

## Lecture Notes on Neural Network: Part 2A, Excel Data Column Extraction

Nick V. Flor, nickflor@unm.edu

Now that you know the basics of Neural Networks (NNs), let's apply what you've learned to create a NN that tries to predict if a tweet will go viral. This consists of several parts.

### PART A. SETUP INPUT-OUTPUT DATA SET

#### Step 1. Scrape your Data & Extract a Spike (only end-result shown)

- Scrape a hashtag during a high-activity day, e.g., #TheMagicians on Thursday.
- Extract the spike and put it in a new sheet:

FileHomeInsertPage LayoutFormulasDataReviewViewDeveloperHelp

Visual BasicMacro SecurityCodeAdd-insExcel Add-insCOM Add-insInsertDesign ModeRun DialogControlsSourceMap PropertiesExpansion PacksImportRefresh Data

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

W

3994

(continued on next page)

## Step 2. Remove Stop-Words

- Highlight* just the data in column P (not the entire column).
- Select* the menu item: Developer > Macros > removeStopWords.
  - Note*: the end of column P has the tweets with stop-words removed.
- Format* the sheet so you don't have weird row heights (not shown).

The screenshot shows the Microsoft Excel interface with the following elements:

- Ribbon:** File, Home, Insert, Page Layout, Formulas, Data, Review, View, Developer, Help.
- Formulas Tab:** Add-ins, Add-ins, Add-ins, Add-ins, Design Mode, Controls, Properties, View Code, Source, Map Properties, Expansion Packs, Import, Refresh Data, XML.
- Formula Bar:** Contains the formula `=IF(AND(A1<>"",B1<>""),C1,D1,E1,F1,G1,H1,I1,J1,K1,L1,M1,N1,O1,P1,Q1,R1,S1,T1,U1,V1,W1,X1,Y1,Z1,AA1)`.
- Worksheet Grid:** Columns A through AA. Rows 1 through 3994. The grid contains a repeating pattern of tweets from @arjanguptabk, @magicianswyr, and @magicians, each followed by a URL. The tweets are: "got magic back well world lived happily ever after... jk, #themagicians. don't know 'happy' is. see guys next season! https://t.co/penny's dj now. bet parties lit.", "thought i remembered q noooooooooooooo #themagicians excuse what!!! #themagicians", "jesus fucking christ is... just... L.. fuck #themagicians want hell guys #themagicians", "rt @lvidgetfactoryco: #themagicians \$seasonfinale https://t.co/o17ddt90ei allot creepy monster thing? ohhhhhh shit. #themagicians inside eliot #themagicians", "oh he's best #themagicians", "rt @magiciansfyf: child like things always creepy #themagicians https://t.co/1e9yc059lg fuckk #themagicians upset", "similar vices #themagicians https://t.co/emBuktncuu become just regular people... they're magicians anymore?#themagicians https://t.co/180d9ydb1o".
- Status Bar:** Processing Tweet #3994/3994, Spike, Count: 3994.

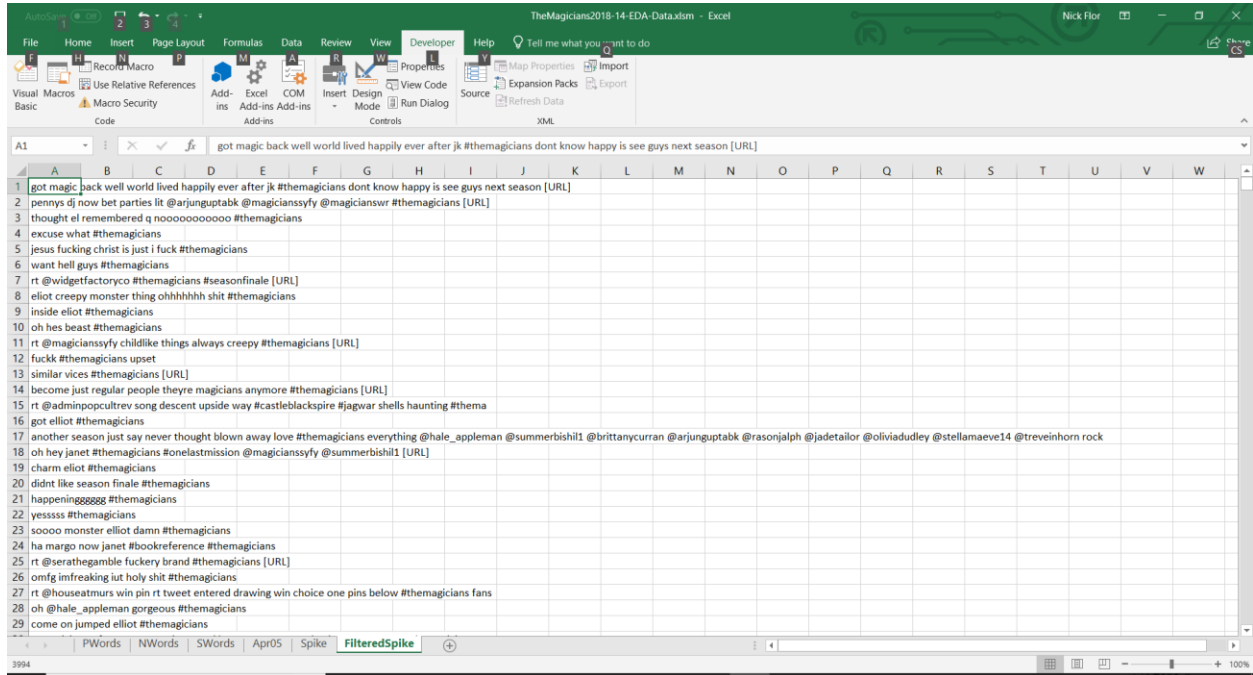
### Step 3. Filter those Tweets

- Select all the tweets with the stop-words removed.
- Select menu-item: Developer > Macros > filterTweets
- Note: Column S has the filtered tweets (no punctuation, lowercase)

[illegible]

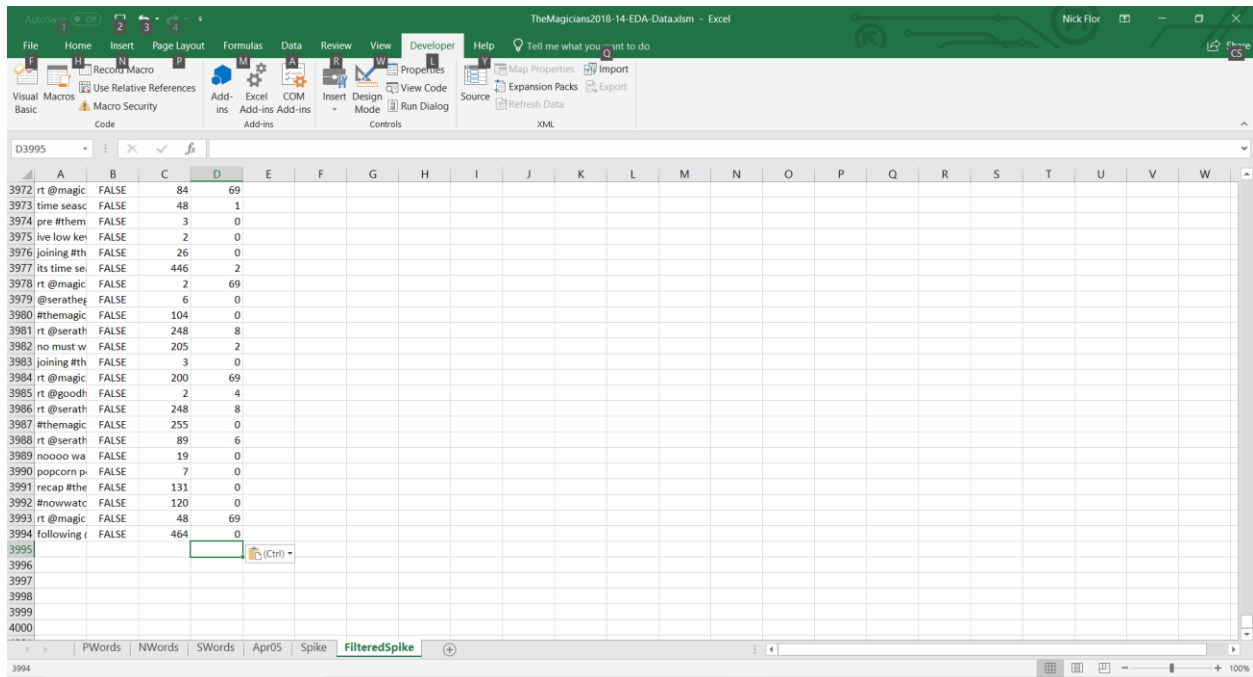
## Step 4. Copy the Filtered Tweets to a New Worksheet

