

REACH Add-in for MS Excel

A Guideline for Excel-based Data Processing

Developed By REACH Afghanistan

version 2.4 (August 2024)

Contents

Installation.....	1
Installation steps	1
Data cleaning.....	2
Setting	2
Download Audit File	3
Time Checking	3
Stop Process	3
Clear Filter	3
Show Empty Columns.....	3
Check Duplicate	3
Consistency Check.....	3
Add Label.....	4
Setup Logical Check.....	4
Import and export logical checks	5
Running logical checks.....	5
Add To Logs	5
Detect Outliers	6
Logs option	6
Replace Logs	7
Data Analysis	8
Weighting	8
Disaggregation levels.....	9
Analysis List	10
Analyze	11
Chart.....	11
Indicator finder.....	12
General Tips.....	13

Installation

Installation steps

REACH utility is a Microsoft Excel add-in that helps you to clean and analyze your data with ease. With this utility, you can perform common data cleaning tasks such as checking the interview duration, applying logical checks, finding missing data, and spotting outliers. You can also create summaries and visualizations of your data with just a few clicks. To install REACH utility, please follow these steps:

1. Download the package and extract the files. You will see two files: REACH add-in and install.
2. Double-click on install and wait for the success message to appear. (Make sure MS Excel application is close)

Now you can see a new tab in your Excel application. (See Figure 1)

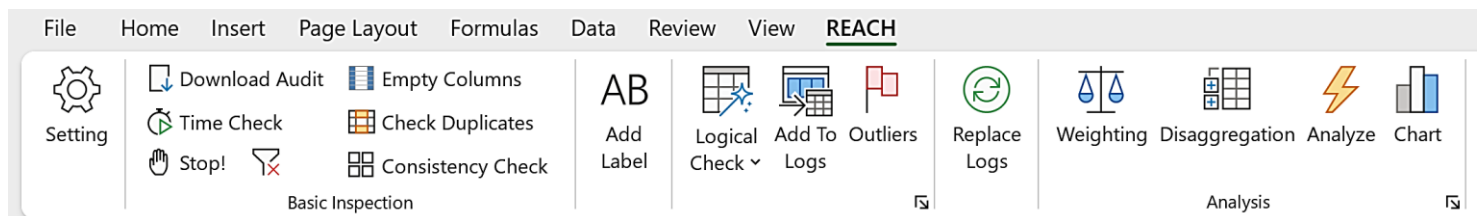


Figure 1, REACH tab

Data cleaning



Before we begin, it's important to note that this utility is specifically designed to handle KOBO data in **XLM format**. If your dataset uses labels instead of XML values, this application may not be suitable for your needs. Additionally, please ensure that your dataset includes a '_uuid' column for optimal functionality and in case your dataset is in '.csv' please convert it to the '.xlsx' file format. It is highly recommended to avoid opening other excel file during data cleaning or analysis using REACH utility.

Setting

In the first step, you need to open your KOBO data in Excel and switch to the REACH tab. In this tab, you will find the basic inspection section, which allows you to calculate the interview duration using the KOBO audit files, identify duplicate records, and add labels for better readability. To use this section, you need to configure the system first by clicking on Settings at the left corner. (See figure 2)

The screenshot shows a 'Setting' dialog box with the following components:

- Info section:** Contains text about Kobo information, Kobo Tool necessity, sheet name constraints, and the integrated tool path. A red circle with the number 6 is next to the file path.
- Version:** 'version 2.3' at the bottom left.
- Input fields:**
 - KOBO user name:** A text input field with a red circle 1 next to it.
 - KOBO password:** A text input field with a red circle 2 next to it.
 - Main Data:** A dropdown menu showing 'Sheet1' with a red circle 3 next to it.
 - Audit URL:** A text input field with a red circle 4 next to it.
- Buttons:**
 - Import Tools:** A button with a red circle 5 next to it.
 - Reset Setting:** A button with a red circle 8 next to it.
 - Save:** A button with a red circle 7 next to it.

Figure 2, Setting form.

To configure the utility, you need to fill in the following information:

1. Your KOBO username if you want to download the audit file using REACH utility. If you don't need the audit file, you can skip this step.
2. Your KOBO password if you entered your username in the previous step.

