



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN

**Licenciatura en Ingeniería en Tecnologías de la
Información**

ETAPA 1: ANÁLISIS

INGENIERÍA DE SOFTWARE

PhD. Alfredo García Suárez

Elaborado por

Romero Lagunes Jossette Yeraldin

Matrícula: 202354620

10 de enero del 2025

1. Traducción y descripción de las variables obtenidas de la base de datos

Variable	Traducción	Descripción
id	Identificación	Identificador único de Airbnb para el anuncio.
listing_url	URL del anuncio	Permite acceder directamente a la publicación del anuncio.
last_scraped	Último borrador	Hora UTC. La fecha y hora en que se "elimino" este listado.
source	Fuente	Uno de "búsqueda de barrio" o "borrado previo". "búsqueda de vecindario" significa que el anuncio se encontró buscando en la ciudad, mientras que "borrado anterior" significa que el anuncio se vio en otro respaldo realizado en los últimos 65 días y se confirmó que el anuncio todavía estaba disponible en el sitio de Airbnb.
name	Nombre	Nombre del listado (anuncio).
description	Descripción	Descripción detallada del listado.
neighborhood_overview	Descripción general del vecindario	Descripción del anfitrión del vecindario.
picture_url	URL de la imagen	URL de la imagen de tamaño anormal alojada en Airbnb para el anuncio
host_id	Identificación del anfitrión	Identificador único de Airbnb para el anfitrión/usuario
host_url	URL del anfitrión	La página de Airbnb para el anfitrión.
host_name	Nombre del anfitrión	Nombre del anfitrión. Por lo general, solo el nombre(s).
host_since	anfitrión _desde	La fecha en que se creó el anfitrión/usuario. Para los anfitriones que son invitados de Airbnb, esta podría ser la fecha en la que se registraron como invitados.
host_location	ubicación_anfitrión	La ubicación auto informada por el anfitrión.
host_about	Sobre_anfitrión	Descripción sobre el anfitrión
host_response_time	Tiempo_respuesta_anfitrión	Tiempo que tarda el anfitrión en responder los mensajes
host_response_rate	tasa de respuesta del anfitrión	Porcentaje de mensajes respondidos por el anfitrión, expresado como un porcentaje del 0% al 100%
host_acceptance_rate	Tasa de aceptación del anfitrión	Esa tarifa a la que un anfitrión acepta solicitudes de reserva.
host_is_superhost	Anfitrión es superanfitrión	Indica si el anfitrión tiene el estatus de Superhost. Valores: "t" (true) o "f" (false).
host_thumbnail_url	url_miniatra_anfitrión	URL de la imagen en miniatura del anfitrión, pequeña foto de perfil.
host_picture_url	url_imagen_anfitrión	URL de la imagen de perfil completa del anfitrión, más grande que la miniatura.

