



CENTRO DE ESTUDIOS DEMOGRÁFICOS, URBANOS Y AMBIENTALES

DETERMINANTES DEL PRECIO DE LA VIVIENDA EN LA ZMCM:
ANÁLISIS DEL EFECTO DE LOS ELEMENTOS DEL ENTORNO URBANO

Tesis presentada por:

FERNANDO JOSÉ SANTIAGO BACILIO

Para optar por el grado de

MAESTRO EN ESTUDIOS URBANOS

Director de tesis:

DR. JAIME SOBRINO

Lector:

DR. ABEL ALEJANDRO CORONADO IRUEGAS

Ciudad de México, 7 de julio de 2025

Agradecimientos

Quiero agradecer a mi familia y amigos por todo el apoyo que me brindaron en esta etapa.

A mis profesores, por su conocimiento, dedicación y calidad humana que hicieron mi estancia muy enriquecedora.

A mi comité, el Dr. Jaime Sobrino y el Dr. Abel Alejandro Coronado, por su guía y disposición en todo el proceso de redacción de la tesis.

A El Colegio de México, que me abrió sus puertas y contribuyó de forma importante en mi desarrollo profesional y personal.

Y finalmente a la SECIHTI, a la República y a la sociedad mexicana, a quienes agradezco profundamente por financiar mis estudios y este trabajo. Espero que en estas páginas encuentren una modesta contribución para avanzar hacia ciudades más justas y equitativas.

Resumen

La tendencia global a la urbanización implica la necesidad de incrementar la dotación y mantenimiento de infraestructura pública esencial, como de movilidad, equipamientos básicos y servicios públicos, que aseguren una calidad de vida mínima aceptable para la población. Al mismo tiempo, representa una mayor presión financiera sobre los gobiernos locales quienes se encargan de su provisionamiento. La realidad del país muestra, sin embargo, que este nivel de gobierno es muy dependiente de las transferencias federales y que sus recursos son insuficientes para costear el cumplimiento de sus obligaciones. Por tal motivo, es necesario incrementar su capacidad recaudatoria y hacer frente a los retos del futuro. Más aún, el proceso de urbanización que experimentó la Zona Metropolitana de la Ciudad de México (ZMCM) dio como resultado una estructura urbana fuertemente polarizada y desigual. Combinado con la lógica de rentabilidad en la provisión de servicios públicos urbanos, resultado de los procesos de neo liberalización, se ha creado un círculo vicioso en donde los elementos urbanos, financiados públicamente, se concentran en zonas mejor dotadas, perpetuando la desigualdad espacial y exclusión social. Los gobiernos locales tienen a su disposición herramientas muy potentes para atender estos problemas, como los impuestos a la propiedad inmobiliaria, tales como el impuesto predial y la captación de plusvalías. Sin embargo, para su implementación se necesitan análisis detallados de las relaciones entre los elementos del entorno urbano y el precio de los inmuebles, ya que permiten identificar los efectos de la inversión pública en el valor. En México son pocos los estudios que analizan estas relaciones, principalmente por la falta de información sobre el mercado inmobiliario. Además, su análisis requiere consideraciones metodológicas debido a la naturaleza espacial de la información. Por tal motivo, este trabajo analiza la relación entre los elementos del entorno urbano y el precio en el mercado de vivienda de la ZMCM, utilizando técnicas de econometría espacial y aprendizaje automático. Los resultados muestran que la estructura, la localización y la relación de dependencia espacial son variables fundamentales en la conformación del valor. Las viviendas de mayor precio se concentran en zonas centrales consolidadas, con larga trayectoria de inversión pública, y se benefician en gran magnitud de la proximidad a infraestructura y equipamientos urbanos. En contraste, las viviendas de menor valor se ubican en las amplias periferias del norte y oriente de la metrópoli, donde la accesibilidad a infraestructuras de movilidad es limitada y la cercanía a los elementos urbanos presenta menores beneficios.

Índice

<i>Introducción</i>	
<i>Planteamiento del problema</i>	
<i>Preguntas y objetivos</i>	
<i>Hipótesis</i>	
1. <i>Marco Teórico</i>	12
1.1 Vivienda.....	12
1.2 Análisis de precio.....	13
1.2.1 Modelos de localización residencial	
1.2.2 Modelo de precios hedónicos	
1.3 Estimación de la Función de Precios Hedónicos.....	22
1.3.1 Método clásico	
1.3.1.1 Supuestos clave	
1.3.1.2 Información espacial	
1.3.2 Métodos espaciales y semi-paramétricos	
1.3.3 Evidencia empírica	
1.4 Conclusiones.....	33
2. <i>El aprendizaje automático en la estimación de la FPH</i>	35
2.1 Funcionamiento.....	36
2.2 Modelos Agnósticos de Interpretabilidad.....	37
2.3 Aplicación en la metodología hedónica.....	39
2.4 Conclusiones.....	41
3. <i>Contextualización de la zona de estudio</i>	43
3.1 Urbanización y dotación de infraestructura.....	43
3.2 Distribución de población y empleo.....	47
3.3 Conclusiones.....	49
4. <i>Metodología</i>	51
4.4 Obtención de datos.....	51
4.5 Construcción de variables.....	53
4.6 Método de análisis.....	58
5. <i>Análisis empírico</i>	60
5.1 Descripción del mercado de vivienda.....	60
5.2 Determinantes del precio.....	65
5.3 Análisis exploratorio de relaciones complejas.....	77
<i>Conclusiones</i>	83
<i>Bibliografía</i>	87

Lista de siglas y acrónimos

AIC	Criterio de Información Akaike
ALE	Efectos Locales Acumulados
BIC	Criterio de Información Bayesiano
BRT	Autobuses de Rápido Tránsito
CETRAM	Centro de Transferencia Modal
CPyV	Censo de Población y Vivienda 2020
CU	Ciudad Universitaria
DARPA	Agencia de Proyectos de Investigación Avanzada de Defensa de Estados Unidos
DCN	Distrito Central de Negocios
DDF	Departamento del Distrito Federal
DENUE	Directorio Estadístico Nacional de Unidades Económicas
ENVI	Encuesta Nacional de Vivienda 2021
FI	Importancia de Variables
FPH	Función de Precios Hedónicos
GAM	Modelos Aditivos Generalizados
GLS	Mínimos Cuadrados Generalizados
GWR	Regresión Geográficamente Ponderada
IMSS	Instituto Mexicano del Seguro Social
INEGI	Instituto Nacional de Estadística y Geografía
INFONAVIT	Instituto del Fondo Nacional de la Vivienda para los Trabajadores
IPN	Instituto Politécnico Nacional
ISSSTE	Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado
LM	Multiplicadores de Lagrange
LIME	Explicaciones Agnósticas de Modelos Locales Interpretables
LISA	Índice Local de Autocorrelación Espacial I Moran
LWR	Regresión Localmente Ponderada

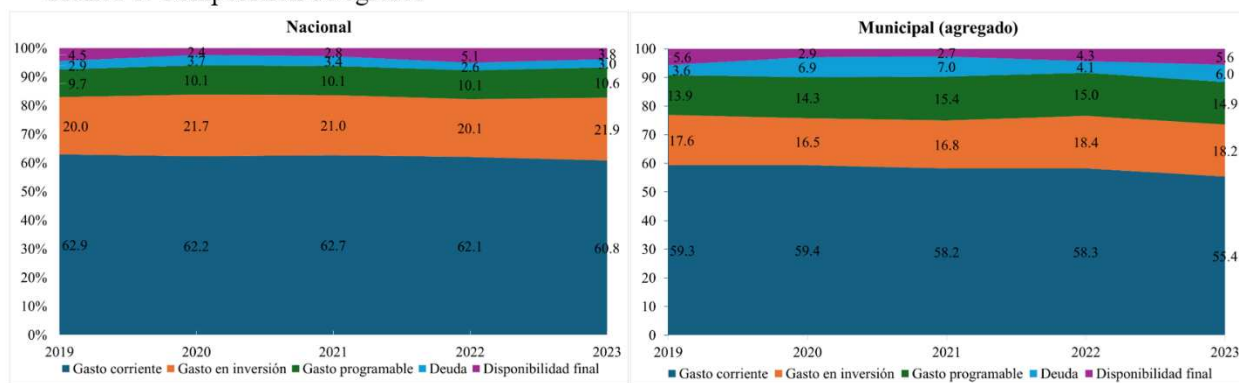
MCO	Mínimos Cuadrados Ordinarios
MGWR	Modelo Mixto de Regresión Geográficamente Ponderada
MV	Máxima Verosimilitud
ONU	Organización de las Naciones Unidas
PD	Dependencia Parcial
RMSE	Raíz del Error Cuadrático Medio
RTP	Red de Transporte de Pasajeros
SAR	Modelo Espacial Autorregresivo
SEM	Modelo de Error Espacial
SHAP	Explicaciones Aditivas de Shapley
SIG	Sistemas de Información Geográfica
STC	Sistema de Transporte Colectivo Metro
UAM	Universidad Autónoma Metropolitana
VIF	Factor de Inflación de la Varianza
XIA	Inteligencia Artificial Explicable
ZMCM	Zona Metropolitana de la Ciudad de México

Introducción

El mundo avanza hacia un futuro cada vez más urbano. De acuerdo con ONU-Hábitat, en 2022 el 56% de la población habitaba en zonas urbanas y para 2050 se estima que la proporción aumentará a 68%. En América Latina, la población urbana alcanzará el 87%, consolidándose como una de las regiones más urbanizadas del mundo (ONU-Hábitat, 2022). Particularmente, en México se proyecta que entre el 70% y 80% de la población vivirá en asentamientos urbanos (ONU-Hábitat, 2017).

Esta tendencia a la urbanización implica una creciente presión sobre la capacidad de los gobiernos locales para financiar infraestructura, bienes y servicios públicos que aseguren una calidad de vida mínima aceptable para sus habitantes. Los recursos públicos suelen ser insuficientes para solventar el costo de la infraestructura, especialmente en contextos de rápido crecimiento poblacional y expansión territorial. En el caso de México, al analizar la composición de egresos nacional y del agregado subnacional (Gráfica 1), se observa que la mayor proporción del gasto se destina a gasto corriente, es decir, gastos administrativos, mientras que solo al rededor del 20% se destina a inversión en infraestructura.

Gráfica 1. Composición de egresos



Fuente: Elaboración propia con datos de INEGI, 2024

Más aún, el proceso de urbanización que experimentó la ZMCM dio como resultado una estructura urbana marcada por una fuerte polarización y desigualdad social. Los grupos de menores ingresos se concentran en áreas periféricas con escaso acceso a infraestructura de movilidad, equipamientos públicos de calidad y oportunidades laborales, en tanto que los grupos de mayores ingresos se localizan en zonas privilegiadas, con alta accesibilidad a los componentes de la ciudad (Montejano Escamilla et al., 2018).

En las últimas décadas, la provisión de servicios públicos urbanos, históricamente responsabilidad estatal, experimentó un creciente proceso de mercantilización dando como resultado que los servicios públicos básicos u obras de infraestructura, como el transporte o equipamientos urbanos, se gestionen como bienes privados sujetos a lógicas de rentabilidad, relegando su carácter de derecho ciudadano. Esto ha reforzado la distribución espacial desigual de los infraestructura y equipamientos, ya que las inversiones públicas se concentran en zonas de mayor nivel socioeconómico, en donde los márgenes de rendimiento económico son más altos. Por ejemplo, un estudio de Flores y Caracheo (2021), sobre la calidad del servicio de agua potable en la Ciudad de México, encontró una relación directa con la capacidad económica de las alcaldías, mostrando patrones de exclusión territorial.

Estas circunstancias generan un círculo vicioso que refuerza la desigualdad urbana. Las áreas privilegiadas se vuelven más atractivas para la escasa inversión pública y privada, lo que revaloriza sus propiedades. En contraste, las periferias marginadas, que carecen de acceso a infraestructura de transporte y equipamientos básicos de calidad, se mantienen en las mismas circunstancias por la poca inversión que se realiza, perpetuando la desigualdad espacial y exacerbando la brecha entre grupos sociales.

Para atender estos problemas, los gobiernos locales tienen a su disposición potentes herramientas como los impuestos inmobiliarios, tales como la captura de plusvalías y el impuesto predial, que permiten capturar la valorización de la inversión pública en la propiedad inmobiliaria y ayudan a incrementar los ingresos locales. Sin embargo, para su implementación se necesitan análisis detallados que muestren las relaciones entre los elementos urbanos y el precio de las viviendas; de esta manera se puede identificar el efecto de la inversión pública en el valor inmobiliario. Esta información podría servir para diseñar mecanismos de financiamiento más justos y equitativos cuyos recursos generados sean utilizados para financiar proyectos en zonas marginadas, promoviendo una redistribución de beneficios y reducción de brechas sociales.

Planteamiento del problema

A pesar de su importancia, hay una falta de investigaciones que analicen los efectos de los elementos urbanos en el precio de las viviendas. Esto se debe, principalmente, a la falta de información sobre el mercado inmobiliario y a la dificultad de conseguir estimaciones fiables. La mayoría de los estudios recientes sobre la vivienda en la ZMCM han abordado problemas como la

división social del espacio, el uso del suelo y las políticas de vivienda, dejando sin atender el análisis de los precios (Sobrino, 2014). La investigación más reciente sobre los efectos de los elementos urbanos fue realizada por Ramírez de la Cruz et al., (2020) quienes exploraron la relación entre la infraestructura pública y la capitalización de sus beneficios por parte de desarrolladores inmobiliarios en la ciudad de México. Dicho trabajo adoptó la teoría hedónica y aplicó un modelo paramétrico de mínimos cuadrados ordinarios (MCO) para medir la valorización de estos elementos de la ciudad.

Las relaciones entre los atributos de la vivienda, en especial los elementos urbanos, y el precio necesitan ser analizadas cuidadosamente. La naturaleza espacial de los datos inmobiliarios y de los elementos urbanos provoca problemas en los modelos tradicionales de MCO para estimar de forma fiable los efectos marginales de los atributos. De no cumplir con ciertos requisitos metodológicos, pueden surgir problemas relacionados con sesgos, inconsistencias y eficiencia de los coeficientes.

La información localizada o con dimensión espacial presenta efectos denominados como dependencia espacial y no estacionariedad a lo largo del territorio, que pueden generar problemas en las estimaciones de los modelos. Estos efectos se refieren a la existencia de patrones espaciales no aleatorios, aglomeraciones que concentran a ciertos tipos homogéneos de población y vivienda. La existencia de aglomeraciones puede provocar que los modelos realicen estimaciones ineficientes y sesgadas, invalidando cualquier análisis inferencial.

Por otro lado, los atributos del entorno, al estar fuertemente vinculados al espacio pueden presentar efectos no lineales o complejos. Por ejemplo, la cercanía a algunas estaciones de transporte público, que concentran actividades que generan externalidades negativas, puede tener un efecto negativo en el precio de las viviendas más cercanas, pero después de cierto umbral de distancia la relación se invierte debido a los beneficios en términos de accesibilidad aun presentes. Otro ejemplo es el de Chen et al., (2024) quienes encontraron, en la ciudad estadounidense de Lucas County, que la relación entre las áreas verdes y los precios de la vivienda presentan una relación no lineal debido a efectos de umbral y a efectos marginales decrecientes. La accesibilidad a espacios verdes se asocia con una fuerte valorización positiva hasta los 400 m, sin embargo, superando esa distancia, la magnitud del efecto se reduce. Este tipo de relaciones son difíciles de capturar por los modelos de regresión tradicionales que asumen un efecto lineal y estable de cada atributo.

Las características de la información espacial y sus implicaciones metodológicas muestran la necesidad de adoptar métodos que permitan dar cuenta de tales complejidades. Por tal motivo, este trabajo explora los efectos de los atributos de la vivienda en el precio, con especial atención a los elementos de entorno urbano, utilizando técnicas de econometría espacial y de aprendizaje automático. La información generada por este análisis podrá ser utilizada para mejorar herramientas como la captura de plusvalías y el impuesto predial, las cuales tienen un alto potencial para ayudar a atender los problemas de debilidad financiera local, desigualdad espacial y polarización social en la ciudad.

Por lo tanto, la pregunta y el objetivo general de la investigación es:

¿Cuáles son los efectos de los atributos de la vivienda, especialmente los elementos del entorno urbano, en la conformación del precio en la ZMCM?

El objetivo general de esta investigación es identificar el efecto que tienen las características de la vivienda, especialmente los elementos del entorno urbano, en la conformación de su valor en el mercado de vivienda de la ZMCM.

Preguntas particulares:

¿Cuáles son los atributos que más contribuyen en la conformación del precio?

¿Cuál es el efecto de los elementos del entorno urbano en el precio?

¿Cuáles son las características de las relaciones existentes entre los atributos y el precio?

Objetivos particulares:

Identificar los atributos más importantes en la conformación del precio.

Estimar el efecto de los elementos del entorno urbano en el precio.

Analizar los detalles de las relaciones existentes entre los atributos de la vivienda y el precio.

Hipótesis:

El efecto de los elementos urbanos en la valorización de las viviendas, además de depender de los beneficios que representa su accesibilidad o las externalidades que puedan generar, también

depende de la localización geográfica y contextualización histórica de su espacio inmediato, así como de las relaciones de dependencia espacial que existen entre las viviendas de la ZMCM.

Marco teórico

Vivienda

La vivienda es una de las necesidades materiales más esenciales de la humanidad; abarca múltiples dimensiones y, por lo tanto, puede ser analizada desde diferentes visiones (Balchin et al., 2000). Desde una visión sociológica, además de representar un refugio, también contiene atributos subjetivos como estatus e identidad social, los cuales motivan la agrupación espacial de individuos con características semejantes, lo que a su vez refleja la división de la estructura social. Tales agrupaciones muestran y definen el lugar de cada individuo dentro de la ciudad y revelan la forma diferenciada de acceso y apropiación de los recursos urbanos. Se argumenta que el lugar donde viven las personas determina en gran medida las oportunidades a las que tienen acceso (Garrido Rodríguez et al., 2023).

Desde la economía espacial, la vivienda, como mercancía, es única de múltiples formas y comparte ciertos aspectos con la visión sociológica. Es durable y espacialmente inmóvil. Su localización implica efectos relacionados con el vecindario en términos de externalidades positivas y negativas, y con la accesibilidad a los centros de empleo, lo que significa que su posición relativa y el entorno urbano son fundamentales en la conformación de su valor. Es heterogénea, existen viviendas de distinto tamaño, calidad, edad, estilo, amenidades, vecindarios y localizaciones. La heterogeneidad implica que el mercado de vivienda puede ser desagregado en diferentes submercados, según sus características estructurales, patrones de accesibilidad y características de la población que las habita. Tiene un precio alto, por lo que su demanda también depende de la disponibilidad de crédito y financiamiento. Mudarse es costoso no solo en términos monetarios, por lo tanto, cuando el ingreso o las preferencias cambian, los consumidores no ajustan instantáneamente su consumo de vivienda. Y su producción está relacionada con el crecimiento económico (Bluestone et al., 2008; Sobrino, 2014).

Estas particularidades caracterizan el funcionamiento del mercado residencial. Su reconocimiento, y el de las complejidades que involucran, condujo al desarrollo de diferentes cuerpos teóricos para explicar la racionalidad, bajo principios económicos, de la conformación de los precios y las dinámicas espaciales del mercado de vivienda (Bluestone et al., 2008).

Análisis de precios

En el mercado de vivienda, como en otros mercados, la interacción entre oferta y demanda puede considerarse como fundamental en la creación de los niveles de precios, en donde estos asignan las viviendas a los consumidores con los cuales sus ingresos se empatan con los costos (Kirby, 1976). La descripción de este equilibrio tradicionalmente se lleva a cabo en términos de la teoría del equilibrio del consumidor, sin embargo, en el caso de la vivienda, esta no permite agregar a los individuos para encontrar una solución de mercado. De forma clásica se puede derivar la curva de demanda individual variando el precio del producto. Si las curvas de demanda individuales se pueden agregar, se obtiene la demanda de mercado. Pero en este caso particular, vivienda y suelo están tan interconectados que no se puede derivar esta curva. La curva de demanda solo se podría obtener para ubicaciones específicas, por lo tanto, la demanda del mismo individuo varía con la localización (Alonso, 1964).

Por tal motivo se han desarrollado distintos ajustes a los conceptos tradicionales usados para formular el equilibrio de mercado con la intención de considerar las características particulares de la vivienda y usar la racionalidad económica para explicar sus procesos y productos. Estos ajustes han sido materializados en dos perspectivas analíticas fundamentales, las cuales están fuertemente relacionadas (Sobrino, 2014).

Modelos clásicos de localización

La primera y más antigua es la relacionada con los modelos de localización residencial desarrollados por Wingo, (1963), Alonso (1964) y Muth, 1969 (O 'Sullivan, 2012; Sobrino, 2014). Estos examinan la selección del lugar de residencia como un procedimiento de maximización de utilidad restringida por el ingreso (Hirsch, 1973). El objetivo, además de explicar cómo es que el suelo adquiere su valor, es saber cómo es que se asigna entre distintas categorías de usuarios, incluyendo empresas y residentes (Bluestone et al., 2008).

El modelo se desarrolló a partir del trabajo de Von Thunen (1826) quien estableció, para el caso agrícola, que la renta de un predio se determina por el excedente económico que puede brindar una localización, el cual es determinado por la diferencia entre los ingresos y los costos totales, que incluyen a los costos de transporte (Alonso, 1964).

Supone que los agentes son idénticos en el sentido de que producen la misma cantidad de productos con los mismos costos, pero difieren en su localización. Por lo tanto, aquellos localizados cerca del mercado central tienen un mayor excedente económico debido a los bajos costos de transporte. Al alejarse del centro, los costos de transporte aumentan y, por lo tanto, el excedente económico disminuye. Al alejarse más se llega a un punto en donde los costos totales se igualan a los ingresos totales, marcando el límite geográfico de esa actividad y el umbral de la ciudad, ya que más allá de ese punto la actividad agrícola se vuelve financieramente inviable (Bluestone et al., 2008). Este proceso es conocido como el principio residual, siendo que el monto de renta es el residuo después de pagar los costos de producción, los costos de transporte y la ganancia media de la actividad (Garrido Rodríguez et al., 2023).

Este mecanismo fue adaptado al entorno urbano en donde no existe solo un tipo de usuario, sino distintos tipos, con actividades diferentes, que compiten por acceder al mercado central, también conocido como Distrito Central de Negocios (DCN). En este caso no solo valoran los costos de transporte, sino que también la accesibilidad y las economías de aglomeración que brinda el DCN. Por ejemplo, en el caso del sector de la información, valoran la accesibilidad debido a que su actividad exige el intercambio de información tácita, lo cual solo se logra a través del contacto cara a cara entre los trabajadores, cuya alta especialización vuelve sus costos de oportunidad muy elevados (O'Sullivan, 2012). Además, la disposición de este sector a pagar depende de las ganancias que reporta, de tal manera que, para este tipo de agentes, la curva de demanda tiene una pendiente negativa pronunciada cerca del centro, lo que resulta en una disposición a pagar por usar suelo central muy alta, pero que al alejarse disminuye rápidamente.

Dentro de este marco de análisis urbano se incluye al sector residencial, el cual compite con los demás usuarios por el uso del suelo (O'Sullivan, 2012). A diferencia de las empresas, el factor locacional clave para los residentes son los costos de traslado, es decir, el costo de los viajes que realizan para llegar a su lugar de trabajo. Cabe decir que el modelo residencial elabora algunas suposiciones y simplificaciones de la realidad para hacer el análisis más claro y sencillo. Estas incluyen: que todos los residentes tienen las mismas preferencias, lo que resulta en una vivienda estándar; que tienen conocimiento perfecto y existe un mercado en equilibrio; que los costos de traslado son estrictamente monetarios, además de incrementar proporcionalmente con la distancia; que al menos un miembro de la familia viaja al DCN; que otros desplazamientos son

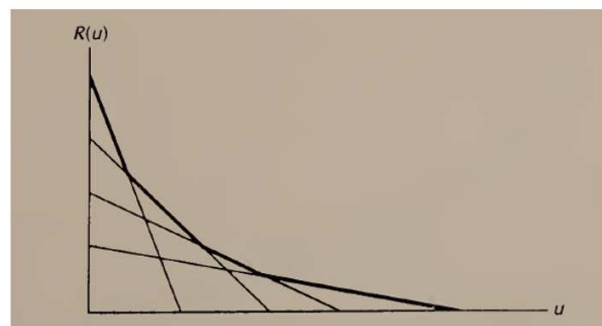
insignificantes; que los servicios públicos y los impuestos son iguales en todos los lugares; y que la calidad ambiental es la misma en todos los lugares. Estas suposiciones convierten al DCN en el punto focal para los residentes de la ciudad (O'Sullivan, 2012).

La curva de demanda de los residentes tiene una pendiente negativa que refleja la menor disposición a pagar por vivienda al alejarse del centro debido a los mayores costos de traslado. También es convexa, ya que los residentes enfrentan una restricción presupuestaria y además de vivienda consumen otros bienes, como alimentos, vestimenta o entretenimiento. Por tanto, eligen la combinación de vivienda y otros bienes, dado el precio de la vivienda a cierta distancia, que le reporte la mayor satisfacción dentro de la restricción de su ingreso. En este contexto, el consumo de vivienda implica un costo de oportunidad, pues al gastar más en vivienda se sacrifica el consumo de otros bienes. Cuanto mayor sea este costo, menos metros cuadrados se estarán dispuestos a consumir, priorizando otros componentes de su bienestar (O'Sullivan, 2012).

Esta lógica influye en la elección de la ubicación residencial. Dado que los precios de la vivienda suelen disminuir con la distancia al centro, los hogares evalúan si el ahorro en renta compensa el mayor esfuerzo en tiempo y dinero que implica trasladarse diariamente. El modelo predice que los hogares se irán desplazando hacia zonas más alejadas mientras el beneficio del menor costo de vivienda supere el sacrificio de accesibilidad. Sin embargo, este proceso tiene un límite y llegará un punto en el que la incomodidad o el costo de transporte iguallen o superen el ahorro en renta, y en ese punto se alcanza un equilibrio espacial (Straszheim, 1975). Así, la localización residencial es un resultado del balance entre accesibilidad, precios del suelo y preferencias individuales.

En este marco de análisis, todos los agentes urbanos compiten por el uso del suelo y, ya que el terrateniente busca maximizar sus ganancias, las distintas curvas de los potenciales usuarios se comparan en todas las localizaciones y en cada localización el suelo se asigna al mayor postor. La superposición de estas disposiciones a pagar genera una curva envolvente de renta

Gráfica 2. Curva envolvente de renta



Fuente: Alonso, 1964

(Gráfica 2), donde en cada ubicación prevalece el uso más productivo (residencial, comercial o industrial), definiendo el patrón espacial de la ciudad (Alonso, 1964). Esta curva representa el

equilibrio locacional y se alcanza cuando los costos y beneficios de cambiar de ubicación se igualan, garantizando que todos los agentes obtengan el mismo nivel de satisfacción, independientemente de su localización (Mills y Hamilton, 1984).

Aplicaciones empíricas

La aplicación de este marco ha demostrado ser útil para describir tendencias generales de la organización espacial de la sociedad. El modelo describe de forma general los patrones espaciales de la mayoría de las ciudades en los países desarrollados en los años de 1980 y 1990, como una mayor densidad de población e intensidad constructiva cerca del centro de las ciudades. Respecto al gradiente de precios, los estudios realizados han mostrado resultados inconsistentes, ya que en algunas ciudades se encuentra una relación positiva con la distancia al DCN, es decir un aumento del precio al alejarse del centro, en lugar de una relación negativa como lo predice el modelo. Esta inconsistencia puede deberse a la zona de estudio o la escala del análisis (Kirby, 1976; Mills y Hamilton, 1984).

El modelo también explica la observación empírica sobre la localización de las familias de bajos ingresos cerca del centro de la ciudad y las de altos ingresos en los suburbios. Esto es resultado de la diferente valorización del costo de transporte en función del nivel de ingresos. En las ciudades estadounidenses, para las familias de bajos estratos representa una proporción significativa de sus ingresos y ante cambios en los costos de transporte las personas responden reduciendo la distancia al centro y compensando el aumento en el precio con un menor consumo de vivienda, incluso llegando al hacinamiento en algunos casos. Las familias de altos ingreso, por otro lado, eligen su ubicación considerando otros factores, como la disponibilidad de espacio y áreas verdes (Bluestone et al., 2008; Mills y Hamilton, 1984; O'Sullivan, 2012).

Cabe destacar que en la mayoría de las ciudades de América Latina el patrón es inverso, ya que las familias de altos ingresos se localizan cerca del centro, mientras que las de bajos ingresos se localizan en las periferias de la ciudad, por lo que este marco de análisis podría no brindar explicaciones claras sobre la localización residencial en las ciudades latinoamericanas (Ruvalcaba y Schteingart, 2012; Sobrino, 2014; Zubicaray, 2015). Además, si bien la teoría urbana predice una relación negativa entre los precios de la tierra y la distancia del DCN, la suavidad y simetría encontrada en los modelos teóricos pueden ser guías imprecisas de los precios locales, ya que

pueden ser influenciados por centros de empleo secundarios y terciarios y por externalidades positivas o negativas segregadas en el espacio (Redfearn, 2009).

