



Trabajo de Grado

**Construcción de un conjunto de microdatos
del mercado de vivienda en Colombia a partir
de información de anuncios web**

José Londoño Botero

Asesor

Santiago Bohorquez Correa

Departamento de Economía y Finanzas

Universidad EAFIT

Medellín

Mayo, 2021

1. Introducción

Actualmente no existen conjuntos de datos públicos que permitan a un investigador monitorear el estado y la evolución del mercado de vivienda en Colombia. La falta de una estructura de información organizada limita la capacidad de entender y estudiar sus dinámicas. Específicamente, un individuo que desea acceder a datos del mercado inmobiliario tiene, actualmente, tres opciones: (i) asociarse con agencias inmobiliarias que le concedan información sobre sus propiedades, (ii) solicitar información a la Cámara Colombiana de la Construcción (Cama-col)¹ para obtener datos de viviendas nuevas, o (iii) hacer uso de los datos de vivienda usada publicados por el Banco de la República de Colombia de forma trimestral. Las tres alternativas con las que cuenta el individuo están limitadas bien sea por su disponibilidad, su frecuencia o su cobertura, lo que finalmente restringe las posibilidades y el alcance de la investigación que este podría realizar.

Nuestro objetivo en este documento es construir un conjunto de microdatos de acceso libre que haga seguimiento en tiempo real a los mercados de vivienda de las principales ciudades de Colombia. Lograr este objetivo implica a su vez dar respuesta a preguntas de índole técnica, metodológica y teórica tales como: ¿Qué fuentes de información proporcionan un volumen constante y consistente de información del sector de vivienda en Colombia?, ¿Qué métodos existen para extraer esta información y cuáles son los más eficientes?, ¿Qué alcances podría tener este sistema y cómo podría articularse con información actualmente disponible? y ¿Qué tipo de información nos permitiría capturar y describir el estado del mercado de vivienda en Colombia?

En línea con nuestro objetivo, a lo largo de este artículo presentamos: (i) un algoritmo para capturar los precios y características de las viviendas en venta a partir de anuncios web, (ii) un sistema para almacenar y actualizar los microdatos capturados, de forma que cualquier persona o entidad pueda acceder a ellos fácilmente y (iii) un análisis aplicado basado en las características coyunturales del mercado inmobiliario en el momento de construcción de la base de datos. Estos sistemas se valen en gran medida del uso de la metodología de Loberto et al. (2018).

Las altas tasas de crecimiento urbano que han experimentado las principales ciudades de Colombia, en conjunto con una clara limitación territorial han incrementado de forma significativa el interés por el desarrollo urbano y la óptima distribución de las ciudades en el País.

¹La Cámara Colombiana de la Construcción es una asociación gremial que comprende empresas y personas naturales relacionadas al sector de la construcción y que tiene como fin velar por los intereses de la misma industria.

En este contexto, es cada vez más necesario ofrecer respuestas que permitan a los reguladores y hacedores de política comprender las relaciones entre los diversos agentes que interactúan en las ciudades para plantear soluciones eficientes a los problemas que pueda conllevar su crecimiento futuro. Con la creación de un conjunto de datos libre con cobertura nacional y alta frecuencia de actualización buscamos dar solución a las problemáticas actuales que limitan el avance de este campo de investigación y que no permiten ofrecer respuestas concretas a las problemáticas de distribución urbana.

Este documento está organizado de la siguiente forma: en la Sección 2 describimos las ramas de investigación del mercado de vivienda que se han destacado en los últimos años. En la Sección 3 presentamos la metodología que usamos para crear y mantener el conjunto actualizado. En la Sección 4 describimos en detalle los datos que capturamos, su fuente y su procesamiento. Finalmente, en las Secciones 5 y 6 presentamos un caso aplicado y las conclusiones de nuestro trabajo.

2. Marco teórico

El “mercado de vivienda” es realmente un conjunto de mercados que se diferencia por su geografía y sus características (Piazzesi & Schneider, 2016). En la literatura, generalmente se asume que estos mercados están compuestos por vendedores que desean deshacerse de un inmueble y por compradores que buscan adquirir una propiedad. También se ha considerado la presencia de agentes inmobiliarios que se encargan de llevar a cabo las negociaciones precisadas para cerrar los intercambios en este mercado (Jud & Frew, 1986; Yavas & Yang, 1995).

La interacción entre los agentes del mercado de vivienda puede ser comprendida a través de la aplicación de la teoría de la búsqueda. El problema del vendedor puede ser visto como uno de muestreo sin reemplazo donde el vendedor elige un comprador entre el grupo de posibles compradores y decide aceptar o rechazar su oferta; este proceso se repite hasta que se encuentra una oferta aceptable. Los compradores, por su lado, están buscando un inmueble entre un grupo de inmuebles y pueden hacer varias ofertas al mismo tiempo. Las ofertas recibidas por los vendedores son las mismas por unidad de tiempo y el dueño del inmueble debe cargar con un costo asociado a cada unidad de tiempo en que su inmueble está en venta pero no es vendido (Haurin, 1988).

En los últimos 40 años se han destacado tres ramas de investigación sobre este mercado. Sus intereses han sido, respectivamente: (i) cuantificar la contribución de las características de un

hogar a su precio de venta a través del desarrollo de modelos de precios hedónicos, (ii) comprender cuales son los factores y características que determinan el tiempo que una propiedad permanece en el mercado de vivienda y (iii) estudiar la dinámica de los mercados de vivienda, especialmente cómo responden los precios agregados del mercado a choques macroeconómicos y qué particularidades determinan la velocidad de estas respuestas. Es particularmente importante para nuestro interés que todas ellas son de carácter empírico.

