

4 - Parte EDA & Limpieza - Sección NHCEP

October 31, 2022

1 Análisis de la Encuesta Multiproposito

1.1 Importación de Paquetes y carga del archivo

```
[1]: import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
from scipy import stats
import numpy as np
```

```
[2]: data = pd.read_excel('Encuesta_Multiproposito_Suba.xlsx')
```

```
[3]: data.shape
```

```
[3]: (24536, 498)
```

Hay en total 24536 encuestados en la Localidad de Suba

2 Cuarta Sección NHCEP

2.1 NPCEP6

2.1.1 7. ¿Cuál es el parentesco de ...con el o la jefe(a) de este hogar?

1. Jefe(a) del hogar
2. Pareja, esposo(a),cónyuge, compañero(a)
3. Hijo(a), hijastro(a)
4. Nieto(a)
5. Padre, madre, padrastro, madrastra
6. Suegro o suegra
7. Hermano(a), hermanastro (a)
8. Yerno, nuera
9. Otro(a) pariente del(a) jefe(a)

10. Empleado(a) del servicio doméstico
11. Parientes del servicio doméstico
12. Trabajador
13. Pensionista
14. Otro(a) no pariente

Datos: 24536

```
[4]: data['NPCEP6'].count()
```

```
[4]: 24536
```

```
[5]: data['NPCEP6'].describe().apply("{0:.2f}".format)
```

```
[5]: count      24536.00
     mean        2.45
     std         1.96
     min         1.00
     25%         1.00
     50%         2.00
     75%         3.00
     max         14.00
     Name: NPCEP6, dtype: object
```

```
[6]: data['NPCEP6'].value_counts()
```

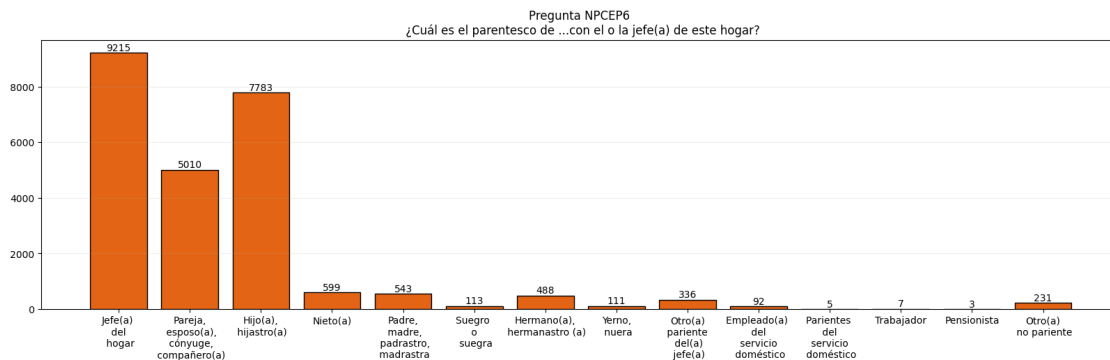
```
[6]: 1      9215
     3      7783
     2      5010
     4       599
     5       543
     7       488
     9       336
    14       231
     6       113
     8       111
    10        92
    12         7
    11         5
    13         3
     Name: NPCEP6, dtype: int64
```

```
[7]: plt.figure(figsize=(20,5))
     bars = plt.bar(data['NPCEP6'].value_counts().index.tolist(),data['NPCEP6'].
     ↪value_counts().tolist(), edgecolor = 'k', color = '#e36414')
     plt.xticks([1,2,3,4,5,6,7,8,9,10,11,12,13,14], ['Jefe(a) \n del \n hogar',
     'Pareja,\n esposo(a),\n cónyuge,\n compañero(a)',
     'Hijo(a),\n hijastro(a)',
     'Nieto(a)',
     'Padre,\n madre,\n padrastro,\n madrastra',
     'Suegro \n o \n suegra',
     'Hermano(a),\n hermanastro (a)',
     'Yerno,\n nuera',
     'Otro(a)\n pariente \n del(a)\n jefe(a)',
```

```

'Empleado(a) \n del \n servicio \n doméstico',
'Parientes \n del \n servicio \n doméstico',
'Trabajador',
'Pensionista',
'Otro(a)\n no pariente'])
plt.title('Pregunta NPCEP6 \n ¿Cuál es el parentesco de ...con el o la jefe(a) de este hogar?')
plt.bar_label(bars)
plt.grid(alpha = 0.2, axis = 'y')
plt.show()

```



2.2 NPCEP7

2.2.1 8. ¿Cuál es el estado civil actual de ...?

