LAB

03 – Cloudera Data Warehouse (Part 1)

Data Lifecycle on CDP Public Cloud

Data Warehouse Lab

The ‘Data Warehouse Lab’ is divided into 2 Parts.
The 1st Part is around ‘Dashboard development’.
The 2nd Part focusses on integration between Data Viz and Cloudera Machine Learning.

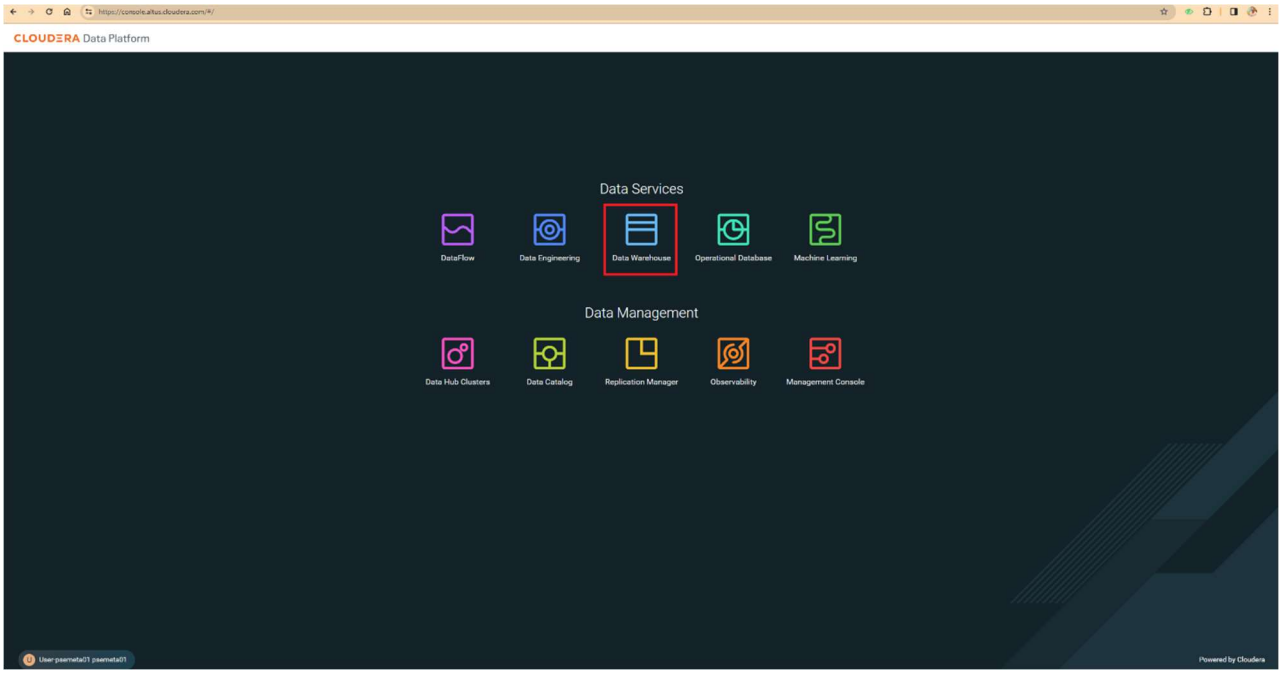
Let’s work on the 1st Part now.

Part 1: Dashboard development

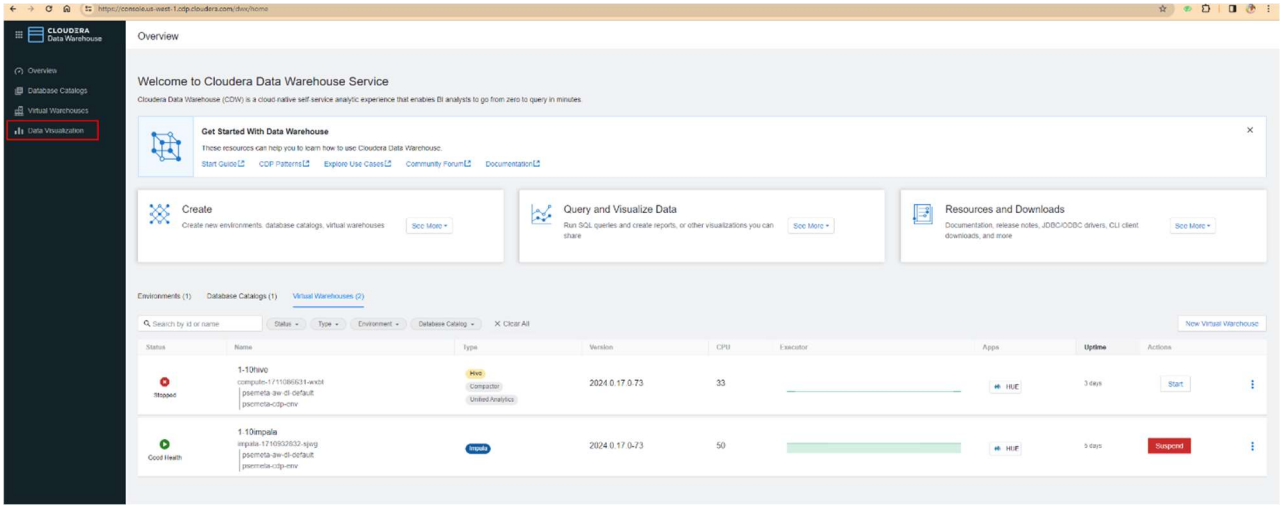
Goals:

- Create a dataset pointing to the table.
- Create a dashboard with metrics and dimensions.

