3 - Parte EDA & Limpieza - Sección NHCDP

December 14, 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)
```

[4]:
```
data.head()
```

[4]:
```
   COD_LOCALIDAD NOMBRE_LOCALIDAD  COD_UPZ_GRUPO NOMBRE_UPZ_GRUPO  \
0             11             Suba           28.0        El Rincón
1             11             Suba           28.0        El Rincón
2             11             Suba           28.0        El Rincón
3             11             Suba           28.0        El Rincón
4             11             Suba           28.0        El Rincón

   ESTRATO2021 NOMBRE_ESTRATO                     NVCBP1 NVCBP2 NVCBP3  \
0     11001122      El Rincon  Sendero o camino en tierra    NaN     Si
1     11001122      El Rincon  Sendero o camino en tierra    NaN     Si
2     11001122      El Rincon  Sendero o camino en tierra    NaN     Si
3     11001122      El Rincon  Sendero o camino en tierra    NaN     Si
4     11001122      El Rincon  Sendero o camino en tierra    NaN     Si

  NVCBP4  … NHCLP29_1L  NHCLP31AA NHCLP31AB NHCLP31AC NHCLP31BA NHCLP31BB  \
0     Si  …        NaN        1.0       1.0       1.0       1.0       1.0
1     Si  …        NaN        1.0       1.0       1.0       1.0       1.0
2     Si  …        NaN        1.0       1.0       1.0       1.0       1.0
3     Si  …        NaN        1.0       1.0       1.0       1.0       1.0
4     Si  …        NaN        1.0       1.0       1.0       1.0       1.0
```

```
  NHCLP31BC NHCLP31CA NHCLP31CB NHCLP31CC
0       1.0       1.0       2.0       1.0
1       1.0       1.0       2.0       1.0
2       1.0       1.0       2.0       1.0
3       1.0       1.0       2.0       1.0
4       1.0       1.0       2.0       1.0

[5 rows x 498 columns]
```

Hay en total 24536 encuestados en la Localidad de Suba

2 Tercera Sección (NHCDP)

2.1 NHCDP2

2.1.1 2. ¿Cuánto pagó este hogar la ÚLTIMA VEZ por el servicio de acueducto?

Datos: 22998

[4]:
```
data['NHCDP2'].count()
```

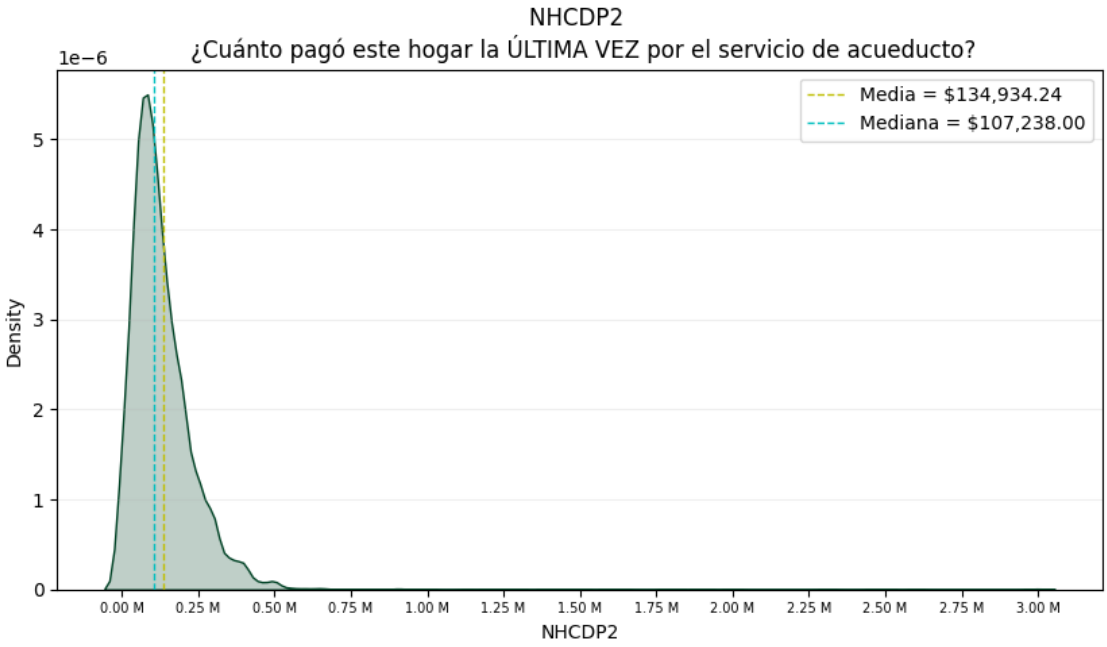
[4]:
```
22998
```

[5]:
```
data['NHCDP2'].describe().apply("{0:.2f}".format)
```

[5]:
```
count      22998.00
mean      134934.24
std       140688.75
min            0.00
25%        60000.00
50%       107238.00
75%       180000.00
max      3000000.00
Name: NHCDP2, dtype: object
```

[11]:
```
plt.figure(figsize=(10,5))
sns.kdeplot(data = data, x = 'NHCDP2', color = '#004225', fill = True)
plt.title('NHCDP2 \n ¿Cuánto pagó este hogar la ÚLTIMA VEZ por el servicio de␣
 ↪acueducto?')
plt.grid(alpha = 0.2, axis = 'y')
plt.axvline(data['NHCDP2'].mean(), color='y', linestyle='dashed', linewidth=1,␣
 ↪label = f'Media = ${data["NHCDP2"].mean():,.2f}')
plt.axvline(data['NHCDP2'].median(), color='c', linestyle='dashed',␣
 ↪linewidth=1, label = f'Mediana = ${data["NHCDP2"].median():,.2f}')
plt.xticks(range(0,3100000,250000),[f'{(i / 1000000):.2f} M' for i in␣
 ↪range(0,3100000,250000)], fontsize=7)
plt.legend()
```

```
plt.show()
```



2.2 NHCDP11

2.2.1 11. ¿Cuánto pagó este hogar la ÚLTIMA VEZ por el servicio de energía eléctrica?

Datos 23186

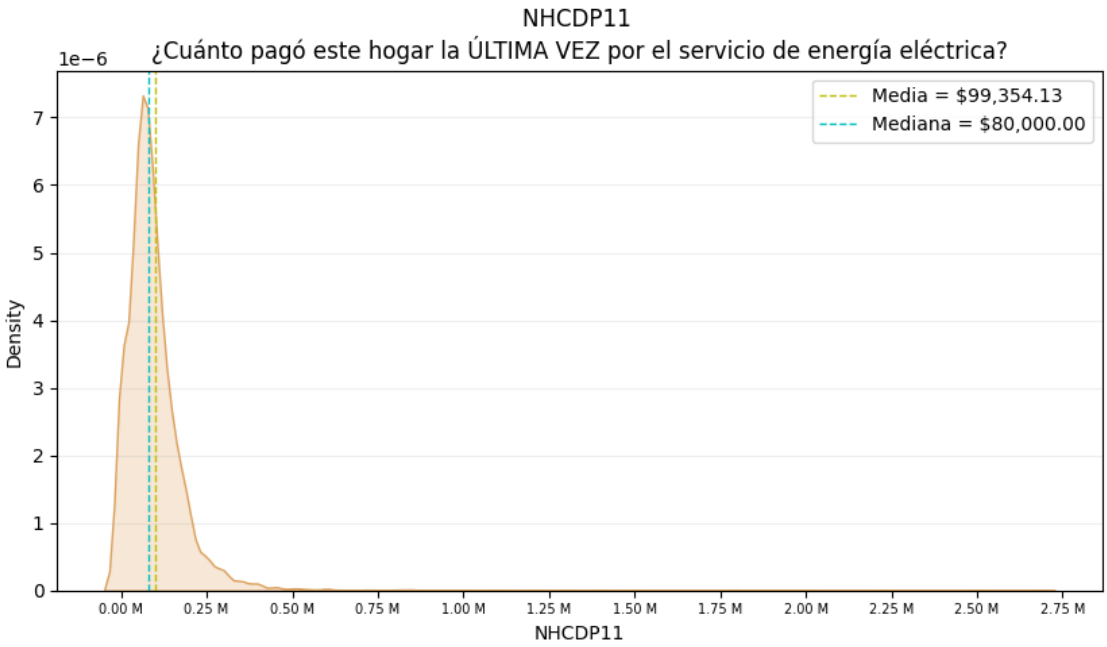
[12]:
```
data['NHCDP11'].count()
```

[12]:
```
23186
```

[13]:
```
data['NHCDP11'].describe().apply("{0:.2f}".format)
```

[13]:
```
count      23186.00
mean       99354.13
std       120412.43
min            0.00
25%        45744.00
50%        80000.00
75%       120000.00
max      2680000.00
Name: NHCDP11, dtype: object
```

[16]:
```
plt.figure(figsize=(10,5))
sns.kdeplot(data = data, x = 'NHCDP11', color = '#dda15e', fill = True)
plt.title('NHCDP11 \n¿Cuánto pagó este hogar la ÚLTIMA VEZ por el servicio de␣
 ↪energía eléctrica?')
plt.grid(alpha = 0.2, axis = 'y')
plt.axvline(data['NHCDP11'].mean(), color='y', linestyle='dashed', linewidth=1,␣
 ↪label = f'Media = ${data["NHCDP11"].mean():,.2f}')
plt.axvline(data['NHCDP11'].median(), color='c', linestyle='dashed',␣
 ↪linewidth=1, label = f'Mediana = ${data["NHCDP11"].median():,.2f}')
plt.xticks(range(0,3000000,250000),[f'{(i / 1000000):.2f} M' for i in␣
 ↪range(0,3000000,250000)], fontsize=7)
plt.legend()
plt.show()
```



2.3 NHCDP11A

2.3.1 ¿A cuántos meses correspondió el pago? (energía)

Datos: 20607

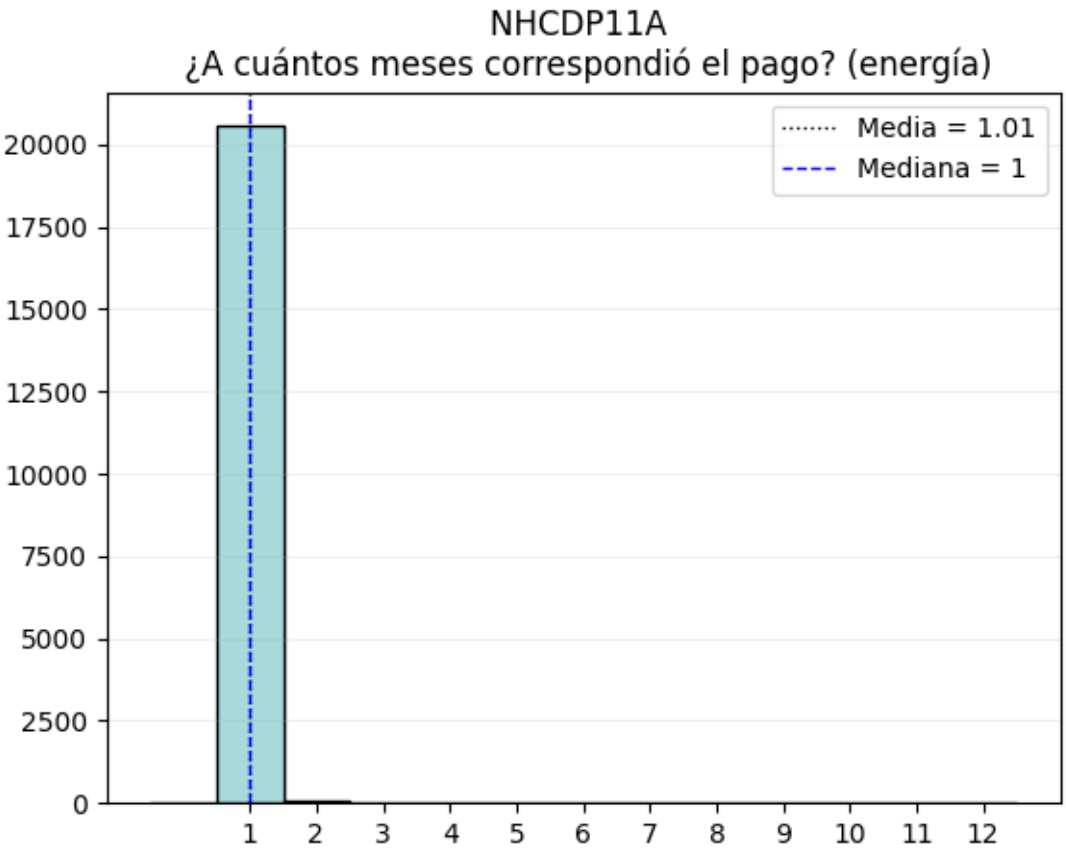
[19]:
```
data['NHCDP11A'].count()
```

[19]: 20607

[20]:
```
data['NHCDP11A'].describe().apply("{0:.2f}".format)
```

[20]:
```
count    20607.00
mean         1.01
std          0.22
min          1.00
25%          1.00
50%          1.00
75%          1.00
max         12.00
Name: NHCDP11A, dtype: object
```

[21]:
```
counts, edges, bars = plt.hist(data['NHCDP11A'], bins = np.arange(14)-0.5 ,␣
 ↪edgecolor = 'black', color = '#a8dadc')
#ticklabels = [i for i in range(5)]
#plt.xticks(range(5), ticklabels)
plt.xticks(range(1,13))
plt.title('NHCDP11A \n ¿A cuántos meses correspondió el pago? (energía)')
#plt.xlim([-1,10])
plt.axvline(data['NHCDP11A'].mean(), color='k', linestyle=':', linewidth=1,␣
 ↪label = f'Media = {data["NHCDP11A"].mean():.2f}')
plt.axvline(data['NHCDP11A'].median(), color='b', linestyle='dashed',␣
 ↪linewidth=1, label = f'Mediana = {data["NHCDP11A"].median():.0f}')
plt.legend()
plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



