



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 2<sup>nd</sup> Part focusses on integration between Data Viz and Cloudera Machine Learning.

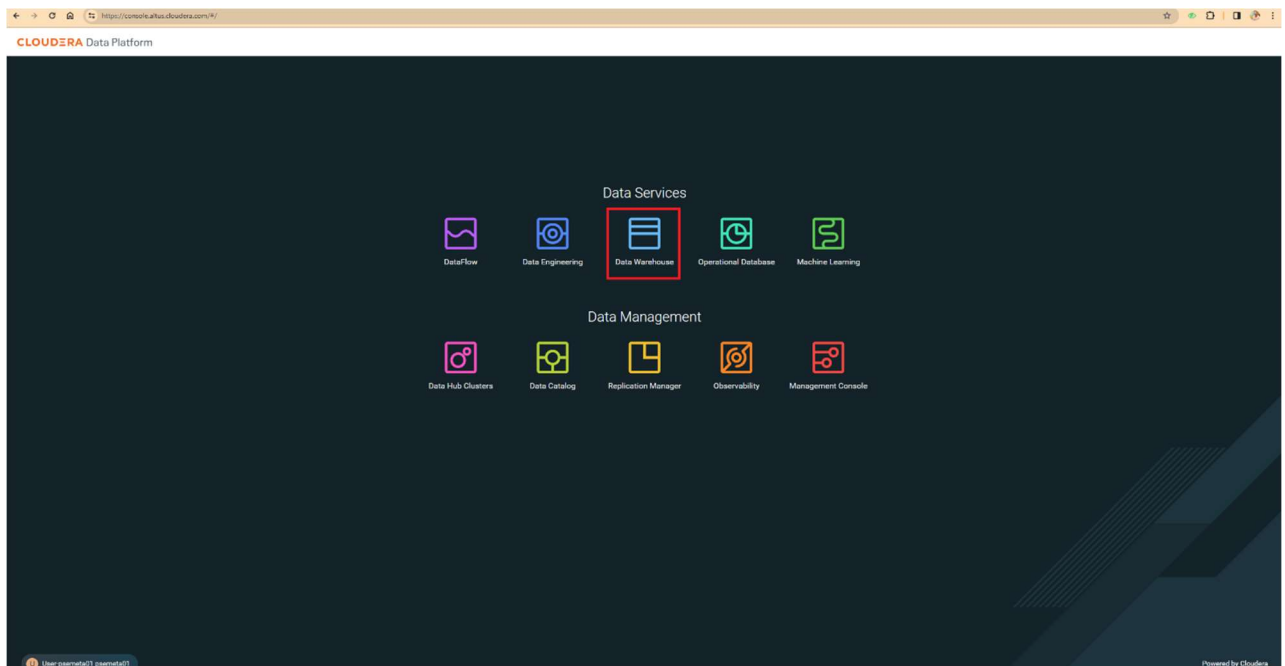
Let's work on the 1<sup>st</sup> Part now.

## Part 1: Dashboard development

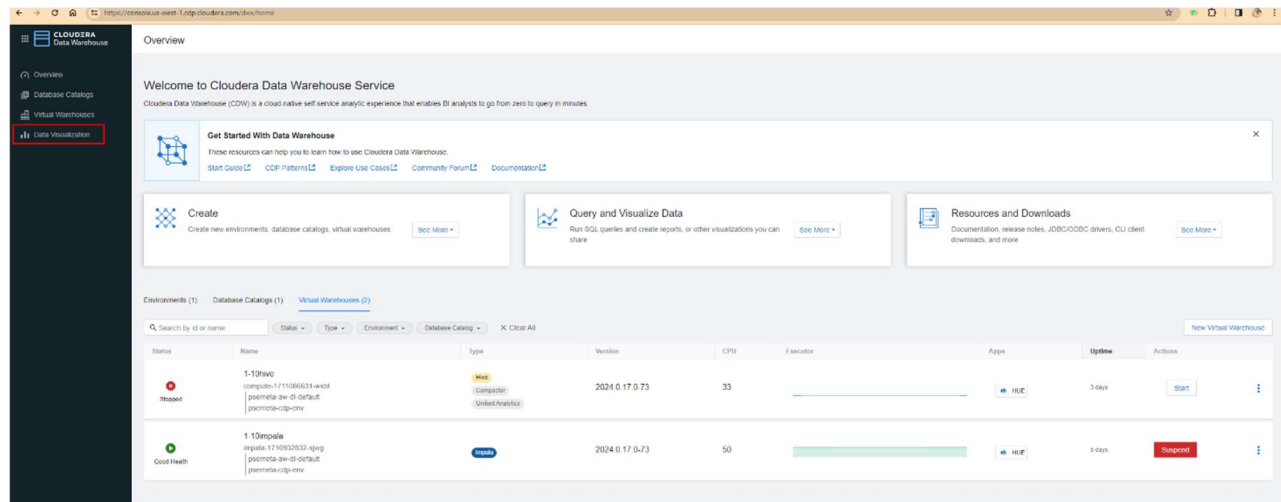
Goals:

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

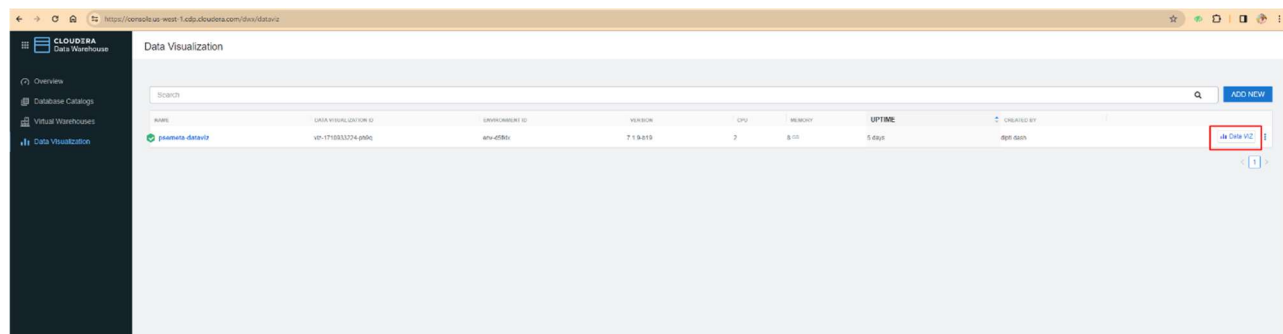
1. Click on Data Warehouse from CDP PC Home.



