



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus: Puebla

Materia: Analítica de datos y herramientas de inteligencia artificial I

Actividad 1 (Entendimiento de los Datos)

Diego Sánchez Márquez

16/02/2025

- Realizar la traducción a español y la descripción de cada una de las variables de la base de datos con ayuda del "Data Dictionary" proporcionado por la compañía Airbnb.
 1. **id** (integer): Identificador único de Airbnb para la propiedad.
 2. **listing_url** (text): URL del anuncio en Airbnb.
 3. **scrape_id** (bigint): ID de la "extracción" de datos de Inside Airbnb en la que se incluyó este anuncio.
 4. **last_scraped** (datetime): Fecha y hora en UTC en la que se extrajo la información de este anuncio.
 5. **source** (text): Indica si la fuente de la extracción fue una "búsqueda por barrio" o una "extracción previa".
 6. **name** (text): Nombre del anuncio.
 7. **description** (text): Descripción detallada del anuncio.
 8. **neighborhood_overview** (text): Descripción del barrio proporcionada por el anfitrión.
 9. **picture_url** (text): URL de la imagen alojada en Airbnb del anuncio.
 10. **host_id** (integer): Identificador único de Airbnb para el anfitrión/usuario.
 11. **host_url** (text): Página de Airbnb del anfitrión.
 12. **host_name** (text): Nombre del anfitrión (generalmente solo el nombre de pila).
 13. **host_since** (date): Fecha en que el anfitrión se registró en Airbnb.
 14. **host_location** (text): Ubicación proporcionada por el anfitrión.
 15. **host_about** (text): Descripción sobre el anfitrión.
 16. **host_response_time** (text): Tiempo de respuesta del anfitrión.
 17. **host_response_rate** (text): Tasa de respuesta del anfitrión.
 18. **host_acceptance_rate** (text): Tasa de aceptación de solicitudes de reserva por parte del anfitrión.
 19. **host_is_superhost** (boolean [t=true; f=false]): Indica si el anfitrión es un "Superhost".
 20. **host_thumbnail_url** (text): URL de la miniatura de la imagen del anfitrión.
 21. **host_picture_url** (text): URL de la imagen del anfitrión.
 22. **host_neighbourhood** (text): Barrio donde reside el anfitrión.
 23. **host_listings_count** (text): Número de propiedades que tiene el anfitrión (según cálculos desconocidos de Airbnb).
 24. **host_total_listings_count** (text): Número total de propiedades que tiene el anfitrión (según cálculos desconocidos de Airbnb).
 25. **host_verifications** (text): Métodos de verificación utilizados por el anfitrión.
 26. **host_has_profile_pic** (boolean [t=true; f=false]): Indica si el anfitrión tiene foto de perfil.
 27. **host_identity_verified** (boolean [t=true; f=false]): Indica si la identidad del anfitrión ha sido verificada.
 28. **neighbourhood** (text): Nombre del barrio según Airbnb.
 29. **neighbourhood_cleansed** (text): Nombre del barrio geocodificado utilizando latitud y longitud con archivos de datos abiertos o públicos.
 30. **neighbourhood_group_cleansed** (text): Grupo de barrios geocodificado utilizando latitud y longitud con archivos de datos abiertos o públicos.
 31. **latitude** (numeric): Latitud en el sistema de coordenadas geográficas WGS84.

