

# Proyecto Final KC Data - The JSON 5

Integrantes: Ana Gallegos, Andrea Duque, Andreina Garcia, Carla Romero y Mariana Fabian.



Como inversoras en el mercado de propiedades para alquileres de Airbnb, nuestra meta de negocio es, por supuesto, lograr maximizar nuestros ingresos. Actualmente, tenemos alojamientos ubicados en dos de las ciudades europeas más importantes: Berlín y Zürich, los cuales han sido sumamente exitosos y cuyos alquileres han generado grandes ingresos. Debido al éxito de estas dos ciudades, decidimos expandirnos a Madrid, una de las capitales más visitadas en el mundo. Por lo tanto, queremos asegurarnos de que estamos estableciendo los precios adecuados para nuestros futuros alojamientos en esta ciudad. Por ello, nos enfrentamos al desafío de analizar una muestra de datos de Airbnb de la capital española para entender mejor el mercado.

## **Dataset**

El dataset del que disponemos es una base de datos scrapeada de 14780 registros/alojamientos. Este dataset procede de **Inside Airbnb**, que es un conjunto independiente y no comercial de herramientas y datos para explorar cómo se está utilizando Airbnb en ciudades de todo el mundo. Los datos en los que se basa Inside Airbnb proceden de la información pública disponible en el sitio web de Airbnb.

### **Limitaciones del dataset:**

- La información recopilada corresponde a un período de tiempo entre 2009 y 2017. Estos datos no son actuales ni significativos a día de hoy, debido a que han pasado 6 años desde que fueron tomados por última vez.
- Hay un gran número de valores null por lo que gran parte de los datos se perderán durante el proceso de limpieza, disminuyendo notablemente el dataset.
- Se encuentran datos con errores de tipografía, que aunque sean corregidos en su gran mayoría durante el proceso de limpieza, puede ser que permanezcan sin detectar, provocando errores en el análisis.
- No hay manera de hacer un análisis temporal de los datos debido a la tipología del dataset.

## **Objetivo**

Nuestro **objetivo** es descubrir patrones y tendencias en los precios de los alquileres de Madrid y predecir un precio adecuado para nuestros alojamientos.

Luego de evaluar nuestro objetivo, decidimos plantearnos las siguientes **interrogantes**:

- ¿Cuáles son los vecindarios más costosos y económicos de Madrid?
- ¿Cuáles son las zonas con mayor y menor número de alojamientos?
- ¿Existe relación entre el precio y el número de habitaciones en los alquileres?
- ¿Qué mes es el que tiene el precio promedio de los alquileres más elevado? ¿Y el más bajo?
- ¿Durante qué fechas (día-mes) suelen ser más costosos los alojamientos en Madrid?
- ¿Cuál es el promedio de precios de los alquileres por año?

### **Suposiciones iniciales:**

Los datos a través del análisis nos guiarán y darán respuesta a las incógnitas previamente planteadas, donde se validará o no, nuestras suposiciones iniciales. Con dichos hallazgos, podremos establecer una estrategia de precios efectiva para nuestros alquileres, de manera que maximicen nuestra inversión. Asimismo, esperamos que la información obtenida sea útil para cualquier persona interesada en entender mejor el mercado de los alquileres de Airbnb en la capital española.

1. Los alojamientos de los vecindarios dentro de Madrid Central son los que tendrán el precio promedio más elevado. Mientras que los alojamientos fuera de esta zona y fuera de la M-30 son los que tienen un precio más económico.

Al realizar la visualización de datos (Fig. 1), pudimos observar que en primer lugar con el precio promedio más elevado encontramos el barrio de **El Plantío** (280,00€), que no se encuentra en las afueras de Madrid ciudad. Este barrio destaca por ser uno de los más ricos y lujosos de la comunidad, por lo que sus precios son bastante elevados.

En segundo lugar se encuentra el barrio **Palomas** (142,83€), no ubicado en Madrid Centro, pero si dentro de la ciudad, además se encuentra muy cerca del aeropuerto de Barajas, Parque de Juan Carlos I y del IFEMA Madrid.

Y en tercer lugar encontramos el barrio de **Recoletos** (135,20€), otro barrio costoso de Madrid, este barrio tampoco se ubica en el centro de la ciudad, sin embargo su ubicación es muy privilegiada ya que se encuentran zonas de alto interés turístico: Goya, Serrano, Velázquez, Banco de España y Retiro.

En esta ocasión nuestra hipótesis no era correcta ya que ninguno de los 3 barrios con alojamientos más costosos se encuentran en Madrid Centro.

Por otro lado, los barrios más económicos son: Rosas (17,50€), Orcasitas (18,50€) y Amposta (20,40€), todos estos se encuentran ubicados fuera de la M-30. En este caso, nuestra suposición sí ha sido correcta.

2. El número de alojamientos dentro de Madrid Central será mayor por el atractivo turístico de la zona y por el fácil acceso a distintos métodos de transporte, que permiten realizar viajes de manera rápida y cómoda.

Para obtener los vecindarios con mayor número de propiedades en Madrid, decidimos sumar todos los zipcodes que corresponden a cada uno de los alojamientos (Fig. 2).

Los tres barrios con mayor número de alojamientos de Madrid son: Embajadores, Universidad y Palacios.

- En el barrio **Embajadores** hay diversidad de ofertas gastronómicas, espacios culturales como el Museo Reina Sofía y galerías de arte contemporáneo, además del mercado de antigüedades de El Rastro.
- El barrio **Universidad** queda próximo a zonas como Gran Vía, Callao, Plaza España y Malasaña.
- En cuanto al barrio de **Palacios**, alberga zonas conocidas como el Real Palacio de Madrid, la Catedral de la Almudena, La Latina y la Puerta de Toledo.

Los tres barrios con menor número de alojamientos de Madrid son: Orcasitas, Cuatro Vientos y El Goloso. Son barrios con una extensión muy pequeña, que además se encuentran alejados de las zonas turísticas.

Luego de realizar el análisis, podemos afirmar que nuestra suposición inicial era correcta.

3. Consideramos que los meses en los cuales los alojamientos tendrán precios más elevados son:

- Junio por la celebración del Orgullo LGBTQ+ en Madrid a lo largo de todo el mes.
- Julio debido a que muchos turistas gozan de vacaciones durante estas fechas.
- Octubre por las condiciones climáticas según la información del Ayuntamiento de Madrid.
- Diciembre por el atractivo turístico del alumbrado navideño y los días festivos.

En cambio, el mes más económico para alquilar un alojamiento en Madrid podría ser:

- Agosto, debido al clima caluroso y la preferencia de los viajeros de hospedarse en ciudades costeras.

Asimismo, consideramos que algunas de las fechas más costosas para hospedarse en uno de los alojamientos en Madrid serán:

- 24 y 25 de Diciembre (Navidad)
- 31 de Diciembre (Nochevieja)
- Fechas donde celebran actividades de ocio como por ejemplo: partidos de fútbol, Madcool, Semana del Orgullo, conciertos, Fiesta de San Isidro, etc.

Nuestras suposiciones sobre las fechas y meses más costosos no pueden ser corroboradas, debido a falta de información en nuestro dataset (Fig. 3).

4. Por último, teniendo en cuenta que a mayor espacio mayor será la comodidad del alojamiento, el número de habitaciones elevará el costo del alquiler. Como hemos podido comprobar tras graficar los datos de precio promedio por número de habitaciones (Fig. 4):

- Conforme aumenta el número de habitaciones, el precio promedio de los alojamientos tiende a aumentar, indicando que son directamente proporcionales.
- Hay algunas excepciones como los alojamientos de 6 habitaciones, que son más económicos que los de 5 y 4 habitaciones. Lo mismo ocurre con los alojamientos de

10 habitaciones, que son menos costosos que los de 8 y 7 habitaciones. Esto se puede deber a distintos factores como el tipo de alojamiento o la ubicación del mismo.

### **Métricas seleccionadas**

Las métricas del proyecto:

**Análisis de precios** en función de las siguientes variables: número de habitaciones, número de aseos, número de camas y número de reviews.

**R2** Se ha obtenido un coeficiente de correlación de 0.31, por lo que nuestro modelo a pesar de no ser fiable a la hora de predecir resultados, nos sirve para tomar decisiones basadas en datos.

### **Lecciones aprendidas y qué haríamos de otra forma**

Durante el desarrollo del proyecto intentamos utilizar la variable “Host Since” como fecha de reserva de los alojamientos, debido a que no se dispone de una variable que indique la fecha de inicio y fin de la reserva. Como hemos visto, Host Since hace referencia a la fecha en la que el host se ha dado de alta en Airbnb y no nos proporciona información significativa. Por lo que, para obtener mejores predicciones de precio, en un futuro Airbnb debe proporcionar estos datos (fechas de reserva).

Lastimosamente, al no contar con los datos necesarios, no pudimos responder qué fechas son las más costosas. Para hacer un análisis más completo, podríamos tomar en cuenta distintos factores: festivos, celebración de deportes, atractivos turísticos, cercanía a transporte público o actividades de ocio. Para ello, podríamos solicitar datos actualizados a Airbnb y añadir otras fuentes de datos como las del Ayuntamiento de Madrid.

