

Data Lifecycle en CDP Public Cloud

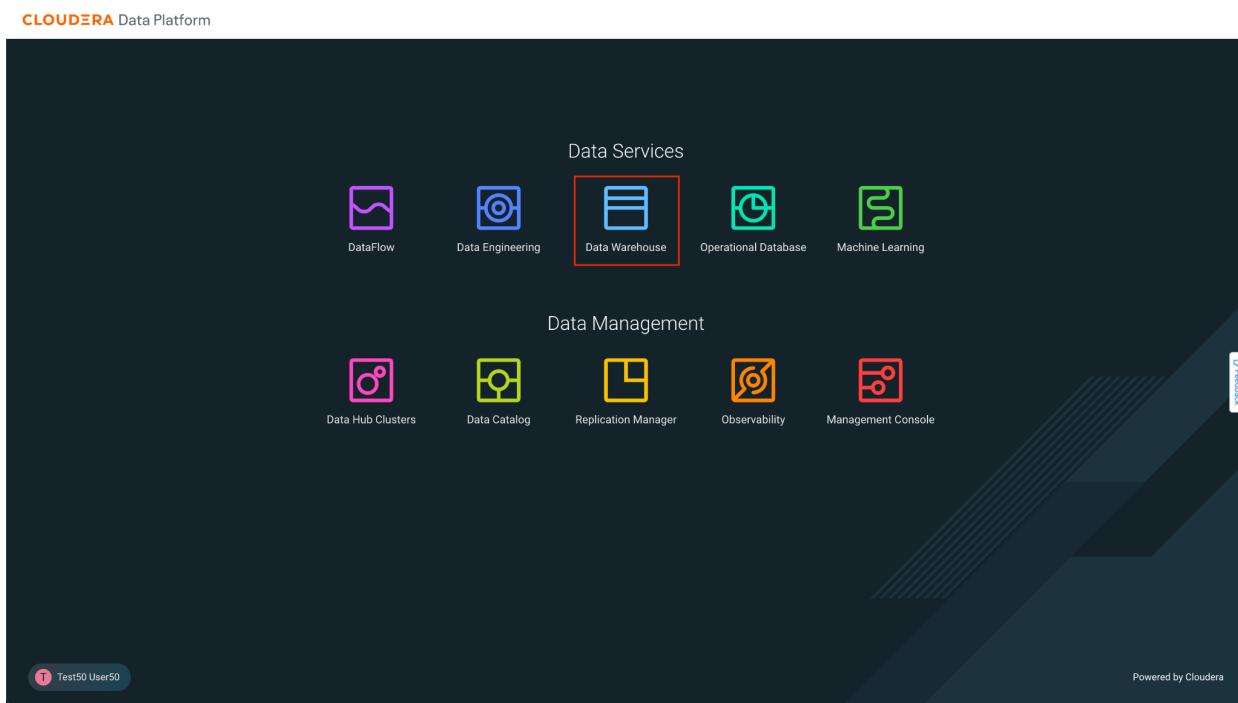
Data Warehouse Lab

Dashboard development

Goals:

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

1. Click on Data Warehouse from CDP PC Home:



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

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 "Database Catalogs" tab shows one entry: "ssa-datalake-de...", which is running and has 9 total cores, 25 GB total memory, and 2 virtual warehouses. The "Virtual Warehouses" tab shows two entries: "impala-vw-0" (running, 2 executors, 63 total cores, 476 GB total memory, type IMPALA) and "hive-vw-0" (stopped, 0 executors, 12 total cores, 56 GB total memory, type HIVE, UNIFIED ANALYTICS, COMPACTOR). A "Management Console" link is also present.

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

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, "dataviz-0", is highlighted with a red border around its "Data Viz" button. The table columns include NAME, DATA VISUALIZATION ID, Environment ID, VERSION, CPU, MEMORY, UPTIME, and CREATED BY. The "dataviz-0" row shows values: viz-1685400615-2kkq, env-rggpp, 7.1.1-b30, 2, 8 GB, an hour, and acampos.

