Project Proposal A

Client: Booze 'R' Us

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

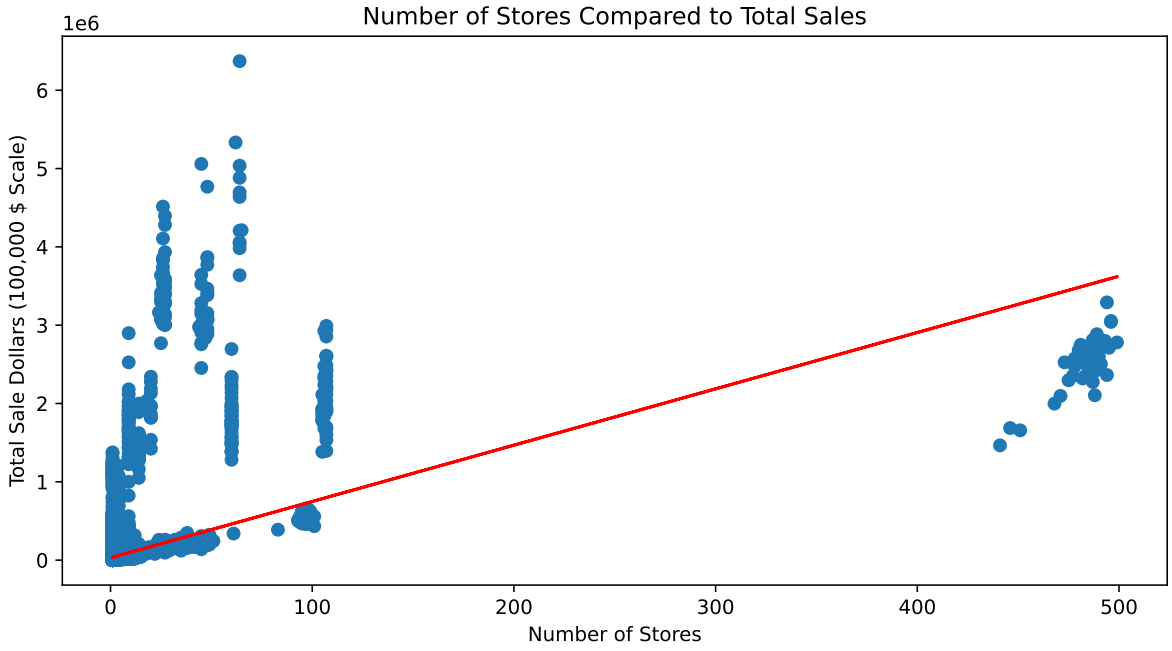
We collected this data with an API (virtual endpoint for data collection) and will be using data from January 1st, 2022 and onward.

