

# ***CODER HOUSE***



---

## **Proyecto Final - Data Analytics**

---

Autor:

**Guido Ivan Simoncini**

Profesor:

**Luciano Julian Gomez Olivera**

Comisión:

**42290**

---

Título del Proyecto:

**Análisis de plazas de Alojamiento via AirBnB en New York 2019**

Iniciativa de análisis:

**Realizar un análisis de la oferta de alojamientos y su variedad para conocer las preferencias y conductas del consumidor**

# 1. Introducción

La ciudad de Nueva York es una de las ciudades más grandes y concurridas del planeta, ya sea con fines turísticos o de negocios la misma posee un tráfico de personas bastante importante y para esta demanda se necesita de una gran cantidad de plazas de alojamiento, con una gran gama de diversidad en los establecimientos que lo ofrecen. Con el avance tecnológico y el apogeo de páginas y aplicaciones para facilitar al huésped la tarea de encontrar alojamiento y al anfitrión la de brindar su espacio se genera la compañía Air BnB, la cual conecta huéspedes con anfitriones sin intermediarios. En un principio la gente no confiaba tanto en suplir la necesidad de alojamiento vía Internet, pero con el correr de los años esta práctica se ha ido magnificando y convirtiéndose en un método confiable para gran parte de la población, quizás en parte atribuido a la llegada de los smartphones y la costumbre de adaptar nuestra vida en base a esta herramienta.

# 2. Descripción de la temática de los datos

Para el presente trabajo se extrajo información de las reviews sobre reservas de alojamiento en la ciudad de Nueva York vía AirBnB, del año 2019.

Se tendrán en cuenta varias fuentes de información pertinente, tales como, la ubicación del alojamiento (contemplando ubicación en específica y el barrio en el cual se encuentra), su categoría, su precio, la cantidad mínima de noches y la cantidad de reseñas que tiene hechas el alojamiento en el sitio web.

# 3. Hipótesis

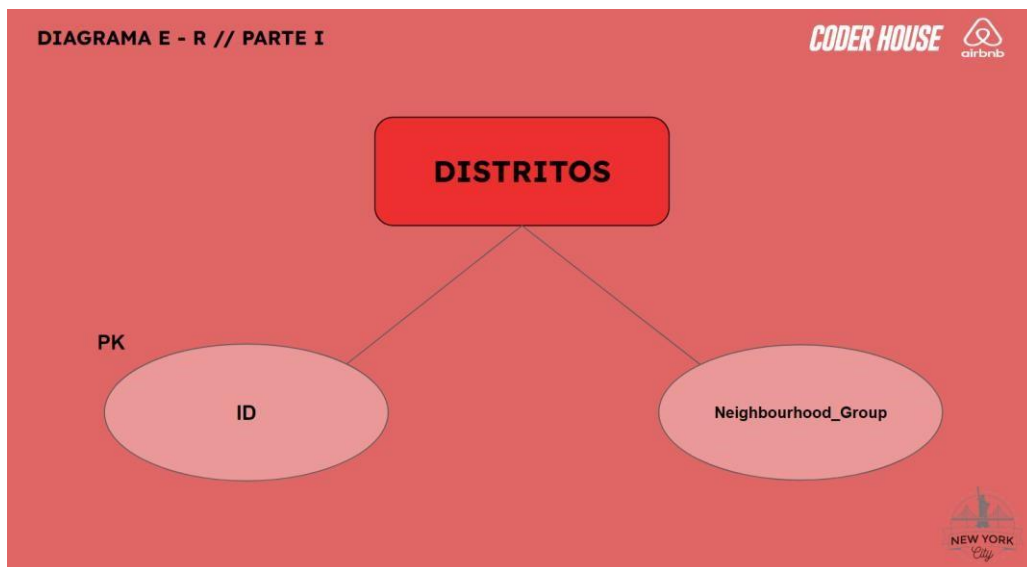
El proyecto se enfocará en conocer las experiencias de los huéspedes y el feedback que estos dan a los alojamientos, además se estudiará el crecimiento del fenómeno de reservas vía web y como se ha ido incrementando con el pasar de los años. Esta información es clave para poder comprender la multiplicidad de aspectos que engloba el fenómeno turístico y los alojamientos. Podemos efectuar varios planteos tales como, cuáles son los barrios más frecuentados, las categorías de cuarto predilectas de los huéspedes. Además puede dar a conocer si hay algún área que sea notablemente más frecuentada que otra y cuál podría llegar a ser la razón de esto. Sí es precio o comodidad lo que se elige o una mezcla de ambas. Podemos ver los datos por distrito, por cantidad de reviews, por valoración, por disponibilidad de camas, podemos identificarlos por su precio y saber cuáles son los más caros y en qué zonas se ubican. Otra alternativa es hacer un análisis tomando de eje a los anfitriones, haciendo un top de los 10 con más camas disponibles, o con mejores reseñas, entre otros tópicos. Podríamos analizar por tipo de habitación en cada distrito, con ellas promediar el precio de tipo de habitación por distrito, o hacer una relación tipo de habitación por precio por valoración. Con estos enfoques podríamos hacernos un panorama del fenómeno turístico en la ciudad de Nueva York.

