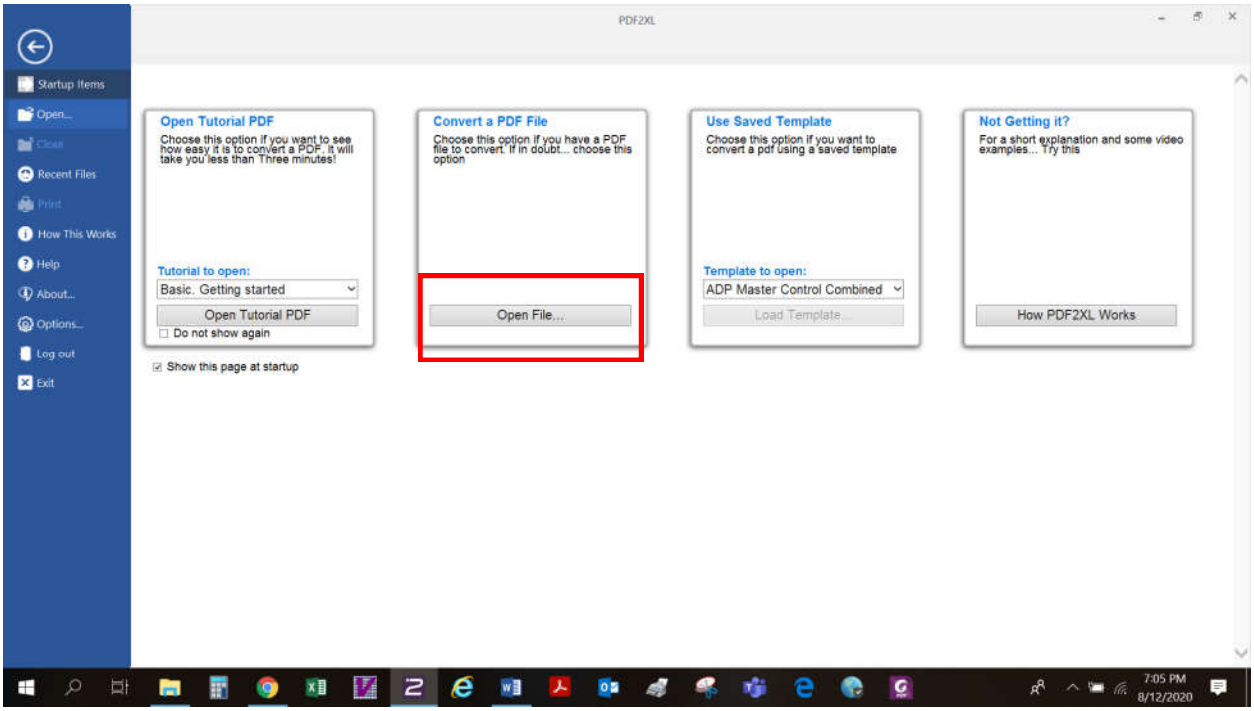
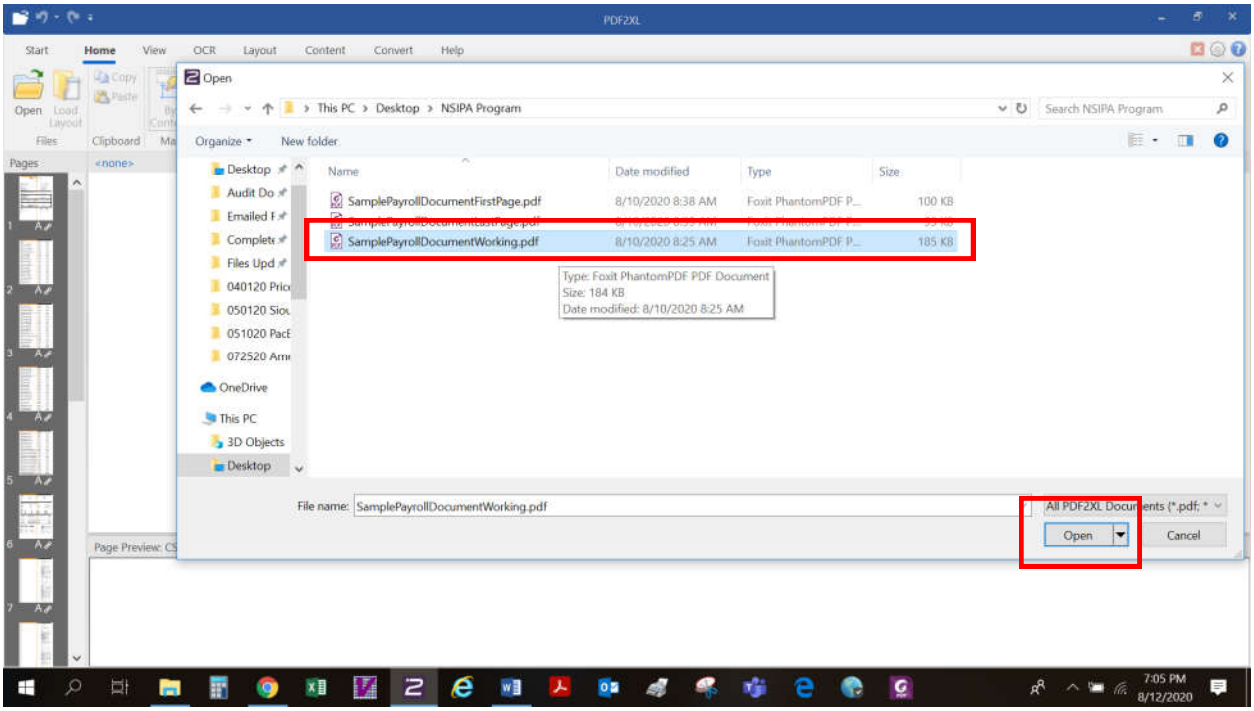
Using Cogniview to Convert a PDF to Excel and Mining Necessary Data to Complete a Premium Audit

By Greg Rohner APA, Senior Auditor—National Accounts, Berkshire Hathaway Homestate Companies

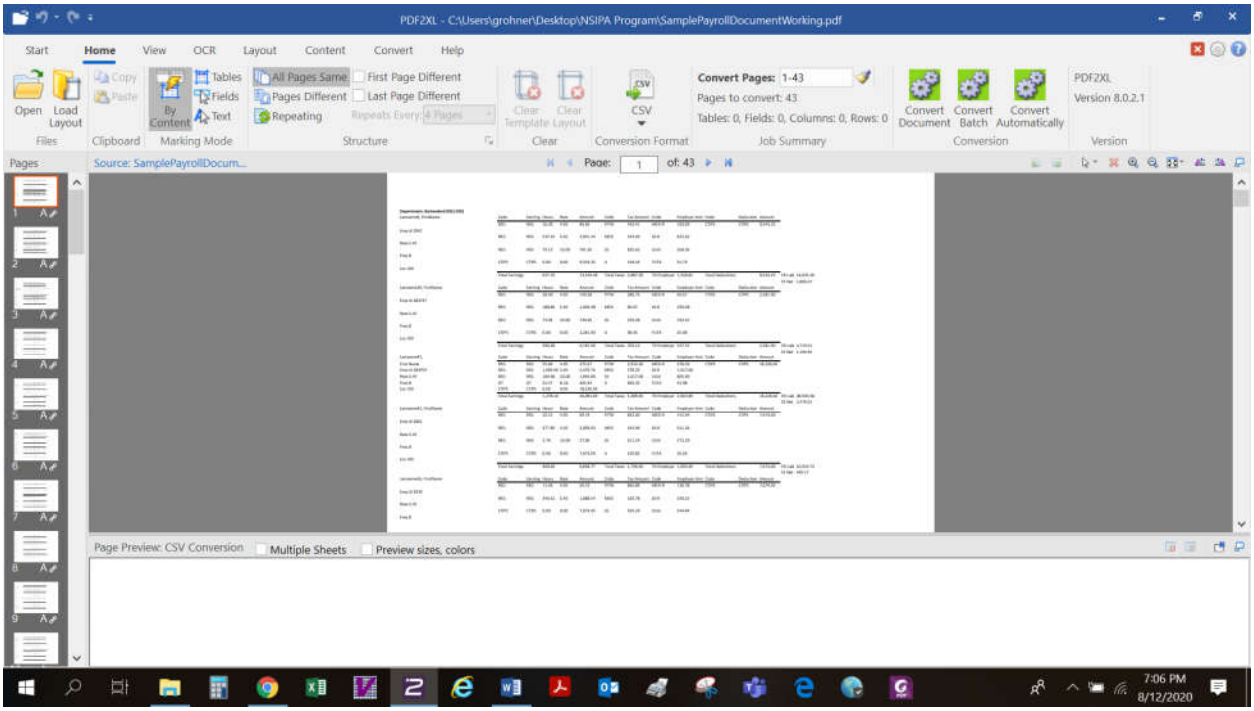
Initial Screen for Cogniview



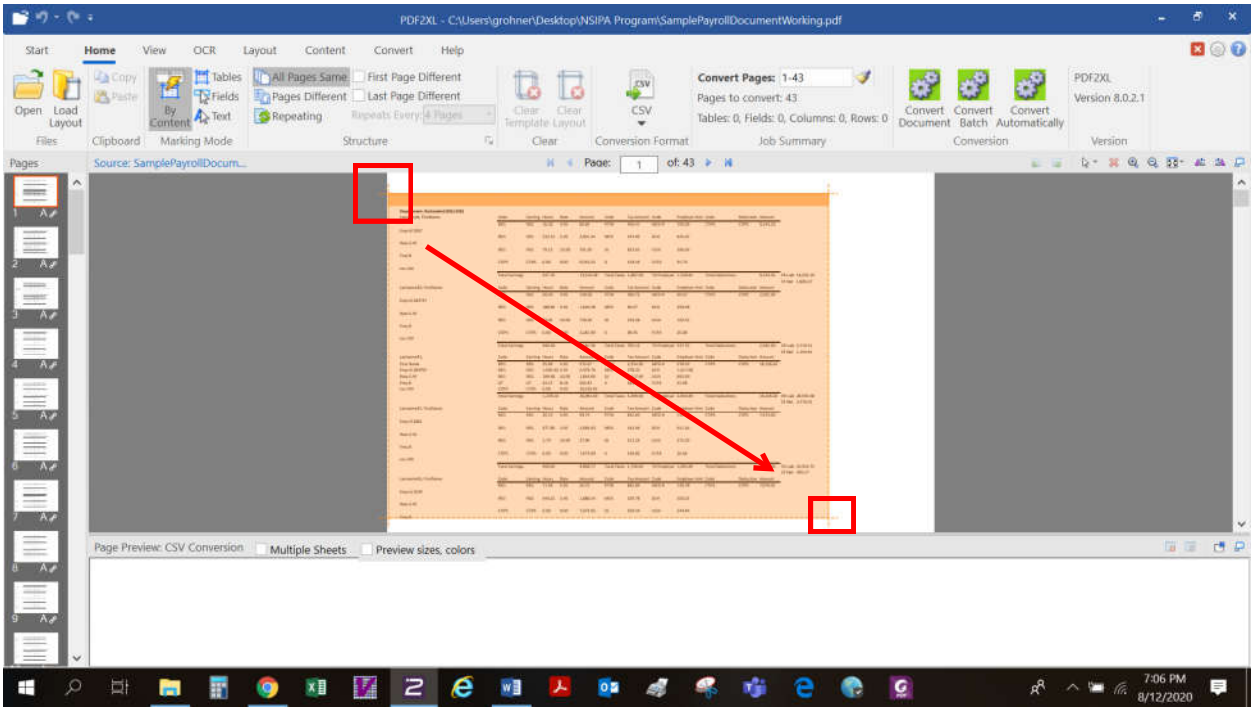
Click on “Open File” and navigate to the location where the pdf is that you want to convert. Click “open” once the file is selected



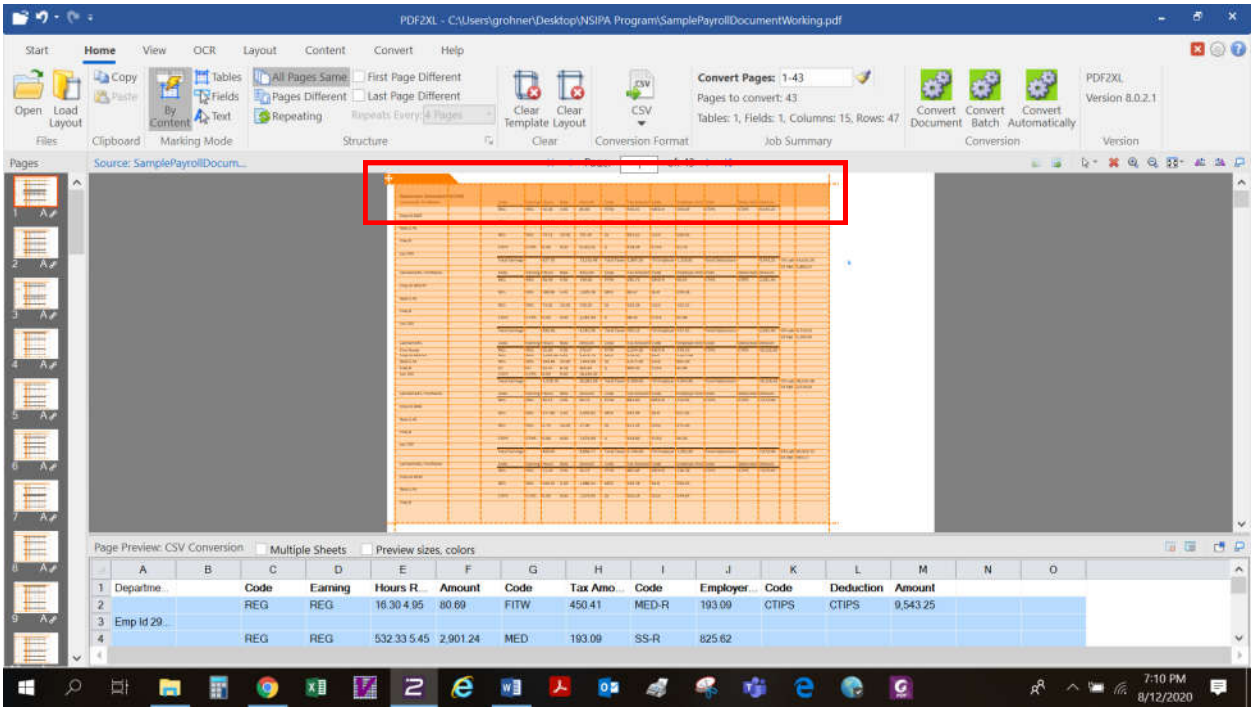
Once a file is open, you will see the first page in the program.



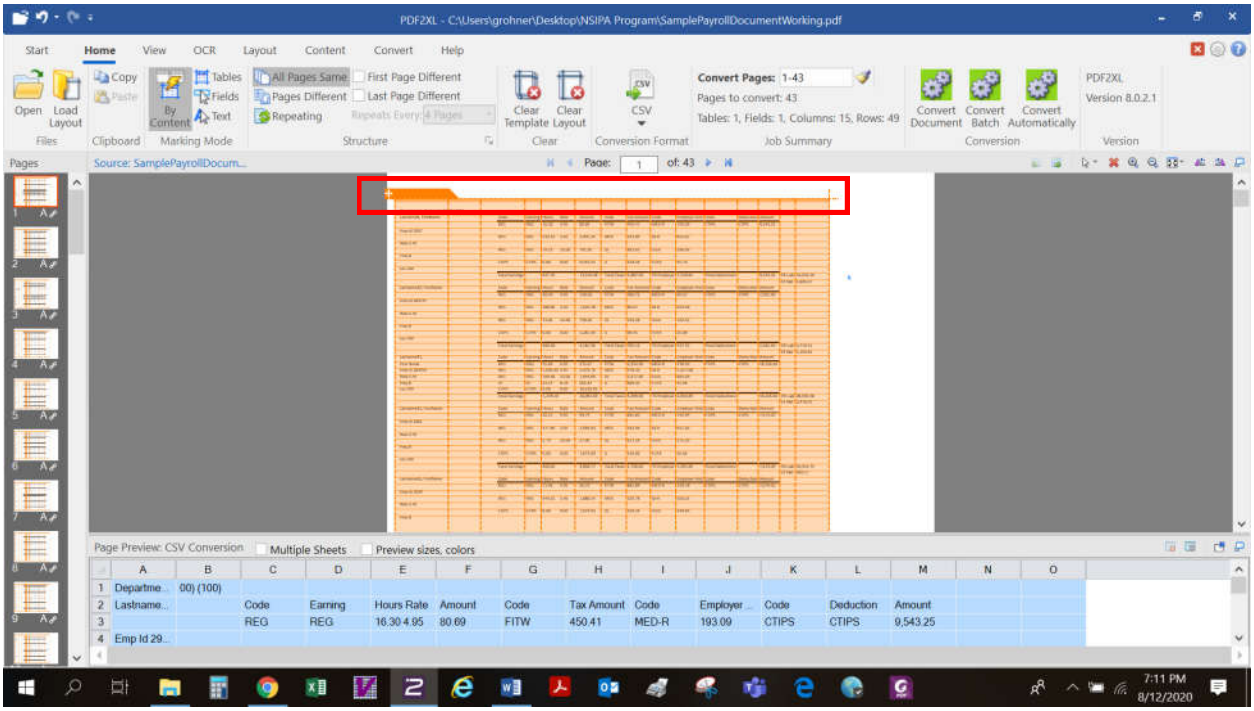
From the upper left corner, click and drag your curser diagonally so that the entire data field on the first page is highlighted.



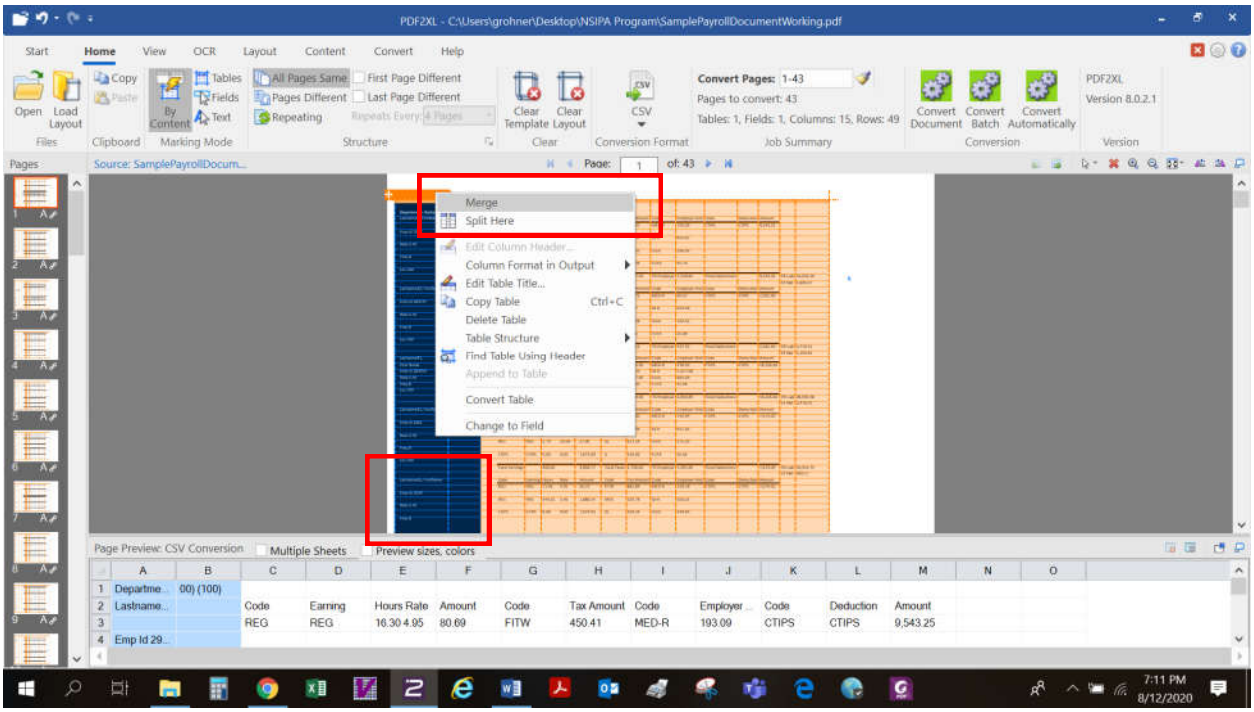
Once the click is released, Cogniview will attempt to identify columns and rows. The program will also attempt to identify a header, which is denoted by the darker shaded orange area. Data in that area will only show once on the exported report



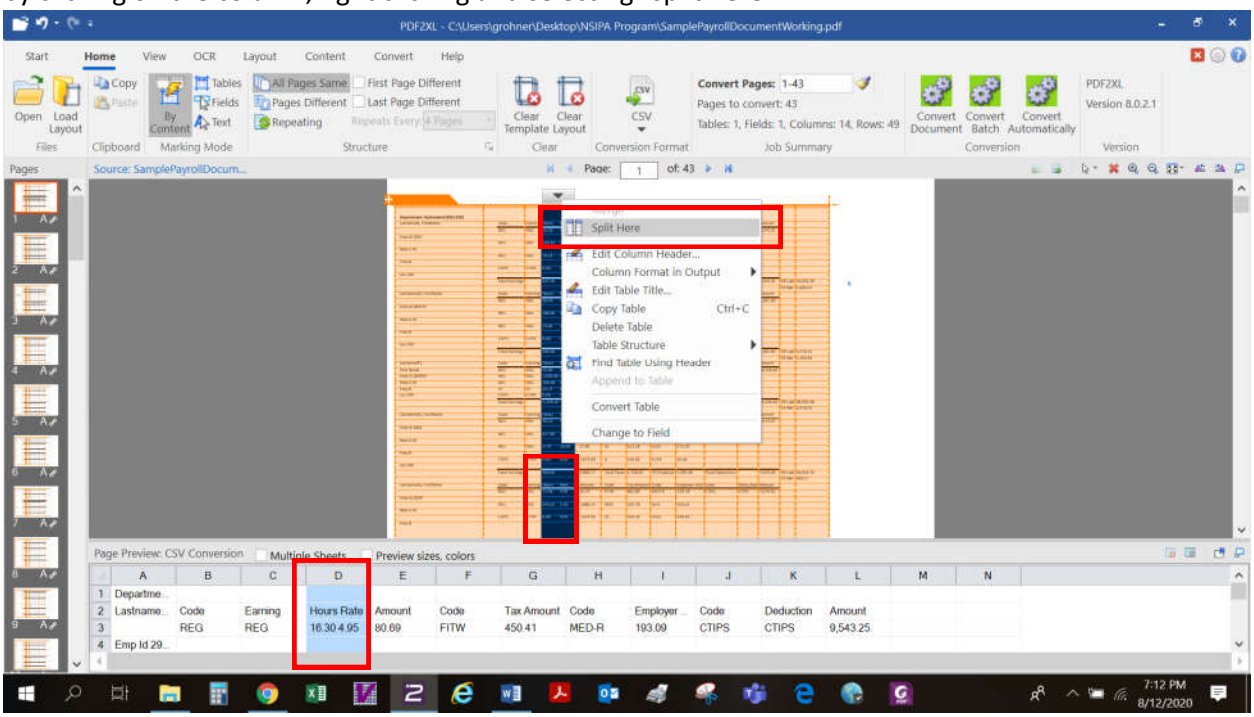
Click on the edge of the shaded area and drag up to eliminate it.



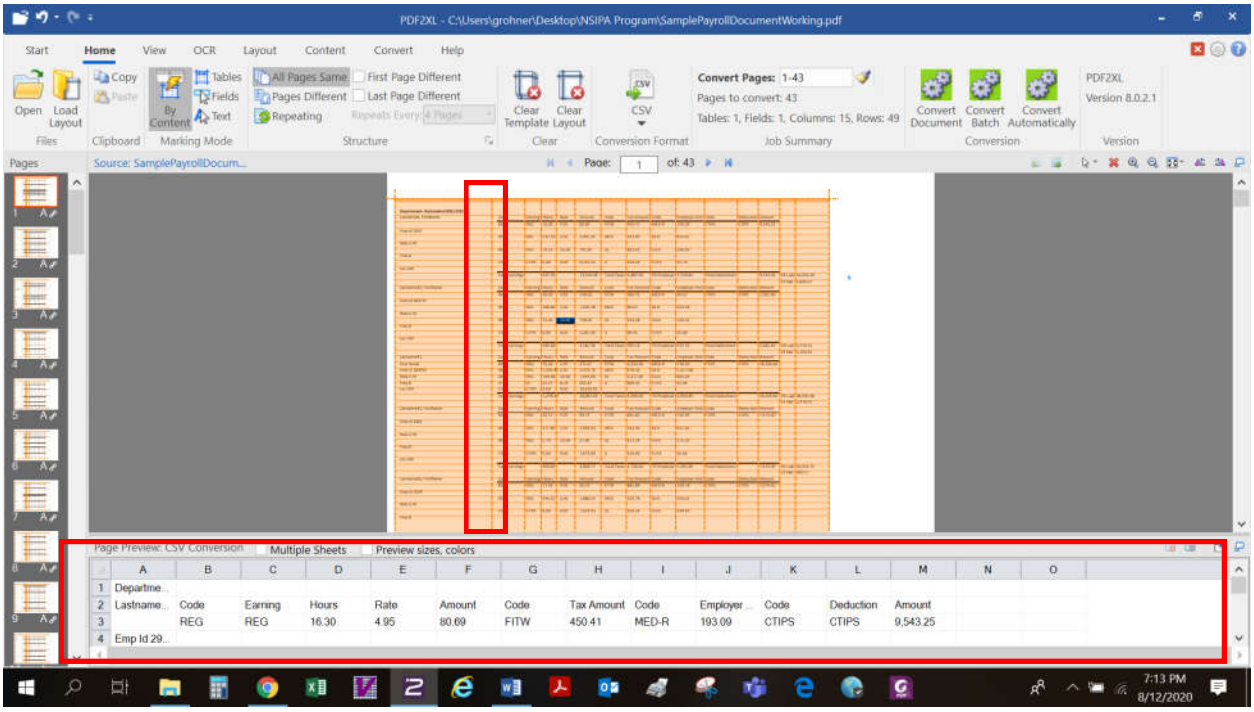
In this case, Cogniview placed two columns for the column where names are located. These need to be merged together, which can be accomplished by clicking on the first column to the left, holding the control key down and clicking on the column next to it. Right click once that's completed and select "merge"



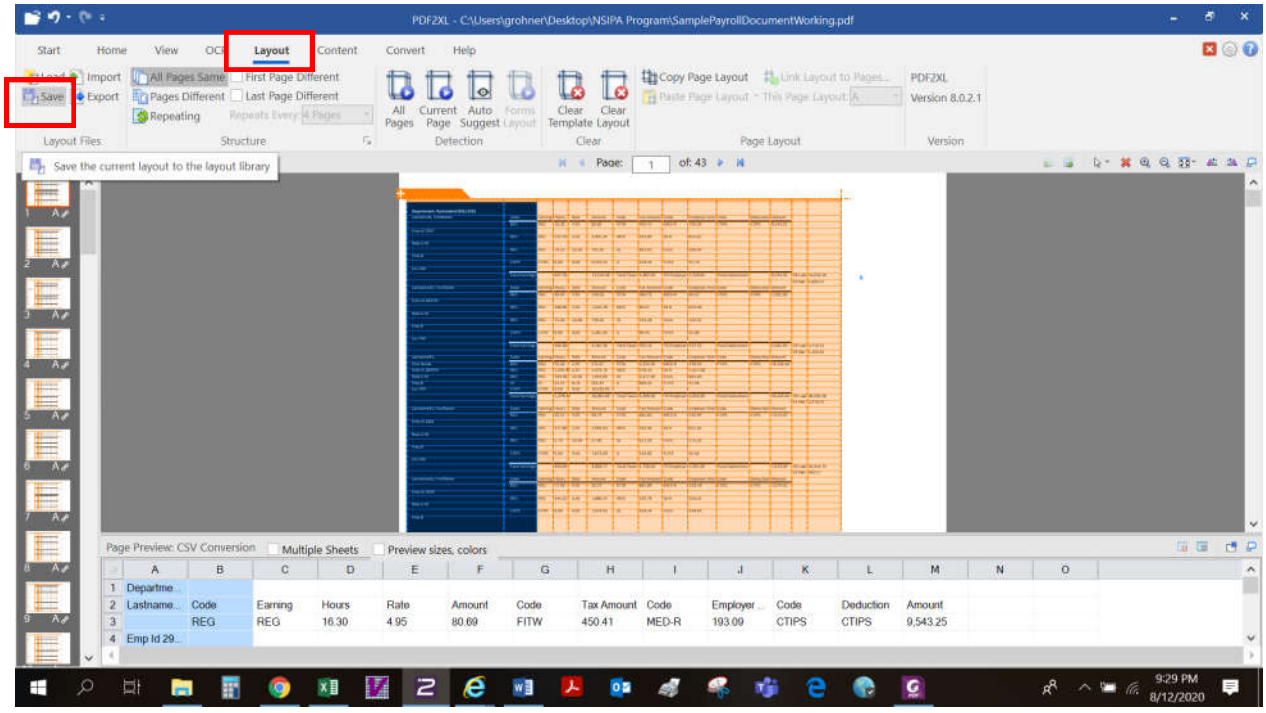
If you find there's a case where data that should be in two columns, but are in one, that data can be split by clicking on the column, right clicking and selecting "split here".



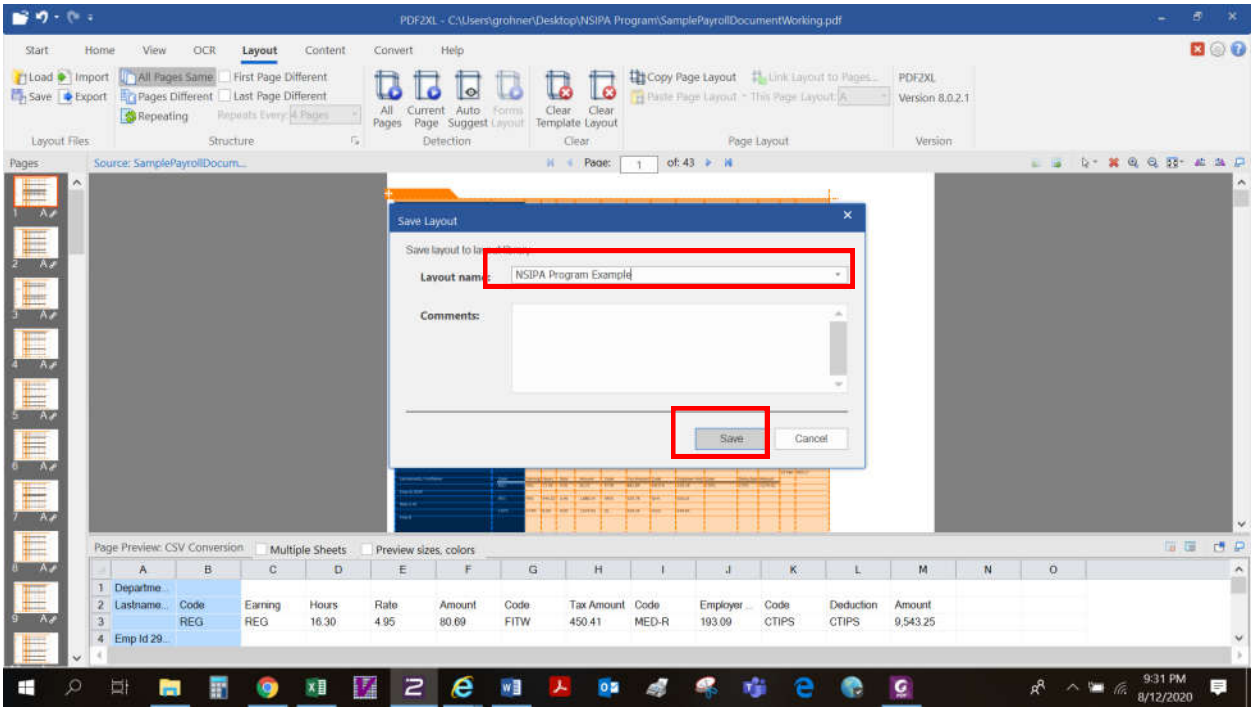
Columns can be adjusted by clicking on the dotted orange lines and dragging the line to the desired position by clicking on it and a double arrow appears. Also note that the lower part of the screen shows a preview of what the output will be. That view can be adjusted or scrolled as well.



