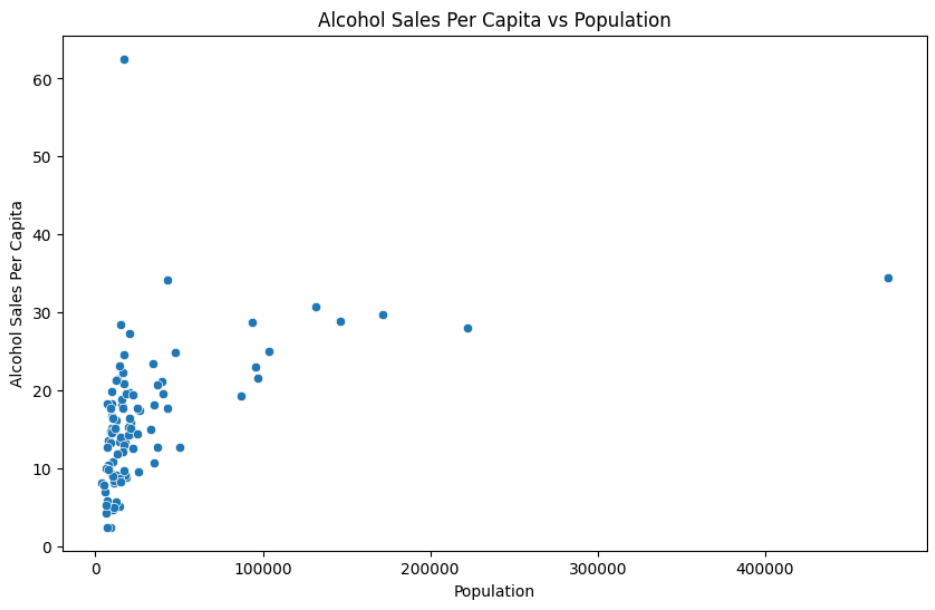
- All these counties seem to be increasing in population
- Polk, the most populous, is increasing at a much faster rate than the rest.