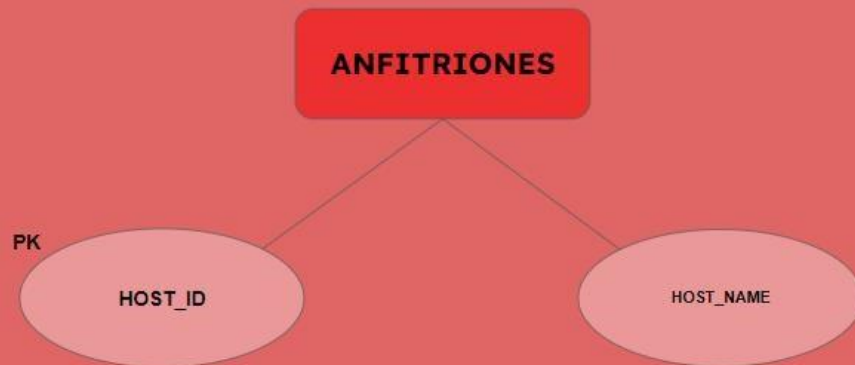
# 4. Herramientas tecnológicas implementadas

Para el presente trabajo se utilizaron los siguientes programas:

- Excel para la lectura de los Datasets.
- Presentaciones de google, para la creación del Diagrama E-R.
- Presentaciones de google, para la generación de los fondos de los tableros.
- Hojas de Cálculo para la generación de la definición de tablas.
- Power BI para la generación de los tableros.
- Logos e imágenes en PNG extraídas de Google y LOGODOWNLOAD y FLATICON.

## 5. Diagrama Entidad – Relación





## Definición de tablas

CODER HOUSE



Tabla I - DISTRITOS

| Campo               | Tipo de Dato | Tipo de Clave |
|---------------------|--------------|---------------|
| ID                  | INT          | PK            |
| Neighbourhood_Group | TEXT         |               |

Tabla II - HABITACIONES

| Campo     | Tipo de Dato | Tipo de Clave |
|-----------|--------------|---------------|
| ID        | INT          | PK            |
| ROOM_TYPE | VARCHAR      |               |

Tabla III - ANFITRIONES

| Campo     | Tipo de Dato | Tipo de Clave |
|-----------|--------------|---------------|
| HOST_ID   | INT          | PK            |
| HOST_NAME | TEXT         |               |

Tabla IV - VALORACIONES

| Campo      | Tipo de Dato | Tipo de Clave |
|------------|--------------|---------------|
| ID         | INT          | PK            |
| VALORACIÓN | TEXT         |               |



Tabla V - ALOJAMIENTOS

| Campo                          | Tipo de Dato | Tipo de Clave |
|--------------------------------|--------------|---------------|
| ID                             | INT          | PK            |
| NAME                           | TEXT         |               |
| HOST_ID                        | INT          | FK            |
| ID_DISTRITO                    | INT          | FK            |
| LATITUD                        | FLOAT        |               |
| LONGITUD                       | FLOAT        |               |
| PRICE                          | MONEY        |               |
| MINIMUM_NIGHTS                 | INT          |               |
| NUMBER_OF_VIEWS                | INT          |               |
| LAST_REVIEWS                   | DATE         |               |
| ID_ROOM                        | INT          | FK            |
| REVIEWS_PER_MONTH              | FLOAT        |               |
| CALCULATED_HOST_LISTINGS_COUNT | INT          |               |
| ID_VALORACION                  | INT          | FK            |

CODER HOUSE



## 6. Generación de fondos para los tableros

Se utiliza PowerPoint para generar los fondos a utilizar en los distintos tableros. A continuación, se muestran los 2 fondos diseñados.

## Portada del informe:



## Fondo para el resto de páginas:



Fondo para ToolTip:



## 7. Aplicación de los fondos creados en Power BI

Una vez generados los fondos en PowerPoint, se aplican en las distintas páginas del archivo de Power BI.

