

Data Lifecycle in CDP Public Cloud

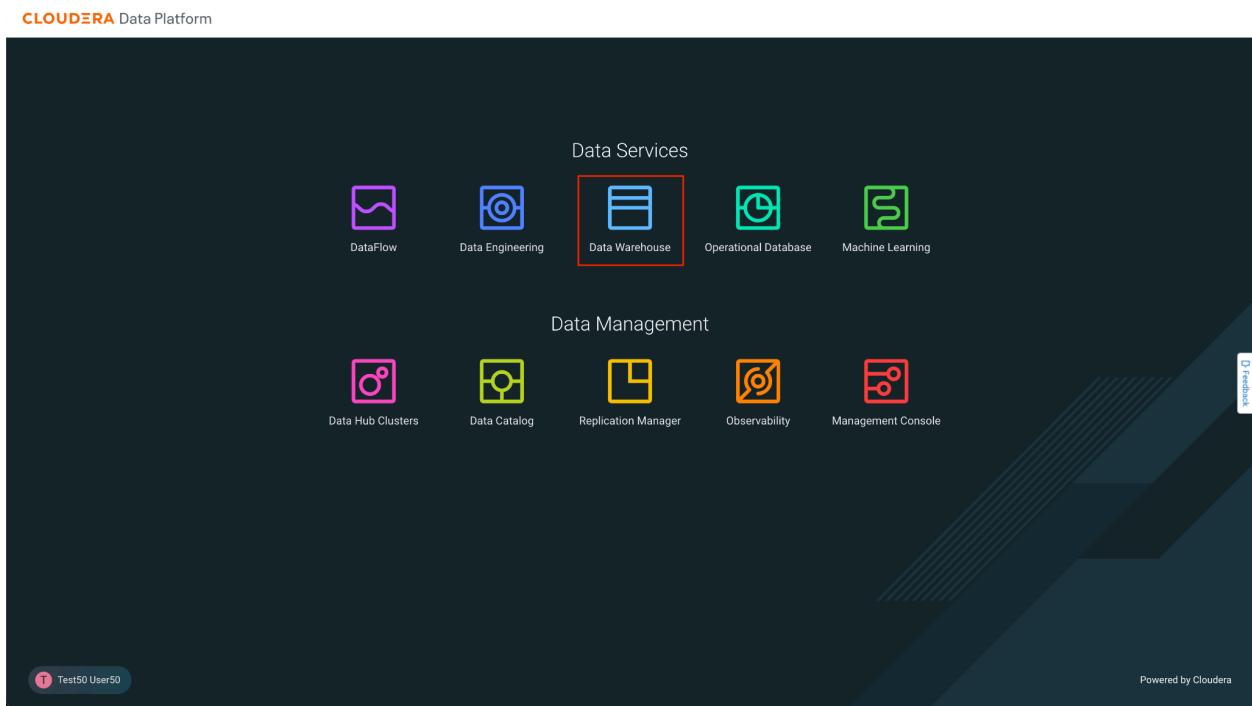
Data Warehouse Lab

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. In the Data Warehouse welcome screen, click on Data Visualization in the left menu.

The screenshot shows the Cloudera Data Warehouse interface. On the left sidebar, under the 'Data Visualization' section, there is a 'Data Viz' button. This button is highlighted with a red box in the image. The main content area displays an 'Overview' dashboard with sections for 'Get started with Data Warehouse', 'Create', 'Query and visualize data', and 'Guides and More'. Below these sections are two tabs: 'Database Catalogs | 1' and 'Virtual Warehouses | 2'. The 'Virtual Warehouses' tab is selected, showing three virtual warehouses: 'ssa-datalake-de...', 'impara-vw-0', and 'hive-vw-0'. Each warehouse entry includes details like 'TOTAL CORES', 'TOTAL MEMORY', 'EXECUTORS', and 'TYPE'.

3. In Data Visualization, click on the **Data Viz** button according to what your user was assigned.

The screenshot shows the 'Data Visualization' page. The 'Data Viz' button from the previous screenshot is visible in the top right corner of the main content area. The main content area displays a table of data visualizations. One row in the table is highlighted with a red box around the 'Data Viz' button in the last column. The columns in the table include NAME, DATA VISUALIZATION ID, Environment ID, VERSION, CPU, MEMORY, UPTIME, CREATED BY, and a 'Data Viz' button.

NAME	DATA VISUALIZATION ID	Environment ID	VERSION	CPU	MEMORY	UPTIME	CREATED BY	Data Viz
dataviz-0	viz-1685400615-2kkq	env-9ggpp	7.1.1-b30	2	8 GB	an hour	acampos	Data Viz

4. Once in Data Visualization, go to the Data option from the top menu, and then to the Connector **ImpalaConn** from the left menu.

The screenshot shows the Data Visualization application's interface. On the left, there is a sidebar with a 'NEW CONNECTION' button, a 'All Connections' section containing an 'ImpalaConn' item (which is highlighted with a red box), and a 'samples' section. The main area is titled 'Datasets' and contains a table with 12 rows. The columns are: Title/Table, ID, Created, Last Updated, Modified By, and # Dashboards. The rows list various datasets: 'Food Stores Inspection in NYC', 'Cereals', 'World Life Expectancy', 'Earthquake Data January 2019', 'US State Populations Over Time', 'US County Population', 'Global Information Security Threats', and 'Restaurant Inspection SF'. Each row also includes a 'Data Connection: samples' link.

5. We have to 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.

The screenshot shows the 'New Dataset' dialog box open in the Data Visualization application. The dialog has a header with 'NEW DATASET' (which is highlighted with a red box), 'ADD DATA', and a close button. Below the header is a 'Datasets' tab and a 'Connection Explorer' tab. The main area of the dialog shows a table with columns: Title/Table, ID, Created, Last Updated, Modified By, and # Dashboards. A single row is present with the status 'No data'.

6. Enter the information for the new data source:

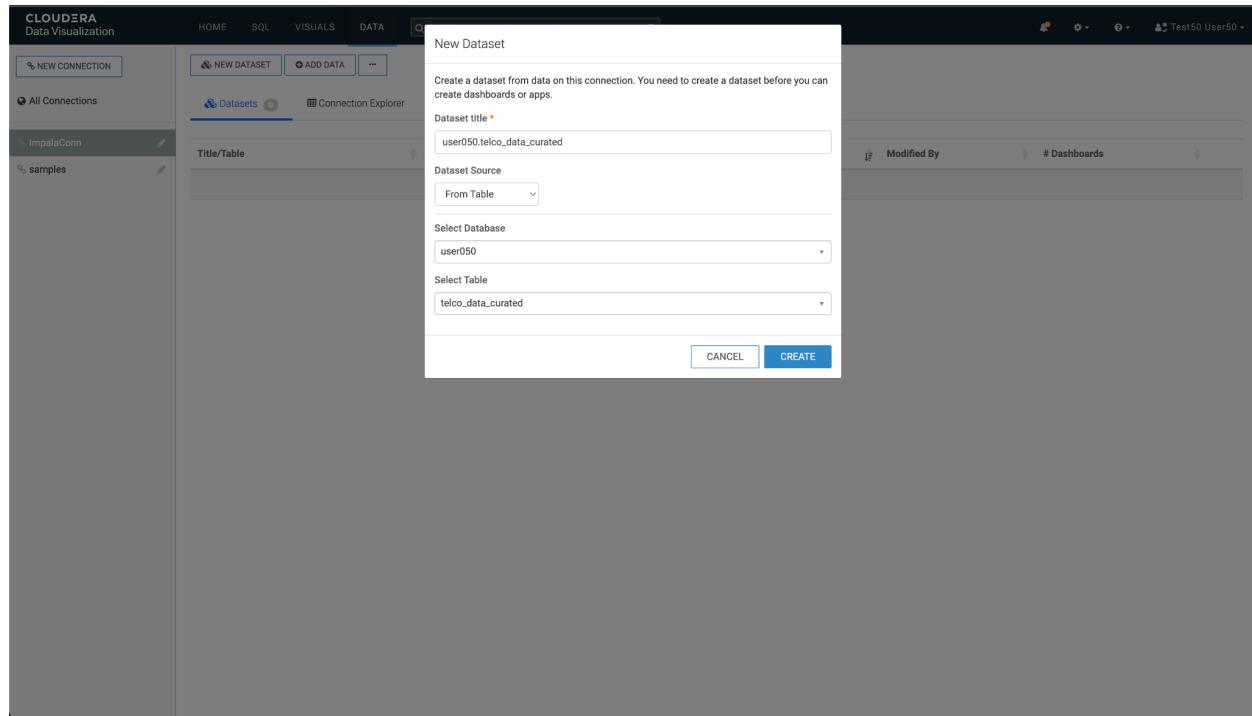
Dataset title: <assigned_user>.telco_curated_data

