

MIS0855: Data Science

In-Class Exercise for Fri, Apr 10 – Sentiment Analysis Using Excel

Objective: Differentiate between positive and negative sentiment in text

Learning Outcomes:

- Perform a sentiment analysis of a Twitter stream using software tools
- Compare automatic and manual sentiment analysis methods
- Explain the limitations of automatic versus manual sentiment analysis



In the last exercise, you examined tweets and classified them as positive, negative, or neutral. In this exercise, we'll use some simple software tools to do the same thing automatically.

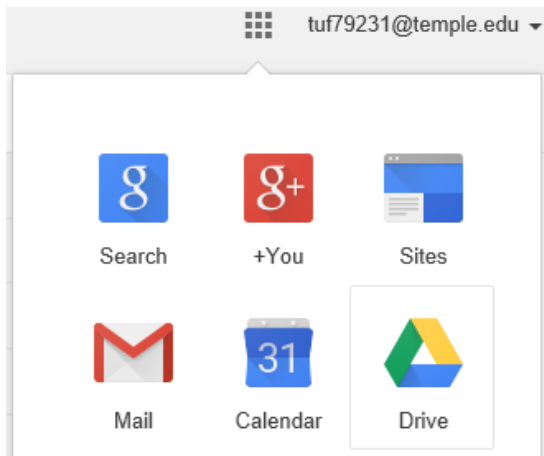
We'll be using Google Docs and Excel to do the analysis. Google Docs will help us gather our tweets and Excel will help us analyze them. While these are somewhat simplified versions of what gets used in industry, they basically work the same way and produce useful results.

Part 0: Create an account at Twitter.

You will also need a Twitter account to complete this exercise. If you don't already have one, go to [Twitter.com](https://twitter.com) and sign up. You don't have to Tweet something for this exercise to work, but you have to know your Twitter username and password.

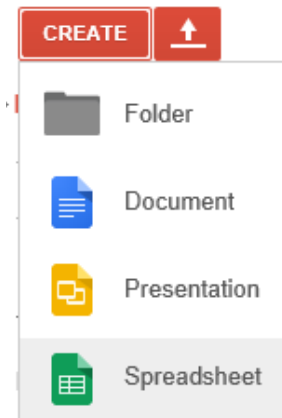
Part 1: Gather Tweets using Google Docs

- 1) Sign into TUMail (<http://tumail.temple.edu>).
- 2) Go to "Google Drive" by clicking on the menu button () next to the search bar and choosing the "Drive" icon ().

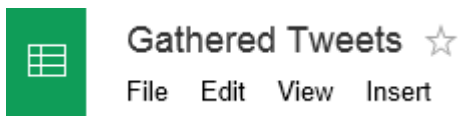


(If you have a Google account, you can log directly on at <http://drive.google.com/>.)

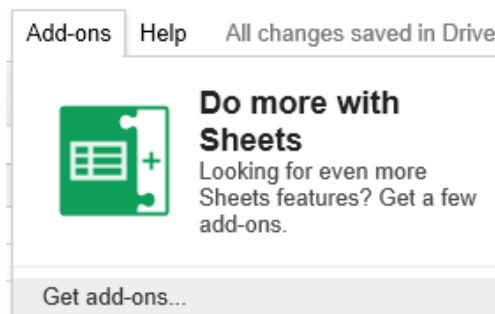
- 3) Click on the “Create” button and select “Spreadsheet”.