2.4 NHCDP12

2.4.1 12. ¿El valor pagado incluyó consumo por negocios de industria, comercio o servicios? (energía)

0. No
1. Si

Datos: 23186

[22]:
```
data['NHCDP12'].count()
```

[22]:
```
23186
```

[23]:
```
data['NHCDP12'].describe().apply("{0:.2f}".format)
```

[23]:
```
count    23186.00
mean         1.97
std          0.16
min          1.00
25%          2.00
```

```
50%          2.00
75%          2.00
max          2.00
Name: NHCDP12, dtype: object
```

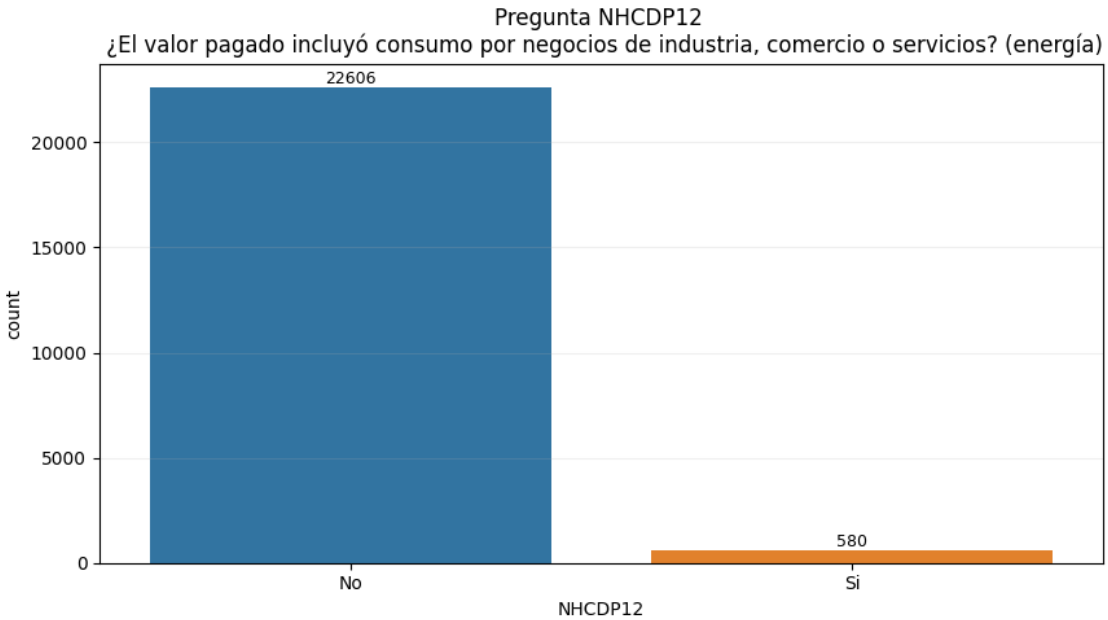
[24]:
```
data = data.replace({'NHCDP12':2},0)
```

[25]:
```
data['NHCDP12'].value_counts()
```

[25]:
```
0.0    22606
1.0      580
Name: NHCDP12, dtype: int64
```

[5]:
```
data['NHCDP12'] = data['NHCDP12'].replace([0,1],["No","Si"])
```

[6]:
```
fig, ax = plt.subplots(figsize=(10, 5))
g = sns.countplot(ax=ax, data = data, x ='NHCDP12')

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

ax.set_title('Pregunta NHCDP12 \n ¿El valor pagado incluyó consumo por negocios␣
 ↪de industria, comercio o servicios? (energía)')
#ax.set_xticklabels(['No', 'Si'])

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



2.5 NHCDP15

2.5.1 15. ¿Este hogar tiene servicio de gas natural conectado a red pública?

0. No
1. Si

Datos: 24536

[27]:
```
data['NHCDP15'].count()
```

[27]:
```
24536
```

[28]:
```
data['NHCDP15'].describe().apply("{0:.2f}".format)
```

[28]:
```
count    24536.00
mean         1.05
std          0.22
min          1.00
25%          1.00
50%          1.00
75%          1.00
max          2.00
Name: NHCDP15, dtype: object
```

[29]:
```
data = data.replace({'NHCDP15':2},0)
```

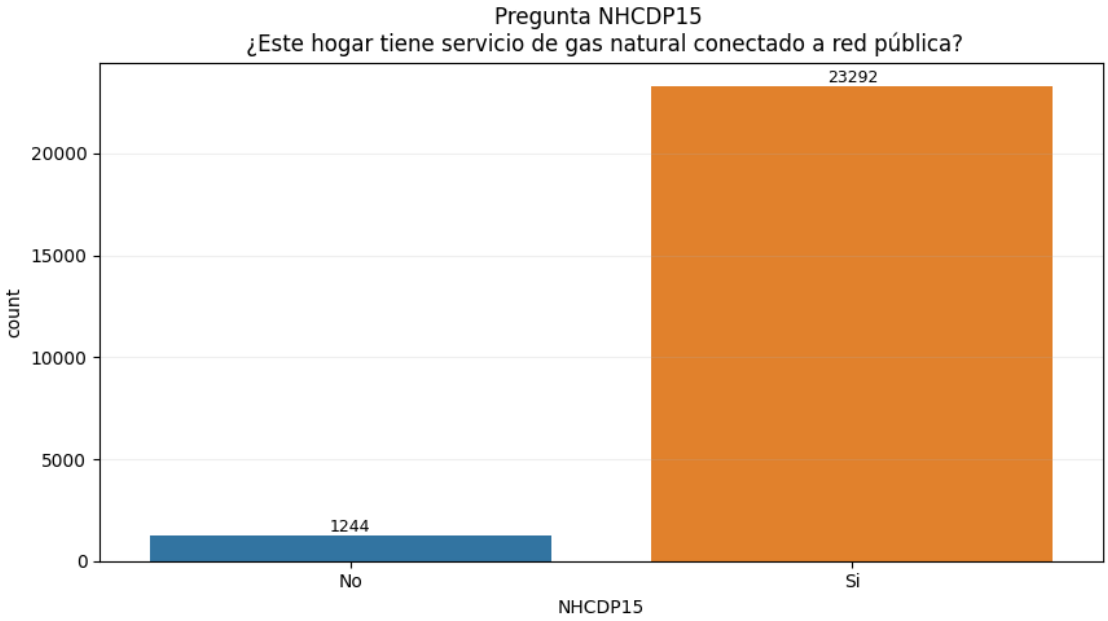
[30]:
```
data['NHCDP15'].value_counts()
```

[30]:
```
1    23292
0     1244
Name: NHCDP15, dtype: int64
```

[7]:
```
data['NHCDP15'] = data['NHCDP15'].replace([0,1],["No","Si"])
```

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

g = sns.countplot(ax=ax, data = data, x ='NHCDP15')

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

ax.set_title('Pregunta NHCDP15 \n ¿Este hogar tiene servicio de gas natural␣
 ↪conectado a red pública?')
ax.set_xticklabels(['No', 'Si'])

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