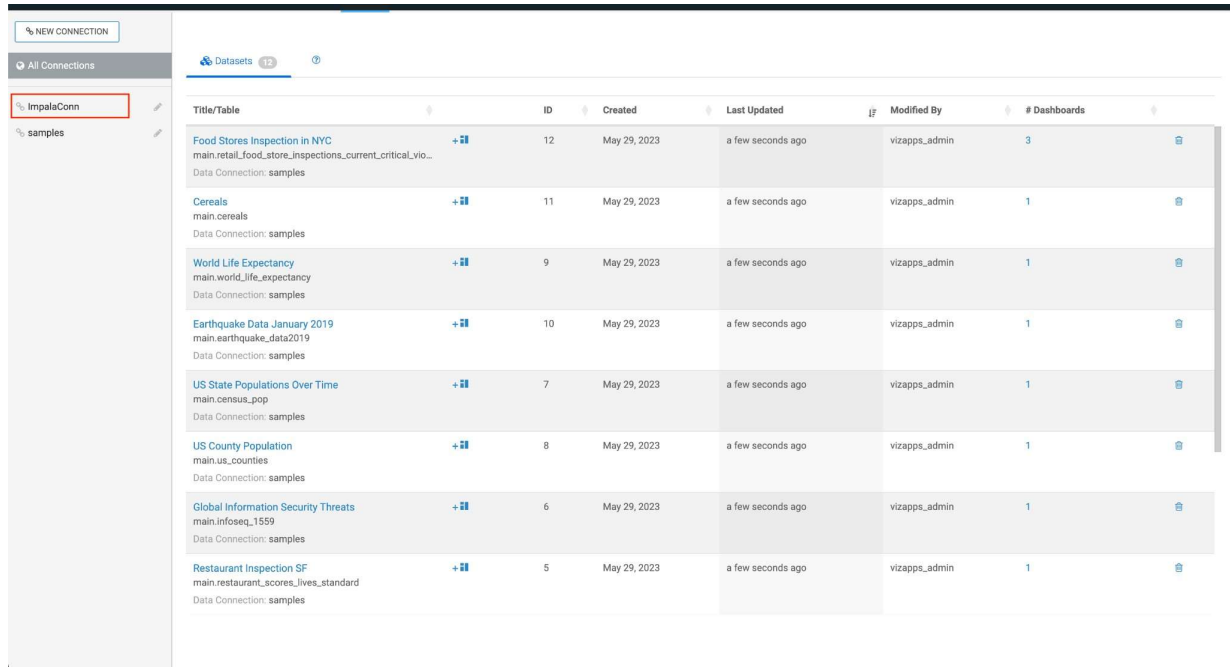
2. Data Warehouse welcome screen. Click on Data Visualization in the left menu.



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

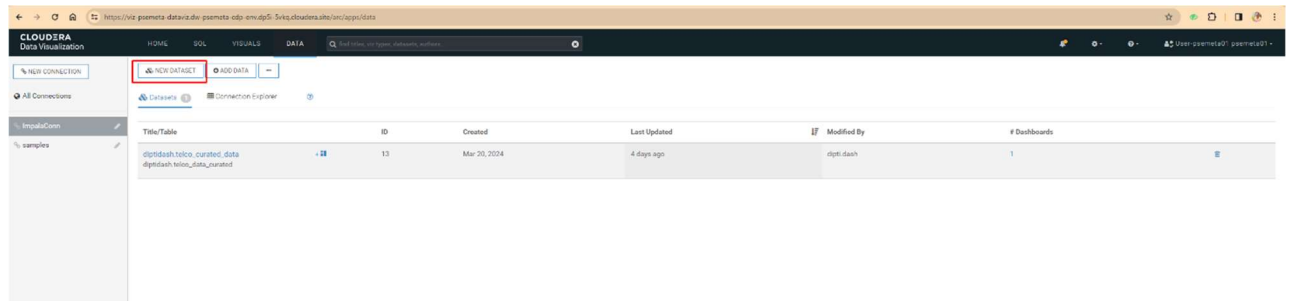


4. Once in Data Visualization, go to the Data option from the top menu, and then to the Connector **ImpalaConn** from the left menu. Or, chose the **connection** that is assigned to you (check with instructor on this).



Title/Table	ID	Created	Last Updated	Modified By	# Dashboards
Food Stores Inspection in NYC main.retail_food_store_inspections_current_critical_vio...	12	May 29, 2023	a few seconds ago	vizapps_admin	3
Cereals main.cereals	11	May 29, 2023	a few seconds ago	vizapps_admin	1
World Life Expectancy main.world_life_expectancy	9	May 29, 2023	a few seconds ago	vizapps_admin	1
Earthquake Data January 2019 main.earthquake_data2019	10	May 29, 2023	a few seconds ago	vizapps_admin	1
US State Populations Over Time main.census_pop	7	May 29, 2023	a few seconds ago	vizapps_admin	1
US County Population main.us_counties	8	May 29, 2023	a few seconds ago	vizapps_admin	1
Global Information Security Threats main.infosec_1559	6	May 29, 2023	a few seconds ago	vizapps_admin	1
Restaurant Inspection SF main.restaurant_scores_lives_standard	5	May 29, 2023	a few seconds ago	vizapps_admin	1

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.



