



EL COLEGIO DE MÉXICO

CENTRO DE ESTUDIOS ECONÓMICOS

MAESTRÍA EN ECONOMÍA

TRABAJO DE INVESTIGACIÓN PARA OBTENER EL GRADO DE
MAESTRO EN ECONOMÍA

ANÁLISIS DE PRECIOS HEDÓNICOS PARA AIRBNB EN LA CDMX

DIEGO ALBERTO LÓPEZ TAMAYO

PROMOCIÓN 2019 - 2021

ASESORA:

DRA. AURORA RAMÍREZ ÁLVAREZ

JULIO 2021

Agradecimientos

A mis padres, por su ejemplo diario y apoyo incondicional en este proceso.

A mis profesores y laboratoristas, por sus enseñanzas y dedicación.

A la Dra. Aurora Ramírez Álvarez, por su dedicada asesoría y consejos.

A mis compañeros, por su amistad y su disposición a siempre trabajar en equipo.

Al Colegio de México, por permitirme ser parte de una gran comunidad académica.

Resumen

Las plataformas digitales han revolucionado las industrias de transporte, comercio y alojamiento en los últimos años. Airbnb representa uno de los casos de mayor éxito y el estudio de los determinantes del precio de los listados son un tema de investigación reciente, donde el mercado mexicano está poco explorado. Para el caso de la Ciudad de México, los anfitriones profesionales (aquellos con más de 2 listados registrados) representan el 41 % de la oferta. Esta heterogeneidad en la oferta y el hecho que la plataforma permite a sus anfitriones establecer libremente cuánto cobrar por noche, convierte a la fijación de precio por noche en un elemento estratégico en la dinámica del mercado.

Utilizando una base de datos novedosa que detalla los precios y atributos de los listados en la CDMX, se identifican una serie de atributos intrínsecos, extrínsecos y de plataforma para aplicar un modelo de precios hedónicos a 13,810 listados que permite explicar las diferencias en los niveles de precios y la contribución marginal de los atributos al precio por noche.

Los resultados muestran que permitir un mayor número de huéspedes, ofrecer más amenities, mayor flexibilidad en la reservación, contar con el estatus de anfitrión profesional, que el contenido de las reseñas de huéspedes anteriores sea positivo y una mayor experiencia (antigüedad) en la plataforma tienen un impacto positivo y significativo en el precio por noche. Por otra parte, la falta de privacidad (una habitación privada o compartida), estar ubicado más lejos de puntos de interés en la ciudad, la incidencia delictiva en la zona, estar cerca de accesos a rutas de transporte colectivo y el volumen de reseñas de huéspedes anteriores tienen un impacto negativo y significativo en el precio. Estos hallazgos confirman, en gran parte, resultados de estudios anteriores y proporcionan un primer acercamiento al mercado mexicano de alojamiento colaborativo.

Keywords: Share economy, Airbnb, Mercados digitales, Fijación de precios, Precios hedónicos

Índice general

Agradecimientos	I
Resumen	III
1. Introducción	1
2. Revisión de Literatura	7
3. Datos y Estrategia Empírica	11
3.1. Fuentes de datos	11
3.2. Modelo de precios hedónicos	17
4. Airbnb en la CDMX	19
4.1. Estadística descriptiva de variables	25
5. Resultados y discusión	27
5.1. Regresiones hedónicas	27
5.2. Análisis por tipo de anfitrión	32
5.3. Análisis con precio compuesto	34
6. Conclusiones	37
A. Anexo	39
A.1. Fundamentos del modelo de regresión hedónica	39
B. Apéndice	43
B.1. Regresiones por tipo de anfitrión	43
B.2. Regresiones con precio compuesto.	43
Bibliografía	47
Índice de tablas	49
Índice de figuras	51

Capítulo 1

Introducción

Fundada en 2008, Airbnb se convirtió rápidamente en la plataforma digital líder en alojamiento *peer-to-peer*, operando actualmente más de 7 millones de listados¹ en 220 países (“2020 Airbnb Update,” 2020). Además, Airbnb es uno de los principales actores de la economía colaborativa o *share economy*. A diferencia de la industria hotelera tradicional, Airbnb no dispone de alojamientos propios ni tampoco los gestiona. Solo permite a sus anfitriones compartir sus viviendas con turistas, teniendo como propuesta de valor el intercambio de experiencias y la oportunidad de hospedarse en lugares no cubiertos por la industria hotelera.

La competencia basada en precios generada a través de intermediarios bilaterales, como Airbnb, puede tener efectos más allá de la plataforma y los agentes involucrados. Por ejemplo, en ciudades turísticas donde Airbnb tiene presencia se observan efectos en los precios de inmuebles y rentas, además de procesos urbanos como la gentrificación. En consecuencia, los gobiernos locales y la academia han desarrollado estudios diagnóstico para promover arreglos institucionales y regulatorios, procurando la mejor inserción de las plataformas digitales en los destinos turísticos. (Guttentag, 2019).

La literatura sobre fijación de precios son un tema reciente en el estudio de plataformas digitales. En términos generales, existen dos modelos de precios en las plataformas de economía colaborativa (Cai et al., 2019). El primer modelo está basado en algoritmos, lo que significa que la empresa tiene el control total sobre los precios de cada solicitud de servicio; ejemplo de este modelo es Uber. El segundo modelo de precios utiliza algoritmos solo como referencia, dejando la decisión de precios abierta a los proveedores, Airbnb está en esta segunda categoría.

La naturaleza de esta plataforma permite al anfitrión fijar libremente el precio por noche

¹Se utiliza el término “listado” para referirse a los anuncios de departamentos completos, habitaciones privadas, compartidas y otros tipos de alojamiento.

de su listado aportando sólo sugerencias e información sobre el precio promedio de listados similares. Esto puede llevar a una fijación de precios ineficiente por parte de anfitriones sin experiencia y otorga cierta ventaja a empresas o anfitriones profesionales (particulares con múltiples inmuebles, agentes de bienes raíces o empresarios inmobiliarios) quienes fijan precios utilizando su mayor conocimiento del mercado.

Para el caso de la Ciudad de México, los anfitriones profesionales (aquellos con más de 2 listados registrados) representan el 41 % de la oferta, una proporción mayor a la de ciudades como Nueva York (35 %), Berlín (24 %) y Amsterdam (21 %). (“Inside Airbnb. Adding Data to the Debate.” 2021) Esta heterogeneidad en la oferta, y el hecho que la plataforma permite a sus anfitriones establecer libremente cuánto cobrar por noche, convierte a la fijación de precio en un elemento estratégico en la dinámica del mercado.

Estudios anteriores se han basado principalmente en ciudades de Estados Unidos, Europa o Asia y el mercado mexicano está poco explorado. La presente investigación tiene como objetivo aportar a la nueva literatura de fijación de precios en plataformas bilaterales estudiando el caso de Airbnb en la Ciudad de México a través de un modelo de precios hedónicos que incorpora atributos intrínsecos utilizados en la literatura tradicional para hoteles e inmuebles (habitaciones, baños, amenidades), atributos extrínsecos (competencia, niveles de delincuencia, acceso a redes de transporte) y atributos de plataforma basadas en los principios del *share economy* (reputación, calificación de otros usuarios, atributos del anfitrión). Utilizando una base de datos novedosa que detalla los precios y atributos de los listados en la CDMX, se aplica una regresión hedónica al precio de 13,810 listados que permite explicar las diferencias en los niveles de precios y la contribución marginal de los atributos sobre el precio por noche.

Si bien existen estudios y metodologías de determinación de precios para los servicios tradicionales de hoteles, estos modelos no son necesariamente adecuados para Airbnb. En primer lugar, como propiedades privadas, los servicios de alojamiento en Airbnb muestran mucha más complejidad que los hoteles en términos de atributos y características del anfitrión, convirtiéndose en un servicio fuertemente diferenciado. En segundo lugar, basado en los principios de la economía colaborativa, tanto anfitriones como huéspedes buscan intercambios más allá de lo económico².

El modelo de precios hedónicos ofrece un marco teórico para comprender las estrategias de precios en un mercado digital peer-to-peer como Airbnb. Este modelo, derivado de la teoría del consumidor de (Lancaster, 1966) y del modelo teórico de (Rosen, 1974), asume que el precio de un bien/servicio diferenciado puede descomponerse en un vector de atributos, donde

²Servicios específicos que un hotel no ofrece, flexibilidad para definir acuerdos, alojarse en zonas en las que no hubiera sido posible de otra forma. (IMCO, 2018)

el aporte marginal de cada uno al precio de venta puede interpretarse como la disposición marginal promedio a pagar (AMWP en inglés) por cada atributo. Dado que las características de un hospedaje no se pueden consumir por separado, existe un mercado implícito para cada una. Las funciones hedónicas descomponen el precio de venta (resultante de la interacción de oferta y demanda) en sus atributos. A cada atributo se le puede asignar un precio “sombra” que muestre tanto la valoración que hacen los consumidores de ese elemento como su voluntad de pagar una prima para disfrutar de su uso o consumo.

Debido a que los servicios de hostelería e inmuebles son productos heterogéneos, la teoría de precios hedónicos se ha utilizado ampliamente en investigaciones sobre precios de inmuebles y hostelería. Además, los servicios de Airbnb tienen la particularidad de tener bajos costos de transacción, alto nivel de interacción entre consumidores y el componente de reputación (reduciendo la asimetría de información) que acercan el consumo de Airbnb aún más a los supuestos del modelo de demanda hedónica que los hoteles tradicionales. (Lladós-Masllórens et al., 2020)

La Ciudad de México es la ciudad con mayor número de listados en el país y uno de los principales destinos turísticos. Para su estudio, se obtienen datos de Inside Airbnb: una organización independiente y no comercial que realiza scraping de la web de Airbnb y ofrece datos abiertos para academia y gobierno. Actualmente es la mejor fuente de datos abiertos para el estudio de esta plataforma.³ Se selecciona el periodo de estudio de enero 2020 con el fin de ofrecer resultados previos a los efectos de la pandemia Covid-19 que comenzó a tener efectos en el turismo global a partir del mes de febrero el mismo año.

Para los datos de niveles de delincuencia y acceso a redes de transporte utilizamos los Datos Abiertos Ciudad de México para identificar la frecuencia y localización de los delitos, así como las zonas de acceso para transporte colectivo Metrobús y Metro. Para elegir los 10 puntos de interés (PDI) respecto a los cuales se toman distancias (fórmula Haversine) nos basamos en “*Las atracciones más populares en Ciudad de México*” de acuerdo con la plataforma Trip Advisor. Esta fuente de datos es utilizada para el mismo propósito en (Chen & Xie, 2017; Guttentag, 2019).

El análisis de las diferencias entre anfitriones profesionales y no profesionales está basado en la hipótesis de que un anfitrión con más listados registrados no es una persona que busque sólo un ingreso adicional rentando ocasionalmente su propiedad (siendo éste el fundamento de *share economy* detrás de Airbnb), sino que es un agente profesional que considera su actividad en Airbnb como un negocio completo administrando múltiples propiedades. Como resultado,

³Una de las principales limitaciones en el estudio de las plataformas digitales (Uber, Airbnb, Facebook) es la falta de datos abiertos. La mayoría de estudios aplicados se realizan en colaboración con la empresa.

tendrá un mayor conocimiento del mercado y su fijación de precios será más informada y eficiente. Se identifica a un anfitrión profesional como aquel con más de dos listados registrados y activos siguiendo a (Cai et al., 2019) y (Arvanitidis et al., 2020). En algunos otros artículos como (Lladós-Masllorens et al., 2020) se considera profesional a cualquier anfitrión con más de un listado registrado. Dado que no existe un consenso al respecto, nos apegamos al criterio más estricto de más de dos listados.

El método de precios hedónicos tiene como supuesto central que el bien o servicio heterogéneo puede ser representado por los atributos que lo conforman. En el caso de Airbnb, una persona que renta un espacio está comprando no solo el acceso y estancia en ese espacio, sino también las características intrínsecas (habitaciones, servicios, comodidades) y extrínsecas (localización, política de alojamiento). Adicionalmente, un consumidor toma en cuenta una serie de características de plataforma que le dan información adicional sobre el anfitrión y lugar que está rentando. Por ejemplo, los comentarios de huéspedes anteriores, las calificaciones que le han otorgado al anfitrión y el estatus en la plataforma del anfitrión (si cuenta con documentos verificados y la categoría de *Superhost*.) Se estima un modelo log lineal incorporando los grupos de características intrínsecas, extrínsecas y de plataforma para identificar el impacto que cada característica tiene en la fijación de precios. La selección de las características se basa en los modelo hedónicos tradicionales para inmuebles e industria hotelera y en literatura reciente de precios hedónicos para Airbnb en las ciudades de Montreal (Gibbs et al., 2017), Barcelona (Lladós-Masllorens et al., 2020) y Hong Kong (Cai et al., 2019).

En contraste con (Lladós-Masllorens et al., 2020), quienes se limitan a estudiar anfitriones no profesionales (sólo anfitriones con un listado registrado), este estudio profundiza en la diferencia entre la fijación de precios de anfitriones profesionales y no profesionales. Además, realizamos una clasificación de variables que distingue entre factores intrínsecos y extrínsecos y de plataforma no presente en (Gibbs et al., 2017) y (Lorde et al., 2019).

Esta investigación contribuye a la literatura de precios hedónicos para el estudio de Airbnb incorporando por primera vez al conjunto de variables explicativas los niveles de delincuencia y acceso a rutas de transporte colectivo, variables relevantes en una ciudad tan compleja como lo es la CDMX. Se analiza por primera vez el sentimiento de las reseñas de huéspedes y no se limita a usar sólo el volumen de reseñas como factor explicativo, que ha tenido resultados inconsistentes en la literatura.

