

Contents

Assignment.....	2
Assignment in brief	2
Due Dates & Checklist.....	2
Report	2
Top-line Summary:.....	3
Main Q&A.....	3
The Data	4
Data Background.....	4
Variable Definitions.....	4
Structure of the Excel Document.....	5
Quirks in the data.....	5
Examples of Creatives from the Web	6
Grading.....	7
Excel Workshop	8
About this workshop.....	8
Sorting data.....	8
Computing numeric day of the week using the WEEKDAY Function.....	10
VLOOKUP Function: Computing <i>written</i> day of the week	11
VLOOKUP Practice with age.....	14
PIVOT Tables: Example 1 (gender x conversions)	15
PIVOT Tables: Example 2 (Number of touch points before conversion)	16
PIVOT Tables: Example 3 (Nesting)	18
Basic IF statements	19
Nested IF statements: Example 1 (weekend or not?).....	19
Advanced Nested IF statements: Example 2 (morning, afternoon, or evening?).....	20
All-Around Advanced: Nested IF statements + filtering + VLOOKUP :)	21
There must be an easier way??	23
A note on descriptive statistics	24

Calculating descriptive statistics in Excel.....	24
Plotting the distribution.....	24
What descriptive(s) are “best”?.....	25
Graphing the distribution	27
Calculating percentiles.....	30
Bonus: adding cumulative percent line to your graph	34
Excel resource if you get stuck.....	37

Assignment

Assignment in brief

Your task is to explore data from a display campaign run in the Irish market, then write a brief report consisting of 3 questions and a topline summary. You may work with a partner of your choosing or alone.

Due Dates & Checklist

Feb 22/23 (depending on class section) anytime before your class section starts. If you are working with a partner in a different section, the due date is the *earlier* of your two sections –
Upload your report & excel document [here](#) ¹:

Checklist

- ☐ Report is a word doc
- ☐ Report has no cover page – jump right in!
- ☐ First line of report has student ID(s) and names (*NOT* in a header, they are difficult to copy when they are in there)
- ☐ Report is named using the following convention: studentD1_studentID2.docx, e.g., 123456_342344.docx (If you did not work with a partner, you'll just have 1 number in your file name).
- ☐ Excel file is named using the same convention as the word doc
- ☐ All word counts are adhered to

Report

The report consists of **2 parts: a topline summary and a Q&A**. The report is meant to be brief. Pay careful attention to word counts and use the space wisely.

¹ If you are working with a partner, only one person needs to upload the report.

Top-line Summary:

The topline summary is in **bullet-point form**. You are to write <1 sentence to 2 sentences on the following:

1. In future campaigns, how many impressions should we “spend” on each individual person² if we assume that the client is on a budget and billed on a cost per mill basis?³ (<1-2 sentences)
2. Make three **more** recommendations for optimizing future display campaigns based on your insights from the dataset (<1-2 sentences each)
 - a.
 - b.
 - c.
3. How would you measure success and what are the most important attributes for success? (1-2 sentences)

Main Q&A

Note: your answers should be supported with the dataset, in class material, and readings. Cite all sources using either APA or Harvard Style (<https://libguides.ucd.ie/academicintegrity/harvardstyle>)

1. Explain the logic behind your recommended impressions/person(s). Remember that the client is billed on an impression basis and is on a budget, so you will have to decide what your strategy should be. For example, should we try to reach a lot of people at the sacrifice of repetition, or should we try to reach a smaller number of people more times? Whatever your strategy, be sure to provide actual numerical recommendations (e.g., 5 impressions (also known as “touch points”) per person) and support your answer with evidence from the data (e.g., tables, figures, et cetera) and learnings from class/readings/online, where appropriate. (max 400 words)
2. Explain the logic behind your three **additional** recommendations for optimizing future display campaigns as noted in the topline summary. Support your answer with evidence from the data (e.g., tables, figures, et cetera) and learnings from class, where appropriate. (max 600 words)
3. The final question ignores the dataset and asks you to do some research on display campaigns more generally. Based on your own research, how would you measure the success of a display campaign. What attributes matter for success? Please be specific and use citations to support your answers. Answers that seem to rely only on intuition or opinion will not score well. (max 700 words).

² If you have different recommendations for different types of people, please specify that in your answer

³ Cost per mill = cost per 1,000 impressions

The Data

Data Background

This is confidential, white-labeled data from [Quantcast](#). The data is from a display campaign run for an Irish bank selling mortgages. A major goal of the campaign was to get people to request more information or a meeting using an online callback form. Hence, in this dataset, a “conversion” occurs when someone fills out the callback form.

The dataset contains information on how many ads were served to each individual (n=235), what gender and age the individuals are (estimated), when and where the ads were served, what targeting was employed (retargeting, prospecting), and what kind of conversion occurred (click, view – explained below under “Targeting Type”).

The campaign was run on a cost per mill basis. That is, advertisers were charged for every 1,000 impression. They were *not* charged based on clicks.

Variable Definitions

The dataset contains 14 unique variables across several tabs.

pUnit Number: There are 235 individuals in this dataset. pUnit number identifies each individual by their cookie (numbered simply 1 – 238 (a few people were removed, hence the total n is 235)).

Event_date: The date and time at which an ad exposure occurred.

Campaign: The campaign associated with this data (i.e., mortgages).

Targeting type: There are four levels here

- a. Prospecting:* at the time of exposure, the individual had *not* visited the bank’s mortgage landing pages in the recent past.
- b. Retargeting:* at the time of exposure, the individual *had* visited the bank’s mortgage landing pages in the recent past.
- c. Click Conversion:* the individual clicked on the ad and filled out the callback form
- d. View Conversion:* the individual did *not* click on the ad, but later went to the site and filled out the callback form.

Event_device: Whether the ad was served on a desktop, mobile device, tablet, app, et cetera.

Event_category: There are four levels here

- a. *First_touch*: the first time an individual is exposed to an ad from this display campaign
- b. *Middle_touch*: the ad exposure(s) that occur between the first and last touch
- c. *Last_touch*: the final ad served before the “conversion” ad (i.e., the 2nd to last ad)
- d. *Conversion*: the final ad served before conversion

Event_domain: The website where the ad was shown

Creativesize: The dimensions of the ad (see “examples of creatives from the web” below for examples of the sizes).

Gender: The estimated gender of the individual

Age_group: The estimated age of the individual

First Touch Date (see tab Mortgages TTC): The date of the first touch

First Touch TTC (see tab Mortgages TTC): The time in days from first touch to conversion

Last Touch Date (see tab Mortgages TTC): The date of the last touch

Last Touch TTC (see tab Mortgages TTC): The time in days from last touch to conversion

Structure of the Excel Document

The Excel document contains 3 tabs:

1. Mortgages – Raw
2. Mortgages – TTC
3. Data Dive Practice

The first tab contains the raw data; all remaining tabs are derived from this tab. The second tab (Mortgages TTC) contains “time to convert” (TTC) data. It is the same data as the “raw” tab, but only the “conversion” line is shown. The final tab (Data Dive Practice) is a truncated version of the raw data that we’ll use for in-class practice.

Quirks in the data

You may notice that some people have no “first touch.” There are several explanations for this:

1. *The “first touch” happened outside the date range of this dataset*
2. *The “first touch” and “conversion” happened so close together that the system thought it more appropriate to label the “first touch” as a “last touch.”*

3. Conversion happened immediately.

While you may want to look at TTC (time to convert) and make recommendations, keep in mind the limitations of your claims – you will be missing data from a significant chunk of people who don't have a "first touch." What impact will that have on your recommendations, if any? Top papers will note the limitations and how those limitations might change their answers, if they choose to explore TTC.

Examples of Creatives from the Web

Note – we do *not* know which Irish bank ran this campaign. However, I thought it might be helpful to show some examples of mortgage-related display ads with their dimensions.

300*250



728*90

160*600



Grading

UCD defines the following four grade brackets as follows: D = adequate, C = good, B = great, and A = excellent. Grades lower than D- indicate a failing grade. Below, I'll give you some more context regarding what each bracket constitutes

Grade	UCD Criteria	Specifics
A range	Excellent	<ul style="list-style-type: none"> Many high-quality insights (of which many are non-obvious, i.e, <i>not</i> low hanging fruit) Compelling recommendations supported not just by the data in the dataset, but also by my lectures/assigned readings and other relevant outside readings. Makes me think (on several occasions), "I wish I'd thought of that!" or "I wish I'd found that reading!" or "I wish I'd noticed that!" etc. Does not fall victim to spurious correlations or overstating insights based on the data
B range	Great	<ul style="list-style-type: none"> Many high-quality insights (though not as nuanced or non-obvious as the A-range) Compelling recommendations that are well-supported, but not at the A-level Rarely, if ever, falls victim to spurious correlations or overstating insights based on the data
C range	Good	<ul style="list-style-type: none"> Good insights, but many fall under the "low hanging fruit" category (i.e., are easy to identify) Recommendations are logically supported but could be fleshed out more/researched more. Some instances of opinion rather than actual support May overstate insights of the data, may miss important insights
D range	Adequate	<ul style="list-style-type: none"> Nearly all insights fall under the "low hanging fruit" category Recommendations may be based on opinion May overstate insights of the data May have some mistakes in data analysis
Below D-	Not passing	<ul style="list-style-type: none"> Incorrect insights Poor support, based mostly on opinion Substantial mistakes in data analysis Thin on data, research, and writing

Excel Workshop

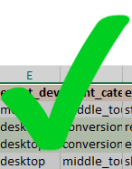
About this workshop

This workshop is meant to cater to multiple levels of Excel. If some of the tasks are too easy for you, skip ahead! And don't be shy about raising your hand 😊

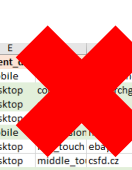
Sorting data

Go to tab “data dive practice” – all examples will be based on this tab. The cookie data (pUnit Number) are out of order. Let's fix this.

1. Select *all* data, not just the cookie column—otherwise, the rest of the data become dissociated. A quick way to select all the data without dragging your mouse around:
 - a. Click on Cell A1 (not *into* the cell, just click so it is selected)
 - b. Hold down CTRL SHIFT (Command Shift on Mac)
 - c. While holding down CTRL SHIFT, hit your right arrow key.
 - d. While still holding down CTRL Shift, hit your down arrow key.

