

Data Lifecycle on CDP Public Cloud

Data Warehouse Lab

This is the 2nd part of the 'Data Warehouse Lab' is divided into 2 Parts.

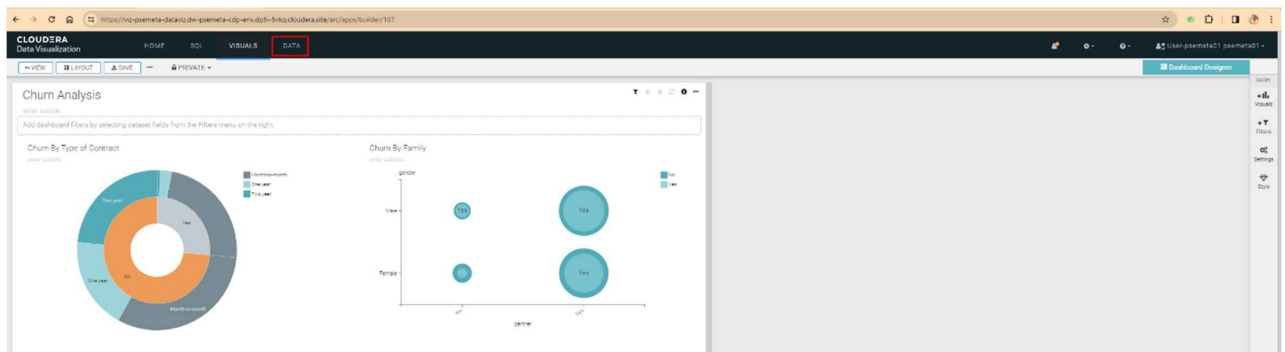
Let's work on the 2nd Part now.

Part 2: Integrating Data Viz & CML

Goals:

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

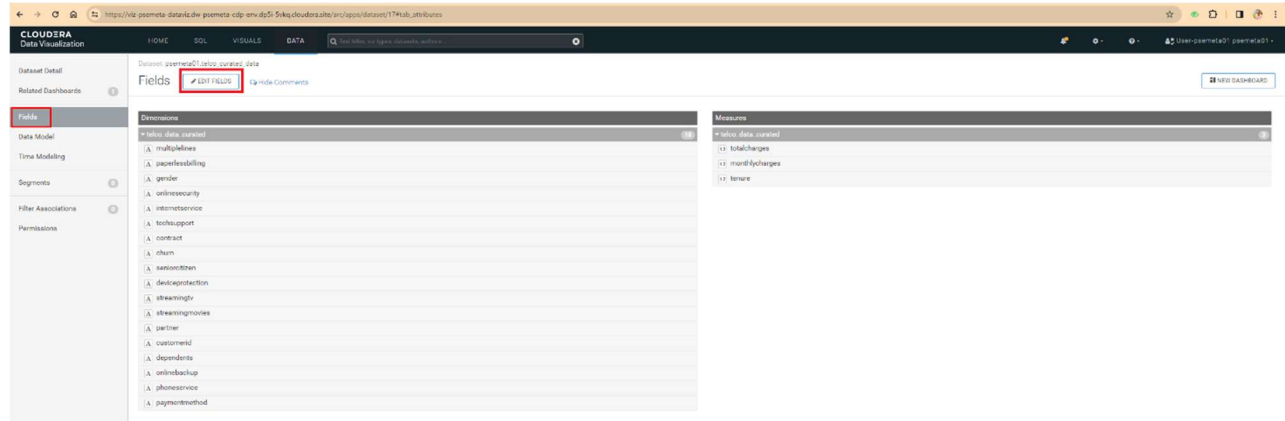
1. Edit the previously created Dataset, in Data -> <user_assigned>.telco_data_curated.



