

```
[1]: # Importamos las librerías pandas y numpy
import pandas as pd
import numpy as np

# Definir la ruta del archivo de datos
ruta_archivo = "C:/Users/Arle/OneDrive/Escritorio/Api-2-Modulo2"

In [2]: # Importamos los archivos csv originales con los datos que vamos a utilizar
# provincia, año, población_total, población_varones, población_mujeres
poblacion = pd.read_csv("poblacion.csv", encoding="latin-1")

In [3]: # Importamos los archivos csv originales con los datos que vamos a utilizar
# esperanza_de_vida (provincia, año, mujeres, varones)
esperanza_de_vida = pd.read_csv("esperanza_de_vida.csv", encoding="latin-1")

In [4]: # Importamos los archivos csv originales con los datos que vamos a utilizar
# superficie (provincia, año, superficie)
hogares_viviendas_superficie = pd.read_csv("hogares_viviendas_superficie.csv", encoding="latin-1")

In [5]: # Exploramos las primeras filas del archivo con datos de población
poblacion.head(15)

Out[5]:
provincia  año  población_total  población_varones  población_mujeres
0  Total País  2010  40788453  19940704  20847749
1  Total País  2011  41261490  20180791  21080699
2  Total País  2012  41733271  20420391  21312880
3  Total País  2013  42202935  20696037  21543988
4  Total País  2014  42669600  20966203  21703397
5  Total País  2015  43131966  21131346  22000620
6  Total País  2016  43603608  21364170  22239438
7  Total País  2017  44048411  21595623  22452788
8  Total País  2018  44496922  21826477  22670445
9  Total País  2019  44897012  22052032  22844980
10 Total País  2020  45276762  22273182  23003580
11 Total País  2021  45698747  22492818  23205929
12 Total País  2022  46234830  22709478  23525352
13 Total País  2023  46654581  22922981  23731700
14 Total País  2024  47067641  23132946  23934695

In [6]: # Exploramos las primeras filas del archivo con datos de esperanza de vida
esperanza_de_vida.head()

Out[6]:
provincia  año  mujeres  varones
0  Buenos Aires  2015  80.22  73.54
1  Buenos Aires  2020  81.34  74.74
2  Buenos Aires  2025  82.32  75.80
3  Buenos Aires  2030  83.20  76.76
4  Buenos Aires  2035  83.98  77.60

In [7]: # Exploramos las primeras filas del archivo con datos de hogares, viviendas, superficie
hogares_viviendas_superficie.head()

Out[7]:
provincia_id  provincia  hogares  viviendas_particulares  viviendas_particulares_habitadas  superficie_km2
0  Capital Federal  1150134  1423973  1032988  200
1  6  Buenos Aires  4789484  5377786  4425193  307571
2  10  Catamarca  96001  111634  89376  102602
3  14  Córdoba  1031843  1232211  978653  166321
4  18  Corrientes  267797  292644  248844  88199

In [8]: # Exploramos las últimas filas del archivo dataset
poblacion.tail()

Out[8]:
provincia  año  población_total  población_varones  población_mujeres
770 Tierra del Fuego  2036  241593  122567  119026
771 Tierra del Fuego  2037  245734  124625  121109
772 Tierra del Fuego  2038  249853  126670  123183
773 Tierra del Fuego  2039  253948  128702  125246
774 Tierra del Fuego  2040  258020  130721  127299

In [9]: # Exploramos las últimas filas del archivo dataset
esperanza_de_vida.tail()

Out[9]:
provincia  año  mujeres  varones
139 Tucumán  2020  81.05  75.11
