

Análisis de características influyentes en los precios de alojamientos de Airbnb en la ciudad de Washington D.C usando modelos de regresión

3 de abril de 2023



Universidad
Internacional
de Valencia

Titulación:

Maestría Oficial en Big Data
y Ciencia de Datos

Curso académico

2022 – 2023

Alumno/a:

Carvajal Barreto, Karen
D.N.I: 1.143.115.303

Director/a de TFM:

Edith Cecilia Macedo

Convocatoria:

Segunda

De:

 Planeta Formación y Universidades

Índice

| | |
|---|----|
| Resumen | 5 |
| 1. Introducción | 6 |
| 2. Objetivos..... | 7 |
| 2.1. Generales | 7 |
| 2.2. Específicos | 7 |
| 3. Estado del Arte y Marco teórico | 8 |
| 4. Desarrollo del proyecto y resultados | 11 |
| 4.1. Ciudad de caso de estudio | 11 |
| 4.2. Fuentes de datos | 11 |
| 4.3. Limpieza y transformación de los datos | 13 |
| 4.4. Descripción de los datos..... | 16 |
| 4.5. Selección de características | 22 |
| 4.6. Modelos de regresión | 25 |
| 4.7. Resultados | 26 |
| 5. Conclusión y trabajos futuros..... | 32 |
| 6. Referencias | 33 |
| Apéndice I..... | 36 |
| Anexos I..... | 37 |

Índice de ilustraciones

| | |
|---|----|
| Ilustración 1. Distribución de precios de los alojamientos. | 14 |
| Ilustración 2. Tipo de habitación..... | 40 |
| Ilustración 3. Tipo de habitación vs precio. | 40 |
| Ilustración 4. Número de personas permitidas en el alojamiento vs precio | 41 |
| Ilustración 5. Correlación de las características del alojamiento. | 41 |
| Ilustración 6. Correlación de las características de servicios adicionales. | 42 |
| Ilustración 7. Calificaciones por tipo de habitación. | 42 |
| Ilustración 8. Correlación de las características de calificaciones. | 43 |
| Ilustración 9. Host es superhost. | 43 |
| Ilustración 10. Identidad y si el host tiene imagen de perfil. | 44 |
| Ilustración 11. Promedio de precios de alojamientos de Airbnb por barrio. | 44 |
| Ilustración 12. Cantidad de alojamientos de Airbnb por barrio en la ciudad. | 45 |
| Ilustración 13. Puntaje promedio de alojamientos de Airbnb por barrio en la ciudad. ... | 46 |
| Ilustración 14. Promedio de disponibilidad por barrio en la ciudad. | 47 |
| Ilustración 15. Promedio de noches mínimas para hacer reserva por barrio. | 48 |
| Ilustración 16. Correlación de variables del host. | 50 |
| Ilustración 17. Correlación de variables de ubicación..... | 51 |

Índice de tablas

| | |
|--|----|
| Tabla 1. Estadística descriptiva de las principales características..... | 38 |
| Tabla 2. Cantidad de tipos de propiedad..... | 39 |
| Tabla 3. Reserva inmediata en los alojamientos. | 39 |
| Tabla 4. Resumen de estadísticas por barrios. | 50 |
| Tabla 5. Determinantes del precio por noche para OLS y QR..... | 54 |
| Tabla 6. Porcentajes de incremento / disminución en el precio..... | 55 |
| Tabla 7. P-valores obtenidos de cada modelo de regresión..... | 56 |

Resumen

El presente trabajo se trata del análisis de características de los alojamientos turísticos de Airbnb en la ciudad de Washington DC usando modelos de Regresión OLS y Regresión Cuantílica que ayudarán a identificar las variables influyentes en los precios.

Inicialmente se obtendrán los datos con el fin ser limpiados y transformados de modo que estos se encuentren listos para que los modelos de regresión arrojen resultados óptimos, luego se hará un análisis estadístico de todas las características encontradas en la literatura en donde se examinarán para poder hacer un resumen de estas y decidir cuáles serán enviadas a los modelos de regresión y obtener los resultados del comportamiento de estas características con respecto al precio de los alojamientos de Airbnb en la ciudad de Washington DC.

Se usará el lenguaje de programación Python el cual es un lenguaje muy usado en ciencia de datos para programar algoritmos de Machine Learning y hacer estadística descriptiva de un conjunto de datos, se usarán una serie de librerías para ciencia de datos que permitirán obtener, limpiar, transformar, crear modelos, entre otras funciones. Esto se hará en un notebook en Jupyter.

Palabras clave: Airbnb, Características del barrio, Washington D.C, Precio de alojamientos, Datos espaciales, Regresión Cuantílica, OLS.

1. Introducción

El sector del turismo actualmente está pasando por un momento de muchos cambios y evolución, en los últimos años ha nacido un nuevo mercado que consiste en alquilar alojamientos o espacios que no se están usando y se le quieran sacar alguna utilidad, este mercado se ha convertido en uno de los más importantes en esta industria, permitiéndole a muchas personas rentar sus propiedades de una forma diferente a la que se había conocido hasta ahora. Es por esta razón que nació Airbnb, la cual es una plataforma que ha revolucionado el mercado de las rentas de este tipo de alojamientos y esto les permite a muchos viajeros acceder precios y a opciones diferentes a las habituales que se manejaban que eran, en su gran mayoría solo hoteles.

El análisis de precios en el mercado de alojamientos es de gran importancia en la industria turística. Por esto, la plataforma Airbnb se ha convertido en una herramienta clave para la renta de alojamientos ya sea de forma temporal o por un tiempo extendido en ciudades alrededor del mundo. En este Trabajo de Fin de Máster se aborda el análisis de características influyentes en los precios de alojamientos de Airbnb en la ciudad de Washington D.C, utilizando modelos de regresión para identificar los factores y características que tienen mayor impacto en la determinación de los precios. El objetivo es proporcionar una mejor comprensión de los factores que influyen en los precios de alojamientos de Airbnb en esta ciudad, ya sea que proporcionen aumento o disminución de estos, lo que puede ser de gran interés para los propietarios de alojamientos y los viajeros que buscan opciones de alojamiento económicas y de alta calidad. A través del uso de modelos de regresión se pretende lograr el análisis de datos y la identificación de las variables que tienen mayor influencia en la determinación de precios en el mercado de alojamientos de Airbnb y poder obtener una mejor comprensión de las variables que influyen en los precios.

En resumen, este trabajo de fin de máster tiene como finalidad contribuir al conocimiento sobre el mercado de alojamientos de Airbnb en Washington D.C. y proporcionar una visión más clara de los factores que influyen en la fijación de precios en esta ciudad. De esta manera, se espera que los resultados obtenidos puedan ser útiles para los propietarios de alojamientos, los viajeros y cualquier persona interesada en este ámbito.

2. Objetivos

Los objetivos establecidos para este proyecto son los siguientes:

2.1. Generales

Determinar qué características influyen en los precios de los alojamientos de Airbnb en la ciudad de Washington D.C.

2.2. Específicos

- Hacer un análisis exploratorio de los datos de Airbnb en la ciudad de Washington D.C usando estadística descriptiva.
- Usar métodos de regresión para analizar el impacto en el precio de las variables independientes.
- Identificar patrones o relación entre las características y los precios de los alojamientos de Airbnb.
- Estimar un modelo de regresión por mínimos cuadrados (OLS) para hallar aquellas características que influyen en el precio
- Estimar un modelo de regresión cuantílica (QR) para hallar aquellas características que influyen en el precio.
- Comparar los resultados del modelo de regresión por mínimos cuadrados y el modelo de regresión cuantílica.

3. Estado del Arte y Marco teórico

Es importante conocer el concepto de economía colaborativa, este se refiere a un sistema económico en el que las personas comparten bienes y servicios en lugar de poseerlos exclusivamente. Según (Dudás et al., 2020), en los últimos años, este tipo de economía se ha vuelto cada vez más popular debido a la globalización y a la crisis económica ya que esto provoca cambios a los que las personas y los mercados van adaptándose y en nuestros días, gracias al acceso a la tecnología y plataformas digitales, se facilita más la conexión entre personas que desean compartir bienes y servicios, se proveen alternativas para el consumo de recursos como el alquiler de bienes propios.

Airbnb es una plataforma fundada en el 2008 para alquiler de inmuebles, la cual está incluida dentro de este concepto de economía colaborativa porque los propietarios comparten o alquilan sus propiedades con personas que por lo general son viajeros, el modelo de negocio consiste en que una persona privada que posea algún inmueble o alojamiento ya sea una casa, un apartamento, una habitación, una cabaña, un edificio, etc., le coloca un precio y lo alquila a otra persona que lo requiera. En Airbnb el dueño del alojamiento es llamado “anfitrión” y la persona interesada en alquilar se le llama “huésped”. La ventaja de este servicio es que es intuitiva y fácil de usar, brinda la facilidad de que una persona que quiera alquilar lo haga rápidamente y sin tanto trámite, en general, esta plataforma brinda una cierta seguridad y confiabilidad en cuanto a los pagos y la comunicación entre el huésped y el anfitrión. Además, al alquilar una casa en lugar de un hotel, los huéspedes podrán experimentar y ver cómo es la vida de los locales. Airbnb ha sido un gran revolucionario de la industria de alquiler de viviendas y es un ejemplo de cómo la economía colaborativa funciona, se ha vuelto bastante popular por la gran cantidad de variedad de alojamientos y presupuestos que se adaptan a cualquier viajero, se ha convertido en el favorito sobre todo para aquellas personas que quieren vivir sus vacaciones en un lugar más natural a la cultura del sitio, esto ha cambiado la forma en que las personas planean sus viajes y también ha traído muchos beneficios para la economía local porque le ha permitido a los propietarios generar ingresos extras alquilando aquellas viviendas que quizá no están usando, incluso algunos anfitriones siguen viviendo en la propiedad junto a los huéspedes y colocan en alquiler solo una parte de esta.

En Airbnb el anfitrión es quien fija el precio del alojamiento por noche, por lo tanto, este varía mucho dependiendo de varios factores, al ser algo de suma importancia definir bien estos precios, han surgido diversos estudios de aquellos factores que influyen en el precio de una propiedad que está basada en la economía colaborativa, (Wang & Nicolau, 2017) menciona que algunas características del anfitrión, características del alojamiento, algunos servicios extra que brindan comodidad al usuario, Número de reseñas y calificaciones que los usuarios colocan en la página, son tenidas en cuenta como indicadores que afectan el precio. Por otro lado, (Teubner et al., 2017) indica que características que implican la reputación del host, el puntaje

que colocan los usuarios, incluso el atributo de si una propiedad tiene una imagen o no, juegan un papel significativo en los precios de los alojamientos. En el estudio realizado por (Toader et al., 2021) confirma que entre más cerca esté la propiedad del centro de la ciudad el precio aumenta, también indica que entre más comodidades y servicios tenga el alojamiento el precio se incrementa, así mismo con las características de la propiedad como baños, camas, etc.

Además de todo lo mencionado anteriormente, también se deben tener en cuenta características externas de la ciudad y su ubicación, así lo indica en su estudio (Perez-Sanchez et al., 2018), en donde se tiene en cuenta factores que rodean el ambiente de la localidad, se tiene en cuenta la distancia del alojamiento con respecto al centro de la ciudad, lugares culturales y recreacionales, incluso en la ciudad elegida se consideró la distancia con respecto a la costa.

Actualmente existen varios modelos para analizar posibles variables que afecten el precio de una vivienda, los modelos de precios hedónicos son modelos económicos que ayudan a establecer el valor de un bien o un servicio, en este se asume que el valor está determinado por sus características individuales (Limsombunchai, 2004).

Existen dos ramas bastante conocidas de aprendizaje automático: aprendizaje supervisado y no supervisado. El aprendizaje no supervisado es una técnica que se enfatiza en analizar y descubrir patrones y relaciones en un conjunto de datos sin tener los datos etiquetados y el modelo se encarga de predecir estas etiquetas. Las principales técnicas de este método es la clusterización de datos, reducción de dimensionalidad, entre otras. Las más conocidas son KMeans y DBSCAN para clustering y análisis de componentes principales para la reducción de la dimensionalidad. Por otro lado, en el aprendizaje supervisado se utilizan datos que están previamente etiquetados, es decir, donde existe la variable que se quiere predecir y con estos datos etiquetados se entrena el modelo para luego predecir o clasificar datos. Algunas aplicaciones del aprendizaje supervisado es la clasificación en la cual la variable dependiente es una variable categórica, y la regresión en donde la variable dependiente es una variable numérica continua, algunos modelos que aplican a ambas son los árboles de decisión, random forest, support vector machines. (Bonaccorso, n.d.)

Existen varias herramientas y lenguajes de programación como Python, el cual es potente cuando se usa para ciencia de datos, y algoritmos de Machine Learning como la regresión por cuantiles, que es una alternativa a la regresión lineal y es un poco más robusta que otros modelos cuando se usan datos que no están distribuidos linealmente y dispersos con respecto a la media, también cuando hay muchos outliers, además permite elaborar varios modelos con distintas condiciones para la variable dependiente (Perez-Sanchez et al., 2018). La regresión por cuantiles estima la función de distribución condicional de la variable dependiente en diferentes puntos del rango, a diferencia de la regresión tradicional, la cual estima la media de la variable dependiente. La regresión por cuantiles explica la relación entre los cuantiles de la

variable de resultado, en cambio la regresión lineal explica es la relación lineal con respecto a la media de las variables.

La regresión por mínimos cuadrados o OLS es un método que supone que la media de una variable dependiente depende del valor que tomen una o varias variables independientes. En algunas ocasiones, la media podría no ser un buen método para explicar los datos, o hay heterocedasticidad, es decir, la varianza no es constante, lo cual va en contra de lo que se requiere para una regresión por mínimos cuadrados, es por esta razón que se puede usar la regresión cuantílica, porque permite determinar cuantiles para analizar la influencia de ciertas variables independientes en un punto específico de la muestra sobre la variable dependiente.

4. Desarrollo del proyecto y resultados

Este apartado está compuesto por el análisis de precios de los alojamientos de Airbnb para la ciudad de Washington D.C, la capital de Estados Unidos, en donde se extraerán los datos de la página de Inside Airbnb (*Inside Airbnb: Get the Data*, n.d.), se procesarán, transformarán en un formato que ayude a su comprensión mediante estadística descriptiva y se le aplicarán modelos de regresión para su respectivo análisis. Los datos fueron obtenidos el día 2 de febrero de 2023.

4.1. Ciudad de caso de estudio

La ciudad seleccionada para el análisis fue Washington D.C ubicada en Estados Unidos, la cual tiene aproximadamente 671.000 habitantes y antes de pandemia del Covid-19 era uno de los destinos turísticos más populares en este país, este suceso afectó de forma significativa el turismo en todo el mundo y el turismo en la ciudad disminuyó significativamente en 2020 debido a las restricciones de viaje y cierre de atracciones turísticas, aunque en el 2021 se espera que estos datos del turismo vuelvan a sus valores normales (*DOMESTIC VISITATION FORECAST THROUGH 2022 Washington, DC (District of Columbia)*, n.d.). Teniendo en cuenta este evento, se analizaron las estadísticas del 2019 en donde se encontró que se recibieron más de 24.6 millones de visitantes siendo este valor el más alto de los últimos años, estos visitantes gastaron aproximadamente 8 millones de dólares, la mayoría gastaron dinero en alojamiento, comida, compras, entretenimiento y transporte, así mismo, este dinero que gastan las personas ayuda a sostener a más de 70.000 empleos (*Investigación de Visitantes de Washington, DC | Washington DC*, n.d.).

Esta ciudad es conocida como uno de los destinos más populares para visitar durante todo el año en este país. Es un destino turístico para las personas que desean conocer la cultura de este país, ya que, al ser la capital, es conocida por su arquitectura, sus innumerables sitios de interés cultural y monumentos como la casa blanca, el capitolio, los distintos monumentos, diversos museos, entre muchas otras actividades (*Experience DC | Seek out the Unexpected in Washington, DC*, n.d.).

Este análisis se enfocará en analizar los principales factores que afectan en los precios de Airbnb en esta ciudad desde el punto de vista geográfico y urbano.

4.2. Fuentes de datos

La principal fuente de datos de alojamientos de Airbnb de la ciudad de Washington DC proviene de la página Inside Airbnb (*Inside Airbnb: Get the Data*, n.d.), la cual es una página web que se encarga recopilar, procesar y mostrar los datos que posee Airbnb

en su sitio web usando las APIS públicas. Estos datos son actualizados regularmente y están disponibles de forma gratuita. La misión principal de Inside Airbnb es facilitarles a los residentes de una ciudad información que les permita evaluar el impacto de Airbnb en su vecindario y tomar decisiones informadas sobre las políticas y regulaciones que afectan el alojamiento.

Los datos de los alojamientos fueron obtenidos en febrero del 2023, este conjunto de datos contiene 6473 observaciones y 76 características. Cada observación es un alojamiento de Airbnb en la ciudad de Washington DC, el cual posee diferentes características en donde se encuentra información diversa acerca de cada alojamiento, tales como: la identificación del alojamiento, el tipo de alojamiento, el precio por noche, número de personas permitidas, los distintos servicios que brindan comodidad, el número de habitaciones, camas y baños, el número mínimo de noches para reservar, si está disponible el alojamiento, también se observa información de geo posición como la latitud y longitud, entre otros. Así mismo, se muestra información del host como: el id del host, si este es superhost, si el host tiene imagen de perfil, si está verificado, cuanto es el tiempo de respuesta, cuantos alojamientos posee el host, etc., incluso, en el dataset se observa información de las reseñas que ha tenido el alojamiento, de las opiniones y calificaciones de los usuarios en general, etc.

Adicionalmente, hay otro dataset que proviene de la misma página Inside Airbnb (*Inside Airbnb: Get the Data*, n.d.), el cual contiene los datos geográficos de los barrios de la ciudad de Washington DC, estos datos sirven para ver gráficamente en un mapa la distribución de los barrios y servirá para calcular el precio promedio, el puntaje promedio que han dado los usuarios a los alojamientos en cada barrio y otras características por barrio. En total se identificaron 39 barrios para la ciudad estudiada.

Otra fuente de datos que servirá para tener información de los lugares cercanos al alojamiento es Opendata DC (*Open Data DC*, n.d.), la cual es una página del gobierno de Washington DC creada especialmente para almacenar datos que tienen que ver con la ciudad, ahí se encuentra información sobre demografía, información geográfica sobre sitios turísticos, recreación y cultura, bancos, transporte, entre otros. Para este análisis se eligieron datos geográficos sobre bancos (*Bank Locations | Open Data DC*, n.d.), sitios para recreación identificados por el Departamento de Parques y Recreación de DC (*Recreation Facilities | Open Data DC*, n.d.), paradas de bus para la línea del metro (*Metro Bus Stops | Open Data DC*, n.d.), datos de algún incidente de crimen reportados por el Departamento de Policía Metropolitana de la ciudad (*Crime Incidents in 2020 | Open Data DC*, n.d.), puntos de interés en general en la ciudad como museos, monumentos, bibliotecas, edificios federales, entre otros (*Points of Interest | Open Data DC*, n.d.) y algunos sitios de interés puntuales como la casa blanca, el zoológico nacional, el monumento a Lincoln, el monumento a Washington, el museo nacional del Aire y el Espacio de Estados Unidos. Estos datos son útiles para calcular la distancia del alojamiento más cercano hacia alguno de estos puntos y con esto poder determinar si que el alojamiento esté cerca de alguno de estos puntos influye en el precio de Airbnb.

4.3. Limpieza y transformación de los datos

El proceso de limpieza y transformación de datos se refiere a que los datos necesitan ser identificados y analizados para eliminar inconsistencias, valores faltantes, cambiar formatos, crear nuevas variables, y corregir errores en estos para así garantizar la calidad y seguridad de que los algoritmos de Machine Learning tengan un buen rendimiento y arrojen datos más precisos.

