

Problem set 3

Predicción de los precios inmobiliarios en el mercado caleño

Introducción

El mercado inmobiliario en Cali no consta de una buena cantidad de información para poder establecer los lineamientos económicos para que las nuevas empresas incursiones sin llegar a bancarrotas al no saber las limitaciones propias del mercado. Dado lo anterior, es inminente lograr predecir de forma correcta los precios de los inmuebles en Cali para que las nuevas empresas puedan establecerse en el mercado con el menor riesgo posible. Por tanto, haciendo uso de la amplia información existente sobre este mercado en Medellín y Bogotá, ciudades similares, incluyendo la idea base de bienes diferenciados de Rosen (1974), se estiman predicciones que le permitan a la empresa interesada en incursionar en este mercado la información necesaria para que su planeación financiera y económica minimice lo más posible el riesgo.

Ahora bien, aunque se encuentren desventajas en nuestro modelo como la posible omisión de variables que puedan llegar a ayudar a predecir el precio de los inmuebles, se tienen unas mayores ventajas como el valor de MSE el cual es, en relación a los otros modelos, mínimo, debido a la fuerte relación entre las variables independientes que si fueron seleccionadas y la variable a predecir. La anterior relación es la que permite entender que las predicciones son lo más acertadas posible y que, con base en nuestros resultados, la empresa en cuestión puede organizarse para entrar al mercado. Para lograr la mejor predicción, en nuestra metodología no solo relacionamos las variables explicativas con la predicción de los precios inmobiliarios por medio de modelo tradicional, sino que realizamos los diferentes y respectivos procedimientos y modelajes para encontrar el modelo con mayor rendimiento esperado.

Para esto se modeló de forma tradicional, por medio de Lasso y Ridge, Elastic Net y, por último, con Superlearner. A pesar de utilizar 5 modelos distintos, una desventaja es, dado los equipos de investigación, se hizo imposible realizar una metodología de predicción de Elastic Net. No obstante, se logró obtener el MSE de este tipo de modelaje, pero al no ofrecer un mejor rendimiento del que ofrece el Superlearner, no es esencial las predicciones del Elastic Net. Por tanto, nuestros resultados están basados en el modelo funcional que mejor predice los precios de los inmuebles para Cali.

Datos

Para la metodología que utilizaremos luego de este apartado, se tienen datos obtenidos de OpenStreetMap que se son relevantes para entender cómo se fijan los precios de los inmuebles en términos generales. Estas variables son escogidas debido a que “las propiedades que ofrecen el mayor número de características deseadas por los compradores tendrán precios más altos, y las propiedades con menos características tendrán precios más bajos” (Robinson 1979) (Sheppard 1999). En primer lugar, una de estas características es la superficie total de la propiedad, entendiendo que por metros cuadrados es razonable que el precio de un inmueble cambie. En suma, relacionado al tamaño de la propiedad, también está el tipo de inmueble y la cantidad de habitaciones, por lo que serán variables relevantes a tomar en cuenta de igual forma. En adición, que una propiedad tenga terraza o balcón es un claro valor agregado que vuelve más atractivo el producto, por lo tanto, claramente impacta el precio. Por último, se tiene que tener en cuenta la ubicación de la propiedad porque si a sus alrededores tienen facilidades de vida, su precio será mayor. Dado lo anterior, se tendrán como variables explicativas las distancias a los supermercados, estaciones de policía, paraderos de bus y colegios; entendiendo que son determinantes para la elección de vivienda, se entendería que de igual forma son determinantes para el precio.