32. **longitude** (numeric): Longitud en el sistema de coordenadas geográficas WGS84.
33. **property_type** (text): Tipo de propiedad seleccionado por el anfitrión.
34. **room_type** (text): Tipo de habitación disponible (casa/apt entero, habitación privada, habitación compartida, hotel).
35. **accommodates** (integer): Capacidad máxima del alojamiento.
36. **bathrooms** (numeric): Número de baños en la propiedad.
37. **bathrooms_text** (string): Descripción textual del número de baños (debido a cambios en el sitio web de Airbnb).
38. **bedrooms** (integer): Número de habitaciones.
39. **beds** (integer): Número de camas.
40. **amenities** (json): Lista de comodidades disponibles en la propiedad.
41. **price** (currency): Precio diario en la moneda local (ignorar el signo "\$").
42. **minimum_nights** (integer): Número mínimo de noches requeridas para reservar.
43. **maximum_nights** (integer): Número máximo de noches permitidas en la reserva.
44. **minimum_minimum_nights** (integer): Valor más bajo del mínimo de noches requerido en los próximos 365 días.
45. **maximum_minimum_nights** (integer): Valor más alto del mínimo de noches requerido en los próximos 365 días.
46. **minimum_maximum_nights** (integer): Valor más bajo del máximo de noches permitido en los próximos 365 días.
47. **maximum_maximum_nights** (integer): Valor más alto del máximo de noches permitido en los próximos 365 días.
48. **minimum_nights_avg_ntm** (numeric): Promedio del mínimo de noches requerido en los próximos 365 días.
49. **maximum_nights_avg_ntm** (numeric): Promedio del máximo de noches permitido en los próximos 365 días.
50. **calendar_updated** (date): Última actualización del calendario.
51. **has_availability** (boolean [t=true; f=false]): Indica si la propiedad está disponible.
52. **availability_30** (integer): Disponibilidad de la propiedad en los próximos 30 días.
53. **availability_60** (integer): Disponibilidad de la propiedad en los próximos 60 días.
54. **availability_90** (integer): Disponibilidad de la propiedad en los próximos 90 días.
55. **availability_365** (integer): Disponibilidad de la propiedad en los próximos 365 días.
56. **calendar_last_scraped** (date): Última vez que se extrajo información del calendario.
57. **number_of_reviews** (integer): Número total de reseñas recibidas.
58. **number_of_reviews_ltm** (integer): Número de reseñas recibidas en los últimos 12 meses.
59. **number_of_reviews_l30d** (integer): Número de reseñas recibidas en los últimos 30 días.
60. **first_review** (date): Fecha de la primera reseña recibida.

61. **last_review** (date): Fecha de la última reseña recibida.
62. **review_scores_rating** (integer): Calificación general del anuncio.
63. **review_scores_accuracy** (integer): Puntuación de precisión de la descripción.
64. **review_scores_cleanliness** (integer): Puntuación de limpieza.
65. **review_scores_checkin** (integer): Puntuación del proceso de check-in.
66. **review_scores_communication** (integer): Puntuación de comunicación del anfitrión.
67. **review_scores_location** (integer): Puntuación de la ubicación.
68. **review_scores_value** (integer): Puntuación de la relación calidad-precio.
69. **license** (text): Número de licencia, permiso o registro del anuncio.
70. **instant_bookable** (boolean [t=true; f=false]): Indica si la reserva puede realizarse de forma automática sin aprobación del anfitrión.
71. **calculated_host_listings_count** (integer): Número de anuncios que tiene el anfitrión en la extracción actual dentro de la ciudad/región.
72. **calculated_host_listings_count_entire_homes** (integer): Número de anuncios de casas/departamentos enteros que tiene el anfitrión en la extracción actual dentro de la ciudad/región.
73. **calculated_host_listings_count_private_rooms** (integer): Número de anuncios de habitaciones privadas que tiene el anfitrión en la extracción actual dentro de la ciudad/región.
74. **calculated_host_listings_count_shared_rooms** (integer): Número de anuncios de habitaciones compartidas que tiene el anfitrión en la extracción actual dentro de la ciudad/región.
75. **reviews_per_month** (numeric): Promedio de reseñas por mes a lo largo de la vida del anuncio.

- Seleccionar 50 variables que consideres más importantes para el análisis de datos priorizando los tipos de variables numéricas y categóricas. (Justificar la selección de cada variable)
 1. **id** (integer): Identificador único del anuncio, útil para unir datos.
 2. **host_id** (integer): Identificador único del anfitrión, clave para análisis de propietarios.
 3. **latitude** (numeric): Ubicación geográfica, necesaria para análisis espaciales.
 4. **longitude** (numeric): Ubicación geográfica, complementa la latitud.
 5. **neighbourhood_cleansed** (text): Nombre del barrio, útil para segmentación.
 6. **neighbourhood_group_cleansed** (text): Agrupación de barrios, permite análisis por zonas.
 7. **property_type** (text): Tipo de propiedad, importante para clasificar alojamientos.
 8. **room_type** (text): Tipo de habitación, clave para segmentar la oferta.
 9. **accommodates** (integer): Capacidad máxima de huéspedes, influye en la demanda.
 10. **bedrooms** (integer): Número de habitaciones, afecta el precio y la ocupación.
 11. **beds** (integer): Número de camas, relevante para la capacidad del alojamiento.