Dataset Source: From table

Select Database: <assigned_user>

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.

The screenshot shows the Databricks Dataset list interface. On the left, there's a sidebar with connection management options like 'NEW CONNECTION', 'All Connections', and 'ImpalaConn'. The main area displays a table of datasets. The first row in the table is highlighted, showing the dataset name 'user050.telco_data_curated', ID 16, creation date 'May 29, 2023', last update 'a few seconds ago', modified by 'user050', and 0 dashboards. There are also edit and delete icons for this row.

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

The screenshot shows the Databricks Dataset Detail page for 'user050.telco_data_curated'. The left sidebar includes sections for 'Dataset Detail' (selected), 'Related Dashboards', 'Fields', 'Data Model', 'Time Modeling', 'Segments' (with 0 items), 'Filter Associations' (with 0 items), and 'Permissions'. The main content area is titled 'Detail' and contains the following information:

- Dataset:** user050.telco_data_curated
- Table:** user050.telco_data_curated
- Connection Type:** Impala
- Data Connection:** ImpalaConn
- Description:** (empty)
- Join Elimination:** Enabled
- Result Cache:** From Connection
- Incremental Results:** Disabled

Below this, a timeline section shows the dataset's history:

- ID: 16
- Created on: May 29, 2023 06:15 PM
- Created by: user050
- Last updated: May 29, 2023 06:15 PM
- Last updated by: user050

At the top right of the detail page, there are 'CLONE DATASET' and 'NEW DASHBOARD' buttons.

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

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes links for HOME, SQL, VISUALS, and DATA, along with a search bar and user information. On the left, a sidebar titled 'Dataset Detail' lists 'Related Dashboards', 'Fields' (which is currently selected), 'Data Model', 'Time Modeling', 'Segments' (with a count of 0), 'Filter Associations' (with a count of 0), and 'Permissions'. The main content area is titled 'Dataset: user050 telco_data_curated'. It features two sections: 'Dimensions' and 'Measures'. The 'Dimensions' section contains 18 items under the 'telco_data_curated' category, including 'multiplelines', 'paperlessbilling', 'gender', 'onlinesecurity', 'internetservice', 'techsupport', 'contract', 'churn', 'seniorcitizen', 'deviceprotection', 'streamingtv', 'streamingmovies', 'partner', 'customerid', 'dependents', 'onlinebackup', 'phoneservice', and 'paymentmethod'. The 'Measures' section contains 3 items under the same category, labeled 'totalcharges', 'monthlycharges', and 'tenure'. There are also 'Edit Fields' and 'Hide Comments' buttons at the top of the Fields section.

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.

The screenshot shows the Cloudera Data Visualization interface. The left sidebar has a 'Data Model' section selected. The main area displays the dataset 'telco_data_curated'. A prominent button labeled 'SHOW DATA' is highlighted with a red box. Other visible buttons include 'EDIT DATA MODEL' and 'Apply Display Format'.

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.

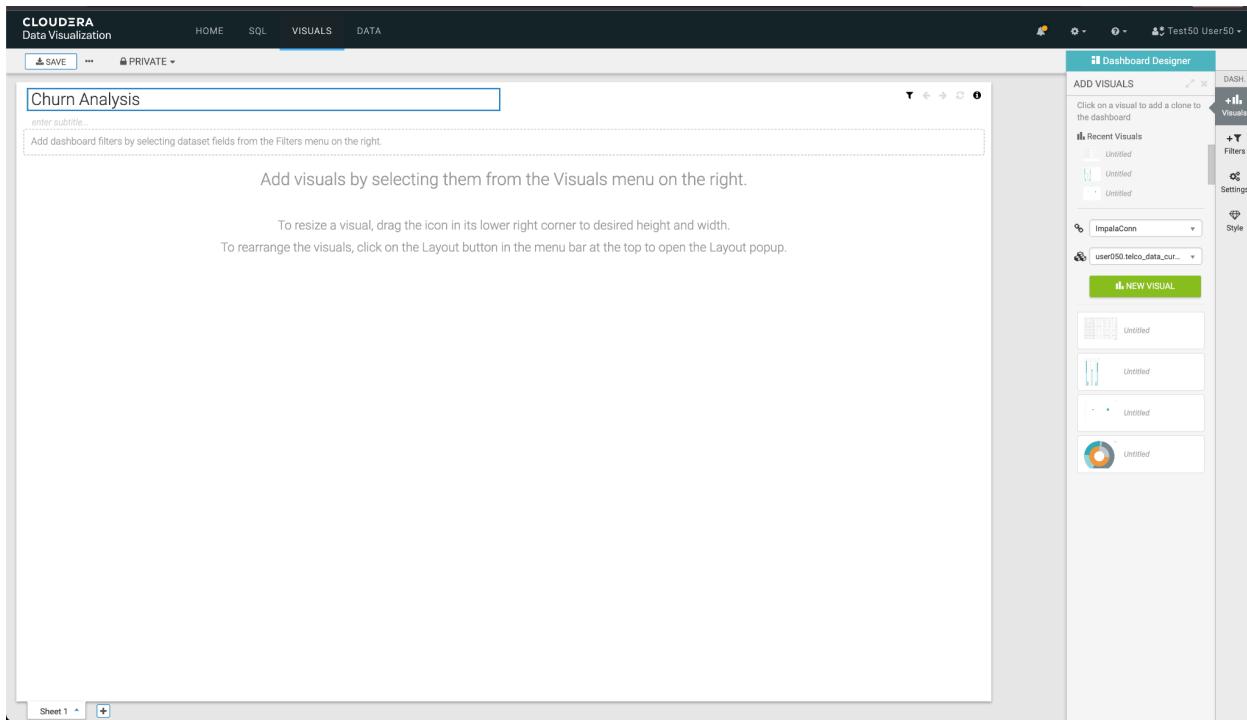
The screenshot shows the same Cloudera Data Visualization interface. The 'NEW DASHBOARD' button is highlighted with a red box. Below the table, the first few rows of the 'telco_data_curated' dataset are shown:

multiplelines	paperlessbilling	gender	onlinesecurity	internetservice	techsupport	contract	churn	seniorcitizen	deviceprotection	streamingtv	streamingmovies	totalcharges	partner	monthlycharges	customerid	dept
No phone service	Yes	Female	No	DSL	No	Month-to-month	No	0	No	No	No	29.85000381469727	Yes	32.602622985839844	7590-VHVEG	N
No	No	Male	Yes	DSL	No	One year	No	0	Yes	No	No	1889.5	No	79.32872009277344	5575-GNVED	N
No	Yes	Male	Yes	DSL	No	Month-to-month	Yes	0	No	No	No	108.1500015258789	No	53.849998474121094	3668-QPYBK	N
No phone service	No	Male	Yes	DSL	Yes	One year	No	0	Yes	No	No	1840.75	No	39.008785247802734	7795-CFOCW	N
No	Yes	Female	No	Fiber optic	No	Month-to-month	Yes	0	No	No	No	151.64999389648438	No	70.69999694824219	9237-HQITU	N
Yes	Yes	Female	No	Fiber optic	No	Month-to-month	Yes	0	Yes	Yes	Yes	820.5	No	99.6500015258789	9305-CDSKC	N
Yes	Yes	Male	No	Fiber optic	No	Month-to-month	No	0	No	Yes	No	1949.4000244140625	No	154.11448669433594	1452-KIOVK	Y
No phone service	No	Female	Yes	DSL	No	Month-to-month	No	0	No	No	No	301.8999938964844	No	46.75687789916992	6713-OKOMC	N
Yes	Yes	Female	No	Fiber optic	Yes	Month-to-month	Yes	0	Yes	Yes	Yes	3046.050048828125	Yes	104.80000305175781	7892-POOKP	N

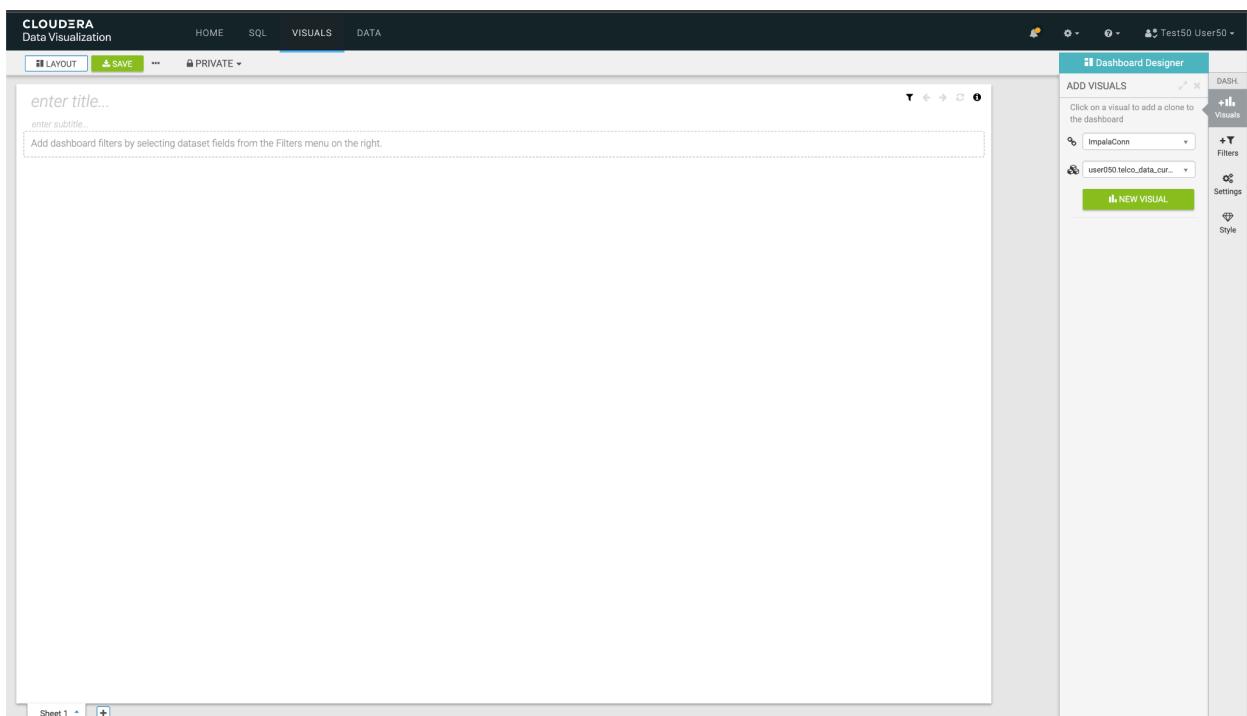
12. When opening the design canvas of a new panel, remove the element that is added by default, by clicking on the three dots (...) button at the top right of the element, and then clicking on the option **Delete Visual**

The screenshot shows the Cloudera Data Visualization interface. On the left, there's a dashboard canvas with a table visual containing data about phone service. A context menu is open over the table, with the 'Delete Visual' option highlighted. The menu also includes options like 'View Data and Queries', 'Download as...', 'Save as Table or Dataset', 'Hide Empty Title & Subtitle', and 'Clone'. To the right of the canvas is the 'Dashboard Designer' sidebar, which includes sections for DATA, Dimensions, Measures, Filters, and Measures. The 'Dimensions' section lists fields like 'multiplelines', 'paperlessbilling', 'gender', and 'onlinesecurity'. The 'Measures' section lists fields like 'Record Count', 'totalcharges', 'monthlycharges', and 'tenure'. At the bottom of the sidebar, there are 'Build' and 'Style' tabs.

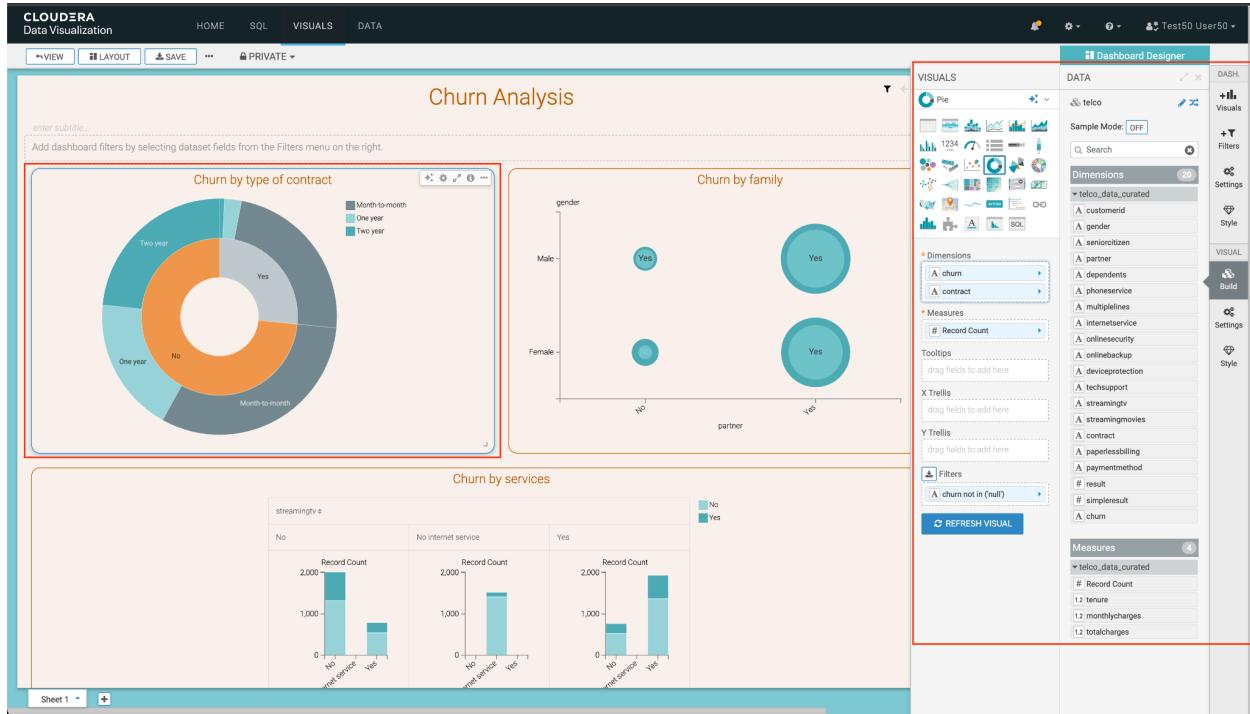
At the top of the canvas, in the enter title field, enter the name *Churn Analysis-<user_assigned>* to identify the dashboard.



