



Revista de Economía del Rosario

ISSN: 0123-5362

ISSN: 2145-454X

luis.gutierrez@urosario.edu.co

Universidad del Rosario

Colombia

Cárdenas Rubio, Jeisson Arley; Chaux Guzmán, Francisco José; Otero, Jesús
Una base de datos de precios y características de vivienda en Colombia con información de Internet
Revista de Economía del Rosario, vol. 22, núm. 1, 2019, Enero-Junio, pp. 75-100
Universidad del Rosario
Colombia

DOI: <https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.7768>

Disponible en: <http://www.redalyc.org/articulo.oa?id=509559056003>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEH
redalyc.org

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Una base de datos de precios y características de vivienda en Colombia con información de Internet

Recibido: octubre, 10 de 2018 - Aprobado: noviembre, 30 de 2018

Doi: <http://dx.doi.org/10.12804/revistas.urosario.edu.co/economia/a.7768>

Jeisson Arley Cárdenas Rubio*
Francisco José Chaux Guzmán†
Jesús Otero‡

Resumen

Este documento ilustra la utilización de técnicas de la ciencia de datos para la creación y análisis de una base con información de precios y características de venta o arriendo de casas o apartamentos en las principales ciudades de Colombia. Se espera que esta nueva base de datos, que sigue siendo actualizada, se pueda convertir en el futuro en una herramienta útil para el diagnóstico temprano del comportamiento y evolución del sector inmobiliario en el país.

Palabras clave: ciencia de datos, extracción de información automatizada.

Clasificación JEL: C80, R31.

* Institute for Employment Research University of Warwick, Reino Unido. Correo electrónico: J.Cardenas-Rubio@warwick.ac.uk

† Facultad de Economía Universidad del Rosario, Colombia. Correo electrónico: cal.chaux@gmail.com

‡ Facultad de Economía Universidad del Rosario, Colombia. Correo electrónico: jesus.otero@urosario.edu.co

Para citar este artículo: Cárdenas Rubio, J. A., Chaux Guzmán, F. J., & Otero, J. (2019). Construcción de una base de datos de precios y características de vivienda para Colombia. *Revista de Economía del Rosario*, 22(1), 75-100. doi: <http://dx.doi.org/10.12804/revistas.urosario.edu.co/economia/a.7768>

An Internet-Based Data Set of Prices and Characteristics of Dwelling in Colombia

Abstract

This paper illustrates the use of data science tools to retrieve and analyze Internet data on dwelling prices and characteristics in the housing markets of the main Colombian cities. We expect that this new data set may become a useful tool for the analysis of the evolution of housing markets in our country.

Keywords: Data science, web scraping.

JEL Classification: C80, R31.

Construção de uma base de dados de preços e características de moradia para a Colômbia

Resumo

Este documento ilustra a utilização de técnicas da ciência de dados para a criação e análise de uma base com informação de preços e características de venda ou aluguel de casas ou apartamentos nas principais cidades da Colômbia. Espera-se que esta nova base de dados, que continua a ser atualizada, possa-se converter no futuro em uma ferramenta útil para o diagnóstico precoce do comportamento e evolução do setor imobiliário no país.

Palavras-chave: ciência de dados, extração de informação automatizada.

Classificação JEL: C80, R31.

Introducción

La vivienda es un bien de consumo durable que en la gran mayoría de países constituye el activo más importante de los portafolios de los hogares (Holmes, Otero & Panagiotidis, 2011). Por este motivo, suele utilizarse como colateral para préstamos, de forma que una gran parte de los activos del sistema financiero están atados a los valores de las viviendas. En este sentido, las fluctuaciones de los precios de la vivienda pueden tener un efecto importante sobre la actividad económica nacional y regional, así como sobre la solidez del sistema financiero, lo que potencialmente puede amplificar los efectos de los choques que sufre la economía (Goodhart & Hofmann, 2007). Además, la localización de las viviendas puede tener efecto sobre el acceso a oportunidades laborales, sociales y a bienes públicos locales (O'Sullivan, 2012).

A pesar de la importancia del sector inmobiliario, es frecuente encontrar deficiencias en la disponibilidad y análisis de los precios de alquiler y venta de las viviendas, lo que dificulta la predicción de los retornos de las inversiones en el sector, así como sus posibles efectos en otros sectores de la economía (Caplin, Chopra, Leahy, LeCun & Thampy, 2008). Según DiPasquale y Wheaton (1996), la vivienda se caracteriza por ser un bien de carácter heterogéneo, en el sentido que se diferencia en términos de tamaño y otras características, así como por su localización. Mientras en otros mercados el precio se define según una cantidad fija del bien (e.g. precio por una libra de naranjas o por un galón de gasolina), en los mercados de vivienda se observa el gasto por cada vivienda individual.

Para el caso latinoamericano, la información sobre precios de vivienda está disponible para periodos relativamente cortos de tiempo en algunos países, y principalmente para grandes áreas metropolitanas. Por otra parte, no hay información detallada sobre el stock de viviendas disponibles, ni sobre la actividad constructora (Cubeddu, Tovar Mora & Tsounta, 2012). De esta forma, el conocimiento sobre el mercado inmobiliario es limitado, lo cual dificulta su análisis.

En Colombia, las principales fuentes de información de acceso público se resumen en el Índice de Precios de Vivienda Usada (IPVU), producido por el Banco de la República, y en el Índice de Precios de Vivienda Nueva (IPVN), producido por el Departamento Administrativo Nacional de Estadística. De acuerdo con Galvis y Carrillo (2013), estos indicadores no tienen en cuenta la heterogeneidad de las viviendas, lo cual dificulta, por ejemplo, la construcción de índices entre distintas ciudades. Además, los índices son calculados a un nivel de agregación bastante alto y no necesariamente son representativos en algunas ciudades o zonas debido a problemas de muestra pequeña.

De acuerdo con lo anterior, es clara la necesidad e importancia asociada a la construcción de una base de datos con información sobre precios y características de las viviendas en el mercado inmobiliario colombiano. Entre los trabajos que han contribuido al respecto, Jaramillo y Cuervo (2014) presentan información de precios de la vivienda en Bogotá para el periodo comprendido entre 1970 y 2013. Estos autores presentan series de precios de alquiler de vivienda, compraventa de vivienda usada y de tierra destinada a vivienda. Para la información de precios, los autores utilizaron las ofertas publicadas en los periódicos, mientras que los precios del suelo provienen de la Lonja de Propiedad Raíz.

Una aproximación similar es la de Garza (2007), en la que se hace uso de la publicidad que realiza la revista *La guía inmobiliaria* desde el mes de mayo de 1992 hasta abril de 2004. Con dicha información, se construye una base de datos georreferenciada, que permite evidenciar una relación de causalidad que va del precio por metro cuadrado al área en metros cuadrados construidos.

En este punto es importante mencionar que los trabajos anteriores tienen las limitaciones propias de utilizar datos recolectados de forma manual, que se refleja en posibles sesgos de selección, limitación en términos del volumen de datos obtenidos, de la frecuencia de recolección, así como de posibles errores al momento de transcribir la información. Para subsanar, o al menos minimizar, estos inconvenientes, en el presente trabajo se utilizan herramientas desarrolladas para el análisis de grandes volúmenes de datos, también conocidas como *big data* (Edelman, 2012). En una ilustración reciente, para estudiar el comportamiento del mercado de la vivienda en diversas áreas metropolitanas en Estados Unidos, Boeing y Waddell (2016) recolectaron información de cerca de 11 millones de viviendas en arriendo del sitio de Internet Craigslist. A partir de un análisis detallado de dicha información, los autores encontraron evidencia de la existencia de correlación a nivel espacial, así como temporal.

Efthymiou y Antoniou (2013), en su análisis de los efectos de la infraestructura y políticas de transporte sobre los precios de venta y alquileres de la vivienda en Atenas, aplican técnicas de extracción de información (*web scraping*) para obtener datos sobre las transacciones inmobiliarias. En este trabajo se especifican y estiman diversos modelos de precios hedónicos mediante técnicas de econometría espacial, que permiten concluir que las variaciones en el precio de las viviendas se ven enormemente afectadas por los diversos tipos de infraestructura de transporte.

Por su parte, Caplin et al (2008) buscan mejorar la predicción de los retornos de la vivienda en la ciudad de Los Ángeles, teniendo en cuenta consideraciones de tipo espacial. Para esto, utilizan técnicas de aprendizaje automático

(*machine learning*), con el fin de eliminar posibles errores de predicción en tiempo real. De lo anterior se concluye que obtener y analizar información de los precios y características de las viviendas, como tipo de inmueble, número de habitaciones y baños, área construida y ubicación geográfica, entre otras, puede llegar a ser de suma utilidad para la investigación y la política pública en temas como la determinación de la calidad de las viviendas, así como para cuantificar la magnitud del déficit cualitativo de vivienda por sector.

También es importante destacar que la posibilidad de que la información de los inmuebles esté georreferenciada facilita su utilización en conjunto con información del mismo tipo disponible en otras bases de datos, de forma tal que se puedan contrastar hipótesis relacionadas con patrones espaciales presentes en el mercado de vivienda. Así, existe el potencial, por ejemplo, de facilitar la determinación de las ubicaciones óptimas de servicios como educación y salud. Estos aspectos se relacionan con el auge del concepto conocido como ciudad inteligente, definida como un área urbana en la que los sectores público y privado colaboran para lograr resultados sostenibles mediante el análisis de información contextual en tiempo real y compartido (Sevillano Pérez, 2015).

El objetivo de este documento es presentar una base de datos con información de precios y características de venta o arriendo de casas o apartamentos en las principales ciudades de Colombia. El análisis de la información contenida en la base de datos ofrece la posibilidad de efectuar diagnósticos tempranos del comportamiento y evolución de la economía colombiana, en lugares donde hay escasez, demoras y altos costos de recolección de este tipo de información. La metodología que se utiliza está basada en la extracción, compilación y limpieza en tiempo real de anuncios publicados en páginas de Internet especializadas en el mercado inmobiliario, lo que ofrece una alternativa rápida y eficiente en comparación con métodos manuales establecidos para el análisis de los niveles de precios en el mercado inmobiliario y su heterogeneidad.

Los registros recolectados de forma automática minimizan errores de medición de las variables y reducen los riesgos asociados a posibles sesgos de selección de la información. Como resultado, la analítica de datos transforma datos generados en tiempo real en información con la que se pueden identificar tendencias y patrones, además de proporcionar capacidades de predicción, mejorando la toma de decisiones en la ciudad.

Los anuncios de estas páginas hacen parte de lo que se conoce como Información Geográfica Voluntaria (*Volunteered Geographic Information*, *vgi*), definida como aquel contenido generado por el usuario y geolocalizado. De esta forma, esta fuente alternativa de información se destaca por ofrecer una

gran cantidad de observaciones, y por ser muy detallada espacial y temporalmente, convirtiéndose en una herramienta adicional con un potencial importante para la planificación e investigación urbana (Boeing & Waddell, 2016).

La estructura de este documento es la siguiente. En la primera sección se explica el proceso de recolección de información, incluyendo las frecuencias de recolección para cada página web. En la segunda, se realiza una descripción de las bases construidas, detallando las variables, la distribución espacial de las observaciones y las principales características estadísticas diferenciadas por ciudad, tipo de inmueble y tipo de negocio. En la tercera sección, se mencionan algunas características adicionales que muestran el potencial que tienen las bases para futuras investigaciones. En la cuarta sección se resaltan las limitaciones de las fuentes de información utilizadas. Finalmente, en la quinta sección se ofrecen algunas conclusiones.

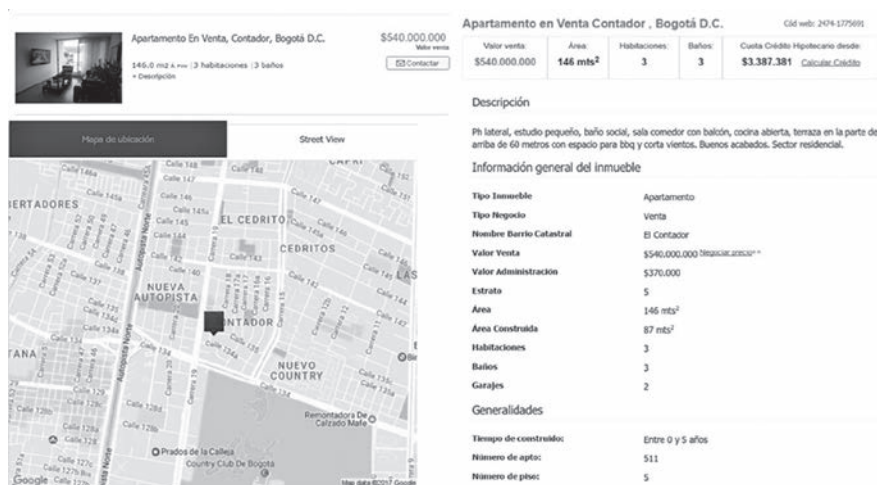
1. Recolección de la información

La extracción de información de precios y características de la vivienda (en venta o alquiler) anunciada en Internet mediante la utilización de métodos automatizados reduce los tiempos de recolección, agilizando los tiempos de respuesta de investigadores, analistas y hacedores de política económica frente al comportamiento y evolución del sector inmobiliario. De esta forma, se tienen datos detallados y georreferenciados en tiempo real a un costo relativo bajo. Estas características permiten realizar análisis precisos y oportunos para lograr diagnósticos tempranos de la economía colombiana.

Debido a que una página web es principalmente un texto semiestructurado, se utiliza la técnica de *web scraping* que permite recolectar la información relevante para una investigación, valiéndose de técnicas de minería de datos y análisis de expresiones regulares (Munzert, Rubba, Meißner, & Nyhuis, 2014). Estas técnicas se aplican para extraer datos sobre precios, ubicación geográfica y características del inmueble como área, estrato, número de habitaciones y baños, estado, antigüedad, tipo de inmueble, entre otras. Para este trabajo, se recolectan y analizan los precios de venta y arriendo de casas y apartamentos nuevos y usados, presentes en las principales páginas web de anuncios de vivienda. Aunque la recolección de la información se hace para las cinco principales ciudades del país (Bogotá, Medellín, Cali, Barranquilla y Bucaramanga), en este documento por motivos de espacio únicamente se presentará información detallada para la ciudad de Bogotá.

En la figura 1 se observa la forma en que es presentada la información en las páginas, con sus respectivos anuncios, así como el contenido al que se accede al seleccionar los vínculos disponibles. Se observa que pueden ex-

traerse los precios, las características de la vivienda e incluso sus coordenadas geográficas. Debido a que ciertos anuncios pueden estar disponibles en un periodo de tiempo corto (por ejemplo, menor a un mes), los datos fueron descargados cada 15 días para recolectar el mayor número de información posible de las páginas de Internet.



