



Tecnológico de Monterrey

Plataformas de Analítica de Negocios Para Organizaciones

Reporte de evidencia

José Luis Spíndola Gómez | A01734737

Erick Antonio Nava Melchy | A01424422

Alexander Penagos Muñoz | A01738144

11 de septiembre de 2025

Tabla de contenido:

Introducción.....	3
Análisis y cálculo de medidas.....	3
Indicadores.....	4
Conclusión.....	6
Link de repositorio:.....	6
Referencias.....	7

Nota: Se subieron los CSV de las actividades realizadas de forma individual para cada paso, sin embargo, el código se unificó para obtener los insights que se presentaron con un mismo formato (se anexan las bases de México con el código general aplicado).

Introducción

Este proyecto tiene como objetivo entender, limpiar y analizar información sobre distintas ciudades en Airbnb para compararlas y obtener insights.

Airbnb fue fundada en 2007 en San Francisco cuando sus fundadores rentaron una habitación a 3 huéspedes; la plataforma ha crecido hasta llegar a los 5 millones de anfitriones en el mundo que tiene actualmente (Airbnb, s. f.).

Nuestro equipo escogió las ciudades de Washington DC, Venecia y Jersey City, las cuales compararemos con la Ciudad de México. Gracias a este análisis encontramos características particulares que hacen única a cada ciudad, lo que podría ayudar a generar estrategias de marketing y oferta de propiedades.

Análisis y cálculo de medidas

Para realizar el análisis, comenzamos con el tratamiento de los valores nulos en las distintas variables usando diferentes métodos, como la moda en el caso de los baños y las camas, la mediana en las recámaras, la media en el precio, valores centinela para fechas que permiten identificarlas manteniendo el formato y sin información, entre otros. Esto permite identificar estos datos sin afectar la base.

Posteriormente, realizamos el tratamiento de valores atípicos para evitar sesgos extremos en cada base. El primer paso fue la identificación visual gracias a un boxplot, lo que facilitó encontrarlos; posteriormente, nos aseguramos de que las calificaciones estuvieran del 0 al 5 para evitar valores erróneos. Después, las bases de datos fueron tratadas con el método IQR de manera homogénea (los outliers encontrados fueron reemplazados por la mediana en lugar de la media para no crear valores decimales en variables que fueran de números enteros) y se procedió con la generación del CSV.

Para finalizar, realizamos la extracción de características, comenzando por el manejo de datos centinela; posteriormente, usamos la función `value_counts()` para generar gráficas

tanto de pastel como de barras que facilitarán la visualización. Es importante mencionar que se decidió optar por un número específico de categorías en lugar de aplicar la regla de Sturges con la finalidad de tener el mismo número de bins en las 4 ciudades.

Indicadores

Escogimos 10 variables finales de la base de datos para realizar el análisis, las cuales se dividen en 3 categorías que consideramos relevantes para medir la experiencia general de los huéspedes; estas son atención al cliente, seguridad y preferencias de hospedaje. A continuación se encuentran todas las variables seleccionadas junto con los insights de cada una:

Superhost: Mide si los anfitriones tienen o no la categoría de superhost, la cual es una insignia que da la plataforma por reseñas positivas, responder rápido y evitar las cancelaciones (Airbnb, s. f.). El análisis indica que la Ciudad de México es la que menor porcentaje tiene de las 4 analizadas, con 36.7% de las propiedades.

Tiempo de respuesta: Son las distintas categorías en las que se encuentran los anfitriones dependiendo de si en promedio contestan los mensajes en una hora, varias horas, un día, varios días o más y sin información. Esto es útil para medir la calidad del servicio; destacamos que Venecia tiene el mayor porcentaje que responde en menos de una hora.

Identidad verificada: Es un elemento de seguridad para los huéspedes, el cual consideramos muy relevante; en este caso descubrimos que Washington tiene la tasa más baja (87.3%) a pesar del buen servicio que brindan.