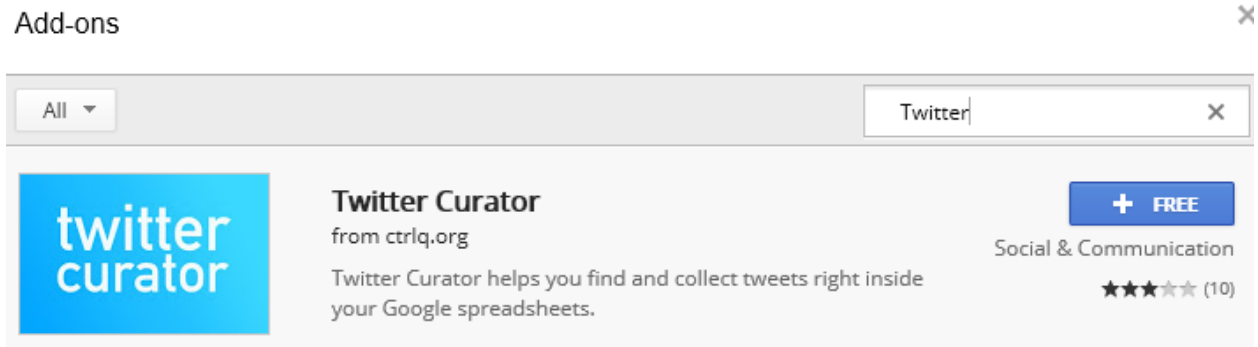
- 4) Give the spreadsheet a name by clicking on “Untitled spreadsheet” and changing it to “Gathered Tweets.”



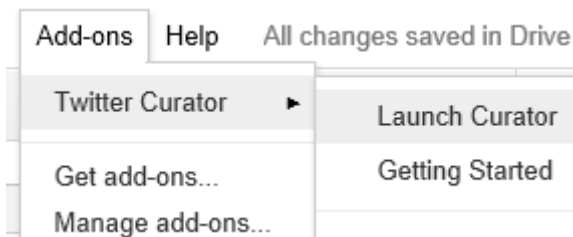
- 5) Go to the Add-ons menu and select “Get add-ons”



- 6) Type “twitter” in the search box and press Enter. You should see “Twitter Curator” in the list. Click on the  button to install it.



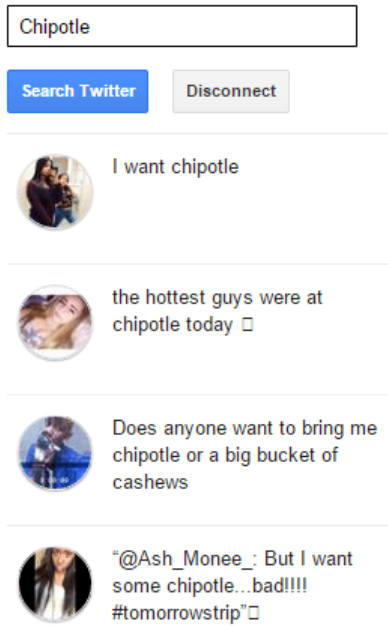
- 7) You’ll see a window asking you to grant the add-on several permissions. Click “Accept.”
- 8) Click on the Add-ons menu and select Twitter Curator/Launch Curator.



- 9) If this is the first time you are using the add-on, it will ask you to sign in using Twitter. Follow the directions to sign in and it will return you to your spreadsheet.
- 10) You’ll see a sidebar appear on the right side of the browser window.
- 11) Search Twitter for a brand. It can be one of the ones you used in the last exercise, or something new.

NOTE: The tool seems to only do exact matches, so @Nike, #Nike, and Nike will return different results. For this exercise, just choose the one that give you the most interesting results.

- 12) The results of your search will appear in the sidebar as a series of tweets.



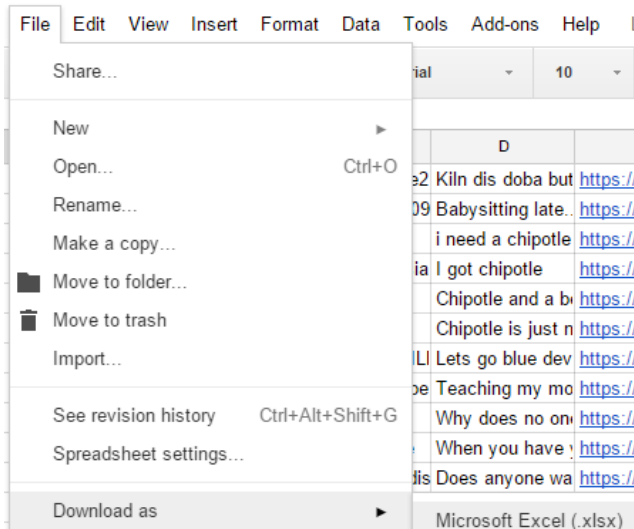
13) Click on a Tweet and the text will be imported into the Google Spreadsheet, along with a lot of other data about the tweet such as the date, the Twitter handle and name of the poster, and the direct URL to the tweet.

