



Tecnológico de Monterrey

Plataformas de analítica de negocios para organizaciones

Reporte ejecutivo de reto:

Actividad 7 (Reto Final)

Integrantes:

Jasiel Guillermo García Añorve - A01424128

Bruno Chávez Meza - A01737301

Braulio Francisco Islas Salas - A01734425

Andre Alvarez Vizcaino -A01734094

Profesor: Alfredo García Suárez

(Gpo 101)

Resumen

Dentro del siguiente reporte se presentan los principales hallazgos y visualizaciones de datos que representan aspectos esenciales de la operación de Airbnb en las ciudades de Amsterdam y Chicago. Este análisis se centra en la extracción, preprocesamiento, y modelado de datos para comprender mejor las variables que influyen en el rendimiento de las propiedades. Se han generado propuestas, identificado áreas de oportunidad y señalado puntos de interés clave a través de la relación y correlación de las variables seleccionadas.

Objetivos

- Extraer y limpiar los datos relevantes para el análisis de Airbnb en Amsterdam y Chicago.
- Realizar un análisis descriptivo para identificar patrones y tendencias en las variables de interés.
- Desarrollar y evaluar modelos de regresión lineal para entender las relaciones entre las variables y mejorar las predicciones.

Metodología

Extracción de Datos: Se aplicaron filtros específicos para seleccionar las variables de interés en las bases de datos de Amsterdam y Chicago.

Preprocesamiento: Se trataron valores nulos y outliers para asegurar la calidad de los datos.

Extracción de Características: Se analizaron columnas categóricas como “host_response_time”, “host_acceptance_rate”, “property_type”, “room_type”, “amenities” y “host_is_superhost”.

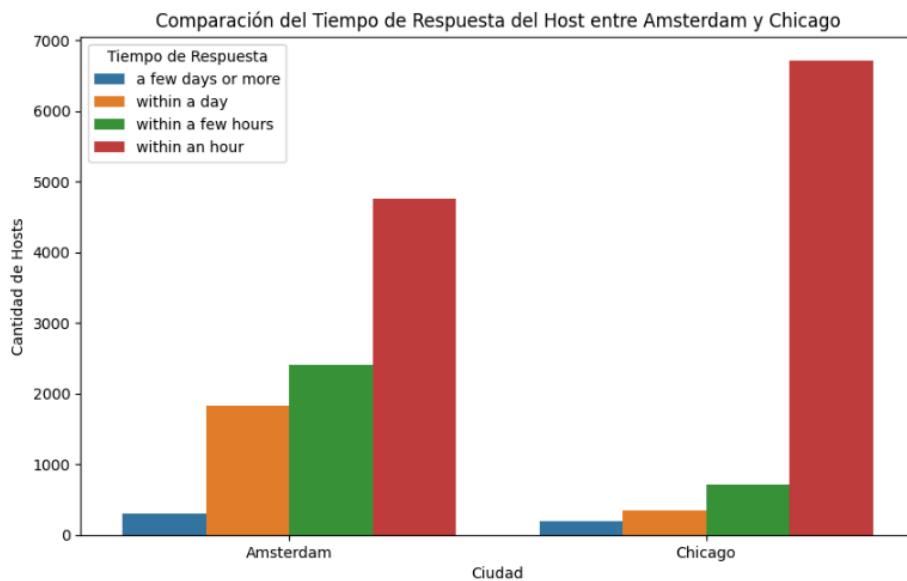
Análisis Descriptivo: Se utilizaron tablas y gráficos para visualizar y resumir los datos.

Modelado: Se construyeron mapas de calor para analizar correlaciones entre variables numéricas y se desarrollaron modelos de regresión lineal simple y múltiple.

Hallazgos y Análisis

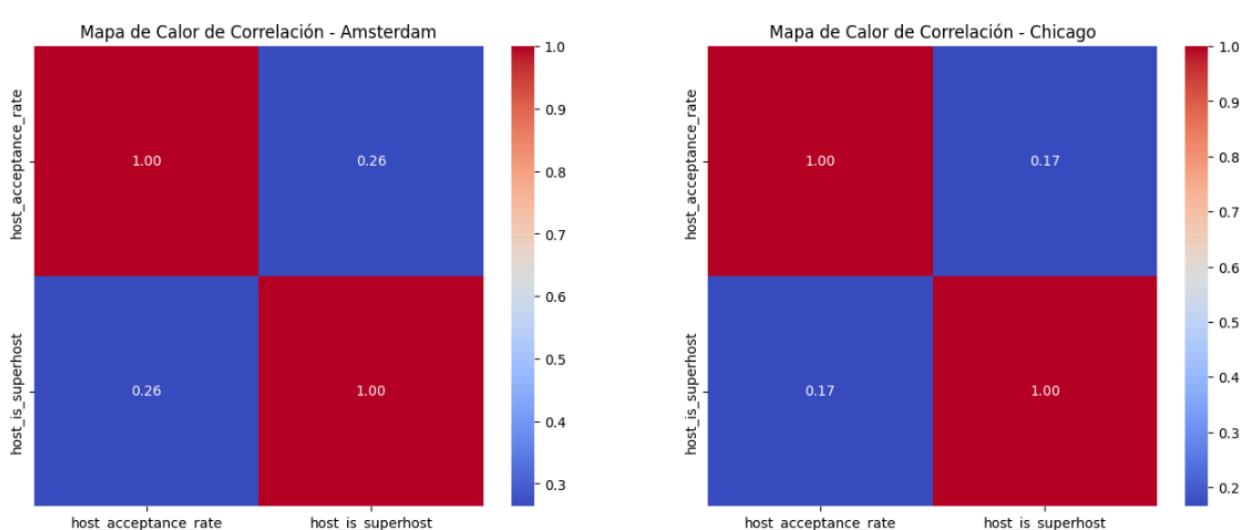
Durante la exploración de los datos y de la manipulación de las variables anteriormente mencionadas se pudieron encontrar caracteres especiales como “\$”, “%” y “/” lo cual en primer instancia perjudicaba la manipulación de la base de datos dentro del lenguaje de programación, para esto desde la copia de la base de datos se realizó el cambio de dichas columnas a tipo “Numérico” para que pudieran ser manipuladas correctamente, posteriormente se realizó una limpieza de Outliers y datos Nulos para obtener resultados más concretos y certeros.

Como primer hallazgo se obtuvo la siguiente gráfica de barras:



Dentro de la gráfica se encuentran la variable de el tiempo de respuesta del host dentro de la aplicación de AIRBNB en donde se puede observar la frecuencia con la que mayormente responden los host a sus clientes a través de la aplicación siendo dominante la respuesta “within an hour” que quiere decir que mayormente responden dentro de solo una hora, como se puede observar en chicago tiene mayor predominancia esta respuesta a comparación que en amsterdam, por otro lado en amsterdam se puede ver que si llegan a tardar más hasta días a comparación de chicago lo cual puede afectar directamente en la satisfacción y comodidad del cliente al igual que puede buscar algún sustituto como hotel, hostal entre otros que pueden robar esta “venta”.

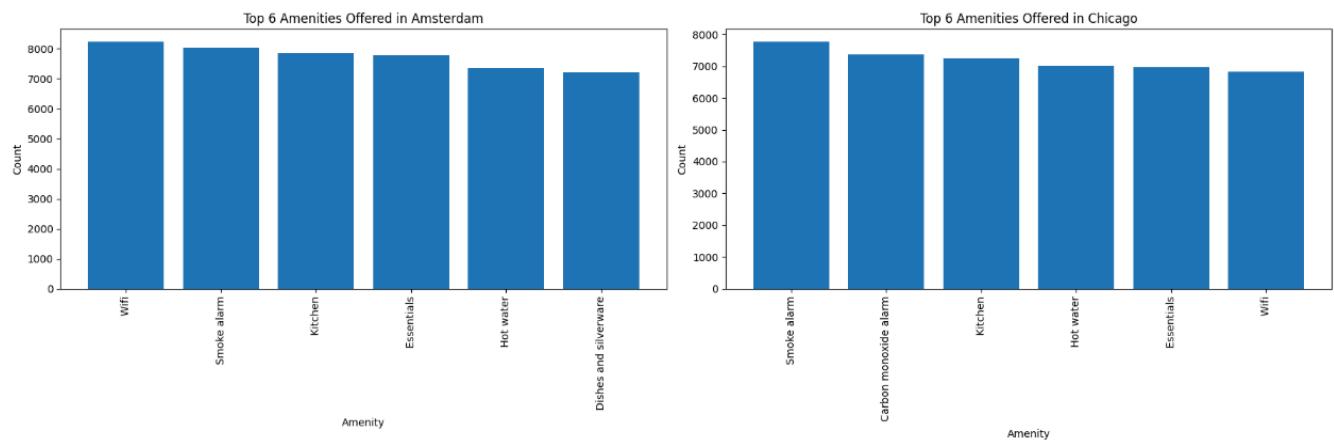
A continuación se presenta un mapa de calor entre dos variables de ambas ciudades:



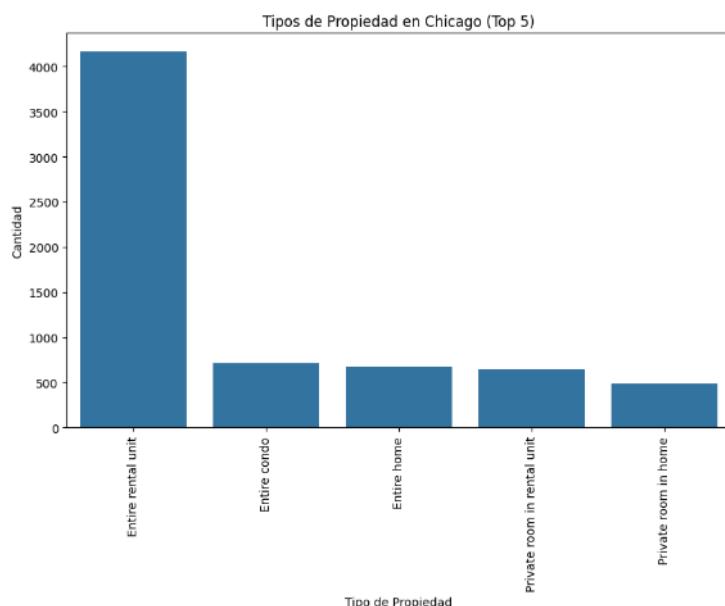
En este caso se puede observar que la cantidad de veces que un host acepta solicitudes de hospedaje y un superhost presentan una correlación de 0.26 en amsterdam y 0.17 en chicago lo que representa una significancia pequeña pero significante viéndolo desde un punto de vista de satisfacción al cliente, en

este caso esta relación afecta más en la ciudad de amsterdam a comparación de chicago lo que nos indica que entre más acepten o rechacen los usuarios pueden considerarlos unos muy buenos host o dejar buenas o malas reseñas sobre la atención que reciben.

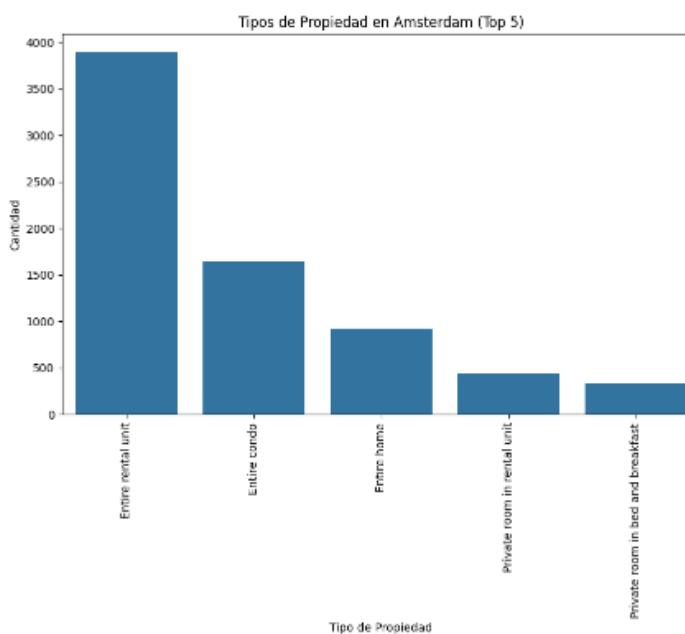
Como siguiente tenemos los tipos de cuarto y las principales amenidades que se ofrecen en estas ciudades :



en Amsterdam como prioridad se tiene al WiFi sin embargo en chicago la alarma de humo es indispensable dentro de los departamentos o casas, esto puede depender por el entorno de y tipo de ambiente que se vive dentro de estas ciudades sin embargo para ser competitivos el WiFi en la actualidad es indispensable por lo que debería ser prioridad en la mayoría de las ciudades en las que esté presente AIRBNB, como menciono anteriormente esto depende del clima, entorno y circunstancias en el que un usuario viaje o el host preste.

