Cloudera Data Visualization 8.0.9

Working with Derived Data

Date published: 2020-10-30
Date modified: 2025-12-16

CLOUDERA

https://docs.cloudera.com/

Legal Notice

Contents

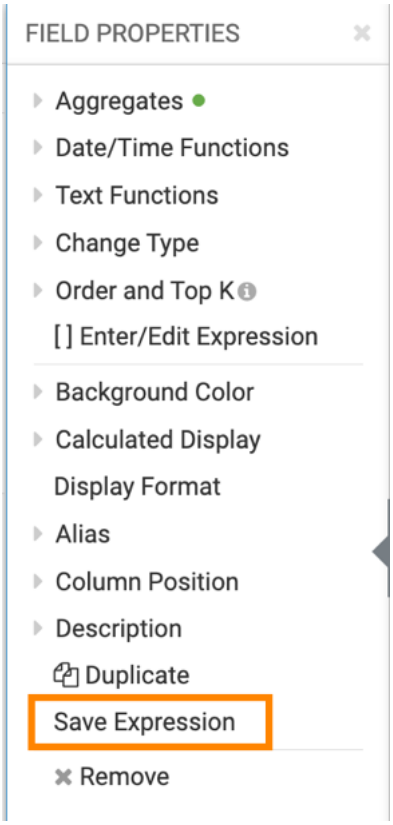
Defining derived data

Procedure

1. Inside a visual that generates a subset of data, select a measure field on a shelf.

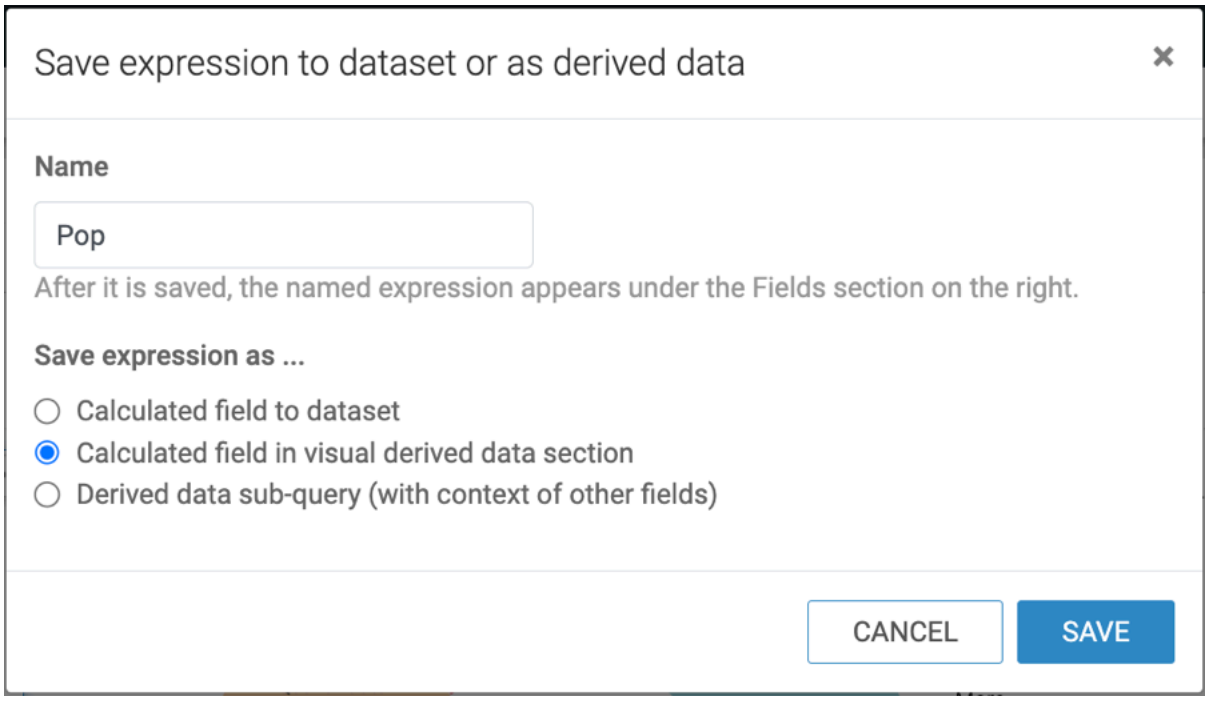
   In this example, sum(population) is used.
2. Click the Right Arrow icon to the right of the field.

