What are regular expressions?

MARKETING ANALYTICS IN SPREADSHEETS

Luke Pajer
Digital Marketing Specialist

Digital marketing table

Source	Campaign Name	Ad Group	Cost	CPC
Google	DataCamp Brand	Data Science	$52.28	$1.80
Facebook	DataCamp Brand	Data Analyst	$42.04	$1.27
Google	R users	Data Science	$47.11	$1.39
Facebook	R users	Data Analyst	$54.99	$3.23
Google	Python Users	Data Science	$47.74	$1.77
Facebook	Python Users	Data Analyst	$49.69	$1.91

Digital marketing table

Source	Campaign Name	Ad Group	Cost	CPC
Google	DataCamp Brand	Data Science	$52.28	$1.80
Facebook	DataCamp Brand	Data Analyst	$42.04	$1.27
Google	R users	Data Science	$47.11	$1.39
Facebook	R users	Data Analyst	$54.99	$3.23
Google	Python Users	Data Science	$47.74	$1.77
Facebook	Python Users	Data Analyst	$49.69	$1.91

Answer: use regular expressions

- What are regular expressions?
  - A special search pattern
  - Made of a sequence of characters
  - Also known as 'regex'
- When are they used?
  - To search within strings
  - Commonly used when filtering by categories
- How to get both `R users` and `Python Users`?

Regular expression for users

Breaking down the regular expression that will match all campaigns ending in users:

```
regular_expression = .*[u|U]sers
```

- `.*` - matches **any number** of characters preceding the word 'users'
- `u|U` - matches **either** the lower or upper case variation of letter 'u'
- `[u|U]sers` - matched terms **must** contain one of the items within the brackets

Three regular expression metacharacter categories: Wildcards, Anchors, and Groups.

Basic regular expression characters: wildcards

Wildcards:

- `.` matches any character
- `*` matches 0 or more times
- `?` matches 0 or 1 time
- `+` matches 1 or more times
- `x|y` matches x OR y
- `\` escapes any special character

Examples:

- `d.g` matches both 'dog' and 'dig'
- `.*` matches **all** the letters in 'dog'
- `dogs?` matches either 'dog' or 'dogs'
- `dog.+` matches 'dogs' but not 'dog'
- `dog|cat` matches either 'dog' or 'cat'
- `who\?` matches 'who?'

[1] https://support.google.com/analytics/answer/1034324?hl=en

Basic regular expression characters: anchors

Anchors

- `^x` the start of a string
- `x$` the end of a string

Examples

- `^T` matches 'The dog likes to dig'
- `g$` matches 'The dog likes to dig'

[1] https://support.google.com/analytics/answer/1034324?hl=en

Basic regular expression characters: groups

Groups

- `[x|X]` matches either 'x' or 'X'
- `{x}` matches `x` number of times

Examples

- `[d|D]og` matches either 'dog' or 'Dog'
- `.{2}` matches 'Do' in 'Dog'

[1] https://support.google.com/analytics/answer/1034324?hl=en

Let's get to work!

MARKETING ANALYTICS IN SPREADSHEETS

Test a string using REGEXMATCH

MARKETING ANALYTICS IN SPREADSHEETS

Luke Pajer
Digital Marketing Specialist

Recall the campaigns of interest...

Source	Campaign Name	Ad Group	Cost	CPC
Google	DataCamp Brand	Data Science	$52.28	$1.80
Facebook	DataCamp Brand	Data Analyst	$42.04	$1.27
Google	R users	Data Science	$47.11	$1.39
Facebook	R users	Data Analyst	$54.99	$3.23
Google	Python Users	Data Science	$47.74	$1.77
Facebook	Python Users	Data Analyst	$49.69	$1.91

Recall the campaigns of interest...

Source	Campaign Name	Ad Group	Cost	CPC
Google	DataCamp Brand	Data Science	$52.28	$1.80
Facebook	DataCamp Brand	Data Analyst	$42.04	$1.27
Google	R users	Data Science	$47.11	$1.39
Facebook	R users	Data Analyst	$54.99	$3.23
Google	Python Users	Data Science	$47.74	$1.77
Facebook	Python Users	Data Analyst	$49.69	$1.91

- Cost: total amount spent by the campaign's ad group
- CPC: the average cost each time someone clicks on an ad

REGEXMATCH()

=REGEXMATCH(`STRING TO TEST`, `REGULAR EXPRESSION`)

- Great for testing existence of a string within a string
  - Returns a boolean (`True` or `False`)
- Often used for filtering tables
  - Creating subtables using `FILTER()`
- Useful when aggregating specific categories
  - `SUM()`
  - `AVERAGE()`

Filter with REGEXMATCH()

```
=FILTER(range, REGEXMATCH(string to test, regular expression))
```

Source	Campaign Name	Ad Group	Cost	CPC
Google	DataCamp Brand	Data Science	$52.28	$1.80
Facebook	DataCamp Brand	Data Analyst	$42.04	$1.27
Google	R users	Data Science	$47.11	$1.39
Facebook	R users	Data Analyst	$54.99	$3.23
Google	Python Users	Data Science	$47.74	$1.77
Facebook	Python Users	Data Analyst	$49.69	$1.91