1. No está casado(a) y vive en pareja hace menos de dos años
2. No está casado(a) y vive en pareja hace dos años o más
3. Está viudo(a)
4. Está separado(a) o divorciado(a)
5. Está soltero(a)
6. Está casado(a)

Datos: 22213

```
[8]: data['NPCEP7'].count()
```

```
[8]: 22213
```

```
[9]: data['NPCEP7'].describe().apply("{0:.2f}".format)
```

```

[9]: count    22213.00
     mean         4.67
     std         1.42
     min         1.00
     25%         4.00

```

```

50%          5.00
75%          6.00
max           6.00
Name: NPCEP7, dtype: object

```

```
[10]: data['NPCEP7'].value_counts()
```

```

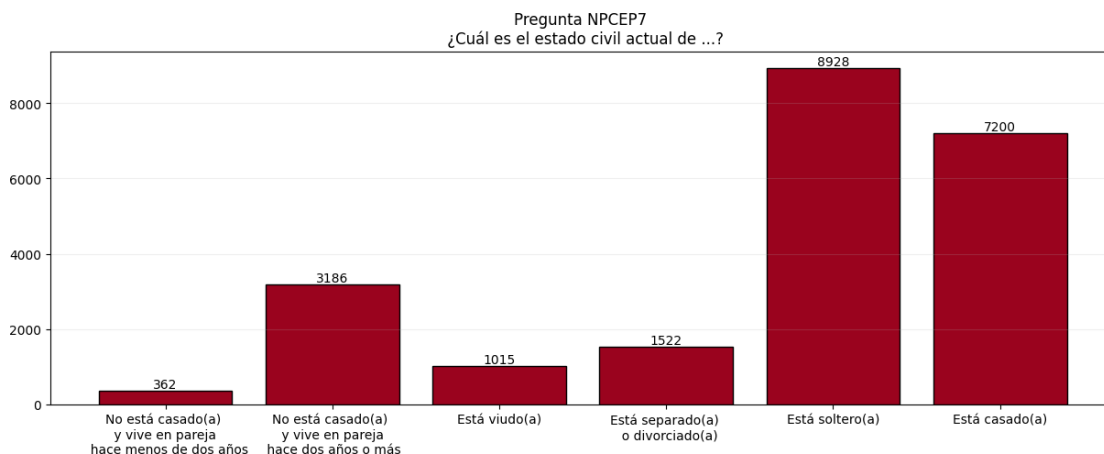
[10]: 5.0      8928
      6.0      7200
      2.0      3186
      4.0      1522
      3.0      1015
      1.0       362
Name: NPCEP7, dtype: int64

```

```

[11]: plt.figure(figsize=(15,5))
      bars = plt.bar(data['NPCEP7'].value_counts().index.tolist(),data['NPCEP7'].
      ↪value_counts().tolist(), edgecolor = 'k', color = '#9a031e')
      plt.xticks([1,2,3,4,5,6], ['No está casado(a) \n y vive en pareja \n hace_
      ↪menos de dos años',
      'No está casado(a) \n y vive en pareja \n hace dos años o más',
      'Está viudo(a)',
      'Está separado(a) \n o divorciado(a)',
      'Está soltero(a)',
      'Está casado(a)'])
      plt.title('Pregunta NPCEP7 \n ¿Cuál es el estado civil actual de ...?')
      plt.bar_label(bars)
      plt.grid(alpha = 0.2, axis = 'y')
      plt.show()

```



2.3 NPCEP8

2.3.1 9. ¿El (la) cónyuge de ...vive en este hogar?

