

CMBI 2018 ©

BUSINESS INTELLIGENCE SERVICES



POWER BI SAMPLES

# AUSTRALIAN CENSUS POWER BI SAMPLE

NOTES



**BUSINESS INTELLIGENCE CONSULTANCY**

## Contents

1.1.	Power Pivot Census Sample Overview .....	3
1.2.	Download the Power Pivot Census Sample .....	3
1.3.	Power Pivot Sample Data Model .....	3
1.4.	About the ABS Data Packs .....	4
1.5.	Integrating the data pack files with Power Query .....	5
1.6.	Query Dependencies .....	5
1.7.	Power Query functions.....	6
1.8.	About the sample census reports .....	7
1.8.1.	Suburb Compare report .....	7
1.8.2.	Single Suburb report .....	8
1.9.	Conditional Formatting in Pivot Tables.....	9
1.10.	Importing the Excel Power Pivot model into Power BI .....	11
1.11.	Power BI Census Model on the web.....	12
1.12.	Download links .....	13
1.12.1.	Links to fully processed models .....	13
1.12.2.	Links to component files for reprocessing the data models .....	13
1.13.	Understanding the Power Pivot SHELL file .....	14
1.13.1.	Download data pack files .....	14
1.13.2.	Set the pLoadFiles parameter.....	14
1.13.3.	Set the other model parameters .....	15
1.13.4.	Process the Census Pivot query .....	15

## 1.1. Power Pivot Census Sample Overview

This CMBI Census Power Pivot sample model uses Power Query and Power Pivot to integrate two complete ABS Census data packs from the 2011 and 2016. The integration makes it much easier to analyse a wide range of census data across and within different suburbs.

	A	B	C	D	E	F	G	H	I	J	K
1											
2	State	NSW									
3	by Census Category	All									
4	Census Metric	Total Number of bedrooms	One bedroom								
5	Check the data for 2016 at the ABS website <a href="http://www.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/SSC12001">http://www.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/SSC12001</a>										
	Census Metric	# Data Points	# Census 2016	# Census 2011	% Population 2016	% Population 2011	% Population Change 2011 to 2016	# Suburb Rank 2016	# Suburb Rank 2011	# Rank Change 2011 to 2016	# Population 2016
7	Surry Hills	2	2,734	2,512	16.66%	16.37%	0.29%	1	1	-	16,412
8	Sydney	2	2,097	1,946	12.16%	13.60%	-1.45%	2	2	-	17,252
9	Darlinghurst	2	2,061	1,924	18.21%	19.13%	-0.92%	3	3	-	11,320
10	Camperdown	2	1,883	1,195	18.21%	15.19%	3.02%	4	11	(7)	10,341
11	Potts Point	2	1,808	1,557	19.19%	22.69%	-3.50%	5	5	-	9,423
12	Waterloo	2	1,652	1,111	11.30%	10.43%	0.88%	6	16	(10)	14,616
13	Randwick	2	1,648	1,483	5.50%	5.35%	0.15%	7	6	1	29,986

We can use the Power Pivot model to analyse 1000s of statistical data points across all the suburbs in Australia. For each data point and suburb, we can find the movement between the two census dates, calculate the data point percentage of the suburb's population, and find the suburb rank and rank movement.

In total, the Power Pivot model contains over 40 million data points.

## 1.2. Download the Power Pivot Census Sample

We designed the model as an engaging Power BI training aid. We have made all the sample files available for download. Please feel free to download, extend, and distribute them.

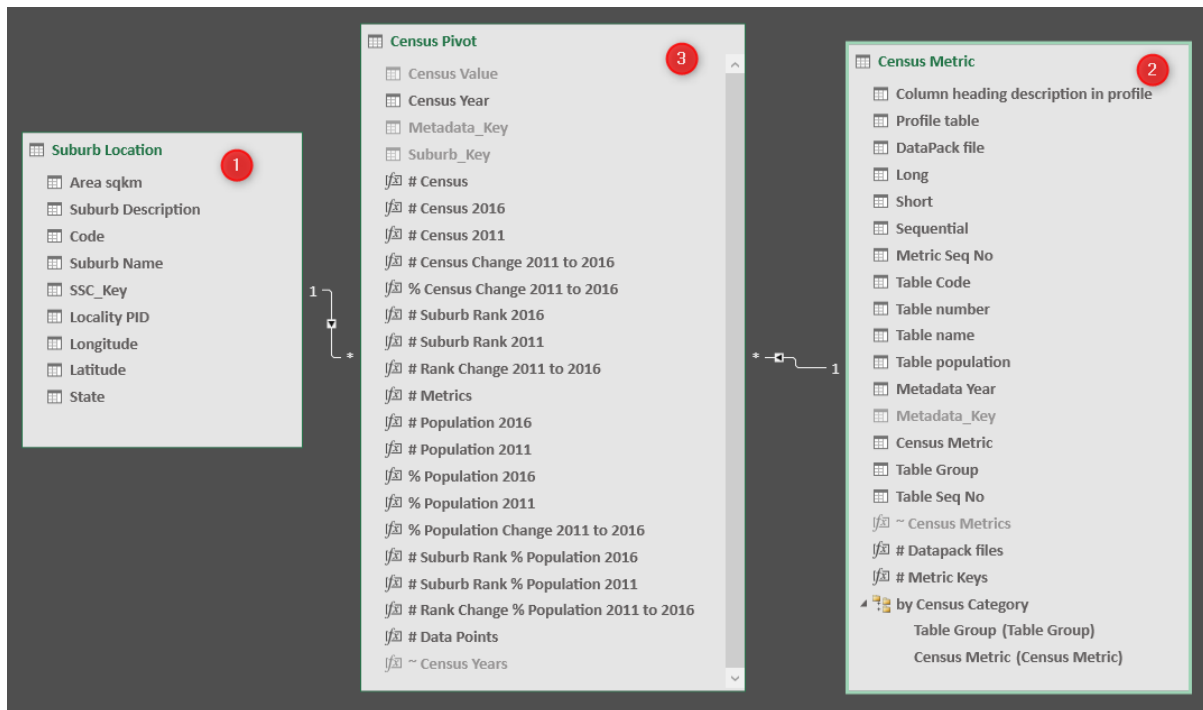
When using the samples we kindly make the following requests

