

## Actividad 1

La información obtenida de los almacén de datos se encuentra en idioma inglés, por lo que para entender y comprender los datos es necesario traducir las columnas. Por lo que a continuación se presentarán las columnas en orden y traducidas que contiene el almacén de datos junto a la descripción de la información que contiene cada columna.

1. **id**: identificador único para el anuncio.
2. **anuncio\_url**: dirección web del anuncio en airbnb.
3. **extraccion\_id**: identificador de la extracción de datos.
4. **ultima\_extraccion**: utc fecha y hora en la que se realizó la última extracción de datos.
5. **fuelle**: origen de los datos. " city scrape " significa que el anuncio se encontró buscando en la ciudad, mientras que "previous scrape " significa que el anuncio se vio en otro raspado realizado en los últimos 65 días y se confirmó que el anuncio todavía estaba disponible en el sitio de airbnb.
6. **nombre**: nombre/título del anuncio de la propiedad.
7. **descripción**: texto descriptivo que detalla la propiedad y sus características.
8. **vecindario\_descripcion**: texto descriptivo que detalla el vecindario y sus características.
9. **imagen\_url**: dirección web de la imagen alojada en airbnb del anuncio.
10. **anfitrión\_id**: identificador único del anfitrión/usuario.
11. **anfitrión\_url**: dirección web del perfil airbnb del anfitrión.
12. **anfitrión\_nombre**: nombre del anfitrión. usualmente es el primero.
13. **anfitrión\_desde**: fecha en que se creó el host/usuario. para los anfitriones que son invitados de airbnb, esta podría ser la fecha en la que se registraron como invitados.
14. **anfitrión\_ubicacion**: ubicación dada por el anfitrión.
15. **anfitrión\_informacion**: texto descriptivo acerca del anfitrión.

16. **anfitrion\_tiempo\_respuesta**: tiempo promedio que tarda el anfitrión en responder.
17. **anfitrion\_tasa\_respuesta**: porcentaje de mensajes respondidos por el anfitrión.
18. **anfitrion\_tasa\_aceptacion**: porcentaje de solicitudes aceptadas por el anfitrión.
19. **anfitrion\_es\_superanfitrion**: el anfitrión tiene el estatus de super host (sí o no).
20. **anfitrion\_miniatura\_url**: dirección web de la miniatura de la foto del anfitrión.
21. **anfitrion\_foto\_url**: dirección web de la foto completa del anfitrión.
22. **anfitrion\_vecindario**: vecindario donde vive el anfitrión.
23. **anfitrion\_cantidad\_anuncios**: cantidad de anuncios activos que tiene el anfitrión (según cálculos desconocidos de airbnb).
24. **anfitrion\_cantidad\_total\_anuncios**: cantidad total de anuncios que tiene el anfitrión (según cálculos desconocidos de airbnb).
25. **anfitrion\_verificaciones**: métodos de verificación del anfitrión (correo electrónico, teléfono, etc.).
26. **anfitrion\_tiene\_foto\_perfil**: indica si el anfitrión tiene una foto de perfil (sí o no).
27. **anfitrion\_identidad\_verificada**: indica si la identidad del anfitrión ha sido verificada (sí o no).
28. **vecindario**: vecindario.
29. **vecindario\_geocodificado**: nombre del vecindario. el vecindario está geocodificado utilizando la latitud y longitud frente a los vecindarios definidos por archivos de forma digitales abiertos o públicos.
30. **grupo\_vecindarios\_geocodificado**: nombre de la ciudad donde se encuentra. el grupo de vecinos está geocodificado utilizando la latitud y longitud de los barrios definidos por archivos de forma digitales abiertos o públicos.
31. **latitud**: coordenada de latitud de la propiedad. utiliza la proyección del sistema geodésico mundial (wgs84) para latitud y longitud.
32. **longitud**: coordenada de longitud de la propiedad. utiliza la proyección del sistema geodésico mundial (wgs84) para latitud y longitud.
33. **tipo\_propiedad**: indica tipo de propiedad autoseleccionado. los hoteles y bed and breakfast son descritos como tales por sus anfitriones en este campo.