## 8. Generación de Hipervínculos

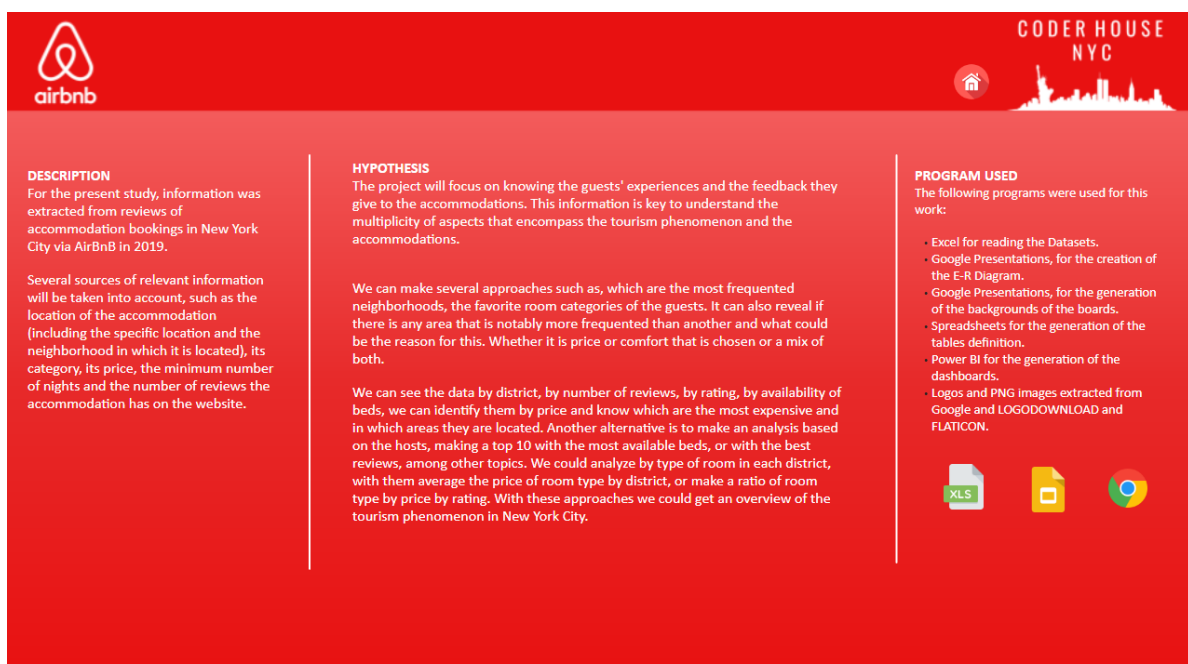
Se insertan 1 imagen en la portada del informe con el logo de LinkedIn y se le genera el hipervínculo correspondiente a un enlace que lleva al perfil de LinkedIn correspondiente en el cual también se facilitaran datos de contacto.

Se insertan 4 imágenes con hipervínculos que permiten pasar a cada una de las páginas del informe.



## 9. Alcance Hipótesis Glosario y herramientas Tecnológicas

En la segunda página del informe, se incluye una descripción del análisis del proyecto, la hipótesis a evaluar, un glosario y un listado de todas las herramientas tecnológicas utilizadas. Además, se incluye una imagen con su correspondiente hipervínculo para poder retornar a la primera página del informe (PORTADA).





## 10. Importación de tablas (Datasets)

1. Se importa la información contenida en el archivo de Excel con el nombre **DATASET**

### **AIRBNB**

Sobre la tabla **Anfitriones** se llevan a cabo las siguientes transformaciones, en el campo **host\_name**:

Se cambian nombres repetidos que están cargados con un aditivo, 'SONDER (NYC)' y 'SONDER', ahora solo figura como 'SONDER'.

Se eliminan espacios en blanco y nombres con caracteres en japonés y chino. Se cambia el tipo de INT64 a TEXT

## 11. Tabla de Fechas

La tabla de fechas se genera en Power BI utilizando la función CALENDARAUTO con la siguiente fórmula:

**Calendario = CALENDARAUTO()**

Una vez generada la tabla CALENDARIO, se la marca como tabla de fechas ejecutando el comando HERRAMIENTAS DE TABLAS > MARCAR COMO TABLA DE FECHAS.

A continuación, se generan distintas columnas para llevar a cabo distintos tipos de gráficos, filtros y cálculos de inteligencia de tiempo.

### **Cálculo de años:**

Año = YEAR(Calendario[Fecha])

### **Cálculo del número del mes:**

Número Mes = MONTH(Calendario[Fecha])

### **Cálculo de meses en letras:**

Mes = UPPER(LEFT(FORMAT(Calendario[Fecha], "MMMM"), 1)) & MID(FORMAT(Calendario[Fecha], "MMMM"), 2, 15)

### **Cálculo de trimestres:**

Trimestre = QUARTER(Calendario[Fecha]) & "º trimestre"

### **Cálculo de estaciones del año:**

Estación =

IF(Calendario[Fecha] >= DATE(YEAR(Calendario[Fecha]), 3, 21) && Calendario[Fecha] <= DATE(YEAR(Calendario[Fecha]), 6, 20),"Otoño",

