Co-funded by the Erasmus+ Programme of the European Union

2019-1-BE02-KA201-060212

MODULE 1

INFORMATION &

DATA LITERACY

Manipulating your dataset

Expert Level Activities

Advanced use of tools for manipulating datasets

DESCRIPTION OF THE ACTIVITY

The raw data in the datasets are separated by a delimiter, usually a comma, semicolon, tab or something similar. In order for these datasets to be retrieved, organized and stored more easily a proper manipulation must take place. It is often the case that the data has been under some pre-processing and reformatting by the authority that publishes them, so in that case this dataset must be examined, if it is in a situation that can be manipulated in the preferred way and if not, proper reformatting has to be made.

A visualisation paradigm of a dataset in Microsoft Excel.

A simple methodology will be followed in our example for the preparation of the data.

1st step: Clean and format the data
2st step: Create a Pivot Table and Insert Slicers for filtering
3st step: Create the preferred graph in a new sheet with all the slicers in it

Our purpose is to use a dataset of worldwide cases and deaths concerning the Covid-19 pandemic. We are going to keep our example as simple and clear as possible. Implement the following steps for the creation of the visualisation.

1. Download .csv file from https://www.ecdc.europa.eu/en/publications-data/data-national-14-day-notification-rate-covid-19
2. 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?”
3. 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”.
4. Rename the sheet you are working on, if you like, to “COVID-19 Data”. Right-Click (1) → Rename (2) (see the figure below)

This project has been funded with support from the European Commission. This publication reflects the views only of the author(s), and the Commission cannot be held responsible for any use which may be made of the information contained therein.

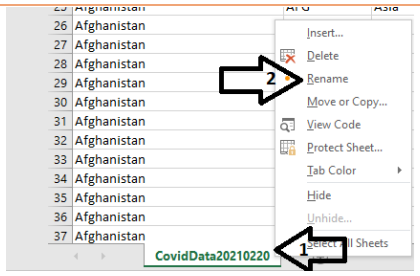

Figure 20: Rename Excel Sheet

5. 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 “country_code”.

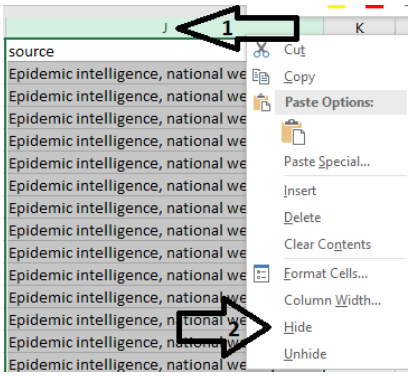

Figure 21: Hide Column J with header name "Source"

6. Delete all the rows in column “country” where the values are equal to the following bullets. Using the Filter (Ctrl + Shift + L) or the Find (Ctrl+F) utilities can make your life easier in this task. Be careful not to leave any empty lines on the table.
   - Africa (total)
   - America (total)
   - Asia (total)
   - EU/EEA (total)
   - Europe (total)
   - Oceania (total)

   We deleted those rows because a Pivot Table does the aggregations automatically for the continents and having the already calculated data from the dataset would produce false results in our analysis (double values particularly in our case).

   Obviously, if you had to deal with another dataset or with another problem using the same dataset, you probably would have chosen to do another formatting or cleaning of data.
7. Select all the data of the table. One way is to “left click” from top of column A and drag-and-drop up to column I.

This project has been funded with support from the European Commission. This publication reflects the views only of the author(s), and the Commission cannot be held responsible for any use which may be made of the information contained therein.

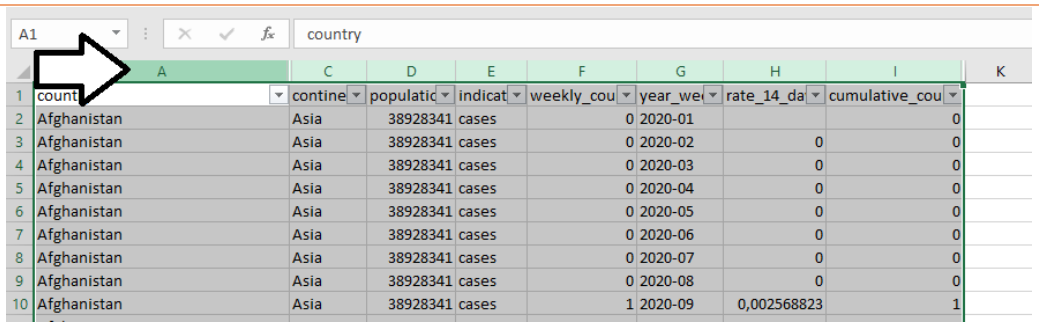

Figure 22: Select all the data of the table

8. Click "Insert → Pivot Table → OK" and the fields of the Pivot Table are going to be imported to a new sheet.
9. Go to the new sheet and drag-and-drop from the field area the following fields
   - year_week to "Rows"
   - weekly_count to "Values"

   Observe the table that is produced. It presents Cases+Deaths per Week.
10. Click "Analyze → Insert Slicer" and click on "country". A slicer is produced that will function as a filter on the presented data. Click on the slicer and from the ribbon "Options" (1) configure your slicer, for example "Columns (2) → (set it to) 6 (or as many as you like that can fit your view area)". Then Right-Click on Slicer → Slicer Settings… → "Hide items with no data".