Variable	Traducción	Descripción
host_neighbourhood	Vecindario_anfitrión	Barrio del anfitrión según su descripción, más general que la ubicación exacta.
host_listings_count	Recuento de listados de anfitriones	La cantidad de anuncios que tiene el anfitrión (según cálculos desconocidos de Airbnb)
host_total_listings_count	Recuento total de listados de anfitrión	
host_verifications	Verificaciones de anfitrión	Lista de métodos de verificación completados por el anfitrión, como correo, teléfono, ID oficial, etc.
host_has_profile_pic	Anfitrión tiene foto de perfil	Indica si el anfitrión tiene una foto de perfil. Valores: "t" (true) o "f" (false).
host_identity_verified	Identidad del anfitrión verificada	Indica si la identidad del anfitrión ha sido verificada. Valores: "t" (true) o "f" (false).
neighbourhood	Vecindario	Barrio donde se encuentra el alojamiento, descripción general de la ubicación.
neighbourhood_cleansed	Barrio_limpio	El vecindario está geocodificado utilizando la latitud y longitud frente a los vecindarios definidos por archivos de forma digitales abiertos o públicos.
neighbourhood_group_cleansed	Grupo de vecinos limpiado	El grupo de vecinos está geocodificado utilizando la latitud y longitud de los barrios definidos por archivos de forma digitales abiertos o públicos.
latitude	Latitud	Utiliza la proyección del Sistema Geodésico Mundial (WGS84) para latitud y longitud.
longitude	longitud	
property_type	tipo_propiedad	Tipo de propiedad autoseleccionado. Los hoteles y Bed and Breakfast son descritos como tales por sus anfitriones en este campo.
room_type	tipo_habitación	<p>[Casa entera/apto Habitación privada Habitación compartida Hotel]. Todas las viviendas se agrupan en los siguientes tres tipos de habitaciones:</p> <ul style="list-style-type: none"> • Todo el lugar • habitación privada • Habitación compartida <p>Todo el lugar Los lugares enteros son mejores si buscas un hogar lejos del hogar. Con un alojamiento completo, tendrás todo el espacio para ti. Esto suele incluir un dormitorio, un baño, una cocina y una entrada independiente y exclusiva. Los anfitriones deben indicar en la descripción si estarán en la propiedad o no (por ejemplo: "El anfitrión ocupa el primer piso de la casa") y proporcionar más detalles sobre el anuncio.</p> <p>Habitaciones privadas Las habitaciones privadas son ideales cuando prefieres un poco de privacidad y aún valoras una conexión local. Cuando reservas una habitación privada, tendrás</p>

Variable	Traducción	Descripción
		<p>tu propia habitación privada para dormir y podrás compartir algunos espacios con otras personas. Es posible que tengas que caminar por espacios interiores que otro anfitrión o huésped pueda ocupar para llegar a tu habitación.</p> <p>Habitaciones compartidas Las habitaciones compartidas son para cuando no te importa compartir un espacio con otras personas. Cuando reservas una habitación compartida, dormirás en un espacio compartido con otras personas y compartirás todo el espacio con otras personas. Las habitaciones compartidas son populares entre los viajeros flexibles que buscan nuevos amigos y estancias económicas.</p>
accommodates	Alojamiento	La capacidad máxima de huéspedes que puede alojar.
bathrooms	baños	Número de baños en el listado.
bathrooms_text	texto de baños	El número de baños en el listado. En el sitio web de Airbnb, el campo de los baños ha evolucionado de un número a una descripción textual. Para raspaduras más antiguas, se utilizan los baños
bedrooms	Dormitorios	Número de dormitorios
beds	Camas	Número de camas
amenities	Comodidades	Lista de comodidades ofrecidas, como WiFi, aire acondicionado, estacionamiento, etc.
price	Precio	Precio diario en moneda local. NOTA: el signo \$ es un artefacto técnico de la exportación, ignórelo
minimum_nights	Noches mínimas	Número mínimo de noches de estadía para el alojamiento (las reglas del calendario pueden ser diferentes)
maximum_nights	Noches máximas	Número máximo de noches de estadía para el alojamiento (las reglas del calendario pueden ser diferentes)
minimum_minimum_nights	Noches mínimas mínimas	El valor mínimo_noche más pequeño del calendario (mirando 365 noches en el futuro)
maximum_minimum_nights	Máximo mínimo de noches	El valor mínimo_noche más grande del calendario (mirando 365 noches en el futuro)
minimum_maximum_nights	noches_mínimas_máximas	El valor máximo_noche más pequeño del calendario (mirando 365 noches en el futuro)
maximum_maximum_nights	maximo_maximo_noches	El valor máximo_noche más grande del calendario (mirando 365 noches en el futuro)
minimum_nights_avg_ntm	noches_mínimo_avg_ntm	El valor promedio de noche mínima del calendario (mirando 365 noches en el futuro)