Ahora bien, en los anexos se pueden ver las estadísticas descriptivas tanto para train como para test. En esta se puede resaltar, en primer lugar, que el precio promedio de inmueble de Bogotá ronda los 870 millones de pesos mientras que en Medellín ronda los 640 millones de pesos. Adicionalmente, en las tres ciudades existen más apartamentos que casas, pero de las 3 es Cali la ciudad más balanceada con un 56% de apartamentos y un 44% de casas. Bogotá cuenta con una relación 77/23 y Medellín 80/20. El tipo de vivienda se ve más emparejado cuando se refiere a si poseen balcón o no, en Cali la proporción es 50/50, muy similares a Bogotá y Medellín que no distan de dicha proporción. En el tema de distancias, Bogotá tiene mayores alrededores que facilitan la vida, seguida de Medellín y, detrás de ambas, Cali. Para finalizar, para representar como se diferencian las ciudades en las variables de interés, se realizó los gráficos referentes a las distancias con respecto a lo que denominamos alrededores con facilidad de vida (colegios, estaciones de bus y de policía, supermercados y oficinas). A través de estas visualizaciones, se puede ver estas facilidades de vida según la abundancia en cada ciudad; los mapas se encuentran en los anexos del presente artículo.

Metodología y resultados

Ahora bien, para poder predecir los precios de los inmuebles en la ciudad de Cali y ofrecer la información necesaria para que la empresa incursione en este mercado sin tanto riesgo, se realizó una metodología extensa para descartar diferentes modelajes y escoger el de mayor rendimiento posible. Para esto, se utilizaron nueve variables independientes referentes al tipo de vivienda, la cantidad de cuartos, la superficie total, si tiene o no balcón, la distancia a supermercados, estaciones de bus, estaciones de policías, oficinas y colegios en 5 modelaciones predictivas diferentes para el entrenamiento. Estas siendo, un tradicional OLS, un Lasso, un Ridge, un Elastic Net y un Superlearner. De estos tipos de modelos se procedería a escoger para la predicción final el modelo con menor MSE, es decir, el que presentara un mejor rendimiento a la hora de predecir. Una vez escogido el modelo ideal, se realizarían las predicciones esperando ningún problema de sobreestimación de precios para que la empresa este aconsejada de la mejor manera.

Con respecto a los resultados de los rendimientos, el modelo que menos rendimiento presenta es el modelo tradicional, lo cual es esperado siendo la modelación más sencilla y presentando un MSE de $6.148223e+17$. Seguido a este, se encuentran los modelos de Ridge y Lasso con MSE muy parecidos al del tradicional. En nuestra primera etapa de la metodología, el segundo mejor rendimiento lo presenta el modelo de Elastic Net con un MSE de $6.147405e+17$. No obstante, el modelo de SuperLearner tiene un mejor rendimiento, presentando un MSE mucho menor que el de sus contrapartes, siendo este de $2.537410e+17$. Por lo tanto, de los 5 modelos que se utilizan para entrenar la muestra, se escoge el SuperLearner para realizar las predicciones del test, es decir, encontrar los precios predichos de los inmuebles en Cali.

Entonces, se realiza el mismo procedimiento del entrenamiento con Superlearner, pero ahora en el test real. Es decir, se utilizan los datos existentes de Cali con las variables anteriormente seleccionadas para poder predecir los diferentes precios. Con los resultados del modelo se crea una nueva base de datos en formato csv que vendrá adjunta a este artículo. Con esta información, se le suministra todos los aspectos indispensables que le hace falta a la empresa para que pueda tomar la decisión correcta dentro de este mercado para no acarrear altos costos. Vale la pena hacer el hincapié de que solo se va a mostrar los precios que prediga el modelo de SuperLearner, ya que a través de la etapa de entrenamiento se corroboró que es la mejor opción posible para estos resultados.

Conclusiones y recomendaciones

Como se ha mencionado al principio, Cali es una ciudad con poca información respecto a los precios de los inmuebles, por lo que para una empresa de este mercado es sumamente difícil tomar decisiones con esta falta de información. Dado lo anterior, para lograr hacer una predicción correcta es necesario entrenar modelos con ciudades que se le asemejen como lo son Bogotá y Medellín que sí cuentan con la información necesaria, Sin embargo, para poder hacer esto es necesario entrenar diferentes modelos para poder escoger el más adecuado, o si no se corre el riesgo de dar resultados sobre o subestimados que lleven a malas decisiones a la empresa.

