



MODULE 1

INFORMATION & DATA LITERACY

Accessing Data in Dataset

Expert Level Activities

Access to data, information and content and navigation between them

DESCRIPTION OF THE ACTIVITY

Access and navigate datasets through **Tableau Public**

If you are not familiar with Tableau, besides the steps described below for our example, you can study the Tutorial from the official website <https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-home.htm> or any other tutorial found on the web. If you want your content to be saved, you should create an account on the Tableau site. You can download the software from the official site of [Business Intelligence and Analytics Software \(tableau.com\)](https://www.tableau.com) - Products - [Tableau Public](https://www.tableau.com).

1. Download the .csv file from <https://www.ecdc.europa.eu/en/publications-data/data-national-14-day-notification-rate-covid-19>
2. Connect → To a File → Text file (choose *All files (*.*)* if you cannot find the file) → select the .csv dataset you downloaded
3. If your data is not properly presented in tabular form make sure you check “comma” as a separator “Text Field Properties... → Field separator: Comma”

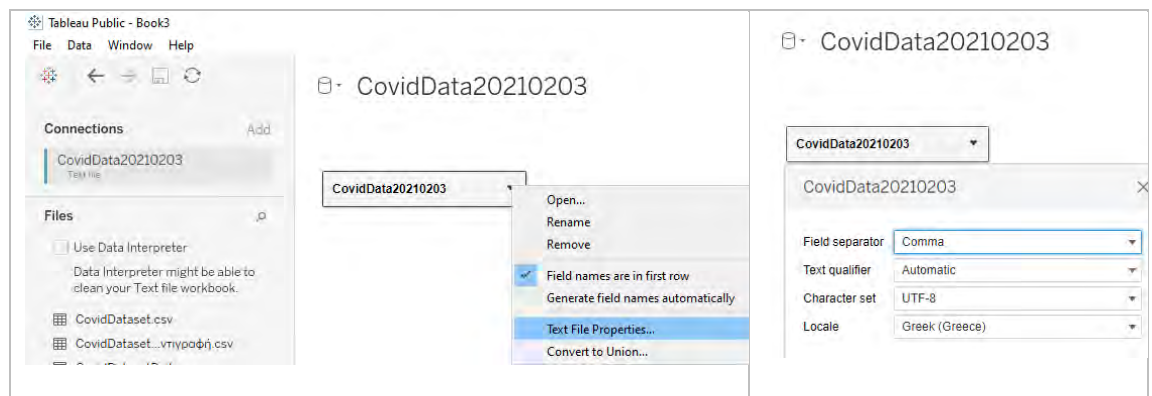


Figure 5: Comma as Text Field Separator

4. From the same position you can “Rename” your Data Source and Sheet if you like. Let’s change it to “COVID-19 Data”

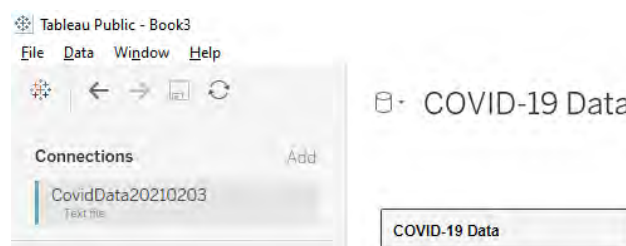


Figure 6: Rename Data Source

5. You can check the type of data per column

country	country_code	continent
Afghanistan	AFG	Asia
Afghanistan	AFG	Asia
Afghanistan	AFG	Asia
Afghanistan	AFG	Asia
Afghanistan	AFG	Asia
Afghanistan	AFG	Asia

Figure 7: Type of Column Data

6. You can hide unneeded columns. We will hide *country_code*, *source*.

country	country_code
Afghanistan	AFG
Afghanistan	AFG
Afghanistan	AFG
Afghanistan	AFG
Afghanistan	AFG
Afghanistan	AFG
Afghanistan	AFG

Figure 8: Hide columns

7. Filtering data, follow “Filters → Add → Add... → Select a field:” and let’s select “United Kingdom”. You can of course make any filtering you like, combination of one or more choices and add more than one filter.

