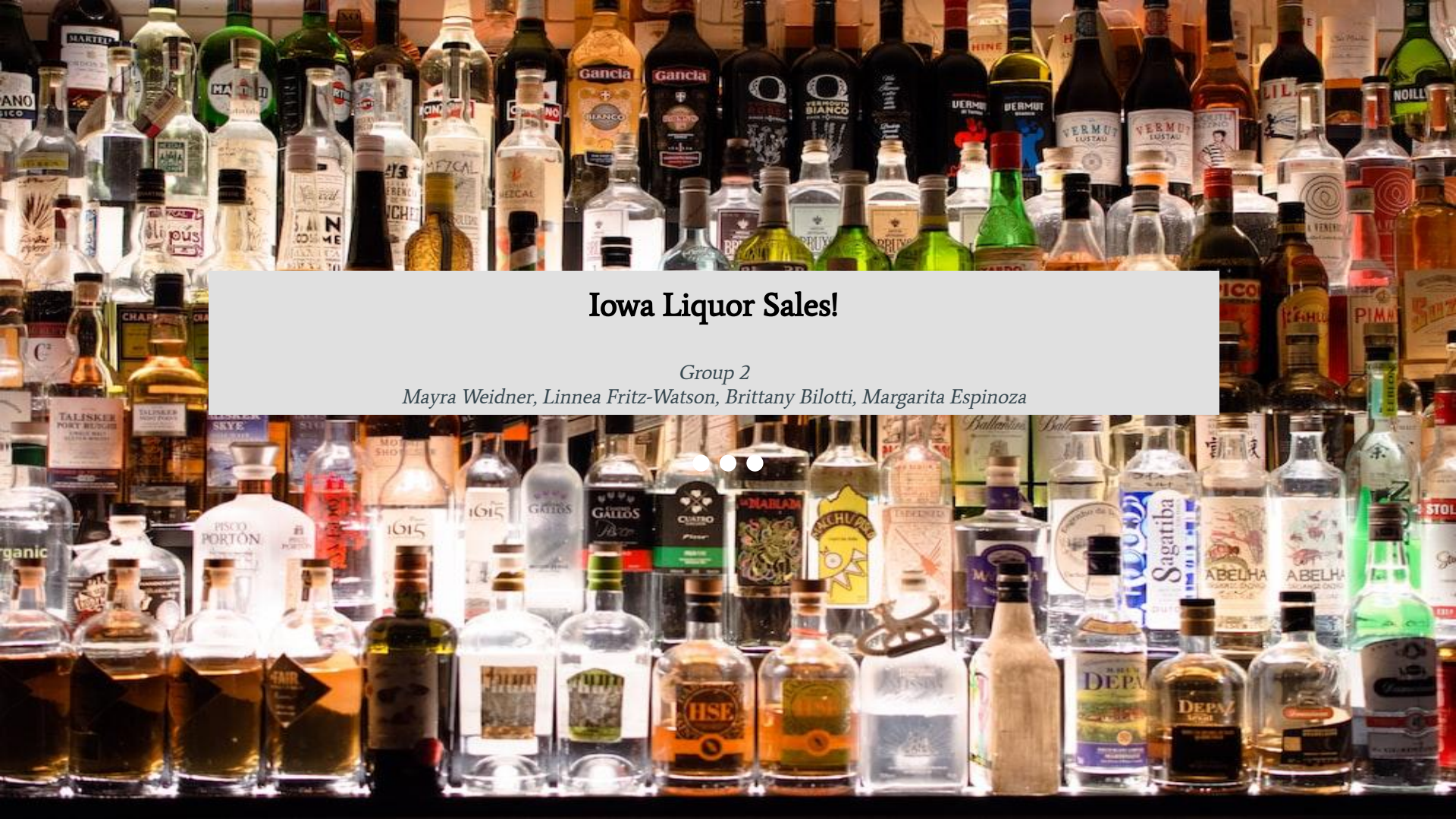


Iowa Liquor Sales!

Group 2
Mayra Weidner, Linnea Fritz-Watson, Brittany Bilotti, Margarita Espinoza

Description

This dataset contains the spirits purchase information of Iowa *Class “E” liquor licensees from January 1, 2012 to current. The following questions we seek to answer focus on trends and anomalies surrounding the sales of liquor in the state of Iowa.