Apartamento En Venta, Contador, Bogotá D.C.

\$540.000.000
Valor venta

146,0 m² 4 m² 3 habitaciones 3 baños
+ Descripción

[Ver Foto](#)

Mapa de Ubicación

Street View

Descripción

Ph lateral, estudio pequeño, baño social, sala comedor con balcón, cocina abierta, terraza en la parte de atrás de 60 metros con espacio para bñq y corta vientos. Buenos acabados. Sector residencial.

Información general del inmueble

Tipo Inmueble	Apartamento
Tipo Negocio	Venta
Nombre Barrio Catastral	El Contador
Valor Venta	\$540.000.000 (Impuesto predial)
Valor Administración	\$370.000
Entrada	5
Área	146 m ²
Área Construida	87 m ²
Habitaciones	3
Baños	3
Garajes	2

Generalidades

Tiempo de construcción:	Entre 0 y 5 años
Número de aptos:	511
Número de pisos:	5

Figura 1. Ejemplo de un anuncio y su contenido

Fuente: Principal página para oferta de viviendas en internet.

Una vez descargada la información de cada mes, se compilaron los datos en una base de datos para cada una de las ciudades (Bogotá, Barranquilla, Bucaramanga, Medellín y Cali), para cada tipo de inmueble (apartamentos y casas) y para cada tipo de negocio (venta y arriendo). Consecuentemente, en total se construyeron veinte bases de datos.

Debido a que la información se descargó cada 15 días, es posible que se encuentre información duplicada por mes. Además, las personas o las empresas pueden publicar el mismo anuncio de venta o alquiler del inmueble en varias páginas simultáneamente. Como consecuencia de la posible existencia de anuncios duplicados, al interior de un mismo sitio web o entre varios sitios web, se realizó un proceso para su eliminación. Para cada una de las bases por ciudad y tipo de inmueble se eliminaron todas aquellas observaciones duplicadas según el precio, área construida, área privada, localización (latitud y longitud), número de habitaciones, mes de descarga de la información y, para el caso de apartamentos, el número del piso. Con este procedimiento minimizamos la posibilidad de registros duplicados en la información de precios de vivienda.

Una vez realizado el proceso de limpieza y eliminación de duplicados obtuvimos las bases de precios de vivienda para su respectivo análisis.

2. Descripción de la información

De las páginas se construyeron, por ciudad, cuatro bases de datos para cada combinación de tipo de vivienda (apartamento o casa) y negocio (venta o arriendo), obteniendo las siguientes variables:

- Mes: mes de recolección.
- Día: día de recolección.
- Año: año de recolección.
- Id : identificador único del inmueble durante una fecha de recolección específica.
- Longitud: coordenada geográfica.
- Latitud: coordenada geográfica.
- Actualización: código interno de la página de la última actualización de la información de la vivienda.
- Código: código interno de la página para identificar la vivienda específica.
- Precio: precio de venta o de alquiler del inmueble (pesos colombianos).
- Área Const: Área construida, i.e. área incluyendo muros divisorios y de fachada, ductos y estructura (columnas o muros estructurales); en caso de compartir paredes con algún vecino, se mide desde la mitad del muro divisor (en metros cuadrados).
- Área privada: Área privada, i.e. área excluyendo muros divisorios y de fachada, ductos y estructura (en metros cuadrados).
- precio m²: precio por metro cuadrado, el cual aparece por defecto en el anuncio en pesos.
- Estrato: estrato al que pertenece la vivienda.
- Habitaciones: número de habitaciones.
- Baños: número de baños.
- Estado: estado en el que se encuentra la vivienda.
- Antigüedad : antigüedad aproximada en años.
- Garajes: número de garajes, si existen.
- Sector: sector de la ciudad en que está la vivienda.
- Piso: piso en el que se ubica la vivienda, usualmente aplicable para apartamentos.
- Administración: si aplica, cuota de administración que debe pagarse mensualmente.

- Tipo: tipo de vivienda, por ejemplo, si es apartamento duplex, penthouse, entre otros.
- Inmobiliaria: empresa inmobiliaria encargada de ofrecer la vivienda.

A partir de la información obtenida, se construyeron los mapas, tablas y las series temporales mensuales para cada tipo de inmueble y ciudad.

2.1. Georreferenciación: distribución espacial de los inmuebles

Como se mencionó, una característica relevante de estas bases de datos es que permiten conocer la localización exacta de los inmuebles ofrecidos para la venta o arriendo en las diferentes ciudades en consideración. Esto facilita el análisis de los precios de los inmuebles por zonas (barrios) y su concentración en tiempo real. Además, con esta información se puede determinar el impacto de ciertos eventos sobre el precio de las viviendas en una zona determinada.

En particular, en las figuras 2, 3, 4 y 5 se puede observar la distribución de los apartamentos y casa en venta o arriendo para la ciudad de Bogotá. Como se puede observar en estos mapas los inmuebles con mayores precios se encuentran en el noreste de la ciudad.

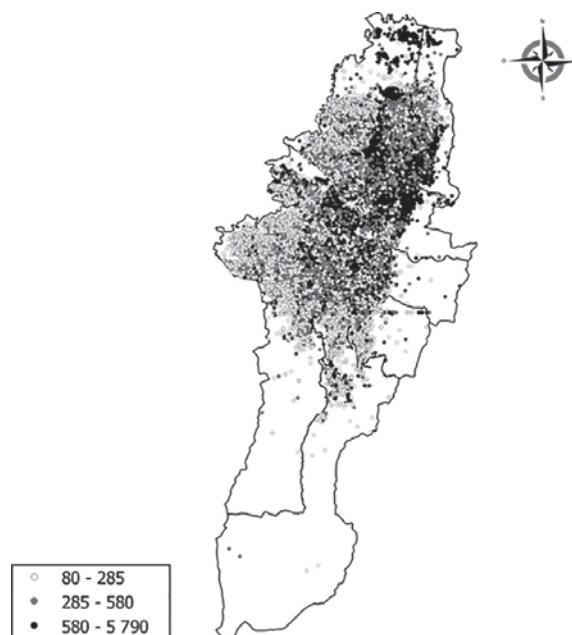


Figura 2. Apartamentos en venta en Bogotá (millones de pesos)

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet.

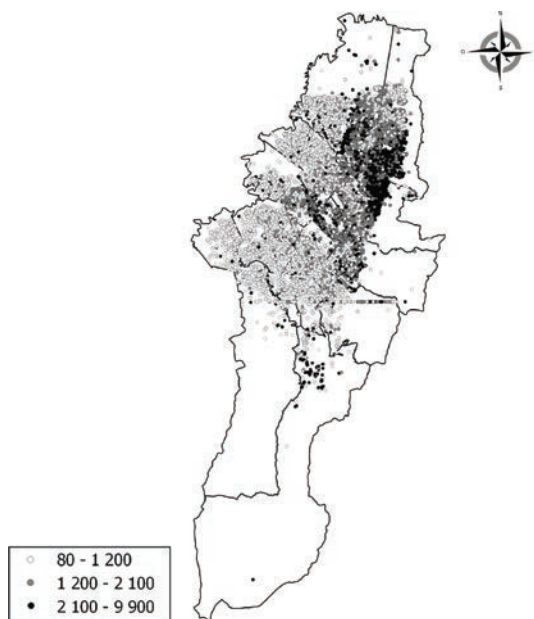


Figura 3. Apartamentos en arriendo en Bogotá (miles de pesos)

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet.

