**Trifacta User Guide**

**Team Data Wranglers**

**GMU DAEN 690 Section 4**

**Professor Robert E Kraig**

[**Table of Contents**](#_30j0zll)**: Page**

1 [Introduction  3](#_1fob9te)

[2](#_3znysh7)  Installation\Sign Up 4

[3](#_2et92p0)  Import Data 6

4 Flow/Recipe Creation 8

5 Run Job 24

6 Output 27

[7](#_4d34og8) Customer Support 28

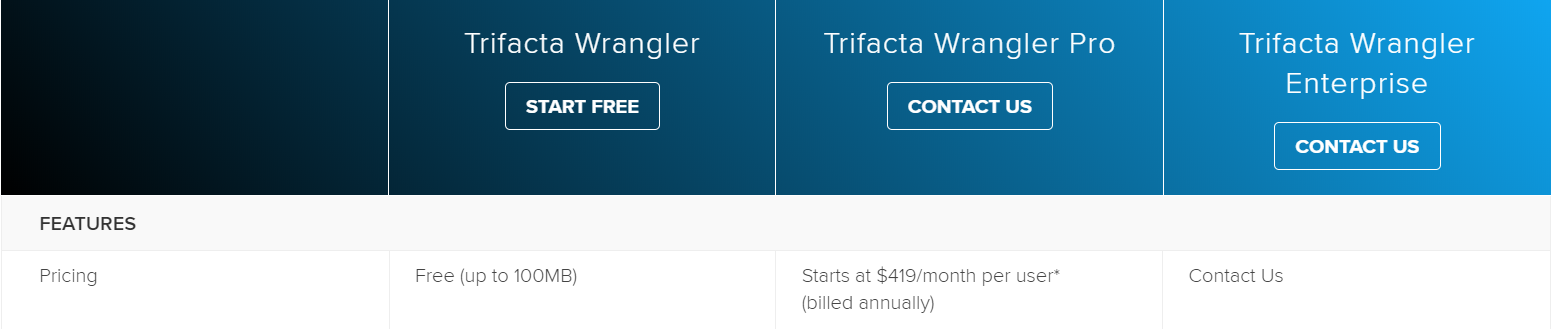
**1 Introduction:**

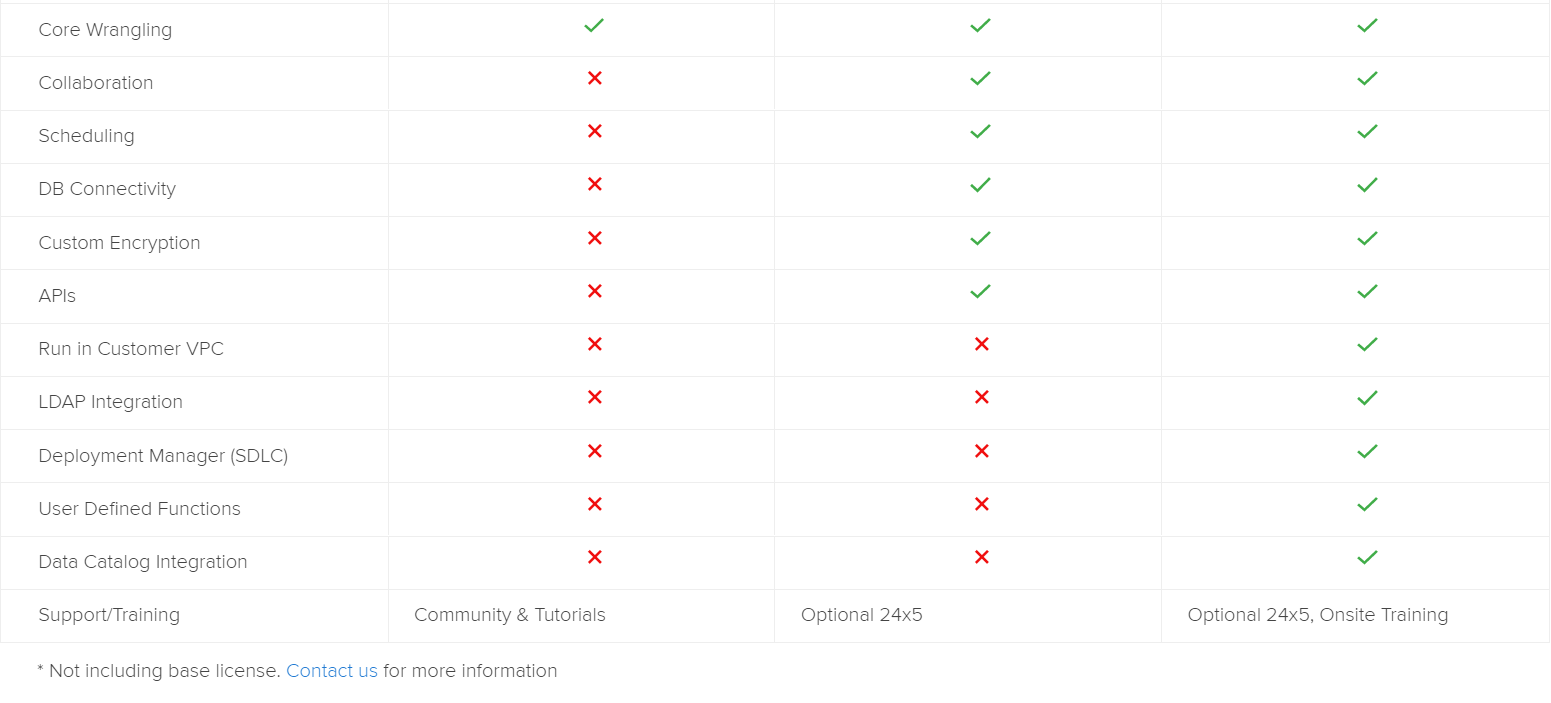
The contents of this document provide an insight to working on the data prep tool “Trifacta”. This document gives a step by step information from the installation of the tool to recipe creation and finally to get the output with the 64 standard attributes as required by the client “Share Our Strength”. The information in this document can be used by the client to understand the tool better and work on it without any difficulties in the recipe creation.

**2 Installation /Sign Up:**

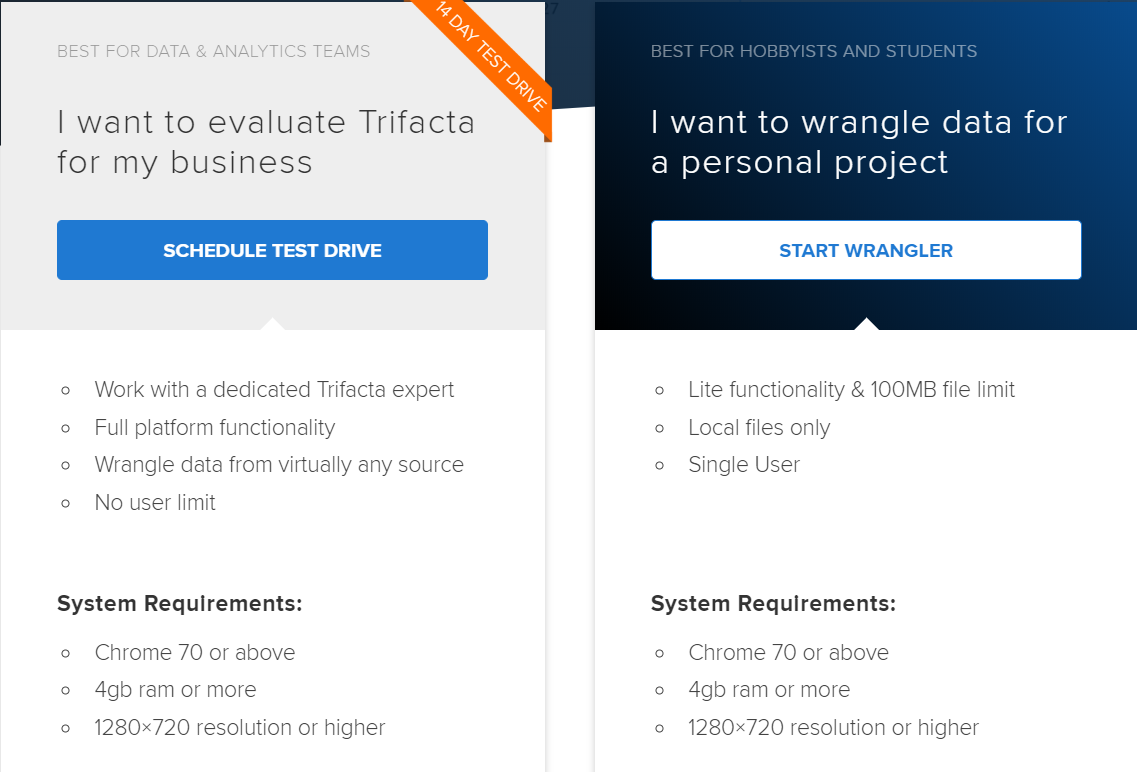
Trifacta is a completely cloud based tool therefore it does not require any installation on your local PC. Initially we have to sign up using the browser and select between the three options provided by Trifacta. The three options are:

* Trifacta Wrangler (Free Version)
* Trifacta Wrangler Pro (Professional Version)
* Trifacta Wrangler Enterprise (Enterprise Version)

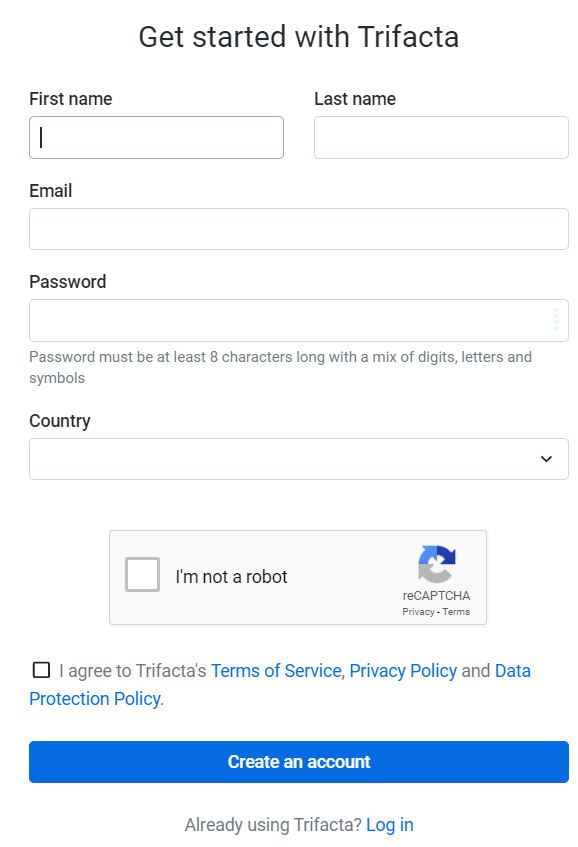




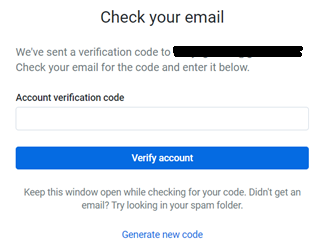
Based on the version selected we have to provide details to Trifacta for sign up. The Free Version can be opted in two variants:



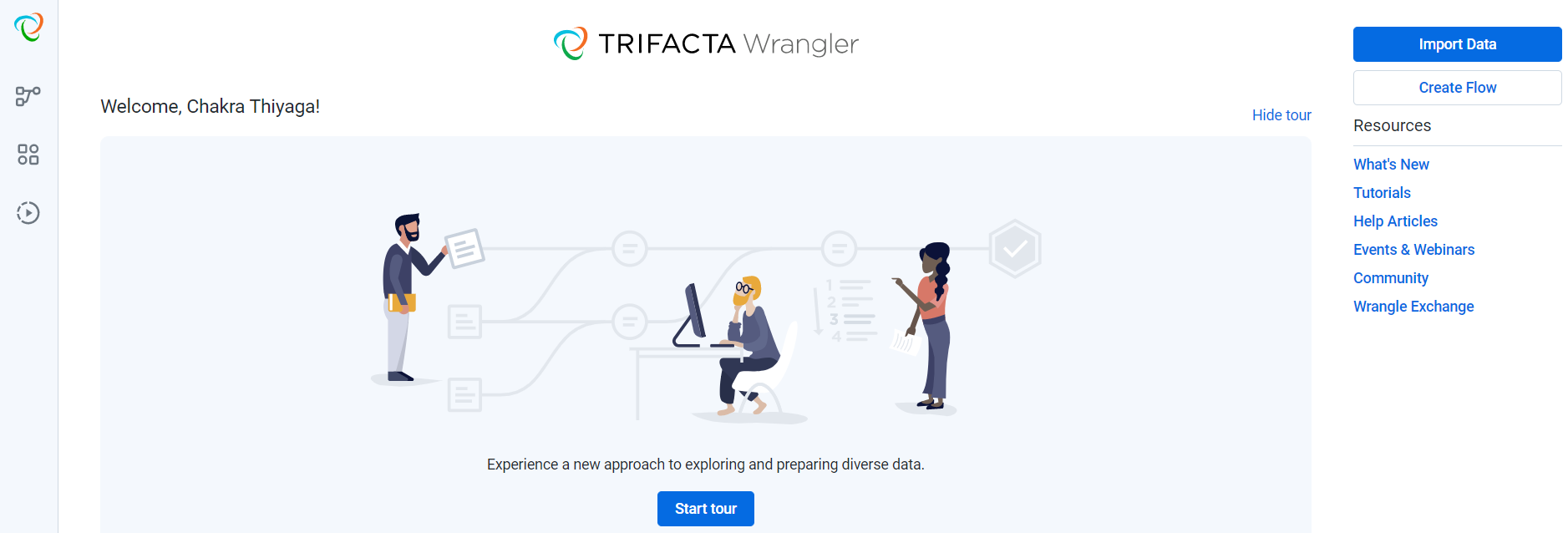
Since we used the free version of the tool: The following details were required for sign up:



Once we provide the details, it sends a verification to the email account in order to verify the email account provided.



