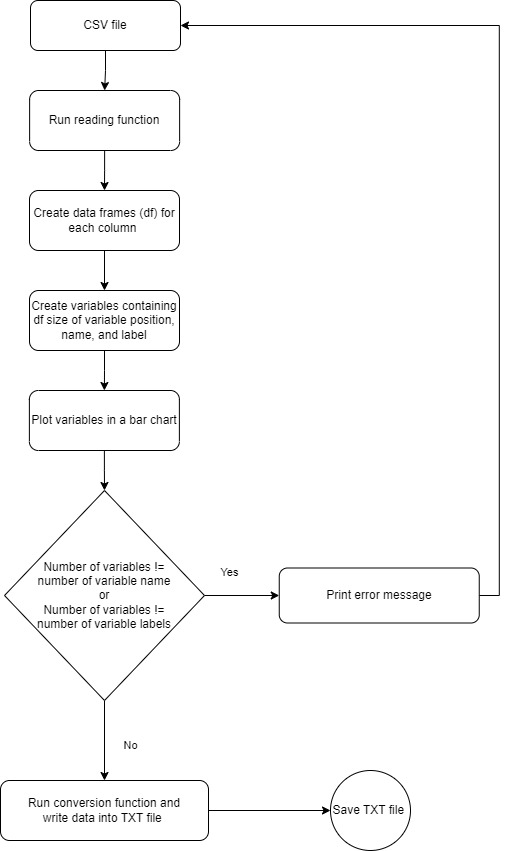
# Programming a CSV to TXT codebook converter with Python

## Requirements

* Reading a CSV file containing all variables of a given dataset and describing these variables with regards to the following attributes: name, label, valid values, and corresponding value labels as well as missing values and corresponding value labels (INPUT).
* Converting the CSV file into a TXT file, wherein the attributes belonging to a given variable are presented in accordance with a predefined row-based schema. In particular, there needs to be a text block for each variable starting with the variable name in the first row, followed by the variable label in the second, the set of all valid and missing values with the corresponding value labels in the remaining rows (OUTPUT).
* There needs to be a function reading the CSV file and plotting number of variables and the number of distinct values for the name and label attribute into a bar chart. If number of variables and number of distinct values for the name and label attribute are unequal, there needs to be an error message asking users to check their CSV file for duplets regarding the name and label column and revise the CSV file accordingly (Test case 1).
* If the CSV file is correct there needs to be a loop for a conversion function reading each row of the CSV file and writing all information included in the row in a text block within the to be created text file following the pre-defined schema. After the loop has ended the TXT file will be saved. (Test case 2)
* Needed libraries: pandas, matplotlib.pyplot
* possible errors:
  + Different separators may be used. Therefore, the separator should be defined in the loading function for the CSV file.
  + To circumvent parsing and encoding errors encoding as well as on\_bad\_lines parameter should be set.

## Design

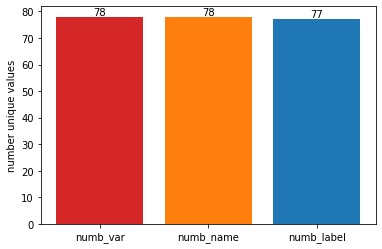


### Conducted tests

**Test case 1:**

**Input:** CSV file “Codebook\_Test.csv” with duplicate variable label

**Output:**



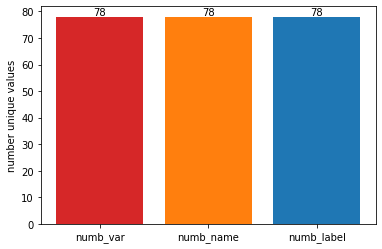
Please check your CSV file for duplets in the columns 'name' and 'label' and revise your file if necessary.

No text file created

**Test case 2:**

**Input:** CSV file “Codebook\_Test\_revised.csv” without duplicate variable label

**Output:**



CSV file ready for conversion.

Completed: Finished converting variable information.

Text file containing variable information in accordance with the pre-defined schema (“Codebook\_converted.txt”)