WELCOME TO THE DATA VISUALIZATION LAB

Part 3 - Data Visualization [25 minutes]

Overview: What is Data Visualization and how do we use it with our data?

Purpose: Create visualization using the flight information answering the question (visually with a density graph):
   ● What were the most number of flights from destination to origin between (1995-2008) - Route Density

1) Open CDP, using the "admin" user within the Test Drive link.

Your link should look something like (remember click the link in your email not the link below)
http://login.trycdp.com/auth/realms/trycdp-trialxx/protocol/saml/clients/samlclient?tn=trialxx_admin@trycdp.com&p=X
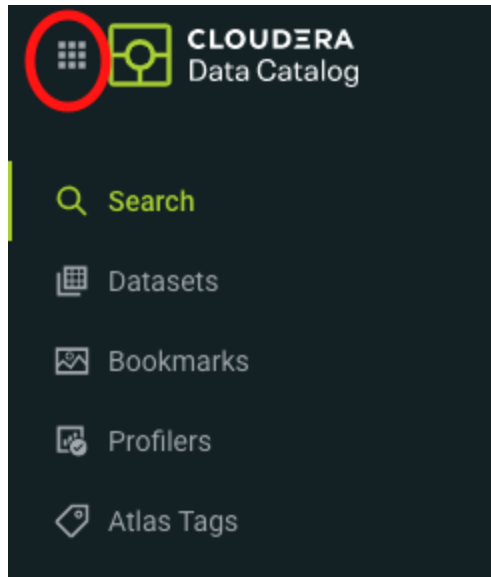*xx represents the trial user #
*X represents the password
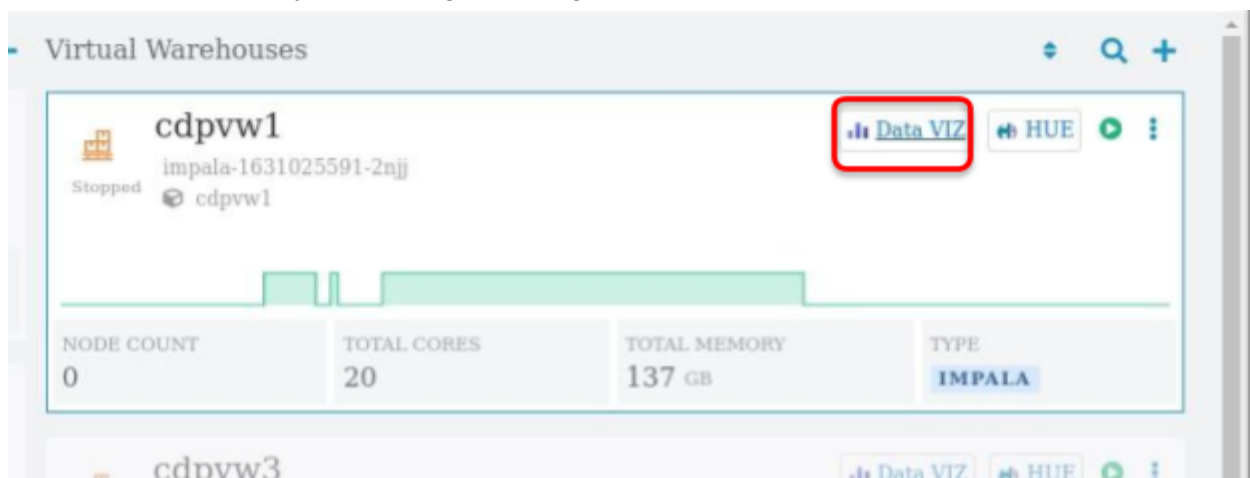
2) Click the "Data Warehouse" within the CDP Home Screen

How do you get to the CDP Home Screen?

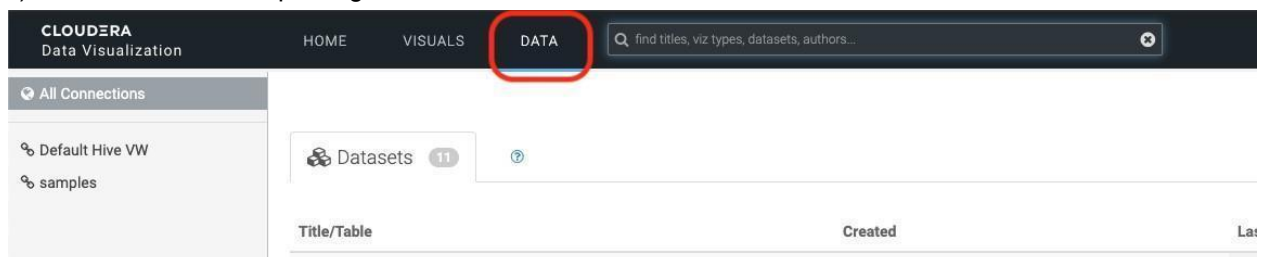- From any experience such as "Data Catalog", click the 9 square at the top left and then click "Home"

3) Click "Data VIZ" on your existing "Running" Virtual Warehouse.

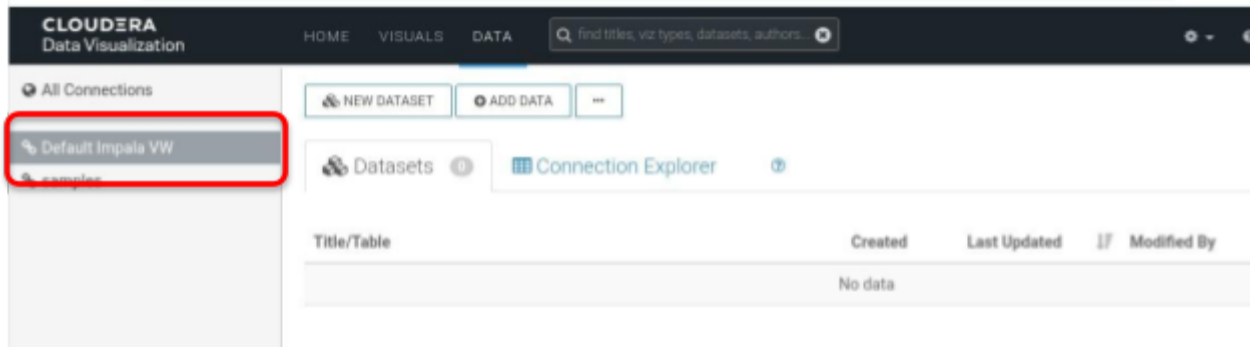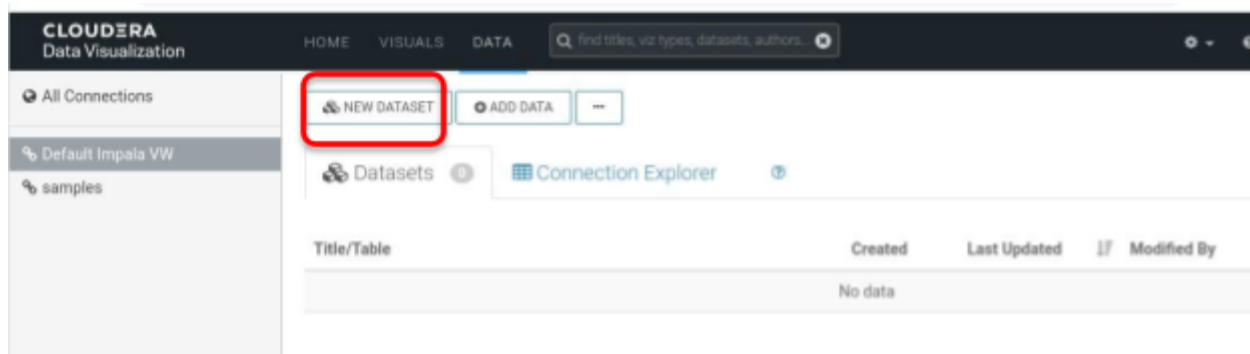
4) You are autologged into Data VIZ

5) Click "DATA" the top navigation bar

6)   Click "Default Impala VW" to add our dataset.

7)   Click "NEW DATASET" to add our "flights" data

8)   Enter a name for the Dataset title naming "mynamehere_airline_new_parquet_flights" Recall Hive prefers ORC and Impala prefers Parquet. Our development team is making both engines like both file formats equally.

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

martyflightsnewparquet

Dataset Source

From Table ⌄

Select Database

airlines_new_parquet_cdpvw1 ▾

Select Table

flights ▾

CANCEL   CREATE

9) Choose the database "airlines_new_parquet_YOURVIRTUALWAREHOUSENUMBER"

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

| martyflightsnewparquet |

Dataset Source

| From Table ▼ |

Select Database

| airlines_new_parquet_cdpvw1 ▼ |

Select Table

| flights ▼ |

CANCEL   CREATE

10) Choose the table "flights"
*Need to import multiple databases and tables? You'd use Dataset Source = SQL

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

martyflightsnewparquet

Dataset Source

From Table

Select Database

airlines_new_parquet_cdpvw1

Select Table

flights

CANCEL   CREATE

11) Click "CREATE"

New Dataset

Create a dataset from data on this connection. You need to create a dataset before you can create dashboards or apps.

Dataset title *

martyflightsnewparquet

Dataset Source

From Table

Select Database

airlines_new_parquet_cdpvw1

Select Table

flights

CANCEL   CREATE

12) Click "+" to create a New Dashboard

13) Choose "TreeMap" under "VISUALS" or explore with the Wizard
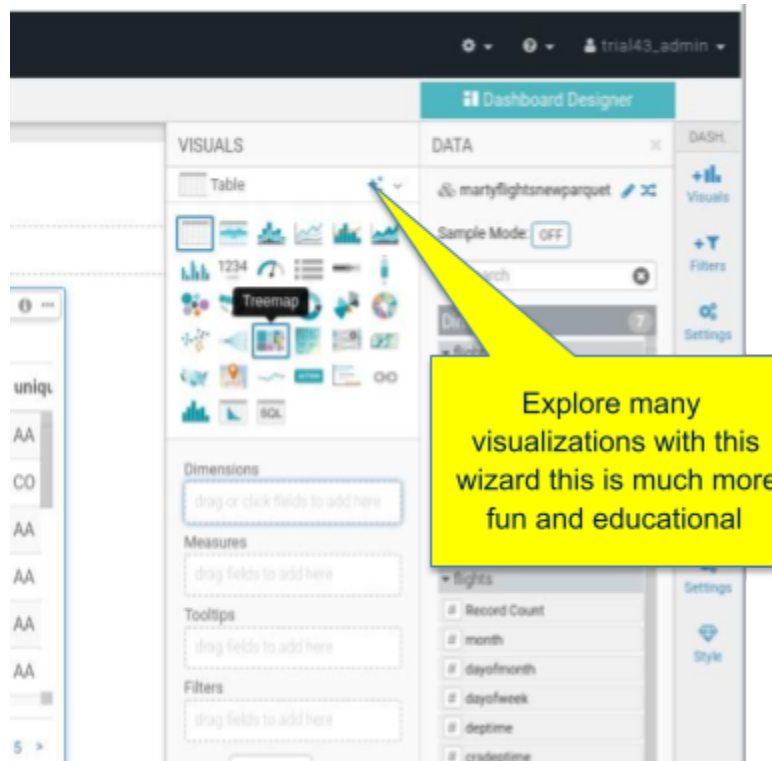
14) Drag-and-drop both "dest" and origin" from Dimensions->Flights into Dimensions under
Visuals

15) Drag-and-drop "Record Count" from Measures->Flights into Measures under Visuals

16) Click the right arrow next to Record Count and select "Descending" under Order and Top K
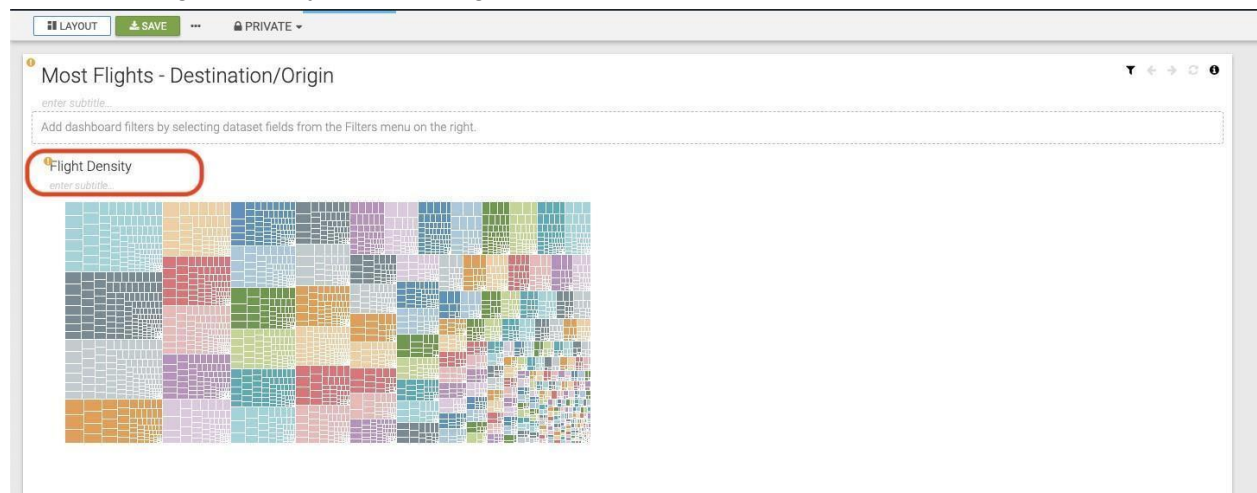
17) Click "REFRESH VISUAL"

*Notice - you can have other Visuals chosen to be displayed with the Dimensions and Measure(s), then click REFRESH VISUALS
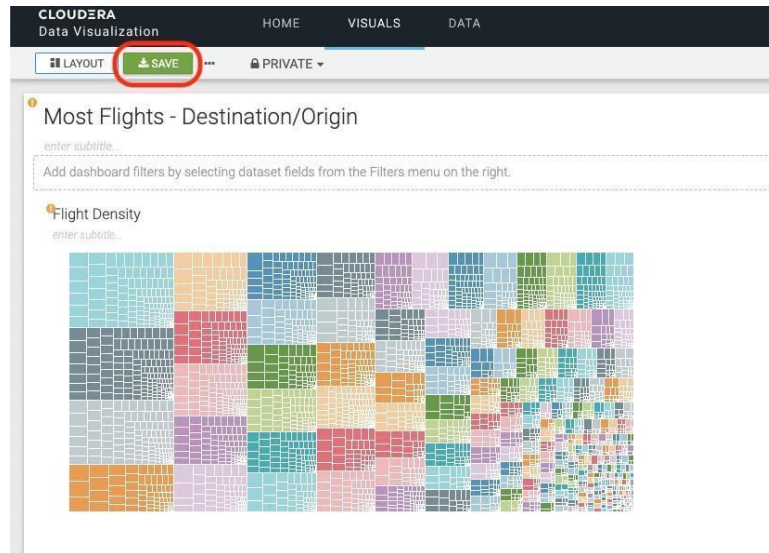
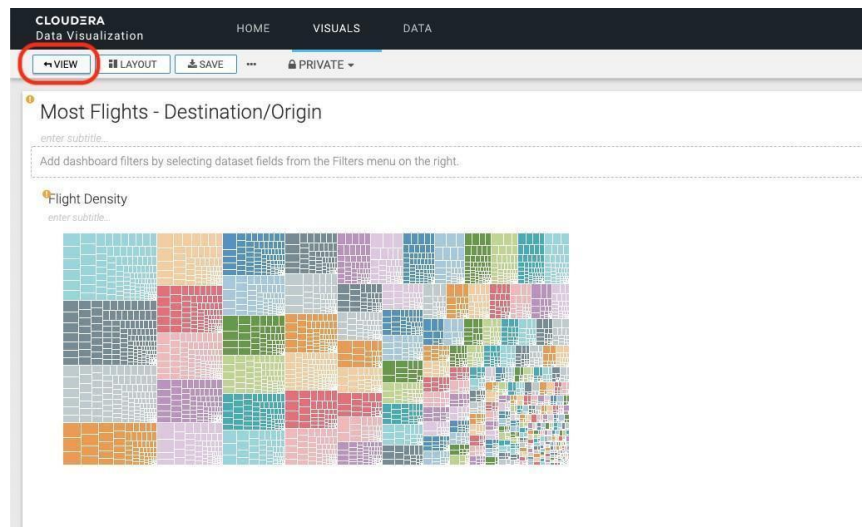**18) Enter a title "Most Flights - Destination/Origin"**
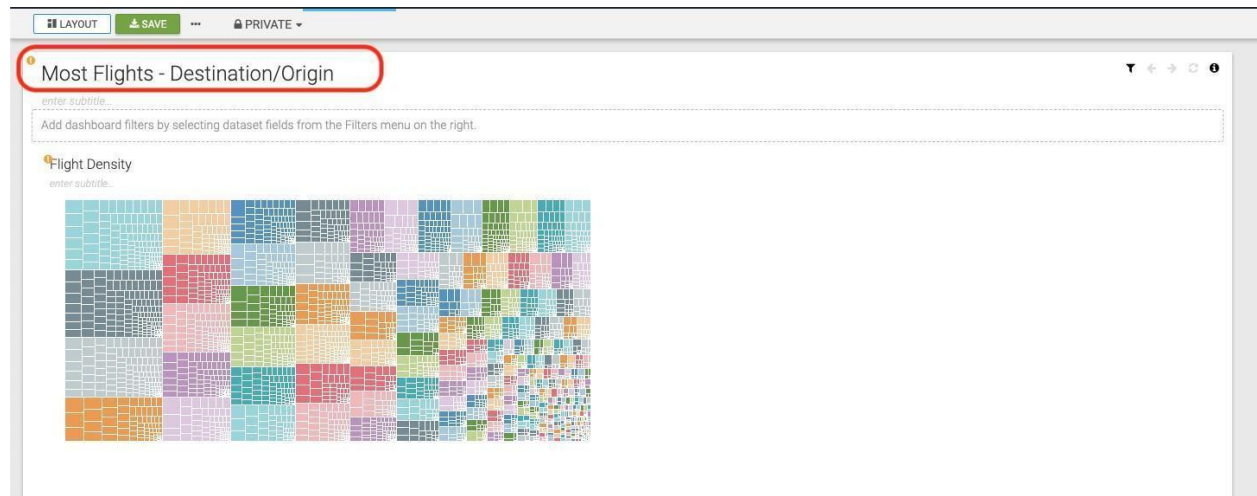
19) Enter "Flight Density" under the graph's title
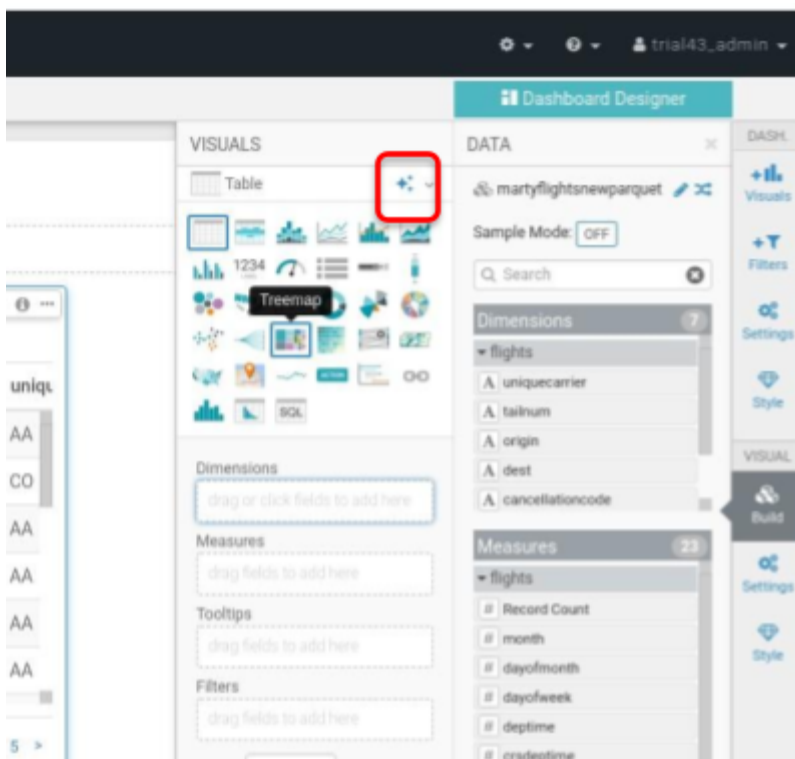
20) Click "SAVE"

21) Click "VIEW"

22) Scroll over the graph and click "Expand Visual"

Destinations are displayed

23) Time to explore the other visualization options.

Please scroll through the options and experiment with some of the additional visuals.