

Análisis alojamientos airbnb

Nueva York

2019



Curso: Data Analytics

Comisión: 64640

Alumna: Agustina Di Pinto



TABLA DE CONTENIDO

| | |
|--|----|
| 1. INTRODUCCIÓN | 2 |
| 1.1 Descripción de la temática de los datos | 2 |
| 1.2 Objetivo..... | 2 |
| 1.3 Hipótesis | 2 |
| 1.4 Alcance del Proyecto | 2 |
| 1.5 Usuario final y nivel de aplicación del análisis | 2 |
| 2. BASE DE DATOS | 3 |
| 2.1 Listado de tablas | 3 |
| 2.2 Listado de columnas de cada tabla | 4 |
| 2.3 Diagrama Entidad – Relación | 5 |
| 3 DOCUMENTACIÓN DEL PROYECTO | 5 |
| 3.1 Pasos Aplicados para realizar el análisis..... | 5 |
| 3.2 Imagen Modelo de Power BI | 6 |
| 3.3 Transformaciones realizadas | 6 |
| 3.4 Medidas Calculadas | 9 |
| 3.5 Columnas y Tablas calculadas con DAX | 10 |
| 3.6 Dashboard..... | 13 |
| 4. CONCLUSIÓN | 17 |

1. INTRODUCCIÓN

1.1 Descripción de la temática de los datos

Este documento tiene como objetivo definir la temática de interés para el desarrollo del Proyecto Final, basada en un conjunto de datos (dataset) relacionado con los alojamientos ofrecidos en la plataforma digital Airbnb. El análisis se centrará en los alojamientos disponibles en diversas zonas y barrios de la Ciudad de Nueva York durante el año 2019.

1.2 Objetivo

El principal objetivo del proyecto es analizar los precios de los alojamientos ofrecidos en la plataforma Airbnb en Nueva York en el año 2019 en función de diversos factores, tales como el tipo de alojamiento, la ubicación geográfica, la disponibilidad de los alojamientos y las reseñas dejadas por los huéspedes.

1.3 Hipótesis

La hipótesis por comprobar al final del desarrollo del trabajo es la siguiente: Los alojamientos ubicados en barrios turísticos, con mayor cantidad de reseñas y que corresponden a "departamentos enteros" son los más costosos.

1.4 Alcance del Proyecto

Como se define anteriormente en el objetivo del proyecto, el mismo abarcará el análisis detallado de los precios de los alojamientos ofrecidos en la plataforma Airbnb en Nueva York durante el año 2019. Dicho análisis se enfocará en identificar cómo variables como el tipo de alojamiento, la ubicación geográfica, la disponibilidad y las reseñas de los huéspedes influyen en el precio de los alquileres.

Como resultado final del proyecto, se desarrollará un dashboard interactivo que permitirá a los usuarios explorar y visualizar de manera dinámica los datos obtenidos. Este dashboard ofrecerá una interfaz intuitiva para filtrar y analizar los precios de los alojamientos en función de diferentes criterios, proporcionando una herramienta útil para la toma de decisiones.

1.5 Usuario final y nivel de aplicación del análisis

El dashboard, producto del análisis, se aplicará en un nivel táctico ya que los usuarios finales de este serán inversores inmobiliarios y consultores de mercado con el fin de identificar oportunidades de inversión o realizar ajustes en sus estrategias de negocio, como la elección de nuevos barrios para invertir, o bien modificar la estrategia de precios en función de la demanda observada.

2. BASE DE DATOS

2.1 Listado de tablas

En este apartado, se mencionan cada una de las tablas junto a una breve descripción de estas. También se definen las clave primarias y foráneas correspondientes a cada tabla:

Tabla_Airbnb_NYC: exhibe los alojamientos localizados en Nueva York ofrecidos a través de AirBnB. Contiene las siguientes columnas.

- ID_airbnb: Clave primaria (PK).
- Nombre_airbnb
- ID_host: Clave foránea (FK).
- ID_zona: Clave foránea (FK).
- ID_barrio: Clave foránea (FK).
- Latitud
- Longitud
- ID_tipo_airbnb: Clave foránea (FK).
- Precio: precio del alojamiento por noche.
- Estancia_minima: cantidad mínima de noches que se debe alquilar el alojamiento.
- Número_reseñas: cantidad de reseñas que posee el alojamiento en la web.
- Última_reseña: fecha de la última reseña que obtuvo el alojamiento.
- Reseñas_por_mes: cantidad de reseñas por mes
- Recuento_listas_host: número de alojamientos que el mismo anfitrión ha publicado en el sitio web de Airbnb.
- Disponibilidad_anual: disponibilidad del alojamiento con respecto a los 365 del año.

Tabla_Hosts: presenta el ID y el nombre de los anfitriones de los hospedajes.

- ID_host: Clave primaria (PK).
- Nombre_host

Tabla_Zonas: muestra el ID y el nombre de las zonas de la ciudad de Nueva York.

- ID_zona: Clave primaria (PK).
- Nombre_zona

Tabla_Barrios: muestra el ID y el nombre de los barrios que se localizan en las distintas zonas de la ciudad de Nueva York.

- ID_barrio: Clave primaria (PK).
- Nombre_barrio

Tabla_Tipo_Airbnb: muestra el ID y el tipo de alojamiento.

