

# ABS 2021 Mesh Block dwelling counts to CSV

As of 31 August 2022, the Australian Bureau of Statistics has only released the [2021 census Mesh Block dwelling counts](#) as an [Excel file](#), with data stratified across 12 worksheets. This is inconvenient for users who wish to link this data with digital boundaries, and not useful for data posterity. In addition to completing a feedback survey for the ABS, suggesting that a CSV download should be provided (which the Excel file itself suggests is the case in its Explanatory Notes, and was the case with the 2011 and 2016 census releases), I thought it would be useful to take the time to compile these and make them public for myself and others until an official release is produced.

Carl Higgs 31 August 2022

```
In [1]: # setup
import urllib.request
import pandas as pd
import numpy as np

In [2]: # Get the data
meshblocks_url = "https://www.abs.gov.au/census/guide-census-data/mesh-block-counts/2021"
file = "ABS 2021 Mesh Block Counts.xlsx"
urllib.request.urlretrieve(meshblocks_url, file)

Out[2]: ('ABS 2021 Mesh Block Counts.xlsx', <http.client.HTTPMessage at 0x2bc1170f8e0>)
```

By manually opening up the download, it is confirmed that the Mesh Block counts data is stratified as follows:

Sheet	State or Territory	Header row	Last row	Footer length
Table 1	New South Wales	7	60000	4
Table 1.1	New South Wales	7	52752	4
Table 2	Victoria	7	60000	4
Table 2.1	Victoria	7	28753	4
Table 3	Queensland	7	60000	4
Table 3.1	Queensland	7	11876	4
Table 4	South Australia	7	28438	4
Table 5	Western Australia	7	43329	4
Table 6	Tasmania	7	13041	4
Table 7	Northern Territory	7	3368	4
Table 8	Australian Capital Territory	7	6669	4
Table 9	Other Territories	7	143	4

The table header was on row 7 of each sheet, and contained the following fields:

Variable	Data type
MB_CODE_2021	str

MB_CATEGORY_NAME_2021	str
AREA_ALBERS_SQKM	float
Dwelling	int
Person	int
State	int

While the MB\_CODE\_2021 variable is technically a really big integer (11 digits long; ie. longer than a 32-bit integer can be represented without an [overflow error](#)); therefore, the ABS represents the Mesh Block code as a string, and so this convention has been maintained in this dataset. Alternatively, the MB\_CODE\_2021 category could be represented using a 64-bit integer, and I expect there could be some optimisation benefits in doing that for indexed queries of large databases. But I won't worry about that here.

Some records may have nulls, so [Int64](#) data type will be used to represent integers, allowing for these occurrences to be correctly retained.

The last row for each table was checked, with any records going beyond row 60,000 cascading to a subsequent table for that State or Territory. It was also confirmed that the length of footer on each page was consistently 4 rows.

So, we'll create a data structure reflecting the above table, so we can iterate over and compile each State and Territories' Mesh Block records into a single master table for export to CSV. I'll then upload that to a public repository for myself and others to use, until an official single CSV release becomes available.