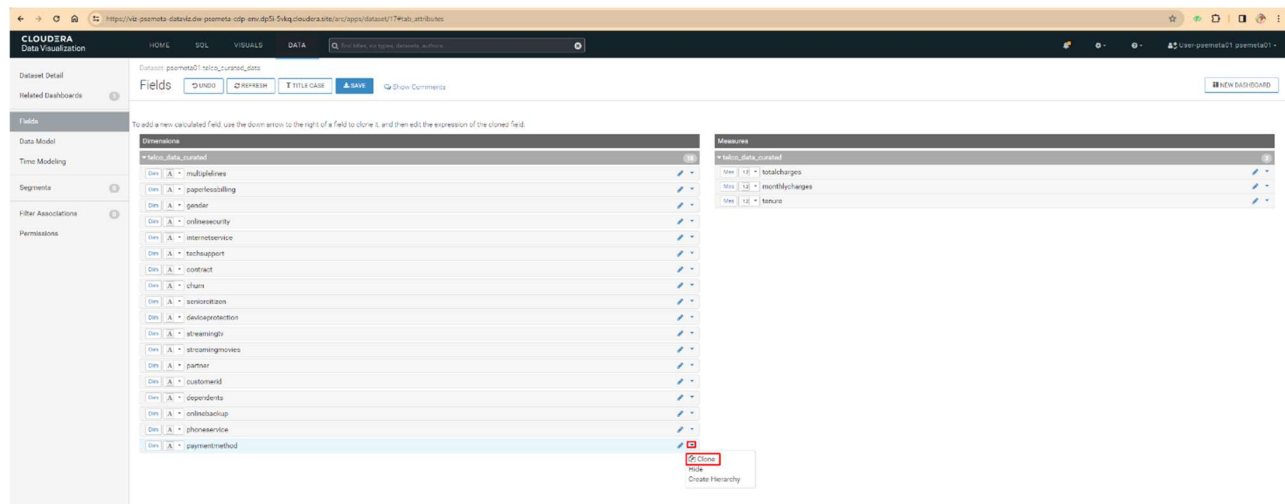
The screenshot shows the Cloudera Data Warehouse interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA', with 'DATA' being the active tab. The main content area displays a list of datasets. The dataset 'psmetad01.telco_data_curated' is highlighted with a red box.

Title/Tabla	ID	Created	Last Updated	Modified By	# Dashboards
psmetad01.telco_data_curated	17	Mar 23, 2024	a minute ago	psmetad01	1
dpltdash.telco_data_curated	13	Mar 20, 2024	4 days ago	dpltdash	1

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.



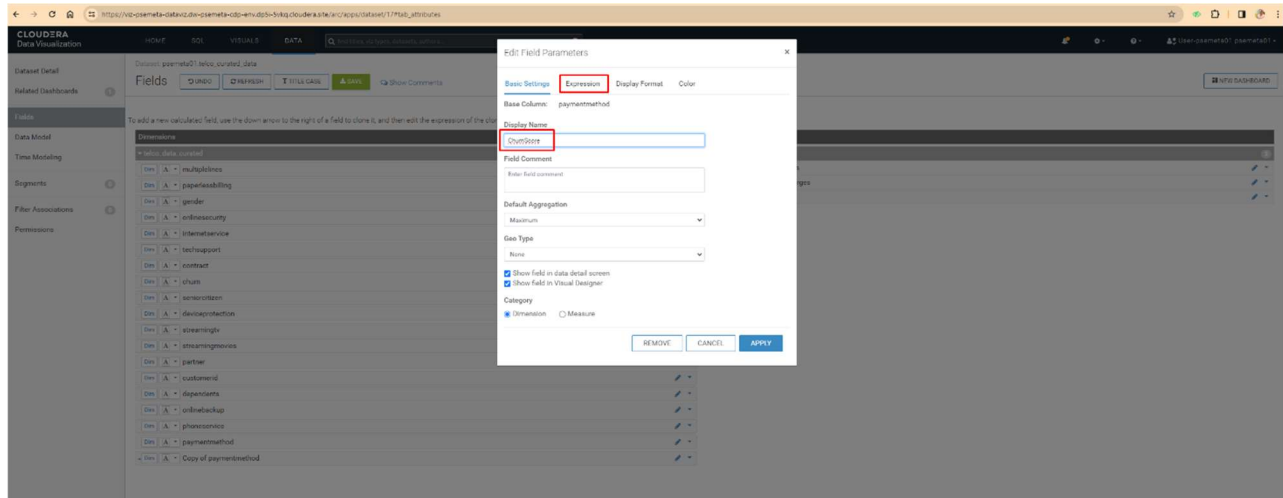
3. In the list of **Dimensions**, click the down arrow of the last field in the list, and select the option **Clone**.