1. Continue to credit the ABS in line with their **terms of use** for census data
2. Credit CMBI and/or continue to include the About page as a visible worksheet
3. CMBI provides the sample AS IS for the purposes of demonstrating Power BI capabilities and provides no warranties explicit or implied for use of the resulting data

The download links are available at the bottom of this page.

## 1.3. Power Pivot Sample Data Model

The Power Pivot model integrates data from over 100 ABS Census files into three tables



1. **Suburb Location** contains the list of Australian suburbs, their State, and their latitude and longitude so that we can visualise statistics on the map.
2. **Census Metric** contains a list of all the statistical census metrics in the census Data Packs. For instance, *Persons over 100 years old* is a census metric.
3. **Census Pivot** contains the 40+ million data points each of which references a Census Metric and a Suburb

## 1.4. About the ABS Data Packs

The ABS data packs are a great resource but are quite complex to understand and integrate. Each data packs comes as a collection of around 60 flattened CSV files. Each of these files contains codes and information about these codes is in various metadata and location data files, also included in the ABS data pack.

We used two suburb level data packs to build this model. You can download the data packs from

<https://datapacks.censusdata.abs.gov.au/datapacks/>

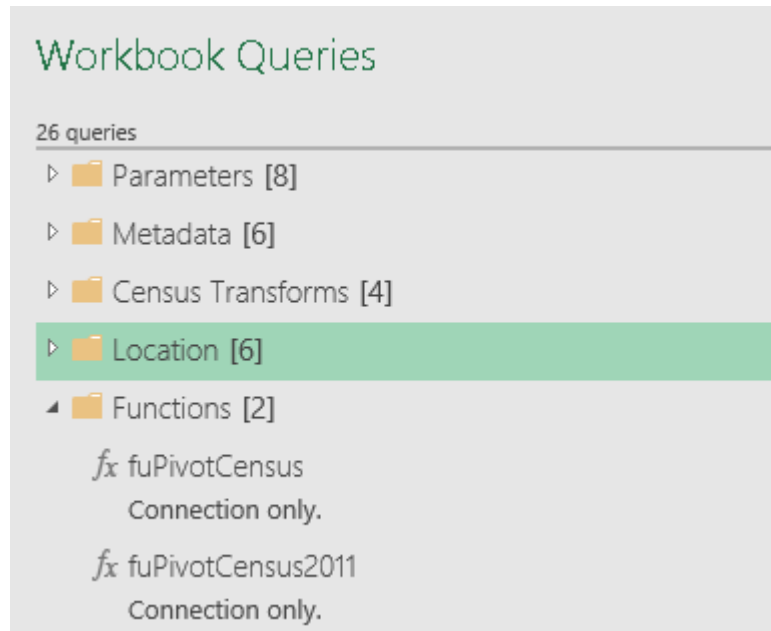
- 2016\_GCP\_SSC\_for\_AUS\_short-header.zip
- 2011\_BCP\_SSC\_for\_AUST\_short-header.zip

We chose suburb level data because it is the most familiar geographic sub division for most people. Postcodes level statistics are also available from the ABS but these are less granular. Other statistical divisions like LGA (Local Government Area) can be volatile and are less familiar to most people.

## 1.5. Integrating the data pack files with Power Query

The Power Pivot model integrates data from over 130 separate files. These files are a combination of metadata, reference data, and the statistical data pack files themselves.