El carácter empírico de las tres ramas mencionadas anteriormente hace necesario comprender que características y variables se han utilizado para acercarse a este tipo de problemas. Sirmans et al. (2006) presentaron un meta-análisis en el que sintetizaron las contribuciones de las ocho características más estudiadas por los modelos de precios hedónicos. Entre estas se encuentran el área construida, la localización del inmueble respecto al centro de la ciudad, el número de cuartos, baños y estacionamientos, entre otras. Sirmans et al. (2005) recolectaron los resultados de modelos de precios hedónicos para características más específicas como el tipo de techado, el sistema de riego del jardín, la presencia de indumentaria específica como hornos dobles y si el inmueble pertenecía a un conjunto cerrado. Investigaciones más recientes como Schlöpfer et al. (2015) y Kong et al. (2007) han evaluado el efecto que tienen características de tipo ambiental y del paisaje sobre la valoración de las viviendas.

El análisis de las *dinámicas del mercado de vivienda* hace uso de estas mismas variables, aunque su objetivo no radica en comprender como estas afectan el precio actual de las viviendas sino comprender cómo estas características determinan la velocidad de los cambios en su precio. Para hacer este contraste, los modelos suelen incluir además variables “fundamentales” como las características macroeconómicas locales y los efectos de las burbujas inmobiliarias (Bourassa et al., 2009; Capozza et al., 2002; Jud & Winkler, 2002). Los modelos de tiempo en el mercado también utilizan estas variables e incluyen una medición de las unidades de tiempo que una vivienda permanece en el mercado (Anglin et al., 2003; Haurin, 1988; Knight, 2002).

3. Metodología

Nuestro conjunto de datos es construido y actualizado en un proceso de tres etapas que se repite cada semana. Las primeras dos fueron adoptadas del artículo de Loberto et al. (2018), la tercera fue diseñada por nosotros. En la primera etapa usamos herramientas de *web scraping* para extraer información de las viviendas publicadas en *fnccaraiz.com* (FR) y *metrocuadrado.com* (MC). Paso seguido, procesamos y limpiamos los datos obtenidos en la extracción. Finalmente, añadimos estos nuevos datos a el conjunto de información recopilada

en semanas anteriores. Cárdenas-Rubio et al. (2019) presentaron un sistema similar para Colombia, pero sus datos no han sido publicados.

Extracción de los datos

Cada semana extraemos un set de características de aproximadamente 100,000 viviendas explotando la información de anuncios publicados en portales de compra y venta de inmuebles. Extraer esta información manualmente es inviable, para garantizar que esta tarea podría ser llevada a cabo consistentemente utilizamos herramientas de *web scraping*. Programamos estas herramientas para dirigirse a cada uno de los anuncios de FR y MC extrayendo las características que la literatura ha mostrado son importantes para determinar el precio y tiempo en el mercado de los inmuebles y construyendo con ellas unidades observacionales individuales. Usando esta técnica podemos recuperar información de aproximadamente 80 viviendas por minuto.

El uso de las metodologías de *web scraping* para la creación de conjuntos de datos es realmente novedoso, pues su aplicación se ha hecho factible únicamente en años recientes gracias a un veloz desarrollo de las tecnologías computacionales. Estas herramientas tienen como objetivo sistematizar la extracción de información de una, o varias, páginas web a altas velocidades². En los últimos años, los investigadores de las ciencias sociales han aprovechado estas capacidades para acceder a un universo de datos no digitalizados ni estructurados (Landers et al., 2016; Teele & Thelen, 2017) .

Procesamiento de los datos

Habiendo adquirido la información de todas las viviendas publicadas en FR y MC, procedemos a procesar los datos. Este paso es necesario para asegurar que ambas fuentes de información sean comparables. En cada extracción semanal es necesario procesar variables de texto, eliminar anuncios duplicados y crear una métrica que permita calcular el tiempo que ha permanecido cada vivienda en el mercado. Las etapas de procesamiento son descritas con mayor profundidad en la Sección 4.

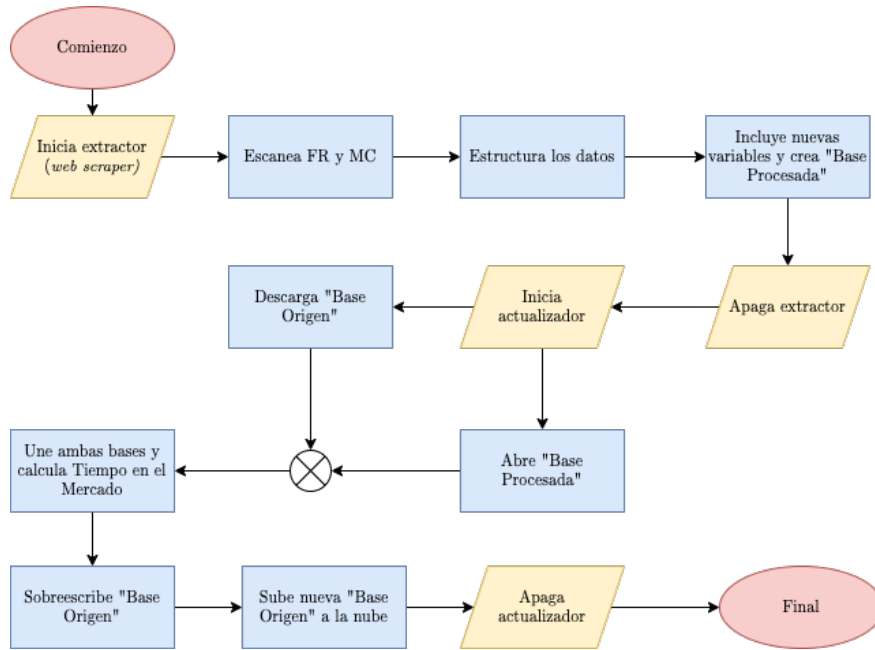
Actualización y retroalimentación del conjunto