Questions:

1. What times of the year have liquor sales been the highest?
2. Are there hotspots in the state where liquor sales have been higher than the average?
3. Can we see if there are preferred liquor types for regions or cities?
    a. Based on the above three questions, can we tailor advertisements to increase sales of a certain liquor?
4. Is there a surprising/unexpected time of year when liquor sales have gone up?
5. Are there purchase trends during holidays and college football season?
6. What have the average liquor prices in cities during holidays been?
    a. With this information we can predict what the highest price a customer is willing to pay before sales start to drop off

** Class “E” liquor license, for grocery stores, liquor stores, convenience stores, etc., allows commercial establishments to sell liquor for off-premises consumption in original unopened containers.*

Prior Work

Iowa Liquor Sales Explorer:

- Explorer presents liquor purchases from grocery stores, liquor stores, and convenience stores by product and date of purchase which users can filter
- https://data.iowa.gov/stories/s/Iowa-Liquor-Sales-Explorer/mke2-7r5k

Data on Liquor Sales Analysis in Iowa:

- Data mining results suggest that Des Moines, Polk would be the ideal location for a new store that would be most profitable if inventory bottle sizes ranged from 450-1750 ml
- https://nycdatascience.com/blog/student-works/data-on-liquor-sales-analysis-in-iowa/

Iowa Retail Liquor Sales Data Analysis Using Linear Regression:

- The linear regression model predicts 2020 liquor sales as $2,626,510 vs. actual sales of $2,588,159 (variance score of 98%)
- https://jonando-baskara.medium.com/iowa-retail-liquor-sales-data-analysis-using-linear-regression-model-bbc22fc150f9

Visualization and Analysis of Liquor Sales in Iowa:

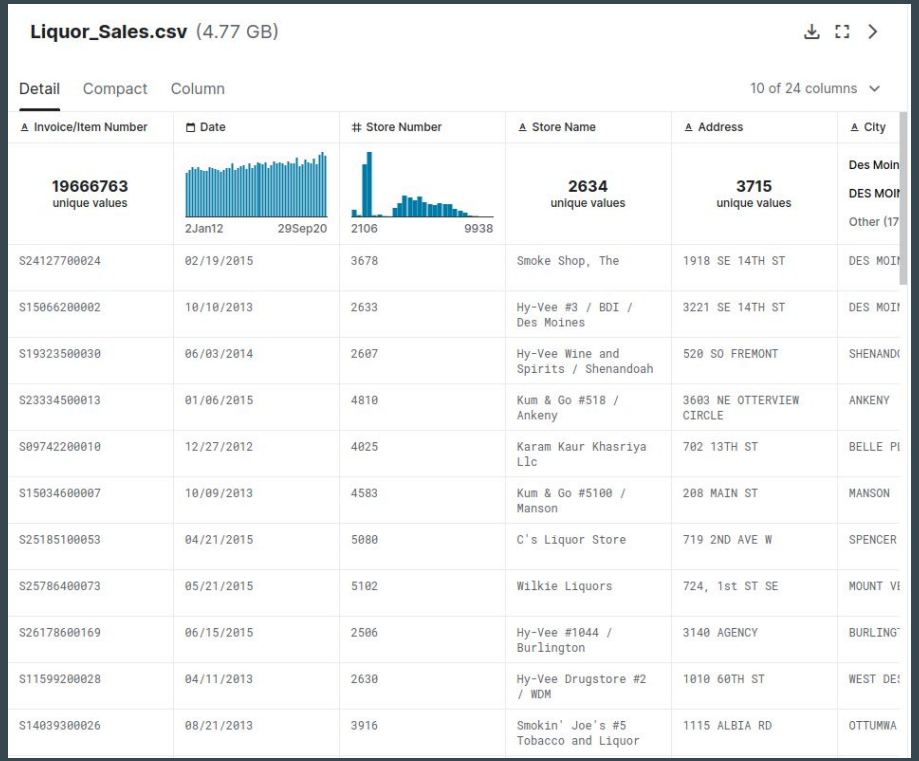
- Heatmap, chart, and plot results indicate areas with high liquor consumption and how seasonality affects sales
- https://richardcmsc320final.github.io/

Dataset Information

- **Dataset**:
  - Liqour_Sales.csv dataset (original): contains 19,666,763 unique values and 24 attributes, putting it at 4.77 GB file size.
  - Due to the large size of the dataset, we had to create a sub-dataset to work with by randomly pulling a little over 1 million rows. The result is nearly 4 consecutive years worth of sales data.

- **URL and data supplier:**
  - URL: https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy
  - Data provided by: Iowa Department of Commerce, Alcoholic Beverages Division

- **Download:**
  - The sub-dataset is downloaded onto each of the team members computers and their respective JupyterLabs.



Liquor_Sales.csv (4.77 GB)

