

Data Lifecycle en CDP Public Cloud

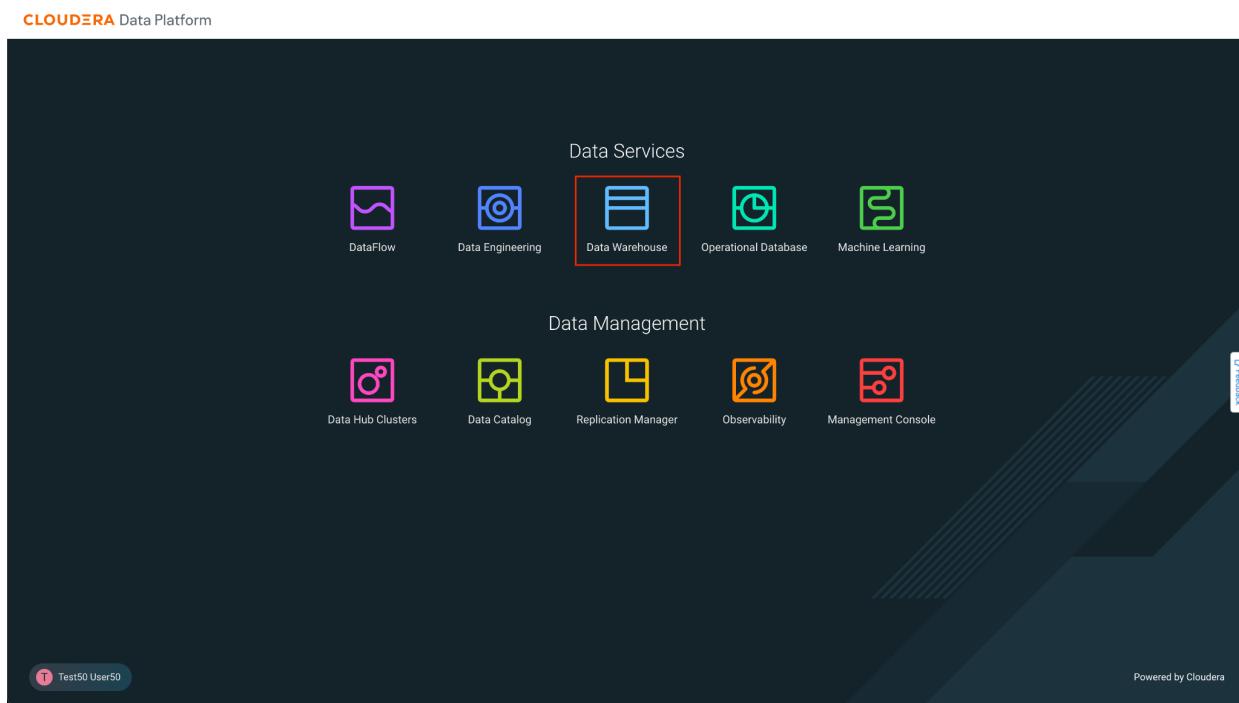
Laboratorio Data Warehouse

Parte 1: Desarrollo dashboard

Objetivos:

- Crear un dataset que apunte a la tabla
- Crear un dashboard con métricas y dimensiones

1. Hacer clic en Data Warehouse desde el Home de CDP PC:



2. Pantalla de bienvenida de Data Warehouse. Hacer clic en Data Visualization en el menú izquierdo.

The screenshot shows the Cloudera Data Warehouse interface. On the left sidebar, there are links for Overview, Database Catalogs, Virtual Warehouses, and Data Visualization. The main area is titled 'Overview' and contains sections for 'Get started with Data Warehouse', 'Create', 'Query and visualize data', and 'Guides and More'. Below these are two tabs: 'Database Catalogs | 1' and 'Virtual Warehouses | 2'. The 'Virtual Warehouses' tab is selected, displaying two entries: 'impala-vw-0' (Running) and 'hive-vw-0' (Stopped). Each entry includes details like total cores, memory, and executors.

3. En Data Visualization, hacer clic en el botón **Data Viz** del cual les fue asignado.

The screenshot shows the 'Data Visualization' page. The left sidebar has the same navigation as the previous screen. The main area displays a table of data visualizations. One row, labeled 'dataviz-0', has its 'Data Viz' button highlighted with a red box. The table columns include NAME, DATA VISUALIZATION ID, Environment ID, VERSION, CPU, MEMORY, UPTIME, and CREATED BY.

NAME	DATA VISUALIZATION ID	Environment ID	VERSION	CPU	MEMORY	UPTIME	CREATED BY
dataviz-0	viz-1685400615-2kkq	env-rgppp	7.1.1-b30	2	8 GB	an hour	acampus

4. Una vez en Data Visualization, ir a la opción Data del menú superior, y posteriormente a la Connector **ImpalaConn** del menú izquierdo.

The screenshot shows the Databricks Data Catalog interface. On the left, there's a sidebar with a 'Connections' section containing 'All Connections' and two entries: 'ImpalaConn' (highlighted with a red box) and 'samples'. The main area is titled 'Datasets (12)' and displays a table of datasets. The columns are: Title/Table, ID, Created, Last Updated, Modified By, and # Dashboards. The datasets listed are: 'Food Stores Inspection in NYC' (ID 12), 'Cereals' (ID 11), 'World Life Expectancy' (ID 9), 'Earthquake Data January 2019' (ID 10), 'US State Populations Over Time' (ID 7), 'US County Population' (ID 8), 'Global Information Security Threats' (ID 6), and 'Restaurant Inspection SF' (ID 5). All datasets were created and last updated on May 29, 2023, by 'vizapps_admin', and each has one dashboard associated with it.

5. Tenemos que crear un nueva fuente de datos, para eso, hacer clic en New Dataset y aparecerá una ventana para ingresar la información de la nueva fuente de datos.

The screenshot shows the Databricks Data Catalog interface with the 'New Dataset' dialog open. The 'Connections' sidebar is visible on the left, showing 'All Connections', 'ImpalaConn' (highlighted with a red box), and 'samples'. The main area has tabs for 'Datasets' and 'Connection Explorer'. The 'Datasets' tab is active, showing a table with one row: 'No data'. At the top of the main area, there are buttons for '% NEW CONNECTION', 'NEW DATASET' (highlighted with a red box), 'ADD DATA', and '...'. The 'NEW DATASET' button is the primary focus of this step.

