



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias de la Computación

Ingeniería en Tecnologías de la Información

Nombre del Alumno: Eduardo Ballinas Ballinas

Matrícula:202321840

Actividad 1 (Etapa 1 Análisis)

3. Realizar la traducción al 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.

En esta sección nos encargaremos de traducir y describir cada una de las variables que se encuentran en la base de datos, esto con el objetivo de tener fácil comprensión de cada uno de ellos.

1. **id (Identificador único):** El id es el identificador único que proporciona Airbnb para el listado de los anfitriones. Es de tipo entero.
2. **listing_url (URL del anuncio):** Es el enlace que dirige al anuncio en Airbnb de la casa/habitación del anfitrión. Es de tipo texto.
3. **scrape_id (Identificador de scraping):** El id de los datos obtenidos por técnica de “Web Scraping” el cual básicamente es una búsqueda, realizados por la sección de Inside Airbnb. Es de tipo entero grande.
4. **last_scraped (Última búsqueda de scraping):** Es la fecha en la que esta lista fue buscada. Es de tipo fecha y hora.
5. **source (Origen):** Es un dato de texto que tiene dos valores: búsqueda previa o búsqueda por ciudad. La búsqueda previa se refiere a que el anuncio fue visto en otra búsqueda realizada en los últimos 65 días y se confirmó que el anuncio aún estaba disponible en el sitio de Airbnb. En la búsqueda por ciudad significa que se encontró el anuncio al buscar en la ciudad.
6. **name (Nombre):** Nombre del anuncio.
7. **description (Descripción):** Descripción detallada del anuncio.
8. **(Descripción general del vecindario):** Descripción del vecindario según el anfitrión.
9. **picture_url (URL de la imagen):** Es el URL que te redirige a la imagen de tamaño regular del lugar de Airbnb del anuncio.

- 10.host_id (Identificador único del anfitrión):** Identificador único del usuario o anfitrión en Airbnb.
- 11.host_url (URL del anfitrión):** La página de Airbnb para el anfitrión.
- 12.host_name (Nombre del anfitrión):** El nombre del anfitrión, generalmente solo el primer nombre.
- 13.host_since (Anfitrión desde):** La fecha en que el usuario del anfitrión creó su cuenta. Para los anfitriones que son huéspedes de Airbnb, esta podría ser la fecha en la que se registraron como huéspedes.
- 14.host_location (Ubicación del anfitrión):** La ubicación reportada por el anfitrión.
- 15.host_about (Sobre el anfitrión):** La descripción del anfitrión.
- 16.host_response_time (Tiempo de respuesta del anfitrión):** Tiempo promedio que tarda en responder el anfitrión.
- 17.host_response_rate (Tasa de respuesta del anfitrión):** Porcentaje de respuesta del anfitrión.
- 18.host_acceptance_rate (Tasa de aceptación del anfitrión):** Porcentaje de aceptación de reserva.
- 19.host_is_superhost (El anfitrión es super anfitrión):** Valor booleano que indica si un anfitrión es super anfitrión o no.
- 20.host_thumbnail_url (URL de la miniatura del anfitrión):** Enlace que muestra la miniatura de la foto del anfitrión.
- 21.host_picture_url (URL de la imagen del anfitrión):** Enlace que muestra la foto del anfitrión.
- 22.host_neighbourhood (Vecindario del anfitrión):** Nombre del vecindario de la ubicación del anfitrión.
- 23.host_listings_count (Cantidad de listados activos del anfitrión):** La cantidad de propiedades activas que tiene el anfitrión (según cálculos desconocidos de Airbnb).
- 24.host_total_listings_count (Cantidad total de listados del anfitrión):** La cantidad de propiedades totales que tiene el anfitrión (según cálculos desconocidos de Airbnb).

- 25.host_verifications (Verificaciones del anfitrión):** Indica si el anfitrión tiene verificado el email, el teléfono o ambos.
- 26.host_has_profile_pic (El anfitrión tiene foto de perfil):** Valor booleano que indica si el anfitrión posee una foto de perfil o no.
- 27.host_identity_verified (Verificación de identidad del anfitrión):** Valor booleano que indica si el anfitrión ha verificado su identidad.
- 28.neighbourhood (Vecindario):** Indica el nombre de la villa, municipio o lugar asociado a la ubicación en la que se encuentra la casa/departamento.
- 29.neighbourhood_cleansed (Vecindario verificado):** El vecindario geocodificado utilizando latitud y longitud de la ubicación.
- 30.neighbourhood_group_cleansed (Grupo de vecindarios verificados):** El grupo de vecindarios geocodificado utilizando latitud y longitud de la ubicación.
- 31.latitude (Latitud):** Utiliza la proyección del Sistema Geodésico Mundial (WGS84) para latitud y longitud.
- 32.longitude (Longitud):** Utiliza la proyección del Sistema Geodésico Mundial (WGS84) para latitud y longitud.
- 33.property_type (Tipo de propiedad):** Tipo de propiedad seleccionada por el propietario.
- 34.room_type (Tipo de habitación):** Es en donde se clasifica el tipo de servicio se ofrece en ese lugar. Existen cuatro definidos: alojamiento completo, habitación privada, habitación compartida y hotel.
- 35.accommodates (Capacidad del alojamiento):** La capacidad máxima del alojamiento.
- 36.bathrooms (Baños):** El número de baños en el alojamiento.
- 37.bathrooms_text (Texto de baños):** El número de baños en el alojamiento. En el sitio web de Airbnb, el campo de baños se convierte de un número a un texto. Para búsquedas anteriores, se usa "bathrooms".
- 38.bedrooms (Dormitorios):** El número de dormitorios en el alojamiento.
- 39.beds (Camas):** El número de camas en el alojamiento.

- 40.amenities (Comodidades):** Comodidades y servicios que se ofrecen en el alojamiento.
- 41.price (Precio):** Precio diario en moneda local.
- 42.minimum_nights (Noches mínimas):** Número mínimo de noches de estancia para el alojamiento (las reglas del calendario pueden ser diferentes).
- 43.maximum_nights (Noches máximas):** Número máximo de noches de estancia para el alojamiento (las reglas del calendario pueden ser diferentes).
- 44.minimum_minimum_nights (Mínimo de noches mínimas):** El valor mínimo de “minimum_nights” del calendario (mirando 365 noches en el futuro).
- 45.maximum_minimum_nights (Máximo de noches mínimas):** El valor máximo de “minimum_nights” del calendario (mirando 365 noches en el futuro).
- 46.minimum_maximum_nights (Mínimo de noches máximas):** El valor mínimo de “maximum_nights” del calendario (mirando 365 noches en el futuro).
- 47.maximum_maximum_nights (Máximo de noches máximas):** El valor máximo de “maximum_nights” del calendario (mirando 365 noches en el futuro).
- 48.minimum_nights_avg_ntm (Promedio de noches mínimas en los próximos 12 meses):** El valor promedio de minimum_night del calendario (mirando 365 noches en el futuro).
- 49.maximum_nights_avg_ntm (Promedio de noches máximas en los próximos 12 meses):** El valor promedio de maximum_night del calendario (mirando 365 noches en el futuro).
- 50.calendar_update (Calendario actualizado):** No se encuentra ni en la descripción ni datos en la base de datos (Lisboa, Portugal).
- 51.has_availability (Tiene disponibilidad):** Un valor booleano que indica si un alojamiento esta disponible o no.
- 52.availability_30 (Disponibilidad de 30 días):** La disponibilidad del alojamiento dentro de 30 días según lo determine el calendario. Tenga en

cuenta que un alojamiento puede no estar disponible porque un huésped lo ha reservado o el anfitrión lo ha bloqueado.

53.availability_60 (Disponibilidad de 60 días): La disponibilidad del alojamiento dentro de 60 días según lo determine el calendario. Tenga en cuenta que un alojamiento puede no estar disponible porque un huésped lo ha reservado o el anfitrión lo ha bloqueado.

54.availability_90 (Disponibilidad de 90 días): La disponibilidad del alojamiento dentro de 90 días según lo determine el calendario. Tenga en cuenta que un alojamiento puede no estar disponible porque un huésped lo ha reservado o el anfitrión lo ha bloqueado.

55.availability_365 (Disponibilidad de 365 días): La disponibilidad del alojamiento dentro de 365 días según lo determine el calendario. Tenga en cuenta que un alojamiento puede no estar disponible porque un huésped lo ha reservado o el anfitrión lo ha bloqueado.