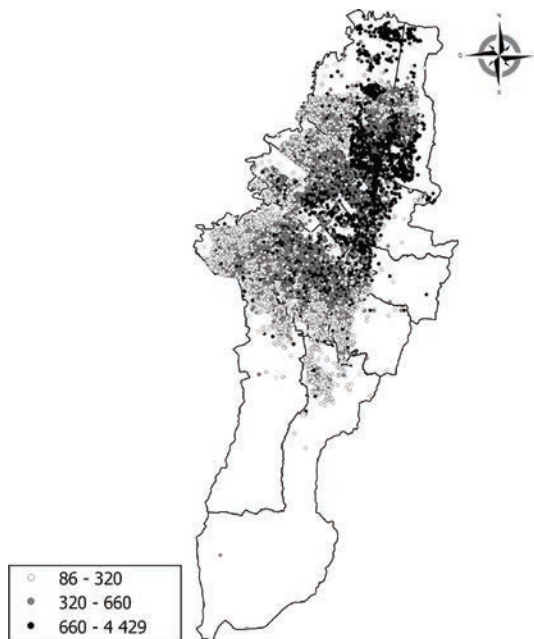


Figura 4. Casas en venta en Bogotá (millones de pesos)

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet.

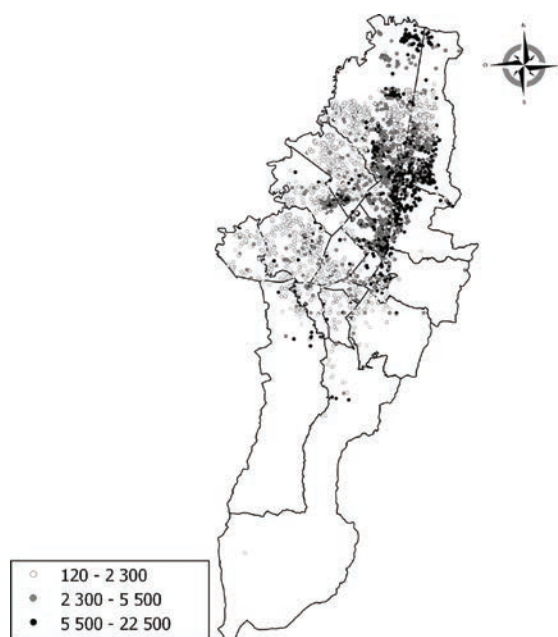


Figura 5. Casas en arriendo en Bogotá (miles de pesos)

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet.

Es importante resaltar que el mayor número de viviendas ofrecidas se encuentra en aquellas partes donde el precio de la vivienda es relativamente mayor. Esto es debido a que las viviendas con mayores precios tienden a ser ofrecidas más frecuentemente en sitios web, mientras que las viviendas con menores precios o en sectores más informales, los anuncios se publican a través de otros medios, como, por ejemplo, el voz a voz). A pesar de esto, se tiene disponible un número considerable de inmuebles ofrecidos incluso en zonas de menores ingresos.

2.2. *Análisis de los precios de vivienda*

2.2.1. Apartamentos en venta en Bogotá

Para la ciudad de Bogotá, el número total de observaciones para apartamentos en venta, correspondiente al periodo de diciembre de 2016 a junio de 2018, fue de 403.378 inmuebles (ver cuadro 1). Así mismo, la mediana del precio de los apartamentos en venta para este periodo fue de 400 millones de pesos, la del área construida fue de 92 m² y la del número de habitaciones fue de 3.

En la figura 6 se puede observar la evolución de los precios de los apartamentos en venta para las diferentes ciudades. En el caso de Bogotá, el precio de los apartamentos en venta ha tendido a aumentar ligeramente entre diciembre de 2016 y junio de 2018.

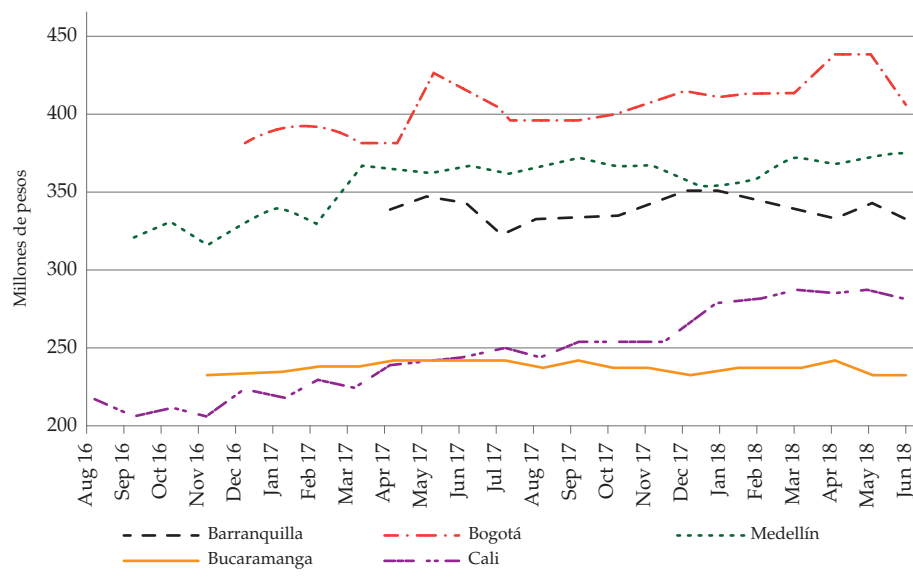


Figura 6. Series de mediana del precio de apartamentos en venta

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet. A medida que se elaboraban los códigos se empezó la recopilación de los datos. Para algunas ciudades los códigos se hicieron más temprano que para otras, esto explica por qué para algunas ciudades las series de precios comienzan antes o después de otras ciudades.

En la figura 7 se presenta la evolución de la mediana de los precios por metro cuadrado de los apartamentos en venta para las diferentes ciudades. En la ciudad de Bogotá, el precio por metro cuadrado para un apartamento en venta está alrededor de los 4 millones de pesos, el cual es relativamente más alto en comparación con las demás ciudades analizadas; sin embargo, no se observa una tendencia creciente o decreciente en el nivel de precios de apartamentos en venta por metro cuadrado en la ciudad de Bogotá en el periodo analizado.

2.2.2. Apartamentos en arriendo en Bogotá

El número total de observaciones para apartamentos en arriendo en Bogotá, correspondiente al periodo de diciembre de 2016 a junio de 2018, fue de 187.305 inmuebles (ver cuadro 1). La mediana del precio de los apartamentos

