

Taller 2

October 8, 2023

1 Carga de los datos

2 1. Entendimiento de los datos

2.1 Diccionario

country: Unique Identifier

incomeperperson: 2010 Gross Domestic Product per capita in constant 2000 US\$. The inflation but not the differences in the cost of living between countries has been taken into account.

alconsumption: 2008 Alcohol consumption per adult (age 15+), litres Recorded and estimated average alcohol consumption, adult (15+) per capita consumption in litres pure alcohol

armedforcesrate: Armed forces personnel (% of the total labour force)

breastcancerper100TH: 2002 breast cancer new cases per 100,000 females. Number of new cases of breast cancer in 100,000 female residents during a certain year.

co2emissions: 2006 cumulative CO2 emission (metric tons), Total amount of CO2 emission in metric tons since 1751.

femaleemployrate: 2007 female employees age 15+ (% of the population). Percentage of female population, age above 15, that has been employed during the given year.

HIVrate: 2009 estimated HIV Prevalence % - (Ages 15-49). Estimated number of people living with HIV per 100 population of age group 15-49.

Internetuserate: 2010 Internet users (per 100 people). Internet users are people with access to the worldwide network.

lifeexpectancy: 2011 life expectancy at birth (years). The average number of years a newborn child would live if current mortality patterns were to stay the same.

oilperperson: 2010 oil Consumption per capita (tonnes per year and person).

polityscore: 2009 Democracy score (Polity). Overall polity score from the Polity IV dataset, calculated by subtracting an autocracy score from a democracy score. The summary measure of a country's democratic and free nature. -10 is the lowest value, 10 the highest.

relectricperperson 2008 residential electricity consumption, per person (kWh). The amount of residential electricity consumption per person during the given year, counted in kilowatt-hours (kWh).

suicideper100TH: Suicide, age-adjusted, per 100,000 Mortality due to self-inflicted injury, per 100,000 standard population, age-adjusted.

employrate: 2007 total employees age 15+ (% of the population). Percentage of the total p population, age above 15, that has been employed during the given year.

urbanrate: Urban population (% of total) Urban population refers to people living in urban areas as defined by national statistical offices (calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects)

2.2 1.1 Tipos de datos todas las columnas son indicadores numéricos

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype
---  -
0   country               178 non-null   object
1   incomeperperson       175 non-null   float64
2   alccconsumption       178 non-null   float64
3   armedforcesrate      174 non-null   float64
4   breastcancerper100th  177 non-null   float64
5   co2emissions          176 non-null   float64
6   femaleemployrate      177 non-null   float64
7   hivrate               155 non-null   float64
8   internetuserate       174 non-null   float64
9   lifeexpectancy        178 non-null   float64
10  oilperperson          67 non-null    float64
11  polityscore           167 non-null   float64
12  relectricperperson    141 non-null   float64
13  suicideper100th       178 non-null   float64
14  employrate            177 non-null   float64
15  urbanrate             178 non-null   float64
16  region                178 non-null   object
17  income_groups         174 non-null   object
dtypes: float64(15), object(3)
memory usage: 25.2+ KB
```

Como se observa en la tabla anterior, y de acuerdo con el diccionario de datos, con excepción de la columna `country` todas las columnas son indicadores numéricos

2.3 1.3 Análisis univariado

	count	mean	std	min \
incomeperperson	175.0	7.553392e+03	1.107457e+04	103.775857
alccconsumption	175.0	6.865257e+00	5.111631e+00	0.050000
armedforcesrate	171.0	1.376918e+00	1.510972e+00	0.000000
breastcancerper100th	174.0	3.781092e+01	2.336893e+01	3.900000
co2emissions	173.0	5.845393e+09	2.757598e+10	850666.666700

femaleemployrate	174.0	4.765805e+01	1.469791e+01	12.400000
hivrate	154.0	1.862662e+00	4.288853e+00	0.060000
internetuserate	172.0	3.490039e+01	2.815022e+01	0.210066
lifeexpectancy	175.0	6.952811e+01	9.753006e+00	47.794000
oilperperson	66.0	1.355786e+00	1.669981e+00	0.032281
polityscore	164.0	3.646341e+00	6.379361e+00	-10.000000
relectricperperson	137.0	1.154900e+03	1.552701e+03	9.192395
suicideper100th	175.0	1.005070e+01	6.540393e+00	0.201449
employrate	174.0	5.874195e+01	1.033769e+01	34.900002
urbanrate	175.0	5.636137e+01	2.232546e+01	10.400000
log_incomeperperson	175.0	7.830242e+00	1.598949e+00	4.651823
log_alcconsumption	175.0	1.801144e+00	7.877582e-01	0.048790
log_armedforcesrate	171.0	7.341038e-01	4.740197e-01	0.000000
log_breastcancerper100th	174.0	3.481636e+00	6.097279e-01	1.589235
log_co2emissions	173.0	1.987728e+01	2.370644e+00	13.653777
log_hivrate	154.0	6.299275e-01	7.419928e-01	0.058269
log_oilperperson	66.0	7.319055e-01	4.491709e-01	0.031771
log_relectricperperson	139.0	6.200701e+00	1.576966e+00	0.000000
log_suicideper100th	175.0	2.228666e+00	6.195432e-01	0.183528

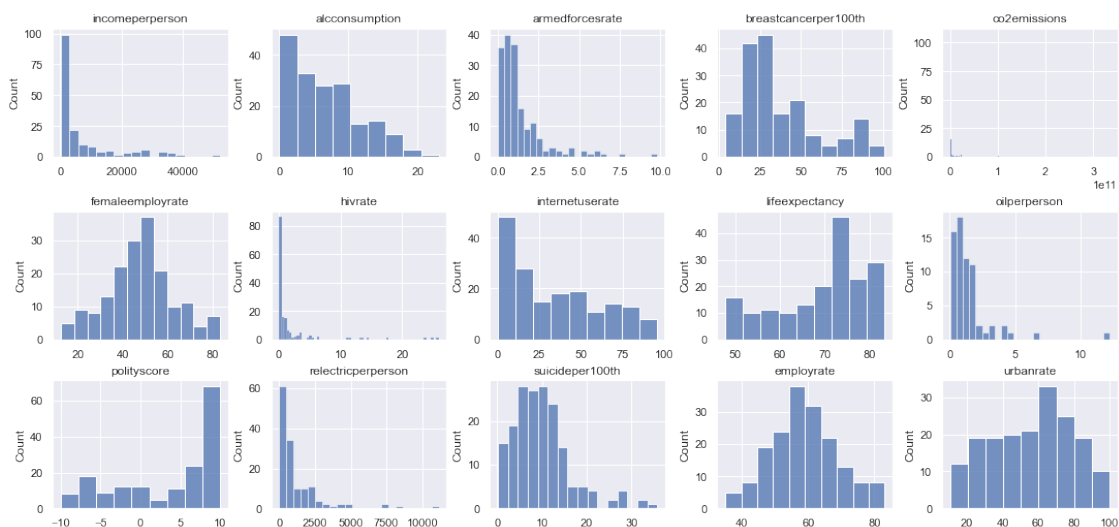
	25%	50%	75% \
incomeperperson	6.097443e+02	2.425471e+03	8.880432e+03
alcconsumption	2.560000e+00	6.080000e+00	1.003500e+01
armedforcesrate	4.848023e-01	9.311688e-01	1.565301e+00
breastcancerper100th	2.045000e+01	3.015000e+01	5.075000e+01
co2emissions	8.797067e+07	3.009343e+08	2.406741e+09
femaleemployrate	3.890000e+01	4.845000e+01	5.587500e+01
hivrate	1.000000e-01	4.000000e-01	1.200000e+00
internetuserate	9.973925e+00	2.993993e+01	5.630003e+01
lifeexpectancy	6.274700e+01	7.297400e+01	7.633700e+01
oilperperson	5.702306e-01	8.908132e-01	1.576598e+00
polityscore	-2.000000e+00	6.000000e+00	9.000000e+00
relectricperperson	2.614340e+02	6.179256e+02	1.566106e+03
suicideper100th	5.802861e+00	9.127511e+00	1.264050e+01
employrate	5.132500e+01	5.870000e+01	6.480000e+01
urbanrate	3.755000e+01	5.962000e+01	7.346000e+01
log_incomeperperson	6.414678e+00	7.794193e+00	9.091395e+00
log_alcconsumption	1.269761e+00	1.957274e+00	2.401064e+00
log_armedforcesrate	3.952812e-01	6.581254e-01	9.420610e-01
log_breastcancerper100th	3.065717e+00	3.438803e+00	3.946412e+00
log_co2emissions	1.829251e+01	1.952240e+01	2.160154e+01
log_hivrate	9.531018e-02	3.364722e-01	7.884574e-01
log_oilperperson	4.511645e-01	6.370070e-01	9.464677e-01
log_relectricperperson	5.515861e+00	6.423096e+00	7.333571e+00
log_suicideper100th	1.917330e+00	2.315256e+00	2.612902e+00

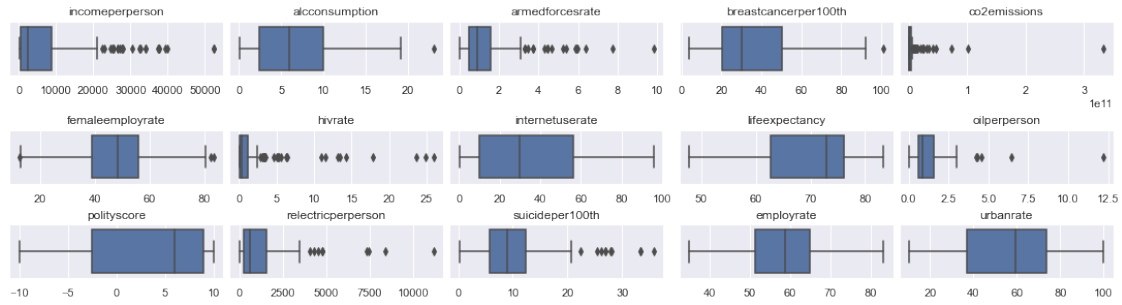
max

incomeperperson	5.230159e+04
alcoholconsumption	2.301000e+01
armedforcesrate	9.820127e+00
breastcancerper100th	1.011000e+02
co2emissions	3.340000e+11
femaleemployrate	8.330000e+01
hivrate	2.590000e+01
internetuserate	9.563811e+01
lifeexpectancy	8.339400e+01
oilperperson	1.222864e+01
polityscore	1.000000e+01
relectricperperson	1.115476e+04
suicideper100th	3.575287e+01
employrate	8.320000e+01
urbanrate	1.000000e+02
log_incomeperperson	1.086480e+01
log_alcoholconsumption	3.178470e+00
log_armedforcesrate	2.381408e+00
log_breastcancerper100th	4.625953e+00
log_co2emissions	2.653441e+01
log_hivrate	3.292126e+00
log_oilperperson	2.582385e+00
log_relectricperperson	9.319711e+00
log_suicideper100th	3.604216e+00

En la tabla anterior se observan algunos valores posiblemente erroneos, en particular parece extraño que existan países en que el consumo electrico por persona sea cero, aunque puede que sea aproximadamente cero .

