

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

2 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 | 7 |
| 2. Objetivos..... | 10 |
| 2.1. Generales | 10 |
| 2.2. Específicos | 10 |
| 3. Estado del Arte y Marco teórico | 11 |
| 4. Desarrollo del proyecto y resultados | 16 |
| 4.1. Ciudad de caso de estudio | 16 |
| 4.2. Fuentes de datos | 16 |
| 4.3. Limpieza y transformación de los datos | 18 |
| 4.4. Descripción de los datos..... | 21 |
| 4.5. Selección de características | 38 |
| 4.6. Modelos de regresión | 41 |
| 4.7. Resultados | 42 |
| 5. Conclusión y trabajos futuros..... | 52 |
| 6. Referencias | 54 |
| Anexos I..... | 57 |

Índice de ilustraciones

| | |
|---|----|
| Ilustración 1. Distribución de precios de los alojamientos. | 19 |
| Ilustración 2. Proporción de alojamientos por tipo de habitación. | 24 |
| Ilustración 3. Tipo de habitación vs precio. | 24 |
| Ilustración 4. Número de personas permitidas en el alojamiento vs precio | 27 |
| Ilustración 5. Correlación de las características del alojamiento. | 28 |
| Ilustración 6. Correlación de las características de servicios adicionales. | 29 |
| Ilustración 7. Calificaciones por tipo de habitación. | 30 |
| Ilustración 8. Correlación de las características de calificaciones. | 31 |
| Ilustración 9. Proporción de superhost. | 32 |
| Ilustración 10. Identidad y si el host tiene imagen de perfil. | 32 |
| Ilustración 11. Correlación de variables del host. | 33 |
| Ilustración 12. Cantidad de alojamientos de Airbnb por barrio. | 34 |
| Ilustración 13. Promedio de precios de alojamientos de Airbnb por barrio. | 35 |
| Ilustración 14. Correlación de variables de ubicación. | 37 |
| Ilustración 15. Puntaje promedio de alojamientos de Airbnb por barrio en la ciudad. ... | 57 |
| Ilustración 16. Promedio de disponibilidad por barrio en la ciudad. | 58 |
| Ilustración 17. Promedio de noches mínimas para hacer reserva por barrio. | 59 |

Índice de tablas

| | |
|--|----|
| Tabla 1. Estadística descriptiva de las principales características..... | 22 |
| Tabla 2. Cantidad de alojamientos por tipo de propiedad. | 25 |
| Tabla 3. Reserva inmediata en los alojamientos. | 26 |
| Tabla 4. Resumen de estadísticas por barrios. | 36 |
| Tabla 5. Determinantes del precio por noche para OLS y QR..... | 43 |
| Tabla 6. P-valores obtenidos de cada modelo de regresión..... | 44 |
| Tabla 7. Porcentajes de incremento / disminución en el precio..... | 46 |

Resumen

Airbnb ha contribuido a la evolución de la industria del alojamiento turístico, ofreciendo una amplia variedad de opciones de alojamiento que anteriormente no se ofrecían, esto ha aumentado la competencia en el mercado y ha mejorado la calidad y la diversidad del alojamiento turístico en general, por lo tanto, definir un buen precio para rentar el alojamiento se ha convertido en un tema de interés.

El objetivo principal del presente trabajo fue analizar las características que influyen en los precios de los alojamientos turísticos de Airbnb en la ciudad de Washington DC. Los datos fueron obtenidos del sitio público de Inside Airbnb, en donde se extrajeron 6473 observaciones y 76 características, estos datos fueron limpiados y transformados de modo que se encuentren listos para que los modelos de regresión arrojen resultados óptimos, luego se realizó un análisis estadístico de todas las características encontradas en la literatura para poder hacer un resumen de estas, 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 utilizó 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 emplearon librerías para ciencia de datos para obtener, limpiar, transformar, crear modelos, entre otras funciones.

Los resultados del análisis arrojaron que los alojamientos que son casas o apartamentos tienen mayor valor, el número de personas permitidas, las habitaciones, la cantidad de baños, la disponibilidad también influye significativamente en el aumento del precio, se identificó principalmente que los precios aumentan hasta en un 38% dependiendo de la cantidad de baños que tenga el alojamiento y hasta un 61% dependiendo de la cantidad de habitaciones que posea el alojamiento, el número de personas que permite el alojamiento también influye en los precios y el incremento puede ser de hasta un 18%. Las calificaciones de los usuarios que tienen que ver con limpieza y localización se encontraron que influyen positivamente en el precio, representando un aumento desde un 9% hasta un 13% respectivamente. Se comprobó que el precio incrementa hasta en un 15% en alojamientos que estén cercanos a lugares como el Zoológico Nacional y en general a algún lugar de interés en la ciudad hasta un 19%. Por otro lado, se determinó que los precios de los alojamientos disminuyen a medida que el número de camas aumenta y en cierta medida, los precios disminuyen cuando los alojamientos tienen piscina y cocina.

Los resultados indican que la ubicación es un factor clave en la determinación de los precios de los alojamientos de Airbnb, aquellos ubicados en las zonas más turísticas y céntricas de la ciudad suelen tener precios más altos que los que se encuentran en las zonas más alejadas. También, los apartamentos o casas completos suelen tener un precio mayor que las habitaciones privadas y compartidas, además, las propiedades de lujo suelen tener precios muchísimo más altos. Esto comprueba lo visto en la teoría



y confirma que en la ciudad de Washington DC los alojamientos que cumplen con las condiciones anteriormente expuestas pueden tener un mayor precio y esto podría servir de guía para que aquellos hosts o propietarios tengan una referencia de los precios que podría tener el alojamiento.

Este trabajo se hizo en un notebook en Jupyter el cual está ubicado en un repositorio de github público (Carvajal, 2023).

Palabras clave: Airbnb, Características del barrio, Washington D.C, Precio de los alojamientos, Regresión por mínimos cuadrados, Regresión Cuantílica.

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, ya que permite a los usuarios ofrecer y reservar propiedades en diferentes ciudades del mundo, esta plataforma 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 y debido a esto 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 Airbnb, los hosts (o anfitriones) son los que le colocan el precio por noche al alojamiento, por lo tanto, estos precios no son fijos ni son decididos por Airbnb, cada persona puede colocar el precio que considere que es el correcto. Un artículo hecho por Airbnb (*Cómo Fijar El Precio de Tu Alojamiento - Centro de Ayuda de Airbnb*, n.d.) en su centro de ayuda muestra a los cómo deben fijar los precios de sus alojamientos, en él se muestra que lo ideal es guiarse de los precios de otros alojamientos que se encuentren en la misma zona y muchos anfitriones han tomado esto como un estándar o una guía a la hora de fijar sus precios, pero esto no siempre es óptimo ya que hay alojamientos que tienen diferentes características y esto se ha convertido en un reto, de ahí que, el análisis de precios en el mercado de alojamientos es de gran importancia en esta industria turística.

Muchos estudios se han enfocado en estudiar las características que determinan los precios. La investigación de (Gómez et al., 2020) consistió en analizar los alojamientos de la Ciudad de México para poder determinar el precio, la principal técnica usada para lograr esto fue la implementación de métodos estadísticos y técnicas de aprendizaje automático como regresión cuantílica y regresión logística. Se usaron características del host como el tiempo de respuesta, el número de alojamientos que posee el host, del alojamiento se incluyeron variables como el número de personas permitidas en el alojamiento, el número de baños, camas, habitaciones, también características que tienen que ver con las calificaciones como el número de calificaciones, entre otras. Se concluyó que el tipo de habitación, el barrio, el mínimo y máximo número de noches son las variables más importantes que influyen en el precio.

Otro estudio hecho por (Wang & Nicolau, 2017) en donde se aplicó regresión OLS y regresión cuantílica, usando atributos del host, de ubicación, amenidades y servicios, reglas de renta y calificaciones. En donde los resultados mostraron que la característica que indica si un host es superhost hace que los precios aumenten, igual con el número de alojamientos que posee el host. En general, se comprueba que las

técnicas de regresión OLS y QR son útiles para determinar los precios de los alojamientos y que las características más importantes son las relacionadas con la ubicación, atributos del alojamiento, amenidades, servicios, reglas de renta y comentarios y calificaciones de los usuarios.

De modo similar, en el estudio realizado por (Perez-Sanchez et al., 2018) en la ciudad de Valencia, Alicante, Elche y Castellón de la Plana, se halló que la ubicación del alojamiento con respecto a sitios que son turísticos aumenta significativamente el precio y que los precios se van reduciendo gradualmente a medida que aumenta la distancia desde la costa, así mismo, se usaron atributos del alojamiento, anuncios, características ambientales y de ubicación. En este estudio se usó regresión cuantílica y regresión por mínimos cuadrados y los resultados se compararon entre modelos.

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 DC, capital de Estados Unidos, y por ser la capital y centro político de este país, es un atractivo turístico para muchos visitantes. Para este análisis los datos principales de los alojamientos provienen de la página Inside Airbnb, la cual contiene información pública sobre alojamientos de Airbnb de varias ciudades en el mundo, en este caso se elegirá la ciudad anteriormente mencionada, y en ese mismo sitio web se obtendrán los datos geográficos de los barrios de la ciudad de estudio. Adicionalmente se tendrán otros dataset que se obtendrán de sitios web públicos del gobierno de esta ciudad y que contienen información geográfica de algunos puntos de interés, incidentes de crímenes, sitios de recreación y localización de bancos en la ciudad. Una vez se obtengan esos datos se pretenderá hacer un proceso de limpieza y transformación para luego analizar estas variables y con la ayuda de mapas de calor determinar la correlación entre las características con el fin de elegir aquellas que proporcionen información valiosa y no redundante para los dos modelos de regresión que se aplicarán y así reducir su complejidad y mejorar su efectividad. La correlación se mide en una escala entre -1 y 1, siendo -1 una correlación negativamente perfecta y 1 una correlación positivamente perfecta, entre más se acerque el valor a 0, menor será la relación entre las variables, si dos características están altamente correlacionadas, es posible que una de ellas no sea necesaria. Principalmente se propone el uso de dos modelos de regresión que, como se mencionó anteriormente en otros estudios, son bastante usados en este tipo de mercado: regresión por mínimos cuadrados (OLS) y regresión cuantílica (QR). Con la primera se estimará la relación lineal entre las variables independientes y el precio medio esperado y la segunda permitirá estimar la relación entre las variables independientes y diferentes niveles del precio (cuantiles). Estos modelos ayudarán a identificar los factores y características que tienen mayor impacto en la determinación de los precios, usando como variable dependiente el precio por noche en dólares y en las variables independientes estarán varias características que fueron incluidas dentro de cinco categorías: características del alojamiento, servicios adicionales, calificaciones del huésped, atributos del host y de ubicación. Los resultados de estos dos modelos se compararán para obtener una visión más completa sobre la relación entre el precio y las características elegidas.

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.

Este trabajo de fin de máster tiene como finalidad contribuir al conocimiento sobre el mercado de alojamientos de Airbnb en Washington DC, esto con ayuda de la literatura existente y estudios anteriormente realizados, 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 y ayudar a comprender mejor el comportamiento de este tipo de mercado, adicionalmente, aportar evidencia empírica sobre los factores que influyen en el precio de los alojamientos de Airbnb en una ciudad tan importante como Washington DC y comparar los resultados obtenidos con la regresión por mínimos cuadrados (OLS) y regresión cuantílica. De igual manera, se espera que este estudio pueda servir como base para futuras investigaciones que amplíen el análisis a otras ciudades o que incorporen otras variables o métodos. También se pretende probar la hipótesis de que el precio de los alojamientos de Airbnb en Washington DC depende positivamente de la cantidad de habitaciones, calificaciones de los usuarios, la cercanía a algunos puntos de interés en la ciudad, de comodidades y servicios adicionales de los alojamientos. De igual manera, revisar si el precio de los alojamientos de Airbnb en Washington D.C. depende negativamente de la distancia al centro de la ciudad y a los principales puntos de interés turístico. Para lograr esto, se realizará un análisis estadístico de las características obtenidas en los datos, las cuales se analizarán y se elegirán aquellas que se mencionan en algunos estudios, se hará una limpieza y transformación de algunas variables y se elegirán aquellas que sean consideradas relevantes y que la correlación no sea demasiado alta. Estos resultados pueden ayudar a los hosts o anfitriones a establecer el precio óptimo para su alojamiento.

Este documento se organiza de la siguiente forma: En la siguiente sección se enumeran los objetivos generales y específicos de este trabajo. Seguidamente, se menciona el estado del arte y marco teórico en donde se explica de manera general los conceptos que se estarán usando a lo largo de todo el trabajo como la regresión por mínimos cuadrados o OLS y regresión cuantílica o QR, también se mencionan con más detalle algunos estudios realizados anteriormente. Después se muestra el desarrollo del proyecto y los resultados, en donde se detalla la ciudad de caso de estudio, las fuentes de datos, el proceso de limpieza y transformación de los datos, descripción de los datos, modelos de regresión empleados (OLS y QR) y por último los resultados. Por último, se encuentran las conclusiones basadas en los resultados obtenidos.

