Instructions for running the programs:

The rules can be generated by executing the **ruleGen.py** program using the following command. This is the only program that has to be run if the input is already in the form of 1s and 0s indicating the presence of items, and a columns file with the same number of names as there are columns in the input file.

python ruleGen.py [input file name] [column file name] [minsup] [itemset generation method] [choice of interestingness measure] [confidence or lift]

[input file name] is the file name with 1s and 0s separated by ‘,’

[column file name] is the file with column names for the above file separated by ‘,’

[itemset generation method] takes one of the following values for choosing the frequent itemset generation method:

‘b’ for Brute force method

‘f1’ for Fk-1 \* F1

‘fk’ for Fk-1 \* Fk-1

[choice of interestingness measure] takes the following values

‘c’ for confidence

‘l’ for lift

[confidence or lift] takes a number to specify the minimum threshold of either lift or confidence whichever is chosen in the previous argument.

## Program descriptions:

1. **freqItemsetGen.py:** Generates item sets that are frequent. It is called from the program **ruleGen.py.** But it can also be run independently by appropriately changing the main function.
2. **ruleGen.py:** This will generate the rules, by first calling the program above and then a function to create the rules.
3. **convertColsNum2Cat.py:** This is a simple program to convert numerical data into categorical data. It will convert the numerical columns into exactly 3 categories based on the distribution.

Input filename has to be changed on line 5.

Line 4 has the list of column numbers that are to be converted into categorical.

1. **convertCat2Bin.py:** Converts categorical data into binary data.

Takes 2 inputs: file name to be converted and the file with columns corresponding to the categorical data.

The conversion procedure is as follows:

For each unique value in each column, an ID(number starting from 0) is assigned. Values are considered uniquely in each column independent of the values in other columns. Once all IDs have been assigned, these ID numbers become the columns in the binary dataset. And 1 is assigned to a row and a column, if that row has the value corresponding to that ID the column corresponds to.

A new columns file is created where the column names take the form *originalColumnName:uniqueValue.*

So, it gives 2 output files: Binary data file, new column file.

These are inputs to the **ruleGen.py** or **freqItemsetGen.py**