

## TRANSFORMING ROWS IN POWER QUERY – DEMO NOTES

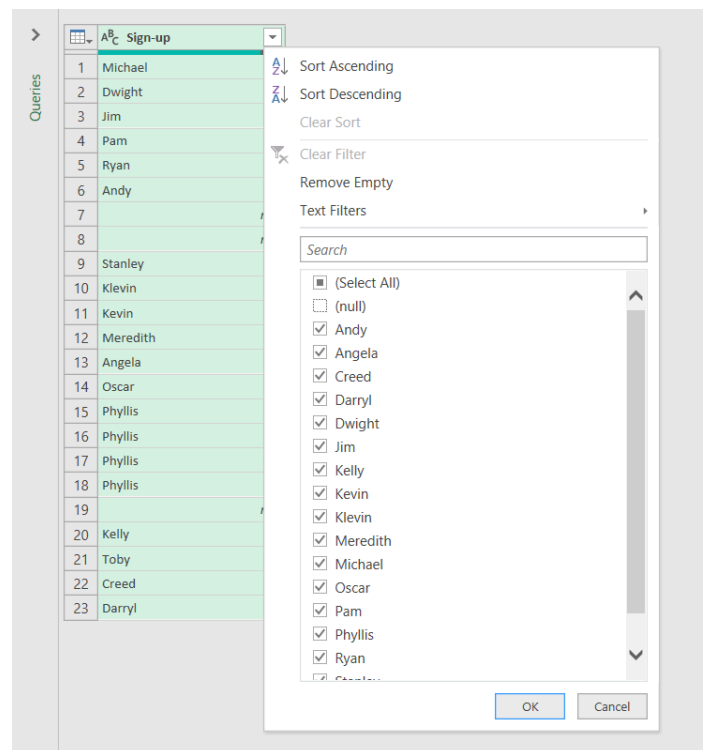
Demo: office-rsvps.xlsx

Worksheet: signups

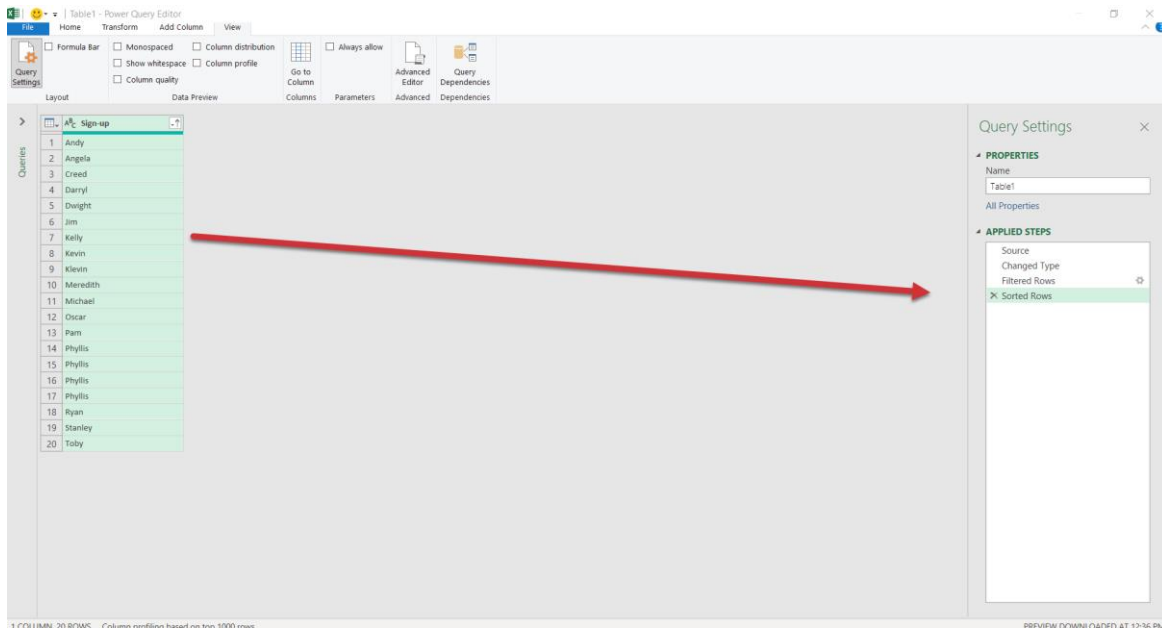
1. You are consulting with the Party Planning Committee to clean up a list of RSVP's to a party. We would like to have the list sorted alphabetically, with duplicates, blanks and misprints removed.
2. This could be accomplished easily enough in classic Excel, but we would like to track each step of the data cleaning process, and we would like a solution that continues to work as more people RSVP to the list. These requirements make Power Query an excellent choice.
3. Create the connection from the range. Your data will be converted into a table.
4. You will see that blank values have been populated as null in Power Query. This is a special value indicating a missing value. It's not the same thing as zero!

7	null
---	------

5. We want to filter out missing records, so select the drop-down on the column label and de-select null. This will remove them.

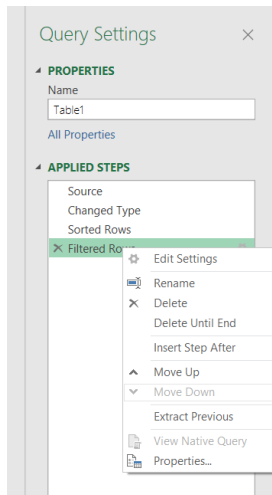


6. We can also sort the list A-Z with the same menu.
7. You will begin to see a running list of the steps we have taken on the right-hand side of the editor (Applied Steps).

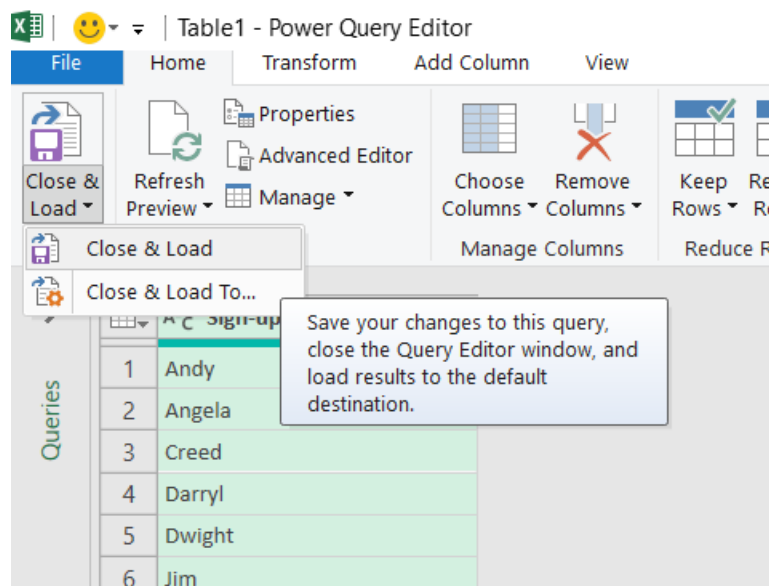


8. Let's remove the third step, Filtered Rows. Our dataset remains sorted A-Z, but nulls are no longer filtered out.
  - a. **Careful: There is no "undo" for removing an applied step!**
9. Go ahead and re-filter the nulls from the data. You will see that becomes the last Applied Step.
10. You can modify the ordering of an Applied Step by right-clicking it.



