



## Actividad 4.1: Regresion Logistica

Carlos Fonseca Olivares

A01734538

Profesor: Alfredo Garcia Suarez

Tecnológico de Monterrey; Campus Puebla

26/04/2025

En este trabajo se hizo un análisis utilizando **regresión logística** para entender cómo se relacionan diferentes características de alojamientos en Belice. Se usó la base de datos llamada Belize\_datoslimpios.csv.

Se revisaron 10 combinaciones diferentes de variables para ver si había una relación importante entre ellas.

## **Procedimiento**

### **Limpieza de datos**

Se abrió la base de datos y se quitaron los registros que tenían espacios vacíos.

La columna price tenía precios escritos con símbolos de dólar, así que se limpiaron y se convirtieron en números normales.

Se crearon nuevas columnas donde los datos son 0 o 1. Estas columnas ayudan a que el modelo entienda mejor los datos. Las nuevas columnas fueron:

- host\_is\_superhost\_bin
- instant\_bookable\_bin
- room\_type\_bin
- superhost\_response\_bin

### **Modelado de los casos**

Para cada uno de los 10 casos, se eligieron las variables que se iban a analizar (independientes y dependientes). Se separaron los datos, una parte para entrenar el modelo y otra para probar qué tan bien funcionaba. De igual forma, aplicó un escalado para que los valores estén en un rango similar.

Se creó un modelo de regresión logística con LogisticRegression. Por consiguiente, se entrenó el modelo con los datos de entrenamiento y luego se hicieron predicciones con los datos de prueba.

### **Evaluación**

Para ver qué tan bueno era cada modelo se midieron tres cosas:

- Precisión: qué tan bien predijo los positivos.

- Exactitud: qué porcentaje total de aciertos tuvo.
- Sensibilidad: qué tan bien detectó los casos positivos reales.

## **Hallazgos y Datos Más Importantes**

### **Patrones encontrados**

Cuando se usaron precio y cantidad de personas que caben en el alojamiento (accommodates), los resultados fueron aceptables, pero no siempre muy buenos. En algunos modelos, sobre todo cuando se quiso predecir si un alojamiento podía ser reservado al instante (instant\_bookable\_bin), la precisión fue muy baja o incluso cero. Esto muestra que algunas variables no son suficientes solas para hacer buenas predicciones.

### **Calidad de los modelos**

Algunos modelos lograron alta exactitud (más del 70%), pero baja precisión o baja sensibilidad. Esto significa que el modelo acertó en general, pero falló cuando intentaba predecir casos especiales, como distinguir superhosts o tipo de habitación.

### **Importancia de preparar los datos**

Fue muy importante convertir las variables en 0 y 1, porque si no, el modelo no podría entenderlas. También fue muy necesario limpiar el precio, porque tener signos de dólar causaría errores en el análisis.

## **Conclusión**

Este análisis ayudó a entender cómo se relacionan algunas características de los alojamientos en Belice. Aunque algunos modelos no lograron resultados muy buenos, el proceso completo de limpiar datos, convertir variables, entrenar modelos y medir resultados se hizo de forma correcta y organizada.

Se reafirma que, en análisis de datos reales, preparar bien la información es casi tan importante como usar un buen modelo. Tener los datos en el formato adecuado hace que todo el análisis sea posible y que los resultados sean más confiables.