Cleaning text data

DATA ANALYSIS IN EXCEL

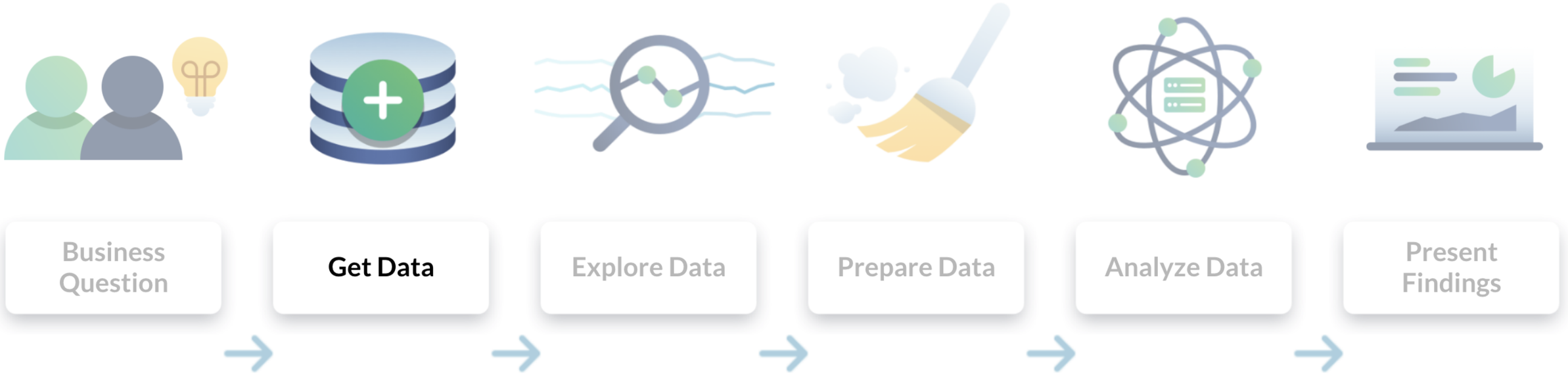
Jen Bricker
Head of Instructor Success, DataCamp

datacamp

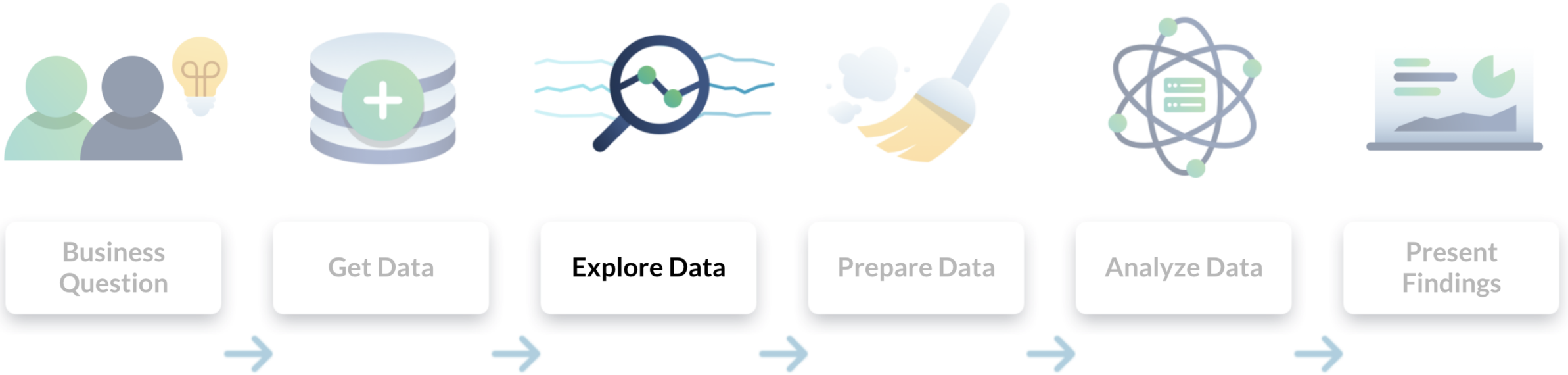
Data analysis process review

Business Question → Get Data → Explore Data → Prepare Data → Analyze Data → Present Findings