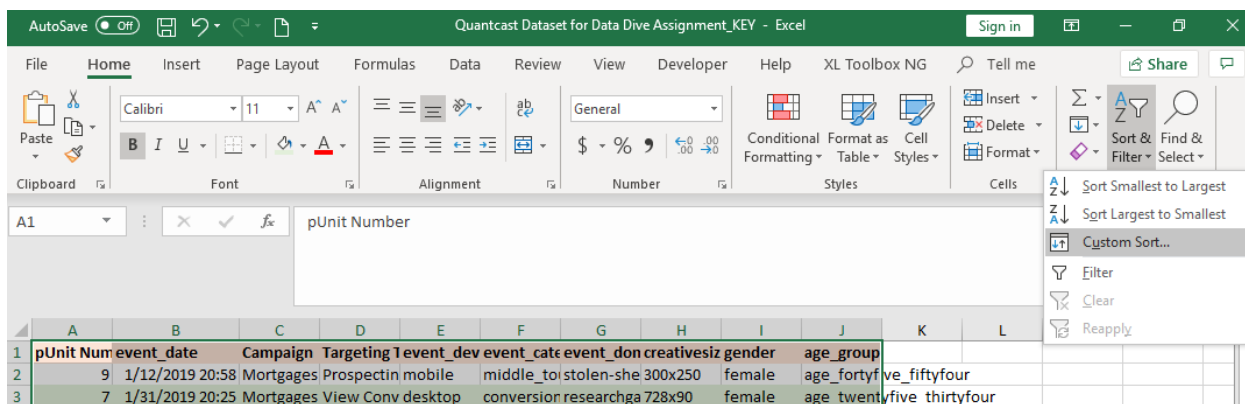


	A	B	C	D	E	F	G	H	I	J	K	L
1	pUnit Num	event_date	Campaign	Targeting	Event dev	event_cate	event_don	creativesiz	gender	age_group		
2	9	1/12/2019 20:58	Mortgages Prospectin mobile	middle_toi	stolen-she	300x250	female	age_fortyfive	fiftyfour			
3	7	1/31/2019 20:25	Mortgages View Conv desktop	conversion	researchg	728x90	female	age_twentyfive	thirtyfour			
4	5	1/7/2019 9:58	Mortgages View Conv desktop	conversion	ebay.ie	300x250	male	age_twentyfive	thirtyfour			
5	5	1/1/2019 19:10	Mortgages Retargetin desktop	middle_toi	skyscanner	300x250	male	age_twentyfive	thirtyfour			
6	6	1/22/2019 9:59	Mortgages View Conv mobile	conversion	mumsnet.i	300x250	male	age_twentyfive	thirtyfour			
7	5	1/6/2019 15:58	Mortgages Retargetin desktop	last_touch	ebay.ie	300x250	male	age_twentyfive	thirtyfour			
8	5	1/5/2019 22:24	Mortgages Retargetin desktop	middle_toi	csfd.cz	160x600	male	age_twentyfive	thirtyfour			
9	4	1/14/2019 8:25	Mortgages View Conv desktop	conversion	daft.ie	728x90	male	age_fiftyfive	sixtyfour			
10	1	1/2/2019 13:50	Mortgages View Conv desktop	conversion	daft.ie	728x90	male	age_twentyfive	thirtyfour			
11	7	1/29/2019 20:03	Mortgages Retargetin desktop	last_touch	researchg	728x90	female	age_twentyfive	thirtyfour			
12	10	1/8/2019 23:28	Mortgages Retargetin mobile	middle_toi	whattoexp	300x250	female	age_twentyfive	thirtyfour			
13	9	1/10/2019 20:14	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
14	4	1/6/2019 18:15	Mortgages Prospectin desktop	last_touch	daft.ie	728x90	male	age_fiftyfive	sixtyfour			
15	9	1/5/2019 14:18	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
16	9	1/13/2019 12:27	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
17	9	1/6/2019 17:36	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
18	5	1/5/2019 22:24	Mortgages Retargetin desktop	middle_toi	b.cpx.cz	728x90	male	age_twentyfive	thirtyfour			
19	9	1/6/2019 0:02	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
20	9	1/16/2019 23:21	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
21	5	1/1/2019 19:05	Mortgages Retargetin desktop	first_touch	skyscanner	300x250	male	age_twentyfive	thirtyfour			
22	9	1/18/2019 20:48	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
23	9	1/6/2019 0:02	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
24	9	1/15/2019 18:37	Mortgages Prospectin mobile	middle_toi	stolen-she	300x250	female	age_fortyfive	fiftyfour			
25	1	1/2/2019 12:53	Mortgages Prospectin desktop	last_touch	daft.ie	728x90	male	age_twentyfive	thirtyfour			



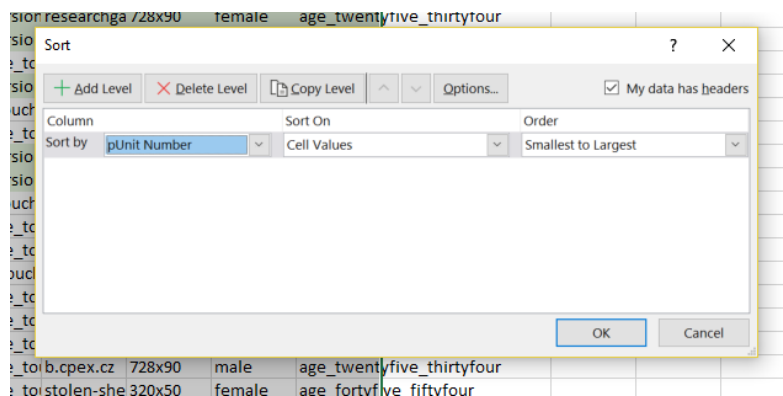
	A	B	C	D	E	F	G	H	I	J	K	L
1	pUnit Num	event_date	Campaign	Targeting	Event dev	event_cate	event_don	creativesiz	gender	age_group		
2	9	1/12/2019 20:58	Mortgages Prospectin mobile	middle_toi	stolen-she	300x250	female	age_fortyfive	fiftyfour			
3	7	1/31/2019 20:25	Mortgages View Conv desktop	conversion	researchg	728x90	female	age_twentyfive	thirtyfour			
4	5	1/7/2019 9:58	Mortgages View Conv desktop	conversion	ebay.ie	300x250	male	age_twentyfive	thirtyfour			
5	5	1/1/2019 19:10	Mortgages Retargetin desktop	middle_toi	skyscanner	300x250	male	age_twentyfive	thirtyfour			
6	6	1/22/2019 9:59	Mortgages View Conv mobile	conversion	mumsnet.i	300x250	male	age_twentyfive	thirtyfour			
7	5	1/6/2019 15:58	Mortgages Retargetin desktop	last_touch	ebay.ie	300x250	male	age_twentyfive	thirtyfour			
8	5	1/5/2019 22:24	Mortgages Retargetin desktop	middle_toi	csfd.cz	160x600	male	age_twentyfive	thirtyfour			
9	4	1/14/2019 8:25	Mortgages View Conv desktop	conversion	daft.ie	728x90	male	age_fiftyfive	sixtyfour			
10	1	1/2/2019 13:50	Mortgages View Conv desktop	conversion	daft.ie	728x90	male	age_twentyfive	thirtyfour			
11	7	1/29/2019 20:03	Mortgages Retargetin desktop	last_touch	researchg	728x90	female	age_twentyfive	thirtyfour			
12	10	1/8/2019 23:28	Mortgages Retargetin mobile	middle_toi	whattoexp	300x250	female	age_twentyfive	thirtyfour			
13	9	1/10/2019 20:14	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
14	4	1/6/2019 18:15	Mortgages Prospectin desktop	last_touch	daft.ie	728x90	male	age_fiftyfive	sixtyfour			
15	9	1/5/2019 14:18	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
16	9	1/13/2019 12:27	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
17	9	1/6/2019 17:36	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
18	5	1/5/2019 22:24	Mortgages Retargetin desktop	middle_toi	b.cpx.cz	728x90	male	age_twentyfive	thirtyfour			
19	9	1/6/2019 0:02	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
20	9	1/16/2019 23:21	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
21	5	1/1/2019 19:05	Mortgages Retargetin desktop	first_touch	skyscanner	300x250	male	age_twentyfive	thirtyfour			
22	9	1/18/2019 20:48	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
23	9	1/6/2019 0:02	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
24	9	1/15/2019 18:37	Mortgages Prospectin mobile	middle_toi	stolen-she	300x250	female	age_fortyfive	fiftyfour			
25	1	1/2/2019 12:53	Mortgages Prospectin desktop	last_touch	daft.ie	728x90	male	age_twentyfive	thirtyfour			
26	9	1/4/2019 22:11	Mortgages Prospectin mobile	middle_toi	stolen-she	320x50	female	age_fortyfive	fiftyfour			
27	3	1/16/2019 17:08	Mortgages View Conv mobile	conversion	mumsnet.i	300x250	male	age_twentyfive	thirtyfour			
28	4	1/10/2019 17:09	Mortgages Prospectin desktop	middle_toi	daft.ie	728x90	male	age_fiftyfive	sixtyfour			
29	10	1/7/2019 22:25	Mortgages Retargetin mobile	first_touch	property.i	728x90	female	age_twentyfive	thirtyfour			
30	5	1/6/2019 15:22	Mortgages Retargetin desktop	middle_toi	tpc.google	300x250	male	age_twentyfive	thirtyfour			
31	8	1/16/2019 18:18	Mortgages View Conv mobile	conversion	despreco	160x600	female	age_twentyfive	thirtyfour			

2. Click on “custom sort.” The sort function is located here. Always opt for “custom sort.” Otherwise, Excel will choose a sorting column and method for you, which may not be the one you want.



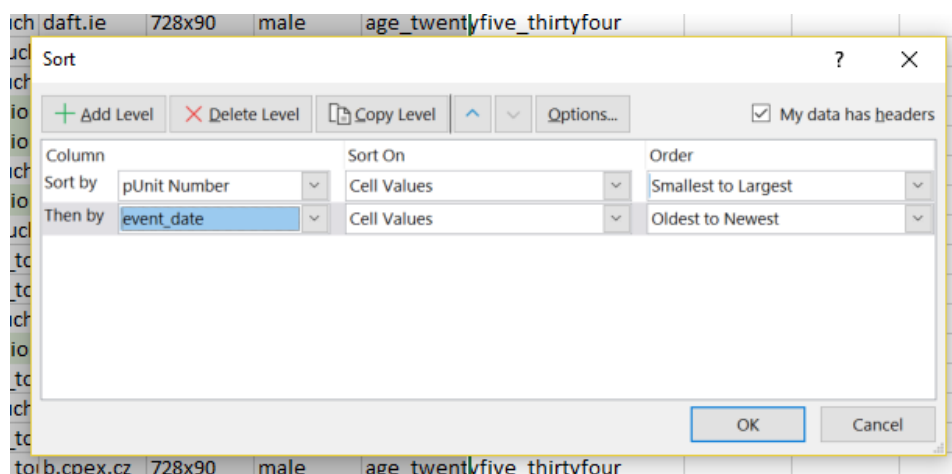
The screenshot shows the Excel ribbon with the 'Sort & Filter' dropdown menu open. The 'Custom Sort...' option is highlighted. The background shows a portion of the data table with columns A through L.

- Click on “my data has headers” if you’ve selected the heading row like I have above.
Sort by pUnit_Number.



- The data should now look like this. Note that the event categories (first touch, middle touch, et cetera) are out of order. Let’s fix this while holding cookie order constant. Select all data, go back to custom sort, then click “add level” and select “event_date.” This will tell Excel to first sort by cookie, then sort by cookie number then sort by event_date.

pUnit	event_date	Campaign	Targeting	event_dev	event_cate	event_don	creativesiz	gender	age_group
1	1/2/2019 13:50	Mortgages	View Conv	desktop	conversion	daft.ie	728x90	male	age_twentyfive_thirtyfo
1	1/2/2019 12:53	Mortgages	Prospectin	desktop	last_touch	daft.ie	728x90	male	age_twentyfive_thirtyfo
1	1/2/2019 10:54	Mortgages	Prospectin	desktop	first_touch	hoganstan	320x50	male	age_twentyfive_thirtyfo
2	1/14/2019 11:08	Mortgages	Retargetin	desktop	last_touch	daft.ie	728x90	male	age_fortyfive_fiftyfour
2	1/14/2019 11:13	Mortgages	View Conv	desktop	conversion	daft.ie	728x90	male	age_fortyfive_fiftyfour
3	1/16/2019 17:08	Mortgages	View Conv	mobile	conversion	mumsnet.c	300x250	female	age_twentyfive_thirtyfo
3	1/5/2019 18:38	Mortgages	Prospectin	mobile	last_touch	mumsnet.c	300x250	female	age_twentyfive_thirtyfo
4	1/14/2019 8:25	Mortgages	View Conv	desktop	conversion	daft.ie	728x90	male	age_fiftyfive_sixtyfour
4	1/6/2019 18:15	Mortgages	Prospectin	desktop	first_touch	daft.ie	728x90	male	age_fiftyfive_sixtyfour
4	1/10/2019 17:09	Mortgages	Prospectin	desktop	middle_to	daft.ie	728x90	male	age_fiftyfive_sixtyfour
4	1/10/2019 17:12	Mortgages	Prospectin	desktop	middle_to	daft.ie	728x90	male	age_fiftyfive_sixtyfour
4	1/11/2019 8:06	Mortgages	Prospectin	desktop	last_touch	daft.ie	728x90	male	age_fiftyfive_sixtyfour



5. Check to make sure your heading is still in the first row. If not, something went awry when you sorted. You should hit CTRL Z (undo) and try again. Most likely, you did not tick the box “my data has headers.”