a. *Note: I named the new worksheet **FilteredTweets***



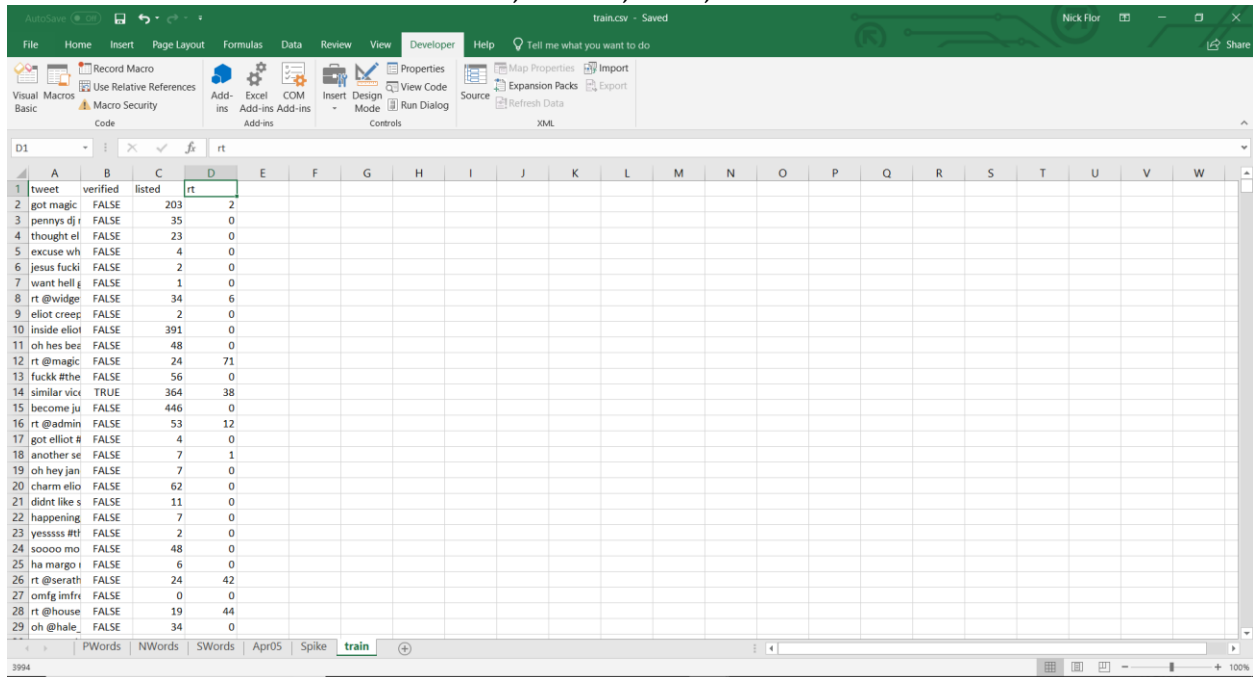
## Step 5. Copy Columns M (verified user) & J (listed) & L (# retweets) in FilteredSpike

a. *Note: The column lengths should line up perfectly (e.g., all are 3994 rows long).*



## Step 6. Add Column Headers

a. *Note:* I added the columns: tweet, verified, listed, rt



The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	tweet	verified	listed	rt																			
2	got magic	FALSE	203	2																			
3	pennys dj r	FALSE	35	0																			
4	thought el	FALSE	23	0																			
5	excuse wh	FALSE	4	0																			
6	jesus fucki	FALSE	2	0																			
7	want hell g	FALSE	1	0																			
8	rt @widge	FALSE	34	6																			
9	eliot creep	FALSE	2	0																			
10	inside eliot	FALSE	391	0																			
11	oh hes bee	FALSE	48	0																			
12	rt @magic	FALSE	24	71																			
13	fuckk ttthe	FALSE	56	0																			
14	similar vici	TRUE	364	38																			
15	become ju	FALSE	446	0																			
16	rt @admin	FALSE	53	12																			
17	got elliot #	FALSE	4	0																			
18	another se	FALSE	7	1																			
19	oh hey jan	FALSE	7	0																			
20	charm elio	FALSE	62	0																			
21	didnt like s	FALSE	11	0																			
22	happening	FALSE	7	0																			
23	yesssss tt!	FALSE	2	0																			
24	soooo mo	FALSE	48	0																			
25	ha margo i	FALSE	6	0																			
26	rt @serath	FALSE	24	42																			
27	omfg imfr	FALSE	0	0																			
28	rt @house	FALSE	19	44																			
29	oh @hale	FALSE	34	0																			

## Step 7. Export (e.g., save as train.csv)

b. *Note:* I did a save as train.csv