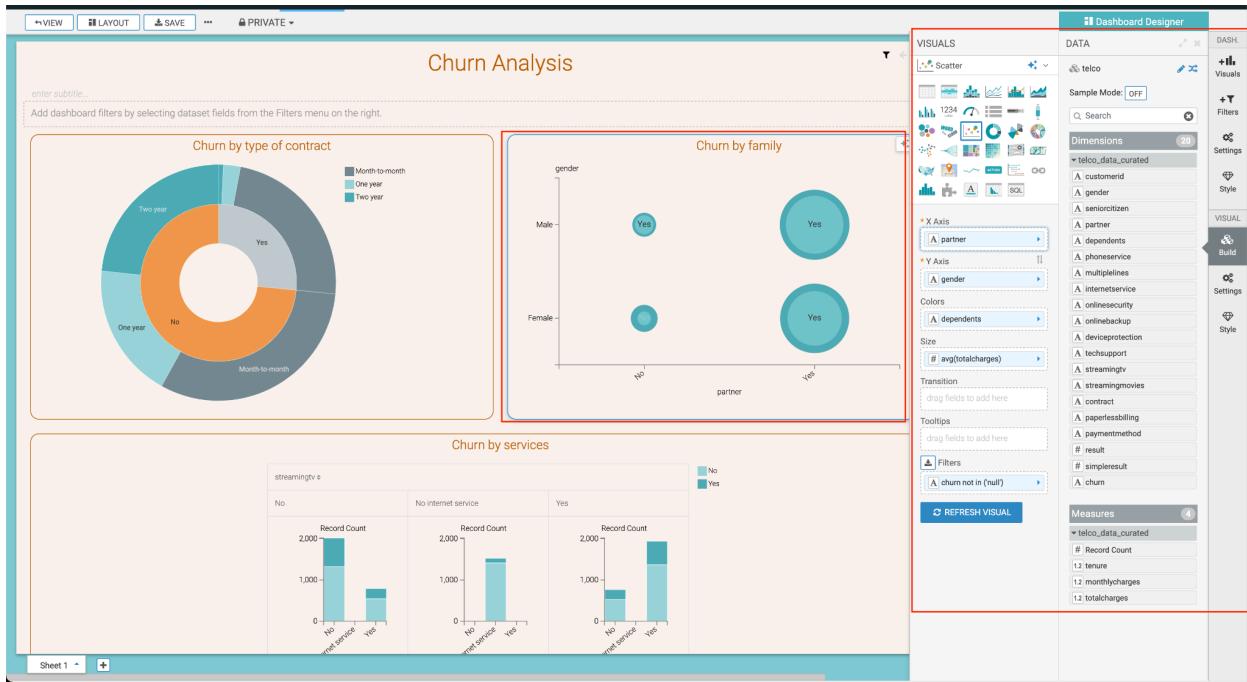
13. To add a new visual element, click on the button **Visuals** from the right menu, select the dataset that corresponds to them (telco_data_curated), and click on the button **New Visual**.



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



15. Add the second visual element (Visuals -> New Visual), which is a scatter chart with the dimension **partner** for X Axis, **gender** for 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** for X Axis, **Record Count** for Y Axis and **churn** for Colors. Once finished, click the button **Refresh Visual**.

The screenshot shows the Cloudera Data Visualization interface. At the top, there's a navigation bar with 'HOME', 'SQL', 'VISUALS', and 'DATA' tabs. Below the navigation is a toolbar with 'VIEW', 'LAYOUT', 'SAVE', and a 'PRIVATE' dropdown.

The main area contains three visual elements:

- A donut chart titled 'Month-to-month' with segments for 'partner' (blue) and 'no' (orange).
- A bar chart titled 'Churn by services' with three panels for 'streamingtv' (No, Yes), 'streamingmovies' (No, Yes), and 'internet service' (No, Yes). The Y-axis is 'Record Count' (0 to 2,000) and the X-axis categories are 'No internet service' and 'Yes'. The bars are colored by 'churn' (blue for No, orange for Yes).
- A table titled 'Scoring - Churn Probability' with columns: result, customerid, tenure, monthlycharges, totalcharges, gender, dependents, onlinesecurity, multiplelines, internetservice. The table shows several rows of error data.

The right side of the interface features a 'Dashboard Designer' sidebar with various tools and settings. A red box highlights the configuration panel for the bar chart, which includes sections for 'DATA', 'Dimensions', 'X Axis', 'Y Axis', 'Colors', 'Tooltips', 'Drill', 'Labels', 'Filters', and 'Measures'. A blue button labeled 'REFRESH VISUAL' is visible at the bottom of this panel.

result	customerid	tenure	monthlycharges	totalcharges	gender	dependents	onlinesecurity	multiplelines	internetservice
Error fetching data	7590-VHVEG	1	32.602622985839844	29.850000381469727	Female	No	No	No phone service	DSL
Error fetching data	5575-GNVDE	34	79.32872009277344	1,889.5	Male	No	Yes	No	DSL
Error fetching data	3668-QPYBK	2	53.849998474121094	108.1500015258789	Male	No	Yes	No	DSL
Error fetching data	7795-CFOCW	45	39.008785247802734	1,840.75	Male	No	Yes	No phone service	DSL
Error fetching data	9237-HQITU	2	70.69999694824219	151.64999389648438	Female	No	No	No	Fiber optic

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 18 dimensions and 3 Measures to the **Dimensions**. Once finished, click the button **Refresh Visual**.

The screenshot shows the Cloudera Data Visualization interface. On the left, there are three bar charts side-by-side, each titled "Record Count". The first chart is for "No streamingmovies" and has two bars: "No internet service" at approximately 1,800 and "Yes" at approximately 500. The second chart is for "streamingmovies" and has two bars: "No internet service" at approximately 1,500 and "Yes" at approximately 500. The third chart is for "Yes streamingmovies" and has two bars: "No internet service" at approximately 1,000 and "Yes" at approximately 1,500. Below these charts is a table titled "Scoring - Churn Probability". The table has 11 columns: customerid, tenure, monthlycharges, totalcharges, gender, dependents, onlinesecurity, multipiplines, internetservice, seniorcitizen, and a final column for churn probability. The data consists of 10 rows of customer information. To the right of the charts is a sidebar titled "Dashboard Designer" which includes sections for Dimensions, Measures, Tooltips, Filters, and a "Build" section with settings for style and filters.

Scoring - Churn Probability										
customerid	tenure	monthlycharges	totalcharges	gender	dependents	onlinesecurity	multipiplines	internetservice	seniorcitizen	
7590-VHVEG	1	32.60262298589844	29.850000381469727	Female	No	No	No phone service	DSL	0	
5575-GNVEDE	34	79.32872009277344	1,889.5	Male	No	Yes	No	DSL	0	
3668-QPYBK	2	53.849998474121094	108.1500015258789	Male	No	Yes	No	DSL	0	
7795-CFOCW	45	39.008785247802734	1,840.75	Male	No	Yes	No phone service	DSL	0	
9237-HQITU	2	70.69999694824219	151.64999389648438	Female	No	No	No	Fiber optic	0	
9305-CDSKC	8	99.6500015258789	820.5	Female	No	No	Yes	Fiber optic	0	
1452-KIOVK	22	154.11448669433594	1,949.4000244140625	Male	Yes	No	Yes	Fiber optic	0	
6713-OKOMC	10	46.756877899166992	301.8999938964844	Female	No	Yes	No phone service	DSL	0	

Save the dashboard by clicking the button **Save** from the top menu.

