

# **Data Challenge – Econometría Aplicada I**

## **Identificación de las colonias con mayor rendimiento potencial para invertir en Airbnb en la CDMX**

Aldo Muller Quintero

Alexis Capdevielle Harrison

Liliana Vargas Lugo

Noviembre 2025

Este documento presenta un análisis econométrico aplicado al mercado de rentas de corta estancia en la Ciudad de México con el objetivo de identificar las tres colonias con mayor rendimiento potencial para invertir en propiedades destinadas a Airbnb. Utilizamos las bases de Inside Airbnb para Ciudad de México, complementadas con información externa a nivel de colonia, y estimamos dos modelos, uno modelo de precios hedónicos por colonia en forma semilogarítmica para explicar el precio por noche en función de atributos intrínsecos del alojamiento y características de la colonia, y después calculamos la rentabilidad con un lapso de tiempo de 15 años y rentabilidad del 10 por ciento (VPN-Costo de adquisición). Y el otro enfoque que utilizamos fue el de rentabilidad por propiedad en función de dichos atributos considerando la variable de Gross Operating Profit como la variable dependiente, calculada como un porcentaje del 40 por ciento de los ingresos promedios anuales (producto de la ocupación y el precio promedio por noche). A partir de los resultados, construimos un criterio de ranking por colonia y proponemos recomendaciones de inversión para la empresa. Recomendando así un portafolio con propiedades en las siguientes colonias: Roma Norte, Pedregal de Carrasco y Cuauhtémoc y con el otro enfoque Roma Norte, Bellavista y Condesa.

# 1. Introducción y motivación

En este informe, se lleva a cabo un análisis del mercado de Airbnb en la Ciudad de México con el objetivo de identificar las colonias más convenientes para invertir en propiedades destinadas a rentas de corta estancia. Para ello, se adoptan dos enfoques diferentes que permiten estimar la rentabilidad de estas inversiones.

Desde la perspectiva de una empresa de inversión en bienes raíces, es fundamental no solo identificar zonas con precios altos, sino comprender las características que explican esas diferencias de precio y cómo se relacionan con el potencial de rendimiento de una inversión inmobiliaria centrada en Airbnb. Complementariamente, desde una visión hotelera, es esencial analizar el Gross Operating Profit (GOP) para determinar cuáles colonias son las más rentables.

Para este análisis, se emplearon dos modelos de Mínimos Cuadrados Ordinarios (MCO) con estimación robusta a heteroscedasticidad. El primer enfoque se basa en estimar el GOP en función de las características intrínsecas de las propiedades y de su entorno. El segundo enfoque utiliza un modelo de precios hedónicos por colonia, donde el precio promedio por noche se interpreta como el valor que, en promedio, los huéspedes asignan a distintos atributos de las propiedades y su contexto, permitiendo cuantificar el aporte marginal de características como el número de huéspedes, el tipo de alojamiento y las amenidades disponibles.

Se utilizaron dos métricas de rentabilidad. Una para cada enfoque: la métrica de rentabilidad del primer enfoque se aproximó utilizando ingresos anuales promedio, considerando la ocupación de cada espacio y restando los gastos operativos relacionados con la industria hotelera, que incluyen servicios básicos, comisiones de Airbnb y costos de limpieza y mantenimiento (estos gastos operativos se calcularon como el 60 por ciento de los ingresos anuales promedio que es equivalente a decir que el GOP equivale al 40 por ciento de éstos, según estados financieros de una cadena hotelera (información no pública)). Además, para el segundo enfoque, se calculó el costo de adquisición aproximado restando dicho gasto de los ingresos potenciales promedio por colonia, utilizando el precio de venta promedio por metro cuadrado y el tamaño promedio de las propiedades en cada colonia. Para el análisis de estas regresiones se recurrió al trabajo de López Tamayo D. (2021), quien publica un análisis de precios hedónicos para Airbnb en la Ciudad de México, del cual se extrajeron valiosas variables explicativas.