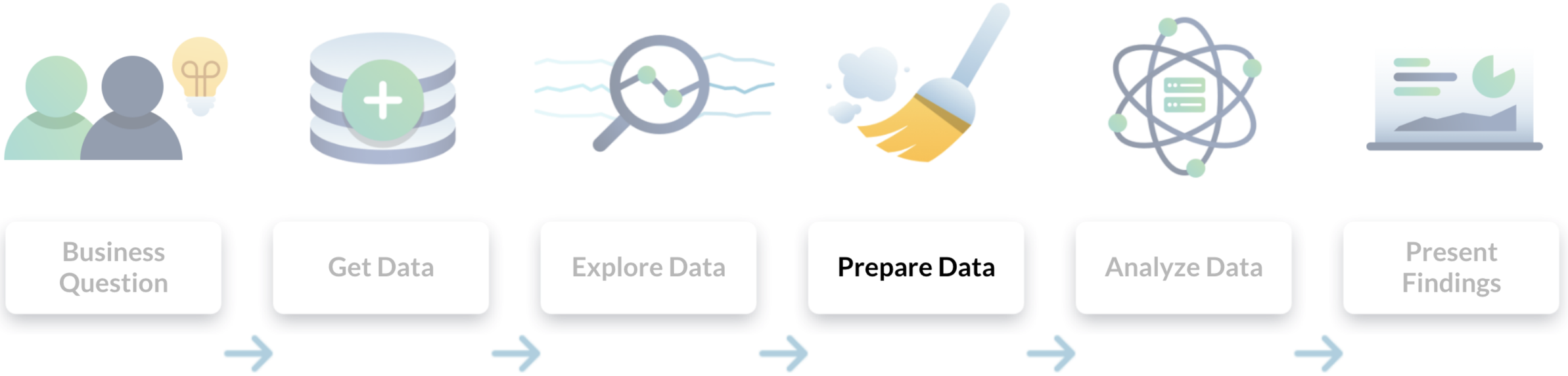
Data analysis process review



Data analysis process review



Data analysis process review



Concatenating text strings

	A	B	C
1	First Name	Last Name	Full Name
2	Ryan	Ahmed	Ryan Ahmed
3	Martynas	Anees	Martynas Anees
4	Catherine	Jonan	Catherine Jonan
5	Jeremy Ryan	Gharavi	Jeremy Ryan Gharavi
6	Marie	Kalmey	Marie Kalmey
7	Nicola	Lyu	Nicola Lyu

CONCATENATE() function syntax

Syntax:

```
=CONCATENATE(value1, [value2],...)
```

Example:

```
=CONCATENATE(A2,B2)
```

	A	B	C
1	id	name	CONCATENATE
2	952844010	Hydrate Edge | Hydration Monitoring Wearable (Canceled)	=CONCATENATE(A2,B2)
3	1299500496	Flyfit - Smart Ankle Tracker For Fitness, Cycling & Swimming	1299500496Flyfit - Smart Ankle Tracker For Fitness, Cycling & Swimming
4	1613449735	KOR-FX Gaming Vest: 4DFX Haptic Feedback System	1613449735KOR-FX Gaming Vest: 4DFX Haptic Feedback System
5	273890115	ProfileMyRun: Run the Right Way, Run the Natural Way	273890115ProfileMyRun: Run the Right Way, Run the Natural Way
6	353045033	365 DAZE - Apple Watch minimal wallpapers	353045033365 DAZE - Apple Watch minimal wallpapers
7	159764729	KYMIRA Sport: The Most Advanced Reactive Sports Apparel	159764729KYMIRA Sport: The Most Advanced Reactive Sports Apparel
8	906749645	UV Anti-Cancer Melanoma Skin Protection James Prattas Co	906749645UV Anti-Cancer Melanoma Skin Protection James Prattas Co
9	661560435	DAZLN: NFC Nails that Light Up Holiday Parties!	661560435DAZLN: NFC Nails that Light Up Holiday Parties!

CONCATENATE() improved

Example:

```
=CONCATENATE(A2," - ",B2)
```

	A	B	C
1	id	name	CONCATENATE improved
2	952844010	Hydrate Edge | Hydration Monitoring Wearable (Canceled)	=CONCATENATE(A2," - ",B2)
3	1299500496	Flyfit - Smart Ankle Tracker For Fitness, Cycling & Swimming	1299500496 - Flyfit - Smart Ankle Tracker For Fitness, Cycling & Swimming
4	1613449735	KOR-FX Gaming Vest: 4DFX Haptic Feedback System	1613449735 - KOR-FX Gaming Vest: 4DFX Haptic Feedback System
5	273890115	ProfileMyRun: Run the Right Way, Run the Natural Way	273890115 - ProfileMyRun: Run the Right Way, Run the Natural Way
6	353045033	365 DAZE - Apple Watch minimal wallpapers	353045033 - 365 DAZE - Apple Watch minimal wallpapers
7	159764729	KYMIRA Sport: The Most Advanced Reactive Sports Apparel	159764729 - KYMIRA Sport: The Most Advanced Reactive Sports Apparel
8	906749645	UV Anti-Cancer Melanoma Skin Protection James Prattas Co	906749645 - UV Anti-Cancer Melanoma Skin Protection James Prattas Co
9	661560435	DAZLN: NFC Nails that Light Up Holiday Parties!	661560435 - DAZLN: NFC Nails that Light Up Holiday Parties!

