



# Tecnológico de Monterrey

**Campus Puebla**

Curso:

Plataformas de analítica de negocios para organizaciones (Gpo 101)

Actividad:

Actividad 1 (Entendimiento de los Datos)

Elaborado Por:

Samantha Mayrin Martínez Balbuena A01733837

Fecha:

19/08/2025

## Selección y justificación de variables

### Variables numéricas

1. `id` → Identificador único de la propiedad. Sirve como clave primaria para unir o diferenciar observaciones.
2. `scrape_id` → Marca temporal de la extracción de datos; útil para comparar con diferentes cortes de información.
3. `host_id` → Identificador del anfitrión, necesario para analizar actividad y comportamiento por anfitrión.
4. `host_listings_count` → Número de propiedades activas del host; ayuda a diferenciar entre anfitriones individuales y empresas.
5. `host_total_listings_count` → Total de propiedades listadas por host, incluso inactivas; importante para medir experiencia del anfitrión.
6. `neighbourhood_group_cleansed` → Variable geográfica categorizada; permite agrupar áreas más amplias de la ciudad.
7. `latitude` → Coordenada geográfica esencial para análisis espacial y mapas.
8. `longitude` → Complementa la localización precisa en el mapa, útil para georreferenciación.

9. accommodates → Número de huéspedes que puede alojar; factor clave en segmentación de propiedades.
10. bathrooms → Número de baños, indicador fuerte de comodidad y precio.
11. bedrooms → Número de habitaciones; directamente relacionado con capacidad y tarifas.
12. beds → Número de camas, ayuda a medir la capacidad real de hospedaje.
13. minimum\_nights → Restricciones de estadía mínima; útil para analizar políticas de reserva.
14. maximum\_nights → Límite de estadía máxima, también refleja estrategia del anfitrión.
15. minimum\_minimum\_nights → Valor mínimo histórico para estadías cortas; permite ver flexibilidad del host.
16. maximum\_minimum\_nights → Restricciones más estrictas históricas en estadías mínimas.
17. minimum\_maximum\_nights → Restricciones mínimas sobre la estadía máxima.
18. maximum\_maximum\_nights → Restricciones máximas sobre la estadía máxima.

19. `minimum_nights_avg_ntm` → Promedio de noches mínimas, ajustado por calendario.

Métrica refinada de política de reservas.

20. `maximum_nights_avg_ntm` → Promedio de noches máximas. Útil para analizar comportamiento de largo plazo.

21. `availability_30` → Días disponibles en los próximos 30 días. Indicador de ocupación a corto plazo.

22. `availability_60` → Disponibilidad en 60 días, permite ver planificación a mediano plazo.

23. `availability_90` → Disponibilidad en 90 días, útil para predicciones de temporada alta.

24. `availability_365` → Disponibilidad en un año; mide constancia de la propiedad en la plataforma.

25. `number_of_reviews` → Cantidad total de reseñas, indicador de popularidad y confianza.

26. `number_of_reviews_ltm` → Número de reseñas en los últimos 12 meses; mide actividad reciente.

27. `number_of_reviews_l30d` → Reseñas en los últimos 30 días; refleja demanda actual.

28. `availability_eoy` → Disponibilidad al final del año, útil para medir cierres por temporada.
29. `number_of_reviews_ly` → Número de reseñas en el último año calendario.  
Complementa la variable anterior.
30. `estimated_occupancy_l365d` → Tasa estimada de ocupación en 365 días. Clave para análisis de performance.
31. `estimated_revenue_l365d` → Ingreso estimado anual. Una de las métricas principales para análisis financiero.
32. `review_scores_rating` → Calificación promedio global. Factor decisivo en percepción del cliente.
33. `review_scores_accuracy` → Evalúa si la descripción corresponde con la realidad.  
Impacta en la confianza.
34. `review_scores_cleanliness` → Calificación de limpieza. Fundamental en hospitalidad.
35. `review_scores_checkin` → Puntaje de proceso de entrada. Influye en satisfacción inicial del cliente.
36. `review_scores_communication` → Evalúa la comunicación host-huésped. Clave en experiencia del usuario.

37. review\_scores\_location → Puntaje de ubicación; importante para precios y elección del huésped.

38. review\_scores\_value → Relación valor-precio según el huésped. Afecta la competitividad del listing.

39. calculated\_host\_listings\_count → Número de propiedades activas del host calculado por Airbnb.

40. calculated\_host\_listings\_count\_entire\_homes → Número de casas/apartamentos completos listados por el host.

41. calculated\_host\_listings\_count\_private\_rooms → Cantidad de cuartos privados listados.

42. calculated\_host\_listings\_count\_shared\_rooms → Cantidad de cuartos compartidos listados.

43. reviews\_per\_month → Promedio mensual de reseñas; aproxima la frecuencia de reservas.

## **Variables categóricas**

44. listing\_url → URL de la propiedad. Identificador externo útil para verificar registros.

45. last\_scraped → Fecha de última extracción; asegura vigencia de la información.

46. source → Fuente de los datos; importante si se comparan datasets distintos.

47. name → Nombre del anuncio, contiene keywords que pueden usarse en análisis de texto (NLP).

48. description → Descripción del host, útil para análisis semántico y extracción de insights.

49. neighborhood\_overview → Texto descriptivo del vecindario, clave para marketing y segmentación geográfica.

50. host\_name → Nombre del anfitrión. Útil para explorar la repetición de nombres, detectar anfitriones con múltiples cuentas y enriquecer el análisis social/demográfico del mercado.