

The Sales Situation of Liquors in Different Regions

Lei Teng
lete3485@colorado.edu

Mingxuan Zhang
mizh1382@colorado.edu

Yuxiang Wang
yuwa4103@colorado.edu

ABSTRACT

Nowadays, alcoholic beverage becomes one of the most important things in our life. People use liquors in many different ways. Such as cooking, medicine, etc. But for most of us, alcoholic beverage is used for drinking and then release our pressure. And what we will do in this project is to analysis the sales situation of liquors in different regions.

KEYWORDS

Liquors, alcoholic beverage

1 INTRODUCTION

More specifically, our project works for analyzing the name, date, kind, price, quantity, and location of sales of individual containers or packages of containers of alcoholic beverages from 2015 to get the sales situation and drinking behaviors of people in different regions.

2 Problem Statement

2.1 Sales Situation

2.1.1 Description

For the sales situation, we can analyze the data to gain some correlation results. Such as the total amount of alcoholic beverage sold and consumed in months, years and regions. The alcoholic beverage sells best in different regions. Which region has the most liquors' store. By analyzing the correlation of price and locations, we can get the region which has the highest price. By analyzing the correlation of date (months, years), and sales volume, we can get the tendency of months and years in different regions. Analyzing the correlation of name (or kind) and price to get the alcoholic beverage which gains the maximum profit. Analyzing the correlation of date and price to get the distribution in different regions, etc.

2.1.2 Specific Questions

- Which kind of alcoholic beverage gains the maximum profit?
- The tendency of sales volume of different alcoholic beverage.
- How much alcoholic beverage is consumed?
- The distribution of prices.

2.2 Drinking Behavior

2.2.1 Description

For the drinking behavior, according to the results of sales situation, such as the total sales amount of alcoholic beverage, can tell us people in which region drink the most alcoholic beverage or in which season people drink more. Also, from the dataset, we can get different sales rate of different alcoholic beverage in a specific region, then we will know which alcoholic beverage is the most popular in that region.

2.2.2 Specific Questions

- Which alcoholic beverage is the most popular in different regions?
- People in which region drink more alcoholic beverage?
- Figure out some other drinking behaviors for people in different regions.

3 Literature survey/Prior work

3.1 Prior work describe

3.1.1 Previous work (1)

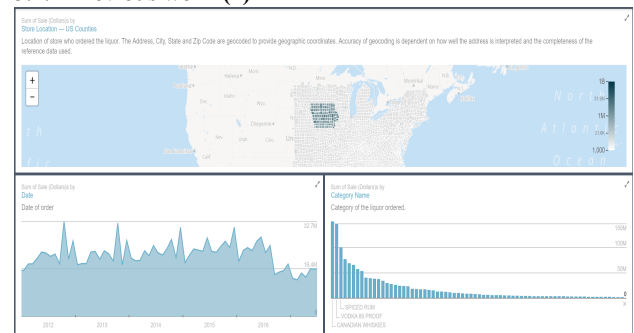


Figure 1 ^[1]

Figure 1: This work is about Iowa Liquor Sales in Dollars. It contains a map and two plots, which give us the sales in dollars respect to locations and times and categories.

3.1.2 Previous work (2)

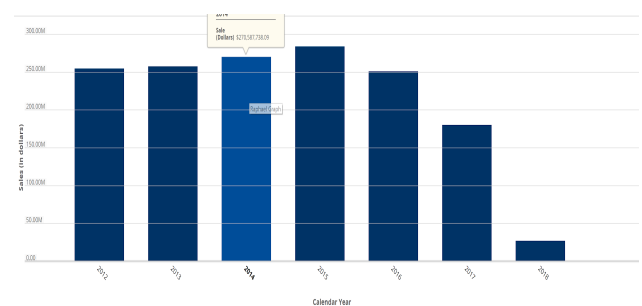


Figure 2 ^[2]

Figure 2: This work gives us the histogram about Iowa Liquor Sales in Dollars by Year.