```
13      Iraq
175     Namibia
Name: country, dtype: object
```





De las gráficas anteriores se pueden observar valores atípicos en bastantes de las variables. Para intentar mitigar este efecto se propone aplicar logaritmo a estas variables con lo cual se espera obtener un distribución más cercana a una normal:



Vemos que con esta transformación se mitiga el efecto de los datos extremos en varias de las variables

y se obtienen distribuciones más cercanas a una normal. Aunque aun persisten algunos valores atípicos.

3 Análisis bivariado

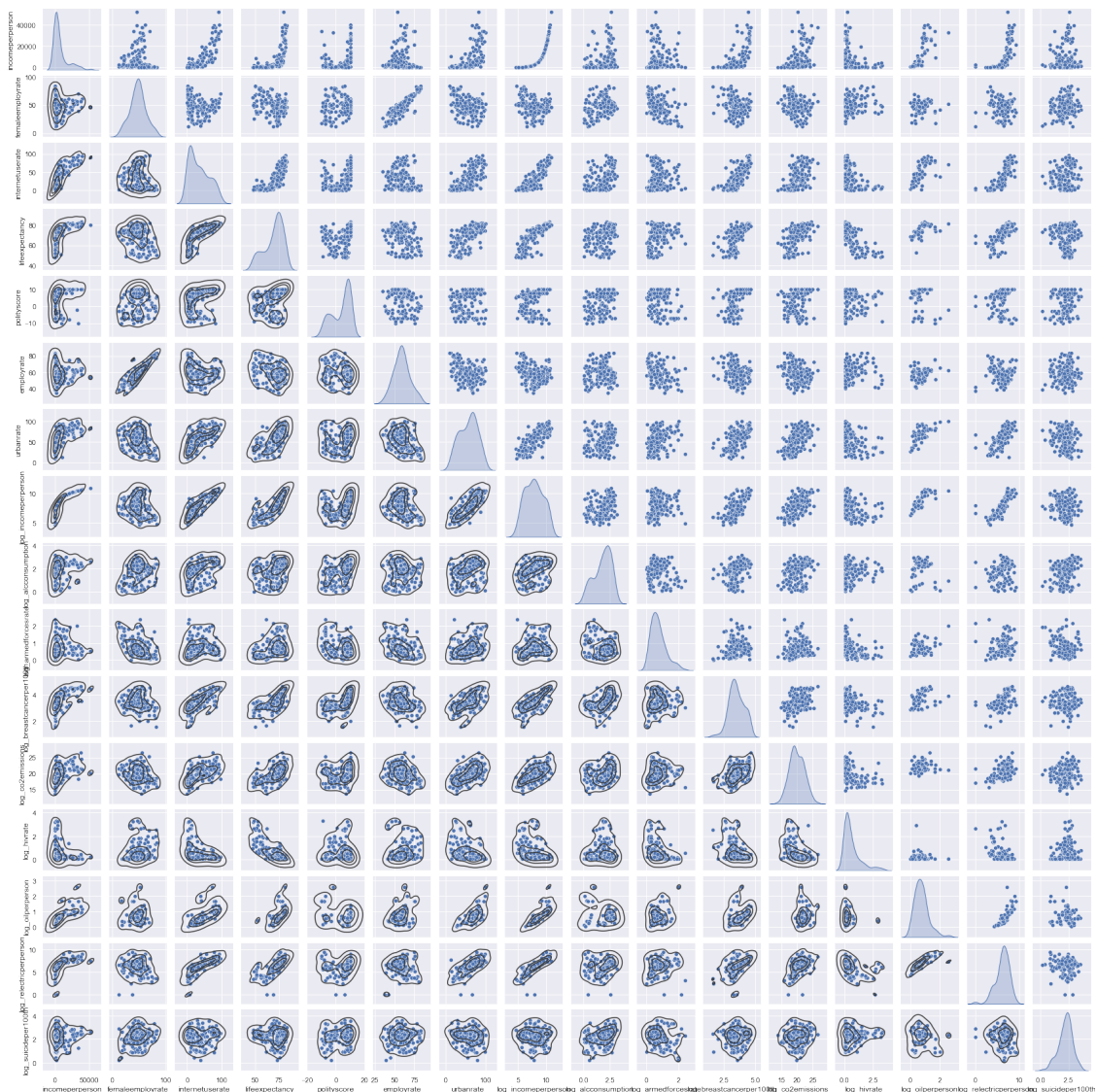
```
C:\Users\yacha\AppData\Local\Programs\Python\Python39\lib\site-
packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to
tight
```

```
self._figure.tight_layout(*args, **kwargs)
```

```
C:\Users\yacha\AppData\Local\Programs\Python\Python39\lib\site-
packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to
tight
```