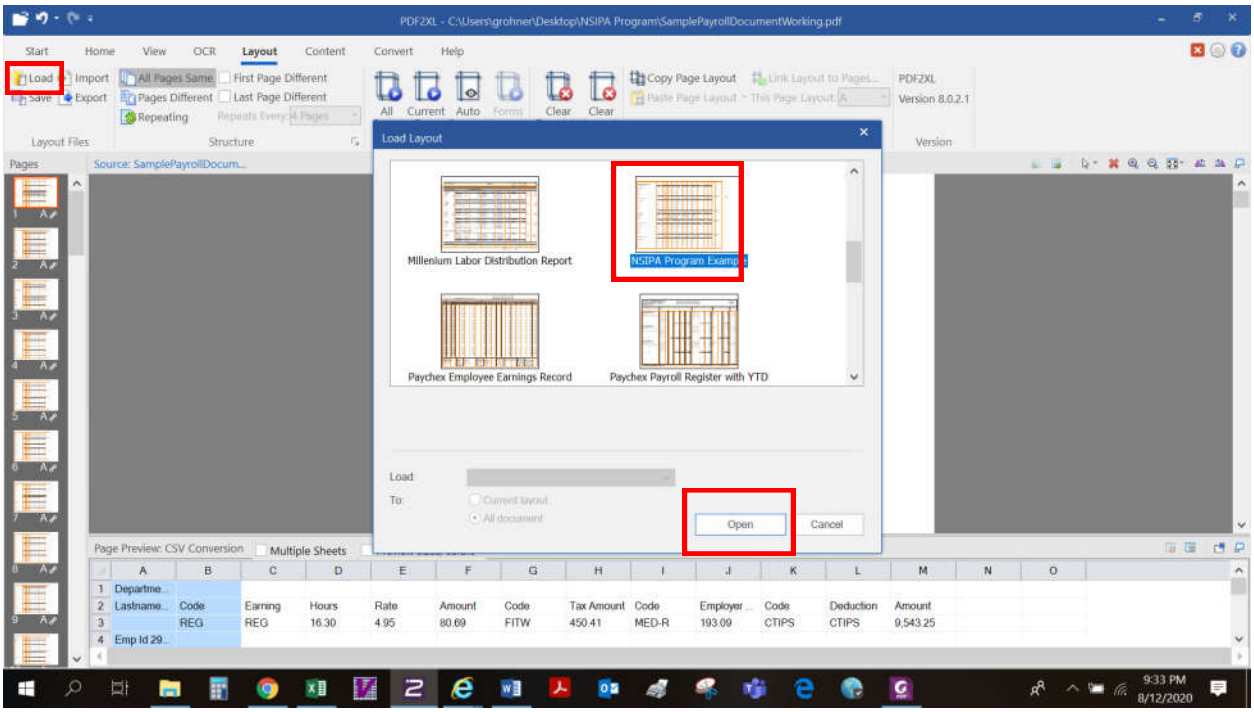
Once you get the data mapped to the desired position, you can save this layout as a template for use on future conversions. On the “Layout” tab, click “save”



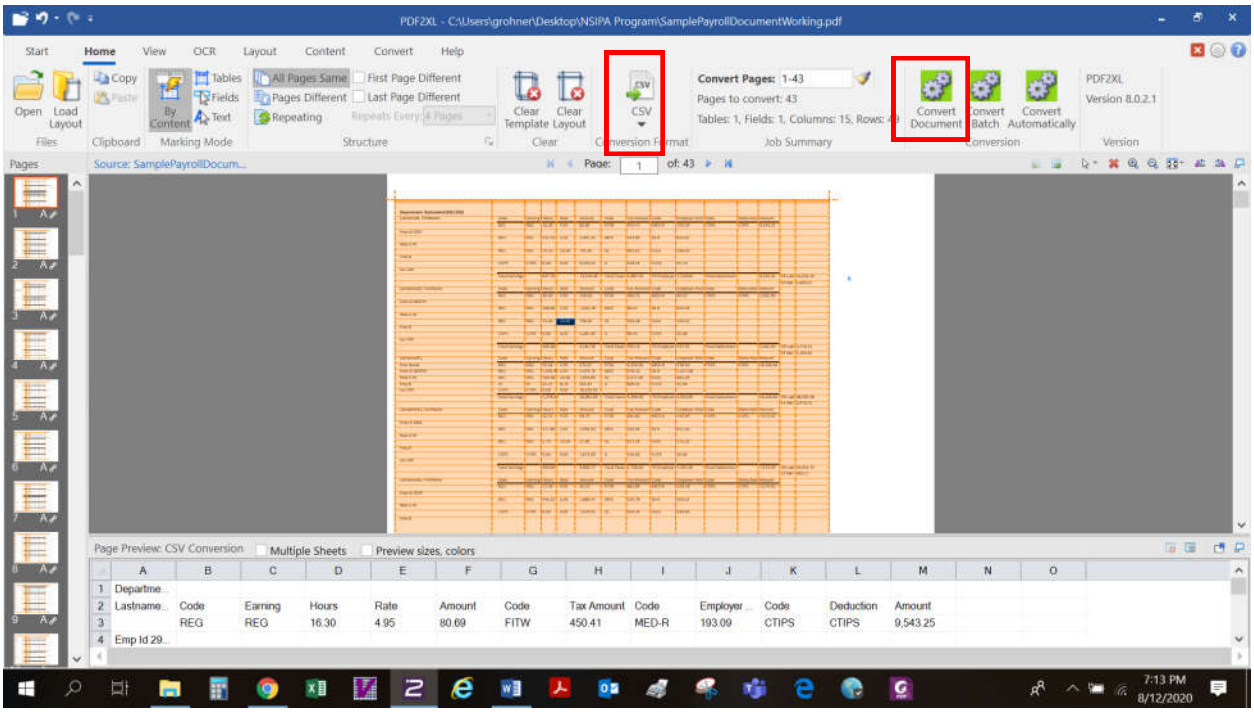
When you click “save” a dialog box comes up to name the template



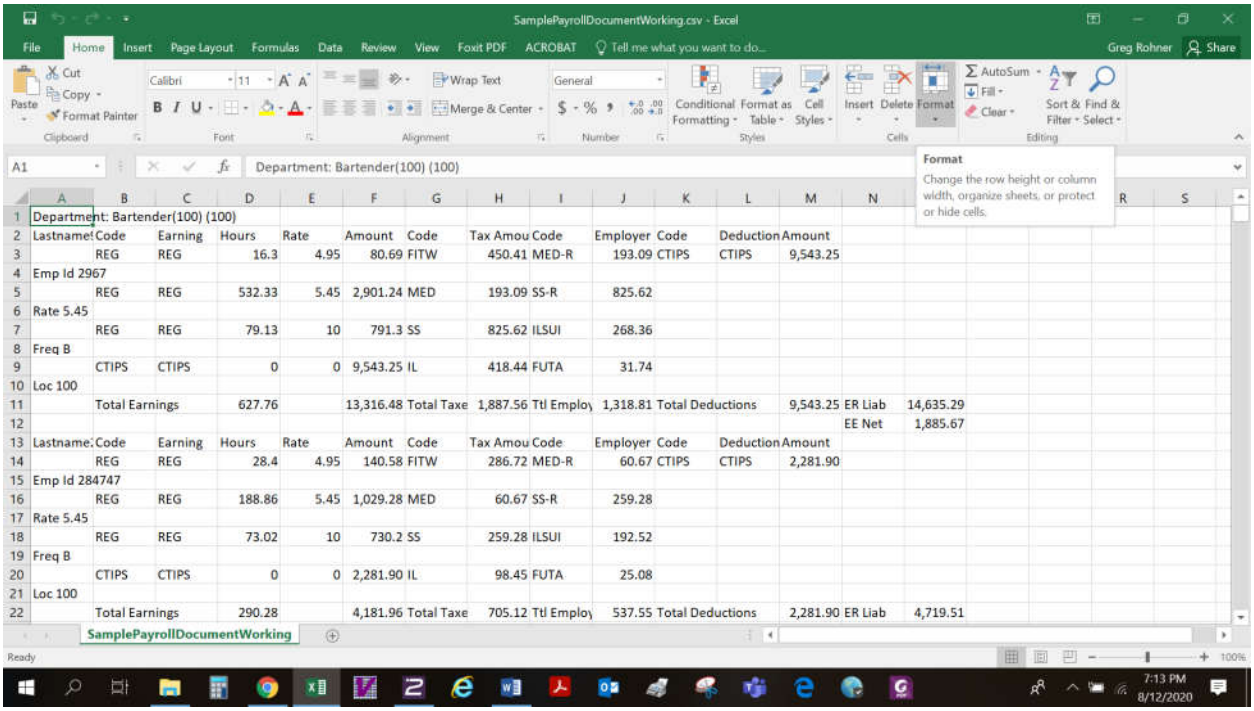
Layouts can be loaded by clicking the “load” option on the “layout” tab and selecting the desired layout



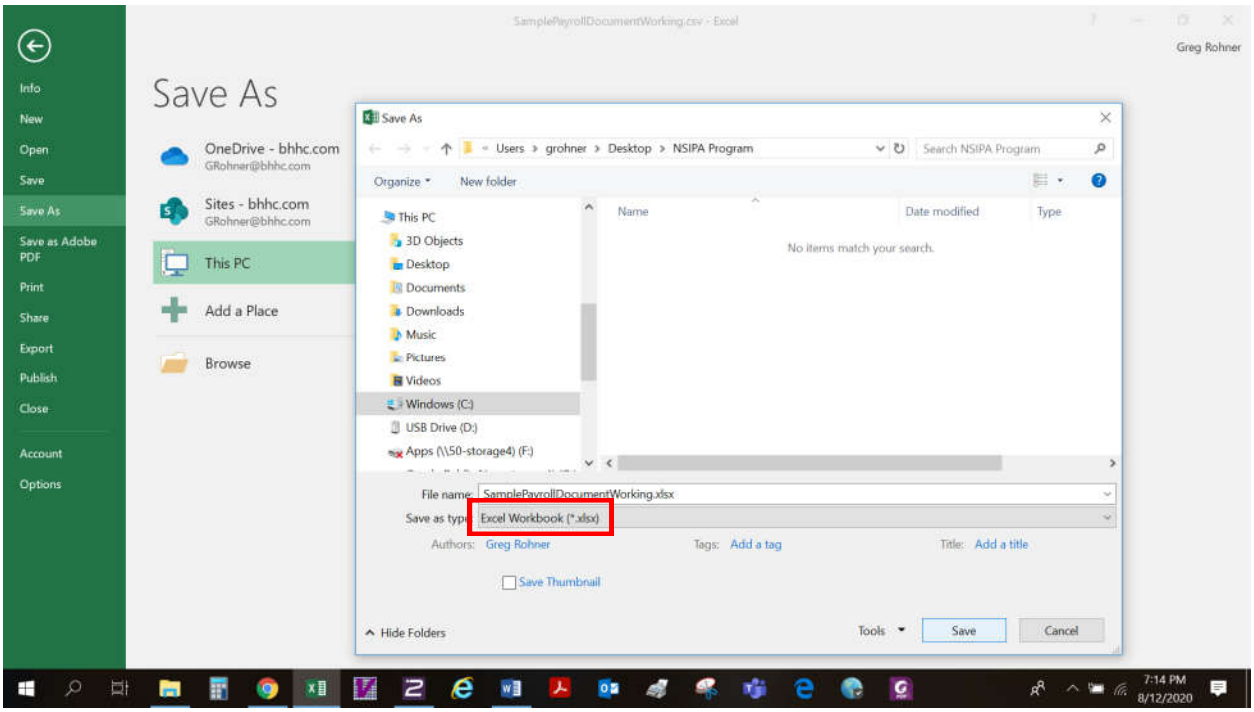
To do your conversion, ensure the “Conversion Format” is set to CSV and then click the “Convert Document” option.



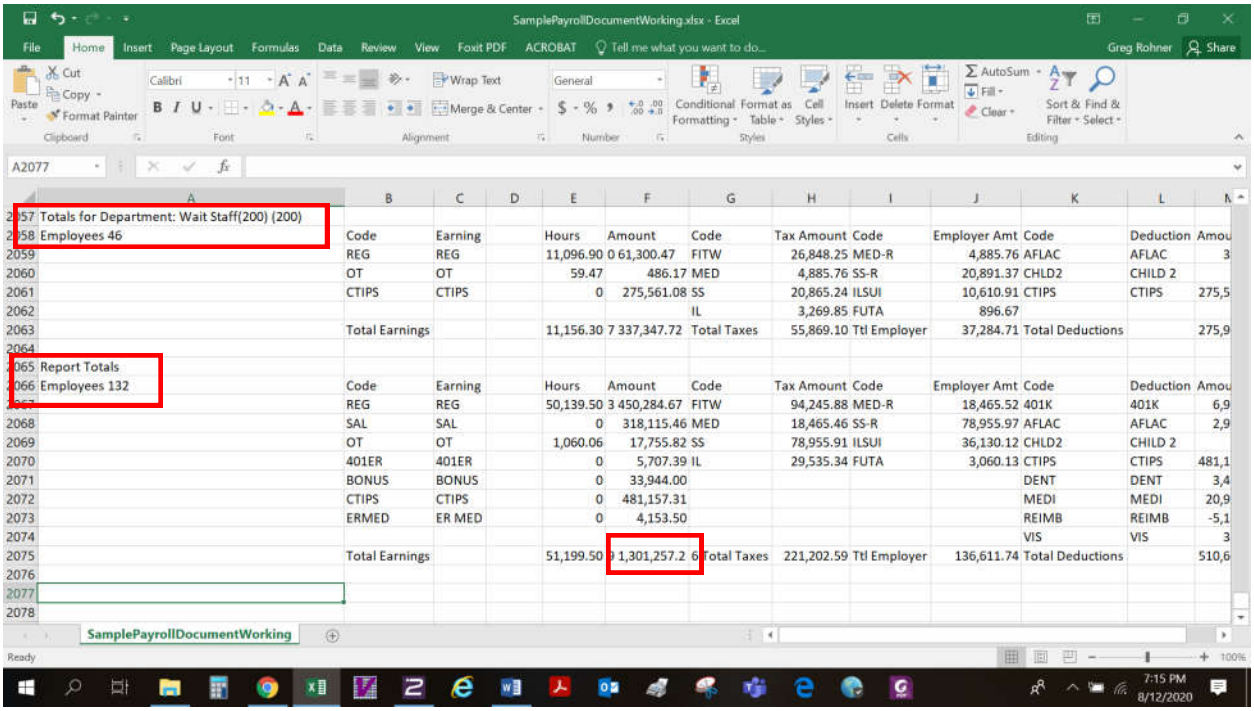
This is the output



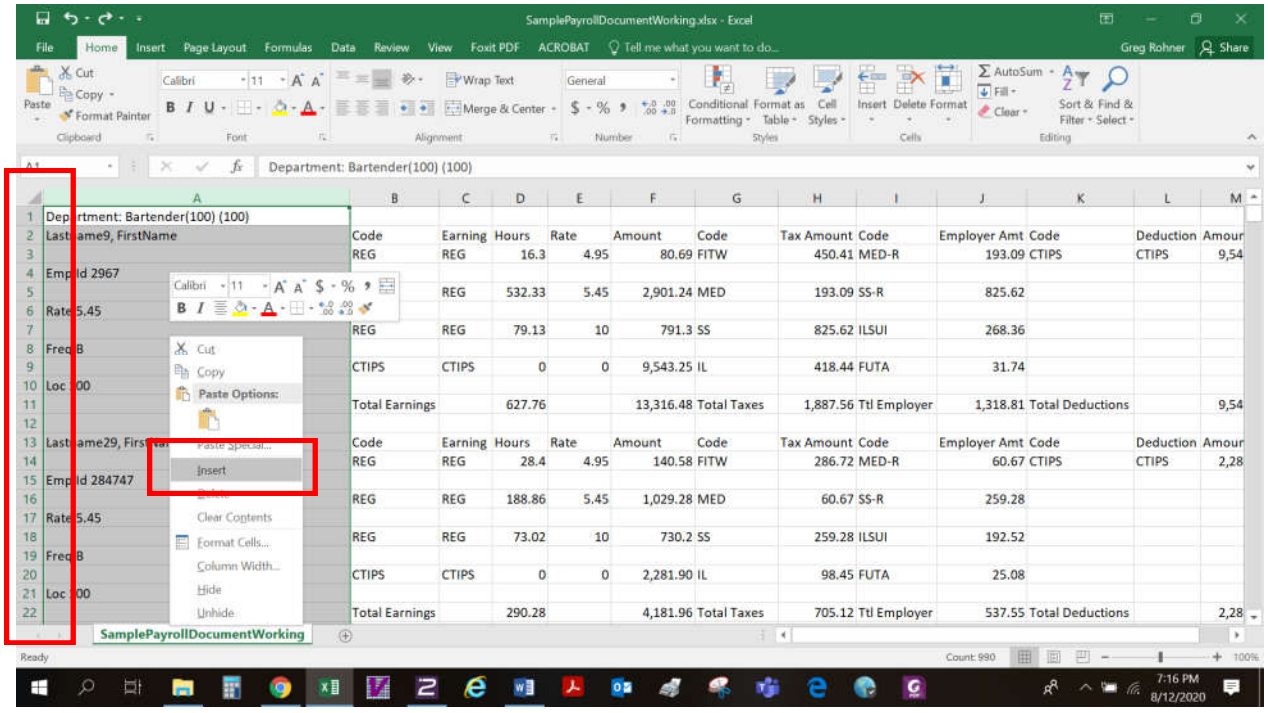
When working with CSV, it's important to save the file as an Excel file. This process involves using multiple pages on the workbook and CSV will only save 1 page whereas Excel files will save multiple pages.



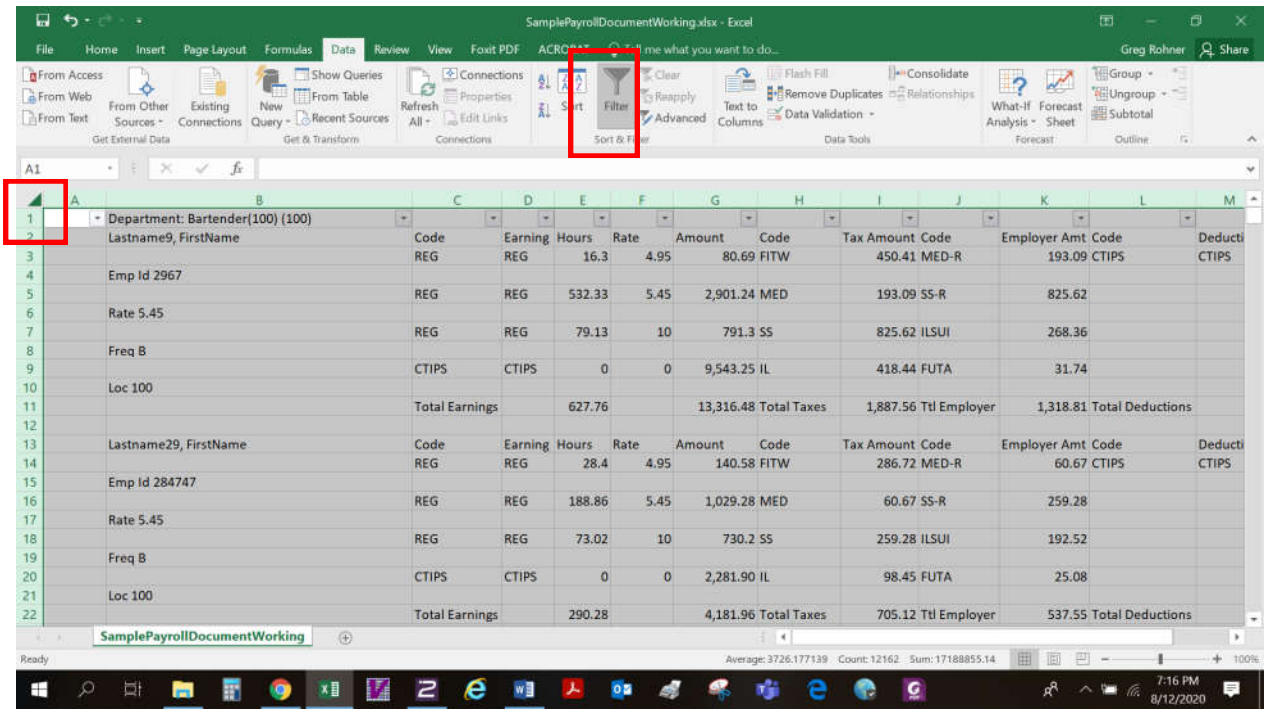
Once the report is converted, scroll to the end of the data. Take note of the totals (in many cases they may split between columns) and also note if there are subtotals imbedded in the report. Part of the data mining process is to ensure that those subtotals don't get included with your totals.



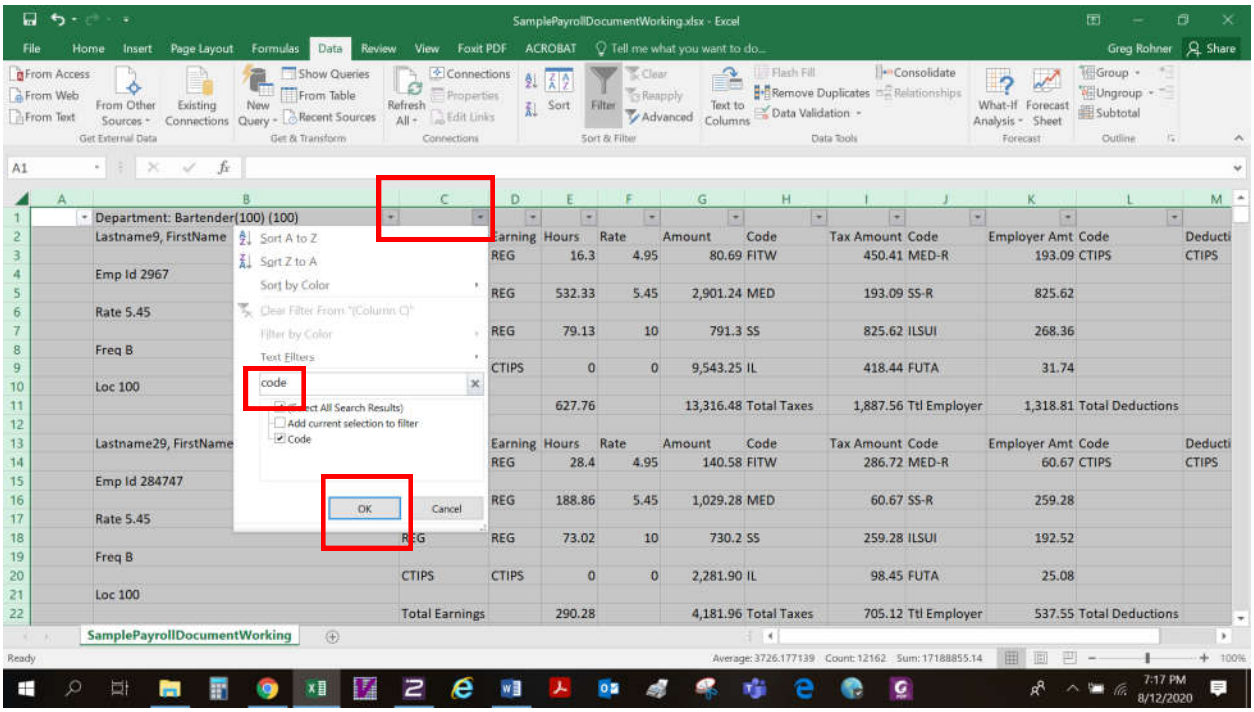
Insert a column on the left side of the data field.



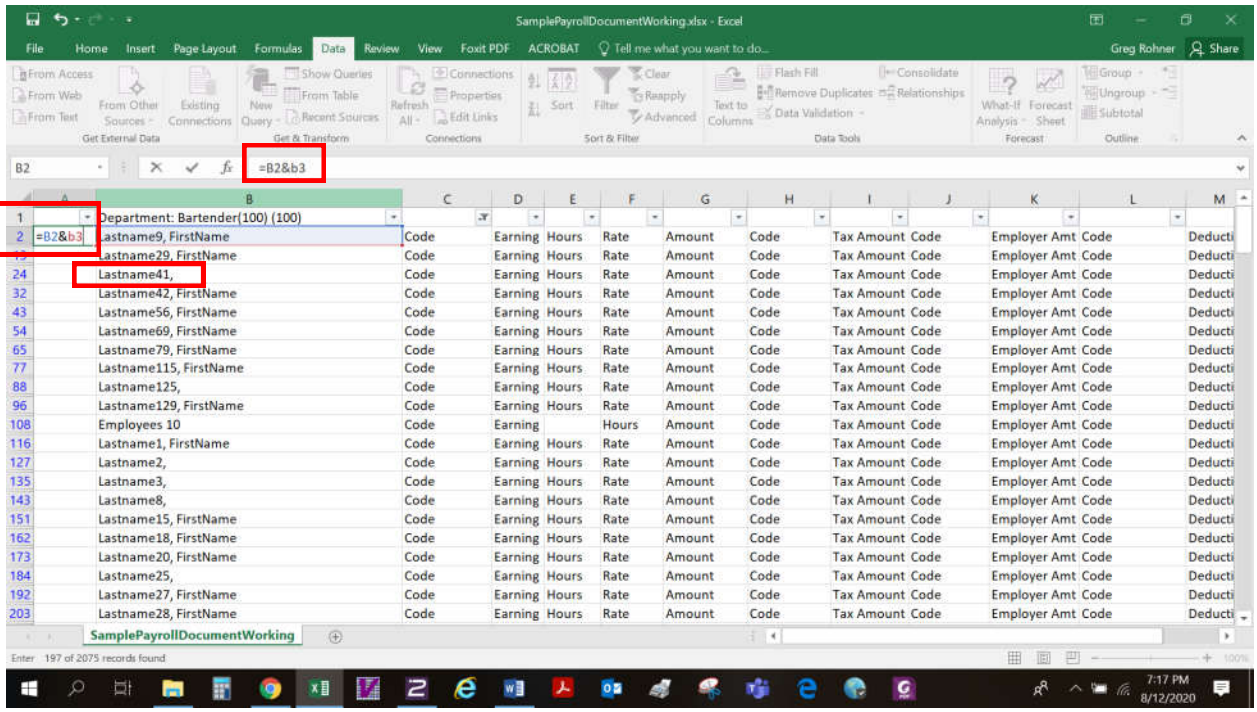
Highlight the entire sheet and click on “filter”



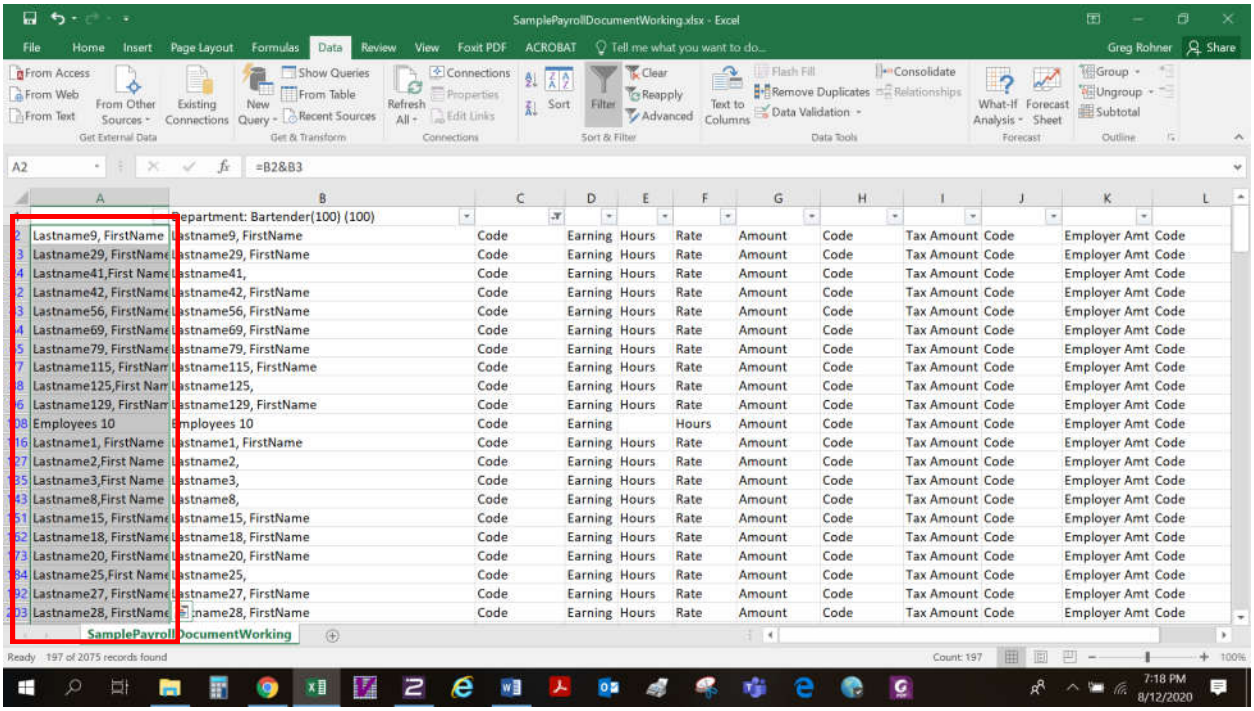
In column C on this example, click the drop down box on the filter and type the word “code” in the diaglog box. With this particular report, names will line up with the headers for each employee’s data



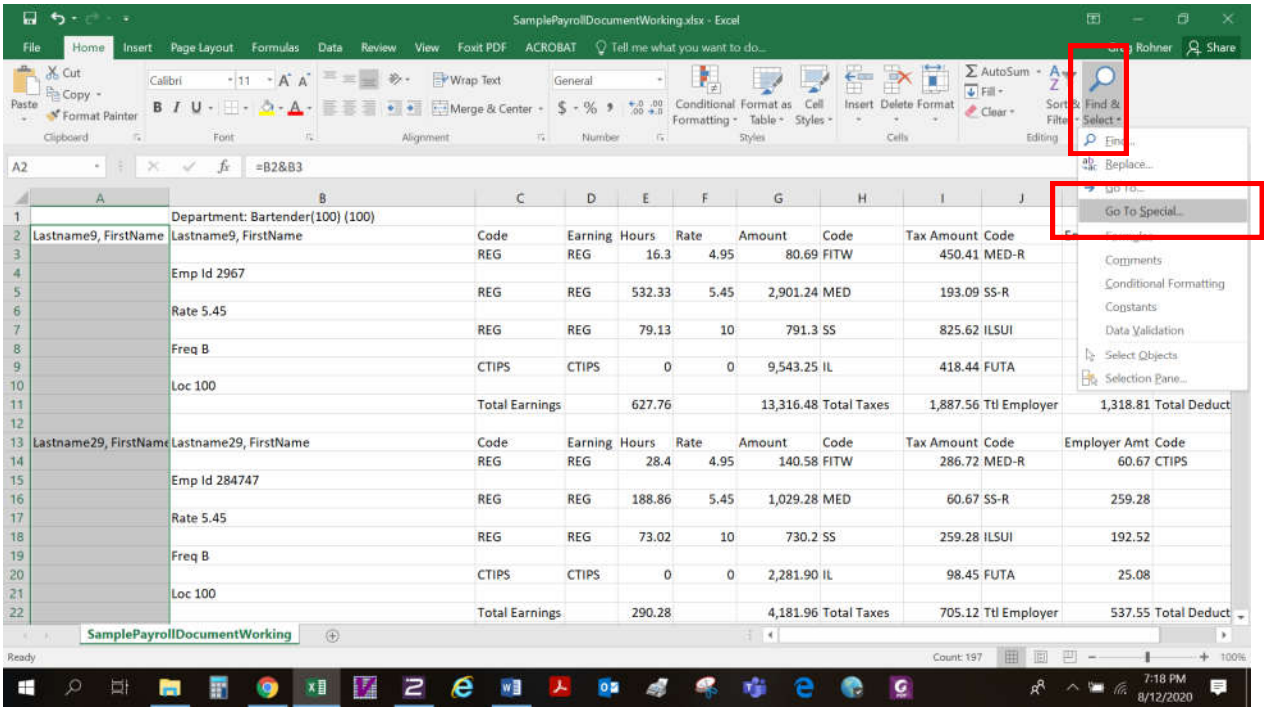
One characteristic of this report is that often longer names will carry to two lines. (see the example of “Lastname41”). In this example, you concatenate the name by combining the first row of the name and the second row of the name. In the example below, the formula =b2&b3 will accomplish that.



