

Valoración de inversiones inmobiliarias

Proyecto fin de curso Data & Desarrollo en Python



Autora: Sandra requena Filiu

Contenido

[estudioDatos 1](#_Toc146619101)

[regresionLineal 5](#_Toc146619102)

[testing 6](#_Toc146619103)

# 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:

Gráfico, Gráfico de líneas

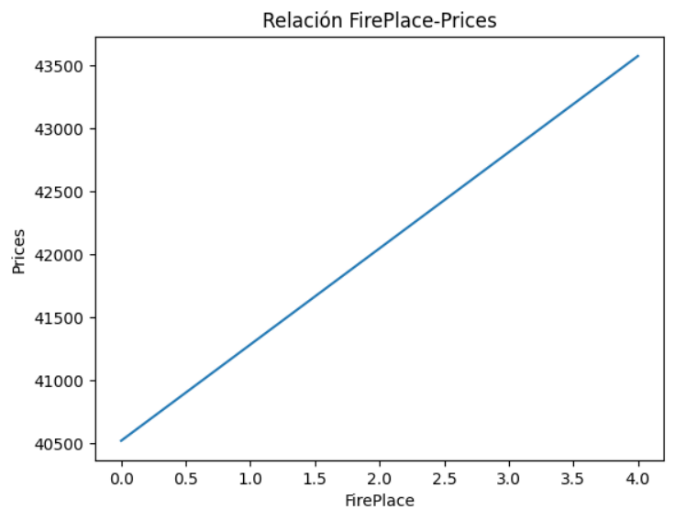
Descripción generada automáticamente

* **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:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

* **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:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

* **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:

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

* **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:

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

* **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:

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

* **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:

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

* **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:

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

* **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:

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

* **Swiming 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:

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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  | - | -6 |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  | - |  |  |

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:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| 25 | 1500 | 750 | 1250 | 7667 | -1333 | -6333 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| 15000 | 3500 | 250 | 1250 | 11750 | 4450 | 7333 |

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

# 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:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

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:

