JOIN BEYOND THE BASICS – DEMO NOTES

Demo: `championships.xlsx`

We would like to find what cities can claim *only* a baseball or football championship.

0. Preface: This data has been wrangled using Column From Examples. This is a powerful way to add a conditionally-formatted column to a table.
    a. To do this, open the football query, select WINNER field and head to the query editor and Add Column > Column From Examples > From Selection.
    b. What we want to do is start typing the name of the team in the new column. Power Query will start to use conditional logic to begin to complete the field for us.
    c. This is an iterative process. Power Query might get things right at first and then not later. Eventually it should get to "the truth" as determined by you. You can then click OK and use the column in your query.

0. Back to the task at hand: We want to find what teams have a baseball championship and not a football championship.
1. Open up the baseball query in the editor and go to Home > Merge Queries > Merge Queries as New.
    a. This way we don't write over this current query, we make a new query.

b. This will be a left anti join, to get the cities that have a baseball and not a football win.

Merge

Select tables and matching columns to create a merged table.

baseball

SEASON	WINNER	Team	City
2018	Boston Red Sox	Red Sox	Boston
2017	Houston Astros	Astros	Houston
2016	Chicago Cubs	Cubs	Chicago
2015	Kansas City Royals	Royals	Kansas City
2014	San Francisco Giants	Giants	San Francisco

football

SEASON	WINNER	City	Team
2020	Kansas City Chiefs	Kansas City	Chiefs
2019	New England Patriots	New England	Patriots
2018	Philadelphia Eagles	Philadelphia	Eagles
2017	New England Patriots	New England	Patriots
2016	Denver Broncos	Denver	Broncos

Join Kind

Left Anti (rows only in first)

☐ Use fuzzy matching to perform the merge

▷ Fuzzy matching options

✔ The selection excludes 83 of 114 rows from the first table.

OK   Cancel

2. Click OK. You are going to see a new column "football" in our query which we can expand, however since we are only keeping the baseball records, this is going to be all blank.
   a. Since it's a blank field, let's delete it.

SEASON	WINNER	City	Team
null	null	null	null

5 COLUMNS, 31 ROWS   Column profiling based on top 1000 rows

3. Here we can see all the cities that have a baseball win but not a football win. We could clean this up further if we wanted by removing the other fields and then going to Home > Remove Rows > Remove Duplicates.
4. Let's rename this query as baseball_only.
   a. One quick thing to notice about our data – we see for example that "Florida" is listed as a city because that is the name of the team. Currently, the Florida Marlins are the Miami Marlins – and the Miami *Dolphins* have won a Super Bowl, so we could dispute whether this one should be on the list.
      i. There are lots of other ways to nitpick our results, what else can you think of?

Let's now find cities that have a football but not a baseball win.
1. Go back to the baseball query and select Home > Merge Queries > Merge Queries as New.
2. This time we will want a right anti-join, to get only the cities with just a football championship.

Merge

Select tables and matching columns to create a merged table.

baseball ▾

SEASON	WINNER	Team	City
2018	Boston Red Sox	Red Sox	Boston
2017	Houston Astros	Astros	Houston
2016	Chicago Cubs	Cubs	Chicago
2015	Kansas City Royals	Royals	Kansas City
2014	San Francisco Giants	Giants	San Francisco

football ▾

SEASON	WINNER	City	Team
2020	Kansas City Chiefs	Kansas City	Chiefs
2019	New England Patriots	New England	Patriots
2018	Philadelphia Eagles	Philadelphia	Eagles
2017	New England Patriots	New England	Patriots
2016	Denver Broncos	Denver	Broncos

Join Kind

Right Anti (rows only in second) ▾

☐ Use fuzzy matching to perform the merge

▷ Fuzzy matching options

✔ The selection excludes 30 of 54 rows from the second table.

[OK]   [Cancel]

3. This time it looks like we didn't get any data, however that's because all of it is "hidden" in that "football" field. Go ahead and click on it to expand. We can then get rid of the null baseball records.
   a. We now have a list of cities who have a football but no baseball championship.
   b. Let's name this query football_only.

Drill: `championships-2.xlsx`

Which cities can claim *only* a hockey or basketball championship?
(Just fill out the city name, you don't need to create a team name column.)

Demo: `office-employees.xlsx`
Worksheet: `get-to-know-you`

HR wants to set up a get-to-know-you activity for the sales team. You need to set up a table so that each salesperson can fill out their favorite color, food, sport to play and sport to watch.

We can do this with a cross join in Power Query:

5. We've already loaded each of these tables in as queries. Click into the `names` query to edit.
6. Copy and paste the `names` query and rename it `get_to_know_you`.
7. We want to add a custom column (Add Column > Custom Column). We will name this column `favorite_things`.
   a. The formula for our column will be `favorite_things`. This is another query that shows up in the Intellisense.

Custom Column

Add a column that is computed from the other columns.