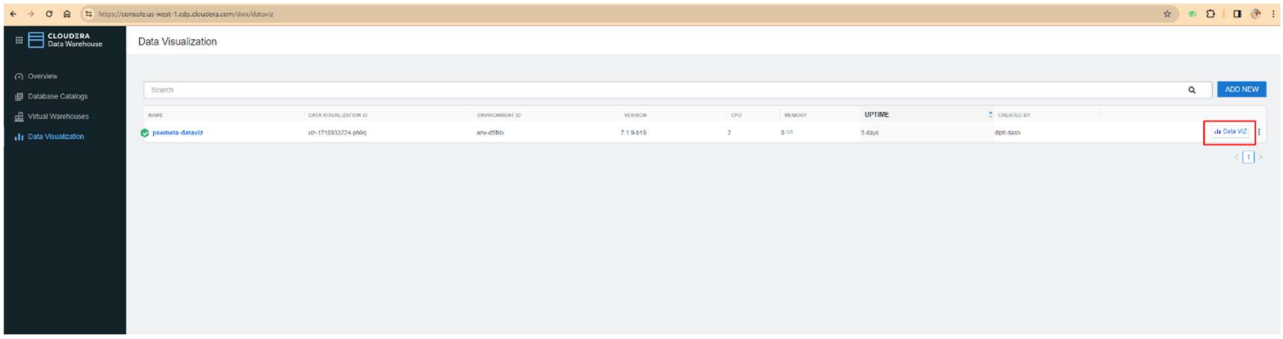
1. Click on Data Warehouse from CDP PC Home.



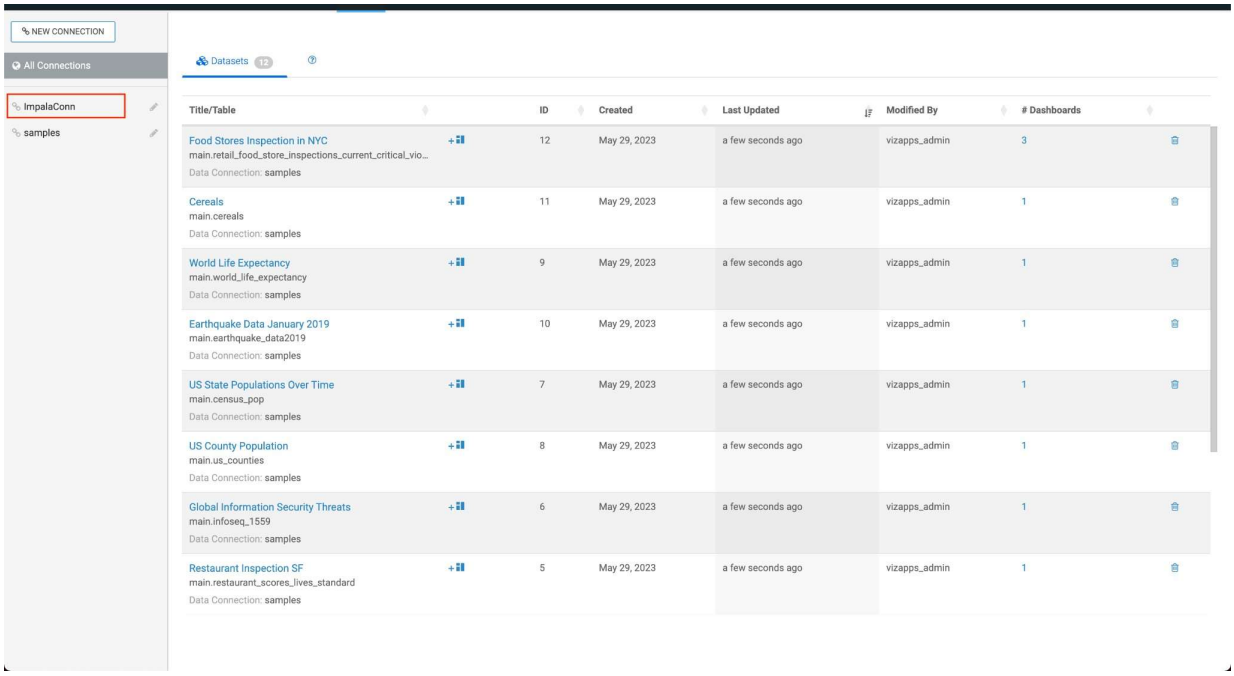
2. Data Warehouse welcome screen. Click on Data Visualization in the left menu.



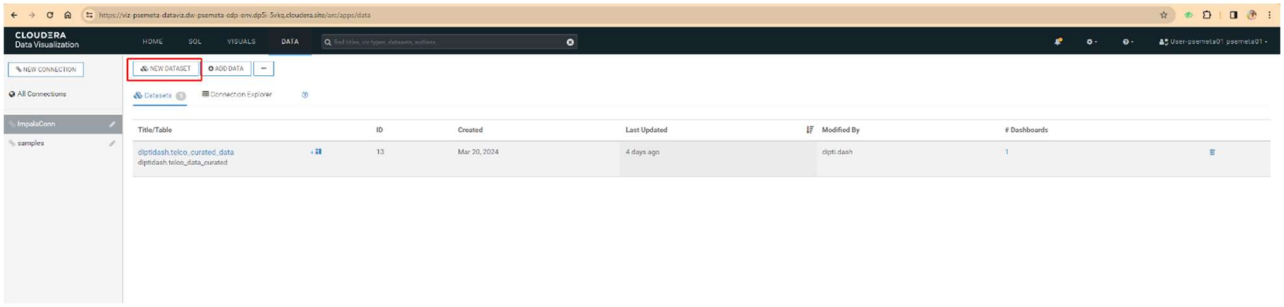
3. In Data Visualization, click on the button **Data Viz** from which they were assigned.



4. Once in Data Visualization, go to the Data option from the top menu, and then to the Connector **ImpalaConn** from the left menu. Or, chose the **connection** that is assigned to you (check with instructor on this).



5. We must create a new data source, for that, click on **NEW DATASET** and a window will appear to enter the information of the new data source.