En este caso en específico, fue esencial entrenar los datos con 5 diferentes modelos (OLS, Ridge, Lasso, EN y SuperLearner) para encontrar el que mejor rendimiento tuviera. Una vez entrenados, se concluye que los modelos de OLS, Ridge, Lasso e incluso Elastic net tienen un rendimiento similar e inferior al modelo de Superlearner; por lo que este último fue el escogido para predecir. Sumado a esto, la elección de variables se realizó con precisión de igual forma para poder realizar una buena predicción. Por lo que, con apoyo de la literatura, se escogieron las variables más pertinentes que tengan que ver con aspectos obvios como la superficie, el tipo de la vivienda o la cantidad de cuartos, pero también aquellas que se denominaron alrededores de facilidad de vida. Lo anterior, se basó en el entendimiento de que el precio del inmueble no solo es afectado por condiciones intrínsecas de este, sino por alrededores donde este puede estar ubicado que facilitan la vida del propietario.

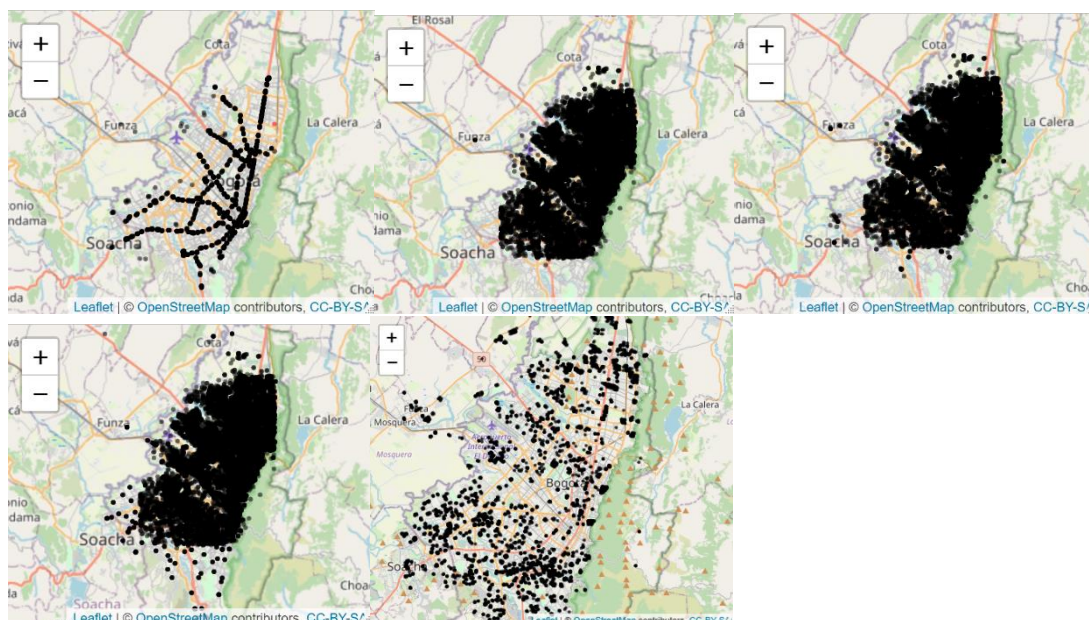
En pocas palabras, hay varios aspectos directos o indirectos que afectan el precio de un inmueble que, para poder tener una buena predicción, deben ser tenidos. Esto debe ser siempre conjunto a un modelo previamente entrenado que permita llegar a estimaciones correctas. De esta forma, se suministra la información necesaria para que la empresa minimice su riesgo y logre romper las barreras de entrada en el mercado inmobiliario caleño.

