

Actividad 1 – Entendimiento de los datos

Tecnológico de Monterrey, Campus Puebla

Plataformas de analítica de negocios para organizaciones

Karyme Pérez Chatú A01174367

Índice

1. Introducción
2. Metodología de selección de variables
3. Selección y justificación de las 50 variables
4. Traducción y descripción

1. Introducción

Este reporte resume el entendimiento de los datos de Airbnb para el dataset asignado.

Incluye la selección de 50 variables clave, su justificación y la metodología usada para priorizarlas, con apoyo del Data Dictionary de Airbnb e investigación individual.

2. Metodología de selección de variables

Se priorizaron variables numéricas (precio, reseñas, disponibilidad, puntuaciones) y categóricas (tipo de propiedad, tipo de habitación, host).

Se descartaron campos de texto libre, URLs, imágenes y licencias porque no aportan valor directo al análisis cuantitativo.

El objetivo es contar con un conjunto balanceado de variables predictoras y descriptivas para modelos de predicción de precios, ingresos y calidad del servicio.

3. Selección y justificación de las 50 variables

- Identificación

id: es básicamente el “folio” único de cada anuncio, sin él sería imposible dar seguimiento.

scrape_id: me sirve para saber de qué extracción salió la información.

host_id: ayuda a identificar a los anfitriones y ver si tienen un solo anuncio o varios.

- Ubicación

latitude y longitude: son esenciales porque permiten mapear la ubicación exacta y analizar proximidad.

neighbourhood_cleansed: muestra el barrio ya estandarizado.

neighbourhood_group_cleansed: agrupa barrios por distritos, lo que facilita un análisis más macro.

property_type: diferencia si es una casa, un depa, un hotel, etc.

room_type: indica si es un alojamiento entero, solo una habitación o algo compartido.

- Anfitrión

host_since: me dice desde cuándo está activo el anfitrión; entre más tiempo, más experiencia.

host_location: muestra dónde dice vivir el anfitrión.

host_response_time: indica qué tan rápido responde.

host_response_rate: refleja qué porcentaje de mensajes contesta.

host_acceptance_rate: porcentaje de solicitudes aceptadas.

host_is_superhost: me parece clave porque un superhost genera confianza inmediata.

host_listings_count y host_total_listings_count: revelan si el anfitrión es individual o más bien alguien con perfil empresarial.

host_has_profile_pic y host_identity_verified: son señales claras de confianza y transparencia.

- Alojamiento

accommodates: número máximo de huéspedes, súper útil para comparar espacios.

bathrooms y bathrooms_text: muestran tanto la cantidad como el tipo de baño.

bedrooms y beds: variables que reflejan el tamaño y comodidad del lugar.

amenities: lista de servicios, desde WiFi hasta cocina o aire acondicionado, que influyen mucho en la decisión de los usuarios.

- Precio y disponibilidad

price: la variable más directa, cuánto cuesta por noche.

minimum_nights y maximum_nights: reglas sobre la estancia.

minimum_minimum_nights, maximum_minimum_nights, minimum_maximum_nights, maximum_maximum_nights: estos valores me sirven para ver la flexibilidad o rigidez del anfitrión.

minimum_nights_avg_ntm y maximum_nights_avg_ntm: reflejan promedios más estables, que ayudan a suavizar los extremos.

has_availability: indica si la propiedad realmente está disponible.

availability_30, availability_60, availability_90, availability_365 y availability_eoy: permiten medir la disponibilidad a diferentes horizontes de tiempo.

- Reseñas y ratings

number_of_reviews: total de reseñas, clave para medir popularidad.

number_of_reviews_ltm, number_of_reviews_l30d y number_of_reviews_ly: muestran la evolución de la actividad reciente.

reviews_per_month: frecuencia promedio de reseñas, indicador de movimiento constante.

review_scores_rating: calificación global, refleja la percepción general.

review_scores_accuracy, review_scores_cleanliness, review_scores_communication y review_scores_location: son subpuntuaciones que ayudan a entender en qué destaca o falla un anuncio.

4. Traducción y descripción

Las variables seleccionadas fueron traducidas al español y descritas con base en el Data Dictionary oficial de Airbnb, y en algunos casos investigación individual.