



**Benemérita Universidad Autónoma de Puebla**

**Facultad de Ciencias de la Computación**

**Ingeniería en Tecnologías de la Información**

**Inteligencia de Negocios**

**Reporte 1: Entendimiento de los datos**

**Alumna: Diana Gabriela Victorino Osorio**

**202165678**

**Primavera 2025**

## Entendimiento de los datos

Para poder trabajar con los datos de la ciudad que yo escogí, Zurich, necesitamos entender la información que la página de Airbnb nos proporciona, es por eso que la primera actividad que llevamos a cabo es la traducción de cada uno de los campos y una pequeña descripción que nos permita entender y descubrir la información que nos será útil.

| No. | Nombre en Inglés      | Traducción al Español       | Descripción  |
|-----|-----------------------|-----------------------------|--|
| 1   | id                    | Identificador               | Identificador único de Airbnb para el anuncio.                         |
| 2   | listing_url           | URL del Anuncio             | URL única de Airbnb para el anuncio..                                  |
| 3   | scrape_id             | ID del Scrape               | 'Scrape' de Inside Airbnb del que formaba parte.                       |
| 4   | last_scraped          | Última Extracción           | UTC. La fecha y hora en que se 'extrajo' este anuncio.                 |
| 5   | source                | Fuente                      | La fuente del anuncio: 'búsqueda de vecindario' o 'extracción previa'. |
| 6   | name                  | Nombre                      | Nombre del anuncio.  |
| 7   | description           | Descripción                 | Descripción detallada del anuncio.                                     |
| 8   | neighborhood_overview | Descripción del Vecindario  | Descripción del vecindario por el anfitrión.                           |
| 9   | picture_url           | URL de Imagen               | URL de la imagen alojada en Airbnb para el anuncio.                    |
| 10  | host_id               | ID del Anfitrión            | URL de la imagen alojada en Airbnb para el anuncio.                    |
| 11  | host_url              | URL del Anfitrión           | La página de Airbnb para el anfitrión.                                 |
| 12  | host_name             | Nombre del Anfitrión        | Nombre del host  |
| 13  | host_since            | Anfitrión Desde             | La fecha en que se creó el anfitrión.                                  |
| 14  | host_location         | Ubicación del Anfitrión     | La ubicación reportada del anfitrión.                                  |
| 15  | host_about            | Acerca del Anfitrión        | Descripción sobre el anfitrión.  |
| 16  | host_response_time    | Tiempo de Respuesta         | Tiempo de respuesta típico del anfitrión.                              |
| 17  | host_response_rate    | Tasa de Respuesta           | Tasa de respuesta del anfitrión a consultas.                           |
| 18  | host_acceptance_rate  | Tasa de Aceptación          | Tasa de aceptación del anfitrión para reservas.                        |
| 19  | host_is_superhost     | ¿Es Superhost?              | Si el anfitrión es un superhost (verdadero/falso).                     |
| 20  | host_thumbnail_url    | URL Miniatura del Anfitrión | URL de la imagen miniatura del anfitrión.                              |

|    |                              |                              |  |
|----|------------------------------|------------------------------|--|
| 21 | host_picture_url             | URL Foto del Anfitrión       | URL de la foto de perfil del anfitrión.                                      |
| 22 | host_neighbourhood           | Vecindario del Anfitrión     | Vecindario reportado del anfitrión.  |
| 23 | host_listings_count          | Número de Anuncios           | El número de anuncios que tiene el anfitrión.                                |
| 24 | host_total_listings_count    | Total de Anuncios            | Número total de anuncios del anfitrión..                                     |
| 25 | host_verifications           | Verificaciones del Anfitrión | Métodos de verificación para el anfitrión.                                   |
| 26 | host_has_profile_pic         | ¿Tiene Foto de Perfil?       | Si el anfitrión tiene foto de perfil.  |
| 27 | host_identity_verified       | Identidad Verificada         | Si la identidad del anfitrión está verificada.                               |
| 28 | neighbourhood                | Vecindario                   | Vecindario según se informa.   |
| 29 | neighbourhood_cleanse        | Vecindario Limpio            | Vecindario limpio  |
| 30 | neighbourhood_group_cleansed | Grupo de Vecindarios         | Grupo de vecindarios geocodificado.  |
| 31 | latitude                     | Latitud                      | Latitud del anuncio.   |
| 32 | longitude                    | Longitud                     | Longitud del anuncio   |
| 33 | property_type                | Tipo de Propiedad            | Tipo de propiedad seleccionada por el anfitrión.                             |
| 34 | room_type                    | Tipo de Habitación           | Tipo de habitación (Casa entera, Habitación privada, Habitación compartida). |
| 35 | accommodates                 | Capacidad                    | La capacidad máxima del anuncio.   |
| 36 | bathrooms                    | Baños                        | El número de baños en el anuncio.  |
| 37 | bathrooms_text               | Texto de Baños               | Descripción textual de los baños.  |
| 38 | bedrooms                     | Habitaciones                 | El número de habitaciones en el anuncio.                                     |
| 39 | beds                         | Camas                        | El número de camas en el anuncio.  |
| 40 | amenities                    | Servicios                    | Servicios disponibles en el anuncio.   |
| 41 | price                        | Precio                       | Precio diario en moneda local.   |
| 42 | minimum_nights               | Noches Mínimas               | Noches mínimas requeridas.   |
| 43 | maximum_nights               | Noches Máximas               | Noches máximas permitidas.   |
| 44 | availability_30              | Disponibilidad (30 días)     | Disponibilidad para los próximos 30 días.                                    |
| 45 | availability_60              | Disponibilidad (60 días)     | Disponibilidad para los próximos 60 días.                                    |
| 46 | availability_90              | Disponibilidad (90 días)     | Disponibilidad para los próximos 90 días.                                    |

