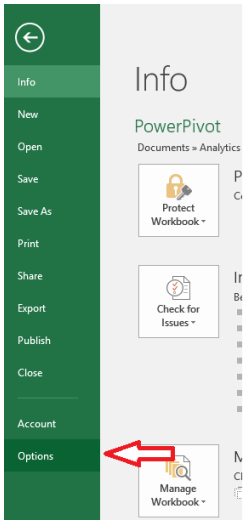
Lab 03 – Microsoft Excel

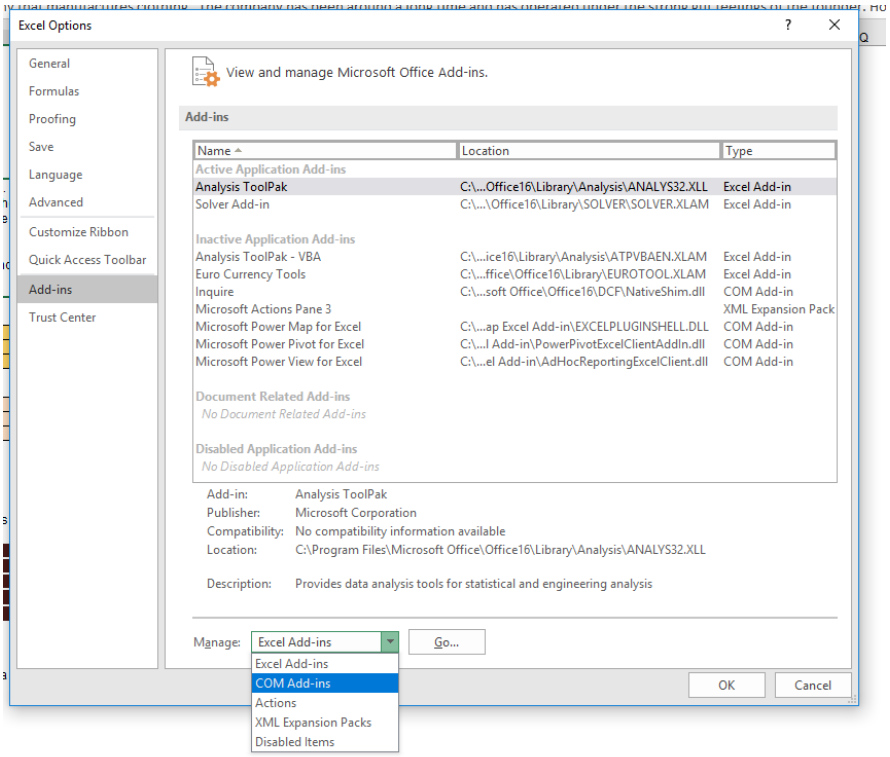
PowerPivot

There are many ways to analyse and summarise data in Excel, one feature is called PowerPivot. It is similar to the PivotTable feature you covered in the previous labs. However, PowerPivot is able to run pivot tables by merging several big datasets and run pivot tables off the consolidated data.

To use PowerPivot you must be familiar with the Internal Data Model (IDM). IDM is the analytical engine that Excel uses behind the scenes to construct the PowerPivot. It is basically a database in which Excel organises information. Two features that are useful: first, it can establish relationships between multiple databases. Second, it can hold unlimited number of rows and columns. You are only limited by a 2 gigabyte workbook and the memory available on the machine where Excel is running. (IDM is only available in Excel 2013 or newer)



To activate PowerPivot we need to add it to our available tabs in Excel. Go to the File menu and choose options.

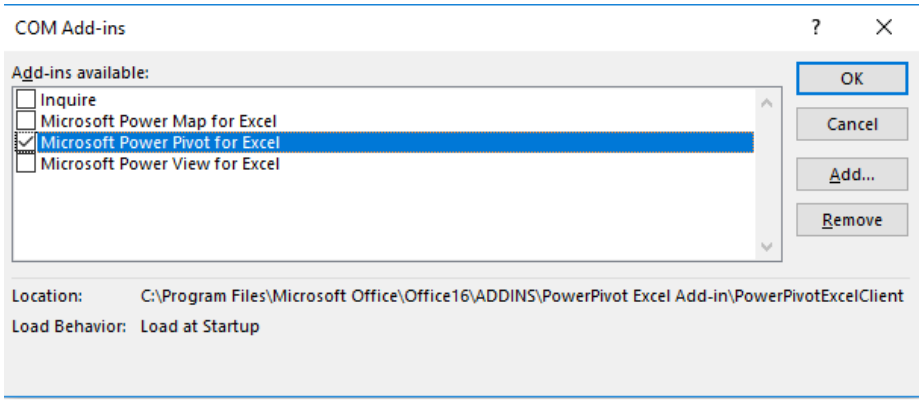


Excel options box opens, choose Add-ins from the menu on the left and then choose COM Add-ins from the drop down list Label Managed:

Click Go.