```
plt.show()
```



2.6 NHCDP17

2.6.1 17. ¿El servicio de gas natural lo pagan entre varios hogares de esta u otras viviendas?

0. No
1. Si

Datos: 24536

[32]:
```
data['NHCDP17'].count()
```

[32]:
```
22219
```

[33]:
```
data['NHCDP17'].describe().apply("{0:.2f}".format)
```

[33]:
```
count    22219.00
mean         1.95
std          0.22
min          1.00
25%          2.00
50%          2.00
75%          2.00
max          2.00
Name: NHCDP17, dtype: object
```

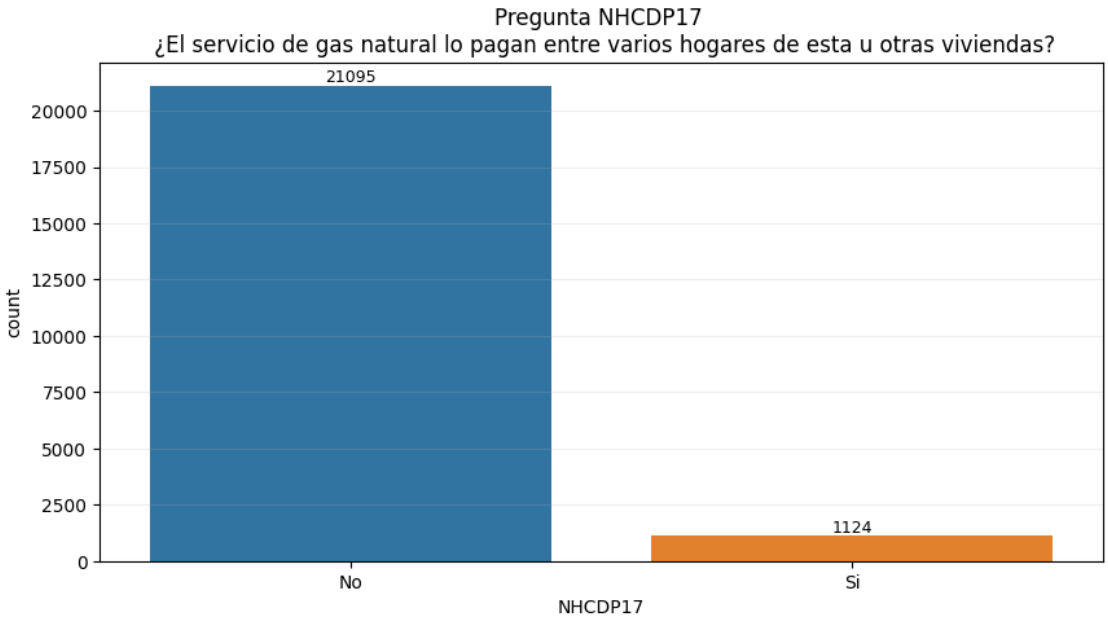
[34]:
```
data = data.replace({'NHCDP17':2},0)
```

[35]:
```
data['NHCDP17'].value_counts()
```

[35]:
```
0.0    21095
1.0     1124
Name: NHCDP17, dtype: int64
```

[9]:
```
data['NHCDP17'] = data['NHCDP17'].replace([0,1],["No","Si"])
```

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

g = sns.countplot(ax=ax, data = data, x ='NHCDP17')

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

ax.set_title('Pregunta NHCDP17 \n ¿El servicio de gas natural lo pagan entre␣
 ↪varios hogares de esta u otras viviendas?')
#ax.set_xticklabels(['No', 'Si'])

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



2.7 NHCDP17A

2.7.1 ¿Entre cuántos hogares? (gas natural)

Datos: 1124

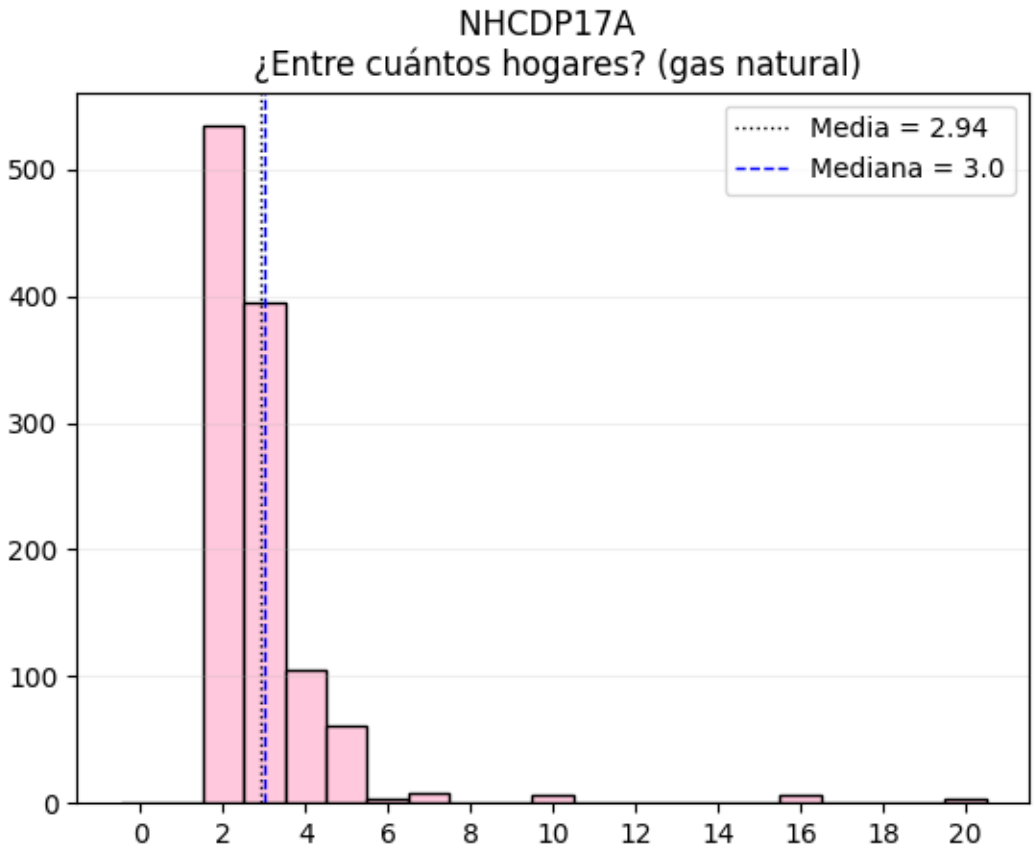
[37]:
```
data['NHCDP17A'].count()
```

[37]:
```
1124
```

[38]:
```
data['NHCDP17A'].describe().apply("{0:.2f}".format)
```

[38]:
```
count    1124.00
mean        2.94
std         1.78
min         2.00
25%         2.00
50%         3.00
75%         3.00
max        20.00
Name: NHCDP17A, dtype: object
```

[39]:
```
counts, edges, bars = plt.hist(data['NHCDP17A'], bins = np.arange(22)-0.5 ,␣
 ↪edgecolor = 'black', color = '#ffc8dd')
#ticklabels = [i for i in range(5)]
#plt.xticks(range(5), ticklabels)
plt.xticks(range(0,22,2))
#plt.bar_label(bars)
plt.title('NHCDP17A \n ¿Entre cuántos hogares? (gas natural)')
#plt.xlim([-1,10])
plt.axvline(data['NHCDP17A'].mean(), color='k', linestyle=':', linewidth=1,␣
 ↪label = f'Media = {data["NHCDP17A"].mean():.2f}')
plt.axvline(data['NHCDP17A'].median(), color='b', linestyle='dashed',␣
 ↪linewidth=1, label = f'Mediana = {data["NHCDP17A"].median()}')
plt.legend()
plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



2.8 NHCDP18

2.8.1 18. ¿Cuánto pagó este hogar la ÚLTIMA VEZ por el servicio de gas natural?

