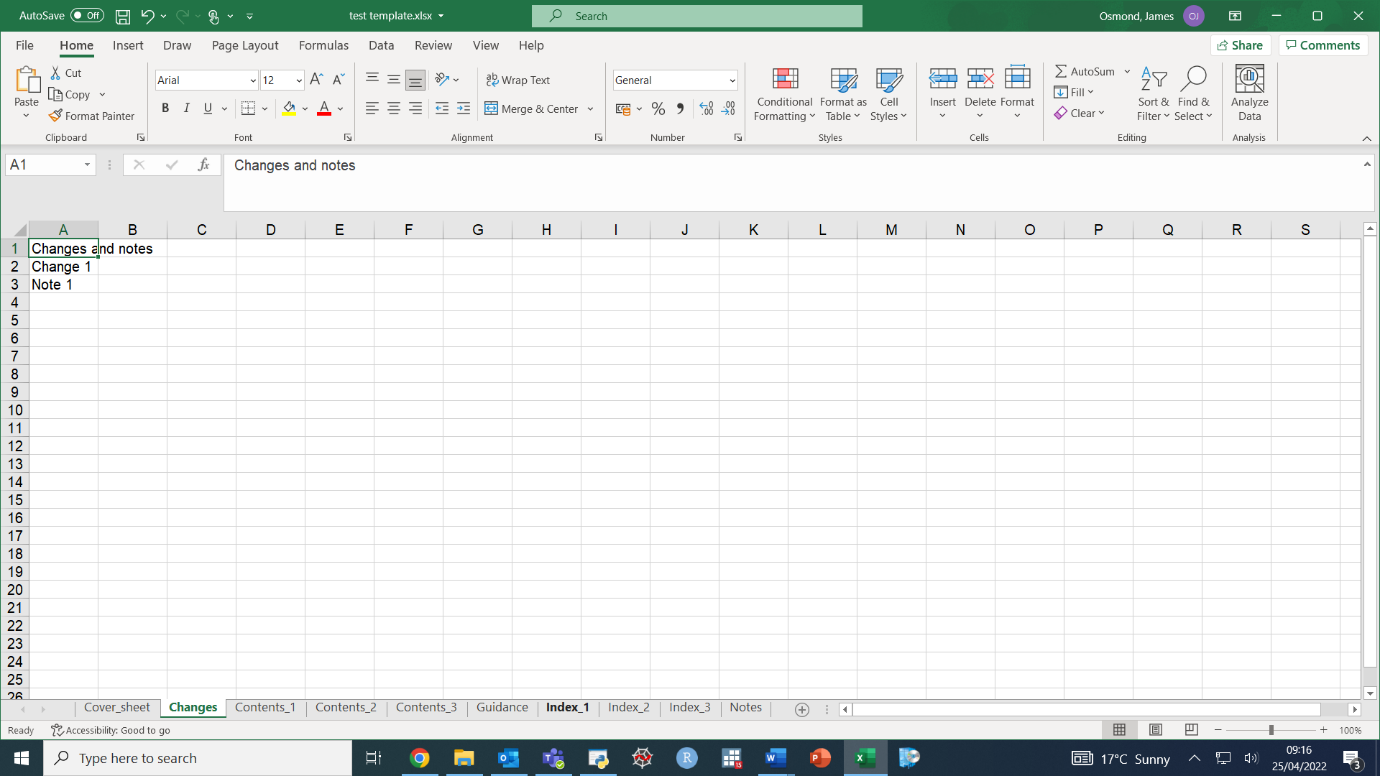
# Initialisation template

The *produce\_dataset()* function makes use of what is called an initialisation template, which determines key characteristics of the presentation of the data and the Excel workbook that will be produced. Changes to certain aspects of formatting can therefore be made easily and intuitively inside this template.

## Cover sheet

A list of changes since the previous release can also be placed on the cover sheet of the dataset. These are optional. These notes should be placed in the ‘Changes’ sheet of the initialisation template, in cells A2, A3,.... Do *not* remove the ‘Changes’ sheet, nor the content of cell A1, simply leave cells A2 onwards blank if there are no changes.

Hyperlink to dataset release page or the URL of the associated release. For readability purposes, it is usually appropriate to give this link another name that is displayed instead of the URL itself. This will appear in cell A3.



