Mod_300_empleo_funciones

February 12, 2023

Base de datos : ENAHO

Modulo: Educacion

El proyecto del presente notebook, esta centrado en poder extraer información de la encuesta nacional de hogares (ENAHO), generando una base a nivel de ingresos y años de educacion

Para generar la queña base, la cual sera denominado Mincer, se debe trabajar con variables en las bases de de diferentes modulos de la ENAHO, los cuales son:

- Moudlo 300: educación

1 Instalacion de Librerias

[1]:
```
import pandas
import os
import numpy
import sys
import pyreadstat
```

[2]:
```
#Ruta de carpetas
ruta = 'D:/Dropbox/BASES/ENAHO'
output = 'D:/Dropbox/BASES/ENAHO/Python_scripts'
```

2 Base de Educacion

[3]:
```
info = pandas.read_stata(os.path.join(ruta,"2021","enaho01a-2021-300.dta"),␣
↪convert_categoricals=False )
```

[4]:
```
def variable_reduca(frame, target, var1, var2, var3):

    """
    El comando variable_reduca, genera la variable de reduca, el cual permitira␣
↪la variable cuantiativa de los
    años de educacion de la persona, considerando las varuables: p301a, p301b,␣
↪p301c
```

```
    """

    frame[var1] = pandas.to_numeric(frame[var1])
    frame[var2] = pandas.to_numeric(frame[var2])
    frame[var3] = pandas.to_numeric(frame[var3])
    zeros = frame[var3].min()

    frame[target] = frame[var2]

    frame.loc[(frame[var1]>=1) & (frame[var1]<=4),target] = (frame[target] + 0)
    frame.loc[(frame[var1]>=5) & (frame[var1]<=6),target] = (frame[target] + 6)
    frame.loc[(frame[var1]>=7) & (frame[var1]<=10),target] = (frame[target] + 11)
    frame.loc[(frame[var1]==11),target] = (frame[target] + 16)

    return frame

def variable_rpersona(frame, target, var1, var2, var3, var4):
    frame[target] = frame[var1]+frame[var2]+frame[var3]+frame[var4]
    return frame
```

[5]:
```
# Codigo de persona
base_educa = variable_rpersona(info, "rpersona",␣
 ↪"conglome","vivienda","hogar","codperso")

# Variabl de años de educacion
base_educa = variable_reduca(info,'reduca',"p301a","p301b","p301c")
```

```
C:\Users\edinson\AppData\Local\Temp\ipykernel_11916\533699019.py:23:
PerformanceWarning: DataFrame is highly fragmented.  This is usually the result
of calling `frame.insert` many times, which has poor performance.  Consider
joining all columns at once using pd.concat(axis=1) instead. To get a de-
fragmented frame, use `newframe = frame.copy()`
  frame[target] = frame[var1]+frame[var2]+frame[var3]+frame[var4]
C:\Users\edinson\AppData\Local\Temp\ipykernel_11916\533699019.py:13:
PerformanceWarning: DataFrame is highly fragmented.  This is usually the result
of calling `frame.insert` many times, which has poor performance.  Consider
joining all columns at once using pd.concat(axis=1) instead. To get a de-
fragmented frame, use `newframe = frame.copy()`
  frame[target] = frame[var2]
```

[6]:
```
base_educa['reduca'].describe()
```

[6]:
```
count    102582.000000
mean          7.494882
std           5.661735
min           0.000000
25%           1.000000
```

```
50%           9.000000
75%          11.000000
max          18.000000
Name: reduca, dtype: float64
```

[7]:
```
help(variable_reduca)
```

```
Help on function variable_reduca in module __main__:

variable_reduca(frame, target, var1, var2, var3)
    El comando variable_reduca, genera la variable de reduca, el cual permitira
la variable cuantiativa de los
    años de educacion de la persona, considerando las varuables: p301a, p301b,
p301c
```

[8]:
```
base_educa[['p301a','p301b','p301c','reduca']].sample(19).transpose()
```

[8]:

	94091	102530	68944	102892	26144	9422	56449	57704	\
p301a	6.0	5.0	2.0	8.0	4.0	7.0	8.0	4.0	
p301b	5.0	3.0	1.0	3.0	0.0	1.0	3.0	0.0	
p301c	NaN	NaN	NaN	NaN	6.0	NaN	NaN	6.0	
reduca	11.0	9.0	1.0	14.0	0.0	12.0	14.0	0.0	

	47371	76775	24832	12196	83573	27299	39908	20866	\
p301a	8.0	5.0	6.0	8.0	3.0	6.0	5.0	1.0	
p301b	3.0	3.0	5.0	3.0	0.0	5.0	4.0	NaN	
p301c	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	
reduca	14.0	9.0	11.0	14.0	0.0	11.0	10.0	NaN	

	39319	32039	34196
p301a	5.0	1.0	8.0
p301b	4.0	NaN	3.0
p301c	NaN	NaN	NaN
reduca	10.0	NaN	14.0

[9]:
```
base_final_educa = base_educa[['rpersona','reduca']]
base_final_educa.head(4)
```

[9]:

	rpersona	reduca
0	0050070031101	17.0
1	0050070121101	11.0
2	0050070221101	16.0
3	0050070221102	14.0

3 Exportar Exxcel

[ ]:
```
# Observamos al direccion actual
os.chdir(output)
os.getcwd()

base_final_educa.to_csv('BD_Educacion_2021.csv')
base_final_educa.to_excel('BD_Educacion_2021.xlsx')
```

[ ]:
```

```