3. In the dropdown, click Save Expression.

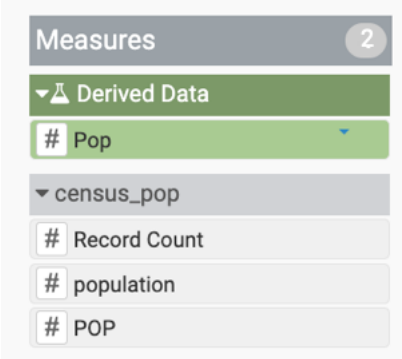


4. In the Save expression to dataset or as derived data modal window, under Name, type POP.

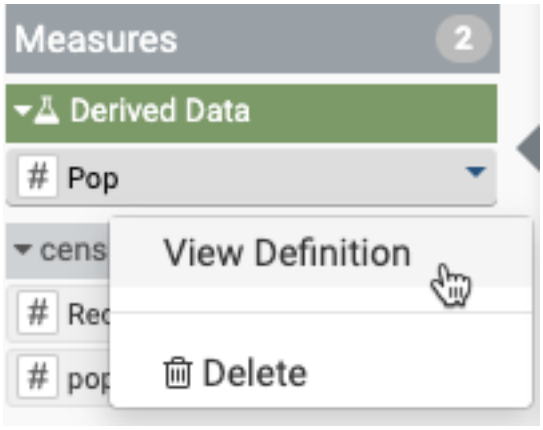
5. Select Calculated field in visual derived data section and click SAVE.



You can see that in the right navigation under Measures, you have a new entry, POP.



6. To check the definition of the saved derived data POP, click the down arrow to the right of its name, and click View Definition.



The Derived Data Definition modal window appears. You can see that the definition includes a parameterized filter for year.

7. Click SAVE to save both the visual and the new derived data definitions.

Results

You can now use Pop as any other measure.

Defining additional derived data

About this task

Next, define other derived data in the same shelf set-up and save it as derived data LifeExpXPop.

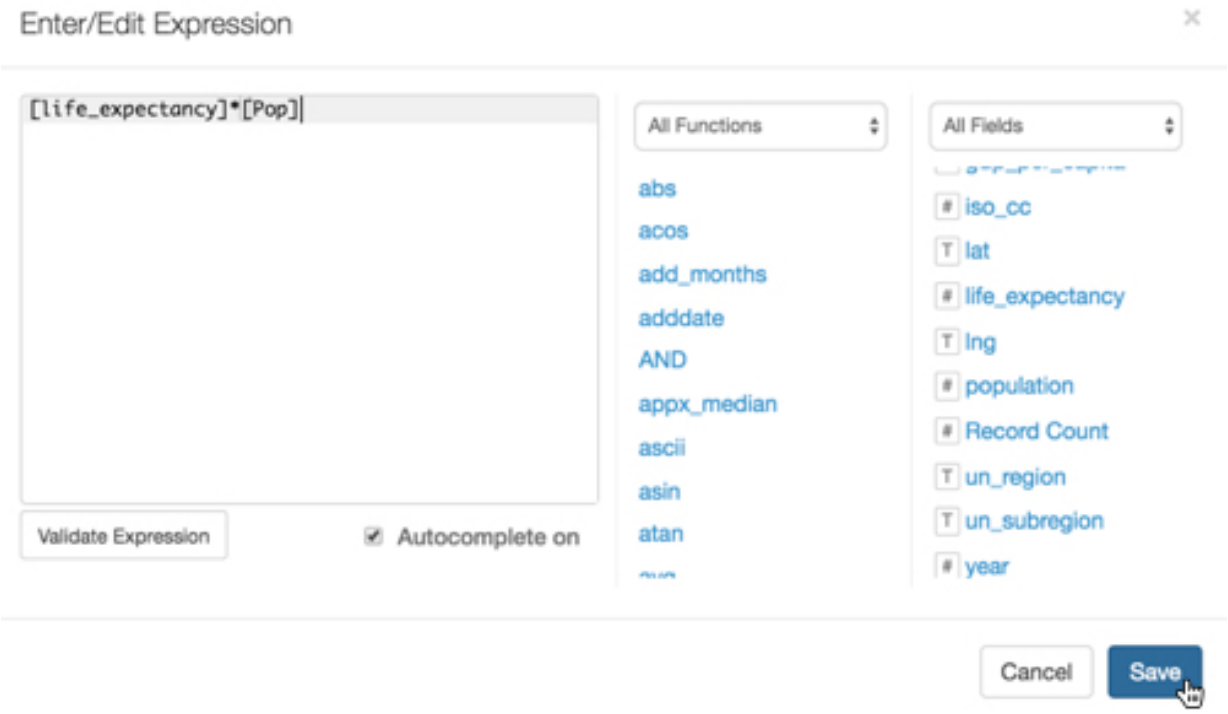
Procedure

1. On the Dimensions shelf, place the following fields: un_region, un_subregion, and country.
2. Add the derived data Pop to the Dimensions shelf.

   You should do this because fields on the Measure shelf must be aggregated, but the definition of Pop already includes the summation over the population field.
3. Place the field life_expectancy on the Measures shelf and change the aggregate to avg(life_expectancy).

4. Place the field year on the Filters shelf and change the expression to [year]=<<year_param:2000>>.

5. Duplicate Derived data Pop in place, on the Dimensions shelf.
6. Click the Down Arrow icon to the right of the cloned field and click Enter/Edit Expression in the dropdown.
7. In the Enter/Edit Expression modal window, change the expression to [life_expectancy]*[Pop] and save it.



8. Save the same field as derived data, named LifeExpXPop.

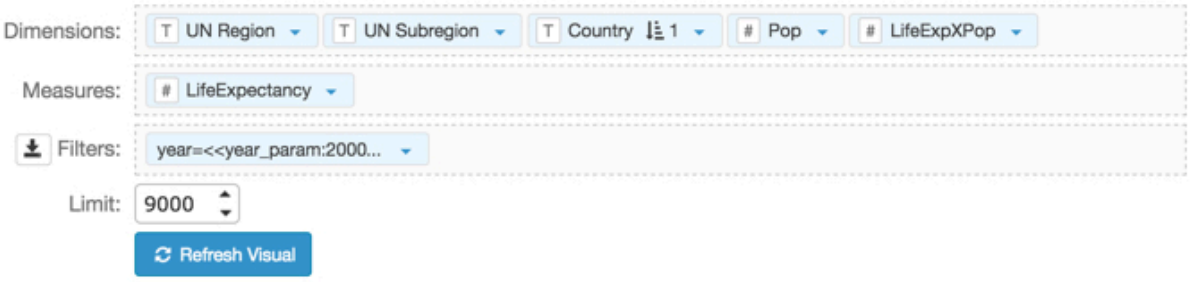
   You can see that in the Data menu, under Dimensions, you have a new category Derived Data that contains Life ExpXPop.

9. To check the definition of the saved derived data LifeExpXPop, click the down arrow to the right of its name and click View Definition.

   The Derived Data Definition modal window appears. Note that the definition includes a parametrized filter for year, and the Pop derived data field.

10. Replace the field on the Dimensions shelf with the new derived data field, click Refresh Visual, and Save the application.

    The visual should look similar to this one. Note that we aliased the fields and made some changes to display formats:



UN Region	UN Subregion	Country	Pop	LifeExpXPop	LifeExpectancy
Asia	Southern Asia	Afghanistan	22,856,302	1,254,811,040	54.90
Europe	Southern Europe	Albania	3,071,856	227,931,710	74.20
Africa	Northern Africa	Algeria	30,533,828	2,106,834,175	69.00
Africa	Middle Africa	Angola	13,926,373	629,472,083	45.20
Americas	Caribbean	Antigua and Barbuda	77,656	5,707,716	73.50
Americas	South America	Argentina	36,930,708	2,725,486,419	73.80
Asia	Western Asia	Armenia	3,076,098	219,941,012	71.50
Americas	Caribbean	Aruba	90,271	6,662,000	73.80
Oceania	Australia and New Zealand	Australia	19,164,352	1,531,231,785	79.90

< 1 2 3 4 5 >

Using derived data

About this task

To use the derived data measurements in a calculation, treat it as any other measurement. In this example, the derived data LifeExpXPop is used to accurately calculate the weighted life expectancy for an aggregation dimension, such as UN Region or UN Subregion. You can do this by dividing the sum of Life Expectancy X Population by sum(populati on) at the appropriate aggregate level.

Procedure

1. Place the following fields on the Dimensions shelf: un_region and un_subregion.
2. Add the derived data field Pop.
3. Place the field year on the Filters shelf , and change the expression to [year]=<<year_param:2000>>.
4. Place the derived data field LifeExpXPop onto the Meausres shelf. Notice that it appears as an aggregation, sum(LifeExpXPop).

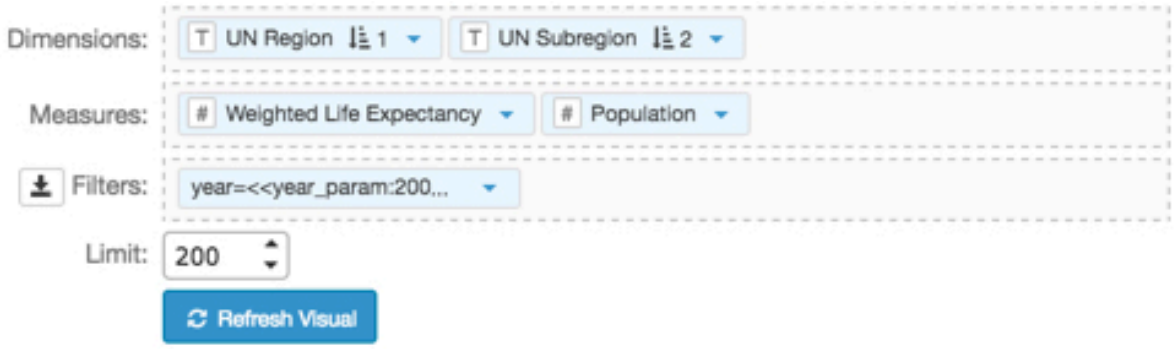
5. Using the Enter/Edit Expression window modal, change the definition of the field as follows,and click Save. You can see that the field has been aliased.

```
(sum([LifeExpXPop])/sum([population])) as 'Weighted Life Expectancy'
```

6. Use the Alias option to change the column names to UN Region, UN Subregion

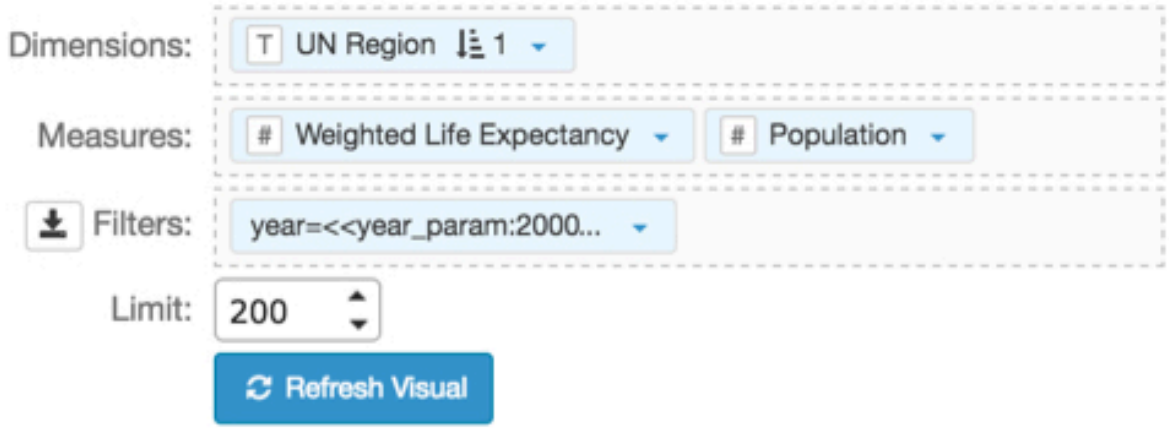
7. Click Refresh Visual.

   The visual should look similar to this one:



UN Region	UN Subregion	Weighted Life Expectancy	Population
Africa	Eastern Africa	52.37	108,167,918
Africa	Middle Africa	47.42	89,208,823
Africa	Northern Africa	68.72	98,182,244
Africa	Southern Africa	50.60	1,757,925
Africa	Western Africa	54.14	49,296,775
Americas	Caribbean	68.85	23,593,157
Americas	Central America	70.25	27,565,249
Americas	Northern America	79.38	30,723,560
Americas	South America	71.35	279,783,195
Asia	Central Asia	64.17	19,911,544
Asia	Eastern Asia	72.97	1,401,620,301
Asia	South-Eastern Asia	67.01	225,842,357
Asia	Southern Asia	62.31	1,206,917,948
Asia	Western Asia	71.56	54,154,366
Europe	Eastern Europe	70.78	28,274,513
Europe	Northern Europe	75.83	18,353,582
Europe	Southern Europe	78.82	79,244,298
Europe	Western Europe	78.50	159,577,216
Oceania	Australia and New Zealand	79.90	19,164,352
Oceania	Melanesia	67.60	811,718
Oceania	Micronesia	71.35	239,157

8. [Optional] To see the correctly calculated Life Expectancy for a UN Region, remove UN Subregion from the Dimensions shelf.



UN Region	Weighted Life Expectancy	Population
Africa	55.97	346,613,685
Americas	71.79	361,665,161
Asia	68.00	2,908,446,516
Europe	77.65	285,449,609
Oceania	79.22	20,452,865