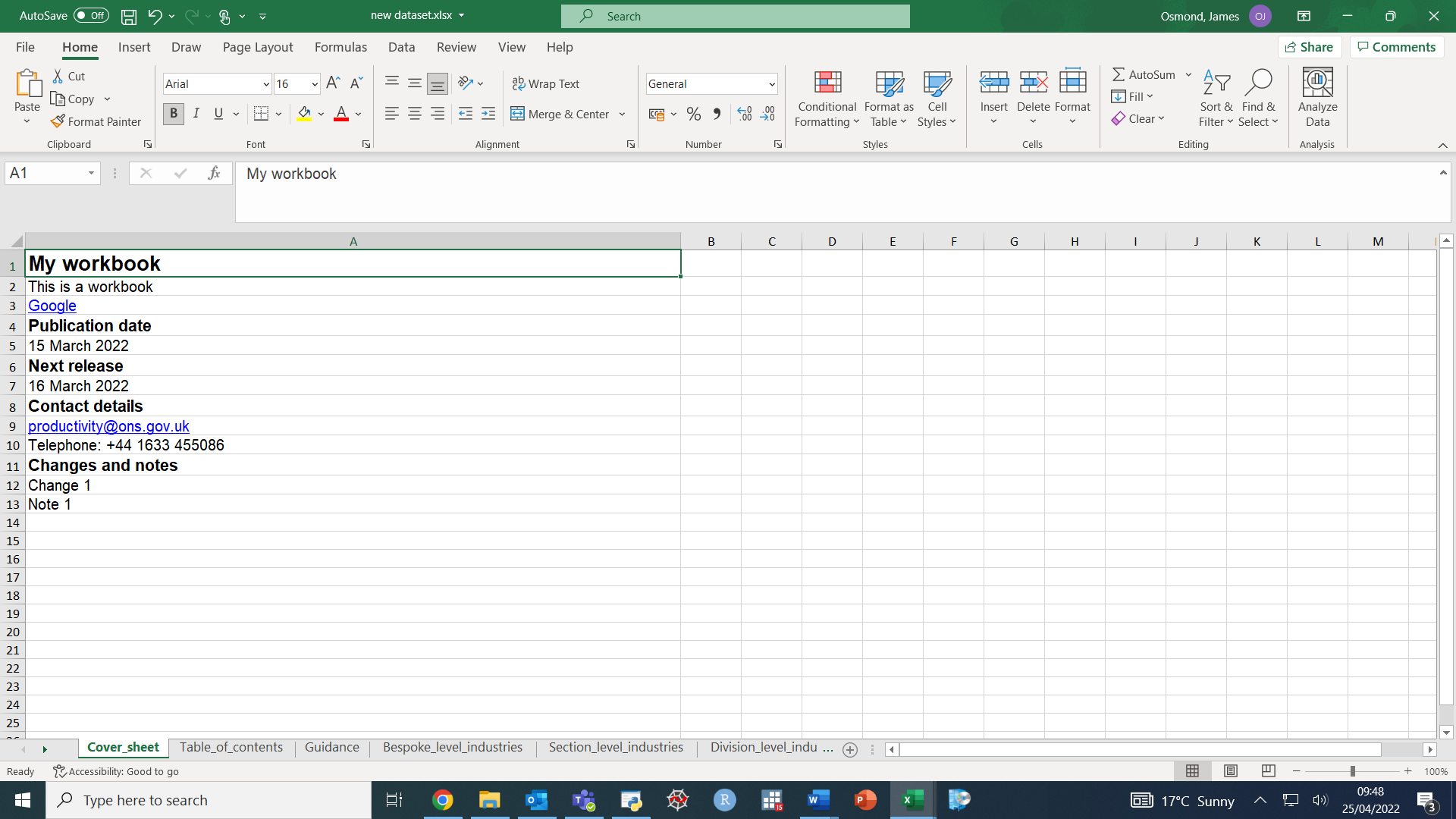
Contact details for this dataset.

Dates of this and the next release. Can be entered as TBC or similar.

Summary of the workbook’s contents and its associated release if applicable. Cell A2 on worksheet.

Title of workbook – appears on coversheet in cell A1

The above template produces the following final cover sheet:



## Contents

In this package, the contents table initialisation sheets not only direct how the contents tables will be formatted, but also dictate certain aspects of the formatting of the worksheets referenced (such as the precision to which data points are rounded).

