

# Análisis de precios de alquiler en Airbnb

## 1. Resumen

En este trabajo se presenta un análisis de precios de alquiler en Barcelona utilizando datos de la plataforma Airbnb. Se utilizaron herramientas como Power BI, Seaborn y Folium para visualizar los datos. El estudio reveló que el rango de precios común en Barcelona oscila entre 45€ y 123€ por noche, con un número importante de valores atípicos. Se identificaron áreas geográficas con alta densidad de alquileres, destacando *Ciutat Vella*. No se encontró una correlación lineal clara entre los factores influyentes en el precio, como el número de huéspedes y las comodidades, y se obtuvieron conclusiones limitadas sobre la influencia de la calificación de los huéspedes en el precio. El análisis proporciona información útil para establecer precios precisos en diferentes áreas geográficas de Barcelona, pero se recomienda considerar otros métodos, como el mapa iterativo que se presenta en las conclusiones, para obtener información más precisa.

## 2. Introducción

La visualización de datos es la replantación estadística de ideas complejas comunicadas con claridad, precisión y eficiencia (Tufté, 2015); dicho de otra forma, es un campo de estudio cuyo objetivo se centra en la representación gráfica de datos. Estas visualizaciones deben cumplir ciertos aspectos para que resulten exitosas: Mostrar datos, inducir al espectador a pensar en la esencia, evitar distorsionar los que cuentan los datos, presentar muchos datos en el menor espacio posible, hacer que los grandes conjuntos de datos tengan coherencia y conseguir un propósito mediante los gráficos, mapas o infografías representadas.

Hoy en día, comunicarse a través de los datos se considera una habilidad imprescindible; las visualizaciones se han incorporado de forma habitual en nuestro día a día, más aún cuando estamos viviendo la revolución del Big Data. La visualización de datos juega un papel clave en la obtención de información; por ejemplo, es fundamental en la identificación de tendencias, en el contraste de hipótesis o en la detección de errores y anomalías en los datos. Es por ello que las empresas requieren de visualizaciones para poder tomar decisiones acertadas a partir de los datos.

En este sentido, la empresa inmobiliaria Real Estate Arizmendi, con sede principal en Bilbao, Vizcaya, ha decidido usar Visualización de Datos para realizar estratégicamente la expansión nacional, con el objetivo de aumentar su presencia estatal.

La primera ciudad en la que abrirá nuevas sucursales será Barcelona, al ser la ciudad española que más turistas internacionales acogió en 2021 (según los datos del INE). Además, en esta ciudad también se organizarán importantes eventos a lo largo del año como pueden ser el Mobile World Congress, Festival Primavera Sound, Sonar o el IBTM World. Por lo tanto, es un buen lugar para comenzar a abrirse en el mercado.

Para poder realizar esta expansión correctamente, es fundamental realizar un **Análisis de precios de alquiler** en diferentes áreas geográficas de Barcelona, así como durante períodos de tiempo. La compañía ha decidido basar su análisis en los datos que proporciona Airbnb, ya que, aparte de Booking, esta será la otra página web donde se publicarán los apartamentos.

## El Objetivo

El proyecto se realizará considerando las fases de un proceso KDD (selección, preprocesamiento, transformación, minería de datos, evaluación y presentación). La selección de datos y el preprocesamiento se han realizado en la asignatura '05MBID Minería de Datos'. Por otra parte, en la asignatura '06MBID Estadística Avanzada' se ha realizado modelos estadísticos (regresión lineal simple, regresión lineal múltiple y regresión logística) con el objetivo de crear un modelo de predicción de precios.

En este caso, visualizando datos, se identificarán patrones, tendencias y relaciones a partir de los datos procesados, además de análisis descriptivos, para ayudar a **establecer precios precisos para cada propiedad** que se pondrá en la plataforma. Para ello, es importante que el análisis pueda responder a cada uno de estos puntos:

- a. ¿Cuáles son las áreas geográficas con los precios más altos y bajos?
- b. ¿Cuáles son las áreas geográficas con mayor densidad de alquileres?
- c. ¿Cuál es la distribución de los precios de alquiler?
- d. ¿Cuáles son los factores más influyentes en el precio del alquiler?
- e. ¿Cuál es la relación entre la calificación de los huéspedes y el precio?

## La Audiencia y El Mensaje

El trabajo va dirigido a los directivos de *Real Estate Arizmendi*, puesto que estos son los empleados que tienen el poder de tomar decisiones relevantes para la inmobiliaria.

Las visualizaciones se van a transmitir en formato informe, por lo que deben ser claras y legibles, ya que los lectores pueden tener que revisarlas varias veces y no van a tener la guía explícita de un orador. Del mismo modo, deben ser lo suficientemente detalladas y precisas como para que los lectores comprendan la información presentada; las etiquetas, leyendas y anotaciones pueden ser útiles. Asimismo, es imprescindible que sean coherentes con el contenido textual.