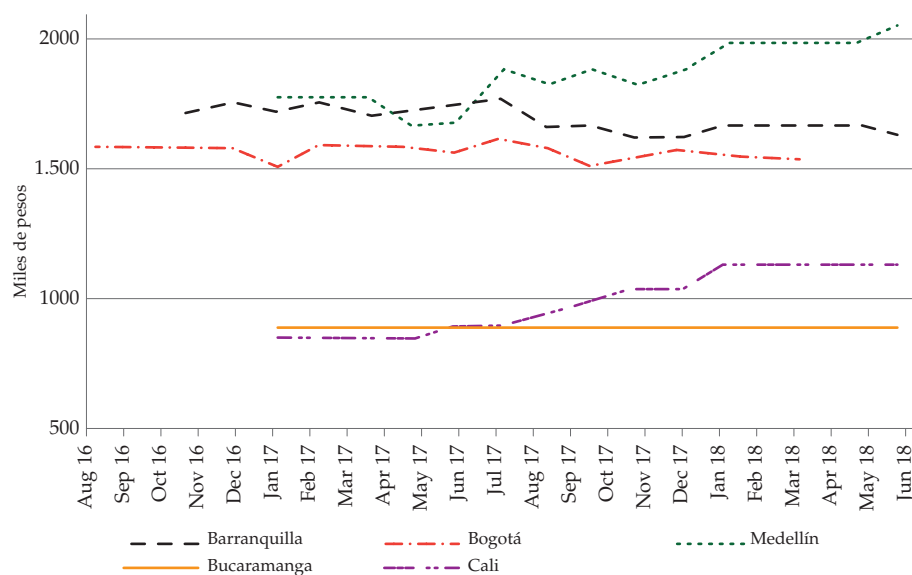


Figura 7. Series de mediana del precio/m² de apartamentos en venta

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet. A medida que se elaboraban los códigos se empezó la recopilación de los datos. Para algunas ciudades los códigos se hicieron más temprano que para otras. Esto explica por qué para algunas ciudades las series de precios comienzan antes o después de otras ciudades.

en arriendo fue de 1'592.000 pesos, la del área construida fue de 72 metros cuadrados y la del número de habitaciones fue de 3.

En la figura 8 se puede observar la evolución de los precios de apartamentos en arriendo. En el caso de Bogotá, el precio de los apartamentos en arriendo se ha mantenido relativamente constante alrededor de 1'592.000 pesos mensuales para el periodo de diciembre de 2016 a junio de 2018.

Según la figura 9, en la ciudad el precio por metro cuadrado para un apartamento en arriendo está alrededor de 22 millones de pesos, el cual es relativamente elevado en comparación con las demás ciudades analizadas. Además, no se observa una tendencia creciente o decreciente en el nivel de precios de arriendos por metro cuadrado durante el periodo analizado.

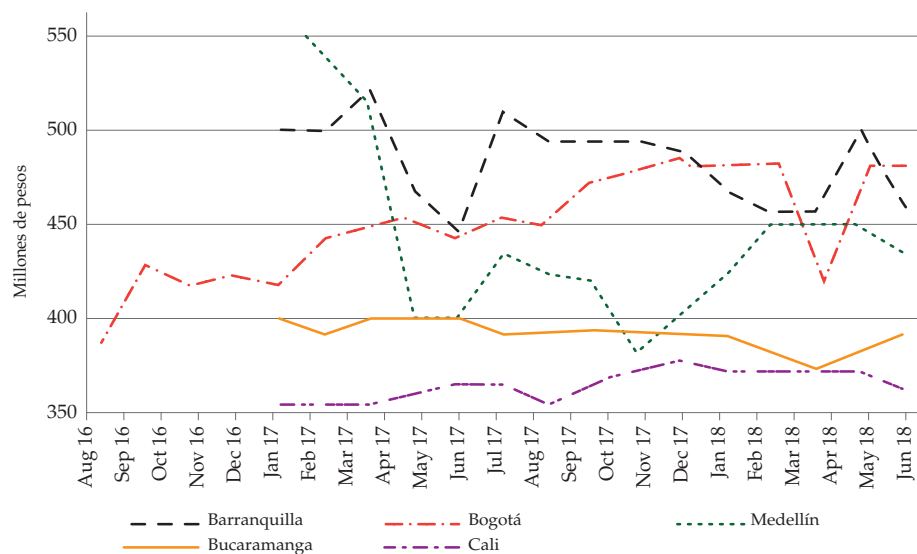


Figura 8. Series de mediana del precio de apartamentos en arriendo

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet. A medida que se elaboraban los códigos se empezó la recopilación de los datos. Para algunas ciudades los códigos se hicieron más temprano que para otras. Esto explica porque para algunas ciudades las series de precios comienzan antes o después de otras ciudades.

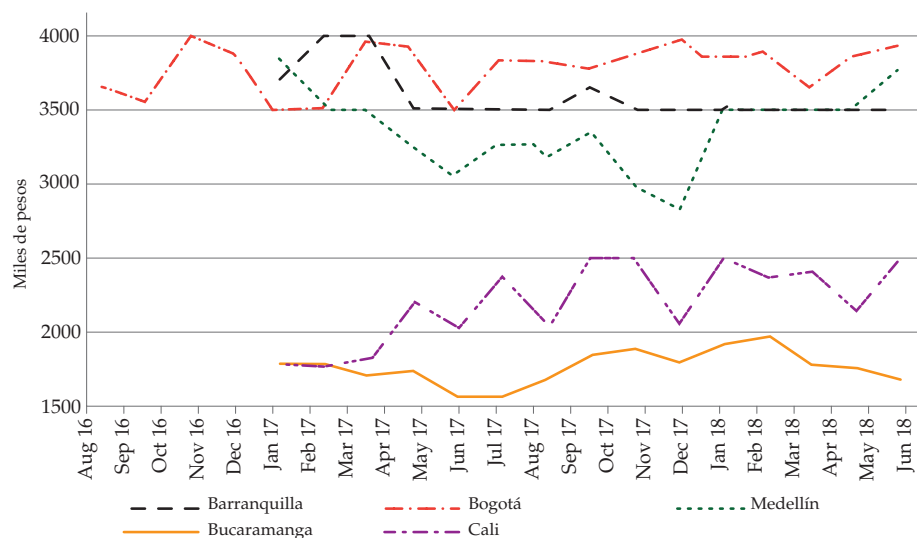


Figura 9. Series de mediana del precio/m² de apartamentos en arriendo

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet. A medida que se elaboraban los códigos se empezó la recopilación de los datos. Para algunas ciudades los códigos se hicieron más temprano que para otras. Esto explica por qué para algunas ciudades las series de precios comienzan antes o después de otras ciudades.

2.2.3. Casas en venta en Bogotá

Con respecto a las casas, el número total de observaciones para las casas en venta en Bogotá, correspondiente al periodo de diciembre de 2016 a junio de 2018, fue de 170.051 inmuebles (ver tabla 1). La mediana del precio de las casas en venta fue de 450 millones de pesos, la del área construida fue de 175 metros cuadrados y la del número de habitaciones fue de 4.

En la figura 10 se puede observar la evolución de los precios de las casas en venta para las distintas ciudades. En el caso de Bogotá, el precio ha tendido a incrementarse en el periodo analizado.

Cuadro 1: Estadísticas de tipos y cánones de vivienda en Bogotá

Vivienda y negocio	Número de observaciones	Mediana del precio	Mediana área construida	Mediana de habitaciones
Apartamentos en venta	403 378	400 000 000	92	3
Apartamentos en arriendo	187 305	1 592 000	72	3
Casas en venta	170 051	450 000 000	175	4
Casas en arriendo	26 916	3 710 000	220	4

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet.