- ID_tipo_airbnb: Clave primaria (PK).
- Tipo_airbnb

2.2 Listado de columnas de cada tabla

A continuación, se definirán las columnas que posee cada tabla junto con su tipo de campo y clave:

| Tabla_airbnb_NYC_2019 | | |
|-----------------------|--------------|---------------|
| Campo | Tipo de Dato | Tipo de Clave |
| ID_airbnb | Int | PK |
| Nombre_airbnb | VarChar(100) | - |
| ID_host | Int | FK |
| ID_zona | Int | FK |
| ID_barrio | Int | FK |
| Latitud | Decimal(9,6) | - |
| Longitud | Decimal(9,6) | - |
| ID_tipo_airbnb | Int | FK |
| Precio | Float | - |
| Estancia_minima | Int | - |
| Número_reseñas | Int | - |
| Última_reseña | Date | - |
| Reseñas_por_mes | Decimal(6,2) | - |
| Recuento_listas_host | Int | - |
| Disponibilidad_anual | Int | - |

| Tabla_hosts | | |
|-------------|--------------|---------------|
| Campo | Tipo de Dato | Tipo de Clave |
| ID_host | Int | PK |
| Nombre_host | Varchar(100) | - |

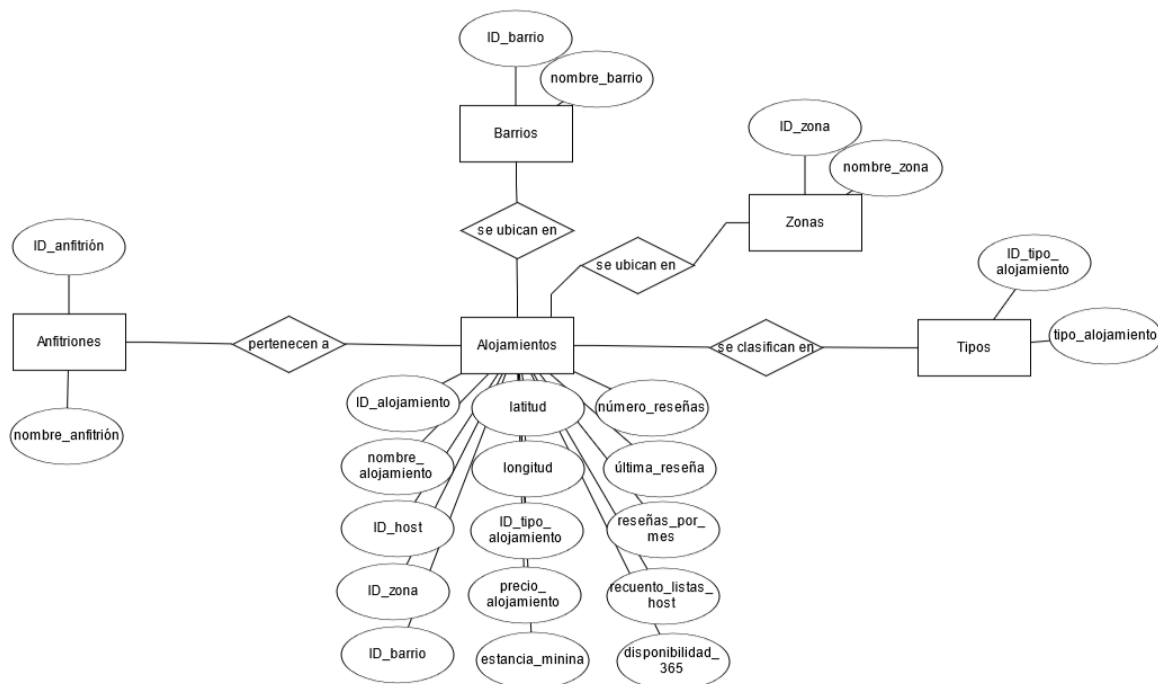
| Tabla_Zonas | | |
|-------------|--------------|---------------|
| Campo | Tipo de Dato | Tipo de Clave |
| ID_zona | Int | PK |
| Nombre_zona | Varchar(100) | - |

| Tabla_Barrios | | |
|---------------|--------------|---------------|
| Campo | Tipo de Dato | Tipo de Clave |
| ID_barrio | Int | PK |
| Nombre_barrio | Varchar(100) | - |

| Tabla_Tipo airbnb | | |
|-------------------|--------------|---------------|
| Campo | Tipo de Dato | Tipo de Clave |
| ID_Tipo_airbnb | Int | PK |
| Tipo_airbnb | Varchar(50) | - |

2.3 Diagrama Entidad – Relación

A continuación, se presenta el diagrama Entidad- Relación de los datos del proyecto:



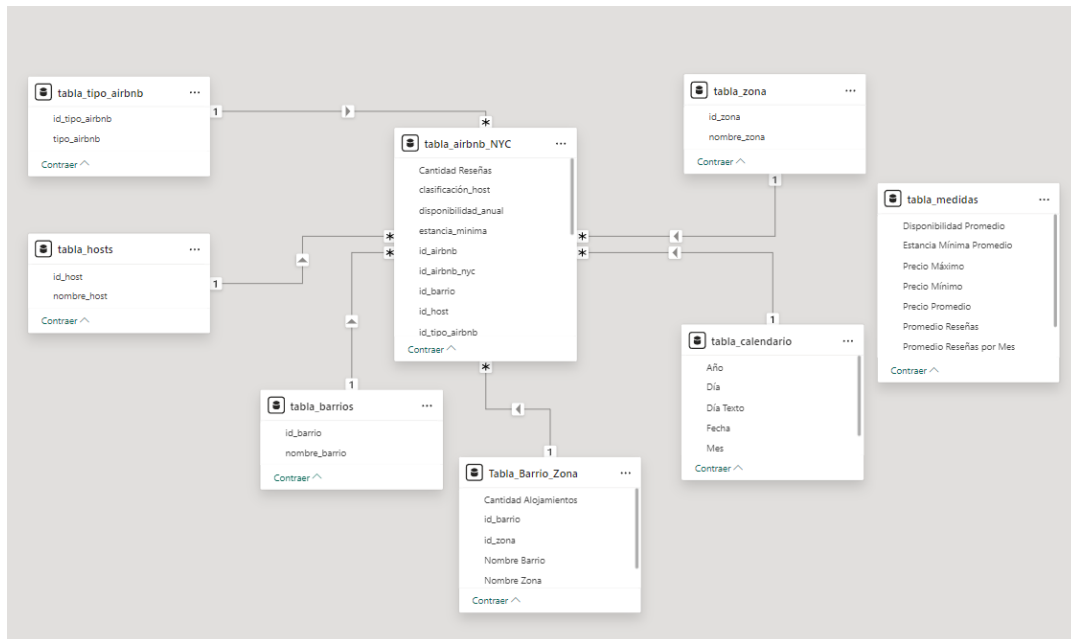
3 DOCUMENTACIÓN DEL PROYECTO

3.1 Pasos Aplicados para realizar el análisis

- 1- Se descargaron los datos registrados en el documento “AIRBNB NYC 2019” disponible en la carpeta Datasets de CoderHouse. El mismo se encuentra en formato “Hoja de cálculo de Google”.
- 2- Se creó la base de datos correspondiente en SQL Server y se importaron los datos. Se verificó que no existían duplicados en el registro.
- 3- Se procedió a trabajar en Power BI Desktop con los datos de la base creada en SQL Server para hacer las transformaciones necesarias y formar la visualización del análisis.

3.2 Imagen Modelo de Power BI

No fue necesario generar correcciones en relaciones de “muchos a muchos”, sí se activó manualmente la relación entre la tabla calendario y la tabla_airbnb_NYC. Y se generó una tabla calculada para llevar a cabo ciertas visualizaciones, se generó también de forma manual la relación entre dicha tabla (Tabla_Zona_Barrio) y la tabla_airbnb_NYC.



3.3 Transformaciones realizadas

Tabla Airbnb NYC

La única transformación fue del tipo de dato del precio. Se cambió a un número decimal fijo.



Tabla Barrios

No sufrió ninguna transformación luego de su importación.

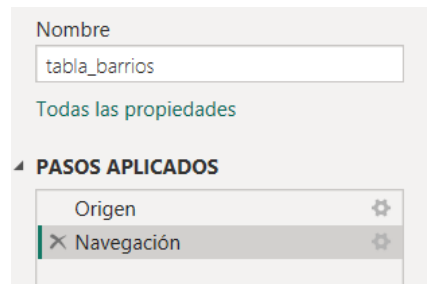


Tabla hosts

No sufrió ninguna transformación luego de su importación.



Tabla tipo airbnb

Se reemplazaron los valores de la columna tipo_airbnb por su traducción al español.

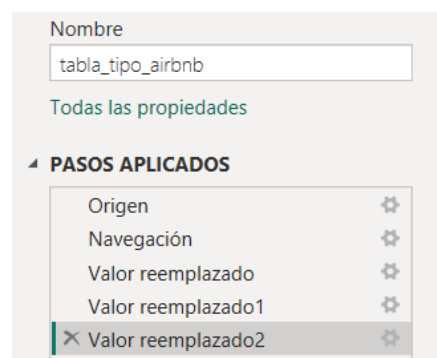


Tabla zona

No sufrió ninguna transformación luego de su importación.



Relación entre Tabla Calendario y Tabla Airbnb NYC

Se generó manualmente la relación entre las dos tablas a partir de las columnas última_reseña y fecha. La relación es de muchos a uno y su dirección es única.

Desde la tabla
tabla_airbnb_NYC

| | longitud | nombre_airbnb | numero_rese... | recuento_lista... | reseñas_por_... | ultima_reseña |
|---|-----------|-------------------|----------------|-------------------|-----------------|---------------|
| 2 | -73.98473 | Lovely Large ... | 0 | 1 | 0 | null |
| 8 | -73.97137 | Best Central P... | 0 | 1 | 0 | null |
| 1 | -73.9694 | Uws Manhatt... | 0 | 1 | 0 | null |

A la tabla
tabla_calendario

| Año | Día | Día Texto | Fecha | Mes | Mes texto |
|------|-----|-----------|------------|-----|-----------|
| 2011 | 28 | lunes | 28/03/2011 | 3 | marzo |
| 2011 | 29 | martes | 29/03/2011 | 3 | marzo |
| 2011 | 30 | miércoles | 30/03/2011 | 3 | marzo |

Cardinality
Varios a uno (*:1)

Dirección de filtro cruzado
Único

☒ Activar esta relación

☐ Aplicar filtro de seguridad en ambas direcciones

Relación entre Tabla Zona Barrio y Tabla Airbnb NYC

Se generó manualmente la relación entre las dos tablas a partir de las columnas id_barrio e id_barrio fecha. La relación es de muchos a uno y su dirección es única.

