# Context

This dataset contains the spirits purchase information of Iowa Class “E” liquor licenses by product and date of purchase from January 2021 to January 2022. This project is to find insights about liquor sales from multiple geolocation aspects, i.e., state, city, and county. I will then see if there were any noticeable trends for the year and see if there are any detectable correlations to showcase.

# Data Source

This data set was sourced from [Kaggle](https://www.kaggle.com/datasets/gabrielramos87/iowa-sales-liquor-jan-2021jan-2022), which is also published on [Iowa Data’s website](https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-qhgy). The data is provided by the Iowa Department of Commerce, Alcoholic Beverages Division, so it is a reputable source. This dataset is updated on the first day of each month, which is reliable. The only ethically questionable biases I have is the collection and measurement of the Quantitative fields of the data set since there is no documentation on how they collect and measure the fields.

# Limitations

The only apparent limitation I see is that there are some Null values (blank cells) mainly in the store\_location column, 12% of it in fact. I will have to decide whenever I use geolocation if I want to remove all null values or not. There are other missing values in other columns, but it is less than 5% so I will ignore those columns. There are no duplicated entries in the data set. Some columns were renamed for legibility.

# Data Profile

The dataset has 2,805,307 rows and 24 columns.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Variable* | *Description* | *Quant/Qual* | *Structured/*  *Unstructured* | *Time-variant/Invariant* | *Nominal/Ordinal/*  *Discrete/Continuous* |
| invoice\_and\_item\_number | Unique identifier for the individual liquor order that includes the invoice and item number. | Qualitative | Structured | Invariant | Nominal |
| date | Date of products order | Quantitative | Structured | Variant | Discrete |
| store\_number | Unique number assigned to store | Qualitative | Structured | Invariant | Ordinal |
| store\_name | Name of store who ordered the liquor | Qualitative | Structured | Invariant | Nominal |
| address | Address of store who ordered the liquor | Qualitative | Structured | Invariant | Nominal |
| city | City where the store who ordered the liquor | Qualitative | Structured | Invariant | Nominal |
| zip\_code | Zip code where the store who ordered the liquor | Qualitative | Structured | Invariant | Ordinal |
| store\_location | Geolocation of store who ordered the liquor | Qualitative | Structured | Invariant | Ordinal |
| county\_number | Iowa county number of the liquor order | Qualitative | Structured | Invariant | Ordinal |
| county | County where the store who ordered the liquor | Qualitative | Structured | Invariant | Nominal |
| category | Category code associated with the liquor ordered | Qualitative | Structured | Invariant | Ordinal |
| category\_name | Category of liquor ordered | Qualitative | Structured | Invariant | Nominal |
| vendor\_number | The vendor number of the company for the brand of liquor ordered | Qualitative | Structured | Invariant | Ordinal |
| vendor\_name | The vendor name of the company for the brand of liquor ordered | Qualitative | Structured | Invariant | Nominal |
| item\_number | Item number for the individual liquor product ordered. | Qualitative | Structured | Invariant | Ordinal |
| item\_description | Description of the individual liquor product ordered. | Qualitative | Structured | Invariant | Nominal |
| pack | The number of bottles in a case for the liquor ordered | Quantitative | Structured | Invariant | Discrete |
| bottle\_volume\_ml | Volume of each liquor bottle ordered in milliliters. | Quantitative | Structured | Invariant | Discrete |
| state\_bottle\_cost | The amount that Alcoholic Beverages Division paid for each bottle of liquor ordered | Quantitative | Structured | Invariant | Continuous |
| state\_bottle\_retail | The amount the store paid for each bottle of liquor ordered | Quantitative | Structured | Invariant | Continuous |
| bottles\_sold | The number of bottles of liquor ordered by the store | Quantitative | Structured | Invariant | Discrete |
| sale\_dollars | Total cost of liquor order (number of bottles multiplied by the state bottle retail) | Quantitative | Structured | Invariant | Continuous |
| volume\_sold\_liters | Total volume of liquor ordered in liters. (i.e. (Bottle Volume (ml) x Bottles Sold)/1,000) | Quantitative | Structured | Invariant | Continuous |
| volume\_sold\_gallons | Total volume of liquor ordered in gallons. (i.e. (Bottle Volume (ml) x Bottles Sold)/3785.411784) | Quantitative | Structured | Invariant | Continuous |

# Questions

1. Which liquor companies are popular in the state, county, and city level on average? Any changes during the 4 seasons? The top 3 liquor products in each season?
2. When is revenue the lowest during the year? Are there any correlations to find the trends in sales over the year (I.e., Football games, promotions, etc....)?
3. What is the average number of bottles of liquor ordered and the average cost from all stores? Any recommendations to market to stores?
4. What will the forecast be for January and February bottles ordered in 2023 to the top 20 stores?