Check the box for Microsoft

PowerPivot for Excel and click OK.



Now you should see the PowerPivot tab appear on the top ribbon in Excel.

Linking Data Tables

We will work on some Airline data for this step. Download the file Power Pivot Lesson. There are four data tables in this workbook each containing different kinds of airline data.

The first is Flights, it contains information relating to flights out of Chicago, including:

- destination
- the ID of the aircraft
- the scheduled departure time
- pricing information
- date of the flight

The second is Routes, it provide insight into:

- the specific departure and arrival airports
- the distance
- list price

Note that all flights are coming out of ORD with is the airport code for O'Hare International Airport Chicago.

The third data table is the Aircraft tab, it provides specs for all aircraft in the fleet:

- type of aircraft,
- seat capacity

- fuel cost per seat per mile

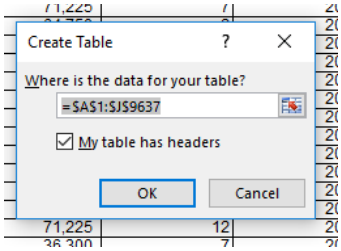
The last data table is the Airports tab which provides:

- airport full name
- number of available gates

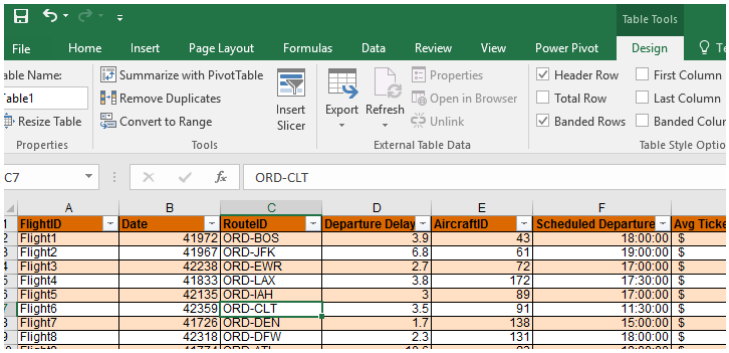
This lesson will address how to analyse this data, it is no simple task as the data is spread over several tabs. For example if the business asked you to analyse the types of aircraft that frequent certain routes and have delays, it would require you to combine data from the flights, aircraft and routes tabs. Instead of going back and forth between tabs we can use the relational data model and PowerPivot to analyse data across multiple tabs.

First we convert the flights data into a table format so that we can use it in our PowerPivot data model.

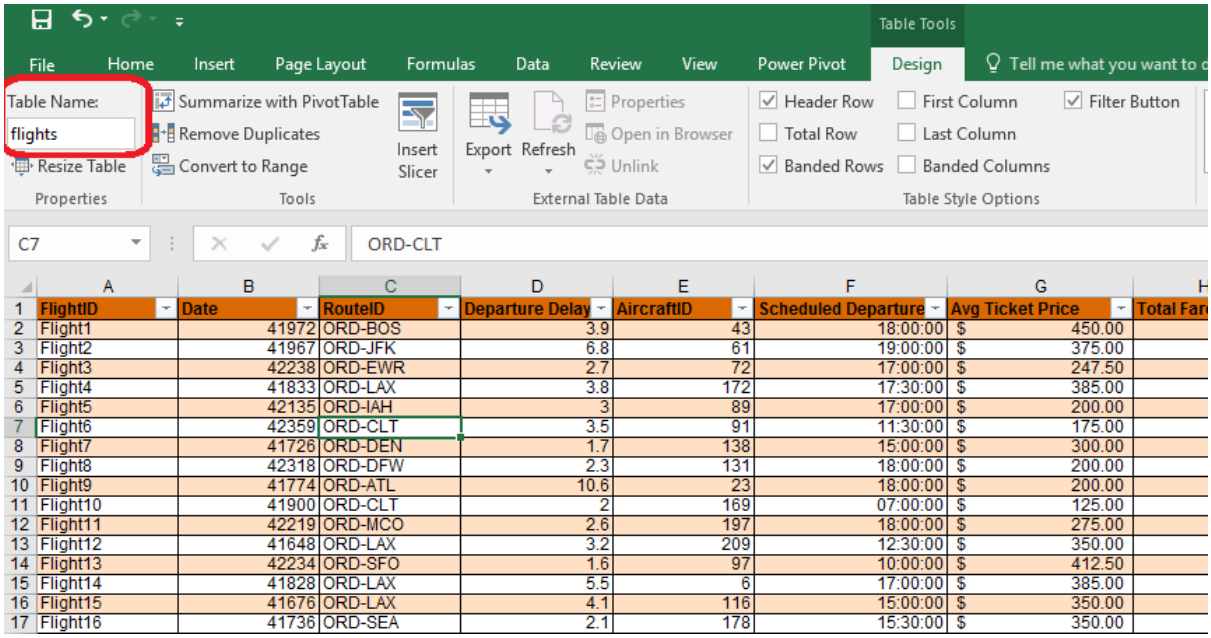
Click on the flights tab, select the entire dataset, navigate to the insert menu ribbon and choose Table from the options available. You will see a dialog box open to confirm the data range and confirm that the data has a row of headers. It is considered best practice to include headers when selecting a data set. This allows Excel to add informative field names in the data model and clarity to the user.



