



# Tecnológico de Monterrey

## **Actividad 1: Entendimiento de los Datos**

Lucero Jannete López García A01736938

**Analítica de datos y herramientas de inteligencia artificial  
I (Gpo 101)**

**Fecha**  
16/02/2025

**\*\*Las variables subrayadas son porque tienen similitud de justificación pero difieren en números, características o tiempos**

#Col	Variable	Tipo de dato	Descripción	Justificación
1	ID	Numérica (int)	Identificador único de la propiedad en Airbnb	Esta variable es fundamental porque permite diferenciar cada propiedad de manera única. Sin un identificador exclusivo, se generarían problemas de duplicación y falta de precisión en el análisis de datos.
2	last_scraped	Texto (categórica)	Fecha y hora en la que se obtuvo la información	Me ayuda a evaluar qué tan actualizados están los datos y detectar cambios en la información de los listados a lo largo del tiempo y analizar por qué algunas variables han desaparecido o si se generaron nuevas y así saber las tendencias del consumidor.
3	Source	Texto (categórica)	Una de las opciones de "búsqueda de barrio" o "búsqueda previa". "Búsqueda de barrio" significa que el anuncio se encontró al buscar en la ciudad, mientras que "búsqueda previa" significa que el anuncio se vio en otra búsqueda realizada en los últimos 65 días y se confirmó que el anuncio aún estaba disponible en el sitio de Airbnb.	Es relevante para conocer el origen de los datos y determinar si la información proviene de un scraping reciente o de registros anteriores, lo que impacta su confiabilidad.
4	host_id	Texto (categórica)	Identificador único de Airbnb para el anfitrión/usuario	Permite identificar de manera única a cada anfitrión y analizar su actividad dentro de la plataforma. Esto es útil para detectar patrones de comportamiento y evaluar la oferta de alojamientos según la cantidad de propiedades que maneja cada anfitrión.
5	host_response_time	el programa la pone como objeto pero es un porcentaje	Tiempo que tarda en responder el anfitrión	Me da una idea de qué tan rápido responde un anfitrión a las solicitudes de los huéspedes. Esto puede influir en la experiencia del usuario y en la posibilidad de que una reserva se concrete.
6	host_response_rate	el programa la pone como objeto pero es un porcentaje	Tasa de respuesta del host	Es clave para medir qué tan receptivo es un anfitrión ante las solicitudes. Un porcentaje bajo puede indicar que el anfitrión no está muy activo en la plataforma, lo que podría afectar la disponibilidad de su propiedad.
7	host_acceptance_rate	el programa la pone como objeto pero es un porcentaje	Tarifa a la que un anfitrión acepta solicitudes de reserva.	Me ayuda a entender la frecuencia con la que un anfitrión acepta solicitudes de reserva. Un anfitrión con una tasa de aceptación baja puede representar una opción poco confiable para los huéspedes.