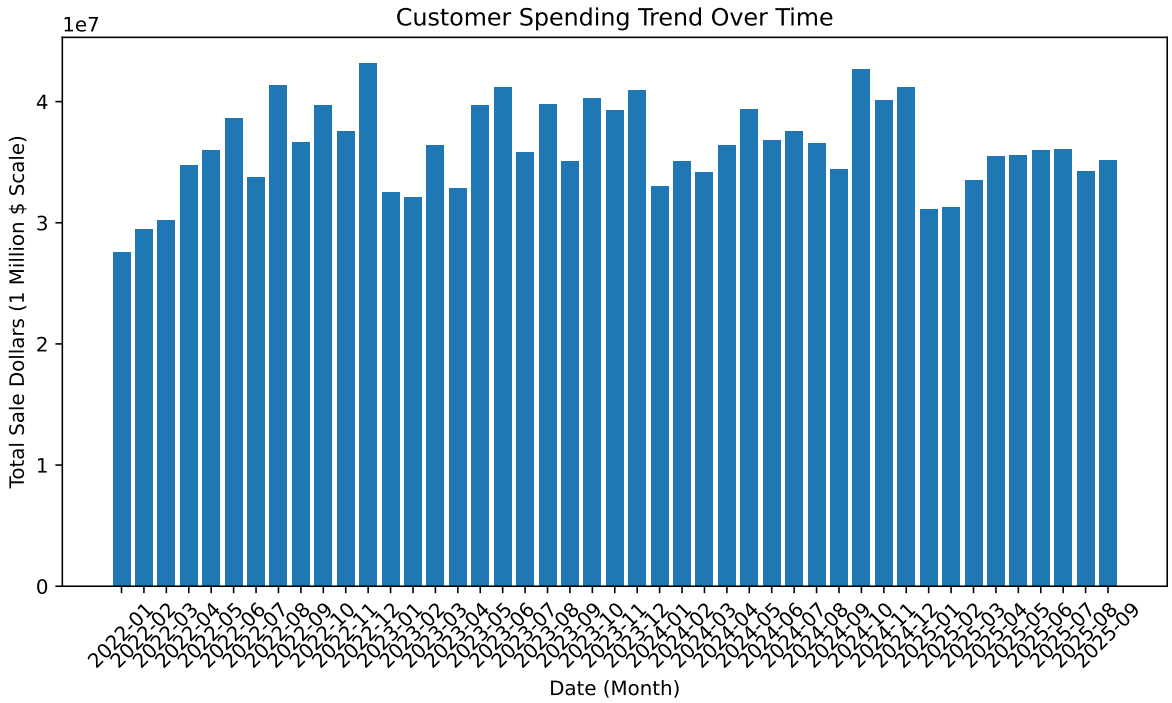
Model Creation, Selection, and Validation

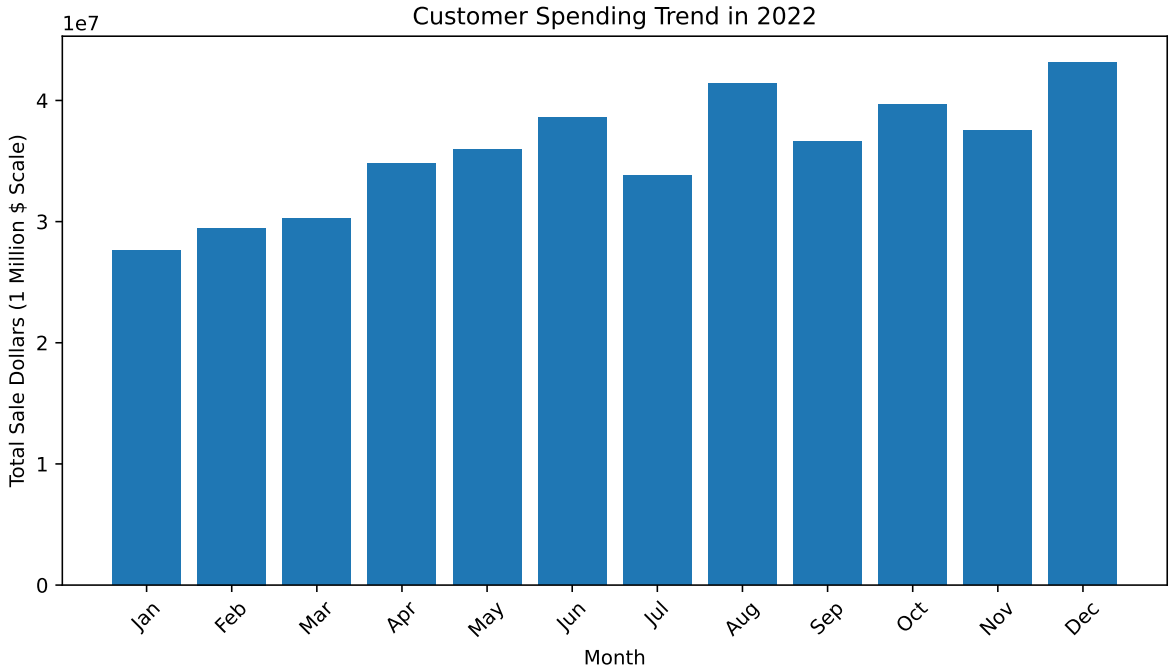
In order to help Booze 'R' Us accurately project sales and confidently expand into other states, we need to understand the characteristics and sales of other chains that sell liquor. To achieve this goal, we have grouped the data to see each chain's total sale dollars, average sale dollars, number of bottles sold, number of transactions, and number of stores all organized by month and year. Some visualizations of this data are shown here:

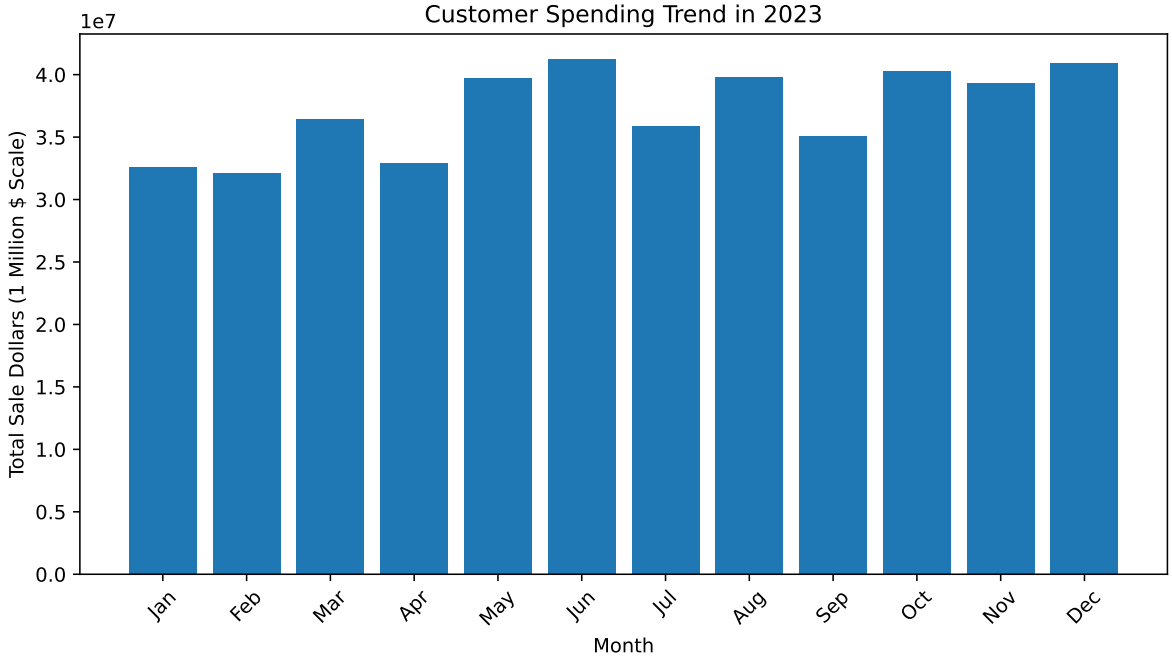
```
# imports
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
grouped_data_booze = pd.read_csv("grouped_data_booze.csv")

# Plot Number of Stores vs Total Sales
slope, intercept = np.polyfit(grouped_data_booze['num_stores'],
```