Title/Table	ID	Created	Last Updated	Modified By	# Dashboards
dpdash.telco_curated_data dpdash.telco_data_curated	13	Mar 30, 2024	4 days ago	dpdash	1

6. Enter the information for the new data source:

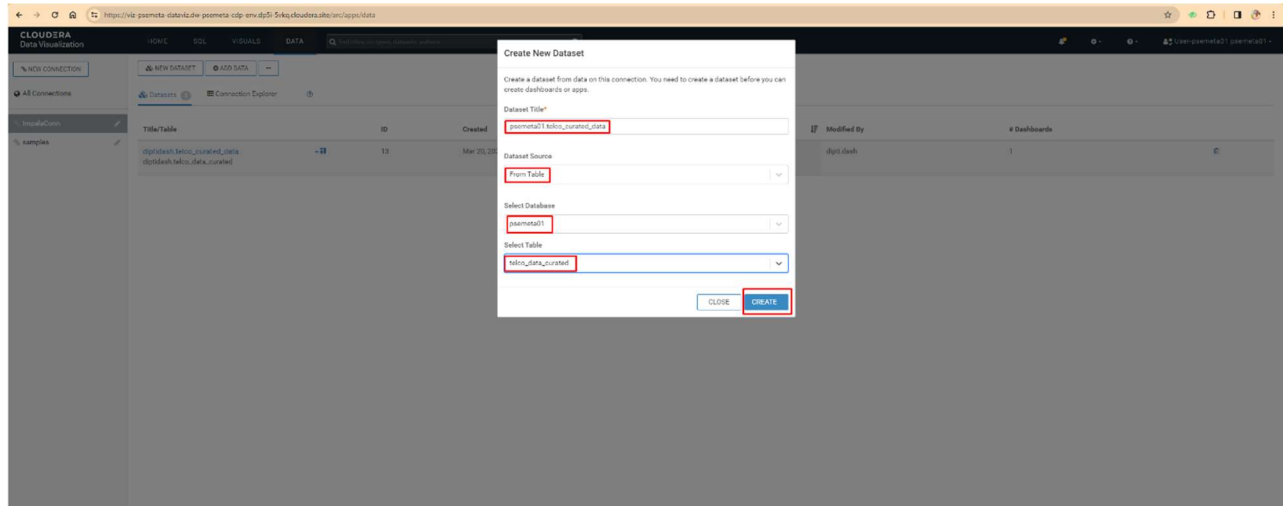
**Dataset title:** <assigned\_user>.telco\_curated\_data Ex: **psemeta01.telco\_curated\_data**

**Dataset Source:** From table

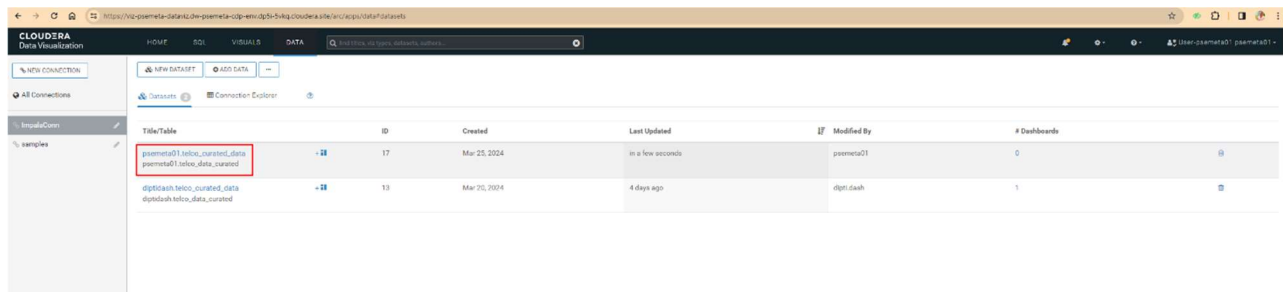
**Select Database:** <assigned\_user> Ex: **psemeta**

**Select Table:** telco\_data\_curated

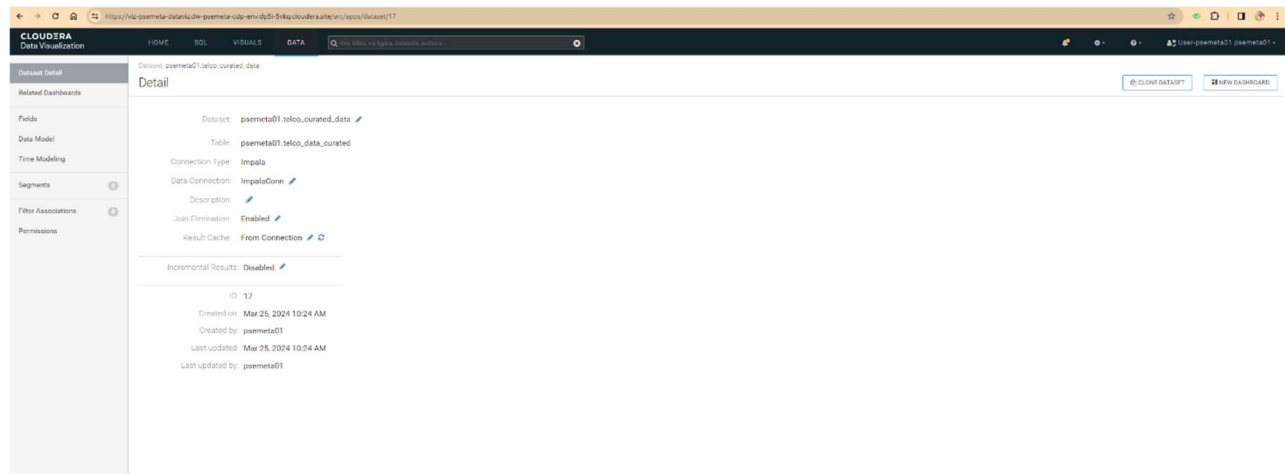
Click on Create to create the new Dataset.