6. Enter the information for the new data source:

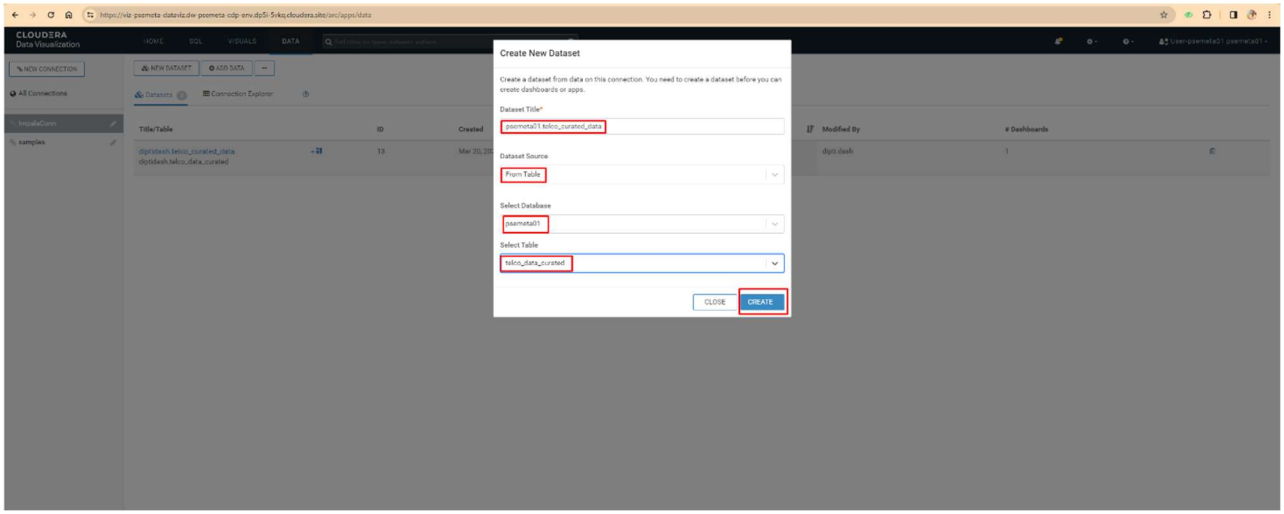
Dataset title: <assigned_user>.telco_curated_data Ex: ***psemeta01.telco_curated_data***

Dataset Source: From table

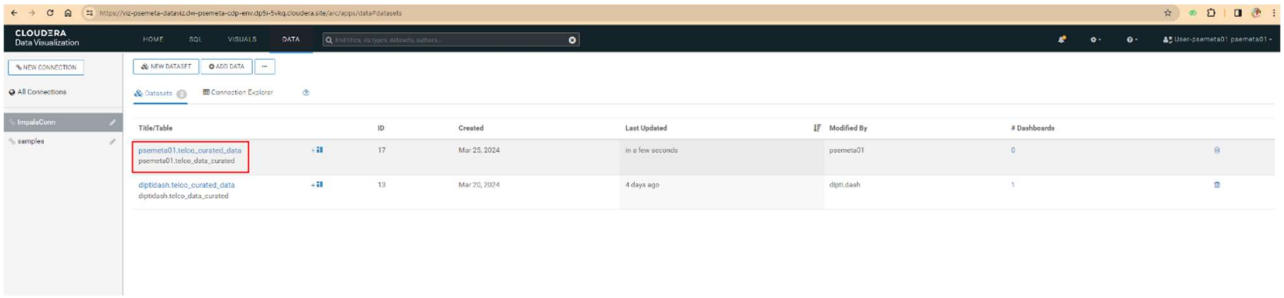
Select Database: <assigned_user> Ex: ***psemeta***

Select Table: telco_data_curated

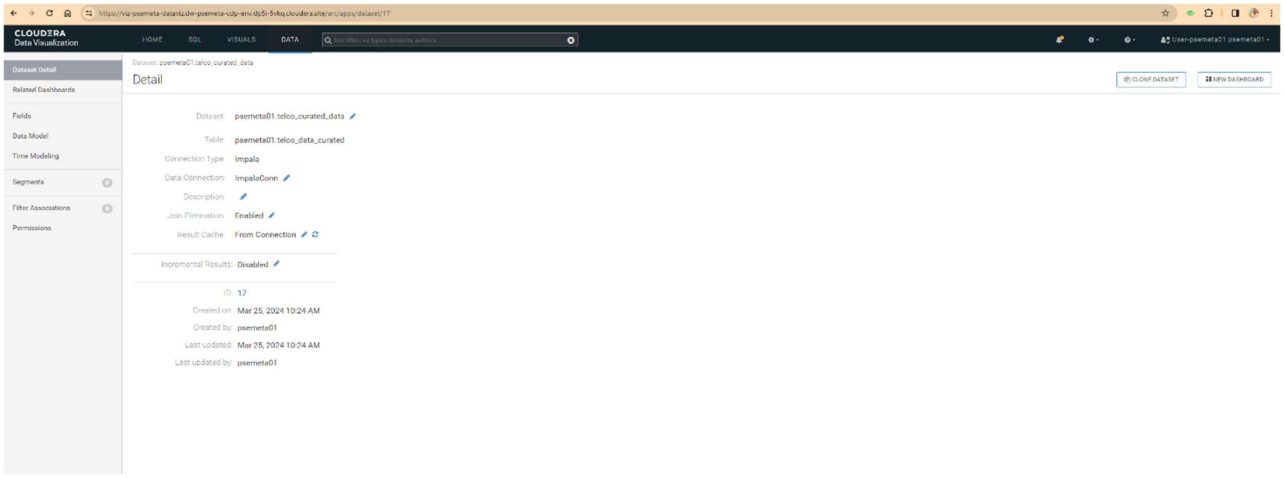
Click on Create to create the new Dataset.



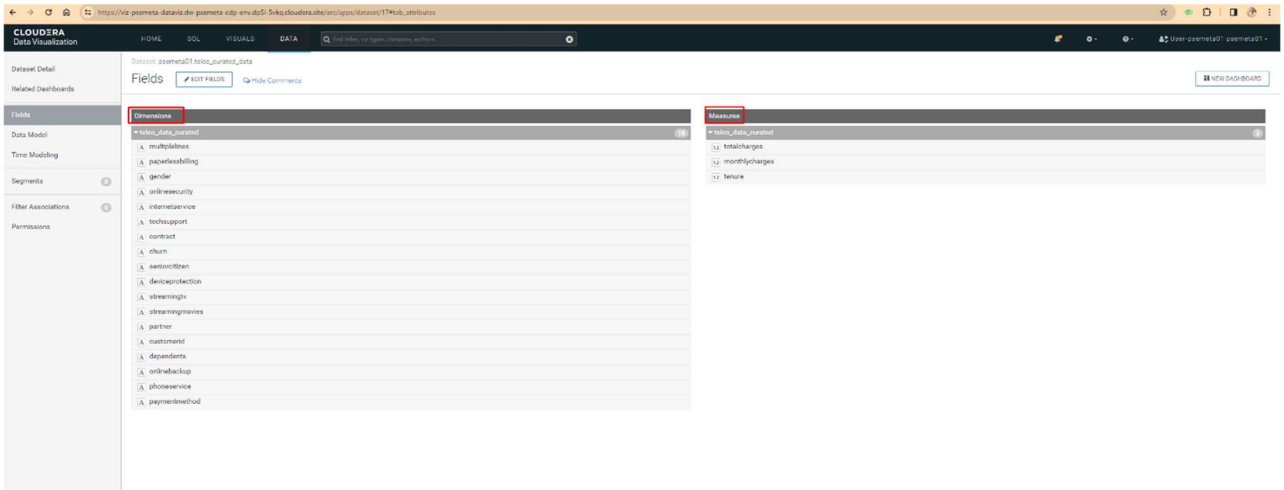
7. The new Dataset should appear in the list. Click on the dataset that you just created.