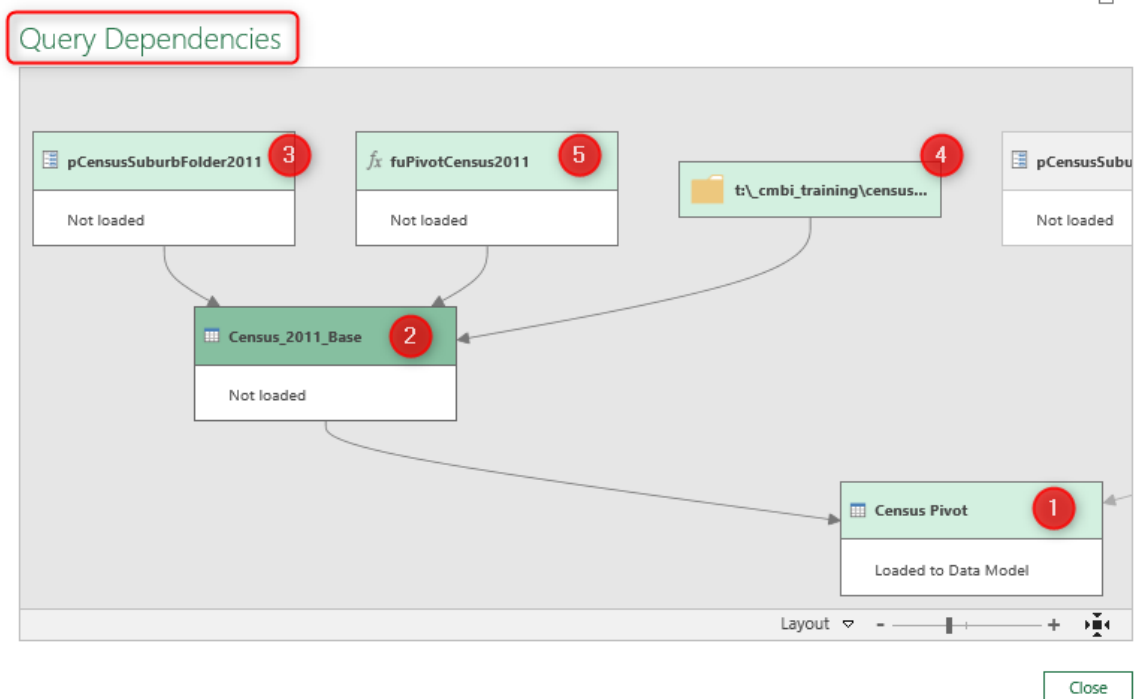
We have categorised the Power Queries in the workbook using folders.



- **Parameters** folder contains the Power Queries parameters that hold the location of the census files we integrate. (if you want to rerun the queries you will need to adjust these parameters to the location where you have placed the census files)
- **Metadata** folder contains the queries that extract the Census metadata for the two Census years. This metadata is the list of statistical data points and categories.
- **Census Transform** contains the queries that integrate the 128 Census Data Pack csv files
- **Functions** contains two similar Power Query functions that take a file name as a parameter and perform a number of transformations on the files to make it ready for integration into the Power Pivot model

## 1.6. Query Dependencies

A useful feature of Power Query is the ability to see Query Dependencies.



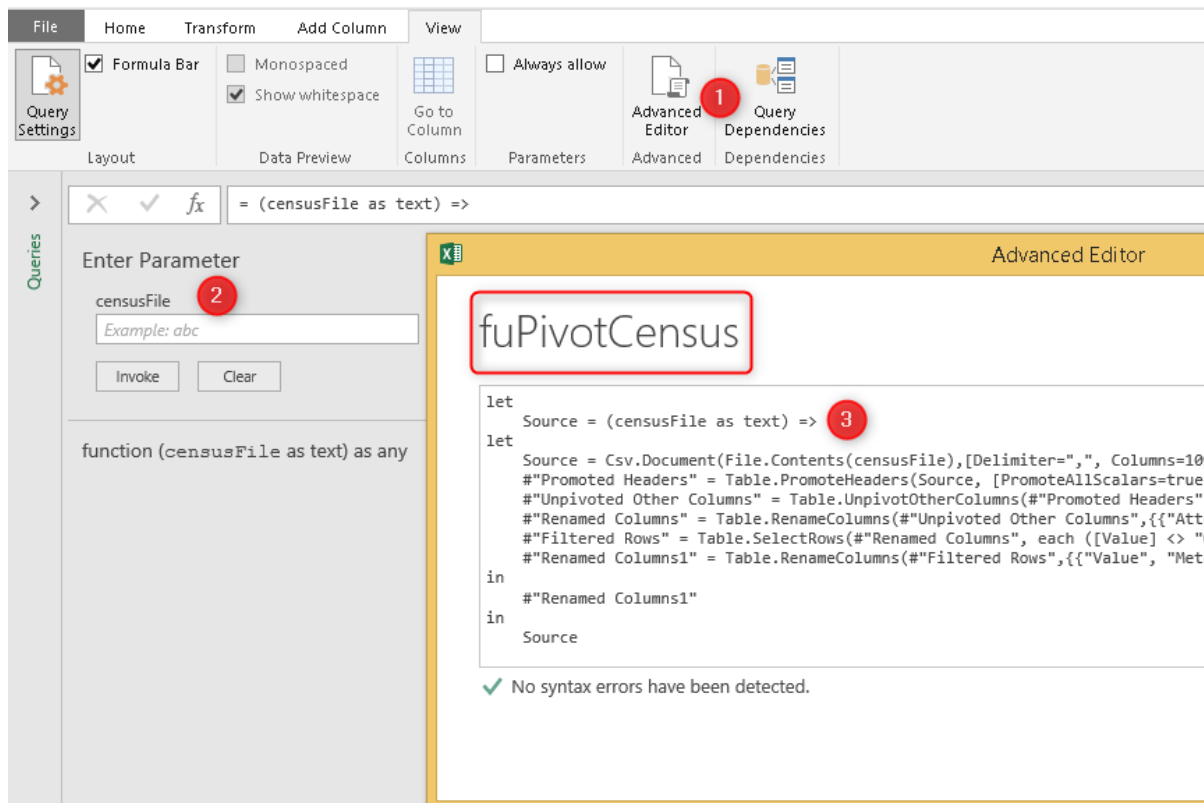
The figure above illustrates the dependencies between the final Power Pivot Table: Census Pivot, and the underlying Power Queries and data sources.