Variable	Traducción	Descripción
maximum_nights_avg_ntm	máximas_noches_avg_ntm	El valor promedio de máxima_noche del calendario (mirando 365 noches en el futuro)
calendar_updated	Actualización del calendario	Frecuencia de actualización del calendario del anuncio por el anfitrión.
has_availability	Tiene disponibilidad	[t=verdadero; f=falso]
availability_30	Disponibilidad_30	Disponibilidad_x. La disponibilidad del listado x días en el futuro según lo determine el calendario. Tenga en cuenta que es posible que un anuncio no esté disponible porque ha sido reservado por un huésped o bloqueado por el anfitrión.
availability_60	Disponibilidad_60	
availability_90	Disponibilidad_90	
availability_365	Disponibilidad_365	
calendar_last_scraped	Último borrador del calendario	Fecha en que los datos del calendario fueron extraídos por el sistema (AAAA-MM-DD).
number_of_reviews	Numero_de_reseñas	El número de reseñas que tiene el anuncio.
number_of_reviews_ltm	número_de_reseñas_ltm	El número de reseñas que tiene el anuncio (en los últimos 12 meses)
number_of_reviews_l30d	número_de_reseñas_l30d	El número de reseñas que tiene el anuncio (en los últimos 30 días)
first_review	Primera_revisión	La fecha de la primera/más antigua revisión.
last_review	Ultima_revisión	La fecha de la última/más reciente revisión.
review_scores_rating	calificación de puntuaciones de revisión	Calificación general del anuncio basada en reseñas, en una escala de 0 a 100.
review_scores_accuracy	revisar la precisión de las puntuaciones	Calificación de precisión de la descripción del anuncio según los huéspedes, en una escala de 0 a 10.
review_scores_cleanliness	revisar puntuaciones limpieza	Calificación de limpieza del alojamiento según los huéspedes, en una escala de 0 a 10.
review_scores_checkin	revisar puntajes de registro	Calificación de facilidad de check-in según los huéspedes, en una escala de 0 a 10.
review_scores_communication	Revisar puntuaciones de comunicación	Calificación de comunicación del anfitrión según los huéspedes, en una escala de 0 a 10.
review_scores_location	Revisar la ubicación de las puntuaciones	Calificación de la ubicación del alojamiento según los huéspedes, en una escala de 0 a 10.
review_scores_value	Revisar el valor de las puntuaciones	Calificación del valor percibido por el precio pagado, en una escala de 0 a 10.
license	Licencia	El número de licencia/permiso/registro
instant_bookable	reservable_instantáneamente	[t=verdadero; f=falso]. Si el huésped puede reservar automáticamente el alojamiento sin que el anfitrión tenga que aceptar su solicitud de reserva. Un indicador de una cotización comercial.
calculated_host_listings_count	recuento calculado de listados de anfitriones	La cantidad de listados que el anfitrión tiene en el borrador actual, en la geografía de la ciudad/región.

Variable	Traducción	Descripción
calculated_host_listings_count_entire_homes	listados de anfitriones calculados cuentan casas enteras.	La cantidad de listados completos de casas/apto. que el anfitrión tiene en el borrador actual, en la geografía de la ciudad/región.
calculated_host_listings_count_private_rooms	listados de anfitriones calculados cuentan las habitaciones privadas.	La cantidad de listados de habitaciones privadas que el anfitrión tiene en el borrador actual, en la geografía de la ciudad/región.
calculated_host_listings_count_shared_rooms	listados de anfitriones calculados cuentan las habitaciones compartidas.	La cantidad de listados de habitaciones compartidas que el anfitrión tiene en el borrador actual, en la geografía de la ciudad/región.
reviews_per_month	reseñas_por_mes	El número promedio de reseñas por mes que tiene el anuncio durante su vida útil. Seudocoe/~SQL: SI fecha_scrape - primera_revisión <= 30 ENTONCES número_de_revisiones ELSE número_de_revisiones / ((scrape_date - primera_revisión + 1) / (365/12))