8	host_is_superhost	Texto (categórica)	El anfitrión es superhost o no	Es una variable importante porque permite identificar si el anfitrión tiene un estatus destacado dentro de la plataforma. Los superhosts suelen ofrecer una mejor experiencia al huésped, por lo que este dato es relevante para la toma de decisiones.
9	host_listings_count	Numérica (int)	La cantidad de anuncios que tiene el anfitrión (según cálculos desconocidos de Airbnb)	Me ayuda a conocer si un anfitrión gestiona varias propiedades, lo que puede indicar si es un anfitrión profesional o alguien que solo alquila una vivienda ocasionalmente. Esto puede influir en la calidad del servicio y la disponibilidad de los alojamientos.
10	host_total_listings_count	Numérica (int)	La cantidad de anuncios que tiene el anfitrión (según cálculos desconocidos de Airbnb)	
11	host_verifications	Texto (categórica)	Cuenta con email, teléfono, mail de trabajo etc.	Es importante saber qué tipo de verificación tiene un anfitrión, ya que esto aumenta la confianza de los huéspedes y la seguridad en la plataforma.
12	neighbourhood	Texto (categórica)	Lugar dónde se encuentra el apartamento, habitación, casa etc.	Conocer la ubicación de la propiedad me permite analizar la oferta de alojamientos por zona, identificar tendencias de preferencia y evaluar los precios en distintas áreas de la ciudad.
13	latitude	Numérica (float)	Utiliza la proyección del Sistema Geodésico Mundial (WGS84) para latitud y longitud.	Estas coordenadas me permiten ubicar con precisión cada propiedad en un mapa y hacer análisis espaciales sobre la distribución de los alojamientos en la ciudad.
14	longitude	Numérica (float)		
15	property_type	Texto (categórica)	Tipo de propiedad seleccionada por el propietario. Los hoteles y los Bed and Breakfasts se describen como tales por sus anfitriones en este campo	Es clave para diferenciar los distintos tipos de alojamientos disponibles. No es lo mismo una casa, un departamento o un hotel, y esta información influye en la elección de los huéspedes.
16	room_type	Texto (categórica)	Todas las viviendas se agrupan en los siguientes tres tipos de habitaciones:  Alojamiento completo Habitación privada Habitación compartida	Me permite conocer qué tipo de espacio se ofrece (alojamiento completo, habitación privada o compartida), lo que es relevante para los huéspedes al momento de buscar una opción acorde a sus necesidades.
17	accommodates	Numérica (Int)	La capacidad máxima del lugar	Es importante porque determina cuántas personas pueden alojarse en el lugar, lo que influye en la oferta y en el precio.
18	bathrooms	Numérica (Int)	Baños máximos del lugar	Estas variables ayudan a evaluar la comodidad del alojamiento y su capacidad real. Son datos clave para los huéspedes al momento de elegir una propiedad.
19	bedrooms	Numérica (Int)	El número de habitaciones	

20	beds	Numérica (Int)	El número de camas	
21	amenities	Texto (categórica)	Tipo de facilidades con las que cuenta (TV, refrigerador, elevador, etc.)	Me permite analizar qué servicios y comodidades ofrece cada alojamiento, lo que puede ser un factor determinante en la elección de los huéspedes y en el precio del alquiler.
22	price	Numérica (Int)	Precio diario en moneda local.	Es uno de los datos más importantes porque permite hacer comparaciones entre alojamientos y analizar la oferta y demanda dentro de la plataforma.
23	minimum_nights	Numérica (Int)	Número mínimo de noches de estancia para el alojamiento (las reglas del calendario pueden ser diferentes)	Estas variables ayudan a entender la flexibilidad del anfitrión respecto a las reservas y a identificar qué tipo de clientes puede atraer cada alojamiento (viajeros de corta o larga estancia).
24	maximum_nights	Numérica (Int)	Número máximo de noches de estancia para el alojamiento (las reglas del calendario pueden ser diferentes)	
25	minimum_minimum_nights	Numérica(Int)	El valor mínimo de noche más pequeño del calendario (mirando 365 noches en el futuro)	Identificar cuál es la menor cantidad de noches que un anfitrión ha permitido en su calendario en el último año. Esto es útil para entender la flexibilidad en las reservas y detectar listados que permiten estadías muy cortas, lo cual puede atraer a viajeros que solo necesitan hospedaje por pocos días.
26	maximum_minimum_nights	Numérica (Int)	El valor mínimo de noche más grande del calendario (mirando 365 noches en el futuro)	Cuál ha sido el número mínimo de noches más alto que un anfitrión ha establecido. Puede indicar estrategias como evitar estancias de corta duración para reducir costos de limpieza o enfocarse en huéspedes que se queden por períodos más largos.
27	minimum_maximum_nights	Numérica (Int)	el valor máximo de noche más pequeño del calendario (mirando 365 noches en el futuro)	Menor número máximo de noches permitido en un calendario. Es útil para analizar si hay anfitriones que prefieren solo estadías cortas,
28	maximum_maximum_nights	Numérica (Int)	el valor máximo de noche más grande del calendario (mirando 365 noches en el futuro)	Mayor número de noches que un anfitrión ha permitido. Esto es relevante porque algunos anfitriones pueden enfocarse en alquileres a largo plazo, y entender esta tendencia me ayuda a diferenciar listados dirigidos a turistas de los que apuntan a huéspedes que buscan estancias extendidas.
29	minimum_nights_avg_ntm	Numérica (Int)	el valor mínimo de noche promedio del calendario (mirando 365 noches en el futuro)	Identificar tendencias en la flexibilidad de los anfitriones a lo largo del año y observar si hay cambios en la duración mínima por temporadas.

