

Analisis Inicial - Tarea 1

Nombre: Paulina Aldape Bretado

Importing Libraries

```
In [1]: import numpy as np # linear algebra
import pandas as pd # data processing
import matplotlib.pyplot as plt
import seaborn as sns
# Input data files are available in the "../input/" directory.
import os
```

Loading the Dataset

```
In [30]: empleo=pd.read_csv('/content/BD Tarea 1 Analisis inicial.csv')
```

First 10 values

```
In [31]: empleo.head(10)
```

Out[31]:

	t_loc_tri	h_mud	sex	eda	cs_p13_1	n_hij	e_con	zona	clase1	clase2	domestico	anios_e
0	1	0	1	43	7		6	2	1	1	3	
1	1	0	2	67	4	2	2	2	1	1	3	
2	1	0	2	26	7	0	6	2	1	1	3	
3	1	0	2	28	7	0	6	2	1	1	3	
4	1	0	2	30	7	0	6	2	1	1	3	
5	1	0	2	26	7	0	6	2	1	1	3	
6	1	0	1	59	2		6	2	1	1	3	
7	1	0	2	47	6	1	2	2	1	1	3	
8	1	0	1	64	4		3	2	1	1	3	
9	1	0	1	37	3		5	2	1	1	4	

Shape of the dataset

```
In [4]: empleo.shape
```

```
Out[4]: (185627, 15)
```

Print the name of columns

```
In [5]: empleo.columns
```

```
Out[5]: Index(['t_loc_tri', 'h_mud', 'sex', 'eda', 'cs_p13_1', 'n_hij', 'e_con',  
              'zona', 'clase1', 'clase2', 'domestico', 'anios_esc', 't_tra',  
              'hrsocup', 'ingocup'],  
             dtype='object')
```

Check for NULL Values

```
In [6]: empleo.isnull().sum()
```

```
Out[6]: t_loc_tri      0  
        h_mud         0  
        sex           0  
        eda           0  
        cs_p13_1      0  
        n_hij         0  
        e_con         0  
        zona          0  
        clase1        0  
        clase2        0  
        domestico     0  
        anios_esc     0  
        t_tra         0  
        hrsocup       0  
        ingocup       0  
        dtype: int64
```

Check unique values

```
In [7]: empleo.nunique()
```

```
Out[7]: t_loc_tri      4
        h_mud         4
        sex           2
        eda           86
        cs_p13_1      11
        n_hij         24
        e_con         7
        zona          2
        clase1        1
        clase2        2
        domestico     5
        anios_esc     25
        t_tra         2
        hrsocup       117
        ingocup       2037
        dtype: int64
```

Check for Duplicate values

```
In [8]: empleo.duplicated().sum()
```

```
Out[8]: 11278
```

Existen 11,278 valores duplicados que se requieren remover de la base de datos.

Make a copy of the dataset

```
In [32]: df = empleo.copy()
```

```
In [33]: df.shape
```

```
Out[33]: (185627, 15)
```

Drop NULL values

```
In [34]: df=df.dropna()
        df.shape
```

```
Out[34]: (185627, 15)
```

Print first 10 values

In [35]: `df.head(10)`

Out[35]:

	t_loc_tri	h_mud	sex	eda	cs_p13_1	n_hij	e_con	zona	clase1	clase2	domestico	anios_e
0	1	0	1	43	7		6	2	1	1	3	
1	1	0	2	67	4	2	2	2	1	1	3	
2	1	0	2	26	7	0	6	2	1	1	3	
3	1	0	2	28	7	0	6	2	1	1	3	
4	1	0	2	30	7	0	6	2	1	1	3	
5	1	0	2	26	7	0	6	2	1	1	3	
6	1	0	1	59	2		6	2	1	1	3	
7	1	0	2	47	6	1	2	2	1	1	3	
8	1	0	1	64	4		3	2	1	1	3	
9	1	0	1	37	3		5	2	1	1	4	

Revisar los tipos de datos de las variables

In [36]: `df.dtypes`

Out[36]:

t_loc_tri	int64
h_mud	int64
sex	int64
eda	int64
cs_p13_1	int64
n_hij	object
e_con	int64
zona	int64
clase1	int64
clase2	int64
domestico	int64
anios_esc	int64
t_tra	int64
hrsocup	int64
ingocup	int64
dtype:	object

La mayoría de las variables no están bien catalogadas, esto debido a que la mayoría son variables categóricas con excepción de la variable de edad, número de hijos, años de escolaridad, horas trabajadas y ingreso. Por lo que se deben de categorizar de la forma correcta para realizar la limpieza de la base de datos.

```
In [37]: # Convertir los campos correspondientes en string
df['t_loc_tri'] = pd.Series(df['t_loc_tri'], dtype="string")
df['sex'] = pd.Series(df['sex'], dtype="string")
df['cs_p13_1'] = pd.Series(df['cs_p13_1'], dtype="string")
df['e_con'] = pd.Series(df['e_con'], dtype="string")
df['zona'] = pd.Series(df['zona'], dtype="string")
df['clase1'] = pd.Series(df['clase1'], dtype="string")
df['clase2'] = pd.Series(df['clase2'], dtype="string")
df['clase1'] = pd.Series(df['clase1'], dtype="string")
df['domestico'] = pd.Series(df['domestico'], dtype="string")
df['t_tra'] = pd.Series(df['t_tra'], dtype="string")
```

```
In [45]: df['h_mud'] = pd.Series(df['h_mud'], dtype="int64")
```

```
In [51]: df['n_hij'] = pd.Series(df['n_hij'], dtype="int64")
```

```
In [52]: df.dtypes
```

```
Out[52]: t_loc_tri    string
h_mud          int64
sex            string
eda            int64
cs_p13_1       string
n_hij          object
e_con          string
zona           string
clase1         string
clase2         string
domestico      string
anios_esc      int64
t_tra          string
hrsocup        int64
ingocup        int64
dtype: object
```

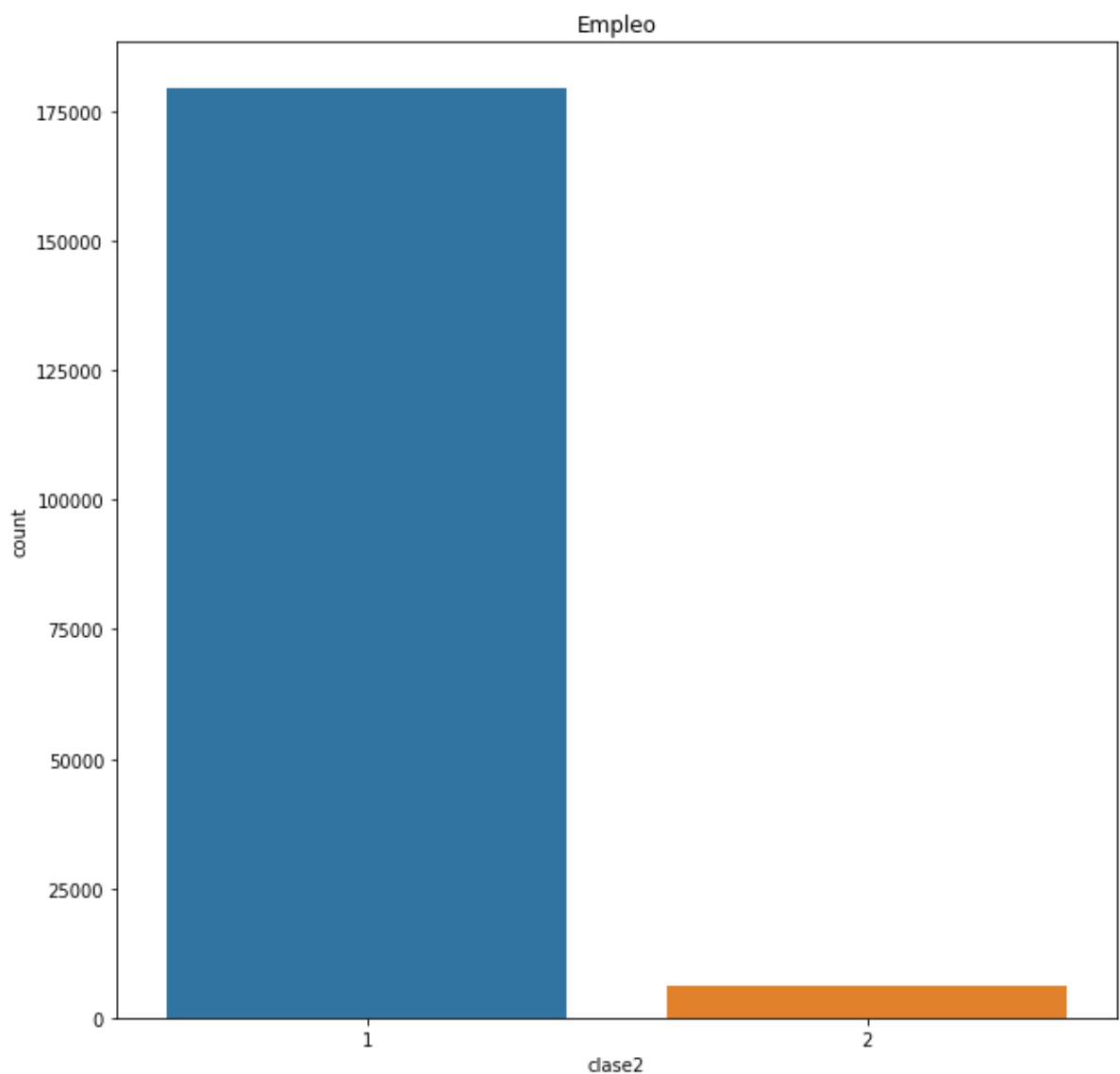
Data Visualization

Revisar la variable de clase 2 que será la variable de interes dentro del analisis

```
In [41]: sns.countplot(df['clase2'])  
fig = plt.gcf()  
fig.set_size_inches(10,10)  
plt.title('Empleo')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.  
warnings.warn(
```

```
Out[41]: Text(0.5, 1.0, 'Empleo')
```



A primera instancia se puede observar que existe un desbalanceo de las clases, que se tendrá que corregir.