Por lo tanto, se analizarán los datos para poder ver los tipos de datos y aplicar técnicas de limpieza de datos y valores faltantes con el objetivo de prepararlos para los modelos de regresión que se aplicarán más adelante, para ello, se cambió el formato de algunas características que están en formato texto o tienen caracteres especiales a número, ya que idealmente los modelos de regresión necesitan variables cuantitativas y las variables que sean categóricas se deben convertir a variables dummies las cuales tomarán valores numéricos dependiendo de cada categoría, generalmente estas variables dummies son dicotómicas, es decir, que toman valores de 0 o 1 para indicar “sí” o “no” (Montero Granados, 2016), adicionalmente se revisaron los duplicados y se encontró que no habían duplicados en el dataset de los alojamientos. También se convirtieron algunas variables categóricas booleanas a numéricas, es decir aquellas características binarias donde los valores sean “T” y “F”, estas se cambiaron a 1 y 0.

En cuanto a la revisión de nulos o valores faltantes, en el dataset de los alojamientos se encontraron múltiples características con valores vacíos, es importante no tener valores vacíos en los datos ya que esto puede afectar el modelo, por lo tanto, en algunas características se rellenaron con 0, el cual, para cada variable significará que no existe o no aplica, en otras se rellenaron los valores vacíos con el promedio.

La característica que contiene el precio del alojamiento contenía un signo de dólar y comas en sus valores, por lo tanto, se procedió a eliminar estos valores y convertir a un formato decimal, adicionalmente habían varios alojamientos donde el precio era 0 lo cual podría indicar un error en los datos por lo tanto, estos registros se eliminaron, así mismo, se revisaron los valores atípicos y se determinó que la distribución del precio está altamente sesgada, esto indica que hay un alto número de alojamientos económicos y pocos alojamientos costosos, así que se transformó el precio a logaritmo para hacer un suavizado de los datos y evitar la heterocedasticidad para ayudar a tener una mejor interpretación de los coeficientes. (Perez-Sanchez et al., 2018), esto también ayudará a observar mejor la comparación de esta variable con otras variables del dataset. Por lo tanto, se creará una nueva columna llamada “log_price” con el precio convertido a logaritmo natural. La **Ilustración 1** evidencia lo anteriormente explicado.

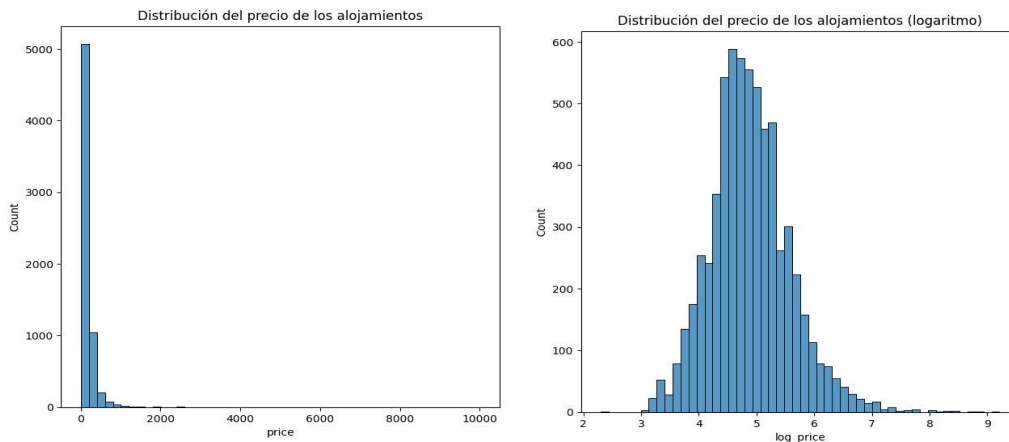


Ilustración 1. Distribución de precios de los alojamientos.

Fuente: Elaboración propia

La característica `bathrooms_text` contiene información del número de baños y el tipo de baño, por lo tanto, se creó una nueva columna llamada `"bathrooms_quantity"` la cual tendrá nada más número de baños que posee el alojamiento, esta será convertida a tipo numérico, esta variable contenía valores vacíos y 0, por lo tanto, estos valores se rellenaron con el promedio.

La característica `"room_type"`, la cual contiene el tipo de alojamiento se dividió en cuatro variables nuevas, una por cada tipo de alojamiento, quedando el dataset con cuatro variables nuevas: `"shared_room"` la cual quiere decir que la habitación es compartida con otros huéspedes, `"private_room"` es una habitación privada en la mayoría de los casos está ubicada donde el host vive, `"hotel_room"` indica si es una habitación de un hotel y `"entire_home_apartment"` que es una propiedad privada que es rentada en su totalidad, estas se rellenaron con valores de 1 (si es de este tipo) o 0 (no es de este tipo).

También existe una característica con el tipo de propiedad `"property_type"`, la cual contiene 51 posibles valores entre los cuales se encuentran que, si un alojamiento es una casa entera, un piso, una villa, una habitación en un hotel, una habitación privada, entre otras. A esta variable se le aplicó el modelo no supervisado llamado KModes, el cual es un algoritmo de Clustering usado para datos categóricos y se encarga de agrupar objetos de acuerdo con sus características categóricas en clústeres homogéneos, este usa modas como medida de distancia, es bastante eficiente cuando se desea agrupar datos categóricos (Jiang & Liu, 2020). A este se le aplicó el método del codo para obtener el número óptimo para k y se determinaron dos clústeres para esta variable que fueron dos.

Así mismo, de la variable `"amenities"` que contiene las diferentes comodidades o servicios adicionales que posee el alojamiento se extrajeron algunas de las comodidades más representativas que los huéspedes por lo general tienen en cuenta

como si el alojamiento tiene aire acondicionado, refrigerador, secador de cabello, piscina, si posee seguridad, entre otras (*Las Comodidades Que Quieren Los Huéspedes - Centro de Recursos de Airbnb - Airbnb*, n.d.), comodidades de si en el alojamiento hay internet, si tiene cocina, si tiene lavadora, secadora, si el desayuno viene incluido, si se admiten mascotas, parqueadero gratis en el alojamiento fueron tenidas en cuenta (Dogru & Pekin, 2017), también se consideró si el alojamiento posee detectores de humo (Jones et al., 2019). En total se identificaron 20 comodidades, las cuales fueron agregadas al conjunto de datos.

Los valores vacíos de las características que tienen que ver con calificaciones o reseñas de los huéspedes se rellenaron con el promedio.

Por otro lado, se convirtieron algunas variables que tienen que ver con el host de categóricas a numéricas, para ello se rellenaron con 1 o 0 aquellas características que contengan texto, esto es asumiendo que la información que contenga el texto no influye en nada que tenga que ver con el precio, pero si influye en que, si se proporciona esa información, el usuario que quiere alquilar el alojamiento tenga la confianza para hacerlo. Algunas características de la calificación del host estaban en formato texto, por lo tanto, se cambiaron a formato decimal.

- Se considerará que un host no tiene imagen de perfil si el campo `host_has_profile_pic` está vacío.
- Se considerará que un host no tiene la identidad verificada si el campo `host_identity_verified` está vacío.
- Se considerará que un host no es superhost si el campo `host_is_superhost` está vacío.
- Las variables `host_response_rate` y `host_acceptance_rate` serán rellenadas con 0 cuando haya valores vacíos. Ambas variables estaban en formato texto y contenía el símbolo de porcentaje (%), por lo tanto, este fue removido y la variable transformada a valor decimal.
- La variable “`host_verifications`”, la cual tiene una lista de valores en las que el host puede ser contactado por el usuario, fue convertida a número, indicando el número de verificaciones que tiene ese host en el alojamiento

Las características `accommodates`, la cual es el número máximo de personas permitidas en el alojamiento, `beds`, que es el número de camas que posee el alojamiento y `bedrooms` que contiene el número de habitaciones que posee el alojamiento, contenían valores vacíos y ceros, por lo tanto, se rellenaron con el promedio.

Además, se obtuvieron datos de ubicación de algunos puntos clave en la ciudad y se crearon nuevas características usando los dataset de lugares cercanos al alojamiento mencionados anteriormente, en estas variables se tiene la distancia mínima del alojamiento hasta ese punto, es decir, se calcularon diez nuevas variables las cuales contienen la distancia mínima en pies del alojamiento a las siguientes ubicaciones:

- Banco más cercano.
- Algún sitio para recreación más cercano.
- A la parada de bus más cercana.
- A algún punto de interés en general en la ciudad.
- A la Casa Blanca.
- Al monumento de Lincoln.
- Al monumento de Washington.
- Al museo de Aire y Espacio Nacional.
- Al Zoológico Nacional.
- A algún incidente de crimen que haya ocurrido cerca del alojamiento.

El dataset final de los alojamientos resultó con 6470 observaciones y 87 características, lo cual indicaría que estos son los alojamientos en Airbnb que se analizarán para la ciudad de Washington DC.

Al dataset de barrios, se le agregaron nuevas columnas que contienen el precio promedio de un alojamiento en ese barrio, el promedio de mínimo de noches por alojamiento en ese barrio, la disponibilidad promedio de un alojamiento en ese barrio, el número total de alojamiento del barrio y también la calificación promedio de los alojamientos en ese barrio.

4.4. Descripción de los datos

El dataset final de los alojamientos contiene más de 80 características y para que trabajar con el conjunto de datos sea menos engorroso, estas se dividieron en varias categorías que ayudarán a tener una mejor visión de los datos y hacer un mejor análisis:

- Variable dependiente: es el precio de los alojamientos de Airbnb.
- Características del alojamiento: las cuales van a incluir todo lo que tenga que ver con la propiedad como: el número de baños, número de camas, número de habitaciones, tipo de habitación, tipo de propiedad, etc.
- Características de servicios adicionales: si el alojamiento tiene amenidades adicionales como lavadora, cocina, wifi, piscina, etc.
- Características de calificaciones del huésped: son las calificaciones que el usuario le da al alojamiento en distintas categorías como limpieza, comunicación, ubicación, etc.
- Características del host: incluye datos de si el host tiene foto de perfil, si está verificado, entre otras.
- Características de ubicación: aquí se tendrán en cuenta características que tengan que ver con la ciudad, como la distancia mínima al lugar más cercano para recreación, bancos, algunos puntos de interés, entre otros.

Variable dependiente

La variable dependiente para este estudio es el precio de los alojamientos de Airbnb en la ciudad de Washington DC. En general, el precio promedio en toda la ciudad es de \$179 USD, los alojamientos tienen precios entre \$10 USD la noche el más barato y el precio más caro es de \$10.000 USD la noche. De hecho, la mayoría de los alojamientos tienen un precio inferior a \$500 por noche (6235 en total), aunque si hay algunos alojamientos que tienen precios muy elevados, pero son muy pocos.

Características del alojamiento

Esta categoría incluye las características físicas relacionadas con el alojamiento como tal, en las estadísticas de estos, se encontró que el promedio de personas que se pueden alojar es de cuatro, el mínimo es una persona y el máximo dieciséis personas, el promedio de camas por alojamientos es de dos, el mínimo es uno y el máximo veinte y más del 50% de los alojamientos poseen entre una y dos camas, el promedio de habitaciones por alojamientos es de 1.5, el mínimo es 1 y el máximo quince y que más del 70% de los alojamientos poseen entre una y dos habitaciones, el promedio noches mínimas por alojamientos es de veintiuno, el mínimo es 1 y el máximo 1125, el promedio de baños que poseen los alojamientos es de 1.3 y se observa que más del 70% de los alojamientos tienen entre uno y dos baños, en la **Tabla 1** se observan más detalladamente los datos estadísticos de las principales características.

Seguidamente, se analizaron más detalladamente estas variables y su influencia en el precio, algunas de ellas se describen a continuación:

Cada alojamiento está incluido dentro de una categoría que contiene el tipo de del alojamiento (room_type), esta característica se refiere principalmente a como fue arrendado el alojamiento y posee cuatro posibles valores: habitación privada, habitación compartida, habitación de hotel o casa/apartamento completo, La **Ilustración 2** muestra que el 76% de los alojamientos son casas o apartamentos, más de la mitad de los alojamientos son de este tipo, teniendo así un significativo número en los datos, el 22% son habitaciones privadas, esto puede indicar que el alojamiento es la casa donde el anfitrión vive, el 2% son habitaciones compartidas y la cantidad mínima la poseen las habitaciones de hoteles, lo que muestra que en Airbnb hay pocos hoteles ofreciendo servicios, ya que para los hoteles hay otras plataformas especializadas solo en este tipo de alojamientos. Haciendo una comparación de esta característica con el precio, se determinó que el precio es más alto en los alojamientos que son apartamentos o casas rentados en su totalidad (Entire home/apto) seguido de habitaciones privadas (private room), también se evidencian valores atípicos sobre todo en los alojamientos en estos dos tipos **Ilustración 3**.

También existe otra categoría que muestra el tipo de propiedad, en estos datos se identificaron 51 posibles valores, la **Tabla 2** muestra que el tipo de propiedad que

posee más alojamientos es “Entire rental unit” con 2301 propiedades, le sigue “Entire home”.

Hay una variable que indica si el huésped puede hacer la reserva sin necesidad de aprobación del host, solamente el 41% de los alojamientos aceptan ser rentados inmediatamente, esto implica mayor rapidez al momento de reservar el alojamiento, también se observa que los precios de alojamientos donde si se requiere autorización del host son más altos en comparación con los que no requieren autorización **Tabla 3**.

Estudios indican que el número de personas permitidas en el alojamiento influye en el precio en el que el host le coloca a la renta, (Tsai et al., 2023). En la **Ilustración 4** se observa que, en general, el precio incrementa entre más sea el máximo número de personas que permite el alojamiento y hay outliers o valores atípicos en algunos de los valores, sobre todo cuando el número máximo de personas es 4 y 6, y cuando es 11 el precio por alguna razón no aumenta, esto podría significar que no hay muchos alojamientos que acepten 11 personas y no haya suficientes datos para demostrar un aumento en el precio. Se observa que hay una relación positiva entre "accommodates" y el precio, lo que indica que, en promedio, los alojamientos más grandes tienen precios más altos.

la variable que indica el número mínimo de noches que un huésped puede reservar el alojamiento, muestra que la mayoría de estos tienen un número mínimo de noches de 1, pero hay algunos alojamientos con números mínimos de noches muy altos, hasta de 1125, esto podría significar que el alojamiento puede ser arrendado para personas que vayan a tener una estadía alta en la ciudad o a vivir en ella. El percentil 50 es de 14 noches, lo que indica que la mayoría de los alojamientos tienen un requisito mínimo de catorce noches, mientras que el percentil 25 es de 2 noches, lo que sugiere que alrededor del 25% de los alojamientos tienen un requisito mínimo de dos noches.

Otra variable analizada es "availability_365", que indica la disponibilidad del alojamiento a lo largo del año, en ella se observa que el promedio de disponibilidad de todos los alojamientos es de 175 días.

Se realizó un mapa de calor para ver correlación de las variables (**Ilustración 5**) y analizar el comportamiento con respecto al precio y se identificó que el número máximo de personas permitidas en el alojamiento, la cantidad de baños, el número de camas están correlacionada de alguna manera con el precio, adicionalmente los tipos de alojamiento que son apartamentos o casas completas tienen relevancia con respecto al precio. Se observa que hay una correlación positiva fuerte entre el precio y las variables que tienen que ver con el tipo de habitación, también entre el número de personas permitidas en el alojamiento, número de camas y número habitaciones, esto indica que a medida que entre más aumenten estas variables, el precio tenderá a ser más alto.

Además, se examinaron de cerca los valores atípicos de diferentes variables para comprender las diferentes formas en que los propietarios publicaban sus alojamientos. Se observa que algunas variables, como el número de baños y la cantidad de reseñas,

tienen valores extremos que podrían ser valores atípicos, sin embargo, dado que estos valores son relativamente pocos en comparación con el tamaño total de la muestra, es posible que no afecten significativamente en los resultados.

Características de servicios adicionales

Las amenidades se refieren a las comodidades extras que se ofrecen en cada uno de los alojamientos. Diversos estudios indican que las amenidades o servicios adicionales son tenidas en cuenta por los usuarios al momento de reservar un alojamiento en Airbnb, por lo tanto, en este estudio se tuvieron en cuenta algunas de las más importantes (*Las Comodidades Que Quieren Los Huéspedes - Centro de Recursos de Airbnb - Airbnb*, n.d.). Hay comodidades lujosas como cocina, detector de alarma, piscina y gimnasio o servicios como desayuno, están disponibles en pocos alojamientos, lo que podría indicar que estos servicios exclusivos influyan en el precio, es decir, que los alojamientos que incluyan alguno de estos servicios tienden a ser más costosos que los que no tienen.

En la **Ilustración 6** se observa que algunas comodidades están relacionadas entre sí, como lavadora y secadora, gimnasio y piscina, nevera y productos para limpieza, aire acondicionado y calefacción, entre otras. Esto indica que los alojamientos que ofrecen una comodidad específica tienden a ofrecer otras comodidades relacionadas.

Características de calificaciones del huésped

Según varios estudios, las calificaciones juegan un papel importante en los precios de los alojamientos, es decir, la reputación y las calificaciones son tenidas en cuenta por los usuarios al momento de elegir un alojamiento (Teubner et al., 2017). En los datos obtenidos de los alojamientos están presente siete de estas calificaciones, en ellas están incluidas calificaciones de ubicación, check in, comunicación, limpieza y precisión. Según (Zhang et al., 2017) la experiencia y calificación de los usuarios influye en la reputación de los alojamientos y a su vez esto puede influir en los precios, ya que a medida que haya mejores calificaciones, los usuarios se sentirán más seguros y confiados en que tendrán una buena experiencia con el alojamiento.

Para la variable que indica la calificación promedio de los huéspedes para cada alojamiento se observa que en promedio los alojamientos tienen una calificación de 4.7 y los valores de las calificaciones están entre 0 y 5, además, hay algunos alojamientos con calificaciones por debajo de 4 (162 alojamientos) pero la gran mayoría de los alojamientos tienen calificaciones mayores a 4 (6308). Analizando esta misma característica en función del tipo de habitación se observa que los alojamientos de tipo casa o apartamentos que son rentados en su totalidad tienen un promedio de calificación de 4.7 (**Ilustración 7**).

Todas las variables de calificaciones están altamente correlacionadas entre sí, con coeficientes de correlación que van desde 0.5 hasta 0.8 (**Ilustración 8**), esto podría sugerir que las opiniones y calificaciones de los huéspedes son importantes para la reputación del alojamiento de Airbnb en Washington DC y que los alojamientos con mejores evaluaciones pueden atraer más reservas.

Características del host

Las principales características del host juegan un papel importante en el precio del alojamiento, entre estas características se encuentran: si el host está verificado, ya que las personas tienden a sentirse más seguros que con un host que no está verificado y la influencia de si el usuario es superhost implica mayor calidad y confianza en que los usuarios hagan reserva, así lo indica (Teubner et al., 2017). Otras variables como las imágenes de perfil de los hosts y que la confianza de los usuarios puede verse afectada solamente por la variable superhost (Zhang et al., 2017).

En este análisis se tuvo en cuenta si un host es superhost que quiere decir que aquellos con esta característica son usuarios reconocidos que brindan las mejores experiencias y servicios en Airbnb, es una variable binaria, 1 indica si, 0 indica que no es superhost, en ella se encontró que solamente el 30% son superhost **Ilustración 9**. Lo mismo aplica para la característica que indica si el host está verificado o no, en ella se identificó que el 82% de los hosts están verificados. También se tuvo en cuenta si el host tenía imagen de perfil o no, esta variable también es binaria en donde 1 indica que, si tiene imagen de perfil y 0 indica que no, en ella se encontró que el 97% de los hosts si tienen imagen de perfil **Ilustración 10**.