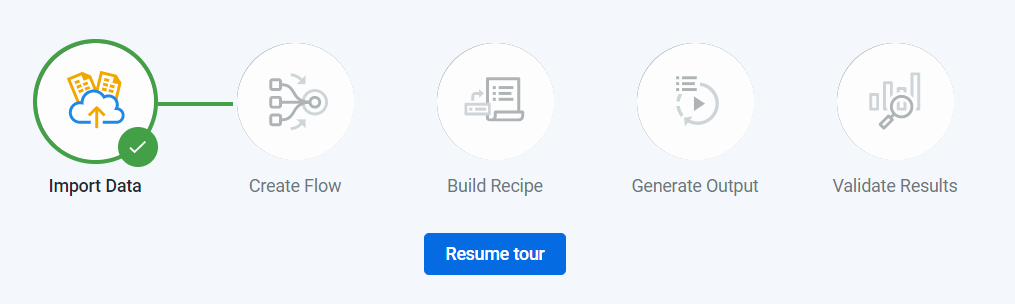
Then you enter the home screen of Trifacta tool:



You can start the tour where it gives a short but very informative glimpses of how the Trifacta tool operates from importing data to validate results.

**Note:**

Please do not skip this tour as it can be very helpful in understanding the operation structure of the tool.



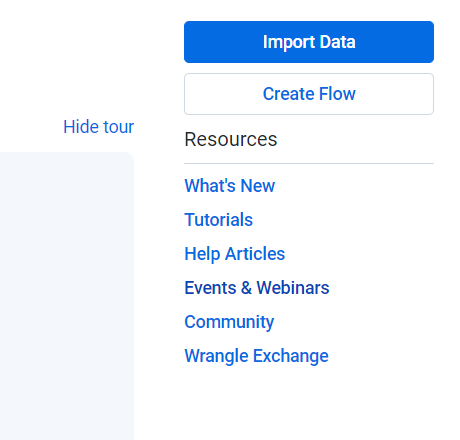
**2. Import Data:**

Importing the data into the Trifacta is an important process which enables us to bring in data from our personal computer. Based on the work done with No Kid Hungry data, every input dataset where in “Excel” format.

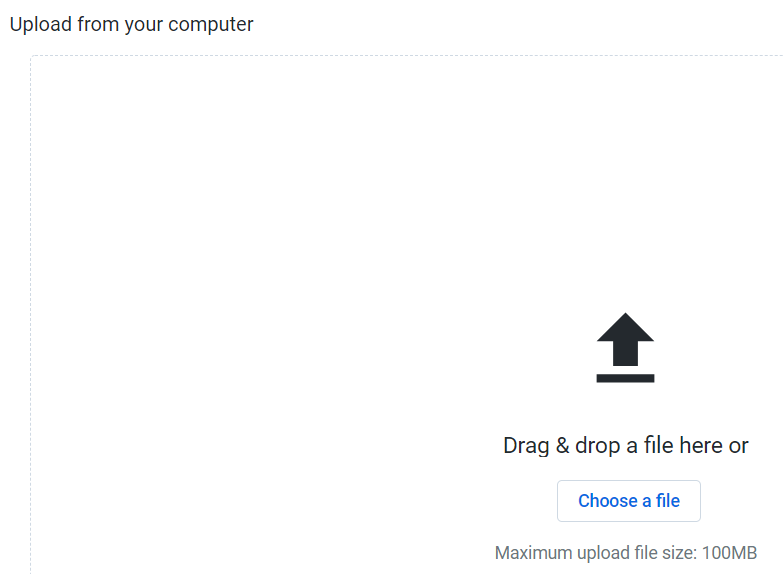
**Note:**

Trifacta considers each sheet in the Excel Workbook as a separate dataset. For Example: If we import an Excel workbook with three sheets, the Trifacta will consider it to have three datasets.

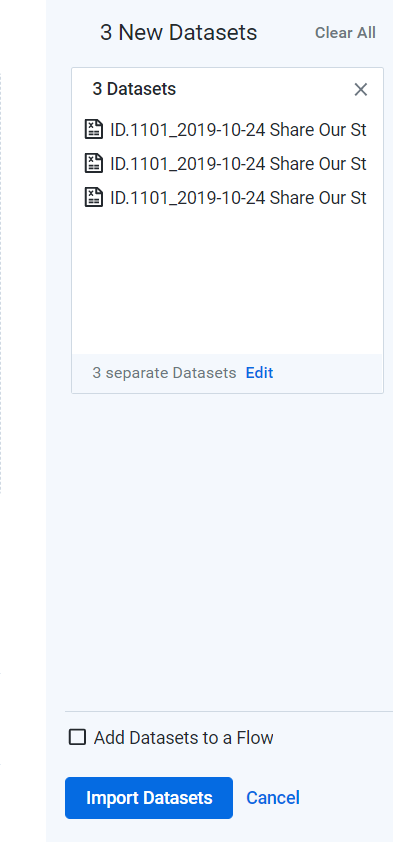
Trifacta can also handle different data sources other than Personal Computer. It can handle data sources such as “Hadoop File Distribution System”, Relational Databases (SQL server, IBM, Oracle, MySQL) and Cloud based databases (Tableau Server, Amazon AWS, Azure, Google Cloud)



Since we utilized the free version, we can upload the data from the local computer.

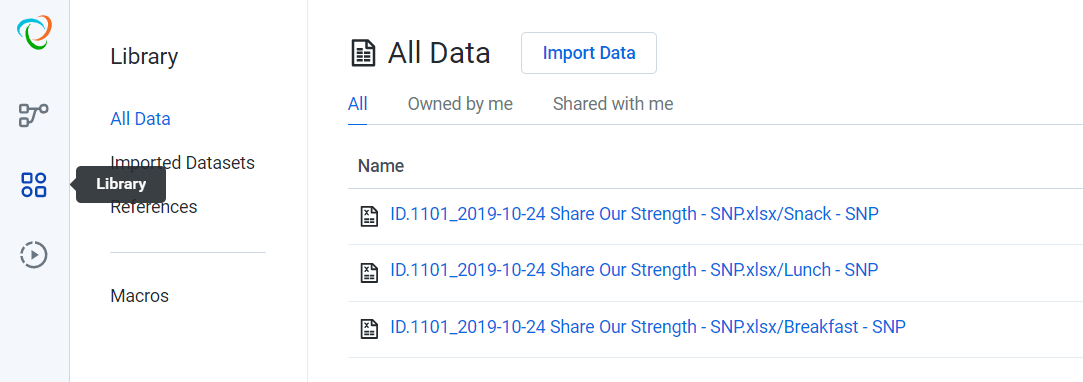


As mentioned above, I uploaded a raw data file that contained three different sheets, which is recognized by Trifacta as three different datasets.



Once Uploaded, click on Import Dataset in the bottom right corner.

If you check the “Library” Icon in the top left corner of the tool, we can notice the dataset that have been uploaded to the Trifacta tool.



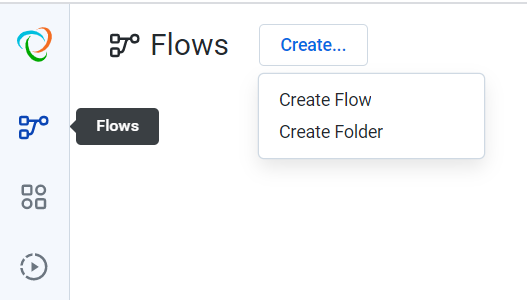
**Note:**

If we create any macros in the recipe, these macros will be represented under the Macros tab in the library. We can also import Macros from other Trifacta flows.

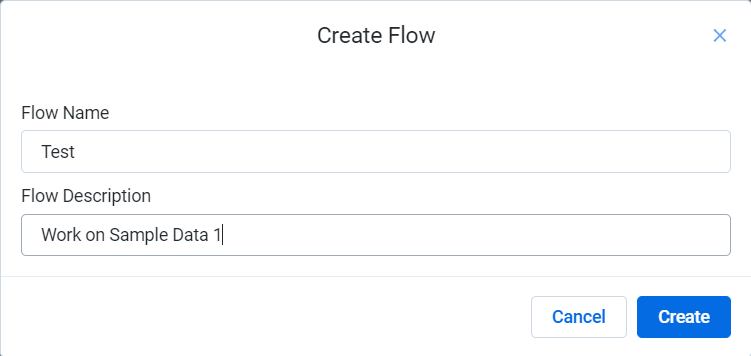


**3 Flow/Recipe Creation:**

Flow represents the step by step process in which we create recipes on the imported data in order to get the standard 64 fields.

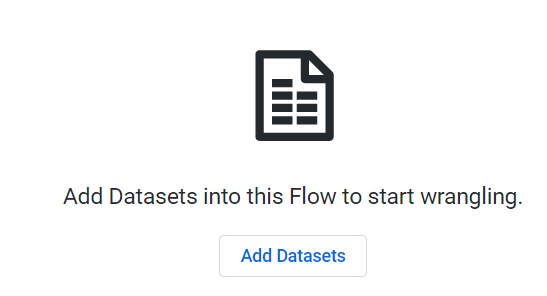


We can create a folder to contain all the flows, if in case a recipe has multiple flows. Otherwise click on “Create Flow”



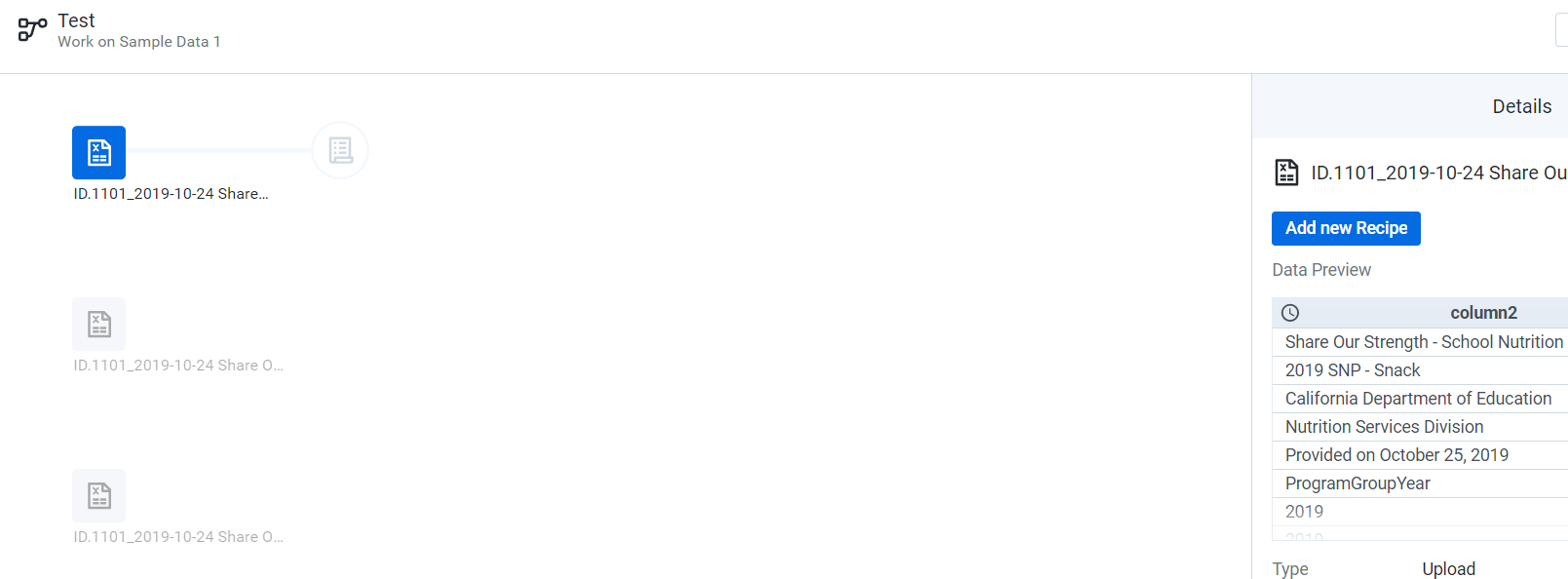
Fill in the requested details so that anyone else referring the created flow can understand the functionality of the flow.

Once the flow has been created, we have to add the datasets on which the recipe has to be created.

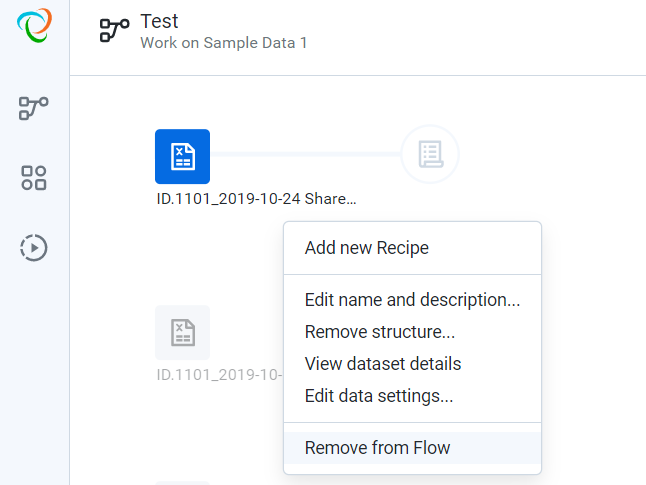




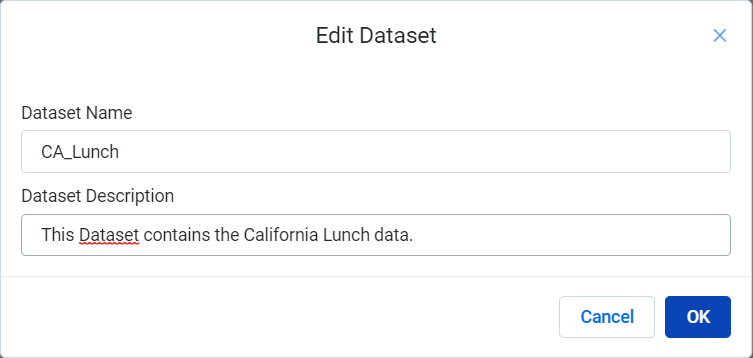
We can select the datasets the are required in the Flow. If you want to import new datasets, we can use the “Import Datasets” option on the left bottom of the window.