3. The name of your data sheet. Make sure the name is shorter than 15 characters and does not contain spaces or special characters like @ and \$.
4. The option that matches your data's audit URL, if you have one. If you don't need to calculate the interview duration, you can leave this blank.
5. The Kobo tool that corresponds to your data. You need to import the tool to add labels and perform data analysis.
6. Show the imported Kobo tool.
7. Click Save to complete the configuration.
8. Click Reset to clear the tool and all the information you entered.

Download Audit File

Click on Download Audit File. The utility will begin to download the audit files from the Kobo server. This may take some time, depending on the size of your data and your internet speed. You can see the number of downloaded files in the Excel status bar at the bottom of the application. Make sure your data set has a “_uuid” column for this to work.

Note: If it is possible to download the audit files directly from Kobo server, it would be a better option, but if the tool has photo or other media types, then downloading audit files can be challenging directly from the Kobo server.

Time Checking

To check the interview duration, click on this option. A progress window will appear and add two columns at the end of your data set. The duration column will show the total time spent on each interview in minutes, and the duration remarks column will indicate if the audit file is available or not. If there is no audit file, the utility will use the start and end values for the interview duration. You can then remove or flag the data that falls outside of a predefined time range, according to your assessment criteria. Note: For this step, your data set must have a 'start' and 'end' column, otherwise the utility cannot calculate the interview duration.

Stop Process

You can cancel all actions like downloading audit file by clicking on the stop button.

Clear Filter

By clicking on this button, all active filters will be disabled.

Show Empty Columns

By clicking on this button, a list of empty columns appears. This is handy when we want to check if the relevancy in the tool is working fine or not.

Check Duplicate

To check for duplicate data, click on this button. The utility will scan the “_uuid” column and show the result at the end of your data set. If you find any duplicate records, make sure they are the same, then you can delete one of them.

Consistency Check

It is a good practice to check the categorical values in the data. By clicking this option all the select_one data type will be checked and in case of finding any discrepancies a logbook will be generated automatically. This process is time consuming and may take a couple of minutes to accomplish.

Add Label

Sometimes, your data may contain coded values that are easy to process and analyze, but hard to read and understand. To make your data more readable, you can use this button to add labels in a new column next to the coded values.

Setup Logical Check

A useful way to check your data quality is to compare different sections and see if their values are matching and consistent. The system offers a feature that allows you to perform logical checks between two columns or within a single column. (See figure 3)

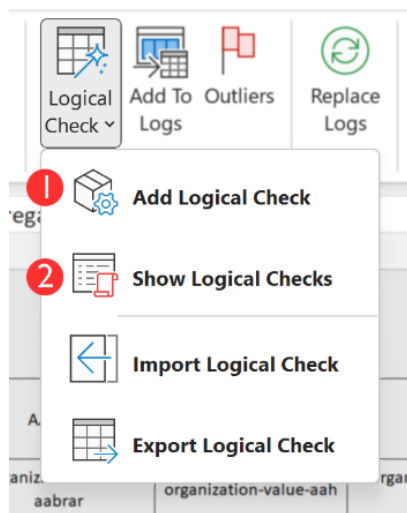


Figure 3, Set logical checks

By clicking Set Logical Checks a new window appears. You can write a rule based on one or two columns. (See figure 4)

A screenshot of a dialog box titled 'Logical Data Cleaning'. It contains a section 'Define a rule for logical data cleaning:' with two rows of dropdown menus. The first row has 'respondent_age', 'is less than', and '18'. The second row has 'respondent_age', 'is greater than', and '75'. Below this are radio buttons for 'simple check', 'and', and 'or', with 'or' selected. A 'Message:' section contains a text box with the text 'respondant age is not acceptable.' At the bottom right are 'Cancel' and 'Save' buttons. Red circles with numbers 1, 2, 3, and 4 highlight the first dropdown, the message text box, the 'Save' button, and the 'Cancel' button respectively.

Figure 4, Adding logical checks.

- 1) In this section fill the checking role
- 2) In the message section we can put the desired flagging message
- 3) by clicking Save button the checking role will be stored.

4) by clicking Cancel button the window will be closed.

