

## PROBLEM SET 3: HOUSING PRICES

MARIA CAMILA CELY MORENO – SARA OSPINA GIRALDO

Repositorio Github: [https://github.com/camilacely/ProblemSet3\\_Cely\\_Ospina](https://github.com/camilacely/ProblemSet3_Cely_Ospina)

### INTRODUCCIÓN

Las áreas de Chapinero, en Bogotá, y El Poblado, en Medellín, se caracterizan por concentrar propiedades altamente apetecidas por los compradores de las respectivas ciudades, pero también por concentrar los precios más altos. El portal Properati, especializado en asesoría inmobiliaria en Latinoamérica, identificó que para comprar un apartamento en Chapinero y pagarlo con un crédito a 10 años, se requieren ingresos mensuales cercanos a los 28 millones de pesos. Así mismo, para comprar un apartamento en El Poblado en el mismo plazo se requieren ingresos mensuales superiores a los 19 millones de pesos.<sup>1</sup> Al ser las áreas con mayores precios inmobiliarios en Bogotá y en Medellín, resultan muy atractivas como opción de inversión, por lo cual se resalta la utilidad de generar modelos predictivos que permitan aproximarse al precio de las mismas, sobre todo teniendo en cuenta la disparidad entre el avalúo catastral oficial y el precio de mercado. Mediante el presente taller se analizan, por un lado, características inherentes a la vivienda, y, por otro lado, características de su entorno urbano, para aproximar la mejor predicción posible. Pese a que, en promedio, Chapinero y El Poblado concentran propiedades de altísimo valor comercial, se presentan variaciones al interior de dichos barrios que es importante identificar para no sobreestimar propiedades. Como contribución relevante del ejercicio se encuentra que, en los barrios analizados, la proximidad a los sistemas masivos de transporte (Transmilenio y Metro de Medellín, respectivamente) está relacionada con un menor precio de las viviendas, lo cual es coherente con los hallazgos de Munoz-Raskin (2010)<sup>2</sup> y Bocarejo, Portilla & Pérez (2012)<sup>3</sup>, quienes encuentran que los sectores de ingresos medios valoran positivamente la cercanía a Transmilenio, pero los sectores de ingresos altos castigan esta cercanía debido a su preferencia a utilizar vehículo privado y a las externalidades negativas, como ruido y seguridad, asociadas con el transporte masivo.

### DATOS

La base de datos suministrada para el desarrollo del ejercicio proviene del portal Properati y contiene más de 107.000 observaciones de entrenamiento, localizadas tanto en Bogotá como en Medellín (no solamente en los barrios analizados). La información básica asociada a las características de la vivienda incluye área, número de habitaciones y número de baños. Por medio de la inspección del texto de las descripciones de la vivienda, se recuperan características adicionales que mejoran la predicción del precio, tales como si la propiedad tiene balcón, si es dúplex, si tiene ascensor, si tiene parqueadero y si está remodelada. Adicionalmente, por medio del análisis del entorno de las propiedades se recuperaron variables relevantes para la predicción del precio, tales como la distancia a las estaciones de transporte masivo, la distancia a vías principales, la distancia a hospitales y la distancia a parques. (**Ver Tabla 1**). Es importante resaltar que se generó una variable relacionada con la distancia al Central Business District de cada ciudad. Teniendo en cuenta que el CBD de dichas ciudades no se corresponde con el centro tradicional y fundacional de las ciudades, los puntos de CBD se determinan según la localización de grandes clústeres de torres de

---

<sup>1</sup> <https://www.eltiempo.com/economia/finanzas-personales/salario-que-debe-ganar-para-comprar-vivienda-en-bogota-o-medellin-630469>

<sup>2</sup> <https://doi.org/10.1016/j.tranpol.2009.11.002>

<sup>3</sup> <https://doi.org/10.1016/j.retrec.2012.06.030>

oficinas, alrededor de la Carrera Séptima en Bogotá y de la Avenida El Poblado en Medellín, identificados mediante Open Street Maps.