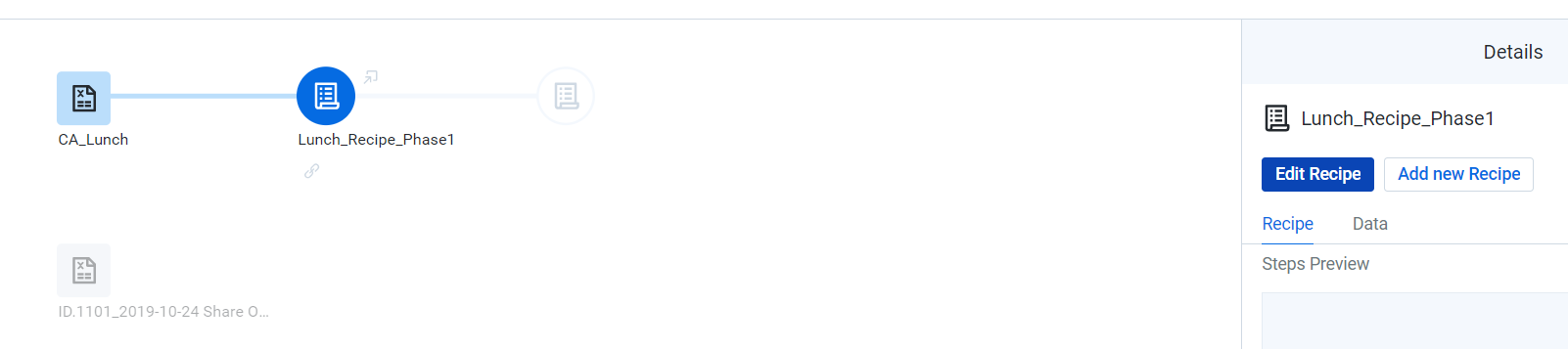
In case we want to remove any datasets from the flow:



We can edit the name and description on the dataset, so that it is understandable.

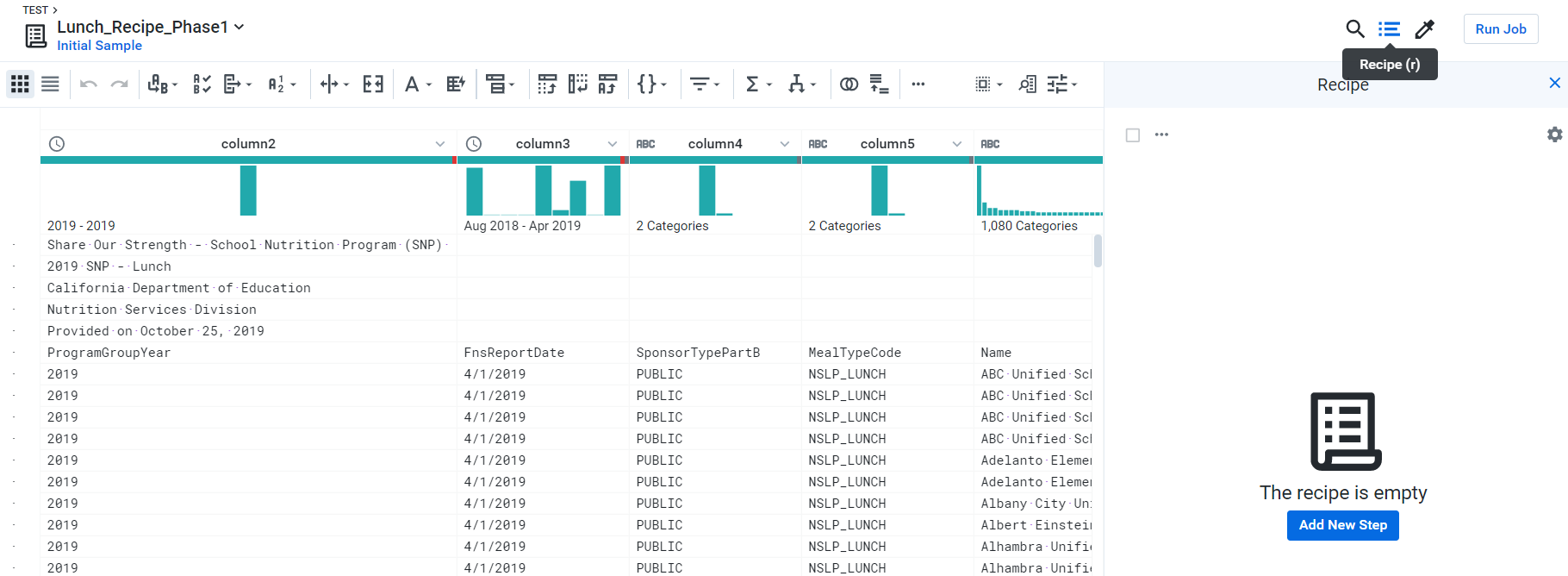


We can add a recipe by selecting the “Add New Recipe”. Then select “Edict Recipe” to add functions to the recipe.



We are provided with the below window:

If we click on the “Recipe” Tab, we can new step or function to the recipe using “Add New Step”.



We can also add functions by clicking on the rows or columns in the dataset.

**Delete Rows:**

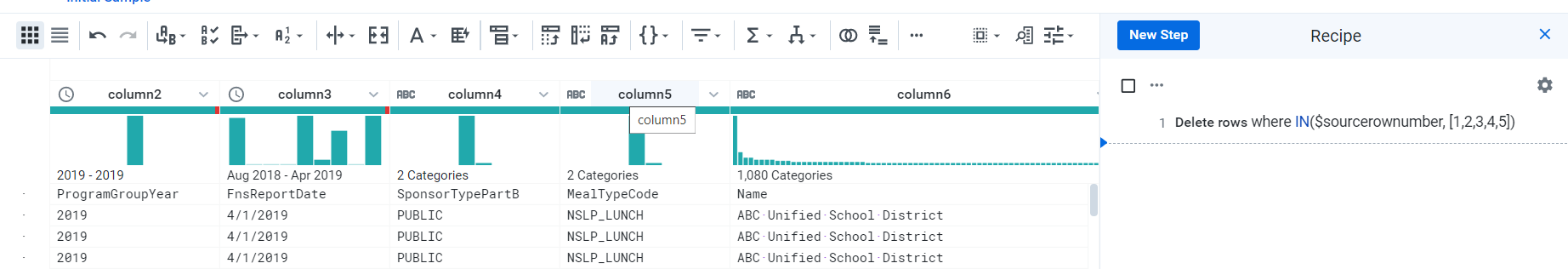
When we select the rows or columns, we are suggested with functions as shown below, we can perform “Rename”, “Delete rows” and “Keep Rows” functions.

We can add the function, by selecting the suggested function and click “Add”.



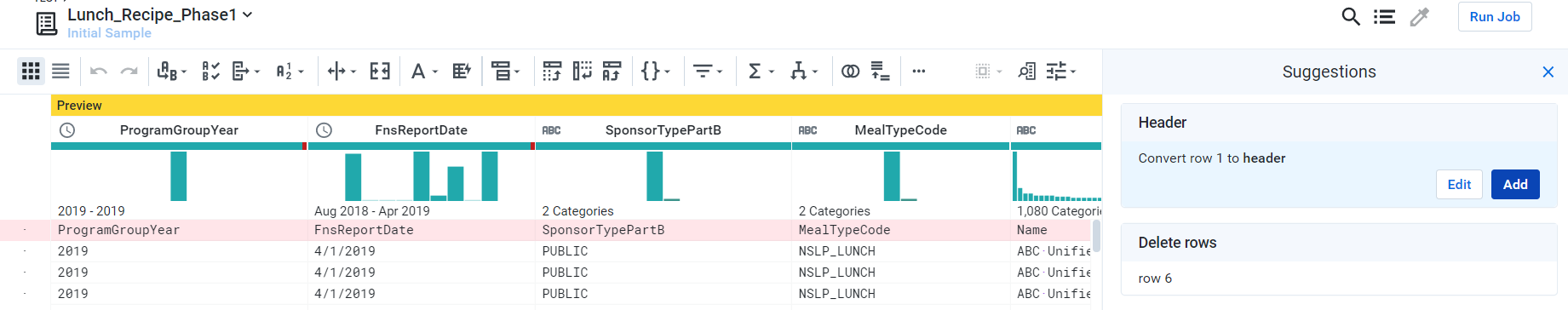
**Result:**

The function is added as a step under the “Recipe” tab.

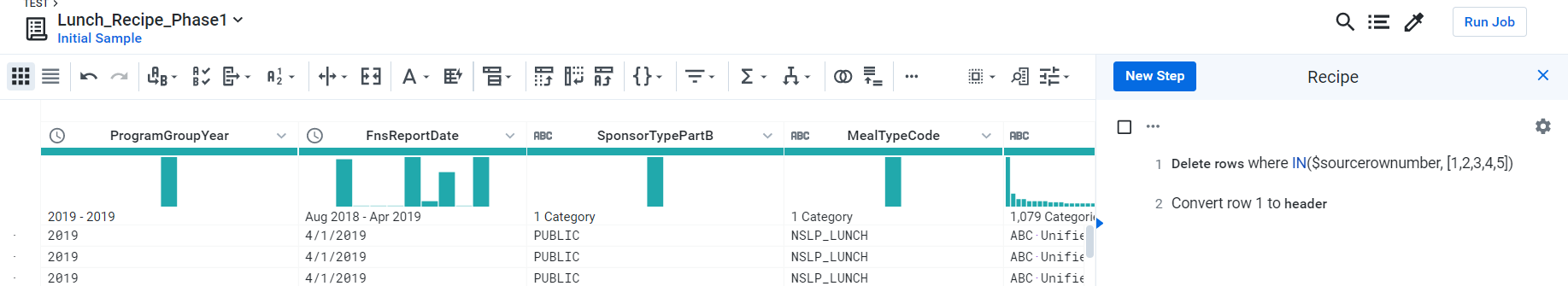


**Add Header:**

Select the row which is the header for the dataset. Trifacta tool suggest for “Add Header” function, click add to make the row as a header.

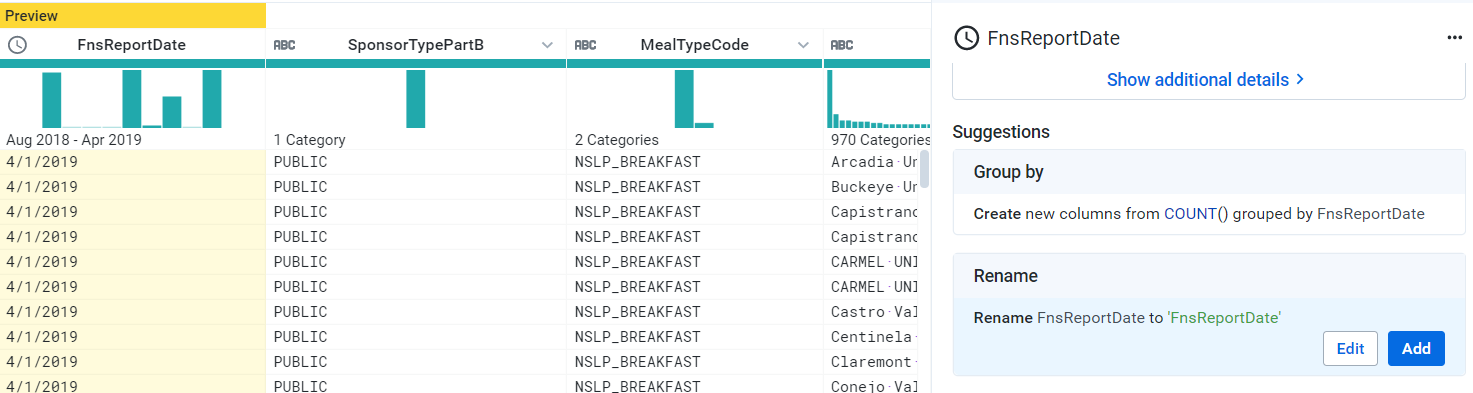


**Result:**

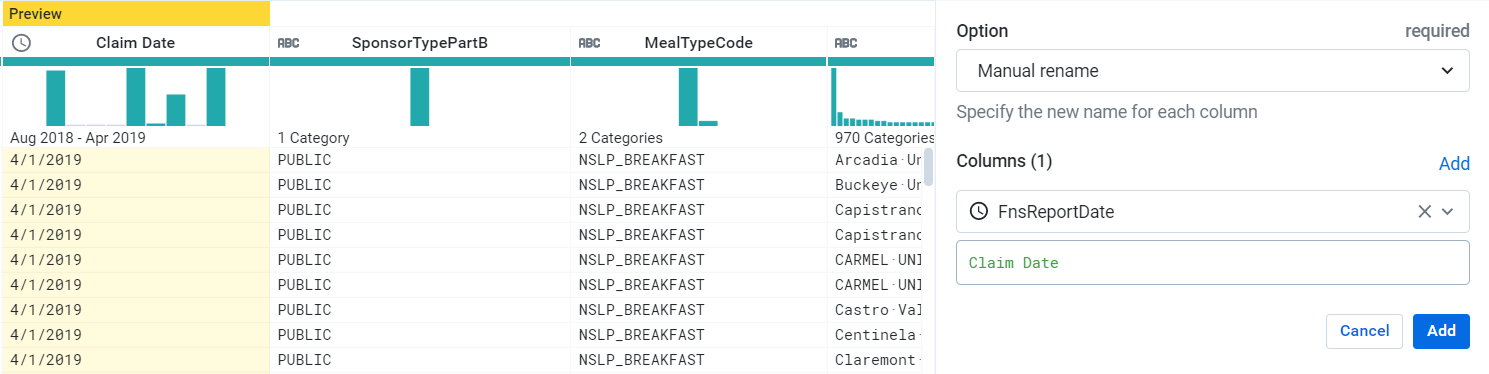


**Rename Column:**

Select the columns we want to rename. Click on rename function and select “Edit”. Provide the new name of the column in the space provided below under the old column name.

