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.

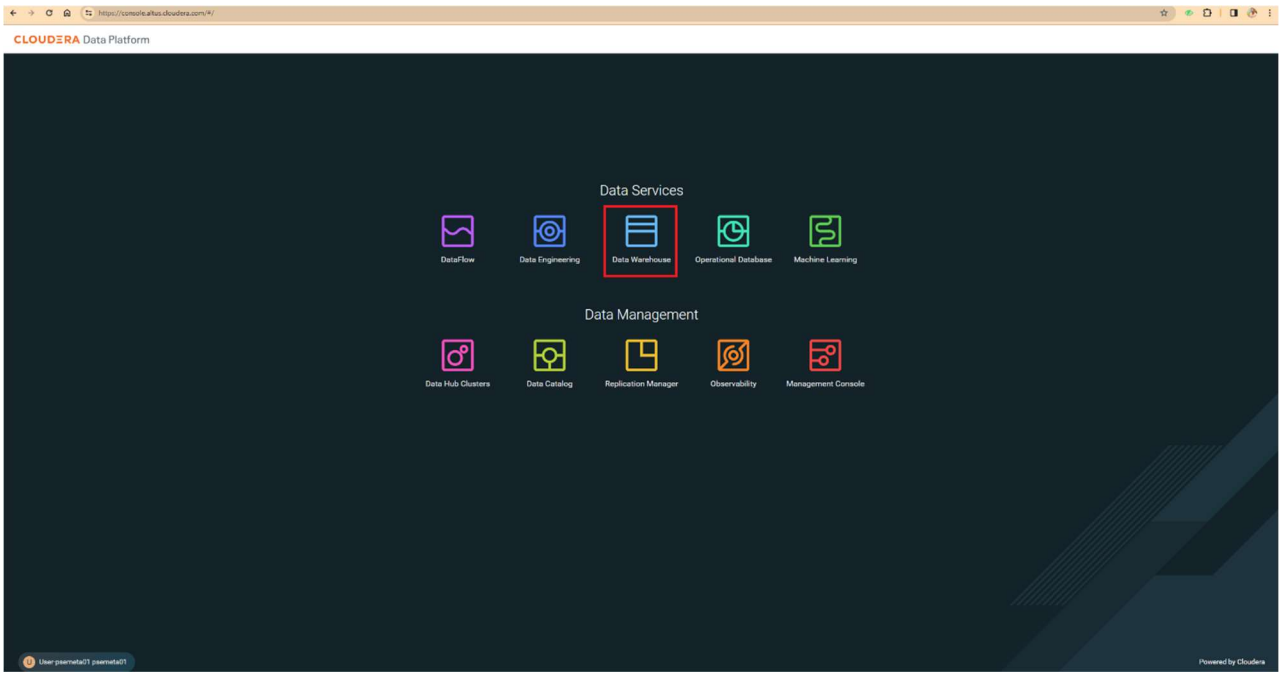
Let’s work on the 1st Part now.

Part 1: Dashboard development

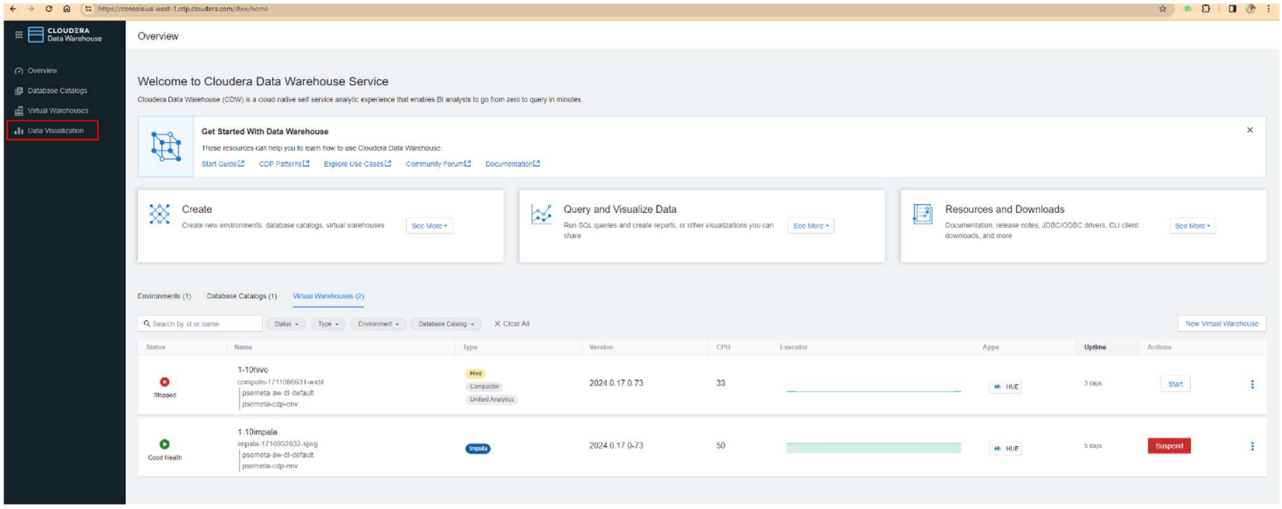
Goals:

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

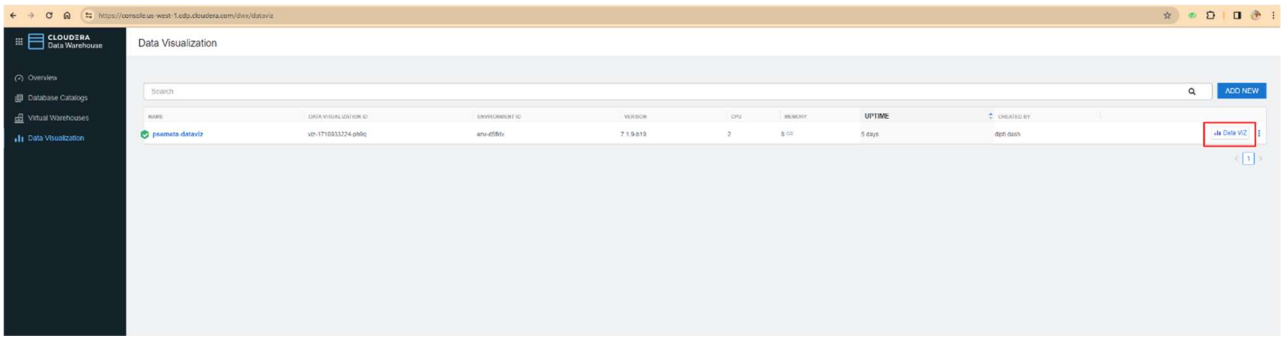
1. Click on Data Warehouse from CDP PC Home.



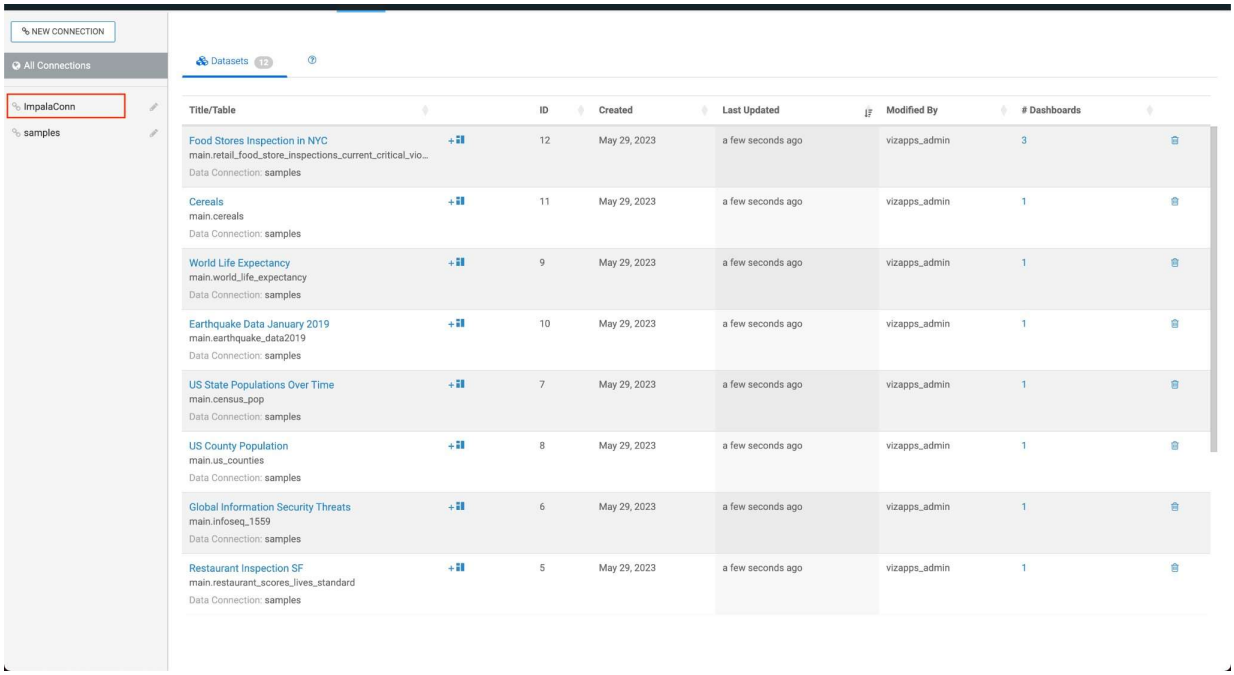
2. Data Warehouse welcome screen. Click on Data Visualization in the left menu.