Datos 10748

```
[12]: data['NPCEP8'].count()
```

```
[12]: 10748
```

```
[13]: data['NPCEP8'].describe().apply("{0:.2f}".format)
```

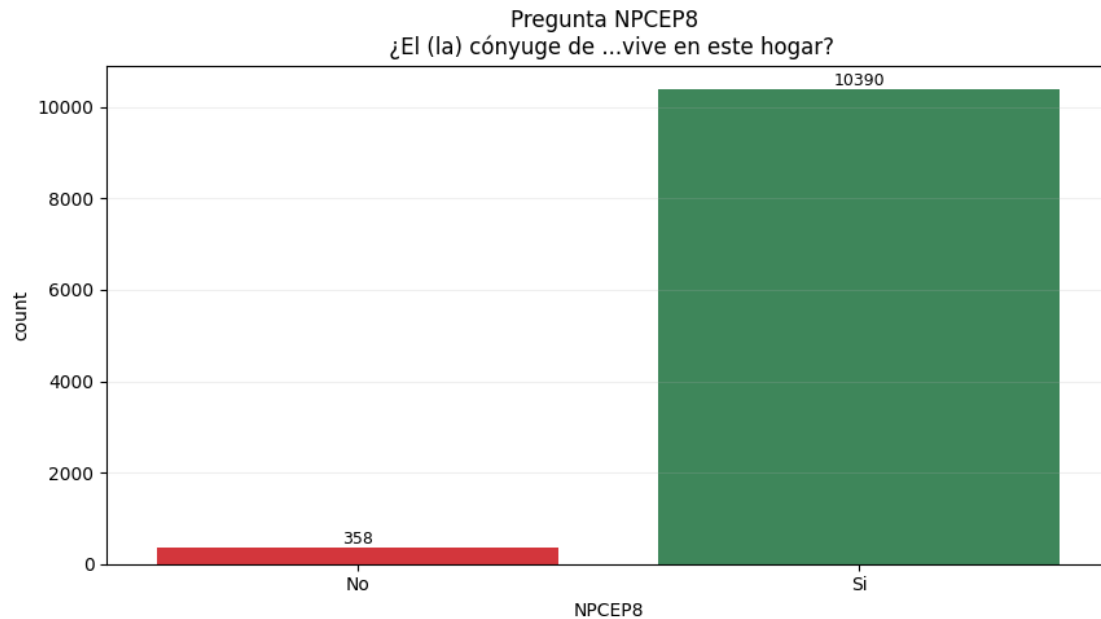
```
[13]: count      10748.00  
      mean        1.03  
      std         0.18  
      min         1.00  
      25%         1.00  
      50%         1.00  
      75%         1.00  
      max         2.00  
      Name: NPCEP8, dtype: object
```

```
[14]: data = data.replace({'NPCEP8':2},0)
```

```
[15]: data['NPCEP8'].value_counts()
```

```
[15]: 1.0      10390  
      0.0       358  
      Name: NPCEP8, dtype: int64
```

```
[16]: fig, ax = plt.subplots(figsize=(10, 5))  
  
g = sns.countplot(ax=ax, data = data, x = 'NPCEP8', palette= ["#ED1C24", "#329257"])  
  
for bars in ax.containers:  
    ax.bar_label(bars, fmt='%.0f', fontsize=9)  
  
ax.set_title('Pregunta NPCEP8 \n ¿El (la) cónyuge de ...vive en este hogar?')  
ax.set_xticklabels(['No', 'Si'])  
  
plt.grid(alpha = 0.2, axis = 'y')  
plt.show()
```



2.4 NPCEP8A

2.4.1 9a. Núm. de orden

Contestan 10390

```
[17]: data['NPCEP8A'].count()
```

```
[17]: 10390
```

```
[18]: data['NPCEP8A'].describe().apply("{0:.2f}".format)
```

