

Data Lifecycle in CDP Public Cloud

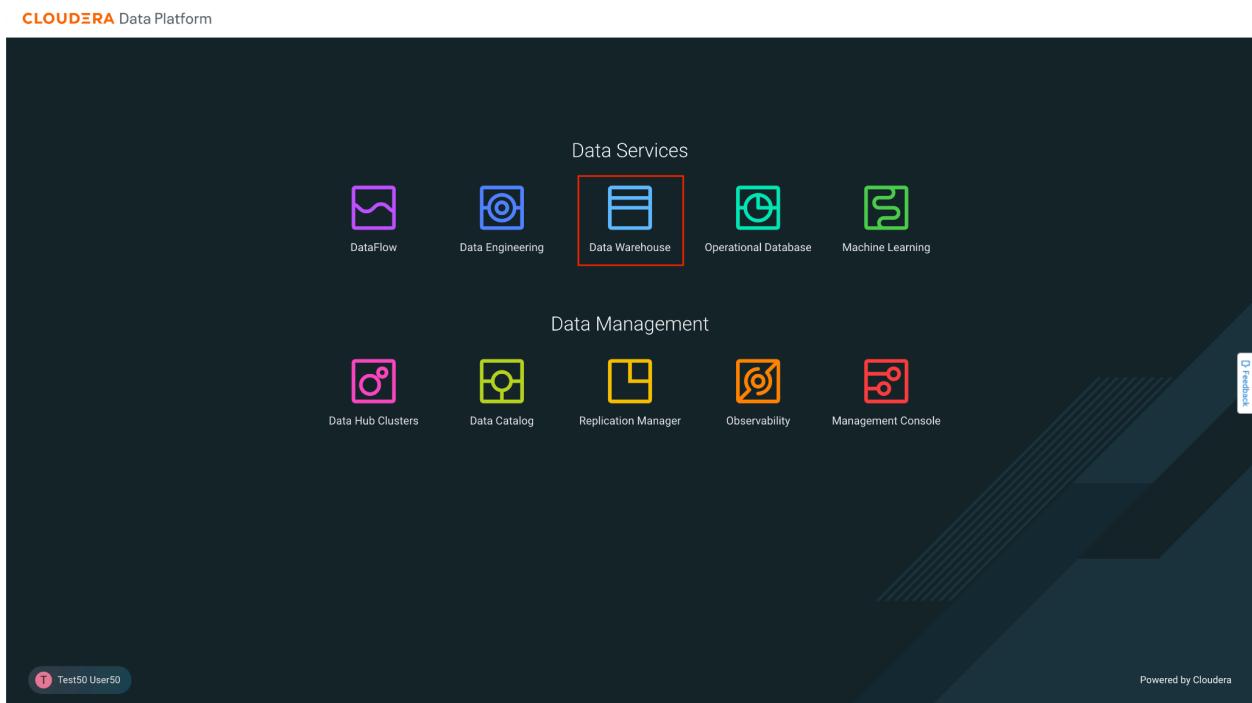
Data Warehouse Lab

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 Overview page. On the left sidebar, under the 'Data Visualization' section, there is a 'Data Viz' button. The main content area displays various data visualization components:

- Get started with Data Warehouse:** A callout box with 'See More' links for 'Start Guide' and 'CDP Patterns'.
- Create:** A box for creating new environments, databases, catalogs, and virtual warehouses, with a 'See More' link.
- Query and visualize data:** A box for running SQL queries and creating reports, with a 'See More' link.
- Virtual Warehouses:** A list showing two entries:
 - ssa-datalake-de...**: Running, environment ID 1685386885-5ndk, with 9 cores, 25 GB memory, and 2 virtual warehouses.
 - impara-vw-0**: Running, environment ID 1685398129-7pxf, with 2 executors, 63 cores, 476 GB memory, and IMPALA type.
- hive-vw-0**: Stopped, environment ID 1685387629-w2rw, with 0 executors, 12 cores, 56 GB memory, and HIVE, UNIFIED ANALYTICS, COMPACTOR types.

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

The screenshot shows the Data Visualization page. On the left sidebar, under the 'Data Visualization' section, there is a 'Data Viz' button, which is highlighted with a red box. The main content area displays a table of data visualizations:

NAME	DATA VISUALIZATION ID	Environment ID	VERSION	CPU	MEMORY	UPTIME	CREATED BY	Actions
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 same Data Visualization interface, but the 'Datasets' tab is now active, indicated by a blue underline. A red box highlights the 'NEW DATASET' button in the top navigation bar. Below the navigation bar, there is a 'Connection Explorer' section. The main content area displays a table with the same columns as before ('Title/Table', 'ID', 'Created', 'Last Updated', 'Modified By', and '# Dashboards'), but it shows 'No data'.