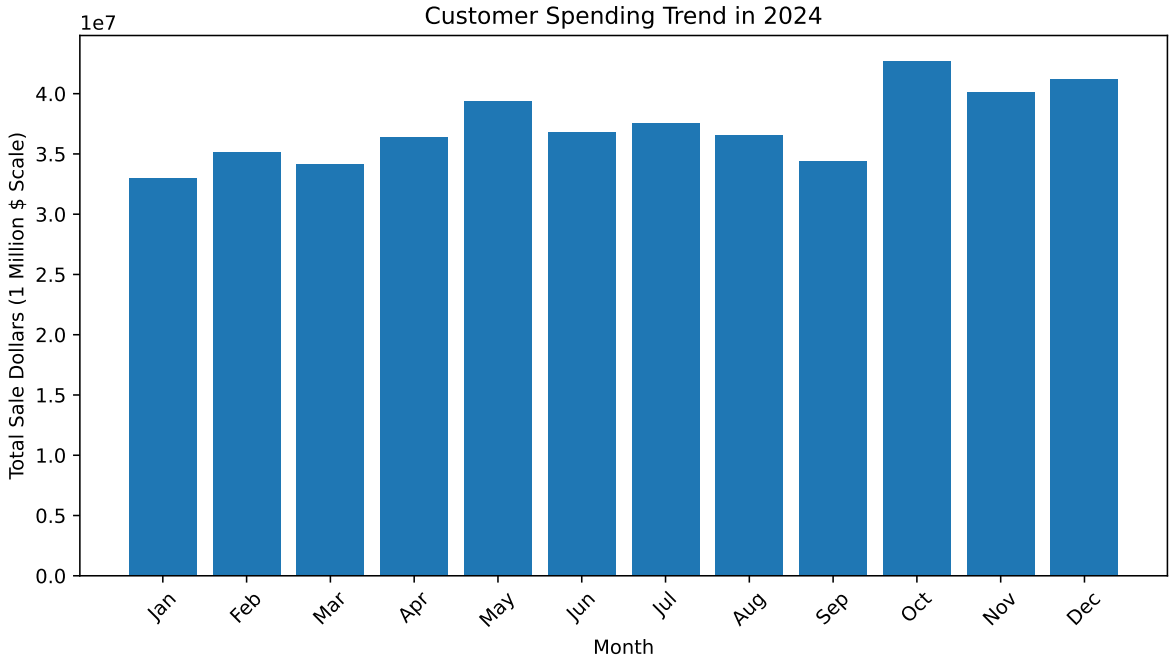
```
    grouped_data_booze['sum_sale_dollars'], 1)
line_of_best_fit = slope * grouped_data_booze['num_stores'] + intercept
plt.figure(figsize=(10,5))
plt.scatter(grouped_data_booze['num_stores'],
    grouped_data_booze['sum_sale_dollars'])
plt.plot(grouped_data_booze['num_stores'], line_of_best_fit,
    color='red', label='Line of Best Fit')
plt.ylabel("Total Sale Dollars (100,000 $ Scale)")
plt.xlabel("Number of Stores")
plt.title("Number of Stores Compared to Total Sales")
plt.show()


# Plot total sale dollars over time
grouped_data_graphing = grouped_data_booze.reset_index()
plot_data = (grouped_data_graphing.groupby('date')
             ['sum_sale_dollars'].sum().reset_index())
plt.figure(figsize=(10,5))
plt.bar(plot_data['date'], plot_data['sum_sale_dollars'])
plt.ylabel("Total Sale Dollars (1 Million $ Scale)")
plt.xlabel("Date (Month)")
plt.title("Customer Spending Trend Over Time")
plt.xticks(rotation=45)
plt.show()

# Plot total sale dollars over time (Seperate Years)
plot_data['date'] = pd.to_datetime(plot_data['date'], errors='coerce')
plot_data['year'] = plot_data['date'].dt.year
plot_data['month'] = plot_data['date'].dt.strftime('%b')

for yr in sorted(plot_data['year'].unique()):
    yearly_data = plot_data[plot_data['year'] == yr]
    plt.figure(figsize=(10,5))
    plt.bar(yearly_data['month'], yearly_data['sum_sale_dollars'])
    plt.ylabel("Total Sale Dollars (1 Million $ Scale)")
    plt.xlabel("Month")
    plt.title(f"Customer Spending Trend in {yr}")
    plt.xticks(rotation=45)
    plt.show()
```