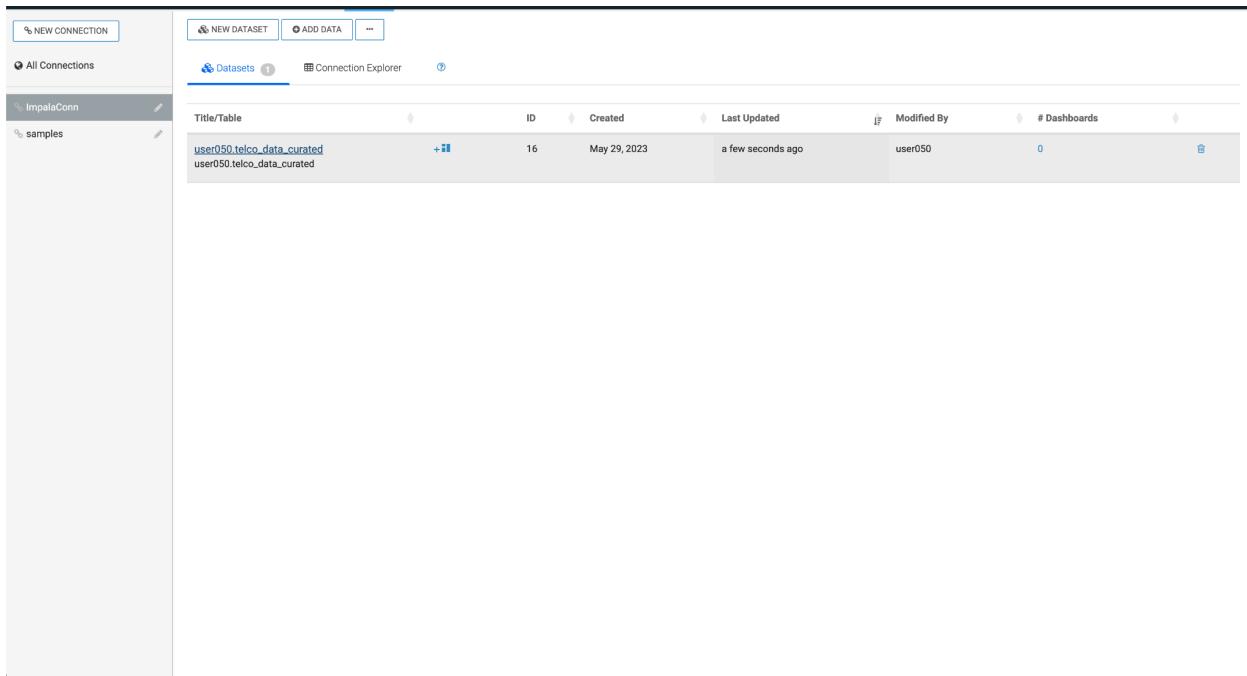
End of Part 1 - we will come back to run Part 2 later

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.



The screenshot shows the Databricks Data page. On the left, there's a sidebar with 'NEW CONNECTION' and 'All Connections' sections, and a single entry 'ImpalaConn' with a 'samples' link. The main area has tabs for 'Datasets' (which is selected) and 'Connection Explorer'. Below is a table with columns: Title/Table, ID, Created, Last Updated, Modified By, and # Dashboards. Two rows are listed: 'user050.telco_data_curated' (ID 16, Created May 29, 2023, Last Updated 'a few seconds ago', Modified by user050, 0 dashboards) and 'user050.telco_data_curated' (another row with identical details). There are also '+' and edit icons for creating new datasets.

Title/Table	ID	Created	Last Updated	Modified By	# Dashboards
user050.telco_data_curated	16	May 29, 2023	a few seconds ago	user050	0
user050.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.

CLOUDERA Data Visualization

Dataset Detail

Related Dashboards

Fields

Dataset: user050 telco_data_curated

Dimensions

Measures

EDIT FIELDS Hide Comments NEW DASHBOARD

telco_data_curated

- multipelines
- paperlessbilling
- gender
- onlinesecurity
- internetservice
- techsupport
- contract
- churn
- seniorcitizen
- deviceprotection
- streamingtv
- streamingmovies
- partner
- customerid
- dependents
- onlinebackup
- phoneservice
- paymentmethod

totalcharges

monthlycharges

tenure

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

CLOUDERA Data Visualization

Dataset Detail

Related Dashboards

Fields

Dataset: user050 telco_data_curated

Dimensions

Measures

EDIT FIELDS UNDO REFRESH TITLE CASE SAVE Show Comments NEW DASHBOARD

telco_data_curated

- Dim | A - multipelines
- Dim | A - paperlessbilling
- Dim | A - gender
- Dim | A - onlinesecurity
- Dim | A - internetservice
- Dim | A - techsupport
- Dim | A - contract
- Dim | A - churn
- Dim | A - seniorcitizen
- Dim | A - deviceprotection
- Dim | A - streamingtv
- Dim | A - streamingmovies
- Dim | A - partner
- Dim | A - customerid
- Dim | A - dependents
- Dim | A - onlinebackup
- Dim | A - phoneservice
- Dim | A - paymentmethod

totalcharges

monthlycharges

tenure

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

The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes links for HOME, SQL, VISUALS, and DATA, along with a search bar and user authentication information. The left sidebar contains sections for Dataset Detail, Related Dashboards, Fields, Data Model, Time Modeling, Segments, Filter Associations, and Permissions. The main panel displays the 'Fields' section for the dataset 'user050.telco_data_curated'. It is divided into 'Dimensions' and 'Measures' sections. In the Dimensions section, there is a list of fields such as 'multiplelines', 'paperlessbilling', 'gender', 'onlinesecurity', etc. At the bottom of this list, there is a field named 'Copy of paymentmethod'. To the right of this field is a small black rectangular button with the white text 'Edit Field'.

