



airbnb

ANÁLISIS DEL MERCADO EN AMÉRICA LATINA



EQUIPO:

FLORENTINA GALLO

FLORENCIA LÓPEZ

FRANCISCO SCOGNAMILLO

15 de Julio de 2021



INTRODUCCIÓN

Airbnb es una compañía que ofrece una plataforma digital dedicada a la oferta de alojamientos a particulares y turísticos (alquiler vacacional) mediante la cual los anfitriones pueden publicitar y contratar el arriendo de sus propiedades, y tanto anfitriones como huéspedes pueden valorarse mutuamente como referencia para futuros usuarios. El nombre es un acrónimo de *airbed and breakfast*. Airbnb tiene una oferta de unas 2.000.000 propiedades en 192 países y 33.000 ciudades.

En cuanto a su funcionamiento, para poder hacer uso como usuario, lo primero que se debe hacer es registrarse para poder comenzar a buscar a través de distintos filtros como: tipo de habitación, rango de precio, servicios, tipo de alojamiento, etc. Una vez filtrado, lo siguiente es seleccionar el hospedaje que más haya gustado, desde la fachada hasta las comodidades y servicios. La plataforma tiene un sistema de mensajes diseñado para seleccionar fechas y hacer la reserva. Se llenan todos los datos del perfil que se piden, verificando cada uno de ellos para que el anfitrión tenga la seguridad necesaria al hospedar. Al anfitrión se le notificará la solicitud y podrá aceptarla o cancelarla en las próximas 24 horas de la petición.

Para la seguridad, una de las características de Airbnb es que, tanto un anfitrión como un huésped confiable, cuentan con verificación de identidad. Es fácilmente identificable porque aparece un ícono en su perfil.

OBJETIVO

A partir de un set de datos de la compañía buscamos analizar y describir cómo es el mercado de propiedades en esta plataforma en América Latina para los años comprendidos entre 2014 y 2019. Nos centraremos en dos países con disponibilidad de datos para el continente: Brasil y México.

Intentaremos ver qué relación puede existir entre los diferentes tipos de vivienda, las localizaciones, los rangos de precios y las calificaciones que tienen los anfitriones. Podremos ver cómo fueron cambiando los precios y el desempeño de los anfitriones en el tiempo, como así también ver cómo evolucionan las reservas de los diferentes tipos de propiedades.

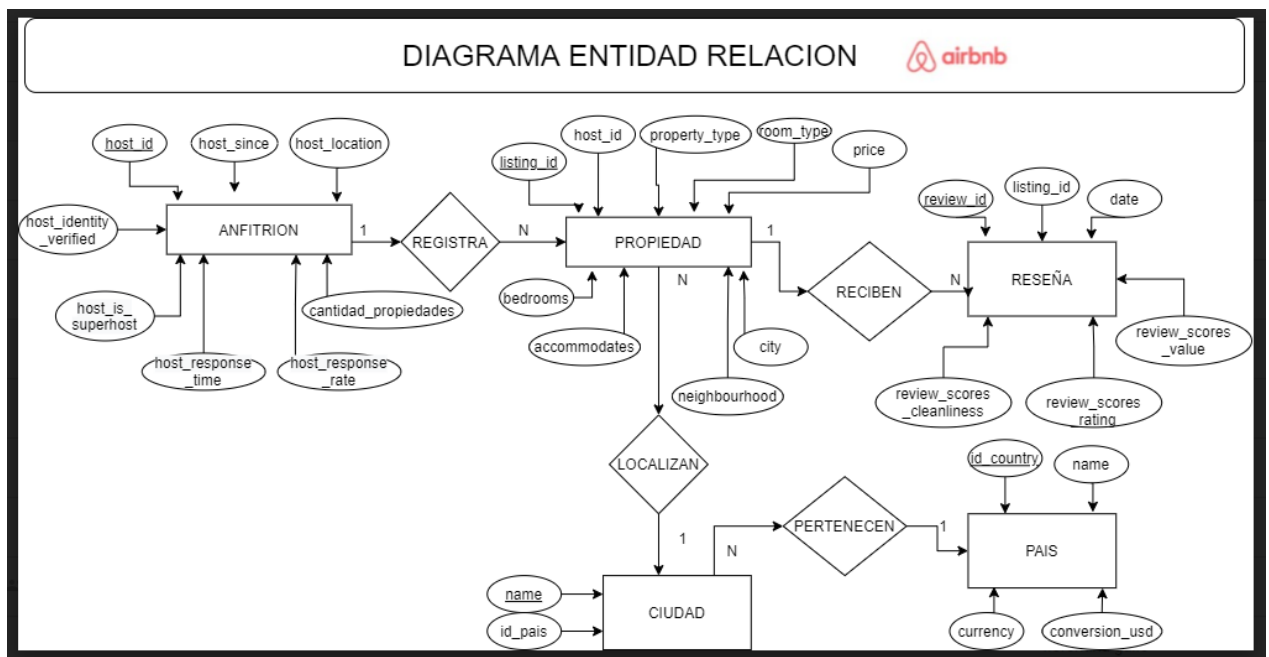
Este trabajo con sus respectivos datos y tableros interactivos permiten realizar un análisis a nivel táctico y estratégico, detectando puntos fuertes y/o débiles o posibles potencialidades para que el negocio pueda seguir desarrollándose a futuro con una mejor performance.

CONTENIDO

1. Diagrama Entidad- Relación (DER)
2. Tablas
3. Descripción de tablas y variables

4. Modelo relacional en Power BI
5. Transformación de datos
6. Análisis funcional del tablero
7. Medidas calculadas
8. Consideraciones a futuro

DIAGRAMA ENTIDAD-RELACIÓN



TABLAS

1) Host (anfitrión):

Host_id (PK) (Int): código de identificación del anfitrión

Host_identity:verified (Varchar): verificación de identidad del anfitrión (inglés)

Host_is_superhost (Varchar): tipo de host (host o superhost)

Host_location (Varchar): ubicación del anfitrión

Host_response_rate (Decimal): rango de respuesta del anfitrión

Host_response_time (Varchar): tiempo de respuesta del anfitrión

Host_since (Datetime): fecha de inicio de actividad del anfitrión

Host_total_listings_count(Int): total de propiedades del anfitrión

Name (Varchar): nombre de publicación

Tipo de anfitrión (Varchar): tipo de anfitrión (anfitrión normal o super anfitrión)

Verificada_NoVerificada (Varchar): verificación de identidad del anfitrión (español)

2) Propiedades:

Accomodate (Int): comodidades de la propiedad

Apartment_type (Varchar): tipo de propiedad

Bedrooms (Int): número de habitaciones de la propiedad

City (FK) (Varchar): nombre de la ciudad

Host_id (FK) (Int): código de identificación del anfitrión

Listing_id (PK) (Int): código de identificación de la propiedad

Neighbourhood (Varchar): barrio de la propiedad

Price (Int): precio de la propiedad por noche

Price_usd: precio de la propiedad por noche en dólares estadounidenses

Price_usd_pc: precio de la propiedad por noche en dólares estadounidenses per cápita

Rango precio (Varchar): medida que evalúa la propiedad según su precio (bajo, medio, alto, muy alto)

Room_type (Varchar): tipo de habitación de la propiedad

3) Ciudades:

City (PK) (Varchar): nombre de la ciudad

Conversion_usd (Decimal): conversión de la moneda a dólar estadounidense

Id_city (Int): código de identificación de la ciudad

Id_country (FK) (Int): código de identificación de país

Population_millions (Decimal): cantidad de habitantes

4) Países:

Conversion_usd (Decimal): conversión de la moneda a dólar estadounidense

Currency (Varchar): divisa o moneda de curso legal del país

Id_country (PK) (Int): código de identificación del país

Name (Varchar): nombre del país

Population_millions(Decimal): cantidad de habitantes

5) Reviews (reseñas):

Anfitrión: medida que relaciona las reseñas con cada anfitrión de la propiedad que fue calificada

Date (Datetime): fecha de la reseña

Listing_id (FK) (Int): código de identificación de la propiedad
Review_id (PK) (Int): código de identificación de la reseña
Reviews_score_accuracy (Int): calificación de precisión de las características de la propiedad
Review_scores_cleanliness (Int): calificación de la limpieza de la propiedad
Review_scores_communication (Int): calificación de la comunicación del anfitrión
Review_scores_location (Int): calificación de la ubicación de la propiedad
Review_scores_rating(Int): calificación general de la propiedad, rating total.
Review_scores_value (Int): calificación precio/calidad de la propiedad
Reviewer_id (Int): código de identificación del usuario que hace la reseña

DESCRIPCIÓN DE TABLAS Y VARIABLES

La tabla **Host** cuenta con los campos necesarios para analizar de forma descriptiva al dueño del alojamiento en alquiler. La misma cuenta con código de identificación del anfitrión, fecha de inicio de actividad, ubicación, tiempo promedio de respuesta, la verificación de la identidad y tipo de anfitrión (host o superhost). La clave primaria es el código de identificación (host_id) a través del cual se vincula con la tabla *Propiedades*.

Por su parte, **Propiedades** lleva el registro de todos los alojamientos presentes en la plataforma. La tabla cuenta con un código de identificación de la propiedad, código de identificación del anfitrión, tipo de apartamento, tipo de habitaciones que se ofrecen en alquiler, cantidad de habitaciones en alquiler, capacidad máxima de huéspedes, precio, ciudad y barrio donde se ubica. La clave primaria es el código de identificación (listing_id) y una clave foránea (host_id) a través de la cual se vincula con la tabla *Host*.

La tabla **Ciudades** se conforma de cuatro variables: código de identificación, nombre, código del país al cual pertenece y población (en millones de habitantes). La clave primaria es el nombre de la ciudad (city) a través de la cual se vincula con la tabla *Propiedades* y la clave foránea es country_id a partir de la cual se vincula con la tabla *Países*.

La tabla **País** está compuesta por el código de identificación, nombre, moneda, conversión en dólares americanos, y población (en millones de habitantes). La clave primaria es el código de identificación (id_country) a través de la cual se vincula con la tabla *Ciudades*.

La tabla **Reviews** es la que brinda información respecto de la evolución del negocio, ya que en ella podemos encontrar los registros de las reseñas realizadas por los huéspedes para el periodo de análisis. Cuenta con los campos fecha, código de identificación de la reseña, código de identificación de quien reseña, código de identificación de la propiedad sobre la que realiza la reseña. A su vez, algunas de las reseñas evalúan limpieza, ubicación, comunicación con el host y un puntaje general en relación al precio y la calidad de la propiedad. La clave primaria es el código de identificación (review_id) y cuenta con dos claves foráneas, la primera es listing_id a través de la cual se vincula con la tabla *Propiedades*, y la segunda es date a través de la cual se vincula con la tabla *Calendario*.