2. By clicking the Show Logical Checks, you can see the list of checking roles. (See figure 5)

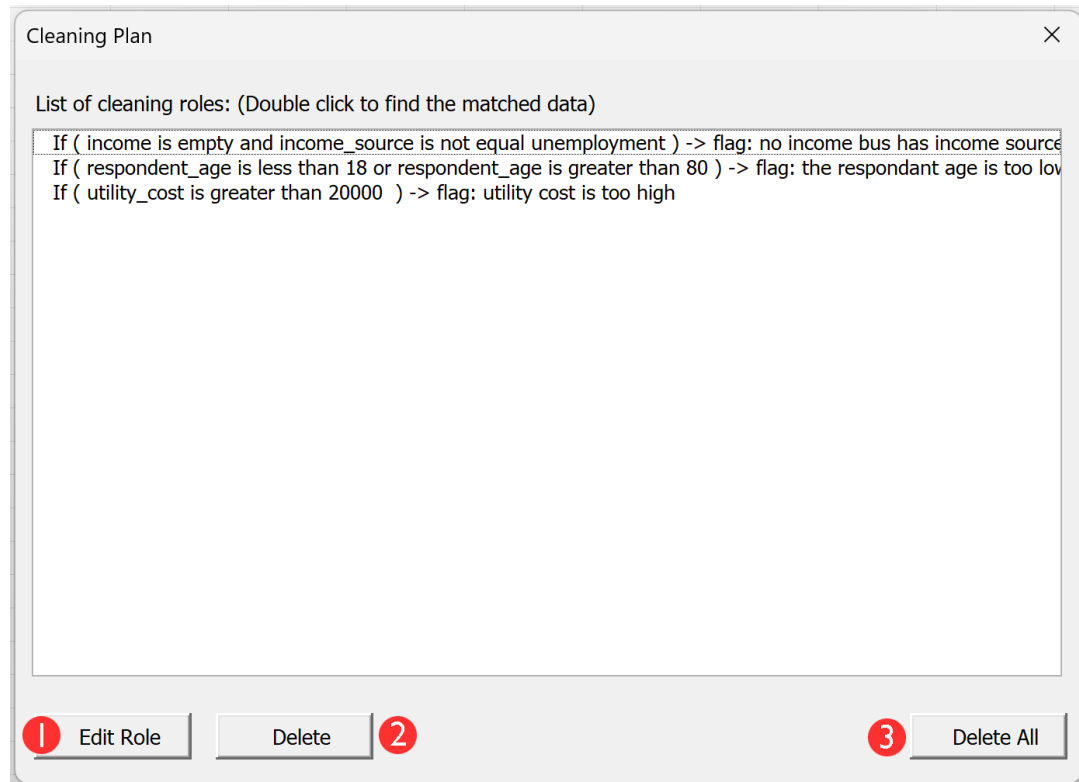


Figure 5, Show logical checks.

2.1) Select a role then click edit role for modifying the role.

2.2 by clicking this button the selected role will be deleted.

2.3) by clicking this button all roles will be deleted.

Import and export logical checks

To save time, you can export your logical checks and share it with others or import it later into another file when you need to use them again.

Running logical checks

To apply your logical checks to your data, click on the logical check button. The utility will scan your data and look for any patterns that match your checks. If it finds any, it will create a logbook for you.

Add To Logs

If you need to check a specific part of the data manually, for example filtering missed income and selecting all the empty cells in the income column, you can flag them into the logbook by clicking on the Add to logs.

Detect Outliers

To detect outliers in your data, follow these steps:

- Select a numerical column in your data set.
- Click on the Detect Outliers button. A pop-up message will show you the minimum and maximum values based on the IQR method.
- The utility will filter your data and show you the outliers in the selected column. You can flag them into the logbook if you want.

Logs option

If you want to add more information to your logbook or check the logbook, you can click on the corner button below the outliers. (See figure 6)

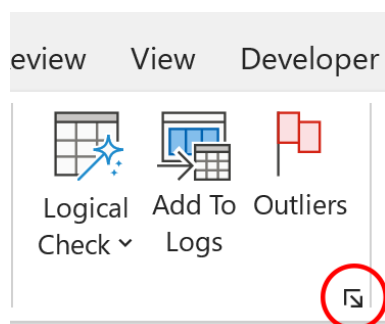


Figure 6, Logs option

A new window will be appeared that allows you to include any relevant details and check the duplicated logs. (See figure 7)