08-Mar-2015	catalina	@Catalinaa_c	and dunkin & chi	https://twitter.com/Catalinaa_c/status/574386900028420096
08-Mar-2015	erick espinoza	@erickes360078	BLACK MAN AN	https://twitter.com/erickes360078/status/574386876162818048
08-Mar-2015	Jake Ryon	@Jake_Ryon	bath bombs at lu	https://twitter.com/Jake_Ryon/status/574386839483650050
08-Mar-2015	Stephane Lo Pre	@StephaneWith	Chipotle saved n	https://twitter.com/StephaneWithAnE/status/574386772857253889
08-Mar-2015	Dominick Mr.Tay	@justNickTaylor	Any of my NWA	https://twitter.com/justNickTaylor/status/574386797796540416
08-Mar-2015	phallus	@Nelliee_3870	@_AntCasper r	https://twitter.com/Nelliee_3870/status/574386748173717504
08-Mar-2015	kEEM	@kxmcorona	IM CRAVING CH	https://twitter.com/kxmcorona/status/574386730297638912
08-Mar-2015	Marcus Vieta	@marcusvieta13	Chipotle anyone'	https://twitter.com/marcusvieta13/status/574386718339547136
08-Mar-2015	M A R C H 16th	@Letsbeyanest	At chipotle.	https://twitter.com/Letsbeyanest/status/574386615906336768
08-Mar-2015	Matt Jones	@DamnesiaVu	6. Say "This One	https://twitter.com/DamnesiaVu/status/574386720528957441
08-Mar-2015	Pauline	@PaulineBelle14	Sitting in chipotle	https://twitter.com/PaulineBelle14/status/574386703206645760

Once you import the Tweet it will disappear from the results list.

14) Collect 30 to 40 tweets this way. Choose a combination of positive, negative, and neutral tweets. Also, make sure you only choose Tweets written in English!

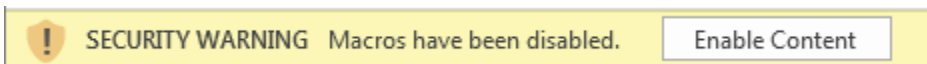
15) When you are done, download the spreadsheet to your computer by going to the File menu and selecting Download as/Microsoft Excel (.xlsx).



Make sure the file is saved in a place where you can find it. (It is automatically saved to your Google Drive account.)

Part 2: Analyze Sentiment Using Excel

- 1) Download “Sentiment Analysis Tools.xlsx” file from the class site and save it to your computer in the same location where you downloaded your Google Docs file.
- 2) Open the Sentiment Analysis Tools workbook. Select the “Sentiment Analysis” tab.
- 3) If you see this warning in Excel:




Click “Enable Content”


NOTE: This spreadsheet has some embedded code that computes average sentiment using a dictionary of positive and negative words to analyze the text.

It gives an overall score from -5 (extremely negative) to 5 (extremely positive) based on (1) the frequency of positive versus negative words in the text and (2) the strength of the feeling conveyed by those words. For example, “great” is a +3, “amazing” is a +4, and “sucks” is a -3.


- 4) To see how it computes sentiment, look at the first sample comment:

This is the best hotel ever.  3


Now change the text to: "This is the most awesome hotel ever."

This is the most awesome hotel ever.  4

Now change the text to: "This is the most awesome hotel ever, but some things are bad."

This is the most awesome hotel ever, but some things are bad.  0.5

Finally, change the text to: "This hotel is horrible."