Un obstáculo mayor del análisis neoclásico del mercado de vivienda es la heterogeneidad en la vivienda (Witte et al., 1979). El modelo supone que las preferencias de los consumidores son iguales y, por tanto, supone una vivienda homogénea en términos de calidad. Sin embargo, una de las características de la vivienda es que es heterogénea en relación con el tamaño, calidad, amenidades, equipamientos e infraestructuras disponibles. Al necesitar suelo para su producción, está fija en el espacio, por lo que las familias no solo compran los atributos estructurales del inmueble, sino también una variedad de servicios en una ubicación particular (Kain y Quingley, 1975; Wilkinson, 1973). Esto convierte a la vivienda en un bien complejo cuya percepción de valor no solo proviene de sus atributos internos sino de la conjunción con los atributos externos. Los modelos de accesibilidad ignoran la evidencia del mundo real de que los precios de viviendas en áreas igualmente accesibles pueden esconder amplias variaciones por diferencias en el atractivo del vecindario (Richardson et al., 1974).

El reconocimiento de estas particularidades provocó una transición en la literatura internacional de las teorías de localización a nuevas aproximaciones metodológicas para analizar el mercado de vivienda en las ciudades modernas. Para considerar la inherente heterogeneidad de la vivienda, muchos análisis recientes han visto la vivienda en términos hedónicos o de atributos (Witte et al., 1979).

Modelo de precios hedónicos

La segunda perspectiva de análisis del mercado de vivienda constituye una aproximación distinta de los modelos de localización, empleada para introducir las complejidades asociadas a la heterogeneidad del bien. El modelo fue desarrollado por Rosen, (1974) a partir de los avances de Lancaster (1966), sobre la teoría del equilibrio del consumidor. Su principal innovación consiste en romper con la visión tradicional de que los bienes son objetos directos de utilidad, y establece que son las propiedades de los atributos del bien, de donde surge la utilidad (Lancaster, 1966). De tal forma, el modelo de Rosen descompone a la vivienda en un conjunto de características, cada una de las cuales ofrece un flujo particular de servicios al residente (Griliches, 2013; Richardson et al., 1974; Rosen, 1974). El objetivo del modelo es describir el equilibrio competitivo en un

mercado de bienes heterogéneos en el que vendedores y compradores interactúan (Rosen, 1974; Taylor, 2008).

La vivienda se introduce como un bien diferenciado, descrito por un vector de atributos $V = (v_1, v_2, \dots, v_n)$. Cada vivienda tiene un precio de mercado observado asociado a un conjunto particular de valores del vector V , lo que permite revelar implícitamente una función de precios $P_i = P(v_{i1}, v_{i2}, \dots, v_{in})$. La derivada parcial de $P(\cdot)$ respecto a la característica v_j , $\partial P / \partial v_j$, representa el precio marginal implícito de dicha característica, manteniendo las demás constantes (Brown y Rosen, 1982; Greenstone, 2017; Rosen, 1974).

El equilibrio competitivo se fundamenta en las decisiones de compradores y vendedores quienes, bajo los supuestos de competencia perfecta e información completa, maximizan su utilidad o beneficios, sujetos a las restricciones de su ingreso y del mercado. Para los consumidores, la utilidad depende de dos elementos: el consumo de un bien compuesto, X , que representa todos los demás bienes; y el conjunto de atributos que integran la vivienda, V . Así, su función de utilidad se define como:

$$U = U(X, V)$$

Si se asume que el consumidor solo compra una vivienda, entonces la restricción presupuestaria se expresa como $I = P + X$, donde I es el ingreso. Sustituyendo la restricción presupuestaria en la función de utilidad resulta en:

$$U = U(I - P, v_1, v_2, \dots, v_n)$$

Para analizar la demanda respecto a un atributo particular, se invierte la función de utilidad, manteniendo constantes todas las características salvo v_j . Esto permite derivar la disposición marginal a pagar por ese atributo:

$$D_j = D_j(I - P, v_j; V_{-j}^*, u^*)$$

Aquí, u^* es el máximo nivel de utilidad alcanzable dado el ingreso y V_{-j}^* es el vector óptimo de atributos, excluyendo la característica v_j . Esta función representa la curva de demanda o de indiferencia, ya que muestra la cantidad máxima que un individuo está dispuesto a pagar por diferentes cantidades de v_j , manteniendo la utilidad constante (Greenstone, 2017; Taylor, 2008).

El modelo es simétrico, por lo que del lado de la oferta la curva de oferta se construye de forma análoga. Se asume que los productores son heterogéneos debido a diferencias en sus funciones de costos. Invertiendo la función de ganancia del productor se puede derivar la curva de oferta para la característica v_j :

$$O_j = O_j(I - P, v_j; V_{-j}^*, \pi^*)$$

Donde π^* es la ganancia máxima disponible dada su función de costos y los precios de mercado.

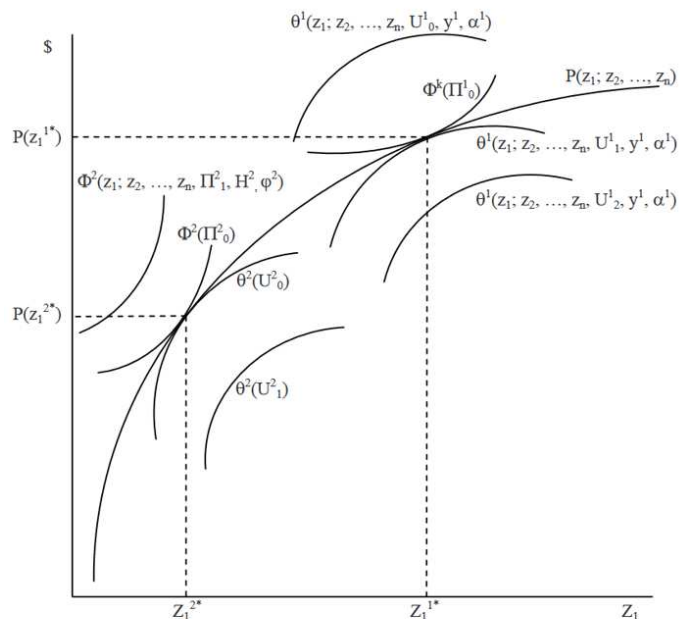
La condición de equilibrio del consumidor se alcanza cuando se maximiza la utilidad sujeta a la restricción presupuestaria. Esta condición requiere que el consumidor elija una combinación óptima del bien compuesto X y de cada atributo de la vivienda v_j , de manera que la tasa marginal de sustitución entre v_j y X sea igual al precio marginal implícito del atributo v_j . Matemáticamente, esto se expresa como:

$$\frac{\frac{\partial U}{\partial v_j}}{\frac{\partial U}{\partial X}} = \frac{\partial P}{\partial v_j}$$

Esta igualdad representa una situación de equilibrio en la que el beneficio adicional de obtener una unidad extra del atributo v_j , en relación con el bien numerario X , se iguala al costo adicional que impone el mercado por esa unidad extra, es decir, su precio marginal implícito. En este punto, el consumidor no tiene incentivos para modificar su elección, ya que cualquier cambio resultaría en una menor utilidad o mayor costo (Greenstone, 2017; Taylor, 2008).

En equilibrio, la interacción entre compradores y vendedores en el mercado de vivienda genera una coincidencia entre sus respectivas decisiones óptimas. Las funciones de demanda de los consumidores y de oferta de los

Gráfica 3. Función de Precios Hedónicos



Fuente: Taylor, 2008

productores se intersecan en un conjunto de puntos en los que ambos maximizan su utilidad o beneficios. Estas intersecciones comparten un gradiente común que define la pendiente de la función de precios implícitos del mercado. Como resultado, las observaciones empíricas de precios de vivienda, P_i , conforman una envolvente conjunta de curvas de oferta y demanda individuales, reflejando las condiciones de tangencia que se presentan en cada punto de equilibrio (Gráfica 3) (Rosen, 1974; Taylor, 2008).

Esta envolvente, que describe cómo varía el precio de mercado en función de los atributos del bien, recibe el nombre de Función de Precios Hedónicos (FPH). Debido a la heterogeneidad en preferencias e ingresos entre los consumidores, la FPH puede ser tangente a múltiples curvas de demanda individual, cada una correspondiente a un tipo distinto de comprador con su propia disposición marginal a pagar (Greenstone, 2017; Taylor, 2008).

Método de aplicación

La aplicación de la metodología hedónica se realiza mediante un proceso de dos etapas interrelacionadas. En la primera etapa, se estima la FPH o P_i mediante un análisis de regresión múltiple, utilizando datos transversales de los precios de venta de las viviendas y sus atributos (Griliches, 1971; Taylor, 2008). Esta etapa permite calcular los precios marginales implícitos de los atributos de la vivienda, los cuales revelan la disposición marginal a pagar por una unidad adicional de un atributo.

Sin embargo, la FPH por sí misma tiene un limitado rango de análisis, ya que no es suficiente para conducir un análisis del bienestar para cambios no marginales, debido a que la información de precios observada solo revela un punto de cada función de demanda individual. Otros puntos son observados para otros individuos que, se presume, tienen diferentes gustos y niveles de ingresos. Por lo tanto, la FPH no puede ser usada para determinar los cambios no marginales de los atributos de interés (Greenstone, 2017).

La solución de Rosen fue una segunda etapa que vincula los precios implícitos con las funciones de demanda. En esta, los precios marginales implícitos se usan como variables dependientes, incorporando factores como ingresos, características demográficas y disponibilidad de sustitutos a las covariables (Witte et al., 1979). Sin embargo, la propuesta enfrenta obstáculos prácticos. La identificación de las funciones de demanda requiere datos detallados sobre las preferencias

individuales y variabilidad exógena en atributos, condiciones raramente cumplidas en mercados inmobiliarios. Además, la segunda etapa se considera un jardín fértil de problemas de identificación, específicamente respecto a la distinción entre efectos de oferta y demanda (Witte et al., 1979; Taylor, 2008).

A pesar de esta limitación, el trabajo de Rosen inició una rica literatura empírica y teórica, explorando el rol de los atributos de la vivienda en la toma de decisiones del consumidor. Como se mencionó, la elección de vivienda, además de relacionarse con el consumo de las características estructurales, también se relaciona con las características locacionales resultado de las externalidades que enfrenta una ubicación particular. Por tanto, las elecciones observadas sobre vivienda revelan al investigador información sobre las preferencias subyacentes por estos elementos u otras características de interés (Taylor, 2008).

De hecho, muchas mercancías, como la vivienda, pueden ser vistas como una canasta de atributos individuales para los cuales no existen mercados explícitos. El modelo de Rosen delimita un procedimiento para estimar la disposición a pagar de los individuos por bienes y servicios públicos lo que permite estimar relaciones de extraordinaria importancia para la determinación de una política óptima. Ejemplos clásicos incluyen la calidad del medio ambiente, de las escuelas, de equipamientos e infraestructura urbana, así como de fenómenos sociales como el crimen y la segregación (Greenstone, 2017).

La gran mayoría de aplicaciones del modelo hedónico se enfocan únicamente en la primera fase del modelo, es decir, la estimación de la función de precios hedónicos (Taylor, 2008). Sin embargo, en la época del artículo de Rosen, la profesión económica no apreciaba adecuadamente la dificultad de obtener estimaciones causales de las relaciones empíricas, en particular de corte transversal. Por tanto, el trabajo prestó muy poca atención a los retos metodológicos de estimaciones consistentes de la FPH. De ser inconsistentes resulta en una estimación errónea de la disposición marginal a pagar por los atributos, invalidando cualquier análisis del bienestar ante cambios no marginales (Greenstone, 2017).

Estimación de la Función de Precios Hedónicos

La estimación de la función de precios se realiza de una variedad de formas. Entre los métodos más usados destaca el uso de regresión lineal múltiple mediante Mínimos Cuadrados Ordinarios (MCO) o máxima verosimilitud (MV) (Taylor, 2008). Comúnmente se emplea una especificación semilogarítmica:

$$\ln(P_i) = \alpha + \sum_{j=1}^J \beta_j Z_{ji} + \varepsilon_i,$$

Donde el logaritmo natural del precio de la *i*-ésima vivienda es una función de la *j*-ésima característica que se asume, influye en el precio, Z_{ji} son sus atributos, e es el término de error, α y β son los coeficientes estimados.

La estimación mediante este método tiene diferentes ventajas. En primer lugar, el análisis de regresión múltiple es un dispositivo analítico conceptualmente sólido y muy potente que combina la teoría de la probabilidad con el cálculo, lo que permite ordenar las influencias cruzadas que afectan a los valores inmobiliarios. En segundo lugar, este método se ajusta perfectamente a la definición de valor de mercado, expresado como el precio más probable que debiera pagarse en mercados competitivos. En tercer lugar, el enfoque no se limita a producir estimaciones de valor, sino que añade perspectivas muy útiles sobre las causalidades del valor de la propiedad. Por lo tanto, se considera una herramienta de toma de decisiones que aporta inteligencia de mercado, más cuando se utiliza en combinación con sistemas de información geográfica (SIG). Este método se posiciona como una herramienta idónea para cuantificar el valor de mercado al ofrecer un enfoque estandarizado, replicable y respaldado por evidencia estadística, lo que garantiza transparencia en la interpretación de señales del mercado y robustez en las conclusiones (Kauko y D' Amato, 2008).

Supuestos clave

Sin embargo, para poder obtener todas estas ventajas, los modelos deben cumplir con ciertos supuestos relacionados con el proceso de generación de la información. El más importante se refiere a errores esféricos, es decir, homoscedasticidad, ausencia de autocorrelación, así como una distribución normal de los residuos ($\varepsilon \sim N(0, \sigma^2)$). La homoscedasticidad se refiere a una varianza constante de los residuos. La ausencia de autocorrelación se refiere a que los errores son independientes, es decir, a que el error de una observación no está correlacionado con el error de otra observación. También tiene que ver con el grado de asociación de las variables independientes,

en este caso atributos de la vivienda, o de la variable dependiente, es decir los precios de las viviendas. La normalidad en errores permite realizar inferencias válidas mediante pruebas t , F e intervalos de confianza, ya que el teorema de Gauss-Márkov asegura que, bajo estas condiciones, los estimadores MCO sean los mejores, lineales, insesgados y eficientes (McCluskey y Borst, 2007; Turizo, 2024). También, el uso de MCO supone estabilidad en los parámetros, lo que significa que los supuestos se cumplen para todas las observaciones, independientemente de su ubicación territorial (Basu y Thibodeau, 1998).

Además, asume que la relación entre las covariables y la variable de respuesta es lineal en los parámetros. Esto implica que el efecto de cada variable es constante, no depende de otras variables o de otras observaciones. Por lo tanto, el modelo no puede capturar relaciones complejas entre variables porque está basado en supuestos de linealidad estrictos y su estructura matemática solo permite modelar efectos aditivos y proporcionales (Potrawa y Tetereva, 2022).

Estos supuestos aseguran que los estimadores MCO sean confiables y las inferencias estadísticas válidas. El incumplimiento de tales restricciones puede provocar que las estimaciones de los parámetros, es decir, los precios marginales implícitos de los atributos dejen de ser fiables (Kauko y D' Amato, 2008; Turizo, 2024).

Información espacial

La información sobre viviendas, y en general la información georreferenciada o con dimensión espacial se caracteriza por presentar efectos espaciales, los cuales son decisivos y deben ser considerados en análisis econométricos para evitar invalidar los resultados obtenidos. Estos efectos son dos: dependencia espacial, también llamada autocorrelación espacial, y, heterogeneidad espacial o no estacionariedad (Anselin, 1988).

La dependencia espacial se refiere a una relación de dependencia entre observaciones espacialmente cercanas y obedece a la primera ley de Tobler la cual establece que “todo está relacionado con todo, pero las cosas cercanas están más relacionadas” (Tobler, 1970, p.236). En el caso de la vivienda, la dependencia espacial se puede producir por diferentes razones, por un lado, los atributos subjetivos y la idiosincrasia de la colonia o barrio pueden inducir la aglomeración de población con características similares, lo que implica la agrupación de viviendas similares (Wittowsky et al., 2020).

Por otro lado, dado que los barrios suelen desarrollarse al mismo tiempo, las viviendas pueden tener características estructurales muy similares, como el tamaño o el diseño. Además, las viviendas pertenecientes a una misma zona comparten externalidades, tanto positivas como negativas, relacionadas con su ubicación (Basu y Thibodeau 1998; Turizo, 2024). Por lo tanto, las viviendas ubicadas cerca una de la otra frecuentemente están autocorrelacionadas, es decir, sus precios también dependen de los precios de sus vecinos (Cohen y Coughlin, 2008; Wittowsky et al., 2020).

Si los residuos hedónicos están autocorrelacionados espacialmente, la varianza de los coeficientes es subestimada, debido a que la varianza residual y la matriz de varianza-covarianzas no reflejan la varianza real del fenómeno. Por lo tanto, los errores estándar estarán subestimados, en caso de autocorrelación positiva, y las pruebas de significancia como las pruebas t-student o la prueba F tendrán valores inflados, lo que puede provocar inferencias espurias (Basu y Thibodeau, 1998; Petkov, 2020; Wittowsky et al., 2020).

Respecto a la heterogeneidad espacial, ésta se refiere a la falta de estabilidad en el comportamiento del fenómeno estudiado a lo largo del espacio. Implica que las formas funcionales y los parámetros cambian con la localización y no son homogéneos en todo el conjunto de datos. Esta condición puede ser provocada por los mismos mecanismos que generan la autocorrelación espacial, aunque no son fenómenos iguales (Anselin, 1988).

Uno de los primeros en señalar las implicaciones de la heterogeneidad fue Straszheim (1975), quien identificó que los residentes no tienen preferencias homogéneas y que, además, la oferta de vivienda es difícil de ajustar en el corto plazo, es decir, relativamente inelástica. Cuando una demanda específica e inelástica se enfrenta a una oferta también rígida, el mercado tiende a segmentarse en submercados diferenciados. Por tal razón, el espacio no debe tratarse como un continuo homogéneo, sino como un conjunto de áreas discretas e irregulares, con condiciones propias y relaciones particulares entre variables (Basu y Thibodeau, 1998; Redfearn, 2009; Yuan et al., 2020).

La heterogeneidad puede provocar que los errores no sean homoscedásticos debido a que pueden existir diferentes varianzas en cada aglomeración. La presencia de heteroscedasticidad afecta la eficiencia de los estimadores, ya que los coeficientes obtenidos por MCO dejan de ser los mejores estimadores, y existen alternativas más adecuadas que corrigen la estructura de varianza no

constante. Además, al igual que en el caso de la autocorrelación de los errores, la heteroscedasticidad distorsiona la estimación de la varianza de los errores, lo que genera errores estándar sesgados y pruebas de hipótesis poco confiables (Redfearn, 2009).

Finalmente, la existencia de los efectos espaciales provoca relaciones complejas entre los atributos y el precio. Por ejemplo, las externalidades negativas causadas por la presencia de actividad comercial en estaciones de transporte público, como ruido, contaminación o crimen, pueden provocar que el valor sea menor cerca de la estación, pero después de cierto umbral de distancia el sentido de la relación se invierte debido a la accesibilidad que sigue representado su cercanía (Soltani et al., 2024).

La teoría hedónica anticipa la existencia de relaciones no lineales entre los atributos y el precio de la vivienda. En el modelo de Rosen la curva de demanda se deriva de la disposición marginal a pagar por ese atributo, la cual refleja su utilidad marginal. En la racionalidad económica, la ley de rendimientos decrecientes establece que la utilidad marginal de un atributo puede ser menor a medida que aumenta su cantidad. Esto genera que la disposición a pagar marginal por un atributo no sea constante, creando una relación no-lineal entre el atributo y el precio. Por ejemplo, una vivienda puede aumentar su valor al incrementar su tamaño, sin embargo, después de cierto umbral de tamaño, el efecto de un metro adicional será cada vez menor debido a la ley de rendimientos decrecientes (Francke y Van De Minne, 2024; Potrawa y Tetereva, 2022).

La aplicación de modelos lineales bajo relaciones no lineales subyacentes implica problemas de especificación errónea de la forma funcional, provocando estimaciones sesgadas e interpretaciones erróneas. En tales casos, los coeficientes no capturan los efectos marginales reales de las variables y las predicciones pierden precisión en rangos donde la no linealidad es pronunciada. Además, los residuos exhiben patrones sistemáticos violando los supuestos de independencia, homoscedasticidad y normalidad requerida para inferencias válidas (McCluskey y Borst, 2007; Potrawa y Tetereva, 2022).

Por estas razones existe una necesidad de incorporar los efectos espaciales y relaciones no lineales en los modelos para poder realizar cualquier análisis inferencial fiable (Greenstone, 2017; Kauko y D'Amato, 2008; Redfearn, 2009). En los últimos años, muchas investigaciones se han enfocado en desarrollar técnicas que permitan superar las limitaciones del método de MCO para el estudio

de precios de la vivienda en presencia de efectos espaciales, mejorando su capacidad predictiva y explicativa (Turizo, 2024).

Modelos espaciales globales

Los campos de la econometría espacial y la geoestadística han proporcionado algunas de las herramientas más utilizadas para modelar efectos espaciales. Entre las más destacadas se encuentran los modelos de autorregresión espacial (SAR) y los modelos de error espacial (SEM), en el caso de la dependencia espacial; y modelos de regresión local y geográficamente ponderada (LWR, GWR), en el caso de la heterogeneidad espacial (Anselin, 1988; Potrawa y Tetereva, 2022; Taylor, 2008).

El modelo SAR, también conocido como modelo de rezago espacial, incorpora explícitamente la dependencia espacial en la variable dependiente mediante un término de rezago espacial. Este término, ρWy , representa el efecto del precio promedio de las viviendas adyacentes sobre el precio de la vivienda analizada. Su coeficiente, ρ , tiene un valor estimado y un nivel de significación, como cualquier otra variable explicativa (Anselin, 1988; Wittowsky et al., 2020). Por su parte, el modelo SEM introduce la dependencia espacial en los errores, lo que permite resolver los problemas derivados de la autocorrelación residual (Liu et al., 2024). La especificación de ambos modelos es la siguiente:

$$SAR: y = \rho Wy + X\beta + \varepsilon$$

$$SEM: y = X\beta + u$$

$$u = \lambda Wu + \varepsilon$$

Donde y es el vector de la variable dependiente, X es la matriz ($n \times k$) con k variables explicativas, β es el vector de parámetros, W es la matriz de pesos espaciales, Wy es el vector y con rezago espacial, u es el vector de los residuos, Wu el vector de residuos con rezago espacial, λ el parámetro asociado a los residuos rezagados y ε es vector residuos distribuidos normalmente.

En estos modelos, para poder manejar la autocorrelación espacial se necesita especificar la forma y el grado de vecindad entre los datos. Para ello, se utiliza una matriz de pesos espaciales, W , que ayuda a establecer relaciones multidireccionales, es decir, la interdependencia entre regiones u observaciones (Anselin, 1988; Turizo, 2024; Taylor, 2008).

La matriz de pesos espaciales es similar al operador de rezago en las series de tiempo, solo que en este caso es multidimensional. Esta define el sentido en el cual las viviendas se consideran vecinas y determina la importancia de las observaciones adyacentes en la conformación del precio. En los estudios de vivienda, las matrices de vecindad y distancia son las especificaciones más comunes de matriz de pesos (Taylor, 2008). Aun así, la especificación de la matriz de pesos continúa siendo uno de los aspectos más controversiales de la econometría espacial. Existe evidencia de que una suposición inapropiada de un error espacial de primer orden conduce a una inferencia sustancialmente defectuosa (Anselin, 1988; McCluskey y Borst, 2007; Taylor, 2008). Además, al necesitar una considerable capacidad de cómputo, un número muy grande de observaciones puede volverse computacionalmente muy tardado o incluso inviable.

Cabe decir que en el modelo de rezago espacial (SAR), los coeficientes estimados β no pueden interpretarse directamente como efectos marginales, debido a la retroalimentación entre unidades vecinas inducida por el término espacial $\rho W y$. El efecto de un cambio en una variable independiente sobre la variable dependiente no solo afecta a la observación en cuestión, sino que también se propaga hacia las observaciones vecinas y retorna parcialmente a la unidad original a través de la red espacial definida por la matriz de pesos W (Anselin, 2002; LeSage y Pace, 2009).

Por tanto, los coeficientes β representan únicamente los parámetros estructurales del modelo, pero no los impactos marginales reales. Para dar cuenta del efecto real es necesario calcular los efectos directos, indirectos y totales. Los efectos directos se obtienen como el promedio de los impactos que un cambio en la variable independiente x_i tiene sobre la variable dependiente y_i en la misma unidad, incluyendo los efectos de retroalimentación. Es decir, el efecto directo refleja no solo el efecto propio, sino también el efecto que el incremento en la variable dependiente de una unidad genera en sus vecinas y que, a su vez, retorna a la unidad original (LeSage y Pace, 2009).

Por su parte, los efectos indirectos reflejan el impacto que se transmite espacialmente entre observaciones debido a la estructura de dependencia definida por la matriz de pesos espaciales. Captura los efectos de externalidades, es decir, como un cambio en x_i afecta a y_j , (con $j \neq i$) a través de las interconexiones espaciales (LeSage y Pace, 2009).

Los efectos totales son la suma de los efectos directo e indirecto y reflejan el impacto completo en todo el sistema ante un cambio en la variable independiente, teniendo en cuenta las dependencias espaciales. En el modelo SAR, estos impactos se calculan a partir de la siguiente expresión:

$$\partial y / \partial X_r = (I - \rho W)^{-1} \beta_r$$

Esta matriz indica que el efecto total de un cambio marginal en x_r sobre y es igual al producto del coeficiente β_r por el multiplicador espacial $(I - \rho W)^{-1}$. A partir de esta matriz se pueden obtener el efecto directo y el efecto indirecto. Esta descomposición ofrece una interpretación más adecuada del impacto espacial en fenómenos como la formación de precios de la vivienda, donde las condiciones del entorno tienen un papel relevante (LeSage y Pace, 2009).

Modelos semi-paramétricos

Los modelos de rezago espacial (SAR) se consideran enfoques globales porque no toman en cuenta las variaciones locales de los parámetros. Para poder manejar la heterogeneidad espacial, un enfoque comúnmente usado es la regresión localmente ponderada (LWR), desarrollada por Cleveland y Devlin, (1988). Esta involucra la estimación de un vector de precios implícitos de cada observación, usando un subconjunto de observaciones que están cerca. Cercanía se refiere a proximidad espacial o temporal. El subconjunto de observaciones usado en cada regresión local se define por aquellas observaciones dentro de una ventana alrededor de la observación. La ventana puede ser definida como un número fijo de observaciones cercanas, un conjunto a una distancia fija o alguna otra función. (Redfearn, 2009).

En 1998, Fotheringham et al., proponen la regresión geográficamente ponderada (GWR), que es una aplicación específica del modelo LWR. Este es un modelo que incluye ubicaciones geográficas en los parámetros del modelo y considera los diferentes efectos de la misma variable en distintas ubicaciones (Fotheringham et al., 2015; Yang et al., 2018).

La especificación de este modelo es de la siguiente forma:

$$Y_i = \sum_{j=0}^K \beta_j(x_i, y_i) X_{ji} + \varepsilon_i$$

Donde x_i y y_i son las coordenadas espaciales de la observación i , y $\beta_j(x_i, y_i)$ representan el valor del coeficiente estimado para la variable j en ese punto específico. X_{ji} denota el valor de la j -ésima variable explicativa para la observación i , mientras que ε_i es el término de error aleatorio asociado a dicha observación.

En comparación con un único modelo global o incluso con unos cuantos modelos sobre submercados, el GWR puede ser intensivo desde el punto de vista computacional. Sin embargo, independientemente de los requisitos técnicos, GWR es una herramienta analítica muy potente y tiene la capacidad de revelar patrones que varían espacialmente dentro de un modelo hedónico (McCluskey y Borst, 2007).

Otro método frecuentemente utilizado para controlar la heterogeneidad espacial es dividir el conjunto de datos en subconjuntos que representan submercados. Para corroborar su existencia, primero se establecen submercados a priori en los cuales se aplican modelos MCO. Luego, se comparan los modelos en busca de diferencias estructurales. Las diferencias se detectan mediante la prueba Chow y de encontrarse, se confirma la existencia de submercados.

También es frecuente el uso de técnicas de análisis de conglomerados, como K medias o DBSCAN que, usando información sobre las características de la vivienda, su precio, así como características de la población, y considerando la contigüidad espacial, buscan aquellas observaciones más parecidas al interior de su submercado y más diferentes entre submercados. Es una forma objetiva de establecer submercados; sin embargo, tienen la desventaja de suponer el número de submercados a priori, además, según la semilla utilizada para generar los submercados, las aglomeraciones pueden ser distintas de una aplicación a otra, incluso usando las mismas variables (Basu y Thibodeau, 1998; Redfearn, 2009; Sobrino, 2014).

