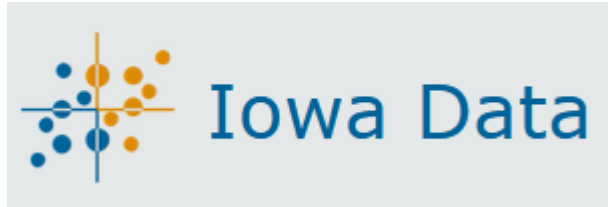


Iowa Liquor Sales



Assignment 1

- Perform Data Profiling (with Alteryx)
- Load Data into Staging Tables
 - Two Staging DBMS
 - Load one staging dbms with Alteryx, other with Talend
- Create Dimensional Model
- Outline data cleaning tasks
 - Identify what needs to be cleaned & explain how you are going to do it



Part 1: Input Files for Loading and Analysis

File	Description	As of
Iowa_Liquor_Sales.tsv	Sales 2012-2022 YTD	10/25/2022
Iowa_Liquor_Products.tsv	Products	10/25/2022
Iowa_Liquor_Stores.tsv	Stores	10/25/2022
Total_City_Population_by_Year.tsv	Iowa City Population by Year	
County_Population_in_Iowa_by_Year.tsv	Iowa Country Population by Year	
Iowa_Liquor_Sales_20221002_notCleaned.tsv	Sales 2012-2022 YTD, No Data Cleansing	10/3/2022
Iowa_Liquor_Products_20221002_notCleaned.tsv	Products, No Data Cleansing	10/3/2022
Iowa_Liquor_Stores_20221002_notCleaned.tsv	Stores, No Data Cleansing	10/3/2022
Iowa_Liquor_Sales_by_Year_and_County.tsv	Total Sales by County and Year	
Iowa_Liquor_Sales_by_Year.tsv	Total Sales by Year	
2012_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
2013_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
2014_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
2015_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
2016_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
2017_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
2018_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
2019_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
2020_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
2021_Iowa_Liquor_Sales.tsv	Sales for a Specific Year	
BigQuery Public Datasets - Iowa Liquor Sales.sql	Big Query SQL Query	

Alternate source for Iowa Liquor Sales data:

bigquery-public-[data](#).iowa_liquor_sales.sales

Part 1: Data source to perform data profiling

Data Source	Data Profile with Alteryx
Iowa_Liquor_Sales.tsv	Yes
Iowa_Liquor_Products.tsv	Yes
Iowa_Liquor_Stores.tsv	Yes
bigquery-public-data.iowa_liquor_sales.sales	Yes

- Data Profiling using Alteryx
- Deliverables: (upload the following)
 - Screen shots of completed profiling
 - Time each profiling job takes
 - Completed Alteryx jobs

Part 1: Load into separate staging tables

Data Source	Stage Table Name
Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
Iowa_Liquor_Products.tsv	stg_iowa_liquor_products
Iowa_Liquor_Stores.tsv	stg_iowa_liquor_stores
bigquery-public- data .iowa_liquor_sales.sales	stg_iowa_liquor_sales_bigquery
Iowa_Liquor_Sales_20221002_notCleaned.tsv	stg_iowa_liquor_sales_notcleaned
Iowa_Liquor_Products_20221002_notCleaned.tsv	stg_iowa_liquor_products_notcleaned
Iowa_Liquor_Stores_20221002_notCleaned.tsv	stg_iowa_liquor_stores_notcleaned
Total_City_Population_by_Year.tsv	stg_iowa_city_population_by_year
County_Population_in_Iowa_by_Year.tsv	stg_iowa_county_population_by_year
Iowa_Liquor_Sales_by_Year_and_County.tsv	stg_iowa_liquor_sales_by_year_and_county
Iowa_Liquor_Sales_by_Year.tsv	stg_iowa_liquor_sales_by_year

- Load staging tables with Talend
 - (see page for alternate loading approach)
- Staging tables options (you choose):
 - Azure SQL, SQL Server, MySQL or Azure MySQL
- Deliverables: (upload the following)
 - Screen shots of completed jobs
 - Time each profiling job takes
 - Completed Talend jobs
 - List of tables with rows counts

Part 1: Load into separate staging tables

Data Source	Stage Table Name
Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales_alteryx
Iowa_Liquor_Products.tsv	stg_iowa_liquor_products_alteryx
Iowa_Liquor_Stores.tsv	stg_iowa_liquor_stores_alteryx
bigquery-public-data.iowa_liquor_sales.sales	stg_iowa_liquor_sales_bigquery_alteryx
Total_City_Population_by_Year.tsv	stg_iowa_city_population_by_year_alteryx
County_Population_in_Iowa_by_Year.tsv	stg_iowa_county_population_by_year_alteryx