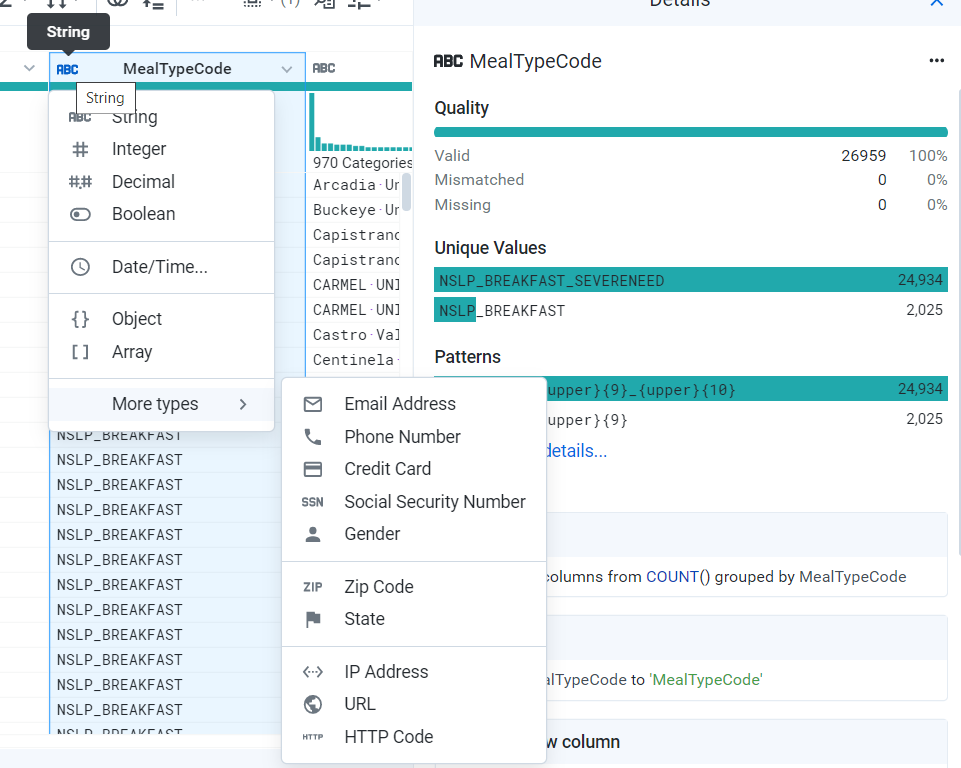


Result: Click “Add” to add the function to the recipe.



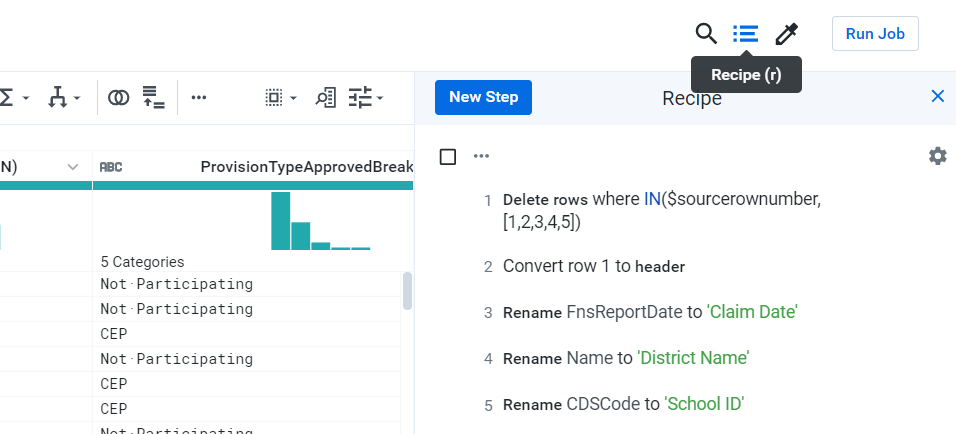
**Change Data Type in a column:**

In order to change the type of data in the column, we can select the column and click on top left corner of the column name as shown below. Trifacta provides suggestions in the drop down box which can selected to change the data type of the column.

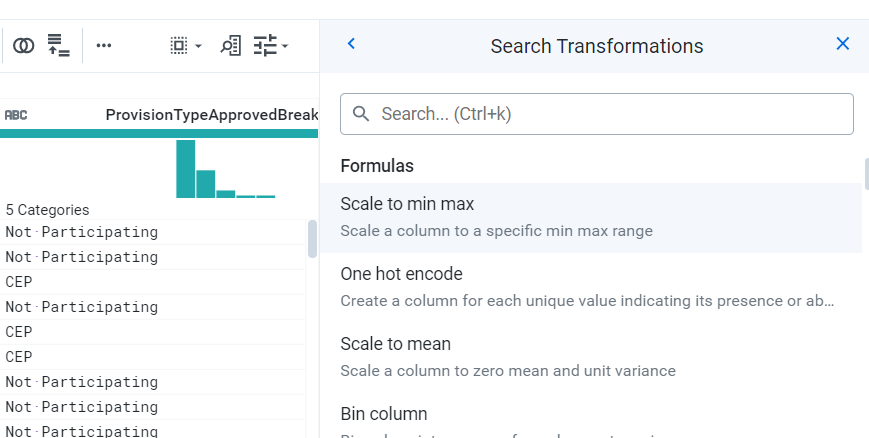


**Functions:**

To add a New Function in the recipe, select the recipe and then click on “New Step”



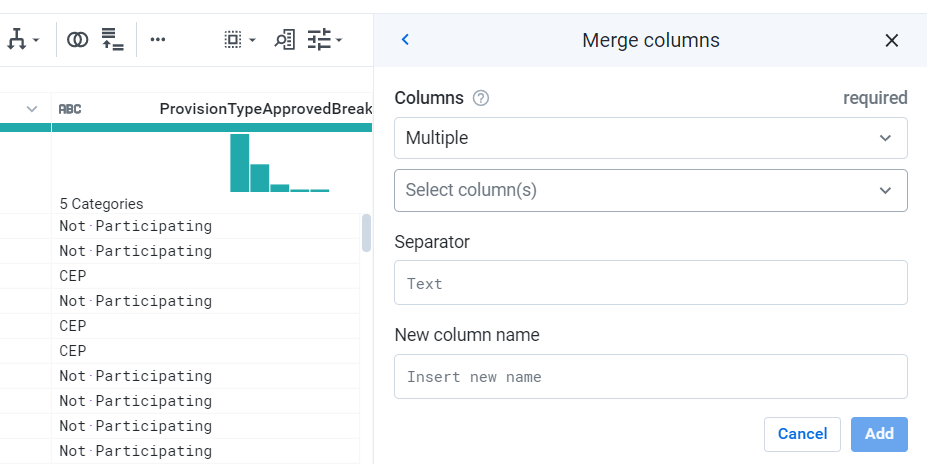
A search tool bar with some function listed below the search bar appears.



We can search for any formulas in the search bar and select it to add the function to the recipe. The functions come with syntax and we can select the columns on which the function is to be implemented. For Eg: **Merge Function**

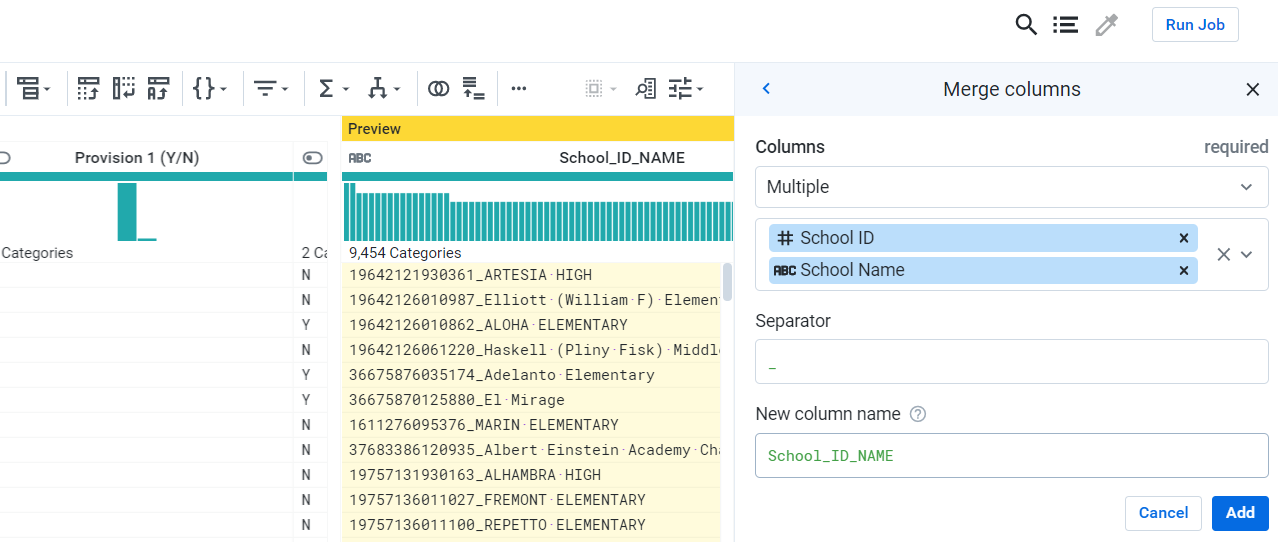
**Merge Function:**



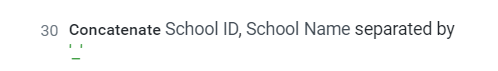
Select Merge Columns Function from the drop down suggestions. 

We can select the columns that are required to merged under the “Select column(s)” . It can be any number of columns but the order in which the columns are to be merged should be given systematically.

The “Seperator” represents the value you want to provide between the merges of each column. You can provide the name for the new column after perfoming the merge function under the “New Column Name” tab.

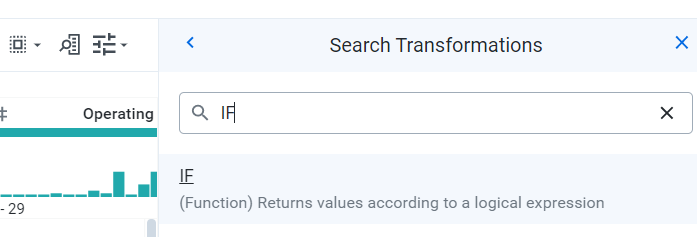


Once, the Preview of the New column is correct, click on Add to add the function to the existing recipe.

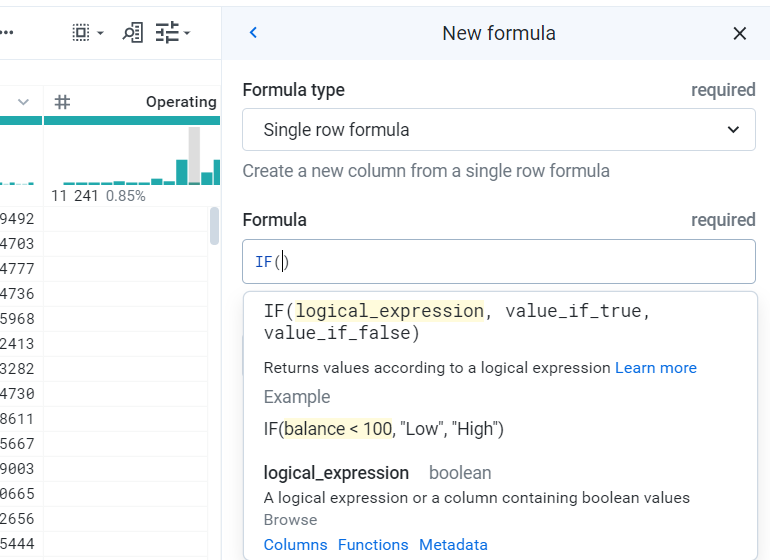


**IF Function:**

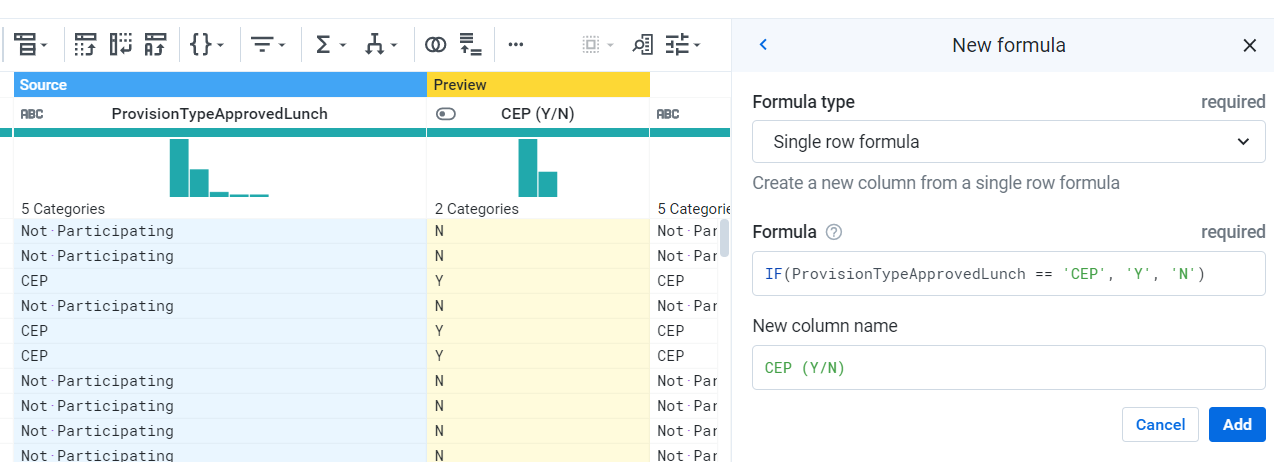
Select “New Step” and in the search bar type “IF” and click on the “IF” function.



You will be prompted with the syntax of the function. We can select the columns, functions from the bottom of the syntax.



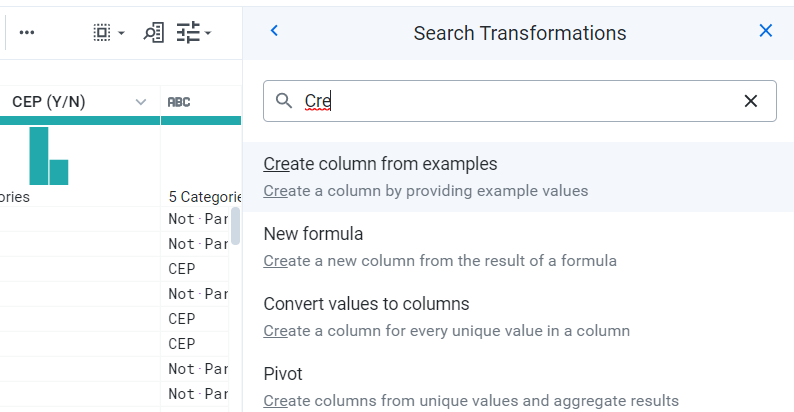
Enter the formula according to the syntax with the appropriate column and values. Once the results in the Preview are correct, then click on “Add” to add the function to the recipe.