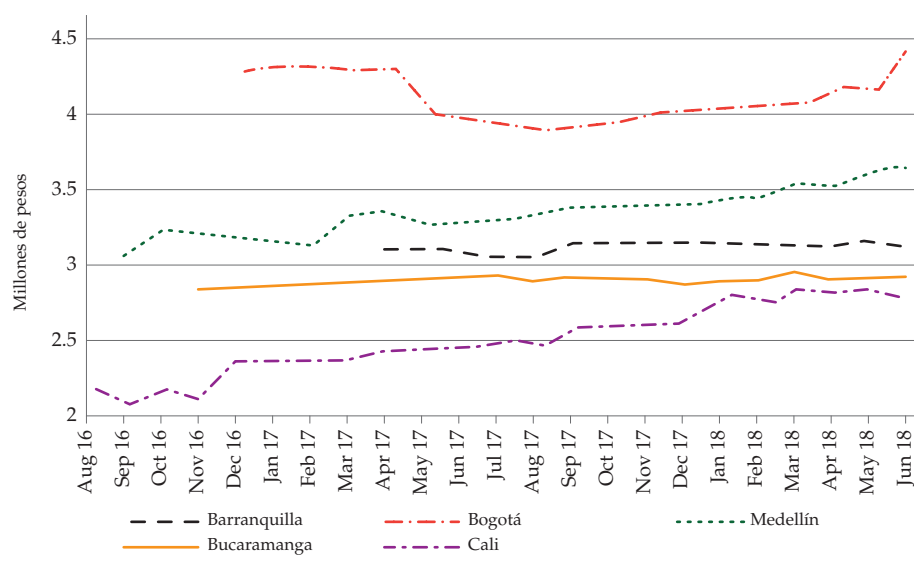


Figura 10. Series de mediana del precio de casas en venta

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet. A medida que se elaboraban los códigos se empezó la recopilación de los datos. Para algunas ciudades los códigos se hicieron más temprano que para otras. Esto explica por qué para algunas ciudades las series de precios comienzan antes o después de otras ciudades.

Según la figura 11, en la ciudad el precio por metro cuadrado para una casa en venta es relativamente elevado en comparación con las demás ciudades analizadas. Además, se observa una tendencia creciente en el nivel de precios de las casas en venta por metro cuadrado en el periodo analizado.

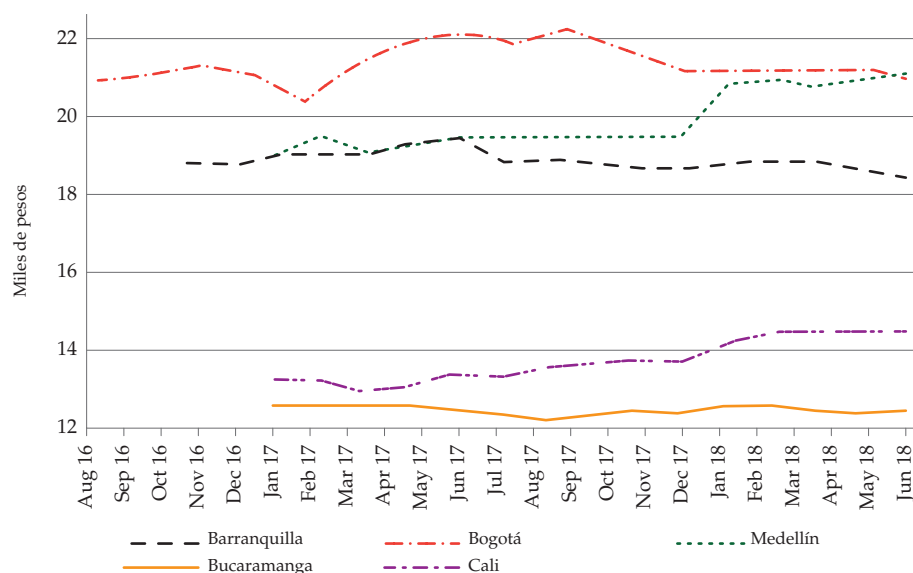


Figura 11. Series de mediana del precio/m² de casas en venta

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet. A medida que se elaboraban los códigos se empezó la recopilación de los datos. Para algunas ciudades los códigos se hicieron más temprano que para otras. Esto explica por qué para algunas ciudades las series de precios comienzan antes o después de otras ciudades.

2.2.4. Casas en arriendo en Bogotá

Para las casas en arriendo en Bogotá, el número total de observaciones fue de 26.916 en el periodo de diciembre de 2016 a junio de 2018 (ver tabla 1). La mediana del precio de las casas en arriendo fue de 3'710.000 pesos, la del área construida fue de 220 metros cuadrados y la del número de habitaciones fue de 4.

En la figura 12 se puede observar la evolución de los precios de las casas en arriendo para las distintas ciudades. En el caso de Bogotá, el precio de las casas en arriendo se ha mantenido relativamente constante alrededor de 3'710.000 pesos mensuales en el periodo comprendido entre diciembre de 2016 y junio de 2018.

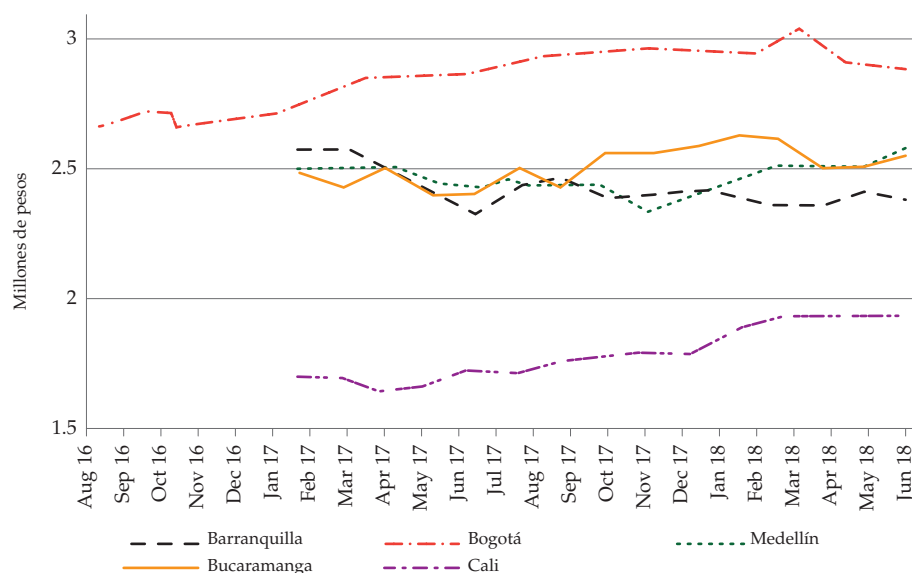


Figura 12. Series de mediana del precio de casas en arriendo

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet. A medida que se elaboraban los códigos se empezó la recopilación de los datos. Para algunas ciudades los códigos se hicieron más temprano que para otras. Esto explica por qué para algunas ciudades las series de precios comienzan antes o después de otras ciudades.

Según la figura 13, el precio por metro cuadrado para una casa en arriendo está alrededor de 17 millones de pesos mensuales, el cual es relativamente elevado en comparación con las demás ciudades analizadas. Además, no se observa una tendencia creciente o decreciente en el nivel de precios de arriendos por metro cuadrado para la ciudad de Bogotá durante el periodo analizado.