## **2. Metodología aplicada**

### **2.1. Bases de datos**

Las bases de datos de Airbnb y las externas, se unificaron utilizando el código de cada colonia de geojson, se eliminaron anuncios y se construyeron las variables necesarias para realizar las regresiones de los modelos que se plantean más adelante. Se utilizan solamente las observaciones de las propiedades de la categoría de entire home/apt con lo que al final cuenta con la información de 17,235 propiedades de Airbnb. Un inversionista no puede ampliar su portafolio comprando partes de recámaras o de cuartos de hoteles como tal por lo que eliminar esas observaciones de la base de datos no genera ningún sesgo.

### **2.2. Variables principales**

Cuadro 1: Descripción breve de las variables utilizadas

Variable	Descripción
<b>Variables dependientes</b>	
log(GOP)	Log(Gross Operating Profit): $0,4 \times$ precio promedio $\times$ ocupación.
log(precio)	Logaritmo del precio promedio por noche por colonia.
<b>Características del alojamiento</b>	
Habitaciones	Número de habitaciones.
Baños	Número de baños.
Huéspedes	Capacidad máxima permitida.
Amenidades	Conteo de amenidades.
Reviews	Número total de reseñas.
Superhost	Dummy: anfitrión con estatus Superhost.
<b>Características de la colonia</b>	
Distancia turística	Distancia promedio a 10 puntos turísticos principales.
Robos	Número de robos registrados por colonia.
Metro	Número de estaciones de Metro en la colonia.
Metrobús	Número de estaciones de Metrobús.
Competencia	Número de alojamientos Airbnb en la colonia.
<b>Variables de rentabilidad</b>	
VPN	Ingreso esperado a 15 años descontado al 10 %.
Costos de adquisición	$m^2$ estimados $\times$ precio promedio del $m^2$ por colonia.

### 3. Estimación econométrica: modelos de Mínimos Cuadrados Ordinarios semilogarítmicos

#### 3.1. Enfoque 1: Modelo de rentabilidad semilogarítmico

Un modelo que utiliza los datos de las propiedades en donde estima el GOP en función de características intrínsecas de las propiedades y de Airbnb así como las características de la zona, realizando estimación robusta a heteroscedasticidad.

La forma funcional básica que emplearemos es:

$$\ln(GOP_i) = \beta_0 + X_i^{\text{intr}} \beta_{\text{intr}} + X_i^{\text{colonia}} \beta_{\text{colonia}} + \varepsilon_i, \quad (1)$$

donde:

- $GOP_i$  es el Gross Operating Profit anual del alojamiento  $i$ ,
- $X_i^{\text{intr}}$  agrupa los atributos intrínsecos del alojamiento,
- $X_i^{\text{colonia}}$  contiene las variables externas a nivel de colonia,
- $\varepsilon_i$  es el término de error.

La forma semilogarítmica permite interpretar los coeficientes como aproximaciones a cambios porcentuales en el GOP frente a variaciones en cada característica.

### 3.2. Enfoque 2: Modelo de precios hedónicos por colonia semilogarítmico

Un modelo de precios hedónicos interpreta el precio de un bien heterogéneo como la suma del valor que los consumidores asignan a cada una de sus características. En nuestro caso, el bien es una noche de alojamiento en una propiedad de Airbnb, y sus características incluyen atributos intrínsecos a las propiedades y de la colonia.

La forma funcional básica que emplearemos es:

$$\ln(p_i) = \beta_0 + X_i^{\text{intr}} \beta_{\text{intr}} + X_i^{\text{colonia}} \beta_{\text{colonia}} + \varepsilon_i, \quad (2)$$

donde:

- $p_i$  es el precio por noche promedio de la colonia  $i$ ,
- $X_i^{\text{intr}}$  agrupa los promedios de los atributos intrínsecos del alojamiento de cada colonia,
- $X_i^{\text{colonia}}$  contiene las variables externas a nivel de colonia,
- $\varepsilon_i$  es el término de error.

En este enfoque se realizó después de la estimación del modelo, el cálculo del VPN a quince años con una tasa de rendimiento del 10 por ciento y de la rentabilidad utilizando los valores ajustados del  $\log(\text{precio})$  del modelo que estimamos restándole el costo de adquisición que se calculó como el producto del precio de venta promedio por colonia y los metros cuadrados aproximados por colonia.