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

Datasets (12)

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

5. We 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.

NEW DATASET

Datasets (1)

Title/Table	ID	Created	Last Updated	Modified By	# Dashboards
No data					

6. Enter the information for the new data source:

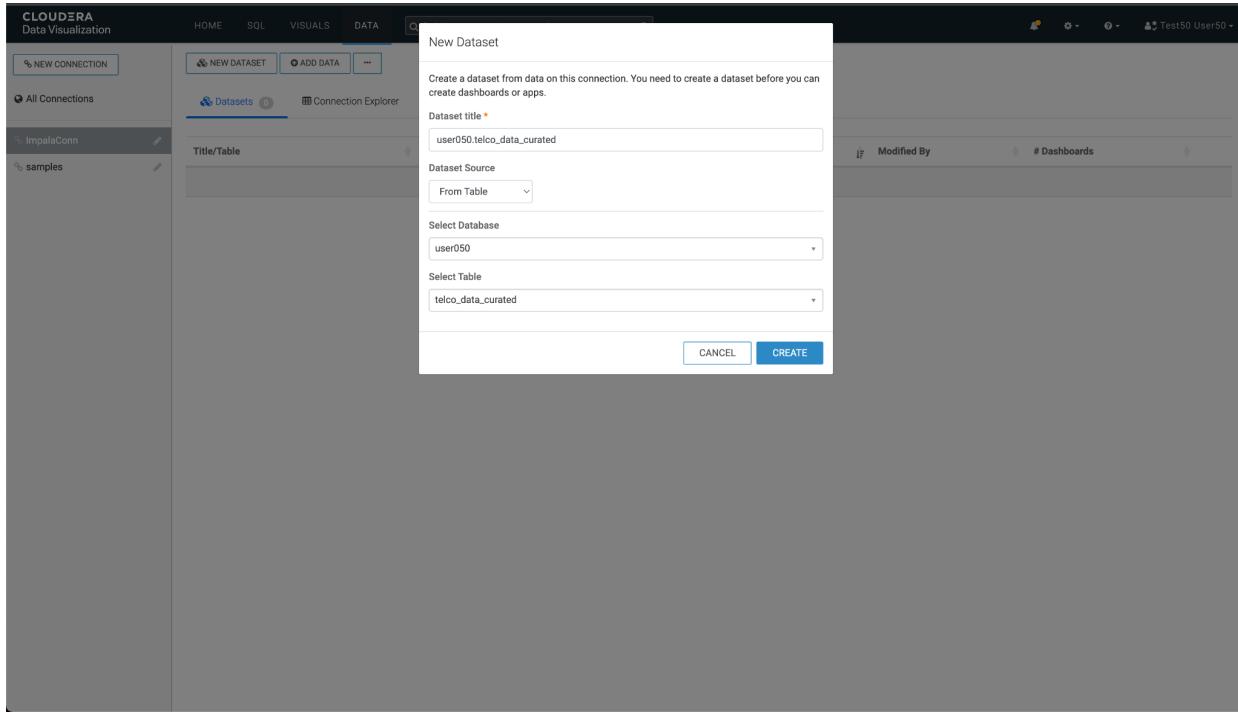
Dataset title: <assigned_user>.telco_curated_data . For example, **cdpuser50**

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.

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					

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 contains navigation links: Dataset Detail, Related Dashboards, Fields, Data Model, Time Modeling, Segments (0), Filter Associations (0), and Permissions. The main content area displays 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 these are creation and update logs: 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 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 'Fields' page for the dataset 'user050.telco_data_curated'. The left sidebar is identical to the previous screenshot. The main content area shows the 'Fields' section with a 'Dimensions' panel containing 19 items: multiplelines, paperlessbilling, gender, onlinesecurity, internetservice, techsupport, contract, churn, seniorcitizen, deviceprotection, streamingtv, streamingmovies, partner, customerid, dependents, onlinebackup, phoneservice, and paymentmethod. To the right is a 'Measures' panel with 3 items: totalcharges, monthlycharges, and tenure. At the top right are 'Edit Fields' and 'New Dashboard' buttons. The top navigation bar includes HOME, SQL, VISUALS, DATA, and a search bar.

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. In the main area, there is a dataset named 'telco_data_curated'. A prominent blue button labeled 'SHOW DATA' is centered in the main content area, with a red box drawn around it. Below this button is a checkbox labeled 'Apply Display Format'. At the top right of the main area, there is a 'NEW DASHBOARD' button.

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.

