How to build a filter for data storage

A data storage include a range of data with a range of features. Filters are to help users to quickly and clearly find data with certain features, it will significantly improve the efficiency of data analyzing and make it more convenient to do this.

Use a csv file as example, the data about information of athletes. The information include sex, age, year, team and sport.

First, import pandas and the data

import pandas as pd

df = pd.read\_csv("athlete\_events.csv")

def filtering(df):

then ask users to choose what features they want

print('Please, enter the numbers of the filters you would like to use (e.g. 234 if you want to filter by age, team and year):')

print('1. Sex')

print('2. Age')

print('3. Team')

print('4. Year')

print('5. Sport')

check availability

try:

    selection = list(input(":")) # ask the user the entered numbers

except (ValueError, SyntaxError):

    print('Please enter an integer.')

    continue # if the user did not enter an integer, continue printing the menu and ask the user to re-enter

create empty list

else:

            list1 = []

            list2 = [] # creating empty list for use

            if len(selection) <= 5: # to check if the user entered more than 5 numbers (e.g.123456)

record the user’s input to each filter

for n in range(len(selection)): # to iterate user input

                    if selection[n]=="1":

                        list1.append("Sex") # if user input "1", then add "Sex" into list1

                        a = input("Enter F for female, M for male:") # ask the user to chose "F" or "M",

                        list2.append(a) # and add "F" or "M" into list2

                    elif selection[n] == "2":

                        list1.append("Age")

                        a = int(input("Enter age in years:"))

                        list2.append(a)

                    elif selection[n] == "3":

                        list1.append("Team")

                        a = input("Enter the name of team:")

                        list2.append(a)

                    elif selection[n] == "4":

                        list1.append("Year")

                        a = int(input("Enter year:"))

                        list2.append(a)

                    elif selection[n] == "5":

                        list1.append("Sport")

                        a = input("Enter the name of sport:")

                        list2.append(a)

                    else: # if the user entered a number>5, return 0 number of records and the original dataframe

                        print("%s is an invalid number" % selection[n])

                        print("Number of records is zero")

                        number=0

                        return df,number

if user enter more than 5 numbers

else: # if the user entered more than 5 numbers, e.g.123456, return 0 number of records and the original dataframe

                print("Too many options")

                number=0

                return df,number # return to filtering()

check availability

for i in list1: # to check duplicates

                if list1.count(i) > 1: # if the user entered a number more than once, e.g. 122:

                    number = 0

                    return df, number # return 0 records and original dataframe

                else:

                    pass

if list empty

if not list1: # execute if list1 is empty

                number=0

                return df, number # return 0 records and original dataframe to filtering()

provide results

else:

                new\_df = df[(df[str(list1[0])]==list2[0])]

                for n in range(1,len(list1)): # filter the dataframe using the condition entered by the user

                    new\_df = new\_df[(new\_df[str(list1[n])]==list2[n])]

                number = len(new\_df) # number of records retrieved

                return new\_df,number # return the filtered new dataframe and the number of records

new\_df is the data after filtered