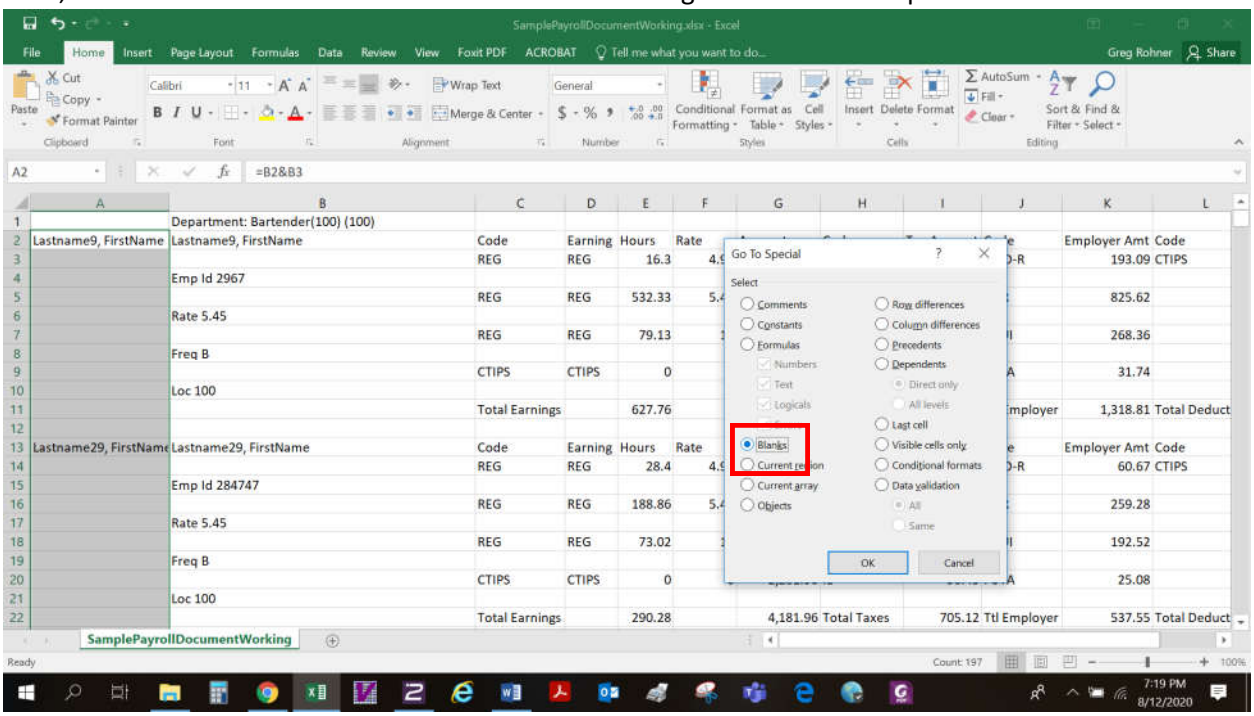
Copy your formula for the entire data field



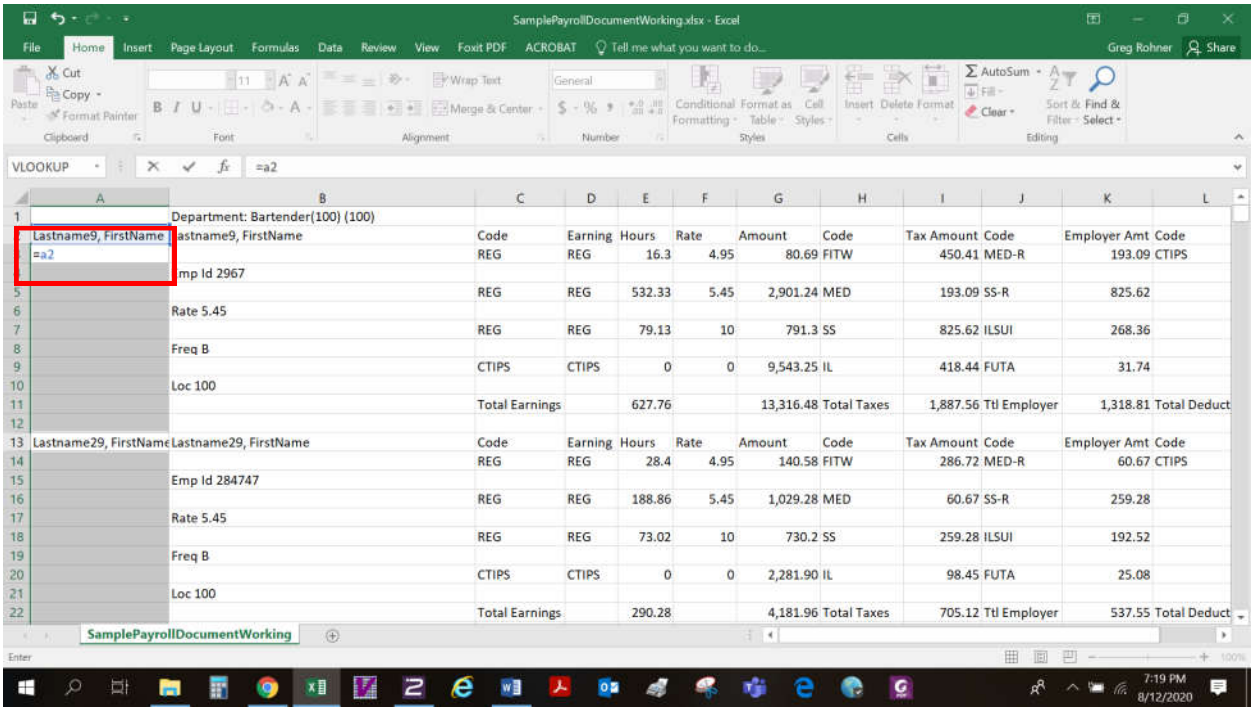
Release the filter and highlight the area (if it does not stay highlighted) from cell A2 to the end of the data set



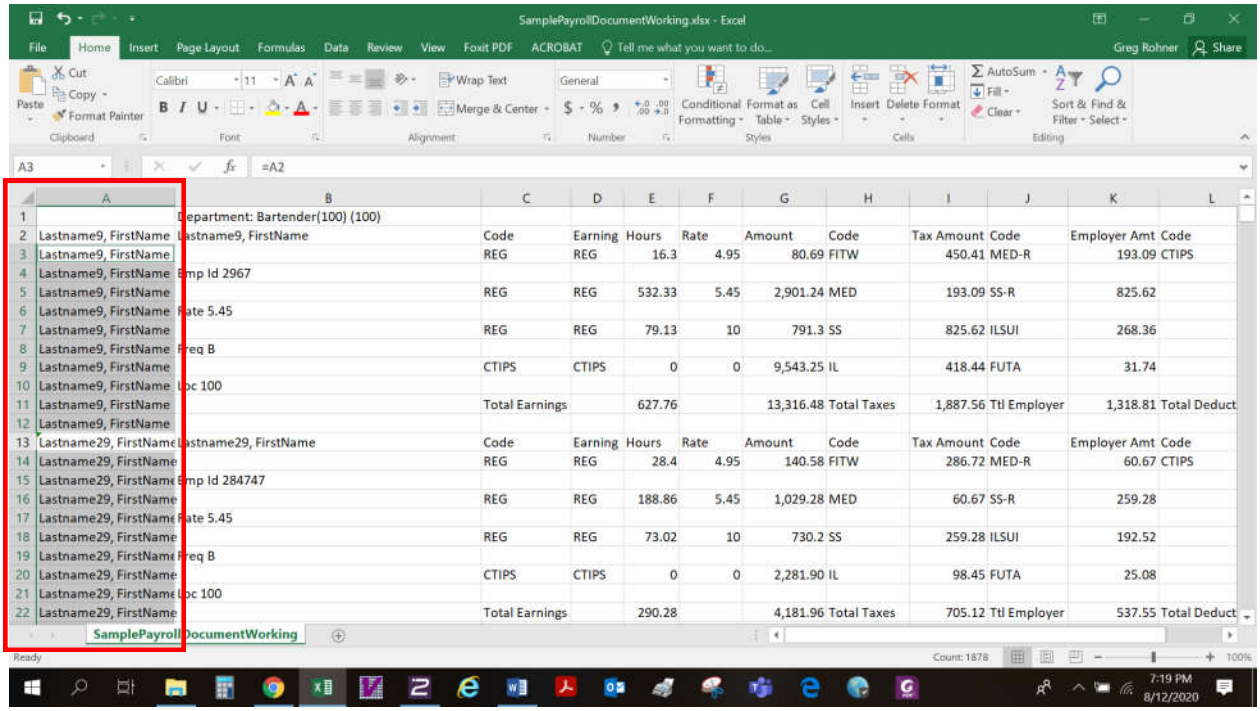
Next, click “find & select” and select “blanks” in the dialog box that comes up.



When you click “ok”, the first cell below the first with data will be where the cursor is. Type = and the cell coordinate immediately above (=a2 in this example) and hit the control and enter keys

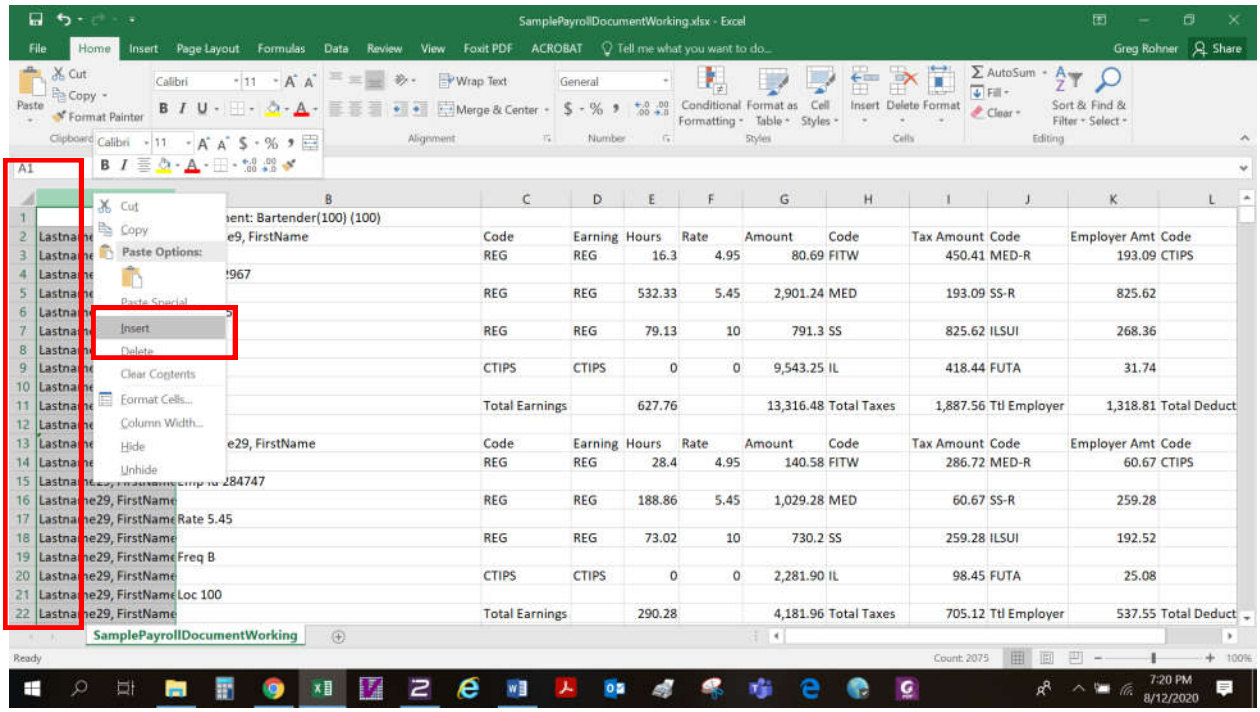


The end result is that you will have your employee names lined up with each piece of information for the subject employee

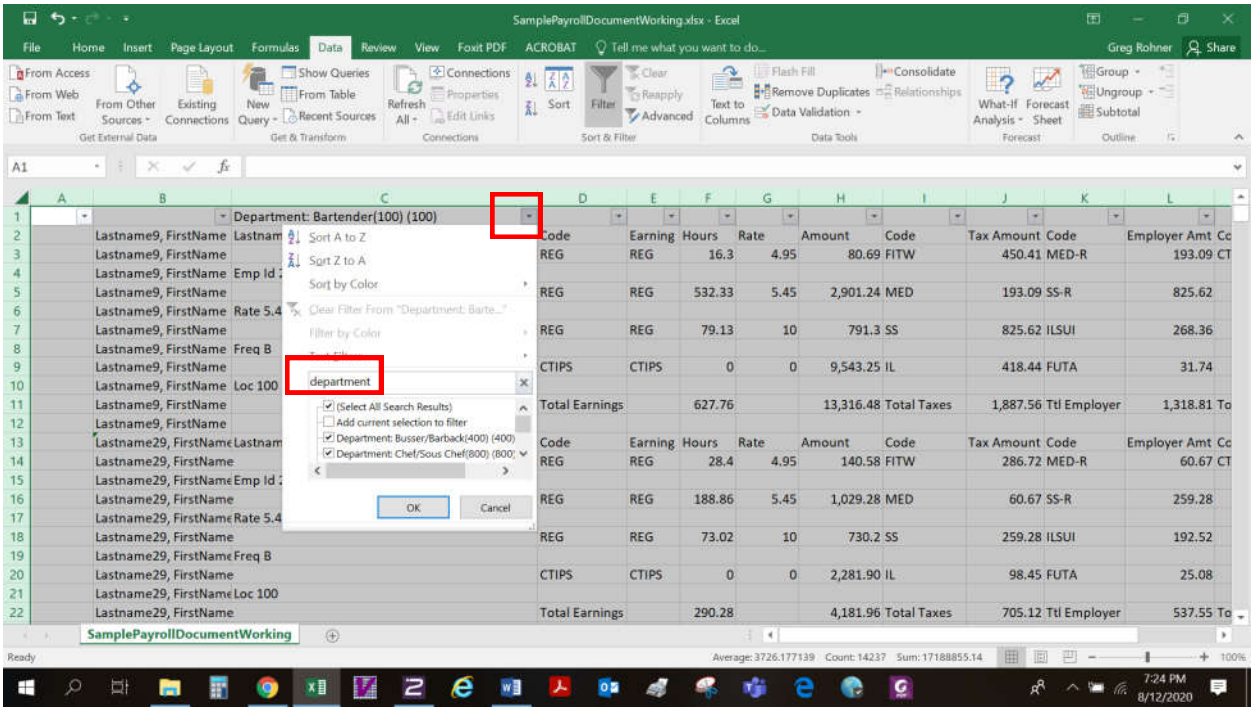


	A	B	C	D	E	F	G	H	I	J	K	L
1		Department: Bartender(100) (100)										
2	Lastname9, FirstName	Lastname9, FirstName	Code	Earning	Hours	Rate	Amount	Code	Tax Amount	Code	Employer Amt	Code
3	Lastname9, FirstName		REG	REG	16.3	4.95	80.69	FITW	450.41	MED-R	193.09	CTIPS
4	Lastname9, FirstName	Emp Id 2967										
5	Lastname9, FirstName		REG	REG	532.33	5.45	2,901.24	MED	193.09	SS-R	825.62	
6	Lastname9, FirstName	Rate 5.45										
7	Lastname9, FirstName		REG	REG	79.13	10	791.3	SS	825.62	ILSUI	268.36	
8	Lastname9, FirstName	Freq B										
9	Lastname9, FirstName		CTIPS	CTIPS	0	0	9,543.25	IL	418.44	FUTA	31.74	
10	Lastname9, FirstName	Loc 100										
11	Lastname9, FirstName		Total Earnings		627.76		13,316.48	Total Taxes	1,887.56	Ttl Employer	1,318.81	Total Deduct
12	Lastname9, FirstName											
13	Lastname29, FirstName	Lastname29, FirstName	Code	Earning	Hours	Rate	Amount	Code	Tax Amount	Code	Employer Amt	Code
14	Lastname29, FirstName		REG	REG	28.4	4.95	140.58	FITW	286.72	MED-R	60.67	CTIPS
15	Lastname29, FirstName	Emp Id 284747										
16	Lastname29, FirstName		REG	REG	188.86	5.45	1,029.28	MED	60.67	SS-R	259.28	
17	Lastname29, FirstName	Rate 5.45										
18	Lastname29, FirstName		REG	REG	73.02	10	730.2	SS	259.28	ILSUI	192.52	
19	Lastname29, FirstName	Freq B										
20	Lastname29, FirstName		CTIPS	CTIPS	0	0	2,281.90	IL	98.45	FUTA	25.08	
21	Lastname29, FirstName	Loc 100										
22	Lastname29, FirstName		Total Earnings		290.28		4,181.96	Total Taxes	705.12	Ttl Employer	537.55	Total Deduct

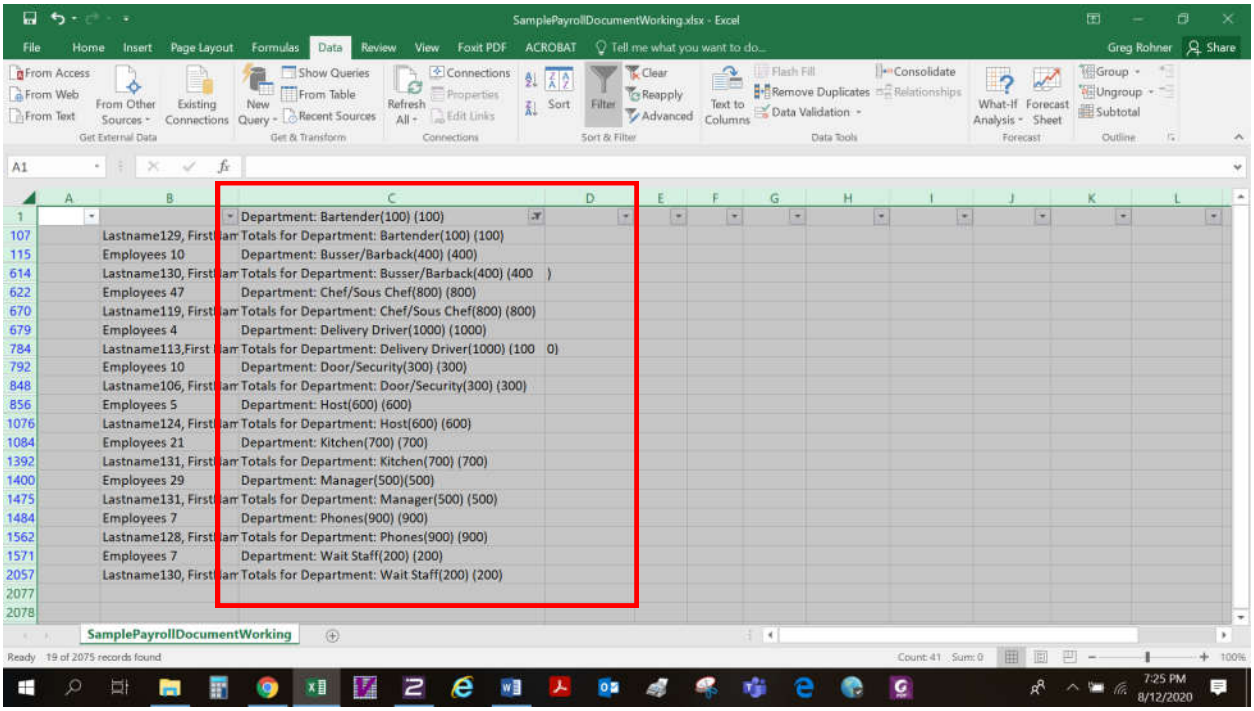
Because this report is segregated by department, bringing this information into our data mining process will help expedite the classification process by mining this information as well. Insert another column on the left side of the data set



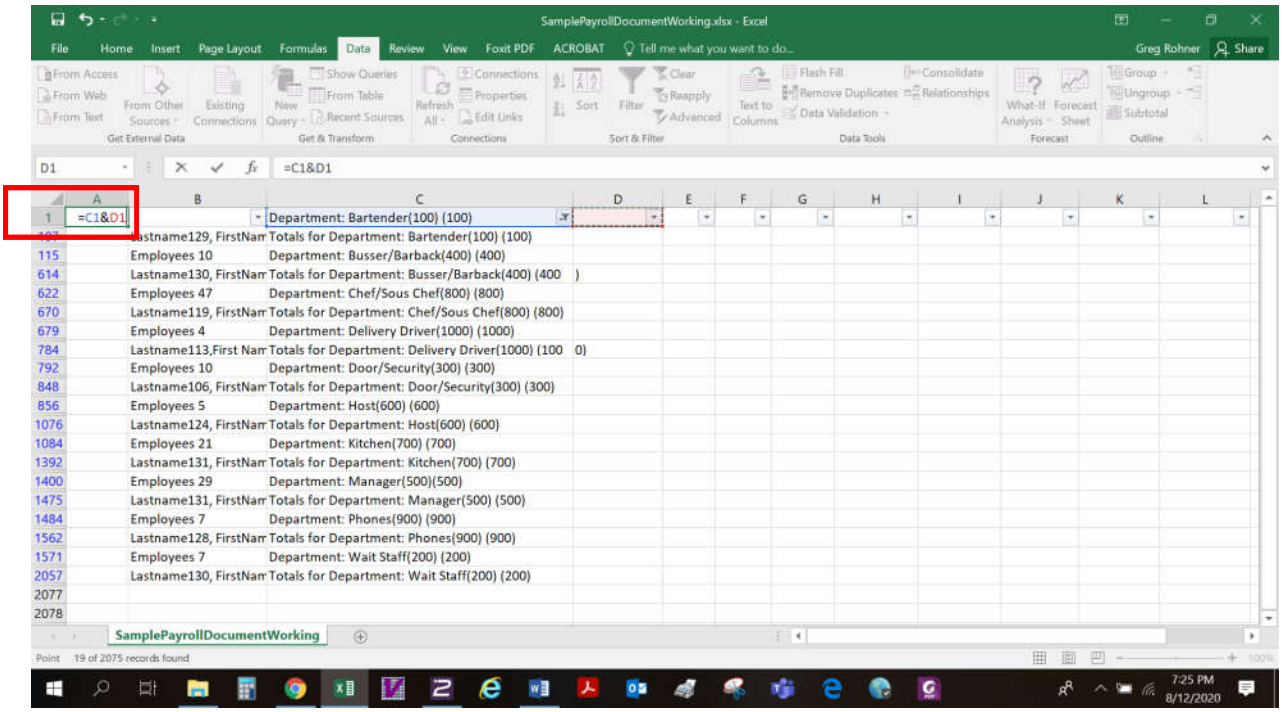
Highlight the entire spreadsheet and click on “filter” to put filters on the dataset. In column C, select the dropdown box and type “department” in the dialog box



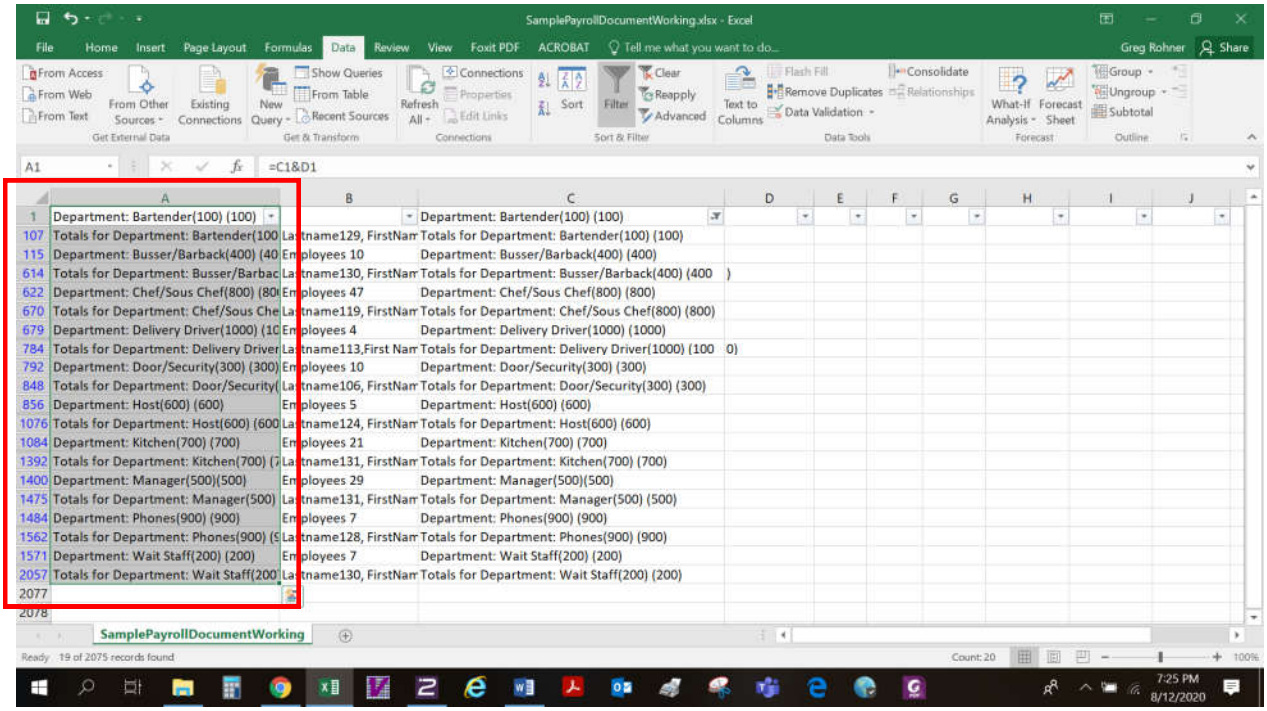
This will filter all of the department names, some of which may be split between columns



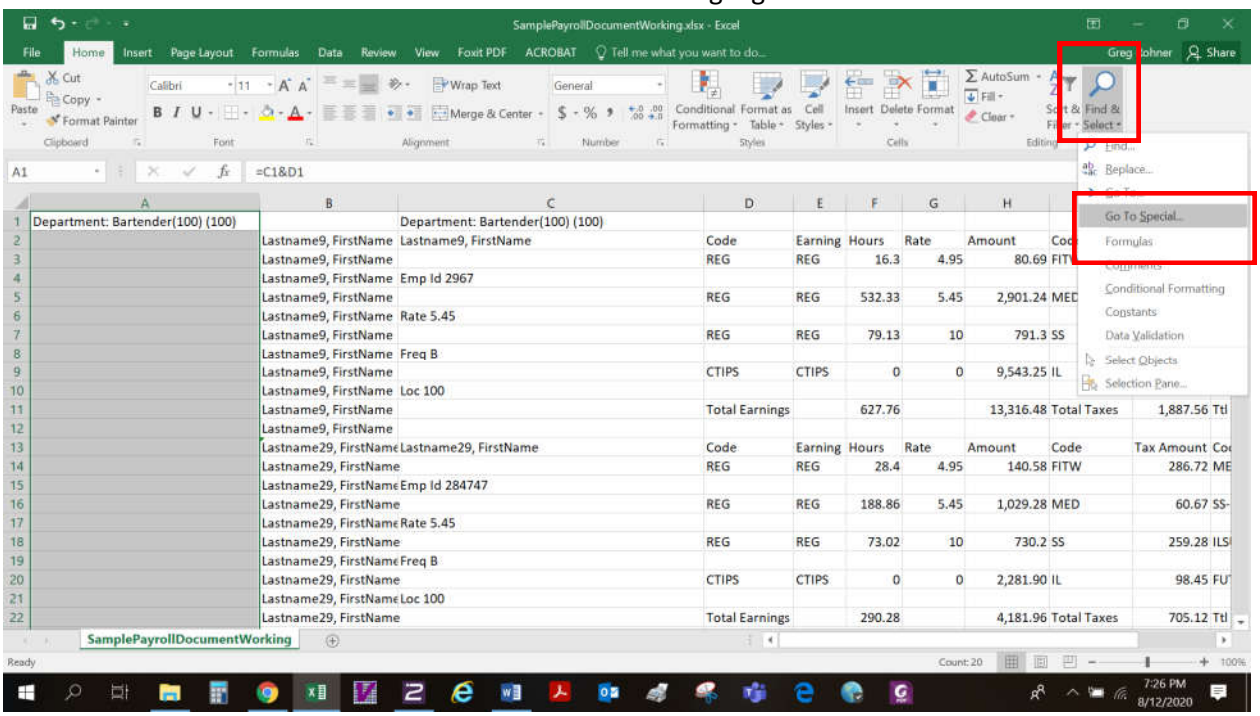
In cell A1, concatenate the department name information (in this example, =c1&d1) and copy this formula down



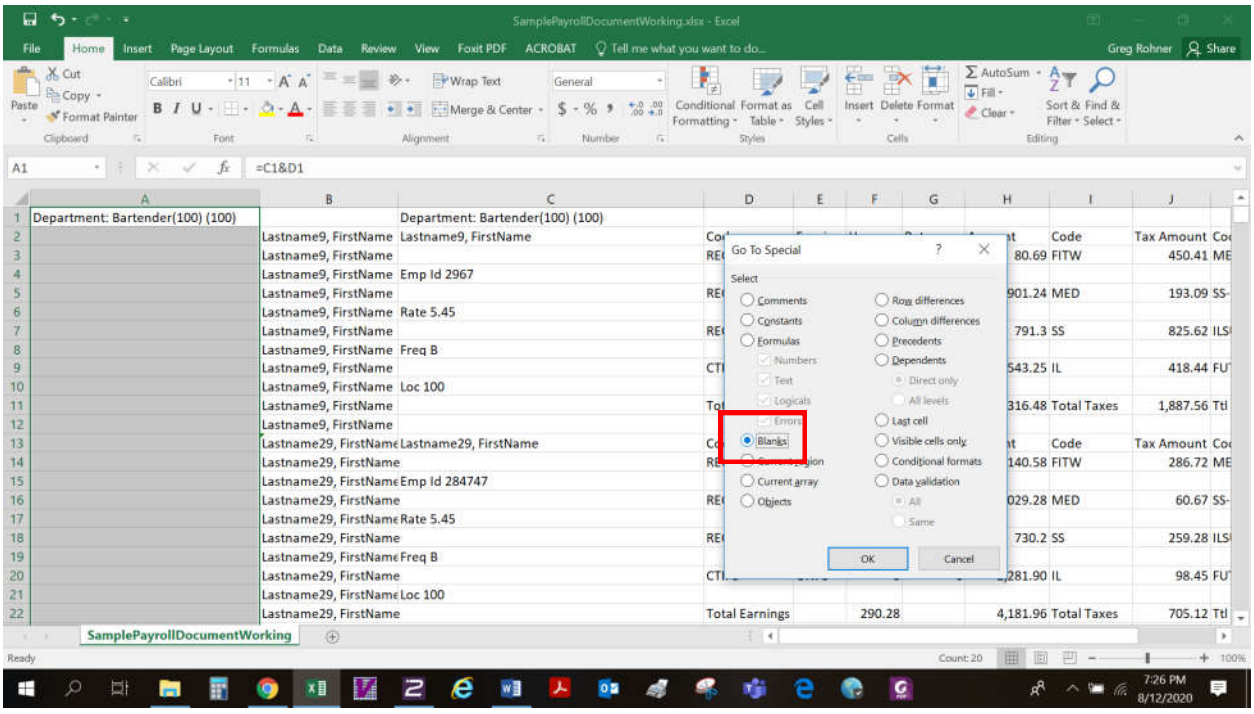
This is the result



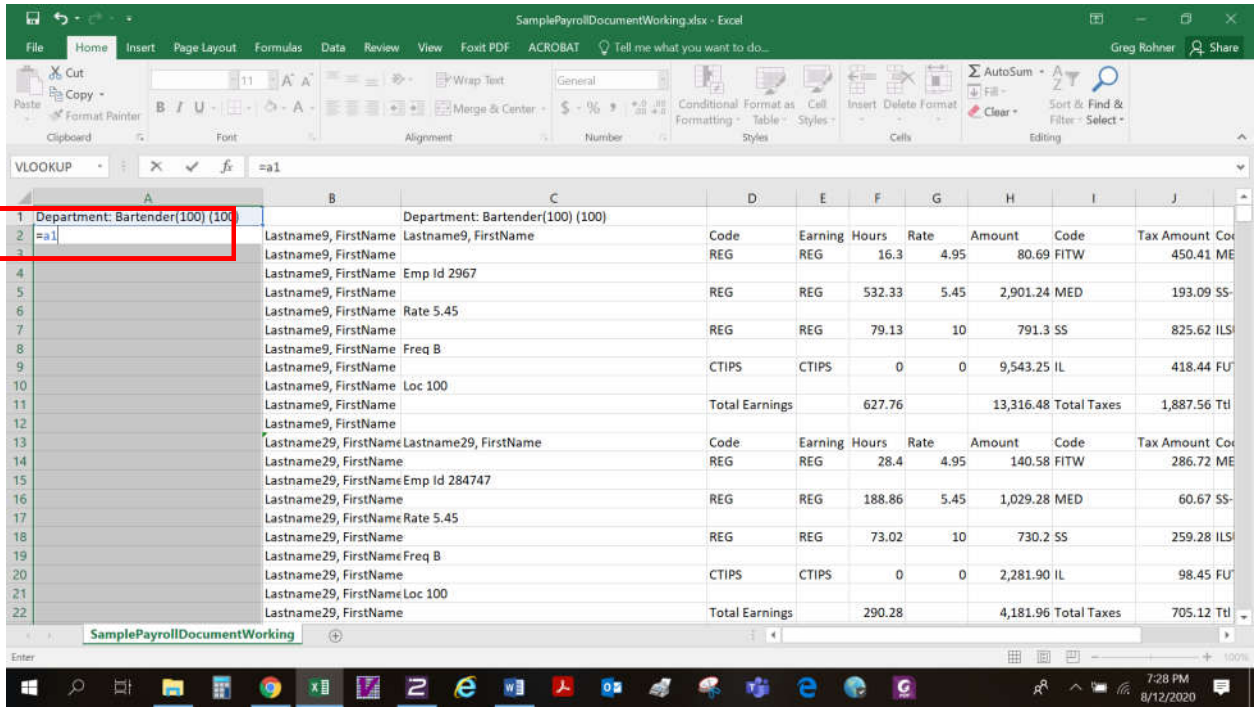
Release the filter and ensure that the data in column A is highlighted