## 4. Resultados principales

A continuación se presentan los resultados de los modelos especificados para ambos enfoques con los mapas en donde colores más oscuros implican colonias más rentables, para el enfoque 1 y el 2:

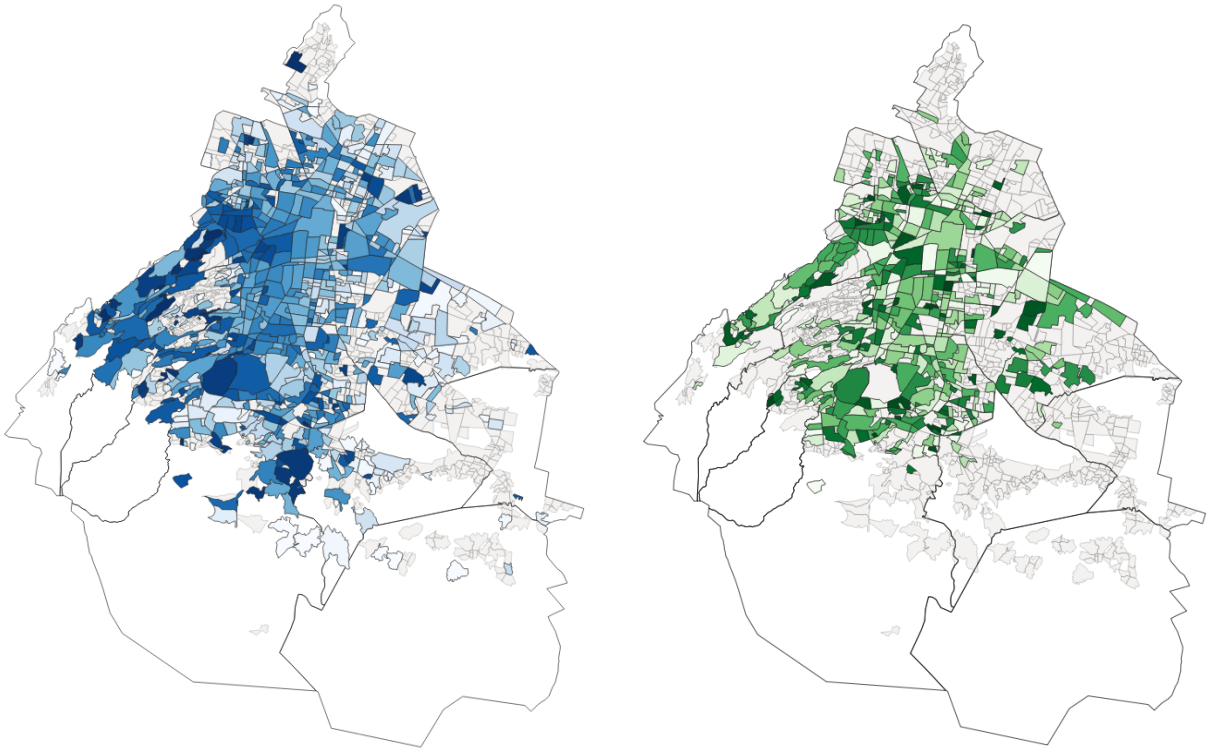


Figura 1: Comparación: mapa GOP y mapa de rentabilidad de Airbnb en CDMX.

## 5. Recomendaciones de inversión y ranking de colonias

Después de analizar los resultados del primer enfoque, se llevó a cabo un promedio ponderado por colonia para identificar las tres colonias con mayor Gross Operating Profit (GOP) ajustado. Sin embargo, se destaca una limitante importante: las escasas observaciones por colonia podrían indicar que las propiedades son rentables, pero no necesariamente la colonia en sí. Por lo tanto, se consideraron solo las colonias representativas, es decir, aquellas con un número de propiedades superior al promedio, para determinar en qué sectores conviene invertir.