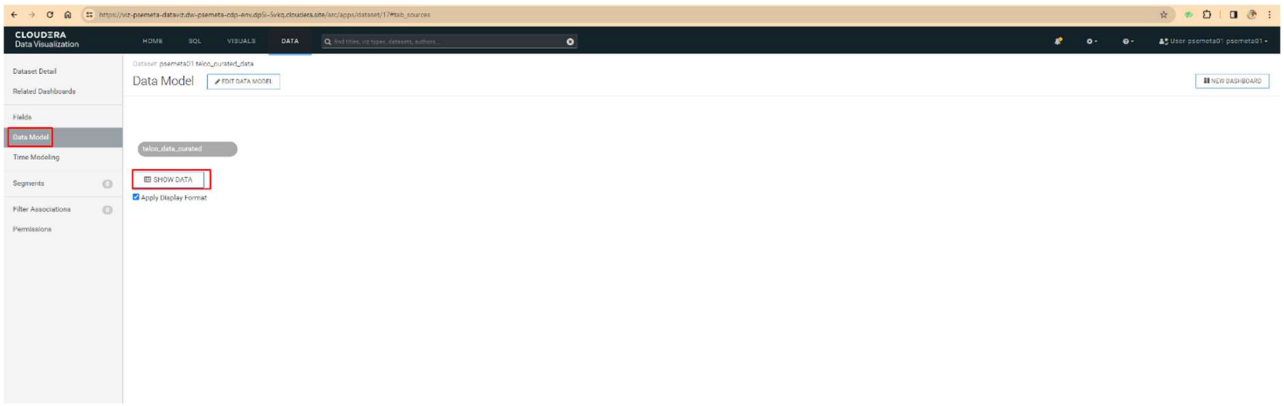
8. Here you will see the details of the dataset.



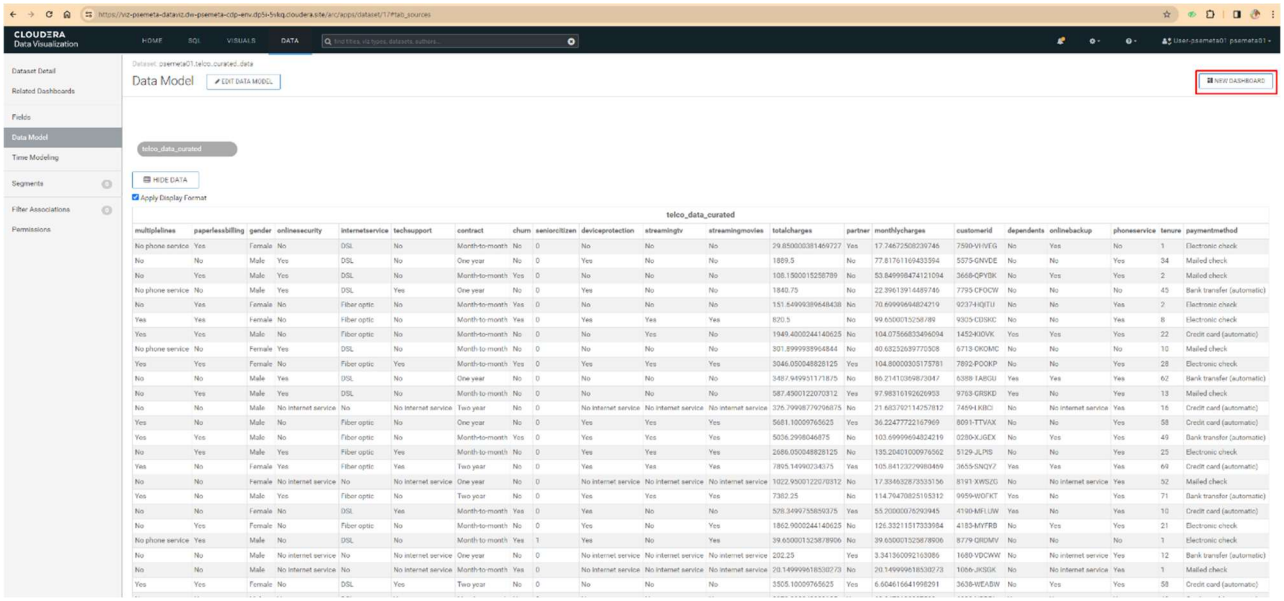
9. Click on **Fields** (left menu) to see the fields automatically captured during the dataset creation process.



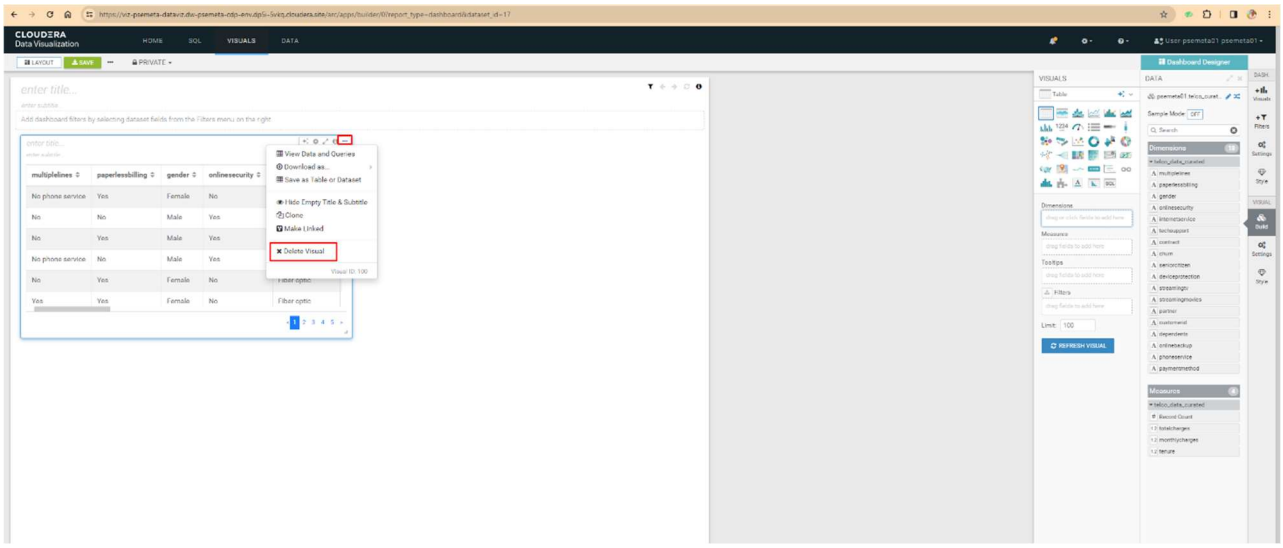
10. You can also preview the data from this screen. Click on **Data Model** (left menu) and then on the button **SHOW DATA** that appears in the center.



