



Tecnológico de Monterrey
Escuela de Negocios

Instituto Tecnológico y de Estudios Superiores de Monterrey

Escuela de Negocios

Campus Puebla

Gestión de proyectos de plataformas tecnológicas (Gpo 201)

Actividad 1 (Regresión Lineal Simple y Múltiple)

Isaac Miguel Barrón Portillo

A01737199

Docente: PhD Alfredo García

Fecha: 4 de octubre de 2025

Análisis diagnóstico Madrid

Análisis por Room Type

Entire Home/Apt

Palanca primaria: elevar `host_response_rate`; el $r \approx 0.68$ con aceptación indica que SLAs de respuesta ($\leq 1h$) pueden traducirse en más reservas y volumen de reseñas ($r \approx 0.17$).

Gestión de inventario: la correlación negativa ($r \approx -0.06$) entre disponibilidad anual y reseñas sugiere que los listados con buena conversión pasan más días ocupados; conviene pricing dinámico y mínimos de noches para capturar demanda pico sin erosionar ADR.

Precio apenas se asocia con aceptación ($r \approx 0.09$): el valor percibido se mueve más por tiempo de respuesta, limpieza y ubicación que por tarifa.

Private Room

Relación operativa fuerte: $r \approx 0.53$ entre respuesta y aceptación; la disciplina en mensajería y preaprobaciones impulsa conversión y reseñas ($r \approx 0.21$).

Precio con aceptación es muy débil ($r \approx 0.07$): competir por rapidez, claridad de reglas y seguridad más que por tarifa pura.

Estrategia: fences por ocupación (1–2 vs. 3–5), check-in flexible y transparencia en normas para capitalizar el perfil de viajero solo/pareja.

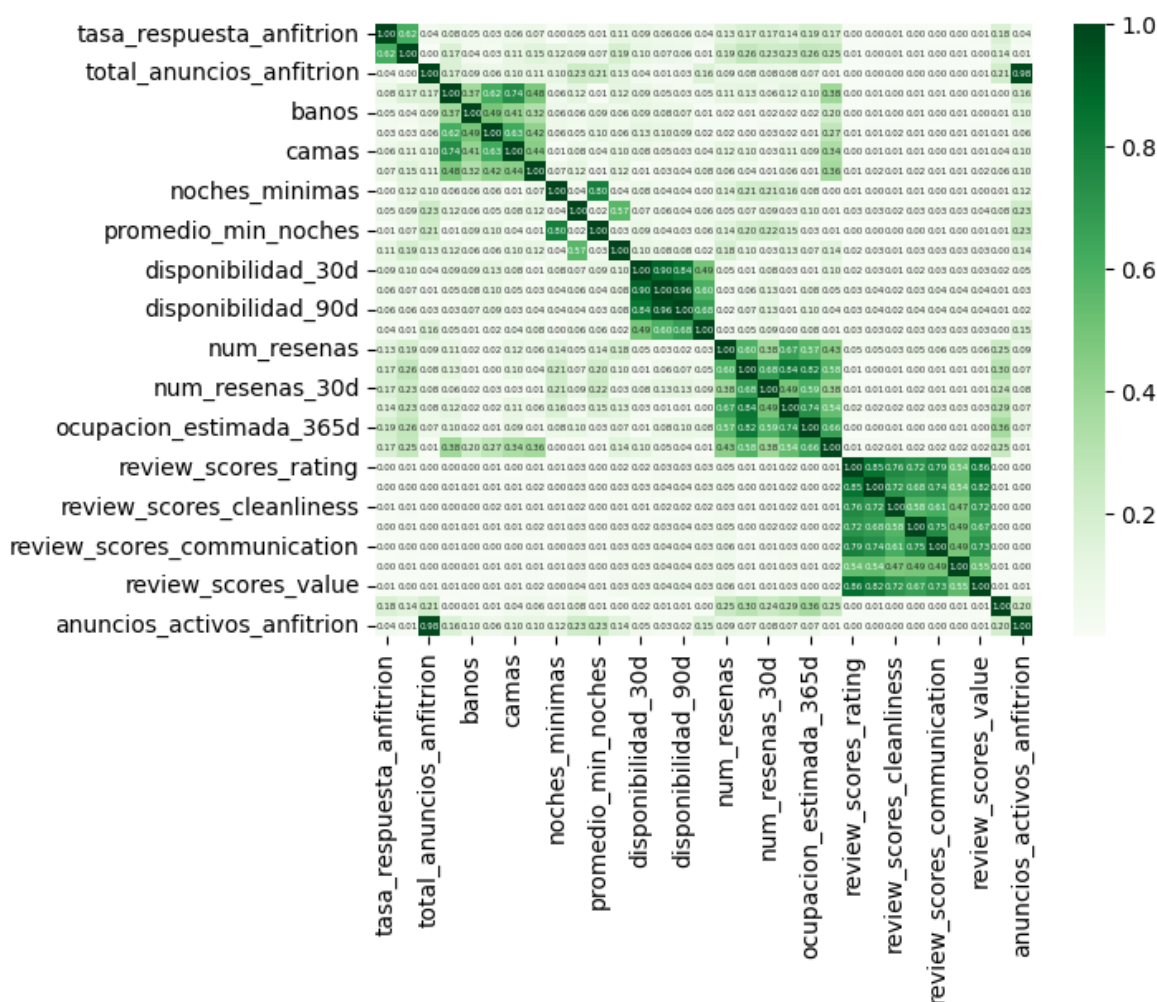
Hotel Room

Operativo: respuesta–aceptación ($r \approx 0.37$), aceptación–precio ($r \approx 0.20$) y disponibilidad–reseñas ($r \approx 0.36$). Señala un circuito donde operación consistente \rightarrow aceptación alta \rightarrow mayor producción \rightarrow capacidad para sostener precio.

Advertencia de calidad: `review_scores_rating` vs `total_listings` ($r \approx -0.21$) sugiere que expandir cartera sin control puede degradar puntuaciones; priorizar estándares y auditorías por propiedad.

Shared Room

Correlaciones muy débiles en todas las combinaciones (respuesta–aceptación $r \approx 0.06$, aceptación–precio $r \approx -0.10$): el desempeño depende de otras palancas (ubicación ultra céntrica, comunidad, reglas de convivencia). Enfocar pruebas A/B en amenities básicos, seguridad y normas claras.



Lectura del heatmap (agrupamientos y alertas de modelado). Se ven clusters claros: (a) demanda/producción: *num_resenas* con *num_resenas_30d* y ocupación estimada; (b) calendario: *disponibilidad_30d–90d* asociadas entre sí y con ocupación en sentido opuesto según segmento; (c) tamaño del alojamiento: *camas*, *baños* y *capacidad* correlacionan moderadamente; (d) estancia: *noches_minimas* con *promedio_min_noches*. En contraste, *host_response_rate* y los *review_scores* tienen correlaciones bajas con el resto, por lo que funcionan como señales independientes de experiencia/operación. Para decisiones: i) enfocar palancas en respuesta (Entire/Private) y calidad + aceptación (Hotel) evitar combinar en un

mismo modelo variables muy colineales del mismo cluster (p. ej., *num_resenas* vs *num_resenas_30d* o *noches_mínimas* vs *promedio_min_noches*), y iii) diseñar pricing y mínimos de noches a partir del binomio disponibilidad–reseñas/ocupación por zona y temporada.

Regresión lineal múltiple

review_scores_rating ~ value + accuracy + cleanliness ($R^2=0.8206$). La calificación global se explica muy bien por tres palancas: accuracy ($\beta \approx 0.392$) y value ($\beta \approx 0.389$) son casi equivalentes y dominan la percepción final; cleanliness ($\beta \approx 0.210$) también suma de forma significativa. Implicación: actualizar descripciones/fotos y gestionar expectativas (accuracy) junto con pricing justo y extras visibles (value) eleva el rating más rápido que invertir solo en limpieza; sin descuidarla, las primeras dos generan el mayor retorno marginal.

tasa_aceptacion_anfitrión ~ tasa_respuesta_anfitrión ($R^2=0.3803$). Cada punto de respuesta se asocia con ≈ 0.83 puntos de aceptación y existe un nivel base $\approx 9.7\%$; es una palanca operativa clara. Establecer SLA de respuesta ≤ 1 hora, plantillas de mensajes y alertas móviles debería traducirse en más solicitudes aceptadas y, por arrastre, mayor conversión a reserva.

es_superanfitrión ~ respuesta + aceptación + rating + ingresos ($R^2=0.0809$). El estatus de superhost apenas se explica por estas variables: es un outcome discreto condicionado por reglas de Airbnb (cancelaciones, cumplimiento, periodos de evaluación) no capturadas aquí. Recomendación: no usar este modelo para decisión; priorizar features de proceso (cancelaciones, *on-time response*, disputas) y modelado no lineal/árboles.

total_anuncios_anfitrión ~ disp_365d + reseñas + ingresos ($R^2=0.0325$). El tamaño de cartera del anfitrión es estructural y casi independiente del desempeño del listing: disponibilidad suma levemente ($\beta \approx 0.089$), reseñas restan marginalmente ($\beta \approx -0.111$, proxy de ocupación), e ingresos aportan poco. Implicación: segmentar hosts por arquetipo (profesionales vs. individuales) en lugar de intentar predecir su cartera con métricas de un anuncio.

capacidad ~ camas + habitaciones ($R^2=0.5899$). La capacidad se ajusta bien con inventario: camas ($\beta \approx 0.882$) pesa más que habitaciones ($\beta \approx 0.534$). Útil para validación automática de

fichas (detección de outliers de capacidad) y para construir reglas de precio por huésped adicional consistentes con el equipamiento real.

habitaciones ~ camas + baños + capacidad ($R^2=0.4974$). El número de baños ($\beta \approx 0.406$) es el predictor más informativo de dormitorios, seguido de camas ($\beta \approx 0.213$) y capacidad ($\beta \approx 0.136$); refleja que tipologías más amplias suelen incorporar más baños. Sirve para sanity checks de inventario y para planificar *bundles* (p. ej., familias) donde baños es un driver de conversión.

precio ~ camas + habitaciones + capacidad ($R^2=0.2612$). El tamaño explica parte del precio ($\beta_{\text{habitaciones}} \approx 21.0$ el más alto), pero la varianza es grande: faltan ubicación, temporada, amenities premium y calidad. Usar este modelo como piso base y superponer capas de pricing dinámico (barrio/fecha/evento) para capturar disposición a pagar.

review_scores_value ~ accuracy + cleanliness ($R^2=0.7080$). La percepción de valor está anclada en accuracy ($\beta \approx 0.717$) y reforzada por cleanliness ($\beta \approx 0.301$). Prioridades: descripciones precisas, fotos actuales y equipamiento realmente disponible; después, estándares de limpieza repetibles. Subir *value* tiende a mejorar el rating global (ver primer modelo).

baños ~ habitaciones + camas ($R^2=0.2536$). Los baños crecen con habitaciones ($\beta \approx 0.235$) más que con camas ($\beta \approx 0.072$); la baja R^2 sugiere que depende de antigüedad/tipo de propiedad y normas de reforma. Úsese para detección de inconsistencias (p. ej., muchas habitaciones con muy pocos baños).

num_resenas_30d ~ num_resenas_12m + ocupación_365d + num_resenas_1y ($R^2=0.4836$). La reciente generación de reseñas aumenta con el histórico 12m ($\beta \approx 0.0746$) y la ocupación estimada ($\beta \approx 0.0030$), mientras num_resenas_1y entra negativo ($\beta \approx -0.0271$) por solapamiento temporal/colinealidad. Insight: monitorear reseñas 12m + ocupación como *nowcast* de tracción; simplificar el feature set para evitar redundancias y ganar estabilidad.