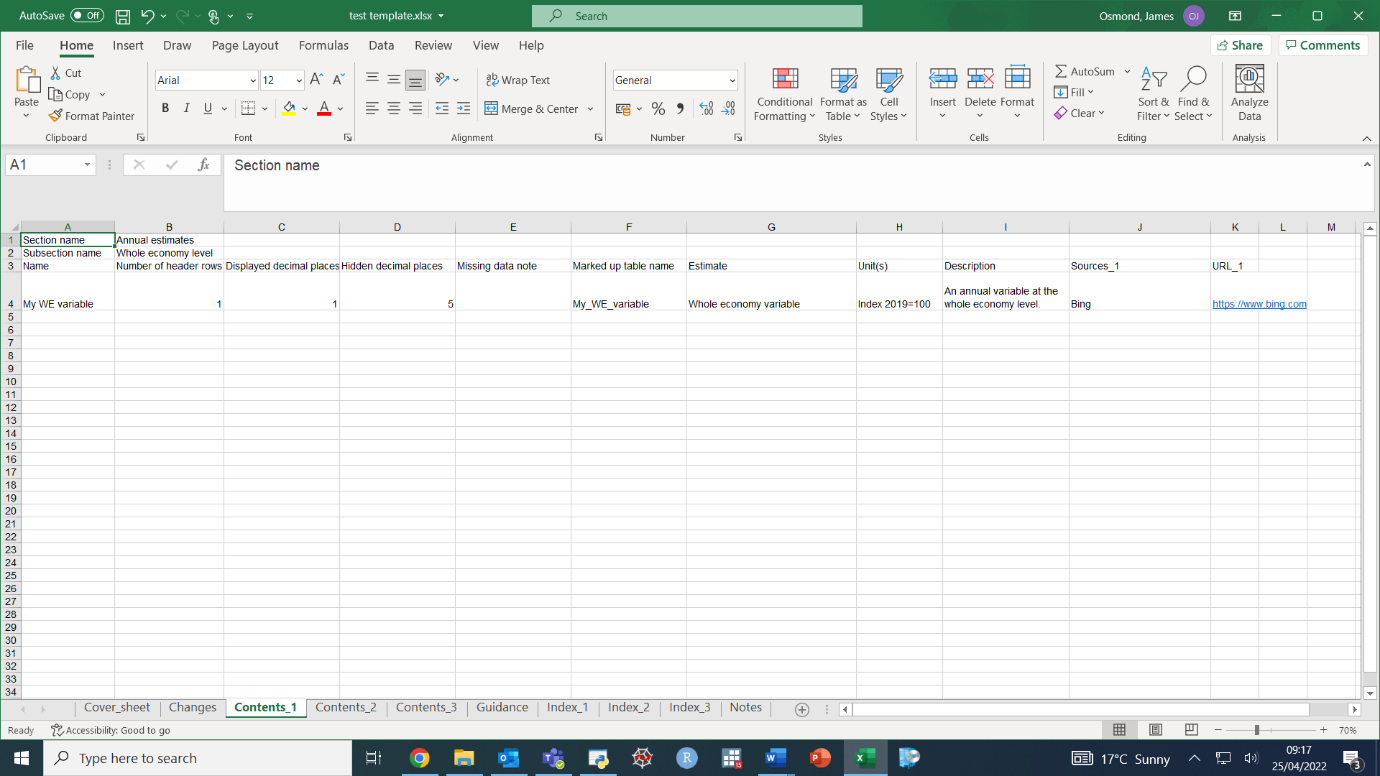
Contents sheet can be divided up into sections, each section containing multiple tables. This is an optional feature to add – if cell B1 is left blank in every contents sheet, no sections will be added.

Appears on contents sheet. Tells user what variable the table contains.

How many decimal places will be available (e.g. by clicking on a data point)

How many decimal places will be displayed by default in the table.

Number of header rows of the table (e.g. how many rows at the top of the data table need to be formatted in bold)



Appears on contents sheet. A source can be given, with an optional URL that it links to.

Appears on contents sheet. Gives user a description of the table.