- Load staging tables with Alteryx
 - (see page alternate loading approach)
- Staging tables options (you choose):
 - Azure SQL, SQL Server, MySQL, Azure MySQL, PostgreSQL or Azure PostgreSQL – BUT has to be different dbms than database used with Talend
- Deliverables: (upload the following)
 - Screen shots of completed jobs
 - Time each profiling job takes
 - Completed Talend jobs
 - List of tables with rows counts

Part 1: Load into separate staging tables

Data Source	Stage Table Name
2012_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
2013_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
2014_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
2015_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
2016_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
2017_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
2018_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
2019_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
2020_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
2021_Iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
bigquery-public- data .iowa_liquor_sales.sales	stg_iowa_liquor_sales_bigquery

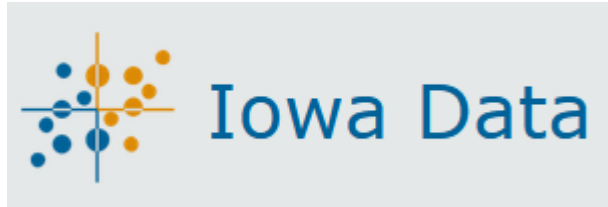
- Alternative process for loading stg_iowa_liquor_sales, instead of reading Iowa_Liquor_Sales.tsv loading data year by year using YYYY_Iowa_Liquor_Sales.tsv files
- Alternative process for loading stg_iowa_liquor_sales_bigquery, instead of reading bigquery-public-[data](#).iowa_liquor_sales.sales all at once iteratively load a data range, such as a year, at once

Part 1: Create a Dimensional Data Model

Stage Tables to be used for Dimensional Data Model
stg_iowa_liquor_sales
stg_iowa_liquor_products
stg_iowa_liquor_stores
stg_iowa_city_population_by_year
stg_iowa_county_population_by_year
stg_iowa_liquor_sales_by_year_and_county
stg_iowa_liquor_sales_by_year

- Create a dimensional data model in ER/Studio or Navicat (MacOS) using the stage tables listed
- Create the SQL DDL script for dimensional data model for same database as Stage tables are loaded into
- Deliverables: (upload the following)
 - Screen shot of data model
 - Data model file
 - SQL DDL script

Iowa Liquor Sales



Assignment 2

- Load integration schema (dimensional model)
- Query integration schema to answer business questions
- Use Alteryx data prep tool to load data



Assignment 2

- Load data from STG tables to dimensional model & perform data cleansing
 - Load another dbms using Talend (same as Assignment 1)
 - Load one dbms using Alteryx (same as Assignment 1) – optional. It will be counted as a bonus but ONLY if the Talend work is correct.
- Document data cleansing tasks and results
- Query dimensional data model with the listed business questions as a minimum
- Upload
 - Screen shots of completed loads
 - Time each load took (should be on screen shots above)
 - Completed Talend
 - Completed Alteryx jobs (bonus)
 - List of tables with rows counts
 - Results of each query

Part 2: Load Dimensional Model

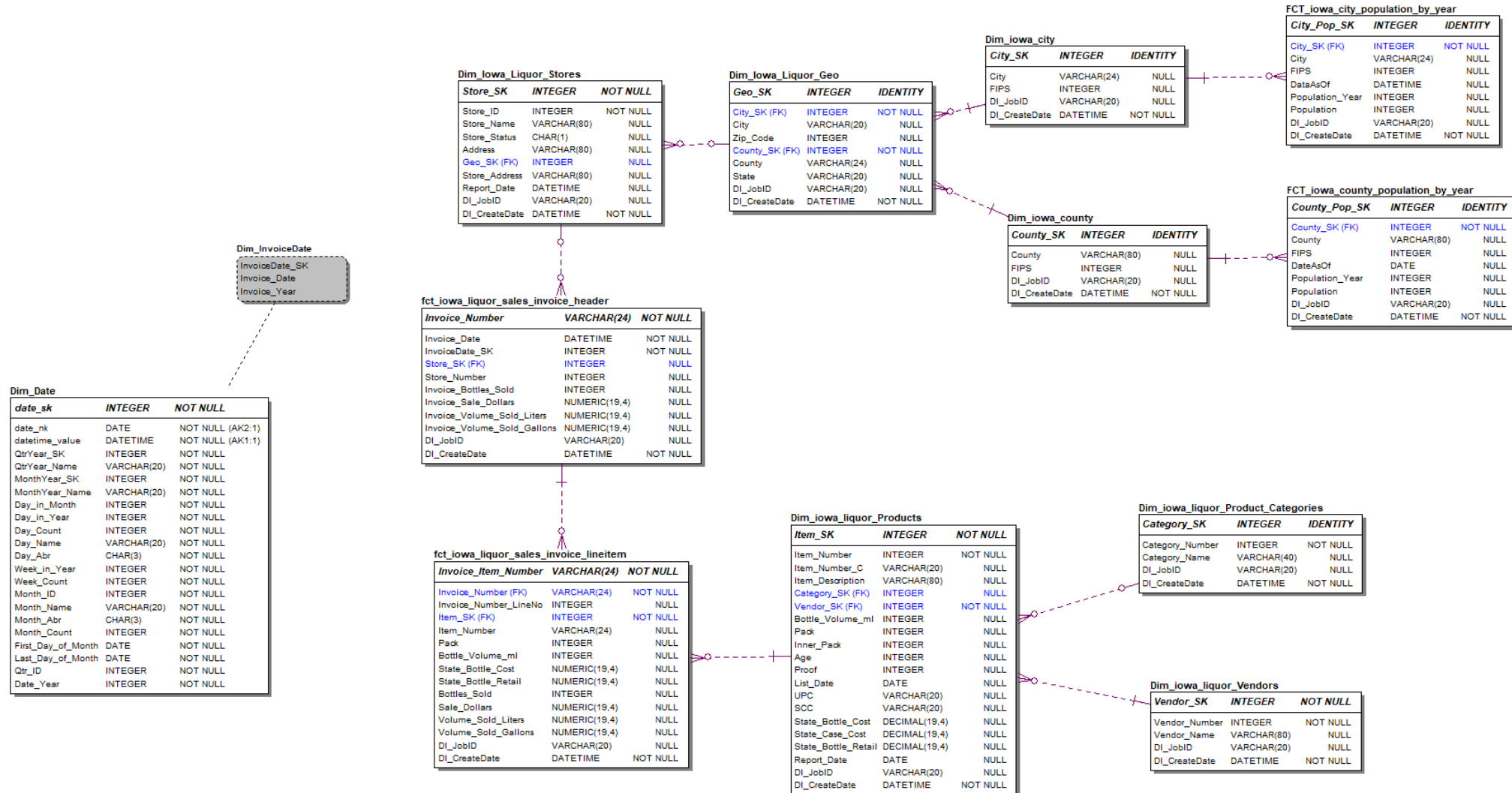
Stage Tables
stg_iowa_liquor_sales
stg_iowa_liquor_products
stg_iowa_liquor_stores
stg_iowa_city_population_by_year
stg_iowa_county_population_by_year
stg_iowa_liquor_sales_by_year_and_county
stg_iowa_liquor_sales_by_year

- Load staging tables with Talend
- Dimension tables options (you choose):
 - Azure SQL, SQL Server, MySQL or Azure MySQL
- Deliverables: (upload the following)
 - Screen shots of completed jobs
 - Time each profiling job takes
 - Completed Talend jobs
 - List of tables with rows counts

Iowa Liquor Sales: Business questions

- Liquor sales: Measures
 - Sales \$
 - Sales volume (gallons)
 - Sales volume (bottles)
 - Gross profit (retail price – cost)
 - Sales \$ per Capita
- Liquor Sales by time
 - Total
 - Year
 - Year, Month
 - Year over Year (YOY)
- Liquor sales by dimension
 - Store
 - County
 - City
 - Category
 - Item
 - Vendor

Dimensional Model



Iowa Liquor Sales: Business questions

- Liquor sales: Measures
 - Sales \$
 - Sales volume (gallons)
 - Sales volume (bottles)
 - Gross profit (retail price – cost)
 - Sales \$ per Capita
- Liquor Sales by time
 - Total
 - Year
 - Year, Month
 - Year over Year (YOY)
- Liquor sales by dimension
 - Store
 - County
 - City
 - Category
 - Item
 - Vendor