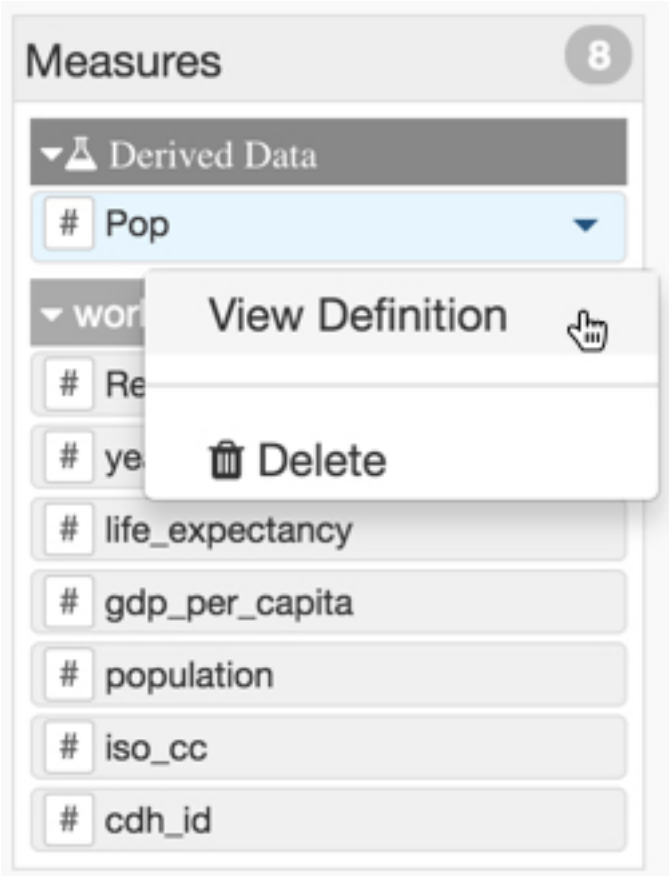
Viewing derived data definitions

Procedure

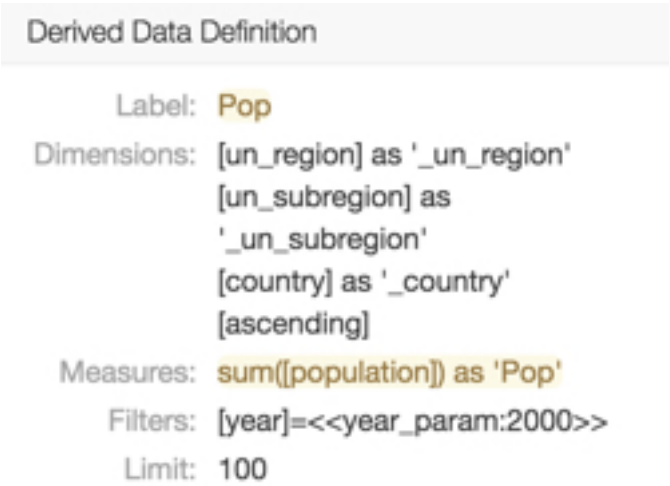
1. In the left navigation, under Measures Derived Data , select the derived measure you want to check.

   In this example, it is Pop.
2. Click the Down Arrow icon to the right of the field.

3. In the drop-down, select View Definition.



The Derived Data Definition modal window appears.



Note: You cannot edit the definition of derived data. If you want to change it, you must first delete it and then create and save a new definition. For more information, see *Deleting derived data definitions* and *Defining derived data*.

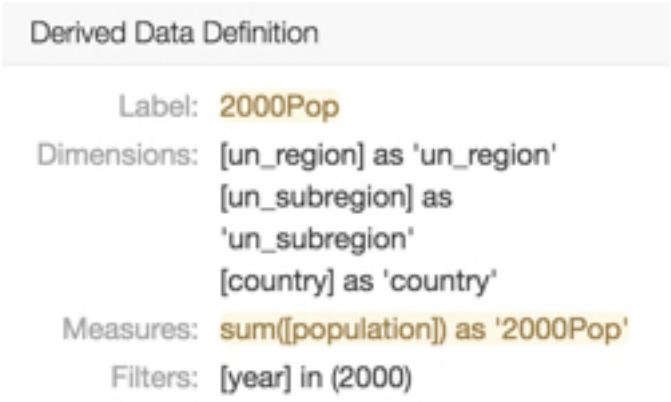
Related Information

Deleting derived data definitions

Defining derived data

Saving derived data with full context

This is the default option, when the entire context of the derived data expression can be captured and re-used. In he earlier example, the definition of derived data includes Label, Dimensions, Measures, and Filters:



Saving derived data expression only

Procedure

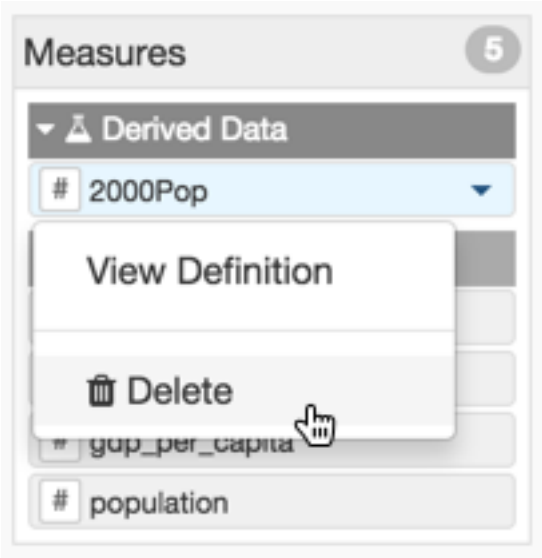
Deleting derived data definitions

Procedure

1. In the left navigation under Measures Derived Data , identify the derived measure you want to delete.

   In this example, it is 2000Pop.
2. Click the Down Arrow icon to the right of the field.

3. In the drop-down, select Delete.



4. Click Save to save both the visual and the changes to derived definitions.