Click OK. The dataset is now formated in a default table style. You will notice when the table is active that another tab called Table Tools Design appears on the main ribbon and the name of the table is presented to you.



Now when we select any cell the design tab appears. We can change the name of the table as it is good practice to change it from the default. Let's call this set of data flights.

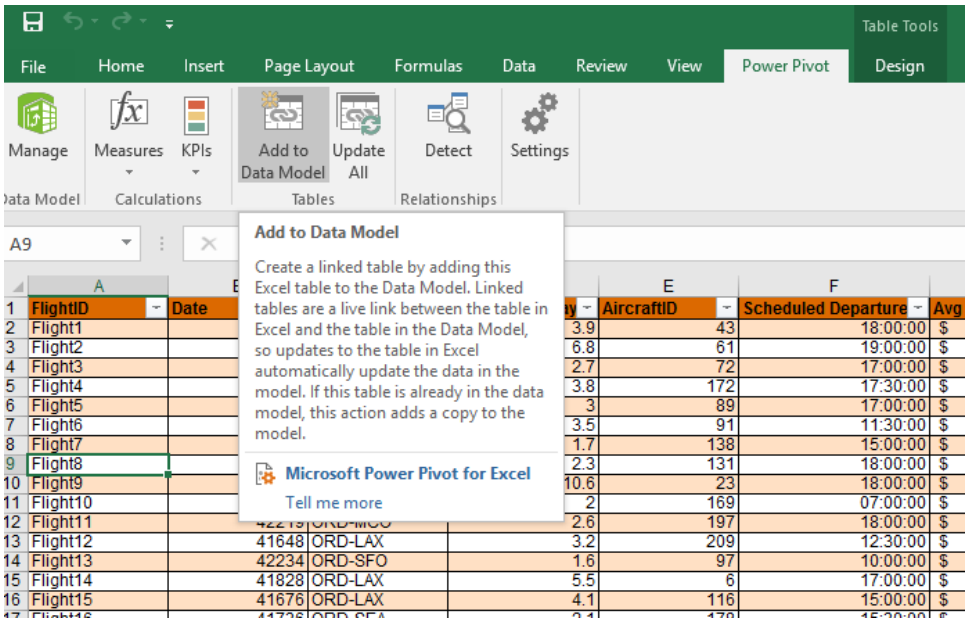


	A	B	C	D	E	F	G	H
1	FlightID	Date	RouteID	Departure Delay	AircraftID	Scheduled Departure	Avg Ticket Price	Total Far
2	Flight1	41972	ORD-BOS	3.9	43	18:00:00	$ 450.00	
3	Flight2	41967	ORD-JFK	6.8	61	19:00:00	$ 375.00	
4	Flight3	42238	ORD-EWR	2.7	72	17:00:00	$ 247.50	
5	Flight4	41833	ORD-LAX	3.8	172	17:30:00	$ 385.00	
6	Flight5	42135	ORD-IAH	3	89	17:00:00	$ 200.00	
7	Flight6	42359	ORD-CLT	3.5	91	11:30:00	$ 175.00	
8	Flight7	41726	ORD-DEN	1.7	138	15:00:00	$ 300.00	
9	Flight8	42318	ORD-DFW	2.3	131	18:00:00	$ 200.00	
10	Flight9	41774	ORD-ATL	10.6	23	18:00:00	$ 200.00	
11	Flight10	41900	ORD-CLT	2	169	07:00:00	$ 125.00	
12	Flight11	42219	ORD-MCO	2.6	197	18:00:00	$ 275.00	
13	Flight12	41648	ORD-LAX	3.2	209	12:30:00	$ 350.00	
14	Flight13	42234	ORD-SFO	1.6	97	10:00:00	$ 412.50	
15	Flight14	41828	ORD-LAX	5.5	6	17:00:00	$ 385.00	
16	Flight15	41676	ORD-LAX	4.1	116	15:00:00	$ 350.00	
17	Flight16	41736	ORD-SEA	2.1	178	15:30:00	$ 350.00	

Repeat this step for the other three datasets.

Next we complete our data model in PowerPivot.