6. Ingresar la información de la nueva fuente de datos:

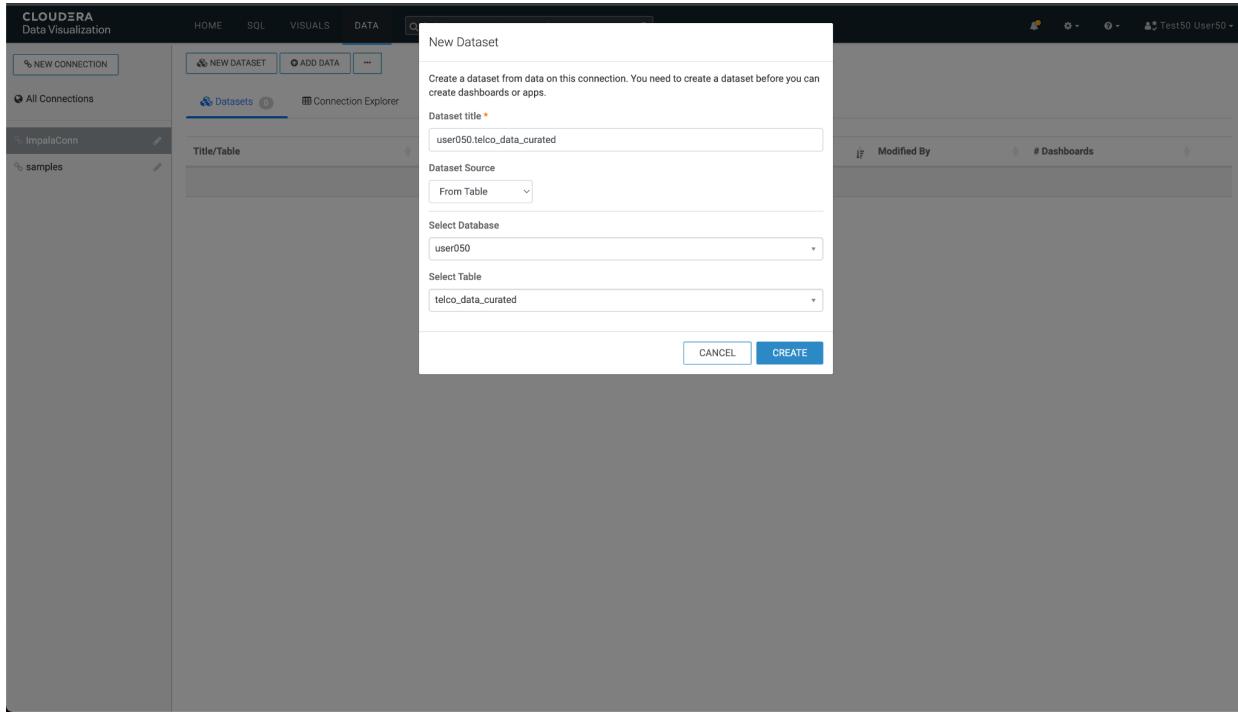
Dataset title: <usuario_asignado>.telco_data_curated

Dataset Source: From table

Select Database: <usuario_asignado>

Select Table: telco_data_curated

Hacer clic en Create para crear el nuevo Dataset.



7. El nuevo Dataset deberá aparecer en el listado. Hacer clic sobre el dataset que usted recién creó.

The screenshot shows the Databricks Dataset Explorer interface. On the left, there's a sidebar with connection management options like 'NEW CONNECTION', 'All Connections', and a selected 'ImpalaConn' entry which includes a 'samples' folder. The main area displays a table titled 'Datasets'. The table has columns for 'Title/Table', 'ID', 'Created', 'Last Updated', 'Modified By', and '# Dashboards'. A single dataset row is listed: 'user050.telco_data_curated' with ID 16, created on May 29, 2023, last updated 'a few seconds ago' by user050, and 0 dashboards.

8. Acá verás el detalle del dataset.

The screenshot shows the 'Dataset Detail' page for the dataset 'user050.telco_data_curated'. The left sidebar lists various dataset configurations: Fields, Data Model, Time Modeling, Segments, Filter Associations, and Permissions. The main panel displays detailed information about the dataset, including its connection type (Impala), data connection (ImpalaConn), and various configuration settings like 'Join Elimination' and 'Result Cache'. At the bottom, it shows the dataset's ID (16) and creation history: created on May 29, 2023 at 06:15 PM by user050, and last updated on the same day at 06:15 PM by user050.

9. Hacer clic en **Fields** (menú izquierdo) para ver los campos automáticamente capturados durante el proceso de creación del dataset.

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. The left sidebar has sections for Dataset Detail, Related Dashboards, Fields, Data Model (selected), Time Modeling, Segments, Filter Associations, and Permissions. The main content area is titled 'Dataset: user050.telco_data_curated'. It shows two sections: 'Fields' (with 'EDIT FIELDS' and 'Hide Comments' buttons) and 'Dimensions' and 'Measures'. The 'Dimensions' section lists fields like 'multipleslines', 'paperlessbilling', 'gender', etc., while the 'Measures' section lists 'totalcharges', 'monthlycharges', and 'tenure'.

10. También se puede pre-visualizar los datos desde esta pantalla. Hacer clic en **Data Model** (menú izquierdo) y después en el botón **Show Data** que aparece en el centro.

This screenshot shows the Cloudera Data Visualization interface again, but from a different perspective. The top navigation bar and sidebar are similar. The main content area is titled 'Dataset: user050.telco_data_curated' and shows the 'Data Model' section. A red box highlights the 'SHOW DATA' button in the center. There is also a checked checkbox for 'Apply Display Format'.

11. En este momento se ejecuta una consulta al Virtual Warehouse para recuperar los datos del data set. Observe las columnas y valores. Hacer clic en New Dashboard para crear un nuevo panel de comando.

The screenshot shows the Cloudera Data Visualization interface. On the left, there's a sidebar with various navigation options: Dataset Detail, Related Dashboards, Fields, Data Model (which is selected and highlighted in grey), Time Modeling, Segments, Filter Associations, and Permissions. The main area displays a preview of the 'telco_data_curated' dataset with several rows of data. At the top right of the main area, there is a red-bordered button labeled 'NEW DASHBOARD'.

multipelines	paperlessbilling	gender	onlinesecurity	internetservice	techsupport	contract	churn	seniorcitizen	deviceprotection	streamingtv	streamingmovies	totalcharges	partner	monthlycharges	customerid	dc
No phone service	Yes	Female	No	DSL	No	Month-to-month	No	0	No	No	No	29.85000381469727	Yes	32.602622985839844	7590-VHVEG	Ni
No	No	Male	Yes	DSL	No	One year	No	0	Yes	No	No	1889.5	No	79.32872009277344	5575-GNVDE	Ni
No	Yes	Male	Yes	DSL	No	Month-to-month	Yes	0	No	No	No	108.1500015258789	No	53.849998474121094	3668-QPYBK	Ni
No phone service	No	Male	Yes	DSL	Yes	One year	No	0	Yes	No	No	1840.75	No	39.008785247802734	7795-CFCOW	Ni
No	Yes	Female	No	Fiber optic	No	Month-to-month	Yes	0	No	No	No	151.64999389648438	No	70.69999694824219	9237-HQITU	Ni
Yes	Yes	Female	No	Fiber optic	No	Month-to-month	Yes	0	Yes	Yes	Yes	820.5	No	99.6500015258789	9305-CDSKC	Ni
Yes	Yes	Male	No	Fiber optic	No	Month-to-month	No	0	No	Yes	No	1949.4000244140625	No	154.11448669433594	1452-KIOVK	Ye
No phone service	No	Female	Yes	DSL	No	Month-to-month	No	0	No	No	No	301.8999938964844	No	46.7568778916992	6713-OKOMC	Ni
Yes	Yes	Female	No	Fiber optic	Yes	Month-to-month	Yes	0	Yes	Yes	Yes	3046.050048828125	Yes	104.80000305175781	7892-POOKP	Ni