6. Enter the information for the new data source:

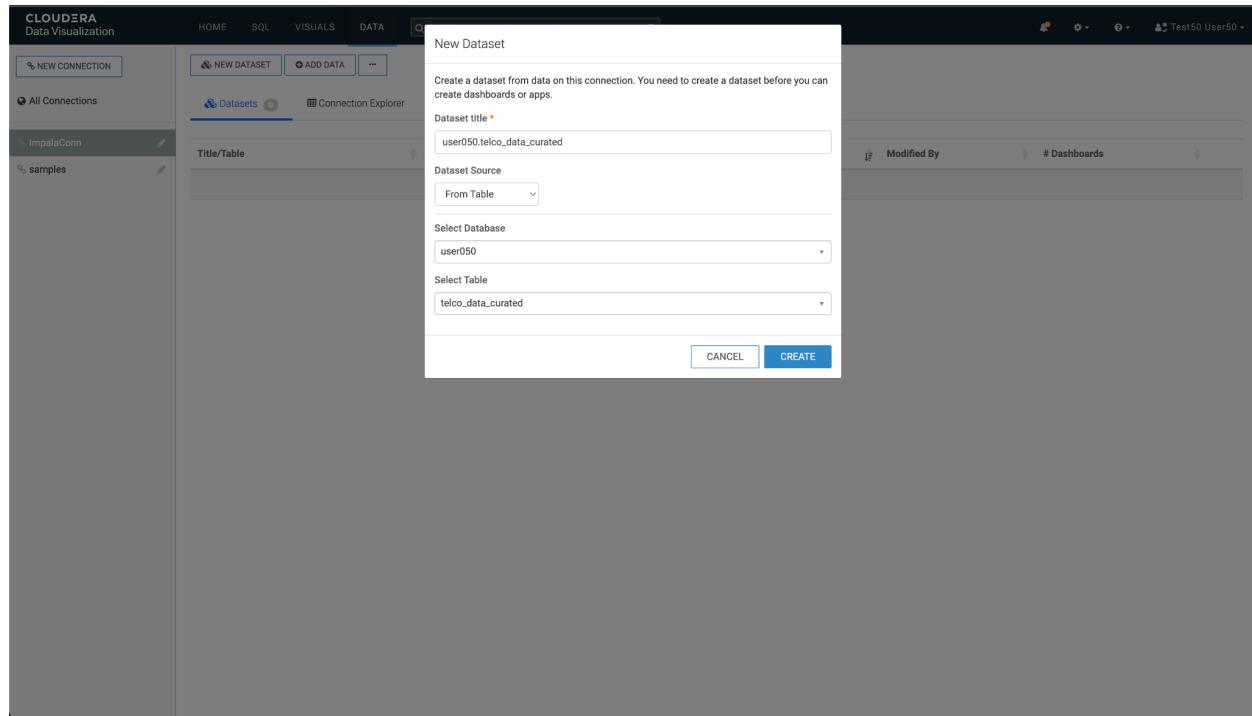
Dataset title: <assigned_user>.telco_curated_data

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 a user interface for managing datasets. On the left, there's a sidebar with connection management options like 'NEW CONNECTION', 'All Connections', and a selected 'ImpalaConn' which has a 'samples' folder. The main area is titled 'Datasets' and contains a table with columns: Title/Table, ID, Created, Last Updated, Modified By, and # Dashboards. There are two entries: 'user050.telco_data_curated' (ID: 16, Created: May 29, 2023, Last Updated: a few seconds ago, Modified By: user050, # Dashboards: 0).

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

The screenshot shows the 'Dataset Detail' page for the dataset 'user050.telco_data_curated'. The left sidebar lists sections: Dataset Detail (selected), Related Dashboards, Fields, Data Model, Time Modeling, Segments (0), Filter Associations (0), and Permissions. The main content area shows dataset details: 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, it shows metadata: 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, there are buttons for 'CLONE DATASET' and 'NEW DASHBOARD'.

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. On the left, there's a sidebar with various options like Dataset Detail, Related Dashboards, Fields (which is currently selected), Data Model, Time Modeling, Segments, Filter Associations, and Permissions. The main area is titled "Dataset: user050 telco_data_curated". It has tabs for HOME, SQL, VISUALS, and DATA. A search bar at the top right says "find titles, viz types, datasets, authors...". Below the search bar are buttons for "EDIT FIELDS" and "Hide Comments", and a "NEW DASHBOARD" button. The "Fields" section is divided into two main sections: "Dimensions" and "Measures". The "Dimensions" section contains 18 items: 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: totalcharges, monthlycharges, and tenure.

Dimensions	
▼ telco_data_curated	(18)
A multiplelines	
A paperlessbilling	
A gender	
A onlinesecurity	
A internetservice	
A techsupport	
A contract	
A churn	
A seniorcitizen	
A deviceprotection	
A streamingtv	
A streamingmovies	
A partner	
A customerid	
A dependents	
A onlinebackup	
A phoneservice	
A paymentmethod	

Measures	
▼ telco_data_curated	(3)
I totalcharges	
I monthlycharges	
I tenure	

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'.

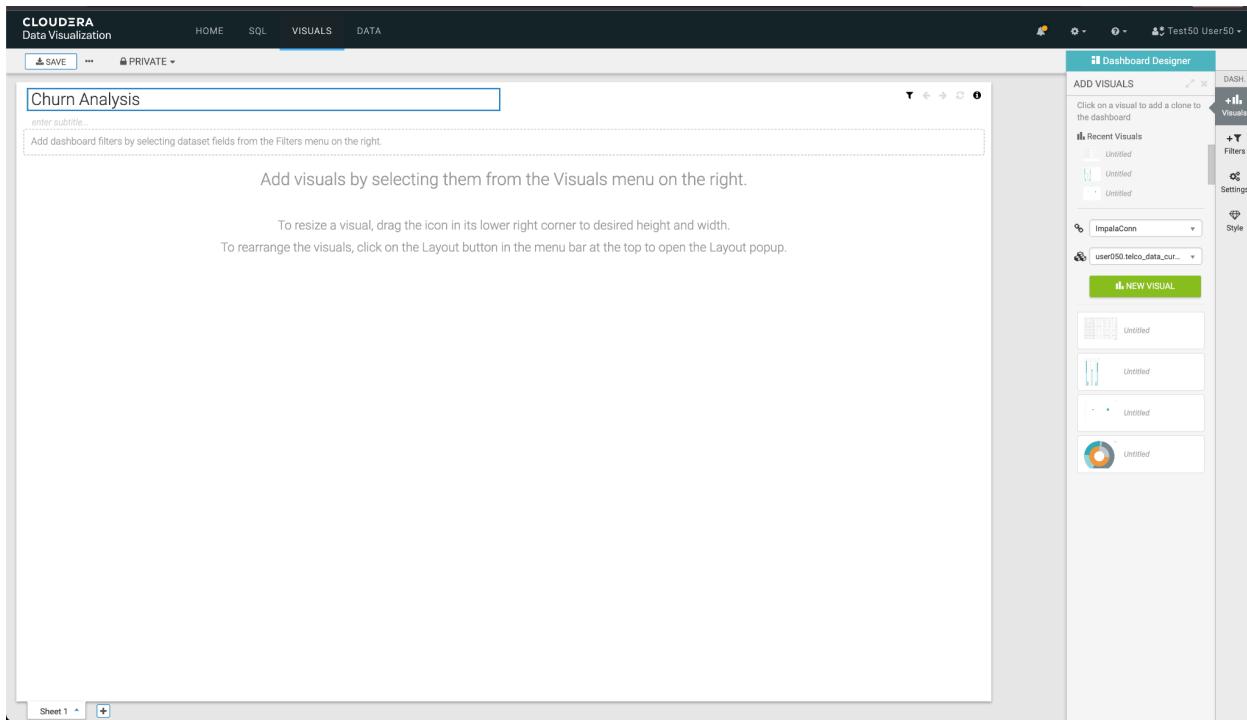
11. At this moment, a query to the Virtual Warehouse is executed to retrieve the data from the data set. Notice the columns and values. Click New Dashboard to create a new dashboard.