En el caso del modelo de regresión lineal multivariable, es importante tener en cuenta que se está tratando de modelar una relación compleja entre múltiples variables, por lo que podemos llegar a la conclusión de que con un coeficiente de correlación de 0,31 la aceptabilidad del modelo es moderada-baja.

Además, al existir variables cualitativas importantes que no estén incluidas en el modelo, puede afectar su capacidad para explicar la varianza en los datos. En estos casos, el coeficiente de correlación moderadamente bajo está justificado.

Es importante también considerar que los modelos de regresión lineal multivariable se utilizan a menudo como herramientas para tomar decisiones basadas en datos, y no necesariamente para hacer predicciones precisas. En este caso, el modelo nos serviría para hacernos una idea aproximada del precio que deberíamos ponerle a nuestros apartamentos en 2017, por lo que para una mayor precisión deberíamos trabajar con datos actualizados.

### **Conclusión**

En base a lo expuesto anteriormente, después de haber realizado el análisis y estudio de los datos, podemos tener mayor conocimiento de los precios de los alojamientos en las zonas más populares y de esta forma intentar estimar el precio más adecuado con respecto a la ubicación de la propiedad.

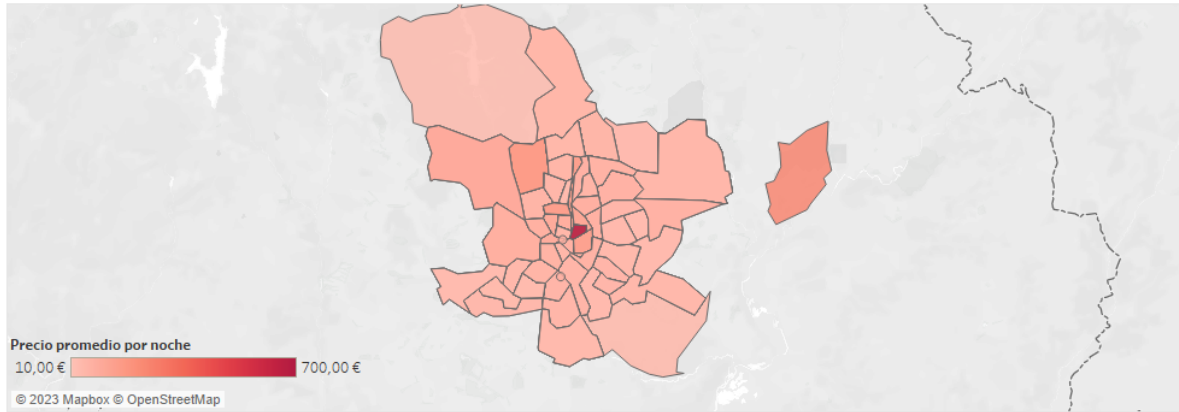
Por otra parte, una de las lecciones más relevantes que hemos aprendido es sin duda la importancia de trabajar con una base de datos completa, limpia y actualizada, ya que durante la exploración de datos se tuvo que excluir columnas cuyos valores pudiesen haber sido relevantes para nuestro proyecto, sin embargo, sus filas al estar prácticamente llenas de valores nulos no las pudimos utilizar. Asimismo, al contar con un dataset que no ha sido actualizado desde el 2017, carecemos de información que no permite realizar un correcto estudio de los precios de los alojamientos de Airbnb en la actualidad.

Para finalizar, consideramos que es fundamental no adelantarse a los resultados ni emitir opiniones personales antes de realizar el análisis de los datos, manteniendo la parcialidad en todo momento, ya que, por ejemplo, en nuestro caso no estuvimos en lo correcto en todas nuestras suposiciones iniciales, por lo que es importante formular preguntas e incógnitas, pero antes de formar nuestras opiniones e intentar responder debemos dejar que los datos hablen y reflejen los hechos.

