



**Tecnológico
de Monterrey**

JOSÉ FRANCISCO VERA JIMÉNEZ

ELECCIÓN DE VARIABLES

1. Introducción

El análisis de datos de Airbnb requiere seleccionar un conjunto representativo de variables que permitan comprender la oferta, demanda, calidad del servicio y desempeño económico de los alojamientos. Se seleccionaron **50 variables** priorizando aquellas de tipo **numérico y categórico**, que aportan información relevante para estudiar tendencias, segmentar el mercado y realizar análisis estadísticos.

2. Metodología de Selección

La elección de variables se basó en tres criterios principales:

- **Relevancia analítica:** aquellas que inciden directamente en precios, ocupación y reseñas.
- **Diversidad de tipo de datos:** combinación de variables numéricas (ej. precio, reseñas) y categóricas (ej. tipo de propiedad, barrio).
- **Disponibilidad de información:** se descartaron variables con alto porcentaje de valores nulos o duplicadas en significado.

3. Variables Seleccionadas y Justificación

Identificación y ubicación

- **id:** Permite identificar de manera única cada anuncio.
- **name:** Apoyo para reconocimiento y validación.
- **description:** Fuente de análisis cualitativo y texto.
- **neighbourhood_cleansed:** Ubicación estandarizada, útil para segmentación geográfica.
- **latitude, longitude:** Coordenadas para mapas y análisis espacial.

Tipo de propiedad y características físicas

- **property_type:** Clasifica la oferta (casa, departamento, etc.).
- **room_type:** Determina la modalidad (entero, privado, compartido).
- **accommodates:** Capacidad de huéspedes, clave para precios y segmentación.
- **bathrooms, bedrooms, beds:** Miden infraestructura y comodidad.

- **amenities:** Diferenciador de servicios ofrecidos.

Precios y reglas de estadía

- **price:** Variable central de análisis económico.
- **minimum_nights, maximum_nights:** Reglas de estancia que condicionan la oferta.

Disponibilidad

- **availability_30, availability_60, availability_90, availability_365:** Miden la frecuencia con que un alojamiento está abierto a reservas.

Reseñas y popularidad

- **number_of_reviews, number_of_reviews_ltm, number_of_reviews_l30d:** Indicadores de actividad y confianza.
- **first_review, last_review:** Permiten medir antigüedad y vigencia.
- **review_scores_rating:** Calificación general del alojamiento.
- **review_scores_accuracy, cleanliness, checkin, communication, location, value:** Subdimensiones que explican la calidad percibida.
- **reviews_per_month:** Flujo de retroalimentación continua.

Información del anfitrión

- **host_id, host_name:** Identificación del propietario.
- **host_since:** Antigüedad como anfitrión.
- **host_location:** Ubicación declarada.
- **host_response_time, host_response_rate:** Capacidad de comunicación.
- **host_acceptance_rate:** Disposición a aceptar reservas.
- **host_is_superhost:** Indicador de reputación y servicio superior.
- **host_listings_count, host_total_listings_count:** Escala de operación del anfitrión.
- **host_identity_verified:** Confianza y seguridad.

Reserva y automatización

- **instant_bookable:** Facilita reservas inmediatas, mejora competitividad.

Métricas calculadas de anuncios

- **calculated_host_listings_count, entire_homes, private_rooms, shared_rooms:** Desglose de la oferta del anfitrión.

Indicadores económicos

- **estimated_occupancy_l365d:** Estimación de ocupación anual.
- **estimated_revenue_l365d:** Estimación de ingresos anuales.

4. Conclusión

Las 50 variables seleccionadas permiten un **análisis integral** del mercado de Airbnb, combinando aspectos de:

- **Oferta** (tipo de propiedad, capacidad, amenidades).
- **Demanda** (reseñas, ocupación).
- **Calidad** (calificaciones de huéspedes).
- **Economía** (precio e ingresos).
- **Gestión de anfitriones** (superhost, número de anuncios, tiempos de respuesta).

Con este conjunto es posible desarrollar análisis descriptivos, modelos predictivos y comparativos por zonas, tipos de alojamiento y desempeño de anfitriones.