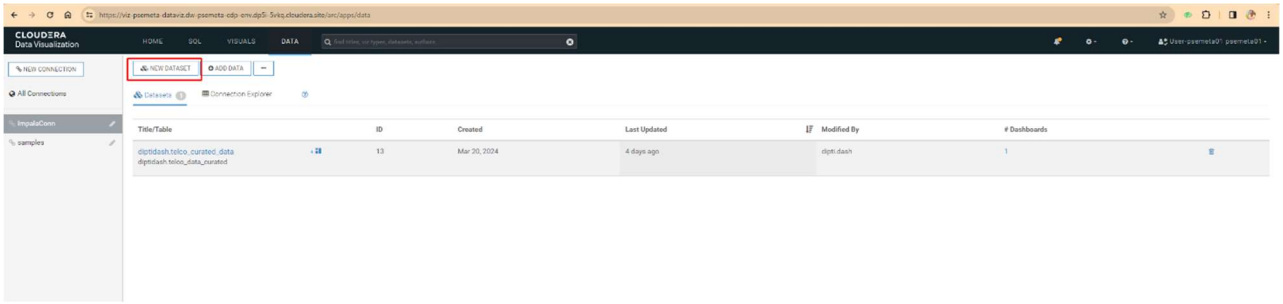
3. In Data Visualization, click on the button **Data Viz** from which they were assigned.



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.

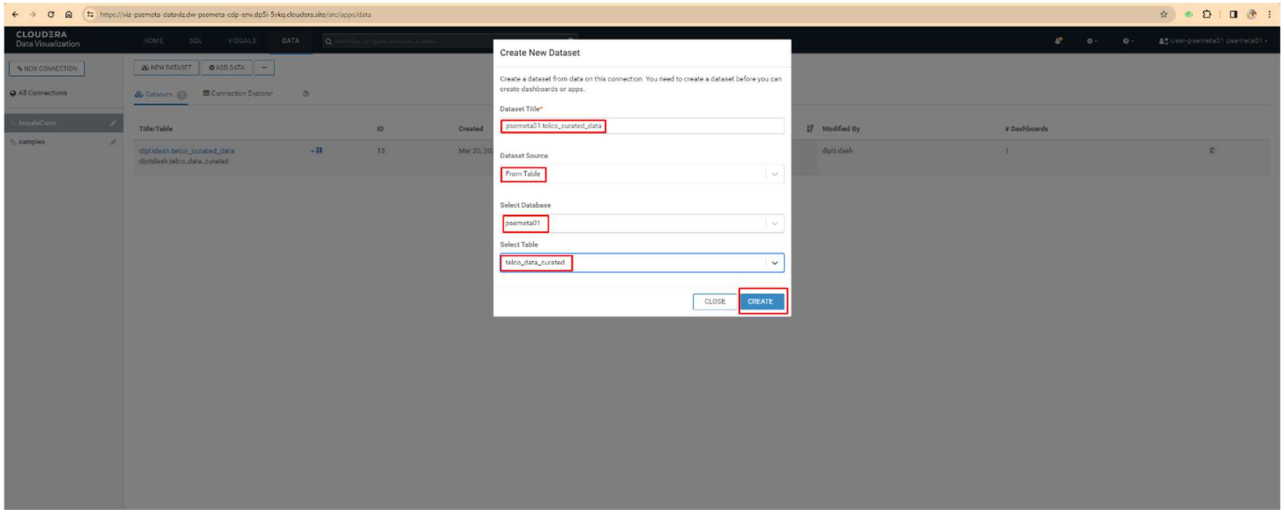


6. Enter the information for the new data source:

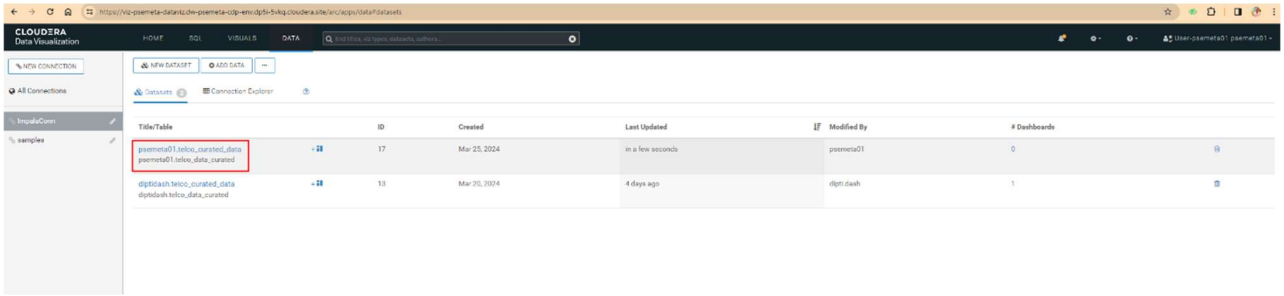
Dataset title: <assigned_user>.telco_curated_data Ex: ***psemeta.telco_curated_data***
Dataset Source: From table

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

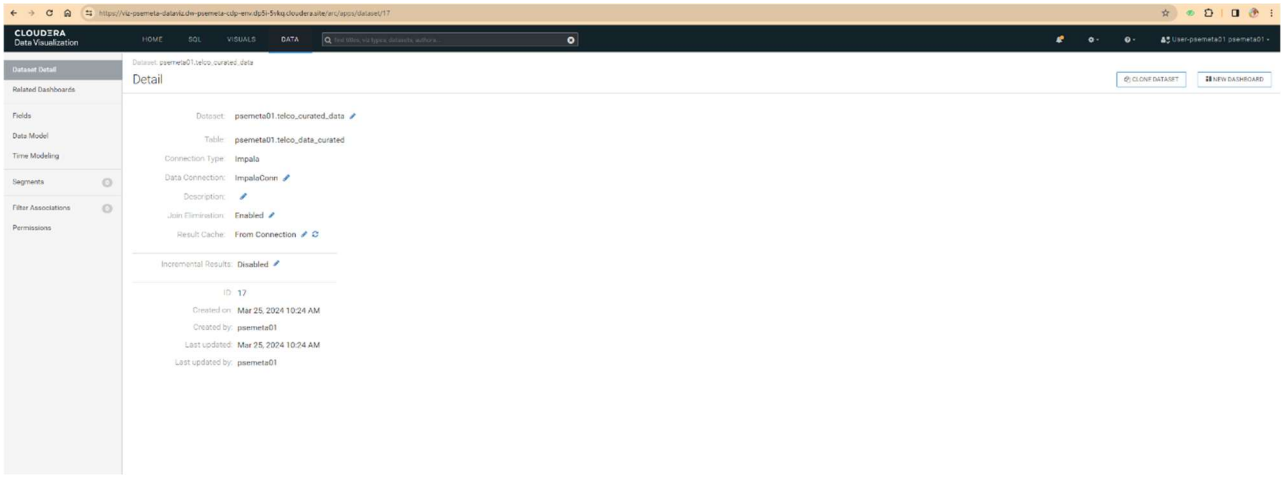
Click on Create to create the new Dataset.



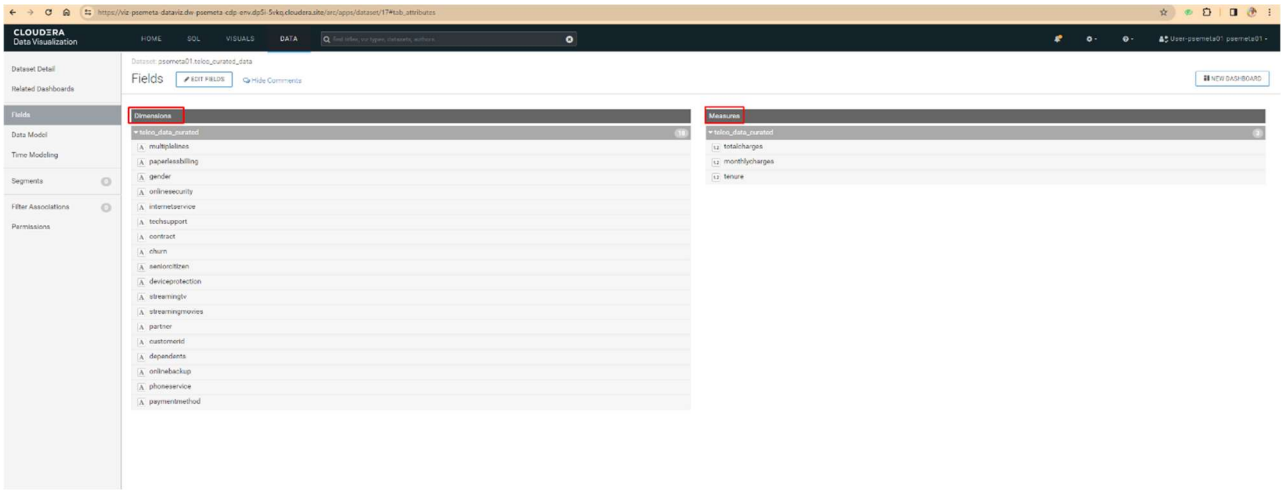
7. The new Dataset should appear in the list. Click on the dataset that you just created.



