

Avance 1: Carga y transformación de los datos

```
In [17]: import pandas as pd
import numpy as np
```

Se carga el archivo csv

```
In [19]: datos = pd.read_csv('data_latinoamerica.csv')
```

Se muestran las primeras 10 filas del dataset

```
In [20]: datos.head(10)
```

```
Out[20]:
```

	location_key	date	country_code	country_name	new_confirmed	new_deceased	cumu
0	AR	2020-01-01	AR	Argentina	3.0	0.0	
1	AR	2020-01-02	AR	Argentina	14.0	0.0	
2	AR	2020-01-03	AR	Argentina	3.0	0.0	
3	AR	2020-01-04	AR	Argentina	7.0	0.0	
4	AR	2020-01-05	AR	Argentina	5.0	0.0	
5	AR	2020-01-06	AR	Argentina	9.0	0.0	
6	AR	2020-01-07	AR	Argentina	4.0	0.0	
7	AR	2020-01-08	AR	Argentina	3.0	0.0	
8	AR	2020-01-09	AR	Argentina	0.0	0.0	
9	AR	2020-01-10	AR	Argentina	1.0	0.0	

10 rows × 50 columns



Se muestra la cantidad de filas y columnas

```
In [21]: datos.shape
```

Out[21]: (12216057, 50)

```
In [22]: datos['country_name'].unique()
```

```
Out[22]: array(['Argentina', 'Bolivia', 'Brazil', 'Chile', 'Colombia',  
               'Costa Rica', 'Cuba', 'Dominican Republic', 'Ecuador', 'Guatemala',  
               'Honduras', 'Mexico', 'Nicaragua', 'Panama', 'Peru', 'Puerto Rico',  
               'Paraguay', 'El Salvador', 'Uruguay', 'Venezuela'], dtype=object)
```

```
In [91]: paises_interes = ('Argentina', 'Chile', 'Mexico', 'Peru', 'Colombia', 'Brazil')
```

```
In [24]: datos_paises_interes = datos[datos ['country_name'].isin (paises_interes)]
```

```
In [25]: datos_paises_interes.shape
```

Out[25]: (11970289, 50)

```
In [26]: datos_paises_interes['country_name'].unique()
```

```
Out[26]: array(['Argentina', 'Brazil', 'Chile', 'Colombia', 'Mexico', 'Peru'],  
               dtype=object)
```

```
In [27]: datos_paises_interes = datos_paises_interes[datos_paises_interes ['location_key'].i
```

```
In [28]: datos_paises_interes.shape
```

Out[28]: (5946, 50)

```
In [29]: datos_paises_interes = datos_paises_interes[datos_paises_interes ['date'] >= '2021
```

```
In [30]: datos_paises_interes.shape
```

Out[30]: (3750, 50)

```
In [31]: datos_paises_interes = datos_paises_interes.dropna(axis=0, how='all' )  
datos_paises_interes = datos_paises_interes.dropna(axis=1, how='all' )
```

```
In [32]: datos_paises_interes.shape
```

Out[32]: (3750, 50)

```
In [33]: datos_paises_interes.isnull().sum()
```

```

Out[33]: location_key      0
         date              0
         country_code      0
         country_name      0
         new_confirmed     21
         new_deceased      21
         cumulative_confirmed 21
         cumulative_deceased 21
         cumulative_vaccine_doses_administered 589
         population        0
         population_male   0
         population_female 0
         population_rural  0
         population_urban  0
         population_density 0
         human_development_index 0
         population_age_00_09 0
         population_age_10_19 0
         population_age_20_29 0
         population_age_30_39 0
         population_age_40_49 0
         population_age_50_59 0
         population_age_60_69 0
         population_age_70_79 0
         population_age_80_and_older 0
         gdp_usd           0
         gdp_per_capita_usd 0
         latitude          0
         longitude         0
         area_sq_km        0
         smoking_prevalence 0
         diabetes_prevalence 0
         infant_mortality_rate 0
         nurses_per_1000    0
         physicians_per_1000 0
         average_temperature_celsius 42
         minimum_temperature_celsius 41
         maximum_temperature_celsius 41
         rainfall_mm        90
         relative_humidity  42
         population_largest_city 0
         area_rural_sq_km   0
         area_urban_sq_km  0
         life_expectancy    0
         adult_male_mortality_rate 0
         adult_female_mortality_rate 0
         pollution_mortality_rate 0
         comorbidity_mortality_rate 0
         new_recovered      2119
         cumulative_recovered 2740
         dtype: int64

```

```

In [34]: datos_paises_interes = datos_paises_interes.drop(columns=['new_recovered', 'cumulat

```

```

In [35]: datos_paises_interes.shape

```

Out[35]: (3750, 48)

```
In [36]: datos_paises_interes.head(10)
```

Out[36]:

	location_key	date	country_code	country_name	new_confirmed	new_deceased	cur
366	AR	2021-01-01	AR	Argentina	2685.0	140.0	
367	AR	2021-01-02	AR	Argentina	7767.0	166.0	
368	AR	2021-01-03	AR	Argentina	4934.0	157.0	
369	AR	2021-01-04	AR	Argentina	13953.0	157.0	
370	AR	2021-01-05	AR	Argentina	14085.0	160.0	
371	AR	2021-01-06	AR	Argentina	14496.0	131.0	
372	AR	2021-01-07	AR	Argentina	13722.0	160.0	
373	AR	2021-01-08	AR	Argentina	13932.0	162.0	
374	AR	2021-01-09	AR	Argentina	9959.0	158.0	
375	AR	2021-01-10	AR	Argentina	5174.0	165.0	

10 rows × 48 columns



```
In [37]: datos_paises_interes.describe()
```

Out[37]:

	new_confirmed	new_deceased	cumulative_confirmed	cumulative_deceased	cumulative_vaccine_doses_administered
count	3729.000000	3729.000000	3.729000e+03	3729.000000	3.729000e+03
mean	13924.598284	277.300885	6.777284e+06	192974.495307	1.929745e+06
std	24258.010004	508.206954	7.995728e+06	182936.954037	1.829370e+06
min	-573.000000	0.000000	9.710000e+02	1.000000	9.710000e+02
25%	1531.000000	26.000000	2.125355e+06	59146.000000	2.125355e+06
50%	5164.000000	100.000000	3.629796e+06	139621.000000	3.629796e+06
75%	14911.000000	314.000000	6.223497e+06	215028.000000	6.223497e+06
max	298408.000000	11447.000000	3.456883e+07	685203.000000	3.456883e+07

8 rows × 44 columns



```
In [38]: for pais in paises_interes:
          datos_paises_interes['new_confirmed'] = datos_paises_interes['new_confirmed']
          datos_paises_interes['new_deceased'] = datos_paises_interes['new_deceased'].fillna(0)
          datos_paises_interes['cumulative_confirmed'] = datos_paises_interes['cumulative_confirmed']
          datos_paises_interes['cumulative_deceased'] = datos_paises_interes['cumulative_deceased']
          datos_paises_interes['cumulative_vaccine_doses_administered'] = datos_paises_interes['cumulative_vaccine_doses_administered']
          datos_paises_interes['average_temperature_celsius'] = datos_paises_interes['average_temperature_celsius']
          datos_paises_interes['minimum_temperature_celsius'] = datos_paises_interes['minimum_temperature_celsius']
          datos_paises_interes['maximum_temperature_celsius'] = datos_paises_interes['maximum_temperature_celsius']
          datos_paises_interes['rainfall_mm'] = datos_paises_interes['rainfall_mm'].fillna(0)
          datos_paises_interes['relative_humidity'] = datos_paises_interes['relative_humidity']
```

```
In [39]: datos_paises_interes.isnull().sum()
```

```

Out[39]: location_key      0
         date              0
         country_code      0
         country_name      0
         new_confirmed     0
         new_deceased      0
         cumulative_confirmed 0
         cumulative_deceased 0
         cumulative_vaccine_doses_administered 108
         population        0
         population_male   0
         population_female 0
         population_rural  0
         population_urban  0
         population_density 0
         human_development_index 0
         population_age_00_09 0
         population_age_10_19 0
         population_age_20_29 0
         population_age_30_39 0
         population_age_40_49 0
         population_age_50_59 0
         population_age_60_69 0
         population_age_70_79 0
         population_age_80_and_older 0
         gdp_usd           0
         gdp_per_capita_usd 0
         latitude          0
         longitude         0
         area_sq_km        0
         smoking_prevalence 0
         diabetes_prevalence 0
         infant_mortality_rate 0
         nurses_per_1000    0
         physicians_per_1000 0
         average_temperature_celsius 0
         minimum_temperature_celsius 0
         maximum_temperature_celsius 0
         rainfall_mm        0
         relative_humidity  0
         population_largest_city 0
         area_rural_sq_km   0
         area_urban_sq_km   0
         life_expectancy    0
         adult_male_mortality_rate 0
         adult_female_mortality_rate 0
         pollution_mortality_rate 0
         comorbidity_mortality_rate 0
         dtype: int64

```

```

In [40]: for pais in paises_interes:
         datos_paises_interes['cumulative_vaccine_doses_administered'] = datos_paises_in

```

```

In [41]: datos_paises_interes.isnull().sum()

```

```

Out[41]: location_key      0
         date              0
         country_code      0
         country_name      0
         new_confirmed     0
         new_deceased       0
         cumulative_confirmed 0
         cumulative_deceased 0
         cumulative_vaccine_doses_administered 0
         population        0
         population_male   0
         population_female  0
         population_rural  0
         population_urban  0
         population_density 0
         human_development_index 0
         population_age_00_09 0
         population_age_10_19 0
         population_age_20_29 0
         population_age_30_39 0
         population_age_40_49 0
         population_age_50_59 0
         population_age_60_69 0
         population_age_70_79 0
         population_age_80_and_older 0
         gdp_usd           0
         gdp_per_capita_usd 0
         latitude          0
         longitude         0
         area_sq_km        0
         smoking_prevalence 0
         diabetes_prevalence 0
         infant_mortality_rate 0
         nurses_per_1000    0
         physicians_per_1000 0
         average_temperature_celsius 0
         minimum_temperature_celsius 0
         maximum_temperature_celsius 0
         rainfall_mm        0
         relative_humidity  0
         population_largest_city 0
         area_rural_sq_km   0
         area_urban_sq_km   0
         life_expectancy    0
         adult_male_mortality_rate 0
         adult_female_mortality_rate 0
         pollution_mortality_rate 0
         comorbidity_mortality_rate 0
         dtype: int64

```

```
In [42]: datos_paises_interes.to_csv('datos_paises_interes.csv', index=False)
```

```
In [43]: datos_paises_interes = pd.read_csv('datos_paises_interes.csv')
```

```
In [44]: datos_paises_interes.head(10)
```

Out[44]:

	location_key	date	country_code	country_name	new_confirmed	new_deceased	cumu
0	AR	2021-01-01	AR	Argentina	2685.0	140.0	
1	AR	2021-01-02	AR	Argentina	7767.0	166.0	
2	AR	2021-01-03	AR	Argentina	4934.0	157.0	
3	AR	2021-01-04	AR	Argentina	13953.0	157.0	
4	AR	2021-01-05	AR	Argentina	14085.0	160.0	
5	AR	2021-01-06	AR	Argentina	14496.0	131.0	
6	AR	2021-01-07	AR	Argentina	13722.0	160.0	
7	AR	2021-01-08	AR	Argentina	13932.0	162.0	
8	AR	2021-01-09	AR	Argentina	9959.0	158.0	
9	AR	2021-01-10	AR	Argentina	5174.0	165.0	

10 rows × 48 columns



In [45]:

```
for variable in datos_paises_interes.columns:
    if datos_paises_interes[variable].dtype == 'int64' or datos_paises_interes[variable].dtype == 'float64':
        print(f"La mediana de {variable} es: {datos_paises_interes[variable].median}")
        print(f"La variancia de {variable} es: {datos_paises_interes[variable].var}")
        print(f"El valor minimo de {variable} es: {datos_paises_interes[variable].min}")
        print(f"El valor maximo de {variable} es: {datos_paises_interes[variable].max}")
        print(f"El rango de {variable} es: {datos_paises_interes[variable].max() - datos_paises_interes[variable].min()}")
        print(f"El promedio de {variable} es: {datos_paises_interes[variable].mean}")
        print(f"La desviacion estandar de {variable} es: {datos_paises_interes[variable].std}")
    print("-----"*10)
```


La mediana de new_confirmed es: 5221.5
La variancia de new_confirmed es: 585160750.1670985
El valor minimo de new_confirmed es: -573.0
El valor maximo de new_confirmed es: 298408.0
El rango de new_confirmed es: 298981.0
El promedio de new_confirmed es: 13918.83207342995
La desviacion estandar de new_confirmed es: 24190.09611735965

La mediana de new_deceased es: 102.0
La variancia de new_deceased es: 256941.35128844946
El valor minimo de new_deceased es: 0.0
El valor maximo de new_deceased es: 11447.0
El rango de new_deceased es: 11447.0
El promedio de new_deceased es: 276.50057326892113
La desviacion estandar de new_deceased es: 506.8938264453903

La mediana de cumulative_confirmed es: 3640785.5
La variancia de cumulative_confirmed es: 64019086123943.49
El valor minimo de cumulative_confirmed es: 971.0
El valor maximo de cumulative_confirmed es: 34568833.0
El rango de cumulative_confirmed es: 34567862.0
El promedio de cumulative_confirmed es: 6786495.4992
La desviacion estandar de cumulative_confirmed es: 8001192.793824149

La mediana de cumulative_deceased es: 139614.0
La variancia de cumulative_deceased es: 33456831054.01476
El valor minimo de cumulative_deceased es: 1.0
El valor maximo de cumulative_deceased es: 685203.0
El rango de cumulative_deceased es: 685202.0
El promedio de cumulative_deceased es: 192863.5696
La desviacion estandar de cumulative_deceased es: 182912.08558762528

La mediana de cumulative_vaccine_doses_administered es: 47713014.0
La variancia de cumulative_vaccine_doses_administered es: 8747794082611842.0
El valor minimo de cumulative_vaccine_doses_administered es: 18.0
El valor maximo de cumulative_vaccine_doses_administered es: 347868481.0
El rango de cumulative_vaccine_doses_administered es: 347868463.0
El promedio de cumulative_vaccine_doses_administered es: 80698755.43866667
La desviacion estandar de cumulative_vaccine_doses_administered es: 93529642.8016906

4

La mediana de population es: 47910798.0
La variancia de population es: 4507476512829145.0
El valor minimo de population es: 17574003.0
El valor maximo de population es: 212559409.0
El rango de population es: 194985406.0
El promedio de population es: 77721474.16666667
La desviacion estandar de population es: 67137742.833887

La mediana de population_male es: 22254165.0
La variancia de population_male es: 1101239470229229.0
El valor minimo de population_male es: 8972014.0
El valor maximo de population_male es: 104435783.0
El rango de population_male es: 95463769.0
El promedio de population_male es: 37870352.5
La desviacion estandar de population_male es: 33184928.3595615

La mediana de population_female es: 23245825.0
La variancia de population_female es: 1189150601199526.5
El valor minimo de population_female es: 8601989.0
El valor maximo de population_female es: 108123626.0
El rango de population_female es: 99521637.0
El promedio de population_female es: 39271616.5
La desviacion estandar de population_female es: 34484063.00306747

La mediana de population_rural es: 8316127.5
La variancia de population_rural es: 101621319832048.36
El valor minimo de population_rural es: 2341903.0
El valor maximo de population_rural es: 27807886.0
El rango de population_rural es: 25465983.0
El promedio de population_rural es: 12554975.833333334
La desviacion estandar de population_rural es: 10080740.043868225

La mediana de population_urban es: 41083436.5
La variancia de population_urban es: 3398339529217889.0
El valor minimo de population_urban es: 16610135.0
El valor maximo de population_urban es: 183241641.0
El rango de population_urban es: 166631506.0
El promedio de population_urban es: 68339307.83333333
La desviacion estandar de population_urban es: 58295278.79012064

La mediana de population_density es: 25.7345
La variancia de population_density es: 283.56777021094507
El valor minimo de population_density es: 16.515
El valor maximo de population_density es: 66.325
El rango de population_density es: 49.81
El promedio de population_density es: 34.266833333333334
La desviacion estandar de population_density es: 16.839470603642653

La mediana de human_development_index es: 0.755
La variancia de human_development_index es: 0.001321574642126788
El valor minimo de human_development_index es: 0.747
El valor maximo de human_development_index es: 0.832
El rango de human_development_index es: 0.08499999999999996
El promedio de human_development_index es: 0.7773333333333333
La desviacion estandar de human_development_index es: 0.036353468089396754

La mediana de population_age_00_09 es: 7066513.0
La variancia de population_age_00_09 es: 95173712081376.69
El valor minimo de population_age_00_09 es: 2428079.0
El valor maximo de population_age_00_09 es: 29076910.0
El rango de population_age_00_09 es: 26648831.0
El promedio de population_age_00_09 es: 12059083.0
La desviacion estandar de population_age_00_09 es: 9755701.51661974

La mediana de population_age_10_19 es: 7582694.5
La variancia de population_age_10_19 es: 107108210833415.16
El valor minimo de population_age_10_19 es: 2493879.0
El valor maximo de population_age_10_19 es: 31160446.0
El rango de population_age_10_19 es: 28666567.0
El promedio de population_age_10_19 es: 12636944.666666666
La desviacion estandar de population_age_10_19 es: 10349309.67907595

La mediana de population_age_20_29 es: 7638859.5
La variancia de population_age_20_29 es: 117224240949289.08
El valor minimo de population_age_20_29 es: 2995538.0
El valor maximo de population_age_20_29 es: 34104643.0
El rango de population_age_20_29 es: 31109105.0
El promedio de population_age_20_29 es: 12658900.666666666
La desviacion estandar de population_age_20_29 es: 10827014.40607193

La mediana de population_age_30_39 es: 6793723.5
La variancia de population_age_30_39 es: 120736455701022.78
El valor minimo de population_age_30_39 es: 2945404.0
El valor maximo de population_age_30_39 es: 34476762.0
El rango de population_age_30_39 es: 31531358.0
El promedio de population_age_30_39 es: 12025529.166666666
La desviacion estandar de population_age_30_39 es: 10988014.183692282

La mediana de population_age_40_49 es: 5478956.0
La variancia de population_age_40_49 es: 87592296542412.97
El valor minimo de population_age_40_49 es: 2578404.0
El valor maximo de population_age_40_49 es: 29462006.0
El rango de population_age_40_49 es: 26883602.0
El promedio de population_age_40_49 es: 9932776.0
La desviacion estandar de population_age_40_49 es: 9359075.62435591

La mediana de population_age_50_59 es: 4727961.5
La variancia de population_age_50_59 es: 58772956406172.875
El valor minimo de population_age_50_59 es: 2352271.0
El valor maximo de population_age_50_59 es: 24421202.0
El rango de population_age_50_59 es: 22068931.0
El promedio de population_age_50_59 es: 7993603.333333333
La desviacion estandar de population_age_50_59 es: 7666352.222939726

La mediana de population_age_60_69 es: 3342804.5
La variancia de population_age_60_69 es: 27781880922376.86
El valor minimo de population_age_60_69 es: 1791787.0
El valor maximo de population_age_60_69 es: 16896862.0
El rango de population_age_60_69 es: 15105075.0
El promedio de population_age_60_69 es: 5442399.333333333
La desviacion estandar de population_age_60_69 es: 5270852.01104877

La mediana de population_age_70_79 es: 1898659.0
La variancia de population_age_70_79 es: 7306165053627.3
El valor minimo de population_age_70_79 es: 993126.0
El valor maximo de population_age_70_79 es: 8801551.0
El rango de population_age_70_79 es: 7808425.0
El promedio de population_age_70_79 es: 2964913.0
La desviacion estandar de population_age_70_79 es: 2702991.8708030367

La mediana de population_age_80_and_older es: 969818.5
La variancia de population_age_80_and_older es: 1567927236717.9304
El valor minimo de population_age_80_and_older es: 537721.0
El valor maximo de population_age_80_and_older es: 4159027.0
El rango de population_age_80_and_older es: 3621306.0
El promedio de population_age_80_and_older es: 1451953.166666667
La desviacion estandar de population_age_80_and_older es: 1252169.0128404913

La mediana de gdp_usd es: 386733127531.0
La variancia de gdp_usd es: 3.680673752713903e+23
El valor minimo de gdp_usd es: 226848050819.0
El valor maximo de gdp_usd es: 1839758040765.0
El rango de gdp_usd es: 1612909989946.0
El promedio de gdp_usd es: 730112870585.6666
La desviacion estandar de gdp_usd es: 606685565405.4993

La mediana de gdp_per_capita_usd es: 9290.0
La variancia de gdp_per_capita_usd es: 7651020.744420735
El valor minimo de gdp_per_capita_usd es: 6432.0
El valor maximo de gdp_per_capita_usd es: 14896.0
El rango de gdp_per_capita_usd es: 8464.0
El promedio de gdp_per_capita_usd es: 9481.833333333334
La desviacion estandar de gdp_per_capita_usd es: 2766.047856495027

La mediana de latitude es: -11.7
La variancia de latitude es: 400.8457810971814
El valor minimo de latitude es: -34.0
El valor maximo de latitude es: 23.0
El rango de latitude es: 57.0
El promedio de latitude es: -10.566666666666666
La desviacion estandar de latitude es: 20.021133361954845

La mediana de longitude es: -72.125
La variancia de longitude es: 222.52635425002225
El valor minimo de longitude es: -102.0
El valor maximo de longitude es: -53.0
El rango de longitude es: 49.0
El promedio de longitude es: -73.20833333333333
La desviacion estandar de longitude es: 14.917317260486962

La mediana de area_sq_km es: 1624797.5
La variancia de area_sq_km es: 7096394976860.637
El valor minimo de area_sq_km es: 756700.0
El valor maximo de area_sq_km es: 8515770.0
El rango de area_sq_km es: 7759070.0
El promedio de area_sq_km es: 2740702.1666666665
La desviacion estandar de area_sq_km es: 2663905.9624657617

La mediana de smoking_prevalence es: 13.95
La variancia de smoking_prevalence es: 114.5385991820041
El valor minimo de smoking_prevalence es: 4.8
El valor maximo de smoking_prevalence es: 37.8
El rango de smoking_prevalence es: 33.0
El promedio de smoking_prevalence es: 16.883333333333336
La desviacion estandar de smoking_prevalence es: 10.70227074886466

La mediana de diabetes_prevalence es: 8.0
La variancia de diabetes_prevalence es: 6.647328176402597
El valor minimo de diabetes_prevalence es: 5.9
El valor maximo de diabetes_prevalence es: 13.5
El rango de diabetes_prevalence es: 7.6
El promedio de diabetes_prevalence es: 8.733333333333334
La desviacion estandar de diabetes_prevalence es: 2.578241295224828

La mediana de infant_mortality_rate es: 11.05
La variancia de infant_mortality_rate es: 5.007168578287543
El valor minimo de infant_mortality_rate es: 6.2
El valor maximo de infant_mortality_rate es: 12.8
El rango de infant_mortality_rate es: 6.600000000000005
El promedio de infant_mortality_rate es: 10.35
La desviacion estandar de infant_mortality_rate es: 2.2376703462055225

La mediana de nurses_per_1000 es: 2.5197000000000003
La variancia de nurses_per_1000 es: 21.214228169734156
El valor minimo de nurses_per_1000 es: 1.3309
El valor maximo de nurses_per_1000 es: 13.3248
El rango de nurses_per_1000 es: 11.9939
El promedio de nurses_per_1000 es: 5.368366666666668
La desviacion estandar de nurses_per_1000 es: 4.605890594633589

La mediana de physicians_per_1000 es: 2.28375
La variancia de physicians_per_1000 es: 0.6432865277963014
El valor minimo de physicians_per_1000 es: 1.3048
El valor maximo de physicians_per_1000 es: 3.9901
El rango de physicians_per_1000 es: 2.6853
El promedio de physicians_per_1000 es: 2.4363166666666665
La desviacion estandar de physicians_per_1000 es: 0.8020514495942896

La mediana de average_temperature_celsius es: 21.590741
La variancia de average_temperature_celsius es: 36.823139740808166
El valor minimo de average_temperature_celsius es: 3.432099
El valor maximo de average_temperature_celsius es: 39.138889
El rango de average_temperature_celsius es: 35.70679
El promedio de average_temperature_celsius es: 21.023542656
La desviacion estandar de average_temperature_celsius es: 6.068207292175191

La mediana de minimum_temperature_celsius es: 15.856790499999999
La variancia de minimum_temperature_celsius es: 49.11473648230403
El valor minimo de minimum_temperature_celsius es: -5.383333
El valor maximo de minimum_temperature_celsius es: 33.0
El rango de minimum_temperature_celsius es: 38.383333
El promedio de minimum_temperature_celsius es: 14.967169933866666
La desviacion estandar de minimum_temperature_celsius es: 7.008190671086513

La mediana de maximum_temperature_celsius es: 27.405092500000002
La variancia de maximum_temperature_celsius es: 27.59834112586727
El valor minimo de maximum_temperature_celsius es: 6.950617
El valor maximo de maximum_temperature_celsius es: 41.944444
El rango de maximum_temperature_celsius es: 34.993826999999996
El promedio de maximum_temperature_celsius es: 27.058436530666665
La desviacion estandar de maximum_temperature_celsius es: 5.253412331605741

La mediana de rainfall_mm es: 0.0
La variancia de rainfall_mm es: 18.931911982072346
El valor minimo de rainfall_mm es: 0.0
El valor maximo de rainfall_mm es: 46.736
El rango de rainfall_mm es: 46.736
El promedio de rainfall_mm es: 1.4456042250666665
La desviacion estandar de rainfall_mm es: 4.351081702527814

La mediana de relative_humidity es: 65.237704
La variancia de relative_humidity es: 254.0996018894969
El valor minimo de relative_humidity es: 10.296407
El valor maximo de relative_humidity es: 94.817706
El rango de relative_humidity es: 84.521299
El promedio de relative_humidity es: 62.27568529386665
La desviacion estandar de relative_humidity es: 15.940501933424082

La mediana de population_largest_city es: 12918324.5
La variancia de population_largest_city es: 32604963649152.105
El valor minimo de population_largest_city es: 6723516.0
El valor maximo de population_largest_city es: 21846507.0
El rango de population_largest_city es: 15122991.0
El promedio de population_largest_city es: 14438882.0
La desviacion estandar de population_largest_city es: 5710075.625519517

La mediana de area_rural_sq_km es: 1543881.0
La variancia de area_rural_sq_km es: 6680878694738.18
El valor minimo de area_rural_sq_km es: 709418.0
El valor maximo de area_rural_sq_km es: 8241430.0
El rango de area_rural_sq_km es: 7532012.0
El promedio de area_rural_sq_km es: 2636579.5
La desviacion estandar de area_rural_sq_km es: 2584739.5796749387

La mediana de area_urban_sq_km es: 45582.0
La variancia de area_urban_sq_km es: 2036293653.224193
El valor minimo de area_urban_sq_km es: 12027.0
El valor maximo de area_urban_sq_km es: 134981.0
El rango de area_urban_sq_km es: 122954.0
El promedio de area_urban_sq_km es: 59502.5
La desviacion estandar de area_urban_sq_km es: 45125.31056097224

