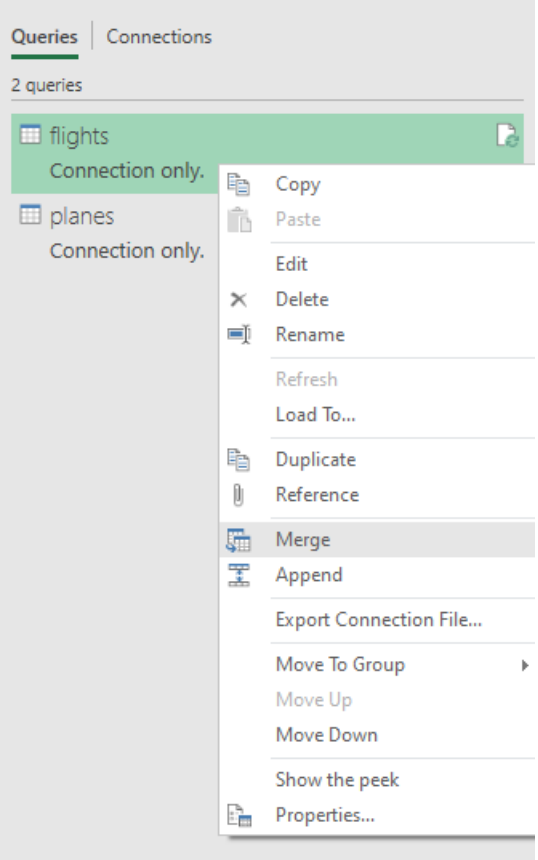
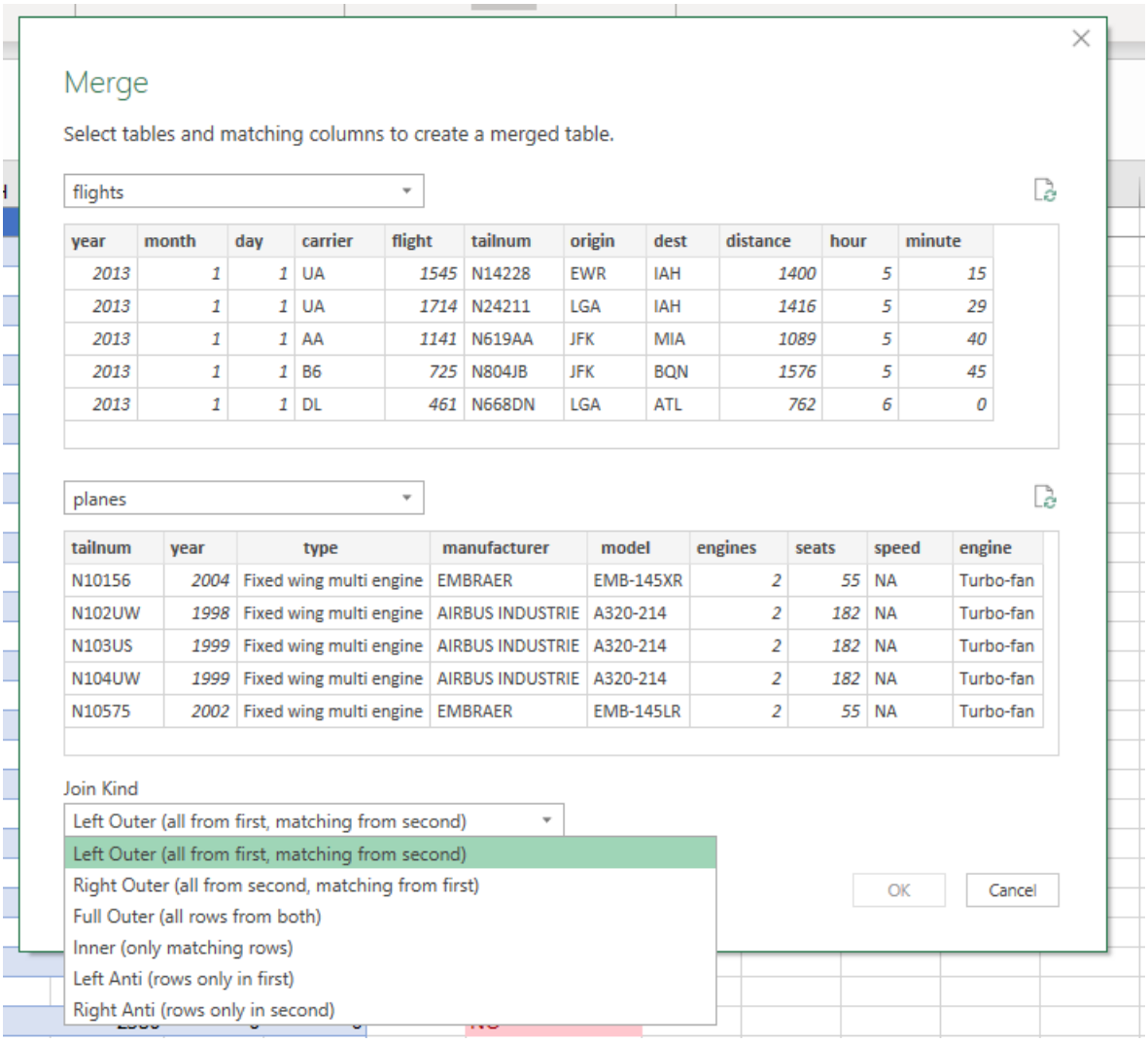
VLOOKUP(), MEET JOIN – DEMO NOTES