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

The screenshot shows the Cloudera Data Visualization (CDV) interface. The top navigation bar includes 'HOME', 'SQL', 'VISUALS', and 'DATA'. The left sidebar contains 'Dataset Detail', 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Segments', 'Filter Associations', and 'Permissions'. The main area is titled 'Fields' and shows a list of dimensions and measures. The 'Dimensions' list includes fields like 'info: date: current', 'dim: A: multiplexes', 'dim: A: paperlessbilling', 'dim: A: gender', 'dim: A: onlinesecurity', 'dim: A: internetaccess', 'dim: A: techsupport', 'dim: A: contract', 'dim: A: churn', 'dim: A: sessionlength', 'dim: A: deviceprotection', 'dim: A: streamingtv', 'dim: A: streamingmovies', 'dim: A: partner', 'dim: A: customerid', 'dim: A: dependents', 'dim: A: onlinebackup', 'dim: A: phonenumber', 'dim: A: paymentmethod', and 'dim: A: Copy of paymentmethod'. The 'Measures' list includes 'info: date: current', 'dim: A: totalcharges', 'dim: A: monthlycharges', and 'dim: A: tenure'. The 'Copy of paymentmethod' field is highlighted with a red box, and a pencil icon is visible next to it.

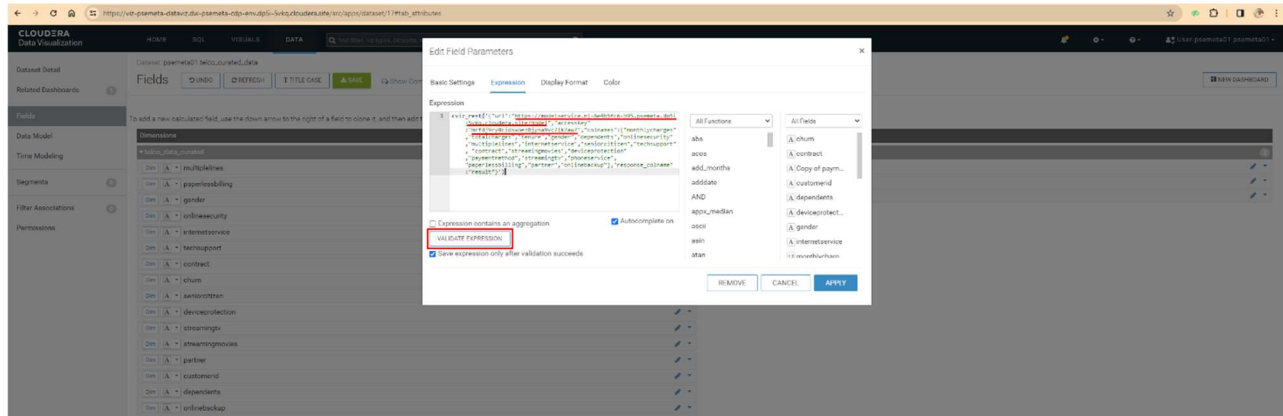
5. In the popup window that appears, enter the name of the new field in **Display Name**. We suggest that you enter *ChurnScore*.



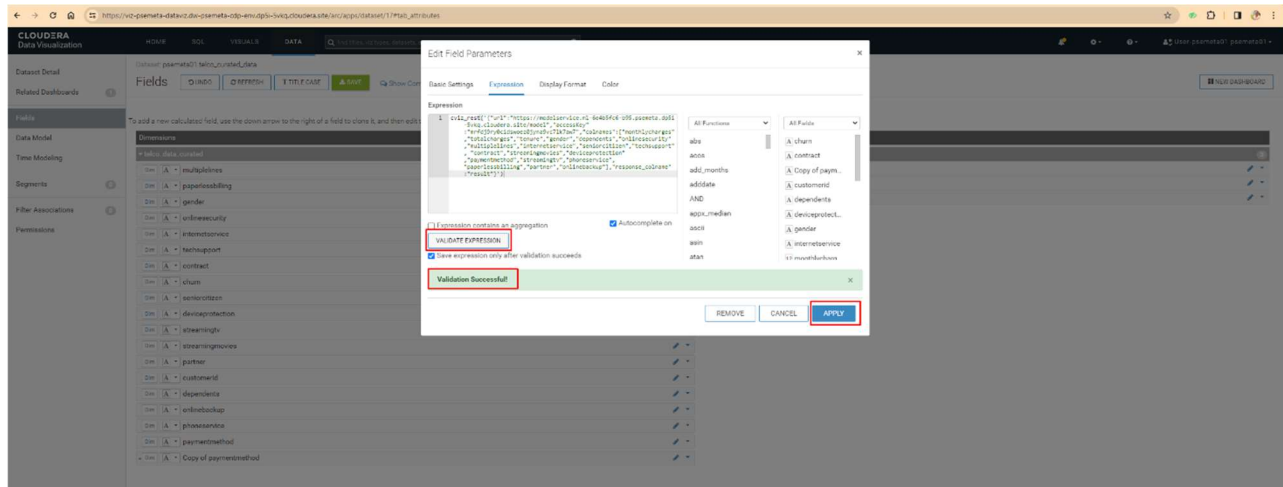
6. Go to the Expressions tab and enter the following value in the Expression field. This will allow you to call the REST API of the Model you have previously deployed.