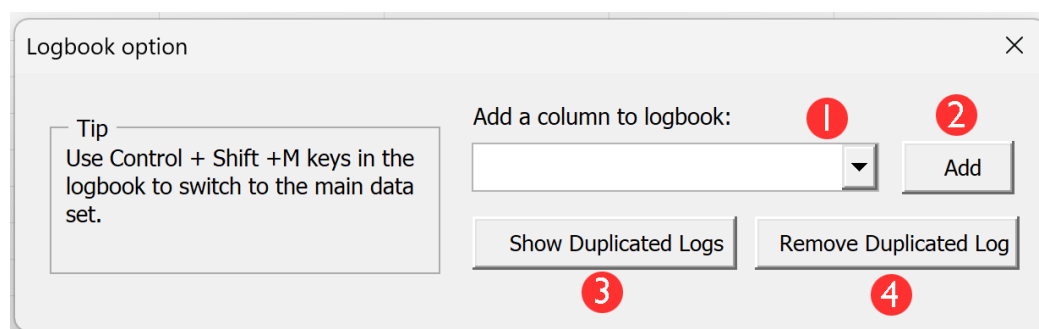


Figure 7, Logs option

1. Select the column you want to add.
2. Click Add bouton to add the column to the logbook.
3. By clicking this option, the system checks the duplicated logs. If there are any duplicated logs it shows as a highlighted value.
4. If there are any duplicated logs, all of them will be removed by clicking this button.

Tip: When you are in the logbook you can select an issue cell, then press **Control + Shift + M** simultaneously to go to the main dataset cell that has the issue.

Replace Logs

To update your data with the changes you made in the logbook, click on this button. The utility will replace the old values with the new ones in the main data set. Make sure the logbook has a column named “changed” and it has a “yes” value for the rows that you want to update.

Data Analysis

Data analysis contains three steps:

- Weighting (optional)
- Setup the disaggregation levels
- Set up the list of indicators or questions.

Weighting

Weighting is an optional step for data analysis. If you don't need weighting in your analysis, skip this step and go to the disaggregation part.

To adjust your analysis according to the sampling framework, it's essential to integrate the sampling details into your dataset. If your sampling information is stored in a separate file, you'll need to copy the 'strata' and 'population' columns and paste them in the top-left cell (A1) of a new sheet within your clean data file. (See figure 8 as an example)

	A	B
1	Settlement_code	hh_population
2	area_01	20000
3	area_02	34000
4	area_03	8000
5	area_04	19700
6	area_05	47500
7	area_06	12000
8		
9		
10		
11		
12		
13		
14		
15		
16		

< > clean_data sampling

Figure 8, sampling information

Then you can click on the weighting button. (See figure 9)

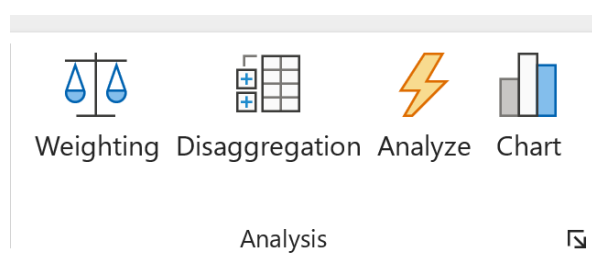


Figure 9, Analysis utilities

After clicking the weighting, the below window will appear. (See figure 10)

The image shows a 'Weighting Setting' dialog box with a close button (X) in the top right corner. On the left, there is a 'Note' section with the text: 'Please make sure the sampling information is available. In the sampling information two columns are necessary, "strata" and "population".' On the right, there are five dropdown menus and two buttons, each with a red circular callout number: 1. 'sampling sheet:' dropdown with 'sampling' selected. 2. 'sampling strata:' dropdown with 'Settlement_code' selected. 3. 'sampling population:' dropdown with 'hh_population' selected. 4. 'main sheet:' dropdown with 'data' selected. 5. 'strata in main sheet:' dropdown with 'village' selected. 6. 'Test Strata' button. 7. 'Add Weight' button.

Figure 10. weighting options

1. Select the sheet that contains the sampling information.
2. Select the column that contains the strata in the sampling sheet. Each stratum should have a unique value in the sampling frame.
3. Select the column that contains the population size for each stratum in the sampling sheet.
4. Select the sheet that contains the data that you want to analyze.
5. Select the column that contains the strata in the data sheet.
6. Click on this button to check if all the strata in your data are present in the sampling frame. If there are any missing or mismatched strata, the utility will show you a list of them.
7. Click on the Add Weight button to create a new column at the end of your data sheet and sampling sheet. This column will contain the weights for each record in your data. Do not rename this column, as it will be used for data analysis.