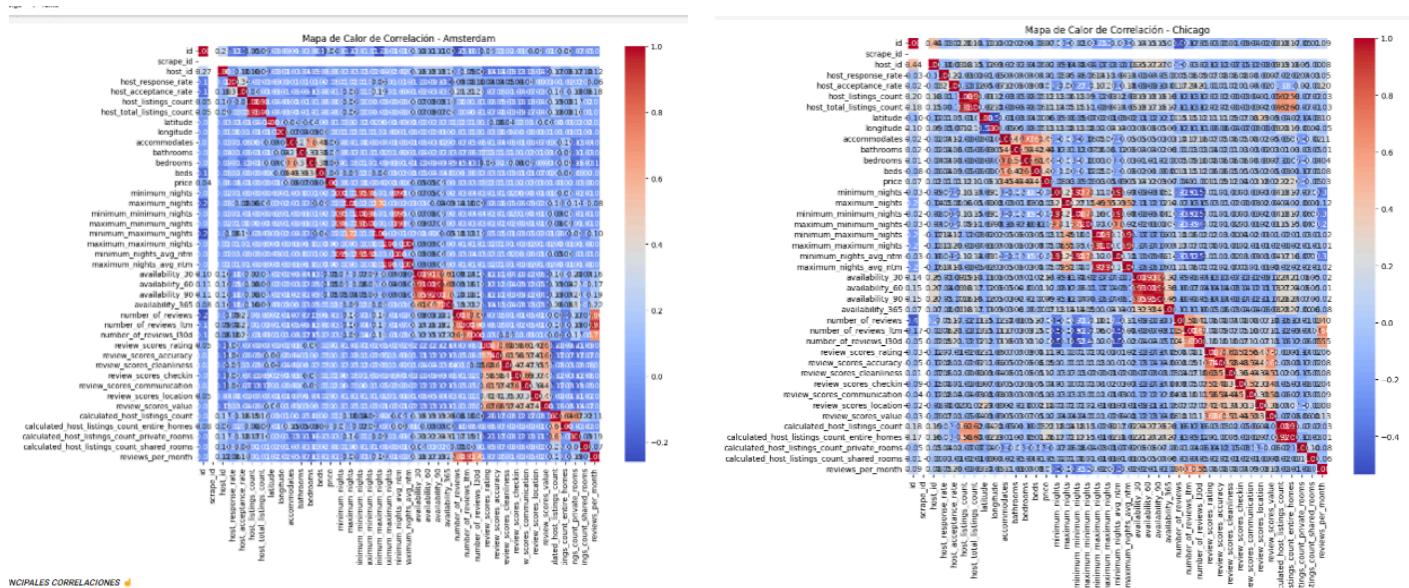


Lo siguiente son dos gráficas de barras que muestran los tipos de propiedades que prestan sus servicios a los usuarios de AIRBNB en el que fueron filtradas las 5 primeras más famosas o rentadas, la respuesta más destacada fue “Entire rental unit” lo que quiere decir una unidad completa (que incluya todos los

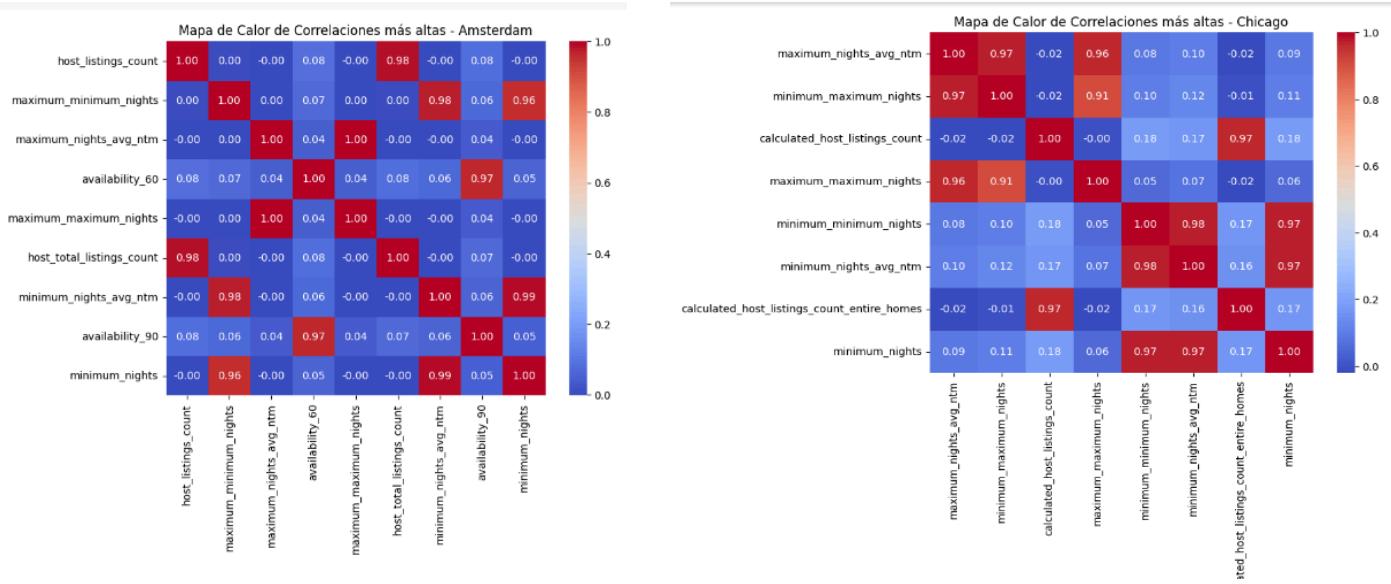


servicios y amenidades) sin embargo en ambas ciudades el lugar 5 cambia a “Private room in home” en chicago y a “Private room in bed and breakfast” en Ámsterdam. A simple vista se puede observar que estas conductas en la elección de tipos de propiedades es más cambiante en la ciudad de Ámsterdam que en Chicago que las 4 anteriores a la primera son casi constantes a comparación que en Ámsterdam que presenta una línea negativa exponencialmente.

De acuerdo a las instrucciones dadas se obtuvo el siguiente mapa de color que muestra TODAS las variables cuantitativas de ambas ciudades.



Sin embargo para mayor claridad y mejor visualización se filtraron las principales correlaciones de cada ciudad para una mejor interpretación y se obtuvo lo siguiente:



Con esta gráfica podemos identificar como ciertas variables tienen cierto impacto dentro de otras variables de manera positiva o negativa, gracias a sus correlaciones y el color podemos distinguir las más importantes y afectadas.

De igual manera se realizó un análisis descriptivo en donde las correlaciones encontradas en los datos de Amsterdam y Chicago proporcionan una visión clara de las relaciones entre diferentes variables

Correlaciones más altas para Amsterdam:		
maximum_maximum_nights	maximum_nights_avg_ntm	1.000000
minimum_nights_avg_ntm	minimum_nights	0.990217
host_listings_count	host_total_listings_count	0.984839
minimum_nights_avg_ntm	maximum_minimum_nights	0.975424
availability_60	availability_90	0.968158
minimum_nights	maximum_minimum_nights	0.958905
dtype: float64		

Correlaciones más altas para Chicago:		
minimum_nights_avg_ntm	minimum_minimum_nights	0.975574
minimum_nights	minimum_minimum_nights	0.974889
maximum_nights_avg_ntm	minimum_maximum_nights	0.973785
minimum_nights	minimum_nights_avg_ntm	0.972430
calculated_host_listings_count	calculated_host_listings_count_entire_homes	0.967846
maximum_maximum_nights	maximum_nights_avg_ntm	0.961243
dtype: float64		

relacionadas con la disponibilidad y las políticas de noches en Airbnb. En ambas ciudades, se observa una fuerte consistencia entre las variables que miden noches mínimas y máximas, así como entre

los listados de los anfitriones. Estas relaciones sugieren que las políticas de disponibilidad y noches mínimas/máximas son implementadas de manera consistente por los anfitriones y reflejan prácticas estandarizadas en cada ciudad. Este análisis puede ser útil para la toma de decisiones estratégicas relacionadas con la gestión de propiedades y la optimización de listados en diferentes mercados.

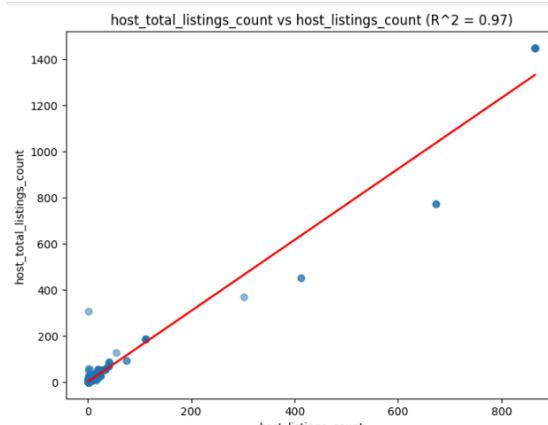
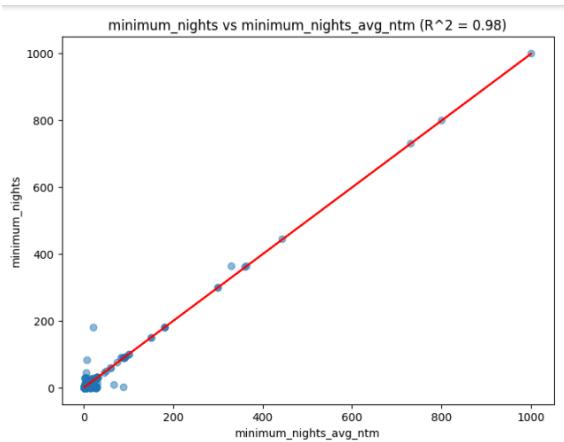
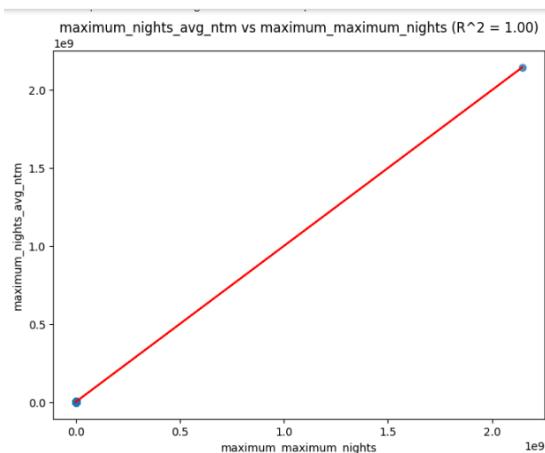
Gracias a lo anterior podemos llegar a algunas de las siguientes conclusiones:

La alta correlación indica que el promedio de noches mínimas en el tiempo (`minimum_nights_avg_ntm`) está muy bien alineado con el número mínimo de noches mínimas (`minimum_minimum_nights`). Esto sugiere una consistencia en la aplicación de políticas de noches mínimas a lo largo del tiempo.

La alta correlación sugiere que las noches mínimas generales (`minimum_nights`) están estrechamente relacionadas con el número mínimo de noches mínimas permitidas (`minimum_minimum_nights`). Esto refleja que los anfitriones que establecen un número mínimo general tienden a establecer valores mínimos consistentes.

La alta correlación muestra que el número total de listados calculados por anfitrión (`calculated_host_listings_count`) está estrechamente relacionado con el número de listados de casas enteras (`calculated_host_listings_count_entire_homes`). Esto sugiere que la mayoría de los listados de los anfitriones son casas enteras.

A continuación se muestran las 3 principales correlaciones (o las mejores) de acuerdo al modelo de regresión lineal con sus respectivas R cuadradas y correlaciones las cuales se explicaron anteriormente (de acuerdo a los datos anteriores). De igual manera en la parte inferior se muestran las características de algunas de estas y de otras correlaciones de ambas ciudades.





```
Modelo: maximum_nights_avg_ntm ~ maximum_maximum_nights
Coeficiente de correlación: 1.00
Coeficiente de determinación (R^2): 1.00
Intersección: -7.35
Pendiente: 1.00

Modelo: minimum_nights ~ minimum_nights_avg_ntm
Coeficiente de correlación: 0.99
Coeficiente de determinación (R^2): 0.98
Intersección: -0.12
Pendiente: 1.00

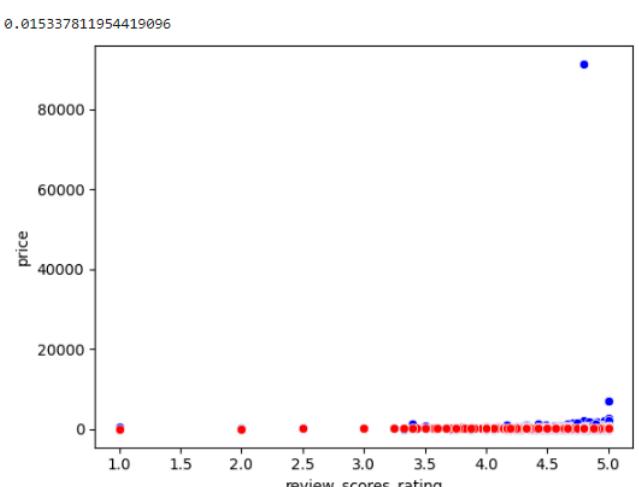
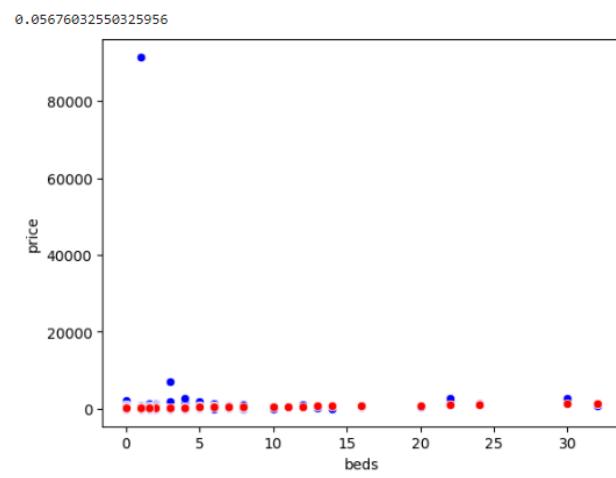
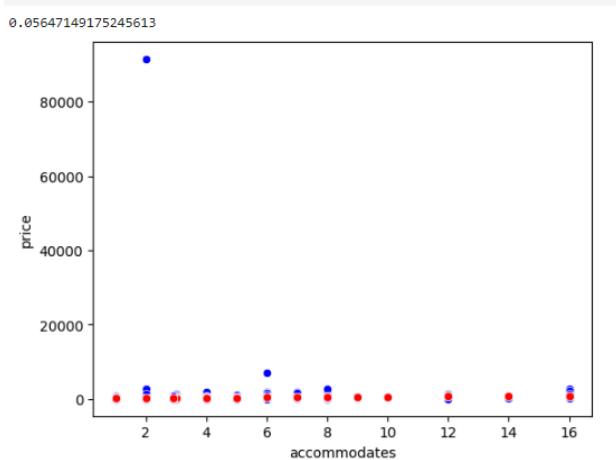
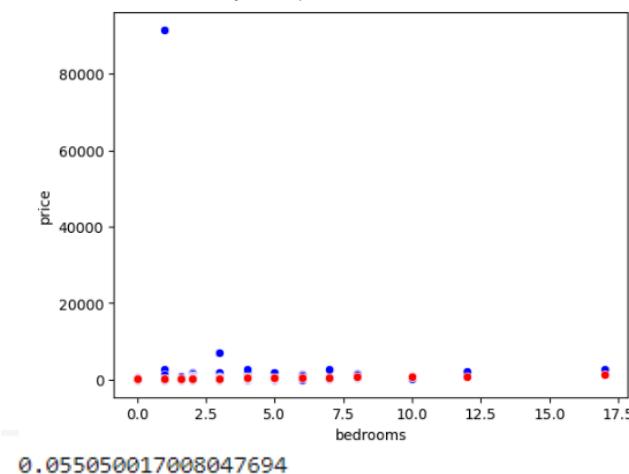
Modelo: host_total_listings_count ~ host_listings_count
Coeficiente de correlación: 0.98
Coeficiente de determinación (R^2): 0.97
Intersección: -0.21
Pendiente: 1.54
```

```
Modelo: minimum_minimum_nights ~ minimum_nights_avg_ntm
Coeficiente de correlación: 0.98
Coeficiente de determinación (R^2): 0.95
Intersección: 0.02
Pendiente: 0.93

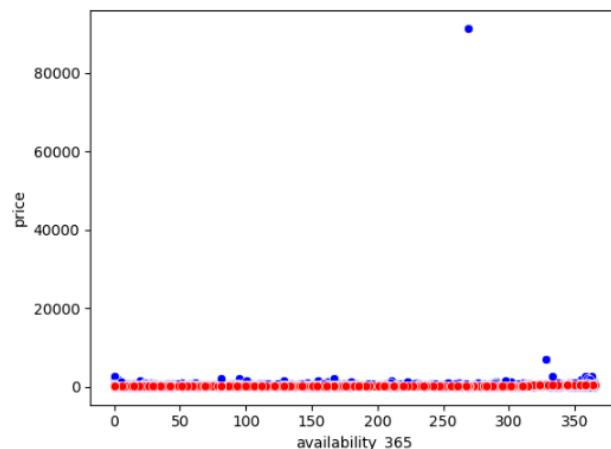
Modelo: minimum_minimum_nights ~ minimum_nights
Coeficiente de correlación: 0.97
Coeficiente de determinación (R^2): 0.95
Intersección: 0.40
Pendiente: 0.96

Modelo: minimum_maximum_nights ~ maximum_nights_avg_ntm
Coeficiente de correlación: 0.97
Coeficiente de determinación (R^2): 0.95
Intersección: -23.07
Pendiente: 1.01
```