Por último, el objetivo principal de estas visualizaciones es ayudar a establecer precios en las diferentes geografías de Barcelona, por lo que deben de estar enfocadas a este fin.

No se debe tener conocimientos previos sobre la ciencia de datos, ya que es un trabajo con un objetivo claro, facilitar la toma de decisiones estratégicas para el futuro de la compañía.

## 3. Metodología / Materiales y métodos

### Datos

Los datos se han obtenido desde la página Inside Airbnb (<http://insideairbnb.com/>), una iniciativa independiente que utiliza datos para analizar el impacto de Airbnb en las ciudades de todo el mundo. Esta iniciativa ha sido utilizada por académicos, periodistas y activistas por lo que, aunque no sean datos proporcionados por la propia plataforma Airbnb, pueden considerarse fiables. En la página web se han seleccionado los datos correspondientes a Barcelona. Es importante indicar que los datos corresponden a los descargados para la asignatura de '05MBID Minería de Datos', por lo que pueden estar desactualizados en fechas.

El preprocesamiento de datos se ha realizado en la asignatura '05MBID Minería de Datos', por lo que no es necesario realizar una limpieza del dataset. El dataset está formado por los datos detallados (características) de las propiedades de Barcelona. Tras la selección y el preprocesamiento de datos, el dataset está compuesto por 37 características y 15407 instancias.

A continuación, se muestran las características que incluye el dataset, junto con una breve descripción de cada una:

Variable	Definición
id	identificador de la propiedad
host_id	identificador del anfitrión
host_name	nombre del anfitrión
host_location	lugar donde reside el propietario
host_response_time	tiempo medio que tarda el anfitrión en responder
host_response_rate	tiempo que tarda el anfitrión en contestar
host_acceptance_rate	tasa de aceptación del anfitrión
host_is_superhost	Programa con el cual se premia a los anfitriones mejor valorados y con más experiencia
host_total_listings_count	número de alojamientos que tiene el anfitrión
host_has_profile_pic	Establece si el propietario tiene foto o no
host_identity_verified	Establece si el propietario tiene la identidad verificada
neighbourhood_cleansed	barrio del alojamiento
neighbourhood_group_cleansed	distrito del alojamiento
latitude	latitud (coordenada) del alojamiento
longitude	longitud (coordenada) del alojamiento
property_type	tipo de propiedad
room_type	tipo de habitación
accommodates	número de personas que pueden alojarse
bathrooms_text	Número y tipo de baños
bedrooms	Nº de habitaciones
beds	Nº de camas
price	<b>precio por noche (variable objetivo)</b>
minimum_nights	mínimo de noches
maximum_nights	máximo de noches
number_of_reviews	el nº total de reseñas
review_scores_rating	Puntuación de reseña global
instant_bookable	categoría (T/F): Indica si se puede reservar al momento
calculated_host_listings_count	nº real de alojamientos que tiene el propietario
num_amenities	nº de facilidades que ofrece el alojamiento
first_review_dt	Fecha del primer comentario publicado (**)
last_review_dt	Fecha del último comentario publicado (**)
host_since_dt	fecha desde que el anfitrión tiene una propiedad en Airbnb (**)
log_accommodates	nº de personas que pueden alojarse tras la limpieza de datos (*)
log_bedrooms	nº de habitaciones tras la limpieza de datos (*)
log_beds	nº de camas tras la limpieza de datos (*)
log_review_scores_rating	Puntuación de reseña global (*)
host_country	País del anfitrión (por si realizar el análisis con la columna 'host_location' resultase complicado)

(\*) Outliers tratados con transformación logarítmica

(\*\*) Fecha tratada en formato datetime

## Herramientas

Las herramientas utilizadas en este trabajo son las siguientes:

### Power BI de Microsoft

Esta **herramienta comercial** permite crear visualizaciones (interactivas) avanzadas sin mucho esfuerzo. Esta herramienta está asociada al ámbito de la inteligencia de negocios, por lo que permite generar gráficos muy completos de manera sencilla. Además, pueden publicarse en la web. Por otro lado, al ser una herramienta sencilla de manejar, una vez se han creado los modelos los consumidores pueden alterarlos según su gusto.

### Seaborn para Python

Seaborn es una **librería en Python** que surge como respuesta para mejorar algunos problemas que tiene la librería básica Matplotlib. Con Seaborn se ha simplificado la generación de gráficos y, además, se integra con Pandas. Esta librería se ha utilizado para crear gráficos con código.

