



Análisis General

Buenos Aires

New York

Singapur



Elaborado por Juan Ignacio Lerena Para el curso Data Analytics de CoderHouse

Contenido - Índice

1. Descripción de la Temática de Datos. Pág 2
2. Objetivos Del Proyecto. Pág 2
3. Nivel de alcance, usuario final y nivel de aplicación. Pág 2
4. Herramientas utilizadas. Pág 3
5. Listado de Tablas. Pág 3
6. Diagrama de Entidad – Relación. Pág 4
7. DataSet. Pág 5
8. Transformaciones realizadas. Pág 8
9. Medidas calculadas generadas y fórmulas. Pág 9
10. Tablero. Pág 10

1. DESCRIPCIÓN DE LA TEMÁTICA DE DATOS

En la tema elegida se abordarán temáticas relacionadas a la estadísticas obtenidas con respecto a los alquileres de AirBnb en tres grandes ciudades elegidas de diferentes continentes.

Los datos se obtuvieron de <https://www.kaggle.com/>, conformados por tres datasets diferentes, uno para cada ciudad. En todos, los parámetros a analizar son los mismo lo que dará facilidad para la comparación de las estadísticas entre las ciudades.

Se busca ver si existe igual correlación para todas las ciudades, entre los precios de los alquileres (por persona) y la cantidad de días al año que fueron alquilados, es decir si el porcentaje de ocupación anual es similar para las diferentes ciudades en sus respectivas escalas de precios.

2. OBJETIVOS DEL PROYECTO

El objetivo principal del proyecto es tomar los datos obtenidos del DataSet para una final toma de decisiones de inversión. Para ellos se tendrán en cuenta todos los datos de las ciudades para analizar rango de precios, porcentaje de ocupación de los hosts, cantidad de hosts y distribución por barrios, evaluación de los tipos de hosts en cada ciudad; y finalmente si existe o no relación, causalidad o correlación entre las diferentes ciudades.

3. NIVEL DE ALCANCE, USUARIO FINAL Y NIVEL DE APLICACIÓN

Los resultados obtenidos serán de utilidad tanto para miembros de la aplicación como parte propietaria para evaluar posibles postulaciones de sus propiedades en la plataforma; así como también es estudio por parte de AirBnb para la futura toma de decisiones sobre posibles refuerzos en las campañas de marketing, lanzamientos de promociones y descuentos para alquileres, entre otras medidas que aumenten el nivel tanto de uso de la App en la ciudad como el porcentaje de ocupación de los Hosts postulados.

4. HERRAMIENTAS UTILIZADAS

Para el trabajo presentado se utilizaron las siguientes herramientas informáticas:

- Microsoft Excel para la visualización y preparado de las tablas preliminares.
- ERDPlus para la creación del diagrama entidad-relación.
- Microsoft Power BI Desktop para la creación del tablero de control y visualización.

5. LISTADO DE TABLAS

- Tabla de Reservas: En esta tabla encontraremos todos los datos correspondientes a cada reserva que se hizo, con su respectivo Host y las características de este.

Tipo de Clave	Campo	Tipo de Campo	Detalle
PK	id_reserva	INT	Número único de identificación de reserva
FK	id_host	INT	Número único de identificación del Host
	ciudad	VARCHAR(20)	Ciudad correspondiente
FK	id_barrio	VARCHAR(10)	Código identificador del barrio
	barrio	VARCHAR(20)	Barrio correspondiente
	latitud	DECIMAL (5,5)	Ubicación de la propiedad
	longitud	DECIMAL (5,5)	Ubicación de la propiedad
	tipo_host	VARCHAR(20)	Tipo clasificadorio de la propiedad
	precio	INT	Precio de la reserva
	minimo_noches	INT	Mínimo de noches a reservas
	numero_resenias	INT	Cantidad de reseñas
	ultima_resenia	DATE	Fecha de la última reseña
	reseñas_por_mes	INT	Promedio de reseñas por mes
	suma_host_dueño	INT	Cantidad de Host que administra
	ocupacion_año	DECIMAL (5,5)	Promedio de noches por año de ocupación

- Tabla de Propiedades: En esta tabla encontraremos estrictamente la información relacionada con la posición y el tipo de propiedad del cual se trata.

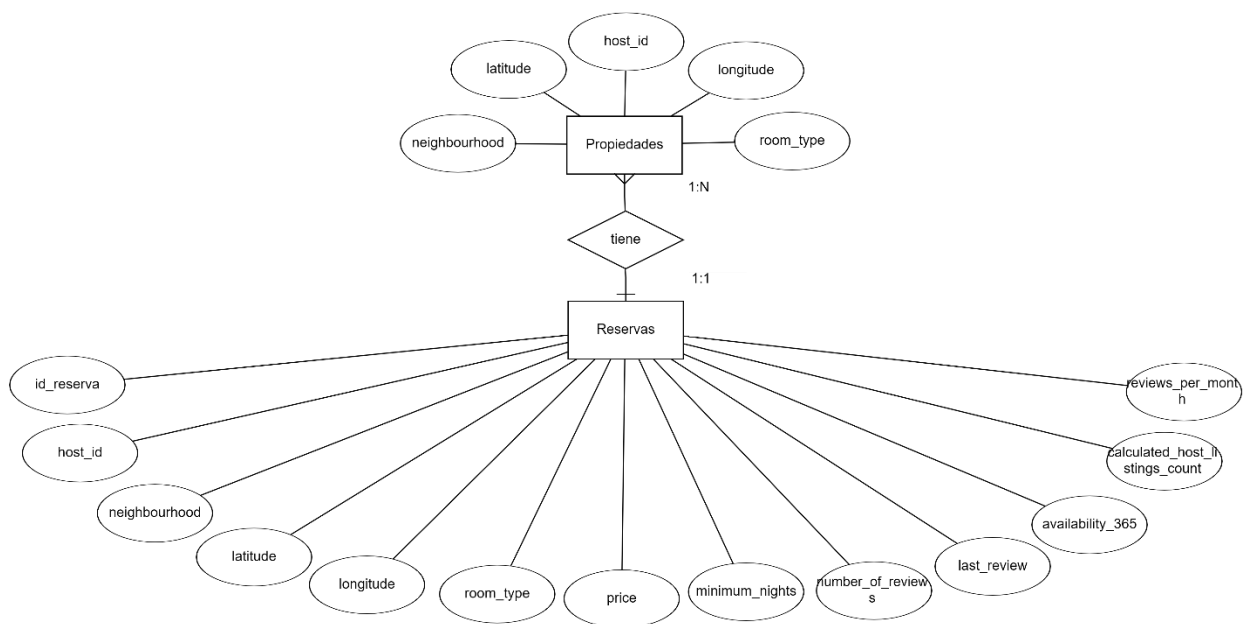
Tipo de Clave	Campo	Tipo de Campo	Detalle
PK	id_host	INT	INT
	ciudad	VARCHAR(20)	VARCHAR(20)
	id_barrio	VARCHAR(10)	VARCHAR(10)
	barrio	VARCHAR(20)	VARCHAR(20)
	tipo_host	VARCHAR(20)	VARCHAR(20)

- Tabla de Barrios: Es esta tabla se encuentran identificados todos los barrios con su código correspondiente y ciudad a la cual pertenecen.

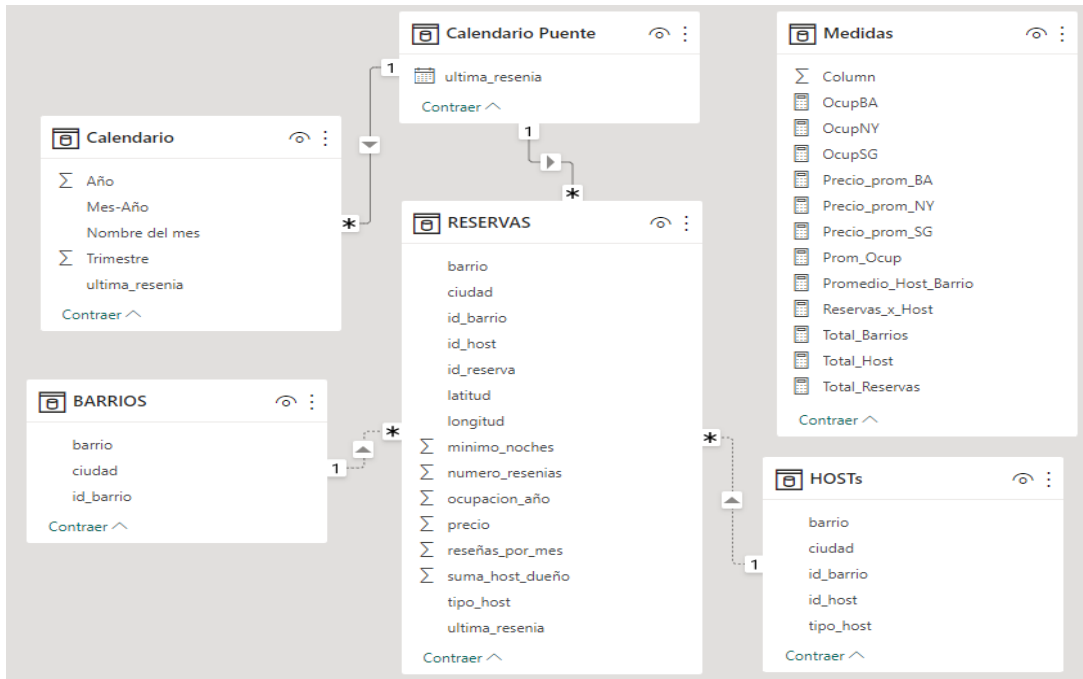
Tipo de Clave	Campo	Tipo de Campo	Detalle
PK	id_barrio	VARCHAR(10)	VARCHAR(10)
	barrio	VARCHAR(20)	VARCHAR(20)
	ciudad	VARCHAR(20)	VARCHAR(20)

6. DIAGRAMA DE ENTIDAD – RELACIÓN

6.1.1 DIAGRAMA INICIAL



6.1.2 DIAGRAMA FINAL



7. DATASET

7.1.1 VISTA INICIAL DEL DATASET

En esta primera vista al descargar los DataSet de la página antes mencionada, se observaban tres archivos Excel distintos (uno para cada ciudad) los cuales contaban con una sola solapa con todas las columnas de datos.

S13																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	id	name	host_id	host_name	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_availability	365	
2	7270	Amazing Great Loft	18822	Toni	Palermo	-34.58241	-58.43382	Entire home/apt	2033	20	14	30/04/2018	0.15	1	363	
3	11508	Amazing Loft	42762	Candela	Palermo	-34.58225	-58.4246	Entire home	3588	2	25	13/08/2019	0.28	1	281	
4	12463	Room in Recoleta	48799	Matt	Recoleta	-34.59761	-58.39468	Private room	1435	1	20	11/12/2017	0.17	1	365	
5	13095	Standard Room	50994	Ariel	Palermo	-34.59382	-58.42994	Private room	2691	1	1	18/12/2018	0.09	7	360	
6	13096	Standard Room	50994	Ariel	Palermo	-34.59398	-58.42853	Private room	2691	1	0			7	360	
7	13097	Standard Room	50994	Ariel	Palermo	-34.59381	-58.43007	Private room	2691	1	65	19/10/2019	2.19	7	358	
8	13099	Lofty room	50994	Ariel	Palermo	-34.59362	-58.42984	Private room	3588	1	0			7	343	
9	13100	Master Room	50994	Ariel	Palermo	-34.59288	-58.4301	Private room	4185	1	1	23/11/2017	0.04	7	354	
10	13349	Room #301	52100	Yanina	Montserrat	-34.61564	-58.38823	Private room	4185	1	0			6	365	
11	13350	Room #302	52100	Yanina	Montserrat	-34.6178	-58.3864	Private room	3588	1	0			6	365	
12	13351	Hotel boutique	52100	Yanina	Montserrat	-34.61715	-58.38832	Private room	4185	1	0			6	365	
13	13352	Room #304	52100	Yanina	Montserrat	-34.61575	-58.38667	Private room	3887	1	0			6	365	
14	13367	Hotel Boutique	52100	Yanina	Montserrat	-34.61704	-58.38673	Private room	3588	1	0			6	365	
15	14222	RELAX IN HAVANA	55825	Manuel	Palermo	-34.58777	-58.41106	Entire home	2332	5	87	03/11/2019	0.97	3	359	
16	14711	SPECTACULAR	57770	Cintia	Montserrat	-34.60913	-58.37372	Entire home	7175	1	0			1	365	
17	15074	ROOM WITH	59338	Monica	Nuñez	-34.53724	-58.4658	Private room	1794	29	0			1	365	
18	15889	ROOM WITH	62183	Sonia	Balvanera	-34.6029	-58.4114	Private room	2093	3	6	01/06/2014	0.06	2	365	
19	16695	DUPLEX LOFT	64880	Elbio Marian	Montserrat	-34.61385	-58.37706	Entire home	3887	1	45	26/10/2019	0.42	12	360	
20	16696	Duplex Loft	64880	Elbio Marian	Montserrat	-34.61469	-58.37421	Entire home	2990	1	50	16/07/2019	0.45	12	359	
21	16697	Small Duplex	64880	Elbio Marian	Montserrat	-34.61427	-58.37488	Entire home	2691	1	22	16/04/2019	0.21	12	342	

7.1.2 VISTA FINAL DEL DATASET

A continuación se muestra como quedó finalmente, luego de algunas modificaciones, el DataSet en Excel para la próxima carga y análisis en Power BI.

Unificado_AirBnB.xlsx - Excel																
Q5																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	id_reserva	id_host	ciudad	id_barrio	barrio	latitud	longitud	tipo_host	precio	minimo_noches	numero_resenas	ltima_resen	reseñas_por_mes	suma_host_dueño	ocupacion_año	
2	11508	HBA0009	Buenos Aires	BA021	Palermo	-34.58225	-58.4246	Entire home/apt	3588	2	25	13/08/2019	0.28	1	281	
3	12463	HBA0010	Buenos Aires	BA028	Recoleta	-34.59761	-58.39468	Private room	1435	1	20	11/12/2017	0.17	1	365	
4	13095	HBA0011	Buenos Aires	BA021	Palermo	-34.59382	-58.42994	Private room	2691	1	1	18/12/2018	0.09	7	360	
5	13096	HBA0011	Buenos Aires	BA021	Palermo	-34.59398	-58.42853	Private room	2691	1	0			7	360	
6	13097	HBA0011	Buenos Aires	BA021	Palermo	-34.59381	-58.43007	Private room	2691	1	65	19/10/2019	2.19	7	358	
7	13099	HBA0011	Buenos Aires	BA021	Palermo	-34.59362	-58.42984	Private room	3588	1	0			7	343	
8	13100	HBA0011	Buenos Aires	BA021	Palermo	-34.59288	-58.4301	Private room	4185	1	1	23/11/2017	0.04	7	354	
9	13349	HBA0012	Buenos Aires	BA017	Montserrat	-34.61564	-58.38823	Private room	4185	1	0			6	365	
10	13350	HBA0012	Buenos Aires	BA017	Montserrat	-34.6178	-58.3864	Private room	3588	1	0			6	365	
11	13351	HBA0012	Buenos Aires	BA017	Montserrat	-34.61715	-58.38832	Private room	4185	1	0			6	365	
12	13352	HBA0012	Buenos Aires	BA017	Montserrat	-34.61575	-58.38667	Private room	3887	1	0			6	365	
13	13367	HBA0012	Buenos Aires	BA017	Montserrat	-34.61704	-58.38673	Private room	3588	1	0			6	365	
14	14222	HBA0016	Buenos Aires	BA021	Palermo	-34.58777	-58.41106	Entire home/apt	2332	5	87	03/11/2019	0.97	3	359	
15	14711	HBA0018	Buenos Aires	BA017	Montserrat	-34.60913	-58.37372	Entire home/apt	7175	1	0			1	365	
16	15074	HBA0020	Buenos Aires	BA019	Nuñez	-34.53724	-58.4658	Private room	1794	29	0			1	365	

En esta preparación se dejaron todos los datos en un solo archivo Excel. Se separaron los datos de cada ciudad en tres solapas, dejando una principal con todos los datos de las reservas, y otras dos secundarias con los datos específicos de los Hosts y de los Barrios de cada ciudad.

