Description of my algorithm:

First of all, I have read the ICD10 GEM file to find all of the one-to-many mappings into a dataframe by groupby() attribute pandas of datafarme.

Then, converted the original California state ICD10 .xlsx file into a pure .txt file.

Thereafter, for each of the one-to-many mappings of the ICD10 GEM file, I have checked the frequency (TotalDiag) of each corresponding icd9 code and chose the icd9 code with highest frequency as the elite code to convert any one-to-many mapping to a one-to-one mapping.

My code takes the input of the two text files, i.e., “2018\_I10gem.txt” and “Q1-Q3-ICD-9-CM.txt” as input and writes out the output which is the result of final one-to-one mappings.