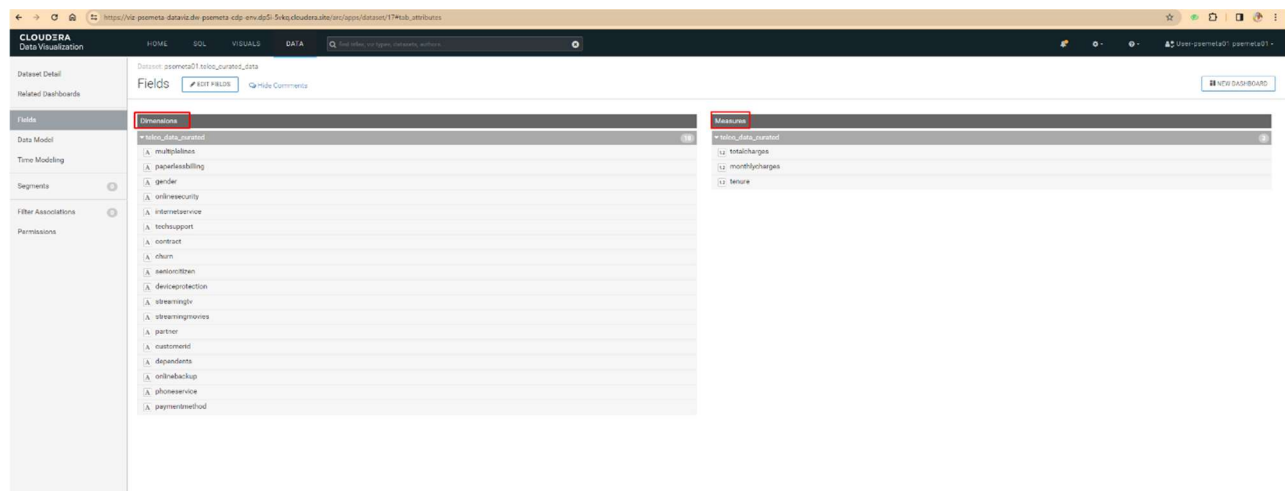
7. The new Dataset should appear in the list. Click on the dataset that you just created.



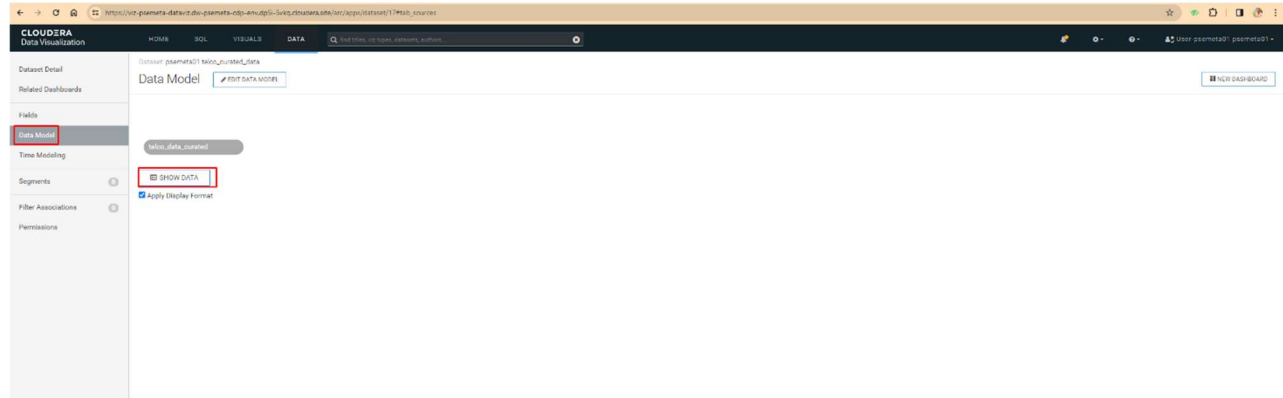
8. Here you will see the details of the dataset.



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.

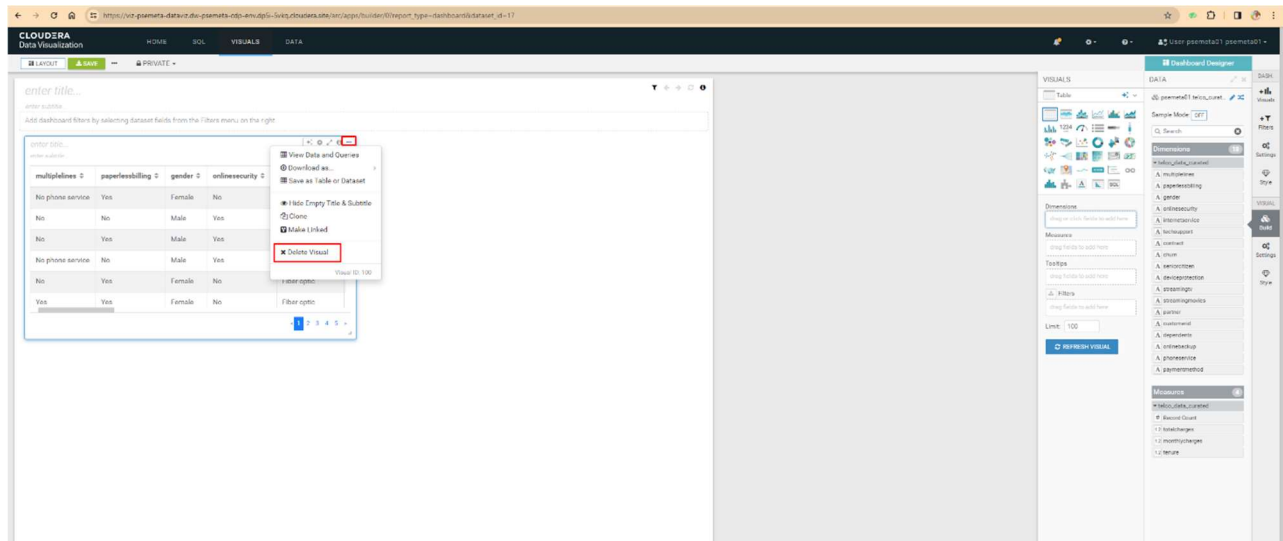


11. At this moment, a query to the Virtual Warehouse is executed to retrieve the data from the data set. Notice the columns and values. Click **NEW DASHBOARD** to create a new dashboard.