- 34. **tipo\_habitacion**: indica el tipo de espacio disponible (habitación privada, compartida, todo el lugar).
- 35. **capacidad**: número máximo de huéspedes que la propiedad puede alojar.
- 36. **baños**: número de baños en la propiedad.
- 37. **baños\_descripcion**: descripción textual de los baños (compartido, privado, etc.).
- 38. **dormitorios**: número de dormitorios.
- 39. **camas**: número de camas disponibles.
- 40. **servicios**: lista de servicios disponibles en la propiedad.
- 41. **precio**: precio por noche de la propiedad.
- 42. **noches\_minimas**: cantidad de noches mínimas requeridas para solicitar alojamiento.
- 43. **noches\_maximas**: cantidad de noches máximas requeridas para solicitar alojamiento.
- 44. **noches\_minimas\_minimo**: valor más bajo permitido para noches mínimas.
- 45. **noches\_minimas\_maximo**: valor más alto permitido para noches mínimas.
- 46. **noches\_maximas\_minimo**: valor más bajo permitido para noches máximas.
- 47. **noches\_maximas\_maximo**: valor más alto permitido para noches máximas.
- 48. **noches\_minimo\_promedio**: promedio de noches mínimas en el mes siguiente.
- 49. **noches\_maximo\_promedio**: promedio de noches máximas en el mes siguiente.
- 50. **calendario\_actualizado**: fecha de cuándo se actualizó el calendario del anuncio por última vez.
- 51. **tiene\_disponibilidad**: indica si la propiedad está disponible (sí o no).
- 52. **disponibilidad\_30**: indica si la propiedad está disponible en los próximos 30 días (sí o no).
- 53. **disponibilidad\_60**: indica si la propiedad está disponible en los próximos 60 días (sí o no).
- 54. **disponibilidad\_90**: indica si la propiedad está disponible en los próximos 90 días (sí o no).
- 55. **disponibilidad\_365**: indica si la propiedad está disponible en los próximos 365 días (sí o no).

- 56. **ultima\_extraccion\_calendario**: fecha de última recopilación de datos del calendario.
- 57. **numero\_de\_reseñas**: cantidad de reseñas que tiene el anuncio.
- 58. **numero\_de\_reseñas\_udm**: cantidad de reseñas de los últimos 12 meses que tiene el anuncio.
- 59. **numero\_de\_reseñas\_u30d**: cantidad de reseñas de los últimos 30 días que tiene el anuncio.
- 60. **primera\_reseña**: fecha de la primera reseña (la reseñas más vieja) echa.
- 61. **ultima\_reseña**: fecha de la última reseña (reseña más nueva) echa.
- 62. **reseñas\_puntuacion\_general**: puntuación promedio de todas las reseñas.
- 63. **reseñas\_puntuacion\_precision**: puntuación promedio sobre la precisión de la información del anuncio.
- 64. **reseñas\_puntuacion\_limpieza**: puntuación promedio sobre la limpieza de la propiedad.
- 65. **reseñas\_puntuacion\_registro**: puntuación promedio sobre la facilidad para registrarse.
- 66. **reseñas\_puntuacion\_comunicacion**: puntuación promedio sobre la comunicación del anfitrión.
- 67. **reseñas\_puntuacion\_ubicacion**: puntuación promedio sobre la ubicación de la propiedad.
- 68. **reseñas\_puntuacion\_valor**: puntuación promedio sobre la relación calidad-precio.
- 69. **licencia**: número de licencia o permiso para alquilar la propiedad.
- 70. **reserva\_inmediata**: la propiedad permite la reserva inmediata sin que el anfitrión tenga que aceptar su solicitud de reserva.
- 71. **anfitrión\_cantidad\_calculada\_anuncios**: número calculado de propiedades activas del anfitrión.
- 72. **anfitrión\_cantidad\_calculada\_casas\_enteras**: número de propiedades completas (no compartidas) del anfitrión.
- 73. **anfitrión\_cantidad\_calculada\_cuartos\_privados**: número de habitaciones privadas publicadas por el anfitrión.

74. **anfitrión\_cantidad\_calculada\_habitaciones\_compartidas**: número de habitaciones compartidas publicadas por el anfitrión.

75. **reseñas\_por\_mes**: promedio de reseñas recibidas por mes.

Para entender el rendimiento de las propiedades, la calidad del servicio, la segmentación del mercado, y la estrategia de precios y disponibilidad. Por lo que continuación se enlistarán las 50 columnas más importantes junto a su justificación, para el análisis de datos respecto a los requerimientos del cliente **Airbnb**.

1. **nombre**: es fundamental para identificar y diferenciar que impacto tienen los nombres de los anuncios.
2. **descripción**: proporciona información clave sobre las características de la propiedad, lo que es importante para análisis cualitativos.
3. **vecindario\_descripcion**: ayuda a entender la ubicación y el contexto del vecindario, crucial para la segmentación y análisis de la demanda.
4. **anfitrión\_nombre**: es importante para ver el impacto que tienen los nombres de los anfitriones.
5. **anfitrión\_desde**: permite conocer la experiencia del anfitrión en la plataforma, lo cual puede ser relevante para análisis de confiabilidad y desempeño.
6. **anfitrión\_ubicacion**: localización general del anfitrión, útil para análisis de la proximidad a la propiedad, confiabilidad, etc.
7. **anfitrión\_informacion**: puede ser útil proporcionar detalles adicionales sobre el anfitrión en un análisis cualitativo.
8. **anfitrión\_tiempo\_respuesta**: importante para entender el impacto que tiene la eficiencia en la comunicación del anfitrión con los huéspedes.
9. **anfitrión\_tasa\_respuesta**: refleja la efectividad del anfitrión en la comunicación, lo que es útil para análisis de atención al cliente.
10. **anfitrión\_tasa\_aceptacion**: indica la fiabilidad del anfitrión y puede influir en la toma de decisiones de los huéspedes.
11. **anfitrión\_es\_superanfitrión**: este atributo es crucial para evaluar la reputación y el rendimiento de los anfitriones.