Finalmente, para modelar efectos no lineales e interacciones entre variables se han empleado distintos métodos y aproximaciones paramétricas y semi-paramétricas. El más usado ha sido la transformación logarítmica de la variable dependiente que linealiza las relaciones entre los atributos y el precio. Sin embargo, esta transformación puede ocultar relaciones complejas existentes entre los datos. Por su parte los modelos aditivos (GAM) con regresión spline permiten capturar no linealidades al dividir variables en segmentos, mientras que los modelos lineales generalizados (GLM), como la regresión por cuantiles, estima efectos diferenciales en distintos puntos de la distribución de la variable dependiente. Sin embargo, estos métodos enfrentan problemas significativos. En primer lugar, la interpretación se vuelve compleja debido a los coeficientes no lineales. En segundo, sigue existiendo una dependencia crítica a los supuestos de MCO. Estos límites explican el auge de métodos modernos como la combinación de modelos locales GWR y modelos aditivos (MGWR), modelos multinivel o modelos de aprendizaje

automático que automatizan la detección de patrones no lineales y espaciales. (Lorenz et al., 2023; Potrawa y Tetereva, 2022).

Evidencia empírica

Este amplio abanico de herramientas ha sido usado para estimar la primera etapa de la metodología hedónica, es decir, la estimación de los precios marginales implícitos de los atributos que componen la vivienda. Sus resultados muestran la importancia de incluir en los modelos los efectos espaciales y las relaciones complejas y no linealidades, con el objetivo de obtener estimadores confiables e información más detallada sobre los fenómenos subyacentes. En la práctica, los estudios hedónicos se han concentrado en la estimación de los efectos de tres categorías principales de atributos: los atributos estructurales de la vivienda, las características del vecindario y las características de su localización relativa. Debido a que el objetivo de este trabajo es analizar los efectos de los elementos urbanos, la evidencia empírica de esta sección describirá con mayor detalle los resultados de tales atributos.

Respecto a las variables estructurales, en general los estudios han encontrado una relación positiva entre el tamaño y la calidad de la vivienda con su precio. Potrawa y Tetereva (2022) en Róterdam, encontraron una relación no lineal entre el tamaño y el precio de propiedades en renta; descubrieron que el precio marginal del tamaño de la vivienda es lineal por partes. Cada metro adicional, de 1 a 136 m, tiene un costo estimado de 7.26 euros, de 137 a 191 m 1.65 euros, y más de 191 m 5.7 euros.

Por otro lado, la antigüedad, que es usualmente empleada como variable aproximada de la calidad, presenta efectos mixtos. Los estudios generalmente han encontrado efectos negativos, pero Lorenz et al., (2023), en la ciudad de Frankfurt encontraron una relación no lineal en forma de “U” que refleja una alta valorización en viviendas nuevas y antiguas. En este caso, el fenómeno es explicado por el contexto histórico de la ciudad. Por su parte, Caridad et al., (2008) en la ciudad de Porto encontraron una relación negativa entre la edad de la propiedad y su valor para los primeros 16 a 20 años. Posteriormente, la relación se invierte. Estos hallazgos son consistentes con Sabella (1974) quien propone que el valor de una propiedad se eleva en los últimos años debido al aumento en el valor del suelo (Durán et al., 2012).

En lo tocante a sistemas de transporte masivo como trenes, metro o sistemas BRT, los resultados han sido mixtos. Estudios como los de Khan (2007) y Grass (1992) destacan efectos positivos de la cercanía a estas infraestructuras, como un precio mayor de las viviendas o fenómenos asociados a esta, como gentrificación, mientras que otros autores como Kilpatrick et al (2007) señalan efectos negativos resultado de la presencia de ruido o contaminación. Salon et al (2014) en Guangzhou, China, no encontraron relación negativa en la cercanía a estaciones de Metro, pero sí para la cercanía con los sistemas BRT asociada con inseguridad, ruido y contaminación.

Por su parte, Diao et al (2016) en Singapur, identificaron la existencia de externalidades negativas asociadas con el ruido del Metro para viviendas a menos de 400 m y más allá de este umbral un incremento de precios del 13.7%. En comparación, Strand y Vagnes (2001) observaron que los precios de las propiedades en Noruega incrementaban cerca de 10% a una distancia de 100 m de las vías del tren. Respecto al Tren ligero, Seo et al (2014) en Arizona encontraron una relación positiva para viviendas cercanas a las estaciones, con una valorización máxima a los 300 m, mientras que Cervero y Duncan (2002), encontraron un aumento de 23% relacionado con la cercanía a este medio. Gadziński y Radzimski (2016) Yang et al (2020) y Wardrip (2011) encontraron efectos heterogéneos, una relación positiva en suburbios y negativa en áreas centrales (Potrawa y Tetereva, 2022; Seo et al., 2019; Soltani et al., 2024).

En cuanto a otro tipo de infraestructura, como autopistas o aeropuertos, los resultados muestran en general efectos negativos. Por ejemplo, Liu et al (2024) en China y Seo et al (2019) en Seúl encontraron que la proximidad a autopistas está vinculada con una disminución de precios por contaminación y congestión. Soltani et al (2024) en su trabajo aplicado a Australia, destacan beneficios económicos a cierta distancia de aeropuertos y efectos negativos a menor distancia debido al ruido, y restricciones de zonificación.

Respecto a los equipamientos urbanos y ambientales, los análisis se han concentrado en equipamientos educativos, espacios verdes y parques, así como lugares de consumo y recreación. Respecto a los primeros, se ha encontrado una relación positiva con la calidad educativa. Feng et al (2018) encontraron un aumento de 17.1% relacionado con la cercanía a escuelas clave en Shanghái, por su parte Fack et al (2010) encontraron en París una valorización de entre 1.45% y 2.4% resultado de la cercanía a escuelas de calidad. Zheng et al (2016) observaron primas del 6.8% en Pekín por la proximidad a escuelas primaria clave.

Sobre los espacios verdes y parques se han encontrado efectos mixtos. Por un lado, Jim et al (2009) identifican un aumento del 16.88% en Hong Kong debido a la combinación de accesibilidad y la existencia de vistas a los espacios verdes. Wen et al (2017) en Hangzhou encuentran que la distancia al Gran Canal se relaciona con una disminución de precios de 0.016% por cada 1% de alejamiento. Debido a estos efectos mixtos, Xu et al (2025) cuestionan el impacto universal de los espacios verdes, sugiriendo una interacción con factores socioeconómicos y el tipo y calidad de equipamiento. Respecto a los lugares de consumo y entretenimiento, también se han encontrado efectos no lineales. Por ejemplo, Chiang et al (2023) encuentran que la accesibilidad a tiendas comestibles varía según segmentos de precios. Respecto al crimen y la contaminación, Ramírez de la Cruz et al (2020) vinculan altos niveles de delincuencia con disminución de precios en la ciudad de México.

En cuanto a la localización relativa de la vivienda en la ciudad, se han encontrado resultados mixtos. En ciudades con tendencias monocéntricas, en general, se ha encontrado un efecto negativo al aumentar la distancia, tal como lo predicen los modelos clásicos de localización. Por ejemplo, Jang y Kang (2015) en Seúl, encontraron que la proximidad al DCN aumenta los precios de las viviendas. Sin embargo, la presencia de otros centros puede invertir esta relación (Basu & Thibodeau, 1998).

Finalmente, se ha encontrado que los efectos de estos atributos pueden ser diferentes según las características de la población, como su nivel socioeconómico, su estructura familiar, rangos de edad o nivel educativo; así como de otras variables. Por ejemplo, Quigley (1985) y Harrison y Rubinfeld (1978) encontraron que en zonas de bajos ingresos la accesibilidad es altamente valorizada, mientras que en áreas ricas se valoran más los atributos estructurales. Por su parte, Chen et al (2023) encontraron que compradores de bajos ingresos priorizan el transporte público y los atributos del entorno, mientras que grupos de altos ingreso valoran más el estado de la vivienda.

Además, Yuan et al (2020) encontraron que los efectos de sistemas como el BRT son mayores en los suburbios, mientras que, Kilpatrick et al (2007), mostraron que en áreas centrales la proximidad a estaciones de metro puede generar externalidades negativas. Por su parte, Potrawa y Tetereva (2022) encontraron que el efecto de cada atributo no solamente depende de su cantidad y calidad, sino de la cantidad de otros atributos. Por ejemplo, el valor de una vista panorámica en Róterdam tiene un valor de más de 100 euros, pero solo en propiedades amplias y bien localizadas.

Esta evidencia empírica muestra que los efectos de los atributos en el precio de las viviendas son complejos, no lineales y están influenciados por múltiples factores interconectados. Las variables de estructura, como el tamaño y la edad, muestran efectos no lineales y efectos de umbral. La infraestructura de transporte y los equipamientos urbanos exhiben efectos mixtos, algunos positivos por la accesibilidad y oportunidades económicas que brindan, pero otros negativos por externalidades ambientales como contaminación, crimen y ruido. Además, los efectos de los atributos varían según la localización espacial, las características socioeconómicas y la cantidad y calidad de otros atributos.

Conclusiones

Esta revisión bibliográfica brinda un panorama general de los marcos analíticos históricamente utilizados para describir el proceso de conformación del precio de las viviendas. Primero se describe los modelos clásicos de localización, sus aportaciones empíricas y sus límites. Después se introduce la metodología hedónica, que surge como una alternativa para dar cuenta de la heterogeneidad intrínseca de las viviendas. Se detalla su método y etapas de aplicación, además, se describen las consideraciones metodológicas concernientes a la estimación de la primera fase o Función de Precios Hedónicos, relacionadas con la naturaleza espacial de la información del mercado urbano de viviendas. Finalmente, se describen las herramientas metodológicas usadas para superar los límites y se muestran sus resultados de su aplicación.

La evidencia empírica mostrada da cuenta de la existencia de efectos espaciales, relaciones complejas y sus implicaciones en los análisis de la capitalización de los elementos del entorno urbano. Sugiere la necesidad de incluir estos efectos en los modelos para evitar resultados espurios y poco fiables. Muestra que a pesar de la claridad que brindan los modelos autorregresivos, aún contienen ciertas limitaciones, la principal de ellas es la imposibilidad de dar cuenta de relaciones no lineales o complejas. Por su parte, en los modelos locales como GWR, también existen limitaciones relacionadas con el costo computacional elevado que representa el uso de muestras grandes.

Recientemente, algunos investigadores han considerado utilizar la habilidad natural de los modelos de aprendizaje automático de capturar no linealidades y entender con mayor detalle las relaciones existentes. En la última década, se han desarrollado avances técnicos que habilitan la interpretación de las relaciones aprendidas por los algoritmos, brindando la posibilidad de una descripción más

detallada de las relaciones que existen entre las características de la vivienda y su precio. En la siguiente sección se describirá de forma breve el funcionamiento de los métodos de aprendizaje automático, su aplicación tradicional en el mercado de vivienda y los desarrollos recientes sobre interpretabilidad, concluyendo con ejemplos de aplicaciones en el marco de la metodología hedónica.

El aprendizaje automático en la estimación de la Función de Precios Hedónicos.

El avance tecnológico de las últimas tres décadas ha dado como resultado el surgimiento de nuevas fuentes de datos y herramientas de gran utilidad para analizar las dinámicas del mercado inmobiliario. Por un lado, la gran cantidad de datos generados por los usuarios de internet y publicados en sitios web representa una nueva fuente de información que puede reflejar dinámicas sociales muy importantes. Por ejemplo, los portales web de anuncios clasificados ponen a disposición pública información actualizada sobre la oferta del mercado inmobiliario. Esta información contiene datos sobre los precios, las características estructurales de las viviendas y su localización geográfica, lo que enriquece de forma sustancial su utilidad, ya que permiten asociar la ubicación precisa de una vivienda con la disponibilidad y accesibilidad a diferentes elementos de la ciudad (Conway, 2018).

Por otro lado, la creciente cantidad de información también ha traído consigo el desarrollo y despliegue de diferentes herramientas metodológicas capaces de procesar y analizar grandes cantidades de datos. Específicamente, los algoritmos de aprendizaje automático representan una nueva metodología con gran desempeño en tareas predictivas. Su aplicación en el análisis inmobiliario no es reciente, sin embargo, se ha limitado simplemente a estimar los valores inmobiliarios, sin analizar las relaciones subyacentes. Estos algoritmos tradicionalmente han sido denominados “cajas negras” debido a la complejidad para interpretar sus resultados y por esta razón no han sido adoptados en la metodología hedónica. Avances metodológicos recientes habilitan la interpretación de las relaciones aprendidas por estos algoritmos, lo que incrementa su utilidad en los análisis de las relaciones subyacentes (Rico-Juan y Taltavull De La Paz, 2021; Lorenz et al., 2023; Potrawa y Tetereva, 2022).

La aparición de regulaciones motivadas por la necesidad de transparencia en los modelos de aprendizaje automático, sobre todo en un contexto de creciente adopción en la toma de decisiones de políticas públicas, ha impulsado el desarrollo de herramientas interpretativas que permiten mayor conocimiento y transparencia de los resultados. Estas nuevas técnicas abren la posibilidad de revelar los efectos que las diferentes variables explicativas tienen en la variable de respuesta, beneficiándose de la capacidad de los algoritmos de aprender relaciones complejas y no lineales.

Su adopción podría potenciar la metodología hedónica, específicamente la estimación de la primera fase o Función de Precios Hedónicos (MacBride, 2024; Du et al., 2025). En este apartado se describe de forma breve su funcionamiento y algunos ejemplos de aplicación con el objetivo de mostrar los potenciales beneficios de su adopción en los análisis inmobiliarios.

Funcionamiento

En comparación con los métodos econométricos, que dependen de supuestos clave sobre el proceso de generación de datos para generar inferencias confiables, los algoritmos de aprendizaje automático se consideran métodos no paramétricos en el sentido de no depender de ningún supuesto (Francke y Van De Minne, 2024). Se dividen en dos tipos: aprendizaje supervisado y aprendizaje no supervisado. En el aprendizaje no supervisado, los algoritmos aprenden relaciones entre los datos sin conocer los resultados con anticipación. Algunos ejemplos de estos son las herramientas de análisis de conglomerados, como k medias, o el análisis de componentes principales, que identifican observaciones similares entre sí, sin conocimiento previo de las relaciones o patrones (Conway, 2018).

Por su parte, en el aprendizaje supervisado los resultados se conocen con anterioridad y tienen una función fundamental en la operación de los algoritmos (Mullainathan y Spiess, 2017). La información es utilizada para comparar las predicciones del modelo y ajustarse automáticamente hasta que la diferencia entre la predicción y los resultados observados sea la mínima. Este tipo de funcionamiento tiene la ventaja de obtener una gran precisión predictiva, por lo que los algoritmos son muy útiles en tareas de regresión o de clasificación (Conway, 2018).

Los algoritmos de aprendizaje supervisado ajustan diferentes relaciones funcionales, las cuales se optimizan mediante un proceso iterativo en función de la precisión de la predicción y la reducción en la tasa de errores. Utilizan regularización para limitar su complejidad y evitar el riesgo de sobreajuste. Para elegir el nivel óptimo de regularización, se utiliza ajuste empírico, es decir la creación de experimentos fuera de la muestra del algoritmo, comúnmente dividiendo los datos en 80% datos de entrenamiento y 20% datos de prueba. Los datos ajustan las relaciones funcionales en el conjunto de entrenamiento hasta encontrar una función generalizable que posteriormente se emplea en el conjunto de datos de prueba para determinar su precisión. (Mullainathan y Spiess, 2017; Yazdani, 2021).

La gran mayoría de algoritmos de aprendizaje automático funciona siguiendo estos principios metodológicos, desde las redes neuronales artificiales hasta los modelos de árboles de decisión o bosques aleatorios, aunque con diferencias en la estructura de aprendizaje de cada modelo. Los beneficios de esta metodología se reportan en su gran precisión predictiva, superando significativamente a los modelos econométricos tradicionales. Sin embargo, debido a que la estimación de la FPH no representa una tarea predictiva, su adopción en la metodología hedónica ha sido limitada. En estudios sobre el mercado inmobiliario estos algoritmos son utilizados, sobre todo, para estimar el valor del parque inmobiliario, por lo que su adopción ha sido en gran medida por parte de agencias hipotecarias, bancos y haciendas locales con fines catastrales (Rico-Juan y Taltavull De La Paz, 2021; Yazdani, 2021).

Modelos Agnósticos de Interpretabilidad

En los últimos años, han surgido leyes y reglamentos que exigen que los algoritmos de aprendizaje automático sean justos, transparentes y explicables. Algunos ejemplos son la Regulación General de Protección de Datos de la Unión Europea, publicada en 2016 o el programa de Inteligencia Artificial Explicable (XIA), de la Agencia de Proyectos de Investigación Avanzada de Defensa de Estados Unidos (DARPA), publicado en 2017, que fueron elaborados con el objetivo de avanzar en las técnicas de aprendizaje automático para producir modelos interpretables sin sacrificar el rendimiento y permitir que los usuarios comprendan, confíen y gestionen adecuadamente la generación de sistemas de inteligencia artificial (Li, 2022).

La emergencia de los modelos interpretables abre la posibilidad de conseguir indicios más completos y matizados sobre las relaciones existentes en el mercado de vivienda, los cuales no pueden ser capturados por enfoques paramétricos tradicionales. Con su avance, los algoritmos de aprendizaje automático dejan de ser cajas negras y se vuelve posible aprovechar su flexibilidad y habilidad predictiva para realizar análisis detallados sobre las relaciones entre variables, por ejemplo, en la estimación de la primera fase hedónica y beneficiarse de su capacidad de aprender relaciones complejas entre los atributos y el precio (Rico-Juan & Taltavull De La Paz, 2021).

Existen dos formas de interpretar los resultados obtenidos con los modelos de aprendizaje automático. La primera es definida como interpretabilidad intrínseca y se refieren a modelos de aprendizaje automático tradicionales reducidos en complejidad, lo suficiente como para que las relaciones funcionales puedan ser interpretadas directamente. Ejemplos de estos son los modelos

basados en árboles de decisión con una profundidad limitada a dos o tres particiones o redes neuronales con una o dos capas (Breiman, 2001). Sin embargo, la reducción en la complejidad significa renunciar a la precisión predictiva, la cual es su mayor ventaja.

La segunda forma es mediante el uso de marcos explicativos separados de los modelos de aprendizaje automático, también denominados modelos agnósticos de interpretabilidad. Estos métodos permiten preservar la capacidad predictiva de los modelos, y al mismo tiempo extraer información interpretable de las relaciones aprendidas. Son muy flexibles ya que no dependen de un algoritmo de aprendizaje automático específico, sino que pueden ser aplicados a cualquiera, independientemente de su mecanismo de funcionamiento (Du et al., 2025; Molnar, 2020; Rico-Juan y Taltavull De La Paz, 2021). Estos modelos permiten construir explicaciones sobre las relaciones aprendidas por los algoritmos de aprendizaje automático, utilizando los valores de las predicciones realizadas por estos.

Existen tres tipos de datos que se pueden extraer con los modelos agnósticos de interpretabilidad. En primer lugar, la importancia de las variables (FI), la cual permite identificar la contribución de cada una de las variables en la predicción del modelo. Mediante la comparación de la magnitud del efecto en la predicción, ante cambios en los valores de las variables de interés, permite identificar las características más importantes en la conformación del valor generado por la predicción. En segundo lugar, la dependencia parcial (PD), la cual es una herramienta gráfica que muestra los efectos marginales de una característica sobre el valor predicho por el modelo. Al marginar la predicción del modelo sobre la distribución de todas las características, excepto la de interés, se puede estimar la relación marginal entre la variable y la predicción (Friedman, 2001). Una forma usada para calcular esta dependencia parcial es mediante los Efectos Locales Acumulados (ALE), los cuales tienen la ventaja, ante otros métodos, de ser robustos ante la presencia de multicolinealidad, común en la información espacial (Y. Wu y Han, 2025).

En tercer lugar, existen los modelos de interpretación local, como las explicaciones Shapley aditivas, (SHAP) o las explicaciones locales interpretables, (LIME), los cuales muestran el efecto de cada variable en una observación particular (Jin et al., 2023; Lundberg y Lee, 2017; Molnar 2020; Ribeiro et al., 2016). En el caso de los valores SHAP, para su estimación se utilizan conceptos analíticos de la teoría de juegos que permiten identificar la contribución individual de cada variable en el resultado final, por lo que es considerado un método confiable y robusto para extraer efectos

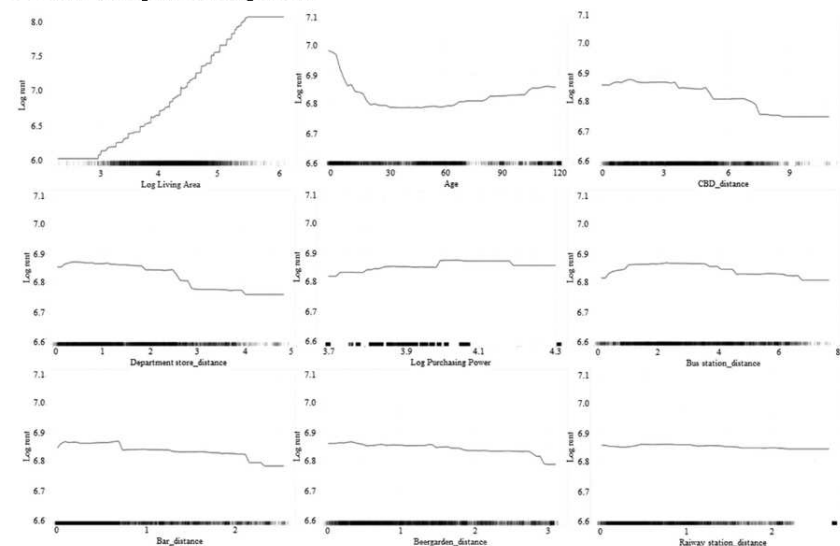
individuales. Por su parte, los valores LIME utilizan modelos locales lineales para estimar los efectos de las variables. Este último tipo tiene ciertas similitudes con los modelos GWR en el sentido de tener que definir una ventana de observaciones, aunque en este caso, son observaciones cercanas en un plano vectorial (Lundberg y Lee, 2017; Molnar, 2020).

Aplicación en la metodología hedónica

En el marco de la metodología hedónica, estos modelos representan nuevas formas de estimar la función de precios hedónicos o los efectos marginales de las características de las viviendas e incluso la posibilidad conocer sus impactos en viviendas individuales. En la última década, algunos investigadores han experimentado con su uso en la estimación de la primera fase hedónica, mostrando resultados prometedores respecto a la identificación de relaciones complejas.

Por ejemplo, Lorenz et al (2023), analizando la ciudad de Frankfurt, utilizaron gráficas de dependencia parcial calculadas con ALE (Gráfica 4), con el espíritu de explorar los efectos complejos. Encontraron diferentes relaciones, entre ellas, una relación en forma de “U” de la edad con el precio y mostraron que los

Gráfica 4. Dependencia parcial



Se muestra la dependencia parcial del tamaño en metros cuadrados, la edad, la distancia al distrito financiero (DCN), al centro comercial, a estaciones de autobús, a bares, a jardines de cerveza, a estaciones de ferrocarril y sobre el poder adquisitivo. El eje vertical indica los valores de la característica del nivel de renta en escala logarítmica, mientras que el eje horizontal representa los valores de las covariables. Las líneas negras superpuestas muestran el número de observaciones en cada valor. Fuente: Lorenz et al., 2023

métodos de interpretación pueden revelar a los investigadores conocimientos más profundos sobre el funcionamiento del mercado.

Por su parte, Zhao y Hastie (2019) en Shenzhen, mediante gráficos de dependencia parcial (Gráfica 5), midieron el impacto de la calidad ambiental en el precio mostrando efectos no lineales. De igual forma, Potrawa y Tetereva (2022), usando modelos de interpretación local LIME (Gráfica 6), analizaron la influencia de las vistas panorámicas en la renta de departamentos en Róterdam mostrando una mayor precisión predictiva, así como información más detallada sobre la relación

entre las vistas y las rentas, revelando efectos diferenciados de la valorización en función de la altura y ubicación de los departamentos.

Estos ejemplos empíricos, si bien limitados, muestran que los métodos de aprendizaje automático, más complejos, pero más precisos, pueden ser tan interpretables como los métodos de estimación hedónica tradicionales. Además, tienen la ventaja de capturar las relaciones no constantes de los atributos de la vivienda, lo que no puede capturarse con métodos convencionales MCO o incluso con las variantes espaciales.

Tampoco están sujetos a los supuestos clave del proceso de generación de información que deben cumplir los modelos econométricos tradicionales. Sin embargo, a pesar de sus ventajas, también están limitados en ciertos aspectos.

En primer lugar, son métodos no paramétricos, meramente exploratorios y descriptivos. Aunque pueden evaluar qué variables contribuyen más a la predicción y ofrecen métricas de desempeño como el R^2 o la raíz del error cuadrático medio (RMSE) para estimar la bondad de ajuste del modelo, no están

diseñados para realizar inferencia estadística en el sentido tradicional. Es decir, no proporcionan estimaciones puntuales de coeficientes acompañadas de errores estándar, intervalos de confianza o valores p que permitan evaluar la significancia estadística a de los predictores. Estos componentes están relacionados con la inferencia estadística y permiten valorar, por ejemplo, que la probabilidad

Gráfica 5. Dependencia parcial

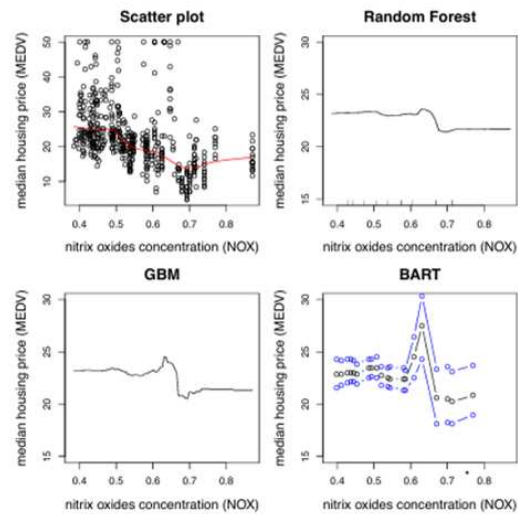
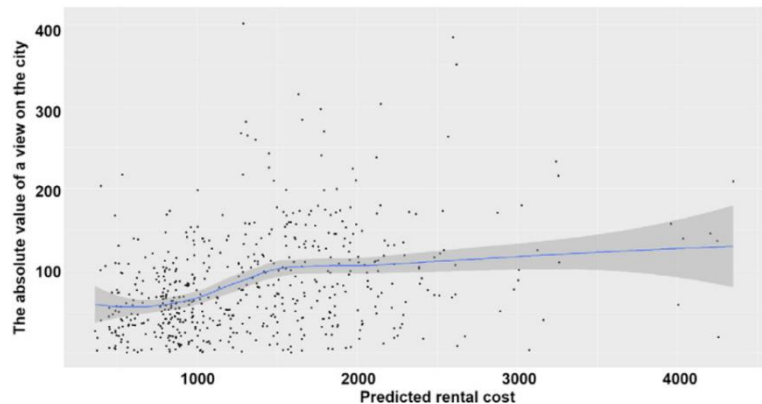


Diagrama de dispersión (superior izquierdo) y gráficos de dependencia parcial (superior derecho e inferior izquierdo) utilizando diferentes algoritmos de “caja negra”. Las curvas azules en la gráfica BART son intervalos de confianza bayesianos. Fuente: Zhao y Hastie, 2019

Gráfica 6. Dependencia parcial con LIME



Valor de la vista contra la renta predicha. Fuente: Potrawa y Teterova, 2022

real en la población sea distinta de cero. Por este motivo, estos modelos no permiten realizar inferencias causales directas y su aplicación debe considerarse principalmente exploratoria (Iskhakov et al., 2020).

En segundo lugar, el cálculo de estos modelos, principalmente los relacionados con los efectos en observaciones individuales, implican un elevado costo computacional, lo que puede resultar problemático para grandes conjuntos de datos. En tercer lugar, algunos de ellos se ven afectados por la correlación entre variables, como por ejemplo el modelo LIME o la dependencia parcial, aunque métodos de cálculo como ALE superan estos límites (Li, 2022; Molnar, 2020; Potrawa y Teterova, 2022).