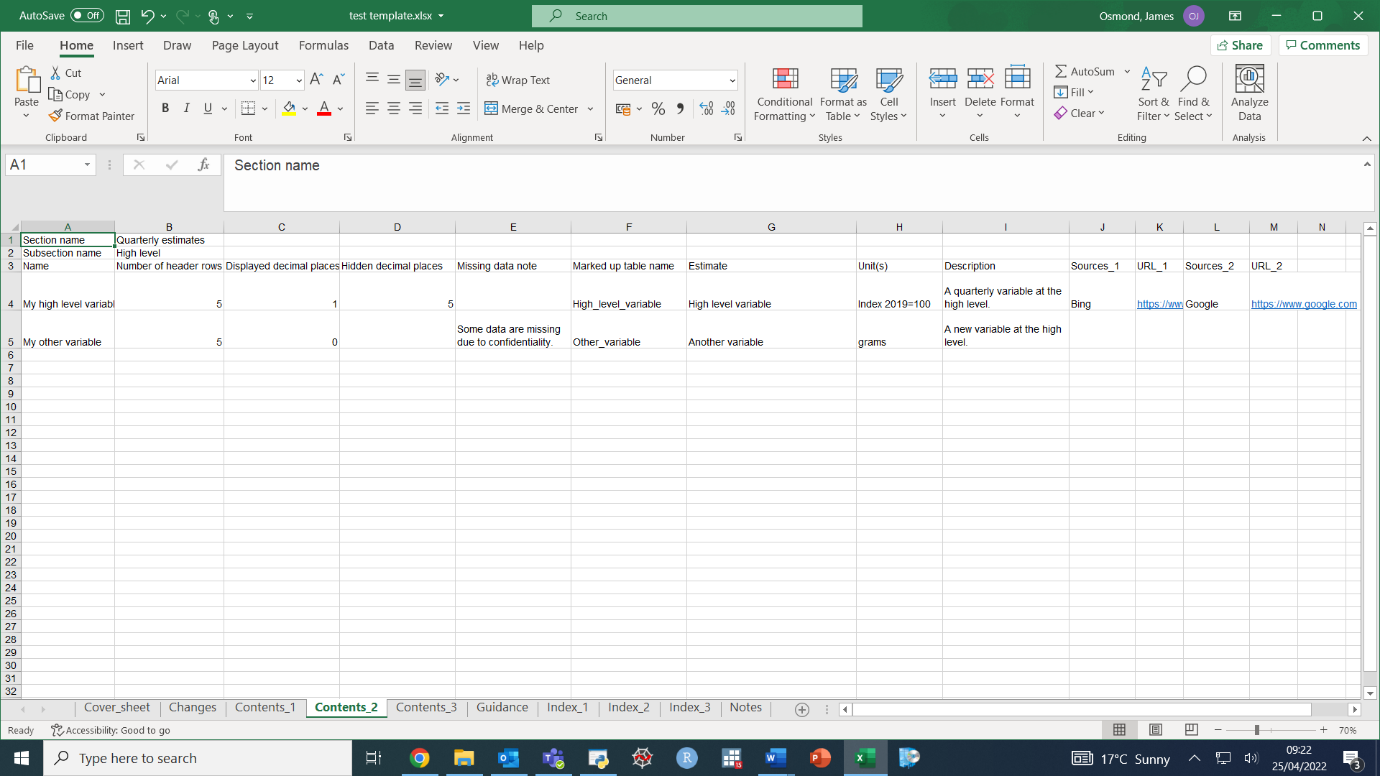
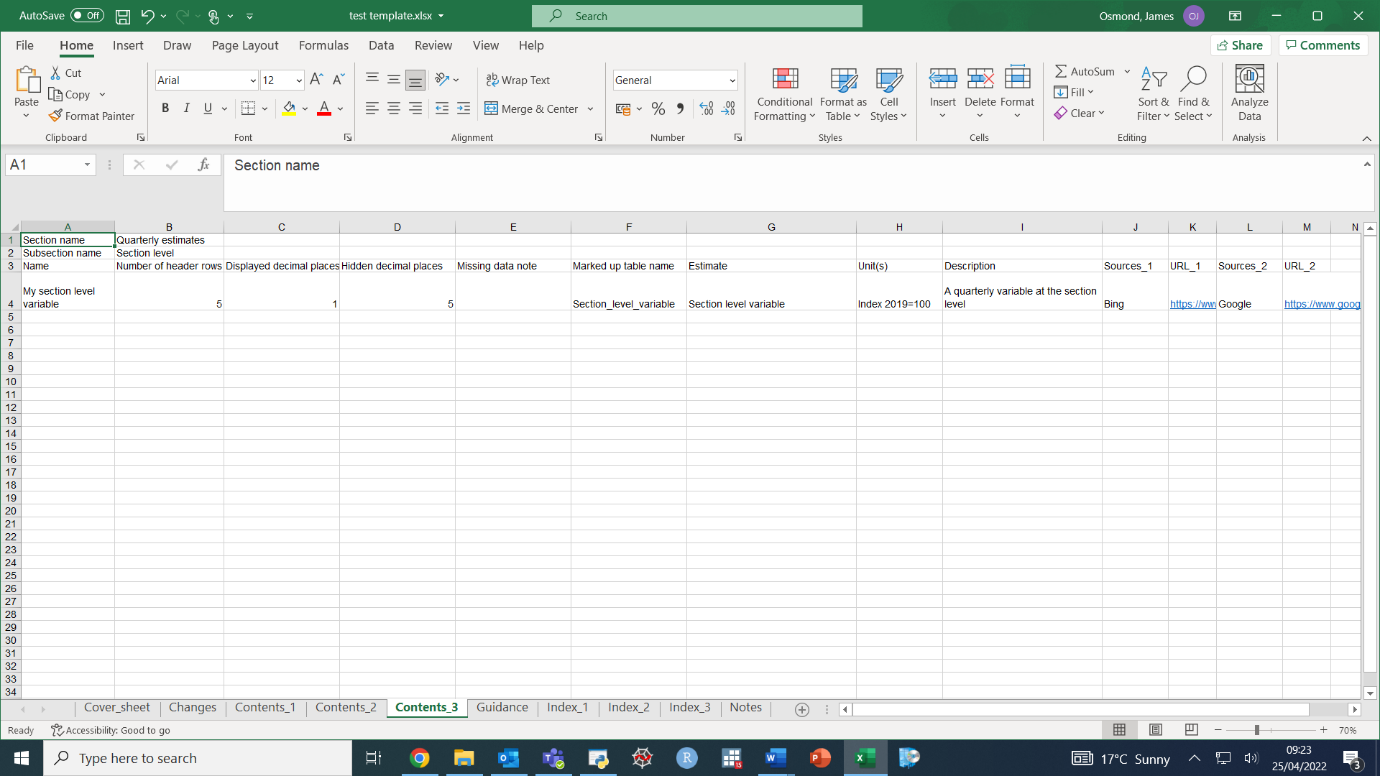
Appears on contents sheet. Tells user the unit of the table.

Each table should be marked up with a name that explains the content of the table (no blank spaces, etc) This can be used by users of screen reader software.

If there are missing data, or an explanatory note is otherwise needed, this can be defined here. (if there are missing data but no note, a note will be written by default.