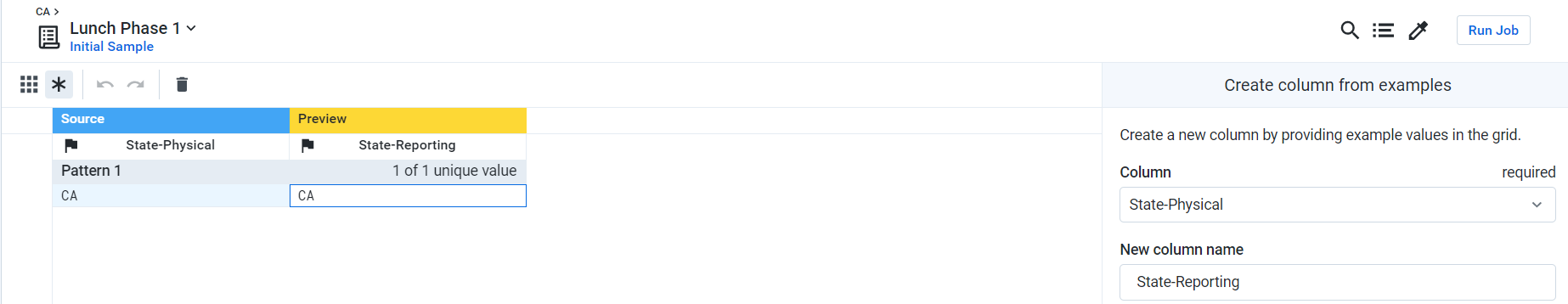
The function will be added to the existing list of functions in the recipe as a next step.

**CREATE COLUMN Function:**

This function can be used to create new columns based on the values from other columns.

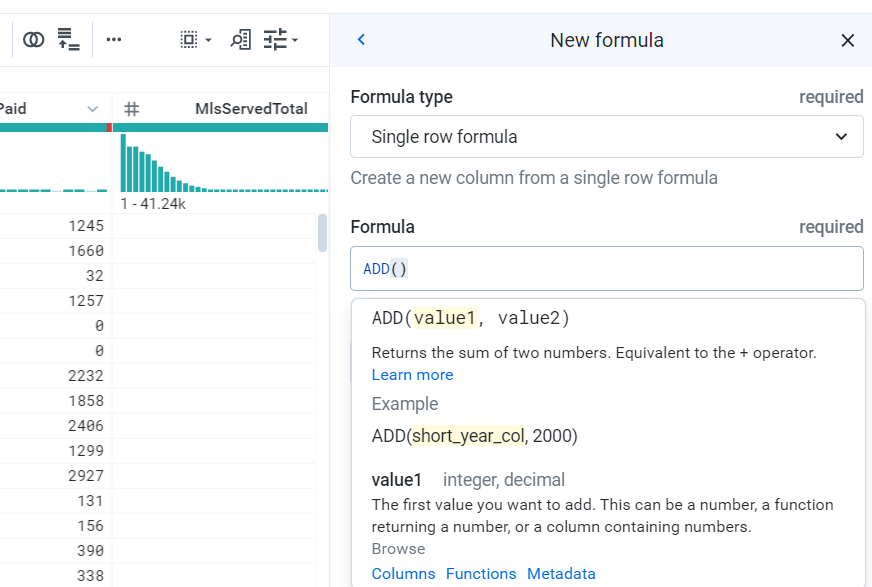


Provide the pattern of how the values in the new column should change based on the values present in the original column. The values can be provided under the Source and Preview tab is left corner. The click on “Add To Recipe” to add the function to the recipe.

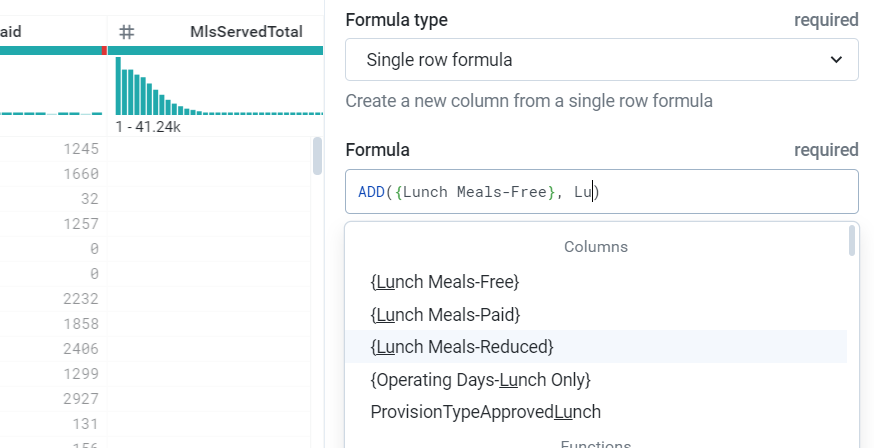


**ADD Functions:**

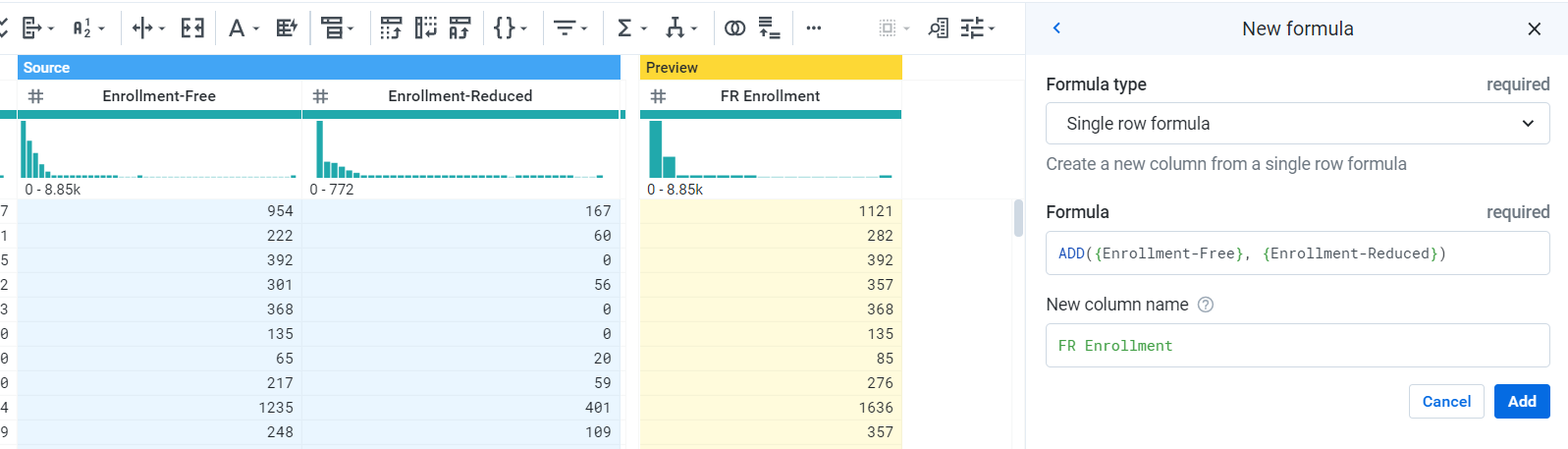
It sums the values present in the two or more selected columns.



**Based on the syntax, type in the column names separated by a ‘,’ sign.**



**Example For Add Function:**



We can notice that the source columns are highlighted in **blue** color and the preview column of the result is highlighted in **yellow** color. If the results are collect, click on “Add” to add the function to the recipe.

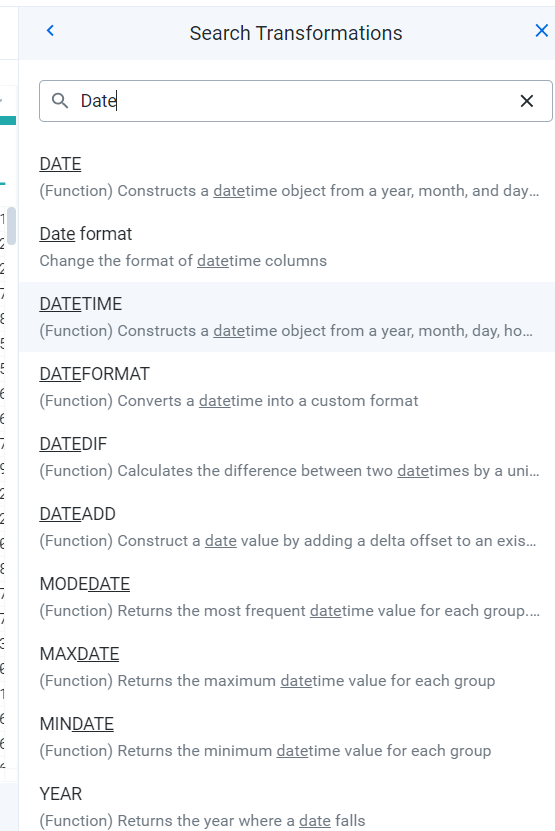
**Note:**

**Similarly, other Arithmetic function such as subtract, division, percentage, multiplication, min, max etc…can be implemented by following the predefined syntax for each function as per Trifacta guidelines.**

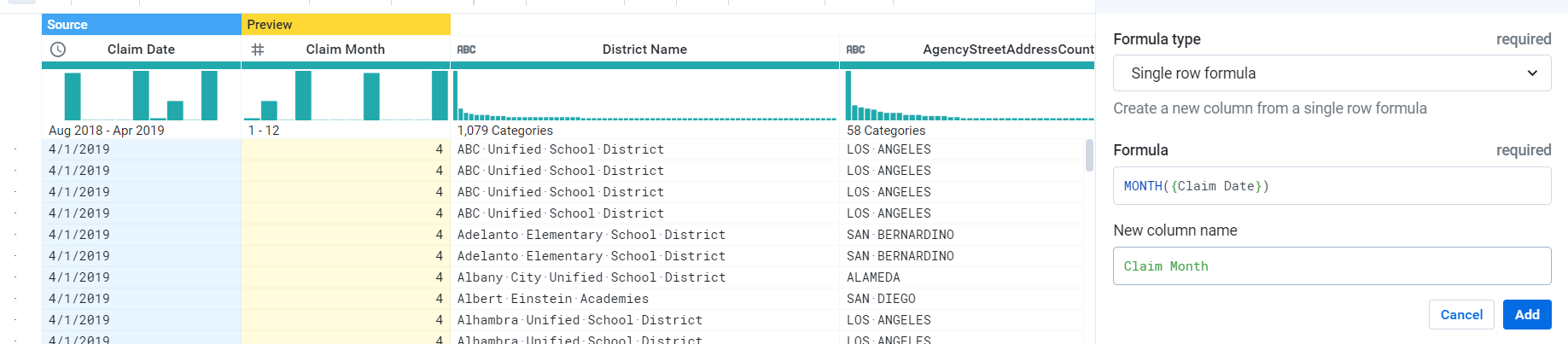
**DATE Functions:**

Date functions are inbuilt in the Trifacta tool, which can be used based on the information that is to be extracted from a date field or functions that can be implemented on a date field. Syntax is important for the date functions to be implemented.

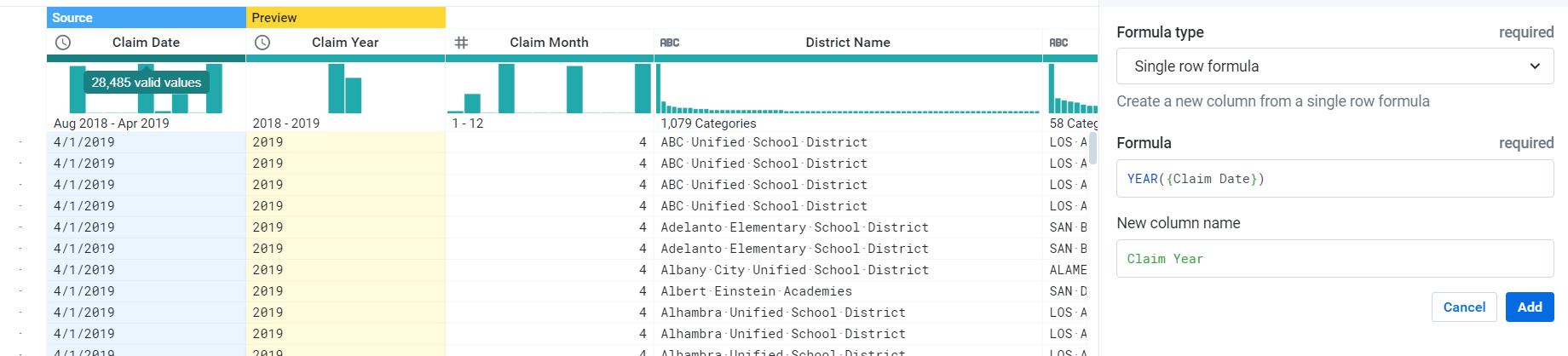
**Some of the functions are listed below:**



