

Excel Power Query Introduction (Get & Transform)

By Neale Blackwood

www.a4accounting.com.au

www.excelyourself.com.au

a4@iinet.net.au



A4 ACCOUNTING
EXCEL CONSULTING, TRAINING AND WEBINARS

Contents

Introduction.....	3
Extracting from CSV Files	4
Another CSV File Example	10
Unpivot Data?.....	12
MYOB Report Layouts Conquered	14
Where to From Here.....	18

Introduction

Power Query first became available in Excel 2010 and Excel 2013 as an add-in. Power Query is included in Excel 2016 and later versions.

In Excel 2016 it is called Get & Transform and is in the Data ribbon tab. In previous versions it is called Power Query and has its own ribbon tab. I will use the term Power Query throughout the manual as it the more common term.

Power Query allows you to import data from various sources. During the importation process you can also perform data cleansing operations. Data cleansing involves amending the data so that Excel can work with it. For example you might need to amend the date format used in the data so that Excel can recognise the dates.

Power Query records the actions you take during the data cleansing operation and it will repeat them the next time the data is refreshed. Power Query acts like the macro recorder. It records your actions so that they can be replayed. You can also edit the actions that are recorded to fix any issues.

As an example you could have a CSV file that contains the dump of your year to date transactions. Power Query can import that CSV file into a data table and perform data cleansing. To update the data all you need to do is to paste a new CSV file over the top of the old CSV file and refresh the data. This will automatically update the data table in Excel with the new data.

Power Query also has the ability to combine multiple files into a single database. For example you could have 12 monthly CSV sales files that you combine into a single full year's sales data file. This will be covered in a future webinar.

In this session we will work with a couple of different csv file examples and see how we can create data tables from them. The common file formats used for data are .txt and .csv files as well as Excel files.

Built-in Features

Power Query has a lot of built-in advanced features to enable you to convert many different data structures into a table format. The problems with data have been around for decades and Power Query can handle the most common data issues.

Excel works well with table layouts. Many common data cleansing processes have been captured in single operations to make the data cleansing process easier in Power Query.

In the past data cleansing has been done with a combination of formulas and macros (VBA).

Formulas

Our last example will introduce Power Query formulas. These are similar to Excel cell-based formulas but have a number of differences that can make them hard to learn and understand. The example provided works for certain MYOB reports. I will not go into detail about these formulas, they will be included in a future webinar.

Extracting from CSV Files

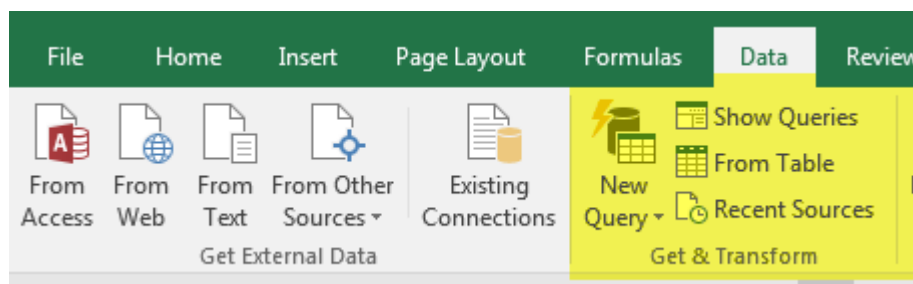
CSV stands for Comma Separated Values. CSV files are typically used to dump data out of one system so that it can be used in other systems. Most systems these days can download directly into Excel, but many systems also have the ability to download into CSV file. CSV files have no formatting and can hold huge amounts of data.

Excel works very well with CSV files. You can open a CSV file in Excel and Excel will read the data. The problem arises when the data has issues. For examples the date format might be wrong which requires you to amend the data once it is opened in Excel.

The idea behind Power Query is that you can import a CSV file and then apply data cleansing procedures to that file to create a final data table layout. Once you have created the Power Query import routine then it is as simple as replacing the CSV file with the latest version of the data (retain the name) and then opening your Excel file and clicking refresh to update the data table.

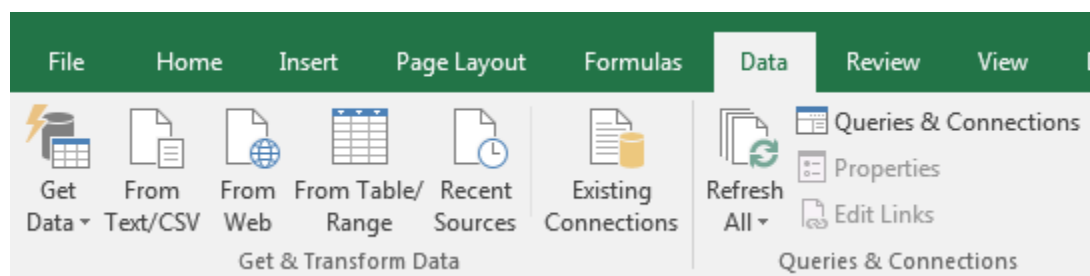
The data table can then be used by formulas; a pivot table or a PowerPivot report to summarise the data.

Power Query has its own tab in the ribbon in Excel 2010 and 2013 (see image on right). Power Query is part of the Data ribbon in Excel 2016 and called Get &



Transform. This session will focus on the latest Excel ribbon interface.