To *remove* or *edit* a filter follow “Filters → Edit → Select your filter → Edit... or Remove”

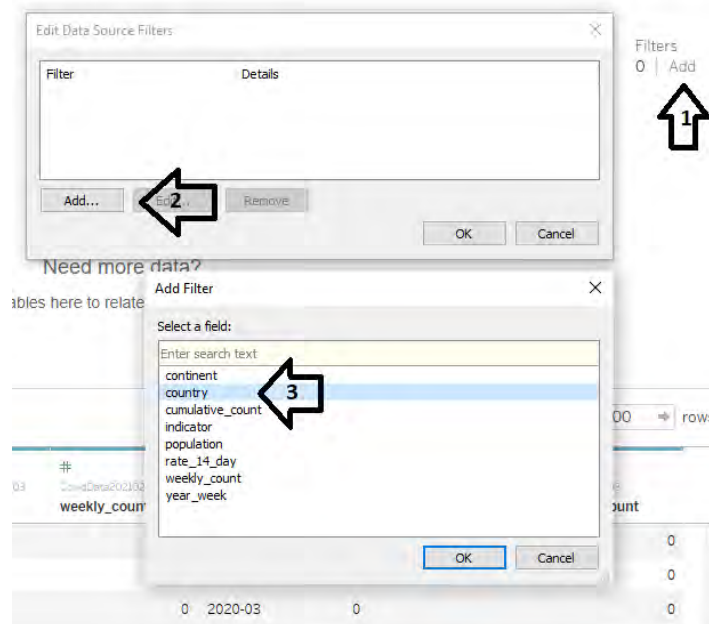


Figure 9: Add filtering

Filtering Activity:

- i. Cases and Deaths in France and UK for the 3rd week of 2021
 - Add filter by country: France and UK,
 - Add another filter by year_week: choose “2021-03”
 - How many rows do they appear?
 - Which column shows the cases and deaths up until that week?
 - Do you have the same results with the following photo? (Normally you should because the historic data in a dataset are not meant to be changed)

	Abc	#	Abc	#	Abc	Abc	#
	CovidData20210203	CovidData20210203	CovidData20210203	CovidData20210203	CovidData20210203	CovidData20210203	CovidData20210203
	country	continent	population	indicator	weekly_count	year_week	rate_14_day
France	Europe	67.012.883	cases	142.628	2021-03	403.446304496406	3.053.617
United Kingdom	Europe	66.647.112	cases	251.504	2021-03	862.924112900796	3.647.463
France	Europe	67.012.883	deaths	2.766	2021-03	79.0743475400096	73.049
United Kingdom	Europe	66.647.112	deaths	8.678	2021-03	247.692653209039	97.939

Figure 10: Filtering data of France and UK for year 2021 and 3rd week

- ii. Show deaths in European countries in descending order for the current week if you have the latest dataset, otherwise choose 3rd week of 2021.
 - You need 3 filters (by continent, by indicator and by year_week)
 - Order by cumulative count

continent	country	indicator	rate_14_day	year_week	cumulative_count		
Europe	Europe (total)	deaths	91.7347078894602	2021-03	701.9	COVID-19 Data.csv.cumulative_count	39.293
Europe	EU/EEA (total)	deaths	103.173552948714	2021-03	449.395		23.405
Europe	United Kingdom	deaths	247.692653209039	2021-03	97.939		8.678
Europe	Italy	deaths	111.100901918646	2021-03	85.461		3.284

Figure 11: Deaths in European countries from COVID-19 until 3rd week of 2021 in descending order

- In order to save the current changes, you made and avoid in the future many of the steps to format the raw data from the initial dataset you downloaded, you can export your current data to CSV file format “Data → Export Data to CSV → Choose name and location for your CSV file”.