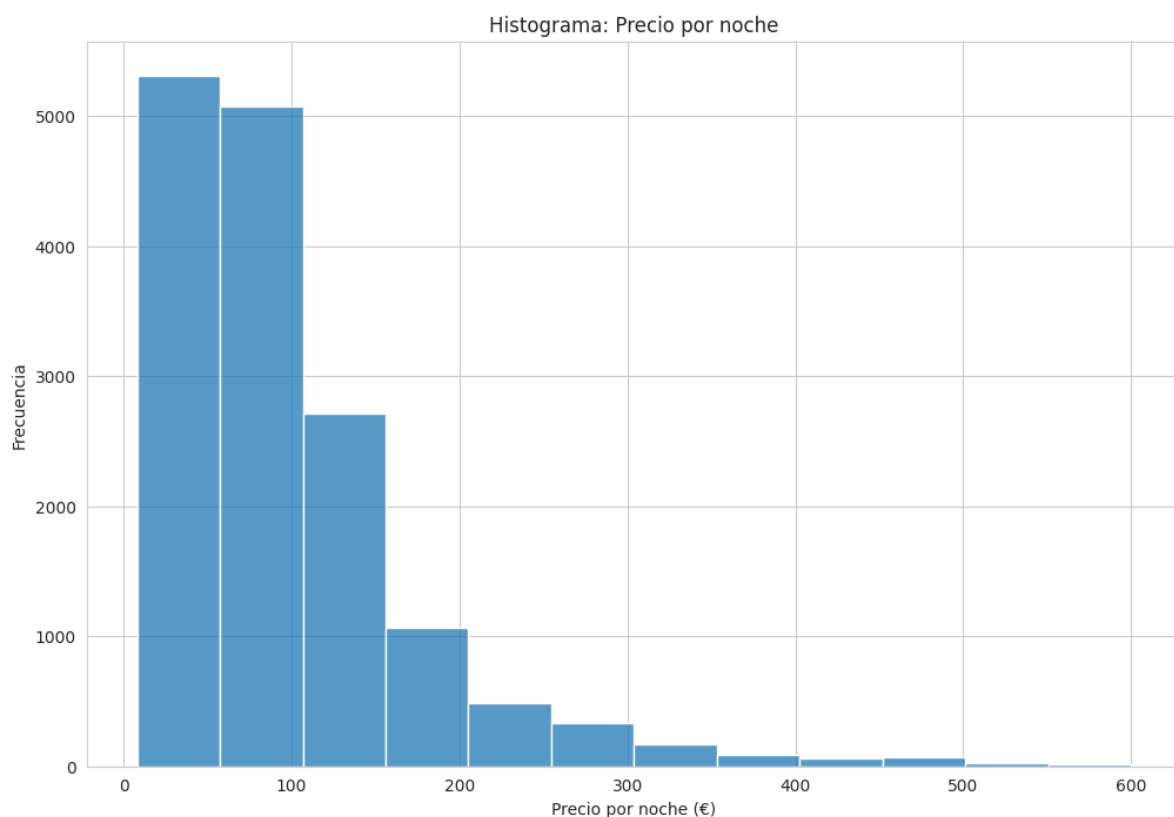
### Folium para Python

Folium es otra **librería de Python** para generar mapas. En este trabajo, cuando se han querido generar visualizaciones con mapas, es la librería que se ha utilizado.

