Evaluación del efecto de la pandemia Covid-19 sobre el precio de la vivienda en la ciudad de Santiago de Cali

Juan Diego Gutiérrez Gil

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2022

Evaluación del efecto de la pandemia Covid-19 sobre el precio de la vivienda en la ciudad de Santiago de Cali

Juan Diego Gutiérrez Gil

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico

Director:
M.Sc. David Arango Londoño
Codirector:
Ph.D. Jaime Mosquera Restrepo

Universidad del Valle
Facultad de Ingeniería, Escuela de Estadística
Santiago de Cali, Colombia
2022

Dedicatoria

Este logro se lo dedico a mi familia. A mi padre, por ser mi pilar en las matemáticas, por ayudarme a comprender su naturaleza con paciencia, amor y dedicación me enseñaste algo más puro que las matemáticas, tu amor hacia nuestra familia. A mi madre, que siempre estuvo presente, ayudándome a entender mucha de la inteligencia emocional de la cual siempre debemos trabajar, enseñándome con amor, mucho del granito de arena que debemos aportar. A mi hermano, que siempre estuvo presente en los altibajos que tenía y le añade una pizca de humor a mi día a día. A Leidy Johanna, que desde el día que la conocí siempre me ha impulsado a ser mejor, a no decaer por cometer errores y aprender que con esfuerzo y empeño todo se puede lograr y al resto de mi familia que siempre me impulsa a ser una mejor persona, que el conocimiento no solo prevalece en nuestras ideas, sino, que las podamos compartir con los demás.

Resumen

La pandemia por el COVID-19 que empezó a afectar a Colombia y Santiago de Cali en marzo de 2020 llevó a que los gobiernos nacional y municipal impusieran medidas restrictivas para contener el contagio del virus, a través de diversas disposiciones sanitarias y de distanciamiento social. Estas decisiones afectaron fuertemente los sectores económicos que tienen como necesidad principal la interacción humana, siendo el sector inmobiliario uno de los más afectados.

En esta investigación se pretende evaluar el impacto de las restricciones determinadas por los gobiernos nacional y municipal para evitar la propagación del COVID-19 sobre los precios de la vivienda en la ciudad de Cali. Los datos fueron extraídos mediante un algoritmo de Web Scraping desde página web de OLX Colombia en la prepandemia y pospandemia. Haciendo uso de estos registros se construye para cada periodo un conjunto de datos con todas las filas que tienen datos completos. Adicionalmente, mediante la localización espacial del bien se complementaron los datos con covariables del entorno. Posterior a esto, se realiza el análisis exploratorio de datos y se ajusta un modelo de regresión PLS, corrigiendo la multicolinealidad presentada en las variables predictoras. Los resultados sugieren una disminución tanto del precio como de la oferta de la vivienda en el periodo pospandemia, influenciado por las covariables del entorno de la vivienda (Acceso a estaciones del MIO, cercanía a centros comerciales y cercanía a zonas verdes). Como conclusión, se tiene que en el contexto de la pandemia COVID-19 las covariables del entorno de la vivienda juegan un papel fundamental como criterio de adición de valor al precio del inmueble, lo cual impactó de manera negativa el precio de la vivienda en Santiago de Cali.

Palabras clave: Pandemia, COVID-19, precio de la vivienda, covariables de entorno, Web Scraping, PLS(Partial least Square), Regresión.

Abstract

The pandemic caused by COVID-19 that began to affect Colombia and Santiago de Cali in March 2020 led national and municipal governments to impose restrictive measures that sought to contain the spread of the virus through various health provisions and social distancing. These decisions strongly affected the sectors that have as main need the human interaction, being the real estate sector one of the most affected.

This research seeks to relate the influence of the restrictions determined by the national and municipal governments with respect to the COVID-19 on the prices of housing. The data were extracted using the Web Scraping technique on the website of OLX Colombia

for the years 2019 and 2020. Making use of these records is built for each year a dataset with all rows having complete data. Additionally, by locating the records, the data sets were supplemented with environment covariates. After this, the exploratory data analysis was performed and a PLS regression model with two components was adjusted, correcting the multicollinearity presented in the predictor variables and obtaining an optimal percentage of explained variance for the dependent variable. The results show a decrease in both the price and the supply of housing, as well as the impact on their own characteristics and their environment in the post-transitional period. As a conclusion, the restrictions implemented because of COVID-19 negatively impacted the price of housing in Santiago de Cali.

Keywords: Pandemic, COVID-19, housing price, environmental covariates, Web Scraping, PLS, Regression

Contenido

1 Introducción

El COVID-19 es la enfermedad causada por el nuevo coronavirus conocido como SARS-CoV-2 cuyos síntomas son: fiebre, tos seca, cansancio, pérdida del gusto u olfato, congestión nasal, dolor de cabeza, entre otros síntomas leves. Otros síntomas más graves son la disnea, la perdida de apetito y la confusión. Según la OMS, solo el 85 % de las personas que contraen esta enfermedad se recuperan sin necesidad de tratamiento hospitalario. El resto desarrollan una enfermedad grave donde es necesario oxígeno, inclusive pueden llegar a un estado crítico dependiendo de cuidados intensivos y empeorar hasta la muerte.

A finales de diciembre del 2019 se reportó el primer caso de COVID-19 en la República Popular de China. Hasta ese momento existía una total ignorancia sobre el daño que este virus podría llegar a causar. Su propagación fue tan rápida que China comenzó a implementar estrategias de aislamiento obligatorio y desinfección de las calles, así como el cierre total de todos los aeropuertos para evitar el contagio en otros países. Sin embargo, esto no fue suficiente, ya que se comenzaron a reportar miles de casos a nivel mundial.

La situación en Colombia no fue ajena a lo sucedido en China. El primer caso reportado en este país fue el 6 de marzo de 2020. Veinte días después se implementaron las mismas medidas de aislamiento preventivo obligatorio nombradas anteriormente para China, como también estrategias de disminución del contagio y aumento del distanciamiento, tales como: la toma de temperatura de las personas antes de ingresar a un lugar, el uso obligatorio del tapabocas, el confinamiento preventivo obligatorio en el cual se restringía la salida a solo una persona por hogar, el pico y cédula, donde las cédulas terminadas en números pares salían en días pares y así para las que terminaban en impares. Como consecuencia de estas medidas restrictivas las actividades económicas sufrieron un fuerte impacto, principalmente aquellas relacionadas con bienes o servicios no escenciales ya que tuvieron un congelamiento total de su operación Ferrari and González (2021).

Este congelamiento de las actividades económicas reflejaron en el PIB[1] del año 2020 un decrecimiento del 6.8 % respecto al año 2019. Según el DANE, la construcción fue una actividad económica que contribuyó a la disminución del PIB para el periodo pospandemia. Dicha actividad presentó un decrecimiento del 27.7 % respecto al periodo anterior, afectando directamente a los proyectos de vivienda nueva propuestos. Así pues, no solo ese sector de

[1] Producto Interno Bruto

actividad económica fue el más afectado, en el boletín de micronegocios entregado por el DANE, se registra que 509.370 micronegocios cerraron o quebraron, de los cuales el 26.9 % se ubicaban en una vivienda. Además, se señala que el 82.9 % de los micronegocios disminuyó sus ventas durante ese período de tiempo de la misma forma que aumentó el desempleo y disminuyó el salario que las personas ganaban un año antes por la cantidad de horas trabajadas. Todo este congelamiento y decrecimiento de la actividad económica afectó también a la compra y venta de la vivienda usada o nueva en Colombia, particularmente en la ciudad de Santiago de Cali, capital del departamento del Valle del Cauca.

Conocida la situación anterior, esta investigación busca relacionar la influencia de las restricciones determinadas por los gobiernos nacional y municipal para mitigar el efecto del COVID-19 sobre los precios de la vivienda usada, evaluando su efecto para los periodos prepandemia y pospandemia, los cuales corresponden a los años 2019 y 2020 respectivamente. En este trabajo de grado se construye una hoja de datos con características propias de la vivienda y su localización mediante la técnica de Web Scraping, la cual es usada para recopilar información de la página web OLX Colombia. Adicionalmente, mediante la localización espacial de los registros, se complementó la hoja de datos con las covariables del entorno del bien tales como: cercanía a estaciones del bus, centros comerciales y zonas verdes.

Para comparar los resultados obtenidos en los periodos prepandemia y pospandemia, se realizó un análisis exploratorio y descriptivo de datos, haciendo uso de las características propias de la vivienda. Posteriormente, se ajusta el modelo de regresión PLS, añadiendo las covariables de entorno de la vivienda.

Finalmente, se comparan las estimaciones obtenidas para los periodos prepandemia (Mayo - Septiembre 2019) y pospandemia (Mayo - Octubre 2020) y se evalúa el cambio que sufre el precio de la vivienda ocasionado por las restricciones implementadas en el marco de la pandemia.

2 Definición del problema

2.1. Planteamiento del problema

En marzo de 2020, el Ministerio de Salud y Protección Social declaró la emergencia sanitaria en todo el territorio colombiano, con la cual se ordenó implementar estrictas medidas de bioseguridad para evitar el contagio y la propagación del COVID-19. Conforme fue creciendo la cifra de contagiados con el virus en el país, se ordenó el aislamiento preventivo obligatorio, que restringió la circulación de las personas a casos de emergencia y al desarrollo de las actividades productivas consideradas esenciales. Las medidas tomadas por el gobierno nacional y los gobiernos locales para enfrentar la emergencia sanitaria afectaron el desempeño económico del país y la calidad de vida de sus habitantes. El sector inmobiliario, el cual representa un pilar fundamental en la economía del país, fue uno de los principales afectados por esta recesión económica, lo cual se reflejó en el precio de la vivienda.

Esta no fue la primera vez que el precio de la vivienda en Colombia sufrió una fuerte caída. En 1997 se vivió una de las crisis más grandes que ha tenido el país en el sector inmobiliario, teniendo importantes pérdidas en el índice del precio de la vivienda. De acuerdo con Aguirre (2007) el ritmo de desembolsos de créditos de vivienda por parte de los bancos especializados en cartera hipotecaria cayó dramáticamente, de niveles de 5.9 billones de pesos en 1998 a 1.2 billones de pesos en el 2003. En consecuencia, las personas que querían obtener casa propia o invertir en una vivienda, no tenían posibilidades de acceder a un subsidio, por lo tanto, el sector inmobiliario seguiría en crisis.

Actualmente, el gobierno ha implementado subsidios de vivienda que no son constituidos por bancos privados, si no por el Fondo Nacional de Vivienda (Fonvivienda). Sin embargo, estos subsidios no fueron suficientes para soportar el rezago que la pandemia estaba dejando por sus largas jornadas de confinamiento. Según el informe especial sobre el mercado inmobiliario y cartera en Colombia, generado en el segundo semestre del año 2020 Eduardo et al. (2020), se puede evidenciar como la pandemia causó para este sector un fuerte golpe económico, donde el indicador de intención de compra de vivienda nueva o usada disminuyó drásticamente en -18.6 % y su variación real anual presenta un rezago en su crecimiento a causa de la pandemia. Cabe resaltar que esta fue específicamente analizada para las tres ciudades más importantes del país, entre ellas estuvo Cali, que según el informe no tuvo rezagos en su crecimiento a causa de las normas de confinamiento que se habían propuesto

para ese entonces.

Eduardo et al. (2020) no fueron los únicos en modelar la problemática de la pandemia sobre el mercado inmobiliario, Urrea-Ríos and Piraján (2020) tomaron la iniciativa de estudiar todos los sectores que abarcan una actividad económica en el país y que por las medidas de confinamiento se vieron afectados, entre ellos el sector inmobiliario. En su estudio ponen en evidencia la caída tan fuerte y muy cercana al mínimo histórico presentado en el 2008 (1.1 %) siendo este el 2 % para los meses de marzo, abril y mayo, generado por una parálisis de la posible compra de inmuebles y materiales para construcción, como también la restricción de mudanzas.

Sin embargo, en la mayoría de estos estudios se utilizan modelos de precios de vivienda hedónicos, es decir, el planteamiento de las variables que explican la variación real anual y el indicador de intención de compra de vivienda no interactúan con ninguna variable de entorno que pueden estar asociadas a la pandemia. Por ejemplo, bajo el supuesto de que el confinamiento era reglamentario y el desconocimiento de la pandemia generó pánico, una vivienda que se encuentre cerca a un centro comercial o a una estación de bus, puede estar influenciada en su precio, por la cercanía a estos centros de aglomeración.

En la actualidad, existen pocas investigaciones en el país sobre el mercado inmobiliario relacionado con el impacto del COVID-19 en el precio de la vivienda. Para la ciudad de Santiago de Cali no se han encontrado investigaciones que estudien este impacto sobre el precio de la vivienda, como tampoco estudios que tengan en cuenta covariables de su entorno relacionadas a este contexto. Teniendo en cuenta estos factores, resulta necesario plantear la pregunta que direcciona el siguiente trabajo: ¿Cuál ha sido el impacto de las covariables de entorno en los precios de la oferta de vivienda para la ciudad de Cali, a causa de la pandemia COVID-19?

2.2. Justificación

La mayoría de estudios que se realizaron sobre el precio de la vivienda antes y después de la pandemia, tomaron como variables únicas e importantes sus características propias como lo son: el tipo de vivienda, los metros cuadrados, la cantidad de habitaciones, etc. Es por esto que el estudio se debe efectuar, no solo por la medición de los precios de la vivienda, sino por la influencia de las covariables de entorno en el contexto de la pandemia. El mercado inmobiliario guarda una fuerte asociación sobre la economía del país, por ende su estudio puede contribuir a identificar los efectos de la pandemia sobre su estructura económica.

En Colombia no existe un registro oficial del precio de la vivienda, por lo cual conocer el precio de la vivienda implica realizar un ejercicio manual de exploración en los lugares de

comercio de viviendas (páginas web). La técnica de Web Scrapping facilita esta labor, permitiendo disponer de una referencia de toda la oferta inmobiliaria. Además, como consecuencia de las restricciones determinadas por los gobiernos nacional y municipal para mitigar el efecto del COVID-19, la mayor parte del movimiento en el mercado inmobiliario en Colombia se presentó mediante plataformas digitales. La principal limitante que afrontan los ofertantes de viviendas debido a estas restricciones es el desconocimiento de la afectación del valor de su vivienda dado por su entorno, las cuales pueden sesgar la decisión de compra del demandante.