La mediana de life_expectancy es: 76.518
La variancia de life_expectancy es: 2.5516663610296075
El valor minimo de life_expectancy es: 74.992
El valor maximo de life_expectancy es: 80.042
El rango de life_expectancy es: 5.049999999999997
El promedio de life_expectancy es: 76.8085
La desviacion estandar de life_expectancy es: 1.5973936149332786

La mediana de adult_male_mortality_rate es: 149.351
La variancia de adult_male_mortality_rate es: 729.0075832444206
El valor minimo de adult_male_mortality_rate es: 107.669
El valor maximo de adult_male_mortality_rate es: 188.528
El rango de adult_male_mortality_rate es: 80.859
El promedio de adult_male_mortality_rate es: 154.27466666666666
La desviacion estandar de adult_male_mortality_rate es: 27.000140430087036

La mediana de adult_female_mortality_rate es: 82.149
La variancia de adult_female_mortality_rate es: 139.25465157375305
El valor minimo de adult_female_mortality_rate es: 59.035
El valor maximo de adult_female_mortality_rate es: 95.815
El rango de adult_female_mortality_rate es: 36.78
El promedio de adult_female_mortality_rate es: 81.42800000000001
La desviacion estandar de adult_female_mortality_rate es: 11.800620813065432

```
-----  
La mediana de pollution_mortality_rate es: 33.3  
La variancia de pollution_mortality_rate es: 169.71748021694674  
El valor minimo de pollution_mortality_rate es: 25.3  
El valor maximo de pollution_mortality_rate es: 63.9  
El rango de pollution_mortality_rate es: 38.599999999999994  
El promedio de pollution_mortality_rate es: 36.56666666666667  
La desviacion estandar de pollution_mortality_rate es: 13.027566166285503  
-----  
La mediana de comorbidity_mortality_rate es: 15.75  
La variancia de comorbidity_mortality_rate es: 2.77546234551436  
El valor minimo de comorbidity_mortality_rate es: 12.4  
El valor maximo de comorbidity_mortality_rate es: 16.6  
El rango de comorbidity_mortality_rate es: 4.200000000000001  
El promedio de comorbidity_mortality_rate es: 14.816666666666666  
La desviacion estandar de comorbidity_mortality_rate es: 1.6659718921741626  
-----
```

Avance 2:

```
In [1]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
In [2]: datos_paises_interes = pd.read_csv('datos_paises_interes.csv')
```

```
In [3]: datos_paises_interes.head(10)
```

Out[3]:

	location_key	date	country_code	country_name	new_confirmed	new_deceased	cumu
0	AR	2021-01-01	AR	Argentina	2685.0	140.0	
1	AR	2021-01-02	AR	Argentina	7767.0	166.0	
2	AR	2021-01-03	AR	Argentina	4934.0	157.0	
3	AR	2021-01-04	AR	Argentina	13953.0	157.0	
4	AR	2021-01-05	AR	Argentina	14085.0	160.0	
5	AR	2021-01-06	AR	Argentina	14496.0	131.0	
6	AR	2021-01-07	AR	Argentina	13722.0	160.0	
7	AR	2021-01-08	AR	Argentina	13932.0	162.0	
8	AR	2021-01-09	AR	Argentina	9959.0	158.0	
9	AR	2021-01-10	AR	Argentina	5174.0	165.0	

10 rows × 48 columns



```
In [8]: total_muertes = datos_paises_interes['new_deceased'].sum()
print(f"Total de muertes: {total_muertes}")
```

Total de muertes: 1036877.1497584542

```
In [4]: print(datos_paises_interes.groupby('country_name').describe())
```


	new_confirmed					
	count	mean	std	min	25%	\
country_name						
Argentina	625.0	12894.917874	23508.795780	0.0	1239.0	
Brazil	625.0	43069.839737	39555.782092	-573.0	14314.0	
Chile	625.0	6148.639474	6809.363543	273.0	1932.0	
Colombia	625.0	7438.746674	9225.670055	0.0	365.0	
Mexico	625.0	8885.231606	12005.345352	192.0	2116.0	
Peru	625.0	5075.617074	8018.170303	0.0	944.0	

	new_deceased					...	\
	50%	75%	max	count	mean	...	
country_name						...	
Argentina	5205.0	14269.000000	174174.0	625.0	134.388084	...	
Brazil	32321.0	62504.000000	298408.0	625.0	784.836442	...	
Chile	4239.0	7574.000000	38446.0	625.0	71.602484	...	
Colombia	2594.0	12894.917874	40415.0	625.0	185.079284	...	
Mexico	4253.0	10529.000000	81002.0	625.0	286.333063	...	
Peru	2244.0	6787.000000	58128.0	625.0	196.764084	...	

	pollution_mortality_rate		comorbidity_mortality_rate			\
	75%	max		count	mean	
country_name						
Argentina	26.6	26.6		625.0	15.8	
Brazil	29.9	29.9		625.0	16.6	
Chile	25.3	25.3		625.0	12.4	
Colombia	37.0	37.0		625.0	15.8	
Mexico	36.7	36.7		625.0	15.7	
Peru	63.9	63.9		625.0	12.6	

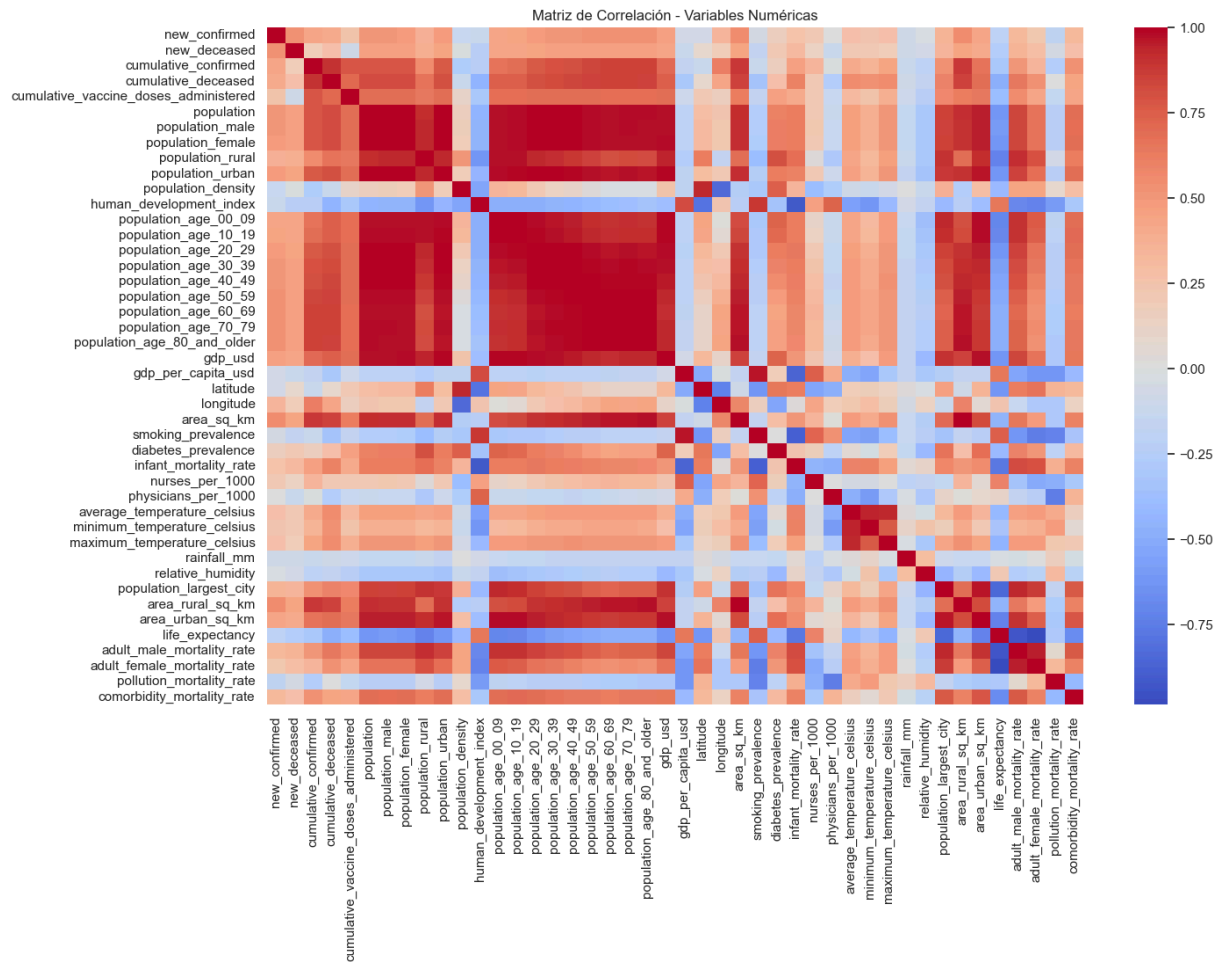
	std	min	25%	50%	75%	max
country_name						
Argentina	0.000000e+00	15.8	15.8	15.8	15.8	15.8
Brazil	3.555559e-15	16.6	16.6	16.6	16.6	16.6
Chile	1.777780e-15	12.4	12.4	12.4	12.4	12.4
Colombia	0.000000e+00	15.8	15.8	15.8	15.8	15.8
Mexico	0.000000e+00	15.7	15.7	15.7	15.7	15.7
Peru	1.777780e-15	12.6	12.6	12.6	12.6	12.6

[6 rows x 352 columns]

Se crea una matriz de correlaciones para comparar las variables

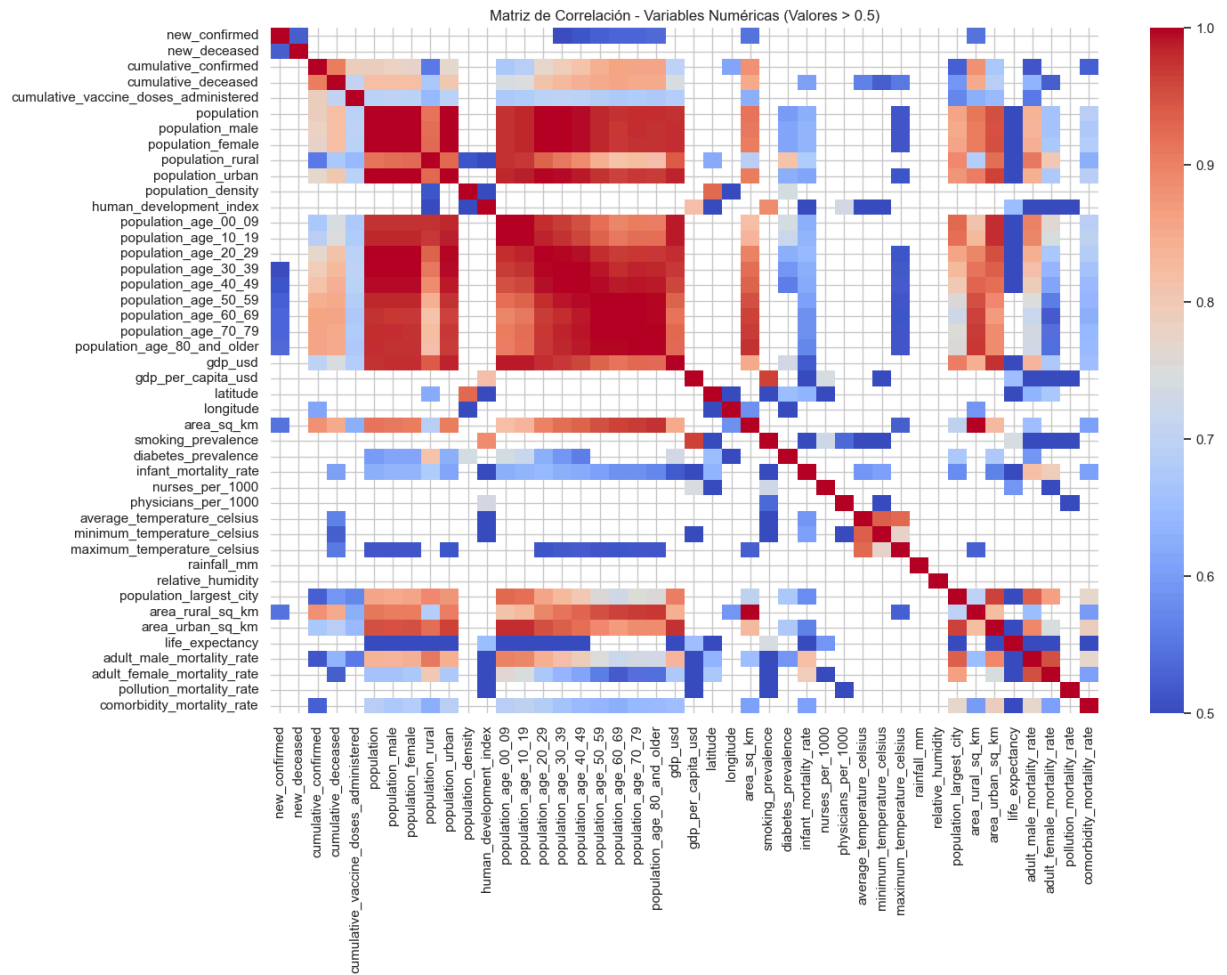
```
In [79]: sns.set_theme(style="whitegrid")
```

```
In [95]: matriz_correlacion = datos_paises_interes.select_dtypes(include=[np.number]).corr()
plt.figure(figsize=(15, 10))
sns.heatmap(matriz_correlacion, cmap='coolwarm')
plt.title('Matriz de Correlación - Variables Numéricas')
plt.show()
```



```
In [96]: correlacion_filtrada = matriz_correlacion.copy()
correlacion_filtrada[abs(correlacion_filtrada) < 0.5] = np.nan
```

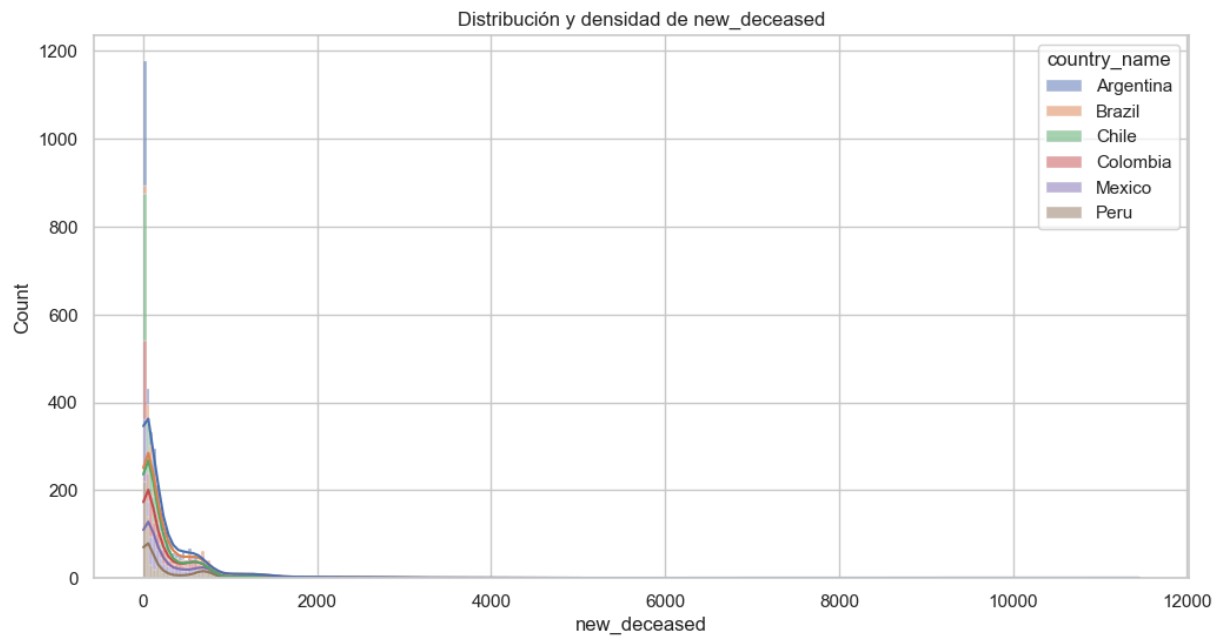
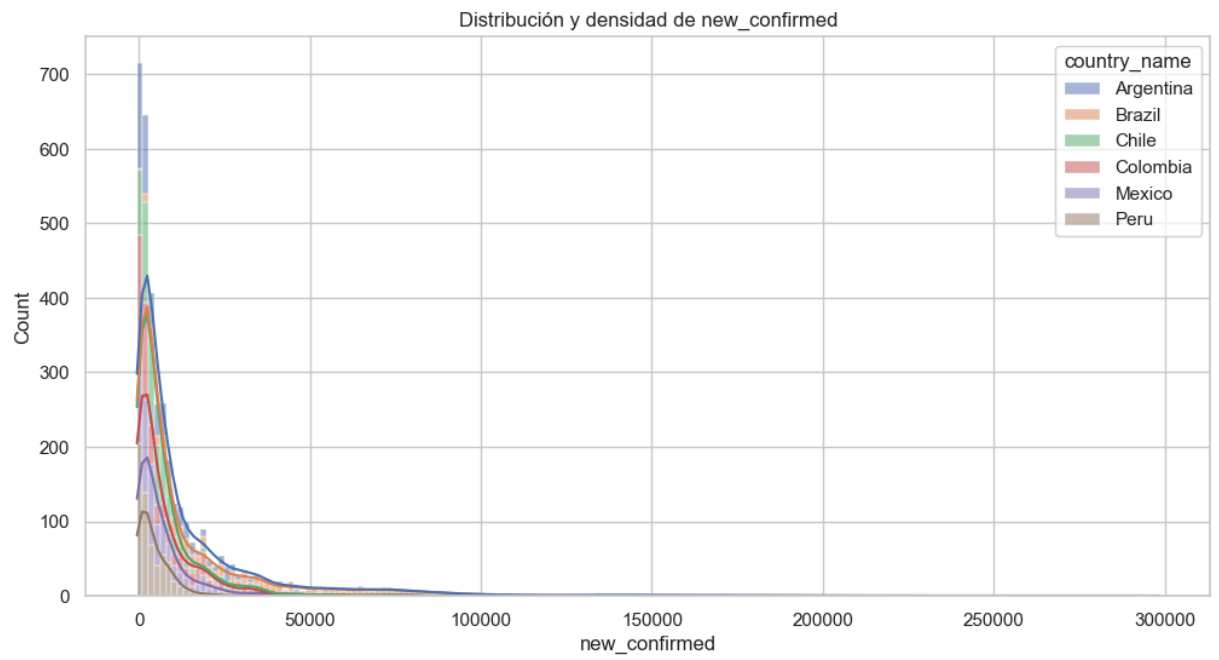
```
In [97]: plt.figure(figsize=(15, 10))
sns.heatmap(correlacion_filtrada, cmap='coolwarm', vmin=0.5, vmax=1)
plt.title('Matriz de Correlación - Variables Numéricas (Valores > 0.5)')
plt.show()
```

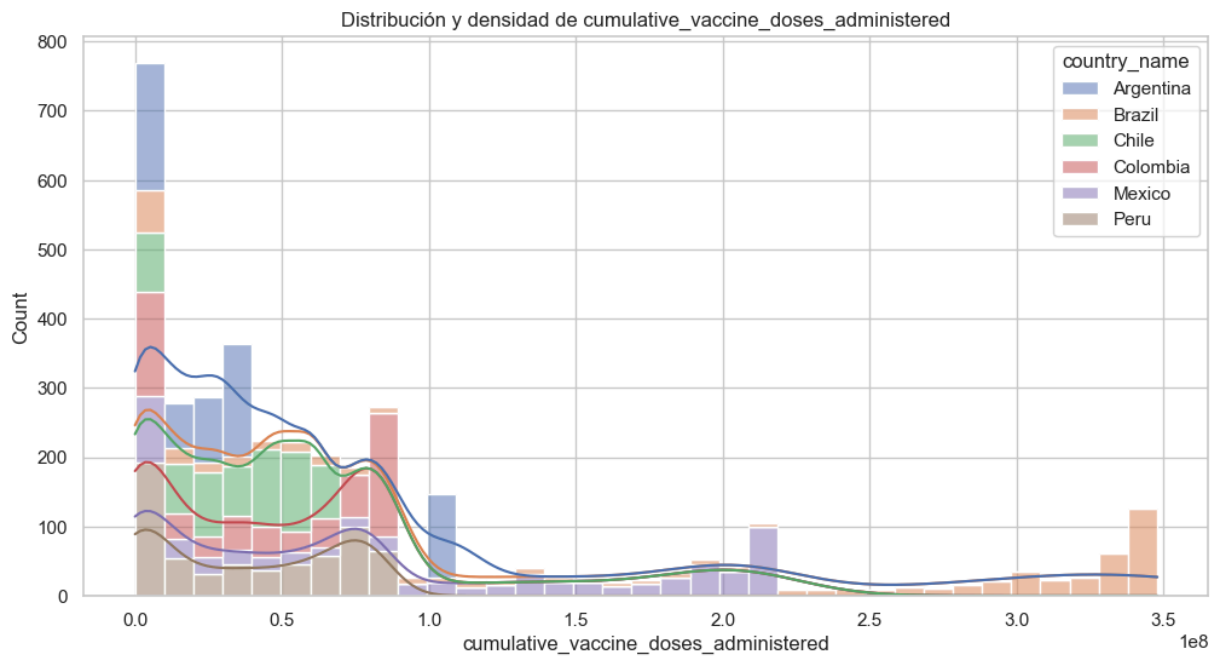


Histogramas

```
In [82]: variables = ['new_confirmed', 'new_deceased', 'cumulative_vaccine_doses_administere

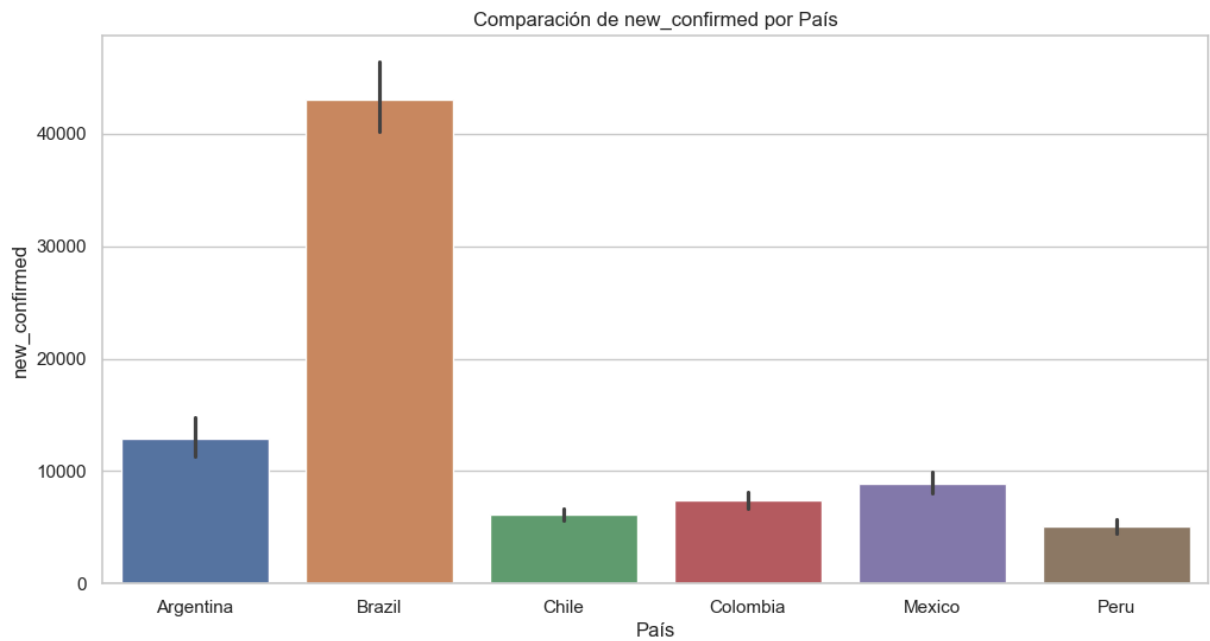
for variable in variables:
    plt.figure(figsize=(12,6))
    sns.histplot(data=datos_paises_interes, x=variable, hue='country_name', kde=True)
    plt.title(f'Distribución y densidad de {variable}')
    plt.xlabel(variable)
    plt.show()
```

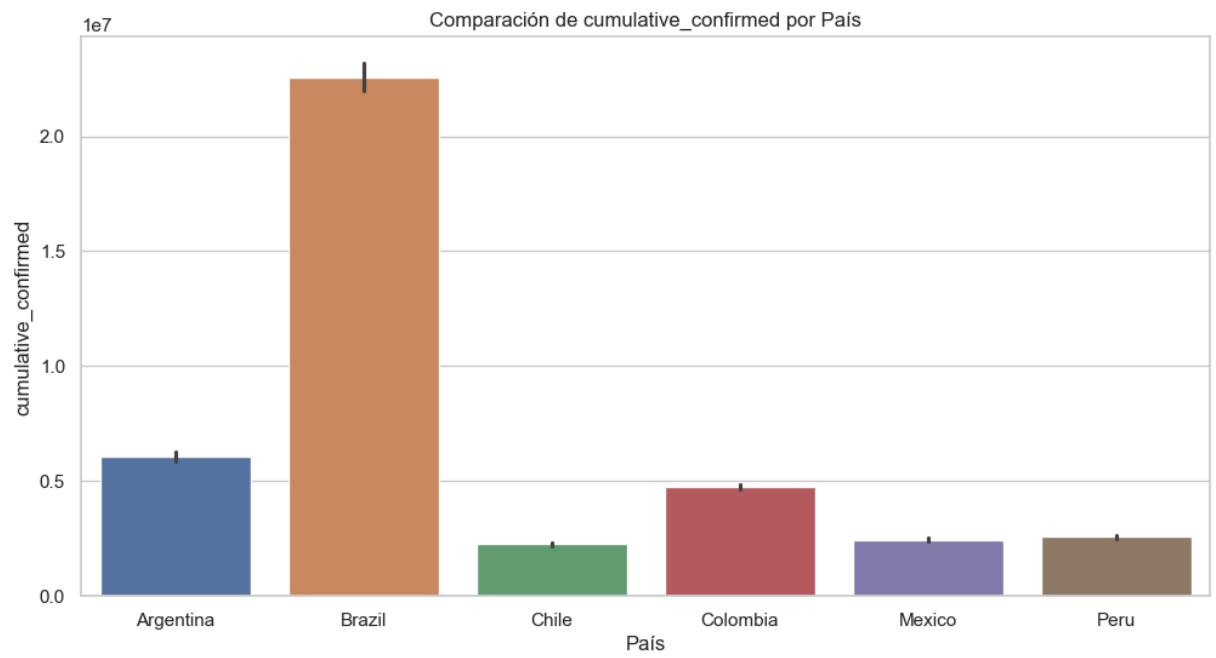
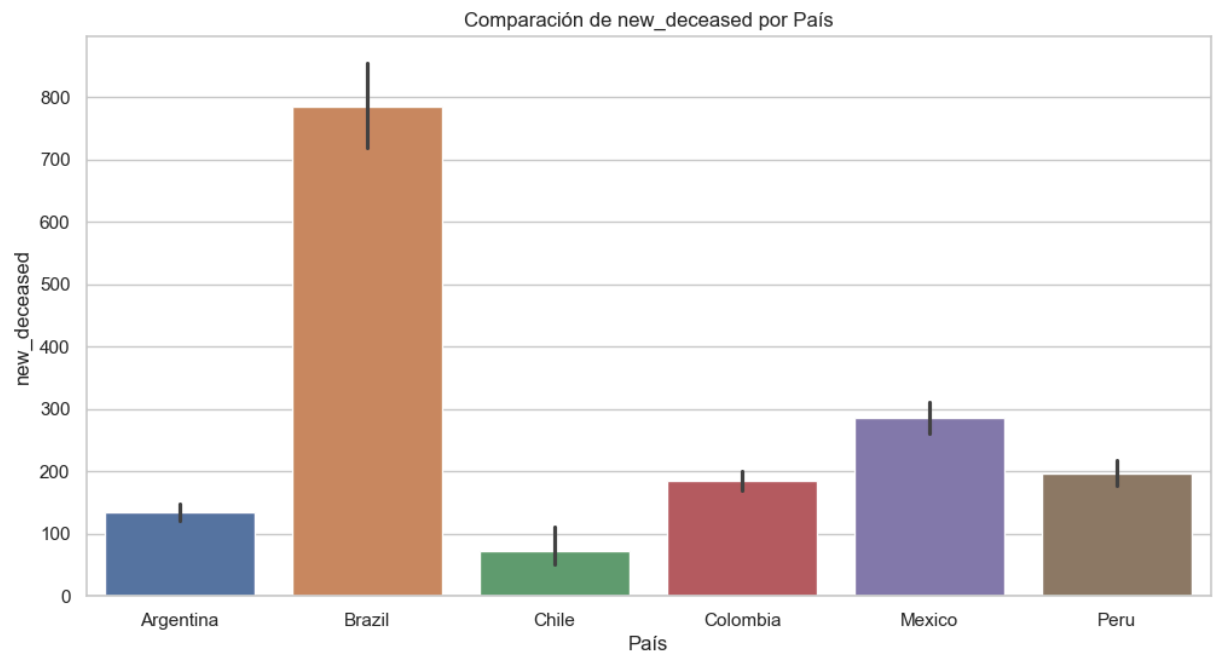