Referencias

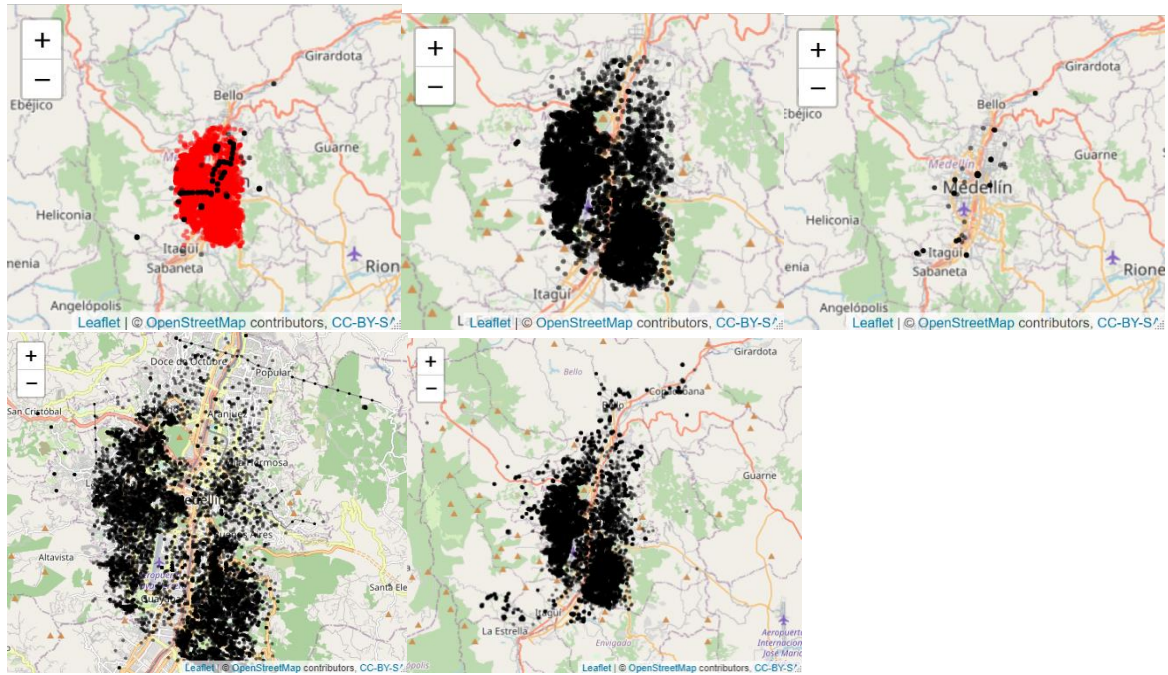
Sheppard, S. (1999). Hedonic analysis of housing markets, In: Cheshire, P. C. y Mills, E. S. (eds.), Handbook of applied urban economics, v.3, chap.41, 1595-1635. New York: Elsevier. (Disponible en: <http://www.sciencedirect.com/science/article/pii/S1574008099800108>)

Rosen (1974). Precios hedónicos y mercados implícitos: diferenciación de productos en competencia pura. Recuperado de: <http://www.ucipfg.com/Repositorio/MAES/PED/Semana4/PreciosHedonicos.pdf>

