

Create Column by Example

You can create a new column of data from an existing one by providing example values for the new column for values in the source column. With each successive example value, Transformation by Example (TBE) improves the quality of the output values, until you have the desired set of values for your newly generated column.

Limitations:

- Transformation by Example works best for text-based inputs. Non-text inputs are treated as String type by the feature.

NOTE: Multi-value inputs, such as Object or Array data types, must be converted to String data type prior to transformation by example.

- In the Transformer page, TBE is applied across the currently displayed sample. In the entire dataset, there may be outlier values that do not match any of the examples that you have provided.

Tip: If your column data is quite varied, you should collect additional samples to verify that your TBE is properly matching all values in the column.

For more information, see *Overview of TBE*.

Steps:

- In the Transformer page, locate the column to use as your source. From the column menu, select **Create column from examples**.
- In the Transform Builder, enter the new column name.
- In the following example, a new column called `zip` is being created from the `Addresses` column:

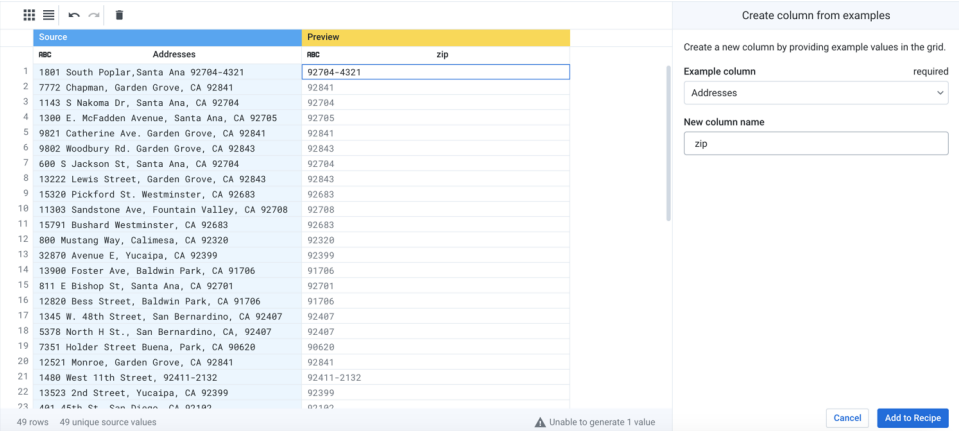


Figure: Selected column and first value is specified

- Double-click an empty cell in the Preview column to populate it with an example. In the above, the zip code from the first value has been entered into the Preview column: 92704-4321.

Tip: You can copy values from the source column and paste them into the Preview column.

- While many of the zip code values from other rows have been accurately populated, there are still some values that need fixing. In the following, you can see that one zip code was not properly extracted. Double-click in the Preview column for the third row and fix the value: 91935:

The screenshot shows a data transformation tool interface. On the left, a table with columns 'Source', 'Address', '#', and 'Zip' is displayed. The third row has a zip code that needs fixing. On the right, a 'Create column from examples' dialog is open, showing a grid with 'Address' and 'Zip' columns. The 'Zip' column has a value of '91935' entered.

Figure: Populating multiple example rows improves the overall quality of transformation across all rows

- A quick scroll through the rest of the rows in the sample indicate that you have properly extracted the zip code values for all rows.
- Click **Add to Recipe**.
- The new zip column is added to the dataset.

The screenshot shows a data transformation tool interface with a transformed dataset. The table has columns 'RBC', 'column2', 'RBC', 'column3', 'RBC', 'column4', 'RBC', 'Address', '#', 'Zip', '#', and 'column6'. The 'Zip' column is populated with values like 92704, 92841, 92704, etc.

Figure: Transformed example column

For more information on previewing changes, see *Transform Preview*.