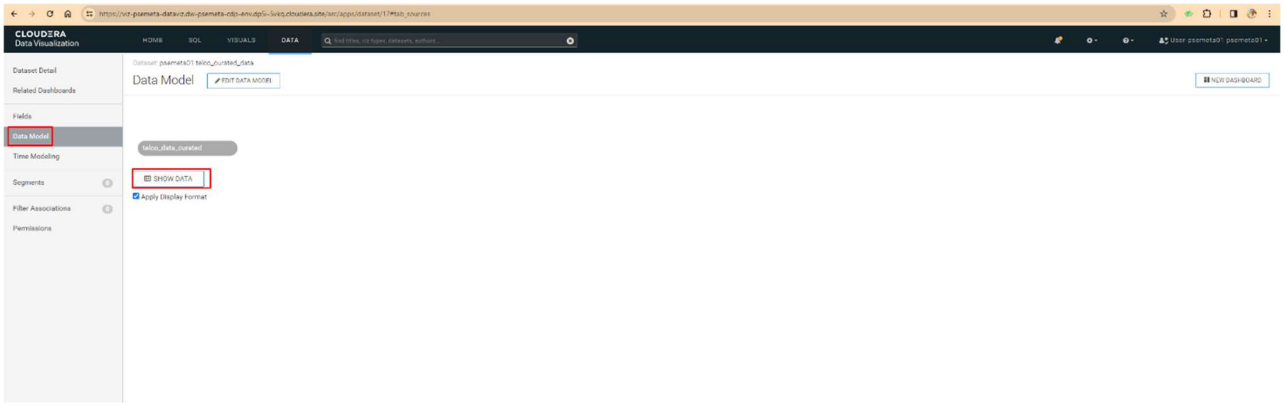
8. Here you will see the details of the dataset.



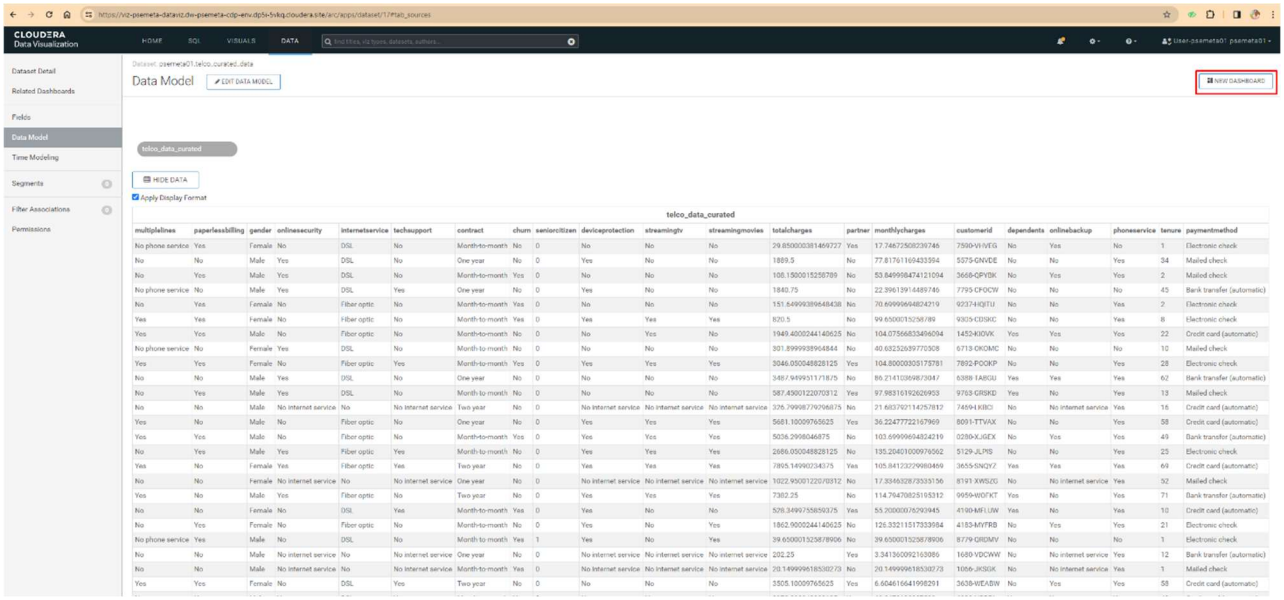
9. Click on **Fields** (left menu) to see the fields automatically captured during the dataset creation process.



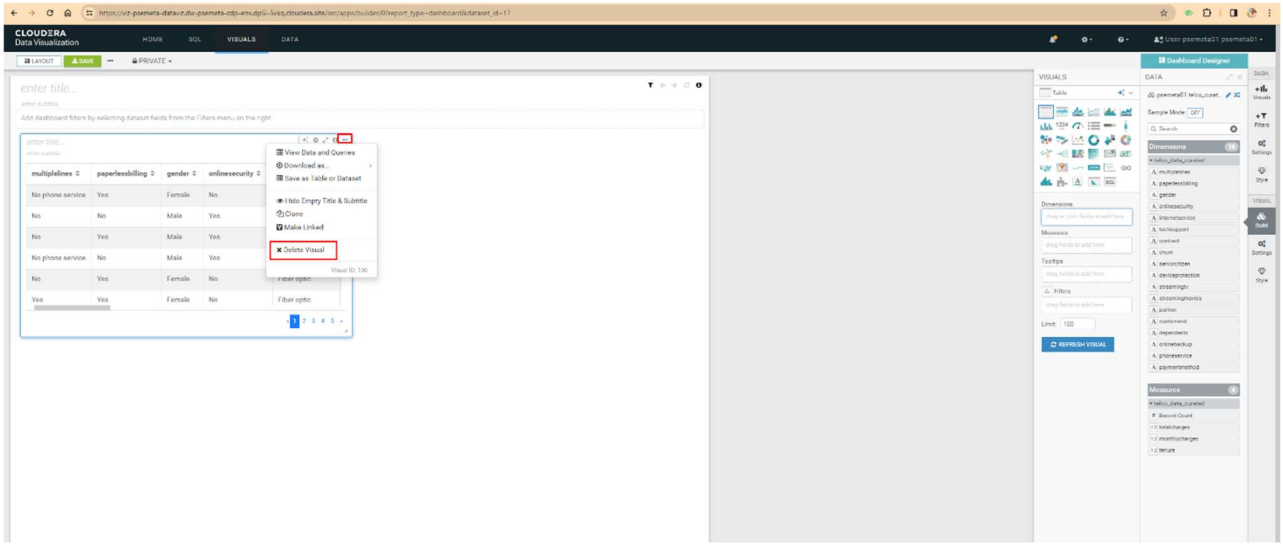
10. You can also preview the data from this screen. Click on **Data Model** (left menu) and then on the button **SHOW DATA** that appears in the center.



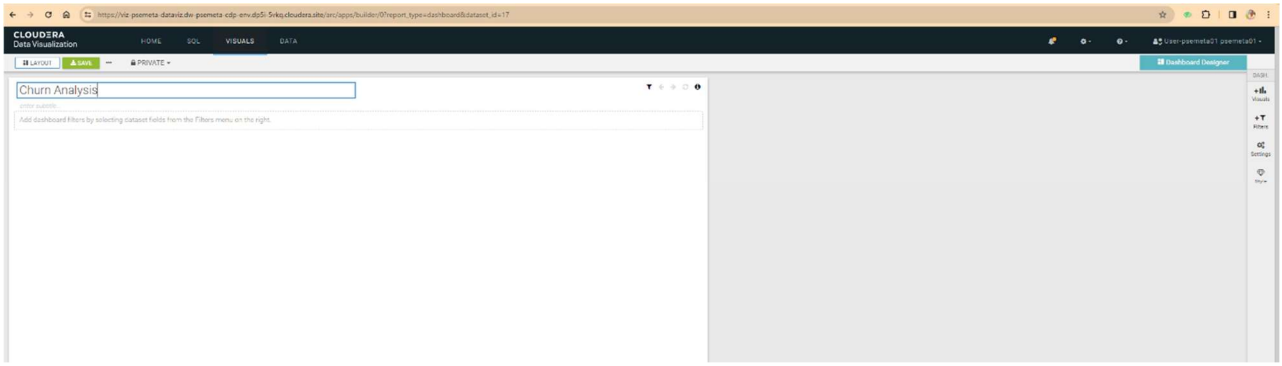
11. At this moment, a query to the Virtual Warehouse is executed to retrieve the data from the data set. Notice the columns and values. Click **NEW DASHBOARD** to create a new dashboard.



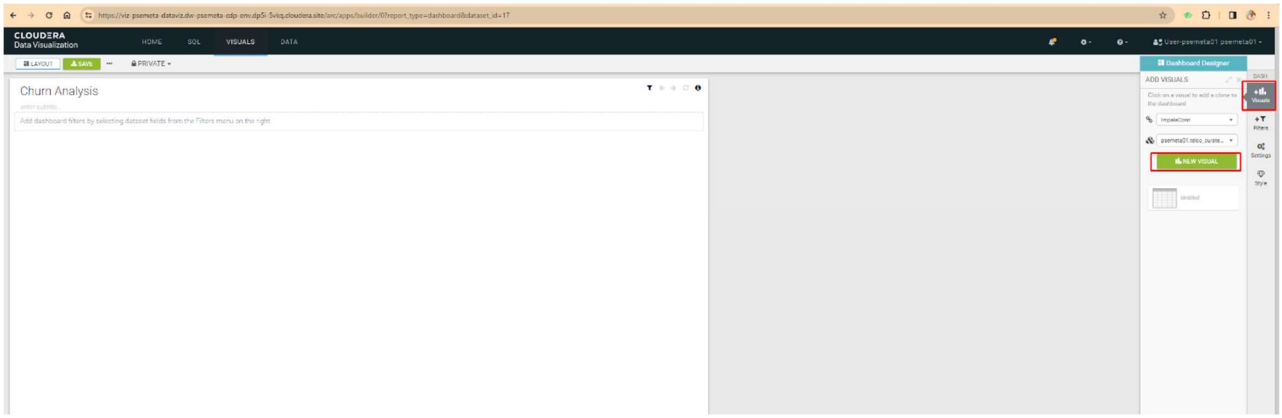
12. When opening the design canvas of a new panel, remove the element that is added by default, by clicking on the three dots (...) button at the top right of the element, and then clicking on the option **Delete Visual**