The screenshot shows the same Cloudera Data Visualization interface as the previous one, but now the 'telco_data_curated' dataset is fully loaded. The 'NEW DASHBOARD' button is highlighted with a red box. Below the table, several rows of data are visible, including columns like 'multiplelines', 'paperlessbilling', 'gender', 'onlinesecurity', 'internetservice', 'techsupport', 'contract', 'churn', 'seniorcitizen', 'deviceprotection', 'streamingtv', 'streamingmovies', 'totalcharges', 'partner', 'monthlycharges', 'customerid', and 'dept'.

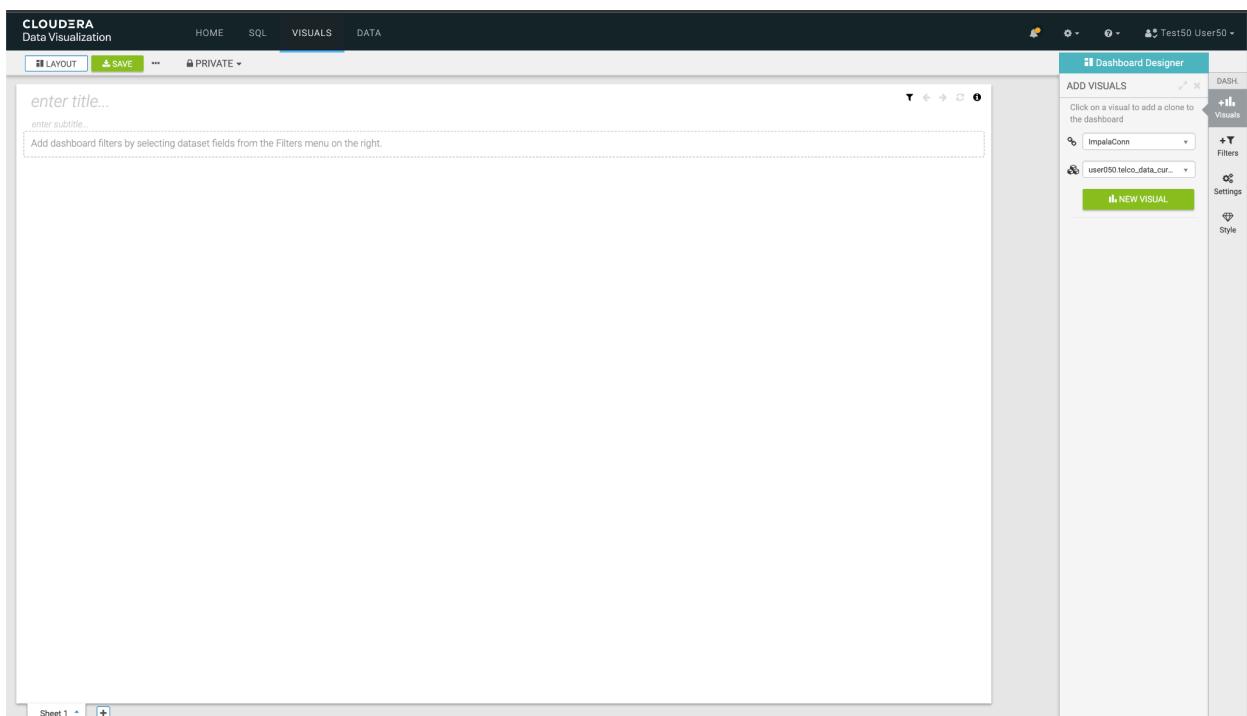
12. When opening the design canvas of a new panel, remove the element that is added by default, by clicking on the three dots (...) button at the top right of the element, and then clicking on the option **Delete Visual**

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.

