

Documentación Proyecto Data Analytics

Proyecto: Airbnb - CABA

Integrantes:
Yamila Galiano
Comisión CH: 43510

Octubre- 2022



Contenido

1. Introducción.
2. Descripción de la temática de los datos.
3. Tipos de análisis.
4. Alcance.
5. Usuario final y nivel de aplicación.
6. Herramientas utilizadas.
7. Datasets.
8. Glosario.
9. Diagrama de entidad-relación.
10. Listado de tablas.
11. Listado de columnas por tabla.
12. Segmentación elegida.
13. Modelo relacional en Power BI.
14. Transformación de tablas.
15. Medidas calculadas.
16. Visualización de datos.
17. Conclusión.



1 - Introducción

Airbnb puede considerarse hoy en día la cadena hotelera más grande y está revolucionando la industria del turismo en todo el mundo.

Airbnb se convirtió en un popular mercado de alquileres en línea y una aplicación móvil que conecta a los anfitriones y huéspedes que buscan un alojamiento alternativo con alojamiento a corto plazo y actividades relacionadas con el turismo.

La compañía ha revolucionado el turismo mundial y está impulsando la competencia con los hoteles tradicionales, al ofrecer una excelente alternativa al servicio de los hoteles caros y de baja calidad.

Ciudad de Buenos Aires

En Buenos Aires, la ciudad podría observar un boom desde 2011, año en que Airbnb aterrizó en Argentina. Más de 900.000 viajeros se alojaron en ubicaciones de Airbnb dentro de Argentina en 2018 y las cifras siguen aumentando a lo largo de los años.

Una búsqueda rápida de ubicaciones de Airbnb en Buenos Aires y sus alrededores arroja cientos de lugares donde los viajeros pueden quedarse.

En el presente trabajo quisimos abordar el crecimiento de los alojamientos de esta zona en particular para abordarlo con más detalles y encontrar tendencias o patrones en toda la información.



2 - Descripción de la temática de los datos.

En el presente trabajo, se obtuvieron los datos de los alojamientos temporarios ofrecidos en la Ciudad de Buenos Aires de la plataforma de Airbnb, en [Inside Airbnb](#). Inside es un conjunto de herramientas y datos independientes, no comerciales, que permiten a la comunidad explorar cómo se utiliza Airbnb en ciudades de todo el mundo.

En particular, se tuvieron en cuenta las características propias de anfitriones ("Host"), su validación en dicha plataforma ("Verification"), las propiedades a alquilar ("Property"), los tipos de alojamientos que se ofertan ("Type room"), el barrio en el que se ubican ("Neighborhood"), los servicios que ofrecen ("Services") y otras variables relativas al puntaje ("Score") y las reseñas que reciben de parte de los usuarios ("Review").

A partir del tipo del modelo denominado "Entidad-Relación" se estructuró y organizó la base de datos escogida en sus respectivas tablas, las cuales se encuentran divididas en solapas.

3 - Tipos de análisis seleccionado.

- Descriptivo

Con base en la pregunta ¿Cuáles son los alojamientos disponibles en CABA según AIRBNB? Se describen los patrones claves en los datos existentes, lo que permite observar la situación actual de las plazas ofrecidas por la plataforma en la Ciudad de Buenos Aires.

Además, entendemos que la presente base de datos permitirá reflejar si hubo aumento en la cantidad de alojamientos disponibles en CABA, según los registros de años anteriores. Con ello, examinar si el eventual crecimiento guarda relación con la problemática existente en torno a los alquileres en la Ciudad. También podremos detectar posibles concentraciones en ciertos sectores de la ciudad como su precio promedio.

Demás tipos de análisis:

- Diagnósticos

A partir de ciertos hitos y momentos donde se cambiaron las leyes de alquiler en alojamientos de larga estadía, podremos validar si se condice con el crecimiento de alquileres temporales.

- Predictivos

Validando la concentración por barrios o zonas de la ciudad, podremos detectar un patrón de zonas de turistas y una tasa de crecimiento promedio.

- Prescriptivos

Esta información podría generar posibles acciones de la ciudad para brindar



mayor seguridad y limpieza en dichas zonas, así como activar un comercio mayor.

4 - Alcance

El dashboard planea ser parte de un análisis táctico y en parte estratégico para el sector turístico.

Táctico, para que los interesados puedan obtener información que impulsen a tomar medidas y proyectar el crecimiento que demande la cantidad de alojamientos cumpliendo con las expectativas de los huéspedes.

