End-User Instructions

Updated 4/15/22
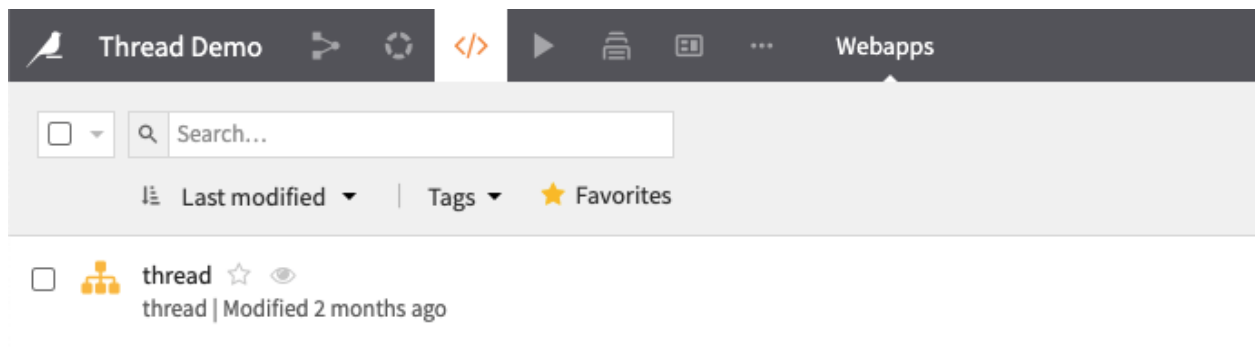
Getting Started

You must be logged into Dataiku to access Thread. To access Thread Data while in Dataiku go to your Thread project and then Webapps
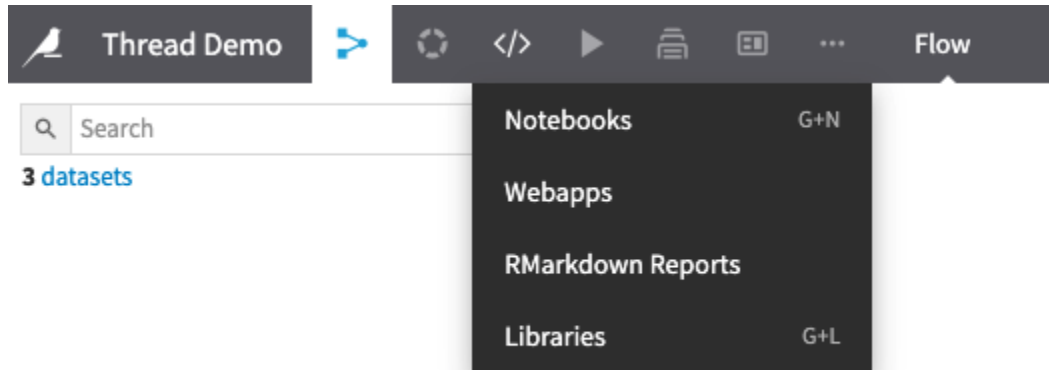
Select Thread Data

To open this as a public webapp click the link

Home Screen

Search

This search bar allows you to search for columns, datasets, projects and definitions. Uncheck the boxes to filter the results.