Una de las razones que hace pertinente estudiar el efecto generado por las covariables de entorno en el contexto de la pandemia COVID-19 sobre el precio de una vivienda, es poder tener una medida de referencia confiable para el inicio de su negociación, además de permitir al demandante identificar criterios estratégicos adicionales para la selección de una vivienda. En este orden de ideas, se genera el propósito de identificar los factores que más influyen en el precio de la vivienda en el contexto del COVID-19.

Es necesario profundizar en esta investigación debido al papel fundamental que tienen las covariables del entorno como criterio de adición al precio de la vivienda en el contexto del COVID-19. La cercanía a centros de aglomeración y comercio deja de ser apetecida y la tendencia de compra puede cambiar, generando valor a bienes ubicados en lugares tranquilos, con baja densidad poblacional y con buena capacidad de abastecimiento.

En resumen, con la inclusión de variables que describen el entorno de la vivienda, como lo son la cercanía a las estaciones de bus, a los centros comerciales y a las zonas verdes y la comparación de los periodos que son prepandemia (2019) y pospandemia (2020), se busca evidenciar la influencia de la pandemia sobre el precio de la vivienda, su oferta y demanda.

2.3. Objetivos

2.3.1. Objetivo General

Evaluar el efecto de la pandemia COVID-19 sobre los precios de vivienda usada en la ciudad de Cali, observando su evolución en el periodo 2019-2020 y la estructura de asociación con las covariables de entorno.

2.3.2. Objetivos Específicos

- Estudiar el comportamiento del precio de la vivienda, sus características y covariables de entorno para el periodo de prepandemia COVID-19.

- Estudiar el comportamiento del precio de la vivienda, sus características y covariables de entorno para el periodo de pospandemia COVID-19.
- Evaluar el nivel de afectación del precio de la vivienda en Cali en el contexto de la pandemia, utilizando el modelo de regresión PLS.

3 Antecedentes

Con el fin de conocer lo que se ha investigado acerca de la influencia de la pandemia COVID-19 sobre los precios de la vivienda en Colombia, se realizó una revisión bibliográfica. A continuación, se presentan algunas de las investigaciones más relevantes.

Urrea-Ríos and Piraján (2020) Realizaron un informe que demuestra el impacto de la pandemia sobre las actividades económicas más importantes del país, entre ellas el mercado inmobiliario. Realizaron el análisis de la serie temporal de la tasa de crecimiento anual para las actividades inmobiliarias, calculada mediante el indicador de seguimiento a la economía (ISE) proporcionada por el DANE. Adicional a esto, realizaron la comparación del valor mínimo histórico del índice de precio de vivienda nueva en áreas urbanas y metropolitanas frente al valor obtenido durante la pandemia. Esta metodología les permitió evidenciar que aunque las actividades inmobiliarias tuvieron un comportamiento positivo en el año 2019 frente al año 2018, para el periodo de la pandemia este sector sufrió una gran desaceleración y que esta reducción se vio explicada por las medidas de confinamiento ordenadas por el Gobierno Nacional.

Finalmente, los autores concluyen que el declive en las actividades inmobiliarias es explicado por el congelamiento de los cánones de arrendamiento, *la restricción en las mudanzas, la parálisis en los arrendamientos de locales comerciales y el cierre de negocios, restaurantes y bares.*

Castelblanco Rodriguez (2021) *en su proyecto de grado llamado 'Análisis de los efectos del COVID-19 que afectaron el mercado de inmuebles residenciales en el municipio de Cajicá Cundinamarca'*, identificó los factores que influenciaron el declive del mercado inmobiliario por la pandemia COVID-19. Haciendo uso de las proyecciones en el mercado inmobiliario según datos del informe del Banco de la República (2020) para este sector, además de información obtenida de manera directa de 15 constructoras seleccionadas de manera aleatoria en el municipio de Cajicá, construyó tablas comparativas entre los periodos prepandemia y pospandemia de la tendencia de compra inmobiliaria y la rentabilidad del sector.

Con la información recogida de las constructoras, el autor evidencia que debido al crecimiento de la modalidad de trabajo en casa causada por las restricciones impuestas por el Gobierno Nacional las personas mostraron una preferencia por los proyectos de vivienda ubicados en

sitios aledaños a la ciudad, con espacios verdes y cercanos al campo. Finalmente concluyó que la pandemia trajo consigo cambios en las preferencias de los compradores, reflejadas principalmente en la compra de inmuebles lo cual se reflejó en un declive de la tendencia de compra inmobiliaria.

Álvarez Zuluaga et al. (2022) realizaron un trabajo investigativo sobre el impacto del COVID-19 en la oferta y demanda de la vivienda nueva No VIS[1] en Colombia. En esta investigación plantean un modelo de oferta y demanda con estimaciones por mínimos cuadrados ordinarios en tres etapas (MCE3), el cual permite analizar diferentes variables independientes originando ecuaciones simultáneas además de controlar posibles problemas de endogeneidad. Para el modelo de oferta, añadieron una variable correspondiente al número de casos de COVID-19 en Colombia durante el periodo de estudio.

Al finalizar este trabajo, los autores concluyen que conforme empezó el aislamiento preventivo obligatorio y el número de contagios por COVID-19 fue incrementando rápidamente, además de las restricciones impuestas por el Gobierno Nacional, se presentó en las estimaciones un impacto negativo en la oferta de la vivienda.

Por último, **Vega-Vilca and Guzman (2011)** plantean dos metodologías para solucionar el problema de multicolinealidad en regresión múltiple, las cuales son: regresión por mínimos cuadrados parciales (PLS) y regresión por componentes principales (PCR). Los autores comparan e ilustran ambas metodologías mediante ejemplos de aplicación sobre un mismo conjunto de datos y evidencian la eficiencia de la regresión PLS sobre la regresión PCR. Este artículo es importante para este trabajo de grado porque posibilita el entendimiento de la regresión PLS en variables que presentan multicolinealidad.

En los antecedentes aquí presentados, se evidencia que se han realizado diversas aproximaciones estadísticas que intentan reflejar el impacto que tuvo la pandemia COVID-19 sobre la economía nacional y los principales sectores económicos, entre ellos el sector inmobiliario. Un hallazgo importante es que hasta el momento no existe un estudio enfocado al impacto de la pandemia sobre los precios de vivienda usada. Por otra parte, se observa que la regresión PLS es una metodología óptima para tratar problemas de multicolinealidad, lo cual la hace idónea para el presente trabajo de grado.

[1] Vivienda de Interés Social

4 Marco Teórico

4.1. Marco Conceptual

En esta sección se definen de manera conceptual las variables hedónicas y las covariables de entorno, como también el algoritmo Web Scraping utilizado para la extracción de los datos.

4.1.1. Covariables del entorno

Definidas así, porque explican el entorno en el que se encuentra un objeto de estudio, así pues, estas ayudan a entender la percepción que tiene este objeto con sus alrededores y la influencia que puede llegar a tener con respecto a la respuesta de este, por ejemplo: la cercanía a centros comerciales y el estrato socioeconómico son covariables del entorno de una vivienda.

4.1.2. Variables Hedónicas

Las variables hedónicas describen un conjunto de atributos que puede tener una propiedad inmobiliaria Poeta et al. (2019), el cual se puede resumir en dos principales grupos que representan atributos de la propia construcción (elementos físicos) y la ubicación. Los atributos físicos de la vivienda están relacionados con el tamaño (cantidad de habitaciones, baños, superficie total), diseño y el estándar de construcción, entre otros. Los aspectos de la ubicación representan las condiciones de calidad y accesibilidad del barrio.

4.1.3. Oferta

En economía, se define la oferta como aquella propiedad dispuesta a ser intercambiada libremente a cambio de un precio Mankiw et al. (2012).

4.1.4. Demanda

La demanda puede ser definida como la cantidad de bienes y servicios que son adquiridos por consumidores a diferentes precios en una unidad de tiempo específica Mankiw et al. (2012).

4.1.5. Mercado inmobiliario

El mercado inmobiliario es el marco en el cual se desarrollan todas aquellas transacciones económicas, que tienen por objeto inmediato la propiedad o el disfrute de un bien inmueble, es decir, tienen como finalidad el derecho a gozar o disponer de un bien que tiene una situación fija en el espacio y no puede desplazarse, como son los terrenos, locales comerciales, viviendas, fincas, etc Molina García (2014).

4.1.6. Web Scraping

Web Scraping es una técnica de minería de datos utilizada para extraer datos de sitios web, preferiblemente usando un programa que simula la exploración humana mediante el envío de peticiones por protocolo de transferencia de hipertexto (HTTP) simples o emulando un navegador web completo. Esta técnica se enfoca principalmente en la transformación de datos no estructurados en la web, en datos estructurados que pueden ser almacenados y analizados Chris (2013).El objetivo es automatizar el proceso de recolección de información de interés que no puede ser descargada manualmente, por ejemplo, si se quiere obtener las características de las ofertas de bienes raíces de una determinada página web, la comparación de precios en tiendas de automóviles o cualquier clasificado de interés.

Existen limitaciones legales relacionadas con el uso del Web Scraping, ya que, algunos países reconocen los derechos de bases de datos y limitan la reutilización de la información que se obtiene de sitios web publicados. De hecho, algunos sitios web se protegen declarando en sus condiciones legales la prohibición de realizar Web Scraping sobre la página.

4.2. Marco Teórico Estadístico

4.2.1. Estimación por Mínimos Cuadrados Ordinarios

Uno de los métodos mayormente utilizados para estimar los parámetros de un modelo de regresión lineal múltiple es el método de Mínimos Cuadrados Ordinarios. Según Montgomery et al. (2006) esto se puede realizar minimizando la suma del residuo cuadrático entre los valores observados y los valores estimados. Para explicarlo mejor supongamos que tenemos el siguiente sistema:

$$S(\beta) = \sum_{i=1}^{n} \epsilon_i = \epsilon^{'}\epsilon = (y - X\beta)^{'}(y - X\beta) \tag{4-1}$$

De esta Ecuación 4-1 se pueden desglosar los valores transpuestos, con el fin de obtener los valores cuadráticos, por lo tanto, $S(\beta)$ se puede expresar como:

$$S(\beta) = y^{'}y - \beta^{'}X^{'}y - y^{'}X\beta + \beta^{'}X^{'}X\beta = y^{'}y - 2\beta^{'}X^{'}y + \beta^{'}X^{'}X\beta \tag{4-2}$$

Seguidamente, se obtiene la derivada con respecto a β.

$$\frac{\delta S}{\delta \beta}|_{\hat{\beta}} = -2X^{'}y + 2X^{'}X\hat{\beta} = 0 \tag{4-3}$$

Por último, igualando a cero se obtiene como resultado las ecuaciones normales de mínimos cuadrados:

$$X^{'}X\hat{\beta} = X^{'}y \tag{4-4}$$

Para resolver la Ecuación 4-4 y obtener el estimador por Mínimos Cuadrados Ordinarios, debemos multiplicar a ambos lados de la ecuación por la inversa de $X^{'}X$. Esto se obtiene como resultado el estimador por (MCO):

$$\hat{\beta} = (X^{'}X)^{-1}X^{'}y \tag{4-5}$$

Por último, esta estimación se puede obtener solo si existe la matriz inversa de $X^{'}X$. Esto quiere decir que los regresores no son linealmente dependientes, en pocas palabras ninguna columna de la matriz X es una combinación lineal de las demás columnas.

4.2.2. Multicolinealidad

Según Montgomery et al. (2006) la multicolinealidad implica una dependencia casi lineal entre los regresores, los cuales son las columnas de la matriz X, por lo que es claro que una dependencia lineal exacta causaría una matriz $X^{'}X$ singular. La presencia de dependencia lineal puede influir en forma dramática sobre la capacidad de estimar coeficientes de regresión, lo cual implica que no existe una solución única para el sistema de ecuaciones de la matriz $X^{'}X$.

Supongamos que tenemos de la Ecuación $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$ con dos variables regresoras X. Por lo tanto, se tendrán 3 parámetros para este modelo y de lo cual podemos extraer una muestra de tamaño n, con variables estandarizadas que llamaremos W, por lo cual tendremos el modelo $Y = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + E$ y la siguiente matriz de correlaciones $W^{'}W$:

$$W^{'}W = \begin{pmatrix} 1 & \sum_{i=1}^{n} w_{1i}w_{2i} = r_{12} \\ \sum_{i=1}^{n} w_{2i}w_{1i} = r_{21} & 1 \end{pmatrix} \tag{4-6}$$

De la cual se puede definir r_{12} y r_{21} como la correlación que tiene W_1 y W_2. El siguiente paso es calcular las estimaciones por mínimos cuadrados de los parámetros del modelo que en la Ecuación 4-5 se pueden reemplazar:

$$\hat{\beta} = \frac{1}{1 - r_{12}^2} \begin{pmatrix} w_1^{'}y - r_{12}w_2^{'}y \\ -r_{12}w_1^{'}y + w_2^{'}y \end{pmatrix} \tag{4-7}$$

Por lo cual si definimos una alta correlación entre X_1 y X_2, r_{12} podrá aproximarse a uno, generando así un incremento en la estimación del parámetro, reflejada en la anterior ecuación como un resultado infinito, por lo cual no solo incrementara su estimación, también probablemente su varianza y covarianza, ocasionando un impacto en la representación de lo que es la muestra, ya que, al realizarse el ejercicio de sacar una nueva muestra, los valores en sus estimaciones serán muy diferentes.

4.2.3. Regresión por mínimos cuadrados parciales-(PLSR)

El PLSR o Regresión por Mínimos Cuadrados Parciales, nace de la lógica aplicada en el Análisis de Componentes Principales cuyo objetivo es establecer relaciones existentes entre individuos y variables. Es decir reducir la dimensionalidad de los datos Márquez Ruiz (2017), describiendo los valores p de variables en un subconjunto q de estructuras latentes incorrelacionadas. Al subconjunto q se le conoce como componentes principales y es evidente que $q \leq p$.