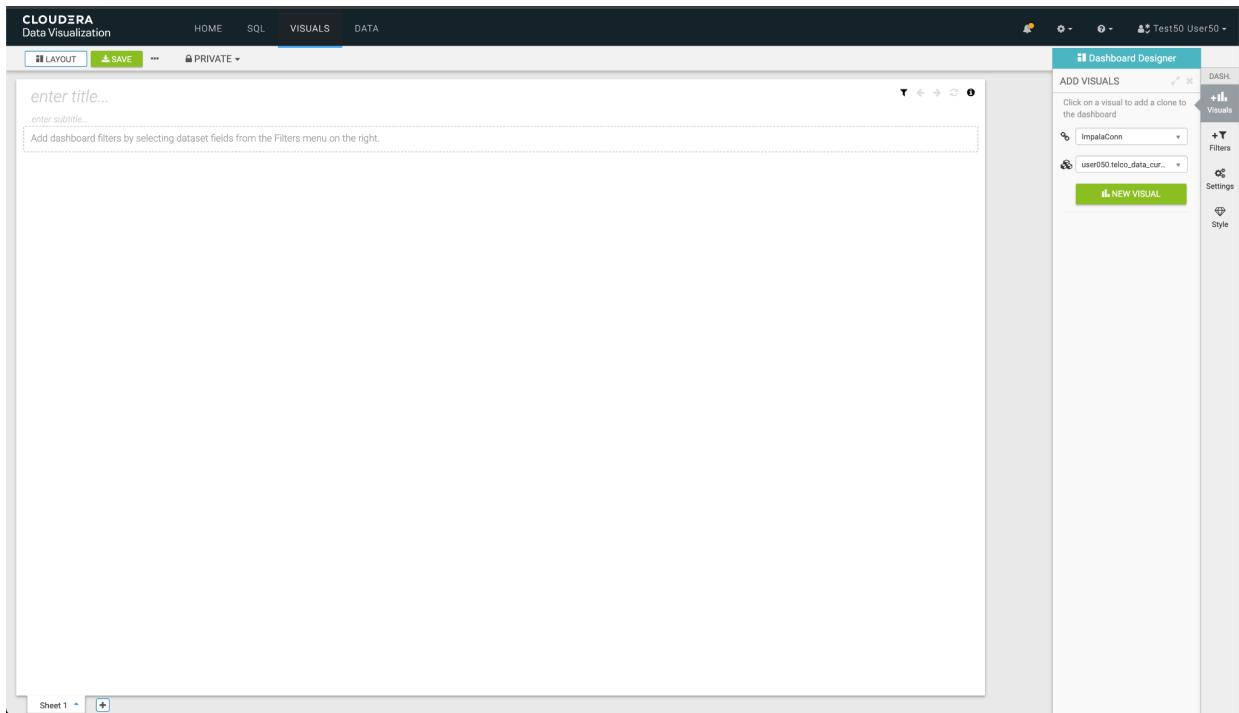
12. Al abrir el canvas de diseño de un nuevo panel, eliminar el elemento que se agrega por defecto, haciendo clic en el botón de tres puntos (...) que hay en la parte superior derecha del elemento, y después clic en la opción **Delete Visual**

The screenshot shows the Cloudera Data Visualization interface. A context menu is open over a table visual titled "Churn Analysis". The menu includes options like "View Data and Queries", "Download as...", "Save as Table or Dataset", "Hide Empty Title & Subtitle", "Clone", "Make Linked", and "Delete Visual". The table visual displays data from the "telco_data_catered" dataset with columns: multiplelines, paperlessbilling, gender, and onlinesecurity. The right side of the screen features the Dashboard Designer sidebar with sections for Dimensions, Measures, and Filters.

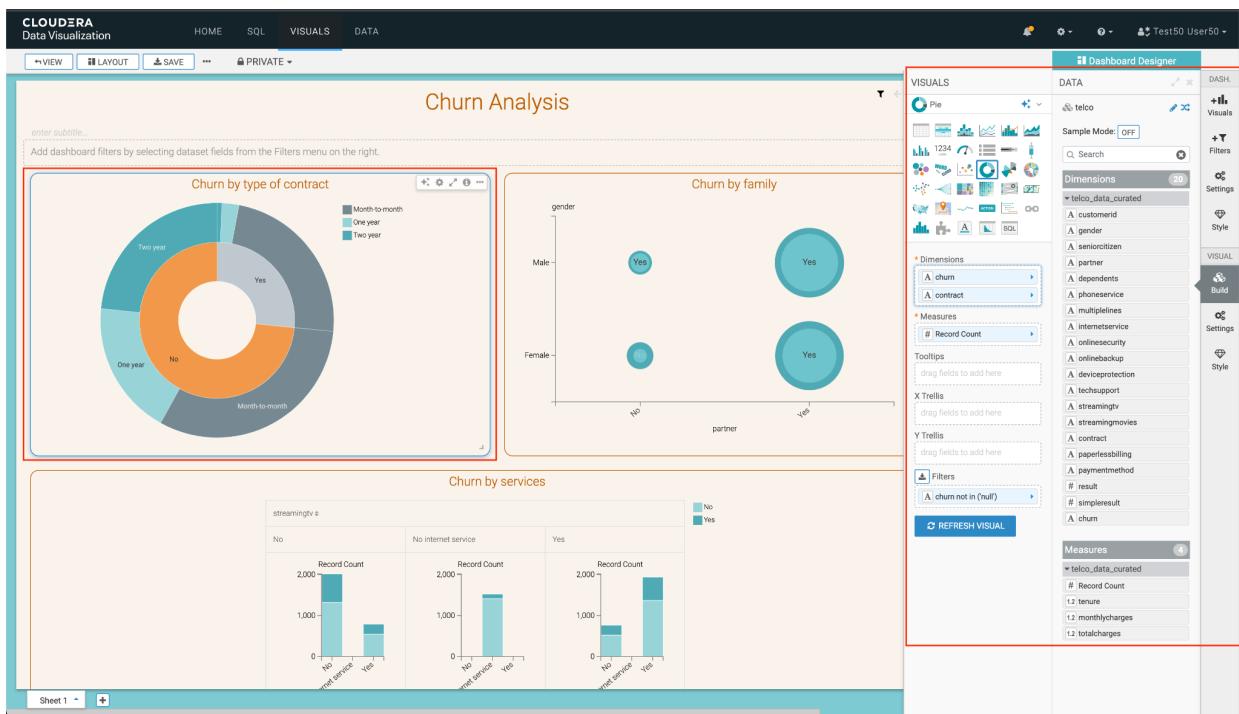
En la parte superior del canvas, en el campo enter title, ingresar el nombre *Churn Analysis* para el identificar el dashboard.