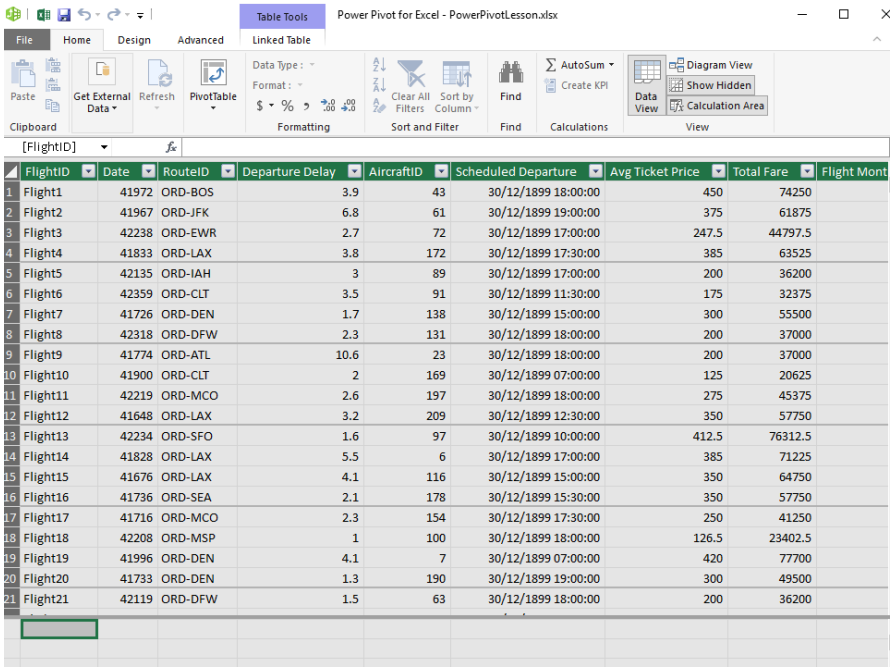
- Start with the flights table, click on a cell in the table
- Navigate to the PowerPivot tab and click on the Add to Data Model button



	A	Date	E	F	Avg
1	FlightID	Date	AircraftID	Scheduled Departure	Avg
2	Flight1		43	18:00:00	$
3	Flight2		61	19:00:00	$
4	Flight3		72	17:00:00	$
5	Flight4		172	17:30:00	$
6	Flight5		89	17:00:00	$
7	Flight6		91	11:30:00	$
8	Flight7		138	15:00:00	$
9	Flight8		131	18:00:00	$
10	Flight9		23	18:00:00	$
11	Flight10		169	07:00:00	$
12	Flight11		197	18:00:00	$
13	Flight12	41648	209	12:30:00	$
14	Flight13	42234	97	10:00:00	$
15	Flight14	41828	6	17:00:00	$
16	Flight15	41676	116	15:00:00	$

The PowerPivot table is then displayed in a separate window.

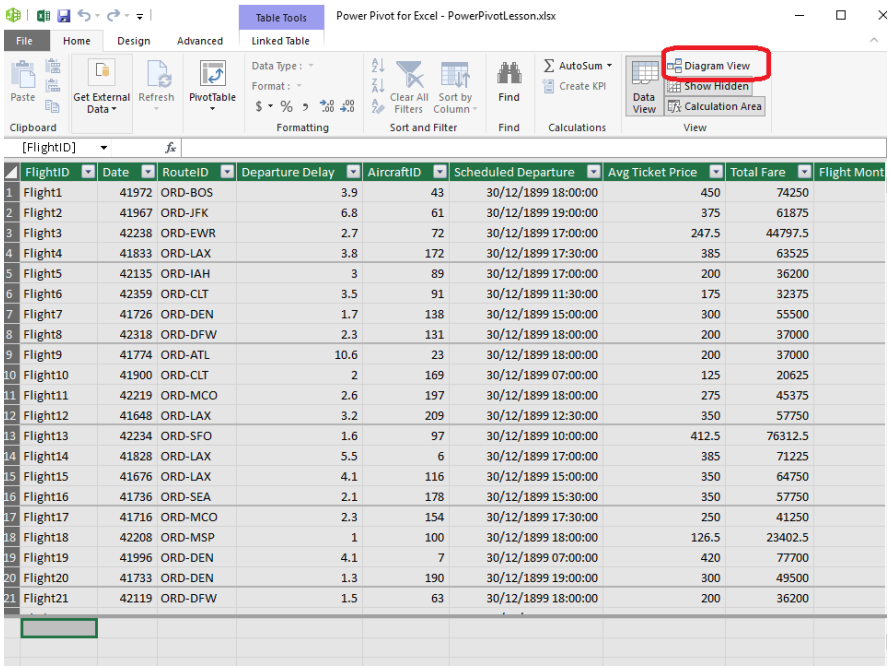
- repeat these steps for each table until all four are added.