Disaggregation levels

Please make sure you have already imported the relevant Kobo tool into the REACH utility first then you can set the disaggregation level by clicking on the Disaggregation button. A new window appears. (See figure 11)

Analysis Disaggregations

Tips

Make sure you have a weight column in the main data set if you need weighted analysis.

Disaggregation levels are usually based on geographic divisions and other categorical indicators.

Since producing analysed information is a time-consuming process, please avoid adding too many disaggregation levels.

Integrated Tool:
C:\Users\Mohammad AZIMI\Desktop\test45\ram_tool.xlsx

Select Clean Data:

data

Select Disaggregation Level:

Weight:

Disaggregation Level	Weight
ALL	no
province	no

☐ Add all questions for analysis

Remove Save

Figure 11, add disaggregation levels.

1. Select the sheet that contains the data that you want to analyze.
2. Select the column that you want to use for disaggregation and the option that you want to use for weighting.
3. If you need to apply weighting, select “yes” otherwise select “no”
4. Click on the + button to add the selected column and option to the list window.
5. If you want to add all the questions from the tools to the analysis result, select the check box.
6. Click on the Save button to save your settings.
7. Click on the Remove button to clear the disaggregation list.

Analysis List

You also need to list all desired indicators or list of questions in the analysis_list sheet. (See figure 12)

	A	B
1	question	type
2	income	integer
3	debt_amount	integer
4	hhh_gender_calc	calculate
5	assist_satisf	select_one
6	repairs_unable	select_multiple
7		
8		
9		
10		
11		
12		
13		

< > ram sampling analysis_list +

Figure12, Analysis List

In the first column enter the question name and in the second column use “=question_type(A2)” formula to get the question type. Please note that only integer, decimal, select_one and select_multiple data types are acceptable, other data types will be ignored during data analysis. You can manually set an indicator as select_one or integer if you wish.

Analyze

To perform the data analysis, click on the Run Analysis button. The system will show you a progress report as it calculates the indicators and disaggregates the data. Please do not open other excel file during analysis. (See figure 13)

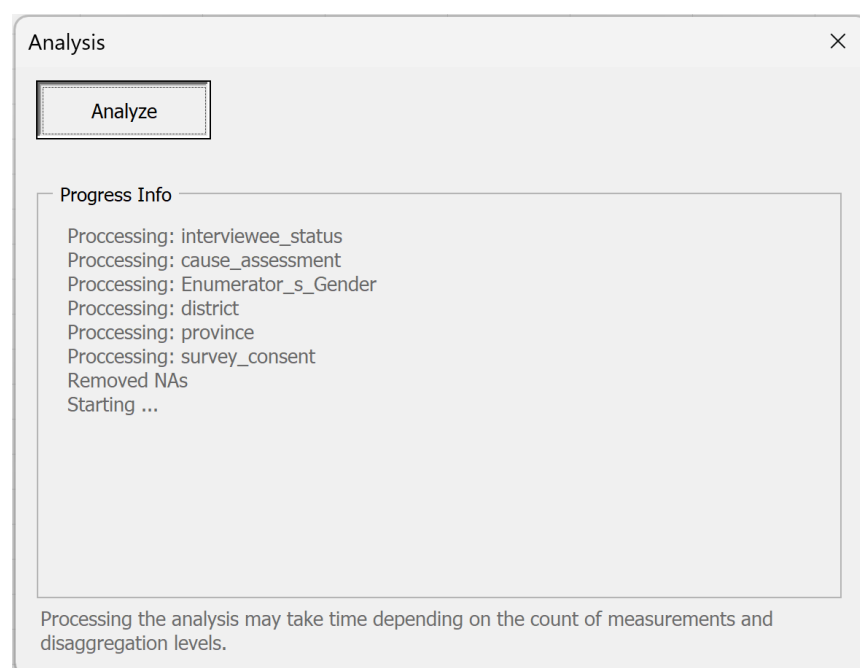


Figure 13, Analysis progress

When the analysis is done, the utility will create two new sheets: “result” and “datamerge”. These sheets have the same content, but in different formats. Please note that the result sheet is hidden.