A primera vista, se evidencia que el valor promedio de las propiedades en Chapinero es mayor que en El Poblado, lo cual es coherente con las conclusiones del informe realizado por Properati<sup>1</sup>. En general, se observa que las áreas de los apartamentos en Chapinero son mayores, lo cual sería coherente con la intuición de que, a mayor área, mayor precio; hay mayor distancia promedio al transporte masivo, lo cual puede indicar que hay mayor número de propiedades “resguardadas” de las externalidades negativas asociadas a este (**ver Mapa 1**); hay menor distancia promedio a parques, lo cual puede ser considerado un ammenity importante tanto paisajístico como recreativo y estar correlacionado con mayor precio de las viviendas aledañas (**ver Mapa 2**); hay mayor número de baños en promedio, lo cual puede estar elevando el valor de las propiedades; y hay mayor proporción de propiedades dúplex o con altillo, lo cual da valor agregado al área de las propiedades. Por su parte, se observa que en El Poblado hay mayor proporción de apartamentos con balcón o terraza, lo cual es coherente con la diferencia de temperatura promedio de la ciudad; y también hay mayor proporción de apartamentos con ascensor, lo cual da cuenta de la diferencia de regulación de alturas que hay en las dos ciudades, así como la diferencia de edad de los edificios construidos. Es importante resaltar que la base suministrada cuenta con menos observaciones en Medellín que en Bogotá, por lo cual es posible que esta diferencia en la proporción de información de la base pueda llegar a impactar los resultados predictivos.

Para el ejercicio predictivo se generan las variables mencionadas principalmente para las observaciones localizadas directamente en Chapinero y El Poblado, ya que como se mencionó, se considera que la capacidad predictiva dependerá de poder encontrar las diferencias “sutiles” que determinan el precio al interior de los barrios, y no necesariamente de las diferencias que tienen estas propiedades cuando se las compara con el resto de la ciudad. Sin embargo, con el fin de poder comparar y sustentar esta hipótesis, se generan algunas variables para la base de datos completa, obteniendo los resultados presentados en la Tabla 2. Se observa que el precio promedio de las propiedades disminuye, especialmente en Bogotá, lo cual estaría dando cuenta de mayores diferencias de poder adquisitivo de la población que reside en la capital, comparado con quienes viven en Medellín. La proporción de casas en venta aumenta por fuera de Chapinero, lo cual indica que en Chapinero hay mayor concentración de apartamentos que en el resto de la ciudad, y esto a su vez indica que el suelo es más caro, por lo cual hay que hacerlo más rentable optimizando el área construida. La distancia promedio a las vías principales aumenta para ambas ciudades, así como la distancia promedio al CBD, ya que de hecho el CBD en ambas ciudades está localizado dentro de los barrios analizados. Por este motivo, se observa que la distancia al CBD es la variable más relevante para predecir precios en la muestra completa, pero no es la única variable relevante en Chapinero y El Poblado (**ver Mapa 3**).

## MODELOS Y RESULTADOS

Teniendo en cuenta lo anterior, y a partir de las experiencias del Problem Set pasado, se decidió aplicar la metodología de Random Forests para realizar una primera aproximación a cuáles son las variables que mejor explican el precio de las propiedades analizadas. Ya que para esta metodología se requiere que las variables proporcionadas sean dummies, se realizó una transformación que tomaba valor de 1 si la distancia era superior al promedio y 0 de lo contrario, para el caso de todas las variables espaciales suministradas (distancia a bares, a transporte público, a cafés, a restaurantes, a bares, a cines, a teatros, a hospitales, a colegios, a universidades, a parques, a vías principales, al CBD y a las montañas). Así mismo, para las variables obtenidas de la descripción de los anuncios, se generaron variables dicótomas que tomaban el valor de 1 si la vivienda estaba por encima del promedio y 0 de lo contrario, en aspectos tales como el número de habitaciones, si la vivienda es penthouse, si es dúplex, si tiene altillo, si es remodelada, si tiene vista al exterior, si tiene ascensor y si tiene parqueadero.