El fin de este conjunto de datos es finalmente permitir a investigadores, y personas interesadas, conocer el estado del mercado de vivienda en cualquier momento del tiempo. Para

²Las extracciones de nuestro algoritmo son hasta 150 veces más rápidas que los métodos manuales.

asegurar que esto sea posible, hacemos mediciones semanales del mercado de vivienda y las incluimos en el conjunto de datos. En la Figura 1 presentamos el esquema general de nuestro algoritmo completo.

Figura 1: Flujo de trabajo para extraer y actualizar los datos



4. Datos

4.1. Fuente de los datos

Nuestro conjunto de datos es construido a partir de la información de anuncios publicados en los portales de venta y alquiler de inmuebles *metrocuadrado.com* (Figura 2) y *fincaraiz.com* (Figura 3). Estos sitios alojan diariamente más de 100,000 y 220,000 anuncios, respectivamente y administran las publicaciones de 1,040 y 2,488 inmobiliarias, respectivamente. Al momento de escribir este artículo, la tasa semanal de entrada y salida de anuncios en estos sitios web es todavía desconocida. Para efectos prácticos usamos como referencia una medida de 5%, que es consistente con los resultados de Loberto et al. (2018) para el mercado italiano.

Publicar un inmueble en cualquiera de estas dos plataformas requiere de la creación de un usuario, bien sea personal o de agencia inmobiliaria. Tanto FR como MC ofrecen dos publicaciones gratuitas a las cuentas personales³. Una vez superado este límite los usuarios

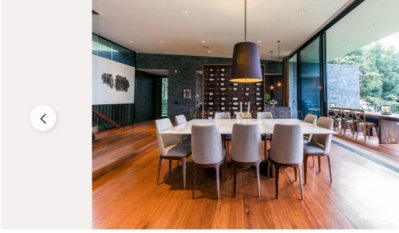
³En MC esta oferta no es válida para publicaciones ubicadas en la ciudad de Bogotá ni sus alrededores.

deben realizar pagos por cada nueva publicación para alojar su anuncio en el portal durante 90 días.

Figura 2: Anuncio en *metrocuadrado.com*

Casa en Venta, EL POBLADO LAS PALMAS
Medellín Área 734 m²

Favorito Precio **\$6.000.000.000**



1/18

Área construida **734 m²** | 4 Habitaciones | 5 Baños | 6 Estratos

Conoce este inmueble de MI HABITAT SAS

A solo 10 minutos del Centro Comercial El Tesoro se encuentra esta casa de arquitectura moderna y minimalista, ubicada en el mejor rincón de una de las más exclusivas y seguras unidades campesinas de la ciudad de Medellín. Rodeada de árboles y naturaleza, en donde podrás disfrutar de su inmenso jardín con 5.162 m² de terreno mayormente plano y sin registro. ¡Llámanos para coordinar una visita!

Reportar Compartir

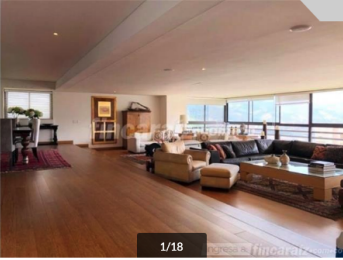
Datos principales del inmueble

| | | | |
|---------------------------------------|---------------------------------------|-----------------------------------|--------------------------------|
| Código Inmueble 3137-M2395020 | Barrio común EL POBLADO LAS PALMAS | Precio \$6.000.000.000 | Antigüedad Entre 0 y 5 años |
| Área construida 734 m ² | Área privada 734 m ² | Valor administración \$980.000 | Parqueaderos 4 |

Figura 3: Anuncio en *fincaraiz.com*

Apartamento en Venta
Medellín El Poblado **\$ 2.700.000.000**

Favoritos Usado



1/18

\$10,00 m² | Habitaciones: 4 | Baños: 5 | Parqueaderos: 4

[Ver Video](#)

Calcula tu crédito Selección un banco para calcular tu crédito

Calcular

Comparar Ver teléfono WhatsApp Ubicación

| | | |
|--|---------------------------------------|---|
| Área privada: 510,00 m ² | Área Const.: 510,00 m ² | Precio m ² : 5.294.118/m ² |
| Admón: \$1.850.000 | Estrato: 6 | Piso No: 7° |
| Sector: SurOriente | | |

[Recibir inmuebles similares](#)

Contacta al vendedor

*Nombre
*Email
Ciudad / Municipio
Teléfono

Estoy interesado/a en este Inmueble que vi en fincaraiz.com.co...

