



Lesson 4 – Advanced Transforms

Chapter 4A | Header and Rows | 5 Minutes

Chapter Goals

In this Chapter, you will:

- Understand how to use the following transforms:
 - Row and Source Row
 - Header
 - Sourcerownumber

Chapter Instructions

YOUR GOAL: Use the header and sourcerownumber transformations to create a header row and delete specific rows from your data.

1.	<p>Add a new dataset to the Training Flow</p> <p>Click on the Training Flow.</p> <p>Click Add Datasets.</p> <p>Click Import Datasets.</p> <p>Drag and drop the <code>state_stats2.txt</code> file into Trifacta's import interface.</p> <p>Create a new dataset using the following information:</p> <ul style="list-style-type: none"> • Dataset Name: State Statistics • Dataset Description: Training Use Case <p>Click Import Datasets.</p> <p>Click on the Edit Recipe to open up the transformer.</p>	
2.	<p>Clean up the junk data</p> <p>The first thing you'll notice in the dataset is there are quote (") delimiters in every single column. Let's start by removing these from the entire dataset.</p> <p>With your mouse, brush over any of the quotes in the dataset. Choose the Replace suggestion card. You can see in the preview that the transformation is removing the quotes from only column2. Click the second option dot on the suggestion card. This card shows you that it will remove the quotes from all</p>	<pre>replace col: * with: '' on: `"` global: true</pre>



	<p>columns. Click the Modify button.</p> <p>Below the column box in the Transformation Builder, you can see the wildcard or asterisk (*) character. The wildcard will perform a transformation across all columns.</p> <p>Click Add to Recipe.</p>	
3..	<p>Row and Source Row</p> <p>Hover over the dot to the left of the first row of data. You can see that the Row and Source Row are 1. The Row indicates the position of the data on the Trifacta transformation page. The Source Row represents the row position of the data in the source or the imported dataset.</p> <p>To illustrate how this works, click on the dot for the first row. From the Suggestion Cards at the bottom, choose the Delete card and add it to your recipe. Now hover back over the first row of data. Notice how the Row is now 1 and the Sourcerownumber is 2. If we were to look at the source data, this line would be row 2 in that data sample. However, on the Trifacta transformer page, it is now in the number 1 position. This goes back to Lesson 1 where we discussed that Trifacta never modifies the source data.</p> <p>Click Undo to remove the last step from the Recipe</p>	<pre>delete row: sourcerownumber() == 1</pre>
4.	<p>Promote the header row.</p> <p>Click on the first row of data again.</p> <p>This time, from the suggestion cards, select the transform that says Convert row 1 to header. This will take the values of the row and promote them to the column headers.</p> <p>Click Add to Recipe.</p>	<pre>header</pre>
5.	<p>Promote a new header row using the sourcerownumber parameter.</p> <p>You can promote any row to the header using the sourcerownumber parameter. Clicking on the row number will offer suggestions of delete, keep, and header.</p>	



	<p>The following transforms might make the original row information invalid or unavailable: aggregate, flatten, join, lookup, union, unnest, and unpivot. After these transformations you may not be able to generate suggestion cards for row transformations.</p> <p>Click on Row 4 Source Row 5. Scroll to find the Header suggestion card and click modify.</p> <p>In the transform Builder, you can see that you can choose any Row number to promote to the header, by clicking through the arrows or typing in the number.</p> <p>Click Cancel on the transform Builder.</p>	
<p>6.</p>	<p>Delete a row using the sourcerownumber parameter.</p> <p>We're going to delete Source Rows 10 through 20. There are a couple of different ways that you can accomplish this.</p> <p>The first way is to delete rows that are contained in a list. Click on Source Row number 10. The first card that appears suggests to Delete rows. Now, click on each Source Row until you get to Source Row 20. Each time you select a row it will be highlighted in red. You'll see the Delete suggestion card has generated a list of 10 through 20. Click Cancel.</p> <p>You can also delete all rows less than or equal to a particular value or a range of values.</p> <p>Click on Source Row number 10. Select the Delete rows card and click Modify.</p> <p>In the Builder, you see that the transformation is going to delete the sourcerownumber equal to 10. Click on the equal signs and this will generate a list of symbols. Click on the greater than or equals to symbol. Add two ampersands (&&), which represents AND. Continue to include another Sourrownumber condition and this time set it to less than or equal to 20.</p> <p>The syntax for this statement reads: Delete Source Rows greater than or equal to 10 and less than or equal to 20:</p>	<pre>delete row: sourcerownumber() >= 10 && sourcerownumber() <= 20</pre>



	<pre>delete row: sourcerownumber() >= 10 && sourcerownumber() <= 20</pre> <p>It's no more difficult than your middle school math class.</p> <p>Click Add to Recipe.</p> <p>That's all for Chapter 4A. In the next chapter we will continue to learn more advanced transformations.</p>	
--	---	--