El análisis por medio de esta metodología se realizó con dos bases de datos: con la base completa de observaciones de la ciudad, y con la submuestra que solo incluía las observaciones de Chapinero y El Poblado. De forma relevante, se obtiene que la variable más representativa en la muestra completa es la distancia al CBD, pero la variable más representativa en la submuestra de los barrios en cuestión es la distancia al transporte público, lo cual indica que la predicción de los precios en barrios tan cercanos al CBD de la ciudad debe especificarse con mucho más detalle que la predicción general de la ciudad, pues entre viviendas con poca distancia al CBD hay mucha variación de características tanto internas como urbanas.

Para el ejercicio de predicción del precio de los inmuebles, con base en lo encontrado en el ejercicio anterior, se corrieron cuatro modelos con las variables que se consideraron más significativas. En primer lugar, se corrieron dos modelos de mínimos cuadrados ordinarios, donde se comparó entrenar el modelo para toda la ciudad y solo para los barrios a predecir. El modelo que se entrenó con información de adentro de los barrios obtuvo un mejor R cuadrado (0.56) por lo cual los siguientes dos modelos se entrenaron con esta muestra. Posteriormente, se corrieron modelos Lasso y Ridge, con lo cual se encontró que efectivamente la variable de cuarto tiene un gran peso dentro del precio del inmueble. Ambos modelos valoraron de forma diferente las variables, el Ridge consideró más importante las variables de distancia.

Con base en los resultados de los MSE, se seleccionó el modelo de mínimos cuadrados ordinarios con la muestra de los barrios de Chapinero y El Poblado. De esta forma se predijeron los valores de precios estimados dentro de la base de test suministrada.

## **CONCLUSIONES Y RECOMENDACIONES**

- Por las características de ambos barrios, es fácil sobreestimar los precios porque si se comparan las observaciones con el resto de la ciudad estas tienden a ser más altas por la cercanía a la batería de amenities de ambas ciudades.
- Un modelo predictivo debe ahondar en las características internas del barrio, como las que se tuvieron en cuenta para los modelos de este trabajo. Esto se debe a que al interior de cada uno de los barrios se presenta mucha variación de precios, lo cual, puede llevar a malas estimaciones si no se identifican apropiadamente los canales principales que tienen mayor influencia en los precios del mercado.
- Este tipo de aproximaciones son útiles tanto para inmobiliarias como para los formuladores de política pública, ya que el hecho de estar cerca de universidades y parques, por ejemplo, aumenta el precio del suelo, lo que hace que solo personas con cierto nivel adquisitivo puedan vivir en estos sectores. Otras personas de distintas características socioeconómicas deben vivir mucho más alejadas de este tipo de amenities, por lo cual su calidad de vida se ve directamente afectada, y esta segregación socioespacial se perpetúa al dejarla solamente en manos del mercado.

## ANEXOS

**Tabla\_1**

Characteristic	Bogotá D.C, N = 14,738 <sup>†</sup>	Medellín, N = 1,633 <sup>†</sup>
price	990,000,000 (624,000,000, 1,670,000,000)	380,000,000 (262,000,000, 685,000,000)
property_type		
Apartamento	13,849 (94%)	1,178 (72%)
Casa	889 (6.0%)	455 (28%)
area	142 (122, 165)	130 (77, 157)
dist_bus	741 (400, 1,225)	630 (283, 1,158)
dist_playground	219 (139, 321)	1,192 (992, 1,362)
dist_highway	1,020 (493, 1,454)	1,530 (1,104, 1,736)
dist_cbd	1,204 (753, 1,577)	2,894 (2,335, 3,252)
Num de cuartos	3.00 (2.00, 3.00)	3.00 (2.00, 3.00)
balcon	5,297 (36%)	788 (48%)
extras	555 (3.8%)	12 (0.7%)
ascen	2,662 (18%)	406 (25%)
parq	7,846 (53%)	807 (49%)
renov	1,979 (13%)	60 (3.7%)