The screenshot shows the CloudPaaS Data Visualization interface with a table of data. The table has 25 columns and 25 rows. The columns are: multipleline, paperlessbilling, gender, onlinesecurity, internetaccess, techsupport, contract, churn, servicecitizen, deviceprotection, streamingtv, streamingmovies, totalshapes, partner, monthlycharges, customerid, dependents, onlinebackup, phoneservice, tenure, and paymentmethod. The data is displayed in a table format. A 'NEW DASHBOARD' button is highlighted in the top right corner.

multipleline	paperlessbilling	gender	onlinesecurity	internetaccess	techsupport	contract	churn	servicecitizen	deviceprotection	streamingtv	streamingmovies	totalshapes	partner	monthlycharges	customerid	dependents	onlinebackup	phoneservice	tenure	paymentmethod
No	phone service	Yes	Female	No	DSL	No	One year	No	0	Yes	No	29.850003281409777	Yes	17.2462502299348	7859-VHVS	No	Yes	No	1	Electronic check
No	No	Male	Yes	DSL	No	One year	No	0	Yes	No	No	1899.5	No	77.81761105423254	5275-ONVCE	No	No	Yes	34	Mailed check
No	Yes	Male	Yes	DSL	No	Month-to-month	Yes	0	No	No	No	102.1900015256789	No	53.84999847421094	3658-OPDK	No	Yes	Yes	2	Mailed check
No	phone service	No	Female	Yes	DSL	Yes	One year	No	0	Yes	No	1842.75	No	22.39613914485746	7793-CPQCV	No	No	No	43	Bank transfer (automatic)
No	Yes	Female	No	Fiber optic	No	Month-to-month	Yes	0	No	No	No	151.8499939648438	No	70.89999668242719	92374Q2TJ	No	No	Yes	2	Electronic check
Yes	Yes	Female	No	Fiber optic	No	Month-to-month	Yes	0	Yes	Yes	Yes	820.5	No	99.6500013236789	9323-K3NKC	No	No	Yes	8	Electronic check
Yes	Yes	Male	No	Fiber optic	No	Month-to-month	No	0	No	Yes	No	1949.4002244140635	No	104.07566833486054	1453-K3NKC	Yes	Yes	Yes	22	Credit card (automatic)
No	phone service	No	Female	Yes	DSL	No	Month-to-month	No	0	No	No	301.899998964844	No	40.6323269775308	6713-CKMNC	No	No	No	10	Mailed check
Yes	Yes	Female	No	Fiber optic	Yes	Month-to-month	Yes	0	Yes	Yes	Yes	3046.05004828125	Yes	104.8000033173781	7892-FOQKP	No	No	Yes	28	Electronic check
No	No	Male	Yes	DSL	No	One year	No	0	No	No	No	3487.949931171875	No	86.2141036873847	6388-1ABQJ	Yes	Yes	Yes	62	Bank transfer (automatic)
No	Yes	Male	Yes	DSL	No	Month-to-month	No	0	No	No	No	567.8500322070312	Yes	97.9831519262993	9793-GRSKD	Yes	No	Yes	13	Mailed check
No	No	Male	No	No internet service	No	No internet service	Two year	No	0	No internet service	No internet service	326.7999877026875	No	21.603702114257832	74894-K3NKC	No	No internet service	Yes	16	Credit card (automatic)
Yes	No	Male	No	Fiber optic	No	One year	No	0	Yes	Yes	Yes	5681.10029765625	Yes	36.22477722167969	8893-TYVXV	No	No	Yes	58	Credit card (automatic)
Yes	Yes	Male	No	Fiber optic	No	Month-to-month	Yes	0	Yes	Yes	Yes	5026.5999964821219	No	103.6999964821219	0255-KJGZ	No	Yes	Yes	48	Bank transfer (automatic)
Yes	Yes	Male	Yes	Fiber optic	Yes	Month-to-month	No	0	Yes	Yes	Yes	2086.05004828125	No	135.20621020976502	9129-JLPH	No	No	Yes	23	Electronic check
Yes	No	Female	Yes	Fiber optic	Yes	Two year	No	0	Yes	Yes	Yes	3855.14990734375	Yes	105.84712379680486	3655-ONVCE	Yes	Yes	Yes	40	Credit card (automatic)
No	No	Female	No internet service	No	No internet service	One year	No	0	No internet service	No internet service	No internet service	1022.9002122270312	No	17.334632873333336	8193-K9K5C	No	No internet service	Yes	32	Mailed check
Yes	No	Male	Yes	Fiber optic	No	Two year	No	0	Yes	Yes	Yes	7832.25	No	114.75478825155312	9659-K0VKT	Yes	No	Yes	71	Bank transfer (automatic)
No	No	Female	No	DSL	Yes	Month-to-month	Yes	0	Yes	No	No	526.349973589375	Yes	55.20000376269845	4190-K8LJW	Yes	No	Yes	10	Credit card (automatic)
No	Yes	Female	No	Fiber optic	No	Month-to-month	No	0	Yes	No	Yes	1862.9002244140635	No	126.33211517333954	4185-MVFRB	No	Yes	Yes	21	Electronic check
No	phone service	Yes	Male	No	DSL	No	Month-to-month	Yes	1	Yes	No	39.650001323678906	No	39.650001323678906	8779-GRQMV	No	No	No	1	Electronic check
No	No	Male	No internet service	No	No internet service	One year	No	0	No internet service	No internet service	No internet service	202.25	Yes	3.34150092163086	1680-VQCVW	No	No internet service	Yes	12	Bank transfer (automatic)
No	No	Male	No internet service	No	No internet service	Month-to-month	Yes	0	No internet service	No internet service	No internet service	20.149999618330773	No	20.149999618330773	1066-JKQKJ	No	No internet service	Yes	1	Mailed check
Yes	Yes	Female	No	DSL	Yes	Two year	No	0	No	No	No	3505.10029765625	Yes	6.06461564199291	3638-VEJ5W	No	Yes	Yes	58	Credit card (automatic)