Acepto la [Política de tratamiento de datos y Términos y condiciones](#)

Quiero guardar la información de este inmueble en mi correo.

Enviar y ver teléfono

GetHome
gestión inmobiliaria

Aviso Profesional
GET HOME GESTION INMOBILIARIA
★★★★★

Cada anuncio contiene información sobre el precio y las características de un inmueble particular. Una particularidad de estos anuncios es que la información que se decide compartir

en ellos puede estar incompleta –e.g. que no se incluya el número de pisos– o, en mucha menor proporción, mal medida –e.g. que la localización del inmueble no coincida con la real–. *A priori* esto no debería generar problemas dado el alto número de observaciones y la baja proporción de anuncios con errores de medición. Para aclarar al lector, dividimos las características de los anuncios en tres grupos según el grado de precisión de medida que se puede esperar. En el primer grupo tenemos características que reflejan las cualidades exactas del inmueble o transacción, en él distinguimos el área construida, el número de habitaciones, el número de baños, el número de estacionamientos, los años del inmueble, su estrato⁴ y el ente publicador. En el segundo grupo observamos menos precisión, está compuesto por información sobre las comodidades con las que cuenta el inmueble (i.e. estacionamientos cubiertos, cuarto de servicios, aire acondicionado) y su localización (en coordenadas y por barrios, zonas y ciudades). El tercer grupo aporta información subjetiva de la vivienda como el número de fotos que fueron publicadas y una descripción general del inmueble. En el Cuadro 3 presentamos una lista completa de la clasificación de cada variable.

Las diferencias en la precisión de las medidas en cada grupo son determinadas por el sistema de incentivos de los entes publicadores. En el primer grupo están las primeras características que evalúa un posible comprador, por lo tanto no puede existir margen de error en su publicación (i.e. una familia con tres carros y cuatro hijos no estará interesada en una vivienda de dos estacionamientos y dos habitaciones). El segundo grupo de características también es objetivo, pero tiene un mayor margen de error, esto ocurre por ser variables secundarias, con un costo de medición alto, cuyas desviaciones de la realidad no tendrán un efecto preponderante en la primera impresión de los potenciales compradores. En el tercer grupo se encuentran la descripción del inmueble, que es totalmente subjetiva, y el número de imágenes que el publicador decide compartir.

4.2. Variables

Nuestro conjunto de datos captura información sobre los precios, las características y la localización de cada vivienda publicada en FR y MC. También recoge información sobre los detalles de la publicación, como los comentarios, el número de fotos publicadas y la fecha de entrada y salida del anuncio. Algunos autores incluyen el número de visitas y el número de veces que una publicación ha sido guardada, sin embargo, los portales que utilizamos en nuestro estudio no hacen pública esta información. En el Cuadro 1 presentamos un resumen de las variables capturadas.

⁴En Colombia, las viviendas son categorizadas por estratos socioeconómicos que van desde 1 hasta 6, donde 1 es la clasificación “bajo-bajo” y 6 es la clasificación “alto”.

Cuadro 1: Variables capturadas

| Tipo | Variable |
|-------------|--|
| Numérica | salePrice, rentPrice, areaBuilt, rooms, bathrooms, garages, floor, adminPrice, numPictures |
| Catagórica | source, propID, propType, businessType, stratum, cityID, cityName, zoneID, zoneName, neighborhood, commonNeighborhood, companyId, companyName, propertyState, builtTime |
| Binaria | hasBalcony, hasChimney, hasServiceRoom, hasStorageSpace, hasInterphone, hasAirConditioner, extColsedComplex, extVigilance, extGreenZones, extCoveredGarage, publishedSectorAmenities |
| Geográfica | propAddress, latitude, longitude |
| Temporal | timeMarket, firstCapture, lastCapture |
| Textual | comment, amenitiesInteriors, amenitiesExteriors, amenitiesCommonZones, amenitiesSector |

Las variables presentadas en esta tabla son explicadas en el Anexo I, Cuadro 3.

4.3. Procesamiento de los datos

Las variables observadas en nuestra base de datos no tienen la misma forma que cuando son extraídas de FR y MC. Esto se debe a que antes de ser publicados, nuestros datos pasan por un procesamiento que les dan una estructura que responde a las necesidades de investigadores e individuos interesados. Estas etapas son: (i) filtrado del tipo de información, (ii) procesamiento de las variables de texto, (iii) unión de las extracciones de FR y MC y (iv) generación de nuevas variables.

En la primera etapa incluimos dos filtros. El filtro geográfico se encarga de delimitar la extracción a viviendas localizadas en los departamentos de Boyacá, Antioquia, Valle del Cauca y Atlántico. Dada la alta concentración de publicaciones en sus ciudades capitales, estos filtros finalmente ofrecen información significativa para las ciudades de Bogotá, Medellín, Cali y Barranquilla. El filtro por tipo de inmueble asegura que solo extraemos casas y apartamentos. En la segunda etapa homogeneizamos las variables con el fin de hacer comparables las extracciones de FR y de MC. En este proceso eliminamos los caracteres especiales y trajimos variables categóricas de inglés a español. En la tercera etapa creamos un grupo de

datos a partir de la unión, y posterior *check* de duplicados, de las extracciones de FR y MC. Nuestro criterio para determinar la unicidad de una publicación basó en la comparación de las variables presentadas en el Cuadro 2. Estas variables resultan idóneas para la detección de duplicados pues están presentes en todos los anuncios son muy precisas. En la cuarta etapa generamos variables binarias que capturan la presencia de una comodidad (*amenity*) en un anuncio, el precio por metro cuadrado y variables relacionadas al Tiempo en el Mercado (*Time on Market*).