Se puede observar que la variable `host_response_rate` tiene una mediana de 100%, lo que indica que la mayoría de los hosts responden a las consultas de los huéspedes. La variable `host_listings_count` tiene una mediana de 2 alojamientos por host, lo que indica que la mayoría de estos tienen solo un par de propiedades en Airbnb. Aunque, se evidencia un rango amplio en estas dos variables.

La mayoría de los hosts tienen un alto nivel de respuesta y aceptación de reservas y como una proporción significativa de anfitriones tienen el estado de superhost podría significar una alta calidad en la atención a los huéspedes y más confiabilidad para los huéspedes porque les suministra información sobre la identidad de este y los ayuda a identificar hosts confiables y que ya tengan experiencia con alojamientos de Airbnb.

En cuanto a la correlación, la cual se puede apreciar en la **Ilustración 16**, se observa que las variables de respuesta y aceptación del host están correlacionadas positivamente y también las variables que indican si el host tiene imagen en miniatura, si el host tiene un enlace a su foto y si el host tiene imagen de perfil están altamente correlacionadas, en este caso se podría dejar solo una de ellas.

Características de ubicación

Adicionalmente a todas las variables mencionadas anteriormente, la ubicación del alojamiento y de algunos sitios en la ciudad causan un impacto importante en el precio de estos (Teubner et al., 2017), por lo tanto, se analizaron algunos aspectos con respecto a la ubicación del alojamiento. En total se identificaron 39 barrios en la ciudad de Washington DC.

Se observa que la mayoría de los alojamientos están cercanos al centro de la ciudad, cerca de monumentos y atracciones turísticas como la Casa Blanca, el Monumento a Washington o el Capitolio **Ilustración 12**. La cantidad de alojamientos varía bastante dependiendo del barrio, Capitol Hill, Lincoln Park es el que más alojamientos tiene con un total de 561 junto a Union Station, Stanton Park, Kingman Park que contiene 550 alojamientos, estos dos barrios son los más cercanos al centro y a lugares como el Capitolio. También está el otro extremo, barrios como Woodland/Fort Stanton, Garfield Heights, Knox Hill y Eastland Gardens, Kenilworth contienen solo 12 alojamientos cada uno, estos dos barrios se encuentran en la parte más alejada en el este de la ciudad.

En cuanto a los barrios más caros se encontró que Southwest Employment Area, Southwest/Waterfront, Fort McNair, Buzzard Point ubicado muy cerca al Capitolio y el centro de la ciudad es el más costoso, con un promedio de \$294 USD por noche, y el más barato es Congress Heights, Bellevue, Washington Highlands, con un valor promedio de \$100 USD por noche, este barrio está ubicado un poco más en el extremo sureste de la ciudad. En la **Ilustración 111** se observa en general, que los alojamientos ubicados en barrios más populares y cercanos al centro de la ciudad tienen un valor más elevado en el precio por noche que los ubicados en barrios más alejados.

En la **Tabla 4** se observan algunos valores por barrio, hay barrios donde el promedio de noches mínimas de reserva de un alojamiento es de más de 30 días, pero en la gran mayoría de barrios se pueden reservar en menos de 20 días. En cuanto a la calificación promedio, los barrios tienen por lo menos una calificación mayor a 4.4, el barrio mejor calificado es Hawthorne, Barnaby Woods, Chevy Chase, con una calificación de 4.8 este barrio tiene un promedio de precios de 226 USD la noche, es uno de los más costosos a pesar de que no queda cerca al noroeste de la ciudad y posee solamente 68 alojamientos. El barrio peor calificado es North Cleveland Park, Forest Hills, Van Ness con una calificación de 4.4, curiosamente este barrio se encuentra justo al lado del barrio mejor calificado y su precio promedio por noche es de \$111 USD, no es de los más caros, pero tampoco es de los más baratos y posee 32 alojamientos.

Por otra parte, la correlación entre la latitud, longitud y el precio es negativa, lo que sugiere que los alojamientos ubicados en áreas con valores de latitud y longitud más altos tienen precios más altos, esto quiere decir que los alojamientos ubicados cerca al centro de la ciudad son más caros. De hecho, la distancia mínima entre el alojamiento y lugares como bancos, puntos de interés, puntos de recreación están correlacionados

entre sí y se observa también una correlación negativa con el precio, lo cual indicaría que a medida que disminuye la cercanía del alojamiento hacia el lugar el precio es menor **Ilustración 17**. A pesar de que se encontraron variables correlacionadas as variables correlacionadas son importantes y relevantes para el modelo, puede ser justificable mantenerlas en el modelo.

En resumen, la ubicación es un ingrediente importante que los huéspedes tienen en cuenta al reservar un alojamiento en Washington, DC, ya que los precios pueden variar considerablemente dependiendo del barrio en el que se encuentre el alojamiento, su proximidad a ciertos puntos de la ciudad e incluso al centro de la ciudad.

4.5. Selección de características

Después de analizar los datos e identificar todas las características de los alojamientos, se elegirán aquellas que se consideren relevantes para el modelo y que no estén altamente correlacionadas para evitar mucha multicolinealidad. A continuación, se muestra un resumen de todas estas variables elegidas que fueron ya transformadas y explicadas con detalle anteriormente.

Algunas características que tienen que ver con servicios adicionales fueron eliminadas ya que se determinó que la presencia de una de ellas incluía a la otra, como es el caso de `has_Dryer` con `has_Washer`, los datos indicaban que en la gran mayoría de alojamientos cuando había lavadora ya venía incluida la secadora, el mismo caso se determinó con las variables `has_Heating` y `has_Airconditioning`. Igualmente, con algunas características del alojamiento que tenían una alta correlación, la variable `private_room` tenían una correlación muy alta positiva cuando estaba en presencia de la variable `entire_home_apt`, por esta razón, la última fue elegida para el modelo, también la variable `property_type_cluster` estaba correlacionada negativamente con `entire_home_apt`, de modo que, se eligió esta última. En las características del host se observó que las variables `host_response_rate` y `host_acceptance_rate` estaban altamente correlacionadas y que al parecer brindaban casi la misma información, por lo tanto, se eligió la última mencionada. Desafortunadamente la característica del tipo de propiedad calculada con el algoritmo KModes presentaba una correlación demasiado alta con otras variables como `entire_home_apt` y podía afectar mucho el modelo, esto podría indicar que, en este caso, esta variable está incluida dentro de los alojamientos de tipo casa o apartamento y por lo tanto `property_type_cluster` fue descartada del modelo.

Las variables definidas como tipo dummy indican que la característica posee valores de 0, que significa no y 1 significa sí.

Variable dependiente:

- `log_price`: Indica el precio del alojamiento en dólares (USD), como se ha explicado anteriormente, esta variable fue transformada usando el logaritmo para suavizar los datos. (Tipo: Número).

Características del alojamiento:

- `shared_room`: Indica si un alojamiento es de tipo habitación compartida (Tipo: Dummy).
- `hotel_room`: Indica si un alojamiento es una habitación de hotel (Tipo: Dummy).
- `entire_home_apartment`: Indica si un alojamiento es una casa o un apartamento (Tipo: Dummy).
- `accommodates`: Es el número máximo de huéspedes por noche que permite el alojamiento. (Tipo: Número).
- `bedrooms`: Número de habitaciones que posee el alojamiento. (Tipo: Número).
- `beds`: Número de camas que tiene el alojamiento. (Tipo: Número).
- `bathrooms_quantity`: Cantidad de baños que posee el alojamiento. (Tipo: Número).
- `minimum_nights`: Número mínimo de noches que puede ser reservado el alojamiento. (Tipo: Número).
- `instant_bookable`: Indica si el alojamiento puede ser reservado inmediatamente o si requiere alguna aprobación o revisión previa del host. (Tipo: Dummy).
- `availability_365`: Número de días que el alojamiento está disponible en el año. (Tipo: Número).

Características de servicios adicionales:

- `has_breakfast`: El alojamiento ofrece desayuno incluido. (Tipo: Dummy).
- `has_washer`: El alojamiento tiene lavadora. (Tipo: Dummy).
- `has_dishwasher`: Hay lavavajillas (Tipo: Dummy).
- `has_tv`: Tiene televisor. (Tipo: Dummy).
- `has_freezer`: Hay refrigerador en el alojamiento. (Tipo: Dummy).
- `has_airconditioning`: Hay aire acondicionado. (Tipo: Dummy).
- `has_hotwater`: Hay agua caliente. (Tipo: Dummy).
- `has_cleaning_products`: Indica si se ofrecen productos para limpieza. (Tipo: Dummy).
- `has_wifi`: Hay internet en el alojamiento. (Tipo: Dummy).
- `has_hairdryer`: Tiene secador de Cabello. (Tipo: Dummy).
- `has_kitchen`: Tiene cocina. (Tipo: Dummy).
- `has_dedicated_workspace`: Tiene algún espacio dedicado para trabajar. (Tipo: Dummy).
- `has_pets_allowed`: Indica si el alojamiento acepta mascotas. (Tipo: Dummy).

- has_Pool: Tiene piscina el alojamiento. (Tipo: Dummy).
- has_Freestreetparking: Indica si se puede parquear gratis en la calle. (Tipo: Dummy).
- has_Freeparkingonpremises: Indica si en el área hay algún tipo de parqueadero gratis. (Tipo: Dummy).
- has_Smokealarm: Tiene alarma de humo. (Tipo: Dummy).
- has_Securitycamerasonproperty: Tiene cámaras de seguridad en el alojamiento. (Tipo: Dummy).
- has_Gym: Tiene gimnasio. (Tipo: Dummy).

Características de calificaciones del huésped:

- number_of_reviews: Número de reseñas que tiene el alojamiento. (Tipo: Número).
- review_scores_rating: Es el puntaje general del alojamiento, los valores van de 1 que es la calificación más baja a 5 que indica la más alta. (Tipo: Número decimal).
- review_scores_accuracy: Los huéspedes califican cuando se sienten bien con el alojamiento y que eso que pidieron al llegar concuerda con lo ofrecido. (Tipo: Número decimal).
- review_scores_cleanliness: El alojamiento cumple con un alto nivel de limpieza. (Tipo: Número decimal).
- review_scores_checkin: El proceso de registro es fácil y el huésped está contento cuando llega al alojamiento. (Tipo: Número decimal).
- review_scores_communication: Los huéspedes califican cuando la comunicación es buena. (Tipo: Número decimal).
- review_scores_location: Esta calificación es sobre la ubicación del alojamiento. (Tipo: Número decimal).
- review_scores_value: Calificación sobre la calidad del alojamiento. (Tipo: Número decimal).

Características del host:

- host_response_time_hour: Número de horas que tarda el host en responder alguna inquietud por parte de un huésped. (Tipo: Número).
- host_about: Indica si el host posee información extra sobre él o ella que ayude al huésped a conocer más sobre el host. (Tipo: Dummy).
- host_is_superhost: Indica si el host es un superhost. Esta característica es dada solamente a los hosts con mejores calificaciones, por encima de 4.8, índice de respuesta de más del 90% y la tasa de cancelación de menos del 1% en 1 año. (Tipo: Dummy).
- host_verifications: Número de formas de verificación que posee el host. (Tipo: Número).
- host_has_profile_pic: Indica si el host tiene imagen de perfil. (Tipo: Dummy).

- `host_identity_verified`: La identidad del host está verificado por Airbnb. (Tipo: Dummy).
- `host_acceptance_rate`: Aceptación del host. (Tipo: Número decimal).
- `calculated_host_listings_count`: Cantidad de alojamientos que posee el host para rentar en Airbnb. (Tipo: Número).

Características de ubicación:

- `min_dist_to_bank_ft`: Distancia mínima en pies del alojamiento al banco más cercano de la ciudad. (Tipo: Número decimal).
- `min_dist_to_recreation_ft`: Distancia mínima en pies del alojamiento al punto para recreación más cercano de la ciudad. (Tipo: Número decimal).
- `min_dist_to_metro_bus_stop_ft`: Distancia mínima en pies del alojamiento a la parada de bus del metro de la ciudad. (Tipo: Número decimal).
- `min_dist_to_points_interest_ft`: Distancia mínima en pies del alojamiento a algún punto de interés de la ciudad. (Tipo: Número decimal).
- `min_dist_to_white_house_ft`: Distancia mínima en pies del alojamiento a la Casa Blanca. (Tipo: Número decimal).
- `min_dist_to_lincoln_memorial_ft`: Distancia mínima en pies del alojamiento al monumento de Lincoln. (Tipo: Número decimal).
- `min_dist_to_national_space_museum_ft`: Distancia mínima en pies del alojamiento al Museo Nacional del Espacio. (Tipo: Número decimal).
- `min_dist_to_washington_memorial_ft`: Distancia mínima en pies del alojamiento al monumento de Washington. (Tipo: Número decimal).
- `min_dist_to_national_zoo_ft`: Distancia mínima en pies del alojamiento al Zoológico Nacional. (Tipo: Número decimal).
- `min_dist_to_crime_incident_ft`: Distancia mínima en pies del alojamiento a algún incidente de crimen que haya ocurrido en la ciudad. (Tipo: Número decimal).

4.6. Modelos de regresión

Para poder pasar los datos a los modelos de regresión primero se normalizaron los datos de las variables independientes usando la librería Scikit-Learn, la función `MinMaxScaler` es usada para estandarizar los datos y que las diferentes medidas y la diferencia de rangos de valores entre las variables estén en una escala semejante entre sí, esto ayudará a que el modelo arroje valores más precisos. Se usó en Python la librería `statsmodels.formula.api` para hacer este cálculo y se le enviaron las variables que se encuentran en la sección anterior, en ella se obtuvieron los valores del intercepto, los p-valores y los coeficientes. Estos modelos fueron elegidos para poder determinar si hay una relación entre el precio de los alojamientos de Airbnb y las distintas características anteriormente descritas.

Regresión por mínimos cuadrados (OLS)

La regresión por mínimos cuadrados estima la relación lineal entre una variable dependiente y una o varias independientes, su principal objetivo es encontrar la recta que mejor se ajuste a los datos. La regresión OLS funciona mediante el cálculo de la suma de los cuadrados de las diferencias entre los valores observados de la variable dependiente y los valores predichos por la línea de regresión. Luego, se ajustan los coeficientes de la línea de regresión para minimizar esta suma de cuadrados. Los coeficientes de la línea de regresión representan la pendiente y la intersección de la línea.

Regresión cuantílica (QR)

La regresión cuantílica es un modelo de regresión que estima diferentes cuantiles de una variable dependiente en función de una o varias independientes y a diferencia de la regresión OLS, la regresión cuantílica se enfoca en encontrar los valores que separan a la variable dependiente en diferentes percentiles. Es útil cuando la variable dependiente tiene una distribución irregular donde la media y la mediana no describen bien los datos, es decir, cuando los datos no cumplen las condiciones que necesita la regresión lineal. La regresión cuantílica se basa en la minimización de una función de pérdida que tiene en cuenta los errores en diferentes cuantiles de la variable dependiente, para ello, se utilizan los percentiles como puntos de corte en la función de pérdida y se busca encontrar los coeficientes de regresión que minimicen la suma ponderada de los errores absolutos entre los valores observados y los valores predichos para cada cuantil. Se calcularon varios modelos usando los cuantiles 0.10, 0.25, 0.50, 0.75 y 0.90.

4.7. Resultados

Los datos obtenidos por los modelos de regresión OLS (Mínimos cuadrados ordinarios) y QR (Regresión cuantílica) se encuentran en la **Tabla 5**, en ella están los coeficientes, la significancia estadística, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, en donde p es el p-valor obtenido y entre paréntesis están los valores para los errores estándares, y en la **Tabla 7** se encuentran los p-valores. Para el modelo OLS y los cuantiles 0.10, 0.25, 0.50, 0.75 y 0.90 se encontró que hay variables que aportan significativamente a los precios de los alojamientos de Airbnb y hay unas que aportan incremento al precio y hay otras que resultan en una reducción de estos.

Como el precio se transformó a logaritmo, la interpretación de los resultados del modelo puede ser un poco diferente al de una variable sin transformar. Cuando la variable está sin transformar, el coeficiente obtenido representa el aumento en la

variable dependiente cuando la independiente aumenta una unidad y en la variable transformada a logaritmo natural representa el cambio pero de modo porcentual ya que al hacer esta transformación los valores muy grandes son más fáciles de comprender. La fórmula para poder calcular ese porcentaje es la siguiente: $(\exp(\text{coeficiente}) - 1) * 100$, la función "exp" es la función exponencial la cual transforma el logaritmo natural de un número en su valor original. Por lo tanto, un aumento del 1% en la variable independiente se relaciona con un aumento del porcentaje que arroje la fórmula, en la variable dependiente (Ortiz Pinilla & Gil, 2014). A los resultados obtenidos se le aplicó esta fórmula para poder hacer mejor las comparaciones, esto se puede ver en la **Tabla 6**.

Los resultados serán analizados por las categorías definidas anteriormente.

Características del alojamiento

Para esta categoría, en el modelo OLS solamente la variable `instant_bookable` no es estadísticamente significativa, el resto sí influye en el precio del alojamiento. Como se esperaba, las características más importantes que hacen que el precio aumente son el número de habitaciones (`bedrooms`) que aporta en un 47% en el aumento del precio, el número de baños (`bathroom_quantity`) incrementa el precio en un 21% y el número de personas permitidas (`accommodates`) aumenta en un 17% el precio del alojamiento. Se identificó que los alojamientos que son casas o apartamentos hacen que el precio se incremente en un 4%, en comparación con las habitaciones compartidas y las habitaciones de hoteles que muestran valores negativos, es decir hacen que el precio disminuya en más de un 7% cada una. Inesperadamente el número de camas influye negativamente en el precio, quiere decir que los hosts consideran que esta variable no hace que aumenten el precio.

En cuanto a la regresión cuantílica se observa que en la gran mayoría de cuantiles las variables son estadísticamente significativas, excepto el número de camas en los cuantiles 0.5 en adelante, el número mínimo de noches en el cuantil 0.9, `instant_bookable` en los cuantiles 0.75 y 0.9 y en las habitaciones de hotel en el cuantil 0.9, siguiendo esta idea, se puede decir que en los precios más altos estas variables no son relevantes para determinar el precio de un alojamiento. Se podría decir que las variables que más influyen en incremento del precio son el número de baños, el cual se observa que en el cuantil 0.9 tiene un incremento de hasta un 38% en comparación con el resto de cuantiles, de hecho, el número de habitaciones también influye en el incremento del precio, en los precios más bajos se ve un incremento de hasta el 61% y en los alojamientos que tienen precios más altos aumenta en un 46%, igualmente con el número de personas permitidas en el alojamiento, en ella se observa un valor muy parecido en los primeros tres cuantiles y se ve que en los precios más altos (cuantiles 0.75 y 0.9) el número de personas hace que el precio aumente en por lo menos un 18%. La disponibilidad es estadísticamente significativa en los dos modelos, se observa un valor casi constante en todos los cuantiles.

Al igual que OLS, la variable que indica el número de camas influye negativamente en el precio, lo que concluye que no es relevante para determinar el precio de un alojamiento en ninguno de los dos modelos. Con el número mínimo de noches para poder reservar se observa que tiene una influencia negativa en el precio, esto quiere decir que los hosts están teniendo en cuenta otros factores para poder determinar el precio por noche. Con respecto al tipo de alojamiento, las habitaciones compartidas y las habitaciones de hotel en todos los cuantiles se observan valores negativos, lo cual confirma lo anteriormente estudiado, los precios para estas propiedades son menores a los que puede contener una casa o apartamento entero los cuales significan que son los que tienen los precios más altos, significando que los cuantiles que están por debajo de la mediana son los que más tienen aumento de precio de más de 5% en los alojamientos que son casas o apartamentos.

Características de servicios adicionales

En el modelo OLS, se encontraron múltiples variables que no son estadísticamente significativas, entre ellas están, si un alojamiento tiene o no tiene TV, Wifi, refrigerador, aire acondicionado, productos para limpieza, si permite mascotas y si tiene cámara de seguridad.

De las variables más significativas en los dos modelos se encuentra que si un alojamiento posee lavavajillas en la regresión OLS aumenta el precio hasta un 1.1% y en la regresión cuantílica en los cuantiles más bajos de la mediana los precios aumentan a más de 1.3%, siendo el cuantil 0.1 el que posee el porcentaje más alto de todos los cuantiles con un 1.5% de aumento en el precio, lo que quiere decir que entre más barato es un alojamiento, tener una cocina aumentará más su precio y en aquellos alojamientos más caros quizá ya venga incluido y por eso no significa un aumento en el precio.

Con los alojamientos que poseen cocina, se observan valores negativos tanto en OLS como en QR, y en el cuantil 0.1 no es estadísticamente significativa esta variable. Este comportamiento lo que podría estar queriendo decir es que la mayoría de los alojamientos ya poseen cocina y esto no repercute o tiene poca importancia en el precio para los hosts.

En cuanto a los alojamientos que poseen agua caliente se encontró que es estadísticamente significativa en todos los cuantiles y la regresión OLS pero se observan valores negativos, lo cual significa que los precios disminuyen entre 0.5% y 2.3%. Por otro lado, los alojamientos que poseen secador de cabello se observa un valor positivo en los precios en la regresión OLS el precio incrementa un 1.2% y en la QR en todos los cuantiles se observa un aumento de por lo menos un 1%, excepto en el cuantil 0.9 en donde esta variable no es estadísticamente significativa.

El Gimnasio es la variable la importante que afecta el precio de un alojamiento en OLS presenta un aumento del 3.6% y en QR en el cuantil 0.1 es donde más se presenta un

aumento del precio (3.7%), en el cuantil 0.25 el aumento es del 3.3%, en el cuantil 0.5 el aumento es del 3.2%, en el 0.75 el aumento es del 3.4% y en el 0.9 el aumento es de 3.1%, es decir, los hosts consideran que el precio de los alojamientos que posean esta característica debería ser mayor a los que no tienen gimnasio.

En cuanto a variables de si un alojamiento tiene la lavadora se observa una significancia en los cuantiles 0.75 y 0.9 y se presentan valores negativos, lo que quiere decir que probablemente los alojamientos ya tengan este servicio incluido y no representa para los hosts un aumento en el precio. Por otro lado, cuando se posee cámara de seguridad se observa significancia nada más en los extremos de los cuantiles (0.1 y 0.9), en el cuantil 0.1 se observa un valor positivo, lo que indica que el precio aumenta un 0.5% cuando el alojamiento tiene cámaras de seguridad, por otro lado, en el cuantil 0.9 se observa un valor negativo, lo que quiere decir que el precio disminuye en un 0.8%.

Los alojamientos con espacios dedicados para trabajo se presentan valores negativos en todos los modelos, pero solamente se presenta significancia en OLS y los cuantiles 0.25 y 0.9. Con el parqueadero ocurre que aquellos parqueaderos que están en la calle en los alojamientos más caros representan una disminución del precio, pero en los alojamientos más baratos representa un aumento del 0.4% y en cuanto a los parqueaderos gratuitos en el área presenta significación en los valores más altos de los alojamientos e incluso representa más de un 0.5% de aumento en el precio.

En alojamientos que tienen alarma de humo se observa una significancia en los cuantiles por encima de la mediana con valores negativos, significando una reducción en el precio hasta de un 6% en los precios más altos y los que ofrecen desayuno también poseen significancia en los dos últimos cuantiles y el precio aumenta entre 1.6% y 2.9%.

Características de calificaciones del huésped

En definitiva, la característica sobre calificaciones que más influyen en los precios son la de ubicación, limpieza y el puntaje general del alojamiento, donde en esta última nada más el cuantil 0.1 no es estadísticamente significativo, el resto si lo son. Sobre la ubicación se observan aumentos hasta del 13% en los precios (cuantil 0.75). En la limpieza se ven aumentos desde el 7.7% hasta el 9.4% en el precio, siendo la regresión OLS la que mayor porcentaje de aumento representa. En el puntaje general el aumento del precio más bajo se encuentra en el cuantil 0.25 (4.4%) y el más alto en el cuantil 0.75 (11%). El checkin y la calidad poseen valores negativos y son estadísticamente significativos en los cuantiles por encima de la mediana. Las calificaciones de comunicación y precisión no son estadísticamente significativas para ninguno de los modelos.

Características del host

En OLS la característica que indica si es superhost no es estadísticamente significativa lo que indica que en esta distribución de los datos esta variable no influye en el precio, el resto de las variables si lo son y varias de ellas tiene valores positivos y negativos. La tasa de aceptación del host, el número de verificaciones y si el host tiene imagen de perfil aportan un incremento del precio desde el 0.6 hasta más del 3%, siendo la última mencionada la que más aporta. Por otro lado, las que contienen valores negativos son aquellas que tienen que ver con la identidad verificada, el total de alojamientos del host, información extra del host y tiempo de respuesta implican una reducción en los precios que van desde 0.4% hasta 2.8%, siendo la cantidad de host la que más reducción aporta al precio y la información extra del host la que menos reducción del precio aporta.

Con respecto a los cuantiles hay diferencias en varios cuantiles con respecto a OLS. La cantidad de alojamientos que tiene el host es significativa en todos los cuantiles excepto en el 0.25, solamente aporta un incremento en el precio de 1.9% en los alojamientos que tienen precios más bajos (Q0.1), en los cuantiles que van del 0.5 en adelante puede haber una reducción en los precios desde 1.1% hasta 4.2% en los alojamientos con precios más altos.

La tasa de aceptación del host en el cuantil 0.75 no es estadísticamente significativa y en el 0.9 aporta reducción al precio de hasta 2%, en los cuantiles de 0.5 hacia abajo aporta un incremento entre 0.7% y 1%.

La identidad verificada del host es significativamente estadística en el cuantil 0.75 y aporta una reducción del 0.5% en el precio, en el resto no se presenta significancia.

En resumen, la variable más importante para determinar un aumento en el precio en esta sección es "host_has_profile_pic".

Características de ubicación

Alojamientos cercanos a algún incidente de crimen muestran una disminución en los precios en los modelos, OLS presenta una reducción de más de 16% y en los alojamientos más baratos la reducción es menor (6.5%) pero en el 0.25 aumenta al doble la reducción del precio (12%), en los alojamientos más caros se presenta hasta una reducción del 10%, en general, se deduce que los incidentes de crimen hacen que los precios vayan disminuyendo en todas las zonas de la ciudad.

La distancia hacia el Zoológico Nacional es significativa en todos los modelos con valores positivos, siendo que en los alojamientos más caros (Q0.9) el precio aumenta hasta un 15% y en el resto entre 8.5% y 9.9%. En OLS se observa que los precios aumentan también, en por lo menos 10.6%. Así mismo con los puntos de interés, estos influyen positivamente en el precio siendo en la regresión OLS donde más se

presenta este aumento (19%), en el cuantil más bajo se presenta un 8% de incremento y en el 0.25 se duplica hasta en un 16% y ya en los otros cuantiles más altos el incremento en el precio se da entre 10% y 13%.

La distancia al banco también presenta valores significativos, pero son negativos, lo que quiere decir que los alojamientos que están cercanos a un banco por lo general son más baratos que otros. En los alojamientos más baratos se presenta una reducción de 4.2% y en los precios más altos se presenta una reducción más grande (7%).

Por otro lado, la distancia a la Casa Blanca presenta valores negativos y no presenta significancia en los cuantiles 0.5 y 0.75. En OLS se presenta una reducción de hasta 21% en el precio y en los alojamientos con los precios más altos se presenta una disminución del precio de hasta 26%.

Los puntos de recreación en OLS es estadísticamente significativo y se presenta un crecimiento en el precio de 2.3%, en los cuantiles es un poco diferente, solamente los cuantiles 0.5 y 0.75 presentan significancia y los precios aumentan en 2% y 3% respectivamente.

Alojamientos cercanos a paradas de bus solamente presentan significancia en los cuantiles más altos presentando un incremento de más de 4% en el precio. Algo parecido pasa con alojamientos cercanos al Museo Nacional del Espacio, solamente se presentan significancias en los cuantiles 0.25 y 0.5 pero el aumento del precio es bastante significativo, de 23% y 30% respectivamente.

Cercanía a lugares como el Monumento de Washington y Monumento de Lincoln no muestran significancia.

5. Conclusión y trabajos futuros

Se ha conseguido analizar los alojamientos turísticos de Airbnb en la ciudad de Washington DC. Además, se ha podido tener una más clara idea de que características son determinantes en el aumento de los precios, ya que a los hosts les interesa siempre poder cobrar lo justo y sacar ganancias.

En cuanto a las características del alojamiento, se determinó que los tipos que sean casas o apartamentos influyen bastante en el precio que por lo general son los más altos, el número de personas permitidas, las habitaciones, la cantidad de baños, la disponibilidad también influye significativamente en el aumento del precio. Amenidades como si el alojamiento posee gimnasio, secador de cabello y lavavajillas también afectan el precio positivamente. En cuanto a la ubicación, la cercanía a los principales lugares de interés turístico y al Zoológico Nacional contribuye al incremento en los precios.

Debe señalarse que muchas de las características que se pensaban que iban a influir positivamente en los precios, resultaron siendo no significativas o contribuyendo a la disminución de estos, por lo menos con los datos obtenidos estos fueron los resultados. No obstante, lo ideal sería hacer pruebas eliminando estas características y probando nuevamente el modelo.

6. Referencias

- Bank Locations | Open Data DC.* (n.d.). Retrieved February 16, 2023, from <https://opendata.dc.gov/datasets/DCGIS::bank-locations/about>
- Bonaccorso, G. (n.d.). *Machine Learning Algorithms: Popular algorithms for data science and machine learning.* Retrieved March 3, 2023, from https://books.google.com.co/books?hl=es&lr=&id=_EZsDwAAQBAJ&oi=fnd&pg=P1&dq=machine+learning+data+science&ots=cIVpZWd55E&sig=ure1e_r87iZ8jJl mKYFIRQqrY1M&redir_esc=y#v=onepage&q=machine%20learning%20data%20science&f=false
- Crime Incidents in 2020 | Open Data DC.* (n.d.). Retrieved February 16, 2023, from <https://opendata.dc.gov/datasets/DCGIS::crime-incidents-in-2020/about>
- Dogru, T., & Pekin, O. (2017). *What do guests value most in Airbnb accommodations? An application of the hedonic pricing approach.* www.bu.edu/bhr
- DOMESTIC VISITATION FORECAST THROUGH 2022 Washington, DC (District of Columbia).* (n.d.).
- Dudás, G., Kovalcsik, T., Vida, G., Boros, L., & Nagy, G. (2020). Price determinants of Airbnb listing prices in Lake Balaton Touristic Region. *Hungary. European Journal of Tourism Research*, 24, 2410.
- Experience DC | Seek out the unexpected in Washington, DC.* (n.d.). Retrieved February 24, 2023, from <https://washington.org/experiencedc>
- Inside Airbnb: Get the Data.* (n.d.). Retrieved February 3, 2023, from <http://insideairbnb.com/get-the-data/>
- Investigación de visitantes de Washington, DC | Washington DC.* (n.d.). Retrieved February 24, 2023, from <https://washington.org/es/research/washington-dc-visitor-research#>
- Jiang, Z., & Liu, X. (2020). A Novel Consensus Fuzzy K-Modes Clustering Using Coupling DNA-Chain-Hypergraph P System for Categorical Data. *Processes* 2020, Vol. 8, Page 1326, 8(10), 1326. <https://doi.org/10.3390/PR8101326>
- Jones, V. C., Kennedy, R. D., Welding, K., Gielen, A. C., & Frattaroli, S. (2019). The prevalence of fire and CO safety amenities in Airbnb venues that permit smoking – Findings from 17 countries. *Preventive Medicine*, 123, 8–11. <https://doi.org/10.1016/J.YPMED.2019.02.021>
- Las comodidades que quieren los huéspedes - Centro de recursos de Airbnb - Airbnb.* (n.d.). Retrieved February 16, 2023, from

https://www.airbnb.com.co/resources/hosting-homes/a/the-amenities-guests-want-25?_set_bev_on_new_domain=1672953879_YjhiZGIzODUyODQ5

Limsombunchai, V. (2004). *House Price Prediction: Hedonic Price Model vs. Artificial Neural Network*. <https://doi.org/10.22004/AG.ECON.97781>

Metro Bus Stops | Open Data DC. (n.d.). Retrieved February 16, 2023, from <https://opendata.dc.gov/datasets/DCGIS::metro-bus-stops-2/about>

Montero Granados, R. (2016). Modelos de regresión lineal múltiple. *Documentos de Trabajo En Economía Aplicada. Universidad de Granada. España*.

Open Data DC. (n.d.). Retrieved February 4, 2023, from <https://opendata.dc.gov/>

Ortiz Pinilla, J., & Gil, D. (2014). Transformaciones logarítmicas en regresión simple. *Comunicaciones En Estadística*, 7(1), 80. <https://doi.org/10.15332/s2027-3355.2014.0001.06>

Perez-Sanchez, V. R., Serrano-Estrada, L., Marti, P., & Mora-Garcia, R. T. (2018). The What, Where, and Why of Airbnb Price Determinants. *Sustainability* 2018, Vol. 10, Page 4596, 10(12), 4596. <https://doi.org/10.3390/SU10124596>

Points of Interest | Open Data DC. (n.d.). Retrieved February 16, 2023, from <https://opendata.dc.gov/datasets/DCGIS::points-of-interest/about>

Recreation Facilities | Open Data DC. (n.d.). Retrieved February 16, 2023, from <https://opendata.dc.gov/datasets/DCGIS::recreation-facilities-1/about>

Teubner, T., Hawlitschek, F., & Dann, D. (2017). PRICE DETERMINANTS ON AIRBNB: HOW REPUTATION PAYS OFF IN THE SHARING ECONOMY. *Journal of Self-Governance and Management Economics*, 5(4), 53. <https://doi.org/10.22381/JSME5420173>

Toader, V., Negrușă, A. L., Ruxandra, O., Rozalia, B. & Rus, V., Letiția, A., Negrus, L., Bode, R., & Rus, R. V. (2021). Analysis of price determinants in the case of Airbnb listings. <http://www.tandfonline.com/action/authorSubmission?JournalCode=rero20&page=instructions>, 35(1), 2493–2509. <https://doi.org/10.1080/1331677X.2021.1962380>

Tsai, W.-H., Zhao, C., Wu, Y., Chen, Y., & Chen, G. (2023). Multiscale Effects of Hedonic Attributes on Airbnb Listing Prices Based on MGWR: A Case Study of Beijing, China. *Sustainability* 2023, Vol. 15, Page 1703, 15(2), 1703. <https://doi.org/10.3390/SU15021703>

Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com.



International Journal of Hospitality Management, 62, 120–131.
<https://doi.org/10.1016/J.IJHM.2016.12.007>

Zhang, Z., Chen, R. J. C., Han, L. D., & Yang, L. (2017). Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach. *Sustainability* 2017, Vol. 9, Page 1635, 9(9), 1635. <https://doi.org/10.3390/SU9091635>



Apéndice I

Anexos I

Estadística descriptiva de las principales características

| Variable | Mean | Min | Max | STD | 25% | 50% | 75% |
|--------------------------------------|-----------|----------|-----------|----------|-----------|-----------|-----------|
| price | 179.202 | 10 | 10000 | 271.895 | 86 | 125 | 199 |
| log_price | 4.891 | 2.303 | 9.21 | 0.691 | 4.454 | 4.828 | 5.293 |
| host_response_rate | 0.743 | 0 | 1 | 0.419 | 0.71 | 1 | 1 |
| host_acceptance_rate | 0.746 | 0 | 1 | 0.378 | 0.66 | 0.96 | 1 |
| host_is_superhost | 0.305 | 0 | 1 | 0.46 | 0 | 0 | 1 |
| host_verifications | 2.117 | 1 | 3 | 0.528 | 2 | 2 | 2 |
| host_identity_verified | 0.825 | 0 | 1 | 0.38 | 1 | 1 | 1 |
| latitude | 38.912 | 38.81 | 39.007 | 0.029 | 38.896 | 38.909 | 38.925 |
| longitude | -77.017 | -77.134 | -76.891 | 0.035 | -77.038 | -77.019 | -76.996 |
| accommodates | 3.66 | 1 | 16 | 2.347 | 2 | 3 | 4 |
| bedrooms | 1.523 | 1 | 15 | 0.935 | 1 | 1 | 2 |
| beds | 1.96 | 1 | 20 | 1.376 | 1 | 2 | 2 |
| minimum_nights | 21.41 | 1 | 1125 | 40.722 | 2 | 14 | 31 |
| has_availability | 0.931 | 0 | 1 | 0.254 | 1 | 1 | 1 |
| availability_365 | 175.436 | 0 | 365 | 134.241 | 41.25 | 163 | 314 |
| number_of_reviews | 49.631 | 0 | 875 | 83.457 | 2 | 14 | 60 |
| review_scores_rating | 4.745 | 1 | 5 | 0.337 | 4.73 | 4.79 | 4.94 |
| review_scores_accuracy | 4.806 | 1 | 5 | 0.312 | 4.806 | 4.86 | 4.97 |
| review_scores_cleanliness | 4.758 | 1 | 5 | 0.324 | 4.75 | 4.805 | 4.96 |
| review_scores_checkin | 4.861 | 1 | 5 | 0.291 | 4.861 | 4.92 | 5 |
| review_scores_communication | 4.845 | 1 | 5 | 0.307 | 4.845 | 4.91 | 5 |
| review_scores_location | 4.788 | 1 | 5 | 0.301 | 4.78 | 4.83 | 4.98 |
| review_scores_value | 4.704 | 1 | 5 | 0.346 | 4.68 | 4.74 | 4.89 |
| instant_bookable | 0.41 | 0 | 1 | 0.492 | 0 | 0 | 1 |
| calculated_host_listings_count | 23.577 | 1 | 235 | 54.515 | 1 | 2 | 9 |
| reviews_per_month | 1.82 | 0.01 | 105 | 2.189 | 0.46 | 1.72 | 2.25 |
| bathrooms_quantity | 1.384 | 0.5 | 15 | 0.751 | 1 | 1 | 1.5 |
| min_dist_to_bank_ft | 2237.189 | 8.176 | 13350.622 | 2077.969 | 841.068 | 1564.923 | 2856.735 |
| min_dist_to_recreation_ft | 1967.884 | 20.241 | 10886.456 | 1564.443 | 989.076 | 1552.188 | 2337.291 |
| min_dist_to_metro_bus_stop_ft | 434.585 | 2.767 | 4958.527 | 358.265 | 216.575 | 354.12 | 553.159 |
| min_dist_to_points_interest_ft | 541.98 | 0.62 | 7379.596 | 1145.431 | 90.788 | 170.569 | 354.061 |
| min_dist_to_white_house_ft | 14114.563 | 968.813 | 41443.221 | 8317.53 | 7751.037 | 12106.858 | 18245.82 |
| min_dist_to_lincoln_memorial_ft | 17188.283 | 2641.538 | 45186.85 | 8496.096 | 10936.539 | 15421.941 | 21078.916 |
| min_dist_to_national_space_museum_ft | 14798.973 | 2061.859 | 43972.982 | 8222.306 | 8985.463 | 12167.353 | 18918.784 |
| min_dist_to_washington_memorial_ft | 15435.325 | 2991.984 | 43035.809 | 8146.185 | 9609.556 | 12857.577 | 19012.778 |
| min_dist_to_national_zoo_ft | 15799.359 | 409.169 | 47269.862 | 9043.097 | 8788.844 | 13787.657 | 21216.149 |

| | | | | | | | |
|-------------------------------|---------|-------|----------|----------|---------|---------|---------|
| min_dist_to_crime_incident_ft | 497.238 | 1.369 | 7864.576 | 1190.607 | 100.166 | 146.608 | 220.764 |
|-------------------------------|---------|-------|----------|----------|---------|---------|---------|

Tabla 1. Estadística descriptiva de las principales características.