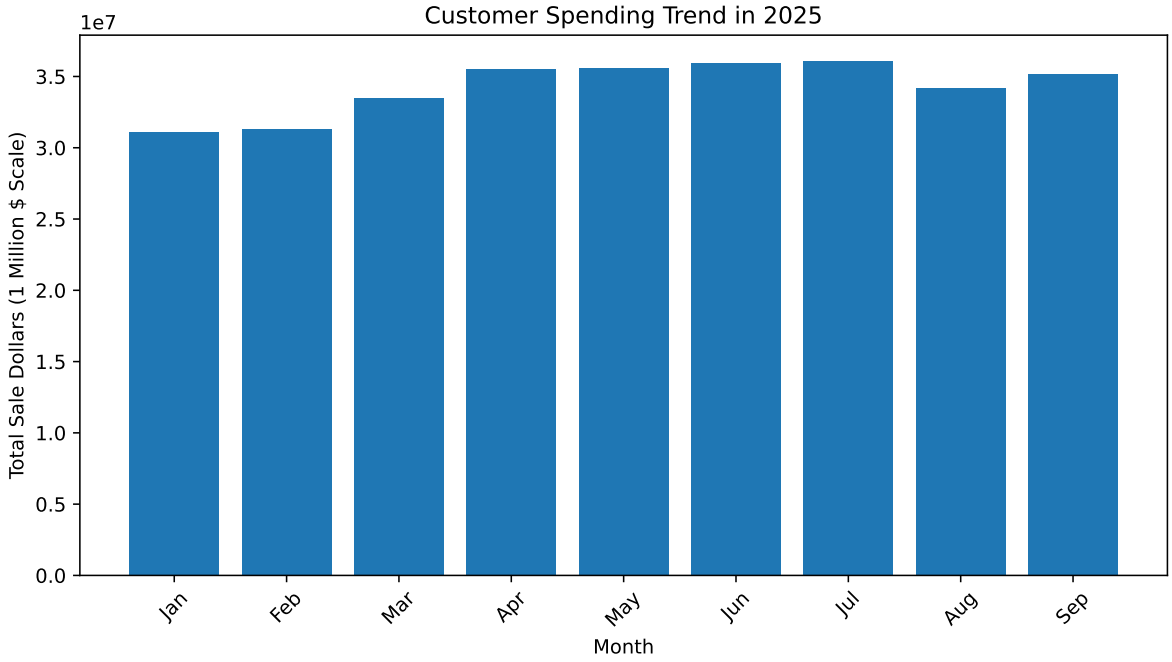

Number of Stores Compared to Total Sales
1e6
Total Sale Dollars (100,000 $ Scale)
Number of Stores
0
1
2
3
4
5
6
0
100
200
300
400
500


Customer Spending Trend Over Time
1e7
Total Sale Dollars (1 Million $ Scale)
Date (Month)
0
1
2
3
4


Customer Spending Trend in 2022
1e7
Total Sale Dollars (1 Million $ Scale)
0
1
2
3
4
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month


Customer Spending Trend in 2023
1e7
Total Sale Dollars (1 Million $ Scale)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month





With this data, we will be able to fit various linear models that predict total alcohol sales from different combinations of chain characteristics. Once we have fitted these models, we can compare them using cross-validation and select the ones that perform the best. This method

will ensure that we provide the strongest, most accurate model(s) that will best help Booze 'R' Us predict its sales and expand into other states.

Conclusions and Deliverables

After our model creation and analysis, Booze 'R' Us will be able to accurately predict sales data. Additionally, Booze 'R' Us will have an improved understanding of the characteristics that affect chain sales so that it can formulate a reliable and comprehensive plan to expand its footprint. We appreciate the opportunity to present our proposal and look forward to working with Booze 'R' Us.