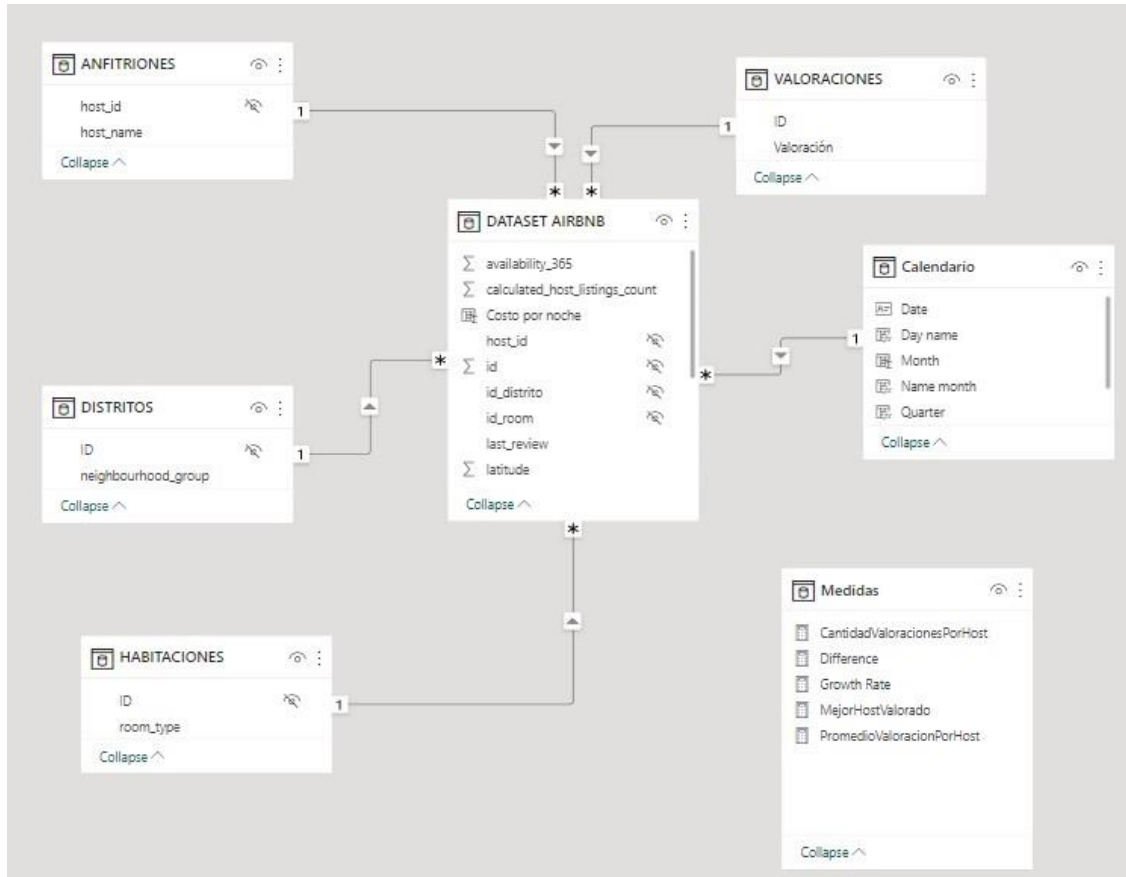
IF(Calendario[Fecha] >= DATE(YEAR(Calendario[Fecha]), 6, 21) && Calendario[Fecha] <= DATE(YEAR(Calendario[Fecha]), 9, 20),"Invierno",

IF(Calendario[Fecha] >= DATE(YEAR(Calendario[Fecha]), 9, 21) && Calendario[Fecha] <= DATE(YEAR(Calendario[Fecha]), 12, 20),"Primavera",

"Verano"))))

## 12. Relación entre Tablas

### Modelo de datos Power BI



## 13. Medidas

Se genera una tabla con el nombre MEDIDAS en la que se almacenarán todas las medidas a incluir en el proyecto.

- Se genera una medida rápida para conocer la cantidad de valoraciones:

**CantidadValoracionesPorHost =**  
**COUNTROWS('DATA SET AIRBNB')**

- Se genera una MEDIDA RÁPIDA que permita calcular la tasa de crecimiento anual según la facturación en AIRBNB Nueva York. La fórmula obtenida a través de esta medida rápida es:

```

Growth Rate =
VAR __PREV_YEAR =
    CALCULATE(
        SUM('DATASET AIRBNB'[price]),
        DATEADD('Calendario'[Date], -1, YEAR)
    )
RETURN

```

```

IF(SELECTEDVALUE(Calendario[Year]) IN VALUES(Calendario[Year]),
    DIVIDE(SUM('DATASET AIRBNB'[price]) - __PREV_YEAR, __PREV_YEAR), BLANK())