Detail | Compact | Column | 10 of 24 columns

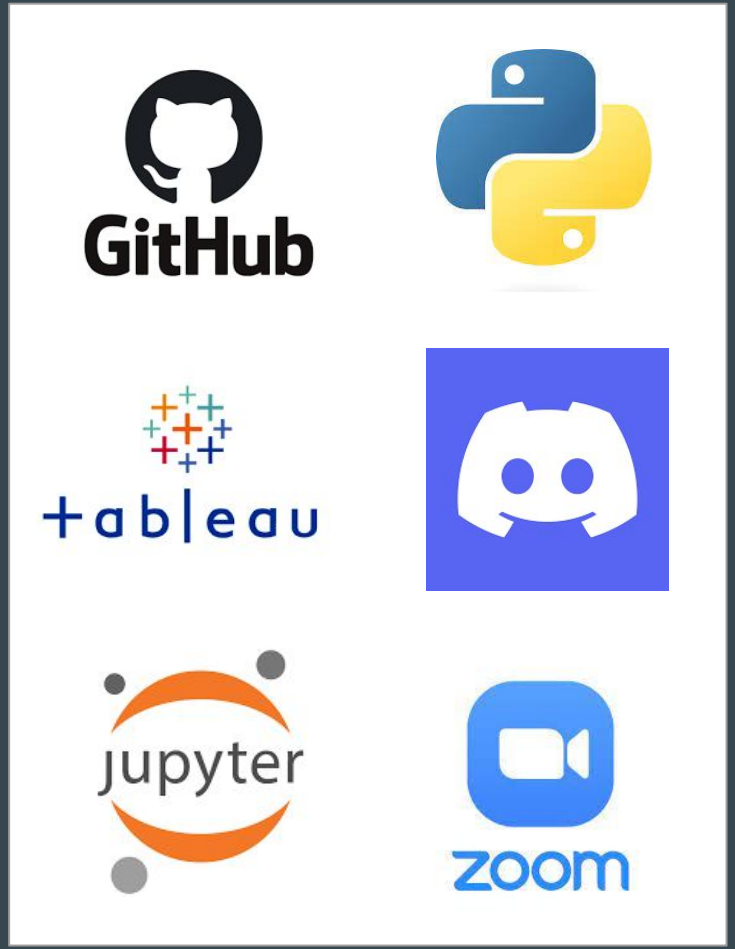
Invoice/Item Number	Date	Store Number	Store Name	Address	City
19666763 unique values	2Jan12 – 29Sep20	2106 – 9938	2634 unique values	3715 unique values	Des Moin… DES MOI… Other (17
S24127700024	02/19/2015	3678	Smoke Shop, The	1918 SE 14TH ST	DES MOI…
S15066200002	10/10/2013	2633	Hy-Vee #3 / BDI / Des Moines	3221 SE 14TH ST	DES MOI…
S19323500030	06/03/2014	2607	Hy-Vee Wine and Spirits / Shenandoah	520 SO FREMONT	SHENANDO…
S23334500013	01/06/2015	4810	Kum & Go #518 / Ankeny	3603 NE OTTERVIEW CIRCLE	ANKENY
S09742200010	12/27/2012	4025	Karam Kaur Khasriya Llc	702 13TH ST	BELLE PL…
S15034600007	10/09/2013	4583	Kum & Go #5100 / Manson	208 MAIN ST	MANSON
S25185100053	04/21/2015	5080	C's Liquor Store	719 2ND AVE W	SPENCER
S25786400073	05/21/2015	5102	Wilkie Liquors	724, 1st ST SE	MOUNT VI…
S26178600169	06/15/2015	2506	Hy-Vee #1044 / Burlington	3140 AGENCY	BURLING…
S11599200028	04/11/2013	2630	Hy-Vee Drugstore #2 / WDM	1010 60TH ST	WEST DES…
S14039300026	08/21/2013	3916	Smokin' Joe's #5 Tobacco and Liquor	1115 ALBIA RD	OTTUMWA

Proposed Work

- **Data cleaning:**
  - Go through row values that are null and omit those entities
- **Data integration:**
  - Not needed as all data is unified into one source
- **Data Reduction:**
  - Reducing attributes: if grouping by city then columns store number, store name, address of store, zip code, store location, and county number may be redundant
  - Dimensionality reduction: might be needed if grouping together by one attribute
- **Data transformation:**
  - Attributes listed under data cleaning could be transformed into ranges:
    - Bottle volume: from ml into small, large, medium
    - State bottle retail: $1-$5, $5-$15, $15-$30, etc
    - Sale (Dollars): range similar to State bottle retail
    - Item Description: may generalize liquor from “Laws Rye Whiskey” to “Whiskey”
- **Data Processing:**
  - Choose attributes for correlation testing
  - Find min support and confidence for attribute correlation
  - Determine best algorithms like Apriori, x^2 correlation, FP growth to use for finding correlation between attributes

Tools

- **GitHub** - Project Repository
- **Microsoft**
  - **Excel** - Spreadsheet Program
  - **Access** - Database Management System
- **Tableau/Power BI** - Visualization Tools
- **Python** - Programming Language
  - Pandas
  - Numpy
- **JupyterLab** - Web Based IDE
- **Zoom** - Video Conferencing Tool
- **Discord** - Team Communication and Project Management
- **Google Docs/Slides** - Documentation Management



Evaluation

We plan on using the following techniques to evaluate the results:

- Pattern tracking
- Association
- Clustering
- Regression
- Prediction
- Sequential patterns
- Statistical techniques
- Visualization

We will further assess the quality of any correlations, patterns and anomalies by using the following evaluation methods:

- Accuracy
- Classification Accuracy
- Clustering Accuracy
    - Assessment & Tendencies
- Visual Inspection
- Support
- Confidence
- Lift
- Prediction Accuracy