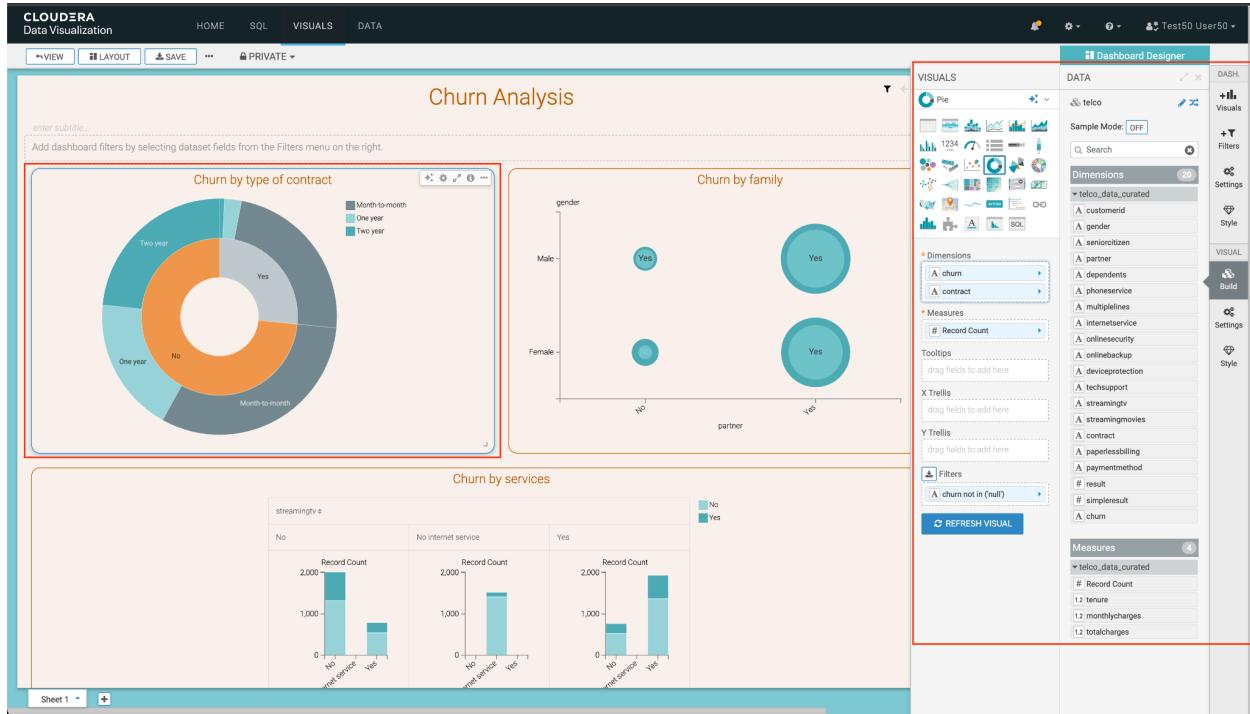
At the top of the canvas, in the enter title field, enter the name *Churn Analysis-<user_assigned>* to identify the dashboard.



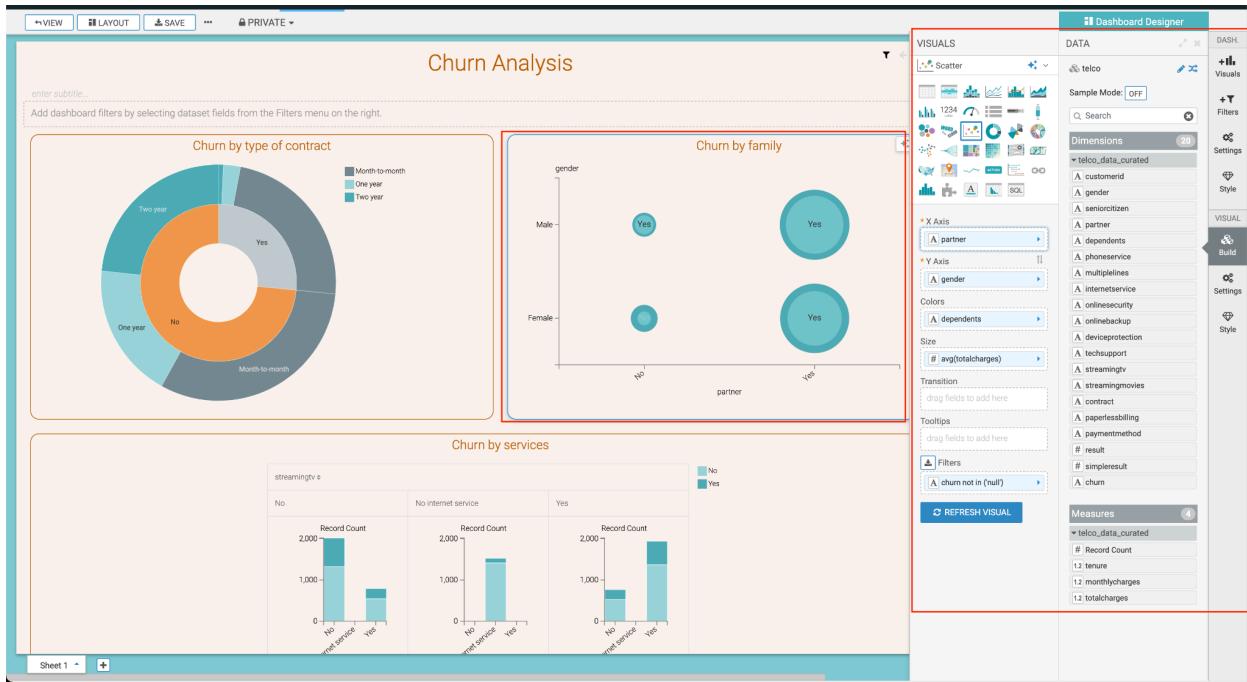
13. To add a new visual element, click on the button **Visuals** from the right menu, select the dataset that corresponds to them (telco_data_curated), and click on the button **New Visual**.



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 options like HOME, SQL, VISUALS, and DATA. Below the navigation is a header bar with buttons for VIEW, LAYOUT, SAVE, and PRIVATE mode.

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" ranging from 0 to 2,000. The bars are colored by "churn" (No is light blue, Yes is teal).
- 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 bar chart's configuration panel, which includes sections for DATA, Dimensions, X Axis, Y Axis, Colors, and Filters. 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 to the Dimensions section and 3 metrics (Measures) to the Measures section of the table. Once finished, click the button **Refresh Visual**.

The screenshot shows the Cloudera Data Visualization interface. At the top, there are three bar charts side-by-side, each titled "Record Count". The first chart has categories "No streamingmovies" and "Yes streamingmovies" with values approximately 1800 and 200 respectively. The second chart has categories "No internet service" and "Yes" with values approximately 1800 and 200 respectively. The third chart has categories "No streamingmovies" and "Yes" with values approximately 1800 and 200 respectively. Below these charts is a table titled "Scoring - Churn Probability". The table has columns for customerid, tenure, monthlycharges, totalcharges, gender, dependents, onlinesecurity, multipipelines, internetservice, and seniorcitizen. The data rows are as follows:

customerid	tenure	monthlycharges	totalcharges	gender	dependents	onlinesecurity	multipipelines	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

The right side of the interface shows the "Dashboard Designer" panel with sections for DATA, Dimensions, Measures, and Settings. A red box highlights the "Scoring - Churn Probability" table.