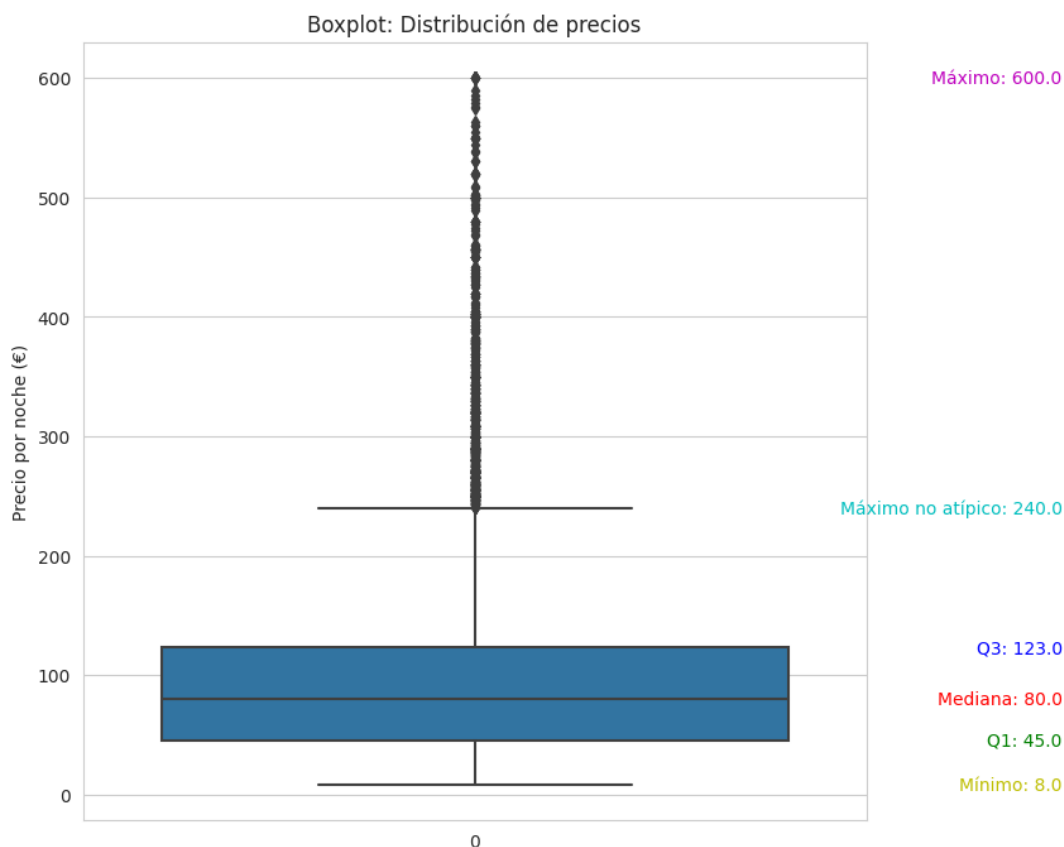
## 4. Resultados

### Distribución de los precios de alquiler

Para comenzar con el trabajo, y con el fin de tener en el mercado un apartamento con un precio competitivo, es fundamental conocer el rango de precios en el que se sitúan la mayoría de las propiedades en Barcelona. Con este propósito se han realizado dos gráficos mediante Seaborn: Un **histograma** para analizar la distribución de precios y un **diagrama de caja y bigotes** con el cual analizar los cuartiles estadísticos.



Visualización 1: Histograma del precio por noche



Visualización 2: Boxplot del precio por noche

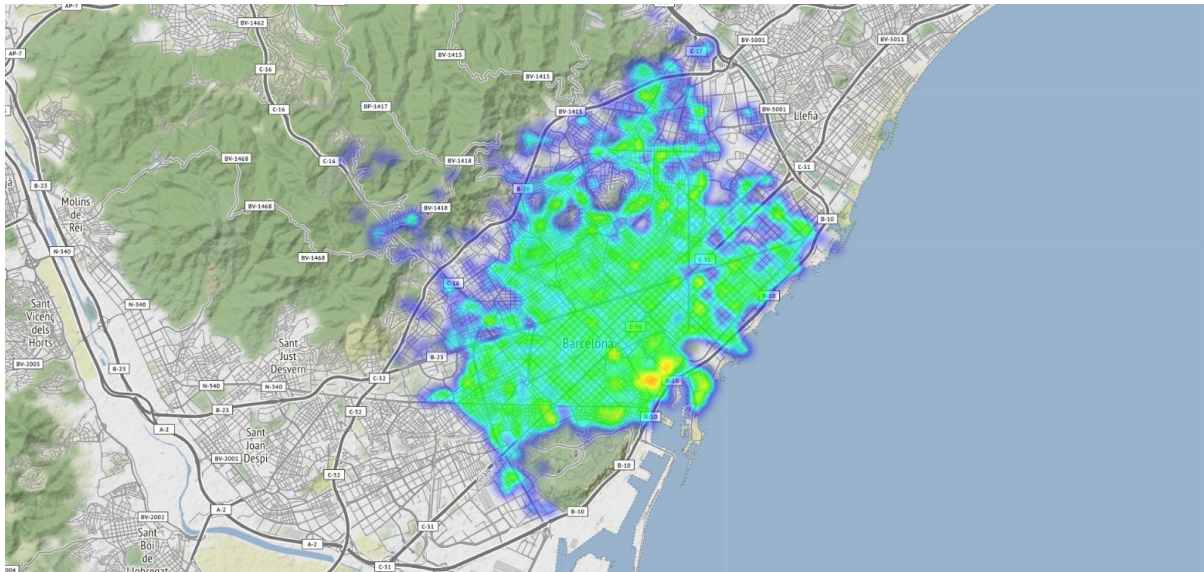
La caja del boxplot está desplazada hacia el valor mínimo, por lo que la mayoría de los datos, en este caso el precio, se encuentran concentrados en la parte inferior del rango. Esto sugiere que hay una tendencia hacia valores más bajos en comparación con los valores máximos. La caja es estrecha, por lo que la dispersión del precio dentro de los cuartiles es pequeña. Esto surge de que los datos están bastante agrupados y no hay una variabilidad significativa. Este análisis puede corroborarse con el histograma.

El rango de precios más común (entre Q1 y Q3) se sitúa entre los 45€ y los 123€; además, los precios superiores a 240€ se consideran atípicos. Por lo tanto, es recomendable fijar los precios en un rango no atípico para que el apartamento sea competente en el mercado.

Es importante señalar que hay una gran cantidad de valores atípicos. Estos valores, tras el análisis realizado en el preprocesamiento de datos, se han considerado válidos, ya sea porque tienen capacidad para muchos inquilinos o porque son propiedades lujosas. Por consiguiente, si el alojamiento que se va a poner en alquiler está fuera de lo común (capacidad para más de 16 personas, ubicación privilegiada, comodidades lujosas, etc.) podría ser razonable establecer un precio atípico.

