



QLAB-Courses / summer_python_econ

Q Type [/] to search



<> Code Issues 9 Pull requests 1 Actions Projects Wiki Security Insights



main

summer_python_econ / assignments / assignment_2 / Assignment_2.ipynb



Q Go to file



1DanielFP new assignments

def68a9 · 2 days ago



History



main

summer_python_econ / assignments / assignment_2 / Assignment_2.ipynb

↑ Top

Preview

Code

Blame

373 lines (373 loc) · 10 KB

Raw



Assignment_2

If you copy codes from previous answers, explain each step. No explanation is 0 points.


Part 1. Pandas

1. Go to [INEI's](#) page.
2. Search for "Consulta por encuestas" and click on ENAHO metodología ACTUALIZADA.
3. Choose the one that says "Condiciones de Vida y Pobreza - ENAHO" (the one that isn't panel).
4. Choose the year 2022, and in period choose yearly..
5. Download the **CSV files** from the Code 1 (housing and home characteristics) and module 34 (Summary - Calculated variables).

(The file of the summary is Sumaria-2022).

This task is meant to be developed using only the CSV files.

1. Read both csv files with pandas and label them with a name that allows you to work with them as databases. (If there is an error, it's because our surveys are taken on spanish and have some characteristics such as ~ from the ñ or the accent. That's why in the options of reading the csv file add the next option `encoding = "latin-1"` .

(Don't forget to import pandas )

In case your columns have their labels with uppercase , you can run the next code to switch to lowercase.

```
df.columns= df.columns.str.lower()
```

Where `df` is the name of the data frame.

In []:

In []:

2.
 - What are the dimensions of the summary(or the number of columns and rows)
 - What are the dimensions of the housing module (or the number of columns and rows).

In []:

In []:

3. We wont need all the variables from ENAHO.

- Keep the next variables from the summary CONGLOME VIVIENDA HOGAR MIEPERHO PERCEPHO POBREZA GASHOG2D INGHOG2D UBIGEO
- Keep the next variables from the housing module CONGLOME VIVIENDA HOGAR RESULT NBI1 NBI2 NBI3 NBI4 NBI5 .

Note: each one of the variables is :

- MIEPERHO: number of members per home..
- PERCEPHO: number of members that earn some income
- POBREZA: monetary poverty according to the poor line.
- GASHOG2D: gross anual household expenditure.
- INGHOG2D: anual net household incomer.
- UBIGEO: 6 digit code that indicates the district the persons belongs to.

Basic needs::

- NBI1: inadequate housing
- NBI2: overcrowding
- NBI3: housing without hygienic services
- NBI4: households with children who do not attend school.
- NBI5: households with high economic dependence

CONGLOME, VIVIENDA y HOGAR are the unique's household identifiers (like a key id).

In []:

4. The housing module includes all the houses that got into the sampling frame even the ones that didn't participate in the ENAHO for various reasons. To delete these houses just keep the ones with RESULT is equal to 1 or 2. Keep this information for later

In []:

5. We want to calculate a series of new variables(or new columns) :

In the summary :

- We want the number of household dependent. These are the ones that don't perceive any income or the difference between the total household members and the total of perceivers.
- We want the average monthly income from the household perceivers.
- We want the difference between income and expenditure of the households.

In []:

In []:

In []:

6. We want the poverty variable in the summary to be codified as an integer 1, 2, 3 according to: 1 : extreme poor home 2 : not extreme poor home 3 : not poor home

Using this variable , make two more:

1. A variable that says 'extreme poor' , ' not extreme poor' and 'not poor ' instead of 1,2,3.
2. A variable that has 'poor' and 'not poor' (or one that combines the poor and not extreme poor households).

In []:

In []:

7. We want to now the average household income (inghog2d) per poverty level. Make a `groupby` to find this.

(For example, the extreme poor households would only have one mean and it's the same for the others two poor categories).

In []:

8. We whould now merge the data

But before a sanity check: Verify that the number of rows in both data bases are the same.(remember we filtrate the observations by `result`). The ENAHO recolects the information of these 34,490 homes trough all their modules. In order to use all the modules at once, we would merge them.

Merge both bases with an *inner merge*. Remember that the id that identifies each home its componed by `conglome vivienda hogar` .

In []:

In []:

In []:

Part 2. If and loops

1.

a. Do a for-loop that goes from 22 to 42, but that only take the odd numbers. Print each of the elements as you iterate through them. Hint :
you can use `ranges`

you can see ranges:

b. Modify your program so that as you iterate through them, you add them to a variable called `sum_total`, defined before the loop begins. Use a operator that has a higher allocation.

In []:

In []:

2. Write a while loop that begins in 1 and ends in 100, but each loop, the counter `i` is increasing `+=2`.

In []:

3. Write a program that asks the user for their age and determines if they are of legal age (18 years or older). If it is, it displays a message indicating that you can vote. If not, it displays a message indicating that you cannot vote yet.

In []:

In []: