



Entendimiento de los datos

Atenas, Grecia

Actividad 1

Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias de la Computación

Profesor
Alfredo García Suárez

Materia
Inteligencia de Negocios

Matricula
202256748

Alumno
Haciel Antonio Viveros Reyes

Periodo
Primavera 2025

Traducción y descripción

En esta sección traduciremos el nombre de las 75 variables presentes en nuestra base de datos, y le agregaremos una descripción de esa fila, todo basándose en el “Diccionario de Datos” proporcionado por la misma compañía Airbnb.

La tabla con las traducciones de las variables y la descripción de estas es la siguiente:

Variables	Variables y traducción	Descripción
id	id	Identificador único del anuncio en Airbnb.
listing_url	url_del_anuncio	URL del anuncio en Airbnb.
scrape_id	id_de_extracción	Identificador del "Scrape" en Airbnb (parte del proceso de extracción de datos).
last_scraped	última_extracción	Fecha y hora en formato UTC en la que se realizó la extracción del anuncio.
source	fuentes	Fuente de los datos, ya sea por "búsqueda por vecindario" o "búsqueda previa". Una búsqueda por vecindario significa que el anuncio fue encontrado buscando la ciudad; mientras que una búsqueda previa indica que el anuncio ya existía y se confirmó que seguía disponible.
name	nombre	Nombre del anuncio.
description	descripción	Descripción detallada del anuncio.
neighborhood_overview	resumen_del_vecindario	Descripción del vecindario proporcionada por el anfitrión.
picture_url	url_de_imagen	URL de la imagen del anuncio alojada en Airbnb.
host_id	id_del_anfitrión	Identificador único del anfitrión en Airbnb.
host_url	url_del_anfitrión	URL del perfil del anfitrión en Airbnb.
host_name	nombre_del_anfitrión	Nombre del anfitrión.
host_since	anfitrión_desde	Fecha en la que el anfitrión creó su cuenta. Para anfitriones que también son huéspedes, podría ser la fecha en que se registraron como huéspedes.
host_location	ubicación_del_anfitrión	Ubicación autoinformada por el anfitrión.

host_about	acerca_del_anfitrión	Descripción proporcionada por el anfitrión sobre sí mismo.
host_response_time	tiempo_de_respuesta_del_anfitrión	Tiempo promedio de respuesta del anfitrión a solicitudes de reserva.
host_response_rate	tasa_de_respuesta_de_l_anfitrión	Porcentaje de solicitudes de reserva a las que el anfitrión responde.
host_acceptance_rate	tasa_de_aceptación_del_anfitrión	Porcentaje de solicitudes de reserva aceptadas por el anfitrión.
host_is_superhost	anfitrión_es_superhost	Indica si el anfitrión es un "Superhost" (valor booleano Sí o No).
host_thumbnail_url	miniatura_del_anfitrión	URL de la imagen en miniatura del anfitrión.
host_picture_url	imagen_del_anfitrión	URL de la imagen de perfil del anfitrión.
host_neighbourhood	vecindario_del_anfitrión	Vecindario del anfitrión.
host_listings_count	número_de_anuncios_del_anfitrión	Número de anuncios que tiene el anfitrión.
host_total_listings_count	total_de_anuncios_del_anfitrión	Número total de anuncios del anfitrión.
host_verifications	verificaciones_del_anfitrión	Métodos de verificación del anfitrión.
host_has_profile_pic	anfitrión_tiene_foto_de_perfil	Indica si el anfitrión tiene foto de perfil (valor booleano Sí o No).
host_identity_verified	identidad_verificada_del_anfitrión	Indica si la identidad del anfitrión ha sido verificada (valor booleano Sí o No).
neighbourhood	vecindario	Vecindario, según geocodificación usando la latitud y longitud.
neighbourhood_cleansed	vecindario_normalizado	Vecindario limpiado o normalizado, según geocodificación.
neighbourhood_group_cleansed	grupo_de_vecindarios_normalizado	Grupo de vecindarios normalizado, según geocodificación.
latitude	latitud	Latitud del anuncio, usando el sistema geodésico mundial (WGS84).
longitude	longitud	Longitud del anuncio, usando el sistema geodésico mundial (WGS84).
property_type	tipo_de_propiedad	Tipo de propiedad seleccionada por el anfitrión.
room_type	tipo_de_habitación	Tipo de habitación (Casa/Departamento