7.2.1 METADATA INICIAL

- "id": referencia el código único de cada reserva.
- "name": referencia el nombre asignado al host al cual pertenece la reserva.
- "host_id": es el código único identificador del host de la reserva.
- "host_name": se refiere al nombre del administrador del Host correspondiente.
- "neighbourhood": nombre del barrio al cual pertenece el Host.
- "latitude": latitud de la ubicación del Host.
- "longitude": longitud de la ubicación del Host.
- "room_type": se refiere al tipo de Host según la clasificación asignada por AirBnb.
- "price": es el importe al cual se registró la reserva correspondiente.
- "minimum_nights": número que indica el número mínimo de noches que se alquila ese Host.
- "number_of_reviews": indica la cantidad de referencias que tiene el Host.
- "last_review": es la fecha de la última reseña que obtuvo el Host.
- "reviews_per_month": es un promedio calculado de reseñas que ha obtenido el Host desde el inicio de su actividad al momento de la extracción de los datos.
- "calculated_host_listings_count": este número indica la cantidad de Hosts que administra la misma persona encargada de la administración del Host correspondiente a la reserva en cuestión.
- "availability_365": este indicador nos muestra el promedio de noches por año que estuvo alquilado el host desde el inicio de su actividad al momento de la extracción de los datos.

7.2.2 METADATA FINAL

Se realizaron cambios en los nombres asignados a las columnas para una mejor identificación. Y se añadieron algunas para un mejor análisis dada la unificación de los archivos.

- "id" → "id_reserva"
- "host_id" → "id_host"
- "ciudad": referencia la ciudad a la que pertenece el Host.

- "id_barrio": se asignó un código único identificador a los barrios de cada ciudad.
- "neighbourhood" → "barrio"
- "latitude" → "latitud"
- "longitude" → "longitud"
- "room_type" → "tipo_host"
- "price" → "precio"
- "minimum_nights" → "minimo_noches"
- "number_of_reviews" → "numero_resenias"
- "last_review" → "ultima_resenia"
- "reviews_per_month" → "resenias_por_mes"
- "calculated_host_listings_count" → "suma_host_dueño"
- "availability_365" → "ocupación_año"

8. TRANSFORMACIONES REALIZADAS

8.1 TRANSFORMACIONES INICIALES

- Inicialmente y luego de realizar los arreglos antes mencionados se crea la tabla "Medidas", como contenedor de todas las medidas luego calculadas.
- Se revisan todos los tipos de datos de los campos, encontrando que eran acordes con la variable que albergan.
- Se utilizó como color primario el color característico de la marca AirBnb, cuyo código hexadecimal es #FF5A5F.
- Se creó la tabla Calendario, donde se visualizan las fechas de las últimas reseñas recibidas. Esto se logró duplicando la tabla principal, dejando solo la columna "ultima_resenia" y luego se eliminaron los duplicados. Posteriormente, y en base a la columna obtenida se desmembraron las columnas que indican de cada fecha, el año, el nombre del mes, el número de trimestre y por último, en formato "MM/AAAA".

9. MEDIDAS CALCULADAS GENERADAS Y FÓRMULAS

Todas las medidas calculadas se hicieron en la tabla "Medidas". Se crearon las siguientes medidas:

- "OcupBA" que indica el promedio de ocupación de todos los Hosts de Buenos Aires.

```
OcupBA = CALCULATE(AVERAGE(RESERVAS[ocupacion_año]),RESERVAS[ciudad]="Buenos Aires")
```

La misma fórmula pero indicando las distintas ciudades (New York y Singapur) se usó para calcular el promedio de ocupación de estas dos, siendo "OcupNY" y "OcupSG" los nombres de las medidas obtenidas.