2. Variables importantes para el análisis de datos respecto a los requerimientos del cliente Airbnb

No.	Variable	Justificación
Información del anuncio		
1	id	Al ser el identificador único de cada anuncio, considero que es esencial para los seguimientos que requieran evaluaciones de propiedades dependiendo de la exclusividad en características.
2	listing_url	Debido a su funcionamiento como referencia directa al anuncio en la plataforma, su papel en el análisis de datos puede basarse en la validación de estos, así como la muestra de resultados para información detallada.
3	last_scraped	Cuando nos encontramos realizando el análisis es importante tener en cuenta la fecha en que se han extraído los datos trabajados para garantizar que se encuentren actualizados y pueda ser detectada la obsolescencia en la información.
4	source	Esta variable nos será de utilidad para confirmar el origen de los datos detectados, de la misma forma contribuye a la comparación de calidad y consistencia de la información en fuentes diferentes.
5	name	Tomar en cuenta el nombre como un parámetro de análisis nos servirá para poder identificar palabras claves o tendencias que los anfitriones utilizan para clasificar sus propiedades y que puedan ser destacadas.
6	description	El análisis de texto es útil cuando se proporcionan los detalles del arrendamiento, pues nos permite identificar los puntos clave que son esenciales al momento de que el cliente considere una decisión, de esta manera se entiende como un anfitrión describe su propiedad para diferentes mercados.
7	neighbourhood_overview	Cuando se tiene información adicional sobre la ubicación de un sitio se facilita el uso de análisis cualitativos, en este caso se pueden realizar comparaciones en los vecindarios de acuerdo con las características destacables por el anfitrión mismas que ayudaran al cliente a tomar una decisión.
8	picture_url	Cuando se trata del cliente es de utilidad para poder comparar anuncios entre sí, si nos referimos al análisis de datos nos permitirá vincular

		gráficos y resultados visuales con imágenes del anuncio y de esta manera mejorar la presentación de los reportes.
9	neighbourhood_cleansed	En mi consideración, el análisis de esta variable nos facilitaría el agrupamiento de los datos para proporcionar análisis geoespaciales con una mayor precisión, pues la variable detalla una versión “limpia” del vecindario que contiene el anuncio.
10	latitud	Al representar la ubicación geográfica del anuncio, su análisis facilitará la representación de detalles geoespaciales en mapas interactivos.
11	longitude	Nos sirve como el complemento de la variable que proporciona la latitud del anuncio, es crucial para determinar ubicaciones exactas y visualizaciones cartográficas.
12	property_type	Cuando se proporcionan detalles específicos de la propiedad como lo es en este caso el “tipo” facilita la segmentación de anuncios de acuerdo con la preferencia del usuario y de la misma forma proporciona tendencias en el tipo de alojamiento preferido. Esto es de utilidad para la mejora en el mercado.
13	room_type	Debido a que esta variable describe el tipo de habitación con la que cuenta el alojamiento se pueden obtener claves para comprender que opciones prefieren los huéspedes mejorando de esta manera la oferta.
14	accommodates	Su fundamento, como en algunas variables antes mencionadas, se basa en analizar la oferta de acuerdo con las necesidades del huésped, cuando se proporciona la cantidad de huéspedes que pueden alojarse (en conjunto con demás especificaciones) presenta mayor eficiencia el análisis ante la capacidad de respuesta del mercado.
15	bathrooms_text	La relevancia de esta variable, en mi punto de vista se basa con la importancia que tienen determinadas áreas de un hospedaje debido a la necesidad que un cliente representa como punto de consideración ante una posible reserva.
16	bedrooms	El considerar el número de habitaciones disponibles es esencial para analizar la capacidad del mercado en términos de espacio de esta manera se aportan especificaciones para segmentar anuncios por tamaño.
17	beds	Cuando un cliente busca un lugar adecuado a su consideración, es de suma importancia contar dentro de un análisis con una métrica clave para determinar la capacidad real del alojamiento y la adecuación para grupos más grandes.
18	amenities	Contar con una lista de servicios y comodidades que ofrece el alojamiento es crucial dentro del análisis para identificar y entender como los factores de cada uno de ellos diferencian con cada anuncio existente y cuales son más valorados por el usuario.
Información sobre el anfitrión		
19	host_id	El identificador único del anfitrión nos permite vincular anuncios múltiples hacia un solo anfitrión, esto frente a operadores nos facilita el análisis del desempeño individual que evalúa la distribución de estos.
20	host_name	Como he mencionado con anterioridad, su análisis es útil para identificar patrones en anfitriones destacados y poder obtener análisis cualitativos que nos faciliten la presentación de los resultados dentro de un reporte.
21	host_since	Con esta variable podremos evaluar la experiencia del host y con ella su correlación ante métricas que permitan definir que tan destacado es. Nos es útil al momento de realizar segmentaciones entre ofertadores y analizar su tendencia de desempeño.
22	host_location	La ubicación detallada del anfitrión es importante para analizar si es local o solo gestiona sus propiedades de forma remota, nos ayuda a explorar posibles diferencias en la calidad del servicio ante el mercado.
23	host_about	Las descripciones personales, así como la información adicional de un arrendatario es importante para analizar cual es la presentación que