```

- Se genera una medida a partir de la medida anterior con el nombre DIFFERENCE que permita calcular, en dinero, la diferencia de facturación entre un año y otro:

```

Difference =
VAR __PREV_YEAR =
    CALCULATE(
        SUM('DATASET AIRBNB'[price]),
        DATEADD('Calendario'[Date], -1, YEAR)
    )
RETURN

```

```

IF(NOT(SELECTEDVALUE(Calendario[Year]) IN VALUES(Calendario[Year])) || [Growth Rate] =
    BLANK(), BLANK(),
    SUM('DATASET AIRBNB'[price]) - __PREV_YEAR)

```

- Se genera una medida para conocer al mejor Host Valorado

```

MejorHostValorado =
VAR MaxPromedioValoracion = MAXX(ALL('ANFITRIONES'[host_name]),
[PromedioValoracionPorHost])
VAR MaxCantidadValoraciones = MAXX(FILTER(ALL('ANFITRIONES'[host_name]),
[PromedioValoracionPorHost] = MaxPromedioValoracion), [CantidadValoracionesPorHost])
RETURN
IF(
    AND([PromedioValoracionPorHost] = MaxPromedioValoracion, [CantidadValoracionesPorHost]
= MaxCantidadValoraciones),
    FIRSTNONBLANK('ANFITRIONES'[host_name], 1),
    BLANK()
)

```

- Se genera una nueva medida de promedio de valoración por host

**PromedioValoracionPorHost =**

**AVERAGEX('DATASET AIRBNB', 'DATASET AIRBNB'[valoración])**

- Se generaron 3 medidas para calcular la cantidad de tipos de habitación individualmente, para ser asignada a medidores:

**Shared Rooms =**

**CALCULATE(COUNT('DATASET AIRBNB'[id\_room]),'DATASET AIRBNB'[id\_room]=3)**

**Private Room =**

**CALCULATE(COUNT('DATASET AIRBNB'[id\_room]),'DATASET AIRBNB'[id\_room]=1)**

**Entire Home/Apt =**

**CALCULATE(COUNT('DATASET AIRBNB'[id\_room]),'DATASET AIRBNB'[id\_room]=2)**

- Se generaron 2 medidas para seleccionar valores y asignarlos a tarjetas:

**Year Selected =**

**IF( SELECTEDVALUE(Calendario[Year]) = BLANK(), "All Years",  
SELECTEDVALUE(Calendario[Year]))**

**Season =**

**IF( SELECTEDVALUE(Calendario[Seasons]) = BLANK(), "All",  
SELECTEDVALUE(Calendario[Seasons]))**

- Se generaron 2 medidas de promedio, para calcular en precio una es por año y la otra es promedio total

**average income = DIVIDE(SUM('DATASET AIRBNB'[price]), COUNTROWS('DATASET AIRBNB'))**

**Average Income per Year = CALCULATE(DIVIDE(SUM('DATASET AIRBNB'[price]),DISTINCTCOUNT(Calendario[Year])))**

- Se genero una medida para calcular el primer cuartil y otra para el tercero

**First quartile = PERCENTILE.INC('DATASET AIRBNB'[price],0.25)**

**First quartile = PERCENTILE.INC('DATASET AIRBNB'[price],0.25)**

## 14. Visualizaciones

En el Informe se generan las siguientes visualizaciones:

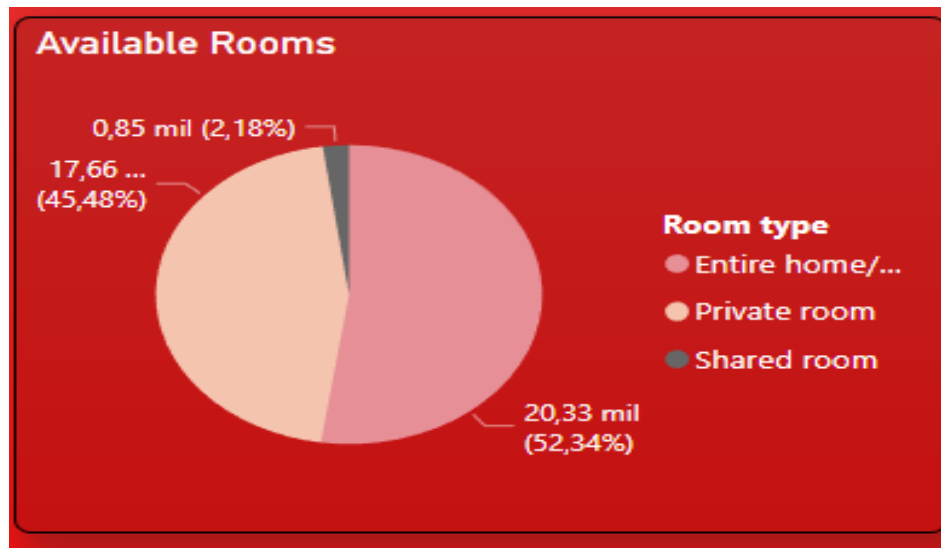
- 1- **Matriz** en la cual podemos ver la cantidad de dinero generado por los anfitriones a lo largo de los años y sus fluctuaciones.