Datos: 22219

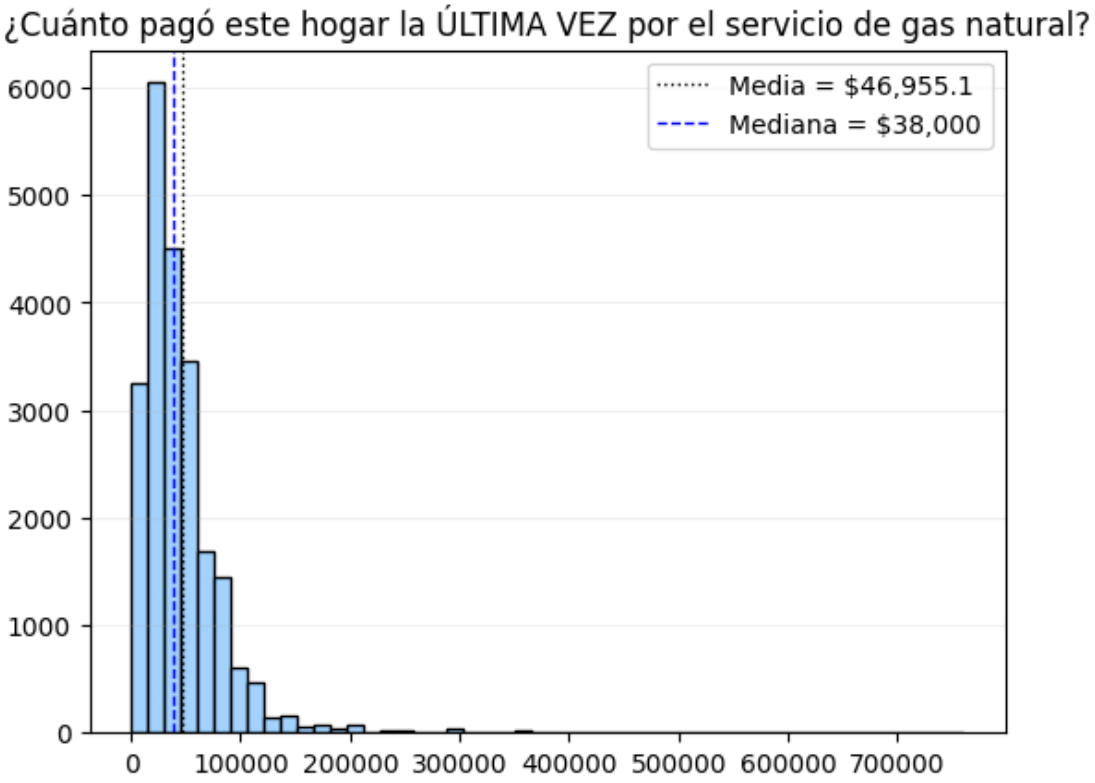
[40]:
```
data['NHCDP18'].count()
```

[40]:
```
22219
```

[41]:
```
data['NHCDP18'].describe().apply("{0:.2f}".format)
```

[41]:
```
count     22219.00
mean      46955.08
std       45457.12
min           0.00
25%       22000.00
50%       38000.00
75%       60000.00
max      760000.00
Name: NHCDP18, dtype: object
```

[43]:
```
counts, edges, bars = plt.hist(data['NHCDP18'], bins = 50 , edgecolor =␣
↪'black', color = '#a2d2ff')
#ticklabels = [i for i in range(5)]
#plt.xticks(range(5), ticklabels)
#plt.xticks(range(9))
#plt.bar_label(bars)
plt.title('¿Cuánto pagó este hogar la ÚLTIMA VEZ por el servicio de gas natural?
↪')
#plt.xlim([-1,10])
plt.axvline(data['NHCDP18'].mean(), color='k', linestyle=':', linewidth=1,␣
↪label = f'Media = ${data["NHCDP18"].mean():,.1f}')
plt.axvline(data['NHCDP18'].median(), color='b', linestyle='dashed',␣
↪linewidth=1, label = f'Mediana = ${data["NHCDP18"].median():,.0f}')
plt.legend()
plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



2.9 NHCDP18A

2.9.1 ¿A cuántos meses correspondió el pago? (gas natural)

Datos: 21523

[44]:
```
data['NHCDP18A'].count()
```

[44]:
```
21523
```

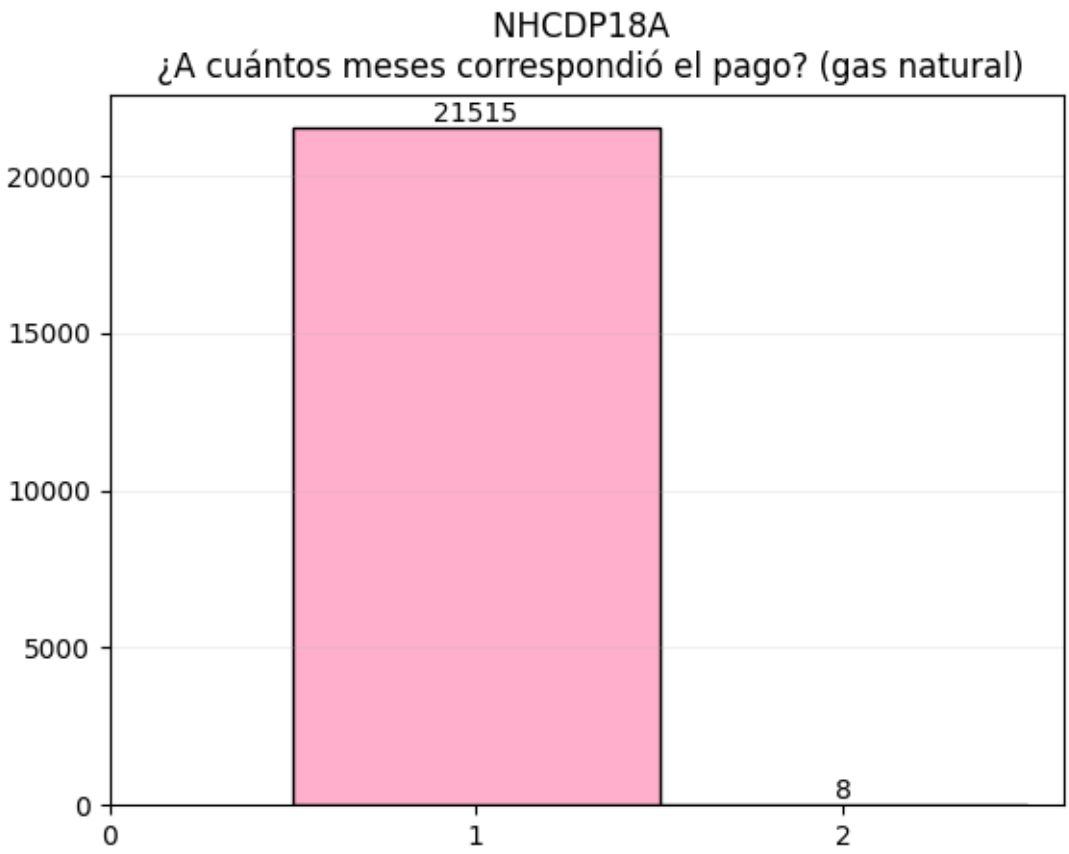
[45]:
```
data['NHCDP18A'].describe().apply("{0:.2f}".format)
```

[45]:
```
count    21523.00
mean         1.00
std          0.02
min          1.00
25%          1.00
50%          1.00
75%          1.00
max          2.00
Name: NHCDP18A, dtype: object
```

[46]:
```
data['NHCDP18A'].value_counts()
```

[46]:
```
1.0    21515
2.0        8
Name: NHCDP18A, dtype: int64
```

[47]:
```
counts, edges, bars = plt.hist(data['NHCDP18A'], bins = np.arange(1,3.5)-0.5 ,␣
 ↪edgecolor = 'black', color = '#ffafcc')
#ticklabels = [i for i in range(5)]
#plt.xticks(range(5), ticklabels)
plt.xticks(range(3))
plt.bar_label(bars)
plt.title('NHCDP18A \n ¿A cuántos meses correspondió el pago? (gas natural)')
#plt.xlim([-1,10])
#plt.axvline(data['NHCDP18A'].mean(), color='k', linestyle=':', linewidth=1,␣
 ↪label = f'Media = {data["NHCDP18A"].mean()}')
#plt.axvline(data['NHCDP18A'].median(), color='b', linestyle='dashed',␣
 ↪linewidth=1, label = f'Mediana = {data["NHCDP18A"].median()}')
#plt.legend()
plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



2.10 NHCDP19

2.10.1 19. ¿El valor pagado incluyó consumo por negocios de industria, comercio o servicios? (gas natural)

0. No
1. Si

Datos: 24536

[48]:
```
data['NHCDP19'].count()
```

[48]:
```
22219
```

[49]:
```
data['NHCDP19'].describe().apply("{0:.2f}".format)
```

[49]:
```
count    22219.00
mean         1.99
std          0.11
min          1.00
25%          2.00
```

```
50%          2.00
75%          2.00
max          2.00
Name: NHCDP19, dtype: object
```

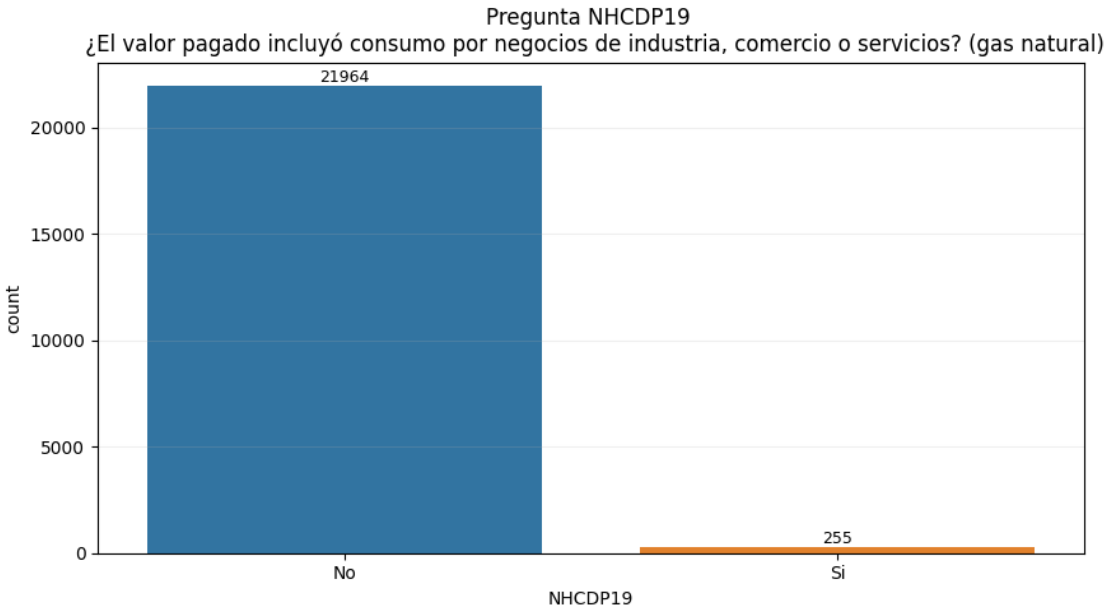
[50]:
```
data = data.replace({'NHCDP19':2},0)
```

