



## Análisis de los Alquileres de Alojamientos Turísticos No Hosteleros en Málaga: Factores, Tendencias Y Excesos

Un Estudio por Barrios sobre los Determinantes del Precio, Tendencias de Mercado y Detección de Alojamientos con Precios Excesivos

Título en inglés: Analysis of Non-Hotel Tourist Accommodation Rentals in Málaga: Factors, Trends, and Excesses

Nº de palabras 8729

Curso académico: 2023/2024

Autor: Jose Luis Vicaria Cañaveras

Correo electrónico: [JoseLuis.Vicaria@alu.uclm.es](mailto:JoseLuis.Vicaria@alu.uclm.es)

Tutor: Miguel Ángel Tarancón Morán

Fecha 29/09/2024

Firma autor:

*Jose Luis Vicaria Cañaveras*

Firma tutor:



# Trabajo Fin de Grado

máster universitario en modelización  
y análisis de datos económicos

*“Como la hierba, la idea busca la luz,  
ama las multitudes,  
el mestizaje la enriquece,  
crece más vigorosa cuando se la pisa.”*

*Úrsula K. Le Guin*



## Índice

Análisis de los Alquileres de Alojamientos Turísticos No Hosteleros en Málaga: Factores, Tendencias Y Excesos.....	1
Abstract.....	4
Marco, justificación y objetivos del trabajo.....	5
Metodología Empleada .....	7
Desarrollo y resultados .....	8
Análisis de la oferta.....	9
Análisis descriptivo.....	12
Modelo de regresión .....	17
Análisis Clúster .....	21
Modelo logit.....	30
Conclusiones .....	37
Bibliografía .....	42
Anexo I – Tratamiento de las variables.....	44
Anexo II – Descriptivo de las variables .....	46



# Trabajo Fin de Grado

máster universitario en modelización  
y análisis de datos económicos

## Abstract

Este Trabajo Fin de Máster analiza los alquileres de alojamientos turísticos no hosteleros en Málaga, con el objetivo de identificar factores que afectan sus precios. La investigación se justifica por el impacto creciente del turismo en la economía local. Se emplea una metodología que incluye análisis descriptivo, modelos de regresión y análisis clúster, así como un modelo logit para determinar si el precio de un cierto alojamiento es o no excesivo. Los resultados revelan patrones significativos en la oferta y la demanda, proporcionando información valiosa para la gestión del sector turístico y la formulación de políticas locales.

This Master's thesis analyzes the rental prices of non-hotel tourist accommodations in Málaga, aiming to identify factors influencing these prices. The research is justified by the growing impact of tourism on the local economy. A methodology is employed that includes descriptive analysis, regression models, and cluster analysis, as well as a logit model to assess whether the price of a given accommodation is excessive. The results reveal significant patterns in supply and demand, providing valuable insights for tourism management and local policy formulation.

## Marco, justificación y objetivos del trabajo

El sector turístico es uno de los pilares más fuertes de la provincia de Málaga y de su ciudad.

Durante el siglo XX, especialmente en el primer tercio del siglo y posteriormente a partir de los años 60 con el aperturismo del régimen franquista, hubo un incentivo al crecimiento del turismo, viéndose la Costa del Sol fuertemente influenciada debido a sus características climáticas y sus playas (Pellejero Martínez, s.f.), (Fu, 2023)

Desde entonces, el impacto y el peso del turismo en Málaga no han dejado de crecer. Durante el primer cuatrimestre de 2024, la Costa del Sol registró un incremento del 8,1% en la estancia de viajeros en establecimientos reglados y un 13,5% en alojamientos turísticos (Diputación de Málaga, 2023). Es sobre este último porcentaje donde se centrará este análisis. La inmensa mayoría de estos alojamientos turísticos, cuyo número en la ciudad de Málaga no para de crecer, se encuentran anunciados en la plataforma Airbnb. Estos se conocen como alojamientos turísticos no hoteleros.

Este crecimiento no está exento de controversia. El pasado mes de julio se llevaron a cabo grandes manifestaciones en contra del turismo masivo, culpando a Airbnb del aumento de los precios del alquiler (Euronews, 2024).

Se ha de destacar a su vez que el precio de los alojamientos turísticos no hosteleros en Málaga tampoco ha dejado de subir. No es solo que estos hayan aumentado de número, sino que se da un crecimiento de los precios que hace indicar que, pese a que la oferta es cada vez mayor, quizá aún no sea suficiente para cubrir toda la demanda.

Y bien ¿En qué debería fijarse un posible cliente a la hora de buscar un alojamiento no turístico en Málaga? ¿En el número de habitaciones?, ¿En el barrio en que se encuentra el alojamiento? Pese a que al viajar se suele tener una idea general de las necesidades y de lo que se busca, realmente no se suelen tener las mismas completamente claras, ni las herramientas para poder saber si lo que se encuentra es adecuado a las mismas o no.

Lo que es innegable es que el turismo es un sector muy importante en Málaga, y que los posibles clientes se ven afectados directamente por los precios y características de los alojamientos en que van a alojarse, sin tener ningún medio con el cual saber, más allá de su intuición, si están reservando un alojamiento con buenas, malas, o incluso excesivas condiciones.



No cabe duda de que los alojamientos turísticos no hoteleros son un tema de actualidad en Málaga.

En este contexto de actualidad e importancia del sector del alquiler turístico en la ciudad de Málaga, cabe preguntarse, desde un punto de vista económico ¿Cómo está dicho mercado?, ¿Cuáles son las características del mismo?, ¿Qué factores determinan los precios? Entre muchas otras.

El objetivo de este trabajo es realizar un análisis del mercado de alquiler de alojamientos turísticos no hoteleros en la ciudad de Málaga, específicamente aquellos que se anuncian en la plataforma Airbnb. Este análisis busca identificar y comprender qué características presentes en los anuncios de la plataforma resultan más interesantes, importantes y significativas desde la perspectiva del posible cliente. Asimismo, se pretende desarrollar un modelo que permita responder a la pregunta: ¿Es el precio de un determinado anuncio adecuado o excesivo?

Para alcanzar estos objetivos, se llevará a cabo un estudio detallado que incluirá varias fases. En primer lugar, se procederá a la recopilación y limpieza de datos, extrayendo los mismos de la base de datos de AirBNB (Inside Airbnb (<https://insideairbnb.com/get-the-data/>)). Los datos recogidos incluirán variables como la ubicación, el precio por noche, la capacidad de alojamiento, las comodidades ofrecidas, tasas de respuestas a los anuncios y otros factores que puedan influir en la percepción del valor por parte de los clientes.

Una vez obtenidos los datos, se llevará a cabo un análisis descriptivo para identificar patrones y tendencias en el mercado de alquiler de alojamientos turísticos en Málaga. Se prestará especial atención a la relación entre el precio y las características de los alojamientos, buscando determinar cuáles de estas características tienen un mayor impacto en la fijación de precios.

Posteriormente, se desarrollará un modelo econométrico que permita evaluar la adecuación de los precios de los anuncios en función de las características identificadas. Este modelo tendrá como objetivo ofrecer una herramienta práctica para los potenciales clientes, ayudándoles a determinar si el precio de un anuncio es justo o si, por el contrario, es excesivo.

En resumen, este trabajo pretende proporcionar una visión integral del mercado de alquiler de alojamientos turísticos no hoteleros en Málaga, aportando feedbacks valiosos sobre las características que más valoran los clientes y ofreciendo herramientas para la evaluación justa de los precios en este mercado dinámico y en constante crecimiento.



## Metodología Empleada

Este trabajo consiste en un análisis de una base de datos de AirBNB para el cual se han empleado las siguientes técnicas:

### Análisis descriptivo:

- Cálculo de la frecuencia de anuncios por barrio y presentación comparativa en forma de gráfico de burbuja.
- Histograma de precios por barrio
- Bloxplot de precios por barrio
- Media y varianza de las distintas variables por barrio
- Histograma de precios por barrio, con y sin limitaciones de propiedades a 200.

### Modelo de regresión:

- Creación de un modelo de regresión con las variables que aparecen en los anuncios de AirBNB para clientes con un conjunto de entrenamiento (Se detallará más a fondo en el siguiente apartado).
- Comparación con los valores del conjunto de test.

### Análisis Clúster

- Método del codo para determinación del n° óptimo de clusters.
- Análisis clúster del conjunto de datos.
- Obtención y representación de las componentes principales

### Modelo Logit:

- Creación de un modelo logit con las variables que aparecen en los anuncios de AirBNB para clientes con un conjunto de entrenamiento (Se detallará más a fondo en el siguiente apartado)
- Comparación con los valores del conjunto de test.
- Caracterización de la curva ROC del modelo.

## Desarrollo y resultados

Para este trabajo se ha comenzado extrayendo las variables más interesantes para estudiar de la base de datos, siendo estas:

- id: Identificador del alojamiento
- accommodates: N° de plazas disponibles en el alojamiento
- antigüedad: Antigüedad del anuncio
- host\_response\_rate: Tasa de respuesta del anunciante
- host\_response\_time: Tiempo medio del anunciante en responder
- neighborhood\_overview\_flag: Descripción del vecindario
- bathrooms: Número de baños completos (Los aseos cuentan como 0,5 unidades, mientras que los baños completos cuentan como 1 unidad).
- beds: N° de camas disponibles en el alojamiento
- price: Precio por noche en euros.
- host\_is\_superhost: Muestra si el anunciante entra en la categoría de superhost (Anunciantes con más de 4,8 de valoración, host response rate superior al 90% y tasa de cancelación inferior al 1%).
- host\_listings\_count: Número de alojamientos anunciados en Airbnb del mismo propietario
- host\_has\_profile\_pic: Disponibilidad de foto en el perfil del anunciante
- host\_identity\_verified: Verificación, o no, de la identidad del propietario.
- neighbourhood\_cleansed: Barrio en que se encuentra el alojamiento
- property\_type: Tipo de propiedad anunciada
- room\_type: Tipo de los cuartos de la propiedad
- minimum\_nights: Mínimo de noches de las que se puede disponer el alojamiento
- maximum\_nights: Máximo de noches de las que se puede disponer el alojamiento

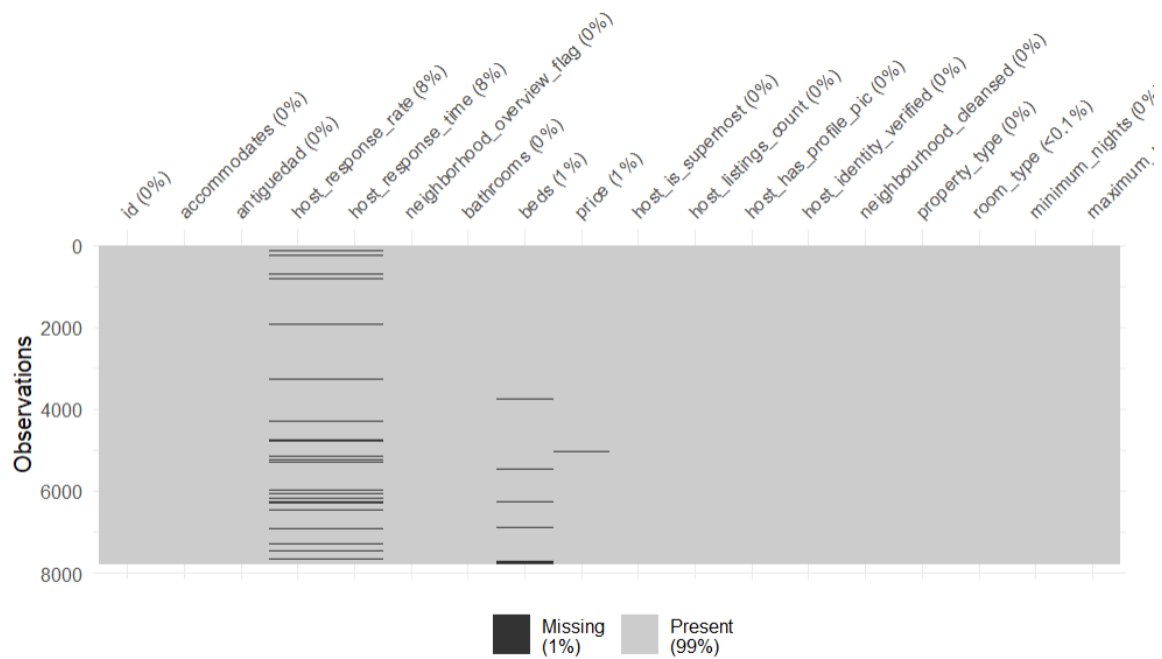
A priori, no era posible trabajar con todas las variables tal y como eran presentadas en la base de datos. Para ello primero fue necesario realizar algunas transformaciones en las mismas que se pueden encontrar detalladas en el Anexo I.

Una vez adecuadas las variables para poder trabajar con ellas, se procedió a la limpieza de los datos.

Se comenzó con la identificación y eliminación de los **missing values**.