PLSR es una extensión del análisis de regresión múltiple y de componentes principales en el que se analizan los efectos de combinaciones lineales de varios predictores sobre una variable de respuesta Carrascal et al. (2009). Se establecen asociaciones con factores latentes extraídos de variables predictoras que maximizan la varianza explicada sobre la variable dependiente. Estos factores latentes se definen como combinaciones lineales entre variables predictores y de respuesta, cuyo fin es reducir la multidimensionalidad original a un menor número de factores ortogonales para detectar la estructura entre las relaciones de las variables predictoras, la variable de respuesta y los factores latentes.

El PLSR también permite trabajar con multicolinealidad, datos faltantes y un mayor número de individuos que de variables Carrascal et al. (2009), de no tener estos problemas en el conjunto de datos, entonces no se considera necesario usar este metodo.

PLS1

Existen dos tipos de PLS definidos como PLS1 y PLS2. El primero es el método PLS cuando se tiene una sola variable dependiente y el segundo es cuando se tienen dos o más variables dependientes, para esta problemática se abordó el método PLS1.

Supongamos que tenemos una matriz de variables predictoras X de orden $n \times p$ y un vector respuesta Y de orden $n \times 1$, PLSR transforma la matriz X teniendo en cuenta el vector Y, en una matriz de variables latentes $T = [t_1, t_2, ...t_p]$ llamadas componentes del PLS.

El objetivo de la regresión PLS es maximizar el cuadrado de la varianza entre la componente $t_i = xw$ y la variable de respuesta y, sujeta a la restricción $W^TW = 1$, donde $W = (w_1, w_2, ...w_p)^T$ es el vector tal que cada w_j es la covarianza entre la variable de respuesta con cada variable predictora Gaviria Peña (2016).

Construcción de la primera componente

La primera componente t_1 se define de la siguiente manera:

$$t_1 = w_{11}x_1 + w_{12}x_2 + ... + w_{1p}x_p \tag{4-8}$$

así pues $t = \sum_{j=1}^{p} w_{1j}x_j$, donde:

$$w_{1j} = \frac{cov(x_j, y)}{\sqrt{\sum_{j=1}^{p} cov^2(x_j, y)}} = \frac{< x_j, y >}{\sqrt{\sum_{j=1}^{p}(< x_j, y >)^2}} \quad (j = 1, 2, ..., p) \tag{4-9}$$

Denotando $< x_j, y >= cov(x_j, y)$. Por lo cual ya podemos obtener la ecuación lineal de predicción estimada:

$$y^* = \hat{\beta_1^*} t_1 \tag{4-10}$$

Donde $\hat{\beta}_1^*$ se calcula a partir de $\frac{cov(y,t_1)}{\|t_1\|^2}$ que es igual a $\frac{\sqrt{n-1}}{\|t_1\|} r_{y,t_1}$. Ahora se pueden calcular los residuos asociados a la recta de regresión:

$$e_1 = y - y^* \tag{4-11}$$

Construcción de la segunda componente

Se construye una segunda componente t_2 que sea combinación lineal de x_j, no correlacionada con la componente t_1 y explicando bien el residuo. Esta componente t_2 es combinación lineal de los residuos e_{1j} de las regresiones de las variables x_j sobre la componente t_1.

Se obtiene t_2 mediante la siguiente ecuación:

$$t_2 = w_{2,1}e_{1,1} + w_{2,2}e_{1,2} + ... + w_{2,p}e_{1,p} \tag{4-12}$$

Donde:

$$w_{2j} = \frac{cov(e_{1,j}, e_1)}{\sqrt{\sum_{j=1}^{p} cov^2(e_{1,j}, e_1)}} = \frac{< e_{1,j}, e_1 >}{\sqrt{\sum_{j=1}^{p}(< e_{1,j}, e_1 >)^2}} \quad (j = 1, 2, ..., p) \tag{4-13}$$

para el cálculo de los residuales $e_{1,j}$ para $j = 1, 2, ..., p$ se realizan las regresiones entre x_j sobre t_1 y se obtienen las predicciones estimadas:

$$x_j^* = \hat{\alpha_j^*} t_1, \quad (j = 1, 2, ..., p) \tag{4-14}$$

Donde la estimación de los coeficientes de regresión se calculan a partir de $\alpha_j^* = \frac{cov(x_j,t_1)}{\|t_1\|^2} = \frac{\sqrt{n-1}}{\|t_1\|} r_{x_j,t_1}$. Los residuales asociados a la recta de regresión están dados por:

$$e_{1,j} = x_j - x_j^* \tag{4-15}$$

Ahora bien, la estimación de y sobre la segunda componente se calcula utilizando la misma Ecuación 4-16 para la primera componente, donde $\hat{\beta}_2^*$ se calcula a partir de $\frac{\sqrt{n-1}}{\|t_2\|} r_{y,t_2}$ y sus residuos asociados a la regresión se dan mediante la Ecuación 4-11.

Regresión lineal múltiple sobre dos componentes

La ecuación lineal múltiple de y respecto a dos componentes es:

$$y^* = \hat{\beta_1^*} t_1 + \hat{\beta_2^*} t_2 \tag{4-16}$$

De lo cual, las estimaciones de los coeficientes de regresión se expresan de la siguiente forma:

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sqrt{n-1}}{\|t_1\|} \left[\frac{r_{y,t_1} - r_{y,t_2} r_{t_1,t_2}}{1 - r_{t_1,t_2}^2} \right] \\ \hat{\beta}_2 &= \frac{\sqrt{n-1}}{\|t_2\|} \left[\frac{r_{y,t_2} - r_{y,t_1} r_{t_1,t_2}}{1 - r_{t_1,t_2}^2} \right] \end{aligned} \tag{4-17}$$

dado que t_1, t_2 son ortogonales entonces $r_{t_1,t_2} = 0$, por lo tanto los estimadores se reducen en:

$$\hat{\beta}_1 = \frac{\sqrt{n-1}}{\|t_1\|} r_{y,t_1}; \quad \hat{\beta}_2 = \frac{\sqrt{n-1}}{\|t_2\|} r_{y,t_2} \tag{4-18}$$

Por último el residuo asociado a este modelo de regresión múltiple con dos componentes es:

$$e_2 = y - y^* = e_1 - y^* \tag{4-19}$$

En dado caso que la regresión lineal con las dos componentes no tenga un poder explicativo fuerte, se puede plantear un modelo con una tercera componente o con la cantidad de componentes óptima. Sin embargo en la mayoría de modelos esto no sucede, la obtención de la cantidad de componentes para el modelo PLSR se realiza a través de diferentes técnicas que se especifican en la metodología de este documento.

4.2.4. Validación Cruzada

Según Amat(2016) los métodos de validación cruzada son estrategias que permiten estimar la capacidad predictiva de los modelos frente a nuevas observaciones, es decir, se ajusta el modelo sobre un subconjunto de datos llamado conjunto de entrenamiento. Sobre el conjunto de datos de entrenamiento se evalúa la capacidad predictiva del modelo mediante métricas

de ajuste como lo es el coeficiente de determinación (R^2), la Raíz del Error Cuadrático Medio (RECM) o Porcentaje de Error medio Absoluto (PEMA), este proceso se repite múltiples veces hasta lograr obtener una distribución empírica de la métrica escogida. La diferencia entre los métodos de validación cruzada radica en el proceso de selección de los datos de entrenamiento y validación. A continuación se muestran las características principales del método empleado en esta investigación:

Método Leave One Out

El método de validación cruzada Leave One Out es un proceso iterativo que consiste en seleccionar como conjunto de entrenamiento todas las observaciones de la base de datos exceptuando una, por lo cual esta será utilizada como validación. el proceso se repite hasta lograr recorrer todas las observaciones de la base de datos, es decir para cada interacción se obtiene una observación diferente utilizada para la validación, por lo tanto, la métrica escogida para definir el desempeño de predicción del modelo se calculará a partir de todos los k errores calculados.

5 Metodologia

5.1. Introducción

En este capítulo se describe la metodología empleada para la presente investigación. Inicialmente se denotan las características de la población de estudio, seguido del procedimiento para la creación de la hoja de datos y la técnica utilizada para la extracción de la información. Posterior a esto, se presenta el proceso de depuración y validación de datos, así como el cálculo de las covariables de entorno. Finalmente, se detallan los procedimientos utilizados para el análisis exploratorio de los datos y la modelación.

5.2. Población de estudio

Esta investigación se plantea sobre la oferta de vivienda publicada en la página web OLX, específicamente para la ciudad de Santiago de Cali, capital del departamento del Valle del Cauca, Colombia. Santiago de Cali se encuentra territorialmente dividida por 15 corregimientos en su zona rural, 22 comunas y un total de 249 barrios en su perímetro urbano.(DAP 2019).

5.3. Creación de la hoja de datos

Uno de los principales retos de la investigación fue la obtención de una muestra de la oferta de vivienda en la ciudad de Cali, dado que en la actualidad no existen hojas de datos consolidadas con esta información. Por esta razón, se recurre a la técnica de Web Scraping, ya que realizar este ejercicio manualmente sería muy dispendioso dada la gran cantidad de registros.

5.4. Recolección de los precios de la vivienda y sus características

La obtención de los datos se delimita a la oferta de vivienda de la página web OLX para la ciudad de Cali en el año 2020, ya que se cuenta con una hoja de datos consolidada con información obtenida de esta página para el año 2019. Además de esto, dicha página permite

extraer las características propias, el precio de venta y la ubicación exacta de las viviendas, lo cual es fundamental para obtener las covariables de entorno. A continuación se muestra una captura de pantalla de como se muestra esta información en la página web.

Figura 5-1: Captura de la Página OLX/Properati (2022)

En la Figura **5-1**, se identifican muchas de las características que se extraerán para efectos de la investigación, como lo es el precio, la cantidad de baños, habitaciones, etc.

El uso del Web Scraping se hizo a partir de una extensión del navegador Google Chrome llamada "webscraper". El proceso consiste en ingresar a cualquier anuncio de oferta de vivienda en Cali, especificar qué tipo de nodo se requiere extraer de la página y dar çlic"sobre los elementos de interés para la extracción. Finalmente, se seleccionan todos los anuncios oferta de vivienda en Cali, generando como resultado una hoja de datos en Excel. Durante la implementación de este proceso se encontró que los anuncios de oferta de vivienda publicados en la página web OLX estaban siendo migrados a una nueva página creada por OLX dedicada exclusivamente a la compra y venta de bienes raíces llamada "Properati", razón por la cual se tuvo que exportar el archivo .json que contenía la URL desde cual se realizaba la extracción de los datos y añadir todas las publicaciones de la página Properati. A continuación, se muestra un resumen de las variables extraídas:

Variable	Tipo de variable	Comentario
Precio de la vivienda	Cuantitativa Continua	Extraída en millones de pesos*
Área Construida	Cuantitativa Continua	Extraída en m^2
Tipo de vivienda	Cualitativa Dicotómica Nominal	Define el tipo de vivienda se está midiendo: Casa o Apartamento
Estrato	Cualitativa Ordinal	Define el estrato socioeconómico 1,2,3,4,5,6 de la vivienda
Baños	Cuantitativa Discreta	La cantidad de baños que tiene la vivienda ofertada
Habitaciones	Cuantitativa Discreta	La cantidad de habitaciones que tiene la vivienda ofertada
Ubicación	URL de georeferencia	Obtención del URL utilizado en la API de Google Maps*
Descripción	Texto	Información extra donde se puede obtener las demás variables*

Tabla 5-1: Variables extraídas por Web Scraping

En la Tabla **5-1** se resumen las variables incluidas en la primera versión de la hoja de datos. Para el año 2019 la extracción de los datos se realizó en el mes de septiembre, obteniendo registros de mayo a septiembre de 2019 y en el año 2020 la extracción fue realizada en el mes de octubre, obteniendo registros desde mayo a octubre de 2020. A continuación, se muestra la cantidad de registros obtenidos para cada periodo.

Periodo	**Registros**
2019	8328
2020	1504

Tabla 5-2: Cantidad de registros obtenidos mediante Web Scraping.

5.4.1. Depuración de la hoja de datos

Organización de los datos

Luego de extraer los datos de la oferta de vivienda, el primer paso para la depuración de la hoja de datos es organizar los registros en sus respectivas variables, esto es necesario ya que los datos provienen de una página web en la cual no todos los vendedores diligencian correctamente los campos de los anuncios. Utilizando la herramienta Excel, se aplicaron los siguientes procesos:

- Definir la estructura de las variables.
- Eliminar los valores duplicados, publicaciones con el mismo título, descripción y ubicación.
- Reorganizar los registros mal diligenciados en sus respectivas variables.
- Incorporar la información obtenida en las páginas web OLX y Properati.

Validación de los datos

Una vez tenemos los datos organizados, se implementaron los siguientes criterios de inclusión y exclusión fundamentales para el estudio:

1. Los registros deben presentar valores no nulos en su ubicación (latitud y longitud), ya que es primordial obtener esta información para la creación de las covariables de entorno. Para la hoja de datos del 2020 se eliminaron 4 viviendas sin ubicación, obteniendo 1500 registros, mientras que en la hoja de datos del 2019 no se encontraron registros nulos.

2. Las viviendas deben ubicarse en la ciudad Santiago de Cali, por lo cual se grafican los registros de ambos periodos en el mapa de Cali utilizando el software R Core Team (2020) logrando así, para el año 2020 eliminar 377 viviendas y para el 2019 un total de 52 viviendas que se encontraban por fuera de los límites geográficos de la ciudad.

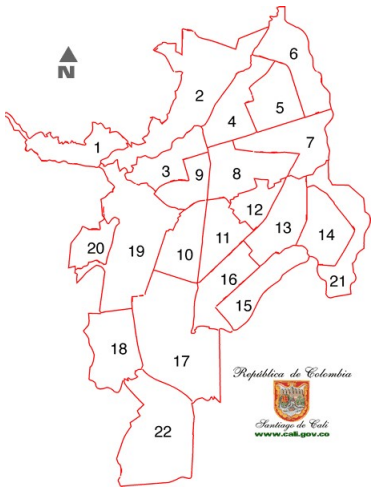

Figura 5-2: Límites geográficos establecidos por la Alcaldía (2022)

3. Viviendas que tengan un área total inferior a $1000m^2$, esto con el fin de excluir condominios o fincas que pueden inflar el precio de la vivienda. Para el año 2019 se se encontraron 15 viviendas que superan los $1{,}000m^2$, por lo cual se tienen 8.313 registros. Para el año 2020

no se encontraron valores superiores a esta área, teniendo un total de 1500 registros.