This hotel is horrible.  -3

- 5) So now put your collected tweets into the spreadsheet. Open the file you just downloaded from Google Docs. It should be an Excel file called "Gathered Tweets."
- 6) All you care about is column D because it contains the tweet text. Highlight the cells in that column and click copy. *Yours will look different from mine since you have different tweets.*

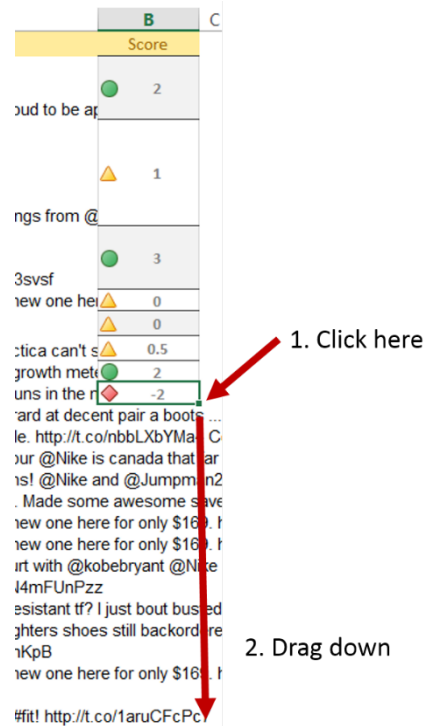
	A	B	C	D	E
1	8/17/2014	DJ SEAN MACN	@TheKidSeanMa	Congrats to Chic	https://t
2	8/17/2014	Patricia Davidsor	@ShopaholicGui	@stylemeupcou	https://t
3	8/17/2014	LIN	@876048018	i love it so much	https://t
4	8/17/2014	Parker Reed	@CouponHorse	Move Into Spring	https://t
5	8/17/2014	mhari	@mhariblakemor	could buy thousa	https://t
6	8/17/2014	nevaeh kowalews	@raynev4eva	"@EmilyOsmen	https://t
7	8/17/2014	Tom Hagan	@Tom_Hagan83	@Nike continue	https://t
8	8/17/2014	Repub Prep	@RepubPrep	Upset because I	https://t
9	8/17/2014	tommypMUFC	@tommypMUFC	@Nike yous wou	https://t
10	8/17/2014	Mark Moons	@markmoons	Google buys Jet	https://t
11	8/17/2014	Gord Mac	@MVPgear	How come I can	https://t
12	8/17/2014	Nick	@TheCreatorND	They are killing C	https://t
13	8/17/2014	Alberto Ruiz	@albertojuiz	The full review of	https://t
14	8/17/2014	The Coupon Twin	@CouponTwin	Move Into Spring	https://t
15	8/17/2014	The Coupon Giraffe	@CouponGiraffe	Move Into Spring	https://t
16	8/17/2014	John Challinor	@JohnChallinor	Nike creates the	https://t
17	8/17/2014	The Coupon Twin	@CouponTwin	I want the new @	https://t
18	8/17/2014	Tay Hizzle	@Ayyelts_Tay	Yo @Nike I thou	https://t
19	8/17/2014	Chad Riggs	@ChadRiggs2	@Nike @nikesor	https://t
20	8/17/2014	Kirk Hughes	@dovermonkey	@Nike Daughter	https://t
21	8/17/2014	I've Got Coupons	@ivegotcoupons	Move Into Spring	https://t
22	8/17/2014	K a t e l y n	@GroneKatelyn	Why does is tak	https://t
23	8/17/2014	One They Call K	@OneTheyCallK	I love my @Nike	https://t

7) Switch back to the Sentiment Analysis spreadsheet, click Cell A2 and paste the text.

8) You'll notice that it only computed the scores for the first 8 tweets. To compute the scores for the rest of the tweets, copy the cell down by clicking on the bottom right corner of Cell B9 and dragging it to the last row in your spreadsheet (see image at right).

9) Now look at the text for each tweet (Column A) and the score (Column B). In some cases, it will categorize the tweets in the way you expect. In other cases, you might disagree with its score.

Things that tend to confuse sentiment analysis tools like this are sarcasm, slang, misspellings, and abbreviations.