Dataiku Instance Stats

This is a summary of all projects, datasets, columns, and definitions in **only this node** of Dataiku that have been indexed.
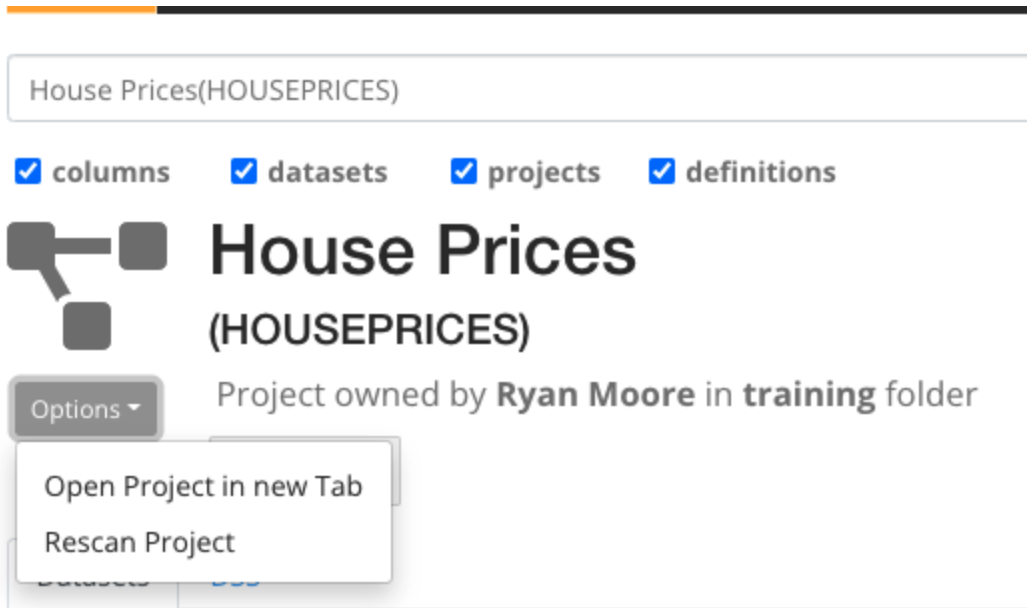
View Search Results from Home Page

Project

You can see on this page
1. Project Name
2. Project ID
3. Project owner
4. Project folder
5. Project Tags
6. Datasets tab
   a. Name: All datasets in this project. Clicking this will take you to this dataset's result page
   b. Documentation Status:How many columns have been documented in each data set
7. DSS tab
   a. Project flow in Dataiku
8. Documentation Status of all columns in the entire project
9. Options
   a. Open project in new tab: new tab to view this project information page
   b. Rescan project: for bringing in new datasets and columns from Dataiku

Dataset

You can see on this page
1. Dataset Name
2. Storage type of the dataset
3. Project Name of this data set. You can click this to view the project results page
4. Columns
   a. Name: Column Name of all columns in this dataset
   b. Type: Data type
   c. Description: Column Description
5. Lineage

a. Source dataset (even if in a different project)
b. Current dataset highlighted in blue
c. Final dataset(s)
d. Numbers between the boxes indicate count of datasets between these datasets
e. The blue arrow in the top right corner of each box allows you to jump to that datasets results page
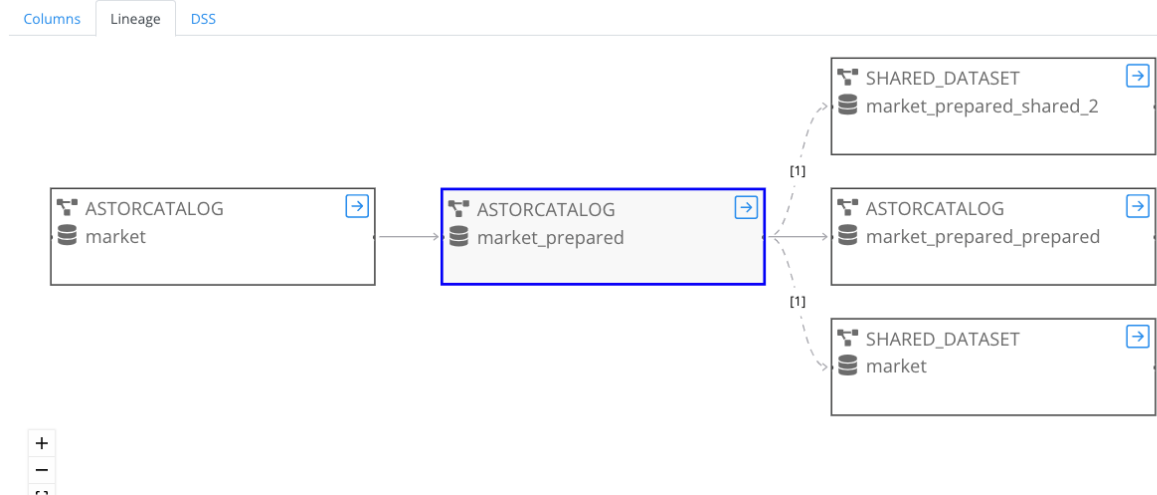
6. DSS: View the dataset explore page in Dataiku
7. Documentation Status: Total number and percentage of columns in this dataset that have been defined
8. Options
    a. Open Dataset in a new Tab

