

Modulo_500_funciones

February 12, 2023

Base de datos : ENAHO

Modulo: Empleo

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:

- Modulo 500 : ingresos y empleo

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'
```

```
[3]: #Ejemplos de bucles
fechas =[2020, 2021]
for x in fechas:
    print(f"periodos de enaho: {x}")
    print(x)
    print(os.path.join(ruta,f"{x}",f'enaho01a-{x}-500.dta'))
```

periodos de enaho: 2020

2020

D:/Dropbox/BASES/ENAHO\2020\enaho01a-2020-500.dta

periodos de enaho: 2021

2021

D:/Dropbox/BASES/ENAHO\2021\enaho01a-2021-500.dta

2 Base de Empleo

```
[4]: data = pandas.read_stata(os.path.join(ruta, "2021", "enaho01a-2021-500.dta"))
data.shape
```

```
[4]: (86806, 1459)
```

```
[5]: data['p301a'].value_counts()
```

```
[5]: secundaria completa                21989
secundaria incompleta                 15376
primaria incompleta                  11499
primaria completa                     9954
superior no universitaria completa    6757
superior universitaria completa       6381
superior universitaria incompleta     5279
sin nivel                             4626
superior no universitaria incompleta  3586
maestria/doctorado                   1182
básica especial                       79
inicial                              58
Name: p301a, dtype: int64
```

```
[6]: def variable_r6(frame, target, varx, subdata):
    frame[target] = numpy.where(frame[varx]=="ocupado", subdata.sum(axis=1), 0)
    frame[target] = frame[target] / 12
    return frame

def variable_rDpto(frame, target, varx):
    frame[target] = frame[varx].str[0:2]
    frame[target] = frame[target].replace({"01": "Amazonas", "02": "Ancash", "03":
↪ "Apurimac", "04": "Arequipa", "05": "Ayacucho",
                                           "06": "Cajamarca", "07": "Callao", "08":
↪ "Cusco", "09": "Huanavelica", "10": "Huanuco",
                                           "11": "Ica", "12": "Junin", "13": "La
↪ Libertad", "14": "Lambayeque", "15": "Lima",
                                           "16": "Loreto", "17": "Madre de
↪ Dios", "18": "Moquegua", "19": "Pasco", "20": "Piura",
                                           "21": "Puno", "22": "San Martin", "23":
↪ "Tacna", "24": "Tumbes", "25": "Ucayali"
                                           })
    return frame

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

```

def variable_redad(frame, target, varx ):
    frame[target] = frame[varx]
    return frame

def variable_rmujer(frame, target, varx):
    frame[target] = numpy.where(frame[varx]=='mujer', 1,0)
    return frame

def variable_r3(frame, target, varx, vary):

    frame[target] = frame[varx].replace({"ocupado": "ocupado",
                                         "no pea": "desempleo",
                                         "desocupado abierto": "inactivo",
                                         "desocupado oculto": "inactivo"})

    frame[vary] = frame[target].replace({"ocupado": "pea",
                                         "desempleo": "pea",
                                         "inactivo": "no pea",
                                         "0": "no pea"})

    return frame

def variable_rneduca(frame, target, varx ):
    frame[target] = frame[varx].replace({"sin nivel": "inicial",
                                         "básica especial": "inicial",
                                         "inicial": "inicial",
                                         "primaria completa": "primaria",
                                         "primaria incompleta": "primaria",
                                         "secundaria completa": "secundaria",
                                         "secundaria incompleta": "secundaria",
                                         "superior no universitaria completa": "
↪"tecnic",
                                         "superior no universitaria incompleta": "
↪"tecnic",
                                         "superior universitaria completa": "Univ.",
                                         "superior universitaria incompleta": "Univ.",
                                         "maestria/doctorado": "Univ."})

    return frame

```

```

[7]: # Variables de empleo
df =
↪data[['i524a1', 'd529t', 'i530a', 'd536', 'i538a1', 'd540t', 'i541a', 'd543', 'd544t']]
base_empleo = variable_r6(data, "r6", "ocu500", df)

# Variables demograficas

```

```

base_empleo = variable_rDpto(data, "rDpto", "ubigeo")
base_empleo = variable_rmujer(data, "rmujer", "p207")
base_empleo = variable_redad(data, "redad", "p208a")
base_empleo = variable_r3(data, "r3", "ocu500", "rpea")

base_empleo = variable_rneduca(data, 'rneduca', "p301a")

# Código de persona
base_empleo = variable_rpersona(data, "rpersona", "
↳"conglome", "vivienda", "hogar", "codperso")

```

```

[8]: base_empleo['rneduca'].value_counts()
#base_empleo.sample(4).transpose()

```

```

[8]: secundaria    37365
     primaria      21453
     Univ.         12842
     tecnica       10343
     inicial        4763
     Name: rneduca, dtype: int64

```

```

[9]: base_final_empleo =
     ↳base_empleo[['rpersona', 'r6', 'rDpto', 'redad', 'rmujer', 'rneduca', 'r3']]
     base_final_empleo.head(4)

```

```

[9]:
      rpersona      r6      rDpto  redad  rmujer      rneduca      r3
0  0050070031101  6433.593913  Amazonas    56      0      Univ.  ocupado
1  0050070121101  1062.165853  Amazonas    21      1  secundaria  ocupado
2  0050070221101  2224.665202  Amazonas    62      0      Univ.  ocupado
3  0050070221102  1619.210449  Amazonas    61      1      tecnica  ocupado

```

3 Exportar excel

```

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

     base_final_empleo.to_csv('BD_Empleo_2021.csv')
     base_final_empleo.to_excel('BD_Empleo_2021.xlsx')

```

```

[ ]:

```