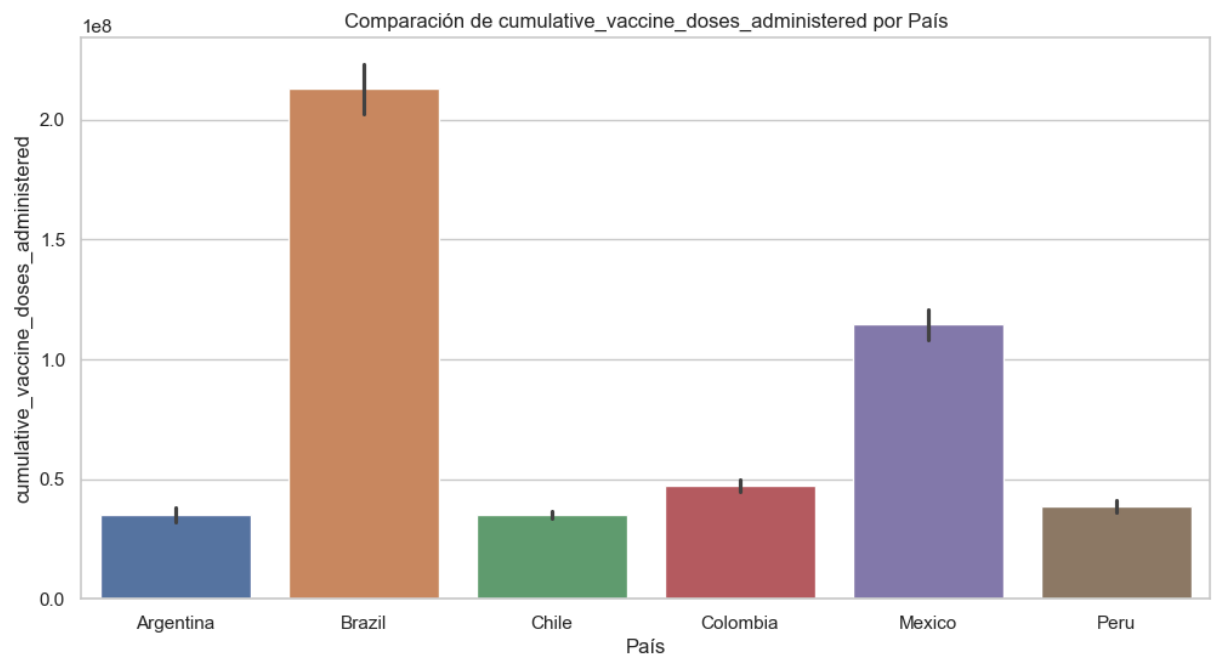
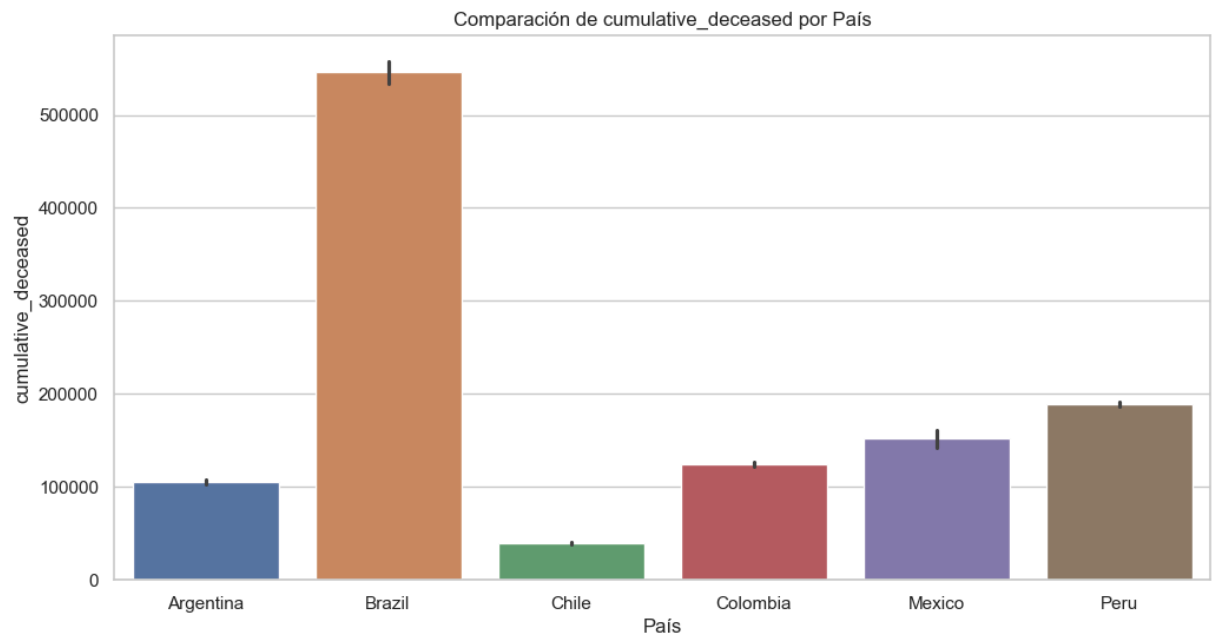


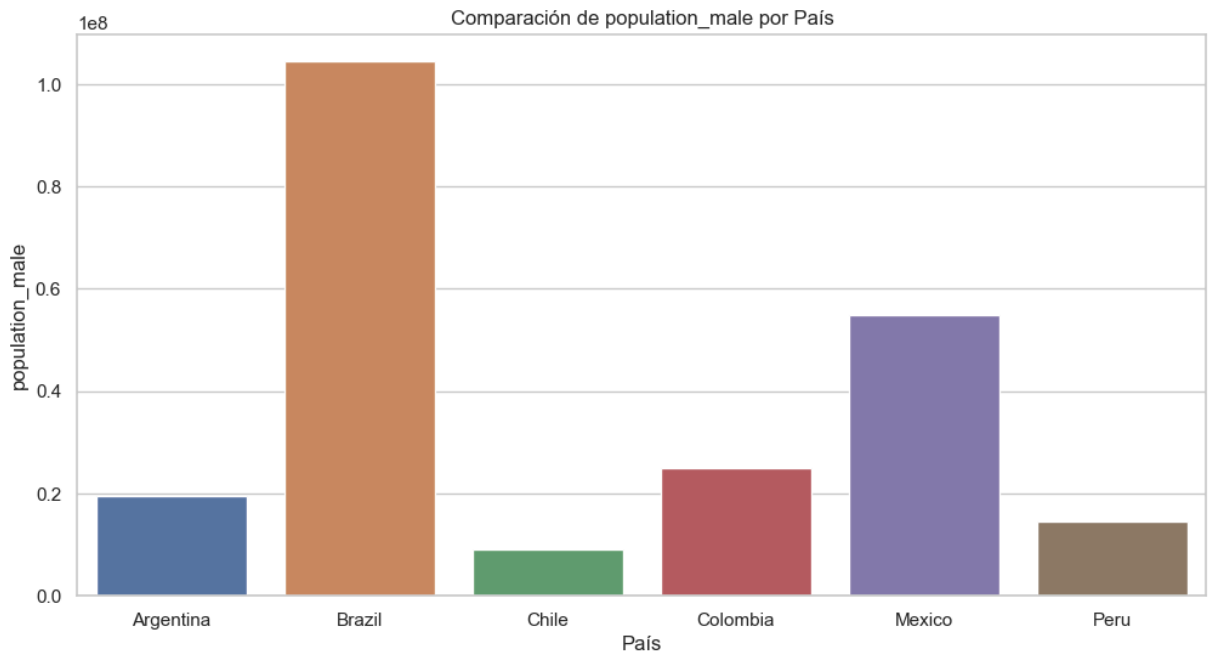
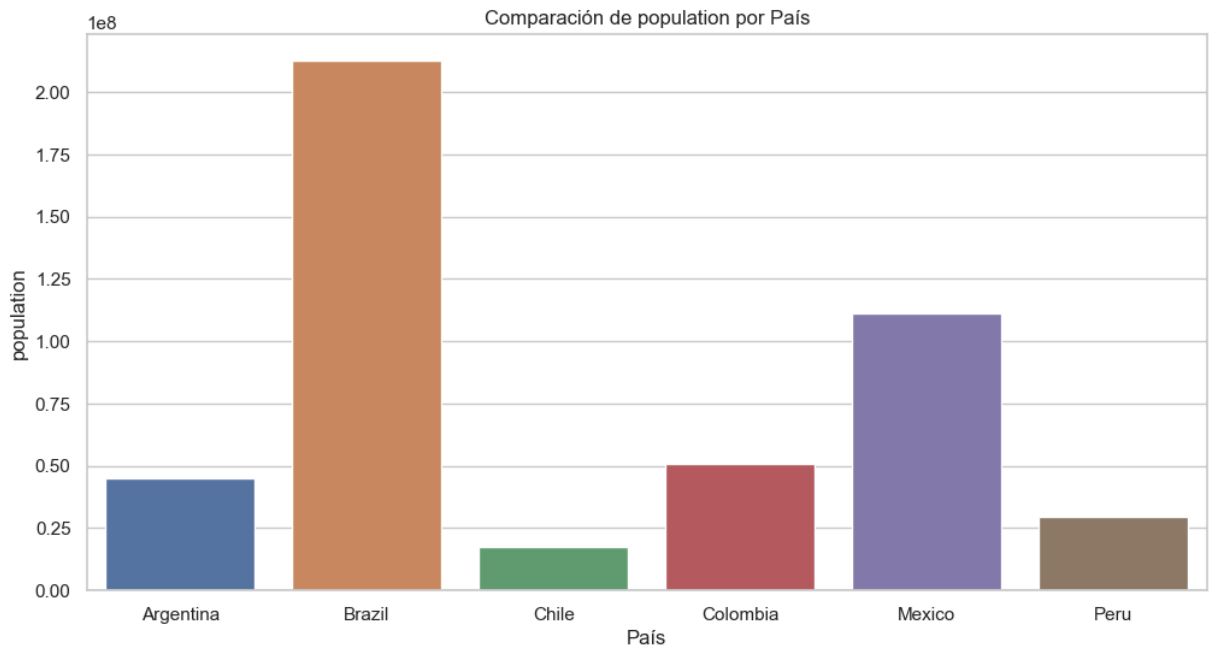
```
In [109...] variables_numericas = datos_paises_interes.select_dtypes(include=[np.number]).columns
```

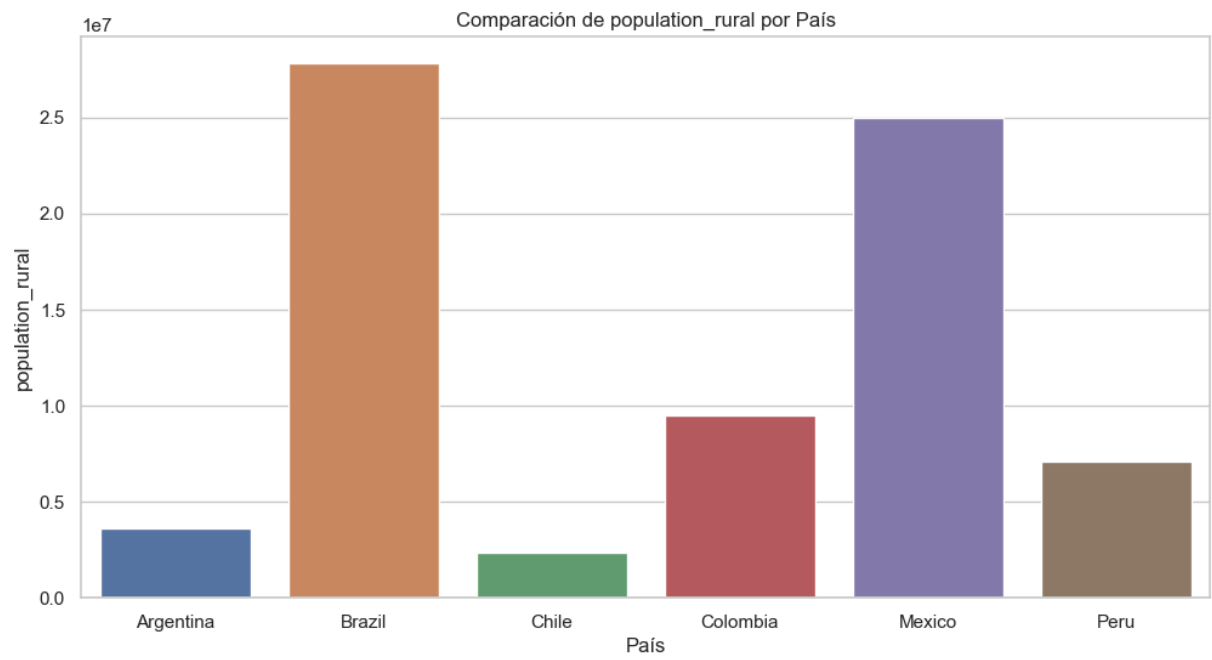
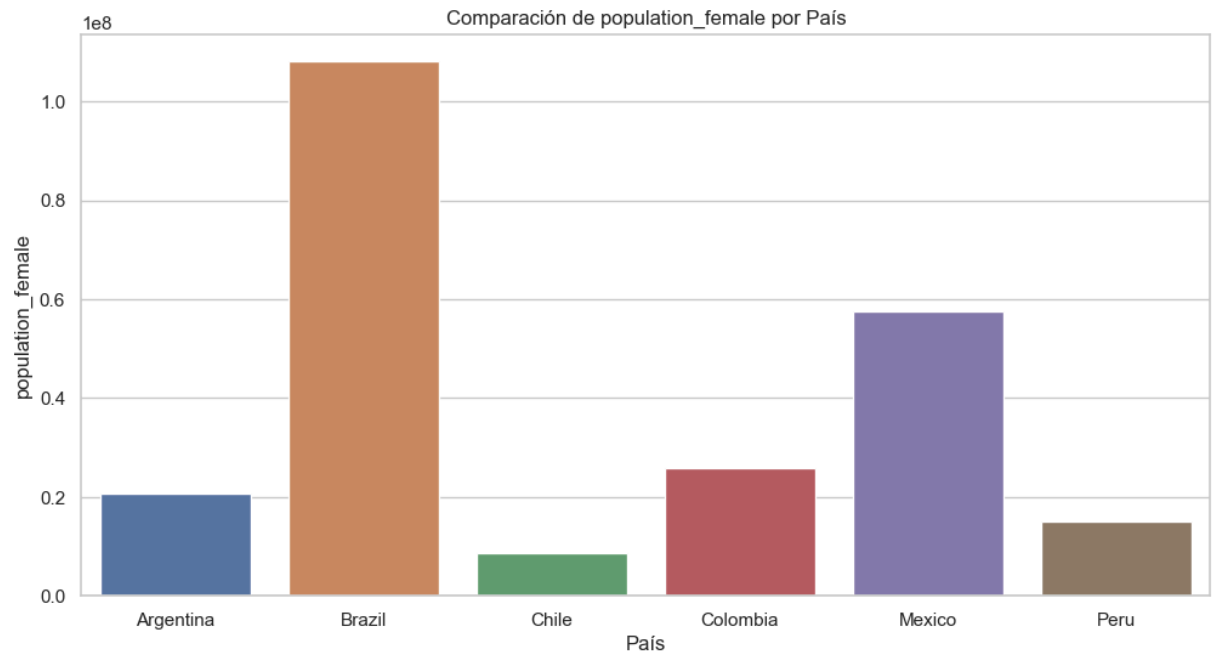
```
In [112...] for variable in variables_numericas:
    plt.figure(figsize=(12, 6))
    sns.barplot(data= datos_paises_interes, x='country_name', y=variable, hue='country_name')
    plt.title(f'Comparación de {variable} por País')
    plt.xlabel('País')
    plt.ylabel(variable)
    plt.show()
```

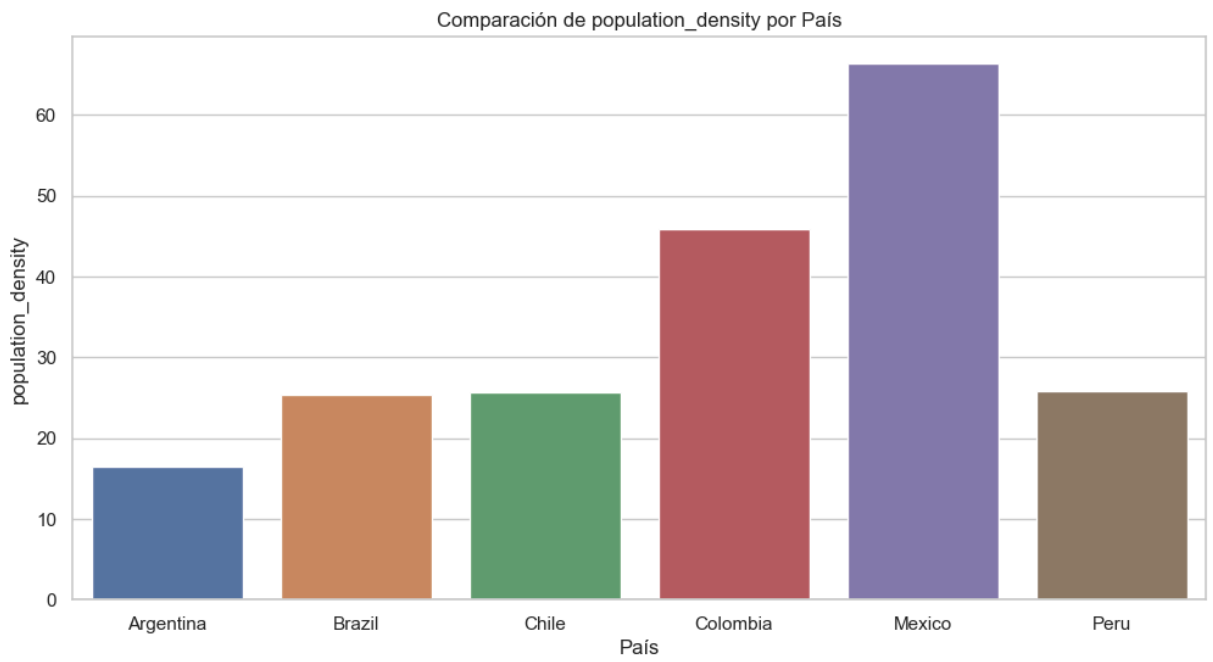
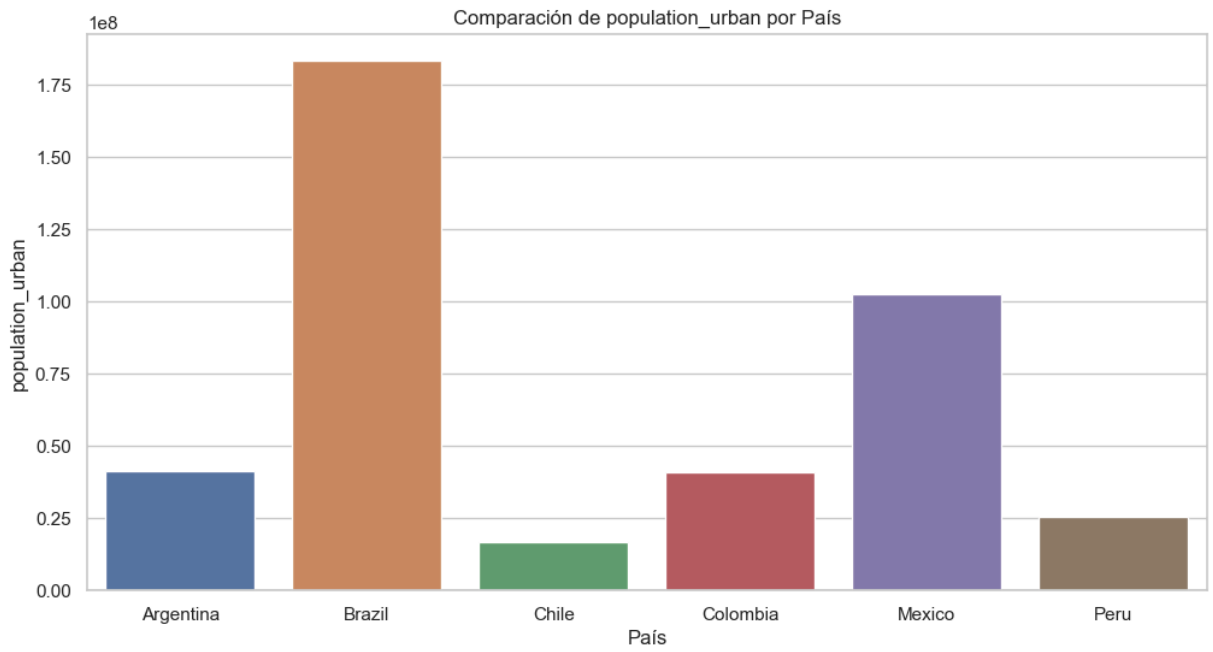


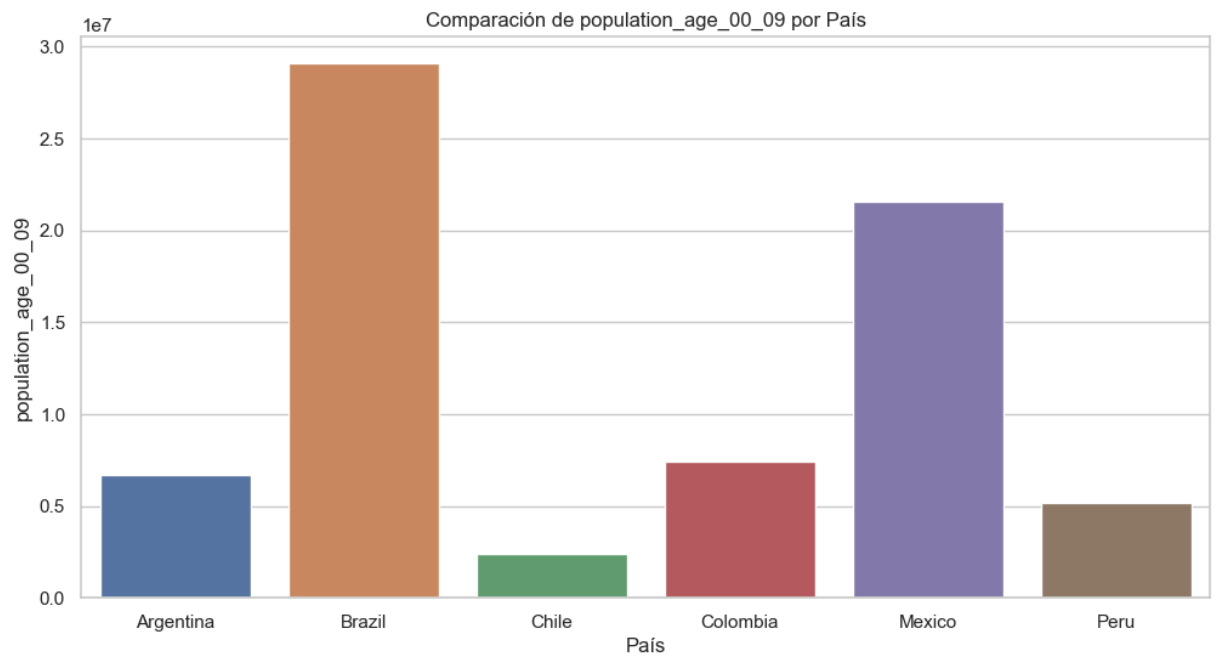
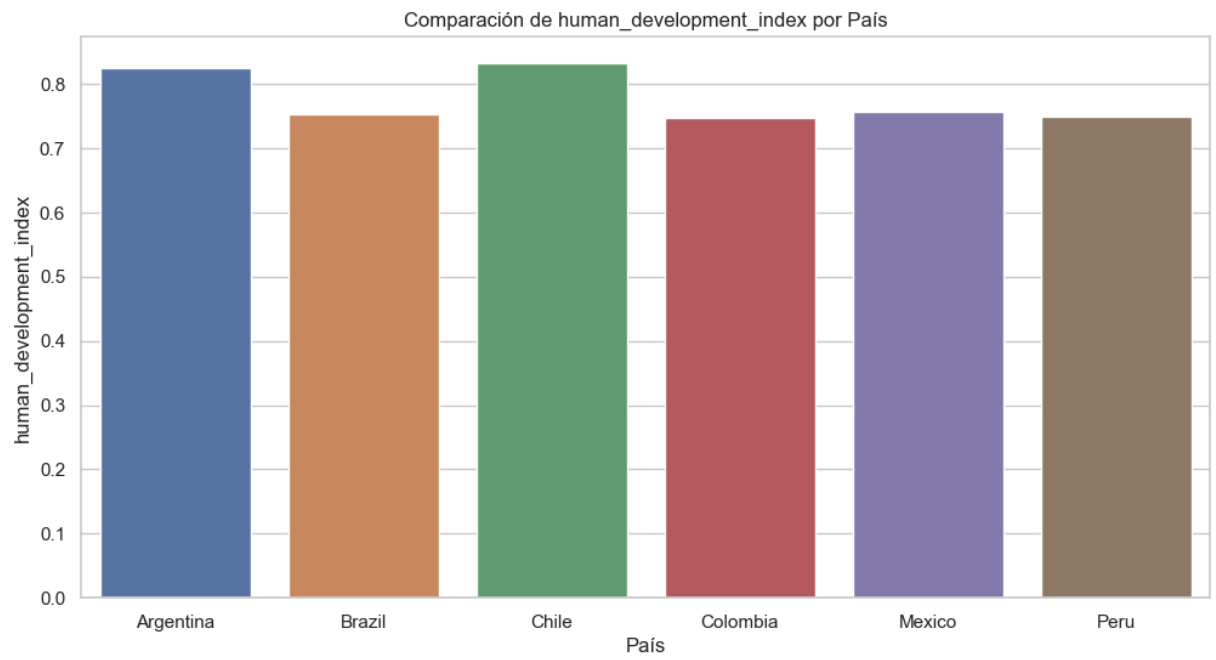


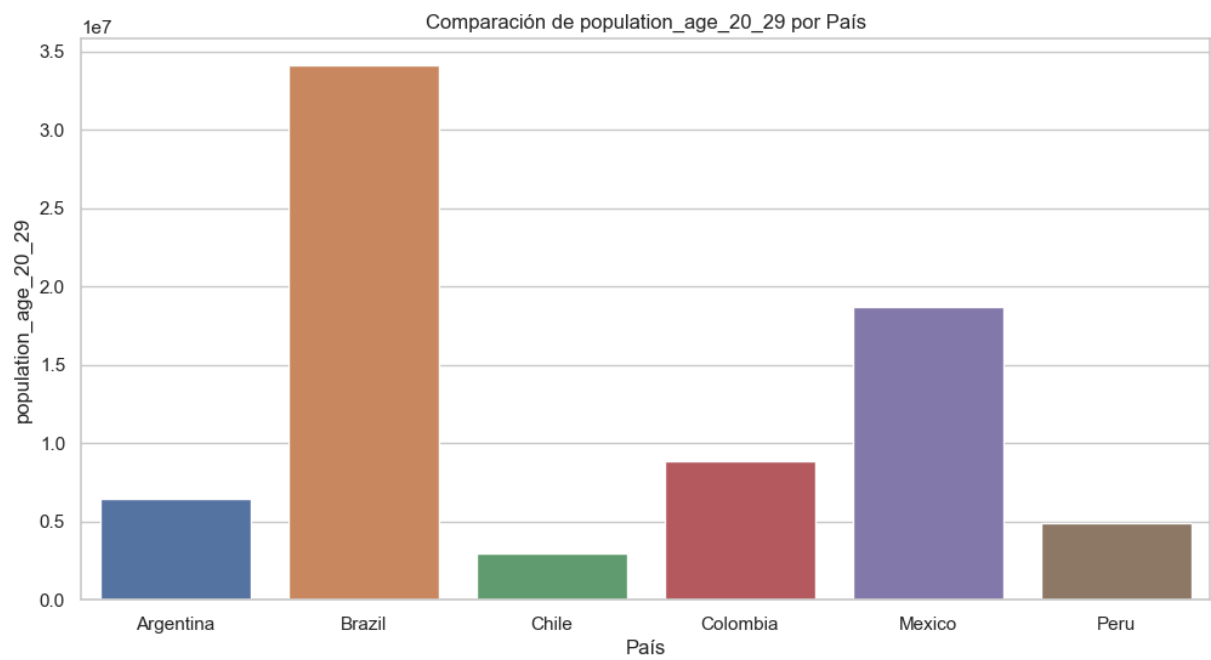
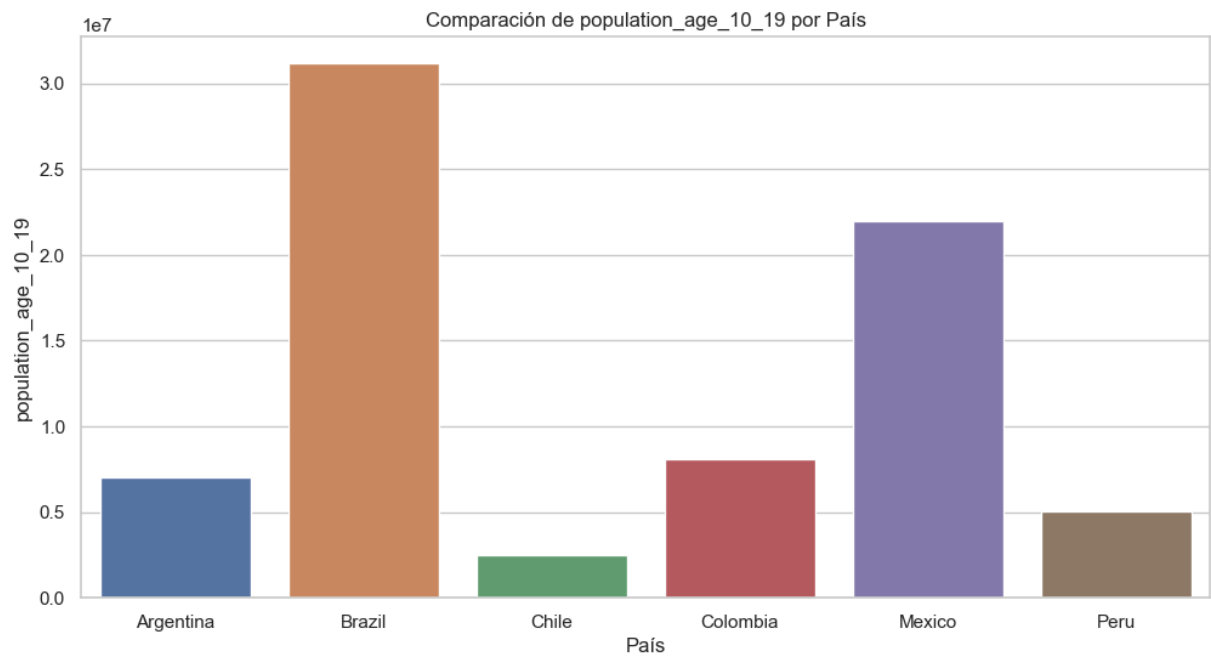


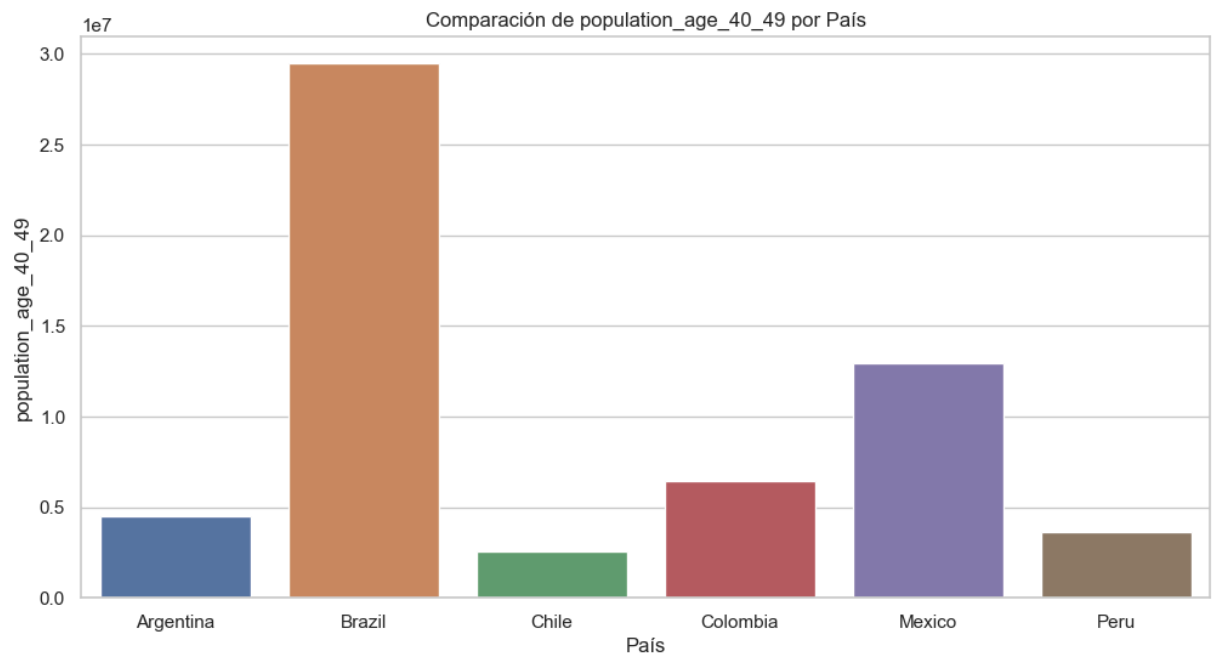
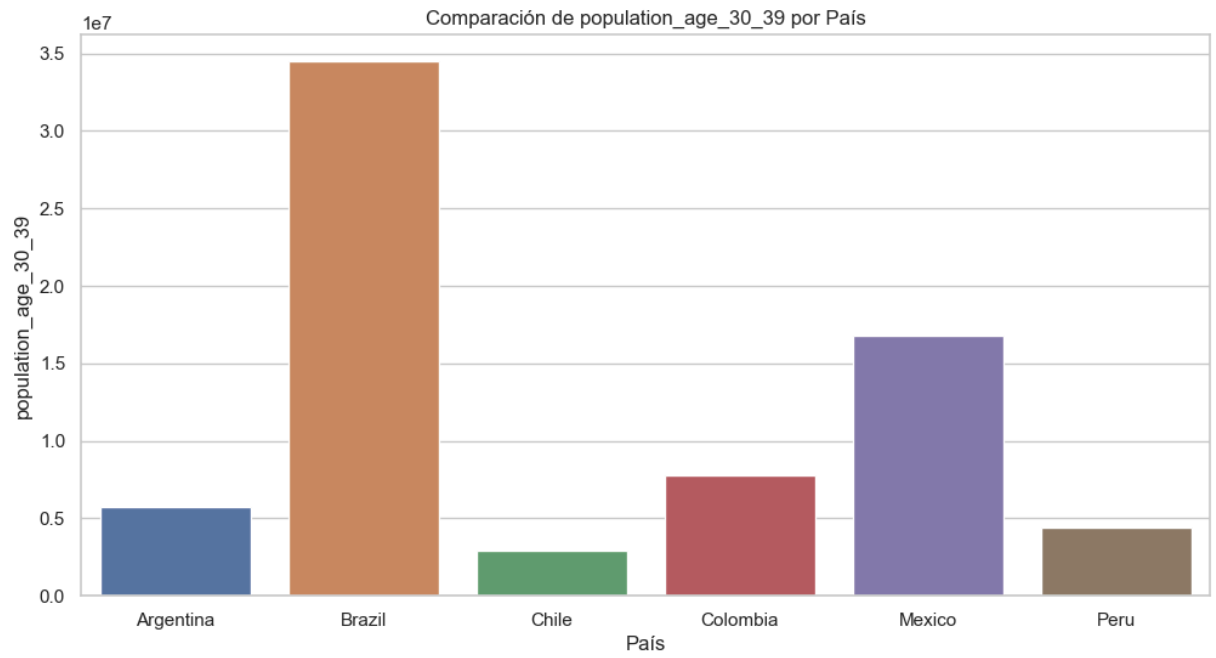


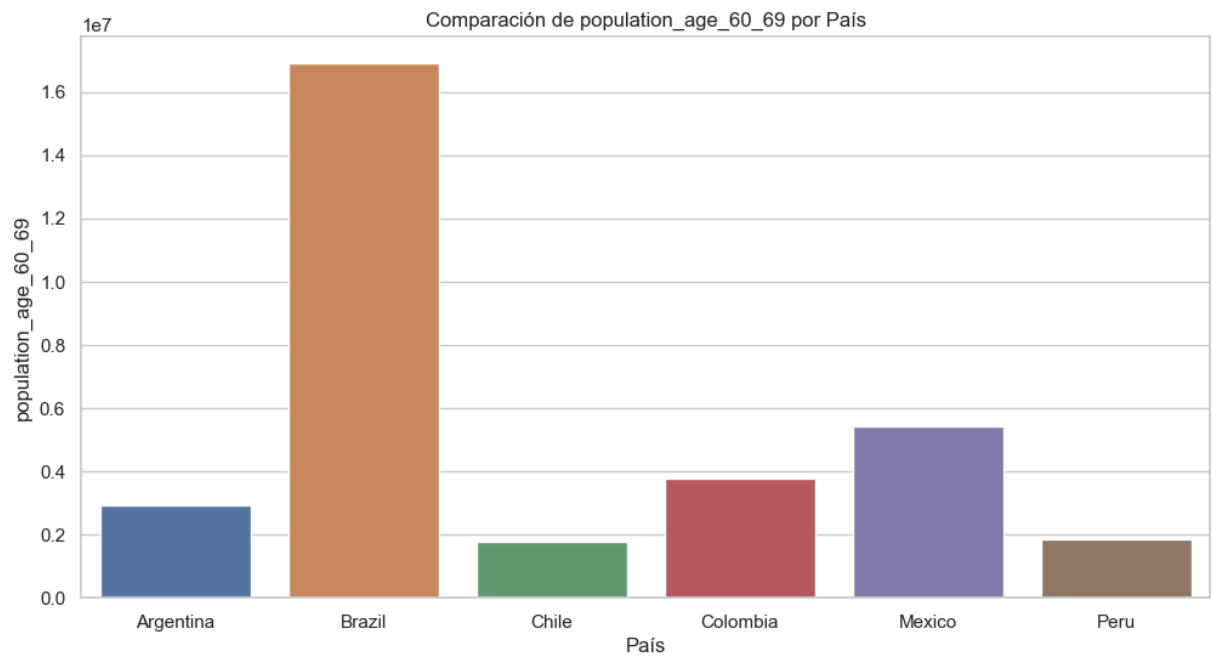
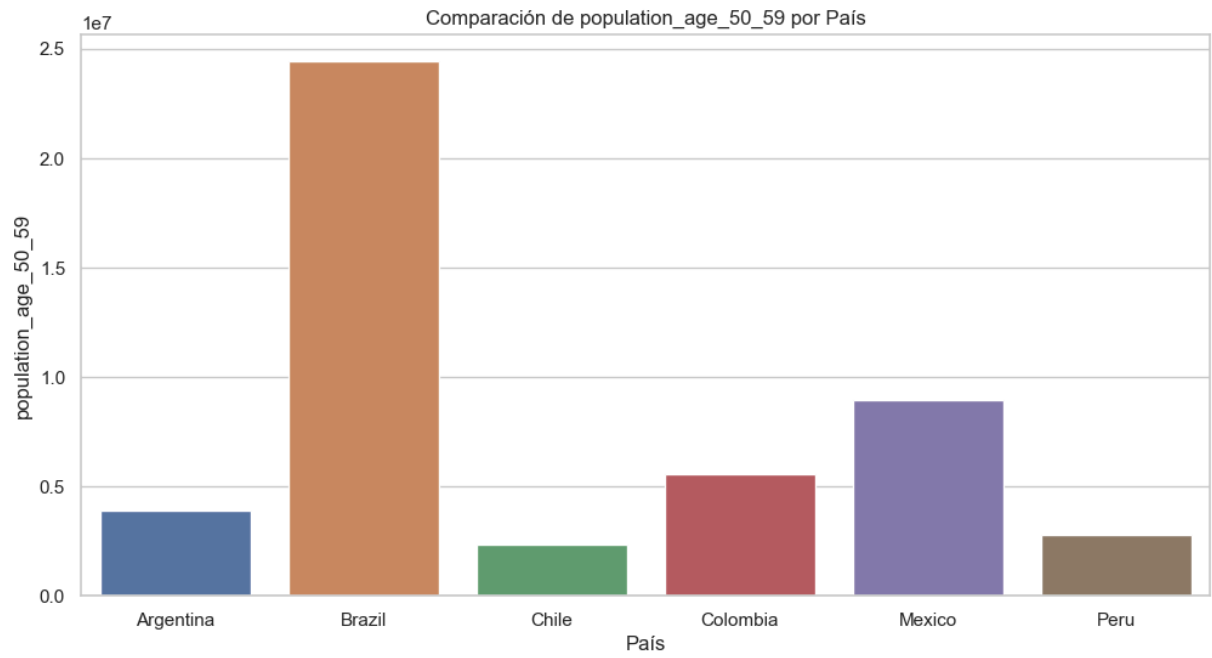


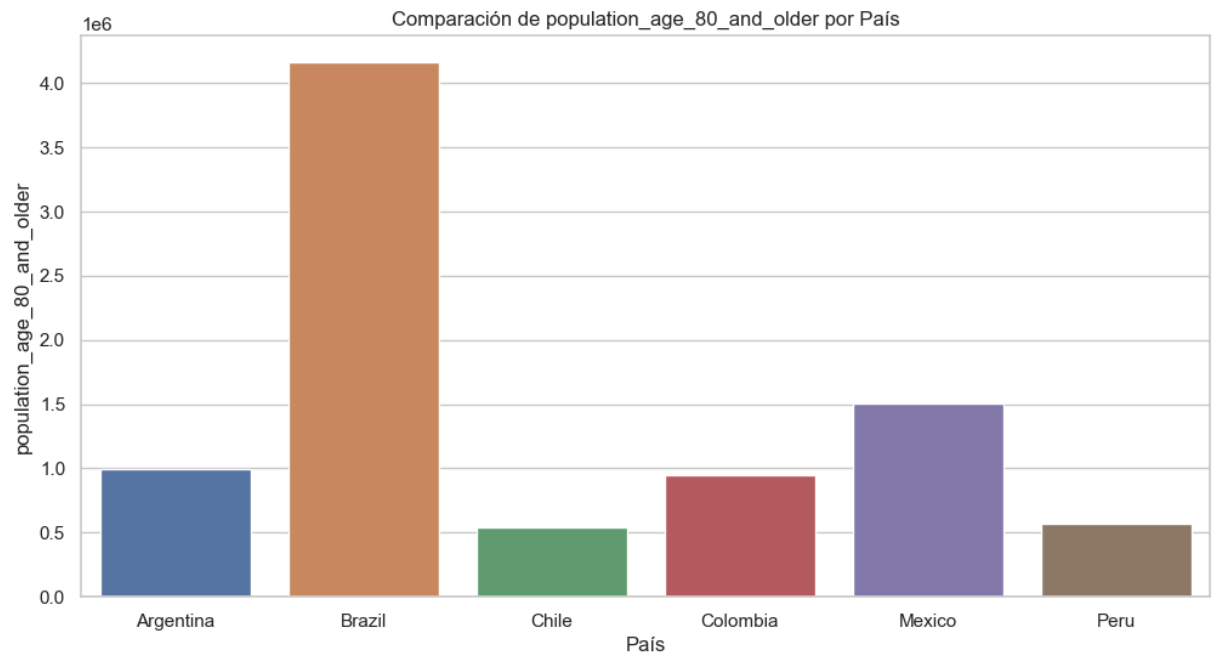
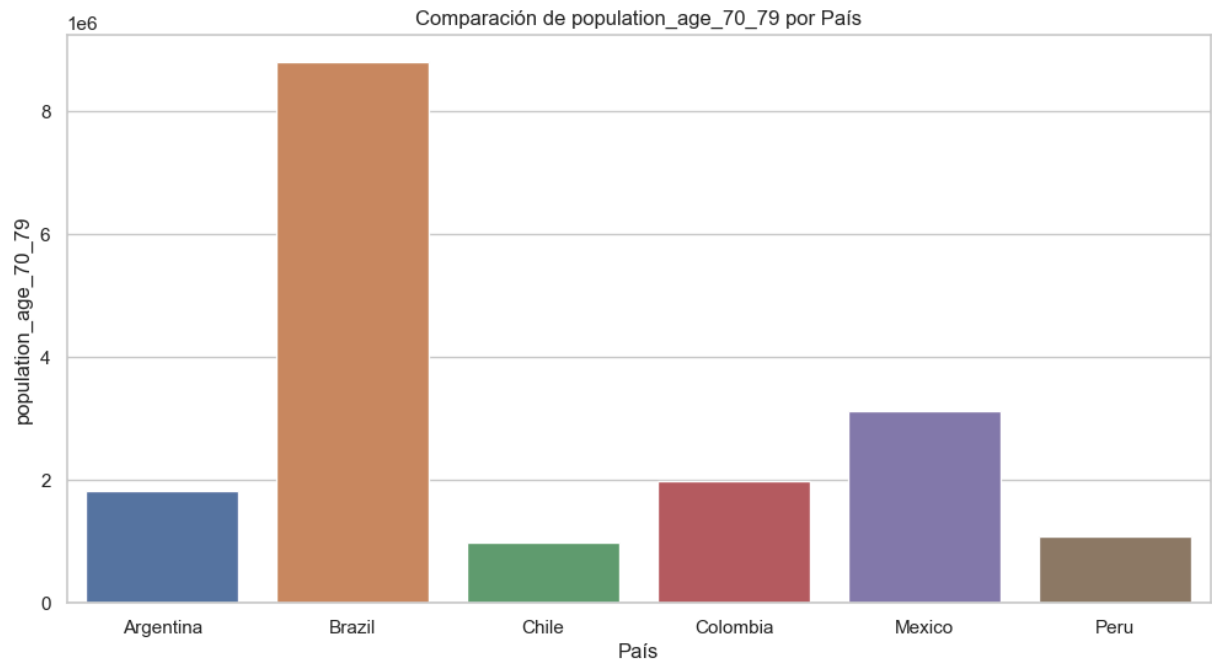


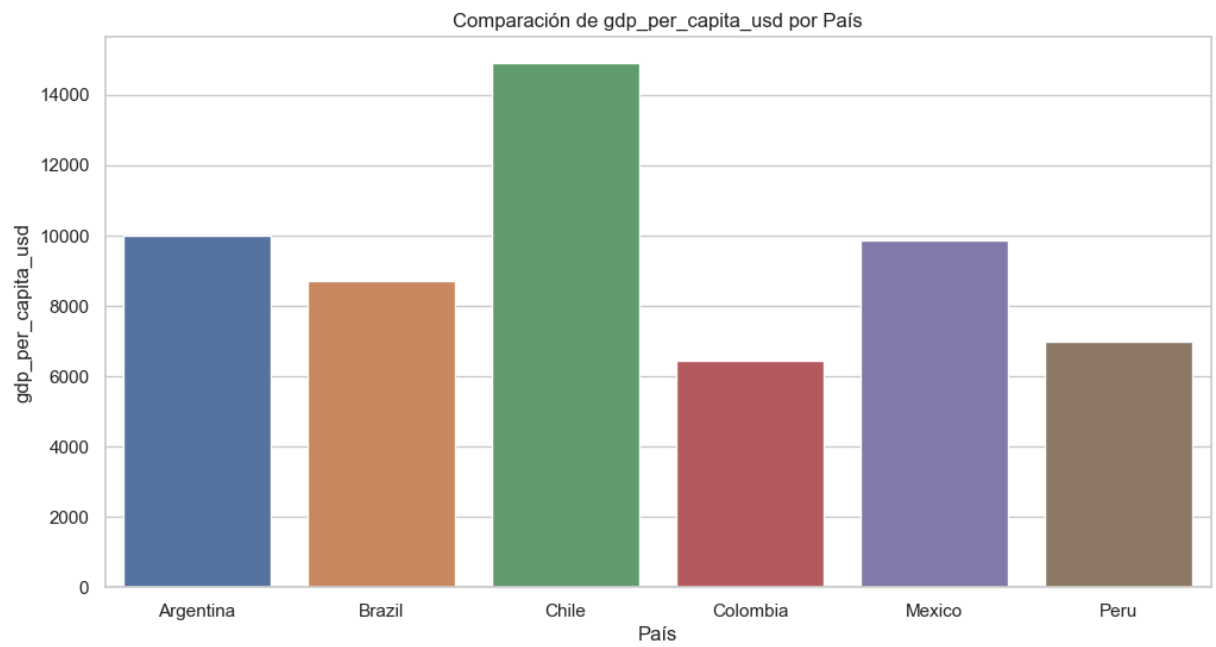
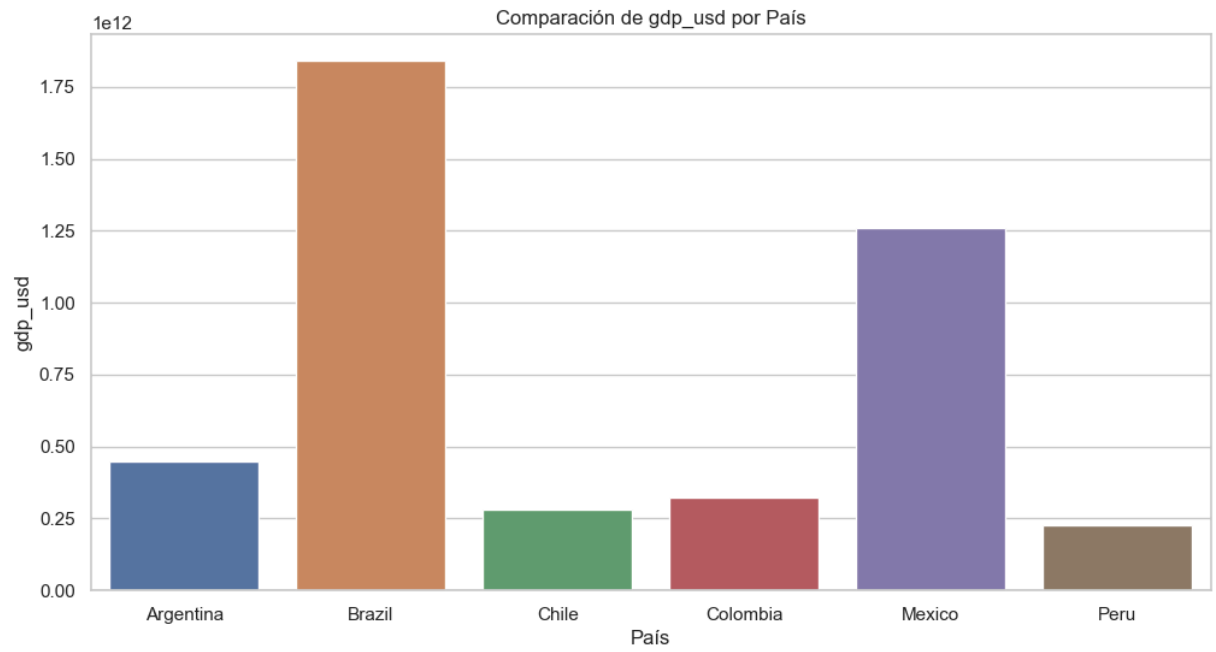


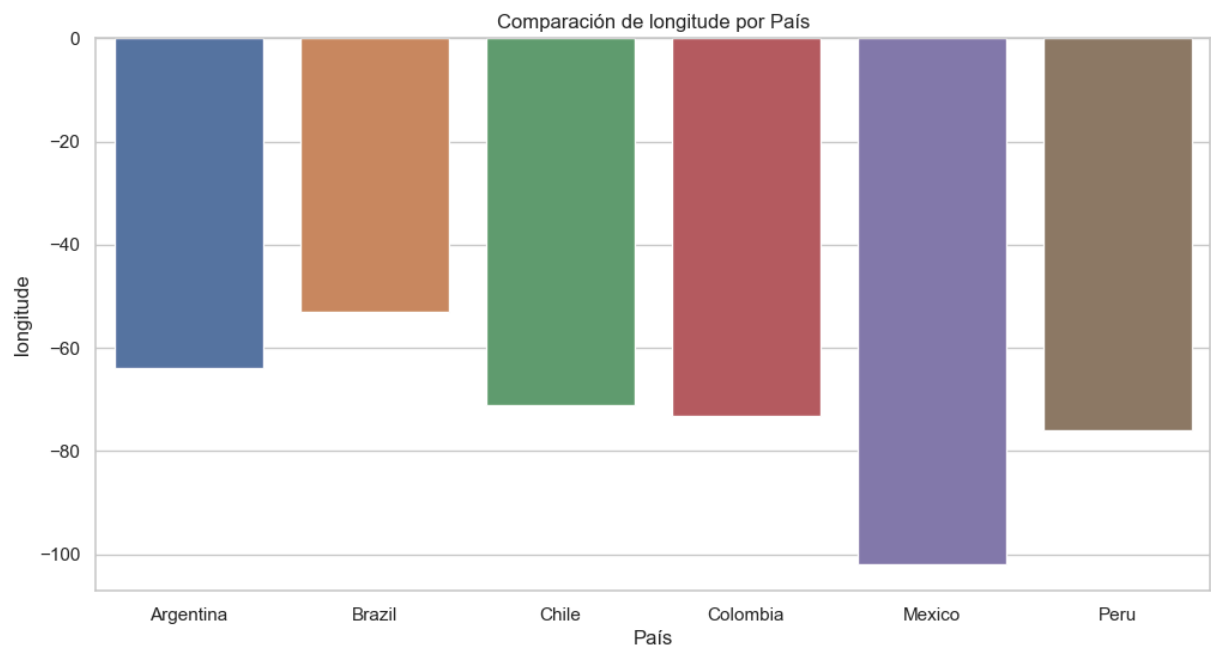
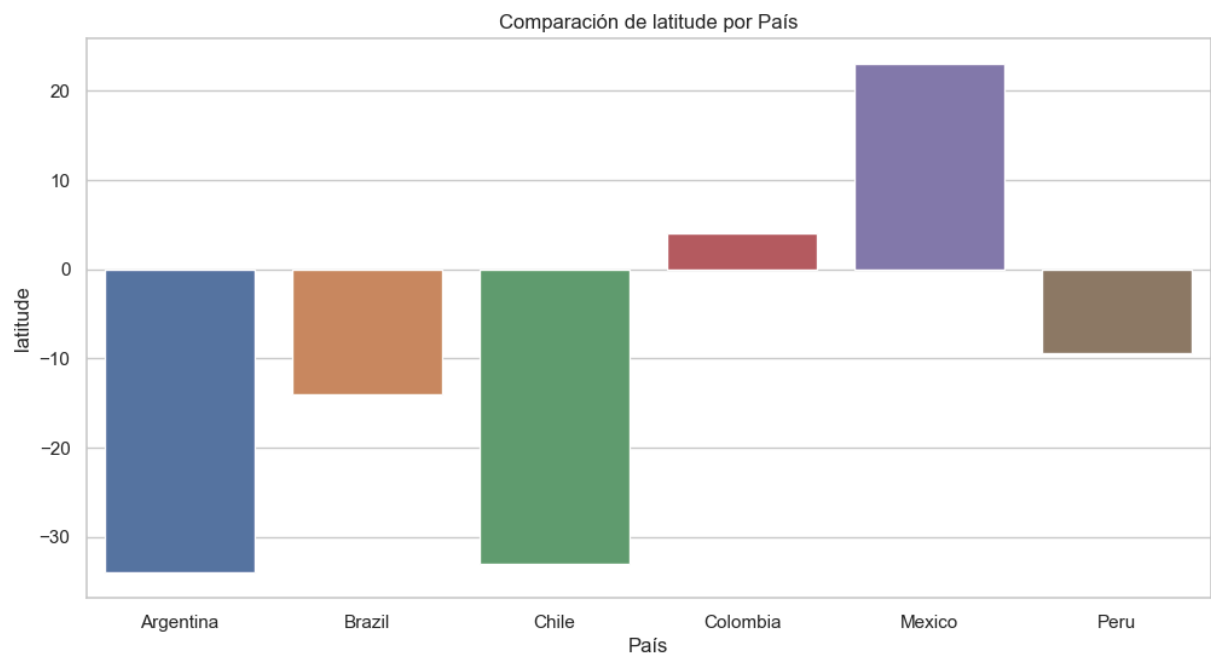


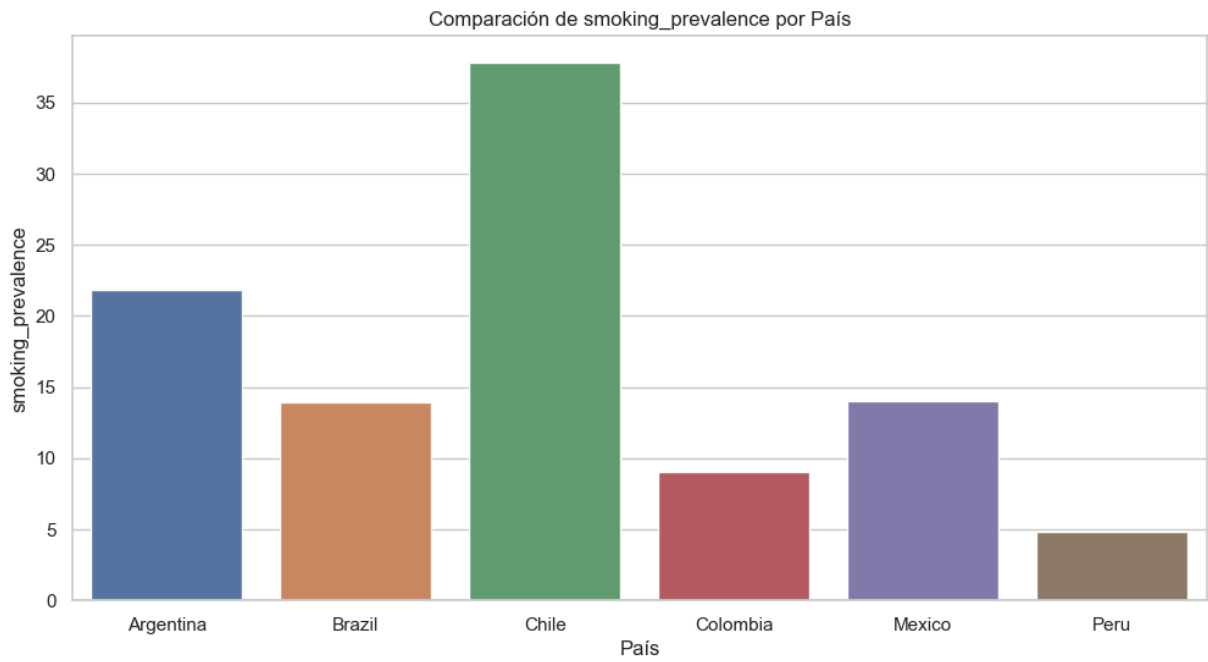
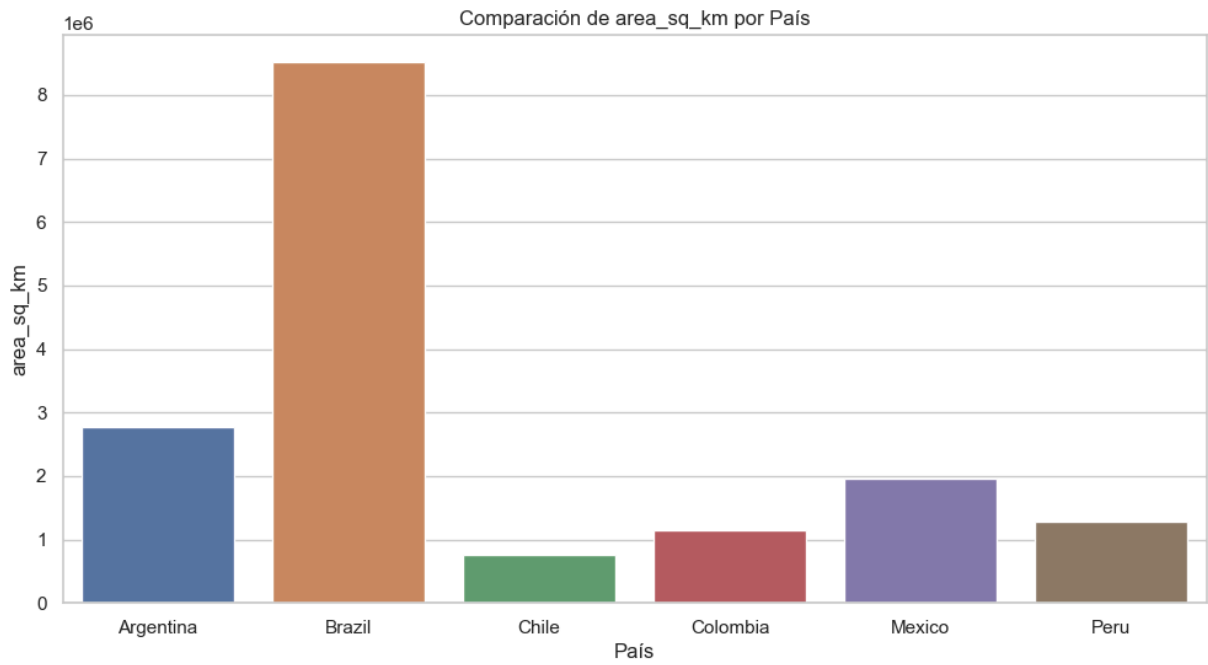


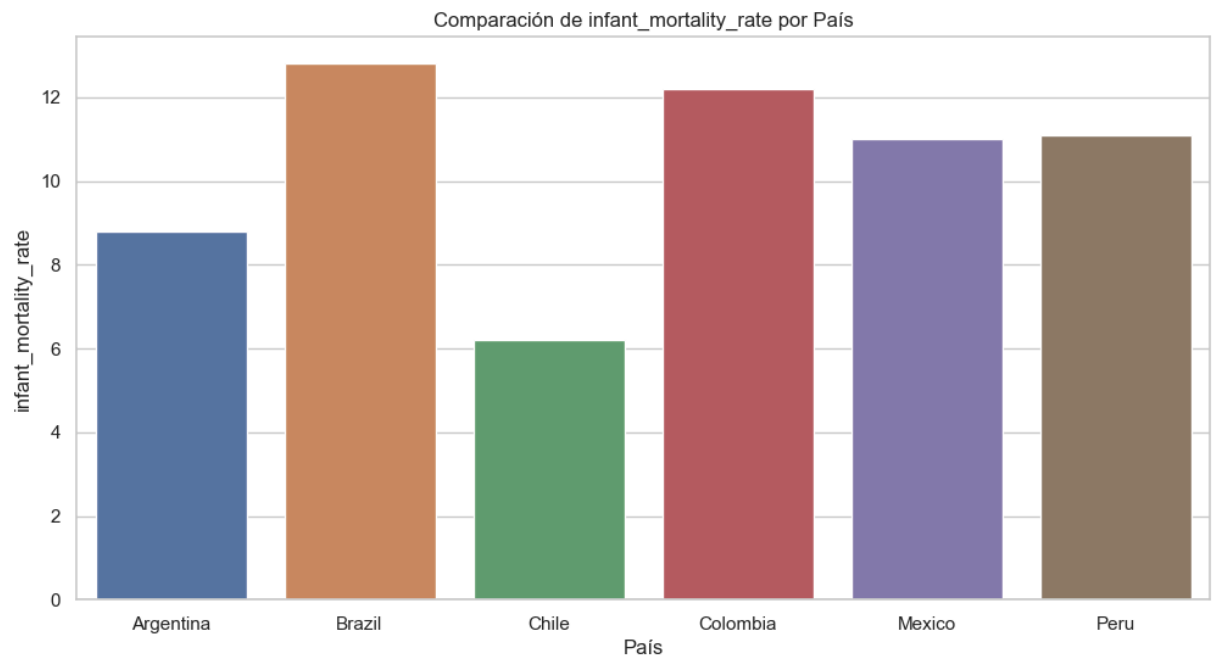
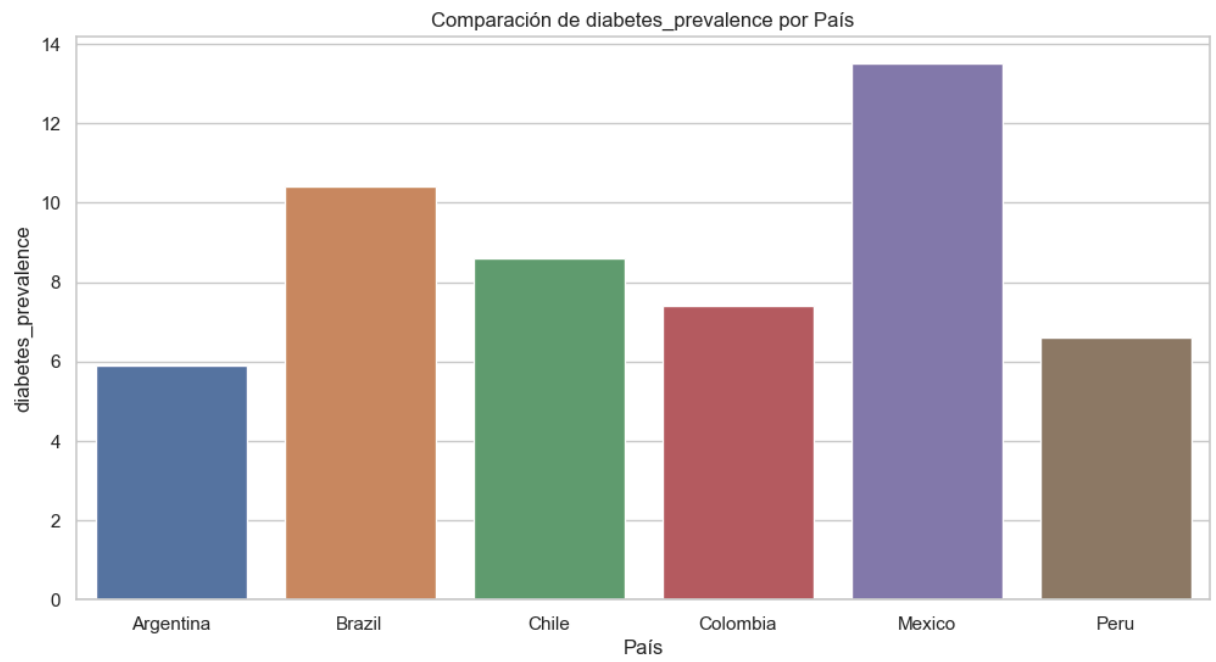


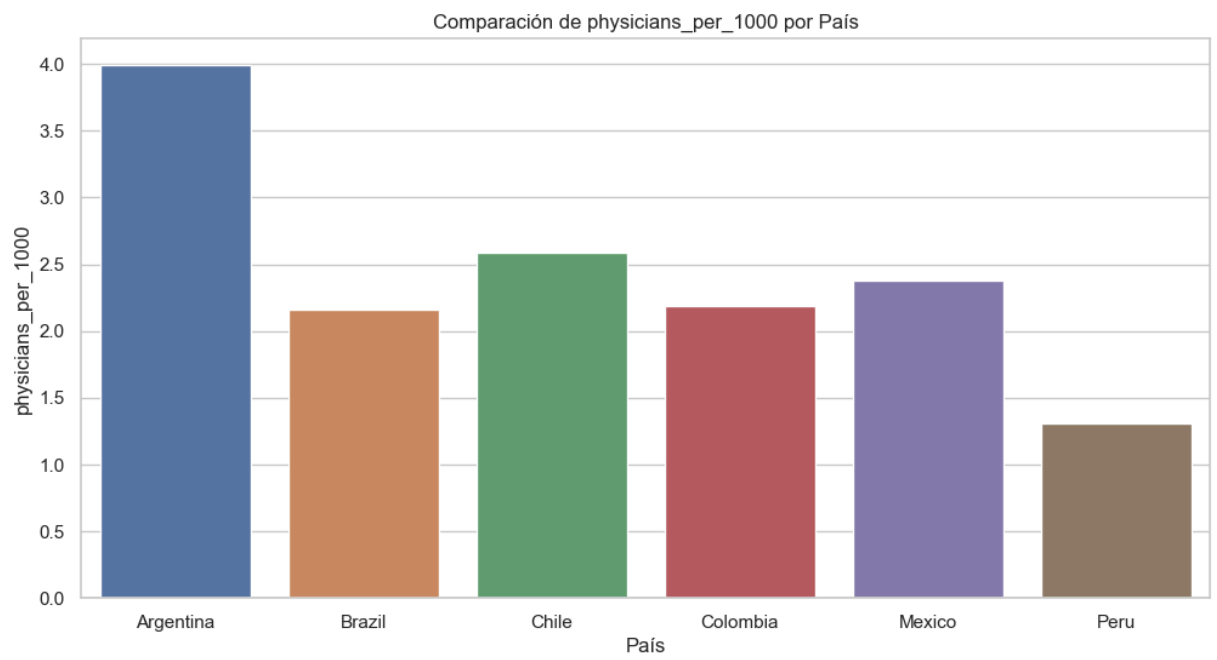
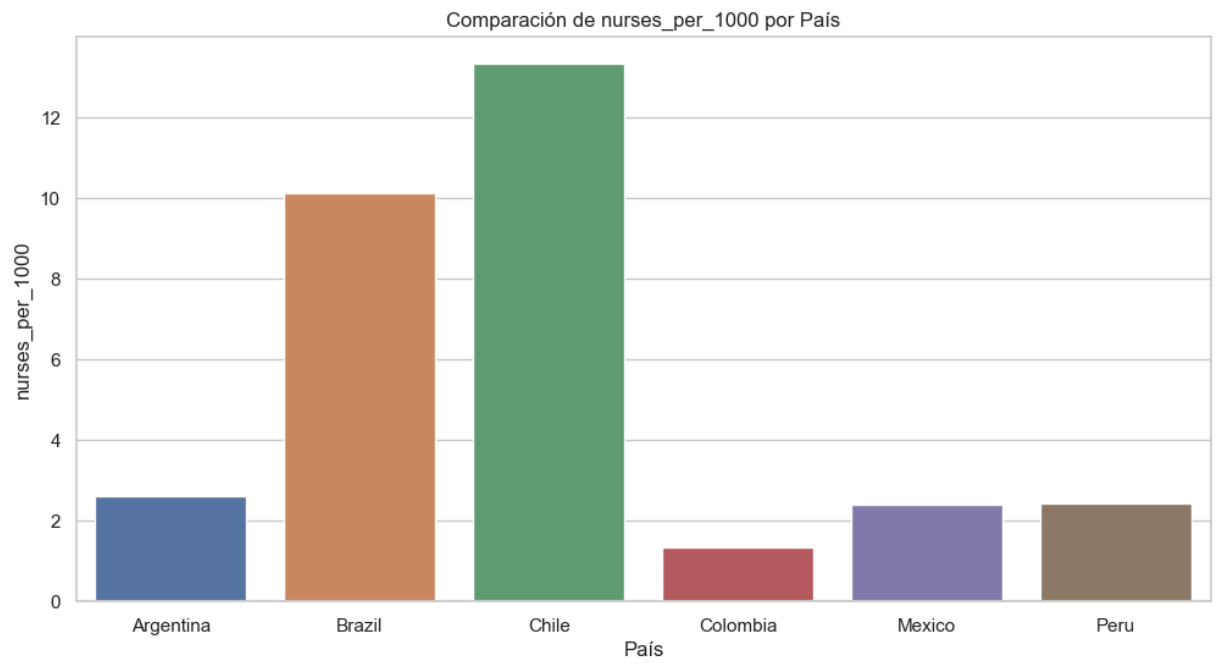


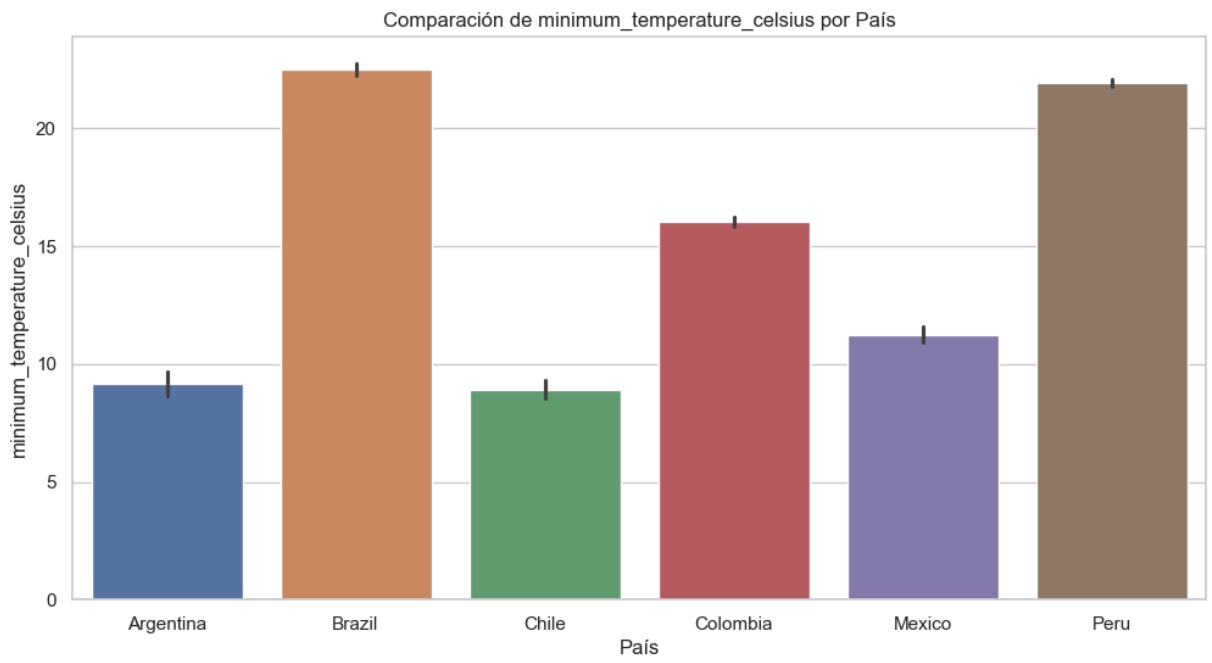
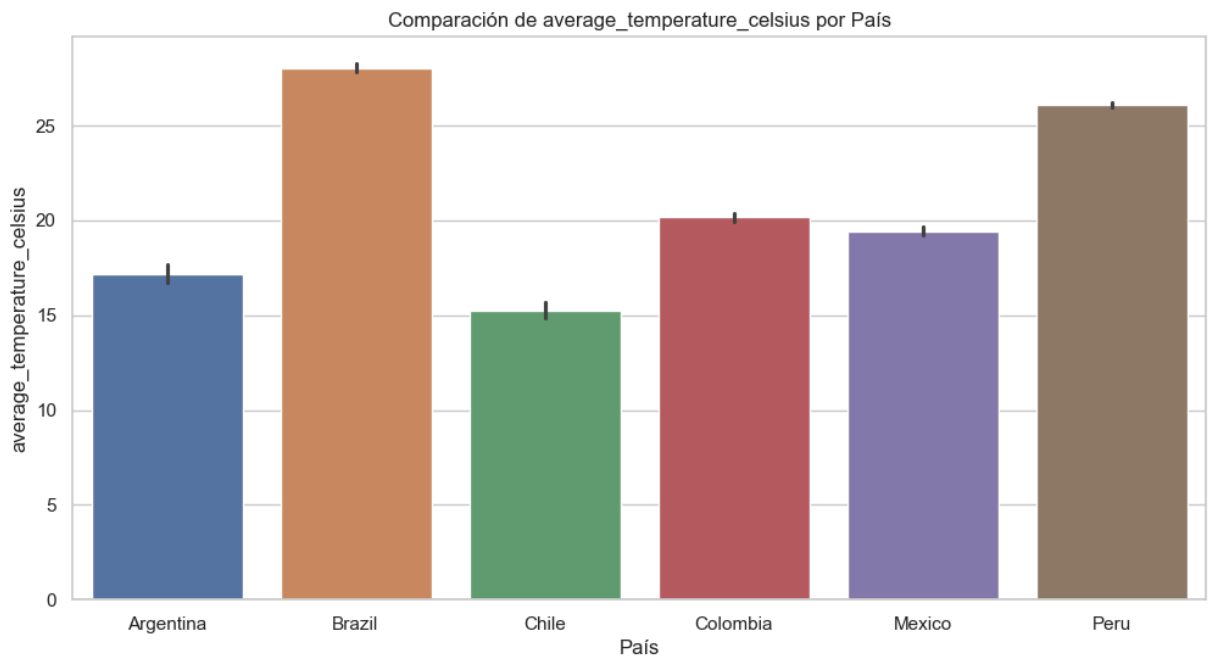


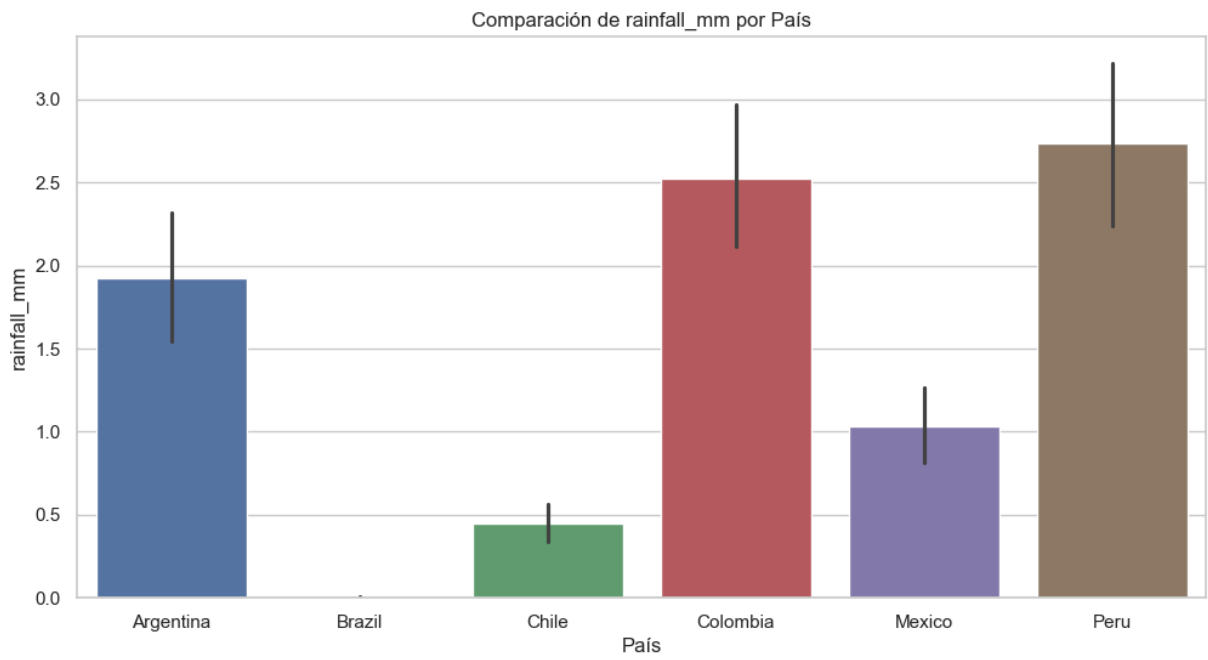
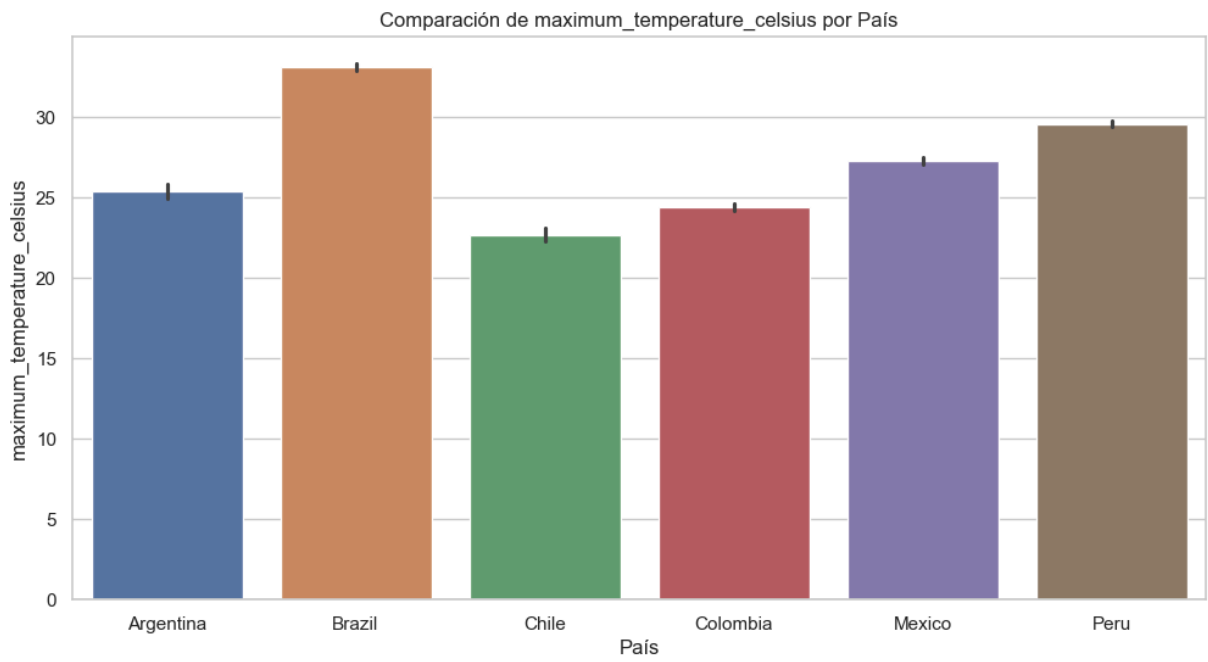