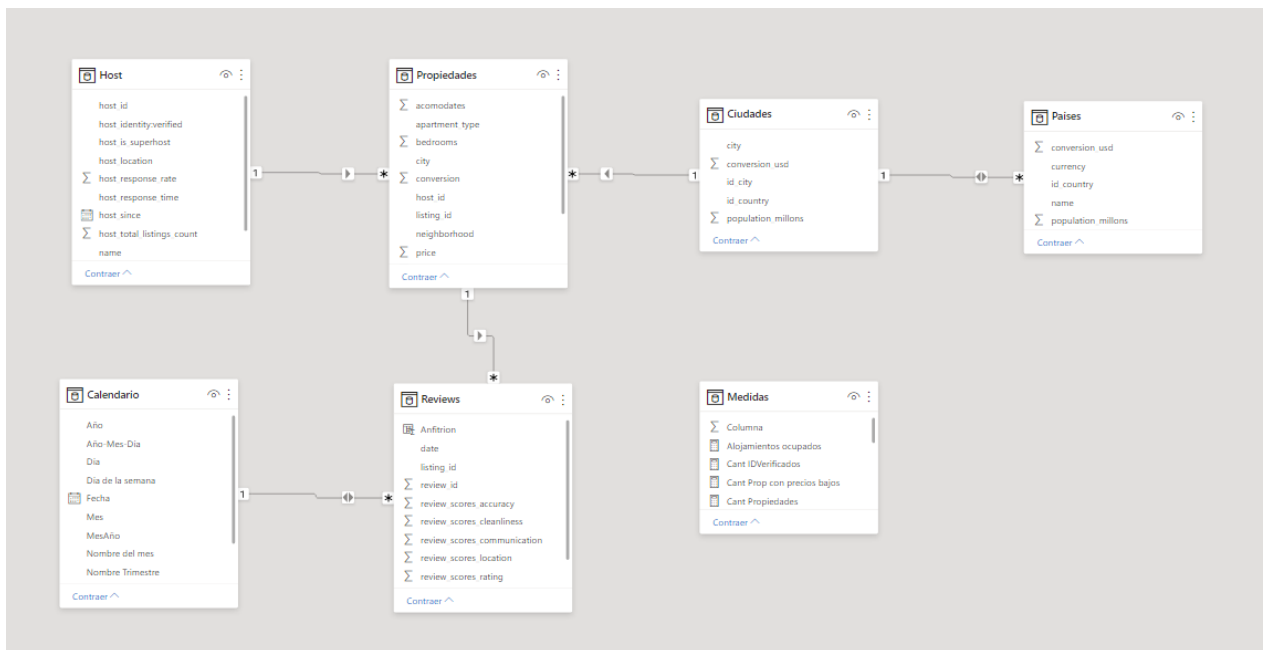
MODELO RELACIONAL EN POWER BI

Para la conexión a la fuente de datos del Proyecto Final se aplicaron los siguientes pasos:

- Origen
- Encabezados Promovidos
- Tipo cambiado

Asimismo, el modelo relacional fue generado de forma automática en Power BI.

A continuación, se adjunta una captura del modelo.



TRANSFORMACIÓN DE DATOS

Pasos aplicados en la tabla **Propiedades**:

- Se reemplazaron los valores “.” por “,” y luego se cambió el tipo de datos de texto a número decimal en la columna Price (precio).
- Luego se cambiaron los tipos de datos del resto campos (los que figuraban como texto a número y fecha según correspondía).
- A través de la función related se buscó la conversión a dólar americano para cada registro, ya que a futuro la intención es estimar el precio en dólares de cada alojamiento.

Pasos aplicados en la tabla **Reviews**:

- Se quitaron columnas en blanco.
- Se cambió el tipo de dato de la columna ‘date’ a fecha.

Pasos aplicados en la tabla **Host**:

- Columna condicional agregada: Tipo de anfitrión (se modificó el texto 't/f' al texto 'Host/Superhost').
- Columna condicional agregada: Identidad Verificada (se modificó el texto 't/f' al texto Verificada/No Verificada).
- Columna condicional agregada: Tipo de anfitrión (se modificó el texto 'Host/Superhost' al texto 'Anfitrión/Super anfitrión')

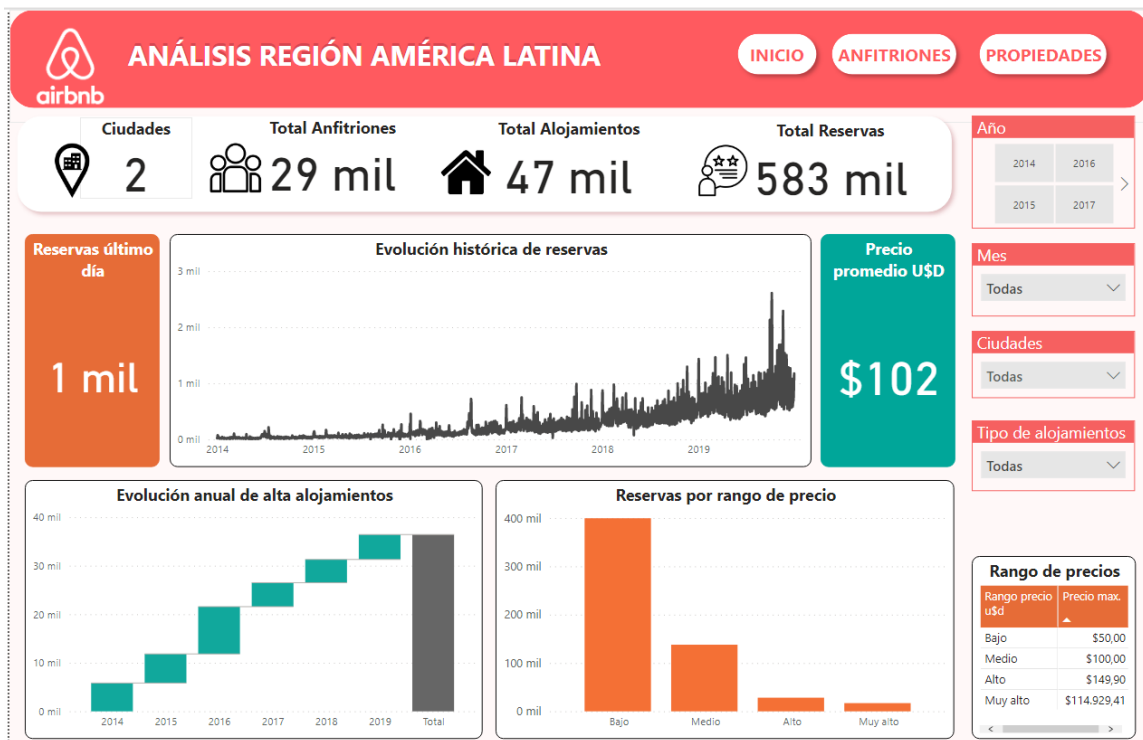
Creación de la **tabla Calendario**: se agregó una tabla Calendario donde estén expresadas las fechas y variables de tiempo (día, mes, año, etc.) en orden de poder realizar un análisis con funciones de inteligencia de tiempo. La misma se realizó a través del lenguaje DAX.