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

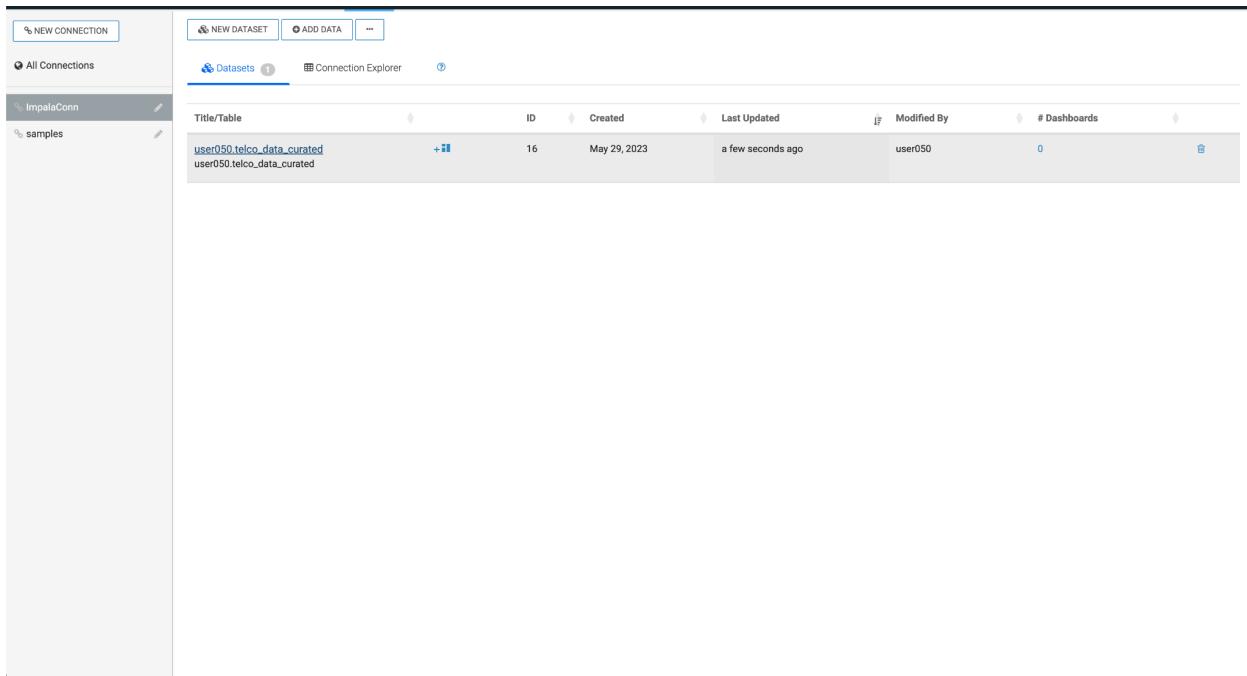
End of Part 1 - we will come back to run Part 2 later

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' dataset. 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 datasets are listed: 'user050.telco_data_curated' (ID: 16, Created: May 29, 2023, Last Updated: a few seconds ago, Modified By: user050, # Dashboards: 0) and 'user050.telco_data_curated' (ID: 16, Created: May 29, 2023, Last Updated: a few seconds ago, Modified By: user050, # Dashboards: 0). The second row is highlighted in blue.

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	16	May 29, 2023	a few seconds ago	user050	0

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.

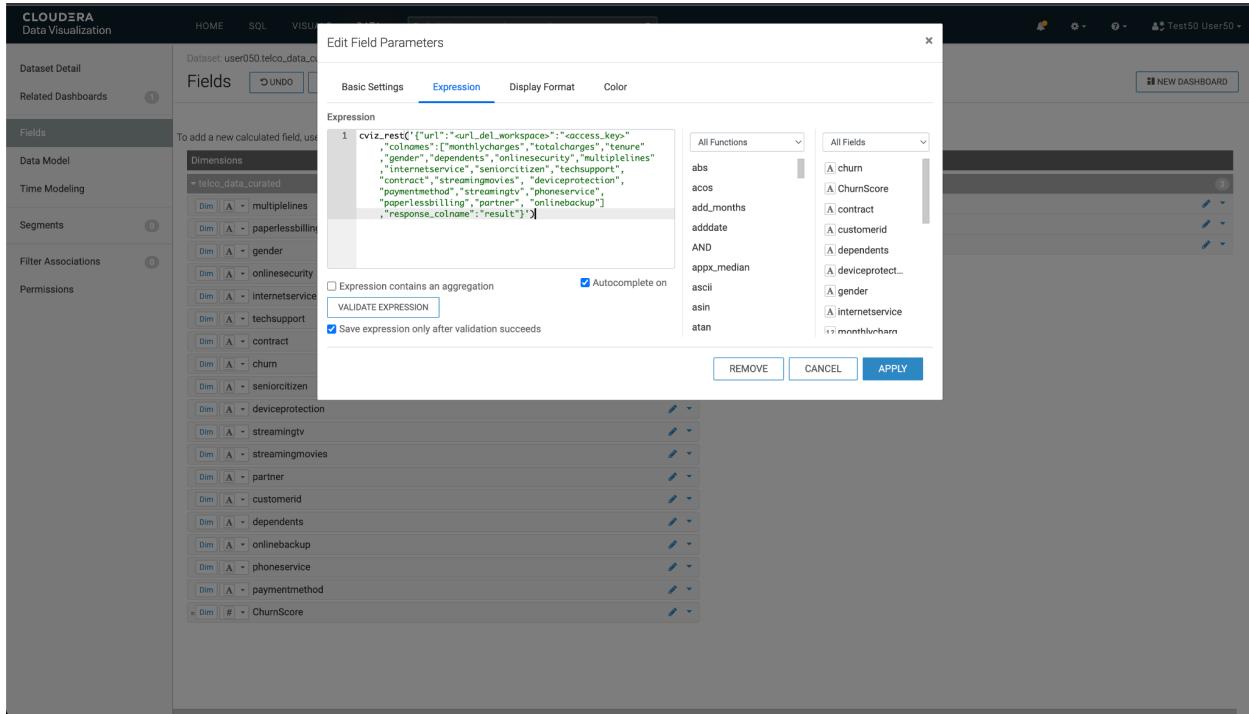
The screenshot shows the Cloudera Data Visualization interface. The top navigation bar includes HOME, SQL, VISUALS, DATA, and a search bar. The left sidebar has sections for Dataset Detail, Related Dashboards, Fields, Data Model, Time Modeling, Segments, Filter Associations, and Permissions. The main area is titled 'Dataset: user050:telco_data_curated' and shows the 'Fields' section. It displays two columns: 'Dimensions' and 'Measures'. In the Dimensions column, there is a list of fields under 'telco_data_curated', including 'multipleinles', 'paperlessbilling', 'gender', 'onlinesecurity', 'internetservice', 'techsupport', 'contract', 'churn', 'seniorcitizen', 'deviceprotection', 'streamingtv', 'streamingmovies', 'partner', 'customerid', 'dependents', 'onlinebackup', 'phoneservice', 'paymentmethod', and 'Copy of paymentmethod'. The 'Copy of paymentmethod' field is highlighted with a blue selection bar and has a small edit icon ('Edit Field') next to it. The Measures column contains 'totalcharges', 'monthlycharges', and 'tenure'.