**Example of Date Functions: Month Function:**

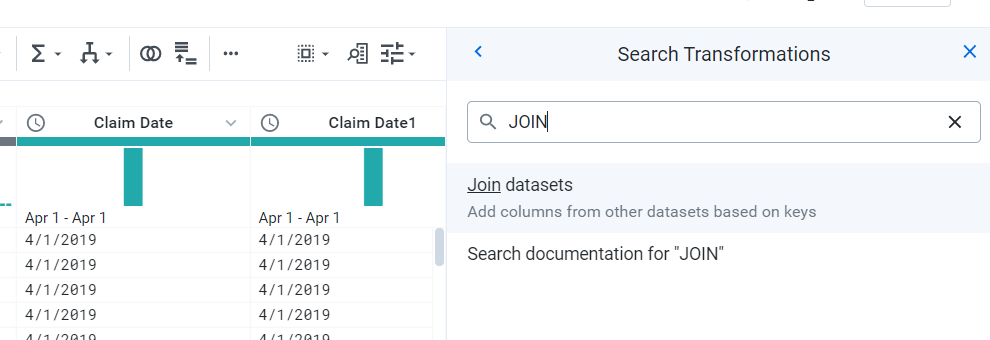


**Year Function:**



**JOIN OPERATION:**

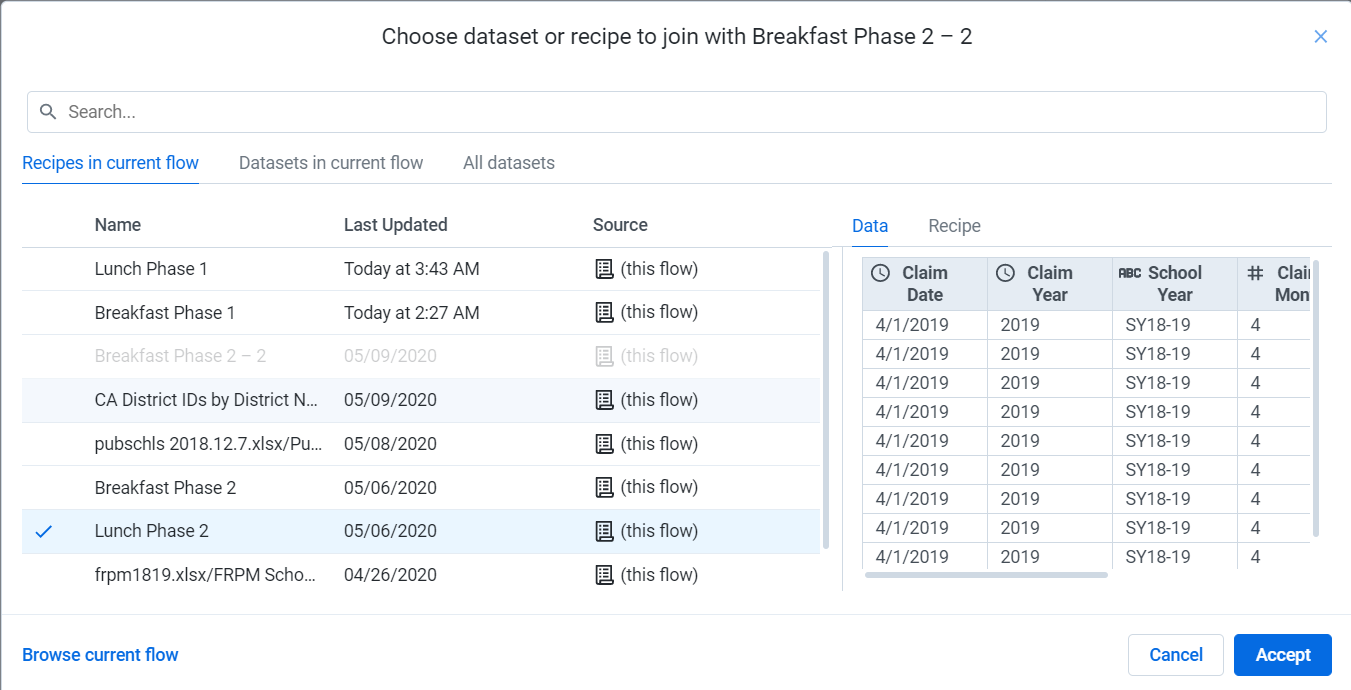
**Join function is used to combine the results from different recipes or combine datasets. Trifacta.**



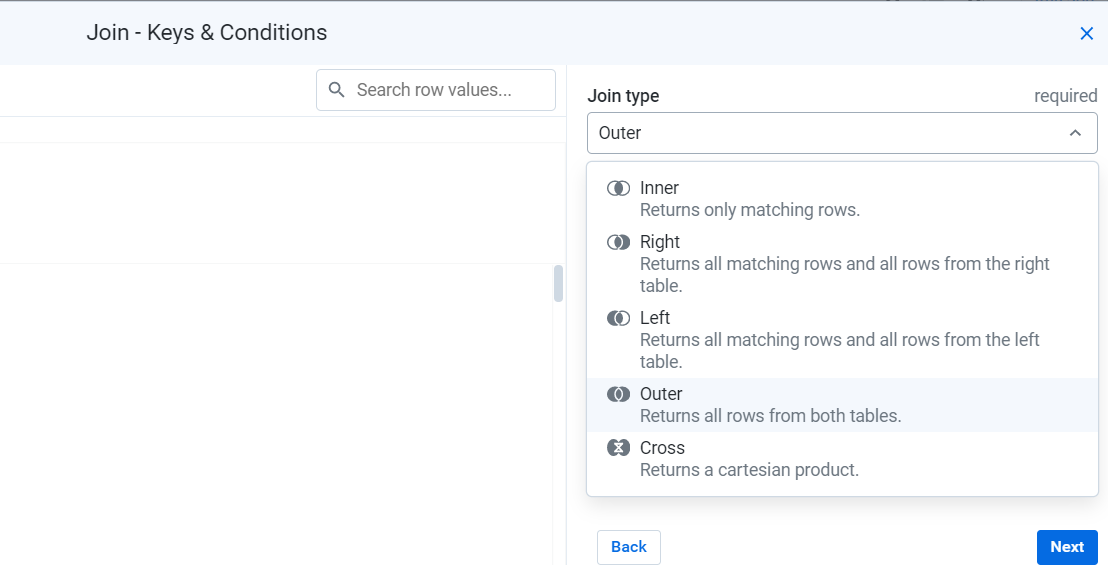
**Select the Join function and a new window opens for the Join function.**

**This window provides options to select datasets or recipes from the current flow or from the entire library uploaded to the Trifacta tool.**

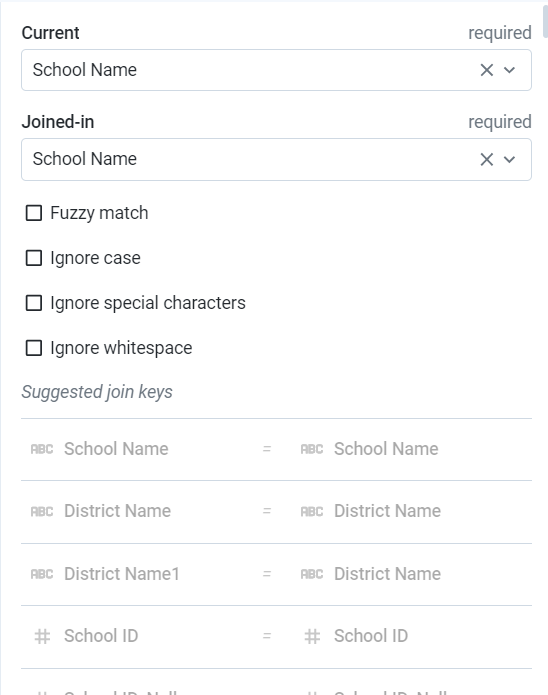
**Once the dataset is selected click on “Accept”**



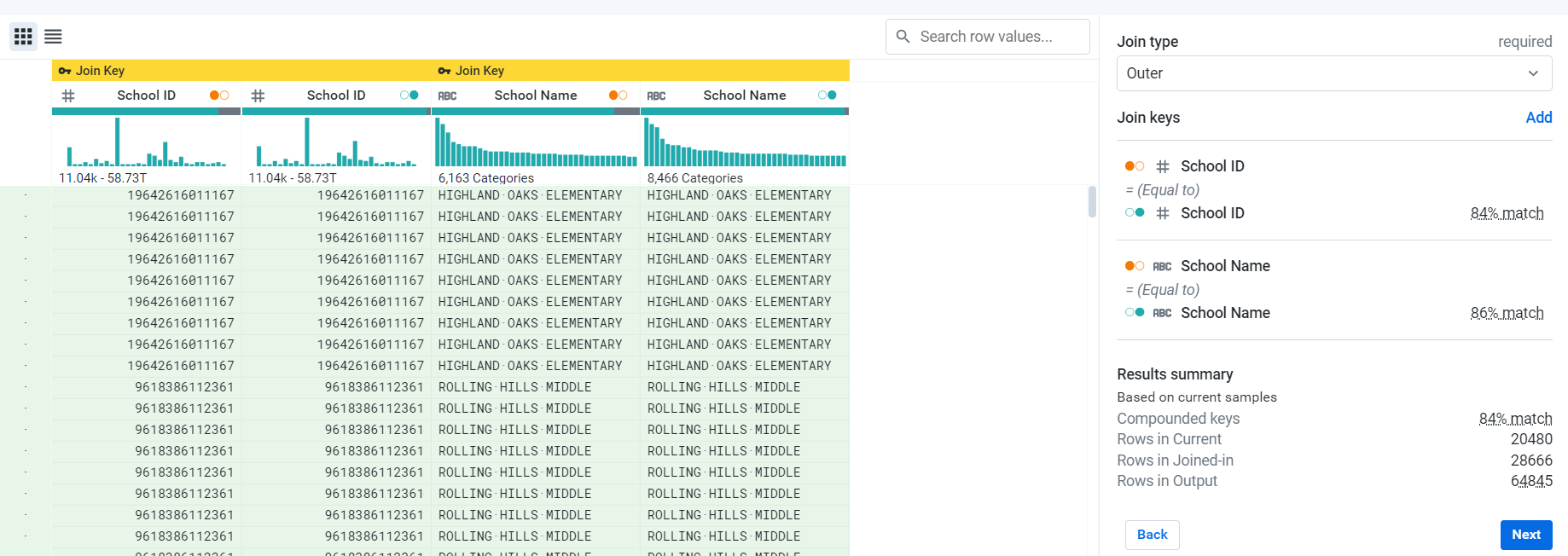
We can select the type of “Join” to be performed on the selected datasets. Then click “Next”



Once the type of join is selected, we can then select from the “Suggested Join Keys” or a combination of multiple fields acting as a unique key (Composite key). The limitations on how accurate or exact the join values should be accepted by the Join function also can be altered by selecting any one of the options “Fuzzy Match”/“Ignore Case”/Ignore Special Characters/Ignore Whitespace.

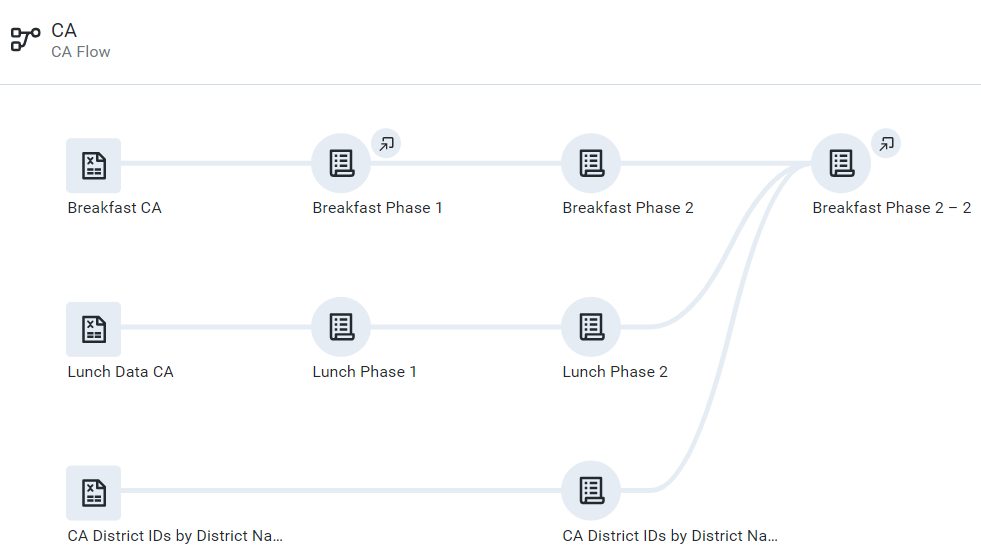


Once the Join function parameters are set and Join Keys are selected. A preview of the Joins are provided on the left corner of the Join window.



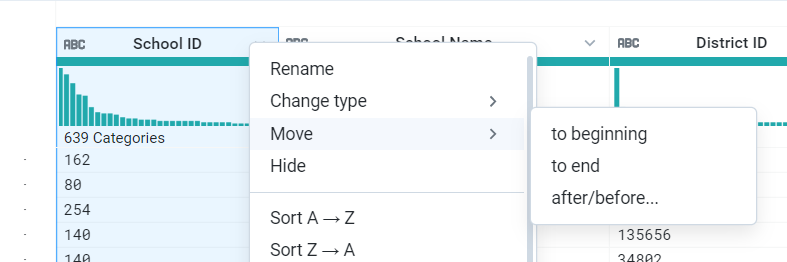
You can review the results by selecting the “Review” option and if the Join operations are correct, select “Add to recipe” to add the Join function to recipe.

If you check the workflow after the Join operation, there is a line connecting the datasets which are part of the Join operation.

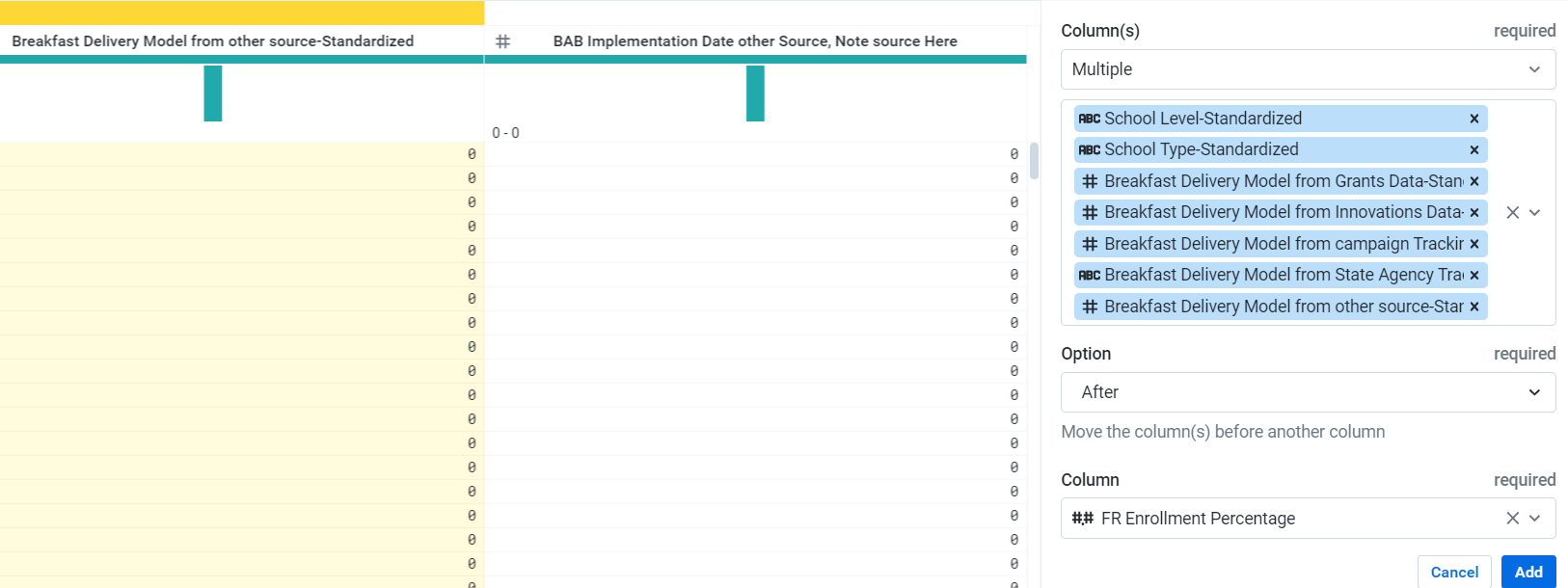


**Move Function:**

Move function can be used to alter the order of the columns present in the dataset. This can be used to get the exact same order of columns as present in the clean data file generated by Tableau Prep. Move function can also be implemented on multiple columns.



