If we first open the CSV file into excel to view the data, we notice that there are some lines that need to be cleaned up. There are some lines without a country and without a state attached to them. Since the file contains around 80k lines we can do the cleanup in Microsoft excel.

If we apply a filter in each column, we can see that there are a couple of lines in which states are abbreviated with 2 letters but have no country. Looking at the data we see that these abbreviations are used for states/provinces in the US and Canada. Therefore, we conclude that this column is only used for sightings in these countries. We can easily add the country to these lines using the SWITCH excel function. We will use the switch function in this case. We will create a new column called Country Clean for this.

The code we used id referenced in the Excel\_code document.

Graphical user interface, application

Description automatically generated

If we select only “au”,”gb”,”de” in the countries we see that there are a couple of entries of states that have a country outside the US and Canada. Since the quantity is so small, we manually change those:

Graphical user interface, application, table

Description automatically generated