Access and navigate datasets through Microsoft Excel

On the Web you can find a lot of tutorials on Microsoft Excel with a simple search on a search engine, YouTube or Vimeo.

- Download the .csv file from <https://www.ecdc.europa.eu/en/publications-data/data-national-14-day-notification-rate-covid-19>
- Open file from Excel and choose *All files (*.*)* if you cannot find the file. If Excel finds your file format not appropriate or says your file is corrupted, ignore it and click “Yes” to the question “Do you want to open it anyway?”
- A wizard will come up to prepare your data. Check in Step 1: “Delimited” and “My data has headers” → Next → Step 2: “Comma” → Next → Step 3: “Advanced” and choose the right settings to recognise numeric data in case in your country you use differently (concerns mainly column “rate_14_day” in our csv, confirm it’s correct representation), otherwise just click “Finish”.
- Rename the sheet you are working on, if you like, to “COVID-19 Data”. Right-Click (1) → Rename (2) (see the figure below)

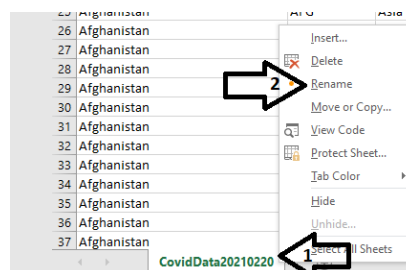


Figure 12: Rename Excel Sheet

- You can hide unneeded columns. We will hide columns “country_code” and “source”. Right-Click on top of column J (1) → Hide (2). Do the same for the column of “country_code”.

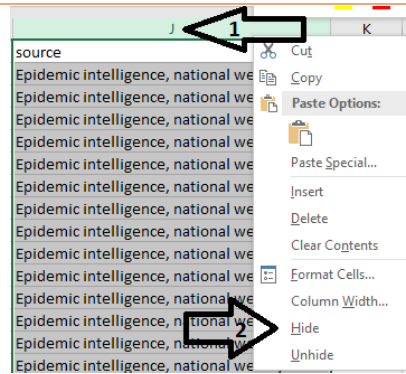


Figure 13: Hide Column J with header name "Source"

6. **Filtering data** in Excel. Let's filter the data that concerns the United Kingdom. First, one way to ensure that we don't lose any data while filtering is to click on the upper-left corner of the Excel sheet so all data is selected, see Figure.



Figure 14: Select all data

Then, click "Sort & Filter" → Filter.

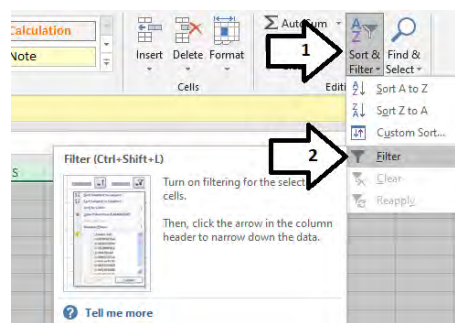


Figure 15: Excel Filtering

Click on "Country" → Select All (in order all countries to be deselected) → Scroll down and click "United Kingdom" → OK

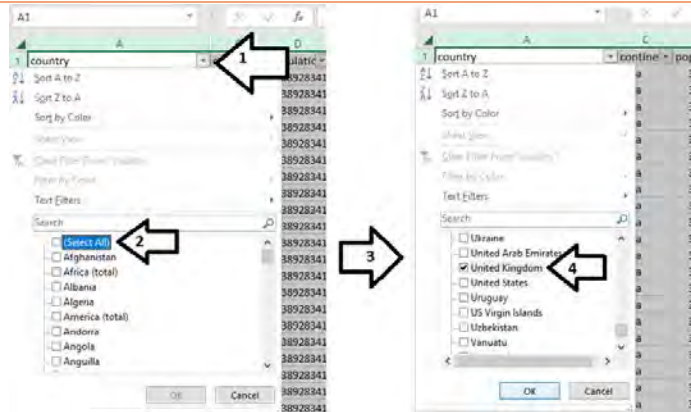


Figure 16: Filtering specific country

The result should be something like the following Figure. Specifically, the data in columns F and I, weekly_count and cumulative_count respectively, should be exactly the same because they are historical data not meant to be altered, except in the case they were not properly calculated in the first place.