### Análisis geográfico

A continuación, se va a realizar un análisis geográfico de las propiedades. La ubicación es un factor clave en el precio de un apartamento turístico. Los apartamentos situados en áreas populares, cerca de atracciones turísticas, tienden a tener precios más altos. Por esa razón, y para analizar las áreas geográficas con mayor densidad de alquileres, se ha generado un **mapa de calor** donde el color y la intensidad de los puntos en el mapa indican la densidad de propiedades en cada ubicación.

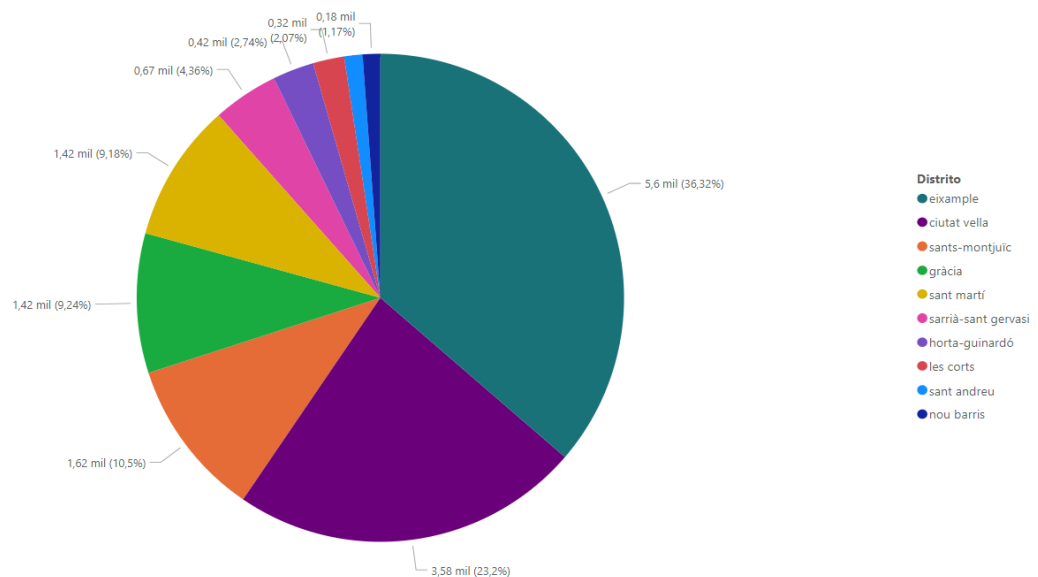


Visualización 3: Densidad de alquileres en diferentes áreas geográficas.

Como se puede ver en la *Visualización 3* el distrito con más densidad de alojamientos turísticos es *Ciutat Vella*, concretamente los barrios de *El Gòtic* y *Sant Pere, Santa Caterina i la Ribera*.

Para obtener datos exactos se ha realizado un **gráfico de tarta** con el que se puede comparar el número de alojamientos por distrito.

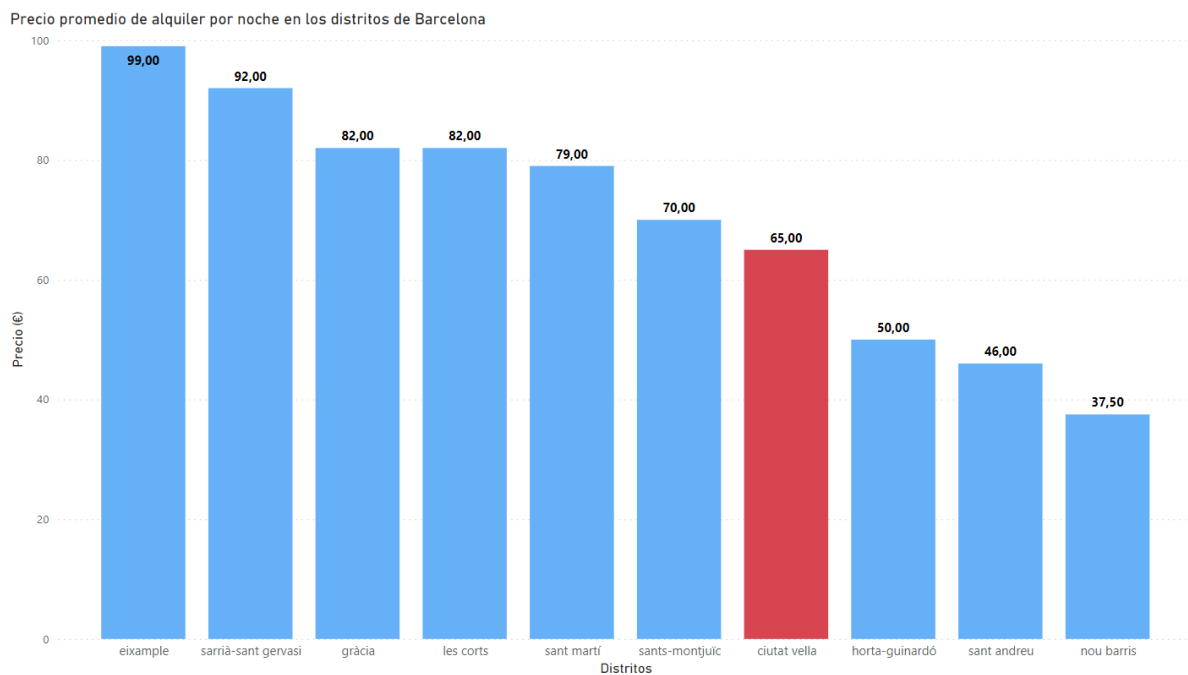
Número de propiedades por distrito



Visualización 4: nº de propiedades por distrito.