Se generó una tabla que presenta las seis colonias más rentables considerando únicamente las colonias con una representatividad suficiente en los datos (para lo que se eliminaron las

colonias que tenían menos propiedades que la media).

Para los inversionistas que priorizan el flujo de caja, se recomienda el enfoque 1, sugiriendo ampliar el portafolio en las colonias de Bellavista, Condesa y Roma Norte. En cambio, aquellos que se centran en la rentabilidad basada en el costo de adquisición deberían considerar colonias como Roma Norte, Pedregal de Carrasco y Cuauhtémoc.

La colonia más recomendada es Roma Norte, dada su densidad de propiedades Airbnb y su popularidad en la ciudad. Las opciones segunda y tercera dependerán del indicador de rentabilidad preferido, con Condesa y Cuauhtémoc como alternativas esperadas, y Bellavista o Pedregal de Carrasco como opciones menos típicas pero igualmente atractivas para los inversionistas.

Cuadro 2: Colonias con mayor rentabilidad

<b>Enfoque 1 (GOP)</b>	<b>Enfoque 2 (Costo de adquisición)</b>
Bellavista	Roma Norte
Condesa	Pedregal del Carrasco
Roma Norte	Cuauhtémoc

Si los inversionistas le dan prioridad al flujo al invertir en una propiedad, deben considerar el enfoque 1 de nuestras estimaciones y considerar ampliar el portafolio en las colonias: Bellavista, Condesa y Roma Norte.

Si por el otro lado, consideran más un enfoque de rentabilidad basado en el costo de adquisición, deben considerar ampliar su portafolio con propiedades en las siguientes colonias: Roma Norte, Pedregal de Carrasco y Cuauhtémoc.

Por lo que la colonia que más recomendamos, siendo consistentes con la densidad de Airbnbs en la zona y la popularidad en la ciudad es la Roma Norte. La segunda y tercera opción dependerán de qué indicador de rentabilidad por el que tengan preferencia pudiendo elegir una opción esperada como Condesa y Cuauhtémoc o algo más insospechado como resultó Bellavista o Pedregal de Carrasco que pueden ser atractivas y tal vez menos típicas de las que se suele considerar típicamente.

## 6. Conclusiones

Se utilizaron dos enfoques diferentes en cuanto a la definición de rentabilidad para contestar la pregunta de interés de conocer cuáles son las colonias más rentables. Hicimos varios supuestos que la literatura y el sentido común sustentan para poder modificar un modelo tradicional de precios hedónicos obteniendo resultados claros y rotundos.

Como se puede observar, ambos enfoques de rentabilidad son confiables y consistentes ya que nos presentan una opción igual de colonia (Roma Norte) iguales entre ambos por lo que es una decisión segura invertir en ella, y las otras alternativas también suenan consistentes con lo que se conoce en el mercado inmobiliario de la CDMX.

## **7. Anexo metodológico**

Debido a que el enfoque de este documento no es académico, sino más bien para inversión, decidimos detallar en un anexo los componentes que consideramos relevantes en el plano académico pero que no afectan el análisis realizado sobre rentabilidad.

### **7.1. Manejo y creación de la base de datos**

La base de datos proviene de Inside Airbnb para CDMX, DATOS de crimen, datos de trip advisor. De la base de datos de Airbnb, se utilizan solamente las observaciones de la categoría de entire home/apt con lo que al final cuenta con la información de 17, 235 propiedades de Airbnb. Para la estimación por colonia, se agregan de manera ponderada las observaciones de las propiedades a nivel colonia para contar con 750 colonias. Para el análisis se construyeron diversas variables a nivel colonia utilizando fuentes oficiales y bases de datos públicas. La variable Robos corresponde al número total de robos registrados por colonia, obtenidos a partir del conteo de víctimas reportadas en las carpetas de investigación de la Fiscalía General de Justicia (FGJ), disponibles en el portal de datos abiertos de la Ciudad de México. La variable Metro mide la accesibilidad al transporte masivo mediante el número de estaciones de Metro ubicadas dentro de cada colonia, utilizando la información de líneas y estaciones publicada por el STC en datos abiertos. De manera equivalente, la variable Metrobús captura el número de estaciones de este sistema presentes en cada colonia, a partir de la base de líneas y estaciones del Metrobús en datos abiertos.

