



Tecnológico de Monterrey

Actividad 3 Regresión Logística

Raymundo Díaz Tijera A01664497

15 oct 2025

Plataformas de analítica de negocios para organizaciones

Profesor: Maria Luisa Gomez Barrios

Co-Titular: Alfredo García Suárez

Co-Titular: Elena García y García

Tabla de coeficientes

Caso	Var. Dependiente	Var. Independientes con su correlación	Coef. Precisión	Coef. Exactitud	Coef. Sensibilidad
1	host_is_superhost	host_response_rate (0.30) number_of_reviews_ltm (0.44) estimated_occupancy_l365d (0.51)	'f' = 0.86 't' = 0.60	0.81	'f' = 0.89 't' = 0.52
2	host_has_profile_pic - desbalanceo	availability_30 (-0.11) review_scores_value (0.13) review_scores_checkin (0.13)	'f' = 0.05 't' = 0.98	0.69	'f' = 0.59 't' = 0.69
3	host_identity_verified - desbalanceo	availability_365 (0.17) maximum_minimum_nights (0.21) estimated_occupancy_l365d (0.16)	'f' = 0.20 't' = 0.95	0.71	'f' = 0.72 't' = 0.71
4	has_availability - desbalanceo	minimum_nights (-0.26) availability_365 (0.25) host_identity_verified_num (0.69)	'f' = 0.20 't' = 0.96	0.72	'f' = 0.72 't' = 0.72
5	instant_bookable	host_acceptance_rate (0.34) availability_365 (0.25) estimated_revenue_l365d (0.20)	'f' = 0.82 't' = 0.69	0.78	'f' = 0.88 't' = 0.53
6	host_response_rate_cat - desbalanceo	host_acceptance_rate (0.33) reviews_per_month (0.24) estimated_revenue_l365d (0.19)	'Good response' = 0.99 'Bad response' = 0.14	0.87	'Good response' = 0.87 'Bad response' = 0.80
7	review_scores_rating_cat	reviews_per_month (0.37) estimated_occupancy_l365d (0.43) minimum_nights_avg_ntm (-0.40)	'Good reviews' = 0.64 'Bad reviews' = 0.68	0.65	'Good reviews' = 0.74 'Bad reviews' = 0.54
8	availability_365_cat	availability_eoy (0.90) reviews_per_month (0.18) calculated_host_listings_count (0.23)	'Great av.' = 0.82 'Low av.' = 0.89	0.86	'Great av.' = 0.81 'Low av.' = 0.90
9	price_cat	accommodates (0.29) bedrooms (0.25) estimated_revenue_l365d (0.23)	'High price' = 0.63 'Low price' = 0.72	0.70	'High price.' = 0.36 'Low price' = 0.88
10	beds_cat	accommodates (0.55) bathrooms (0.32) price (0.25)	'Few beds' = 0.93 'Many beds' = 0.85	0.92	'Few beds' = 0.98 'Many beds' = 0.64

Hallazgos

- Ajuste de balanceo / submuestreo para los casos: 3,4 y 6 (variable dicotómica creada), todos presentaron un coeficiente de precisión igual o menor de 0.14 a 0.20 para una categoría y mayor al 0.95 para la otra categoría, que podría ser causado este desbalance original, debido al resultado para el conteo para cada categoría total.

host_has_profile_pic (peor caso después del balanceo)	host_identity_verified	has_availability	host_response_rate_cat
# t 13776 # f 411	t 12900 f 1287	# t 13389 # f 798	# Good response 13852 # Bad response 335

Este balanceo fue fundamental para mejorar los coeficientes de precisión y sensibilidad en cada caso identificado, a excepción del caso 2, siendo host_has_profile_pic el peor caso, con un coeficiente de precisión del 0.05 y 0.98 correspondiente, que podría deberse a la gran diferencia entre la clase mayoritaria y clase minoritaria.

Antes del balanceo, los modelos tendían a predecir siempre la clase mayoritaria 't' y 'Good response' tomando en cuenta que el 70% de los datos son para entrenamiento y el 30% para las pruebas.

- Las variables con correlaciones moderadas (0.20 a 0.55) fueron las variables con el mayor coeficiente de exactitud, estando arriba de 0.80, lo que significa modelos más estables, en donde la multicolinealidad de las variables no afecta y no afecta a los coeficientes de precisión y sensibilidad.
- Los casos con mejores coeficientes de exactitud y sensibilidad son casos donde están presentes métricas operativas de los host, a excepción del caso 10, con métricas de características del listing/propiedad.