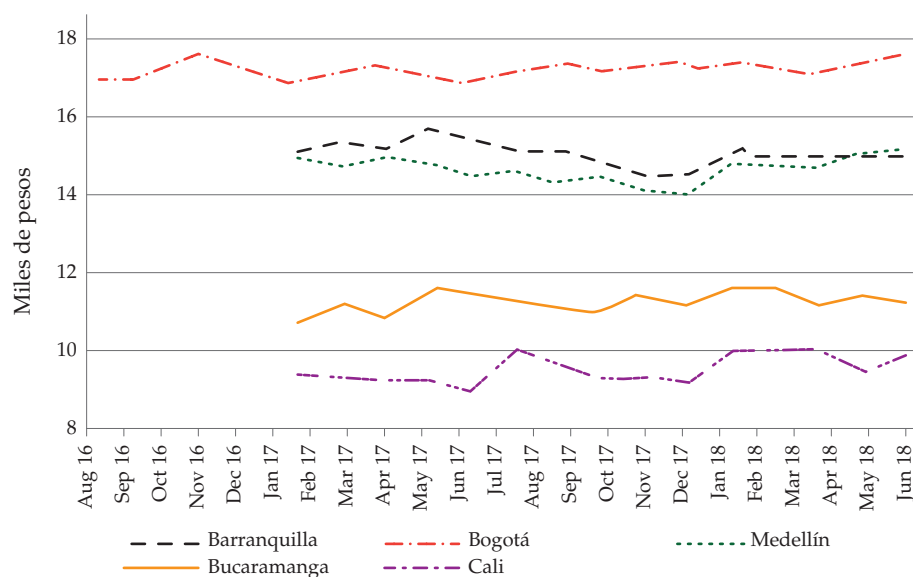


Figura 13. Series de mediana del precio/m² de casas en arriendo

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet. A medida que se elaboraban los códigos se empezó la recopilación de los datos. Para algunas ciudades los códigos se hicieron más temprano que para otras. Esto explica porque para algunas ciudades las series de precios comienzan antes o después de otras ciudades.

4. Otras características de vivienda

Como se ya se mencionó, con la información consolidada se pueden analizar diferentes características de las viviendas ofrecidas para la venta o arriendo en las diferentes ciudades estudiadas. Por ejemplo, las figuras 14, 15, 16 y 17 muestran la proporción de inmuebles según el número de habitaciones construidas a nivel de localidad en Bogotá. En la figura 14 se observa que en localidades como Bosa, Fontibón y Tunjuelito hay una mayor proporción de apartamentos para arriendo con tres habitaciones, mientras que en las localidades de Santa Fe y Chapinero predominan los apartamentos en arriendo con una habitación.

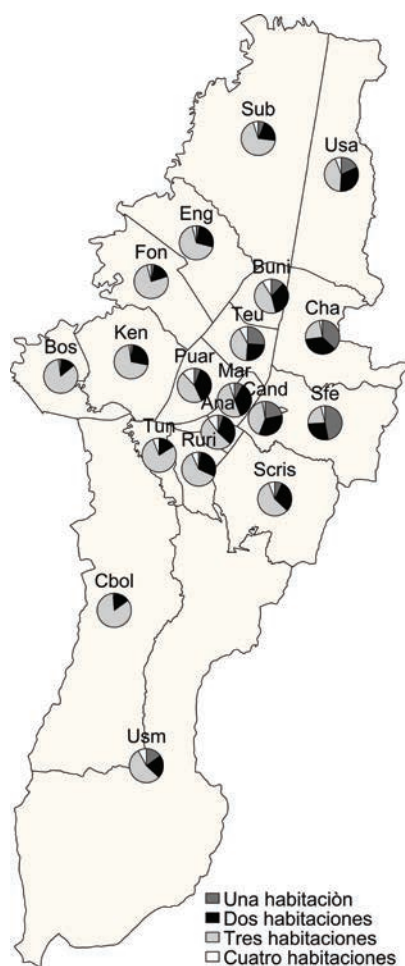


Figura 14. Apartamentos en arriendo en Bogotá por habitaciones

Nota: Suba (Sub), Engativá (Eng), Usaquén (Usa), Chapinero (Cha), Santa Fé (Sfe), San Cristóbal (Scr), Usme (Usm), Tunjuelito (Tun), Bosa (Bos), Kennedy (Ken), Fontibón (Fon), Barrios Unidos (Buni), Teusaquillo (Teu), Los Mártires (Mar), Antonio Nariño (Ana), Puente Aranda (Puar), La Candelaria (Cand), Rafael Uribe (Ruri) y Ciudad Bolívar (Cbol).

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet.

Con respecto a los apartamentos en venta se observan diferentes patrones (ver figura 16). En general, se nota que los apartamentos con tres o más habitaciones se ofrecen en mayor proporción que los apartamentos con menor número de habitaciones. De hecho, en las localidades de Bosa, Barrios Unidos, Antonio Nariño, Mártires, Puente Aranda y San Cristóbal, los apartamentos con cuatro habitaciones representan más de la mitad de los apartamentos para venta en esas localidades.

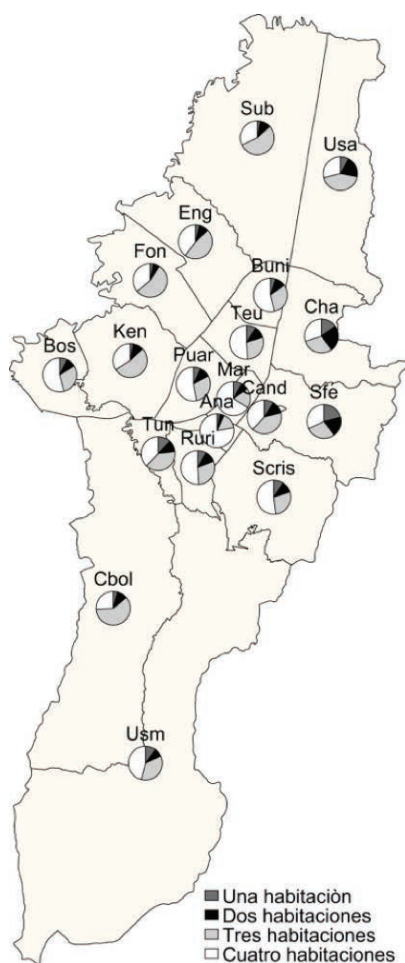


Figura 15. Apartamentos en venta en Bogotá por habitaciones

Nota: Suba (Sub), Engativá (Eng), Usaquén (Usa), Chapinero (Cha), Santa Fé (Sfe), San Cristóbal (Scr), Usme (Usm), Tunjuelito (Tun), Bosa (Bos), Kennedy (Ken), Fontibón (Fon), Barrios Unidos (Bun), Teusaquillo (Teu), Los Mártires (Mar), Antonio Nariño (Ana), Puente Aranda (Puar), La Candelaria (Cand), Rafael Uribe (Ruri) y Ciudad Bolívar (Cbol).

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet.

Por otra parte, como era de esperarse, tanto las casas en arriendo como las casas en venta tienen un mayor número de habitaciones. Por ejemplo, en localidades como San Cristóbal, Antonio Nariño, Teusaquillo, Barrios Unidos y Santa Fe más del 75 % de las casas disponibles para arrendar tenían 4 o más habitaciones (ver figura 17). Por su parte, en todas las localidades las casas en venta con 4 o más habitaciones representaron más de la mitad de los inmuebles disponibles para este tipo de negocio.