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.
- Identificar patrones o relación entre las características y los precios de los alojamientos de Airbnb.
- Identificar la correlación entre el precio del alojamiento y las características para no tener información redundante.
- Usar métodos de regresión para analizar el impacto en el precio de las variables independientes.
- 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.
- Encontrar las principales características que influyan en los precios de los alojamientos de Airbnb.
- Contribuir a futuros análisis de características que influyan en los precios de los alojamientos.

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.

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). Específicamente en Airbnb, en donde los precios de los alojamientos dependen de varios factores como, por ejemplo, la ubicación, el tamaño, el número de habitaciones, la disponibilidad de servicios, etc., estos modelos cobran mucha importancia para determinar qué características de un alojamiento influyen en su precio de alquiler. En la revisión bibliográfica de (Sandoval, 2008) se expone que el uso de los modelos de

precios hedónicos se remonta a la década de 1970, cuando el economista Sherwin Rosen publicó un artículo sobre el tema en donde propuso que el precio de un bien o servicio depende de sus características y que estos modelos ofrecen ventajas en comparación con otros enfoques parecidos ya que estos explican y predicen muy bien el precio de las viviendas. Hasta nuestros días, esta idea se ha utilizado mucho para estimar el valor de los bienes inmuebles, como los alquileres de Airbnb.

Hay varios estudios que han utilizado modelos de precios hedónicos para analizar el mercado de Airbnb. Por ejemplo, en un estudio realizado por (Gibbs et al., 2018) donde se aplicó un modelo de precios hedónicos a 15.716 alojamientos de Airbnb arrojó como resultado que las características físicas como el tipo de habitación, el número de personas aceptadas por noche, amenidades como piscina, gimnasio, otras características como la ubicación y las características de los hosts tienen un impacto significativo en el precio. Otro estudio hecho por (Cerimelo & Porto, 2022) usando datos de Booking y Airbnb mostró que los precios de Airbnb se establecen principalmente por aquellas características relacionadas con la ubicación, el tamaño y la calidad del alojamiento. En el estudio presentado por (Poeta et al., 2019) se examinaron los modelos hedónicos pero enfocados en la valoración de la vivienda, en él se concluyó que estos modelos son útiles para estimar los precios de los bienes inmobiliarios basándose en los atributos como el área de la vivienda, el número de habitaciones, el barrio, entre otros.

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,. Numero 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, 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.

Existen dos ramas bastante conocidas de aprendizaje automático: aprendizaje supervisado y no supervisado. El aprendizaje no supervisado es una técnica que se

enfatisa 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 y otros algoritmos como la regresión por mínimos cuadrados y la regresión cuantílica (Bonaccorso, n.d.).

La regresión por mínimos cuadrados (OLS) es usada para analizar la relación entre una variable dependiente y una o varias independientes, este método supone que la media de una variable dependiente depende del valor que tomen una o varias variables independientes, el objetivo de esta regresión es encontrar la línea que mejor se ajuste a los datos, minimizando la suma de los errores cuadrados entre la línea y los puntos, es decir, la relación debe ser lineal. Cabe recalcar que esta regresión supone que los errores de la regresión tienen una distribución normal y que la varianza es constante, además de la linealidad de la relación entre las variables, y es importante que los datos cumplan con estos supuestos para que el algoritmo arroje resultados más precisos. La ecuación de la regresión OLS se escribe como: $Y = \beta_0 + \beta_1 X + \epsilon$, donde Y es la variable dependiente (precio), X es la variable independiente, β_0 es la intersección de la línea de regresión, β_1 es la pendiente de la línea de regresión y ϵ es el término de error. Para estimar los valores de β_0 y β_1 , se minimiza la suma de los errores cuadráticos entre los valores observados y los valores predichos, esto se logra encontrando los valores de β_0 y β_1 que hacen que la suma de los errores cuadráticos sea la más pequeña posible. Diversos estudios han aplicado regresión OLS para obtener los principales factores que afectan los precios de los alojamientos de Airbnb, (Samwel, 2022) usó este algoritmo y obtuvo como resultado que las principales variables que determinan los precios en la ciudad de estudio que fue New York son el número de camas, el número de reviews, los tipos de habitación, tipo de propiedad y el barrio.

La regresión cuantílica (QR) 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, es robusta frente a valores extremos en las mediciones de respuesta, 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. Explica la

relación entre los cuantiles de la variable de resultado, por otro lado, la regresión lineal explica es la relación lineal con respecto a la media de las variables. En el caso de Airbnb, la regresión cuantílica es una técnica que aporta información valiosa a los hosts o anfitriones, ya que los ayuda a determinar los precios de sus propiedades de una forma más exacta dependiendo de las características de estas. En el estudio realizado por (Wang & Nicolau, 2017) se utiliza la regresión cuantílica para determinar la influencia de ciertas características y factores que influyen en el precio de los alojamientos de Airbnb, en él se concluye que la regresión cuantílica es una herramienta útil para analizar los precios de Airbnb debido a su capacidad para capturar la heterogeneidad de los efectos de los factores en diferentes cuantiles de los precios y se verifica cómo diferentes características como los atributos del host, la localización, de la propiedad como tal, amenidades y servicios, calificaciones de los usuarios y reglas de renta, afectan los precios de Airbnb en diferentes cuantiles.

La regresión cuantílica tiene ciertas ventajas con respecto a otros tipos de regresión como OLS ya que la primera proporciona una estimación más precisa y puede manejar mejor la heterogeneidad en los datos de precios de Airbnb, algo que es muy común en este tipo de mercado ya que los precios de los alojamientos pueden variar dependiendo de varios factores como la ubicación, la temporada, el número de huéspedes, etc., y puede ayudar a los anfitriones a entender cómo ajustar sus precios en estas situaciones. 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. Cabe mencionar que en la mayoría de los estudios encontrados se usa OLS junto con algún otro método de regresión para poder ser comparados los resultados.

Existen varias herramientas y lenguajes de programación como Python, el cual es potente cuando se usa para ciencia de datos, en donde existen una gran cantidad de librerías y paquetes que facilitan el manejo y procesamiento de los datos permitiendo realizar análisis estadísticos y facilitando la ejecución de algoritmos de Machine Learning. Entre las librerías más usadas están NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, entre otros.

NumPy facilita la realización de cálculos numéricos y operaciones de matriz. También sirve para hacer operaciones matemáticas en datos, como calcular datos de estadística descriptiva y realizar operaciones de álgebra lineal.

Pandas proporciona herramientas para la manipulación y análisis de datos. Se usa para trabajar con los datos en formato de tabla y ejecutar tareas como la limpieza de datos, selección de subconjuntos de datos y agregación de datos.



Matplotlib y Seaborn son librerías de visualización de datos de alta calidad. Permiten crear una variada cantidad de gráficos como gráficos de barras, dispersión, mapas de calor, etc.

Scikit-learn es una librería de aprendizaje automático que brinda herramientas para la creación de modelos predictivos. Incluye una amplia variedad de algoritmos de aprendizaje automático, como regresión lineal, regresión logística, clasificación de árboles de decisión, entre otras.

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.

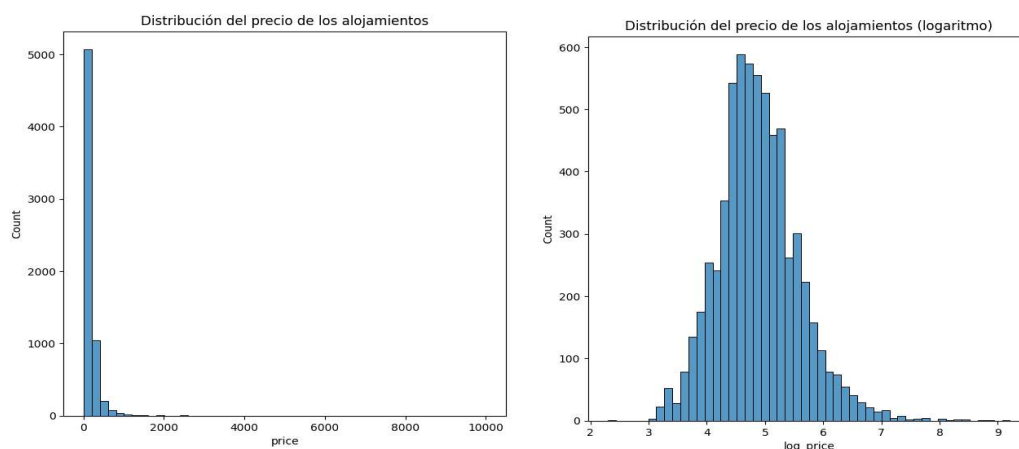


Ilustración 1. Distribución de precios de los alojamientos.

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.

Tabla 1. 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 |

Fuente: Elaboración propia.

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.

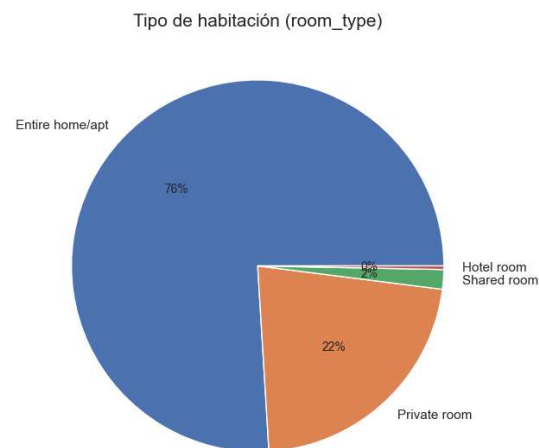


Ilustración 2. Proporción de alojamientos por tipo de habitación.

Fuente: Elaboración propia

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.**

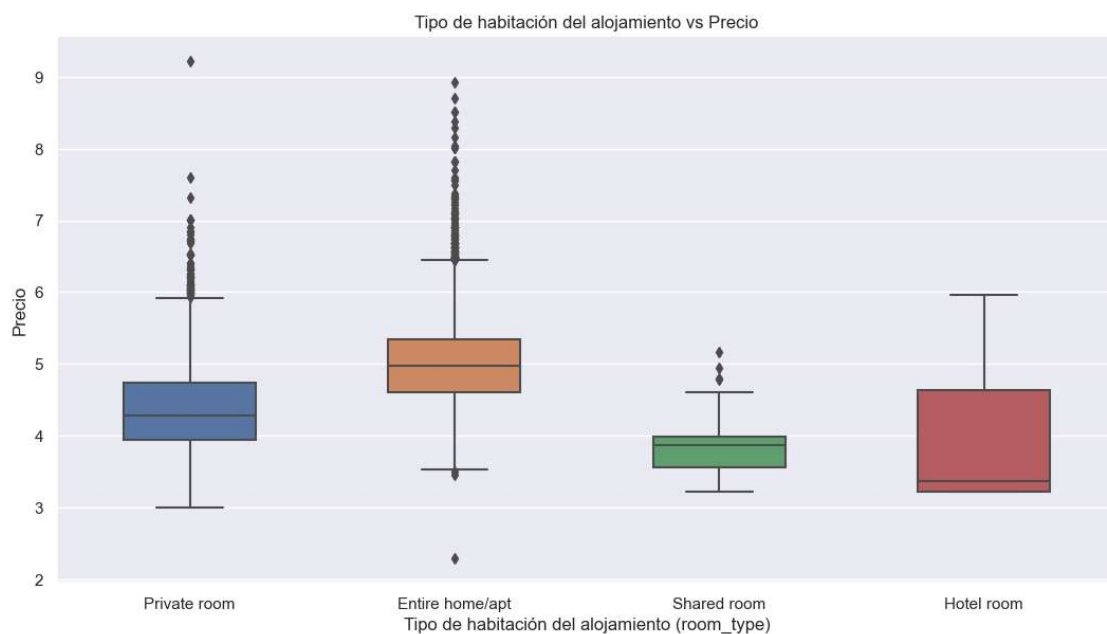


Ilustración 3. Tipo de habitación vs precio.

Fuente: Elaboración propia.

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”.

Tabla 2. Cantidad de alojamientos por tipo 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 |

Fuente: Elaboración propia.

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**.

Tabla 3. 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 |

Fuente: Elaboración propia.

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.