- "Precio_Prom_BA" - "Precio_Prom_NY" - "Precio_Prom_SG" : Estos son los nombres asignados para el cálculo del precio promedio para cada ciudad. Para una mejor comparación y dado que solo los datos de Buenos Aires estaban expresados en otra moneda, se supuso un tipo de cambio de 1USD=100PESOS para poder expresar todas bajo una misma denominación.

```
Precio_prom_SG = CALCULATE(AVERAGE(RESERVAS[precio]),RESERVAS[ciudad]="Singapur")
```

```
Precio_prom_NY = CALCULATE(AVERAGE(RESERVAS[precio]),RESERVAS[ciudad]="New York")
```

```
Precio_prom_BA = CALCULATE((AVERAGE(RESERVAS[precio])/100),RESERVAS[ciudad]="Buenos Aires")
```

- "Promedio_Host_Barrio" es una medida que indica el promedio de Host por barrio. Se toma la cantidad de Host y se divide por la cantidad de Barrios. Luego aplicando una segmentación podemos obtener el dato para cada ciudad.

```
Promedio_Host_Barrio = CALCULATE((COUNT(HOSTs[id_host]))/(COUNT(BARRIOS[id_barrio])))
```

- "Reservas_x_Host" es una medida que indica el promedio de la cantidad de reservas que tiene cada Host. Se obtiene dividiendo el total de reservas por el total de Host. Con la segmentación podemos ver el promedio para cada ciudad.

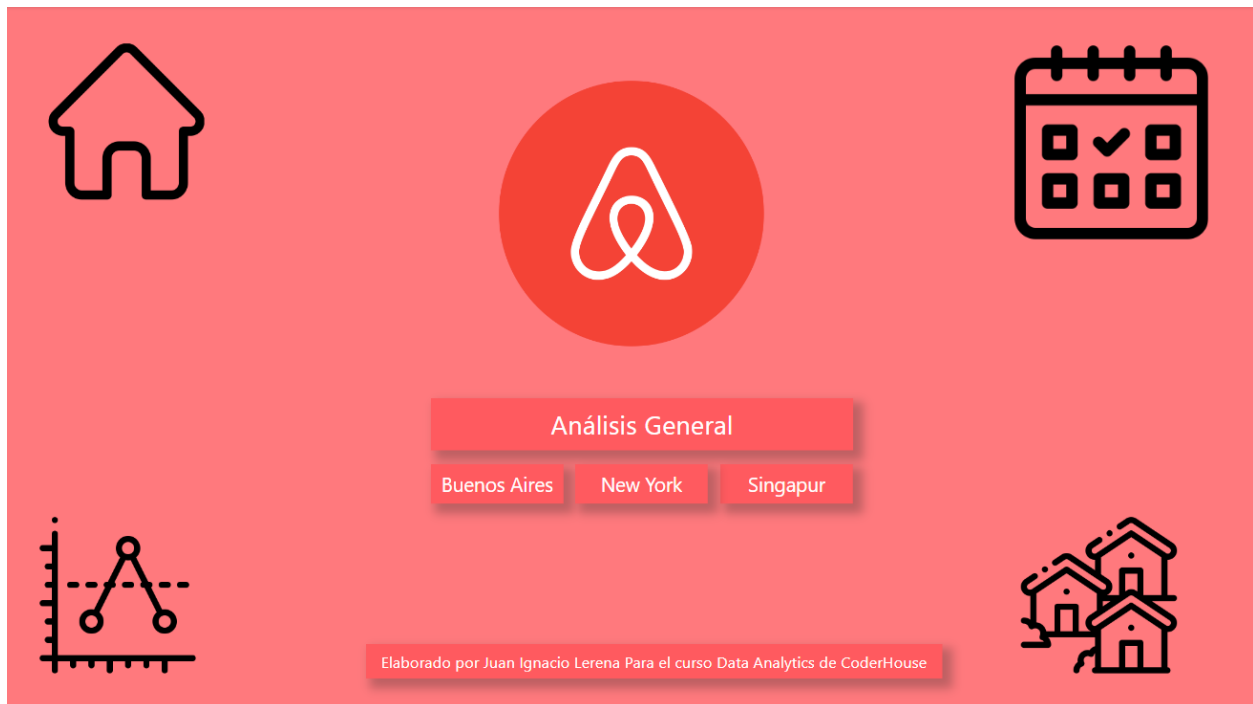
```
Reservas_x_Host = [Total_Reservas]/[Total_Host]
```

- Por último, con las medidas “Total_Barrios”, “Total_Host” y “Total_Reservas” se obtuvieron mediante un conteo, el índice de la totalidad de los Barrios, Hosts y Reservas respectivamente; que luego mediante la segmentación podremos apreciar dichos índices para cada ciudad en análisis.