Casos divididos por Confiables / Casos con área de mejora, en base a sus coeficientes: En general todos los casos presentan un comportamiento conservador, pues prefieren marcar

falsos positivos sobre falsos negativos, lo que podría ser de utilidad si Airbnb prefiere evitar etiquetar algo malo que en realidad era bueno, a menos que sea el caso tenga características que puedan afirmar con mucha confiabilidad el negativo.

- **Confiables**

1 host_is_superhost

Hallazgo: La ocupación alta y la experiencia del anfitrión (número de reseñas) son buenos indicadores para ser “superhost”.

Coefficientes: precisión y sensibilidad equilibradas (0.81 y 0.86), y es mejor prediciendo ‘f’ que ‘t’.

3 host_identity_verified (desbalanceo)

Hallazgo: el balanceo mejoró drásticamente la sensibilidad (0.71) y la precisión se estabilizó.

La disponibilidad anual (availability_365) y las noches mínimas están moderadamente relacionadas con la verificación de identidad.

4 has_availability (desbalanceo)

Variable más influyente: host_identity_verified_num (0.69).

Hallazgo: los hosts verificados tienden a mantener sus anuncios activos con mayor disponibilidad; el modelo está bien calibrado (0.72 en todas las métricas).

6 host_response_rate_cat (desbalanceo)

Hallazgo: tras balanceo, el modelo predice casi perfectamente las respuestas ‘Good response’ (precisión 0.99, sensibilidad 0.87).

Hallazgo: la tasa de aceptación y los ingresos estimados son indicadores sólidos de eficiencia en la comunicación del anfitrión.

8 availability_365_cat

Hallazgo: fuerte correlación entre availability_eoy (0.90) y la disponibilidad anual predicha.

Un modelo muy balanceado y preciso (exactitud 0.86, sensibilidad 0.81–0.90).

Indica relación entre la disponibilidad futura y la actual del host.

10 beds_cat

Hallazgo: El modelo con los coeficientes más altos (exactitud 0.92, sensibilidad 0.98 para “Few beds”).

Por lo que las variables físicas (accommodates, bathrooms, price) son muy predictivas, con correlaciones altas (0.55–0.32).

Casos con área de mejora

2 host_has_profile_pic (desbalanceo)

Variables débiles: correlaciones bajas (menor 0.15), indicando baja capacidad predictiva y explicativa, siendo las correlaciones más altas identificadas en la tabla de correlaciones.

Hallazgo: El modelo logra alta sensibilidad (0.69), pero la precisión para clase ‘f’ es demasiado baja (0.05), por lo que presenta overfitting a la clase ‘t’.

5 instant_bookable

Hallazgo: los anfitriones con alta aceptación y disponibilidad tienden a activar reservas instantáneas.

Precisión alta para ‘f’ (0.82) pero baja para ‘t’ (0.69), presente el modelo conservador, pues penaliza más a la automatización.

7 review_scores_rating_cat

Hallazgo: el modelo logra buena sensibilidad (0.65) con correlaciones consistentes (entre 0.37–0.43), con coeficientes por debajo del 0.70.

Los hosts con más reseñas mensuales y ocupación elevada obtienen mejores calificaciones; sin embargo, noches mínimas largas afectan negativamente.

Al definir 'Good reviews' como clase positiva, el modelo muestra un comportamiento más optimista, pues tiende a clasificar muchas reseñas como buenas, lo que aumenta la sensibilidad pero reduce la precisión, por lo que prefiere equivocarse marcando reseñas malas como buenas, FP igual a 955, antes que dejar pasar reseñas buenas sin reconocerlas, FN igual a 526, de la matriz de confusión.

9 price_cat

Hallazgo: aunque la exactitud global es aceptable (0.70), la sensibilidad para precios altos es baja (0.36).

Se evidencia un sesgo hacia precios bajos, posiblemente por su frecuencia mayoritaria en el dataset.

Las variables de tamaño (accommodates, bedrooms) y rentabilidad influyen, pero requieren mejor balanceo.

Indica que la capacidad total está fuertemente alineada con el número de camas ofrecidas.