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 overlying the Cloudera Data Visualization interface. The dialog has tabs for 'Basic Settings', 'Expression', 'Display Format', and 'Color'. The 'Basic Settings' tab is active, showing 'Base Column: paymentmethod'. The 'Display Name' field is set to 'ChurnScore'. The 'Field Comment' field contains 'Enter field comment'. Under 'Default Aggregation', 'Maximum' is selected. Under 'Geo Type', 'None' is selected. There are three checked checkboxes at the bottom: 'Show field in data detail screen', 'Show field in Visual Designer', and 'Use as a partition column for Analytical Views'. 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.

```
cviz_rest('{"url":"<url_del_workspace>","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. Being in 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 <url_del_workspace> of the Expression field.

The screenshot shows the Cloudera Data Visualization interface. A modal window titled 'Edit Field Parameters' is open, specifically for an 'Expression' field. The expression itself is a curl command:

```
curl -H "Content-Type: application/json" -X POST https://modelerService.ml-369883c-99e.ssa-hol.yuit-vbzg.cloudera.site/model/d -d '{"accessKey": "urn:la1kvu47a19guwrd4xhm9w3k9cb", "request": {"date": {"colnames": ["monthlycharges"], "totalcharge": "tenure", "gender": "dependents", "onlineservice": "seniorcitizen", "techsupport": "contract", "streamingmovies", "deviceprotection", "paymentmethod", "streamingtv", "phoneservice", "paperlessbilling", "partner", "onlinebackup"], "response_colname": "result"}}
```

The 'Expression' tab is selected in the top navigation bar. Below it, there are tabs for 'Basic Settings', 'Display Format', and 'Color'. On the right side of the dialog, there are dropdown menus for 'All Functions' and 'All Fields', both currently set to 'churn'. At the bottom, there are three buttons: 'REMOVE', 'CANCEL', and 'APPLY', with 'APPLY' being the active one.

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

The screenshot shows the Cloudera Machine Learning interface. On the left, there's a sidebar with various project management options like 'All Projects', 'Overview', 'Sessions', 'Data', 'Experiments', 'Models', 'Jobs', 'Applications', 'Files', 'Collaborators', and 'Project Settings'. The 'Models' option is selected.

The main area is titled 'ModelViz_user050 / Overview'. It shows the 'Model Details' section with information like Source (Code), Model Id (8), and Model CRN (cm.cdp.ml/us-west-1:508fd88f-8076-498a-acfb-6f8765cd5e8workspace814194cb-1c7e-48cd-9989-b499a79ed5f6/dae534c1-b214-45eb-acd0-101e651f68d). It also shows Deployment details (Deployment Id 10, Deployment CRN cm.cdp.ml/us-west-1:508fd88f-8076-498a-acfb-6f8765cd5e8workspace814194cb-1c7e-48cd-9989-b499a79ed5f6/cf985a5d-9870-4533-9f9a-d42add0b56ed, Build Id 10, Build CRN cm.cdp.ml/us-west-1:508fd88f-8076-498a-acfb-6f8765cd5e8workspace814194cb-1c7e-48cd-9989-b499a79ed5f6/0e00e2d9-80cb-4ee8-8304-7998763d3e32).

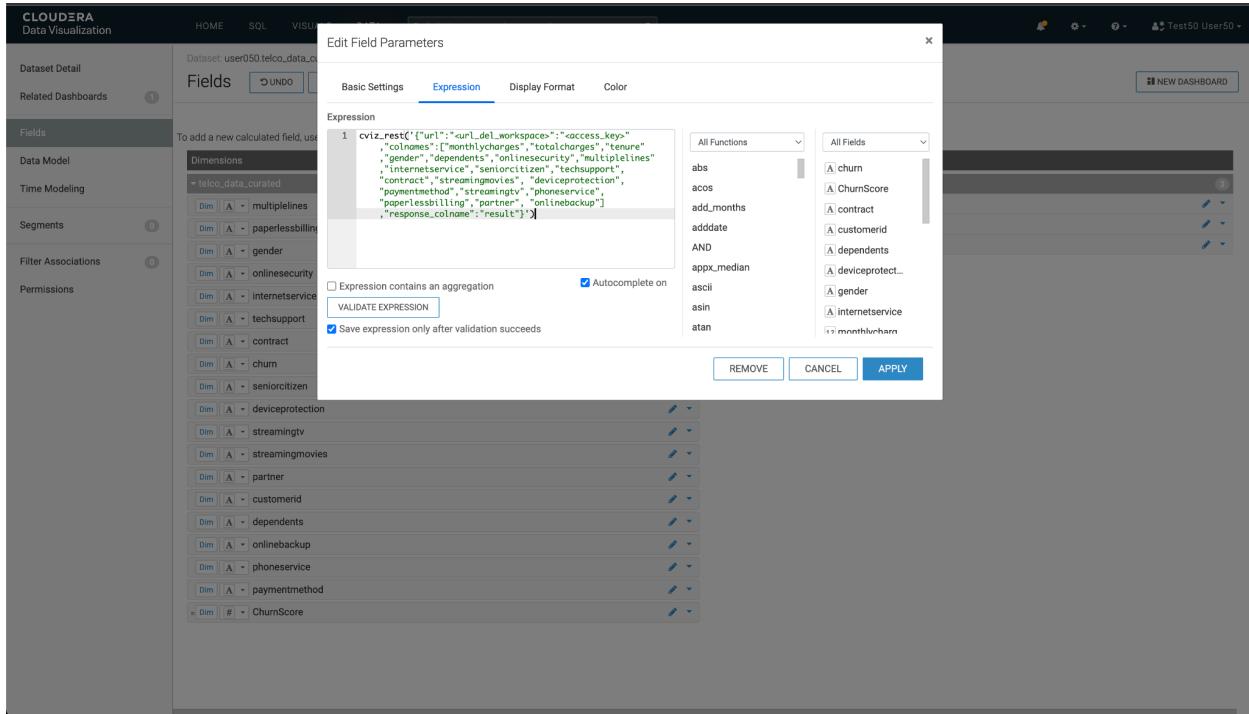
The 'Sample Code' section contains a 'Shell' tab with a curl command to interact with the model. The command uses the copied accessKey value:

```
curl -H "Content-Type: application/json" -X POST https://modelerService.ml-369883c-99e.ssa-hol.yuit-vbzg.cloudera.site/model/d -d '{"accessKey": "urn:la1kvu47a19guwrd4xhm9w3k9cb", "request": {"date": {"colnames": ["monthlycharges"], "totalcharge": "tenure", "gender": "dependents", "onlineservice": "seniorcitizen", "techsupport": "contract", "streamingmovies", "deviceprotection", "paymentmethod", "streamingtv", "phoneservice", "paperlessbilling", "partner", "onlinebackup"], "response_colname": "result"}}
```

The 'Test Model' section at the bottom shows the workspace as 'ssa-cml-workspace' and the cloud provider as 'aws (AWS)'.

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"})
```



10. Finish the process of copying the *url del workspace* 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.

The screenshot shows the Cloudera Data Visualization interface. In the center, a modal dialog titled "Edit Field Parameters" is open, specifically on the "Expression" tab. The expression input field contains a complex JSON object. Below the input field are two checkboxes: "VALIDATE EXPRESSION" (which is checked) and "Save expression only after validation succeeds". At the bottom of the dialog, a green bar displays the message "Validation Successful!". To the right of the dialog, there are dropdown menus for "All Functions" and "All Fields", each listing various data fields. The background of the main interface shows a dataset detail view with tabs for HOME, SQL, VIZ, and a list of related dashboards.

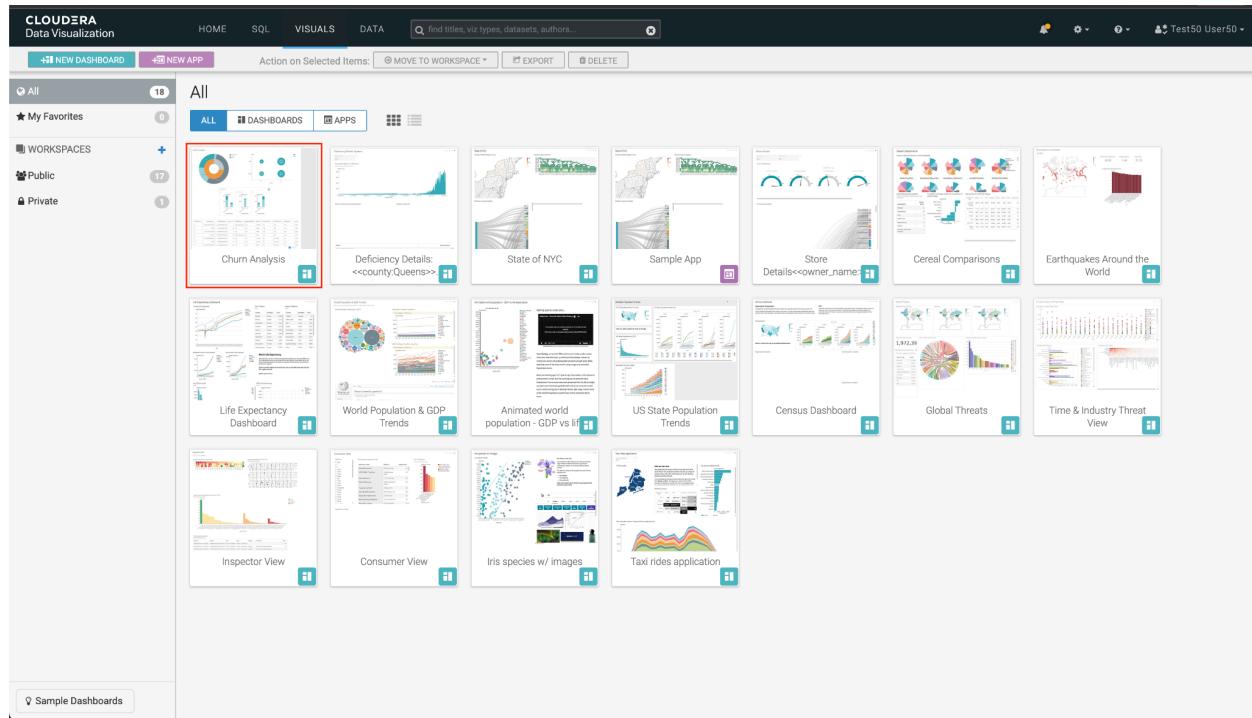
11. The new field should appear in the list of fields. Change the data type, selecting the type *Integer*, which is represented by the symbol #

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.