	FlightID	Date	RouteID	Departure Delay	AircraftID	Scheduled Departure	Avg Ticket Price	Total Fare	Flight Mont
1	Flight1	41972	ORD-BOS	3.9	43	30/12/1899 18:00:00	450	74250	
2	Flight2	41967	ORD-JFK	6.8	61	30/12/1899 19:00:00	375	61875	
3	Flight3	42238	ORD-EWR	2.7	72	30/12/1899 17:00:00	247.5	44797.5	
4	Flight4	41833	ORD-LAX	3.8	172	30/12/1899 17:30:00	385	63525	
5	Flight5	42135	ORD-IAH	3	89	30/12/1899 17:00:00	200	36200	
6	Flight6	42359	ORD-CLT	3.5	91	30/12/1899 11:30:00	175	32375	
7	Flight7	41726	ORD-DEN	1.7	138	30/12/1899 15:00:00	300	55500	
8	Flight8	42318	ORD-DFW	2.3	131	30/12/1899 18:00:00	200	37000	
9	Flight9	41774	ORD-ATL	10.6	23	30/12/1899 18:00:00	200	37000	
10	Flight10	41900	ORD-CLT	2	169	30/12/1899 07:00:00	125	20625	
11	Flight11	42219	ORD-MCO	2.6	197	30/12/1899 18:00:00	275	45375	
12	Flight12	41648	ORD-LAX	3.2	209	30/12/1899 12:30:00	350	57750	
13	Flight13	42234	ORD-SFO	1.6	97	30/12/1899 10:00:00	412.5	76312.5	
14	Flight14	41828	ORD-LAX	5.5	6	30/12/1899 17:00:00	385	71225	
15	Flight15	41676	ORD-LAX	4.1	116	30/12/1899 15:00:00	350	64750	
16	Flight16	41736	ORD-SEA	2.1	178	30/12/1899 15:30:00	350	57750	
17	Flight17	41716	ORD-MCO	2.3	154	30/12/1899 17:30:00	250	41250	
18	Flight18	42208	ORD-MSP	1	100	30/12/1899 18:00:00	126.5	23402.5	
19	Flight19	41996	ORD-DEN	4.1	7	30/12/1899 07:00:00	420	77700	
20	Flight20	41733	ORD-DEN	1.3	190	30/12/1899 19:00:00	300	49500	
21	Flight21	42119	ORD-DFW	1.5	63	30/12/1899 18:00:00	200	36200	

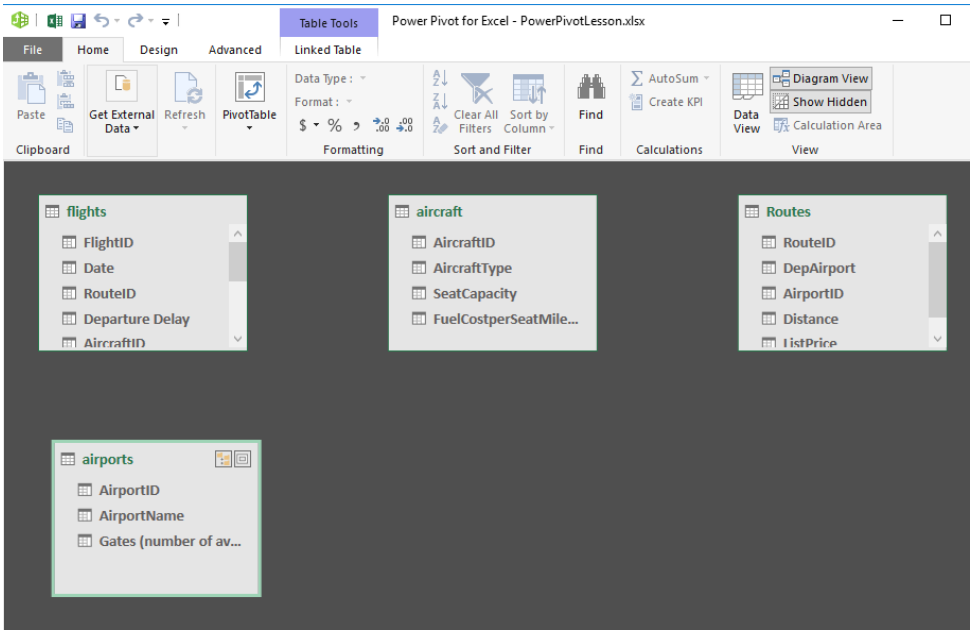
Now that we have added all 4 of the tables to our data model, we need to form the required relationships between the tables to create our PowerPivot.

- Click on the tab called Pivot 1a. Next navigate to the PowerPivot tab, and select the Manage option. It may take a couple of seconds to load the data.
- Click on the Diagram View option