## Visualización de Datos Faltantes



Fuente: Elaboración Propia

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 1

Es en las variables *host\_response\_rate* y *host\_response\_time* donde se dan la mayor parte de los missing values, siendo estos el 8% de los datos de cada una, y en las variables *beds* y *price* también se dan missing values, siendo estos un 1% de los datos de las mismas. El conjunto el total de missing values es un 1% de los datos.

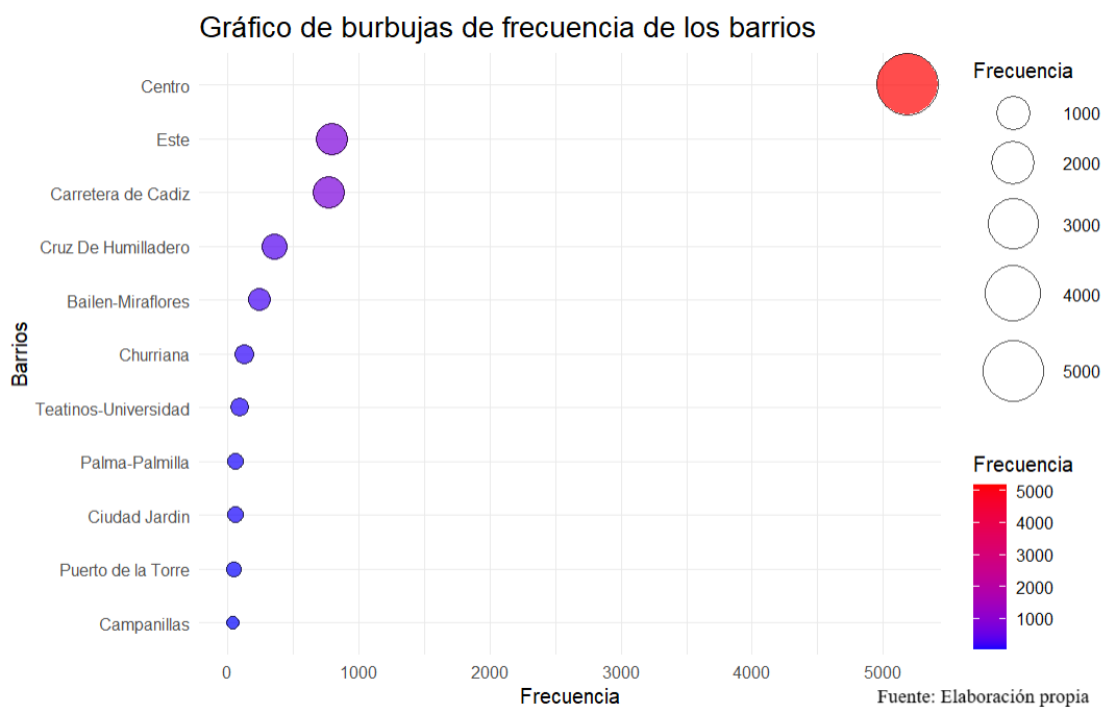
Las filas de los datos con estos missing values fueron eliminadas, con lo que la muestra de datos resultante ya era adecuada para comenzar a trabajar con ella. Esta depuración de los datos garantiza que el análisis de la oferta, así como el resto de análisis posteriores, se basen en información completa y precisa.

## Análisis de la oferta

Para comenzar, se quiere responder a la pregunta “¿Qué barrios son más atractivos turísticamente?”.

Para ello, el procedimiento que se sigue en este trabajo será estudiar en que barrios hay más anuncios, ya que, es en aquellos barrios con más demanda de alojamientos donde habrá mayor oferta de los mismo.

A continuación, se muestra un gráfico de burbuja en el que se representa el número de anuncios en los barrios de Málaga de forma que a mayor número de anuncios, mayor tamaño y mayor tendencia al rojo tendrá la burbuja representada.



Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 2

Representados todos los barrios, se aprecia una amplia diferencia del barrio **Centro** frente al resto.

Excepto **Centro**, todos los barrios tienen menos de 1000 anuncios, mientras que **Centro** supera los 5000. Esta diferencia ya de por sí es significativa, pero se desarrollará más a fondo a continuación, una vez presentados los resultados numéricos.

Tabla de frecuencia por barrios

barrio	frecuencia
Centro	5,189
Este	792
Carretera de Cadiz	774
Cruz De Humilladero	353
Bailen-Miraflores	245
Churriana	125
Teatinos-Universidad	92
Ciudad Jardin	64
Palma-Palmilla	64
Puerto de la Torre	46
Campanillas	39

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 1

El número de anuncios presente en **Centro** es de 5189, frente a 792 anuncios que hay presente en **Este**, siendo este el segundo barrio más poblado.

Visto desde otra perspectiva, de los 7783 anuncios de alojamientos turísticos no hosteleros presentes en la plataforma AirBNB, 5189 son solamente en el barrio **Centro**, por tanto, este barrio acapara el 66.67% de la oferta de anuncios presentes en toda la ciudad de Málaga.

El que la mayoría de la oferta se acumule en este barrio ya muestra que es el barrio más demandado de toda Málaga para los turistas que busquen una estancia en alojamientos turísticos no hosteleros en la ciudad.

Además, debe tenerse en cuenta que, al ser el barrio más demandado por una amplia diferencia, los precios de los alojamientos en él tenderán a ser más altos que en el resto, por lo que parece que puede suponerse con estas pistas que el barrio en que se encuentre

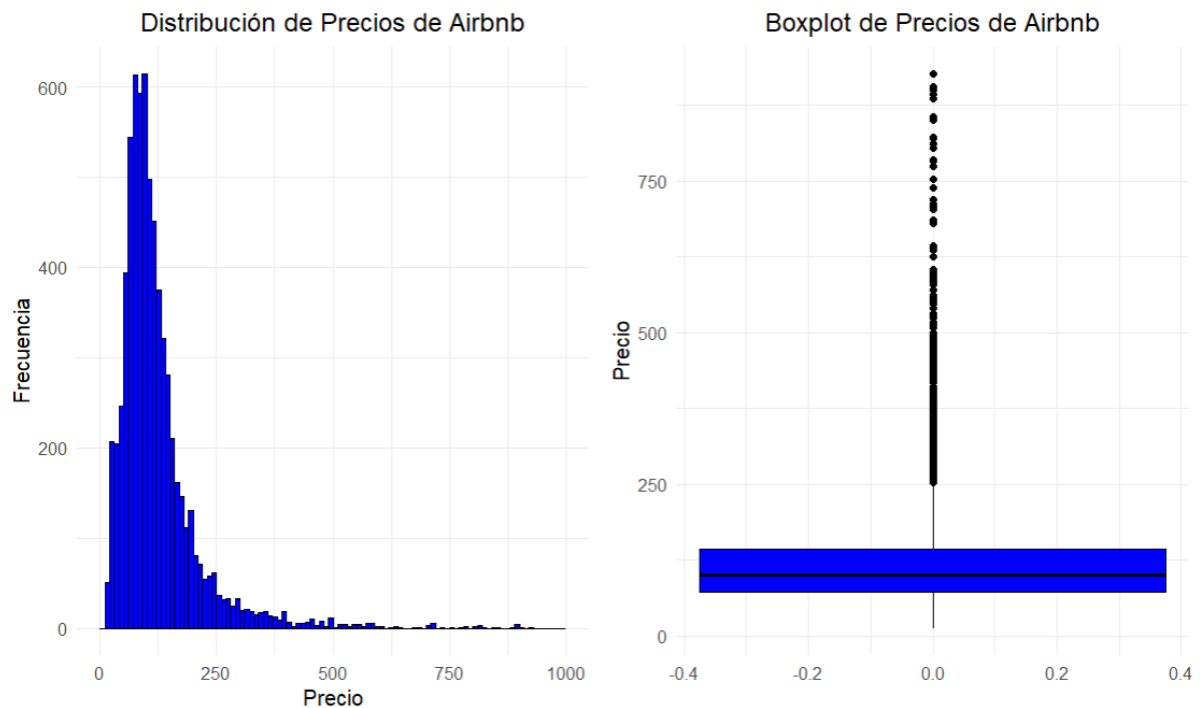
el alojamiento turístico no hostelero anunciado será uno de los factores importantes en la determinación del precio.

Teniendo en cuenta la concentración significativa de anuncios en el barrio Centro y su impacto potencial en los precios, en el siguiente apartado se va a mostrar el análisis realizado al conjunto de datos enfocado en el precio, para ver de forma más detallada lo que ocurre (Wachsmuth & Weisler, 2018).

## Análisis descriptivo

En este apartado, se va a mostrar el análisis realizado al conjunto de datos. La variable sobre la que se va a seguir el estudio en todo momento es el precio.

Para comenzar, se estudia la distribución del precio según la frecuencia de anuncios, así como el boxplot del precio considerando todos los casos (Cano, 2024):



Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

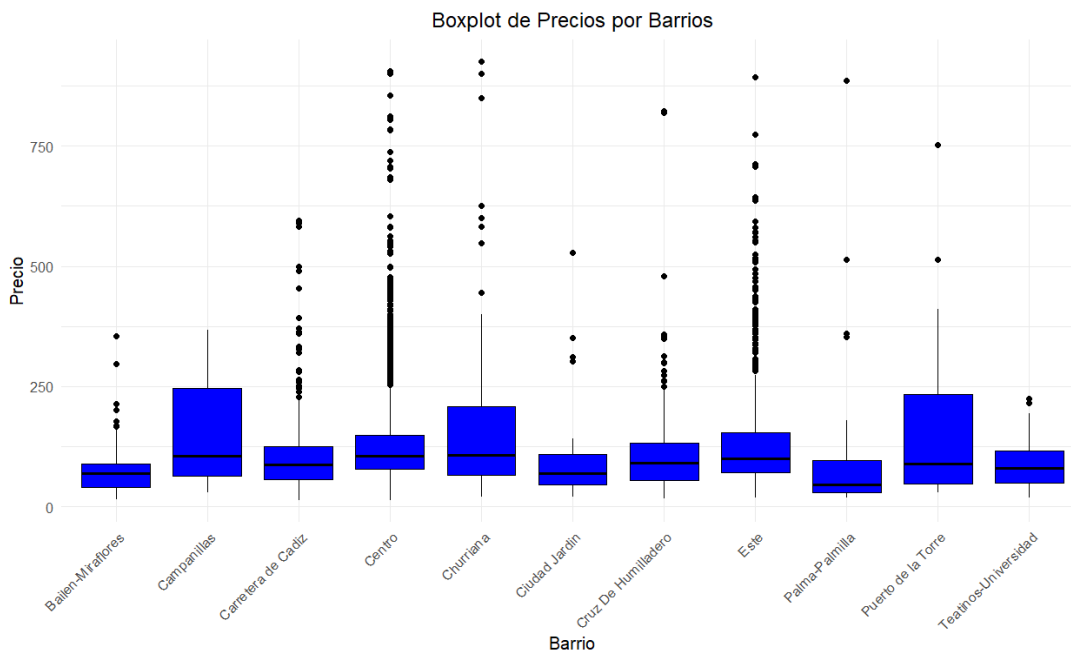
Figura 3

La distribución de precios permite apreciar que la mayor parte de los anuncios tienen precios inferiores a 250 euros por noche, siendo la concentración de los mismos significativamente menor a partir de los 250 euros por noche.

El gráfico boxplot permite precisar aún más, y ver que los outliers comienzan a partir de los 250 euros por noche.

Según Kimes y Wirtz (2015), la fijación de precios en el mercado de alojamientos tiende a presentar outliers cuando los precios superan significativamente el valor de mercado debido a varios factores, como la percepción de valor o servicios adicionales. En el caso que se está estudiando, y se verá en el resto del análisis más claramente, todos los outliers se dan por precio superior, es decir, en toda la muestra de datos, no existe ningún caso en el que un alojamiento se oferte por debajo del precio de mercado, sino que los casos que destacan son porque están por encima del mismo, por tanto, todo parece indicar que los precios de los alojamientos turísticos no hosteleros están por encima del precio de mercado en Málaga.