Tamaño de localidad

```
In [42]: sns.countplot(df['t_loc_tri'])
sns.countplot(df['t_loc_tri']).set_xticklabels(sns.countplot(df['t_loc_tri']).
get_xticklabels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('Tamaño de localidad')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

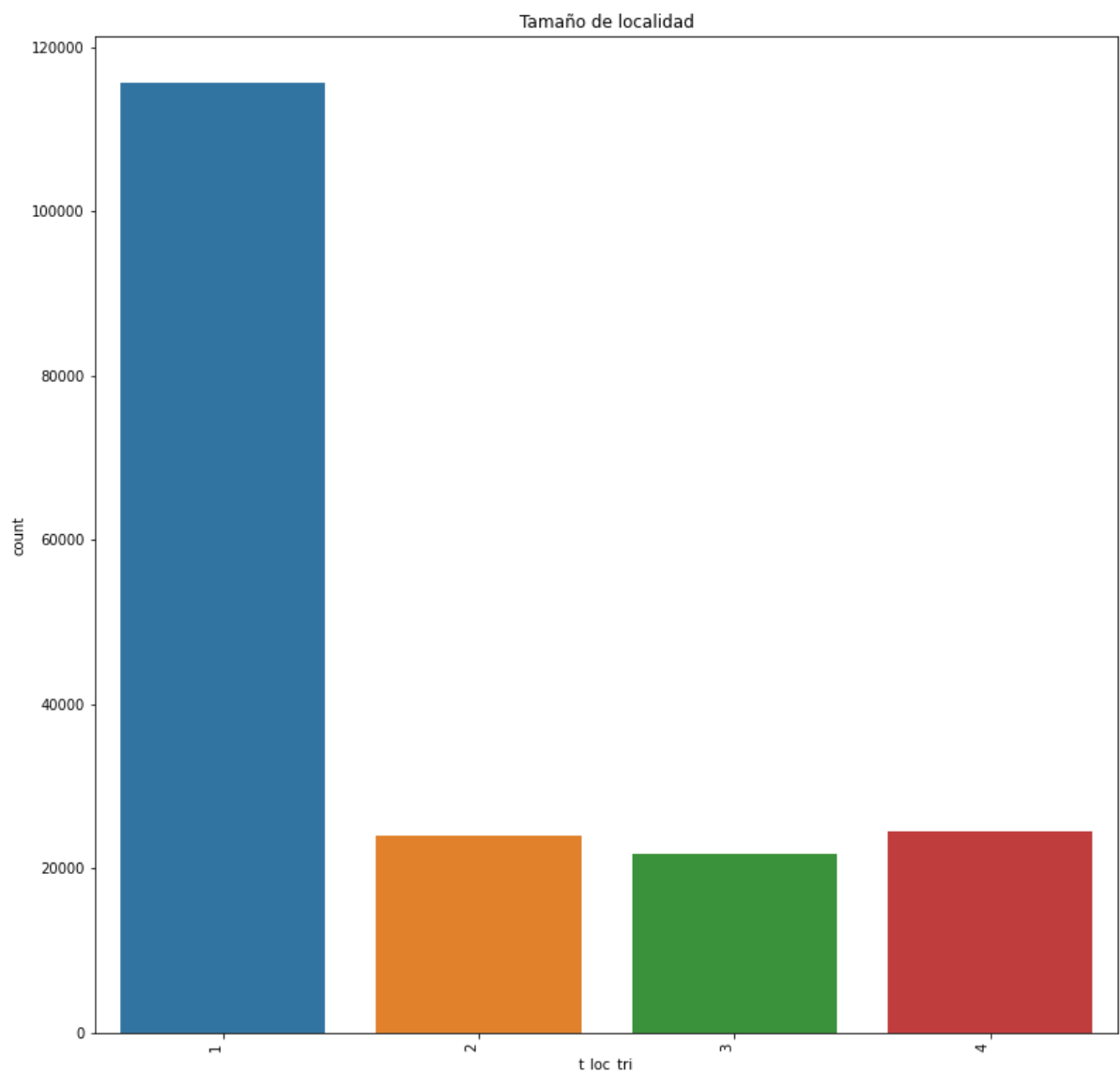
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[42]: Text(0.5, 1.0, 'Tamaño de localidad')



Numero de veces que se han mudado del hogar

```
In [47]: sns.countplot(df['h_mud'])
sns.countplot(df['h_mud']).set_xticklabels(sns.countplot(df['h_mud']).get_xtic
klabels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('Numero de veces que se han mudado')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

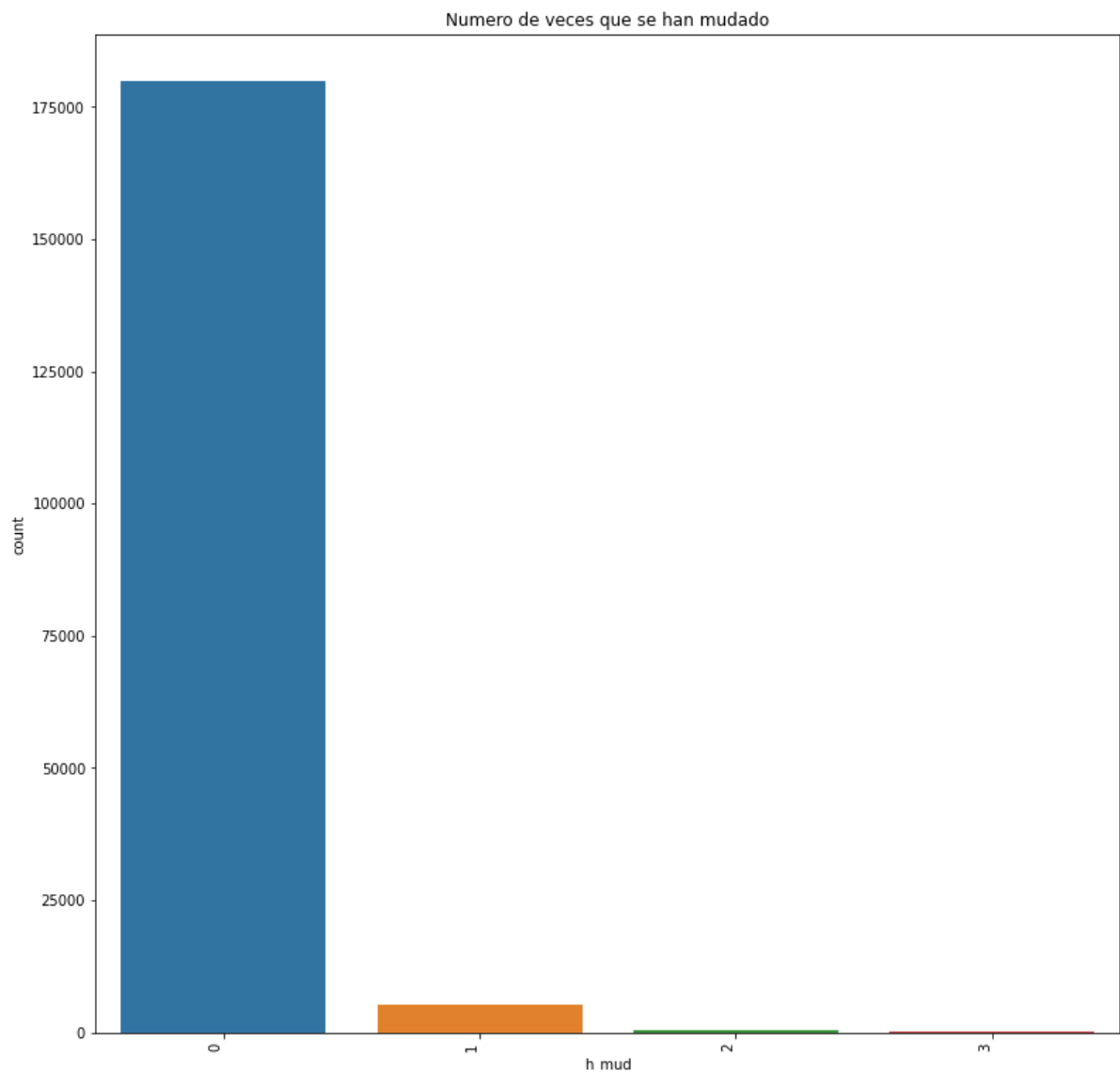
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[47]: Text(0.5, 1.0, 'Numero de veces que se han mudado')