La variable Competencia se aproxima mediante el número de propiedades listadas en plataformas de alojamiento dentro de cada colonia, utilizando la base de datos de Inside Airbnb para la Ciudad de México. Para estimar la rentabilidad esperada, se construyó una variable de VPN, definida como el valor presente neto del ingreso proyectado a 15 años, calculado multiplicando el precio por noche por los días estimados de ocupación. Estos flujos futuros se trajeron a valor presente utilizando una tasa de descuento estándar del 10 por ciento, ampliamente empleada en evaluaciones financieras y económicas para reflejar el costo de oportunidad del capital y el riesgo asociado a inversiones inmobiliarias.

Finalmente, los Costos de adquisición se estimaron mediante la construcción de un valor



aproximado de metros cuadrados para cada alojamiento. Este estimado se obtuvo utilizando información sobre el tipo de propiedad (casa o departamento) y el número de cuartos disponibles, lo que permitió aproximar su superficie. Posteriormente, este valor se multiplicó por el precio promedio por metro cuadrado de cada colonia, empleando los datos del artículo “Gentrification and access to housing in Mexico City during 2000–2022”

## 7.2. Estadística descriptiva de variables relevantes

Aquí se incluyen los histogramas de las variables dependientes y las tablas de las estadísticas descriptivas de los modelos.