| Year-on-year growth of the platform |                 |             |              |                  |           |                   |            |
|-------------------------------------|-----------------|-------------|--------------|------------------|-----------|-------------------|------------|
| Year                                | Money Generated | Growth rate | Difference   | Average bookings | Best sale | Amount of booking | Percentage |
| 2011                                | \$1.183         |             |              | \$169            | \$299     | 7                 | 0,02%      |
| 2012                                | \$3.953         | 234,15 %    | 2.770,00     | \$158            | \$700     | 25                | 0,07%      |
| 2013                                | \$12.298        | 211,11 %    | 8.345,00     | \$256            | \$5.000   | 48                | 0,22%      |
| 2014                                | \$31.905        | 159,43 %    | 19.607,00    | \$160            | \$800     | 199               | 0,58%      |
| 2015                                | \$218.561       | 585,04 %    | 186.656,00   | \$157            | \$9.999   | 1391              | 3,95%      |
| 2016                                | \$410.435       | 87,79 %     | 191.874,00   | \$152            | \$10.000  | 2704              | 7,43%      |
| 2017                                | \$434.004       | 5,74 %      | 23.569,00    | \$135            | \$10.000  | 3205              | 7,85%      |
| 2018                                | \$840.647       | 93,70 %     | 406.643,00   | \$139            | \$8.500   | 6049              | 15,21%     |
| 2019                                | \$3.574.080     | 325,16 %    | 2.733.433,00 | \$142            | \$7.500   | 25209             | 64,67%     |
| Total                               | \$5.527.066     |             |              | \$142            | \$10.000  | 38837             | 100,00%    |

- 2- **Mapa** que muestra el nivel de facturación por Distrito. La misma tiene también un **Scroller** debajo.



3- **Gráfico Circular** que nos indica la división de los tipos de alojamiento



4- **Tarjetas** una que muestra cantidad de Reviews y otra cantidad de camas disponibles

**1 mill.**  
Reviews Number

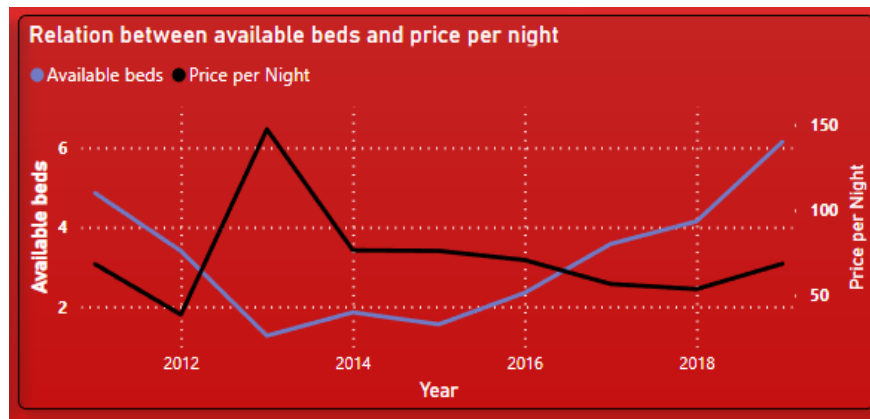
**4 mill.**  
Available Beds

5- **Segmentaciones** que permitirán filtrar el tablero por Temporada y por Distrito en páginas 1 y 2

**Seasons**  
Todas ▼

**District**  
Todas ▼

6- Se genera un **Gráfico de Líneas** que nos indicara la relación entre camas disponibles y el valor de la habitación por noche en el transcurso de los años.



7- Se genera una página con el nombre **TOOLTIP**

En la misma se genera un gráfico de embudo para conocer el top 10 de los Anfitriones con más camas disponibles.

Y dos tarjetas que muestran el número de habitaciones reservadas y el precio promedio por noche.