[51]:
```
data['NHCDP19'].value_counts()
```

[51]:
```
0.0    21964
1.0      255
Name: NHCDP19, dtype: int64
```

[11]:
```
data['NHCDP19'] = data['NHCDP19'].replace([0,1],["No","Si"])
```

[12]:
```
fig, ax = plt.subplots(figsize=(10, 5))

g = sns.countplot(ax=ax, data = data, x ='NHCDP19')

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

ax.set_title('Pregunta NHCDP19 \n ¿El valor pagado incluyó consumo por negocios␣
 ↪de industria, comercio o servicios? (gas natural)')
#ax.set_xticklabels(['No', 'Si'])

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



2.11 NHCDP24

2.11.1 24. ¿En este hogar tienen computador de escritorio (PC - Desktop) o portátil (laptop)?

0. No
1. Si

Datos: 24536

[53]:
```
data['NHCDP24'].count()
```

[53]:
```
24536
```

[54]:
```
data['NHCDP24'].describe().apply("{0:.2f}".format)
```

[54]:
```
count    24536.00
mean         1.20
std          0.40
min          1.00
25%          1.00
50%          1.00
75%          1.00
max          2.00
Name: NHCDP24, dtype: object
```

[55]:
```
data = data.replace({'NHCDP24':2},0)
```

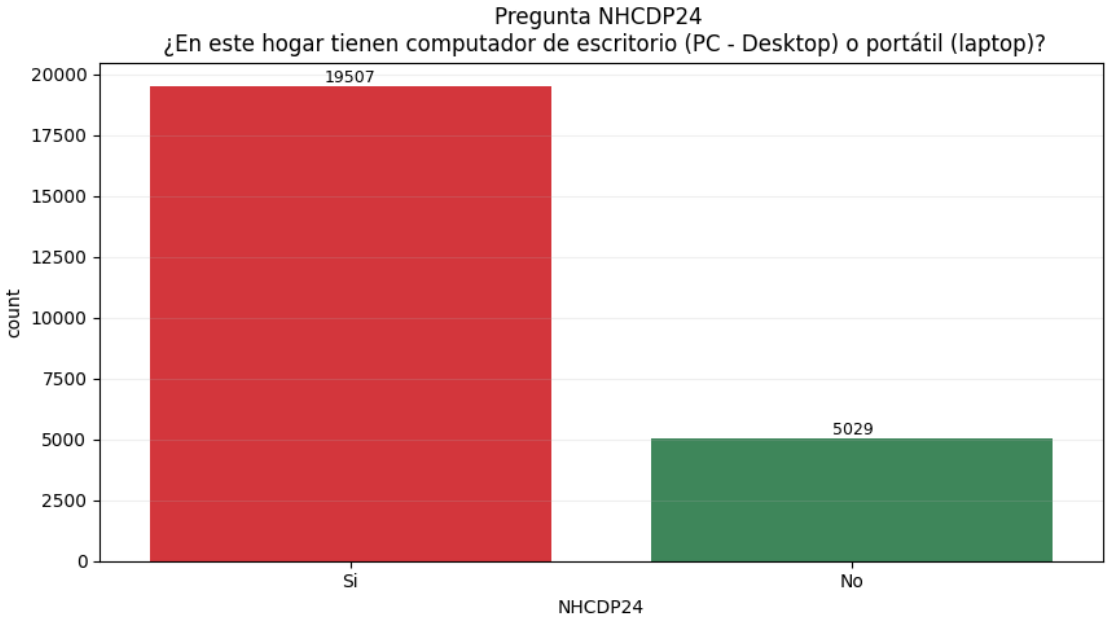
[56]:
```
data['NHCDP24'].value_counts()
```

[56]:
```
1    19507
0     5029
Name: NHCDP24, dtype: int64
```

[13]:
```
data['NHCDP24'] = data['NHCDP24'].replace([0,1],["No","Si"])
```

[14]:
```
fig, ax = plt.subplots(figsize=(10, 5))

g = sns.countplot(ax=ax, data = data, x ='NHCDP24', palette= ["#ED1C24",␣
 ↪"#329257"])

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

ax.set_title('Pregunta NHCDP24 \n ¿En este hogar tienen computador de␣
 ↪escritorio (PC - Desktop) o portátil (laptop)?')
#ax.set_xticklabels(['No', 'Si'])
```

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



2.12 NHCDP24A

2.12.1 24a. ¿Cuántos?

Datos: 19507

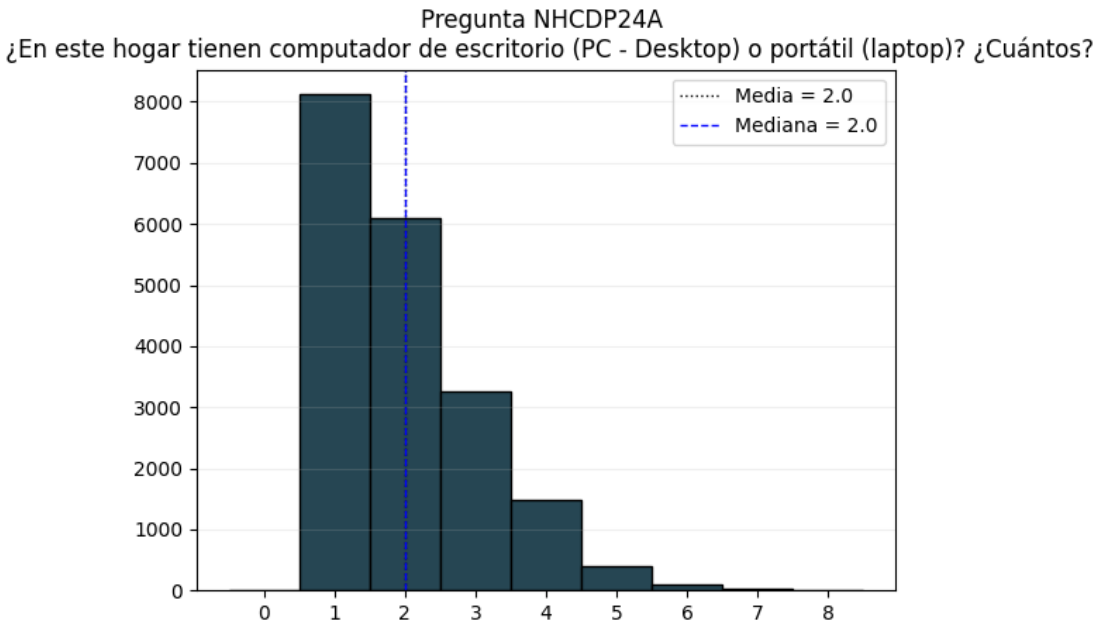
[59]:
```
data['NHCDP24A'].count()
```

[59]:
```
19507
```

[60]:
```
data['NHCDP24A'].describe().apply("{0:.2f}".format)
```

[60]:
```
count    19507.00
mean         2.00
std          1.10
min          1.00
25%          1.00
50%          2.00
75%          3.00
max          8.00
Name: NHCDP24A, dtype: object
```

[61]:
```
counts, edges, bars = plt.hist(data['NHCDP24A'], bins = np.arange(10)-0.5 ,␣
 ↪edgecolor = 'black', color = '#264653')
#ticklabels = [i for i in range(5)]
#plt.xticks(range(5), ticklabels)
plt.xticks(range(9))
#plt.bar_label(bars)
plt.title('Pregunta NHCDP24A \n ¿En este hogar tienen computador de escritorio␣
 ↪(PC - Desktop) o portátil (laptop)? ¿Cuántos?')
#plt.xlim([-1,10])
plt.axvline(data['NHCDP24A'].mean(), color='k', linestyle=':', linewidth=1,␣
 ↪label = f'Media = {data["NHCDP24A"].mean():.1f}')
plt.axvline(data['NHCDP24A'].median(), color='b', linestyle='dashed',␣
 ↪linewidth=1, label = f'Mediana = {data["NHCDP24A"].median()}')
plt.legend()
plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



2.13 NHCDP26

2.13.1 25. ¿En este hogar tienen tabletas digitales (tablets)?

0. No
1. Sí

Datos: 24536

[87]:
```
data['NHCDP26'].count()
```

[87]:
```
24536
```

[88]:
```
data['NHCDP26'].describe().apply("{0:.2f}".format)
```

[88]:
```
count    24536.00
mean         0.27
std          0.45
min          0.00
25%          0.00
50%          0.00
75%          1.00
max          1.00
Name: NHCDP26, dtype: object
```

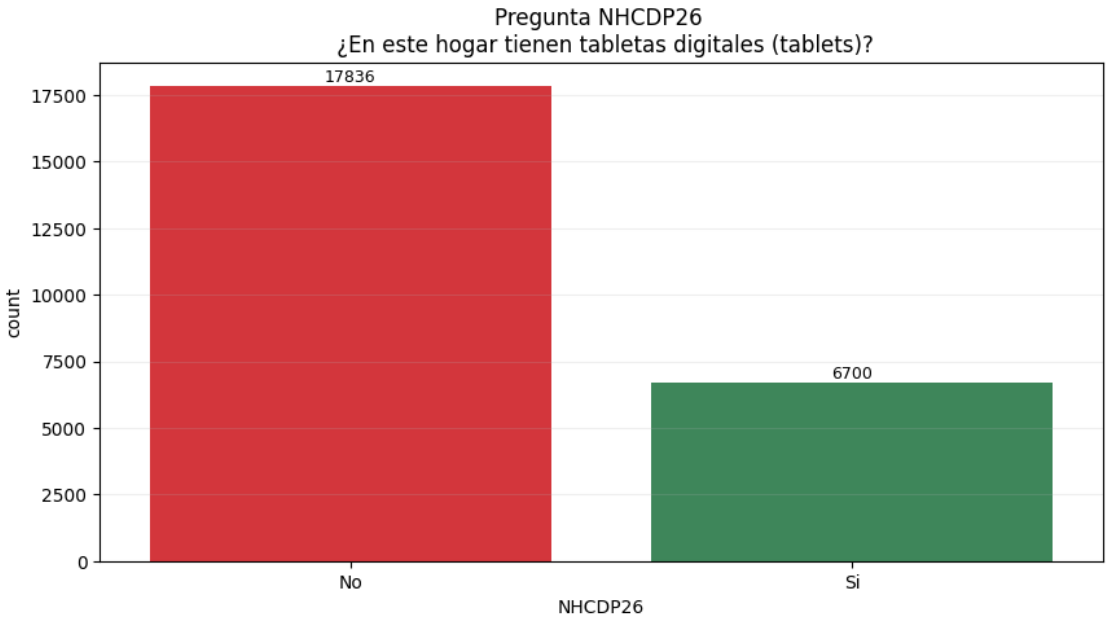
[89]:
```
data = data.replace({'NHCDP26':2},0)
```