ANÁLISIS FUNCIONAL DEL TABLERO

El tablero se divide en tres partes: Inicio, Anfitriones y Propiedades. Los accesos a estas solapas se encuentran en el extremo superior derecho.

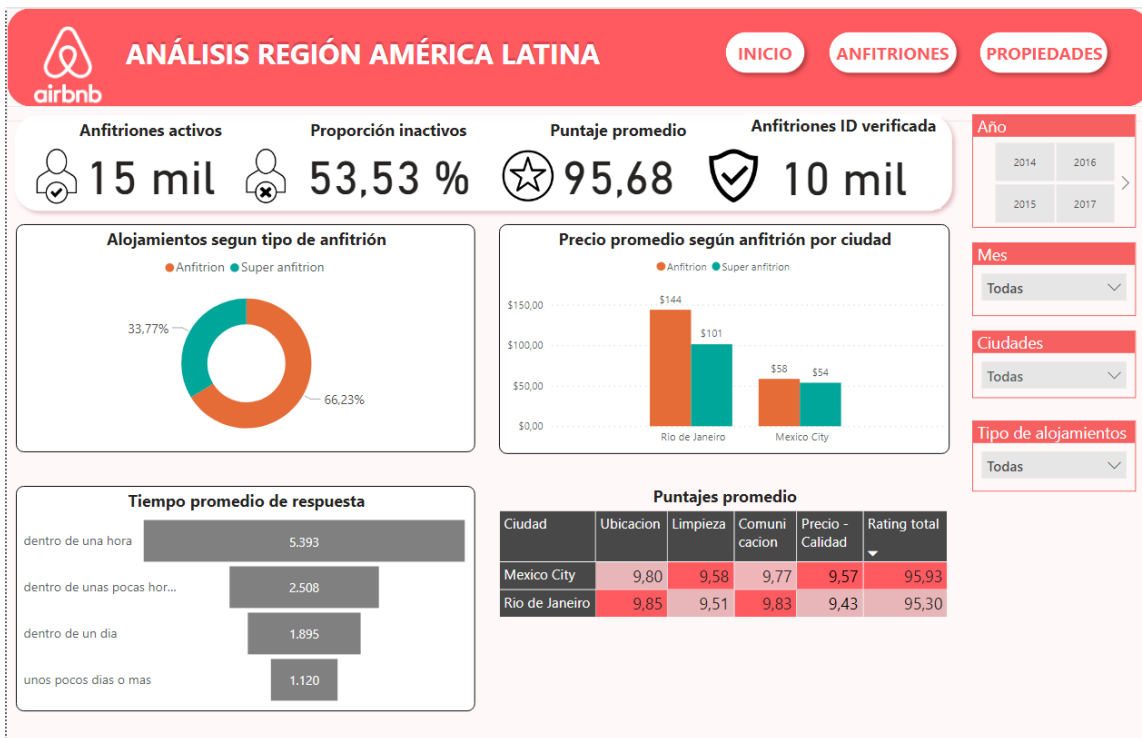
En la solapa **Inicio** se muestran los gráficos y etiquetas de:

- Evolución histórica de reseñas: se muestra cómo evoluciona la cantidad de reseñas realizadas a lo largo de los años en análisis (del 2014 al 2019).
- Evolución anual de alta de alojamientos: muestra las altas de nuevos alojamientos a través de los años.
- Reservas por rango de precio: marca la cantidad de reservas hechas en el total de los años según el segmento de precio al que pertenece la vivienda. Estos segmentos de precios se ven en la gráfica Rango de precios, pudiendo ser la propiedad en cuestión de precio bajo, medio, alto o muy alto dependiendo del costo de su alquiler.
- Precio promedio: es el precio promedio del total de los alojamientos.
- Ciudades: cantidad de ciudades donde hay propiedades para analizar.
- Total anfitriones: número total de anfitriones en la base de datos del tablero.
- Total alojamientos: número total de alojamientos en la base de datos del tablero.
- Total reservas: número total de las reservas realizadas en el período de análisis tomada en cuenta de la base de datos del tablero. Se utiliza como un *proxy* de la variable **reseñas**, dado que para cada reserva que se realiza es obligatorio dejar una reseña una vez brindado el servicio.



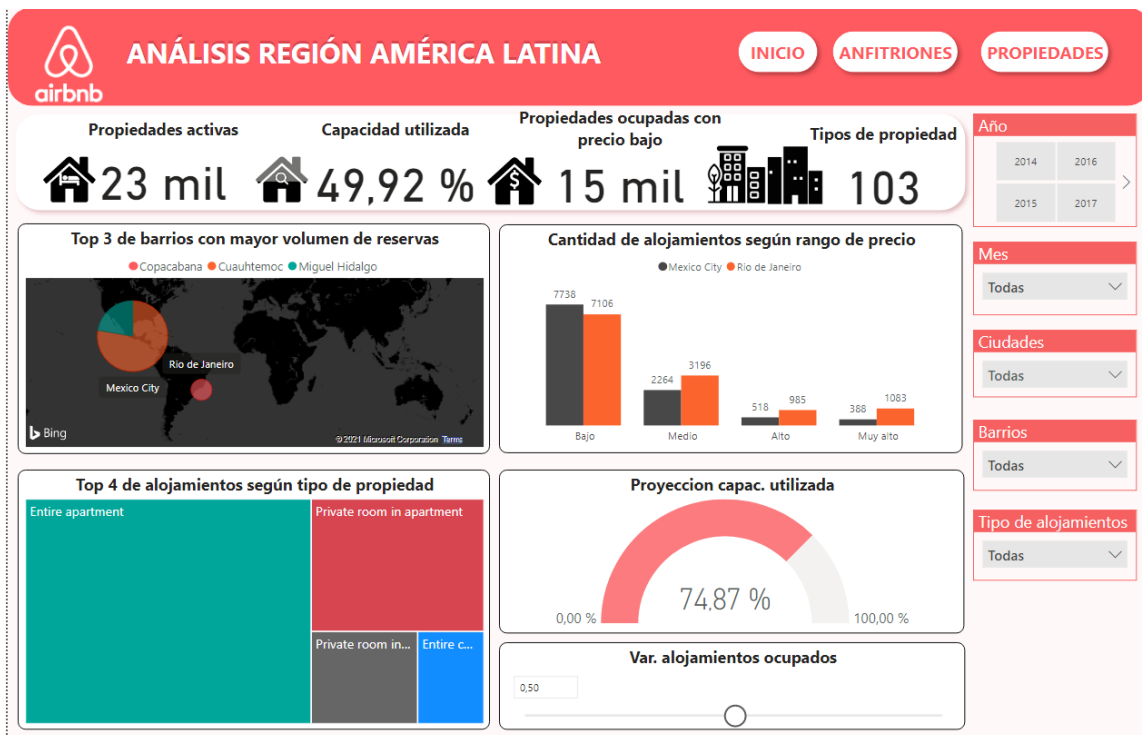
En la solapa **Anfitriones** se encuentran:

- Alojamientos por tipo de anfitrión: muestra qué porcentaje de los alojamientos pertenecen a un Host normal o a un Superhost.
- Precio promedio según anfitrión por ciudad: marca los precios de las propiedades según sea un Superhost o Host normal y al mismo tiempo clasificando por la ciudad donde se encuentra.
- Tiempo promedio de respuesta: expone cada cuánto responden los anfitriones de las propiedades ante diversas consultas.
- Puntajes promedio: tabla que presenta un promedio de los distintos puntajes de las reviews (limpieza, comunicación, ubicación, precisión, precio-calidad, rating total) para calificar a los hosts y sus propiedades, clasificando a su vez por la ciudad que representa cada puntaje.
- Anfitriones activos: cantidad de anfitriones que cuentan con una cuenta activa.
- Proporción inactivos: proporción de anfitriones que poseen una cuenta inactiva.
- Puntaje promedio: puntaje promedio del total de las reseñas.
- Anfitriones con ID verificada: cantidad de anfitriones con su identificación verificada.