```
self._figure.tight_layout(*args, **kwargs)
```

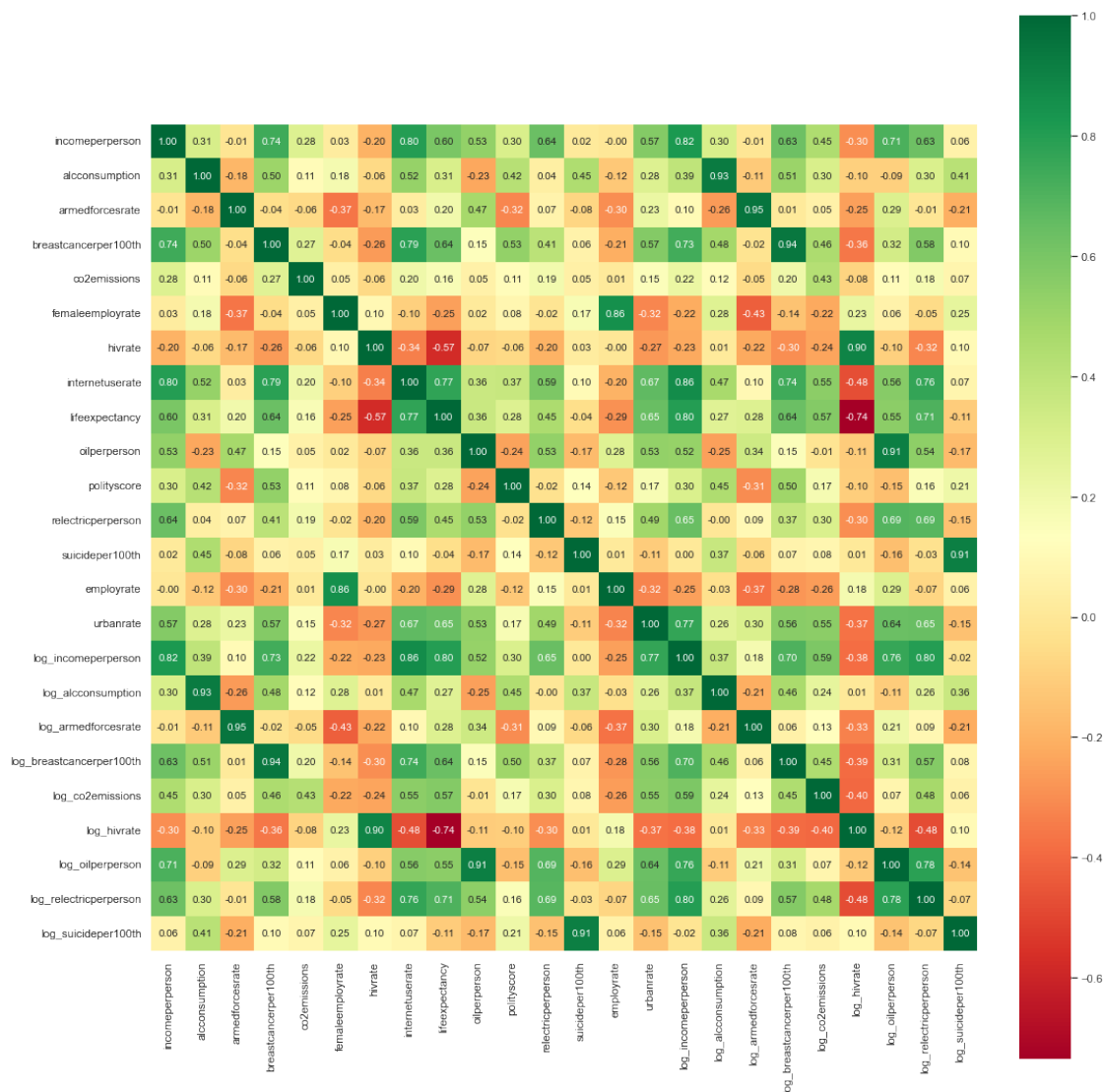
```
<seaborn.axisgrid.PairGrid at 0x1b383532610>
```



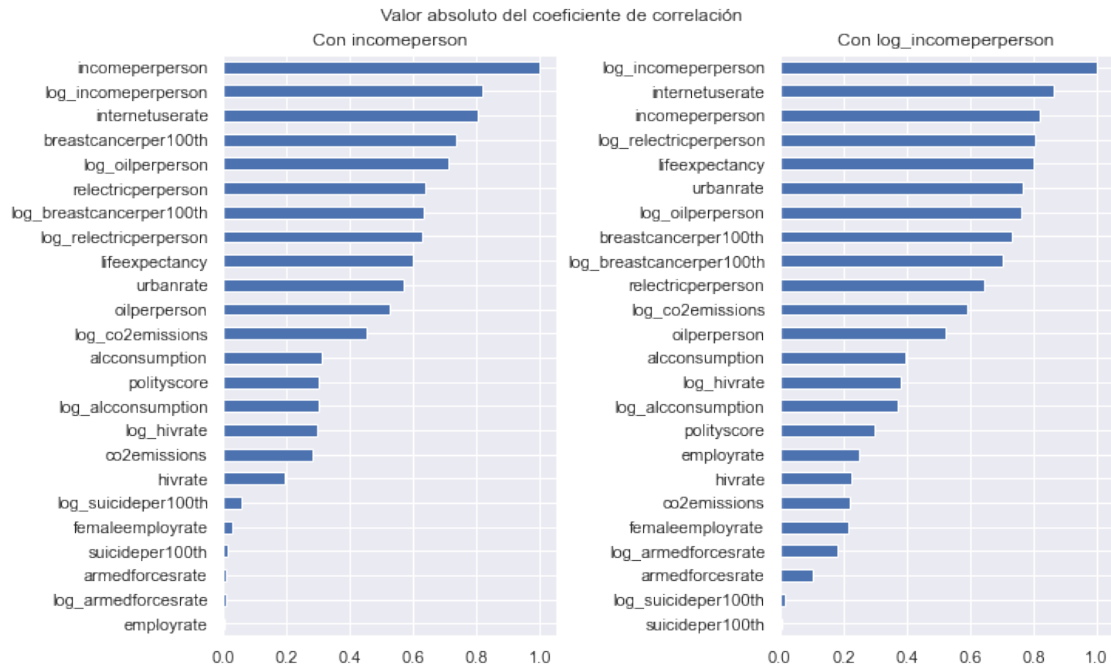
A partir de la gráfica anterior es posible observar que existen varias variables que tienen una fuerte relación con en el ingreso per cápita y que esta relación es aproximadamente lineal con el logaritmo de esta variable. Lo cual es particularmente notorio en el caso de: - **internetuserate** - **log relectricperperson** - **breastcancerper100th** - **urbanrate** - **log oilperperson**

3.0.1 Análisis de correlaciones

<Axes: >



<Axes: title={'center': 'Con log_incomeperperson'}>



De las dos gráficas anteriores se seleccionan las cinco variables que tienen mayor correlación con el ingreso o logaritmo del ingreso:

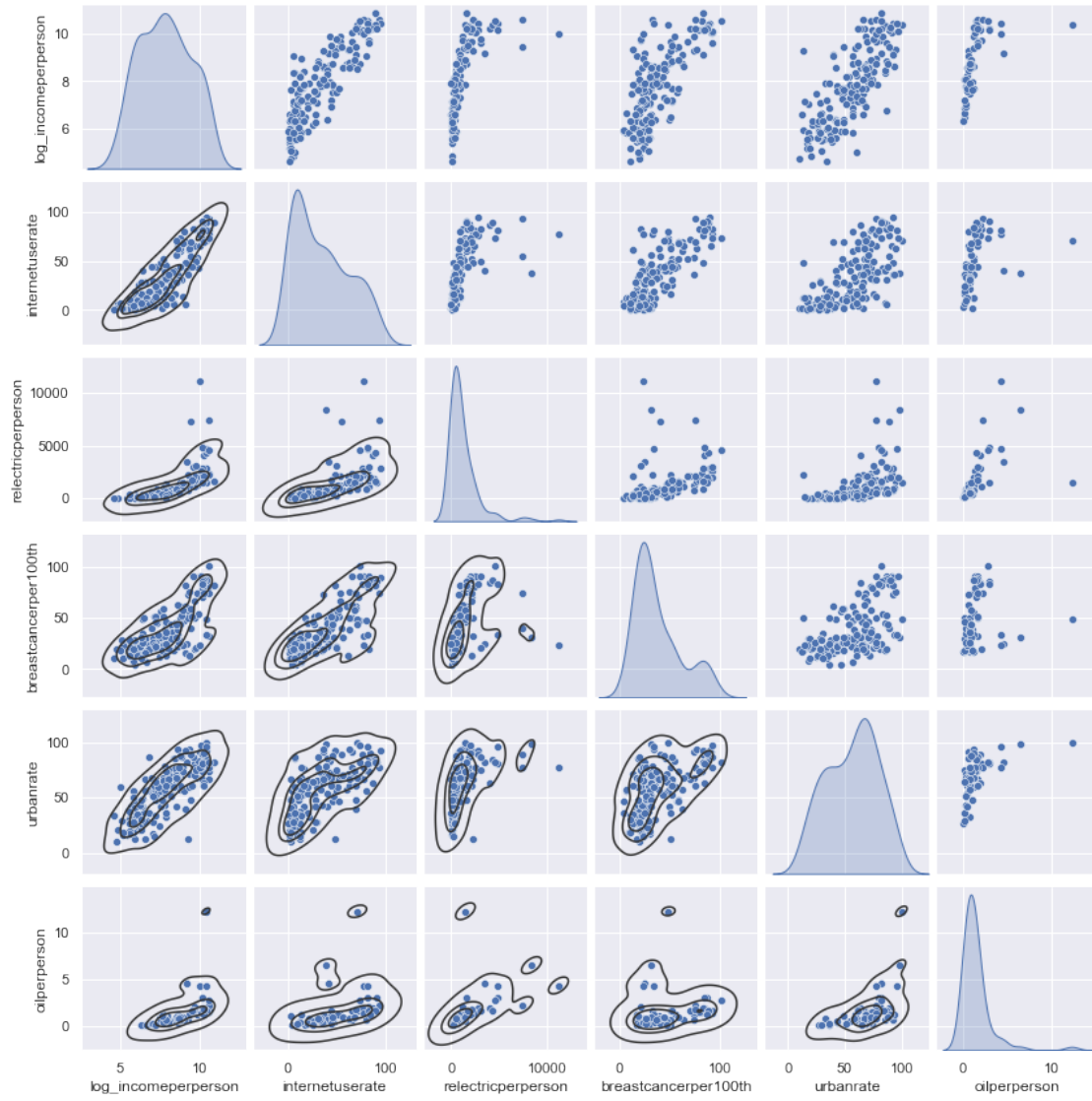
- **internetuserate**
- **relectricperson**
- **breastcancerper100th**
- **urbanrate**
- **oilperson**