Variable de genero

```
In [48]: sns.countplot(df['sex'])
sns.countplot(df['sex']).set_xticklabels(sns.countplot(df['sex']).get_xticklabels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('Genero')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

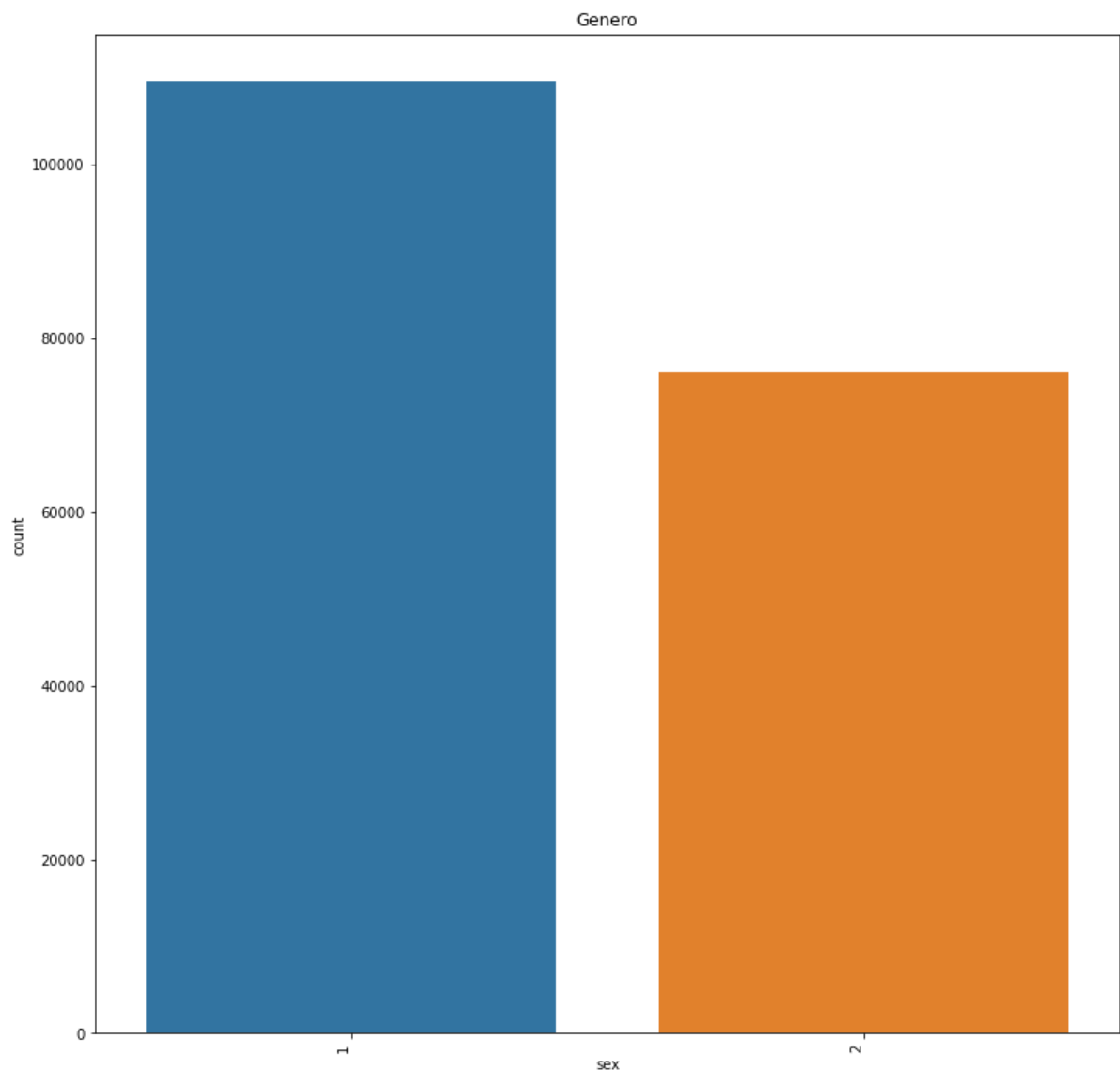
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[48]: Text(0.5, 1.0, 'Genero')



Variable de edad

```
In [49]: sns.countplot(df['eda'])
sns.countplot(df['eda']).set_xticklabels(sns.countplot(df['eda']).get_xticklabels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('Edad')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

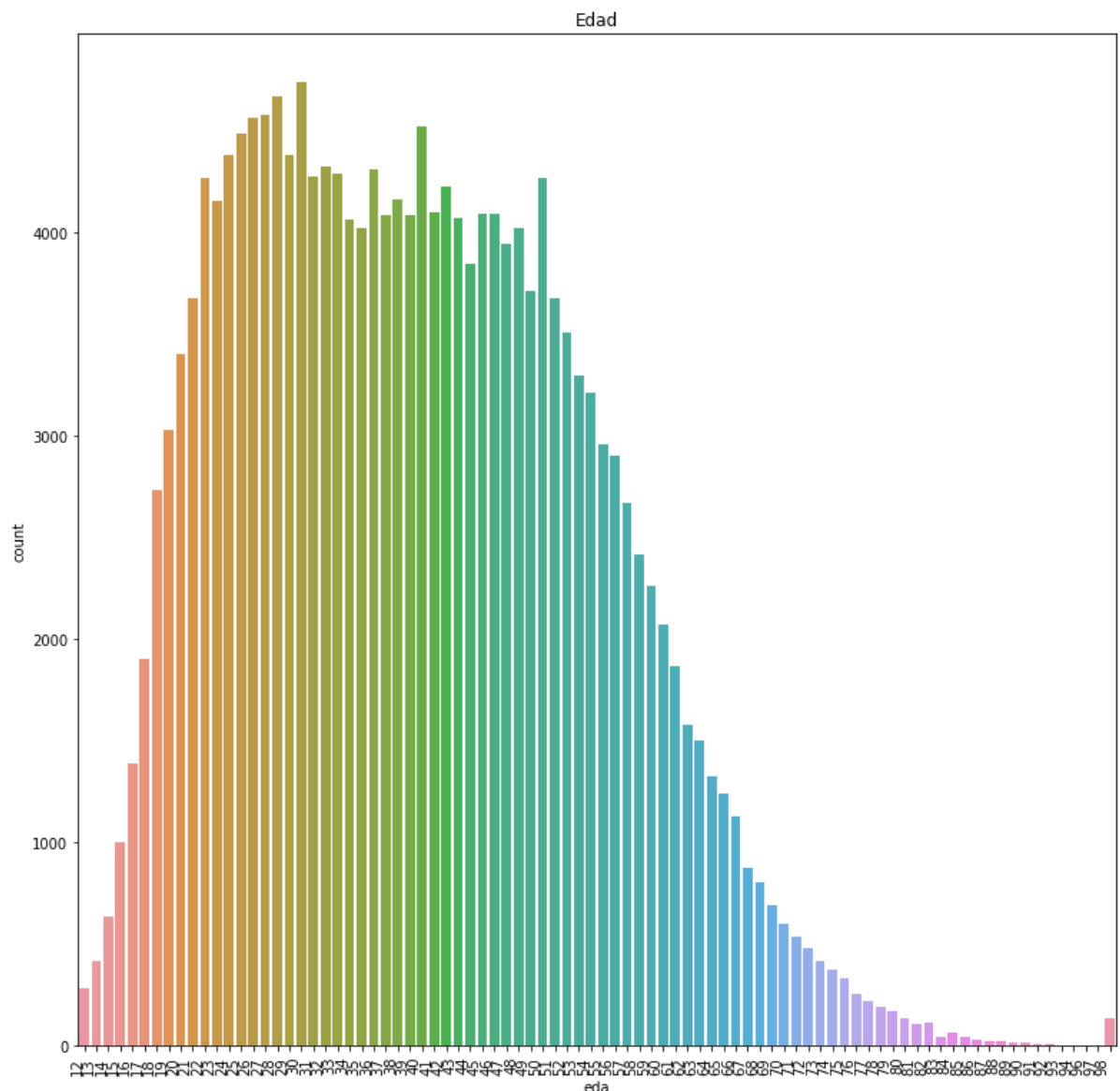
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[49]: Text(0.5, 1.0, 'Edad')



Se puede observar que la variable de edad tiene una distribucion cargada a la izquierda debido a que no hay muchas personas con edades tan avanzadas.