30	maximum_nights_avg_ntm	Numérica (Int)	el valor máximo de noche promedio del calendario (mirando 365 noches en el futuro)	Se pueden detectar patrones de estadías a largo plazo.
31	availability_30	Numérica (Int)	Disponibilidad. La disponibilidad del alojamiento x días en el futuro según lo determine el calendario. Tenga en cuenta que un alojamiento puede no estar disponible porque ha sido reservado por un huésped o bloqueado por el anfitrión.	Son clave para medir la disponibilidad de los alojamientos en diferentes periodos de tiempo y entender cómo varía la oferta según la época del año.
32	availability_60	Numérica (Int)		
33	availability_90	Numérica (Int)		
34	availability_365	Numérica (Int)		
35	number_of_reviews	Numérica (Int)	El número de reseñas que tiene el anuncio.	Me ayudan a medir la popularidad y reputación de un alojamiento. Si un anuncio tiene muchas reseñas, es más probable que sea confiable y que haya tenido una alta demanda.
36	number_of_reviews_ltm	Numérica (Int)	La cantidad de reseñas que tiene el anuncio (en los últimos 12 meses)	
37	number_of_reviews_l30d	Numérica (Int)	La cantidad de reseñas que tiene el anuncio (en los últimos 30 meses)	
38	review_scores_rating	Numérica (float)	Puntuación de calificación	Estas puntuaciones me permiten evaluar la calidad del alojamiento y del servicio del anfitrión desde la perspectiva de los huéspedes. Son fundamentales para el análisis de satisfacción del cliente.
39	review_scores_accuracy	Numérica (float)	Puntuación de precisión	
40	review_scores_cleanliness	Numérica (float)	Puntuación de limpieza	
41	review_scores_checkin	Numérica (float)	Puntuación de check in	
42	review_scores_communication	Numérica (float)	Puntuación de comunicación	
43	review_scores_location	Numérica (float)	Puntuación de ubicación	
44	review_scores_value	Numérica (float)	Puntuación de valor	
45	license	Texto (categórica)	Tipo de licencia	Conocer si un alojamiento tiene licencia me ayuda a verificar si cumple con las normativas locales y si está operando de manera legal.
46	calculated_host_listings_count	Numérica (Int)	La cantidad de anuncios que el anfitrión tiene en el scraping actual, en la geografía de la ciudad/región.	Estas variables me permiten conocer cuántos alojamientos gestiona un anfitrión en la región, lo que puede ayudar a identificar patrones de inversión en alquileres turísticos y analizar la competencia dentro del mercado.

47	calculated_host_listings_count_entire_homes	Numérica (Int)	La cantidad de anuncios de casas/apartamentos completos que el anfitrión tiene en el scraping actual, en la geografía de la ciudad/región.	
48	calculated_host_listings_count_private_rooms	Numérica (Int)	La cantidad de anuncios de habitaciones privadas que el anfitrión tiene en el scraping actual, en la geografía de la ciudad/región.	
49	calculated_host_listings_count_shared_rooms	Numérica (Int)	La cantidad de anuncios de habitaciones compartidas que el anfitrión tiene en el scraping actual, en la geografía de la ciudad/región.	
50	reviews_per_month	Numérica(float)	La cantidad promedio de revisiones por mes que tiene el anuncio durante su vigencia.	Es una métrica útil para entender la frecuencia con la que los huéspedes dejan reseñas, lo que me permite estimar la actividad y popularidad de un alojamiento en la plataforma.