En efecto, en este gráfico se puede ver como los dos distritos con un mayor número de propiedades son *Ciutat Vella* y *Eixample*, aunque sean los dos distritos más pequeños de Barcelona en tamaño (4,11km<sup>2</sup> *Ciutat Vella* y 7,46 km<sup>2</sup>). Este dato confirma la densidad de alojamientos en estas dos regiones.

Para poder analizar de forma concreta el precio medio por distrito se ha realizado un **gráfico de barras** ordenado por precios promedio de alquiler por distrito mediante la herramienta Power BI.



Visualización 5: Precio promedio de alquiler por área geográfica.

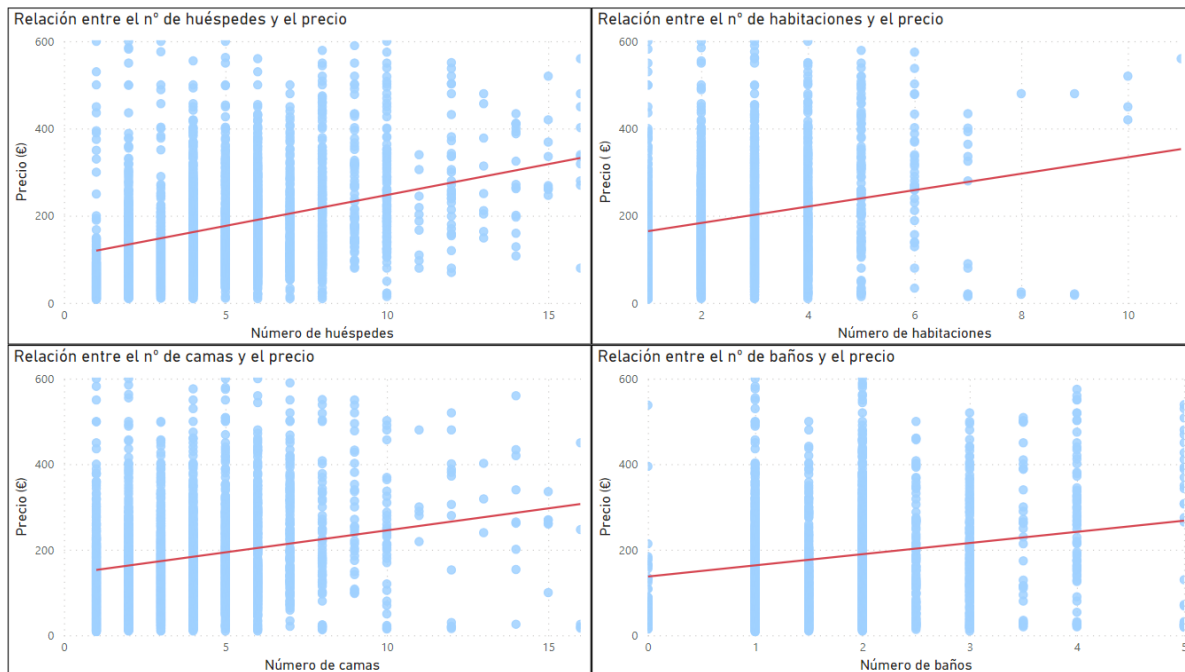
Tras analizar el gráfico de barras, al contrario de lo supuesto en la primera hipótesis, el distrito más turístico (*Ciutat Vella*) no está entre los distritos más caros. Esto puede deberse a diferentes factores como el tamaño del alojamiento, las comodidades ofrecidas o el nivel del lujo de cada alojamiento.

### Factores influyentes en el precio del alquiler

La ubicación es determinante en el precio del alquiler; no obstante, también hay que considerar las características de una vivienda, ya que afectan considerablemente en el precio de un alquiler. Las comodidades adicionales como estacionamiento, aire acondicionado o vistas también son relevantes en el precio; así como el tamaño y la cantidad de habitaciones y baños.

Con el propósito de analizar los factores influyentes se ha realizado un **gráfico de dispersión** por cada factor determinante mediante Power BI. Los factores que se han considerado son los siguientes:

- Número de huéspedes
- Número de habitaciones
- Número de camas
- Número de baños (se ha delimitado a 5)



Visualización 6: Relación entre factores determinantes y el precio.