Variable de nivel de escolaridad

```
In [50]: sns.countplot(df['cs_p13_1'])  
sns.countplot(df['cs_p13_1']).set_xticklabels(sns.countplot(df['cs_p13_1']).ge  
t_xticklabels(), rotation=90, ha="right")  
fig = plt.gcf()  
fig.set_size_inches(13,13)  
plt.title('Nivel de escolaridad')
```



```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

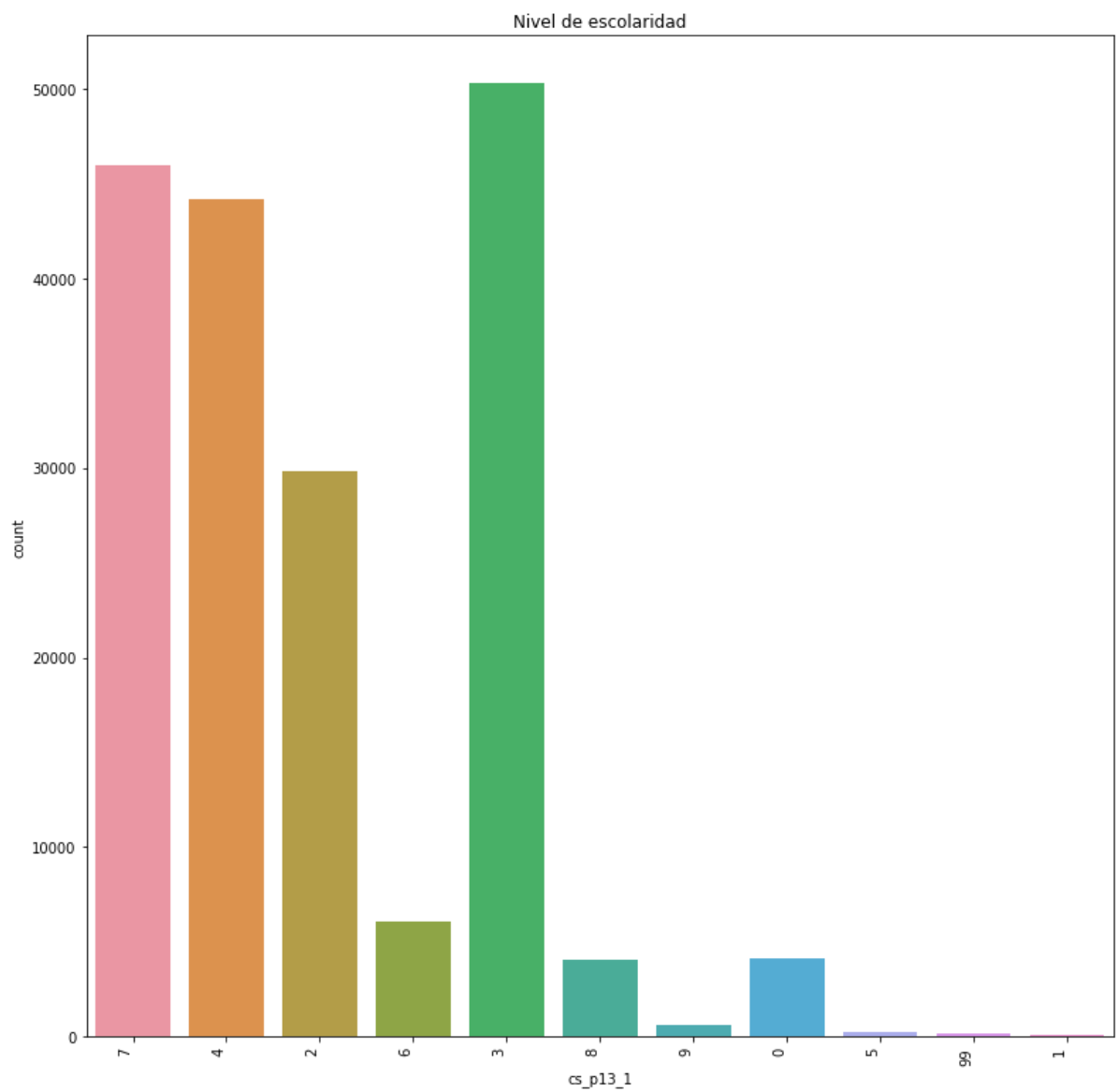
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[50]: Text(0.5, 1.0, 'Nivel de escolaridad')



Se observa que la categoría con mas registros es la de secundaria seguida por el nivel de profesional.

Numero de hijos

```
In [53]: sns.countplot(df['n_hij'])  
sns.countplot(df['n_hij']).set_xticklabels(sns.countplot(df['n_hij']).get_xtic  
klabels(), rotation=90, ha="right")  
fig = plt.gcf()  
fig.set_size_inches(13,13)  
plt.title('Numero de hijos')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

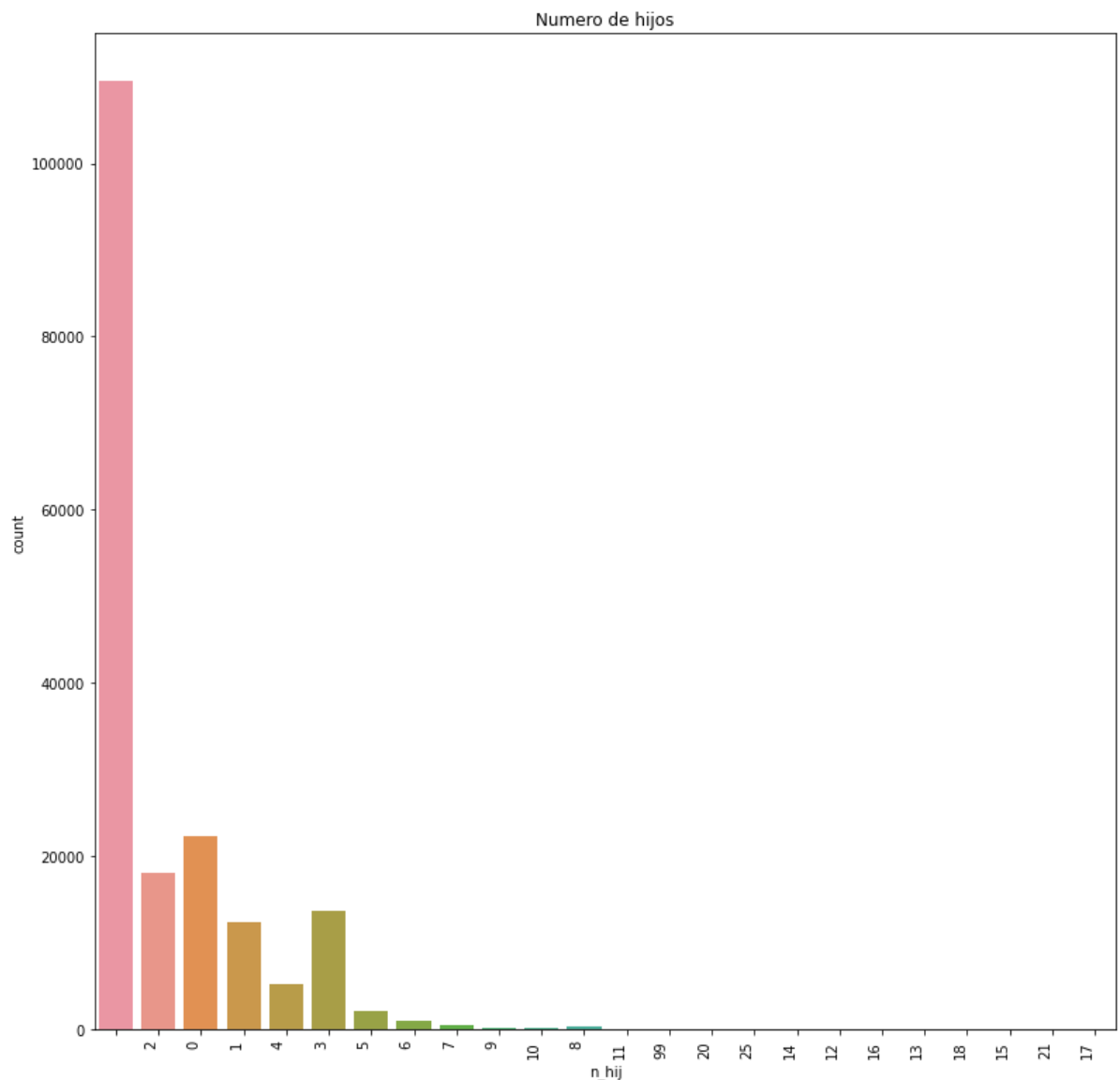
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[53]: Text(0.5, 1.0, 'Numero de hijos')



Con el grafico nos podemos dar cuenta que existe un gran numero de registros que no tienen valor en esta variable que no lo detectaba el comando de isnull debido a que la celda cuenta con espacios en blanco, lo cual lo cuenta como caracter.

Estado conyugal

```
In [54]: sns.countplot(df['e_con'])
sns.countplot(df['e_con']).set_xticklabels(sns.countplot(df['e_con']).get_xtic
klabels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('Estado conyugal')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

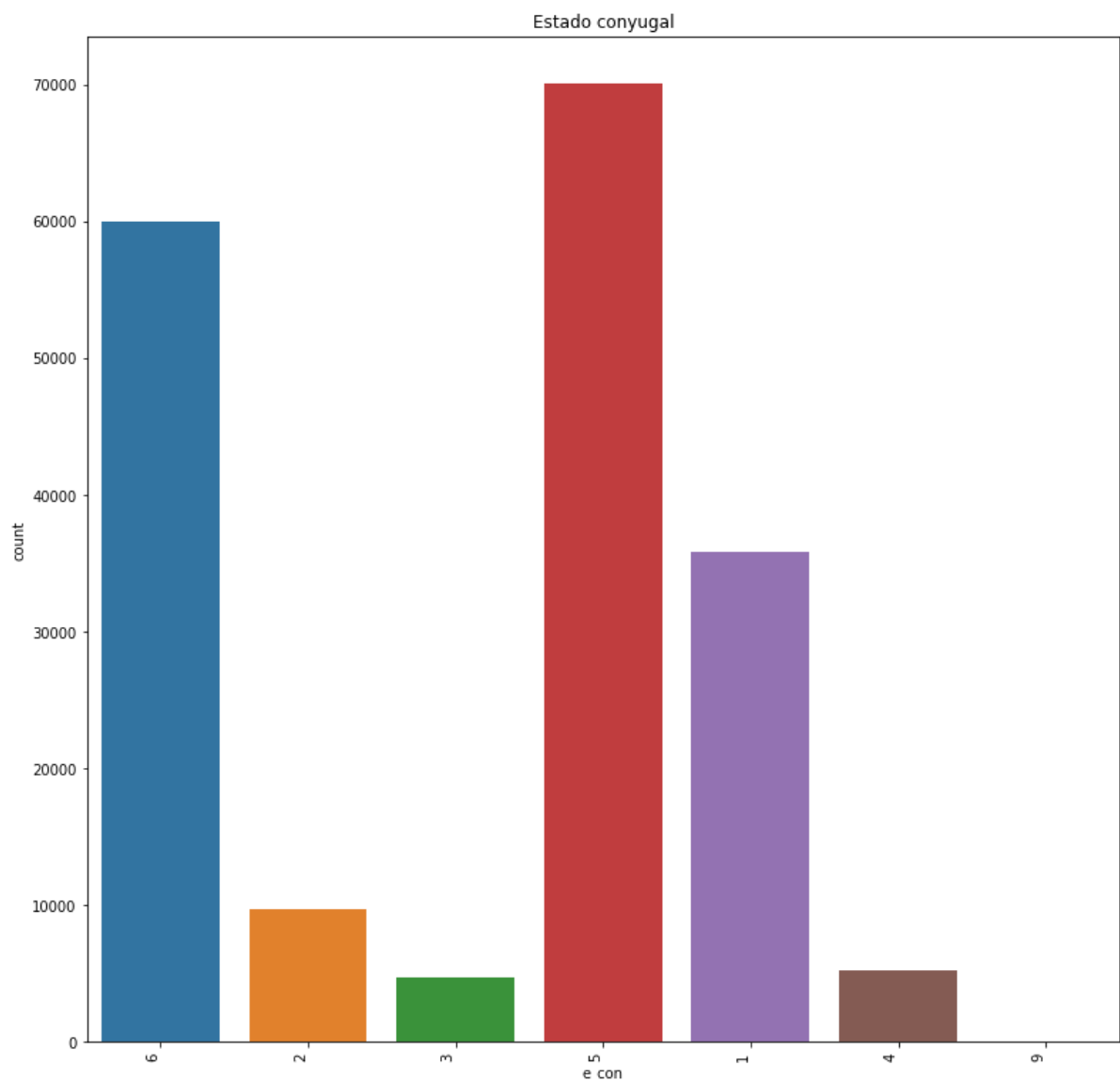
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[54]: Text(0.5, 1.0, 'Estado conyugal')



Las dos primeras categorías que resaltan son las de casado y soltero.

Zona salarial