Finalmente, a pesar de la flexibilidad y capacidad de aprender relaciones en los datos, los modelos de aprendizaje automático no consideran ningún efecto espacial en los datos (Kopczewska, 2022). No atender a los efectos espaciales genera problemas como ajustes demasiado optimistas de los modelos, información omitida o predicciones subóptimas (Heng et al., 2018), aspectos discutidos en párrafos anteriores. Sin embargo, para capturar las relaciones espaciales basta con introducir en el modelo variables que den cuenta de las relaciones en el espacio. Algunos ejemplos pueden ser coordenadas geográficas, distancias a puntos específicos de la ciudad, e incluso una variable espacialmente rezagada idéntica a la utilizada en los modelos de econometría espacial. De hecho, la dimensión espacial puede abordarse en cada etapa del proceso de modelización y la combinación de diferentes soluciones ha demostrado mejorar la calidad del modelo (Yoshida y Seya, 2021). Además, atender el problema de autocorrelación permite no solo reproducir los datos de entrenamientos, sino también realizar predicciones en nuevas ubicaciones fuera del conjunto de datos (Jin et al 2023).

Conclusiones

Los algoritmos de aprendizaje automático poseen una mayor capacidad predictiva, pero tradicionalmente se ha limitado su uso debido a la nula capacidad de revelar las relaciones subyacentes en los datos. El desarrollo de modelos agnósticos de interpretabilidad supera estas limitaciones y brinda nuevas formas de analizar las relaciones complejas que mantienen las características de las viviendas y su precio. La flexibilidad natural de los modelos de aprendizaje automático, combinado con los modelos de interpretabilidad, permiten aportar mayor detalle y

riqueza en los análisis de la capitalización de la inversión pública. Sin embargo, su adopción con fines inferenciales debe ser tomada con precaución.

A pesar de que los desarrollos de interpretabilidad habilitan conocer la magnitud y forma de las relaciones, no permite establecer causalidad, lo que significa que solamente pueden ser utilizados con fines exploratorios para visualizar las relaciones complejas o analizar las formas funcionales. Además, al igual que en la econometría clásica, la dependencia espacial tiene que ser especificada para ser tomada en cuenta. De no hacerlo, podría causar problemas en las predicciones de los modelos.

Estos desarrollos representan herramientas valiosas que puede enriquecer el análisis sobre la capitalización de la infraestructura pública y apoyar decisiones de política pública mejor informadas. Por tales motivos, en este trabajo se adoptaron algunas de estas herramientas con el espíritu de explorar las formas funcionales aprendidas por los modelos y compararlas con resultados de modelos econométricos tradicionales.

Contextualización de la zona de estudio

Urbanización y dotación de infraestructura

La Zona Metropolitana de la Ciudad de México es una región ubicada en el suroeste del Valle de México, en la meseta alta del centro del país. Se encuentra a una altitud de más de 2,000 metros sobre el nivel del mar, en una cuenca semicerrada. Su superficie es de 7,866 km² y en ella habitan poco más de 21 millones de personas. Está constituida por 76 municipios: 16 correspondientes a la Ciudad de México, 59 al estado de México y uno al estado de Hidalgo (INEGI, 2018). Ocupa, el cuarto puesto del ranking de ciudades más pobladas del mundo, después de Tokio, Delhi y Shanghái (Naciones Unidas, 2014). Los patrones de desigualdad y segregación que muestra su estructura espacial no es una novedad de los últimos años, sino estructuras heredadas desde inicios del siglo XX, las cuales se han ido acentuando y profundizando (Márquez-López, 2016; Sosa, 1993).

En la década de 1920 ya se mostraban rasgos de desigualdad en la ciudad. Existían colonias acomodadas, como Lomas de Chapultepec, pero también miles de viviendas construidas con materiales de desperdicio. Durante los años 1930, se dio una etapa de desarrollo de planes reguladores, cuyos objetivos fueron atender los problemas relacionados con la extensión urbana de la ciudad. Para tales objetivos se implementaron acciones urbanas sobre el núcleo central de la ciudad, justificadas en ideas de modernización, aunque al mismo tiempo, el Departamento del Distrito Federal (DDF) legitimó de diferentes maneras el proceso de urbanización irregular, por lo que dio lugar a una expansión significativa de la ciudad (Sosa, 1993).

Debido a las limitaciones financieras y la intensificación de diversos problemas, el gasto en el periodo se orientó a la construcción de infraestructura primordial como pozos de agua, tuberías de agua potable y estaciones de bombeo. También se construyeron algunos equipamientos de salud, educativos, y monumentos muy importantes, como el monumento a la Revolución, El Instituto Politécnico Nacional (IPN) en el viejo Casco de Santo Tomás y el Colegio de México, ubicado en la colonia Roma (Sosa, 1993).

Durante la Segunda Guerra Mundial se produjeron cambios importantes en la ciudad. La caída en las importaciones fue considerada una oportunidad para la industrialización del país, por lo que diversas políticas de exención de impuestos y de adecuación de zonas industriales fueron aplicadas en el periodo. Se instalaron decenas de industrias coordinando su localización con las vías de

ferrocarriles y con la Terminal del Valle de México, que sería la principal terminal de carga ferroviaria, dando como resultado la instalación de 11 zonas industriales. Además, en ciertas zonas, como Paseo de la Reforma y avenida Insurgentes, grandes edificios de oficinas surgieron y revalorizaron esas vialidades, a la vez que modificaron la estructura tradicional de la ciudad que ubicaba como distrito financiero y comercial al antiguo centro histórico (Sosa, 1993).

A finales de los años de 1940 y principios de la década siguiente, la planeación global fue prácticamente olvidada y sustituida por enormes proyectos arquitectónicos. Algunos de ellos fueron el Hospital de la Raza, en avenida Insurgentes, el Conservatorio Nacional en Polanco, el Auditorio Nacional, en Chapultepec, la cárcel de Mujeres, en Iztapalapa, y la instalación del aeropuerto internacional en el ex vaso regulador de Texcoco. Otras obras colosales tuvieron que ver con el drenaje de la ciudad, como el entubamiento de ríos y plantas de bombeo. Sin embargo, la obra más importante del periodo fue la construcción, sobre el Pedregal de San Ángel, de Ciudad Universitaria (CU), que significó no solo la modernización de las instalaciones de enseñanza superior más importantes del país, sino también la salida de la universidad de lo que entonces era la ciudad de México, lo que influyó de forma importante en la extensión de la ciudad hacia el sur (Sosa, 1993).

Respecto a la política de vivienda, la orientación del capital privado hacia el mercado inmobiliario de alta rentabilidad dejó de lado la construcción de viviendas para la clase trabajadora, y dio como resultado un marcado contraste entre, por un lado, viviendas que carecían de servicios básicos y estaban construidas de madera y cartón, cuya población vivía en condiciones precarias y de hacinamiento, y, por otro lado, colonias como Polanco, Del Valle y Pedregal de San Ángel, en donde las clases acomodadas comenzaron a residir. Este problema, junto con la alta migración provocada por la industrialización, hizo que, durante la presidencia de Miguel Alemán, la producción de vivienda se considerara de alta prioridad. El Estado, a través de instituciones como el ISSSTE, construyeron conjuntos multifamiliares del tipo de supermanzanas inspiradas por Le Corbusier en colonias como Del Valle, Santa Cruz Atoyac, Tlatelolco y Guerrero. En la práctica, sin embargo, implicaban elevados costos de mantenimiento, así como un rechazo a la ciudad, la imposición de un modelo que negaba lo existente y buscaba ser un ejemplo espectacular de lo que debería ser el espacio urbano moderno (Sosa, 1993).

En la década de 1950, el IMSS y el ISSSTE construyeron un gran número de viviendas de alquiler en zonas centrales y al sur, en Tacubaya, Tlalpan, CU y Doctores. Por su parte, el equipamiento de la ciudad se enriqueció con diversas obras. Se construyó el Museo de Historia Natural, el Museo de Arte Moderno, el Museo del Caracol, el Museo Nacional de Antropología, la Escuela Nacional de Antropología, así como la Unidad Profesional Zacatenco del IPN, al norte de la ciudad. A pesar de esta numerosa dotación de servicios, muchas colonias populares seguían teniendo graves problemas de escasez de servicios públicos, en buena medida a causa de su origen ilegal. Estas colonias se convirtieron en espacios marginados, tanto por la incapacidad de gestión de los habitantes que les permitiera introducir los servicios que necesitaban, como por el rechazo de las autoridades de la ciudad a instalar servicios, con el argumento de detener la reproducción de este tipo de colonias (Sosa, 1993).

A finales de los años de 1960 se inauguró la primera línea del metro, conectando a las estaciones Zaragoza y Chapultepec. En el decenio siguiente se inauguraron las líneas dos y tres, por lo que el Metro representó la mayor inversión estatal en la transportación masiva del país y de la ciudad (Schteingart e Ibarra, 2016). Diversas estaciones incrementaron los valores de la tierra y fomentaron la creación de centros comerciales, polos de comercio fijo y ambulante que aprovechaban las grandes afluencias de personas. Al mismo tiempo, otros sistemas de transporte masivo fueron sustituidos, como las líneas de tranvía y de trolebuses. Además, gracias a las olimpiadas de 1968 se construyeron diversos equipamientos en toda la ciudad, como el Canal de Cuemanco, la Alberca Olímpica, el Palacio de los Deportes, el Velódromo Olímpico, la Sala de Armas de la Magdalena, la Villa Olímpica y la Torre de Telecomunicaciones (Sosa, 1993).

En la década de 1970 los equipamientos educativos y de salud se desconcentraron hacia la periferia sur, oriente y norte. En este periodo se construyó el Hospital del IMAN, entre Periférico e Insurgentes; el Hospital Psiquiátrico, el Instituto Nacional de Cardiología y el Hospital de Nutrición, en Tlalpan. Tres unidades de la UAM, el nuevo edificio de El Colegio de México, en camino al Ajusco, y los reclusorios Norte y Oriente. En la década de 1980 se construyeron las líneas 8, A y B del STC Metro que conectan con el Estado de México, agregando 49 km a la red (Schteingart e Ibarra, 2016).

A finales de los años 1980 y principios de los noventa, el escenario económico global impulsó a las ciudades a implementar estrategias de planificación basadas en la transformación urbana, con

el objetivo de consolidar su relevancia internacional. Según Ziccardi (2004), la competitividad urbana dependió de la capacidad para crear condiciones óptimas que atrajeran inversiones y actividades económicas productivas. Como respuesta, las principales metrópolis promovieron iniciativas de modernización de su infraestructura y fomentaron la creación de núcleos especializados en innovación y gestión (Aguayo, 2016).

En la Ciudad de México comenzó un proceso de renovación y revitalización de espacios centrales o estratégicos para el desarrollo financiero y turístico de la ciudad. El interés por convertirla en una ciudad global promovió el reforzamiento de los nodos financiero-comerciales existentes sobre Paseo de la Reforma, Polanco, Las Lomas y el sur de la ciudad. La creación de Santa Fe, como nuevo nodo financiero, permitiría posicionar a México en la competencia por la atracción de inversión extranjera. Al mismo tiempo, se pusieron en marcha distintos proyectos de renovación urbana en lugares como el Centro Histórico, San Ángel, Coyoacán, Condesa, Nápoles, Chimalistic, Roma y Polanco. Centros financieros, centros comerciales, áreas residenciales cerradas, colonias renovadas y conjuntos urbanos de grandes dimensiones se multiplicaron desde la década de 1990 en puntos nodales de la ciudad, formando áreas con infraestructura y servicios de primer nivel que contrastan con otras áreas menos favorecidas por las políticas de revitalización y mejora urbana (Aguayo, 2016; Flores y Caracheo, 2021; Márquez-López, 2016).

Desde los 2000, frente al cúmulo de problemas viales y de transporte, y a los siempre insuficientes recursos públicos, el gobierno de la Ciudad de México y del Estado de México han realizado un conjunto de obras de infraestructura vial y de transporte que se proponen privilegiar el transporte colectivo, solucionar los nodos viales más conflictivos y modernizar de manera prioritaria algunos corredores viales (Delgadillo, 2010). De hecho, en los últimos años, en México el tema de movilidad sustentable ha destacado en la agenda federal, considerándose como parte de los objetivos de la reforma urbana en 2016 del programa Nacional de Desarrollo 2014-2018 y de la Estrategia Nacional de Movilidad Urbana Sustentable. Por tal motivo, se han destinado recursos de los fondos internacional, federal y local para la implementación de obras de movilidad sustentable, con la finalidad de transformar los hábitos de desplazamiento de los mexicanos, y con ello, desincentivar el uso del automóvil. Algunos ejemplos son los siguientes.

En 2003 se inauguró la primera línea del Metrobús, un sistema de autobuses híbridos articulados de gran capacidad que se desplaza en un carril de uso exclusivo. Hasta la fecha, 2025, existen siete

rutas que recorren las principales vialidades de la ciudad y conectan algunos centros nodales periféricos. En 2012 se inauguró la línea 12 de metro, que vincula al sur de la ciudad con la red de metro en sentido oriente-poniente, completando así una infraestructura que cuenta con 195 estaciones y poco más de 225 km de vías en servicio. Por su parte, en 2013 se destinó un importe de 266 millones de pesos para la construcción del sistema BRT Mexibús corredor 3 Nezahualcóyotl - Chimalhuacán. Además, se han desarrollado diferentes corredores de transporte público eléctrico, como el Trolebús, que ha sustituido numerosas rutas de transporte público concesionado y que actualmente cuenta con 13 rutas; el Cablebus, que actualmente cuenta con 4 líneas; y el tren suburbano que cuenta con una línea que conecta con la periferia norte de la ciudad. La infraestructura ciclista también se ha desarrollado, aunque de forma más moderada y concentrada en colonias como la Condesa, Roma, y Polanco (Delgadillo, 2010; Naranjo, 2013).

Distribución de la población y empleo

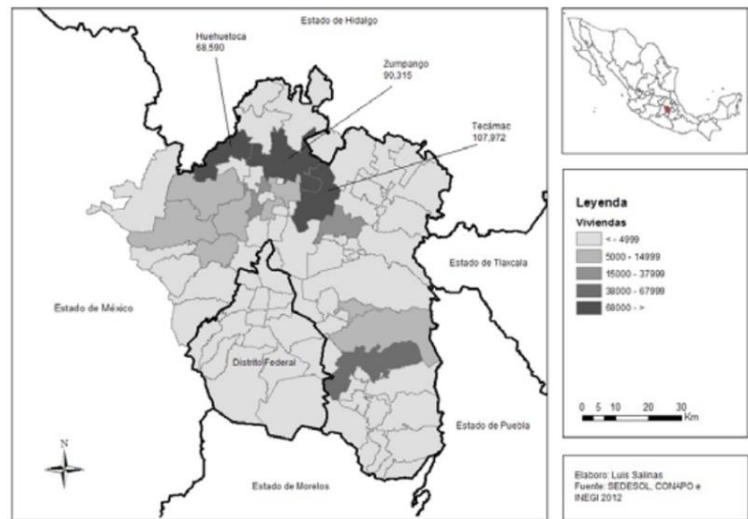
En la actualidad existen concentraciones de población muy altas al oriente, en Iztapalapa, Nezahualcóyotl y Ecatepec, al occidente Ciudad Satélite y Naucalpan y al centro Tlalnepantla y Cuautitlán Izcalli. Estos espacios se consolidaron como centros de población de alta densidad entre los años de 1970 y 1980 (Aguayo, 2016). Además, tras el sismo de 1985 que dañó severamente la zona centro de la ciudad, especialmente el Centro Histórico y áreas aledañas, la zona sur y poniente de la ciudad, asentadas en suelos más firmes, recibieron una fuerte presión inmobiliaria. Mientras que la parte central de la aglomeración perdió al rededor 1.2 millones de personas entre 1970 y 2000, los municipios periféricos experimentaron tasas de crecimiento muy altas (Schteingart e Ibarra, 2016).

Durante el primer decenio del siglo XXI, la política de desarrollo de vivienda social en México promovió la construcción masiva de nuevos conjuntos habitacionales en municipios de las periferias urbanas y favoreció una baja mezcla de usos de suelo (Montejano et al, 2018). Esto provocó el desarrollo de conjuntos habitacionales distantes al área urbana, que al no formar parte integral de la estructura vial rompen con la continuidad espacial de las ciudades y dificulta la accesibilidad a otros satisfactores, como los centros de empleo, el equipamiento educativo, salud, cultura, recreación etc.

Entre 2000 y 2015, en los municipios del Estado de México e Hidalgo, se autorizaron la construcción de 687,263 viviendas, siendo Huehuetoca, Tecámac y Zumpango tres de los

municipios metropolitanos con el mayor número de viviendas autorizadas para su construcción, las cuales ascienden a 278,474, es decir, en esos tres municipios se concentran el 40.5% de la oferta total de viviendas de los municipios que integran la ZMCM. En Huehuetoca se incrementó la población de 38,458 a 128,486 de 2000 a 2015, un aumento que

Mapa 1. Construcción de vivienda 2004-2014



Fuente: Salinas Arreortua, 2016

representa el 234%. Tecámac pasó de 172,813 habitantes a 446,008, un incremento del 158%. Y Zumpango pasó de 99,774 a 199,069 habitantes, representando un 100% de incremento (Montejano et al, 2018). Por su parte, Salinas Arreortua (2016) encontró que entre 2004 y 2015 el INFONAVIT otorgó 88,086 créditos para la adquisición de vivienda en la Ciudad de México, mientras que en los municipios conurbados se otorgaron 349,509. Los municipios que más concentraron créditos fueron Huehuetoca con 40,905, Tecámac con 49,943 y Zumpango con 66,317 (Mapa 1).

Este desarrollo masivo de viviendas en grandes conjuntos habitacionales tiene diversos problemas. Se ubican en lejanas periferias, donde el suelo es barato y obligan a sus residentes a realizar grandes traslados para acceder al trabajo, los servicios y equipamientos básicos esenciales. También tienen impactos negativos en los territorios donde se ubican y en la infraestructura de los asentamientos vecinos. No se diseñan en el marco de un plan municipal ni se planifican en el ámbito urbano o regional. Por tanto, constituyen una forma de exclusión social, ya que contribuyen a la formación de *guetos* propensos a la generación de conflictos y resentimientos sociales por estar alejados de las ciudades y de carecer de servicios y equipamientos básicos (Aguayo, 2016).

Por su parte, el empleo también se ha descentralizado, principalmente a causa del cambio de modelo económico de la década de 1990, incluso con más intensidad que la población, aunque no ha alcanzado el mismo grado de descentralización. Muñiz et al (2015) analizaron la configuración espacial del empleo con información del Censo Económico de 2009 y el Directorio Estadístico Nacional de Unidades Económicas, e identificaron siete subcentros de empleo y un DCN. Estos se

localizan en Cuautitlán, Tlalnepantla, Santa Fe, Tlalpan, C. Abastos, Pantitlán, Aeropuerto, y Ecatepec. Una característica en común es que cinco de ellos son municipalidades que históricamente han sido polos concentradores de la actividad económica y de población, (Mapa 2). Sus resultados reflejan una concentración del empleo mayor que la de población, mostrando una fuerte tendencia monocentrica en la ciudad, además de una creciente desconexión entre las áreas residenciales y los centros de empleo.

Mapa 2. Concentración del empleo



Fuente: Muñiz et al., 2015

Conclusiones

Esta breve descripción histórica sobre la urbanización y dotación de infraestructura revela que la configuración espacial de la ciudad no es un producto de fenómenos recientes, sino de estructuras históricas heredadas que se han intensificado bajo un patrón de polarización urbana. Los sectores más desfavorecidos se han concentrado en áreas con oportunidades limitadas de acceso a servicios esenciales como educación, salud, transporte y vivienda, reproduciendo así círculos de pobreza difícilmente reversibles, mientras que la población acomodada se ha concentrado en lugares con gran disponibilidad de equipamientos de calidad y objetos de diversas acciones de renovación urbana. Esta situación no es estática, ya que las zonas de pobreza han crecido y densificado en las últimas décadas, particularmente al oriente de la metrópoli, donde habita cerca del 40% de la población.

La ciudad tiene una configuración compleja en la que sectores sociales, con diferencias significativas en sus niveles de ingresos, coexisten en proximidad espacial, aunque sin interacción social (Bayón, 2008). Las políticas neoliberales adoptadas a partir de los años de 1980 han provocado la expulsión de los sectores populares hacia la periferia, donde predomina la informalidad urbana y laboral. Más que una nueva configuración urbana, el escenario actual refleja una agudización de patrones preexistentes.

En este contexto, la ciudad se divide entre una ciudad con acceso a servicios y oportunidades, y otra carente de cualidades básicas, habitada por una población informalizada, con menores niveles educativos y oportunidades laborales, cuyo acceso a los lugares de empleo depende de largos

desplazamientos diarios. Esta herencia histórica condiciona las dinámicas actuales del mercado de vivienda, así como las formas en que se distribuyen el valor del suelo, el acceso a servicios y las oportunidades urbanas. Comprender este devenir histórico resulta crucial para interpretar los patrones espaciales contemporáneos que este trabajo analiza.

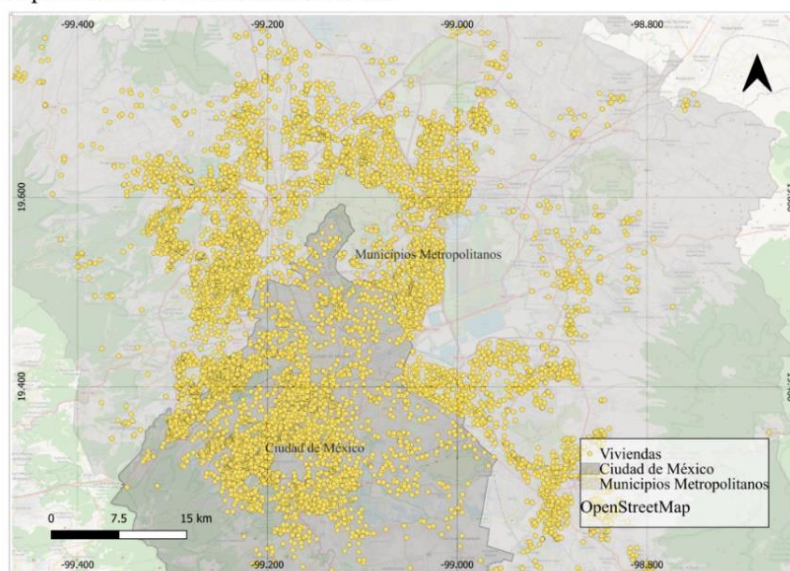
Metodología

Una vez tratadas las consideraciones teóricas y conceptuales, y después de haber brindado una caracterización general de la urbanización, distribución de la población y el empleo en la ZMCM, en esta sección se abordan los aspectos metodológicos que permitirán contestar a la pregunta general de investigación planteada al inicio del trabajo, a saber, ¿Cuáles son los efectos de los atributos de la vivienda, especialmente los elementos del entorno urbano, en la conformación de su precio en la ZMCM? Para tal propósito, en la primera sección se describe el método de obtención de los datos inmobiliarios, la fuente, el tamaño de la muestra, sus características y una breve discusión sobre sus alcances y límites. En el segundo apartado se muestra el proceso de construcción de las variables del entorno urbano, así como las fuentes de datos utilizadas. Finalmente, en el tercer apartado se describe el proceso y las herramientas analíticas usadas para contestar la pregunta de investigación. De forma breve, se anticipa que, dadas las complejidades metodológicas generadas por la naturaleza espacial de la información, se emplearon modelos de regresión espacial (SAR), y de aprendizaje automático, así como modelos agnósticos de interpretabilidad, cuyos resultados se contrastan y discuten a la luz de las circunstancias históricas de la metrópoli.

Obtención de datos

Los datos sobre el mercado inmobiliario se obtuvieron mediante el uso de la técnica rascado de datos o *webscrapping*. La página usada como fuente de información es “*Propiedades.com*”, debido al gran número de publicaciones inmobiliarias que contiene. La información obtenida del portal consiste únicamente en publicaciones de casas en venta, disponibles en la página del 16 al 20 de abril de 2025. La información corresponde a la Ciudad de México y los municipios que conforman a la ZMCM. La

Mapa 3. Oferta de vivienda en la ZMCM



Fuente: Elaboración propia.

muestra final resultante después de organizar y limpiar los datos es un conjunto de 10,802 viviendas, 4051 ubicadas en la ciudad de México y 6751 en los municipios metropolitanos (Mapa 3).

La información obtenida incluye el identificador de la vivienda en la página web, dirección, precio y moneda, la cual puede ser pesos mexicanos o dólares. También contiene información sobre las características del inmueble en términos de tamaño y calidad, como metros cuadrados de construcción y de terreno, edad, número de habitaciones, número de baños y presencia de distintas amenidades y características, como jardín, sótano, chimenea, calefacción, terraza, entre otras. Además, incluye las coordenadas geográficas en grados decimales que permiten conocer su ubicación exacta en el territorio y complementar su caracterización con información del entorno urbano.

La información estructural de las viviendas y su localización en coordenadas permiten capturar una gran diversidad de características físicas y contextuales. Esta heterogeneidad, tanto espacial como tipológica, representa una fortaleza para el análisis, ya que ofrece la posibilidad de explorar diferencias en función del entorno urbano, la localización y las condiciones materiales de los inmuebles.

Sin embargo, la muestra presenta ciertas limitaciones que deben ser consideradas al interpretar los resultados. En primer lugar, al extraerse exclusivamente de publicaciones activas en un portal inmobiliario en línea, y representar un momento específico en el tiempo, la muestra podría estar sesgada hacia el mercado formal y medio-alto, excluyendo porciones significativas del mercado informal o de menores ingresos, especialmente en zonas de menor cobertura digital. Además, se excluyen departamentos, viviendas en renta y otros tipos de inmuebles, por lo que los hallazgos no son generalizables al conjunto total del parque habitacional. Finalmente, la calidad de la información depende de la veracidad del autorreporte de los anunciantes, lo que puede generar inconsistencias o errores en variables clave como superficie construida, edad o características específicas.

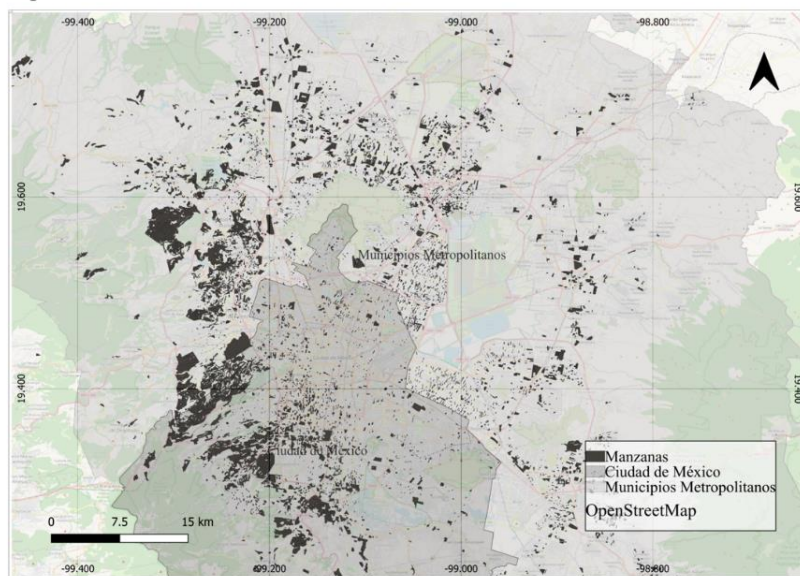
Si bien la muestra no representa la totalidad del mercado ni permite análisis longitudinales, su tamaño, riqueza informativa y diversidad estructural constituyen una base pertinente para examinar los determinantes espaciales y estructurales del precio de la vivienda en la ZMCM.

Construcción de variables

Las variables usadas para potencialmente caracterizar el vecindario se construyeron utilizando información del Censo de Población y Vivienda, 2020 (CPyV) del INEGI. Con herramientas de análisis espacial, se extrajo información sociodemográfica sobreponiendo la capa de puntos de viviendas a la información vectorial a nivel manzana del censo, de tal forma que se agregó la información del censo, de la

manzana en donde se localiza la vivienda, a la base de datos de puntos como una nueva columna (Mapa 4). Los campos extraídos del censo fueron: población total y grado promedio de escolaridad. Esta información fue utilizada para construir dos variables, las cuales son densidad de población y grado promedio de escolaridad, que se incluyeron en la categoría localización.

Mapa 4. Manzanas



Fuente: Elaboración propia.

Respecto a la caracterización del entorno urbano, puede realizarse mediante diferentes métodos. En este trabajo se seleccionó el método multibanda por las siguientes razones: Permite capturar relaciones no lineales entre los atributos y el precio, ya que mide el efecto de la presencia de un equipamiento a diferentes distancias. Otros métodos, como el análisis de densidad de núcleo o la medición directa de la distancia euclidiana a cada elemento de interés también permiten estimar esta relación; sin embargo, en el caso de la primera la interpretación directa se vuelve complicada, ya que no solo mide la cercanía, sino también la densidad o la cantidad de equipamientos disponibles. Respecto a la segunda, mantiene la facilidad interpretativa, pero implica un elevado costo computacional, si el número de elementos urbanos es alto.

La elección del método multibanda permite mantener la capacidad de estimar relaciones complejas, manteniendo simple la interpretación y reduciendo el costo computacional de su implementación.