```
C:\Users\yacha\AppData\Local\Programs\Python\Python39\lib\site-
packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to
tight
```

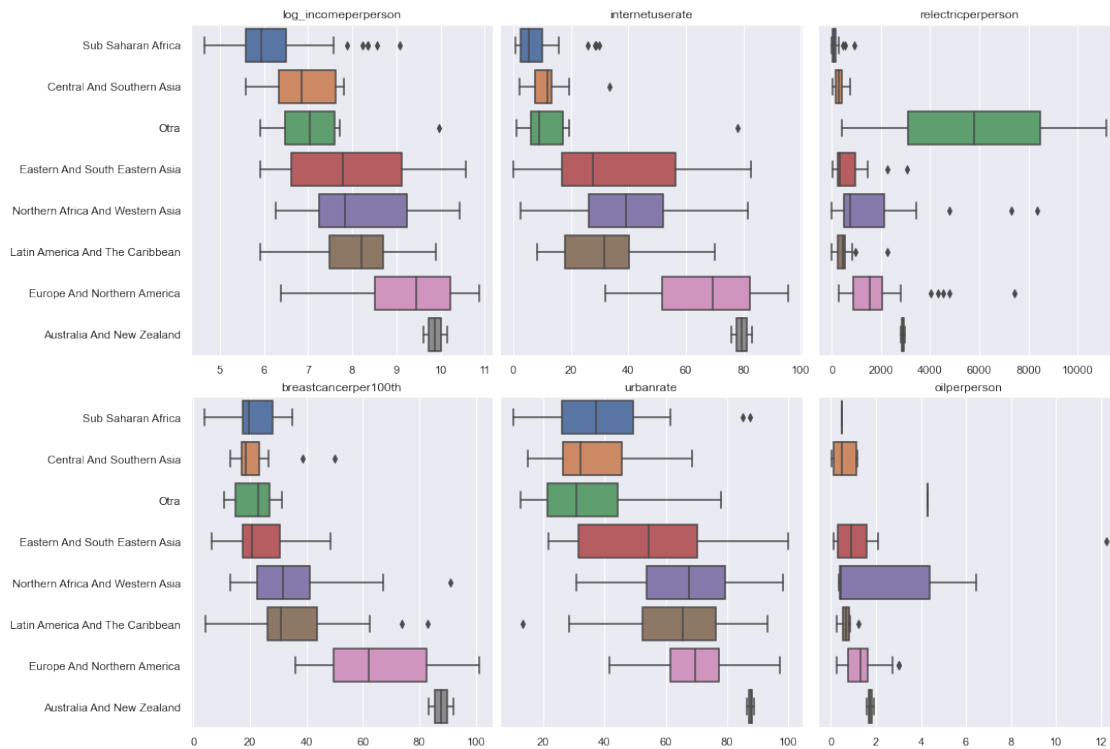
```
self._figure.tight_layout(*args, **kwargs)
```

```
C:\Users\yacha\AppData\Local\Programs\Python\Python39\lib\site-
packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to
tight
```