```
In [55]: sns.countplot(df['zona'])
sns.countplot(df['zona']).set_xticklabels(sns.countplot(df['zona']).get_xtickl
abels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('Zona salarial')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

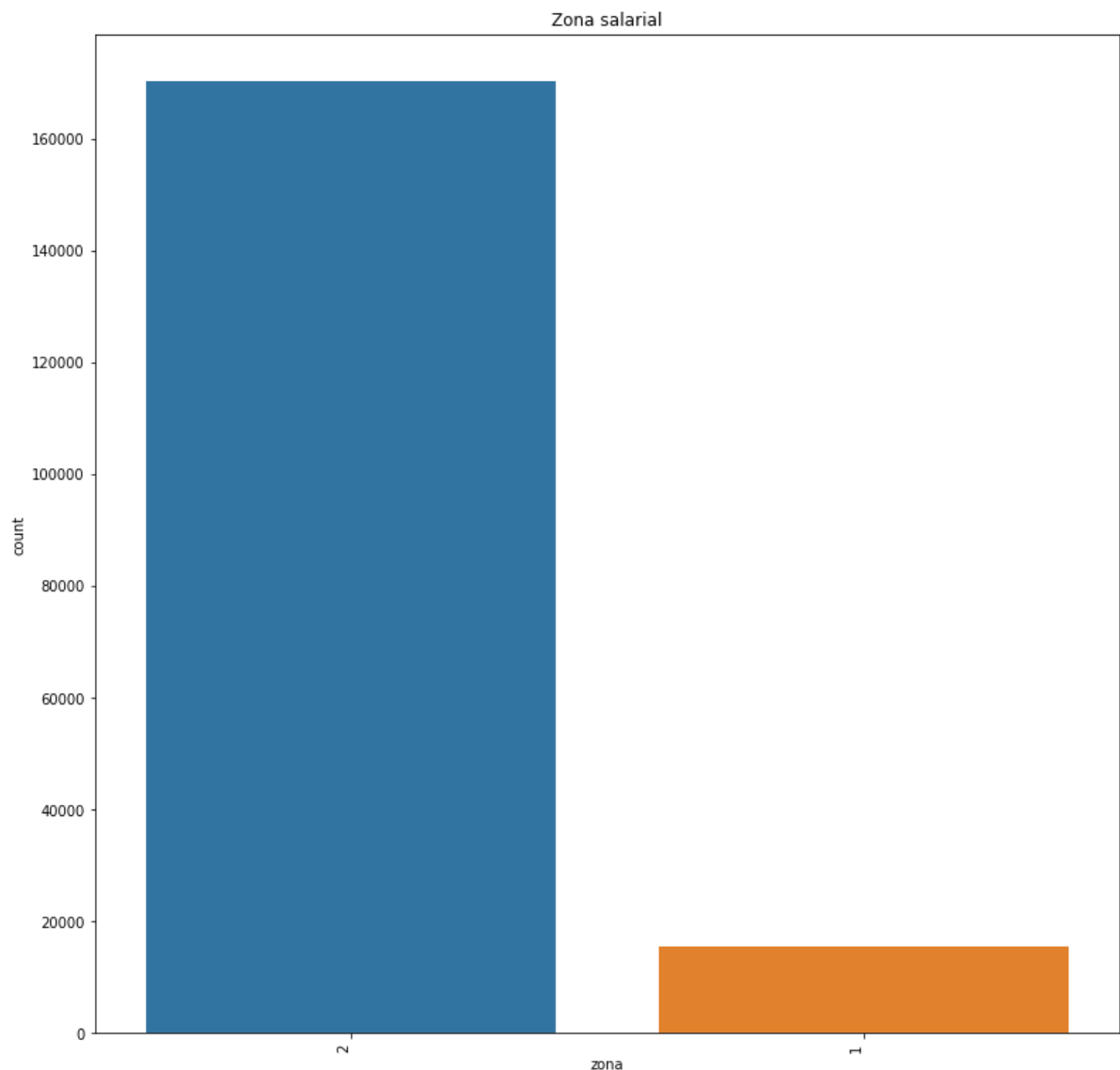
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[55]: Text(0.5, 1.0, 'Zona salarial')



La zona salarial que predomina es la de "Resto del pais" debido a que son pocos los estados que se encuentran en la zona fronteriza.

Variable de clase

```
In [56]: sns.countplot(df['clase1'])
sns.countplot(df['clase1']).set_xticklabels(sns.countplot(df['clase1']).get_xticklabels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('Clase')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

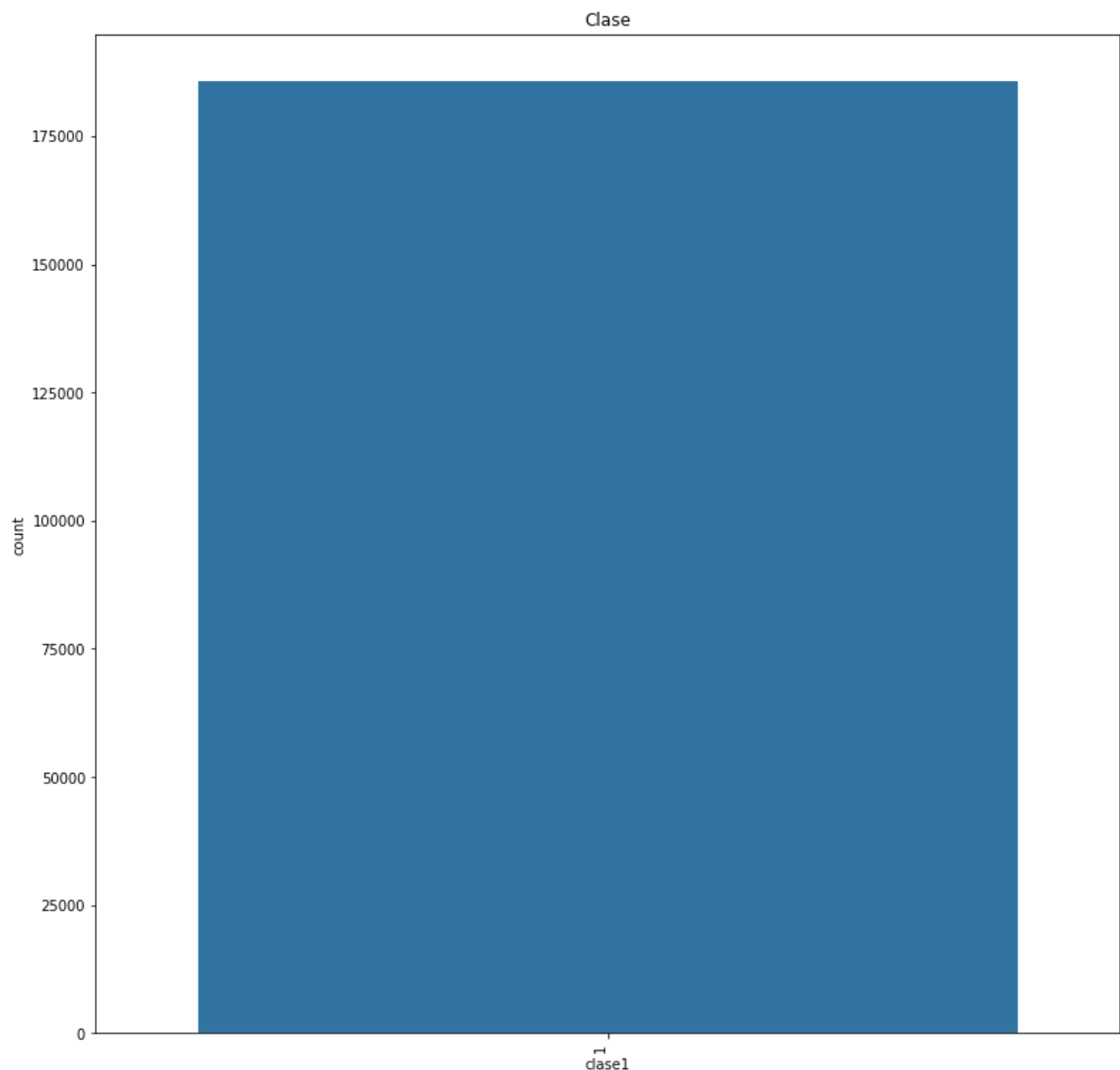
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[56]: Text(0.5, 1.0, 'Clase')



Podemos observar que la variable de clase solo tiene un valor, esto debido a que el grupo que se esta analizando solo corresponde a la poblacion economicamente activa.

Clasificacion por condicion de actividad