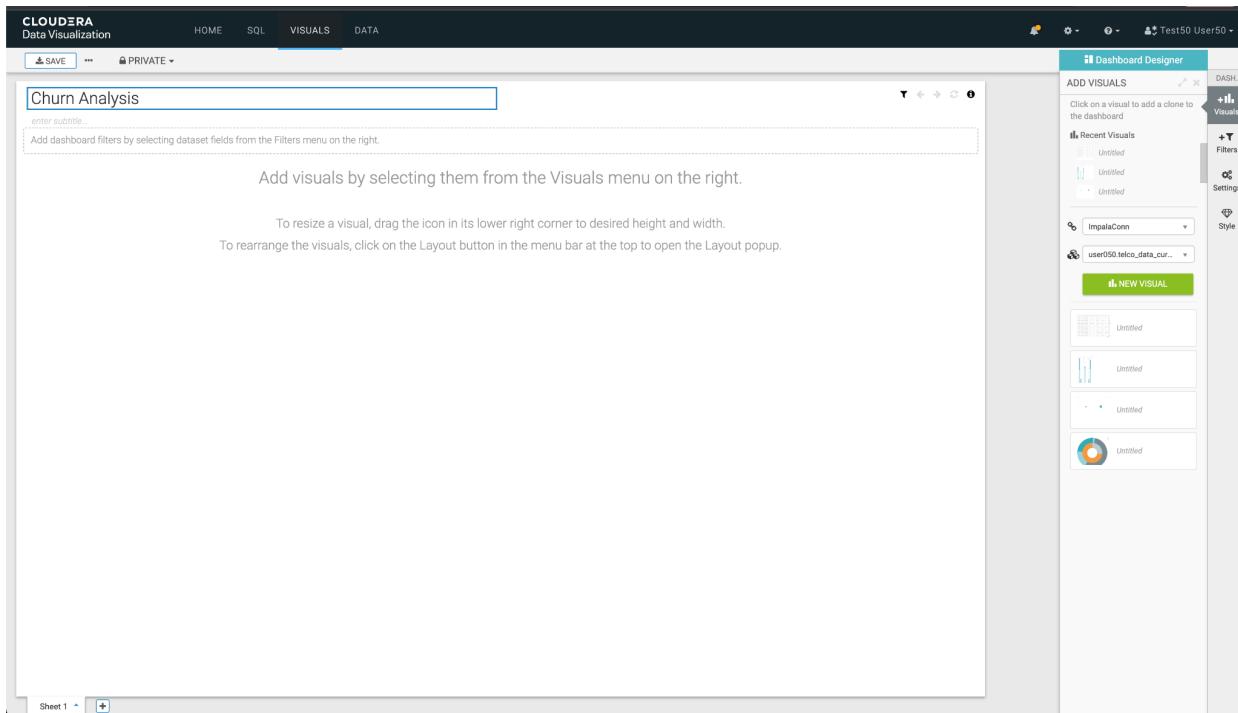
The screenshot shows the same Cloudera Data Visualization interface as the previous one, but now the 'telco_data_curated' dataset is fully visible as a data preview. The 'HIDE DATA' button is highlighted with a red box. The data preview shows various columns such as 'multiplelines', 'paperlessbilling', 'gender', 'onlinesecurity', 'internetservice', 'techsupport', 'contract', 'churn', 'seniorcitizen', 'deviceprotection', 'streamingtv', 'streamingmovies', 'totalcharges', 'partner', 'monthlycharges', 'customerid', and 'de'. The data consists of multiple rows of customer information.

12. When opening the design canvas of a new panel, remove the Table visual 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**