Loading Dimensional Model

Rules of the Road

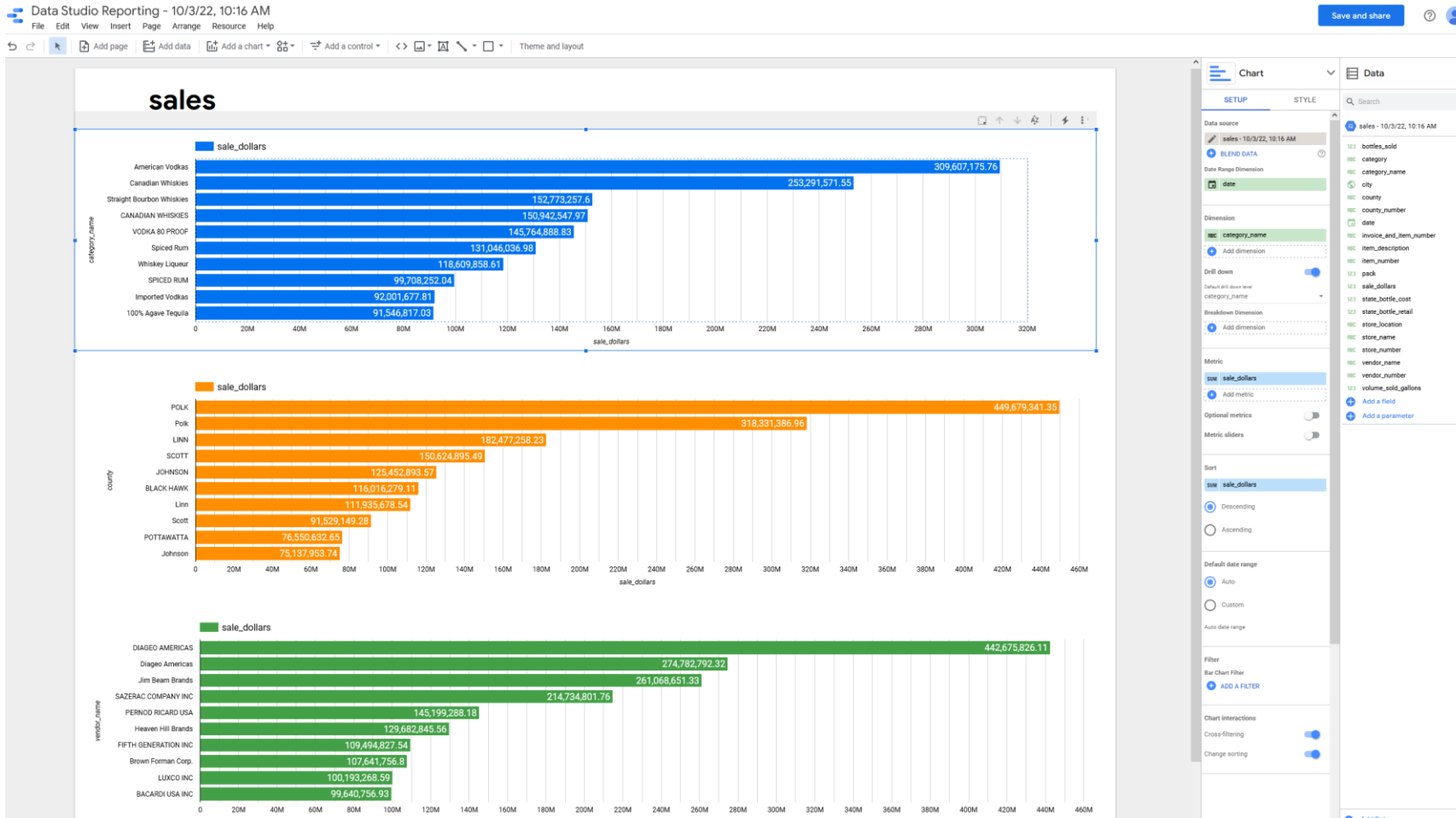
Iowa Data

- System of Record (SOR) for Facts & dimensions below.
- Iowa data reflects real-world conditions where the SOR for dimensions may NOT have all the historical values reflected in the sales transactions.
 - Supplement SOR with Sales values
- There are some inconsistencies in data, such as city and country names, so a slight amount of data cleansing is necessary. You may add reject tables or other WIP tables to complete.