```
In [57]: sns.countplot(df['domestico'])  
sns.countplot(df['domestico']).set_xticklabels(sns.countplot(df['domestico']).  
get_xticklabels(), rotation=90, ha="right")  
fig = plt.gcf()  
fig.set_size_inches(13,13)  
plt.title('Clasificacion por condicion de actividad')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

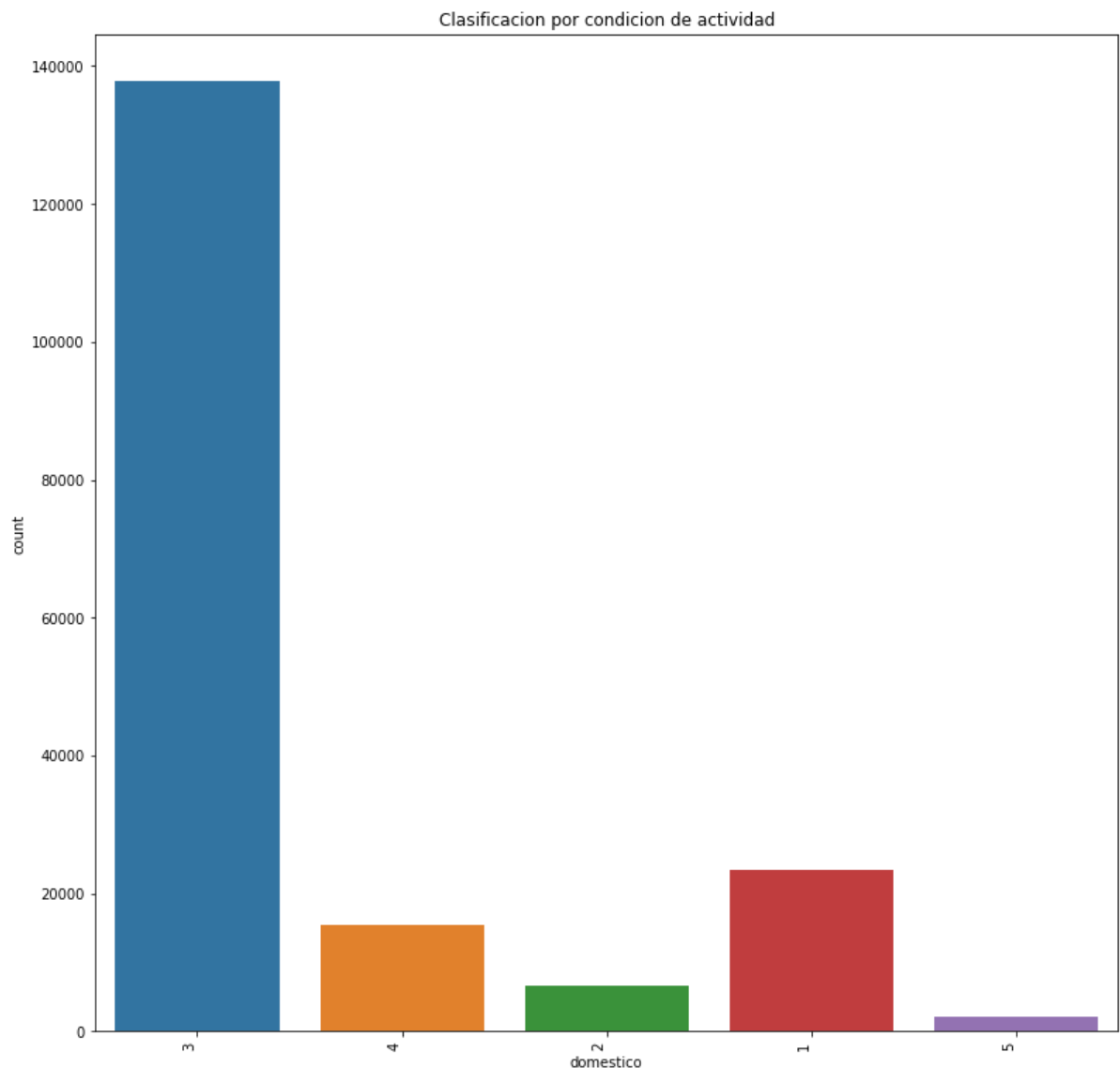
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[57]: Text(0.5, 1.0, 'Clasificacion por condicion de actividad')



La actividad que sobresale de las demás es que además de su ocupación se dedican a los quehaceres domésticos.

Años de escolaridad

```
In [58]: sns.countplot(df['anios_esc'])
sns.countplot(df['anios_esc']).set_xticklabels(sns.countplot(df['anios_esc']).
get_xticklabels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('Años escolaridad')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

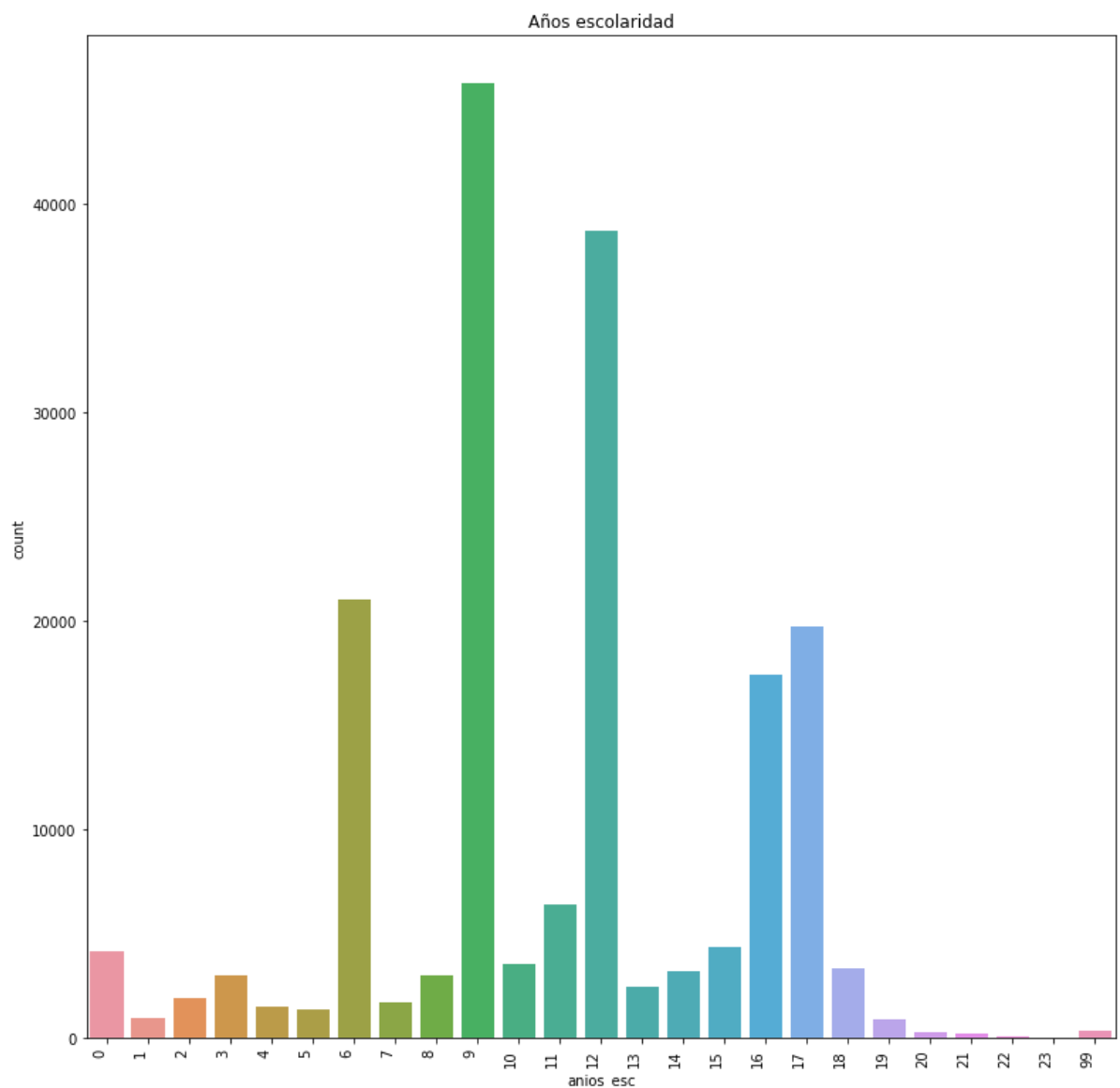
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[58]: Text(0.5, 1.0, 'Años escolaridad')



Los años de escolaridad que mas se repiten son 9, que corresponde a un nivel de secundaria.

Total de trabajos

```
In [60]: sns.countplot(df['t_tra'])
sns.countplot(df['t_tra']).set_xticklabels(sns.countplot(df['t_tra']).get_xtic
klabels(), rotation=90, ha="right")
fig = plt.gcf()
fig.set_size_inches(13,13)
plt.title('total de trabajos')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