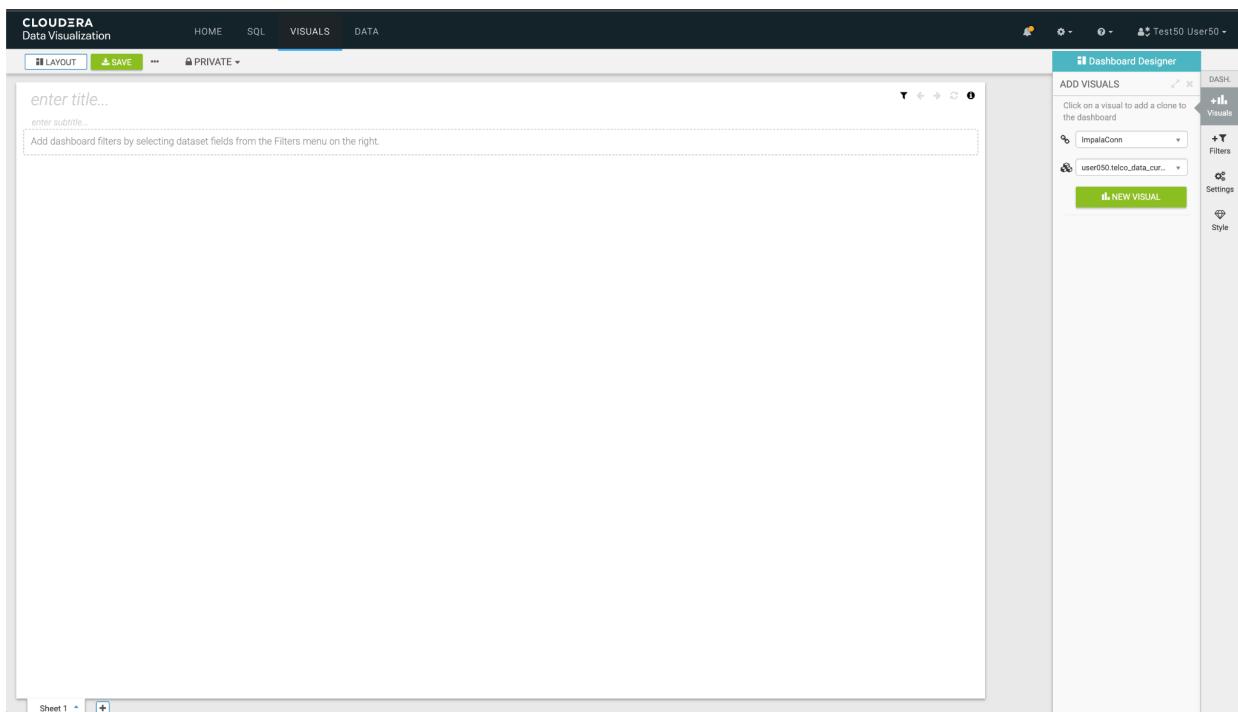
The screenshot shows the Cloudera Data Visualization interface. On the left, there's a dashboard canvas with a table visual containing data about phone service. A context menu is open over the table, with the 'Delete Visual' option highlighted. To the right is the 'Dashboard Designer' sidebar, which includes sections for DATA, DASH, VISUAL, and BUILD. The DATA section shows a dataset named 'user050.telco_data_cur...' with various dimensions and measures listed.

multiplelines	paperlessbilling	gender	onlinesecurity
No phone service	Yes	Female	No
No	No	Male	Yes
No	Yes	Male	Yes
No phone service	No	Male	Yes
No	Yes	Female	No
Yes	Yes	Female	No

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



13. To add a new visual element, click on the button **Visuals** from the right menu, select the appropriate dataset, and click on the button **New Visual**.



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

Cloudera Data Visualization

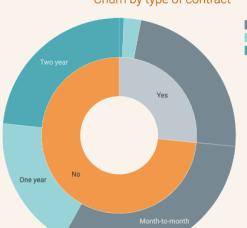
HOME SQL VISUALS DATA

VIEW LAYOUT SAVE PRIVATE

Churn Analysis

enter subtitle... Add dashboard filters by selecting dataset fields from the Filters menu on the right.

Churn by type of contract



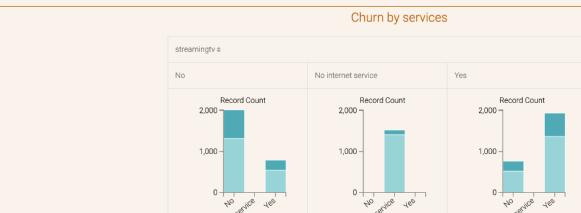
Contract Type	Count
Month-to-month	1500
One year	1000
Two year	500

Churn by family



Gender	partner	Churn (Yes)
Male	High	150
Female	Low	100

Churn by services



Service	No	Yes
streamingv2	2000	1000
no_internet_service	2000	1500
yes_internet_service	2000	1500

VISUALS DATA DASH. Dashboard Designer

Pie telco

Sample Mode OFF

Search

Dimensions 26

- telco_data_captured
- customer
- gender
- seniorcitizen
- partner
- dependents
- phoneservice
- multipleservices
- internetservice
- onlinesecurity
- onlinebackup
- deviceprotection
- techsupport
- streamingv2
- streamingmovies
- contract
- paperlessbilling
- paymentmethod

Measures 4

- # Record Count
- # Record Count
- # tenure
- # monthlycharges
- # totalcharges

Filters

- A churn not in ('null')

REFRESH VISUAL

DASH. Filters

Visuals Build

Style Settings

Style Build

Settings Build

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

VIEW LAYOUT SAVE PRIVATE

Churn Analysis

enter subtitle... Add dashboard filters by selecting dataset fields from the Filters menu on the right.