Depending on your version of Excel 2016 you may have the following Data Ribbon tab.



This is my current version and the images in the rest of the manual will use this ribbon.

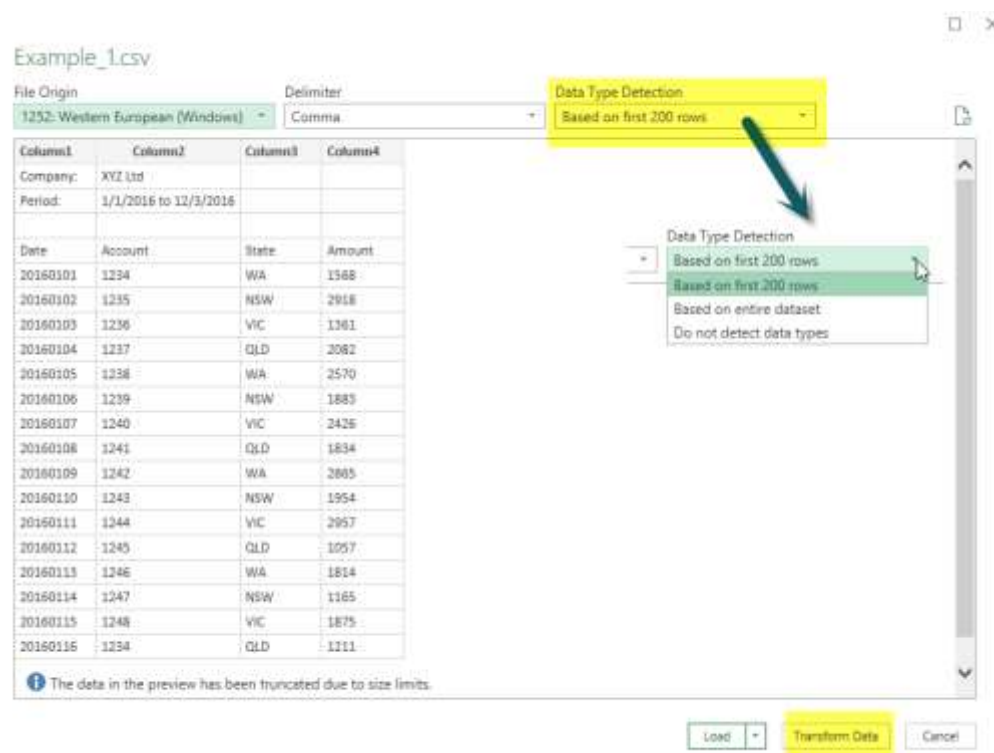
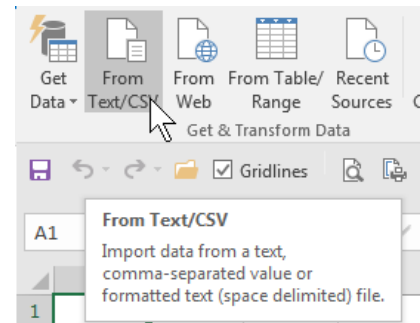
Be aware that Power Query is constantly being updated so screens, buttons and icons may change as updates occur.

Some workarounds may not be needed in the future as new features are added. Keep an eye out for changes in options and ribbons.

In the first example we will handle three data cleansing issues: dates; leading zeroes and spaces.

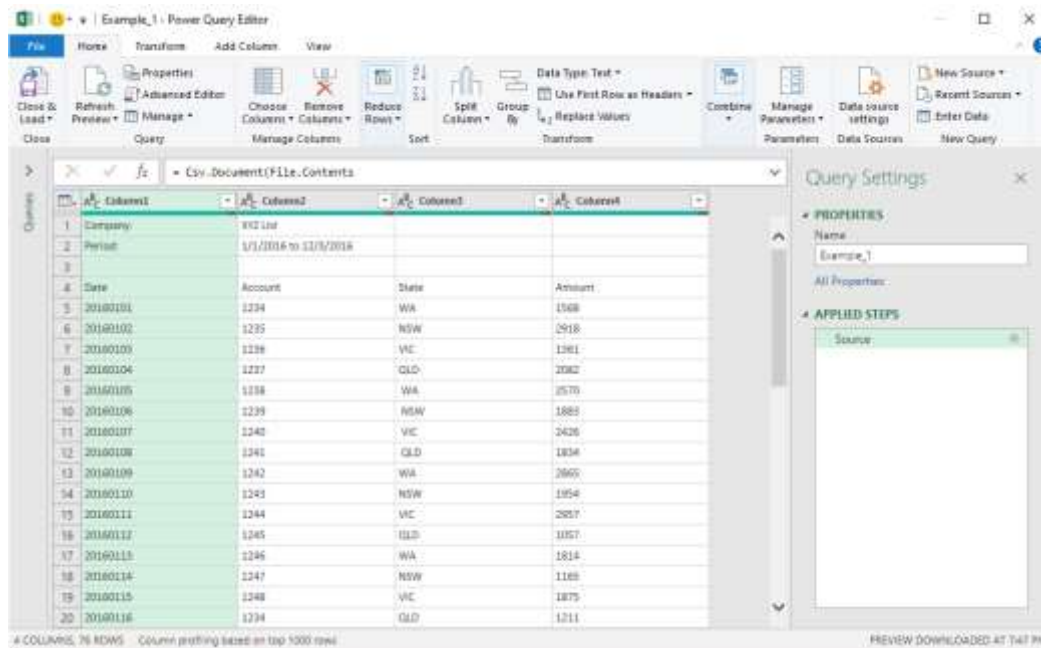
PRACTICE:

1. Click the Data ribbon tab and click the From Text/CSV icon.
2. Navigate to the PQ_Files folder and select Example_1.csv file and click Import.
3. A new dialog will open. Click the Transform Data button in the bottom, right corner. If the data in the file varies in layout you might want to use the Data Type Detection drop down and select Based on entire dataset.