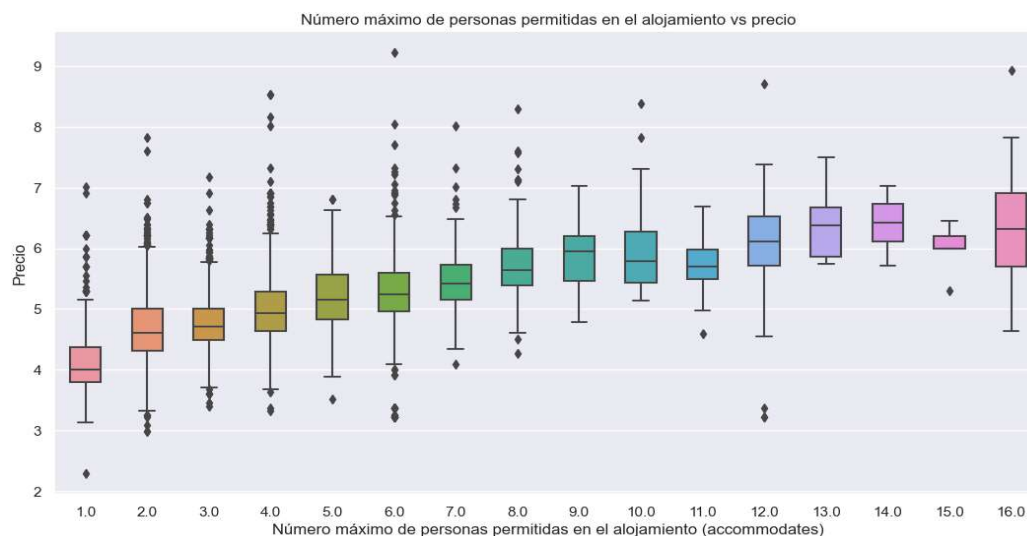


Ilustración 4. Número de personas permitidas en el alojamiento vs precio

Fuente: Elaboración propia.

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.

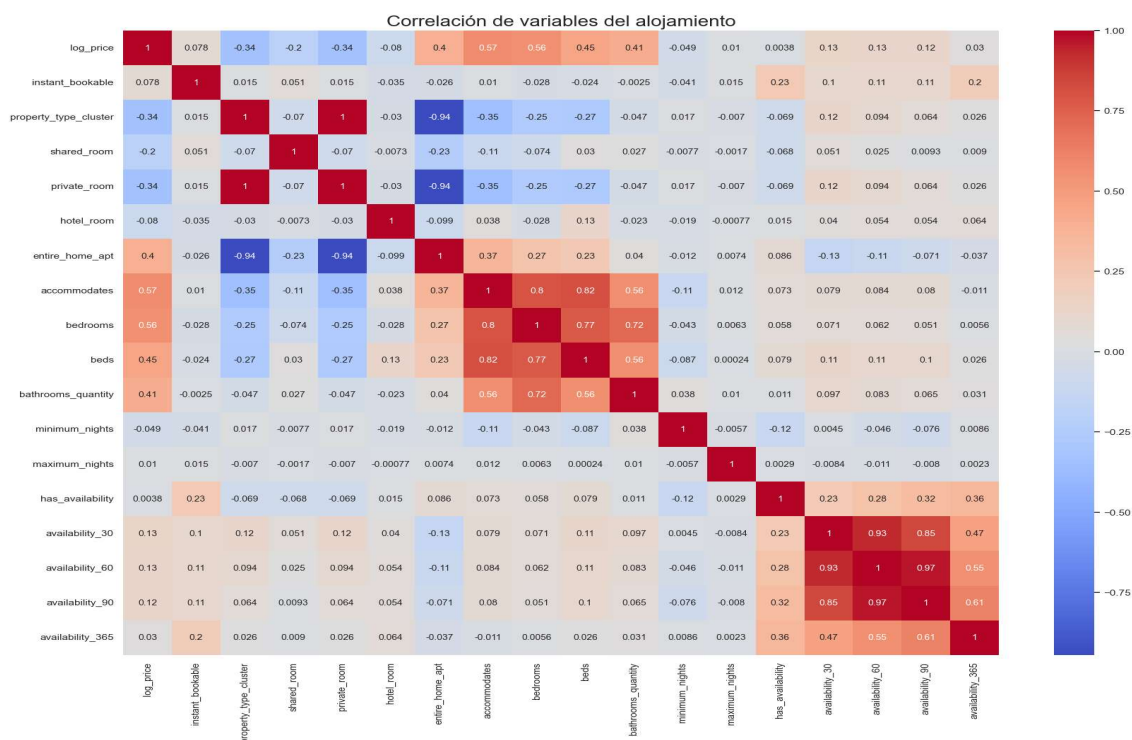


Ilustración 5. Correlación de las características del alojamiento.

Fuente: Elaboración propia.

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.

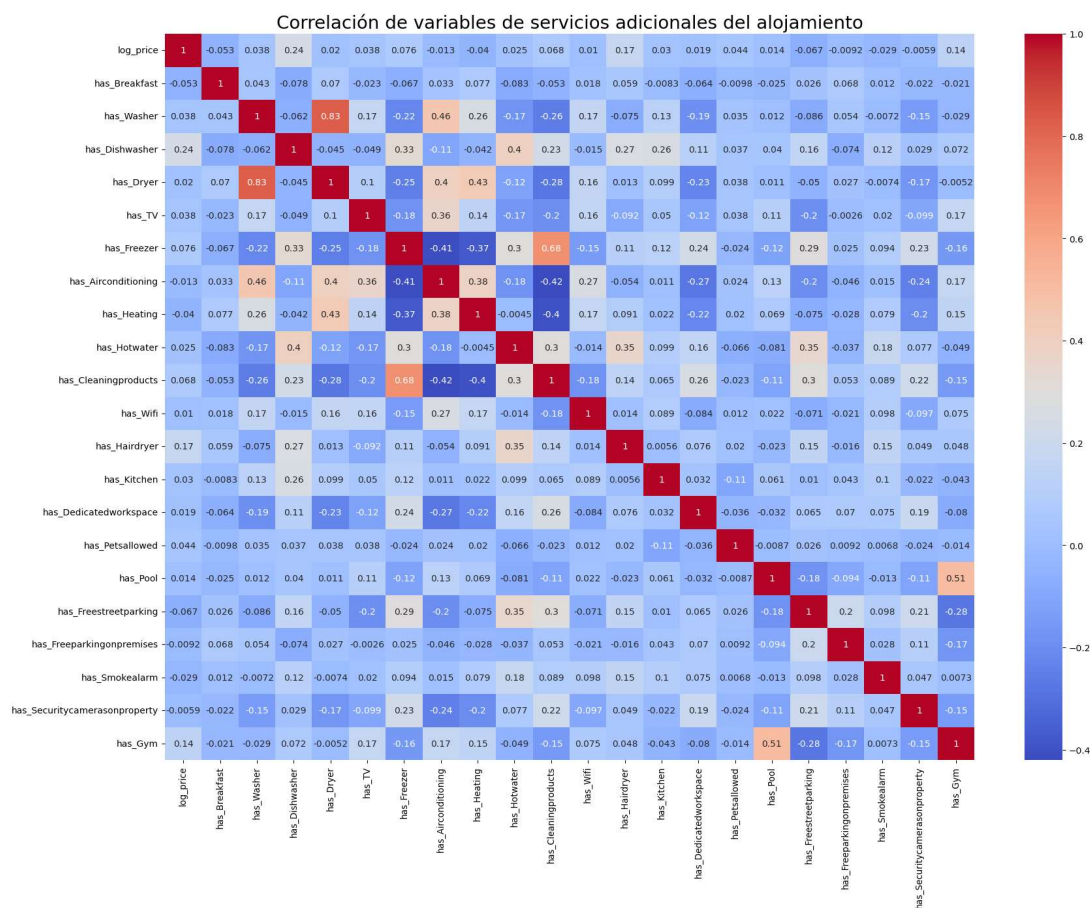


Ilustración 6. Correlación de las características de servicios adicionales.

Fuente: Elaboración propia.

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**).

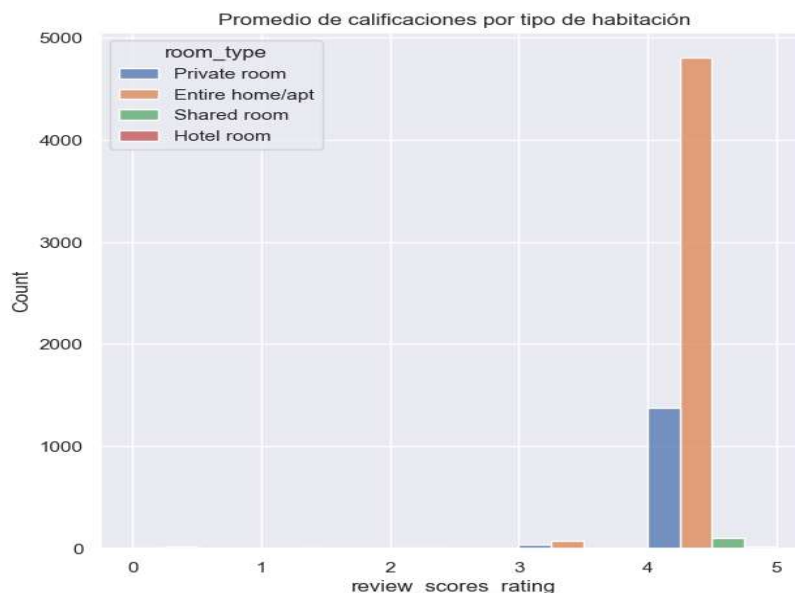


Ilustración 7. Calificaciones por tipo de habitación.

Fuente: Elaboración propia.

Al revisar la correlación entre las variables se observa que 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.

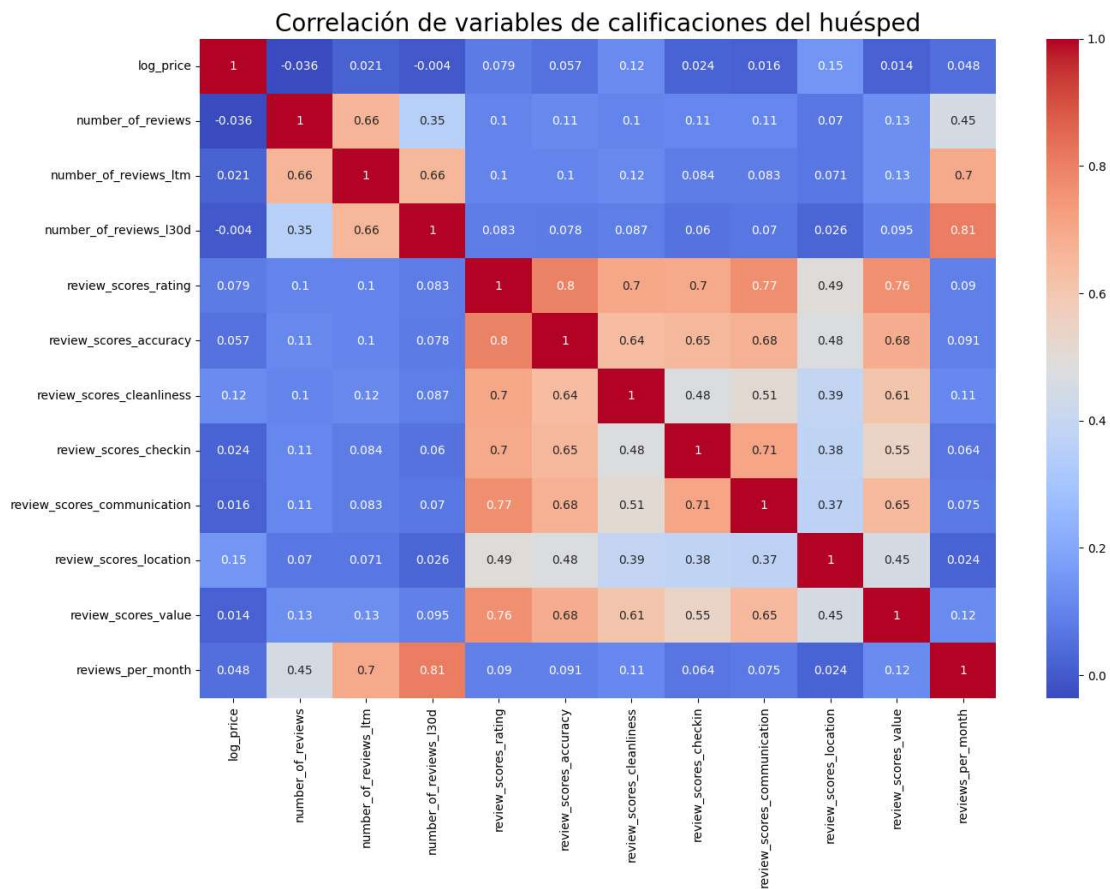


Ilustración 8. Correlación de las características de calificaciones.

Fuente: Elaboración propia

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**.