Desde la tabla
tabla_airbnb_NYC

| Cantidad Res... | clasificación_... | disponibilida... | estancia_mini... | id_airbnb | id_airbnb_nyc | id_barrio |
|-----------------|-------------------|------------------|------------------|-----------|---------------|-----------|
| 0-50 | 1-10 alojamie... | 0 | 1 | 2243769 | 27417 | 203 |
| 0-50 | 1-10 alojamie... | 0 | 1 | 7798930 | 29186 | 203 |
| 0-50 | 1-10 alojamie... | 0 | 1 | 32669305 | 29334 | 203 |

A la tabla
Tabla_Barrio_Zona

| Cantidad Aloj... | id_barrio | id_zona | Nombre Barrio | Nombre Zona |
|------------------|-----------|---------|-----------------|-------------|
| 1971 | 203 | 2 | Upper West S... | Manhattan |
| 252 | 104 | 2 | Inwood | Manhattan |
| 899 | 207 | 2 | Washington ... | Manhattan |

Cardinality
Varios a uno (*:1)

Dirección de filtro cruzado
Único

☒ Activar esta relación

☐ Aplicar filtro de seguridad en ambas direcciones

☐ Asumir integridad referencial



3.4 Medidas Calculadas

Se calcularon las siguientes medidas con el objetivo de generar las visualizaciones necesarias para llevar a cabo el análisis.

Total Alojamientos: alojamientos totales registrados en el dataset.

Total Alojamientos =

COUNTROWS(tabla_airbnb_NYC)

Precio máximo: valor máximo del precio de los alojamientos por noche.

Precio Máximo =

```
CALCULATE(  
    MAX(tabla_airbnb_NYC[precio]),  
    NOT(ISBLANK(tabla_airbnb_NYC[precio])) && tabla_airbnb_NYC[precio] > 0  
)
```

Precio mínimo: valor mínimo del precio de los alojamientos por noche.

Precio Mínimo =

```
CALCULATE(  
    MIN(tabla_airbnb_NYC[precio]),  
    NOT(ISBLANK(tabla_airbnb_NYC[precio])) && tabla_airbnb_NYC[precio] > 0  
)
```

Precio promedio: promedio del precio de los alojamientos por noche.

Precio Promedio =

```
CALCULATE(  
    AVERAGE(tabla_airbnb_NYC[precio]),  
    NOT(ISBLANK(tabla_airbnb_NYC[precio])) && tabla_airbnb_NYC[precio] > 0  
)
```

Promedio reseñas: promedio de las reseñas totales recibidas por los alojamientos.

Promedio Reseñas = AVERAGE(tabla_airbnb_NYC[numero_reseñas])

Promedio reseñas por mes: promedio de las reseñas por mes recibidas por los alojamientos.

```
Promedio Reseñas por Mes = CALCULATE(AVERAGE(tabla_airbnb_NYC[reseñas_por_mes]),  
    NOT(ISBLANK(tabla_airbnb_NYC[reseñas_por_mes])) &&  
    tabla_airbnb_NYC[reseñas_por_mes] > 0  
)
```

Disponibilidad Promedio: Promedio de la disponibilidad anual de los alojamientos.

Disponibilidad Promedio =

```
CALCULATE(  
    AVERAGE(tabla_airbnb_NYC[disponibilidad_anual]),  
    NOT(ISBLANK(tabla_airbnb_NYC[disponibilidad_anual])) &&  
    tabla_airbnb_NYC[disponibilidad_anual] > 0  
)
```

Estancia mínima promedio: promedio de la estancia mínima que se ofrecen los alojamientos.

```
Estancia Mínima Promedio = AVERAGEX(  
    FILTER(tabla_airbnb_NYC, NOT(ISBLANK(tabla_airbnb_NYC[estancia_minima]))),  
    tabla_airbnb_NYC[estancia_minima]  
)
```

3.5 Columnas y Tablas calculadas con DAX

Columnas en tabla_airbnb_NYC

- Rango_Disponibilidad: agrupa alojamientos en base a su disponibilidad en distintos rangos. Se crea también una columna llamada orden_rango_disponibilidad para lograr un mejor resultado en las visualizaciones.