```
self._figure.tight_layout(*args, **kwargs)
```

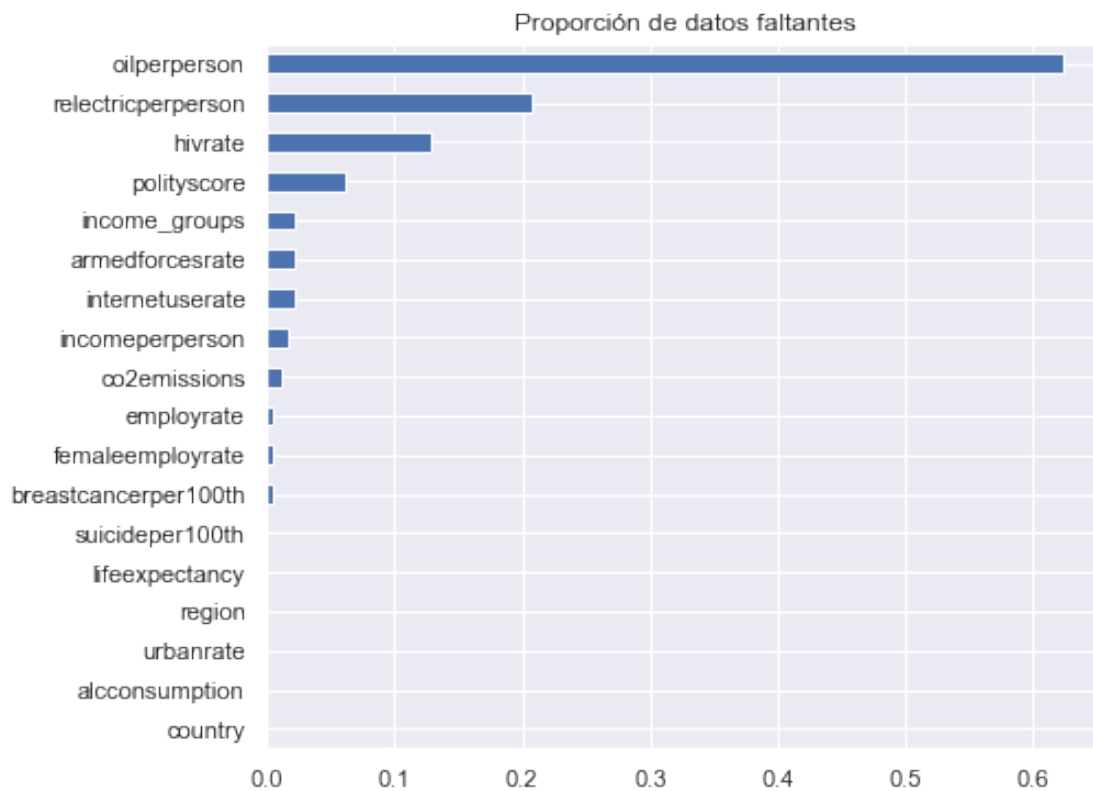
3.1 Comportamiento de los indicadores por región



En la gráfica anterior podemos observar que la región es una variable que se relaciona con todos los indicadores seleccionados. En particular es posible observar que Australia, Nueva Zelanda, Europa y Norte América tienden a tener un desempeño más alto. Aunque en los grupos se encuentran países atípicos, como es el caso de los países con relativamente altos ingresos en África

4 2. Preparación de datos

4.1 Completitud (Datos Faltantes)



	Faltantes
country	0.000000
alcoholconsumption	0.000000
urbanrate	0.000000
region	0.000000
lifeexpectancy	0.000000
suicideper100th	0.000000
breastcancerper100th	0.005618
femaleemployrate	0.005618
employrate	0.005618
co2emissions	0.011236
incomeperperson	0.016854
internetuserate	0.022472
armedforcesrate	0.022472
income_groups	0.022472
polityscore	0.061798
hivrate	0.129213
relectricperperson	0.207865

oilperperson 0.623596

Como se puede observar en la gráfica anterior, se encuentran bastantes datos faltantes, principalmente en las variables `oilperperson` (62.4%), `relectricperperson` (20.8%) y `hivrate` (12.9%).

Consistencia : No se identifican problemas de consistencia en la medida en que las variables incluidas no permiten establecer contradicciones entre sí.

Procesabilidad : Los datos son procesables en la medida en que todos se encuentran en un formato que permite su uso inmediato.

Disponibilidad : Los datos se encuentran disponibles y dado que se trata de un análisis puntual no existe un problema en términos de la latencia para acceder a los mismos. Sin embargo, los tiempos de actualización de los datos implican que no es fácil mantener estos análisis actualizados en caso de que se requiera.

Temporalidad : De acuerdo al diccionario es posible observar que no todas las variables están actualizadas al mismo periodo, sin embargo, se considera que esto puede no ser un problema mayor en la medida en que no se esperan cambios abruptos en estos indicadores, aunque es algo a tener en cuenta y que no se puede descartar. Adicionalmente el dataset es relativamente viejo 2007-2011, por lo que algunas de las relaciones que se identifiquen pueden ya no mantenerse.

Credibilidad : De acuerdo con el portal GapMinder: “Most of our data are not good enough for detailed numeric analysis” ([enlace](#)). Por lo que es posible que las fuentes de información puedan tener algunos problemas de credibilidad, pues los datos son recolectados de fuentes diversas y GapMinder intenta llenar los vacíos que encuentra.

Complejidad : Se identifican datos faltantes en varias variables (ver más arriba) por lo que será necesario eliminar algunas de ellas o hacer imputaciones en los casos en que el porcentaje de datos faltantes no es tan elevado.

4.2 Remediación de datos faltantes

De las variables consideradas claves se propone lo siguiente para remediar los problemas de complejidad:

- **incomeperperson**: Eliminar las filas con datos faltantes, dado que esta es la variable que se pretende explicar.
- **oilperson**: Eliminar esta columna, dado que el porcentaje de datos faltantes es demasiado alto.
- **relectricperperson, breastcancerper100th, internetuserate**: imputar estos valores con la mediana de los países de la misma región y grupo de ingresos. Se utilizará la mediana, dado que se observan algunos valores extremos en algunas de estas variables y se quiere utilizar un método robusto.

log_incomeperperson	0
internetuserate	3
relectricperperson	38
breastcancerper100th	1
urbanrate	0
oilperperson	109

dtype: int64

Antes de avanzar con la imputación separamos los datos entre test y train, para garantizar que las estadísticas que se utilicen en la imputación no utilicen los datos de test.

		median_internetuserate \
region	income_groups	
Australia And New Zealand	high_income	83.002584
	low_income	9.739743
	lower_middle_income	13.598876
	upper_middle_income	13.000111
Eastern And South Eastern Asia	high_income	77.638535
	lower_middle_income	18.949975
	upper_middle_income	45.338912
Europe And Northern America	high_income	75.873096
	lower_middle_income	42.353795
	upper_middle_income	40.020095
Latin America And The Caribbean	high_income	45.000000
	low_income	8.370207
	lower_middle_income	11.090765
	upper_middle_income	32.934351
Northern Africa And Western Asia	high_income	59.206265
	low_income	16.506453
	lower_middle_income	36.562553
	upper_middle_income	31.004378
Otra	lower_middle_income	1.280050
	upper_middle_income	14.830736
Sub Saharan Africa	low_income	2.450362
	lower_middle_income	9.837459
	upper_middle_income	17.616353
		median_relectricperperson
region	income_groups	
Australia And New Zealand	high_income	2993.092660
	low_income	218.659975
	lower_middle_income	164.774591
	upper_middle_income	618.654395
Eastern And South Eastern Asia	high_income	2257.572233
	lower_middle_income	302.725655
	upper_middle_income	603.780028
Europe And Northern America	high_income	1736.697704
	lower_middle_income	489.262488
	upper_middle_income	614.907287
Latin America And The Caribbean	high_income	532.515177
	low_income	9.192395
	lower_middle_income	256.099151

	upper_middle_income	457.644842
Northern Africa And Western Asia	high_income	2233.099055
	low_income	472.544461
	lower_middle_income	209.094517
	upper_middle_income	661.619794
Otra	lower_middle_income	NaN
	upper_middle_income	NaN
Sub Saharan Africa	low_income	38.222943
	lower_middle_income	72.225843
	upper_middle_income	NaN
median_breastcancerper100th \		
region	income_groups	
Australia And New Zealand	high_income	
91.90		
Central And Southern Asia	low_income	
17.50		
	lower_middle_income	
17.30		
	upper_middle_income	
17.50		
Eastern And South Eastern Asia	high_income	
32.70		
	lower_middle_income	
16.20		
	upper_middle_income	
24.75		
Europe And Northern America	high_income	
77.35		
	lower_middle_income	
44.75		
	upper_middle_income	
44.30		
Latin America And The Caribbean	high_income	
43.90		
	low_income	
4.40		
	lower_middle_income	
23.90		
	upper_middle_income	
32.75		
Northern Africa And Western Asia	high_income	
79.00		
	low_income	
39.95		
	lower_middle_income	
22.50		

31.70	upper_middle_income	
Otra	lower_middle_income	
17.30		
	upper_middle_income	
31.20		
Sub Saharan Africa	low_income	
18.25		
	lower_middle_income	
20.05		
	upper_middle_income	
28.15		
		median_urbanrate
region	income_groups	
Australia And New Zealand	high_income	86.56
Central And Southern Asia	low_income	21.85
	lower_middle_income	34.48
	upper_middle_income	63.20
Eastern And South Eastern Asia	high_income	66.48
	lower_middle_income	27.84
	upper_middle_income	56.73
Europe And Northern America	high_income	73.56
	lower_middle_income	54.87
	upper_middle_income	71.10
Latin America And The Caribbean	high_income	88.44
	low_income	46.84
	lower_middle_income	56.74
	upper_middle_income	72.95
Northern Africa And Western Asia	high_income	80.78
	low_income	42.43
	lower_middle_income	43.44
	upper_middle_income	68.68
Otra	lower_middle_income	12.54
	upper_middle_income	52.36
Sub Saharan Africa	low_income	33.07
	lower_middle_income	45.37
	upper_middle_income	39.66

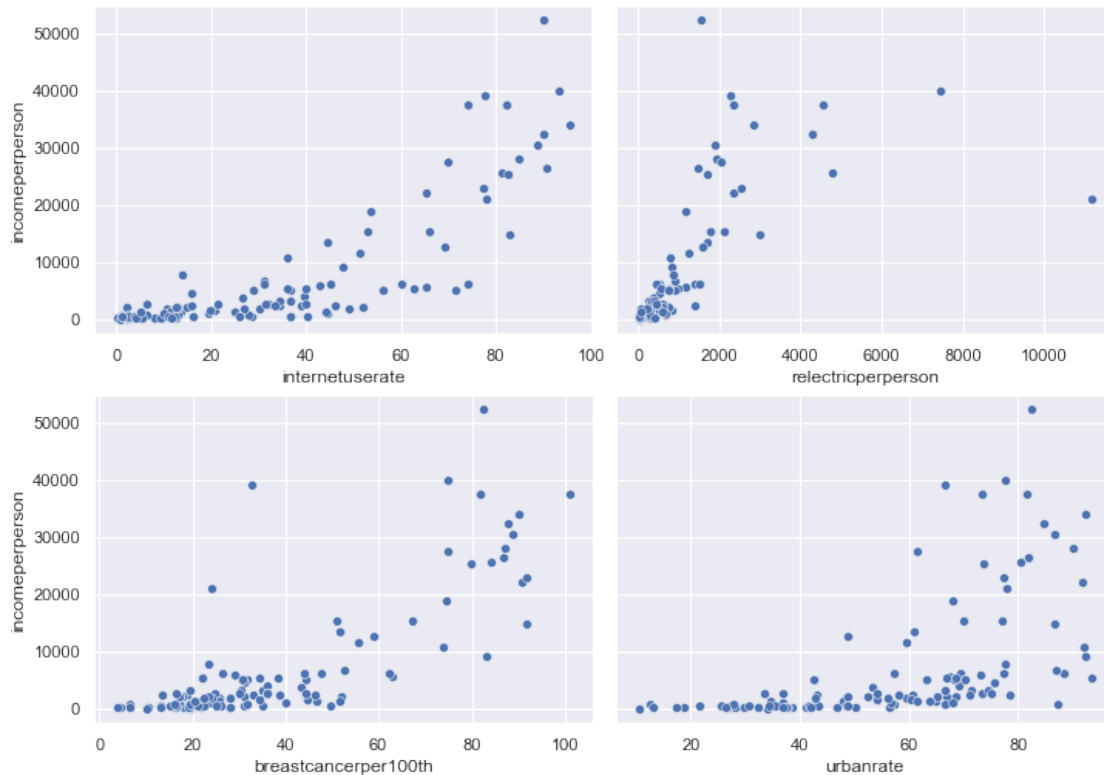
En los casos en que no se contó con un valor en la misma región y grupo de ingresos se imputó el valor utilizando únicamente el nivel de ingresos:

5 Entrenamiento del modelo de regresión:

5.1 Codificación de variable categórica (Región)

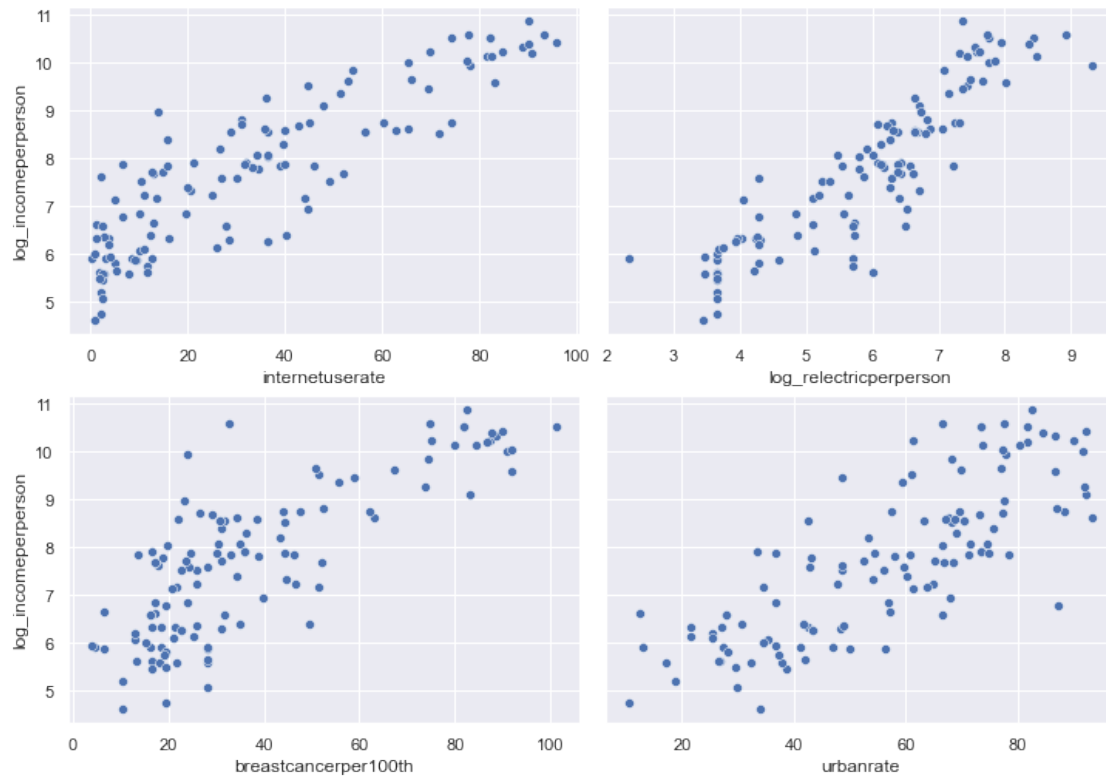
Dado que la región es una variable categórica se utiliza OneHot Encoder para codificarla:

5.2 Verificando si la relación es aproximadamente lineal



De la gráfica anterior se puede inferir que la relación no es realmente lineal. De hecho en algunos de los casos la relación parece exponencial. Por ello se propone aplicar algunas transformaciones a los datos:

- incomeperperson \rightarrow log
- relectricperperson \rightarrow log



Con estas transformaciones observamos ahora unas relaciones aproximadamente lineales

5.3 Entrenamiento

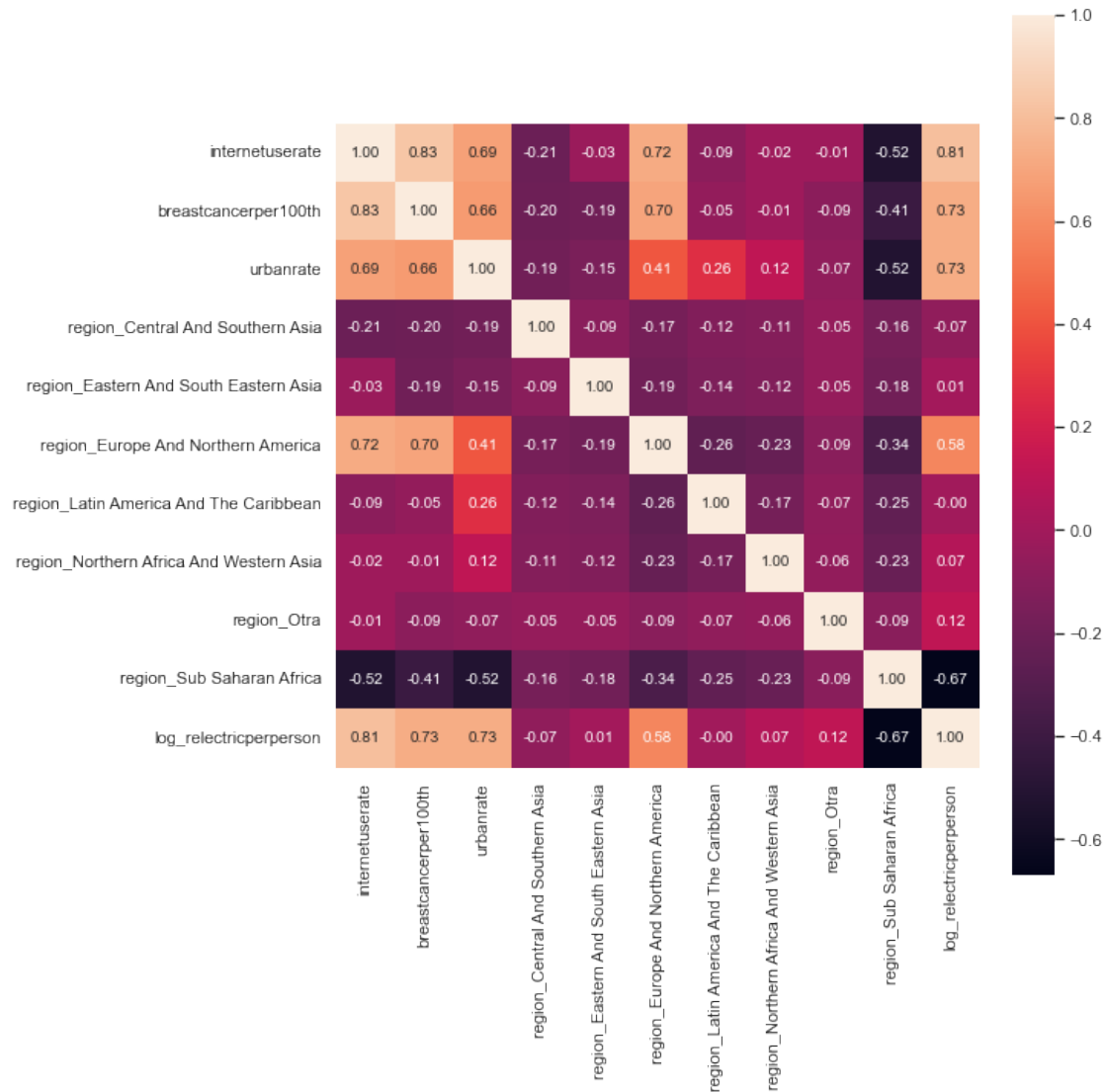
5.4 Evaluación

MAE: 3256.652934975392
 RMSE: 5981.011550765352
 R^2 : 0.7097036385819444

En las métricas anteriores se observa un error relativamente grande, lo cual es de esperarse en la medida en que el PIB es una variable compleja pero probablemente este influenciada por muchos otros factores, sin embargo, el R^2 parece relativamente alto para la sencillez del modelo que se está ajustando.

5.5 Verificación de supuestos

5.5.1 No Multicolinealidad



Correlación

internetuserate

0.834875

breastcancerper100th

0.834875

log_relectricperperson

0.809913

internetuserate

0.809913

log_relectricperperson

breastcancerper100th

internetuserate

internetuserate

log_relectricperperson

urbanrate

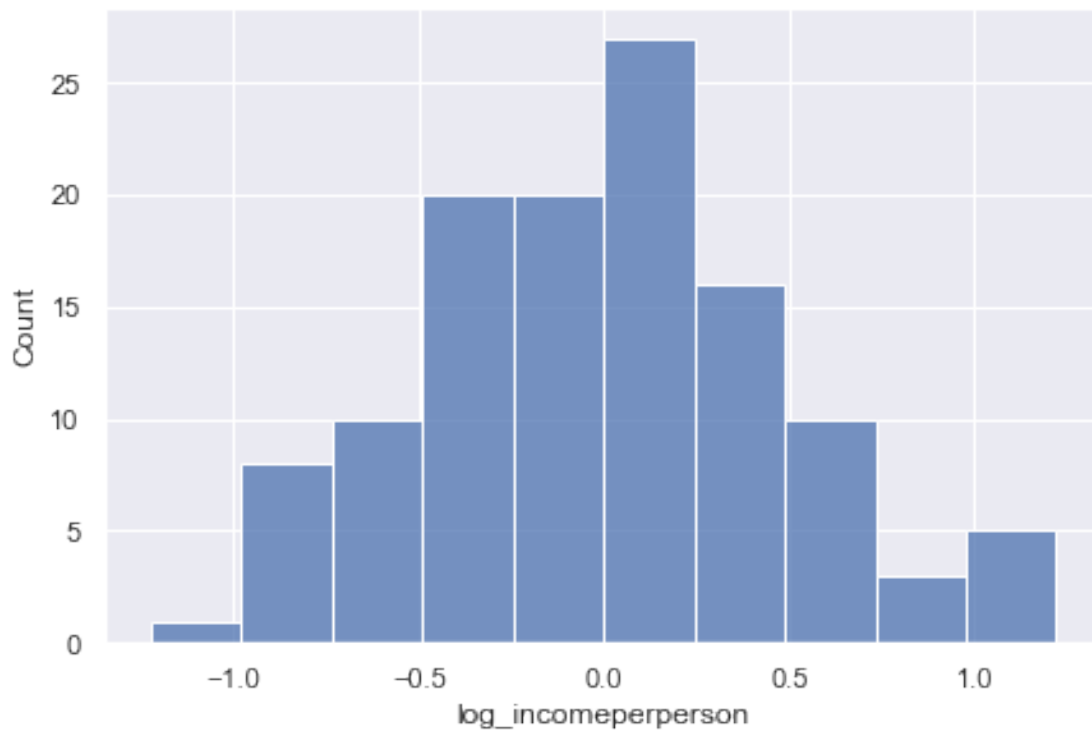
```