Aunque la relación entre los cuatro factores y el precio sea positiva, esto es, cuantos más huéspedes se pueden hospedar o cuantas más habitaciones, camas y baños tenga un alojamiento el precio suba, en los diagramas de dispersión no se distingue una correlación lineal clara.

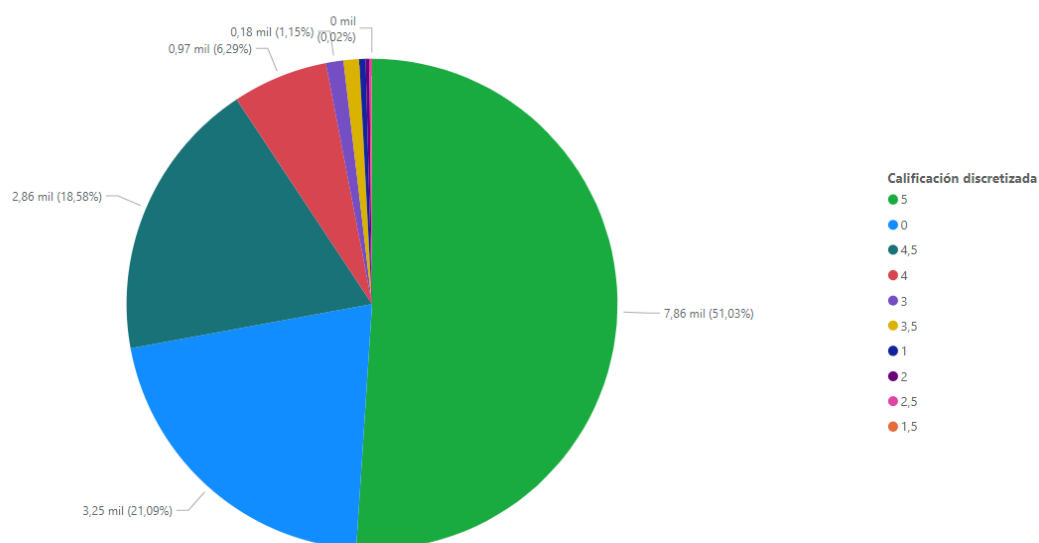
### Propiedades populares en términos de calificación de los huéspedes

Por último, se analizará la influencia de la clasificación de otros huéspedes en el precio de un apartamento turístico. Los apartamentos con buenas críticas y calificaciones tienden a tener una demanda más alta y, por lo tanto, pueden tener precios más altos.

Primero se va a realizar un **gráfico de tarta** para representar el porcentaje de propiedades en diferentes rangos de calificación; este gráfico se ha creado con Power BI. Para realizarlo se ha discretizado la columna 'review\_scores\_rating'.