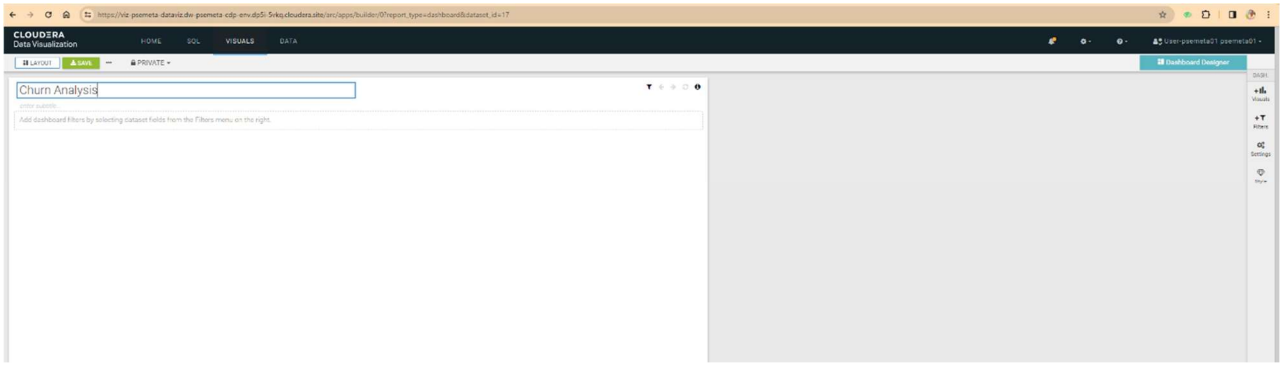
11. At this moment, a query to the Virtual Warehouse is executed to retrieve the data from the data set. Notice the columns and values. Click **NEW DASHBOARD** to create a new dashboard.



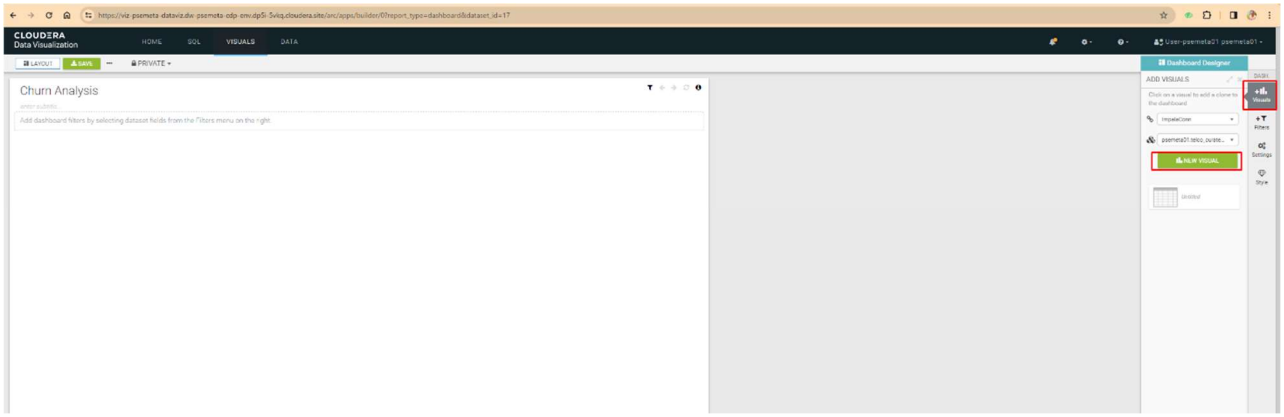
12. When opening the design canvas of a new panel, remove the element that is added by default, by clicking on the three dots (...) button at the top right of the element, and then clicking on the option **Delete Visual**



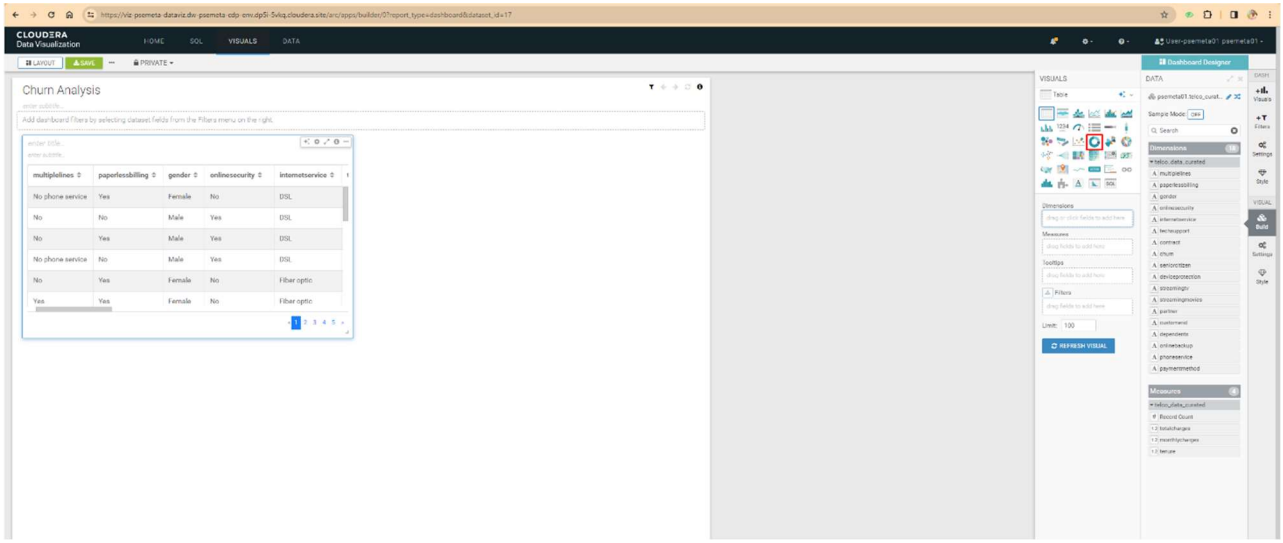
At the top of the canvas, in the enter title field, enter the name ***Churn Analysis*** to identify the dashboard.