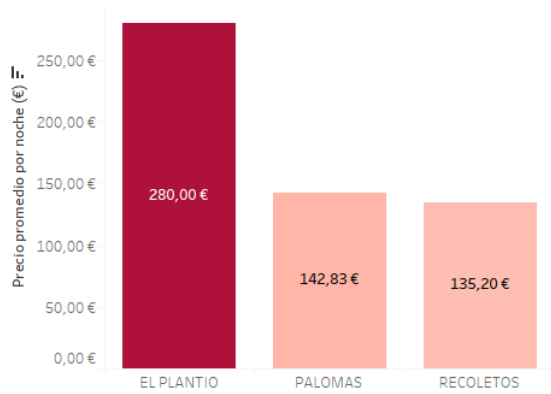
Figura 1

*Dashboard con el mapa de Madrid por código postal, el gráfico de los 3 vecindarios con el precio promedio de alquiler más elevado y el gráfico de los 3 vecindarios con el precio más económico.*

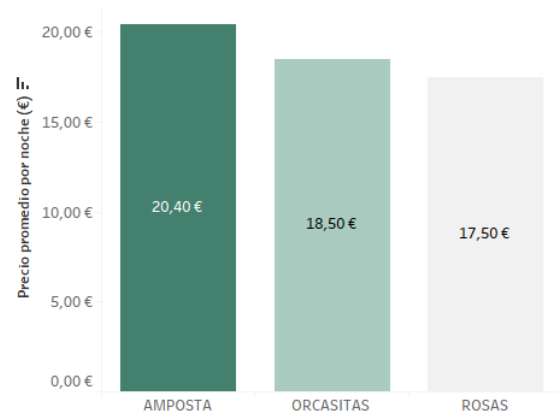
Mapa de Madrid



**Vecindarios más caros (Top 3)**

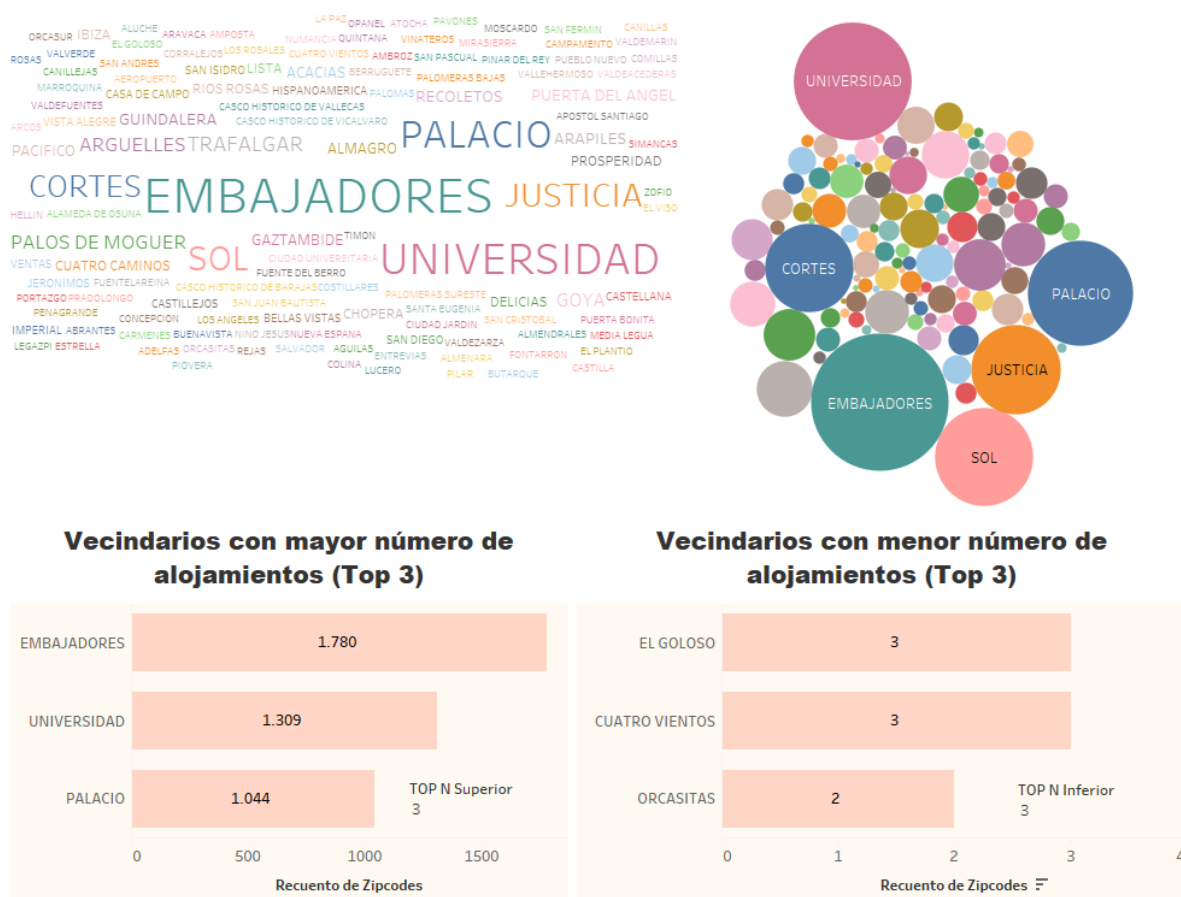


**Vecindarios más baratos (Top 3)**



**Figura 2**

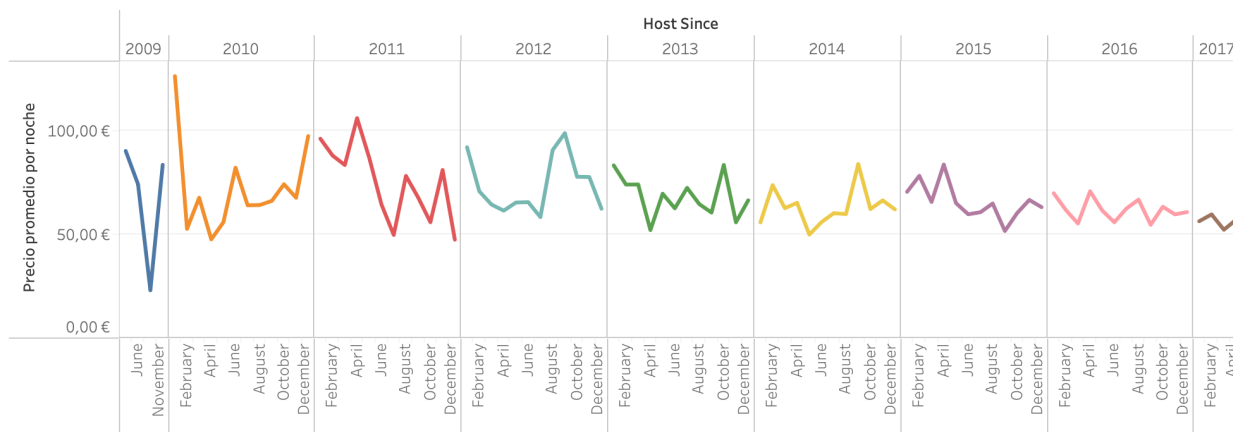
*Dashboards con un gráfico de tipo nube de palabras y otro de tipo burbuja en el que el tamaño del vecindario se corresponde con el número de alojamientos. También se muestra un gráfico con los tres vecindarios