0.729658
urbanrate                                log_relectricperperson
0.729658
breastcancerper100th                    log_relectricperperson
0.728971
log_relectricperperson                    breastcancerper100th
0.728971
internetuserate                          region_Europe And Northern America
0.720756
region_Europe And Northern America internetuserate
0.720756

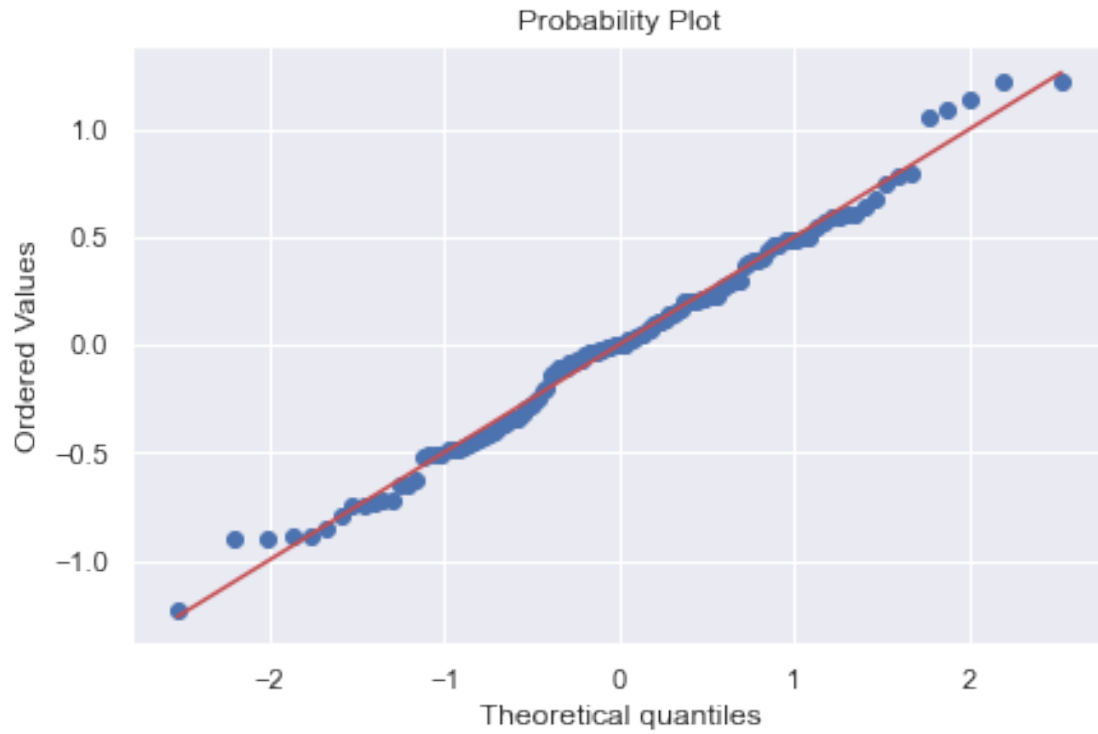
```

Aunque existe correlación entre los regresores el peor de los casos ocurre entre `internetuserate` y `breastcancerper100th` con un coeficiente de correlación de 0.83

5.6 Normalidad de los errores

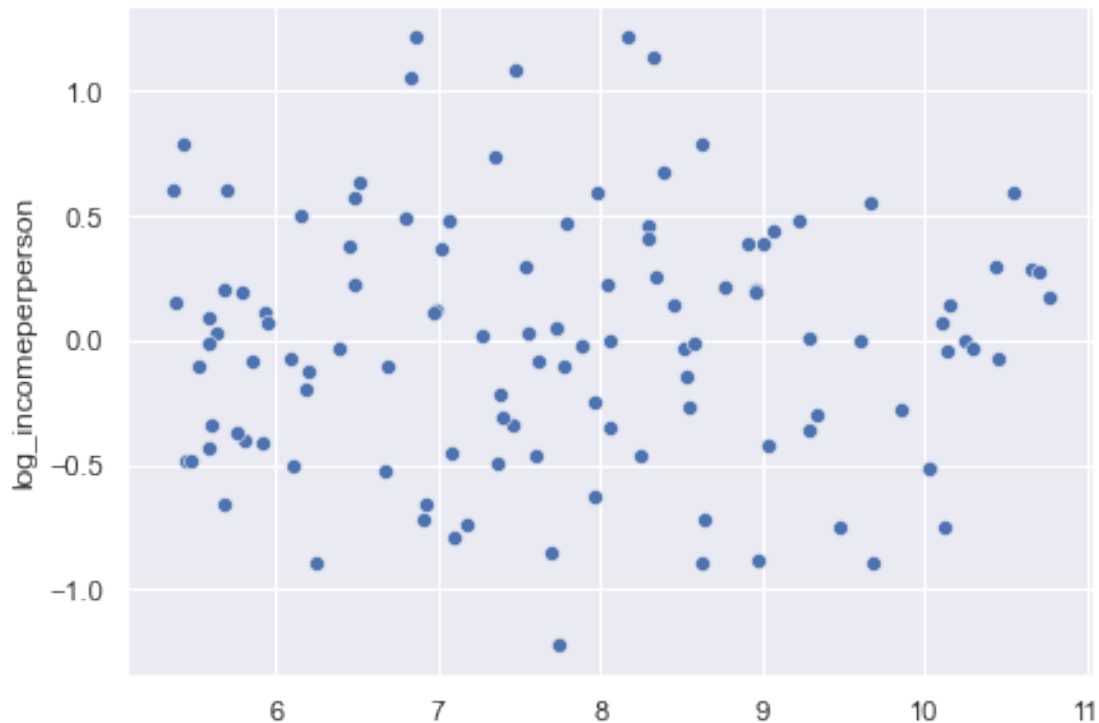