Fuente: Elaboración propia.

Cantidad de tipos de propiedad

| Tipo de propiedad | Cantidad |
|-----------------------------------|----------|
| Entire rental unit | 2301 |
| Entire home | 681 |
| Private room in home | 571 |
| Entire condo | 545 |
| Entire townhouse | 499 |
| Entire guest suite | 473 |
| Private room in townhouse | 282 |
| Entire serviced apartment | 259 |
| Private room in rental unit | 232 |
| Room in boutique hotel | 112 |
| Private room in condo | 72 |
| Entire guesthouse | 68 |
| Room in hotel | 60 |
| Shared room in rental unit | 47 |
| Private room in guest suite | 35 |
| Entire loft | 33 |
| Private room in bed and breakfast | 32 |
| Shared room in townhouse | 25 |
| Shared room in hostel | 16 |
| Shared room in home | 16 |
| Entire vacation home | 15 |
| Room in aparthotel | 11 |
| Entire bungalow | 10 |
| Room in hostel | 10 |
| Private room in guesthouse | 9 |
| Private room in loft | 6 |
| Entire place | 6 |
| Room in bed and breakfast | 5 |
| Private room in hostel | 4 |
| Room in serviced apartment | 4 |
| Private room | 3 |
| Private room in villa | 3 |
| Casa particular | 3 |
| Shared room in guesthouse | 3 |
| Camper/RV | 2 |

| | |
|------------------------------------|---|
| Private room in bungalow | 2 |
| Shared room in loft | 1 |
| Shared room in hotel | 1 |
| Shared room in serviced apartment | 1 |
| Tent | 1 |
| Tiny home | 1 |
| Boat | 1 |
| Private room in serviced apartment | 1 |
| Private room in resort | 1 |
| Private room in casa particular | 1 |
| Houseboat | 1 |
| Floor | 1 |
| Entire villa | 1 |
| Entire cottage | 1 |
| Campsite | 1 |
| Tower | 1 |

Tabla 2. Cantidad de tipos de propiedad

Fuente: Elaboración propia.

Tabla de reserva inmediata en los alojamientos

| Reserva inmediata (Instant_bookable) | Cantidad de alojamientos | Cantidad (%) | Precio promedio (USD) |
|---|-----------------------------|-----------------|--------------------------|
| No | 3820 | 59 | 176.076 |
| Si | 2650 | 41 | 183.708 |

Tabla 3. Reserva inmediata en los alojamientos.

Fuente: Elaboración propia.

proporción del tipo de alojamiento

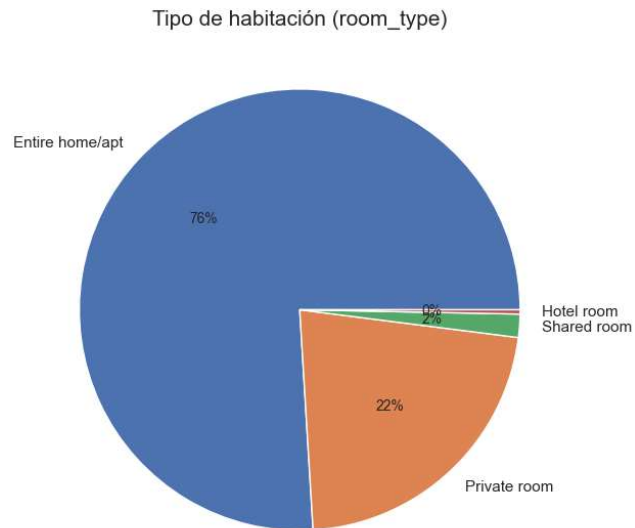


Ilustración 2. Tipo de habitación.

Fuente: Elaboración propia

Precio por tipo de habitación

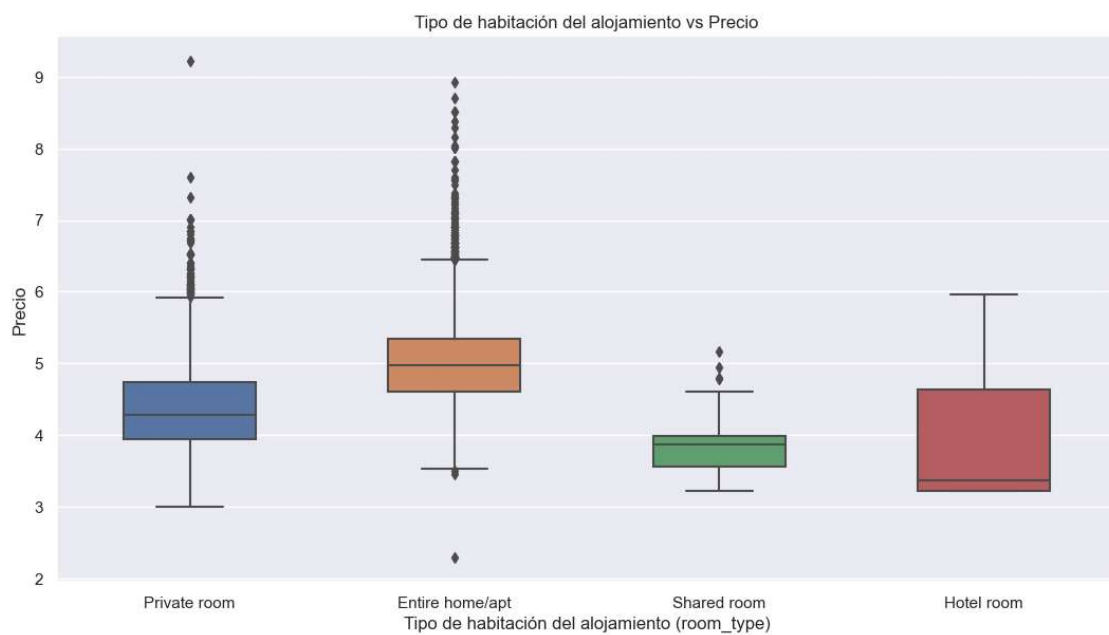


Ilustración 3. Tipo de habitación vs precio.

Fuente: Elaboración propia.

Precio por número de personas permitidas en el alojamiento

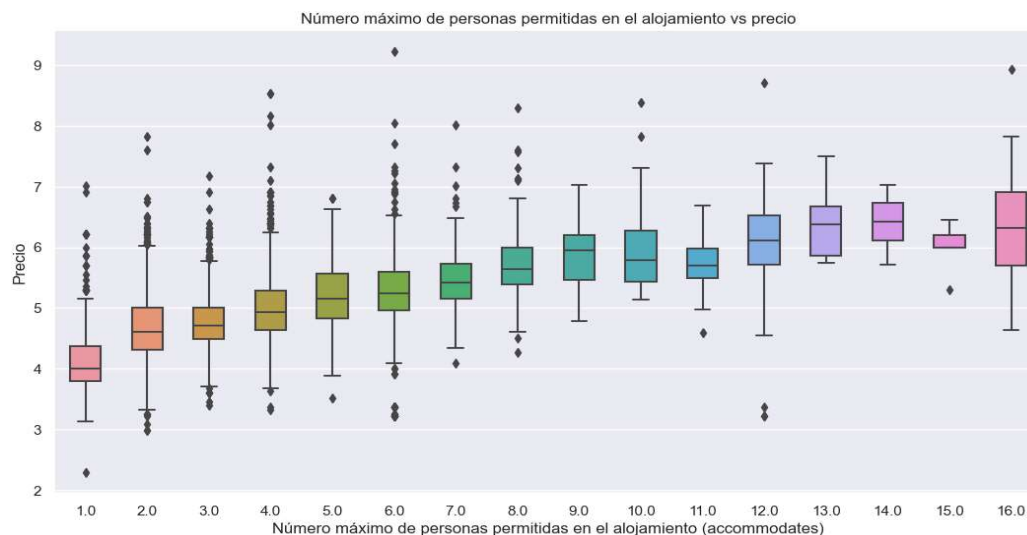


Ilustración 4. Número de personas permitidas en el alojamiento vs precio

Fuente: Elaboración propia.

Correlación de las características del alojamiento

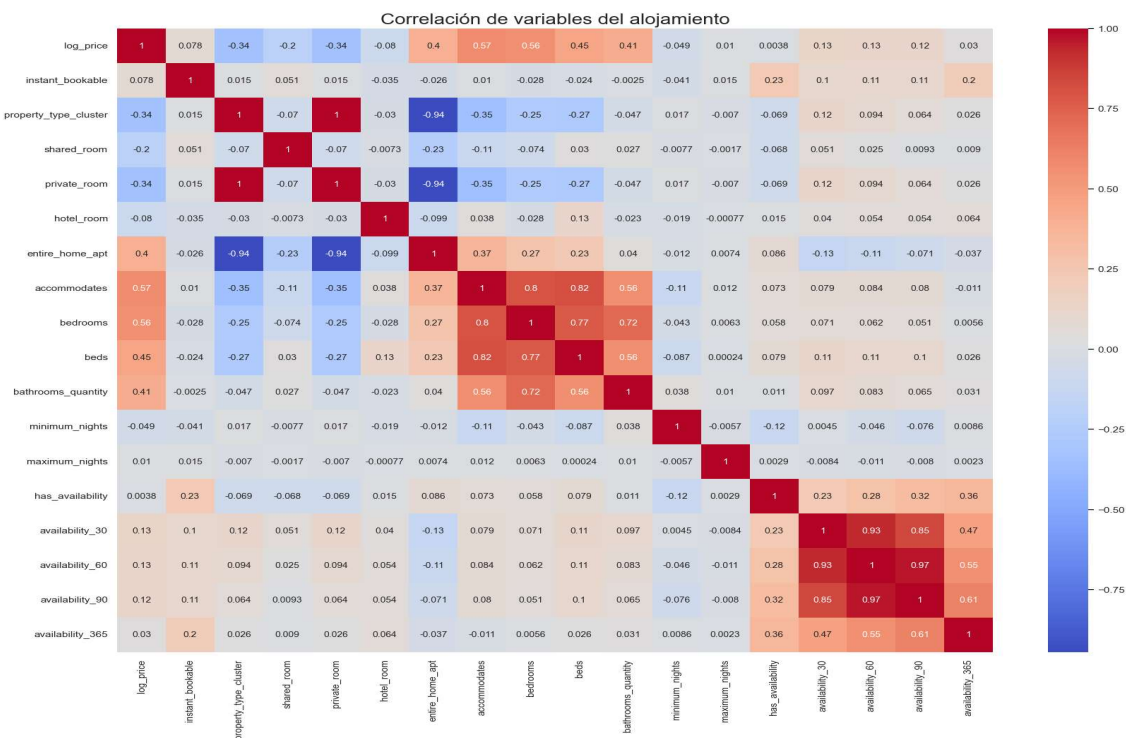


Ilustración 5. Correlación de las características del alojamiento.

Fuente: Elaboración propia.

Correlación de características de servicios adicionales

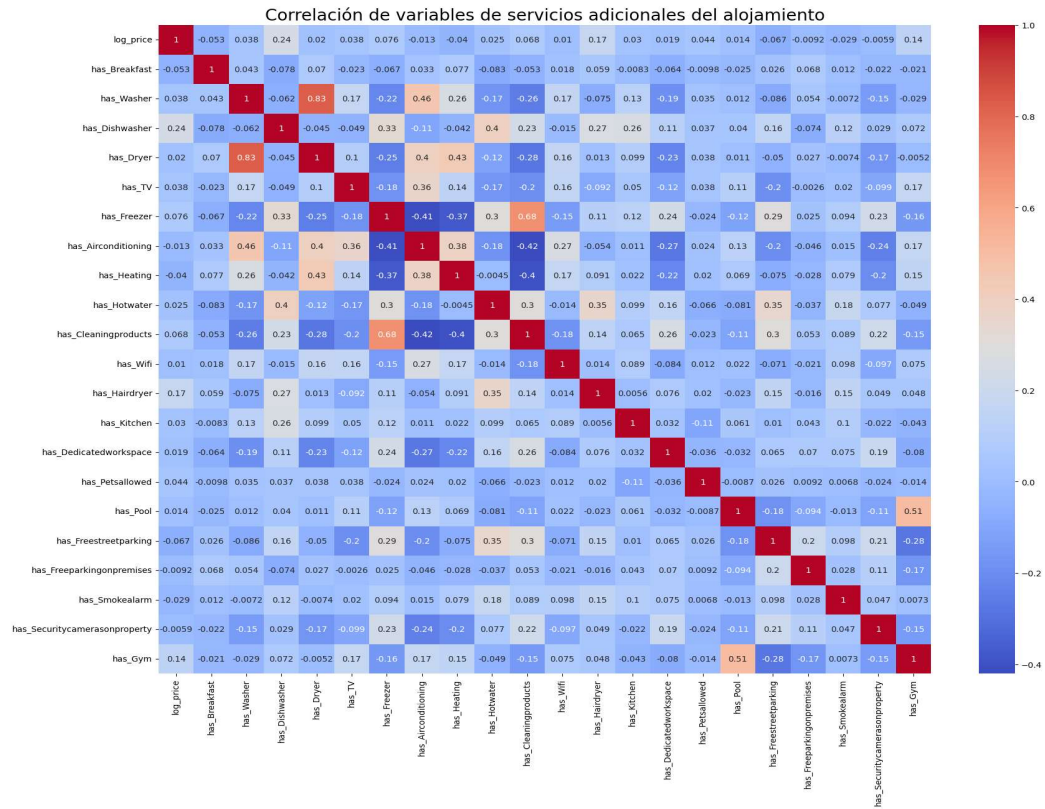


Ilustración 6. Correlación de las características de servicios adicionales.

Fuente: Elaboración propia

Calificaciones por tipo de habitación

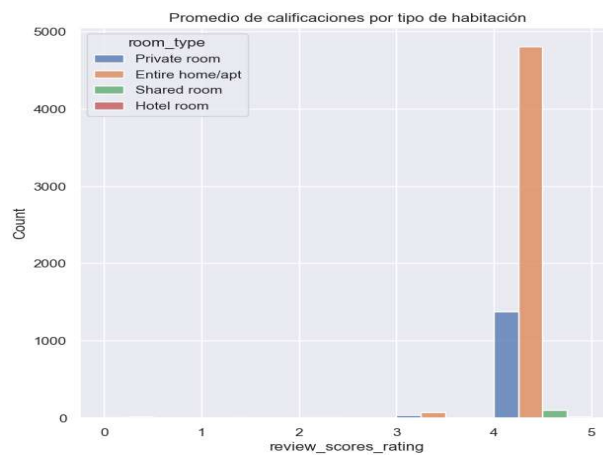


Ilustración 7. Calificaciones por tipo de habitación.

Fuente: Elaboración propia.

Correlación de características de calificaciones de huéspedes

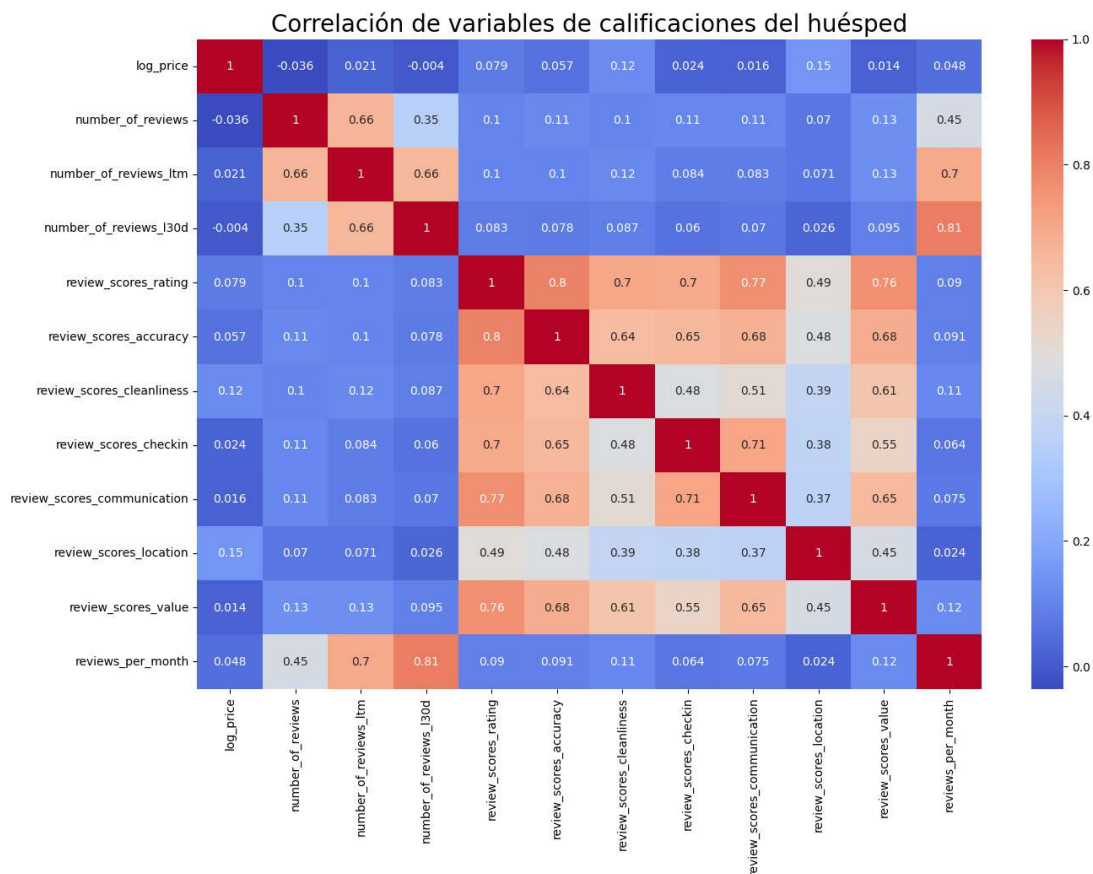


Ilustración 8. Correlación de las características de calificaciones.

Fuente: Elaboración propia

proporción del si un host es superhost

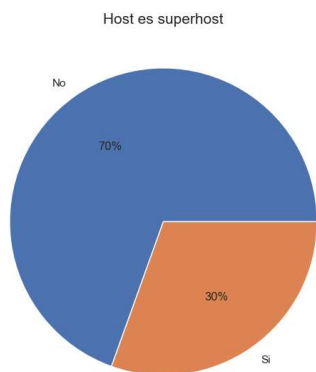


Ilustración 9. Host es superhost.

Fuente: Elaboración propia.

Proporción de la identidad del host y de imagen de perfil

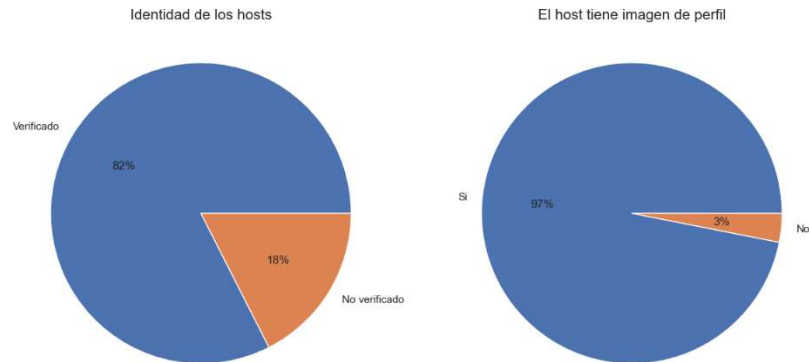


Ilustración 10. Identidad y si el host tiene imagen de perfil.

Fuente: Elaboración propia.

Mapa de promedio de precios de alojamientos por barrio

Promedio de precios de alojamientos de Airbnb por barrio en la ciudad de Washington DC

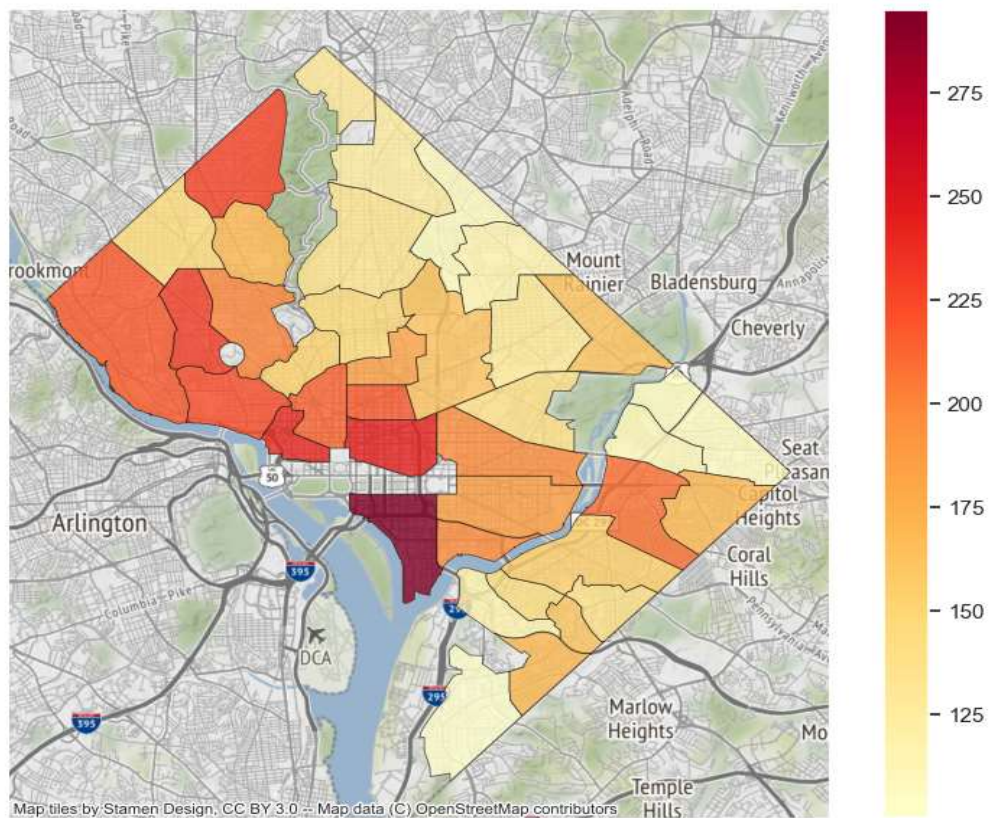


Ilustración 11. Promedio de precios de alojamientos de Airbnb por barrio.

Fuente: Elaboración propia

Mapa de cantidad de alojamientos por barrio

Cantidad de alojamientos de Airbnb por barrio en la ciudad de Washington DC

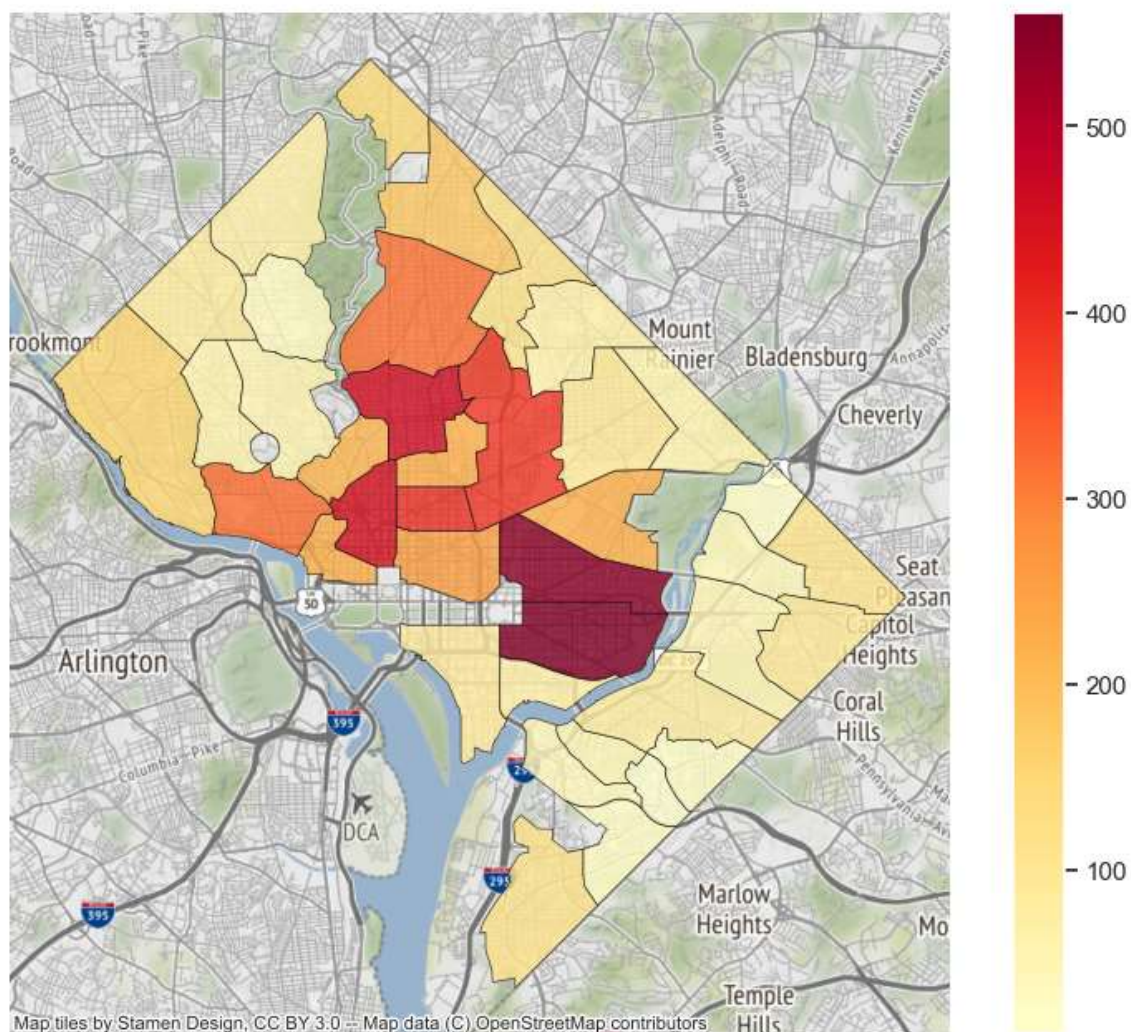


Ilustración 12. Cantidad de alojamientos de Airbnb por barrio en la ciudad.

Fuente: Elaboración propia

Mapa de promedio de puntaje de alojamientos por barrio

Puntaje promedio de alojamientos de Airbnb por barrio en la ciudad de Washington DC

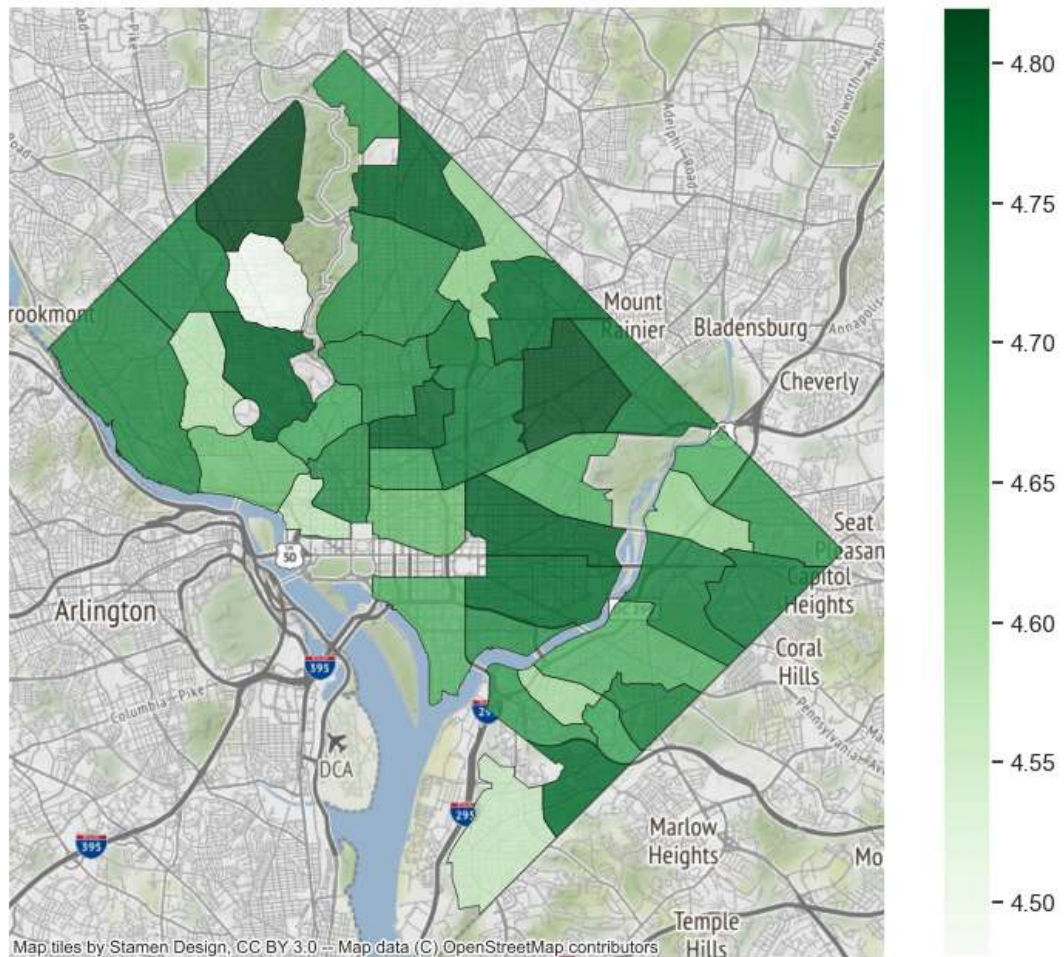


Ilustración 13. Puntaje promedio de alojamientos de Airbnb por barrio en la ciudad.

Fuente: Elaboración propia.

Mapa de promedio de puntaje de alojamientos por barrio

Promedio de disponibilidad por barrio en la ciudad de Washington DC

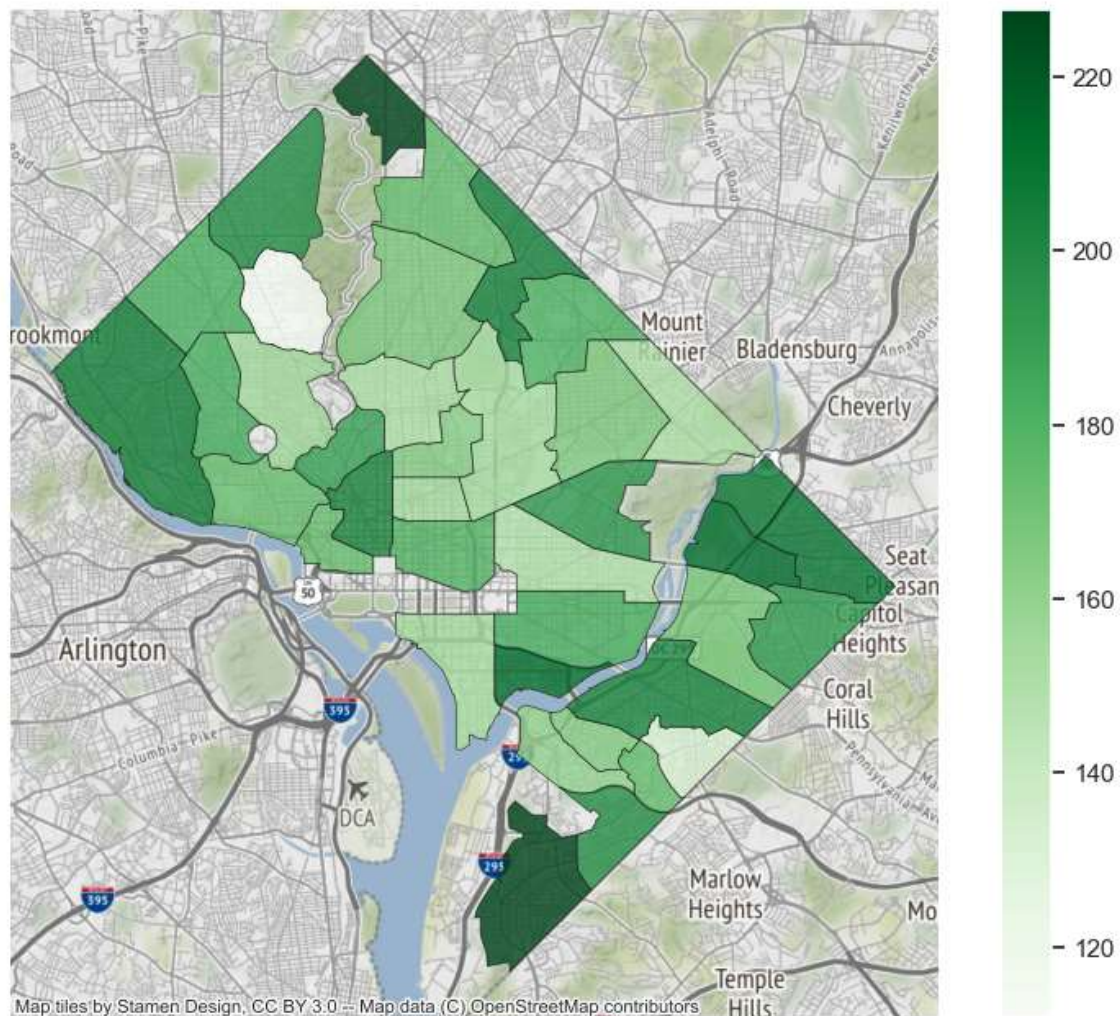


Ilustración 14. Promedio de disponibilidad por barrio en la ciudad.

Fuente: Elaboración propia.

Mapa de promedio de noches mínimas de alojamientos por barrio

Promedio de noches mínimas para hacer reserva por barrio en la ciudad de Washington DC

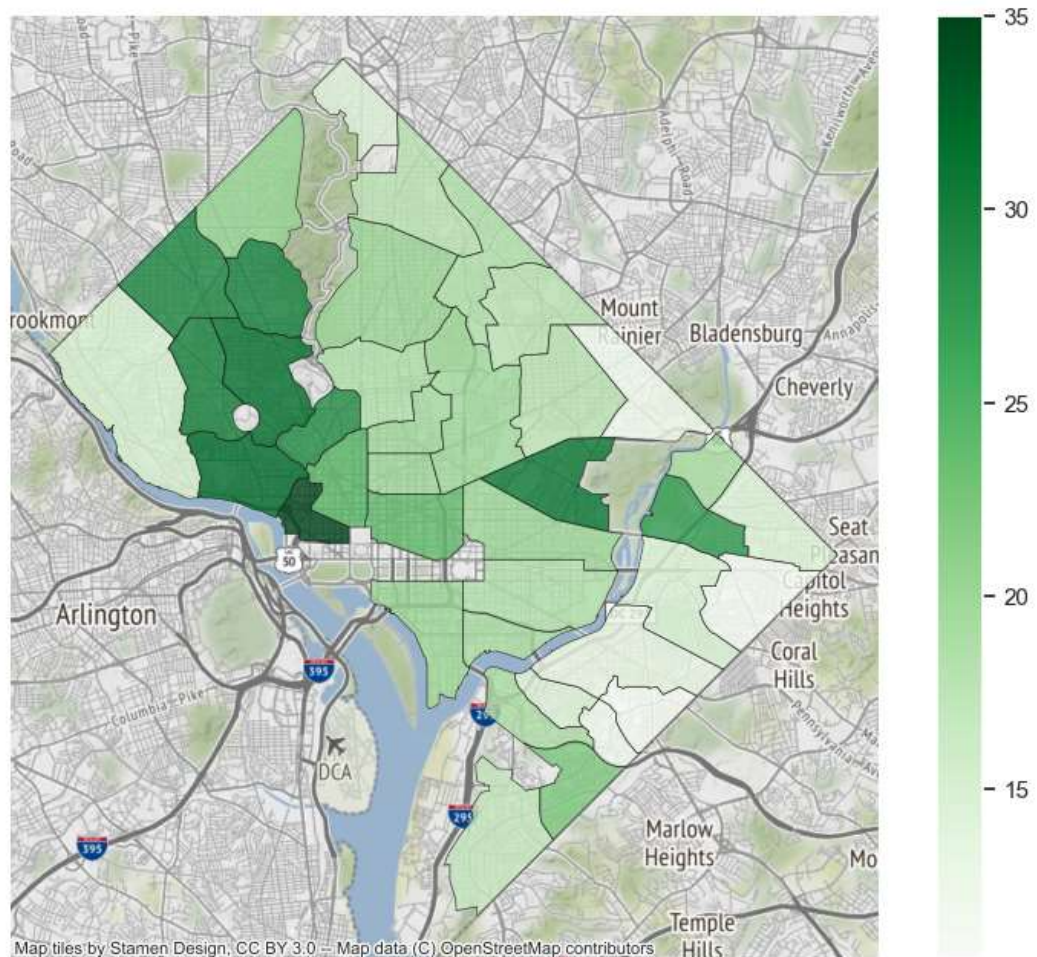


Ilustración 15. Promedio de noches mínimas para hacer reserva por barrio.