Se encuentra que permitir un mayor número de huéspedes, ofrecer más amenidades, mayor flexibilidad en la reservación, el estatus de anfitrión profesional, que el contenido de las reseñas de huéspedes anteriores sea positivo y una mayor experiencia (antigüedad) en la plataforma se valoran en el mercado y tienen un impacto positivo y significativo en el precio. Por otra

parte, la falta de privacidad (ofrecer una habitación privada o compartida), estar ubicado más lejos de puntos de interés en la ciudad, la incidencia delictiva en la zona, estar cerca de accesos a rutas de transporte colectivo y el volumen de reseñas de huéspedes anteriores tienen un impacto negativo y significativo en el precio.

Respecto a este último resultado, (Gibbs et al., 2017) encuentra que un mayor número de reseñas tienen un impacto negativo en el precio. En contraste, (Lladós-Masllorens et al., 2020) encuentran un impacto positivo de las reseñas en el precio. Con el fin de aportar una posible explicación sobre estos resultados inconsistentes, se realiza un análisis de sentimiento a través de la construcción de diccionarios con términos positivos y negativos dentro de las reseñas de los listados seleccionados. Se encuentra que los comentarios positivos aumentan significativamente el precio (4.97%), mientras que los comentarios negativos reducen el precio de forma no significativa y de menor magnitud (-0.24%).

En relación con el tipo de anfitrión, controlando por atributos mediante la regresión hedónica encontramos que ser un anfitrión profesional permite fijar un precio 16.2% por encima de un listado equivalente administrado por un anfitrión no profesional. Corroborando la hipótesis inicial sobre los anfitriones profesionales teniendo una ventaja sobre el resto al momento de fijar un precio.

El resto de la investigación sigue la siguiente estructura: el capítulo 2 presenta una revisión de la literatura sobre fijación de precios para Airbnb y los principales resultados obtenidos en estudios anteriores. El capítulo 3 presenta las fuentes de datos y el modelo de precios hedónicos a implementar. El capítulo 4 realiza una descripción completa de Airbnb en la Ciudad de México, presentando estadística descriptiva de las variables principales, mapas que ilustran la distribución espacial de los listados y se identifican las principales diferencias entre anfitriones profesionales y no profesionales. El capítulo 5 presenta y discute los resultados de las regresiones hedónicas y el capítulo 6 concluye e identifica las limitaciones del estudio.

Capítulo 2

Revisión de Literatura

La creciente adopción de plataformas digitales como una nueva forma de organizar las actividades económicas y sociales ha provocado un interés creciente en su investigación desde la academia. Como resultado, la literatura sobre plataformas digitales ha crecido significativamente en la última década. Ver (de Reuver et al., 2018) y (López-Tamayo, 2019) para una revisión del concepto de plataforma digital, las distintas clasificaciones en la literatura y las ramas de investigación actuales.

Airbnb es una plataforma digital a través de la cual los consumidores (huéspedes) pueden alquilar espacios de otros usuarios de la plataforma (anfitriones) como alojamiento privado. Estos espacios generalmente son un “alquiler completo” (casa, departamento), una habitación privada en una residencia donde el anfitrión también está presente o una habitación compartida. La oferta de Airbnb es amplia e incluye listados para todas las preferencias y presupuestos. Para los huéspedes, el proceso de búsqueda y reserva de alojamiento en Airbnb es relativamente similar al uso de una agencia de viajes en línea (por ejemplo, Expedia). Para los agentes que poseen y/o administran propiedades, la plataforma ofrece la posibilidad de convertirse de manera libre y a un costo casi nulo en un proveedor de alojamiento turístico, publicando fotografías y la descripción de su alojamiento, comunicándose con huéspedes potenciales y aceptando reservas y pagos de todo el mundo. Como otras plataformas digitales, el objetivo es reducir los costos de transacción entre los dos lados del mercado y obtener una comisión o tarifa por cada transacción realizada.

(Guttentag, 2019) realiza un meta análisis de la literatura en inglés para Airbnb y encuentra que la mayoría de las investigaciones se han publicado a partir de 2014, principalmente en revistas arbitradas de hotelería, turismo y administración para EE. UU., Canadá y Europa. Dentro de la categoría de estudios enfocados al anfitrión se encuentra el estudio de la fijación

de precios. A diferencia de otras plataformas digitales (por ejemplo, Uber), los anfitriones de Airbnb tienen la libertad de establecer sus propios precios y se han utilizado modelos de regresión hedónica para comprender cómo los diferentes atributos del listado influyen en el precio.

Los principales análisis de precios hedónicos (Cai et al., 2019; Chen & Xie, 2017; Gibbs et al., 2017; Lladós-Masllorens et al., 2020; Lorde et al., 2019) encuentran que el precio está asociado positivamente con un listado completo y privado (en comparación con un alojamiento compartido), el recuento de habitaciones y baños, el número de huéspedes permitidos, el estatus de *Superhost*, la experiencia del anfitrión, amenidades (por ejemplo, estacionamiento, cocina) y el recuento de fotos en el anuncio del listado. En contraparte, el precio se ha asociado negativamente con la distancia desde el centro de la ciudad, que un anfitrión tenga varios listados (anfitrión profesional), políticas de cancelación más flexibles, disponibilidad de reserva instantánea y el recuento de reseñas. Es importante tener en cuenta que estos estudios analizaron múltiples ciudades y se han encontrado diferencias notables entre ellas. Además, los hallazgos de algunos estudios no se han ajustado a estos patrones generales.

Dentro de los resultados irregulares, (Chen & Xie, 2017) estudiando Texas para 2017 no encuentran un impacto significativo del estado de *Superhost* en el precio, explicando que el mercado analizado tiene una proporción excepcionalmente alta de anfitriones Superhost comparado con otras ciudades de EE. UU. Por otra parte, (Cai et al., 2019) estudiando Hong Kong para 2016 encuentran que en comparación con estudios anteriores el efecto de ofrecer un listado completo en el precio de Airbnb es excepcionalmente alto, debido al “lujo del espacio” en esta ciudad sobrepoblada. También encuentran que solo los listados de “baja calidad” se benefician de la cercanía al centro de la ciudad, indicando efectos heterogéneos de la ubicación sobre los precios de Airbnb.

Un resultado interesante en la literatura es el del impacto del número de reseñas en el precio. (Gibbs et al., 2017) encuentran que más reseñas están asociadas significativamente con una reducción en el precio, explicando este resultado a través del siguiente mecanismo: más reseñas puede significar menor asimetría de información. En consecuencia, una menor capacidad de fijar un sobreprecio. Otro mecanismo es que los anfitriones saben que cada noche que su propiedad permanece vacía no obtienen ningún ingreso, por tanto, es probable que cobren precios más bajos para “llenar sus propiedades.” Este resultado negativo del número de reseñas se encuentra en la mayoría de estudios de precios hedónicos (Cai et al., 2019; Lladós-Masllorens et al., 2020; Lorde et al., 2019). Esta investigación busca aportar a esta discusión identificando el sentimiento de las palabras en las reseñas (positivas o negativas) y no solo evaluar su volumen.

Por otra parte, algunos estudios de precios para Airbnb han analizado variables adicionales más allá de las características intrínsecas del listado. Por ejemplo, (Kakar et al., 2018) estudiando San Francisco para 2015 utilizan esta metodología para observar la discriminación racial que ocurre en Airbnb y encontraron que, en igualdad de condiciones, los anfitriones asiáticos e hispanos en San Francisco tendían a cobrar entre un 8 - 10 % menos que los anfitriones blancos. No obstante, las tasas de ocupación fueron las mismas y ni el género ni la orientación sexual tuvieron un impacto similar en el precio. De esta forma podemos observar que el análisis de precios hedónicos puede incorporar variables extrínsecas al listado y descubrir fenómenos asociados a la plataforma digital. Esta investigación incorpora por primera vez el nivel de delincuencia y el acceso a redes de transporte público como características extrínsecas del listado que pueden impactar significativamente en el precio.

Para el caso de México, no existen estudios previos que utilicen esta metodología en el estudio de Airbnb y la literatura sobre la plataforma es escasa. Destaca (IMCO, 2018), un estudio privado realizado en colaboración directa con Airbnb donde se hace una comparación de la oferta de listados en la plataforma con la oferta hotelera tradicional en la CDMX para identificar las zonas de la ciudad donde ha penetrado más la plataforma. Estos resultados muestran que la alcaldía de Coyoacán tiene el mayor crecimiento de listados Airbnb y una baja oferta hotelera. Otro avance importante es (López-Gay et al., 2019), quienes estudian desde una perspectiva demográfica la penetración de Airbnb en las ciudades de Sao Paulo y CDMX identificando las características demográficas de las zonas donde se instala la oferta de Airbnb y los fenómenos urbanos resultantes, como la gentrificación.

La mayor parte de la investigación sobre Airbnb ha sido realizada en EE. UU., Canadá y Europa y se ha centrado en estas mismas regiones, mientras que los destinos de América Latina, África y Asia han recibido poca atención. Es en esta área de oportunidad donde esta investigación aporta al debate mostrando las particularidades del mercado mexicano e incorporando una serie de variables explicativas no utilizadas anteriormente en el análisis hedónico de precios intentando llenar, al menos parcialmente, ese vacío en la literatura.

Capítulo 3

Datos y Estrategia Empírica

3.1. Fuentes de datos

La fuente principal de datos sobre listados proviene del sitio Inside Airbnb, una organización independiente y no comercial de autodenominados “científicos de datos con interés social” que realiza periódicamente scraping de la plataforma para las principales ciudades del mundo, ofreciendo datos abiertos para academia y gobierno. Los datos para CDMX comienzan a estar disponibles a partir de marzo 2019 con periodicidad mensual. Esta fuente de datos ha sido utilizada en estudios previos (Cai et al., 2019; Chattopadhyay & Mitra, 2019; Lladós-Masllorens et al., 2020).

Se obtienen datos de 21,477 listados en la CDMX para enero 2020. Siguiendo a (Cai et al., 2019) se filtran los listados con al menos una reseña en los últimos 3 meses garantizando que los listados estuvieran “activos” al momento del estudio y con al menos una reseña en el periodo de estudio para asegurar que el precio del listado sea un precio realizado en el mercado. Además, se remueven todas las observaciones con valores nulos en alguna variable, resultando en 13,810 observaciones que conforman la muestra del estudio.

Estos datos incluyen un identificador del listado, su ubicación con una precisión de 100 m, la fecha de incorporación a la plataforma, el número de habitaciones, el número de baños, el tipo de inmueble, el número máximo de personas que puede admitir una estancia, las puntuaciones de los clientes y el precio medio por noche en moneda nacional (MXN peso mexicano) entre otras variables. En el proceso de limpieza de datos se crearon variables nuevas para poder realizar el análisis de regresión hedónica. Se extrae del texto de descripción del listado la lista de amenidades con las que cuenta. Se utiliza la ubicación geográfica para ubicar el listado en la alcaldía/colonia correspondiente y para calcular distancias respecto a

puntos de interés. Se utiliza la fecha de registro de listado y registro de anfitrión para obtener la antigüedad y experiencia del anfitrión en meses al momento de la extracción de datos.

Para el análisis de sentimiento de las reseñas se realizó una limpieza y clasificación de palabras. Debido a que las reseñas se encuentran en el idioma original que fueron escritas, se limita a clasificar palabras positivas y negativas en inglés y español. Para ello se quitan *stopwords* y se realiza un proceso de *stemming* para ambos idiomas. La clasificación se realiza utilizando tres fuentes: 1) el diccionario Harvard-IV para términos positivos y negativos en inglés, 2) la versión en español del diccionario NRC Emotion Lexicon agrupando *positive*, *joy*, *trust* como términos positivos y *negative*, *anger*, *disgust* como términos negativos. 3) Un diccionario propio basado en el análisis exploratorio de frecuencia de palabras en las reseñas.

Para la selección de puntos de interés (PDI) respecto a los cuáles se calculan distancias se consultó la plataforma Trip Advisor. Esta plataforma proporciona reseñas y contenido relacionado con viajes, siendo una fuente de primera mano para conocer los principales destinos turísticos de la CDMX. Esta fuente de datos es utilizada para el mismo propósito en (Chen & Xie, 2017; Guttentag, 2019). La Tabla 3.1 detalla los sitios seleccionados y su ubicación.

Tabla 3.1: Puntos de interés

Punto de interés (PDI)	Coordenadas en grados decimales (DD)
Museo Nacional de Antropología	(19.42612, -99.18628)
Castillo de Chapultepec	(19.42067, -99.18165)
Basílica de Guadalupe	(19.48513, -99.11741)
Palacio de Bellas Artes	(19.43563, -99.14117)
Coyoacán (Centro)	(19.34839, -99.16317)
Museo Frida Kahlo	(19.35546, -99.16205)
Plaza de la Constitución (Zócalo)	(19.43282, -99.13292)
Museo Soumaya	(19.44097, -99.20467)
Museo del Templo Mayor	(19.43488, -99.13132)
Bosque de Chapultepec	(19.41982, -99.18914)

Se obtiene la distancia de los listados a cada PDI en unidades de 100 metros a través de la fórmula Haversine, que calcula la distancia de entre dos puntos x, y de una esfera conociendo su longitud y su latitud de acuerdo a la siguiente fórmula:

$$D(x, y) = 2 \arcsin[\sqrt{\sin^2((\theta_y - \theta_x)/2) + \cos(\theta_x) \cos(\theta_y) \sin^2((\lambda_y - \lambda_x)/2)}]$$

Donde θ_x, θ_y representan las latitudes y λ_x, λ_y las longitudes de los puntos. Se utiliza la fórmula Haversine en particular por el buen nivel de precisión en distancias menores a un kilómetro. De las distancias calculadas se obtiene la media para la variable localización descrita en la Tabla 3.3. La Figura 3.1 muestra la distribución espacial de los PDI.

Para las variables sobre delincuencia y acceso a transporte utilizamos los Datos Abiertos Ciudad de México. Este portal ofrece distintas bases de datos de interés público agrupadas por categoría. Para nuestro propósito, utilizamos tres bases disponibles para Enero 2020:

- **Estaciones de Metrobús**
- **Estaciones de Metro**
- **Carpetas de investigación FGJ de la Ciudad de México**

De las primeras dos bases se extraen los datos de ubicación de los accesos a las redes de transporte colectivo para utilizarlos como característica de la colonia. La Figura 3.1 muestra la ubicación de los accesos a redes de transporte colectivo Metrobús y Metro. Se ubican 429 estaciones en conjunto para ambas redes de transporte en enero 2020.

Con los datos de “Carpetas de investigación FGJ de la Ciudad de México” se obtienen los delitos registrados por alcaldía de enero 2019 a enero 2020. Obtenemos también del Censo de Población y Vivienda 2020 la población mayor de 18 años por alcaldía. Con estos datos se calcula una **tasa de incidencia delictiva** como el cociente entre el número de delitos en el periodo especificado y la población adulta (18 años y más). Se expresa como los delitos por cada 100 habitantes en este rango de edad. La Tabla 3.2 muestra estos datos y la Figura 3.2 muestra en un mapa la distribución espacial en la CDMX.

Se agrupan las variables en factores intrínsecos, extrínsecos y de plataforma tomando como base a (Cai et al., 2019; Lladós-Masllorens et al., 2020) pero separando en su propio grupo a los atributos de plataforma para una mejor identificación. La Tabla 3.3 describe el conjunto de variables seleccionadas, la fuente y el procesamiento.

CDMX Enero 2020. Listados Airbnb: azul (13,810)
 Acceso a redes de transporte colectivo Metrobús y Metro: morado (429)
 Puntos de Interés Trip Advisor: verde (10)

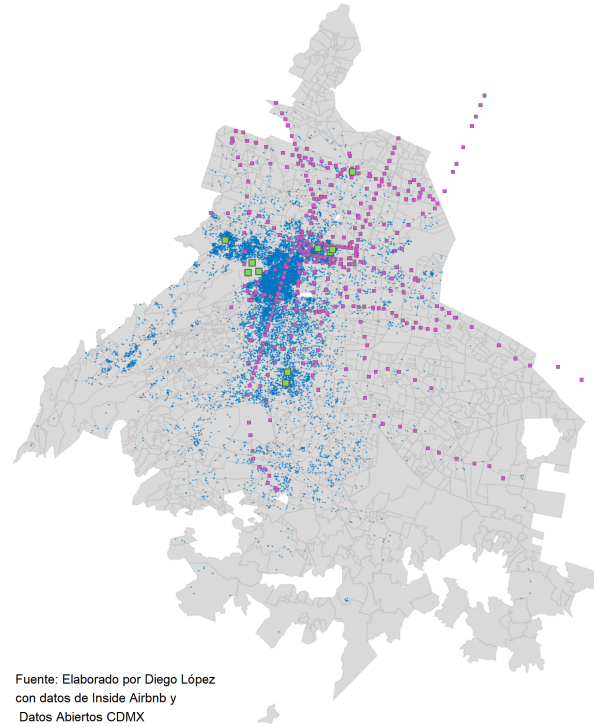


Figura 3.1: Mapa de listados, accesos a transporte y puntos de interés.

Tabla 3.2: Delitos en CDMX. Carpetas de investigación.

Alcaldía	Población 18 años y más	Delitos registrados	Tasa de incidencia delictiva
Azcapotzalco	345,123	10,820	3.14
Coyoacán	500,374	15,722	3.14
Cuajimalpa de Morelos	162,395	3,785	2.33
Gustavo A. Madero	911,482	24,229	2.66
Iztacalco	319,021	10,240	3.21
Iztapalapa	1,383,459	38,185	2.76
La Magdalena Contreras	189,047	4,004	2.12
Milpa Alta	108,278	1,641	1.52
Álvaro Obregón	590,251	17,983	3.05
Tláhuac	288,877	6,000	2.08
Tlalpan	539,464	14,524	2.69
Xochimilco	332,325	7,432	2.24
Benito Juárez	372,085	20,595	5.54
Cuauhtémoc	444,024	37,045	8.34
Miguel Hidalgo	332,956	16,227	4.87
Venustiano Carranza	347,576	14,418	4.15
Totales	7,166,737	242,850	3.36 (Promedio)

Nota: Población del Censo 2020. | Delitos registrados entre enero 2019-2020 | Delitos por cada 100 habitantes mayores de edad.

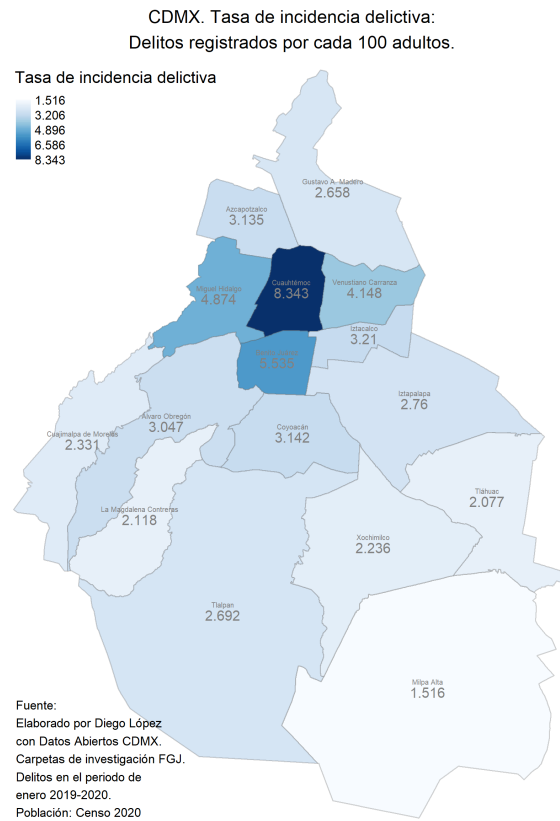


Figura 3.2: Mapa CDMX. Tasa de incidencia delictiva.

Tabla 3.3: Grupos y descripción de variables

Grupo	Variable	Descripción	Fuente / Proceso
Variable dependiente	Precio	Precio por noche de reserva.	Inside Airbnb
Factores intrínsecos	Tipo de listado	Variables dummies: Listado entero: casa o depto. Habitación privada en casa o depto. Habitación compartida en casa o depto.	Inside Airbnb. Creada a partir del tipo de habitación/casa
	Huéspedes permitidos	Número máximo de huéspedes permitidos	Inside Airbnb
	Habitaciones	Número de habitaciones	Inside Airbnb
	Baños	Número de baños completos y medios	Inside Airbnb
	Amenidades	Índice aditivo de hasta 92 amenidades del listado	Inside Airbnb Extracción de texto en descripción de listado
Factores extrínsecos	Flexibilidad (Política de reservación)	Índice aditivo: Flexibilidad de cancelación + Sin noches mínimas + Sin Tarifa de limpieza + No requerir documentos verificados	Inside Airbnb Índice creado a partir de dummies de requerimientos.
	Localización	Distancia media (fórmula Haversine) a Puntos de Interés (PDI). Selección de 10 puntos más visitados.	Trip Advisor para elección de PDI y Google Maps para coordenadas.
	Competencia	Número de competidores con mismo número de huéspedes permitidos en la colonia.	Inside Airbnb. Cálculo a partir de ubicación de listado.
	Delincuencia	Tasa de incidencia delictiva: cociente del número de delitos enero 2019 - 2020 entre población adulta (18 años y más)	Portal de datos de la CDMX. Carpetas de investigación FGJ Censo de Población y Vivienda 2020
	Acceso transporte	Número de accesos a estaciones de transporte colectivo Metrobús o Metro en colonia	Portal de datos de la CDMX. Estaciones de Metrobús y Metro
Factores de plataforma	Anfitrión profesional	Anfitrión con más de dos listados registrados	Inside Airbnb
	Estatus Superhost	Variable dicotómica: 1 = Certificado Superhost 0 = Sin certificado Superhost	Inside Airbnb
	Promedio de calificación	Clasificación global del listado (Promedio sobre 100)	Inside Airbnb
	Descripción	Presentación del listado por parte del anfitrión (medido en caracteres)	Inside Airbnb. Creado a partir de extracción de texto.
	Antigüedad	Tiempo transcurrido desde que el listado se registró en la plataforma (medido en meses)	Inside Airbnb. Diferencia entre fecha de registro y recolección.
	Experiencia	Tiempo transcurrido desde que el anfitrión se registró en la plataforma (medido en meses)	Inside Airbnb. Diferencia entre fecha de registro y recolección
	Número de reseñas	Reseñas realizadas por huéspedes y disponibles en el anuncio del listado.	Inside Airbnb
	Sentimiento de comentarios	Palabras positivas y negativas por reseña. (Análisis de sentimiento basado en diccionario Harvard-IV para inglés, NRC Emotion Lexicon para español y diccionario propio)	Inside Airbnb. Creado a partir de análisis de texto para las reseñas.

3.2. Modelo de precios hedónicos

La teoría de la utilidad hedónica (Lancaster, 1966) da sustento teórico al modelo de precios hedónicos (Rosen, 1974) en el cual se basa la metodología del estudio. El supuesto central del modelo es que el precio de venta de un bien o servicio diferenciado puede descomponerse en un vector de atributos que se consumen en conjunto. Creando así un mercado implícito para los atributos del bien heterogéneo (como lo son los inmuebles, la hostelería y los listados de Airbnb), en el cual a cada atributo se le puede asignar un precio “sombra” e interpretarse como la disposición marginal a pagar del consumidor por esta característica.

El punto de partida es la función de precios hedónicos, donde i identifica el listado y $Z_i = (z_{i1}, z_{i2}, \dots, z_{ij})$ es un vector de las j características que componen al bien heterogéneo i .

$$P_i = F(Z_i)$$

La derivada parcial de la función hedónica con respecto a cada característica j proporciona el precio marginal implícito o precio sombra, que representa la disposición marginal media de los compradores a pagar y la disposición marginal media de los vendedores a aceptar por un atributo. Ver Anexo 1 para un fundamento microeconómico del modelo de precios hedónicos bajo un contexto de equilibrio general.

$$\frac{\partial P_i}{\partial z_{ij}} = \frac{\partial F(Z_i)}{\partial z_{ij}}$$

Aunque varias formas funcionales son compatibles con el análisis de precios hedónicos, la forma semilogarítmica recomendada por (Rosen, 1974) se utiliza con mayor frecuencia en esta corriente de investigación. (Cai et al., 2019; Gibbs et al., 2017; Lorde et al., 2019)

$$\ln P_i = \beta_0 + \sum_{j=1}^k \beta_j Z_{ij} + \epsilon_i$$

Donde $\ln P_i$ es el logaritmo natural del precio de alquiler del listado i ; Z_{ij} es un vector de j atributos asociados con el listado i medidos en niveles; β_0 es la ordenada y ϵ_i es un término de error aleatorio con las propiedades habituales $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Para una mejor interpretación del modelo semilogarítmico se transforman los coeficientes de la siguiente forma:

$$\% \Delta P = (e^{\beta_j} - 1) \cdot 100$$

Lo que indica que un cambio unitario en la característica j (o la presencia de j si la variable es dummy) tiene un impacto de $\% \Delta P$ en el precio por noche de un listado.

Es importante mencionar que el supuesto clave detrás del modelo de precios hedónicos es que el precio a explicar es resultado de la interacción de oferta y demanda. Por tanto, revela las preferencias de los consumidores y su disposición a pagar. Este supuesto se cumple para nuestro estudio al restringir la muestra original a listados activos que han sido reservados al precio registrado, tomando como proxy de la demanda las reseñas realizadas, como lo hacen estudios similares. (Dudás et al., 2017; Gunter & Onder, 2018)

La estimación del modelo de precios hedónicos tiene dos objetivos. Primero, se revisa el ajuste (R^2) del modelo de regresión del precio de los listados en función de los atributos como variables independientes. En segundo lugar, se analizan los efectos marginales (es decir, los coeficientes) de las variables independientes sobre el precio de mercado de los listados de Airbnb. Para ello, se estiman una serie de modelos de regresión multivariada incorporando los grupos de características intrínsecas, extrínsecas y de plataforma descritas en la Tabla 3.3 de forma secuencial analizando el ajuste del modelo y los efectos marginales. Para cada conjunto de modelos, verificamos la robustez de los resultados utilizando el método de mínimos cuadrados ordinarios (MCO) con errores estándar robustos a heterocedasticidad agrupados en el nivel de colonia (clusterizados).

Los principales problemas econométricos asociados con la estimación del modelo hedónico son la heterocedasticidad en los términos de error y la multicolinealidad en las variables independientes. Estimamos la significancia de coeficientes con errores estándar robustos sensibles a heterocedasticidad, además se mide el factor de inflación de la varianza (VIF) para determinar el grado de multicolinealidad.

Capítulo 4

Airbnb en la CDMX

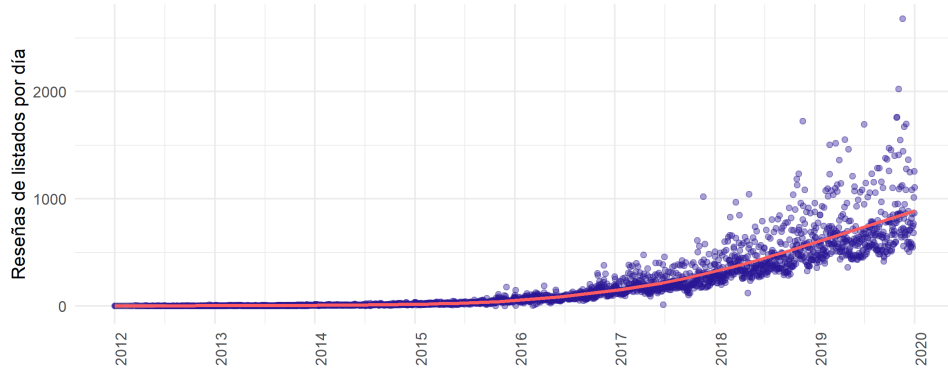
De acuerdo con la revisión de literatura, no hay un estudio previo que describa las características del mercado de Airbnb en la Ciudad de México de manera profunda y desagregada. Por esta razón, este capítulo describe las características intrínsecas, extrínsecas y de plataforma de los listados, además identifica las principales diferencias entre los tipos de anfitriones (profesionales y no profesionales) con el fin de presentar una imagen completa y actual de la plataforma en el mercado mexicano.

La base de datos contiene las reseñas realizadas por huéspedes de los listados seleccionados desde su registro de plataforma, esto nos permite utilizar el número de reseñas por día como proxy de la demanda de listados Airbnb, como lo hacen estudios similares (Dudás et al., 2017; Gunter & Onder, 2018). Con esto podemos presentar la evolución de la demanda que han recibido los listados seleccionados hasta enero de 2020, momento de la recolección de datos.

La Figura 4.1 nos muestra un aumento casi exponencial en la demanda en los últimos 8 años, con un despegue importante a partir de 2015-2016, esto coincide con lo reportado por (IMCO, 2018) quienes comparan la reserva de Airbnb con las reservaciones hoteleras. Basados en datos proporcionados por la plataforma, informan que de diciembre de 2015 a diciembre de 2016 el número de noches reservadas en Airbnb creció a más del doble.

Para enero 2020, la Figura 4.2 nos muestra la distribución de listados a nivel colonia y alcaldía y la Figura 4.3 nos muestra la misma distribución por tipo de anfitrión.

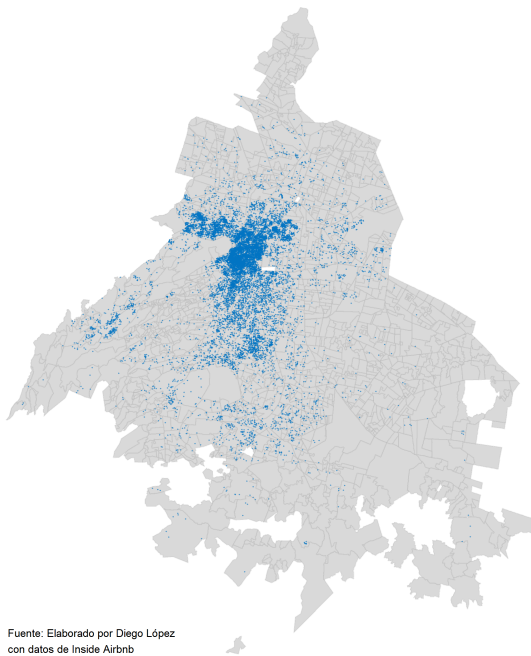
Listados de Airbnb en la CDMX Marzo 2012 - Enero 2020
 Número de reseñas como proxy de la demanda



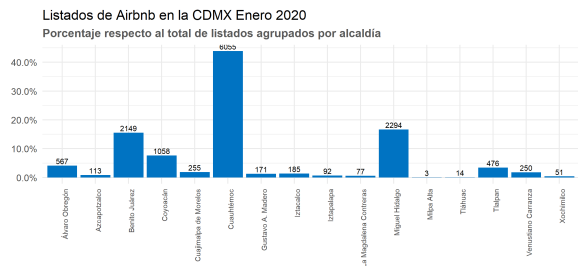
Elaborado por Diego López con datos de InsideAirbnb.

Figura 4.1: Evolución de demanda 2012 - 2020 Airbnb CDMX.

CDMX Enero 2020 Listados Airbnb: azul (13,810)



Fuente: Elaborado por Diego López con datos de Inside Airbnb



Elaborado por Diego López con datos de InsideAirbnb.

Figura 4.2: Listados por alcaldía y colonia

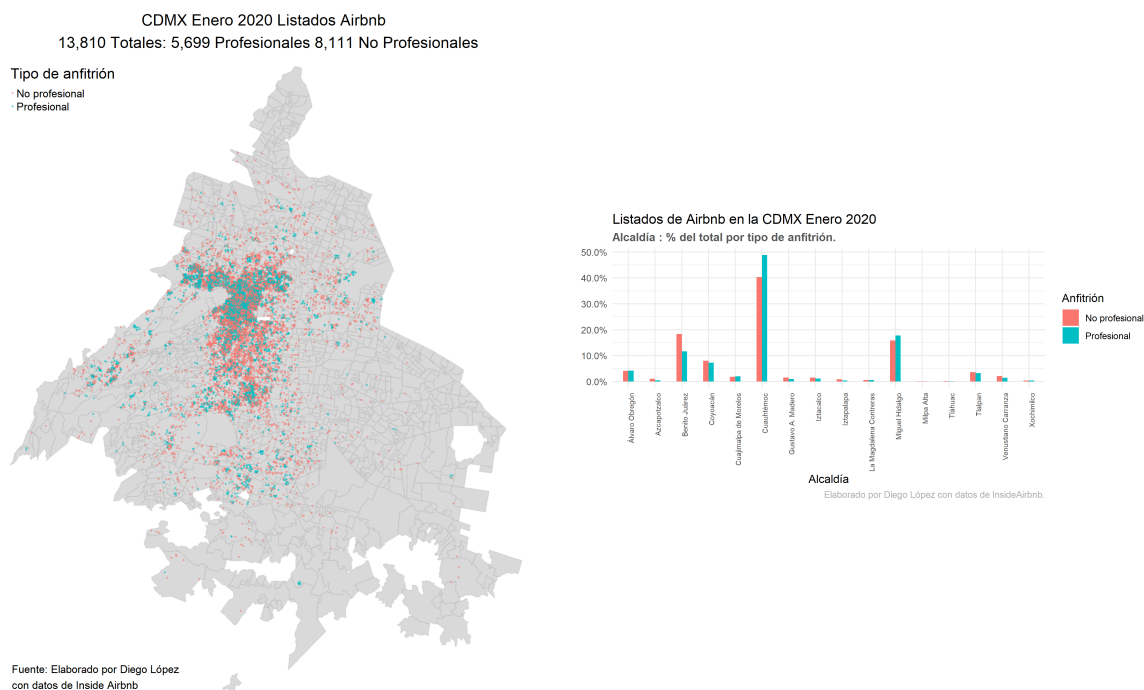


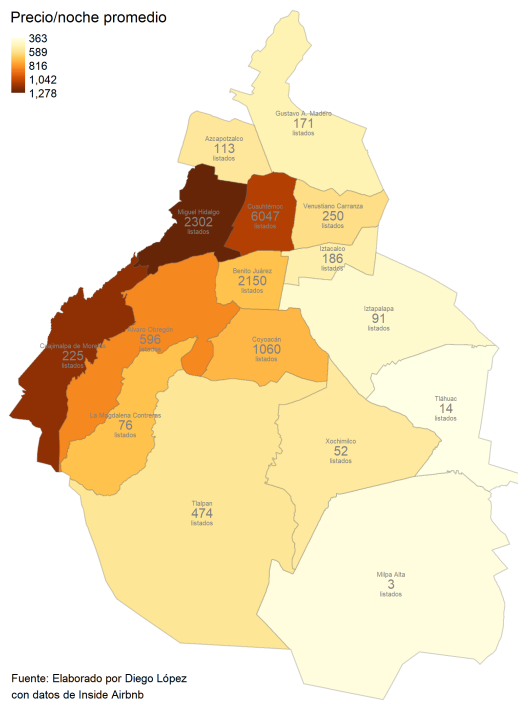
Figura 4.3: Tipo de Anfitrión: Listados por alcaldía y colonia

Podemos observar que la mayoría de los listados se concentran en 4 alcaldías: Cuauhtémoc, Miguel Hidalgo, Benito Juárez y Coyoacán en orden de relevancia. La preponderancia de la alcaldía Cuauhtémoc y Miguel Hidalgo se explican por abarcar la zona centro de la CDMX y 6 de los 10 puntos de interés seleccionados (ver Figura 3.1).

Al separar por tipo de anfitrión podemos observar que los anfitriones profesionales se concentran en las alcaldías Cuauhtémoc y Miguel Hidalgo, concentrando el 66 % de los anfitriones profesionales. También destaca que las alcaldías Benito Juárez y Coyoacán tienen una mayor proporción de anfitriones no profesionales, esto indica que los anfitriones profesionales tienden a concentrarse en zonas altamente turísticas (zona centro de la ciudad).

A continuación, la Figura 4.4 presenta la distribución del precio promedio de los listados agrupando por alcaldía y colonia. Podemos observar que, si bien la alcaldía Cuauhtémoc tiene la mayor concentración de listados, los precios promedio más altos se encuentran en la alcaldía Miguel Hidalgo. Esto se puede explicar porque esta alcaldía concentra algunas de las colonias de mayor ingreso medio por hogar (i.e. Chapultepec Morales, Polanco Reforma, Los Morales Polanco) y ser colonias con la mayor concentración de puntos de interés.

CDMX Enero 2020 Listados de Airbnb
 Precio promedio por noche a nivel alcaldía y PDI Trip Advisor: verde (10)



CDMX Enero 2020 Listados de Airbnb
 Precio promedio por noche a nivel colonia y PDI Trip Advisor: verde (10)

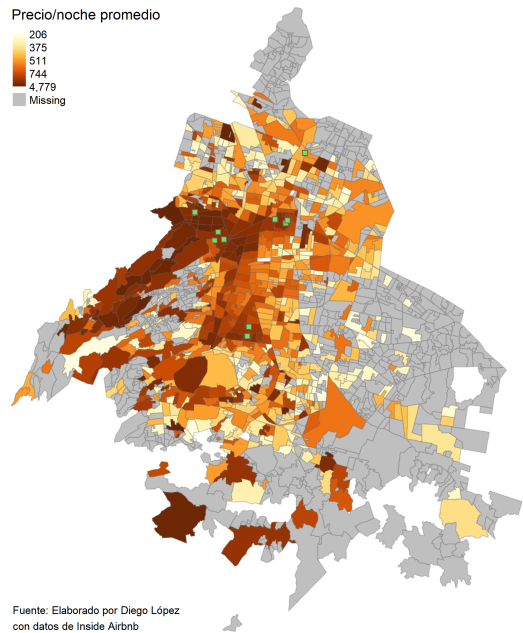


Figura 4.4: Mapa Precio promedio por alcaldía y colonia

La Figura 4.5 nos muestra que, para todas las alcaldías de la ciudad, los departamentos son el principal tipo de propiedad en la oferta de Airbnb seguido de las casas. Los condominios se distinguen de los departamentos por el tipo de propiedad del inmueble, pero para fines prácticos de la demanda podemos agruparlos con los departamentos. Esto refleja en parte la oferta de inmuebles y el desarrollo urbano que existe en la Ciudad de México.

Destaca que las alcaldías de Coyoacán y Tlalpan tienen una proporción similar de departamentos y casas. Esto se puede explicar en parte porque es un distrito histórico en donde el Instituto Nacional de Antropología e Historia (INAH) está encargado de supervisar cualquier tipo de construcción. Las leyes protegen su uso residencial y prohíben el cambio de uso en las propiedades, la remodelación de fachadas, la altura de nuevos edificios y la cantidad de espacios de estacionamiento resultando en una menor oferta de edificios departamentales/condominios. Lo mismo sucede en la zona centro de Tlalpan.

Tabla 4.1: Top 10 listados por número de reseñas. Airbnb Enero 2020.

Alcaldía	Colonia	Tipo	Precio/noche	Huéspedes	Superhost	Profesional	Reservaciones	Transporte	Competencia	Antig.Anfitrión	Antig.Listado
Cuauhtémoc	Condesa	Apartment	1124	4	0	1	208	1	133	73.87	52.40
Coyoacán	Del Carmen	Loft	1087	2	1	1	208	1	100	43.10	42.10
Cuauhtémoc	Condesa	Apartment	1687	6	0	1	207	1	53	73.87	61.30
Iztacalco	Granjas Mexico Ii	Apartment	487	2	0	1	207	3	10	60.20	59.60
Cuauhtémoc	Hipodromo Condesa	Apartment	1106	4	0	0	207	0	42	59.43	59.13
Benito Juárez	Vertiz Narvarte	Condominium	506	2	1	1	207	0	20	47.10	46.80
Cuauhtémoc	Hipodromo Ii	Apartment	375	2	1	0	207	3	152	95.23	43.47
Cuauhtémoc	Hipodromo I	House	300	2	1	0	207	2	181	52.53	41.73
Cuauhtémoc	Juarez	Loft	1649	5	1	0	206	6	39	88.53	71.33
Coyoacán	San Lucas (Barr)	Guesthouse	543	2	0	1	206	0	12	50.80	46.27

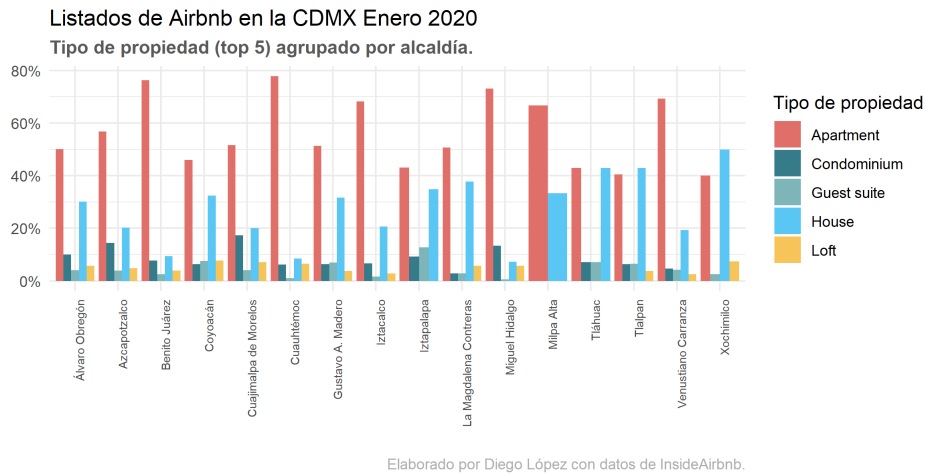


Figura 4.5: Listados por tipo de propiedad y alcaldía.

La Tabla 4.1 nos muestra el “top 10” listados por reseñas (demanda). Destaca el listado presente en la alcaldía Iztacalco con un precio de \$487/noche, relativamente menor al de sus competidores. Los principales listados en la ciudad pertenecen a anfitriones profesionales y no se observa que la categoría de *Superhost* sea un factor relevante para que se posicionen listados principales. En el otro extremo, la Tabla 4.2 nos muestra el “bottom 10” donde vemos una menor proporción de anfitriones profesionales, menor competencia (listados con mismo número de huéspedes en la colonia) y listados más recientes (menor antigüedad de listado), lo que puede explicar en parte su baja demanda.

Por último, el análisis de texto puede ilustrar los factores que influyen en la toma de decisiones del consumidor. Para enero 2020 se cuentan con 508,870 reseñas completas para los listados utilizados en este estudio, después de remover stopwords en inglés y español y realizar un steaming para ambos idiomas se cuenta con poco más de 19 millones de palabras para ser procesadas. Un primer acercamiento exploratorio utilizado en la literatura de análisis de sentimiento son las nubes de palabras. La Figura 4.6 nos muestra en una nube de palabras

4.1. Estadística descriptiva de variables

La Tabla 4.3 muestra la estadística descriptiva de las variables del estudio. La primera columna describe el conjunto completo de listados y las siguientes dos columnas de acuerdo con el tipo de anfitrión. Se realiza también una prueba de igualdad en medias.

Podemos observar que los anfitriones profesionales tienen en promedio 12 listados registrados y representan una proporción alta de los anfitriones operando en la CDMX (41 %). Estos agentes administran una cantidad considerable de inmuebles y con ello pueden obtener más información del mercado para fijar precios de manera más eficiente respecto a los anfitriones no profesionales. Las pruebas de igualdad en medias muestran que para la mayoría de variables, estos grupos son significativamente diferentes. Cabe destacar algunas características de los listados Airbnb en CDMX:

- Encontramos diferencias significativas en los listados separando por tipo de anfitrión. Los listados profesionales muestran un mayor número de reservas y atributos más favorables en comparación con los ofrecidos por los no profesionales. En particular, tienen listados más grandes (pueden acomodar a más personas, 1.91 respecto 1.72 de no profesionales), tienen baños y dormitorios adicionales, brindan más comodidades (25.89 respecto 24.04 de no profesionales) y se encuentran mejor ubicados.
- Más de la mitad (58 %) de los listados registrados en la CDMX son listados completos y no hay una diferencia significativa entre el tipo de anfitrión. Tampoco se encuentra una diferencia entre anfitriones en la antigüedad promedio de los listados, que a nivel general es de 20.6 meses. Sin embargo, sí hay una diferencia significativa en la experiencia entre anfitriones profesionales y no profesionales, teniendo los primeros en promedio 4 meses más de antigüedad en la plataforma.
- Coincidiendo con (Chen & Xie, 2017; Gibbs et al., 2017), la calificación promedio de los listados es particularmente alta, una media de 95.66 sobre 100. La literatura atribuye este resultado a que, en general, las interacciones que genera Airbnb son positivas y existe una tendencia de los anfitriones a calificar positivamente una vez que toman la decisión de escribir una reseña.
- Respecto al sentimiento de las reseñas, podemos observar la predominancia de los términos positivos en el texto de las reseñas (3.69 por reseña promedio) sobre los términos negativos (0.66 por reseña promedio). Este fenómeno se puede explicar por la misma razón de las calificaciones altas, existe una tendencia de calificación positiva por parte de los huéspedes.

Tabla 4.3: Estadística descriptiva de variables

Variable	Total / Promedio	Profesionales	No Profesionales	t-test medias	Min	Max
Listados	13,810	5,699 41%	8,111 59%	NA	NA	NA
Listados registrados	5.45 (0.08)	11.36 (0.17)	1.30 (0.01)	p-value 0.000	1	78
Precio	976.91 (6.81)	1,116.32 (12.15)	878.95 (7.65)	p-value 0.000	206	5,809
Precio/ persona	318.67 (1.81)	342.08 (2.93)	302.22 (2.28)	p-value 0.000	12.875	5,322
Listado entero: casa o depto	0.58 (0.00)	0.58 (0.01)	0.59 (0.01)	p-value 0.191	0	1
Habitación privada en casa o depto	0.41 (0.00)	0.41 (0.01)	0.41 (0.01)	p-value 0.585	0	1
Habitación compartida en casa o depto	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)	p-value 0.000	0	1
Huéspedes permitidos	1.80 (0.01)	1.92 (0.02)	1.72 (0.01)	p-value 0.000	1	16
Habitaciones	1.49 (0.01)	1.53 (0.02)	1.47 (0.01)	p-value 0.001	1	50
Baños	1.42 (0.01)	1.48 (0.02)	1.37 (0.01)	p-value 0.000	0.5	50
Amenidades del listado	24.81 (0.08)	25.90 (0.14)	24.04 (0.11)	p-value 0.000	1	92
Flexible (1=flexible,0=estricto)	0.87 (0.00)	0.90 (0.00)	0.86 (0.00)	p-value 0.000	0	1
Distancia media a PDI (unidad: 100 m)	66.60 (0.25)	64.99 (0.39)	67.73 (0.34)	p-value 0.000	45.56	291.69
Competencia en colonia (listados misma capacidad)	49.07 (0.54)	54.13 (0.88)	45.52 (0.68)	p-value 0.000	1	251
Delincuencia en alcaldía Tasa de incidencia delictiva	6.048 (0.02)	6.230 (0.03)	5.920 (0.02)	p-value 0.000	1.516	8.343
Acceso a estaciones de transporte colectivo	1.97 (0.03)	2.24 (0.06)	1.79 (0.04)	p-value 0.000	0	21
Superhost	0.47 (0.00)	0.54 (0.01)	0.42 (0.01)	p-value 0.000	0	1
Calificación (máx. 100)	95.66 (0.05)	94.98 (0.07)	96.15 (0.06)	p-value 0.000	20	100
Descripción anfitrión (caracteres)	192.48 (2.82)	244.66 (5.17)	155.82 (3.07)	p-value 0.000	0	5,443
Antigüedad listado (meses)	20.61 (0.14)	20.58 (0.23)	20.63 (0.18)	p-value 0.849	0.167	116.633
Antigüedad anfitrión (meses)	44.00 (0.20)	46.56 (0.33)	42.20 (0.26)	p-value 0.000	0.367	133.567
Reviews totales	31.90 (0.32)	34.84 (0.55)	29.84 (0.40)	p-value 0.000	2	208
Reviews / mes	1.91 (0.02)	2.06 (0.03)	1.80 (0.02)	p-value 0.000	0.03	17.8
Palabras positivas / reseña	3.69 (0.02)	3.59 (0.02)	3.75 (0.02)	p-value 0.000	0	25
Palabras negativas / reseña	0.66 (0.01)	0.69 (0.01)	0.64 (0.01)	p-value 0.000	0	15.5

Nota: Error estándar en paréntesis.

Capítulo 5

Resultados y discusión

Nuestro modelo seleccionado contiene 20 variables explicativas para el precio por noche. Algunas se extrajeron directamente de la versión final del conjunto de datos provenientes de Inside Airbnb (por ejemplo, huéspedes permitidos, baños, experiencia del anfitrión) mientras que otras se construyeron utilizando una o más variables del conjunto de datos original de acuerdo a lo descrito en la Tabla 3.3.

En primer lugar, estudiamos la correlación de nuestras las variables explicativas dado que uno de los principales problemas a nivel econométrico de los modelos regresión hedónica es la multicolinealidad de las variables. Las Tablas 5.1 y 5.2 muestran que ningún atributo se correlaciona perfectamente. Posteriormente en el análisis de regresión se revisa el factor de inflación de la varianza (VIF) para cuantificar la intensidad de la multicolinealidad bajo los supuestos de MCO.

A continuación presentamos los resultados de las regresiones hedónicas donde la variable dependiente es el logaritmo del precio por noche del alojamiento (expresado en pesos mexicanos MXN), y las variables independientes se agrupan en los tres bloques descritos en la Tabla 3.3: Factores intrínsecos, extrínsecos y de plataforma.

5.1. Regresiones hedónicas

En la Tabla 5.3, el modelo 1 estima el efecto de los atributos intrínsecos del listado sobre el precio. El modelo 2 agrega variables relacionadas con los atributos extrínsecos al listado y el modelo 3 agrega atributos de plataforma. Para facilitar la interpretación de coeficientes la Tabla 5.4 muestra los resultados aplicando la fórmula $\% \Delta P = (e^{\hat{\beta}_j} - 1) \cdot 100$. Lo que indica que un cambio unitario en la característica j (o la presencia de j si la variable es dummy)

Tabla 5.1: Matriz de correlación 1/2

	Huéspedes	Habitaciones	Baños	Amenidades	Flexibilidad	Distancia PDI	Competencia	Delincuencia	Acceso transp.
Listado entero	0.471	0.303	0.129	0.255	-0.109	-0.136	-0.052	0.139	0.087
Huéspedes		0.652	0.479	0.222	0.014	-0.057	-0.22	0.086	0.075
Habitaciones			0.751	0.132	-0.013	-0.001	-0.178	0.016	-0.005
Baños				0.133	-0.01	0.008	-0.106	0.01	-0.016
Amenidades					-0.038	-0.06	-0.001	0.051	0.038
Flexibilidad						0.051	-0.019	-0.047	-0.023
Distancia PDI							-0.401	-0.756	-0.256
Competencia								0.532	0.175
Delincuencia									0.387
Acceso transp.									
Profesional									
Superhost									
Calificación									
Descr. Anfitrión									
Antigüedad(mes)									
Experiencia(mes)									
No. reseñas									
Palabras +/reseña									
Palabras -/reseña									

Tabla 5.2: Matriz de correlación 2/2

	Profesional	Superhost	Calificación	Descr. Anfitrión	Antigüedad(mes)	Experiencia(mes)	No. reseñas	Palabras +/reseña	Palabras -/reseña
Listado entero	-0.011	0.109	-0.004	-0.026	-0.017	0.045	0.177	0.163	0.233
Huéspedes	0.094	0.041	-0.051	0.024	-0.037	0.01	0.104	0.041	0.074
Habitaciones	0.027	0.009	-0.029	-0.009	-0.003	0.011	0.025	0.038	0.054
Baños	0.057	0.001	-0.015	0.031	0.012	0.022	-0.011	0.039	0.045
Amenidades	0.093	0.277	0.14	0.148	0.13	0.124	0.325	0.172	0.124
Flexibilidad	0.058	-0.011	-0.015	-0.069	-0.209	-0.161	0.026	-0.194	-0.174
Distancia PDI	-0.045	-0.023	0.028	-0.036	-0.045	-0.142	-0.171	-0.271	-0.26
Competencia	0.067	0.022	0.001	0.031	0.078	0.142	0.117	0.25	0.215
Delincuencia	0.07	0.017	-0.038	0.047	0.037	0.141	0.184	0.311	0.308
Acceso transp.	0.056	0.035	-0.007	0.057	0.023	0.04	0.117	0.081	0.111
Profesional		0.111	-0.101	0.132	-0.002	0.089	0.065	-0.043	0.037
Superhost			0.286	0.079	-0.008	0.023	0.244	0.136	0.027
Calificación				0.028	-0.009	0.005	0.052	0.17	-0.136
Descr. Anfitrión					0.167	0.197	0.099	0.148	0.117
Antigüedad(mes)						0.477	0.422	0.286	0.27
Experiencia(mes)							0.19	0.273	0.247
No. reseñas								0.152	0.156
Palabras +/reseña									0.717
Palabras -/reseña									

tiene un impacto de $\% \Delta P$ en el precio por noche de un listado.

Los 3 modelos tienen un buen desempeño pasando la prueba de significancia global F al 95 % y el R^2 ajustada (penalizando por la adición de variables adicionales) incrementa de 0.524 a 0.597, lo que significa que nuestros modelos pueden explicar más del 50 % de la varianza en el logaritmo del precio por noche y las variables adicionales que se incorporan en cada etapa aportan al poder explicativo del modelo. Se utilizan errores estándar robustos a heterocedasticidad agrupados por colonia en todos los modelos para tomar en cuenta la posible correlación en la varianza de errores. Ninguna de las variables utilizadas en los modelos rebasa un VIF de 3.21, por debajo del umbral de 5 establecido por la literatura para considerar problemas de multicolinealidad.

Al observar en el Modelo 1 los atributos intrínsecos de los listados queda claro que ofrecer un listado completo (departamento/casa completa) tiene un impacto muy importante en el precio de un anuncio. Ofrecer en cambio una habitación privada reduce el precio en 50 % y una habitación compartida en 69.5 % respecto al listado completo. Este patrón fue constante en el resto de los modelos al incorporar el resto de los atributos. Los demás atributos intrínsecos tienen en general un impacto positivo y significativo. El número de huéspedes es el más relevante con un incremento de 8.7 % en el precio por cada huésped adicional permitido. Destaca que el número de habitaciones no resulta significativo, contrario al consenso en la literatura. El precio de listados de Airbnb también está influenciado positivamente por la provisión de una amplia gama de amenidades diseñados para ofrecer estadías más convenientes, cómodas y agradables, teniendo un incremento de 1.03 % por cada amenidad adicional que se ofrezca.

Incorporando los atributos extrínsecos del listado en el Modelo 2 observamos que una mayor flexibilidad en la reserva (no solicitar depósitos de garantía, no tener un número mínimo de noches, no solicitar documentos del huésped) está asociado con un incremento de 3.54 % en el precio. Esto indica que el mercado mexicano valora la flexibilidad de reserva y es un resultado opuesto al obtenido por (Cai et al., 2019; Lladós-Masllorens et al., 2020) quienes argumentan que los huéspedes valoran positivamente un anfitrión más estricto como señal de profesionalismo. Por otra parte, la distancia promedio respecto a puntos de interés y el nivel de competencia en la zona tienen impacto bajo pero significativo en el precio. Aumentar en 100 m la distancia promedio respecto a los puntos de interés (ver Tabla 3.1) tiene un impacto negativo de -0.34 % en el precio de un listado y un competidor adicional con la misma capacidad en la colonia aumenta el precio en 0.2 %. Esto puede ser consecuencia de la distribución de puntos de interés en la CDMX (ver Figura 3.1) ; un listado de Airbnb en estas colonias tiene una mayor probabilidad de atraer huéspedes, a pesar de la mayor densidad de competidores directos.

Dos variables que no han sido incorporadas en estudios anteriores son el nivel de delincuencia (medido como la tasa de incidencia delictiva, ver Tabla 3.3) y el acceso a rutas de transporte colectivo en la colonia. Un aumento en el número de delitos por cada 100 personas mayores de 18 años tiene un efecto negativo de - 2.7 % en el precio por noche al incorporar todas las variables explicativas. Este resultado es interesante dado que ningún estudio anterior ha medido el efecto de la delincuencia en el precio de Airbnb y concuerda con lo encontrado por (Nuñez et al., 2017) quienes utilizan una metodología de precios hedónicos para estimar el efecto de la criminalidad como una “desamenidad” que refleja niveles de precios de renta más bajos para distintos municipios en México, incluida la CDMX.

El coeficiente del acceso a rutas de transporte nos muestra que contar con un acceso adicional a redes de transporte colectivo (Metrobús o Metro) en la colonia reduce el precio en - 0.65 % al incorporar todas las variables explicativas, un resultado inesperado ya que intuitivamente estar mejor comunicado con las redes de transporte aumentaría el precio del listado. Dado que esta variable no ha sido incorporada anteriormente en la literatura, no se puede afirmar que es un resultado particular de la CDMX o existen no observables influyendo en el resultado, una posible explicación es la correlación positiva (0.39) entre la delincuencia y el acceso a rutas de transporte en la CDMX (ver Tabla 5.1), por tanto el efecto negativo de la delincuencia se refleja en las colonias con mayor acceso a transporte público.

El Modelo 3 añade los atributos de plataforma al modelo y muestra el mayor coeficiente de ajuste $R^2 = 0.59$. Se encuentra que, controlando por el resto de los atributos, ser un anfitrión profesional permite fijar un precio 16.2 % por encima de un listado equivalente administrado por un anfitrión no profesional. Corroborando la hipótesis inicial sobre los anfitriones profesionales teniendo una ventaja sobre el resto al momento de fijar un precio. Una unidad adicional la calificación global (sobre 100) tiene un impacto positivo de 0.33 % en el precio y contar con el status de *Superhost* no muestra un impacto significativo en el precio de los listados. Este último resultado coincide con (Chen & Xie, 2017) estudiando Texas para 2017, tampoco encuentran un impacto significativo del estado de *Superhost* en el precio, explicando que el mercado analizado tiene una proporción excepcionalmente alta de anfitriones Superhost comparado con otras ciudades de EE. UU. En el caso de México no se cuenta con información de otras ciudades para hacer una comparación similar.

Destaca que el número de reseñas tiene un impacto negativo en el precio. Este resultado, aun cuando es inesperado, es consistente con la literatura. En este sentido (Gibbs et al., 2017) argumenta que puede explicarse por fenómeno de asimetría de información, donde un mayor número de reseñas reduce la capacidad de establecer un sobreprecio y, por tanto, tiene un impacto negativo. Otra posible explicación es que sea un resultado de la demanda, donde listados con menor precio se reservan más frecuentemente y por tanto tienen mayores reseñas.

El análisis de sentimiento, tampoco presente anteriormente en la literatura, identifica el promedio de palabras positivas y negativas por reseña que tiene un listado. Una palabra positiva adicional refleja un incremento de 4.97 % en el precio. Las palabras negativas muestran un impacto negativo pero no significativo y de menor magnitud de las positivas (-0.24 %). Esto es interesante porque muestra que los usuarios de Airbnb valoran más las palabras positivas en las reseñas y los anfitriones explotan esta valoración con un incremento en el precio. En cambio, las palabras negativas parecen no afectar en el precio.

Tabla 5.3: Regresiones hedónicas

	Variable dependiente: logaritmo de precio por noche.		
	Intrínsecos (Modelo 1)	Modelo 1 + Extrínsecos (Modelo 2)	Modelo 2 + Plataforma (Modelo 3)
Habitación privada c.r.a. listado entero	-0.694*** (0.021)	-0.671*** (0.018)	-0.680*** (0.016)
Habitación compartida c.r.a. listado entero	-1.189*** (0.094)	-1.112*** (0.093)	-1.150*** (0.091)
Huéspedes permitidos	0.084*** (0.009)	0.097*** (0.010)	0.097*** (0.010)
Habitaciones	0.001 (0.020)	0.016 (0.021)	0.018 (0.020)
Baños	0.073*** (0.022)	0.063*** (0.024)	0.051** (0.021)
Amenidades	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)
Flexibilidad reserva		0.035* (0.020)	0.105*** (0.019)
Distancia PDI (100m)		-0.003*** (0.001)	-0.003*** (0.001)
Competencia		0.002*** (0.0003)	0.002*** (0.0003)
Delincuencia		-0.023 (0.015)	-0.028** (0.014)
Acceso transporte		-0.008*** (0.003)	-0.007*** (0.002)
Profesional			0.150*** (0.013)
Superhost			-0.00003 (0.012)
Calificación			0.003*** (0.001)
Descripción Anfitrión			-0.0001* (0.00003)
Antigüedad (meses)			0.002*** (0.0005)
Experiencia (meses)			0.001** (0.0004)
Número reseñas			-0.002*** (0.0002)
Palabras positivas/reseña			0.049*** (0.007)
Palabras negativas/reseña			-0.002 (0.010)
Observations	13,810	13,810	13,810
Adjusted R ²	0.524	0.565	0.597
F Statistic	2,535.397***	1,634.269***	1,025.504***

Note:

*p<0.1; **p<0.05; ***p<0.01
 Errores estándar robustos agrupados a nivel colonia.
 Tipo HC1. Clusters = 803 colonias.
 Se omite el intercepto.

Tabla 5.4: Coeficientes como aporte porcentual al precio por noche.

Grupo	Variable	Modelo 1	Modelo 2	Modelo 3
Factores intrínsecos	Habitación privada	-50.027*	-48.876*	-49.36*
	Habitación compartida	-69.562*	-67.103*	-68.328*
	Huéspedes permitidos	8.729*	10.227*	10.173*
	Habitaciones	0.099	1.577	1.8
	Baños	7.535*	6.471*	5.207*
	Amenidades	1.03*	0.973*	0.965*
Factores extrínsecos	Flexibilidad reserva		3.544*	11.052*
	Distancia PDI (100m)		-0.339*	-0.332*
	Competencia		0.19*	0.164*
	Delincuencia		-2.272	-2.725*
	Acceso transporte		-0.793*	-0.652*
Factores de plataforma	Profesional			16.197*
	Superhost			-0.003
	Calificación			0.328*
	Descripción Anfitrión			-0.005*
	Antigüedad (meses)			0.206*
	Experiencia (meses)			0.075*
	Número reseñas			-0.211*
	Palabras positivas/reseña			4.977*
	Palabras negativas/reseña			-0.249
Observaciones		13,810	13,810	13,810
Coeficiente de determinación ajustado R_{adj}^2		0.524	0.565	0.597

Nota: * Coeficientes significativos. Ver Tabla 5.3

5.2. Análisis por tipo de anfitrión

De forma complementaria, la Tabla 5.5 muestra los resultados de la especificación del Modelo 3 aplicado a las dos submuestras de anfitriones profesionales y no profesionales. (Ver en Apéndice las Tablas B.1 y B.2 las regresiones completas). Podemos observar que los coeficientes se mantienen en general estables al restringir para estas dos submuestras. Destaca que los coeficientes de determinación ajustados R_{adj}^2 aumentan respecto al modelo 3 y que los coeficientes para las habitaciones, antes no significativos ahora lo son particularmente para los anfitriones no profesionales con un aumento de 6% en el precio con una habitación adicional.

La flexibilidad de reservación se vuelve más relevante para la submuestra de anfitriones

Tabla 5.5: Coeficientes como aporte porcentual al precio por noche y por tipo de anfitrión.

Grupo	Variable	Modelo 4 Profesional	Modelo 5 No Profesional
Factores intrínsecos	Habitación privada	-49.813*	-48.222*
	Habitación compartida	-69.046*	-65.353*
	Huéspedes permitidos	10.875*	6.677*
	Habitaciones	4.175*	6.79*
	Baños	-0.758	16.73*
	Amenidades	0.977	0.913*
Factores extrínsecos	Flexibilidad reserva	19.423*	6.426*
	Distancia PDI (100m)	-0.48*	-0.251*
	Competencia	0.171*	0.155*
	Delincuencia	-4.52*	-1.474
	Acceso transporte	-0.857*	-0.316
Factores de plataforma	Superhost	3.333	-2.403*
	Calificación	0.53*	0.193*
	Descripción Anfitrión	-0.005	-0.004*
	Antigüedad (meses)	0.234*	0.171*
	Experiencia (meses)	-0.005	0.128*
	Número reseñas	-0.224*	-0.174*
	Palabras positivas/reseña	6.299*	3.754*
	Palabras negativas/reseña	-0.157	-0.007
Observaciones		5,699	8,111
Coefficiente de determinación R_{adj}^2		0.598	0.603

Nota: * Coeficientes significativos. Ver Tablas B.1 y B.2

profesionales reforzando el argumento de que el mercado mexicano valora la flexibilidad de reserva y es un resultado opuesto al obtenido por (Cai et al., 2019; Lladós-Masllorens et al., 2020) quienes, estudiando Hong Kong y Barcelona respectivamente, argumentan que los huéspedes valoran positivamente a un anfitrión más estricto como señal de profesionalismo. Nuestros resultados muestran un fenómeno en el sentido opuesto, donde la flexibilidad se valora positivamente.

La variable de delincuencia parece de igual forma afectar en mayor proporción a la submuestra de anfitriones profesionales donde el aumento en el número de delitos por cada 100 personas mayores de 18 años tiene un efecto negativo de 4.5% en el precio por noche comparado con 2.7% del modelo 3. El resto de coeficientes no muestra cambios significativos respecto al modelo 3 con la muestra completa, robusteciendo los resultados generales del estudio.

5.3. Análisis con precio compuesto

Con el fin de complementar el análisis y robustecer los resultados generales del estudio, se reestiman los modelos 1-3 (ver Tabla 5.3) modificando la variable dependiente. Esta nueva variable dependiente incorpora al precio por noche una tarifa fija de limpieza (independiente del número de noches reservadas). Este cargo adicional es observado por el consumidor hasta el momento de realizar la reserva y no ha sido considerado por estudios previos. Sin embargo, esta tarifa fija adicional es una herramienta para los anfitriones al separar los costos fijos de mantenimiento y limpieza del ingreso objetivo por noche, por tanto es importante considerarla en el análisis de fijación de precio.

El cargo de limpieza es de \$253.88 en promedio. La Figura 5.1 nos muestra que los anfitriones profesionales tienen en promedio cargos mayores por limpieza que los no profesionales. También podemos observar que existen mucha dispersión para ambos grupos, llegando a máximos de \$7,496, superando incluso al precio por noche para el 3% de los listados.

Dado que este cargo es único por reserva, se debe dividir entre el número de noches reservadas para poder sumarlo al precio por noche. Desafortunadamente no se cuentan con datos sobre el número de noches reservadas y no hay datos publicados de Airbnb que indiquen el número de noches promedio de una reserva en la CDMX o México. Para resolver la falta de datos específicos, tomamos el promedio de 4.3 noches por reserva a nivel internacional publicado por Airbnb para 2020 (“Airbnb Statistics User Growth,” 2021). A partir de estos datos creamos la variable: **precio compuesto** sumando al precio por noche la tarifa de limpieza dividida entre 4.3 noches. La Figura 5.2 nos muestra las distribuciones del precio por noche y el precio compuesto por noche, podemos observar que el precio compuesto hace más pesada la cola de la distribución.

La Tabla 5.6 muestra los coeficientes transformados para la especificación del Modelo 3 con el nuevo precio compuesto (ver en Apéndice la Tabla B.3 la regresión completa). Podemos observar que los coeficientes se mantienen en general estables al utilizar la variable dependiente alternativa. Destaca que los coeficientes de determinación ajustados R_{adj}^2 de los modelos 7 y 8 aumentan respecto a los modelos 2 y 3 y se reduce en dos puntos porcentuales el aporte de la flexibilidad de reserva al precio compuesto por noche respecto al precio regular. El resto de coeficientes no muestra cambios significativos respecto a las regresiones de la Tabla 5.3, mostrando que los resultados son robustos ante especificaciones alternativas del precio por noche.