3.1.3 Previous work (3)

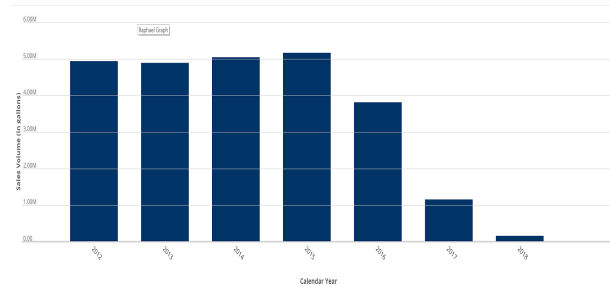


Figure 3 ^[3]

Figure 3: This work gives us the histogram about Iowa Liquor Sales in Gallons by Year.

3.1.4 Previous Work (4)

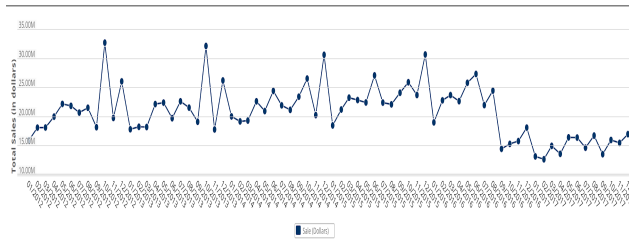


Figure 4 ^[4]

Figure 4: This work gives us the histogram about Iowa Liquor Sales in Gallons by Year.

3.1.5 Previous Work (5)

	Date	County	Sale (Dollars)	Volume Sold (Gallons)
1	2018	Adair	\$22,292.57	203.18
2	2018	ADAIR	\$25,368.84	156.89
3	2018	ADAMS	\$4,709.95	86.47
4	2018	ALLAMAKEE	\$76,742.45	575.00
5	2018	APPANOOSE	\$84,414.18	469.24
6	2018	AUDUBON	\$14,764.40	349.15
7	2018	BENTON	\$95,766.11	735.35
8	2018	Black Hawk	\$43,930.93	466.45
9	2018	BLACK HAWK	\$1,599,401.10	9,119.71
10	2018	Boone	\$18,989.58	226.71
11	2018	BOONE	\$213,670.17	1,228.03
12	2018	BREMER	\$183,483.74	1,389.14
13	2018	Buchanan	\$26,826.39	304.41
14	2018	BUCHANAN	\$112,582.86	953.62
15	2018	BUENA VIST	\$176,043.82	1,425.38
16	2018	Butler	\$2,946.21	40.47
17	2018	BUTLER	\$36,219.57	262.02
18	2018	CALHOUN	\$44,428.77	293.93
19	2018	CARROLL	\$179,503.17	1,165.40

Figure 5 ^[5]

Figure 5: This work gives us the table about Iowa Liquor Sales by Year and County. In this table, we can get the sales in gallons and dollars in different time and county.

4 Proposed Work

4.1 Data Cleaning

For missing data, our approach is to Fill in it automatically with attribute mean.

4.2 Data Integration

Since we only have one database, we don't have to do the data integration.

4.3 Data Reduction

There are 24 attributes in the dataset. Since we don't need all 24 attributes, we have to do Dimensionality reduction. We will remove irrelevant attributes like Invoice/Item Number. We also have to remove redundant attributes, for example, Volume Sold (Liters) and Volume Sold (Gallons) are redundant attributes, we will use Volume Sold (Liters) instead of Volume Sold (Gallons).

4.4 Graphic Analysis

We will use histogram to analysis Iowa Liquor Sales in Dollars by Year, which is similar to previous (2), and we also use histogram to analysis Iowa Liquor Sales in Gallons by Year, which is similar to previous work (3).

We also can get the distribution for different liquors by using histogram.

And we will plot bar charts for total liquors sales respect to different cities.

Finally, plot the histogram for the distribution of prices.

4.5 Pattern Finding

We will get what is the best liquor in some major cities, we will also get the support and confidence.

For pattern finding, we focus on the following two questions:

- (1) What is the most popular liquors in the whole data set.
- (2) We will use different min supports to find frequent 1-item set.

And we believe as the course is going on, we will have more skills/goals for our project. These works are our basic goals for now.

5 Data Set

URL: <https://www.kaggle.com/residentmario/iowa-liquor-sales/data>^[6]

This dataset contains information on the name, kind, price, quantity, and location of sale of sales of individual containers or packages of containers of alcoholic beverages.

Our datasets have around 12 million objects and 24 different attributes^[7]:

Invoice/Item Number: Concatenated invoice and line number associated with the liquor order.