140 Tucumán  2025  82.21  76.15
141 Tucumán  2030  83.23  77.07
142 Tucumán  2035  83.84  77.88
143 Tucumán  2040  84.54  78.58

In [10]: # Exploramos las últimas filas del archivo dataset
hogares_viviendas_superficie.tail()

Out[10]:
provincia_id  provincia  hogares  viviendas_particulares  viviendas_particulares_habitadas  superficie_km2
19  76  Santa Cruz  81796  93881  76233  242943
20  82  Santa Fe  1023777  1143651  948369  133007
21  86  Santiago del Estero  218025  242034  197906  136361
22  90  Tucumán  368538  396040  338621  22524
23  94  Tierra del Fuego  38956  43360  36669  100245

In [11]: # Usamos shape para ver cuántas filas y columnas tiene y print para agregarle texto
print("Las filas y columnas son:")

Out[11]:
Las filas y columnas son:
(1795, 5)

In [12]: # Usamos shape para ver cuántas filas y columnas tiene y print para agregarle texto
print("Las filas y las columnas son:")

Out[12]:
Las filas y las columnas son:
(144, 4)

In [13]: # Usamos shape para ver cuántas filas y columnas tiene y print para agregarle texto
print("Las filas y las columnas son:")
hogares_viviendas_superficie.shape

Out[13]:
Las filas y las columnas son:
(24, 6)

In [14]: # Chequeamos los tipos de variables
poblacion.dtypes

Out[14]:
provincia      object
año            int64
poblacion_total  int64
poblacion_varones  int64
poblacion_mujeres  int64
dtype: object

In [15]: # Chequeamos los tipos de variables
esperanza_de_vida.dtypes

Out[15]:
provincia      object
año            int64
mujeres        float64
varones        float64
dtype: object

In [16]: # Chequeamos los tipos de variables
hogares_viviendas_superficie.dtypes

Out[16]:
provincia_id      int64
provincia         object
hogares           int64
viviendas_particulares  int64
viviendas_particulares_habitadas  int64
superficie_km2    int64
dtype: object

In [17]: # Importamos de vuelta el archivo sin las últimas 3 filas
poblacion = pd.read_csv("poblacion.csv", skipfooter=3, engine="python", encoding="latin-1")
# Filtramos las observaciones/ filas que corresponden a las provincias y año
poblacion = poblacion[poblacion["provincia"].isin(["Corrientes", "Chaco", "Total País"]) &
poblacion["año"] <= 2015]]

Out[17]:
provincia  año  población_total  población_varones  población_mujeres
0  Total País  2010  40788453  19940704  20847749
1  Total País  2011  41261490  20180791  21080699
2  Total País  2012  41733271  20420391  21312880
3  Total País  2013  42202935  20696037  21543988
4  Total País  2014  42669600  20966203  21703397
5  Total País  2015  43131966  21131346  22000620
155 Corrientes  2010  1017731  501452  516279
156 Corrientes  2011  1028248  506702  521546
157 Corrientes  2012  1038786  511969  526817
158 Corrientes  2013  1049325  517240  532085
159 Corrientes  2014  1059836  522500  537336
160 Corrientes  2015  1070283  527731  542552
186 Chaco  2010  1080017  534347  545670
187 Chaco  2011  1092625  540391  552234
188 Chaco  2012  1105280  546471  558670
189 Chaco  2013  1117953  552570  565670
190 Chaco  2014  1130608  558670  571938
191 Chaco  2015  1143201  564746  578455

In [18]: # Importamos de vuelta el archivo sin las últimas 3 filas