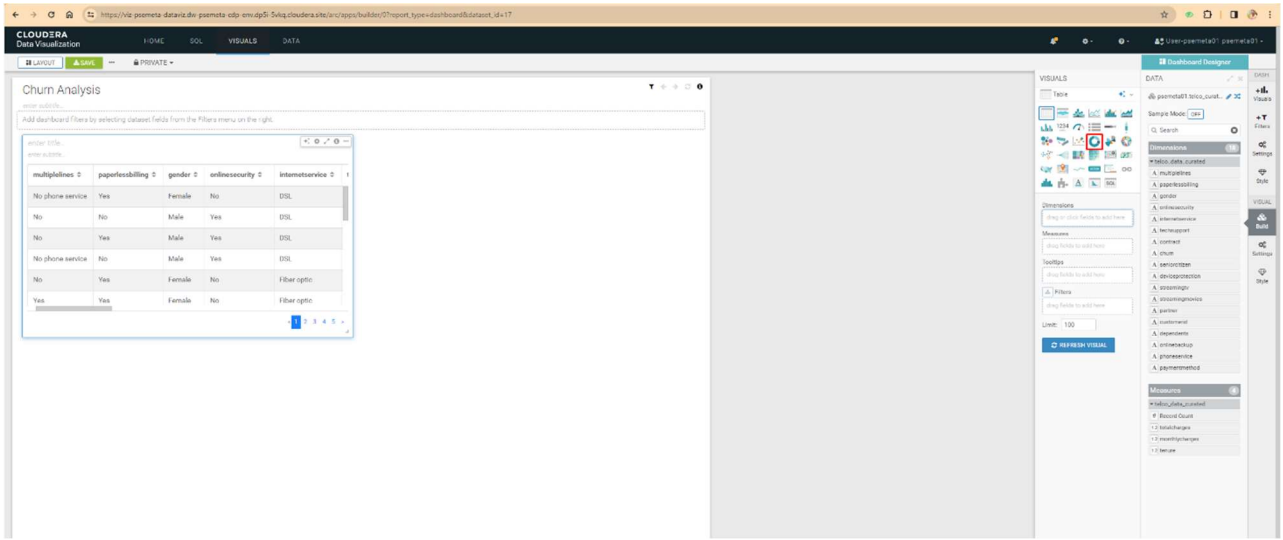
At the top of the canvas, in the enter title field, enter the name ***Churn Analysis*** to identify the dashboard.

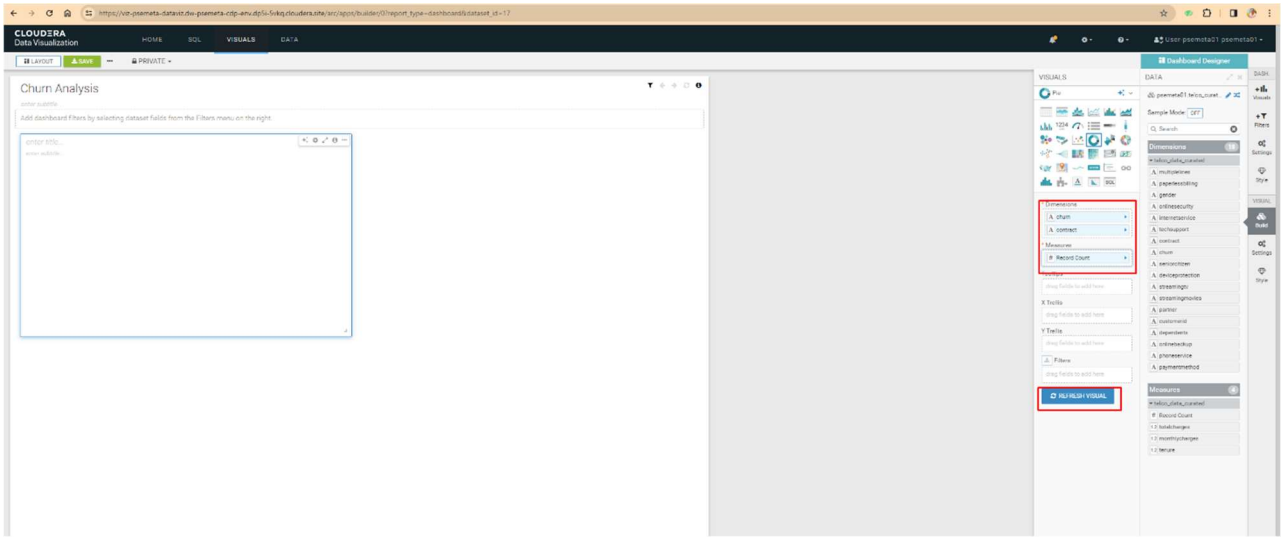


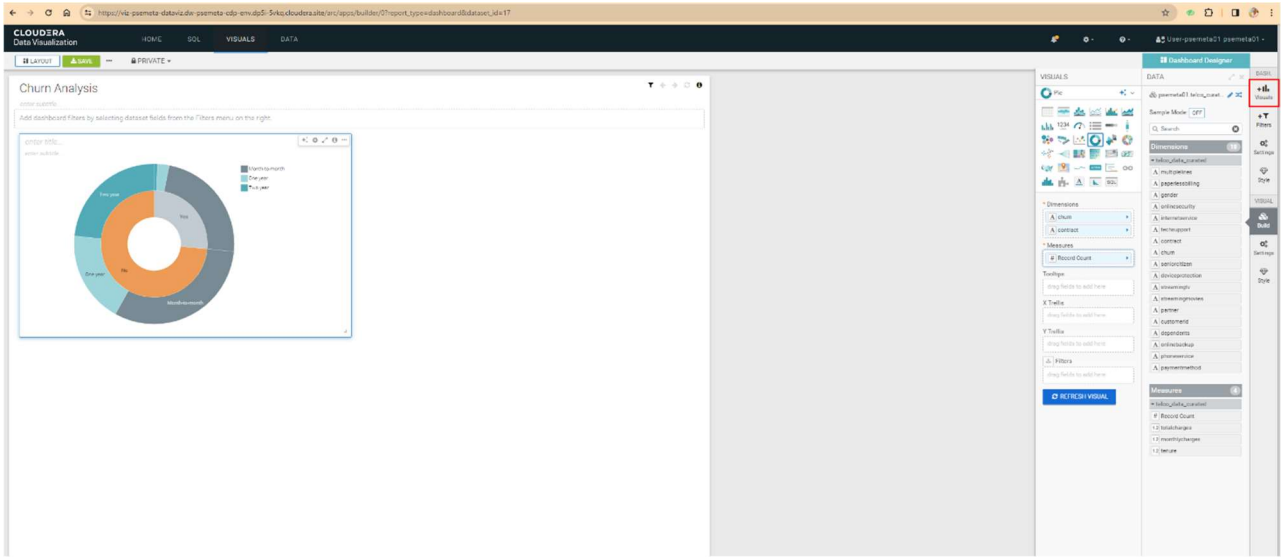
13. To add a new visual element, click on the button **Visuals** from the right menu, select the dataset that corresponds to them, and click on the button **New Visual.**



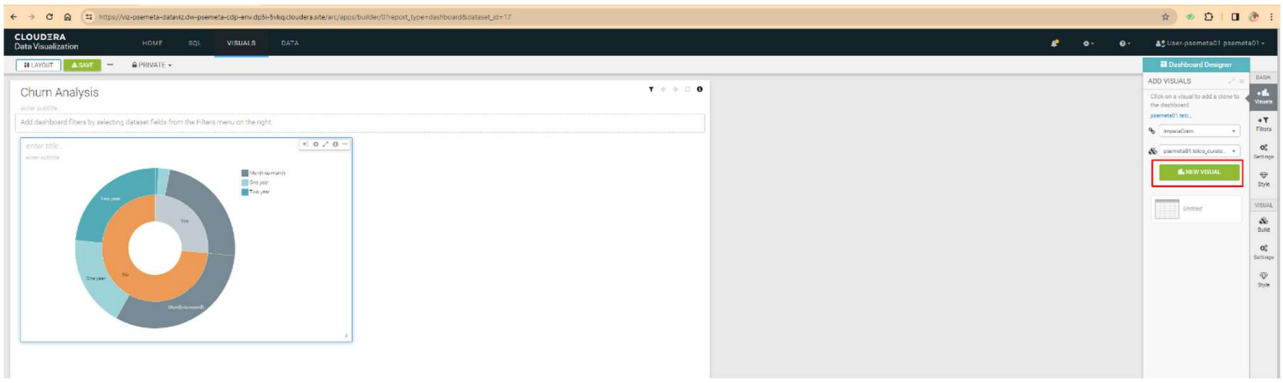
14. Add the first visual element, which is a pie chart with the dimensions **churn** and **contract**, with the metric of **Record count**. Once finished, click the button **Refresh Visual**.