Chart

This feature simplifies the process of generating charts tailored to your needs. When utilizing the "ALL" disaggregation level for analysis, choosing the first option automatically generates separate tables and charts for each measurement in your dataset. Additionally, you have the flexibility to create charts based on other disaggregated data or to generate a chart specifically for a chosen indicator. (See the figure 14)

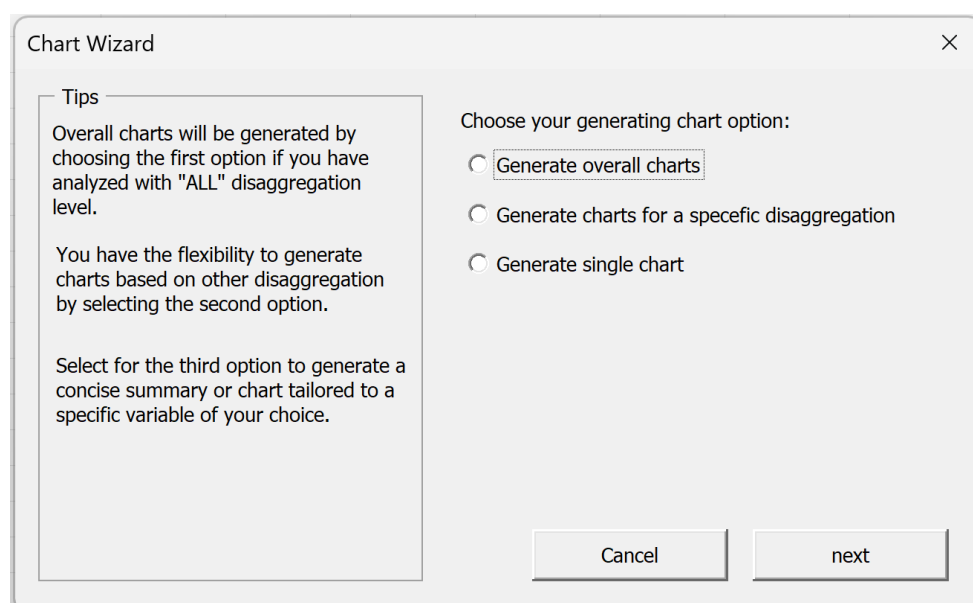


Figure 14, Chart Wizard