Computing numeric day of the week using the WEEKDAY Function

It may be useful to know which day of the week certain events occur on. Excel makes this easy with the **WEEKDAY** function. *More information on this function [here](#).*

1. Insert a new column C and name it “weekday numeric”
2. Enter the following formula: =WEEKDAY(B2). This function returns a number that represents day of the week according to the following convention:

Numeric	Written
1	Sunday
2	Monday
3	Tuesday
4	Wednesday
5	Thursday
6	Friday
7	Saturday

You may notice that you don’t get a number at all, but rather, a weirdly formatted date. That’s because Excel is confused about what the cell’s formatting should be. Change formatting from “custom” to “number.” Click on the “00s” button to remove any decimals. See red boxes below.

The screenshot shows the Excel interface with the 'Home' tab selected. The ribbon includes sections for Clipboard, Font, Alignment, and Number. In the Number section, the 'Custom' dropdown is highlighted with a red box, and the '00s' button is also highlighted with a red box. Below the ribbon, the formula bar shows the formula '=WEEKDAY(B2)' in cell C2. The spreadsheet below shows a table with columns A through I. Column C is labeled 'Weekday' and contains the value '1/4/1900 0:00' for row 2, which is highlighted with a green border. The table data is as follows:

	A	B	C	D	E	F	G	H	I
1	pUnit Num	event_date	Weekday	Campaign	Targeting Type	event_dev	event_cate	event_don	creativesiz ge
2	1	1/2/2019 10:54	1/4/1900 0:00	Mortgages	Prospecting	desktop	first_touch	hoganstan	320x50 m
3	1	1/2/2019 12:53		Mortgages	Prospecting	desktop	last_touch	daft.ie	728x90 m
4	1	1/2/2019 13:50		Mortgages	View Conversion	desktop	conversion	daft.ie	728x90 m

3. Drag the formula all the way down, or hover your mouse in the bottom right of the cell until a black cross appears, then double click.

VLOOKUP Function: Computing *written* day of the week

VLOOKUP allows you to look up and return data that matches an index you specify. For example, our variable “weekdays” is in the form of 1,2,3,4,5,6,7. I’d prefer to have this information in *string* form, that is written out as “Monday,” “Tuesday,” et cetera. With VLOOKUP, I can ask Excel to look in the table below and return the written word that corresponds to the numeric day of the week. Below, I’ll show you how.

Numeric	Written
1	Sunday
2	Monday
3	Tuesday
4	Wednesday
5	Thursday
6	Friday
7	Saturday

While there are other ways to solve this conundrum, I’m introducing VLOOKUP because it is a crucial one for you to know. VLOOKUP is the function I use more than any other in excel. It is also a function that I saw tested via an assignment for a job application. It’s worth the effort to learn it, even though it’s fussy and can be frustrating. So let’s go:

1. Create a new column D and name it “Weekday.”
2. Add a new tab called “weekday.”
3. In the new tab “weekday,” enter the following legend

	A	B
1	Numeric	Written
2	1	Sunday
3	2	Monday
4	3	Tuesday
5	4	Wednesday
6	5	Thursday
7	6	Friday
8	7	Saturday
9		

4. “Now return to the tab Data Dive Practice.” In cell D, we’ll enter the following formula:

=VLOOKUP(value,table,col_index,[range_lookup])

where

value= The value you’d like to look up

table=The range of data you'd like to look in
col_index=The column of the value you would like to return.
[range_lookup]= TRUE is an approximate match and FALSE is exact match. Use FALSE.

[Help I don't get this function!](#)

[Help, I'm always getting errors with VLOOKUP ARRRHG!](#)

[23 Things You Should Know about VLOOKUP](#)

VLOOKUP can be a fussy function. I've linked a nice troubleshooting guide above. Here's what the formula in column D will look like:

=VLOOKUP(C2,Weekday!\$A\$1:\$B\$8,2,FALSE)

Let's break this down.

C2— this is the value I'd like to look up. We are going to drag this formula down, and because there are no "\$", the formula will update as we drag (e.g., C3, C4, C5, et cetera)

Weekday!\$A\$1:\$B\$8—the first piece Weekdays! tells excel we're going to a new tab. \$A\$1:\$B\$7 is the array of data that has the values I'm indexing on (i.e., the numeric day of the week) and the value I would like to be returned if excel finds a match. The "\$" locks the array. This means that when I drag the formula down (or sideways), the array I've selected remains the same. This is critical so that when I drag the formula down, the reference array stays the same. There is a \$ in front of the letter to lock the columns, and a \$ in front of the number to lock the rows. It is possible to lock just one, but don't worry about that right now.

2 – This tells excel that I when it finds a match between the two arrays, I want the value in column 2 (i.e., Column B), returned. Excel always counts the left most column *in the array* as "1". So even if your table starts in column C, this formula treats it as a 1 and column D as 2, et cetera.

FALSE – This tells excel to return an exact match.

Conceptually, here's what is happening.

STEP 2

Weekday!\$A\$1:\$B\$8

C
Weekday Numeric
4
4
4
2
2
7
4
1
5
5
6
2
3

STEP 1 (C2)

Excel is going to look for "4" wherever we tell it to (in this case, the weekdays tab)

Excel looks in this table for a 4 and finds it!

1	Sunday
2	Monday
3	Tuesday
4	Wednesday
5	Thursday
6	Friday
7	Saturday

Mortgages - Raw	Mortgages - Conversions	Mortgages TTC	Mortgages Touch Points	Data Dive Practice	Weekday
-----------------	-------------------------	---------------	------------------------	--------------------	---------

STEP 3

2,FALSE

"Column 1" "Column 2"

	A	B
1	Numeric	Written
2	1	Sunday
3	2	Monday
4	3	Tuesday
5	4	Wednesday
6	5	Thursday
7	6	Friday
8	7	Saturday
9		

Excel returns whatever value is in column 2 (which we specified in the formula with a "2") The FALSE tells Excel to only return a value if there is an exact match versus a partial match (e.g., "4" not "44" or "14" et cetera)

The result:

D2

⌵

⌵

✖

✔

fx

=VLOOKUP(\$C\$2:\$C\$52,Weekday!\$A\$1:\$B\$8,2,FALSE)

	A	B	C	D	E	F	G	H	I
1	pUnit Num	event_date	Weekday Numeric	Weekday	Campaign	Targeting Type	event_dev	event_cat	event_domain
2	1	1/2/2019 10:54	4	Wednesday	Mortgages	Prospecting	desktop	first_touch	hoganstan
3	1	1/2/2019 12:53	4		Mortgages	Prospecting	desktop	last_touch	daft.ie
4	1	1/2/2019 13:50	4		Mortgages	View Conversion	desktop	conversion	daft.ie
5	2	1/14/2019 11:08	2		Mortgages	Retargeting	desktop	last_touch	daft.ie
6	2	1/14/2019 11:13	2		Mortgages	View Conversion	desktop	conversion	daft.ie
7	3	1/5/2019 18:38	7		Mortgages	Prospecting	mobile	last_touch	mumsnet
8	3	1/16/2019 17:08	4		Mortgages	View Conversion	mobile	conversion	mumsnet
9	4	1/6/2019 18:15	1		Mortgages	Prospecting	desktop	first_touch	daft.ie
10	4	1/10/2019 17:09	5		Mortgages	Prospecting	desktop	middle_touch	daft.ie

5. Now drag the formula down to populate all cells.

VLOOKUP Practice with age

Let's try another VLOOKUP example, this time with age. I find it confusing to have "age" written out. Let's change it to numeric.

age_twentyfive_thirtyfour	25_34
age_fortyfive_fiftyfour	45_54
age_fiftyfive_sixtyfour	55_64
age_thirtyfive_fortyfour	35_44
age_eighteen_twentyfour	18_24
age_sixtyfive_plus	65

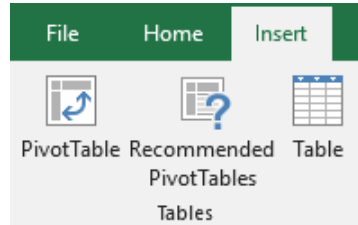
See if you can get the following results. If not, raise your hand and I'll come help.

L	M	
age_group	Age_Numeric	
age_twent	25_34	
age_twent	25_34	
age_twent	25_34	
age_fortyfi	45_54	
age_fortyfi	45_54	
age_twent	25_34	
age_twent	25_34	
age_fiftyfi	55_64	
age_fiftyfi	55_64	
age_fiftyfi	55_64	
age_fiftyfi	55_64	
age_fiftyfi	55_64	
age_twent	25_34	
age_twent	25_34	

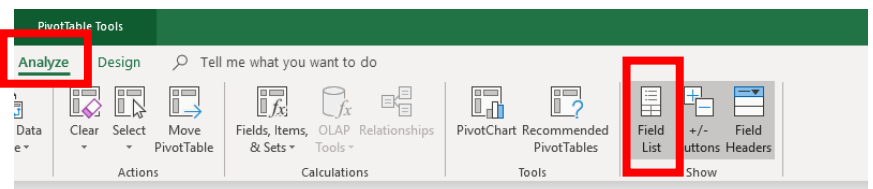
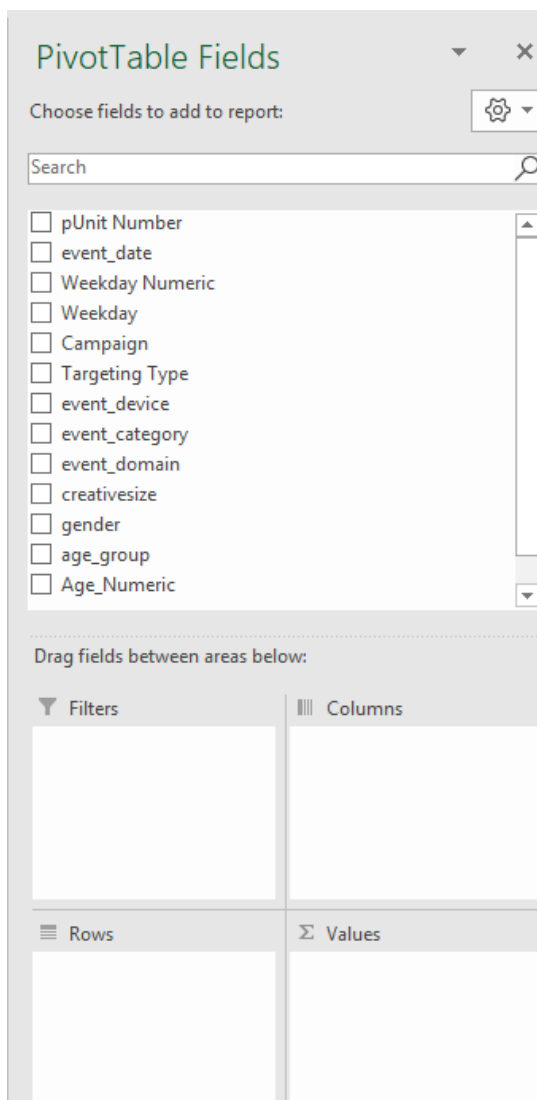
PIVOT Tables: Example 1 (gender x conversions)

Pivot tables allow us to create easy cross-tabs on the data. First, select all data as you do when you sort.