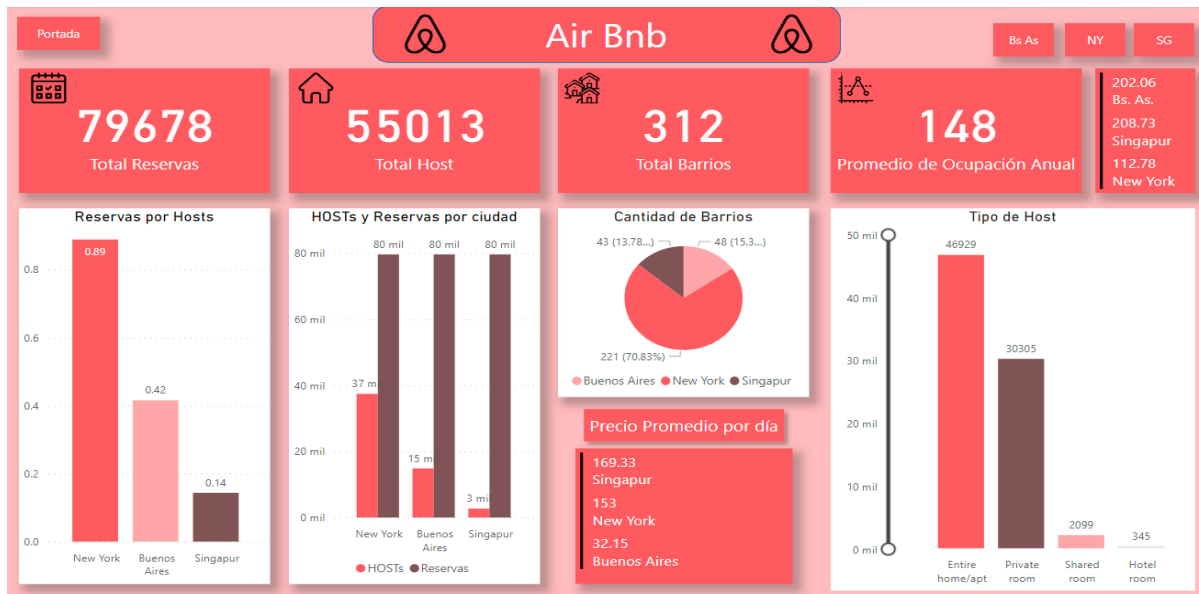
```
Total_Barrios = COUNT(BARRIOS[id_barrio])  
Total_Host = COUNT(HOSTs[id_host])  
Total_Reservas = COUNT(RESERVAS[id_reserva])
```

10. TABLERO

- 10.1 PORTADA: Esta es la solapa que presenta una puerta de entrada al trabajo referenciando botones hacia todas las otras solapas del trabajo. Una vez dentro de cada solapa se podrán encontrar botones dinámicos para poder dirigirse rápidamente hacia cualquiera de las solapas del tablero. Además se incluyen los datos del autor (con link referenciado al perfil de LinkedIn) y entorno de realización.



10.2 ANÁLISIS GENERAL: Esta solapa es la principal del trabajo, contiene información agregada de todas las ciudades en análisis reflejando todos los indicadores para una correcta comparación entre ciudades en análisis.



10.3 TABLERO POR CIUDAD: Luego del análisis general tenemos un tablero para cada ciudad donde podemos encontrar los mismos indicadores de la solapa antes mencionada pero segmentados por la ciudad correspondiente para así tener datos específicos de cada ciudad. Además, se agrega mapa de concentración de los Hosts segmentados en colores por barrios.