1. **Census Pivot** is the final table loaded to Power Pivot
2. Census Pivot is the combination of data from the **Census\_2011\_Base** and the **Census\_2016\_Base** queries.
3. The **Census\_2011\_Base** relies on the parameter **pCensusSuburbFolder2011** to locate the data pack files.
4. The **pCensusSuburbFolder2011** points to **this file location**
5. The **Census\_2011\_Base** also relies on the **fuPivotCensus2011** function which applies a set of transformations to each file in the data pack directory

## 1.7. Power Query functions

One very useful feature of Power Query is the ability to convert a set of transformation into a function and reuse them many times. In our Census model, we have many census data pack files and we want to apply the same set of transformations to each of these files.

The easiest way to create a Power Query function is to go through the normal process of connecting to some data and doing transformations using the regular Power Query window. Having done this, we can then parameterise the query and make it generic. It is a very similar process to recording a macro in Excel and then editing it using VBA Editor.



To create a Power Query function from an existing query

1. Duplicate the original query and name it appropriately then go to **Advanced Editor**
2. Type an additional couple of lines at the top of the query: **let Source = (parameter as text) =>**
3. After this additional code follows the original Power Query and substitute areas we want to parameterise with the parameter name we created at the top of the query

There are plenty of blogs on creating Power Query functions. They are definitely worth exploring.

## 1.8. About the sample census reports

We have created two sample reports to illustrate the breadth and depth of the data and to demonstrate some useful Power Pivot and Pivot Table functionality

### 1.8.1. Suburb Compare report

The **Suburb Compare** Excel worksheet contains a list of suburbs. We have chosen to filter this Pivot Table for NSW suburbs. The Census Metric in the Filter section of the Pivot Table contains a long list of Census Metrics. You can search using the text search box or simply browse using the scrollbars.

	A	B	C	D	E	F	G	H	I	J	K
1	State	NSW									
2	by Census Category	All									
3	Census Metric	Total Number of bedrooms One bedroom									
4	Check the data for 2016 at the ABS website <a href="http://www.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/SSCI2001">http://www.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/SSCI2001</a>										
5	Census Metric	# Data Points	# Census 2016	# Census 2011	% Population 2016	% Population 2011	% Population Change 2011 to 2016	# Suburb Rank 2016	# Suburb Rank 2011	# Rank Change 2011 to 2016	# Population 2016
6	Surry Hills	2	2,734	2,512	16.66%	16.37%	0.29%	1	1	-	16,412
7	Sydney	2	2,097	1,946	12.16%	13.60%	-1.45%	2	2	-	17,252
8	Darlinghurst	2	2,061	1,924	18.21%	19.13%	-0.92%	3	3	-	11,320
9	Camperdown	2	1,883	1,195	18.21%	15.19%	3.02%	4	11	(7)	10,341
10	Potts Point	2	1,808	1,557	19.19%	22.69%	-3.50%	5	5	-	9,423
11	Waterloo	2	1,652	1,111	11.30%	10.43%	0.88%	6	16	(10)	14,616
12	Randwick	2	1,648	1,483	5.50%	5.35%	0.15%	7	6	1	29,986

We have applied various settings to this Pivot Table to enhance usability

- Suburbs rows filtered to suburbs with two # Data Points. This means they have data for 2011 and 2016 and which makes them more interesting
- Sorted by # Census 2016 to see the most significant suburb contributors at the top of the list
- Conditional formatting on % of Population Change 2011 to 2016 (see metric explanation below)
- Ranks of each Suburb and change in rank between the two census dates

### 1.8.2. Single Suburb report

Whereas the Suburb Compare report compares many suburbs against a single metric, the **Single Suburb** report compares many metrics for one suburb.