4. Viviendas que superen los 2,000 millones de pesos, con el fin de excluir condominios o fincas. Para este ejercicio solo se encontró un registro que superó este valor en la hoja de datos del año 2020.

Variable	Limite inferior	Limite superior
Precio por millón	$78 Mill	<$2.000 Mill
Área por m2	65 m2	<1.500 m2

Tabla 5-3: Límites del precio y área de la vivienda.

5. Eliminar registros que contengan más de dos variables con datos faltantes. Para el periodo del 2020 se encontraron 56 registros que contenían valores nulos en el tipo de vivienda, área por m^2 y estrato socioeconómico. Para el periodo 2019 no se encontraron registros con esta problemática.
6. Imputación de datos faltantes:

Variable	**Método**
Estrato socioeconomico	Se cuenta con 313 datos faltantes. El proceso de imputación se realizó concatenando la variable barrio obtenida mediante la API de Google Maps y posteriormente cruzando esta variable con la información de estrato socioeconómico por barrio en el documento publicado por el Departamento administrativo de planeacion municipal, de los cuales solamente se pudieron imputar 293.
Baños	Se cuenta con 50 datos faltantes. Se realizó imputación de la media por m2 a cada uno de estos datos

Tabla 5-4: Técnicas de imputación en variables con datos faltantes.

Terminado el proceso de depuración, se obtuvo la hoja de datos para el año 2019 con un total de 8261 registros y la hoja de datos del año 2020 con 948 registros. A continuación, se muestra gráficamente la distribución de los registros obtenidos para cada periodo en la ciudad de Cali.
En la Figura **5-3** se puede evidenciar que la cantidad de anuncios extraídos para el año 2020 fue muy poca comparada con la cantidad de observaciones obtenidas en el 2019, lo cual puede deberse a que el 36,26 % de la hoja de datos del año 2020 contenía valores faltantes en el estrato socioeconómico y que además fue imposible completarlos con la técnica empleada en la imputación de datos faltantes.

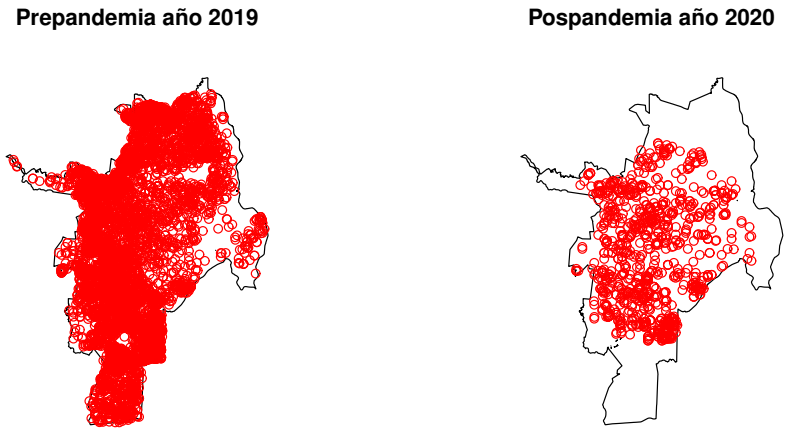

Figura 5-3: Registros por periodo

Creación de una nueva variable

Por último, se creó la variable precio por metro cuadrado, la cual genera realizando un cálculo simple: dividiendo el precio de la vivienda entre el área total en metros cuadrados. Esto se hizo con el fin de observar el comportamiento del precio por metro cuadrado en cada registro para los periodos de estudio.

5.4.2. Cálculo de las covariables de entorno

Por medio de la localización de los registros (latitud y longitud), se realiza la construcción de covariables de entorno que podrían ayudar a describir el efecto de la pandemia COVID-19 sobre el precio de la vivienda en la ciudad de Cali. Para este estudio se eligieron las siguientes covariables:

- Estrato socioeconómico
- Acceso a estaciones del MIO
- Cercanía a centros comerciales
- Cercanía a zonas verdes

Los datos utilizados para la construcción de las covariables fueron obtenidos del área de planeación de Cali (Alcaldía de Cali, 2018).

Estrato socioeconómico

La estratificación socioeconómica es un valor altamente influyente en el precio de la vivienda, además de permitir una homogeneidad en las características físicas y productivas de las viviendas. Esta covariable se construye mediante información extraída del Departamento administrativo de planeación municipal (Alcaldía de Santiago de Cali, 2018) en el documento que incluye información sobre los barrios, veredas y áreas de la ciudad de Cali. Este proceso se lleva a cabo concatenando las hojas de datos por medio de la variable barrio y extrayendo el estrato socioeconómico correspondiente.

Acceso a estaciones del MIO

Dado que el MIO (Masivo Integrado de Occidente) es el medio de transporte urbano más utilizado por los habitantes de la ciudad de Cali, resulta muy interesante conocer la influencia de esta variable sobre la oferta de vivienda. Esta variable se construye mediante la ubicación de los registros de las viviendas, calculando la distancia euclidiana en kilómetros entre las coordenadas de cada registro y la estación de MIO más cercana.

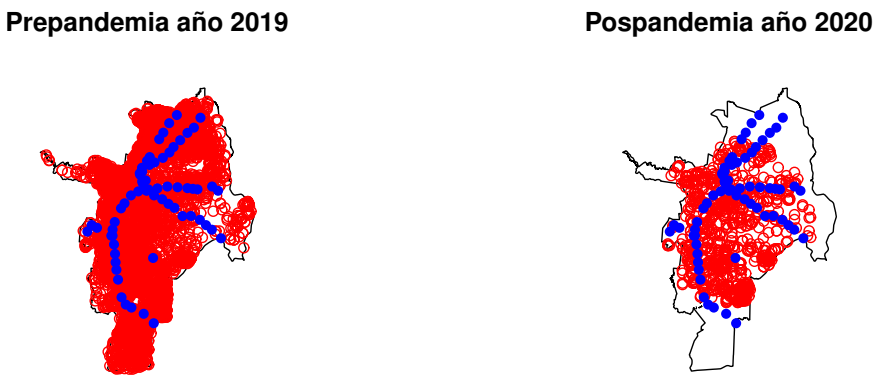

Figura 5-4: Estaciones del mio

La Figura **5-4** muestra la ubicación geográfica de las viviendas (color rojo) y de las estaciones de MIO (color azul).

Cercanía a centros comerciales

Teniendo en cuenta que los centros comerciales son de los lugares más concurridos en la ciudad de Cali tanto por recreación como por abastecimiento, se toma como variable de

entorno. Para el cálculo de esta variable, se halla la distancia euclidiana en kilómetros entre las coordenadas de cada registro de vivienda y el centro comercial más cercano.

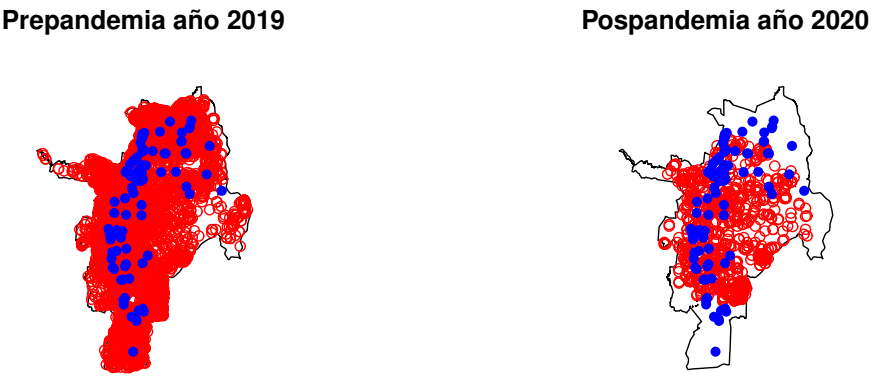

Figura 5-5: Centros Comerciales

La Figura **5-5** muestra la ubicación geográfica de las viviendas (color rojo) y de los centros comerciales de la ciudad de Cali (color azul).

Cercanía a zonas verdes

En esta oportunidad resulta interesante conocer cómo la cercanía a las zonas verdes afecta el precio de la vivienda. Dada la dificultad para hallar la ubicación exacta de todas las zonas verdes de la ciudad de Cali, se toma como zonas verdes los parques de la ciudad. Para la construcción de esta covariable se calcula la distancia euclidiana en kilómetros entre las coordenadas de cada registro de vivienda y el parque más cercano.

La Figura **5-6** muestra la ubicación geográfica de las viviendas (color rojo) y de los parques de la ciudad de Cali (color verde).

5.5. Variables objeto de estudio

Finalmente, después de depurar, validar y agregar las covariables de entorno a la hoja de datos, se tienen las siguientes variables como objeto de estudio para la investigación:

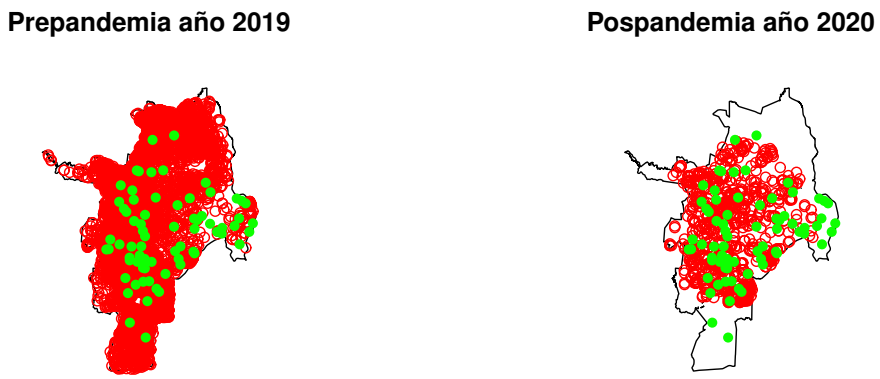

Figura 5-6: Zonas verdes

Tipo	Variable
Características propias de la vivienda	Precio de la vivienda
	Área por metro cuadrado
	Precio del metro cuadrado
	Tipo de vivienda
	Estrato socioeconómico
	Baños
	Habitaciones
Localización	Acceso a estaciones del MIO
	Cercanía a centros comerciales
	Cercanía a zonas verdes

Tabla 5-5: Variables objeto de estudio.

5.6. Ajuste del modelo PLSR

Con las hojas de datos correspondientes a los periodos prepandemia y pospandemia, se procede a estimar un modelo de regresión por mínimos cuadrados parciales (PLS), empleando el precio de la vivienda como variable de respuesta y las variables correspondientes a características propias de la vivienda y las covariables de entorno como variables predictoras, usando el algoritmo PLS1 propuesto en la Sección 4.2.5, el cual dará como resultado el siguiente modelo:

$$Preciovivienda = \beta_0^* + \beta_1^* Area + \beta_2^* Estrato + \beta_3^* Tipo + \beta_4^* Banos + \beta_5^* Habitaciones + \beta_6^* Precioarea + \beta_7^* Accesomio + \beta_8^* Acentrocomerciales + \beta_9^* Azonasverdes$$

Donde las estimaciones de los coeficientes β_j^* provienen de la cantidad de componentes que mejor expliquen el precio de la vivienda como variable dependiente. Así, entonces, se plantea un primer modelo con 8 componentes para ambos periodos utilizando validación cruzada en método LOO (leave-one-out) propuesta en la Sección 4.2.4. Una vez planteado el modelo, se busca reducir el RMSEP (Root Mean Squared of Prediction) o raíz cuadrada media de la predicción, ya que con este se evalúa el número ideal de componentes que explican mejor la regresión de la variable dependiente. Posterior a esto, se utilizó la función de R (coefficients) para obtener los coeficientes del modelo y extraer la máxima información comparativa para ambos modelos

6 Resultados

En este capítulo se presentan los resultados obtenidos de esta investigación. En la primera parte se realiza un análisis exploratorio de datos, con el fin de analizar el comportamiento de cada una de las variables y covariables objeto de estudio, además se evalúan las correlaciones existentes entre ellas y se comparan estos resultados para los periodos definidos como prepandemia (año 2019) y pospandemia (año 2020). Seguidamente, para la modelación estadística se utiliza un modelo de regresión por mínimos cuadrados parciales haciendo uso del precio de la vivienda como variable de respuesta y las variables correspondientes a características propias de la vivienda y covariables de entorno como variables predictoras.

6.1. Análisis exploratorio

Con el fin de analizar el comportamiento del precio de la vivienda y la influencia de cada una de las variables y covariables de estudio sobre este, se realiza un análisis exploratorio de datos, además se evalúa la correlación existente entre estas variables y se realiza el comparativo de los resultados para los periodos pre y pospandemia.

6.1.1. Análisis del precio de la vivienda

Se realizó un análisis de la distribución del precio de la vivienda con el fin de tener una vista general del comportamiento de esta variable para cada periodo de estudio.

En la Figura (**6-1**) se ilustra el histograma de frecuencia del precio de la vivienda en el cual se observa para ambos periodos una distribución asintótica positiva donde la mayoría de las observaciones se concentran dentro de un rango de precios entre los 59 y 800 millones de pesos colombianos. Por lo general, los precios de la vivienda se distribuyen de esta manera y presentan colas pesadas cuando el valor supera los 1000 millones de pesos, esto debido a la influencia de los metros cuadrados de la vivienda, su tipo y ubicación.