Caracterizando barrio a barrio, se obtiene (Cano, 2024):



Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 4

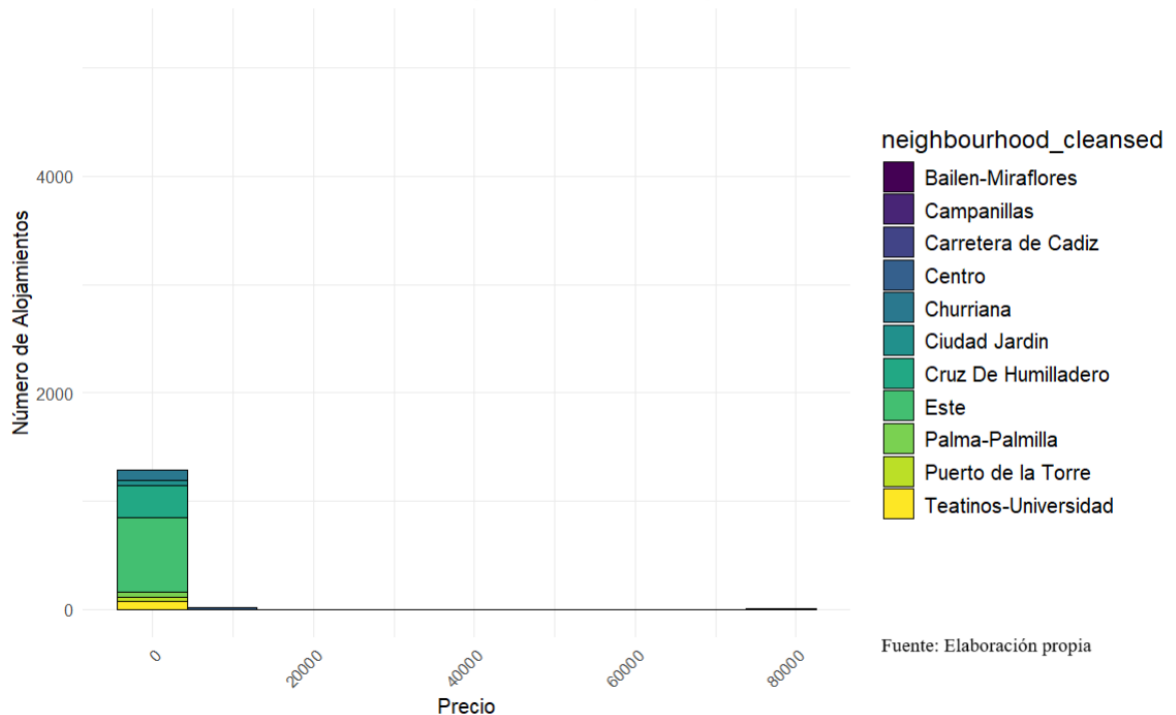
- Bailen-Miraflores: Tiene un rango de precios bajo, con algunos outliers destacables.
- Campanillas: Es el barrio con más variabilidad de precios dentro del rango intercuartílico recogido por el boxplot. Destaca por ser el único barrio sin outliers.
- Carretera de Cádiz: Rango de precios bajos, presencia de outliers hasta cerca de los 625 euros por noche.

- Centro: El rango intercuartílico no es de los más elevados, sin embargo, destaca por ser el que tiene una mayor presencia de outliers. Debe recordarse que este es el barrio con el mayor nº de anuncios con una amplia diferencia.
- Churriana: Rango intercuartílico amplio y pocos outliers en él.
- Ciudad Jardín: Su rango intercuartílico no es muy amplio, pero sus outliers, pese a ser pocos, tienen una amplia diferencia en el precio con respecto a sus precios normales.
- Cruz de Humilladero: Análogo a Ciudad Jardín
- Este: Rango intercuartílico destacable. Gran presencia de outliers que llegan a alcanzar el rango de precios más elevado.
- Palma-Palmilla: Rango de precios bajo. Sin embargo, destacan sus outliers, los cuales guardan una gran diferencia en los precios, llegando uno de ellos al rango más alto.
- Puerto de la Torre: Rango intercuartílico bastante amplio y con presencia de 2 outlier destacables, cuyos precios son altos.
- Teatinos-Universidad: Rango de precios no medio. Tiene 2 outliers, pero a diferencia del resto de barrios, estos no se alejan demasiado de su rango de precios.

En este gráfico es posible apreciar que, de acuerdo con lo expuesto por Kimes y Wirtz (2015), son los barrios **Centro** y **Este** los que están más por encima del precio de mercado.

Estos resultados, expuestos en forma de distribución, y calculando 10 intervalos entre el precio mínimo y el máximo dan (Cano, 2024):

Distribución del Precio por Barrio (Sin Filtrar)



Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 5

Visto así, no es posible extraer demasiada información. Para ello, en visos de una mejor apreciación, se ha extraído una muestra aleatoria de 200 alojamientos de aquellos barrios que superaban dicho número para apreciar mejor la distribución de los precios.

Dichos barrios son:

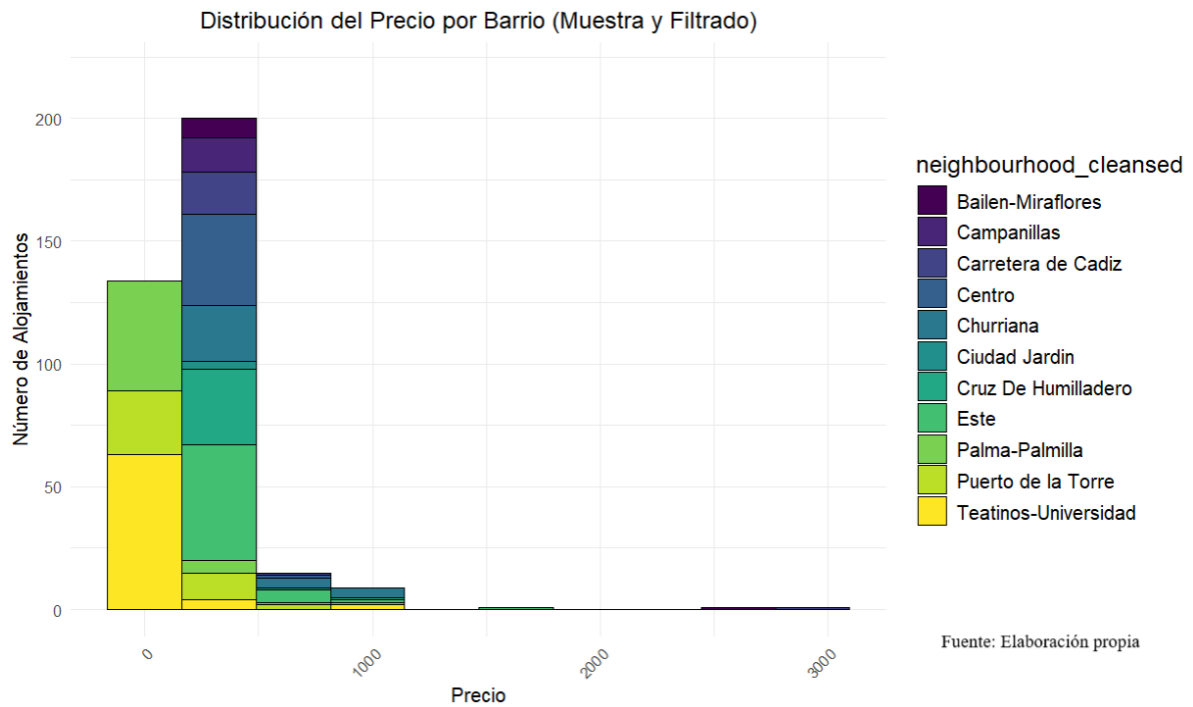
Tabla de barrios con más de 200 anuncios y excesos

Barrio	Nº de anuncios	Exceso sobre 200
Bailen-Miraflores	206	6
Carretera de Cádiz	652	452
Centro	4806	4606
Cruz de Humilladero	293	93
Este	690	490

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 2

La distribución ahora queda (Cano, 2024):

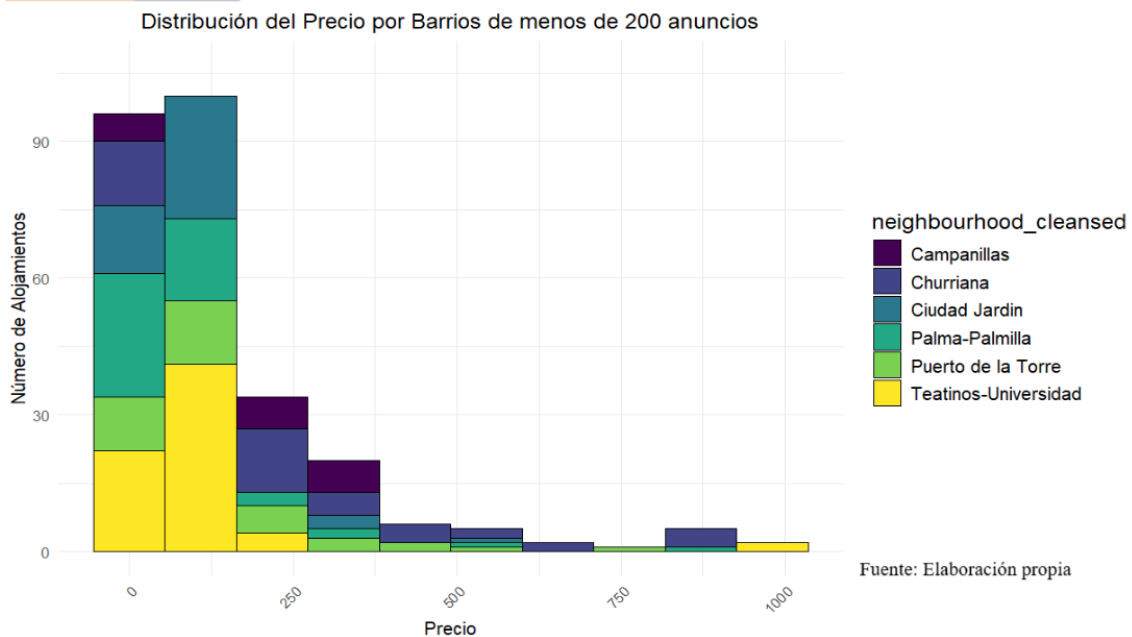


Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 6

Aun así, sigue siendo difícil sacar algunas conclusiones, por lo que, con el objetivo de poder comparar, se ha realizado también la distribución sin contar con los barrios de más de 200 propiedades, resultando (Cano, 2024):





Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 7

De la comparativa entre las figuras 4, 5 y 6 se extrae lo siguiente: Los barrios con mayor número de anuncios de propiedades son a su vez aquellos con los mayores precios por noche.

Esto indica que la demanda se concentra en ellos, por lo que hay más oferta, y al mismo tiempo que la demanda parece seguir siendo superior a la oferta en dichos barrios, **Centro** principalmente, al tener la mayor presencia de outliers y a su vez ser estos los de precios superiores. De nuevo, según Kimes y Wirtz (2015), estos van a ser los barrios cuyos precios estén más por encima del precio de mercado.

Con el análisis realizado, se han identificado los barrios más atractivos para los clientes (Y por ende donde aquellos que compran vivienda con fines turísticos). Una vez se tienen los barrios más interesantes, la pregunta a responder es: ¿Qué es lo que hace a un alojamiento interesante? Será en el siguiente apartado donde se intentará dar respuesta a esta pregunta.

## Modelo de regresión

Este trabajo está enfocado desde el punto de vista de un posible cliente, por lo que, las variables que se consideran para la construcción del modelo son aquellas que se pueden ver desde los anuncios de AirBNB. El objetivo no es tanto hallar un modelo de regresión (Cano, 2024) que explique el precio de los anuncios de los alojamientos turísticos no



# Trabajo Fin de Grado

máster universitario en modelización  
y análisis de datos económicos

hosteleros (Aunque esto se hará), sino encontrar cuales de estas variables son significativas para explicar los precios de los anuncios.

El modelo no será solo creado, sino que por supuesto también será testado. Para ello, primeramente, se limpiarán los outliers, mostrados en el apartado anterior, pasando el conjunto de datos de **6984** filas a **6073** filas (Un 13.04% de los datos).

A continuación, se dividirá el conjunto de datos entre **datos de entrenamiento** (80% de los datos) y **datos de prueba** (20% de los datos restantes).

Las variables a considerar para la creación del modelo de regresión son:

Variable explicada:

- Price

Variables explicativas:

- accommodates
- host\_response\_rate
- bathrooms
- beds
- minimum\_nights
- host\_has\_profile\_pic
- host\_identity\_verified
- property\_type
- room\_type
- maximum\_nights
- minimum\_nights
- neighbourhood\_cleansed

Estas son las variables que se pueden extraer de un alojamiento a partir de un anuncio de un alojamiento turístico no hostelero en AirBNB.

El modelo resultante, así como sus métricas de evaluación se encuentran en el archivo **Regresión\_model\_summary\_performance.txt**<sup>1</sup>.