Rango_Disponibilidad =

```
SWITCH(  
    TRUE(),  
    'tabla_airbnb_NYC'[disponibilidad_anual] <= 30, "0-30 días",  
    'tabla_airbnb_NYC'[disponibilidad_anual] <= 90, "31-90 días",  
    'tabla_airbnb_NYC'[disponibilidad_anual] <= 180, "91-180 días",  
    'tabla_airbnb_NYC'[disponibilidad_anual] <= 270, "181-270 días",  
    "271-365 días"  
)
```

Orden_Rango_Disponibilidad =

```
SWITCH(  
    TRUE(),  
    'tabla_airbnb_NYC'[disponibilidad_anual] <= 30, 1,
```

```
'tabla_airbnb_NYC'[disponibilidad_anual] <= 90, 2,  
'tabla_airbnb_NYC'[disponibilidad_anual] <= 180, 3,  
'tabla_airbnb_NYC'[disponibilidad_anual] <= 270, 4,  
'tabla_airbnb_NYC'[disponibilidad_anual] <= 365, 5  
)
```

- Cantidad de reseñas: agrupa los alojamientos en distintos rangos según la cantidad de reseñas que poseen los mismos. Se crea también una columna llamada orden_rango_reseñas para lograr un mejor resultado en las visualizaciones.

Cantidad Reseñas =

```
SWITCH(  
  TRUE(),  
  'tabla_airbnb_NYC'[numero_reseñas] <= 50, "0-50",  
  'tabla_airbnb_NYC'[numero_reseñas] <= 100, "51-100",  
  'tabla_airbnb_NYC'[numero_reseñas] <= 150, "101-150",  
  'tabla_airbnb_NYC'[numero_reseñas] <= 200, "151-200",  
  'tabla_airbnb_NYC'[numero_reseñas] <= 250, "201-250",  
  'tabla_airbnb_NYC'[numero_reseñas] <= 300, "251-300",  
  'tabla_airbnb_NYC'[numero_reseñas] <= 350, "301-350",  
  'tabla_airbnb_NYC'[numero_reseñas] <= 400, "351-400",  
  'tabla_airbnb_NYC'[numero_reseñas] <= 450, "401-450",  
  'tabla_airbnb_NYC'[numero_reseñas] <= 500, "451-500",  
  "501-629"  
)
```

Orden_Rango_Reseñas =

```
SWITCH(  
  TRUE(),  
  'tabla_airbnb_NYC'[numero_reseñas] <= 50, 1,  
  'tabla_airbnb_NYC'[numero_reseñas] <= 100, 2,  
  'tabla_airbnb_NYC'[numero_reseñas] <= 150, 3,  
  'tabla_airbnb_NYC'[numero_reseñas] <= 200, 4,  
  'tabla_airbnb_NYC'[numero_reseñas] <= 250, 5,
```



```
'tabla_airbnb_NYC'[numero_reseñas] <= 300, 6,  
'tabla_airbnb_NYC'[numero_reseñas] <= 350, 7,  
'tabla_airbnb_NYC'[numero_reseñas] <= 400, 8,  
'tabla_airbnb_NYC'[numero_reseñas] <= 450, 9,  
'tabla_airbnb_NYC'[numero_reseñas] <= 500, 10,  
'tabla_airbnb_NYC'[numero_reseñas] <= 629, 11  
)
```

- Clasificación host: clasifica a los alojamientos según el tipo de host, en base a cuantos alojamientos ofrece un mismo host.

clasificación_host =

```
SWITCH(  
  TRUE(),  
  'tabla_airbnb_NYC'[recuento_listas_host] <= 10, "1-10 alojamientos",  
  'tabla_airbnb_NYC'[recuento_listas_host] <= 30, "11-30 alojamientos",  
  'tabla_airbnb_NYC'[recuento_listas_host] <= 50, "31-50 alojamientos",  
  'tabla_airbnb_NYC'[recuento_listas_host] <= 70, "51-70 alojamientos",  
  'tabla_airbnb_NYC'[recuento_listas_host] <= 100, "71-100 alojamientos",  
  "más de 100 alojamientos"  
)
```

Creación de Tabla_Zona_Barrio y Tabla Calendario

- Tabla_Zona_Barrio: creada para contener los datos barrio, nombre barrio, zona, nombre zona, cantidad alojamientos por barrio, precio promedio por barrio.

Tabla_Barrio_Zona =

```
SUMMARIZE(  
  tabla_airbnb_NYC,  
  tabla_airbnb_NYC[id_barrio],  
  tabla_airbnb_NYC[id_zona],  
  "Nombre Barrio", MAXX(RELATEDTABLE(tabla_barrios),  
  tabla_barrios[nombre_barrio]),  
  "Nombre Zona", MAXX(RELATEDTABLE(tabla_zona), tabla_zona[nombre_zona]),  
  "Cantidad Alojamientos", COUNT(tabla_airbnb_NYC[id_airbnb]),  
  "Precio Promedio", CALCULATE(  

```

```
AVERAGE(tabla_airbnb_NYC[precio]),  
NOT(ISBLANK(tabla_airbnb_NYC[precio])) && tabla_airbnb_NYC[precio] > 0  
)  
)
```

- Tabla_Calendario: creada para establecer una línea temporal entre las fechas contenidas en el dataset.

```
tabla_calendario = CALENDAR(FIRSTDATE(tabla_airbnb_NYC[ultima_reseña]),  
LASTDATE(tabla_airbnb_NYC[ultima_reseña]))
```

Se agregan las columnas año, mes día, mes texto y día texto.

