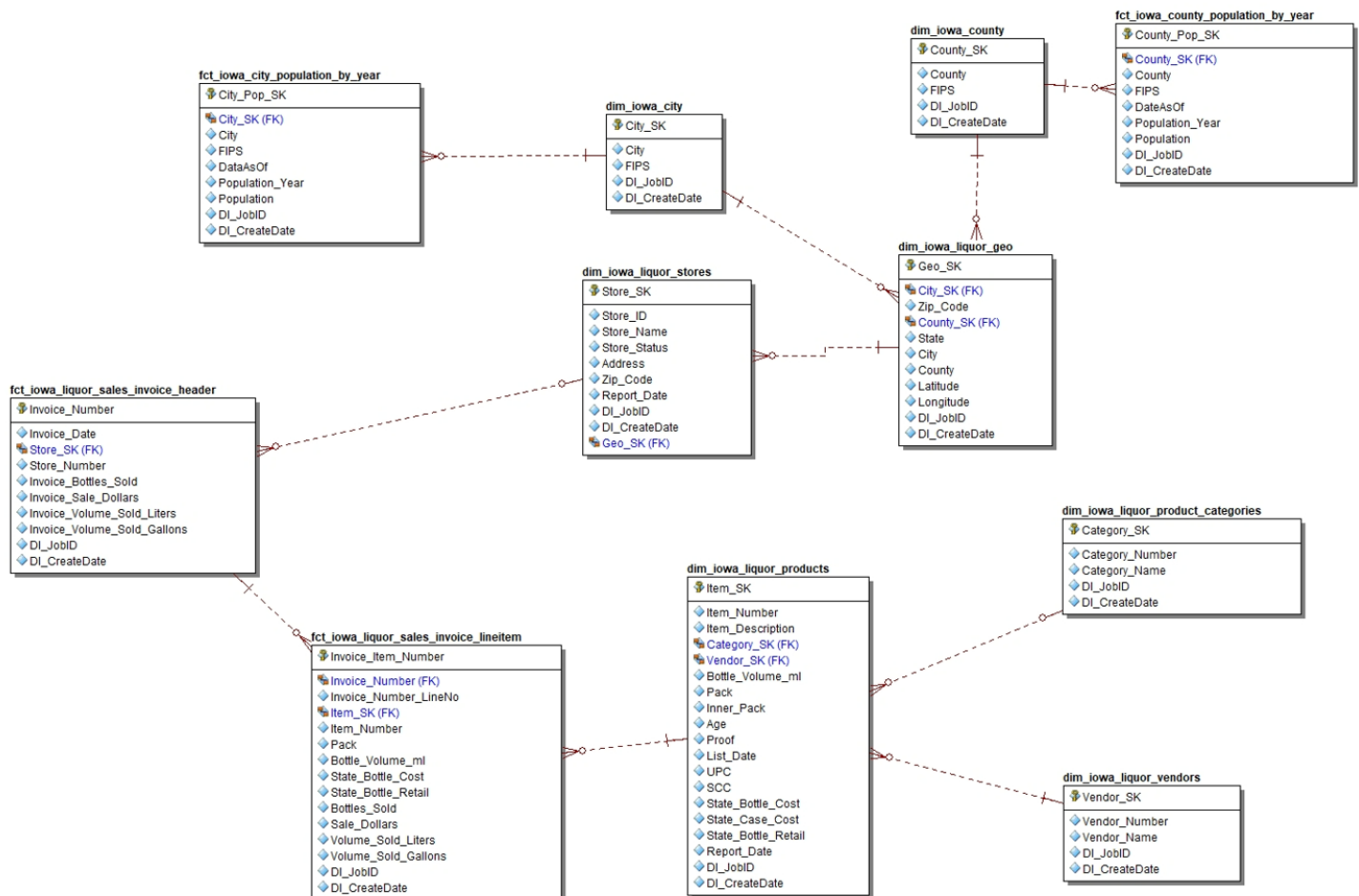# IOWA Liquor Sales – Part 1 & 2

Data Source:

The IOWA liquor datasets were provided which gives information on the sales of liquor in the state of Iowa. The datasets include information such as the date of the sale, the type of liquor sold, the quantity sold, and the price. Further we generated a dimensional model to load the data and populate the database. The Data also includes year wise population of counties and cities in Iowa State.

Dim Model Outline:

As part of this project, we prepared an outline for the dimension model. Analyzed the data and created dimensional tables namely **dim_iowa_liquor_address, dim_iowa_city, dim_iowa_county, dim_iowa_liquor_stores, dim_iowa_liquor_products, dim_iowa_liquor_categories, dim_iowa_liquor_vendors, fct_iowa_city_population_by_year, fct_iowa_county_population_by_year, fct_iowa_liquor_sales_invoice_header, fct_iowa_liquor_sales_invoice_lineitem**. The model includes dimensions such as time, location, product, and customer. The dimensions are designed to capture key information about the sales of liquor in Iowa, such as the location of the sale, the type of liquor sold, and the customer who made the purchase.