Dimension or Fact	Table	Supplemental
Sales	stg_iowa_liquor_sales	
Items (Products)	stg_iowa_liquor_products	stg_iowa_liquor_sales
Vendors	stg_iowa_liquor_products	stg_iowa_liquor_sales
Stores with Address	stg_iowa_liquor_stores	stg_iowa_liquor_sales
Geo	stg_iowa_liquor_stores	stg_iowa_liquor_sales
City	stg_iowa_city_population_by_year	stg_iowa_liquor_sales, stg_iowa_liquor_stores
Country	stg_iowa_county_population_by_year	stg_iowa_liquor_sales, stg_iowa_liquor_stores

Data Matching & Cleansing

- Compare strings both in UPPER case



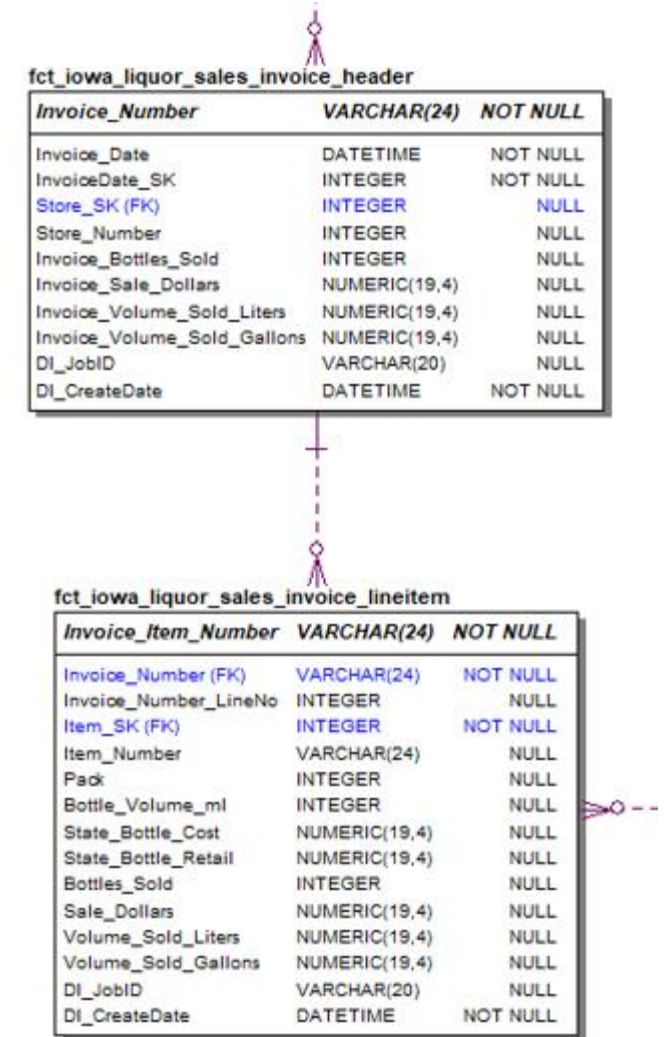
Fyi: Prior to 10/24/22
Several attributes were
entered in as mixed case
and uppercase (see
queries below)

Iowa Data: SCD

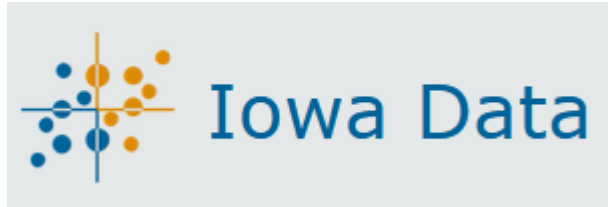
- SCD is NOT part of this assignment. Use the latest values for stores, store addresses, products, vendors, etc.

Iowa Data: Invoices

- In Sales data, Invoice_Item_Number is an invoice line item.
- Stores purchase alcohol via an invoice with one or more invoice line items.
- Need to split sales data between header and line items.
- Need to generate an invoice_number (header)



Iowa Liquor Sales



Assignment 3

- Query Integration Schema to answer business questions (not source files or tables)
 - SQL
- Create dashboards using Integration Schema to answer business questions
 - Power BI Service
 - Tableau Online



Iowa Liquor Sales: Business questions

- Liquor sales: Measures
 - Sales \$
 - Sales volume (gallons)
 - Sales volume (bottles)
 - Gross profit (retail price – cost)
 - Sales \$ per Capita
- Liquor Sales by time
 - Total
 - Year
 - Year, Month
 - Year over Year (YOY)
- Liquor sales by dimension
 - Store
 - County
 - City
 - Category
 - Item
 - Vendor
- Suggestions for above (but not limited to)
 - Ranked
 - Top 10
 - Distribution
 - Trends

Iowa Liquor Sales: Business questions

- Liquor sales by sales \$
 - Yearly sales including 2022 YTD
 - Top 20 stores (all-time)
 - Top 20 cities (all-time)
 - Top 10 counties (all-time)
 - Top 20 categories (all-time)
 - Top 50 items (all-time)
 - Top 20 vendor (all-time)