4. The Power Query window opens – see next page. This is a separate window to Excel.

It displays the top section of the data.



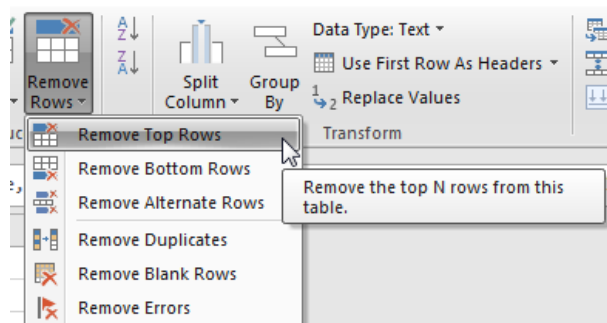
On the right of screen the Query name defaults to the file name.

You can click in the Name box and change it. All the icons/options mentioned in this example are on the Home ribbon.

The Applied Steps section lists the processes already performed and it will list, in sequence all the changes we are about to make to the layout.

5. The first three rows at the top of the data are not required.

We can click the Remove Rows icon drop down and specify Remove Top Rows. See image on right.



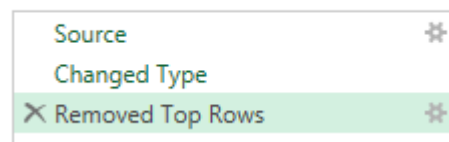
6. Enter the number of rows to remove. In this case type 3 and click OK.



7. In the Applied Steps section an extra step has been added.

The small gear icon on the right means you can adjust the setting to this step.

APPLIED STEPS



WARNING: THERE IS NO UNDO COMMAND IN THIS WINDOW. IF YOU MAKE A MISTAKE, DELETE THE STEP AND RE-DO IT. IF THERE IS A GEAR ICON ON THE RIGHT OF THE STEP YOU MAY BE ABLE TO ADJUST THE SETTINGS TO ACHIEVE WHAT YOU WANT.

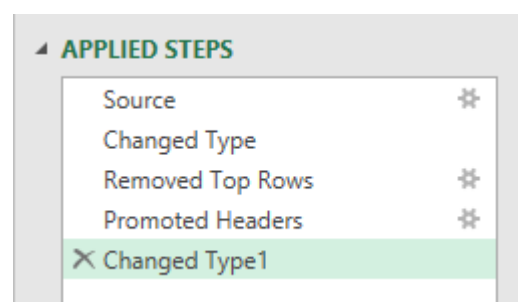
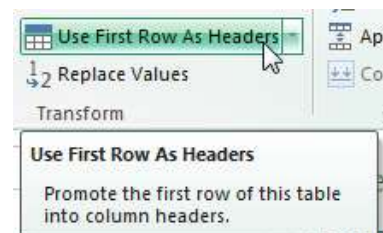
8. Now that the layout looks more like a table we can tell Power Query to use the first row as the column headings.

Click the Use First Row As Headers icon.

Another two steps will be added to the Applied Steps on the right of screen.

The first step is the Promoted Headers.

The second is another Changed Type which happens after you promote the headers.



9. Click the small icon on the left of the header row in the Date column. Choose Date.

A dialog below will display. It doesn't matter whether you click Replace current or Add new step. I will use Replace current. I will use this option on future dialogs as well. This relates the previous Change Type step that was automatically added.

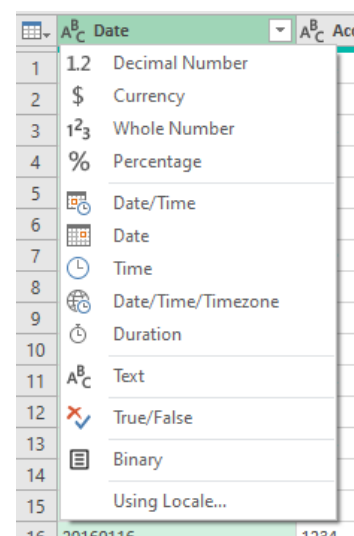
Change Column Type

The selected column has an existing type conversion. Would you like to replace the existing conversion, or preserve the existing conversion and add the new conversion as a separate step?

Replace current

Add new step

Cancel



Power Query will convert all the dates into the correct format.

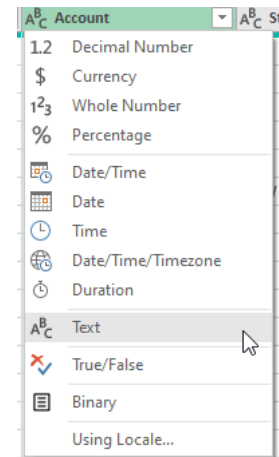
Note the icon in the header row of the column changes to a calendar icon.