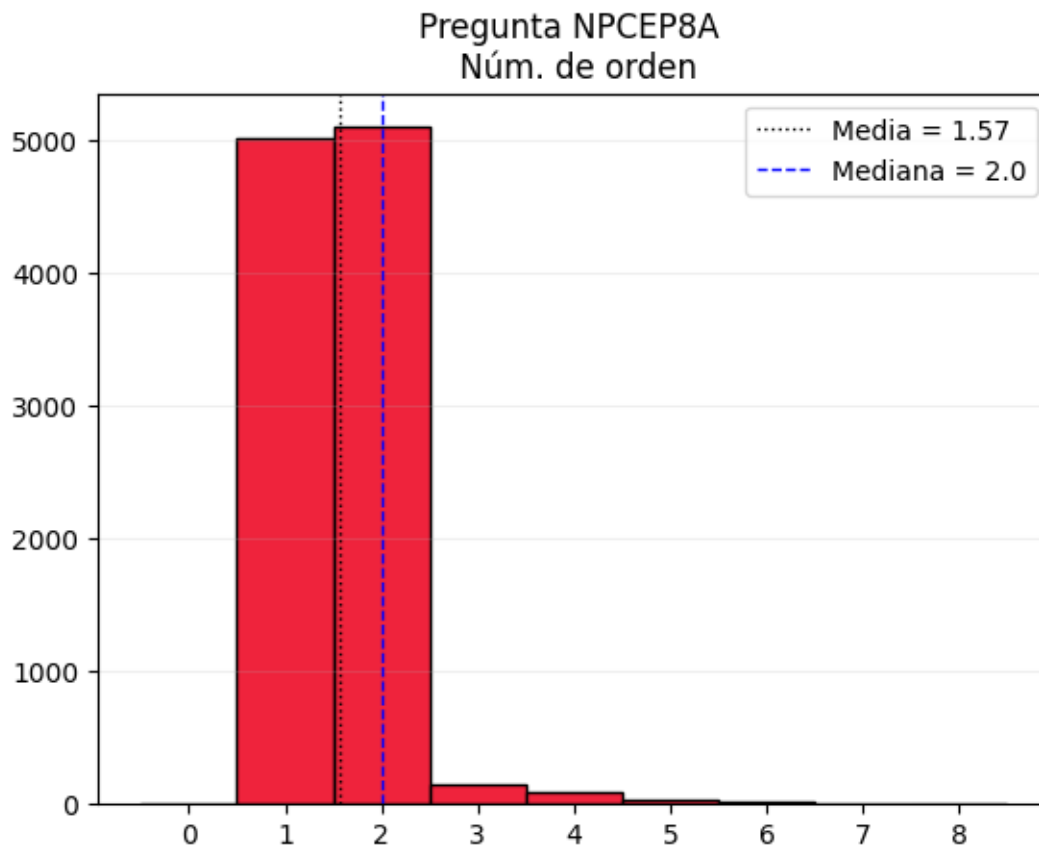
```
[18]: count      10390.00
      mean         1.57
      std          0.63
      min          1.00
      25%          1.00
      50%          2.00
      75%          2.00
      max           8.00
      Name: NPCEP8A, dtype: object
```

```
[19]: counts, edges, bars = plt.hist(data['NPCEP8A'], bins = np.arange(10) - 0.5 ,
      ↪edgecolor = 'black', color = '#ef233c')
      #ticklabels = [i for i in range(5)]
      plt.xticks(range(9))
      #plt.xticks(np.linspace(0, data['NPCEP8A'].max(),6),rotation = 45)
```

```

plt.bar_label(bars)
plt.title('Pregunta NPCEP8A \n Núm. de orden')
plt.xlim([-1,10])
plt.axvline(data['NPCEP8A'].mean(), color='k', linestyle=':', linewidth=1,
            label = f'Media = {data["NPCEP8A"].mean():.2f}')
plt.axvline(data['NPCEP8A'].median(), color='b', linestyle='dashed',
            linewidth=1, label = f'Mediana = {data["NPCEP8A"].median():.2f}')
plt.legend()
plt.grid(alpha = 0.2, axis = 'y')
plt.show()

```



2.5 NPCEP9

2.5.1 10. ¿Dónde vivía la madre de ...cuando ...nació?