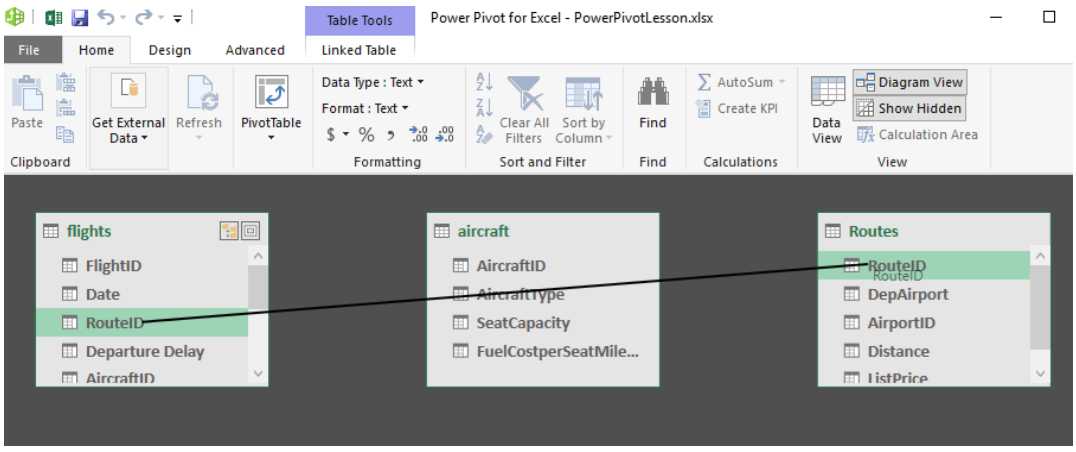
	FlightID	Date	RouteID	Departure Delay	AircraftID	Scheduled Departure	Avg Ticket Price	Total Fare	Flight Mont
1	Flight1	41972	ORD-BOS	3.9	43	30/12/1899 18:00:00	450	74250	
2	Flight2	41967	ORD-JFK	6.8	61	30/12/1899 19:00:00	375	61875	
3	Flight3	42238	ORD-EWR	2.7	72	30/12/1899 17:00:00	247.5	44797.5	
4	Flight4	41833	ORD-LAX	3.8	172	30/12/1899 17:30:00	385	63525	
5	Flight5	42135	ORD-IAH	3	89	30/12/1899 17:00:00	200	36200	
6	Flight6	42359	ORD-CLT	3.5	91	30/12/1899 11:30:00	175	32375	
7	Flight7	41726	ORD-DEN	1.7	138	30/12/1899 15:00:00	300	55500	
8	Flight8	42318	ORD-DFW	2.3	131	30/12/1899 18:00:00	200	37000	
9	Flight9	41774	ORD-ATL	10.6	23	30/12/1899 18:00:00	200	37000	
10	Flight10	41900	ORD-CLT	2	169	30/12/1899 07:00:00	125	20625	
11	Flight11	42219	ORD-MCO	2.6	197	30/12/1899 18:00:00	275	45375	
12	Flight12	41648	ORD-LAX	3.2	209	30/12/1899 12:30:00	350	57750	
13	Flight13	42234	ORD-SFO	1.6	97	30/12/1899 10:00:00	412.5	76312.5	
14	Flight14	41828	ORD-LAX	5.5	6	30/12/1899 17:00:00	385	71225	
15	Flight15	41676	ORD-LAX	4.1	116	30/12/1899 15:00:00	350	64750	
16	Flight16	41736	ORD-SEA	2.1	178	30/12/1899 15:30:00	350	57750	
17	Flight17	41716	ORD-MCO	2.3	154	30/12/1899 17:30:00	250	41250	
18	Flight18	42208	ORD-MSP	1	100	30/12/1899 18:00:00	126.5	23402.5	
19	Flight19	41996	ORD-DEN	4.1	7	30/12/1899 07:00:00	420	77700	
20	Flight20	41733	ORD-DEN	1.3	190	30/12/1899 19:00:00	300	49500	
21	Flight21	42119	ORD-DFW	1.5	63	30/12/1899 18:00:00	200	36200	

This enables us to see each of the tables we have saved within the data model. For each table we must identify the links to the other tables by dragging lines between the tables, in order to create relationships. In order for this linking to work the columns must have exactly the same name.



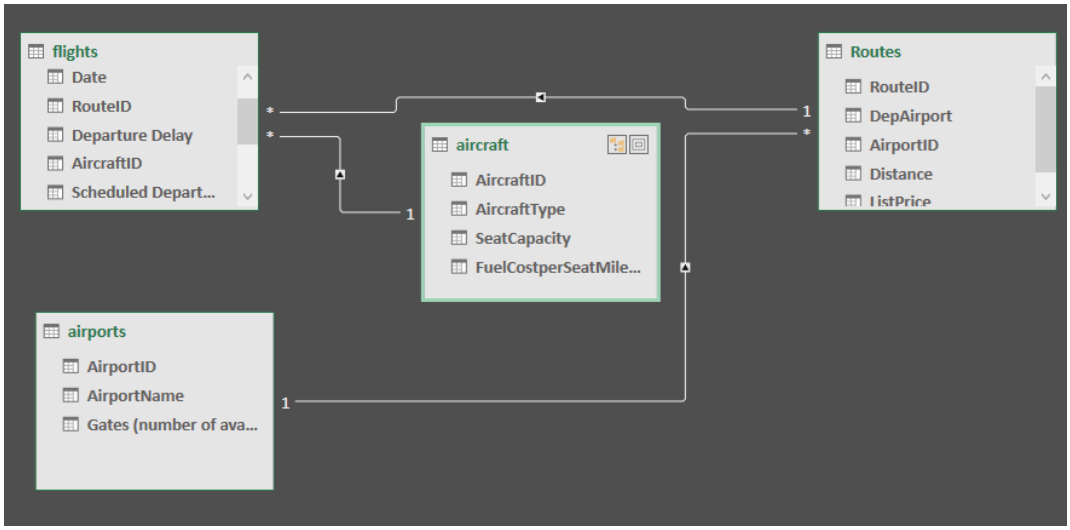
Route ID appears in the routes table and the flights table, this means it is a foreign key in the flights table and a primary key in the routes table.

