



Designing Advanced Data Architectures for Business Intelligence

DAMG 7370

Northeastern University

Final Project

[Iowa Liquor Sales | data.iowa.gov](https://data.iowa.gov)

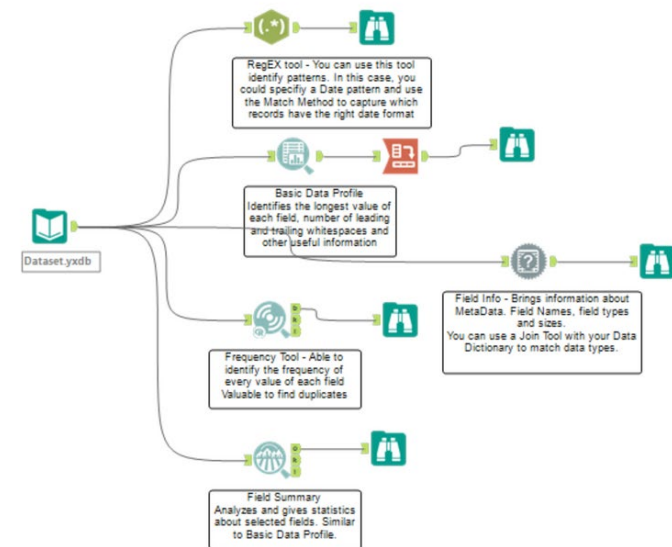
Iowa Liquor Sales – Business Questions

- 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
- Overall consumption of liquor year wise
 - Write your inferences on the data
- Which type of sprit/liquor had more sales
 - Top 3 names
- What bottle size is favorable?
 - Had more sales

Iowa Liquor Sales – Business Questions

- Liquor sales by sales \$
 - Yearly sales including 2024 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)

Iowa Liquor Sales



Final Team Project

Part – 1

The screenshot shows the Talend Data Preparation interface. On the left is a sidebar with navigation options: Preparations, Datasets, Connections, and Semantic types (which is highlighted). The main area displays a table of semantic types. A circular callout highlights a subset of these types, which are also shown in a larger, detailed view on the right.

Name	Description	Type	Modified by	Published on	Published by	Status
<input type="checkbox"/> Company	Company Names	Dictionary	N/A	2016-12-08 15:38:49	Talend	Published
<input type="checkbox"/> Continent	Continent name (multi...	Dictionary		2016-12-08 15:38:49	Talend	Published
<input checked="" type="checkbox"/> Continent Code	Continent code	Dictionary		2016-12-08 15:38:49	Talend	Published
<input checked="" type="checkbox"/> Country	Country Name (EN+FR)	Dictionary		2016-12-08 15:38:49	Talend	Published
<input type="checkbox"/> Country Code ISO2	Country code of ISO31...	Dictionary		2020-10-06 09:25:00		Published
<input type="checkbox"/> Country Code ISO3	Country code of ISO31...	Dictionary		2016-12-08 15:38:49	Talend	Published
<input type="checkbox"/> Currency Code	Currency alphabetic c...			N/A	N/A	Draft

Part –1A: Iowa Liquor Sales – Data Profiling and Fixing

(A) Profiling and Fixing (with Alteryx/Talend)

- Find data gap on sales data between
 - iowa_Liquor_Sales.tsv
 - iowa_Liquor_Sales_notCleaned.tsv
- Identify data differences between each respective files
- stage the data using Talend and find the data differences.
 - Create schema called **profiling** and under this schema create the stage tables
 - You can use any database of your choice
 - All best practices of Talend job to be used as discussed in the class
- Deliverables
 - Screen shots of completed profiling
 - Time metrics each profiling job takes
 - Time metrics of each talend job for staging
 - Completed Alteryx jobs
- Outline data cleaning tasks
 - Identify what needs to be cleaned & explain how you are going to do it using Talend
 - Create talend jobs to clean/merge/fix the data
 - If you want, you can fix data while staging using Alteryx

Data Source	Stage Table Name Schema: profiling
iowa_Liquor_Sales.tsv	stg_iowa_liquor_sales
iowa_Liquor_Products.tsv	stg_iowa_liquor_products
iowa_Liquor_Stores.tsv	stg_iowa_liquor_stores
iowa_Liquor_Sales_notCleaned.tsv	stg_iowa_liquor_sales_notcleaned
iowa_Liquor_Products_notCleaned.tsv	stg_iowa_liquor_products_notcleaned
iowa_Liquor_Stores_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