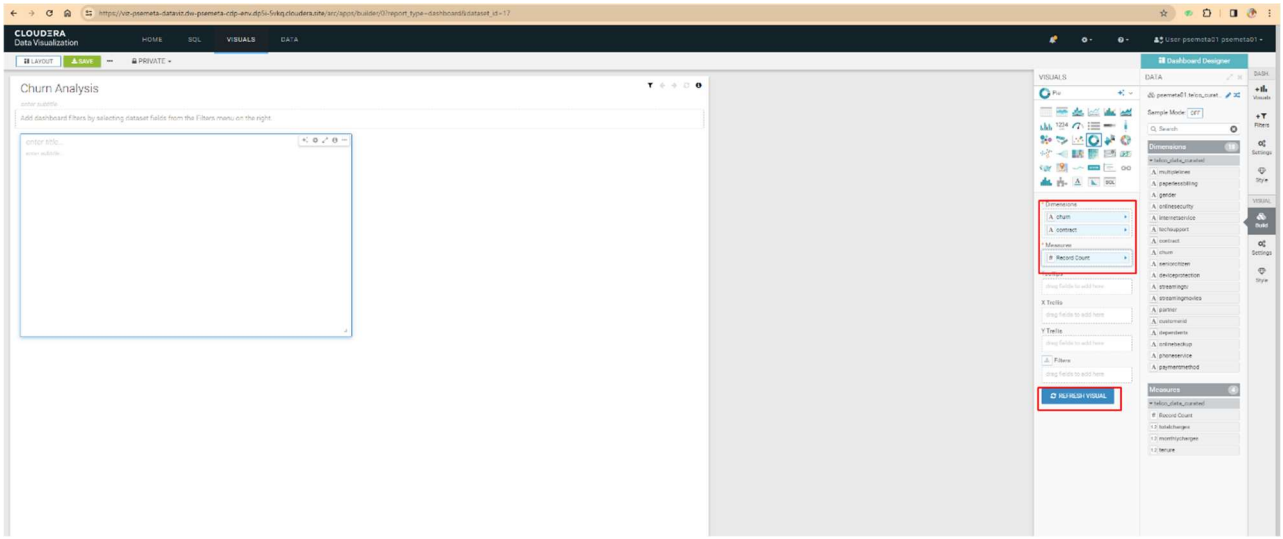


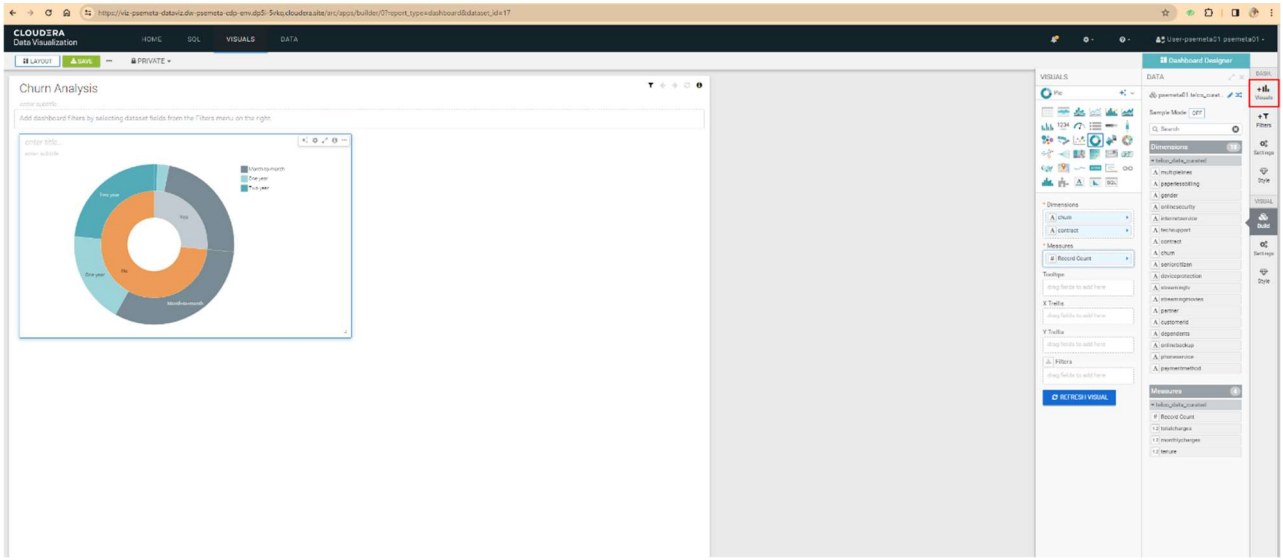
13. To add a new visual element, click on the button **Visuals** from the right menu, select the dataset that corresponds to them, and click on the button **New Visual.**