Demo: `flights-and-planes.xlsx`

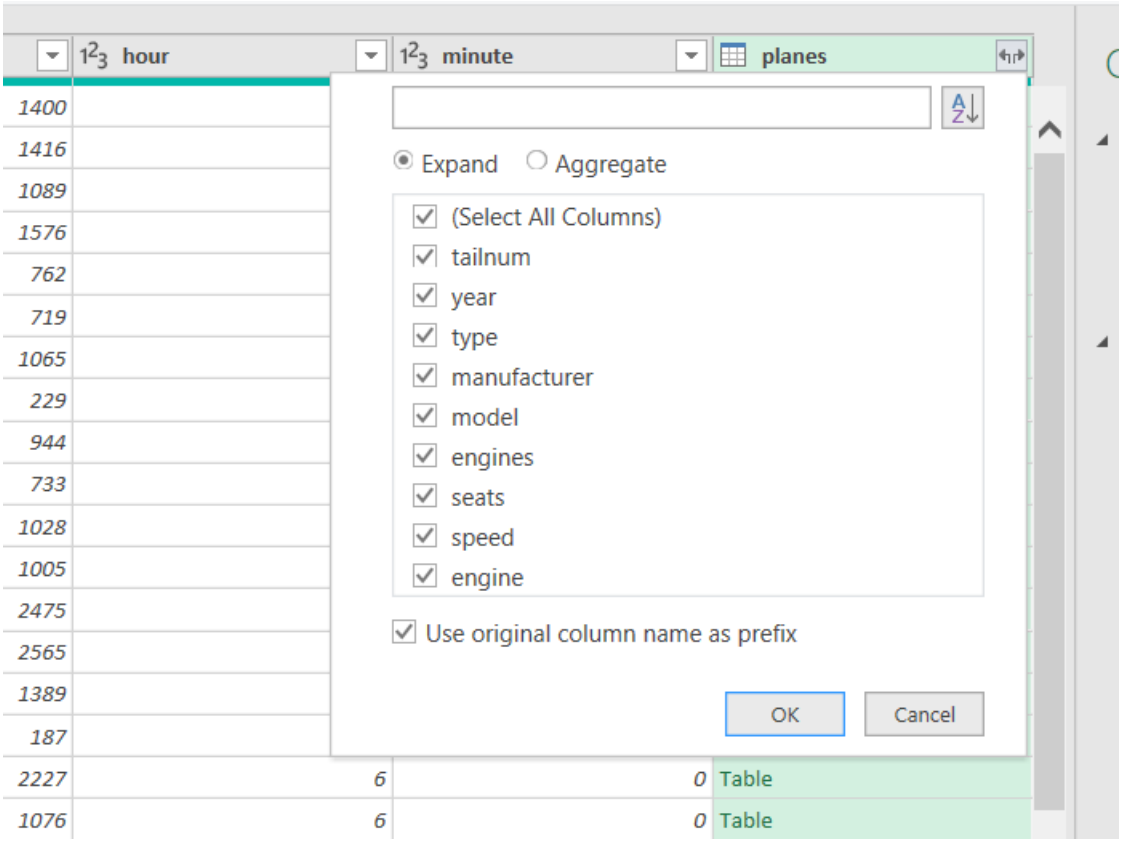
1. We have a table of flights and tables of planes. The "lookup value" is `tailnum` but there is not a "match" for all of them (See `Found in planes?` column to confirm.)
    a. So, when we "look up" this plane information into our flights table, do we want to keep the information about the records without a match? Essentially we are asking, when we join `flights` on `planes` do we want to use a left outer join or an inner join?
2. Load both tables in Power Query and create only a connection for each.
3. In the Queries & Connections menu, right-click on flights and select Merge.