Además, es un método ampliamente usado y recomendado en los análisis de precios hedónicos que estiman la relación con los elementos externos a la vivienda (Basu y Thibodeau, 1998).

Los elementos urbanos considerados son, por un lado, infraestructura de movilidad como estaciones y rutas del sistema de transporte público de la ZMCM. La información es de tipo vectorial y fue obtenida de la página de datos abiertos del gobierno de la Ciudad de México y la página de la Secretaría de Transporte y Movilidad del Estado de México. Las infraestructuras abarcan estaciones de Metro, Tren Ligero, Tren Suburbano, Teleférico (Cablebus y Mexicable), RTP, BRT (Metrobús y Mexibús), CETRAM, Ecobici, Biciestacionamientos y Trolebús; Rutas de Transporte Concesionado, Líneas de Ciclovía; y Aeropuertos.

Por otro lado, se consideran los equipamientos urbanos cuya información también es de tipo vectorial y se obtuvo del Directorio Estadístico Nacional de Unidades Económicas (DENUE) del INEGI. Las actividades tomadas en cuenta para el análisis son aquellos equipamientos de carácter público que proveen servicios como: Educación, Salud, Recreación, Cultura y Deporte y Oficinas de Gobierno. Las claves de las unidades económicas seleccionadas fueron agrupadas de la siguiente forma:

Cuadro 1. Equipamientos Urbanos

<i>Categoría</i>	<i>Código</i>	<i>Descripción</i>
Educación Básica	6111112	Preescolar público
	6111122	Primaria público
	6111132	Secundaria general público
	6111142	Secundaria técnica público
Educación Media	6111152	Media técnica terminal público
	6111162	Media superior público
Educación Superior	6111312	Educación superior público
Clinicas	6221112	Clinicas
Hospitales generales	6221122	Hospitales generales público
Hospitales de especialidad	6221132	Hospitales especialidades público
Cultura y Patrimonio	7121112	Museos público
	7121190	Grutas, parques naturales y otros sitios del patrimonio cultural de la nación
Parques, Recreación y Ocio	7121322	Jardines botánicos y zoológicos público
	7131112	Parques de diversiones y temáticos públicos
	7131142	Parques acuáticos y balnearios públicos
	7131942	Clubes deportivos del sector público
Actividades Gubernamentales	931210	Administración pública general

Fuente: Elaboración propia con datos de INEGI, 2025

Algunas actividades se dividieron y algunas unidades económicas se omitieron debido a posibles errores de clasificación o porque interferían con el análisis. Por ejemplo, en actividades gubernamentales y educación superior, se omitieron todas aquellas unidades cuya razón social contuviera la palabra “Bodega”, pues no se considera que sea una unidad económica que provea el

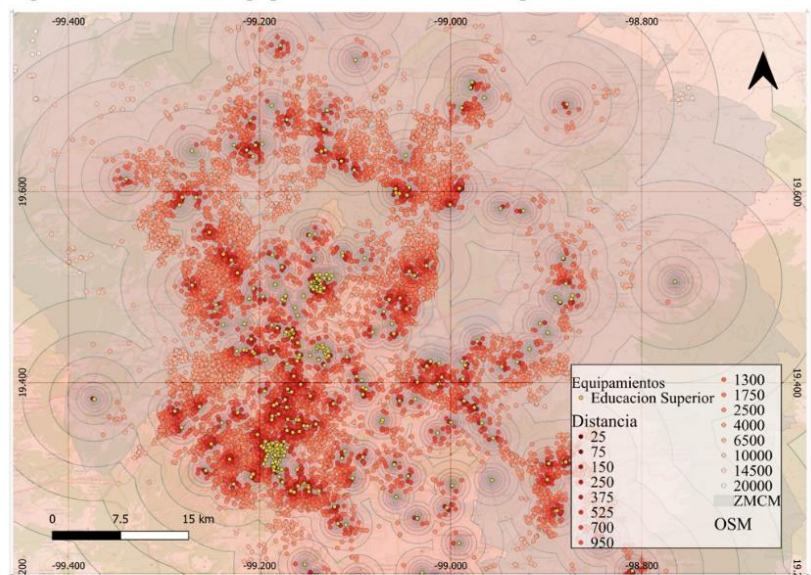
servicio de la categoría a la que pertenece. En el caso de la actividad “622112”, Hospitales Generales, debido a que las unidades presentes en esa categoría incluyen tanto clínicas como hospitales, se separaron aquellas unidades económicas cuya razón social contuviera la palabra “Hospital”. En este caso, ambos grupos se incluyeron en el análisis como categorías distintas, Clínicas Médicas y Hospitales Generales, respectivamente.

Las bandas se construyeron mediante *buffers* alrededor de los elementos urbanos, de tal forma que cada variable de accesibilidad a un elemento urbano específico contiene 20 anillos o rangos de distancia, cuyos valores son: 0-50, 100, 200, 300, 450, 600, 800, 1,100, 1,500, 2,000, 3,000, 5,000, 8,000, 12,000, 17,000, 23,000, 30,000, 38,000, 47,000 y 60,000 metros. La elección de estas distancias se basa en la expectativa teórica de que a distancias cortas los efectos de los elementos son más intensos y variados, siguiendo la primera ley de Todler, por lo que los anillos capturan más variaciones a distancias cortas.

Al igual que con la información vectorial del censo de población, la capa de puntos se sobrepuso a la capa de bandas y se extrajo el valor de la banda en que se localiza cada vivienda. Cabe decir que, en la creación de las bandas, los bordes se diluyeron, por lo que no existe sobreposición de bandas y la variable resultante representa la distancia al elemento urbano de interés más cercano. El valor que se extrae es la distancia media entre el límite superior e inferior de la banda.

Por ejemplo, el Mapa 5 muestra la accesibilidad de las viviendas a equipamientos de educación superior. Como se puede observar, las viviendas más cercanas se presentan en colores más intensos, lo que refleja su cercanía a este elemento urbano, mientras que, estar más lejos, disminuye la intensidad del color. Lo mismo sucede con el Mapa 6 que muestra la accesibilidad a estaciones de Metro.

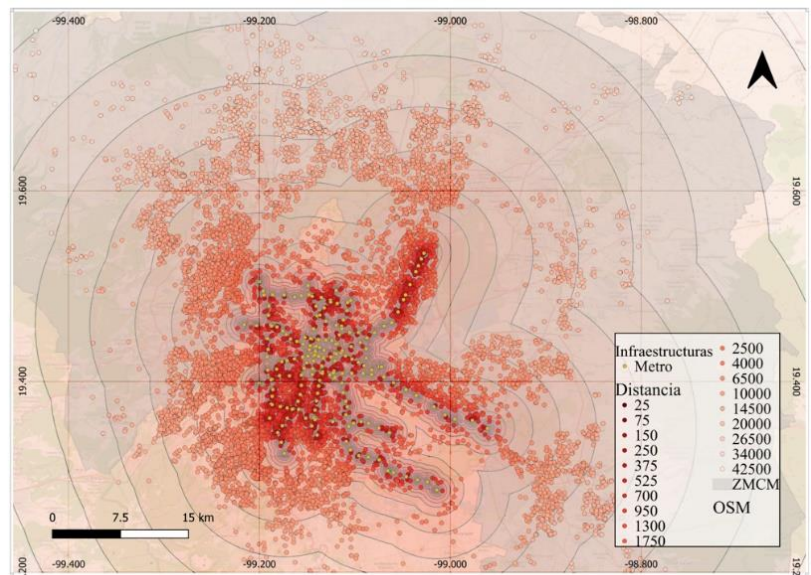
Mapa 5. Accesibilidad a equipamientos de educación superior



Fuente: Elaboración propia.

Una característica importante para destacar al comparar ambos mapas es la diferente concentración espacial de los elementos urbanos. Se puede ver una mayor concentración de las estaciones de Metro en el centro de la metrópoli y una mayor cobertura de los equipamientos de educación superior en toda la metrópoli, que además se refleja en los

Mapa 6. Accesibilidad a estaciones de Metro



Fuente: Elaboración propia.

valores de distancia que toman las viviendas. En el caso de las estaciones de Metro, más concentradas espacialmente, el rango de valores de distancia es mayor, hasta 42 km, ya que hay viviendas más alejadas a este tipo de elemento urbano. Contrariamente, los equipamientos de educación superior, al estar más extendidos, muestran un menor rango de distancia, hasta 20 km. Estas diferencias no se limitan a estos ejemplos, sino que, en general, todos los elementos de infraestructura de movilidad se localizan más concentrados y, por lo tanto, hay viviendas localizadas a distancias mayores. Por su parte, los equipamientos urbanos tienen una mayor cobertura, que se refleja en valores de distancia menores.

Como en estos mapas, todas las infraestructuras y equipamientos urbanos tuvieron el mismo tratamiento, lo que resultó en un total de 22 variables que dan cuenta de la estimación potencial de accesibilidad a 22 tipos de elementos urbanos distintos, 13 de infraestructura de movilidad y nueve de equipamientos urbanos.

Finalmente, para las variables de localización se incluyó la distancia al DCN y al centro financiero de Santa Fe. Estos fueron tomados del trabajo de Muñiz et al, 2015, quienes realizaron un ejercicio de identificación de centros y subcentros de empleo mediante métodos estadísticos rigurosos. Su trabajo identificó siete subcentros y un mercado central. El mercado central fue ubicado en el Ángel de la Independencia y los subcentros en Santa Fe, Aeropuerto, Central de Abastos, Pantitlán, Cuautitlán, Ecatepec, Villa Coapa, Tlalnepantla y Tlalpan. Sin embargo, en el presente trabajo solo

se consideraron el DCN y el subcentro de Santa Fe para simplificar el análisis. También se incluyó una variable espacialmente rezagada con la variable dependiente, tanto en los modelos de econometría espacial, como los modelos de aprendizaje automático, la cual fue calculada con una matriz de pesos espaciales construida con el método k vecinos más cercanos = 10.

La variable de respuesta utilizada en el análisis es la transformación logarítmica del precio de la vivienda, debido a que la distribución de la variable “precio” está fuertemente sesgada a la derecha y a que la variable logarítmica mostró un mejor ajuste en todos los modelos. Por su parte, las variables independientes fueron estandarizadas en los modelos paramétricos MCO y SAR para facilitar su comparabilidad. Sin embargo, en los modelos de aprendizaje automático, debido a la flexibilidad que los caracteriza y con el objetivo de simplificar la interpretación de los resultados en el contexto de relaciones no lineales, se utilizaron los valores absolutos de las variables independientes. Las variables usadas se resumen en el siguiente cuadro:

Cuadro 2. Variables

<i>Clase</i>	<i>Variable</i>	<i>Tipo</i>	<i>Descripción</i>
Dependiente	Ln_Precio	N Numérica	Logaritmo natural del precio. Interpretación en términos porcentuales.
Estructurales	Área construida	N Numérica	Área construida en metros cuadrados
	Recámaras	N Numérica	Número de habitaciones en la vivienda
	Baños	N Numérica	Número de baños
	Niveles	N Numérica	Número de niveles en la vivienda
	Edad	N Numérica	Años de antigüedad del inmueble
	Jardín	B Binaria	Presencia de jardín = 1
	Terraza	B Binaria	Presencia de terraza = 1
Localización	Densidad de Pob	N Numérica	Resultado de dividir población total entre tamaño de la manzana en Ha
	GRAPROES	N Numérica	Promedio de años de estudio
	Dist_DCN	N Numérica	Distancia al Distrito Central
	Dist_Santa_Fe	N Numérica	Distancia al subcentro económico Santa Fe
Entorno Urbano	Metro	N Numérica	Distancia a la estación de metro más cercana
	Tren ligero	N Numérica	Distancia a la estación de tren ligero más cercana
	Tren suburbano	N Numérica	Distancia a la estación de tren suburbano más cercana
	Teleferico	N Numérica	Distancia a la estación de Teleferico más cercana
	BRT	N Numérica	Distancia a la estación de BRT más cercana
	RTP	N Numérica	Distancia a la estación de RTP más cercana
	Trolebús	N Numérica	Distancia a la estación de trolebús más cercana
	Ecobici	N Numérica	Distancia a la estación de Ecobici más cercana
	Ciclovías	N Numérica	Distancia a la ciclovía más cercana
	Bicestacionamientos	N Numérica	Distancia al bicestacionamiento más cercano
	CETRAM	N Numérica	Distancia al centro de transferencia modal más cercano
	Aeropuerto	N Numérica	Distancia al aeropuerto más cercano
	Transp_Conc	N Numérica	Distancia a la ruta de transporte concesionado más cercana
	Clin_Med	N Numérica	Distancia a la clínica o unidad médica más cercana
	Hosp_Gen	N Numérica	Distancia al hospital general más cercano
	Hosp_Esp	N Numérica	Distancia al hospital de especialidades más cercano
	Edu_Bas	N Numérica	Distancia a la escuela de educación básica más cercana
	Edu_Med	N Numérica	Distancia a la escuela de educación media más cercana
	Edu_Sup	N Numérica	Distancia a la escuela de educación superior más cercana
	Cut_Patrim	N Numérica	Distancia a equipamientos culturales o turísticos patrimoniales cercanos
	Parq_Rec_Jar	N Numérica	Distancia a áreas de parques recreativos más cercanos
	Ofi_Gob	N Numérica	Distancia a oficina de administración general de gobiernos más cercana

Fuente: Elaboración propia

Método de Análisis

El análisis empírico de los determinantes del precio se divide en tres secciones. En la primera se describen las características del mercado de vivienda, sus rangos de precios, la estructura espacial y los valores promedios de las variables. Algunas variables se comparan con información de bases de datos como la Encuesta Nacional de Vivienda (ENVI) y el CPyV con la finalidad de brindar una perspectiva más amplia y crítica sobre el alcance y límites de los resultados obtenidos.

En la segunda sección se aplican modelos de regresión para estimar la primera fase de la metodología hedónica o estimación de la Función de Precios Hedónicos. Para una estimación consistente se utilizaron diversas herramientas estadísticas. Primero, se estimaron pruebas de multicolinealidad y se seleccionaron aquellas con menor correlación. Posteriormente, se estimó un modelo base de MCO bajo los principios de parsimonia con las variables seleccionadas, cuyos errores se analizaron mediante diversas pruebas estadísticas para buscar indicios autocorrelación espacial y verificar el cumplimiento de los supuestos clave para estimar coeficientes confiables.

Con los resultados obtenidos en esta etapa se justificó el uso de modelos espaciales y se estimaron modelos SAR y SEM los cuales se compararon mediante las medidas de bondad de ajuste R^2 , AIC, BIC, y Log-likelihood para determinar el mejor modelo. Los coeficientes del modelo seleccionado se interpretaron considerando las implicaciones de la dependencia espacial. Por tal motivo, se calcularon los efectos directos, indirectos y totales, que reflejan el impacto completo de las variables explicativas, teniendo en cuenta las interdependencias espaciales, y se interpretaron conforme a lo expuesto en la sección teórica (LeSage y Pace, 2009).

Adicionalmente, con un espíritu exploratorio, se estimaron modelos de aprendizaje automático y se aplicaron modelos agnósticos de interpretabilidad para mostrar la dependencia parcial y las contribuciones locales de las variables. Los modelos de interpretabilidad permiten explorar los detalles y matices de los efectos de los atributos en el precio. La dependencia parcial de las variables se calculó mediante ALE, que permite identificar efectos marginales promedio a lo largo del rango de cada variable, ofreciendo un puente entre los coeficientes globales de los modelos econométricos y las relaciones complejas capturadas por los modelos de aprendizaje automático. Por su parte, las contribuciones locales se calcularon con el uso de valores SHAP, los cuales permiten observar las contribuciones individuales de cada variable en las predicciones del modelo, descomponiendo su valor estimado para cada vivienda.

Cabe señalar que, si bien existen distintos enfoques tradicionales para analizar estas relaciones, como los modelos GAM para relaciones no lineales y la regresión geográficamente ponderada GWR o la estimación de submercados mediante análisis de conglomerados para la heterogeneidad espacial, estos no fueron implementados en el presente trabajo debido a las limitaciones prácticas y de alcance. Por un lado, la estimación de un modelo GWR para una muestra de 10,802 observaciones implicaría un costo computacional muy elevado y requeriría una calibración intensiva del ancho de banda o ventana de observaciones. Por otro lado, la segmentación del mercado en subgrupos mediante técnicas de clustering implicaría un desarrollo metodológico y analítico adicional que excede el enfoque exploratorio de esta investigación.

En su lugar, el enfoque adoptado permite identificar patrones no lineales y variaciones locales sin necesidad de fragmentar el espacio urbano ni imponer supuestos paramétricos. De esta manera, mientras los modelos espaciales permiten captar estructuras y dependencias globales, las técnicas de aprendizaje automático aportan una perspectiva más flexible, detallada y localizada, revelando matices y complejidades que suelen quedar fuera del alcance de los métodos paramétricos tradicionales.

Finalmente, en la última sección se comparan y discuten los resultados obtenidos tanto por los modelos de econometría espacial como los modelos agnósticos de interpretabilidad, a la luz de las características estructurales históricas de la metrópoli y se brindan conclusiones sobre los efectos de los atributos de la vivienda, especialmente de los elementos del entorno urbano, en su precio.

Análisis empírico

Descripción del mercado de vivienda

La muestra representa a 10,802 viviendas en oferta, con precios que van desde los \$501,000 hasta los \$180 millones de pesos. El precio promedio es de 9.85 millones, pero este valor está influido por viviendas de lujo con precios muy altos. La mayoría de las viviendas tiene precios mucho menores: la mediana es de \$4.85 millones, lo que significa que la mitad de las viviendas cuesta menos que eso. Además, el 25% de las viviendas tienen un valor menor a \$1.85 millones, mientras que el 75% está por debajo de \$11.5 millones. La desviación estándar es alta (\$14.66 millones), lo que indica que los precios varían mucho entre sí (Cuadro 3).

Cuadro 3. Características estructurales

<i>Variable</i>	<i>Media</i>	<i>Desv. Est</i>	<i>Min</i>	<i>25%</i>	<i>50%</i>	<i>75%</i>	<i>Max</i>
Precio	\$9,853,823	\$14,656,771	\$501,000	\$1,855,000	\$4,850,000	\$11,500,000	\$180,000,000
Área Construida	288.38	243.97	45	122	227	367	5657
Recámaras	3.45	1.19	1	3	3	4	8
Niveles	2	0.46	1	2	2	2	4
Baños	2.55	1.3	1	2	2	3	9
Edad	22.44	12.33	1	18	20	26	110

Fuente: Elaboración propia

Las viviendas tienen, en general, entre tres y cuatro recámaras, dos niveles y un promedio de 2.5 baños. Al comparar el número de recámaras con información del CPyV sobre las características de las viviendas en las manzanas donde se ubica la muestra (Cuadro 4), observamos que en promedio el 89% de las viviendas en dichas manzanas tienen tres o más cuartos, lo cual coincide parcialmente con los resultados obtenidos.

Cuadro 4. No. de cuartos

<i>Variable</i>	<i>Promedio</i>
Un cuarto	2%
Dos cuartos	7%
Tres o más	89%

Fuente: Elaboración propia con datos de INEGI, 2020

En cuanto al área construida, ésta varía considerablemente: la media es de 288 m², la mediana de 227m², y existen valores extremos superiores a los 5,000 m². Si comparamos estos datos con los resultados de la Encuesta Nacional de Vivienda, 2021 del INEGI, se observa una diferencia sustancial (Cuadro 5). De acuerdo con la ENVI, la mitad del parque habitacional en la Ciudad de México tiene un tamaño promedio de hasta 75m², y solamente el 12% tiene un tamaño superior a los 150m². En el Estado de México, el 64% del parque habitacional tiene un tamaño de construcción

inferior a los 100m² y únicamente el 16% excede los 150m². Estas diferencias podrían indicar que la muestra obtenida mediante web scrapping está sesgada hacia segmentos de vivienda correspondientes a estratos socioeconómicos medios y altos.

Cuadro 5. M² de construcción, ENVI

<i>Entidad</i>	<i>Hasta 45 m²</i>	<i>46–75 m²</i>	<i>76–100 m²</i>	<i>101–150 m²</i>	<i>>150 m²</i>	<i>No especificado</i>
CDMX	21%	33%	16%	11%	12%	6%
México	18%	25%	23%	13%	16%	6%

Fuente: Elaboración propia con datos de INEGI, 2021

Respecto a la antigüedad de las construcciones, la edad promedio es de 22 años, con una mediana de 20 años. Algunas viviendas superan incluso los 100 años de antigüedad. El 50% de los inmuebles se concentran en un rango de entre 18 y 26 años. Al contrastar con los datos de la ENVI (Cuadro 6), encontramos una mayor coincidencia: en la Ciudad de México, aproximadamente el 40% de las viviendas tiene entre 11 y 30 años de antigüedad, mientras que en el Estado de México este porcentaje asciende al 50%. Finalmente, 3,934 viviendas cuentan con jardín y 2499 con terraza.

Cuadro 6. Edad, ENVI

<i>Entidad</i>	<i>≤5 años</i>	<i>6–10 años</i>	<i>11–20 años</i>	<i>21–30 años</i>	<i>≥31 años</i>	<i>No especificado</i>
CDMX	3%	9%	20%	20%	41%	7%
México	8%	11%	26%	24%	25%	7%

Fuente: Elaboración propia con datos de INEGI, 2021

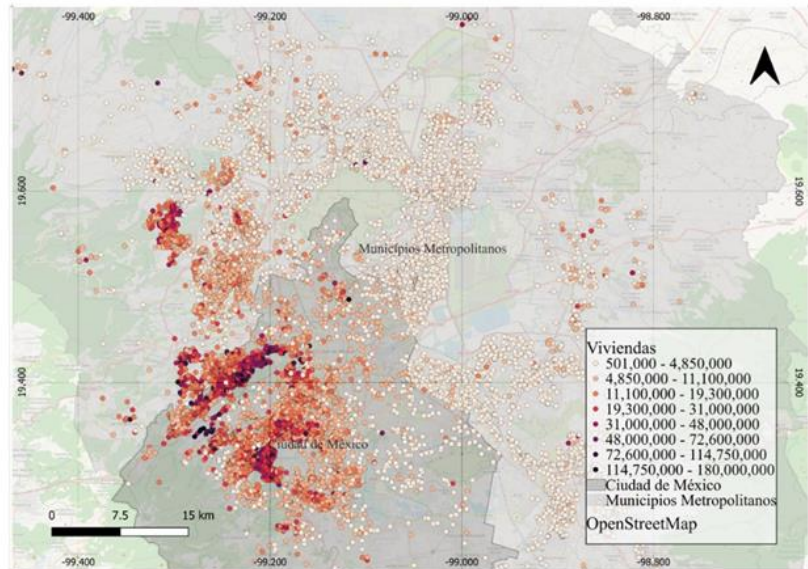
Esta caracterización estructural sugiere que muchas de las viviendas de la muestra poseen atributos asociados a segmentos de clase media y alta. No obstante, la base de datos también incluye viviendas de menor tamaño y valor, algunas desde 45m² y con precios desde \$501,000, así como propiedades distribuidas a lo largo de prácticamente todo el territorio urbano, incluyendo zonas populares o con cierto grado de marginación.

En este sentido, aunque la muestra obtenida mediante web scrapping refleja principalmente la oferta inmobiliaria formal activa, y tiende a sobre representar zonas de características de vivienda más amplias, no se limita exclusivamente a estos segmentos. Por tanto, los resultados del estudio deben interpretarse como representativos del mercado inmobiliario visible en portales digitales,

con un sesgo probable hacia viviendas de nivel medio y alto, pero sin excluir completamente la diversidad del parque inmobiliario en oferta.

La distribución espacial de los precios revela una estructura urbana fuertemente desigual. En el Mapa 7 se observa que las viviendas de más valor se concentran en la zona superpuesta de la ZMCM, y dentro de esta zona se forman enclaves de alto valor en colonias como Jardines del Pedregal, San Ángel, Coyoacán, Condesa, Polanco,

Mapa 7. Distribución espacial de precios



Fuente: Elaboración propia.

Lomas de Chapultepec, Bosques de las Lomas, Condado de Sayavedra y Hacienda de Vallescondido. En estas colonias, los valores de las viviendas oscilan entre los \$48 millones hasta \$180 millones. En contraste, en las extensas zonas periféricas al norte y oriente de la metrópoli, las viviendas no rebasan los \$4.85 millones, con valores promedio de \$2.15 millones. Esta profunda brecha de valores coincide con el patrón histórico de urbanización característico de la ciudad, marcado por una concentración de infraestructura y equipamiento de calidad en los centros tradicionales y carencias significativas en las periferias.

Las viviendas se localizan en vecindarios con perfiles socio demográficos y territoriales marcadamente diversos (Cuadro 7). La escolaridad promedio es de 12.4 años, equivalente al nivel medio superior, con una desviación estándar de 2.07 años. Sin embargo, existen zonas donde apenas alcanza los 4.6 años y otras donde se aproxima a los 19.5, lo que refleja una significativa disparidad educativa. La densidad poblacional también muestra una alta variabilidad: aunque el promedio es de 165 habitantes por hectárea, el valor mínimo es de apenas 0.4 y el máximo supera los 2,000 habitantes (Culhuacán IX A), con una desviación estándar de 158.2, lo que evidencia distintos grados de concentración urbana.

Cuadro 7. Variables de localización

<i>Variable</i>	<i>Media</i>	<i>Desv. Est</i>	<i>Min</i>	<i>25%</i>	<i>50%</i>	<i>75%</i>	<i>Max</i>
Prom Escolaridad	12	2	5	11	12	14	19
Densidad Pob (ha)	165	158	0	51	122	239	2191
Dist_Santa_Fe (m)	24555	15306	569	11395	22687	36366	81766
Dist_DCN (m)	19350	11260	359	11580	16333	25761	69968

Fuente: Elaboración propia con datos de INEGI, 2020

En cuanto a la localización respecto a los centros de empleo, la distancia media a Santa Fe es de 24.6 km, con un rango que va desde 569 metros hasta más de 81 km. Para el Distrito Central de Negocios (DCN), la distancia media es de 19.4 km, con valores que oscilan entre 359 metros y casi 70 km, lo que refleja la gran extensión territorial de la metrópoli.

Las viviendas presentan una marcada heterogeneidad respecto a su accesibilidad y proximidad a distintos elementos urbanos (Cuadro 8). En promedio, están ubicadas a 17.2 km de algún aeropuerto, algunas estando muy próximas, mientras que otras a más de 50 km de distancia. En cuanto al sistema de transporte público, la distancia media a una estación de BRT es de 5.6 km; a un CETRAM, 10.7 km; al Teleférico, 10.2 km; al Metro, 8.9 km; al RTP, 6.8 km; y al Trolebús 9.9 km, reflejando una distribución diversa en términos de conectividad modal, pero también mostrando que el BRT es el modo con mayor cobertura territorial.

Respecto a la movilidad sustentable, la distancia promedio a Biciestacionamientos es de 13.2 km; a Ecobici, 14.2 km; y a Ciclovías de 8.5 km, lo que sugiere una distribución más concentrada y accesibilidad más restringida de estas infraestructuras respecto a la movilidad masiva. La distancia media al Tren Ligero es de 21.5 km y al Tren Suburbano, de 14.5 km, lo que de igual forma sugiere que estos modos de transporte están concentrados en el territorio.

En cuanto a servicios educativos, la distancia media a escuelas de educación básica es de solo 0.52 km, mientras que para educación media superior es de 1.3 km y para educación superior es 2.2 km, lo que sugiere una amplia cobertura de estos servicios. Respecto a los servicios de salud, la accesibilidad a clínicas médicas es de 2.9 km, mientras que a hospitales generales y de especialidad es de 3.7 km y 5.2 km en promedio, respectivamente.

En términos de acceso a espacios de recreación y cultura, la distancia promedio a jardines, parques o espacios recreativos es de 2.7 km, y a museos y sitios patrimoniales es de 4.5 km. El acceso a

oficinas de gobierno está más extendido, con una distancia media de 0.98 km, lo que indica una mayor cobertura de los servicios administrativos.

