



Tecnológico  
de Monterrey

*Reporte Regresiones Logísticas*

*René*

*Enrique*

*Patricio*

Este reporte expone los hallazgos derivados del estudio de correlación logística entre varias variables categóricas de una base de datos empleando el método de "Regresión Logística". El propósito de este análisis es detectar vínculos relevantes entre variables categóricas, lo que facilitará una mejor comprensión de su comportamiento en conjunto y su impacto en los resultados de interés.

Adicionalmente, se ha realizado la transformación de cinco variables numéricas en variables dicotómicas, utilizando el instrumento de "Intervalos por categorías". Este procedimiento ha sido imprescindible para llevar a cabo un análisis más exacto y consistente, ya que ciertas variables numéricas necesitaban ser modificadas para poder ser evaluadas a través de la regresión logística. Las categorías escogidas para cada variable transformada se establecieron en función de su distribución e importancia en el marco del análisis.

Este informe presenta tanto los hallazgos de las correlaciones logísticas de las variables categóricas iniciales como las derivadas de la transformación de las variables numéricas, resaltando las relaciones más relevantes y su potencial efecto en la interpretación global de los datos.

Dentro de la siguiente tabla se expondrán los valores obtenidos de nuestros análisis y posteriormente una interpretación de los mismos:

México:

	Coeficiente de Precisión	Coeficiente de Exactitud	Coeficiente de Sensibilidad
has_availability	0.99	0.99	0.99
host_is_superhost	0.56	0.59	0.59
host_is_verified	0.77	0.79	0.79
host_has_profile_pic	0.96	0.98	0.98
instant_bookable	0.57	0.61	0.61
price	0.98	0.98	1

accommodates	0.95	0.92	0.96
beds	0.92	0.91	0.99
bathrooms	0.92	0.91	0.98
host_total_listings_count	0.97	0.97	1

Hong Kong:

	Coeficiente de Precisión	Coeficiente de Exactitud	Coeficiente de Sensibilidad
has_availability	0.99	0.99	0.99
host_is_superhost	0.74	0.86	0.86
host_is_verified	0.70	0.74	0.74
host_has_profile_pic	0.94	0.97	0.97
instant_bookable	0.63	0.79	0.79
price	0.99	0.99	1
accommodates	0.95	0.93	0.97
beds	0.93	0.93	1
bathrooms	0.99	0.99	1
host_total_listings_count	0.79	0.79	1

En el caso de nuestra base de datos de Tokyo solo pudimos obtener algunos coeficientes pero no se pudo realizar un análisis completo ya que la base de datos no presenta los datos necesarios en ciertas categorías, teniendo solo datos únicos imposibilitando el uso o creación de variables dicotómicas.

Ahora para la interpretación de los resultados:

## *México*

has\_availability (Disponibilidad): Los coeficientes son altos (0.99 en todas las métricas), lo que indica que el modelo predice de manera casi perfecta la disponibilidad de propiedades. Esta variable tiene un impacto muy fuerte y consistente.

host\_is\_superhost (Superhost): Con coeficientes alrededor de 0.56-0.59, esta variable tiene un rendimiento moderadamente bajo, sugiriendo que el modelo no es tan preciso para predecir si un anfitrión es Superhost en México.

host\_is\_verified (Anfitrión verificado): Este coeficiente (0.77-0.79) muestra un rendimiento moderado, con una precisión y exactitud algo más alta que en el caso de Superhost, pero aún lejos de las variables más influyentes.

host\_has\_profile\_pic (Tiene foto de perfil): Coeficientes altos (0.96-0.98), lo que indica que esta variable está bien representada por el modelo, siendo una característica fácilmente predecible.

instant\_bookable (Reservas inmediatas): Con coeficientes alrededor de 0.57-0.61, la capacidad del modelo para predecir si una propiedad es reservable de inmediato es baja, similar al caso de Superhost, mostrando que esta variable no está claramente diferenciada por el modelo.

price (Precio): Coeficientes muy altos (0.98-1), lo que sugiere que el modelo predice con gran precisión el precio de las propiedades, reflejando que el precio es una variable crucial y bien representada.

accommodates (Capacidad de alojamiento): Esta variable tiene coeficientes relativamente altos (0.92-0.96), lo que significa que el modelo predice con precisión cuántas personas puede alojar una propiedad.

beds (Camas): Coeficientes de 0.91-0.99, lo que refleja un rendimiento fuerte, especialmente en sensibilidad (0.99), lo que indica que el modelo puede predecir correctamente la cantidad de camas en la mayoría de los casos.

bathrooms (Baños): Los coeficientes (0.91-0.98) son similares a los de camas, lo que indica un buen rendimiento en la predicción del número de baños en las propiedades.

host\_total\_listings\_count (Número total de listados del anfitrión): Con coeficientes altos (0.97-1), el modelo predice con precisión el número total de listados de los anfitriones, siendo una variable muy bien representada.

### ***Hong Kong***

has\_availability (Disponibilidad): Similar a México, esta variable tiene coeficientes perfectos (0.99), mostrando que el modelo predice con gran precisión la disponibilidad de propiedades en Hong Kong.

host\_is\_superhost (Superhost): En Hong Kong, esta variable tiene un rendimiento significativamente mejor (0.74-0.86) que en México, lo que sugiere que en este mercado el modelo tiene más éxito en la predicción de Superhosts.

host\_is\_verified (Anfitrión verificado): Con coeficientes de 0.70-0.74, esta variable presenta un rendimiento moderado en Hong Kong, muy similar al desempeño observado en México.

host\_has\_profile\_pic (Tiene foto de perfil): Coeficientes altos (0.94-0.97), lo que indica un excelente rendimiento, igual que en México, para predecir si los anfitriones tienen fotos de perfil.

instant\_bookable (Reservas inmediatas): Con coeficientes de 0.63-0.79, esta variable tiene un rendimiento más alto en Hong Kong que en México, lo que sugiere que el modelo predice mejor esta característica en el contexto de Hong Kong.

price (Precio): Al igual que en México, los coeficientes (0.99-1) son extremadamente altos, lo que indica que el precio es una variable clave en ambos mercados.

accommodates (Capacidad de alojamiento): Similar a México, los coeficientes (0.93-0.97) son altos, lo que indica que el modelo predice de manera precisa la capacidad de alojamiento en las propiedades de Hong Kong.

beds (Camas): Coeficientes elevados (0.93-1), lo que indica que el modelo predice con precisión la cantidad de camas en las propiedades de Hong Kong.

bathrooms (Baños): En Hong Kong, los coeficientes de 0.99-1 muestran un rendimiento perfecto, lo que indica que el modelo predice con precisión el número de baños.

host\_total\_listings\_count (Número total de listados del anfitrión): En este caso, los coeficientes son más bajos (0.79-1) en comparación con México, lo que sugiere que el modelo tiene menos éxito al predecir el número total de listados en este mercado.

En conclusión, las variables de disponibilidad y precio son las más consistentemente predecibles en ambos mercados, con coeficientes cercanos a 1 en todas las métricas, lo que refleja un alto rendimiento del modelo. Sin embargo, variables como `host_is_superhost` e `instant_bookable` presentan un mejor desempeño en Hong Kong que en México, lo que sugiere que en ese mercado estas características son más relevantes o distinguibles. Por otro lado, el número total de listados del anfitrión es mucho más predecible en México que en Hong Kong, lo que podría estar relacionado con diferencias en la estructura de los anfitriones en cada región. Finalmente, la variable `host_is_verified` muestra un rendimiento moderado en ambos mercados, indicando que la verificación del anfitrión no es un fuerte predictor en ninguno de los dos contextos.