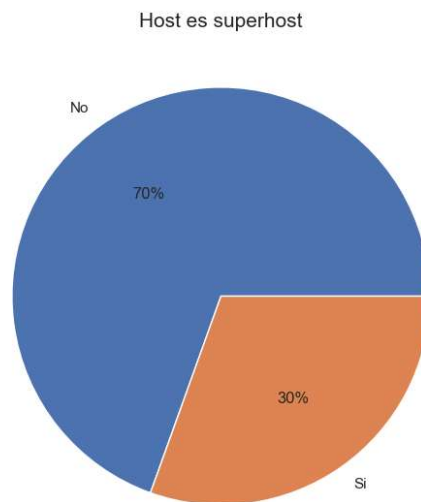


Ilustración 9. Proporción de superhost.

Fuente: Elaboración propia.

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**.

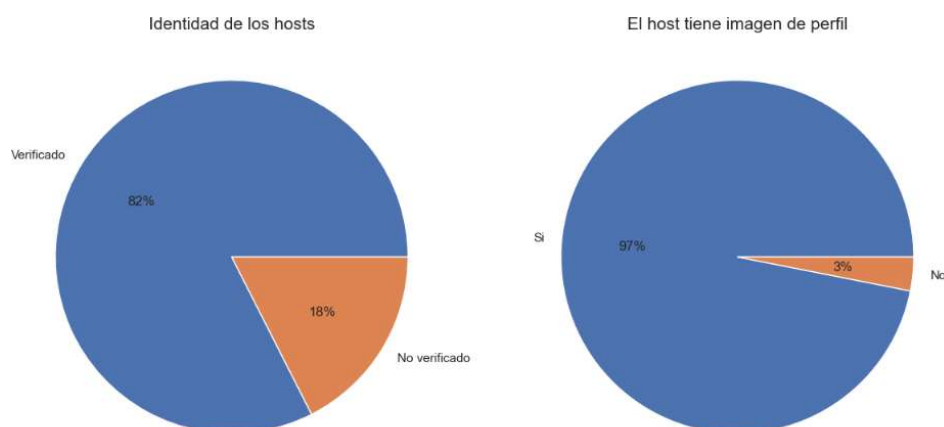


Ilustración 10. Identidad y si el host tiene imagen de perfil.

Fuente: Elaboración propia.

Se puede percibir 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 11**, 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.

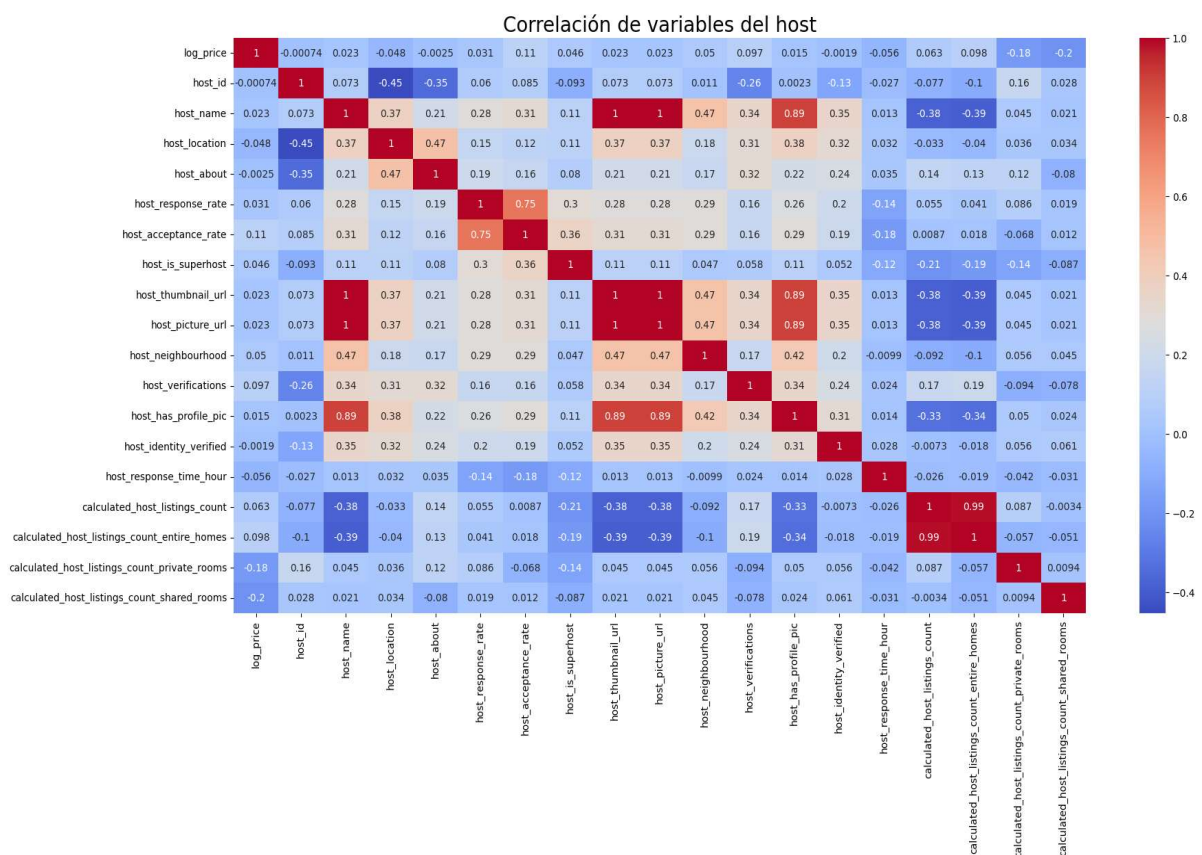


Ilustración 11. Correlación de variables del host.

Fuente: Elaboración propia.

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.

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

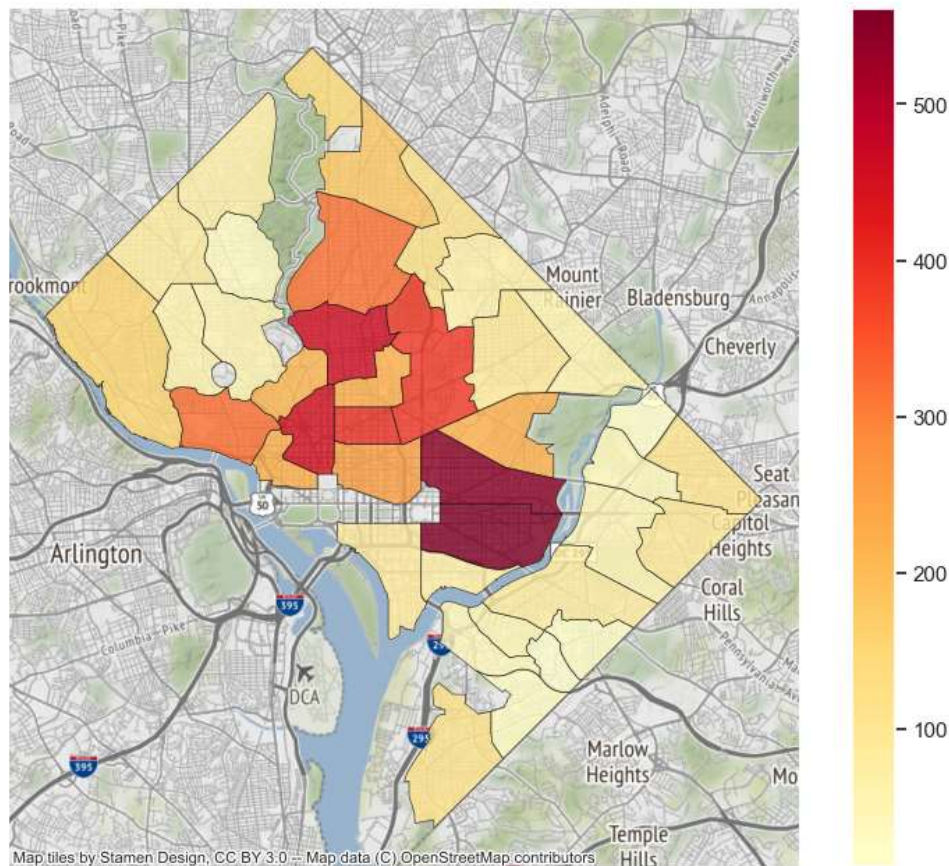


Ilustración 12. Cantidad de alojamientos de Airbnb por barrio.

Fuente: Elaboración propia

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 133** 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.

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

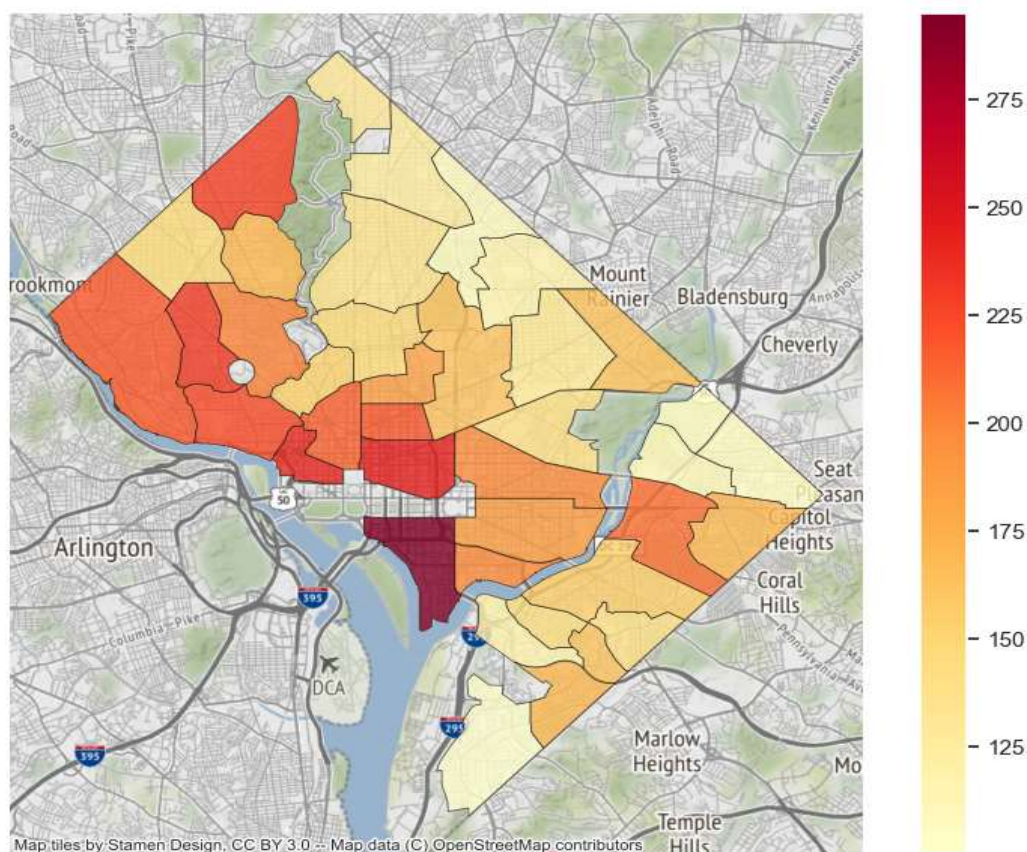


Ilustración 133. Promedio de precios de alojamientos de Airbnb por barrio.

Fuente: Elaboración propia

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.

Tabla 4. Resumen de estadísticas por barrios.

| Barrio | Precio promedio | Cantidad de alojamientos | Calificación promedio | Noches mínimas promedio | Disponibilidad promedio |
|---|-----------------|--------------------------|-----------------------|-------------------------|-------------------------|
| Southwest Employment Area, Southwest/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, Massachusetts 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, Mahaning 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 |

Fuente: Elaboración propia.

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 144**).