1. En este municipio
2. En otro municipio
3. En otro país

Datos: 24536

```
[20]: data['NPCEP9'].count()
```

```
[20]: 24536
```

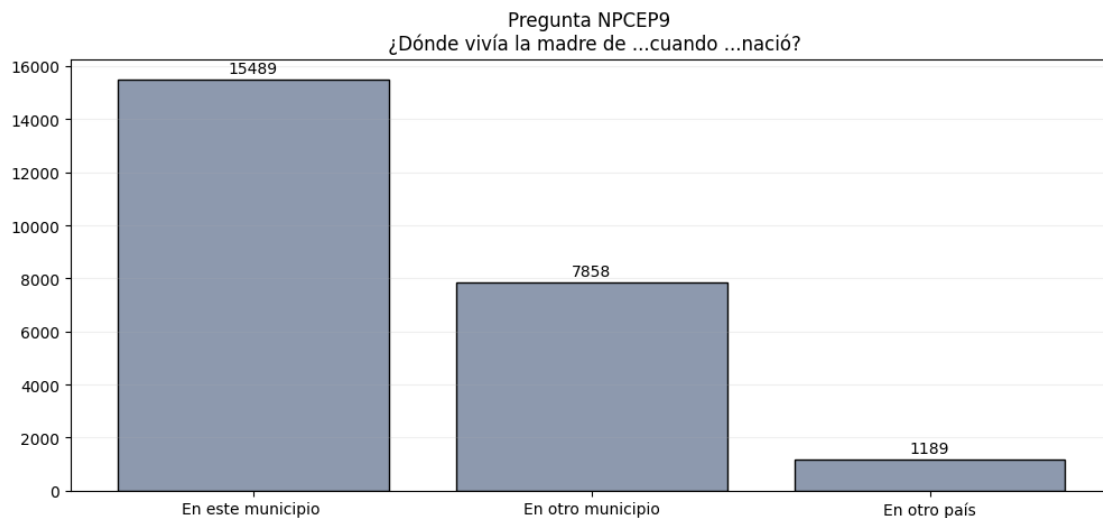
```
[21]: data['NPCEP9'].describe().apply("{0:.2f}".format)
```

```
[21]: count      24536.00
      mean        1.42
      std         0.58
      min         1.00
      25%         1.00
      50%         1.00
      75%         2.00
      max         3.00
      Name: NPCEP9, dtype: object
```

```
[22]: data['NPCEP9'].value_counts()
```

```
[22]: 1      15489
      2       7858
      3       1189
      Name: NPCEP9, dtype: int64
```

```
[23]: plt.figure(figsize=(12,5))
      bars = plt.bar(data['NPCEP9'].value_counts().index.tolist(),data['NPCEP9'].
      ↪value_counts().tolist(), edgecolor = 'k', color = '#8d99ae')
      plt.xticks([1,2,3], ['En este municipio ', 'En otro municipio', 'En otro país'])
      plt.title('Pregunta NPCEP9 \n ¿Dónde vivía la madre de ...cuando ...nació?')
      plt.bar_label(bars, padding = 2)
      plt.grid(alpha = 0.2, axis = 'y')
      plt.show()
```



2.6 NPCEP9A

2.6.1 10a. ¿Dónde vivía la madre de ...cuando ...nació? ¿Departamento?

Datos: 7858

Agregar departamento por código

```
[24]: data['NPCEP9A'].count()
```

```
[24]: 7858
```

```
[25]: data['NPCEP9A'].value_counts()
```

```
[25]: 15.0    1642
      25.0    1464
      68.0     752
      73.0     630
       5.0     308
      76.0     302
      54.0     272
      41.0     264
       8.0     261
      17.0     248
      13.0     222
      23.0     215
      47.0     187
      50.0     154
      70.0     148
      52.0     138
      20.0     137
      63.0      92
      66.0      77
      19.0      70
      44.0      70
      18.0      66
      85.0      48
      27.0      25
      81.0      23
      86.0      18
      88.0      10
      91.0       6
      95.0       4
      97.0       3
      94.0       1
      99.0       1
```

Name: NPCEP9A, dtype: int64

2.7 NPCEP9B

2.7.1 10b. ¿Dónde vivía la madre de ...cuando ...nació? ¿Municipio?