[90]:
```
data['NHCDP26'].value_counts()
```

[90]:
```
0    17836
1     6700
Name: NHCDP26, dtype: int64
```

[15]:
```
data['NHCDP26'] = data['NHCDP26'].replace([0,1],["No","Si"])
```

[16]:
```
fig, ax = plt.subplots(figsize=(10, 5))

g = sns.countplot(ax=ax, data = data, x ='NHCDP26', palette= ["#ED1C24",␣
 ↪"#329257"])

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

ax.set_title('Pregunta NHCDP26 \n ¿En este hogar tienen tabletas digitales␣
 ↪(tablets)?')
#ax.set_xticklabels(['No', 'Si'])

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



2.14 NHCDP26A

2.14.1 25a. ¿Cuántas?

Datos: 6700

[92]:
```
data['NHCDP26A'].count()
```

[92]:
```
6700
```

[93]:
```
data['NHCDP26A'].describe().apply("{0:.2f}".format)
```

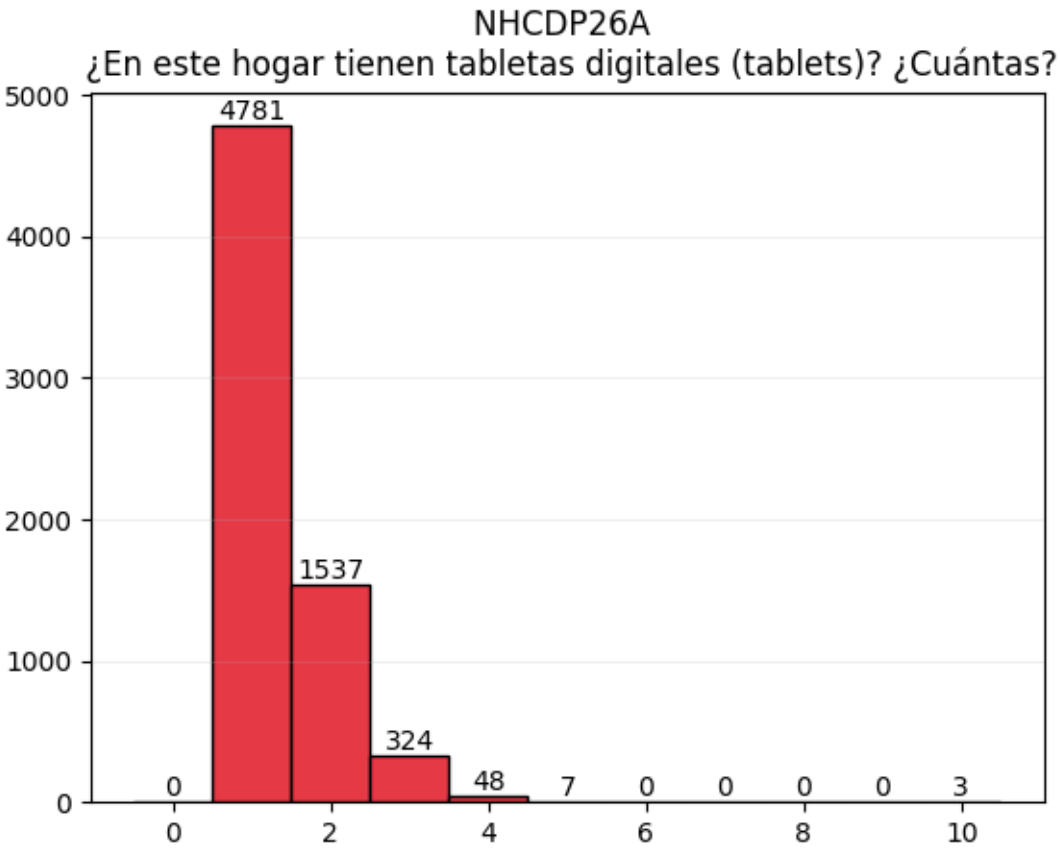
[93]:
```
count    6700.00
mean        1.36
std         0.64
min         1.00
25%         1.00
50%         1.00
75%         2.00
max        10.00
Name: NHCDP26A, dtype: object
```

[94]:
```
data['NHCDP26A'].value_counts()
```

[94]:
```
1.0     4781
2.0     1537
3.0      324
```

```
4.0       48
5.0        7
10.0       3
Name: NHCDP26A, dtype: int64
```

[70]:
```
counts, edges, bars = plt.hist(data['NHCDP26A'], bins = np.arange(12)-0.5 ,␣
↪edgecolor = 'black', color = '#e63946')
plt.bar_label(bars)
plt.title('NHCDP26A \n ¿En este hogar tienen tabletas digitales (tablets)?␣
↪¿Cuántas?')
plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



2.15 NHCDP27 *INTERNET*

2.15.1 26. ¿El hogar tiene conexión a internet?

0. No
1. Si

Datos: 24536

[71]:
```
data['NHCDP27'].count()
```

[71]:
```
24536
```

[72]:
```
data['NHCDP27'].describe().apply("{0:.2f}".format)
```

[72]:
```
count    24536.00
mean         1.09
std          0.29
min          1.00
25%          1.00
50%          1.00
75%          1.00
max          2.00
Name: NHCDP27, dtype: object
```

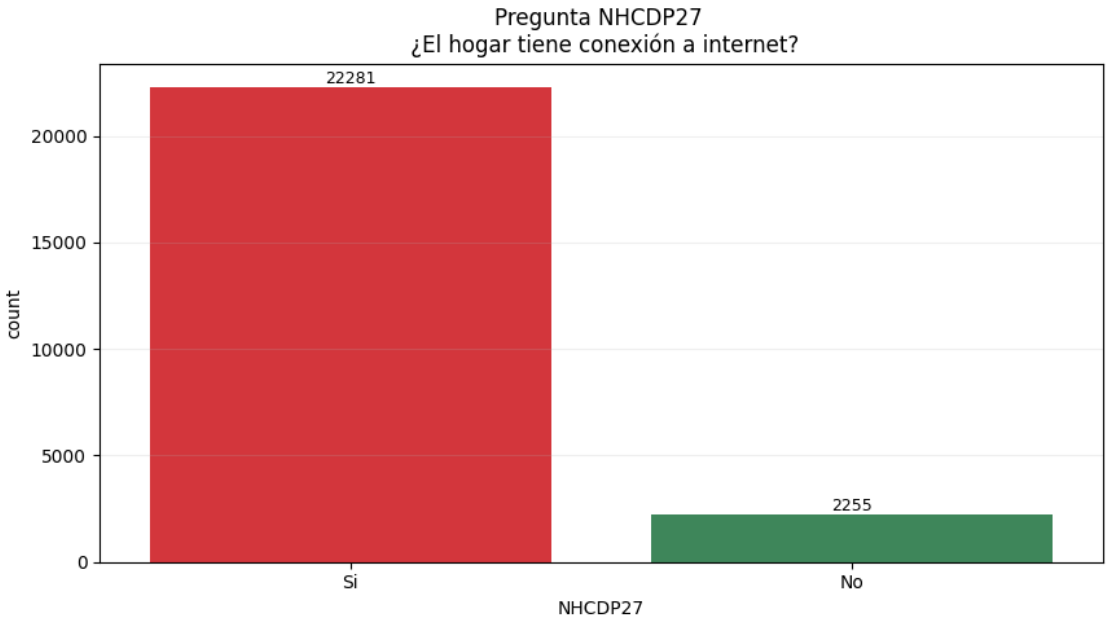
[73]:
```
data = data.replace({'NHCDP27':2},0)
```