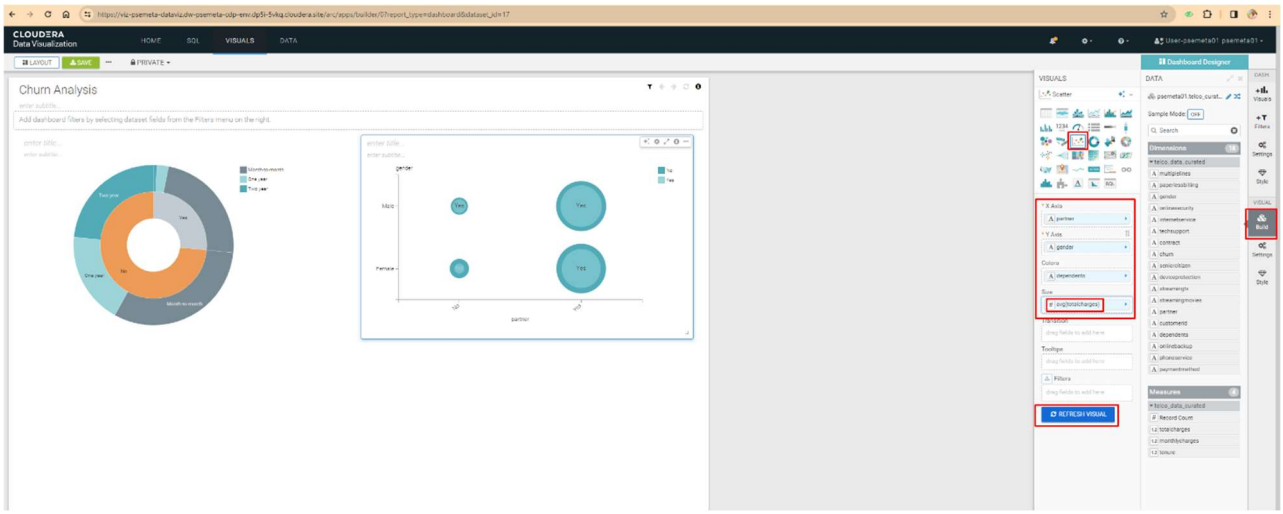



CLOUDERA
Data Visualization
HOME
SQL
VISUALS
DATA
Churn Analysis
Add dashboard filters by selecting dataset fields from the Filters menu on the right.
enter title...
Dashboard Designer
VISUALS
DATA
Dimensions
churn
contract
Measures
Record Count
Tooltips
X Trellis
Y Trellis
Filters
REFRESH VISUAL

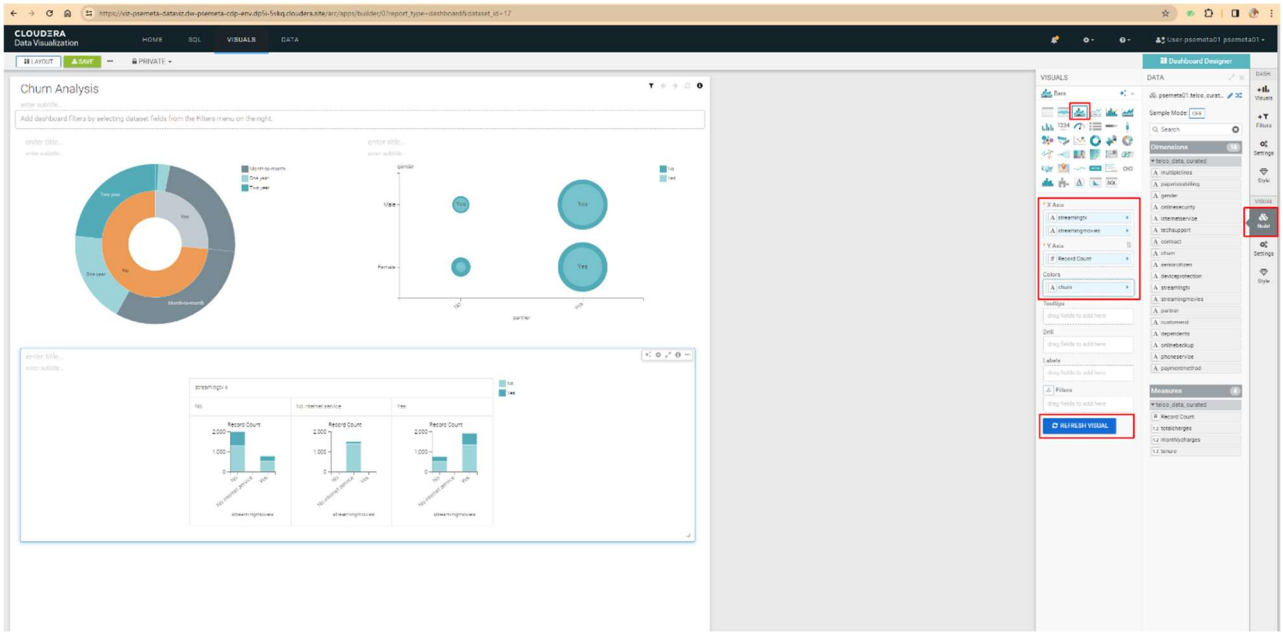

CLOUDERA
Data Visualization
HOME
SQL
VISUALS
DATA
Churn Analysis
Add dashboard filters by selecting dataset fields from the Filters menu on the right.
enter title...
Month-to-month
One year
Two year
Yes
No
Dashboard Designer
VISUALS
DATA
Dimensions
churn
contract
Measures
Record Count
Tooltips
X Trellis
Y Trellis
Filters
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**.

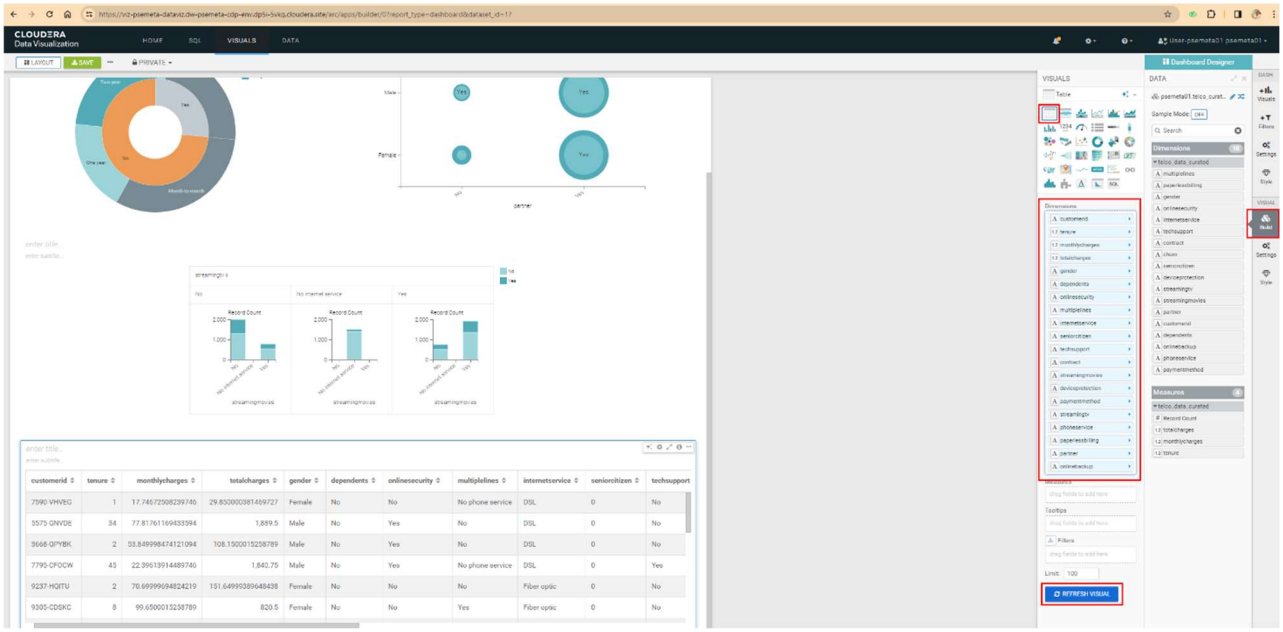




15. Add the third visual element, which is a bar chart with the dimensions **streamingtv** and **streamingmovies** like X Axis,
Record Count how Y Axis and **churn** how Colors. Once finished, click the button **Refresh Visual**.

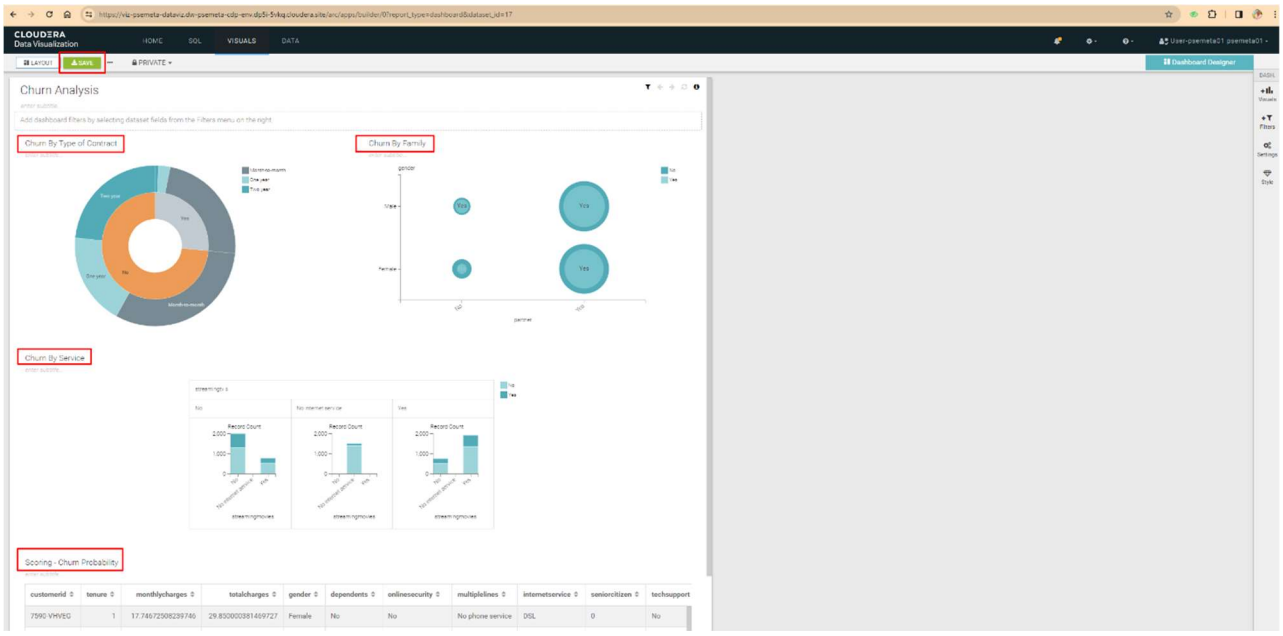


16. Add the fourth and last visual element, which is a table with the dimensions and metrics of the dataset. Be sure to add all 17 dimensions and 3 metrics to the table, so in total 20 elements. Once finished, click the button **Refresh Visual**.



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



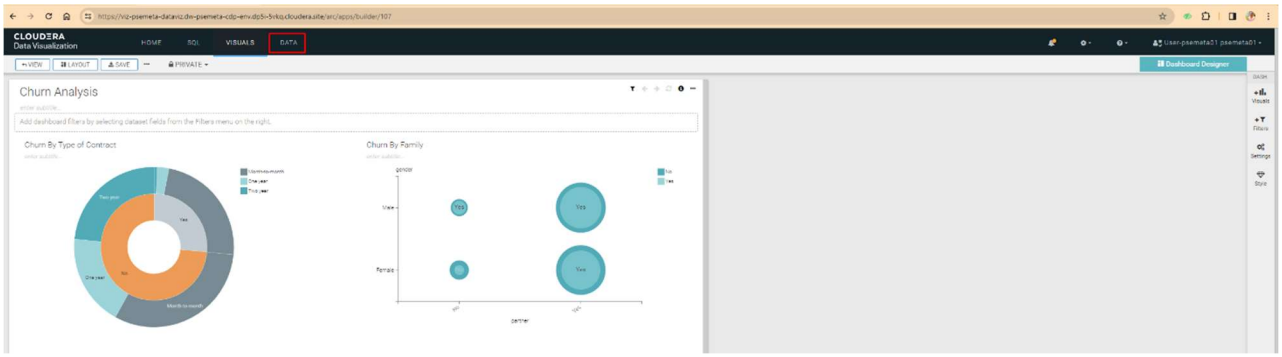
At this point we will go ahead and complete the lab on CML and then come back to the next section.

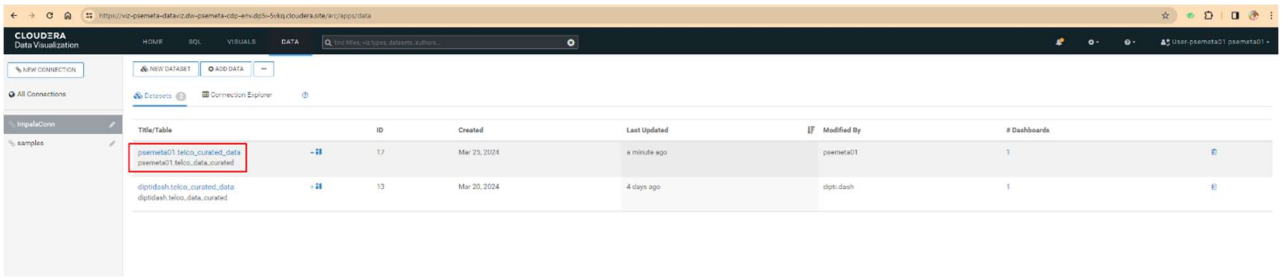
Part 2: Add New field

Goals:

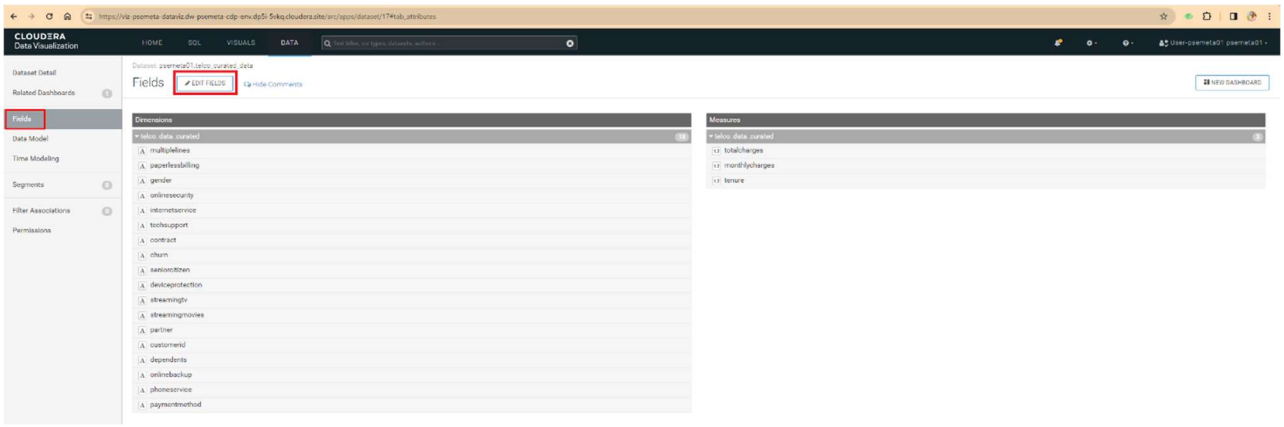
- Add a new field that makes calls to the ML model
- Add the new field to the dashboard

1. Edit the previously created Dataset, in Data -> <user_assigned>.telco_data_curated.

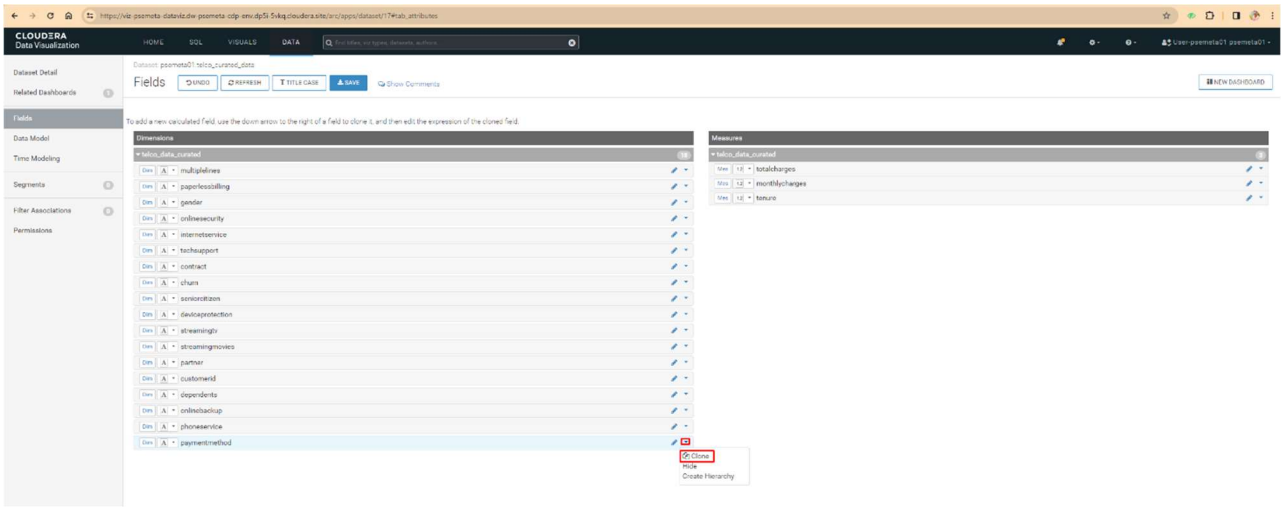




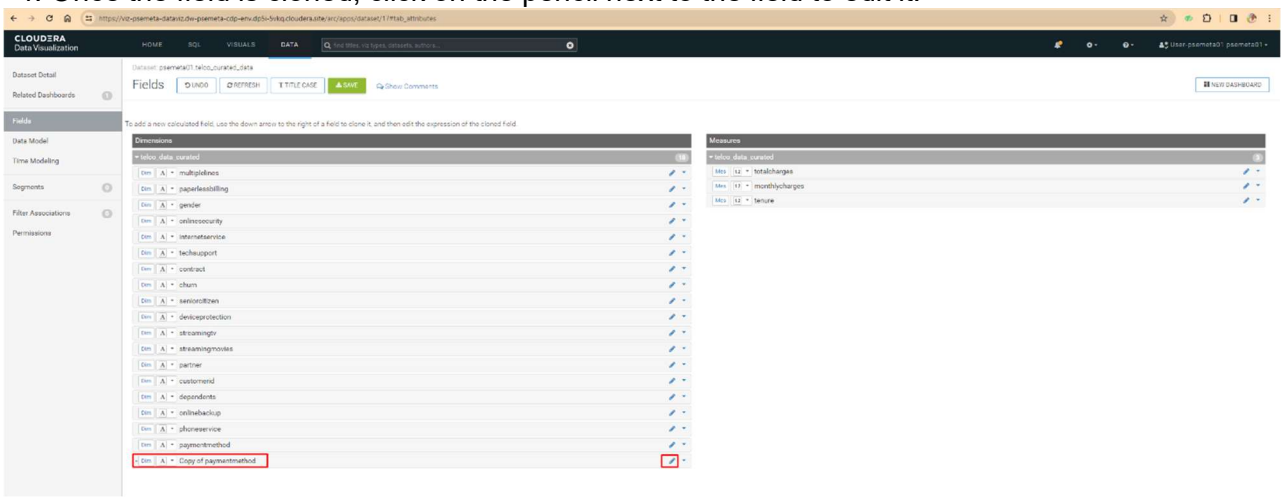
2. Once in the Dataset, go to **Fields** in the left menu and then click on **Edit Field** to edit the fields of your dataset.