En la Figura (**6-2**) se puede observar cómo el precio de la vivienda está condicionado por el tipo de inmueble, además evidencia que los apartamentos son los que más generan valores atípicos, pero al mismo tiempo una menor variabilidad de precios. Por otra parte, se destaca

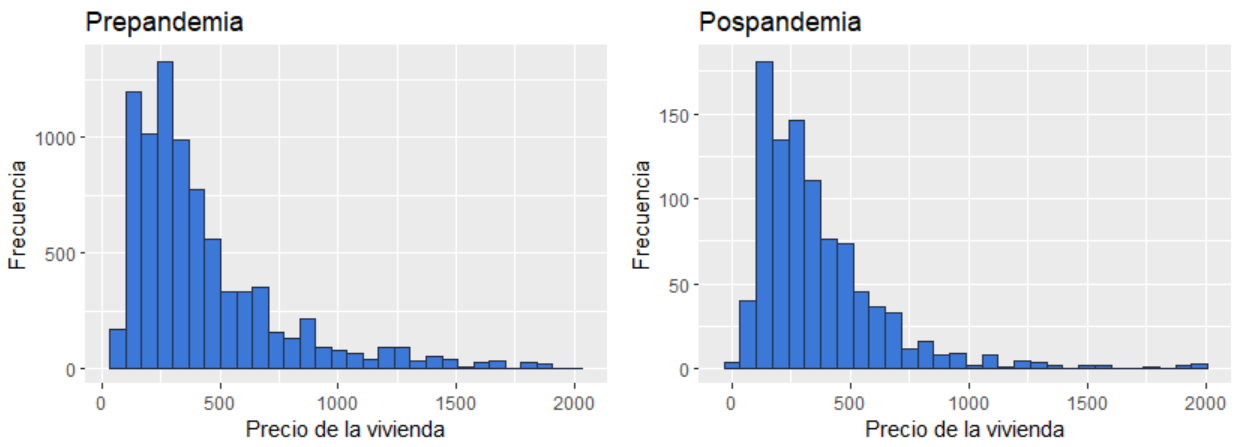

Figura 6-1: Distribución del precio de la vivienda para los periodos pre y pospandemia

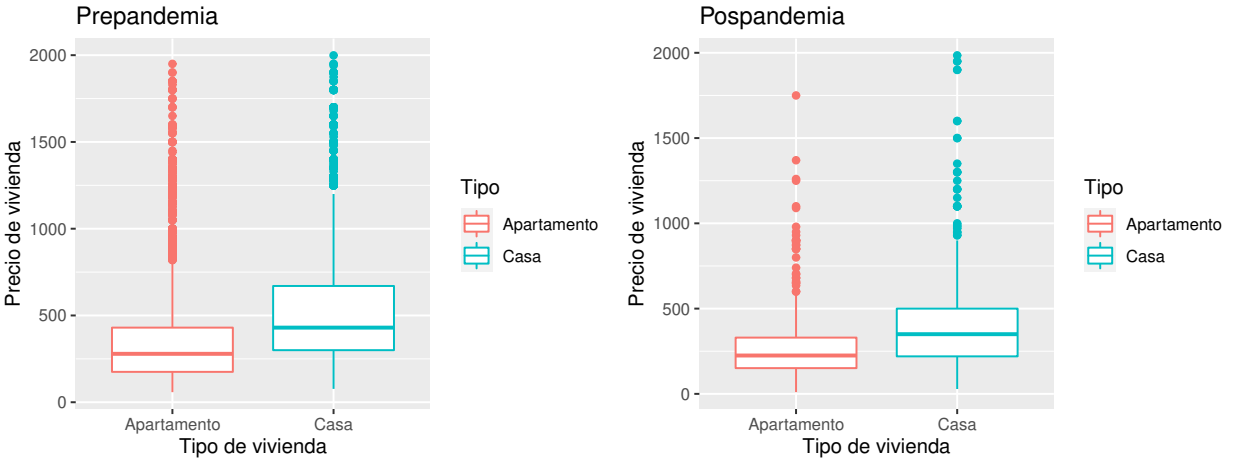

Figura 6-2: Precio de la vivienda por tipo de vivienda

que las casas son naturalmente más costosas, tienen una mayor variabilidad de precios y poseen una menor cantidad de datos atípicos.

Comparando los resultados obtenidos para los periodos pre y pospandemia, observamos un leve cambio en la distribución con respecto al tipo de inmueble, específicamente en los apartamentos. Las medianas de cada periodo son notablemente distintas, en el periodo prepandemia el valor es de 279 millones de pesos, en cambio, para el periodo pospandemia el valor fue de 195 millones de pesos, generando una diferencia de 84 millones de pesos. Por otra parte, es importante conocer la distribución del precio por tipo de vivienda en relación con su estrato socioeconómico.

Como se observa en la Figura(**6-3**), dado que el estrato socioeconómico es una variable de

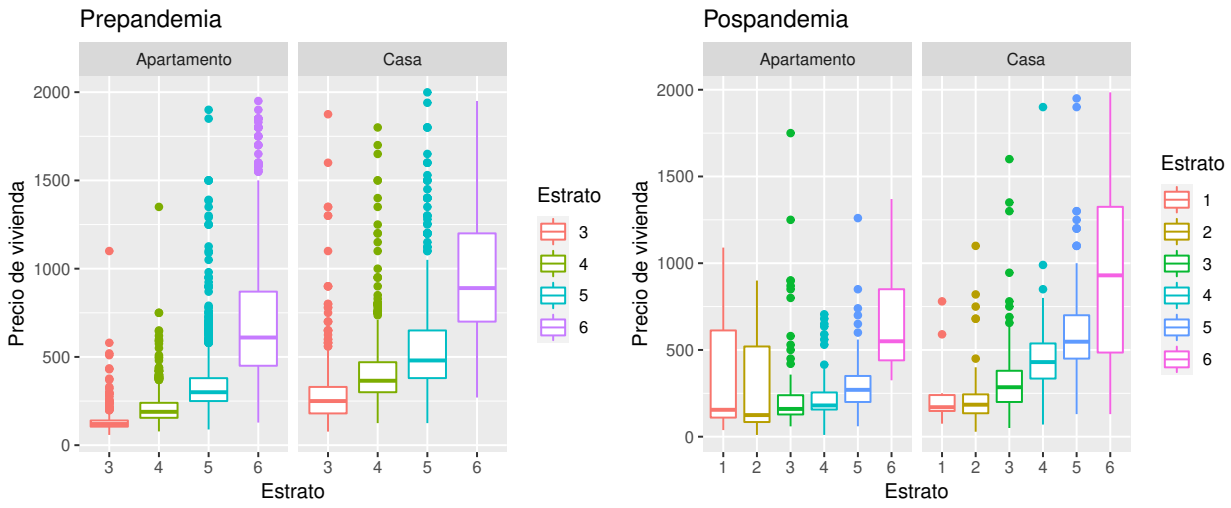

Figura 6-3: Precio de la vivienda por tipo de vivienda y estrato

tipo ordinal se evidencia una relación lineal positiva entre el precio de la vivienda y su estrato socioeconómico, de manera que si aumenta el estrato socioeconómico, el precio también lo hará. Sin embargo, la verdadera causal de la relación entre el precio de la vivienda y el estrato socioeconómico es el nivel de ingresos y la capacidad de endeudamiento de las personas que pueden vivir en un estrato socioeconómico alto.

Por otra parte, en análisis de la Figura(**6-3**) se muestra una diferencia significativa en la mediana del precio de las viviendas tipo apartamento de estrato 6, la cual en el periodo pospandemia se encuentra muy por debajo del valor obtenido en el periodo anterior. Así mismo, se tiene que para el periodo prepandemia los apartamentos de estrato 5 y 6 presentan una gran cantidad de datos atípicos, caso totalmente contrario al periodo pospandemia. Por último, vale la pena resaltar que para el periodo prepandemia no se encontraron registros de ofertas de vivienda con estrato socioeconómico inferior a 3.

Una de las variables con mayor influencia sobre el precio de la vivienda es el área en metros cuadrados. Usualmente estas variables presentan una relación lineal positiva, por lo cual antes de ver su interacción resulta interesante conocer cómo las dimensiones de la vivienda pueden verse influenciadas por su tipo y estrato socioeconómico.

En este orden de ideas, la Figura (**6-4**) ilustra la interacción entre el área en metros cuadrados, el tipo de vivienda y el estrato socioeconómico. Se puede inferir que para ambos periodos de estudio el área en metros cuadrados para los apartamentos es considerablemente menor comparado con las casas de su mismo estrato, además se tiene que para el periodo prepandemia el promedio de área en metros cuadrados es de 174 m2, el cual resulta ser inferior al periodo pospandemia que presenta un promedio de 193 m2; sin embargo, este promedio

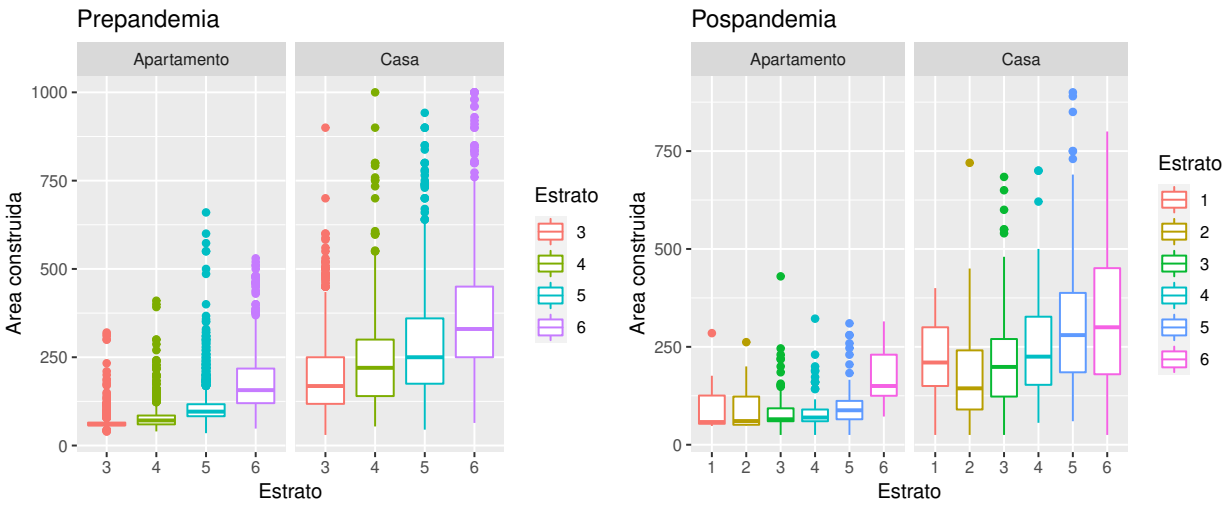

Figura 6-4: Área por metro cuadrado por tipo de vivienda y estrato

no permite sacar ninguna conclusión ya que el periodo pospandemia contiene viviendas tipo apartamento estrato 1 y 2 con áreas considerablemente grandes, por lo cual se debe evaluar la relación entre el área en metros cuadrados de la vivienda y su precio.

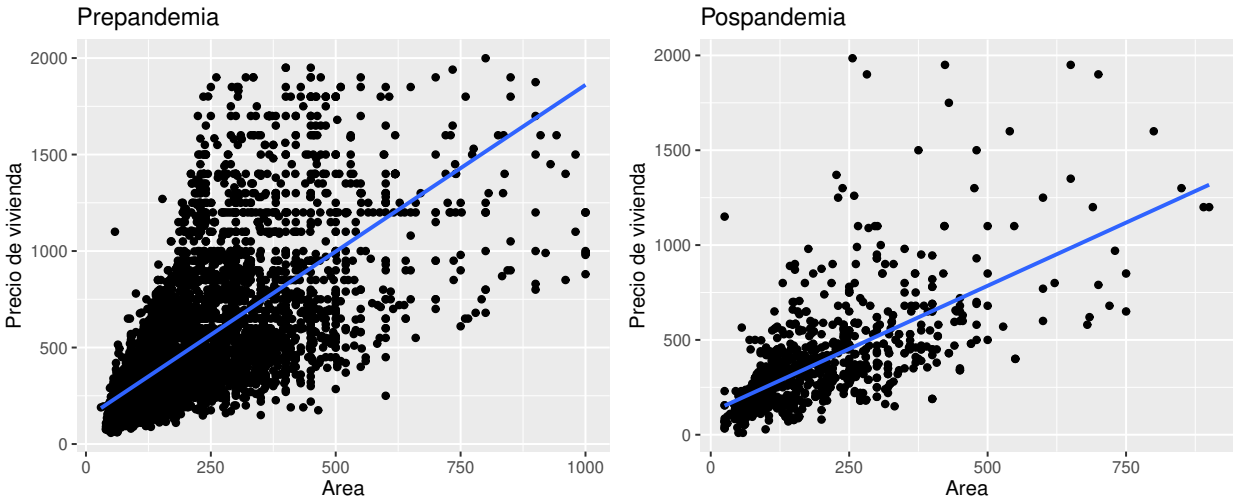

Figura 6-5: Precio de la vivienda por área en metros cuadrados

En la Figura (**6-5**) realizando la comparación de los valores presentados para los periodos pre y pospandemia se observa un cambio significativo en la tendencia del precio de la vivienda, evidenciando que para el periodo prepandemia la tendencia alcanza un valor máximo de 2000 millones de pesos para viviendas que superan los 1000 m2, además de que en este periodo la mayoría de los puntos se encuentra por debajo de los 500 m2 y los 1000 millones de pesos, lo cual confirma la existencia de una correlación positiva entre las variables. En

el periodo pospandemia se tiene la misma correlación entre las variables; sin embargo, se observa que la mayoría de las viviendas presentan precios por debajo de los 500 millones de pesos y áreas inferiores a los 750 m2. Adicionalmente, es importante destacar que aunque algunas observaciones intentan mantener el comportamiento del periodo prepandemia, se evidencia una afectación considerable en la distribución.

Puesto que la figura anterior solo permite comparar el comportamiento entre el precio de la vivienda y su área para los periodos pre y pospandemia, es necesario evaluar la interacción de estas variables con el tipo vivienda y el estrato.