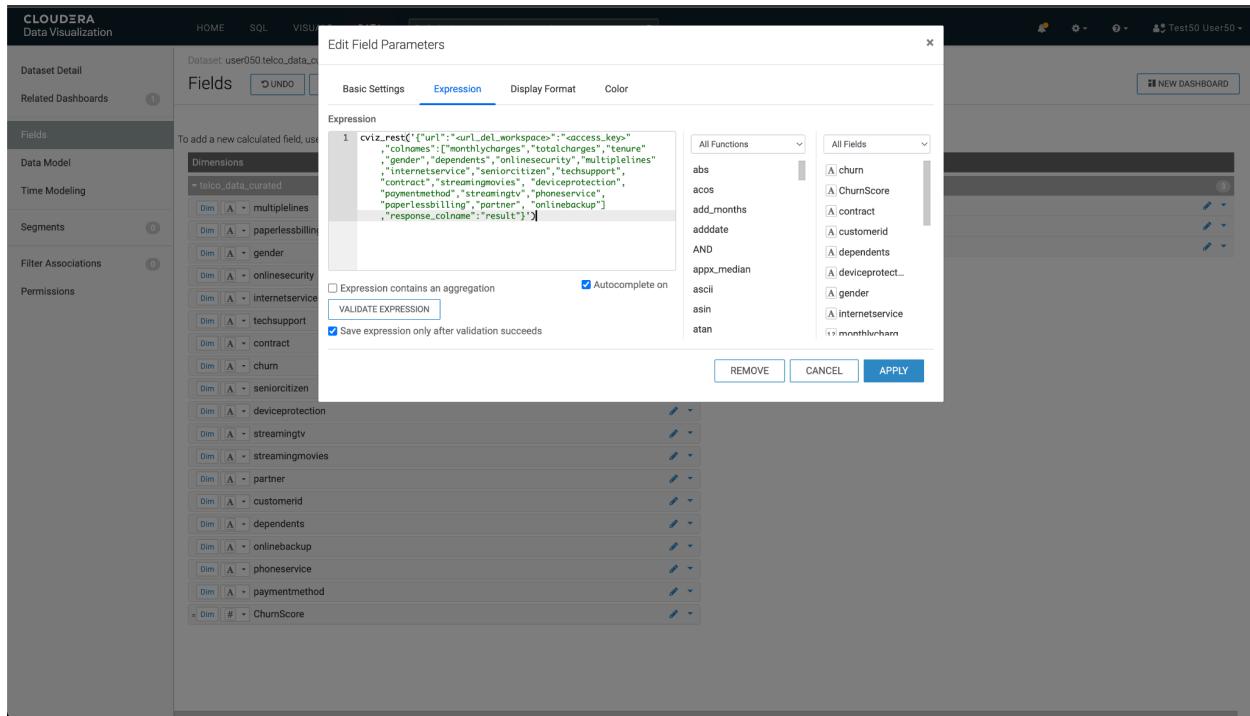
5. In the popup window that appears, enter the name of the new field in **Display Name**. We suggest that you enter *ChurnScore*.

The screenshot shows the 'Edit Field Parameters' dialog box overlaid on the Cloudera Data Visualization interface. The dialog has tabs for 'Basic Settings', 'Expression', 'Display Format', and 'Color'. The 'Basic Settings' tab is active, showing the 'Base Column' as 'paymentmethod'. The 'Display Name' field is filled with 'ChurnScore'. Other settings include 'Field Comment' (empty), 'Default Aggregation' (set to 'Maximum'), 'Geo Type' (set to 'None'), and checkboxes for 'Show field in data detail screen' (checked), 'Show field in Visual Designer' (checked), and 'Use as a partition column for Analytical Views' (unchecked). The 'Category' section has a radio button for 'Dimension' selected. At the bottom of the dialog are 'REMOVE', 'CANCEL', and 'APPLY' buttons.

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: Please make sure this is pasted as a single line. If not, you can either delete the lines or copy this from the github repo's README page.

```
cviz_rest('{"url":"<workspace_url>","accessKey":"<access_key>","colnames":["monthlycharges", "totalcharges", "tenure", "gender", "dependents", "onlinesecurity", "multiplelines", "internetservice", "seniorcitizen", "techsupport", "contract", "streamingmovies", "deviceprotection", "paymentmethod", "streamingtvtv", "phoneservice", "paperlessbilling", "partner", "onlinebackup"], "response_colname": "result"}')
```



7. Open CML in another tab of the web browser, go to the section of **Models** of your project, and click on the Model that begins with the name *ModelViz*, followed by your assigned username.

8. In the Overview tab, copy the URL that allows you to interact and call the workspace API.

Replace the copied value in the attribute <workspace_url> of the Expression field.

The screenshot shows the Cloudera Data Visualization interface with the 'Edit Field Parameters' dialog open. The 'Expression' tab is selected, showing the following CQL code:

```

1 cvtz_rest('{"url":"<url_of_workspace>","access_key":"
2 "cohort","monthlycharges","totalchurn","tenure"
3 "gender","dependents","partner","multiplelines"
4 "internetservice","seniorcitizen","techsupport",
5 "contract","streamingmovies","deviceprotection",
6 "paymentmethod","streamingtv","phoneservice",
7 "paperlessbilling","partner","onlinebackup"]
8 "response_colname":"result"}]

```

Below the code, there are two checkboxes: 'Expression contains an aggregation' (unchecked) and 'Autocomplete on VALIDATE EXPRESSION' (checked). A 'VALIDATE EXPRESSION' button is also present. At the bottom right of the dialog are 'REMOVE', 'CANCEL', and 'APPLY' buttons.

9. Returning to the CML, copy the accessKey of the model.

The screenshot shows the Cloudera Machine Learning interface with the 'ModelViz_user050' overview page open. The left sidebar shows navigation options like 'All Projects', 'Overview', 'Sessions', 'Data', 'Experiments', 'Models', 'Jobs', 'Applications', 'Files', 'Collaborators', and 'Project Settings'. The main content area shows the following details:

- Model Details:**
 - Source: Code
 - Model Id: 8
 - Model CRN: cm.cdp.ml.us-west-1:508fd88f-8076-498a-acfb-6f8765cd5e8 workspace:814194cb-1c7e-48cd-9989-b499a79ed5f6/dae534c1-b214-45eb-acd0-101e651f68d
 - Deployment Id: 10
 - Deployment CRN: cm.cdp.ml.us-west-1:508fd88f-8076-498a-acfb-6f8765cd5e8 workspace:814194cb-1c7e-48cd-9989-b499a79ed5f6/cf985a5d-9870-4533-9f9a-d42add0b56ed
 - Build Id: 10
 - Build CRN: cm.cdp.ml.us-west-1:508fd88f-8076-498a-acfb-6f8765cd5e8 workspace:814194cb-1c7e-48cd-9989-b499a79ed5f6/0e00e2d9-80cb-4ee8-8304-7998763d3e32
 - Deployed By: user050
 - Comment: Initial revision.
 - Runtime Image: Python 3.7 (Standard)
 - File: 13_model_viz.py
 - Function: predict
- Model Resources:**
 - Replicas: 1
 - Total CPU: 1 vCPUs
 - Total Memory: 2.00 GB

The 'Sample Code' section contains a 'curl' command for interacting with the model via its API endpoint. The 'Sample Response' section shows a JSON object: {}.

Replace the copied value in the attribute <access_key> of the Expression field. The format should be as follows, e.g.

```
cviz_rest('{"url":"https://modelservice.ml-b200bd6f-fb9.za-mtn-l.yu1t-vbzg.cloudera.site/model","accessKey":"mjy1fowabqiwpfjb19s9ht6xmuvy0f2j","colnames":["monthlycharges","totalcharges","tenure","gender","dependents","onlinesecurity","multiplelines","internetservice","seniorcitizen","techsupport","contract","streamingmovies","deviceprotection","paymentmethod","streamingtvtv","phoneservice","paperlessbilling","partner","onlinebackup"],"response_colname":"result"})
```

The screenshot shows the Cloudera Data Visualization interface. On the left, there's a sidebar with 'Dataset Detail' (Dataset: user050.telco_data.curated), 'Related Dashboards', 'Fields' (selected), 'Data Model', 'Time Modeling', 'Segments', 'Filter Associations', and 'Permissions'. The main area shows a list of dimensions under 'telco_data_curated': multiplelines, paperlessbilling, gender, onlinesecurity, internetservice, techsupport, contract, churn, seniorcitizen, deviceprotection, streamingtv, streamingmovies, partner, customerid, dependents, onlinebackup, phoneservice, paymentmethod, and ChurnScore. A modal window titled 'Edit Field Parameters' is open, showing the 'Expression' tab. The expression field contains the code: `1 cviz_rest('{"url":"https://modelservice.ml-b200bd6f-fb9.za-mtn-l.yu1t-vbzg.cloudera.site/model","accessKey":"mjy1fowabqiwpfjb19s9ht6xmuvy0f2j","colnames":["monthlycharges","totalcharges","tenure","gender","dependents","onlinesecurity","multiplelines","internetservice","seniorcitizen","techsupport","contract","streamingmovies","deviceprotection","paymentmethod","streamingtvtv","phoneservice","paperlessbilling","partner","onlinebackup"],"response_colname":"result"})`. Below the expression, there are three checkboxes: 'Expression contains an aggregation' (unchecked), 'Autocomplete on' (checked), and 'Save expression only after validation succeeds' (checked). To the right of the expression field are two dropdown menus: 'All Functions' and 'All Fields', each listing various data fields like churn, ChurnScore, contract, customerid, etc. At the bottom of the dialog are 'REMOVE', 'CANCEL', and 'APPLY' buttons.