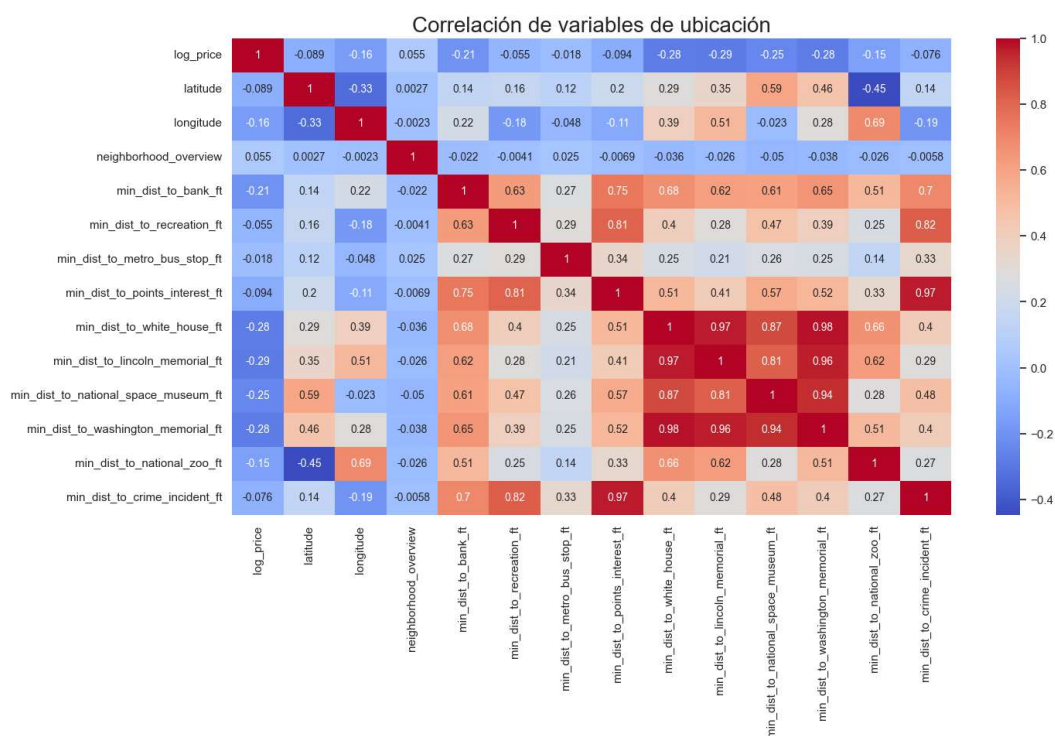


Ilustración 144. Correlación de variables de ubicación.

Fuente: Elaboración propia.

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_apartment`, por esta razón, la última fue elegida para el modelo.

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_apartment` 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.

Sobre las variables de ubicación se identificó una alta correlación entre varias de ellas, `min_dist_to_recreation_ft` estaba muy correlacionada con la variable que indica los puntos de interés, por lo tanto, la primera fue descartada.

`min_dist_to_crime_incident_ft` y `min_dist_to_points_interest_ft` estaban correlacionada al 0.97, esto indica que la mayoría de los incidentes de crímenes por lo general ocurren en sitios cercanos a algún punto de interés turístico en la ciudad, por lo tanto la primera fue descartada también ya que ambas variables están indicando casi la misma información.

`min_dist_to_lincoln_memorial_ft` y `min_dist_to_washington_memorial_ft` al tener una correlación alta con `min_dist_to_white_house_ft` se descartaron, ya que esto indica

que ya con tener la distancia mínima a la casa blanca se incluyen ya sitios como el Monumento de Lincoln y el Monumento de Washington, y esto tiene sentido ya que estos tres sitios quedan en ubicaciones muy cercanas, están casi dentro de la misma zona.

En general, se descartaron variables que estaban muy correlacionadas, a más de 0.8, algunas que tienen correlación son importantes y relevantes para el modelo, puede ser justificable mantenerlas en el modelo.

En las siguientes definiciones 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_aprt`: 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_Cleaningproducts: 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_Dedicatedworkspace: Tiene algún espacio dedicado para trabajar. (Tipo: Dummy).
- has_Petsallowed: 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_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_national_space_museum_ft`: Distancia mínima en pies del alojamiento al Museo Nacional del Espacio. (Tipo: Número decimal).
- `min_dist_to_national_zoo_ft`: Distancia mínima en pies del alojamiento al Zoológico Nacional. (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 **¡Error! No se encuentra el origen de la referencia.**, 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.

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

| 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) | -0.0679 *** (0.007) | -0.0563 *** (0.0064) | -0.065 *** (0.0059) | -0.0927 *** (0.0078) | -0.1254 *** (0.0109) |
| | beds | -0.0403 (0.0224) | -0.1225 *** (0.0265) | -0.0823 *** (0.0212) | -0.0026 (0.0187) | 0.0063 (0.0245) | 0.0103 (0.0367) |
| | availability_365 | 0.0158 *** (0.0025) | 0.0176 *** (0.0024) | 0.0163 *** (0.0022) | 0.0136 *** (0.0021) | 0.0112 *** (0.0028) | 0.0161 *** (0.004) |
| | instant_bookable | 0.0036 (0.0019) | 0.006 ** (0.0021) | 0.0095 *** (0.0018) | 0.0041 * (0.0016) | 0.0008 (0.002) | -0.0018 (0.0027) |
| | bathrooms_quantity | 0.1972 *** (0.0244) | 0.1603 *** (0.0289) | 0.2123 *** (0.0238) | 0.2113 *** (0.0204) | 0.189 *** (0.0239) | 0.2937 *** (0.0307) |
| | minimum_nights | -0.0665 ** (0.0246) | -0.6802 *** (0.0133) | -0.5366 *** (0.0155) | -0.2556 *** (0.0206) | -0.1294 *** (0.0372) | 0.0874 (0.056) |
| | bedrooms | 0.3861 *** (0.0266) | 0.4723 *** (0.0348) | 0.3941 *** (0.0267) | 0.407 *** (0.0222) | 0.4114 *** (0.0278) | 0.4027 *** (0.0402) |
| | accommodates | 0.1588 *** (0.0112) | 0.1089 *** (0.0126) | 0.1224 *** (0.0109) | 0.1149 *** (0.0093) | 0.1644 *** (0.0119) | 0.1685 *** (0.0169) |
| | entire_home_apartment | 0.0402 *** (0.0025) | 0.0569 *** (0.0025) | 0.0556 *** (0.0022) | 0.0523 *** (0.0021) | 0.0327 *** (0.0026) | 0.0225 *** (0.0035) |
| | hotel_room | -0.1075 *** (0.0157) | -0.0948 *** (0.0153) | -0.1298 *** (0.0141) | -0.1569 *** (0.0131) | -0.0503 ** (0.018) | 0.0451 (0.0246) |
| Características de servicios adicionales | has_Hotwater | -0.0129 *** (0.0025) | -0.0134 *** (0.0027) | -0.0121 *** (0.0023) | -0.0056 ** (0.0021) | -0.0115 *** (0.0026) | -0.0259 *** (0.0033) |
| | has_Petsallowed | 0.005 (0.0226) | -0.0279 (0.0238) | 0.0382 (0.0199) | 0.0273 (0.0188) | 0.0169 (0.0267) | 0.0079 (0.0339) |
| | has_Dedicatedworkspace | -0.0043 * (0.0019) | -0.0039 * (0.0019) | -0.0049 ** (0.0017) | -0.0016 (0.0016) | -0.0028 (0.002) | -0.0049 (0.0028) |
| | has_Kitchen | -0.0195 *** (0.003) | -0.0071 * (0.0033) | -0.0082 ** (0.0028) | -0.014 *** (0.0025) | -0.0245 *** (0.0031) | -0.0198 *** (0.004) |
| | has_Hairdryer | 0.0124 *** (0.0024) | 0.0124 *** (0.0025) | 0.0119 *** (0.0022) | 0.0087 *** (0.002) | 0.0106 *** (0.0026) | 0.0059 (0.0036) |
| | has_Wifi | 0.0046 (0.0031) | -0.0014 (0.0032) | 0.0008 (0.0028) | 0.0006 (0.0026) | 0.0041 (0.0033) | 0.0158 *** (0.0044) |
| | has_Cleaningproducts | 0.0023 (0.0027) | 0.0031 (0.0028) | 0.0041 (0.0025) | 0.0024 (0.0022) | 0.0005 (0.0028) | -0.0019 (0.0038) |
| | has_Washer | -0.0045 * (0.0021) | -0.0022 (0.0021) | -0.0013 (0.0019) | 0.0004 (0.0017) | -0.0034 (0.0022) | -0.0064 * (0.0031) |
| | has_Airconditioning | 0.0044 (0.0024) | -0.0015 (0.0025) | -0.0005 (0.0022) | 0.0007 (0.002) | 0.0081 ** (0.0026) | 0.0088 * (0.0035) |
| | has_Freezer | 0.0028 (0.0027) | 0.0021 (0.0027) | 0.0051 * (0.0024) | 0.003 (0.0022) | 0.0086 ** (0.0029) | 0.0101 ** (0.0038) |
| | has_TV | 0.0001 (0.0019) | 0.0049 * (0.002) | 0.0038 * (0.0017) | -0.001 (0.0016) | -0.0054 ** (0.002) | -0.0071 * (0.0028) |
| | has_Dishwasher | 0.0117 *** (0.0021) | 0.0159 *** (0.0022) | 0.0133 *** (0.002) | 0.013 *** (0.0018) | 0.0117 *** (0.0023) | 0.0086 ** (0.003) |
| | has_Freestreetparking | -0.0065 ** (0.0021) | 0.0046 * (0.0021) | 0.002 (0.0019) | -0.0031 (0.0017) | -0.0099 *** (0.0022) | -0.0137 *** (0.0028) |
| | has_Breakfast | 0.0076 * (0.0035) | -0.0082 * (0.0035) | -0.0004 (0.0032) | 0.0028 (0.0029) | 0.0159 *** (0.0039) | 0.0298 *** (0.0054) |
| | has_Pool | -0.0194 *** (0.0046) | -0.0287 *** (0.0041) | -0.0297 *** (0.0039) | -0.017 *** (0.0039) | -0.0016 (0.0053) | -0.0019 (0.0072) |
| | has_Freeparkingonpremises | 0.0075 *** (0.002) | 0.002 (0.002) | 0.002 (0.0018) | 0.0022 (0.0017) | 0.0049 * (0.0022) | 0.0077 * (0.003) |
| | has_Securitycamerasonproperty | -0.0027 (0.002) | 0.0047 * (0.0021) | 0.0004 (0.0019) | -0.0025 (0.0017) | -0.0039 (0.0022) | -0.0079 ** (0.003) |
| | has_Gym | 0.0354 *** (0.0032) | 0.0358 *** (0.0035) | 0.0327 *** (0.003) | 0.0316 *** (0.0027) | 0.036 *** (0.0037) | 0.034 *** (0.005) |
| | has_Smokealarm | -0.0331 *** (0.0049) | -0.0081 (0.0051) | -0.0079 (0.0045) | -0.0158 (0.0041) | -0.0412 *** (0.0052) | -0.0686 *** (0.0072) |
| Características de calificaciones del huésped | review_scores_location | 0.0933 *** (0.0143) | 0.0879 *** (0.0144) | 0.0954 *** (0.0125) | 0.0937 *** (0.012) | 0.1279 *** (0.0173) | 0.0961 *** (0.0228) |
| | number_of_reviews | -0.0734 *** (0.0101) | -0.0087 (0.01) | -0.0247 ** (0.0087) | -0.0457 *** (0.0084) | -0.0812 *** (0.0123) | -0.0927 *** (0.0195) |
| | review_scores_rating | 0.0818 *** (0.0227) | 0.0247 (0.024) | 0.0352 (0.0213) | 0.1021 *** (0.019) | 0.0983 *** (0.026) | 0.117 *** (0.0346) |
| | review_scores_accuracy | 0.0031 (0.0188) | -0.0028 (0.0181) | 0.0034 (0.0179) | -0.0147 (0.0157) | -0.0121 (0.0201) | 0.0172 (0.024) |
| | review_scores_cleanliness | 0.0912 *** (0.0151) | 0.0922 *** (0.0156) | 0.0898 *** (0.014) | 0.0723 *** (0.0126) | 0.0786 *** (0.0171) | 0.072 ** (0.0235) |
| | review_scores_checkin | -0.0524 ** (0.0175) | -0.0579 ** (0.0195) | -0.029 (0.0167) | -0.0408 ** (0.0146) | -0.0687 *** (0.019) | -0.0707 ** (0.0234) |