The screenshot shows the Cloudera Data Visualization interface with the dashboard titled "Churn Analysis". The dashboard canvas is currently empty, with a placeholder message: "Add visuals by selecting them from the Visuals menu on the right." The right sidebar shows the "ADD VISUALS" section with a "New Visual" button highlighted in green, and a list of recent visual clones.

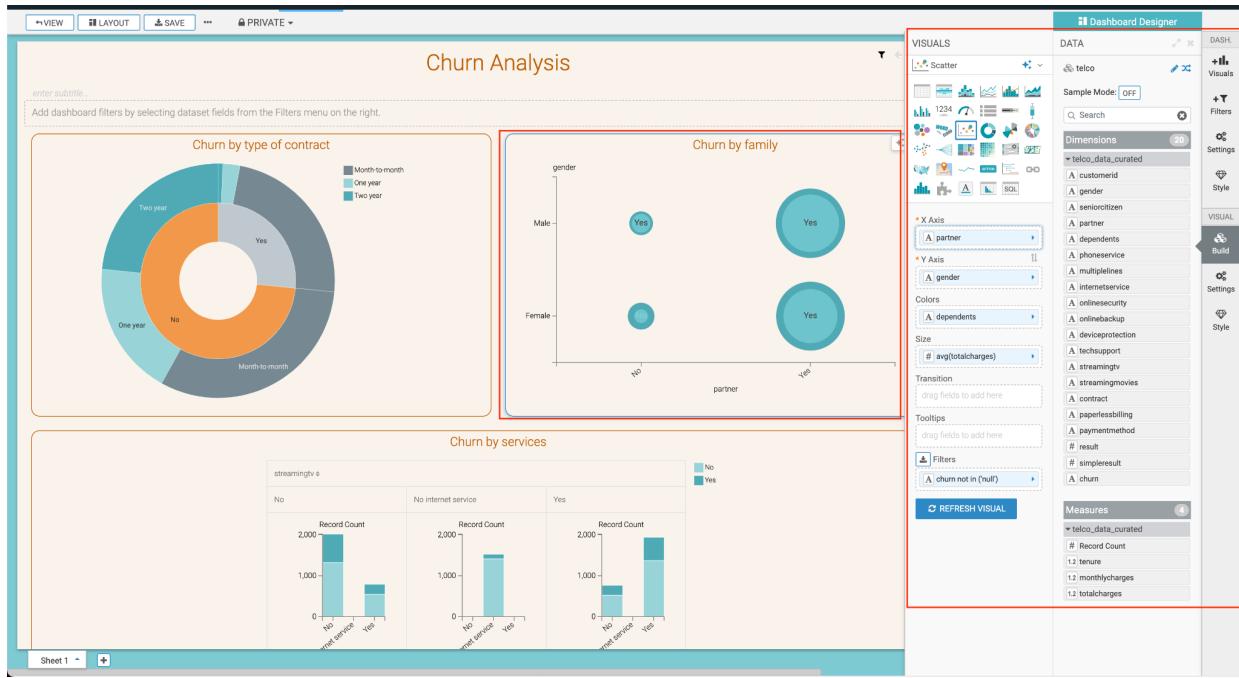
13. Para agregar un nuevo elemento visual, hacer clic en el botón **Visuals** del menú derecho, seleccionar el dataset que les corresponda, y hacer clic en el botón **New Visual**.



14. Agregar el primer elemento visual, que es un gráfico de torta con las dimensiones **churn** y **contract**, con la métrica de **Record count**. Una vez terminado, hacer clic en el botón **Refresh Visual**.



15. Agregar el segundo elemento visual, que es un gráfico tipo Scatter con la dimensión **partner** como X Axis, **gender** como Y Axis, **dependents** como Colors y **avg(totalcharges)** como Size. Una vez terminado, hacer clic en el botón **Refresh Visual**.



15. Agregar el tercer elemento visual, que es un gráfico de barras con la dimensiones **streamingtv** y **streamingmovies** como X Axis, **Record Count** como Y Axis y **churn** como Colors. Una vez terminado, hacer clic en el botón **Refresh Visual**.

The screenshot shows a dashboard titled "Churn by services" containing three bar charts. The first chart, "streamingtv", compares "Record Count" for "No" and "Yes" categories across "No internet service" and "Yes". The second chart, "streamingmovies", does the same. The third chart, also for "streamingmovies", compares "Record Count" for "No" and "Yes" across "No internet service" and "Yes".

Below these charts is a table titled "Scoring - Churn Probability" with columns: result, customerid, tenure, monthlycharges, totalcharges, gender, dependents, onlinesecurity, multiplelines, internetservice, and seniorcitizen. The table lists five rows of data.

The right side of the screen displays the "Dashboard Designer" interface, which includes sections for Dimensions, Visuals, and Measures, along with a sidebar for Settings and Style.

customerid	tenure	monthlycharges	totalcharges	gender	dependents	onlinesecurity	multiplelines	internetservice	seniorcitizen
7590-VHVEG	1	32.602622985839844	29.850000381469727	Female	No	No	No phone service	DSL	0
5575-GNVDE	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

16. Agregar el cuarto y último elemento visual, que es una table con las dimensiones y métricas del dataset. Asegurarse de agregar las 17 dimensiones y 3 métricas a la tabla. Una vez terminado, hacer clic en el botón Refresh Visual.