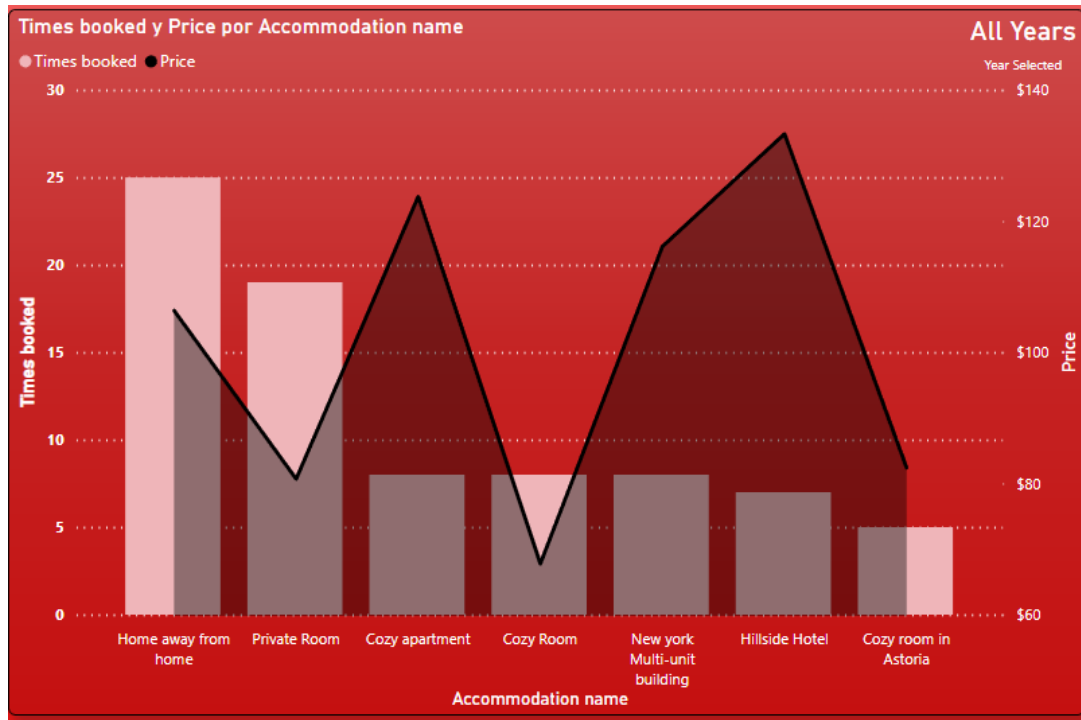
El mismo está asociado a los globos del mapa ubicado en la primer página informativa.



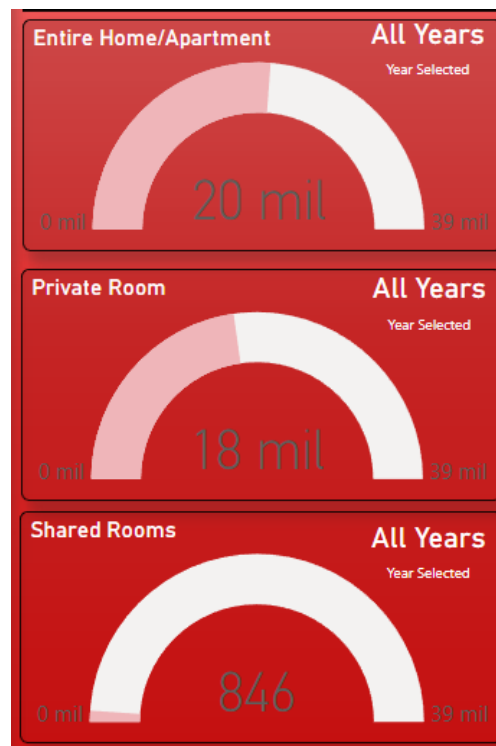
8- Se inserta una Imagen para usar como borrador de todos los filtros



9- Se crea un **Gráfico de columnas agrupadas y de líneas** que nos indicara un top 7 de los alojamientos más reservados y en la línea nos mostrara su precio por noche promedio.



10- Se generan 3 **Medidores** que nos indicaran la cantidad de habitaciones por cada tipo de habitación.