Dimensions that are connected to fct_iowa_liquor_sales_invoice_header:
- dim_iowa_liquor_stores
Dimensions that are connected to fct_iowa_liquor_sales_invoice_products:
- dim_iowa_liquor_product_categories

-dim_iowa_liqour_vendors

Data Cleansing and Data Staging:

- **Sales -**
  - Null values in Address, City, Category Name, County were replaced with NA
  - Null values in Zipcode, county number and category replaced with 99999 and -1 respectively
  - Store location is divided into longitude and latitude, also null values are replaced with
- **Products -** UPC is generally a **12 digit number** but in the dataset we found some with **11 digits** hence they were **replaced with -99999**, enhanced List date and rectified all errors by converting the date to "%m/%d/%Y" in this format. Initially we were missing 3% of date data after enhancing 100% data was available.
- **Stores -** Null values in Store_Address are replaced with NA. Further store address is splitted into two parts Longitude and Latitude. Store address 21 and store address 22 are renamed to Longitude and Latitude respectively.
- **City population by year -** Null values replaced NA, enhanced census date and rectified all errors by converting the date to "%m/%d/%Y"
- **County population by year -** Null values replaced NA, enhanced census date and rectified all errors by converting the date to "%m/%d/%Y"**.** No null values were present in this data. Primary Point is renamed to Longitude and Latitude respectively.
- **Sales by year and county -** Null values in county are replaced with NA
- **Invoice_Number_LineNo** Field in fct_iowa_liqour_sales_invoice_lineitem- This field was extracted after separating the last two digits of the Invoice Number given.
- **City and County Names**- Inconsistencies were removed in the City and County names present in the population data set.

  For example: The county 'Franklin' was differently named as 'Franklin' in one and 'Franklin County' in another. In here, we made the field values consistent by removing 'county' in the end. Same is done for cities.

Brief Explanation about the Dimensions:

- **dim_iowa_liquor_address -** This dimension likely contains information about the addresses of liquor stores in Iowa, such as street address, city, state, and zip code.
- **dim_iowa_city -** This dimension likely contains information about cities in Iowa, such as city name, county, population, and geographic location.

- **dim_iowa_county** - This dimension likely contains information about counties in Iowa, such as county name, population, and geographic location.
- **dim_iowa_liquor_stores** - This dimension likely contains information about liquor stores in Iowa, such as store name, address, owner, and license number.
- **dim_iowa_liquor_products** - This dimension likely contains information about liquor products sold in Iowa, such as product name, manufacturer, price, and volume.
- **dim_iowa_liquor_categories** - This dimension contains information about liquor product categories, such as whiskey, vodka, beer, and wine.
- **dim_iowa_liquor_vendors** - This dimension contains information about liquor vendors, such as vendor name, address, and contact information.
- **fct_iowa_city_population_by_year** - This dimension likely contains information about city populations in Iowa by year, such as population estimates or census data.
- **fct_iowa_county_population_by_year** - This dimension likely contains information about county populations in Iowa by year, such as population estimates or census data.

Brief Explanations of Field:

- **FIPS(County)**- FIPS codes are numbers which uniquely identify geographic areas. The number of digits in FIPS codes vary depending on the level of geography. State-level FIPS codes have two digits, county-level FIPS codes have five digits of which the first two are the FIPS code of the state to which the county belongs.
- **FIPS(City)**- FIPS codes are assigned alphabetically by geographic name for states, counties, core based statistical areas, places, county subdivisions, **consolidated cities** and all types of American Indian, Alaska Native, and Native Hawaiian (AIANNH) areas.
- **Report_Date**- Report Date refers to the date on which the sales transactions were reported to the Iowa Alcoholic Beverages Division (ABD).
- **Store Status**: Given information of store status in "I" and "A" which represents Active and Inactive respectively.
- **Pack**- Pack typically refers to the number of containers of a specific alcoholic beverage product that were sold as part of a single transaction.
- **Inner Pack**- "Inner pack" is a term that refers to the number of individual containers of a specific alcoholic beverage product that are packaged together within a larger package or case. In the Iowa liquor sales dataset, "inner pack" is typically listed as a numeric value representing the number of containers per case.
- **Age**- The age of alcohol generally refers to the length of time that a specific type of alcoholic beverage has been aged or matured before it is consumed. Aging is a process that can affect the flavor, aroma, and texture of the beverage, and it is often used to create complex and nuanced flavors in certain types of alcohol

- **List_Date**- "List date" is a field in the Iowa liquor sales dataset that indicates the date when a specific product was added to the list of available products for sale in Iowa. This information is important for tracking sales trends and understanding the availability of specific products in the market.
- **UPC**- In the Iowa liquor sales dataset, "UPC" stands for Universal Product Code, which is a standardized barcode used to identify and track individual products.
- **SCC**- In the Iowa liquor sales dataset, "SCC" stands for State Code Category, which is a code used to classify the type of product being sold. The SCC is a three-digit code assigned by the Iowa Alcoholic Beverages Division (ABD) based on the type of product