The screenshot shows the same dashboard as the previous one, but the table "Scoring - Churn Probability" now includes the 17 dimensions and 3 metrics specified in the task. The table has 10 rows of data.

customerid	tenure	monthlycharges	totalcharges	gender	dependents	onlinesecurity	multiplelines	internetservice	seniorcitizen
7590-VHVEG	1	32.602622985839844	29.850000381469727	Female	No	No	No phone service	DSL	0
5575-GNVDE	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	46.75687789916992	301.8999938964844	Female	No	Yes	No phone service	DSL	0
									100

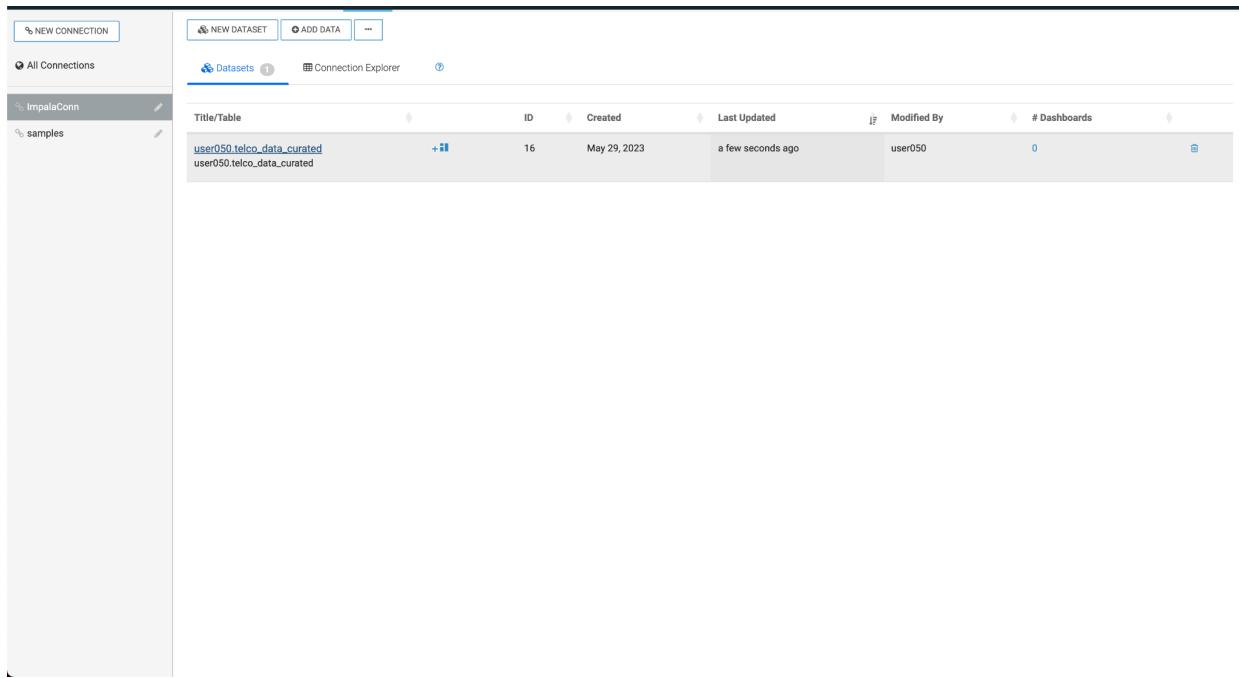
Guarda el dashboard haciendo clic en el botón **Save** del menú superior.

Parte 2: Agregar nuevo campo

Objetivos:

- Agregar un nuevo campo que hace llamadas al modelo de ML
- Agregar el nuevo campo al dashboard

1. Editar el Dataset creado anteriormente, en Data -> <usuario_asignado>.telco_data_curated.



The screenshot shows the Databricks Data interface. On the left, there's a sidebar with connection management options like 'NEW CONNECTION', 'All Connections', and specific connections for 'ImpalaConn' and 'samples'. The main area is titled 'Datasets' and includes a 'Connection Explorer' tab. A table lists datasets, with one entry highlighted: 'user050.telco_data_curated'. The table columns are: Title/Table, ID, Created, Last Updated, Modified By, and # Dashboards. The dataset details are: ID 16, Created May 29, 2023, Last Updated 'a few seconds ago', Modified By user050, and 0 dashboards.

2. Una vez en el Dataset, ir a **Fields** en el menú izquierdo y posteriormente hacer clic en **Edit Field** para editar los campos de su dataset.

CLOUDERA Data Visualization

Dataset Detail

Related Dashboards

Fields

Dataset: user050.telco_data_curated

Fields EDIT FIELDS Hide Comments NEW DASHBOARD

Dimensions

- telco_data_curated
 - 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
 - Iz totalcharges
 - Iz monthlycharges
 - Iz tenure

3. En la lista de **Dimensions**, hacer clic en la flecha hacia abajo del último campo de la lista, y seleccionar la opción **Clone**.

CLOUDERA Data Visualization

Dataset Detail

Related Dashboards

Fields UNDO REFRESH TITLE CASE SAVE Show Comments NEW DASHBOARD

To add a new calculated field, use the down arrow to the right of a field to clone it, and then edit the expression of the cloned field.

Dimensions

- telco_data_curated
 - Dim | A - multiplelines
 - 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

Measures

- telco_data_curated
 - Mes | Iz totalcharges
 - Mes | Iz monthlycharges
 - Mes | Iz tenure

4. Una vez clonado el campo, hacer clic en el lápiz al lado del campo para editarlo.

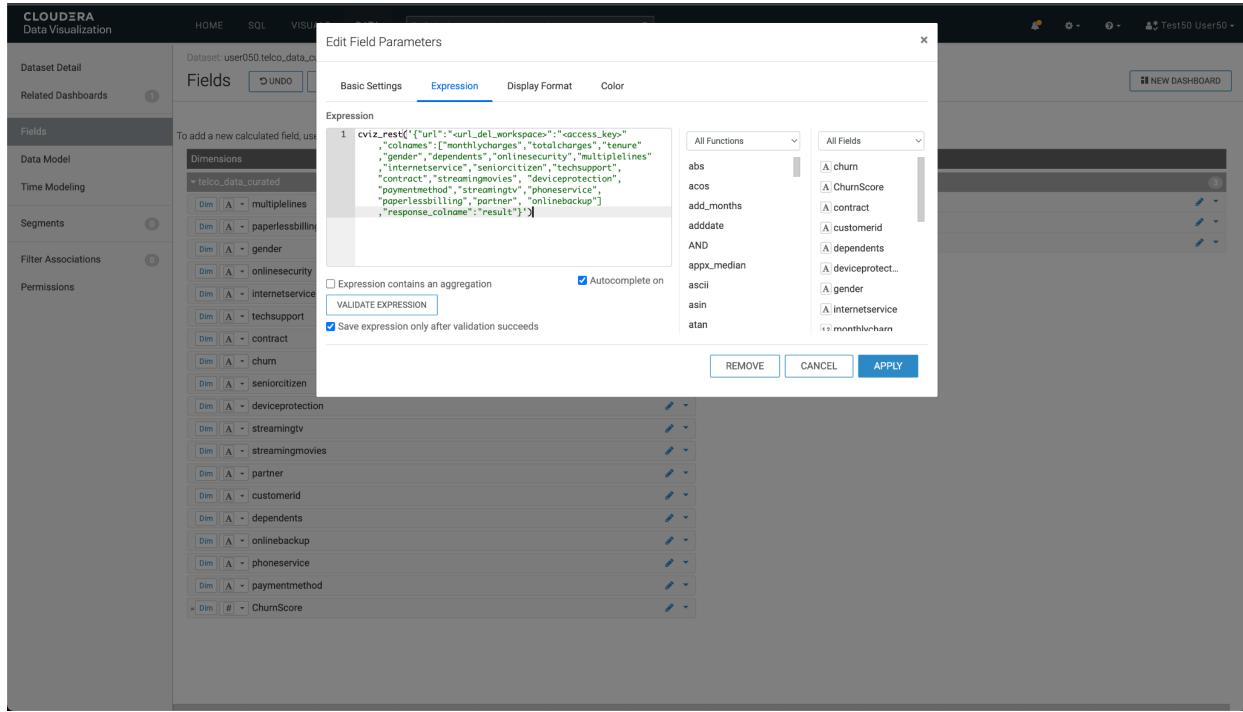
The screenshot shows the 'Fields' section of the Cloudera Data Visualization interface. The 'Dimensions' section contains 19 items, including 'multiplelines', 'paperlessbilling', 'gender', 'onlinesecurity', 'internetservice', 'techsupport', 'contract', 'churn', 'seniorcitizen', 'deviceprotection', 'streamingtv', 'streamingmovies', 'partner', 'customerid', 'dependents', 'onlinebackup', 'phoneservice', and 'paymentmethod'. The 'Measures' section contains 3 items: 'totalcharges', 'monthlycharges', and 'tenure'. A sidebar on the left lists 'Dataset Detail', 'Related Dashboards', 'Segments', 'Filter Associations', and 'Permissions'.

