Standardize Using Patterns

Contents:

This section describes techniques to standardize values in your datasets using patterns. From the Column Details panel in the Trifacta® application, you can review and select patterns in the column's data. These selections can be used as the basis for converting all applicable values to the selected format.

> **NOTE:** Pattern-based conversions can be applied to any data type.

In the Patterns tab, click the whitespace around a pattern and then review the Convert suggestion to define how the pattern matches can be converted to a single standardized format.

> **Tip:** To select, click the whitespace around the pattern and example values.

> **NOTE:** The application does not suggest pattern-based conversions that add or remove alphanumeric characters.

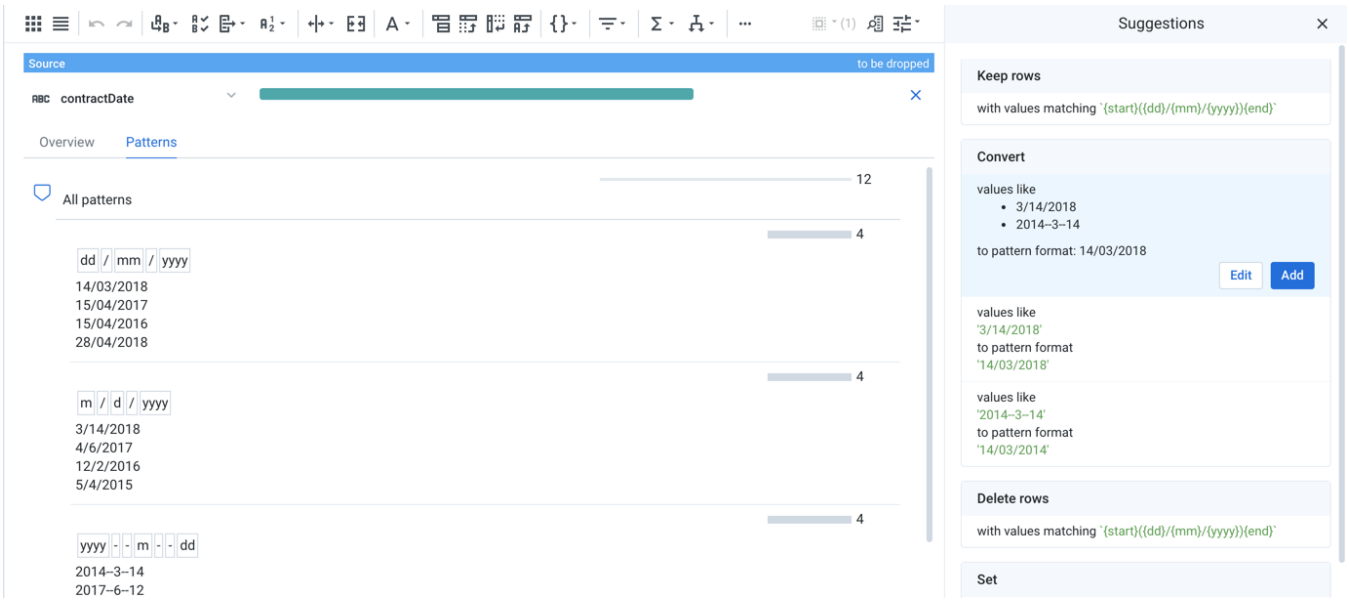


Figure: Selecting Datetime patterns in the Patterns tab

In the above, the pattern block prompts suggestions for Convert tasks based on the selected patterns.

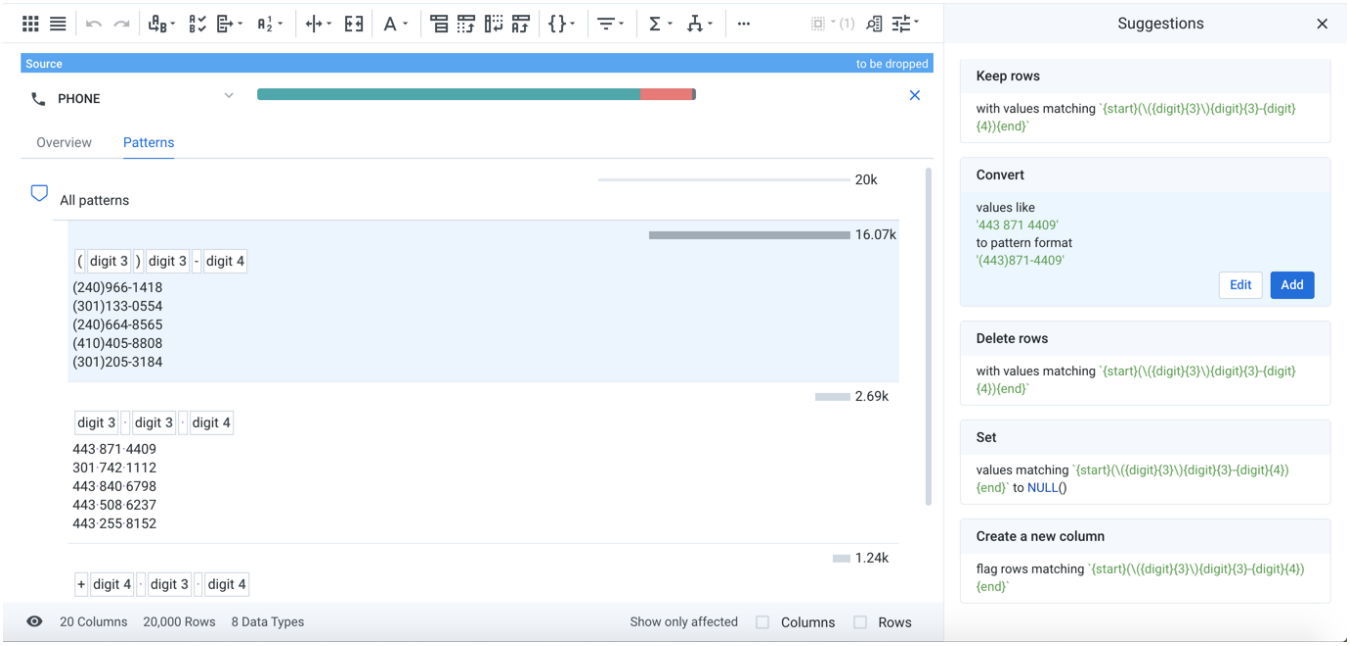
- Click **Edit** to modify the task.
- Click **Add** to add the task as a step to your recipe.

Example - Phone number patterns

For columns containing phone number data, you can use the Patterns tab to standardize formatting options. Consider the following values, which are valid phone numbers. Next to each value is a pattern representing the value:

PhoneNum	Trifacta pattern
(415) 555-1212	`\(({digit}{3})\) ({digit}{3})\-({digit}{4})`
415-555-1212	`({digit}{3})\-({digit}{3})\-({digit}{4})`
415.555.1212	`({digit}{3}).({digit}{3}).({digit}{3})`
415 555-1212	`({digit}{3}) ({digit}{3})\-({digit}{4})`
1+415-555-1212	`1\+{digit}{3}\-{digit}{3}\-{digit}{4}`

In the Patterns tab, you can select the patterns to which you would like the other patterns in the same pattern group to be converted. Below, the selected **target pattern** becomes the pattern to which other patterns in the column values are converted:



NOTE: You may have to modify the phone number values before attempting the conversion, as they may contain extra alphanumeric values. For example, international country codes (such as `044`) or a preceding `1+` required in long-distance numbers, may need to extracted or removed from the column values prior to conversion.

Generic Conversions

Below are types of conversions that are supported and not supported.

Supported:

Example Source Value	Example Target Value	Notes
123.456.7890	123-456-7890	Changing symbolic characters
(123) 456-7890	123 456-7890	Removing symbolic characters
(123)456-7890	(123)-456-7890	Adding symbolic characters
1234567890	123-456-7890	Splitting a long character group and adding symbolic characters
123-456-7890	1234567890	Merging multiple character groups and removing symbolic characters

Not supported:

Example Source Value	Example Target Value	Notes
123.456.7890	+1.123.456.7890	Adding a new character group
+1.123.456.7890	123.456.7890	Deleting a character group (alphanumeric characters cannot be deleted through pattern standardization)
Adam Wilson	A Wilson	Partial deletion of data from a character group
+1 (123) 456-7890	+001 (123) 456-7890	Prepending or appending a character group with specified characters

Datetime Patterns

For columns of Datetime type, the available Convert mappings are based upon the supported date formats in the platform. Standardization of Datetime patterns is a specific implementation.

Notes on Datetime patterns:

Two-digit years (YY) do not yield four-digit year (YYYY) suggestions due to ambiguity. For example, it is unclear if `50` should map to `1950` or `2050`.

For performance reasons, a maximum of two semantic standardizations can be applied at once. Examples:

Source Value	Possible Standardization	Semantic Mappings	Status
`Jan 1, 1981`	`01/01/1981`	• `Jan 01` • `1 01`	ok (2 mappings)

Jan 1, 1981	01/01/81	• Jan 01 • 1 01 • 1981 81	Not suggested (3 mappings)

For more information on supported formats, see *Datetime Data Type*.

For more information on converting Datetime values to a different format, see *DATEFORMAT Function*.

Patterns by Example

You can generate a new column of values based on pattern matches from a source column. When you enter example values to match with source values, other values with similar patterns may also be matched based on your entered example value.

Tip: This method provides an easy way to build pattern-based matching for values in a source column.

For more information on transformation by example, see *Overview of TBE*.