

# Actividad 1 (Entendimiento de los Datos)



Inteligencia de Negocios

Benemérita Universidad Autónoma de Puebla

Ingeniería en Tecnologías de la Información

Alfredo García Suárez

Primavera 2025

202133193

Diego Bustamante Domínguez

## Diccionario de variables

### **Id**

Identificador único de Airbnb para el anuncio.

### **isting\_url**

URL del listado.

### **scrape\_id**

Dentro de Airbnb, "Scrape" esto fue parte de.

### **last\_scraped**

Fecha y hora en UTC en que se "extrajo" este listado.

### **Source**

Fuente. Puede ser uno de los siguientes: "neighbourhood search" o "previous scrape". "Neighbourhood search" significa que el listado fue encontrado buscando en la ciudad, mientras que "previous scrape" significa que el listado fue visto en un scrape realizado en los últimos 65 días y se confirmó que aún está disponible en el sitio de Airbnb.

### **Name**

Nombre del listado.

### **Description**

Descripción detallada del listado.

### **neighborhood\_overview**

Descripción del vecindario por parte del anfitrión.

### **picture\_url**

URL de la imagen alojada en Airbnb del listado.

### **host\_id**

Identificador único de Airbnb para el anfitrión/usuario.

### **host\_url**

Página de Airbnb del anfitrión.

### **host\_name**

Nombre del anfitrión. Generalmente solo el primer nombre.

**host\_since**

Fecha en que el anfitrión/usuario fue creado. Para los anfitriones que son huéspedes de Airbnb, podría ser la fecha en que se registraron como huéspedes.

**host\_location**

Ubicación reportada por el anfitrión.

**host\_about**

Descripción sobre el anfitrión.

**host\_response\_time**

Tiempo de respuesta del anfitrión.

**host\_response\_rate**

Tasa de respuesta del anfitrión.

**host\_acceptance\_rate**

Tasa de aceptación de solicitudes de reserva por parte del anfitrión.

**host\_is\_superhost**

Indica si el anfitrión es un "Superhost".

**host\_thumbnail\_url**

URL de la miniatura del anfitrión.

**host\_picture\_url**

URL de la foto del anfitrión.

**host\_neighbourhood**

Vecindario del anfitrión.

**host\_listings\_count**

Número de anuncios que tiene el anfitrión (según cálculos desconocidos de Airbnb).

**host\_total\_listings\_count**

Número total de anuncios que tiene el anfitrión (según cálculos desconocidos de Airbnb).

**host\_verifications**

Verificaciones del anfitrión.

**host\_has\_profile\_pic**

Indica si el anfitrión tiene foto de perfil.

**host\_identity\_verified**

Indica si la identidad del anfitrión está verificada.

**neighbourhood**

Vecindario del listado.

**neighbourhood\_cleansed**

Vecindario geocodificado utilizando latitud y longitud, de acuerdo con las definiciones de vecindarios basadas en archivos de forma digitales abiertos o públicos.

**neighbourhood\_group\_cleansed**

Grupo de vecindarios geocodificado utilizando latitud y longitud, de acuerdo con las definiciones de vecindarios basadas en archivos de forma digitales abiertos o públicos.

**Latitude**

Latitud según el sistema geodésico mundial (WGS84).

**Longitude**

Longitud según el sistema geodésico mundial (WGS84).

**property\_type**

Tipo de propiedad seleccionado por el anfitrión. Los hoteles y los Bed and Breakfast son descritos como tales por sus anfitriones en este campo.

**room\_type**

Todos los hogares están agrupados en los siguientes tres tipos de habitación:

- **Entire place:** Propiedades enteras para tu uso exclusivo.
- **Private room:** Habitación privada para dormir, con espacios compartidos.
- **Shared room:** Habitación compartida con otros.

**accommodates**

Capacidad máxima del listado.

**Bathrooms**

Número de baños en el listado.

**bathrooms\_text**

En el sitio web de Airbnb, el campo de los baños ha evolucionado de un número a una descripción textual. Para scrapes antiguos, se utiliza el número.

**Bedrooms**

Número de dormitorios.

**Beds**

Número de camas.

**Amenities**

Servicios o comodidades del listado.

**Price**

Precio diario en la moneda local. Nota: el signo "\$" es un artefacto técnico de la exportación, por favor ignóralo.

**minimum\_nights**

Número mínimo de noches para la estadía (las reglas del calendario pueden ser diferentes).

**maximum\_nights**

Número máximo de noches para la estadía (las reglas del calendario pueden ser diferentes).

**minimum\_minimum\_nights**

El valor mínimo de noches en el calendario (mirando 365 noches en el futuro).

**maximum\_minimum\_nights**

El valor máximo de noches mínimas en el calendario (mirando 365 noches en el futuro).

**minimum\_maximum\_nights**

El valor mínimo de noches máximas en el calendario (mirando 365 noches en el futuro).