5. En la ventana emergente que aparece, introducir el nombre del nuevo campo en **Display Name**. Sugerimos que introduzcas *ChurnScore*.

The screenshot shows the 'Edit Field Parameters' dialog box for the 'paymentmethod' field. The 'Basic Settings' tab is selected, showing the 'Display Name' field filled with 'ChurnScore'. Other tabs include 'Expression', 'Display Format', and 'Color'. The 'Field Comment' and 'Default Aggregation' fields are also visible. At the bottom, there are 'REMOVE', 'CANCEL', and 'APPLY' buttons.

6. Ir a la pestaña Expressions e introducir el siguiente valor en el campo Expression. Esto permitirá hacer llamada a la API REST del Modelo que anteriormente ha desplegado.

```
cviz_rest('{"url":"<url_del_workspace>":"<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. Estando en CML en otra pestaña del navegador web, ir a la sección de **Models** de su proyecto, y hacer clic sobre el Model que empieza con el nombre *ModelViz*, seguido de su nombre de usuario asignado.

The screenshot shows the Cloudera Machine Learning interface for the user050 project. The left sidebar contains navigation links: All Projects, Overview, Sessions, Data, Experiments, Models, Jobs, Applications, Files, Collaborators, and Project Settings. The main area displays three sections: Models, Jobs, and Files.

Models

Model	Source	Status	Replicas	CPU	Memory	Last Deployed	Actions
ModelViz_user050	13_mod...	Deployed	1 / 1	1	2.00 GiB	May 29, 2023, 03:54 PM	<button>Stop</button>
ModelOpsChurn_user050	11_best...	Deployed	1 / 1	1	2.00 GiB	May 29, 2023, 03:53 PM	<button>Stop</button>

Jobs

Name	Runs / Failures	Duration	Status	Latest Run	Actions
deploy_best_model	0 / 0	00:00	Not Yet Run	-	<button>Run</button>
retrain	0 / 0	00:00	Not Yet Run	-	<button>Run</button>
avisoPerformance	0 / 0	00:00	Not Yet Run	-	<button>Run</button>
Check Model	0 / 0	00:00	Not Yet Run	-	<button>Run</button>

Files

Name	Size	Last Modified	Actions
__pycache__	-	15 hours ago	
flask	-	15 hours ago	
images	-	15 hours ago	
models	-	15 hours ago	
raw	-	15 hours ago	
O_bootstrap.py	1.95 kB	15 hours ago	<button>Edit</button>
O_create_jobs.py	5.60 kB	15 hours ago	<button>Edit</button>

Workspace: ssa-cml-workspace
Cloud Provider: AWS (AWS)

8. En la pestaña de Overview, copiar la URL que permite interactuar y hacer llamada a la API del workspace.

Reemplazar el valor copiado en el atributo `<url_del_workspace>` del campo Expression.

9. Regresando a CML, copiar la accessKey del modelo.

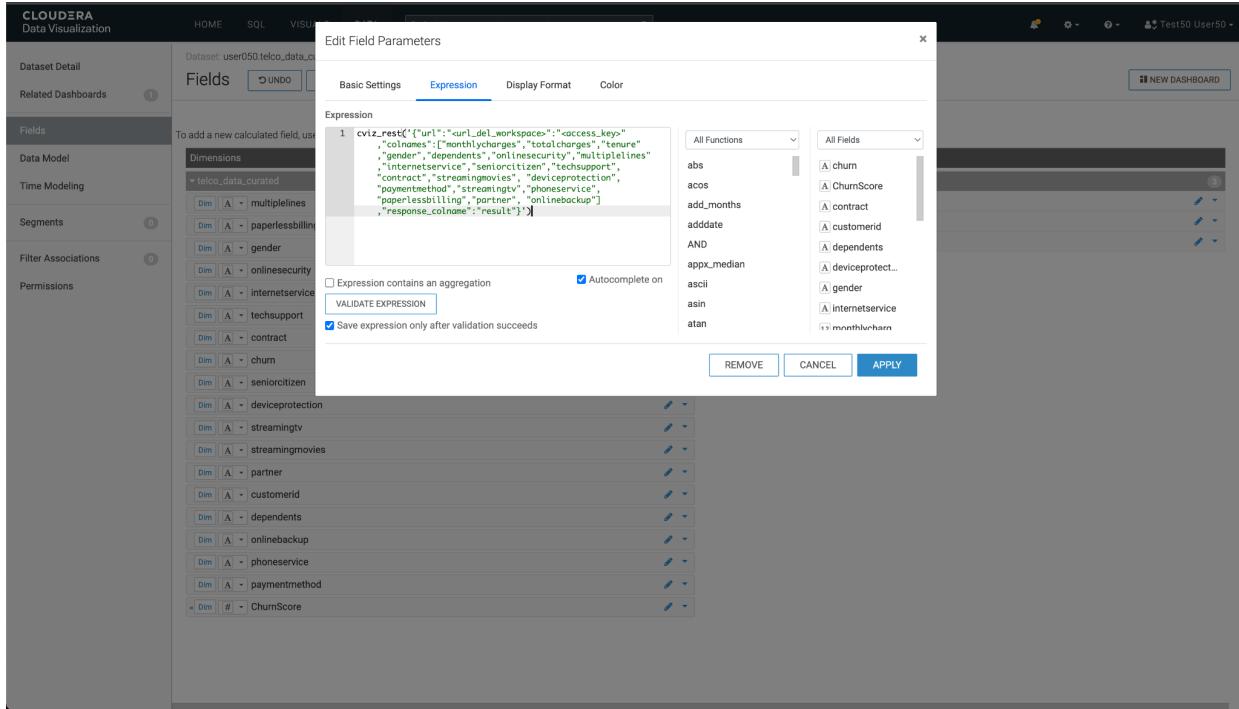
Model Details

- Source: cm:cdp:ml:us-west-1:508fd88f-8076-49ba-acfb-6f8765cd3d5e8
- Code: 8
- Model CRN: crn:cdp:ml:us-west-1:508fd88f-8076-49ba-acfb-6f8765cd3d5e8:workspace:814194cb-1c7e-48cd-9989-b499a79ed5f6/dae534c1-b214-45eb-acd0-101e651ff68d
- Deployment Id: 10
- Deployment CRN: cm:cdp:ml:us-west-1:508fd88f-8076-49ba-acfb-6f8765cd3d5e8:workspace:814194cb-1c7e-48cd-9989-b499a79ed5f6/cf985a5d:9870-4533-9f9a-d42ad0db56ed
- Build Id: 10
- Build CRN: cm:cdp:ml:us-west-1:508fd88f-8076-49ba-acfb-6f8765cd3d5e8:workspace:814194cb-1c7e-48cd-9989-b499a79ed5f6/0e00c2d9-80cb-4ee8-8304-79987673de32
- Deployed By: user050
- Comment: Initial revision.