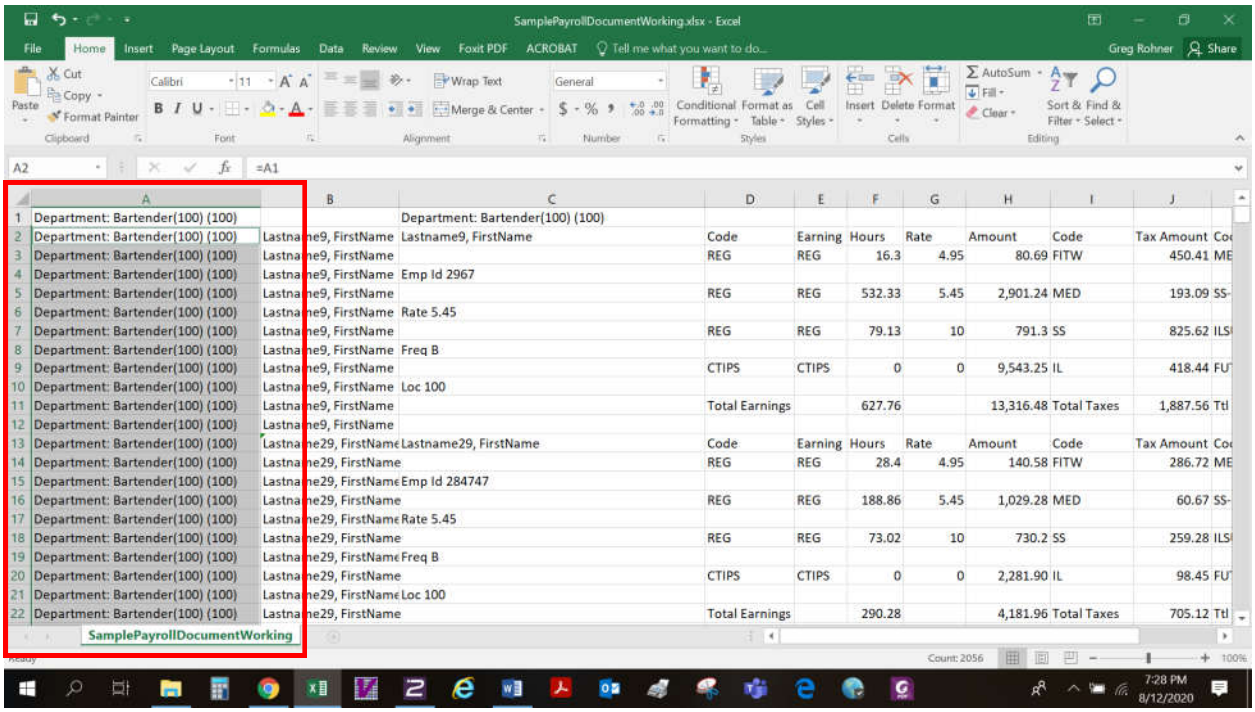
Click “find & select” and select “blanks”



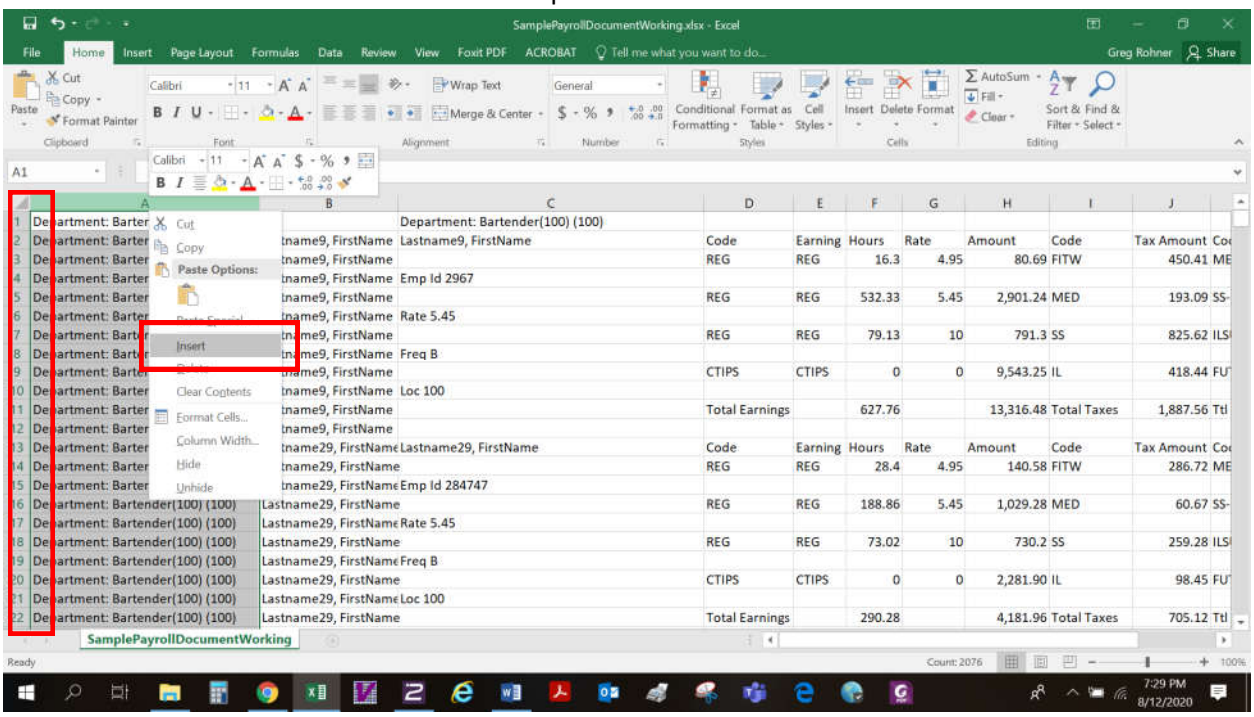
In the first open cell of that field, type the address of the cell above with data (in this example =a1) and hit the control and enter keys



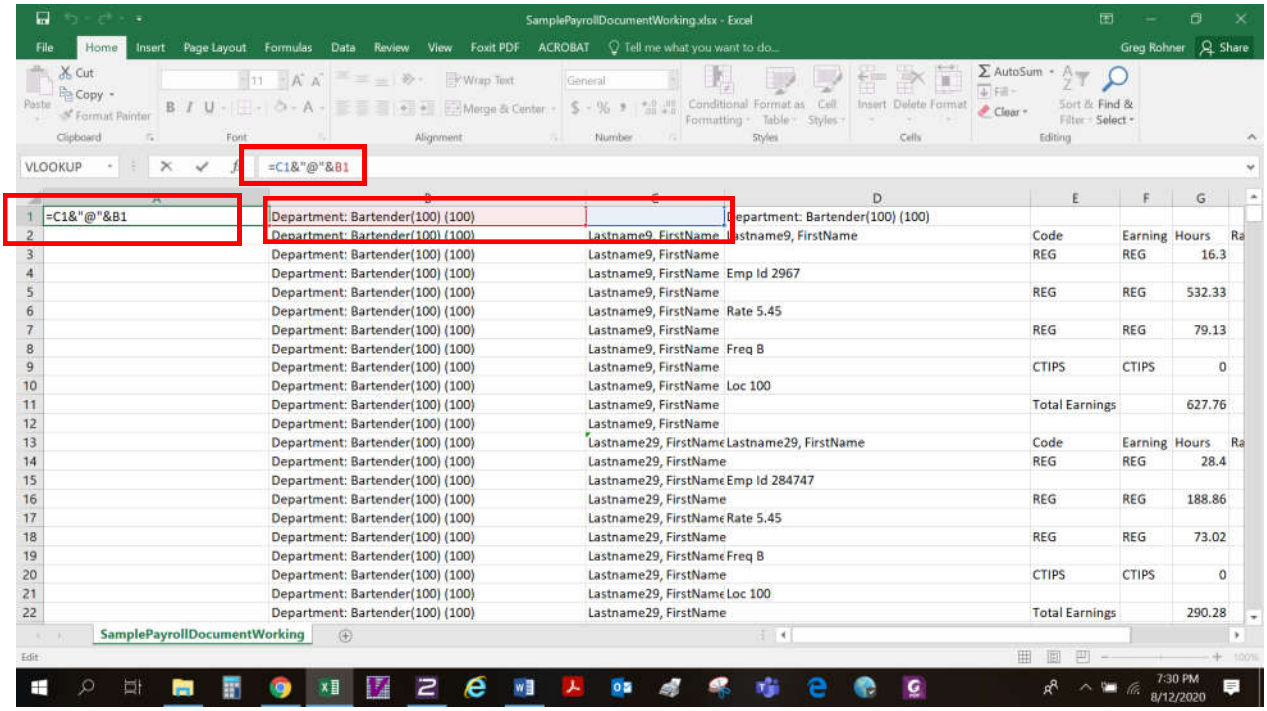
This is the result



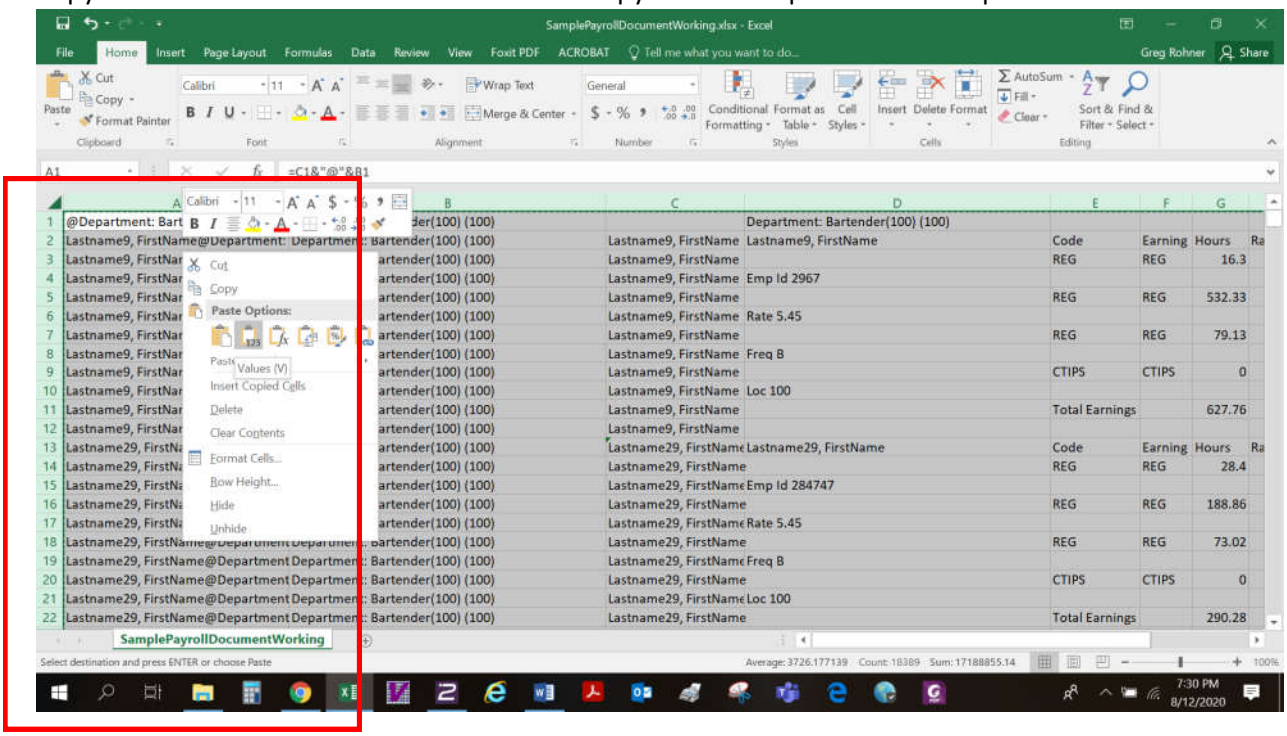
Insert another column on the left side of the spreadsheet



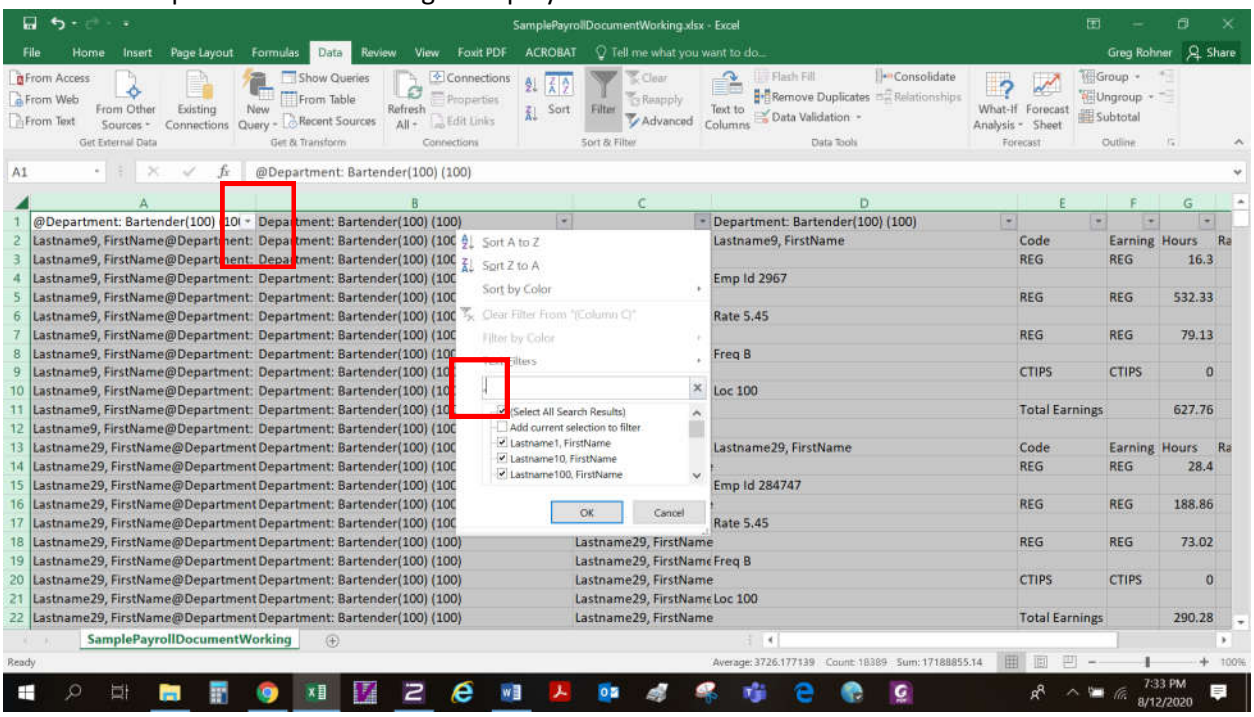
Concatenate the name and department number in the newly added column separating the two with the @ sign. You do this in order to separate this information later in the process. In the example below, the formula =c1&"@"&b1 will do this



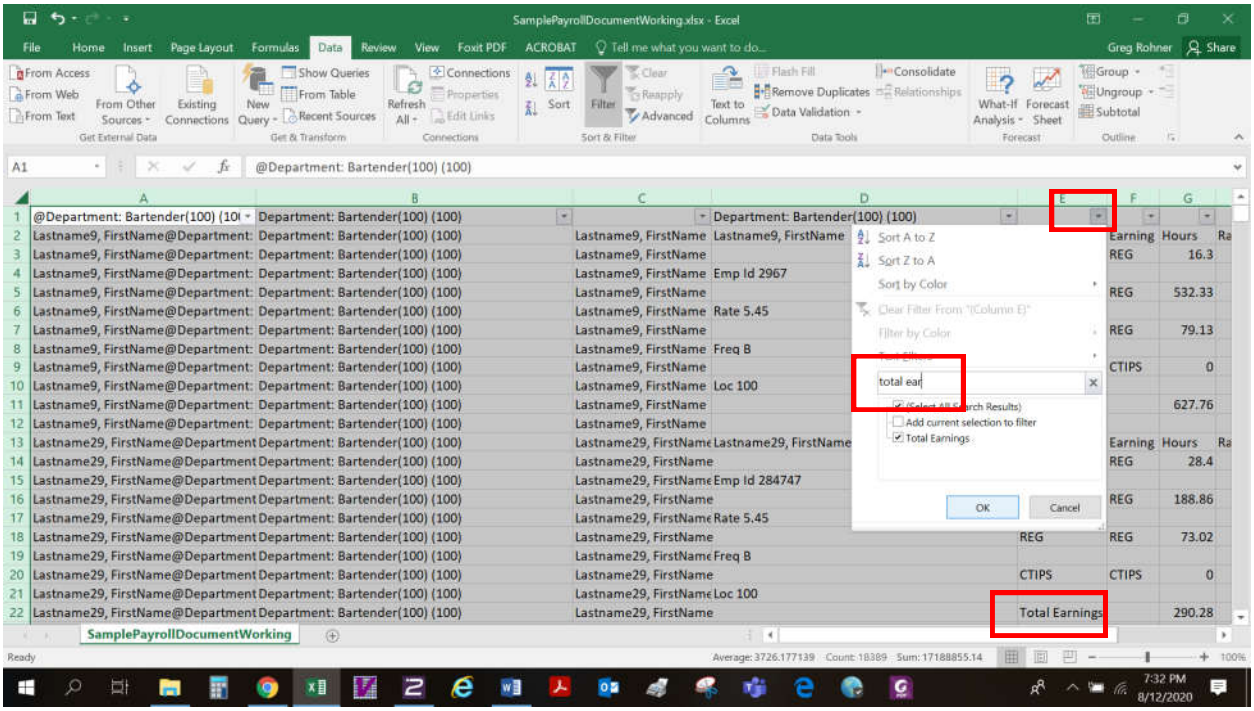
Copy that formula for the entire data set then copy the entire spreadsheet and paste as a value.



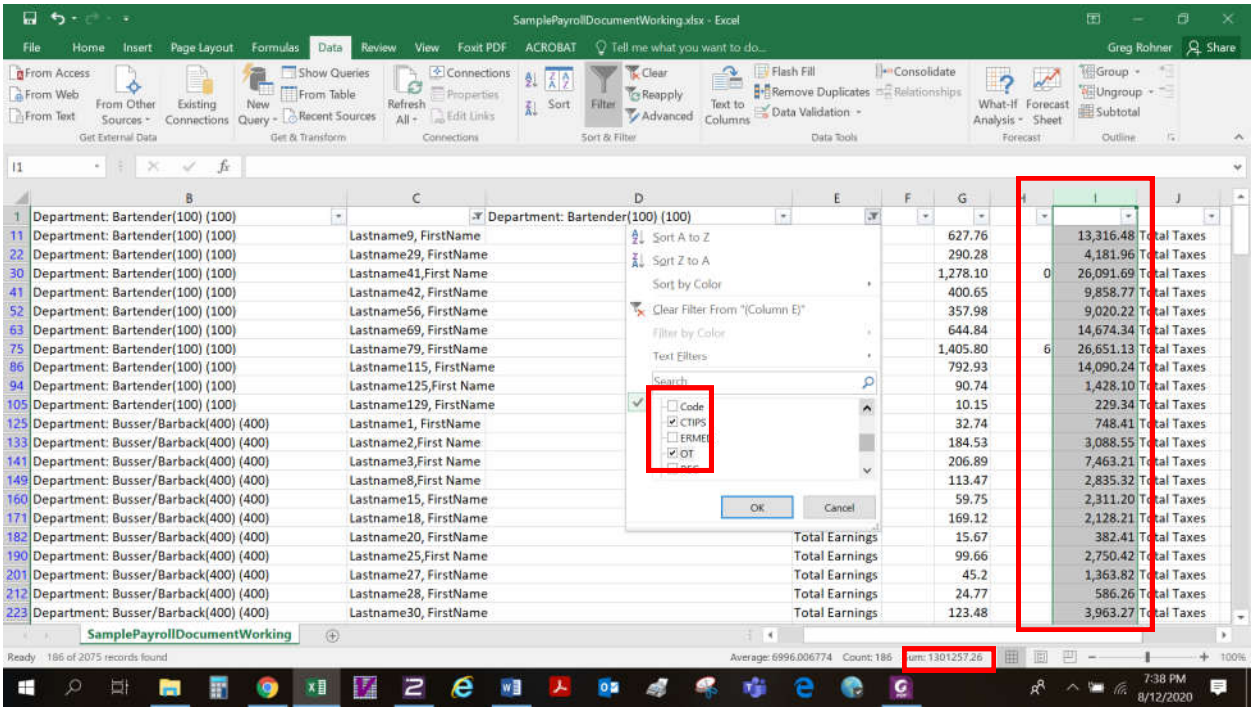
At this point, you will filter out the subtotals. This can be done either by entering a comma as in the example below in the dialog box and hitting ok. Also, in the text filters options, you can use the "does not include" option with the verbiage "employee" in it as that identifies the subtotals.



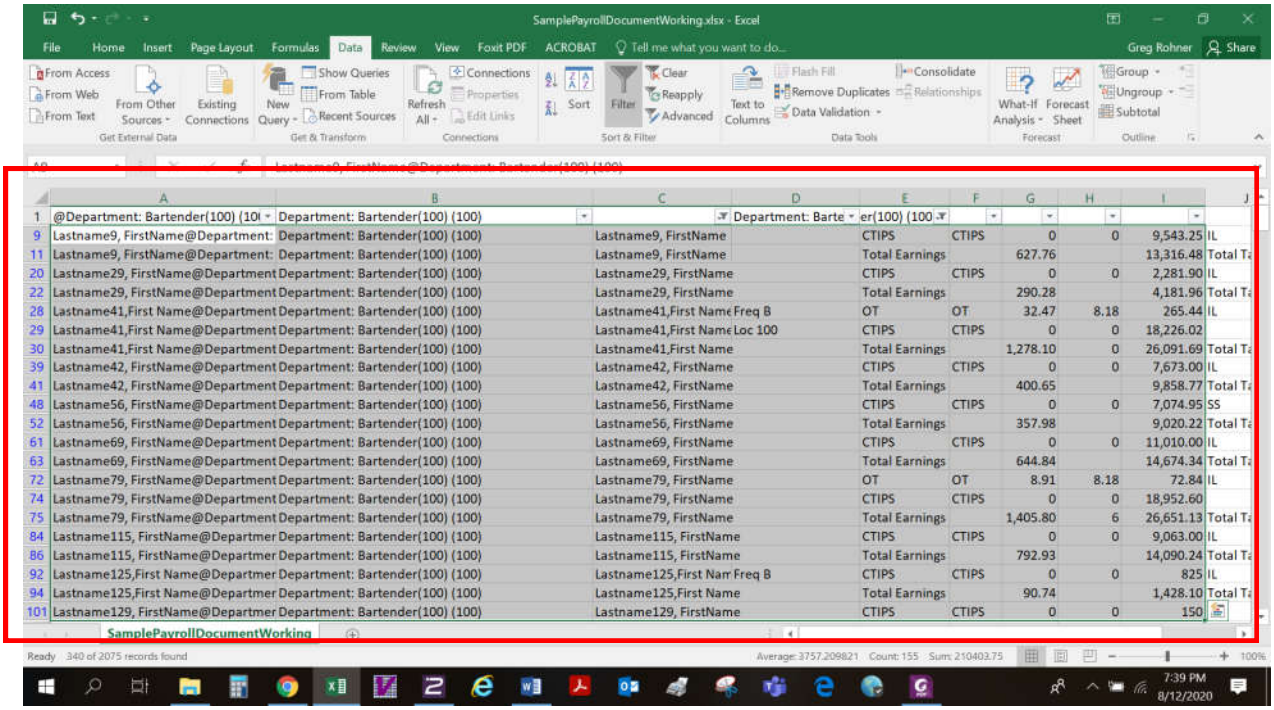
Click on the dropdown box for column E and type the word total earnings to filter just the values for the total earnigns to be found in column I



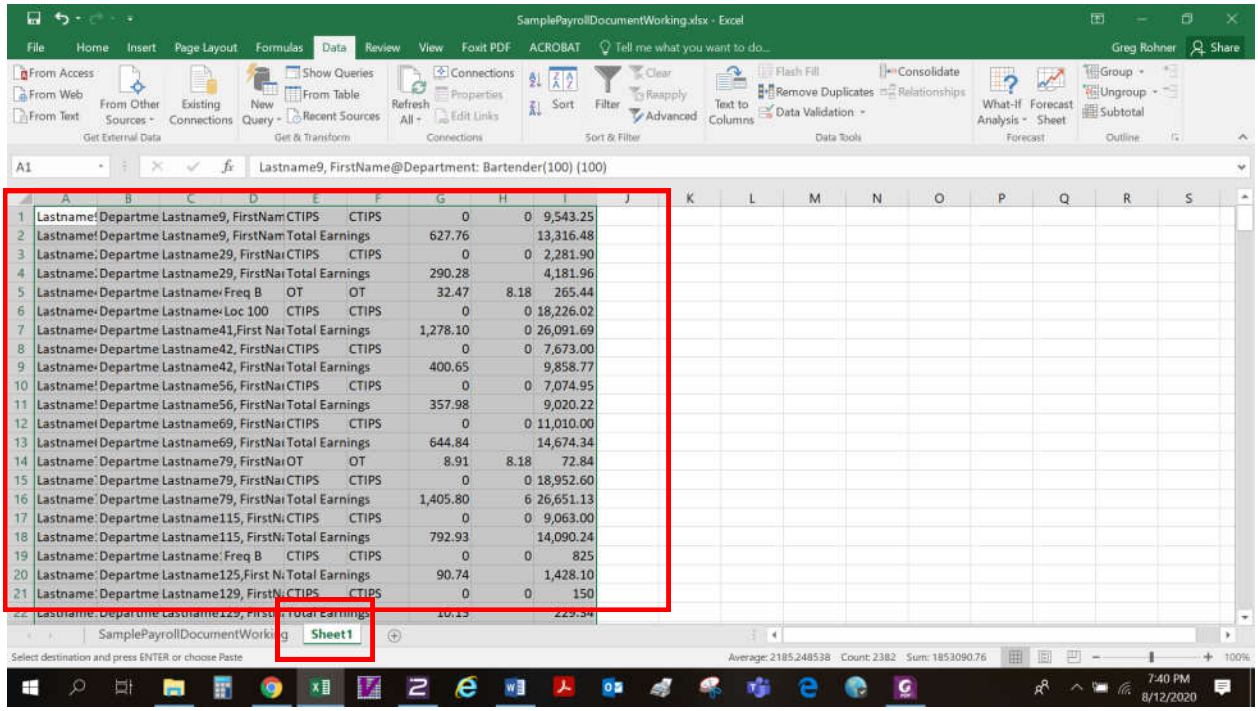
Highlight the values in Column I. If the subtotal at the bottom matches the total earnings for the report, you will know that you have the data converted and identified correctly



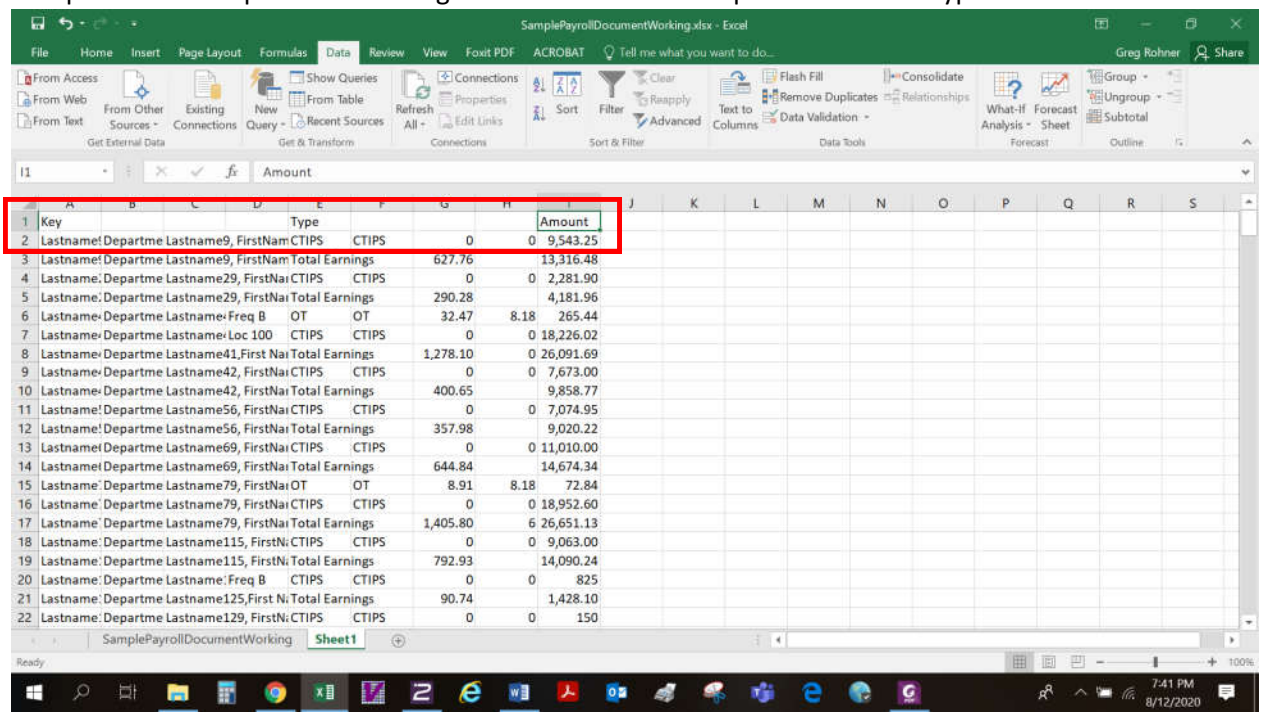
At this point, from the drop down box in Column E, select the items you want to extract. In this example, that would be “Total Earnings”, tips “CTIPS” and overtime denoted as “OT”. Hit control C to copy this data