	A	B	C	D	E	F	G	H	I	J
1	State	NSW								
2	Suburb Name	Jannali								
3	by Census Category	All								
4	Check the data for 2016 at the ABS website <a href="http://www.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/SSCI2001">http://www.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/SSCI2001</a>									
5	Census Metric	# Data Points	# Census 2016	# Census 2011	% Population 2016	% Population 2011	% Population Change 2011 to 2016	# Suburb Rank 2016	# Suburb Rank 2011	# Rank Change 2011 to 2016
6	Total Persons Persons	2	6,184	5,955	100.00%	100.00%	0.00%	370	360	10
7	Counted at home on Census Night Total	2	5,963	5,779	96.43%	97.04%	-0.62%	363	353	10
8	Counted on Census Night At home Persons	2	5,963	5,780	96.43%	97.06%	-0.64%	363	353	10
9	Total private dwellings Persons	2	5,790	5,571	93.63%	93.55%	0.08%	365	355	10

The settings we have applied to this Pivot Table are

- Census Metrics rows with two # Data Points. Again, this means they have data for 2011 and 2016 which makes them more interesting
- Sorted by # Census 2016 to see the most significant metrics as a % of the suburb population at the top of the list
- Conditional formatting on % of Population Change 2011 to 2016
- Ranks of each Suburb and change in rank between the two census dates

### Suburb name uniqueness

A suburb name is not necessarily unique within Australia or even within a State. The most unambiguous column in the model is Suburb Description. However, sometimes the same suburb has slightly different labels between the two census periods. The # Data Points is the best indicator that the rows contains multiple physical suburbs. If it is greater than 2 then analyse the data using the Suburb Description.



## 1.9. Conditional Formatting in Pivot Tables

One of the nice features of Pivot Tables is that we can associate conditional formatting with the Pivot Table value rather than a fixed range. This means that if the Pivot Table number of rows change – when we change a filter value for instance – the conditional formatting remains intact for all the Pivot Table rows.

**Edit Formatting Rule**

Apply Rule To:

☐ Selected cells

☒ All cells showing "% Population Change 2011 to 2016" values

☐ All cells showing "% Population Change 2011 to 2016" values for "Census Metric"

Select a Rule Type:

- Format all cells based on their values
- Format only cells that contain
- Format only top or bottom ranked values
- Format only values that are above or below average
- Use a formula to determine which cells to format

Edit the Rule Description:

**Format all cells based on their values:**

Format Style:

Icon Style:  ☐ Show Icon Only

Display each icon according to these rules:

Icon	Value	Type
	>= 0.05	Number
	>= 0.01	Number
	>= -0.01	Number
	>= -0.05	Number

OK Cancel

1. Associate the rule with all the cells in the Pivot Table for a specific Pivot Table value
2. You can edit the number range for each of the arrows. In our two sample reports we have used slightly different ranges for the conditional formatting to highlight significant variances

## 1.10. DAX Calculations

We have added a number of DAX calculations to the model which highlight the significance of and change in census metrics. These calculations also highlight some of the key features of the DAX calculation language.

In the sections below, we have included a selection of the calculations in the model that broadly represent the features used.

### 1.10.1. # Census:=SUM('Census Pivot'[Census Value])

**# Census** measure aggregates the Census Values for a given selection in the Pivot Table. Not all aggregations of Census Value will be useful calculations. For instance, the addition of Total Females metric and Total Persons metric would not give a sensible value. However, the additional of Total Females 0 – 4 and Total Females 10 – 14 would.

In short, we should use this measure with caution and ensure we know what underlying values we are aggregating. That is one reason why we include the # Data Points calculation in the Pivot Tables, so we can see how many discrete Census Values we are aggregating.

### 1.10.2. # Census 2016:=CALCULATE([# Census],'Census Pivot'[Census Year]=2016)

The **# Census 2016** builds on the # Census measure described in the previous section. The Census Pivot table contains some rows for Census 2011 and some rows for Census 2016. The # Census 2016 aggregates only rows relating to the 2016 Census. There is a similar measure for 2011. We use these two measures in further model calculations that find the difference in value between the two census dates.

### 1.10.3. # Suburb Rank 2016:=RANKX(ALL('Suburb Location'[Suburb Name]),[# Census 2016])

The **# Suburb Rank 2016** calculates the rank of the suburb for 2016 Census Values. So for instance, if we selected the Total Persons metric and then selected State=NSW, and Suburb Name=Castle Hill then we would get # Suburb Rank 2016 of 3. This is indicating that Castle Hill has the third largest Total Persons in NSW.

Power Pivot calculates the Rank value in the context of other filters. For example, if we removed the NSW filter then we would get a Rank of 11 for Castle, showing that Castle Hill is the 11<sup>th</sup> largest suburb, Australia wide.

### 1.10.4. # Rank Change 2011 to 2016:=[# Suburb Rank 2016] - [# Suburb Rank 2011]

The model contains a number of calculations that compare 2011 and 2016 census values. The **# Rank Change 2011 and 2016** shows the difference in Rank between 2011 and 2016. This gives an indication of whether the suburb has changed relative to the change in other suburbs.

1.10.5.      # Population 2016:=IF(ISBLANK( [# Data Points]),BLANK(),  
CALCULATE([# Census 2016], ALL('Census Metric'),  
'Census Metric'[Census Metric]= "Total Persons Persons"))

The **# Population 2016** calculates the Total Persons metric even if we have selected another metric. This is useful because it allows us to compare a metric to the overall population of the suburb or selection (see below).