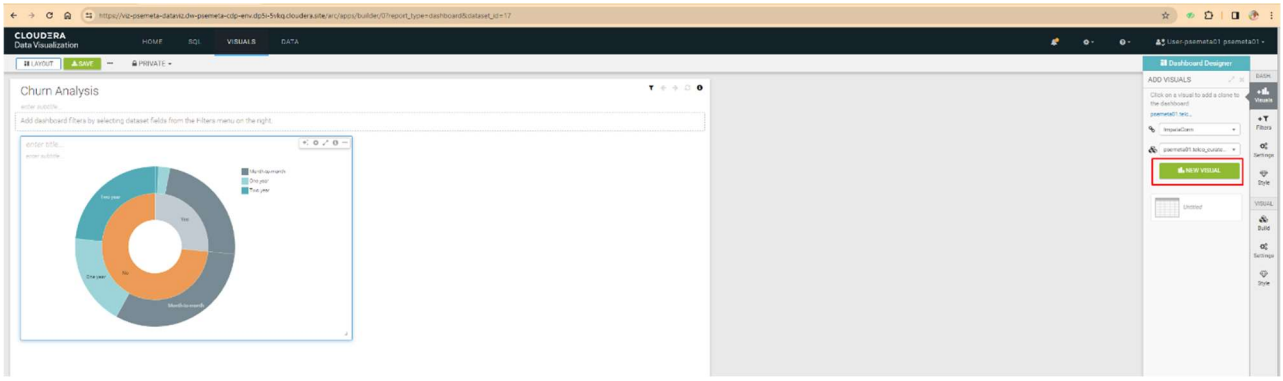
14. Add the first visual element, which is a pie chart with the dimensions **churn** and **contract**, with the metric of **Record count**. Once finished, click the button **Refresh Visual**.

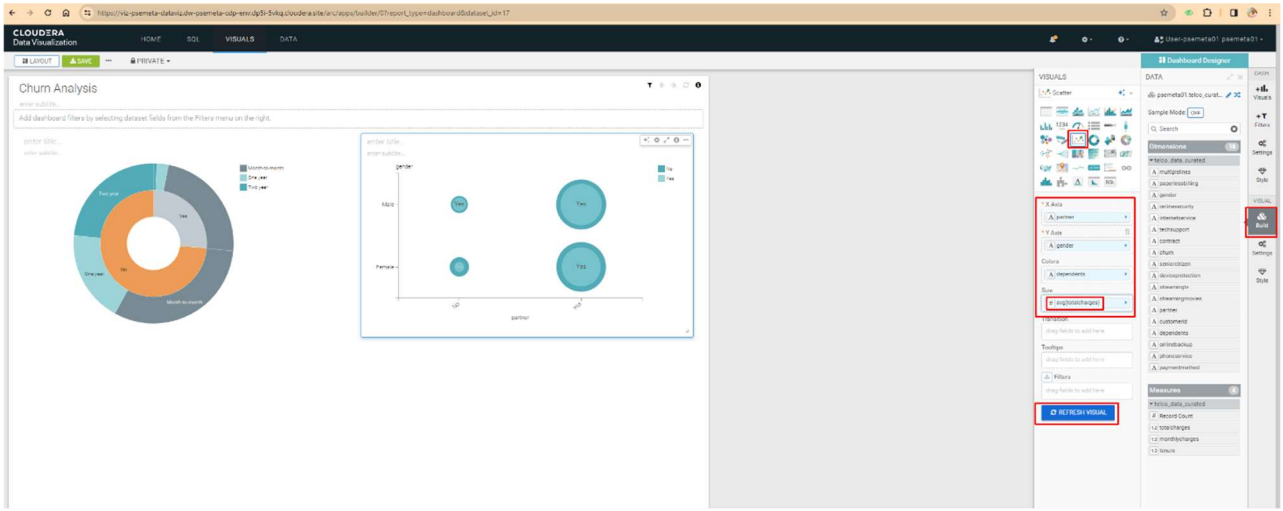



CLOUDERA
Data Visualization
HOME
SQL
VISUALS
DATA
Churn Analysis
REFRESH VISUAL


CLOUDERA
Data Visualization
HOME
SQL
VISUALS
DATA
Churn Analysis
REFRESH VISUAL

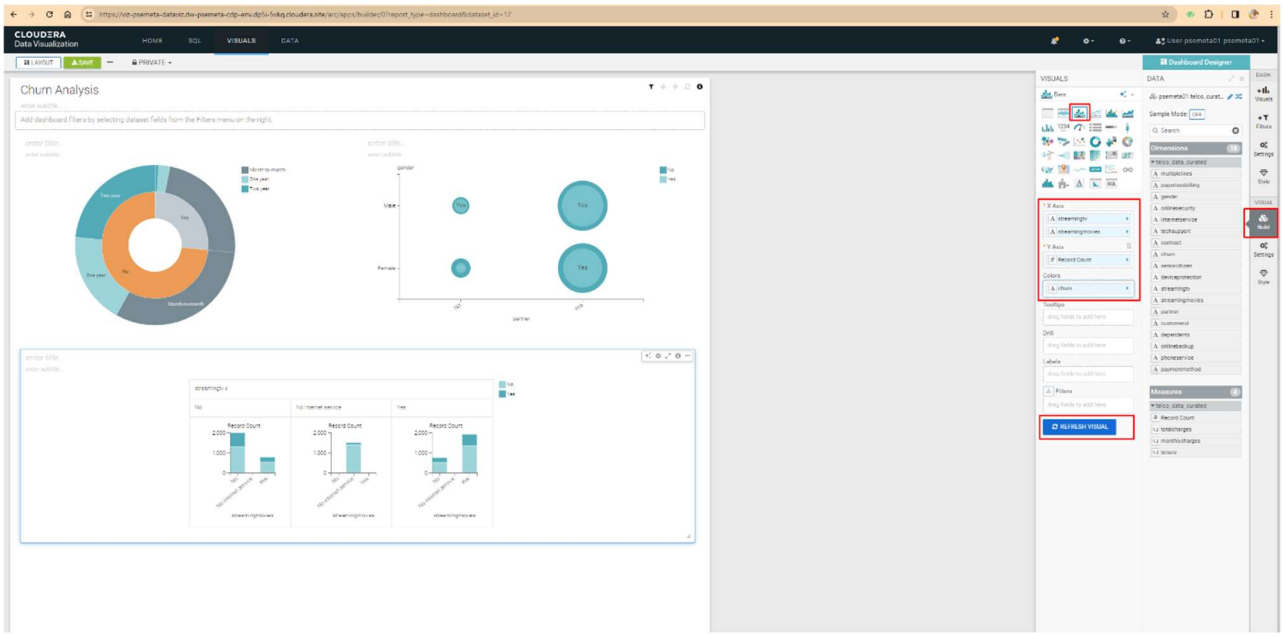
15. Add the second visual element, which is a scatter chart with the dimension **partner** like X Axis, **gender** how Y Axis, **dependents** as Colors and **avg (total charges)** as Size. Once finished, click the button **Refresh Visual**.