12. `bathrooms_text` (string): Número de baños, importante para la comodidad.
13. `price` (currency): Precio por noche, clave en análisis de rentabilidad.
14. `minimum_nights` (integer): Estancia mínima, afecta la flexibilidad del alquiler.
15. `maximum_nights` (integer): Estancia máxima, influye en reservas de largo plazo.
16. `availability_30` (integer): Disponibilidad en los próximos 30 días, útil para análisis de ocupación.
17. `availability_60` (integer): Disponibilidad en los próximos 60 días, complementa la anterior.
18. `availability_90` (integer): Disponibilidad en los próximos 90 días, ayuda a prever tendencias.
19. `availability_365` (integer): Disponibilidad en el año, mide la frecuencia de alquiler.
20. `has_availability` (boolean): Indica si la propiedad está activa o no.
21. `host_since` (date): Fecha en que el anfitrión se unió, mide su experiencia.
22. `host_response_time` (text): Tiempo que tarda en responder, influye en la satisfacción del cliente.
23. `host_response_rate` (text): Porcentaje de respuestas, mide la actividad del anfitrión.
24. `host_acceptance_rate` (text): Porcentaje de solicitudes aceptadas, muestra disponibilidad real.
25. `host_is_superhost` (boolean): Indica si el anfitrión es "Superhost", útil para evaluar su desempeño.
26. `host_listings_count` (integer): Número de propiedades que tiene el anfitrión, diferencia particulares de grandes operadores.
27. `host_identity_verified` (boolean): Indica si la identidad del anfitrión está verificada, relevante para confianza.
28. `number_of_reviews` (integer): Cantidad total de reseñas, indica popularidad del anuncio.
29. `number_of_reviews_ltm` (integer): Reseñas en el último año, útil para medir actividad reciente.
30. `number_of_reviews_l30d` (integer): Reseñas en el último mes, muestra tendencias recientes.
31. `first_review` (date): Fecha de la primera reseña, indica antigüedad del anuncio.
32. `last_review` (date): Fecha de la última reseña, ayuda a detectar listados activos.
33. `review_scores_rating` (integer): Puntuación general del anuncio, mide satisfacción.
34. `review_scores_accuracy` (integer): Puntuación sobre precisión de la descripción.
35. `review_scores_cleanliness` (integer): Puntuación de limpieza, clave en la experiencia del huésped.
36. `review_scores_checkin` (integer): Puntuación del proceso de check-in, afecta comodidad.
37. `review_scores_communication` (integer): Puntuación sobre la comunicación con el anfitrión.

38. review_scores_location (integer): Puntuación de la ubicación, relevante para viajeros.
39. review_scores_value (integer): Puntuación de relación calidad-precio, clave para la demanda.
40. minimum_minimum_nights (integer): Estancia mínima más baja en el año, afecta disponibilidad.
41. maximum_minimum_nights (integer): Estancia mínima más alta en el año, importante para restricciones.
42. minimum_maximum_nights (integer): Estancia máxima más baja en el año, clave para reservas largas.
43. maximum_maximum_nights (integer): Estancia máxima más alta en el año, muestra límites en alquiler.
44. minimum_nights_avg_ntm (numeric): Promedio de noches mínimas, útil para tendencias.
45. maximum_nights_avg_ntm (numeric): Promedio de noches máximas, analiza patrones de reserva.
46. calculated_host_listings_count (integer): Número de anuncios del anfitrión en la ciudad, distingue operadores grandes.
47. calculated_host_listings_count_entire_homes (integer): Número de propiedades enteras, clave para analizar oferta.
48. calculated_host_listings_count_private_rooms (integer): Número de habitaciones privadas, segmenta la oferta.
49. calculated_host_listings_count_shared_rooms (integer): Número de habitaciones compartidas, muestra opciones económicas.
50. reviews_per_month (numeric): Promedio de reseñas mensuales, indica nivel de actividad del anuncio.