12. **anfitrión\_vecindario**: ayuda a conocer la ubicación del anfitrión y puede ser útil para análisis geográficos.
13. **anfitrión\_cantidad\_anuncios**: indica el número de propiedades activas, importante para analizar la capacidad del anfitrión y su desempeño.
14. **anfitrión\_cantidad\_total\_anuncios**: permite conocer la totalidad de propiedades del anfitrión, importante para analizar la diversidad de su oferta.
15. **anfitrión\_verificaciones**: aporta información sobre la confiabilidad del anfitrión, un aspecto importante para la confianza del huésped.
16. **anfitrión\_identidad\_verificada**: indica la confianza y la seguridad del anfitrión, fundamental para análisis de reputación y seguridad.
17. **vecindario\_geocodificado**: localiza geográficamente la propiedad, crucial para análisis de ubicación y proximidad.
18. **grupo\_vecindarios\_geocodificado**: ayuda a agrupar las propiedades por ciudad o región, útil para análisis más amplios a nivel de localización.
19. **latitud**: esencial para análisis geoespacial y mapeo preciso de la ubicación de la propiedad.
20. **longitud**: complementa a la latitud para análisis geoespaciales y mapeo.
21. **tipo\_propiedad**: determina el tipo de propiedad, crucial para segmentar el mercado y entender la oferta disponible.
22. **tipo\_habitacion**: permite conocer el tipo de espacio disponible, útil para segmentar la demanda de acuerdo con las preferencias de los huéspedes.
23. **capacidad**: número máximo de huéspedes, importante para entender la capacidad de alojamiento y la demanda.
24. **baños\_descripcion**: proporciona detalles sobre los baños, lo cual puede ser relevante para quienes buscan servicios específicos.
25. **dormitorios**: es importante para determinar la capacidad de la propiedad y la segmentación del mercado.
26. **camas**: información clave para entender la capacidad de alojamiento de la propiedad.
27. **servicios**: listado de servicios, crucial para diferenciar las propiedades y entender su competitividad en el mercado.

28. **precio**: importantes para el análisis de la oferta y demanda, a igual que para la toma de decisiones de precios.
29. **noches\_minimas**: permite analizar las restricciones de la propiedad y su impacto en la disponibilidad y el mercado.
30. **noches\_maximas**: similar a las noches mínimas, permite entender el impacto y la oferta de la propiedad.
31. **disponibilidad\_30**: indica la disponibilidad a corto plazo, crucial para evaluar la demanda y las operaciones.
32. **disponibilidad\_60**: permite analizar la demanda en el mediano plazo.
33. **disponibilidad\_90**: también útil para la planificación y análisis de la demanda a mediano plazo.
34. **disponibilidad\_365**: permite evaluar la estrategia de ocupación a largo plazo y la planificación anual.
35. **numero\_de\_reseñas**: indica la popularidad y el nivel de interacción de la propiedad con los huéspedes.
36. **numero\_de\_reseñas\_udm**: analiza el rendimiento reciente de la propiedad en cuanto a la recepción de comentarios.
37. **reseñas\_puntuacion\_general**: es una de las métricas más relevantes para evaluar la calidad general de la propiedad.
38. **reseñas\_puntuacion\_precision**: ayuda a evaluar si la propiedad se corresponde con la descripción, importante para la confianza del huésped.
39. **reseñas\_puntuacion\_limpieza**: muy relevante para la experiencia del huésped, una de las puntuaciones más consultadas.
40. **reseñas\_puntuacion\_registro**: la facilidad para registrarse es un factor relevante para la experiencia del huésped.
41. **reseñas\_puntuacion\_comunicacion**: mide la calidad de la interacción entre anfitrión y huésped, crucial para la satisfacción general.
42. **reseñas\_puntuacion\_ubicacion**: la ubicación es uno de los factores más valorados por los huéspedes, importante para análisis de competitividad.
43. **reseñas\_puntuacion\_valor**: mide la relación calidad-precio, un factor decisivo para la decisión de los huéspedes.

- 44. **licencia**: información sobre la legalidad de la propiedad, importante para análisis de cumplimiento normativo.
- 45. **reserva\_inmediata**: la capacidad de reservar sin intervención del anfitrión es relevante para la flexibilidad y conveniencia para los huéspedes.
- 46. **anfitrión\_cantidad\_calculada\_anuncios**: permite evaluar la capacidad y desempeño del anfitrión a través de sus propiedades activas.
- 47. **anfitrión\_cantidad\_calculada\_casas\_enteras**: ayuda a segmentar las propiedades del anfitrión en términos de tipo de alojamiento.
- 48. **anfitrión\_cantidad\_calculada\_cuartos\_privados**: información útil para entender la oferta de habitaciones privadas.
- 49. **anfitrión\_cantidad\_calculada\_habitaciones\_compartidas**: complementa la oferta del anfitrión en términos de habitaciones compartidas.
- 50. **reseñas\_por\_mes**: una métrica importante para analizar el ritmo de crecimiento y la interacción continua con los huéspedes.