10. Click the small icon on the left of the header row in the Account column and choose Text.

See image on right.

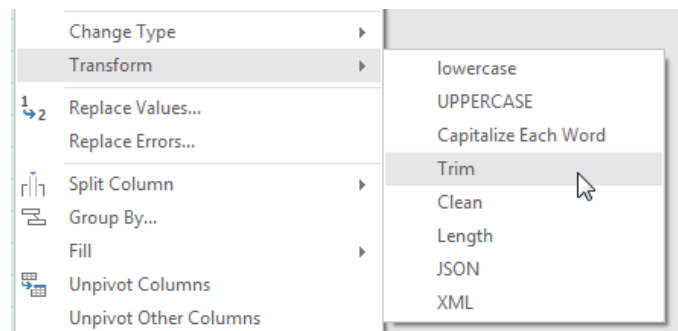
This will ensure the leading zeroes are maintained when the data is loaded into Excel.

When you open a CSV file in Excel leading zeroes are usually removed. Using Power Query solves that issue.



11. Right Click the State column and choose Transform and then click the Trim option.

This removes leading and trailing spaces from the entries in the column.



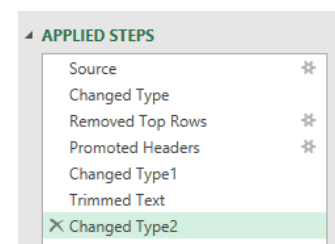
12. Click the small icon on the left of the header row in the Amount column and choose Currency.

Note these changes relate to data types, not formats. Sometimes the format will change, but it is defining a data type for the whole column. This may generate an error if an entry cannot be converted into a specific data type.

13. The final Applied Steps are shown on the right.

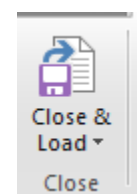
Note: sometimes when you change column types they are combined into one step.

All these steps will be performed in sequence each time the data is refreshed.



14. We are now ready to create an Excel table.

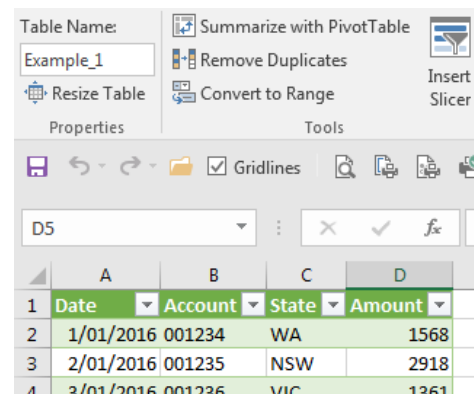
Click the Close & Load icon – top left of Power Query window.



15. This creates a formatted table in Excel in a new sheet.

The new sheet name will have the query name.

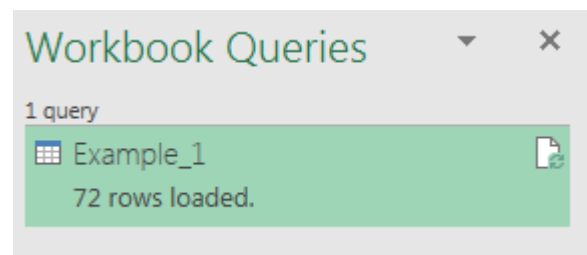
You can change the sheet name to a more descriptive name and it won't affect the Query.



	A	B	C	D
1	Date	Account	State	Amount
2	1/01/2016	001234	WA	1568
3	2/01/2016	001235	NSW	2918
4	3/01/2016	001236	VIC	1361

16. When the table is selected, the Query task pane is shown on the right. It has the Query name and how many rows have been loaded.

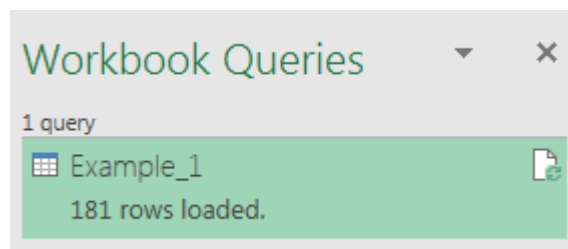
You can double click the Query name (Example_1 in this case) to view the Query.



17. If you save a new CSV file, which has extra rows at the bottom of the data, over the top of the existing file, the extra rows will automatically be included in the table when you refresh the Query.

The file Example_1_Extras has more rows than Example_1. I will use it to overwrite Example_1 and then refresh the Query.

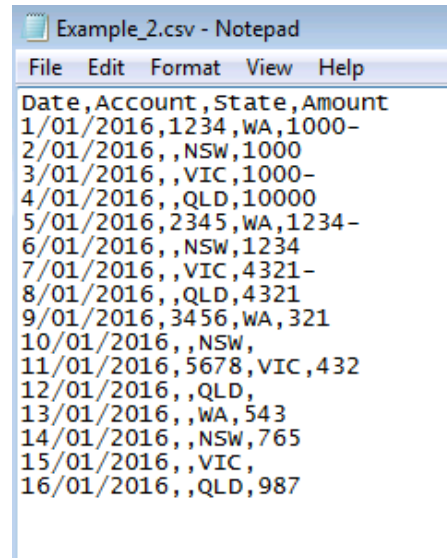
Extra rows at the top of the file would will affect the data cleansing steps we captured and the refresh would not work.