Datos: 7858

Agregar Municipio por código

```
[26]: data['NPCEP9B'].count()
```

```
[26]: 7858
```

```
[27]: data['NPCEP9B'].value_counts()
```

```
[27]: 8001.0      233
      68001.0   231
      15001.0   184
      73001.0   176
      5001.0    170
      ...
      54599.0     1
      54660.0     1
      81065.0     1
      15114.0     1
      23672.0     1
      Name: NPCEP9B, Length: 789, dtype: int64
```

2.8 NPCEP9C

2.8.1 10c. ¿Dónde vivía la madre de ...cuando ...nació? En otro país

1. Venezuela
2. Estados Unidos
3. Ecuador
4. España
5. Perú
6. Otro

Datos: 1189

```
[28]: data['NPCEP9C'].count()
```

```
[28]: 1189
```

```
[29]: data['NPCEP9C'].describe().apply("{0:.2f}".format)
```

```
[29]: count      1189.00
      mean        1.69
      std         1.65
      min         1.00
      25%         1.00
```

```
50%          1.00
75%          1.00
max           6.00
Name: NPCEP9C, dtype: object
```

```
[30]: data['NPCEP9C'].value_counts()
```

```
[30]: 1.0    988
      6.0    140
      2.0    26
      4.0    15
      3.0    14
      5.0     6
      Name: NPCEP9C, dtype: int64
```

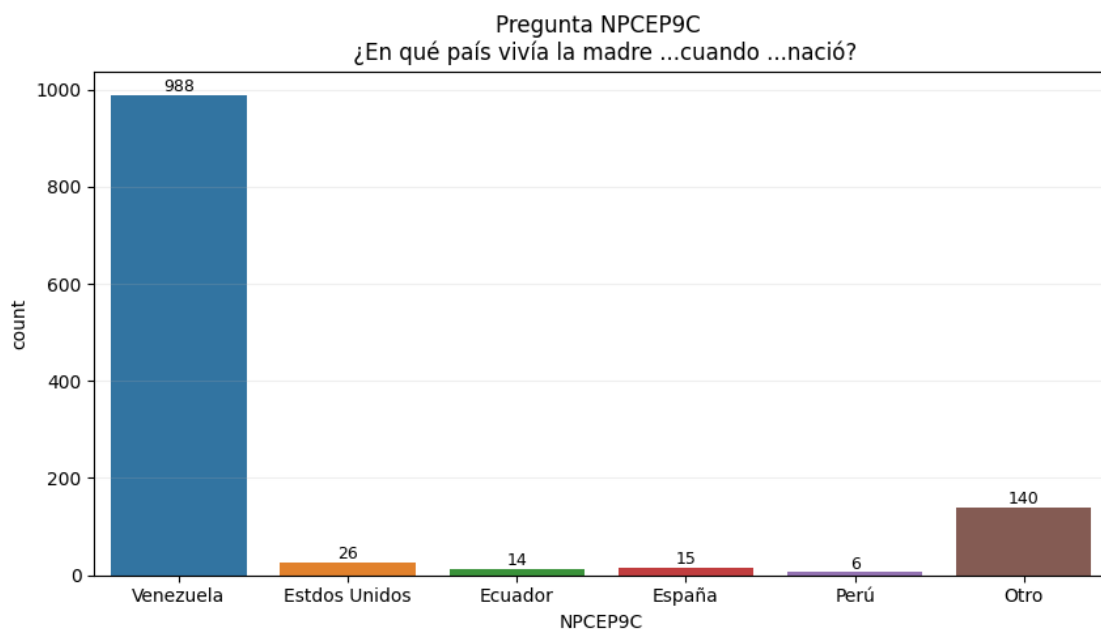
```
[31]: fig, ax = plt.subplots(figsize=(10, 5))

g = sns.countplot(ax=ax, data = data, x = 'NPCEP9C')

for bars in ax.containers:
    ax.bar_label(bars, fmt='%.0f', fontsize=9)

ax.set_title('Pregunta NPCEP9C \n ¿En qué país vivía la madre ...cuando ...
↳ nació?')
ax.set_xticklabels(['Venezuela', 'Estdos Unidos', '↳
↳ Ecuador', 'España', 'Perú', 'Otro'])

plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



```
[32]: data.to_excel('Encuesta_Multiproposito_Suba.xlsx', index = False)
```