Column

You can see on this page
1. Column Name
2. Data type
3. Project Name. You can click this to view this project results page
4. Dataset name. You can click this to view the dataset results page
5. Definition
   a. Definition Name
   b. Definition Tags
   c. Description of definition
6. Lineage
   a. Source dataset of this column (even if in a different project)
   b. Dataset the column you are viewing is located in highlighted in blue
   c. Final dataset(s) this column goes to
   d. Numbers between the boxes indicate count of datasets between these datasets
   e. The blue arrow in the top right corner of each box allows you to jump to that dataset results page
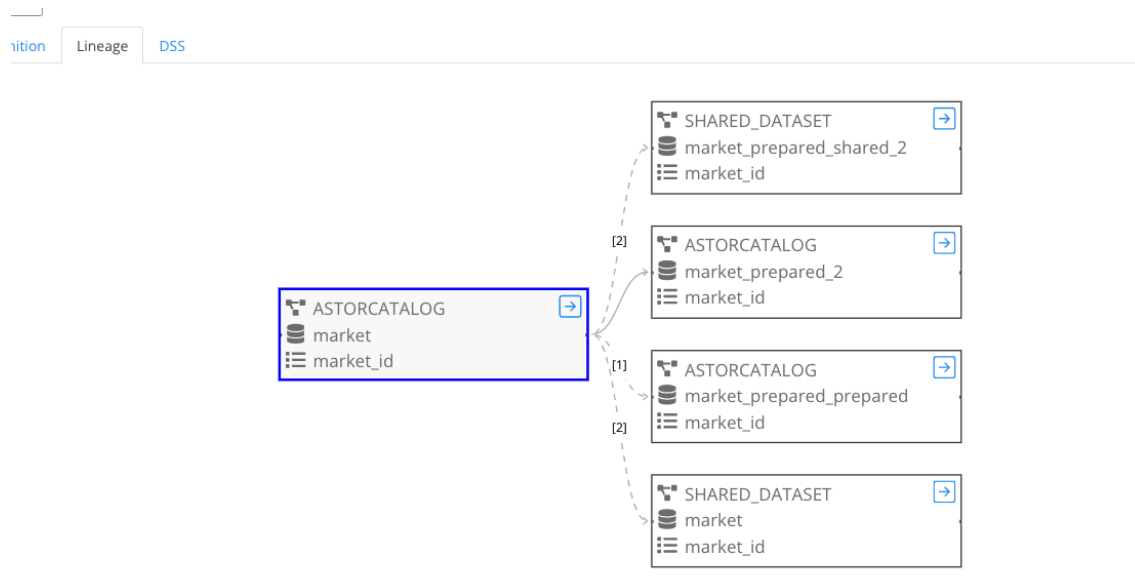
7. DSS: View the dataset with this column in Dataiku explore page
8. Documentation Status: Whether or not this column has a definition
9. Options
   a. Open Dataset in a new Tab
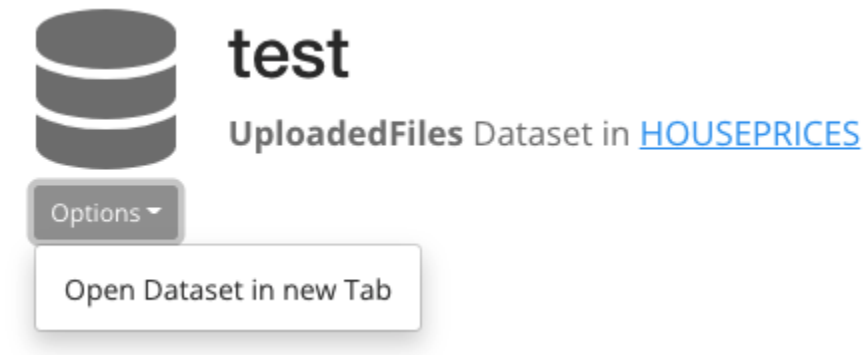
Create & Edit Definitions