1. Go to the “insert” tab and select “pivot table”



2. Select “new worksheet” as the destination for the Pivot table then press OK. Name this new tab “Pivot 1.”
3. You should see a box called “Pivot Tables Fields.” If at any point this box disappears, you can get it back by clicking “analyze” then “field list”



4. Let's build a simple table of demographics x conversion.
 - a. Drag "gender" into "rows"
 - b. Drag "gender" into "values"
 - c. Drag "event_category" into filters

You'll get a table like the one below. This shows the number of events that occurred for females and males (e.g., first touch, last touch, conversion, et cetera). Now, we only want to look at conversions to see if one gender converts more often.

event_category (All)	
Row Labels	Count of gender
female	28
male	23
Grand Total	51

- d. Since we only want to look at conversions, click on "event_category (All)" and select "conversion." Your table should look like this one:

event_category conversion	
Row Labels	Count of gender
female	5
male	5
Grand Total	10

PIVOT Tables: Example 2 (Number of touch points before conversion)

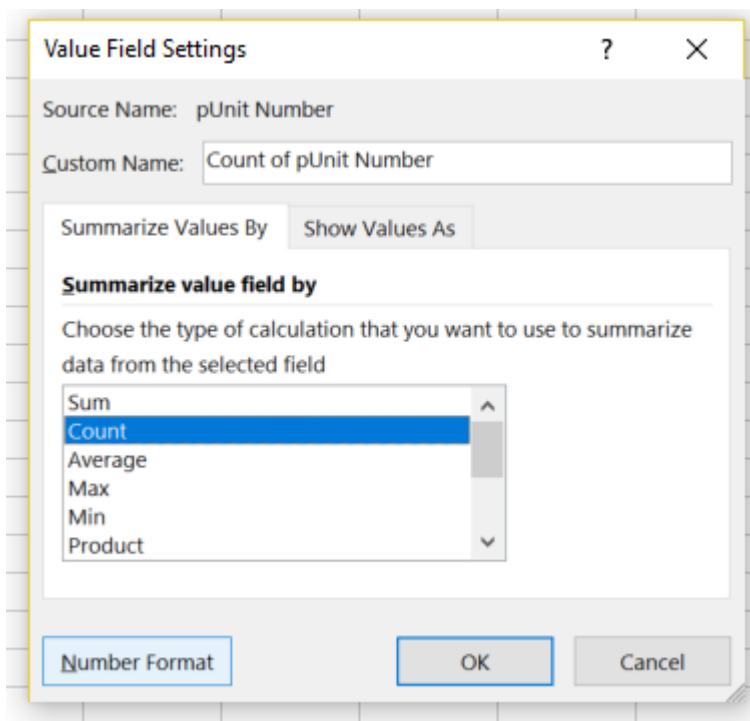
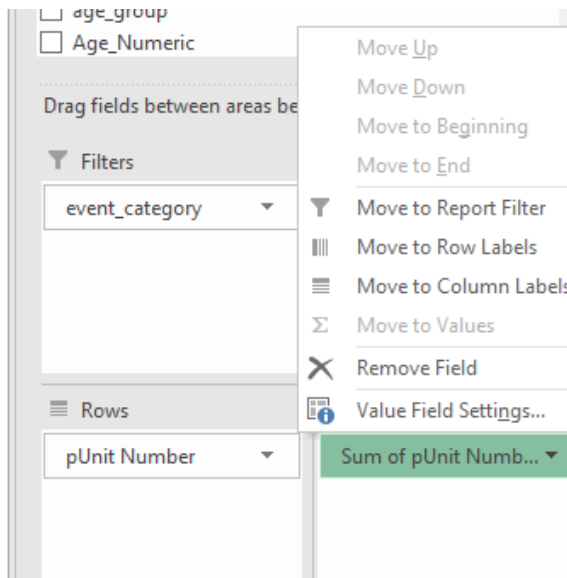
I'd like to know how many impressions (i.e., ads or "touch points") people saw before they converted.

1. Insert a new pivot table, name the tab "Pivot 2"
2. Drag "pUnit" into rows, "event_category" into filters, and "pUnit" into values. Your table should look like this:

event_category (All)	
Row Labels	Sum of pUnit Number
1	3
2	4
3	6
4	20
5	55
6	12
7	14
8	16
9	153
10	50
Grand Total	222

Note that Excel defaults to *summing* the variable pUnit.⁴ That doesn't make sense for what we're interested in knowing. We want to know the number of touchpoints before conversion. So we need to do two things.

3. We need to change the “sum” to “count.” Click on pUnit in the Values box, then “Value Field Settings.” Change “sum” to “count”



⁴ Any time you enter a numeric variable into the Values box, Excel will default to summing that variable. Oftentimes, you DO NOT WANT THE SUM. Often, you'll want an average or a count. Be sure to keep a close eye on this issue when you are doing your pivot tables.

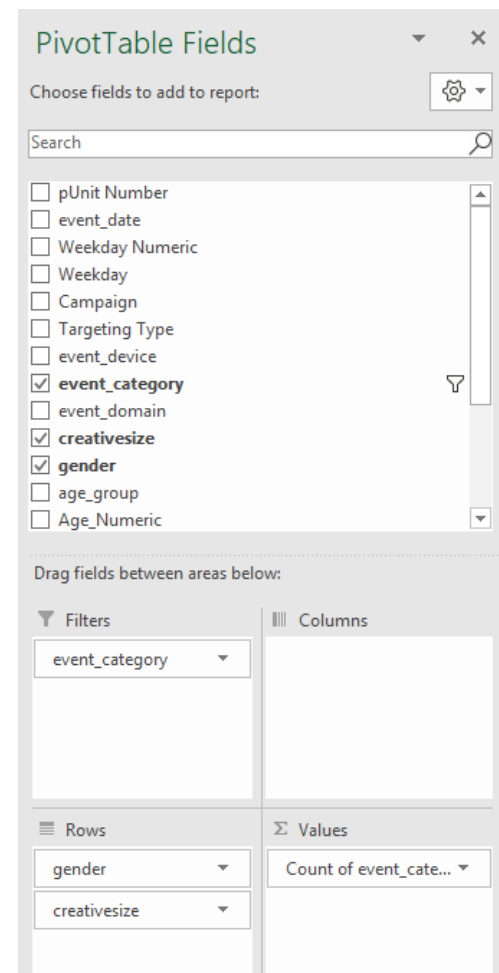
PIVOT Tables: Example 3 (Nesting)

Suppose you would like to know whether males and females react differently to ad size. To begin examining this, we could create a pivot table that nests creative size under gender:

event_category conversion	
Row Labels	Count of event_category
female	5
160x600	1
300x250	2
320x50	1
728x90	1
male	5
300x250	2
728x90	3
Grand Total	10

To build this table you will:

1. Select all data in tab "data dive practice" and create a pivot table in a new tab
2. Label this new tab "Pivot 3"
3. Drag "gender" into rows
4. Drag "creativesize" into rows *below gender*
5. Drag "event category" into "values"
6. Drag "event category" into "filters"
7. Filter event_category by "conversions"



Basic IF statements

First, I want to show you how a basic IF statement works.

=IF (logical_test, [value_if_true], [value_if_false])

where

logical_test = Some statement that can be true or false (e.g., 10>2?)

value_if_true=The value to return when true

value_if_false=The value to return if false

[Help I don't get this function!](#)

Let's do a simple one. I want to compute a new column that is "yes" if the ad was shown on daft.ie and a 0 if not. In column N, you'll write the following function:

=IF(I2="daft.ie","yes",0)

Note that any text is in quotations, while numbers are not.

Your results should look like this:

I	J	K	L	M	N
event_don	creativesiz	gender	age_group	Age_Num	Daft.ie?
hoganstan	320x50	male	age_twent	25_34	0
daft.ie	728x90	male	age_twent	25_34	yes
daft.ie	728x90	male	age_twent	25_34	yes
daft.ie	728x90	male	age_fortyf	45_54	yes
daft.ie	728x90	male	age_fortyf	45_54	yes
mumsnet.c	300x250	female	age_twent	25_34	0
mumsnet.c	300x250	female	age_twent	25_34	0
daft.ie	728x90	male	age_fiftyfi	55_64	yes
daft.ie	728x90	male	age_fiftyfi	55_64	yes
daft.ie	728x90	male	age_fiftyfi	55_64	yes

I realize this column isn't super useful with "yes" and 0. But I wanted to make the point that string variables (i.e., non-numeric data) always needs to be in quotes. ALWAYS.

Nested IF statements: Example 1 (weekend or not?)

While Vlookup is my favorite function, nested IF statements are a close second. A nested function is when there is a function wrapped in a function. Let's do a simple one. I want to create a new variable that is coded "Weekend" if the event date occurs on a weekend and "Weekday" if it does not. To do this, we will nest [the "OR" function](#).

In column O, write the following function:

=IF(OR(D2="Saturday",D2="Sunday"),"Weekend","Weekday")

The “or” statement is nested in the logical test portion of the IF statement. Your results should look like this:

D	E	F	G	H	I	J	K	L	M	N	O
Weekday	Campaign	Targeting Type	event_dev	event_cat	event_don	creativesiz	gender	age_group	Age_Num	Daft.ie?	Weekend?
Wednesday	Mortgages	Prospecting	desktop	first_touch	hoganstan	320x50	male	age_twent	25_34	0	Weekday
Wednesday	Mortgages	Prospecting	desktop	last_touch	daft.ie	728x90	male	age_twent	25_34	yes	Weekday
Wednesday	Mortgages	View Conversion	desktop	conversion	daft.ie	728x90	male	age_twent	25_34	yes	Weekday
Monday	Mortgages	Retargeting	desktop	last_touch	daft.ie	728x90	male	age_fortyfi	45_54	yes	Weekday
Monday	Mortgages	View Conversion	desktop	conversion	daft.ie	728x90	male	age_fortyfi	45_54	yes	Weekday
Saturday	Mortgages	Prospecting	mobile	last_touch	mumsnet.c	300x250	female	age_twent	25_34	0	Weekend
Wednesday	Mortgages	View Conversion	mobile	conversion	mumsnet.c	300x250	female	age_twent	25_34	0	Weekday
Sunday	Mortgages	Prospecting	desktop	first_touch	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekend
Thursday	Mortgages	Prospecting	desktop	middle_to	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekday
Thursday	Mortgages	Prospecting	desktop	middle_to	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekday
Friday	Mortgages	Prospecting	desktop	last_touch	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekday
Monday	Mortgages	View Conversion	desktop	conversion	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekday
Tuesday	Mortgages	Retargeting	desktop	first_touch	skyscanner	300x250	male	age_twent	25_34	0	Weekday
Tuesday	Mortgages	Retargeting	desktop	middle_to	skyscanner	300x250	male	age_twent	25_34	0	Weekday
Tuesday	Mortgages	Retargeting	desktop	middle_to	skyscanner	300x250	male	age_twent	25_34	0	Weekday
Tuesday	Mortgages	Retargeting	desktop	middle_to	daft.ie	160x600	male	age_twent	25_34	yes	Weekday

Advanced Nested IF statements: Example 2 (morning, afternoon, or evening?)

I would like to create a variable that specifies whether an ad or conversion happened in the morning, in the afternoon, or at night.

1. Create a new column called “Time”
2. Use the “hour” function to return the hour of the day from “event_date” =HOUR(B2)
3. Create a new column called “Time of Day”
4. Determine your cutoffs for morning, afternoon, and night. I’ll propose the following:
 - a. Morning: 6 – 11
 - b. Afternoon: 12 – 4
 - c. Evening: 5 +
5. Time for a nested if function!

=IF(AND(P2>=6,P2<=11),"Morning",IF(AND(P2>=12,P2<=16),"Afternoon","Evening"))

The function first asks whether P2 is greater than or equal to 6 *and* less than 11. If so, then the function should return the word “Morning.” If not, Excel should execute a second IF statement that asks whether P2 is greater than or equal to 12 and less than or equal to 16. If so, Excel should return the word “Afternoon.” Otherwise, Excel should return the word “Evening.”

H	I	J	K	L	M	N	O	P	Q
event_cat	event_don	creativesiz	gender	age_group	Age_Num	Daft.ie?	Weekend?	Time?	Time of Day
first_touch	hoganstan	320x50	male	age_twent	25_34	0	Weekday	10	Morning
ast_touch	daft.ie	728x90	male	age_twent	25_34	yes	Weekday	12	Afternoon
onversion	daft.ie	728x90	male	age_twent	25_34	yes	Weekday	13	Afternoon

All-Around Advanced: Nested IF statements + filtering + VLOOKUP :)

I would like to create a new variable that denotes whether a cookie was exposed to retargeting at any point before conversion. Here is one way to do this.