```
Jarque_beraResult(statistic=0.41241489461691916, pvalue=0.8136642663119029)
```



De la las gráficas anteriores concluimos que los errores siguen una distribución aproximadamente normal, lo cual podemos también confirmarlo a través del test de Jarque-Bera, que arroja un p-valor de 0.81, con lo cual podemos concluir que no hay evidencia para rechazar la hipótesis de que los errores provienen de una distribución normal.

6 Homoscedasticidad



En la gráfica anterior no se observa una cambio significativo en la varianza de los errores para diferentes nivel de la variable predicha, por lo cual los errores parecen se homoscedasticos.

6.1 Interpretación de los coeficientes

	Coeficientes
constante	2.141113
internetuserate	0.020817
breastcancerper100th	0.010168
urbanrate	0.012433
region_Central And Southern Asia	1.214445
region_Eastern And South Eastern Asia	1.174415
region_Europe And Northern America	0.816835
region_Latin America And The Caribbean	1.412258
region_Northern Africa And Western Asia	0.978910
region_Otra	1.233913
region_Sub Saharan Africa	1.136813
log_relectricperperson	0.464915

Dado que la función que se ajustó tiene transformaciones logarítmicas los coeficientes no pueden interpretarse directamente. Recordemos que la función que se ha ajustado es

$$\ln(y) = w_0 + w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + w_4 * x_4 + w_5 * x_5 + w_6 * x_6 + w_7 * x_7 + w_8 * x_8 + w_9 * x_9 + w_{10} * x_{10} + w_{11} * \ln(x_{11})$$

6.2 Variables lineales (log-lin):

En estas variables podemos interpretar que un aumento de una unividad en el regresor se relaciona con un incremento de $(\exp(w_i) - 1)\%$ en el ingreso per cápita. Esto se puede probar de la siguiente manera:

$$\ln(y) = w_i x_i + \dots$$

Despejando y :

$$y = \exp(w_i x_i) \exp(\dots)$$

Si se incrementa x_i en una unidad:

$$y' = \exp(w_i(x_i + 1)) \exp(\dots) = \exp(w_i) \exp(w_i x_i) \exp(\dots) = \exp(w_i) y$$

Luego

$$\frac{y' - y}{y} = \exp(w_i) - 1$$

De manera que por cada unidad en que se incrementen estas variables, el incremento en el ingreso sería:

	Coeficientes
internetuserate	0.021035
breastcancerper100th	0.010220
urbanrate	0.012510

De manera que: * Un incremento en el use de internet en una de cada 100 personas implicaría un aumento del 2.1% en el ingreso per cápita. * Un incremento en la identificación de casos de cancer de mama en una de cada 100.000 mujeres se relaciona con un incremento del 1.0% del ingreso per cápita. * Un aumento en una de cada cien personas en zonas urbanas implica un incremento de 1.3% en el PIB per cápita.

En todo caso, hay que tener cuidado en el sentido de que correlación no implica causalidad. Es posible por ejemplo, que el aumento del PIB lleve a qué mas personas utilicen internet y no al contrario. O que un mayor ingreso implica mayor detección de los casos de cancer, 'no tendría mucho sentido concluir que para aumentar el ingreso es necesario aumentar las personas con cancer de mama.

6.3 Variables lineales (log-log):

En este caso podemos interpretar que un aumento del $\delta\%$ en el regresor se relaciona con un incremento del $((1 + \delta)^{w_i})\%$ en la variable objetivo. Esto se puede probar de la siguiente manera:

$$\ln(y) = w_i \ln(x_i) + \dots$$

Despejando y :

$$y = \exp(w_i \ln(x_i)) \exp(\dots) = x_i^{w_i} \exp(\dots)$$

Si se incrementa x_i en un 1%:

$$y' = (x_i * (1 + \delta))^{w_i} \exp(\dots) = (1 + \delta)^{w_i} y$$

$$\frac{y' - y}{y} = (1 + \delta)^{w_i} - 1$$

De manera que por cada 1% en que se incremente esta variable, el incremento en el ingreso sería:

Coeficientes	
relectricperperson	0.004637

Luego un incremento del 1% en el uso de electricidad por persona se relaciona con un incremento del 0.5% en el ingreso per cápita.

7 Generación de resultados

7.1 Principales hallazgos

De acuerdo con los resultados de la regresión aplicada se encontró que: la detección de cancer de seno, el uso de internet, el porcentaje de población urbana y el uso de energía por persona resultan buenos predictores del PIB per cápita. En particular se encontró que un incremento del 1% en el uso de internet se relaciona con un incremento del 2.1% en el PIB per cápita; un aumento en la detección de cancer en una de cada 100.000 mujer se relaciona con un incremento del PIB del 1% (lo cual podría interpretarse como una proxy de la calidad del sistema de salud del país); un aumento de un punto en la tasa de población urbana se relaciona con un incremento del 1.3% del PIB per capita, y un incremento del 1% en el uso de energía con un incremento del 0.5%.

¿qué conjunto de políticas públicas recomendaría implementar, a partir de la premisa de que la mejora en estas áreas indicaría al Banco Mundial que el país es estable, está en una trayectoria de desarrollo sostenible y tiene la capacidad de administrar y reembolsar préstamos de manera efectiva?

Estas conclusiones deben ser tomadas con cuidado, pues la regresión no explica la causalidad entre los fenómenos, por lo que puede ser que el incremento del PIB lleve a aumentar estos indicadores y no al contrario. Sin embargo, de estos resultados sería razonable esperar que los siguientes tipos de políticas puedan llevar a un aumento del PIB per cápita:

- Inversión para mejorar el sistema de salud.
- Aumento de la cobertura de internet.
- Reducción del costo de la energía eléctrica.
- Mejoras en el desarrollo urbano y migración campo-ciudad.

De todas estas políticas, la cobertura de internet parece tener el mayor efecto por lo que podría ser una política estrella, en particular si puede resultar relativamente menos costos incrementar la cobertura de internet que reducir los costos de la energía o mejorar el sistema de salud.