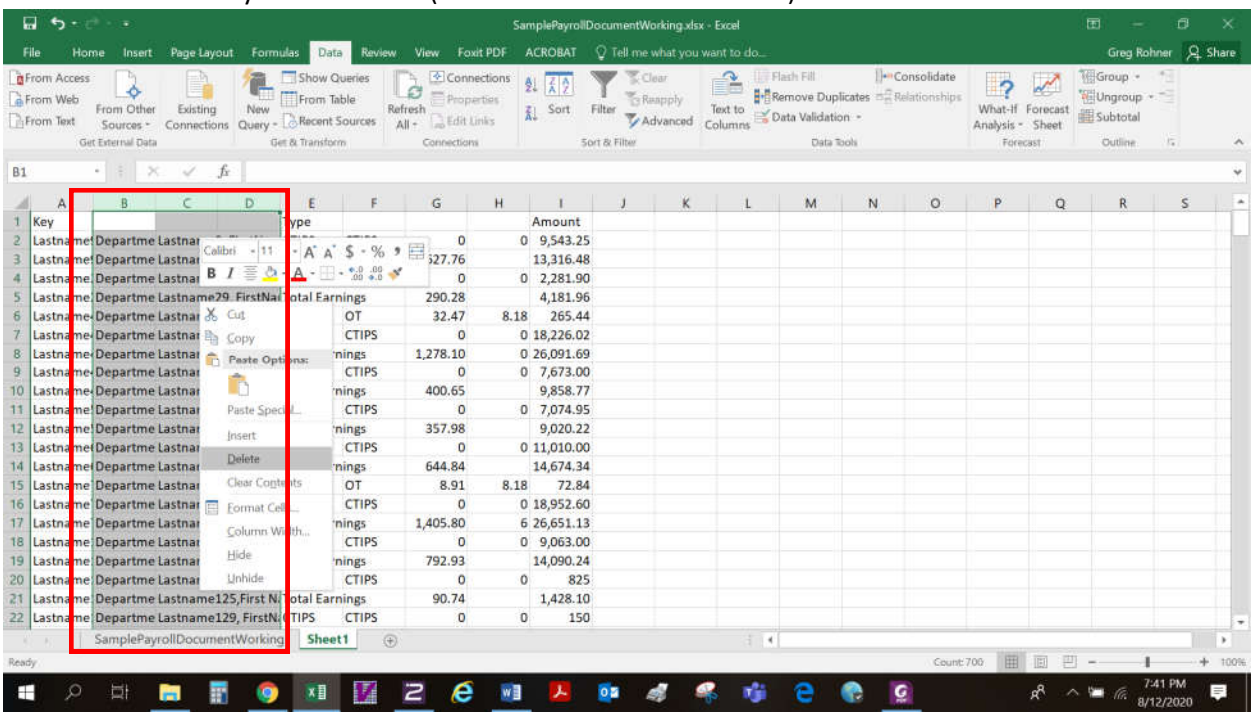
Paste this data into a new page on your workbook



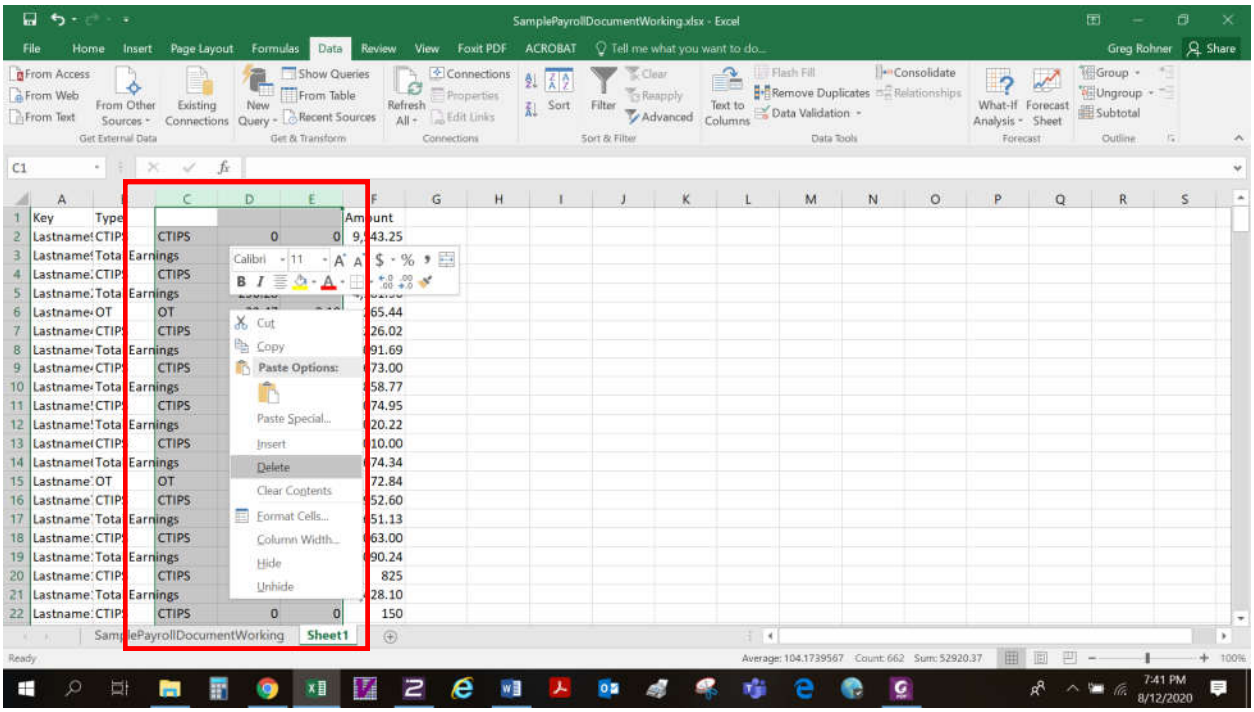
This process uses a pivot table to organize the data. That requires headers. Type headers as below.



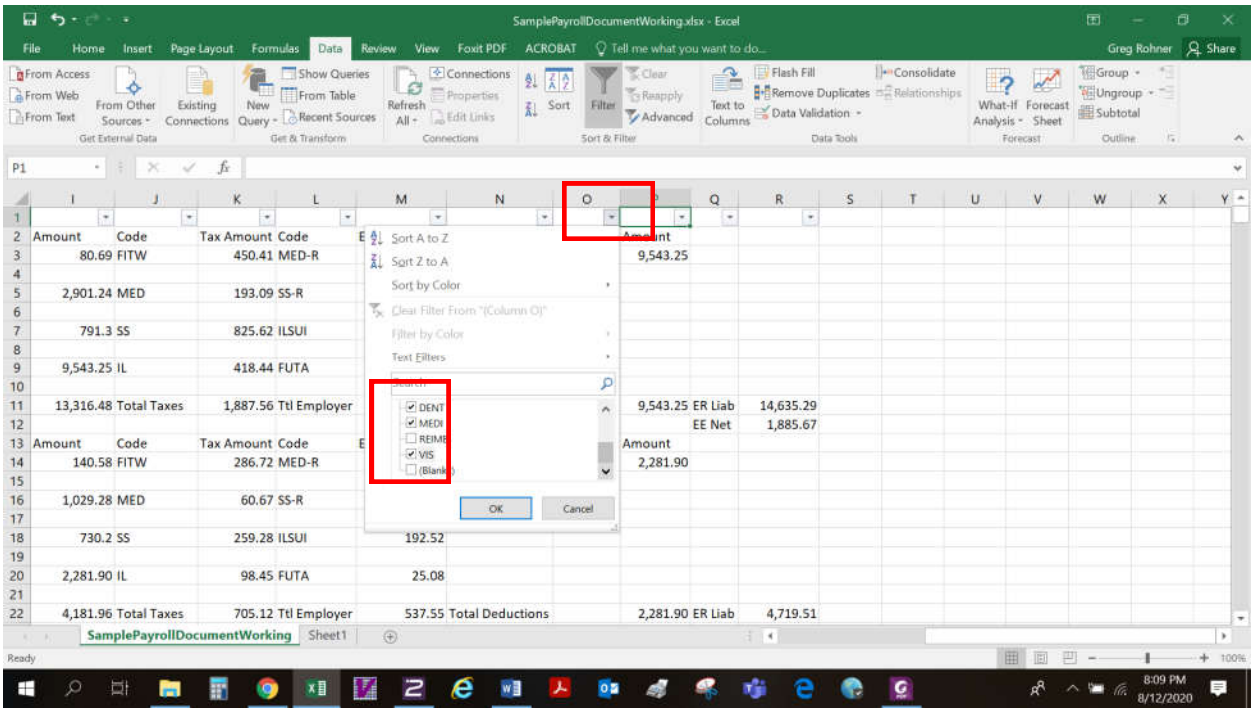
Delete the columns you don't need (ones that don't have headers)



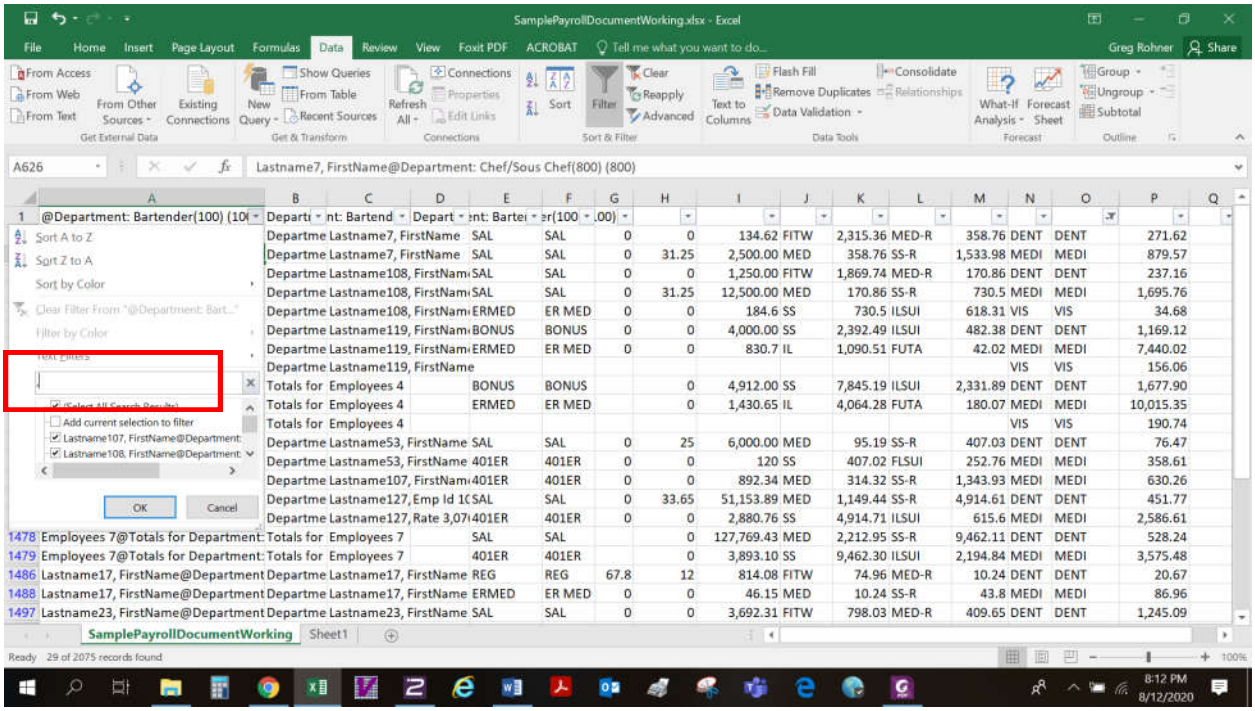
Columns not needed



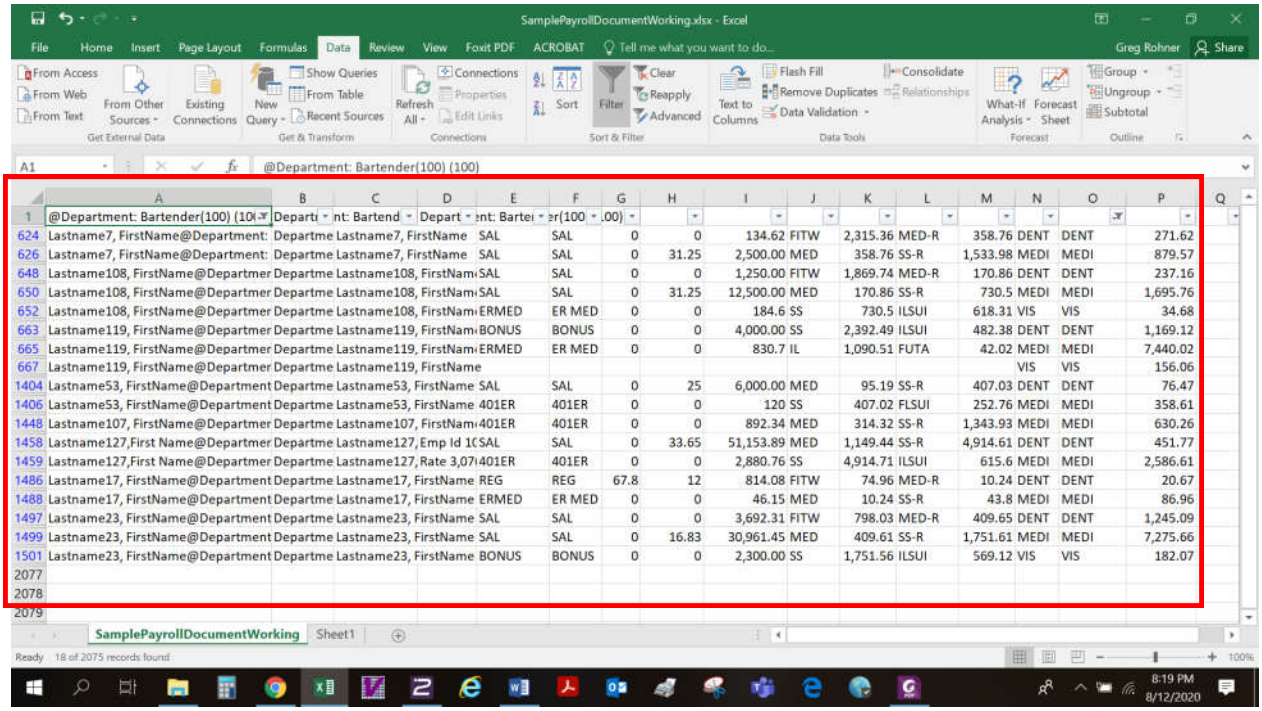
In California, Section 125 contributions are needed for an audit as well. This information can be completed by selecting the various itesm which can be done with the drop down box for column O or column N. Click the dropdown box and select the items that you wish to mine out.



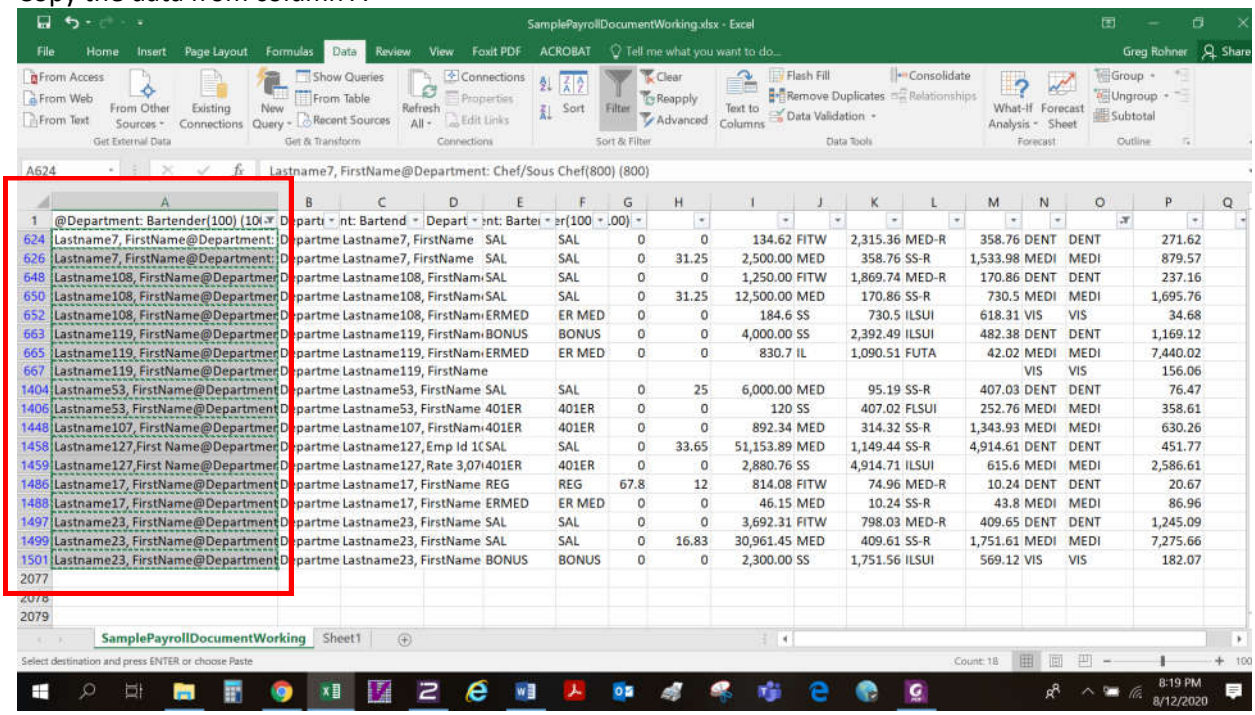
Next ensure you only have names by entering a comma in the dialog box from the dropdown box on column A



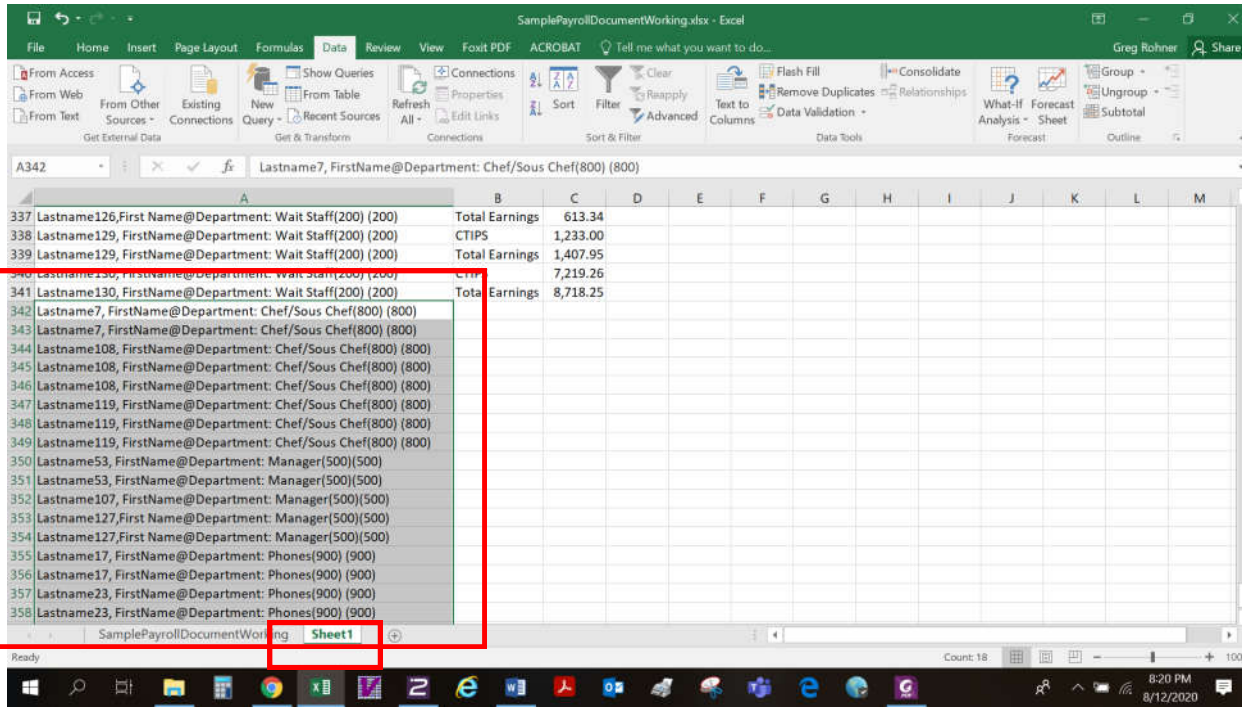
This is what the data will look like



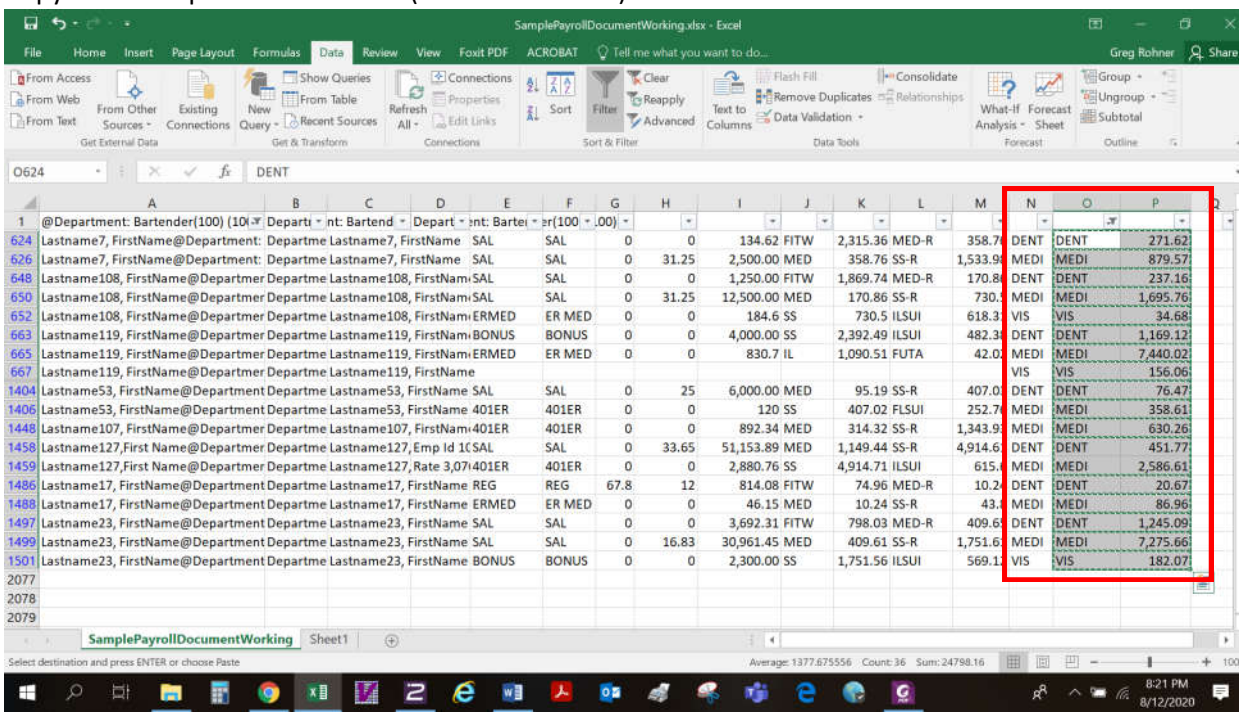
Copy the data from column A



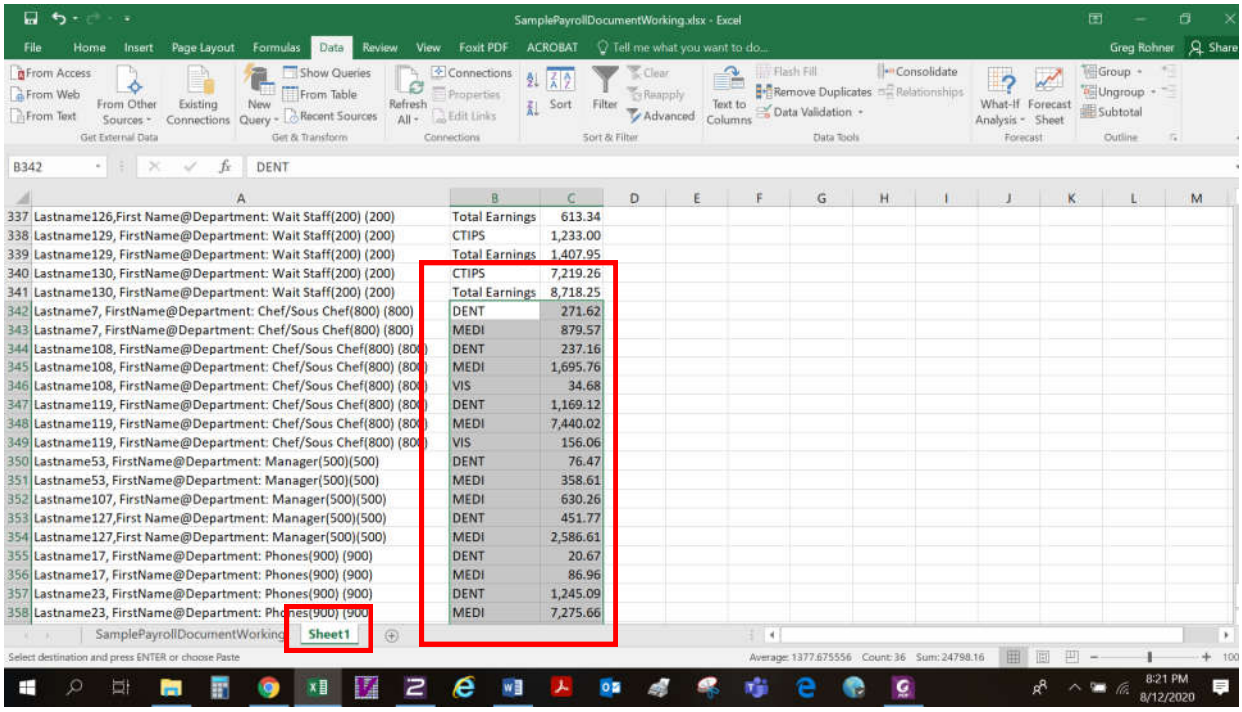
And paste under the data in column A on the page where you placed the other data.



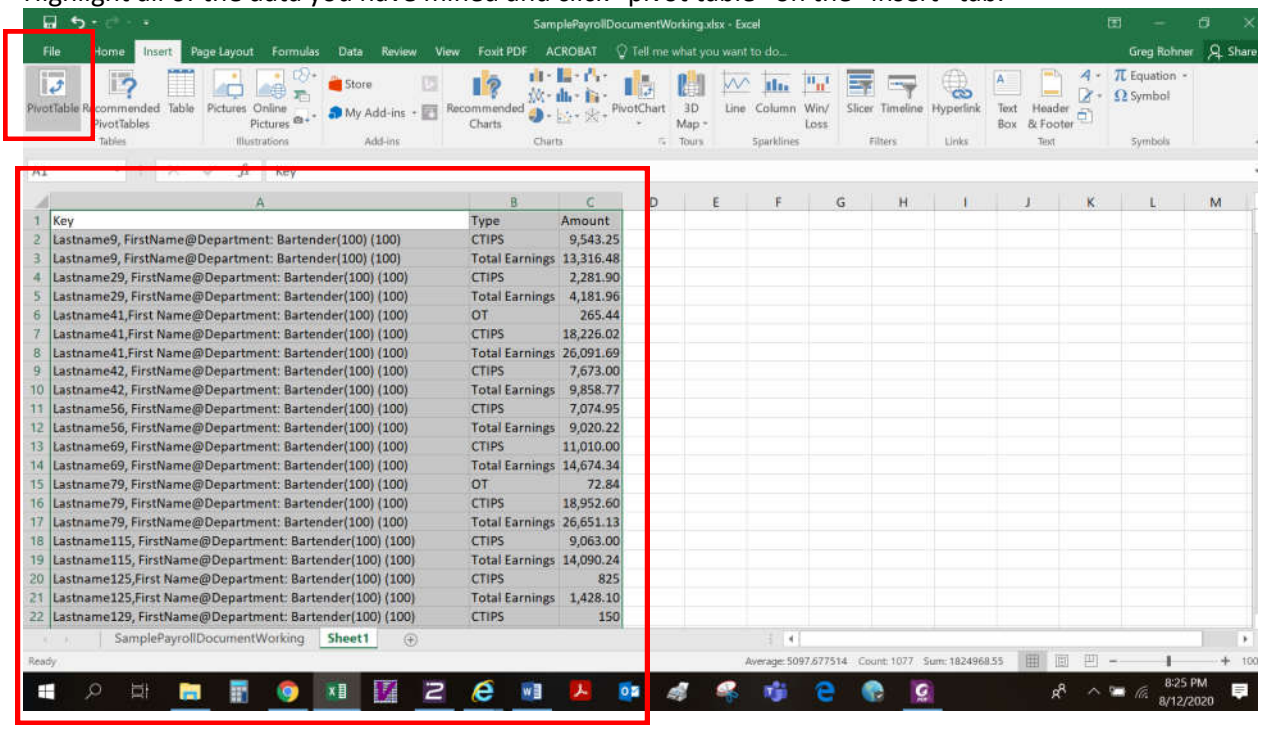
Copy the description and amount (columns O and P).



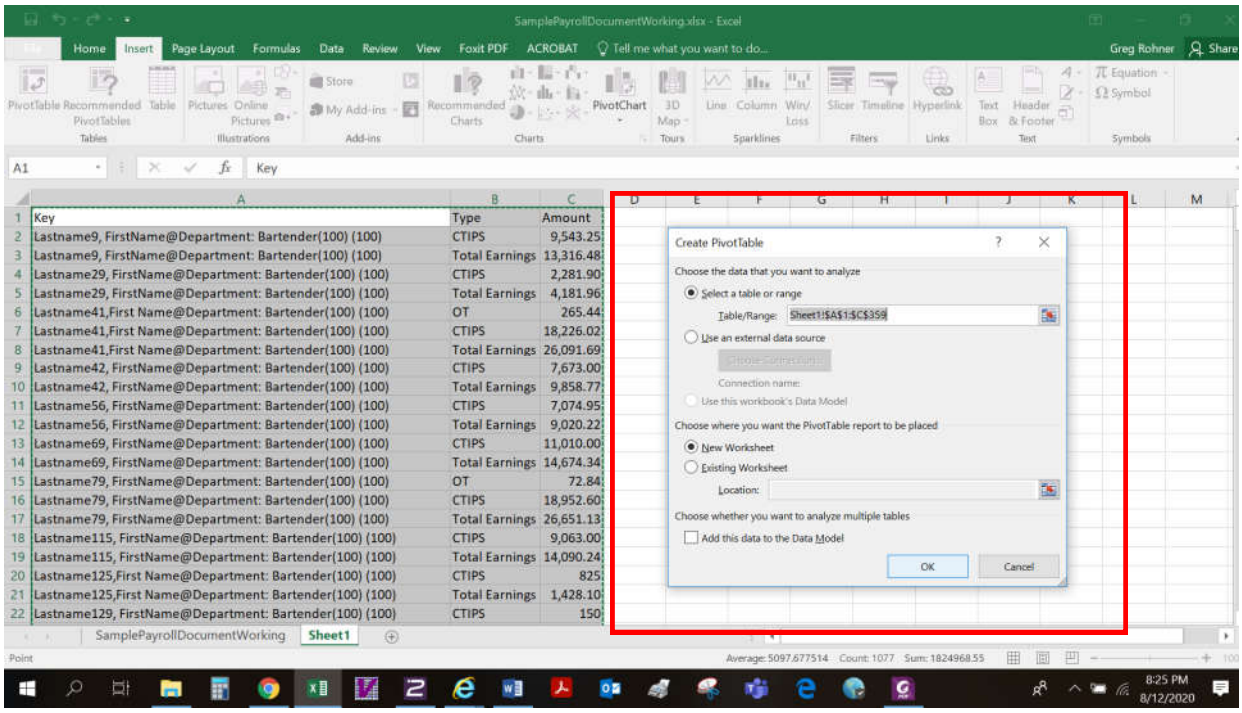
And paste next to the combined name and department on the page where you placed the mined data



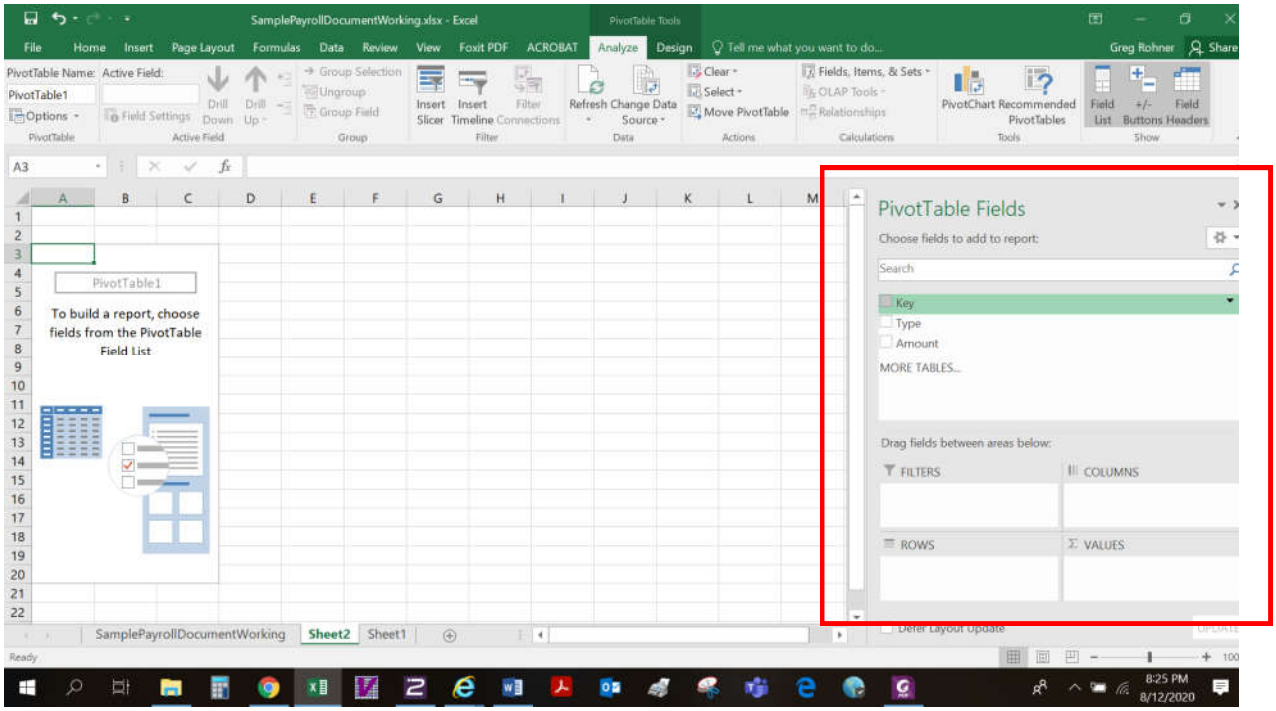
Highlight all of the data you have mined and click “pivot table” on the “insert” tab.