To create, edit or remove a definition from a column find the column using the home screen search.

Apply a Definition to a Column

Click Add Definition

If a definition for this column already exists, "Search for Definition" in the bottom left corner. Find the correct definition and apply to this column.

If this is a brand new definition enter:
1. Name of definition: This is a business-friendly name for the definition that may be applied to many columns that do not have the same name.
2. Tags: Apply new tags or existing tags by typing the tag name
3. Description: definition of the word

Save and Apply
1. Save and Apply to column: only applies the definition to this one column
2. Save and apply to lineage: will apply this definition to all columns in this lineage
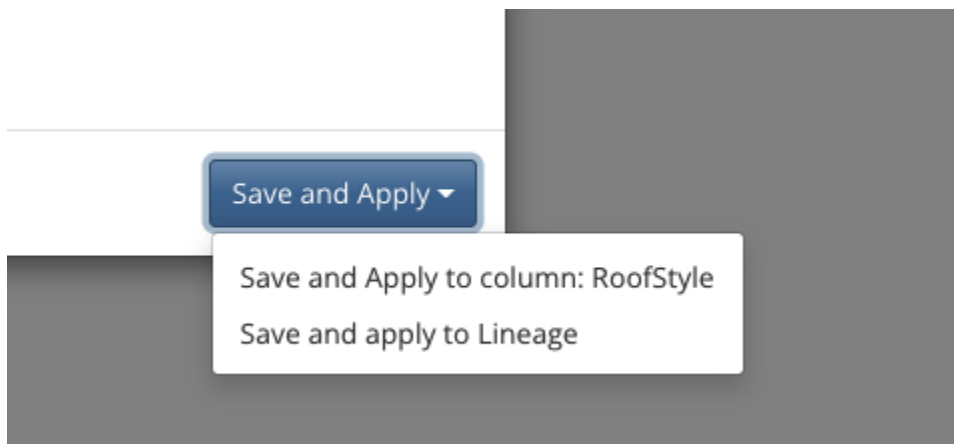
3. If you save and apply to lineage, select which datasets you would like this to apply to in the lineage