3.6 Dashboard

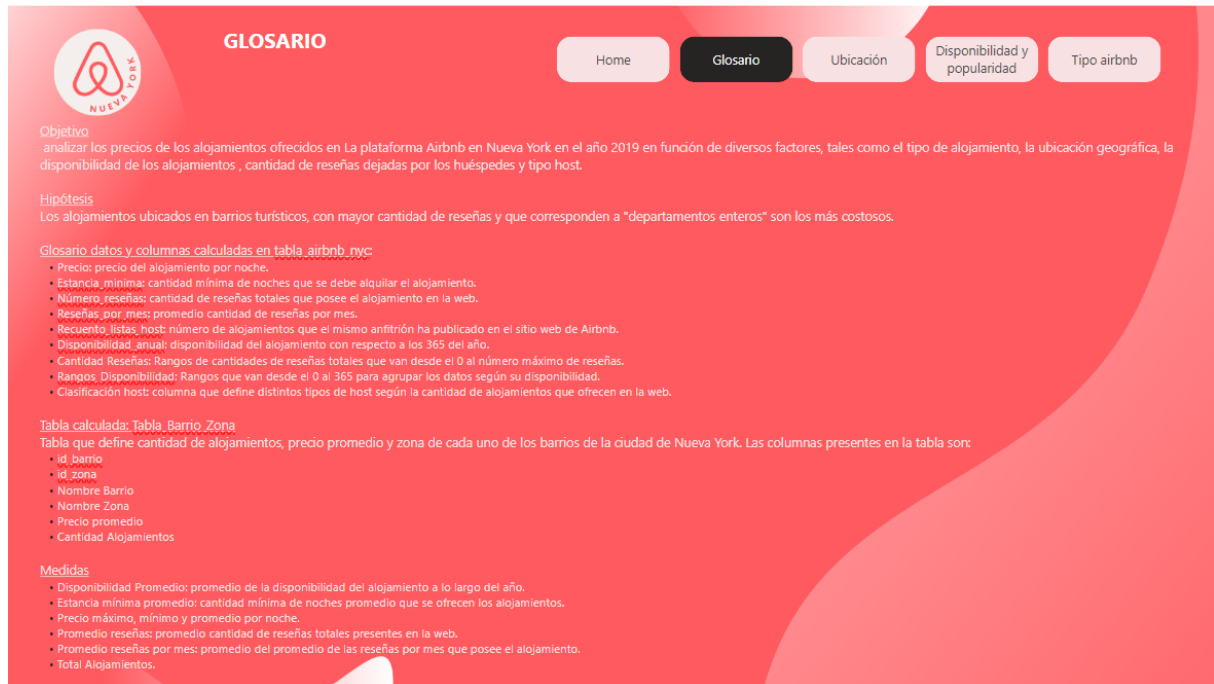
Solapa 1: Home

Portada informativa con título del proyecto, botones de navegación entre solapas hora y fecha de la última actualización.



Solapa 2: Glosario

Glosario general del análisis que contiene todos los conceptos necesarios para interpretar los resultados.



GLOSARIO

Objetivo
analizar los precios de los alojamientos ofrecidos en La plataforma Airbnb en Nueva York en el año 2019 en función de diversos factores, tales como el tipo de alojamiento, la ubicación geográfica, la disponibilidad de los alojamientos, cantidad de reseñas dejadas por los huéspedes y tipo host.

Hipótesis
Los alojamientos ubicados en barrios turísticos, con mayor cantidad de reseñas y que corresponden a "departamentos enteros" son los más costosos.

Glosario datos y columnas calculadas en tabla airbnb_nyc:

- Precio: precio del alojamiento por noche.
- Estancia mínima: cantidad mínima de noches que se debe alquilar el alojamiento.
- Número reseñas: cantidad de reseñas totales que posee el alojamiento en la web.
- Reseñas por mes: promedio cantidad de reseñas por mes.
- Recuento listas host: número de alojamientos que el mismo anfitrión ha publicado en el sitio web de Airbnb.
- Disponibilidad anual: disponibilidad del alojamiento con respecto a los 365 del año.
- Cantidad Reseñas: Rangos de cantidades de reseñas totales que van desde el 0 al número máximo de reseñas.
- Rangos Disponibilidad: Rangos que van desde el 0 al 365 para agrupar los datos según su disponibilidad.
- Clasificación host: columna que define distintos tipos de host según la cantidad de alojamientos que ofrecen en la web.

Tabla calculada: Tabla Barrio Zona
Tabla que define cantidad de alojamientos, precio promedio y zona de cada uno de los barrios de la ciudad de Nueva York. Las columnas presentes en la tabla son:

- id_barrio
- id_zona
- Nombre Barrio
- Nombre Zona
- Precio promedio
- Cantidad Alojamientos