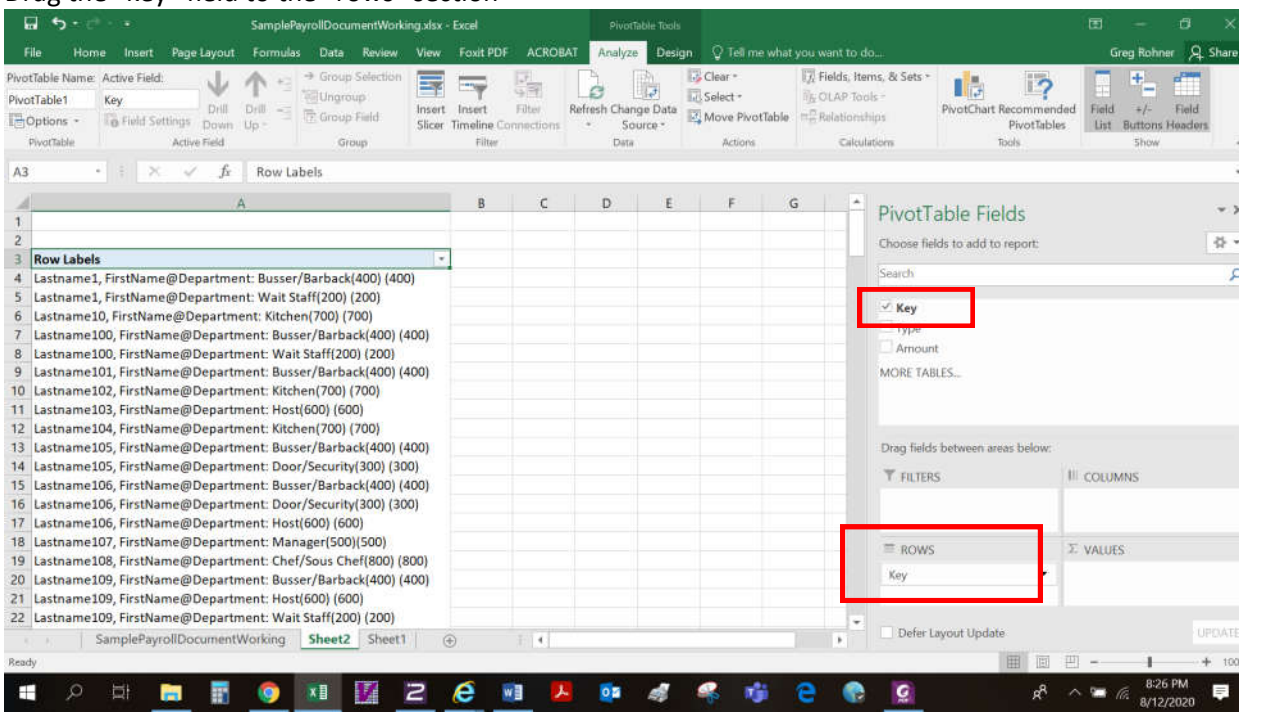
This is the dialog box that will come up when you do that. Click OK



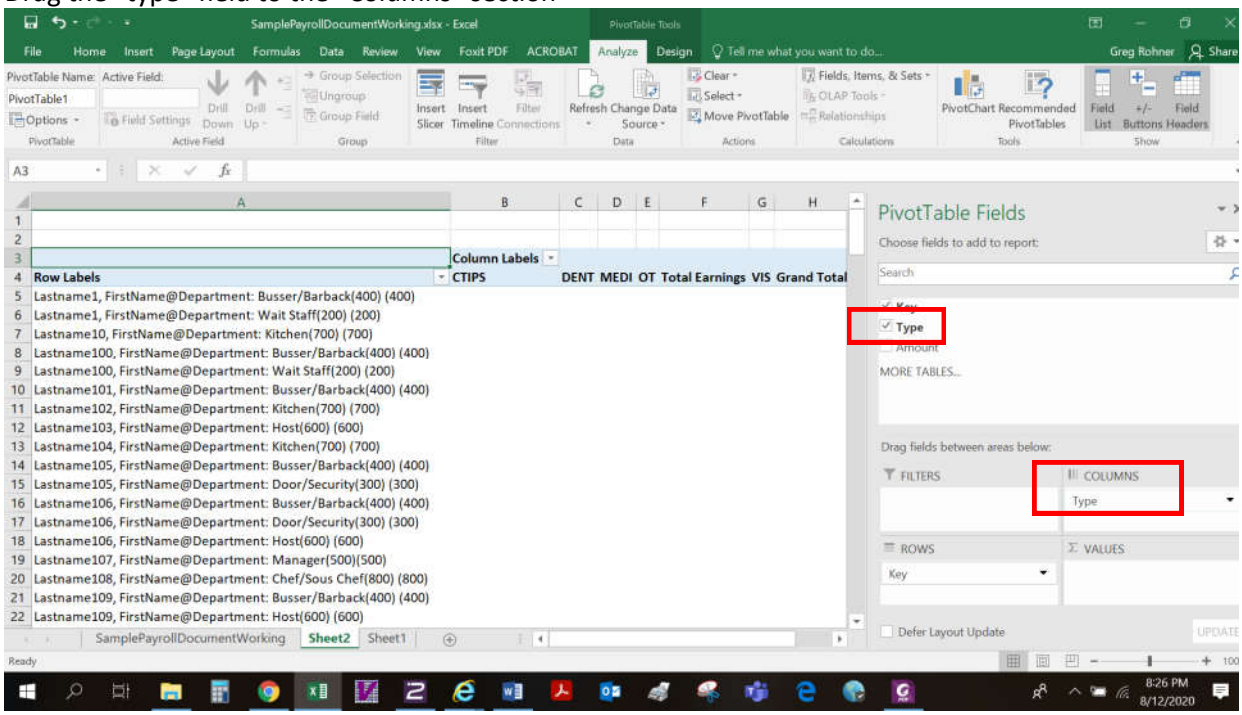
This is the result



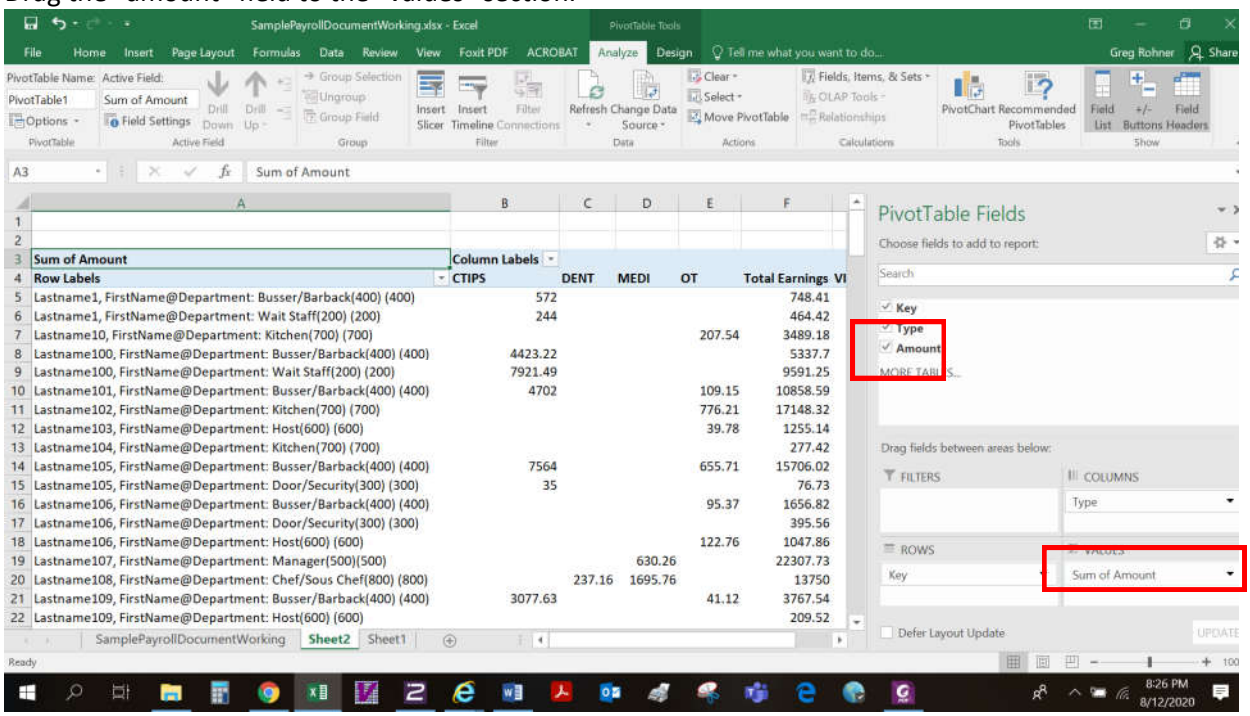
Drag the “key” field to the “rows” section



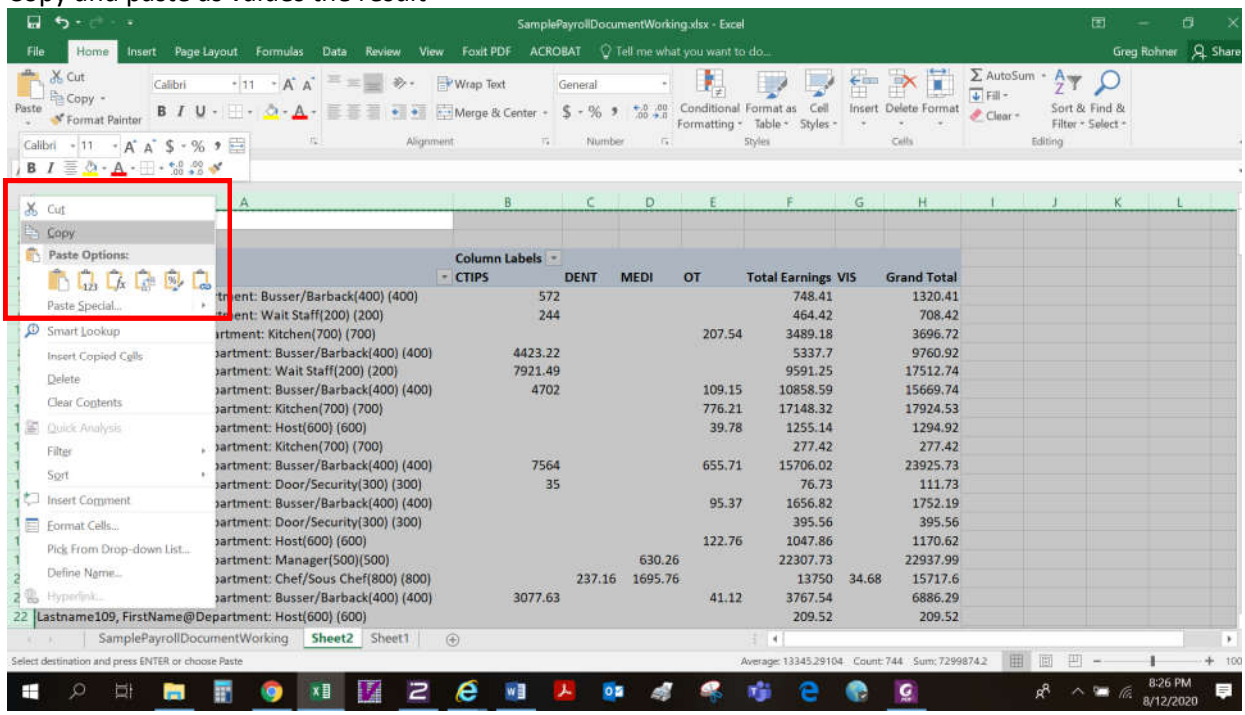
Drag the “type” field to the “Columns” section

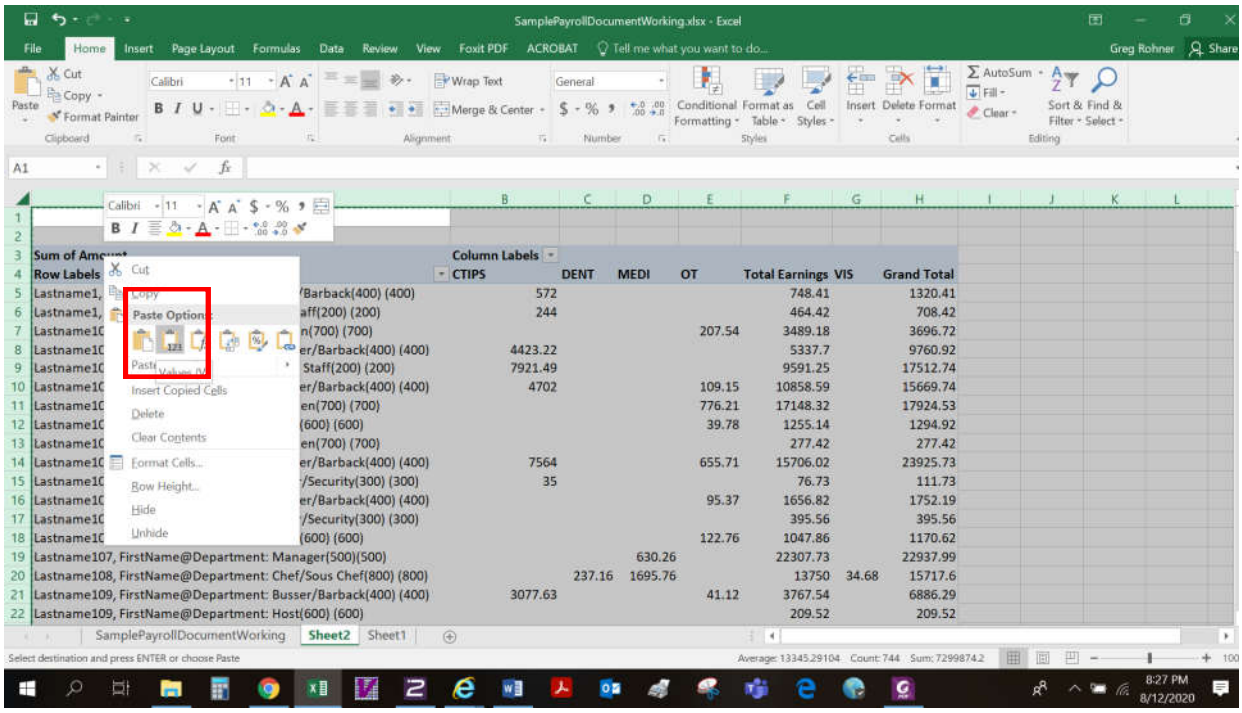


Drag the “amount” field to the “values” section.

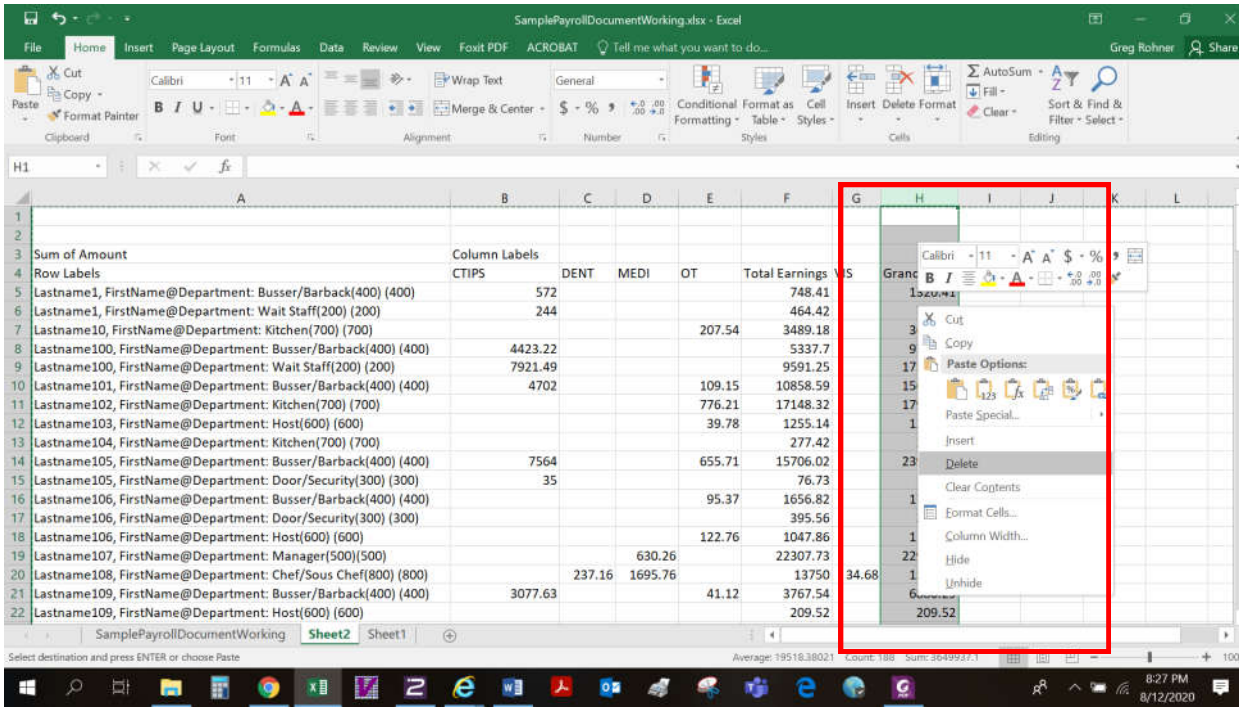


Copy and paste as values the result

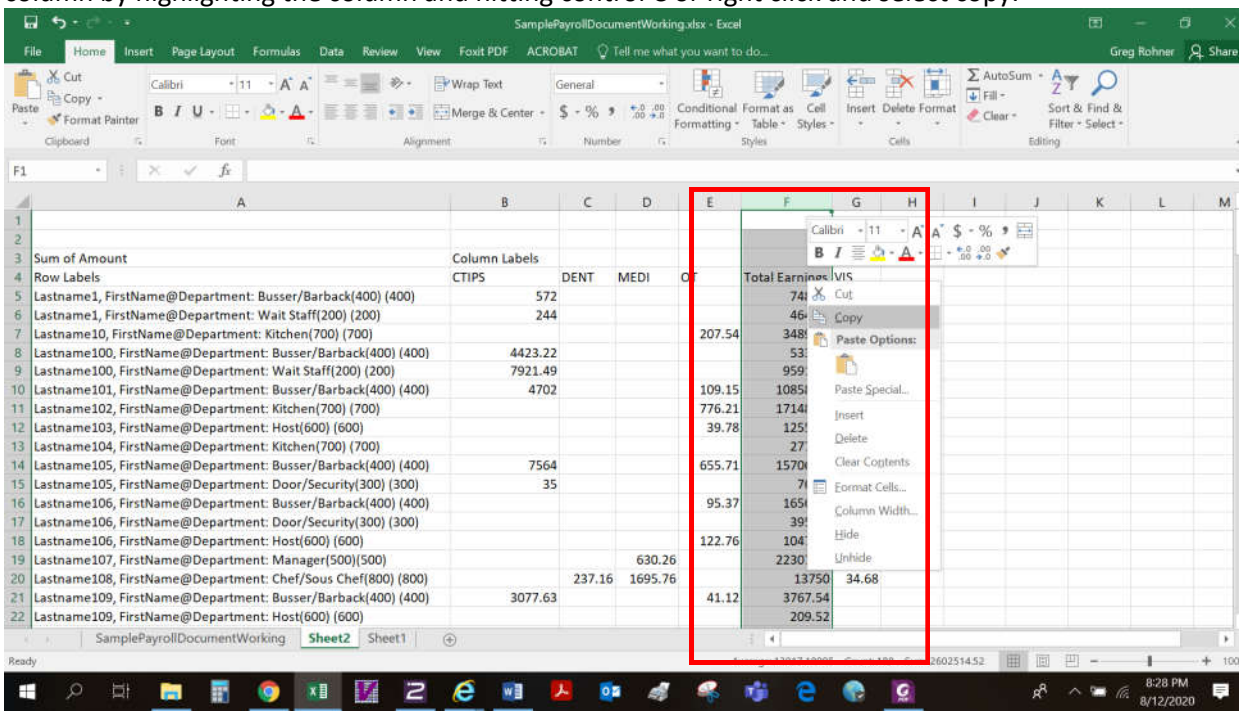




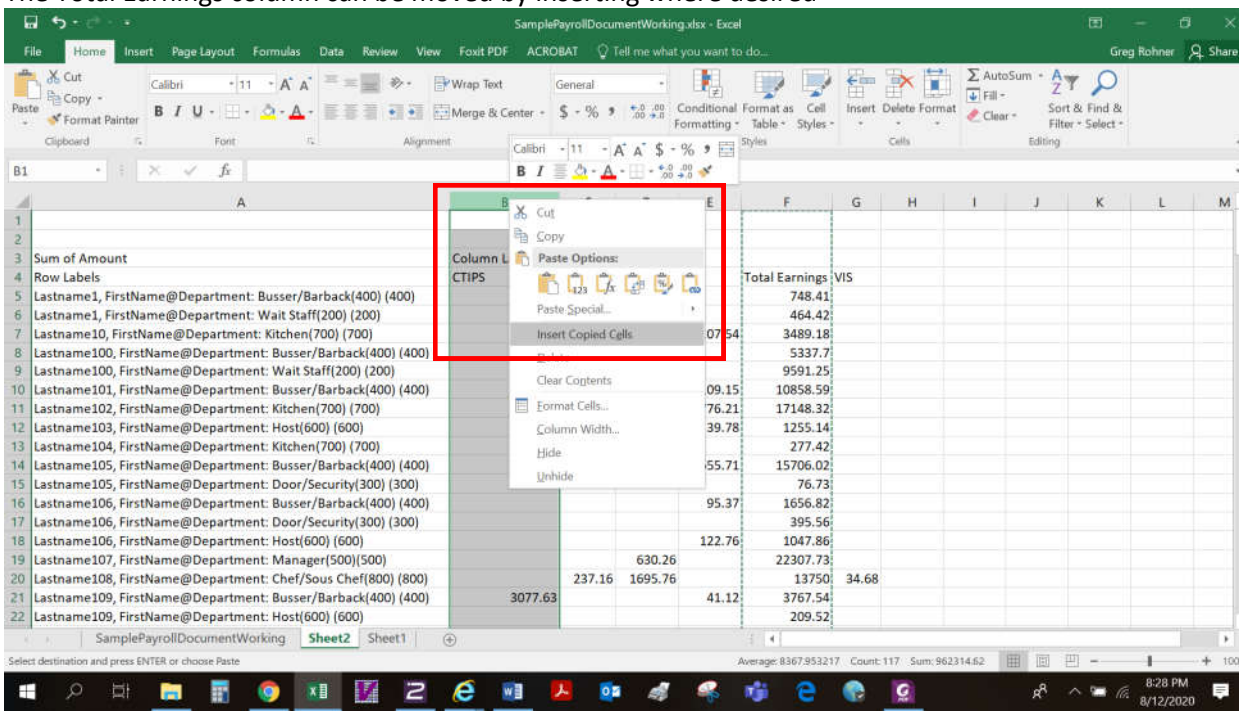
Delete the Grand Total column. This is a total of the grand total plus all of the other items you wanted to mine out.



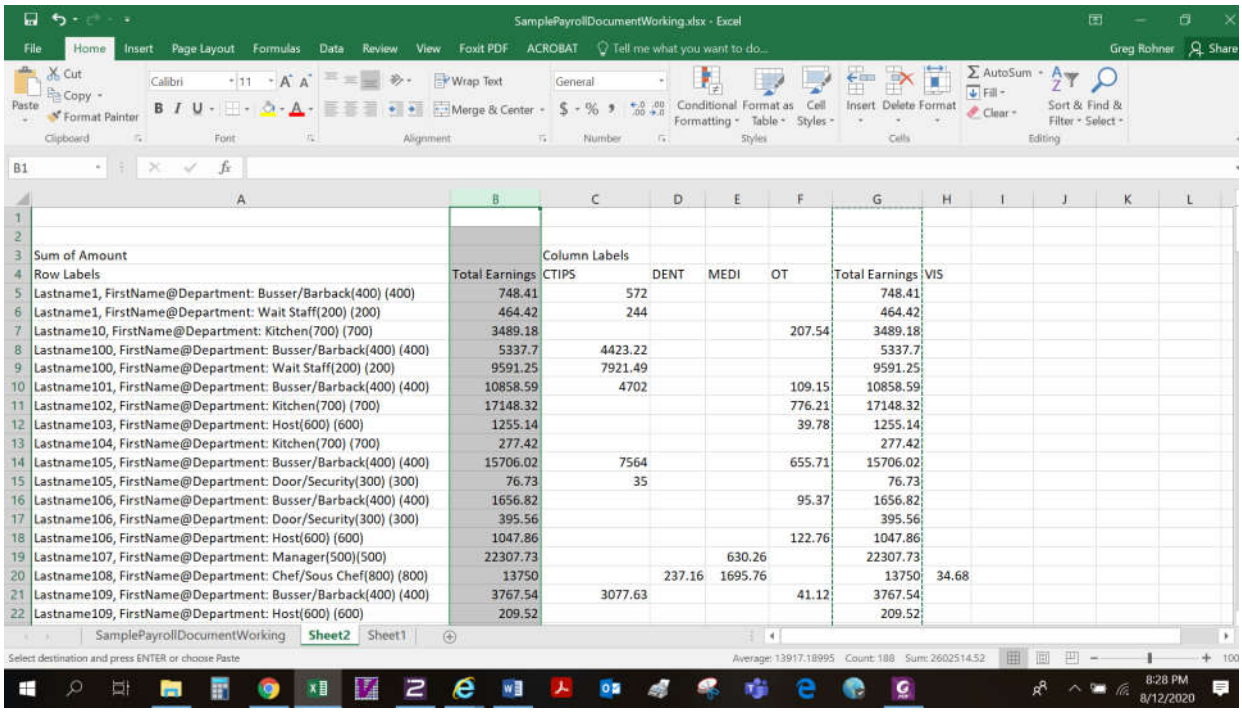
As you have saved this sheet as values, you can move columns without any difficulty. Typically on audit worksheets, the totals are show first then the deductions. You can move the total earnings column by highlighting the column and hitting control C or right click and select copy.



