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!



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

Graphical user interface, table

Description automatically generated

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

Graphical user interface, application, Word

Description automatically generated

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

Graphical user interface, application

Description automatically generated

1. Remove duplicates by going to Home on the ribbon, then Remove Rows -> Remove Duplicates.
   1. You’ll also see there is an option here to remove blank rows, this would have been another way to filter out nulls.
2. 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.
3. On the upper left of the Home tab, there is a Close & Load menu. Click the drop-down and select Close & Load.

Graphical user interface, application

Description automatically generated

1. The result of our query has been *loaded* back into Excel (the L part of ETL!).
2. 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.
   1. If you want to close out this menu, you can open it again under Data -> Queries & Connections.

Application, table, Excel

Description automatically generated

1. 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.
2. For an example, I am going to insert two lines into my table, Roy and a blank row.

Table

Description automatically generated

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

Graphical user interface, application, table, Excel

Description automatically generated

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

Worksheet: office-rsvps

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.”

Graphical user interface, text, application, email

Description automatically generated

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

Worksheet: regional-sales

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.
   1. 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.

Graphical user interface, table

Description automatically generated

1. 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:
   1. By default, Power Query loads into an Excel Table.
   2. We can also load it into a PivotTable or PivotChart. (PivotTable Report = PivotTable)
   3. 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.
   4. 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.

Graphical user interface, text, application

Description automatically generated

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

Worksheet: populations

1. Load into Power Query
2. First, create a concatenated field in the format: Name (Abbreviation)
   1. Add Column on the ribbon, then Custom Column.
   2. Name this column state-full. Use ampersands to concatenate strings:

Graphical user interface, text, application

Description automatically generated

* 1. Click link at the bottom of this menu to “Learn about Power Query formulas”: – this is M code.
  2. Our new column is added to Applied Steps. We can view the formula using the gear box.

1. Move the column to the front of this dataset by holding down Control and dragging it to the front.
2. Delete the two columns we had referred to in our formula. We can delete them and not break our calculated column.
3. We want to calculate population density. Rather than calculate the density for each year, we can “tidy” this dataset to get one, “population” variable, then calculate the density for each year in one fell swoop.
4. To create a “Year” column, select all but the 2010-2016 columns, then right-click and select Unpivot Other Columns.

Table

Description automatically generated

1. We can re-name the Attribute and Value columns to Year and Population, respectively.
2. Now we can create another custom column formula, Density, which is Population/state-size.

Graphical user interface, text, application

Description automatically generated

1. Finished! Close & load.

**Demo: dvdrentals.xlsx**

1. Create the query from the source table.
2. Convert Title and Artist Name to proper case by right-clicking the column and selecting Transform -> Capitalize Each Word.
3. There are no spaces after commas or colons. Add them by right-clicking on the headers and selecting Replace Values. Replace commas, then colons with each character followed by a space.

Graphical user interface

Description automatically generated

1. Split Item # into two columns based on the space delimiter by right-clicking on the column and selecting Split Column -> By Delimiter -> Space.
2. The UPC and ISBN 13 columns are probably better classified as strings than numbers. Change their types by clicking on the number icons to the left of their column headers and changing to text.
3. We don’t need the BTKey column. Simply select it and hit Delete on your keyboard
4. Now, change the Retail column from Decimal to Currency.
5. Finally, convert the Release Date column into three columns, Year, Month and Day:
   1. Right-click the column label and select Duplicate column. Do this twice so there are three Release Date columns in total.
   2. Right-click the first one, and select Transform -> Year -> Year.
   3. Do the same for the remaining columns, but for Month and Day.
   4. Now, rename these columns as Year, Month and Day