10) If you want to compute the average sentiment for all of your tweets, click on Cell E10 and type this formula:

`=AVERAGE(B:B)`

A number greater than 0 means that average sentiment was positive, and a number less than 0 means that average sentiment was negative. Of course, the value matters too. An average value of 0.5 basically means sentiment was either neutral, or there were equal numbers of positive and negative tweets.

11) You could also count how many positive comments there were. In Cell E11, type this formula:

`=COUNTIF(B:B,">=2")`

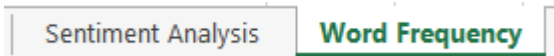
This will tell you how many tweets had a sentiment score of 2 or greater.

Part 3: Word Frequency Analysis Using Excel

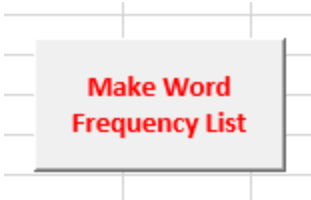
Another useful method of text analysis is finding which words appear most frequently within a collection.

There is some Excel code embedded into our “Sentiment Analysis Tools” worksheet that will compute that for you.

- 1) Click on the “Word Frequency” tab.



- 2) Column A contains the list of text snippets to be analyzed. The test data is the titles from all 287 Beatles songs. Around Column H/I is a button:



Click that button.

- 3) You’ll see a new worksheet (Sheet1).

Column A has every word listed on a separate line.

Column C has every unique word listed on a separate line.

Column D contains the number of times that word appears in the collection of song titles.

	A	B	C	D
1	All Words		Row Labels	Count of All Words
2	12-BAR		1	1
3	ORIGINAL		9	1
4	A		909	1
5	BEGINNING		SUN	3
6	A		SEPTEMBER	1
7	DAY		12-BAR	1
8	IN		A	24
9	THE		ACROSS	1
10	LIFE		ACT	1
11	A		AFTER	1
12	HARD		AGAIN	1
13	DAYS		AINT	1

So from this you learn that SUN appears three times and SEPTEMBER appears only once.
























- 4) Right-click somewhere inside Column D and select Sort/Sort Largest to Smallest. You now see that the most frequently occurring word is YOU (31 times) followed by THE (28 times). You’ll notice many of the most frequent words are what you’d expect, like THE, A, TO, AND, etc.
- 5) Now let’s try it with your tweets.

First, switch to the “Word Frequency” worksheet and click at the top of Column A (right on the “A”).

	A
1	12-Bar Original
2	A Beginning
3	A Day in the Life
4	A Hard Day's Night
5	A Shot of Rhythm and Blues
6	A Taste of Honey
7	Across the Universe
8	Act Naturally
9	Ain't She Sweet
10	All I've Got to Do

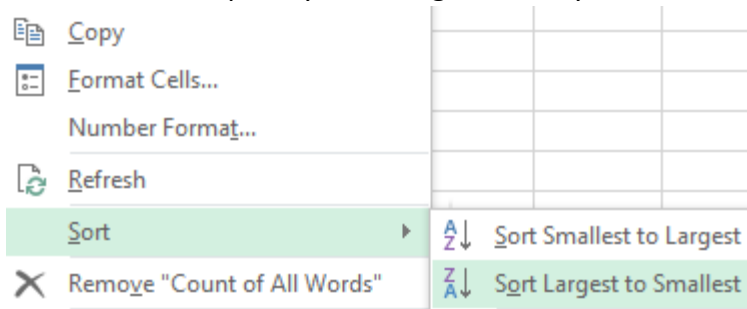
Press delete. This should clear all the text in that column.

- 6) Go back to the “Sentiment Analysis” worksheet and select all the tweets in column A (A2 to the last tweet in that column). Copy the cells.