Cuadro 2: Variables de referencia para determinar duplicados

| Variables | |
|-----------|-------------|
| propType | cityName |
| rooms | salePrice |
| bathrooms | areaBuilt |
| stratum | companyName |

4.4. Periodicidad de los datos

El alto volumen de anuncios que contienen FR y MC hace del proceso de extracción una tarea exhaustiva en tiempo. Teniendo en cuenta este factor y asumiendo una tasa de rotación semanal de las propiedades del 5% encontramos que hacer extracciones semanales nos permitía mantener un alto nivel de precisión, al mismo que hacía factible ejecutar el programa en casi cualquier sistema. Este hecho implica que la variable “Tiempo en el Mercado” es medida en unidades de tiempo semanales. Debemos resaltar que *a priori* no deberían existir errores sistemáticos asociados a la selección discrecional de una unidad de tiempo específica.

4.5. Conexión con otras bases de datos

Como nuestra base de datos contiene información geoespacial esta puede ser conectada con cualquier otra base que tenga esta característica. Gracias a esta cualidad, investigadores y entidades públicas pueden analizar el mercado de vivienda en relación con información demográfica, sobre criminalidad, transporte, infraestructura pública o calidad de vida.

4.6. Limitaciones

Nuestra base de datos representa un gran avance en la disponibilidad de datos para la investigación sobre el mercado de vivienda, sin embargo, esta no está exenta de limitaciones.

Primero, los datos de estrato 1 y 2 y los datos de construcciones nuevas se encuentran subrepresentados. Para superar esta dificultad, esperamos incluir observaciones de portales con mayor presencia de estratos 1 y 2, como *porperati.com*, y portales con mayor presencia de construcciones nuevas como *coordenadaurbana.com*. Segundo, para usar nuestra medida de Tiempo en el Mercado debemos suponer que dejar de observar un determinado anuncio implica que su inmueble asociado fue vendido o arrendado. Tercero, dado que los datos de las bases originales son imputados de forma manual y que las agencias inmobiliarias no tienen incentivos para ser precisos en variables secundarias, hemos notado que existen errores en la geolocalización de algunos inmuebles.

5. Aplicación

Usando el conjunto de datos que hemos descrito en las secciones 3 y 4, un investigador o entidad interesada podrá conocer el estado y la evolución del mercado de vivienda en cualquiera de las principales ciudades de Colombia. Usando esta información podrá responder a una plétora de preguntas que respondan a necesidades públicas o personales. A continuación simulamos un corto caso aplicado de una posible aplicación.

Caso: Restricciones de movilidad en Medellín

A raíz de una “tercera ola” de contagios de Covid-19 en Medellín, el 7 de abril las autoridades comunicaron de forma abrupta la implementación de nuevas, y más estrictas, medidas de restricción de la movilidad. Dado que conocemos la fecha en la que las medidas se hicieron públicas (7 de abril) y que nuestro conjunto de datos contiene registros de una fecha anterior (22 de marzo) y una fecha posterior (15 de abril) podemos preguntarnos ¿Qué efectos tuvieron estas medidas sobre la oferta y la demanda de casas en Medellín? y ¿cómo se vieron reflejadas estos cambios en la valuación de las mismas?

Nuestra hipótesis inicial es que las nuevas restricciones de movilidad podrían reducir las expectativas de ingresos futuros de los hogares de Medellín. Esta situación podría llevar a hogares con obligaciones corrientes impagas a liquidar activos a un valor descontado para solventar sus deudas. Para evaluar esta hipótesis utilizamos datos de 4,791 casas ubicadas en la ciudad de Medellín, de las cuales 3,333 permanecieron en el mercado en ambas semanas y 1,458 solo en una de ellas. En la Figura 4 presentamos la distribución de las observaciones y los precios promedio del m² en la fecha de referencia (22 de marzo). En esta figura podemos ver que las casas en nuestra muestra están agrupadas en el suelo del valle. A través del color de relleno, que va de más claro a más oscuro a medida que el precio aumenta, vemos

que las casas en el sur tienen una valoración más alta que las casas en el norte. Esto se debe a que el estrato de las construcciones en el norte suele ser más bajo que en el sur. En la figura 5 podemos ver esta característica.

Una vez claro el panorama inicial del problema, procedemos a calcular cómo cambiaron las valuaciones de la vivienda después de que se hicieran públicas las nuevas medidas de restricción. En la Figura 6 presentamos estos resultados. En el Cuadro 4 se presenta la misma información de forma detallada. Estos datos demuestran que, contrario a nuestra hipótesis inicial, las valoraciones del mercado de la vivienda aumentaron. Este aumento generalizado en los precios podría deberse a aumentos en la demanda de hogares más amplios, fruto de expectativas de que las restricciones sean prolongadas. También podría pensarse que este incremento en la valuación de la vivienda se deba a la recuperación de la economía Colombiana. Para testear esta hipótesis analizamos los cambios en los precios del 15 de abril al 28 de abril, fechas posteriores a nuestro análisis, donde se mantuvieron las restricciones pero no hubo otro tipo de choques. En la Figura 7 presentamos estos resultados, notando que durante la segunda referencia de tiempo no hubo una dirección generalizada del cambio en los precios. Una vez planteadas estas dos hipótesis un investigador podría dar continuidad a la pregunta usando el set de datos completos para llevar a cabo pruebas estadísticas sobre ambas hipótesis.

