NTUWS Filter Newmember

Introduction

The modules are developed for processing and cleaning the data which generated by the Web Survey platform of National Taiwan University. It's desgined to filter the members of the platform who has filled in the "Welcome Questionnaire" and their recommenders were the interviewers recruited by NTUWS in 2023. The filtered members' list would be compared to the "General members" file, which recorded the detail information of the whole members. Users could get the result of each of members merely recommended by interviewers and the key features of them such as "age" or "area" etc. It enables users to track the successful samples each interviewers got and also evaluate their performance.

Methods and Functions

There are many different medthods below the main module class "admember()". The 5 important functions are:

- "get_filter()"
- "get_problem()"
- "duplicate_number()"
- "sample_merge()"
- "interviewers_performance()"

get_filter()

This is the major method below "admember()". The function would return the two objects: "**wqs**" and "**tts**", representing the filtered data of "welcome questionnaire samples" and "total members samples" respectively. The first one "wqs" is a pandas.dataframe which processed by the "Welcome Questionnaire" file based on the "recommenders" column, only the samples who recommended by interviewers would selected from the origin data. Simmilar to "wqs", the second one "tts" is data frame tpye too. It's selected from the "General members" file, based on the "phone" column which is consistent with the same numbers of the "phone" column in "wqs", "tts" is undoubtedly the data of members the interviewers aqcuired in "General members" file. Both of these two objects are defined as the **arguements** of following methods.

get_problem(wqs, tts)

Both "wqs" and "tts" are the arguements of this function. By comparing each numbers of "phone" whithin the two dataframes, users could filter the numbers exist in "wqs" but not in "tts". (It means such samples are **problematic**) The return type is a dataframe.

duplicate_number(wqs)

There's only one arguement "wqs". This function would count the frequency of each numbers of "phone" in "wqs", and then return a dataframe that shows all the numbers of which frequency are more than one.

sample_merge(wqs, tts)

The two arguements are "wqs" and "tts". This function would filter out the samples who have duplicate and problematic "phone" numbers in "wqs", those left behind and "tts" would be merged into a new dataframe. The result shows the entire qualified and successful samples every interviewers have recommended to become the new members.

interviewers_performance(wqs, tts)

There're also two arguements "wqs" and "tts". This function would count the members based on the specific values of "age" and "city" columns. The **elder** people and **non Great Taipei area** residents are the target samples. Interviers could fit the bill only when the members they recommended whose age is over 50 and not living in Great Taipei. It would return a table that shows the amount of target samples and general samples respectively each interviewers finally aqcuired so far.

Running

In addtion to running the codes by editting the script file, the modules also support running by GUI. Clone the whole file of this repository, make a new directory which is upper the project folder and then create a "main.py" file to run the modules. The codes whithin "main" should be:

```
import os
os.chdir(os.path.dirname(__file__))
from filtmem.filter_member_gui import filmem_gui
filmem_gui().mainloop()
```