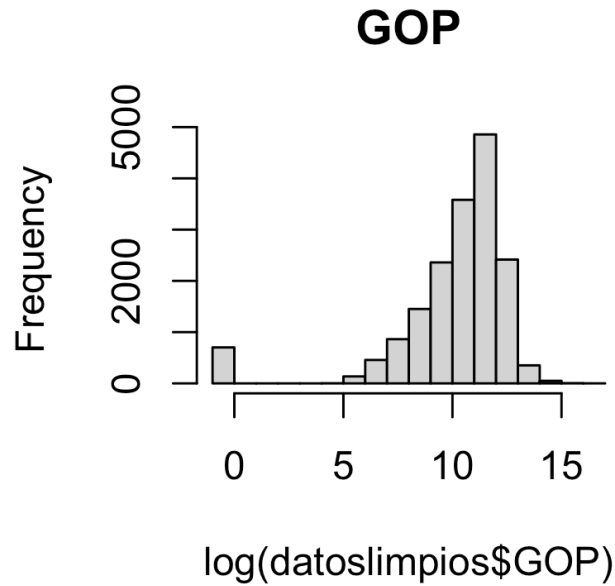


Figura 2: Histograma de la variable dependiente del enfoque 1

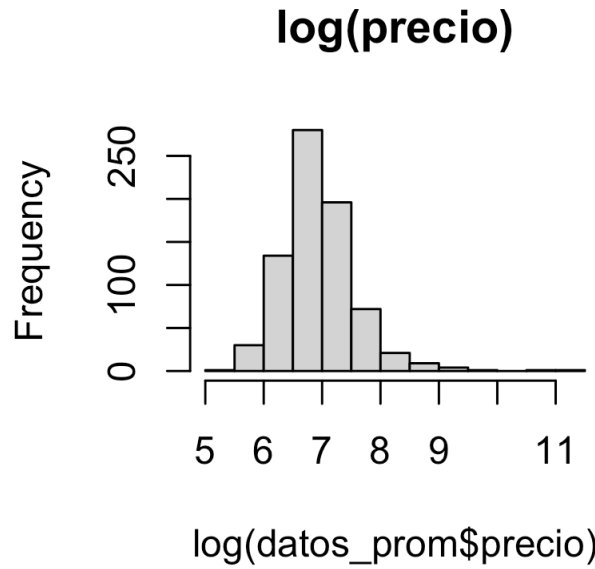


Figura 3: Histograma de la variable dependiente del enfoque 2

Cuadro 3: Estadísticos descriptivos de las propiedades

Variable	Mín	Q1	Mediana	Media	Q3	Máx
baños	1	1	1	1.523	2	49
precio	165	910	1287	1870	1944	135000
superhost	0	0	0	0.4112	1	1
disponibilidad	0.0	145	269	233.7	341	365
amenidades	1.00	25	36	35.94	47	110
reviews	0	0.42	1.33	1.832	2.65	80.38
huéspedes	1	2	4	4.006	5	16
promedio_distancia_turística_km	4.905	5.165	5.619	6.846	7.049	32.536
metro	0	0	1	1.24	1	11
metrobús	0	0	2	4.311	3	35
n_Airbnb_colonia	17235					
robos	1.0	128	286	574.9	689	2517
habitaciones	0	1	2	1.744	2	50
competencia	1.0	53.0	268	593.6	793	2138
ca	165760	1535840	2539020	3010504	3533600	45978400
ocupación	0.0	24	96	131.3	220	365
IAP	1	28442	121759	234058	299209	22437001
GOP	0	11377	48704	93623	119684	8974800
GOP ajustado	3.425	9.6	10.043	10.086	10.509	18.855

Cuadro 4: Estadísticos descriptivos de las colonias

Variable	Mín	Q1	Mediana	Media	Q3	Máy
baños	1	1	1.205	1.427	1.569	8
precio	220.0	695.5	983.3	1395.4	1384.9	67500
superhost	0	0	0.25	0.3129	0.5	1
disponibilidad	0.0	180	239.8	231.4	303.5	365
amenidades	1	23	29.59	29.21	36.1	69.21
reviews	0	0.4738	1.0845	1.2846	1.8258	6.97
huéspedes	1	3	4	4.069	4.6	16
promedio_distancia_turística_km	4.962	7.479	10.103	11.428	14.859	32.536
metro	0	0	0	0.2307	0	11
metrobus	0	0	1	0.824	1	35
robos	1.00	20.00	43.00	87.49	93.00	2517
habitaciones	0	1.325	1.736	1.775	2	9.3
competencia	1.00	1.00	3.00	25.03	9.00	2138
ca	276006	1740149	2567702	2876311	3600830	10947105
ocupación	0.00	61.54	125.17	133.61	185	365
peso	1.00	1.00	3.00	22.98	8.00	2138
precio ajustado	6.236	6.744	6.892	6.936	7.044	9.031
iap	0	60395	130658	146888	196656	2207813
vpn	0	459371	993794	1117241	1495781	16792801
rentabilidad	-10855917	-2470827	-1518988	-1702189	-588120	3044122

### 7.3. Estimación de los modelos econométricos

Para concluir qué variables considerar en las regresiones finales de los modelos se realizó el siguiente proceso:

- (I) **Modelo hiperspecificado:** con todos los posibles atributos intrínsecos y por colonia del alojamiento y a nivel colonia.
- (II) **Modelo depurado:** atributos intrínsecos y atributos por colonia significativos.
- (III) **Modelo depurado final:** en el caso del segundo enfoque, se tuvo que realizar la depuración dos veces por lo que hay 3 regresiones, atributos intrínsecos y atributos por colonia significativos.

Hay algunas consideraciones importantes como el signo de robos que está presentando un sesgo porque En todos los casos, la estimación se realizó mediante Mínimos Cuadrados Ordinarios (MCO) y se emplearán errores estándar robustos. Los promedios están siempre ponderados por el número de observaciones de cada colonia.

## 8. Resultados descriptivos

A continuación se presentan los resultados de los modelos especificados para ambos enfoques:

### 8.1. Limitaciones del análisis y extensiones futuras

Una extensión natural de este enfoque sería incorporar efectos fijos por colonia en la regresión, de manera que se absorba la heterogeneidad no observada entre zonas (como reputación, calidad urbana o trayectoria turística). Esto permitiría interpretar los efectos fijos como una medida del valor intrínseco de cada colonia para el mercado de Airbnb. Sin embargo, en este trabajo nos concentramos en especificaciones MCO simples, dejando el análisis con efectos fijos como una posible extensión futura ya que se necesitarían datos panel. Otras posibilidades son modelos de econometría espacial como Durbin Watson Model y el GWR como lo estiman para varias ciudades europeas y Buenos Aires respectivamente Gyodi y Nawaro (2021) y Alvarez-Herranz, Macedo-Ruiz y Quiroga (2025) pero están fuera del alcance de esta investigación.