1.10.6.      % Population 2016:=DIVIDE([# Census 2016],[# Population  
2016])

The **% Population 2016** calculates the current metric/selection as a % of the Total Population.

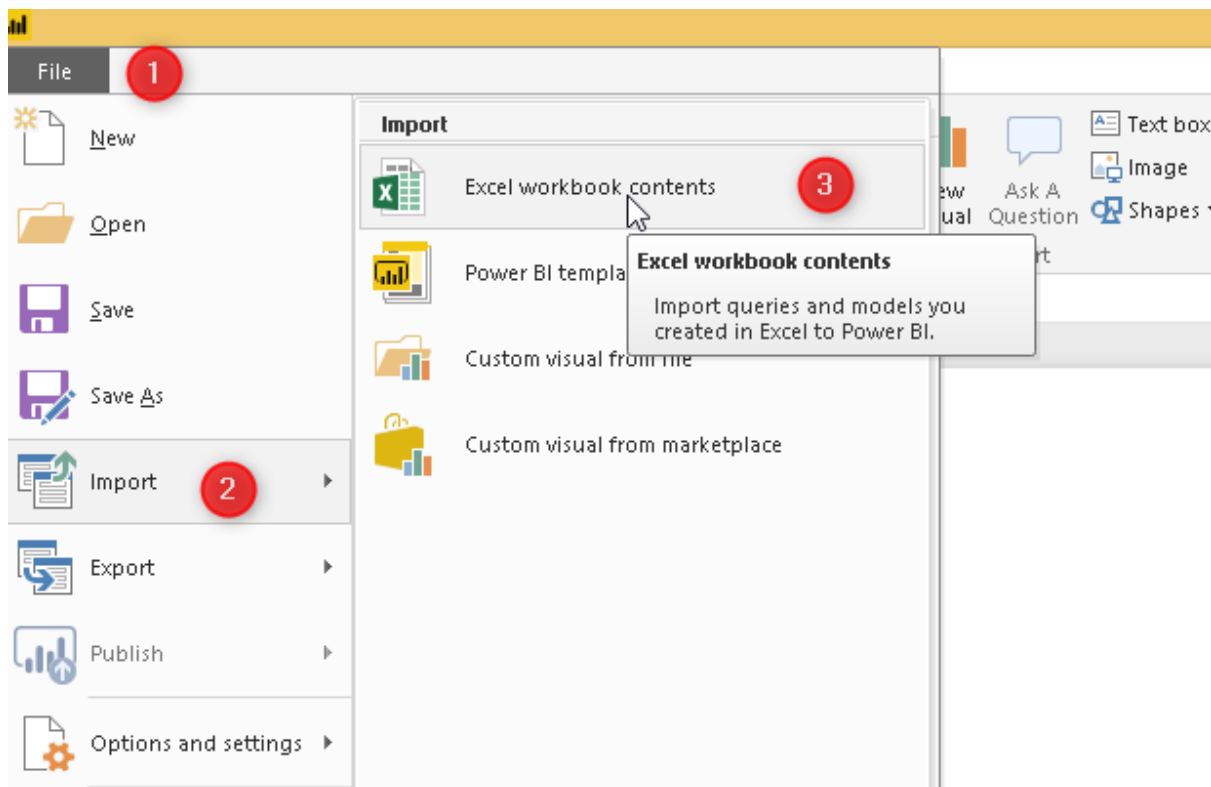
For instance, for Castle Hill it is useful to know that 13,799 persons speak Other Language at home but it provides further context to know that this accounts for 34.85% of Castle Hill's population.

1.10.7.      # Suburb Rank % Population 2016:=RANKX(ALL('Suburb  
Location'[Suburb Name]), [% Population 2016])

The **# Suburb Rank % Population 2016** calculates the suburb rank for % of population of the metric. Following the Castle Hill example above, we see that Castle Hill is ranked 261 as a % of Person speaking another language at home within NSW.

## 1.11. Importing the Excel Power Pivot model into Power BI

One of the advantages of developing Power Pivot models in Excel is that you can import them into Power BI using Power BI Desktop. You cannot import a Power BI Desktop model into Excel so if you want to have a model that you can use in both tools then we recommend you develop the initial model in Excel



Having developed the Excel Power Pivot model it is just a few clicks to create a Power BI Desktop version of the model. After opening Power BI Desktop

1. Select the File menu
2. Select Import
3. Select Excel workbook contents

You then select the path of the existing Excel Power Pivot model and Power BI Desktop imports the Data Model, DAX calculations, and Power Queries. You can now take advantage of all the Power BI features including

- Interactive visualisations and cross filtering
- Publishing to mobile devices and the web
- Performing Q & A natural language queries
- Extending and enhancing the model with the latest DAX functions

## 1.12. Power BI Census Model on the web

We converted the Power Pivot model to a Power BI model using the process described above. Having done this we created some visualisations and published the model to the Power BI web service and then to the web where any web user worldwide can view and interact with the reports.