---

<sup>1</sup> El archivo de este modelo desglosado junto a las métricas de evaluación se encuentra en el siguiente enlace: <https://github.com/Joseluisvic/Trabajo-Final-De-M-ster>

## Resumen del Modelo de Regresión Lineal

### Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	84.858361	8.431956	10.064	< 2e-16	***
accommodates	8.428982	0.613818	13.732	< 2e-16	***
host_response_rate	-8.489902	4.339938	-1.956	0.050496	.
bathrooms	22.966566	1.507984	15.230	< 2e-16	***
beds	2.241662	0.732291	3.061	0.002217	**
host_is_superhost	4.124859	1.201438	3.433	0.000601	***
host_has_profile_pic	5.964513	4.432752	1.346	0.178509	
host_identity_verified	-4.087940	4.163043	-0.982	0.326169	
property_type	0.612420	0.090260	6.785	1.30e-11	***
room_type	-45.902395	1.968049	-23.324	< 2e-16	***
minimum_nights	-5.068263	0.494107	-10.257	< 2e-16	***
maximum_nights	0.002662	0.001218	2.185	0.028912	*
neighbourhood_cleansedCampanillas	15.792626	9.830713	1.606	0.108239	
neighbourhood_cleansedCarretera de Cadiz	8.360414	3.592567	2.327	0.019999	*
neighbourhood_cleansedCentro	23.622185	3.215579	7.346	2.38e-13	***
neighbourhood_cleansedChurriana	10.227657	6.040234	1.693	0.090471	.
neighbourhood_cleansedCiudad Jardin	0.232186	7.105772	0.033	0.973935	
neighbourhood_cleansedCruz De Humilladero	9.399863	4.058368	2.316	0.020591	*
neighbourhood_cleansedEste	15.164686	3.619346	4.190	2.84e-05	***
neighbourhood_cleansedPalma-Palmilla	3.089663	7.291779	0.424	0.671790	
neighbourhood_cleansedPuerto de la Torre	-15.079387	8.870133	-1.700	0.089192	.
neighbourhood_cleansedTeatinos-Universidad	5.743922	6.818348	0.842	0.399594	

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.01 on 4836 degrees of freedom  
Multiple R-squared: 0.3946, Adjusted R-squared: 0.392  
F-statistic: 150.1 on 21 and 4836 DF, p-value: < 2.2e-16

Fuente: Elaboración propia

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 3

### Resumen del modelo:

- **Residual standard error (Error estándar residual):** El valor del error estándar residual de 37.01 indica la desviación estándar de los residuos, es decir, la diferencia entre los valores observados y los valores predichos por el modelo. Este valor nos da una idea de qué tan bien el modelo está ajustado a los datos; en este caso, una desviación promedio de 37.01 unidades en el precio sugiere que las predicciones del modelo pueden estar fuera de los valores reales en esta magnitud. Aunque no es el único indicador de ajuste, es útil para medir la precisión del modelo.

- **Multiple  $R^2$  (R cuadrado múltiple): 0.3946** El valor de  $R^2$  de 0.3946 indica que el 39.46% de la variabilidad en el precio puede ser explicada por las variables independientes incluidas en el modelo. Este valor nos da una idea del poder explicativo del modelo: cuanto más alto sea  $R^2$ , mayor es la proporción de variabilidad en la variable dependiente (precio) que está siendo capturada por el modelo. En este caso, el 39.46% de la variación en el precio es explicada por las variables del modelo, lo que significa que aún hay una porción significativa de variabilidad (el 60.54%) que no está explicada.
- **Adjusted  $R^2$  (R cuadrado ajustado): 0.392** El  $R^2$  ajustado es una versión modificada de  $R^2$  que tiene en cuenta el número de variables del modelo. En este caso, el valor de 0.392, o 39.2%, muestra cuánta variabilidad en el precio se explica después de ajustar por el número de variables incluidas. A diferencia del  $R^2$  simple, que puede aumentar al agregar variables incluso si no son relevantes, el  $R^2$  ajustado penaliza la inclusión de variables no significativas. Aquí, la diferencia entre el  $R^2$  simple (39.46%) y el  $R^2$  ajustado (39.2%) es pequeña, lo que sugiere que las variables adicionales no están generando sobreajuste y aportan valor al modelo.
- **F-statistic: 150.1 y p-valor asociado  $< 2.2e-16$**  El estadístico F de 150.1 evalúa si el modelo en su conjunto tiene un poder explicativo significativo. El valor de F alto indica que el conjunto de variables independientes tiene un impacto significativo en la variable dependiente (*price*). Además, el p-valor asociado es extremadamente pequeño ( $< 2.2e-16$ ), lo que nos dice que la probabilidad de que estas variables no tengan un efecto significativo en la variable dependiente es prácticamente nula. Por tanto, podemos concluir que el modelo es estadísticamente significativo en su conjunto, y que al menos una de las variables explicativas tiene un efecto relevante sobre el precio.

Tomando un 95% de confianza, las variables significativas devueltas por el modelo son:

- accommodates
- host\_response\_time
- bathrooms
- beds
- host\_is\_superhost
- property\_type
- room\_type
- maximum\_nights
- minimum\_nights

- neighbourhood\_cleansedCarretera de Cádiz
- neighbourhood\_cleansedCentro
- neighbourhood\_cleansedCruz de Humilladero
- neighbourhood\_cleansedEste

Mirando más en detalle las variables significativas y prestando atención a sus coeficientes:

- *accommodates*, *bathrooms*, *beds*, *host\_is\_superhost*, *property\_type*, *maximum\_nights*, *neighbourhood\_cleansedCarretera de Cádiz*, *neighbourhood\_cleansedCentro*, *neighbourhood\_cleansedCruz de Humilladero*, *neighbourhood\_cleansedEste* tienen coeficientes positivos, por lo que tienen un efecto creciente en el precio por noche, sobre todo *bathrooms*, que es la que tiene un mayor coeficiente, siendo este 22.9666.
- *host\_response\_rate*, *room\_type*, *minimum\_nights* tienen coeficientes negativos, por lo que tienen un efecto decreciente en el precio por noche, sobre todo *room\_type*, con un coeficiente de -45.9204, que es la que tiene mayor coeficiente. Destacar que no es solo la variable con mayor coeficiente negativo, sino que en valor absoluto es la variable con el mayor coeficiente, por lo que es la variable con más peso de todas.

A este punto del trabajo, ya se tienen los barrios más interesantes desde el punto de vista de clientes y por compradores de vivienda con fines turísticos, y las variables significativas en la determinación del precio por noche en base a las variables que pueden observarse en un anuncio de alojamiento en AirBNB. El siguiente paso natural es preguntarse si existen patrones y/o agrupaciones dentro de los datos que permitan extraer aún más información, lo que permitiría su vez una mejor comprensión de las diferentes ofertas disponibles y facilitando la toma de decisiones para los posibles clientes, y como consecuencia, para aquellos que invierten en vivienda con fines turísticos (Aunque el objetivo de este trabajo es analizar desde la perspectiva el cliente). Para ello, se procede a realizar un análisis clúster de los datos.

## Análisis Clúster

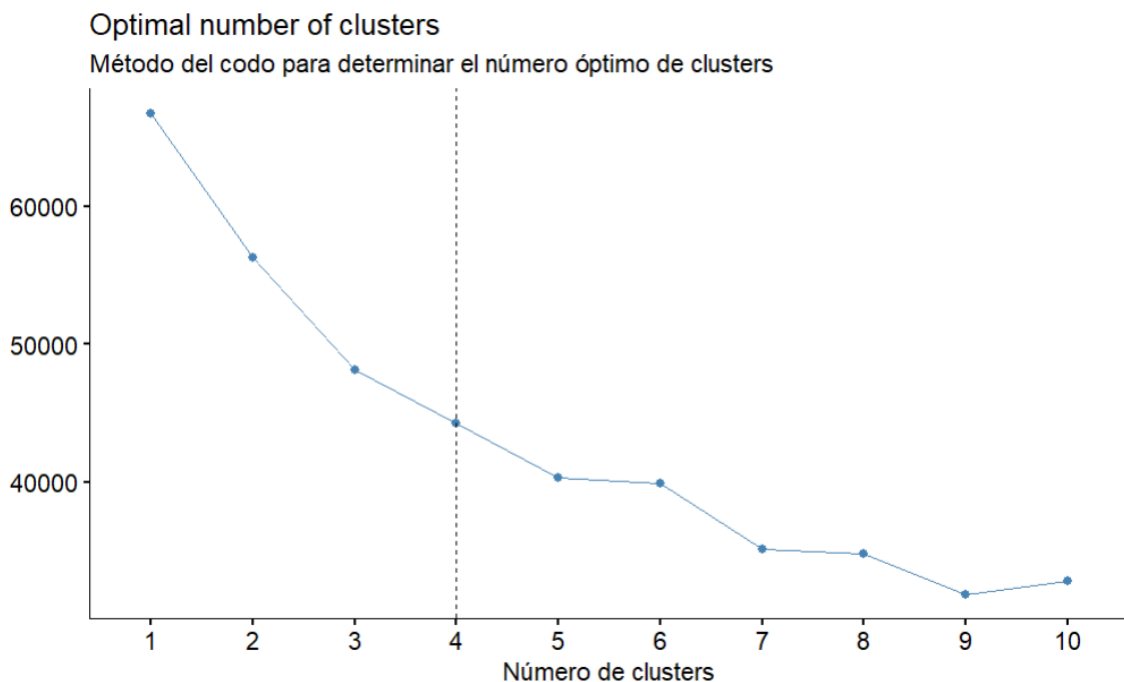
Una vez calculado el modelo y determinadas las variables significativas, se utiliza esta información para realizar un análisis de clúster (Tarancón Morán & Quintana Rojo, 2023-2024) en el conjunto de datos. El objetivo de este análisis es investigar si el barrio donde se encuentra un anuncio afecta de manera significativa la distribución de los datos según su ubicación. Al comprobarse si existe una distinción significativa (o no) que pueda

determinarse y diferenciarse por el barrio de localización de un determinado alojamiento turístico no hostelero, será posible obtener conclusiones del impacto (o no) de este efecto.

Se añade la columna de *price* al conjunto tomado de variables significativas para tenerlo en cuenta para el análisis clúster, ya que es una variable de mucha importancia en el análisis.

También, antes de la realización de la división por clústers, se procede a normalizar las variables para poder trabajar mejor con ellas (Tarancón Morán & Quintana Rojo, 2023-2024).

Para determinar el número óptimo de clúster en que dividir el conjunto, se realiza el **Método del codo**, al no conocerse la cantidad de óptima de centroides  $k$  a utilizar de antemano (Delgado, 2018), por lo que este método es el más adecuado para determinar dicho valor  $k$ .



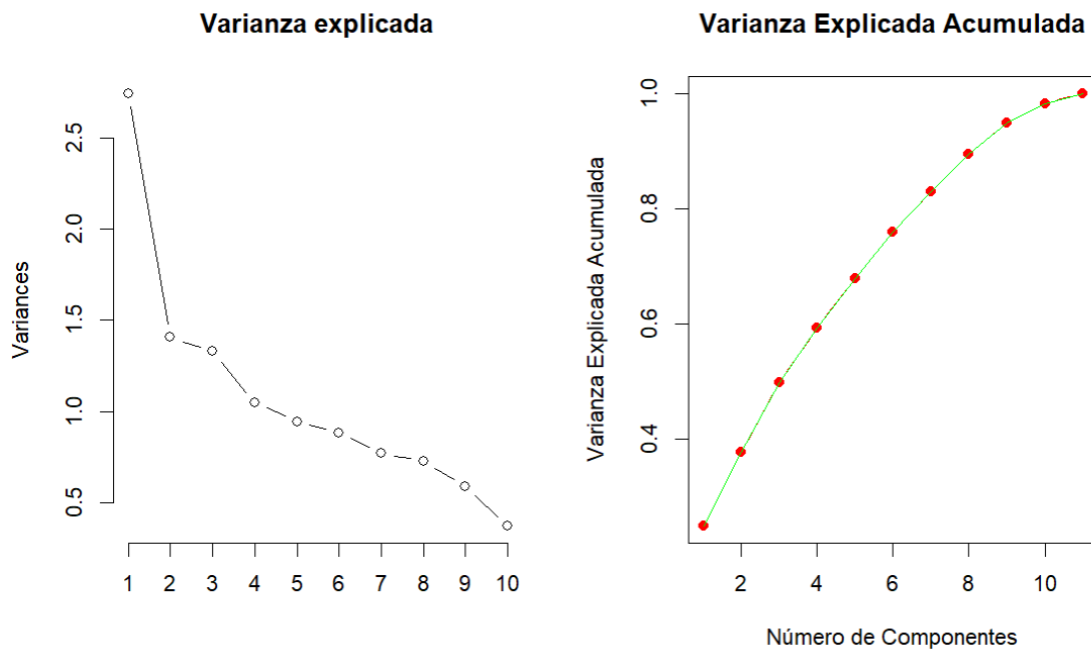
Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 8

El método del Codo indica que la distribución óptima en clusters es de 1 clúster (Al tener el barrio **Centro** la mayoría de los anuncios es probable que el peso de estos provoque que el número óptimo sea 1). Sin embargo, a efectos de estudiar cómo reaccionan los

datos ante divisiones, y si la misma puede proveer alguna idea interesante al estudio, se van a realizar 4 clusters.

