

Actividad 1

Entendimiento de datos

Definición de variables relevantes
en Airbnb

Ciudad: Nueva York

País: Estados Unidos

Gutiérrez Pacheco Ángel de Jesús Elvis
BUAP

Diccionario traducido.

A continuación veremos el diccionario traducido que podemos encontrar en la página oficial de Airbnb.

Campo	Tipo	Calculado	Descripción
id	entero		Identificador único de Airbnb para el anuncio (alojamiento)
URL_del_anuncio	texto	S	
id_extraido	Entero grande	S	ID que contiene la fecha de búsqueda.
ultima_extraccion	fecha y hora	S	Fecha y hora UTC en que se "extrajo" este listado.
fuelle	texto		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.
nombre	texto		Nombre del listado (alojamiento)
descripcion	texto		Descripción detallada del alojamiento.
descripcion_general_del_vecindario	texto		Descripción del vecindario por parte del anfitrión
URL_de_la_imagen	texto		URL de la imagen de tamaño normal alojada en Airbnb para el anuncio
id_del_host	entero		Identificador único de Airbnb para el anfitrión/usuario
URL_del_host	texto	S	La página de Airbnb para el anfitrión

nombre_de_host	texto		Nombre del anfitrión. Generalmente, solo el nombre de pila.
anfitrión_desde	fecha		La fecha en la que se creó el anfitrión o usuario. En el caso de los anfitriones que son huéspedes de Airbnb, esta podría ser la fecha en la que se registraron como huéspedes.
ubicación_del_host	texto		La ubicación informada por el anfitrión
anfitrión_acerca_de	texto		Descripción sobre el anfitrión
tiempo_de_respuesta_del_host			
tasa_de_respuesta_del_host			
tasa_de_aceptación_del_host			Porcentaje en que un anfitrión acepta solicitudes de reserva.
el_host_es_superhost	booleano [t=verdadero; f=falso]		Anfitrión que se distingue por brindar un servicio de calidad y una excelente hospitalidad.
URL_de_la_miniatra_del_host	texto		
URL_de_la_imagen_del_host	texto		
vecindario_anfitrión	texto		
recuento_de_anfitriones_en_listado	texto		La cantidad de anuncios que tiene el anfitrión (según cálculos desconocidos de Airbnb)
recuento_total_de_anfitriones	texto		La cantidad de anuncios que tiene el anfitrión (según cálculos desconocidos de Airbnb)
Verificaciones_del_host			
Host_tiene_una_foto_de_perfil	booleano [t=verdadero; f=falso]		

Identidad_del_host_verificada	booleano [t=verdadero; f=falso]		
vecindario	texto		
barrio_limpio	texto	S	El vecindario geocodificado utilizando la latitud y longitud contra vecindarios definidos por shapefiles digitales abiertos o públicos.
grupo_de_vecindario_limpio	texto	S	El grupo de vecindarios se geocodifica utilizando la latitud y la longitud contra los vecindarios definidos por archivos de formas digitales abiertos o públicos.
latitud	numérico		Utiliza la proyección del Sistema Geodésico Mundial (WGS84) para latitud y longitud.
longitud	numérico		Utiliza la proyección del Sistema Geodésico Mundial (WGS84) para latitud y longitud.
tipo_propiedad	texto		Tipo de propiedad seleccionada por el propietario. Los hoteles y los Bed and Breakfasts se describen como tales por sus anfitriones en este campo
tipo_de_habitación	texto		<p>Todas las casas están agrupadas en los siguientes tres tipos de habitaciones:</p> <p>Alojamiento completo</p> <p>Habitación privada</p> <p>Habitación compartida</p> <p>Los alojamientos completos son la mejor opción si buscas un hogar lejos de casa. Con un alojamiento completo, tendrás todo el espacio para ti. Esto</p>

			<p>suele incluir un dormitorio, un baño, una cocina y una entrada independiente. 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 en el anuncio.</p> <p>Las habitaciones privadas son ideales para cuando prefieres un poco de privacidad y aun así valoras la conexión local. Cuando reservas una habitación privada, tendrás tu propia habitación privada para dormir y podrás compartir algunos espacios con otras personas. Es posible que tengas que atravesar espacios interiores que otro anfitrión o huésped pueda ocupar para llegar a tu habitación.</p> <p>Las habitaciones compartidas son ideales 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 estadías económicas.</p>
capacidad	entero		La capacidad máxima del lugar
baños	numérico		El número de baños en el lugar
baños_texto	cadena		Número de baños en el anuncio. En el sitio web de Airbnb, el campo de baños ha