LOWER(), UPPER(), and PROPER() functions

Syntax:

```
=LOWER(text1)
```

Syntax:

```
=UPPER(text1)
```

Syntax:

```
=PROPER(text1)
```

Output

	A	B	C	D
1	city	LOWER	UPPER	PROPER
2	Salt Lake City	salt lake city	SALT LAKE CITY	Salt Lake City
3	San Francisco	san francisco	SAN FRANCISCO	San Francisco
4	Cambridge	cambridge	CAMBRIDGE	Cambridge
5	Palo Alto	palo alto	PALO ALTO	Palo Alto
6	Toronto	toronto	TORONTO	Toronto
7	Reading	reading	READING	Reading
8	Captain Cook	captain cook	CAPTAIN COOK	Captain Cook
9	Hong Kong	hong kong	HONG KONG	Hong Kong
10	San Diego	san diego	SAN DIEGO	San Diego

Let's practice!

DATA ANALYSIS IN EXCEL

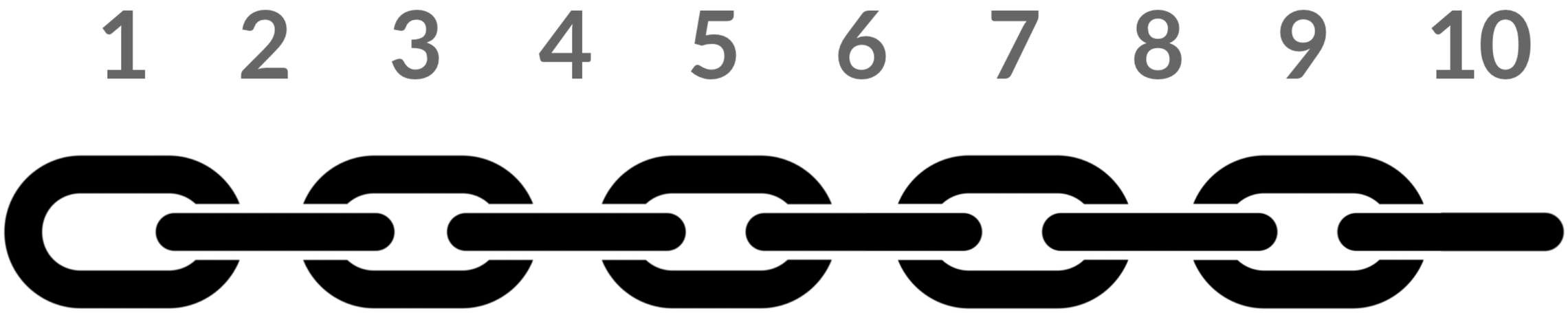
Extracting text from cells

DATA ANALYSIS IN EXCEL

Jen Bricker
Head of Instructor Success, DataCamp

datacamp

What is a string?



Counting characters

LEN() function syntax

Syntax:

```
=LEN(text)
```

Example:

```
=LEN("3D My Kicks - 3Dmykicks.com")
```

or

```
=LEN(B2)
```

Extracting text

{3D My Kicks} - 3Dmykicks.com

LEFT() function syntax

Syntax:

```
=LEFT(text,number_of_characters)
```

Example:

```
=LEFT(B2,LEN("3D My Kicks"))
```

	B	C
1	name	LEFT and LEN
2	3D My Kicks - 3Dmykicks.com	3D My Kicks

Extracting text continued

	M
1	state
2	UT,"type"
3	CA,"type"
4	MA,"type"
5	CA,"type"
6	ON,"type"
7	England,"type"
8	HI,"type"
9	Hong Kong Island,"type"
10	CA,"type"
11	AZ,"type"
12	LA,"type"

What is the length of the bad string?

```
=LEN(",'type'")
```

result: 7

Nesting LEFT() and LEN()