[http://www.cmbi.com.au/Power\\_BI\\_ABS\\_Census\\_Model\\_Sample.html](http://www.cmbi.com.au/Power_BI_ABS_Census_Model_Sample.html)

For ease of use, we have published the report pages as separate displays down the page but you can navigate between the different reports from any of the visualisations using the Power BI toolbar.

Way Way	2	8	4	0.04%	0.08%	0.04%	10	-49
Bellevue Hill	2	7	6	0.06%	0.07%	0.01%	15	-10
<b>Total</b>	<b>457</b>	<b>765</b>	<b>1,087</b>	<b>0.04%</b>	<b>0.02%</b>	<b>-0.01%</b>	<b>1</b>	<b>0</b>

Microsoft Power BI

< 1 of 4 >

1 2

LinkedIn Twitter Facebook Print

1. You can navigate between the report pages using the > controls
2. You can share this report with a friend or colleague using social media (please do!)

## 1.13. Download links

We have made available all the files for this sample.

If you want to explore the existing model and do not want to reprocess the data or extend the data model then you will only require the final sample files

### 1.13.1. Links to fully processed models

These fully processed model files include all the data. No further work required to analyse all 44m data points. However, they are quite large files. The Excel file is 120 MB and the Power BI file is 78 MB. They both contain the same model and data. Power BI Desktop files tend to be smaller than the equivalent Excel Power Pivot models.

#### Excel Power Pivot sample link

- [http://www.cmbi.com.au/CMBI\\_Census\\_Model/CMBI\\_ABS\\_Australian\\_Census\\_Power\\_BI\\_Sample\\_V1.0.xlsx](http://www.cmbi.com.au/CMBI_Census_Model/CMBI_ABS_Australian_Census_Power_BI_Sample_V1.0.xlsx)

#### Power BI sample link

- [http://www.cmbi.com.au/CMBI\\_Census\\_Model/CMBI\\_ABS\\_Australian\\_Census\\_Power\\_BI\\_Sample\\_V1.0.pbix](http://www.cmbi.com.au/CMBI_Census_Model/CMBI_ABS_Australian_Census_Power_BI_Sample_V1.0.pbix)

### 1.13.2. Links to component files for reprocessing the data models

If you would like to explore the models in more detail and reprocess them then you will need the following files

#### Additional reference data for location and census metrics

- [http://www.cmbi.com.au/CMBI\\_Census\\_Model/SCC\\_Locality\\_Map.csv](http://www.cmbi.com.au/CMBI_Census_Model/SCC_Locality_Map.csv)
- [http://www.cmbi.com.au/CMBI\\_Census\\_Model/Census\\_Table\\_Group.txt](http://www.cmbi.com.au/CMBI_Census_Model/Census_Table_Group.txt)

#### Unprocessed version of Power Pivot model

- [http://www.cmbi.com.au/CMBI\\_Census\\_Model/CMBI\\_ABS\\_Australian\\_Census\\_Power\\_BI\\_Sample\\_V1.0%20SHELL.xlsx](http://www.cmbi.com.au/CMBI_Census_Model/CMBI_ABS_Australian_Census_Power_BI_Sample_V1.0%20SHELL.xlsx)
- [http://www.cmbi.com.au/CMBI\\_Census\\_Model/CMBI\\_ABS\\_Australian\\_Census\\_Power\\_BI\\_Sample\\_V1.0.pbix](http://www.cmbi.com.au/CMBI_Census_Model/CMBI_ABS_Australian_Census_Power_BI_Sample_V1.0.pbix)

## 1.14. Understanding the Power Pivot SHELL file

The Excel file CMBI\_ABS\_Australian\_Census\_Power\_BI\_Sample\_V1.0 SHELL.xlsx is a much smaller file than the full model because it contains an unprocessed Census Pivot table.

If you want to process the data using this file then there are four steps:

1. Download the relevant data packs from the ABS website
2. Set the Power Query parameter pLoadFiles to the second option: .csv
3. Set the other parameters to points to the location where you placed the data folders and files
4. Process the Census Pivot Power Query query

### 1.14.1. Download data pack files

We used two suburb level data packs to build this model. You can download the data packs from

