

Wilson Mario Aranda Vique  
Fabian Emmanuel Osorio Varela  
Juan Esteban Castrillón  
Alejandro Borja

MACC



Universidad del  
**Rosario**

Escuela de Ingeniería, ciencia y tecnología  
Ingeniería de datos  
Javier Casas  
Semestre 2023-2

---

### **Base de datos Airbnb**

Para tener acceso a la base de datos use el siguiente enlace:

<https://www.kaggle.com/datasets/deeplearner09/airbnb-listings/data>

Para entrar al repositorio en github use el siguiente enlace:

<https://github.com/MariOxar/Proyecto-Ing-datos-/projects?query=is%3Aopen>

A continuación, se presentan las reglas de la base de datos:

1. Un anfitrión tiene un único ID.
2. El ID es único para cada hospedaje.
3. El lugar cuenta con un Geocode.
4. El lugar presenta latitudes y longitud en cada hospedaje, pero dos hospedajes no pueden tener la misma latitud y longitud.
5. Cada hospedaje tiene un número mínimo de disponibilidad anual.
6. Solo hay dos tipos de hospedaje habitación/casa completa, los cuales se guardarán en una relación aparte.
7. Cada hospedaje tiene una cierta disponibilidad anual.
8. Cada hospedaje tiene un precio definido, sin importar el tipo de hospedaje.
9. Se tiene una relación “Airbnb” que cuenta con las reseñas, cuantas se han hecho, el número de reseñas anual, el promedio de estrellas y la última reseña realizada.
10. Se debe guardar la información del anfitrión con un id y un nombre.
11. Se debe guardar la información del hospedaje en donde se contemplan el nombre, el id del anfitrión, el número de habitaciones, el precio, el número de baños y el número de camas y la ubicación (es decir el Geocode, la latitud y la longitud).
12. Un anfitrión puede tener varios hospedajes.

13.

Posibles escenarios:

1)

Explorar la distribución geográfica de los hospedajes y verificar si ciertas áreas geográficas tienen una mayor concentración de calificaciones altas.

2)

Analizar si hay algún patrón en la cantidad de reviews por mes. ¿Existen meses con más actividad de reviews que otros?

3)

Investigar si hay alguna relación entre la identidad del anfitrión y las calificaciones de los hospedajes.

¿Los hospedajes de ciertos anfitriones tienden a recibir calificaciones más altas que otros?

4)

Determinar qué tipos de hospedaje son más populares entre los usuarios, basándonos en el número total de reviews y la calificación promedio.

5)

Evaluar la relación entre la disponibilidad y la calificación promedio de los hospedajes.

¿Los lugares más disponibles tienden a tener mejores calificaciones o viceversa?

El uso de cada variable se encuentra expuesto a continuación:

Entidad

- Anfitrión: Es el propietario de la habitación.

## Atributos

- Nombre: Nombre del Anfitrión.
- ID: ID asignado al anfitrión.

## Entidad

- Airbnb: Son los datos almacenados en la plataforma sobre la habitación.

## Atributos

- Número reviews: Número de veces que ha sido calificado en total.
- Reviews por mes: Número de veces que ha sido calificado por mes.
- Ultima review: Fecha de la última vez que ha sido calificada.
- Disponibilidad: Que días está disponible en 365 días.
- Estrellas: promedio de estrellas que ha recibido.

## Entidad

- Hospedaje: Contiene los datos relevantes sobre la habitación/hospedaje.

## Atributos

- Tipo de hospedaje: Si es habitación, aparta estudio, casa completa, etc.
- Jefe: ID del anfitrión del hospedaje.
- ID: ID asignado al hospedaje.
- Nombre: Nombre del hospedaje.
- Habitaciones: Número de habitaciones que tiene el hospedaje.
- Precio: Valor en dólares que cuesta hospedar.
- Baños: Cantidad de baños que tiene el hospedaje.
- Camas: Cantidad de camas que tiene el hospedaje.
- Geocode: Código del barrio donde se encuentra el punto.
- Latitud y Longitud: Códigos de posicionamiento global del punto.