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

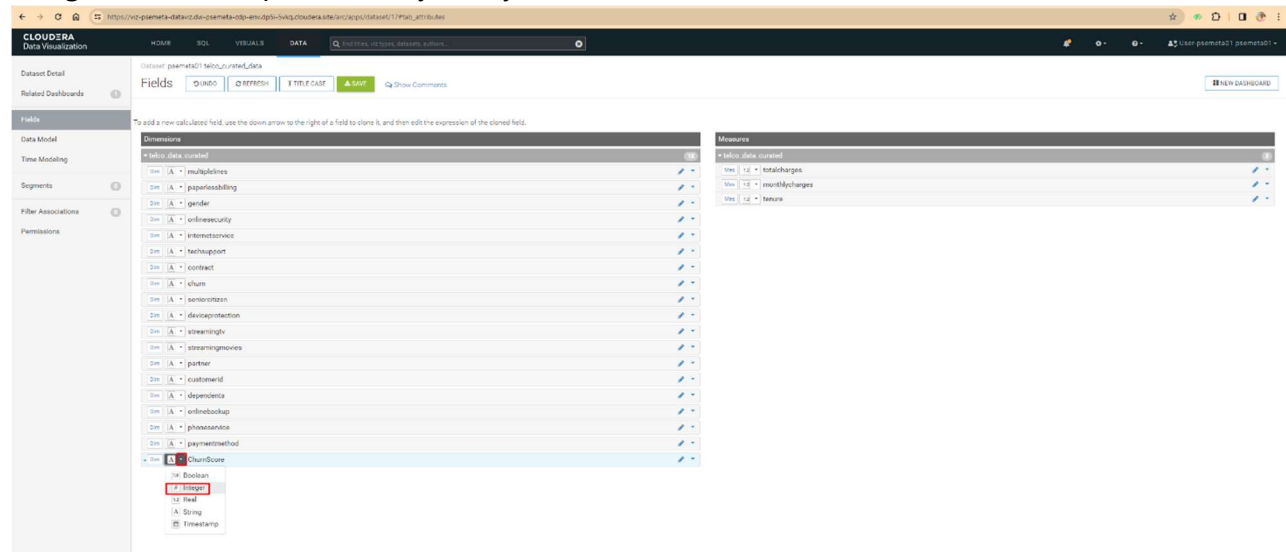
```
cviz_rest({'url':"https://modelservice.ml-6e4b5fc6-b95.psemeta.dp5i-5vkq.cloudera.site/model","accessKey":"miqnv6bxiwontz84ozoxfyc5rt6p6grb","colnames":["monthlycharges","totalcharges","tenure","gender","dependents","onlinesecurity","multiplelines","internetservice","seniorcitizen","techsupport","contract","streamingmovies","deviceprotection","paymentmethod"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



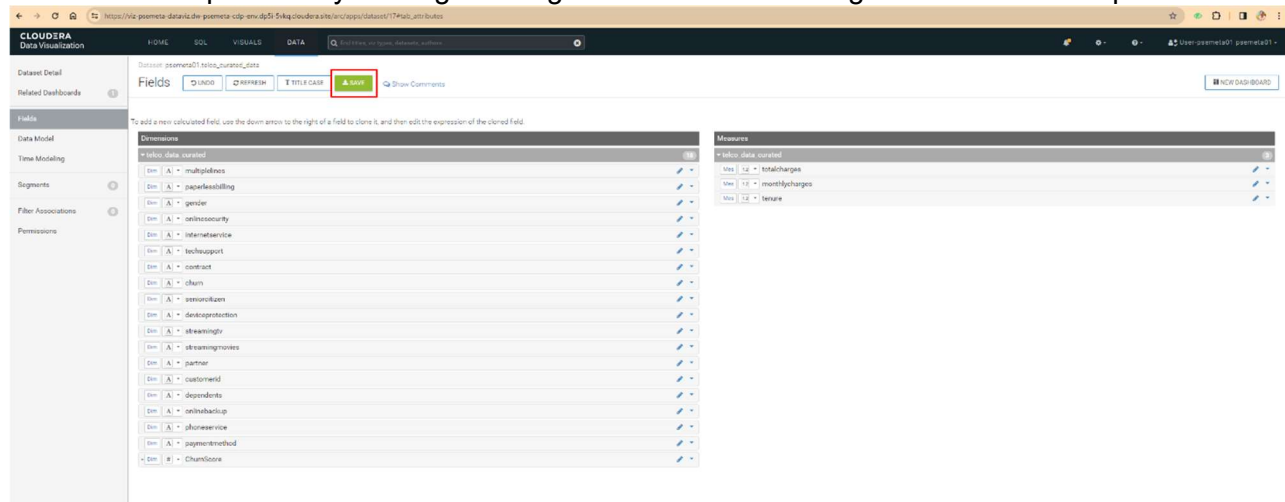
7. Finish the process of copying the **url** and the **accessKey**, click the **Validate Expression** button at the top of the window. If the message appears in green **Validation Successful**, Click on **Apply** to save the settings made.



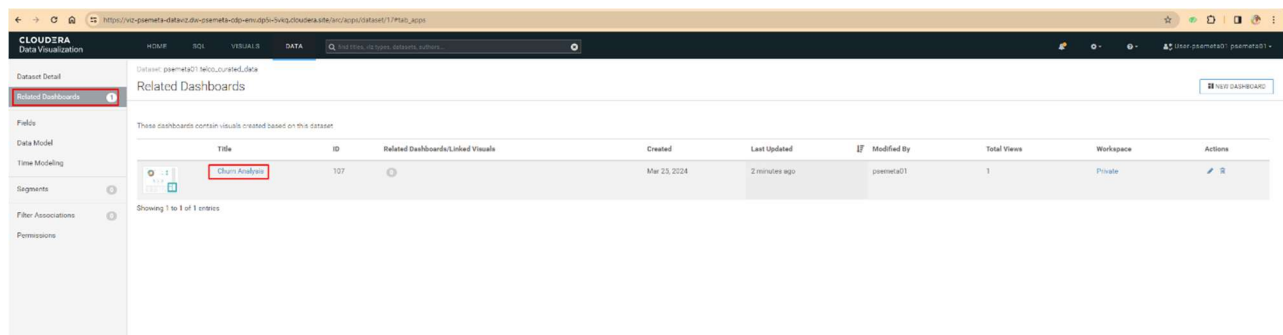
8. The new field should appear in the list of fields. Change the data type, selecting the type **Integer**, which is represented by the symbol #



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.

The screenshot shows the CloudSRA Data Visualization interface. The dashboard includes a donut chart titled 'Churn By Service', a bubble chart, and a table titled 'Scoring - Churn Probability'. The table has columns for customer ID, tenure, monthly charges, total charges, gender, dependents, online security, multiple phones, internet service, senior citizen, and tech support. A red box highlights the table, and a red arrow points to the 'Build' button in the right-hand menu.

customerid	tenure	monthlycharges	totalcharges	gender	dependents	onlinesecurity	multiphones	internetservice	seniorcitizen	techsupport
7590-WHFG	1	17.746726297046	29.850200381469727	Female	No	No	No phone service	DSL	0	No
5575-QNVSE	34	77.8176116433254	1,889.5	Male	No	Yes	No	DSL	0	No
3668-QPYBK	2	53.84998647472104	108.1500015258789	Male	No	Yes	No	DSL	0	No
1795-CHOCW	45	22.39513871485946	1,040.75	Male	No	Yes	No phone service	DSL	0	Yes
9237-HGBTU	2	75.80999698324219	151.54999399548338	Female	No	No	No	Fiber optic	0	No

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.

