

Project Proposal B

Client: Drinking Excess Alcohol is Dangerous (DEAD)

Hailey Ernest, James Irwin, Elliot Kunz, Kaatje Matthews-vanKoetsveld

2025-10-03

Data Summary

The data that we are using is provided by the state of Iowa and contains information about wholesale spirit purchases made by Iowa Class “E” liquor licensees. Class “E” licenses are for grocery stores, liquor stores, and convenience stores, among other establishments.

The data source is linked here: https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy/about_data

We collected this data with an API (virtual endpoint for data collection) and will be using data from January 1st, 2022 and onward.

Additionally, we will be using Iowa population data provided by the Iowa Government: <https://catalog.data.gov/dataset/city-population-in-iowa-by-county-and-year>

Model Creation, Selection, and Validation

In order to help Drinking Excess Alcohol is Dangerous (DEAD) accurately analyze patterns in sales to help them understand what factors drive higher or lower alcohol purchases, we need to understand the patterns across different counties and time periods. To achieve this goal, we have grouped the data to see each county’s total sale liters, average sale liters, number of transactions, and sales by alcohol category, all organized by year. Some visualizations of this data are shown here:

```
# imports
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
grouped_data_dead = pd.read_csv("grouped_data_dead.csv")
```

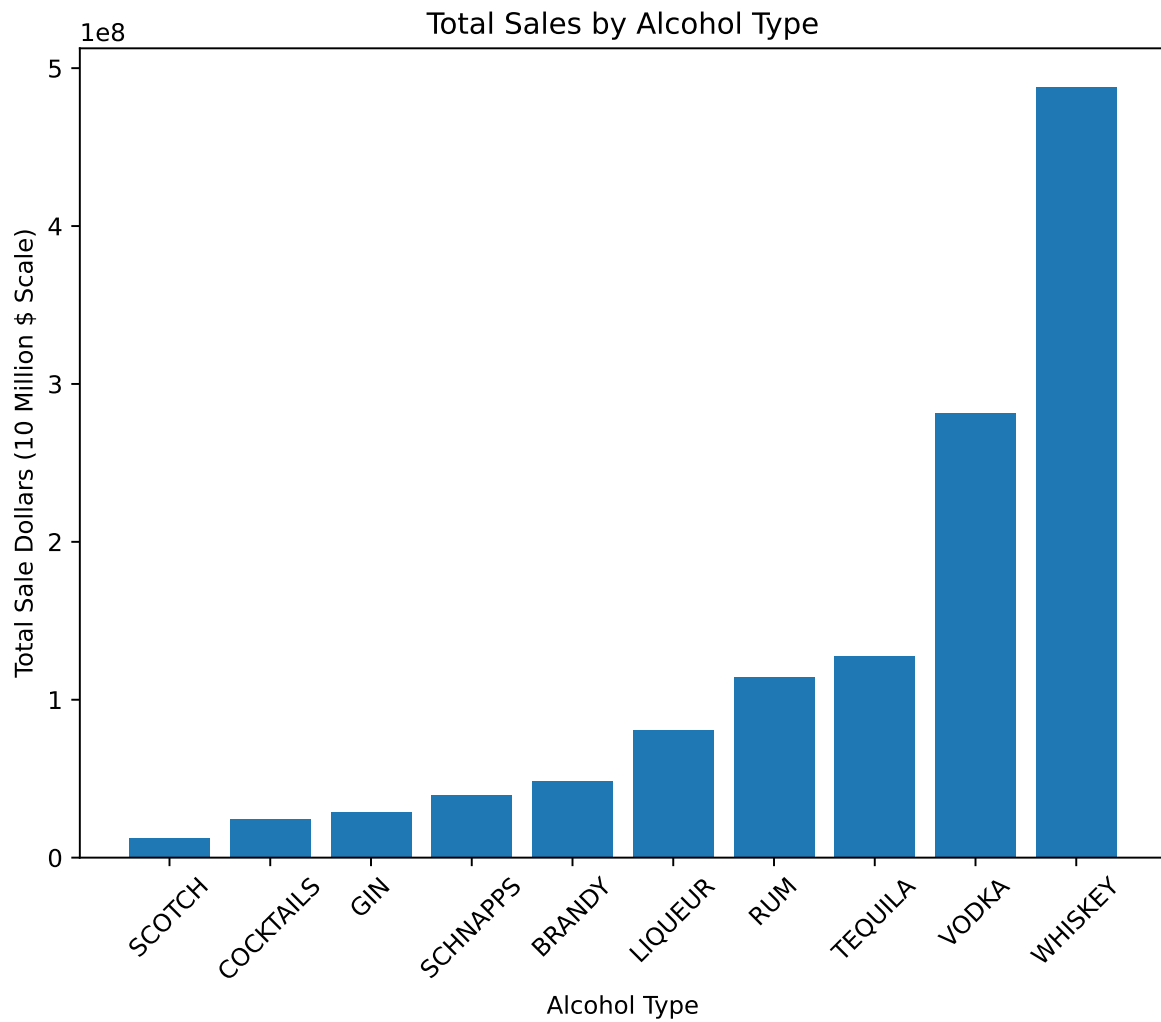
```

# Total Sales by Each Catagory of Alcohol
grouped_data_graphing = grouped_data_dead.reset_index()
plot_data = (grouped_data_graphing.groupby('alcohol_type')
             ['sum_sale_dollars'].sum().reset_index()
             .sort_values(by='sum_sale_dollars'))
plt.figure(figsize=(8,6))
plt.bar(plot_data['alcohol_type'], plot_data['sum_sale_dollars'])
plt.ylabel("Total Sale Dollars (10 Million $ Scale)")
plt.xlabel("Alcohol Type")
plt.title("Total Sales by Alcohol Type")
plt.xticks(rotation = 45)
plt.show()

# Plot Total Sale Dollars to Each County
plot_data_2 = (grouped_data_graphing.groupby('county')
              ['sum_sale_dollars'].sum().reset_index()
              .sort_values(by='sum_sale_dollars'))
plt.figure(figsize=(10,4))
plt.bar(plot_data_2['county'],
        plot_data_2['sum_sale_dollars'])
plt.ylabel("Total Sale Dollars (10 Million $ Scale)")
plt.xlabel("County")
plt.title("County Spending")
plt.xticks(rotation=90)
plt.show()

# Plot only the 25 highest Counties
plt.figure(figsize=(10,4))
plt.bar(plot_data_2['county'].tail(25),
        plot_data_2['sum_sale_dollars'].tail(25))
plt.ylabel("Total Sale Dollars (10 Million $ Scale)")
plt.xlabel("County")
plt.title("County Spending (Top 25 Counties)")
plt.xticks(rotation=90)
plt.show()

```



determine the impact they have and find the drive for alcohol sales. This will provide the best and full analysis of patterns on high purchases to inform DEAD and help make alcohol purchases safer in Iowa.

Conclusions and Deliverables

After our analysis, you will have a complete understanding of what factors drive high alcohol purchases in the counties of Iowa. Also, you will have detailed results of trends by month, by region, and by type of alcohol to create better safety laws on alcohol purchases. We appreciate the opportunity to working with you and are looking forward to performing our in depth analysis.