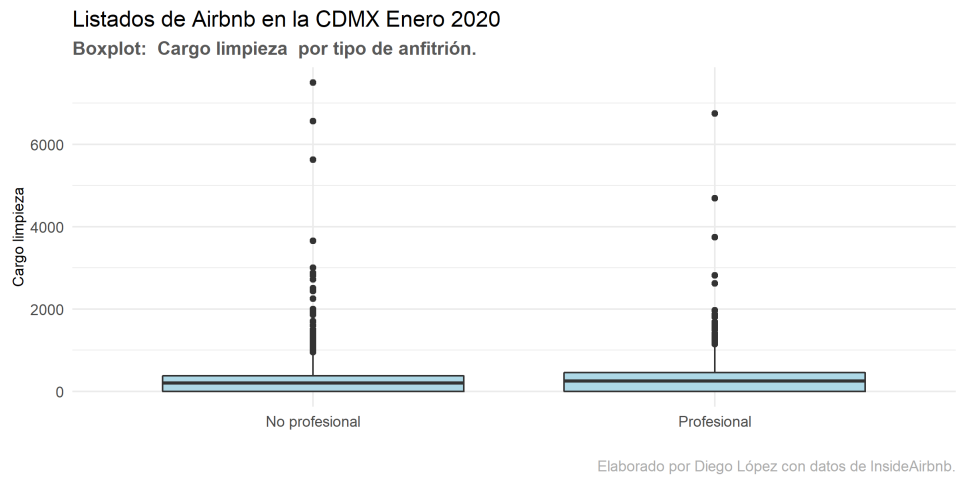


Figura 5.1: Cargo de limpieza por tipo de anfitrión.

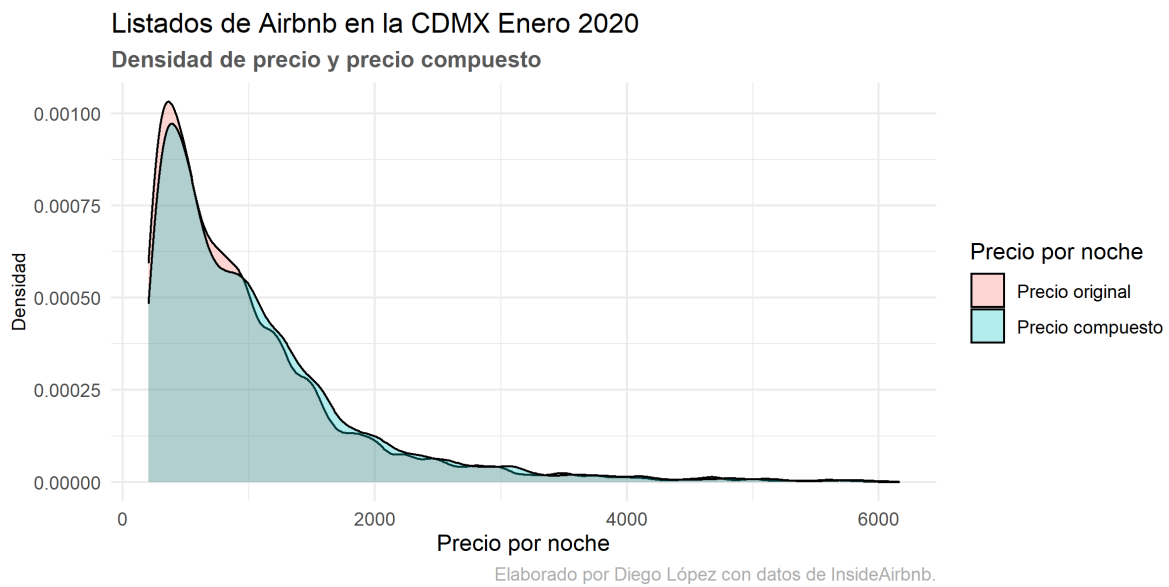


Figura 5.2: Distribución de precio y precio compuesto.

Tabla 5.6: Coeficientes como aporte porcentual al precio compuesto por noche.

Grupo	Variable	Modelo 6	Modelo 7	Modelo 8
Factores intrínsecos	Habitación privada	-50.385*	-49.095*	-49.549*
	Habitación compartida	-70.459*	-67.984*	-69.183*
	Huéspedes permitidos	8.703*	10.215*	10.194*
	Habitaciones	0.049	1.517	1.739
	Baños	7.29*	6.209*	4.866*
	Amenidades	1.101*	1.043*	1.035*
Factores extrínsecos	Flexibilidad reserva		1.174	9.314*
	Distancia PDI (100m)		-0.341*	-0.333*
	Competencia		0.187*	0.16*
	Delincuencia		-2.145	-2.632*
	Acceso transporte		-0.802*	-0.653*
Factores de plataforma	Profesional			15.94*
	Superhost			0.301
	Calificación			0.319*
	Descripción Anfitrión			-0.005*
	Antigüedad (meses)			0.219*
	Experiencia (meses)			0.092*
	Número reseñas			-0.227*
	Palabras positivas/reseña			5.091*
	Palabras negativas/reseña			0.349
Observaciones		13,810	13,810	13,810
Coeficiente de determinación R_{adj}^2		0.542	0.583	0.619

Nota: * Coeficientes significativos. Ver Tabla B.3

Capítulo 6

Conclusiones

La presente investigación aporta a la literatura de precios hedónicos para Airbnb mostrando las particularidades del mercado de la Ciudad de México e incorporando por primera vez al conjunto de variables explicativas los niveles de delincuencia, el acceso a rutas de transporte colectivo y el sentimiento de las reseñas.

Se encuentra que ofrecer un espacio completo tiene un impacto importante en el precio de un listado. Ofrecer, en cambio, una habitación privada reduce el precio en 50 % y una habitación compartida en 69 % respecto al listado completo. Una mayor flexibilidad en la reserva (no solicitar depósitos de garantía, no tener un número mínimo de noches, no solicitar documentos del huésped) está asociado con un incremento de 3.5 % en el precio. Esto indica que el mercado mexicano valora la flexibilidad de reserva y es un resultado opuesto al obtenido por (Cai et al., 2019; Lladós-Masllorens et al., 2020) quienes, estudiando Hong Kong y Barcelona respectivamente, argumentan que los huéspedes valoran positivamente un anfitrión más estricto.

Ser un anfitrión profesional (con más de dos listados registrados) permite fijar un precio 16.2 % por encima de un listado equivalente administrado por un anfitrión no profesional, corroborando la hipótesis inicial de que los anfitriones profesionales cuentan con cierta ventaja al momento de fijar un precio, dado que obtienen en promedio más reservas por mes a un mayor precio para listados con los mismos atributos.

El aumento de una unidad en el número de delitos por cada 100 adultos (tasa de incidencia delictiva) tiene un efecto negativo de 2.7 % en el precio por noche. Este resultado muestra que la delincuencia es una variable relevante en la fijación de precio en la CDMX, que tiene una tasa promedio de 3.36 delitos por cada 100 adultos. Contar con un acceso adicional a redes de transporte colectivo (Metrobús o Metro) en la colonia reduce el precio en 0.65 %, Una

posible explicación de este resultado es la correlación positiva (0.39) entre la delincuencia y el acceso a rutas de transporte en la CDMX.

El análisis de sentimiento de reseñas identifica que una palabra positiva adicional refleja un incremento de 4.97% en el precio. Las palabras negativas muestran un impacto negativo pero no significativo y de menor magnitud (-0.24%). Esto indica que los usuarios de Airbnb valoran más las palabras positivas en las reseñas y los anfitriones explotan esta valoración con un incremento en el precio. En cambio, las palabras negativas parecen no afectar en el precio.

Los resultados del modelo principal son robustos al estimar por separado con las submuestras por tipo de anfitrión y al utilizar el precio compuesto (que incorpora la tarifa fija de limpieza) como variable independiente.

En relación a las limitaciones del estudio, la metodología implementada no es dinámica y no captura efectos estacionales. De acuerdo con los datos disponibles al momento de la recolección (marzo - 2019 - enero 2020), la demanda de Airbnb en la CDMX se mantiene relativamente estable durante el año y no se observa un aumento significativo en verano o fin de año. No se espera que efectos estacionales influyan significativamente en los resultados. La mejor forma de atender esta limitación es contar con un periodo de estudio más amplio (cuando menos de un año) para identificar variaciones temporales, tomando en cuenta el impacto en la industria del turismo y hospedaje que representa la pandemia Covid-19. Una limitación adicional es que los atributos seleccionados no incorporan características sociales-demográficas de los anfitriones. Pueden existir factores específicos al anfitrión que afecten al proceso de fijación de precios, como puede ser la raza, género y edad.

Se mantiene prudencia respecto a si los hallazgos del estudio se pueden generalizar a otras ciudades de México, donde la dinámica de Airbnb podría ser diferente. Estas diferencias pueden incluir la cuota de mercado de anfitriones profesionales y no profesionales y el grado en que los hoteles compiten con Airbnb, lo que puede distorsionar el precio de mercado de los listados.

Las limitaciones identificadas presentan una oportunidad de investigación a futuro para tener una mejor comprensión del fenómeno alrededor de Airbnb y del proceso de fijación de precios en mercados bilaterales.

Anexo A

Anexo

A.1. Fundamentos del modelo de regresión hedónica

La teoría de precios hedónicos (Lancaster, 1966) tiene como hipótesis que el valor de un bien se puede descomponer en función de los atributos que lo componen, (Rosen, 1974) formaliza esta teoría como un problema de optimización en la cual las decisiones del consumidor y productor en el espacio de los atributos de un bien heterogéneo son determinadas por los precios implícitos o precios hedónicos. A pesar de que no existe un mercado para cada uno de los atributos porque se consumen en conjunto, es posible inferir estos precios implícitos como la disposición marginal a pagar por el atributo.

Esta interacción de consumidores y productores en un espacio de atributos se describe en un modelo competitivo de equilibrio general de la siguiente forma:

Las decisiones del consumidor y productor permiten modelar la relación del bien con sus atributos a través de la función de precios hedónicos $P = F(Z)$ donde Z representa el bien heterogéneo. La forma funcional de P depende de las preferencias de los consumidores y de la función de costos del productor de la siguiente forma:

El bien heterogéneo Z está compuesto de j atributos, por tanto es un vector j -dimensional $Z = (z_1, z_2, \dots, z_j)$ donde z_j representa la cantidad de atributo j en el bien heterogéneo y el precio del bien en el mercado está en función de los j atributos $P(z_1, z_2, \dots, z_j)$ creciente en todos sus argumentos. El vector de precios de equilibrio P^* que vacía el mercado es aquel que iguala la oferta y demanda de los atributos y está determinado por las preferencias de consumidores y costos de los productores.

El problema del consumidor

El consumidor tiene una utilidad $U(z_1, z_2, \dots, z_j, x)$ por el consumo del bien heterogéneo y un bien compuesto x , es decir un bien representativo de todos los demás bienes con precio unitario $P_x = 1$. Se supone que U es estrictamente cóncava. Para maximizar su utilidad debe elegir la combinación de atributos y bien compuesto que maximice U sujeto a un ingreso M .

El problema del consumidor es

$$\begin{aligned} & \max_{z_1, z_2, \dots, z_j, x} U(z_1, z_2, \dots, z_j, x) \\ \text{s.a.} \quad & P(z_1, z_2, \dots, z_j) + x = M \end{aligned}$$

El problema se resuelve con las siguientes j condiciones de primer orden:

$$\frac{U_{z_j}}{U_x} = P_{z_j}, \forall j$$

La derivada de la función de precios respecto a el atributo j es el precio de el atributo j . Las condiciones de segundo orden se cumplen bajo el supuesto de U estrictamente cóncava.

Siguiendo a (Rosen, 1974), definimos una función de demanda $D(Z; u, M)$ implícitamente como:

$$U(Z, M - D) = u$$

Donde $D(Z; u, M)$ define una familia de planos de indiferencia que relacionan el atributo z_j con las unidades de x a las que se renuncian. $D(Z; u, M)$ se interpreta como la cantidad que un consumidor está dispuesto a pagar por distintas combinaciones de características dado un nivel de utilidad u y su ingreso M , mientras que $P(Z)$ es el precio menor que debe pagar por la combinación Z en el mercado.

La utilidad se maximiza en $D(Z^*; u^*, M) = P(Z^*)$ y se cumple:

$$D_{z_j}(Z^*; u^*, M) = P_{z_j}(Z^*) \forall j$$

La tasa a la que el consumidor paga por el atributo z_j es igual al precio P_{z_j} . En el óptimo, $P(Z)$ es tangente a $D(Z; u^*, M)$. Dado que estamos en un entorno competitivo, la función $P(Z)$ que enfrentan los consumidores es la misma e independiente de M . La heterogeneidad de los consumidores viene de las preferencias sobre el bien diferenciado Z .

El problema del productor

El productor debe decidir el conjunto de atributos Z del bien a producir y la cantidad, sea $N(Z)$ la cantidad de bienes heterogéneos Z . Simplificando el modelo, cada empresa produce sólo un bien Z . Los costos de que la firma enfrenta vienen dados por la siguiente función de costos que relaciona N, Z y la tecnología asociada representada por un parámetro β :

$$C(N, Z; \beta)$$

Se asume que C es convexa con $C(0, Z) = 0$ y $C_N > 0$, $C_{z_j} > 0$.

Cada empresa busca maximizar sus beneficios resolviendo:

$$\max_{z_1, z_2, \dots, z_j, N} \pi = NP(Z) - C(N, Z; \beta)$$

Dado el supuesto de competencia las empresas son agentes atómicos en el mercado de bienes heterogéneos y $P(Z)$ es independiente de N . Las decisiones óptimas de Z y N implican:

$$P_{z_j}(Z) = \frac{C_{z_j}(N, Z)}{N}, \forall j$$

$$P(Z) = C_N(N, Z)$$

En el óptimo, el ingreso marginal de un atributo adicional es igual al costo marginal unitario del atributo. Y se producen unidades hasta que el ingreso unitario $P(Z)$ se iguala al costo marginal de producción evaluado en la combinación óptima de atributos.

