Main.py is the current backend code.

The top level code is toward the end, starting line 1203

Some parameters are hardcoded right now, which you may need to modify. Filename, colList and labelcolId must be modified based on the input data, other parameters may be fine to use default settings.

fileName = **"unsw-train-1k.csv" ------ replace this file with the data file user loaded. This is anonymized data file. First line is column names**

eps = 1 *--- This is the privacy budget, provide value user input*

numInstances = 20 *# this is the cluster size, user input*

sep = **','** *# separator, by default is ,*

priv = **'y'** *-- whether to compute privacy, default yes*

acc = **'n'** *--- whether run a classifier on anonymized data and compute accuracy, default is no, requires a test file as well*

headerExists = **'y'** *--- whether the csv file has header, default yes*

colList = **'0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41' --- list of columns that user selected to anonymize, start with 0 (these are index of columns)**

labelcolId = **'42' ---- index of the column that contains class label, usually the last column**

useContext = **'n' --- whether to use context based recommender for columns to anonymize, this part is not fully done so leave it as no.**

Output: a file named remade\_perturbed.txt will be generated. It has the same layout as input file except that values are now anonymized.

Privacy.txt stores privacy values. There are a few different metrics, the last number is what we use.

There are a number of other files that may be generated but they are used for computing privacy etc., so you don’t need to worry about them.