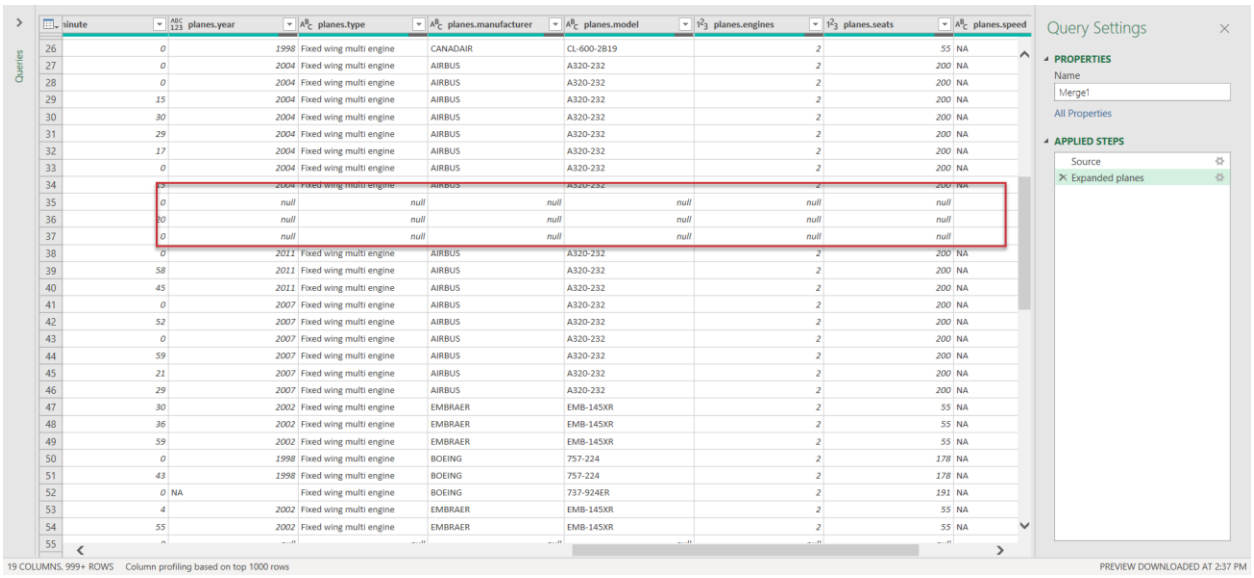
4. We will now create a merged table. We will merge flights on planes. Leave the Join Kind as Left Outer, but check out all the options available on the drop-down.



5. We can't hit OK until we specify *what* we want to join on. In VLOOKUP( )-ese, this would be our "lookup value" which in this case is `tailnum`.
6. We'll get a green check-mark saying it's matched X out of Y rows from the first table. We knew there were going to be some non-matches, so this number makes sense.
7. Hit OK, we get a new query, now we have an accordion-style menu here where we can select any of the returned fields into our merged table. We already have tailnum included in the table since that's what we joined on, so probably we don't need that one.