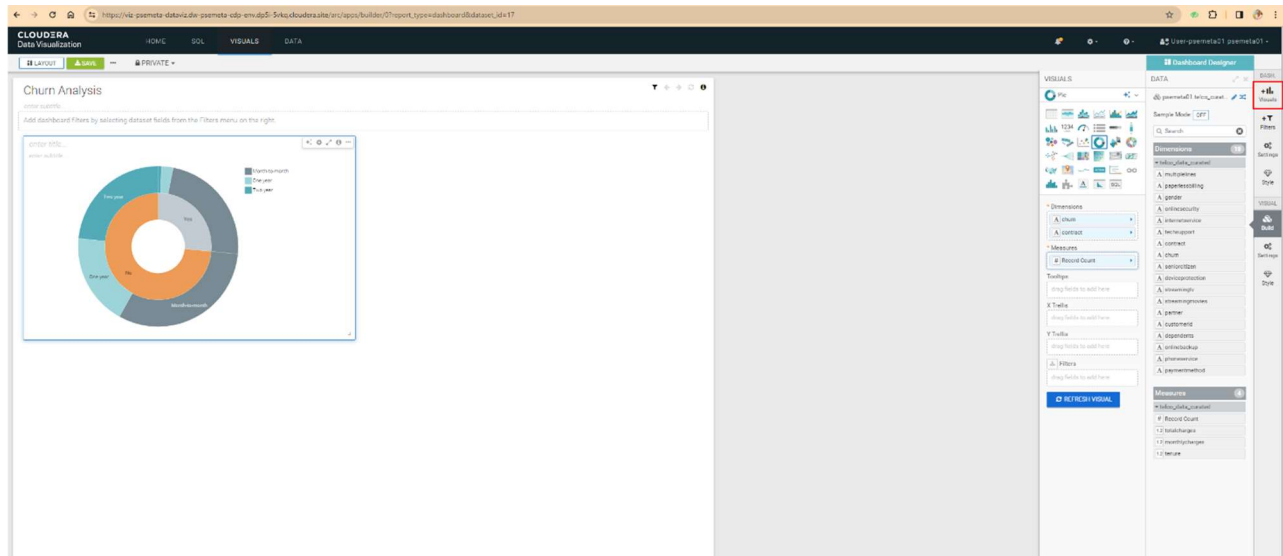
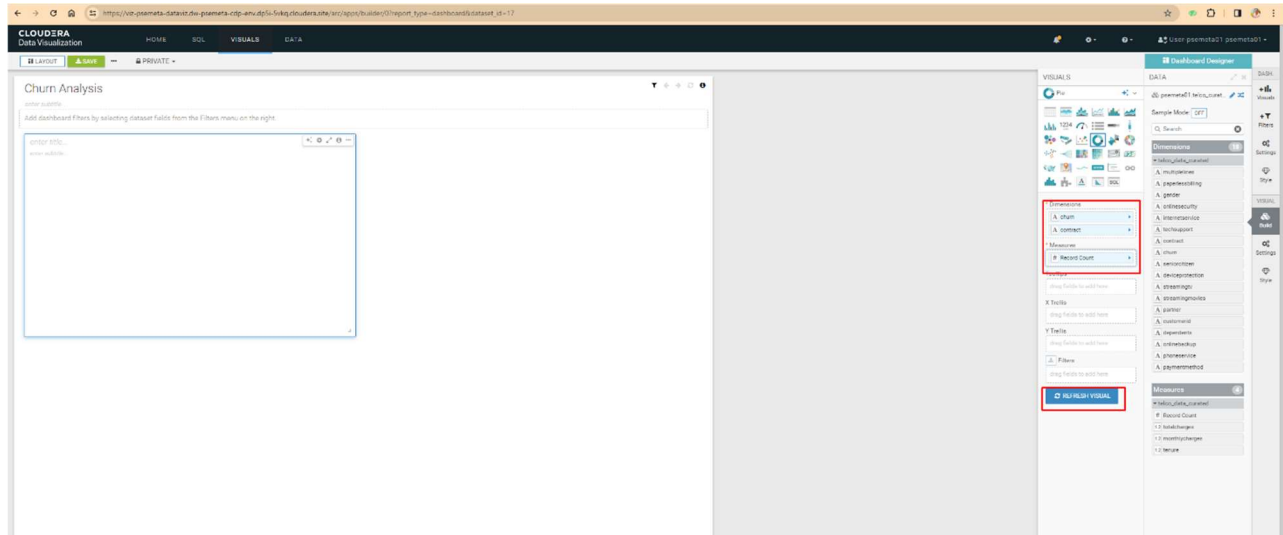
12. When opening the design canvas of a new panel, remove the element that is added by default, by clicking on the three dots (...) button at the top right of the element, and then clicking on the option **Delete Visual**