Estratégico para analizar si esto puede desencadenar mayores zonas de comercios o nuevos negocios.

5 - Usuario final y nivel de aplicación

El usuario final corresponde al sector de turismo, ya sean arrendadores o analistas del sector turístico. En el mismo podrá obtener información relevante para mejorar sus hospedajes y tener en cuenta cual es el funcionamiento actual del sector turístico.

6 - Herramientas utilizadas

- Google sheet: Para el análisis y limpieza del dataset obtenido en un archivo .csv.
- [Diagrams.net](#): Para realizar el diagrama de entidad-relación.
- Power BI Desktop: Para realizar análisis, transformación de datos y creación de dashboard.
- [Canva](#): Para generar un fondo uniforme.
- [Flaticon](#): Para obtener los iconos de la presentación.

7 - Datasets

Se adjunta el documento Google Sheet con las bases de datos organizadas y estructuradas utilizadas en el presente trabajo. [Link](#)

Dicha información ha sido extraída de la web [Inside Airbnb](#) en un archivo .csv para su posterior transformación en excel.

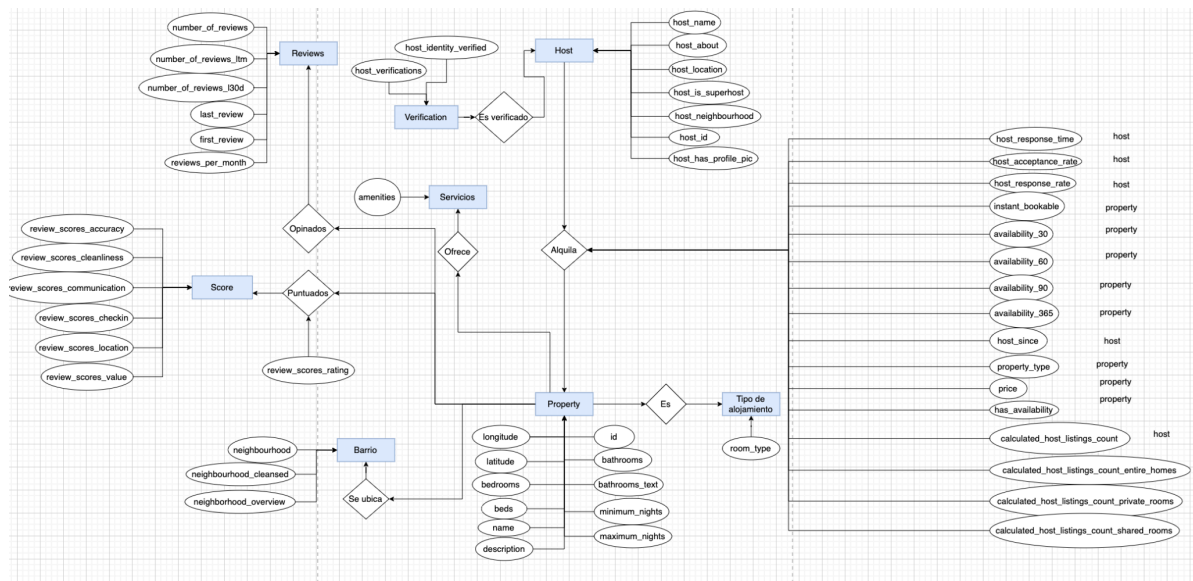
Se quitaron datos irrelevantes para este análisis como imágenes del perfil del arrendatario, links a la publicación, imágenes de la propiedad, entre otros.

8 - Glosario

El glosario se encuentra detallado en el siguiente archivo [google sheet](#).

9 - Diagrama de entidad-relación

Para validar el diagrama entidad-relación con mayor detalle, ingresar al siguiente [link](#).



10 - Listado de tablas

- Host:

Esta tabla contiene los datos del propietario como su id, nombre, ubicación, tipo de host, tiempo de respuesta, cantidad de propiedades y su fecha de alta en la plataforma. Clave PK: host_id.

- Verification:

Esta tabla contiene las verificaciones realizadas por el host/propietario de identidad, email, teléfono y mail de trabajo. Clave PK: id_verification. Clave FK: host_id.

- Property:

Esta tabla contiene información de la propiedad como la ubicación, baños, habitaciones, precio, entre otros. Clave PK: id_property. Clave FK: host_id, id_room.

- Type_room:



Esta tabla contiene el tipo de habitación de la propiedad: Entire places, Private rooms y Shared rooms. Clave PK: id_room.

- **Neighborhood:**

Detalle del barrio en donde se encuentra ubicada la propiedad. Clave PK: id_neighbourhood. Clave FK: id_property.

- **Score:**

Puntuación de la propiedad en distintos aspectos: checkin, comunicación, limpieza, ubicación, costo-beneficio y veracidad. Clave PK: id_score. Clave FK: id_property.

- **Review:**

Cantidad de comentarios que tiene una propiedad, por parte de sus huéspedes. Clave PK: id_review. Clave FK: id_property.

- **Services:**

Tabla con la descripción de los servicios disponibles en la propiedad. Clave PK: id_services. Clave FK: id_property.

Más detalle en el siguiente [link](#).

11 - Listado de columnas por tabla

A continuación se muestra el detalle de columnas por tabla. Para más información ingresar al Link antes mencionado de google sheet.



Host		
Nombre Columna	Tipo de datos	Tipo de clave
host_id	Int	Clave PK
host_name	Varchar	
host_about	Varchar	
host_location	Varchar	
host_is_superhost	Bit	
host_neighbourhood	Varchar	
host_response_time	Varchar	
host_response_rate	Decimal	
host_acceptance_rate	Decimal	
host_since	Datetime	
calculated_host_listings_count	Int	
calculated_host_listings_count_entire_homes	Int	
calculated_host_listings_count_private_rooms	Int	
calculated_host_listings_count_shared_rooms	int	
host_has_profile_pic	Bit	

Property		
Nombre Columna	Tipo de datos	Tipo de clave
id_property	Int	Clave PK
latitude	Int	
longitude	Int	
bathrooms	Decimal	
bathrooms_text	Varchar	
bedrooms	Int	
beds	Int	
minimum_nights	Int	
maximum_nights	Int	
instant_bookable	Bit	
availability_30	Int	
availability_60	Int	
availability_90	Int	
availability_365	Int	
property_type	Varchar	
price	Decimal	
has_availability	Bit	
host_id	Int	Clave FK
name	Varchar	
description	Varchar	
id_room	Int	Clave FK

Verification		
Nombre Columna	Tipo de datos	Tipo de clave
id_verification	Int	Clave PK
host_id	Int	Clave FK
host_verifications	Varchar	
host_identity_verified	Bit	

Type_room		
Nombre Columna	Tipo de datos	Tipo de clave
id_room	Int	Clave PK
type_room	Varchar	

Score		
Nombre Columna	Tipo de datos	Tipo de clave
id_score	Int	Clave PK
id_property	Int	Clave FK
review_scores_rating	Decimal	
review_scores_accuracy	Decimal	
review_scores_cleanliness	Decimal	
review_scores_checkin	Decimal	
review_scores_communication	Decimal	
review_scores_location	Decimal	
review_scores_value	Decimal	

Neighborhood		
Nombre Columna	Tipo de datos	Tipo de clave
id_neighborhood	Int	Clave PK
id_property	Int	Clave FK
neighbourhood_cleaned	Varchar	
neighbourhood	Varchar	
neighborhood_overview	Varchar	



Review		
Nombre Columna	Tipo de datos	Tipo de clave
id_review	Int	Clave PK
id_property	Int	Clave FK
number_of_reviews	Int	
number_of_reviews_ltm	Int	
number_of_reviews_l30d	Int	
first_review	Datetime	
last_review	Datetime	
reviews_per_month	Decimal	

Services		
Nombre Columna	Tipo de datos	Tipo de clave
id_services	Int	Clave PK
id_property	Int	Clave FK
amenities	Varchar	

12 - Segmentación elegida

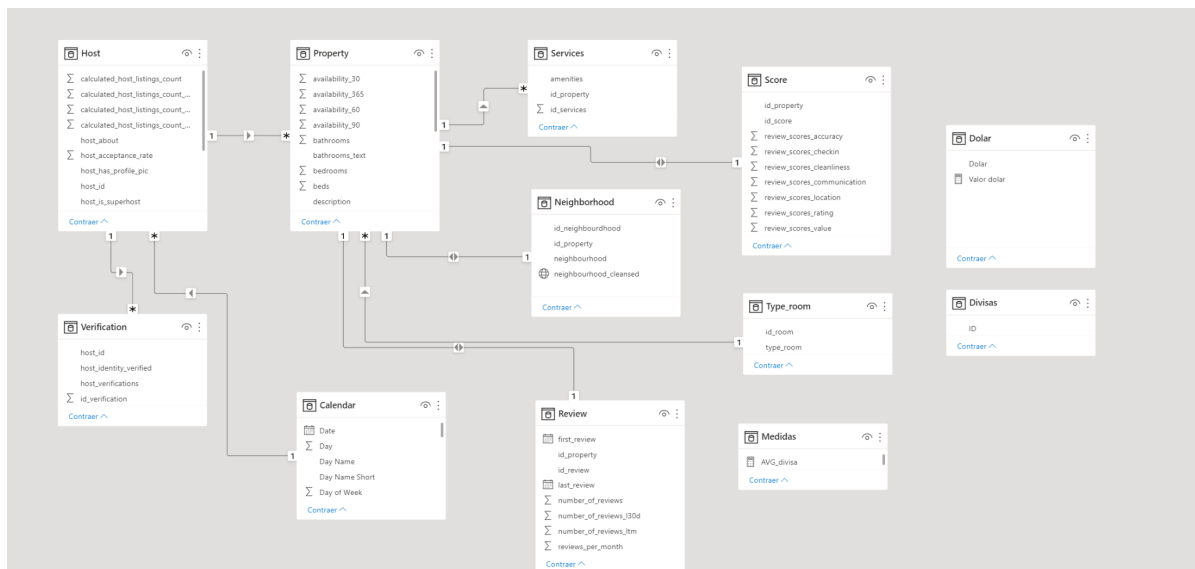
La segmentación elegida en todos los gráficos es por:

- Tipo de propiedad.
- Barrio.

Luego en algunos casos especiales se seleccionó filtros como:

- Tipo de divisa/moneda.
- Puntuación de la propiedad.
- Período de tiempo - Año.

13 - Modelo relacional en Power BI



14 - Transformación de tablas

- Tabla Host:

Se realizó una limpieza de valores únicos antes de subir esta tabla a Power BI.



Se cambió el tipo de dato "host_is_superhost" a lógico (True o False) y con esto se generó una nueva columna llamada 'Superanfitrión' para mostrar los datos más concisos, "host_response_rate" a porcentaje, "host_acceptance_rate" a porcentaje, "host_since" a fecha, "calculated_host_listings_count" a integer, "calculated_host_listings_count_entire_homes" a integer, "calculated_host_listings_count_private_rooms" a integer, "calculated_host_listings_count_shared_rooms" a integer, "host_has_profile_pic" a lógico.

- Tabla Neighborhood:

Se eliminó la columna neighborhood_overview ya que no era utilizada.

- Tabla Property:

Se modificó el tipo de dato a nivel regional de price, se asignó el tipo de dato latitud y longitud a sus respectivas columnas.

Se modificaron los tipos de datos de las columnas "bathrooms" a número, "bedrooms" a número, "beds" a integer, "minimum_nights" en integer, "maximum_nights" a integer, "instant_bookable" a lógico, "availability_30" a integer, "availability_60" a integer, "availability_90" a integer, "availability_365" a integer, "has_availability" a lógica.

- Tabla Review:

Se modificó el tipo de dato de los siguientes campos: {"number_of_reviews", Int64.Type}, {"number_of_reviews_ltm", Int64.Type}, {"number_of_reviews_l30d", Int64.Type}, {"first_review", type date}, {"last_review", type date}, {"reviews_per_month", type number}}.

- Tabla Score:

Se modificó el tipo de dato de los siguientes campos: {"review_scores_rating", type number}, {"review_scores_accuracy", type number}, {"review_scores_cleanliness", type number}, {"review_scores_checkin", type number}, {"review_scores_communication", type number}, {"review_scores_location", type number}, {"review_scores_value", type number}}.

- Tabla Services:

Se modificó la columna Amenities ya que vienen delimitadas por coma.

Primero se dividió columnas por delimitador, cambiamos a tipo texto, se anuló la dinamización de columnas, se quitó la columna 'Atributos' generada y por último se agregó un índice nuevo.

- Type_room:

Se modificaron los tipos de campos: {"id_room", Int64.Type}, {"type_room", type text}}.

- Verification:

Se modificó la columna Host_verifications ya que vienen delimitadas por coma.

Primero se dividió columnas por delimitador, cambiamos a tipo texto, se anuló la dinamización de columnas, se quitó la columna 'Atributos' generada y por último se agregó un índice nuevo.

