

## INFORME FUTUREKEEPCODERS

### 1. Objetivo del análisis

El objetivo principal del análisis ha sido conocer el precio medio del alquiler en Madrid, así como, desarrollar un modelo predictivo utilizando datos de Airbnb para estimar el precio de los alojamientos en función de variables como el vecindario, el tipo de habitación y el número de reseñas.

### 2. Suposiciones Iniciales

#### 2.1. Suposiciones formuladas

- El precio de los alojamientos puede ser predicho utilizando el vecindario, el tipo de habitación y el número de reseñas.
- Los datos de Airbnb son representativos y suficientemente completos para construir un modelo predictivo robusto.
- No hay sesgos significativos en los datos que puedan afectar la precisión del modelo.

#### 2.2. Validación de suposiciones

- El vecindario y el tipo de habitación como predictores: Estas variables han demostrado ser válidas ya que, tienen un impacto significativo en el precio de los alojamientos, lo cual fue corroborado por los coeficientes del modelo y sus valores p.
- Representatividad y completitud de los datos: Esta suposición resultó ser parcialmente válida. Algunos vecindarios y tipos de habitaciones tenían menos datos disponibles, lo cual podría haber introducido cierto sesgo en el modelo.
- Ausencia de sesgos significativos: Se identificaron algunos sesgos potenciales relacionados con la estacionalidad y la variabilidad en la cantidad de reseñas, lo que sugiere que los datos podrían no ser completamente representativos de todas las condiciones posibles.

### 3. Métricas seleccionadas

#### 3.1. Métricas utilizadas

Precisión y exhaustividad (Precisión y recall): Para evaluar la calidad de las predicciones en el caso de un modelo de clasificación binaria.

#### 3.2. Evaluación de métricas

Precisión y exhaustividad: Estas métricas fueron relevantes para el modelo de clasificación binaria, ayudando a evaluar el rendimiento del modelo en términos de verdaderos positivos y falsos negativos.

### 4. Lecciones aprendidas

#### 4.1. Aspectos positivos

- Selección de variables: La elección de variables como el vecindario y el tipo de habitación fue acertada, ya que estas variables demostraron ser significativas en el modelo predictivo.
- Uso de glm para regresión y clasificación: La utilización de glm fue adecuada tanto para modelos de regresión lineal como para modelos de clasificación binaria, proporcionando flexibilidad en el análisis.

#### 4.2. Áreas de mejora

- Tratamiento de datos ausentes y sesgos: Se podría mejorar el tratamiento de datos ausentes y abordar posibles sesgos en los datos de entrenamiento.
- Métricas adicionales: Incorporar métricas adicionales como MAE para tener una evaluación más completa de validación cruzada para asegurar la robustez del modelo y su capacidad de generalización.

### 5. Conclusiones y "Lessons Learned"

#### 5.1. Conclusiones

- Impacto de variables: El vecindario y el tipo de habitación tienen un impacto significativo en el precio de los alojamientos de Airbnb.
- Precisión del modelo: El modelo de regresión lineal mostró un buen ajuste y las predicciones fueron razonablemente precisas, aunque hay margen para mejorar el manejo de datos y la selección de métricas.

#### 5.2. "Lessons Learned"

- Importancia de la limpieza de datos: La calidad de los datos es fundamental para el rendimiento del modelo. Mejorar la limpieza y el preprocesamiento de datos puede llevar a modelos más precisos.
- Evaluación de métricas: Es esencial utilizar una combinación de métricas para evaluar los modelos de manera integral.
- Flexibilidad en modelos: La flexibilidad de glm permite abordar diferentes tipos de problemas (regresión y clasificación con el mismo enfoque metodológico).
- Validación rigurosa: La validación cruzada es crucial para asegurar que los resultados del modelo sean generalizables y no están sobreajustados a los datos de entrenamiento.