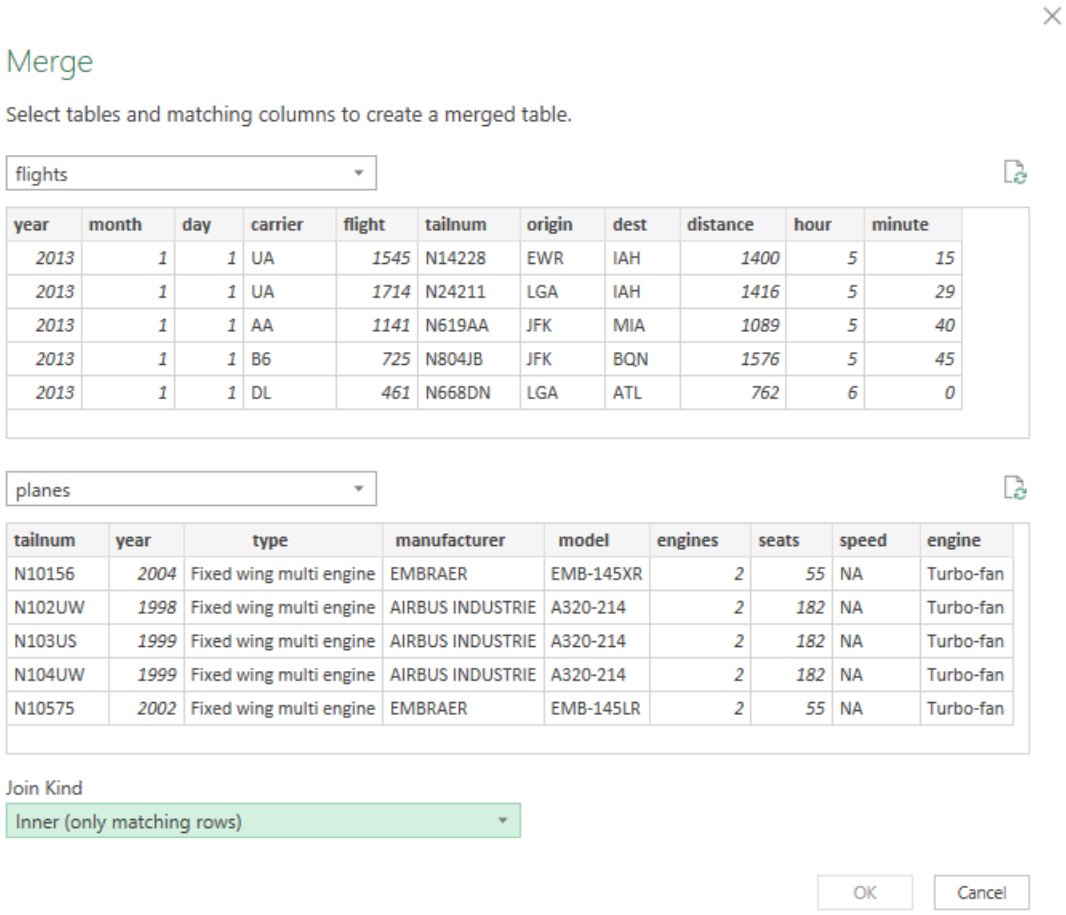


8. You'll see that each of these are named planes.field name. Undo our Expanded step to see why: Hit the accordion again. You'll see the option to "Use original column name as prefix" is checked on.
    a. This is not a terrible idea, for example there is a `year` field for the planes data and a `year` field for the flights field (one for when the plane was built, one for when the flight took place). So this way we easily know which is which.
9. Scroll down the resulting table and we can see there are rows of null's where there was no match for the planes data:



	minute	planes.year	planes.type	planes.manufacturer	planes.model	planes.engines	planes.seats	planes.speed
26	0	1998	Fixed wing multi engine	CANADAIR	CL-600-2B19	2	55	NA
27	0	2004	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
28	0	2004	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
29	15	2004	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
30	30	2004	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
31	29	2004	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
32	17	2004	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
33	0	2004	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
34	15	2004	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
35	0	null	null	null	null	null	null	
36	30	null	null	null	null	null	null	
37	0	null	null	null	null	null	null	
38	0	2011	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
39	58	2011	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
40	45	2011	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
41	0	2007	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
42	52	2007	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
43	0	2007	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
44	59	2007	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
45	21	2007	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
46	29	2007	Fixed wing multi engine	AIRBUS	A320-232	2	200	NA
47	30	2002	Fixed wing multi engine	EMBRAER	EMB-145XR	2	55	NA
48	36	2002	Fixed wing multi engine	EMBRAER	EMB-145XR	2	55	NA
49	59	2002	Fixed wing multi engine	EMBRAER	EMB-145XR	2	55	NA
50	0	1998	Fixed wing multi engine	BOEING	757-224	2	178	NA
51	43	1998	Fixed wing multi engine	BOEING	757-224	2	178	NA
52	0	NA	Fixed wing multi engine	BOEING	737-924ER	2	191	NA
53	4	2002	Fixed wing multi engine	EMBRAER	EMB-145XR	2	55	NA
54	55	2002	Fixed wing multi engine	EMBRAER	EMB-145XR	2	55	NA
55								

10. Now we can close and load the table and I am going to name it `left_join`.
11. Take the same steps except this time we will do an inner join of flights on planes.



12. Another green light.
13. Same steps, expand the resulting columns and load the table.
14. Check it out, this time there are only 284K rows loaded. Why? Well we can take a look here, there are no more NULL's for the plane info, those have been removed from the join. So it stands to reason there would be fewer rows this time.
15. Name the query `inner_join`.

Drill: `hof.csv, people-a-thru-m.csv`

1. What is the result of a left outer join of `hof` on `people-a-thru-m`?
2. What about an inner join?