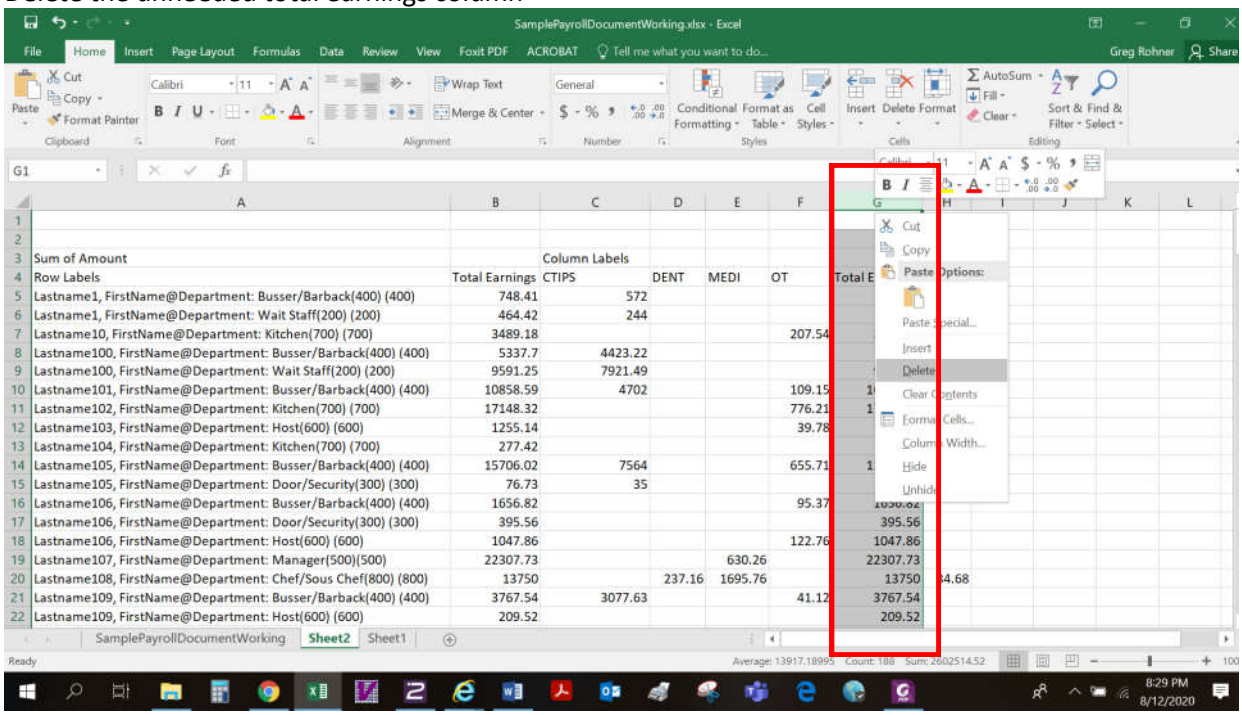
The Total Earnings column can be moved by inserting where desired



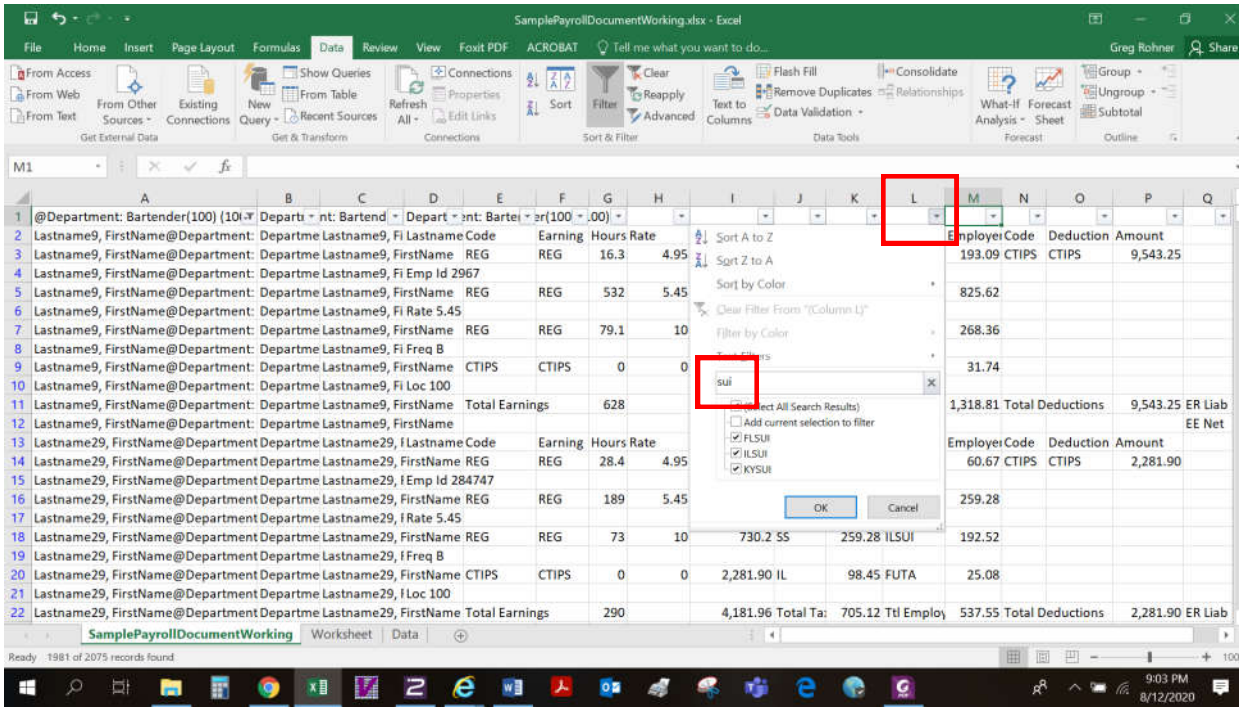
This is the result



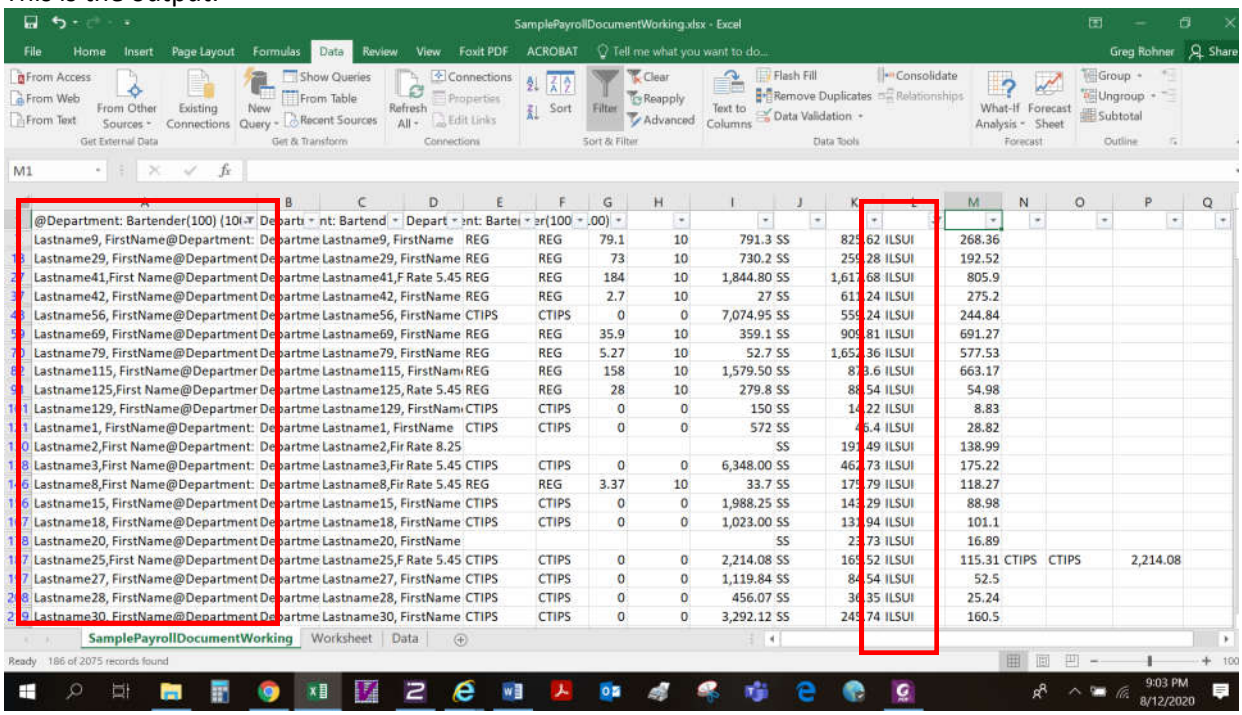
Delete the unneeded total earnings column



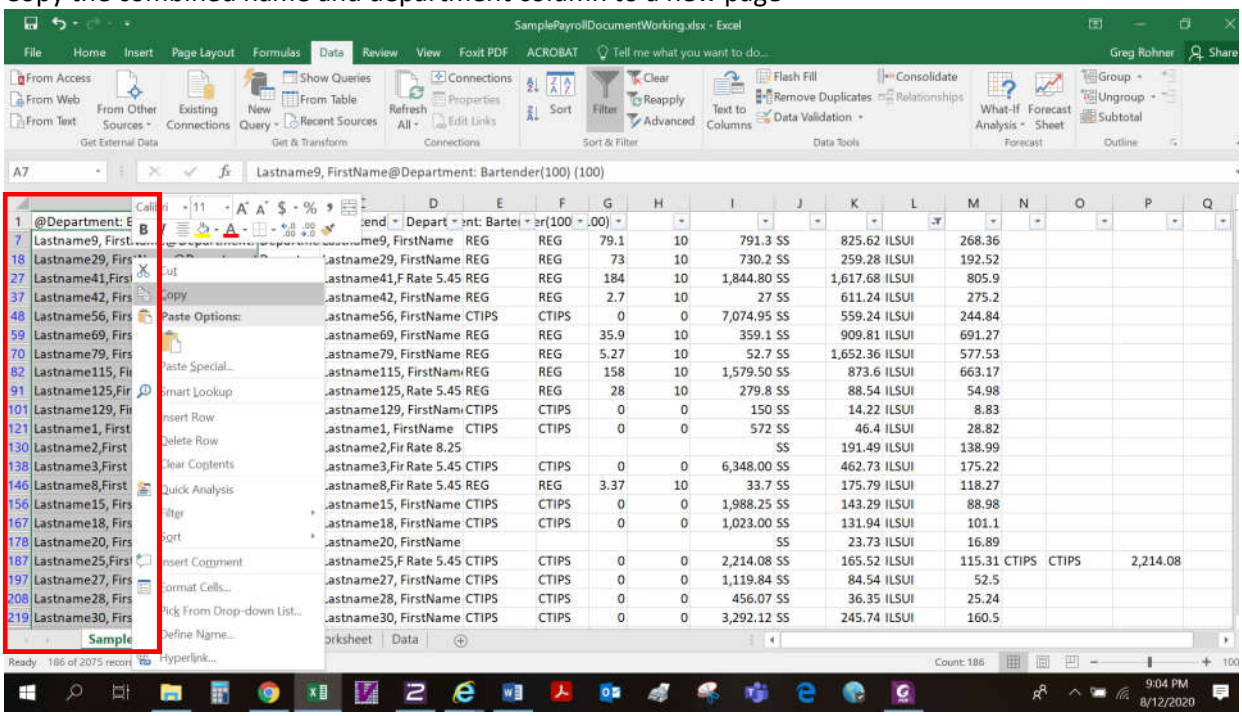
Next we want to pull states. Often this can be done by using the SUI information which this report show. In the dropdown box in column L, type SUI in the dialog and click OK

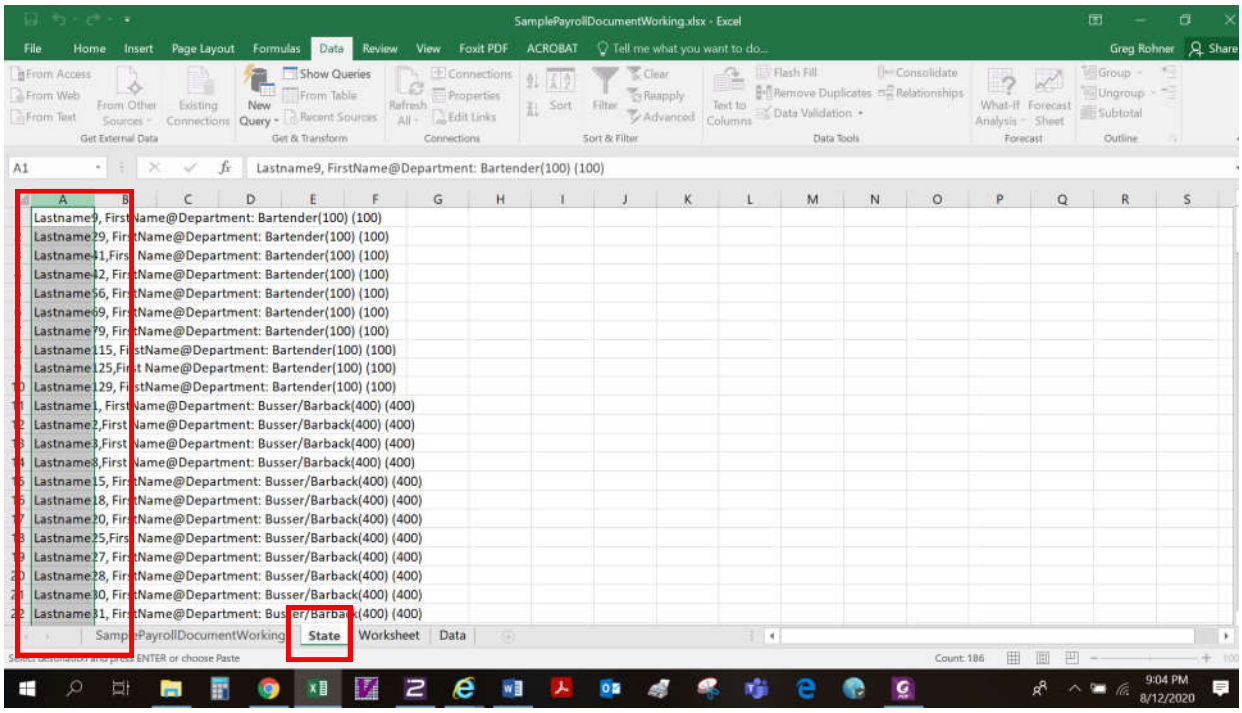


This is the output.

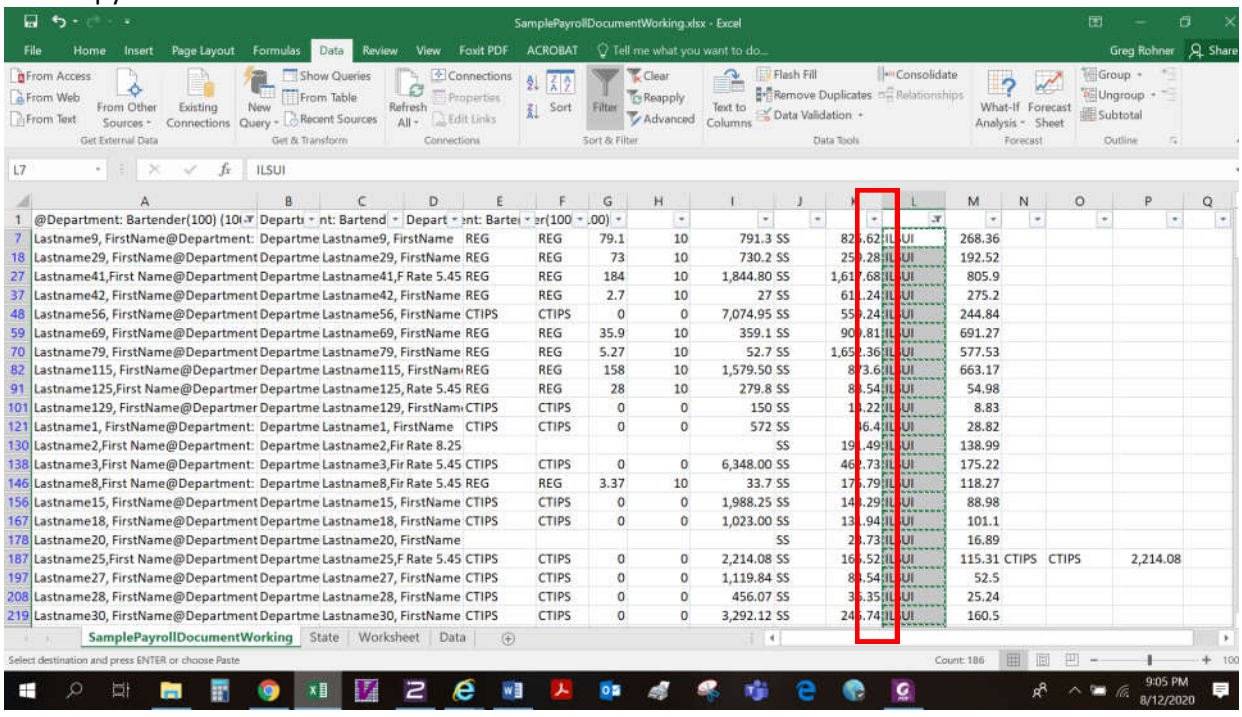


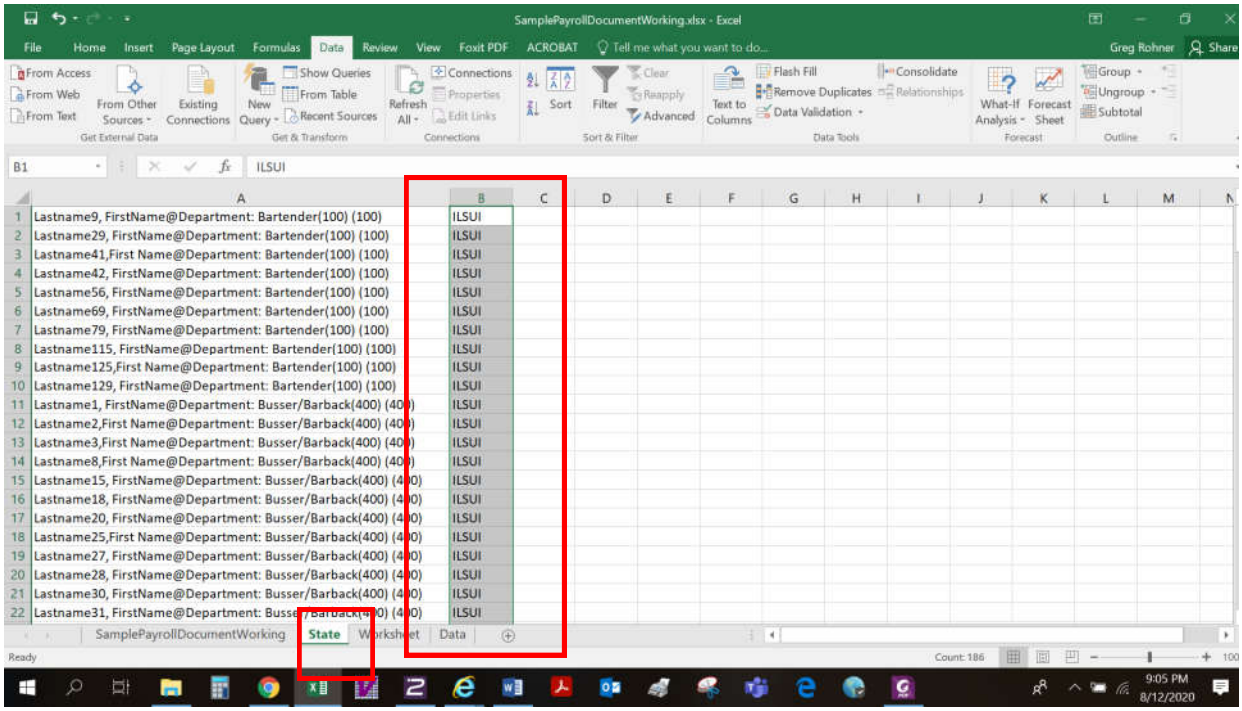
Copy the combined name and department column to a new page



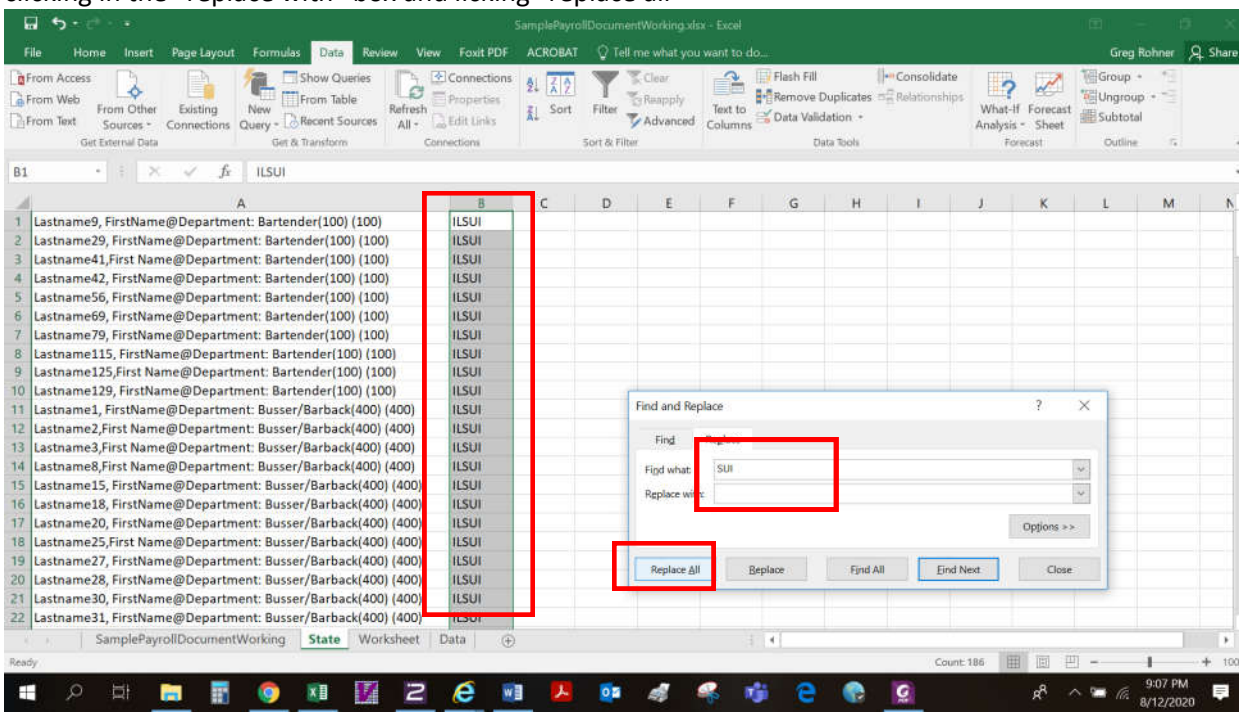


Now copy the SUI column label to the same sheet

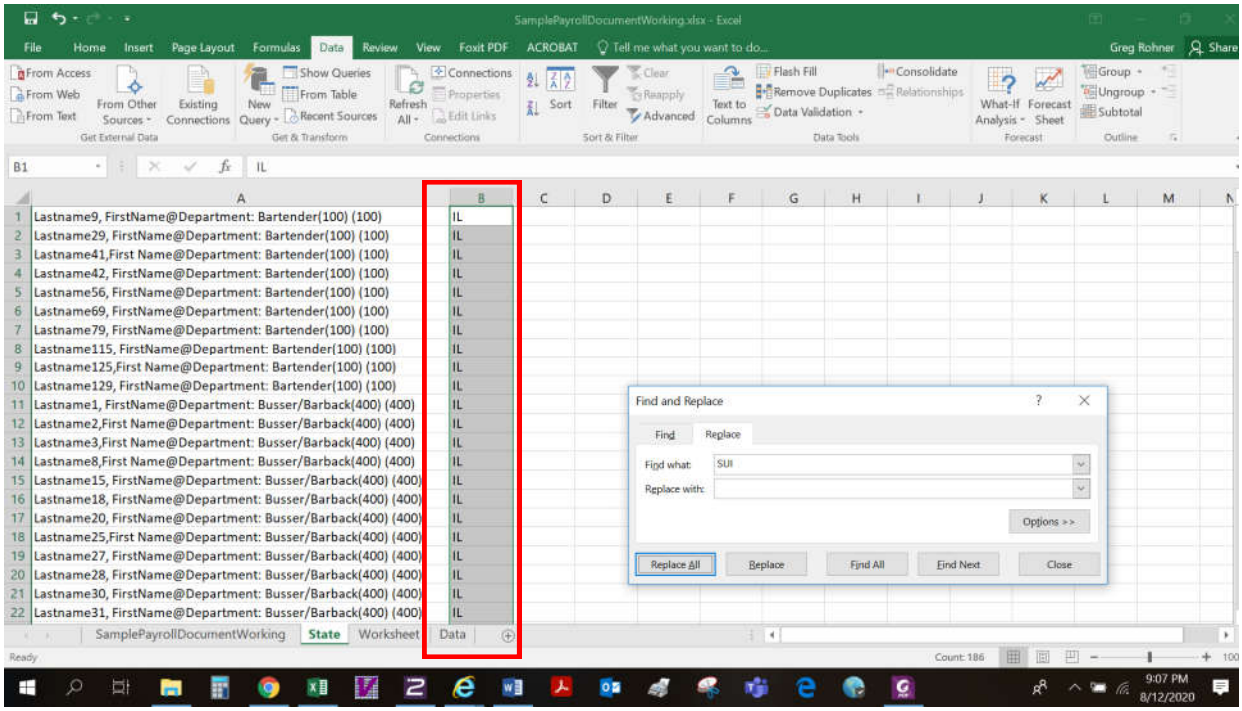




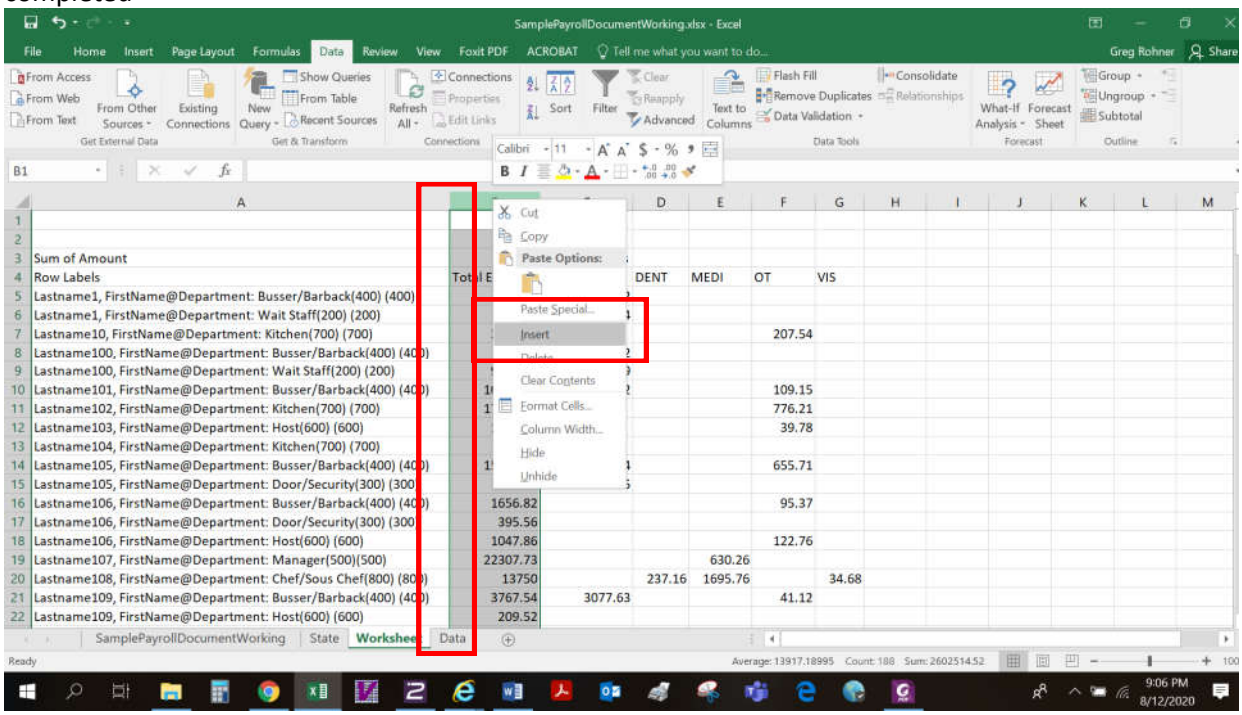
You can get rid of the verbiage “SUI” using control F, typing SUI in the “Find what” box and just clicking in the “replace with” box and licking “replace all”



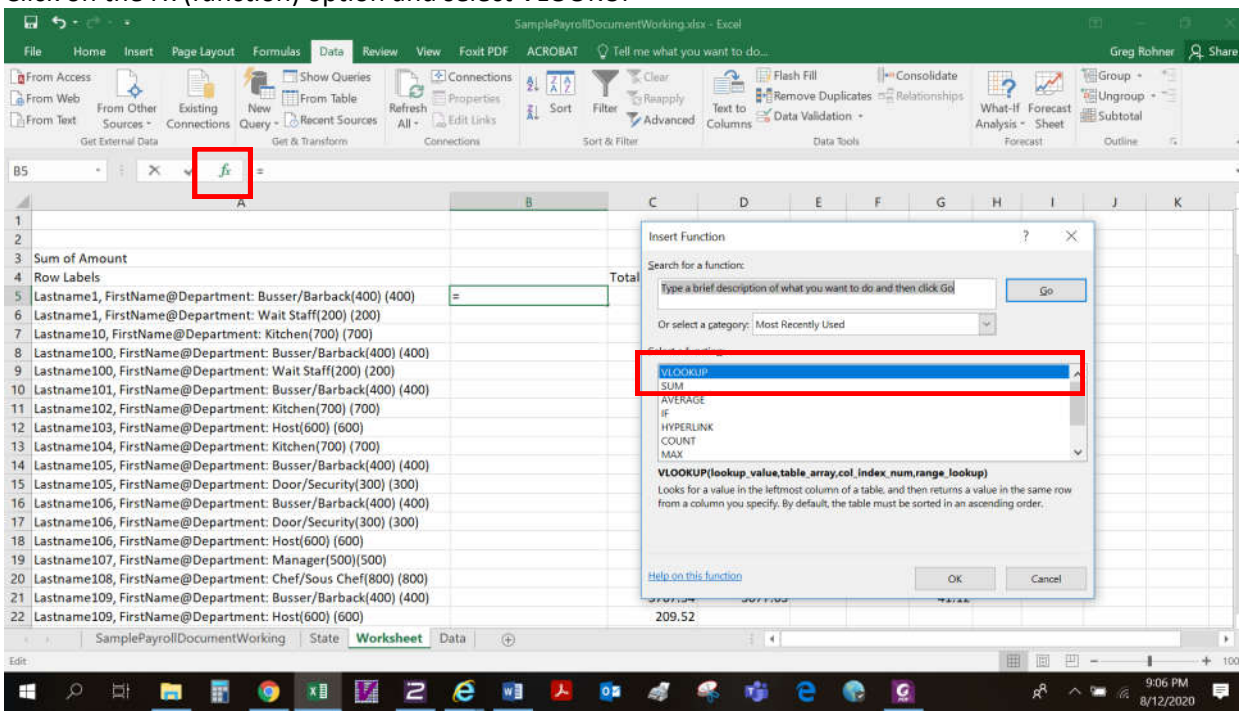
This is the result



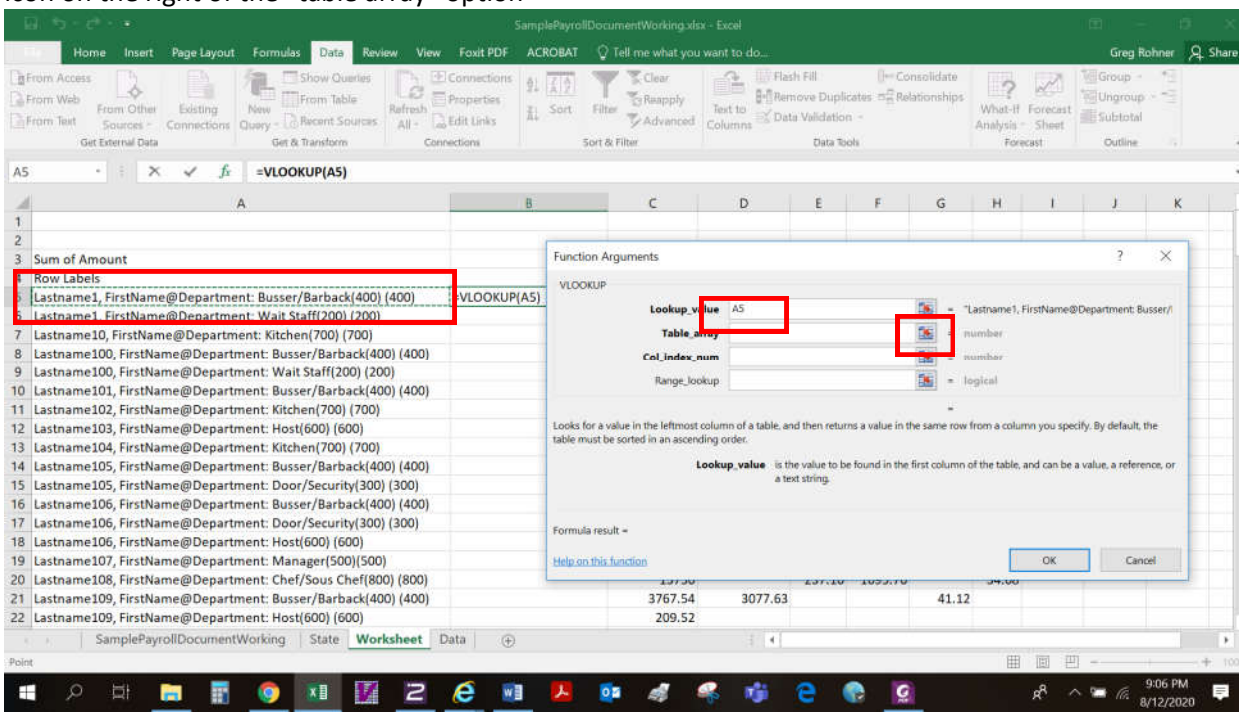
Insert a column next to the combined name and department information on the pivot table you just completed



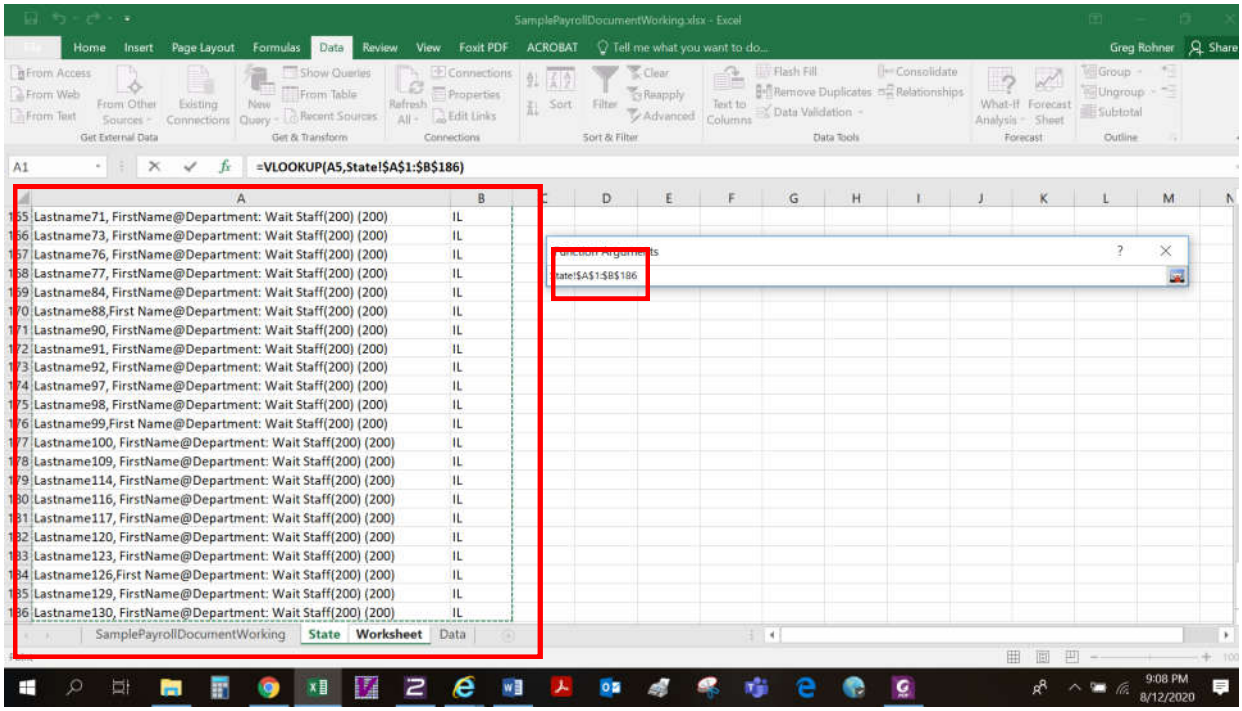
Click on the Fx (function) option and select VLOOKUP