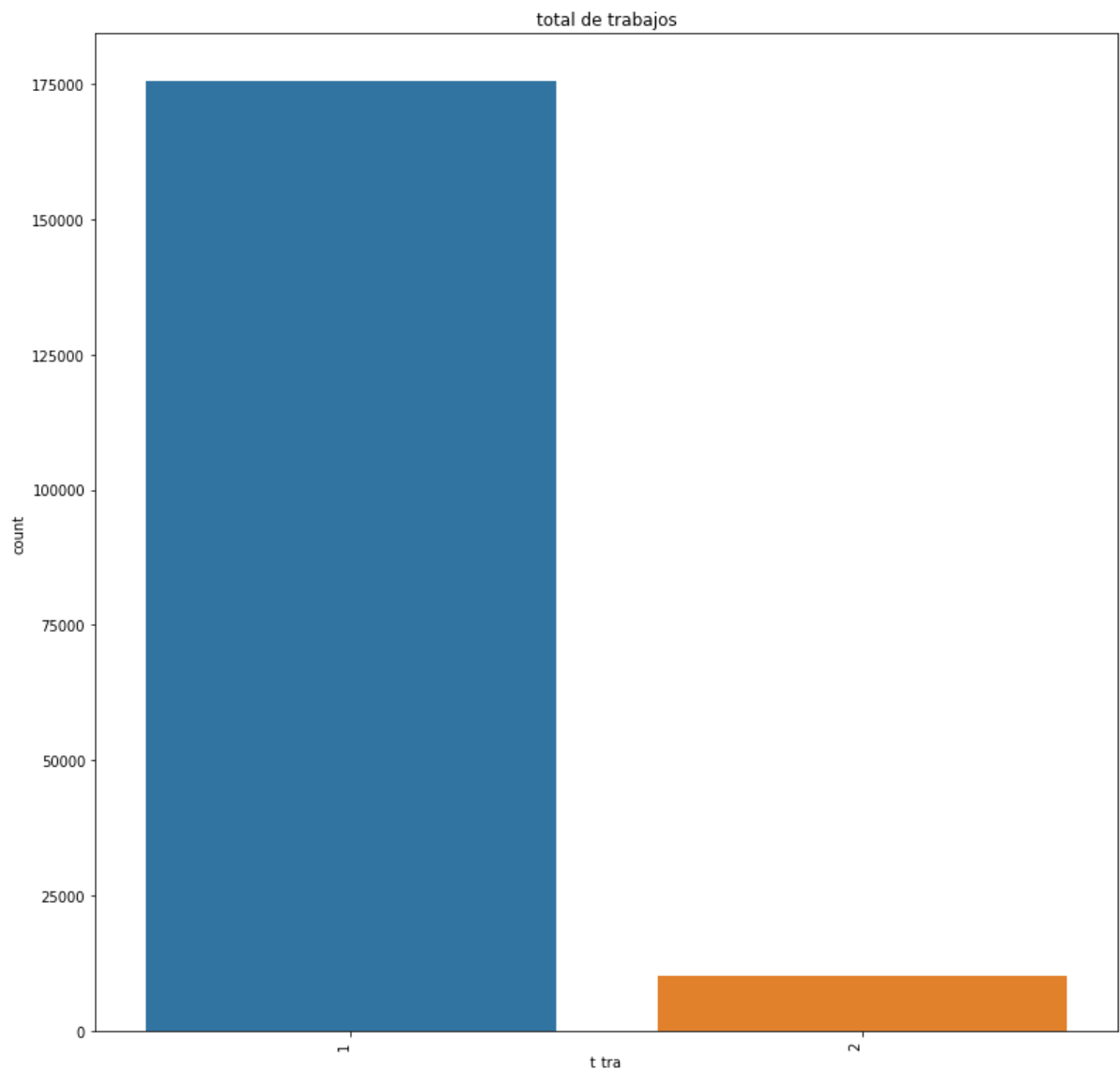
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[60]: Text(0.5, 1.0, 'total de trabajos')



La mayoría de las personas entrevistadas solo tiene un trabajo.

Horas ocupadas

```
In [61]: sns.countplot(df['hrsocup'])  
sns.countplot(df['hrsocup']).set_xticklabels(sns.countplot(df['hrsocup']).get_  
xticklabels(), rotation=90, ha="right")  
fig = plt.gcf()  
fig.set_size_inches(13,13)  
plt.title('horas ocupadas')
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

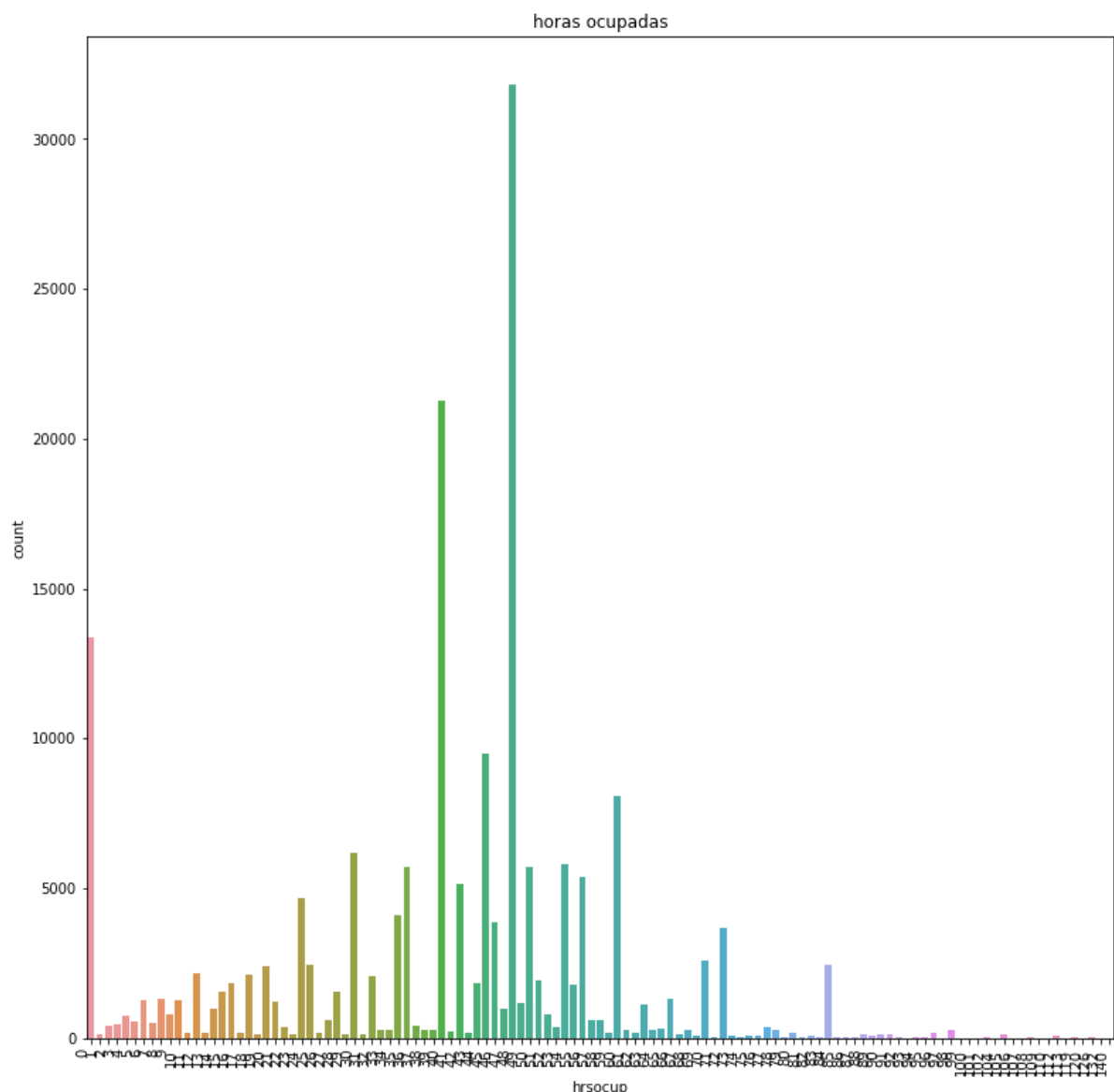
```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
```

```
warnings.warn(
```

Out[61]: Text(0.5, 1.0, 'horas ocupadas')

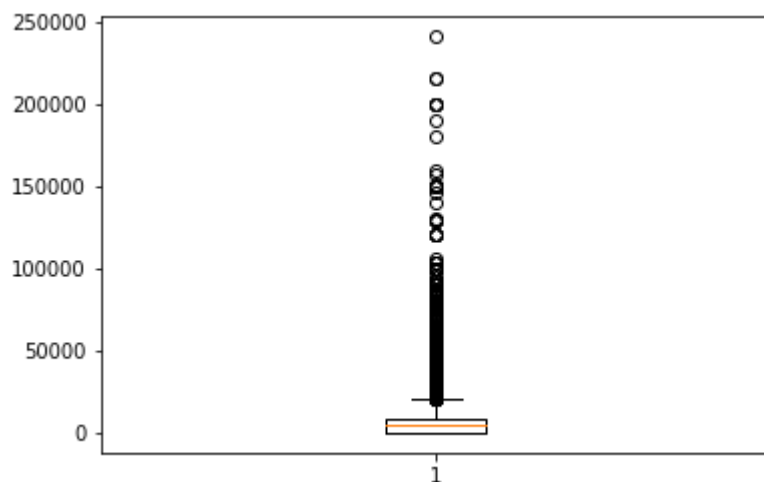


Las horas ocupadas mas frecuentes son de una jornada de 49 horas trabajadas a la semana.

Ingreso mensual