	A	C	D	E	F	G	H	I
1	country	contine	populati	indicat	weekly_cou	year_we	rate_14_da	cumulative_cou
19998	United Kingdom	Europe	66647112	cases	0	2020-01		0
19999	United Kingdom	Europe	66647112	cases	0	2020-02		0
20000	United Kingdom	Europe	66647112	cases	0	2020-03		0
20001	United Kingdom	Europe	66647112	cases	0	2020-04		0
20002	United Kingdom	Europe	66647112	cases	2	2020-05	0.00300088	2
20003	United Kingdom	Europe	66647112	cases	6	2020-06	0.012003521	8
20004	United Kingdom	Europe	66647112	cases	2	2020-07	0.012003521	10
20005	United Kingdom	Europe	66647112	cases	1	2020-08	0.00450132	11
20006	United Kingdom	Europe	66647112	cases	36	2020-09	0.055516284	47
20007	United Kingdom	Europe	66647112	cases	365	2020-10	0.601676484	412
20008	United Kingdom	Europe	66647112	cases	2193	2020-11	3.838125799	2605
20009	United Kingdom	Europe	66647112	cases	6329	2020-12	12.78675061	8934
20010	United Kingdom	Europe	66647112	cases	17889	2020-13	36.33765856	26823
20011	United Kingdom	Europe	66647112	cases	30357	2020-14	72.3902335	57180
20012	United Kingdom	Europe	66647112	cases	32205	2020-15	93.87053411	89385
20013	United Kingdom	Europe	66647112	cases	32027	2020-16	96.37628909	121412
20014	United Kingdom	Europe	66647112	cases	32830	2020-17	97.31404416	154242
20015	United Kingdom	Europe	66647112	cases	31282	2020-18	96.19621627	185524
20016	United Kingdom	Europe	66647112	cases	22867	2020-19	81.24733147	208391
20017	United Kingdom	Europe	66647112	cases	19857	2020-20	64.10480322	228248
20018	United Kingdom	Europe	66647112	cases	16360	2020-21	54.34143943	244608
20019	United Kingdom	Europe	66647112	cases	10907	2020-22	40.91250045	255515
20020	United Kingdom	Europe	66647112	cases	8524	2020-23	29.15505176	264039
20021	United Kingdom	Europe	66647112	cases	7136	2020-24	23.49689211	271175
20022	United Kingdom	Europe	66647112	cases	6680	2020-25	20.73008055	277855
20023	United Kingdom	Europe	66647112	cases	5240	2020-26	17.8852461	283095
20024	United Kingdom	Europe	66647112	cases	4026	2020-27	13.90307805	287121
20025	United Kingdom	Europe	66647112	cases	4271	2020-28	12.44915159	291392
20026	United Kingdom	Europe	66647112	cases	4310	2020-29	12.87527658	295702
20027	United Kingdom	Europe	66647112	cases	4549	2020-30	13.29239893	300251
20028	United Kingdom	Europe	66647112	cases	4444	2020-31	13.4934579	304695
20029	United Kingdom	Europe	66647112	cases	6130	2020-32	15.86563371	310825
20030	United Kingdom	Europe	66647112	cases	7659	2020-33	20.68956866	318484
20031	United Kingdom	Europe	66647112	cases	7158	2020-34	22.2320211	325642
20032	United Kingdom	Europe	66647112	cases	8825	2020-35	23.98153426	334467
20033	United Kingdom	Europe	66647112	cases	12685	2020-36	32.27446675	347152

Figure 17: Result of filtering

Filtering Exercise:

- i. Cases and Deaths in France and UK for 3rd week of 2021
 - Add filter by country: France and UK,
 - Add another filter by year_week: choose "2021-03"
 - How many rows do they appear?