6. Conclusiones

Antes de que creáramos este conjunto de datos no existían en Colombia fuentes de información pública que dieran la capacidad a investigadores, o individuos interesados, de monitorear el estado y la evolución del mercado de vivienda a través del tiempo. Nuestro conjunto de datos logra recuperar información sobre los precios y características de cada uno de los inmuebles publicados por los portales *metrocuadrado.com* y *fincaraiz.com* en las ciudades de Bogotá, Medellín, Cali y Barranquilla en tiempo real. Además, creamos un sistema de almacenamiento que asegura que nuestro conjunto estará actualizado y disponible para el público en todo momento. Como nuestras unidades de observación están georreferenciadas este conjunto puede ser fácilmente complementado con información de gran interés para las entidades gubernamentales, como las características demográficas de las ciudades y la información de criminalidad y calidad de vida. Con este conjunto de datos hemos buscado ofrecer a la comunidad científica una herramienta ágil y completa que le permita comprender y solucionar problemas asociados a la evolución de las ciudades Colombianas. Al mismo tiempo buscamos aumentar la transparencia del mercado de vivienda, ofreciendo a los individuos

y organizaciones interesados información sobre las valoraciones del mercado en tiempo real. En el futuro esperamos complementar esta base de datos con información de más sitios web con el fin de resolver las limitaciones de representatividad de algunos grupos.

Referencias

- Anglin, P. M., Rutherford, R. & Springer, T. M. (2003). The Trade-off between the Selling Price of Residential Properties and Time-on-the-Market: The Impact of Price Setting. *Journal of Real Estate Finance and Economics*, 26(1), 95-111. <https://doi.org/10.1023/A:1021526332732>
- Bourassa, S. C., Haurin, D. R., Haurin, J. L., Hoesli, M. & Sun, J. (2009). House price changes and idiosyncratic risk: The impact of property characteristics. *Real Estate Economics*, 37(2), 259-278. <https://doi.org/10.1111/j.1540-6229.2009.00242.x>
- Capozza, D. R., Hendershott, P. H., Mack, C. & Mayer, C. J. (2002). Determinants of real house price dynamics. *National Bureau of Economic Research*, 9262(October), 1-35. <http://www.nber.org/papers/w9262.pdf>
- Cárdenas-Rubio, J. A., Chaux-Guzmán, F. J. & Otero, J. (2019). Una base de datos de precios y características de vivienda en Colombia con información de Internet. *Revista de Economía del Rosario*, 22(1), 25. <https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.7768>
- Haurin, D. R. (1988). The Duration of Marketing Time of Residential Housing. *Real Estate Economics*, 16(4), 396-410. <https://doi.org/10.1111/1540-6229.00463>
- Jud, G. D. & Frew, J. (1986). Real Estate Brokers, Housing Prices, and the Demand for Housing. *Urban Studies*, 23(1), 21-31. <https://doi.org/10.1080/00420988620080031>
- Jud, G. D. & Winkler, D. (2002). The Dynamics of Metropolitan Housing Prices. *Journal of Real Estate Research*, 23(1), 29-45. <https://doi.org/10.1080/00420988620080031>
- Knight, J. R. (2002). Listing price, time on market, and ultimate selling price: Causes and effects of listing price changes. *Real Estate Economics*, 30(2), 213-237. <https://doi.org/10.1111/1540-6229.00038>
- Kong, F., Yin, H. & Nakagoshi, N. (2007). Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China. *Landscape and Urban Planning*, 79(3-4), 240-252. <https://doi.org/10.1016/j.landurbplan.2009.01.009>
- Landers, R. N., Brusso, R. C., Cavanaugh, K. J. & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for

- use in psychological research. *Psychological Methods*, 21(4), 475-492. <https://doi.org/10.1037/met0000081>
- Loberto, M., Luciani, A. & Pangallo, M. (2018). The potential of big housing data: an application to the Italian real-estate market. *Temì di discussione*. https://www.bancaditalia.it/pubblicazioni/temi-discussione/2018/2018-1171/en%7B%5C_%7Dtema%7B%5C_%7D1171.pdf
- Piazzesi, M. & Schneider, M. (2016). *Housing and Macroeconomics* (1.^a ed., Vol. 2). Elsevier B.V. <https://doi.org/10.1016/bs.hesmac.2016.06.003>
- Schl pfer, F., Waltert, F., Segura, L. & Kienast, F. (2015). Valuation of landscape amenities: A hedonic pricing analysis of housing rents in urban, suburban and periurban Switzerland. *Landscape and Urban Planning*, 141, 24-40. <https://doi.org/10.1016/j.landurbplan.2015.04.007>
- Sirmans, G. S., MacDonald, L., Macpherson, D. A. & Zietz, E. N. (2006). The value of housing characteristics: A meta analysis. *Journal of Real Estate Finance and Economics*, 33(3), 215-240. <https://doi.org/10.1007/s11146-006-9983-5>
- Sirmans, G. S., Macpherson, D. & Zietz, E. N. (2005). The composition of hedonic pricing models. *Journal of real estate literature*, 13(1), 1-44.
- Teele, D. L. & Thelen, K. (2017). Gender in the Journals: Publication Patterns in Political Science. *PS - Political Science and Politics*, 50(2), 433-447. <https://doi.org/10.1017/S1049096516002985>
- Yavas, A. & Yang, S. (1995). The Strategic Role of Listing Price in Marketing Real Estate: Theory and Evidence. <https://doi.org/10.1111/1540-6229.00668>