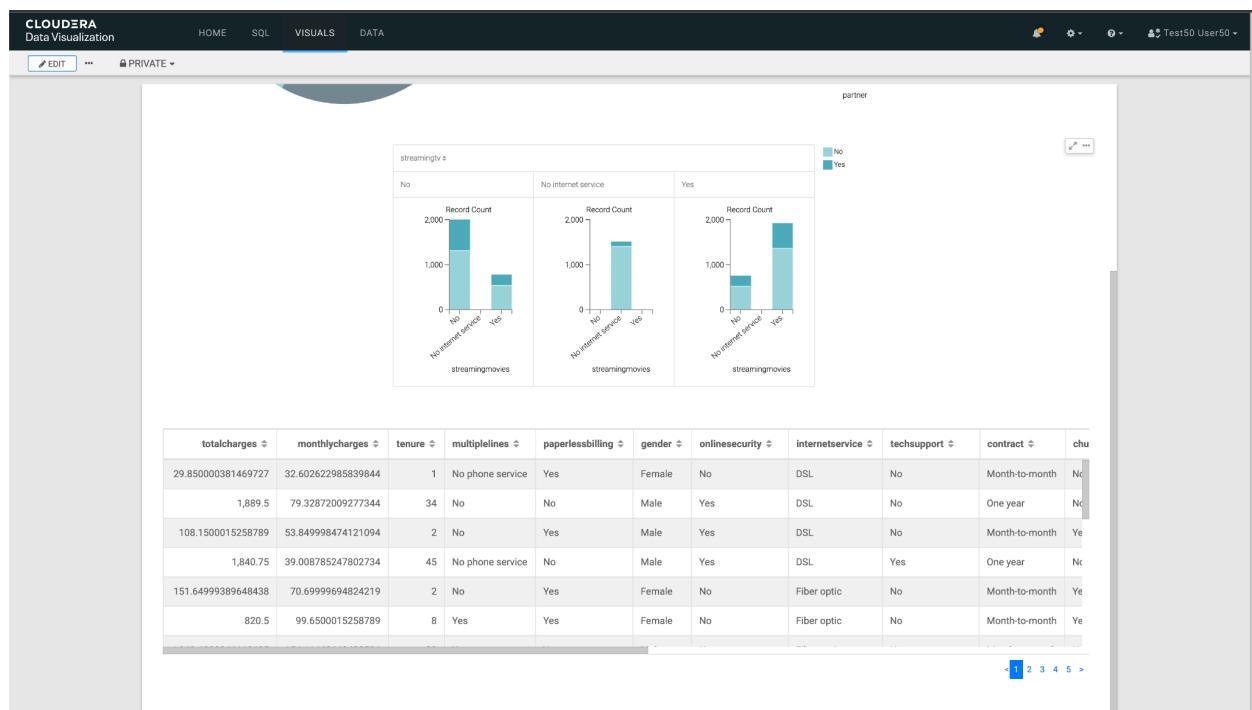
This screenshot shows the same Cloudera Data Visualization interface after the 'SAVE' button was clicked. The 'Fields' page now displays the newly created 'ChurnScore' field in the list of dimensions. The rest of the interface remains the same, with the dataset 'user050 telco_data_curated' selected and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

13. Return to the dashboard, 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 'HOME', 'SQL', 'VISUALS', and 'DATA' tabs. Below the navigation bar is a search bar and a toolbar with buttons for 'MOVE TO WORKSPACE', 'EXPORT', and 'DELETE'. On the left, there's a sidebar with sections for 'All', 'My Favorites', 'WORKSPACES' (Public and Private), and a 'Sample Dashboards' link. 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'.

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



The screenshot shows the 'Edit' view of the 'streamingtv' dashboard. At the top, there are buttons for 'EDIT' and 'PRIVATE'. The dashboard itself has three bar charts under the heading 'No internet service'. The first chart shows 'Record Count' for 'streamingmoves' with categories 'No' and 'Yes'. The second chart shows 'Record Count' for 'streamingmovies' with categories 'No' and 'Yes'. The third chart shows 'Record Count' for 'streamingmoves' with categories 'No' and 'Yes'. Below these charts is a detailed data table with columns: 'totalcharges', 'monthlycharges', 'tenure', 'multiplelines', 'paperlessbilling', 'gender', 'onlinesecurity', 'internetservice', 'techsupport', 'contract', and 'churn'. The table contains several rows of data. At the bottom right of the dashboard, there are page navigation buttons (1, 2, 3, 4, 5).

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**, at the beginning of the table, by clicking and dragging from the option **Dimensions** available.

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 "Dimensions" panel with a list of fields. The "ChurnScore" field is highlighted with a red box. Other dimensions listed include totalcharges, monthlycharges, tenure, multiplelines, paperlessbilling, gender, onlinesecurity, internetservice, techsupport, and contract.

Below the charts, there is a table with several rows of data. The columns are: totalcharges, monthlycharges, tenure, multiplelines, paperlessbilling, gender, onlinesecurity, internetservice, techsupport, and contract. The data includes various service details like "DSL", "Fiber optic", and "Month-to-month".

The interface also features a "Build" tab in the top right corner, which is currently selected. The "Dimensions" and "Measures" panels are visible on the right side of the screen.

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". Each chart has "Record Count" on the y-axis (0 to 2,000) and "No internet service" (No, Yes) on the x-axis. The first chart is for "partner", the second for "streamingmovies", and the third for "streamingtv". A legend indicates "No" is light blue and "Yes" is dark blue. Below the charts is a table with the following data:

	totalcharges	monthlycharges	tenure	multipliness	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 interface includes a sidebar with "Dimensions", "Measures", and "Tools" sections, and a "Refresh Visual" button at the bottom right of the table area.