



Tecnológico de Monterrey

Campus Puebla

José Manuel Flores Pérez | A01733838

Actividad 4.1: Reporte

Alfredo García Suárez

27 de abril de 2025

En el código se realizó un análisis de regresión logística aplicando 10 diferentes combinaciones de variables, el objetivo fue evaluar el comportamiento de diversas variables dependientes relacionadas a características de hospedajes, basándonos en indicadores como tasa de respuesta, disponibilidad, aceptación de solicitudes, entre otros.

Se inició el proceso realizando una limpieza de datos, la cual incluyó la conversión de columnas como `price`, `bathrooms_text`, `host_acceptance_rate` y `host_response_rate` a formatos numéricos adecuados; además, las variables categóricas `host_is_superhost` e `instant_bookable`, originalmente codificadas como texto ("t" y "f"), fueron convertidas a valores binarios (1 y 0) para permitir su análisis en los modelos de regresión logística.

Posteriormente, se definieron 10 casos de análisis, cada uno con su propia combinación de variables independientes y dependientes, para garantizar la correcta aplicación de los modelos, se implementó una conversión a formato dicotómico para aquellas variables dependientes que no fueran originalmente binarias, esto permitió que las métricas de evaluación se calcularán de forma adecuada en cada caso.

Cada uno de los modelos se construyó utilizando `train_test_split` para dividir los datos en entrenamiento y prueba, `StandardScaler` para normalizar las variables independientes, y `LogisticRegression` para realizar el ajuste del modelo, al finalizar el proceso, se evaluaron las predicciones generadas utilizando las métricas estándar: precisión, exactitud y sensibilidad.

El análisis permitió observar diferencias en el desempeño de los modelos dependiendo de las combinaciones de variables seleccionadas, algunas combinaciones presentaron mejores niveles de precisión y sensibilidad, indicando que ciertas variables, como `availability_365`, `reviews_per_month` o `number_of_reviews`, aportan información más relevante para predecir los comportamientos analizados.

En conclusión, se comprobó que algunas variables como `availability_365`, `reviews_per_month`, y `number_of_reviews` aporta un mayor valor predictivo en los modelos analizados; no obstante, la baja sensibilidad en algunos casos refleja la necesidad de seguir optimizando los modelos.