	A	B
1	Comment	Score
2	Congrats to Chicago for taking home the trophy! Very proud to be a	 2
3	@stylemeupcouk I invested in some @Nike printed leggings from @	 1
4	i love it so much @Nike @NikeRunning http://t.co/40ZPp3svsf	 3
5	Move Into Spring with the @NIKE Fuelband SE. Get the new one he	 0
6	could buy thousands of things from @Nike	 0
7	"@EmilyOsment: Thanks for all the swag @Nike ! Antarctica can't s	 0.5
8	.@Nike continues to help find ways to raise the creative growth met	 2
9	Upset because I still haven't gotten my custom Roshe Runs in the r	 -2
10	@Nike yous would think you guys would lend Steven Gerard at dece	 0
11	Google buys Jetpac and intensifies competition with Apple. http://t.co/	 1
12	How come I can't get any gear in canada ? @UnderArmour @Nike i	 0
13	They are killing Chicago with these events and promotions! @Nike a	 0.5
14	The full review of the awesome @Nike GK Premier SGT. Made sorr	 4
15	Move Into Spring with the @NIKE Fuelband SE. Get the new one he	 0
16	Move Into Spring with the @NIKE Fuelband SE. Get the new one he	 0
17	Nike creates the world's first full-size LED basketball court with @kc	 0
18	I want the new @Nike Fuel Band SE so bad! http://t.co/iN4mFUnPz	 1
19	Yo @Nike I thought yall shoes was supposed to be slip resistant tf?	 -1
20	@Nike @nikesoccer @soccerdotcom WHY are my daughters shoe	 -1
21	@Nike Daughter loves her new Nikes http://t.co/IXOUK1hKpB	 0
22	Move Into Spring with the @NIKE Fuelband SE. Get the new one he	 0
23	Why does it take @Nike so long to make shoes?	 0
24	I love my @Nike Fuelband SE. its the push i need to get #fit! http://t.co/	 3

- 7) Switch back to the “Word Frequency” worksheet and click in Cell A1. Paste the text. It will look messy. That’s ok.

- 8) Click the “Make Word Frequency List” button. You’ll see Sheet2, containing a new word frequency analysis.
- 9) Right-click somewhere inside Column D and select Sort/Sort Largest to Smallest. You’ll now see the most frequently occurring words in your tweets.



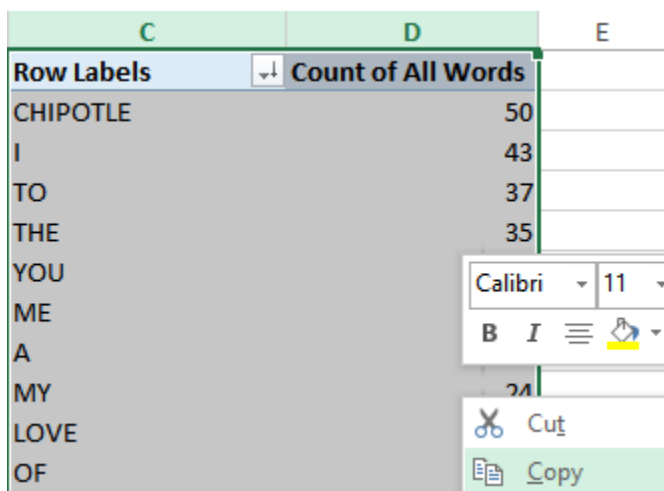
It will be messier than the count of words in Beatles titles, because tweets are messier in general. But if you ignore the nonsense words and common words like “a,” “an,” and “the” you can get a sense of popular terms among people tweeting about your selected brand.


- 10) Save your Excel workbook.

Part 4: Visualizing Your Word Frequency Analysis in Tableau

A table of word frequency counts is fine, but why not use Tableau to create an easy to navigate visualization?

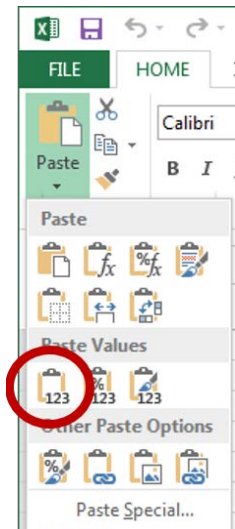
- 1) Open your “Sentiment Analysis Tools” spreadsheet in Excel if it’s not still open.
- 2) Highlight Columns C and D, and copy the columns.



- 3) Create a new worksheet in the workbook (remember, click the  icon to do that).