- Which column shows the cases and deaths up until that week?

Do you have the same results with the following photo? (Normally you should because the historic data in a dataset are not meant to be changed)

	A	B	C	D	E	F	G	H	I	K
1	country	country_code	continent	population	indicator	weekly_count	year_week	rate_14_day	cumulative_count	
7311	France	FRA	Europe	67012883	cases	142628	2021-03	403.4463045	3053617	
7367	France	FRA	Europe	67012883	deaths	2766	2021-03	79.07434754	73049	
20053	United Kingdom	GBR	Europe	66647112	cases	251504	2021-03	862.9241129	3647463	
20109	United Kingdom	GBR	Europe	66647112	deaths	8678	2021-03	247.6926532	97939	

Figure 18: Filtering data of France and UK for year 2021 and 3rd week

- ii. Show deaths in European countries in descending order for the current week if you have the latest dataset, otherwise choose 3rd week of 2021.
 - You need 3 filters (by continent, by indicator and by year_week)
 - Select all data (click upper-left corner) → “Sort & Filter” → Custom Sort... → Sort by “cumulative_count” and Order “Largest to Smallest” → OK

	A	B	C	D	E	F	G	H	I	K
1	country	country_code	continent	population	indicator	weekly_count	year_week	rate_14_day	cumulative_count	
319	Europe (total)		Europe	848075955	deaths	39293	2021-03	91.73470789	701991	
635	EU/EEA (total)		Europe	452548145	deaths	23405	2021-03	103.1735529	449395	
1109	United Kingdom	GBR	Europe	66647112	deaths	8678	2021-03	247.6926532	97939	
1533	Italy	ITA	Europe	60359546	deaths	3284	2021-03	111.1009019	85461	
1645	France	FRA	Europe	67012883	deaths	2766	2021-03	79.07434754	73049	
2145	Russia	RUS	Europe	145934460	deaths	3881	2021-03	52.38653023	69918	
2257	Spain	ESP	Europe	46937060	deaths	2439	2021-03	83.79306245	56208	
2883	Germany	DEU	Europe	83019213	deaths	5454	2021-03	137.3296564	52087	
3355	Poland	POL	Europe	37972812	deaths	1994	2021-03	108.9463693	35401	
4979	Turkey	TUR	Europe	82003882	deaths	1076	2021-03	27.63283816	25073	
5249	Ukraine	UKR	Europe	43733759	deaths	1055	2021-03	47.76630337	21924	
5361	Belgium	BEL	Europe	11455519	deaths	368	2021-03	62.06615344	20837	
5565	Romania	ROU	Europe	19414458	deaths	570	2021-03	57.4829336	17841	
6465	Czechia	CZE	Europe	10649800	deaths	1004	2021-03	204.7925783	15453	
6763	Netherlands	NLD	Europe	17282163	deaths	577	2021-03	67.57626877	13564	

Figure 19: Deaths in European countries from COVID-19 until 3rd week of 2021 in descending order

7. In order to save the current changes, you have made and avoid in the future many of the steps to format the raw data from the initial dataset you downloaded, you should save your data as an “.xlsx” file.

Reflection

Dealing with open data and open datasets is not an easy thing to do. Digital skills over using spreadsheet software are useful to access and manipulate open datasets. The above activity is a small paradigm of how you can do that for a real-life problem. This activity could include more steps and be more exhaustive but such an activity is beyond the scopes of the course. It’s main purpose is for the reader to start experimenting with open-data, stop feeling “absolute fear” about them and with some practice to manage to use some simple datasets for his educational purposes.



TOOLS DATA & RESOURCES NEEDED

- Web Browser
- Tableau Public
- Microsoft Excel

TIME REQUIRED

- 20 minutes for each Tableau/Excel depending on the experience in the use of the software.