- Runtime Image: Python 3.7 (Standard)
- File: 13_model_viz.py
- Function: predict

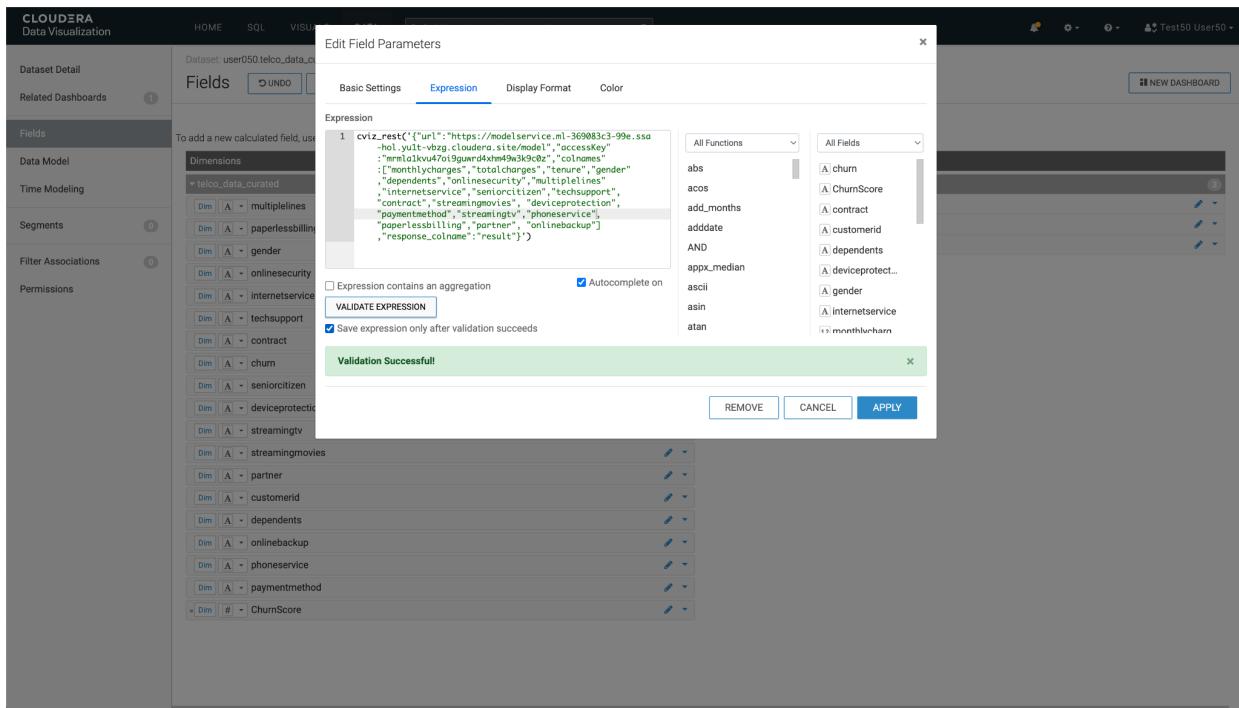
Model Resources

- Replicas: 1
- Total CPU: 1 vCPUs
- Total Memory: 2.00 GiB

Reemplazar el valor copiado en el atributo <access_key> del campo Expression.



10. Finalizado el proceso de copiar la *url del workspace* y la *accessKey*, hacer clic en el botón Validate Expression, que se encuentra en la parte superior de la ventana. Si aparece el mensaje en verde *Validation Successful*, hacer clic en **Apply** para guardar la configuración realizada.



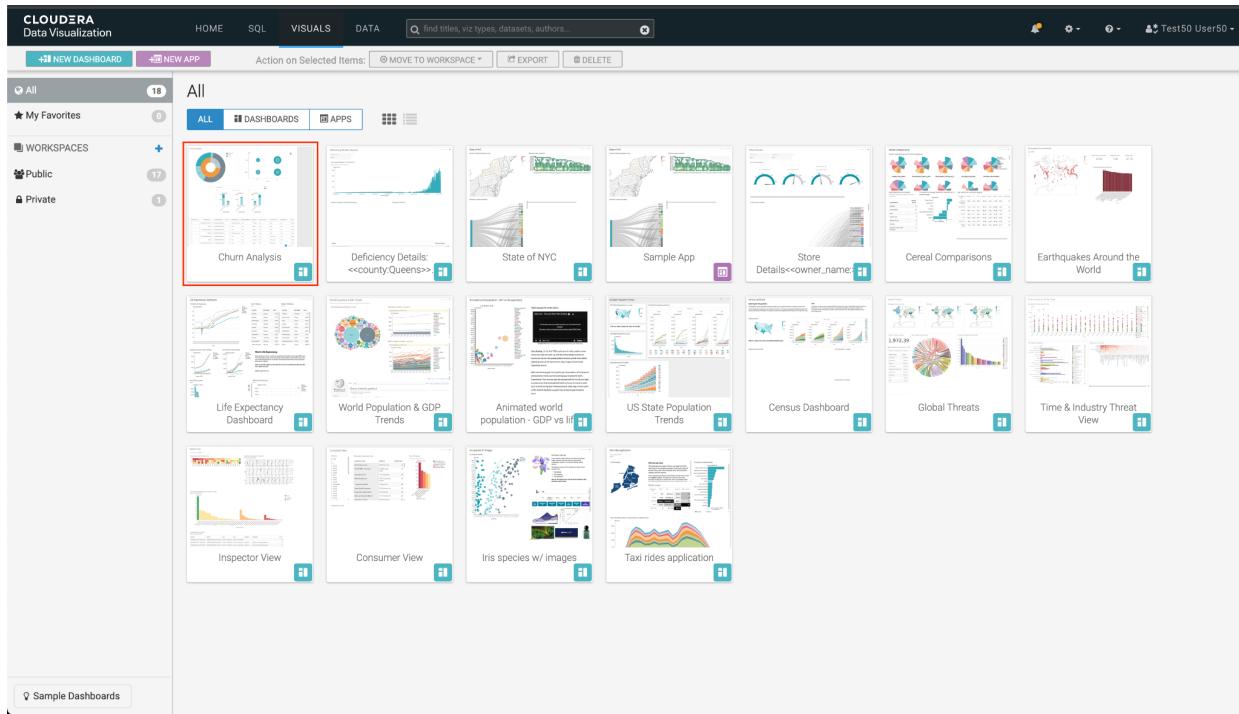
11. El nuevo campo deberá aparecer en el listado de campos. Cambiar el tipo de dato, seleccionando el tipo *Integer*, que se representa por el símbolo #

The screenshot shows the Cloudera Data Visualization interface. In the top navigation bar, 'DATA' is selected. Below it, the dataset 'user050.telco_data_curated' is chosen. On the left sidebar, 'Fields' is selected. In the main area, under 'Dimensions', there is a list of fields like 'multipleinles', 'paperlessbilling', 'gender', etc. A modal dialog is open over this list, showing a list of data types: Boolean, Integer, Real, String, Timestamp, Remove CAST, and #. The '# Integer' option is highlighted. To the right, under 'Measures', there are three fields: 'totalcharges', 'monthlycharges', and 'tenure'.

12. Finalizar el proceso haciendo clic en el botón verde con leyenda **SAVE** en el menú superior.

The screenshot shows the same Cloudera Data Visualization interface after the process from the previous step. The 'Dimensions' list now includes the 'ChurnScore' field, which is listed with the data type '# Integer'. The other fields in the list remain the same: 'multipleinles', 'paperlessbilling', 'gender', etc.

13. Regresar al dashboard, seleccionando la opción **VISUALS** del menú superior, y haciendo clic sobre el nombre del dashboard que se creó anteriormente.



The screenshot shows the Cloudera Data Visualization interface. At the top, there is a navigation bar with links for HOME, SQL, VISUALS, and DATA, along with a search bar and user information. Below the navigation bar is a toolbar with buttons for NEW DASHBOARD, NEW APP, and various actions like MOVE TO WORKSPACE, EXPORT, and DELETE. On the left side, there is a sidebar with sections for All, My Favorites, Workspaces (Public and Private), and Sample Dashboards. The main area displays a grid of sample dashboards, each with a preview image and a title. One dashboard, titled "Churn Analysis", is highlighted with a red box and a white edit icon in the bottom right corner of its preview. 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 life expectancy", "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. Una vez en el dashboard, hacer clic en el botón **Edit** que se encuentra en la parte superior izquierda.

CLOUDERA Data Visualization

HOME SQL VISUALS DATA

EDIT PRIVATE

partner

streamingtv #

	No	No internet service	Yes
Record Count	2,000	2,000	2,000
streamingmovies	No internet service Yes	No internet service Yes	No internet service Yes