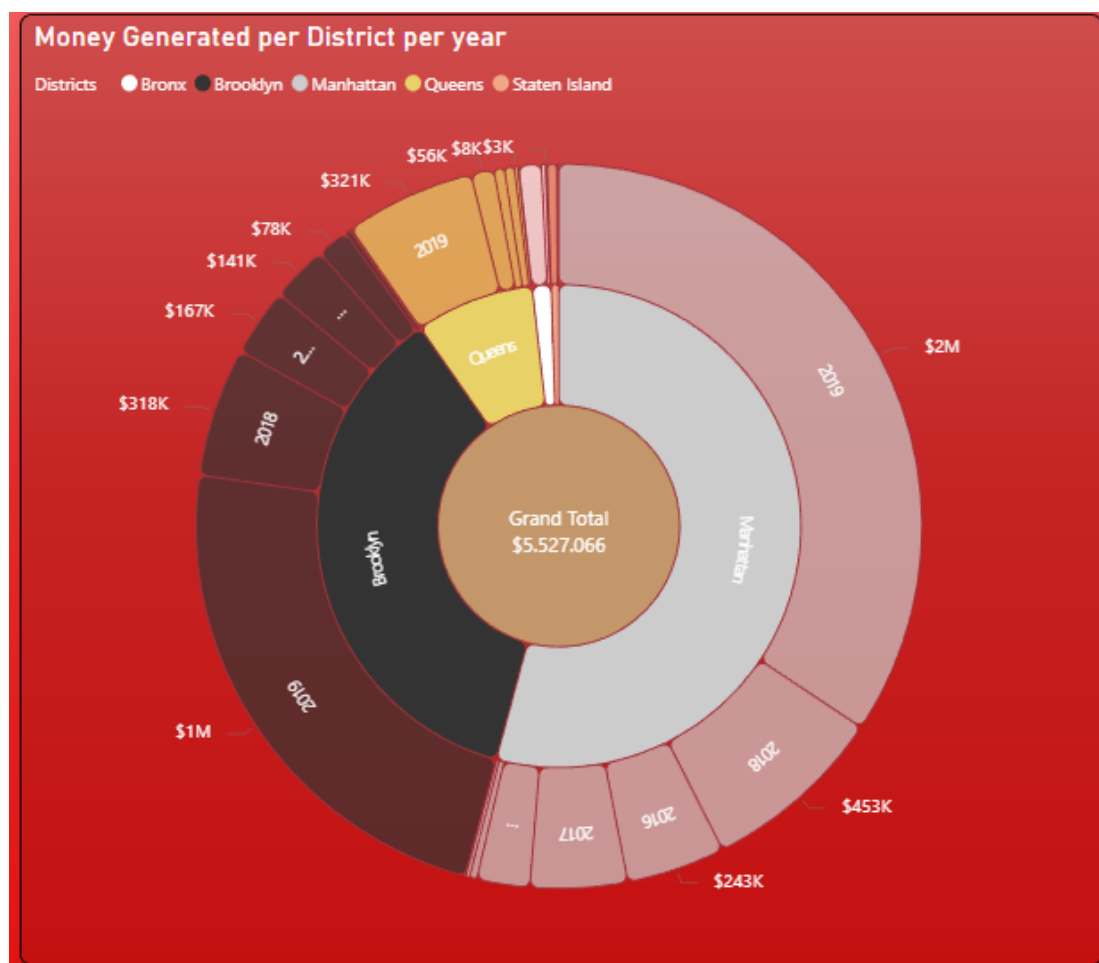




11- Se agregó un **campo de obtención de detalles** en la segunda página del informe, con un drill through, el mismo es por años, también se agregó un **marcador** para borrar el filtrado y el drill through.



12- Se generó un **SunBurst** el cual en el centro tendrá el total de ingresos históricos, en el segundo anillo los distritos y en el tercero los años, para conocer como fluctúan los ingresos en estas variables.



13- Se agregó un **Scroller** y debajo las **nuevas tarjetas** para mostrar diversos indicadores y apoyar la conclusión



## 15. Conclusión

El análisis de los datos de alojamiento de Airbnb en Nueva York proporciona ideas valiosas sobre la evolución de la dinámica del mercado. Una observación significativa es el notable crecimiento en la generación de ingresos, que ascendió a \$5.527.066 millones desde 2011. Es particularmente destacable el marcado aumento en los ingresos en los últimos años, lo que indica un incremento exponencial tanto en los proveedores de la plataforma como en la actividad de reservas. Esta tendencia puede atribuirse a la creciente confianza del consumidor en las reservas en línea, amplificada aún más por la proliferación de teléfonos inteligentes y aplicaciones especializadas.

Además, la prominencia de Manhattan y Brooklyn como los barrios más frecuentados se alinea con su estatus como las regiones de mayor precio. Esta correlación subraya la interacción entre la demanda y el costo, ya que estas áreas no solo atraen un volumen sustancial de reservas, sino que también cobran precios elevados debido a su alta demanda.

Por último, los principios fundamentales de la dinámica del mercado se manifiestan de manera destacada en el conjunto de datos. Durante los períodos de mayor disponibilidad, se observa una clara presión a la baja sobre los precios. Por el contrario, cuando la disponibilidad disminuye, los precios tienden a aumentar. Esta observación reafirma el modelo económico clásico en el que la oferta y la demanda moldean intrincadamente las tendencias de precios.

En resumen, el análisis de los datos de Airbnb ilumina la trayectoria del mercado a lo largo del tiempo. El aumento en los ingresos, la concentración de actividad en distritos específicos y la

alineación de los precios con la demanda denotan los cambios fundamentales en el comportamiento del consumidor y la respuesta del mercado. A medida que la industria continúa evolucionando, estas ideas desempeñarán un papel crucial en la formulación de estrategias para anfitriones, viajeros y partes interesadas por igual.