Cuadro 8. Variables del entorno urbano

<i>Elemento Urbano</i>	<i>Media</i>	<i>Desv. Est</i>	<i>Min</i>	<i>25%</i>	<i>50%</i>	<i>75%</i>	<i>Max</i>
Aeropuerto	17233.7	7071.5	950	14500	20000	20000	53500
BRT	5628.8	5866.8	25	1300	4000	6500	53500
CETRAM	10695.3	9932.6	75	4000	6500	14500	53500
BiciEstacionam	13202.1	8866.5	75	6500	10000	14500	53500
Teleferico	10171.8	7700.3	75	6500	10000	14500	53500
Ecobici	14219.2	11779.7	25	6500	10000	20000	53500
Metro	8931.6	8561.2	25	2500	6500	14500	53500
RTP	6819.7	8596.2	25	375	4000	10000	53500
Tren_ligero	21468.6	13287.0	150	10000	20000	34000	53500
Tren_Suburbano	14500.7	8900.8	150	10000	14500	20000	53500
Ciclovia	8531.5	9631.3	25	1300	4000	14500	53500
Transp_Conce	2384.6	5070.8	25	150	375	2500	42500
Trolebus	9876.3	9337.3	25	2500	6500	14500	53500
Clin_Med	2930.2	2275.2	25	1300	2500	4000	20000
Hosp_Gen	3650.7	3189.1	25	1750	2500	4000	34000
Hosp_Esp	5215.1	3537.5	75	2500	4000	6500	34000
Edu_Bas	521.1	878.1	25	150	375	525	14500
Edu_Med	1301.3	1158.6	25	700	950	1750	20000
Edu_Sup	2207.7	2135.3	25	950	1750	2500	26500
Cult_Patrim	4486.9	3879.6	25	2500	4000	6500	34000
Jard_Parq_Rec	2679.5	2239.0	25	1300	2500	4000	20000
Ofi_Gob	980.2	1022.7	25	375	700	1300	20000

Fuente: Elaboración propia

En conjunto, la caracterización presentada muestra una notable heterogeneidad en el mercado de vivienda. Las unidades ofertadas varían considerablemente en términos de tamaño, localización territorial, contexto sociodemográfico y nivel de accesibilidad a equipamientos e infraestructura urbana. Estas diferencias se reflejan en una amplia dispersión de precios y en una distribución espacial desigual, donde los valores más altos se concentran en zonas con mayor dotación de bienes urbanos, mientras que los más bajos predominan en áreas periféricas o con menor consolidación.

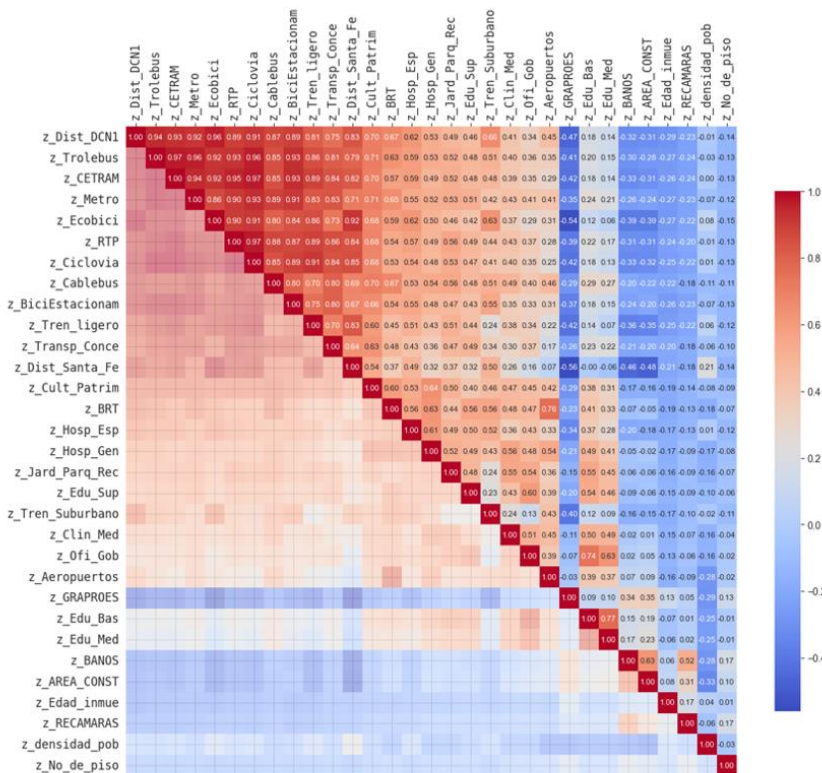
Asimismo, la comparación con fuentes oficiales como la ENVI y el CPyV sugiere que la muestra obtenida mediante web scrapping tiende a sobre representar segmentos medios y altos del mercado formal. No obstante, se conserva cierta diversidad en cuanto a tamaño, valor y ubicación de las viviendas, lo que permite una representación parcial pero útil del mercado inmobiliario visible en plataformas digitales.

Esta diversidad estructural, espacial y social constituye un punto de partida suficiente para el análisis de los factores asociados a la determinación del precio de la vivienda. En la siguiente sección, se abordará esta cuestión mediante la estimación de modelos econométricos y de aprendizaje automático, incorporando explícitamente la dimensión espacial del mercado.

Determinantes del precio

En primer lugar, para evaluar la posible colinealidad entre las variables y reducir problemas en la estimación de los coeficientes, se calculó una matriz de correlación (Figura 1) y se estimaron los factores de inflación de la varianza (VIF) (Cuadro 9). Como se mencionó en el apartado metodológico, las variables usadas en los modelos paramétricos fueron estandarizadas para permitir la comparabilidad entre ellas. Los resultados revelaron elevados niveles de colinealidad en variables relacionadas con la distancia al DCN, al centro financiero Santa fe y a elementos urbanos como ciclovías, estaciones de metro, RTP, Trolebús, Ecobici y CETRAM. Por tal motivo, dichas variables fueron excluidas de la estimación del modelo base. La selección final de variables se presenta en el Cuadro 10.

Figura 1. Matriz de correlación



Fuente: Elaboración propia

Cuadro 9. VIF

Variable	VIF
z_Dist_DCN1	57.35
z_Ciclovia	56.64
z_Ecobici	49.67
z_CETRAM	45.70
z_Trolebus	43.13
z_Metro	33.13
z_RTP	31.93
z_Dist_Santa_Fe	29.97
z_BiciEstacionam	25.06
z_Tren_ligero	17.21
z_Teleférico	11.32
z_Tren_Suburbano	9.62
z_Transp_Conce	7.00
z_Aeropuertos	5.92
z_BRT	5.81
z_Edu_Bas	4.15
z_Hosp_Gen	3.04
z_Ofi_Gob	2.95
z_Edu_Med	2.89
z_Hosp_Esp	2.73
z_Cult_Patrim	3.00
z_Edu_Sup	2.39
z_BANOS	2.33
z_Jard_Parq_Rec	2.32
z_Clin_Med	2.08
z_AREA_CONST	1.97
z_GRAPROES	1.85
z_RECAMARAS	1.57
terracea	1.39
z_densidad_pob	1.38
Jardin	1.29
z_Edad_inmue	1.15
z_No_de_piso	1.07

Fuente: Elaboración propia

Cuadro 10. Selección final de variables

<i>Estructura</i>	<i>Localización</i>	<i>Infraestructura</i>	<i>Equipamientos</i>	<i>Total</i>
z_RECAMARAS	z_GRAPROES	z_Aeropuertos	z_Clin_Med	
z_No_de_piso	z_densidad_pob	z_BiciEstacionam	z_Hosp_Gen	
z_BANOS		z_BRT	z_Hosp_Esp	
z AREA_CONST		z_Teleférico	z_Edu_Bas	
z_Edad_inmue		z_Tren_ligero	z_Edu_Med	
Jardin		z_Tren_Suburbano	z_Edu_Sup	
terrazza		z_Transp_Conce	z_Cult_Patrim	
			z_Jard_Parq_Rec	
			z_Ofi_Gob	
	7	2	7	9
				25

Fuente: Elaboración propia

Con esta selección se estimó un modelo base de regresión lineal mediante Mínimos Cuadrados Ordinarios cuyos resultados se presentan en el Cuadro 11. El modelo muestra un buen ajuste global, con un R^2 de 0.8283 y un R^2 ajustado de 0.8279, lo que indica que aproximadamente el 83% de la variabilidad del \ln_precio de la vivienda es explicada por las variables incluidas. Además, el modelo, a nivel global, es estadísticamente significativo y presenta niveles aceptables de colinealidad.

Para verificar el cumplimiento de los supuestos clásicos y la validez del modelo, se aplicaron diversas pruebas estadísticas, las cuales se muestran en el Cuadro 12. La prueba de Jarque-Bera revela una fuerte desviación de los errores respecto a la normalidad, con alta significancia estadística, lo que indica que los residuos no siguen una distribución normal. En cuanto a la homoscedasticidad, tanto la prueba de Breusch-Pagan como la de Koenker-Bassett fueron estadísticamente significativas, lo que revela la presencia de heteroscedasticidad, es decir, una varianza no constante en los errores.

Cuadro 11. Resultados MCO

<i>Variable / Métrica</i>	<i>Valor</i>
Var dependiente	\ln_precio
Observaciones	10802
Número de variables	26
Grados de libertad	10776
R-cuadrado (R^2)	0.8283
R^2 ajustado	0.8279
Varianza del error	0.241
Error estándar de regresión	0.491
Estadístico F	2079.5183
Prob (F)	0
Log-verosimilitud	-7623.585
Criterio de Akaike (AIC)	15299.169
Criterio de Schwarz (BIC)	15488.644
Multicolinealidad	9.049

Fuente: Elaboración propia

Cuadro 12. Pruebas de diagnostica

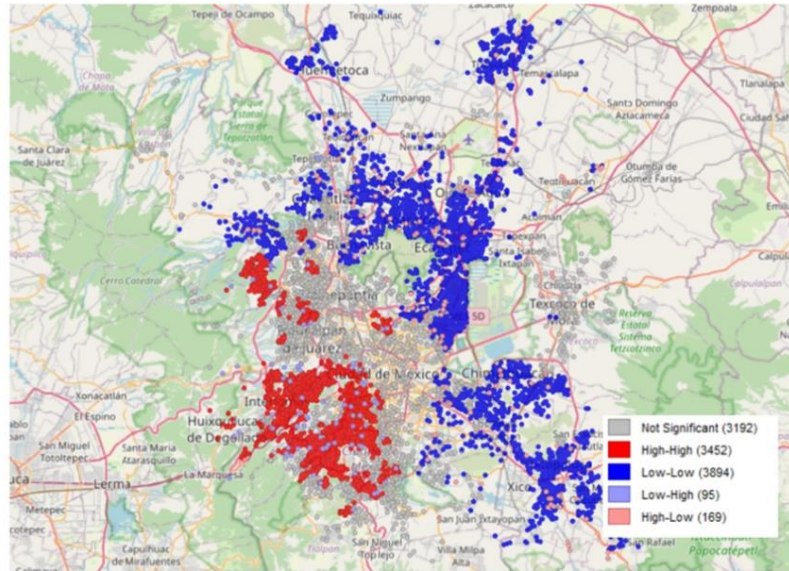
<i>Prueba</i>	<i>MI/GL</i>	<i>Valor</i>	<i>Probabilidad (p-valor)</i>
<i>Test de normalidad de errores</i>			
Jarque-Bera	2	36,976.71	0
<i>Diagnóstico de heterocedasticidad</i>			
Breusch-Pagan	25	34,340.32	0
Koenker-Bassett	25	6,262.19	0
<i>Diagnóstico de dependencia espacial</i>			
Moran's I (error)	0.2321	58.084	0
Multiplicador de Lagrange (rezago)	1	3,874.93	0
Multiplicador de Lagrange robusto (rezago)	1	1,375.16	0
Multiplicador de Lagrange (error)	1	3,272.46	0
Multiplicador de Lagrange robusto (error)	1	772.695	0

Fuente: Elaboración propia

Respecto a la autocorrelación espacial, los resultados también fueron significativos. El índice local Moran aplicado a los errores mostró un valor positivo y significativo, mientras que las pruebas de multiplicadores de Lagrange (LM), tanto para el modelo con rezago espacial (SAR) como para el modelo de error espacial (SEM), resultaron todas significativas, mostrando evidencia de autocorrelación en la variable dependiente y en los residuos.

Adicionalmente, el mapa del índice local de autocorrelación espacial (LISA) I de Moran aplicado a la variable dependiente (Mapa 8), revela aglomeraciones de viviendas de precios altos en la zona suponiendo de la ciudad, de forma concentrada, y de viviendas de precios bajos en las zonas periféricas al norte y oriente, de forma más dispersa. Este

Mapa 8. Índice Local Moran I_n precio



Fuente: Elaboración propia.

patrón espacial es coherente con la distribución histórica desigual de la infraestructura y equipamiento urbano en la ciudad. En conjunto, estos resultados justifican la necesidad de emplear modelos espaciales para capturar adecuadamente la estructura de dependencia espacial presente en los datos.

Por lo tanto, se estimaron modelos espaciales globales con las mismas variables y se compararon usando distintas métricas para determinar el más robusto (Cuadro 13). En el cuadro comparativo se observa que el modelo SAR muestra el mejor desempeño global, con una Pseudo R^2 de 0.870 y Pseudo R^2 Espacial de 0.848. Además, muestra el mayor log-likelihood (-6232.22) y los menores valores de AIC

Cuadro 13. Comparativa entre modelos de regresión

<i>Variable / Métrica</i>	<i>MCO</i>	<i>SAR(MV)</i>	<i>SEM (MV)</i>
Número de variables	26	27	26
Grados de libertad	10776	10775	10776
R-cuadrado (R^2)	0.8283	-	-
R^2 ajustado	0.8279	-	-
Pseudo R-cuadrado	-	0.8704	0.8165
Pseudo R^2 espacial	-	0.8489	-
Varianza del error	0.241	0.1813	0.1858
Error estándar de regresión	0.491	0.4258	0.4311
Log-verosimilitud	-7623.585	-6232.2202	-6549.3374
Criterio de Akaike (AIC)	15299.169	12518.44	13150.675
Criterio de Schwarz (BIC)	15488.644	12715.203	13340.149

Fuente: Elaboración propia

(12518.44) y BIC (12715.20), lo que indica un mejor balance entre ajuste y parsimonia. Esta evidencia apoya la elección del modelo SAR como el más robusto para explicar la variabilidad espacial del precio de la vivienda, ofreciendo una mejor representación del efecto de los atributos estructurales, del entorno y de localización, incluidos los efectos espaciales.

Cabe recordar que, en el modelo SAR, debido a la presencia explícita de dependencia espacial en la variable dependiente, los coeficientes estimados, β , no reflejan los impactos marginales directamente como en un modelo de MCO. En cambio, estos deben ser multiplicados por un operador espacial, $(I - \rho W)^{-1}$, para poder dar cuenta de los efectos de retro propagación derivados de la inclusión de la matriz de vecindad W . Por ello, se estimaron los efectos directos, indirectos y totales, los cuales permiten descomponer cómo influye cada variable tanto en la unidad de análisis como en sus vecinas y se interpretaron siguiendo lo expuesto en el apartado teórico. Los resultados se presentan en el Cuadro 14.

Cuadro 14. Coeficientes modelo de rezago espacial (SAR)

<i>Variable</i>	<i>Coficiente</i>	<i>Error estándar</i>	<i>Estadístico Z</i>	<i>Efecto directo</i>	<i>Efecto indirecto</i>	<i>Efecto total</i>
CONSTANT	7.97152***	0.12836	62.10476	-	-	-
z_RECAMARAS	0.0443***	0.00566	7.82589	0.0443	0.0414	0.0857
z_No_de_piso	0.01633***	0.00439	3.71899	0.0163	0.0153	0.0316
z_BANOS	0.11591***	0.00648	17.8906	0.1159	0.1083	0.2242
z AREA_CONST	0.40655***	0.00646	62.90103	0.4066	0.3798	0.7864
z_Edad_inmue	0.02437***	0.00426	5.71683	0.0244	0.0228	0.0471
Jardin	0.01442	0.00891	1.6186	0.0144	0.0135	0.0279
terraza	0.08745***	0.01076	8.12713	0.0875	0.0817	0.1692
z_GRAPROES	0.06977***	0.00561	12.43936	0.0698	0.0652	0.1349
z_densidad_pob	-0.06298***	0.00516	-12.20433	-0.063	-0.0588	-0.1218
z_Aeropuertos	0.02612***	0.00767	3.4051	0.0261	0.0244	0.0505
z_BiciEstacionam	0.05157***	0.01006	5.12352	0.0516	0.0482	0.0997
z_BRT	0.02809***	0.00879	3.19341	0.0281	0.0262	0.0543
z_Teleférico	-0.02891***	0.00998	-2.89671	-0.0289	-0.027	-0.0559
z_Tren_ligero	-0.16968***	0.01018	-16.66774	-0.1697	-0.1585	-0.3282
z_Tren_Suburbano	-0.09084***	0.00738	-12.31333	-0.0908	-0.0849	-0.1757
z_Transp_Conce	0.05003***	0.00962	5.20253	0.05	0.0467	0.0968
z_Clin_Med	-0.00108	0.00585	-0.18388	-0.0011	-0.001	-0.0021
z_Hosp_Gen	0.01864***	0.00686	2.71703	0.0186	0.0174	0.0361
z_Hosp_Esp	0.01131*	0.00648	1.74469	0.0113	0.0106	0.0219
z_Edu_Bas	0.01851**	0.00837	2.21155	0.0185	0.0173	0.0358
z_Edu_Med	-0.00329	0.00699	-0.4708	-0.0033	-0.0031	-0.0064
z_Edu_Sup	-0.03785***	0.00633	-5.97764	-0.0379	-0.0354	-0.0732
z_Cult_Patrim	-0.01665**	0.00703	-2.36736	-0.0166	-0.0156	-0.0322
z_Jard_Parq_Rec	0.02641***	0.00619	4.26405	0.0264	0.0247	0.0511
z_Ofi_Gob	-0.01291*	0.00699	-1.84698	-0.0129	-0.0121	-0.025
W_ln_precio	0.48301***	0.00833	58.00712	-	-	-

*** si $p \leq 0.01$ ** si $0.01 < p \leq 0.05$ * si $0.05 < p \leq 0.10$

Fuente: Elaboración propia

El modelo SAR revela una fuerte dependencia espacial positiva en los precios de las viviendas, reflejado en el alto valor del término espacial (0.483), y significancia estadística al 99%. Esto quiere decir que el precio de una vivienda está positivamente correlacionado con los precios de las viviendas vecinas, es decir, las viviendas con valores similares tienden a aglomerarse en el espacio.

El área construida muestra el mayor impacto total (0.79), lo que confirma que el tamaño es el principal determinante del precio. Este efecto se descompone en un impacto directo de 0.40 y un impacto indirecto de 0.38. Esto significa que, ante un incremento de una desviación estándar en el área construida, aproximadamente 243m², el precio de la vivienda aumenta en promedio un 40% en la misma unidad, y genera, en promedio, un incremento del 38% en el precio de las demás viviendas conectadas espacialmente, debido a los efectos de dependencia espacial. La suma de ambos efectos representa el efecto total de esta variable en el sistema, es decir, un aumento promedio del 79% en el precio, considerando tanto el impacto directo sobre la propia vivienda como el efecto que dicho cambio provoca sobre las viviendas vecinas.

Este fenómeno refleja la capitalización de atributos y la interdependencia de precios locales, características propias de los mercados inmobiliarios, donde el valor de cada unidad se influye mutuamente a través de la dependencia espacial definida en la matriz de pesos, que en este caso corresponde a los 10 vecinos más cercanos.

Respecto a las demás variables estructurales, ante incrementos en una desviación estándar, el número de baños muestra un efecto total positivo en el precio de 22%, mientras que el número de recámaras un efecto del 9% y el número de niveles del 3%. La edad del inmueble también muestra un efecto positivo del 5%. En este caso, si bien en la literatura, en general, se ha encontrado un efecto negativo, este resultado es consistente con los hallazgos de Sabella (1974) quien propone que el valor de una propiedad se eleva en los últimos años debido al aumento en el valor del suelo. Finalmente, la presencia de terraza tiene un efecto total positivo en el precio de 16% mientras que la presencia de jardín eleva el precio en 2%.

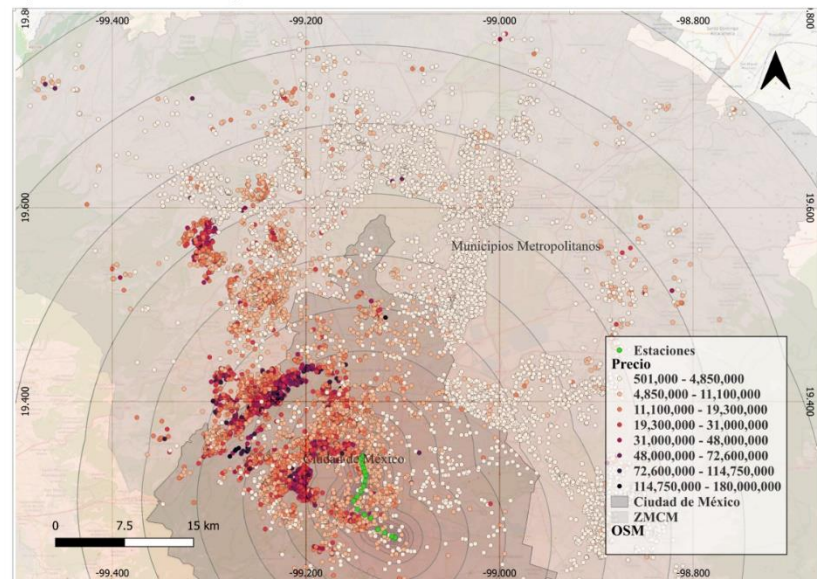
Por su parte, la densidad poblacional muestra un efecto total negativo, indicando que zonas una desviación estándar más densas se asocian con precios 12% menores. Esto podría estar relacionado con efectos de saturación urbana, como tráfico o contaminación. En cambio, zonas con una desviación estándar mayor de años promedio de escolaridad se asocian con precios de vivienda 13% más altos.

Respecto a los elementos urbanos, la infraestructura de movilidad muestra efectos mixtos. La cercanía a estaciones de Tren ligero, Tren suburbano y Teleférico se relacionan con aumentos en el precio de las viviendas. Una vivienda ubicada una desviación estándar más cerca, respecto de la media, de estos elementos se asocia, en promedio, con incrementos en el precio del 33%, 18% y 6% respectivamente. Estos resultados coinciden con lo encontrado por Cervero y Duncan en Santa Clara, California, y Seo et al (2014), en Arizona, en el caso del Tren Ligero; y por Strand y Vagnes (2001) sobre los beneficios en términos económicos de la cercanía a vías de tren.

En contraste, la cercanía con aeropuertos, biciestacionamientos, rutas de transporte concesionado y estaciones de BRT se relaciona con disminuciones en el precio de las viviendas. Una reducción de una desviación estándar en la distancia a estos elementos se asocia con reducciones en el valor de -12%, -10%, -10% y -5% respectivamente. El efecto de los aeropuertos coincide con lo encontrado por Soltani et al (2024), en Australia, quienes argumentan estar relacionados con el ruido y restricciones de zonificación, mientras que la relación negativa con estaciones de BRT coincide con lo encontrado por Salon et al (2014), en Guangzhou, China, asociado a ruido y contaminación.

Los resultados muestran que la cercanía a los sistemas de transporte público ferroviario se asocia con los mayores incrementos en el precio de las viviendas, mientras que la cercanía a los aeropuertos, los biciestacionamientos y las rutas de transporte concesionado con los mayores decrementos. Las relaciones estimadas para estos elementos

Mapa 8. Accesibilidad Tren Ligero



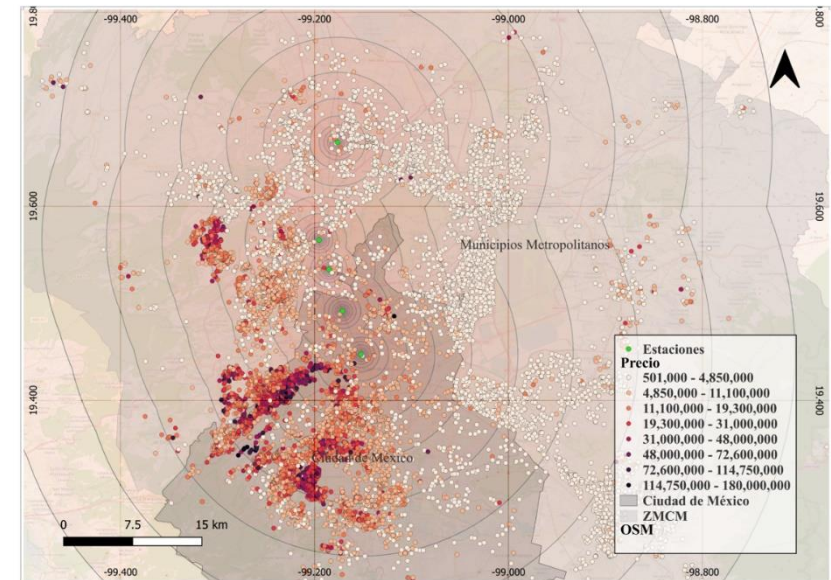
Fuente: Elaboración propia

podrían asociarse con los beneficios en términos de accesibilidad que representan estas infraestructuras, o con las externalidades negativas que pudieran generar o restricciones de

zonificación, como argumentan algunos autores. Sin embargo, también pueden estar relacionadas con la localización de estos elementos en el territorio.

Por ejemplo, en el caso del Tren Ligero, si miramos su ubicación (Mapa 8), vemos que se concentra al sur de la ciudad, en zonas cuyos valores residenciales son medios-altos. Esto podría sugerir que el efecto de esta infraestructura refleja no solo el beneficio de la accesibilidad que representa su cercanía, sino también su localización espacial en zonas de viviendas de alto valor.

Mapa 9. Accesibilidad Tren Suburbano

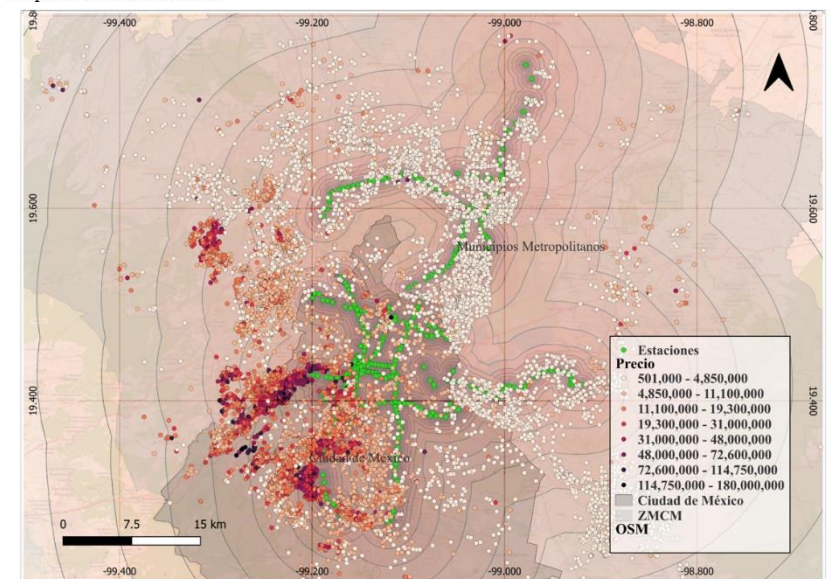


Fuente: Elaboración propia

Por su parte, al analizar la ubicación del Tren Suburbano, (Mapa 9), se puede observar que algunas estaciones se localizan en lugares centrales, donde los precios son medios-altos, mientras que otras se localizan en la periferia norte de la metrópoli, donde los precios son bajos. En este caso, la presencia de viviendas heterogéneas alrededor de las estaciones podría sugerir un efecto positivo asociado a los beneficios de la accesibilidad que representa este transporte, más que a su concentración espacial en zonas de alto valor.

En contraste, si observamos la ubicación de las estaciones de BRT (Mapa 10), observamos que muchas se localizan en las zonas periféricas al norte hasta Tecámac y al Oriente hasta

Mapa 10. Accesibilidad BRT



Fuente: Elaboración propia

Chimalhuacán; y por tal motivo hay un gran número de viviendas con valores medios-bajos alrededor de las estaciones. Esto podría sugerir que el efecto negativo que el modelo captura se asocia, no a externalidades negativas que generan esta infraestructura, sino a la presencia de viviendas de bajo precio.

Dada la estructura fuertemente polarizada de la metrópoli, estos ejemplos invitan a tomar con precaución los resultados, ya que, si bien el modelo permite estimar las dependencias estructurales en los precios, no permite diferenciar si los efectos observados se deben a los beneficios o externalidades generadas o simplemente a su ubicación territorial. No obstante, estos hallazgos podrían reforzar la evidencia de una relación entre la concentración histórica de inversión pública en algunas zonas de la metrópoli y la concentración de vivienda de precios altos, principalmente en el sur poniente de la ciudad.

A diferencia de las infraestructuras de movilidad, que se localizan de forma concentrada dentro del tejido metropolitano, los equipamientos urbanos presentan una distribución mucho más extensa y dispersa, con presencia en prácticamente todo el territorio. Esta mayor cobertura territorial implica que su efecto sobre el valor de la vivienda no está asociado a lógicas de concentración espacial ni a zonas de alto o bajo valor, como ocurre con la infraestructura de movilidad. En estas, es más probable que el efecto estimado esté mediado por las externalidades, positivas o negativas, que los equipamientos generan en su entorno inmediato.