[74]:
```
data['NHCDP27'].value_counts()
```

[74]:
```
1    22281
0     2255
Name: NHCDP27, dtype: int64
```

[17]:
```
data['NHCDP27'] = data['NHCDP27'].replace([0,1],["No","Si"])
```

[18]:
```
fig, ax = plt.subplots(figsize=(10, 5))

g = sns.countplot(ax=ax, data = data, x ='NHCDP27', palette= ["#ED1C24",␣
 ↪"#329257"])

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

ax.set_title('Pregunta NHCDP27 \n ¿El hogar tiene conexión a internet?')
#ax.set_xticklabels(['No', 'Si'])

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



2.16 NHCDP27A

2.16.1 26a. ¿El hogar tiene conexión a internet Fijo? (la conexión solo puede realizarse en el interior o cercanía de la vivienda, e incluye conexión por WiFi)

Datos: 22281

[77]:
```
data['NHCDP27A'].count()
```

[77]:
```
22281
```

[78]:
```
data['NHCDP27A'].describe().apply("{0:.2f}".format)
```

[78]:
```
count    22281.00
mean         1.04
std          0.20
min          1.00
25%          1.00
50%          1.00
75%          1.00
max          2.00
Name: NHCDP27A, dtype: object
```

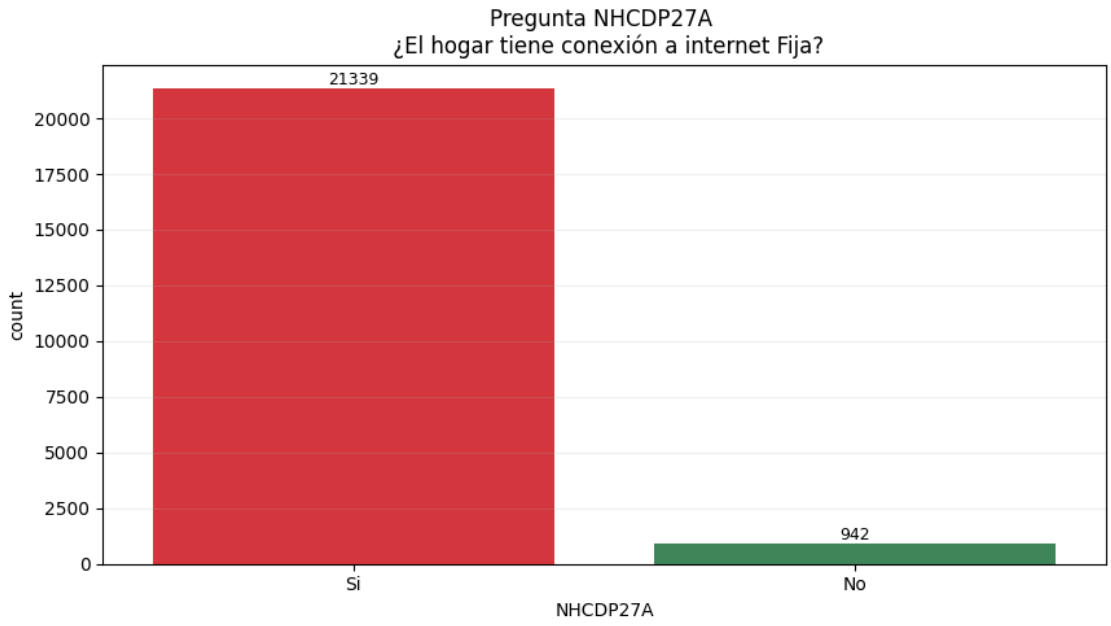
[79]:
```
data = data.replace({'NHCDP27A':2},0)
```

[80]:
```
data['NHCDP27A'].value_counts()
```

[80]:
```
1.0    21339
0.0      942
Name: NHCDP27A, dtype: int64
```

[19]:
```
data['NHCDP27A'] = data['NHCDP27A'].replace([0,1],["No","Si"])
```

[20]:
```
fig, ax = plt.subplots(figsize=(10, 5))

g = sns.countplot(ax=ax, data = data, x ='NHCDP27A', palette= ["#ED1C24",␣
 ↪"#329257"])

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

ax.set_title('Pregunta NHCDP27A \n ¿El hogar tiene conexión a internet Fija?')
#ax.set_xticklabels(['No', 'Si'])

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



2.17 NHCDP27B

2.17.1 26b. ¿El hogar tiene conexión a internet Movil? (conexión mediante red celular o modem inalámbrico o USB por cualquiera de los integrantes del hogar)

Datos: 22281

[82]:
```
data['NHCDP27B'].count()
```

[82]:
```
22281
```

[83]:
```
data['NHCDP27B'].describe().apply("{0:.2f}".format)
```

[83]:
```
count    22281.00
mean         1.52
std          0.50
min          1.00
25%          1.00
50%          2.00
75%          2.00
max          2.00
Name: NHCDP27B, dtype: object
```

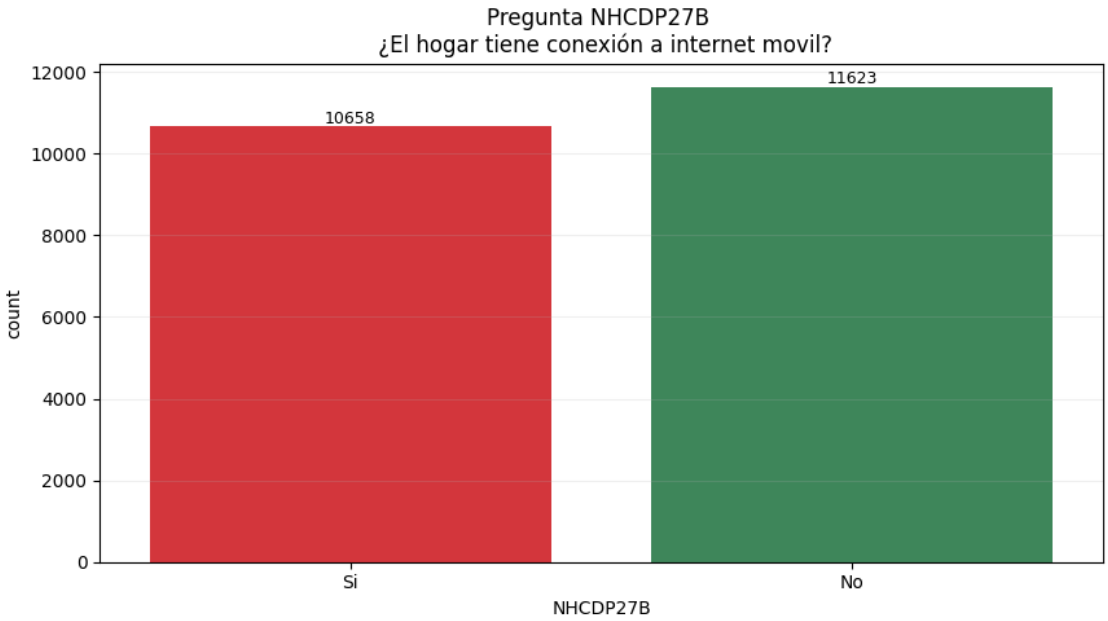
[84]:
```
data = data.replace({'NHCDP27B':2},0)
```

[85]:
```
data['NHCDP27B'].value_counts()
```

[85]:
```
0.0    11623
1.0    10658
Name: NHCDP27B, dtype: int64
```

[21]:
```
data['NHCDP27B'] = data['NHCDP27B'].replace([0,1],["No","Si"])
```

[22]:
```
fig, ax = plt.subplots(figsize=(10, 5))

g = sns.countplot(ax=ax, data = data, x ='NHCDP27B', palette= ["#ED1C24",␣
 ↪"#329257"])

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

ax.set_title('Pregunta NHCDP27B \n ¿El hogar tiene conexión a internet movil?')
#ax.set_xticklabels(['No', 'Si'])

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



2.18 NHCDP28

2.18.1 27. Este hogar paga por el servicio de internet?

1. Sí
2. Sí, con el arriendo
3. No pagan

Datos: 22281

[96]:
```
data['NHCDP28'].count()
```

[96]:
```
22281
```

[97]:
```
data['NHCDP28'].describe().apply("{0:.2f}".format)
```

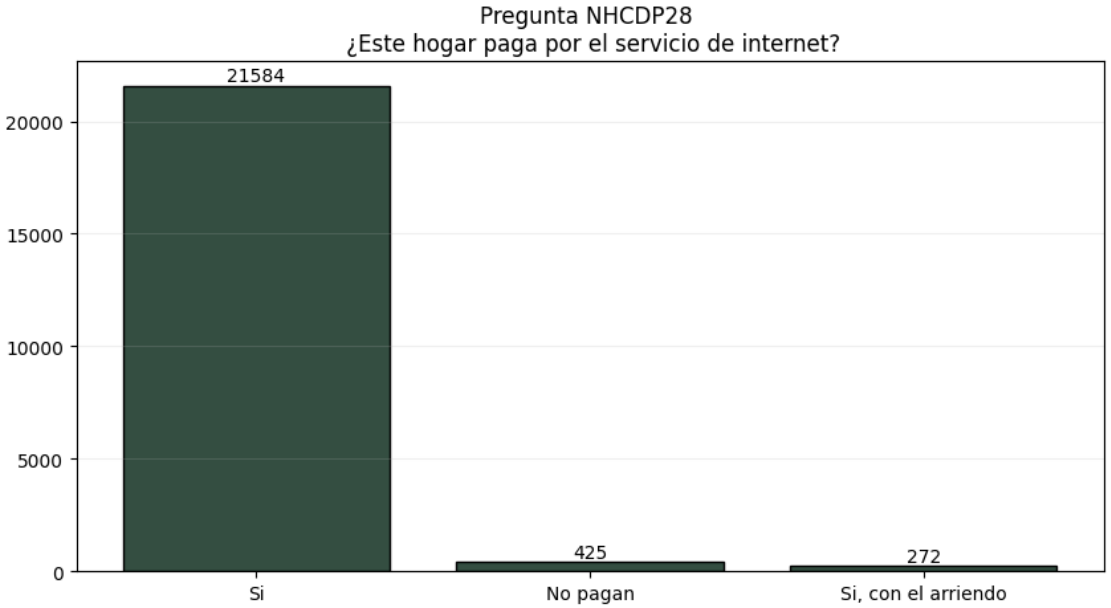
[97]:
```
count    22281.00
mean         1.05
std          0.29
min          1.00
25%          1.00
50%          1.00
75%          1.00
max          3.00
Name: NHCDP28, dtype: object
```

[98]:
```
data['NHCDP28'].value_counts()
```

[98]:
```
1.0    21584
3.0      425
2.0      272
Name: NHCDP28, dtype: int64
```

[23]:
```
data['NHCDP28'] = data['NHCDP28'].replace([1,2,3],["Si","Si, con el␣
↪arriendo","No pagan"])
```

[24]:
```
plt.figure(figsize=(10,5))
bars = plt.bar(data['NHCDP28'].value_counts().index.tolist(),data['NHCDP28'].
↪value_counts().tolist(), edgecolor = 'k', color = '#344e41')
#plt.xticks([1,2,3], ['Si ','Si, con el arriendo', 'No pagan',])
plt.title('Pregunta NHCDP28 \n ¿Este hogar paga por el servicio de internet?')
plt.bar_label(bars)
plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



2.19 NHCDP29

2.19.1 28. ¿Cuánto pagó este hogar la ÚLTIMA VEZ por el servicio de internet?

