Iowa Liquor Sales

Iowa Data

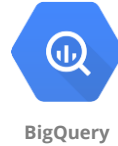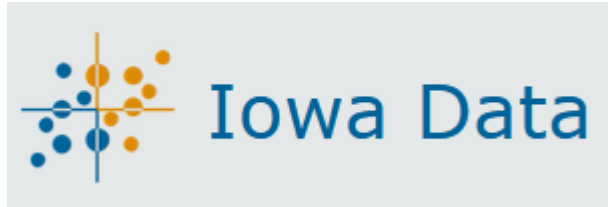**BigQuery**

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
- <u>Outline</u> data cleaning tasks
  - Identify what needs to be cleaned & explain how you are going to do it

ATHENA
IT SOLUTIONS

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

ATHENA
IT SOLUTIONS

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

ATHENA
IT SOLUTIONS

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

ATHENA
IT SOLUTIONS

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

ATHENA
IT SOLUTIONS

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

ATHENA
IT SOLUTIONS

Part 1: Create a Dimensional Data Model

Stage Tables to be used for Dimenional 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

ATHENA
IT SOLUTIONS