Source	Campaign Name	Ad Group	Cost	CPC
=FILTER(H8:L13,REGEXMATCH(I8:I13,".*[u]sers"))			$47.11	$1.39
Facebook	R users	Data Analyst	$54.99	$3.23

Filter with REGEXMATCH()

```
=FILTER(range, REGEXMATCH(string to test, regular expression))
```

regular expression: .*Users

Source	Campaign Name	Ad Group	Cost	CPC
Google	DataCamp Brand	Data Science	$52.28	$1.80
Facebook	DataCamp Brand	Data Analyst	$42.04	$1.27
Google	R users	Data Science	$47.11	$1.39
Facebook	R users	Data Analyst	$54.99	$3.23
Google	Python Users	Data Science	$47.74	$1.77
Facebook	Python Users	Data Analyst	$49.69	$1.91

Source	Campaign Name	Ad Group	Cost	CPC
Google	Python Users	Data Science	$47.74	$1.77
Facebook	Python Users	Data Analyst	$49.69	$1.91

Aggregate with regular expression

To get a sum of the total ad spend or total cost:

```
=SUM(FILTER(range to sum, REGEXMATCH(string to test, regular expression)))
```

To get the average CPC or Cost-per-Click:

```
=AVERAGE(FILTER(range to average, REGEXMATCH(string to test, regular expression)))
```

Let's get to work!

MARKETING ANALYTICS IN SPREADSHEETS

Modify a string using REGEXEXTRACT and REGEXREPLACE

MARKETING ANALYTICS IN SPREADSHEETS

Luke Pajer
Digital Marketing Specialist

datacamp

Recall the digital marketing table...

Source	Campaign Name	Ad Group	Cost	CPC
Google	DataCamp Brand	Data Science	$52.28	$1.80
Facebook	DataCamp Brand	Data Analyst	$42.04	$1.27
Google	R users	Data Science	$47.11	$1.39
Facebook	R users	Data Analyst	$54.99	$3.23
Google	Python Users	Data Science	$47.74	$1.77
Facebook	Python Users	Data Analyst	$49.69	$1.91

REGEXREPLACE()

=REGEXREPLACE(`STRING`, `REGULAR EXPRESSION`, `REPLACEMENT STRING`)

- Returns: original string and replacement string
  - Excludes the matched part of the string
- Often used to clean up categorical data
- `.[u|U]sers` : Replaces the word 'Users' after the strings R and Python
- Enter `''` to delete the word entirely

```
=REGEXREPLACE(`string`, `regular expression`, '')
```

REGEXEXTRACT()

=REGEXEXTRACT(`STRING`, `REGULAR EXPRESSION`)

- Returns: matched portion of the original string
- Great for extracting a specific portion of different strings
  - Parenthesis `()` indicate the group of characters to extract
- To extract: `Python Users`
  - `(.*).Users` returns `Python`
- `(.*).[u|U]sers` would extract 'Python' and 'R' from the campaign names
  - `(.*)` matches the part of the string to extract
  - `.[u|U]sers` is the part of the string to match, but not extract

Using REGEXREPLACE() and REGEXEXTRACT()

To *replace* only a certain categorical string:

```
=IF(REGEXMATCH(test, regex), REGEXREPLACE(string, regex, replacement string), if False)
```

To *extract* only a certain categorical string:

```
=IF(REGEXMATCH(test, regex), REGEXEXTRACT(string, regex), if False)
```

Let's get to work!

MARKETING ANALYTICS IN SPREADSHEETS

Cleaning Campaign Names

MARKETING ANALYTICS IN SPREADSHEETS

Luke Pajer
Digital Marketing Specialist

Recall: REGEXMATCH()

```
=REGEXMATCH(`string to test`, `regular expression`)
```

Filter tables

```
=FILTER(range, REGEXMATCH(string to test, regular expression))
```

Aggregate by categories

```
=SUM(FILTER(range, REGEXMATCH(string to test, regular expression)))
=AVERAGE(FILTER(range, REGEXMATCH(string to test, regular expression)))
```

Recall: REGEXREPLACE()

Used to replace a portion of a string

```
=REGEXREPLACE(string, regular expression, replacement string)
```

To replace only a certain categorical string

```
=IF(REGEXMATCH(test, regex), REGEXREPLACE(string, regex, replacement string), if False)
```

Recall: REGEXEXTRACT()

Used to extract portions of strings

```
=REGEXEXTRACT(`string`, `regular expression`)
```

To extract only a certain categorical string

```
=IF(REGEXMATCH(test, regex), REGEXEXTRACT(string, regex), if False)
```

Chapter wrap-up

- **Marketing director needs the following changes**
  - The source, campaign, and ad group names to be condensed into a unique id
  - Filter and aggregate the performance metrics using the newly created id
- **Keep in mind**
  - Use regular expressions to match, extract, and replace strings
  - Recall the filtering and aggregation techniques learned in this chapter

Let's get to work!

MARKETING ANALYTICS IN SPREADSHEETS