Una vez realizado, se introducen los resultados de los clusters a los datos y se procede a calcular las componentes principales ellas (Tarancón Morán & Quintana Rojo, 2023-2024), obteniéndose los siguientes resultados del análisis:



Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 9

Siendo los valores numéricos:

Porcentaje de varianza explicado por cada componente principal:

CP1: 24.97%  
CP2: 12.78%  
CP3: 12.07%  
CP4: 9.53%  
CP5: 8.57%  
CP6: 8.03%  
CP7: 6.98%  
CP8: 6.60%  
CP9: 5.36%  
CP10: 3.38%  
CP11: 1.72%

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 4

De estos resultados se extraen las siguientes conclusiones:

- La primera componente principal, **CP1**, captura el 24.97% de la varianza total de los datos, lo que implica que casi una cuarta parte de la información de los datos originales se encuentra representada por ella.
- La segunda componente principal, **CP2**, captura el 12.78% de la varianza. Sumando al acumulado (**CP1** solamente en este punto), capturan el 37.75% de la varianza total. Estas dos componentes ya explican por sí solas una parte significativa de la variabilidad.
- La tercera componente, **CP3**, representa el 12.07%, y la cuarta componente, **CP4**, el 9.53%. Cada componente sucesiva va capturando cada vez menos varianza. En este caso, las cuatro primeras componentes principales explican el 59.35% de la variabilidad total de los datos, lo que implica que estos ya se pueden representar razonablemente bien con ellas.
- El resto de componentes explican fracciones progresivamente más pequeñas y, por ende, menos importantes. Para **CP10** y **CP11**, la varianza explicada es muy pequeña (3.38% y 1.72% respectivamente), lo que indica que apenas aportan información adicional sobre la variabilidad de los datos.

Ya que se ha concluido que las dos primeras componentes principales explican por sí solas una parte significativa de la variabilidad total de los datos, se procederá a continuar el análisis con ellas:



## Cargas de las variables originales en las componentes principales:

### Cargas para CP1:

```
accommodates: -0.5257
host_response_time: 0.0385
bathrooms: -0.2816
beds: -0.4784
host_is_superhost: 0.0194
property_type: 0.2177
room_type: 0.3779
maximum_nights: -0.1229
minimum_nights: 0.0114
neighbourhood_cleansed: 0.1075
price: -0.4435
```

### Cargas para CP2:

```
accommodates: -0.1830
host_response_time: -0.3081
bathrooms: -0.4282
beds: -0.2920
host_is_superhost: 0.0186
property_type: -0.3874
room_type: -0.4315
maximum_nights: 0.2197
minimum_nights: -0.0727
neighbourhood_cleansed: -0.4590
price: 0.0459
```

2

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 5

La explicación de los resultados de las cargas de la componente principal 1 (CP1) es:

- **accommodates (-0.5257):** Esta variable tiene una carga negativa fuerte. Significa que a medida que la capacidad de alojamiento aumenta, el valor de la CP1 disminuye significativamente.
- **host\_response\_time (0.0385):** Esta variable tiene una carga positiva muy pequeña, lo que indica que tiene un impacto mínimo en la CP1.
- **bathrooms (-0.2816):** La cantidad de baños tiene una carga negativa moderada, contribuyendo a una disminución de la CP1.
- **beds (-0.4784):** Similar a la variable de alojamiento, un mayor número de camas reduce considerablemente el valor de la CP1.

<sup>2</sup> Las cargas para el resto de componentes pueden encontrarse en el documento *resultados\_pca.txt* disponible en <https://github.com/Joseluisvic/Trabajo-Final-De-M-ster>

- **host\_is\_superhost (0.0194):** Ser un superanfitrión tiene una carga positiva muy baja, indicando una contribución casi insignificante a la CP1.
- **property\_type (0.2177):** El tipo de propiedad tiene una carga positiva moderada, lo que sugiere que diferentes tipos de propiedad aumentan el valor de la CP1.
- **room\_type (0.3779):** El tipo de habitación tiene una carga positiva, indicando que esta variable aumenta el valor de la CP1.
- **maximum\_nights (-0.1229):** El número máximo de noches tiene una carga negativa pequeña, sugiriendo una leve disminución en la CP1 con un mayor número de noches.
- **minimum\_nights (0.0114):** El número mínimo de noches tiene una carga positiva muy baja, indicando un impacto casi nulo en la CP1.
- **neighbourhood\_cleansed (0.1075):** La variable que representa el vecindario tiene una carga positiva pequeña, sugiriendo una ligera influencia en el aumento de la CP1.
- **price (-0.4435):** El precio tiene una carga negativa considerable, indicando que precios más altos tienden a reducir el valor de la CP1.

Y de la componente principal 2 (CP2):

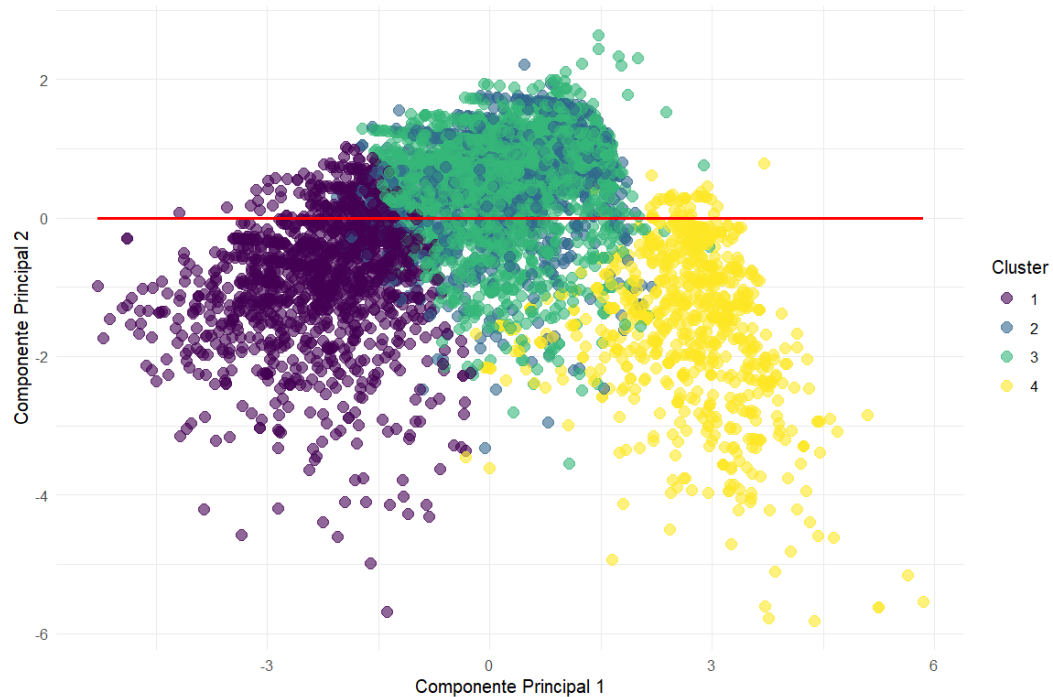
- **accommodates (-0.1830):** Esta variable tiene una carga negativa pequeña, lo que sugiere que a medida que la capacidad de alojamiento aumenta, el valor de la CP2 disminuye ligeramente.
- **host\_response\_time (-0.3081):** El tiempo de respuesta del anfitrión tiene una carga negativa moderada, indicando que tiempos de respuesta más largos reducen el valor de la CP2.
- **bathrooms (-0.4282):** El número de baños tiene una carga negativa significativa, lo que indica que más baños están asociados con una disminución considerable en la CP2.
- **beds (-0.2920):** El número de camas tiene una carga negativa moderada, sugiriendo que más camas reducen el valor de la CP2.
- **host\_is\_superhost (0.0186):** Ser un superanfitrión tiene una carga positiva muy baja, indicando una contribución insignificante a la CP2.
- **property\_type (-0.3874):** El tipo de propiedad tiene una carga negativa considerable, indicando que ciertos tipos de propiedad reducen el valor de la CP2.
- **room\_type (-0.4315):** El tipo de habitación tiene una carga negativa significativa, lo que sugiere que ciertos tipos de habitaciones disminuyen el valor de la CP2.
- **maximum\_nights (0.2197):** El número máximo de noches tiene una carga positiva pequeña, indicando que un mayor número de noches aumenta ligeramente el valor de la CP2.

- **minimum\_nights (-0.0727):** El número mínimo de noches tiene una carga negativa muy pequeña, sugiriendo un impacto leve en la disminución de la CP2.
- **neighbourhood\_cleansed (-0.4590):** El vecindario tiene una carga negativa significativa, indicando que ciertos vecindarios reducen considerablemente el valor de la CP2.
- **price (0.0459):** El precio tiene una carga positiva muy baja, indicando una contribución mínima a la CP2.

En resumen, las variables que más influyen en la primera componente principal son "*accommodates*" y "*beds*" (negativamente), así como "*price*" (también negativamente). Esto sugiere que la **CP1** puede estar representando un factor que combina elementos de capacidad y costo, donde una mayor capacidad de alojamiento y precio más alto se asocian con valores más bajos de **CP1**. En cambio, variables como el tipo de propiedad y el tipo de habitación tienen una influencia positiva moderada, sugiriendo que diferentes tipos de propiedades y habitaciones están asociados con valores más altos de **CP1**, mientras que las variables que más influyen en la segunda componente principal (**CP2**) son "*bathrooms*", "*room\_type*", y "*neighbourhood\_cleansed*" (todas negativamente). Esto sugiere que la **CP2** puede estar representando un factor relacionado con la calidad o el tipo de alojamiento y la ubicación, donde un mayor número de baños, ciertos tipos de habitaciones y ciertos vecindarios están asociados con valores más bajos de **CP2**. Por otro lado, variables como el número máximo de noches y el precio tienen una influencia positiva muy pequeña, sugiriendo un impacto mínimo en el aumento de la **CP2**.

Si se representan estas 2 primeras componentes principales, representando los clusters se tiene:

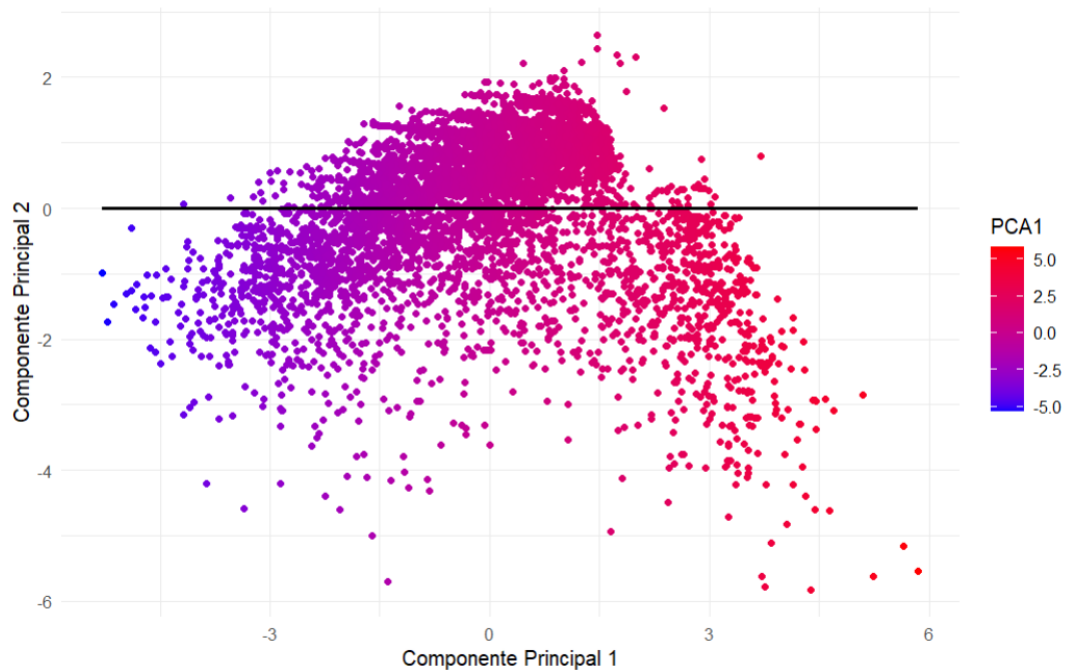
Datos según Componentes Principales



Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 10

Gráfico de las dos primeras componentes principales



Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 11

La representación de las dos componentes principales diferenciando por cluster aparece de esta forma. Se aprecian con cierta facilidad dos clusters, uno en la parte derecha de los gráficos y otro en la izquierda, habiendo 2 clusters en la parte central superior de los mismos superpuestos difíciles de percibir de forma visual.

La independencia lineal de las componentes principales es una verificación clave en el análisis de componentes principales, que puede observarse mediante la horizontalidad de la línea de tendencia entre las componentes (Jolliffe & Cadima, 2016). Simplemente a modo de comprobación extra de que el proceso de creación de las componentes principales se ha realizado correctamente, se les ha añadido dicha línea, cuya horizontalidad indica que no existe relación lineal significativa entre la componente 1 y la componente 2, por lo que no son dependientes la una de la otra.

Ya que las variables más importantes de las variables con más peso en las componentes principales afectan todas negativamente, implica que aquellas con más mayor precio, pero a su vez mayor valores del resto (*accommodates*, *beds*, *bathrooms*, *room\_type*, *neighbourhood\_cleansed*), es decir, los alojamientos más equilibrados, serán los que se encuentren en la zona más cercana al origen en las figuras 8 y 9, por tanto, los alojamientos pertenecientes al primer clúster.

