# MIP Quality Control Tool Quick Guide (for version 0.0.3)

## mipqctool as python script

**Requirements**

* Python 3 with pip
* Installed version of Perl (needed for pdf export)
* Installed LaTex compiler (needed for pdf export)

For windows systems the LaTex Compiler can be downloaded from here: <https://miktex.org/download>

And the Perl from here: <https://www.perl.org/get.html>

**Installation**

Clone the github repository <https://github.com/aueb-wim/DataQualityControlTool/tree/master/qctool>

For Linux OS we run the install.sh script where the setup.py is located

For Windows OS in command prompt we run in the folder where the setup.py is located the command:

pip install -e .

**Usage**

For csv data we run the command

qctool --input\_csv *[dataset csv path]* --meta\_csv *[metadata csv path]* --col\_val *[metadata column name for variable codes/names]* --col\_type *[metadata column name for variable types]*

“col\_val“ and “col\_type” are obligatory. They are referred to columns names of the metadata csv. The “col\_val” is the name of the column that contains the variables codes used as columns in the datatset csv and the “col\_type” is the name of the column in metadata csv that contains the variables types.

At the moment the tool needs the metadata file to detect the nominal variables. So, the “col\_type” column must be filled with the value “nominal” (is not case sensitive) for the categorical variables in order to work properly. Other types like “int”, “float”, “text”, “numerical”, “date” are not taken into account at the moment.

After the execution, three files will be produced:

* a csv file <dataset\_file> + ‘\_dataset\_report.csv’ containing the Statistical Report of the given dataset.
* A csv file <dataset\_file> + ‘\_report.csv’ containing the Statistical Reports of the variables of the given dataset.
* A pdf file <dataset\_file>+’\_report.pdf’ containg the above two reports in a readable pdf format.

Notes: For CLM hospital the file Variables\_CLM\_2018\_11\_02\_annotated.xlsx can be used as a metadata file if it is converted to csv format first. Then use as col\_val *VariableCode* and col\_type *Type*.

## mipqctool as standalone windows executable

**Requirements**

* 64bit Windows OS
* Installed version of Perl (see previous section)
* Installed LaTex compiler (see previous section)

**Installation**

Download the windows executable qctab.exe from:

<https://github.com/aueb-wim/DataQualityControlTool/tree/master/windows_executables>

Optionally, we can add the folder where the qctab.exe is located to the system’s PATH variable in order to run the application globally.

**Usage**

For csv data we run in the command prompt:

qctab --input\_csv *[dataset csv path]* --meta\_csv *[metadata csv path]* --col\_val *[metadata column name for variable codes/names]* --col\_type *[metadata column name for variable types]*

The arguments are the same with the ones that are described above (python script execution).

Notes: This version of qctab.exe is tested in Windows 7 and Windows 10