Apply Definition To Lineage ✕

Select the up and downstream datasets to apply this definition

This Column
🔵 RoofStyle

Downstream
⚪ HOUSEPRICES|test_train_joined|RoofStyle

⚪ HOUSEPRICES|test_train_joined_prepared|RoofStyle

Cancel   Apply
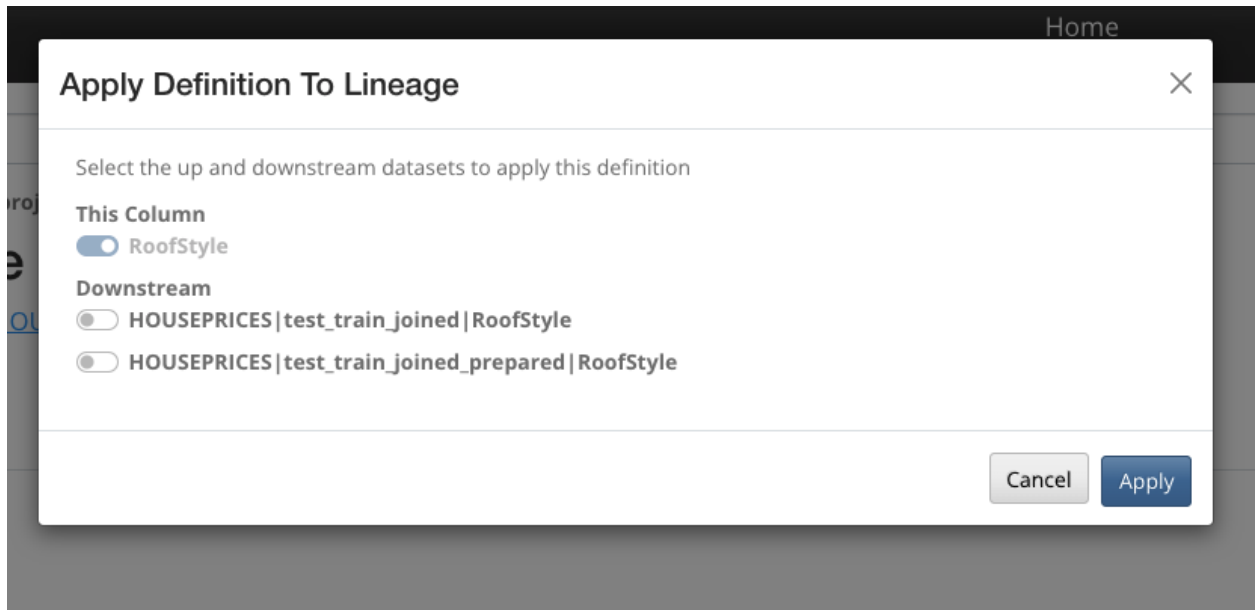
Catalog Screen

View all definitions on this node of Dataiku. One definition may be applied to many columns, even if they don't have the same name.

You are able to view
1. Definition name: Sort this alphabetically ascending or descending by pressing the arrow next to the title. Click this link to edit or delete (using options) the definition
2. Description: Sort this alphabetically ascending or descending by pressing the arrow next to the title
3. Tag: Customizable metadata added when assigned to a column

Navigation
1. Search by name, definition or tag using the top right search bar
2. Filter by tag in the filter bar
3. Filter by tag by clicking a tag

Click on the name to edit or delete
From this page you can view or edit
1. Column Name
2. Name of the definition (this does not need to match the column name)
3. Tags
4. Applied to: if you click this will take you to that dataset results
5. Description: definition in business term

Phone Number
Definition

Options ▾

Details

Name

Phone Number

Tags

pii ✕   Add new tag

Applied To

DKU_CHURN | Churn_Operations | Phone   DKU_CHURN | customers1 | Phone   DKU_CHURN | unlabeled_customers_stacked_by_Phone_joined | Phone   DKU_CHURN | unlabeled_customers_stacked_by_Phone | Phone

DKU_CHURN | unlabeled_customers_stacked | Phone   DKU_CHURN | unlabeled_customers | Phone   DKU_CHURN | customers | Phone   API_DEMO | api_output | Phone

API_DEMO | DKU_Churn_Operations_prepared | Phone

Description

The Phone Number

Will appear in the Dataiku Dataset's column description.

Save

To delete this definition options > delete

☑ columns   ☑ datasets   ☑ projects   ☑ definitions

Phone Number

Definition

Options ▾

Delete

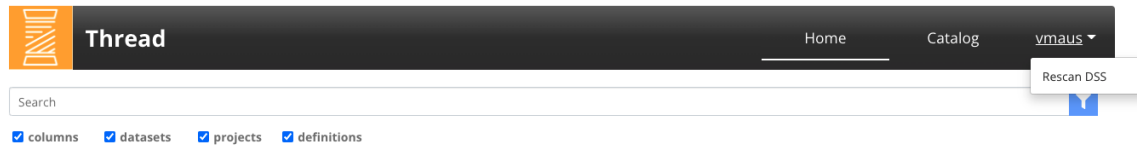## Name

Phone Number

Tags

Adding New Data

In order to add new projects, datasets and columns from Dataiku into thread you need to rescan DSS using the arrow next to your username.