- Click on the route ID in the flights table and drag it onto the route ID in the routes table.



This creates a link in the diagram showing that these fields are the same. aircraft ID also appears in two tables, create the link as before. Lastly link the two occurrences of airport ID.

Tip: always link from the foreign key side to the primary key side in a relationship.



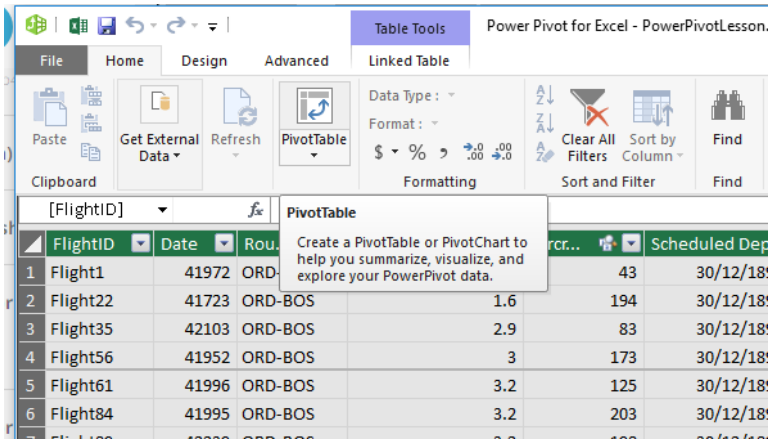
All tables in the diagram are now connected, no single table stands on its own. Close the PowerPivot view.

PowerPivot to Visualise Data

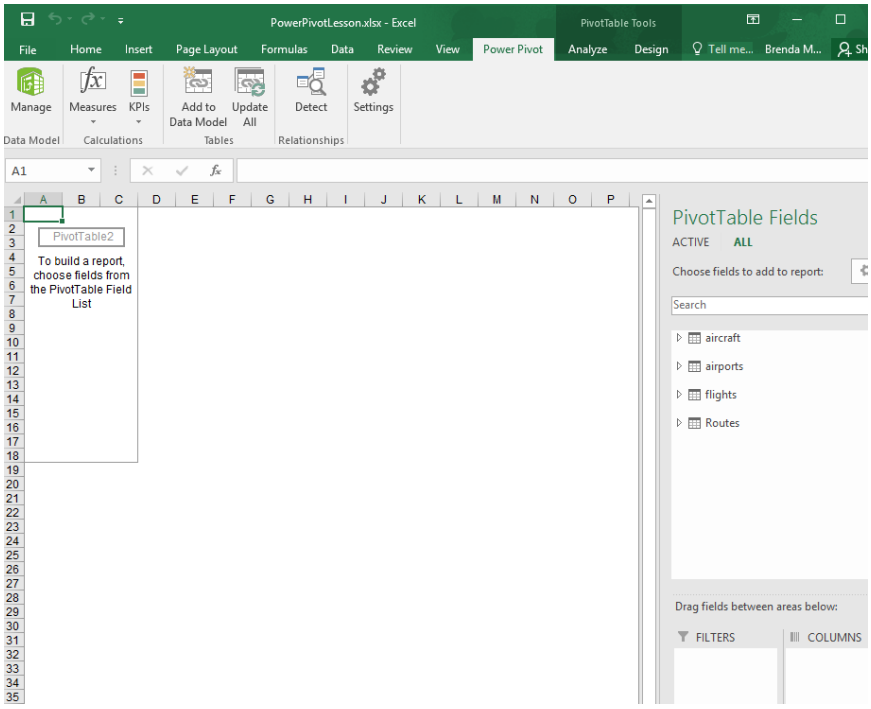
For this exercise you are asked to look into the specific metrics with regard to flights and the aircraft that are being used. We are going to use the data model that we created in the last step.

If we want to answer the qustion: How many flights use the A319 aircraft? We know that flight and aircraft data are stored in two different tabs of data. However we have used the aircraft ID to link those data sets together.