| | | | | | | | |
|---------------------------------|--|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | re-view_scores_communi- cation | -0.0193 (0.0187) | 0.0223 (0.0179) | -0.0269 (0.0176) | -0.0157 (0.0156) | -0.0384 (0.0203) | -0.0445 (0.0252) |
| | review_scores_value | -0.0837 *** (0.0161) | -0.0383 * (0.0184) | -0.0356 * (0.0163) | -0.0855 *** (0.0134) | -0.1014 *** (0.0174) | -0.1136 *** (0.0227) |
| Características del host | host_has_profile_pic | 0.03 *** (0.0066) | 0.0274 *** (0.0061) | 0.0274 *** (0.0057) | 0.0354 *** (0.0055) | 0.0418 *** (0.008) | 0.0506 *** (0.0122) |
| | calcula- ted host listings count | -0.0277 *** (0.0053) | 0.024 *** (0.0051) | 0.0051 (0.0046) | -0.0089 * (0.0044) | -0.0323 *** (0.0062) | -0.0467 *** (0.009) |
| | host_acceptance_rate | 0.0066 * (0.003) | 0.0088 ** (0.0029) | 0.0066 * (0.0027) | 0.0092 *** (0.0025) | 0.0086 * (0.0034) | -0.0183 *** (0.0047) |
| | host_identity_verified | -0.0076 ** (0.0024) | -0.0019 (0.0024) | -0.0042 (0.0022) | -0.0048 * (0.002) | -0.0063 * (0.0026) | -0.0102 ** (0.0037) |
| | host_response_time_h our | -0.0265 *** (0.0059) | 0.0054 (0.0059) | -0.007 (0.0051) | -0.0108 * (0.0049) | -0.0292 *** (0.0066) | -0.0465 *** (0.0094) |
| | host_verifications | 0.0167 *** (0.0037) | 0.0087 * (0.0036) | 0.0121 *** (0.0033) | 0.0129 *** (0.0031) | 0.0163 *** (0.004) | 0.0151 ** (0.0054) |
| | host_is_superhost | 0.0035 (0.0022) | 0.0092 *** (0.0022) | 0.0049 * (0.002) | 0.0029 (0.0018) | -0.0025 (0.0023) | -0.0038 (0.0031) |
| | host_about | -0.0044 * (0.0019) | -0.0022 (0.0019) | -0.0021 (0.0018) | -0.0033 * (0.0016) | -0.0062 ** (0.0021) | -0.0047 (0.003) |
| | min_dist_to_points_inte rest ft | 0.0161 (0.0093) | 0.0131 (0.0089) | 0.0223 ** (0.0081) | 0.0198 * (0.0078) | 0.026 * (0.0104) | 0.0231 (0.0139) |
| Características de ubicación | min_dist_to_white_hou se_ft | -0.2812 *** (0.0211) | -0.2219 *** (0.0203) | -0.2311 *** (0.0188) | -0.2662 *** (0.0176) | -0.2947 *** (0.0239) | -0.3852 *** (0.0325) |
| | min_dist_to_national_s pace museum ft | 0.135 *** (0.0177) | 0.1014 *** (0.0167) | 0.0981 *** (0.0155) | 0.1275 *** (0.0148) | 0.1466 *** (0.0204) | 0.2112 *** (0.028) |
| | min_dist_to_metro_bus stop ft | 0.0354 ** (0.0124) | 0.0162 (0.0114) | 0.0007 (0.0108) | 0.0108 (0.0103) | 0.0415 ** (0.0138) | 0.0539 ** (0.0188) |
| | min_dist_to_bank_ft | -0.0531 *** (0.0095) | -0.045 *** (0.0097) | -0.0358 *** (0.0086) | -0.0413 *** (0.008) | -0.0589 *** (0.0104) | -0.0733 *** (0.0139) |
| | min_dist_to_national_z oo ft | 0.1042 *** (0.0115) | 0.0742 *** (0.012) | 0.0797 *** (0.0107) | 0.0997 *** (0.0096) | 0.117 *** (0.0125) | 0.1635 *** (0.0163) |

Fuente: Elaboración propia.

Coefficientes y entre paréntesis los errores estándares. Significancia estadística *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

En la **¡Error! No se encuentra el origen de la referencia.** 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.

Tabla 6. P-valores obtenidos de cada modelo de regresión.

| Categoría | Característica | OLS | Q 0.1 | Q 0.25 | Q 0.5 | Q 0.75 | Q 0.9 |
|--|------------------------|-------|-------|--------|-------|--------|-------|
| características del alojamiento | shared_room | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | beds | 0.071 | 0.000 | 0.000 | 0.888 | 0.797 | 0.779 |
| | availability_365 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | instant_bookable | 0.057 | 0.004 | 0.000 | 0.011 | 0.692 | 0.511 |
| | bathrooms_quantity | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | minimum_nights | 0.007 | 0.000 | 0.000 | 0.000 | 0.001 | 0.119 |
| | bedrooms | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | accommodates | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | entire_home_aprt | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | hotel_room | 0.000 | 0.000 | 0.000 | 0.000 | 0.005 | 0.068 |
| Características de servicios adicionales | has_Hotwater | 0.000 | 0.000 | 0.000 | 0.006 | 0.000 | 0.000 |
| | has_Petsallowed | 0.826 | 0.242 | 0.054 | 0.148 | 0.527 | 0.815 |
| | has_Dedicatedworkspace | 0.022 | 0.038 | 0.005 | 0.320 | 0.169 | 0.077 |
| | has_Kitchen | 0.000 | 0.031 | 0.004 | 0.000 | 0.000 | 0.000 |
| | has_Hairdryer | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.100 |

| | | | | | | | |
|--|--------------------------------------|-------|-------|-------|-------|-------|-------|
| | has_Wifi | 0.138 | 0.659 | 0.790 | 0.804 | 0.219 | 0.000 |
| | has_Cleaningproducts | 0.385 | 0.256 | 0.098 | 0.289 | 0.848 | 0.625 |
| | has_Washer | 0.030 | 0.301 | 0.490 | 0.796 | 0.135 | 0.039 |
| | has_Airconditioning | 0.070 | 0.536 | 0.805 | 0.732 | 0.002 | 0.013 |
| | has_Freezer | 0.300 | 0.439 | 0.035 | 0.181 | 0.003 | 0.009 |
| | has_TV | 0.952 | 0.014 | 0.029 | 0.512 | 0.008 | 0.010 |
| | has_Dishwasher | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.005 |
| | has_Freestreetparking | 0.002 | 0.033 | 0.290 | 0.072 | 0.000 | 0.000 |
| | has_Breakfast | 0.029 | 0.019 | 0.905 | 0.334 | 0.000 | 0.000 |
| | has_Pool | 0.000 | 0.000 | 0.000 | 0.000 | 0.764 | 0.789 |
| | has_Freeparkingonpremises | 0.000 | 0.327 | 0.287 | 0.196 | 0.027 | 0.011 |
| | has_Securitycamerasonproperty | 0.174 | 0.025 | 0.820 | 0.135 | 0.079 | 0.008 |
| | has_Gym | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | has_Smokealarm | 0.000 | 0.115 | 0.080 | 0.000 | 0.000 | 0.000 |
| Características de calificacio- nes del hués- ped | review_scores_location | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | number_of_reviews | 0.000 | 0.385 | 0.005 | 0.000 | 0.000 | 0.000 |
| | review_scores_rating | 0.000 | 0.304 | 0.099 | 0.000 | 0.000 | 0.001 |
| | review_scores_accuracy | 0.871 | 0.879 | 0.851 | 0.349 | 0.547 | 0.475 |
| | review_scores_cleanliness | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 |
| | review_scores_checkin | 0.003 | 0.003 | 0.083 | 0.005 | 0.000 | 0.003 |
| | review_scores_communication | 0.301 | 0.213 | 0.126 | 0.316 | 0.058 | 0.077 |
| | review_scores_value | 0.000 | 0.038 | 0.029 | 0.000 | 0.000 | 0.000 |
| Características del host | host_has_profile_pic | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | calculated_host_listings_count | 0.000 | 0.000 | 0.265 | 0.044 | 0.000 | 0.000 |
| | host_acceptance_rate | 0.029 | 0.003 | 0.015 | 0.000 | 0.011 | 0.000 |
| | host_identity_verified | 0.001 | 0.416 | 0.052 | 0.017 | 0.016 | 0.005 |
| | host_response_time_hour | 0.000 | 0.363 | 0.170 | 0.027 | 0.000 | 0.000 |
| | host_verifications | 0.000 | 0.015 | 0.000 | 0.000 | 0.000 | 0.005 |
| | host_is_superhost | 0.103 | 0.000 | 0.012 | 0.105 | 0.291 | 0.225 |
| | host_about | 0.024 | 0.253 | 0.228 | 0.042 | 0.004 | 0.111 |
| Características de ubicación | min_dist_to_points_of_interest_ft | 0.083 | 0.142 | 0.006 | 0.011 | 0.013 | 0.095 |
| | min_dist_to_white_house_ft | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | min_dist_to_national_space_museum_ft | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | min_dist_to_metro_bus_stop_ft | 0.004 | 0.155 | 0.950 | 0.294 | 0.003 | 0.004 |
| | min_dist_to_bank_ft | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | min_dist_to_national_zoo_ft | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Fuente: Elaboración propia.

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 formula, 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 [Figura 1](#). No se encuentra el origen de la referencia..

Tabla 7. Porcentajes de incremento / disminución en el precio.

| | | | | | | | |
|---|-----------------------------------|---------|---------|---------|---------|---------|---------|
| características del alojamiento | shared_room | -7.946 | -6.565 | -5.474 | -6.293 | -8.853 | -11.786 |
| | beds | -3.950 | -11.529 | -7.900 | -0.260 | 0.632 | 1.035 |
| | availability_365 | 1.593 | 1.776 | 1.643 | 1.369 | 1.126 | 1.623 |
| | instant_bookable | 0.361 | 0.602 | 0.955 | 0.411 | 0.080 | -0.180 |
| | bathrooms_quantity | 21.799 | 17.386 | 23.652 | 23.528 | 20.804 | 34.138 |
| | minimum_nights | -6.434 | -49.348 | -41.527 | -22.555 | -12.138 | 9.133 |
| | bedrooms | 47.123 | 60.368 | 48.305 | 50.230 | 50.893 | 49.586 |
| | accommodates | 17.210 | 11.505 | 13.021 | 12.176 | 17.869 | 18.353 |
| | entire_home_apartment | 4.102 | 5.855 | 5.717 | 5.369 | 3.324 | 2.276 |
| Características de servicios adicionales | hotel_room | -10.192 | -9.045 | -12.173 | -14.521 | -4.906 | 4.613 |
| | has_hotwater | -1.282 | -1.331 | -1.203 | -0.558 | -1.143 | -2.557 |
| | has_pets_allowed | 0.501 | -2.751 | 3.894 | 2.768 | 1.704 | 0.793 |
| | has_dedicated_workspace | -0.429 | -0.389 | -0.489 | -0.160 | -0.280 | -0.489 |
| | has_kitchen | -1.931 | -0.707 | -0.817 | -1.390 | -2.420 | -1.961 |
| | has_hair_dryer | 1.248 | 1.248 | 1.197 | 0.874 | 1.066 | 0.592 |
| | has_wifi | 0.461 | -0.140 | 0.080 | 0.060 | 0.411 | 1.593 |
| | has_cleaning_products | 0.230 | 0.310 | 0.411 | 0.240 | 0.050 | -0.190 |
| | has_washer | -0.449 | -0.220 | -0.130 | 0.040 | -0.339 | -0.638 |
| | has_airconditioning | 0.441 | -0.150 | -0.050 | 0.070 | 0.813 | 0.884 |
| | has_freezer | 0.280 | 0.210 | 0.511 | 0.300 | 0.864 | 1.015 |
| | has_tv | 0.010 | 0.491 | 0.381 | -0.100 | -0.539 | -0.707 |
| | has_dishwasher | 1.177 | 1.603 | 1.339 | 1.308 | 1.177 | 0.864 |
| | has_freestreet_parking | -0.648 | 0.461 | 0.200 | -0.310 | -0.985 | -1.361 |
| | has_breakfast | 0.763 | -0.817 | -0.040 | 0.280 | 1.603 | 3.025 |
| | has_pool | -1.921 | -2.829 | -2.926 | -1.686 | -0.160 | -0.190 |
| | has_freeparking_on_premises | 0.753 | 0.200 | 0.200 | 0.220 | 0.491 | 0.773 |
| | has_security_camera_on_property | -0.270 | 0.471 | 0.040 | -0.250 | -0.389 | -0.787 |
| | has_gym | 3.603 | 3.645 | 3.324 | 3.210 | 3.666 | 3.458 |
| | has_smoke_alarm | -3.256 | -0.807 | -0.787 | -1.568 | -4.036 | -6.630 |
| Características de calificaciones del huésped | review_scores_location | 9.779 | 9.188 | 10.010 | 9.823 | 13.644 | 10.087 |
| | number_of_reviews | -7.077 | -0.866 | -2.440 | -4.467 | -7.799 | -8.853 |
| | review_scores_rating | 8.524 | 2.501 | 3.583 | 10.749 | 10.329 | 12.412 |
| | review_scores_accuracy | 0.310 | -0.280 | 0.341 | -1.459 | -1.203 | 1.735 |
| | review_scores_cleanliness | 9.549 | 9.658 | 9.396 | 7.498 | 8.177 | 7.466 |
| | review_scores_checkin | -5.105 | -5.626 | -2.858 | -3.998 | -6.639 | -6.826 |
| | review_scores_communication | -1.911 | 2.255 | -2.654 | -1.558 | -3.767 | -4.352 |
| Características del host | review_scores_value | -8.029 | -3.758 | -3.497 | -8.195 | -9.643 | -10.739 |
| | host_has_profile_picture | 3.045 | 2.778 | 2.778 | 3.603 | 4.269 | 5.190 |
| | calculated_host_listings_count | -2.732 | 2.429 | 0.511 | -0.886 | -3.178 | -4.563 |
| | host_acceptance_rate | 0.662 | 0.884 | 0.662 | 0.924 | 0.864 | -1.813 |
| | host_identity_verified | -0.757 | -0.190 | -0.419 | -0.479 | -0.628 | -1.015 |
| | host_response_time_hour | -2.615 | 0.541 | -0.698 | -1.074 | -2.878 | -4.544 |
| | host_verifications | 1.684 | 0.874 | 1.217 | 1.298 | 1.643 | 1.521 |
| | host_is_superhost | 0.351 | 0.924 | 0.491 | 0.290 | -0.250 | -0.379 |
| | host_about | -0.439 | -0.220 | -0.210 | -0.329 | -0.618 | -0.469 |
| Características de ubicación | min_dist_to_points_of_interest | 1.623 | 1.319 | 2.255 | 2.000 | 2.634 | 2.337 |
| | min_dist_to_white_house | -24.512 | -19.900 | -20.634 | -23.371 | -25.525 | -31.969 |
| | min_dist_to_national_space_museum | 14.454 | 10.672 | 10.307 | 13.598 | 15.789 | 23.516 |
| | min_dist_to_metro_bus_stop | 3.603 | 1.633 | 0.070 | 1.086 | 4.237 | 5.538 |
| | min_dist_to_bank | -5.171 | -4.400 | -3.517 | -4.046 | -5.720 | -7.068 |
| | min_dist_to_national_zoo | 10.982 | 7.702 | 8.296 | 10.484 | 12.412 | 17.763 |

Fuente: Elaboración propia.