esperanza_de_vida = pd.read_csv("esperanza_de_vida.csv", skipfooter=3, engine="python", encoding="latin-1")
# Filtramos las observaciones/ filas que corresponden a las provincias y año
esperanza_de_vida = esperanza_de_vida[esperanza_de_vida["provincia"].isin(["Buenos Aires", "Catamarca"]) &
esperanza_de_vida["año"] <= 2040]]

Out[18]:
provincia  año  mujeres  varones
0  Buenos Aires  2015  80.22  73.54
1  Buenos Aires  2020  81.34  74.74
2  Buenos Aires  2025  82.32  75.80
3  Buenos Aires  2030  83.20  76.76
4  Buenos Aires  2035  83.98  77.60
5  Buenos Aires  2040  84.66  78.32
6  Catamarca  2015  80.36  74.74
7  Catamarca  2020  81.45  75.78
8  Catamarca  2025  82.45  76.70
9  Catamarca  2030  83.32  77.55
10 Catamarca  2035  84.08  78.28
11 Catamarca  2040  84.75  78.91

In [19]: # Importamos de vuelta el archivo sin las últimas 3 filas
hogares_viviendas_superficie = pd.read_csv("hogares_viviendas_superficie.csv", skipfooter=3, engine="python", encoding="latin-1")
# Filtramos las observaciones/ filas que corresponden a las provincias
hogares_viviendas_superficie = hogares_viviendas_superficie[hogares_viviendas_superficie["provincia"].isin(["Corrientes"])]

Out[19]:
provincia_id  provincia  hogares  viviendas_particulares  viviendas_particulares_habitadas  superficie_km2
4  18  Corrientes  267797  292644  248844  88199

In [20]: # Eliminamos columnas
poblacion = poblacion.drop(columns=["poblacion_mujeres"])

Out[20]:
provincia  año  población_total  población_varones
0  Total País  2010  40788453  19940704
1  Total País  2011  41261490  20180791
2  Total País  2012  41733271  20420391
3  Total País  2013  42202935  20696037
4  Total País  2014  42669600  20966203
5  Total País  2015  43131966  21131346
155 Corrientes  2010  1017731  501452
156 Corrientes  2011  1028248  506702
157 Corrientes  2012  1038786  511969
158 Corrientes  2013  1049325  517240
159 Corrientes  2014  1059836  522500
160 Corrientes  2015  1070283  527731
186 Chaco  2010  1080017  534347
187 Chaco  2011  1092625  540391
188 Chaco  2012  1105280  546471
189 Chaco  2013  1117953  552570
190 Chaco  2014  1130608  558670
191 Chaco  2015  1143201  564746

In [21]: # Eliminamos columnas
esperanza_de_vida = esperanza_de_vida.drop(columns=["varones"])

Out[21]:
provincia  año  mujeres
0  Buenos Aires  2015  80.22
1  Buenos Aires  2020  81.34
2  Buenos Aires  2025  82.32
3  Buenos Aires  2030  83.20
4  Buenos Aires  2035  83.98
5  Buenos Aires  2040  84.66
6  Catamarca  2015  80.36
7  Catamarca  2020  81.45
8  Catamarca  2025  82.45
9  Catamarca  2030  83.32
10 Catamarca  2035  84.08
11 Catamarca  2040  84.75

In [22]: # Eliminamos columnas
hogares_viviendas_superficie = hogares_viviendas_superficie.drop(columns=["viviendas_particulares", "viviendas_particulares_habitadas"])

Out[22]:
provincia_id  provincia  hogares  superficie_km2
4  18  Corrientes  267797  88199

In [23]: # Transponemos de formato ancho a largo
df_wide = pd.melt(poblacion, id_vars=["provincia"], value_vars=["año", "poblacion_total", "poblacion_varones"])

In [24]: # Renombramos la variable año por año
esperanza_de_vida = esperanza_de_vida.rename(columns={"año": "año"})

Out[24]:
provincia  año  población_total  población_varones