La cercanía a escuelas de educación superior, equipamientos culturales y oficinas de gobierno se asocia con incrementos en el precio. Una reducción en la distancia a estos elementos, en una magnitud de una desviación estándar, se asocia con un incremento en el precio de 7%, 3% y 3% respectivamente. En contraste, la cercanía a los hospitales generales, de especialidad, escuelas de educación básica, parques y equipamientos recreativos se asocian con efectos negativos en el precio. Una reducción en la distancia se asocia con una disminución en el precio de -4%, -2%, -4% y -5%, respectivamente.

En este caso, las universidades son los equipamientos más valorizados por el mercado de vivienda en la ZMCM, confirmando su posición simbólica positiva en el mercado. Por su parte, la cercanía a los parques y equipamientos recreativos aparecen como los menos deseables en el mercado, posiblemente debido a externalidades negativas, provocadas por la actividad económica o afluencia que genera en su entorno, en términos de tráfico, ruido, contaminación o inseguridad. Este resultado

contradice algunos trabajos realizados en otras latitudes que estiman una relación positiva con su cercanía (Jim et al (2009); Wen et al (2017)). Sin embargo, no hay un consenso sobre el efecto de este tipo de elementos urbanos.

En conjunto, los resultados muestran que el mercado de vivienda presenta una fuerte dependencia espacial, donde los precios de las viviendas están interconectados y se influyen mutuamente. El área construida aparece como el principal determinante del precio, mientras que la valoración de elementos urbanos y de infraestructura varía de manera significativa. Los sistemas de transporte público ferroviario, las universidades y los equipamientos culturales son los elementos urbanos más valorizados por el mercado, mientras que infraestructuras como los aeropuertos, las rutas de transporte concesionado, las estaciones de BRT y ciertos equipamientos como hospitales, escuelas de educación básica y parques tienden a depreciar el valor de la vivienda.

No obstante, estos efectos deben interpretarse con cautela, pues podrían reflejar tanto los beneficios o externalidades asociados a estos elementos, como su localización dentro de una metrópoli fuertemente polarizada, donde la concentración histórica de inversión pública en ciertas zonas influye decisivamente en los patrones de valorización inmobiliaria. Además, los efectos promedio estimados podrían ocultar dinámicas más complejas y matizadas. Esto se debe a que el modelo capta únicamente relaciones lineales, aditivas y de carácter global. Esta limitación implica que posibles efectos no lineales o valoraciones diferenciadas entre distintas áreas o segmentos del mercado queden sin identificar, lo cual puede llevar a interpretaciones parciales o incompletas de los resultados.

Por este motivo, y con un espíritu exploratorio, se estimó un modelo de aprendizaje automático al cual se le aplicaron herramientas agnósticas de interpretabilidad. El propósito es analizar tanto la dependencia parcial de cada variable sobre el precio, como su efecto en cada observación particular. Para esto, se utilizaron dos de las herramientas descritas en la revisión teórica, que son: los efectos locales acumulados ALE, para la dependencia parcial; y los valores SHAP, para las contribuciones locales. Esta estrategia permite profundizar en la comprensión de las relaciones complejas y heterogéneas que podrían estar presentes en el mercado inmobiliario, complementando así las evidencias obtenidas a partir del modelo SAR y aportando un análisis más detallado y flexible de los determinantes del valor de la vivienda.

Análisis exploratorio de relaciones complejas

Las variables empleadas en el modelo de aprendizaje automático son las mismas que las del modelo paramétrico. Sin embargo, en lugar de valores estandarizados, se emplearon valores absolutos para facilitar la interpretación de los resultados en un contexto de relaciones no lineales. Cabe señalar que esta decisión no altera la consistencia de los hallazgos.

El algoritmo de aprendizaje automático se seleccionó mediante la librería PyCaret de Python, la cual permite probar diferentes modelos, ajustar automáticamente sus hiperparámetros y elegir el más adecuado en función de las métricas de comparación definidas. Dado que estos modelos se optimizan según su precisión predictiva, se utilizó como métrica de comparación la raíz del error cuadrático medio (RMSE), ampliamente empleada para medir la calidad de las predicciones de variables numéricas (Yazdani, 2021; MacBride, 2024; Y. Wu y Han, 2025). El modelo seleccionado resultó ser un regresor de árboles extra (Extra Trees Regressor), que alcanzó un R^2 de 0.9023 y un RMSE de 0.372, superando en términos predictivos al modelo SAR.

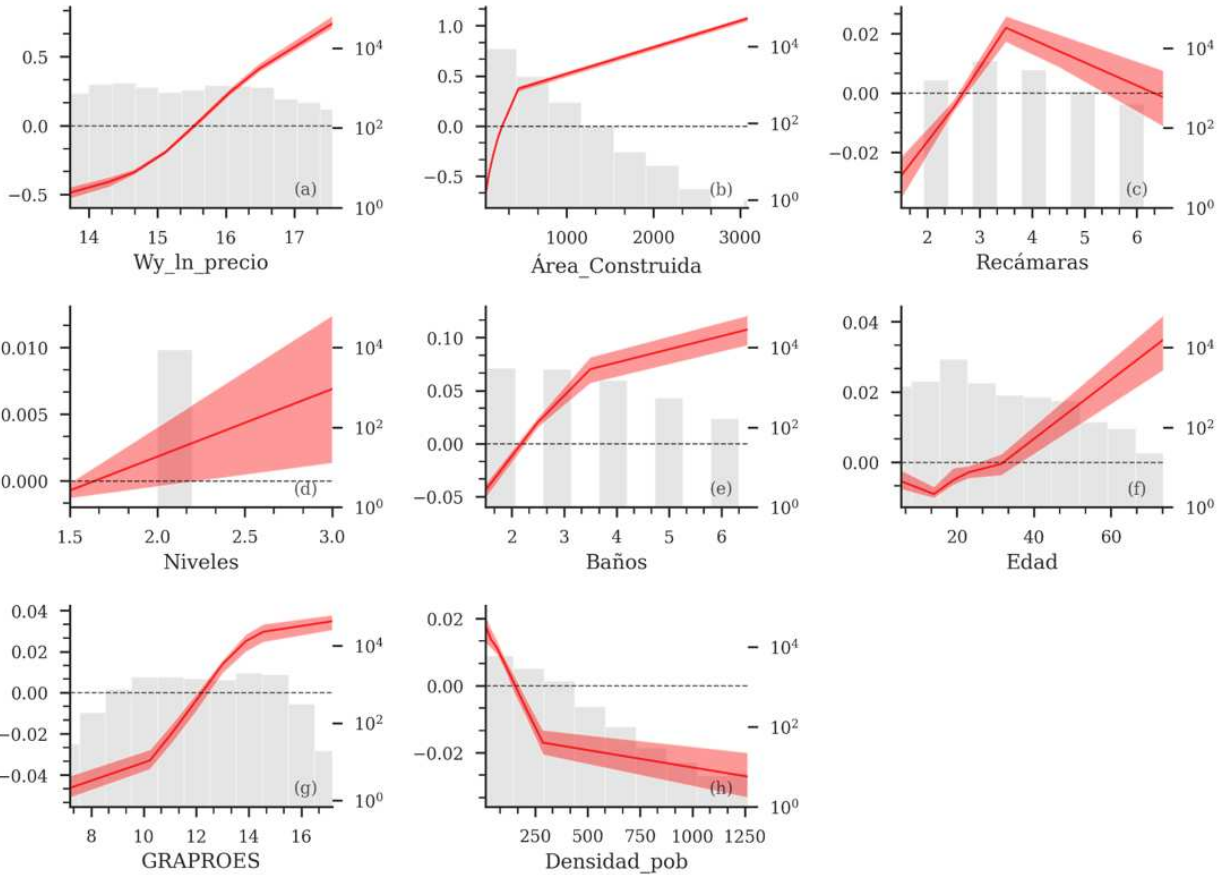
Las gráficas de dependencia parcial calculadas con ALE se leen de la siguiente manera: en el eje X se representa el rango de valores de la variable independiente. En el eje Y izquierdo se muestra el efecto promedio estimado que esos valores ejercen sobre la predicción del modelo, en este caso, sobre el \ln_precio . Es decir, refleja cuanto se espera que cambie \ln_precio cuando varía la variable independiente, manteniendo constantes las demás variables del modelo.

La línea roja indica la magnitud y el sentido de este efecto, y permite identificar relaciones no lineales si la forma de la curva se desvía de una línea recta. Una pendiente positiva implica que, al aumentar la variable independiente, tiende a aumentar el \ln_precio , y viceversa. El área sombreada en rojo representa un intervalo de confianza, calculado mediante Bootstrap, que refleja la certidumbre de la estimación. Al fondo, en color gris, se presenta el histograma de la variable independiente, cuyos valores se leen en el eje y derecho, lo que ayuda a interpretar en qué rangos hay más datos observados. Dado que se trata de una herramienta gráfica, la interpretación de los resultados se realiza de manera visual.

Las variables estructurales y de localización se presentan en la Gráfica 7. En términos generales, los efectos estimados por el algoritmo coinciden en su dirección con los del modelo SAR. El área construida y la variable espacialmente rezagada muestran los efectos positivos más fuertes, lo cual

se refleja tanto en la magnitud del eje Y izquierdo como en la certeza de sus estimaciones, dado el intervalo de confianza. Sin embargo, en ambos casos el efecto no es estrictamente lineal.

Gráfica 7. ALE: Variables de estructura y localización



Fuente: Elaboración propia

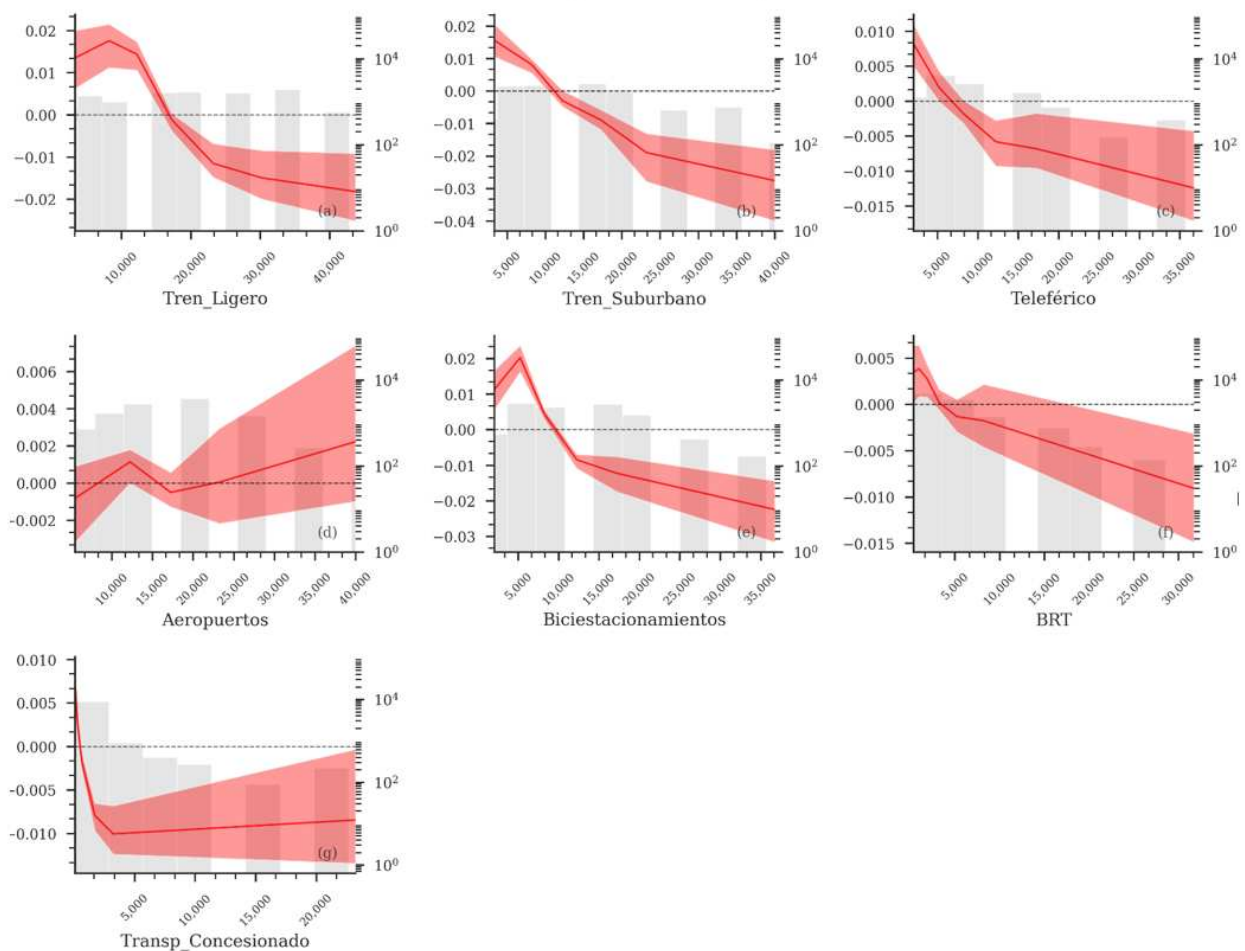
Para la variable rezagada se observa una pendiente positiva muy pronunciada, con una relación cercana a una línea recta, mientras que, para el área construida, se observa un efecto positivo que disminuye su magnitud a partir de aproximadamente 450m² de construcción. Es decir, más allá de este tamaño, el incremento adicional en área construida tiene un efecto menor sobre el precio. Este resultado concuerda con lo encontrado por Potrawa y Tetereva (2022), quienes encontraron que el efecto de cada metro adicional no tiene una relación constante y progresiva con el precio.

Por su parte, el número de recámaras muestra una relación no lineal ya que su efecto cambia de dirección. Un aumento en el número de recámaras hasta llegar a tres se asocia con incrementos en el valor de la vivienda, pero superando ese umbral, continuar añadiendo recámaras se vincula con

disminuciones en el valor. El número de baños, número de pisos y la edad del inmueble presentan una relación positiva con el precio, coincidiendo en términos generales con lo estimado por el modelo espacial. Sin embargo, tanto en el caso de los baños como en la edad del inmueble, esta relación no es lineal.

En particular, la edad muestra inicialmente un efecto negativo sobre el precio hasta aproximadamente los 15 años; superando ese punto, el efecto se revierte y se vuelve positivo. Este patrón coincide con lo observado por Lorenz et al, (2022) en la ciudad de Frankfurt, quienes encontraron una relación en forma de “U”; o por Caridad et al., (2008), en la ciudad de Porto quienes encontraron una relación negativa entre la edad de la propiedad y su valor para los primeros 16-20 años. En estos análisis se ha asociado este efecto con el valor patrimonial e histórico de las edificaciones. Aquí, es posible que la misma lógica aplique.

Gráfica 8. ALE: Variables del entorno urbano (infraestructura)



Fuente: Elaboración propia

Las variables de localización como densidad de población y grado promedio de escolaridad también muestran, en términos generales, coincidencia con las estimaciones del modelo SAR. No obstante, al igual que en los casos anteriores, sus efectos no parecen ser estrictamente lineales.

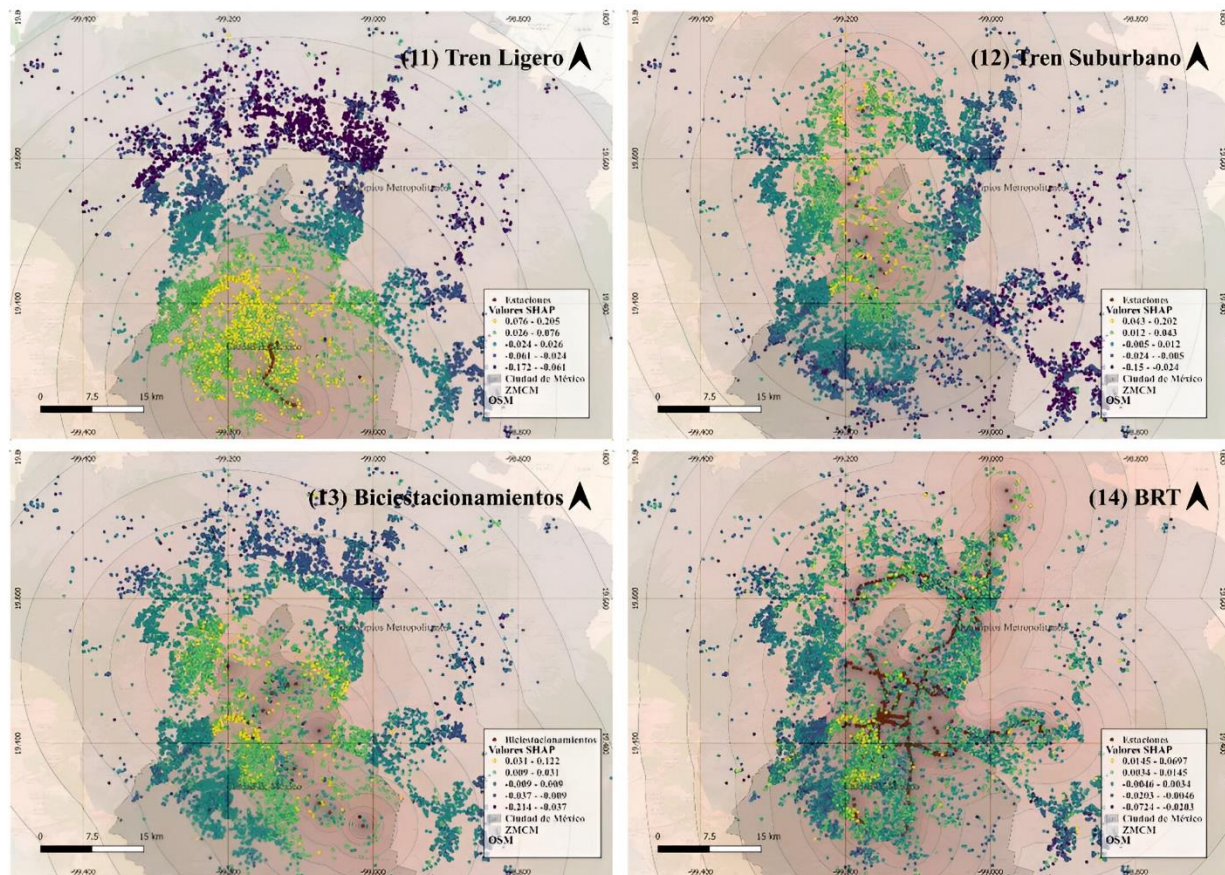
En cuanto a la infraestructura de movilidad, las relaciones identificadas por el algoritmo se presentan en la Gráfica 8. Por un lado, se observan efectos positivos asociados con la cercanía a estaciones de Tren Ligero, Tren Suburbano y Teleférico lo cual coincide con las estimaciones del modelo SAR. De manera similar, la proximidad a aeropuertos muestra un efecto negativo, también en línea con dicho modelo. Sin embargo, las relaciones capturadas por el algoritmo no son constantes, sino que revelan patrones más complejos y matices adicionales. En el caso del Tren ligero la gráfica revela un pico de valorización a los 9km, después de los cuales, alejarse de esta infraestructura disminuye el valor. En el caso del Teleférico y del Tren suburbano se aprecia una relación cercana a una línea para los primeros 10 y 15 km respectivamente, después de estas distancias el efecto en ambos es más incierto dado el intervalo de confianza

Por otro lado, el algoritmo también identificó efectos distintos a los estimados por el modelo SAR en relación con la cercanía a biciestacionamientos, estaciones de BRT y rutas de transporte concesionado. Para los biciestacionamientos, la cercanía entre 0 y 5 km se asocia con reducciones en el valor de los inmuebles, posiblemente por la presencia de externalidades negativas, efecto que captura el modelo SAR. Sin embargo, a partir de los 5km, la relación se invierte y alejarse se relaciona con precios más bajos. Los resultados podrían indicar que los biciestacionamientos generan externalidades negativas en un radio de hasta 5km, pero, dado que suelen estar ubicados en zonas centrales, también reflejan la concentración de viviendas de alto valor, lo que explicaría que alejarse posteriormente se vincule con caídas en los precios.

En cuanto al BRT, su cercanía muestra un efecto positivo sobre el precio, siendo los primeros 4km los de mayor intensidad. Algo similar ocurre con el transporte concesionado, cuya cercanía presenta una fuerte relación positiva con el precio los primeros 2.5km, con un alto nivel de certidumbre en el efecto, aunque más allá de esa distancia el efecto se vuelve menos claro.

Con la intención de profundizar en estos hallazgos, la Figura 2 muestra las contribuciones individuales de la cercanía a cuatro infraestructuras distintas. Estos mapas permiten visualizar, mediante valores SHAP, la magnitud y dirección del efecto que la proximidad a cada elemento ejerce sobre el precio de cada vivienda, lo que permite enriquecer y matizar el análisis.

Figura 2. Mapas de valores SHAP (11-14)



Fuente: Elaboración propia

En general, los patrones espaciales mostrados en los mapas coinciden con los efectos estimados mediante las gráficas ALE. Por ejemplo, el mayor efecto positivo de la cercanía al Tren Ligero se percibe en un radio de 9km, destacando viviendas ubicadas en zonas como Polanco, Condesa, Roma y Coyoacán, que no son necesariamente las más próximas a las estaciones. Esto sugiere que los efectos capturados reflejan más bien la concentración de viviendas de alto valor en el poniente y sur de la ciudad, más que beneficios directos de la accesibilidad.

En contraste, el efecto del Tren Suburbano es mayor en las viviendas más cercanas, y disminuye conforme aumenta la distancia, en línea con el patrón lineal y progresivo identificado por la gráfica ALE. Aquí, el impacto positivo parece reflejar beneficios concretos de accesibilidad y no una lógica de concentración de valor inmobiliario.

En el caso de los biciestacionamientos, la relación con el precio de la vivienda es más compleja. Al igual que lo muestran las gráficas ALE, la cercanía a estos elementos, dentro de un radio de 5

km, se asocia de manera negativa con el valor de las viviendas. Este efecto es más intenso en las zonas próximas a biciestacionamientos ubicados cerca de estaciones de metro como Martín Carrera, La Villa/Basílica y La Raza. A los 5 km de distancia se observa la mayor valorización, beneficiando especialmente a viviendas ubicadas en zonas como Polanco y Condesa, al poniente, y Nezhualcóyotl y Bosques de Aragón al oriente, identificadas en color amarillo en los mapas. Más allá de los 5 km, el efecto se invierte y alejarse más disminuye el precio.

Estos resultados refuerzan la idea de que existen externalidades negativas asociadas a los biciestacionamientos, sobre todo en aquellos localizados cerca de las estaciones de metro mencionadas. No obstante, debido a la centralidad de estos elementos, alejarse más de 5 km también se relaciona con una pérdida de valor en el precio de las viviendas.

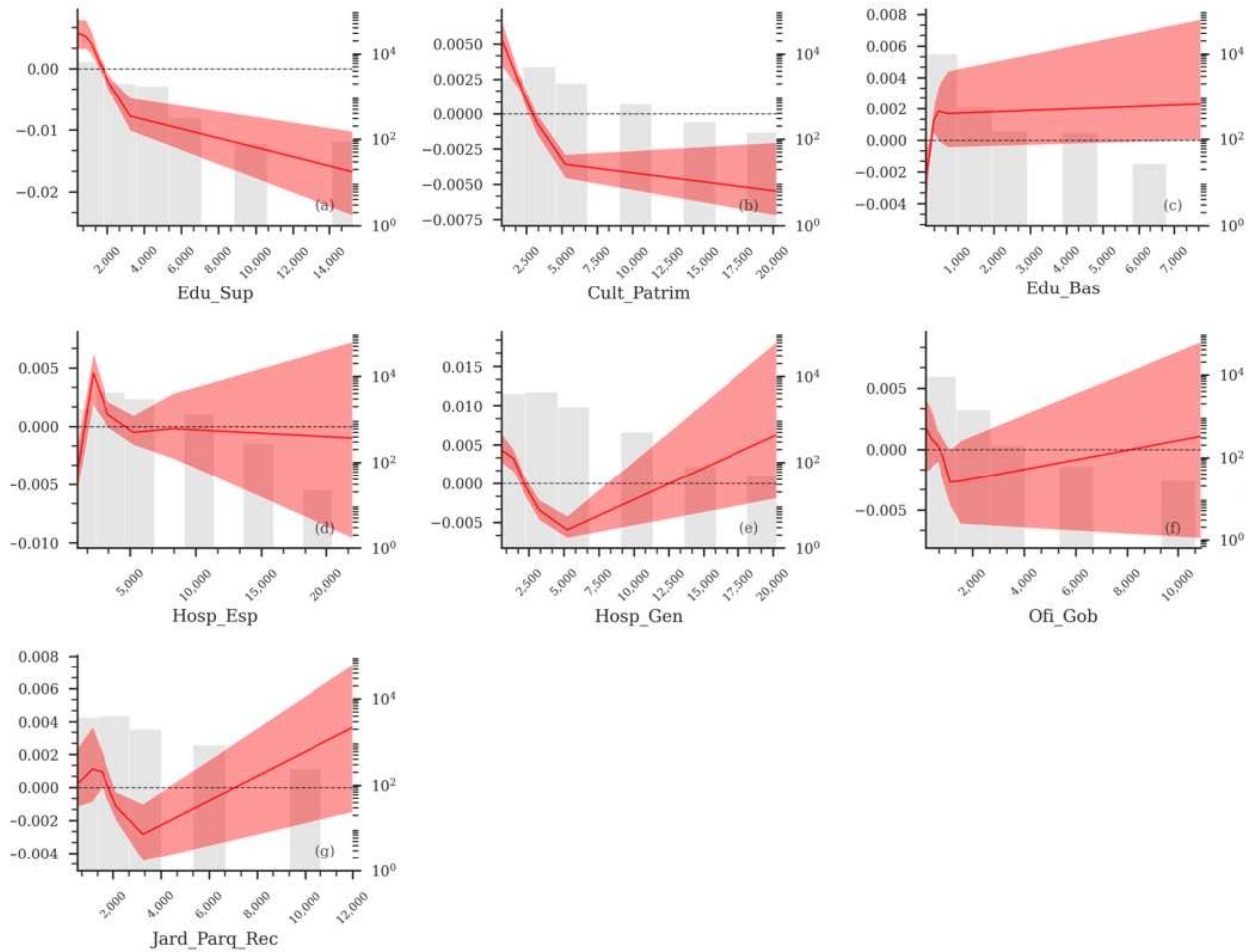
En el caso del BRT, su efecto sobre el valor de la vivienda resulta heterogéneo: se observan las mayores valorizaciones en zonas como Polanco o a lo largo de avenida Insurgentes, mientras que, en áreas periféricas, pese a la proximidad al sistema de transporte, los beneficios son menos intensos. Esto muestra que el efecto de ciertas infraestructuras varía según su localización, beneficiando en mayor medida a zonas históricamente favorecidas por la inversión urbana. Estos resultados coinciden con lo reportado por Gadziński y Radzimski (2016), Yang et al. (2020) y Wardrip (2011), quienes documentan efectos heterogéneos asociados a la ubicación de las viviendas en la ciudad. Así, se refuerza la hipótesis de que las viviendas de mayor valor concentran los mayores beneficios, debido a su localización en zonas históricamente receptoras de inversión pública.

Finalmente, la Gráfica 9 muestra los resultados ALE para los distintos equipamientos urbanos. En términos generales, estos resultados coinciden con lo estimado por el modelo espacial. La cercanía a escuelas de educación superior y equipamientos culturales se asocia con incrementos en el valor de las viviendas, mientras que la cercanía a escuelas de educación básica y hospitales de especialidad presenta un efecto negativo, en línea con lo observado en el modelo SAR.

En contraste, los efectos asociados a la cercanía de hospitales generales, oficinas de gobierno y equipamientos recreativos coinciden solo parcialmente con los del modelo espacial. Por ejemplo, en el caso de los hospitales generales, se observa un patrón no lineal: entre 0 y 5km, alejarse de ellos reduce el valor de la vivienda, pero más allá de los 5km, continuar alejándose se asocia con un aumento en el precio. Dada la certidumbre del efecto los primeros 5km, el resultado podría

sugerir que este equipamiento es valorizado positivamente en el mercado de vivienda, sin embargo, no es posible concluirlo de forma definitiva.

Gráfica 9. ALE: Variables del entorno urbano (equipamientos)



Fuente: Elaboración propia

Respecto a las oficinas de gobierno y los equipamientos recreativos, los efectos también presentan una relación no lineal con el precio de la vivienda. En ambos casos, existe una asociación positiva los primeros 2 y 3 km respectivamente, pero esta se revierte a partir de esa distancia. Sin embargo, los intervalos de confianza se amplían considerablemente más allá de estas distancias, lo que indica un alto grado de incertidumbre sobre la dirección y magnitud de estos efectos.

En síntesis, los hallazgos obtenidos a partir del modelo de aprendizaje automático y de los modelos agnósticos de interpretabilidad muestran, en términos generales, una alta coincidencia con las relaciones estimadas por el modelo SAR. Tal como se resume en el cuadro comparativo (Cuadro

15), de las 22 variables con significancia estadística en el modelo SAR, el efecto de 13 variables coincide en sentido, mientras que 7 variables coinciden parcialmente y otras 2 presentan efectos contrarios. Más allá de estas coincidencias, los modelos de interpretabilidad aportan matices valiosos, revelando detalles sobre la forma funcional de las relaciones y patrones no lineales que enriquecen el análisis de cómo la inversión pública se capitaliza en los precios de la vivienda.