Se quitó los espacios en blanco de la columna host_verifications.



- Tabla Calendario: Se generó a partir del scrip provisto por el profe y se relaciono con la fecha de registro de cada host o anfitrión.

15 - Medidas calculadas

- **AVG_divisa** = SWITCH(
TRUE();
SELECTEDVALUE(Divisas[ID])="ARG"; [AVG_precio_ARG];
SELECTEDVALUE(Divisas[ID])="USD"; [AVG_precio_USD])

Esta medida permite utilizar el filtro de moneda o divisa y que el precio cambie según la cotización.

- **AVG_host_by_year** = [Total_host] / DISTINCTCOUNT('Calendar'[Year])

Esta medida permite delimitar la cantidad de nuevos anfitriones por año y realizar un promedio.

- **AVG_precio_ALL** = AVERAGEX(ALL('Property');('Property'[price]))

Esta medida permite mostrar un promedio del precio total en pesos argentinos, sin filtros.

- **AVG_precio_ALL_USD** = AVERAGEX(ALL('Property');('Property'[price]))/[Valor dolar]

Esta medida permite mostrar un promedio del precio total en dólares, sin filtros.

- **AVG_precio_ARG** = AVERAGE('Property'[price])

Promedio simple de precio en peso argentino, adecuado a los filtros del tablero o gráfico.

- **AVG_precio_USD** = AVERAGE('Property'[price])/Dolar[Valor dolar]

Promedio simple de precio en dólares, adecuado a los filtros del tablero o gráfico.

- **Estrellas** =

```
VAR __MAX_NUMBER_OF_STARS = 5
VAR __MIN_RATED_VALUE = 0
VAR __MAX_RATED_VALUE = 5
VAR __BASE_VALUE = SUM('Score'[review_scores_value])
VAR __NORMALIZED_BASE_VALUE =
    MIN(
        MAX(
            DIVIDE(
                __BASE_VALUE - __MIN_RATED_VALUE;
                __MAX_RATED_VALUE - __MIN_RATED_VALUE
            );
            0
        );
        1
    )
```



```

VAR __STAR_RATING = ROUND(__NORMALIZED_BASE_VALUE
*__MAX_NUMBER_OF_STARS; 0)
RETURN
    IF(
        NOT ISBLANK(__BASE_VALUE);
        REPT(UNICHAR(9733); __STAR_RATING)
        & REPT(UNICHAR(9734); __MAX_NUMBER_OF_STARS
__STAR_RATING)
    )

```

Esta medida se generó de manera automática para mostrar la cantidad de estrellas, respecto al puntaje o score de cada propiedad.

- **Sum_price** = SUM('Property'[price])

Medida generada para sumarizar el precio, acorde a los filtros del entorno.

- **TOP_max_price** = RANKX(ALLSELECTED('Property'[id_property]);[Sum_price])

Se utilizó para realizar un ranking de propiedades dependiendo el barrio que tenían seleccionados. Esto puede verse en el gráfico de 'Propiedades más caras' de la hoja 'Mapa'.

- **TOP_min_price** =
IF(NOT(ISBLANK([Sum_price]));
RANKX(FILTER(ALLSELECTED('Property'[id_property]);
NOT(ISBLANK([Sum_price])));
[Sum_price];ASC))

Se utilizó para realizar un ranking de propiedades dependiendo el barrio que tenían seleccionados. Esto puede verse en el gráfico de 'Propiedades más baratas de la hoja 'Mapa'.

- **Total_host** = COUNTX(ALL(Host);Host[host_id])

Muestra el total de host o anfitriones sin tener en cuenta los filtros del entorno.

- **Total_properties** = COUNTX(ALL('Property');'Property'[id_property])