Another CSV File Example

Let's look at three more common data cleansing issues.

- Blanks cells in the value data columns that need to be populated with zeroes. This can save an annoying problem with pivot table reports and blank value entries
- Blank cells that need to be populated with the value from above
- Trailing minus sign for negatives. One of the large data bases still does this – see file image on right.



PRACTICE:

1. Use the From Text/CSV icon to open the Example_2.csv file.
2. The values in this CSV file had trailing minus signs – Power Query automatically fixed them – no intervention required.

Amount
-1000
1000
-1000

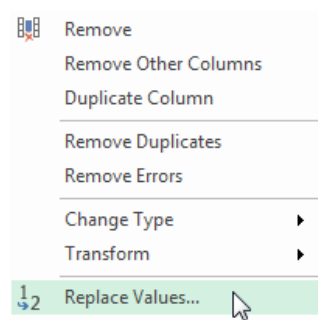
Use the Transform Data button.

3. **Note:** the first row has automatically been identified as the header row.
4. The data in the Amount column has null entries.

These need to be replaced with zeroes.

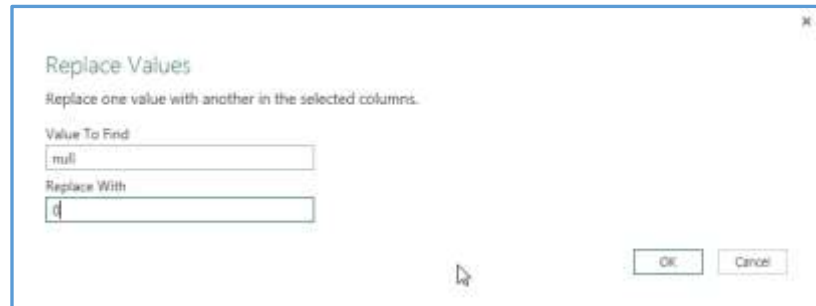
Right click the Amount column and choose Replace Values

	Date	Account	State	Amount
1	1/01/2016	1234	WA	-1000
2	2/01/2016	null	NSW	1000
3	3/01/2016	null	VIC	-1000
4	4/01/2016	null	QLD	10000
5	5/01/2016	2345	WA	-1234
6	6/01/2016	null	NSW	1234
7	7/01/2016	null	VIC	-4321
8	8/01/2016	null	QLD	4321
9	9/01/2016	3456	WA	321
10	10/01/2016	null	NSW	null
11	11/01/2016	5678	VIC	432
12	12/01/2016	null	QLD	null
13	13/01/2016	null	WA	543
14	14/01/2016	null	NSW	765
15	15/01/2016	null	VIC	null
16	16/01/2016	null	QLD	987



- Enter null in the first box and 0 in the second box and click OK.

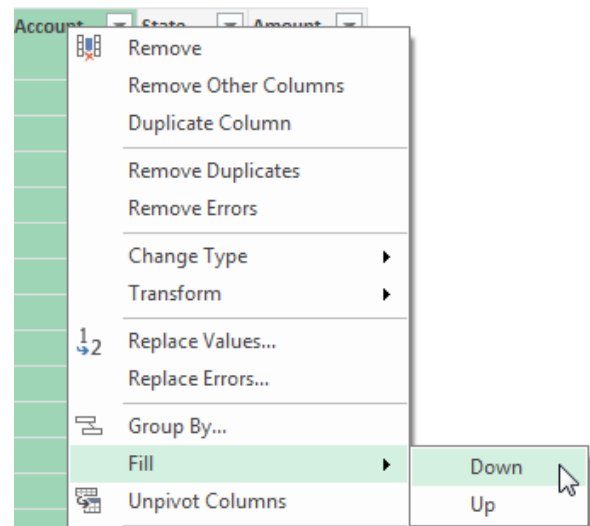
Power Query is case sensitive. Always use lowercase for the word null.



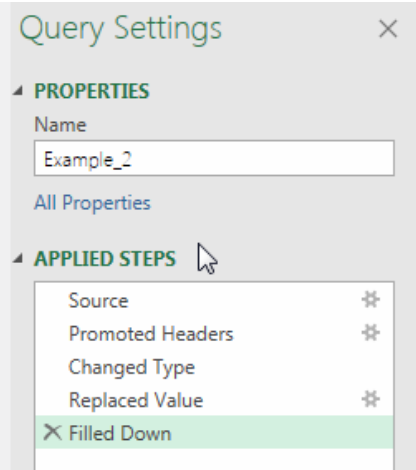
- The null entries in the Account column must be populated with the entries from above. Power Query has an option to do just that.

Right click the Account column and choose Fill, then click the Down option.

- The completed layout is shown below on the left with the Applied Steps on the right.