Example

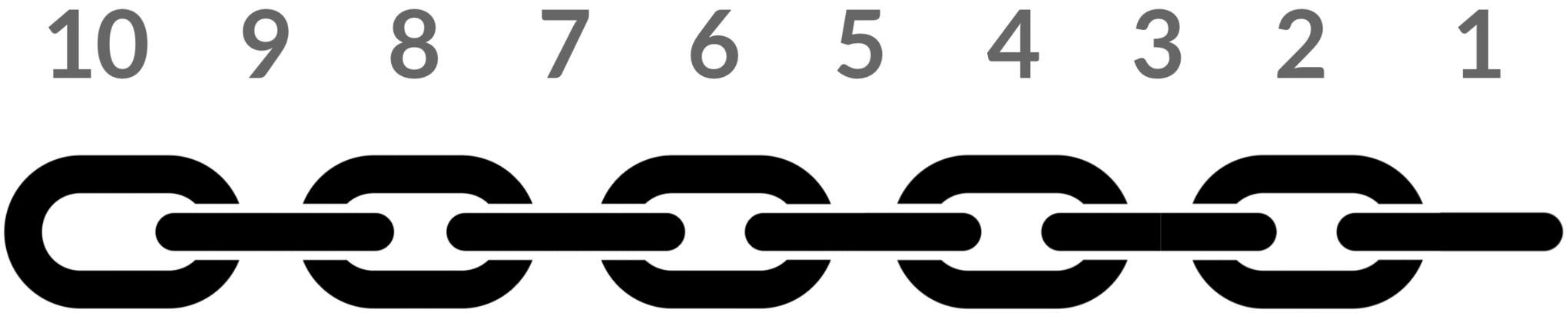
```
=LEFT(M2,LEN(M2)-7)
```

	M	N
1	state	LEFT and LEN combined
2	UT,"type"	=LEFT(M2,LEN(M2)-7)
3	CA,"type"	CA
4	MA,"type"	MA
5	CA,"type"	CA
6	ON,"type"	ON
7	England,"type"	England
8	HI,"type"	HI
9	Hong Kong Island,"type"	Hong Kong Island
10	CA,"type"	CA
11	AZ,"type"	AZ
12	LA,"type"	LA
13	CT,"type"	CT
14	IN,"type"	IN

RIGHT() function syntax

Syntax:

```
=RIGHT(text,number_of_characters)
```



Replacing text

	B
1	name
2	Hydrate Edge | Hydration Monitoring Wearable (Canceled)
3	Flyfit - Smart Ankle Tracker For Fitness, Cycling & Swimming
4	KOR-FX Gaming Vest: 4DFX Haptic Feedback System
5	ProfileMyRun: Run the Right Way, Run the Natural Way
6	365 DAZE - Apple Watch minimal wallpapers
7	KYMIRA Sport: The Most Advanced Reactive Sports Apparel
8	UV Anti-Cancer Melanoma Skin Protection James Prattas Co
9	DAZLN: NFC Nails that Light Up Holiday Parties!
10	Vivir Wearable Technology - Heated Fitness Apparel
11	jmpLite - The first human powered safety light for runners
12	3D My Kicks - 3dmykicks.com

SUBSTITUTE() function syntax

Syntax:

```
=SUBSTITUTE(text, old_text, new_text, [instance])
```

Example:

```
=SUBSTITUTE(B2,":","-")
```

Output

	B	C
1	name	SUBSTITUTE
2	Hydrate Edge | Hydration Monitoring Wearable (Canceled)	=SUBSTITUTE(B2,":","-")
3	Flyfit - Smart Ankle Tracker For Fitness, Cycling & Swimming	Flyfit - Smart Ankle Tracker For Fitness, Cycling & Swimming
4	KOR-FX Gaming Vest: 4DFX Haptic Feedback System	KOR-FX Gaming Vest- 4DFX Haptic Feedback System
5	ProfileMyRun: Run the Right Way, Run the Natural Way	ProfileMyRun- Run the Right Way, Run the Natural Way
6	365 DAZE - Apple Watch minimal wallpapers	365 DAZE - Apple Watch minimal wallpapers
7	KYMIRA Sport: The Most Advanced Reactive Sports Apparel	KYMIRA Sport- The Most Advanced Reactive Sports Apparel
8	UV Anti-Cancer Melanoma Skin Protection James Prattas Co	UV Anti-Cancer Melanoma Skin Protection James Prattas Co
9	DAZLN: NFC Nails that Light Up Holiday Parties!	DAZLN- NFC Nails that Light Up Holiday Parties!
10	Vivir Wearable Technology - Heated Fitness Apparel	Vivir Wearable Technology - Heated Fitness Apparel
11	jmpLite - The first human powered safety light for runners	jmpLite - The first human powered safety light for runners
12	3D My Kicks - 3dmykicks.com	3D My Kicks - 3dmykicks.com

Let's practice!

DATA ANALYSIS IN EXCEL

Preparing date data

DATA ANALYSIS IN EXCEL

Jen Bricker
Head of Instructor Success, DataCamp

datacamp

Interpreting date analysis requests

Look at the project data by date

- Data by launch date?
- How long a project campaign ran?
- Trends by year, month, day of the week?