POPULATION GROWTH RATE



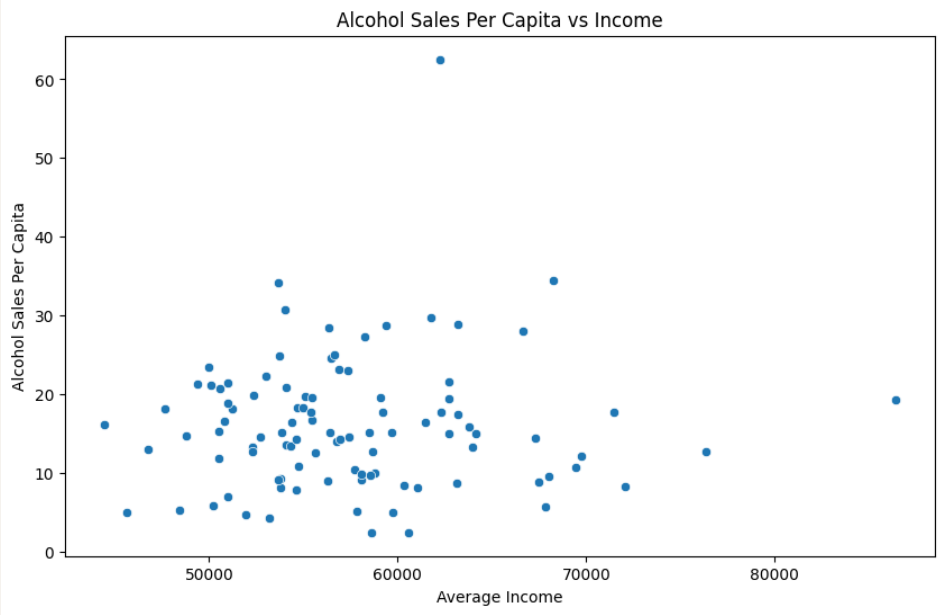


MULTIVARIABLE ANALYSIS



- Correlation between Alcohol Sales per Capita and Population is 0.46

MULTIVARIABLE ANALYSIS



- Correlation between Alcohol Sales per Capita and Income is 0.05

SUPERVISED LEARNING METHODS

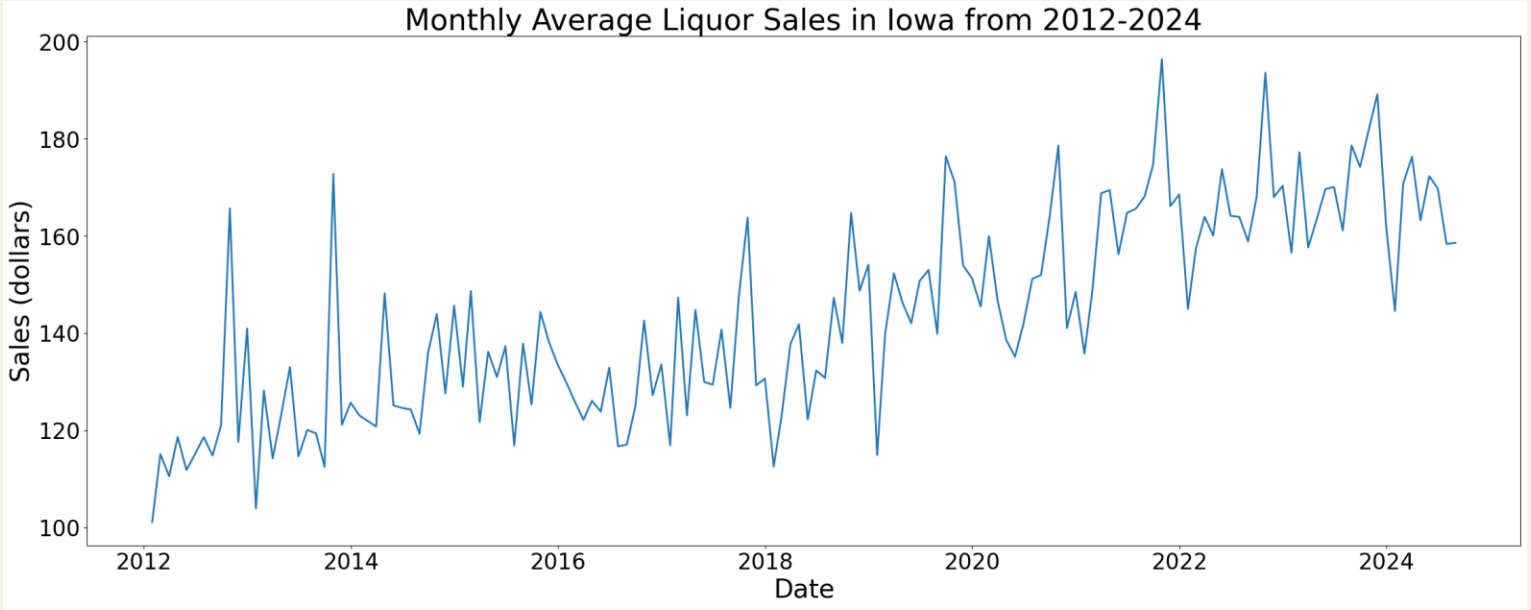
- Regression Models
  - Linear Regression and Random Forest Regression
  - Predict total sales amount within county based on features
    - Income, population, number of vendors, alcohol variety, age, universities, religion, rural vs urban
  - Identify which features were most important indicators in predicting liquor sales
- Time Series Analysis
  - Create a SARIMA model to predict future liquor sales
  - Identify any seasonal components and trends to liquor sales