Part –1B: Iowa Liquor Sales – Staging

(B) Stage data with ADF

- Staging tables DB (you choose)
 - Snowflake, Azure SQL, SQL Server, MySQL
 - Must use metadata based loading
 - Capture all metrics
 - Validate the load row counts (Sanity checks)
 - Load in schema named **stage**
- Deliverables
 - Screen shots of completed jobs
 - Time each profiling job takes to log table
 - Pipeline Json code with data set and link server code
 - List of tables with rows counts captured in log table
- Note: If you are facing issue with loading one big file, then use year wise files to load

Data Source	Stage Table Name Schema: stage
Iowa Liquor Sales.tsv	stg_iowa_liquor_sales
Iowa Liquor Products.tsv	stg_iowa_liquor_products
Iowa Liquor Stores.tsv	stg_iowa_liquor_stores
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

Part –1C: Iowa Liquor Sales – Data Modeling

- Create a dimensional data model in
 - ER/Studio
 - If Mac users have issues with ER/studio then use Navicat
 - stage tables to be used (as defined to the right side)
- Create the SQL DDL script for dimensional data model for same database as Stage tables are loaded into
- Deliverables
 - Screen shot of data model
 - Data model file
 - SQL DDL script
 - Modeled tables to be crated in a schema called **edw**

Stage Tables to be used for Dimensional Data Model from schema stage

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



Iowa Data

Final Team Project

Part – 2



Part –2: Iowa Liquor Sales – Integration schema loading

- Load data to Dimensional model (Integration schema)
 - Use Talend
- Query Integration schema to provide answers related to business
- Deliverables
 - Screen shots of completed loads
 - Time each load took (should be on screen shots above)
 - Completed Talend jobs
 - Job start and end times captured in log table
 - All best practices on talend job to be implemented (Exception handling, Logging, parameters, etc)
 - List of tables with rows counts
 - Results of each query



Iowa Data

Final Team Project

Part – 3



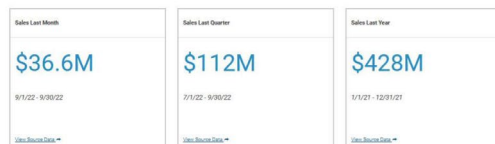
Part –3: Iowa Liquor Sales – Visualization

- Create [dashboards](#) using Integration Schema to answer business questions
 - Power BI Desktop / Online
 - Tableau Desktop / Online
 - Few reports directly captured from Iowa Open site



This explorer presents spirits purchase information of Iowa Class "E" liquor licensees (e.g., grocery stores, liquor stores, convenience stores, etc.) by product and date of purchase beginning January 1, 2012.

Sales Snapshot



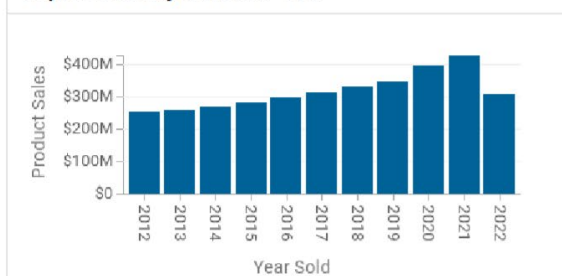
Filtering

Data is filterable by county, city, date of purchase, category name (e.g., vodka, imported gin, etc.), store name (i.e., store that purchased the product), item description (i.e., brand or product), and state bottle retail price. To filter, click the dropdown on a filter, search and select value or values to filter by, then click apply. More than one filter may be applied as desired.

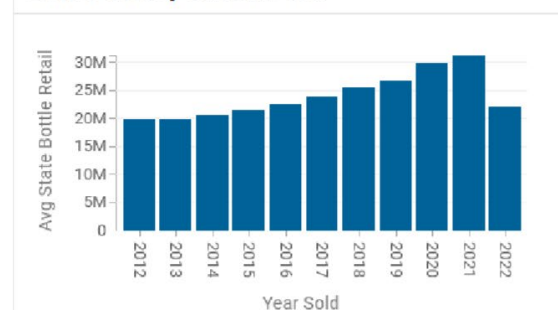
Sales Summary



Liquor Sales by Calendar Year



Bottles Sold by Calendar Year



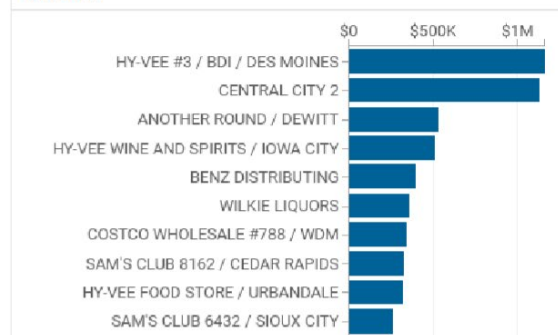
Top Sales by Product

Last Month



Top Sales by Store

Last Month





Questions?