Decision Support Scientist- Intern

I. The given dataset [collaborativeData] is a pattern of each individual user's method of treating a patient (column-Selected) and those recommended by a software (column-Recommended) based on the input Symptoms (Column- Symptom 1 and Column- Symptom 2). For a total of 100 patients, 5 different users (Doctors) selected different Acupoints (Column- Selected) dependent/independent of recommended Acupoints (Column- Recommended).

Task 01. Write a program from scratch to create individual datasets (user 1, user 2, user 3, user 4, user 5) excluding the Chinese characters in column- Recommended and Selected. E.g. The output dataset should consist of Column 0-3, Column 4 – DU20, Column 5- DU20

Task 02. For a given condition, evaluate different user behaviour of choosing acupoints (column-selected) for different given each and every symptom. Statistically evaluate the results and present them visually. For e.g, if user 6 chose Acupoint X for fever, identify if user 5 chose Acupoint X or Y or Z or K and if different find the variances etc.., with what other Acupoints 5 other users treated for the same symptoms fever. At the end, find the same for each and every symptom vs. recommended and selected.

II. Cluster the [given data:Signal Database.xlsx], based on the Gender and present them visually based on the test type and test name. Also, try to understand the other dataset for different subjects available in the same repository and try to explain few lines on the collected dataset, make use of the document: Database_Explanation.docx for hints.