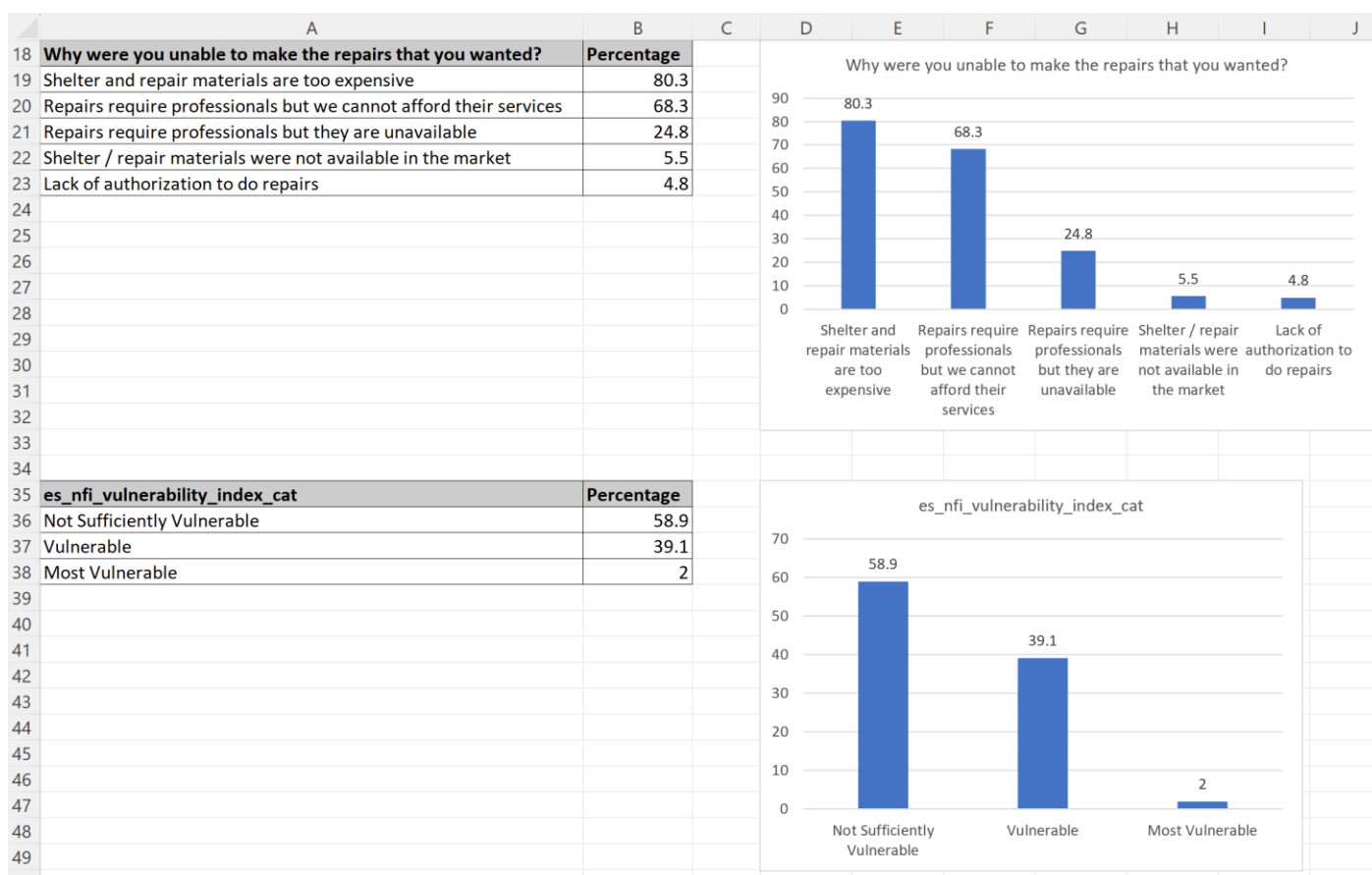


Figure 15, Sample of overall figures.

Indicator finder

If you have lots of indicators in the datamerge, it would be difficult to find a specific indicator, for such reason you can use the finder utility by clicking on the corner button. (See the figure 16)

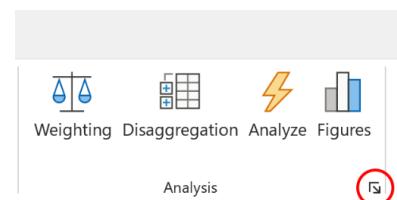


Figure 16, Finder

In the finder, you can double click on the desired indicators to find it in the datamerge or in the overall sheet. (See the figure 17)