0  Total País  2010  40788453  19940704
1  Total País  2011  41261490  20180791
2  Total País  2012  41733271  20420391
3  Total País  2013  42202935  20696037
4  Total País  2014  42669600  20966203
5  Total País  2015  43131966  21131346
155 Corrientes  2010  1017731  501452
156 Corrientes  2011  1028248  506702
157 Corrientes  2012  1038786  511969
158 Corrientes  2013  1049325  517240
159 Corrientes  2014  1059836  522500
160 Corrientes  2015  1070283  527731
186 Chaco  2010  1080017  534347
187 Chaco  2011  1092625  540391
188 Chaco  2012  1105280  546471
189 Chaco  2013  1117953  552570
190 Chaco  2014  1130608  558670
191 Chaco  2015  1143201  564746

In [25]: # Renombramos la variable año por año
esperanza_de_vida = esperanza_de_vida.rename(columns={"año": "año"})

Out[25]:
provincia  año  mujeres
0  Buenos Aires  2015  80.22
1  Buenos Aires  2020  81.34
2  Buenos Aires  2025  82.32
3  Buenos Aires  2030  83.20
4  Buenos Aires  2035  83.98
5  Buenos Aires  2040  84.66
6  Catamarca  2015  80.36
7  Catamarca  2020  81.45
8  Catamarca  2025  82.45
9  Catamarca  2030  83.32
10 Catamarca  2035  84.08
11 Catamarca  2040  84.75

In [27]: # Vemos los tipos de variable de cada columna
poblacion.dtypes

Out[27]:
provincia      object
año            int64
poblacion_total  int64
poblacion_varones  int64
dtype: object

In [28]: # Vemos los tipos de variable de cada columna
esperanza_de_vida.dtypes

Out[28]:
provincia      object
año            int64
mujeres        float64
dtype: object

In [29]: # Vemos los tipos de variable de cada columna
hogares_viviendas_superficie.dtypes

Out[29]:
provincia_id      int64
provincia         object
hogares           int64
superficie_km2    int64
dtype: object

In [30]: # Convertimos algunos de los campos a tipos de variables que sirvan: provincia: string, poblacion_varones: int
poblacion = poblacion.astype({"provincia": str, "poblacion_varones": int})

Out[30]:
provincia      object
año            int64
poblacion_total  int64
poblacion_varones  int32
dtype: object

In [31]: # Convertimos algunos de los campos a tipos de variables que sirvan: provincia: string, año: int
esperanza_de_vida = esperanza_de_vida.astype({"provincia": str, "año": int})

Out[31]:
provincia      object
año            int32
mujeres        float64
dtype: object

In [32]: # En el archivo esperanza_de_vida y hogares_viviendas_superficie no se necesita hacer la conversión

In [33]: # Filtramos las observaciones anteriores a 2015, el último año con datos completos
poblacion = poblacion[poblacion["año"] <= 2015]

Out[33]:
provincia  año  población_total  población_varones
0  Total País  2010  40788453  19940704
1  Total País  2011  41261490  20180791
2  Total País  2012  41733271  20420391
3  Total País  2013  42202935  20696037
4  Total País  2014  42669600  20966203
5  Total País  2015  43131966  21131346
155 Corrientes  2010  1017731  501452
156 Corrientes  2011  1028248  506702
157 Corrientes  2012  1038786  511969
158 Corrientes  2013  1049325  517240
159 Corrientes  2014  1059836  522500
160 Corrientes  2015  1070283  527731
186 Chaco  2010  1080017  534347
187 Chaco  2011  1092625  540391
188 Chaco  2012  1105280  546471
189 Chaco  2013  1117953  552570