- 4) Click in Cell A1 in the blank worksheet.
- 5) Select Paste/Paste Values (selecting “Paste Values” is important – don’t just use regular paste!)



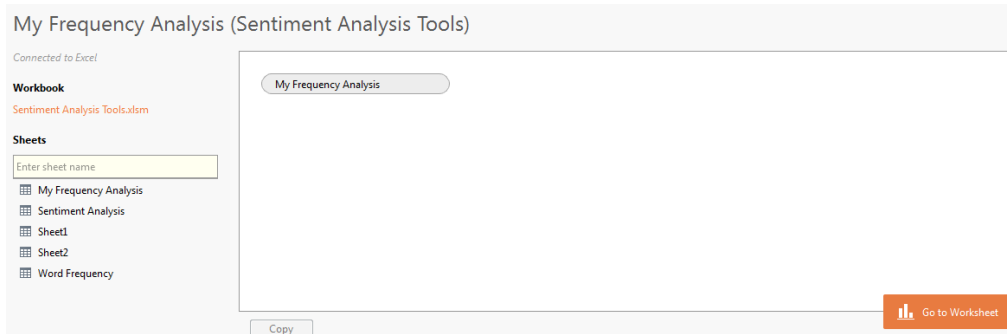
- 6) You’ll see your word frequency table. Again, your tweets are different so the table will look different from mine, but the first row will be the same:

	C	D
Row Labels		Count of All Words
@NIKE		23
THE		19
TO		9
I		9
WITH		7
FOR		7
GET		6
SE		6
NEW		6

- 7) Rename the tab “My Frequency Analysis.” Then save the workbook and close Excel.
- 8) Start Tableau.
- 9) Click “Connect to data.”
- 10) Click “Microsoft Excel”

11) Open your “Sentiment Analysis Tools” workbook.

12) Drag the “My Frequency Analysis” worksheet into the whitespace. Click “Go to worksheet.”



13) Drag the “Row Labels” dimension to the Columns shelf and the “Count of All Words” measure to the Rows shelf.

14) Click the treemap icon under “Show Me.”

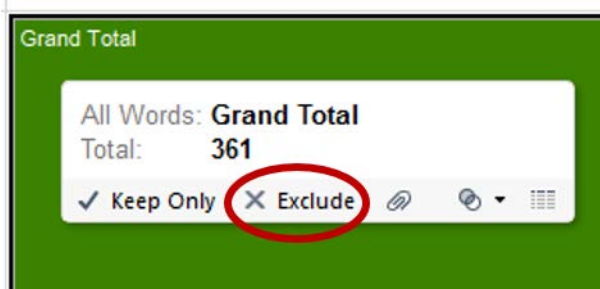


15) You'll see something like this:

Grand Total	@NIKE	THE	I	TO	FOR	WITH				
GET	INTO	OF	ONLY	SPRING	ARE	IN	LOVE	SHOES	SO	
NEW	CANT	FROM	A	BE	SOME	STILL	TAKE	IS		
SE	THEM	BUY								
FUEL	IT	DOES	GK							
BAND	DRIVES									
WHY	EVEN				ITS		LED			
MOVE										
ONE	FAMILY	LONG			NO		NUN			
	FAR	@NBA								
	WOULD	LOVES								
	FIND	MADE	PAIR							
	YOUR	FIRST	POWER	SLIP					TF	
	FIT	MAIL								
	FUEL	MAKE			AIN		UP			
169	BAD	FULL	ME		ALL	THAT	THINK	WAS	AT	
BAND		METER	PROUD		ANY			WAYS		
HERE	GAME	@IBM	MORE	PUSH	STEP	APART		YALL	YET	YO
MY	BOOTS	GEAR				THESE	APPLE	WILL	YOU	
	BUSTED		MUCH	REBOK		THEY	TOO	ASS	AWAY	

16) Having “Grand Total” in there doesn’t make much sense, since that’s the sum of all word frequencies (that’s why it’s the exactly same size as all the rest of the words put together).

So click once on “Grand Total” and click “Exclude”



17) You now have a snapshot of the frequencies of all terms in your tweet collection!