Contestan 21584

Revisar casos de 98

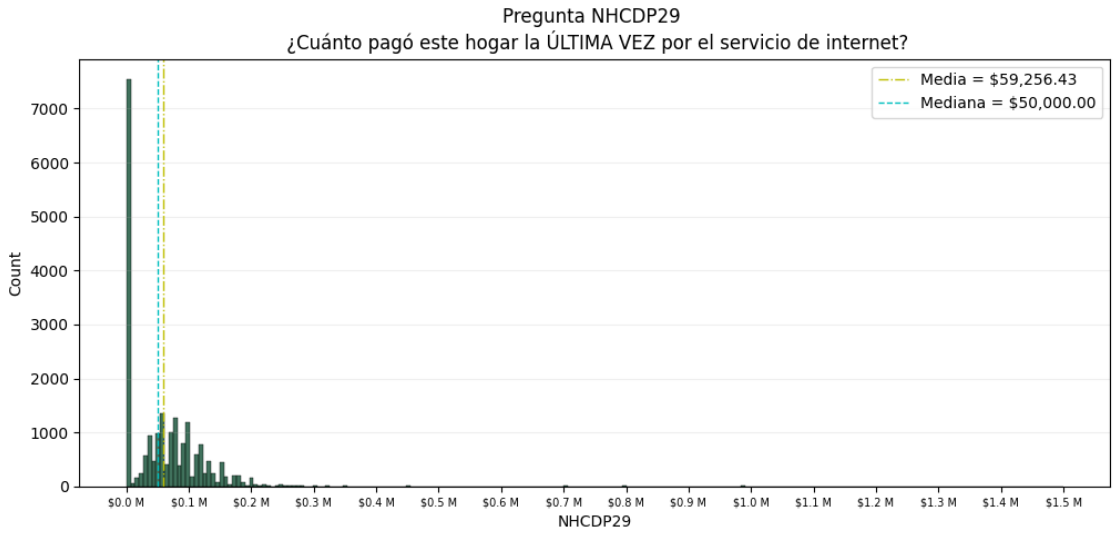
[100]:
```
data['NHCDP29'].count()
```

[100]: 21584

[101]:
```
data['NHCDP29'].describe().apply("{0:.2f}".format)
```

[101]:
```
count      21584.00
mean       59256.43
std        69023.12
min            0.00
25%           98.00
50%        50000.00
75%        95000.00
max      1500000.00
Name: NHCDP29, dtype: object
```

[111]:
```
plt.figure(figsize=(12,5))
sns.histplot(data = data, x = 'NHCDP29', color = '#004225')
plt.xticks(range(0,1600000,100000),[f'${(i / 1000000):.1f} M' for i in␣
 ↪range(0,1600000,100000)], fontsize=7)
plt.title('Pregunta NHCDP29 \n ¿Cuánto pagó este hogar la ÚLTIMA VEZ por el␣
 ↪servicio de internet?')
plt.axvline(data['NHCDP29'].mean(), color='y', linestyle='-.', linewidth=1,␣
 ↪label = f'Media = ${data["NHCDP29"].mean():,.2f}')
plt.axvline(data['NHCDP29'].median(), color='c', linestyle='dashed',␣
 ↪linewidth=1, label = f'Mediana = ${data["NHCDP29"].median():,.2f}')
plt.legend()
plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



2.20 NHCDP33

2.20.1 32. ¿En el último mes pagó por el uso de alguna aplicación para ver películas, series, videos o escuchar música a través de internet?

0. No
1. Sí

Datos: 24536

[112]:
```
data['NHCDP33'].count()
```

[112]:
```
24536
```

[113]:
```
data['NHCDP33'].describe().apply("{0:.2f}".format)
```

[113]:
```
count    24536.00
mean         1.67
std          0.47
min          1.00
25%          1.00
50%          2.00
75%          2.00
max          2.00
Name: NHCDP33, dtype: object
```

[114]:
```
data = data.replace({'NHCDP33':2},0)
```

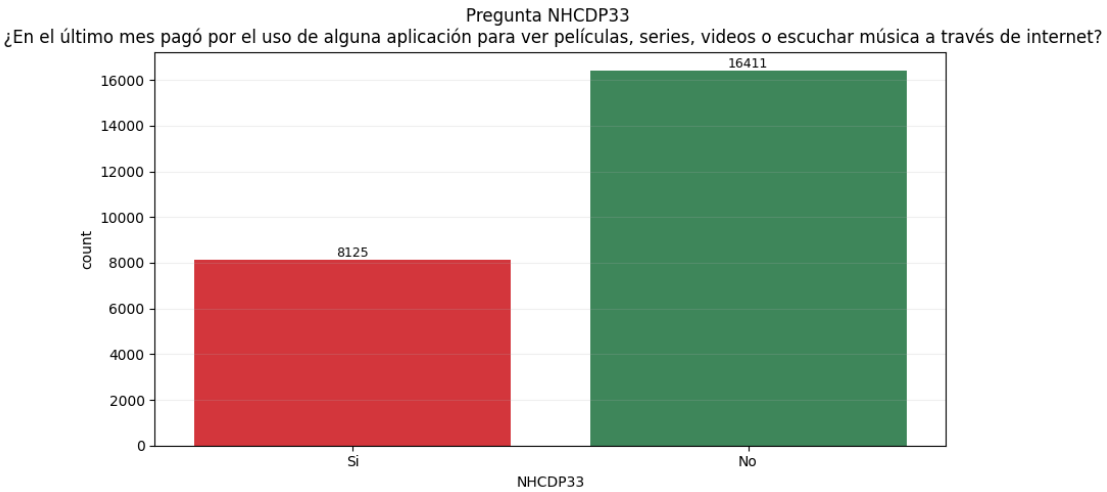
[115]:
```
data['NHCDP33'].value_counts()
```

[115]:
```
0    16411
1     8125
Name: NHCDP33, dtype: int64
```

[25]:
```
data['NHCDP33'] = data['NHCDP33'].replace([0,1],["No","Si"])
```

[26]:
```
fig, ax = plt.subplots(figsize=(10, 5))

g = sns.countplot(ax=ax, data = data, x ='NHCDP33', palette= ["#ED1C24",␣
 ↪"#329257"])

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

ax.set_title('Pregunta NHCDP33 \n ¿En el último mes pagó por el uso de alguna␣
 ↪aplicación para ver películas, series, videos o escuchar música a través de␣
 ↪internet?')
#ax.set_xticklabels(['No', 'Si'])
```

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



2.21 NHCDP33A

2.21.1 32a. ¿Cuánto pagó este hogar por el uso de estas aplicaciones en el ÚLTIMO MES?

Datos: 8125

[117]:
```
data['NHCDP33A'].count()
```

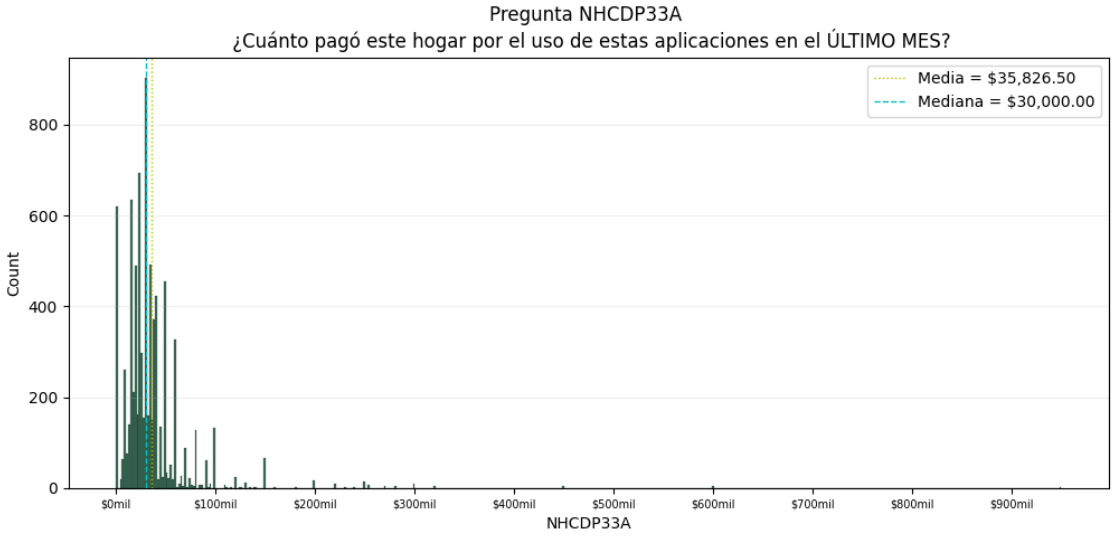
[117]:
```
8125
```

[118]:
```
data['NHCDP33A'].describe().apply("{0:.2f}".format)
```

[118]:
```
count      8125.00
mean      35826.50
std       39903.90
min          98.00
25%       19000.00
50%       30000.00
75%       40000.00
max      950000.00
Name: NHCDP33A, dtype: object
```

[125]:
```
plt.figure(figsize=(12,5))
sns.histplot(data = data, x = 'NHCDP33A', color = '#004225')
plt.xticks(range(0,1000000,100000),[f'${(i / 1000):.0f}mil' for i in␣
 ↪range(0,1000000,100000)], fontsize=7)
#plt.bar_label(bars)
```

```
plt.title('Pregunta NHCDP33A \n ¿Cuánto pagó este hogar por el uso de estas␣
 ↪aplicaciones en el ÚLTIMO MES?')
#plt.xlim([-1,10])
plt.axvline(data['NHCDP33A'].mean(), color='y', linestyle=':', linewidth=1,␣
 ↪label = f'Media = ${data["NHCDP33A"].mean():,.2f}')
plt.axvline(data['NHCDP33A'].median(), color='c', linestyle='dashed',␣
 ↪linewidth=1, label = f'Mediana = ${data["NHCDP33A"].median():,.2f}')
plt.legend()
plt.grid(alpha = 0.2, axis = 'y')
plt.show()
```



[27]:
```
data.to_excel('Encuesta_Multiproposito_Suba.xlsx', index = False)
```

[ ]:
```

```