Fuente: Elaboración propia

Resumen de estadísticas por barrios

| Barrio | Precio pro- medio | Total aloja- mien- tos | Califica- ción prome- dio | Noches mínimas promedio | Disponibi- lidad promedio |
|---|-------------------------|---------------------------------|------------------------------------|-------------------------------|---------------------------------|
| Southwest Employment Area, South- west/Waterfront, Fort McNair, Buzzard Point | 294.94 | 90 | 4.68 | 20.11 | 153.56 |
| Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street | 242.98 | 253 | 4.66 | 23.77 | 178.52 |
| West End, Foggy Bottom, GWU | 239.84 | 225 | 4.57 | 35 | 173.32 |
| Cathedral Heights, McLean Gardens, Glover Park | 230.74 | 58 | 4.58 | 30.1 | 182.33 |
| Hawthorne, Barnaby Woods, Chevy Chase | 226.71 | 68 | 4.82 | 20.66 | 197.9 |
| Georgetown, Burleith/Hillandale | 222.65 | 320 | 4.66 | 32.01 | 173.37 |
| Shaw, Logan Circle | 218.73 | 395 | 4.7 | 19.87 | 153.58 |
| Spring Valley, Palisades, Wesley Heights, Foxhall Crescent, Foxhall Village, Georgetown Reservoir | 218.46 | 153 | 4.74 | 15.14 | 205.08 |
| Dupont Circle, Connecticut Avenue/K Street | 214.06 | 450 | 4.73 | 24.42 | 201.09 |
| River Terrace, Benning, Greenway, Dupont Park | 208.33 | 49 | 4.75 | 14.45 | 168.47 |
| Cleveland Park, Woodley Park, Massa- chusetts Avenue Heights, Woodland- Normanstone Terrace | 201.11 | 53 | 4.79 | 31.57 | 154.98 |
| Near Southeast, Navy Yard | 198.3 | 69 | 4.73 | 19.12 | 213.33 |
| Union Station, Stanton Park, Kingman Park | 197.8 | 550 | 4.79 | 19.91 | 149.1 |
| Capitol Hill, Lincoln Park | 186.47 | 561 | 4.78 | 18.3 | 182.42 |
| Howard University, Le Droit Park, Cardozo/Shaw | 184.34 | 225 | 4.79 | 19.71 | 164.25 |
| North Cleveland Park, Forest Hills, Van Ness | 169.53 | 32 | 4.48 | 30 | 111.84 |
| Capitol View, Marshall Heights, Benning Heights | 168.86 | 95 | 4.76 | 10.93 | 191.92 |
| Douglas, Shipley Terrace | 168.58 | 26 | 4.78 | 22.08 | 196.5 |
| Edgewood, Bloomingdale, Truxton Circle, Eckington | 163.03 | 386 | 4.76 | 19.83 | 153.66 |
| Woodridge, Fort Lincoln, Gateway | 162.51 | 55 | 4.76 | 11.75 | 150.8 |
| Woodland/Fort Stanton, Garfield Heights, Knox Hill | 161.33 | 12 | 4.7 | 11.75 | 174.17 |
| Twining, Fairlawn, Randle Highlands, Penn Branch, Fort Davis Park, Fort Dupont | 155.44 | 82 | 4.68 | 13.23 | 202.44 |
| Kalorama Heights, Adams Morgan, Lanier Heights | 151.84 | 211 | 4.69 | 28.69 | 187.71 |
| Friendship Heights, American University Park, Tenleytown | 146.12 | 83 | 4.75 | 29.49 | 181.29 |
| Columbia Heights, Mt. Pleasant, Pleasant Plains, Park View | 145.74 | 456 | 4.73 | 18.52 | 150.94 |
| Historic Anacostia | 141.54 | 46 | 4.6 | 14.41 | 164.22 |
| Ivy City, Arboretum, Trinidad, Carver Langston | 140.57 | 223 | 4.68 | 31 | 193.1 |

| | | | | | |
|--|--------|-----|------|-------|--------|
| Fairfax Village, Naylor Gardens, Hillcrest, Summit Park | 139.48 | 21 | 4.75 | 10.62 | 132.24 |
| Brightwood Park, Crestwood, Petworth | 134.78 | 316 | 4.73 | 20.23 | 166.9 |
| Brookland, Brentwood, Langdon | 129.82 | 94 | 4.81 | 17.83 | 166.73 |
| Colonial Village, Shepherd Park, North Portal Estates | 127.81 | 140 | 4.72 | 14.39 | 227.71 |
| Takoma, Brightwood, Manor Park | 124.57 | 179 | 4.79 | 17.81 | 173.55 |
| Sheridan, Barry Farm, Buena Vista | 116.84 | 38 | 4.71 | 18.79 | 160.16 |
| North Michigan Park, Michigan Park, University Heights | 115.63 | 67 | 4.78 | 17.78 | 180.61 |
| Mayfair, Hillbrook, Mahanang Heights | 110.32 | 38 | 4.61 | 27.29 | 213.37 |
| Lamont Riggs, Queens Chapel, Fort Totten, Pleasant Hill | 107.13 | 116 | 4.62 | 18.44 | 203.47 |
| Eastland Gardens, Kenilworth | 106 | 12 | 4.68 | 19 | 207.17 |
| Deanwood, Burrville, Grant Park, Lincoln Heights, Fairmont Heights | 105.96 | 110 | 4.72 | 14.46 | 209.04 |
| Congress Heights, Bellevue, Washington Highlands | 100.07 | 113 | 4.56 | 16.84 | 227.69 |

Tabla 4. Resumen de estadísticas por barrios.

Fuente: Elaboración propia.

Correlación de variables del host

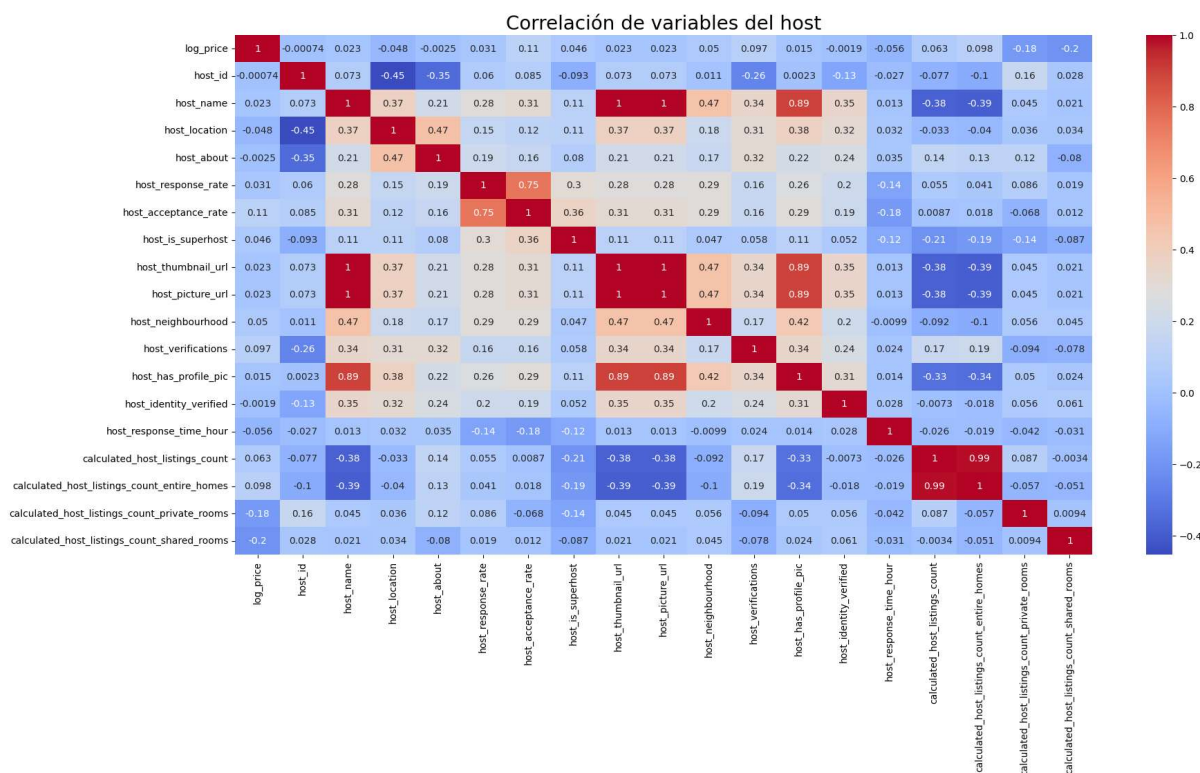


Ilustración 16. Correlación de variables del host.

Fuente: Elaboración propia.

Correlación de variables de ubicación

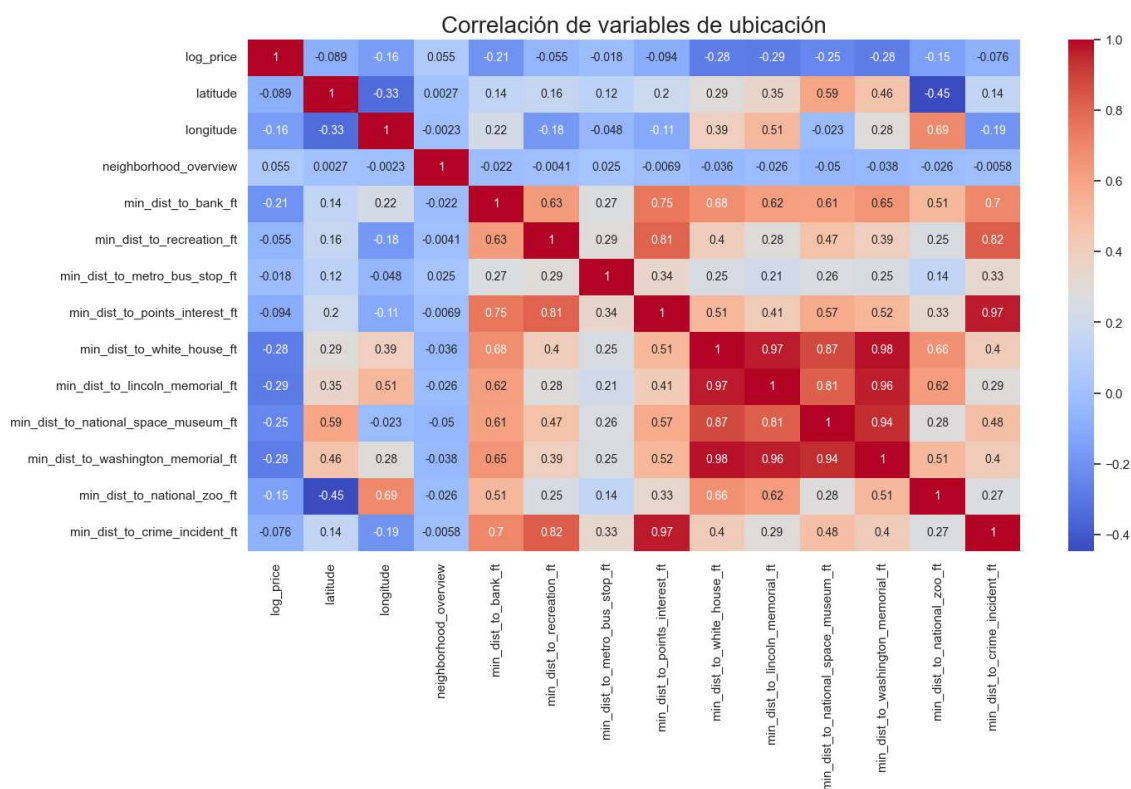


Ilustración 17. Correlación de variables de ubicación.

Fuente: Elaboración propia.

Determinantes del precio por noche para los modelos OLS y QR

Coeficientes y entre paréntesis los errores estándares.

significancia estadística *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

| Categoría | Característica | OLS | Q 0.1 | Q 0.25 | Q 0.5 | Q 0.75 | Q 0.9 |
|--|------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| características del alojamiento | shared_room | -0.0828 *** (0.007) | -0.0668 *** (0.0071) | -0.0558 *** (0.0063) | -0.0643 *** (0.0058) | -0.0943 *** (0.008) | -0.1214 *** (0.0116) |
| | beds | -0.0442 * (0.0223) | -0.1296 *** (0.0266) | -0.0792 *** (0.0207) | -0.0036 (0.0186) | -0.0084 (0.0251) | 0.0059 (0.0376) |
| | availability_365 | 0.0163 *** (0.0025) | 0.0182 *** (0.0024) | 0.0169 *** (0.0022) | 0.0127 *** (0.0021) | 0.0125 *** (0.0029) | 0.0159 *** (0.0041) |
| | minimum_nights | -0.0615 * (0.0246) | -0.6316 *** (0.0131) | -0.4932 *** (0.015) | -0.2707 *** (0.0205) | -0.1285 *** (0.0383) | 0.1012 (0.0694) |
| | bathrooms_quantity | 0.1964 *** (0.0243) | 0.155 *** (0.029) | 0.2209 *** (0.0232) | 0.2054 *** (0.0202) | 0.183 *** (0.0246) | 0.3271 *** (0.0318) |
| | instant_bookable | 0.0033 (0.0019) | 0.0064 ** (0.0021) | 0.0094 *** (0.0018) | 0.0055 *** (0.0016) | 0.0012 (0.0021) | -0.0019 (0.0028) |
| | bedrooms | 0.3872 *** (0.0265) | 0.4779 *** (0.0349) | 0.3902 *** (0.0259) | 0.4051 *** (0.0221) | 0.4158 *** (0.0283) | 0.3848 *** (0.0413) |
| | accommodates | 0.1591 *** (0.0111) | 0.1134 *** (0.0126) | 0.1202 *** (0.0106) | 0.1158 *** (0.0093) | 0.1687 *** (0.0121) | 0.1712 *** (0.0175) |
| | entire_home_apartment | 0.0404 *** (0.0025) | 0.0572 *** (0.0025) | 0.0561 *** (0.0022) | 0.0514 *** (0.002) | 0.0329 *** (0.0027) | 0.023 *** (0.0036) |
| | hotel_room | -0.1071 *** (0.0157) | -0.0931 *** (0.0154) | -0.13 *** (0.0137) | -0.1556 *** (0.0131) | -0.0485 ** (0.0185) | 0.0455 (0.0253) |
| | has_Cleaningproducts | 0.002 (0.0027) | 0.0024 (0.0028) | 0.0051 * (0.0024) | 0.0025 (0.0022) | 0 (0.0029) | -0.0023 (0.004) |
| | has_Freestreetparking | -0.0071 *** (0.0021) | 0.0044 * (0.0022) | 0.0018 (0.0019) | -0.004 * (0.0017) | -0.0095 *** (0.0023) | -0.0139 *** (0.0029) |
| | has_Pool | -0.0177 *** (0.0047) | -0.0247 *** (0.0043) | -0.0254 *** (0.0038) | -0.0194 *** (0.0039) | -0.0065 (0.0056) | -0.0014 (0.0075) |
| | has_Petsallowed | 0.0052 (0.0225) | -0.0202 (0.0239) | 0.039 (0.022) | 0.0268 (0.0187) | 0.0191 (0.024) | 0.0059 (0.0347) |
| | has_Dedicatedworkspace | -0.0043 * (0.0019) | -0.0036 (0.0019) | -0.0053 ** (0.0017) | -0.001 (0.0016) | -0.0022 (0.0021) | -0.0056 * (0.0028) |
| Características de servicios adicionales | has_Kitchen | -0.0193 *** (0.003) | -0.0047 (0.0032) | -0.0082 ** (0.0027) | -0.0128 *** (0.0025) | -0.024 *** (0.0031) | -0.0207 *** (0.0041) |
| | has_Hairdryer | 0.0123 *** (0.0024) | 0.0113 *** (0.0025) | 0.0103 *** (0.0021) | 0.0091 *** (0.002) | 0.0105 (0.0026) | 0.0063 (0.0037) |
| | has_Wifi | 0.0043 (0.0031) | -0.0018 (0.0032) | -0.0004 (0.0027) | 0.0014 (0.0026) | 0.0065 (0.0034) | 0.0156 *** (0.0046) |
| | has_Dishwasher | 0.0117 *** (0.0021) | 0.0156 *** (0.0022) | 0.0131 *** (0.0019) | 0.0133 *** (0.0018) | 0.0114 *** (0.0023) | 0.008 * (0.0031) |
| | has_Hotwater | -0.0129 *** (0.0025) | -0.013 *** (0.0027) | -0.0123 *** (0.0022) | -0.0056 ** (0.002) | -0.0132 (0.0027) | -0.0237 *** (0.0035) |
| | has_Airconditioning | 0.0043 (0.0024) | -0.0008 (0.0025) | 0.0002 (0.0021) | 0.0002 (0.002) | 0.0082 (0.0027) | 0.0091 * (0.0036) |

| | | | | | | | |
|--|--------------------------------------|-------------------------|---------------------------|---------------------------|----------------------------|----------------------------|----------------------------|
| | | | | | (0.0026) | | |
| | has_Freezer | 0.0026 (0.0026) | 0.0023 (0.0027) | 0.0054 * (0.0024) | 0.003 (0.0022) | 0.0078 ** (0.0029) | 0.0101 * (0.004) |
| | has_TV | 0.0001 (0.0019) | 0.0038 (0.002) | 0.0019 (0.0017) | -0.0008 (0.0016) | -0.0056 ** (0.0021) | -0.0068 * (0.0028) |
| | has_Smokealarm | -0.0329 *** (0.0049) | -0.0094 (0.0051) | -0.0083 (0.0044) | -0.0181 *** (0.0041) | -0.0412 *** (0.0054) | -0.0623 *** (0.0074) |
| | has_Washer | -0.0048 * (0.0021) | -0.003 (0.0021) | -0.0005 (0.0018) | 0.0009 (0.0017) | -0.0058 * (0.0023) | -0.0083 ** (0.0032) |
| | has_Breakfast | 0.0074 * (0.0035) | -0.0048 (0.0036) | -0.0029 (0.0031) | 0.0017 (0.0029) | 0.0162 *** (0.0039) | 0.0292 *** (0.0055) |
| | has_Freeparkingonpremises | 0.0067 ** (0.002) | 0.0015 (0.002) | 0.0019 (0.0018) | 0.002 (0.0017) | 0.0054 * (0.0023) | 0.0067 * (0.0031) |
| | has_Securitycamerasonproperty | -0.0027 (0.002) | 0.0055 ** (0.0021) | 0.0004 (0.0018) | -0.0028 (0.0017) | -0.0037 (0.0023) | -0.0082 ** (0.0031) |
| | has_Gym | 0.0354 *** (0.0033) | 0.0367 *** (0.0036) | 0.0326 *** (0.003) | 0.0318 *** (0.0027) | 0.034 *** (0.0038) | 0.0306 *** (0.0052) |
| Características de calificaciones del huésped | number_of_reviews | -0.0743 *** (0.0101) | -0.0064 (0.0099) | -0.0252 ** (0.0085) | -0.0459 *** (0.0084) | -0.0838 *** (0.0124) | -0.1016 *** (0.0199) |
| | review_scores_rating | 0.0812 *** (0.0226) | 0.0242 (0.024) | 0.0432 * (0.0208) | 0.0898 *** (0.0188) | 0.1052 *** (0.0264) | 0.0799 * (0.0363) |
| | review_scores_accuracy | 0.002 (0.0188) | 0.0042 (0.0183) | -0.005 (0.0174) | -0.0017 (0.0156) | -0.0121 (0.0203) | 0.0105 (0.0247) |
| | review_scores_cleanliness | 0.0899 *** (0.0151) | 0.0881 *** (0.0157) | 0.084 *** (0.0138) | 0.0745 *** (0.0125) | 0.0846 *** (0.0173) | 0.0832 *** (0.0247) |
| | review_scores_checkin | -0.0534 ** (0.0175) | -0.0587 ** (0.0196) | -0.0304 (0.0167) | -0.0528 *** (0.0145) | -0.0738 *** (0.0194) | -0.069 ** (0.0238) |
| | review_scores_communication | -0.0161 (0.0186) | 0.0162 (0.0181) | -0.02 (0.0172) | -0.0053 (0.0155) | -0.0314 (0.0207) | -0.0367 (0.0256) |
| | review_scores_location | 0.0908 *** (0.0143) | 0.0882 *** (0.0146) | 0.097 *** (0.0122) | 0.0967 *** (0.0119) | 0.1255 *** (0.0176) | 0.0993 *** (0.0232) |
| | review_scores_value | -0.0824 *** (0.016) | -0.0328 (0.0183) | -0.0353 * (0.016) | -0.0859 *** (0.0133) | -0.1109 *** (0.0174) | -0.1054 *** (0.0231) |
| | host_verifications | 0.0164 *** (0.0037) | 0.0078 * (0.0036) | 0.0122 *** (0.0032) | 0.012 *** (0.0031) | 0.0147 *** (0.0041) | 0.0157 ** (0.0056) |
| | host_has_profile_pic | 0.0309 *** (0.0066) | 0.0275 *** (0.006) | 0.024 *** (0.0055) | 0.0337 *** (0.0055) | 0.045 *** (0.0082) | 0.052 *** (0.0126) |
| Características del host | calculated_host_listings_count | -0.0291 *** (0.0053) | 0.0194 *** (0.0051) | 0.006 ** (0.0045) | -0.0116 *** (0.0044) | -0.0305 *** (0.0064) | -0.0437 *** (0.0094) |
| | host_acceptance_rate | 0.0065 * (0.003) | 0.0105 *** (0.0029) | 0.007 ** (0.0026) | 0.0078 ** (0.0025) | 0.0062 (0.0035) | -0.0204 *** (0.0049) |
| | host_identity_verified | -0.0067 ** (0.0024) | -0.0006 (0.0024) | -0.0025 (0.0021) | -0.0034 (0.002) | -0.0057 * (0.0027) | -0.0063 (0.0038) |
| | host_response_time_hour | -0.0249 *** (0.0058) | 0.0077 (0.0062) | -0.0038 (0.005) | -0.0092 (0.0049) | -0.0309 *** (0.0068) | -0.0494 *** (0.01) |
| | host_is_superhost | 0.0036 (0.0022) | 0.0093 *** (0.0022) | 0.0055 ** (0.0019) | 0.0038 * (0.0018) | -0.0025 (0.0024) | -0.0014 (0.0032) |
| | host_about | -0.0042 * (0.0019) | -0.0022 (0.002) | -0.0024 (0.0017) | -0.0027 (0.0016) | -0.0058 ** (0.0022) | -0.0062 * (0.003) |
| | min_dist_to_white_house_ft | -0.2469 ** (0.0873) | -0.2622 ** (0.0906) | -0.22 ** (0.0792) | -0.1315 (0.0727) | -0.1422 (0.0971) | -0.3077 * (0.1376) |
| | min_dist_to_washington_memorial_ft | -0.0681 (0.239) | -0.079 (0.2489) | -0.2201 (0.2178) | -0.3798 (0.1989) | -0.2021 (0.261) | 0.1526 (0.3618) |
| | min_dist_to_national_space_museum_ft | 0.1576 (0.0906) | 0.1538 (0.0934) | 0.2109 * (0.0822) | 0.2635 *** (0.0754) | 0.1926 (0.0989) | 0.115 (0.1356) |
| | min dist to lincoln memorial ft | -0.0025 | 0.0559 | 0.0984 | 0.1284 | 0.013 | -0.1513 |
| Características de ubicación | | | | | | | |

| | | | | | | |
|--------------------------------|-------------------------|-----------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | (0.096) | (0.0968) | (0.086) | (0.0799) | (0.1052) | (0.1431) |
| min_dist_to_national_zoo_ft | 0.101 *** (0.0172) | 0.0898 *** | 0.0858 *** | 0.0823 *** | 0.0947 *** | 0.148 *** |
| min_dist_to_points_interest_ft | 0.1792 *** (0.0274) | 0.0827 ** (0.0301) | 0.1486 *** (0.0257) | 0.1012 *** (0.0228) | 0.1064 *** (0.0295) | 0.1223 ** (0.0415) |
| min_dist_to_metro_bus_stop_ft | 0.0382 ** (0.0123) | 0.0141 (0.0114) | -0.0097 (0.0105) | 0.0096 (0.0103) | 0.044 ** (0.0144) | 0.0473 * (0.0191) |
| min_dist_to_recreation_ft | 0.0234 * (0.0106) | 0.0026 (0.0107) | -0.0045 (0.0093) | 0.0226 * (0.0088) | 0.031 ** (0.0119) | 0.0298 (0.0167) |
| min_dist_to_bank_ft | -0.0473 *** (0.0096) | -0.0432 *** (0.01) | -0.0288 *** (0.0085) | -0.0371 *** (0.008) | -0.0639 *** (0.0108) | -0.0741 *** (0.0144) |
| min_dist_to_crime_incident_ft | -0.1863 *** (0.0278) | -0.0679 * (0.031) | -0.1288 *** (0.0263) | -0.1003 *** (0.0232) | -0.1017 *** (0.0294) | -0.1154 ** (0.0407) |

Tabla 5. Determinantes del precio por noche para OLS y QR.