UNSUPERVISED LEARNING

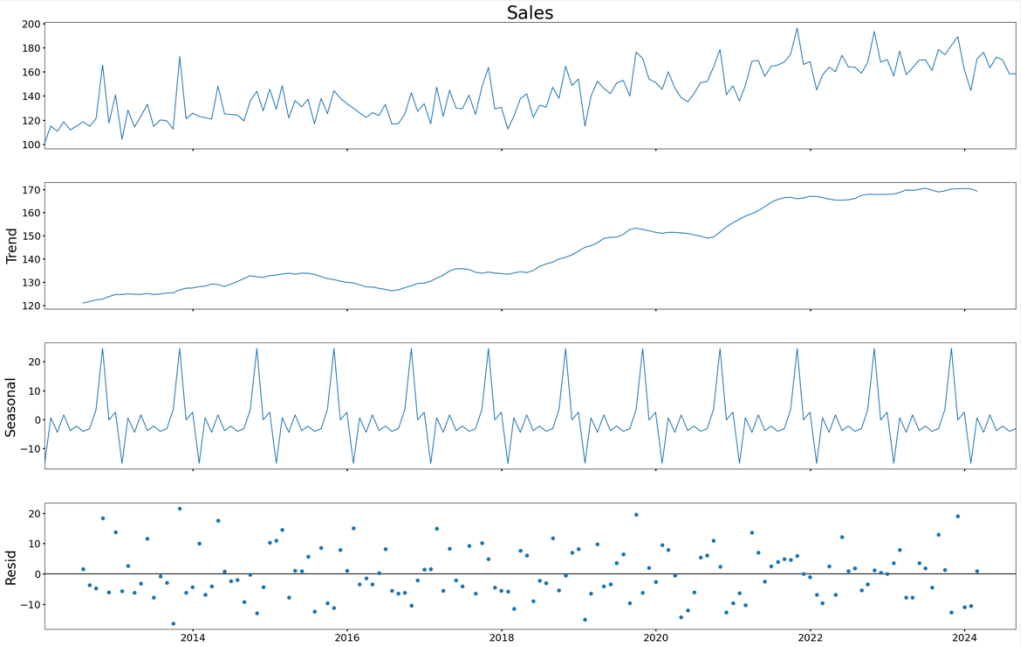
- Principle Component Analysis
  - Dimensionality Reduction
    - Reduce number of features down to 2 for clustering
  - Standard Scaler
    - Different features contain numeric data on different scales
- K-Means Clustering
  - Elbow Method for number of clusters
  - Appended cluster assignments to data frame
  - Looked at descriptive statistics of the clusters