			evolucionado de un número a una descripción textual. En los anuncios más antiguos, se utiliza la palabra baños.
dormitorios	entero		El número de habitaciones
camas	entero		El número de cama(s)
comodidades	json		
precio	divisa		Precio diario en moneda local. NOTA: el signo \$ es un artefacto técnico de la exportación, ignórelo.
minimo_noches	entero		Número mínimo de noches de estancia para el alojamiento
maximo_noches	entero		Número máximo de noches de estancia para el alojamiento
minimo_noches_minimas	entero	S	El valor mínimo de minimo_noches del calendario (para 365 noches en el futuro)
maximo_minimo_noches	entero	S	El valor máximo de minimo_noches del calendario (para 365 noches en el futuro)
minimo_maximo_noches	entero	S	El valor maximo_noches más pequeño del calendario (para 365 noches en el futuro)
maximo_maximo_noches	entero	S	El valor maximo_noches más grande del calendario (para 365 noches en el futuro)
noches_minimas_promedio	numérico	S	El valor promedio de minimo_noches del calendario (para 365 noches en el futuro)
promedio_máximo_de_noches	numérico	S	El valor promedio de maximo_noches del calendario (para 365 noches en el futuro)
calendario_actualizado	fecha		

tiene_disponibilidad	booleano		[t=verdadero; f=falso]
disponibilidad_30	entero	S	La disponibilidad del alojamiento 30 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.
disponibilidad_60	entero	S	La disponibilidad del alojamiento 60 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.
disponibilidad_90	entero	S	La disponibilidad del alojamiento 90 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.
disponibilidad_365	entero	S	La disponibilidad del alojamiento 365 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.
calendario_ultima_busqueda	fecha		
numero_de_reseñas	entero		El número de reseñas que tiene el anuncio.
numero_de_reseñas_ulm	entero	S	La cantidad de reseñas que tiene el anuncio (en los últimos 12 meses)

numero_de_reseñas_u30d	entero	S	El número de reseñas que tiene el anuncio (en los últimos 30 días)
primera_reseña	fecha	S	La fecha de la primera/más antigua revisión
ultima_reseña	fecha	S	La fecha de la última/más reciente revisión
calificación_de_reseñas			
precision_de_reseñas			
reseñas_puntuaciones_limpias			
revision_puntuaciones_verificacion			
revision_puntuaciones_comunicacion			
ubicacion_de_reseñas			
valor_de_reseñas			
licencia	texto		El número de licencia/permiso/registro
reservable_al_instante	booleano [t=true; f=false].		Si el huésped puede reservar automáticamente el alojamiento sin que el anfitrión tenga que aceptar su solicitud de reserva. Indicador de un alojamiento comercial.
recuento_de_anuncios_del_host_calculados	entero	S	La cantidad de listados que tiene el anfitrión en la búsqueda actual, en la geografía de la ciudad/región.
recuento_de_anuncios_del_host_calculados_con_alojamientos_completos	entero	S	La cantidad de anuncios completos de casas/apartamentos que el anfitrión tiene en el lugar actual, en la geografía de la ciudad/región
recuento_de_anuncios_del_host_calculados_con_habitaciones_privadas	entero	S	La cantidad de listados de habitaciones privadas que el

			anfitrión tiene en la búsqueda actual, en la geografía de la ciudad/región
recuento_de_anuncios_del_host_calculados_con_habitaciones_compartidas	entero	S	La cantidad de listados de habitaciones compartidas que el anfitrión tiene en la búsqueda actual, en la geografía de la ciudad/región
reseñas_por_mes	entero	S	El número promedio de reseñas por mes que tiene el anuncio durante su vigencia.

Variables importantes para el análisis.

Ahora, veremos las 50 variables más importantes para el análisis que pide el cliente, junto con su justificación.

Número	Campo	Tipo	Justificación
1	id	entero	Lo usaremos para identificar rápidamente los alojamientos registrados y para hacer conteos por host.
2	fuelle	texto	Puede ser útil para saber si un alojamiento aparece en búsquedas, pero no es solicitado en 2 meses.
3	nombre	texto	Con él, identificamos el nombre del alojamiento, puede ser útil para encontrar patrones en el uso del lenguaje que hagan más atractivo un lugar.
4	descripcion	texto	Con esto podemos saber cómo describen el lugar. Útil para encontrar patrones en el uso del lenguaje que hagan más atractivo un lugar.
5	descripcion_general_del_vecindario	texto	Con esto podemos saber cómo describen la zona. Útil para encontrar patrones en el uso del lenguaje que hagan más atractivo un lugar.
6	id_del_host	entero	Usado para identificar un host. Lo podemos usar para realizar conteos junto con el id del alojamiento.
7	nombre_de_host	texto	Podemos usarlo para identificar al host de una forma más natural.
8	anfitrión_desde	fecha	Para saber la antigüedad del host siendo anfitrión del lugar. Podemos usarlo para saber la experiencia del host.
9	anfitrión_acerca_de	texto	Podemos usarlo para encontrar patrones en el uso del lenguaje que hagan más atractivo un lugar.
10	tiempo_de_respuesta_del_host		Útil para saber si el tiempo de respuesta influye en las reseñas o experiencia.