Porcentajes de incremento / disminución en el precio dependiendo del coeficiente

| Característica | OLS | Q 0.1 | Q 0.25 | Q 0.5 | Q 0.75 | Q 0.9 |
|--------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| shared_room | -7.94648 | -6.46177 | -5.42717 | -6.22764 | -8.99903 | -11.43204 |
| beds | -4.32374 | -12.15533 | -7.61449 | -0.35935 | -0.83648 | 0.59174 |
| availability_365 | 1.64336 | 1.83666 | 1.70436 | 1.27810 | 1.25785 | 1.60271 |
| minimum_nights | -5.96471 | -46.82597 | -38.93309 | -23.71547 | -12.05864 | 10.64979 |
| bathrooms_quantity | 21.70136 | 16.76580 | 24.71987 | 22.80162 | 20.08144 | 38.69402 |
| instant_bookable | 0.33055 | 0.64205 | 0.94443 | 0.55152 | 0.12007 | -0.18982 |
| bedrooms | 47.28510 | 61.26842 | 47.72762 | 49.94524 | 51.55827 | 46.93204 |
| accommodates | 17.24552 | 12.00799 | 12.77224 | 12.27713 | 18.37650 | 18.67281 |
| entire_home_apartment | 4.12272 | 5.88676 | 5.77034 | 5.27439 | 3.34472 | 2.32665 |
| hotel_room | -10.15642 | -8.88976 | -12.19046 | -14.40985 | -4.73427 | 4.65510 |
| has_Cleaningproducts | 0.20020 | 0.24029 | 0.51130 | 0.25031 | 0.00000 | -0.22974 |
| has_Freestreetparking | -0.70749 | 0.44097 | 0.18016 | -0.39920 | -0.94550 | -1.38038 |
| has_Pool | -1.75443 | -2.43975 | -2.50801 | -1.92130 | -0.64789 | -0.13990 |
| has_Petsallowed | 0.52135 | -1.99973 | 3.97705 | 2.71623 | 1.92836 | 0.59174 |
| has_Dedicatedworkspace | -0.42908 | -0.35935 | -0.52860 | -0.09995 | -0.21976 | -0.55843 |
| has_Kitchen | -1.91149 | -0.46890 | -0.81665 | -1.27184 | -2.37143 | -2.04872 |
| has_Hairdryer | 1.23760 | 1.13641 | 1.03532 | 0.91415 | 1.05553 | 0.63199 |
| has_Wifi | 0.43093 | -0.17984 | -0.03999 | 0.14010 | 0.65212 | 1.57223 |
| has_Dishwasher | 1.17687 | 1.57223 | 1.31862 | 1.33888 | 1.14652 | 0.80321 |
| has_Hotwater | -1.28172 | -1.29159 | -1.22247 | -0.55843 | -1.31133 | -2.34214 |
| has_Airconditioning | 0.43093 | -0.07997 | 0.02000 | 0.02000 | 0.82337 | 0.91415 |
| has_Freezer | 0.26034 | 0.23026 | 0.54146 | 0.30045 | 0.78305 | 1.01512 |
| has_TV | 0.01000 | 0.38072 | 0.19018 | -0.07997 | -0.55843 | -0.67769 |
| has_Smokealarm | -3.23647 | -0.93560 | -0.82657 | -1.79372 | -4.03628 | -6.03990 |
| has_Washer | -0.47885 | -0.29955 | -0.04999 | 0.09004 | -0.57832 | -0.82657 |
| has_Breakfast | 0.74274 | -0.47885 | -0.28958 | 0.17014 | 1.63319 | 2.96305 |
| has_Freeparkingonpremises | 0.67225 | 0.15011 | 0.19018 | 0.20020 | 0.54146 | 0.67225 |
| has_Securitycamerasonproperty | -0.26964 | 0.55152 | 0.04001 | -0.27961 | -0.36932 | -0.81665 |
| has_Gym | 3.60340 | 3.73818 | 3.31372 | 3.23110 | 3.45846 | 3.10730 |
| number_of_reviews | -7.16069 | -0.63796 | -2.48851 | -4.48625 | -8.03848 | -9.66092 |
| review_scores_rating | 8.45878 | 2.44952 | 4.41467 | 9.39555 | 11.09328 | 8.31787 |
| review_scores_accuracy | 0.20020 | 0.42088 | -0.49875 | -0.16986 | -1.20271 | 1.05553 |
| review_scores_cleanliness | 9.40649 | 9.20973 | 8.76289 | 7.73453 | 8.82817 | 8.67591 |
| review_scores_checkin | -5.19993 | -5.70104 | -2.99426 | -5.14303 | -7.11426 | -6.66733 |
| review_scores_communication | -1.59711 | 1.63319 | -1.98013 | -0.52860 | -3.09121 | -3.60347 |
| review_scores_location | 9.50500 | 9.22065 | 10.18604 | 10.15299 | 13.37152 | 10.43976 |
| review_scores_value | -7.90965 | -3.22679 | -3.46842 | -8.23140 | -10.49718 | -10.00355 |
| host_verifications | 1.65352 | 0.78305 | 1.22747 | 1.20723 | 1.48086 | 1.58239 |
| host_has_profile_pic | 3.13824 | 2.78816 | 2.42903 | 3.42743 | 4.60279 | 5.33757 |
| calculated_host_listings_count | -2.86807 | 1.95894 | 0.60180 | -1.15330 | -3.00396 | -4.27589 |

| | | | | | | |
|--------------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| host_acceptance_rate | 0.65212 | 1.05553 | 0.70246 | 0.78305 | 0.62193 | -2.01933 |
| host_identity_verified | -0.66776 | -0.05998 | -0.24969 | -0.33942 | -0.56838 | -0.62802 |
| host_response_time_hour | -2.45926 | 0.77297 | -0.37928 | -0.91578 | -3.04275 | -4.81997 |
| host_is_superhost | 0.36065 | 0.93434 | 0.55152 | 0.38072 | -0.24969 | -0.13990 |
| host_about | -0.41912 | -0.21976 | -0.23971 | -0.26964 | -0.57832 | -0.61808 |
| min_dist_to_white_house_ft | -21.87812 | -23.06429 | -19.74812 | -12.32207 | -13.25523 | -26.48642 |
| min_dist_to_washington_memorial_ft | -6.58329 | -7.59601 | -19.75614 | -31.60018 | -18.29868 | 16.48589 |
| min_dist_to_national_space_museum_ft | 17.06978 | 16.62576 | 23.47889 | 30.14773 | 21.23977 | 12.18734 |
| min_dist_to_lincoln_memorial_ft | -0.24969 | 5.74919 | 10.34041 | 13.70077 | 1.30849 | -14.04102 |
| min_dist_to_national_zoo_ft | 10.62766 | 9.39555 | 8.95884 | 8.57815 | 9.93290 | 15.95129 |
| min_dist_to_points_interest_ft | 19.62600 | 8.62159 | 16.02088 | 10.64979 | 11.22667 | 13.00931 |
| min_dist_to_metro_bus_stop_ft | 3.89390 | 1.41999 | -0.96531 | 0.96462 | 4.49824 | 4.84365 |
| min_dist_to_recreation_ft | 2.36759 | 0.26034 | -0.44899 | 2.28573 | 3.14855 | 3.02485 |
| min_dist_to_bank_ft | -4.61988 | -4.22802 | -2.83892 | -3.64202 | -6.19012 | -7.14212 |
| min_dist_to_crime_incident_ft | -16.99754 | -6.56461 | -12.08502 | -9.54340 | -9.66995 | -10.89903 |

Tabla 6. Porcentajes de incremento / disminución en el precio.

Fuente: Elaboración propia.

P-valores obtenidos de cada modelo de regresión

| Característica | OLS | Q 0.1 | Q 0.25 | Q 0.5 | Q 0.75 | Q 0.9 |
|-------------------------------|---------|---------|---------|---------|---------|---------|
| shared_room | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| beds | 0.04770 | 0.00000 | 0.00010 | 0.84650 | 0.73860 | 0.87430 |
| availability_365 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00010 |
| minimum_nights | 0.01240 | 0.00000 | 0.00000 | 0.00000 | 0.00080 | 0.14520 |
| bathrooms_quantity | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| instant_bookable | 0.08390 | 0.00230 | 0.00000 | 0.00050 | 0.56370 | 0.48630 |
| bedrooms | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| accommodates | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| entire_home_apartment | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| hotel_room | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00870 | 0.07190 |
| has_cleaning_products | 0.44220 | 0.37520 | 0.03120 | 0.24970 | 0.99320 | 0.55850 |
| has_freestreetparking | 0.00070 | 0.03940 | 0.32220 | 0.02160 | 0.00000 | 0.00000 |
| has_pool | 0.00010 | 0.00000 | 0.00000 | 0.00000 | 0.24500 | 0.84650 |
| has_pets_allowed | 0.81600 | 0.39820 | 0.07700 | 0.15290 | 0.42510 | 0.86510 |
| has_dedicated_workspace | 0.02100 | 0.05850 | 0.00150 | 0.53090 | 0.28410 | 0.04610 |
| has_kitchen | 0.00000 | 0.14570 | 0.00250 | 0.00000 | 0.00000 | 0.00000 |
| has_hairdryer | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00010 | 0.08760 |
| has_wifi | 0.15790 | 0.56670 | 0.89180 | 0.57660 | 0.05710 | 0.00060 |
| has_dishwasher | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.01040 |
| has_hotwater | 0.00000 | 0.00000 | 0.00000 | 0.00600 | 0.00000 | 0.00000 |
| has_airconditioning | 0.07270 | 0.76420 | 0.93140 | 0.90630 | 0.00170 | 0.01110 |
| has_freezer | 0.33290 | 0.39090 | 0.02180 | 0.16890 | 0.00730 | 0.01180 |
| has_tv | 0.96660 | 0.06370 | 0.26100 | 0.59370 | 0.00680 | 0.01590 |
| has_smokealarm | 0.00000 | 0.06420 | 0.05890 | 0.00000 | 0.00000 | 0.00000 |
| has_washer | 0.02050 | 0.14640 | 0.76790 | 0.60670 | 0.01090 | 0.00950 |
| has_breakfast | 0.03290 | 0.17540 | 0.34390 | 0.55850 | 0.00000 | 0.00000 |
| has_freeparkingonpremises | 0.00110 | 0.45700 | 0.29460 | 0.22710 | 0.01870 | 0.02980 |
| has_securitycamerasonproperty | 0.18400 | 0.00860 | 0.81360 | 0.09600 | 0.09990 | 0.00780 |
| has_gym | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| number_of_reviews | 0.00000 | 0.51530 | 0.00300 | 0.00000 | 0.00000 | 0.00000 |
| review_scores_rating | 0.00030 | 0.31270 | 0.03770 | 0.00000 | 0.00010 | 0.02760 |
| review_scores_accuracy | 0.91340 | 0.81970 | 0.77520 | 0.91420 | 0.55050 | 0.67030 |
| review_scores_cleanliness | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00070 |
| review_scores_checkin | 0.00230 | 0.00280 | 0.06860 | 0.00030 | 0.00010 | 0.00370 |
| review_scores_communication | 0.38920 | 0.37050 | 0.24560 | 0.73340 | 0.13030 | 0.15260 |

| | | | | | | |
|--------------------------------------|---------|---------|---------|---------|---------|---------|
| review_scores_location | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| review_scores_value | 0.00000 | 0.07340 | 0.02720 | 0.00000 | 0.00000 | 0.00000 |
| host_verifications | 0.00000 | 0.02950 | 0.00010 | 0.00010 | 0.00030 | 0.00490 |
| host_has_profile_pic | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| calculated_host_listings_count | 0.00000 | 0.00010 | 0.17770 | 0.00830 | 0.00000 | 0.00000 |
| host_acceptance_rate | 0.03070 | 0.00040 | 0.00710 | 0.00180 | 0.07120 | 0.00000 |
| host_identity_verified | 0.00530 | 0.79310 | 0.23320 | 0.09080 | 0.03440 | 0.09430 |
| host_response_time_hour | 0.00000 | 0.20840 | 0.45020 | 0.05740 | 0.00000 | 0.00000 |
| host_is_superhost | 0.09050 | 0.00000 | 0.00390 | 0.03620 | 0.30230 | 0.65860 |
| host_about | 0.03180 | 0.25720 | 0.16170 | 0.09260 | 0.00770 | 0.04160 |
| min_dist_to_white_house_ft | 0.00470 | 0.00380 | 0.00550 | 0.07050 | 0.14320 | 0.02540 |
| min_dist_to_washington_memorial_ft | 0.77580 | 0.75100 | 0.31230 | 0.05620 | 0.43860 | 0.67310 |
| min_dist_to_national_space_museum_ft | 0.08190 | 0.09990 | 0.01030 | 0.00050 | 0.05170 | 0.39620 |
| min_dist_to_lincoln_memorial_ft | 0.97950 | 0.56390 | 0.25250 | 0.10800 | 0.90150 | 0.29020 |
| min_dist_to_national_zoo_ft | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| min_dist_to_points_interest_ft | 0.00000 | 0.00600 | 0.00000 | 0.00000 | 0.00030 | 0.00320 |
| min_dist_to_metro_bus_stop_ft | 0.00200 | 0.21650 | 0.35710 | 0.35230 | 0.00220 | 0.01310 |
| min_dist_to_recreation_ft | 0.02710 | 0.80450 | 0.63150 | 0.01020 | 0.00940 | 0.07420 |
| min_dist_to_bank_ft | 0.00000 | 0.00000 | 0.00070 | 0.00000 | 0.00000 | 0.00000 |
| min_dist_to_crime_incident_ft | 0.00000 | 0.02860 | 0.00000 | 0.00000 | 0.00060 | 0.00460 |

Tabla 7. P-valores obtenidos de cada modelo de regresión.

Fuente: Elaboración propia.

| has_Breakfast | Cantidad | Precio promedio (USD) |
|---------------------|----------|-----------------------|
| 0 | 6051 | 179.974054 |
| 1 | 419 | 168.052506 |
| has_Washer | Cantidad | Precio promedio (USD) |
| 0 | 3003 | 166.318348 |
| 1 | 3467 | 190.361408 |
| has_Dishwasher | Cantidad | Precio promedio (USD) |
| 0 | 3245 | 163.036055 |
| 1 | 3225 | 195.468217 |
| has_Dryer | Cantidad | Precio promedio (USD) |
| 0 | 3257 | 170.868284 |
| 1 | 3213 | 187.64986 |
| has_TV | Cantidad | Precio promedio (USD) |
| 0 | 3325 | 172.832782 |
| 1 | 3145 | 185.935771 |
| has_Freezer | Cantidad | Precio promedio (USD) |
| 0 | 4562 | 179.737396 |
| 1 | 1908 | 177.921908 |
| has_Airconditioning | Cantidad | Precio promedio (USD) |
| 0 | 1884 | 176.057856 |
| 1 | 4586 | 180.493676 |

| | | |
|--------------------------------------|----------|-----------------------|
| has_Heating | Cantidad | Precio promedio (USD) |
| 0 | 1401 | 183.599572 |
| 1 | 5069 | 177.986585 |
| has_Hotwater | Cantidad | Precio promedio (USD) |
| 0 | 1566 | 199.400383 |
| 1 | 4904 | 172.752039 |
| has_Cleaningproducts | Cantidad | Precio promedio (USD) |
| 0 | 4635 | 179.544121 |
| 1 | 1835 | 178.337875 |
| has_Wifi | Cantidad | Precio promedio (USD) |
| 0 | 575 | 166.937391 |
| 1 | 5895 | 180.398304 |
| has_Hairdryer | Cantidad | Precio promedio (USD) |
| 0 | 1399 | 148.616869 |
| 1 | 5071 | 187.639913 |
| has_Kitchen | Cantidad | Precio promedio (USD) |
| 0 | 723 | 190.694329 |
| 1 | 5747 | 177.756221 |
| has_Dedicatedworkspace | Cantidad | Precio promedio (USD) |
| 0 | 3387 | 184.423679 |
| 1 | 3083 | 173.465456 |
| has_Petsallowed | Cantidad | Precio promedio (USD) |
| 0 | 6461 | 178.984058 |
| 1 | 9 | 335.666667 |
| has_Pool | Cantidad | Precio promedio (USD) |
| 0 | 6135 | 179.341809 |
| 1 | 335 | 176.641791 |
| has_Freestreetparking | Cantidad | Precio promedio (USD) |
| 0 | 3648 | 191.774671 |
| 1 | 2822 | 162.949327 |
| has_Freeparkingonpremises | Cantidad | Precio promedio (USD) |
| 0 | 4341 | 173.896798 |
| 1 | 2129 | 190.019258 |
| has_Smokealarm | Cantidad | Precio promedio (USD) |
| 0 | 207 | 269.362319 |
| 1 | 6263 | 176.222098 |
| has_Securitycamerasonproperty | Cantidad | Precio promedio (USD) |
| 0 | 4588 | 179.982781 |
| 1 | 1882 | 177.298618 |
| Has_Gym | Cantidad | Precio promedio (USD) |
| 0 | 5633 | 172.364459 |
| 1 | 837 | 225.218683 |



Universidad
Internacional
de Valencia