Name of the worksheet as displayed in the worksheet itself (this example will have cell A1 of the worksheet read ‘Worksheet: My WE variable’)

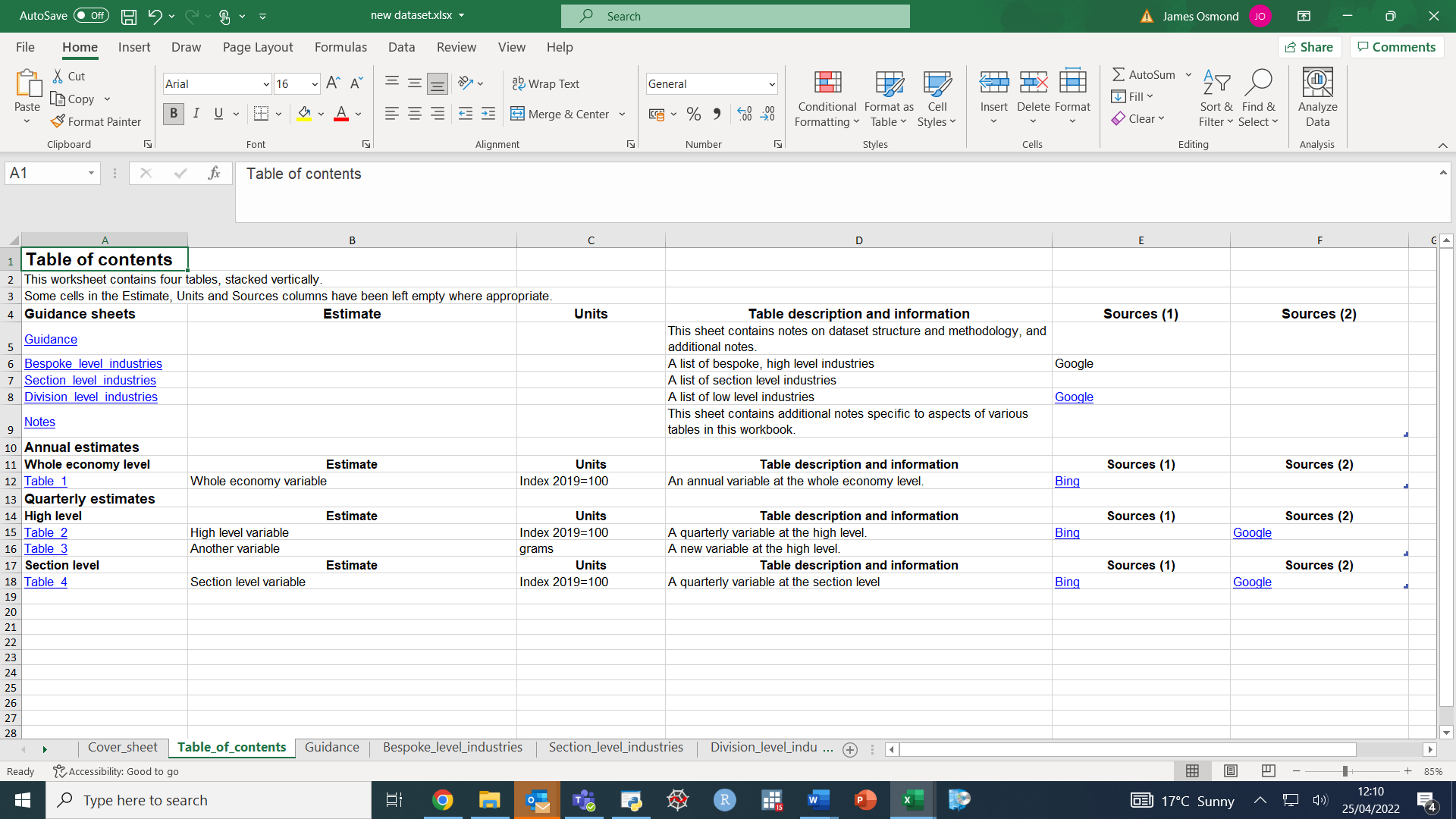
If there is more than one contents table, a unique ‘Subsection name’ is necessary. This will dictate what is in the top left cell of the specified table (e.g. the far left column heading)



Same section name – these two tables will be added to the same section on the final contents sheet.