56.calendar_last_scraped (Último scraping del calendario): La fecha en la que se realizó la última búsqueda de scraping.

57.number_of_reviews (Número de reseñas): El número de reseñas que posee el anuncio.

58.number_of_reviews_ltm (Número de reseñas los últimos 12 meses): El número de reseñas que posee el anuncio (en los últimos 12 meses).

59.number_of_reviews_l30d (Número de reseñas los últimos 30 días): El número de reseñas que posee el anuncio (en los últimos 30 días).

60.first_review (Primera reseña): La fecha de la primera reseña.

61.last_review (Última reseña): La fecha de la última reseña.

62.review_scores_rating (Puntuación general de las reseñas): Calificación promedio del alojamiento.

63.review_scores_accuracy (Puntuación de precisión): Calificación promedio de la similitud de lo ofrecido y lo recibido.

64.review_scores_cleanliness (Puntuación de limpieza): Calificación promedio de limpieza.

- 65.review_scores_checkin (Puntuación de check-in):** Calificación promedio del check-in o entrada al alojamiento.
- 66.review_scores_communication (Puntuación de comunicación):** Calificación promedio de la comunicación entre anfitrión y huésped.
- 67.review_scores_location (Puntuación de ubicación):** Calificación promedio de la ubicación del alojamiento.
- 68.review_scores_value (Puntuación de relación calidad-precio):** Calificación promedio de costo a comparación de lo ofrecido
- 69.license (Licencia):** El número de licencia/permiso/registro.
- 70.instant_bookable (Reserva inmediata):** Valor booleano que indica si el huésped puede reservar el alojamiento automáticamente sin que el anfitrión tenga que aceptar su solicitud. Es un indicador de un alojamiento comercial.
- 71.calculated_host_listings_count (Cantidad calculada de listados del anfitrión):** La cantidad de propiedades u ofertas que tiene el anfitrión en el scrapping actual (la búsqueda actual), en la geografía de la ciudad/región.
- 72.calculated_host_listings_count_entire_homes (Cantidad calculada de listados de casas completas del anfitrión):** La cantidad de alojamientos completos que tiene el anfitrión en el scrapping actual (la búsqueda actual), en la geografía de la ciudad/región.
- 73.calculated_host_listings_count_private_rooms (Cantidad calculada de listados de habitaciones privadas del anfitrión):** La cantidad de habitaciones privadas que tiene el anfitrión en el scrapping actual (la búsqueda actual), en la geografía de la ciudad/región.
- 74.calculated_host_listings_count_shared_rooms (Cantidad calculada de listados de habitaciones compartidas del anfitrión):** La cantidad de habitaciones compartidas que tiene el anfitrión en el scrapping actual (la búsqueda actual), en la geografía de la ciudad/región.
- 75.reviews_per_month (Reseñas por mes):** La cantidad promedio de reseñas por mes que tiene el anuncio o alojamiento durante su vigencia.

4. Seleccionar 50 variables que consideres más importantes para el análisis de datos respecto a los requerimientos del cliente Airbnb. (Justificar la selección de cada variable).

1. id: Para identificar a cada uno de los anfitriones en la base de datos

5. source: Para tener noción de donde vienen la mayoría de los clientes, si de aquellos anuncios que se ofrecen por la búsqueda de ciudades o por aquellos que lo mostraron disponible en lapsos cercanos.

6. name: El nombre del anuncio es la base de la atracción, un buen nombre puede atraer a más personas.

7. description: En la descripción se relatan los detalles que necesita el usuario para determinar si es de su interés el lugar o no.

8. neighborhood_overview: La manera en la que el anfitrión explica los datos del vecindario da una idea el ambiente que se tiene en el lugar.

10. host_id: Con este se puede conocer si un anfitrión posee varias propiedades y conocer cuales son cada una de ellas.

12. host_name: El nombre del anfitrión permite asociar distintas propiedades a un solo anfitrión.

16. host_response_time: El tiempo de respuesta puede impactar de manera positiva o negativa dependiendo de si el tiempo es corto o largo, así que tiene relevancia en la retención de usuarios.

17. host_response_rate: La cantidad de respuestas influye en la cantidad de servicios concretados ya que la probabilidad aumenta mientras los anfitriones respondan la mayor cantidad de ellas.