11	tasa_de_respuesta_del_host		Útil para saber si la tasa de respuesta influye en las reseñas o experiencia.
12	tasa_de_aceptación_del_host		Útil para saber si la tasa de aceptación influye en las reseñas o experiencia.
13	el_host_es_superhost	booleano [t=verdadero; f=falso]	Podemos usarlo para saber cómo influye en la calidad, precio y solicitudes.
14	vecindario_anfitrión	texto	Útil para saber si hay vecindarios más visitados que otros.
15	recuento_total_de_anfitriones	texto	Podemos usarlo para verificar el éxito del host o la rentabilidad que les da la plataforma.
16	verificaciones_del_host		Podemos usarlo para saber si los hosts se verifican adecuadamente y si el método influye en algo.
17	identidad_del_host_verificada	booleano [t=verdadero; f=falso]	Podemos usarlo para saber si los hosts se verifican adecuadamente y si esto influye en algo.
18	latitud	numérico	Podemos usarlo para conocer la ubicación exacta del lugar.
19	longitud	numérico	Podemos usarlo para conocer la ubicación exacta del lugar.
20	tipo_propiedad	texto	Útil para clasificar las propiedades, podríamos identificar las más comunes y exitosas.
21	tipo_de_habitación	texto	Podemos usarlo para encontrar los tipos de habitaciones más solicitadas o mejor valoradas.
22	capacidad	entero	Útil para saber el tamaño de grupos más comunes.
23	baños	numérico	Podemos usarlo para encontrar si esto influye en las reseñas y si se relaciona con la capacidad.

24	dormitorios	entero	Podemos usarlo para encontrar si esto influye en las reseñas y si se relaciona con la capacidad.
25	camas	entero	Podemos usarlo para encontrar si esto influye en las reseñas y si se relaciona con la capacidad.
26	comodidades	json	Podemos usarlo para encontrar si esto influye en las reseñas y en qué medida.
27	precio	divisa	Podemos usar esto para medir la rentabilidad del anuncio y cómo influyen el resto de las variables aquí.
28	minimo_noches	entero	Útil para saber si hay correlación con la cantidad de solicitudes.
29	maximo_noches	entero	Útil para saber si hay correlación con la cantidad de solicitudes.
30	noches_minimas_promedio	numérico	Útil para saber si hay correlación con la cantidad de solicitudes.
31	promedio_máximo_de_noches	numérico	Útil para saber si hay correlación con la cantidad de solicitudes.
32	tiene_disponibilidad	booleano	Útil para conocer qué establecimientos dejaron de ser solicitados (generalmente son lugares que ya no ofrecen el servicio)
33	disponibilidad_30	entero	Podemos usarlo para identificar qué alojamientos están más tiempo ocupados y el flujo de personas que tiene.
34	disponibilidad_60	entero	Podemos usarlo para identificar qué alojamientos están más tiempo ocupados y el flujo de personas que tiene.
35	disponibilidad_90	entero	Podemos usarlo para identificar qué alojamientos están más tiempo ocupados y el flujo de personas que tiene.
36	disponibilidad_365	entero	Podemos usarlo para identificar qué alojamientos están más tiempo

			ocupados y el flujo de personas que tiene.
37	numero_de_reseñas	entero	Útil para saber la popularidad del alojamiento, junto con las demás variables, puede ser útil para saber qué es lo que llama más la atención.
38	numero_de_reseñas_ulm	entero	Útil para saber la popularidad del alojamiento, junto con las demás variables, puede ser útil para saber qué es lo que llama más la atención.
39	numero_de_reseñas_u30d	entero	Útil para saber la popularidad del alojamiento, junto con las demás variables, puede ser útil para saber qué es lo que llama más la atención.
40	primera_reseña	fecha	Podemos usarlo saber qué tanto tiempo lleva activo el alojamiento.
41	ultima_reseña	fecha	Podemos usarlo para identificar si algún alojamiento no es popular junto con las demás variables de reseñas.
42	calificación_de_reseñas		Útil para conocer la popularidad y calidad general del alojamiento.
43	precision_de_reseñas		Útil para conocer la popularidad y calidad general del alojamiento.
44	reseñas_puntuaciones_limpias		Útil para conocer la popularidad y calidad sobre la limpieza del alojamiento.
45	revision_puntuaciones_ingreso		Útil para conocer la popularidad y calidad sobre el acceso del alojamiento.
46	revision_puntuaciones_comunicacion		Útil para conocer la popularidad y calidad sobre la comunicación con el host.
47	ubicacion_de_reseñas		Útil para conocer la popularidad y calidad sobre la ubicación del alojamiento.
48	valor_de_reseñas		Útil para conocer la popularidad y calidad general del alojamiento.

49	reservable_al_instante	booleano [t=true; f=false].	Podemos usarlo para saber si el alojamiento es manejado por una empresa o no.
50	reseñas_por_mes	entero	Podemos usarlo para conocer la popularidad del alojamiento.