		completo, Habitación privada, Habitación compartida, Hotel).
accommodates	capacidad	Número de huéspedes que puede alojar la propiedad.
bathrooms	baños	Número de baños disponibles en la propiedad.
bathrooms_text	texto_de_baños	Información textual sobre los baños
bedrooms	habitaciones	Número de habitaciones en la propiedad.
beds	camas	Número de camas disponibles en la propiedad.
amenities	amenidades	Lista de amenidades proporcionadas.
price	precio	Precio del anuncio, expresado en la moneda configurada.
minimum_nights	noches_mínimas	Número mínimo de noches requerido para reservar.
maximum_nights	noches_máximas	Número máximo de noches permitido para reservar.
minimum_minimum_nights	mínimo_de_noches_mínimas	Mínimo histórico de noches mínimas requeridas para este anuncio.
maximum_minimum_nights	máximo_de_noches_mínimas	Máximo histórico de noches mínimas requeridas para este anuncio.
minimum_maximum_nights	mínimo_de_noches_máximas	Mínimo histórico de noches máximas permitidas para este anuncio.
maximum_maximum_nights	máximo_de_noches_máximas	Máximo histórico de noches máximas permitidas para este anuncio.
minimum_nights_avg_ntm	promedio_noches_mínimas	Promedio de noches mínimas calculado por Airbnb.
maximum_nights_avg_ntm	promedio_noches_máximas	Promedio de noches máximas calculado por Airbnb.
calendar_updated	calendario_actualizado	Fecha de la última actualización del calendario de disponibilidad del anuncio.
has_availability	tiene_disponibilidad	Indica si el anuncio está actualmente disponible (valor booleano Sí o No).
availability_30	disponibilidad_30_días	Número de días disponible del anuncio durante los próximos 30 días
availability_60	disponibilidad_60_días	Número de días disponible del anuncio durante los próximos 60 días.
availability_90	disponibilidad_90_días	Número de días disponible del anuncio durante los próximos 90 días.

availability_365	disponibilidad_365_días	Número de días disponible del anuncio durante los próximos 365 días.
calendar_last_scraped	último_calendario_extraído	Fecha de la última extracción de datos del calendario de este anuncio.
number_of_reviews	número_de_reseñas	Número total de reseñas del anuncio.
number_of_reviews_ltm	número_reseñas_ultimos_12_meses	Número de reseñas recibidas en los últimos 12 meses.
number_of_reviews_l30d	número_reseñas_ultimos_30_días	Número de reseñas recibidas en los últimos 30 días.
first_review	primera_reseña	Fecha de la primera reseña recibida.
last_review	última_reseña	Fecha de la última reseña recibida.
review_scores_rating	puntaje_general_reseñas	Puntaje promedio general basado en las reseñas.
review_scores_accuracy	puntaje_precisión_reseñas	Puntaje promedio sobre la precisión de la descripción, basado en reseñas.
review_scores_cleanliness	puntaje_limpieza_reseñas	Puntaje promedio sobre la limpieza del lugar, basado en reseñas.
review_scores_checkin	puntaje_checkin_reseñas	Puntaje promedio sobre la experiencia de check-in, basado en reseñas.
review_scores_communication	puntaje_comunicación_reseñas	Puntaje promedio sobre la comunicación del anfitrión, basado en reseñas.
review_scores_location	puntaje_ubicación_reseñas	Puntaje promedio sobre la ubicación del anuncio, basado en reseñas.
review_scores_value	puntaje_valor_reseñas	Puntaje promedio sobre la relación calidad-precio, basado en reseñas.
license	licencia	Licencia o registro oficial del anuncio, si aplica.
instant_bookable	reserva_inmediata	Indica si el anuncio puede ser reservado inmediatamente (valor booleano Sí o No).
calculated_host_listings_count	total_calculado_anuncios_anfitrión	Número total de anuncios calculados para el anfitrión.
calculated_host_listings_count_entire_homes	total_anuncios_casas_completas	Número total de anuncios de casas completas calculados para el anfitrión.
calculated_host	total_anuncios_habitac	Número total de anuncios de

t_listings_count_private_rooms	iones_privadas	habitaciones privadas calculados para el anfitrión.
calculated_host_listings_count_shared_rooms	total_anuncios_habitaciones_compartidas	Número total de anuncios de habitaciones compartidas calculados para el anfitrión.
reviews_per_month	reseñas_por_mes	Promedio de reseñas recibidas por mes.

Selección de variables

A partir de la tabla que anteriormente generamos, se seleccionaran 50 variables que se consideraron el análisis de datos, y se realizará una justificación de por que se seleccionó dicha variable.