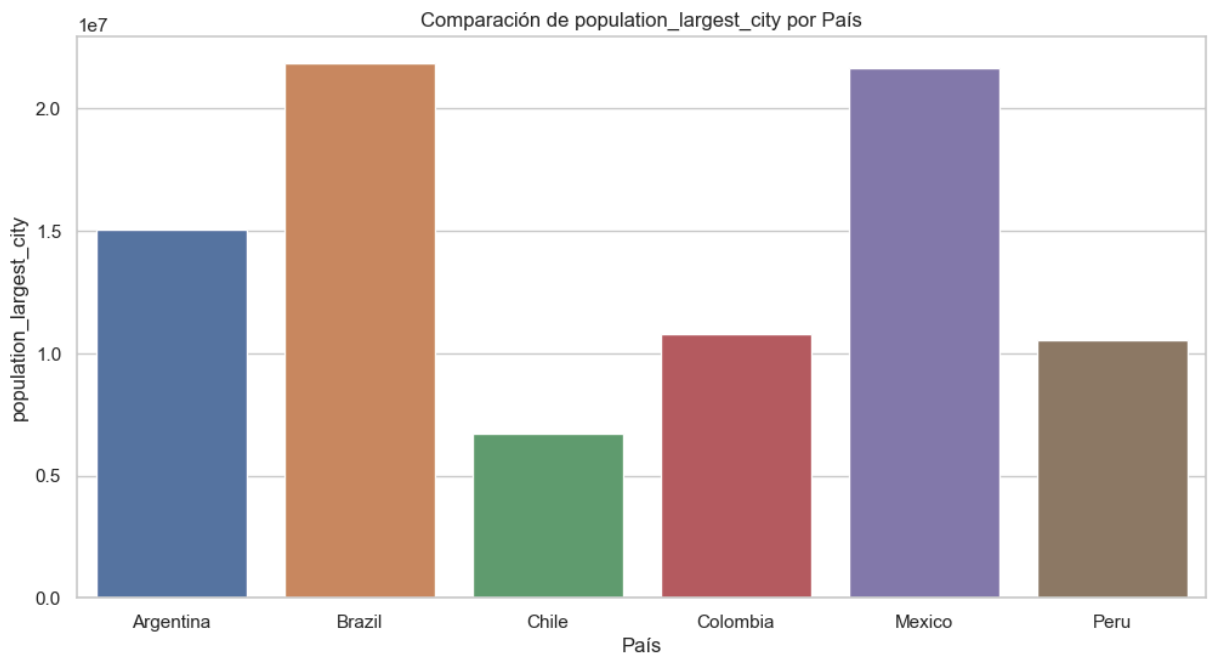
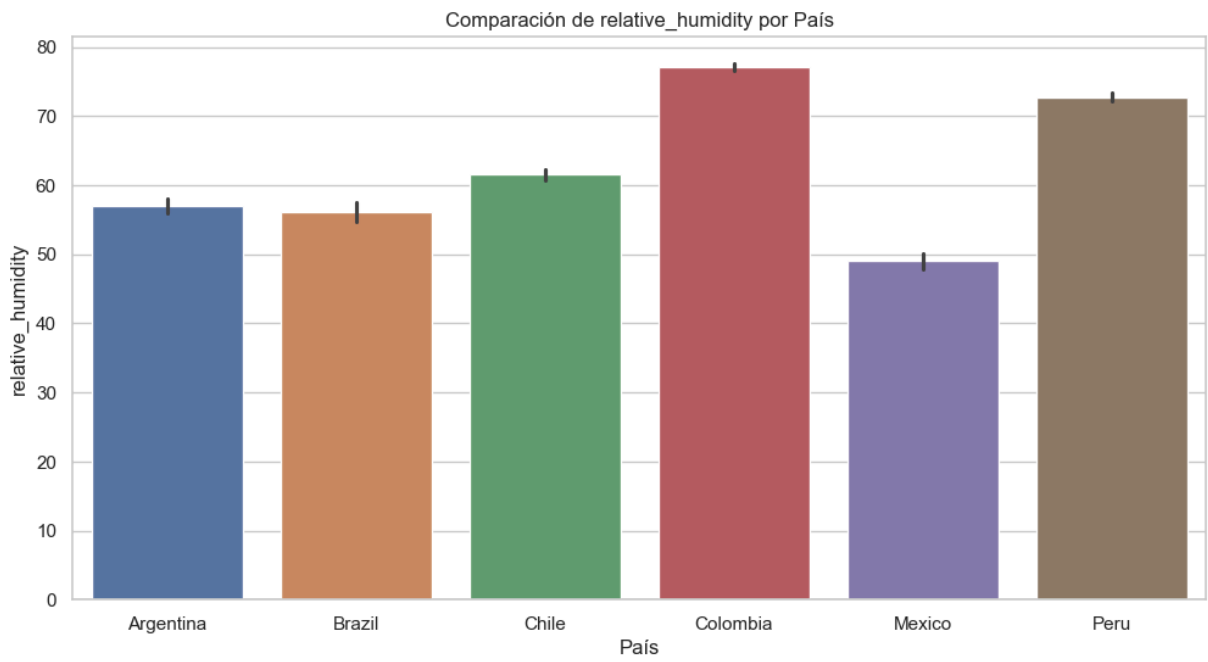


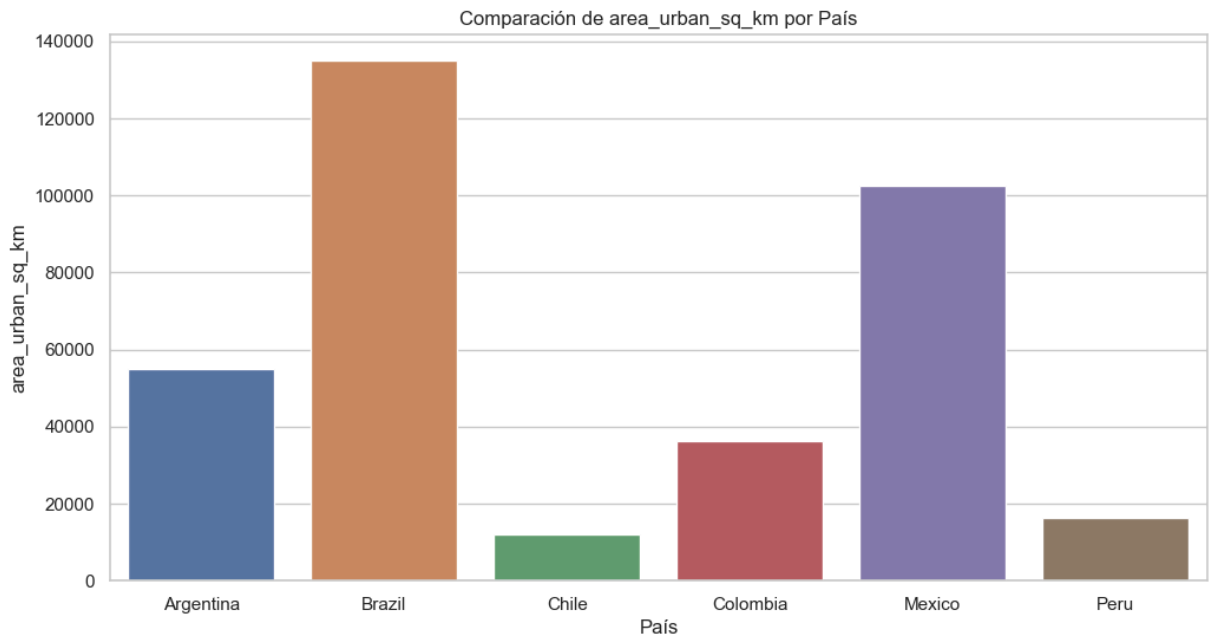
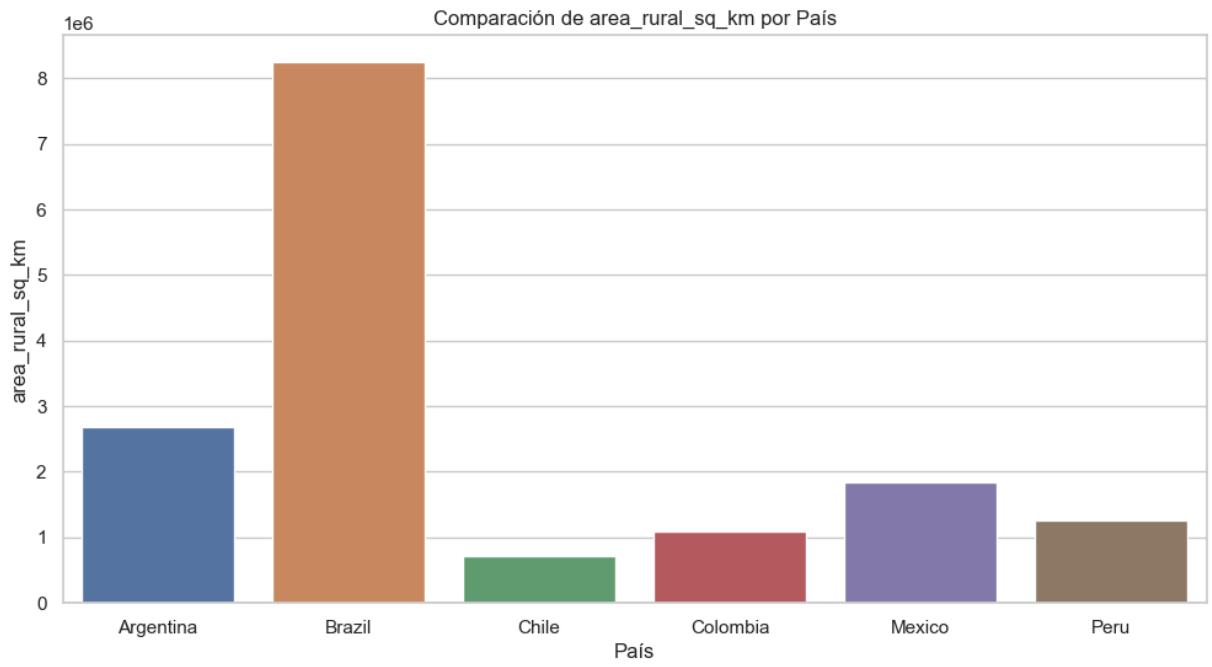


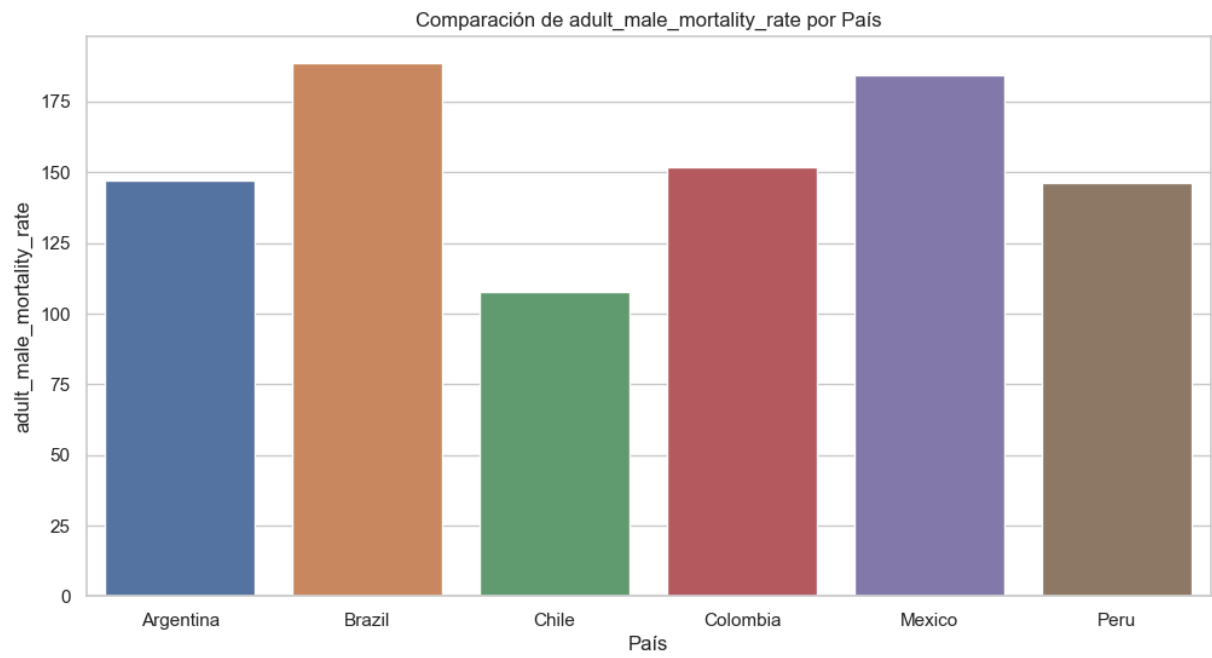
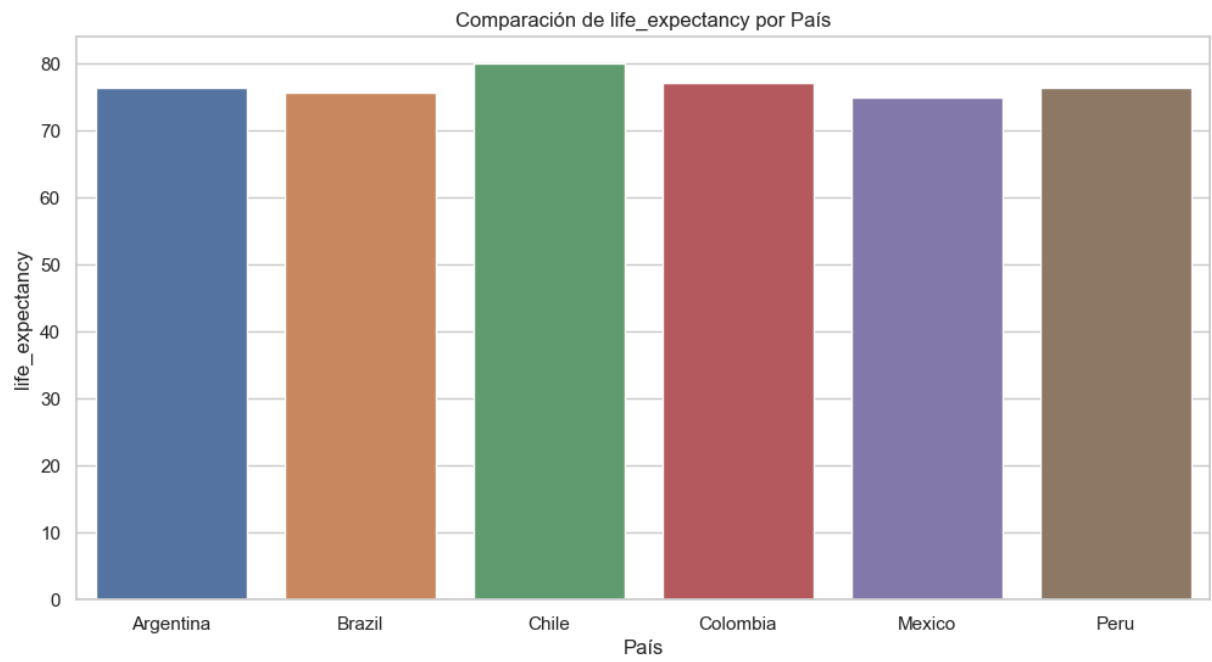


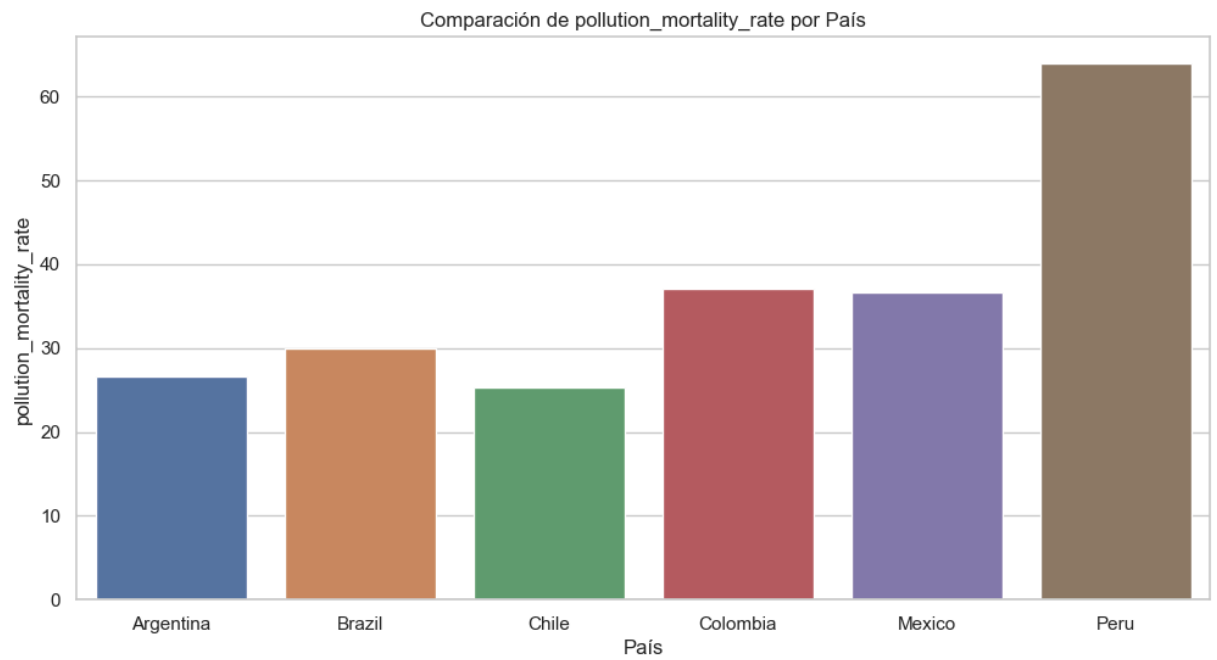
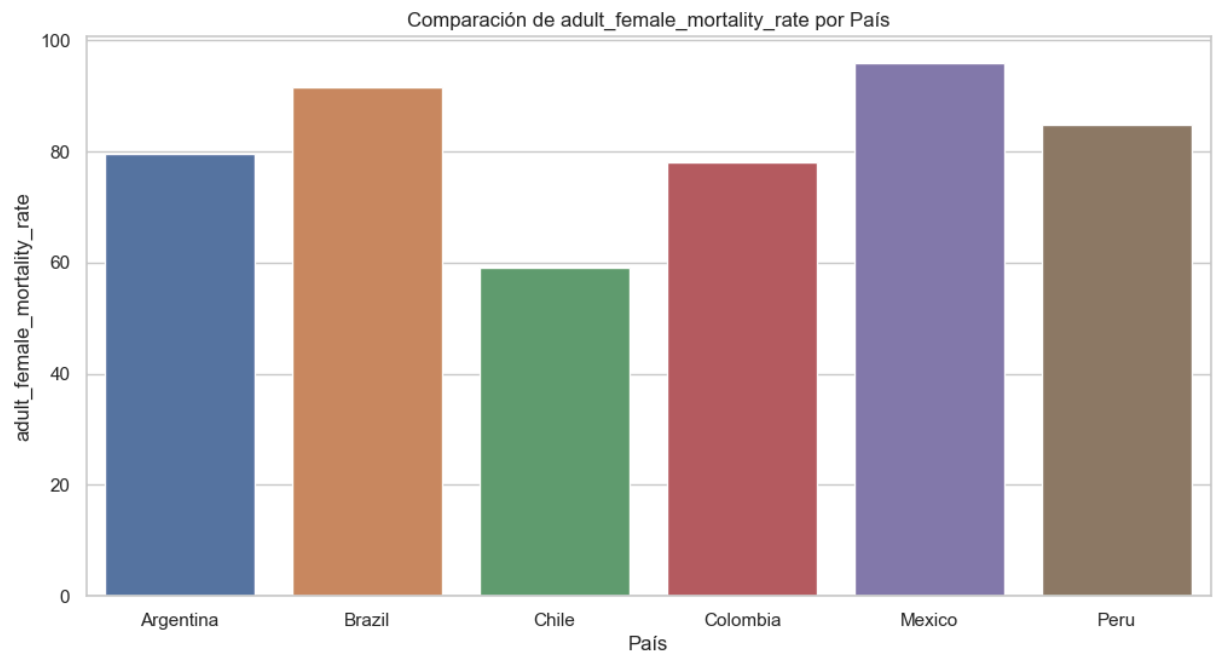