190 Chaco  2014  1130608  558670
191 Chaco  2015  1143201  564746

In [34]: # Filtramos las observaciones anteriores a 2049, el último año con datos completos
esperanza_de_vida = esperanza_de_vida[esperanza_de_vida["año"] <= 2040]

Out[34]:
provincia  año  mujeres
0  Buenos Aires  2015  80.22
1  Buenos Aires  2020  81.34
2  Buenos Aires  2025  82.32
3  Buenos Aires  2030  83.20
4  Buenos Aires  2035  83.98
5  Buenos Aires  2040  84.66
6  Catamarca  2015  80.36
7  Catamarca  2020  81.45
8  Catamarca  2025  82.45
9  Catamarca  2030  83.32
10 Catamarca  2035  84.08
11 Catamarca  2040  84.75

In [35]: # Calculamos las principales estadísticas descriptivas de la variable población_varones
poblacion["poblacion_varones"].describe()

Out[35]:
count    1.800000e+01
mean     7.280727e+06
std      9.787386e+06
min      5.014520e+05
25%      5.238970e+05
50%      5.482020e+05
75%      5.812977e+07
max      2.113135e+07
Name: poblacion_varones, dtype: float64

In [36]: # Calculamos las principales estadísticas descriptivas de la variable mujeres
poblacion["poblacion_mujeres"].describe()

Out[36]:
count    12.000000
mean     82.677580
std      1.678945
min      80.220080
25%      81.422580
50%      82.850980
75%      84.055080
max      84.750080
Name: mujeres, dtype: float64

In [37]: # Calculamos las principales estadísticas descriptivas de la variable hogares
hogares_viviendas_superficie["hogares"].describe()

Out[37]:
count      1.0
mean     267797.0
std      267797.0
min      267797.0
25%      267797.0
50%      267797.0
75%      267797.0
max      267797.0
Name: hogares, dtype: float64

In [38]: # Concatenamos la tabla población con esperanza_de_vida
resultado = pd.concat([poblacion, esperanza_de_vida])
# Combinamos la tabla hogares_viviendas_superficie con la tabla concatenada
resultado = pd.merge(hogares_viviendas_superficie, resultado, left_on=["provincia"], right_on=["provincia"], how="outer")

Out[38]:
provincia_id  provincia  hogares  superficie_km2  año  población_total  población_varones  mujeres
0  18.0  Corrientes  267797.0  88199.0  2010  1017731.0  501452.0  NaN
1  18.0  Corrientes  267797.0  88199.0  2011  1028248.0  506702.0  NaN
2  18.0  Corrientes  267797.0  88199.0  2012  1038786.0  511969.0  NaN
3  18.0  Corrientes  267797.0  88199.0  2013  1049325.0  517240.0  NaN
4  18.0  Corrientes  267797.0  88199.0  2014  1059836.0  522500.0  NaN
5  18.0  Corrientes  267797.0  88199.0  2015  1070283.0  527731.0  NaN
6  NaN  Total País  NaN  NaN  2010  40788453.0  19940704.0  NaN
7  NaN  Total País  NaN  NaN  2011  41261490.0  20180791.0  NaN
8  NaN  Total País  NaN  NaN  2012  41733271.0  20420391.0  NaN
9  NaN  Total País  NaN  NaN  2013  42202935.0  20696037.0  NaN
10 NaN  Total País  NaN  NaN  2014  42669600.0  20966203.0  NaN
11 NaN  Total País  NaN  NaN  2015  43131966.0  21131346.0  NaN
12 NaN  Chaco  NaN  NaN  2010  1080017.0  534347.0  NaN
13 NaN  Chaco  NaN  NaN  2011  1092625.0  540391.0  NaN
14 NaN  Chaco  NaN  NaN  2012  1105280.0  546471.0  NaN
15 NaN  Chaco  NaN  NaN  2013  1117953.0  552570.0  NaN
16 NaN  Chaco  NaN  NaN  2014  1130608.0  558670.0  NaN
17 NaN  Chaco  NaN  NaN  2015  1143201.0  564746.0  NaN
18 NaN  Buenos Aires  NaN  NaN  2015  NaN  NaN  80.22
19 NaN  Buenos Aires  NaN  NaN  2020  NaN  NaN  81.34
20 NaN  Buenos Aires  NaN  NaN  2025  NaN  NaN  82.32
21 NaN  Buenos Aires  NaN  NaN  2030  NaN  NaN  83.20
22 NaN  Buenos Aires  NaN  NaN  2035  NaN  NaN  83.98
23 NaN  Buenos Aires  NaN  NaN  2040  NaN  NaN  84.66
24 NaN  Catamarca  NaN  NaN  2015  NaN  NaN  80.36
25 NaN  Catamarca  NaN  NaN  2020  NaN  NaN  81.45
26 NaN  Catamarca  NaN  NaN  2025  NaN  NaN  82.45
27 NaN  Catamarca  NaN  NaN  2030  NaN  NaN  83.32
28 NaN  Catamarca  NaN  NaN  2035  NaN  NaN  84.08
29 NaN  Catamarca  NaN  NaN  2040  NaN  NaN  84.75

In [39]: # Calculamos la tabla población con densidad
resultado["densidad"] = resultado["poblacion_total"] / resultado["superficie_km2"]

Out[39]:
provincia_id  provincia  hogares  superficie_km2  año  población_total  población_varones  mujeres  densidad
0  18.0  Corrientes  267797.0  88199.0  2010  1017731.0  501452.0  NaN  11.53931
1  18.0  Corrientes  267797.0  88199.0  2011  1028248.0  506702.0  NaN  11.68673
2  18.0  Corrientes  267797.0  88199.0  2012  1038786.0  511969.0  NaN  11.77753
3  18.0  Corrientes  267797.0  88199.0  2013  1049325.0  517240.0  NaN  11.89724
4  18.0  Corrientes  267797.0  88199.0  2014  1059836.0  522500.0  NaN  12.016417
5  18.0  Corrientes  267797.0  88199.0  2015  1070283.0  527731.0  NaN  12.134865
6  NaN  Total País  NaN  NaN  2010  40788453.0  19940704.0  NaN  NaN
7  NaN  Total País  NaN  NaN  2011  41261490.0  20180791.0  NaN  NaN
8  NaN  Total País  NaN  NaN  2012  41733271.0  20420391.0  NaN  NaN
9  NaN  Total País  NaN  NaN  2013  42202935.0  20696037.0  NaN  NaN
10 NaN  Total País  NaN  NaN  2014  42669600.0  20966203.0  NaN  NaN
11 NaN  Total País  NaN  NaN  2015  43131966.0  21131346.0  NaN  NaN
12 NaN  Chaco  NaN  NaN  2010  1080017.0  534347.0  NaN  NaN
13 NaN  Chaco  NaN  NaN  2011  1092625.0  540391.0  NaN  NaN
14 NaN  Chaco  NaN  NaN  2012  1105280.0  546471.0  NaN  NaN
15 NaN  Chaco  NaN  NaN  2013  1117953.0  552570.0  NaN  NaN
16 NaN  Chaco  NaN  NaN  2014  1130608.0  558670.0  NaN  NaN
17 NaN  Chaco  NaN  NaN  2015  1143201.0  564746.0  NaN  NaN
18 NaN  Buenos Aires  NaN  NaN  2015  NaN  NaN  80.22  NaN
19 NaN  Buenos Aires  NaN  NaN  2020  NaN  NaN  81.34  NaN
20 NaN  Buenos Aires  NaN  NaN  2025  NaN  NaN  82.32  NaN
21 NaN  Buenos Aires  NaN  NaN  2030  NaN  NaN  83.20  NaN
22 NaN  Buenos Aires  NaN  NaN  2035  NaN  NaN  83.98  NaN
23 NaN  Buenos Aires  NaN  NaN  2040  NaN  NaN  84.66  NaN
24 NaN  Catamarca  NaN  NaN  2015  NaN  NaN  80.36  NaN
25 NaN  Catamarca  NaN  NaN  2020  NaN  NaN  81.45  NaN
26 NaN  Catamarca  NaN  NaN  2025  NaN  NaN  82.45  NaN
27 NaN  Catamarca  NaN  NaN  2030  NaN  NaN  83.32  NaN
28 NaN  Catamarca  NaN  NaN  2035  NaN  NaN  84.08  NaN
29 NaN  Catamarca  NaN  NaN  2040  NaN  NaN  84.75  NaN

In [40]: # Nos quedamos con las primeras 6 filas
resultado.head(6)

Out[40]:
provincia_id  provincia  hogares  superficie_km2  año  población_total  población_varones  mujeres  densidad  densidad_estandarizada
0  18.0  Corrientes  267797.0  88199.0  2010  1017731.0  501452.0  NaN  11.53931  -1.464607
1  18.0  Corrientes  267797.0  88199.0  2011  1028248.0  506702.0  NaN  11.68673  -0.879017
2  18.0  Corrientes  267797.0  88199.0  2012  1038786.0  511969.0  NaN  11.77753  -0.292257
3  18.0  Corrientes  267797.0  88199.0  2013  1049325.0  517240.0  NaN  11.89724  0.294568
4  18.0  Corrientes  267797.0  88199.0  2014  1059836.0  522500.0  NaN  12.016417  0.87815
5  18.0  Corrientes  267797.0  88199.0  2015  1070283.0  527731.0  NaN  12.134865  1.461507
6  NaN  Total País  NaN  NaN  2010  40788453.0  19940704.0  NaN  NaN  NaN
7  NaN  Total País  NaN  NaN  2011  41261490.0  20180791.0  NaN  NaN  NaN
8  NaN  Total País  NaN  NaN  2012  41733271.0  20420391.0  NaN  NaN  NaN
9  NaN  Total País  NaN  NaN  2013  42202935.0  20696037.0  NaN  NaN  NaN
10 NaN  Total País  NaN  NaN  2014  42669600.0  20966203.0  NaN  NaN  NaN
11 NaN  Total País  NaN  NaN  2015  43131966.0  21131346.0  NaN  NaN  NaN
12 NaN  Chaco  NaN  NaN  2010  1080017.0  534347.0  NaN  NaN  NaN
13 NaN  Chaco  NaN  NaN  2011  1092625.0  540391.0  NaN  NaN  NaN
14 NaN  Chaco  NaN  NaN  2012  1105280.0  546471.0  NaN  NaN  NaN
15 NaN  Chaco  NaN  NaN  2013  1117953.0  552570.0  NaN  NaN  NaN
16 NaN  Chaco  NaN  NaN  2014  1130608.0  558670.0  NaN  NaN  NaN
17 NaN  Chaco  NaN  NaN  2015  1143201.0  564746.0  NaN  NaN  NaN
18 NaN  Buenos Aires  NaN  NaN  2015  NaN  NaN  NaN  NaN  NaN
19 NaN  Buenos Aires  NaN  NaN  2020  NaN  NaN  NaN  NaN  NaN
20 NaN  Buenos Aires  NaN  NaN  2025  NaN  NaN  NaN  NaN  NaN
21 NaN  Buenos Aires  NaN  NaN  2030  NaN  NaN  NaN  NaN  NaN
22 NaN  Buenos Aires  NaN  NaN  2035  NaN  NaN  NaN  NaN  NaN
23 NaN  Buenos Aires  NaN  NaN  2040  NaN  NaN  NaN  NaN  NaN
24 NaN  Catamarca  NaN  NaN  2015  NaN  NaN  NaN  NaN  NaN
25 NaN  Catamarca  NaN  NaN  2020  NaN  NaN  NaN  NaN  NaN
26 NaN  Catamarca  NaN  NaN  2025  NaN  NaN  NaN  NaN  NaN
27 NaN  Catamarca  NaN  NaN  2030  NaN  NaN  NaN  NaN  NaN
28 NaN  Catamarca  NaN  NaN  2035  NaN  NaN  NaN  NaN  NaN
29 NaN  Catamarca  NaN  NaN  2040  NaN  NaN  NaN  NaN  NaN

In [41]: # Otra forma de identificar outliers usando percentiles
p99 = np.percentile(resultado["densidad"], 99)
p99

Out[41]:
nan

In [42]: # Vemos las observaciones por encima del percentil 99
resultado[resultado["densidad"] > p99]

Out[42]:
provincia_id  provincia  hogares  superficie_km2  año  población_total  población_varones  mujeres  densidad  densidad_estandarizada
0  18.0  Corrientes  267797.0  88199.0  2010  1017731.0  501452.0  NaN  11.53931  -1.464607
1  18.0  Corrientes  267797.0  88199.0  2011  1028248.0  506702.0  NaN  11.68673  -0.879017
2  18.0  Corrientes  267797.0  88199.0  2012  1038786.0  511969.0  NaN  11.77753  -0.292257
3  18.0  Corrientes  267797.0  88199.0  2013  1049325.0  517240.0  NaN  11.89724  0.294568
4  18.0  Corrientes  267797.0  88199.0  2014  1059836.0  522500.0  NaN  12.016417  0.87815
5  18.0  Corrientes  267797.0  88199.0  2015  1070283.0  527731.0  NaN  12.134865  1.461507
6  NaN  Total País  NaN  NaN  2010  40788453.0  19940704.0  NaN  NaN  NaN
7  NaN  Total País  NaN  NaN  2011  41261490.0  20180791.0  NaN  NaN  NaN
8  NaN  Total País  NaN  NaN  2012  41733271.0  20420391.0  NaN  NaN  NaN
9  NaN  Total País  NaN  NaN  2013  42202935.0  20696037.0
```