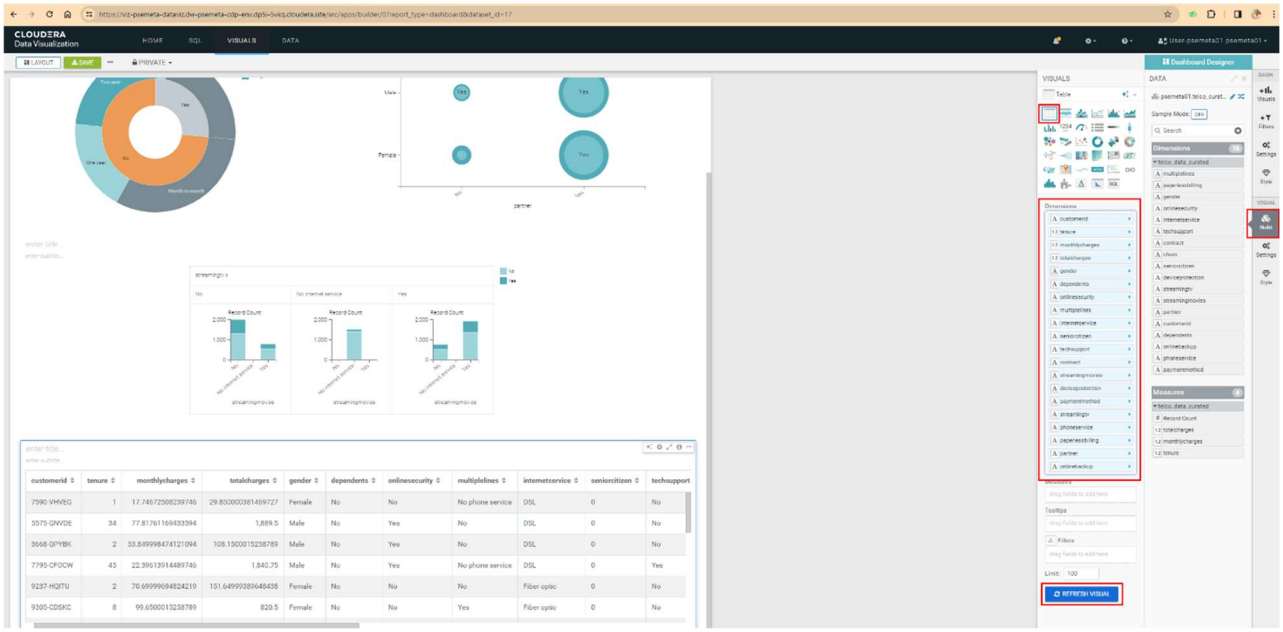
15. Add the third visual element, which is a bar chart with the dimensions **streamingtv** and **streamingmovies** like X Axis,

Record Count how Y Axis and **churn** how Colors. Once finished, click the button **Refresh Visual**.



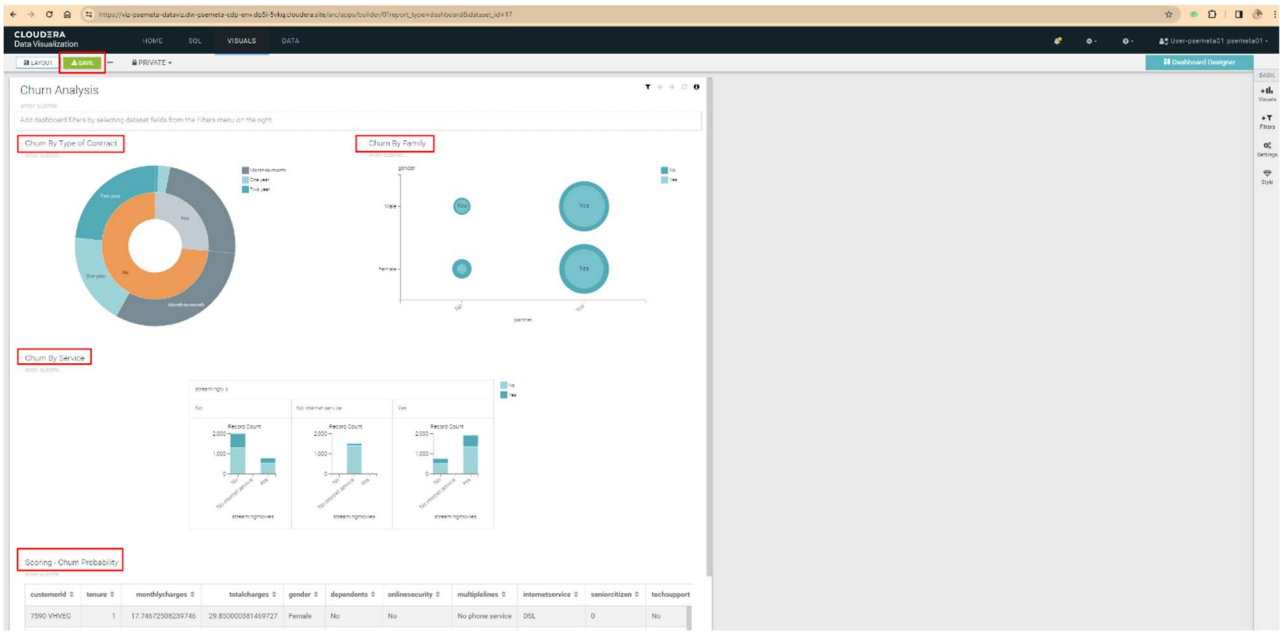
16. Add the fourth and last visual element, which is a table with the dimensions and metrics of the dataset. Be sure to add all 17 dimensions and 3 metrics to the table, so in total 20 elements. Once finished, click the button **Refresh Visual**.


Dimensions
customerid
tenure
monthlycharges
totalcharges
gender
dependents
onlinesecurity
multiplelines
internetservice
seniorcitizen
techsupport
contract
streamingmovies
deviceprotection
paymentmethod
streamingtv
phoneservice
paperlessbilling
partner
onlinebackup
Measures
drag fields to add here



Name the visuals that you have created as follows and then save the dashboard by clicking the button **Save** from the top menu.

1. Churn By Type of Contract
2. Churn By Family
3. Churn By Service
4. Scoring - Churn Probability



At this point we will go ahead and complete the lab on CML and then make a call to the model from Data Viz.