RESULTS

TIME SERIES ANALYSIS

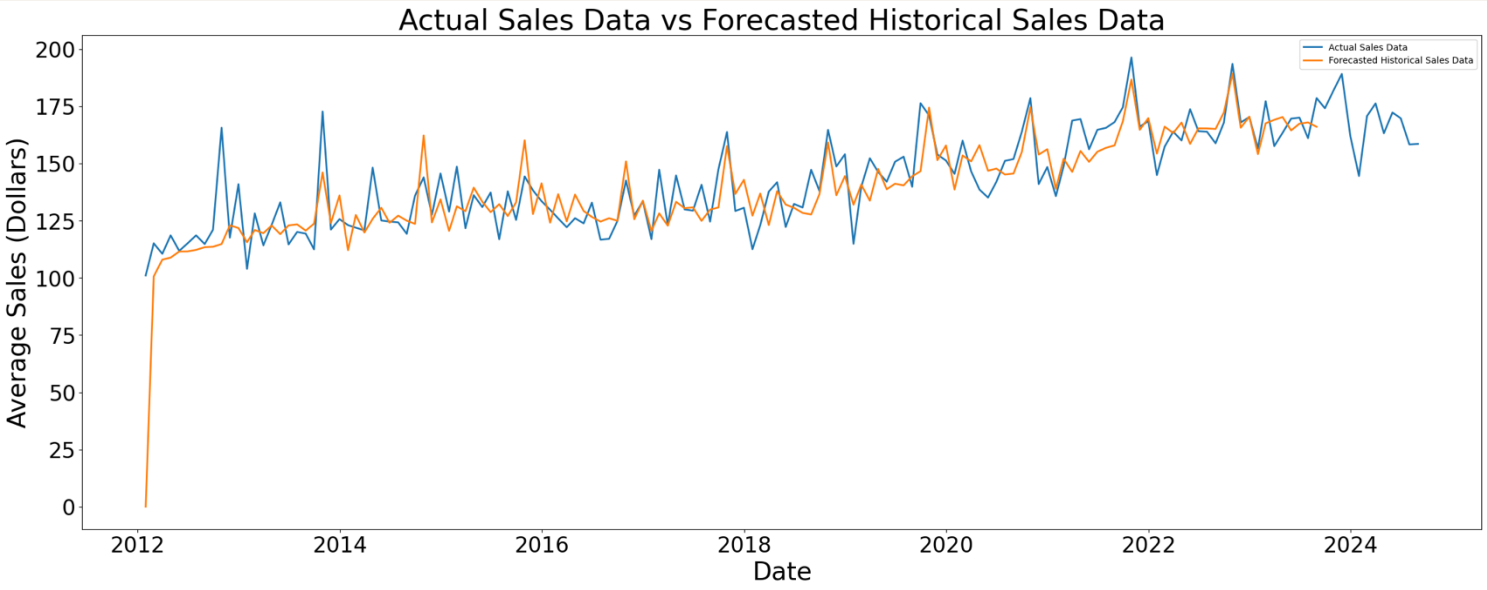


TIME SERIES ANALYSIS

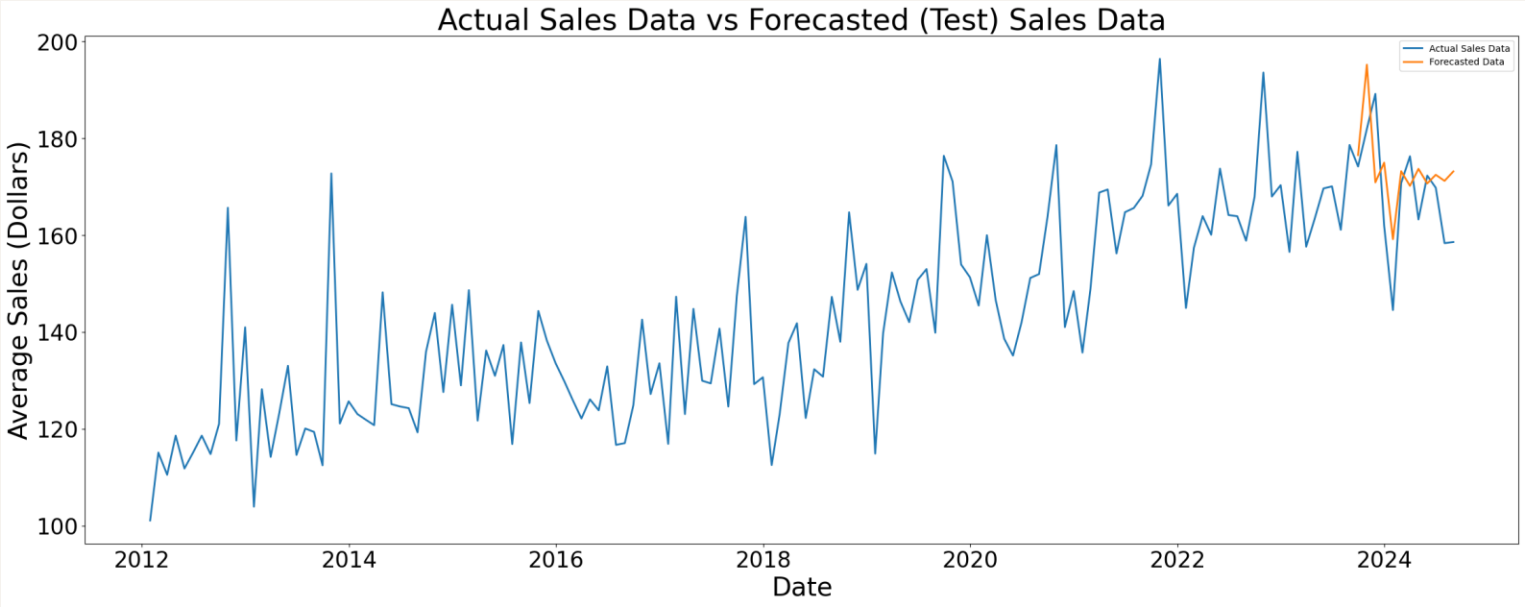


- Liquor sales component breakdown:
  - Upward trend shows increase in overall sales
  - Seasonal component shows yearly spike during holiday season