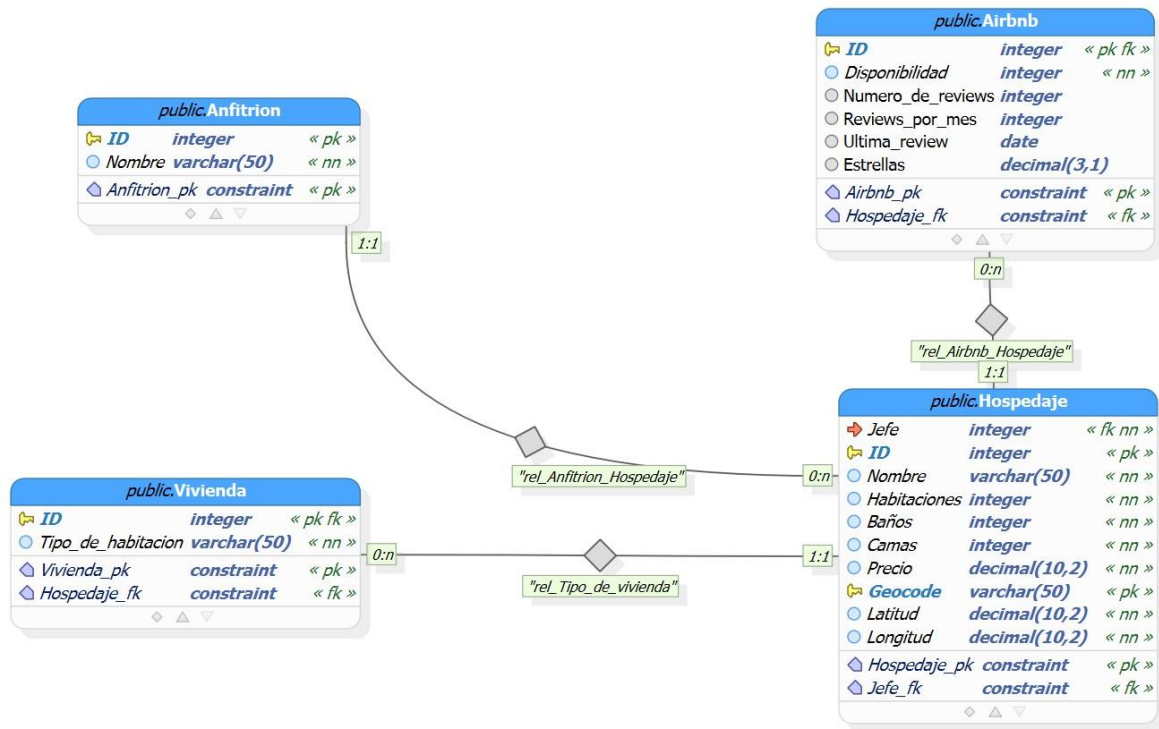
## Entidad

- Vivienda: Contiene el tipo de habitación.

#### Atributos

- Tipo de hospedaje: Si es habitación, aparta estudio, casa completa, etc.
- ID: ID asignado al tipo de hospedaje.

El modelo construido en PgModeler es el siguiente:



Posibles escenarios:

1)

Explorar la distribución geográfica de los hospedajes y verificar si ciertas áreas geográficas tienen una mayor concentración de calificaciones altas.

2)

Analizar si hay algún patrón en la cantidad de reviews por mes. ¿Existen meses con más actividad de reviews que otros?

3)

Investigar si hay alguna relación entre la identidad del anfitrión y las calificaciones de los hospedajes.

¿Los hospedajes de ciertos anfitriones tienden a recibir calificaciones más altas que otros?

4)

Determinar qué tipos de hospedaje son más populares entre los usuarios, basándonos en el número total de reviews y la calificación promedio.

5)

Evaluar la relación entre la disponibilidad y la calificación promedio de los hospedajes.

¿Los lugares más disponibles tienden a tener mejores calificaciones o viceversa?