18. host_acceptance_rate: Es importante conocer que tantas de las reservas son aceptadas para conocer su eficiencia y también asociar el hecho de que pueden ser algo estrictos o libres dependiendo de si aceptan o no a cualquiera.

19. host_is_superhost: Mientras más superhost hayan en la zona, mejor va a ser el servicio promedio de Airbnb en la ciudad.

23. host_listings_count: Conocer la cantidad de propiedades activas del anfitrión es importante para conocer el alcance de cada uno de ellos y su experiencia en el área.

24. host_total_listings_count: Por otro lado conocer la cantidad total de propiedades en la plataforma del anfitrión permite ver si este es activo actualmente o tiene propiedades en bloqueo temporal.

25. host_verifications: Con esto permite identificar la relación de que tan efectivo es tener o ya sea email, teléfono o ambos, y ver si hay un campo de oportunidad ahí que se este olvidando.

29. neighbourhood_cleansed: La muestra inicial de datos sobre cada uno de los vecindarios o barrios que conforman a la ciudad.

30. neighbourhood_group_cleansed: La misma importancia que el punto anterior pero de una forma en la que se agrupan a los vecindarios en grupos más sustanciales y diferenciables.

31 y 32. latitude and longitude: El punto geográfico en específico del alojamiento es un valor a tomar en cuenta de bastante importancia.

33. property_type: El tipo de propiedad también nos muestra la cantidad de cada uno de los elementos que hay en la ciudad según los anfitriones.

34. room_type: Esto define cada uno de los 4 diferentes tipos de habitaciones y eso permite ver el campo que tiene más potencial y ver si alguno de ellos se está desaprovechando.

35. accommodates: Conocer la capacidad de personas en el alojamiento es clave para conocer cuál es la capacidad más usada o buscada en la ciudad.

37. bathrooms_text: Preferí el uso de este en vez de “bathrooms” gracias a sus especificaciones en varios casos de la base de datos y que llena los campos donde el otro no lo hace.

38. bedrooms: La cantidad de dormitorios igual nos muestra si la gente suele preferir casas con pocos o muchos dormitorios.

39. beds: El número de camas igual representa un valor a identificar para ver si el enfoque de estas ciudades es para parejas o para ambiente familiar a través de esta y otras preferencias.

40. amenities: Esto cuenta como otros elementos que los clientes pueden tomar como extra y los cuales generen una mejor experiencia. Se podría rectificar alguna relación de calidad de servicios con la cantidad de comodidades mostradas.

41. price: El dinero es la base de cualquier dinero, por lo tanto es primordial conocer los precios preferidos por la gente y si el servicio ofrecido corresponde a los costos que se ofertan.

42. minimum_nights: El limitar la cantidad de noches mínimas puede ser contraproducente si las noches mínimas son demasiadas, por lo tanto habría que investigar cual es el número que muestra un punto de equilibrio entre rentar más días sin que el cliente no reserve por poner límites muy altos.

43. maximum_nights: Quizás menos importante que el anterior, pero limitar por poco igual puede representar un problema si alguna persona quisiera rentar la casa por más tiempo del que pareciera. Sería perfecto ver si hay clientes potenciales en la ciudad que soliciten muchas noches en el alojamiento.

48. minimum_nights_avg_ntm: Conociendo los datos del pasado podemos predecir el futuro y en este caso podemos ver si el número promedio de noches no se ve limitado por los últimos dos valores seleccionados.

49. maximum_nights_avg_ntm: De la misma manera conocer el promedio de noches máximas nos deja ver en cada uno de estos datos para asociar valores y generar conclusiones.

52, 53, 54 y 55. availability_30, availability_60, availability_90 and availability_365: Estos permiten ver que tanta capacidad tienen cada uno de los alojamientos, si están libres, muy llenos o bloqueados intencionalmente por los anfitriones, lo que nos permite ver la disponibilidad para nuevos clientes y en dado caso de falta de lugares, propuesta de nuevos alojamientos.

57. number_of_reviews: Conocer la cantidad de reseñas nos muestra que tan activo ha sido el anfitrión a lo largo de los años y muestra interés de parte de los usuarios para compartir su opinión. Es de importancia poner más atención a aquellos alojamientos con más reseñas e impulsar a aquellos que están en proceso de ser reconocidos.