Sending Links

You can share a link to a specific page in Thread with anyone who has security access in Dataiku and permission to "read project content" in the Thread project.

FAQ

Why can't I see my project, dataset, or column in the search results, or lineage screen?

Thread needs to index any new projects, datasets, or columns. You can do this for all projects datasets and columns in your instance by going to the arrow in the top right corner next to your username.

Alternatively, you can do this per project by searching for that project on the home screen and using options > Rescan Project. This method is suggested for those instances with a lot of projects.
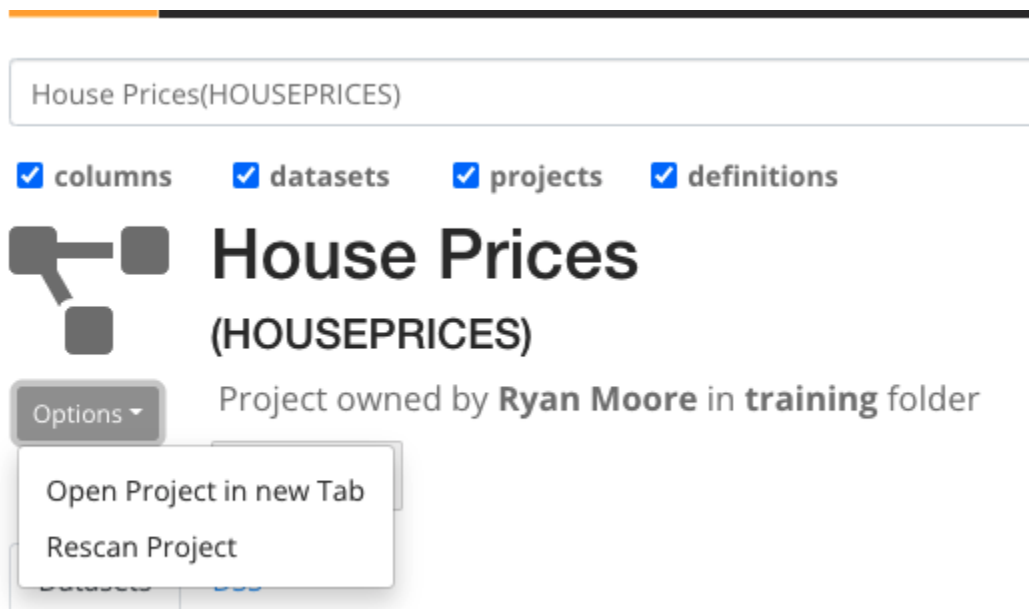
How do I know if a definition already exists for this column?

From the home screen search for the column you would like to define. When you add definition use the "Search for Definition" in the bottom left corner. If you find the definition you can apply it

here.

Alternatively, you can view all existing definitions in the catalog before going to edit the column.

Can everyone see all the data in my instance?
User must have Dataiku access to use Thread and permission of "read project content" to the Thread project. All users can see the column names and definitions for all projects on the instance. If the user doesn't have access to a particular project, they cannot view the DSS tab in the project, dataset or column results screens.

Who can view, edit or delete definitions?
Any Dataiku user with "read project content" permission to the Thread project can view, edit and delete definitions in Thread and the column descriptions in Dataiku.

What if I change the name of a column?
Thread is able to recognize the same column with a different name as the same lineage if the column was renamed in a prep recipe.

How do I know how much is documented?
You can view by project, dataset, or column how much has been documented.

Can I have two Thread webapps running at the same time?
Yes, you can have two Thread webapps but they need to be in separate projects.