

GALO STRAPPINI

TECLAB

API 2

1ER PASO

Carga la base de proyecciones de población por año (en formato .csv) y las otras bases necesarias (ejemplo: expectativa de vida, hogares y vivienda, etc). Tené en cuenta que algunas bases contienen datos de varios años y otras tienen exclusivamente el año del censo 2010. Deberás realizár los chequeos básicos (head(), describe(), info()) para inspeccionar los datos.

```
import pandas as pd
import numpy as np
import os
```

✓ 0.0s

Python

```
poblacion = pd.read_csv('poblacion.csv', encoding='latin1')
esperanza = pd.read_csv('esperanza_de_vida.csv', encoding='latin1')
hogares = pd.read_csv('hogares_viviendas_superficie.csv', encoding='latin1')
```

✓ 0.0s

Python

```
#.HEAD - MOSTRAR PRIMER FILA DE LA TABLA POBLACION
#.TAIL - MOSTRAR ULTIMA FILA DE LA TABLA POBLACION
#.SHAPE - MOSTRAR TAMAÑO DE LA TABLA POBLACION
display(poblacion.head())
display(poblacion.tail())
display(poblacion.shape)
display(poblacion.dtypes)
```

✓ 0.0s

Pythor

2DO PASO

Calculá un nuevo campo que determine la densidad de población (población / superficie) y usá la función describe() sobre este campo nuevo.

```
# - Esta función filtra el DataFrame 'poblacion' para excluir las
# filas donde la columna 'provincia' es igual a 'Total País'.
pobla_filtrado = poblacion[poblacion['provincia'] != 'Total País']
pobla_filtrado

#.MERGE - UNIR TABLAS
# - Une la tabla de poblacion con la tabla de esperanza de vida
# - La columna de unión es 'provincia'
# - El tipo de unión es 'left' (izquierda)
# - La tabla resultante se llama 'poblac_superf'
# - Se utiliza la columna 'provincia' de la tabla de poblacion
# y la columna 'provincia' de la tabla de esperanza de vida
poblac_superf = pd.merge(pobla_filtrado, hogares, left_on=['provincia'],
| | | | | right_on=['provincia'], how='left')
poblac_superf

# - Esta función calcula la densidad poblacional dividiendo la
# población total entre la superficie en km2
poblac_superf['densidad_poblacional'] = (poblac_superf['poblacion_total']
| | | | | / poblac_superf['superficie_km2'])
poblac_superf
```

```
poblac superf['densidad poblacional'].describe()
```

✓ 0.0s

```
count      744.000000
mean       654.743585
std        3063.717821
min         0.131340
25%         4.424877
50%         9.406474
75%        20.322919
max       15437.170000
Name: densidad poblacional, dtype: float64
```

3ER PASO

Identificá si existe algún valor extremo en la densidad de población. De existir deberás explicar a qué podría deberse esto.

```
# - La función 'np.mean' calcula la media
# de la columna 'densidad_poblacional'
# - La función 'np.std' calcula la desviación estándar
# de la columna 'densidad_poblacional'
# - Un valor de 'fuera_de_rango' mayor a 3
# indica que el valor está muy por encima de la media
poblac_superf['fuera_de_rango'] = ((poblac_superf['densidad_poblacional'] -
np.mean(poblac_superf['densidad_poblacional']))
/ np.std(poblac_superf['densidad_poblacional']))
poblac_superf
```

```
poblac_superf[['provincia', 'densidad_poblacional']].drop_duplicates(subset=['provincia'])
```

EXPLICACION

El valor extremo se debe a que áreas urbanas con alta concentración de habitantes, como Capital Federal, tienen una densidad mucho mayor en comparación con regiones menos pobladas, como Tierra del Fuego. Esto refleja diferencias reales en la distribución demográfica.

	provincia	densidad_poblacional
0	Capital Federal	15142.405000
31	Buenos Aires	51.100208
62	Catamarca	3.680981
93	Córdoba	20.402883
124	Corrientes	11.539031
155	Chaco	10.839953
186	Chubut	2.285113
217	Entre Ríos	15.937523
248	Formosa	7.654456
279	Jujuy	12.843402
310	La Pampa	2.279894
341	La Rioja	3.820049
372	Mendoza	11.924832
403	Misiones	37.357102
434	Neuquén	6.079105
465	Río Negro	3.193278
496	Salta	7.969174
527	San Juan	7.764286
558	San Luis	5.784437
589	Santa Cruz	1.129165
620	Santa Fe	24.494252
651	Santiago del Estero	6.448402
682	Tucumán	66.117253
713	Tierra del Fuego	0.131340

4TO PASO

Genera un informe breve sobre una de las variables originales (por ejemplo, superficie), con algunas medidas resumen (percentiles, promedio, etc.) y tus respuestas.

```
pd.options.display.float_format = '{:,.2f}'.format
resumen_superficie = hogares['superficie_km2'].describe(percentiles=[0.25, 0.5, 0.75])

print("Medidas estadísticas de la superficie (en km²):")
print(resumen_superficie)
```

```
Medidas estadísticas de la superficie (en km²):
count          24.00
mean         156,719.75
std          194,090.09
min           200.00
25%           78,272.75
50%          101,117.50
75%          157,946.25
max          1,002,445.00
Name: superficie_km2, dtype: float64
```

INFORME

Las provincias en Argentina tienen tamaños muy distintos. Algunas son muy grandes, como Buenos Aires, mientras que otras, como Capital Federal, son mucho más pequeñas.

- El promedio de superficie es de 156,720 km², pero esto está influenciado por provincias enormes.
- La mediana es de 101,118 km², lo que significa que la mitad de las provincias tienen menos superficie que ese número y la otra mitad, más.
- El 25% de las provincias más pequeñas tiene menos de 78,273 km².
- El 75% de las provincias tiene menos de 157,946 km², lo que indica que solo unas pocas tienen territorios muy grandes.

GRACIAS