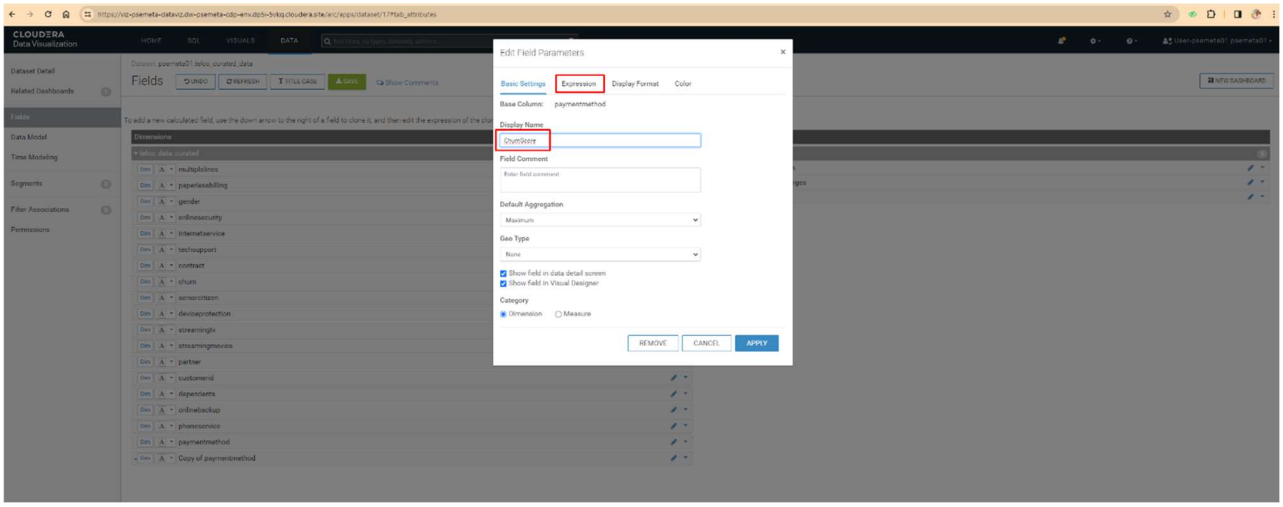
3. In the list of **Dimensions**, click the down arrow of the last field in the list, and select the option **Clone**.



4. Once the field is cloned, click on the pencil next to the field to edit it.



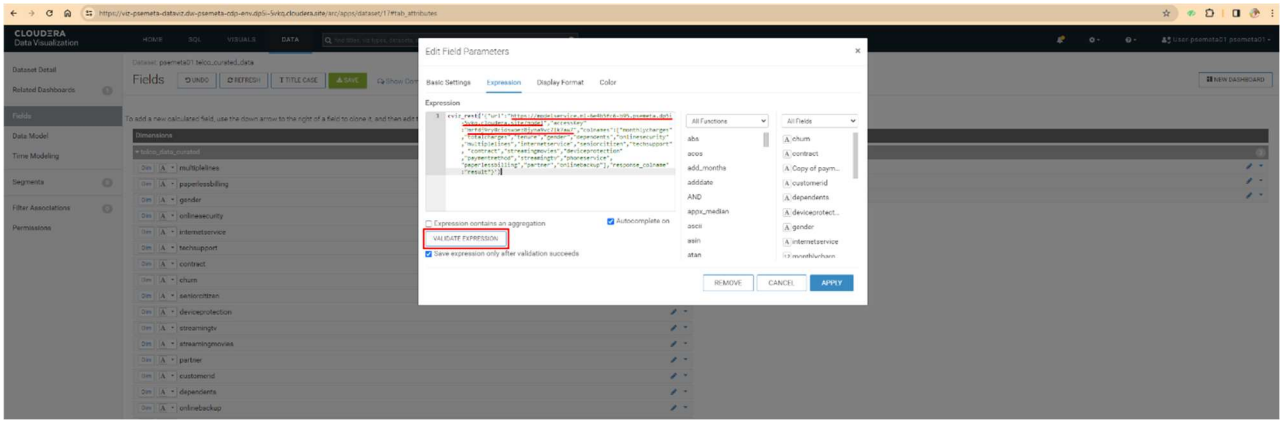
5. In the popup window that appears, enter the name of the new field in **Display Name**. We suggest that you enter *ChurnScore.*



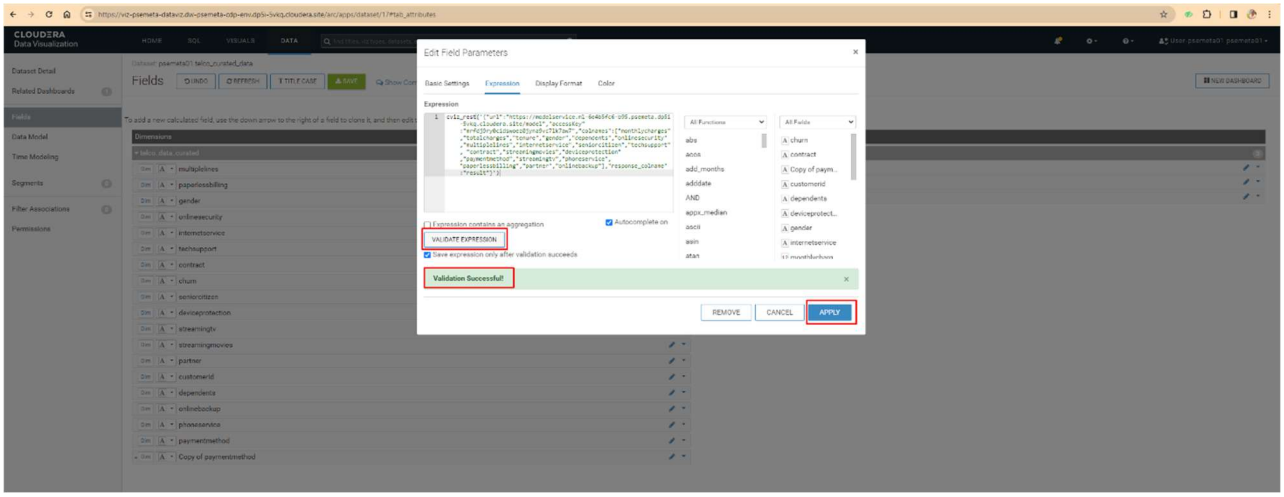
6. Go to the Expressions tab and enter the following value in the Expression field. This will allow you to call the REST API of the Model you have previously deployed.

Note: The url and accessKey is something that you would have noted down as the last step in the CML lab.

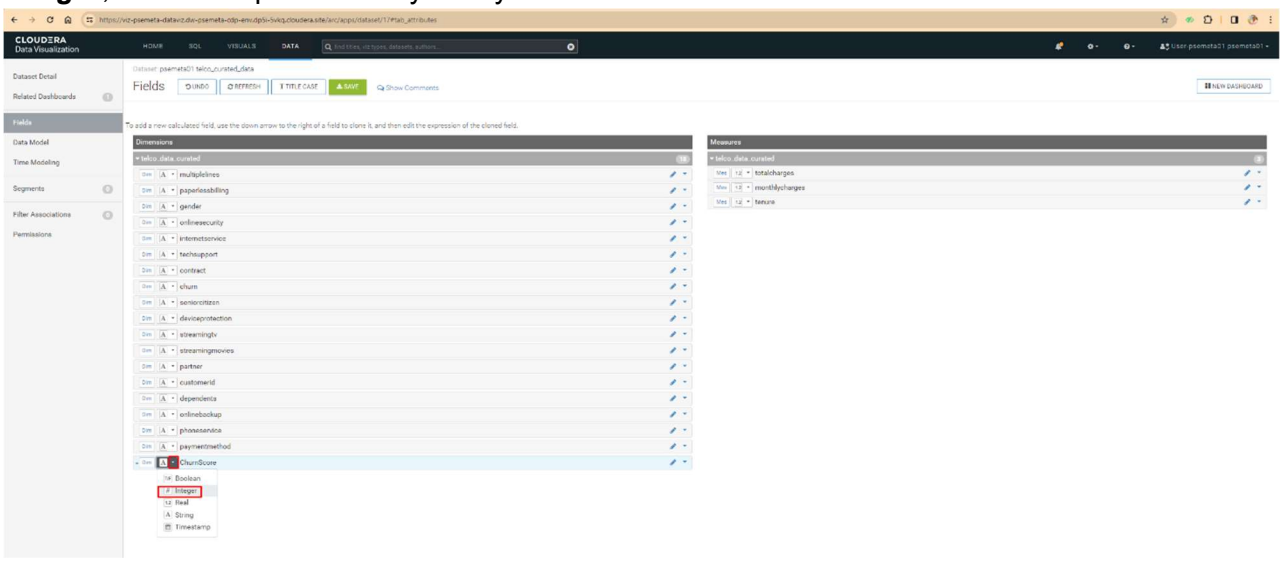
cviz_rest('{"url":"**https://modelservice.ml-6e4b5fc6-b95.psemeta.dp5i-5vkq.cloudera.site/model**","accessKey":"**miqnv6bxiwontz84ozoxfyc5rt6p6grb**","colnames":["monthlycharges","totalcharges","tenure","gender","dependents","onlinesecurity","multiplelines","internetservice","seniorcitizen","techsupport",
"contract","streamingmovies","deviceprotection","paymentmethod","streamingtv","phoneservice",
"paperlessbilling","partner","onlinebackup"],"response_colname":"result"}')