Al igual que la demanda, se define una función de oferta $S(Z; \pi, \beta)$ implícitamente como:

$$\pi = N \cdot S(Z; \pi, \beta) - C(N, Z; \beta)$$

$$C_N(N, Z) = S(Z; \pi, \beta)$$

$S(Z; \pi, \beta)$ define una familia de planos de indiferencia que indican el precio unitario por la combinación de atributos Z que la empresa está dispuesta a aceptar dado un nivel de beneficios π y una tecnología β cuando las cantidades N de cada combinación Z se eligen de forma óptima.

La empresa optimiza en

$$S_{z_j} = \frac{C_{z_j}(N, Z)}{N} > 0 \forall j \text{ y } S_\pi = \frac{1}{N} > 0$$

El precio de reserva marginal para el atributo j dado un nivel de beneficio π es S_{z_j} . Dado que $S(Z; \pi, \beta)$ es el precio de venta que el productor está dispuesto a aceptar por las características Z al nivel de beneficio π , mientras que $P(Z)$ es el máximo precio alcanzable por un bien con características Z en el mercado, por tanto los beneficios se maximizan en:

$$P_{z_j}(Z^*) = S_{z_j}(Z^*; \pi^*, \beta) \quad \forall j$$

El equilibrio del productor está caracterizado por la tangencia del plano de atributos-beneficios $S(Z; \pi^*, \beta)$ y el plano atributos-precio $P(Z)$

Equilibrio de mercado

En equilibrio, el vector de precios hedónicos $P(Z)$ es el plano tangente a la demanda $D(Z; u^*, M)$ y oferta $S(Z; \pi^*, \beta)$ de bienes heterogéneos. Es el vector de precios que vacía los mercados implícitos de los j atributos disponibles. De forma que $Q^D(Z) = Q^S(Z)$.

Es a partir de este modelo competitivo que podemos, a través de diferenciar parcialmente de la función de precios hedónicos $P(Z)$ con respecto a cada característica z_j , obtener el precio marginal implícito o precio sombra que representa la disposición marginal media de los compradores a pagar por una característica particular y la disposición marginal media de los productores a aceptar por el atributo.

$$\frac{\partial P(Z)}{\partial z_j} = \text{Disposición marginal a pagar/aceptar por el atributo } z_j$$

Apéndice B

Apéndice

B.1. Regresiones por tipo de anfitrión

Se presentan las regresiones hedónicas bajo la especificación del modelo 3 (Ver Tabla 5.3) para las submuestras por tipo de anfitrión: Profesional y No Profesional. A partir de estos coeficientes se calculan los aportes porcentuales al precio descritos en la Tabla 5.5

B.2. Regresiones con precio compuesto.

Se presenta la regresión hedónica bajo la misma especificación del modelo 3 (Ver Tabla 5.3) cambiando la variable dependiente por el precio compuesto. A partir de estos coeficientes se calculan los aportes porcentuales al precio descritos en la Tabla 5.6.

Tabla B.1: Regresión hedónica para anfitriones profesionales

	Variable dependiente: Logaritmo de precio por noche. Modelo 4 Profesional
Habitación privada c.r.a. listado entero	-0.689*** (0.028)
Habitación compartida c.r.a. listado entero	-1.173*** (0.129)
Huéspedes permitidos	0.103*** (0.011)
Habitaciones	0.041* (0.025)
Baños	-0.008 (0.016)
Amenidades	0.010*** (0.001)
Flexibilidad reserva	0.178*** (0.038)
Distancia PDI (100m)	-0.005*** (0.001)
Competencia	0.002*** (0.0004)
Delincuencia	-0.046*** (0.016)
Acceso transporte	-0.009*** (0.003)
Superhost	0.033 (0.023)
Calificación	0.005*** (0.002)
Descripción Anfitrión	-0.00005 (0.00004)
Antigüedad (meses)	0.002** (0.001)
Experiencia (meses)	-0.00005 (0.001)
Número reseñas	-0.002*** (0.0003)
Palabras positivas/reseña	0.061*** (0.009)
Palabras negativas/reseña	-0.002 (0.016)
Observations	5,699
Adjusted R ²	0.598
F Statistic	447.113***

Note:

*p<0.1; **p<0.05; ***p<0.01
 Errores estándar robustos agrupados a nivel colonia.
 Tipo HC1. Clusters = 450 colonias.
 Se omite el intercepto.

Tabla B.2: Regresión hedónica para anfitriones no profesionales

	Variable dependiente: Logaritmo de Precio por noche. Modelo 5 No Profesional
Habitación privada c.r.a. listado entero	-0.658*** (0.018)
Habitación compartida c.r.a. listado entero	-1.060*** (0.124)
Huéspedes permitidos	0.065*** (0.009)
Habitaciones	0.066*** (0.024)
Baños	0.155*** (0.029)
Amenidades	0.009*** (0.001)
Flexibilidad reserva	0.062*** (0.014)
Distancia PDI (100m)	-0.003*** (0.001)
Competencia	0.002*** (0.0003)
Delincuencia	-0.015 (0.012)
Acceso transporte	-0.003 (0.002)
Superhost	-0.024** (0.010)
Calificación	0.002* (0.001)
Descripción Anfitrión	-0.00004** (0.00002)
Antigüedad (meses)	0.002*** (0.0004)
Experiencia (meses)	0.001*** (0.0003)
Número reseñas	-0.002*** (0.0002)
Palabras positivas/reseña	0.037*** (0.007)
Palabras negativas/reseña	-0.0001 (0.012)
Observations	8,111
Adjusted R ²	0.603
F Statistic	650.050***

Note:

*p<0.1; **p<0.05; ***p<0.01
 Errores estándar robustos agrupados a nivel colonia.
 Tipo HC1. Clusters = 730 colonias.
 Se omite el intercepto.

Tabla B.3: Regresiones hedónicas con precio compuesto

	Variable dependiente: logaritmo de precio compuesto por noche.		
	Modelo 6	Modelo 7	Modelo 8
Habitación privada c.r.a. listado entero	-0.701*** (0.020)	-0.675*** (0.017)	-0.684*** (0.016)
Habitación compartida c.r.a. listado entero	-1.219*** (0.095)	-1.139*** (0.094)	-1.177*** (0.093)
Huéspedes permitidos	0.083*** (0.008)	0.097*** (0.010)	0.097*** (0.009)
Habitaciones	0.0005 (0.020)	0.015 (0.020)	0.017 (0.019)
Baños	0.070*** (0.022)	0.060*** (0.023)	0.048** (0.021)
Amenidades	0.011*** (0.001)	0.010*** (0.001)	0.010*** (0.001)
Flexibilidad reserva		0.012 (0.019)	0.089*** (0.018)
Distancia PDI (100m)		-0.003*** (0.001)	-0.003*** (0.001)
Competencia		0.002*** (0.0003)	0.002*** (0.0003)
Delincuencia		-0.022 (0.015)	-0.027** (0.013)
Acceso transporte		-0.008*** (0.003)	-0.007*** (0.002)
Profesional			0.148*** (0.013)
Superhost			0.003 (0.011)
Calificación			0.003*** (0.001)
Descripción Anfitrión			-0.00005* (0.00003)
Antigüedad (meses)			0.002*** (0.0005)
Experiencia (meses)			0.001*** (0.0003)
Número reseñas			-0.002*** (0.0002)
Palabras positivas/reseña			0.050*** (0.006)
Palabras negativas/reseña			0.003 (0.010)
Observations	13,810	13,810	13,810
Adjusted R ²	0.542	0.583	0.619
F Statistic	2,719.666***	1,759.450***	1,123.389***

Note:

*p<0.1; **p<0.05; ***p<0.01
 Errores estándar robustos agrupados a nivel colonia.
 Tipo HC1. Clusters = 803 colonias.
 Se omite el intercepto.

Bibliografía

- 2020 Airbnb Update. (2020). In *Airbnb Newsroom*. <https://news.airbnb.com/2020-update/>.
- Airbnb Statistics User and Market Growth Data. (2021). In *iPropertyManagement.com*. <https://ipropertymanagement.com/research/airbnb-statistics>.
- Arvanitidis, P., Economou, A., Grigoriou, G., & Kollias, C. (2020). Trust in peers or in the institution? A decomposition analysis of Airbnb listings' pricing. *Current Issues in Tourism*, 0(0), 1–18. <https://doi.org/10.1080/13683500.2020.1806794>
- Cai, Y., Zhou, Y., Ma, J., & Scott, N. (2019). Price Determinants of Airbnb Listings: Evidence from Hong Kong. *Tourism Analysis*, 24, 227–242. <https://doi.org/10.3727/108354219X15525055915554>
- Chattopadhyay, M., & Mitra, S. K. (2019). Do airbnb host listing attributes influence room pricing homogenously? *International Journal of Hospitality Management*, 81, 54–64. <https://doi.org/10.1016/j.ijhm.2019.03.008>
- Chen, Y., & Xie, K. (2017). Consumer Valuation of Airbnb Listings: A Hedonic Pricing Approach. *International Journal of Contemporary Hospitality Management*, 29. <https://doi.org/10.1108/IJCHM-10-2016-0606>
- de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, 33(2), 124–135.
- Dudás, G., Vida, G., Kovalcsik, T., & Boros, L. (2017). A socio-economic analysis of Airbnb in New York City. *Regional Statistics*, 7, 135–151. <https://doi.org/10.15196/RS07108>
- Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2017). Pricing in the sharing economy: A hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing*, 1–11. <https://doi.org/10.1080/10548408.2017.1308292>
- Gunter, U., & Onder, I. (2018). Determinants of Airbnb Demand in Vienna and their Implications for the Traditional Accommodation Industry. *Tourism Economics*, 24. <https://doi.org/10.1177/1354816617731196>
- Guttentag, D. (2019). Progress on Airbnb: A literature review. *Journal of Hospitality and Tourism Technology*, 10(4), 814–844. <https://doi.org/10.1108/JHTT-08-2018-0075>
- IMCO. (2018). *Airbnb en la Ciudad de México: Caso de estudio para la capital del país* ({Report} No. 1; p. 24). IMCO.
- Inside Airbnb. Adding data to the debate. (2021). In *Inside Airbnb*. <http://insideairbnb.com>.
- Kakar, V., Voelz, J., Wu, J., & Franco, J. (2018). The Visible Host: Does race guide Airbnb rental rates in San Francisco? *Journal of Housing Economics*, 40, 25–40. <https://doi.org/10.1016/j.jhe.2017.08.001>

- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.
- Lladós-Masllorens, J., Meseguer-Artola, A., & Rodríguez-Ardura, I. (2020). Understanding Peer-to-Peer, Two-Sided Digital Marketplaces: Pricing Lessons from Airbnb in Barcelona. *Sustainability*, 12(13), 5229. <https://doi.org/10.3390/su12135229>
- Lorde, T., Jacob, J., & Weekes, Q. (2019). Price-setting behavior in a tourism sharing economy accommodation market: A hedonic price analysis of AirBnB hosts in the caribbean. *Tourism Management Perspectives*, 30, 251–261. <https://doi.org/10.1016/j.tmp.2019.03.006>
- López-Gay, A., Oca, Á. M.-M. de, Sales-Favà, J., & Cunha, J. M. P. da. (2019). Apartamentos turísticos, ciudad y población en América Latina. Los casos de Ciudad de México y São Paulo. *Revista Latinoamericana de Población*, 13(25), 5–35. <https://doi.org/10.31406/relap2019.v13.i2.n25.1>
- López-Tamayo, D. (2019). *Plataformas digitales en México: teoría, competencia y regulación. Ride-sharing en la Ciudad de México* [Undergraduate Thesis]. Universidad Nacional Autónoma de México.
- Núñez, H. M., Paredes, D., & Garduño-Rivera, R. (2017). Is crime in Mexico a disamenity? Evidence from a hedonic valuation approach. *The Annals of Regional Science*, 59(1), 171–187. <https://doi.org/10.1007/s00168-017-0823-8>
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Índice de tablas

3.1. Puntos de interés	12
3.2. Delitos en CDMX. Carpetas de investigación.	14
3.3. Grupos y descripción de variables	16
4.1. Top 10 listados por número de reseñas. Airbnb Enero 2020.	23
4.2. Bottom 10 listados por número de reseñas. Airbnb Enero 2020.	24
4.3. Estadística descriptiva de variables	26
5.1. Matriz de correlación 1/2	28
5.2. Matriz de correlación 2/2	28
5.3. Regresiones hedónicas	31
5.4. Coeficientes como aporte porcentual al precio por noche.	32
5.5. Coeficientes como aporte porcentual al precio por noche y por tipo de anfitrión.	33
5.6. Coeficientes como aporte porcentual al precio compuesto por noche.	36
B.1. Regresión hedónica para anfitriones profesionales	44
B.2. Regresión hedónica para anfitriones no profesionales	45
B.3. Regresiones hedónicas con precio compuesto	46

Índice de figuras

3.1. Mapa de listados, accesos a transporte y puntos de interés.	14
3.2. Mapa CDMX. Tasa de incidencia delictiva.	15
4.1. Evolución de demanda 2012 - 2020 Airbnb CDMX.	20
4.2. Listados por alcaldía y colonia	20
4.3. Tipo de Anfitrión: Listados por alcaldía y colonia	21
4.4. Mapa Precio promedio por alcaldía y colonia	22
4.5. Listados por tipo de propiedad y alcaldía.	23
4.6. Wordcloud de principales 100 términos en reseñas.	24
5.1. Cargo de limpieza por tipo de anfitrión.	35
5.2. Distribución de precio y precio compuesto.	35