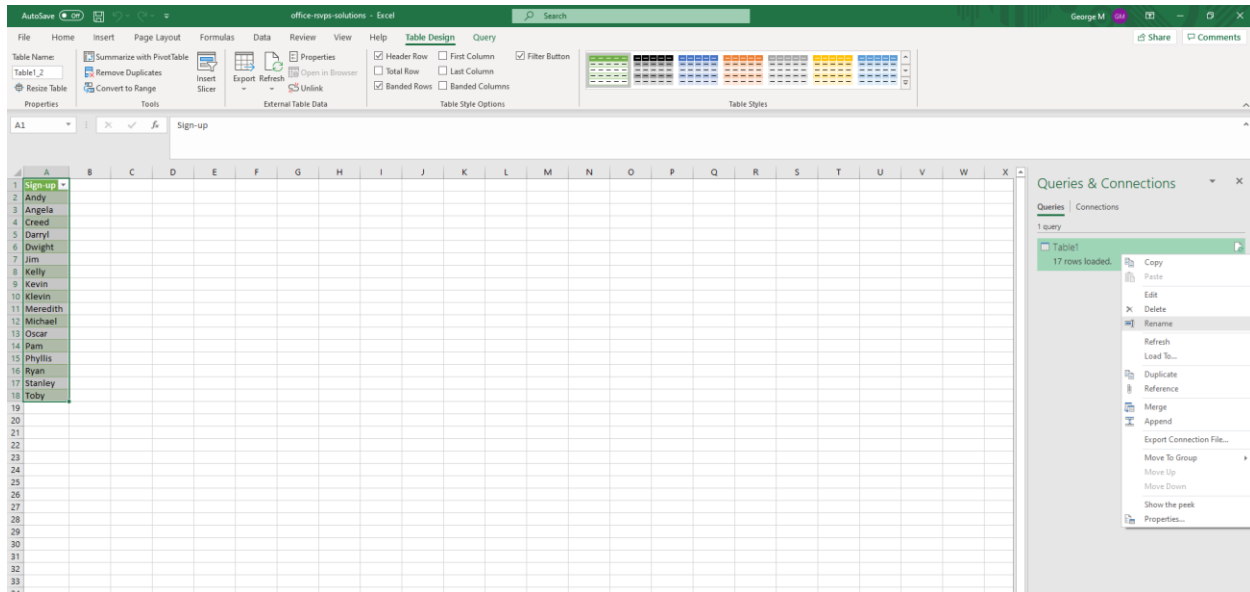


11. Remove duplicates by going to Home on the ribbon, then Remove Rows -> Remove Duplicates.
  - a. You'll also see there is an option here to remove blank rows, this would have been another way to filter out nulls.
12. Last but not least, there is a misprint in the data: a `Klevin` in here. We don't want that either, so filter it out.
13. On the upper left of the Home tab, there is a Close & Load menu. Click the drop-down and select Close & Load.



14. The result of our query has been *loaded* back into Excel (the L part of ETL!).
15. To the right of our table is a Queries & Connections menu. Our query is named Table1. That's not a very descriptive name, so let's rename it to party\_rsvp.
  - a. If you want to close out this menu, you can open it again under Data -> Queries & Connections.



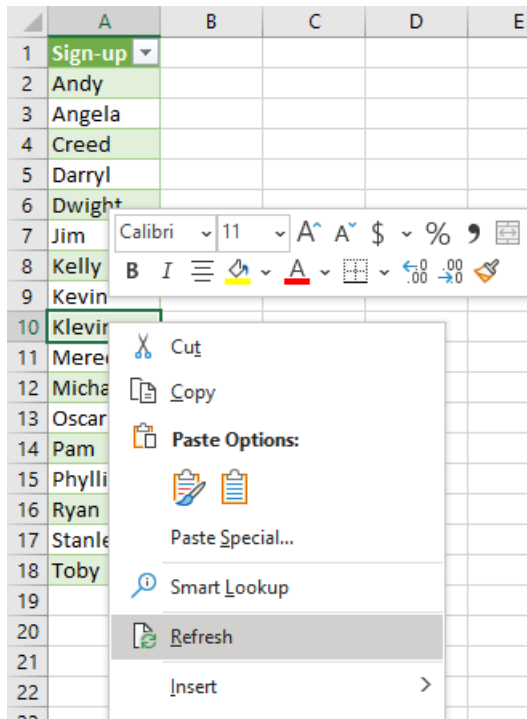


16. Now, any changes made to our source data will be re-loaded into Power Query, go through each step of the data-cleaning process, and be loaded into this new table upon refresh.
17. For an example, I am going to insert two lines into my table, Roy and a blank row.

	A	B
1	Sign-up	
2	Michael	
3	Dwight	
4	Jim	
5	Pam	
6	Ryan	
7	Andy	
8		
9		
10	Stanley	
11	Klevin	
12	Kevin	
13	Meredith	
14	Angela	
15	Oscar	
16	Phyllis	
17	Phyllis	
18	Phyllis	
19	Phyllis	
20		
21	Kelly	
22	Toby	
23	Creed	
24	Roy	
25		
26	Darryl	
27		
28		



18. Go back to the loaded query, right-click and select Refresh.



19. Roy made it into the RSVP, the blank didn't and the results remain sorted alphabetically!

Worksheet: roster

This time, the data has been created with commas separating each name by department. You would like to set up a report to automatically count how many people signed up from each department.

1. Bring the table into Power Query as usual.
2. Click the column and head to Transform > Split Column > By Delimiter.
3. We do want to split by each occurrence of a comma. We also want to click on Advanced Options and select "Split into Rows."



## Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

Comma

Split at

- ☐ Left-most delimiter  
☐ Right-most delimiter  
☒ Each occurrence of the delimiter

Advanced options

Split into

- ☐ Columns  
☒ Rows

Quote Character

"

☐ Split using special characters

Insert special character

OK

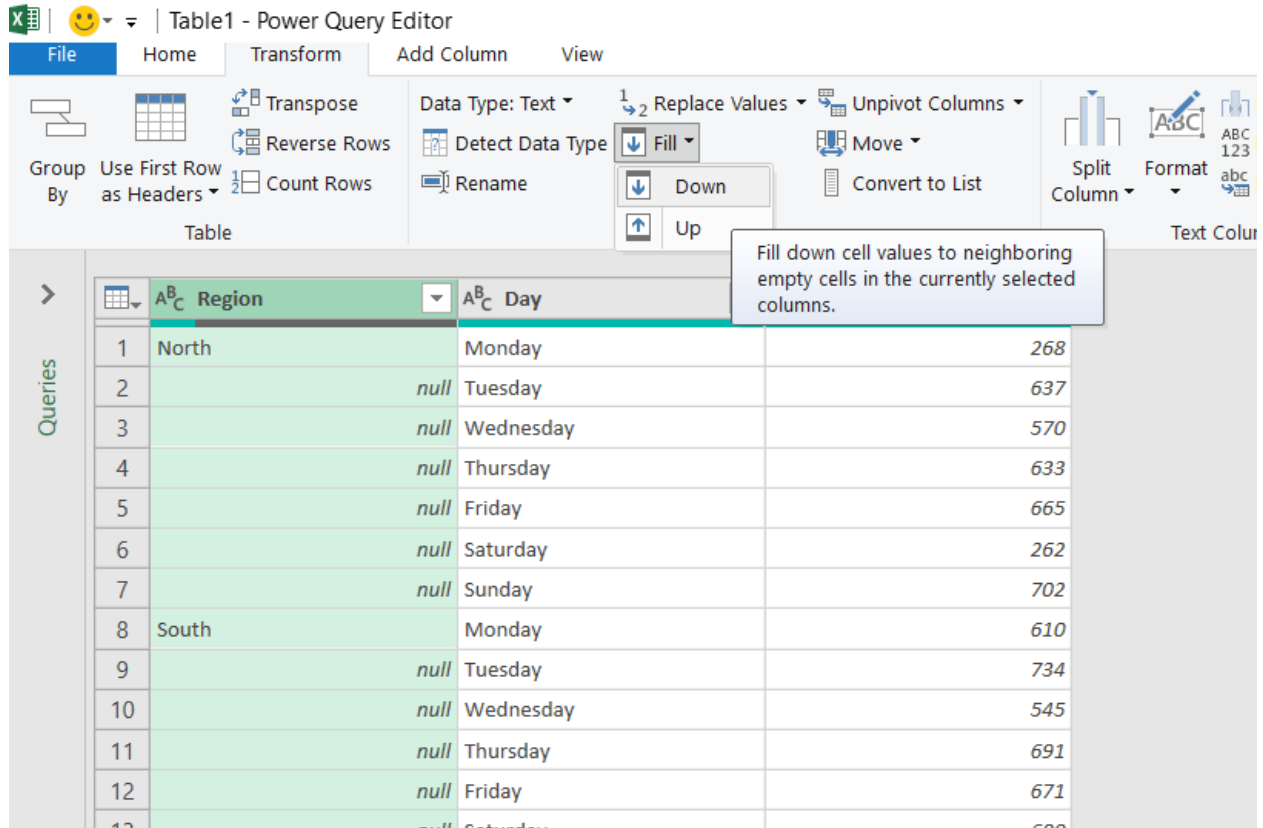
Cancel

4. Click OK.
5. It looks like there is some leftover white space from this delimiting, so let's clean that up.
  - a. Right-click on the column, select Transform and Trim.
6. Close and load. Now our data is tidy.