New column name

favorite_things

Custom column formula ⓘ   Available columns

= favorite_things   name_last
                                   name_first

                                   << Insert

Learn about Power Query formulas

✔ No syntax errors have been detected.   OK   Cancel

8. Click OK. Now if you click on any of the `favorite_things` cells, you can get a preview of the resulting data at the bottom of your screen:

	A^Bc name_last	A^Bc name_first	ABC 123 favorite_things
1	Halpert	Jim	Table
2	Scrhute	Dwight	Table
3	Vance	Phyllis	Table
4	Hudson	Stanley	Table
5	Bernard	Andy	Table

Favorite things
Color
Food
Sport to play
Sport to watch

3 COLUMNS, 5 ROWS   Column profiling based on top 1000 rows

9. Go ahead and expand the data now. We will get this in a tabular form now.
10. To pivot this table to make a checklist, we first need a "values" column to pivot on. This is blank for now so we can insert a blank or null field:

11. Now we need to "pivot" on top of `favorite_things`, based on the `values` column.
    a. Select the `favorite_things` column and go to Transform > Pivot Column.
    b. Select values as the column to pivot on, then select Advanced Options and choose "Don't Aggregate" as your aggregate value function.
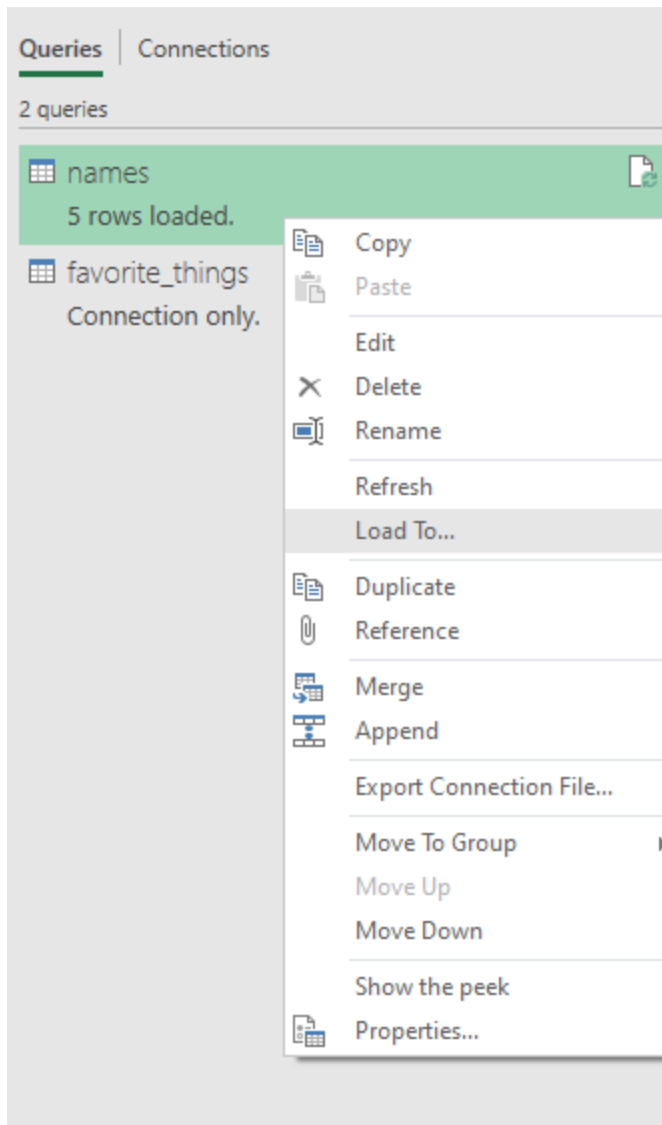
12. We now have a "checklist" table that we can load into Excel.

name_last	name_first	Color	Food	Sport to play	Sport to watch
1	Bernard	Andy	null	null	null	null
2	Halpert	Jim	null	null	null	null
3	Hudson	Stanley	null	null	null	null
4	Scrhute	Dwight	null	null	null	null
5	Vance	Phyllis	null	null	null	null

13. Currently the names query is loading to a connection only. If we want to change that we can right-click on the query and select Load To.

Queries | Connections

2 queries

names
5 rows loaded.

- Copy
- Paste
- Edit
- Delete
- Rename
- Refresh
- Load To...
- Duplicate
- Reference
- Merge
- Append
- Export Connection File...
- Move To Group
- Move Up
- Move Down
- Show the peek
- Properties...

favorite_things
Connection only.

Drill: `states.xlsx`

Create a table to record each state's bird, flower and capital.

Demo note: Note that we can add a new property to our table, and refresh it and get that added, for example we can add the state song to the worksheet.

	A	B	C	...	K	L	M	N	O
1	state		state_object						
2	Alabama		bird						
3	Alaska		flower						
4	Arizona		capital						
5	Arkansas		song						
6	California								
7	Colorado								
8	Connecticut								
9	Delaware								
10	Florida				state	bird	flower	capital	song
11	Georgia				Alabama				
12	Hawaii				Alaska				
13	Idaho				Arizona				
14	Illinois				Arkansas				
15	Indiana				California				
16	Iowa				Colorado				
17	Kansas				Connecticut				
18	Kentucky				Delaware				
19	Louisiana				Florida				
20	Maine				Georgia				
21	Maryland				Hawaii				
22	Massachusetts				Idaho				
23	Michigan				Illinois				
24	Minnesota				Indiana				
25	Mississippi				Iowa				
26	Missouri				Kansas				
27	Montana				Kentucky				
28	Nebraska				Louisiana				
29	Nevada				Maine				
30	New Hampshire				Maryland				
31	New Jersey				Massachusetts				
32	New Mexico				Michigan				
33	New York				Minnesota				
34	North Carolina				Mississippi				
35	North Dakota				Missouri				
36	Ohio				Montana				
37	Oklahoma				Nebraska				
38	Oregon				Nevada				
39	Pennsylvania				New Hampshire				