



Tecnológico de Monterrey

Plataformas de analítica de negocios para
organizaciones

Actividad:

Actividad 1 Entendimiento de los Datos

Elaborado Por:

Héctor Pérez Martínez - A01277777

Fecha: 19/08/2025

Introducción

En este reporte presento mi análisis sobre los datos de Airbnb de Boston. Aquí explico cómo seleccioné 50 variables que considero clave para estudiar la plataforma. También detalló la metodología que seguí, justificó la elección de cada variable y muestro la traducción y descripción con base en el Data Dictionary de Airbnb.

Metodología de selección de variables

Para este trabajo decidí enfocarme en variables numéricas y categóricas, ya que son las que permiten hacer análisis más claros y comparables. Dejé fuera campos como descripciones largas, URLs, imágenes o licencias porque no aportan valor directo al análisis cuantitativo.

Selección y justificación de las 50 variables

1. Identificación básica

id → Identificador único de anuncio (clave primaria).

host_id → Identificador único de anfitrión (para agrupar por host).

Justificación: útiles para trazabilidad y análisis de oferta por anfitrión.

2. Datos de localización

latitude → Ubicación geográfica (numérica).

longitude → Ubicación geográfica (numérica).

neighbourhood_cleansed → Barrio normalizado (categórica).

host_location → Ciudad/región declarada por el host (categórica).

Justificación: permiten analizar patrones espaciales de precios y oferta.

3. Datos del anfitrión

host_since → Año en que el anfitrión se unió (categórica/fecha).

host_response_time → Rapidez de respuesta (categórica).

host_response_rate → Porcentaje de respuestas (numérica).

host_acceptance_rate → Porcentaje de aceptación (numérica).

host_is_superhost → Indicador de superhost (binaria).

host_listings_count → Número de anuncios activos del host (numérica).

host_total_listings_count → Número total de anuncios del host (numérica).

host_has_profile_pic → Binaria (sí/no).

host_identity_verified → Binaria (sí/no).

Justificación: variables de confianza y reputación del host, influyen en reservas y precios.

4. Propiedad y tipo de alojamiento

property_type → Tipo de propiedad (categórica).

room_type → Tipo de habitación (categórica).

accommodates → Número de personas que caben (numérica).

bathrooms → Número de baños (numérica).

bedrooms → Número de habitaciones (numérica).

beds → Número de camas (numérica).

bathrooms_text → Descripción del baño (categórica simplificada).

amenities → Lista de servicios (texto categórico).

Justificación: variables estructurales del alojamiento, clave para explicar precio y satisfacción.

5. Precio y condiciones

price → Precio listado (numérica, después de limpiar formato).

minimum_nights → Noches mínimas por reserva (numérica).

maximum_nights → Noches máximas (numérica).

minimum_minimum_nights → Restricciones mínimas históricas (numérica).

maximum_minimum_nights → Restricciones máximas históricas (numérica).

minimum_maximum_nights → Restricciones mínimas en máximo de noches (numérica).

maximum_maximum_nights → Restricciones máximas en máximo de noches (numérica).

minimum_nights_avg_ntm → Promedio noches mínimas por mes (numérica).

maximum_nights_avg_ntm → Promedio noches máximas por mes (numérica).

Justificación: regulan flexibilidad de reservas y afectan a la competitividad.

6. Disponibilidad del anuncio

has_availability → Binaria (sí/no).

availability_30 → Días disponibles en 30 días (numérica).

availability_60 → Días disponibles en 60 días (numérica).

availability_90 → Días disponibles en 90 días (numérica).

availability_365 → Días disponibles en 365 días (numérica).

availability_eoy → Días disponibles al final del año (numérica).

Justificación: miden la oferta temporal real, clave para estacionalidad y demanda.

7. Reseñas y reputación

number_of_reviews → Número total de reseñas (numérica).

number_of_reviews_ltm → Reseñas en los últimos 12 meses (numérica).

number_of_reviews_l30d → Reseñas en últimos 30 días (numérica).

Justificación: indicadores de popularidad y confianza.

8. Variables de fuente y control (útiles en exploración, limpieza o filtros)

source → Fuente de datos (categórica).

scrape_id → ID de la recolección (para control).

last_scraped → Última fecha de scrapeo (categórica/fecha).

calendar_last_scraped → Última fecha de calendario.

9. Extras de texto estructurado (para NLP o clustering)

name → Nombre del anuncio (texto, pero relevante para extracción de features).

description → Descripción del anuncio (texto).

neighborhood_overview → Descripción del barrio (texto).

10. Variables finales complementarias

traffic_fine_amount (si corresponde en dataset, parece numérica para multas u otro proxy).

time_of_day → Hora/día categórico de registro.

Conclusión

Con la selección de estas 50 variables puedo explicar de forma integral los factores que influyen en la oferta de Airbnb en Boston. Las variables abarcan Identificación básica, Datos de localización, Datos del anfitrión, Propiedad y tipo de alojamiento, Precio y condiciones, Disponibilidad del anuncio, Reseñas y reputación, etc; lo que me permite tener una visión completa y una base sólida para futuros análisis más profundos.