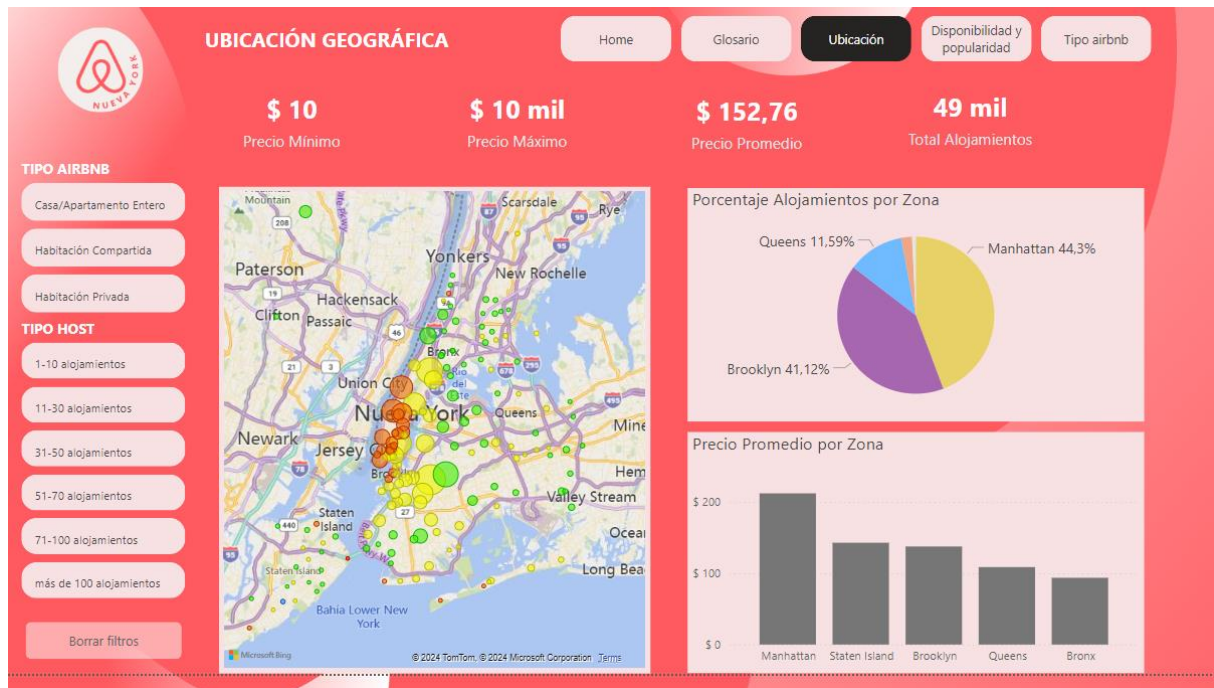
Medidas

- Disponibilidad Promedio: promedio de la disponibilidad del alojamiento a lo largo del año.
- Estancia mínima promedio: cantidad mínima de noches promedio que se ofrecen los alojamientos.
- Precio máximo, mínimo y promedio por noche.
- Promedio reseñas: promedio cantidad de reseñas totales presentes en la web.
- Promedio reseñas por mes: promedio del promedio de las reseñas por mes que posee el alojamiento.
- Total Alojamientos.

Solapa 3: Ubicación

En esta solapa se visualiza la distribución del precio según la ubicación del alojamiento mediante:

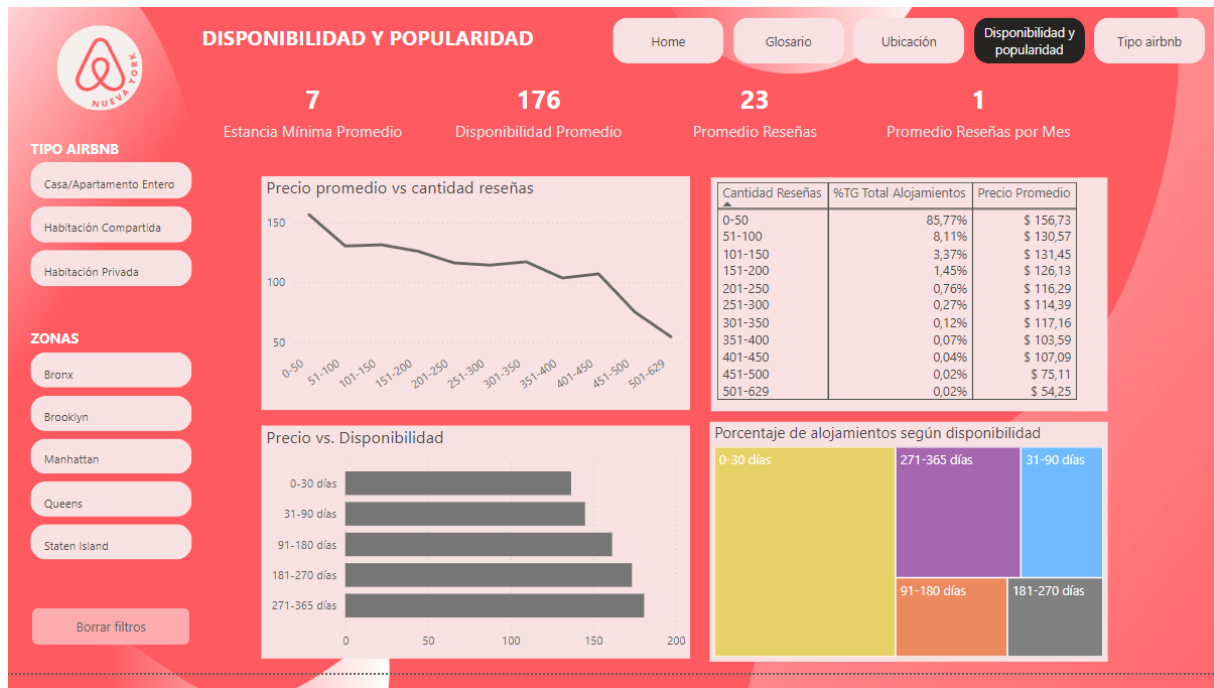
- Mapa que muestra densidad de alojamientos por zona y el precio promedio, siendo rojo el precio promedio más alto, amarillo precio promedio moderado, verde precio promedio más bajo.
- Gráfico de torta que refleja el porcentaje de alojamientos presente en cada zona.
- Gráfico de barras que demuestra el precio promedio en cada zona.
- Filtro según tipo Airbnb y Filtro según tipo host.
- Tarjetas: precio mínimo, precio máximo y precio promedio.