10. 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.

CLOUDERA Data Visualization

HOME SQL VISUALS

Dataset Detail

Related Dashboards

Fields

Dimensions

telco_data_curated

multiplelines

paperlessbilling

gender

onlinesecurity

internetservice

techsupport

contract

churn

seniorcitizen

deviceprotection

streamingtv

streamingmovies

partner

customerid

dependents

onlinebackup

phoneservice

paymentmethod

ChurnScore

Dataset user050.telco_data_curated

Expression

cviz_restC("url":"https://modellservice.ml-369083c3-99e.ssa-hol.yuit-vbzg.cloudera.site/nodl","accessKey":"mmmlolikuw47019gwerd4h4m939cd2","columnes":[{"mem":1000000,"total":1000000,"tenant":1,"gender":1,"dependents":1,"onlinebackup":1,"seniorcitizen":1,"internetservice":1,"seniorcitizen":1,"techsupport":1,"contract":1,"streamingmovies":1,"deviceprotection":1,"paymentmethod":1,"streamingtv":1,"phoneservice":1,"paperlessbilling":1,"partner":1,"onlinebackup":1,"response_colname":1,"result":1})

Expression contains an aggregation

Autocomplete on

VALIDATE EXPRESSION

Save expression only after validation succeeds

Validation Successful

REMOVE CANCEL APPLY

NEW DASHBOARD

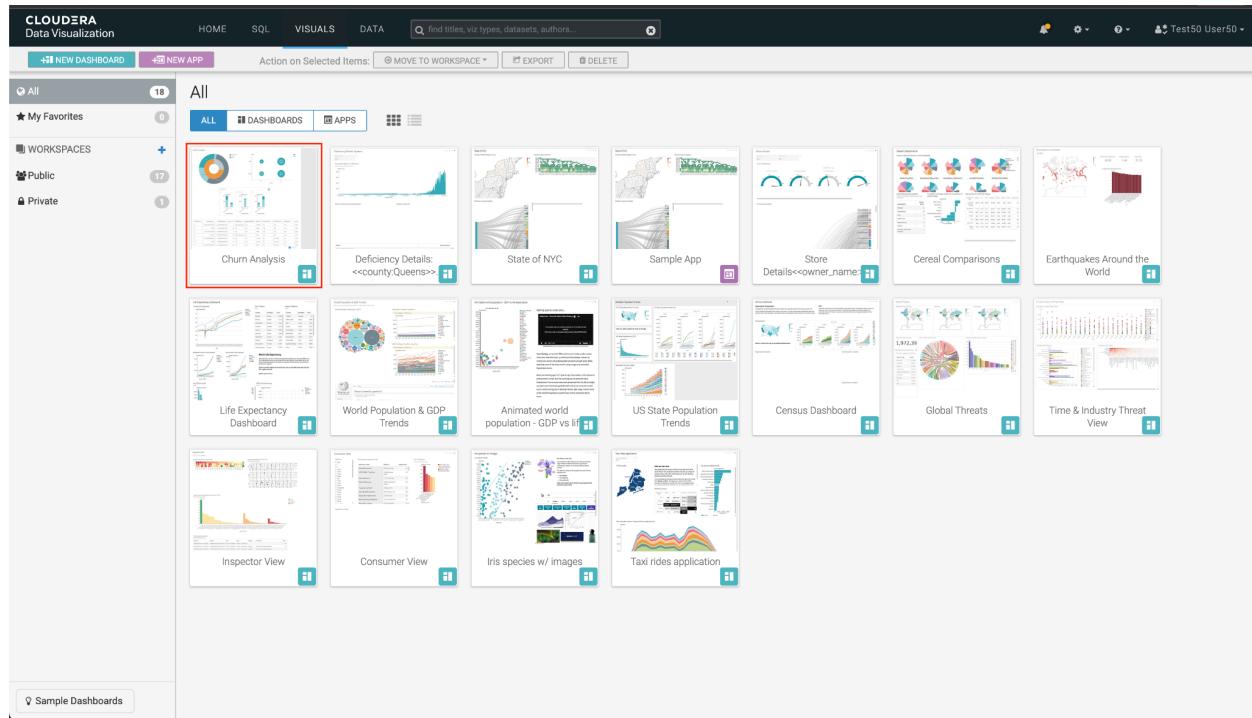
11. The new field should appear in the list of fields. Change the data type, selecting the type *Integer*, which is represented by the symbol #

The screenshot shows the Cloudera Data Visualization interface. The left sidebar has sections for Dataset Detail, Related Dashboards, Fields, Data Model, Time Modeling, Segments, Filter Associations, and Permissions. The main area is titled 'Fields' and shows a dataset named 'user050 telco_data_curated'. It includes tabs for Dimensions and Measures. Under Dimensions, there is a list of fields from the 'telco_data_curated' table, including 'multilines', 'paperlessbilling', 'gender', 'onlinesecurity', 'internetservice', 'techsupport', 'contract', 'churn', 'seniorcitizen', 'deviceprotection', 'streamingtv', 'streamingmovies', 'partner', and several dimensions for Boolean, Integer, Real, String, Timestamp, and Remove CAST. A new field, 'ChurnScore', is being added at the bottom of the list. The 'Measures' section lists 'totalcharges', 'monthlycharges', and 'tenure'. At the top of the Fields page, there are buttons for UNDO, REFRESH, TITLE CASE, SAVE, and Show Comments, along with a search bar and a 'NEW DASHBOARD' button.

