



# Tecnológico de Monterrey

Campus Puebla

José Manuel Flores Pérez | A01733838

**Actividad 2.1:** Reporte

Alfredo García Suárez

6 de abril de 2025

El análisis comenzó con la limpieza de datos y la conversión de variables a formato numérico, seguido de un estudio detallado de la correlación entre pares específicos de variables, tales como `host_acceptance_rate` vs `host_response_rate`, `price` vs `accommodates` y `review_scores_location` vs `review_scores_cleanliness`, evaluadas en cuatro tipos de habitación (Entire home/apt, Private room, Shared room, Hotel room), para cada par, se generaron gráficas de dispersión con línea de regresión y se calculó el coeficiente de correlación para observar la fuerza de la relación.

Posteriormente, el código identificó las combinaciones con mayor correlación y construyó modelos de regresión lineal simple para estas, mostrando los valores de  $R^2$  como medida de ajuste, los mejores resultados se registraron en relaciones como `bedrooms` vs `accommodates`, `review_scores_value` vs `review_scores_cleanliness` y `price` vs `accommodates`, algunas con  $R^2$  por encima de 0.5, lo cual indica una relación lineal moderadamente fuerte; también se elaboró una tabla con los 10 modelos más sólidos seleccionados manualmente, sin utilizar ciclos, resaltando aquellos con mayor capacidad predictiva.

Finalmente, se realizaron modelos de regresión lineal múltiple para diversas variables objetivo (`host_acceptance_rate`, `price`, `bedrooms`, `reviews_per_month`, entre otras), utilizando las variables restantes como predictores, estos modelos permitieron observar cómo múltiples factores explican una sola variable, con valores de  $R^2$  que oscilan entre 0,3 y 0,6, según la situación, en general, el código demostró cómo aplicar herramientas estadísticas básicas para comprender a los clientes en los datos hoteleros, siguiendo una estructura clara y sin bucles, con un código desarrollado paso a paso para facilitar su interpretación.