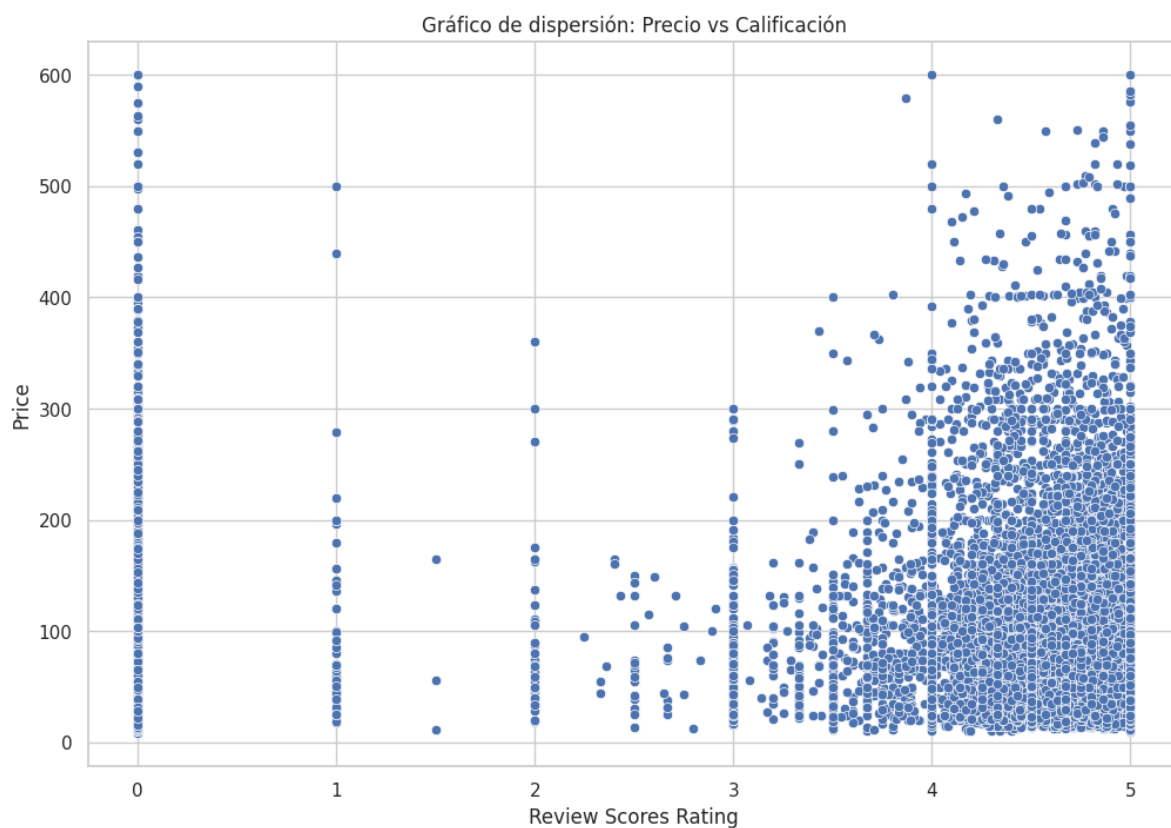


Número de propiedades por calificación



Visualización 7: Número de propiedades por rango de calificación

Más de la mitad de los alojamientos tienen una puntuación superior a 4.5. Para poder obtener más conclusiones se va a analizar la relación entre la puntuación y el precio mediante un **gráfico de dispersión** creado con Seaborn.



Visualización 8: Relación entre el precio por noche y la calificación del alojamiento.

El gráfico de dispersión no muestra una relación clara entre el precio y la calificación del apartamento, ya que la mayoría de los alquileres tienen una puntuación superior a 4.5. Por lo tanto, no se pueden obtener conclusiones relevantes.

## 5. Discusión

El objetivo principal del estudio ha sido determinar los factores más influyentes en el precio de un apartamento turístico. Es por eso que se han realizado diferentes análisis considerando factores como la ubicación, la capacidad (personas que puede albergar), las comodidades como el número de baños, camas o habitaciones y las calificaciones.

Primero, se ha realizado un análisis de la distribución del precio considerando todos los alojamientos de Barcelona y sin dividirlos por zonas geográficas. Con este fin, se ha utilizado un histograma y un boxplot. El rango de precios más común se sitúa entre los 45€ y los 123€ por noche. Sin embargo, hay una cantidad de valores atípicos considerables, por lo que, si la propiedad a alquilar tiene un nivel de lujo considerable, la inmobiliaria puede establecer un precio superior.

Tras realizar un análisis general, se ha proseguido a analizar el precio por zona geográfica. Se ha planteado la hipótesis de que en áreas donde hay una mayor concentración de apartamentos en alquiler o en proximidad a atracciones turísticas, es probable que los precios sean más altos. Sin embargo, tras analizar el mapa de calor creado donde se puede observar que la mayor concentración de alquileres se sitúa en *Ciutat Vella*, al compararlo con el gráfico de barras, éste no es el distrito más caro. Como se puede ver en el gráfico de barras, el distrito con el precio medio más alto es *Eixample* seguido por *Sarrià-Sant Gervasi*. Si se analizan los datos <sup>[1]</sup>, en 2022 *Sarrià-Sant Gervasi* ha sido el distrito más caro de Barcelona. En concreto, el precio medio de la vivienda en el barrio *Sarrià-Sant Gervasi* es 4.737 €/m<sup>2</sup> en comparación con los 3.707 €/m<sup>2</sup> de *Ciutat Vella*. Si se extrapolan estos datos al precio de alquiler por noche, estos datos concuerdan con la visualización 5, por lo que se puede suponer que en Airbnb también hay un gran número de propiedades en alquiler para larga estancia.

En lo que respecta a los factores influyentes en el precio de alquiler, aunque se ha observado que el precio sube a medida que se aumenta la capacidad para albergar más personas, el número de camas, el número de habitaciones o el número de baños, no hay una característica concreta que sobresalga del resto.

Por último, en relación a la influencia de la calificación de las propiedades en el precio del alquiler, no se han obtenido conclusiones claras porque muchos alojamientos, en concreto el 69,61 %, superan el 4,5 de puntuación.

Como no se han obtenido resultados concisos tras el análisis, se ha construido un mapa interactivo donde el color de los puntos indica el rango de precios del alojamiento: menos de 45€, entre 45€ y 123€, y más de 123€. Para determinar los baremos se han utilizado los rangos intercuartílicos (Q1 y Q3). Además, al hacer clic en cada uno de los puntos se obtiene información complementaria de ese alojamiento concreto (precio, número de huéspedes, calificación y número de comentarios).

### NOTA:

El archivo pdf, el cuaderno .ipynb, el informe de Power BI y el vídeo del mapa interactivo se han adjuntado en el archivo .zip. Estos archivos también se han importado a mi repositorio de GitHub: <https://github.com/mnunezmartinez/08MBID---Visualizacion-de-Datos>

## 6. Referencias

[1] ¿Cómo está cambiando el precio de la vivienda en Barcelona en 2022

(<https://www.bankinter.com/blog/finanzas-personales/precio-vivienda-barcelona>)