|    |                      |                           |   |
|----|----------------------|---------------------------|---|
| 47 | availability_365     | Disponibilidad (365 días) | Disponibilidad para el próximo año.                     |
| 48 | calendar_updated     | Calendario Actualizado    | Fecha en que se actualizó el calendario por última vez. |
| 49 | number_of_reviews    | Número de Reseñas         | Número total de reseñas.                                |
| 50 | first_review         | Primera Reseña            | Fecha de la primera reseña.                             |
| 51 | last_review          | Última Reseña             | Fecha de la última reseña.                              |
| 52 | review_scores_rating | Puntuación General        | Puntuación general de la reseña.                        |
| 53 | license              | Licencia                  | Número de registro o permiso.                           |
| 54 | instant_bookable     | Reserva Instantánea       | Si el anuncio es reservable instantáneamente.           |
| 55 | reviews_per_month    | Reseñas por Mes           | Promedio de reseñas por mes.                            |

Una vez realizada la información anterior, haremos una lista con las 50 variables que desde mi punto de vista son las más relevantes para poder hacer un análisis de los datos:

1. Identificador (id): Es el identificador único de Airbnb para cada anuncio. Es crucial para diferenciar entre los anuncios.
2. URL del Anuncio (listing\_url): La URL única de Airbnb para cada anuncio. Útil para verificar detalles específicos de cada anuncio en la plataforma.
3. ID del Scrape (scrape\_id): Permite rastrear la fuente de la información de cada anuncio.
4. Última Extracción (last\_scraped): Indica la fecha y hora en que se extrajo la información, lo que ayuda a mantener los datos actualizados.
5. Fuente (source): Informa sobre el origen de los datos, lo que puede ser relevante para la calidad y el contexto de la información.
6. Nombre (name): El nombre del anuncio puede ser útil para identificar el tipo de propiedad y su atractivo.
7. Descripción (description): Proporciona una descripción detallada del anuncio, importante para entender las características de la propiedad.
8. Descripción del Vecindario (neighborhood\_overview): Ayuda a conocer el entorno del anuncio, lo que puede influir en las decisiones de los huéspedes.