## Referencias

- [1] López Tamayo, Diego Alberto, y Aurora A. Ramírez-Álvarez (2021). *Análisis de Precios Hedónicos para Airbnb en la CDMX*. Documento de Trabajo, Centro de Estudios Económicos, El Colegio de México.
- [2] Merino, Juan José, y Edwin Muñoz-Rodríguez (2024). *Professional Airbnb Hosts in Mexico City: A First Approximation*. Documento de Trabajo, Centro de Estudios Económicos, El Colegio de México.
- [3] Aguilar-Velázquez, D., Rivera Islas, I., Romero Tecua, G., Valenzuela-Aguilera, A. (2024). *Gentrification and access to housing in Mexico City during 2000 to 2022. Proceedings of the National Academy of Sciences of the United States of America, 121(10), Article e2314455121. <https://doi.org/10.1073/pnas.2314455121>*

Cuadro 5

	<i>Dependent variable:</i>	
	log(GOP)	
	(1)	(2)
habitaciones	0.142** (0.064)	0.148* (0.082)
baños	0.008 (0.062)	
huespedes	0.136*** (0.031)	0.136*** (0.029)
amenidades	0.020*** (0.002)	0.020*** (0.002)
reviews	-0.083*** (0.014)	-0.083*** (0.014)
superhost	0.409*** (0.039)	0.410*** (0.039)
promedio_distancia_turistica_km	-0.091*** (0.011)	-0.092*** (0.011)
metro	0.005 (0.027)	-0.015** (0.007)
metrobus	-0.002 (0.006)	
competencia	0.0002*** (0.0001)	0.0002*** (0.00003)
robos	-0.0001 (0.0002)	
Constant	9.078*** (0.120)	9.081*** (0.116)
Observations	17,115	17,115
R <sup>2</sup>	0.078	0.078
Adjusted R <sup>2</sup>	0.077	0.077
Residual Std. Error	2.631 (df = 17103)	2.633 (df = 17106)
F Statistic	131.041*** (df = 11; 17103)	181.582*** (df = 8; 17106)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01  
Niveles de significancia: \* p<0.1, \*\* p<0.05, \*\*\* p<0.01

Cuadro 6

	<i>Dependent variable:</i>		
	log(precio)		
	(1)	(2)	(3)
habitaciones	0.080 (0.096)	0.080 (0.096)	0.078 (0.096)
Baños	0.213*** (0.092)	0.213 (0.092)	0.217** (0.092)
huespedes	0.069** (0.038)	0.069** (0.038)	0.068* (0.038)
amenidades	0.005*** (0.003)	0.005*** (0.003)	0.005** (0.003)
reviews	-0.083*** (0.017)	-0.087*** (0.018)	-0.090*** (0.018)
superhost	-0.053 (0.055)		
promedio_distancia_turistica_km	-0.011 (0.005)	-0.010 (0.005)	-0.009* (0.005)
metro	-0.027 (0.026)	-0.008 (0.022)	
metrobus	-0.029*** (0.014)	-0.020 (0.012)	
competencia	0.001*** (0.0003)	0.001*** (0.00003)	0.001*** (0.0002)
robos	0.0003* (0.0002)		
Constant	6.290*** (0.122)	6.293*** (0.120)	6.268*** (0.119)
Observations	743	743	743
R <sup>2</sup>	0.267	0.265	0.261
Adjusted R <sup>2</sup>	0.256	0.256	0.254
Residual Std. Error	0.536 (df = 731)	0.537 (df = 733)	0.537 (df = 735)
F Statistic	24.151*** (df = 11; 731)	29.295*** (df = 9; 733)	37.067*** (df = 7; 735)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01  
Niveles de significancia: \* p<0.1, \*\* p<0.05, \*\*\* p<0.01