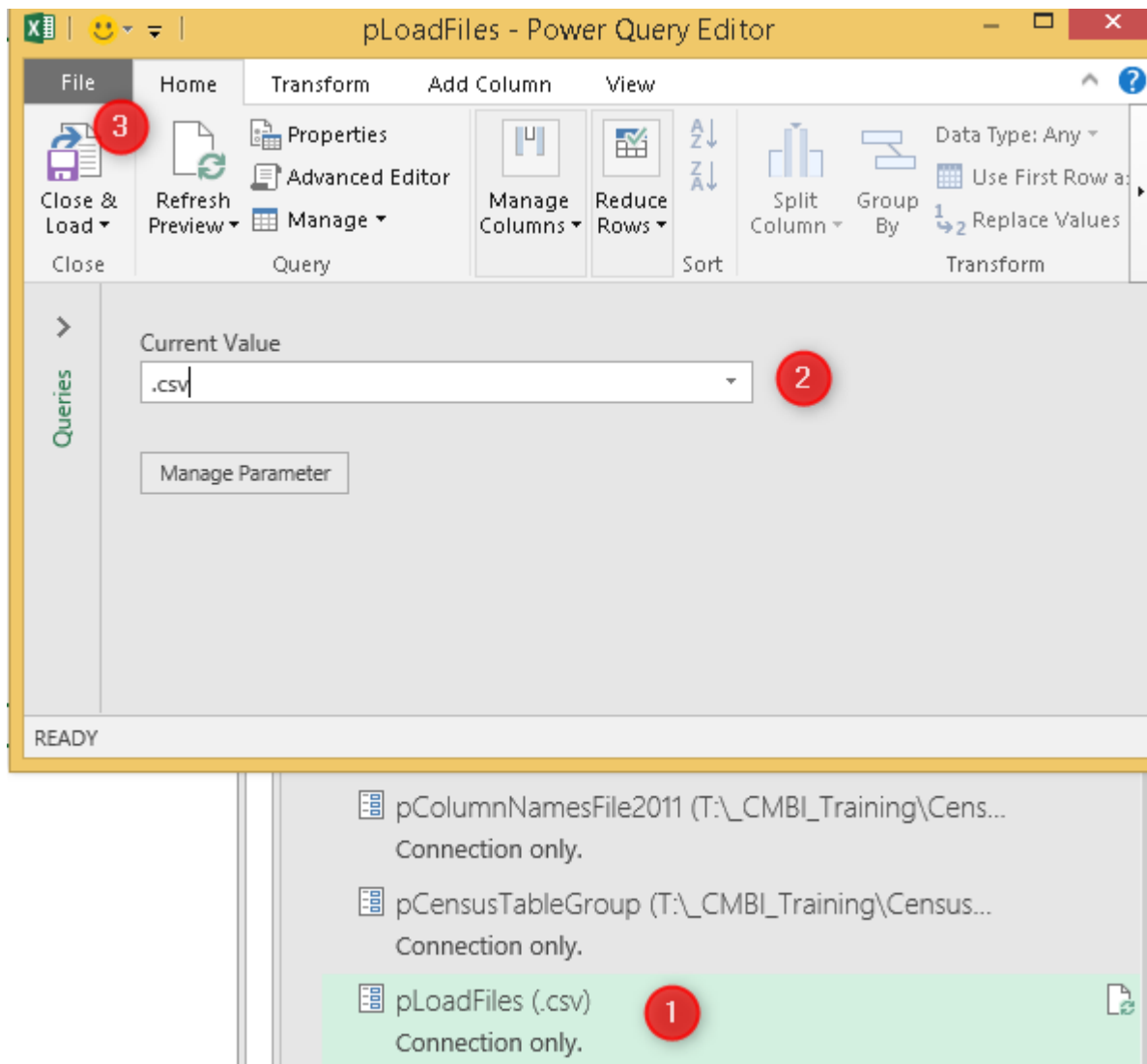
<https://datapacks.censusdata.abs.gov.au/datapacks/>

- 2016\_GCP\_SSC\_for\_AUS\_short-header.zip
- 2011\_BCP\_SSC\_for\_AUST\_short-header.zip

### 1.14.2. Set the pLoadFiles parameter

The Power Query parameter accepts two values. You can change the parameter by following the steps in the screenshot below:

1. Right click edit the pLoadFiles query
2. Select .csv from the Current Value dropdown
3. Close and Load



### 1.14.3. Set the other model parameters

Repeat the process above for each of the other parameters which contains the folder and file paths of each of the source data elements

### 1.14.4. Process the Census Pivot query

Finally, right click the Census Pivot query in the Census Transforms group and click Refresh...

The query is very resource intensive and may take 10 minutes or more to complete.

## ABOUT CMBI

CMBI is a Sydney (NSW) based business intelligence and data analytics consultancy established in 2010. Our clients benefit from our extensive industry experience which spans all stages of the solution lifecycle from business analysis and facilitation through to development, testing and training.

### Meet the CMBI Trainers

#### **Colin McGowan LL.B, PGDip Soft Dev, MSc Computing**

For the last 15 years, Colin has worked as a solution architect and consultant designing BI and data warehouse solutions for multinational organisations in London (UK) and Sydney (Aus). The projects spanned over 50 organisations across a number of industries including banking and finance, market research, international law firms, online media, and government departments.

Connect with [Colin McGowan on LinkedIn](#)



#### **Julie McGowan B.S. Computer Science (USyd)**

Julie is a Sydney-based business analyst and training facilitator who has worked on a diverse array of data-driven business and technical projects within some of the world's most respected financial services organisations in both London and Sydney.

Connect with [Julie McGowan on LinkedIn](#)





## CONTACT DETAILS

**Colin McGowan**

CMBI - Business Intelligence Consultancy

[www.cmbi.com.au](http://www.cmbi.com.au) | 0432 240 260 | [colin@cmbi.com.au](mailto:colin@cmbi.com.au)

**Julie McGowan**

CMBI - Business Intelligence Consultancy

[www.cmbi.com.au](http://www.cmbi.com.au) | 0432 557 893 | [julie@cmbi.com.au](mailto:julie@cmbi.com.au)