Foto de perfil: Esta es otra variable esencial para la seguridad, ya que ayuda a distinguir al anfitrión y darle tranquilidad a los huéspedes; en este caso, Venecia tiene el porcentaje más bajo, lo que se debe trabajar tan pronto como sea posible.

Tipo de propiedad: Esta categoría nos ayuda a conocer las propiedades que prefieren los huéspedes para quedarse y, por lo tanto, que Airbnb busque más alojamientos así. Algo a destacar en este caso es que en las 4 ciudades entire rental unit es la opción más popular.

Tipo de cuarto: Al igual que el tipo de propiedad, los cuartos nos sirven para saber qué prefieren los huéspedes y para conocer mejor las opciones que ofrece cada ciudad. Lo más interesante es que en las 4 bases de datos, entre home/apartment es la más popular, mientras que la segunda opción varía en Venecia, donde en lugar de ser private room como en las demás, es hotel room (seguramente debido a un convenio local).

Precio: Esta variable es muy importante porque permite saber cuánto están dispuestos a pagar los huéspedes y posiblemente aumentar las propiedades en ese rango, ya sea para incrementar las reservas o las ganancias con opciones de lujo. Ciudad de México es la ciudad que ofrece un mayor rango de precios, mientras que Venecia se centra en los más baratos; sin embargo, no descartamos que la base se encuentre en distintas monedas debido al alto costo de algunas propiedades.

Camas: El número de camas nos ayuda a complementar la información de las propiedades y a conocer cuántos huéspedes se podrían estar quedando en las mismas. En general, las 4 ciudades tienen pocas camas en los alojamientos, pero Jersey City y Venecia tienen opciones que llegan a 6 camas, lo cual supera el máximo de 3 de Washington y CDMX.

Reviews por mes: El número de revisiones por mes puede ser útil tanto para saber cuántas reservaciones se están haciendo, como para ver si los huéspedes están evaluando o no a los anfitriones. Las 4 gráficas tienen un comportamiento similar, donde la moda es 1 o 2 y va disminuyendo hasta un máximo de 6 o incluso 7. Es importante mencionar que las reservas de varias noches pueden afectar este número, ya que solo se da un reembolso por estadía.

Score general: Consideramos que esta es la categoría más importante de todas, ya que resume la satisfacción general y permite hacer mejoras cuando es necesario. Afortunadamente, las 4 ciudades tienen la gran mayoría de los reviews en 5, seguidos por 4; esto nos permite decir que los huéspedes se sienten felices en Airbnb.

Conclusión

Este análisis nos permitió conocer mejor la oferta que tiene Airbnb e identificar las similitudes y diferencias en estas ciudades. Estar monitoreando estos indicadores le puede ser muy útil tanto a la plataforma como a los anfitriones para mejorar su servicio y obtener más reservaciones.

Además, la evidencia nos permitió mejorar nuestras habilidades para tratar las bases y analizar los resultados, así como unificar nuestro proceso para que todas las ciudades fueran manejadas con los mismos criterios y gráficos.

Link de repositorio:

- https://github.com/a01424422-hue/Evidencia_PANO/tree/main/Extracci%C3%B3n%20de%20los%20datos

Referencias

- Get the Data. (s.f.). Inside Airbnb. <https://insideairbnb.com/get-the-data/>
- Acerca de Airbnb: qué es y cómo funciona - Centro de ayuda de Airbnb. (n.d.). Airbnb. <https://www.airbnb.mx/help/article/2503>
- Jersey City, New Jersey population 2025. (n.d.). World Population Review. <https://worldpopulationreview.com/us-cities/new-jersey/jersey-city>
- Washington, DC | Data USA. (n.d.). Data USA. <https://datausa.io/profile/geo/washington-dc>
- Sonia. (n.d.). Números de Venecia – Venecia. http://espanol.venice-guide.info/numeros_venecia/
- Ciudad de México. (n.d.). Data México. <https://www.economia.gob.mx/datamexico/es/profile/geo/ciudad-de-mexico-cx>
- *Cómo funciona el programa de Superanfitriones* - Centro de ayuda de Airbnb. (n.d.). Airbnb. <https://www.airbnb.mx/help/article/828>