Aquellos alojamientos con mayor valor de la componente 2 y valor medio de la componente 1 se encuentran en los clústeres 2 y 3, es decir, serán aquellos con menor número de baños y peor tipo de habitación. También es posible que a su vez implique que el alojamiento se encuentre en **Centro**, al tener este barrio un valor bajo asociado y ser aquel con mayor oferta.

Finalmente, los alojamientos con altos valores en la componente 1 y valores bajos en la 2 serían aquellos recogidos en el clúster 4, es decir, aquellos con un número de baños considerable, tipo de habitación bueno y, probablemente no perteneciente a **Centro**. A su vez, serían alojamientos con pocos alojamientos y/o pocas camas. No serían alojamientos con precios bajos y el resto de características porque la tendencia que mostraban los datos por barrio en el análisis descriptivo inicial, sobre todo los outliers, es que los casos excepcionales se daban siempre por precios superiores o muy superiores. Se debe recordar como justificación a esto que no había ni un solo caso de outlier en el precio por barrio por precio bajo. Todos eran por exceso.

Tras todo este análisis, ya se conocen las características del mercado de alquiler de alojamientos turísticos no hosteleros en Málaga, los barrios más interesantes para los clientes, los determinantes del precio por noche y se tienen nociones de algunos patrones

en los datos que ayudan a hacer agrupaciones que facilitan la toma de decisiones. Con todo esto, ya solo queda responder a la pregunta que se planteaba en el apartado

**Definición del objetivo planteado:** “¿Es el precio de un determinado anuncio adecuado o excesivo?”

Con ese objetivo, se desarrolla el siguiente y último apartado de esta parte del trabajo.

## Modelo logit

Teniendo en cuenta que el modelo logit es un modelo de elección binaria (Cameron & Trivedi, 2005), primero hay que definir qué elección se quiere estudiar. Para el caso de este estudio, se va a crear una variable, que se llamará *precio\_alto*, la cual se define en este estudio como 1.5 veces la mediana de *price*.

La idea tras este modelo es, en base a las características que se pueden encontrar de cierto alojamiento turístico no hostelero en el anuncio de AirBNB, responder a la “¿Es el precio de un determinado anuncio adecuado o excesivo?”<sup>3</sup>.

Al igual que con el modelo de regresión calculado anteriormente, se divide el conjunto de datos en **datos de entrenamiento** (80%) y **datos de prueba** (20%), realizándose el modelo con los datos de entrenamiento y el testeo del mismo con los datos de prueba.<sup>4</sup>

---

<sup>3</sup> Se considera solo el caso de que haya exceso porque se ha apreciado en el primer apartado que todos los outliers son por exceso, y ello parece indicar que la tendencia del mercado es al exceso y no al defecto con relación a los precios por noche de los alojamientos.

<sup>4</sup> El archivo de este modelo desglosado junto a las métricas de evaluación se encuentran con los nombres **logit\_model\_evaluation.txt** y **logit\_model\_summary.txt** en el siguiente enlace <https://github.com/Joseluisvic/Trabajo-Final-De-M-ster>

## Resumen del Modelo Logit

-----

### Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-8.507e-01	6.662e-01	-1.277	0.201592	
accommodates	4.058e-01	4.656e-02	8.717	< 2e-16	***
host_response_rate	-7.207e-02	3.773e-01	-0.191	0.848529	
bathrooms	1.276e+00	1.009e-01	12.641	< 2e-16	***
beds	3.649e-03	5.290e-02	0.069	0.945013	
host_is_superhost	1.297e-01	9.440e-02	1.374	0.169575	
host_has_profile_pic	3.504e-01	3.738e-01	0.937	0.348553	
host_identity_verified	-1.063e+00	2.981e-01	-3.565	0.000364	***
property_type	2.033e-02	8.696e-03	2.338	0.019389	*
room_type	-2.028e+00	2.983e-01	-6.798	1.06e-11	***
minimum_nights	-3.487e-01	4.272e-02	-8.163	3.27e-16	***
maximum_nights	5.160e-05	9.467e-05	0.545	0.585707	
neighbourhood_cleansed	-2.051e-01	3.407e-02	-6.020	1.75e-09	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4668.5 on 4857 degrees of freedom  
Residual deviance: 3781.6 on 4845 degrees of freedom  
AIC: 3807.6

Number of Fisher Scoring iterations: 6

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 6

Tomando un 95% de confianza, las variables significativas devueltas por el modelo son:

- accommodates
- bathrooms
- property\_type
- host\_identity\_verified
- room\_type
- minimum\_nights
- neighbourhood\_cleansed

Mirando más en detalle las variables significativas y prestando atención a sus coeficientes:

- *accommodates*, *bathrooms*, *property\_type* tienen coeficientes positivos, por lo que, a mayor número de plazas disponibles, baños o el tipo de vivienda, mayor es el impacto positivo en la probabilidad de éxito. De estos, *bathrooms* es la variable con el mayor peso, con mucha diferencia con respecto a las demás, con un coeficiente de 1.276.
- *host\_identity\_verified*, *room\_type*, *minimum\_nights* tienen coeficientes negativos, por lo que, en caso de que el anunciante tenga su identidad verificada, el tipo de habitación y a mayor número de noches mínimas, menor será la probabilidad de éxito. De estas, *room\_type* es la variable con mayor peso. Al igual que con el modelo de regresión calculado anteriormente, *room\_type* no es solo la variable con mayor coeficiente negativo, sino que es la variable que, en valor absoluto, tiene el mayor valor de coeficiente, siendo este -2.028, indicando que es la variable con mayor peso de todas.

Además

- Deviance Residual: 3781.6. < Null Deviance: 4668.5. El que sea menor la residual que la nula implica que el modelo con los predictores proporciona una mejor representación de los datos que un modelo sin predictores, y además indica una mejora en el ajuste del modelo por la inclusión de las variables independientes.
- AIC:3807.6

Pero, ¿Es este modelo correcto para realizar identificaciones de los casos en qué el precio es alto? Para comprobar esto, se calculan las métricas de rendimiento y la curva de ROC del modelo y su área entre la curva y la línea de 45° (del Corral, 2023-2024).



Métricas del conjunto de prueba:

Exactitud: 0.8312757

Sensibilidad: 0.2757009

Especificidad: 0.95005

Métricas del conjunto de entrenamiento:

Exactitud: 0.8272952

Sensibilidad: 0.2422566

Especificidad: 0.9610521

Tabla de confusión del conjunto de prueba:

	Actual	
Predicted	0	1
0	951	155
1	50	59

Tabla de confusión del conjunto de entrenamiento:

	Actual	
Predicted	0	1
0	3800	685
1	154	219

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 7

Métricas del conjunto de prueba:

- **Exactitud (0.8312757):** La exactitud mide el porcentaje de predicciones correctas sobre el total de predicciones realizadas. En este caso, el modelo tuvo una exactitud del 83.13%, lo que significa que, de todas las predicciones en el conjunto de prueba, el 83.13% fueron correctas.
- **Sensibilidad (0.2757009):** La sensibilidad mide la capacidad del modelo para identificar correctamente los casos positivos (clase 1). Aquí, el modelo identificó correctamente el 27.57% de los casos positivos reales. Esto sugiere que el modelo tiene dificultades para detectar correctamente las observaciones de la clase 1 (positivas).
- **Especificidad (0.95005):** La especificidad mide la capacidad del modelo para identificar correctamente los casos negativos (clase 0). Una especificidad del 95% indica que el modelo es muy bueno para identificar los casos negativos (es decir, clase 0). Predice correctamente la mayoría de los casos negativos.

Métricas del conjunto de entrenamiento:

- **Exactitud (0.8272952):** La exactitud del conjunto de entrenamiento es similar a la del conjunto de prueba, con un 82.73%. Esto indica que el modelo generaliza de manera razonablemente consistente entre el entrenamiento y la prueba, aunque hay una ligera variación.
- **Sensibilidad (0.2422566):** La sensibilidad es aún más baja en el conjunto de entrenamiento, con solo un 24.22% de los positivos identificados correctamente. Al igual que en el conjunto de prueba, el modelo lucha para detectar los positivos.
- **Especificidad (0.9610521):** La especificidad es ligeramente superior en el conjunto de entrenamiento (96.11%), lo que nuevamente indica que el modelo clasifica muy bien los casos negativos en ambas fases (entrenamiento y prueba).

Tabla de confusión del conjunto de prueba:

Tabla de confusión del conjunto de prueba:		
Actual		
Predicted	0	1
0	951	155
1	50	59

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 8

- **951 verdaderos negativos (TN):** El modelo predijo 0 (negativo) y la clase real era 0.
- **59 verdaderos positivos (TP):** El modelo predijo 1 (positivo) y la clase real era 1.
- **50 falsos positivos (FP):** El modelo predijo 1 (positivo) pero la clase real era 0.
- **155 falsos negativos (FN):** El modelo predijo 0 (negativo) pero la clase real era 1.

Tabla de confusión del conjunto de entrenamiento:

Tabla de confusión del conjunto de entrenamiento:		
	Actual	
Predicted	0	1
0	3800	685
1	154	219

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 9

- **3800 verdaderos negativos (TN):** El modelo predijo 0 y la clase real era 0.
- **219 verdaderos positivos (TP):** El modelo predijo 1 y la clase real era 1.
- **154 falsos positivos (FP):** El modelo predijo 1 pero la clase real era 0.
- **685 falsos negativos (FN):** El modelo predijo 0 pero la clase real era 1.

Análisis General de los resultados

➤ **Desequilibrio entre sensibilidad y especificidad:**

Alto desequilibrio entre sensibilidad y especificidad: Aunque la especificidad es muy alta en ambos conjuntos (entrenamiento y prueba), lo que significa que el modelo es muy bueno en predecir los casos negativos (clase 0), la sensibilidad es bastante baja. Esto indica que el modelo no detecta bien los positivos (clase 1), por lo que pueden darse casos de que efectivamente el precio del alojamiento turístico no hostelero estudiado si sea de “Precio alto” y el modelo no lo recoja adecuadamente (Provost & Fawcett, 2001).

➤ **Posible desbalance de clases:**

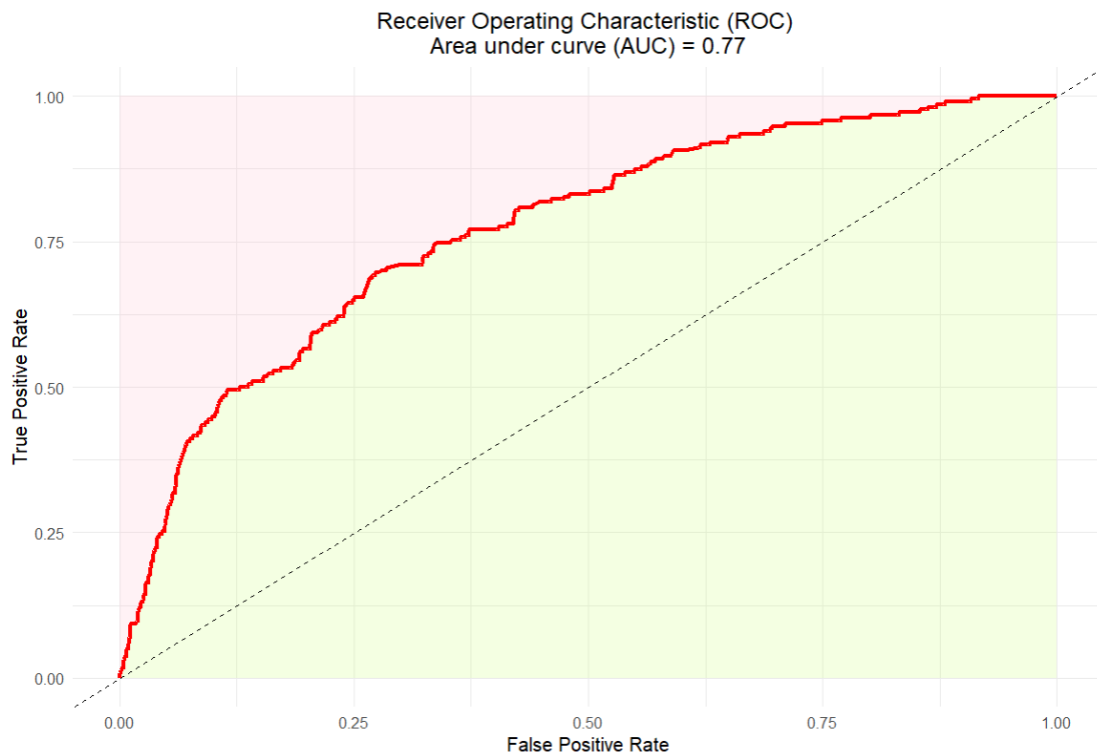
Posible desbalance de clases: La baja sensibilidad sugiere que podría haber un desequilibrio en las clases (más ejemplos de la clase 0 que de la clase 1), lo que podría llevar al modelo a priorizar la predicción de la clase mayoritaria (clase 0) y subestimar la importancia de la clase minoritaria (clase 1), dando más peso por tanto a alojamientos que no son de “Precio alto” (Provost & Fawcett, 2001).

