

Workshop 1 Summary | March 3, 2022

Module site: https://aodhanlutetiae.github.io/dj_prod/

UK sanctions list

- How many entries (rows) concern Russian people or entities in March? in February?
- Filtering variables using 'Choose filter - begins with' and 'ends with'
- ODS formats (just another spreadsheet format)
- Compare: <https://www.bbc.co.uk/news/60591547>

Energy performance in England & Wales

- Lots of questions (and some answers) on buildings, money, energy ratings
- Some data supplied without units: where is the metadata?
- Geographical codes for local authorities and constituencies: W06... and W07...
- Single postcode useful for initial questions but could be expanded to much wider area

Heathrow noise complaints

- One question: what's happened to the annual number of complaints?
- What to do to get bigger tables out of pdfs? Tabula probably the best tool: <https://tabula.technology/>

Google operators for looking up data

- Especially *filetype:xls*
- Calendar via Tools (or syntax like *before:2021-11-11*)

Air quality Birmingham

- Data quality: some data provisional, some data missing
- Data collection: where are the monitors that collect this data?
- Syntax: number of cells with missing data =COUNTIF(C2:C10201, "No data")
- Story? How widespread nationally or regionally is "the non-working monitor"?

Home working in 2020

- Some data comes with a caution (small sample size)
- Some data removed for ethical reasons (people identifiable in small samples?)
- Main geographies: E06..., S12..., W06.... (removing single K03... for Great Britain)
- Exported variable 'Never' in clean, csv. Shetland, Scilly etc. still missing numbers

What to do before the next workshop?

- Task A: do this; won't take long; Task B: do this if you have time

- Assignment: make a start and email a basic outline ahead of next workshop