The data also includes the explanatory note, "Cells in this table have been randomly adjusted to avoid the release of confidential data." and "© Commonwealth of Australia 2022"; users should note this, and that this data was made available under a [Creative Commons Attribution 4.0 International licence](https://www.abs.gov.au/website-privacy-copyright-and-disclaimer#copyright-and-creative-commons) as per <https://www.abs.gov.au/website-privacy-copyright-and-disclaimer#copyright-and-creative-commons>.

```
In [3]: sheets = {
    "New South Wales"      : [{"sheet_name": "Table 1", "header": 7, "skip_footer": 4}, {"
    "Victoria"             : [{"sheet_name": "Table 2", "header": 7, "skip_footer": 4}, {"
    "Queensland"           : [{"sheet_name": "Table 3", "header": 7, "skip_footer": 4}, {"
    "South Australia"      : [{"sheet_name": "Table 4", "header": 7, "skip_footer": 4}],
    "Western Australia"    : [{"sheet_name": "Table 5", "header": 7, "skip_footer": 4}],
    "Tasmania"             : [{"sheet_name": "Table 6", "header": 7, "skip_footer": 4}],
    "Northern Territory"   : [{"sheet_name": "Table 7", "header": 7, "skip_footer": 4}],
    "Australian Capital Territory" : [{"sheet_name": "Table 8", "header": 7, "skip_footer": 4}],
    "Other Territories"     : [{"sheet_name": "Table 9", "header": 7, "skip_footer": 4}],
    }
    column_types = {
    'MB_CODE_2021'          : 'object',
    'MB_CATEGORY_NAME_2021': 'object',
    'AREA_ALBERS_SQKM'     : 'float64',
    'Dwelling'             : 'Int64',
    'Person'               : 'Int64',
    'State'                : 'Int64',
    }
```

```
In [4]: # Load up Mesh Block count dataframes for each State and Territory
dfs = {}
for state in sheets.keys():
    for s in sheets[state]:
        df = pd.read_excel(file,
                           sheet_name=s['sheet_name'],
                           header=s['header']-1,
                           usecols="A:F",
```

```
dtype=column_types,
skipfooter=s['skip_footer'])
if state not in dfs.keys():
    dfs[state]=df.copy()
else:
    dfs[state]=dfs[state].append(df.copy())
```

```
In [5]: # Combine Mesh Block count dataframes
df = pd.concat(dfs)
# drop the within-state sequential index
df = df.droplevel(1).reset_index()
# Rename the State/Territory identifier columns
df.columns = ['STATE_NAME_2021'] + list(column_types.keys())[:-1] + ['STATE_CODE_2021']
# Re-order the columns so State name comes after Mesh Block Category
# State name will be retained so people don't have to mess around looking up State codes
df = df[['MB_CODE_2021',
'MB_CATEGORY_NAME_2021',
'STATE_CODE_2021',
'STATE_NAME_2021',
'AREA_ALBERS_SQKM',
'Dwelling',
'Person']]
```

```
In [6]: df
```

Out[6]:

	MB_CODE_2021	MB_CATEGORY_NAME_2021	STATE_CODE_2021	STATE_NAME_2021	AREA_ALBERS_SQKM
0	10000010000	Residential	1	New South Wales	0.0209
1	10000021000	Commercial	1	New South Wales	0.0829
2	10000022000	Commercial	1	New South Wales	0.0388
3	10000023000	Commercial	1	New South Wales	0.0254
4	10000024000	Residential	1	New South Wales	0.0376
...	...	...	...	...	...
368280	97500000000	Parkland	9	Other Territories	27.6524
368281	97000001777	SHIPPING	9	Other Territories	NaN
368282	97000002777	SHIPPING	9	Other Territories	NaN
368283	97000003777	SHIPPING	9	Other Territories	NaN
368284	90000009499	NOUSUALRESIDENCE	9	Other Territories	NaN

368285 rows × 7 columns

```
In [7]: # output national combined Mesh Block counts dataset
df.to_csv(f"{file.split('.')[0]}.csv",index=False)
```

```
In [8]: # output state specific Mesh Block counts datasets
for state in dfs.keys():
    dfs[state].to_csv(f"{file.split('.')[0]} - {state}.csv",index=False)
```

```
In [13]: df.groupby('STATE_NAME_2021')[['AREA_ALBERS_SQKM','Dwelling','Person']].sum()
```

Out[13]:

	AREA_ALBERS_SQKM	Dwelling	Person
STATE_NAME_2021			
Australian Capital Territory	2.358133e+03	187063	454369

<b>New South Wales</b>	8.007977e+05	3361910	8069912
<b>Northern Territory</b>	1.348134e+06	96348	232509
<b>Other Territories</b>	2.557420e+02	2198	4796
<b>Queensland</b>	1.730171e+06	2193417	5155857
<b>South Australia</b>	9.842314e+05	807782	1781087
<b>Tasmania</b>	6.801754e+04	259156	557368
<b>Victoria</b>	2.274962e+05	2809056	6502955
<b>Western Australia</b>	2.526632e+06	1149491	2659569