		tienen ante sus clientes misma que puede afectar la percepción o reservas identificando patrones o factores clave.
24	host_response_time	Considero el tiempo promedio en el que un anfitrión tarda en responder solicitudes como un factor clave a la experiencia del cliente, pues se sabe que tiempos de respuesta eficientes suelen representar una mayor satisfacción, se evalúa un impacto en reservas.
25	host_response_rate	El porcentaje de una solicitud de reserva nos sirve como análisis de la evaluación de disponibilidad y accesibilidad ante el cliente que proporcionan factores clave en la experiencia de cliente.
26	host_acceptance_rate	Una solicitud aceptada en la reserva es fundamental en el análisis de disposición del anfitrión y su impacto ante la ocupación de sus propiedades mismas que pueden indicar estrategias.
27	host_is_superhost	Esta métrica diferencia anfitriones destacados mismos que presentan experiencias superiores, este estatus permite el análisis ante el impacto de reservas, precios y calificaciones. Útil para su evaluación ante el mercado.
28	host_listings_count	El número de anuncios que se tienen activos en plataforma son útiles para identificar individualidades ante operadores grandes analizando su impacto e el mercado, explora la correlación entre cantidad y calificaciones o ingresos.
29	host_total_listings_count	El número total de propiedades, activas o no, de un host en plataforma es útil para el análisis de experiencia del mismo, así como su comportamiento histórico en el servicio que identifica la tendencia de crecimiento o declive en su actividad.
30	host_verifications	Cuando consideramos los métodos con los que se ha verificado una cuenta se obtiene información importante que evalúa la confiabilidad y el impacto de percepción ante un cliente. Esta métrica nos ayuda a garantizar que los estándares en la plataforma se cumplan.
31	host_identity_verified	Esta medida clave de seguridad influye de manera directa en la percepción del usuario, si nos enfocamos en su análisis podremos obtener información sobre el desempeño de un anfitrión frente a otro. Proporcionando mejores resultados.
Disponibilidad y reservas del anuncio		
32	has_availability	Ahora, si nos enfocamos en la disponibilidad activa de un anuncio obtendremos como resultado de análisis una métrica binaria que funciona como filtro principal ante propiedades en el mercado. Mide la oferta disponible en un momento específico que analiza patrones de disponibilidad en diferentes épocas del año.
33	availability_30	Con esta variable no es posible analizar la ocupación a corto plazo evaluando la demanda inmediata en ciertas localidades, identifica tendencias de planificación en los huéspedes.
34	availability_60	A diferencia de las variables anteriores, proporciona un panorama a mediano plazo identificando periodos de alta o baja demanda y es útil al momento de evaluar la flexibilidad y capacidad de respuesta del anfitrión ante los cambios en la demanda.
35	availability_90	Aquí el análisis es más amplio ante la planificación y la relación de estrategias en precios, temporadas altas o días feriados. Nos ayudará a comprender cómo se comporta la disponibilidad en trimestres.
36	availability_365	Esta variable evalúa la oferta a largo plazo e identifica la disponibilidad que un anfitrión presenta (constante o bloqueo significativo). Nos ayuda a analizar tendencias estacionales y prevé comportamientos en el mercado.
37	calendar_last_scraped	Esta variable garantiza que los datos sean reciente y confiables por lo tanto nos es útil para evaluar la frecuencia con la que los anfitriones actualizan calendarios, factor que afecta directamente la experiencia del