totalcharges ♦ monthlycharges ♦ tenure ♦ multiplelines ♦ paperlessbilling ♦ gender ♦ onlinesecurity ♦ internetservice ♦ techsupport ♦ contract ♦ ch

totalcharges	monthlycharges	tenure	multiplelines	paperlessbilling	gender	onlinesecurity	internetservice	techsupport	contract	chu
29.850000381469727	32.602622985839844	1	No phone service	Yes	Female	No	DSL	No	Month-to-month	No
1,889.5	79.32872009277344	34	No	No	Male	Yes	DSL	No	One year	No
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	Nr
151.64999389648438	70.69999649824219	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

1 2 3 4 5 >

15. Editar la tabla inferior, haciendo clic sobre la misma y posteriormente sobre la opción **Build** del menú vertical derecho. Agregar el nuevo campo, **ChurnScore**, al inicio de la tabla, haciendo clic y arrastrando desde la opción de **Dimensions** disponible.

CLOUDERA Data Visualization

HOME SQL VISUALS DATA

VIEW LAYOUT SAVE PRIVATE

partner

streamingtv #

	No	No internet service	Yes
Record Count	2,000	2,000	2,000
streamingmovies	No internet service Yes	No internet service Yes	No internet service Yes

totalcharges ♦ monthlycharges ♦ tenure ♦ multiplelines ♦ paperlessbilling ♦ gender ♦ onlinesecurity ♦ internetservice ♦ techsupport ♦ contract ♦ ch

totalcharges	monthlycharges	tenure	multiplelines	paperlessbilling	gender	onlinesecurity	internetservice	techsupport	contract	chu
29.850000381469727	32.602622985839844	1	No phone service	Yes	Female	No	DSL	No	Month-to-month	No
1,889.5	79.32872009277344	34	No	No	Male	Yes	DSL	No	One year	No
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	Nr
151.64999389648438	70.69999649824219	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

Sheet 1 +

Dimensions

- # ChurnScore
- 1.2 totalcharges
- 1.2 monthlycharges
- 1.2 tenure
- 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

- drag fields to add here

Tooltips

- drag fields to add here

Filters

- drag fields to add here

Limit: 100

REFRESH VISUAL

16. Hacer clic en el botón Refresh Visual para actualizar los datos. Deberá aparecer la nueva columna *ChurnScore* luego al inicio de la tabla, con un valor de tipo numérico. Finalizar el proceso haciendo clic en botón **SAVE** del menú superior izquierdo.

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 charts have "Record Count" on the Y-axis and "streamingmovies" on the X-axis. The legend indicates "No" in light blue and "Yes" in dark blue. Below the charts is a table with the following data:

	totalcharges	monthlycharges	tenure	multiplelines	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

On the right side of the interface, the "Dashboard Designer" panel is open, showing the "Dimensions" and "Measures" sections. The "Dimensions" section lists fields like ChurnScore, totalcharges, monthlycharges, tenure, multiplelines, paperlessbilling, gender, onlinesecurity, internetservice, techsupport, contract, churn, seniorcitizen, deviceprotection, streamingtv, streamingmovies, partner, customerid, dependents, onlinebackup, phoneservice, paymentmethod, and ChurnScore. The "Measures" section lists Record Count, totalcharges, monthlycharges, and tenure. A "REFRESH VISUAL" button is located at the bottom of the Designer panel.