



Chapter 2–Basic Operations

Lesson B | String Operations | 10 Minutes

Lab Goals

In this Lesson, you will:

- Learn how to use the following Transformations:
 - o Set
 - o Replace
 - o Extract
 - o Countpattern
 - o Split
- Learn how to apply certain Transformations to multiple columns

Lesson Instructions

1.	Changing Case within a Column Click on the column header on the Spring_Day_A column. Select the suggestion to Set the column Spring_Day_A to Uppercase . Select Edit . In the Formula parameter, change upper to proper . Note, the \$col indicator applies the proper function to all of the columns selected (in this case just the Spring_Day_A column). Select Add to add to your recipe. You should notice that Spring_Day_A's histogram went from 14 categories to 7, indicating there are no longer issues with case.	set col: Spring_Day_A value: proper(\$col)
2.	Multi-Column Set Next, Click Columns to enter the columns view. Shift + Click from Spring_Feq_A to Fall_Freq_B . Select Action → Format → To Propercase . You should notice this initiates a builder step with all of the selected columns appearing in the Columns parameter and proper(\$col) in the Formula parameter. Click Add .	set col: Spring_Freq_A, Spring_Day_B, Spring_Freq_B, Summer_Day_A, Summer_Freq_A, Summer_Day_B, Summer_Freq_B, Fall_Day_A, Fall_Freq_A, Fall_Day_B, Fall_Freq_B value: proper(\$col)



3.	Cleaning Irregularities While still in the Columns View , click on the Items column, select Action , then Show in Grid . This brings you to the Items column in the Grid . Highlight over the <code>\$</code> , and select the suggestion to Replace this value. Notice in the preview to the right that the resulting column has been cleaned up. Click Edit to view the transform Builder. Notice that you can apply the Replace transform to multiple columns at once by entering more columns in the boxes below the Column parameter. To apply this transform to all columns we can use the wildcard or asterisk (*) character in the column parameter. This will apply the transform to all of the columns in the dataset. <div data-bbox="313 808 886 1050"><p>The Replace transform finds occurrences of text that match a defined pattern and replaces that text with a specified value. By default, Trifacta replaces the text with an empty string, which will delete the text from your data.</p></div> Let's apply the transformation to only the Items column. So, our transformation step reads as: replace the pattern of '\$ ' with the new value of nothing (we leave the values box blank) on Column Items. Add to your recipe.	<pre>replace col: Items with: " on: '\$ ' global: true</pre>
3.	Select a string to generate a pattern. Brush over the following text in the Items column: wine_alcohol <div data-bbox="313 1493 886 1671"><p>Whenever you directly select data in the transformer grid, Trifacta generates suggestions that apply to occurrences of text that match the same pattern as the selection.</p></div> Notice that the first suggestion card shows the Extract transform. This transform finds occurrences of text that match a defined pattern and places that text into one or more new columns.	



4.	<p>Determine pattern-matching methods for Extract.</p> <p>Look at the first suggestion card. This card uses your selection to define a pattern using Trifacta selection rules. Trifacta selection rules operate like java- style regular expressions, but are designed to be easy to understand.</p> <p>This selection pattern matches exactly 12 alphanumeric characters and underscore characters.</p> <p>Mouse over the different Extract suggestion cards. You can click on each suggestion to update the preview in the transformer grid and display the result of that suggestion.</p> <p>Based on the options displayed on the suggestion card, how can you define patterns in the Extract transform?</p> <ol style="list-style-type: none">1. Trifacta selection patterns.2. Exact string matches.3. By identifying the values that come before and after the pattern that you want. <p>Click on the first Extract suggestion card and click Edit to see the transform in the Builder.</p> <p>If you click on the pattern, you will see the description of this Trifacta pattern. If you highlight the Trifacta pattern, you see a list of all possible pattern options with readable names and descriptions.</p> <p>Note that the Extract transform does not modify the original column.</p>	
5.	<p>Select multiple strings to refine a pattern.</p> <p>You can make multiple text selections in the same column to refine the patterns suggested by Trifacta. When you do this, you teach Trifacta to identify the correct pattern in the column. Note that to refine an existing suggestion, you need to select text that occurs in a different row from your first selection.</p> <p>In the first row, we've brushed over the text: <code>wine_alcohol</code>. In the second row, brush over the text: <code>meat_eggs_seafood</code>. You can expand the column by clicking on the arrow to the right. You can drag the column edge to expand it as well.</p>	<p>Extract col: Items on: `{alphanumeric-underscore}+` limit: 3 before: ``</p>



Notice that the pattern shown on the Extract suggestion card has changed. Instead of reading `{alphanum-underscore}{12}`, the card now reads `{alphanum-underscore}+`.

Since you have selected multiple values in the Items column, Trifacta has generated a more general pattern. `{alphanum-underscore}+` means that Trifacta will match one or more alphanumeric or underscore characters, while `{alphanum-underscore}{12}` will only match exactly 12 alphanumeric or underscore characters.

If you look at the preview in the Transformer Grid, you should also notice that Trifacta will now generate three columns. The Items column in this dataset is a list of item categories that the farmers markets could sell. If there is 'N' after the colon, this market does not sell these items. If there is a 'Y' after the colon, this market does sell these items.

For the purpose of this example, let's assume we only want to extract the item categories.

Click on the second Extract suggestion card. You'll notice that this suggestion adds to the pattern to extract alphanumeric or underscore characters before the colon (':'). In the preview, you will see that Trifacta will now generate two columns.

Click **Edit**.

The **limit** parameter determines the number of times Trifacta will match a pattern, and thus the number of columns that the Extract transform will generate.

Change the number in the limit parameter from 2 to 3 and watch the preview update.

Add to Recipe.



6.	<p>Use the Countpattern transform to count the number of times a delimiter appears</p> <div><p>The Countpattern transform counts the number of times a specified pattern occurs in each record of a column.</p></div> <p>Select the pipe character ' ' in the Items column. Scroll through the suggestion cards until you find the Count values matching suggestion. This uses the Count Pattern transform.</p> <p>Select the Countpattern transform and examine the preview in the grid. You can see that the Countpattern transform creates a new column that contains the number of times the selected pattern appears in each record.</p> <p>Add the Countpattern transform to the recipe</p>	<pre>countpattern col: Items on: ' '</pre>
7.	<p>Generate a Split transform.</p> <p>In the Items column, brush over the pipe character: </p> <p>When you select a delimiter—separator values like commas, pipes, or spaces— Trifacta will suggest a Split transform first. Click on the Split suggestion card and examine the preview.</p> <p>Notice that the Split transform will use the selected delimiter pattern to divide the current column into multiple new columns. The delimiter itself does not appear in the new columns, and the original column is dropped from the dataset.</p>	
8.	<p>Determine pattern-matching methods for Split.</p> <p>Mouse over the different Split transform suggestions to see multiple versions of the suggestion. You can click on each suggestion to update the preview in the transformer grid to display the result of that suggestion. Click on the first Split suggestion and click Edit.</p>	
9.	<p>Control the number of columns generated by the Split transform.</p> <p>Examine the transform in the Builder. Notice that the limit parameter controls the number of columns that will be created. Change the limit parameter value</p>	<pre>split col: Items on: ' ' limit: 5</pre>



	from 4 to 3. How does this change the preview? Add the transform to the recipe.	
10.	Undo the Split transform. Click the undo button at the top of the screen to remove the Split transform from your recipe. 	