

liquid store sales analysis

Project Information and data cleaning

- Data Source
- Data Profile
- Data cleaning steps
- Analysis Questions

Data Source

This data was originally published by the State of Iowa here and has been republished as-is on Kaggle. The author of this data was a data analyst and developer advocate on the Datasets team at Kaggle. Additional he has a good rating on Kaggle. That's why this dataset is **trustworthy**.

Data collection

The Iowa Department of Commerce requires that every store that sells alcohol in bottled form for off-the-premises consumption must hold a class "E" liquor license (an arrangement typical of most of the state alcohol regulatory bodies). All alcoholic sales made by stores registered thusly with the Iowa Department of Commerce are logged in the Commerce department system, which is in turn published as open data by the State of Iowa.

Data content

This data contains 12591077 rows and has 24 columns. 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.

Why this data?

Manly because it fulfills all the criteria given by CareerFoundry.

limitations/ relevancy

This dataset is the only one I'm using in this analysis. This means it is crucial to this analysis.

dropped columns

Store Number, County Number, County Number, Item Number, Vendor Number, Category



Data Profile

<u>Column name</u>	<u>Data Terminology</u>			<u>Description</u>
Invoice/Item Number	Structured data	Quantitative data	Discrete data	Unique ID for every Invoice
Date	Structured data	Quantitative data	Discrete data	Date when the invoice was created
Store Name	Unstructured data	Qualitative data	Nominal data	name of the store the invoice was made in
Address	Unstructured data	Qualitative data	Nominal data	address of the store the invoice was made in
City	Structured data	Qualitative data	Nominal data	city where the invoice was created in
Zip Code	Structured data	Qualitative data	Nominal data	Zip Code of the city the invoice was created in
Store Location	Unstructured data	Qualitative data	Nominal data	location of the store the invoice was created in
County	Structured data	Qualitative data	Nominal data	Country of the store the invoice was made in
Category Name	Unstructured data	Qualitative data	Nominal data	the name of the bottles category

Data Profile

<u>Column name</u>	<u>Data Terminology</u>		<u>Description</u>	
Vendor Name	Unstructured data	Qualitative data	Nominal data	The name of the vendor
Item Description	Unstructured data	Qualitative data	Nominal data	A short description of the article
Pack	Structured data	Quantitative data	Discrete data	amount of bottles in pack
Bottle Volume (ml)	Structured data	Quantitative data	Continuous data	volume of the bottle
State Bottle Cost	Structured data	Quantitative data	Discrete data	The price the store bought the bottle for
State Bottle Retail	Structured data	Quantitative data	Discrete data	The price the store sold the bottle for
Bottles Sold	Structured data	Quantitative data	Discrete data	number of sold bottles
Sale (Dollars)	Structured data	Quantitative data	Discrete data	sum of price of all sold bottles
Volume Sold (Liters)	Structured data	Quantitative data	Continuous data	sum of the volume of liquid that was sold in liters
Volume Sold (Gallons)	Structured data	Quantitative data	Continuous data	sum of the volume of liquid that was sold in Gallons

Data cleaning steps

Mixed data types

I found 13 columns containing mixed data types. The reason for them, are missing values. Since I deleted all columns with null values and changed the dtype for 'Zip Code' to string I had no mixed dtypes left.

null values

I found 13 columns containing null values

I deleted all of them. After this the df had 12495974 rows left.

duplicates

This dataframe does not contain any duplicates.

null value counts

Address	2376
City	2375
Zip Code	2420
Store Location	2375
County Number	79178
County	79178
Category	8020

Category Name	16086
Vendor Number	3
Vendor Name	1
State Bottle Cost	10
State Bottle Retail	10
Sale (Dollars)	10



Analysis Questions

- How much alcohol is sold and consumed in the United States?
- What kind of alcohol is sold and consumed in the United States?
- What are the most popular brands and labels?
- What is the distribution of prices paid in-store?
- Are there any correlations between the variables?
- Does the volume has any impact on the price?