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

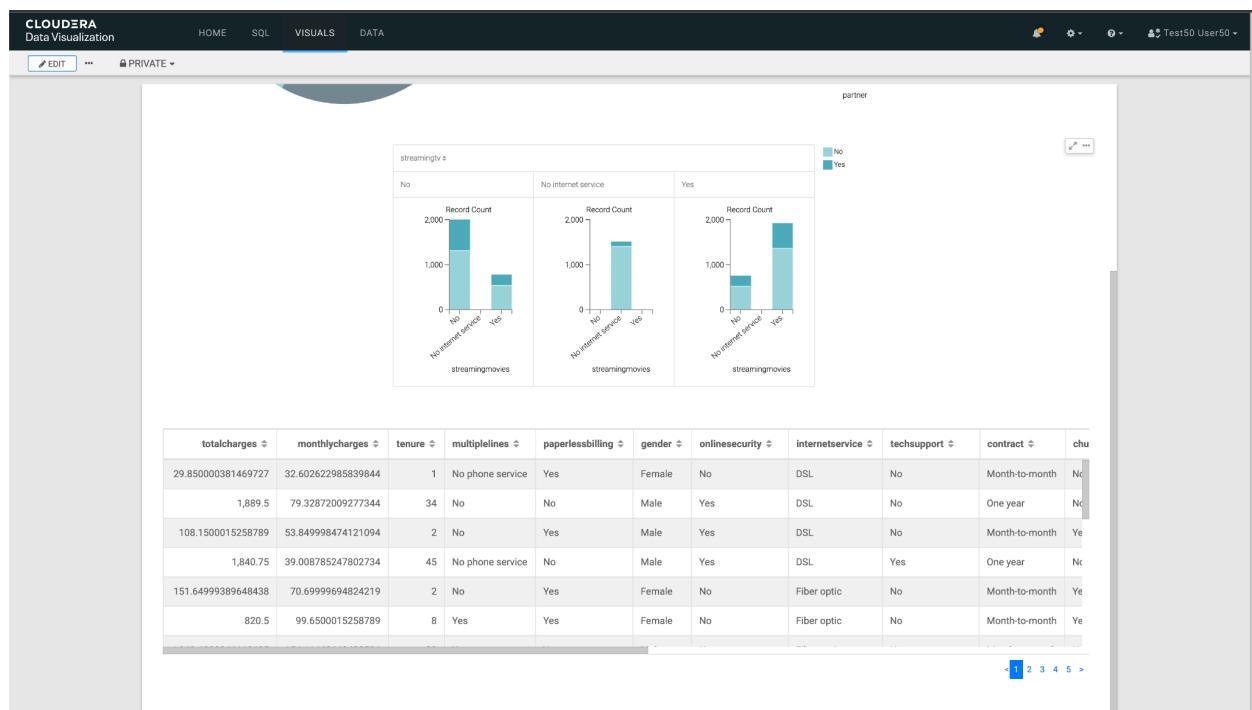
This screenshot shows the same Cloudera Data Visualization interface after the 'SAVE' button was clicked. The 'Fields' page now displays the newly created 'ChurnScore' field in the list of dimensions. The rest of the interface remains the same, with the dataset 'user050 telco_data_curated' selected and the 'Dimensions' and 'Measures' sections visible. The 'SAVE' button is no longer highlighted in green.

13. Return to the dashboard by selecting the option **VISUALS** from the top menu, and clicking on the name of the dashboard that was previously created.



The screenshot shows the Cloudera Data Visualization interface. At the top, there's a navigation bar with links for HOME, SQL, VISUALS, and DATA, along with a search bar and user information. Below the navigation is a toolbar with buttons for NEW DASHBOARD, NEW APP, MOVE TO WORKSPACE, EXPORT, and DELETE. On the left, there's a sidebar with sections for All, My Favorites, Workspaces (Public and Private), and Sample Dashboards. The main area displays a grid of 18 dashboard cards. One card, titled 'Churn Analysis', is highlighted with a red box. Other visible dashboard titles include 'Deficiency Details', 'State of NYC', 'Sample App', 'Store Details', 'Cereal Comparisons', 'Earthquakes Around the World', 'Life Expectancy Dashboard', 'World Population & GDP Trends', 'Animated world population - GDP vs HI', 'US State Population Trends', 'Census Dashboard', 'Global Threats', 'Time & Industry Threat View', 'Inspector View', 'Consumer View', 'Iris species w/ images', and 'Taxi rides application'. Each card shows a preview of the dashboard's contents.

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



The screenshot shows a dashboard view within the Cloudera Data Visualization interface. At the top, there's a header with the Cloudera logo, navigation links for HOME, SQL, VISUALS, and DATA, and a lock icon indicating the dashboard is private. To the right of the lock icon is the user information 'Test50 User50'. Below the header, there's an 'EDIT' button and a 'PRIVATE' dropdown. The main content area contains three bar charts side-by-side, each with 'Record Count' on the y-axis and two categories on the x-axis ('No internet service' and 'Yes'). The charts are titled 'streamingtv', 'streamingmovies', and 'streamingmoves'. Below the charts is a detailed data table with 10 columns: totalcharges, monthlycharges, tenure, multiplelines, paperlessbilling, gender, onlinesecurity, internetservice, techsupport, contract, and chu. The table has 10 rows of data. At the bottom of the table, there are navigation arrows for pagination.

totalcharges	monthlycharges	tenure	multiplelines	paperlessbilling	gender	onlinesecurity	internetservice	techsupport	contract	chu
29.85000381469727	32.602622985839844	1	No phone service	Yes	Female	No	DSL	No	Month-to-month	NK
1,889.5	79.32872009277344	34	No	No	Male	Yes	DSL	No	One year	NK
108.1500015258789	53.849998474121094	2	No	Yes	Male	Yes	DSL	No	Month-to-month	Ye
1,840.75	39.008785247802734	45	No phone service	No	Male	Yes	DSL	Yes	One year	Nc
151.64999389648438	70.69999694824219	2	No	Yes	Female	No	Fiber optic	No	Month-to-month	Ye
820.5	99.6500015258789	8	Yes	Yes	Female	No	Fiber optic	No	Month-to-month	Ye

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

The screenshot shows the Cloudera Data Visualization interface. On the left, there are three stacked bar charts under the title "streamingtv". The first chart is for "No internet service", the second for "No streamingmovies", and the third for "Yes streamingmovies". Each chart has two bars: "No" (light blue) and "Yes" (dark blue). The y-axis is labeled "Record Count" and ranges from 0 to 2,000. The x-axis is labeled "streamingmovies". A legend indicates "No" is light blue and "Yes" is dark blue.

On the right, there is a large table with various columns including "totalcharges", "monthlycharges", "tenure", etc. The "contract" column shows values like "Month-to-month", "One year", and "Month-to-month".

The right side of the interface features a vertical "Build" menu with several sections:

- Dimensions**: Contains a list of fields including "ChurnScore" (highlighted with a red box), "totalcharges", "monthlycharges", "tenure", "multiplelines", "paperlessbilling", "gender", "onlinesecurity", "internetservice", "techsupport", and "contract".
- Measures**: Contains "Record Count", "totalcharges", "monthlycharges", and "tenure".
- Visuals**: Contains "Search", "Filters", "Style", and "Settings".
- Dashboard Designer**: Contains "Sample Mode OFF", "Search", "Dimensions", "Measures", "Visuals", and "Build".

A "REFRESH VISUAL" button is located at the bottom right of the interface.

16. 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.

The screenshot shows the Cloudera Data Visualization interface. On the left, there are three stacked bar charts under the heading "streamingtv". The first chart is for "No internet service", the second for "No", and the third for "Yes". All three charts have "Record Count" on the y-axis (0 to 2,000) and "streamingmovies" on the x-axis. The legend indicates "No" in light blue and "Yes" in dark blue. The middle section contains a table with the following data:

	totalcharges	monthlycharges	tenure	multipelines	paperlessbilling	gender	onlinesecurity	internetservice	techsupport
0	29.850000381469727	32.602622985839844	1	No phone service	Yes	Female	No	DSL	No
0	1,889.5	79.32872009277344	34	No	No	Male	Yes	DSL	No
0	108.1500015258789	53.849998474121094	2	No	Yes	Male	Yes	DSL	No
0	1,840.75	39.008785247802734	45	No phone service	No	Male	Yes	DSL	Yes
6	151.64999389648438	70.6999694824219	2	No	Yes	Female	No	Fiber optic	No
10	820.5	99.6500015258789	8	Yes	Yes	Female	No	Fiber optic	No

The right side of the interface shows the "Dashboard Designer" panel with sections for DATA, Dimensions, Measures, and Visuals. The "Dimensions" section lists various fields like "ChurnScore", "totalcharges", "monthlycharges", "tenure", etc. The "Measures" section lists "Record Count", "totalcharges", "monthlycharges", and "tenure". The "Visuals" section shows the three bar charts. A "REFRESH VISUAL" button is located at the bottom right of the visualization area.

End of Lab 5