# MIP Quality Control Tool Quick Guide (for version 0.0.2)

## Running qctool as python script

Requirements for running the Quality Control tool in command line

* Python 3
* Pandas
* Numpy

Download the qctab.py from github repository <https://github.com/aueb-wim/DataQualityControlTool/tree/master/qctool>

For csv data we run the command

Python qctab.py --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.

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*.

## Running qctab.exe standalone windows executable

Download the windows executable qctab.exe from:

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

For csv data we run the command

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