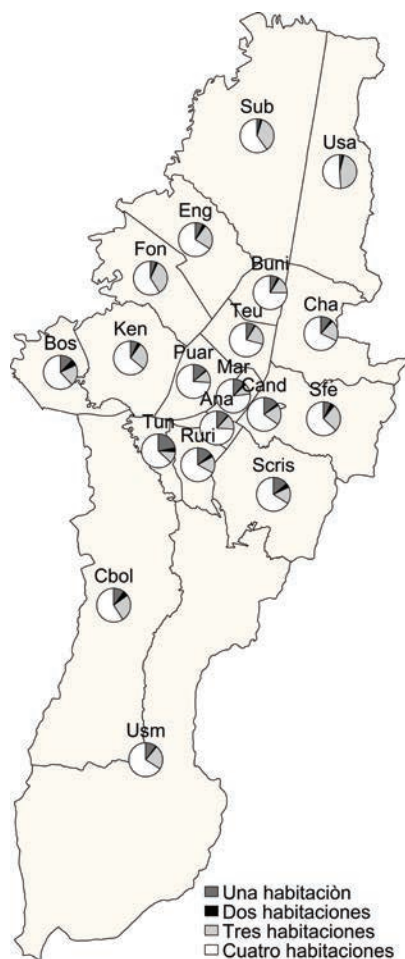


Figura 17. Casas en venta en Bogotá por habitaciones

Nota: Suba (Sub), Engativá (Eng), Usaquén (Usa), Chapinero (Cha), Santa Fé (Sfe), San Cristóbal (Scr), Usme (Usm), Tunjuelito (Tun), Bosa (Bos), Kennedy (Ken), Fontibón (Fon), Barrios Unidos (Bun), Teusaquillo (Teu), Los Mártires (Mar), Antonio Nariño (Ana), Puente Aranda (Puar), La Candelaria (Cand), Rafael Uribe (Ruri) y Ciudad Bolívar (Cbol).

Fuente: Estimaciones a partir de información de las principales páginas para oferta de viviendas en internet.

5. Limitaciones de las fuentes de información

Internet ofrece una valiosa fuente de grandes volúmenes de información detallada para analizar el mercado inmobiliario en nuestro país. Sin embargo, estas fuentes no están exentas de limitaciones. En efecto, a pesar de que se tiene información de las diferentes zonas de las ciudades, la mayoría de viviendas ofrecidas para la venta o arriendo tienden a concentrarse en las

zonas de mayores ingresos, como por ejemplo el nororiente de Bogotá. Esto no necesariamente quiere decir que en las demás zonas de las ciudades se encuentren disponibles menos inmuebles para venta o alquiler, sino que existe la posibilidad de que en estas zonas de ingresos relativamente bajos se utilicen otros medios, como el voz a voz o el uso de carteles para anunciar una vivienda. Como consecuencia, con la información recolectada se podría estar sobrestimando el valor mediano o promedio de las viviendas en arriendo o en venta para cierta ciudad.

Por otra parte, existe una diferencia entre el precio anunciado para arrendar o vender y el precio efectivamente pagado cuando se realiza el negocio. Generalmente, existe un proceso de negociación donde el precio final acordado por el inmueble es menor al anunciado. Por tanto, el nivel de precios de las viviendas usando anuncios de Internet puede sobrestimar el precio al que finalmente se lleva a cabo el negocio.

Conclusiones

La vivienda es uno de los activos más importantes de los portafolios de los hogares. Las fluctuaciones en el precio de este activo tienen efectos significativos sobre la actividad económica a nivel nacional y regional, así como sobre las decisiones de los hogares. A pesar de su importancia, la información que permite analizar la evolución del precio, localización y otras características de las viviendas es relativamente escasa.

Diferentes plataformas en Internet le facilitan al público consultar el precio y las características de las viviendas ofrecidas para la venta o arriendo en diferentes ciudades del país, reduciendo las deficiencias de información en el análisis del mercado inmobiliario. Este documento muestra cómo mediante la utilización de técnicas de recolección automatizadas (*web scraping*) y una depuración adecuada de la información, es posible consolidar una base de datos para estimar el nivel de precios de las viviendas y su evolución a lo largo del tiempo, así como patrones de concentración geográfica, entre otros. En términos del análisis de la información, se encontró que independientemente del tipo de negocio (arriendo o venta) y vivienda (casa o apartamento), Bogotá tiene los mayores precios por metro cuadrado, mientras que Bucaramanga y Cali tienen los menores precios. Sin embargo, para esta última ciudad los precios de los apartamentos en venta y arriendo han aumentado significativamente en el último año. Además, como era de esperarse, las casas tienen, en promedio, una mayor área que los apartamentos, independientemente del tipo de negocio.

Un aspecto que vale la pena destacar es la consistencia de la información contenida en la base de datos resultante. En efecto, no se observan cambios bruscos e inexplicables en las series de precios. Además, las viviendas con mayores precios se encuentran en las zonas de mayores ingresos; por ejemplo, en Bogotá los precios más altos se encontraron en el nororiente de la ciudad. En síntesis, a pesar de algunas limitaciones, la recolección y el análisis de datos de portales de Internet brinda una valiosa información del mercado inmobiliario a un costo relativamente bajo en términos de tiempo y dinero.

Agradecimientos

Los autores agradecen los comentarios de Christopher Baum, Serguei Kaniovsky y de algunos participantes del taller de Big Data organizado por la Universidad del Rosario en junio de 2017. La realización de este trabajo contó con el apoyo financiero de la Universidad del Rosario, mediante el proyecto IV-FGE002.

Descargos de responsabilidad

Los errores y omisiones son responsabilidad de los autores.

Referencias bibliográficas

- Boeing, G., & Waddell, P. (2016). New insights into rental housing markets across the United States: Web scraping and analyzing craigslist rental listing. *Journal of Planning Education and Research*, 37(4), 457-476.
- Caplin, A., Chopra, S., Leahy, J. V., LeCun, Y., & Thampy, T. (2008). Machine learning and the spatial structure of house prices and housing returns. Disponible en <http://dx.doi.org/10.2139/ssrn.1316046>
- Cubeddu, L. M., Tovar Mora, C. E., & Tsounta, E. (2012). Latin America: Vulnerabilities under construction? IMF Working Papers n.º 12/193. International Monetary Fund.
- DiPasquale, D., & Wheaton, W. C. (1996). *Urban economics and real estate markets*. Englewood Cliffs, NJ: Prentice Hall.
- Edelman, B. (2012). Using internet data for economic research. *The Journal of Economic Perspectives*, 26(2), 189-206.
- Efthymiou, D., & Antoniou, C. (2013). How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. *Transportation Research Part A: Policy and Practice*, 52(C), 1-22.

- Galvis, L., & Carrillo, B. (2013). Índice de precios espacial para la vivienda urbana en Colombia: una aplicación con métodos de emparejamiento. *Revista de Economía del Rosario*, 16(1), 25-59.
- Garza, N. (2007). Desempeño del mercado de vivienda nueva en Bogotá 1992-2004. *Cuadernos de Economía*, 26(47), 23-52.
- Goodhart, C., & Hofmann, B. (2007). *House prices and the macroeconomy: Implications for banking and price stability*. Oxford: Oxford University Press.
- Holmes, M. J., Otero, J., & Panagiotidis, T. (2011). Investigating regional house price convergence in the United States: Evidence from a pair-wise approach. *Economic Modelling*, 28(6), 2369-2376.
- Jaramillo, S., & Cuervo, N. (2014). Precios inmobiliarios de vivienda en Bogotá 1970-2013 (Inf. Téc.). Bogotá: Universidad de los Andes-CEDE.
- Lyons, R. C. (2018). *Can list prices accurately capture housing price trends? Insights from extreme markets conditions: Finance research letters*. Doi: <https://doi.org/10.1016/j.frl.2018.10.004>
- Munzert, S., Rubba, C., Meißner, P., & Nyhuis, D. (2014). *Automated data collection with R: A practical guide to web scraping and text mining*. Chichester: John Wiley & Sons.
- O'Sullivan, A. (2012). *Urban economics*. Boston: McGraw-Hill, Irwin.
- Sevillano Pérez, F. (2015). Big Data. *Economía industrial*, 395, 71-86.

