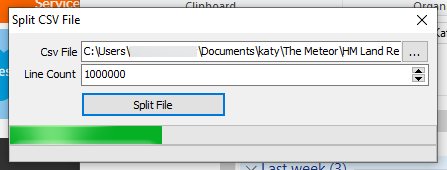
**Using a .csv splitter**

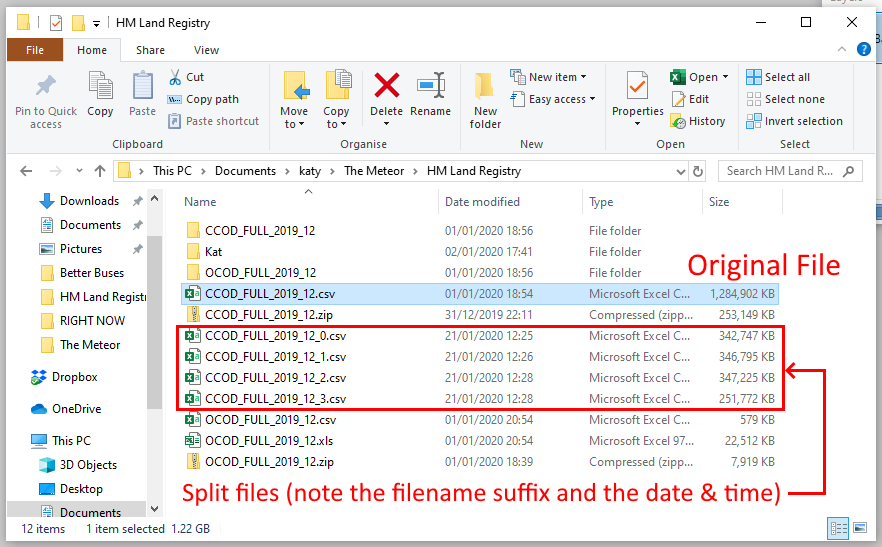
There are various ready-made solutions for breaking .csv files down. The reason I mentioned the ability to open them in text form is that one of my first thoughts was to edit the file by hand and separate it into 3 or 4 other files. As I’ve discovered from text-editing various other files (hello, WordPress!), this is fraught with danger — one character out of place, or delete the wrong line, and the whole file is unusable. Fortunately, .csv splitter programs are better at this than unreliable human operators, so you can just run the file through one of these instead. But that doesn’t mean it’s plain sailing from here on…

I tried a few .csv splitters, with varying success. I just went for the first three that google gave me, stopping at three because the third one was the first I could get to work. I had the best success with [Free Huge CSV Splitter](https://sourceforge.net/projects/splitcsv/files/latest/download), a very simple program that does exactly what you need with no fuss.



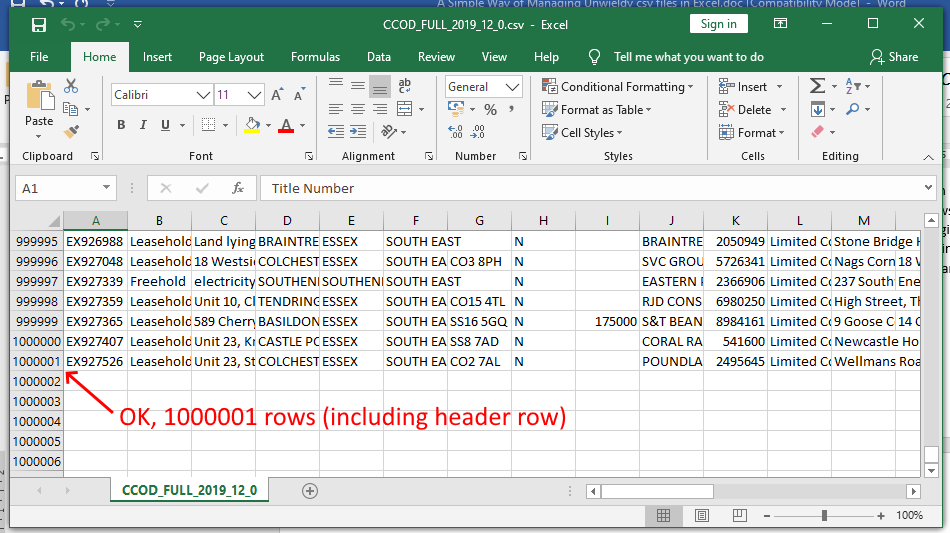
Free Huge CSV Splitter user interface.