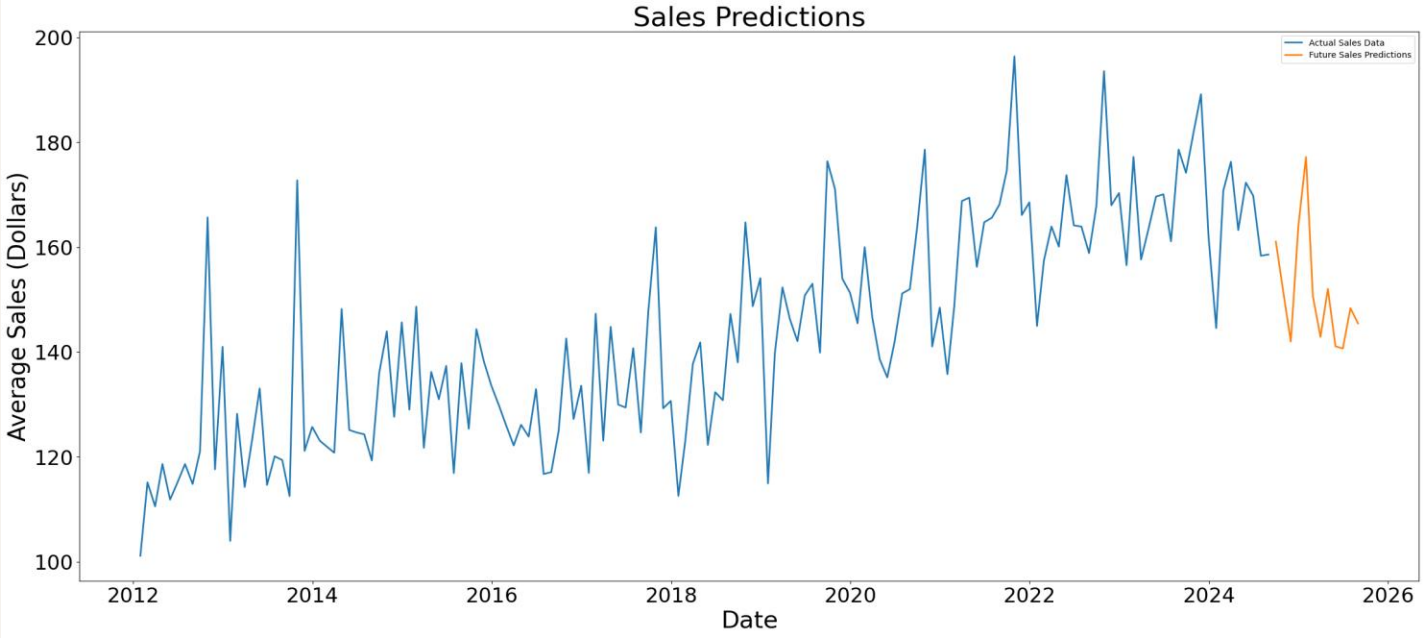
TIME SERIES ANALYSIS



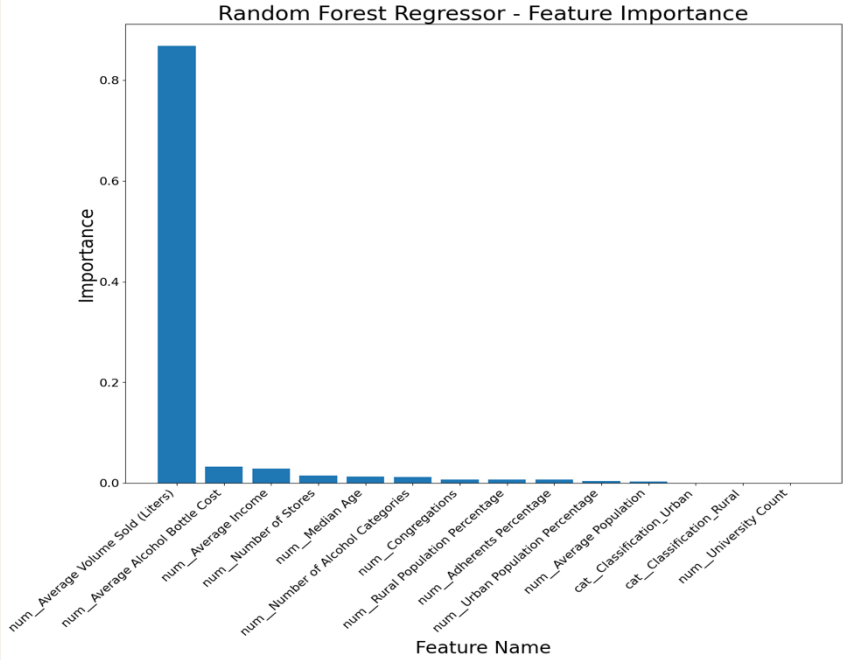
TIME SERIES ANALYSIS



TIME SERIES ANALYSIS

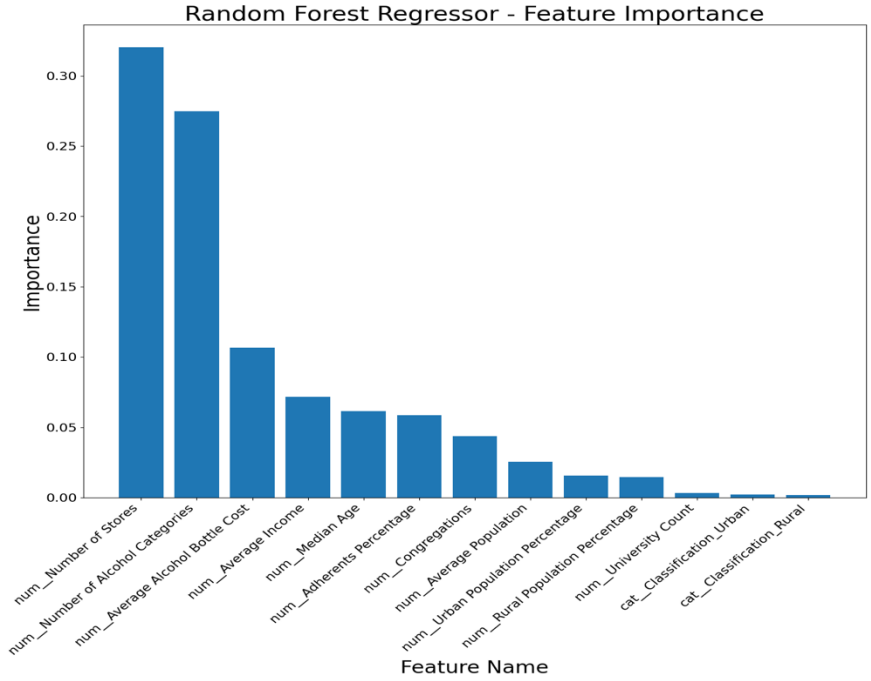


REGRESSION MODELS RESULTS



- Linear Regression
  - RMSE: 11.650
  - R^2 Score: 0.819
- Random Forest:
  - RMSE: 11.705
  - R^2 Score: 0.817

REGRESSION MODELS RESULTS



- Linear Regression
  - RMSE: 25.399
  - R^2 Score: 0.140
- Random Forest:
  - RMSE: 27.072
  - R^2 Score: 0.0224

CLUSTERING

- Main factors differentiating the clusters were sales amount, volume amount, and cost per bottle.
- Shows that there are high volume sales transactions as well as a market for high-end liquor
  - The high-end market is dominated by urban areas as roughly 50% of these sales were in the top 5 most populous counties
  - High-volume transactions likely due to the major urban areas

CONCLUSION

RESEARCH QUESTION AND APPROACH

- **Research Question**: *How can analyzing liquor sales trends provide insights into broader lifestyle changes and shifts in Iowa's population?*
- **Methodology**:
  - Used Iowa liquor sales and census data.
  - Employed exploratory data analysis, supervised learning models, and time series forecasting.

KEY FINDING 1: POPULATION AND URBANIZATION

- **Urban counties** like Polk and Linn report higher alcohol sales.
- **Population size** is a primary driver of consumption.
- **Urbanization and vendor availability** are key factors influencing sales.