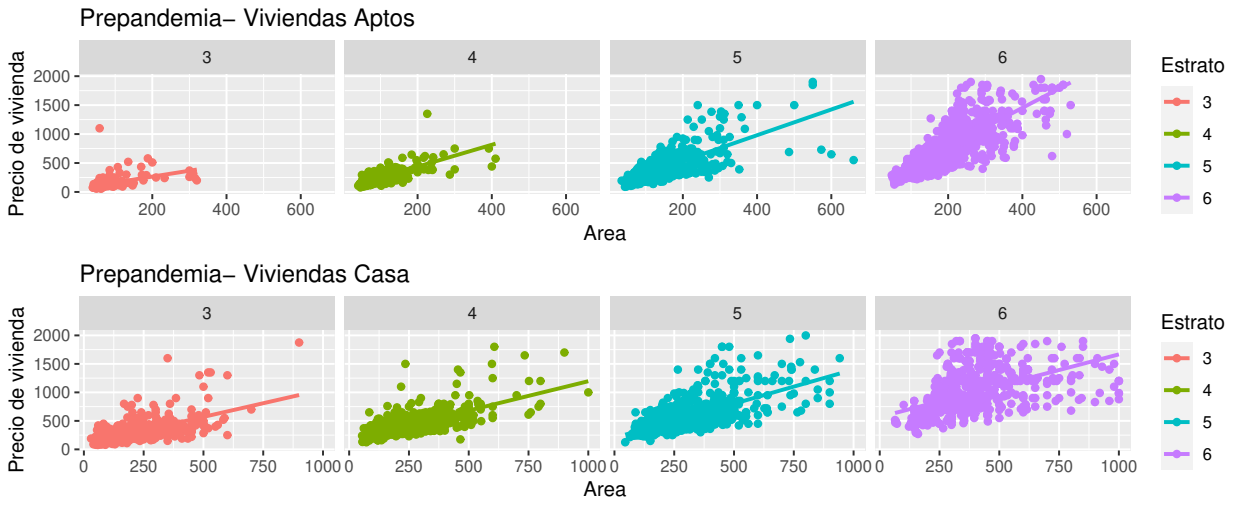

Figura 6-6: Precio de la vivienda por tipo de vivienda, área y estrato en el periodo prepandemia

En la Figura (**6-6**) se observa la influencia que tiene el estrato socioeconómico sobre el precio de la vivienda y su área en metros cuadrados para el periodo prepandemia. Además, se observa que las viviendas de tipo apartamento no superan los 600 m2 y conforme aumenta su estrato así mismo aumenta su área y su precio, el cual alcanza valores cercanos a los 2000 millones de pesos.

Para las viviendas tipo casa las variables presentan la misma relación, aunque se evidencia una gran diferencia en cuanto al área en metros cuadrados, la cual alcanza valores cercanos a los 1000 m2 desde el estrato 3, lo cual es más del doble del área máxima de una vivienda tipo apartamento de su mismo estrato. Esta particularidad puede verse explicada en que la mayoría de los apartamentos se encuentran ubicados en sectores con estrato superior a 3 y que además existen factores del entorno de la vivienda que también afectan su precio.

Por otra parte, en la Figura(**6-7**) se ve que para el periodo pospandemia, la relación entre el estrato socioeconómico, el precio de la vivienda y el área en metros cuadrados es igual

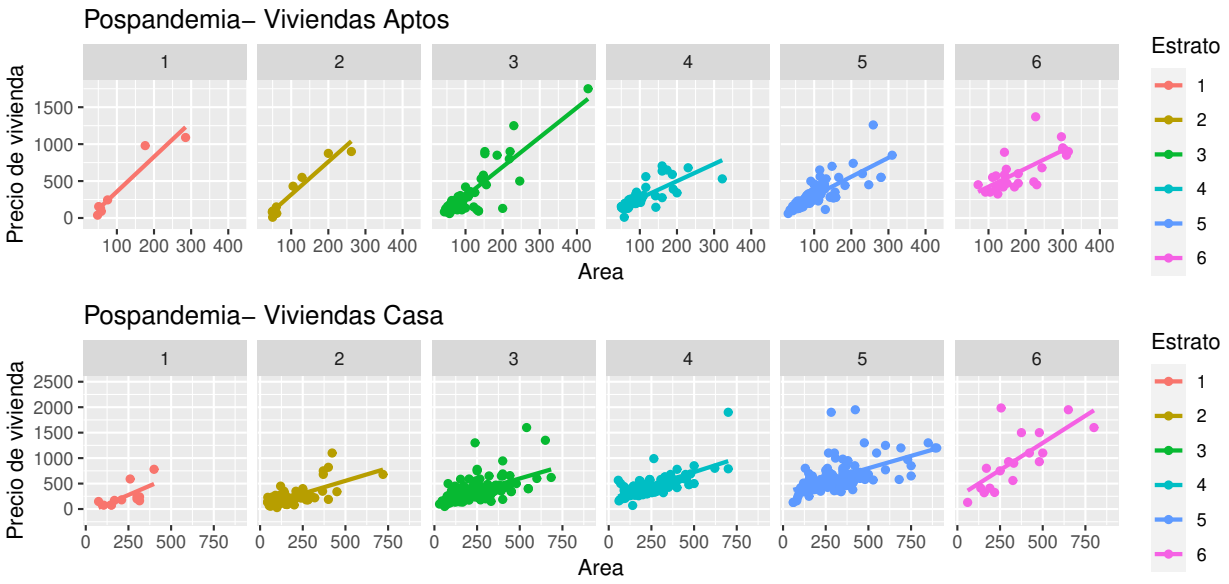

Figura 6-7: Precio de la vivienda por tipo de vivienda, área y estrato en el periodo pospandemia

al periodo anterior. Sin embargo, cabe destacar la disminución en las ofertas de vivienda tipo apartamento y la reducción del área en metros cuadrados. Siguiendo con el análisis, un hallazgo importante es que las ofertas de viviendas tipo casa presentan una reducción significativa del área en metros cuadrados, pues incluso en el estrato más alto el área máxima es de 750m2 y mantiene el precio en valores cercanos a los 2000 millones de pesos, valor con el cual en el periodo prepandemia se conseguían casas de hasta 1000m2.

Análisis del precio por metro cuadrado

Teniendo en cuenta la relación presentada entre el estrato socioeconómico, el área en metros cuadrados y el precio de la vivienda, resulta importante plantear una nueva variable que permita conocer el comportamiento del precio del metro cuadrado en ambos periodos de estudio.

En la Figura(**6-8**) se muestra que no existe una diferencia significativa en el precio del metro cuadrado para los periodos evaluados, también se pudo observar la fuerte influencia del estrato socioeconómico sobre el precio del metro cuadrado y la alta frecuencia de valores atípicos para las viviendas estrato 3 en el periodo prepandemia.

Análisis de las variables discretas

Como se observó anteriormente, el problema de la multicolinealidad está presente dentro del este estudio. Las variables afectadas son la cantidad de baños y habitaciones que tiene la vivienda; sin embargo, no se ha abordado su relación con el precio de la vivienda. En este

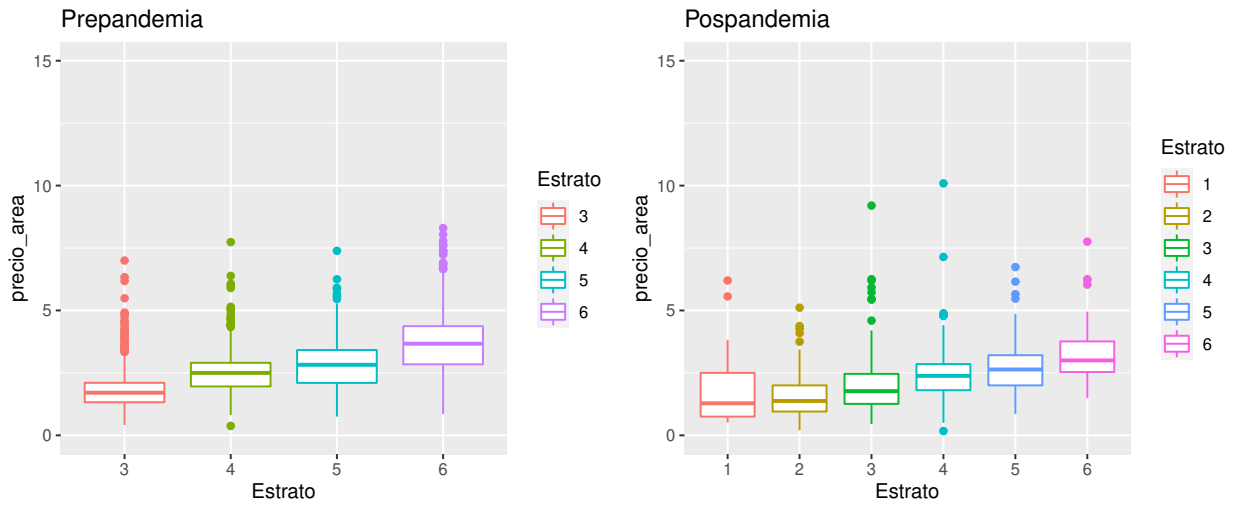

Figura 6-8: Precio del metro cuadrado por estrato

apartado se mostrará la información subyacente y la interacción de estas variables. La figura

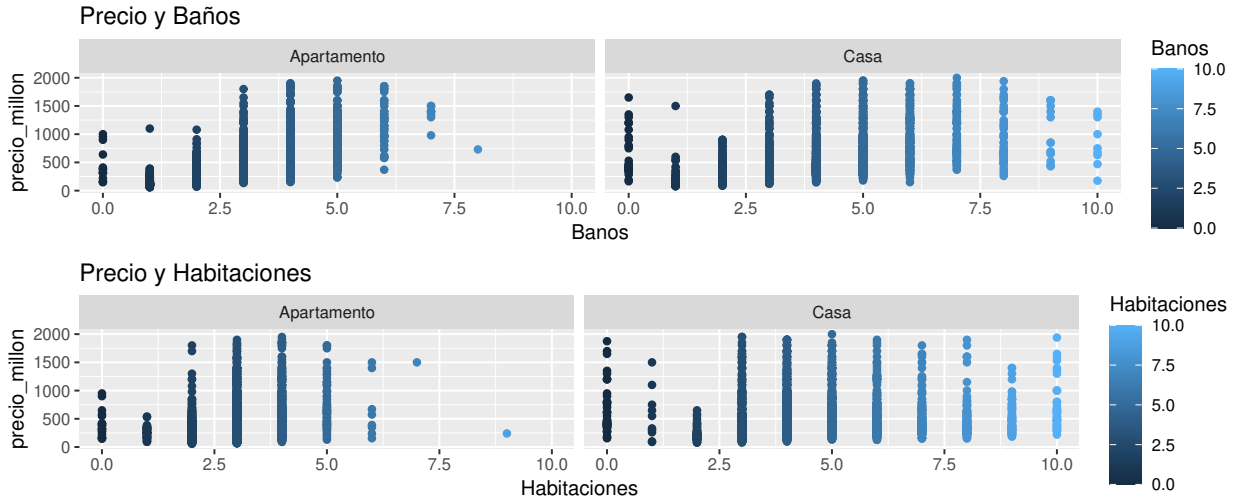

Figura 6-9: Precio por cantidad de baños y habitaciones en el periodo prepandemia

(**6-9**) muestra que para el periodo prepandemia el comportamiento de las variables cantidad de baños y número de habitaciones es natural, de manera que si sus valores aumentan, el precio de la vivienda también aumenta. Además, se evidencia que para las viviendas tipo apartamento dado que su área tiende a ser inferior como se vio anteriormente, está compuesta por una cantidad menor de baños y habitaciones respecto a una casa.

En la Figura (**6-10**) se muestra que aunque las variables conservan la relación presentada en el periodo anterior, los precios se vieron fuertemente afectados. Observamos que son muy pocas las ofertas de vivienda tipo apartamento que superan los 500 millones de pesos. Por

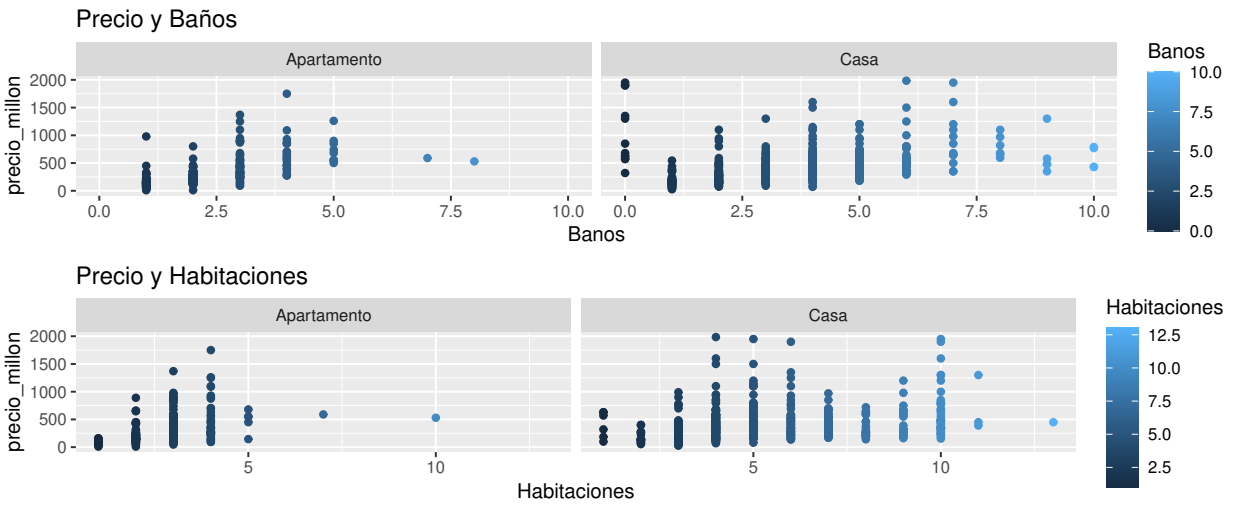

Figura 6-10: Precio por cantidad de baños y habitaciones en el periodo pospandemia

otro lado, las casas presentan un mismo comportamiento aunque más enfocado al número de habitaciones, de manera que si se aumenta la cantidad, el precio de la vivienda puede superar los 1000 millones de pesos.

Análisis de correlación

A continuación se introducen las gráficas que evidencian la existencia de una correlación lineal en las variables objeto de estudio. Además, para graficar correctamente se añadió la variable estrato como un entero y se excluyo la variable tipo de vivienda, ya que es cualitativa cuantificable.

Vemos que en las Figuras(**6-11**,**6-12**) el precio tiene una relación lineal buena con las variables propuestas en el estudio, aunque algunas no destacan por su mínima correlación, como lo son las variables de entorno, por otro lado, se sospecha una multicolinealidad entre el área por metro cuadrado y la cantidad de baños y habitaciones.