con mayor número de alojamientos y otro con los tres vecindarios con menor número de alojamientos.*



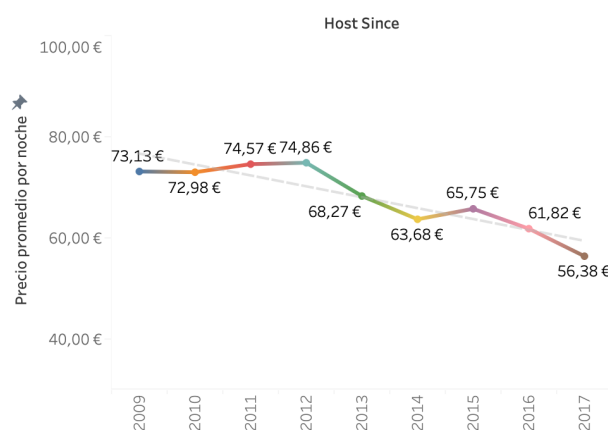
**Figura 3**

Gráfico con los datos de la fecha según “Host since” (fecha desde la que el host se dio de alta en Airbnb) que no nos proporciona información sobre las fechas reales de los alojamientos para poder determinar correctamente los meses más caros.

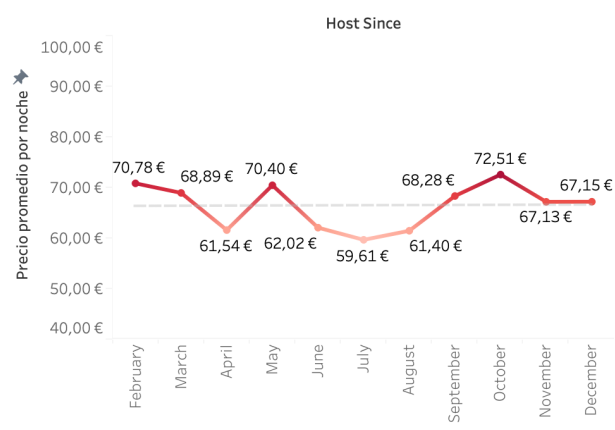
### Precio promedio por año y por mes



### Precio promedio por año



### Precio promedio por mes





**Figura 4**

*Gráfico que muestra cómo varía el precio promedio de los alojamientos según el número de habitaciones que poseen.*