- Social behaviors linked to population density have a larger effect than economic factors.

KEY FINDING 2: ROLE OF INCOME

- Income had a **weaker correlation** with alcohol sales.
- Cultural and social factors were more significant.
- **Wealthier counties** did not always have higher consumption.
- Emphasizes the role of **social environments** in urban areas.

FORECASTED DECLINE IN SALES

- SARIMA time series forecast predicts a **decline** in future alcohol sales.
- Likely driven by increasing **health-conscious behaviors**.
- Impact of **COVID-19** and movements like “sober curious.”
- Social and health trends are reshaping alcohol consumption.

IMPLICATIONS FOR POLICY AND RESEARCH

- **Urbanization and population density** should guide public health strategies.
- Policymakers should regulate **alcohol vendor density** in urban areas.
- Focus on **targeted public health campaigns** to reduce excessive drinking.
- Further research on **social and demographic factors** (age, education, ethnicity).

FINAL THOUGHTS AND RECOMMENDATIONS

- Alcohol consumption in Iowa is shaped by **urbanization, social factors**, and **health trends**.
- Public health interventions are needed to address the risks of high consumption in **urban centers**.
- Promote **health-conscious behaviors** through education and campaigns.
- Future research should explore **per capita trends** and focus on regulating alcohol accessibility.

THANK YOU!