58. number_of_reviews_ltm: Tener noción de la actividad que han tenido los distintos alojamientos a lo largo del último año permite ver la actividad en cada una de las temporadas y conocer en que puntos la gente tuvo tanto reacciones como buenas o malas, pero hubo movimientos en esos alojamientos.

59. number_of_reviews_l30d: Ver las reseñas de los últimos 30 días nos da un panorama más reciente de los acontecimientos dados en los alojamientos de la ciudad y ver si la gente que ha alojado alguno de estos tuvo ganas de compartir su opinión sobre ello. Conocer que piensa la gente nos permite ver que podemos ofrecer y con eso darles lo que quieren aunque no lo conozcan.

62. review_scores_rating: Probablemente este hasta el 68 sean de los datos más importantes para conocer la opinión verdadera sobre el alojamiento en Lisboa, Portugal, ya que muestran lo que la gente realmente piensa de una manera calificativa y eso nos permite ver que alojamientos si hacen las cosas bien y cuales tiendes a tener peores reseñas y con eso, conocer lo que el cliente promedio de la ciudad desea.

63. review_scores_accuracy: Ver si lo prometido en fotos y detalles se cumple genera buenas expectativas en los usuarios. Si los anfitriones cumplen con eso, generaran buenos puntos en estas áreas. Puede que estén relacionadas con su posicionamiento o estética.

64. review_scores_cleanliness: Si la limpieza caracteriza a estos lugares con buenas experiencias, se verán buenas notas en esta área. Las políticas de Airbnb son exigentes con su calidad y se debe cerciorar que se cumplan de manera correcta.

65. review_scores_checkin: La bienvenida es una de las partes primordiales de Airbnb ya que es la primera impresión en persona del cliente sobre el alojamiento. Ver que estas experiencias se cumplan será una tarea de tiempo, pero con estas puntuaciones será más cómodo ver quienes cumplen con esas expectativas.

66. review_scores_communication: La comunicación es clave por cualquier cosa. La atención de anfitrión se verá reflejada en esta categoría.

67. review_scores_location: Quizás la que menos se puede controlar de todas, pero se puede tomar una relación de los sitios favoritos de los clientes dependiendo de las reseñas de este estilo y así impulsar más alojamientos en zonas con buenos posicionamientos.

68. review_scores_value: Hacer valer la experiencia y el dinero del cliente es importante, y esta categoría va a reflejar si lo que se ofrece es acorde a lo que se paga en el servicio.

70. instant_bookable: Si el cliente puede hacer su reservación de inmediato, probablemente dude menos de realizarla y lo haga de manera más frecuente. Habrá que buscar relación entre la cantidad de reservaciones con esta herramienta.

71. calculated_host_listings_count: Conocer los datos específicos de propiedades por la región permite ver si este anfitrión conoce sobre el negocio en esta zona y permite ver resultados de aquellos que llevan más experiencia y aquellos que son nuevos pero tienen potencial.

72. calculated_host_listings_count_entire_homes: Ver cual tipo de habitaciones se ofrecen, en este caso alojamientos completos nos deja ver entre estos tres tipos cual es el favorito y más usado.

73. `calculated_host_listings_count_private_rooms`: En relación con el valor anterior, nos permite ver que tan populares son este tipo de habitación y ver si invertir más en ellas es bueno.

74. `calculated_host_listings_count_shared_rooms`: Probablemente el menos popular de los tipos de habitación pero no por ello olvidado, hay que investigar que puede ocasionar esto, quizás es algo que se puede controlar o quizás es solo parte de la cultura actual.

75. `reviews_per_month`: El ver que tan seguido se generan las reseñas permiten ver que tan activo es el anfitrión en el pasar del tiempo y comparar la cantidad de reseñas con su tiempo en crecimiento.

Conclusión:

Existían diferentes opciones por tomar para seleccionar los valores más importantes, pero considero que esta selección tiene un equilibrio entre valores numéricos importantes como detalles textuales que podrán ser de ayuda en un futuro cercano. Es bueno reflexionar sobre las relaciones entre ellos y ver como cada uno de los datos por si solos puede que no digan mucho pero en conjunto pueden significar más de lo que llegaríamos a creer.