## ➤ Modelo conservador en predicciones positivas:

Modelo conservador en predicciones positivas: El modelo parece ser conservador en cuanto a predecir la clase 1, pues tiene más falsos negativos que verdaderos positivos en ambos conjuntos (Chawla et al., 2004).

A pesar de esto, la capacidad de distinción del modelo entre los casos de “Precio alto” y “No precio alto” es razonablemente buena.

Ello lo muestra el cálculo de su curva ROC, y del área bajo la misma:



Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Figura 12

En este caso se da que el área bajo la curva ROC es de 0.77, encontrándose en el rango que se considera entre Rendimiento aceptable a Bueno (Fawcett, 2006), indicando que el modelo dispone de cierta capacidad de discriminación correcta.

Con todos estos análisis hasta aquí realizados, se proceden a exponer las conclusiones finales del trabajo en el siguiente apartado.

## Conclusiones

En cuanto al mercado del alquiler de alojamientos turísticos no hosteleros en la ciudad de Málaga, la mayor oferta se da en el barrio **Centro**, el cual acumula la mayor parte de los anuncios.

El rango más común de precios de los anuncios de alojamientos turísticos no hosteleros en Málaga se da en el intervalo  $]0, 250]$  euros por noche.

El precio de los anuncios viene determinado por las siguientes variables explicativas al 95% de confianza:

- `accommodates`
- `host_response_time`
- `bathrooms`
- `beds`
- `host_is_superhost`
- `property_type`
- `room_type`
- `maximum_nights`
- `minimum_nights`
- `neighbourhood_cleansedCarretera de Cádiz`
- `neighbourhood_cleansedCentro`
- `neighbourhood_cleansedCruz de Humilladero`
- `neighbourhood_cleansedEste`

De estas variables, *bathroom* y *room\_type* son las dos variables con los coeficientes más grandes (En valor absoluto), implicando que son las dos variables con más peso a la hora de la determinación del precio de un anuncio.

Esto lo que quiere decir es que los dos principales factores que determinan el precio por noche de un alojamiento en la ciudad de Málaga sin tener en cuenta el barrio de localización son: El tipo de habitación que se oferta y el número de baños de los que va a tenerse disponibilidad en el alojamiento.

Dentro de los barrios, el peso del coeficiente asociado al barrio Centro es grande, y positivo, tanto por sí mismo como por comparación con el resto de barrios, demostrando la importancia de este a la hora de determinar el precio de un alojamiento, indicando que el precio por noche de un alojamiento con unas ciertas características será

significativamente mayor si se encuentra en el barrio Centro que el precio por noche de un alojamiento similar en otro barrio.

En relación al análisis clúster, pese a que es posible la división en tantos clusters como se desee, el nº óptimo de clusters en los que dividir estos datos es 1, ya que el gran nº de anuncios en **Centro** en comparación con el resto de barrios eclipsa los efectos del resto.

De las componentes principales calculadas después de este análisis clúster se tiene que:

- **Primera Componente Principal (CP1):** Las variables que más influyen son "*accommodates*" y "*beds*" (ambas negativamente) y "*price*" (también negativamente). Esto indica que la **CP1** refleja un factor combinado de capacidad y costo, donde una mayor capacidad de alojamiento y un precio más alto se asocian con valores más bajos de **CP1**. Además, el tipo de propiedad y el tipo de habitación tienen una influencia positiva moderada, lo que sugiere que diferentes tipos de propiedades y habitaciones están asociados con valores más altos de **CP1**.
- **Segunda Componente Principal (CP2):** Las variables más influyentes son "*bathrooms*", "*room\_type*" y "*neighbourhood\_cleansed*" (todas negativamente). Esto sugiere que la **CP2** representa un factor relacionado con la calidad del alojamiento y la ubicación, donde un mayor número de baños, ciertos tipos de habitaciones y ciertos vecindarios están asociados con valores más bajos de **CP2**. El número máximo de noches y el precio tienen una influencia positiva muy pequeña, indicando un impacto mínimo en el aumento de la **CP2**.

En resumen, la **CP1** está relacionada con la capacidad y el costo del alojamiento, mientras que la **CP2** está más asociada con la calidad del alojamiento y su ubicación.

De esto se pueden sacar varias conclusiones:

- Capacidad y Costo como Factores Clave (**CP1**):
  - Las variables "*accommodates*" y "*beds*" tienen una influencia negativa significativa en la **CP1**, al igual que el "*price*". Esto sugiere que los alojamientos con mayor capacidad (más personas y camas) y precios más altos tienden a estar asociados con valores más bajos de la primera componente principal.
  - Este patrón podría indicar que hay un segmento de mercado donde los alojamientos más grandes y costosos están menos preferidos o se comportan de manera diferente en comparación con otros tipos de alojamientos.

- Influencia del Tipo de Propiedad y Habitación (**CP1**):
  - Las variables "*property\_type*" y "*room\_type*" tienen una influencia positiva moderada en la **CP1**, lo que sugiere que diferentes tipos de propiedades y habitaciones están asociados con valores más altos de esta componente.
  - Esto podría indicar que la variedad en el tipo de alojamiento juega un papel importante en la preferencia de los usuarios o en la manera en que se distribuyen estos datos.
- Calidad del Alojamiento y Ubicación (**CP2**):
  - Las variables "*bathrooms*", "*room\_type*", y "*neighbourhood\_cleansed*" tienen una influencia negativa significativa en la **CP2**. Esto sugiere que los alojamientos con más baños, ciertos tipos de habitaciones y ubicados en ciertos vecindarios tienden a estar asociados con valores más bajos de esta componente.
  - Esto podría reflejar un factor de calidad del alojamiento, donde ciertas características de los alojamientos y su ubicación influyen en cómo se agrupan los datos, posiblemente indicando preferencias de los usuarios o características de mercado específicas.
- Impacto Mínimo de Otras Variables (**CP2**):
  - Las variables como el "*maximum\_nights*" y "*price*" tienen una influencia positiva muy pequeña en la **CP2**, sugiriendo que estos factores tienen un impacto mínimo en esta componente.
  - Esto podría indicar que el número máximo de noches permitidas y el precio no son tan determinantes en la segunda componente principal, quizás porque la calidad del alojamiento y la ubicación tienen un peso mayor en esta dimensión.

Como conclusión, el análisis sugiere que la primera componente principal está dominada por factores de capacidad y coste del alojamiento, mientras que la segunda componente principal está más influenciada por la calidad del alojamiento y la ubicación. Estos resultados pueden ser útiles para segmentar el mercado de alojamientos turísticos no hosteleros y entender mejor las preferencias de los usuarios. Por ejemplo, se podría enfocar en ofrecer diferentes estrategias de marketing o ajuste de precios según las características dominantes de cada segmento.

Finalmente, en relación al modelo Logit calculado también se pueden extraer diversas conclusiones y posibles utilidades del mismo.

El modelo logit desarrollado para identificar alojamientos con precios altos muestra una capacidad razonable de discriminación, con un área bajo la curva ROC (AUC) de 0.77. Este valor indica que el modelo tiene una habilidad moderada para distinguir entre alojamientos con precios altos y bajos. La exactitud general del modelo es buena, con valores similares en el conjunto de prueba (83.13%) y el conjunto de entrenamiento (82.73%), lo que sugiere que el modelo generaliza bien entre los diferentes conjuntos de datos.

El análisis de las variables significativas revela que ciertas características tienen un impacto positivo en la probabilidad de que un alojamiento tenga un precio alto. Las variables como el número de plazas disponibles ("*accommodates*"), el número de baños ("*bathrooms*") y el tipo de propiedad ("*property\_type*") están asociadas con un aumento en la probabilidad de precios altos. De estas, el número de baños es la variable más influyente, destacando su importancia en la evaluación del precio de los alojamientos, es decir, que el número de baños de un alojamiento es el factor más importante en el mercado de alquiler de alojamientos turísticos no hosteleros de Málaga de la plataforma Airbnb.

Por otro lado, variables como la verificación de identidad del anfitrión ("*host\_identity\_verified*"), el tipo de habitación ("*room\_type*") y el número mínimo de noches ("*minimum\_nights*") tienen coeficientes negativos. Esto sugiere que, aunque estas características pueden ser importantes, tienden a estar asociadas con una menor probabilidad de que el precio del alojamiento sea alto. En particular, el tipo de habitación ("*room\_type*") tiene el coeficiente negativo más significativo, lo que indica que su impacto es considerable en la determinación de si el precio por noche de un alojamiento es excesivo o no, por tanto, el tipo de habitación anunciada es determinante también para determinar si el precio es excesivo, o no.

El modelo no está exento de limitaciones. A pesar del buen desempeño general, el modelo presenta desafíos en términos de sensibilidad, que es relativamente baja (27.57% en el conjunto de prueba y 24.22% en el conjunto de entrenamiento). Esto significa que el modelo tiene dificultades para identificar correctamente todos los casos positivos (alojamientos con precios altos), lo que puede ser problemático en situaciones donde es crucial no pasar por alto estos casos. La alta especificidad (95% en el conjunto de prueba y 96.11% en el conjunto de entrenamiento) sugiere que el modelo es muy bueno para identificar los casos negativos, pero su capacidad para detectar casos positivos necesita mejora.



Para acabar, se van a presentar algunos posibles ejemplos de utilidad del modelo Logit calculado, para mostrar su utilidad a pesar de las limitaciones previamente expuestas:

- **Optimización de Estrategias de Precios:** Las plataformas de alquiler de alojamientos pueden utilizar este modelo para identificar propiedades con potencial para precios más altos. Esto puede ayudar a los anfitriones a ajustar sus precios de manera más competitiva y maximizar sus ingresos, basándose en las características significativas identificadas por el modelo, y por el contrario, desde el punto de vista del cliente, puede utilizar este modelo para comprobar si el precio por noche de un alojamiento de su interés es excesivo o no, ayudando en la toma de decisión de alquilar el alojamiento.
- **Segmentación de Mercado:** Las agencias de marketing y las plataformas de alquiler pueden segmentar mejor sus mercados utilizando las predicciones del modelo. Por ejemplo, pueden dirigir campañas de marketing específicas a propiedades con características que se correlacionan positivamente con precios altos, como un mayor número de baños y tipos específicos de propiedades.
- **Evaluación de Nuevas Propiedades:** Los nuevos anfitriones pueden usar el modelo para evaluar el potencial de sus propiedades en términos de precio. Al entender qué características aumentan la probabilidad de precios altos, los anfitriones pueden invertir en mejoras específicas, como aumentar el número de baños o cambiar el tipo de propiedad, para aumentar el valor percibido de su alojamiento.
- **Mejoras en la Plataforma:** Las plataformas pueden utilizar los resultados del modelo para mejorar sus algoritmos de recomendación y búsqueda, destacando propiedades que tienen una alta probabilidad de ser listadas a precios más altos. Esto puede mejorar la experiencia del usuario al proporcionar recomendaciones más precisas y relevantes.

En resumen, aunque el modelo logit presenta ciertas limitaciones en su capacidad para detectar todos los casos de precios altos, su buen desempeño general y su capacidad para identificar características significativas lo hacen una herramienta valiosa para la optimización de precios, segmentación de mercado y evaluación de nuevas propiedades en el sector de alquiler de alojamientos.