El concepto general de la multicolinealidad para este ejercicio puede ser generada por la relación causal entre variables explicativas del modelo, ya que, si enfocamos la atención sobre el área por metro cuadrado de la vivienda y la cantidad de habitaciones y baños que hay en esta, es posible evidenciar esta problemática. Realicemos un ejemplo sencillo, actualmente un apartamento nuevo en obra negra, tiene aproximadamente 75 m^2, distribuidos para tres habitaciones, dos baños, cocina y sala; sin embargo, si aumentamos el área por metro cuadrado, cabe la posibilidad de que se pueda agregar un baño o una habitación de más, siendo esta una relación causal, para explicarlo mejor, a continuación, se realizara un análisis multivariado utilizando ACP (Análisis de componentes principales) donde se ilustrara la gráfica

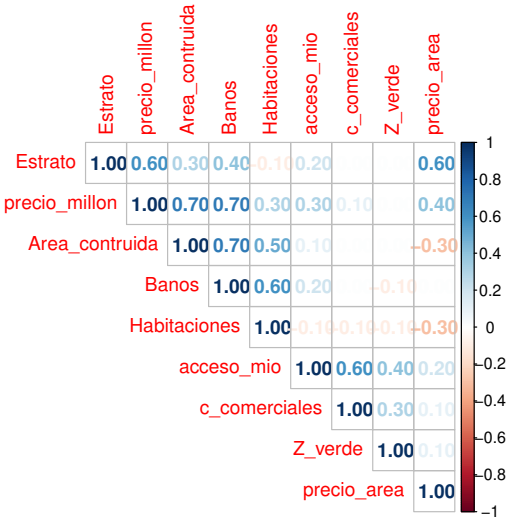

Figura 6-11: Diagrama de correlación de las variables objeto de estudio en periodo prepandemia

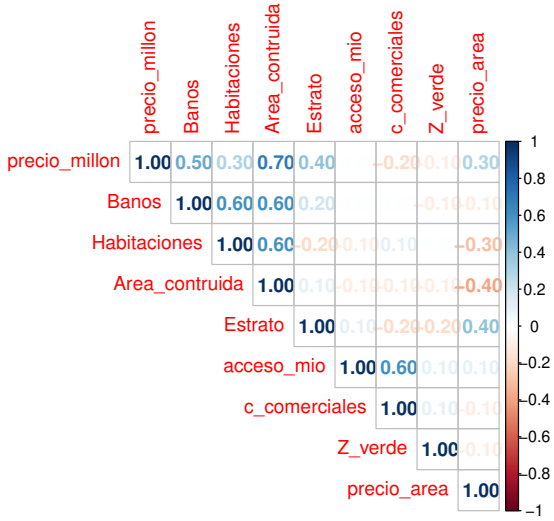

Figura 6-12: Diagrama de correlación de las variables objeto de estudio en periodo pospandemia

de variables y sus relaciones.

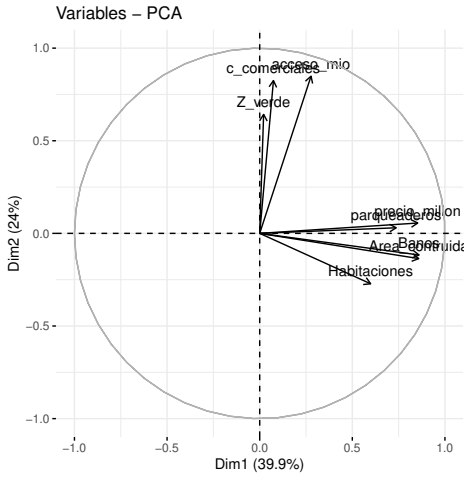

Figura 6-13: Correlación prepandemia

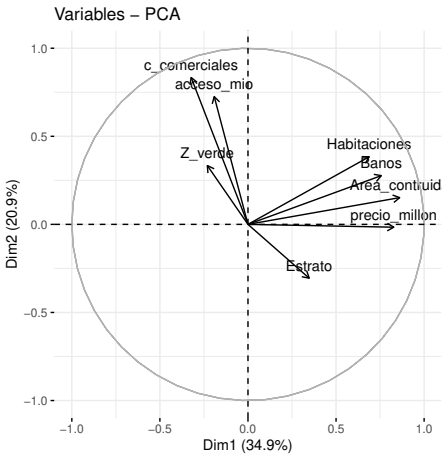

Figura 6-14: Correlación pospandemia

Las Figuras (**6-13**,**6-14**) representan las variables analizadas en el ACP de dos dimensiones con una varianza explicada aproximadamente del 59 %, este exclusivamente se utilizó para afirmar el problema de multicolinealidad, que se puede evidenciar para los dos casos de pre y pospandemia, donde la cantidad de baños y habitaciones están estrechamente relacionadas con el área por metro cuadrado, es decir la posibilidad planteada en el ejemplo anterior no está alejada de la realidad presentada por los datos.

6.2. Regresión por mínimos cuadrados parciales (PLS)

Inicialmente se estandarizan los datos con respecto a la media y varianza para cada variable. Seguidamente se plantearán dos modelos con 8 componentes para cada periodo, esto se realiza con el fin de obtener el número de componentes óptimo que reduzca el error cuadrático medio de predicción y a su vez contengan un buen porcentaje de varianza explicada sobre la variable precio de vivienda.

En la Figura (**6-15**) se puede evidenciar que para ambos modelos planteados para cada periodo, la cantidad de componentes que reducen el error cuadrático medio de predicción es dos. Como también en la Tabla(**6-1**) se resumieron los porcentajes de varianza que contienen máximo 4 componentes, dando como resultado que el número óptimo de componentes para ambos modelos es dos.

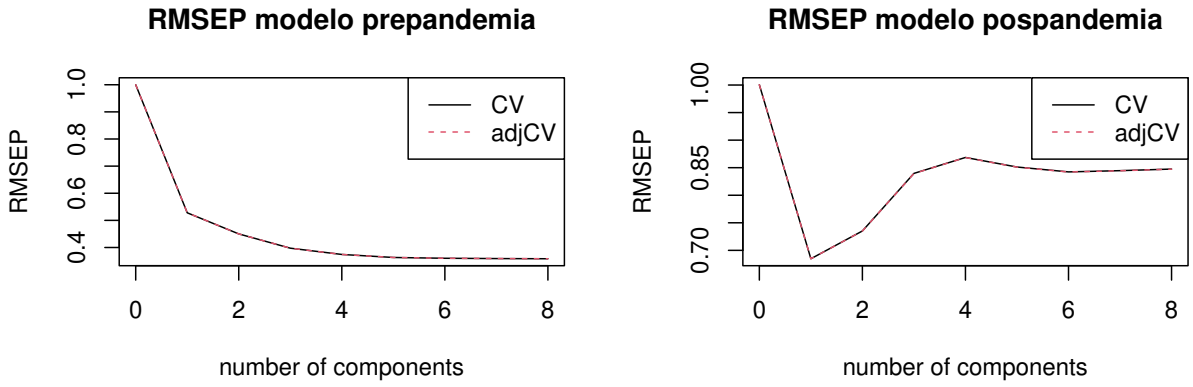

Figura 6-15: Raíz del error cuadrático medio para el modelo pls

La Tabla(**6-1**) muestra el porcentaje de varianza explicada por la cantidad de componentes seleccionada. Para cada modelo se obtuvo un porcentaje diferente, en la prepandemia el porcentaje de varianza explicada sobre la variable dependiente precio de la vivienda es del 80 %, mientras que para la pospandemia se obtuvo un 63 %.

	% de varianza explicada			
Periodos	comp 1	comp 2	comp 3	comp 4
prepandemia	72.18	79.80	84.28	86.07
pospandemia	55.57	63.00	67.24	68.60

Tabla 6-1: Porcentaje de varianza explicada por componente

Ahora bien escogidas las componentes para cada modelo, se procede a graficar las variables y vivienda por su tipo en el plano de ambas componentes.

6.2.1. Gráfico de variables y de viviendas

En las Figuras(**6-13**,**6-14**) se puede observar la gráfica de las variables para los diferentes periodos. Para el modelo ajustado de la prepandemia se puede evidenciar cuáles son las variables mejor explicadas por cada componente, las variables precio de la vivienda, Estrato, precio del metro cuadrado, cercanías a centros comerciales, acceso a estaciones de bus y cantidad de baños, están mayormente explicadas por la primera componente, mientras que en la segunda componente se explica el área en metros cuadrados, la cantidad de habitaciones, el tipo de vivienda y cercanías a zonas verdes.

En cuanto al modelo ajustado para la pospandemia, la primera componente explica el área por metro cuadrado, el precio de la vivienda, Estrato, el precio por metro cuadrado y la cantidad de habitaciones, por otro lado la segunda componente está más asociado a las variables Tipo de vivienda, cercanías a centros comerciales, acceso a estaciones de bus, cercanias a zonas verdes y cantidad de baños.

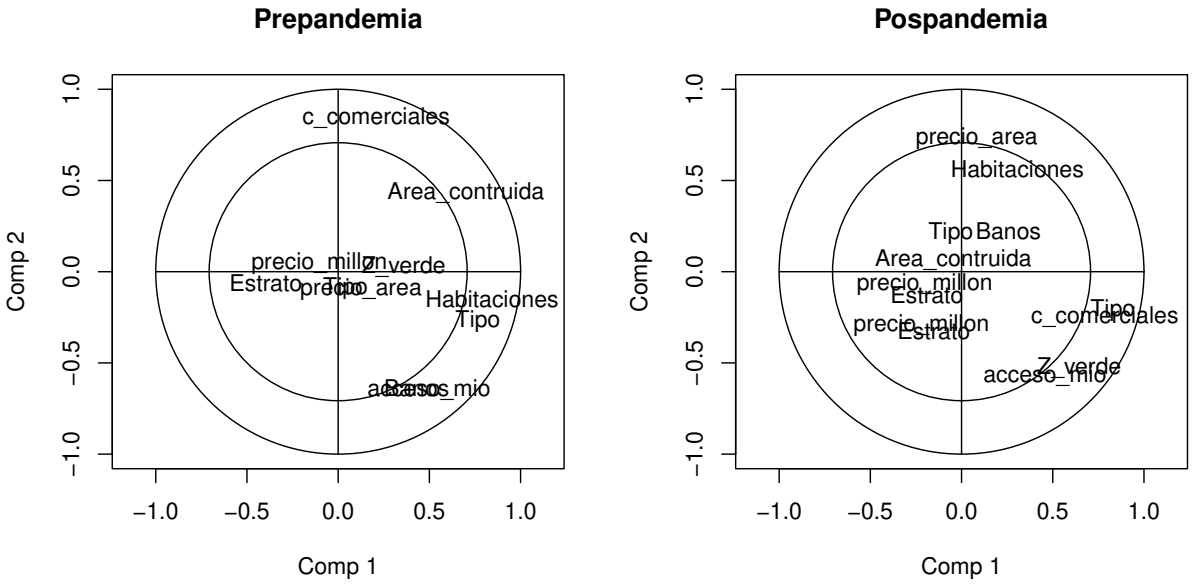

Figura 6-16: Gráfico de variables para los diferentes periodos

Seguidamente, se realiza el gráfico de individuos para ambos modelos, con el fin de obtener detalladamente el comportamiento de viviendas atípicas.

Como se puede evidenciar en la Figura (**6-17**) se destacan las viviendas que contienen valores atípicos para las variables explicadas en cada componente, es decir la vivienda de tipo 424 identificada como apartamento para la prepandemia se encuentra muy próxima a un centro comercial, por lo que a pesar de tener un área de $57m^2$ su precio es de 75 millones de pesos,

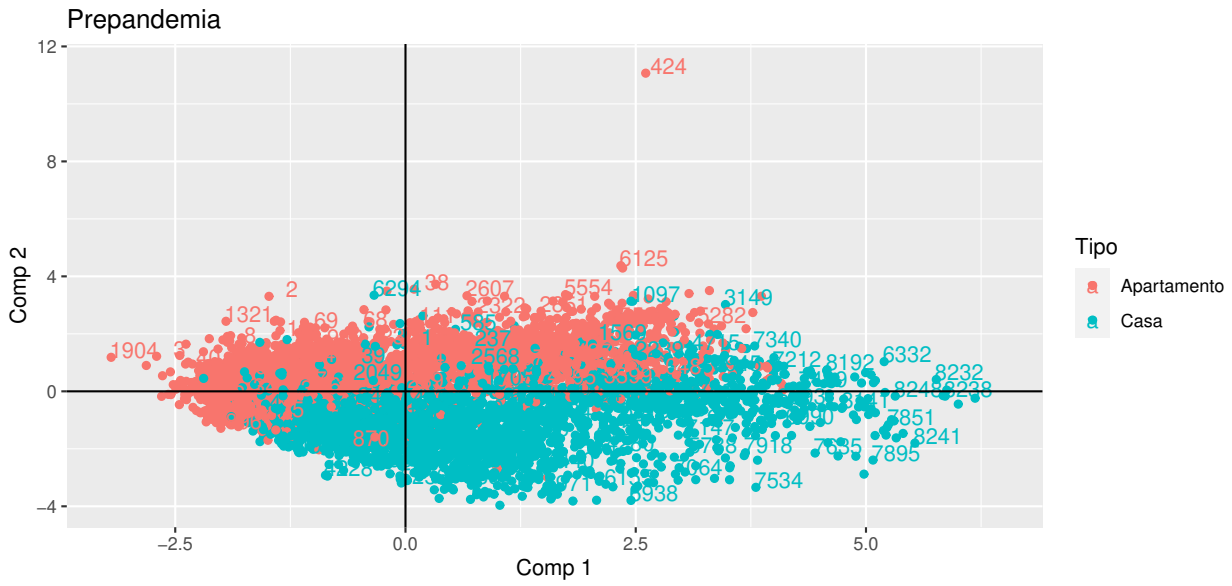

Figura 6-17: Gráfico de individuos sobre las componentes 1 y 2 del modelo PLS prepandemia

mucho menor a lo obtenido sobre el promedio de los precios de la vivienda.