	Date	Account	State	Amount
1	1/01/2016	1234	WA	-1000
2	2/01/2016	1234	NSW	1000
3	3/01/2016	1234	VIC	-1000
4	4/01/2016	1234	QLD	10000
5	5/01/2016	2345	WA	-1234
6	6/01/2016	2345	NSW	1234
7	7/01/2016	2345	VIC	-4321
8	8/01/2016	2345	QLD	4321
9	9/01/2016	3456	WA	321
10	10/01/2016	3456	NSW	0
11	11/01/2016	5678	VIC	432
12	12/01/2016	5678	QLD	0
13	13/01/2016	5678	WA	543
14	14/01/2016	5678	NSW	765
15	15/01/2016	5678	VIC	0
16	16/01/2016	5678	QLD	987



The icon next to the headers represent the data type applied to that column automatically.

Unpivot Data?

The term Unpivot is used to describe converting data from a report layout into a table layout.

Many Excel users will design a report layout to capture their data because it gives them a report once the data is entered. They usually realise that it is hard to change the report, or to report on the data in a different layout. Creating a usable table from the report layout usually requires a lot of copying and pasting to get the right layout.

Table layouts provide the raw materials to create reports in many different layouts.

Power Query can make it easy to convert a report layout into a table layout.

Example

Let's say we have a retail shop that has three branches and it is monitoring client contacts.

They have created a data table (shown on the right) but they want more flexibility.

They need you to convert the report into a proper data table layout.

We will create a table that has columns for Branch, Type, Date and Value.

They want to keep using this table to input the data.

	A	B	C	D	E
1	Branch	Type	1/01/2016	2/01/2016	3/01/2016
2	Joondalup	Walk In	27	10	20
3		TV Ad	46	26	30
4		Radio Ad	12	24	21
5		Email	34	14	28
6		Website	42	14	48
7		Referral	43	39	37
8	Karrinyup	Walk In	40	13	40
9		TV Ad	11	47	31
10		Radio Ad	35	23	46
11		Email	29	50	43
12		Website	33	37	47
13		Referral	26	33	32
14	Innaloo	Walk In	20	27	20
15		TV Ad	10	16	25
16		Radio Ad	19	21	18
17		Email	21	46	10
18		Website	11	16	25
19		Referral	39	41	35

PRACTICE:

1. In the Report tab click inside the table. Click the Data ribbon tab and click the From Table/Range icon. Power Query should guess the correct range so click OK and this will go direct to the Power Query window.
2. We need to fix the Branch column so that the branch name is repeated in all rows. Right click the Branch column heading and select Fill and then Down to populate the other branch entries.
3. Use the Ctrl key to select both the Branch and Type columns. Right click on one of the two column headings and chose Unpivot Other Columns.
4. Right click the Attribute column and rename it to Date.
5. Change the data type of the Date column to a Date data type.

6. Job done! Click the Close & Load icon.
7. Any new data added to the right of the table in the Report sheet will get updated when you click the refresh for the Query.

Try inserting rows in the Report table and add the Other type to each branch with some values. Refresh the Query and see the results.

Below is an image of the final table created. And the Applied Steps are below it.

	Branch	Type	Date	Value
1	Joondalup	Walk In	1/01/2016	27
2	Joondalup	Walk In	2/01/2016	10
3	Joondalup	Walk In	3/01/2016	20
4	Joondalup	Walk In	4/01/2016	23
5	Joondalup	Walk In	5/01/2016	33
6	Joondalup	Walk In	6/01/2016	29
7	Joondalup	Walk In	7/01/2016	35
8	Joondalup	Walk In	8/01/2016	36
9	Joondalup	Walk In	9/01/2016	36
10	Joondalup	Walk In	10/01/2016	35
11	Joondalup	Walk In	11/01/2016	17
12	Joondalup	Walk In	12/01/2016	32
13	Joondalup	TV Ad	1/01/2016	46

APPLIED STEPS
Source
Changed Type
Filled Down
Unpivoted Other Columns
Renamed Columns
X Changed Type1

MYOB Report Layouts Conquered

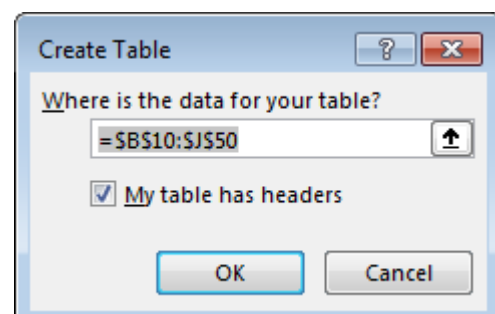
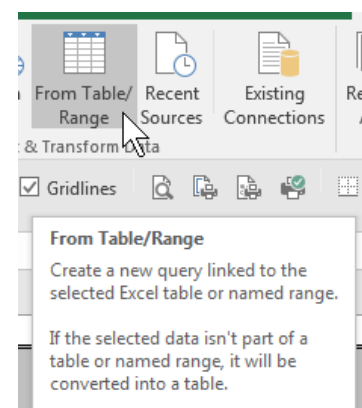
MYOB has been producing below average Excel reports since it first downloaded into Excel. When I am feeling cynical sometimes I think it is on purpose to make it harder to use the report in Excel. Power Query can convert the badly laid out MYOB reports into a data layout. Let's see how. A MYOB report layout shown below.

ABC Pty Limited									
General Ledger [Detail]									
1/01/2014 To 31/05/2014									
ID#	Src	Date	Memor	Debit	Credit	Job No.	Net Activity	Ending Balance	
5-1100	Audit Fee								
Beginning Balance:	\$226,099.65								
GJ002742		GJ 31/05/2014	Prepayment adjustment, Audit accrue	\$8,750.00				\$234,849.65	
GJ002742		GJ 31/05/2014	Prepayment adjustment, Audit accrue	\$3,812.50				\$238,662.15	
			Total:	\$12,562.50	\$0.00		\$12,562.50	\$238,662.15	
5-1200	Bank Charges								
Beginning Balance:	\$92,312.96								
134		CD 10/05/2014	Contractor pre - 12/04/14 - 25/04/14	\$122.42		* 0009		\$92,435.40	
130		CD 10/05/2014	Contractors pre - 12/04/14 - 25/04/14	\$895.91		* 0009		\$93,321.81	
RC210014		CD 10/05/2014	Account fee	\$10.00				\$93,331.81	
SJ000711		SJ 20/05/2014	Finance Charge: XYZ Inc.	\$20.00				\$93,351.81	
SJ000711		SJ 20/05/2014	Finance Charge: XYZ Inc.	\$1.53				\$93,352.54	
CR004818		CR 31/05/2014	Cancelled pre. alic@yahoo.com - 09/05/2014		\$0.00	* 0009		\$98,073.35	
			Total:	\$4,350.84	\$0.27		\$4,350.37	\$98,073.35	
5-1300	Depreciation Expense								
Beginning Balance:	\$227,856.96								
GJ002733		GJ 31/05/2014	Depreciation - May 2014	\$12,028.23				\$240,485.19	
GJ002745		GJ 31/05/2014	Lease Incentive	\$1,383.50				\$241,868.69	
			Total:	\$13,991.73	\$0.00		\$13,991.73	\$241,868.69	

PRACTICE:

- Click inside the MYOB report area, below row 10.
- In the Data ribbon tab click the From Table/Range icon.
- Excel should correctly guess the table range.

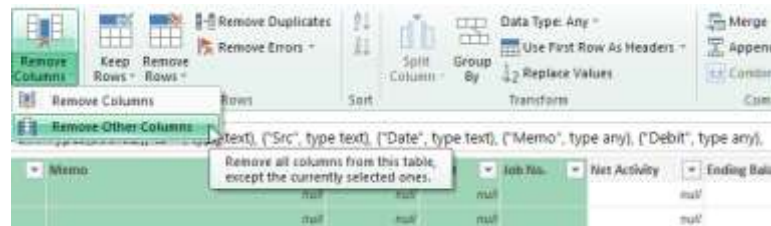
Click OK.



4. We don't require the last two columns.

You have a couple of options.

5. You can select the columns to remove and right click them and choose Remove.

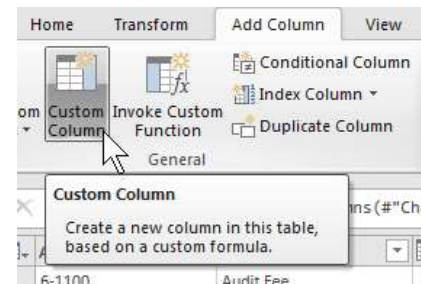


Or select the columns you want to keep in our case

#ID across to Job No. and click the Remove Other Columns option – this removes all columns to the right of the selected columns.

WARNING: Advanced alert – the following technique includes a reasonably advanced feature in the Power Query window involving a specialised formula.

6. Click the Add Column ribbon tab. Click the Custom Column icon.
7. This opens the dialog where you can add formulas.



Change the New column name to Account.

These formulas are not like Excel's normal formulas and there is a steep learning curve involved in using them.

If you can master Excel's formulas these require a few adjustments, but they too can be mastered.



The formula we need to enter is

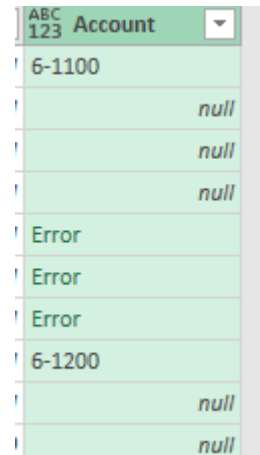
= if Text.Range([#"ID#"],1,1)="-" then[#"ID#"] else null

You have to type in everything except the square brackets and what's between the square brackets. When you get to the square bracket use the Insert button on the right after selecting the ID# column above. This inserts the field name into the formula including the brackets.

The above formula looks for the minus sign as the second character in the entry in the ID# column. When found it includes the entry from the ID# column in the new Account column. Otherwise it displays null.

Note: the Text.Range in the formula mimics the Excel MID function, but it has an important difference in the way it works. Instead of starting the character positions at 1, Power Query formulas treat the first character as zero. Hence, you have to be careful when creating formulas referring to character positions.

The resulting column entries are shown on the right.

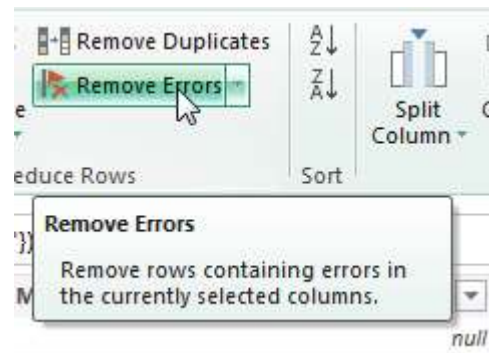


Account
6-1100
null
null
null
Error
Error
Error
6-1200
null
null

8. Right click the Account column and choose Fill and the Down. This will populate the Account numbers in the rows below.