**Tabla\_2**

Characteristic	Bogotá D.C, N = 84,555 <sup>†</sup>	Medellín, N = 21,030 <sup>†</sup>
price	520,000,000 (320,000,000, 900,000,000)	310,000,000 (217,000,000, 450,000,000)
property_type		
Apartamento	64,395 (76%)	16,293 (77%)
Casa	20,160 (24%)	4,737 (23%)
area	100 (100, 110)	100 (75, 100)
dist_highway_ciudades	1,369 (642, 2,317)	1,528 (770, 2,349)
dist_cbd_ciudades	3,483 (1,570, 6,654)	5,576 (4,248, 7,381)

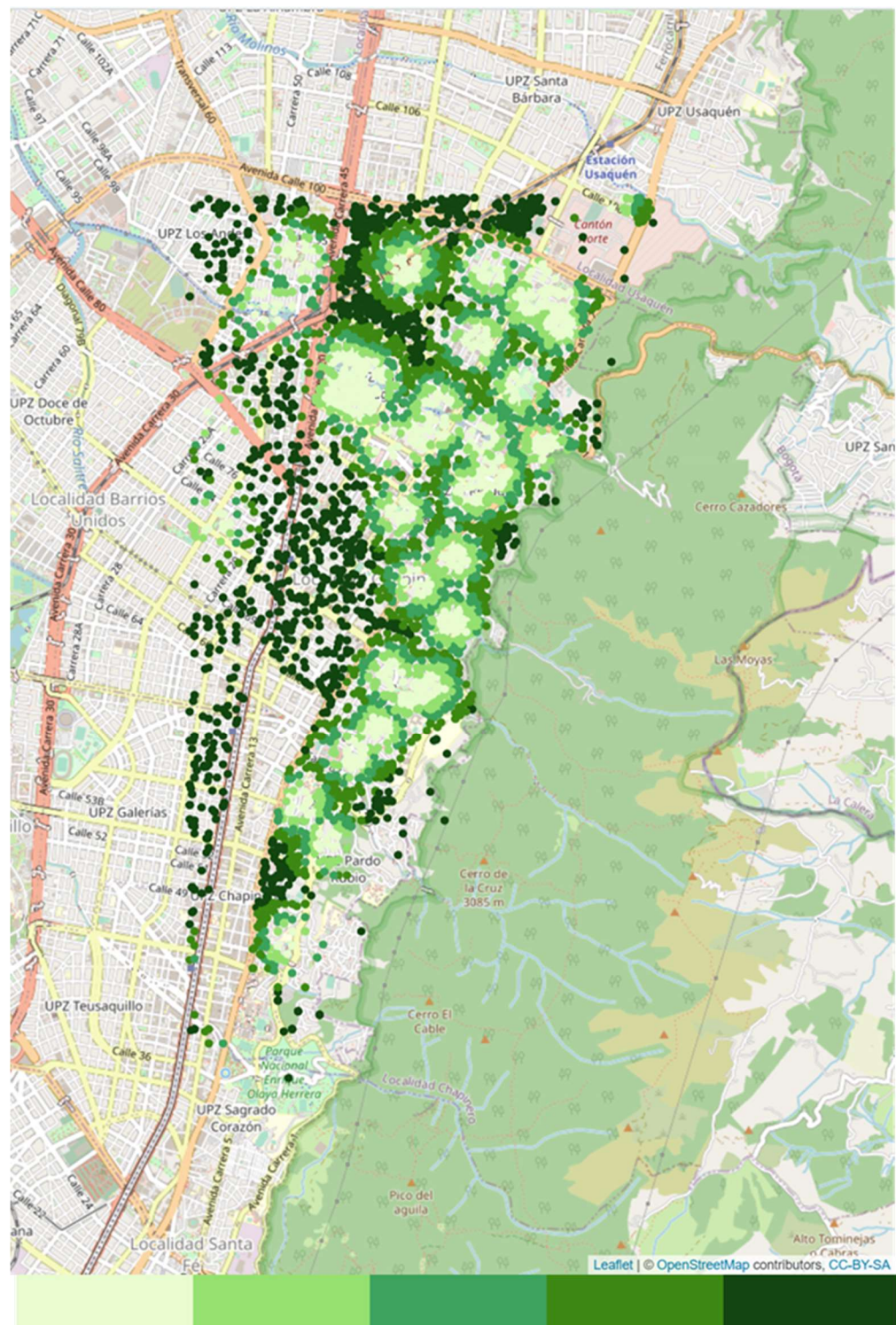
<sup>†</sup> Median (IQR); n (%)