1. Put a filter on the data.

The screenshot shows an Excel spreadsheet with a filter applied to column F, 'Targeting Type'. The filter dropdown is open, showing options like 'All', 'Retargeting', etc. The spreadsheet data includes columns for event details, targeting type, and conversion status.

2. Filter Column F (Targeting Type) to contain only “Retargeting.”

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	pUnit Num	event_date	Weekday Numeric	Weekday	Campaign	Targeting Type	event_d	event_c	event_d	creative	gender	age_gro	Age Nu
5	2	1/14/2019 11:08	2	Monday	Mortgages Retargeting	desktop	last_touch	daft.ie	728x90	male	age_twenty	45_54	
14	5	1/1/2019 19:05	3	Tuesday	Mortgages Retargeting	desktop	first_touch	skyscanner	300x250	male	age_twenty	25_34	
15	5	1/1/2019 19:07	3	Tuesday	Mortgages Retargeting	desktop	middle_toi	skyscanner	300x250	male	age_twenty	25_34	
16	5	1/1/2019 19:10	3	Tuesday	Mortgages Retargeting	desktop	middle_toi	skyscanner	300x250	male	age_twenty	25_34	
17	5	1/1/2019 20:54	3	Tuesday	Mortgages Retargeting	desktop	middle_toi	daft.ie	160x600	male	age_twenty	25_34	
18	5	1/5/2019 22:24	7	Saturday	Mortgages Retargeting	desktop	middle_toi	csfd.cz	160x600	male	age_twenty	25_34	
19	5	1/5/2019 22:24	7	Saturday	Mortgages Retargeting	desktop	middle_toi	b.cpx.cz	728x90	male	age_twenty	25_34	
20	5	1/6/2019 15:22	1	Sunday	Mortgages Retargeting	desktop	middle_toi	tpc.google	300x250	male	age_twenty	25_34	
21	5	1/6/2019 15:51	1	Sunday	Mortgages Retargeting	desktop	middle_toi	doneDeal	728x90	male	age_twenty	25_34	
22	5	1/6/2019 15:52	1	Sunday	Mortgages Retargeting	desktop	middle_toi	doneDeal	728x90	male	age_twenty	25_34	
23	5	1/6/2019 15:58	1	Sunday	Mortgages Retargeting	desktop	last_touch	ebay.ie	300x250	male	age_twenty	25_34	
27	7	1/29/2019 20:03	3	Tuesday	Mortgages Retargeting	desktop	last_touch	researchga	728x90	female	age_twenty	25_34	
48	10	1/7/2019 22:25	2	Monday	Mortgages Retargeting	mobile	first_touch	property.ie	728x90	female	age_twenty	25_34	
49	10	1/7/2019 22:27	2	Monday	Mortgages Retargeting	mobile	middle_toi	property.ie	728x90	female	age_twenty	25_34	
50	10	1/8/2019 23:28	3	Tuesday	Mortgages Retargeting	mobile	middle_toi	whattoexp	300x250	female	age_twenty	25_34	
51	10	1/9/2019 4:40	4	Wednesday	Mortgages Retargeting	mobile	last_touch	virginmedi	300x250	female	age_twenty	25_34	
53													

3. Copy and paste the pUnit column into a new tab. This will give us a list of cookies that were at some point retargeted. Call this tab “retargeting”
4. In column B, type “yes” for each cookie. We are going to use this data as part of a VLookup Function

pUnit Num	Retargeted?
2	Yes
5	Yes
5	Yes
5	Yes
5	Yes
5	Yes
5	Yes
5	Yes
5	Yes
7	Yes
10	Yes
10	Yes
10	Yes
10	Yes

- Return to the tab “data dive practice” and unfilter the data. Name column R “Contained Retargeting?”
- In column R, I want to denote *for each conversion* whether the cookie was retargeted at any point. This means that I only want data to appear in the green shaded rows (where a conversion happened). To do this, we are going to nest a VLookup function inside an IF statement. It will look like this:

=IF(H2="conversion",VLOOKUP(A2,Retargeting!\$A\$2:\$B\$17,2,FALSE),""))

This formula says that if event_category = “conversion,” then go to the Retargeting tab and look up the corresponding cookie, then return whatever is in the 2nd column (in this case, the “yes”). If event_category does *not* = “conversion”, return nothing (i.e., “”).

Your data should look like this:

G	H	I	J	K	L	M	N	O	P	Q	R	S
event_d	event_c	event_d	creative	gender	age_gro	Age_Nu	Daft.ie	Weeker	Time?	Time of	Contained Retargeting	
desktop	first_touch	hoganstan	320x50	male	age_twent	25_34	0	Weekday	10	Morning		
desktop	last_touch	daft.ie	728x90	male	age_twent	25_34	yes	Weekday	12	Afternoon		
desktop	conversion	daft.ie	728x90	male	age_twent	25_34	yes	Weekday	13	Afternoon	#N/A	
desktop	last_touch	daft.ie	728x90	male	age_fortyfi	45_54	yes	Weekday	11	Morning		
desktop	conversion	daft.ie	728x90	male	age_fortyfi	45_54	yes	Weekday	11	Morning	Yes	
mobile	last_touch	mumsnet.	300x250	female	age_twent	25_34	0	Weekend	18	Evening		
mobile	conversion	mumsnet.	300x250	female	age_twent	25_34	0	Weekday	17	Evening	#N/A	
desktop	first_touch	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekend	18	Evening		
desktop	middle_toi	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekday	17	Evening		
desktop	middle_toi	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekday	17	Evening		
desktop	last_touch	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekday	8	Morning		
desktop	conversion	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekday	8	Morning	#N/A	

Now, if you are like me and #N/A makes you feel anxious and unhappy, it’s easy to deal with this. The #N/A is occurring because Excel didn’t find anything in the tab “retargeting” when it looked for the cookie. In other words, the cookie wasn’t there. Why not? Because remember – only cookies that had retargeting are listed in the retargeting tab. This means that any of the #N/A’s conceptually mean “no retargeting. Wouldn’t it be nice if they said that? Enter, the [IFERROR statement](#) (another favorite).

=IFERROR (value, value_if_error)

where

Value – The value or formula you would like to check for an error (e.g., #N/A)

Value_if_error – The value you would like Excel to return instead of the error

- So, let’s amend our formula as follows (new stuff in red):

=**IFERROR**(IF(H2="conversion",VLOOKUP(A2,Retargeting!\$A\$2:\$B\$17,2,FALSE),""),"No")

Now, our data look like this – hurray!

G	H	I	J	K	L	M	N	O	P	Q	R	S
event_d	event_c	event_d	creative	gender	age_gro	Age_Nu	Daft.ie	Weeker	Time?	Time of	Contained	Retargeting?
desktop	first_touch	hoganstan	320x50	male	age_twent	25_34	0	Weekday	10	Morning		
desktop	last_touch	daft.ie	728x90	male	age_twent	25_34	yes	Weekday	12	Afternoon		
desktop	conversion	daft.ie	728x90	male	age_twent	25_34	yes	Weekday	13	Afternoon	No	
desktop	last_touch	daft.ie	728x90	male	age_fortyfi	45_54	yes	Weekday	11	Morning		
desktop	conversion	daft.ie	728x90	male	age_fortyfi	45_54	yes	Weekday	11	Morning	Yes	
mobile	last_touch	mumsnet.	300x250	female	age_twent	25_34	0	Weekend	18	Evening		
mobile	conversion	mumsnet.	300x250	female	age_twent	25_34	0	Weekday	17	Evening	No	
desktop	first_touch	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekend	18	Evening		
desktop	middle_to	daft.ie	728x90	male	age_fiftyfi	55_64	yes	Weekday	17	Evening		

There must be an easier way??

There are SO many ways to do the same thing in Excel. For example, the last exercise (above) could have been done simpler if we had just used the data in the tab “Mortgages – Conversions,” which only contains conversion data. So while the exercises above may not be the simplest way to do things, they give you practice at fundamental skills so that you can build your own formulas and answer your own questions.

A note on descriptive statistics

In order to make your recommendations, you will need to calculate some descriptive statistics for variables of your choosing (e.g., time to convert). These are statistics that summarize the distribution of the data, for example:

Mean: *the average of the data*

Median: *the point at which 50% of the data is above and 50% of the data is below*

Mode: *the data point which appears the most often*

Max: *the highest number in the data set*

Min: *the lowest number in the data set*

If you want more information about this, go [here](#).

Calculating descriptive statistics in Excel

Let's calculate some descriptive statistics for Last Touch TTC. For this, we will use the real data (not the practice data). Go to tab "Mortgages TTC."

Recreate column P in your own dataset.

	J	K	L	M	N	O	P	Q
1	age_group	First Touch Date	First Touch Time	Last Touch Date	Last Touch Time		Descriptives	
2	age_twentyfive_thirtyfour	1/2/2019	0	1/2/2019	0		Mean	
3	age_fortyfive_fiftyfour			1/14/2019	0		Median	
4	age_twentyfive_thirtyfour			1/5/2019	10		Mode	
5	age_fiftyfive_sixtyfour	1/6/2019	7	1/11/2019	3		Min	
6	age_twentyfive_thirtyfour	1/1/2019	5	1/6/2019	0		Max	
7	age_twentyfive_thirtyfour			1/13/2019	8			
8	age_twentyfive_thirtyfour			1/29/2019	2			
9	age_twentyfive_thirtyfour			1/10/2019	6			
10	age_fortyfive_fiftyfour	1/4/2019	15	1/19/2019	0			
11	age_twentyfive_thirtyfour	1/7/2019	1	1/9/2019	0			
12	age_thirtyfive_fortyfour	1/3/2019	5	1/7/2019	1			
13	age_fortyfive_fiftyfour			1/13/2019	0			
14	age_twentyfive_thirtyfour	1/1/2019	13	1/10/2019	4			
15	age_eighteen_twentyfour			1/11/2019	14			
16	age_fortyfive_fiftyfour	1/7/2019	8	1/8/2019	8			
17	age_fortyfive_fiftyfour			1/30/2019	0			
18	age_twentyfive_thirtyfour	1/1/2019	0	1/1/2019	0			
19	age_twentyfive_thirtyfour	1/5/2019	4	1/7/2019	2			
20	age_twentyfive_thirtyfour	1/7/2019	14	1/21/2019	0			
21	age_twentyfive_thirtyfour			1/11/2019	3			
22	age_twentyfive_thirtyfour	1/1/2019	3	1/1/2019	2			

The formulas⁵ will be as follows. Don't copy and paste them from here. Build them yourself.