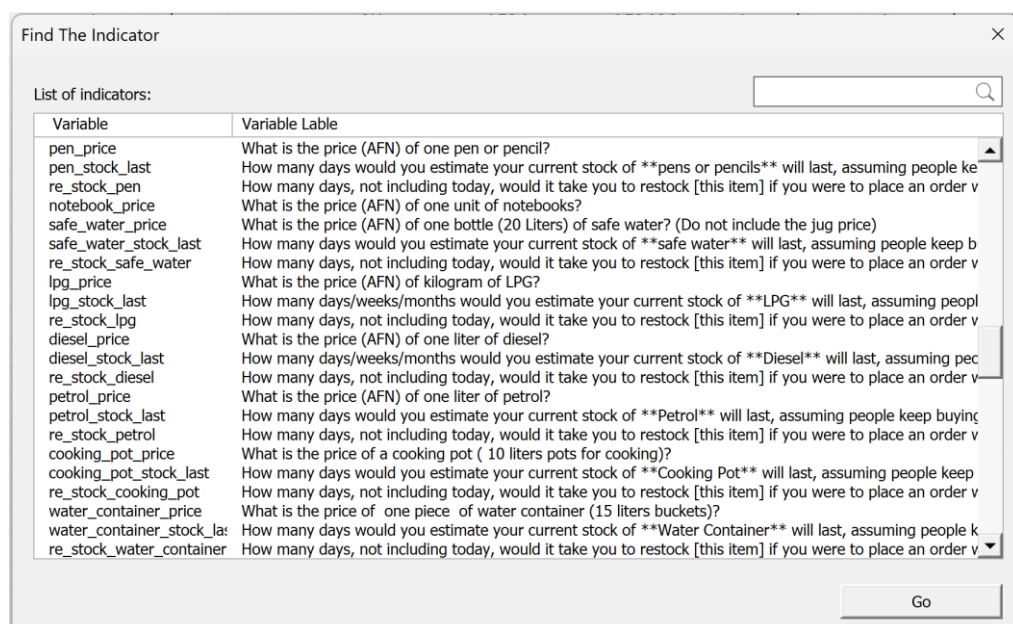


Figure 17, Finder

General Tips

- Please make sure you have installed MS Office 2016 (64 bit) or higher before using the REACH add-in.
- Please make sure your dataset is in XLSX format.
- The data from KOBO server can be in various formats. Please use the XML value and header format instead of labels in the tools.
- Please avoid renaming the header (in the first row) in your dataset.
- Make sure your main dataset has a “_uuid” column.
- Import the relevant KOBO tool before data cleaning and data analysis.
- The KOBO tool should have “label” or “label::English” columns.
- Do not add empty columns or rows in the dataset.
- During data analysis or data cleaning do not open multiple Excel files.
- If you face any trouble, click on the Stop button, also you can reset the system by going to the setting.
- For removing the REACH utility from your MS Excel application, please follow these steps:
 1. Open Excel and go to Options > Add-ins > Go. A pop-up window will appear. (See Figure 18 and Figure 19)
 2. Uncheck the REACH_Addin in Figure 19 for removing REACH tab from the Excel application.

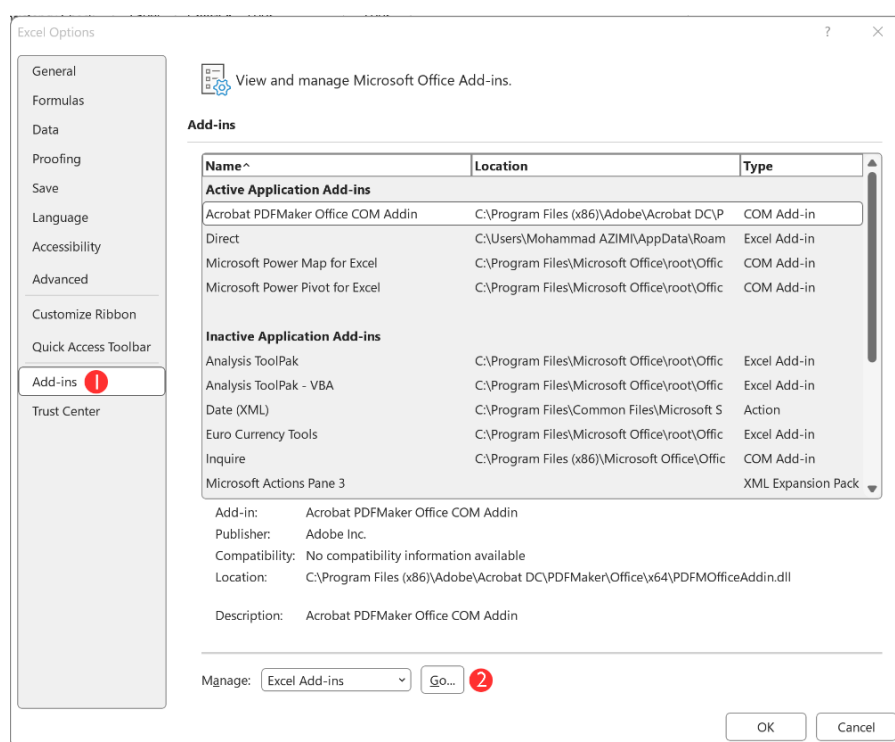


Figure 18, Add new Add-ins.

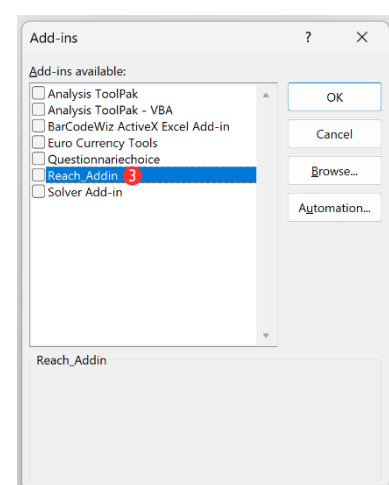


Figure 19, Uncheck marks the REACH Add-in.