Ap ndice I

Cuadro 3: Variables capturadas

| Variable | Descripción | Precisión de medida* |
|--------------------|--|----------------------|
| source | Nombre del portal web del que fue extraído el anuncio | 1 |
| propID | Identificador único del anuncio | 1 |
| propType | Casa, Apartamento” | 1 |
| businessType | ”Venta, Arriendo” | 1 |
| salePrice | Precio bajo el cual se publicó el inmueble | 1 |
| rentPrice | Precio de arriendo (en caso de que esté disponible) | 1 |
| areaBuilt | Área construida en m2 | 1 |
| rooms | Número de habitaciones | 1 |
| bathrooms | Número de baños | 1 |
| garages | Número de estacionamientos | 1 |
| stratum | Estrato de la vivienda | 1 |
| cityID | Identificador único de la ciudad (difere según fuente) | 1 |
| cityName | Nombre de la ciudad | 1 |
| zoneID | Identificador único de la zona (difere según fuente) | 2 |
| zoneName | Nombre de la zona | 2 |
| neighborhood | Nombre de barrio oficial | 2 |
| commonNeighborhood | Nombre de barrio común | 2 |
| propAddress | Dirección de la propiedad (disponible en publicaciones de Fincaraiz) | 2 |
| floor | Número de pisos (en casas) o piso en que se encuentra (apartamentos) | 2 |
| adminPrice | Precio mensual de administración | 2 |
| comment | Descripción textual de la publicación | 3 |
| numPictures | Número de fotos publicadas en el anuncio | 3 |

| | | |
|----------------------------|--|---|
| companyId | Identificador único del individuo o inmobiliaria que publicó | 1 |
| o inmobiliaria que publicó | 1 | |
| propertyState | Üsado.º "Nuevo" | 1 |
| builtTime | Número de años desde la construcción del inmueble | 2 |
| latitude | Latitud | 2 |
| longitude | Longitud | 2 |
| hasBalcony | Propiedad tiene balcón | 2 |
| hasChimney | Propiedad tiene chimenea | 2 |
| hasServiceRoom | Propiedad tiene cuarto de servicio | 2 |
| hasStorageSpace | Propiedad tiene bodega | 2 |
| hasInterphone | Propiedad tiene citófono | 2 |
| hasAirConditioner | Propiedad tiene aire acondicionado | 2 |
| extColsedComplex | Propiedad se encuentra en unidad cerrada | 2 |
| extVigilance | Propiedad cuenta con vigilancia | 2 |
| extGreenZones | Propiedad tiene zonas verdes | 2 |
| extCoveredGarage | Propiedad cuenta con un garage cubierto | 2 |
| publishedSector.Amenities | Publicador ha incluido características del sector | 1 |
| amenitiesInteriors | Listado de características del interior | 2 |
| amenitiesExteriors | Listado de características del exterior | 2 |
| amenitiesCommonZones | Listado de características de zonas comunes | 2 |
| amenitiesSector | Listado de características del sector | 2 |
| timeMarket | Numero de semanas que ha permanecido en linea el anuncio | 1 |
| firstCapture | Fecha de primera captura del anuncio | 1 |
| lastCapture | Fecha de última captura del anuncio | 1 |

*Estos grupos corresponden a los mencionados en la Sección 4.

Cuadro 4: Cambios en el precio promedio del m² por comuna

| Comuna | Referencia | | Posterior | | Cambio | |
|---------------------------------|--------------|------|--------------|------|--------|---------|
| | Precio | Obs. | Precio | Obs. | Precio | Obs. |
| Aranjuez | \$ 2,030,451 | 74 | \$ 2,044,623 | 74 | 0.7 % | 0.0 % |
| Belén | \$ 3,308,113 | 413 | \$ 3,323,345 | 431 | 0.5 % | 4.4 % |
| Buenos Aires | \$ 2,809,650 | 158 | \$ 2,839,568 | 170 | 1.1 % | 7.6 % |
| Castilla | \$ 2,353,696 | 60 | \$ 2,374,920 | 66 | 0.9 % | 10.0 % |
| Corr. Altavista | \$ 3,750,000 | 2 | \$ 3,593,750 | 3 | -4.2 % | 50.0 % |
| Corr. San Antonio de Prado | \$ 2,437,915 | 24 | \$ 2,310,457 | 27 | -5.2 % | 12.5 % |
| Corr. San Cristóbal | \$ 1,843,248 | 7 | \$ 1,843,248 | 7 | 0.0 % | 0.0 % |
| Corr. San Sebastián de Palmitas | \$ 2,142,857 | 1 | \$ 2,650,000 | 1 | 23.7 % | 0.0 % |
| Corr. Santa Elena | \$ 5,224,108 | 88 | \$ 5,284,597 | 90 | 1.2 % | 2.3 % |
| Doce de Octubre | \$ 2,158,162 | 55 | \$ 2,199,266 | 52 | 1.9 % | -5.5 % |
| El Poblado | \$ 4,449,030 | 853 | \$ 4,529,363 | 738 | 1.8 % | -13.5 % |
| Guayabal | \$ 3,225,061 | 83 | \$ 3,236,784 | 86 | 0.4 % | 3.6 % |
| La América | \$ 3,036,978 | 209 | \$ 3,093,837 | 215 | 1.9 % | 2.9 % |
| La Candelaria | \$ 3,097,464 | 235 | \$ 3,084,296 | 243 | -0.4 % | 3.4 % |
| Laureles Estadio | \$ 3,507,021 | 374 | \$ 3,520,203 | 358 | 0.4 % | -4.3 % |
| Manrique | \$ 2,143,997 | 49 | \$ 2,206,344 | 48 | 2.9 % | -2.0 % |
| Popular | \$ 3,179,700 | 54 | \$ 3,188,098 | 50 | 0.3 % | -7.4 % |
| Robledo | \$ 2,260,796 | 144 | \$ 2,307,551 | 132 | 2.1 % | -8.3 % |
| San Javier | \$ 2,270,243 | 43 | \$ 2,419,973 | 40 | 6.6 % | -7.0 % |
| Santa Cruz | \$ 1,553,611 | 9 | \$ 1,571,458 | 12 | 1.1 % | 33.3 % |
| Villa Hermosa | \$ 2,245,163 | 65 | \$ 2,273,609 | 65 | 1.3 % | 0.0 % |