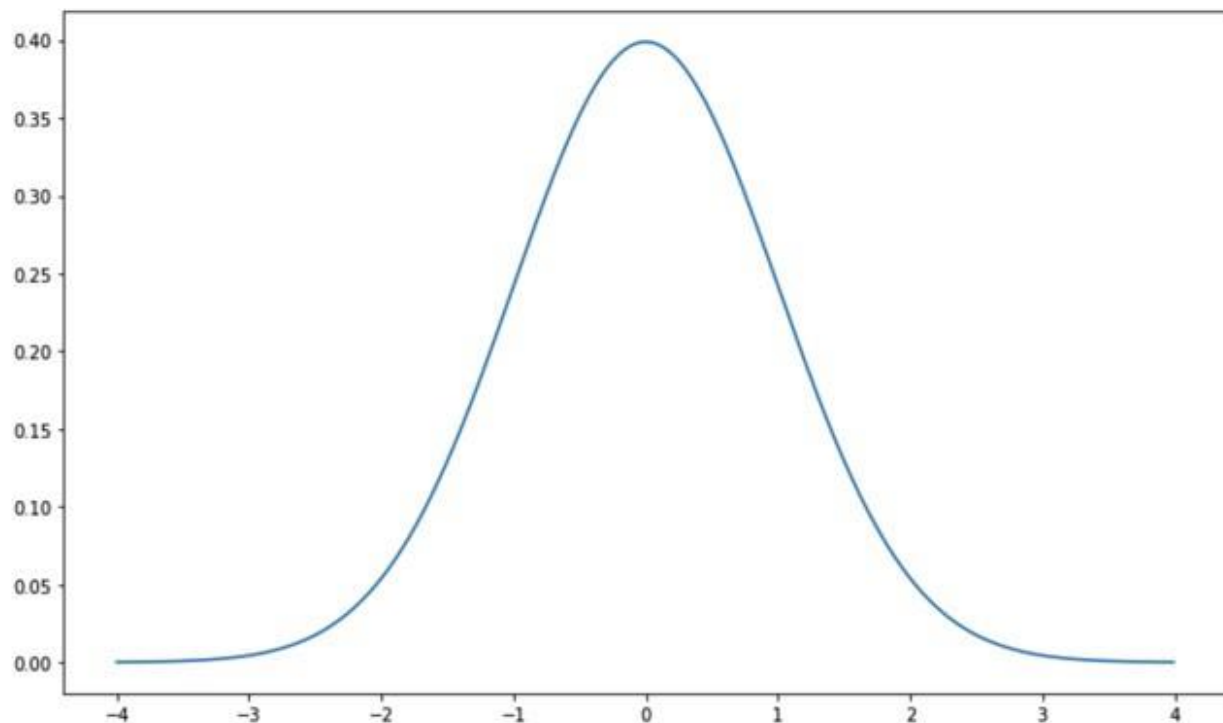
Mean =AVERAGE(N2:N243)
Median =MEDIAN(N2:N243)
Mode =MODE(N2:N243)
Min =MIN(N2:N243)
Max =MAX(N2:N243)

You should get the following answers

Descriptive	
Mean	4.086758
Median	2
Mode	0
Min	0
Max	25

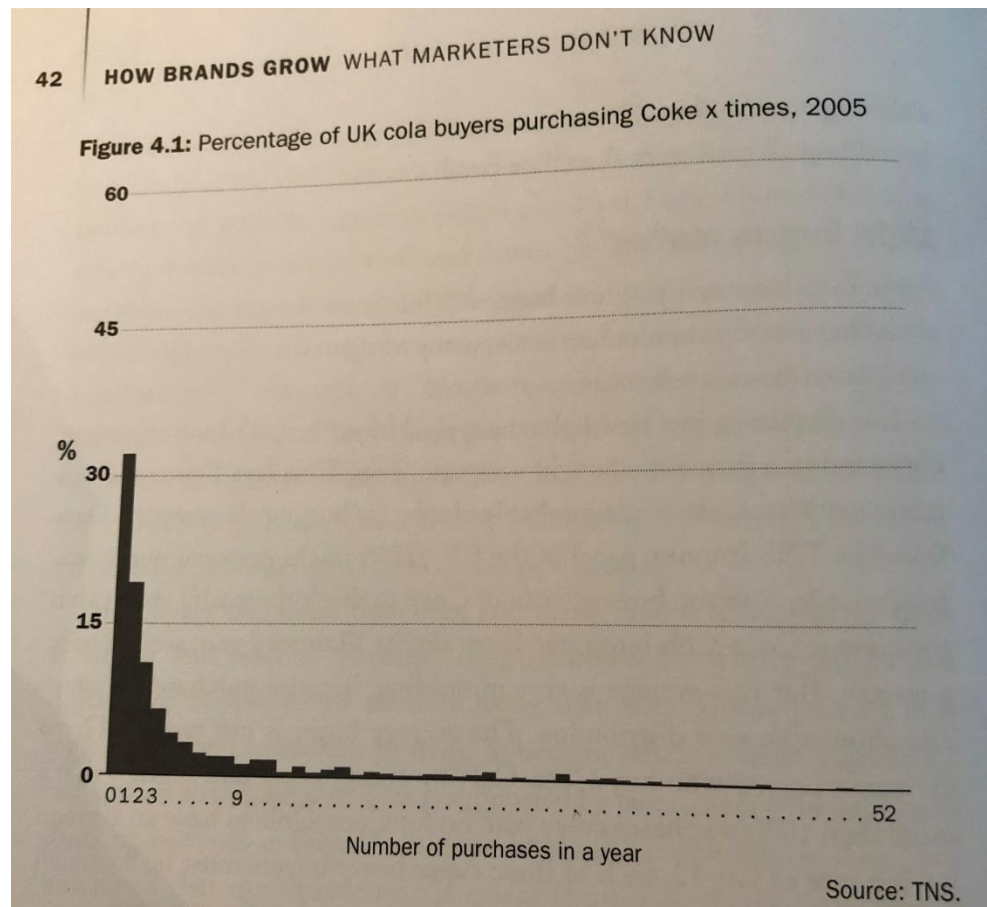
What descriptive(s) are “best”?

This is an important question. You're probably most familiar with the mean. And in many cases, the mean, median, and mode are nearly identical/identical. When, you ask? When data are normally distributed, which looks like this. When data are normally distributed, the mean, median, and mode are all the same number.



⁵ Typo found here and was corrected. Old version stated N9:N243. Should be N2:N2243

However, the mean isn't always the best descriptive for making decisions. Consider the following, for example:



Look familiar? It's a chart from Byron Sharp shown in week 1. (Note – this chart is coming but it will come later class of 2022 😊) On the X axis, we have the number of cokes bought per year. On the Y axis, we have the number of people buying that number of cokes per year. The data is heavily skewed. It does not conform to a normal distribution. Hence, mean, median, and mode will be different.

It's easy to see here that the modal response is 0. That is, most people buy 0 Cokes per year. The bulk of the data is between 0-3 Cokes per year. For skewed data sets, the mean can be quite misleading. If we were to take the mean of this data, it would be around 30. But to focus on the mean here misses the point that *most people only buy a few Cokes per year*. So the "typical" person isn't buying 30 Cokes/year even if 30 is mathematically the average. Make sense? If it doesn't, please raise your hand and I'll come over.

Which descriptive stats would be more useful? Certainly the mode (which is 0). The mode tells us what number is *most common*.

The median would also be helpful, but I would really like to see the data in percentiles. (If you don't know about percentiles, read about them [here](#)). Next, I'll show you how to graph the distribution and calculate the percentiles.

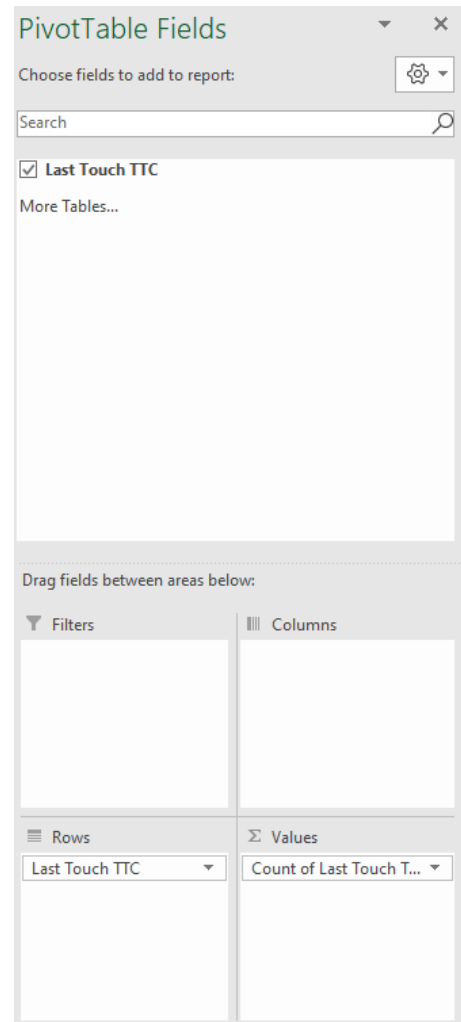
Graphing the distribution

Even though I have the descriptive statistics for last touch time to convert, I would like to see the distribution. To do this, we'll do the following (note, this tutorial shows less steps because I assume you have completed the first tutorial).

Select column containing Last Touch TTC and insert a pivot table (see earlier tutorial if you are having trouble doing this)

Drag Last Touch TTC into "rows" and "values." Change from sum to count. You should see this:

Row Label	Count of Last Touch TTC
0	73
1	33
2	23
3	17
4	5
5	6
6	9
7	6
8	4
9	7
10	11
11	3
12	2
13	2
14	4
15	1
17	1
18	4
19	2
20	2
21	1
22	1
25	2
(blank)	
Grand Total	219



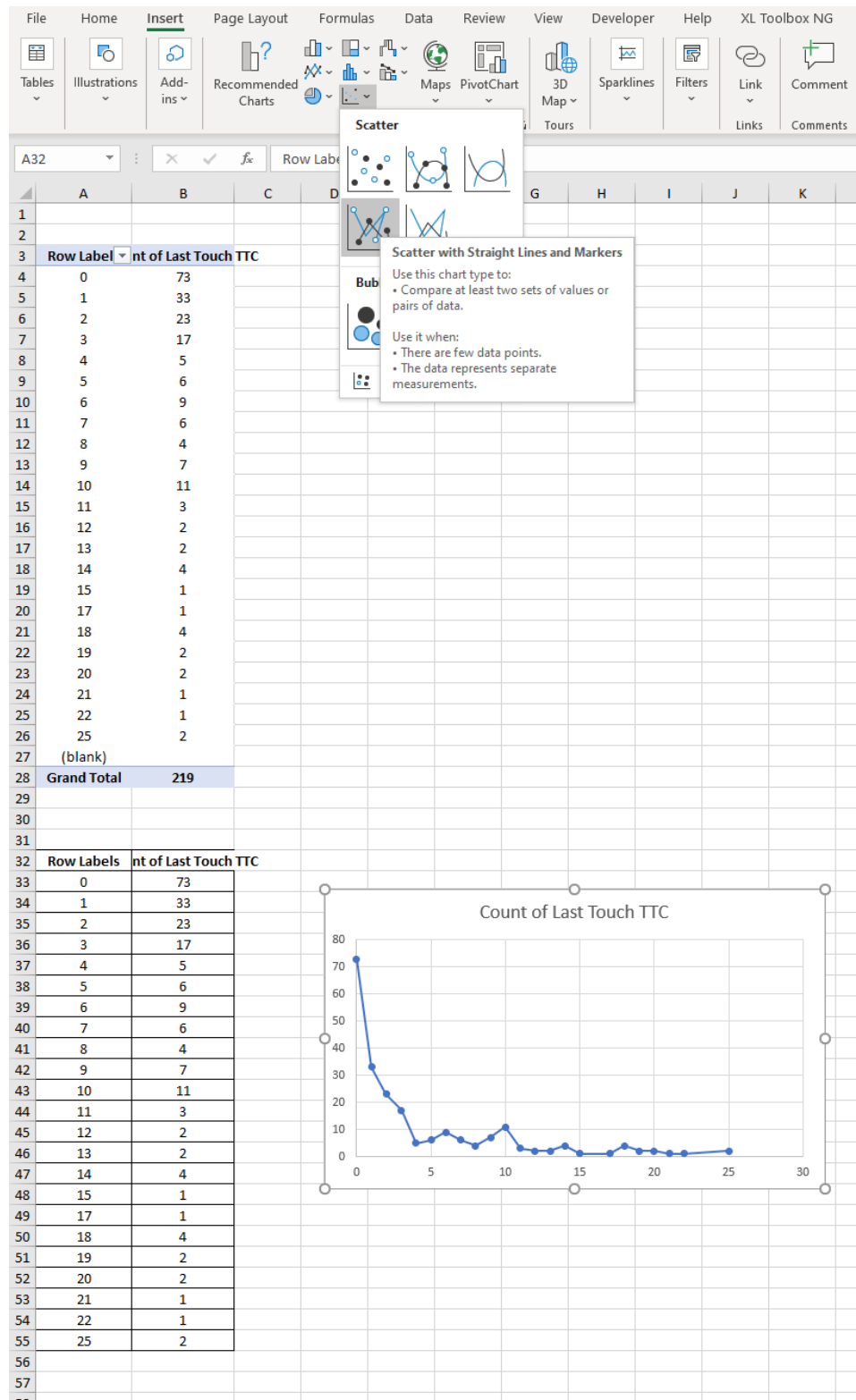
Here is how we can interpret this:

The number of days from the last touch to conversion. Some people converted in 0 days, some people converted after 1 day, et cetera all the way up to 25 days (that matches our “max” calculation from before)

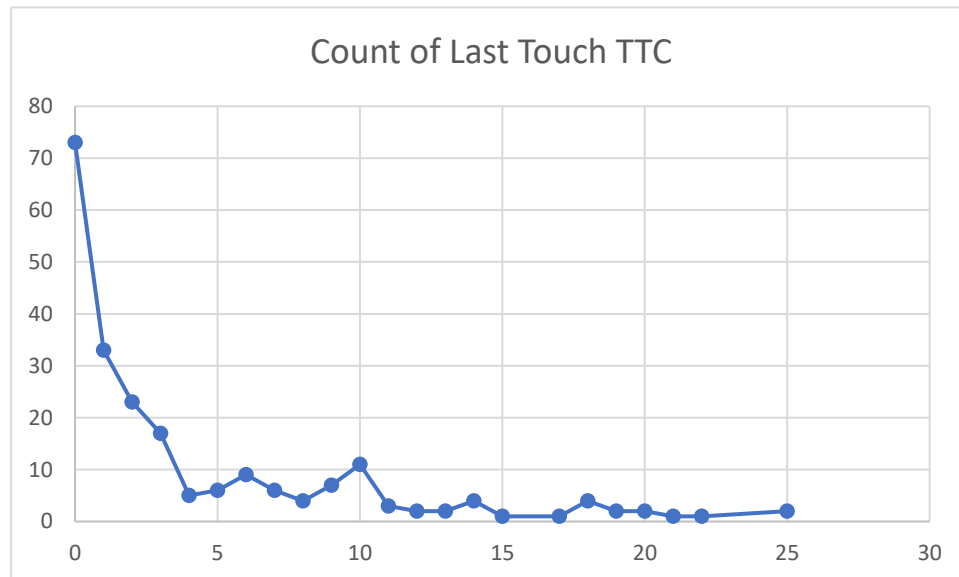
Row Label	Count of Last Touch TTC
0	73
1	33
2	23
3	17
4	5
5	6
6	9
7	6
8	4
9	7
10	11
11	3
12	2
13	2
14	4
15	1
17	1
18	4
19	2
20	2
21	1
22	1
25	2
(blank)	
Grand Total	219

How many people converted for each span of days. For example, 73 people converted in 0 days. 33 took 1 day to convert. 23 took 2 days to convert, et cetera.

What I would really like to see is a distribution. To do this, copy and paste the data in a blank cell. Make sure it pastes as raw data, not as a pivot table (it just makes it easier to work with). Highlight the data and go to insert, then select scatterplot



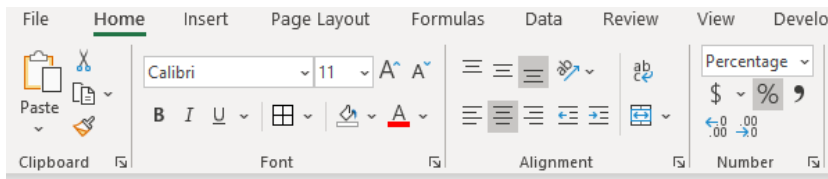
Here it is in all its glory. Now you can do fun cosmetic things to it, if you wish. See some tips [here](#).



Calculating percentiles

It would be nice to know what percentage of people converted in 0 days, 1 day, 2 days, et cetera. How can we do this? It's easy. Let's first just calculate some percents. Type in the following function below. Don't forget to put in the \$ because we are going to drag this formula down. Can you figure out where they need to be added in the below formula?

Row Labels	Int of Last Touch TTC	
0	73	=B33/sum(B33:B55)
1	33	
2	23	
3	17	
4	5	
5	6	
6	9	
7	6	
8	4	
9	7	
10	11	
11	3	
12	2	
13	2	
14	4	
15	1	
17	1	
18	4	
19	2	
20	2	
21	1	
22	1	
25	2	



Change calculation to percent

Last Touch TTC	Count of Last Touch TTC	Percents
0	73	33%

Drag the formula down.

Last Touch TTC	Count of Last Touch TTC	Percentages
0	73	33%
1	33	15%
2	23	11%
3	17	8%
4	5	2%
5	6	3%
6	9	4%
7	6	3%
8	4	2%
9	7	3%
10	11	5%
11	3	1%
12	2	1%
13	2	1%
14	4	2%
15	1	0%
17	1	0%
18	4	2%
19	2	1%
20	2	1%
21	1	0%
22	1	0%
25	2	1%

So this tells us that 33% of people converted in 0 days since the last touch, 15% converted in 1 day from the last touch, et cetera.

To calculate the cumulative percentage (i.e., percentiles), copy the first percentage (33%) into the top row. Then, create the following formula. You will then drag this down.

Last Touch TTC	Count of Last Touch TTC	Percents	Cumulative Percent
0	73	33%	33%
1	33	15%	=C34+D33
2	23	11%	
3	17	8%	
4	5	2%	
5	6	3%	
6	9	4%	
7	6	3%	
8	4	2%	
9	7	3%	
10	11	5%	
11	3	1%	
12	2	1%	
13	2	1%	
14	4	2%	
15	1	0%	
17	1	0%	
18	4	2%	
19	2	1%	
20	2	1%	
21	1	0%	
22	1	0%	
25	2	1%	

Your results should look like this:

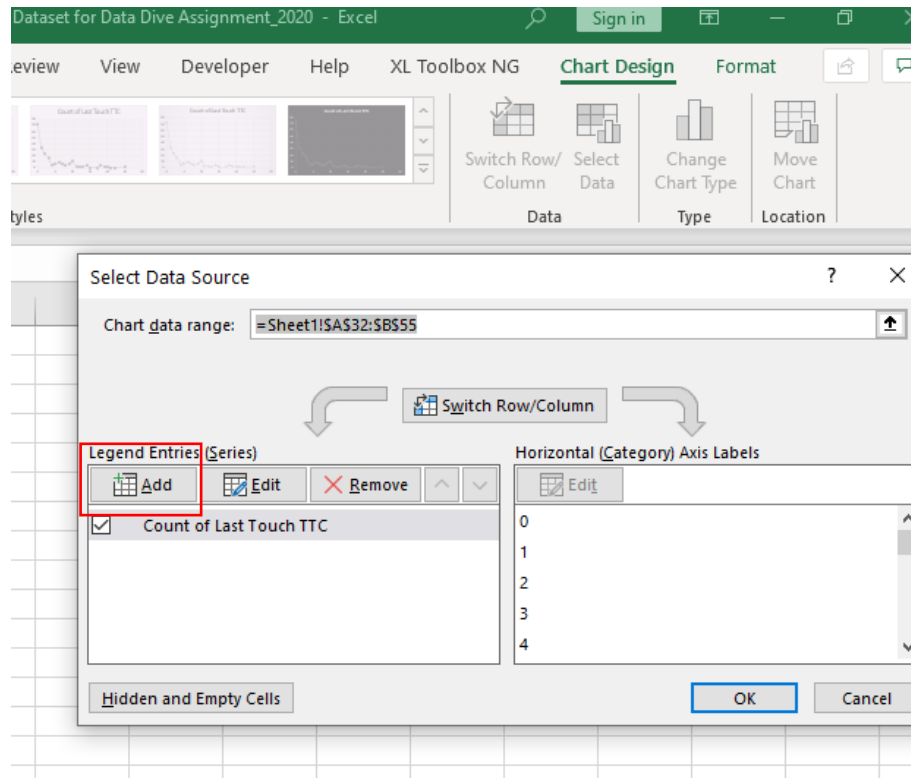
Last Touch TTC	Count of Last Touch TTC	Percents	Cum. Percent
0	73	33%	33%
1	33	15%	48%
2	23	11%	59%
3	17	8%	67%
4	5	2%	69%
5	6	3%	72%
6	9	4%	76%
7	6	3%	79%
8	4	2%	80%
9	7	3%	84%
10	11	5%	89%
11	3	1%	90%
12	2	1%	91%
13	2	1%	92%
14	4	2%	94%
15	1	0%	94%
17	1	0%	95%
18	4	2%	96%
19	2	1%	97%
20	2	1%	98%
21	1	0%	99%
22	1	0%	99%
25	2	1%	100%

So, this tells me that by 8 days after the last touch point, 80% of people in the dataset converted. Of course, from an interpretation standpoint, we have missing data to worry about. But my point here is to show you how to do the calculations.

Bonus: adding cumulative percent line to your graph

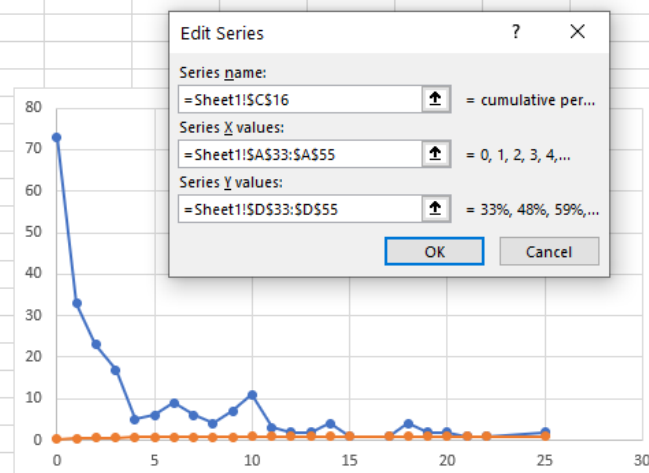
It can be nice to visualize where you start getting a drop in conversions. Here is how you graph the cumulative percent (i.e., percentiles) on your first graph.

Click on the chart you made previously. Go to “chart design” and then to “select data.” You will get this screen. Click on “Add” (see red box).