It is incredibly basic. You download the .exe file, which you can move to somewhere else, or run directly from your Downloads folder. Choose the file you want to split, and enter how many rows you want in each of the output files. Leave it to run, and check back to the folder where the original file is located when it’s done. You’ll see a number of additional files there, named after the original file with \_1, \_2, \_3, etc appended to the filename. These are your bite-size .csv files that Excel can open:



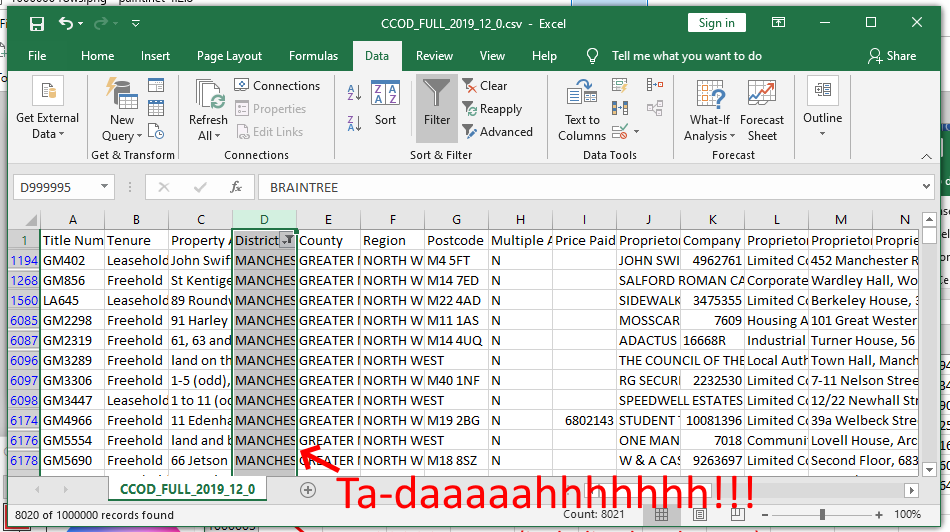
I ended up with four split files. For some reason it starts the numbering at zero with the output filenames. Opening these in Excel was simple and painless, and the files were exactly what I expected, and finished at 1,000,000 rows with some left after.

Annotated Excel spreadsheet showing row 1,000,000.



All that remained to do was the filtering I planned to carry out in the first instance. It was nice to see the data in spreadsheet form at last, and a relief that I’d be able to narrow my search down to just the Manchester area. I ended up with about 40,000 entries for the city of Manchester. The next step was to extract postcode data for each one to plot on a map, but that’s a story for another article.

Annotated Excel spreadsheet showing data filter.



**What didn’t work?**

The previous two google search results were for [CSV Splitter](https://download.cnet.com/CSV-Splitter/3000-2074_4-75910188.html), a very similar program that ran out of memory, and [Split CSV](https://www.splitcsv.com/), an online resource that I was unable to upload my file to.

There are probably alternatives that work fine, but as I said, I stopped once I’d found one that actually worked. I don’t have time to test all the software.

**But Will I Ever Need This Again?**

**Yes.** We are producing data at an astonishing rate, and it’s generating more and more stories. Vast datasets are the perfect vehicle for hiding what one doesn’t want to be found, so investigative journalists are going to have to get used to trawling through massive files to get a scoop. We are carrying out much more of our lives in the digital realm, and it requires new skills in addition to traditional reporting techniques. As this becomes the norm, we’ll develop better solutions for analysing giant datasets, and there will be sophisticated open-source versions available so we won’t have to mess around jumping from program to program to decipher the data. But for now, quick fixes are where it’s at. Just be grateful it’s not a paper copy.