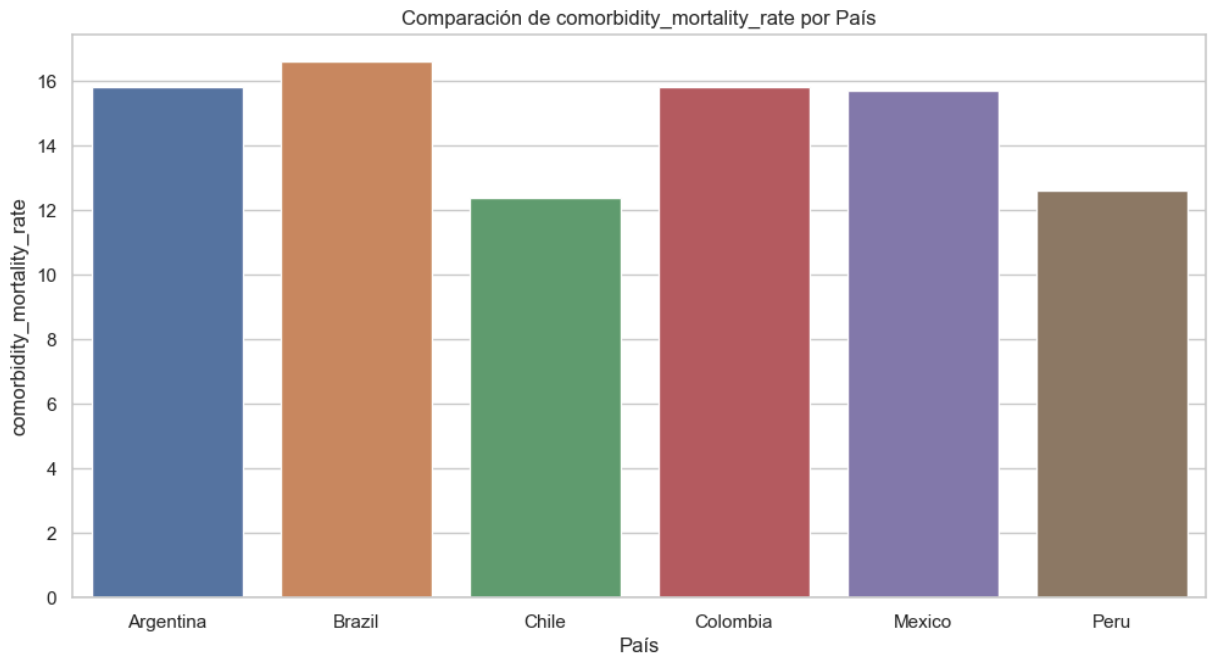






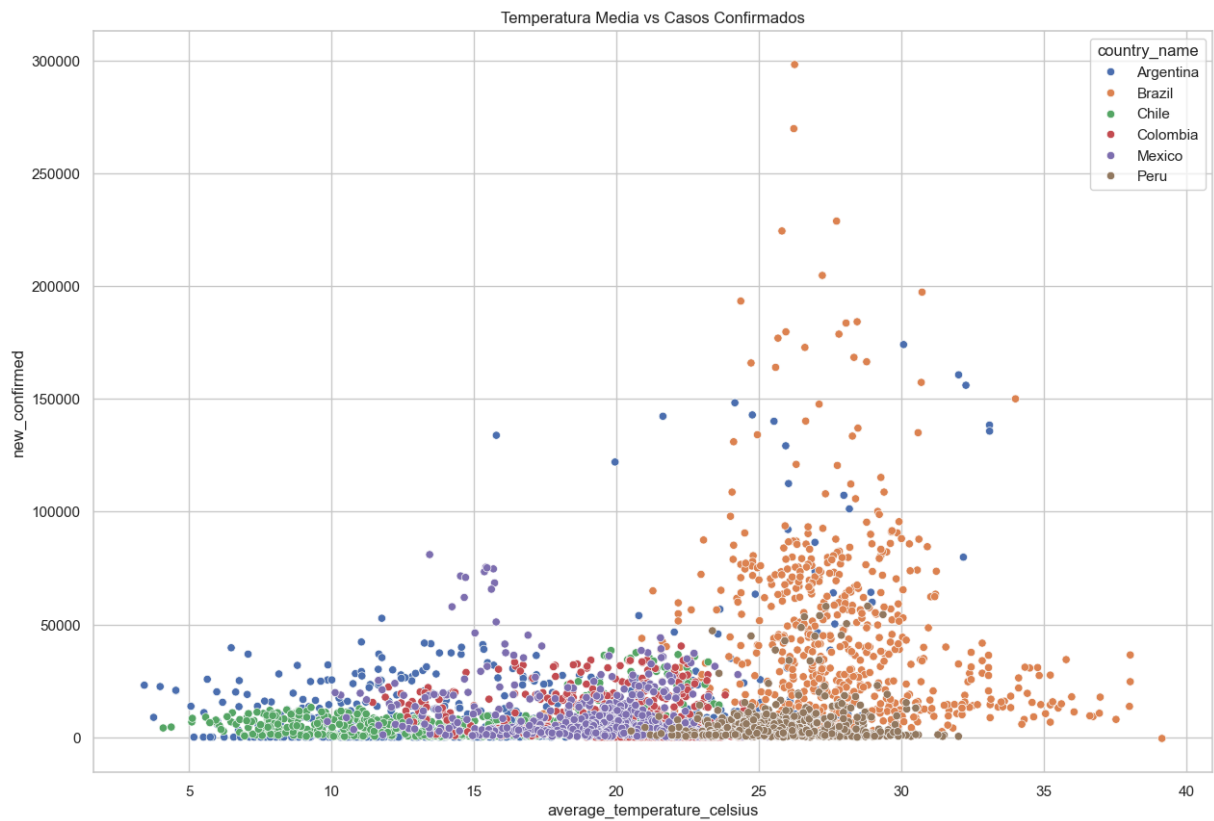






Temperatura media vs Casos

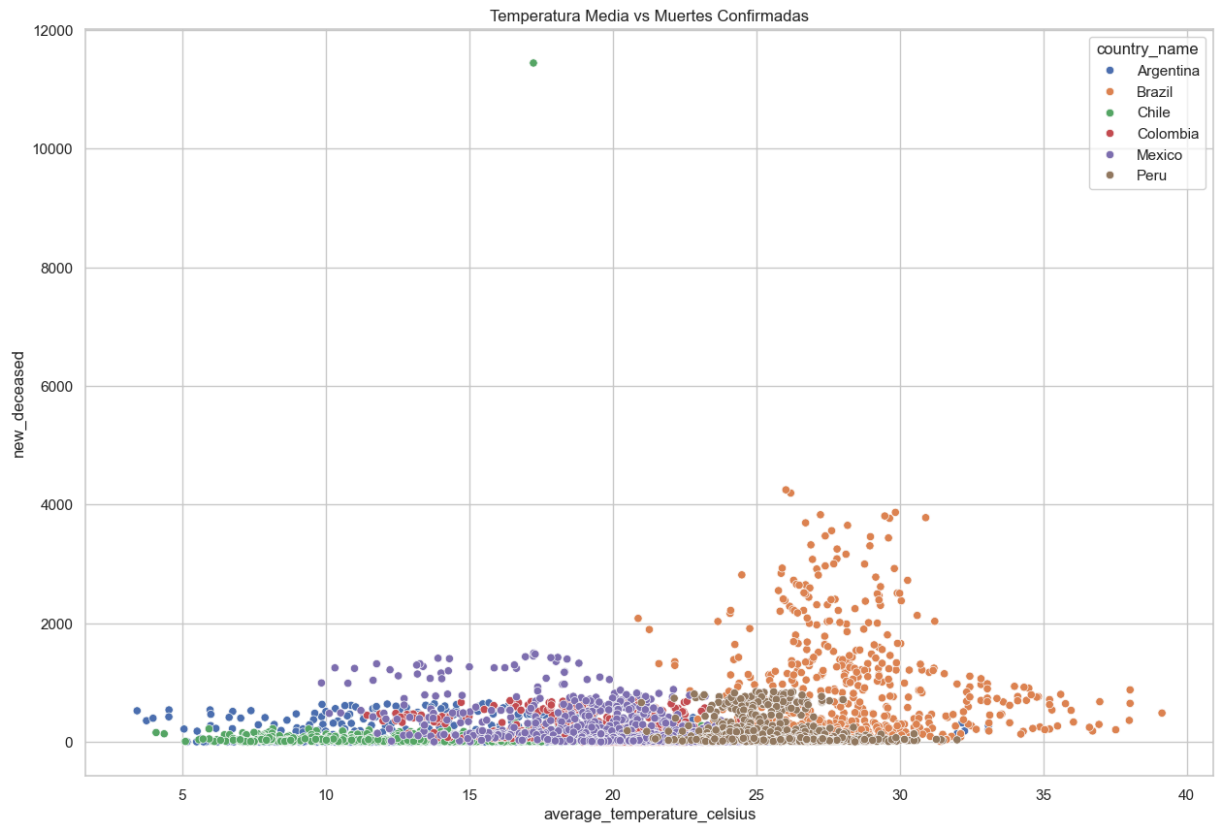
```
In [99]: plt.figure(figsize=(15,10))
sns.scatterplot(data=datos_países_interes, x='average_temperature_celsius', y='new_
plt.title('Temperatura Media vs Casos Confirmados')
plt.show()
```



Temp media vs Muertes

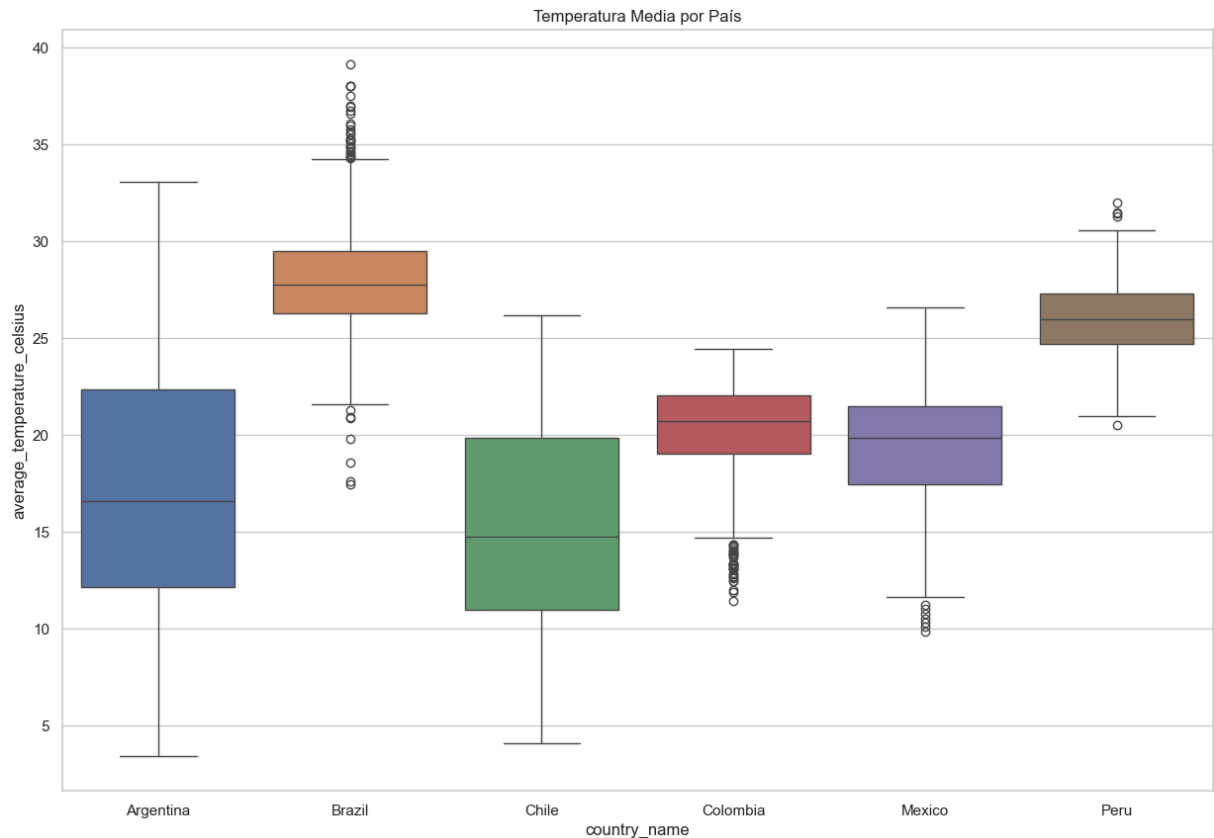
In [100...

```
plt.figure(figsize=(15,10))
sns.scatterplot(data=datos_paises_interes, x='average_temperature_celsius', y='new_
plt.title('Temperatura Media vs Muertes Confirmadas')
plt.show()
```



In [84]:

```
plt.figure(figsize=(15, 10))
sns.boxplot(data=datos_paises_interes, x='country_name', y='average_temperature_cel
plt.title('Temperatura Media por País')
plt.show()
```

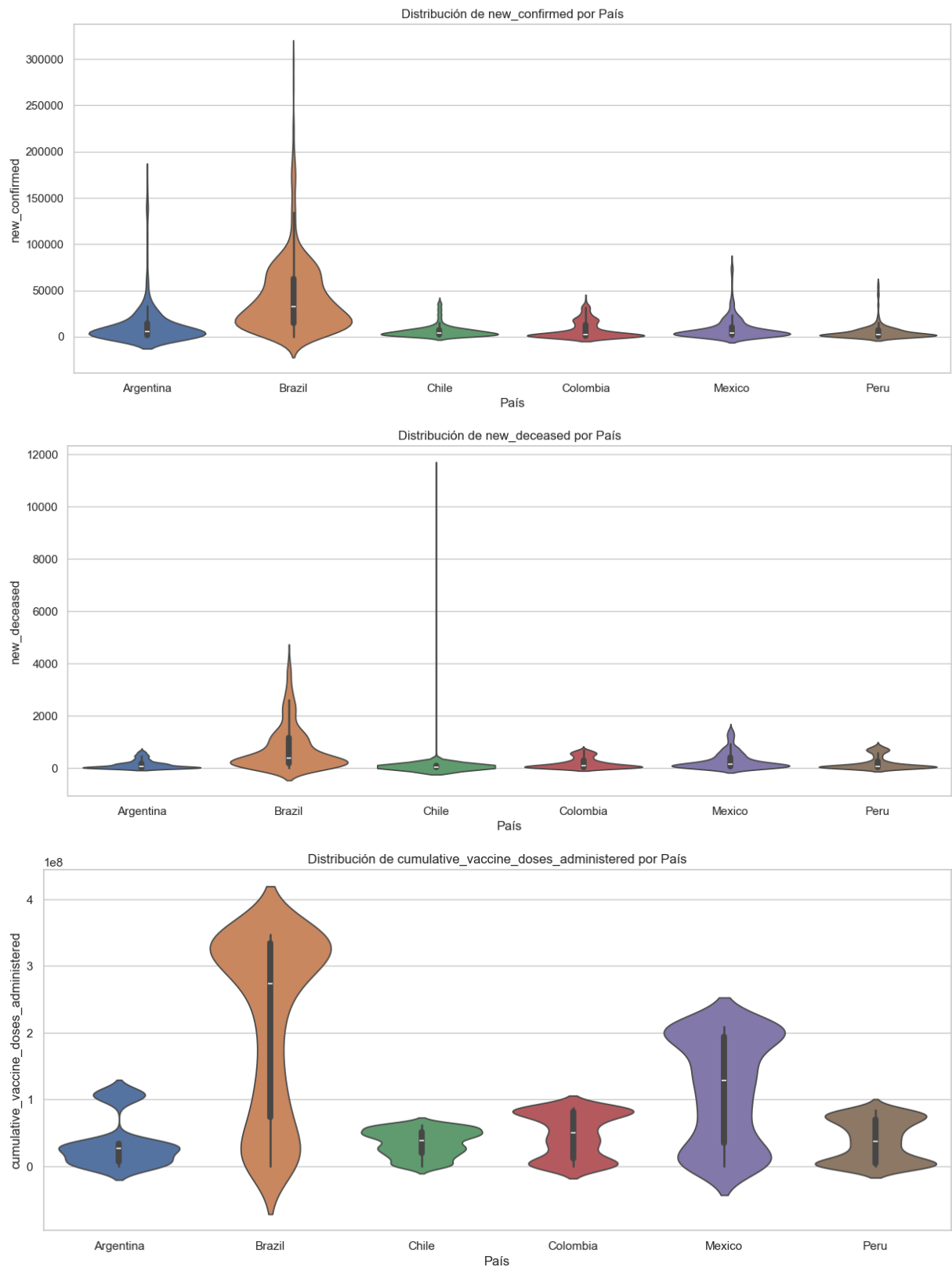


```
In [ ]: variables_interes = ['new_confirmed', 'new_deceased', 'cumulative_vaccine_doses_administered']

for variable in variables_interes:
    plt.figure(figsize=(15, 6))
    sns.violinplot(data=datos_paises_interes, x='country_name', y=variable, hue='country_name')
    plt.title(f'Distribución de {variable} por País')
    plt.xlabel('País')
    plt.ylabel(variable)
    plt.show()
```

```
In [85]: variables_interes = ['new_confirmed', 'new_deceased', 'cumulative_vaccine_doses_administered']

for variable in variables_interes:
    plt.figure(figsize=(15, 6))
    sns.violinplot(data=datos_paises_interes, x='country_name', y=variable, hue='country_name')
    plt.title(f'Distribución de {variable} por País')
    plt.xlabel('País')
    plt.ylabel(variable)
    plt.show()
```



```
In [86]: # Crear variables normalizadas por población
datos_paises_interes['new_confirmed_per_capita'] = datos_paises_interes['new_confirmed'] / datos_paises_interes['population']
datos_paises_interes['new_deceased_per_capita'] = datos_paises_interes['new_deceased'] / datos_paises_interes['population']
```

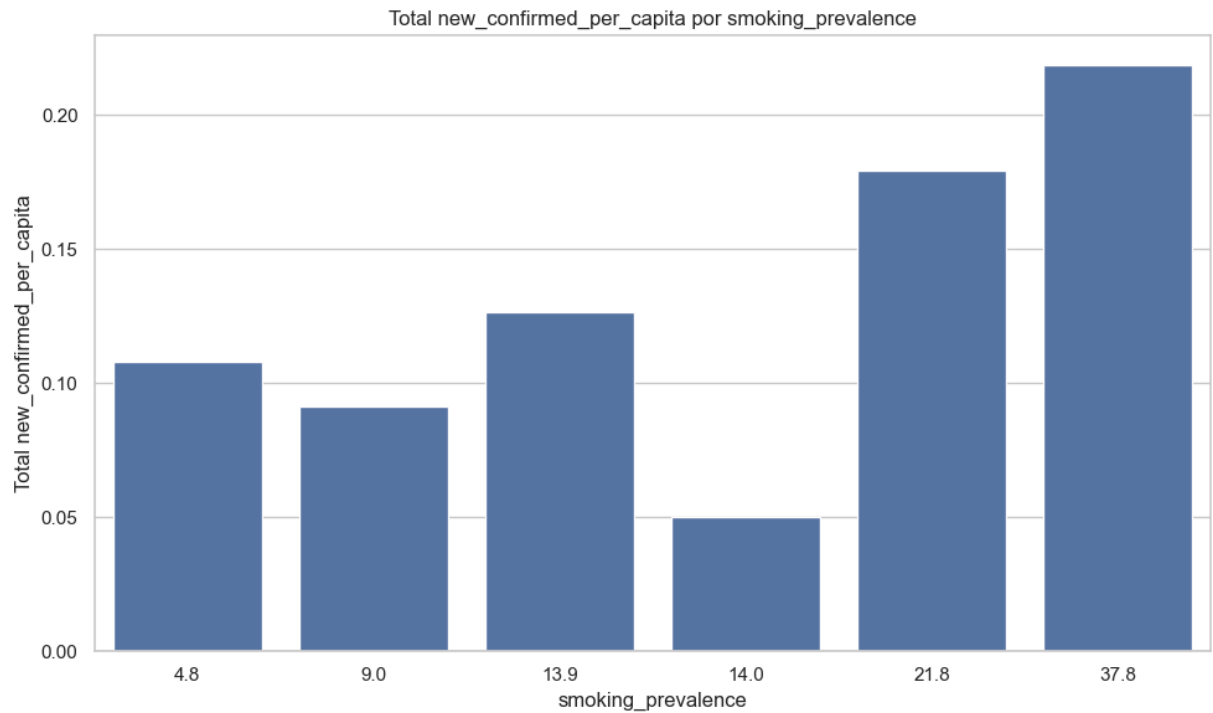
```
In [87]: combinaciones = [
    ('smoking_prevalence', 'new_confirmed_per_capita'),
    ('smoking_prevalence', 'new_deceased_per_capita'),
```

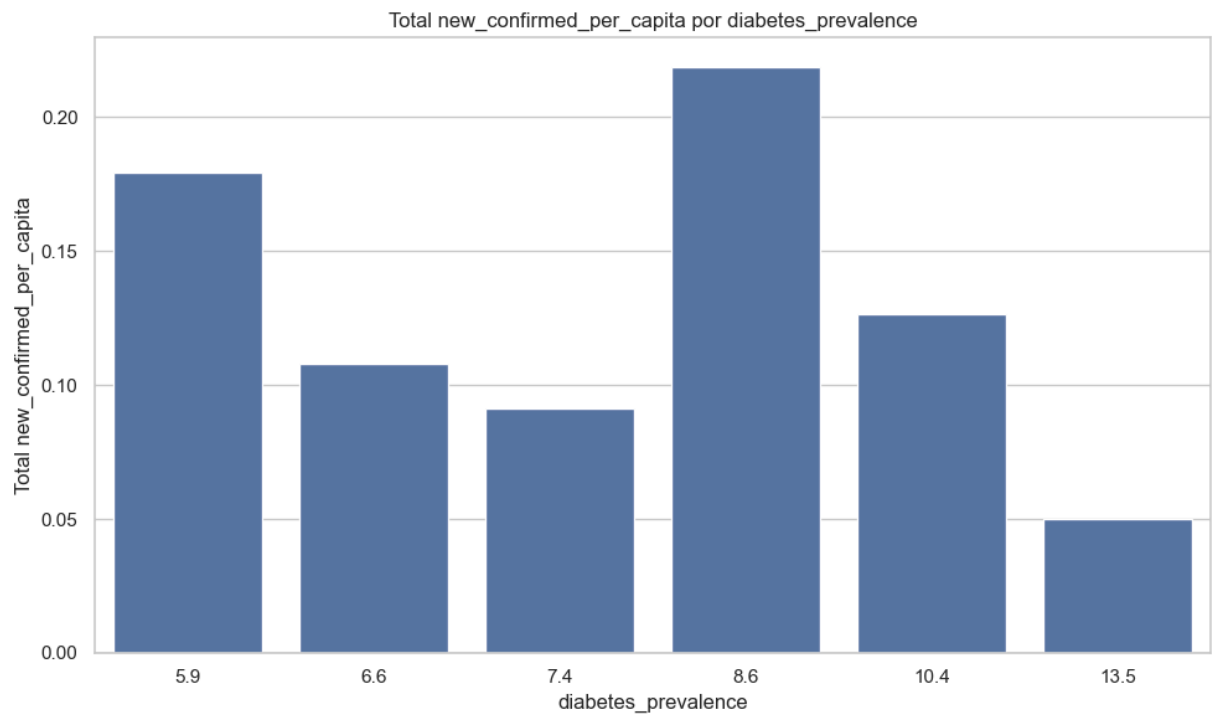
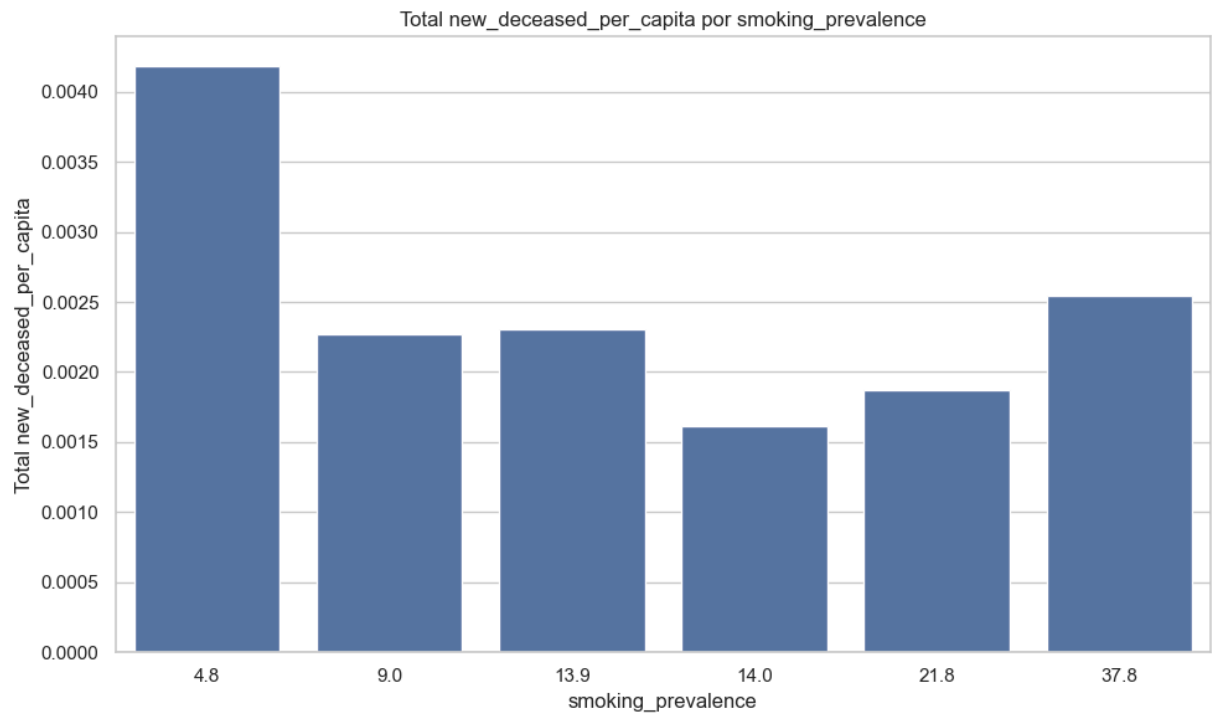
```

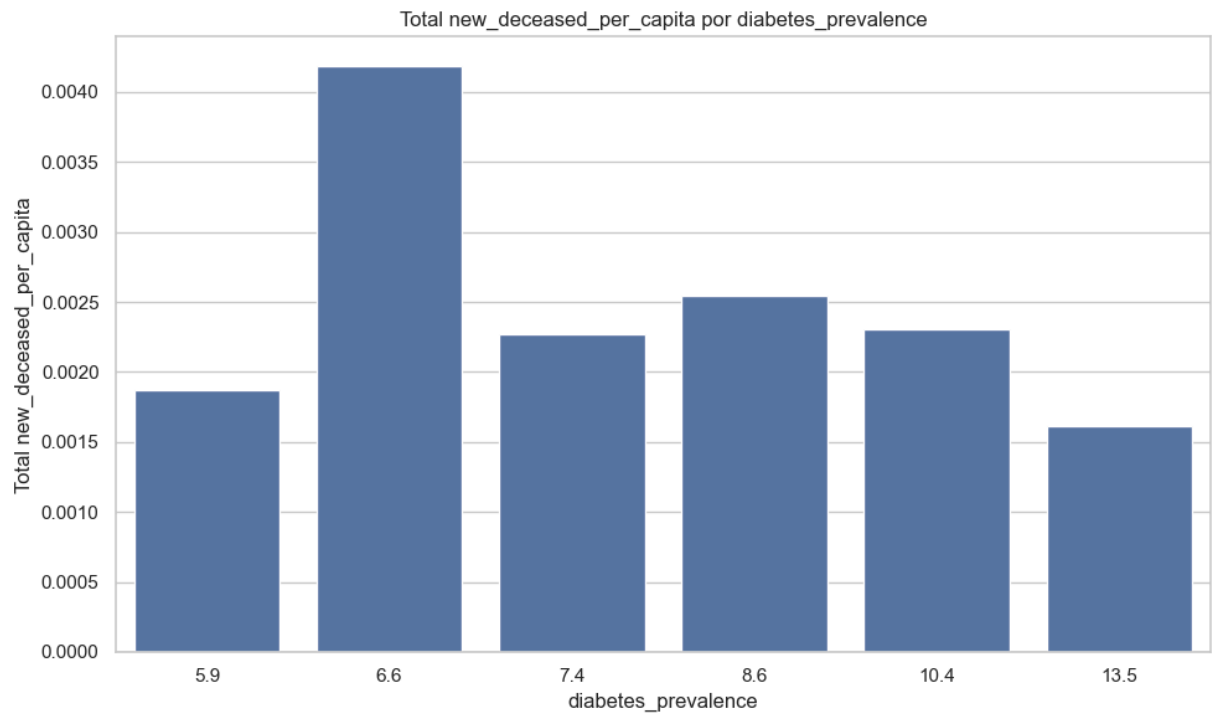
('diabetes_prevalence', 'new_confirmed_per_capita'),
('diabetes_prevalence', 'new_deceased_per_capita')
]

for x_var, y_var in combinaciones:
    datos_agrupados = datos_paises_interes.groupby(x_var)[y_var].sum().reset_index()
    plt.figure(figsize=(10,6))
    sns.barplot(data=datos_agrupados, x=x_var, y=y_var)
    plt.xlabel(x_var)
    plt.ylabel(f"Total {y_var}")
    plt.title(f"Total {y_var} por {x_var}")
    plt.tight_layout()
    plt.show()

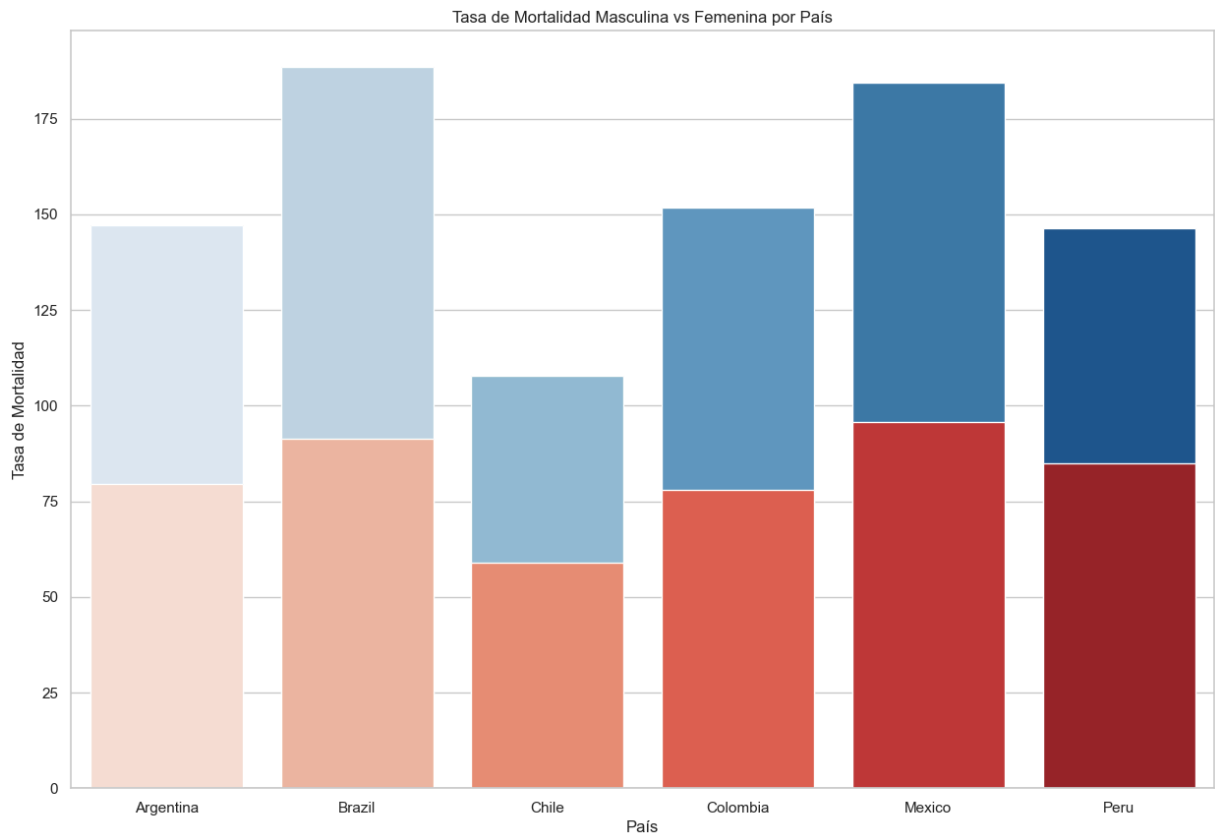
```