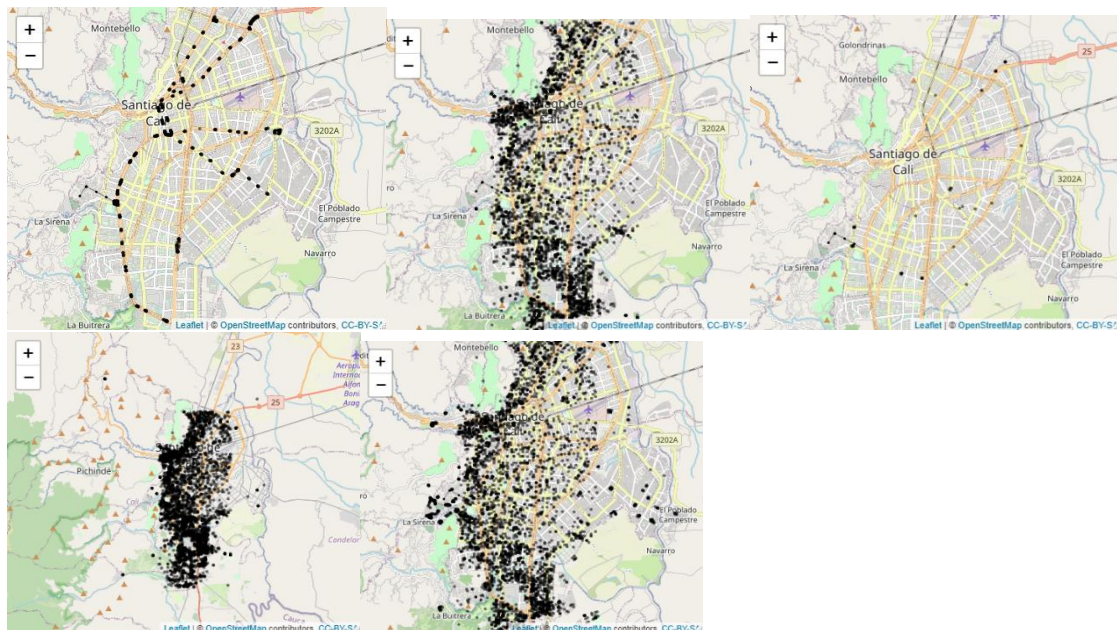
Anexos



De izquierda a derecha: Distancia estaciones de bus en Bogotá, Distancia de oficinas en Bogotá, Distancia de supermercados en Bogotá, Distancia de estaciones de policías en Bogotá y Distancia de colegios en Bogotá



De izquierda a derecha: Distancia estaciones de bus en Medellín, Distancia de oficinas en Medellín, Distancia de supermercados en Medellín, Distancia de estaciones de policías en Medellín y Distancia de colegios en Medellín



De izquierda a derecha: Distancia estaciones de bus en Cali, Distancia de oficinas en Cali, Distancia de supermercados en Cali, Distancia de estaciones de policías en Cali y Distancia de colegios en Cali

- Estadísticas descriptivas

Characteristic	Cali, N = 5,000 [†]
bedrooms	3.75 (1.60)
new_surface	229 (618)
property_type	
Apartamento	2,776 / 5,000 (56%)
Casa	2,224 / 5,000 (44%)
min_dist_bus	1,050 (719)
min_dist_market	2,440 (1,665)
min_dist_colegios	534 (317)
min_dist_oficinas	1,542 (843)
balcon_terr	
Sin Balcón o Terraza	2,504 / 5,000 (50%)
Con Balcón o Terraza	2,496 / 5,000 (50%)
[†] Mean (SD); n / N (%)	

Characteristic	Bogotá D.C, N = 37,985 [†]	Medellín, N = 13,452 [†]
price	869,755,897 (899,818,886)	639,246,711 (623,222,205)
bedrooms	3.08 (1.46)	3.15 (1.02)
new_surface	160 (153)	311 (4,245)
property_type		
Apartamento	29,193 / 37,985 (77%)	10,739 / 13,452 (80%)
Casa	8,792 / 37,985 (23%)	2,713 / 13,452 (20%)
min_dist_bus	879 (610)	1,529 (1,002)
min_dist_market	1,101 (672)	1,562 (1,022)
min_dist_colegios	379 (254)	410 (260)
min_dist_oficinas	1,004 (882)	1,266 (800)
min_dist_policias	677 (378)	1,017 (597)
balcon_terr		
Sin Balcón o Terraza	21,926 / 37,985 (58%)	6,483 / 13,452 (48%)
Con Balcón o Terraza	16,059 / 37,985 (42%)	6,969 / 13,452 (52%)
[†] Mean (SD); n / N (%)		