9. Now we need to clean up the data table by removing those rows that we don't need.

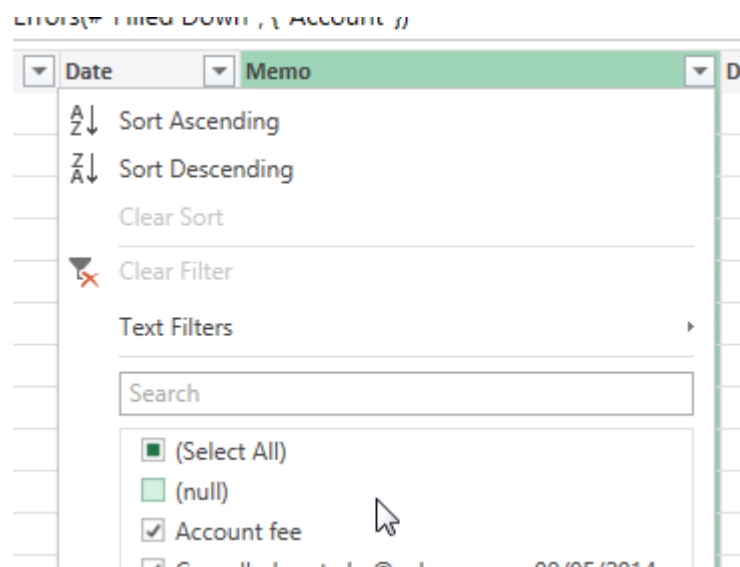
Select the Account column and the open the Home ribbon tab and click the Remove Errors icon.



10. Click on the Filter icon for the Memo column.

Uncheck the null entry and click OK.

Applying a Filter to a column allows you to control what is imported into the final data table.



11. Right click the Date column and use Change Type, then Date. We can ignore the error entries it creates because they are on rows we will not use in our data. This can be a useful technique to eliminate non-date rows which need to be filtered out.

12. Lastly in the Debit and Credit columns you might want to replace the nulls with zeros. Select both columns and click the Replace Values icon. Enter null in the first box and 0 in the second box and click OK – see image on following page.

Replace Values

Replace one value with another in the selected columns.

Value To Find
null

Replace With
0

OK Cancel

13. Finished - you just need to click the Close & Load icon to create a table based on the MYOB report. The final report is shown below. The Applied Steps are shown below that.

	A	B	C	D	E	F	G	H
1	ID#	Src	Date	Memo	Debit	Credit	Job No.	Account
2	GJ002742	GJ	31/05/2014	Prepayment adjustment, Audit accrue	8750			6-1100
3	GJ002742	GJ	31/05/2014	Prepayment adjustment, Audit accrue	3812.5			6-1100
4	134	CD	1/05/2014	Contractor pmt - 12/04/14 - 25/04/14	122.42		9000	6-1200
5	2130	CD	1/05/2014	Contractors pmt - 12/04/14 - 25/04/14	885.61		9000	6-1200
6	SC310514	CD	1/05/2014	Account fee	10			6-1200
7	SJ000711	SJ	2/05/2014	Finance Charge: XYZ Inc.	20			6-1200
8	SJ000711	SJ	2/05/2014	Finance Charge: XYZ Inc.	1.53			6-1200
9	CR004816	CR	31/05/2014	Cancelled pmt abc@yahoo.com - 09/05/2014		0.06	9000	6-1200
10	GJ002733	GJ	31/05/2014	Depreciation - May 2014	12628.23			6-1300
11	GJ002745	GJ	31/05/2014	Lease Incentive	1363.5			6-1300
12	GJ002734	GJ	31/05/2014	Intangible Asset Amort - May 2014	1492.35			6-1350
13	GJ002734	GJ	31/05/2014	Intangible Asset Amort - May 2014	105.73			6-1350
14	GJ002742	GJ	31/05/2014	Prepayment adjustment, Audit accrue	8750			6-1100
15	GJ002742	GJ	31/05/2014	Prepayment adjustment, Audit accrue	3812.5			6-1100

APPLIED STEPS

Source	
Changed Type	
Removed Other Columns	*
Added Custom	*
Filled Down	
Removed Errors	
Filtered Rows	*
✕ Changed Type1	

Where to From Here

Hopefully I have shown you enough to make you want to learn more.

The books listed below are the go to books for Power Query.

The first one has a new edition and name change out later this year.

Power Query is an important skill to have and it can also be applied to Power BI. Getting you data right is the first step to reporting on it.

Don't worry about making mistakes, that's how you learn. Remember to always build in validation checks so you can identify if you have any issues.

Power Query Books

M is for (DATA) MONKEY by Ken Puls and Miguel Escobar – don't be put off by the title, this is a great resource and has helped me learn more about Power Query.

Collect, Combine, and Transform Data Using Power Query in Excel and Power BI by Gil Raviv - this was published more recently than the above book and has more up to date information.