```
In [65]: import matplotlib.pyplot as plt  
plt.boxplot(df['ingocup'])
```

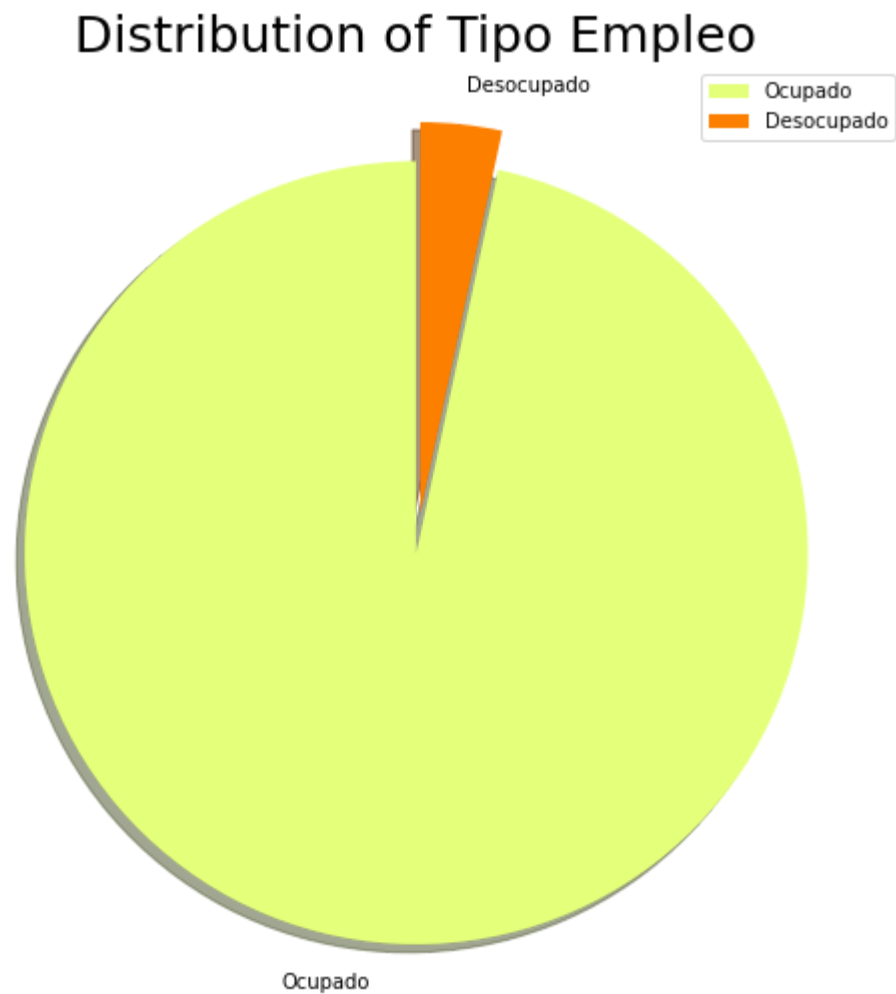
```
Out[65]: {'whiskers': [<matplotlib.lines.Line2D at 0x7f4b1373d5b0>,  
  <matplotlib.lines.Line2D at 0x7f4b137e5760>],  
  'caps': [<matplotlib.lines.Line2D at 0x7f4b137e5580>,  
  <matplotlib.lines.Line2D at 0x7f4b137d64c0>],  
  'boxes': [<matplotlib.lines.Line2D at 0x7f4b1373df10>],  
  'medians': [<matplotlib.lines.Line2D at 0x7f4b137d6820>],  
  'fliers': [<matplotlib.lines.Line2D at 0x7f4b13791610>],  
  'means': []}
```



Se puede observar que hay datos outliers que se ocuparian remover de la base para poder realizar el analisis

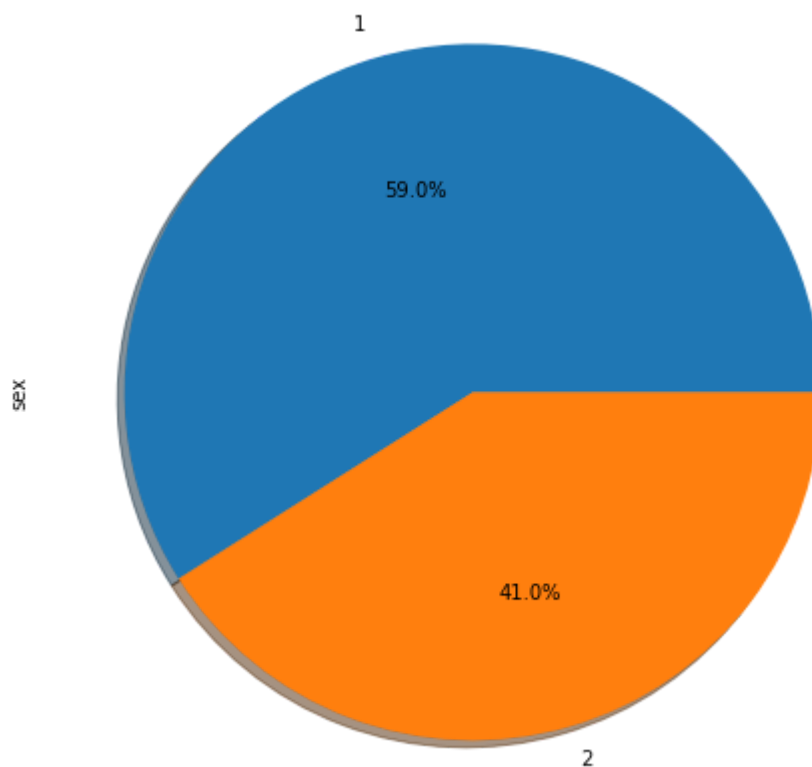
Pie-chart for the Tipo de Empleo: Ocupado o Desocupado

```
In [66]: labels = ['Ocupado', 'Desocupado']
size = df['clase2'].value_counts()
colors = plt.cm.Wistia(np.linspace(0, 1, 2))
explode = [0, 0.1]
plt.rcParams['figure.figsize'] = (9, 9)
plt.pie(size, labels=labels, colors = colors, explode = explode, shadow = True,
startangle = 90)
plt.title('Distribution of Tipo Empleo', fontsize = 25)
plt.legend()
plt.show()
```



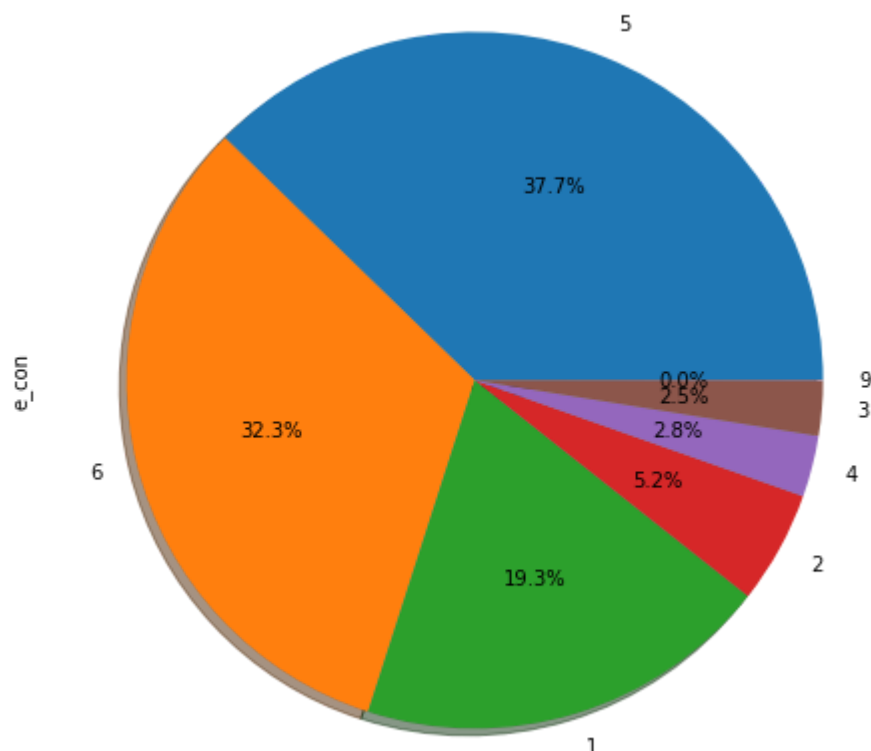
Pie-chart for genero

```
In [67]: df['sex'].value_counts().plot.pie(autopct='%1.1f%%', shadow=True, figsize=(10, 8))  
plt.show()
```



Pay chart de nivel de estado conyugal

```
In [68]: df['e_con'].value_counts().plot.pie(autopct='%1.1f%%', shadow=True, figsize=(10, 8))
plt.show()
```



Primeras impresiones

Realizando este analisis inicial de la base de datos, se pudo observar que la mayoría de las variables son categóricas, que existen datos con espacios que se ocuparían remover. Adicionalmente, la variable de interés, que es el tipo de empleo, es decir, si la persona se encuentra empleada o desempleada, está desbalanceada por lo que se tendría que aplicar una técnica de balanceo. Además, las variables numéricas que se tienen son de diferentes dimensiones por lo que se tendría que aplicar una técnica de estandarización de los datos. Asimismo, se tienen variables que repiten información o no aporta información relevante como es el caso de la variable clase1, por lo que se tendría que revisar con cuáles variables nos quedaríamos para realizar el análisis.

In []: