

EXAMPLE - Extract Values

In this example, you extract one or more values from a source column and assemble them in an Array column.

Suppose you need to extract the hashtags from customer tweets to another column. In such cases, you can use the `{hashtag}` Trifacta pattern to extract all hashtag values from a customer's tweets into a new column.

Source:

The following dataset contains customer tweets across different locations.

User Name	Location	Customer tweets
James	U.K	Excited to announce that we've transitioned Wrangler from a hybrid desktop application to a completely cloud-based service! #dataprep #businessintelligence #CommitToCleanData # London
Mark	Berlin	Learnt more about the importance of identifying issues in your data—early and often #CommitToCleanData #predictivetransformations #realbusinessintelligence
Catherine	Paris	Clean data is the foundation of your analysis. Learn more about what we consider the five tenets of sound #dataprep, starting with #1a prioritizing and setting targets. #startwiththeuser #realbusinessintelligence #Paris
Dave	New York	Learn how #NewYorklife onboarded as part of their #bigdata #dataprep initiative to unlock hidden insights and make them accessible across departments.
Christy	San Francisco	How can you quickly determine the number of times a user ID appears in your data?#dataprep #pivot #aggregation#machinelearning initiatives #SFO

Transformation:

The following transformation extracts the hashtag messages from customer tweets.

Transformation Name	Extract matches into Array
Parameter: Column	customer_tweets
Parameter: Pattern matching elements in the list	`{hashtag}`
Parameter: New column name	Hashtag tweets

Then, the source column can be deleted.

Results:

User Name	Location	Hashtag tweets
James	U.K	["#dataprep", "#businessintelligence", "#CommitToCleanData", "# London"]
Mark	Berlin	["#CommitToCleanData", "#predictivetransformations", "#realbusinessintelligence", "0"]
Catherine	Paris	["#dataprep", "#startwiththeuser", "#realbusinessintelligence", "# Paris"]
Dave	New York	["#NewYorklife", "dataprep", "bigdata", "0"]
Christy	SanFrancisco	["dataprep", "#pivot", "#aggregation", "#machinelearning"]