La tabla resultante junto con su justificación es la siguiente:

Variable	Justificación
id	Importante para identifica de manera única cada registro y realizar análisis específicos.
última_extracción	Importante porque permite conocer la vigencia de los datos y evaluar cambios en el tiempo.
fuelle	Importante para identifica el origen del registro, útil para segmentar análisis.
nombre	Importante para conocer el contexto de lo ofrecido.
descripción	Importante ya que ofrece información detallada que puede ser utilizada para clasificar anuncios.
resumen_del_vecindario	Importante para permitir la evaluación de la percepción del entorno, útil en análisis geoespaciales.
Id_del_anfitrión	Importante para identifica de manera única a cada anfitrión, cosa clave para realizar un análisis de desempeño.
nombre_del_anfitrión	Importante para conocer información básica del anfitrión, útil para segmentar datos.
anfitrión_desde	Importante para conocer y evaluar la

	experiencia del anfitrión.
ubicación_del_anfitrión	Útil para análisis geoespaciales y de alcance de servicios.
acerca_del_anfitrión	Proporciona un contexto sobre el anfitrión, lo cual resulta clave para evaluar descripciones cualitativas.
tiempo_de_respuesta_del_anfitrión	Muestra la calidad en la comunicación entre el cliente y el anfitrión.
tasa_de_respuesta_del_anfitrión	Métrica clave para medir la efectividad en la comunicación del anfitrión.
tasa_de_aceptación_del_anfitrión	Métrica clave para medir la disponibilidad real de las propiedades ofrecidas por el anfitrión.
anfitrión_es_superuser	Indicador de calidad y confianza del anfitrión.
número_de_anuncios_del_anfitrión	Permite identificar anfitriones con una mayor experiencia.
total_de_anuncios_del_anfitrión	Métrica para evaluar alcance y presencia del anfitrión dentro de la plataforma.
verificaciones_del_anfitrión	Evalúa la confiabilidad y legitimidad de la cuenta del anfitrión.
anfitrión_tiene_foto_de_perfil	Indicador de confianza y transparencia de la identidad del anfitrión.
identidad_verificada_del_anfitrión	Clave para evaluar la seguridad en las transacciones con el anfitrión.
vecindario	Ayuda a clasificar los anuncios geográficamente.
vecindario_normalizado	Mejora la calidad de los datos al estandarizar ubicaciones.
latitud	Variable esencial para análisis geoespaciales.
longitud	Complementa los análisis geoespaciales con precisión.
tipo_de_propiedad	Permite segmentar las publicaciones por la característica clave de la propiedad.
tipo_de_habitación	Identifica las opciones de hospedaje disponibles.
capacidad	Mide el tamaño de las propiedades,

	clave para realizar segmentación de los datos.
baños	Indica el nivel de comodidad centrado en la disponibilidad de baños.
texto_de_baños	Detalla características adicionales del baño.
habitaciones	Permite entender cómo se distribuye el espacio de la propiedad.
camas	Útil para complementar el análisis de capacidad y comodidad.
amenidades	Identifica características adicionales valoradas por los usuarios.
precio	Métrica clave para evaluar accesibilidad y segmentar mercados.
noches_mínimas	Indicador de restricciones y políticas del anfitrión en cuanto al tiempo mínimo para arrendar.
noches_máximas	Complementa la evaluación de políticas de hospedaje en cuanto a tiempo de arrendamiento.
disponibilidad_30_días	Mide la demanda a corto plazo. Clave para análisis de demanda.
disponibilidad_60_días	Mide la demanda a mediano plazo. Clave para análisis de demanda.
disponibilidad_90_días	Ayuda a identificar tendencias estacionales y demanda.
disponibilidad_365_días	Clave para análisis de largo plazo y demanda general.
número_de_reseñas	Indicador de popularidad y confiabilidad del anuncio.
numero_reseñas_ultimos_12_meses	Métrica para evaluar desempeño reciente en el último año.
numero_reseñas_ultimos_30_dias	Identifica cambios recientes en la actividad del anuncio.
primera_reseña	Proporciona contexto temporal sobre la antigüedad del anuncio.
última_reseña	Indica la actividad más reciente del anuncio.
puntaje_general_reseñas	Métrica global de calidad percibida.
puntaje_limpieza_reseñas	Clave para evaluar estándares de

	calidad en limpieza.
puntaje_comunicación_reseñas	Indicador de efectividad en la interacción con el anfitrión.
puntaje_ubicación_reseñas	Permite evaluar la conveniencia geográfica del anuncio.
puntaje_valor_reseñas	Mide la relación calidad-precio percibida por los usuarios.
reserva_inmediata	Identifica propiedades con mayor flexibilidad para usuarios espontáneos.
reseñas_por_mes	Mide la actividad promedio del anuncio, útil para evaluar popularidad.