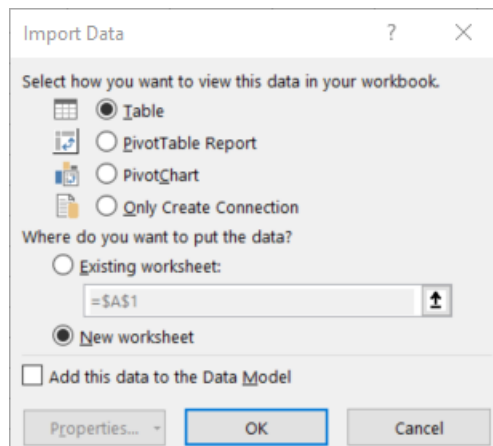
## Demo: regional-sales.xlsx

1. This table does not have a header row and we need to fill down the Region fields. We would like to feed this data into a PivotTable for easy analysis.
2. Import our data into Power Query; remember that this time our Table does *not* have headers.
  - a. We can rename the columns by double-clicking on them in the Query Editor. Name the three columns Region, Day and Amount, respectively.
3. To fill down the blanks for Region, highlight that column by clicking on it, then go to the Transform tab on the ribbon, you will select Fill, Fill Down to fill the nulls down with blanks.





4. We are ready to close and load this data. This time, select Close & Load To. This will give us some options for how to load the data:
  - a. By default, Power Query loads into an Excel Table.
  - b. We can also load it into a PivotTable or PivotChart. (PivotTable Report = PivotTable)
  - c. Finally, there is the connection to only create connection. This means that the query is available in your workbook but not loaded into any worksheet.
  - d. Note the checkmark at the bottom, “Add this data to the Data Model.” This would be if you wanted to build a relational schema in your workbook using Power Pivot.



5. Select PivotTable and we can build a PivotTable from the data just like any other dataset.

Drill: state-populations.xlsx

Worksheet: states

1. Name the query State\_populations.
2. Remove the United States row from the data.
3. Fill down blanks on the Region and Division columns
4. Sort by Population from high to low
5. Load results into a PivotTable

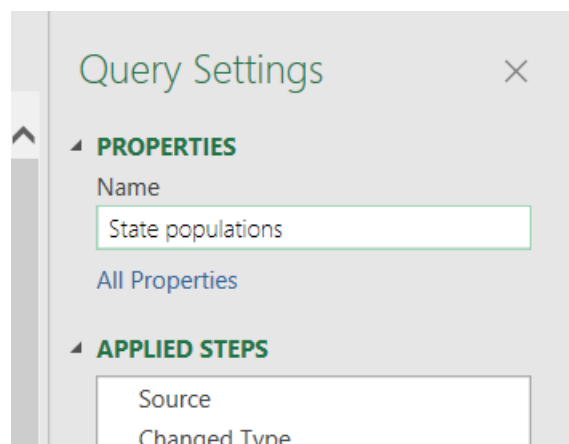
Worksheet: midwest\_cities

These are the 50 largest cities in the Midwest.

1. Convert this data into a table so that each city is in its own row.

Notes for drill:

1. Operate on two columns at a time by holding down Ctrl and selecting each.
2. It's also possible to rename a query in the Query Settings menu within the Query Editor.



Additional demonstration on State populations:





1. It's possible to group/aggregate data in Power Query as you would using SUMIFS or a PivotTable.
2. As an example, right-click on region and select Group By.
3. For example, we can aggregate this data by total population by region by creating a new column total\_population which is the sum of the population field.

The screenshot shows the Power Query Editor interface. On the left, a table with 4 rows and 2 columns is visible. The columns are 'region' and 'total\_population'. The rows are: 1 Midwest (67941429), 2 Northeast (56209510), 3 South (122319574), and 4 West (78657000). A 'Group By' dialog box is open in the center, showing the 'region' column selected for grouping, with 'Sum' as the operation and 'population' as the column to sum. The 'New column name' is 'total\_population'. On the right, the 'Query Settings' pane is open, showing the 'APPLIED STEPS' list. The steps are: Source, Changed Type, Filled Down, Filtered Rows, Sorted Rows, and 'X Grouped Rows'. The 'Grouped Rows' step is highlighted with a red circle, and a red 'X' is next to it.

	region	total_population
1	Midwest	67941429
2	Northeast	56209510
3	South	122319574
4	West	78657000

Group By

Specify the column to group by and the desired output.

\* Basic    ○ Advanced

region

New column name: total\_population    Operation: Sum    Column: population

OK    Cancel

Query Settings

PROPERTIES

Name: State populations

All Properties

APPLIED STEPS

- Source
- Changed Type
- Filled Down
- Filtered Rows
- Sorted Rows
- X Grouped Rows

2 COLUMNS, 4 ROWS    Column profiling based on top 1000 rows    PREVIEW DOWNLOADED AT 1:08 PM

4. To look back at the settings of prior Applied Steps, click on the gear-wheel to the right of that step where applicable.