Churn by type of contract

Type of Contract	Percentage
Month-to-month	40%
One year	30%
Two year	30%

Churn by family

Gender	Churn Status
Male	Yes
Female	Yes

Churn by services

internet service	Record Count
No	~2000
Yes	~1000

VISUALS

DATA

telco

Sample Mode: OFF

Search

Dimensions

- telco_data_curated
- A customer
- A gender
- A seniorcitizen
- A partner
- A dependents
- A phoneline
- A multiplelines
- A internetservice
- A onlinesecurity
- A onlinebackup
- A deviceprotection
- A techsupport
- A streamingtv
- A streamingmovies
- A contact
- A paperlessbilling
- A paymentmethod
- # result
- # simpleresult
- A churn

X-Axis

- A partner

Y-Axis

- A gender

Colors

- A dependents

Size

- # avg(totalcharges)

Transition

- drag fields to add here

Tooltips

- drag fields to add here

Filters

- A churn not in ('null')

REFRESH VISUAL

Visuals

DASH.

FILTERS

Settings

Style

VISUAL

Build

Settings

Style

15. Add the third visual element, which is a bar chart with the dimensions **streamingtv** and **streamingmovies** like X Axis,

Record Count how Y Axis and **churn** how Colors. Once finished, click the button **Refresh Visual**.

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 search bar and a dropdown for 'PRIVATE'. The main area contains several visual elements:

- A large donut chart on the left labeled 'Month-to-month'.
- A smaller chart above it with 'No' and 'Yes' segments.
- A section titled 'Churn by services' containing three bar charts. Each chart has 'streamingtv' on the X-axis with categories 'No' and 'Yes'. The Y-axis is 'Record Count' ranging from 0 to 2,000. The legend indicates 'No' is teal and 'Yes' is light blue. The three charts correspond to different service types: 'No internet service', 'streamingmovies', and 'streamingtv'.
- A table titled 'Scoring - Churn Probability' below the charts. It has columns for result, customerid, tenure, monthlycharges, totalcharges, gender, dependents, onlinesecurity, multiplelines, and internetservice. The data rows show various customer details and service types like DSL or Fiber optic.
- To the right is the 'Dashboard Designer' panel, which includes sections for VIEWS, DATA, DASH, and BUILD. The BUILD section is highlighted with a red border and shows the configuration for the three bar charts, including the X and Y axes, colors, and filters.

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

CLOUDERA Data Visualization

HOME SQL VISUALS DATA

VIEW LAYOUT SAVE PRIVATE

The interface includes a top navigation bar with Home, SQL, Visuals, and Data tabs. Below this are View, Layout, Save, and Private buttons. On the right side, there's a sidebar titled "Dashboard Designer" with sections for Dimensions, Measures, Tooltips, Filters, and a Build tab.

Dimensions:

- customerid
- i2 tenure
- i2 monthlycharges
- i2 totalcharges
- A gender
- A seniorcitizen
- A dependents
- A onlinesecurity
- A multiplelines
- A internetservice
- A seniorcitizen
- A techsupport
- A contract
- A streamingmovies
- A deviceprotection
- A paymentmethod
- A streamingtv
- A phoneservice
- A paperlessbilling
- A partner
- A onlinebackup
- A deviceprotection
- A techsupport
- A streamingtv
- A streamingmovies
- A contract
- A paperlessbilling
- A paymentmethod
- # result
- # simpleresult
- A churn

Measures:

- # Record Count
- i2 tenure
- i2 monthlycharges
- i2 totalcharges

Tooltips:

- drag fields to add here

Filters:

- drag fields to add here

Build:

Limit: 100

REFRESH VISUAL

Scoring - Churn Probability									
customerid	tenure	monthlycharges	totalcharges	gender	dependents	onlinesecurity	multiplelines	internetservice	seniorcitizen
7590-VHVEG	1	32.602622985839844	29.85000381469727	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

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