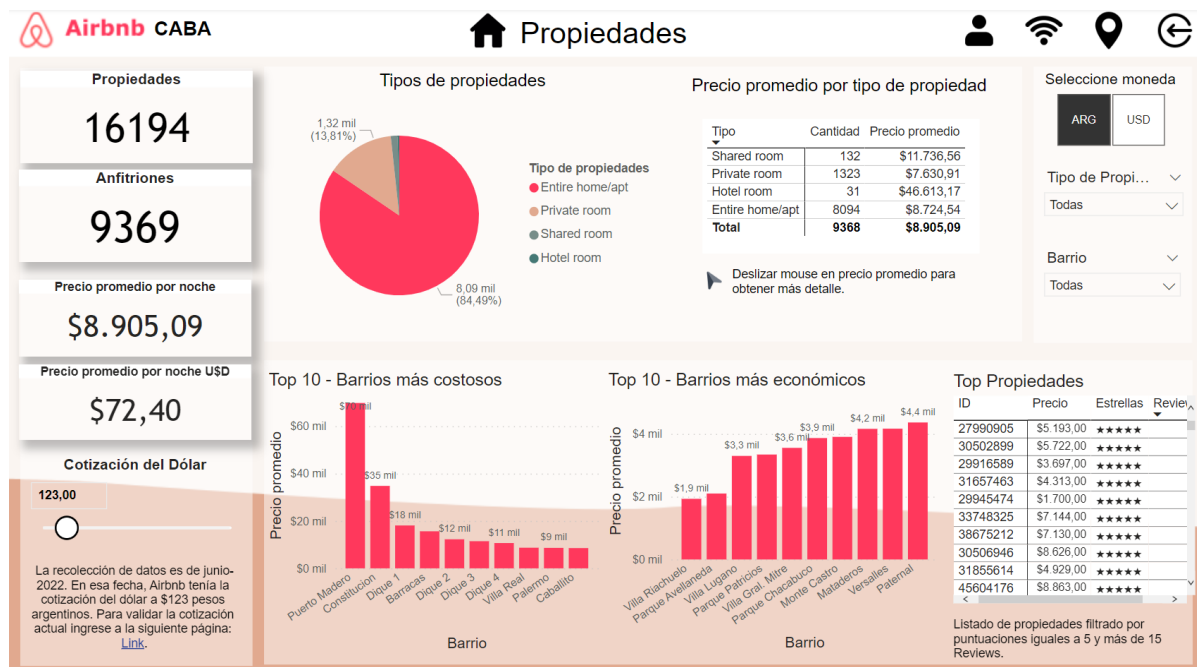
Muestra el total de propiedades sin tener en cuenta los filtros del entorno.



16 - Visualización de datos



1 - Portada: Se desarrolló esta página para mostrar el inicio del tablero, dar una introducción a los posibles grandes temas del dashboard. Tiene acceso a las demás páginas.



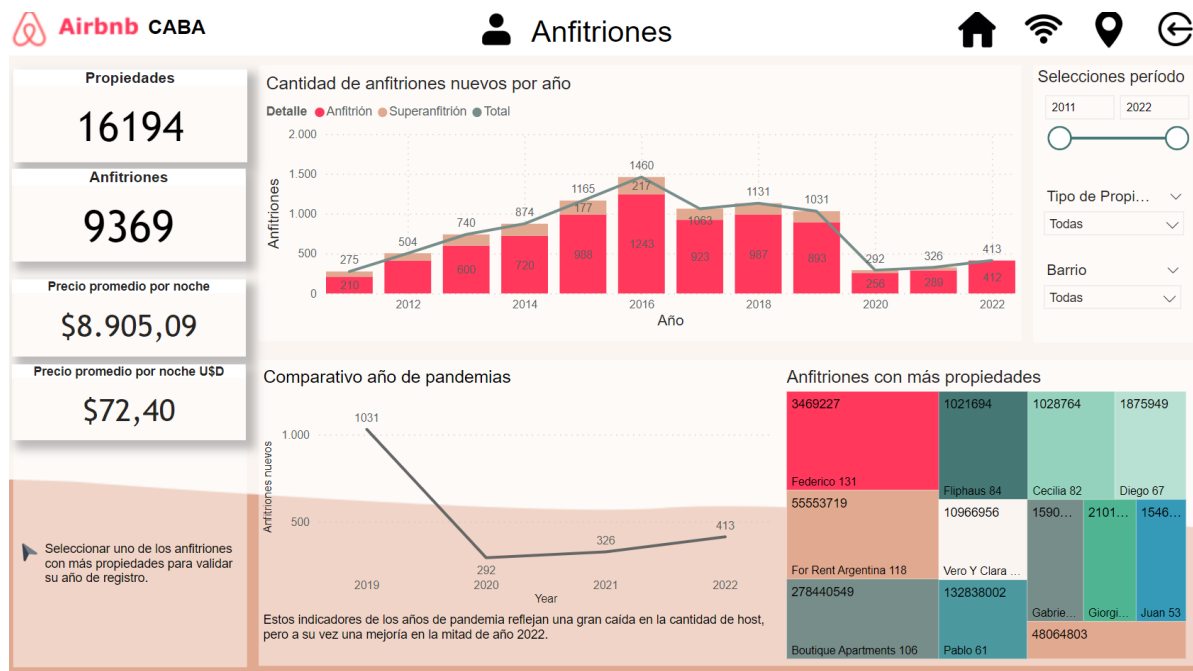
2 - Propiedades: Esta página quiere demostrar la distribución de datos por medio de los tipos de propiedades. Muestra la cantidad de cada uno de ellos y un precio promedio de cada tipo.