Los resultados serán analizados por las categorías definidas anteriormente.

Características del alojamiento

Para esta categoría, en el modelo OLS las variables `beds` e `instant_bookable` no son estadísticamente significativa, el resto si 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 que el precio se incrementa hasta 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% y 10% respectivamente.

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 34% 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 60% y en los alojamientos que tienen precios más altos aumenta en un 49%, 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 17%. La disponibilidad es estadísticamente significativa en los dos modelos, se observa un valor casi constante en todos los cuantiles.

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 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 1.6% 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 tiene un gran aumento en el precio.

Con los alojamientos que poseen cocina, se observan valores negativos tanto en OLS como en QR, esto quiere decir que cuando un alojamiento tiene cocina, el precio tiende a bajar. En OLS los precios disminuyen hasta un 1.9%, el mismo comportamiento se muestra en QR en los alojamientos con mayor precio (cuantil 0.9). 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 2.5%. 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 0.8% hasta 1%, excepto en el cuantil 0.9 en donde esta variable no es estadísticamente significativa.

Cuando un alojamiento posee 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 los cuantiles 0.1 y 0.75 es donde más se presenta un aumento del precio (3.6%), en el cuantil 0.25 el aumento es del 3.3%, en el cuantil 0.5 el aumento es del 3.2% y en el 0.9 el aumento es de 3.4%, 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 el cuantil 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 0.4% 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 0.7%.

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 0.1 y 0.25, en ellos el precio disminuye entre un 0.3% y 0.4%. Con el parqueadero ocurre que aquellos parqueaderos que están en la calle en los alojamientos por arriba de la mediana representan una disminución del precio entre 0.9% y 1.3%, 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 entre 0.4% y 0.7% de aumento en el precio, en el resto de los cuantiles no es significativo este atributo.

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 3%.

Se observa una reducción de precios para aquellos alojamientos que tienen piscina, en los cuantiles 0.75 y 0.9 no se observa significancia, pero en los inferiores a este sí, el precio puede reducir entre 1.6% y 2.9%, esto no concuerda mucho con lo encontrado en otros estudios.

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 los cuantiles 0.1 y 0.25 no son estadísticamente significativos, el resto sí 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 7.4% hasta el 9.6% en el precio, siendo en los precios más bajos (cuantil 0.1) en donde mayor porcentaje de aumento representa. En el puntaje general el aumento del precio más bajo se encuentra en la regresión OLS (8.5%), de resto, se ven aumentos entre 10% y 12% en los cuantiles de la mediana hacia arriba, es decir, los alojamientos con precios mayores que tienen mejor puntaje tienden a aumentar estos porcentajes. Los puntajes para checkin y calidad poseen valores negativos, se observan reducciones de precio entre 3.9% y 6.8% para el checkin y una reducción de los precios entre 3.7% y 10.7% para los

puntajes de calidad. Las calificaciones de comunicación y precisión no son estadísticamente significativas para ninguno de los modelos.

En general se muestra que cuando los huéspedes dan buenas calificaciones de ubicación a los alojamientos, los precios de estos tienden a aumentar, tiene sentido ya que es una referencia para otros huéspedes y entre mejor ubicado esté el alojamiento, mejor será la experiencia en este lugar.

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.

Cuando un host es superhost solamente tiene importancia en los cuantiles 0.1 y 0.25, es decir en los alojamientos con menor precio, representando un aumento de hasta un 0.9% en el precio, en el resto de los cuantiles ni en OLS se observa significancia.

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 5%, siendo la última mencionada la que más aporta, cuando el host tiene imagen de perfil los precios aumentan desde 2.7% hasta 5.1%. La tasa de aceptación del host muestra un aumento de los precios, excepto en los alojamientos más caros (cuantil 0.9), en donde se presenta una reducción de 1.8% del precio.

Por otro lado, las variables 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 0.2% hasta 4.5%, 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 2.4% 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 0.8% hasta 4.5% en los alojamientos con precios más altos.

La tasa de aceptación del host en todos los cuantiles y en OLS es estadísticamente significativa y en el 0.9 aporta reducción al precio de hasta 2%, en los cuantiles de 0.75 hacia abajo aporta un incremento en los precios entre 0.6% y 0.9%.

La identidad verificada del host es significativamente estadística en el cuantil 0.5 hacia arriba y aporta una reducción entre 0.5% y 1% 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", cuando el host tiene imagen de perfil al parecer genera mayor confianza entre los huéspedes y esto repercute en los precios.

Características de ubicación

La distancia hacia el Zoológico Nacional es significativa en todos los modelos con valores positivos, siendo en los alojamientos más caros (Q0.9) donde el precio aumenta hasta un 17% y en el resto entre 7% y 12%. En OLS se observa que los precios aumentan también, en por lo menos 10%, lo que indica que entre más aumente la distancia entre el alojamiento y este sitio, más costoso será el alojamiento.

Así mismo con los puntos de interés, estos influyen positivamente en el precio en los cuantiles 0.25, 0.5 y 0.75, en el resto no se presenta significancia, en el cuantil 0.25 se presenta un 2.25% de incremento y en el 0.5 aumentan en 1.9%, por último, en el cuantil 0.75 es donde los precios aumentan más, hasta 2.6%. En general los alojamientos con precios medios que están más lejanos a algún punto de interés son más costosos, por otro lado, tanto para OLS como los Q 0.1 y 0.9 no es significativo esto.

La distancia al banco también presenta valores significativos en todos los modelos, presentando valores negativos, lo que quiere decir que los alojamientos que están más alejados de un banco por lo general son más baratos. En los alojamientos más baratos se presenta una reducción de 4.4% y en los precios más altos se presenta una reducción más grande (7%).

Por otro lado, la distancia a la Casa Blanca presenta valores significativamente negativos. En OLS se presenta una reducción de hasta 24% en el precio y en los alojamientos con los precios más altos se presenta una disminución del precio de hasta 31%. Quiere decir que entre más distancia haya entre el alojamiento y la Casa Blanca, los precios serán en promedio, más baratos.

Alojamientos cercanos a paradas de bus solamente presentan significancia en los cuantiles más altos presentando un incremento de más de 4% en el precio. En cuanto a los alojamientos cercanos al Museo Nacional del Espacio, se presentan significancias en todos los modelos, el aumento del precio es bastante significativo, desde 10% hasta 23%. Por lo tanto, aquellos alojamientos que están lejanos a este sitio tienden en promedio a tener precios mayores.

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 para atraer a posibles huéspedes y sacar ganancias.

El análisis constó de 51 variables repartidas en 5 categorías: características del alojamiento con 10 variables, Características de servicios adicionales con 19 variables, Características de calificaciones del huésped con 8 variables, Características del host con 8 variables, Características de ubicación con 6 variables.

En la regresión OLS 38 de las 51 variables son estadísticamente significativas, es decir, si influyen en el precio. Por otro lado, en la regresión cuantílica la significancia varía con respecto a todos los cuantiles calculados, se observan significancia en la gran mayoría de cuantiles, solamente 4 variables no son significativas en ninguno de los cinco cuantiles calculados, en el resto de las variables la significancia varía dependiendo al cuantil, hay unos donde si hay significancia y hay otros donde no la hay.

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. Las habitaciones compartidas siempre tendrán un precio menor en comparación a los alojamientos que son casas o apartamentos. Entre más sea la cantidad de personas permitidas en un alojamiento, mayores serán los precios.

Los hosts con una buena tasa de aceptación y que tienen imagen de perfil generalmente tienen precios más altos, gracias a esto los usuarios sienten más confianza y están dispuestos a pagar precios más altos. Contrario a lo que dicen otros estudios, la variable que indica la identidad verificada del host, resultó que influye negativamente en los precios.

Amenidades como si el alojamiento posee gimnasio, secador de cabello y lavavajillas también afectan el precio positivamente, lo cual permite a los hosts cobrar un valor más alto por noche con respecto a los que no ofrecen estos servicios.

Cuanto mayor sea la calificación promedio de los clientes, la calificación de limpieza y localización, mayor será el precio, esto indica que los usuarios se fijan especialmente en estas calificaciones para hacer sus reservas y por lo tanto concluye en aumento de los precios. Además, el número de comentarios influye negativamente en el precio.

En cuanto a la ubicación, la cercanía a los principales lugares de interés turístico y al Zoológico Nacional contribuye al incremento en los precios. En general, la ubicación

del alojamiento en esta ciudad es uno de los factores más importantes que influyen en el precio de estos.

Este análisis indaga en aquellos determinantes de los precios de los alquileres de alojamiento de Airbnb en la ciudad de Washington DC, usando un conjunto de datos con los alojamientos de esta ciudad usando algoritmos de regresión OLS y QR en diferentes rangos de precios. Este análisis aporta información importante sobre los alojamientos de la ciudad estudiada, y es útil para aquellos hosts que quieran tener un referente sobre el precio de sus propiedades y así puedan incrementar sus ganancias. Se logró identificar la correlación entre las variables para poder elegir aquellas que explicaran los datos y no existieran datos redundantes.

Cabe señalar 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 o no arrojando los mismos resultados vistos en otros estudios, 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 y explorar con otras características. Adicionalmente, no se han tenido en cuenta otros factores externos como sociales, climáticos o de temporada (es decir, alguna fiesta o evento importante que ocurra en la ciudad), por lo tanto, sería importante, para futuros estudios, el tener en cuenta estos otros factores.