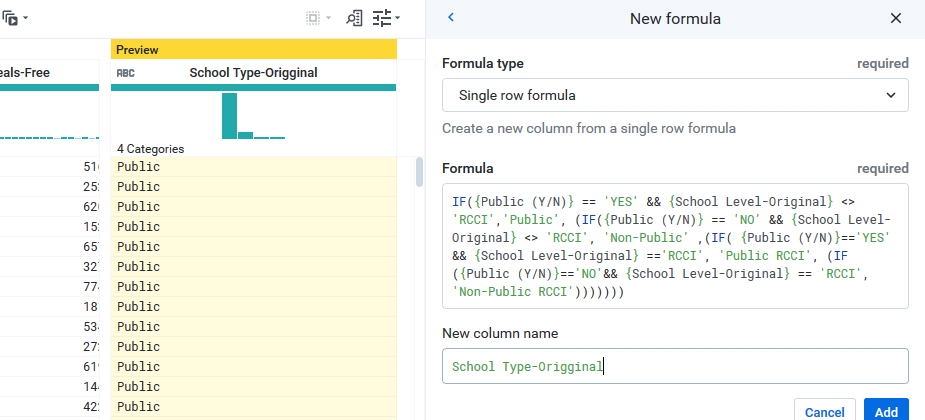
**Example: If the Preview results are correct, click on “Add” to add the step to recipe.**



**Nested IF Functions:**

Derive School Type-Original

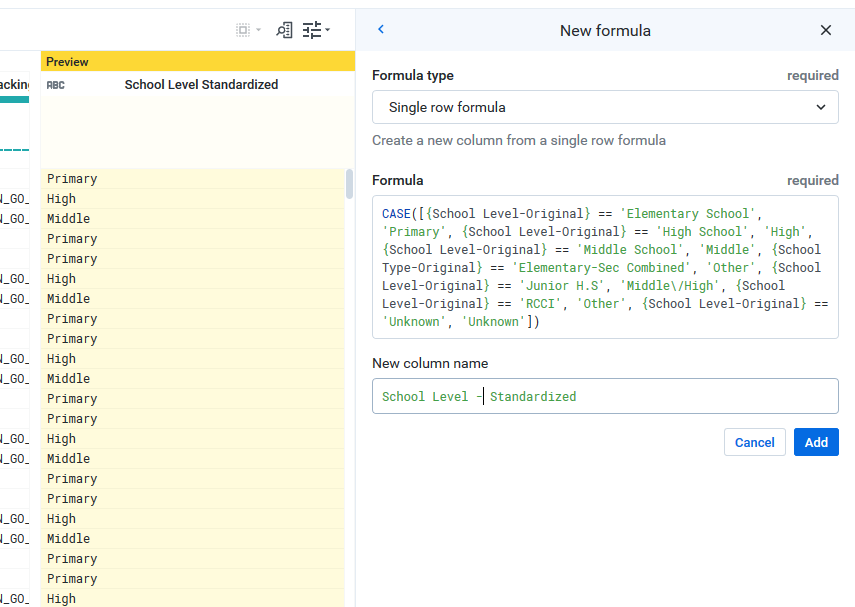
IF ({Public (Y/N)} == 'YES' && {School Type-Original} <> 'RCCI','Public', (IF({Public (Y/N)} == 'NO' && {School Type-Original} <> 'RCCI', 'Non-Public' ,(IF( {Public (Y/N)}=='YES' && {School Type-Original} =='RCCI', 'Public RCCI', (IF ({Public (Y/N)}=='NO'&& {School Type-Original} == 'RCCI', 'Non-Public RCCI')))))))



Click on “Add” to add the function to the recipe.

**Case Function:**

School Level Standardized



The CASE function allows you to perform multiple conditional tests on a set of expressions within a single statement. When a test evaluates to true, a corresponding output is generated. Outputs may be a literal or expression.

**Multiple IF Statements:**

Breakfast Delivery Model from State Agency Tracking-Original

IF(TRADITIONAL\_MODEL == 'Y', 'TRADITIONAL\_MODEL', '')

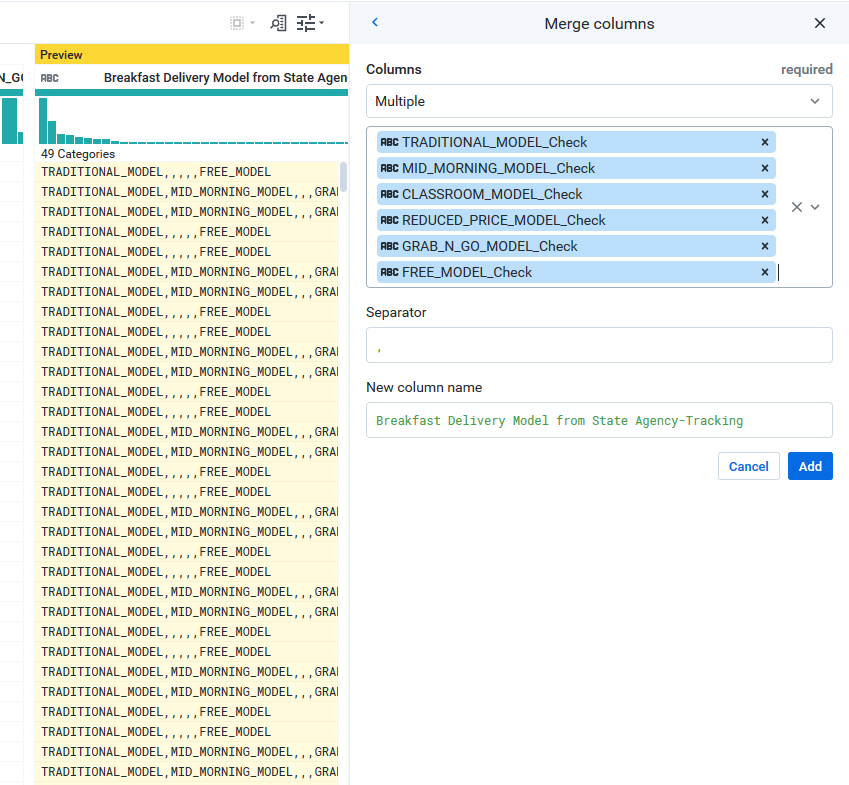
IF(MID\_MORNING\_MODEL == 'Y', 'MID\_MORNING\_MODEL', '')

IF(CLASSROOM\_MODEL == 'Y', 'CLASSROOM\_MODEL', '')

IF(REDUCED\_PRICE\_MODEL == 'Y', 'REDUCED\_PRICE\_MODEL', '')

IF(GRAB\_N\_GO\_MODEL == 'Y', 'GRAB\_N\_GO\_MODEL', '')

IF(FREE\_MODEL == 'Y', 'FREE\_MODEL', '')



**Note:**

The below YouTube channel is created for Trifacta Users, where the Trifacta community adds video, which can be useful in implementing different functions and understanding the tool.

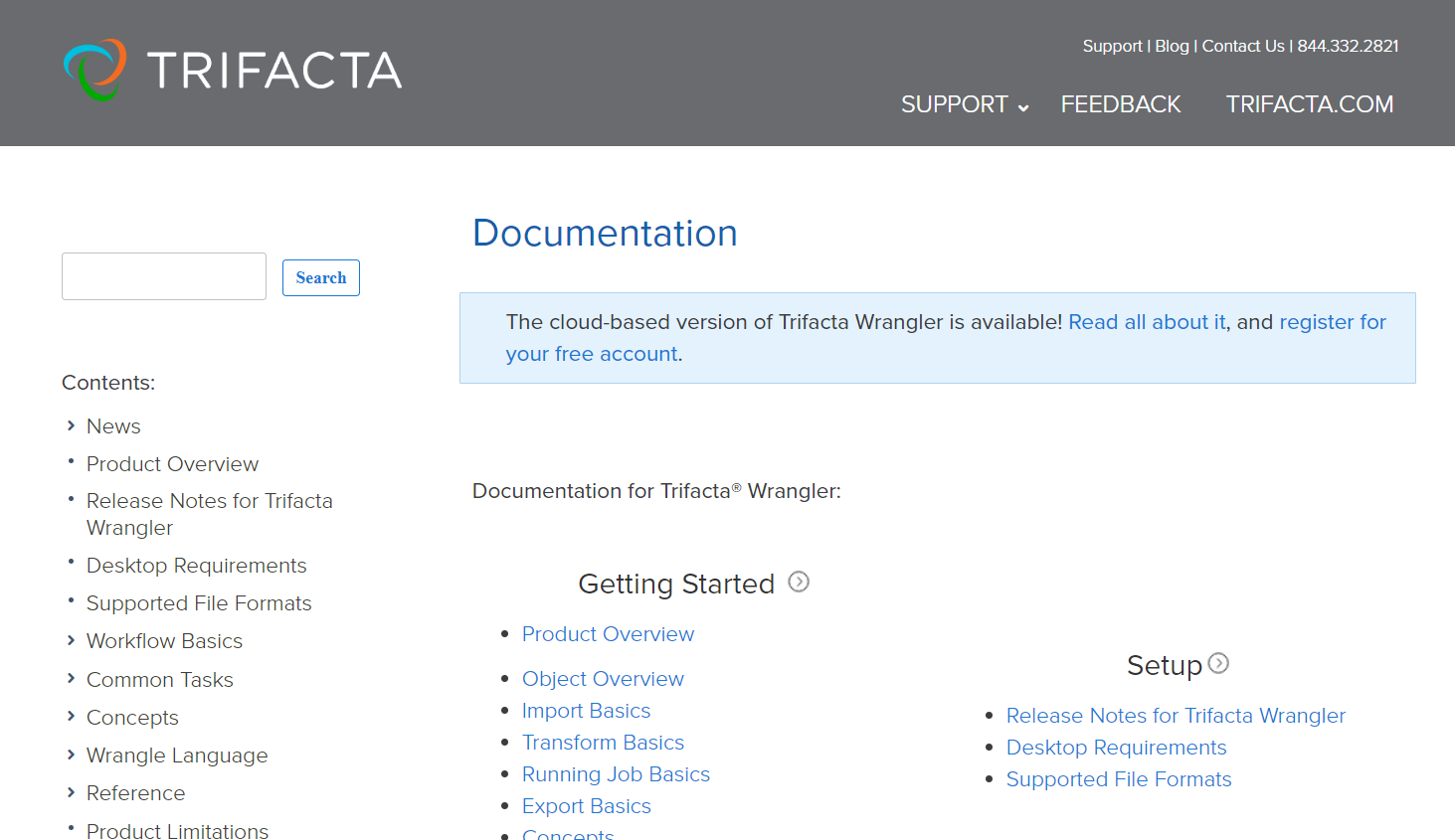
YouTube Channel Name: Trifacta

YouTube Channel Link:<https://www.youtube.com/user/TrifactaInc>

**Document for Functions:**

The below link provides the documentation for the functions in the Trifacta tool.

<https://docs.trifacta.com/display/SS/Documentation>



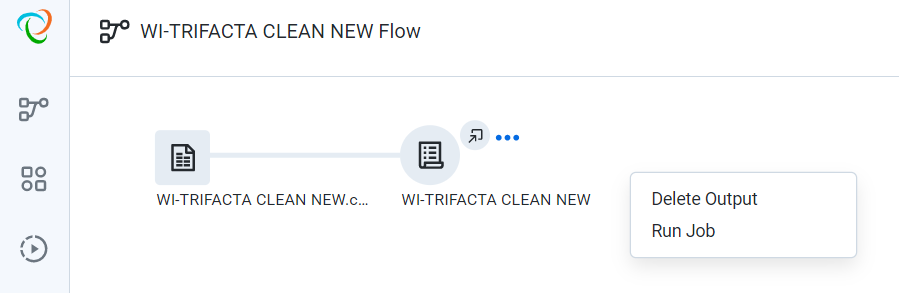
**5 Run Job:**

Once the recipe has been completed with all the functions to transform the imported dataset into the clean data excel workbook with standard 64 attributes.

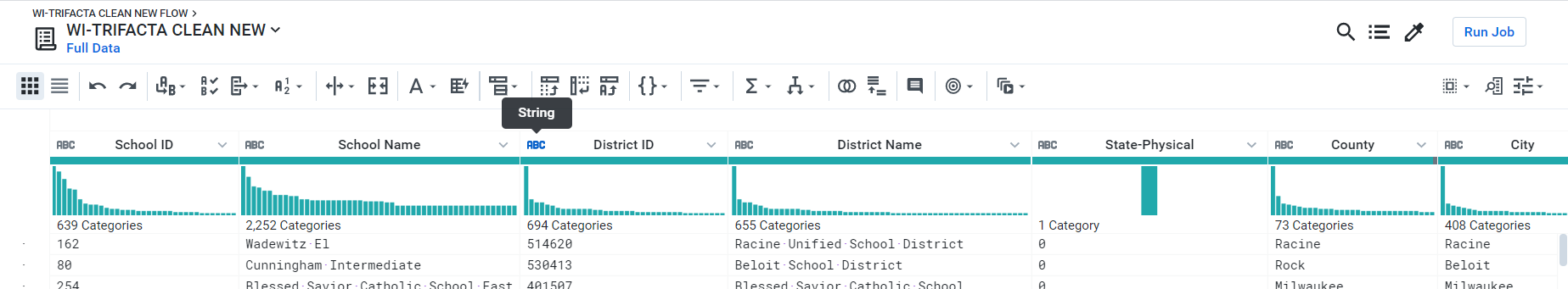
**Note:**

Jobs can be also executed at intermediate stages to check for the results generated from each recipe.

We can run the job in two ways, first one is by selecting the options near the final recipe in the Flow and clicking on “Run Job”



Another method, is to open the final recipe and click on “Run Job” in the top right corner of the recipe window.



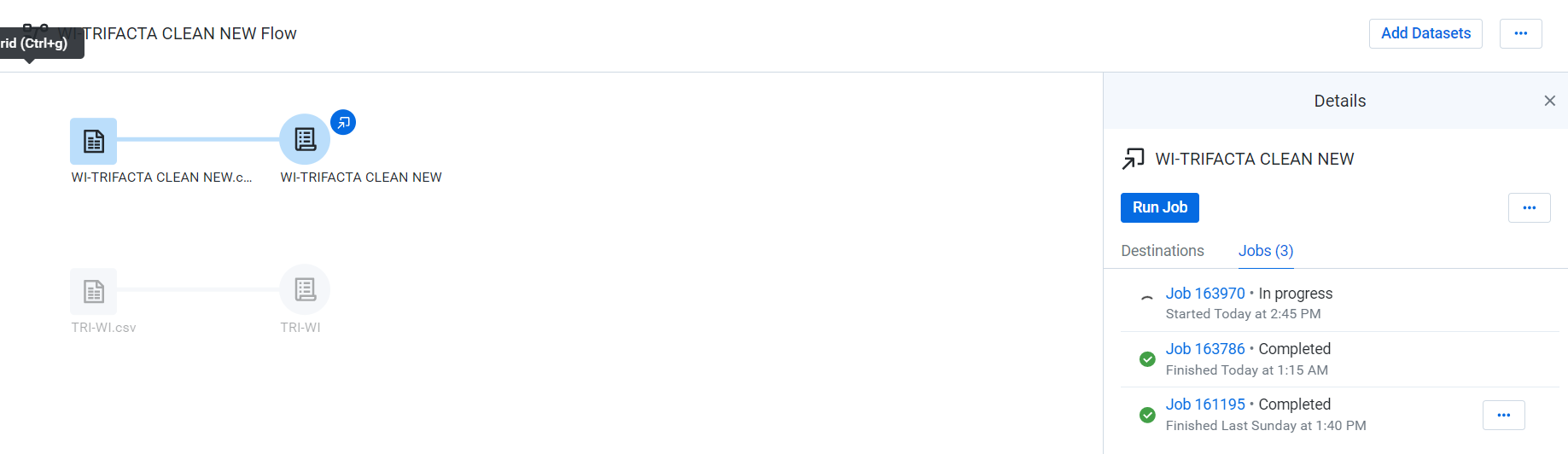
Once the “Run Job” is selected, a new Run Job window opens.

Using this window, we can also create JSON files from the recipe in order to work on different platforms.

We can also turn on or off the “Profile Results”. When this option is selected, we can get validations on the results that is generated in a pdf format.



Once “Run Job” is selected, it executes the recipe and we can notice the progress of the execution of the Job.

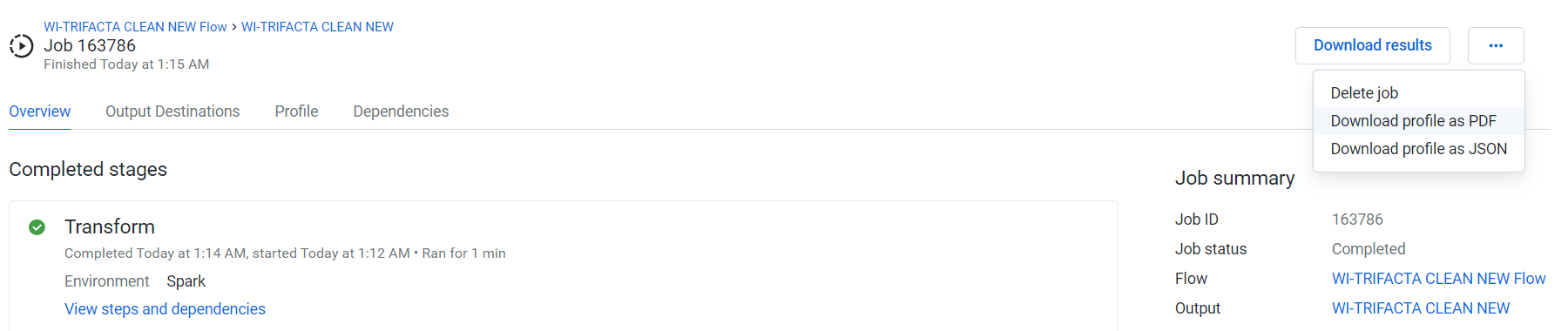


Click on the Job created after the completion of the execution of the Job and you are prompted with a new window.



Next step is to select “Download Results”, You will be able to get the clean data as a csv downloaded to local Personal Computer.

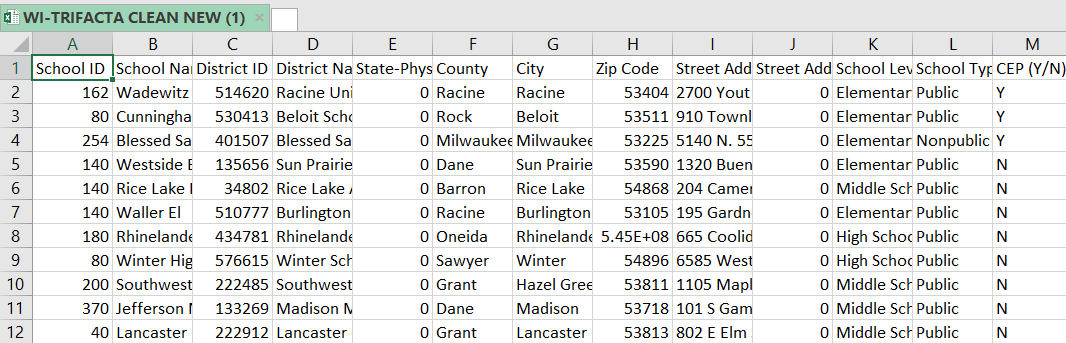
If require any validation provided by the Trifacta Tool. Select on options next to “Download Results” and then click on “Download Profile as PDF”.



A pdf file with the Job Name will be downloaded.

**6 OUTPUT:**

**Clean Data File Sample:**

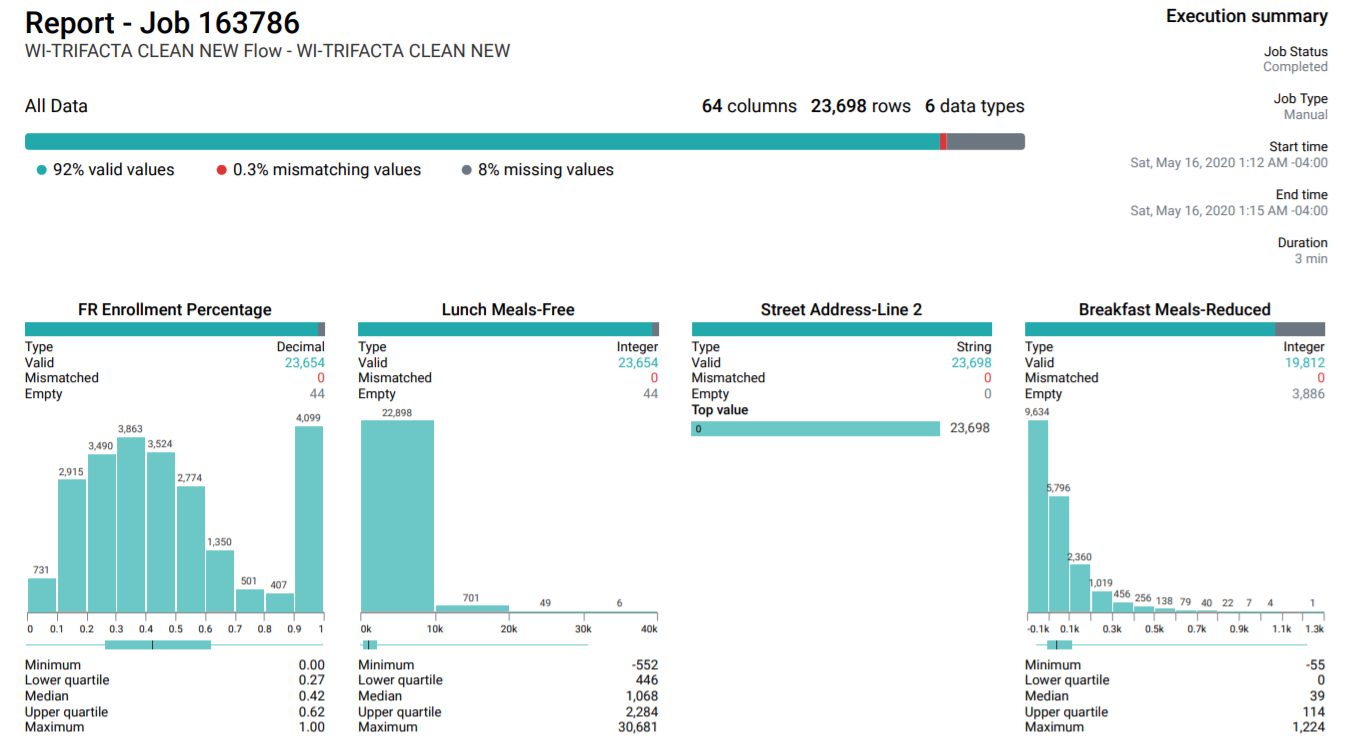


Clean Data file is in CSV format.

**Note:**

In case of any mismatches in the clean data file, changes can be made in the recipe at any point of time to alter the output.

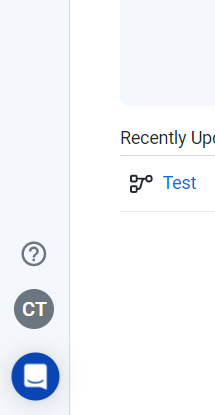
**Profile PDF File:**



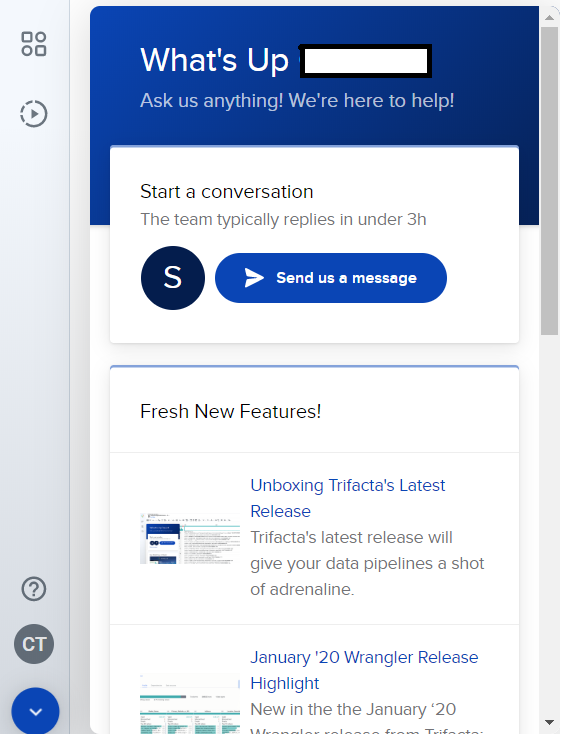
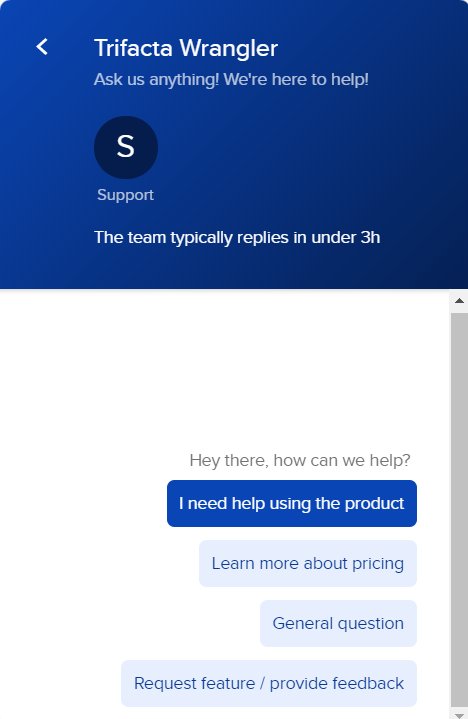
This pdf file can provide you valuable information regarding each column generated in the clean data file. Information can be regarding the datatype, valid values, mismatched values, null values and distribution of values in the column.

**7 Customer Support:**

Trifacta provides excellent customer support. You can access the customer support by clicking on the blue message icon in the left bottom of the Trifacta home window.

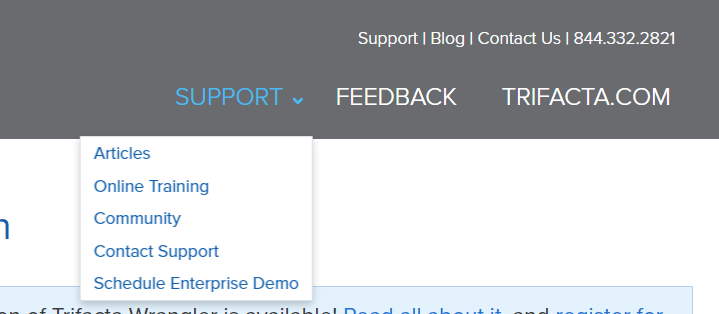


You can enquire about functions that you are trying to implement or regarding any other operations in the Trifacta tool. They respond within the given amount of time.



Other way of accessing customer support is:

Use the attached link: <https://docs.trifacta.com/display/SS/Documentation>



This provides access to articles on Trifacta, online training tools, Trifacta community and schedule a demo for the Enterprise Version of the Trifacta tool.