Además tiene un filtro de tipo de divisa, lo que permite validar el precio en peso argentino y dólar, como la cotización utilizada en el dashboard que se puede modificar.

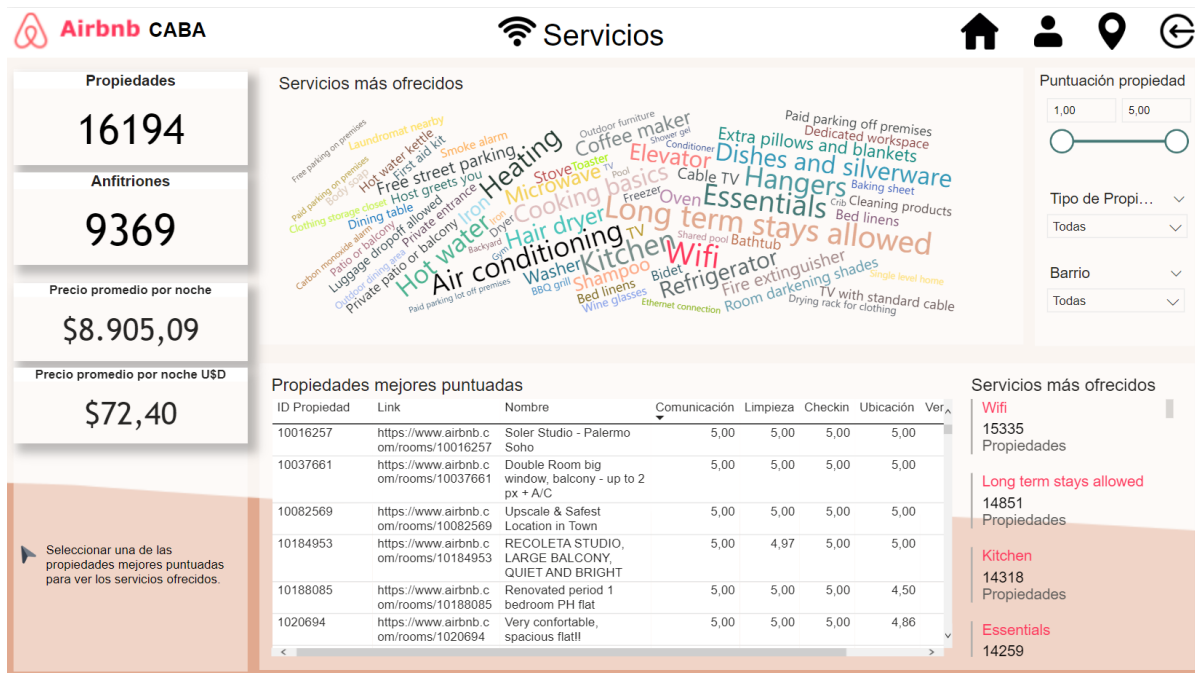
Se muestran además los precios más altos y bajos por barrio con la opción de filtrado por tipo de propiedad y Barrio.

Se muestran los principales KPI que se repiten en cada hoja como el número de propiedades, el número total de anfitriones, precio promedio por noche en peso argentino y en dólares.