**maximum\_maximum\_nights**

El valor máximo de noches máximas en el calendario (mirando 365 noches en el futuro).

**minimum\_nights\_avg\_ntm**

Promedio de noches mínimas en el calendario (mirando 365 noches en el futuro).

**maximum\_nights\_avg\_ntm**

Promedio de noches máximas en el calendario (mirando 365 noches en el futuro).

**calendar\_updated**

Fecha de la última actualización del calendario.

**has\_availability**

Indica si hay disponibilidad.

**availability\_30**

Disponibilidad del listado 30 días en el futuro, determinada por el calendario. Un listado puede no estar disponible si ha sido reservado por un huésped o bloqueado por el anfitrión.

**availability\_60**

Disponibilidad del listado 60 días en el futuro, determinada por el calendario.

**availability\_90**

Disponibilidad del listado 90 días en el futuro, determinada por el calendario.

**availability\_365**

Disponibilidad del listado 365 días en el futuro, determinada por el calendario.

**calendar\_last\_scraped**

Fecha de la última extracción del calendario.

**number\_of\_reviews**

Número total de reseñas del listado.

**number\_of\_reviews\_ltm**

Número de reseñas del listado en los últimos 12 meses.

**number\_of\_reviews\_l30d**

Número de reseñas del listado en los últimos 30 días.

**first\_review**

Fecha de la primera reseña.

**last\_review**

Fecha de la última reseña.

**review\_scores\_rating**

Puntuación general de las reseñas.

**review\_scores\_accuracy**

Puntuación de precisión de las reseñas.

**review\_scores\_cleanliness**

Puntuación de limpieza de las reseñas.

**review\_scores\_checkin**

Puntuación del check-in en las reseñas.

**review\_scores\_communication**

Puntuación de la comunicación en las reseñas.

**review\_scores\_location**

Puntuación de la ubicación en las reseñas.

**review\_scores\_value**

Puntuación de valor en las reseñas.

**license**

Número de licencia/permiso/registro.

**instant\_bookable**

Indica si el listado se puede reservar de manera instantánea sin la necesidad de que el anfitrión acepte la solicitud. Un indicador de un listado comercial.

**calculated\_host\_listings\_count**

Número de anuncios que tiene el anfitrión en el scrape actual, en la ciudad/región.

**calculated\_host\_listings\_count\_entire\_homes**

Número de anuncios de casas completas que tiene el anfitrión en el scrape actual, en la ciudad/región.

**calculated\_host\_listings\_count\_private\_rooms**

Número de anuncios de habitaciones privadas que tiene el anfitrión en el scrape actual, en la ciudad/región.

**calculated\_host\_listings\_count\_shared\_rooms**

Número de anuncios de habitaciones compartidas que tiene el anfitrión en el scrape actual, en la ciudad/región.

**reviews\_per\_month**

Número promedio de reseñas por mes que recibe el listado durante su vida útil.

## Variables más importantes

1. **Id**  
Identificador único de cada anuncio. Es fundamental para la trazabilidad de los datos y el análisis de patrones.
2. **listing\_url**  
URL del listado. Útil para acceder a la página directamente para obtener más detalles si es necesario.
3. **Name**  
Nombre del listado. Ayuda a identificar los anuncios y analizar la influencia del título en las reservas.
4. **Description**  
Descripción detallada del listado. Puede ser analizada para detectar patrones en el lenguaje, identificando características que atraen a los huéspedes.
5. **property\_type**  
Tipo de propiedad (apartamento, casa, etc.), relevante para clasificar y segmentar los datos según el tipo de alojamiento.
6. **room\_type**  
Tipo de habitación (entera, privada o compartida). Es importante para segmentar según las preferencias de los huéspedes.
7. **accommodates**  
Capacidad máxima del listado. Indicador clave para segmentar por tamaño de grupo y tipo de viajero.
8. **Bathrooms**  
Número de baños. Variables como esta permiten análisis de correlación con otros factores como precio y demanda.
9. **Bedrooms**  
Número de dormitorios. Relacionada con la capacidad y el tipo de viajero.
10. **Beds**  
Número de camas disponibles. Relacionado con la capacidad de la propiedad y la demanda de espacios grandes.
11. **Amenities**  
Servicios disponibles (Wi-Fi, piscina, aire acondicionado, etc.). Puede



analizarse para encontrar qué comodidades aumentan la probabilidad de reservas.

**12. Price**

El precio es clave para cualquier análisis de demanda, elasticidad del precio y segmentación de mercado.

**13. minimum\_nights**

Número mínimo de noches para reservar. Importante para calcular patrones de ocupación y segmentar el tipo de reservas.

**14. maximum\_nights**

Número máximo de noches. Permite detectar tendencias en estancias largas y optimizar precios.

**15. availability\_30**

La disponibilidad a 30 días. Relacionada con la oferta y demanda a corto plazo.