9. URL de Imagen (picture\_url): La URL de la imagen del anuncio, útil para verificar visualmente las propiedades.
10. ID del Anfitrión (host\_id): Identificador único para cada anfitrión, crucial para analizar patrones de comportamiento de los anfitriones.
11. URL del Anfitrión (host\_url): Enlace a la página del anfitrión en Airbnb, útil para validar la autenticidad y la información del anfitrión.
12. Nombre del Anfitrión (host\_name): El nombre del anfitrión puede ser relevante para el análisis de datos de satisfacción de los huéspedes.
13. Anfitrión Desde (host\_since): Fecha de creación del anfitrión, importante para analizar la experiencia y trayectoria del anfitrión.
14. Ubicación del Anfitrión (host\_location): La ubicación reportada del anfitrión puede influir en la confiabilidad y disponibilidad del mismo.
15. Acerca del Anfitrión (host\_about): Información sobre el anfitrión, útil para entender mejor al anfitrión y su perfil.
16. Tiempo de Respuesta (host\_response\_time): El tiempo típico de respuesta del anfitrión, importante para evaluar la eficiencia del servicio al cliente.
17. Tasa de Respuesta (host\_response\_rate): Tasa de respuesta a consultas, relevante para medir la proactividad del anfitrión.
18. Tasa de Aceptación (host\_acceptance\_rate): Tasa de aceptación de reservas, clave para analizar la disposición del anfitrión.
19. ¿Es Superhost? (host\_is\_superhost): Indica si el anfitrión es un superhost, lo que puede correlacionarse con la calidad del servicio.
20. URL Miniatura del Anfitrión (host\_thumbnail\_url): URL de la imagen miniatura del anfitrión, útil para verificación visual.
21. URL Foto del Anfitrión (host\_picture\_url): URL de la foto de perfil del anfitrión, relevante para la verificación visual y credibilidad.
22. Vecindario del Anfitrión (host\_neighbourhood): El vecindario reportado del anfitrión, importante para entender el contexto geográfico del anfitrión.

23. Número de Anuncios (host\_listings\_count): El número de anuncios que tiene el anfitrión, clave para evaluar la experiencia del anfitrión.
24. Total de Anuncios (host\_total\_listings\_count): Número total de anuncios del anfitrión, útil para análisis de la escala de operación del anfitrión.
25. Verificaciones del Anfitrión (host\_verifications): Métodos de verificación para el anfitrión, importantes para validar la autenticidad del anfitrión.
26. ¿Tiene Foto de Perfil? (host\_has\_profile\_pic): Indica si el anfitrión tiene foto de perfil, relevante para la confianza del usuario.
27. Identidad Verificada (host\_identity\_verified): Indica si la identidad del anfitrión está verificada, clave para la seguridad del huésped.
28. Vecindario (neighbourhood): Vecindario reportado del anuncio, importante para análisis de ubicación.
29. Vecindario Limpio (neighbourhood\_cleansed): Vecindario geocodificado, útil para análisis de precisión geográfica.
30. Grupo de Vecindarios (neighbourhood\_group\_cleansed): Grupo de vecindarios geocodificado, relevante para análisis de tendencias geográficas.
31. Latitud (latitude): Latitud del anuncio, clave para análisis geoespacial.
32. Longitud (longitude): Longitud del anuncio, relevante para análisis geoespacial.
33. Tipo de Propiedad (property\_type): Tipo de propiedad seleccionada por el anfitrión, importante para segmentación de mercado.
34. Tipo de Habitación (room\_type): Tipo de habitación, útil para análisis de preferencias de los huéspedes.
35. Capacidad (accommodates): La capacidad máxima del anuncio, clave para entender la oferta de alojamiento.
36. Baños (bathrooms): El número de baños en el anuncio, relevante para el confort y conveniencia del huésped.

37. Texto de Baños (bathrooms\_text): Descripción textual de los baños, útil para comprender los detalles de las instalaciones.
38. Habitaciones (bedrooms): El número de habitaciones en el anuncio, importante para la capacidad de alojamiento.
39. Camas (beds): El número de camas en el anuncio, relevante para la planificación de la ocupación.
40. Servicios (amenities): Servicios disponibles en el anuncio, clave para la satisfacción del huésped.
41. Precio (price): Precio diario en moneda local, crucial para análisis de costos y comparaciones.
42. Noches Mínimas (minimum\_nights): Noches mínimas requeridas, relevante para entender las políticas de estancia.
43. Noches Máximas (maximum\_nights): Noches máximas permitidas, importante para la planificación de la estancia.
44. Disponibilidad (30 días) (availability\_30): Disponibilidad para los próximos 30 días, clave para análisis de oferta a corto plazo.
45. Disponibilidad (60 días) (availability\_60): Disponibilidad para los próximos 60 días, útil para análisis de oferta a mediano plazo.
46. Disponibilidad (90 días) (availability\_90): Disponibilidad para los próximos 90 días, relevante para análisis de oferta.
47. Disponibilidad (365 días) (availability\_365): Disponibilidad para el próximo año, clave para análisis de tendencias a largo plazo.
48. Calendario Actualizado (calendar\_updated): Fecha en que se actualizó el calendario, importante para evaluar la frescura de los datos.
49. Número de Reseñas (number\_of\_reviews): Número total de reseñas, útil para medir la popularidad del anuncio.
50. Primera Reseña (first\_review): Fecha de la primera reseña, relevante para analizar la trayectoria del anuncio.