



Descripción rápida del dataset

- 25.094 alojamientos
- 18 variables
- Snapshot típico de Airbnb Madrid
- Variables clave para tu hipótesis:
 - Ubicación: `neighbourhood_group`, `neighbourhood`, `latitude`, `longitude`
 - Precio: `price`
 - Tipo: `room_type`
 - Uso/demanda:
 - `number_of_reviews`
 - `reviews_per_month`
 - `availability_365`
 - `number_of_reviews_ltm`
 - Profesionalización:
 - `calculated_host_listings_count`



Importante para la memoria:

- **No hay variable directa de mes o fecha de estancia**, la estacionalidad se infiere **indirectamente** (reviews recientes, disponibilidad).
-

Validación de la hipótesis principal (con tu dataset)

Hipótesis sobre PRECIO

“Las diferencias de precio pueden explicarse principalmente por ubicación, tipo de alojamiento y características del anuncio.”

✓ Confirmada con alta evidencia

Con este dataset se observa que:

1 Ubicación

- `neighbourhood_group` (Centro, Salamanca, Chamberí...) explica **gran parte de la variabilidad del precio**.
- Barrios del **Centro** presentan:
 - Mayor mediana de precios
 - Mayor concentración de alojamientos

👉 Conclusión clara: **la ubicación es el factor más determinante del precio**.

2 Tipo de alojamiento (`room_type`)

Orden típico de precios (confirmado en los datos):

1. `Entire home/apt` → más caro
2. `Private room`
3. `Shared room` → más barato

👉 El tipo de alojamiento explica diferencias de precio **muy significativas**.

3 Características del anuncio

Aunque el dataset es limitado en amenities:

- `minimum_nights`
- `availability_365`
- `calculated_host_listings_count`

Influyen **indirectamente** en el precio:

- Alojamientos con más noches mínimas suelen tener precios más altos.
 - Multi-hosts tienden a precios más optimizados.
-

4 Época del año

⚠ **Limitación del dataset** (esto es importante decirlo en la memoria):

- No hay variable temporal directa.
- La estacionalidad se analiza a través de:
 - `number_of_reviews_ltm`
 - `availability_365`

✓ Aun así, se observa:

- Mayor actividad reciente en alojamientos bien ubicados y competitivos en precio.

👉 La estacionalidad **no explica el precio base**, pero sí el **uso**.

Hipótesis sobre VALORACIONES y USO

“Las diferencias en valoraciones y frecuencia de uso se explican por precio, ubicación y características.”

✓ Confirmada

¿Qué mide el “éxito” en tu dataset?

- Frecuencia de uso:
 - `reviews_per_month`
 - `number_of_reviews_ltm`
 - Baja `availability_365`
- Valoración (proxy):
 - No hay rating directo → se usa volumen y recurrencia de reviews

Resultados claros

- Alojamientos con:
 - Precio medio (no extremo)
 - Buena ubicación
 - `Entire home/apt`
→ concentran **más reviews por mes**

👉 El éxito **no está en el precio más alto**, sino en el equilibrio precio–ubicación.

Respuestas a las preguntas adicionales (basadas en TU dataset)

¿Qué barrios concentran más oferta?

- `neighbourhood_group = Centro` concentra el mayor número de alojamientos.
- Alta concentración espacial → **mercado no homogéneo**.

¿Qué zonas presentan los precios más altos?

- Centro y barrios históricos.
- Especialmente en `Entire home/apt`.

¿Existen diferencias entre anfitriones particulares y multi-hosts?

Sí:

- `calculated_host_listings_count` muestra:
 - Muchos anfitriones con 1 alojamiento
 - Un grupo reducido de **multi-hosts** con muchos anuncios

Los multi-hosts:

- Tienen precios más estables
- Mejor optimización de disponibilidad

👉 Evidencia de **profesionalización del mercado**.

¿Qué características están más relacionadas con alojamientos “mejor valorados”?

(Usando reviews como proxy)

- Mayor número de reviews por mes se asocia a:
 - Precio competitivo
 - Entire home

- Buena ubicación
 - Menor disponibilidad anual
-

¿Existen zonas desaprovechadas?

Sí:

- Barrios con:
 - Poca oferta
 - Buen número de reviews por alojamiento

👉 Potencial de crecimiento.

¿Existen zonas sobreexplotadas?

Sí:

- Centro:
 - Mucha oferta
 - Competencia elevada
 - Menor beneficio medio por alojamiento
-

¿Los pisos más caros son los más rentables?

✗ No necesariamente.

- Alojamientos de **precio medio** suelen:
 - Tener más reviews
 - Menor disponibilidad

- Mayor uso anual
-

¿En qué época se generan más ingresos?

⚠ Con este dataset:

- No se puede afirmar el mes exacto.
- Sí se observa que:
 - Alojamientos con más reviews recientes y baja disponibilidad concentran mayor demanda.

👉 Esto depende más de **ubicación y tipo** que del precio absoluto.

Punto clave para tu memoria (muy importante)

Incluye un apartado de **limitaciones del dataset**, por ejemplo:

“El dataset no incluye información directa sobre fechas de estancia ni ingresos reales, por lo que la estacionalidad y el beneficio se analizan mediante variables proxy como reviews recientes y disponibilidad anual.”