



---

# VALORACIÓN DE INVERSIONES INMOBILIARIAS

---

Proyecto fin de curso Data & Desarrollo en Python



AUTORA: SANDRA REQUENA FILIU

Contenido

estudioDatos ..... 1

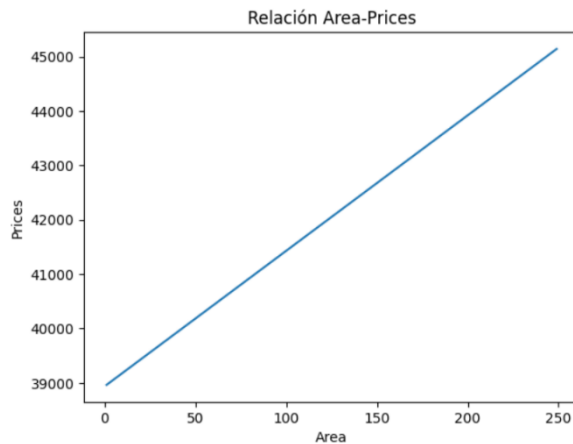
regresionLineal ..... 5

testing..... 6

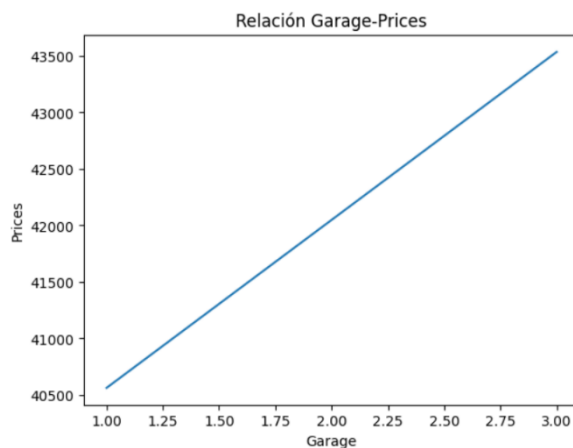
## estudioDatos

En este proyecto se ha realizado una regresión lineal con los datos de 500000 casas facilitados en un dataset. Para cada casa, se indican los valores de las siguientes características:

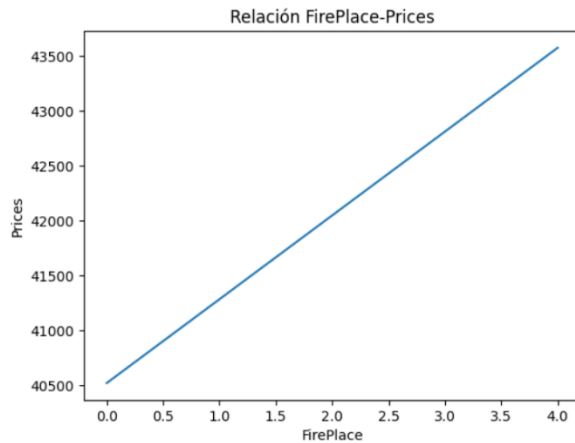
- **Area:** Variable numérica que indica el área de la casa. Tras realizar una regresión lineal, se puede afirmar que el valor de esta característica es directamente proporcional al precio:



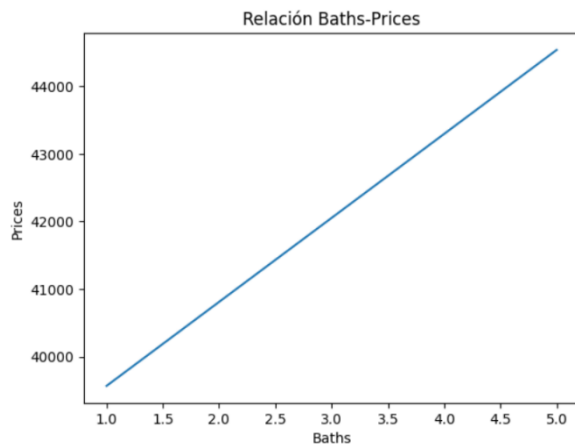
- **Garage:** Variable numérica que indica el número de garajes que tiene la propiedad. Tras realizar una regresión lineal, se puede afirmar que el valor de esta característica es directamente proporcional al precio:



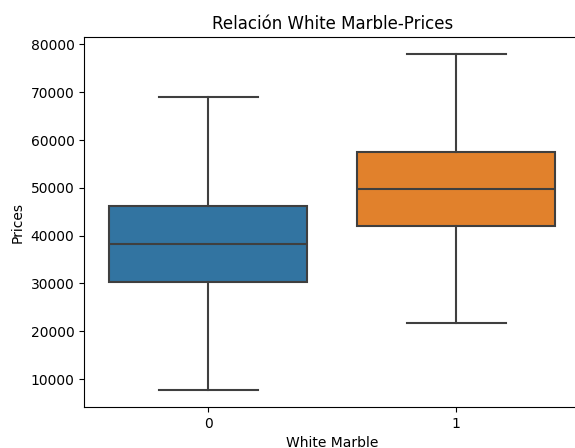
- **FirePlace:** Variable numérica que indica el número de chimeneas presentes en la propiedad. Tras realizar una regresión lineal, se puede afirmar que el valor de esta característica es directamente proporcional al precio:



- **Baths:** Variable numérica que indica el número de baños presentes en la propiedad. Tras realizar una regresión lineal, se puede afirmar que el valor de esta característica es directamente proporcional al precio:

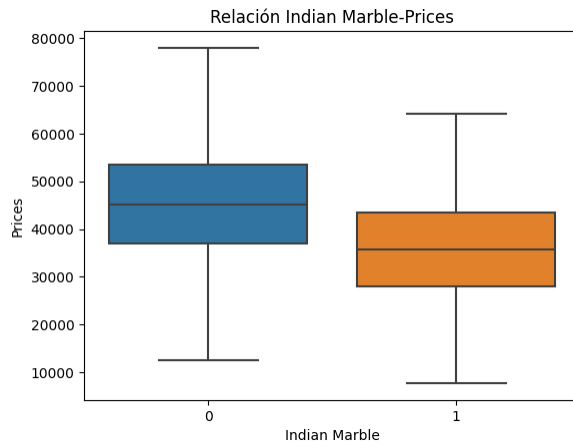


- **White Marble:** Variable categórica que indica la presencia de mármol blanco en la propiedad. Según el siguiente diagrama de cajas, la presencia de este tipo de mármol aumenta el valor de la propiedad:

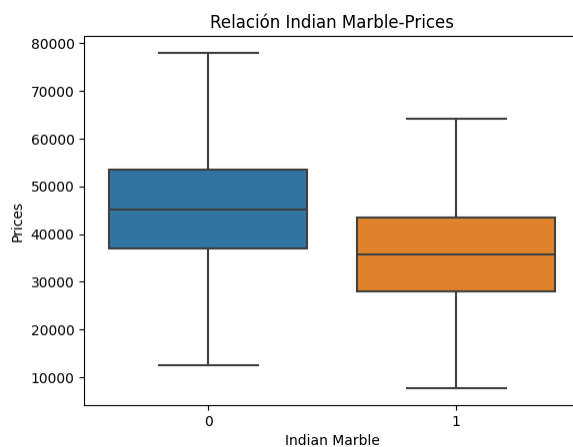


- **Black Marble:** Variable categórica que indica la presencia de mármol negro en la propiedad. Como no se podía distinguir la relación entre esta característica y el precio en el diagrama de cajas, se ha decidido calcular la pendiente de su regresión lineal. La pendiente es -2006.03, por lo que puede afirmar que la presencia de este tipo de mármol disminuye el valor de la propiedad.

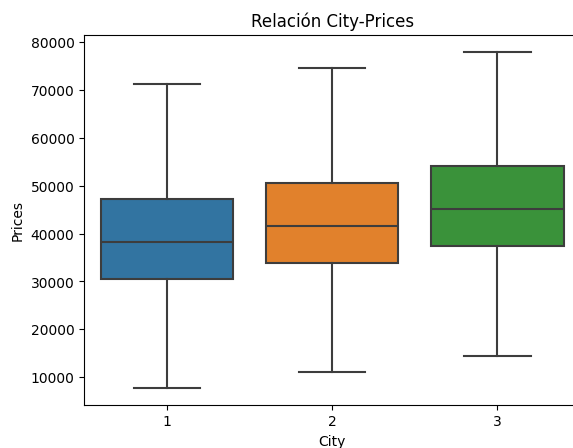
- **Indian Marble:** Variable categórica que indica la presencia de mármol negro en la propiedad. Según el siguiente diagrama de cajas, la presencia de este tipo de mármol disminuye el valor de la propiedad:



- **Floors:** Variable categórica que indica si la propiedad tiene más de una planta. Según el siguiente diagrama de cajas, el número de plantas es directamente proporcional al precio:



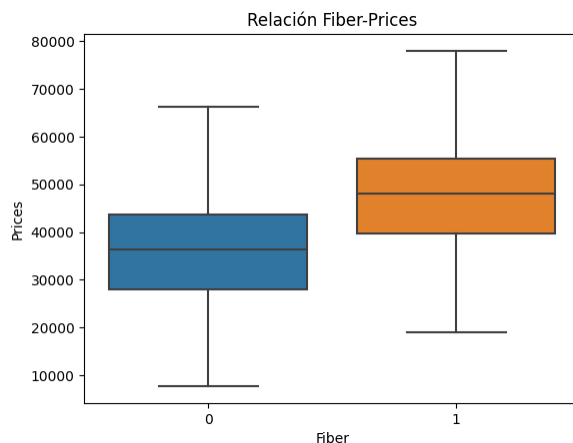
- **City:** Variable categórica que indica la ciudad de la propiedad. Se puede ver el rango de precios en cada ciudad en el siguiente diagrama de cajas:



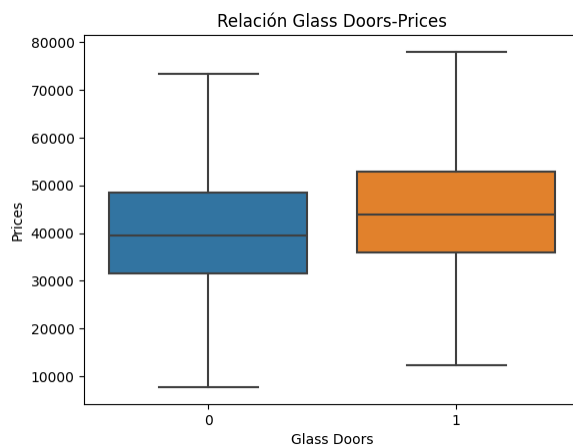
- **Solar:** Variable categórica que indica si la propiedad tiene placas solares. Como no se podía distinguir la relación entre esta característica y el precio en el diagrama de cajas, se ha decidido calcular la pendiente de su regresión lineal.

La pendiente es 204.15, por lo que se puede afirmar que la presencia de placas solares aumenta el valor de la propiedad.

- **Electric:** Variable categórica que indica si la propiedad tiene instalación eléctrica. Como no se podía distinguir la relación entre esta característica y el precio en el diagrama de cajas, se ha decidido calcular la pendiente de su regresión lineal. La pendiente es 1270.18, por lo que se puede afirmar que la presencia de instalación eléctrica aumenta el valor de la propiedad.
- **Fiber:** Variable categórica que indica si la propiedad tiene fibra óptica. Según el siguiente diagrama de cajas, su presencia es directamente proporcional al precio:



- **Glass Doors:** Variable categórica que indica si la propiedad tiene puertas de cristal. Según el siguiente diagrama de cajas, su presencia es directamente proporcional al precio:



- **Swimming Pool:** Variable categórica que indica si la propiedad tiene piscina. Como no se podía distinguir la relación entre esta característica y el precio en el diagrama de cajas, se ha decidido calcular la pendiente de su regresión lineal. La pendiente es 43.27, por lo que se puede afirmar que la presencia de piscina aumenta el valor de la propiedad.
- **Garden:** Variable categórica que indica si la propiedad tiene jardín. Como no se podía distinguir la relación entre esta característica y el precio en el diagrama de cajas, se ha decidido calcular la pendiente de su regresión lineal. La

pendiente es 37.30, por lo que se puede afirmar que la presencia de jardín aumenta el valor de la propiedad.

Todas las casas se han valorado para obtener su precio, por lo que podemos usar sus datos para construir un modelo de regresión lineal para estimar el precio de otras propiedades. Esta estimación se puede comparar con el precio de compra de la vivienda y, si esta estimación es mayor, se recomienda invertir en ese inmueble.