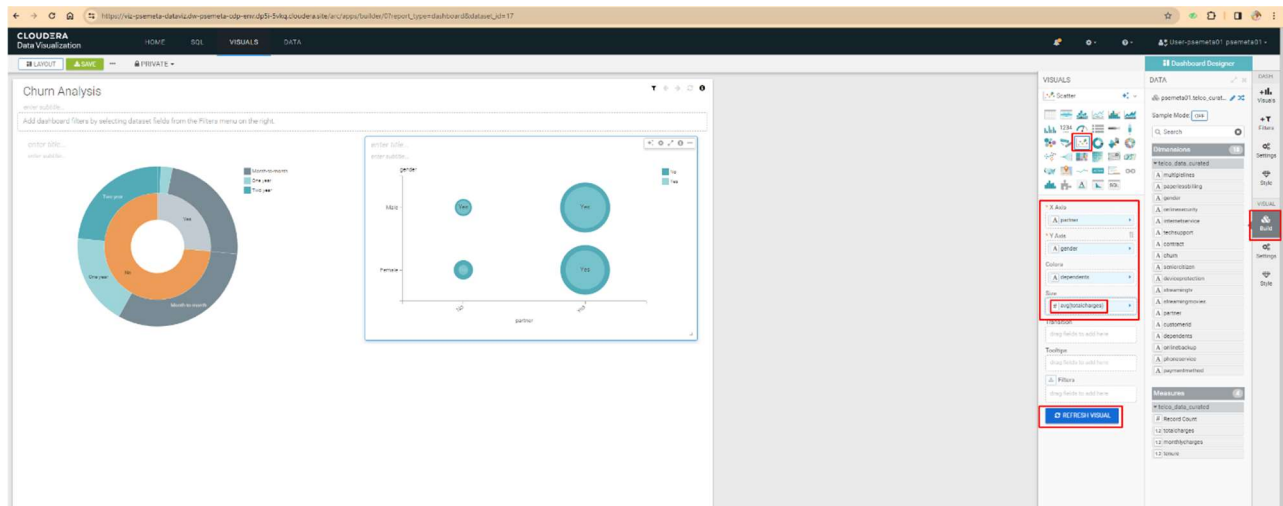
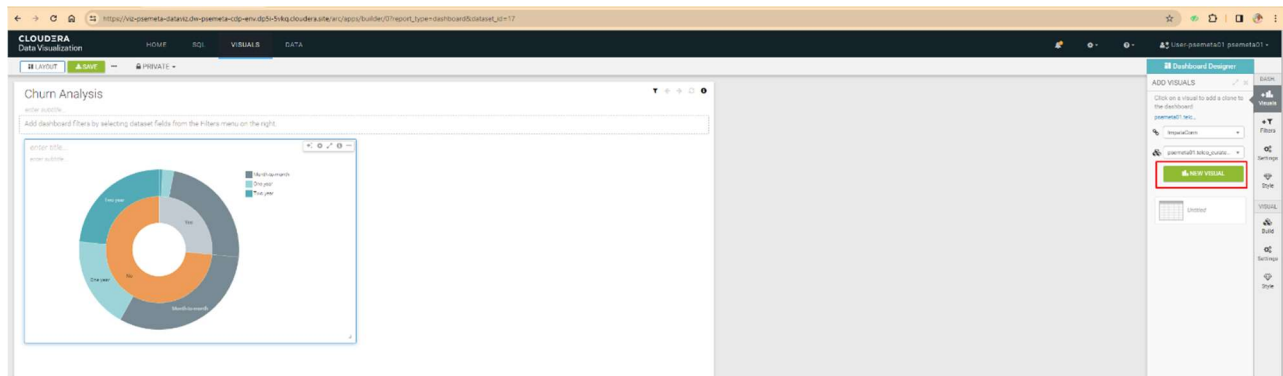
At the top of the canvas, in the enter title field, enter the name **Churn Analysis** to identify the dashboard.





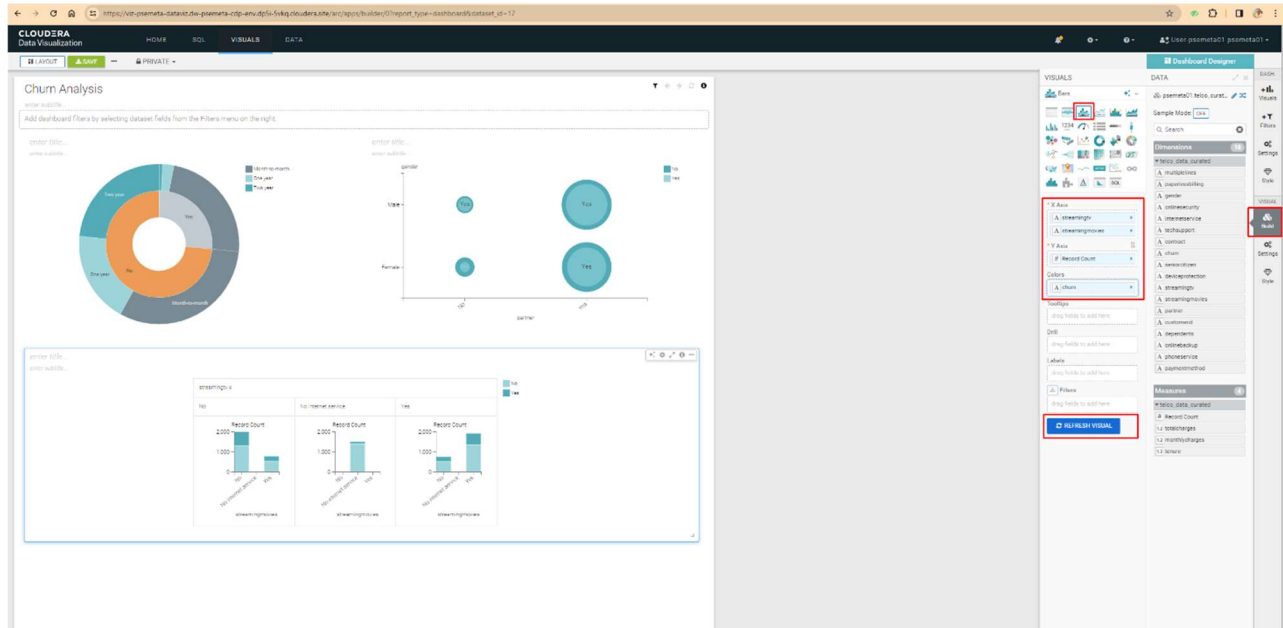


15. Add the second visual element, which is a scatter chart with the dimension **partner** like X Axis, **gender** how Y Axis, **dependents** as Colors and **avg (total charges)** as Size. Once finished, click the button **Refresh Visual**.



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**.