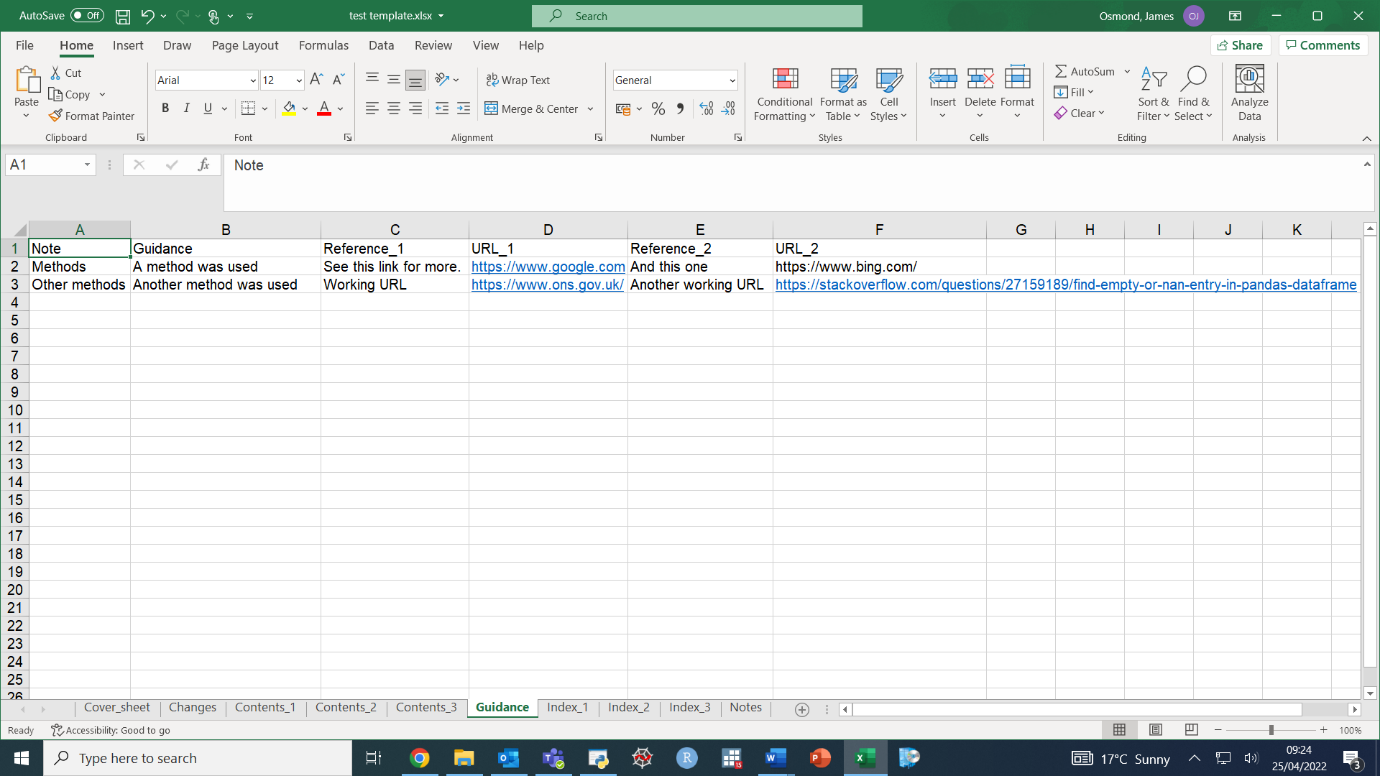
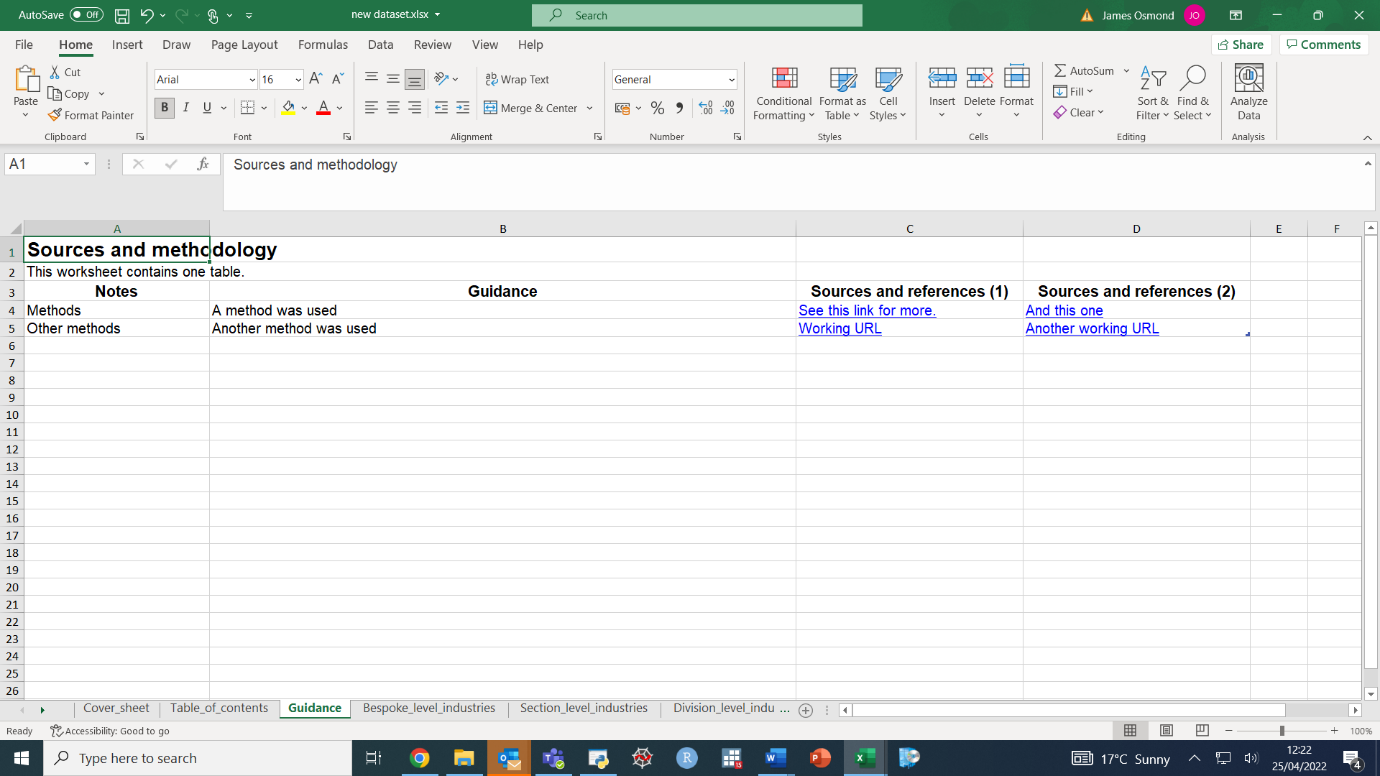
Multiple sources can be added – for each source, simply add two new columns Sources\_X and URL\_X (even if there is no URL, add both columns)