## regresionLineal

Como se ha mencionado anteriormente, se quiere construir un modelo de regresión lineal con la siguiente estructura:

$$\begin{aligned} \text{Price} = & m_A \cdot \text{Area} + m_{Gara} \cdot \text{Garage} + m_{Fir} \cdot \text{FirePlace} + m_B \cdot \text{Baths} + m_{WM} \\ & \cdot \text{White\_Marble} + m_{BM} \cdot \text{Black\_Marble} + m_{IM} \cdot \text{Indian\_Marble} \\ & + m_{Fl} \cdot \text{Floors} + m_C \cdot \text{City} + m_S \cdot \text{Solar} + m_E \cdot \text{Electric} + m_{Fib} \\ & \cdot \text{Fiber} + m_{GD} \cdot \text{Glass\_Doors} + m_{SP} \cdot \text{Swiming\_Pool} + m_{Gard} \\ & \cdot \text{Garden} + a \end{aligned}$$

La regresión lineal del precio frente al conjunto de características dio como resultado las siguientes pendientes y la siguiente ordenada en el origen:

$m_A$	$m_{Gara}$	$m_{Fir}$	$m_B$	$m_{WM}$	$m_{BM}$	$m_{IM}$	$m_{Fl}$
$2.50 \cdot 10^1$	$1.50 \cdot 10^3$	$7.50 \cdot 10^2$	$1.25 \cdot 10^3$	$7.67 \cdot 10^3$	$-1.33 \cdot 10^3$	$-6.33 \cdot 10^3$	$1.50 \cdot 10^4$

$m_C$	$m_S$	$m_E$	$m_{Fib}$	$m_{GD}$	$m_{SP}$	$m_G$	$a$
$3.50 \cdot 10^3$	$2.50 \cdot 10^2$	$1.25 \cdot 10^3$	$1.18 \cdot 10^4$	$4.45 \cdot 10^3$	$-1.15 \cdot 10^{-13}$	$4.96 \cdot 10^{-12}$	$7.33 \cdot 10^3$

Como se puede ver, las pendientes de las características *Swiming Pool* y *Garden* son tan pequeñas que se pueden despreciar, por lo que los datos quedarían ahora de la siguiente manera:

$m_A$	$m_{Gara}$	$m_{Fir}$	$m_B$	$m_{WM}$	$m_{BM}$	$m_{IM}$
25	1500	750	1250	7667	-1333	-6333

$m_{Fl}$	$m_C$	$m_S$	$m_E$	$m_{Fib}$	$m_{GD}$	$a$
15000	3500	250	1250	11750	4450	7333

Con estos datos, se calcula la pendiente y el precio de la siguiente manera:

$$\begin{aligned} \text{pendiente} = & m_A \cdot \text{Area} + m_{Gara} \cdot \text{Garage} + m_{Fir} \cdot \text{FirePlace} + m_B \cdot \text{Baths} \\ & + m_{WM} \cdot \text{White\_Marble} + m_{BM} \cdot \text{Black\_Marble} + m_{IM} \\ & \cdot \text{Indian\_Marble} + m_{Fl} \cdot \text{Floors} + m_C \cdot \text{City} + m_S \cdot \text{Solar} + m_E \\ & \cdot \text{Electric} + m_{Fib} \cdot \text{Fiber} + m_{GD} \cdot \text{Glass\_Doors} \end{aligned}$$

$$\text{precio\_estimado} = \text{pendiente} + a$$

## testing

Se ha creado un dataset con 5 casas, cuyos datos se han generado aleatoriamente usando como referencia los valores mínimos y máximos de cada columna. El dataset creado es el siguiente:

	Area	Garage	FirePlace	Baths	White Marble	Black Marble	Indian Marble	Floors	City	Solar	Electric	Fiber	Glass Doors	Prices
0	116	1	0	4	0	1	1	0	1	1	1	0	1	46522
1	41	2	1	5	0	0	0	0	2	1	1	0	1	24357
2	19	2	4	1	1	0	1	1	2	1	1	0	1	40872
3	226	3	2	2	1	1	1	1	3	1	0	0	0	70577
4	55	2	3	5	0	0	0	1	1	1	1	0	1	10699

Con las fórmulas anteriores, se ha calculado el precio estimado, el cual se ha comparado con los precios de venta para valorar si es buena idea invertir en esa propiedad. El resultado para las casas anteriores es el siguiente:

	Prices	Precios valorados	Buena inversión
0	46522	18516	False
1	24357	31308	True
2	40872	44341	True
3	70577	47233	False
4	10699	44658	True