## Bibliografía

- Pellejero Martínez, C. (s.f.). *Turismo y Economía en la Málaga del siglo XX*. Universidad de Málaga.
- Euronews. (2024, 7 de julio). *Málaga para vivir, no para sobrevivir: Los malagueños protestan contra el turismo en medio de*. Euronews. <https://es.euronews.com/viajes/2024/07/07/malaga-para-vivir-no-para-sobrevivir-los-malaguenos-protestan-contr-el-turismo-en-medio-d>
- Diputación de Málaga. (2023, 20 de septiembre). *El sector turístico de la Costa del Sol evidencia el peso del turismo sobre otros sectores y lanza un potente mensaje contra la turismofobia*. Diputación de Málaga. [https://www.malaga.es/noticias/com1\\_md-3/com1\\_ct-0/com1\\_fb-0/com1\\_cb-0/com1\\_md3\\_cd-50341/el-sector-turistico-de-la-costa-del-sol-evidencia-el-peso-del-turismo-sobre-otros-sectores-y-lanza-un-potente-mensaje-contr-la-turismofobia](https://www.malaga.es/noticias/com1_md-3/com1_ct-0/com1_fb-0/com1_cb-0/com1_md3_cd-50341/el-sector-turistico-de-la-costa-del-sol-evidencia-el-peso-del-turismo-sobre-otros-sectores-y-lanza-un-potente-mensaje-contr-la-turismofobia)
- Cadena SER. (2024, 8 de mayo). *El empleo del sector turístico crece casi un 8% en Málaga, con 125.000 ocupados en la hostelería, alojamientos y agencias*. Cadena SER. <https://cadenaser.com/andalucia/2024/05/08/el-empleo-del-sector-turistico-crece-casi-un-8-en-malaga-con-125000-ocupados-en-la-hosteleria-alojamientos-y-agencias-ser-malaga/>
- Inside Airbnb. (n.d.). Get the data. Inside Airbnb. Retrieved June 23, 2024, from <https://insideairbnb.com/get-the-data/>
- Tarancón Morán, M. Á. & Quintana Rojo, C., (2023-2024). *Técnicas Multivariantes Aplicadas Al Análisis Sectorial*. Máster de Análisis y Modelización de Datos Económicos, Universidad de Castilla-La Mancha.
- Cameron, A. C., & Trivedi, P. K. (2005). In **Microeconometrics: Methods and Applications** (pp. 487-554). Cambridge University Press.
- del Corral, J. (2023-2024). *Microeconometría aplicada*. Tema 3: Modelos de variable dependiente cualitativa: probit/logit. Máster de Análisis y Modelización de Datos Económicos, Universidad de Castilla-La Mancha.
- Kimes, S. E., & Wirtz, J. (2015). *Revenue management: Advanced strategies and tools to enhance firm profitability*. World Scientific.
- Wachsmuth, D., & Weisler, A. (2018). Airbnb and the Rent Gap: Gentrification through the Sharing Economy. *Environment and Planning A: Economy and Space*, 50(6), 1147-1170.

- Fu, Y. (2023). *Nuevas estrategias frente a la intensificación turística en destinos* (Trabajo de Fin de Máster, Universidad de Málaga, Máster Universitario en Dirección y Planificación del Turismo). Universidad de Málaga.
- Cano, E. L. (2024). *Introducción al software estadístico R* ([capítulos 9 y 10]).
- Delgado, R. (2018, junio 23). *Introducción a los modelos de agrupamiento (Clustering) en R*. RPubs. <https://rpubs.com/rdelgado/399475>
- Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, 42(3), 203-231. doi:10.1023/A:1007601015857
- Chawla, N. V., Japkowicz, N., & Kotcz, A. (2004). Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, 6(1), 1-6. doi:10.1145/1007730.1007733
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>

## Anexo I – Tratamiento de las variables

- La variable *host\_response\_time* se dividía en 4 categorías, por lo que a cada una de ellas se le asignó un valor numérico:
  - Within an hour: 1
  - Within a few hours: 2
  - Within a day: 3
  - A few days or more: 4
- Las variables *host\_is\_superhost*, *host\_has\_profile\_pic* y *host\_identify\_verified* son de naturaleza binaria, por lo que a las respuestas que ofrecían (**False** y **True**) se les adjudicaron los valores 0 y 1 respectivamente.
- La variable *room\_type* se dividía en 3 categorías, por lo que a cada una de ellas se le asignó un valor numérico:
  - Entire home/apt: 1
  - Private room: 2
  - Shared room: 3
- La variable *property\_type* se divide en 46 categorías, por lo que a cada una de ellas se le asignó un valor numérico:
  - Entire rental unit : 1
  - Private room in rental unit : 2
  - Shared room in rental unit : 3
  - Private room in townhouse : 4
  - Private room in home : 5
  - Entire loft : 6
  - Entire condo : 7
  - Entire home : 8
  - Entire serviced apartment : 9
  - Entire townhouse : 10
  - Entire cottage : 11
  - Entire guest suite : 12
  - Entire chalet : 13
  - Private room in condo : 14
  - Room in boutique hotel : 15
  - Entire villa : 16
  - Private room in guesthouse : 17
  - Room in serviced apartment : 18
  - Entire guesthouse : 19
  - Floor : 20
  - Camper/RV : 21
  - Tiny home : 22

- Entire vacation home : 23
- Private room in guest suite : 24
- Room in aparthotel : 25
- Private room : 26
- Shared room in earthen home : 27
- Private room in loft : 28
- Private room in chalet : 29
- Private room in hostel : 30
- Private room in casa particular : 31
- Entire cabin : 32
- Private room in serviced apartment : 33
- Entire place : 34
- Shared room in chalet : 35
- Casa particular : 36
- Private room in bed and breakfast : 37
- Dome : 38
- Private room in villa : 39
- Room in hotel : 40
- Private room in vacation home : 41
- Private room in farm stay : 42
- Private room in yurt : 43
- Shared room in casa particular : 44
- Cave : 45
- Shared room in hotel : 46
- La variable *neighbourhood\_cleansed* se divide en 11 categorías, por lo que a cada una de ellas se le asignó un valor numérico:
  - Este : 1
  - Centro : 2
  - Churriana : 3
  - Carretera de Cadiz : 4
  - Bailen-Miraflores : 5
  - Cruz De Humilladero : 6
  - Teatinos-Universidad : 7
  - Puerto de la Torre : 8
  - Ciudad Jardin : 9
  - Campanillas : 10
  - Palma-Palmilla : 11

## Anexo II – Descriptivo de las variables

### Medias de las variables por barrio

Neighbourhood	Number of Properties	Accommodates (Media)	Antigüedad (Media)	Host Response Rate (Media)	Host Response Time (Media)	Neighborhood Overview Flag (Media)	Bathrooms (Media)	Beds (Media)	Price (Media)	Host is Superhost (Media)	Host Listings Count (Media)	Host Has Profile Pic (Media)	Host Identity Verified (Media)	Property Type (Media)	Room Type (Media)	Minimum Nights (Media)	Maximum Nights (Media)
Bailen-Miraflores	205	3.2634	17595.639	0.9363	1.361	0.4829	1.122	2.0146	74.9951	0.2927	23.3512	0.9854	0.9951	3.8976	1.2829	4.9805	473.639
Campanillas	34	5.6176	17935.1765	0.8456	1.5294	0.4412	1.5882	3.5	153.1471	0.2941	147.5882	0.9706	0.9706	10.5882	1.2941	2.0882	526.9412
Carretera de Cadiz	647	4.0479	17675.1731	0.9647	1.2566	0.5085	1.2968	2.6569	99.2334	0.2782	35.8192	0.9784	0.9799	4.1267	1.2612	3.102	583.6213
Centro	4767	4.0023	17356.7017	0.9718	1.1391	0.4789	1.2717	2.4657	126.5049	0.3025	42.6931	0.9904	0.9874	3.0657	1.0967	2.3256	589.3757
Churriana	97	5.7938	17621.8763	0.9669	1.3093	0.4948	1.9536	4.0928	184.1134	0.3299	73.0206	0.9485	0.9897	8.433	1.1443	2.4124	664.6289
Ciudad Jardin	46	4.4565	17656.7391	0.8972	1.5435	0.413	1.3696	2.9348	96.3913	0.4348	54.3261	0.9565	0.9565	5.1739	1.3043	8.8696	332.2391
Cruz De Humilladero	292	3.9075	17606.3836	0.9348	1.2979	0.4418	1.2877	2.5651	108.2603	0.2603	57.7432	0.9829	0.9452	3.8459	1.2637	3.0377	481.6849
Este	684	4.4737	17351.7909	0.9575	1.3553	0.481	1.4737	3.0307	139.3787	0.3056	40.674	0.9737	0.9664	5.4956	1.0892	2.8012	602.7164
Palma-Palmilla	52	3.7885	17245.2692	0.9525	1.2692	0.5577	1.4808	2.6538	100	0.3654	25.3462	0.9615	0.9808	6.3654	1.5577	1.8846	568.6923
Puerto de la Torre	39	6.0513	17330.2308	0.9787	1.4359	0.4359	1.9487	4.359	154.5385	0.2051	75.7949	0.9487	0.9744	8.7436	1.2051	2.6154	481.7436
Teatinos-Universidad	67	4.0896	17188.6567	0.9158	1.6119	0.4627	1.5224	3.1493	84.4478	0.3284	25.2836	0.9403	0.9851	4.4478	1.2687	4.3284	443.6866

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 10

### Medianas de las variables por barrio

Neighbourhood	Number of Properties	Accommodates (median)	Antigüedad (median)	Host Response Rate (median)	Host Response Time (median)	Neighborhood Overview Flag (median)	Bathrooms (median)	Beds (median)	Price (median)	Host is Superhost (median)	Host Listings Count (median)	Host Has Profile Pic (median)	Host Identity Verified (median)	Property Type (median)	Room Type (median)	Minimum Nights (median)	Maximum Nights (median)
Bailen-Miraflores	205	3	17324	1	1	0	1	2	70	0	3	1	1	1	1	2	365
Campanillas	34	6	17755	1	1	0	1	3.5	105	0	2.5	1	1	8	1	2	365
Carretera de Cadiz	647	4	17520	1	1	1	1	2	88	0	4	1	1	1	1	2	365
Centro	4767	4	17205	1	1	0	1	2	106	0	12	1	1	1	1	2	365
Churriana	97	5	17583	1	1	0	1.5	3	108	0	3	1	1	8	1	2	365
Ciudad Jardin	46	3	17547.5	1	1	0	1	1	69	0	2.5	1	1	2	1	2	232.5
Cruz De Humilladero	292	4	17617	1	1	0	1	2	90.5	0	4	1	1	1	1	2	365
Este	684	4	17269.5	1	1	0	1	2	100	0	5	1	1	1	1	2	365
Palma-Palmilla	52	2	17128	1	1	1	1	1	45	0	4	1	1	2	2	1.5	365
Puerto de la Torre	39	4	17058	1	1	0	1.5	3	90	0	2	1	1	8	1	2	365
Teatinos-Universidad	67	4	16979	1	1	0	1.5	3	80	0	3	1	1	2	1	3	365

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 11



# Trabajo Fin de Grado

máster universitario en modelización  
y análisis de datos económicos

## Varianzas de las variables por barrio

Neighbourhood	Number of Properties	Accommodates (variance)	Antigüedad (variance)	Host Response Rate (variance)	Host Response Time (variance)	Neighborhood Overview Flag (variance)	Bathrooms (variance)	Beds (variance)	Price (variance)	Host is Superhost (variance)	Host Listings Count (variance)	Host Has Profile Pic (variance)	Host Identity Verified (variance)	Property Type (variance)	Room Type (variance)	Minimum Nights (variance)	Maximum Nights (variance)
Baileu-Miraflores	205	2.6558	1272634.7514	0.0336	0.7416	0.2509	0.0904	1.2204	2067.9363	0.208	4324.7976	0.0145	0.0049	48.112	0.2039	93.9604	186801.7122
Campanillas	34	10.4857	1263477.7861	0.0928	1.0446	0.254	0.5223	5.5303	11327.2807	0.2139	127124.6132	0.0294	0.0294	94.795	0.2139	1.1738	152050.6025
Carretera de Cadiz	647	4.2562	1266569.261	0.0197	0.4387	0.2503	0.2965	2.5973	5047.399	0.2011	10721.8109	0.0212	0.0197	57.0087	0.2119	53.2496	205181.762
Centro	4767	4.0468	1207217.2945	0.0135	0.2511	0.2496	0.3175	3.2925	7054.8824	0.211	6571.9169	0.0096	0.0124	34.152	0.0891	21.6246	201296.2401
Churriana	97	11.7487	1246638.2762	0.0234	0.4659	0.2526	1.8546	8.2309	38765.5183	0.2234	44806.6454	0.0494	0.0103	54.9356	0.1456	2.5782	220605.8608
Ciudad Jardin	46	14.5647	1697698.686	0.0683	0.7425	0.2478	0.5937	8.4179	9180.2879	0.2512	46419.9135	0.0425	0.0425	66.6802	0.2609	1934.2048	157413.8749
Cruz De Humilladero	292	4.5378	1164737.6943	0.0406	0.6016	0.2475	0.4496	2.4391	8439.1691	0.1932	26966.7826	0.0169	0.052	52.5569	0.1948	46.057	179828.4777
Este	684	6.847	1260842.0133	0.0177	0.4988	0.25	0.8434	6.1586	13814.6749	0.2125	14804.955	0.0257	0.0325	53.9604	0.0843	18.101	204362.4465
Palma-Palmilla	52	13.621	1803038.7104	0.0246	0.3183	0.2515	0.9898	12.5053	21108.7059	0.2364	5456.1523	0.0377	0.0192	77.8051	0.2515	1.359	201762.6878
Puerto de la Torre	39	17.1552	1632864.919	0.0033	0.5682	0.2524	1.4052	15.1835	24567.7814	0.1673	38005.3252	0.0499	0.0256	59.0904	0.1673	6.1377	207378.722
Teatinos-Universidad	67	2.9009	961291.8652	0.0542	0.7865	0.2524	0.2381	2.2804	2165.2813	0.2239	3400.2365	0.057	0.0149	34.948	0.1995	26.8905	206011.2488

Fuente: Elaboración propia en base a los datos extraídos de insideairbnb

Tabla 12