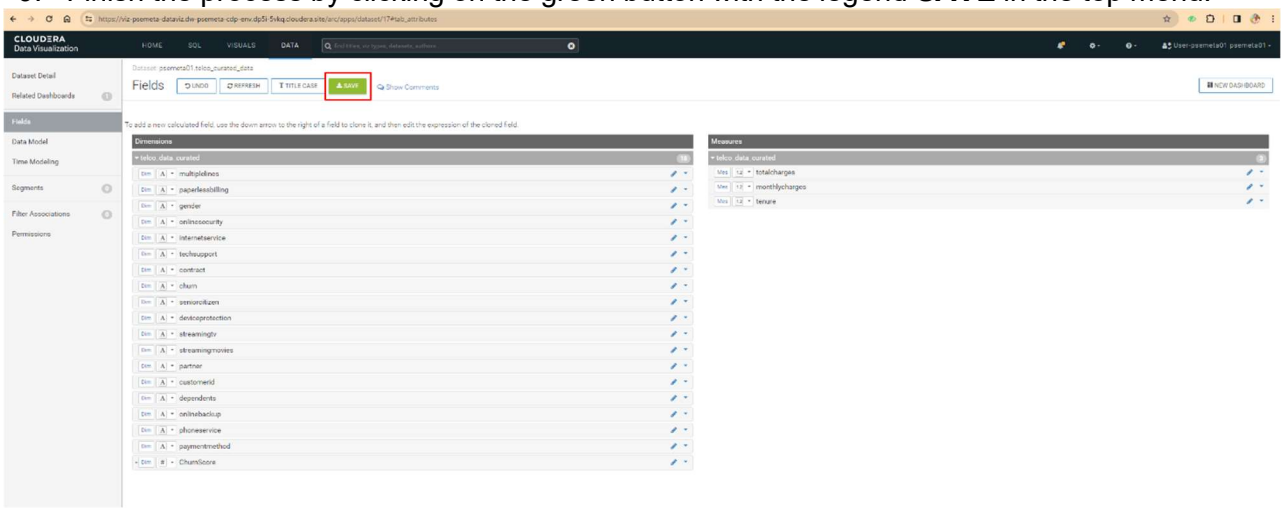
7. Finish the process of copying the ***url*** and the ***accessKey***, click the **Validate Expression** button at the top of the window. If the message appears in green ***Validation Successful***, Click on **Apply** to save the settings made.



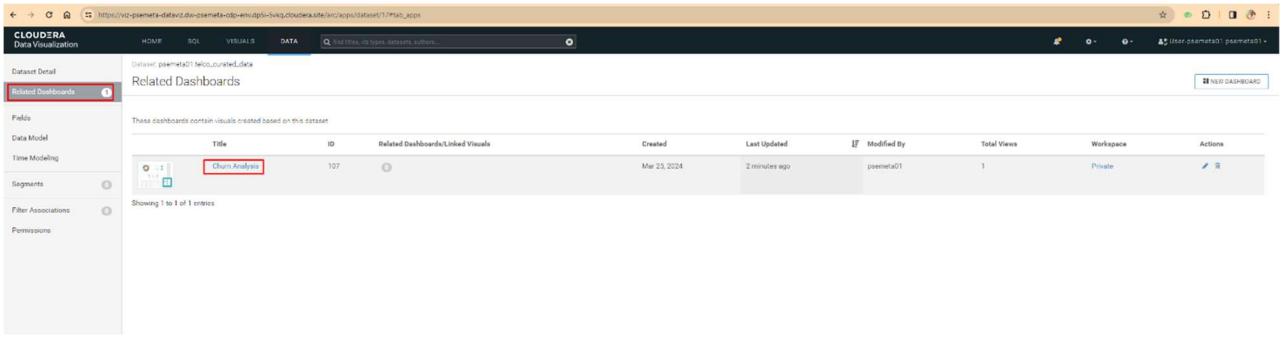
8. The new field should appear in the list of fields. Change the data type, selecting the type ***Integer***, which is represented by the symbol **#**



9. Finish the process by clicking on the green button with the legend **SAVE** in the top menu.

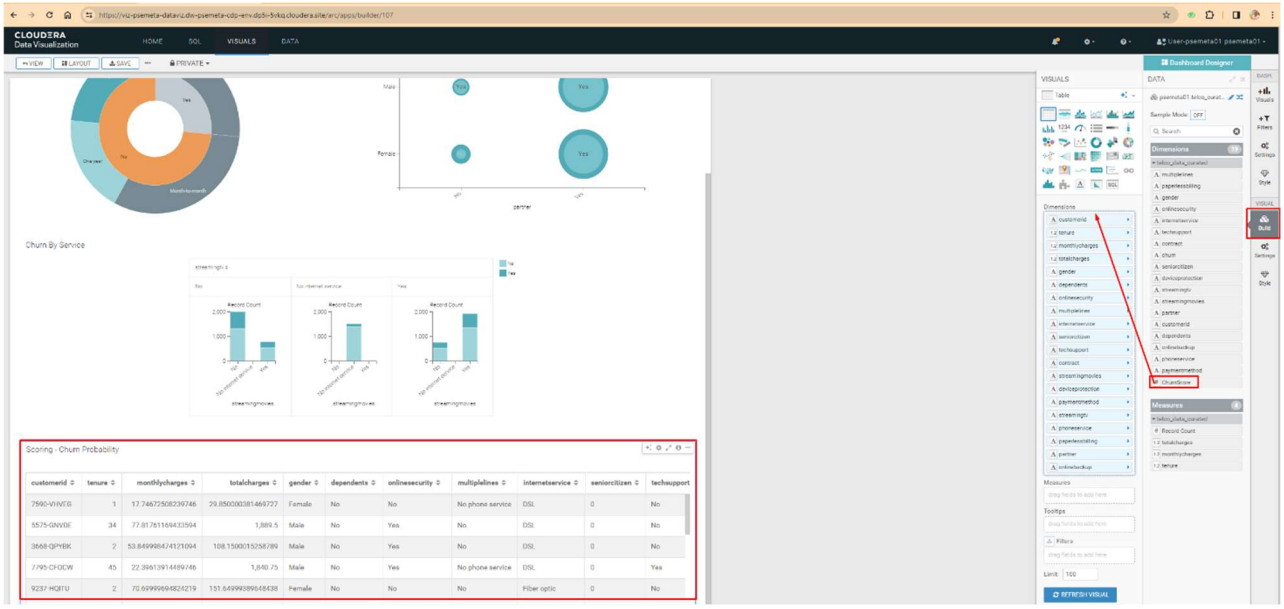


10. Return to the dashboard, select the option **Related Dashboards** from the top left menu, and clicking on the name of the dashboard (**Churn Analysis**) that was previously created.

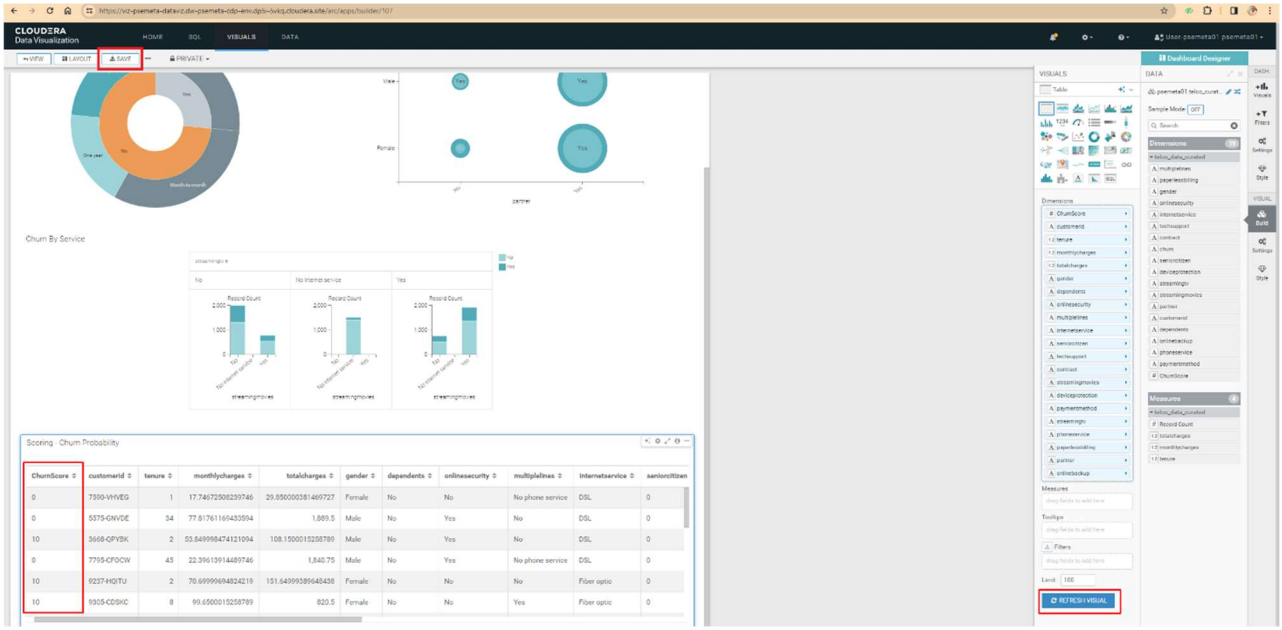


11. Once in the dashboard, click on the button **Edit** which is in the upper left.

12. Edit the lower table by clicking on it and then on the option **Build** from the right vertical menu. Add the new field, **ChurnScore**, at the beginning of the table, by clicking and dragging from the option **Dimensions** available.



13. Click on the Refresh Visual button to update the data. The new column should appear *ChurnScore* then at the beginning of the table, with a value of numeric type. Finish the process by clicking the button **SAVE** from the top left menu.


CLOUDERA
Data Visualization
HOME
SQL
VISUALS
DATA
SAVE
Churn By Service
Scoring - Churn Probability
ChurnScore
customerid
tenure
monthlycharges
totalcharges
gender
dependents
onlinesecurity
multiplelines
internetservice
seniorcitizen
7590-VHVEG
5575-GNVDE
3668-QPYBK
7795-CFOCW
9237-HQITU
9305-CDSKC
DSL
Fiber optic
No phone service
Dashboard Designer
REFRESH VISUAL