Precio/m2  
mayor

● Estación transporte masivo



### Mapa\_2 (Chapinero) - Distancia a parques

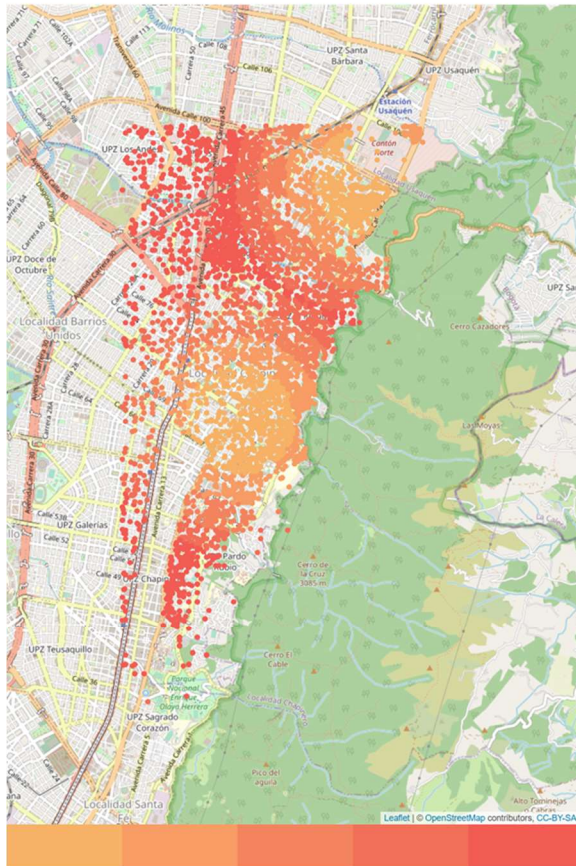


Menor  
distancia  
a parque

Mayor  
distancia a  
parque

### Mapa\_3 (Chapinero y El Poblado) - Distancia a CBD

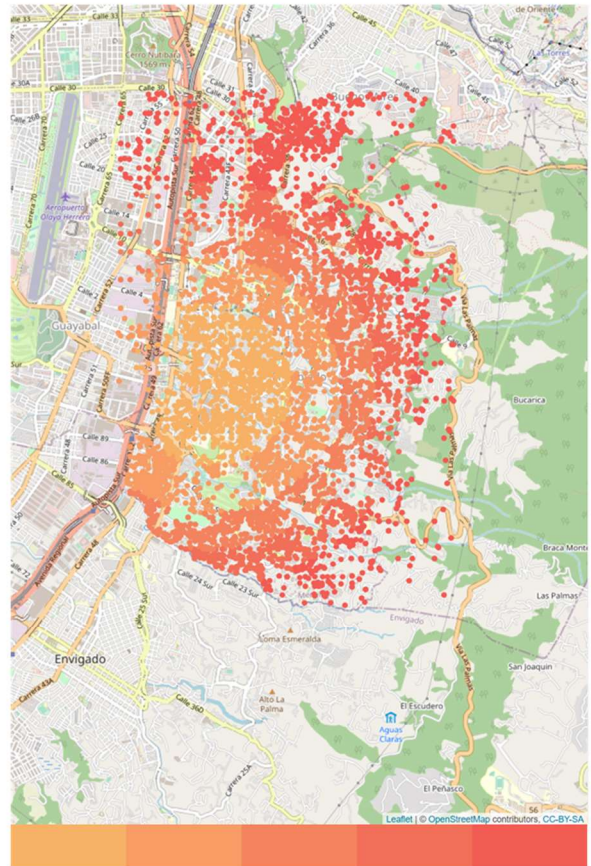
Chapinero



Menor  
distancia  
a CBD

Mayor  
distancia a  
CBD

El Poblado



Menor  
distancia  
a CBD

Mayor  
distancia a  
CBD