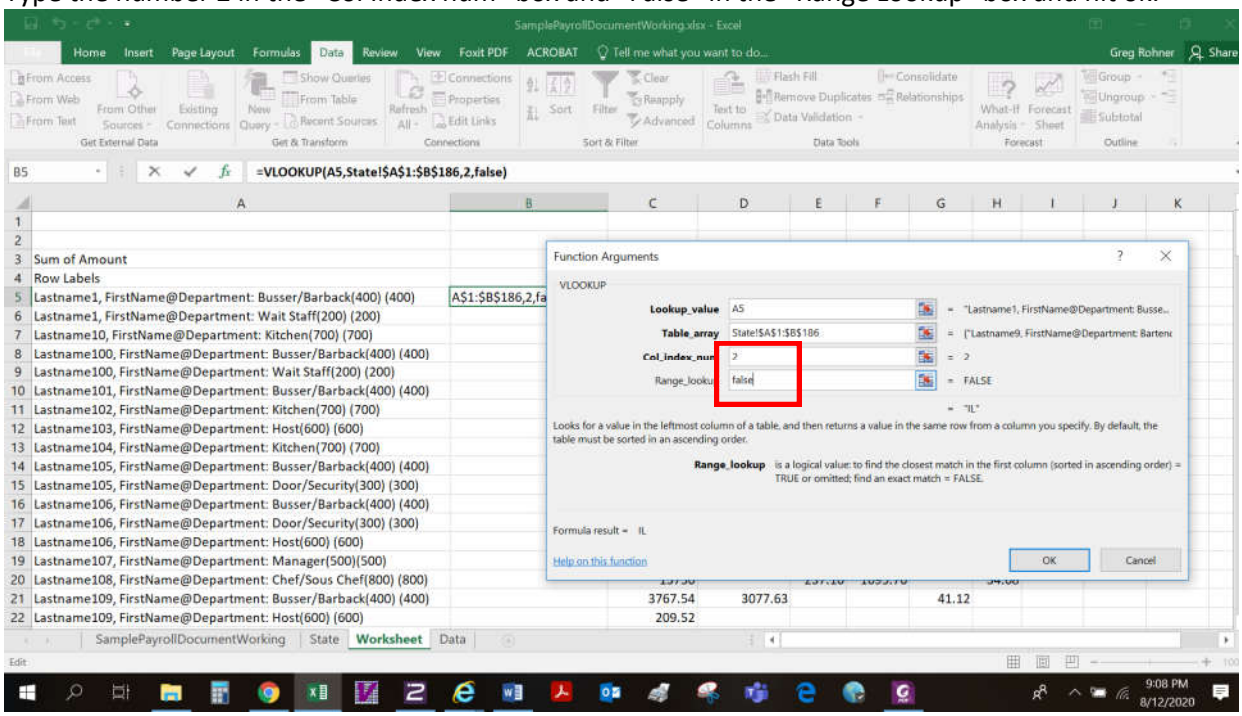
In the resulting box that comes up, type A5 in the lookup value (or click on that cell). Then click on the icon on the right of the “table array” option



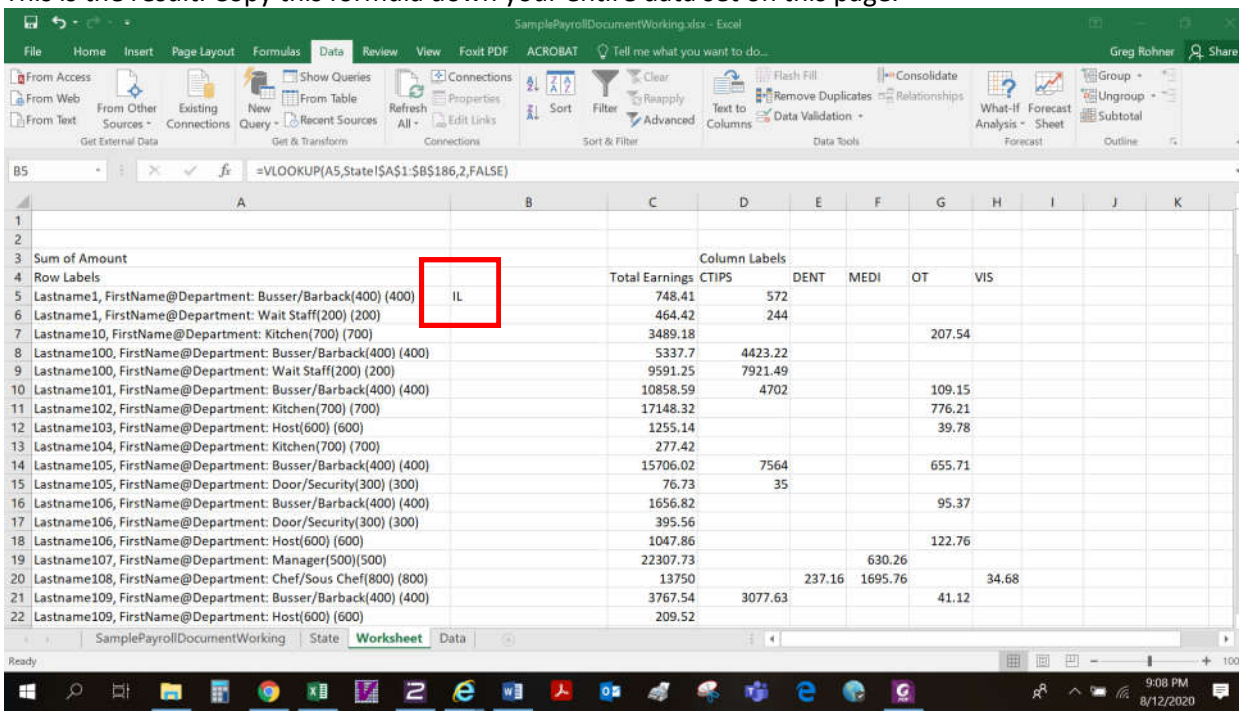
You will get a dialog box like this. Click over to the page where you have the combined names and departments and states information. Highlight the entire data field on this page, hit the f4 key and hit enter.

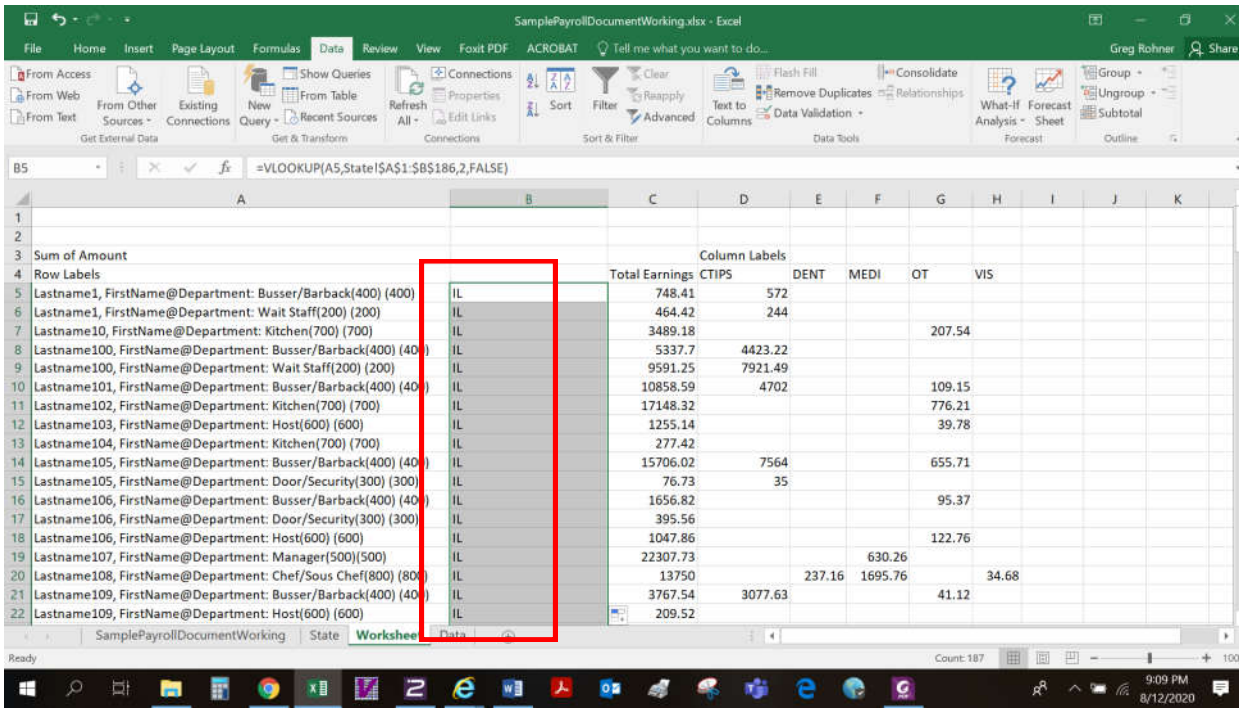


Type the number 2 in the “Col Index num” box and “False” in the “Range Lookup” box and hit ok.

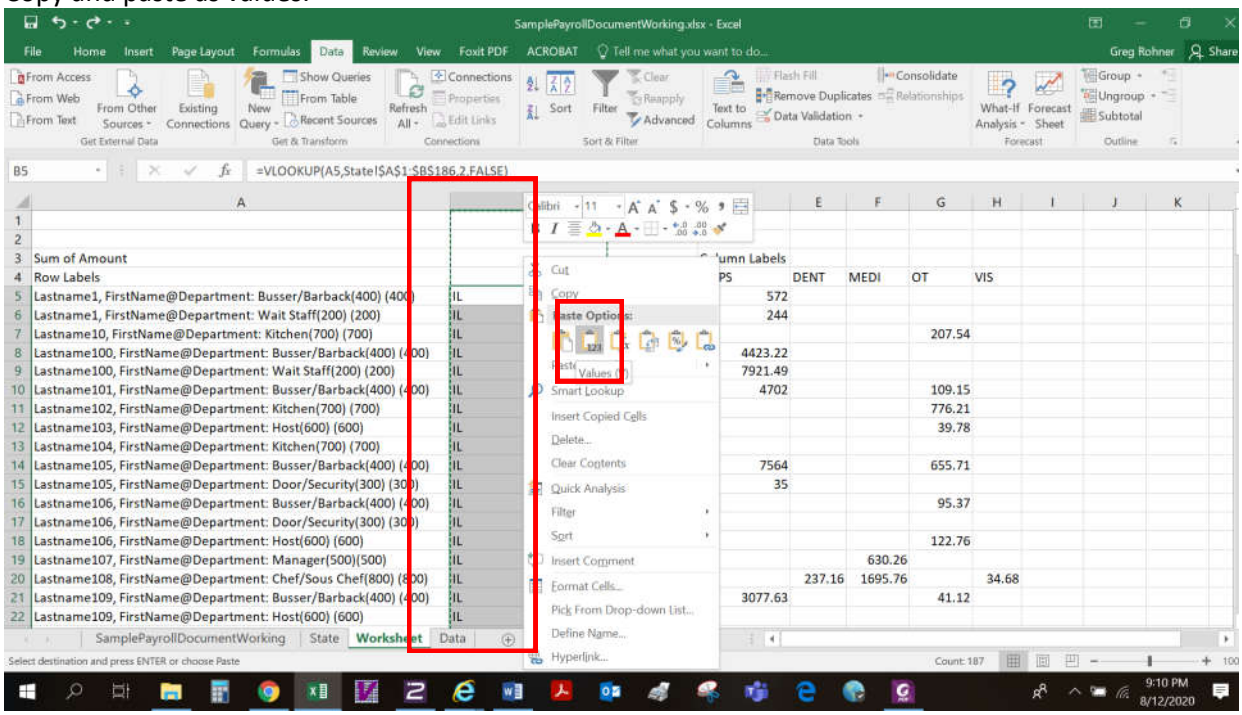


This is the result. Copy this formula down your entire data set on this page.

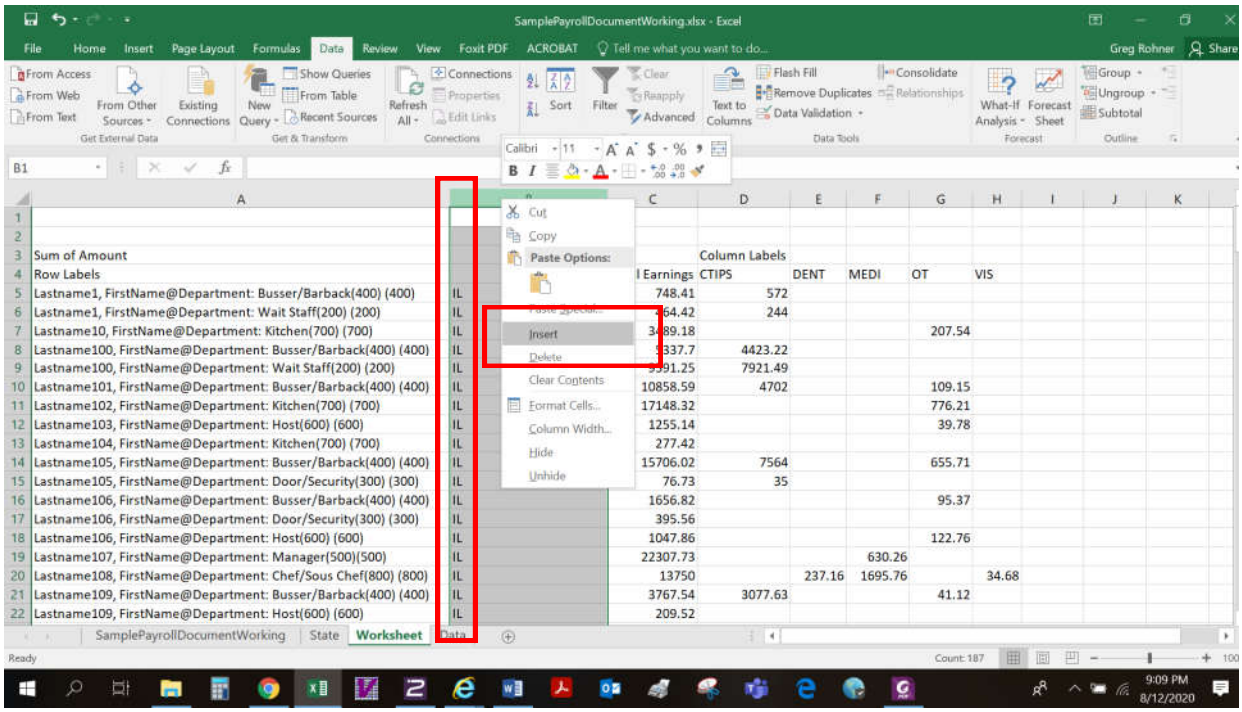




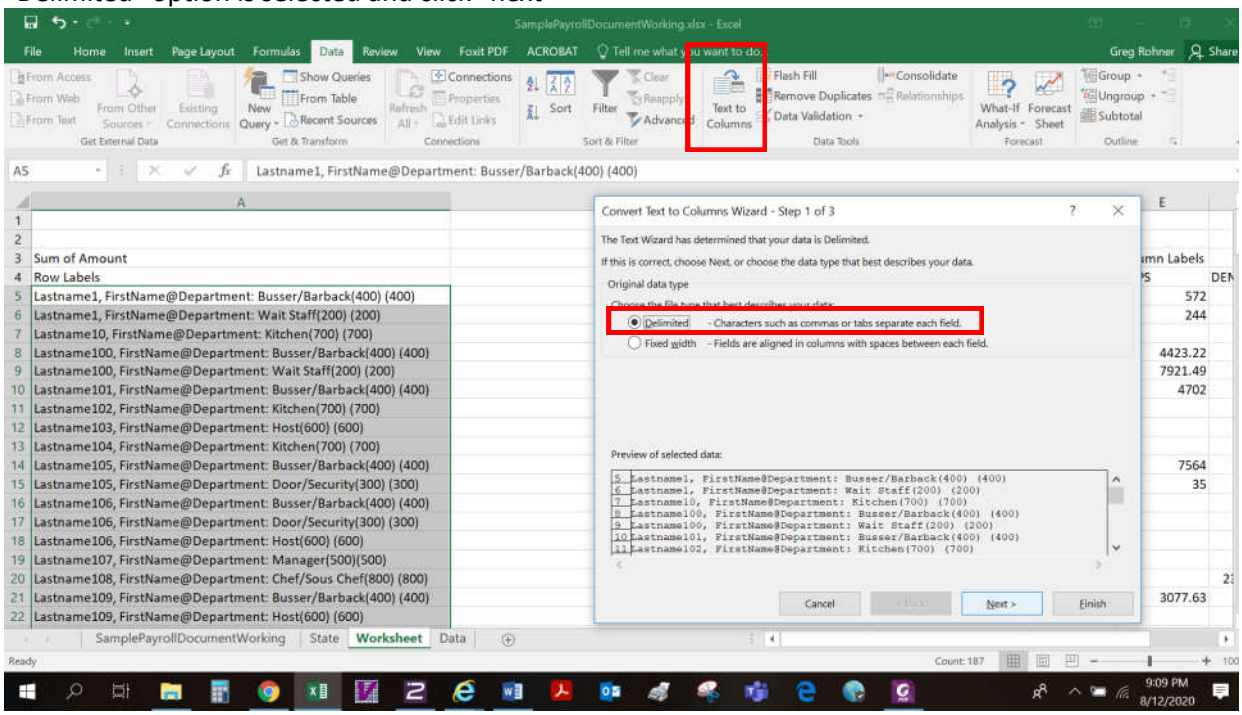
Copy and paste as values.



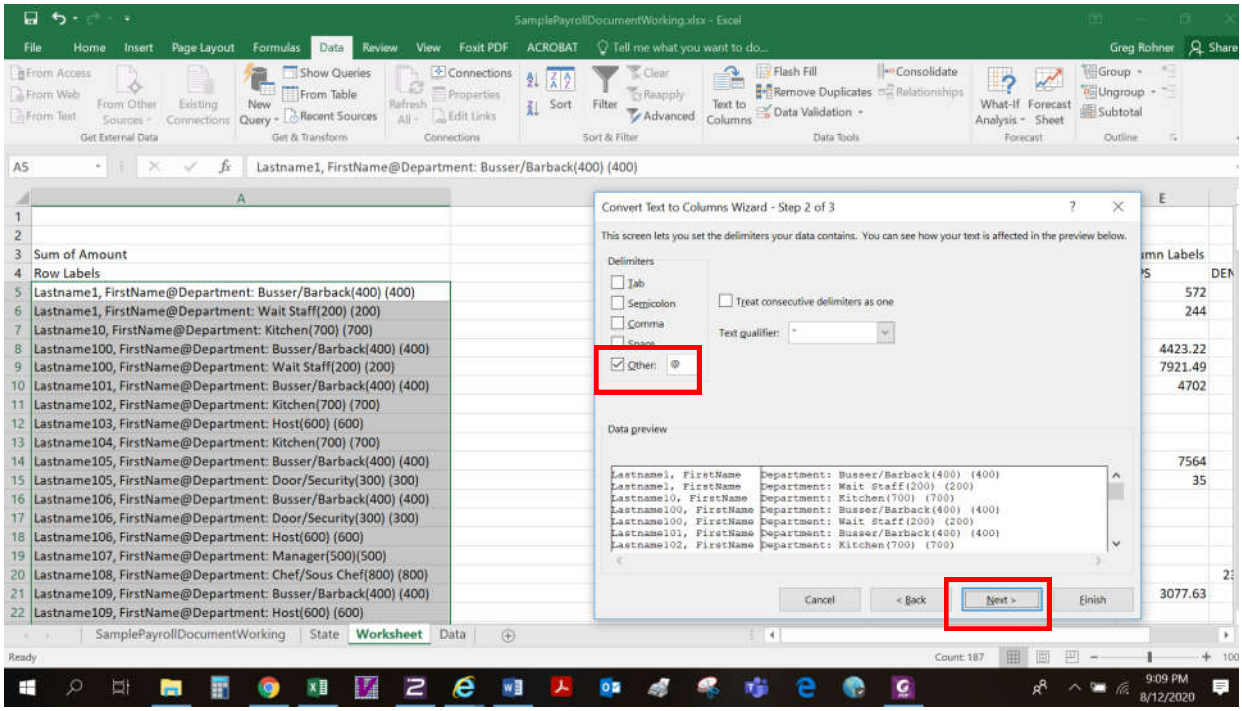
Insert a column to the left of the state information.



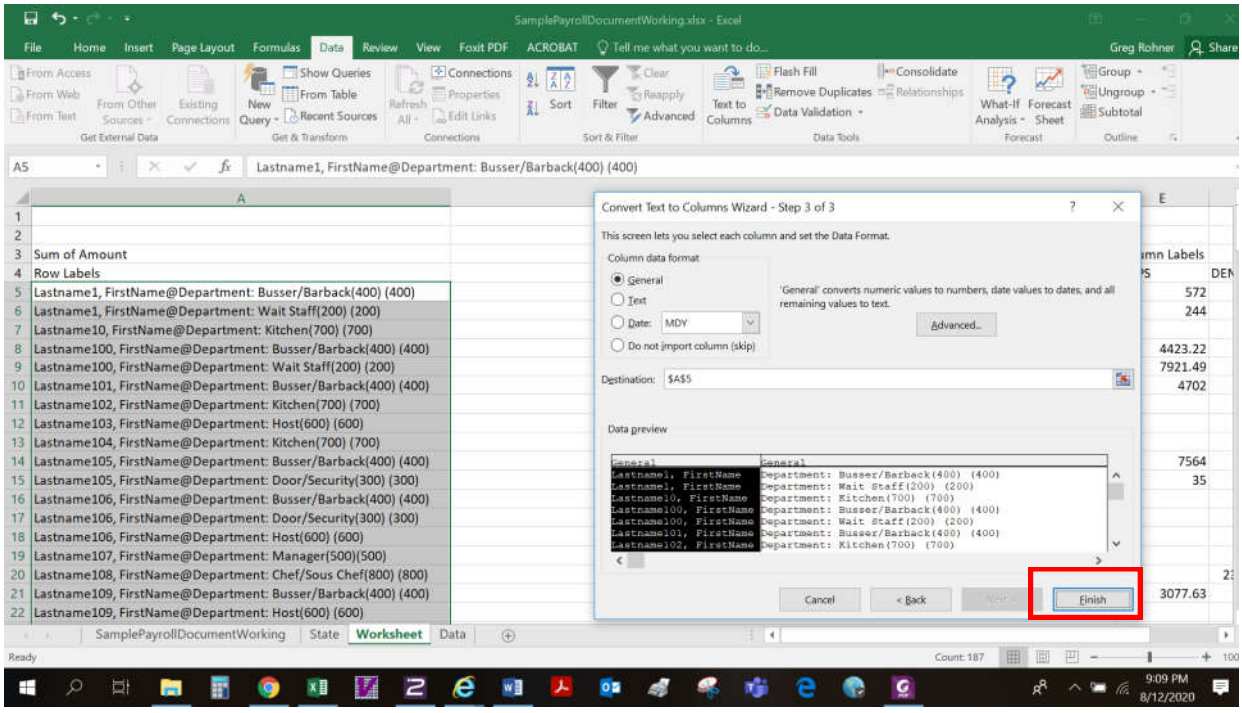
Click on “Text to Columns” on the “Data” tab. This is the box that will come up. Make certain the “Delimited” option is selected and click “next”



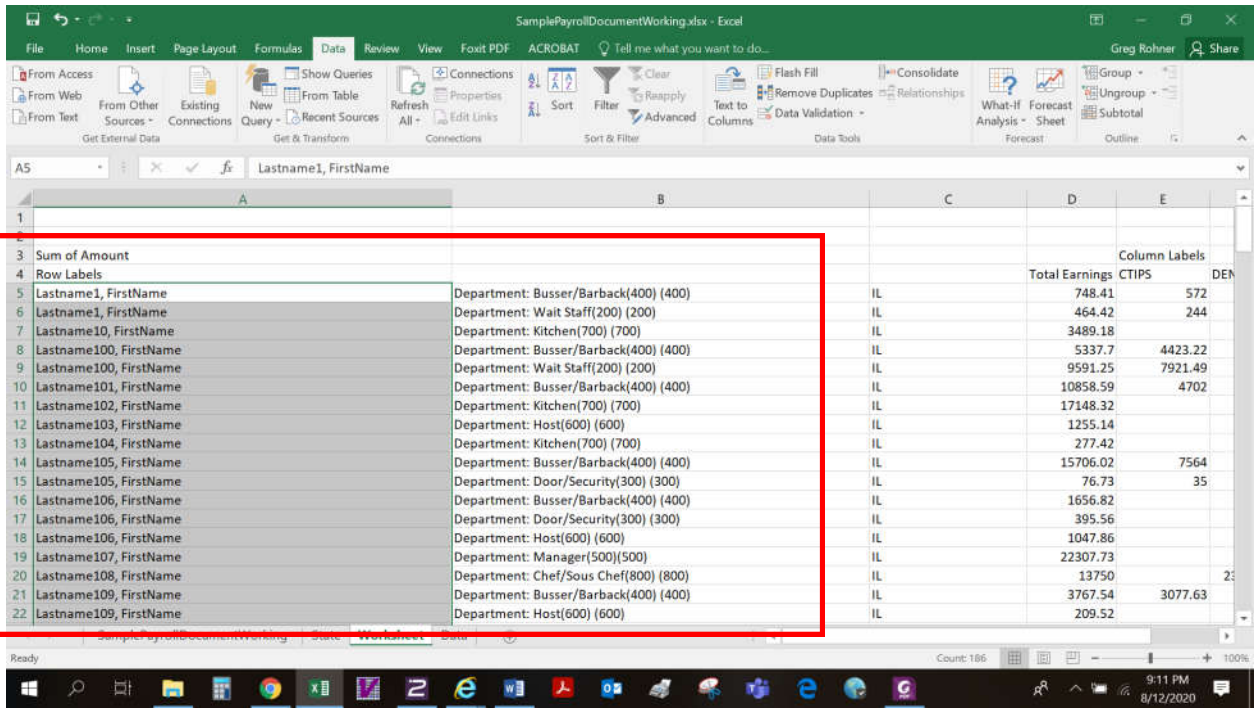
In the next dialog box, deselect any preselected options and select “other” and type “@” in the field next to that. Then click “next”



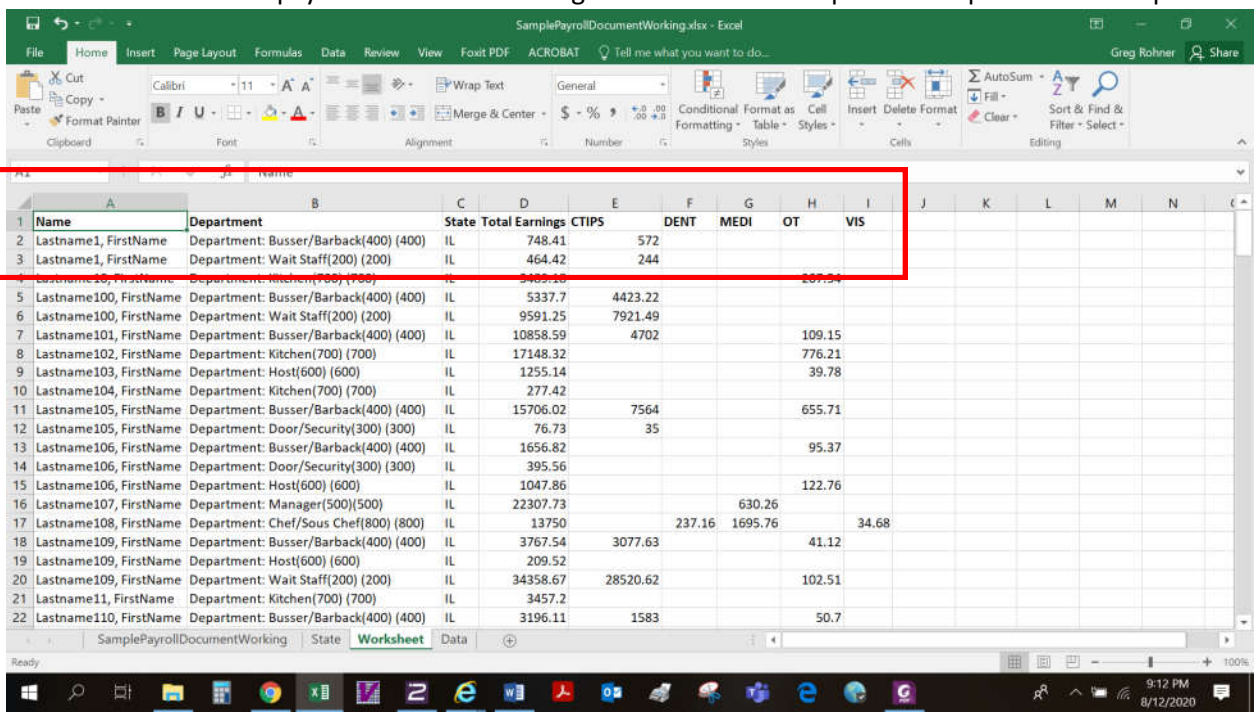
Click “Finish”



This is the result



You can then clean up your data and if using an audit software product import into that product.



	A	B	C	D	E	F	G	H	I
1	Name	Department	State	Total Earnings	CTIPS	DENT	MEDI	OT	VIS
2	Lastname1, FirstName	Department: Busser/Barback(400) (400)	IL	748.41	572				
3	Lastname1, FirstName	Department: Wait Staff(200) (200)	IL	464.42	244				
4	[illegible]	[illegible]	[illegible]	[illegible]				[illegible]	
5	Lastname100, FirstName	Department: Busser/Barback(400) (400)	IL	5337.7	4423.22				
6	Lastname100, FirstName	Department: Wait Staff(200) (200)	IL	9591.25	7921.49				
7	Lastname101, FirstName	Department: Busser/Barback(400) (400)	IL	10858.59	4702			109.15	
8	Lastname102, FirstName	Department: Kitchen(700) (700)	IL	17148.32				776.21	
9	Lastname103, FirstName	Department: Host(600) (600)	IL	1255.14				39.78	
10	Lastname104, FirstName	Department: Kitchen(700) (700)	IL	277.42					
11	Lastname105, FirstName	Department: Busser/Barback(400) (400)	IL	15706.02	7564			655.71	
12	Lastname105, FirstName	Department: Door/Security(300) (300)	IL	76.73	35				
13	Lastname106, FirstName	Department: Busser/Barback(400) (400)	IL	1656.82				95.37	
14	Lastname106, FirstName	Department: Door/Security(300) (300)	IL	395.56					
15	Lastname106, FirstName	Department: Host(600) (600)	IL	1047.86				122.76	
16	Lastname107, FirstName	Department: Manager(500)(500)	IL	22307.73			630.26		
17	Lastname108, FirstName	Department: Chef/Sous Chef(800) (800)	IL	13750		237.16	1695.76		34.68
18	Lastname109, FirstName	Department: Busser/Barback(400) (400)	IL	3767.54	3077.63			41.12	
19	Lastname109, FirstName	Department: Host(600) (600)	IL	209.52					
20	Lastname109, FirstName	Department: Wait Staff(200) (200)	IL	34358.67	28520.62			102.51	
21	Lastname11, FirstName	Department: Kitchen(700) (700)	IL	3457.2					
22	Lastname110, FirstName	Department: Busser/Barback(400) (400)	IL	3196.11	1583			50.7	