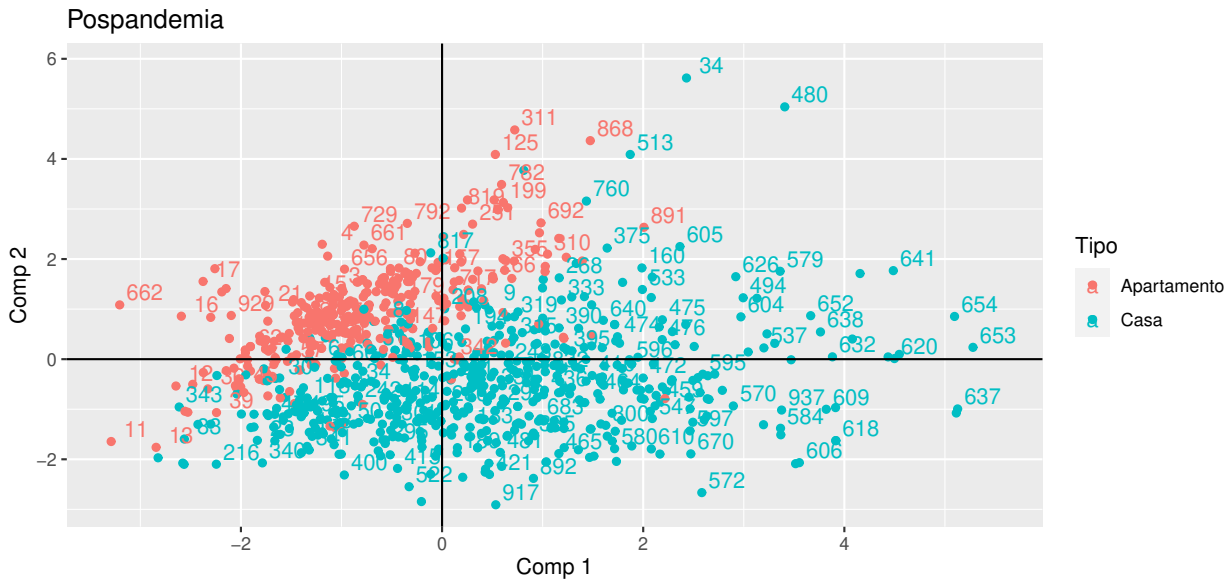

Figura 6-18: Gráfico de individuos sobre las componentes 1 y 2 del modelo PLS año 2020

Posteriormente, en la Figura (**6-18**) se obtienen las viviendas representadas por el modelo planteado para la pospandemia, de lo cual se pudo evidenciar particularidades en las viviendas de tipo Casa. La vivienda 34 se destacó por su lejanía respecto las covariables de

entorno, como también el alto coste de 524 millones de pesos por un área igual a $56m^2$. Entre otras observaciones como la vivienda 480 cuyo valor se encuentra por encima de los 1000 millones de pesos colombianos y se destaca por tener el mismo comportamiento en la lejanía respecto a las covariables de entorno. Teniendo en cuenta todas las características que el modelo pls nos puede ofrecer con respecto a las viviendas y su tipo, a continuación se mostrara el resultado de los coeficientes de cada modelo planteado en ambos periodos.

6.2.2. Coeficientes del modelo de regresión PLS

En la Tabla (**6-2**) se muestra la comparación de los coeficientes estandarizados para los modelos de la prepandemia y pospandemia que siguen el modelo estadístico planteado en la Sección 5.6, donde las variables propuestas para el estudio influencian significativamente el precio de la vivienda. La variable Área en metros cuadrados representa para ambos periodos una relación positiva, esto evidencia lo anteriormente encontrado en la Figura (**6-5**), la comparativa sobre la prepandemia y la pospandemia muestra que para el modelo de la pospandemia su coeficiente aumentó. Ahora bien, al obtener la estimación no estandarizada[1] del coeficiente de la variable Área en metros cuadrados para ambos periodos dio como resultado que en la prepandemia su valor fue de 0,882 y en la pospandemia 0,801 demostrando que no existe cambio significativo entre ambos periodos.

	Prepandemia (2019)	**Pospandemia (2020)**
	Coeficientes Estandarizados	
Area m2	0.36089	0.40344
Estrato 2	-	-0.11603
Estrato 3	-	-0.11515
Estrato4	-0.08614	-0.01562
Estrato5	0.00462	0.11231
Estrato6	0.30081	0.25270
TipoCasa	0.03079	0.03360
Banos	0.26191	0.17879
Habitaciones	-0.00871	0.08690
Precio del metro cuadrado	0.33555	0.27552
Acceso a estaciones de bus	0.08619	0.05221
Acceso a centros comerciales	-0.01203	-0.09892
Acceso a zonas verdes	-0.02050	-0.05139

Tabla 6-2: Comparativa de los coeficientes del modelo PLSR para cada periodo.

[1]Multiplicando el coeficiente β_j estandarizado por $\frac{S_y}{S_{x_j}}$ da como resultado la estimación no estandarizada del coeficiente j

El Estrato que define el entorno socioeconómico de la vivienda sigue la tendencia evidenciada en el análisis exploratorio cuya relación es negativa, entre más estrato se encuentre una vivienda mayor será su valor. Para la comparación de ambos periodos se muestra una disminución de la relación negativa con respecto al estrato 4, mientras que para los estratos 5 y 6 esta relación se mantuvo positiva. El tipo de vivienda casa comparado con el tipo apartamento mantiene la relación positiva con respecto al precio de la vivienda, esto quiere decir que una vivienda de tipo casa puede llegar a costar más que una vivienda de tipo apartamento. en su comparativa no se encuentran diferencias para la prepandemia y pospandemia.

Las variables baños y habitaciones muestran diferencias en sus coeficientes, los baños influencian de manera positiva al precio de la vivienda. Su comparativa para la prepandemia y la pospandemia es de una disminución sobre su coeficiente obteniendo un cambio porcentual del 9 % menos para la pospandemia, la estimación del coeficiente no estandarizado para la variable baños en ambos periodos dio como resultado que, para la prepandemia su valor es de 60 y en la pospandemia 30 confirmando una disminución de 30 millones de pesos. Las habitaciones influencian negativamente el precio solamente en la prepandemia, esto quiere decir que en la pospandemia la relación lineal del precio y las habitaciones tuvo un aumento significativo generando una influencia positiva sobre el precio de la vivienda. Las estimaciones no estandarizadas de la variable habitaciones dieron como resultado en la prepandemia $-1{,}95$ y la pospandemia 10,80 demuestran un cambio significativo importante entre ambos periodos sobre la influencia de la variable habitaciones sobre el precio de la vivienda.

La nueva variable que representa el valor por metro cuadrado de cada vivienda obtuvo una influencia positiva sobre los precios de la misma. En su comparativa se evidenció un cambio significativo para la pospandemia, demostrado a partir de los resultados de las estimaciones no estandarizadas, para la prepandemia su valor fue de 100 y la pospandemia 69 obteniendo una diferencia de 31 lo cual indica qué el precio del metro cuadrado disminuyo significativamente para el periodo de la pospandemia.

Las covariables de entorno presentadas como las cercanías a: estaciones de bus, centros comerciales y zonas verdes presentaron diferentes influencias sobre el precio de la vivienda. Las cercanías a estaciones de bus tienen una menor influencia sobre el precio de la vivienda, que es reflejada para ambos periodos, por lo cual su comparativa no generó cambios significativos. La variable cercanías a centros comerciales influencia de manera negativa el precio de la vivienda, dando como resultado que entre más cercano es un bien a un centro comercial menor será su precio. Comparando los resultados de ambos periodos, en la pospandemia esta influencia negativa aumentó, generando una disminución mayor sobre el precio de la vivienda comparada con la prepandemia. Por ultimo, la variable cercanías a zonas verdes presentó el mismo comportamiento mencionado para la variable cercanías a centros comerciales. Su

comparativa en los resultados para ambos periodos, fue el aumento negativo de la influencia sobre el precio de la vivienda.

7 Conclusiones y recomendaciones

7.1. Conclusiones

En esta investigación tuvo como objetivo principal evaluar el efecto de la pandemia sobre los precios de la vivienda usada en la ciudad de Cali, evaluando su evolución en los periodos definidos como prepandemia (año 2019) y pospandemia (año 2020) y la estructura de asociación con las variables de entorno. Se planteó una metodología de comparaciones mediante análisis estadísticos descriptivos y modelación estadística utilizando específicamente la modelación por regresión de mínimos cuadrados parciales.

Como primera parte, en el análisis exploratorio y descriptivo, destacó la heterogeneidad de varianza entre el precio de la vivienda y las variables correspondientes a las características propias de esta. Esto permite concluir que ninguna metodología podría reducir la varianza presentada en los datos ya que la metodología utilizada en la modelación es la que mejor comportamiento tiene con este factor.

Por otra parte, se observó que la cantidad de viviendas ofertadas en el portal web de OLX para el periodo pospandemia fue muy inferior a la cantidad de registros para el periodo anterior. Además, para el periodo pospandemia los precios de la vivienda disminuyeron significativamente, así como el área en metros cuadrados.

El análisis de la influencia del estrato socioeconómico sobre el precio de la vivienda demuestra que en el año periodo pospandemia la mayoría de viviendas ofertadas se concentraban en los estratos 4,5 y 6, caso contrario a lo encontrado en el periodo prepandemia. De la comparación de estos resultados en ambos periodos se puede concluir que el estrato socioeconómico es un factor altamente influyente sobre el precio de la vivienda y del área en metros cuadrados.

Algunas variables conservan el comportamiento presentado en el periodo anterior a la pandemia, particularmente la distribución del área en metros cuadrados, la cual está fuertemente relacionada con el estrato y el tipo de vivienda. Sin embargo, enfocando el análisis sobre los precios de la vivienda y el área en metros cuadrados por tipo de vivienda, se evidencia que en los apartamentos existe una brecha gigante comparada con los precios obtenidos el periodo prepandemia. Por último, las variables del precio por metro cuadrado, la cantidad de baños y de habitaciones no se vieron tan afectadas por esta comparativa, teniendo en sí

un comportamiento casi equivalente al periodo prepandemia.

En este orden de ideas, con la agregación de las covariables de entorno a este análisis, se evidencia en el modelo PLSR que las estimaciones arrojan diferencias significativas en los precios de la vivienda para los periodos prepandemia y pospandemia, los cuales a su vez tienen en sí una naturaleza casi idéntica sobre la influencia del entorno; sin embargo, para efectos de la contextualización del periodo pospandemia se obtuvo que estas variables generaban un peso aún más negativo de lo que generaron en el periodo anterior, lo cual permite concluir que el entorno de la vivienda si se vio afectado por la pandemia COVID-19 y esto a su vez afectó la percepción de compra y el precio del inmueble, puesto que se evidenció que en el periodo pospandemia las personas mostraban preferencia sobre viviendas alejadas de posibles focos de contagio como lo son las zonas verdes, centros comerciales y estaciones de bus.

7.2. Recomendaciones

Con el fin de complementar el trabajo de grado descrito, se recomienda generar bases de datos con registros de más plataformas web para tener una mayor representatividad de la oferta de vivienda en la ciudad. En la conformación de esta base de datos es recomendable agregar nuevas variables de entorno que permitan tener un mejor contexto del entorno del inmueble, por ejemplo: cercanía a supermercados, cercanía a centros de salud, acceso a vías principales, entre otras. Adicionalmente se recomienda que los registros que conformen la base de datos no sean datos brutos y se tenga una periodicidad para realizar la extracción de los datos con web scraping para garantizar datos actualizados.

Pur último, se recomienda tener precaución al utilizar web scraping, realizando un análisis minucioso de los datos extraídos de las páginas web ya que si no se realiza una correcta depuración de estos, la información e inferencia estadística se generaría a partir de datos contaminados, lo cual conduciría a conclusiones inválidas.

Bibliografía

Aguirre, I. (2007). *El mercado hipotecario en Latinoamerica: Una visión de negocio. Antecedentes y oportunidades de desarrollo.* Libros profesionales de empresa. ESIC.

Alcaldía (2022). Alcaldía de santiago de cali. https://www.cali.gov.co/.

Álvarez Zuluaga, D., Betancur Carvajal, C., and Rendon, J. G. (2022). Impacto del covid-19 sobre la demanda y oferta de vivienda nueva no vis en colombia (impact of covid-19 on the demand and supply of new no vis housing in colombia). *Available at SSRN 4058487*.

Carrascal, L. M., Galván, I., and Gordo, O. (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, 118(5):681–690.

Castelblanco Rodriguez, A. J. (2021). Análisis de los efectos del covid-19 que afectaron el mercado de inmuebles residenciales en el municipio de cajicá cundinamarca.

Chris, H. (2013). Scraping the web for arts and humanities. *Norwich, England: University of East Anglia*.

Eduardo, S. C., Meneses-González, M. F., and Vélez Rodríguez, M. J. (2020). Informe especial de estabilidad financiera: análisis de la cartera y del mercado inmobiliario en colombia-segundo semestre de 2020. *Informes Especiales de Estabilidad Financiera-Segundo semestre de 2020*.

Ferrari, C. and González, J. I. (2021). La cuarentena y las distintas actividades económicas. *Razon publica*.

Gaviria Peña, C. A. (2016). Regresión por mínimos cuadrados parciales pls aplicada a datos variedad valuados. *Escuela de Estadística*.

Mankiw, N. G., M., Staines, M. G., and Carril Villarreal, M. d. P. (2012). *Principios de economía (6a. ed.)*. Cengage Learning.

Márquez Ruiz, C. (2017). Modelo de regresión pls.

Molina García, A. (2014). *Análisis de la rentabilidad del mercado inmobiliario como alternativa de inversión en España*. PhD thesis, Universitat Politècnica de València.

Montgomery, D., Peck, E., and Vining, G. G. (2006). Introducción al análisis de regresión lineal. *México: Limusa Wiley*.

OLX/Properati (2022). https://www.properati.com.co/s/cali-valle-del-cauca/Casa/venta. Accedido en Mayo del 2022.

Poeta, S., Gerhardt, T., and Stumpf Gonzalez, M. (2019). Análisis de precios hedónicos de viviendas. *Revista ingenierÃa de construcciÃ*, 34:215 – 220.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Urrea-Ríos, I. L. and Piraján, J. (2020). Impacto de la pandemia covid-19 sobre la economía colombiana. una pandemia temporal con efectos permanentes (impact of the covid-19 pandemic on the colombian economy. a temporary pandemic with permanent effects). *Una pandemia temporal con efectos permanentes (Impact of the COVID-19 Pandemic on the Colombian Economy. A Temporary Pandemic with Permanent Effects)(August 20, 2020)*.

Vega-Vilca, J. C. and Guzman, J. (2011). Regresion PLS y PCA como solución al problema de multicolinealidad en regresion múltiple. *Revista de Matematica Teorica y Aplicaciones*, 18:09 – 20.