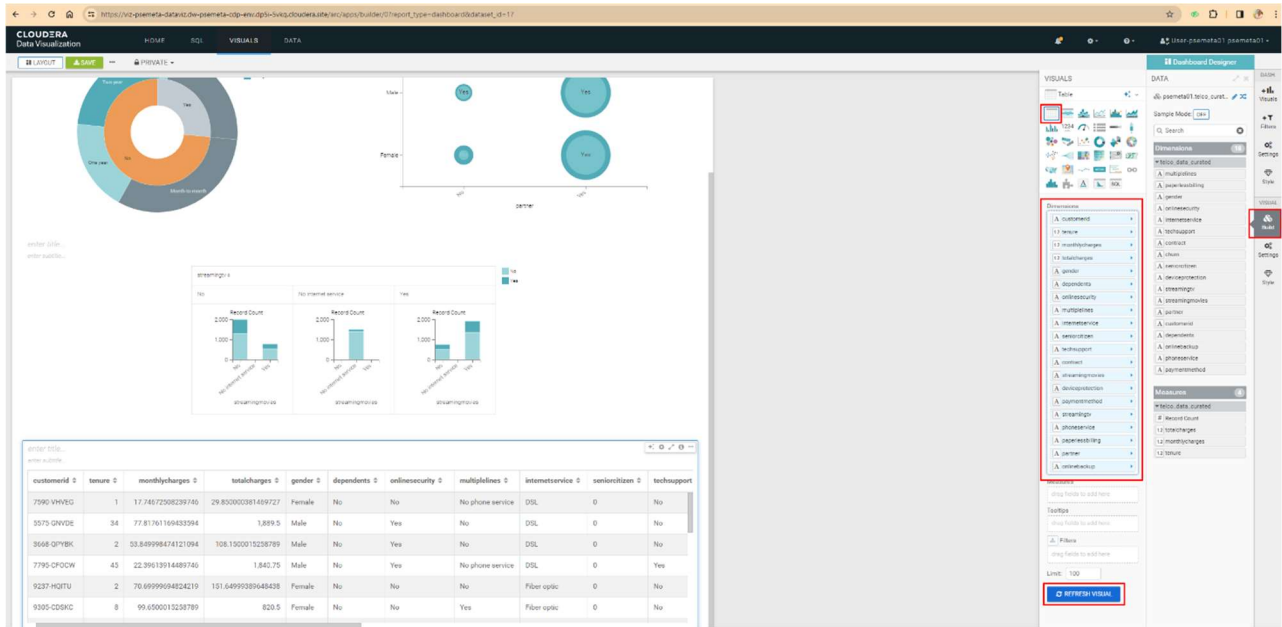
16. Add the fourth and last visual element, which is a table with the dimensions and metrics of the dataset. Be sure to add all 17 dimensions and 3 metrics to the table, so in total 20 elements. Once finished, click the button **Refresh Visual**.

#### Dimensions

- A customerid ▶
- 1.2 tenure ▶
- 1.2 monthlycharges ▶
- 1.2 totalcharges ▶
- A gender ▶
- A dependents ▶
- A onlinesecurity ▶
- A multiplelines ▶
- A internetservice ▶
- A seniorcitizen ▶
- A techsupport ▶
- A contract ▶
- A streamingmovies ▶
- A deviceprotection ▶
- A paymentmethod ▶
- A streamingtv ▶
- A phoneservice ▶
- A paperlessbilling ▶
- A partner ▶
- A onlinebackup ▶

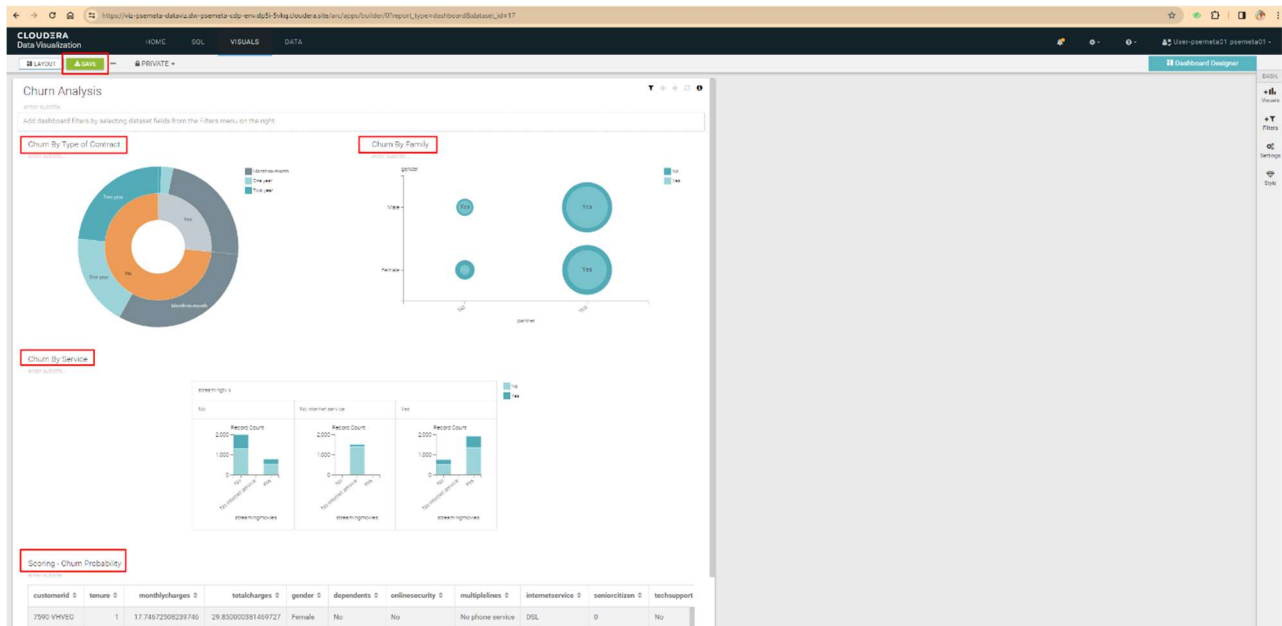
#### Measures

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.