Month of La..	
January	6
February	9
March	6
April	6
May	6
June	10
July	7
August	12
September	10
October	11
November	13
December	7

Project length

Syntax:

```
=deadline-launched_at
```

Example:

```
=E2-D2
```

	D	E	F
1	launched_at	deadline	Project Length
2	8/30/2016	10/14/2016	=E2-D2
3	2/13/2014	3/25/2014	40
4	6/9/2014	7/24/2014	45
5	10/10/2014	11/9/2014	30
6	10/19/2018	11/24/2018	36
7	8/17/2014	9/17/2014	31
8	10/5/2014	11/1/2014	27
9	11/4/2014	11/29/2014	25
10	1/26/2015	3/6/2015	39
11	11/4/2014	12/14/2014	40
12	10/31/2014	11/30/2014	30

Dynamic date and time entry

	E	F	G
1	deadline	NOW	TODAY
2	10/14/2016	1274.47	1274
3	3/25/2014	2208.47	2208
4	7/24/2014	2087.47	2087
5	11/9/2014	1979.47	1979
6	11/24/2018	503.47	503
7	9/17/2014	2032.47	2032
8	11/1/2014	1987.47	1987
9	11/29/2014	1959.47	1959
10	3/6/2015	1862.47	1862
11	12/14/2014	1944.47	1944