Figura 6: Cambios en el precio promedio del m2 (22 de marzo - 15 de abril)

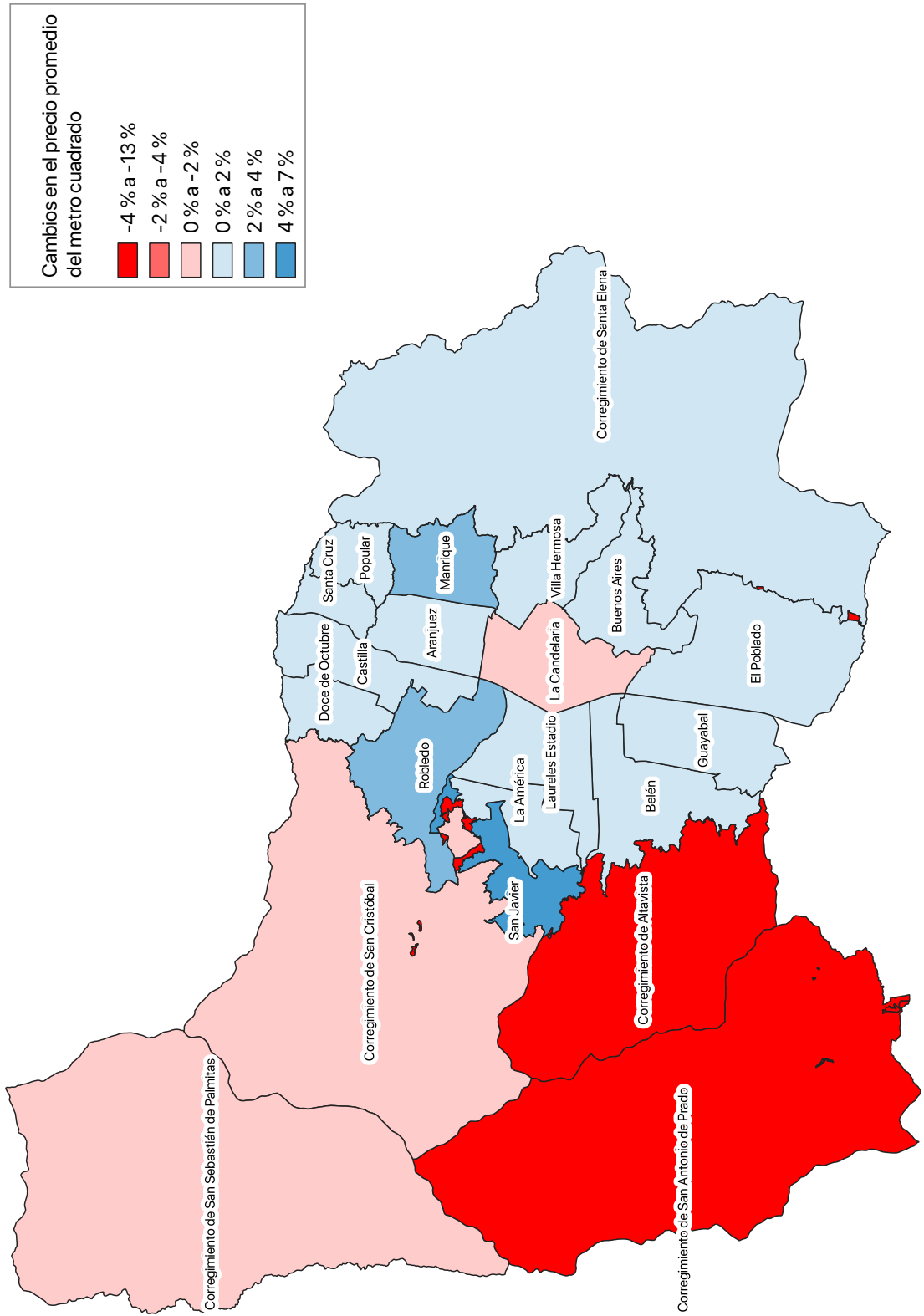


Figura 7: Cambios en el precio promedio del m2 (15 de abril - 28 de abril)