The above contents tables as specified in the initialisation template will produce the following Table\_of\_contents worksheet in the final dataset. Notice the different section headings that we defined, and the top left cells of each table. Also notice that when there is more than one contents table specified, there is a *separate* contents table for all guidance sheets.



## Guidance sheet

In the final dataset, the workbook will have a ‘Guidance’ sheet for explaining methodologies and sources. This sheet will contain a table, which will be ordered in almost the same way in the initialisation template as it will be in the final dataset (but as with the contents sheets, text for references and the corresponding URLs will need to be given separately). There should be no emptycells in the ‘Note’ and ‘Guidance’ columns before the final row. It is okay to have Reference and URL columns be empty (though if there is a URL, there must be a reference).



## Index sheets

Datasets will often need an extra table to define certain terms – in Productivity, these come in the form of defining industry codes.

If no index sheet is necessary, leave sheet ‘Index\_1’ in the initialisation template, but leave it blank below row 6.

The contents page needs to reference any index sheets – therefore each sheet needs a description to put in the contents table, defined in cell B3. Similarly, a source to put on the contents page can be defined below.

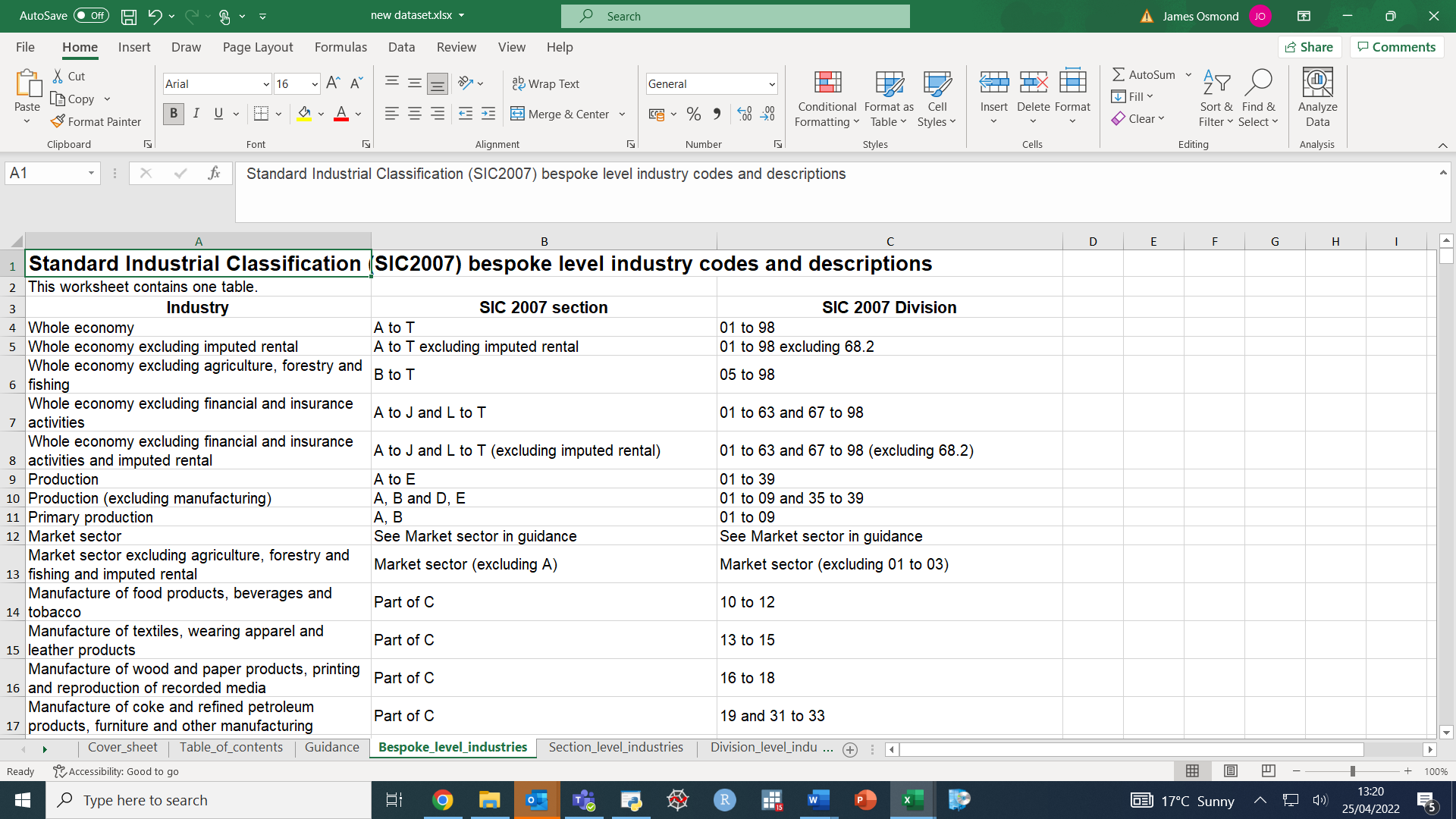
The index sheet will need a title in cell A1. This is defined in cell B2.

