

EXAMPLE - Countpattern Transform

Source:

The dataset below contains fictitious tweet information shortly after the release of an application called, "Myco ExampleApp".

Date	twitterId	isEmployee	tweet
11/5 /15	lawrencetu38141	FALSE	Just downloaded Myco ExampleApp! Transforming data in 5 mins!
11/5 /15	petramkng024	TRUE	Try Myco ExampleApp, our new free data wrangling app! See www.example.com.
11/5 /15	joetri221	TRUE	Proud to announce the release of Myco ExampleApp, the free version of our enterprise product. Check it out at www.example.com.
11/5 /15	datadaemon994	FALSE	Great start with Myco ExampleApp. Super easy to use, and actually fun.
11/5 /15	99redballoon99	FALSE	Liking this new ExampleApp! Good job, guys!
11/5 /15	bigdatadan7182	FALSE	@support, how can I find example datasets for use with your product?

There are two areas of analysis:

- For non-employees, you want to know if they are mentioning the new product by name.
- For employees, you want to know if they are including cross-references to the web site as part of their tweet.

Transformation:

The following counts the occurrences of the string `ExampleApp` in the `tweet` column. Note the use of the `ignoreCase` parameter to capture capitalization differences:

Transformation Name	Count matches
Parameter: Column	tweet
Parameter: Option	Text or pattern
Parameter: Text or pattern to count	'ExampleApp'
Parameter: Ignore case	true

For non-employees, you want to track if they have mentioned the product in their tweet:

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	<code>if(isEmployee=='FALSE' && countpattern_tweet=='1',true,false)</code>
Parameter: New column name	'nonEmployeeExampleAppMentions'

The following counts the occurrences of `example.com` in their tweets:

Transformation Name	Count matches
Parameter: Column	tweet
Parameter: Option	Text or pattern
Parameter: Text or pattern to count	'example.com'
Parameter: Ignore case	true

For employees, you want to track if they included the above cross-reference in their tweets:

Transformation Name	New formula
Parameter: Formula type	Single row formula
Parameter: Formula	if(isEmployee=='TRUE' && countpattern_tweet1 == 1, true, false)
Parameter: New column name	'employeeWebsiteCrossRefs'

Results:

After you delete the two columns tabulating the counts, you end up with the following:

Date	twitterId	isEmployee	tweet	employeeWebsiteCrossRefs	nonEmployeeExampleAppMentions
11/5 /15	lawrencetl u38141	FALSE	Just downloaded Myco ExampleApp! Transforming data in 5 mins!	false	true
11/5 /15	petramktn g024	TRUE	Try Myco ExampleApp, our new free data wrangling app! See www.example.com.	true	false
11/5 /15	joetri221	TRUE	Proud to announce the release of Myco ExampleApp, the free version of our enterprise product. Check it out at www.example.com.	true	false
11/5 /15	datadaemon994	FALSE	Great start with Myco ExampleApp. Super easy to use, and actually fun.	false	true
11/5 /15	99redballo ns99	FALSE	Liking this new ExampleApp! Good job, guys!	false	true
11/5 /15	bigdatada n7182	FALSE	@support, how can I find example datasets for use with your product?	false	false