NOW() and TODAY() functions syntax

Syntax:

```
=NOW()
```

```
=TODAY()
```

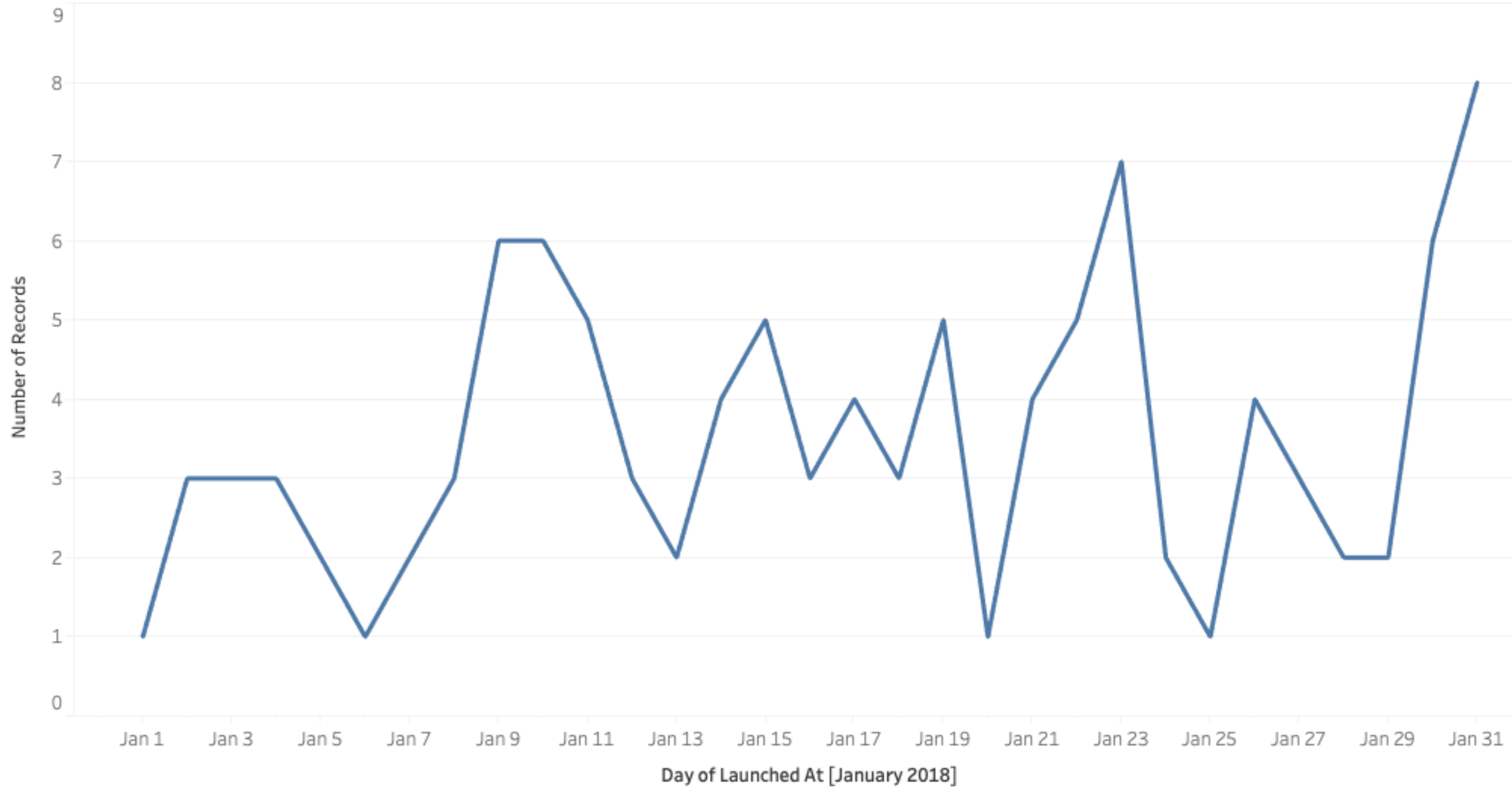
Example:

```
=NOW()-E2
```

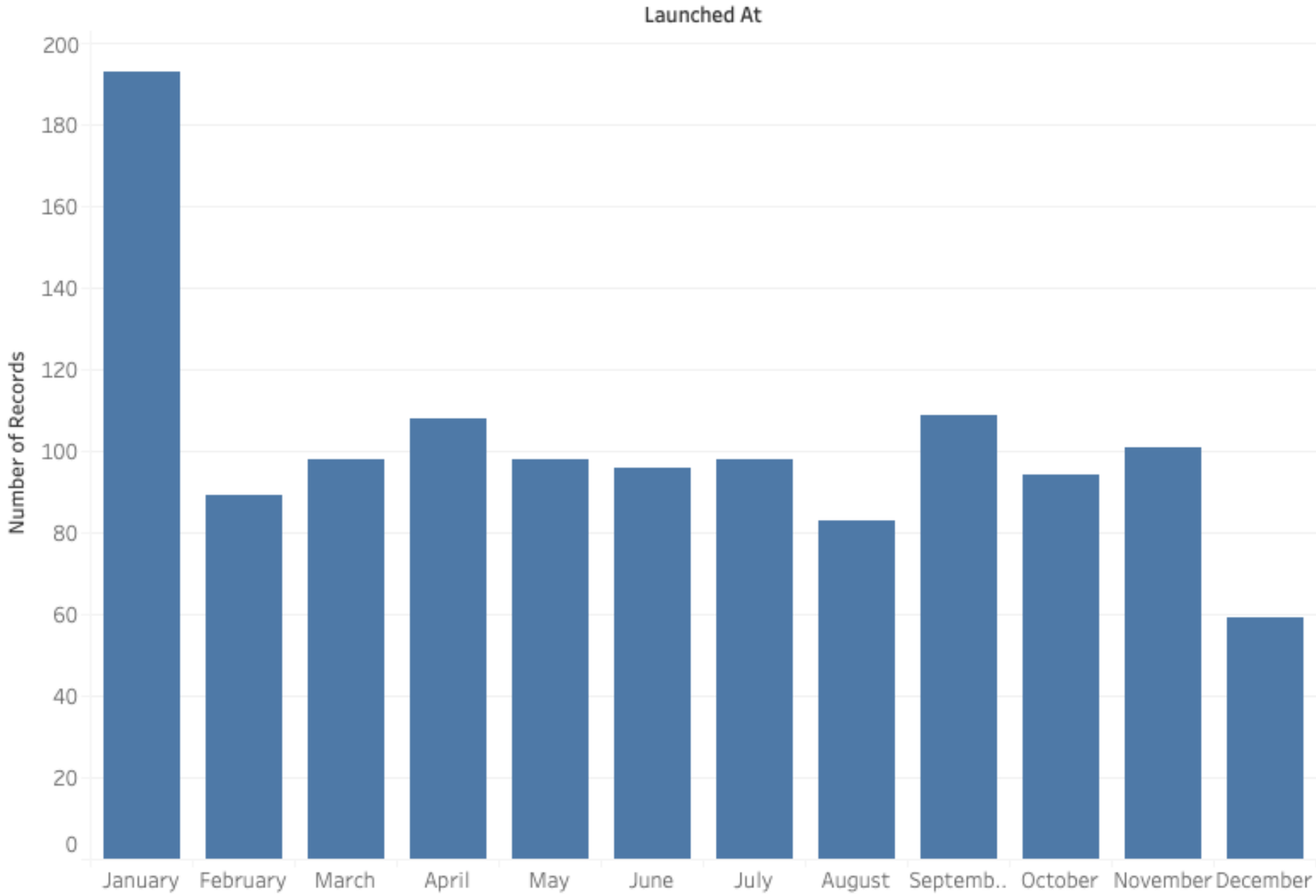
```
=TODAY()-E2
```

	E	F	G
1	deadline	NOW	TODAY
2	10/14/2016	1274.47	=TODAY()-E2
3	3/25/2014	2208.47	2208
4	7/24/2014	2087.47	2087
5	11/9/2014	1979.47	1979
6	11/24/2018	503.47	503
7	9/17/2014	2032.47	2032
8	11/1/2014	1987.47	1987
9	11/29/2014	1959.47	1959
10	3/6/2015	1862.47	1862
11	12/14/2014	1944.47	1944