Go to a new tab name it Exercise 2. We are going to place our pivot table in this new tab so it can expand as needed. - Select the first cell in A1. - Next navigate to the PowerPivot tab on the ribbon - Click Manage - Click Pivot Table

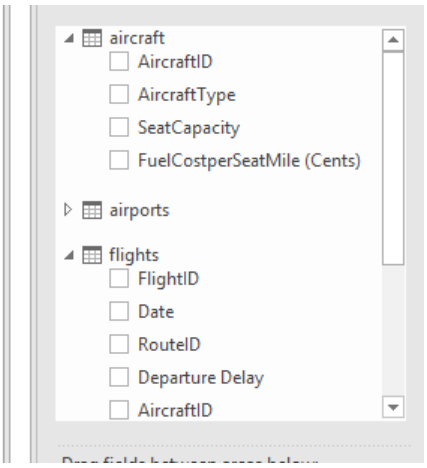


It prompts you to confirm that you want a pivot table created in the existing worksheet

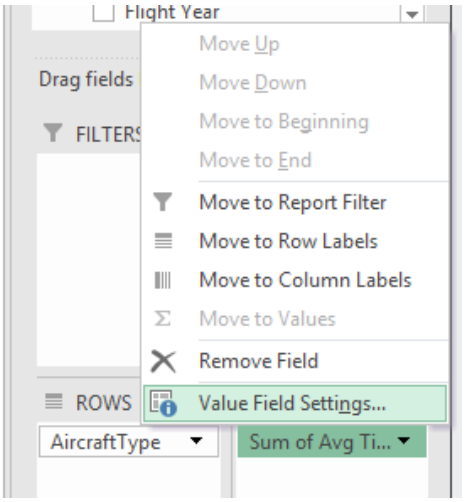


- All four tables in our data model appear on the right hand side.

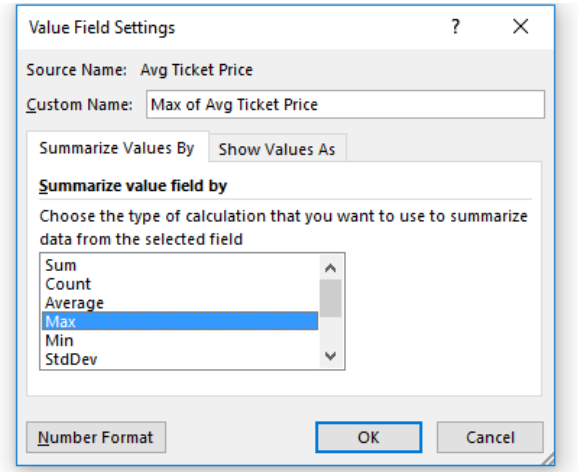
- To answer the question we must combine the aircraft and flight datasets. Expand those two datasets.



- Now we can see flight ID and aircraft type. Drag these two fields into the values and rows fields respectively.
- Now we see the pivot table form.
- It shows a count of the flightIDs which is what we want to show since we are answering the question how many flights use the A319 aircraft.
- Cell B2 shows us there were 3879 flights that used the A319 aircraft.
- The second question we wish to answer is: What is the most common type of aircraft used across all flights? What is the answer?
- The next question is: What is the maximum average ticket price for flights on the A320 aircraft?
- You can create a new tab if you wish to answer this question or re-work your existing power pivot.
- We need the aircraft type (rows) and ticket price (values)
- By default the average ticket price is summed so we want to change this
- Click the dropdown arrow for Sum of Avg Ticket Price in the Values section, choose Value Field Setting.



- Choose Max from the list of options to summarise by.

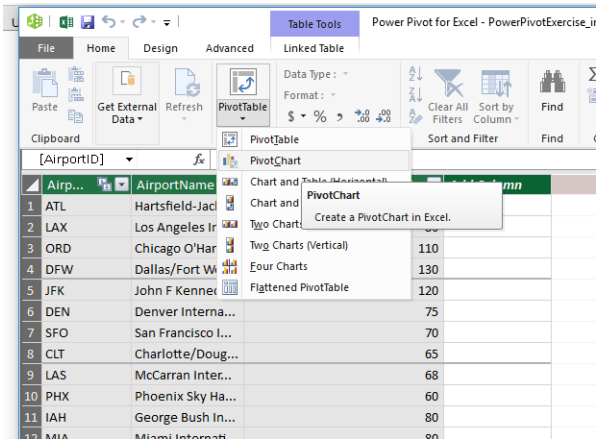


PivotCharts

Another feature of PowerPivot is PivotCharts.

This allows you to visualise data across multiple data sources. In this particular case we are being asked to visualise the number of flights flown for each aircraft type.

- Open a new tab name it PivotChart
- Click in cell D12 and then click on PowerPivot, then the Manage button and then the arrow below PivotTable.
- Choose PivotChart and use the existing workbook.



- You can adjust the height and width of the chart.
- Next select the fields for the x and y axis.
- Aircraft type should be placed in the Axis section.
- FlightID should be placed in the Values section.
- You can change the title by double clicking the text (change it to Flight ID vs Aircraft Type).
- The filter in the bottom left corner of the chart can be used to filter to only two types of aircrafts.