En la solapa **Propiedades** se incluye:

- Top 3 de barrios con mayor volumen de reservas: mapa que muestra el volumen de mayor demanda de viviendas según su barrio y ciudad.
- Cantidad de alojamientos por rango de precio: están las cantidades de cada propiedad según el tipo de segmento de precio que corresponda para cada ciudad.
- Top 4 de propiedades según tipo de alojamiento: informa cuáles son las proporciones de los principales tipos de propiedad (apartamento completo, habitación privada en un apartamento, habitación privada en una casa, condominio entero) sobre el total de alojamientos.
- Proyección de la capacidad utilizada: gráfico que explica cómo puede evolucionar la capacidad utilizada del total de las viviendas en ofertas según cómo varía la cantidad de alojamientos ocupados.
- Propiedades activas: número de propiedades que pertenecen a un anfitrión con cuenta activa.
- Capacidad utilizada: proporción de propiedades siendo utilizadas sobre el total de las mismas.
- Propiedades ocupadas con precio bajo: cantidad de propiedades siendo ocupadas que pertenecen al rango de precios bajos.
- Tipos de propiedad: número de tipos de propiedad en la base de datos del tablero.



SEGMENTACIONES

En la columna derecha pueden encontrarse distintos filtros para un mejor análisis de la información. Estos filtros son: *Año*, *Mes*, *Ciudades*, *Tipo de alojamientos* y *Barrios* (este último solo está disponible para la solapa **Propiedades**).

MEDIDAS CALCULADAS

En el tablero existen diecisiete **medidas** calculadas:

1. Alojamientos ocupados: enumera la cantidad de alojamientos ocupados.
2. Cant ID Verificados: enumera la cantidad de hosts con cuentas activas y verificadas.
3. Cant prop ocupadas con precios bajos: muestra la cantidad de alojamientos que presentan un rango de precio menor a 50 dólares.
4. Cant Propiedades: cuenta la cantidad de tipos de propiedades.
5. Capac utilizada: muestra cuál es el porcentaje de propiedades ocupadas sobre el total de propiedades existentes.
6. Ciudades: muestra la cantidad de ciudades en análisis.
7. Hosts activos: explica qué cantidad de hosts tienen una cuenta activa en la plataforma.
8. Nuevas Reseñas: muestra la cantidad de nuevas reservas que se van realizando a lo largo de los meses o incluso años. También puede filtrarse por la ciudad donde se renuevan.
9. Precio promedio: esta medida calcula el promedio de todos los precios de todas las

propiedades de las ciudades en análisis.

10. Precio promedio ocupadas: calcula el promedio de los precios de las propiedades ocupadas en las ciudades.
11. Propiedades per cápita con precio menor o igual a 50 usd: cantidad de propiedades per cápita pertenecientes al rango de precios bajos.
12. Proporción Hosts inactivos: muestra el porcentaje de hosts sin una cuenta activa en la plataforma.
13. Proyección capacidad utilizada: medida que utiliza un parámetro para descubrir cómo puede ser la capacidad utilizada dependiendo de cuánto puede variar la cantidad de alojamientos ocupados.
14. Score promedio: muestra el puntaje promedio de todas las reseñas.
15. Total alojamientos: realiza un conteo del total de alojamientos existentes en ambas ciudades.
16. Total Hosts: cuenta cuántos anfitriones hay en total de la base de datos para ambas ciudades.
17. Total Reseñas: hace el recuento de todas las reservas realizadas para todas las propiedades de las ciudades.

También se creó un **parámetro**: alojamientos ocupados. El mismo puede tomar valores de 0 a 1 para elegir la posible variación en la cantidad de alojamientos ocupados y así ver cómo esto repercute en la capacidad utilizada del total de viviendas.

CONSIDERACIONES A FUTURO

Teniendo en cuenta que solo se tomaron en consideración dos países (Brasil y México) dada la limitación de la disponibilidad de estos datos, a futuro se espera poder seguir incorporando información de diferentes ciudades y también países. Además, podría alargarse el período de análisis, tanto para años más tarde (de 2019 en adelante), como así también años atrás antes del 2014.