**16. availability\_60**

Disponibilidad a 60 días. Información útil para análisis de patrones de reserva a mediano plazo.

**17. availability\_365**

Disponibilidad a largo plazo. Relevante para análisis de tendencias y proyección de ocupación.

**18. instant\_bookable**

Indica si el listado puede ser reservado automáticamente, lo cual impacta la rapidez de las reservas.

**19. calendar\_updated**

Fecha de última actualización del calendario. Es útil para asegurar que los datos estén actualizados para el análisis.

**20. host\_id**

Identificador único del anfitrión. Es necesario para hacer análisis segmentados por anfitrión y entender su comportamiento.

**21. host\_name**

Nombre del anfitrión, importante para identificar patrones relacionados con la oferta de anfitriones específicos.

**22. host\_since**

Fecha de registro del anfitrión. Permite analizar la antigüedad del anfitrión y correlacionarlo con el desempeño.

**23. host\_location**

Ubicación del anfitrión. Puede ser usada para análisis geoespacial y encontrar patrones según la región.

**24. host\_response\_time**

Tiempo de respuesta del anfitrión. Se correlaciona con la satisfacción del cliente y la tasa de conversión.

**25. host\_response\_rate**

Tasa de respuesta del anfitrión. Relacionado con la calidad de la interacción del anfitrión con los huéspedes.

**26. host\_acceptance\_rate**

Tasa de aceptación de solicitudes de reserva. Información clave para entender la flexibilidad del anfitrión.

**27. host\_is\_superhost**

Indica si el anfitrión es un "Superhost". Esta es una variable clave para analizar la calidad del servicio y su relación con la satisfacción del huésped.

**28. host\_listings\_count**

El número de anuncios de un anfitrión. Útil para análisis sobre la profesionalización del anfitrión.

**29. host\_verifications**

Verificaciones de seguridad del anfitrión. Es crucial para evaluar la confiabilidad del anfitrión.

**30. number\_of\_reviews**

Número total de reseñas. Ayuda a entender la popularidad y el nivel de interacción de un listado.

**31. review\_scores\_rating**

Puntuación general del listado. Fundamental para análisis de calidad y preferencia de los huéspedes.

**32. review\_scores\_accuracy**

Puntuación de precisión de la descripción. Es importante para medir la calidad y coherencia de la información.

**33. review\_scores\_cleanliness**

Puntuación de limpieza. Relacionada con la satisfacción del cliente y la calidad del servicio.

**34. review\_scores\_checkin**

Puntuación del proceso de check-in. Ayuda a identificar problemas en la llegada de los huéspedes.

**35. review\_scores\_communication**

Puntuación de comunicación con el anfitrión. Correlacionada con la satisfacción y calidad del servicio.

**36. review\_scores\_location**

Puntuación de la ubicación. Crucial para los huéspedes que buscan ciertas ubicaciones o barrios específicos.

**37. review\_scores\_value**

Puntuación sobre la relación calidad-precio. Es importante para entender la percepción de valor entre los huéspedes.

**38. first\_review**

Fecha de la primera reseña. Puede ser útil para analizar la antigüedad de un listado.

**39. last\_review**

Fecha de la última reseña. Indica la actividad reciente y puede ser útil para detectar patrones de popularidad.

**40. reviews\_per\_month**

Promedio de reseñas por mes. Es útil para entender la frecuencia de las estancias y la demanda.

**Variables relacionadas con disponibilidad y ocupación:**

**41. neighbourhood\_cleansed**

Vecindario geocodificado. Crucial para análisis espaciales y correlación con las preferencias de los huéspedes.

**42. Latitude**

Latitud geográfica de la propiedad. Importante para análisis de localización y mapas de calor.

**43. Longitude**

Longitud geográfica de la propiedad. También útil para análisis geoespacial y patrones de ubicación.

**44. number\_of\_reviews\_ltm**

Número de reseñas en los últimos 12 meses. Ayuda a analizar la popularidad reciente de los anuncios.

**45. number\_of\_reviews\_l30d**

Número de reseñas en los últimos 30 días. Útil para detectar tendencias y estacionalidad a corto plazo.

**46. neighbourhood\_group\_cleansed**

Grupo de vecindario geocodificado. Permite hacer análisis agrupados por zonas.

**47. minimum\_nights\_avg\_ntm**

Promedio de noches mínimas por reserva. Permite detectar patrones de comportamiento de los huéspedes.

**48. maximum\_nights\_avg\_ntm**

Promedio de noches máximas por reserva. Ayuda a comprender la flexibilidad de los huéspedes al reservar.

**49. calendar\_updated**

Fecha de la última actualización del calendario. Importante para garantizar que los datos estén siempre al día.

**50. availability\_365**

Disponibilidad a 365 días. Relevante para entender la oferta anual y los patrones de ocupación a largo plazo.