Estos resultados sugieren la necesidad de interpretar con cautela las relaciones identificadas y de profundizar, en investigaciones futuras, en las causas que explican las discrepancias observadas. Además, la información obtenida sobre la forma funcional de las relaciones ofrece un insumo relevante para especificar de manera más precisa futuros modelos paramétricos y así mejorar su capacidad explicativa.

Cuadro 15. Comparativa ente resultados

Variable	Modelo paramétrico (SAR)			Aprendizaje automático		Coincidencia
	Coefficiente	Dirección de efecto	Efecto total	ALE	SHAP	
W_ln_precio	0.48301***	↗	-			✓
Recámaras	0.0443***	↗	9%			≈
Niveles	0.01633***	↗	3%			✓
Baños	0.11591***	↗	22%			✓
Área construida	0.40655***	↗	79%			✓
Edad	0.02437***	↗	5%			≈
Graproes	0.06977***	↗	13%			✓
Densidad Pob	-0.06298***	↘	-12%			✓

Fuente: Elaboración propia

Cuadro 15. Comparativa ente resultados (continuación)

Aeropuertos	0.02612***	↘	-5%			☑
Biciestacionamientos	0.05157***	↘	-10%			≈
BRT	0.02809***	↘	-5%			⊗
Teleférico	-0.02891***	↗	6%			☑
Tren ligero	-0.16968***	↗	33%			☑
Tren suburbano	-0.09084***	↗	18%			☑
Transporte concesionado	0.05003***	↘	-10%			⊗
Hospital general	0.01864***	↘	-4%			≈
Hospital especialidad	0.01131*	↘	-2%			≈
Educación básica	0.01851**	↘	-4%			☑
Educación superior	-0.03785***	↗	7%			☑
Cultura y patrimonio	-0.01665**	↗	3%			☑
Jar_Parq_Rec	0.02641***	↘	-5%			≈
Oficinas de gobierno	-0.01291*	↗	3%			≈

*** si $p \leq 0.01$ ** si $0.01 < p \leq 0.05$ * si $0.05 < p \leq 0.10$

Fuente: Elaboración propia

Conclusiones

El presente análisis ofrece una mirada integral sobre la conformación del precio en el mercado de vivienda de la ZMCM, revelando un entramado complejo de factores donde confluyen desigualdades estructurales históricas, dependencia espacial y patrones heterogéneos de valorización. Los hallazgos muestran que el valor de la vivienda no depende únicamente de atributos físicos o de la proximidad a equipamientos, sino también de su inserción histórica y territorial, evidenciando dinámicas urbanas que requieren políticas públicas sensibles a las desigualdades sociales.

Los resultados confirman que existe una fuerte dependencia espacial en los precios de las viviendas, lo que indica que los valores inmobiliarios no se determinan de manera aislada, sino que se influyen mutuamente, consolidando jerarquías urbanas y sociales. El sur-poniente de la ciudad concentra propiedades de muy alto precio, mientras que las extensas periferias norte y oriente viviendas de menor valor. El tamaño emerge como el factor más influyente en el valor de la vivienda, algo intuitivamente esperable. Sin embargo, las gráficas ALE revelaron que el efecto es más intenso hasta los 450m², y después cada metro adicional aporta menor valor relativo. Por su parte, las variables de calidad también contribuyen de forma significativa, con incrementos de hasta 16% en el valor dada la presencia de amenidades como terrazas.

La edad del inmueble muestra un efecto complejo. Aunque parece existir una tendencia a la valorización de edificios antiguos en el mercado de vivienda, los primeros 15 años reducen el precio de los inmuebles. El efecto positivo puede estar relacionado con valorizaciones históricas del suelo o con su importancia en términos de patrimonio histórico, como sucede en otras ciudades del mundo. Las variables de localización también muestran efectos importantes. El incremento en el promedio de años de escolaridad incrementa el valor 13%, mientras que una mayor densidad de población lo disminuye hasta en 12%.

Los resultados de los elementos urbanos reflejan fenómenos diferenciados. Por un lado, algunas infraestructuras y equipamientos muestran relaciones positivas con los precios, explicadas por los beneficios que brinda su cercanía en términos de accesibilidad, como es el caso del Tren Suburbano; o por ser referentes simbólicos positivos como el caso de las universidades y los equipamientos culturales. Por otro lado, algunos elementos se relacionan negativamente, como las escuelas de educación básica, hospitales de especialidad, biciestacionamientos y aeropuertos, lo

que sugiere que sus efectos están medidos por externalidades negativas como ruido, tráfico, contaminación, restricciones de zonificación o percepción de inseguridad.

Particularmente, en el caso de los biciestacionamientos, los resultados mostraron un efecto no lineal: una asociación negativa dentro de un radio de 5km, y positiva superando esa distancia. Este efecto se podría asociar a externalidades negativas presentes dentro del radio definido, mientras que el cambio de sentido en la relación podría reflejar la ubicación, aun céntrica, de esta infraestructura. Este resultado sugiere que las relaciones nos siguen una estructura lineal y aditiva, sino que presentan efectos más complejos que deben ser considerados en las estimaciones.

Adicionalmente, ciertas infraestructuras, como el Tren ligero y estaciones de BRT, muestran un fuerte componente de dependencia espacial y efectos heterogéneos en el territorio. Los resultados obtenidos con los modelos agnósticos de interpretabilidad sugieren una mayor valorización asociada a lugares donde los precios ya son altos. Colonias como Lomas de Chapultepec, Polanco, Jardines del Pedregal, Condesa, Roma, Coyoacán, Del Valle y San Ángel, tradicionalmente asociadas a clases acomodadas, que concentran la mayoría de las viviendas de alto valor y que han sido beneficiadas por una mayor acumulación de inversión pública, se muestran como las más beneficiadas por estas obras de infraestructura. Este fenómeno refuerza la hipótesis de que la magnitud de la capitalización de la inversión pública depende en gran medida de su localización y, particularmente, de su proximidad a zonas donde los precios ya son elevados y donde históricamente se ha concentrado la inversión pública.

Estos resultados sugieren que el efecto de los elementos urbanos en el valor de una vivienda no depende únicamente de la proximidad física, sino también de cómo dicha infraestructura se articula con otras condiciones urbanas. Los efectos diferenciados en función de la localización pueden interpretarse como expresiones de desigualdad urbana. Algunas obras refuerzan centralidades consolidadas, mientras que otras, aunque funcionales, no consiguen superar estigmas territoriales ni condiciones estructurales que limitan su capacidad para valorizar el entorno.

Cabe mencionar que, debido a la acumulación de inversión, especialmente en infraestructura de movilidad, los efectos tienden a superponerse, lo que dificulta la identificación de externalidades individuales asociadas a cada elemento urbano. Esto se refleja en los altos niveles de colinealidad encontrada principalmente en las variables de infraestructura de movilidad.

Desde una perspectiva fiscal, estos hallazgos respaldan la viabilidad de esquemas tributarios basados en la concentración territorial del valor, como el impuesto predial progresivo o los mecanismos de captura de plusvalías. Identificar espacialmente dónde se localiza el valor y cómo afecta al precio la cercanía a los distintos elementos urbanos permite justificar una mayor carga tributaria en aquellas zonas que concentran los mayores beneficios públicos. Este análisis podría mejorar el diseño y recaudación de impuestos como el predial, incrementando los ingresos fiscales y favoreciendo su redistribución hacia otras zonas. Así, se contribuiría a reducir las brechas históricas de desigualdad que caracterizan a la metrópoli.

En conjunto, estos resultados muestran que la valorización del entorno urbano no puede entenderse únicamente desde sus funciones o servicios, sino que está profundamente condicionada por el contexto espacial e histórico en que se inserta, así como por patrones de dependencia espacial que refuerzan desigualdades preexistentes en la ZMCM.

Este trabajo contribuye al entendimiento del mercado de vivienda desde una perspectiva integrada, ya que combina modelos econométricos espaciales con herramientas agnósticas de interpretabilidad. Esta aproximación permitió capturar tanto las dependencias estructurales del precio como las relaciones no lineales y locales entre los atributos de la vivienda y su valor. En este sentido, se propone una metodología replicable para analizar mercados de vivienda urbanos, que puede ser útil no solo para investigadores, sino también para autoridades locales interesadas en diseñar instrumentos de gestión territorial más sensibles a las desigualdades urbanas. Además, el análisis ofrece un punto de partida para el monitoreo de cambios en los precios derivados de nuevas infraestructuras. Esto permitiría, en el futuro, identificar de forma más precisa los efectos individuales de las obras públicas.

No obstante, el análisis también enfrenta ciertas limitaciones. La información utilizada refleja un corte estático en el tiempo, lo cual impide capturar dinámicas temporales en la formación de precios o los efectos de intervenciones recientes. También, presenta cierto sesgo hacia segmentos de medio y alto valor, por lo que es necesario integrar información de otras fuentes de datos para ampliar los alcances de los análisis.

Además, la identificación de efectos individuales se ve dificultada por la fuerte concentración de infraestructura, lo que motiva la necesidad de complementar estos enfoques con métodos cuasi experimentales o longitudinales. Futuros estudios podrían integrar datos panel, evaluar

intervenciones específicas mediante diseños de tipo diferencias en diferencias o profundizar en la valoración de externalidades con modelos causales. Estos enfoques permitirían estimar el impacto específico de nuevas infraestructuras, distinguir efectos temporales y evaluar desigualdades con mayor precisión. Estas líneas de investigación permitirían avanzar hacia una evaluación más precisa del impacto urbano de las infraestructuras y reforzar el vínculo entre investigación y políticas públicas.

Bibliografía:

- Aguayo Ayala, A. (2016). Nuevo Polanco: renovación urbana, segregación y gentrificación en la Ciudad de México. *Iztapalapa. Revista de Ciencias Sociales y Humanidades*, 37(80), 101–123. <https://doi.org/10.28928/revistaiztapalapa/802016/atc4/aguayoayalaa>
- Alonso, W. (1964). *Location and Land Use: Toward a General Theory of Land Rent*. Harvard University Press.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Springer.
- Anselin, L. (2002). Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3), 247–267. <https://doi.org/10.1111/j.1574-0862.2002.tb00120.x>
- Balchin, P. N., Isaac, D., & Chen, J. (2000). *Urban economics: A global perspective*. Palgrave Macmillan.
- Basu, S., & Thibodeau, T. G. (1998). Analysis of spatial autocorrelation in house prices. *The Journal of Real Estate Finance and Economics*, 17(1), 61–85.
- Bayón, M. C. (2008). Desigualdad y procesos de exclusión social. Concentración socioespacial de desventajas en el Gran Buenos Aires y la Ciudad de México. *Estudios Demográficos y Urbanos*, 23(1), 123–150. <https://doi.org/10.24201/edu.v23i1.1305>
- Bluestone, B., Stevenson, M. H., & Williams, R. (2008). *The urban experience: Economics, society, and public policy*. Oxford University Press.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brown, J., & Rosen, H. (1982). On the Estimation of Structural Hedonic Price Models (NBER Working Paper No. t0018). <https://doi.org/10.3386/t0018>
- Cervero, R., & Duncan, M. (2002). Transit's value-added effects: Light and commuter rail services and commercial land values. *Transportation Research Record: Journal of the Transportation Research Board*, 1805(1), 8–15. <https://doi.org/10.3141/1805-02>
- Chen R., et al. (2024). Evaluation of green space influence on housing prices using machine learning and urban visual intelligence. *Cities*, 158, 105661. <https://doi.org/10.1016/j.cities.2024.105661>
- Chen Y., et al. (2023). Disparities in affecting factors of housing price: A machine learning approach to the effects of housing status, public transit, and density factors on single-family housing price. *Cities*, 140, 103458. <https://doi.org/10.1016/j.cities.2023.103458>
- Chin, T.-L., & Chau, K. W. (2002). A critical review of literature on the hedonic price model. *International Journal of Housing Science and Its Applications*, 26(1), 57–70.
- Cisneros Sosa, A. (1993). *La ciudad que construimos: Registro de la expansión de la ciudad de México, 1920–1976*. Universidad Autónoma Metropolitana, Unidad Iztapalapa.

- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403), 596–610. <https://doi.org/10.2307/2289282>
- Cohen, J. P., & Coughlin, C. C. (2008). Spatial hedonic models of airport noise, proximity, and housing prices. *Journal of Regional Science*, 48(5), 859–878. <https://doi.org/10.1111/j.1467-9787.2008.00569.x>
- Connolly, P. (2018). ¿Qué se gobierna en materia de transporte y movilidad? El caso de la Ciudad de México. En P. Le Gales & V. Ugalde (Eds.), *Gobernando la Ciudad de México* (pp. 371–428). El Colegio de México.
- Conway, J. E. (2018). Artificial intelligence and machine learning: Current applications in real estate [Tesis de maestría, Massachusetts Institute of Technology]. DSpace@MIT. <http://hdl.handle.net/1721.1/120609>
- Delgadillo, V. (2010). Territorio, vivienda, infraestructura y transporte: el caso de la Ciudad de México y su área metropolitana. *Foro Iberoamericano y del Caribe sobre Mejores Prácticas, 5 temas selectos del hábitat latinoamericano*. ONU-Hábitat. https://www.academia.edu/4961206/Territorio_Vivienda_e_Infraestructura_Ciudad_de_M%C3%A9xico
- Du, Z., Feng, H., & Arbuckle, J. G. (2025). Exploring the complementarity between traditional econometric methods and machine learning: An application to adoption and disadoption of conservation practices. *Applied Economics*, 1–16. <https://doi.org/10.1080/00036846.2025.2462792>
- Feng Lan, et al. (2018). Spatial Effects of Public Service Facilities Accessibility on Housing Prices: A Case Study of Xi'an, China. *Sustainability*, 10(9), 1439. <https://doi.org/10.3390/su10091439>
- Fernández Durán, L., Llorca Ponce, A., Valero Cubas, S., & Botti Navarro, V. J. (2012). Incidencia de la localización en el precio de la vivienda a través de un modelo de red neuronal artificial: Una aplicación a la ciudad de Valencia. *CT: Catastro*, 74, 7–25. <https://riunet.upv.es/handle/10251/55336>
- Flores, S., & Caracheo, C. (2021). Perfil de la desigualdad urbana a partir de la distribución de la inversión en infraestructura, un análisis interescalar en la Ciudad de México. UNAM.
- Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Exploring, modelling and predicting spatiotemporal variations in house prices. *The Annals of Regional Science*, 54(2), 417–436. <https://doi.org/10.1007/s00168-015-0660-9>
- Francke, M., & Van De Minne, A. (2024). Combining machine learning and econometrics: Application to commercial real estate prices. *Real Estate Economics*, 52(5), 1308–1339. <https://doi.org/10.1111/1540-6229.12483>

- Garrido Rodríguez, L. A., Suárez Lastra, M., Universidad, N. A. de M., Universidad, N. A. de M., & Universidad, N. A. de M. (2023). Mercado inmobiliario, precio de vivienda y segregación socio espacial en la zona metropolitana de la Ciudad de México, 1990–2019.
- Greenstone, M. (2017). The Continuing Impact of Sherwin Rosen’s “Hedonic Prices and Implicit Markets.” *Journal of Political Economy*, 125(6), 1891–1912.
- Griliches, Z. (1971). Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change. En *Price Indexes and Quality Change: Studies in New Methods of Measurement* (pp. 55–87). Harvard University Press.
- Griliches, Z. (2013). *Price Indexes and Quality Change: Studies in New Methods of Measurement*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674592582>
- Harrison, D., Jr., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5(1), 81–102. [https://doi.org/10.1016/0095-0696\(78\)90006-2](https://doi.org/10.1016/0095-0696(78)90006-2)
- Heng, C., Zhang, X., & Li, Y. (2018). Exploring the Spillover Effects of Urban Renewal on Local House Prices Using Multi-Source Data and Machine Learning: The Case of Shenzhen, China. *Land*, 11(9), 1439. <https://doi.org/10.3390/land11091439>
- Hirsch, W. Z. (1973). *Urban Economic Analysis*. McGraw-Hill.
- INEGI. (2018). *Delimitación de las zonas metropolitanas de México 2015 [Cobertura 2015]*. Aguascalientes, México: Instituto Nacional de Estadística y Geografía.
- INEGI. (2020). *Censo de Población y Vivienda 2020*. Fecha de realización: 2-27 de marzo de 2020. Aguascalientes, México: Instituto Nacional de Estadística y Geografía.
- INEGI. (2021). *Encuesta Nacional de Vivienda (ENVI) 2020: diseño conceptual*. Aguascalientes, México: Instituto Nacional de Estadística y Geografía.
- Iskhakov, F., Rust, J., & Schjerning, B. (2020). Machine learning and structural econometrics: Contrasts and synergies. *The Econometrics Journal*, 23(3), S81–S124. <https://doi.org/10.1093/ectj/utaa019>
- Jin, C., Park, S., Ha, H. J., Lee, J., Kim, J., Hutchenreuther, J., & Nara, A. (2023). Predicting households’ residential mobility trajectories with geographically localized interpretable model-agnostic explanation (GLIME). *International Journal of Geographical Information Science*, 37(12), 2597–2619. <https://doi.org/10.1080/13658816.2023.2264921>
- Kain, J. F., & Quigley, J. M. (1975). *Housing Markets and Racial Discrimination: A Microeconomic Analysis*. National Bureau of Economic Research.
- Kauko, T., & D’Amato, M. (Eds.). (2008). *Mass appraisal methods: An international perspective for property valuers*. Wiley-Blackwell.

- Kilpatrick, J. A., Throupe, R., Carruthers, J. I., & Krause, A. (2007). The impact of transit corridors on residential property values. *Journal of Real Estate Research*, 29(3), 303–320. <https://doi.org/10.2139/ssrn.1030006>
- Kirby, A. M. (1976). *Housing Market Studies: A Critical Review*. *Transactions of the Institute of British Geographers*, 1(1), 2. <https://doi.org/10.2307/621308>
- Kopczewska, K. (2022). Spatial machine learning: New opportunities for regional science. *The Annals of Regional Science*, 68(3), 713–755. <https://doi.org/10.1007/s00168-021-01101-x>
- Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2), 132–157.
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press.
- Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems*, 96, 101845.
- Liu, X., Qiao, R., Wu, Z., Yang, T., Zhang, X., Zhang, X., & Zhu, Z. (2024). Unveiling the spatially varied nonlinear effects of urban built environment on housing prices using an interpretable ensemble learning model. *Applied Geography*, 173, 103458. <https://doi.org/10.1016/j.apgeog.2024.103458>
- Lorenz et al. (2023). Interpretable machine learning for real estate market analysis. *Real Estate Economics*, 51. American Real Estate and Urban Economics Association.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. <https://doi.org/10.48550/arXiv.1705.07874>
- MacBride, C. M. (2024). A comparative analysis of machine learning methods and spatial statistical methods for areal unit Scottish property price data [Tesis de maestría, University of Glasgow].
- Márquez-López, L. (2016). Acumulación del capital inmobiliario y apropiación social del espacio público en el Paseo de la Reforma, Ciudad de México. *Economía, Sociedad y Territorio*, 16(50), 71–101. <https://doi.org/10.22136/est0502016760>
- McCluskey, W. J., & Borst, R. A. (2007). Specifying the effect of location in multivariate valuation models for residential properties: A critical evaluation from the mass appraisal perspective. *Property Management*, 25(4), 312–343. <https://doi.org/10.1108/02637470710775185>
- Mills, E. S., & Hamilton, B. W. (1984). *Urban Economics*. Scott, Foresman.
- Molnar, C. (2025). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (3.^a ed.). <https://christophm.github.io/interpretable-ml-book/>
- Montejano Escamilla, J. A., Caudillo Cos, C. A., & Cervantes Salas, M. (2018). Vivienda de interés social, segregación residencial y accesibilidad: Análisis de 121 conjuntos urbanos en el arco nororiente del Valle de México, 2001–2010. *Estudios Demográficos y Urbanos*, 33(1), 187–224.
- Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>

- Muñiz, I., Sánchez, V., & García-López, M. A. (2015). Estructura espacial y densidad de población en la ZMVM 1995–2010: evolución de un sistema urbano policéntrico. *Eure (Santiago)*, 41(122), 75–102.
- O'Sullivan, A. (2012). *Urban economics* (8^a ed.). McGraw-Hill Education.
- ONU-Hábitat. (2017). Tendencias del desarrollo urbano en México. <https://onu-habitat.org/index.php/tendencias-del-desarrollo-urbano-en-mexico>
- ONU-Hábitat. (2022). Informe Mundial de las Ciudades 2022: Visualizando el futuro de las ciudades. Programa de las Naciones Unidas para los Asentamientos Humanos.
- Petkov, M. (2020). Evaluation of spatial data's impact in mid-term room rent price through application of spatial econometrics and machine learning: Case study: Lisbon [Tesis de maestría, NOVA Information Management School]. Repositorio RUN. <https://run.unl.pt/handle/10362/93716>
- Potrawa, T., & Teterewa, A. (2022). How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. *Journal of Business Research*, 144, 50–65. <https://doi.org/10.1016/j.jbusres.2022.01.027>
- Quigley, J. M. (1985). Consumer choice of dwelling, neighborhood and public services. *Regional Science and Urban Economics*, 15(1), 41–63. [https://doi.org/10.1016/0166-0462\(85\)90008-9](https://doi.org/10.1016/0166-0462(85)90008-9)
- Ramírez de la Cruz et al. (2020). La capitalización privada de los bienes públicos. Modelos de precios hedónicos para la vivienda de la Ciudad de México. CIDE.
- Redfearn, C. L. (2009). How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets. *Regional Science and Urban Economics*, 39(3), 297–306. <https://doi.org/10.1016/j.regsciurbeco.2008.11.001>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Richardson, H. W., Vipond, J., & Furbey, R. A. (1974). Determinants of Urban House Prices. *Urban Studies*, 11(2), 189–199. <https://doi.org/10.1080/00420987420080341>
- Rico-Juan, J. R., & Taltavull de la Paz, P. (2021). Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain. *Expert Systems with Applications*, 171, 114590. <https://doi.org/10.1016/j.eswa.2021.114590>
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55. <https://doi.org/10.1086/260169>
- Rubalcava, R. M., & Schteingart, M. (2012). Ciudades divididas: Desigualdad y segregación social en México. El Colegio de México.

- Salinas Arreortua, L. A. (2016). Política de vivienda social y gestión metropolitana en la expansión de la periferia de la zona metropolitana de la Ciudad de México. *Cuadernos Geográficos*, 55(1), 217–237. <https://doi.org/10.30827/cuadgeo.v55i1.3655>
- Salon, D., Wu, J., & Shewmake, S. (2014). Impact of bus rapid transit and metro rail on property values in Guangzhou, China. *Transportation Research Record: Journal of the Transportation Research Board*, 2452(1), 36–45. <https://doi.org/10.3141/2452-05>
- Schteingart, M., & Ibarra, V. (2016). Desarrollo histórico de la ciudad y conformación de las vialidades (1950-2010). En *Desarrollo urbano-ambiental y movilidad en la Ciudad de México: Evaluación histórica, cambios recientes y políticas públicas*. El Colegio de México. <https://www.jstor.org/stable/j.ctt2050ws4.5>
- Seo, K., Golub, A., & Kuby, M. (2014). Combined impacts of highways and light rail transit on residential property values: A spatial hedonic price model for Phoenix, Arizona. *Journal of Transport Geography*, 41, 53–62. <https://doi.org/10.1016/j.jtrangeo.2014.08.003>
- Seo, K., Salon, D., Kuby, M., & Golub, A. (2019). Hedonic modeling of commercial property values: Distance decay from the links and nodes of rail and highway infrastructure. *Transportation*, 46(3), 859–882. <https://doi.org/10.1007/s11116-018-9861-z>
- Sobrino, J. (2014). Precios de la vivienda y submercados en la Ciudad de México: Una evaluación hedónica. *Estudios Económicos de El Colegio de México*, 57–84. <https://doi.org/10.24201/ee.v29i1.74>
- Soltani, A., Zali, N., Aghajani, H., Hashemzadeh, F., Rahimi, A., & Heydari, M. (2024). The nexus between transportation infrastructure and housing prices in metropolitan regions. *Journal of Housing and the Built Environment*, 39(2), 787–812. <https://doi.org/10.1007/s10901-023-10085-3>
- Straszheim, M. R. (1975). *An econometric analysis of the urban housing market*. National Bureau of Economic Research.
- Tamaris Turizo, I. D. (2024). *El efecto de la localización sobre el valor de la vivienda: Los casos de Santa Marta y Atenas* [Tesis doctoral, Universidad de Granada]. Repositorio Institucional de la Universidad de Granada. <https://hdl.handle.net/10481/92833>
- Taylor, L. O. (2008). Theoretical Foundations and Empirical Developments in Hedonic Modeling. En A. Baranzini, J. Ramirez, C. Schaerer, & P. Thalmann (Eds.), *Hedonic Methods in Housing Markets* (pp. 15–37). Springer. https://doi.org/10.1007/978-0-387-76815-1_2
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(Supplement), 234–240. <https://doi.org/10.2307/143141>
- United Nations. (2014). *World Urbanization Prospects: The 2014 Revision*. <https://www.un.org/en/development/desa/population/publications/urbanization/wup2014-report.asp>

- Velandia Naranjo, D. J. (2013). El impacto del sistema de Transporte Rápido de Buses (Metrobús) en los precios del suelo en la Ciudad de México. Lincoln Institute of Land Policy. <https://www.lincolninst.edu/publications/working-papers/el-impacto-del-sistema-transporte-rapido-buses-metrobus-en-los-precios/>
- von Thünen, J. H. (1826). *The Isolated State in Relation to Agriculture and Political Economy*. Friedrich Perthes.
- Wen, H., Xiao, Y., & Zhang, L. (2017). Spatial effect of river landscape on housing price: An empirical study on the Grand Canal in Hangzhou, China. *Habitat International*, 63, 34–44. <https://doi.org/10.1016/j.habitatint.2017.03.00>
- Wilkinson, R. K. (1973). House Prices and the Measurement of Externalities. *The Economic Journal*, 83(329), 72. <https://doi.org/10.2307/2231101>
- Witte, A. D., Sumka, H. J., & Erekson, H. (1979). An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets. *Econometrica*, 47(5), 1151. <https://doi.org/10.2307/1911956>
- Wittowsky, D., Hoekveld, J., Welsch, J., & Steier, M. (2020). Residential housing prices: Impact of housing characteristics, accessibility and neighbouring apartments – a case study of Dortmund, Germany. *Urban, Planning and Transport Research*, 8(1), 44–70. <https://doi.org/10.1080/21650020.2019.1704429>
- Wu, Y., & Han, F. (2025). Ripples create waves: Assessing the catalytic effects of neighborhood micro-regeneration by integrating interpretable machine learning and econometrics. *Cities*, 161, 105860. <https://doi.org/10.1016/j.cities.2025.1058604>
- Xu, Y., Chen, R., Du, H., Chen, M., Fu, C., & Li, Y. (2025). Evaluation of green space influence on housing prices using machine learning and urban visual intelligence. *Cities*, 158, 105661. <https://doi.org/10.1016/j.cities.2024.105661>
- Yang, J., Bao, Y., Zhang, Y., Li, X., & Ge, Q. (2018). Impact of Accessibility on Housing Prices in Dalian City of China Based on a Geographically Weighted Regression Model. *Chinese Geographical Science*, 28(3), 505–515. <https://doi.org/10.1007/s11769-018-0954-6>
- Yazdani, M. (2021). Machine learning, deep learning, and hedonic methods for real estate price prediction. arXiv. <https://arxiv.org/abs/2110.07151>
- Yoshida, T., & Seya, H. (2021). Spatial prediction of apartment rent using regression-based and machine learning-based approaches with a large dataset. *Geographical Analysis*, 54(2), 239–260. <https://doi.org/10.1111/gean.12283>

- Yuan, F., Wei, Y. D., & Wu, J. (2020). Amenity effects of urban facilities on housing prices in China: Accessibility, scarcity, and urban spaces. *Cities*, 96, 102433. <https://doi.org/10.1016/j.cities.2019.102433>
- Zhao, Q., & Hastie, T. (2019). Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 37(1), 1–15. <https://doi.org/10.1080/07350015.2019.1624293>
- Ziccardi, A. (Coord.). (2004). Participación ciudadana y políticas sociales en el ámbito local. Instituto de Investigaciones Sociales, UNAM. <https://repositorio.unam.mx/contenidos/5046853>
- Zubicaray Díaz, G. (2015). Precio de la vivienda y accesibilidad al empleo en el Distrito Federal, 2012 [Tesis de maestría, El Colegio de México]. Repositorio COLMEX. <https://hdl.handle.net/20.500.11986/COLMEX/10005057>