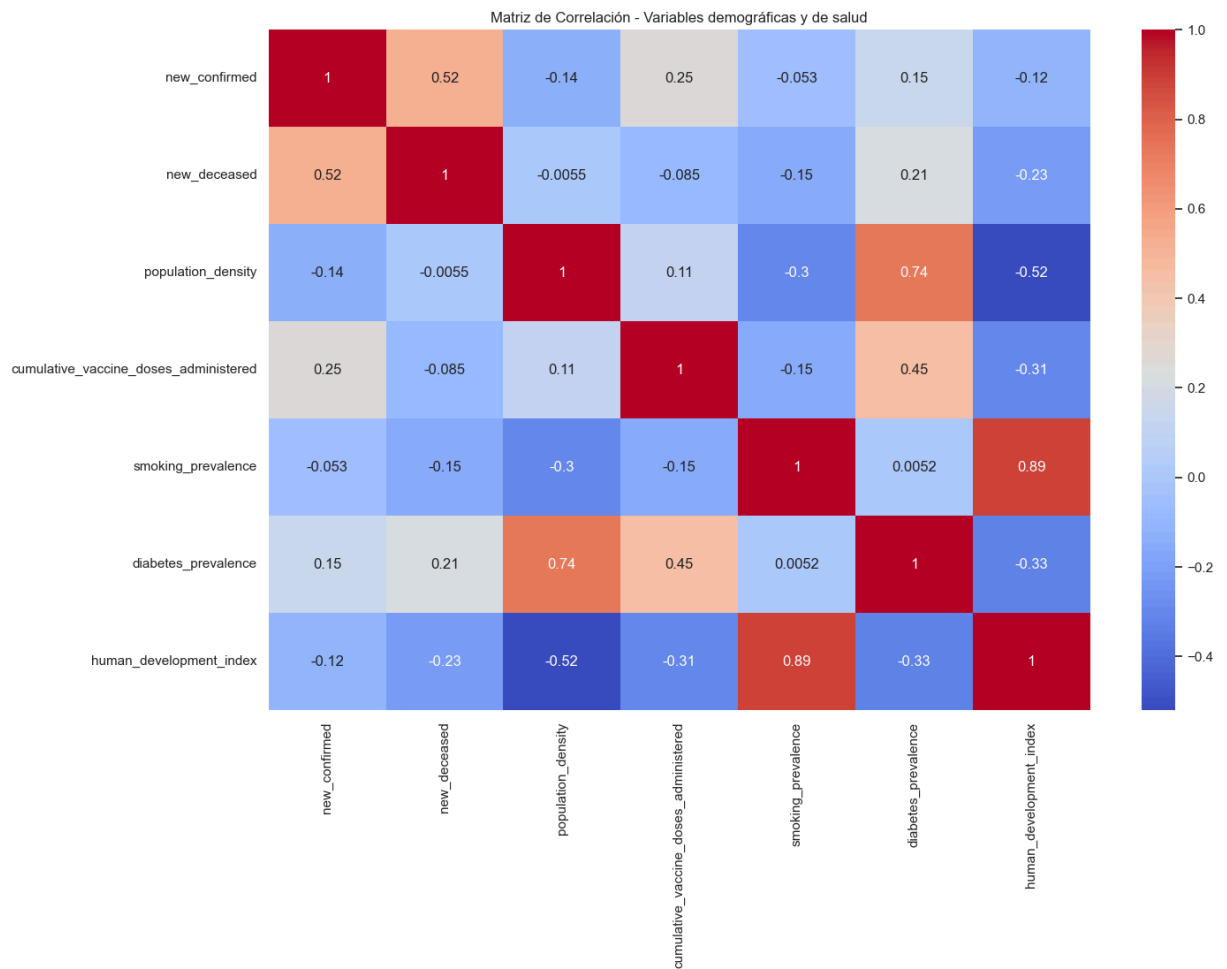




```
In [88]: plt.figure(figsize=(15, 10))
sns.barplot(data=datos_paises_interes, x='country_name', y='adult_male_mortality_rate')
sns.barplot(data=datos_paises_interes, x='country_name', y='adult_female_mortality_rate')
plt.title('Tasa de Mortalidad Masculina vs Femenina por País')
plt.xlabel('País')
plt.ylabel('Tasa de Mortalidad')
plt.show()
```



```
In [89]: variables_interes = ['new_confirmed', 'new_deceased', 'population_density', 'cumulative_vaccine_doses_administered', 'smoking_prevalence', 'diabetes_prevalence', 'human_development_index']
matriz_correlacion = datos_paises_interes[variables_interes].corr()
plt.figure(figsize=(15, 10))
sns.heatmap(matriz_correlacion, annot=True, cmap='coolwarm')
plt.title('Matriz de Correlación - Variables demográficas y de salud')
plt.show()
```



Avance 3:

```
In [3]: import geopandas as gpd
```

```
In [9]: datos_paises_interes['date'] = pd.to_datetime(datos_paises_interes['date'])
```

```
In [10]: datos_paises_interes.head(10)
```

Out[10]:

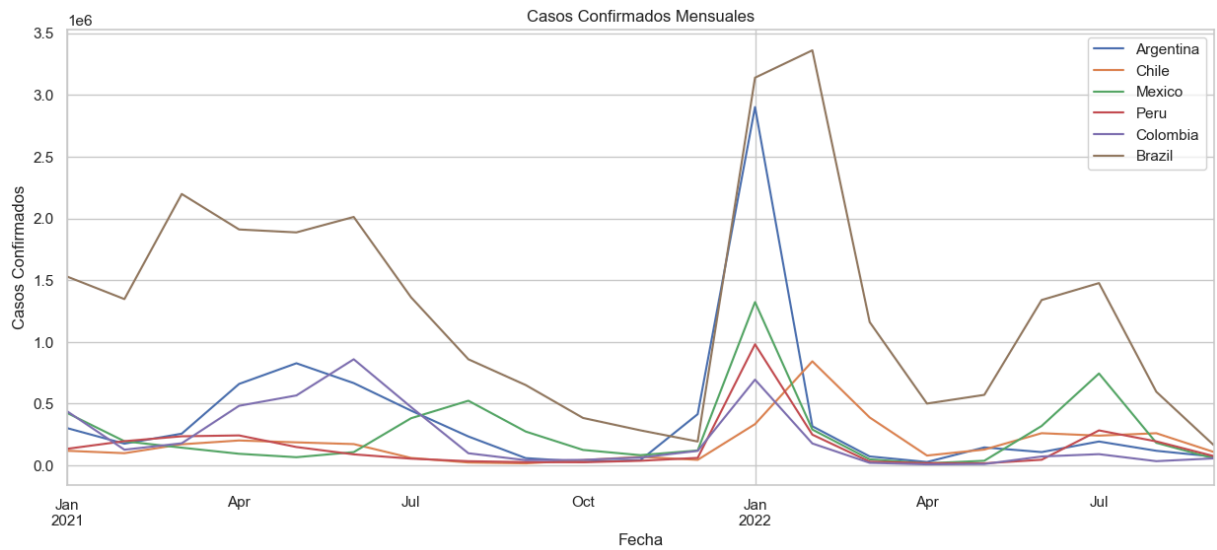
	location_key	date	country_code	country_name	new_confirmed	new_deceased	cumu
0	AR	2021-01-01	AR	Argentina	2685.0	140.0	
1	AR	2021-01-02	AR	Argentina	7767.0	166.0	
2	AR	2021-01-03	AR	Argentina	4934.0	157.0	
3	AR	2021-01-04	AR	Argentina	13953.0	157.0	
4	AR	2021-01-05	AR	Argentina	14085.0	160.0	
5	AR	2021-01-06	AR	Argentina	14496.0	131.0	
6	AR	2021-01-07	AR	Argentina	13722.0	160.0	
7	AR	2021-01-08	AR	Argentina	13932.0	162.0	
8	AR	2021-01-09	AR	Argentina	9959.0	158.0	
9	AR	2021-01-10	AR	Argentina	5174.0	165.0	

10 rows × 48 columns

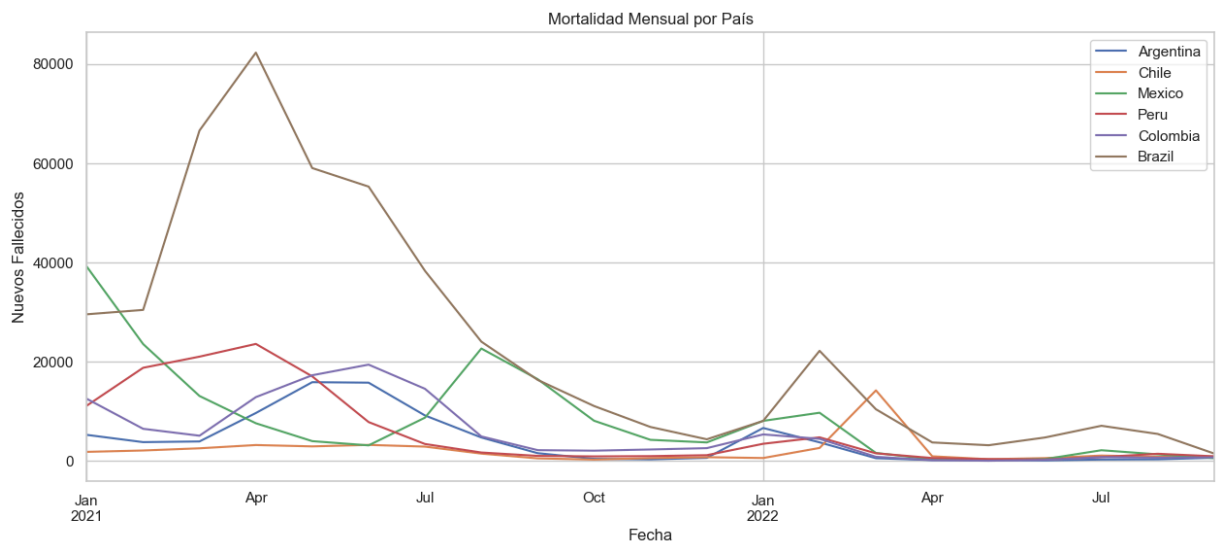


In [108...

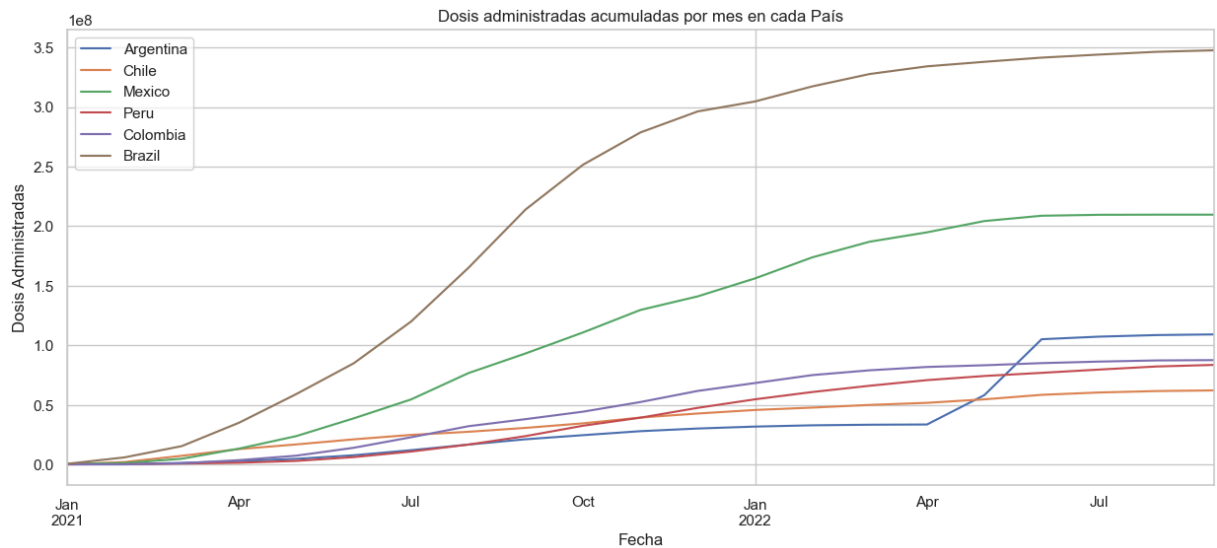
```
for pais in paises_interes:
    datos_paises_interes[datos_paises_interes['country_name'] == pais].set_index('date', inplace=True)
    plt.title('Casos Confirmados Mensuales')
    plt.xlabel('Fecha')
    plt.ylabel('Casos Confirmados')
    plt.legend()
```



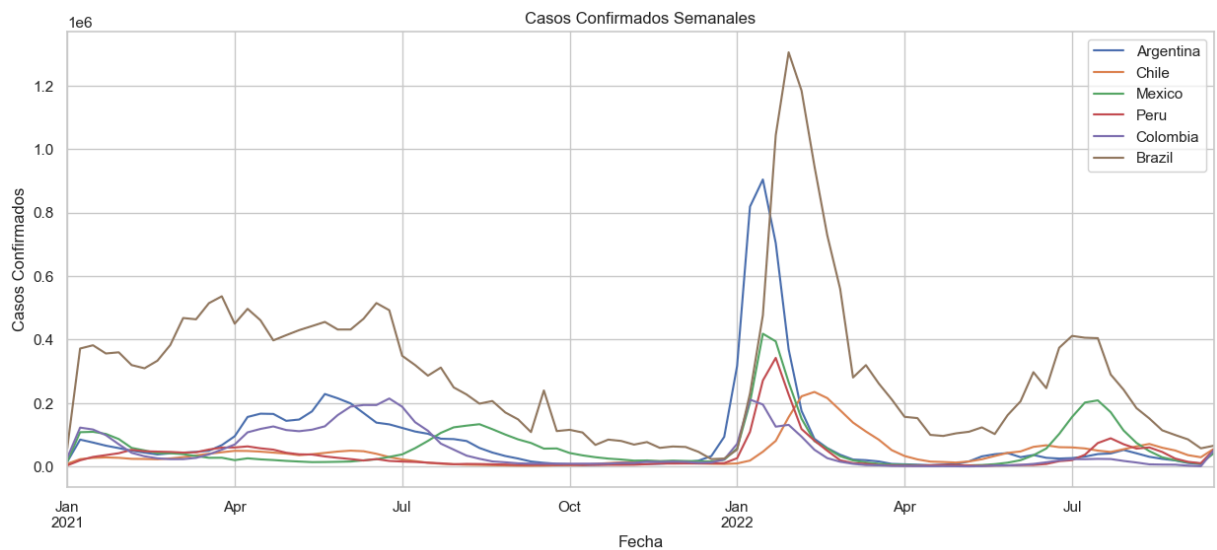
```
In [107... for pais in paises_interes:
    datos_paises_interes[datos_paises_interes['country_name'] == pais].set_index('date')
    plt.title('Mortalidad Mensual por País')
    plt.legend()
    plt.xlabel('Fecha')
    plt.ylabel('Nuevos Fallecidos')
```



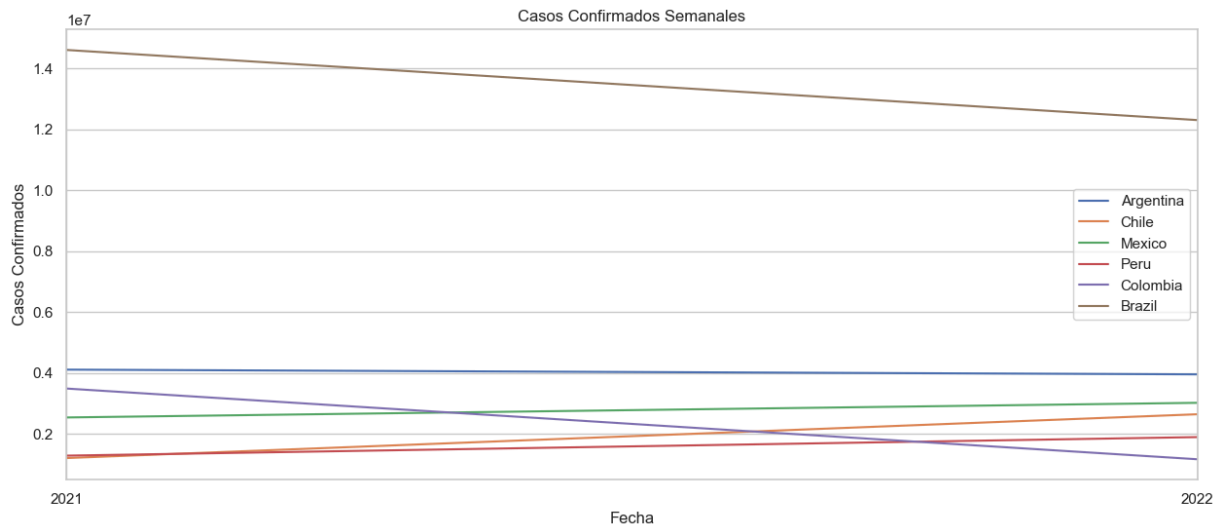
```
In [116... for pais in paises_interes:
    datos_paises_interes[datos_paises_interes['country_name'] == pais].set_index('date')
    plt.title('Dosis administradas acumuladas por mes en cada País')
    plt.legend()
    plt.xlabel('Fecha')
    plt.ylabel('Dosis Administradas')
```



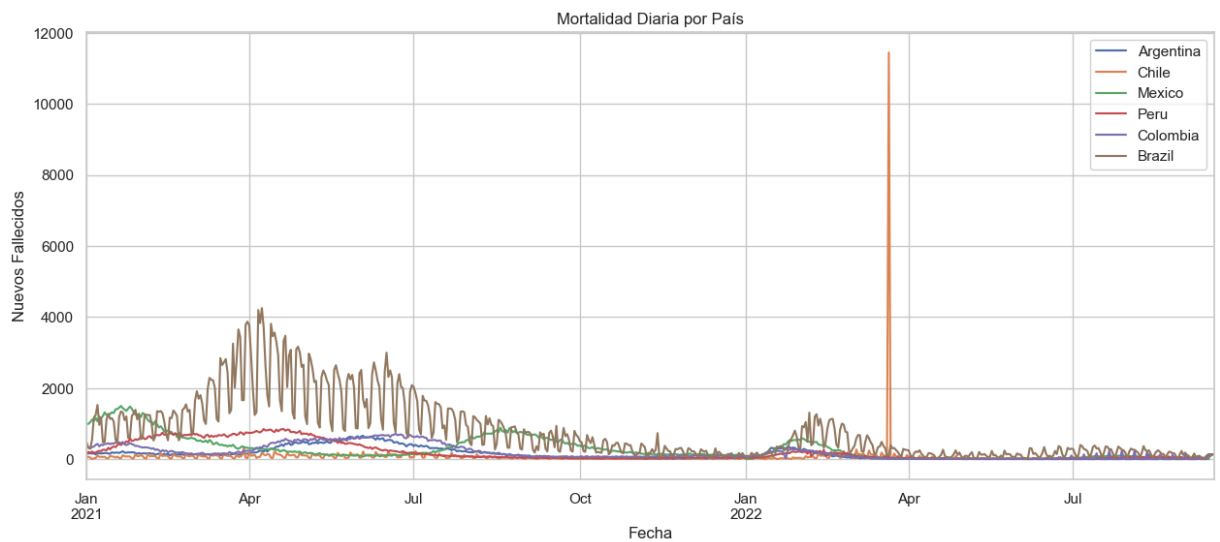
```
In [97]: for pais in paises_interes:
          datos_paises_interes[datos_paises_interes['country_name'] == pais].set_index('date', inplace=True)
          plt.title('Casos Confirmados Semanales')
          plt.xlabel('Fecha')
          plt.ylabel('Casos Confirmados')
          plt.legend()
```



```
In [99]: for pais in paises_interes:
          datos_paises_interes[datos_paises_interes['country_name'] == pais].set_index('date', inplace=True)
          plt.title('Casos Confirmados Semanales')
          plt.xlabel('Fecha')
          plt.ylabel('Casos Confirmados')
          plt.legend()
```

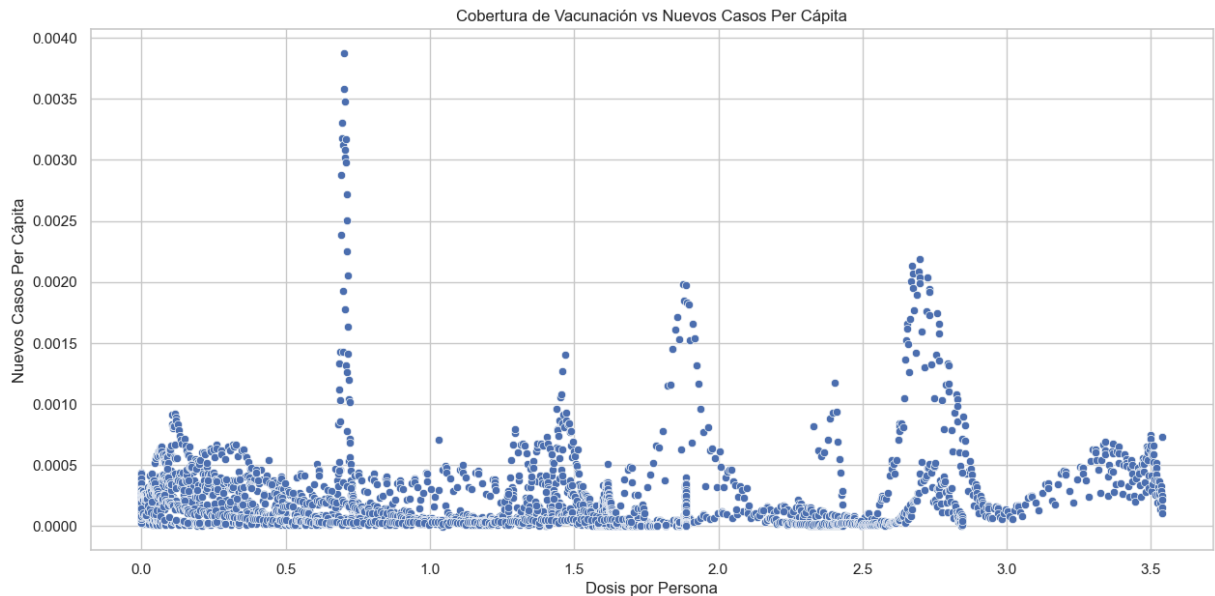


```
In [128... for pais in paises_interes:
    datos_paises_interes[datos_paises_interes['country_name'] == pais].set_index('date')
    plt.title('Mortalidad Diaria por País')
    plt.legend()
    plt.xlabel('Fecha')
    plt.ylabel('Nuevos Fallecidos')
```



```
In [107... datos_paises_interes['vacunacion_per_capita'] = datos_paises_interes['cumulative_vaccination']

plt.figure(figsize=(15, 7))
sns.scatterplot(data=datos_paises_interes, x='vacunacion_per_capita', y='new_confirmed')
plt.title('Cobertura de Vacunación vs Nuevos Casos Per Cápit')
plt.xlabel('Dosis por Persona')
plt.ylabel('Nuevos Casos Per Cápit')
plt.show()
```

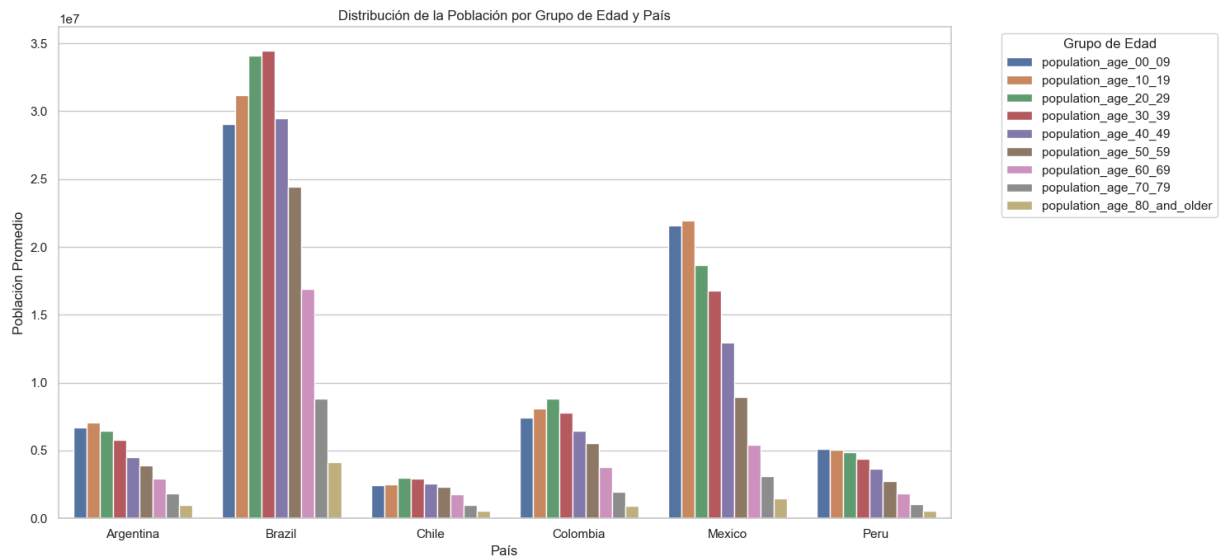


```
In [ ]: variables_edades = ['population_age_00_09', 'population_age_10_19' , 'population_ag

# Agrupar por país (usar promedio si hay varias fechas)
edades_por_pais = datos_paises_interes.groupby('country_name')[variables_edades].me

# Convertir a formato largo
edades_long = edades_por_pais.melt(
    id_vars='country_name',
    value_vars=variables_edades,
    var_name='Grupo_Edad',
    value_name='Poblacion'
)

# Gráfico de barras agrupadas por país y grupo de edad
plt.figure(figsize=(15, 7))
sns.barplot(data=edades_long, x='country_name', y='Poblacion', hue='Grupo_Edad')
plt.title('Distribución de la Población por Grupo de Edad y País')
plt.xlabel('País')
plt.ylabel('Población Promedio')
plt.legend(title='Grupo de Edad', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```



```
In [ ]: import geopandas as gpd
```

```
In [17]: world = gpd.read_file("ne_110m_admin_0_countries.shp")
```

```
In [20]: datos_resumen = datos_paises_interes.groupby('country_name')['new_confirmed'].sum()

world_merged = world.merge(
    datos_resumen,
    how='left',
    left_on='ADMIN',
    right_on='country_name'
)
```

```
In [ ]: fig, ax = plt.subplots(figsize=(20, 15))

world_merged.plot(
    column='new_confirmed',
    cmap='OrRd',
    legend=True,
    ax=ax,
)

ax.set_title('Mapa de Casos Confirmados por País', fontsize=20)
ax.axis('off')

plt.show()
```


Mapa de Casos Confirmados por País