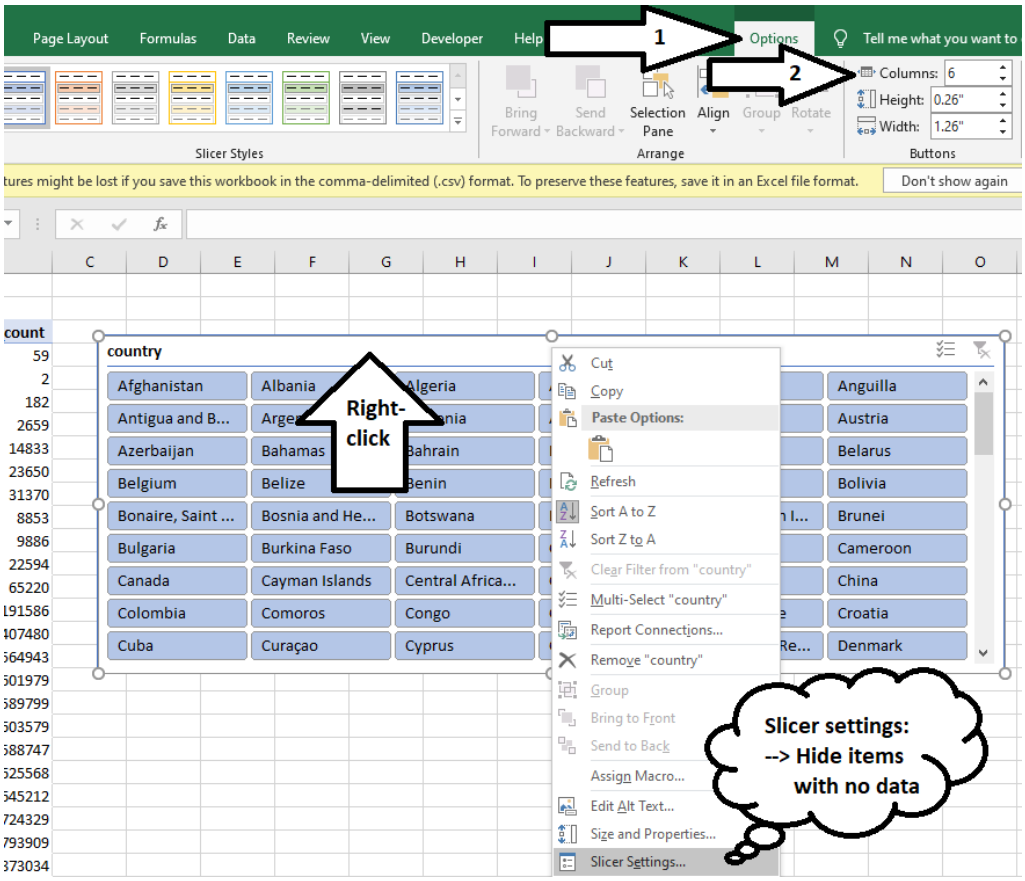

Figure 23: Slicer Configuration

Click on individual countries, for example in the figure below on "Belgium" (1), to see that the table shows the corresponding data. You can make "multiple choices" and "clean the filter" as you can see in the Figure.

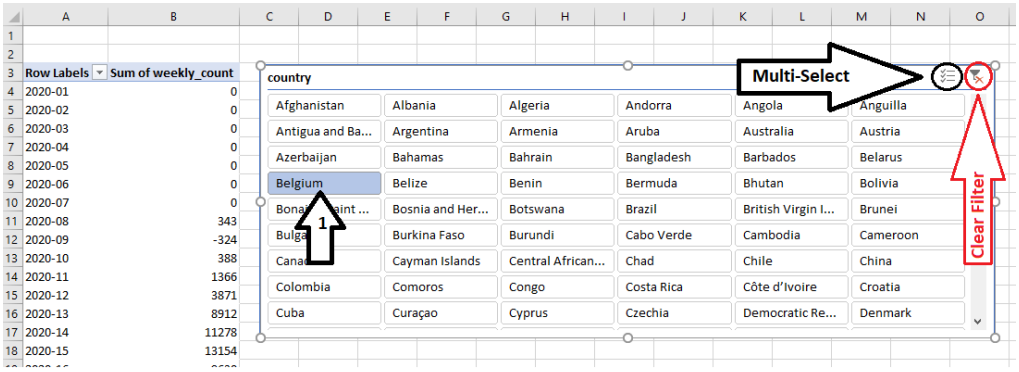

Figure 23: Filtering with Slicer

Create in the same manner the following slicers and make the appropriate configurations as previously:

- continent
- indicator
- year_week

11. Click on the table and then on the ribbon "Insert → Insert Column or Bar Chart → 2-D Column → Clustered Column"

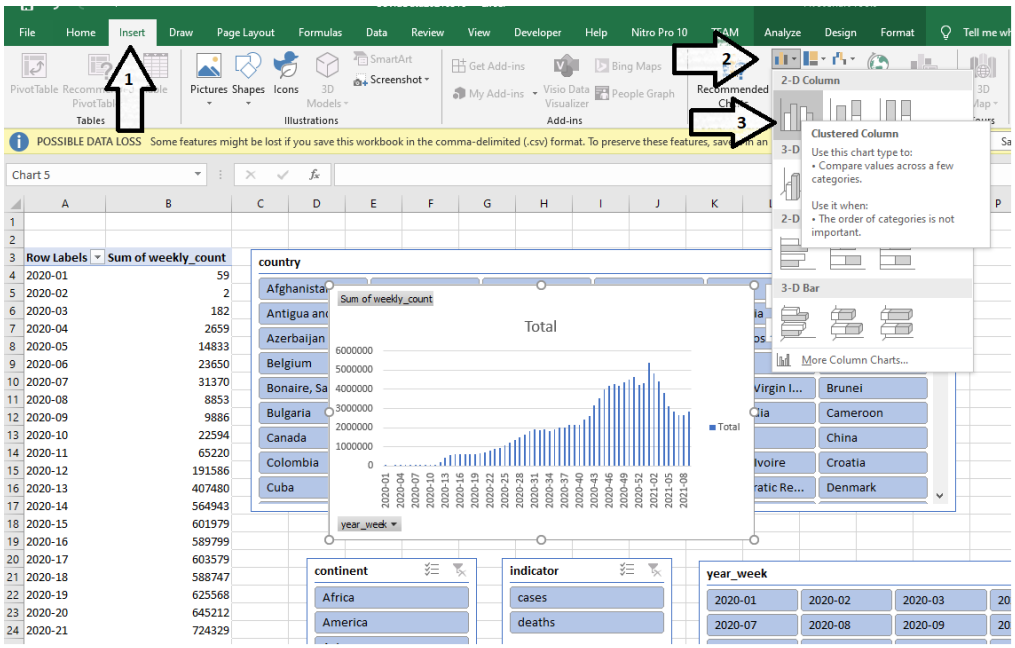

Figure 25: 2D Column Graph

12. Create a new sheet and give the name "Vis – Covid-19 Worldwide"
13. Cut and Paste (transfer) in the new sheet the Graph and the Slicers and arrange them to fit nicely.

This project has been funded with support from the European Commission. This publication reflects the views only of the author(s), and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Figure 26: "Vis - Covid-19 Worldwide" in Action

14. You can make some formatting of your choice in the graph. Experiment in
    - Changing the title to “Worldwide Cases/Deaths from COVID-19 per week”
    - Changing background colour
    - Deleting unnecessary labels and fields
    - Changing the unit of the vertical axis to millions let’s say
15. Select and deselect choices from the four slicers to see the updates on the interactive graph.
16. Experiment with the creation of other Pivot Tables and Graphs.
17. Save it in .xlsx format to be sure that all your formatting, manipulation, graphs and work in general do not get lost.

Reflection and Activities

- Download other datasets of your interest in “.csv” or other formats compatible to Excel and experiment.
- Experiment more with the Pivot Table dragging and dropping fields in the four areas of the Pivot Table (Rows, Values, Filters, Columns).
- Study some tutorials on Excel to become more fluent with Pivot Tables and Graphs
    - https://libguides.com.edu/c.php?g=649573&p=7558211
    - https://uhlibraries.pressbooks.pub/mis3300excel/chapter/6-1-creating-pivot-tables/
    - Search the web for open content (https://www.google.com/advanced_search)
        - “Excel tutorial” or “excel pivot tables”
        - Usage rights → Free to use or share (or “not filtered by licence” if you want it only for your own studying)

This project has been funded with support from the European Commission. This publication reflects the views only of the author(s), and the Commission cannot be held responsible for any use which may be made of the information contained therein.

TOOLS DATA & RESOURCES NEEDED
• Web Browser • Microsoft Excel
TIME REQUIRED
• 40 minutes approximately

This project has been funded with support from the European Commission. This publication reflects the views only of the author(s), and the Commission cannot be held responsible for any use which may be made of the information contained therein.