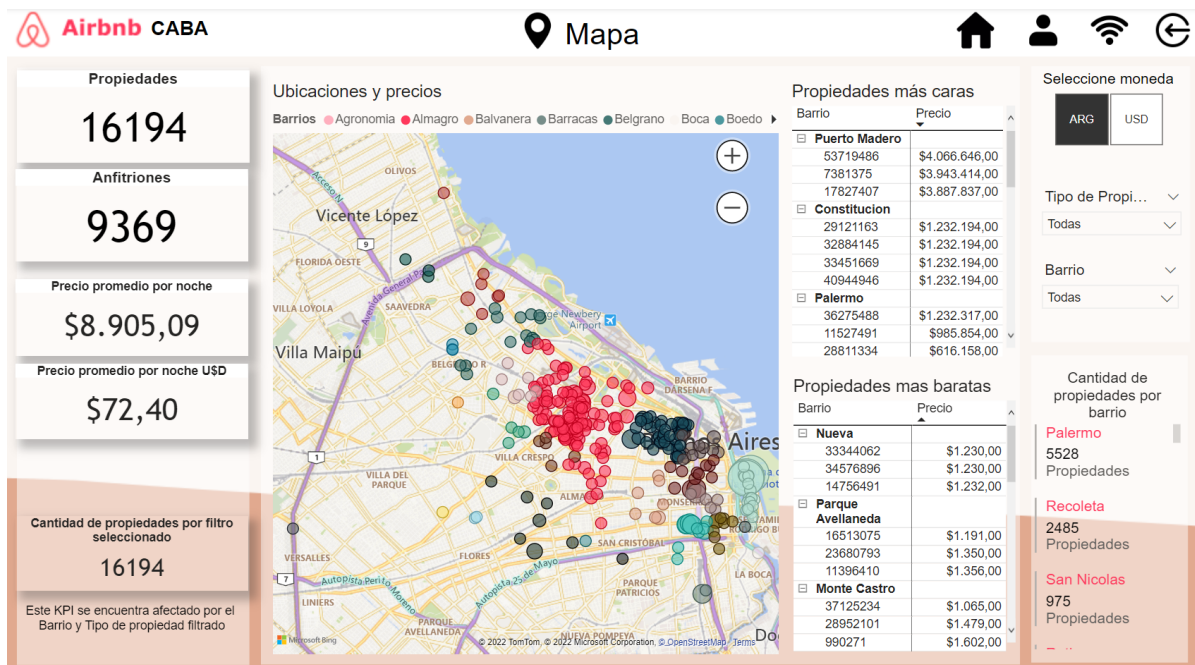


3 - Host o Anfitriones: Esta página permite validar cómo ha ido cambiando el registro de anfitriones al pasar de los años y aquellos que más cantidad de propiedades tienen en Airbnb. Como interacción se puede seleccionar uno de los anfitriones y ver su año de registro.





4 - Servicios: Es una hoja que quiere mostrar los servicios más ofrecidos por las propiedades y que realmente son un esencial por esa misma razón. Tenemos luego una tabla con los id y link de las propiedades mejor puntuadas en los aspectos evaluados en Airbnb. Como interacción extra, podemos seleccionar una de las propiedades y ver los servicios que ofrece.



5 - Mapa: Muestra el detallado de propiedades en el mapa, intentando mostrar donde hay mayor concentración, como el tamaño de la burbuja dependiendo de su precio. Al seleccionar cualquiera de las burbujas será posible validar el precio en peso argentino o dólar, filtrando la moneda.



Además podemos tener un mayor detalle del top propiedades más caras por barrio y las más baratas.

Como último detalle tenemos la cantidad de propiedades por barrio.

17 - Conclusión

Como conclusión podemos decir que se han obtenido datos relevantes en el análisis del dashboard tales como:

- El tipo de propiedad u hospedaje más ofrecido y por ende más solicitado es de un 'Entire home/apt', el cual dispone de un lugar solo para el huésped, sin compartir espacios comunes.
- Palermo es el barrio con más concentración de hospedajes, por lo que se puede impulsar un crecimiento de negocios en el área.
- Los precios más altos y como era de esperarse, se ofrecen en Puerto Madero, por lo que podemos delimitar la clase monetaria de huéspedes que puede recibir dicha zona a la hora de plantear nuevos negocios u nuevos hospedajes.
- Hasta el día de la fecha el año que más registros de anfitriones ha tenido es el 2016, pero lo más drástico a tener en cuenta fue la cantidad de anfitriones por debajo del promedio que se registraron en la pandemia y como esto tuvo un efecto en la oferta de hospedaje. Este dato se encuentra repuntando en el año 2022.
- Los servicios más requeridos y ofrecidos son el Wifi. Ya que estamos en un era moderna que genera la demanda de cantidad de megas en el servicio, es importante que los anfitriones tengan en cuenta que calidad de servicio van a contratar y ofrecer.

Por último consideramos que Buenos Aires tiene un crecimiento mucho mayor esperado y con los valores y detalles mencionados podemos proyectar ese crecimiento de la mejor manera posible.

