



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN

INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

BALLINAS BALLINAS EDUARDO

FLORES OVANDO CHRISTIAN

GARCIA FLORES ANA KAREN

INGENIERÍA DE SOFTWARE 1

REGRESIÓN LOGISTICA

31/03/2025

## TABLA GENERAL MEX vs POR vs SPA vs GER

Métrica	instant_bookable				source				host_is_superhost				has_availability				neighbourhood_group_cleansed			
Precisión T o >	0.548122	0.607281	0.596633	0.77411	0.897589	0.879729	0.622801	1	0.366961	0.39857	0.411364	0.7796	0.99837	0.997246	0.977425	0.998097	0.593409	0.982375	0.687888	0.99935
Precisión F o <	0.625298	0.608762	0.814074	0.65796	0.386973	0.75	0.794908	0.998059	0.613711	0.743285	0.758246	0	0	0	0	0	0.669231	0.979286	0.355212	0.41312
Sensibilidad T o >	0.148916	0.876726	0.779244	0.95316	0.977414	0.999216	0.753191	1	0.19082	0.220465	0.238158	1	1	1	1	1	0.63718	0.990509	0.913873	0.822899
Sensibilidad F o <	0.920438	0.252793	0.647232	0.2447	0.113356	0.01693	0.677108	0.998873	0.796142	0.87154	0.875181	0	0	0	0	0	0.627063	0.961905	0.102679	0.995726
Exactitud	0.617053	0.607546	0.700176	0.76248	0.880878	0.879372	0.708642	0.999286	0.564639	0.690168	0.704409	0.7796	0.99837	0.997246	0.977425	0.998097	0.631724	0.98141	0.657496	0.84213
Métrica	room_type				accommodates				reviews_per_month				review_scores_rating				price			
Precisión T o >	0.893654	0.841621	0.703512	0.48253	0.815201	0.864857	0.875519	0.702494	0.894422	0.878768	0.549669	0.656	0.912368	0.868898	0.975081	0.656378	0.595258	0.725334	0.692699	0.711202
Precisión F o <	0.577518	0.766486	0.930905	0.87366	0.89227	0.790727	0.785276	0.830559	0.87783	0.858432	0.885906	0.853	0.863584	0.923761	0.681818	0.852956	0.657153	0.657429	0.794206	0.704657
Sensibilidad T o >	0.684061	0.943506	0.840741	0.83977	0.793328	0.776334	0.75089	0.87366	0.852223	0.818182	0.364835	0.368	0.846629	0.859758	0.992282	0.36836	0.428315	0.751031	0.705272	0.719708
Sensibilidad F o <	0.841405	0.510858	0.858272	0.55164	0.904926	0.874475	0.895105	0.625997	0.913459	0.907129	0.942857	0.95	0.922719	0.929073	0.394737	0.95	0.79003	0.62687	0.784267	0.695887
Exactitud	0.737429	0.828284	0.853263	0.64741	0.866332	0.824566	0.823633	0.750496	0.885141	0.867116	0.850088	0.83	0.885643	0.904572	0.968254	0.830242	0.638495	0.697329	0.752028	0.708035

## CONCLUSIONES

- Para *instant\_bookable*, la precisión T varía entre 0.548 y 0.774, lo que sugiere diferencias en el desempeño del modelo en función del país.
- En *source*, la exactitud de la prueba es alta, mostrando que la predicción asimila a los datos que son reales.
- En *host\_is\_superhost*, la sensibilidad tiene valores más altos en los valores *false*, lo que nos indica que es más fácil identificar cuándo alguien *no es un superhost* a cuando si lo es.
- En *has\_availability*, la mayoría de los valores de precisión y sensibilidad son 1 o cercanos a 1, lo que indica que el modelo tiene un buen rendimiento para predecir el valor de verdad de la disponibilidad de las propiedades, y esto se debe a que la mayoría de los datos de esta variable son de valor *true*.
- *Neighbourhood\_group\_cleansed* es una variable dependiente que depende su funcionalidad del tamaño de la ciudad, ya que, si su territorio es grande, será más difícil de predecir con las variables independientes latitud y longitud.
- La exactitud general es alta en la mayoría de las métricas, con valores por encima de 0.6 en todas las categorías. Esto significa que el modelo está haciendo predicciones razonablemente correctas en la mayoría de los casos.

- El modelo tiene más dificultades en predecir si un anfitrión es superhost y si los alojamientos tienen reservación instantánea, quizás debido a sus pocas variables independientes en el caso de la primera, o a sus pocas relaciones con otras variables en el caso de la segunda.

De manera general, podemos concluir que estos datos nos podrán ir mostrando el comportamiento de algunas variables dicotómicas que poseamos en nuestro dataframe, lo que nos abre un nuevo panorama de análisis y una oportunidad de ideas. La regresión logística, aunque pueda parecer poco intuitiva, fue muy útil en variables como *accommodates*, *review\_scores\_rating* y *neighbourhood\_group\_cleansed*. Esto puede ser una visión general, pero todavía nos espera un largo análisis final para estos datos.

### **Referencias del modelo:**

- Cada columna representa una variable en la predicción de algún modelo de clasificación sobre reservas de alojamiento en Airbnb.
- Los valores indican cómo se desempeña el modelo en diferentes aspectos, y las ciudades representan diferentes mercados.