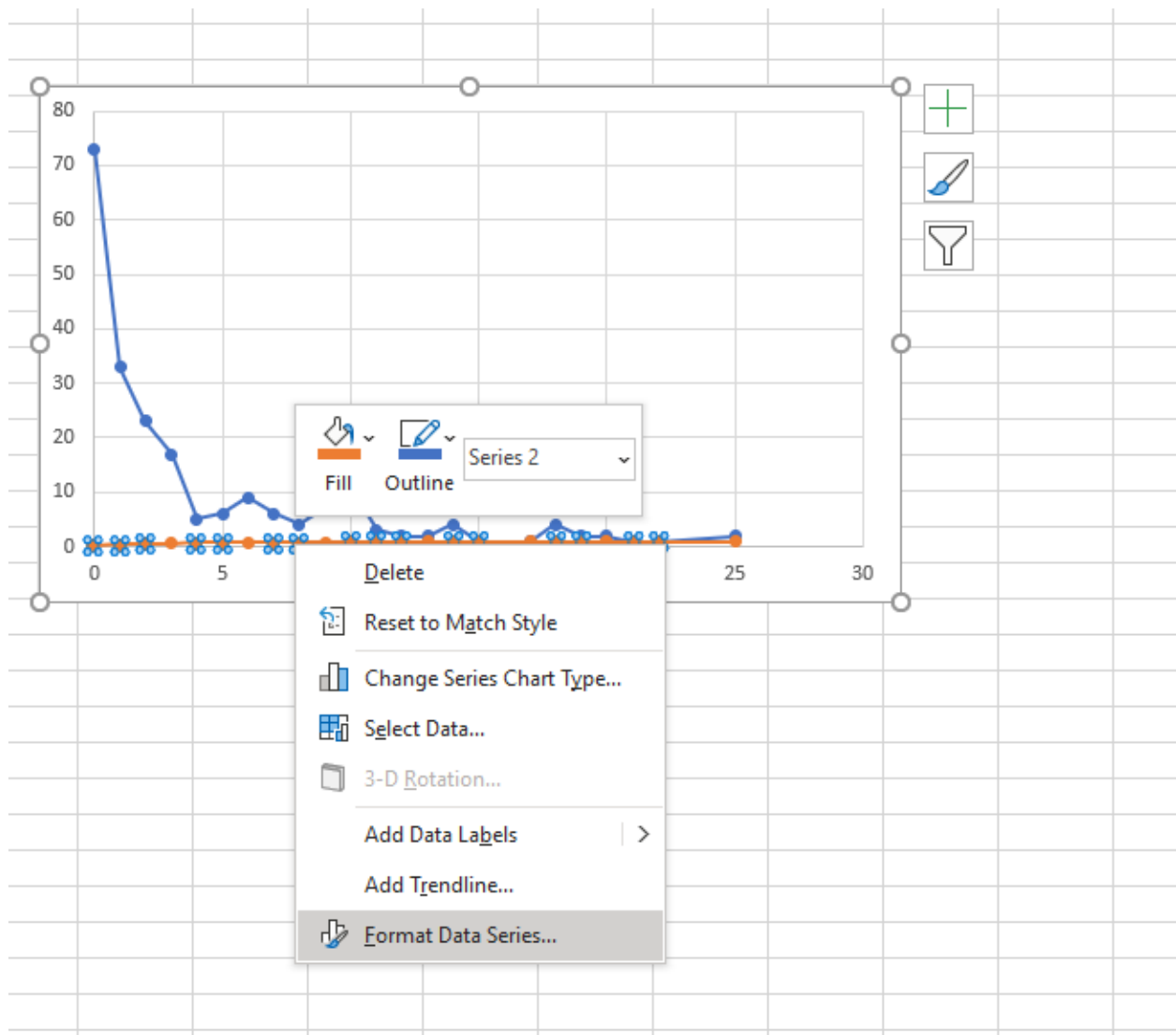


Specify the following (you can select the data with your mouse and Excel will handle writing out the references (e.g., Sheet!.....))

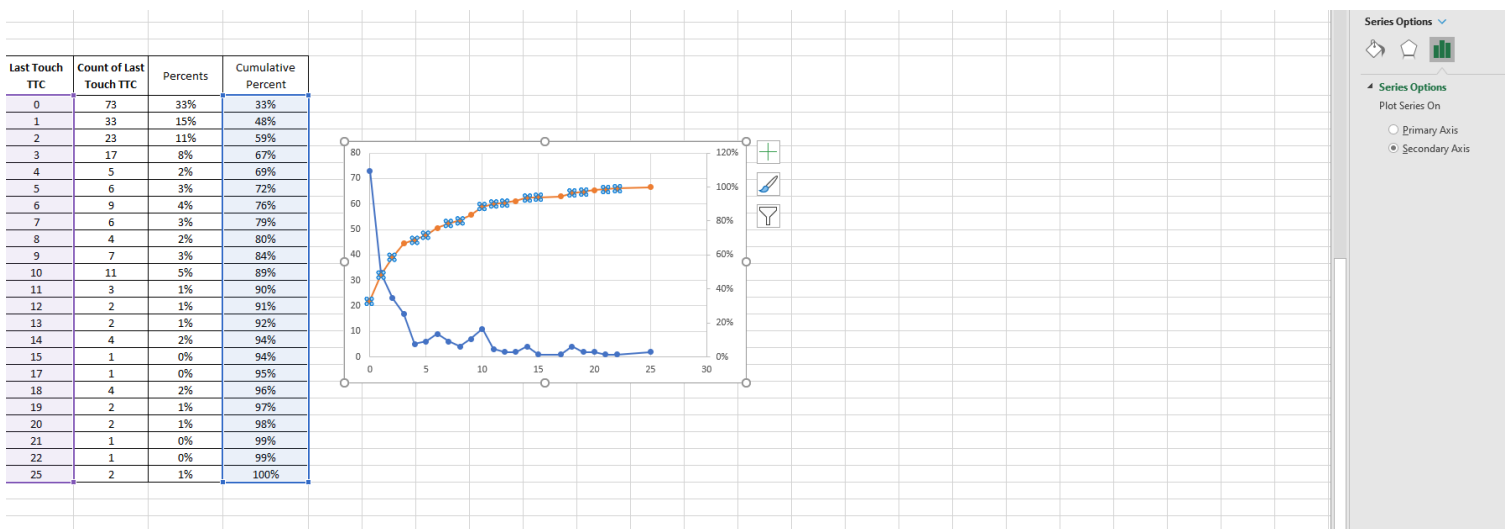
Last Touch TTC	Count of Last Touch TTC	Percents	Cumulative Percent
0	73	33%	33%
1	33	15%	48%
2	23	11%	59%
3	17	8%	67%
4	5	2%	69%
5	6	3%	72%
6	9	4%	76%
7	6	3%	79%
8	4	2%	80%
9	7	3%	84%
10	11	5%	89%
11	3	1%	90%
12	2	1%	91%
13	2	1%	92%
14	4	2%	94%
15	1	0%	94%
17	1	0%	95%
18	4	2%	96%
19	2	1%	97%
20	2	1%	98%
21	1	0%	99%
22	1	0%	99%
25	2	1%	100%



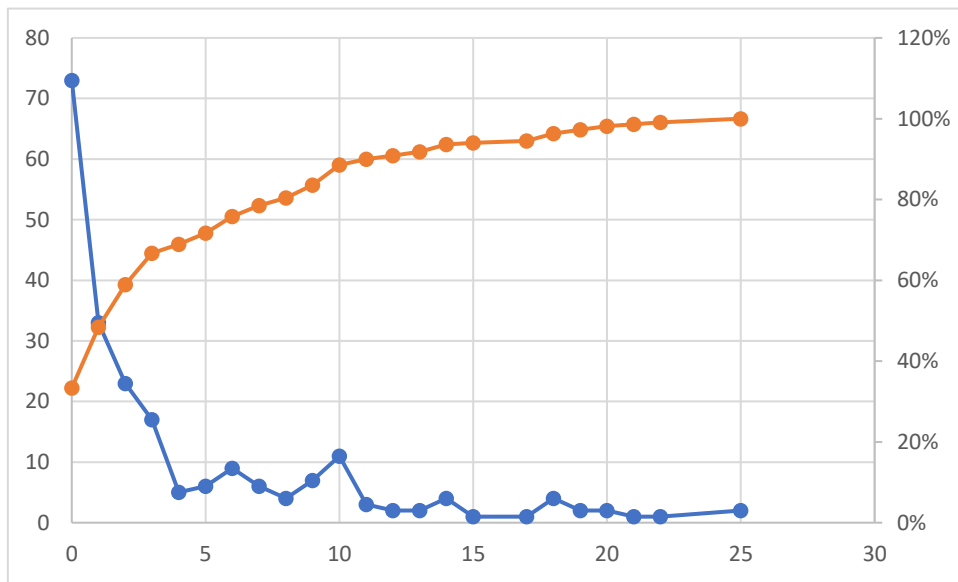
Almost done, but you'll notice that the cumulative percent is basically flat. This is because the Y axis goes from 0 to 80, but our percents go from 0 to 1. We can fix this by adding a second axis. Click on the orange dots and right click, then select "format data series."



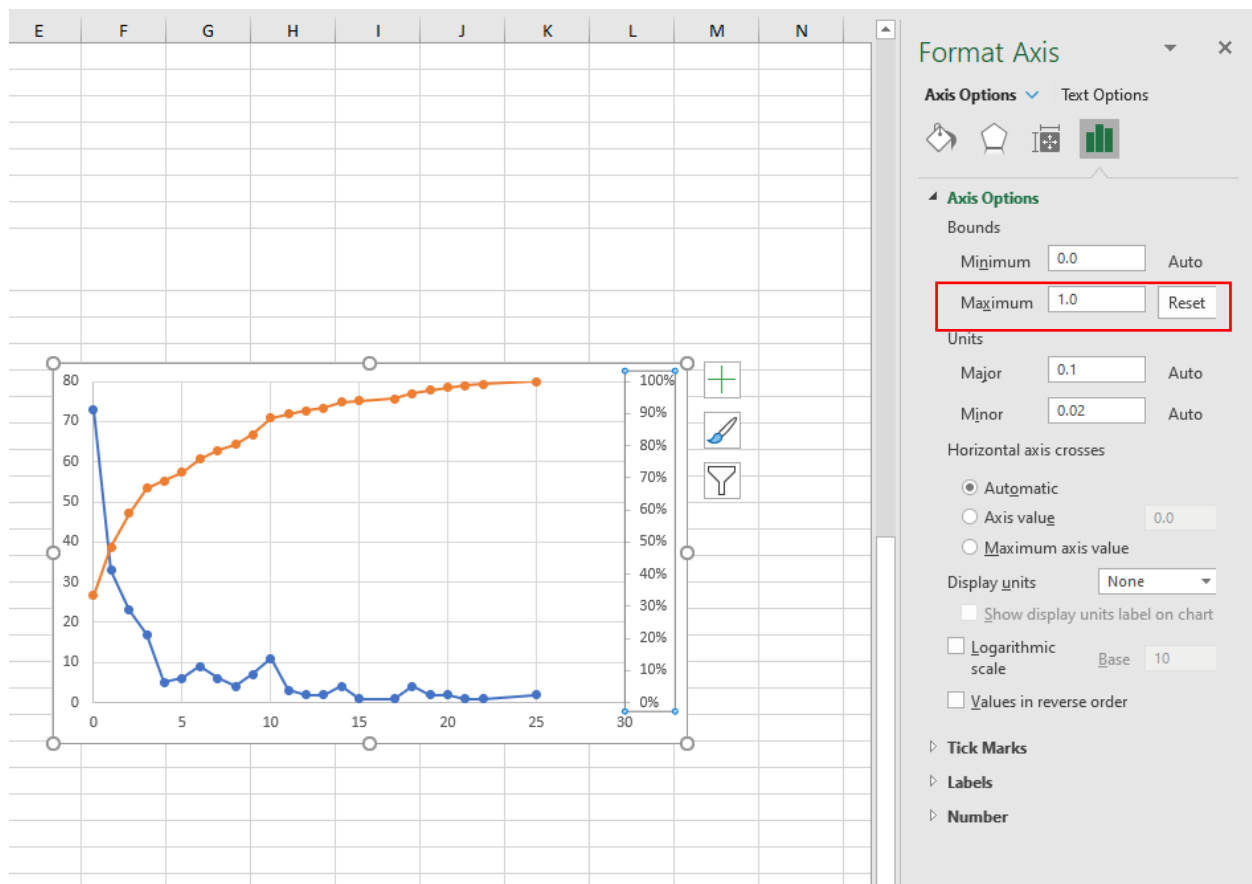
You'll then see this tab open. Click on "secondary axis"



It's almost perfect, but I would like the Y axis to go to 100%, not 120%.



Right click on the righthand axis and select “format axis.” You can then specify the max Y value here.



Can we make the graph prettier? Yes! Play around or Google for tips. But now you have some basic chart building skills, if you didn't have them already.

Excel resource if you get stuck

If you want to learn how to do something in excel, just Google “how do I...in excel.” I do this all the time and I have years of excel experience.

My favorite excel resource is [Excel Jet](#). I've gotten to the point now where if I have an excel question, I just Google “{what I want to do here} excel jet”. Nine times out of ten, Excel Jet pulls through for me. I see that they have YouTube videos, but I find their written answers quicker to understand (they include the code too).

My least favorite resource is support.office.com, so don't worry if you find their explanations confusing.