		huésped, así como la calidad del servicio. Entonces, es importante medir la capacidad del mercado de responder ante necesidades.
Reseñas y calificaciones del anuncio		
38	number_of_reviews	Veamos la variable como un indicador directo de la popularidad y confianza, entonces su análisis permitirá segmentar anuncios según su volumen en reseñas, identifica propiedades con mayor interacción y feedback con el huésped, esto nos ayuda a identificar patrones de éxito o fracaso.
39	number_of_reviews_ltm	Desde mi perspectiva es crucial para analizar tendencias recientes de anuncios que permitirá evaluar la evolución de popularidad en periodos recientes identificando los cambios estacionales y el efecto de las estrategias implementadas por los anfitriones.
40	reviews_per_month	El promedio de reseñas por mes se considera una métrica útil que ayuda a identificar la frecuencia de reservas de una propiedad y la satisfacción continua que oferta. Su estudio ayuda a entender el desempeño constante a lo largo del tiempo.
41	review_scores_rating	Esta métrica es clave para medir la calidad global del alojamiento, nos servirá para comparar el desempeño de diferentes propiedades y a identificar anuncios que sobresalen debido a su alta calificación, a su vez, aquellos que requieran mejoras.
42	review_scores_location	La calificación del anuncio ante la ubicación es importante para identificar si la ubicación es un factor decisivo en la satisfacción del huésped, puede ayudar a segmentar el mercado según preferencias geográficas.
43	review_scores_value	La considero importante para evaluar si los huéspedes consideran que reciben un buen valor ante el precio. Si nos enfocamos en su análisis podemos identificar propiedades que ofertan buena relación calidad-precio.
44	review_scores_cleanliness	Dentro de los factores que influyen en las decisiones del cliente, la limpieza es fundamental. Considero importante esta métrica para identificar propiedades que sobresalen y las que no ante esta necesidad. Identificando de esta manera aquellas áreas que requieren mayor atención.
45	review_scores_communication	La comunicación dentro de cualquier contexto es crucial ante la perspectiva positiva referente a la experiencia, en términos de claridad, disponibilidad y resolución representa decisiones estratégicas ante la retroalimentación directa a los clientes.
Precios, políticas y demás destacables		
46	price	El precio es un factor relevante para comparar la competitividad, permite analizar como se posiciona en términos de tarifas en el mercado. Su análisis nos ayudará a entender la relación entre el valor percibido y precios que han sido cobrados, detecta oportunidades de optimización.
47	minimum_nights	Es relevante para analizar las diferencias en las políticas de alojamiento y como estas afectan la accesibilidad y flexibilidad para huéspedes. Nos ayudará a encontrar tendencias de comportamiento en las clientes mismas que impactan en ocupación y demanda.
48	maximum_nights	Ayuda a identificar la flexibilidad en estancias más largas lo que es relevante en mercados con diferencias en los vaqueros. Nos permitirá ajustar la oferta en función de los diferentes segmentos de clientes.
49	calculated_host_listings_count_entire_homes	El numero de propiedades que un anfitrión ofrece en la plataforma ayuda a entender la concentración de profesionales ante particulares. Esto permite comparar la disponibilidad de vivienda.
50	license	Contar con la información legal es esencial para identificar el cumplimiento de las regulaciones y analizar si es cumplimiento presenta una ventaja competitiva o un obstáculo para el crecimiento.