Date: Date of order.

Store Number: Unique number assigned to the store who ordered the liquor.

Store Name: Name of store who ordered the liquor.

Address: Address of store who ordered the liquor.

City: City where the store who ordered the liquor is located.

Zip Code: Zip code where the store who ordered the liquor is located.

Store Location: Location of store who ordered the liquor.

County Number: Iowa county number for the county where store who ordered the liquor is located.

County: County where the store who ordered the liquor is located.

Category: Category code associated with the liquor ordered.

Category Name: Category of the liquor ordered.

Vendor Number: The vendor number of the company for the brand of liquor ordered.

Vendor Name: The vendor name of the company for the brand of liquor ordered.

Item Number: Item number for the individual liquor product ordered.

Item Description: Description of the individual liquor product ordered.

Pack: The number of bottles in a case for the liquor ordered.

Bottle Volume (ml): Volume of each liquor bottle ordered in milliliters.

State Bottle Cost: The amount that Alcoholic Beverages Division paid for each bottle of liquor ordered.

State Bottle Retail: The amount the store paid for each bottle of liquor ordered

Bottles Sold: The number of bottles of liquor ordered by the store.

Sale (Dollars): Total cost of liquor order (number of bottles multiplied by the state bottle retail).

Volume Sold (Liters): Total volume of liquor ordered in liters.

Volume Sold (Gallons): Total volume of liquor ordered in gallons.

6 Evaluation Methods

6.1 Graphic evaluation

Since we use histogram to analysis Iowa Liquor Sales in Dollars by Year, which is similar to previous work (2), and we also use histogram to analysis Iowa Liquor Sales in Gallons by Year, which is similar to previous work (3). We can compare the graphic to previous work (2) and (3) to check if we are right.

6.2 Pattern Finding evaluation

Compare to the previous work (1) to check if our result is correct. Do some simple online surveys to see if our conclusion for “the best liquors in some major cities” is correct or mostly correct.

7 Tools

7.1 Microsoft Excel

7.2 Pandas

7.3 Numpy

7.4 Python

7.5 Matplotlib

7.6 Delimit

7.7 Jupyter Notebook

7.8 Overleaf

8 Milestones

8.1 Milestone 1

Do data cleaning work

Due date: 3/17

8.2 Milestone 2

Finish data cleaning work and start to do data reduction work

Due date: 3/31

8.3 Milestone 3

Finish data reduction work start to do the graphic analysis

Due date: 4/14

8.4 Milestone 4

Finish the code part of our project

Finish the project progress report

Due date: 4/21

8.4 Milestone 5

Finish the 1st draft of the final report;

prepare for the final presentation

Due date: 4/28

8.6 Milestone 6

Finish the final version of the final report

Due date: 4/30

9 Summary of peer review session

For the presentation last week, we did lack of prior work. We only got a little bit of description and the basic data set.

What’s more, we performed not very well when presentation, and we think the critical problem was that we did not have enough trail presentations and we should practice more next time.

REFERENCES

- [1] Iowa Liquor Sales in Dollars
<https://data.iowa.gov/Economy/Iowa-Liquor-Sales-in-Dollars/8epw->

CSCI 4502 Final Project

u33y

- [2] Iowa Liquor Sales in Dollars by Year
<https://data.iowa.gov/Economy/Iowa-Liquor-Sales-in-Dollars-by-Year/www-7at4>
- [3] Iowa Liquor Sales in Gallons by Year
<https://data.iowa.gov/Economy/Iowa-Liquor-Sales-in-Gallons-by-Year/7uuv-irpi>
- [4] Total Liquor Sales in Iowa by Month
<https://data.iowa.gov/Economy/Total-Liquor-Sales-in-Iowa-by-Month/xiyh-fbv>
- [5] Iowa Liquor Sales by Year and County
<https://data.iowa.gov/Economy/Iowa-Liquor-Sales-by-Year-and-County/ahiv-u4uz>
- [6] Kaggle data set of Iowa Liquor sales
<https://www.kaggle.com/residentmario/iowa-liquor-sales/data>
- [7] Iowa Liquor Sales
<https://data.iowa.gov/Economy/Iowa-Liquor-Sales/m3tr-qhgy>