Continuous vs discrete date analysis



Continuous vs discrete date analysis



Months and days of the week

	E	F	G
1	deadline	MONTH	WEEKDAY
2	10/14/2016	10	6
3	3/25/2014	3	3
4	7/24/2014	7	5
5	11/9/2014	11	1
6	11/24/2018	11	7
7	9/17/2014	9	4
8	11/1/2014	11	7
9	11/29/2014	11	7
10	3/6/2015	3	6
11	12/14/2014	12	1

MONTH() function syntax

Syntax:

```
=MONTH(date)
```

Example:

```
=MONTH(E2)
```

	E	F
1	deadline	MONTH
2	10/14/2016	=MONTH(E2)
3	3/25/2014	3
4	7/24/2014	7
5	11/9/2014	11
6	11/24/2018	11
7	9/17/2014	9
8	11/1/2014	11
9	11/29/2014	11
10	3/6/2015	3
11	12/14/2014	12

WEEKDAY() function syntax

Syntax:

`=WEEKDAY(date,[return_type])`

`return_type` options

- **1 or omitted**: Sunday (1) to Saturday (7)
- **2**: Monday (1) to Sunday (7)

Example:

`=WEEKDAY(E2,2)`

	E	F
1	deadline	WEEKDAY
2	10/14/2016	=WEEKDAY(E2)
3	3/25/2014	3
4	7/24/2014	5
5	11/9/2014	1
6	11/24/2018	7
7	9/17/2014	4
8	11/1/2014	7
9	11/29/2014	7
10	3/6/2015	6
11	12/14/2014	1

Let's practice!

DATA ANALYSIS IN EXCEL

The most important function in Excel

DATA ANALYSIS IN EXCEL

Jen Bricker
Head of Instructor Success, DataCamp

datacamp

VLOOKUP() introduction

	A	B	C	D	E
1	weekday_number	weekday_name		month_number	month_name
2	1	Monday		1	January
3	2	Tuesday		2	February
4	3	Wednesday		3	March
5	4	Thursday		4	April
6	5	Friday		5	May
7	6	Saturday		6	June
8	7	Sunday		7	July
9				8	August
10				9	September
11				10	October
12				11	November
13				12	December
14					
15					
16					
17					
18					

Date Tables | Kickstarter

VLOOKUP

VLOOKUP

VERTICAL

VLOOKUP() syntax

Syntax:

```
=VLOOKUP (lookup_value, table_array, col_num, [range_lookup])
```

- `lookup_value` : what you want to look up
- `table_array` : where you want to look for it
- `col_num` : the column number in the range containing the value to return
- `range_lookup` : return an approximate (`TRUE`) or exact match (`FALSE`)

Columns in common

	E	F	G
1	deadline	MONTH	Project Length
2	10/14/2016	10	45
3	3/25/2014	3	40
4	7/24/2014	7	45
5	11/9/2014	11	30
6	11/24/2018	11	36
7	9/17/2014	9	31
8	11/1/2014	11	27
9	11/29/2014	11	25
10	3/6/2015	3	39
11	12/14/2014	12	40
12	11/30/2014	11	30
13	1/6/2015	1	60
14	11/28/2014	11	30

Kickstarter | Date Tables

	D	E	F
1	month_number	month_name	
2	1	January	
3	2	February	
4	3	March	
5	4	April	
6	5	May	
7	6	June	
8	7	July	
9	8	August	
10	9	September	
11	10	October	
12	11	November	
13	12	December	
14			

Kickstarter | Date Tables

VLOOKUP() applied

Syntax:

```
=VLOOKUP (lookup_value, table_array,
col_num, [range_lookup])
```

Example:

```
=VLOOKUP(F2,'Date Tables'!D:E,2,FALSE)
```

	F	G
1	MONTH	VLOOKUP
2	10	=VLOOKUP(F2,'Date Tables'!D:E,2,FALSE)
3	3	March
4	7	July
5	11	November
6	11	November

Date Tables sheet:

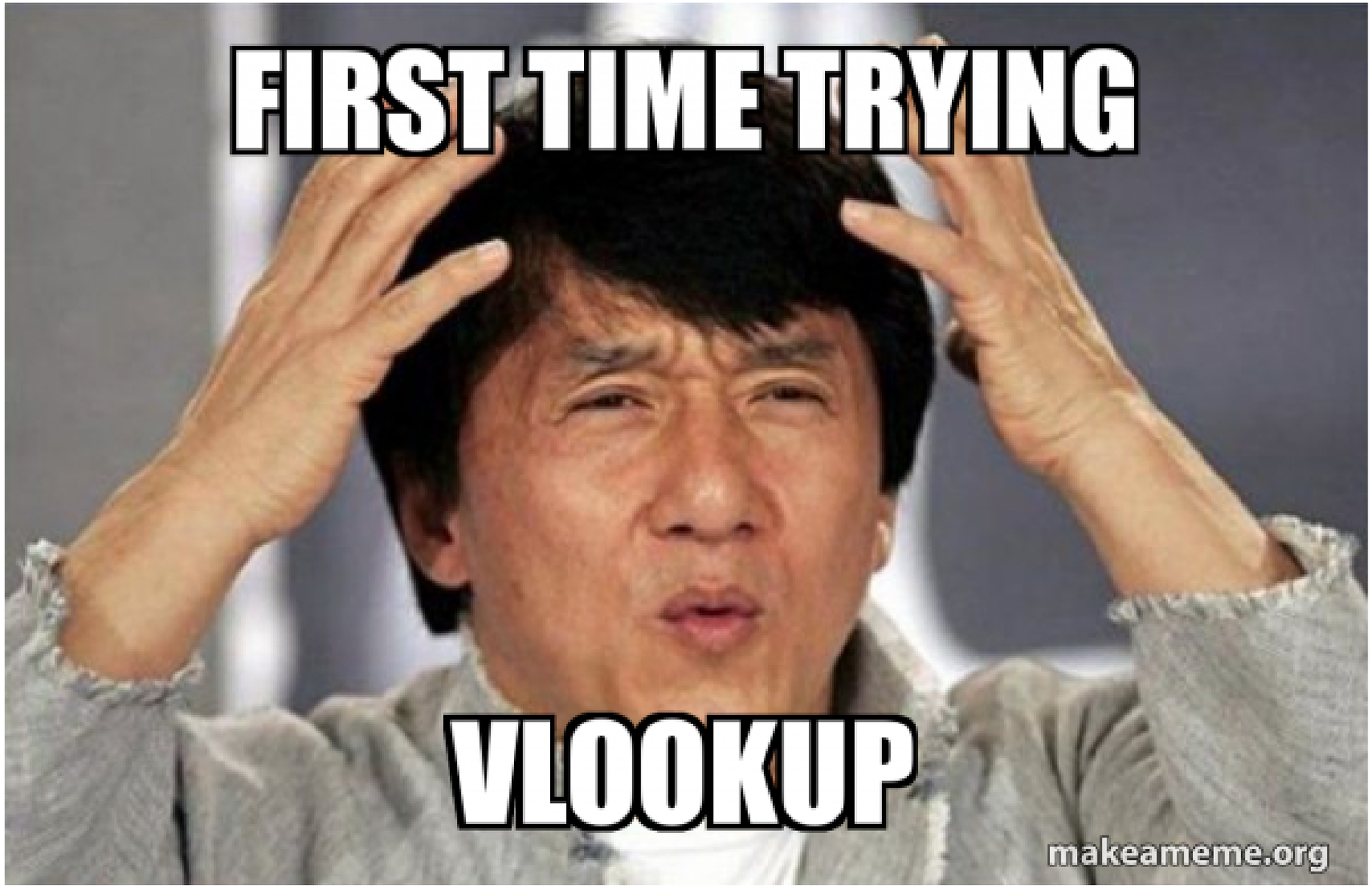
	D	E
1	month_number	month_name
2	1	January
3	2	February
4	3	March
5	4	April
6	5	May
7	6	June
8	7	July
9	8	August
10	9	September
11	10	October
12	11	November
13	12	December

VLOOKUP() and MONTH() combined

Nested Formula Example:

```
=VLOOKUP(MONTH(E2),'Date Tables'!D:E,2,FALSE)
```

	E	F
1	deadline	VLOOKUP
2	10/14/2016	=VLOOKUP(MONTH(E2),'Date Tables'!D:E,2,FALSE)
3	3/25/2014	March
4	7/24/2014	July
5	11/9/2014	November
6	11/24/2018	November
7	9/17/2014	September
8	11/1/2014	November
9	11/29/2014	November


FIRST TIME TRYING
VLOOKUP
makeameme.org

Let's practice!

DATA ANALYSIS IN EXCEL