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

Carvajal, K. (2023, March). *Análisis de características influyentes en los precios de alojamientos de Airbnb en la ciudad de Washington D.C usando modelos de regresión.* <https://github.com/kcarvajalb/TrabajoKaren/blob/main/AIRBNB.ipynb>

Cerimelo, M., & Porto, N. (2022). Atributos, puntuación y precios hedónicos en la Provincia de Buenos Aires y CABA. *Revista Brasileira de Pesquisa Em Turismo*, 16, 2587. <https://doi.org/10.7784/RBTUR.V16.2587>

Cómo fijar el precio de tu alojamiento - Centro de ayuda de Airbnb. (n.d.). Retrieved March 17, 2023, from <https://www.airbnb.com.co/help/article/52>

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>

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2018). Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. *Journal of Travel and Tourism Marketing*, 35(1), 46–56. <https://doi.org/10.1080/10548408.2017.1308292>

Gómez, D., Cantu-Ortiz, F. J., Contreras, V., & Diaz Ramos, R. E. (2020, January). *MEXICO CITY'S AIRBNB LISTING PRICE ANALYSIS USING REGRESSION.*

https://www.researchgate.net/publication/348975238_MEXICO_CITY%27S_AIRBNB_LISTING_PRICE_ANALYSIS_USING_REGRESSION

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_beve_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>

Poeta, S., Gerhardt, T., & Stumpf Gonzalez, M. (2019). Análisis de precios hedónicos de viviendas. *Revista Ingeniería de Construcción*, 34(2), 215–220. <https://doi.org/10.4067/S0718-50732019000200215>

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>

Samwel, M. (2022). What Factors drives the Airbnb Listing's Prices? *International Business & Economics Studies*, 4(1), p26. <https://doi.org/10.22158/IBES.V4N1P26>

Sandoval, J. (2008, June). *REVISIÓN BIBLIOGRÁFICA MODELO DE PRECIOS HEDÓNICOS*. https://www.researchgate.net/publication/355778324_REVISION_BIBLIOGRAFICA_MODELO_DE_PRECIOS_HEDONICOS

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., Negruș, 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>

Anexos I

Mapa de promedio de puntaje de alojamientos por barrio

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

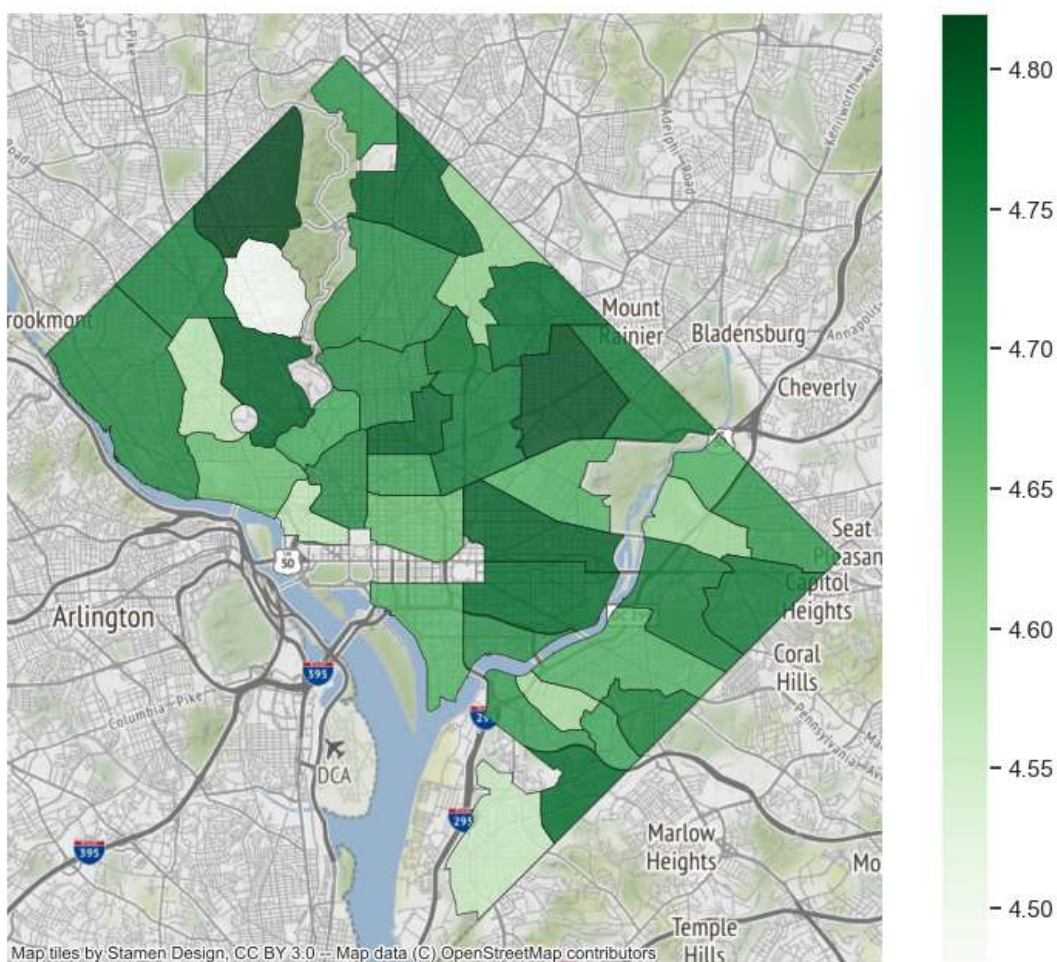


Ilustración 155. Puntaje promedio de alojamientos de Airbnb por barrio en la ciudad.

Fuente: Elaboración propia.

Mapa de promedio de puntaje de alojamientos por barrio

Promedio de disponibilidad por barrio en la ciudad de Washington DC

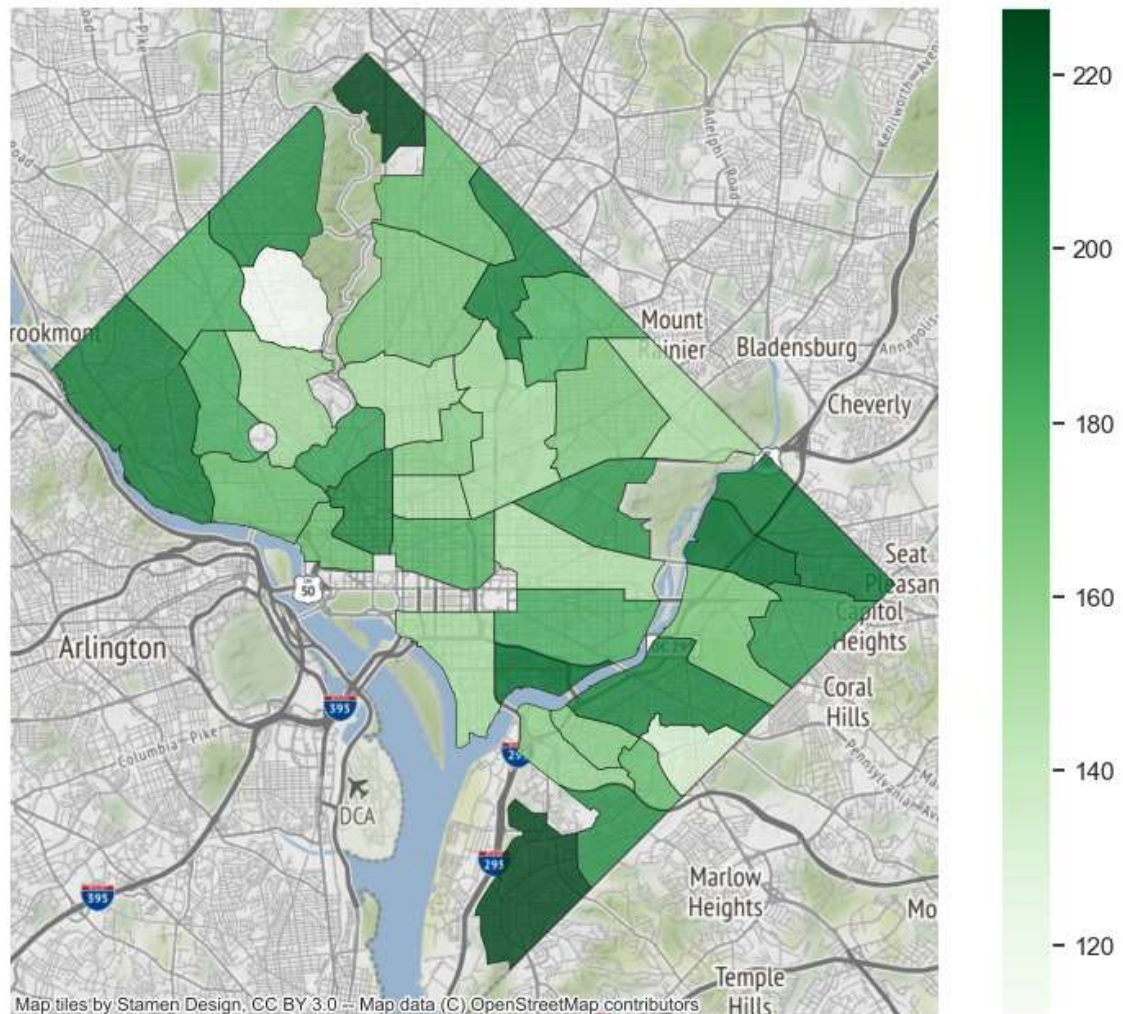


Ilustración 166. Promedio de disponibilidad por barrio en la ciudad.

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

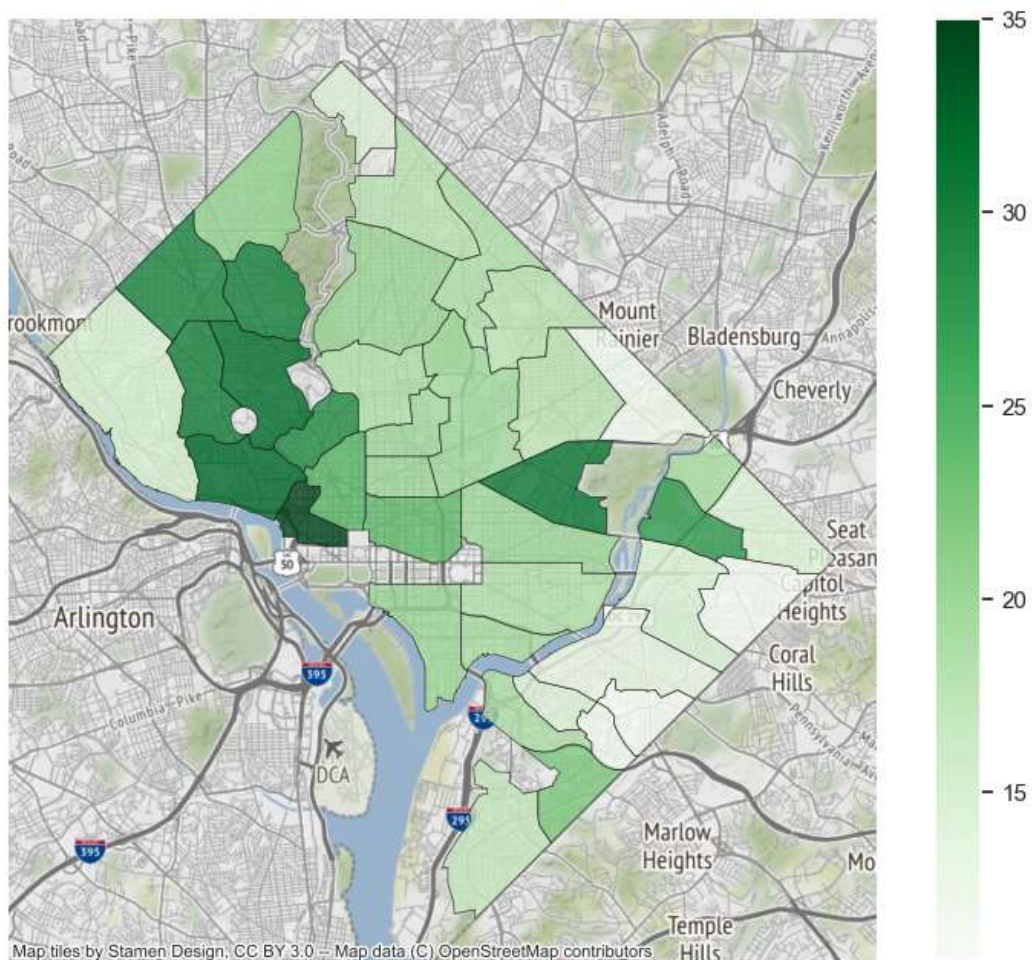


Ilustración 177. Promedio de noches mínimas para hacer reserva por barrio.

Fuente: Elaboración propia

| has_Breakfast | Cantidad | Precio promedio (USD) |
|-------------------------------|-----------------|------------------------------|
| | 0 | 6051 |
| | 1 | 419 |
| has_Washer | Cantidad | Precio promedio (USD) |
| | 0 | 3003 |
| | 1 | 3467 |
| has_Dishwasher | Cantidad | Precio promedio (USD) |
| | 0 | 3245 |
| | 1 | 3225 |
| has_Dryer | Cantidad | Precio promedio (USD) |
| | 0 | 3257 |
| | 1 | 3213 |
| has_TV | Cantidad | Precio promedio (USD) |
| | 0 | 3325 |
| | 1 | 3145 |
| has_Freezer | Cantidad | Precio promedio (USD) |
| | 0 | 4562 |
| | 1 | 1908 |
| has_Airconditioning | Cantidad | Precio promedio (USD) |
| | 0 | 1884 |
| | 1 | 4586 |
| has_Heating | Cantidad | Precio promedio (USD) |
| | 0 | 1401 |
| | 1 | 5069 |
| has_Hotwater | Cantidad | Precio promedio (USD) |
| | 0 | 1566 |
| | 1 | 4904 |
| has_Cleaningproducts | Cantidad | Precio promedio (USD) |
| | 0 | 4635 |
| | 1 | 1835 |
| has_Wifi | Cantidad | Precio promedio (USD) |
| | 0 | 575 |
| | 1 | 5895 |
| has_Hairdryer | Cantidad | Precio promedio (USD) |
| | 0 | 1399 |
| | 1 | 5071 |
| has_Kitchen | Cantidad | Precio promedio (USD) |
| | 0 | 723 |
| | 1 | 5747 |
| has_Dedicatedworkspace | Cantidad | Precio promedio (USD) |
| | 0 | 3387 |
| | 1 | 3083 |

| 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 |