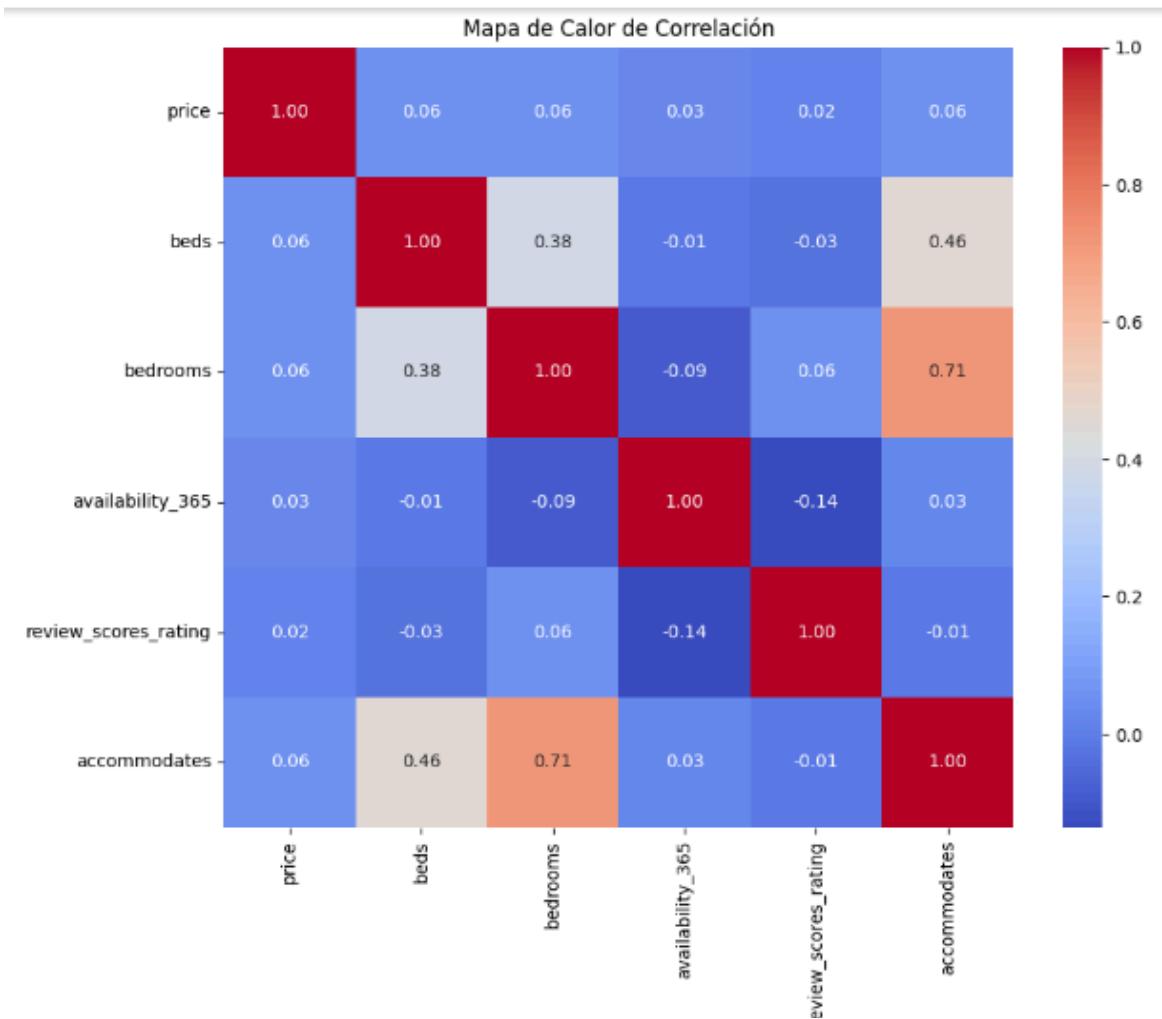
Finalmente se realizaron 5 regresiones lineales con métodos vistos en clase con variables elegidas al azar pero con algo en común, todas las variables fueron relacionadas con el precio.



0.03410294784488439



Podemos notar que a comparación del mapa de calor son parecidas sin embargo se llegan a redondear sin embargo son muy similares, por otro lado también podemos notar que entre las variables seleccionadas tienen una correlación más fuerte aunque no fueron graficadas.



Finalmente podemos concluir que dentro de los servicios de AIRBNB existen muchas variables que gracias a su correlación pueden ser de valor tanto por sus datos actuales como los predictivos que podemos obtener a través de este tipo de modelos lineales regresivos aunque cabe recalcar que no siempre son certeros y siempre existe un margen mínimo, algo que recalcar dentro de este análisis es la atención al cliente y el tiempo de espera que pueden tener para ser atendidos, ajustarse y comprender las necesidades de cada ciudad sería clave para formar estrategias y tomar decisiones.

Podemos decir que la experiencia del usuario y el éxito de los listados están significativamente influenciados por variables específicas como el tiempo de respuesta de los anfitriones, el tipo de propiedad ofrecida, y las amenidades disponibles. La rapidez en la respuesta de los hosts en Chicago, en comparación con Ámsterdam, destaca la importancia de la atención al cliente en la satisfacción y lealtad del usuario.

Además, las correlaciones encontradas entre las políticas de noches mínimas y la consistencia en los listados sugieren que los anfitriones siguen prácticas estandarizadas que pueden ser aprovechadas para mejorar la gestión de propiedades y la optimización de precios en ambos mercados.

El análisis también subraya la necesidad de adaptar las estrategias a las particularidades de cada ciudad, como la prioridad del WiFi en Ámsterdam frente a la alarma de humo en Chicago, lo que refleja diferentes expectativas y necesidades de los huéspedes. Finalmente, los modelos de regresión lineal aplicados indican que, aunque hay fuertes correlaciones que pueden ayudar a predecir precios y comportamientos, es crucial considerar la variabilidad y margen de error inherente a estos modelos al tomar decisiones estratégicas.