Solapa 4: Disponibilidad y popularidad

En esta solapa se visualiza la distribución del precio según la disponibilidad del alojamiento durante el año, y su popularidad en base a la cantidad de reseñas:

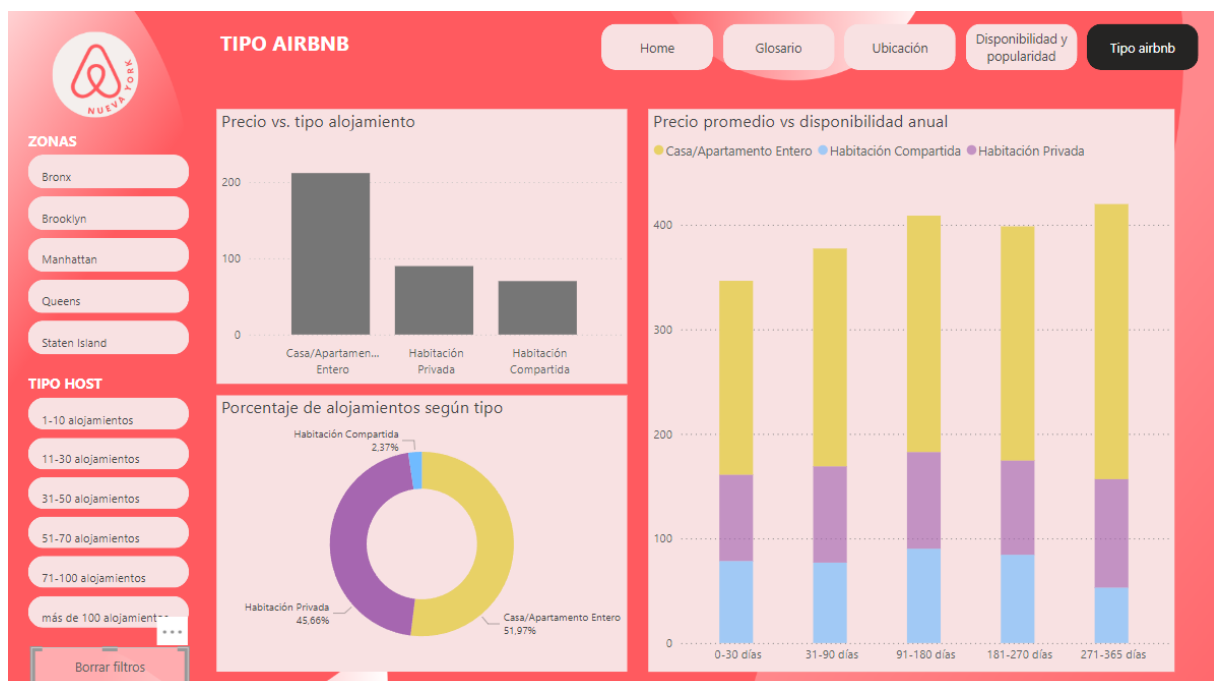
- Gráfico que muestra el precio promedio por cantidad de reseñas.
- Tabla que muestra cantidad de reseñas, porcentaje total de alojamientos según cantidad de reseñas y precio promedio.
- Gráfico de barras agrupadas que muestra precio vs. Disponibilidad del alojamiento.
- Treemap que refleja porcentaje de alojamientos según su disponibilidad.
- Filtro según tipo Airbnb y Filtro según zona.
- Tarjetas: estancia mínima promedio, disponibilidad promedio, promedio reseñas y promedio de reseñas por mes.



Solapa 5: Tipo Airbnb

En esta solapa se visualiza la distribución del precio según el tipo de alojamiento:

- Gráfico de barras que muestra el precio promedio según tipo alojamiento.
- Anillo que muestra porcentaje de alojamientos según el tipo.
- Gráfico de barras que muestra el precio promedio según la disponibilidad anual estratificado según el tipo de alojamiento.





4. CONCLUSIÓN

El análisis realizado permitió confirmar parcialmente la hipótesis planteada, ya que se demuestra que los alojamientos que se ubican en zonas más turísticas, como Manhattan, que son del tipo alojamiento entero son más caros que los que se ubican en zonas menos turísticas y son del tipo habitación compartida. También se demuestra que los alojamientos presentes en zonas turísticas tienen mayor disponibilidad. En contraposición a lo que se planteaba en la hipótesis, los alojamientos que presentaban mayor cantidad de reseñas tienen un precio promedio más bajo.

A partir de estos hallazgos, se identifican nuevas oportunidades para ampliar el análisis, explorando otros factores como la distribución de alojamientos por host.