The index table in its entirety, complete with heading rows, should be placed starting in row 6.

Index sheets may need bespoke names, so the tab name can be defined in cell B1.

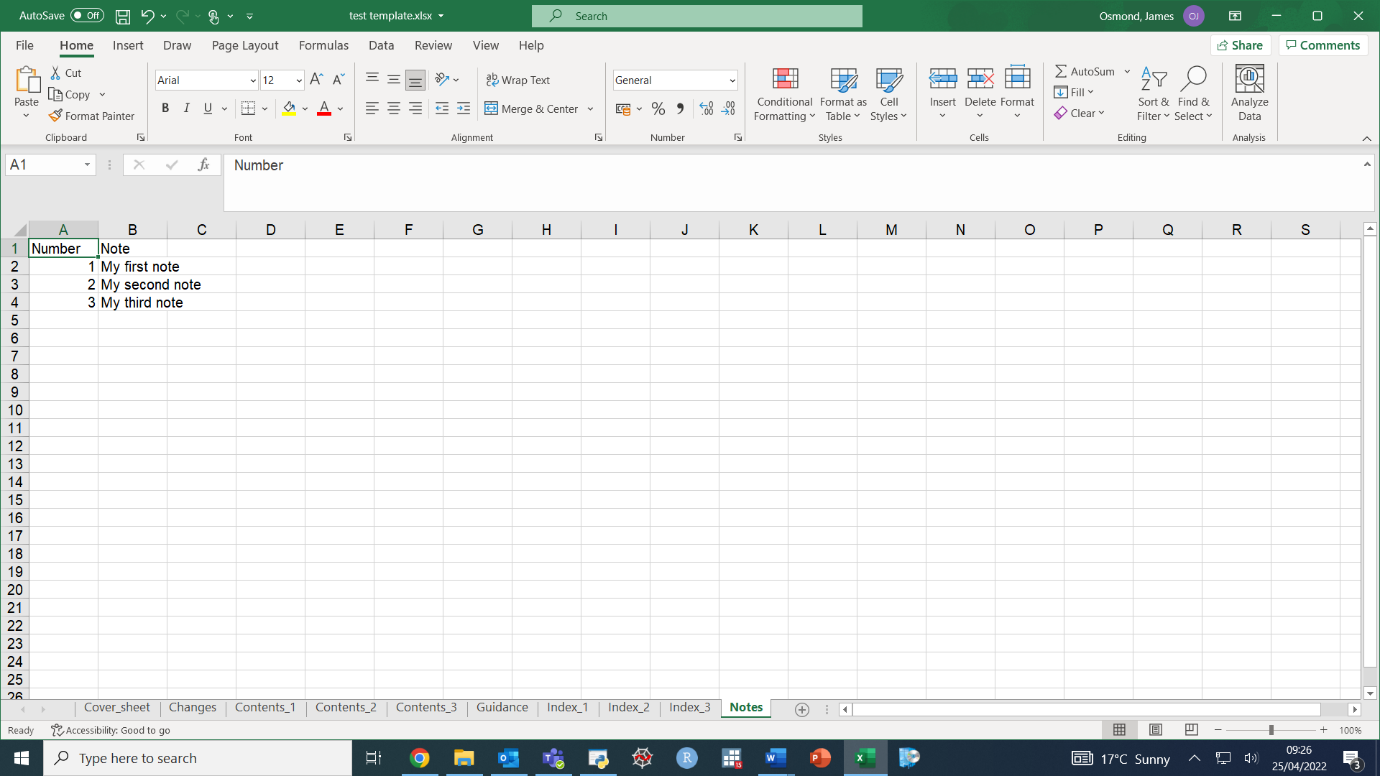
## 

The above sheets will produce index sheets that look like the following:



## Notes

Sometimes specific tables will need certain notes, maybe explaining missing data or changes in methodology for a particular column. The notes need to be appended to row or column labels in data tables, like ‘Column name [note 1]’.

The table of notes, if any notes exist, will be added to the dataset, with two columns. This table will look as it is defined in the initialisation template. If there are no notes, leave the ‘Notes’ sheet in the template, and leave row 1, but remove anything from row 2 onwards.

# Using produce\_dataset()

The function *produce\_dataset()* requires only three arguments. In order, these are:

* The file path to which the final dataset should be written.
* The file path of the initialisation template.
* A Python list of every pandas DataFrame that will be written to the dataset.